



HAL
open science

Aide à la décision en usinage basée sur des règles métier et apprentissages non supervisés

Zhiqiang Wang

► **To cite this version:**

Zhiqiang Wang. Aide à la décision en usinage basée sur des règles métier et apprentissages non supervisés. Sciences de l'ingénieur [physics]. Université de Nantes, 2020. Français. ⟨NNT: 2020NANT4038⟩. ⟨tel-03900893⟩

HAL Id: tel-03900893

<https://hal.science/tel-03900893v1>

Submitted on 19 Dec 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

THESE DE DOCTORAT DE

L'UNIVERSITE DE NANTES

ECOLE DOCTORALE N° 602

Sciences pour l'Ingénieur

Spécialité : « *Génie Mécanique et Productique* »

Par

«Zhiqiang WANG»

«Aide à la décision en usinage basée sur des règles métier et apprentissages non supervisés»

Thèse présentée et soutenue à IUT de Nantes, le 14/12/2020.

Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N - UMR CNRS 6004)

Rapporteurs avant soutenance :

Nabil ANWER Pr., LURPA - Université Paris Saclay
Julien LE DUIGOU Dr. (HDR), laboratoire Roberval - Université de Technologie de Compiègne

Composition du Jury :

Président : Guénaël GERMAIN Pr., laboratoire LAMPA – Arts et Métiers ParisTech Angers
Examineurs : Guénaël GERMAIN Pr., laboratoire LAMPA – Arts et Métiers ParisTech Angers
Mathieu RITOU Dr., LS2N - Université de Nantes
Dir. de thèse : Benoît FURET Pr., LS2N - Université de Nantes
Co-dir. de thèse : Catherine DA CUNHA Pr., LS2N - Ecole Centrale de Nantes

Invité(s)

Virginie CHARBONNIER Responsable projets R&T usinage Airbus Operations

N° d'ordre:

*	*	*
---	---	---

LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES (LS2N, UMR CNRS 6004)
UNIVERSITÉ DE NANTES



Projet de Thèse de Doctorat

en

Génie Mécanique

par

Zhiqiang WANG

Doctorat délivré par l'Université de Nantes

Decision-aid in machining based on business rules and unsupervised machine learning

Aide à la décision en usinage basée sur des règles métier et apprentissages
non supervisés

Présentée publiquement le 14 Décembre 2020 devant le jury d'examen :

Rapporteurs

Pr. Nabil ANWER

LURPA - Université Paris Saclay

Dr. (HDR) Julien LE DUIGOU

laboratoire Roberval - Université de Technologie de Compiègne

Examineurs

Pr. Guénaël GERMAIN

laboratoire LAMPA - Arts et Métiers ParisTech Angers

Directeurs de thèse

Pr. Benoît FURET

LS2N - Université de Nantes

Pr. Catherine DA CUNHA

LS2N - Ecole Centrale de Nantes

Encadrant de thèse

Dr. Mathieu RITOU

LS2N - Université de Nantes

Invitée

Mme. Virginie CHARBONNIER

Responsable projets R&T usinage Airbus Operations

Abstract :

In the general context of Industry 4.0, large volumes of manufacturing data are available on instrumented machine-tools. They are interesting to exploit not only to improve machine-tool performances but also to support the decision making for the operational management. This thesis aims at proposing a decision-aid system for intelligent and connected machine-tools through Data mining.

The first step in a data mining approach is the selection of relevant data. Raw data must, therefore, be classified into different groups of contexts. This thesis proposes a contextual classification procedure in machining based on unsupervised machine learning by Gaussian mixture model. The optimal number of classes is determined by the silhouette method based on the Bayesian information criterion. This method is validated on real data from four different machine-tools in the aerospace industry. Manual data mining and k-fold cross validation confirm that the proposed method provides good contextual classification results. Subsequently, different types of operation and different states of tool engagement can be identified based on the contextual classification.

Based on this contextual classification information, different machining incidents can be detected in real-time. They include chatter, tool breakage and excessive vibration. This thesis introduces a set of business rule for chatter detection. It improves the classification on harmonic / non-harmonic frequencies, enabling a more accurate computation of the chatter criterion Nh , as well as the deciphering of the tool engagement context when chatter occurs. This thesis also proposes a set of business rule combined with two criteria (Ub when *tool is not cutting* and Ub' when *tool is cutting*) for the detection of early and severe cutter breakage. Moreover, the business rule combined with two vibration criteria ($Arms$ and $Vrms$) for excessive vibration detection is proposed in this thesis. These three business rules are validated on real data from one machine-tool above.

The knowledge-based and multi-level approach aggregates the context classification results and the incidents detected into Smart Data. Then, the relevant and appropriate Key Performance Indicators (KPIs) can be proposed based on these Smart Data to support decision making for the operational management. This thesis proposes new aggregation operators and new relevant KPIs for the different department needs. All of these KPIs will become the core of the decision-aid system. Lastly, the impact of contextual classification error on KPI is investigated.

Keywords:

Industry 4.0, machine-tool, Data mining, unsupervised machine learning, GMM, contextual classification, incidents detection, Smart Data, KPIs.

Résumé :

Dans le contexte général de l'industrie 4.0, une entreprise de fabrication moderne dispose de nombreuses données numériques qui pourraient être utilisées pour rendre les machines-outils plus intelligentes et faciliter la prise de décision en matière de gestion opérationnelle. Cette thèse vise à proposer un système d'aide à la décision pour les machines-outils intelligentes et connectées par l'exploitation des données.

L'une des premières étapes en fouille de données est la sélection précise de données pertinentes. Pour ce faire, les données brutes doivent être classées dans différents groupes de contexte. Cette thèse propose un algorithme d'apprentissage automatique non-supervisé, par mélanges gaussiens (GMM), pour la classification contextuelle. Le nombre de clusters optimal est déterminé par le critère BIC (Bayesian Information Criterion). Cette méthode est validée sur des données réelles provenant de quatre machines-outils différentes dans l'industrie aéronautique. Les vérifications par fouilles manuelles et la validation par k-fold montrent ensuite que la méthode GMM permet d'obtenir de bons résultats de classification contextuelle. Ensuite, différents types d'usinage et différents états d'engagement de l'outil peuvent être identifiés sur la base de cette classification contextuelle.

Basé sur les informations contextuelles, différents incidents d'usinage peuvent être détectés en temps réel. Il s'agit notamment de broutement, de bris d'outil et de sur-vibration. Cette thèse présente un ensemble de règles métiers pour la détection du broutement. Elle améliore la classification sur les fréquences harmoniques / non-harmoniques, permettant un calcul plus précis du critère de broutement Nh , ainsi que décrypter le contexte d'usinage lorsque le broutement s'est produit. Cette thèse propose également un ensemble de règles métiers de gestion combinées à deux critères (Ub lorsque l'outil n'usine pas et Ub' lorsque l'outil usine) pour la détection précoce ou temps-réel des bris d'outil. De plus, une règle métier combinant deux critères de vibration ($Arms$ et $Vrms$) pour la détection de sur-vibration est proposée dans cette thèse. Ces trois règles métiers sont validées sur des données réelles provenant d'une machine-outil ci-dessus.

Suite à des agrégations en Smart Data basées sur informations contextuelles et les incidents détectés, des KPIs (Key Performance Indicators) peuvent être évalués afin de soutenir la prise de décision pour la gestion opérationnelle. Cette thèse propose de nouveaux opérateurs d'agrégation et de nouveaux KPIs pertinents pour les besoins spécifiques des différents services d'une entreprise. Tous ces KPIs deviendront le coeur du système d'aide à la décision. Enfin, l'impact d'erreur de classification contextuelle sur les KPI est étudié.

Mots clés :

Industrie 4.0, machines-outils, fouilles de données, apprentissage non-supervisé, GMM, classification contextuelle, détection d'incidents, Smart Data, KPIs

Remerciement

Ce travail de thèse a été réalisé au sein du laboratoire de LS2N (Laboratoire des Sciences du Numérique de Nantes) financé par le projet ANR SmartEmma. J'adresse mes premiers remerciements et ma reconnaissance à mes directeurs de thèse: Benoît Furet, Pr. à l'Université de Nantes; et Catherine Da Cunha, Pr. à l'Ecole Centrale de Nantes. Je tiens à remercier également mon encadrant de thèse, Mathieu Ritou, Dr. à l'Université de Nantes. Je vous remercie pour avoir dirigé et aidé à construire mon projet de thèse au cours de ces trois ans avec vos sens pédagogiques. Merci pour la confiance que vous m'avez accordée, la patience que vous m'avez donnée. J'ai beaucoup apprécié travailler avec vous. En particulier, merci à Benoît de m'avoir accueilli au sein du laboratoire LS2N à Nantes pendant les trois ans de thèse. Merci à Catherine de m'avoir donné beaucoup de conseils sur la rédaction des articles scientifiques. Je voudrais remercier à Mathieu, qui m'avait appris beaucoup de connaissances du process sur l'usinage.

Je tiens à remercier le jury de cette thèse d'avoir assisté à ma soutenance, en particulier le Pr. Nabil Anwer et Dr. (HDR) Julien Le Duigou, qui ont aimablement accepté de rapporter ma thèse. Merci beaucoup pour vos commentaires et remarques constructifs qui m'aident à améliorer la qualité de la thèse. Je voudrais remercier Pr. Guénaël Germain qui a bien dirigé la soutenance de thèse.

Mes grandes remerciements iront à nos partenaires industriels, i.e., Europe Technologie, Mécachrome et Airbus Nantes, qui nous fournissent les données numériques d'usinage en grande volume.

Mes remerciements iront également à l'endroit de mes chers collègues de LS2N, avec qui j'ai pu travailler dans une très bonne ambiance. Merci à mes collègues doctorants (docteurs) dans le projet ANR SmartEmma: Oussama, Zakaria, Asma, Farouk. Merci également mes collègues de LS2N à Nantes de m'accueillir pendant les trois ans de thèse.

Enfin, je voudrais exprimer ma gratitude à ma famille. Je tiens à remercier mes parents Jindun, Shengying et ma soeur Yiran pour leur amour et leurs encouragements. Je tiens à remercier ma femme, Dr. Zhu, pour son soutien. Bien sûr, je n'oublie pas de remercier ma fille, Fabienne, et mon fils, Guillaume qui illuminent ma vie.

Z. WANG

Bouguenais - 18 Décembre 2020.

Contents

1	Introduction	3
1.1	Industrial context	4
1.1.1	Data Mining system in aeronautic machining	4
1.1.2	The ANR SmartEmma Project	5
1.2	State of the art	8
1.2.1	Machining monitoring	8
1.2.2	Data mining - Machine learning in machining monitoring	11
1.2.3	Key performances indicators for smart machining	15
1.2.4	Decision-aid system	15
1.3	Scientific and technological barriers	17
1.4	Structure of this thesis	17
2	Contextual classification for smart machining based on unsupervised machine learning	19
2.1	Motivation of the contextual classification	21
2.2	Comparison between Kmeans and GMM	23
2.2.1	Introduction	23
2.2.2	K-means	24
2.2.3	GMM classification	26
2.2.4	Conclusion	28
2.3	Gaussian mixture model	29
2.4	Determination of cluster number by Bayesian Information Criterion (BIC)	30
2.4.1	BIC on the variation of spindle speed	31
2.4.2	BIC on the variation of feedrate	31
2.4.3	BIC on vibrations and power	31
2.5	Business rules for contextual classification	33
2.5.1	Contextual classification of machine-tool	33
2.5.2	Contextual classification per type of machining operation	35

2.5.3	Contextual classification per type of tool engagement	35
2.6	Contextual classification of the material removal	36
2.6.1	Modeling and classification for machine-tool #2	36
2.6.2	Modeling and classification for machine-tool #3	39
2.7	Method validation	41
2.7.1	Accuracy	41
2.7.2	Sensitivity of the method	43
2.8	Conclusion	44
3	Detection of machining incidents based on GMM and business rules	47
3.1	Data presentation and cleaning	48
3.2	Improvement of chatter identification	53
3.2.1	Classification of harmonic frequency	53
3.2.2	Chatter / non Chatter	57
3.2.3	Test on a new database	59
3.3	Tool breakage	61
3.3.1	Business rule for detection of tool breakage	61
3.3.2	Application	63
3.4	Excessive vibrations	67
3.5	Conclusion	70
4	Proposition and management of KPIs	71
4.1	Knowledge Discovery in Database (KDD) procedure	74
4.1.1	Management and aggregation	75
4.1.2	Proposition of new aggregation operators	76
4.2	KPIs for production organization	77
4.2.1	Advanced Overall Equipment Effectiveness (OEE)	77
4.2.2	Workpieces production	79
4.3	Traceability of Workpiece quality	80
4.3.1	Respect of cutting specifications	80

4.3.2	Detection of incidents	84
4.4	KPIs for production monitoring	86
4.4.1	Traceability of duration of machining operatiосn	86
4.4.2	Traceability of cutting aspects	87
4.4.3	Human factor during production	88
4.4.4	Incidents detected during production	89
4.5	KPIs for tool performance	89
4.5.1	Distribution of chatter per tool	89
4.5.2	Distribution of excessive vibration per tool	91
4.5.3	Tool use duration before tool breakage	92
4.6	KPIs for program performance	92
4.6.1	Distribution of chatter per program	92
4.6.2	Distribution of excessive vibration per program	95
4.6.3	Productivity of CAM programs	95
4.6.4	Slow idle motions	96
4.7	KPIs for monitoring of machine-tool condition	97
4.7.1	Spindle condition monitoring	97
4.7.2	Incidents detection	98
4.8	The impact of the classification errors on the evaluation of KPIs	98
4.9	Conclusion	100
5	Conclusions and Perspectives	103
5.1	Conclusions	104
5.2	Perspectives	107
	References	109

List of Figures

1.1	Specific parts manufactured in the aeronautics field (large dimensions parts, complex parts, thin-walled parts and Fan disc in Titanium)	5
1.2	Chronology of the projects carried out in laboratory LS2N on machining monitoring and smart machining	6
1.3	ANR SmartEmma Consortium	7
1.4	Framework of ANR SmartEmma project	7
1.5	Stability lobe diagram [Quintana and Ciurana, 2011]	10
1.6	Graph representing the different stages of knowledge extraction from the data [Fayyad et al., 1996]	12
1.7	ANN architecture for prediction tool wear in titanium machining in [Krishnakumar et al., 2015]	13
1.8	Architecture of convolutional neural network model in [Wang et al., 2018]	13
1.9	Flow chart of data processing as well as the structure of thesis	18
2.1	Data mining process for machining	21
2.2	Example of the false alarm	22
2.3	Process of classification according to the feedrate V_f and ΔV_f	23
2.4	Process of classification by K-means [Jacques, 2001]	24
2.5	Classification of ΔV_f into 3 classes by K-means	25
2.6	Classification of V_f by K-means during 30 seconds	25
2.7	Density of probability of ΔV_f for one day modeled by GMM (2 Gaussians)	27
2.8	Classification of V_f by GMM during 30 seconds	28
2.9	BIC on ΔN issued from the 4 machine-tools	31
2.10	BIC on ΔV_f issued from the 4 machine-tools	32
2.11	BIC on $Arms$ (when N constant et $V_f > 0$) issued from the 4 machine-tools	32
2.12	BIC on P (when N constant et $V_f > 0$) issued from the 4 machine-tools	33
2.13	Distribution of $Arms$ and P (when N constant and $V_f > 0$) modeled by GMM for machine-tool #2	37
2.14	Classifications of material removal for machine-tool #2	38

2.15	Classifications of different states of tool engagement for machine-tool #2	39
2.16	Distribution of Arms and P (when N constant and Vf>0) modeled by GMM for machine-tool #3	40
2.17	Classifications of material removal for machine-tool #3	41
2.18	The new thresholds of <i>Arms</i> and <i>P</i> obtained in the training sets as well as these thresholds obtained in validation sets for (a) machine-tool #2 and (b) machine-tool #3	44
3.1	Emmatools for process monitoring and data collection	48
3.2	The CRISP-DM life cycle[Wirth and Hipp, 2000]	50
3.3	Accelerometer Y in the rear bearing is missing for the machine-tool #1	50
3.4	Noise from the preparation program 'tool change' for the machine-tool #1	51
3.5	Noise from the lubrication startup (in the period <i>post spindle startup</i>) in the principal machining program for the machine-tool #1	52
3.6	Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for machine-tool #2	54
3.7	Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for machine-tool #2	55
3.8	Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for a) machine-tool #1 and for b) machine-tool #3	56
3.9	Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for ID Tool 6 in machine-tool #4	57
3.10	Chatter in Machine-tool #2 for a) punctual and for b) continuous	58
3.11	Modeling of <i>Nh</i> by GMM (composed of a normal population in green and of a chatter population in black as well as of a severe chatter population in red) and learning of the detection threshold T_{Nh} for chatter	59
3.12	Visualization of chatter criterion <i>Nh</i> (in green) with the classification of different states of tool engagement	60
3.13	Order tracking of the frequency spectrum for ID Tool 50 and for ID prog 99 in the 15 th day for Machine-tool #2	60
3.14	Modeling of <i>Ub</i> distribution by GMM (composed of a normal population in green and of a tool breakage population in red) for Machine-tool #2 and learning of the detection threshold T_{Ub} for tool breakage	62
3.15	Raw data of vibration level <i>Arms</i> during the 136 th day of spindle #1 production for machine-tool #2	63

3.16	Visualization of Ub criterion (in purple) for ID Tool 36, the index of tool instance (in green) and the Boolean for tool breakage detection (in red)	64
3.17	Operational context when tool breakage occurs for ID Tool 36 in ID prog 154 in the 136 th day in the machine-tool #2	64
3.18	Occurrences of tool breakage via ID Tool during 2 spindle lifetimes for machine-tool #2 based on the Old and New tool breakage detections	65
3.19	Operational context when tool breakage occurs for ID Tool 22 in ID prog 108 in the 18 th day during the spindle #1 manufacturing in the machine-tool #2	66
3.20	Modeling of a) $Arms$ and b) $Vrms$ distributions by GMM (with 3 populations: normal population when <i>tool is not cutting</i> , normal population when <i>tool is cutting</i> as well as the excessive vibration) for Machine-tool #2 and learning of the detection threshold T'_{Arms} and T_{Vrms} for the detection of excessive vibration	68
3.21	thresholds recommended by the ISO 10816 standard [ISO 10816, 2009]	69
3.22	Excessive vibration for the ID Tool 36 during the ID prog 249 in the 188 th day for machine-tool #2	69
4.1	Decision-aid framework for intelligent and connected machines-tools	73
4.2	Knowledge-based multi-level data aggregation [Ritou et al., 2019]	74
4.3	Data organization and different potential stages of aggregation	75
4.4	Machine-tool advanced OEE for 17 days database in the machine-tool #1	78
4.5	Occurrence of machined workpieces during 17 days on the machine-tool #1	79
4.6	Time when potentiometers differs from 100%, per production order NumOF on the machine-tool #1	80
4.7	Visualization of the period when $Pot(Vf)$ is not 100% in NumOF 58 on the machine-tool #1	81
4.8	Visualization of the period when $Pot(N)$ is not 100% in NumOF 58 on the machine-tool #1	82
4.9	Time when the spindle speed N didn't respect the recommend speed (except $Pot(N)$ not 100%) for all tools per production order NumOF on the machine-tool #1	82
4.10	The time $T[fz < T_{fz} \mid eng.cst]$ per production order NumOF on the machine-tool #1	83
4.11	Visualization of the period during which fz is insufficient when the ID Tool 19 is <i>milling with $2\frac{1}{2}$ axes</i> in NumOF 31 on the machine-tool #1	83
4.12	The time $T[QL < T_{QL} \mid cutting]$ per production order NumOF on the machine-tool #1	84
4.13	Visualization of the period during which QL is insufficient when the ID Tool 47 is cutting in NumOF 51 on the machine-tool #1	84

4.14	Chatter time $T[Nh > T_{Nh} cutting]$ as well as occurrences of tool breakage per production order NumOF on the machine-tool #1	85
4.15	Visualization of the chatter when the ID Tool 6 is cutting in NumOF 31 on the machine-tool #1	85
4.16	Summary of the traceability of workpiece quality per Production Order NumOF on the machine-tool #1	86
4.17	Distribution of tool use time in spindle via workpiece during 10 days' production in the machine-tool #4	87
4.18	P_{cut} as well as the Tool Instance (in green) for ID Tool 3 during 10 days in the machine-tool #4	88
4.19	Mean P and the time $T[V_{rms} > 8 \text{ mm/s}]$ during cutting per NumOF during 17 days' production in the machine-tool #1	89
4.20	Distribution of chatter per kind of machining operations for the 10 most problematic tools (when spindle #1 is in good condition in machine-tool #2).	90
4.21	Distribution of chatter per kind of machining operations for the 10 most problematic tools (with the end of life of spindle #1 in machine-tool #2)	91
4.22	Distribution of excessive vibration per tool (during the whole spindle #1 lifetime of machine-tool #2)	92
4.23	Tool use duration before tool breakage by tool lifetime in % (in blue) for the problematic ID Tools during the spindle #1 and #2 in the machine-tool #2	93
4.24	Distribution of chatter per kind of machining operations for the 10 most problematic programs (when the spindle is in good condition for machine-tool #2)	93
4.25	Distribution of chatter per kind of machining operations for the 10 most problematic programs (with the end of life of spindle for machine-tool #2)	94
4.26	Distribution of excessive vibration per program (during the whole spindle #1 life time's production in the machine-tool #2)	95
4.27	MRR for each ID Program when the ID Tool 5 is used in machining as well as the mean MRR of ID Tool 5 during the whole spindle #1 lifetime in the machine-tool #2	96
4.28	Mean Vf when the <i>tool is not cutting</i> per ID Program for the machine-tool #2	97
4.29	Chatter time and the occurrences of tool breakage as well as the spindle condition criterion BPFO during the whole spindle #1 lifetime in the machine-tool #2	98
4.30	Machine-tool OEE during 4 hours in the machine-tool #2 based on 3 different contextual classifications: a) Ground truth; and predicted classes obtained by b) New; c) Old contextual classification	100

List of Tables

1.1	Advantages and disadvantages of each algorithm for three modes of machine learning .	16
2.1	Confusion matrix of 4 hours data for the classification of V_f by K-means	26
2.2	Confusion matrix of 4 hours data for the classification of V_f by GMM	28
2.3	Classification of the different states of a machine-tool	34
2.4	Classification of the different operations when the tool is cutting	35
2.5	Classification of the different states of the tool engagement	36
2.6	Confusion matrix for 4 hours data from the machine-tool #2	42
2.7	Confusion matrix for 4 hours data from the machine-tool #3	42
2.8	Confusion matrix for the different tool engagement states during 4h on the machine-tool #2	42
3.1	Frequency theoretical of $BPFO$ and $BPFI$ as well as their tolerances [Godreau, 2017]	54
4.1	Chatter duration (PC in green + CC in yellow) per machining operation per kind of tool engagement for ID Tool 23 in machine-tool #2.	90
4.2	Chatter duration (PC in green + CC in yellow) per machining operation per kind of tool engagement for ID program 50, and for ID Tool 23 in machine-tool #2	94
4.3	Confusion matrix for 4 hours data from the machine-tool #2 based on 'Old contextual classification'	99

Acronyms and Notations

ACC Accuracy of the Classification.

AE acoustic emission.

ANN Artificial Neural Networks.

ANR Agence National de la Recherche.

AO Abnormality Operator.

BIC Bayesian Information Criterion.

CAD Computer-Aided Design.

CAM Computer-Aided Manufacturing.

CNC Computer Numerical Control.

CNN Convolutional Neural Networks.

CO Criticality Operator.

CPS Cyper Physical System.

CRISP-DM CRoss Industry Standard Process for Data Mining.

DM Data Mining.

EM Expectation-Maximization.

FFT Fast Fourier transform.

FN False Negatives.

FoF Factory of the Future.

FP False Positives.

FUI Le Fonds Unique Interministériel.

GEV Generalized Extreme Value.

GMM Gaussian Mixture Model.

HSM High-Speed Machining.

IoT Internet of Things.

KDD Knowledge Discovery in Database.

KPIs Key Performance Indicators.

MLE Maximum Likelihood Estimation.

MRR Material Removal Rates.

MTBF Mean time Between Failures.

OEE Overall Equipment Effectiveness.

RA surface Roughness.

RMS Root Mean Square.

SLD Stability Lobe Diagram.

TCM Tool Condition Monitoring.

TN True Negatives.

TP True Positives.

Notation	Description	Units
N	spindle speed	RPM
Vf	Machine-tool feedrate	m/min
Z	number of tooth of the cutting tool	N/A
fz	feed per tooth	mm/tooth
$Arms$	root mean square (RMS) of vibration acceleration value	m/s^2
P	spindle power	kW
Pot(N)	Potentiometer for spindle speed	%
Pot(Vf)	Potentiometer for machine-tool feedrate	%
Nh	chatter detection criterion	m/s^2
Ub	early tool breakage detection criterion	m/s^2
Ub'	large tool breakage detection criterion	m/s^2
$Apic$	maximum of vibration acceleration value over a given period	m/s^2
$Vrms$	vibration velocity in root mean square	mm/s
$Vpic$	maximum of vibration velocity	mm/s
f	frequency	Hz
f_b	spindle rotation frequency	Hz
f_h	harmonic frequency	Hz
f_{nh}	non-harmonic frequency	Hz
f_{BPFO}	frequency of the contribution at Ball Pass Frequency of Outer ring	Hz
f_{BPFI}	frequency of the contribution at Ball Pass Frequency of Inner ring	Hz
B	number of balls in the bearing	N/A
Φ	ball diameter	mm
D_m	mean diameter of the bearing	mm
α	angle of contact between the ring and the balls	$^\circ$
Δf	harmonic frequency range	Hz
$BPFO$	amplitude of the contribution at Ball Pass Frequency of Outer ring	m/s^2
$BPFI$	amplitude of the contribution at Ball Pass Frequency of Inner ring	m/s^2
FTF	amplitude of the contribution at Fundamental Train Frequency	m/s^2
ΔN	spindle speed variation	tr/s^2
ΔVf	Machine-tool feedrate variation	mm/s^2
ΔP	spindle power variation	kW/s
$T_{\Delta N}$	threshold of the spindle speed variation	tr/s^2
$T_{\Delta Vf}$	threshold of the Machine-tool feedrate variation	mm/s^2
$T_{\Delta P}$	threshold of the spindle power variation	kW/s
T_{Arms}	threshold of the Arms to detect tool cutting or not	m/s^2
T'_{Arms}	threshold of the Arms to detect the excessive vibration	m/s^2
T_{Vrms}	threshold of the $Vrms$ to detect the excessive vibration	mm/s
T_{Nh}	threshold of Nh	m/s^2
T_{Ub}	threshold of Ub	m/s^2
T'_{Ub}	threshold of Ub'	m/s^2
P_{idle}	spindle power corresponding to idle rotation	kW
$P_{finishing}$	threshold of spindle power to distinguish rough milling and finishing	kW
MRR	material removal rate	cm^3/min
D	diameter of the cutting tool	mm
ap	depth of cut	mm
ae	width of cut	mm

Notation	Description	Units
X	absolute position of machine-tool in X axes	mm
Y	absolute position of machine-tool in Y axes	mm
Z	absolute position of machine-tool in Z axes	mm
A	absolute position of machine-tool in A axes	°
C	absolute position of machine-tool in C axes	°
ΔX	absolute position variation of machine-tool in X axes	mm/s
ΔY	absolute position variation of machine-tool in Y axes	mm/s
ΔZ	absolute position variation of machine-tool in Z axes	mm/s
ΔA	absolute position variation of machine-tool in A axes	°/s
ΔC	absolute position variation of machine-tool in C axes	°/s

Preamble

In the history of the industry and of the economy, it is possible to count several successive phases, interspersed with transitions, the industrial revolutions. Generally, each of these revolutions has been made possible by key scientific advances. With this principle, it is possible to separate the industrial history into 4 periods.

The first industrial revolution was triggered by numerous advances in the fields of technology and thermodynamics. One of the emblematic characters of that time was James Watt who, in the 18th century, made great improvements to the steam engine. It thus allowed for the mechanization of work and the development of the routes for train. This first industrial revolution led major changes throughout society and it transformed the production of an organization in workshops to a manufacturing organization. Because of the increase need for processed metals, metallurgy also experienced a great development during this period.

The second industrial revolution, over 40 years (1870 to 1910), was marked by a doubling of world production. The main reason for this explosion was the control of electromagnetism and then of the engine and the electric alternator. This made possible to go from a factory where (steam) energy is grouped together on a single power shaft in the workshop to a factory where energy can be distributed. It thus rationalized the organization of the workstations and energy management. It was at this time that Taylorism and then Fordism made their appearances in the mass production.

The third industrial revolution, also known as the "computer revolution", was in the 1970s. It is marked by the appearance of microprocessors and, later, the Internet. In industrial terms, this revolution enables the robotization of production lines and the creation of Computerized Numerical Control (CNC) machine-tools. Another important aspect which shocked the organization of companies was the appearance of computers. It enabled the Computer-Aided Design (CAD), the Computer-Aided Manufacturing (CAM) etc.

The fourth industrial revolution is currently taking place. It started in Germany around 2013. Its growth is linked to the emergence of the digital sciences and the Artificial Intelligence (AI). The organization of new companies based on these systems, which is called "smart factories", is defined by the interconnection between the machines and the services of the company. It is realized through the instrumentation of manufacturing processes to automate manufacturing processes and through the increase of the information flow between companies, their subcontractors and the customers. The objectives of this change are production flexibility and the optimization, automation of production.

This new industrial revolution creates research needs in many fields, specially in the aeronautic domain. This thesis is in the global context of the fourth industrial revolution. It aims at proposing a decision-aid system for intelligent and connected machines-tools in the aeronautic industry.

The first chapter will specify the industrial context of this thesis, present the state of the art in the corresponding domains, then establish the scientific and technology barriers and the objective of the thesis. Finally, the structure of the manuscript will be presented.

Based on the flow chart of data processing as well as the structure of the thesis in the first chapter, the second chapter introduces a set of business rules for the contextual classification of data based on unsupervised machine learning.

The third chapter proposes the business rules for detecting different incidents (chatter, tool breakage, and excessive vibration) based on unsupervised machine learning and the deciphering of the operational context.

The fourth chapter proposes and computes the new Key Performance Indicators (KPIs) for the different departments needs, based on the contextual classification and the detected incidents. Then, final chapter makes a conclusion and proposes perspectives for future research.

Introduction

Contents

1.1	Industrial context	4
1.1.1	Data Mining system in aeronautic machining	4
1.1.2	The ANR SmartEmma Project	5
1.2	State of the art	8
1.2.1	Machining monitoring	8
1.2.2	Data mining - Machine learning in machining monitoring	11
1.2.3	Key performances indicators for smart machining	15
1.2.4	Decision-aid system	15
1.3	Scientific and technological barriers	17
1.4	Structure of this thesis	17

The first chapter aims to specify the industrial context of the study presented throughout this thesis report. This research focuses on the proposal of a data mining system for intelligent and connected machine-tools in the aeronautics industry. This data mining system will finally support the decision making of operational management.

This chapter will be structured in four distinct parts, the first, as mentioned above, will establish the industrial context of the study. The second part, the state of the art, describes the progress already made and the ones that will have to be made to achieve our objectives. Building from the state of the art, the third part will explain the various scientific and technical barriers that will be addressed in the following chapters. The last part will present the contribution and the structure of this thesis.

1.1 Industrial context

1.1.1 Data Mining system in aeronautic machining

High-Speed Machining (HSM) has significantly increased spindle speed compared to conventional machining. It is especially the case in aeronautic machining, where high power (70 kW) and high-speed spindle (24000 RPM) are used for Aluminum alloys machining. Those aeronautic workpieces are high-value parts. For example, a spar of wing structure can be 12 m long; they are machined for several hours from a raw block and require high Material Removal Rates (MRR). In fact, the volume of the final part can represent only 10% of the raw volume.

Most of time, these aeronautic parts have complex shapes (adapted to aeronautical design) and very low final thicknesses (to limit their weight). Those characteristics make it hard to meet the accuracy requirement after the machining in machine-tools. For hard-to-cut materials such as Titanium or Nickel, machining is characterized by low speeds and high forces. The issues are different according to different situations. Specific parts manufactured in the aeronautics field (large dimensions parts, complex parts, thin-walled parts, and Fan disc in Titanium) can be found in Fig. 1.1.

Production systems are very flexible, especially in the aeronautics industry, where hundreds of tools machine several thousands of parts with different references on the same site. It is not easy for the Manufacturing Department to identify problematic programs or tools. The main problems are chatter (instability during machining), tool breakage and excessive vibrations. Therefore, a Data Mining (DM) system can help to protect the machine-tools and workpieces. In the general context of Industry 4.0, large volumes of manufacturing data are available on instrumented machine-tools. Digital engineering has increasingly taken a more important place in the strategic issues of mechanical engineering companies [Le Duigou et al., 2012]. It is interesting to exploit these large data sets through a data-mining system not only to improve machine-tool performances but also to support the decision making of operational management. This is the object of the 4th industrial revolution that will lead to the digital Factory of the Future (FoF).

The FoF is a significant leverage to increase the competitiveness of the manufacturing companies. In a FoF approach, all the process numerical data can be collected and must be efficiently used to improve the machines' perception and intelligence, for online and offline decision making and for long time learning. It leads to improved processes and increased productivity and quality. Finally, it results in improved competitiveness, which is crucial in the international context.

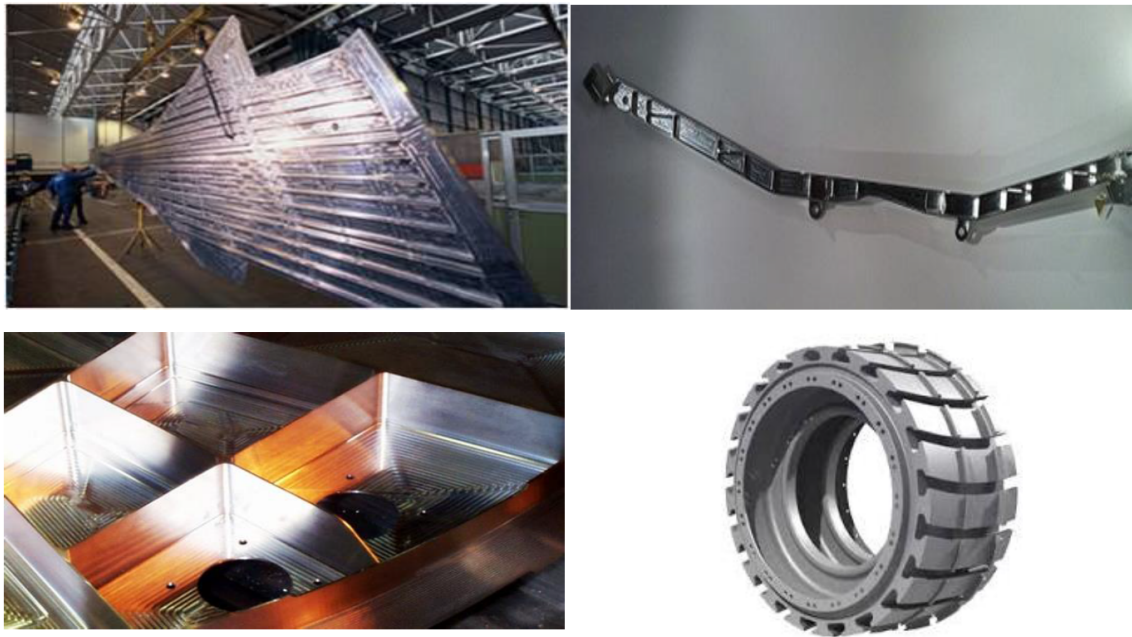


Figure 1.1 – Specific parts manufactured in the aeronautics field (large dimensions parts, complex parts, thin-walled parts and Fan disc in Titanium)

Industry 4.0 was originally initiated in Germany in 2011 [Lu, 2017]. Then USA has been driving industrial Internet, Cyber Physical System (CPS), and Advanced Manufacturing Partnership (AMP) program to advance future manufacturing [Wang and Wang, 2016]. Also there are e-Factory in Japan [Ridgway et al., 2013] and China Manufacturing 2025 [Li et al., 2016]. In France, the prospective workshop FuturProd was supported by Agence National de la Recherche (ANR). Academics and industrialists develop the FoF, in which context the ANR SmartEmma Project was proposed. This project was carried out by the cooperation between academics, i.e., the Laboratory of digital sciences of Nantes (LS2N) and industries : Europe Technologies, Mecachrome and Airbus. This thesis was conducted during this project.

1.1.2 The ANR SmartEmma Project

Several projects have been carried out in the LS2N laboratory to deal with different issues related to the monitoring of machining and the smart machining, as shown in Fig. 1.2. What should be mentioned is that, the SmartEmma project is integrated following the research project FUI QuaUsi (started in 2013) on continuous improvements for the workpiece quality in machining. And both projects were based on the tools developed in the project FUI UsineAE (started in 2008), which addressed machining monitoring and their applications.

These projects are related to each other and contribute to different aspects of machining. They bring the machining system closer to a smart level, which is exactly the objective of the industry 4.0. To better understand the SmartEmma project, it is necessary to take a look at these previous corresponding projects:

- European Idmap project (1998) firstly instrumented the spindle of the machine-tool.

- Then, the FCE project developed the methods for determining the cutting parameters taking into account the vibration problems linked to the dynamics of spindle and tools together.
- Afterwards, the FUI UsinAE project firstly developed the online monitoring system for spindle by the creation of criteria (SB_{BPFO} and SBN). During this project, the data collecting system Emmatools was first developed. It would be used in FUI QuaUsi project and ANR SmartEmma project.
- After these three projects, the FUI QuaUsi project was set up to work on machining part quality. The objective of this project is to reduce as much as possible the time spent on finishing and checking workpiece quality.
- Then, the ANR SmartEmma project started in 2016. The objective of the SmartEmma project is to develop innovative smart and connected machine-tool for High-Speed Machining (HSM). Research will be carried out to define Key Performance Indicators (KPI) and methods for database analysis, to improve HSM process efficiency and to ease the decision making of the operation management, in the context of the Factory of the Future.

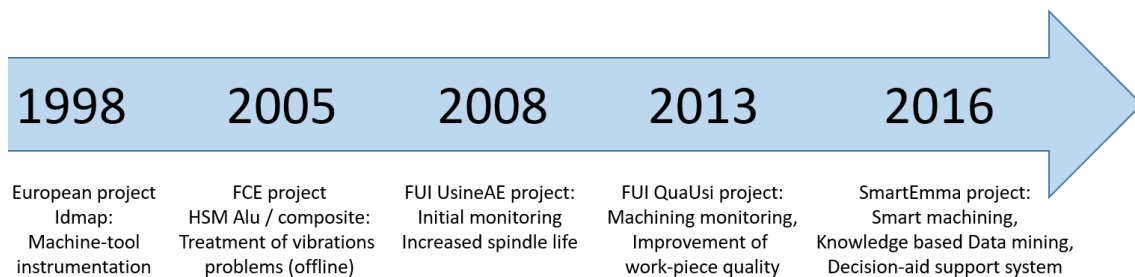


Figure 1.2 – Chronology of the projects carried out in laboratory LS2N on machining monitoring and smart machining

The device Emmatools collects the measurement data during machining and stores it in a database [De Castelbajac et al., 2014]. It collects data every tenth of a second, i.e. 10 Hz. The data comes from three information sources: the Computer Numerical Control (CNC) context information, the sensors integrated into the machine-tool, as well as the added vibrations sensors. By field bus, the CNC provides machining context data such as the tool name, the current program, machine, and spindle motions. The sensors integrated into the machine-tool collect the instantaneous power of the spindle, temperatures, as well as the vibrations from 4 accelerometers integrated into the spindle (radially at each bearing).

The ANR SmartEmma, ANR-16-CE10-0005, is a consortium of academics and industrials, Figure 1.3). The project is structured into four main parts to develop innovative smart and connected machine-tools. More precisely, the four main issues to be solved in this project are as following:

- Modeling and management of HSM process data. (This thesis)
- Research of Key Performance Indicators (KPIs) by Knowledge Discovery in Database (KDD) (This Thesis): this includes the Aluminum alloy cutting issue and the hard-to-cut materials issue (Titanium alloys).



Figure 1.3 – ANR SmartEmma Consortium

- Modeling and management of HSM knowledge. (Thesis of O. Meski)
- Decision-making algorithms for online and offline issues. (collaboration between the two previous theses and the research work of a Postdoc)

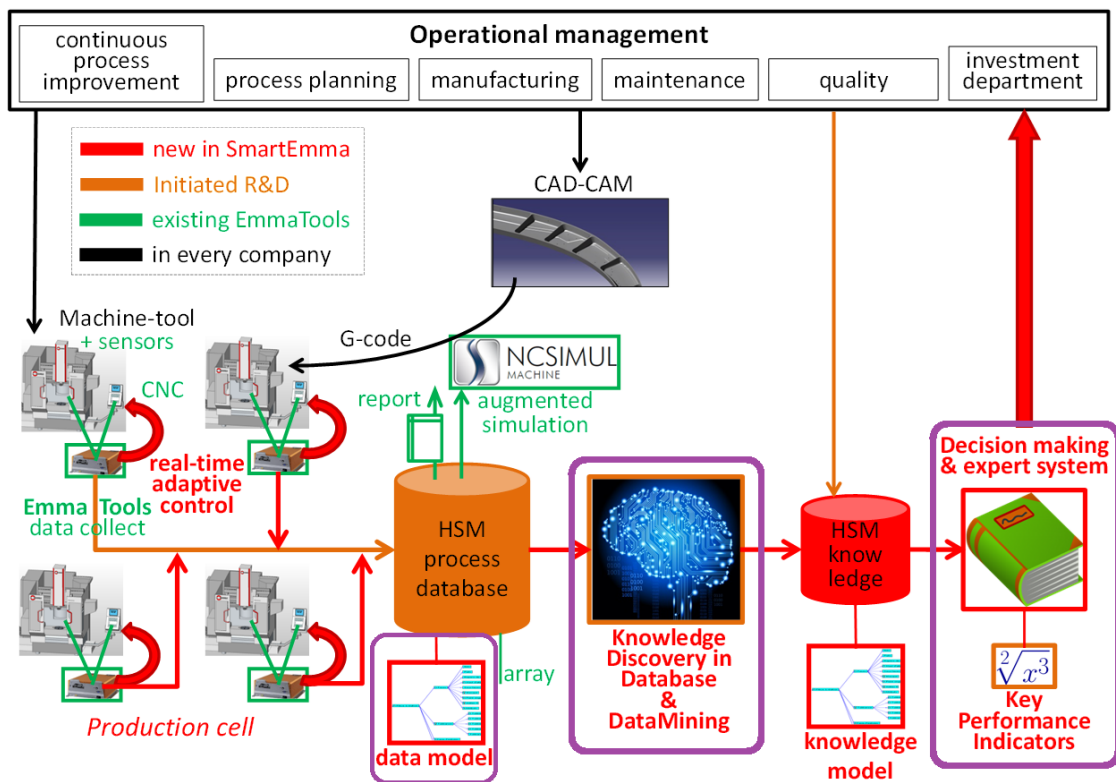


Figure 1.4 – Framework of ANR SmartEmma project

Fig. 1.4 presents the framework of the SmartEmma project. The green parts represent the existing modules such as the EmmaTools, which have been developed in FUI UsinAE project [De Castelbajac et al., 2014]. The orange parts represent the modules which are partly initiated previously. The red parts illustrate the modules needed to be dealt with in this project. It includes the modeling of the raw data collected from Emmatools and data mining from this raw data by Knowledge Discovery in Database (KDD). These two objectives correspond to the first issue mentioned above, which will be solved in this thesis. Then, it is necessary to propose some new KPIs according to the industry's needs, which will play an essential role as the basis of decision-making aid. This objective corresponds to the second issue mentioned above, which will also be dealt with in this thesis. The purple rectangle in Fig. 1.4 presents the research work of this thesis.

After mining from the raw data, the new Smart Data (including more meaningful and smaller volume information) will assist the next step - knowledge modeling, which corresponds to the third issue above. This will be studied in the thesis of O. Meski.

At the end, based on the proposed KPIs and the created knowledge model, the decision-making and expert support system will help machining industry improve its efficiency.

1.2 State of the art

In order to construct a smart and connected machine-tool as well as a decision-aid system, which are the objectives of the project, it is necessary to take knowledge of the state of the art in the corresponding domains. This will help to identify the scientific and technical barriers better.

The first objective is to define the business rules for the detection and the diagnosis of problems in machining. Thus, a literature survey is carried out on machining monitoring, including the sensors used in machining, signal processing, different problematic occurred in machining (surface integrity, chatter, tool breakage, collision and excessive vibration, etc.).

The second part concerns the automation of the data-mining. As mentioned above, the data collected in industry represents a significant volume (For example, 1Go is collected by Emmatools per machine-tool per day). A bibliography concerning the automation of big data processing through algorithms of data mining is conducted.

The third objective is to propose new and adapted KPIs according to the needs of different departments in industry. These KPIs will be the sources for decision-aid indicators. Therefore, a literature review on how to define the adapted and correct KPIs to comply with the demand of industry is essential.

The final objective is about the organization of the data and the integration of the decision-aid indicators into the companies' information systems. A bibliography will be held to review the different existing systems applied in manufacturing, such as CPS, the Internet of Things (IoT), cloud-based manufacturing and digital twins, etc.

1.2.1 Machining monitoring

Manufacturing enterprises face the growing demands of increased product quality, greater product diversity, shorter product development cycles, reduced cost, and global competition [Liang et al.,

2004]. The automation is an effective way to meet these demands in the field of machining. A key issue for an unattended and automated machining system is the development of reliable and robust monitoring systems [Abellan-Nebot and Subirón, 2010]. The objective of machining monitoring is not only to enable shop-floor to track activities and automatically record the performance of machine-tool but also to measure and supervise different defects during machining.

Teti et al. [2010] make an overview of the existing techniques of machining monitoring and list all measurable and studied defects. It includes the sensors used in monitoring, feature extraction and different defects during cutting.

In terms of variables measured by sensors in machining, there are motor power and current measurement, force and torque measurement, acoustic emission measurement and vibration sensors. Al-Sulaiman et al. [2005] presented a universally applicable method for online monitoring of tool condition based on the measurements of differential electrical power consumption. Matsubara and Ibaraki [2009] reviewed monitoring and control schemes of cutting force and torque. Li [2002] make a review on the acoustic emission (AE) method for tool wear monitoring during turning. Plaza and López [2017] applied Singular Spectrum analysis (SSA) to mechanical vibration signals for the in-process monitoring of surface Roughness (RA) in finish turning operations. The choice of sensors depends on the problems needed diagnosis. Sometimes, the multi-sensors would be helpful if there are a lot of incidents that needed to be found. In the case of this thesis, power measurement and vibrations measurement were used to diagnose the problems.

After the signals are collected and stored in the data base, some signal processing methods help eliminate the impurity and feature extraction is necessary to obtain the most adapted feature for diagnostic of incidents. The feature extraction can be divided into time domain and frequency domain:

- In time domain, the most common features are: 1) arithmetic mean, average value, magnitude [Sick, 2002]; 2) effective value (Root Mean Square (RMS)) [Salgado and Alonso, 2006]; 3) variance (or standard deviation) [Zhu et al., 2009]; 4) skewness [Binsaied et al., 2009]; 5) kurtosis [El-Wardany et al., 1996]; 6) peak value [Binsaied et al., 2009].
- In frequency domain, there are: 1) Fast Fourier transform (FFT) [Lee et al., 2017]; 2) Wavelet Transform (WT) [Kamarthi and Pittner, 1997]; 3) Hilbert-Huang Transform (HHT) [Huang, 2014].

Emmatools, used in this thesis, extract the RMS value and the peak value of the vibration signals in time-domain. It also collects the amplitudes and frequency of the vibration information through FFT method. The details of these signals will be presented further in chapter 3.

As mentioned in the review of Teti et al. [2010], the main defects during machining are: chatter, tool wear, collision, and excessive vibration. Abundant bibliography exists on the subject of detecting these incidents. In the framework of this thesis, the following four types of defects will be detailed:

Chatter is a self-excited vibration during machining operations [Altıntaş and Budak, 1995]. It has several negative effects. Firstly, it strongly degrades the surface of the machined workpiece. Secondly, it creates strong vibrations in the spindle that can damage it. Furthermore, it leads to tool wear,

which is harmful and costly. Finally, it reduces the material removal rate (MRR) and therefore reduces the overall productivity.

Many research works were carried out concerning the detection and mitigation of the chatter. Quintana and Ciurana [2011] focus on the state of the art on chatter and existing methods used to detect and to avoid it. For example, out-of-process strategies focus on selecting optimal cutting parameters by seeking stable regions in the Stability Lobe Diagram (SLD) [Altintas and Weck, 2004]. Fig. 1.5 shows how it works when selecting the optimal cutting parameters. The chatter usually occurs in the orange area while in the green zone, less chatter exists.

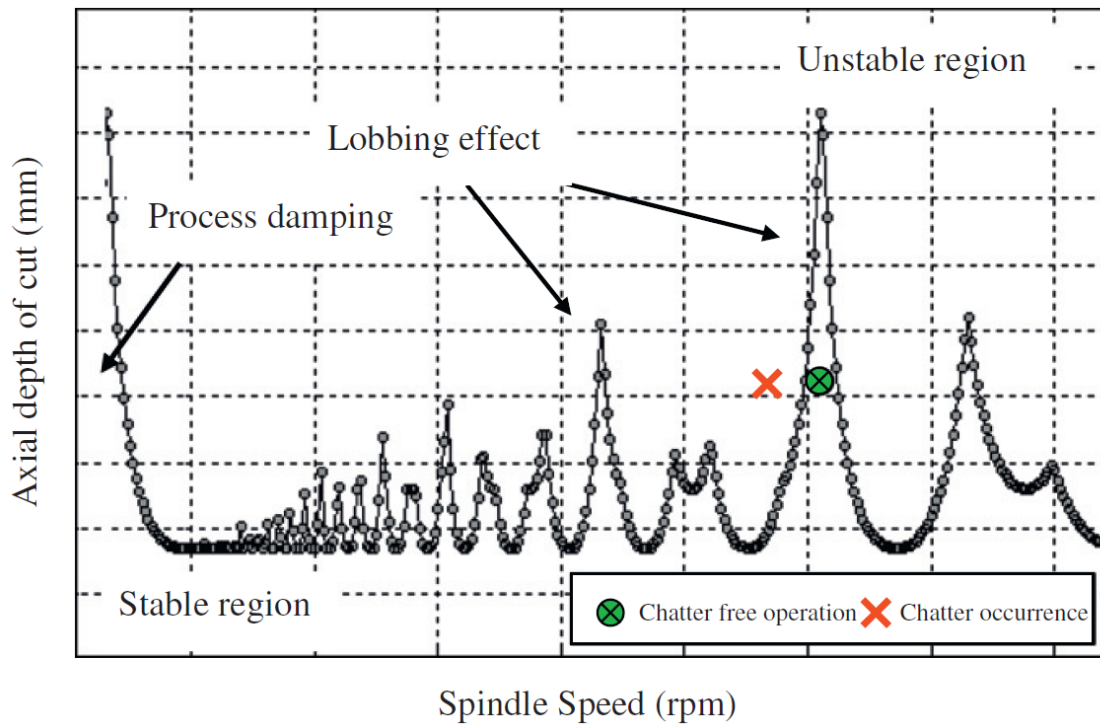


Figure 1.5 – Stability lobe diagram [Quintana and Ciurana, 2011]

In-process strategies focus on identifying or recognizing chatter through the use of several sensor technologies, process monitoring and signal processing. Godreau et al. [2019] propose a criterion based on the sum of the amplitudes of the five dominant non harmonic contributions in FFT (Nh in m/s^2) to identify the chatter during machining. This criterion was proved to be effective when dealing with the big data collected from industry. However, this method is under the assumption of having prior knowledge about material cutting state, which is not the case in reality. As a result, more refined classification should be realized to get rid of this constraint hypothesis. This is the reason for conducting contextual classification in this thesis. There is not yet any bibliography in the current literature for the contextual classification before extracting data from databases.

Tool wear during machining will affect the quality of the workpiece and even damage the spindle. The most important is to detect the tool wear and stop the machine-tool before it generates the defects. Zhou and Xue [2018] make a summary of the methods used to monitor tool wear in milling processes, including sensors, feature extraction and monitoring models. However, the tool condition

for aluminum alloys is totally different from the case of tool condition for titanium alloys which are hard-to-cut materials.

In aluminum machining, [Bhattacharyya et al. \[2007\]](#) showed that cutting force sensors are the most effective means of detecting tool wear. However, cutting force sensors are difficult to apply in industrial environments. This is because they could disturb the cutting process and they are generally expensive. Vibrations sensors are widely employed in Tool Condition Monitoring (TCM) because they are inexpensive, easy to install, and provide a similar periodic signal shape to that of cutting force [\[Sevilla-Camacho et al., 2015\]](#). [Godreau et al. \[2019\]](#) proposed a criterion based on the amplitude of the contribution at the spindle frequency in FFT (Ub in m/s^2) to detect the tool failure in real-time according to the vibration signal. But this criterion can only be applied when the machine tool is not cutting while this last information is generally unknown in reality.

In titanium machining, the tool life is much shorter than that in the aluminum machining due to its low machineability. The poor thermal conductivity of titanium alloys results in the concentration of high temperature at the tool-workpiece and tool-chip interfaces, which accelerates tool wear and consequently increases manufacturing cost [\[Rahman et al., 2006\]](#). Titanium alloys have been widely used in the aerospace because of their good strength-to-weight ratio and superior corrosion resistance. Therefore, it is also an important issue to detect the tool wear in real-time for titanium alloys machining in the context of big data collected in the aeronautic industry.

Collisions between moving and non-moving objects in the workspace cause the highest repair costs and the longest downtime periods [\[Rudolf et al., 2007\]](#). The most frequent sources of collisions are associated with mistakes caused by the operator during setting-up and operation of the machine tool, e.g. incorrect workpiece positioning and tool election as well as programming errors of the numerical control [\[Abele and Korff, 2011\]](#). The most important is to stop the machine-tool whenever load limits on the tool are exceeded to avoid consequential damage. Then, set up reliable business rules to detect the collisions and stop the machine-tool as soon as possible based on the data collected by Emmatools (Power P and vibration signals) in this project is also an important issue.

Excessive vibration, which is different from self-excited vibration (chatter), produced under normal machining conditions. It will also generate the poor quality surface of workpiece and even degrade the spindle. It could be detected by the vibrations signal when it exceeds some limit [\[Godreau, 2017\]](#). However, it should eliminate the case of chatter when the system found that the vibration signal exceeds the threshold.

1.2.2 Data mining - Machine learning in machining monitoring

As mentioned above, a big volume of data was collected by Emmatools for each machine-tool. It is interesting to exploit it through data mining. [Fayyad et al. \[1996\]](#) presented the method of Knowledge Discovery in Database (KDD), which is the first and most general method used in the industry. It consists of a succession of different steps (as shown in Fig. 1.6): selecting the correct data, preprocessing the raw data, transforming the data, data mining and interpreting or evaluating the results.

There are other methods for data mining, such as Cross Industry Standard Process for Data Mining (CRISP-DM) [\[Wirth and Hipp, 2000\]](#). It includes business understanding, data understanding, data

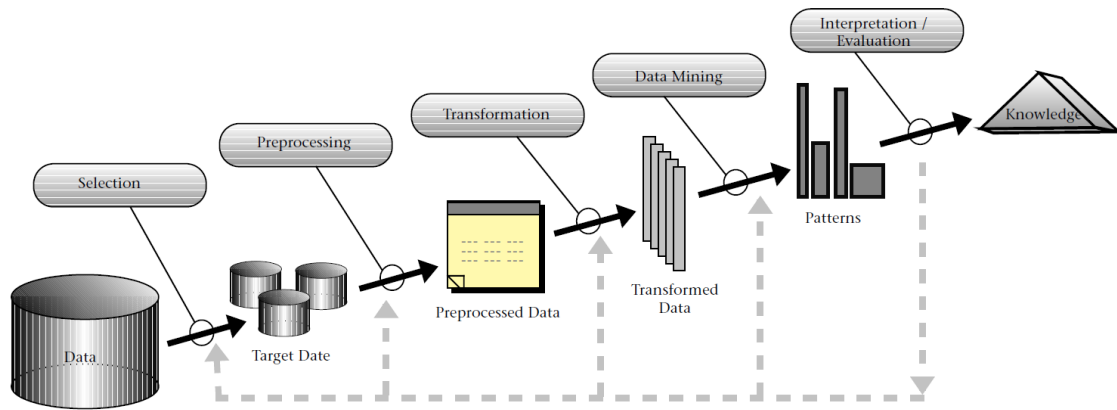


Figure 1.6 – Graph representing the different stages of knowledge extraction from the data [Fayyad et al., 1996]

preparation, modeling, evaluation and deployment. The method SEMMA (Sample, Explore, Modify, Model, Assess) allows an organized and adequate development and maintenance of Data mining (DM) projects [Azevedo and Santos, 2008]. In all these three methods of data mining, the stage at the beginning for each method is very important (selection for KDD, business understanding and data understanding for CRISP-DM, Sample for SEMMA). In order to extract data from databases more efficiently and accurately, contextual classification is necessary to understand the states of the machine-tool and to realize a more relevant calculation of Key Performance Indicators (KPI). That is why the machine learning algorithms will be used in this thesis.

Machine learning is a field of Artificial Intelligence that relies on statistical approaches. It can be applied to machine-tool signals analysis. Kim et al. [2018] list and summarize machining contributions using machine learning algorithms. There are three categories: supervised machine-learning, unsupervised machine learning and reinforcement learning.

Supervised machine learning learns to classify from already labeled output samples. It aims at making correct predictions on the data which are not present in the learning set. This category of algorithms encompasses Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bayesian network, Decision tree learning. Many of these supervised methods have already been used in the machining monitoring domain:

- Artificial neural network (ANN) operates based on biological inspiration. As shown in Fig. 1.7, ANN contains input, hidden and output layers and they are interconnected. These layers contain artificial neurons that have activation function in it [Krishnakumar et al., 2015]. El-Mounayri and Deng [2010] designed and implemented a neural network (ANN) for force prediction to optimize the cutting parameters in optimizing $2\frac{1}{2}$ -axis milling.
- Convolutional neural network (CNN) is a multi-layer neural network trained with the back-propagation algorithm. In CNN, the convolutional layers convolve with raw input data using

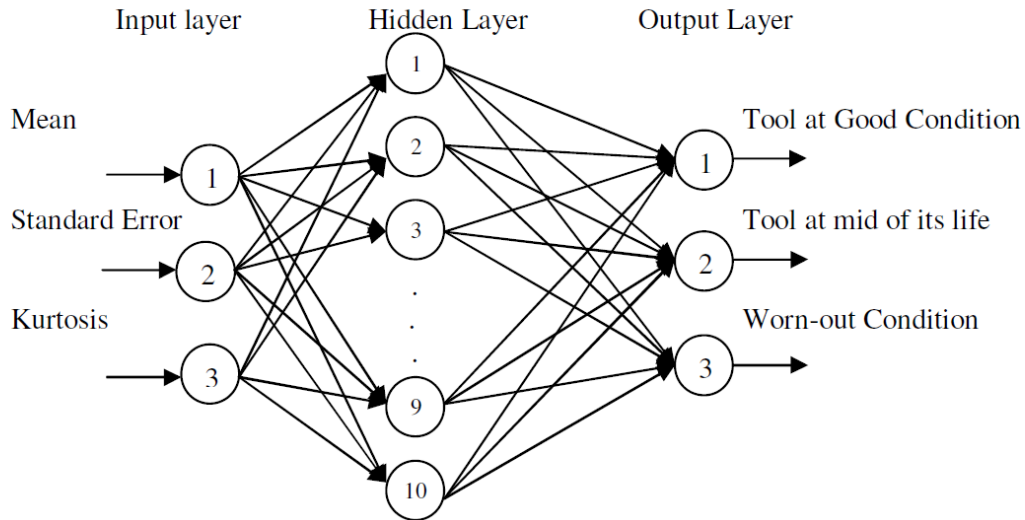


Figure 1.7 – ANN architecture for prediction tool wear in titanium machining in [Krishnakumar et al., 2015]

multiple local kernel filters and generate invariant local features [Wang et al., 2018]. Janssens et al. [2016] proposed a feature learning model for condition monitoring based on convolutional neural networks.

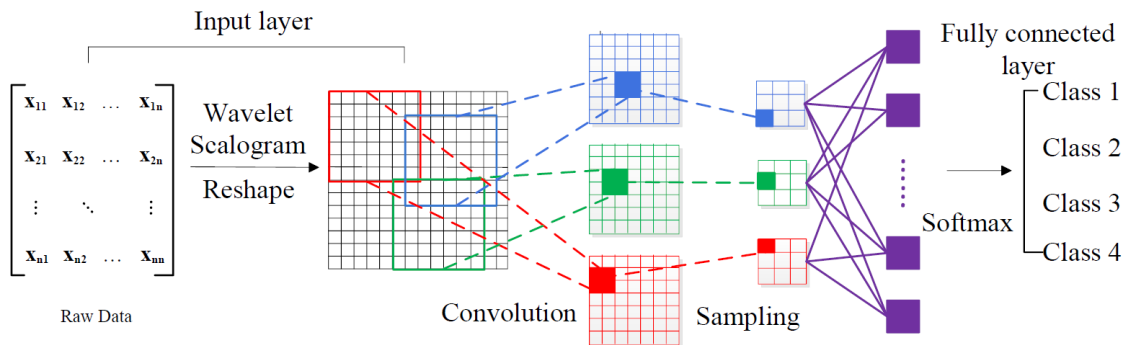


Figure 1.8 – Architecture of convolutional neural network model in [Wang et al., 2018]

- Support vector machine (SVM) constructs a hyperplane or set of hyperplanes in a higher space, which can be used for classification and regression analysis [Cherkassky and Mulier, 2007]. Zhang and Zhang [2016] show that the least square support vector machine (SVM)-based tool wear model is suitable to predict tool wear a certain range of cutting conditions in milling operation.

- K-Nearest Neighbors (KNN) is a non-parametric method proposed by Thomas Cover [Cover and Hart, 1967] used for classification and regression. Andre et al. [2013] applied k-Nearest Neighbor (kNN) to monitor rotational machines using vibration data.
- Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a Directed Acyclic Graph (DAG) [Friedman et al., 1997]. Dey and Stori [2005] tested the Bayesian network into the root cause diagnosis of process variations in machining and found that it is an effective tool for detecting the tool wear by combining the data from multiple sources (spindle power and acoustic emission).
- Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves) [Wu et al., 2008]. Rodríguez et al. [2017] propose a decision-making tool based on decision trees for roughness prediction in face milling.

Unsupervised machine learning learns to classify unlabeled data. There are different types of unsupervised machine learning: K-means, hierarchical classification (agglomerative or divisive), distribution density estimates (e.g., Gaussian Mixture Model-GMM).

- K-means clustering aims at partitioning n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers) [Krishna and Murty, 1999]. Xing et al. [2018] combined the Principal Component Analysis (PCA) method and K-means method to extract the features of Volumetric Errors (VE) vectors and classify the machine-tool accuracy states. They found that K-means can automatically classify the VE feature data and successfully recognize the three faults from the machine-tool normal states.
- Hierarchical classification (agglomerative or divisive) constructs clusters by recursively partitioning the instances in either a top-down (divisive) or bottom-up (agglomerative) fashion [Rokach and Maimon, 2005]. Emec et al. [2016] applied hierarchical agglomerative clustering to group the data with common characteristics when set up a fault-monitoring framework for manufacturing equipment based on data stream mining techniques for online pattern matching in electrical power data streams.
- Distribution density estimates is a statistical tool used to parametrically estimate the distribution of random variables by modeling them as a mixture of known distribution (e.g., Gaussian distribution, Poisson distribution, Bernoulli distribution, etc.) [Bishop, 2006]. Bhinge et al. [2014] build an energy prediction model when machining a component using the Gaussian regression algorithm. Godreau et al. [2019] calculate the criteria thresholds by modeling the distribution of criteria into double Generalized Extreme Value distribution (GEV).

Reinforcement learning is an area of machine learning concerned with how agents ought to take actions in an environment that is typically stated in the form of a Markov Decision Process (MDP) in order to maximize the notion of cumulative reward [Wiering and Van Otterlo, 2012]. The difference between Reinforcement learning and Supervised learning is that it doesn't need labeled input/output pairs, which are 100% correct. And the difference between Reinforcement learning and Unsupervised learning is that it needs labeled input/output pairs.

The Table 1.1 summarize different algorithms of supervised machine learning and unsupervised machine learning as well as reinforcement learning, and compare them by listing their advantages and disadvantages.

In industry, it is challenging to collect labeled output data. As a result, most industry applications use unsupervised machine learning methods.

1.2.3 Key performances indicators for smart machining

A Key Performance Indicator (KPI) is a measurable value that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs at multiple levels to evaluate their success at reaching targets [Parmenter, 2015]. Choosing the right KPIs relies upon a good understanding of what is important to the organization. What is deemed important often depends on the department (e.g., the KPIs useful to quality will differ from the KPIs assigned to manufacturing). For example, the manufacturing department use Overall Equipment Effectiveness (OEE) to measure how well a manufacturing operation is utilized. The maintenance department uses Mean time Between Failures (MTBF) as a predicted elapsed time between inherent failures of a mechanical or electronic system, during normal system operation. Kibira et al. [2018] presented a procedure to select KPIs for manufacturing processes: 1) identifying candidate KPIs from existing sources, 2) defining new candidate KPIs, 3) selecting appropriate KPIs based on KPI criteria, and 4) composing the selected KPIs with assigned weights into a set.

One of the objectives of this thesis is to propose more new KPIs based on the data-mining of the big volume data collected by Emmatools. Especially, the interest of the different departments of a manufacturing company differs, and each one needs adapted KPI to take benefit from the large process databases. The major difficulty will consist in discovering new relevant and simple indicators of performance, comprehensive for non-expert people and resulting from the aggregation of several Mega-octets of process data.

1.2.4 Decision-aid system

Another important aspect is the organization of the KPIs as well as the knowledge obtained through data mining, its integration into the company's information systems [Bernard and Perry, 2003] and the proposition of a decision-aid support system for industry. The global context of this project is the development of Industry 4.0, a company instrumented by the connection of these networks, made it possible to design the transmission and storage of information in the factory of the future. Morgan and O'Donnell [2018] presented the design and development of a cyber-physical process monitoring system (CPS) for intelligent machine-tool. Lenz et al. [2018] propose a holistic approach to machining data analysis in order to combine tasks and to group analysis objectives among different departments within the company. Liu et al. [2018] propose a systematic development method for machine-tool 4.0. Cloud-based manufacturing aims to distribute information to the places where it is needed to improve the efficiency of each department and enable more robust decision-making, exploiting large information and knowledge sharing ([Gao et al., 2015], [Caggiano, 2018]). Chen et al. [2018] proposed an energy efficiency monitoring and management system with the support of the newly emerging Internet of Things (IoT) technology. Tao et al. [2018] applied digital twin (by the convergence between physical

Table 1.1 – Advantages and disadvantages of each algorithm for three modes of machine learning

	Mode	Algorithms	Advantages	Disadvantages
Machine learning	Supervised	ANN	<ol style="list-style-type: none"> 1) Ability to work with incomplete knowledge 2) Storing information on the entire network 3) Having fault tolerance 4) Parallel processing capacity 	<ol style="list-style-type: none"> 1) Convergence is not always assured 2) Unexplained behavior of the network 3) Risk of over-learning 3) Long calculation time for large volumes of data
		CNN	<ol style="list-style-type: none"> 1) Automatic detection the important features without human supervision 2) Computationally efficient 3) Pooling operations and parameter sharing 	<ol style="list-style-type: none"> 1) CNN do not encode the position and orientation of object 2) Lack of ability to be spatially invariant to the input data 3) CNN needs a large data set
		SVM	<ol style="list-style-type: none"> 1) SVM works relatively well when there is a clear margin of separation between classes 2) SVM is more efficient in high dimensional spaces 3) SVM is effective when number of dimensions is greater than the number of samples 4) SVM is relatively memory efficient 	<ol style="list-style-type: none"> 1) SVM is not suitable for large data sets 2) No probabilistic explanation for the classification 3) SVM does not perform well when the data set has more noise
		KNN	<ol style="list-style-type: none"> 1) No training period and much faster 2) New data can be added seamlessly which will not impact the accuracy of the algorithm 3) Suitable for data where each class is represented by several criteria and the boundaries are irregular 	<ol style="list-style-type: none"> 1) KNN does not work well with large data set 2) KNN does not work well with high dimension 3) KNN needs feature scaling 4) KNN is sensitive to noisy data, missing values and outliers
		Bayesian network	<ol style="list-style-type: none"> 1) It is suitable for small and incomplete data sets 2) Structural learning possible 3) Combining different sources of knowledge 4) Explicit treatment of uncertainty and support for decision analysis 5) Fast responses 	<ol style="list-style-type: none"> 1) Computationally expensive to explore a previously unknown network 2) Sensitive to the prior probabilities chosen 3) Requires greater statistical expertise
		Decision tree	<ol style="list-style-type: none"> 1) It requires less effort for data preparation 2) Decision tree does not require normalization data 3) Missing value does not affect the process of building decision tree 4) It is intuitive and easy to explain 	<ol style="list-style-type: none"> 1) A small change in the data will cause a large change in the structure 2) Decision tree involves higher time to train the model 3) Computationally expensive 4) It is not suitable for applying regression and predicting continuous value
	Unsupervised	K-means	<ol style="list-style-type: none"> 1) Computationally faster 2) K-means produces tighter clusters than Hierarchical clustering, especially if the cluster are globular 	<ol style="list-style-type: none"> 1) It is difficult to predict the K value 2) Different initial partitions can result different final clusters 3) It does not work well with clusters of different size and different density
		Hierarchical classification	<ol style="list-style-type: none"> 1) Not need do specify the number of clusters required in algorithm 2) Easy to implement 3) The dendrogram produced is very useful in understanding the data 	<ol style="list-style-type: none"> 1) Long computation time than K-means 2) Difficult to determine the correct number clusters by the dendrogram 3) Partitioning criterion is not global
		Distribution density estimates	<ol style="list-style-type: none"> 1) Each class is characterized by its own probability law 2) It is suitable with clusters of different size and different density 3) Maximum likelihood estimation is used to make the model more approach to the true density of data 	<ol style="list-style-type: none"> 1) Each cluster must follow the probability law 2) It does not work well if the sample data is too small
	Reinforcement	Monte Carlo Methods	<ol style="list-style-type: none"> 1) There is virtually no limit to the analysis. Empirical distributions can be handled. 2) It can be easily extended and developed as required. 3) Easily understood by non-mathematicians 	<ol style="list-style-type: none"> 1) Computationally expensive 2) Solutions are not exact, all outputs are estimates.

and virtual product) for product design, manufacturing, and service driven by big data. Anwer et al. [2018] addressed the classification problem of partitioning operations for the development of ISO GPS (Geometrical Product Specifications and verifications) partitioning standards, which is the potential key enablers of smart manufacturing. Castanier et al. [2005] proposed a decision-aid system for condition-based maintenance of a two-unit system with independent stochastically deteriorating units

These new methods of transferring and sharing information between various departments in the company can be effective in improving overall quality. Here a few studies show what the gains for quality [Köksal et al., 2011], maintenance [Kamsu-Foguem et al., 2013], or workshop organization [Chen, 2003] to improve the efficiency in the industry.

1.3 Scientific and technological barriers

Based on the context of the aeronautical industry, the objectives of the ANR SmartEmma project as well as the state of the art of the current development in the corresponding domain described in the previous sections, three problems appear to remain unsolved to achieve the set objectives:

- Many researches propose the use of machine learning algorithms directly on the raw data (machine-tool signals). However, few researches focus on data pre-processing to identify the context. Contextual classification before data mining is very important not only because it allows classifying the raw data into ordered and meaningful forms which is helpful for the data mining following, but also it could reduce the noise and help calculate the relevant KPIs which became the source for the decision-aid support system in smart machining.
- The criteria for detection of chatter (Nh), tool wear (Ub) were proposed in the thesis of [Godreau, 2017]. However, taking into the consideration context information related to the incidents, the business rules for detection of each incident based on the contextual classification need to be set up. It includes chatter, tool wear, collisions and excessive vibration.
- Although a few KPIs can be found in previous works [Godreau, 2017], a significant effort is necessary to find more relevant KPIs which will be beneficial for different departments of a manufacturing company. These KPIs should not only cover the machine-tool for aluminum alloys, but also incorporate the machine-tool for titanium alloys.

1.4 Structure of this thesis

This thesis is organized as follows:

Chapter 2 introduces the complete business rules for contextual classification of data for smart machining based on unsupervised machine learning. K-means and GMM were compared, and GMM is more adapted for the use in industry. Then BIC method is applied to determine the number of clusters. The classifications are then evaluated using manual data mining and k-fold validation.

Chapter 3 proposes the business rules for detecting different incidents (chatter, tool wear and excessive vibration) based on the contextual classification and the criteria correspondent.

Chapter 4 presents the KPIs calculation for different departments based on the contextual classification and incidents detection. The information of contextual classification and the incidents detected were aggregated into Smart Data. Then the KPIs could be computed based on this Smart Data.

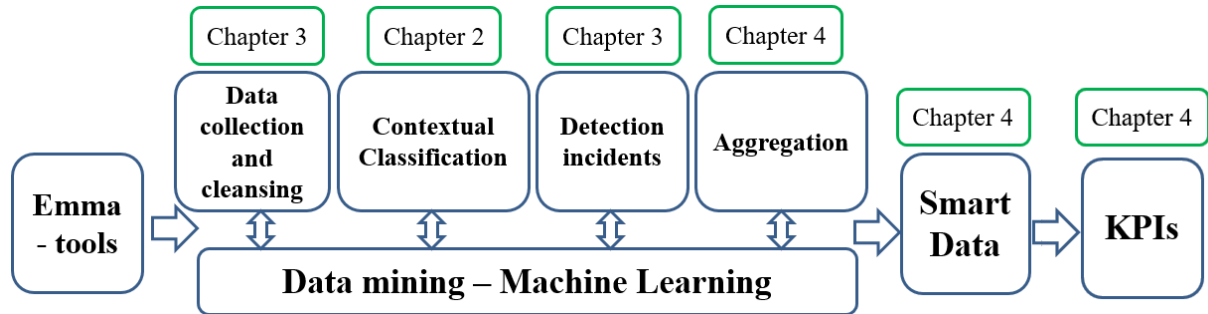


Figure 1.9 – Flow chart of data processing as well as the structure of thesis

Figure 1.9 illustrates the flow chart of data processing as well as the structure of this thesis: first step, data is collected by Emmatools and is cleaned before data mining (in chapter 3); secondly, contextual classification is proposed to identify different states of machine-tool at each instant (in chapter 2); then incidents are detected based on the business rules as well as the contextual classification (in chapter 3); at the end, the contextual information and the detected incidents are aggregated into Smart Data, and the KPIs can be computed based on these Smart Data (in chapter 4).

Contextual classification for smart machining based on unsupervised machine learning

Contents

2.1	Motivation of the contextual classification	21
2.2	Comparison between Kmeans and GMM	23
2.2.1	Introduction	23
2.2.2	K-means	24
2.2.3	GMM classification	26
2.2.4	Conclusion	28
2.3	Gaussian mixture model	29
2.4	Determination of cluster number by Bayesian Information Criterion (BIC)	30
2.4.1	BIC on the variation of spindle speed	31
2.4.2	BIC on the variation of feedrate	31
2.4.3	BIC on vibrations and power	31
2.5	Business rules for contextual classification	33
2.5.1	Contextual classification of machine-tool	33
2.5.2	Contextual classification per type of machining operation	35
2.5.3	Contextual classification per type of tool engagement	35
2.6	Contextual classification of the material removal	36
2.6.1	Modeling and classification for machine-tool #2	36
2.6.2	Modeling and classification for machine-tool #3	39
2.7	Method validation	41
2.7.1	Accuracy	41
2.7.2	Sensitivity of the method	43
2.8	Conclusion	44

Summary

This chapter proposes a contextual classification procedure in machining based on unsupervised machine learning by Gaussian mixture model. The optimal number of classes is determined by the silhouette method, based on Bayesian information criterion. This method is validated on real data from four different machine-tools in aerospace industry to know the material removal. Manual data mining and cross validation confirm that the proposed method provides good contextual classification results. Based on the contextual classification on the material removal, the different operations of machining and the different states of tool engagement during machining could be detected more precisely.

The structure of this chapter is as follows: it begins with the motivation of contextual classification. This is followed by a comparison between two methods of unsupervised machine learning: K-means and Gaussian-mixture method, to determine which method is more effective for contextual classification. Then the Gaussian-mixture model is better. The number of classes in the Gaussian model is determined by the BIC method. Based on this, the business rules for contextual classification are presented. Then, the proposed GMM method is applied to vibration signals and power signals for two machine-tools to know whether the tool is cutting materials or not. Finally, verification the results of classification by manual data mining and cross-validation is done for these two machine-tools.

The work presented in this chapter is published in the journal paper entitled "Contextual classification for smart machining based on unsupervised machine learning by gaussian mixture model" [Wang et al., 2020], International Journal of Computer Integrated Manufacturing, 2020.

2.1 Motivation of the contextual classification

As described previously in the first chapter, abundant digital data have been collected by the Em-matools. The machine learning methods will be developed on the one hand to perform contextual classifications, and on the other, to aggregate the data. 1 GB of data is collected per day and per machine-tool. A knowledge-based multi-level aggregation approach will help decrease the data volume while increasing their significance [Ritou et al., 2019]. Contextual classification permits to classify raw data into ordered and meaningful forms. Using this contextual information, noise is reduced, and relevant KPIs can be calculated. Those KPIs support decisions on the HSM process and the global management of the company, as shown in Fig. 2.1. It is interesting for the industry to know:

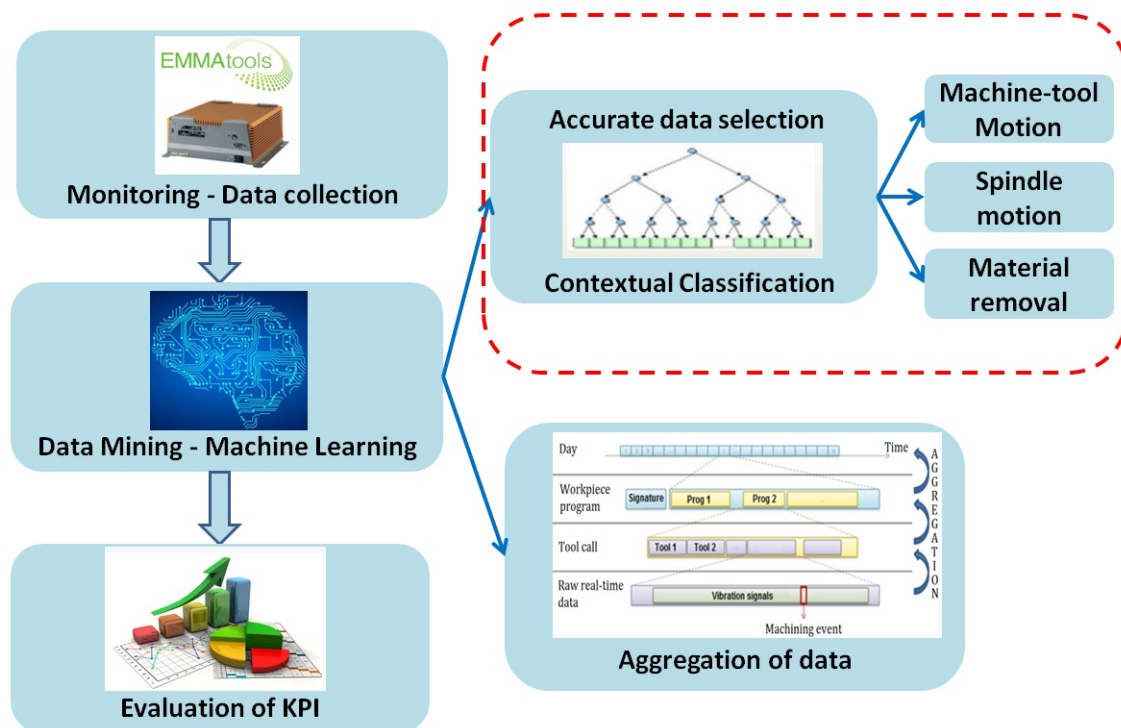


Figure 2.1 – Data mining process for machining

- the machine-tool motion;
- the spindle motion;
- the material removal.

Contextual classification is very important at the beginning stage of data mining [Lemberger et al., 2016]. A wrong contextual classification will conduct to false alarm or missed detection phenomenon. In machining, it is known that if the tool is cutting, its vibration ($Arms$ in m/s^2) is greater than if the tool is not cutting. It is also known that when the tool is cutting, the power (P in kW) will higher than the power when the tool is not cutting. However, before the machining materials, the spindle speed N (in RPM) will accelerate from 0 to the target speed ($N=23720$ RPM in Fig. 2.2). During this *spindle startup* period, the spindle power P (kW) will reach a big value (around 50 kW)

because of the acceleration of the spindle speed, but which is not due to the material removal. Also, in aeronautic machining, the tool is not cutting materials once the N reaches the target speed. There is a period to lubricate the spindle (*post spindle startup*). During the *post spindle startup*, the signals of acceleration $Arms$ (m/s^2) will increase to a big value (around $200 m/s^2$ in the Fig. 2.2), but the tool is not cutting. If the wrong contextual classification occurs, the period of '*spindle startup*' and the period of '*post spindle startup*' will be wrongly labeled as '*tool cutting*', this is called 'false alarm'. As shown in Fig. 2.2. The machine-tool feedrate Vf (in m/min) is in cyan line, the spindle speed N

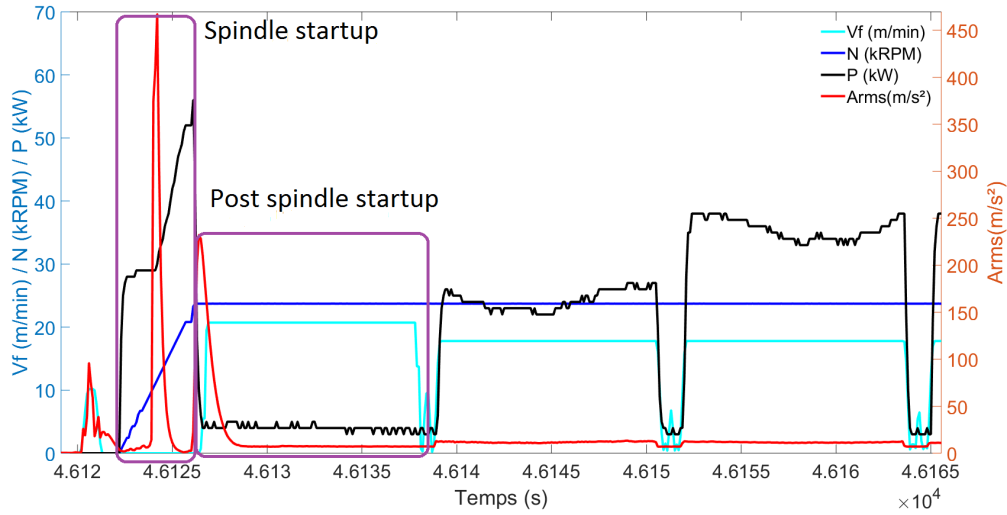
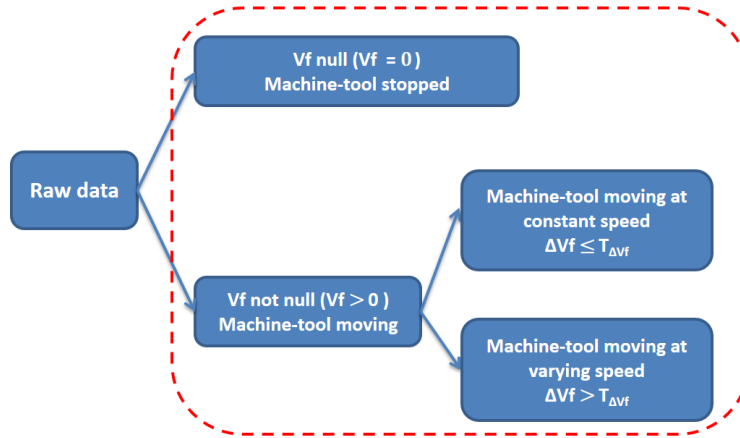


Figure 2.2 – Example of the false alarm

(kRPM equals to x1000 RPM) is in blue line, the acceleration signal $Arms$ (m/s^2) is in red line, and the spindle power P (in kW) is in black line. The first rectangle in purple on the left side represents the period of '*spindle startup*' where the power P reaches 50 kW. While the second rectangle in purple shows the period of '*post spindle startup*' where the acceleration signal $Arms$ reaches a high value (around $200 m/s^2$). The wrong contextual classification will conduct to these two periods wrong labeled as '*tool cutting*', which will lead to the wrong computation of KPIs. And in the end, it will impact the decision-making for the industry.

For the material removal, the tool is cutting materials only when the machine-tool is moving (i.e. feedrate (Vf) of the spindle is not null). Also, when the tool is cutting materials, the tool must be rotating ($N > 0$). Eventually, the machine-tool can (de)accelerate. The varying Vf involves varying cutting conditions. Therefore, classifying the raw data by the feedrate Vf (m/min) into three clusters is necessary: the machine-tool can be stopped or moving at constant or varying speed. The Figure 2.3 represents the flowchart of the method. The data during one day from an aircraft company was taken to test. The data describing the feedrate Vf and its variation ΔVf (in mm/s^2) are used for the classification. Firstly, the data with Vf equals to 0 m/min were removed from the selection and defined as the cluster '*Machine-tool stopped*'. Then, in the cluster where the feedrate Vf is not null, the data could be classified as two classes: *Machine-tool moving at constant speed* and *Machine-tool moving at varying speed*, according to the classification of the variation ΔVf .

The variation of a variable X (e.g., feedrate Vf , spindle speed N or spindle power P) is defined as the

Figure 2.3 – Process of classification according to the feedrate Vf and ΔVf

the simple derivative of the variable X , described as ΔX (*e.g.*, ΔVf , ΔN or ΔP). This is defined as the equation as below:

$$\Delta X_n = \frac{X_n - X_{n-1}}{\Delta t} \quad (2.1)$$

where, X_n is the n^{th} measure; X_{n-1} is the measure at the previous instant and Δt is the data recording period (0.1s for data collected by Emmatools). The classifications are then carried out based on the following rules:

- There are three clusters for machine-tool movements: stopped ($Vf = 0$), moving at constant speed ($Vf > 0$ and $\Delta Vf \leq T_{\Delta Vf}$), or varying speed ($Vf > 0$ and $\Delta Vf > T_{\Delta Vf}$);

The objective is to calculate a classification threshold $T_{\Delta Vf}$ by machine learning methods.

2.2 Comparison between Kmeans and GMM

2.2.1 Introduction

In industry, it is challenging to collect labeled output data. As a result, most industrial applications use unsupervised machine learning methods. There are three different types of unsupervised learning: K-means, hierarchical classification (agglomerative or divisive), distribution density estimates (*e.g.*, Gaussian Mixing Model - GMM). Hierarchical classification (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters [Rokach and Maimon, 2005]. The method to build a hierarchy is to calculate the distances between every observation, which is the same as K-means. The distribution density estimates aim to find a distribution density model the closest to the observation in the true data [Bishop, 2006], which is different from K-means and hierarchical classification. In this section, Kmeans and GMM will be applied to industry data to compare their effectiveness.

2.2.2 K-means

To classify the machine-tool context information into clusters, an unsupervised machine learning method named K-means may be helpful. K-means clustering aims at partitioning n observations into k clusters in which each observation belongs to the cluster with the nearest mean (centroid), serving as a prototype of the cluster [Jacques, 2001].

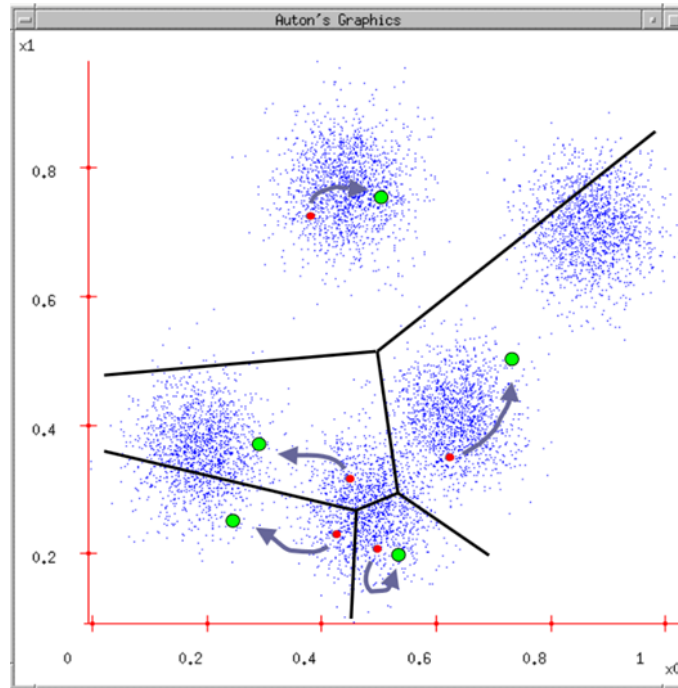


Figure 2.4 – Process of classification by K-means [Jacques, 2001]

The main steps of K-means are represented in Figure 2.4 and are as below:

1. Define the number of clusters.
2. Initial cluster centroids (in red).
3. Assign each point to the nearest cluster.
4. Update the centroids of each cluster (in green).
5. Repeat 3rd and 4th step until the centroids of each cluster don't change.

At the beginning, the data whose feedrate $Vf = 0$ should be removed (as shown in Figure 2.3). Then, for the data whose Vf is not null, K-means was applied for the ΔVf (mm/s^2) to classify ΔVf into 2 states: *machine-tool moving at constant speed*, *machine-tool moving at varying speed*. The Figure 2.5 represents the results of the classification by K-means to ΔVf during 15 hours. Three classes were obtained: '*Machine-tool stopped*' is presented by green diamond, '*Machine-tool moving at constant speed*' is represented by red asterisk, while '*Machine-tool moving at varying speed*' is presented by blue plus.

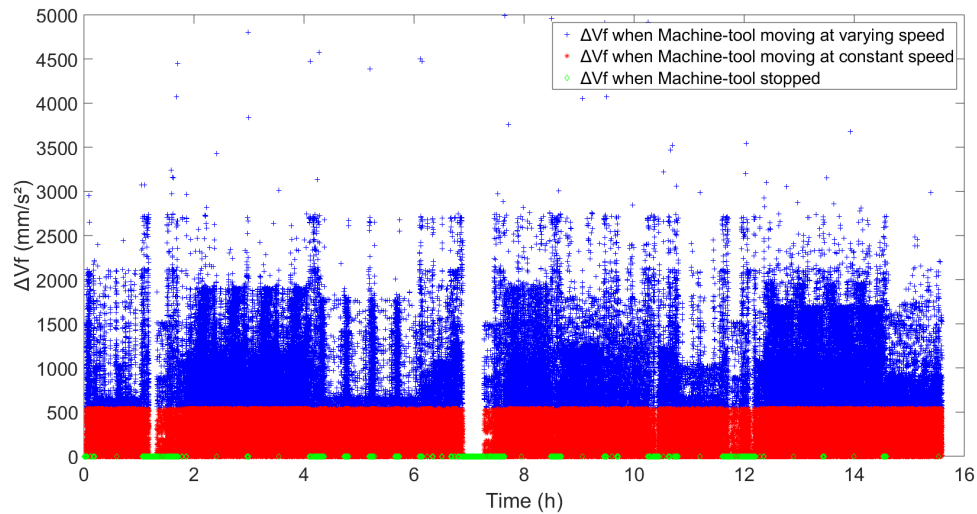


Figure 2.5 – Classification of ΔVf into 3 classes by K-means

The threshold of machine-tool moving at constant or varying speed locates on the boundary between the two clusters: $T_{\Delta Vf} = 500 (mm/s^2)$. Figure 2.6 shows the results of the contextual classification of Vf , during a period of 30s. The Vf is labeled with different colors according to their clusters. *Machine-tool stopped* are presented by green diamond; *Machine-tool moving at constant speed* are presented by red asterisk while *machine-tool moving at varying speed* are presented by blue plus.

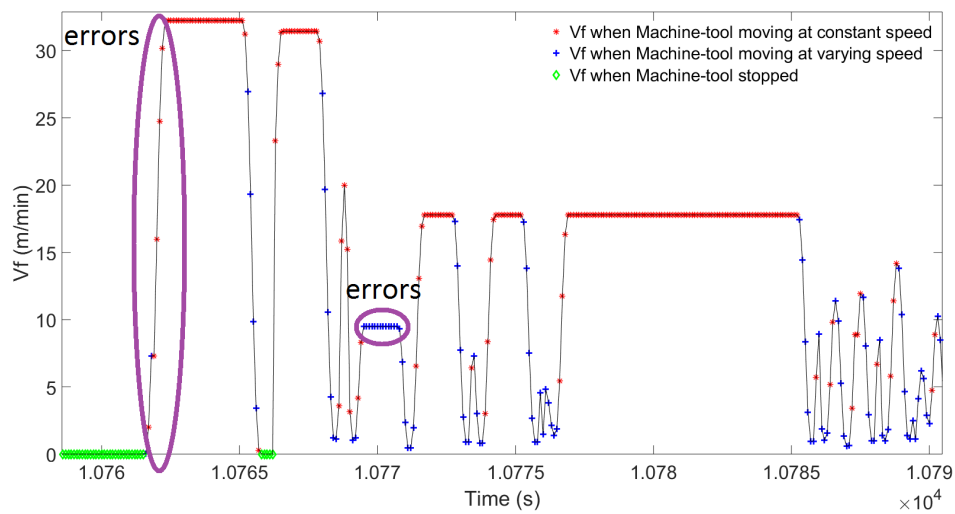


Figure 2.6 – Classification of Vf by K-means during 30 seconds

The performance of this classification can be tested by comparing the automated classification to manual mining. It is found manually that, during the Vf acceleration, there are many red asterisk (cluster *machine-tool moving at constant speed*) in this acceleration period. Which means the machine-tool is accelerating while the K-means suppose that these points are in the cluster *machine-tool moving at constant speed*. And there are many other classification errors in the *machine-tool moving*

at constant speed (cycle purple). It means the threshold found by K-means $T_{\Delta Vf} = 500$ (mm/s^2) is too high. To evaluate the performance of the classification by K-means, a confusion matrix could be useful. In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix [Stehman, 1997] is a specific table layout that allows visualization of the performance of an algorithm (*e.g.* supervised machine learning and unsupervised machine learning). Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or 100% classes). It makes it easy to see if the method is confusing two classes (*e.g.* mislabeling one as another). In this case, the classes labeled as 'Machine-tool moving at varying speed' and 'Machine-tool moving at constant speed' by K-means are the predicted classes (It should be noted that 'Machine-tool stopped' defined as $Vf=0$ is out of the calculation of confusion matrix). Then, go through manual data mining to correct the classification errors, one point by one point, to obtain the actual classes. One period is chosen in this case (4 hours), and the confusion matrix is presented in Table 2.1:

Table 2.1 – Confusion matrix of 4 hours data for the classification of Vf by K-means

Confusion Matrix		Actual Classes	
		Vf Constant	Vf Varying
Predicted Classes	Vf Constant	TP=72 480	FP=18 240
	Vf Varying	FN=16 800	TN=28 800

Accuracy of the Classification (ACC) is calculated as the division between the sum of the correct estimations (True positives + True negatives) and the sum of all estimations (True Positives + True Negatives + False Positives + False Negatives). It is computed as in Eq. (2.2):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

Then, the accuracy of the classification of Vf by K-means during 4 hours is $ACC(Vf)_{K\text{-means}}=74.30\%$ which is not good for the contextual classification.

There is a limitation when using K-means. K-means assumes that it deals with spherical clusters and that each cluster has roughly equal numbers of observations [Wang et al., 2019]. In this case, the distribution of ΔVf is not spherical. And there is much more population of ΔVf around zero, compared to other populations. This means K-means will partition the data where [machine-tool moving at varying speed] into the cluster *machine-tool moving at constant speed*. So other data clustering methods will be tested.

2.2.3 GMM classification

Gaussian mixture model (GMM) [Reynolds, 2009] is a parametric probability density function represented as a weighted sum of Gaussian component densities which maximize the component posterior probability given the data. Its advantage is showing how to fit a GMM to data, cluster using the fitted model, and estimate component posterior probabilities (more information will be given in section 2.3).

Here, the same data as K-means is used as an example: we want to identify 3 states using data describing Vf and ΔVf . The same method (cf. Figure 2.3) is used. Then GMM was to be used to

classify ΔVf into 2 states: *machine-tool moving at constant speed*, *machine-tool moving at varying speed*. It is found that the sum of these 2 Gaussians modeled well our data (see the Figure 2.7). There are 2 great populations of ΔVf : ΔVf around zero (Y1) which is narrow and high is presented in green line; ΔVf distributed along the whole day (Y2) which is wide and low is presented in red line. The model (the sum of the probability of these 2 Gaussians along with the ΔVf (Y3) which is in dotted lines) is drawn to be compared with our true data (density of probability of ΔVf (Histogram of ΔVf)). The threshold $T_{\Delta Vf}$ is chosen at the intersection between the Gaussian Y1 and the other Gaussian Y2 to minimize the false positives (FP) and false negatives(FN), thus obtaining $T_{\Delta Vf} = 2.4$ (mm/s^2).

Figure 2.7 presents the density of probability of ΔVf during this day. The abscissa X is ΔVf , the abscissa Y is the density of probability in scale log.

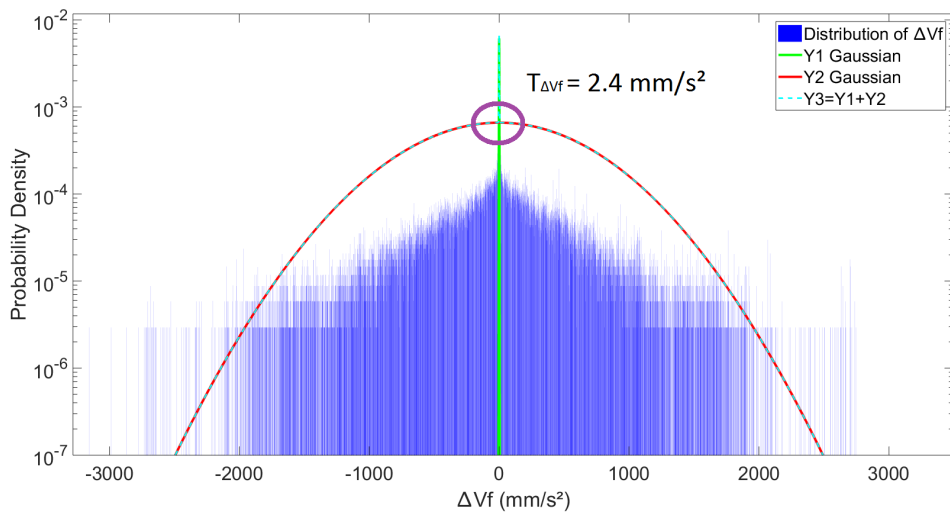


Figure 2.7 – Density of probability of ΔVf for one day modeled by GMM (2 Gaussians)

The threshold is applied to the classification combined with the business rules mentioned above, the same period (30s) of classification result as in K-means is presented in Fig. 2.8. There is no wrong classification which is better than K-means method. Then, to evaluate the performance of classification of Vf by GMM, confusion matrix will be used. The same method as shown in K-means, the classes '*Machine-tool stopped*' defined by $Vf=0$ is excluded out of the calculation. The threshold $T_{\Delta Vf}$ is applied to the classification. Then, the predicted classes by GMM is obtained. The actual classes is obtained by manual data mining as done in the K-means step. The confusion matrix for GMM during the same period (4 hours) as in K-means is shown as in Table 2.2:

The accuracy of classification by GMM is $ACC(Vf)_{GMM}=97.30\%$. It is good enough compared with K-means (74.30%). While there are some errors (FN=368) coming from '*Machine-tool moving at constant speed*', but GMM consumed that they are '*Machine-tool moving at varying speed*'. It is because the threshold $T_{\Delta Vf} = 2.4$ (mm/s^2) is a little high in some period. However, the threshold is approach enough to the exact value.

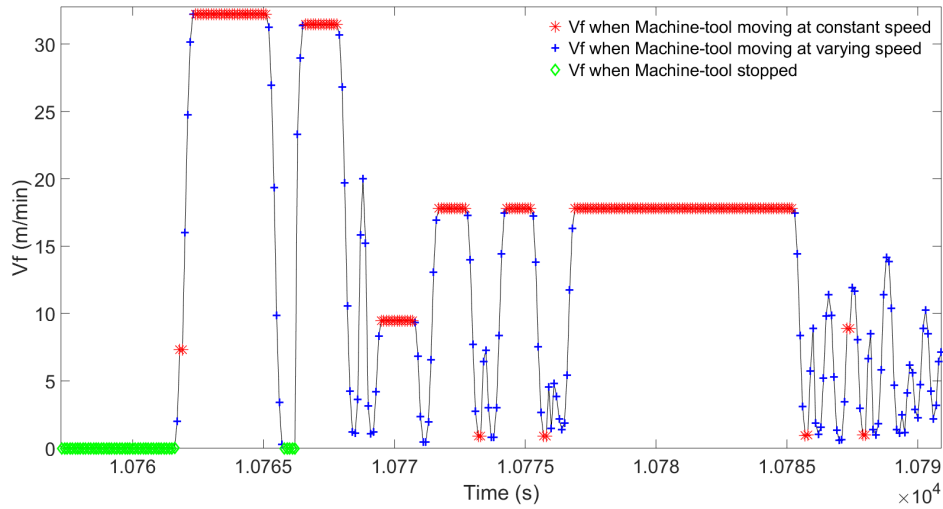


Figure 2.8 – Classification of V_f by GMM during 30 seconds

Table 2.2 – Confusion matrix of 4 hours data for the classification of V_f by GMM

Confusion Matrix		Actual Classes	
		Vf Constant	Vf Varying
Predicted Classes	Vf Constant	TP=90 720	FP=0
	Vf Varying	FN=368	TN=45 232

2.2.4 Conclusion

Two methods of unsupervised machine learning, K-means and GMM, have been tested for contextual classification, based on the machine-tool feedrate V_f , into 3 clusters: machine stopped; constant or varying moving. The results are as below:

- K-means is not suitable because the data is not spherical data (the population of ΔV_f around zero is too large). However, GMM can classify the raw data well.
- The threshold $T_{\Delta V_f}$ is chosen at the intersection between the Gaussian *Machine-tool moving at constant speed* and the other Gaussian *Machine-tool moving at varying speed* to minimize classification errors (false positives and negatives), thus obtaining $T_{\Delta V_f} = 2.4$ (mm/s^2).
- The classification results show that GMM method is better than K-means method in the contextual classification.

Gaussian mixture model will be applied in the whole contextual classification (Machine-tool motion, Spindle motion and Material removal).

2.3 Gaussian mixture model

The Gaussian Mixture Model is a statistical model used to parametrically estimate the distribution of random variables by modeling them as a mixture of Gaussian distributions. It has been used in several areas such as audio recognition, measurement error modeling [Huang et al., 2005]. The principle of GMM is to find an approximation of the probability density f from a sample of n observations of the random vector $X = \{x_1, x_2, x_3 \dots x_n\}$ (where each observation x_i is assumed independent) using a linear combination of several Gaussian components. To evaluate the quality of the model parameter θ , the probability density associated with the observed data is calculated [Maugis et al., 2009]:

$$f(X, \theta) = \sum_{j=1}^k a_j N_j(x_i | \mu_j, \sigma_j) \quad (2.3)$$

where, k is the number of Gaussian; a_j is the weight of the j^{th} Gaussian. It verifies the probability conditions such as $\sum_{j=1}^k a_j = 1$, and $0 \leq a_j \leq 1$; $N_j(x_i | \mu_j, \sigma_j)$ expresses the probability that a measure x_i is associated with the j^{th} Gaussian. In this context, the parameters of the model are collectively represented as $\theta = \{a_j, \mu_j, \sigma_j\}$.

In the literature, there are several methods for estimating these GMM parameters. In the present paper, the Maximum Likelihood method is used to determine the parameter θ .

The main idea of Maximum Likelihood Estimation (MLE) is to find a set of parameter estimates, so that the likelihood of the sample used is maximized. Since the n observations are independent, the likelihood function $L(X|\theta)$ can be written as [Bishop, 2006]:

$$L(X|\theta) = f(x_1, \theta) \times f(x_2, \theta) \times \dots \times f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (2.4)$$

Then substitute Eq. (2.3) into Eq. (2.4) leads to:

$$L(X|\theta) = \prod_{i=1}^n \sum_{j=1}^k a_j N_j(x_i | \mu_j, \sigma_j) \quad (2.5)$$

It is easier to maximize log-likelihood than the likelihood function itself. The log-likelihood function is written as:

$$\ln L(X|\theta) = \sum_{i=1}^n \ln \left(\sum_{j=1}^k a_j N_j(x_i | \mu_j, \sigma_j) \right) \quad (2.6)$$

The problem of estimation by the maximum likelihood method is therefore to find the roots of the following equation:

$$\frac{\partial \ln L(X|\theta)}{\partial \theta} = 0 \quad (2.7)$$

In general, maximizing the likelihood function does not have an analytical solution. It is then necessary to use iterative methods. The most common is to use the EM algorithm (Expectation-Maximization) [Feng et al., 2016]. However, the EM algorithm needs initial parameters; the most important of them is the number of classes k .

In general, the larger the number of parameters to be estimated into the model, the better the model approaches to the data. However, the complexity of the computation will also increase when there

are too many parameters to be estimated. Therefore, a compromise must be found; this is the goal of the the Bayesian Information Criterion (BIC). The BIC is defined as follow [Schwarz et al., 1978]:

$$BIC(k) = 2 \times \text{Max}(\ln(L(X|\theta))) - N_m \ln(n) \quad (2.8)$$

where, $\text{Max}(\ln(L(X|\theta)))$ is the maximized log-likelihood function, N_m is the number of parameters to be estimated in this model (3 per Gaussian: weight, mean and standard deviation), n is the number of observations, and k is the number of classes. In the BIC, the term (N_m) penalizes the complexity of the model. Therefore, BIC is maximized for most parsimonious setting, *i.e.*, with small N_m , and larger likelihood. Consequently, BIC is a good indicator to determine the number of classes required.

2.4 Determination of cluster number by Bayesian Information Criterion (BIC)

In this section, the previously proposed machine learning procedure by GMM is applied to database from four different machine-tools in two factories producing aeronautical structural parts:

- The machine-tool #1 is a Forest Aeromill for aluminum machining, in company #A;
- The machine-tool #2 is a Forest Aerostar for aluminum machining, in company #A;
- The machine-tool #3 is a Mazak Mazatrol for aluminum machining, in company #B;
- The machine-tool #4 is a Mazak i800T for titanium machining, in company #B.

An Emmatools device was installed on each machine-tool, and the machining data were collected for months during industrial productions. The classifications should determine:

- The spindle motion, according to the classification on the descriptor ΔN ;
- The Machine-tool motion, according to the classification on the descriptor ΔVf ;
- The material removal, according to the classification on the descriptors of $Arms$, P and ΔP .

The definition of the simple derivative ΔX ($X = N, Vf$ and P) is presented in section 2.1. For each machine-tool, the determination of the number of context classes is crucial and required for the Gaussian model that follows (in order to identify the classification thresholds for the business rules). The proposed method is:

- Apply GMM for each descriptor ($\Delta N, \Delta Vf, Arms, P$ and ΔP) for different numbers of classes (1, 2, ..., k).
- Compute the BIC, according to Eq. (2.8), in relation to the number of classes for each descriptor.
- Chose the minimal number of context classes required

In the present case, each BIC value converges to a maximum. In this work, a class number based on 90% of the BIC maximum is chosen. In this way, the class number is reduced and enables better interpretability of the results.

2.4.1 BIC on the variation of spindle speed

The Gaussian model analyzed by BIC on the variation of spindle speed ΔN is done on the four machine-tools. Fig. 2.9 shows the results: the BIC value reaches 90% of the BIC maximum with 2 classes. Thus, ΔN was classified into 2 classes, 2 clusters that correspond to the constant and varying spindle rotation.

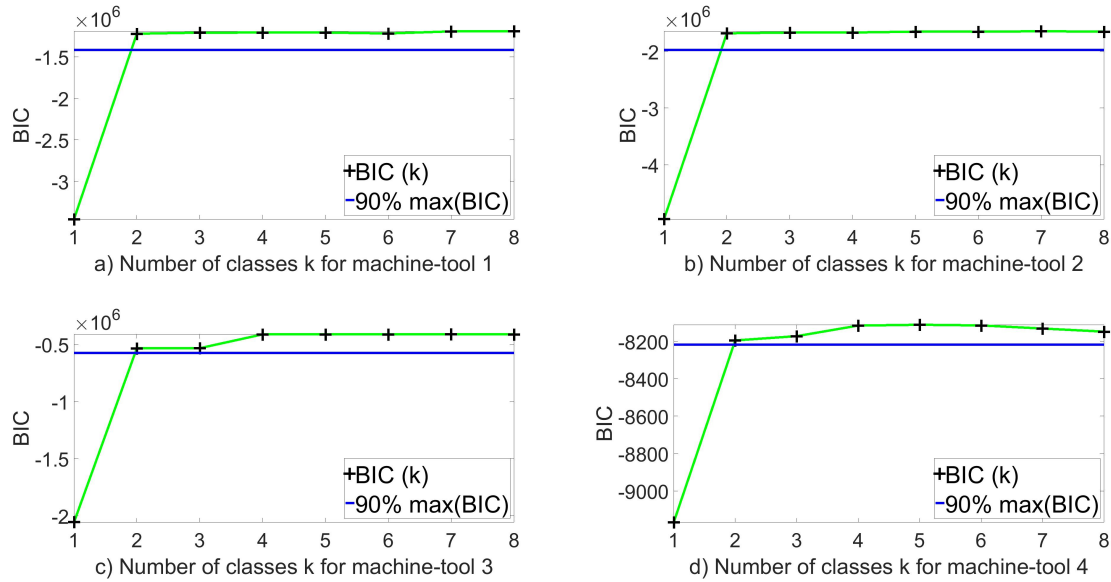


Figure 2.9 – BIC on ΔN issued from the 4 machine-tools

2.4.2 BIC on the variation of feedrate

The Gaussian model analyzed by BIC on the variation on ΔVf are performed on four machine-tools as shown in Fig. 2.10. The figure shows that the BIC value reaches 90% of the BIC maximum with 2 classes. Thus, ΔVf was classified into 2 classes, 2 clusters that correspond to constant and varying motion of the machine-tool.

2.4.3 BIC on vibrations and power

The descriptors $Arms$ (vibration) and P (spindle power) are used to determine if the tool is removing materials or not. To determine the number of classes required, Gaussian model analyses by BIC on $Arms$ as shown in Fig. 2.11 and P Fig. 2.12. In aeronautic industry, tool only cuts when the spindle speed N rotates at constant speed and feedrate Vf it not null. Therefore, the BIC method are applied to $Arms$ and P when N constant and $Vf > 0$. 4 different machine-tools are used in order to validate the results.

The figure shows that the BIC value reaches 90% of the BIC maximum with 3 classes for those 2 descriptors. Thus, $k=3$ and the vibrations $Arms$ and the power P were classified into 3 classes. The 3 classes are interpreted as: not cut, low material removal and high material removal.

The GMM is initialized by random assignment of k classes of data points. Once the number of classes

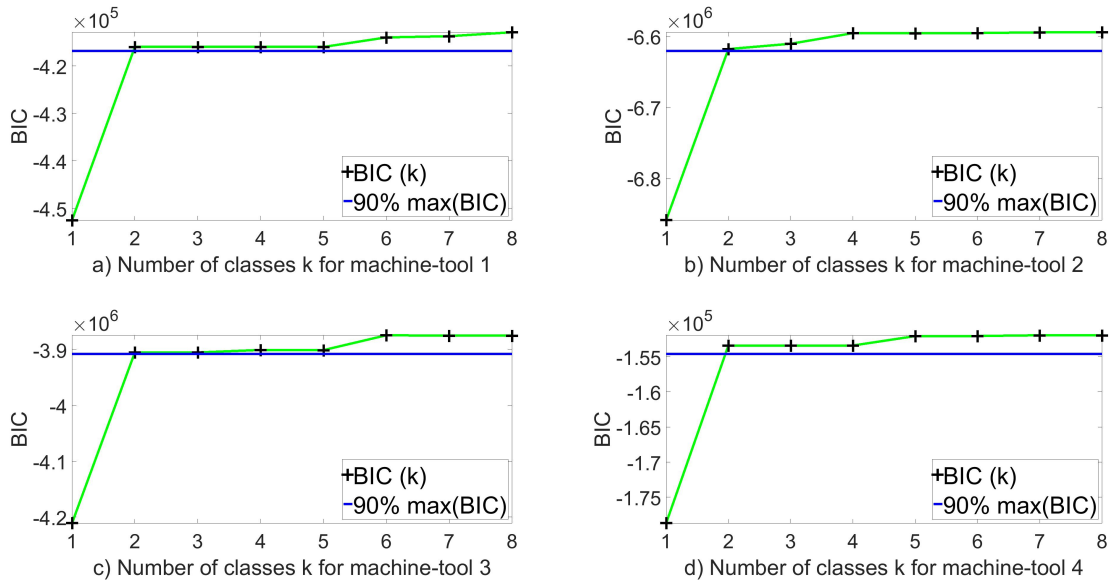


Figure 2.10 – BIC on ΔVf issued from the 4 machine-tools

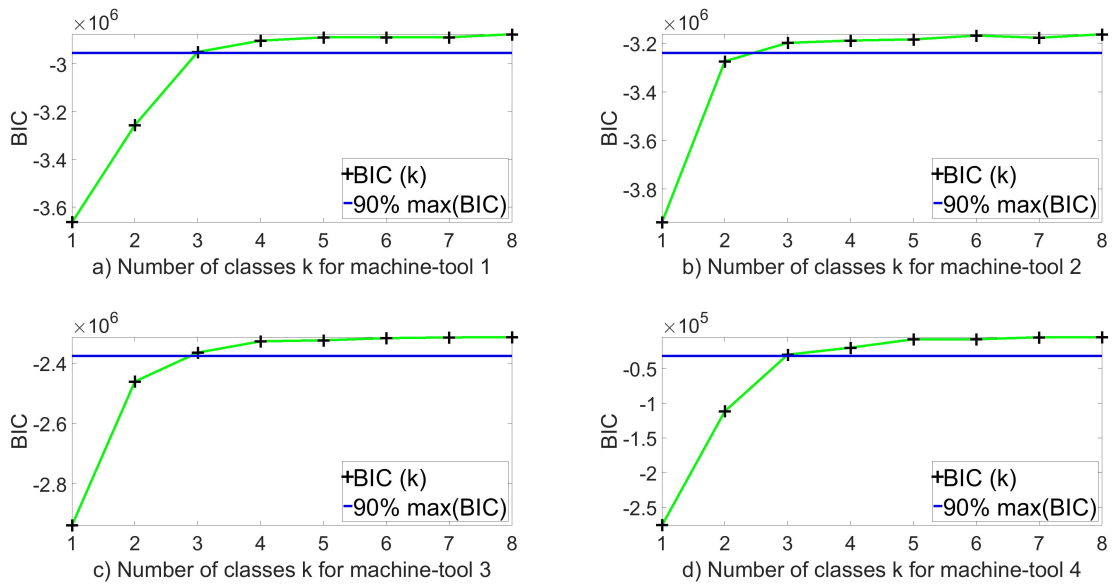


Figure 2.11 – BIC on *Arms* (when N constant et $Vf > 0$) issued from the 4 machine-tools

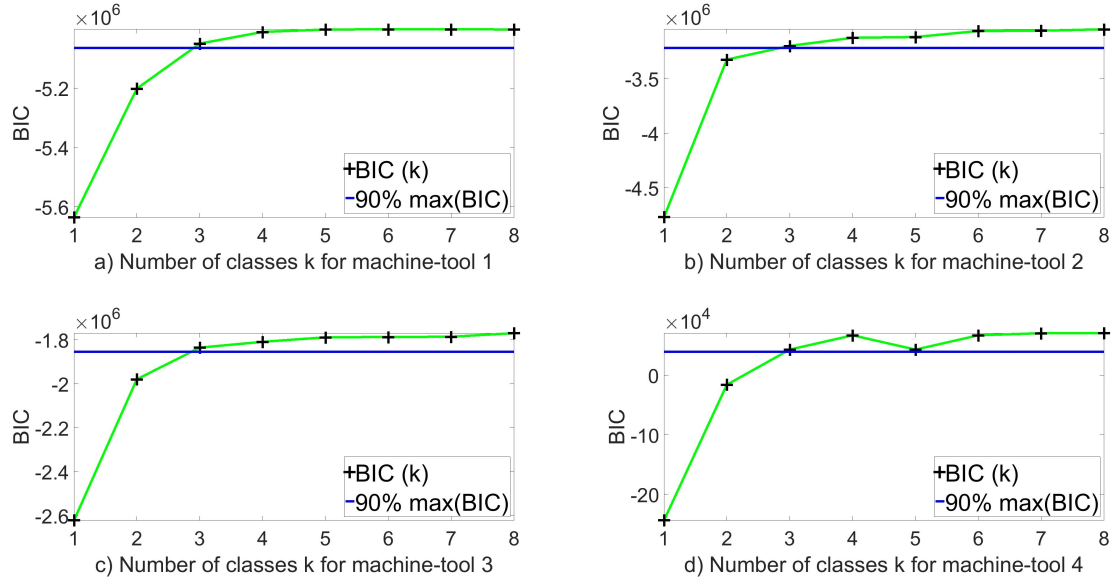


Figure 2.12 – BIC on P (when N constant et $V_f > 0$) issued from the 4 machine-tools

k is determined for GMM, k classes of data points were picked at random, and the mean as well as the standard deviation of each class were taken as the initial values of GMM model, and equal prior probabilities were assigned to the initial mode before running EM.

2.5 Business rules for contextual classification

2.5.1 Contextual classification of machine-tool

As the silhouette method applied to the ΔN , the spindle motion could be classified as 'rotating at constant speed' and 'rotating at varying speed' except the classes 'spindle stopped'. Using the BIC method in the ΔVf , the Machine-tool motion could be classified as 'moving at constant speed' and 'moving at varying speed' except the classes 'Machine-tool stopped'. Applying the BIC method to acceleration $Arms$ and the spindle power P , the material removal could be classified as 'tool not cutting', 'tool cutting'. Combined with the BIC applied to spindle power variation ΔP (in kW/s), the classes 'tool cutting' could be classified into two sub classes: 'tool cutting with constant tool engagement' and 'tool cutting with varying tool engagement'. As a result, the contextual classification for the machine-tool could be defined as below:

- if the spindle is stopped (S) or rotating at constant speed (CS) or varying speed (VS);
- if the machine-tool is stopped or moving at constant or varying speed;
- if the tool is cutting materials or not (N), with a constant (CC) or varying tool engagement (CV).

From these three contextual information, 17 potential states for the machine-tool can be defined (cf. Table 2.3). The objective is to determine the state of the machine-tool at each instant. Indeed, it will

enable accurate data selection and analyses dedicated to the considered state.

Table 2.3 – Classification of the different states of a machine-tool

		Feedrate of the machine-tool Vf						
		Vf=0		Vf>0				
				ΔVf > TΔVf			ΔVf ≤ TΔVf	
				Arms ≤ TArms & P ≤ Pidle	Arms > TArms or P > Pidle	Arms ≤ TArms & P ≤ Pidle	Arms > TArms or P > Pidle	ΔP ≤ TΔP
Spindle speed N	N=0	Spindle: S Machine: S Cutting: N	Spindle: Stopped (S) Machine-tool: Varying Speed (VS) Cutting: No (N)			Spindle: Stopped (S) Machine-tool: Constant Speed (CS) Cutting: No (N)		
	N > 0 ΔN > TΔN	Spindle: VS Machine: S Cutting: N	Spindle: VS Machine: VS Cutting: N	Spindle: VS Machine: VS Cutting: eng. Constant (CC)	Spindle: VS Machine: VS Cutting: eng. Varying (CV)	Spindle: VS Machine: CS Cutting: N	Spindle: VS Machine: CS Cutting: eng. Constant (CC)	Spindle: VS Machine: CS Cutting: eng. Varying (CV)
	N > 0 ΔN ≤ TΔN	Spindle: CS Machine: S Cutting: N	Spindle: CS Machine: VS Cutting: N	Spindle: CS Machine: VS Cutting: CC	Spindle: CS Machine: VS Cutting: CV	Spindle: CS Machine: CS Cutting: N	Spindle: CS Machine: CS Cutting: CC	Spindle: CS Machine: CS Cutting: CV

These classifications will be performed focusing on a variable X (e.g., feedrate Vf , spindle speed N or spindle power P) and also on the variation of X , noted ΔX (e.g., ΔVf or ΔN), which is defined in the 2.2. The objective is to calculate a classification threshold $T_{\Delta X}$ by unsupervised machine learning methods. GMM performs better in this context of research work. The classifications are then carried out based on the following rules:

- There are three clusters for machine-tool movements: stopped ($Vf = 0$), moving at constant speed ($Vf > 0$ and $\Delta Vf \leq T_{\Delta Vf}$), or varying speed ($Vf > 0$ and $\Delta Vf > T_{\Delta Vf}$);
- There are three clusters for spindle rotation: stopped ($N = 0$); rotating at constant speed ($N > 0$ and $\Delta N \leq T_{\Delta N}$); or varying speed ($N > 0$ and $\Delta N > T_{\Delta N}$).

To determine if the tool is cutting materials or not, a classification according to the criteria $Arms$ (root mean square-RMS of vibration acceleration value, in m/s^2) and the spindle power P (kW) is proposed, when the machine is moving (feedrate $Vf > 0$). In machining, it is known that if the tool is cutting, its vibration ($Arms$) is greater than if the tool is not cutting. It is also known that when the tool is cutting, the power (P) will exceed the spindle power corresponding to idle rotation (P_{idle}). However, there is a subclass called *spindle startup* where the P is also greater due to the acceleration of N . Moreover, there is another subclass called *post spindle startup* where the $Arms$ is also greater due to the lubrication spindle startup. These two subclasses should be identified and assigned to the cluster *tool not cutting*. The proposed business rules are:

- if ($P > P_{idle}$ or $Arms > T_{Arms}$) and $Vf > 0$, except the subclass *spindle startup* and the subclass *post spindle startup*, the tool is cutting material;
- if $P \leq P_{idle}$ and $Arms \leq T_{Arms}$, including the subclass *spindle startup* and the subclass *post spindle startup*, the tool is not cutting material.

Table 2.4 – Classification of the different operations when the tool is cutting

Arms $\leq T_{Arms}$ and P $\leq P_{idle}$	Arms $> T_{Arms}$ or P $> P_{idle}$				
	A=0 and C=0			A \neq 0 or C \neq 0	
	$\Delta Z=0$	$\Delta Z \neq 0$		$\Delta A=0$ and $\Delta C=0$	$\Delta A \neq 0$ or $\Delta C \neq 0$
		$\Delta X=0$ and $\Delta Y=0$	$\Delta X \neq 0$ or $\Delta Y \neq 0$		
Cutting : No (N)	Cutting : $2\frac{1}{2}$ milling	Cutting : vertical plunge	Cutting : ramping	Cutting : Pos. 5 axes	Cutting : Con. 5 axes

The following procedure aims to check if the GMM classifications are correct or not:

- Modeling : The contextual classification thresholds are determined, using GMM on different descriptors (ΔN , ΔVf , $Arms$, P and ΔP). Each machine-tool has different characteristics, so it is necessary to perform an initial training for each machine-tool. One-day data of a machining database is used for this purpose.
- Classifying: The classifications are carried out using GMM threshold and the business rules.
- Validation: The quality of classifications is assessed by confusion matrix, using a vector labeled with actual classes of machine-tool states.

2.5.2 Contextual classification per type of machining operation

It is also possible to determine the different types of machining operation according to the classification on the tool position (X , Y , Z , A and C) as well as the position variation (ΔX , ΔY , ΔZ , ΔA and ΔC). (cf. Table 2.4):

- *milling with $2\frac{1}{2}$ axes;*
- *cutting with vertical plunge;*
- *cutting with plunge by ramping;*
- *cutting with positioned 5 axes;*
- *cutting with continuous 5 axes;*

2.5.3 Contextual classification per type of tool engagement

It is even possible to know the different states of the tool engagement when the tool is cutting according to the classification on the spindle power variation ΔP (kW/s) and the feedrate variation ΔVf (mm/s²). (cf. Table 2.5) The business rule are:

- In the class when ($P > P_{idle}$ or $Arms > T_{Arms}$), there is a subclass called *spindle startup* where the P is also greater due to the acceleration of N , there is also another subclass called *post spindle startup* where the $Arms$ is also greater due to the lubrication spindle startup but

Table 2.5 – Classification of the different states of the tool engagement

Arms > T _{Arms} or P > P _{idle}					
P ≤ P _{idle}	P > P _{idle}				
Spindle startup & post spindle startup	ΔP > T _{ΔP}				ΔP ≤ T _{ΔP}
	Cutting: entry mat.	Cutting: exit mat.	Cutting: during turns	Cutting: varying eng.	Cutting: constant eng.

the spindle power is less than the idle rotation power P_{idle} . These two subclass should be identified and assigned to the cluster *tool not cutting*;

- In the class when ($P > P_{idle}$ or $Arms > T_{Arms}$) except the subclass *spindle startup* and the subclass *post spindle startup*, if $\Delta P \leq T_{\Delta P}$, tool is cutting with constant tool engagement.
- if $\Delta P > T_{\Delta P}$, the tool is cutting with varying tool engagement. Moreover, this class could be classified into 4 subclasses. *Tool cutting during entry materials* is defined as the period between the transition point of *not cutting / cutting* until the first point when the spindle power doesn't increase. It is known that, the spindle power will increase during tool entry materials;
- The same method, the subclass *tool cutting during exit materials* is defined as the period between the transition point of *cutting / not cutting* back to the point when the spindle power stops decreasing because the spindle power will decrease during tool exit materials;
- The subclass *tool cutting during turns* is defined as the points when the variation of feedrate ΔVf is greater than the threshold $T_{\Delta Vf}$. It is known that when the tool is cutting during turns, the machine-tool will brake (Vf is much varied)

2.6 Contextual classification of the material removal

Once the class number determined by BIC, the classification thresholds are identified from one-day data by GMM. Then, the business rules are applied, leading to the classification. This section presents the classification. GMM is applied to the collected data in the context of *spindle rotation at constant speed* and *Machine-tool motion* ($Vf > 0$) see Table 2.3. Two datasets illustrate the approach: machine-tools #2 and #3 for aluminum machining.

2.6.1 Modeling and classification for machine-tool #2

The *Arms* distribution is modeled by 3 Gaussians (see section 2.5):

- a Gaussian represents the class *tool not cutting* (Y1 green),
- another the class *tool cutting* (Y3 blue),
- the 3rd Gaussian represents '*severe machining vibrations*' (Y2 red).

The distribution of Y4 (in cyan) is the sum of these 3 Gaussians ($Y4=Y1+Y2+Y3$) as shown in Fig. 2.13(a). The threshold T_{Arms} is defined as the intersection between the Gaussians Y1 and Y3, in

order to minimize classification errors (false positives and negatives). This enables to identify $T_{Arms} = 5.43 \text{ m/s}^2$ for the machine-tool #2.

Applying the same method, P is modeled by 3 Gaussians:

- one Gaussian represents the class *tool not cutting* (Y1 green),
- another represents the class *tool cutting with low material removal* (Y3 blue),
- the 3rd Gaussian (Y2 red) models the class *tool cutting with high material removal*.

The distribution of Y4 (cyan) is the sum of these 3 Gaussians ($Y4=Y1+Y2+Y3$), as shown in Fig. 2.13(b). The threshold is chosen at the intersection between the Gaussian *tool is not cutting* and the other Gaussian *tool cutting with high material removal*. This leads to $P_{idle} = 2.57 \text{ kW}$.

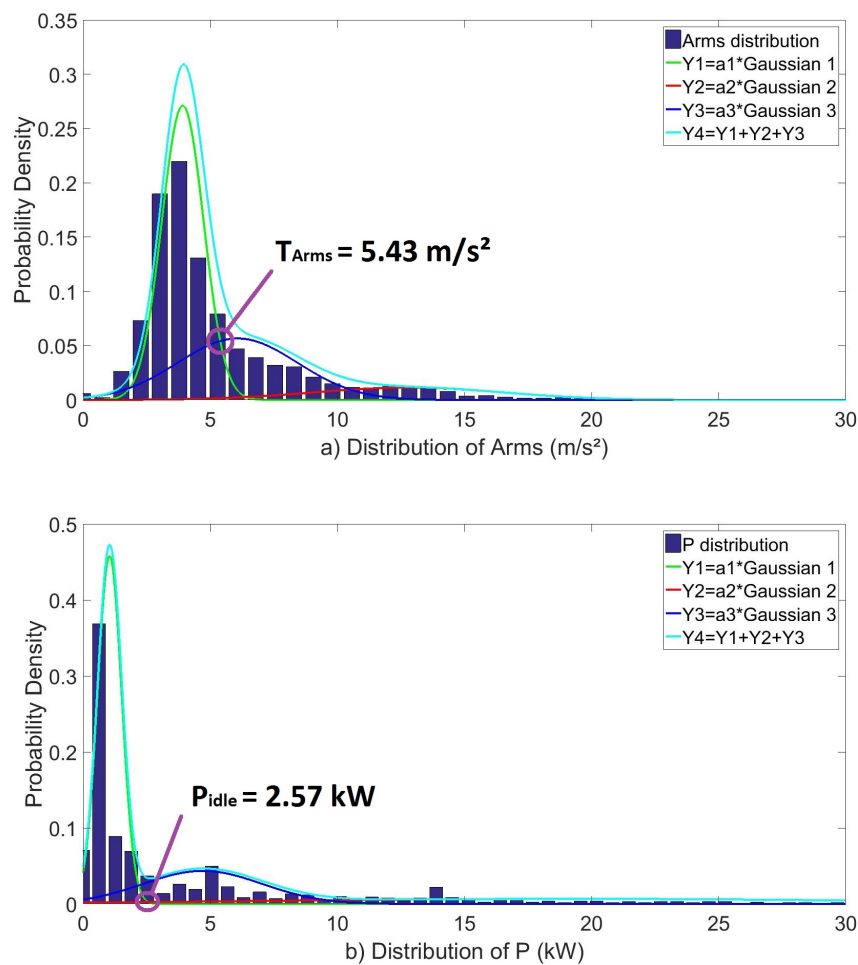


Figure 2.13 – Distribution of Arms and P (when N constant and $V_f > 0$) modeled by GMM for machine-tool #2

In summary, three classes of the tool engagement are defined.

- *tool cutting* when $Arms > T_{Arms}$ or $P > P_{idle}$; it consists of two subclasses : *tool cutting with constant engagement* and *tool cutting with varying engagement*;
- *tool not cutting* when $Arms \leq T_{Arms}$ and $P \leq P_{idle}$ by including the subclass *spindle startup* and the subclass *post spindle startup*.

Fig. 2.14 presents data from machine-tool #2 as well as the result of the classification. The *tool not cutting* class is represented by black '+' signs on the P power curve, the *tool cutting* class is presented by red and blue dots. The *tool cutting* class can be classified into two sub-classes: *tool cutting with constant engagement* (in red points) and *tool cutting with varying engagement* (in blue points) by classification on ΔP (see Table 2.3). There is a subclass called *spindle startup* when the spindle speed N is accelerating to the target value ($N=23720$ RPM in Fig. 2.14). This class is represented by green '+''. Moreover, when the spindle has reached its target speed, the tool does not cut immediately. There is a subclass called *post spindle startup*. This class is also represented by green '+''. It is defined by the following rule: from the reach of the target speed N until the power exceeds the no-load power P_{idle} (which represents the beginning of machining). Identify these two states permits to eliminate parasitic vibrations related to lubrication spindle startup and also eliminate parasitic power related to N acceleration, which would otherwise lead to classification errors. The subclass *spindle startup* and *post spindle startup* are also assigned to the cluster *tool not cutting*.

Cyan line represents Vf , blue line N , pink line $Arms$. The tool engagement is correctly detected.

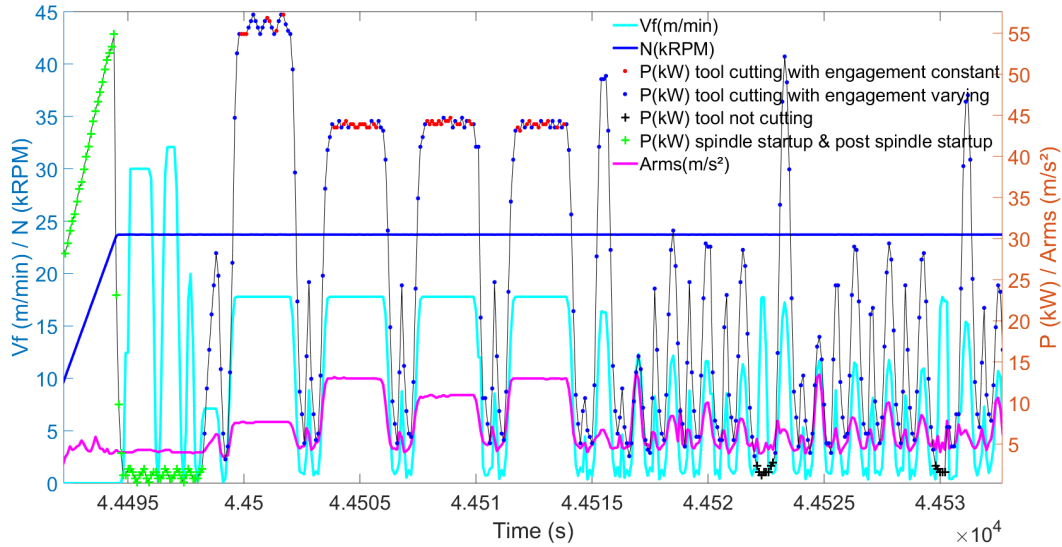


Figure 2.14 – Classifications of material removal for machine-tool #2

It is even possible to know the different states of the tool engagement when the tool is cutting according to the classification on the spindle power variation ΔP (kW/s) and the feedrate variation ΔVf (mm/s^2). (cf. 2.5).

To verify the classification by manual data mining, N (kRPM), Vf (m/min), $Arms$ (m/s^2) and P (kW) will be plotted on the same figure. The seven clusters of power P will be represented in different colors on the same figure. One day of data is plotted and checked, and an area is shown as an example

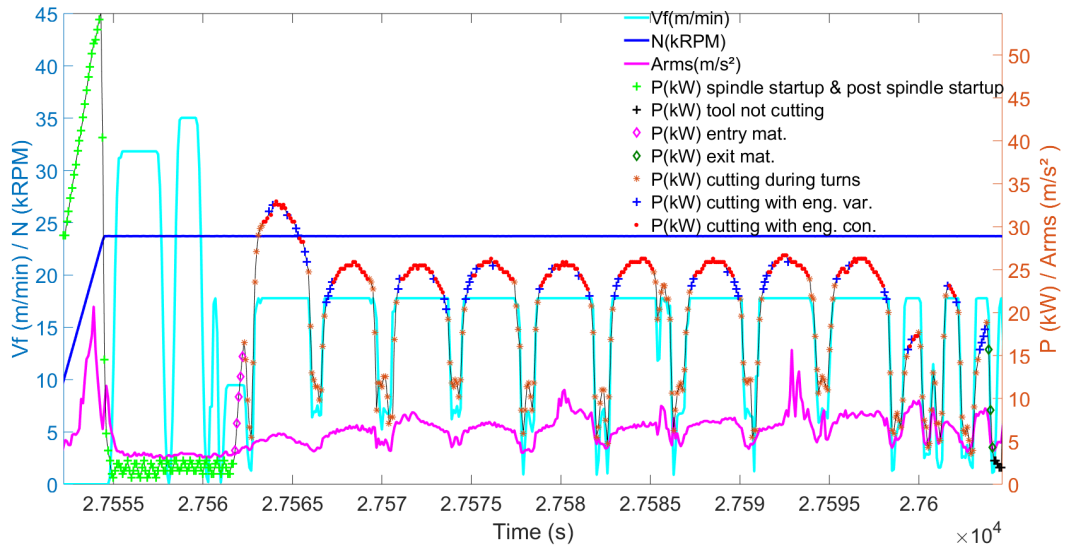


Figure 2.15 – Classifications of different states of tool engagement for machine-tool #2

in Fig. 2.15. V_f is in cyan line, N is in blue line, $Arms$ is in pink line. The cluster 'tool not cutting' are represented by black '+' signs on the power curve P , 'spindle startup' and 'post spindle startup' are presented by green '+', 'tool entry materials' are presented by pink diamonds, 'tool exit materials' are presented by dark green diamonds, 'tool cutting during turns' are presented by brown stars, 'tool cutting with engagement constant' are represented by red dots, and 'tool cutting with engagement varying' are presented by blue '+' signs. The classification of different states of tool engagement when 'tool is cutting' is well classified. The validation will be presented in the section 2.7.

2.6.2 Modeling and classification for machine-tool #3

The distribution of $Arms$ in this class (N constant and $V_f > 0$) is modeled by 3 Gaussians (see section 2.4).

- a Gaussian represents the class *tool not cutting* (Y1 green),
- another the class *tool cutting* (Y3 blue),
- the 3rd Gaussian represents 'severe machining vibrations' (Y2 red).

The distribution of Y4 (in cyan) is the sum of these 3 Gaussians ($Y4=Y1+Y2+Y3$) as shown in Fig. 2.16(a). The threshold T_{Arms} is defined as the intersection between the Gaussians Y1 and Y3, in order to minimize classification errors (false positives and negatives). The threshold is identified: $T_{Arms} = 5.93 \text{ m/s}^2$ is obtained.

Applying the same method, P is modeled by 3 Gaussians:

- one Gaussian represents the class *tool not cutting* (Y1 green),
- another represents the class *tool cutting with low material removal* (Y3 blue),

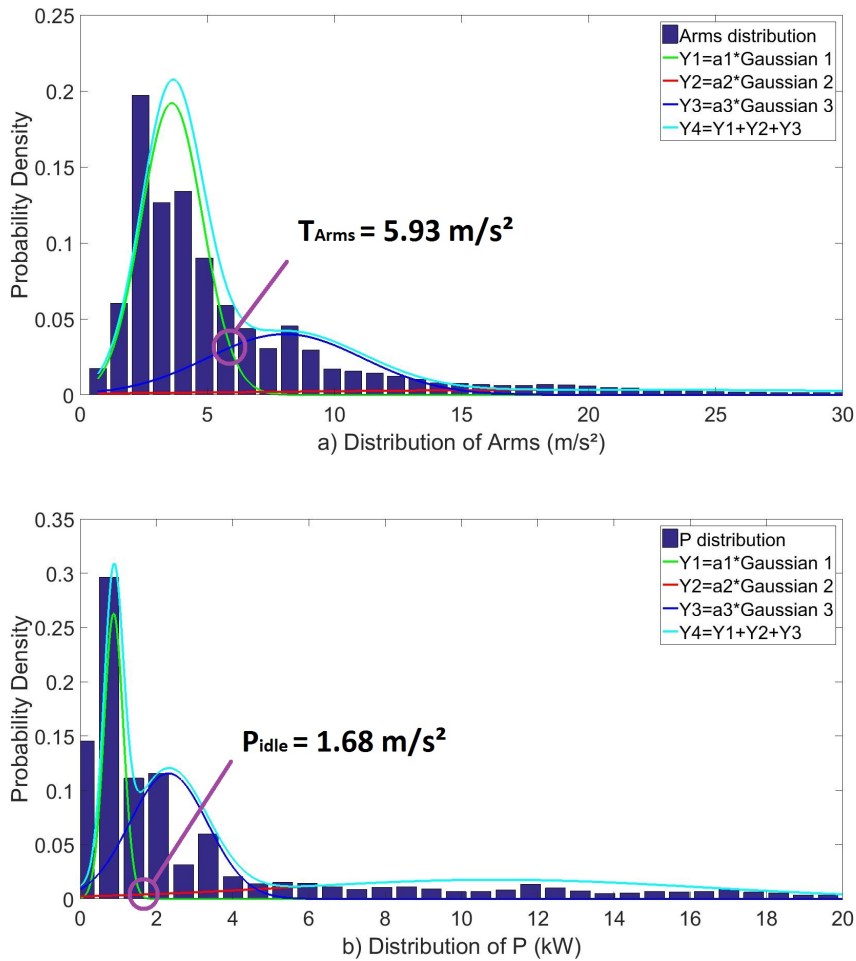


Figure 2.16 – Distribution of Arms and P (when N constant and $V_f > 0$) modeled by GMM for machine-tool #3

- the 3rd Gaussian (Y2 red) models the class *tool cutting with high material removal*.

The distribution of Y4 (cyan) is the sum of these 3 Gaussians ($Y4=Y1+Y2+Y3$), as shown in Fig. 2.16(b). The threshold is chosen at the intersection between the Gaussian *tool is not cutting* and the other Gaussian *tool cutting with high material removal*, thus obtaining $P_{idle} = 1.68 \text{ kW}$.

A combination of the two sensors measurements is performed. Thus, for data where *spindle is rotating at constant speed and axes are moving* ($V_f > 0$) see Table 2.3, if $Arms > T_{Arms}$ or $P > P_{idle}$, except the subclass '*spindle startup*' and the subclass '*post spindle startup*', the data is assigned to the class *tool cutting*.

Fig. 2.17 plots the N (RPM), V_f (m/min), $Arms$ (m/s^2) and P (kW) as well as the classification results in classes (dots and + sign) for machine-tool #3. The material removal is well classified. The evaluation of the classification results will be presented in the section 2.7.

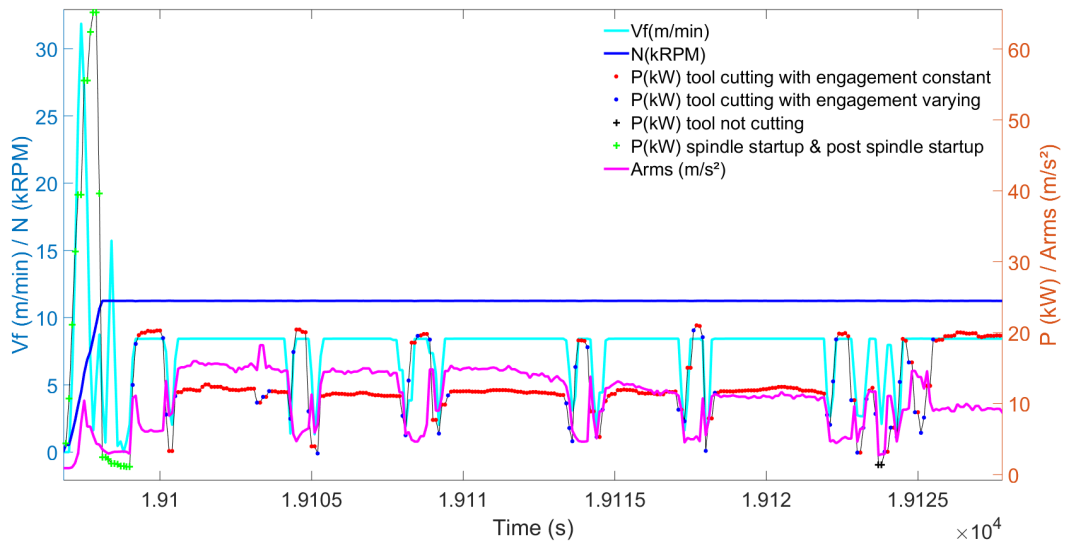


Figure 2.17 – Classifications of material removal for machine-tool #3

2.7 Method validation

The estimation of the quality of the classification by machine learning is essential to choose an optimal descriptor or combining descriptors for a given data set [Wolpert, 1992]. Two criteria constitute the quality: accuracy and sensitivity.

2.7.1 Accuracy

To estimate the accuracy of the classification by GMM, manual data mining was used. In our approach, the classifications based on GMM model conduct to the predicted classes (*tool cutting* or *tool not cutting*). The actual classes correspond to 100% true classes that were constituted by manual data mining. The main steps are as follows:

- Apply GMM for training on a one-day data and identify the thresholds of T_{Arms} and P_{idle} ;
- Choose a data period and use the two thresholds to the set to predict classes: *tool cutting* and *tool not cutting*;
- Go through manual data mining, one point by one point, to obtain the actual classes;
- Compare the predicted with the actual classes and calculate the confusion matrix.

As shown in 2.2, the confusion matrix could be used to evaluate the classification results.

Table 2.6 reports the confusion matrix during 4 hours for the machine-tool #2.

The false positives (FP) in the Table come from the threshold P_{idle} . The threshold P_{idle} here is calculated once for all tools, While some tools consume less power during idle rotation. Which leads to the period '*post spindle startup*' is wrong labeled as '*too cutting*.' During low material removal, the threshold P_{idle} is too high for some tools, which leads to some points labeled as '*tool not cutting*';

Table 2.6 – Confusion matrix for 4 hours data from the machine-tool #2

Confusion Matrix		Actual Classes	
		Tool Cutting	Tool Not Cutting
Predicted Classes	Tool Cutting	TP=54 781	FP=263
	Tool Not Cutting	FN=845	TN=88 111

however, they are 'tool cutting'. They are false negatives (FN). The accuracy for the materials removal is obtained for the machine-tool #2 is: $ACC_{\#2}$ (4h)=99.23%, which is very good enough.

Table 2.7 presents the confusion matrix for 4 hours on the machine-tool #3. The accuracy for the materials removal is obtained for the machine-tool #3: $ACC_{\#3}$ (4h)=99.94%, which is also very good enough for contextual classification.

Table 2.7 – Confusion matrix for 4 hours data from the machine-tool #3

Confusion Matrix		Actual Classes	
		Tool Cutting	Tool Not Cutting
Predicted Classes	Tool Cutting	TP=51 476	FP=18
	Tool Not Cutting	FN=67	TN=92 470

Table 2.8 – Confusion matrix for the different tool engagement states during 4h on the machine-tool #2

Confusion Matrix		Actual Classes						
		Spindle startup & post spindle startup	Entry mat.	Exit mat.	Cutting turns	Cutting eng. var.	Cutting eng. con.	Not Cutting
Predicted Classes	Spindle startup & post spindle startup	786	0	0	0	0	0	0
	Entry mat.	6	8 568	1 212	9	0	0	33
	Exit mat.	0	50	5 199	19	3	2	45
	Cutting turns	0	0	2	1 9130	0	0	1
	Cutting eng. var.	0	0	0	1	2 871	158	0
	Cutting eng. con.	0	0	0	0	20	1 8517	0
	Not Cutting	33	6	2	3	0	0	8 7325

Table 2.8 presents the confusion matrix of the different tool engagement states for 4 hours on the machine-tool #2. There are some classification errors come from the wrong label between 'tool entry materials' and 'tool exit materials'. This is because there are some periods having 3 points continues when 'tool is cutting'. The errors exit when all these 3 points are labeled as 'tool entry materials' or

all these 3 points are labeled as 'tool exit materials'. The errors coming from 'cutting with constant engagement' and 'cutting with varying engagement' is due to the global threshold $T_{\Delta P}$. The threshold $T_{\Delta P}$ is a little high for some tools while it would be a little small for other tools. To evaluate the classification results, the accuracy will be computed. As defined in Eq. (2.2), the accuracy (ACC) is defined as the division between the sum of the correct estimations (TP + TN) and the sum of all estimations (TP + TN + FP + FN). In Table 2.8, the sum of the correct estimations (TP + TN) is equal to the sum of the diagonal values, and the sum of the false estimation (FP + FN) is equal to the sum of the values in the table except for those diagonal values.

Also, the accuracy for the different states of tool engagement during 4 hours for the machine-tool #2 is: $ACC_{Eng.\#2}(4h)=98.89\%$ The classifications obtained using GMM method are very accurate.

2.7.2 Sensitivity of the method

To test whether the method is sensitive to the training data set, the cross-validation method can be used. Kohavi et al. [1995] present different cross-validation methods like 'Holdout', 'Leave-one-out', 'k-fold' and 'Stratification'. Here the original sample is divided into k samples, then one of the k samples is selected as the validation set and the $(k - 1)$ other samples will constitute the training set. The performance score is calculated. The main steps are as follow:

1. Apply GMM to the validation set to obtain its thresholds of $Arms$ and P ,
2. Apply these thresholds to the validation set (one day) to obtain the predicted classes,
3. Go through the manual data mining, check the classification one point by one point, to obtain the actual classes,
4. Compare the predicted classes from the validation set to the actual classes and calculate the confusion matrix. The accuracy for the thresholds from the validation set is obtained,
5. Apply GMM to the $(k - 1)$ sets, to obtain $(k - 1)$ sets of thresholds $Arms$ and P ,
6. Apply these $(k - 1)$ sets of thresholds from training sets to the validation set, to obtain $(k - 1)$ sets of predicted classes.
7. Compare these $(k - 1)$ predicted classes with the actual classes and calculate the $(k - 1)$ confusion matrix separately. The accuracy of the thresholds from each training set is obtained.

Here 11 days were chosen ($k=11$) as the original sample.

The thresholds of $Arms$ and P for each day in the training sets (10 days) as well as corresponding accuracy when applying these thresholds to the validation set for the machine-tool # 2 and the machine-tool # 3 are respectively reported in Fig. 2.18 (a and b). The threshold of $Arms$ and P for the validation set (one day) as well as corresponding accuracy when applying these thresholds to the validation set are represented as the points in the '0' row in the two figures for these two machine-tools.

With 10 training sets, the mean accuracy of the classification for the machine-tool #2: $ACC(k\text{-fold})_{\#2}(4h)=96.28\%$, and for the machine-tool #3: $ACC(k\text{-fold})_{\#3}(4h)=97.60\%$. This shows that the thresholds are not sensitive to which day is chosen to apply GMM. The mean accuracy is lower than the one using the same set for the validation set. It is because the thresholds for *materials*

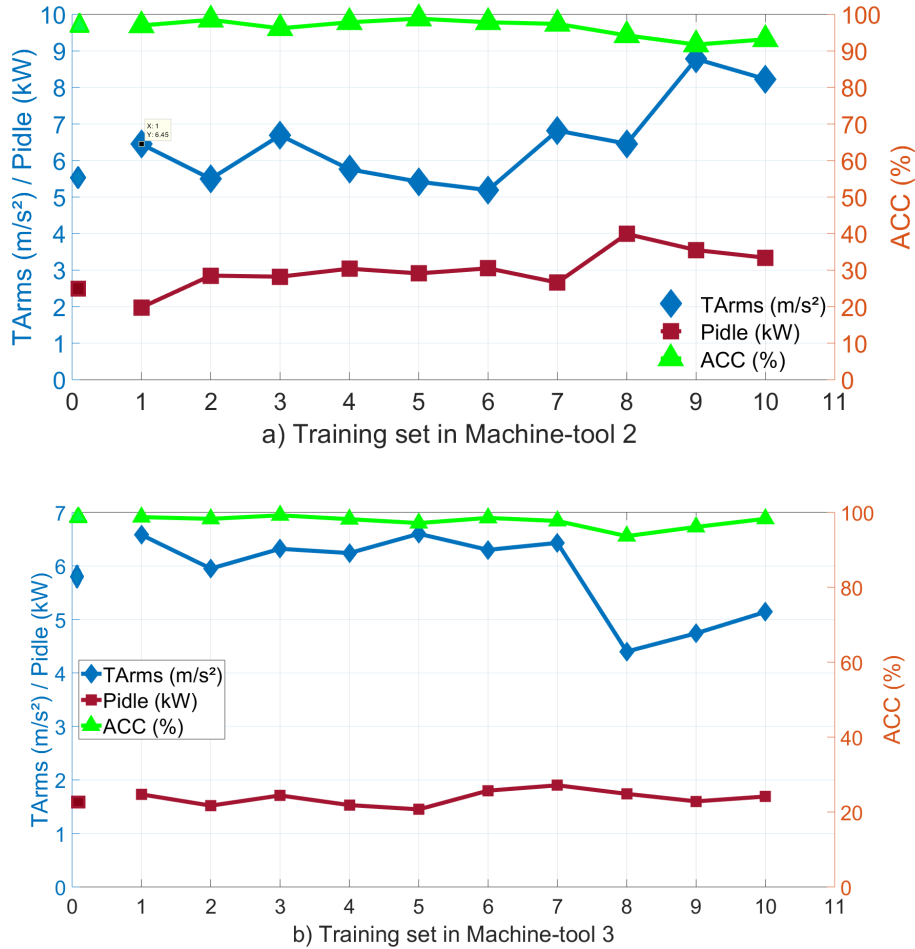


Figure 2.18 – The new thresholds of $Arms$ and P obtained in the training sets as well as these thresholds obtained in validation sets for (a) machine-tool #2 and (b) machine-tool #3

removal or not (T_{Arms} and P_{idle}) may slightly change as time goes by, due to the variation of the machine condition (e.g. wear of the spindle).

Nevertheless, the mean accuracy obtained is high enough that the thresholds learned by GMM during one-day can be used for any other day of production on the same machine-tool.

2.8 Conclusion

In machining, contextual classification is important to:

- Describe the machine-tool working condition;
- Decipher the operational context of machining incident (e.g. chatter), enabling reasoning about incidents and knowledge generation;
- Realize a more relevant calculation of Key Performances Indicators (KPIs).

Classification errors will conduct to false alarm or missed detection. This chapter proposes a contextual classification procedure in machining based on unsupervised machine learning. Two methods of unsupervised machine learning (K-means and GMM) were compared to test the classification of machine-tool movement. Which leads to the conclusion as below:

- K-means is not suitable for the unbalanced data with different size and different density (the population of ΔX around zero is too large). However, GMM can well classify the unbalanced data.
- Despite the fact that GMM can correctly classify data with unbalanced classes, its disadvantage is that the results of GMM directly depend on the choice of parameters (number of Gaussians).

Then, GMM is chosen as the appropriate unsupervised machine learning for the contextual classification. Before applying GMM, the optimal number of classes has been determined by the silhouette method based on the Bayesian information criterion (BIC), which leads to the results as below:

- ΔN should be classified into 2 clusters in GMM to know if the *spindle rotating at constant or varying speed*.
- ΔVf should be classified into 2 clusters in GMM to know if the *machine-tool moving at constant or varying speed*.
- To know the *tool cutting or not cutting materials*, the acceleration *Arms* and the spindle power *P* should be classified into 3 clusters.

Therefore, in relation to the optimal class number obtained by BIC, the meaning of each class has been interpreted, based on manufacturing expertise, and each class has been named. A complete set of business rules for contextual classification has then been proposed:

- Spindle motion: *Stopped (S)* or *rotation at constant speed (CS)* or *varying speed (VS)*.
- Machine-tool movement: *Stopped (S)* or *moving at constant (CS)* or *varying speed (VS)*.
- Material removal: *tool is not cutting (N)* or *cutting with a constant (CC)* or *varying tool engagement (CV)*.

It was also proposed to decipher the operation context of the machining operation and tool engagement. Five different operations of machining could be classified according to classification on the tool position (X, Y, Z, A and C) as well as the position variation ($\Delta X, \Delta Y, \Delta Z, \Delta A$ and ΔC): $2\frac{1}{2}$ *axes milling; vertical plunge; plunge by ramping; cutting with positioned 5 axes and cutting with continuous 5 axes*. Also, five different states of tool engagement could be detected according to the business rules: *tool entry material; tool exit materials; tool cutting during turns; tool cutting with varying engagement and tool cutting with constant engagement*.

A classification of the *tool cutting or not*, by combining GMM thresholds of vibration measurements *Arms* and power measurements *P*, was validated on 2 different machine-tools. The classification of five states of tool engagement when *tool is cutting* is validated by manual data mining on the machine-tool #2. Cross-validation for the tool removal enables us to conclude that the thresholds learned by GMM

during one-day can be applied to any other days for the same machine-tool. It was also found that the thresholds learned by GMM during one-day data may slightly change due to the variation of the machine condition. The learned values of thresholds given in here are specific to a given machine-tool, but the proposed machine learning methodology is replicable to any machine-tool.

This chapter proposes a contextual classification procedure in machining based on unsupervised machine learning by Gaussian mixture model. This method is validated by manual data mining and cross-validation. Then, some monitoring criteria or KPIs can be calculated in different classes based on the contextual classification. For example, the criterion of chatter is calculated in the class when '*tool is cutting materials*'; while the criterion of tool breakage is calculated when '*tool is not cutting*' [Godreau et al., 2019]. The contextual classification will be used in the following chapter for monitoring issues.

Detection of machining incidents based on GMM and business rules

Contents

3.1	Data presentation and cleaning	48
3.2	Improvement of chatter identification	53
3.2.1	Classification of harmonic frequency	53
3.2.2	Chatter / non Chatter	57
3.2.3	Test on a new database	59
3.3	Tool breakage	61
3.3.1	Business rule for detection of tool breakage	61
3.3.2	Application	63
3.4	Excessive vibrations	67
3.5	Conclusion	70

Summary

This chapter will present the business rules for incidents detection based on the contextual classification detailed in chapter 2, including the chatter, tool breakage, and excessive vibrations. The chapter starts with the presentation of the raw data structure collected by Emmatools as well as the identification of abnormal data and noise. Second section presents the improvement of chatter identification and the frequency classification in the frequency spectrum. Then the business rule for tool breakage based on two criteria (Ub when tool is not cutting and Ub' when tool is cutting), and well the adaptive threshold and the application will be presented in third part. Finally, excessive vibrations are considered in the fourth section, beginning with the business rule for excessive vibrations, as well as the application.

Part of the work presented in this chapter is published in the journal paper entitled "A knowledge-based digital shadow for machining industry in a digital twin perspective" [Ladj, Wang, Meski, Belkadi, Ritou, and Da Cunha, 2020], Journal of Manufacturing Systems, 2020.

3.1 Data presentation and cleaning

As described in chapter 1, a device, called Emmatools, collects the measured data during machining and stores it in a database [De Castelbajac et al., 2014]. Fig. 3.1 presents the schematic diagram of the Emmatools. It is installed in four different machine-tools in two factories producing aeronautical structural parts.

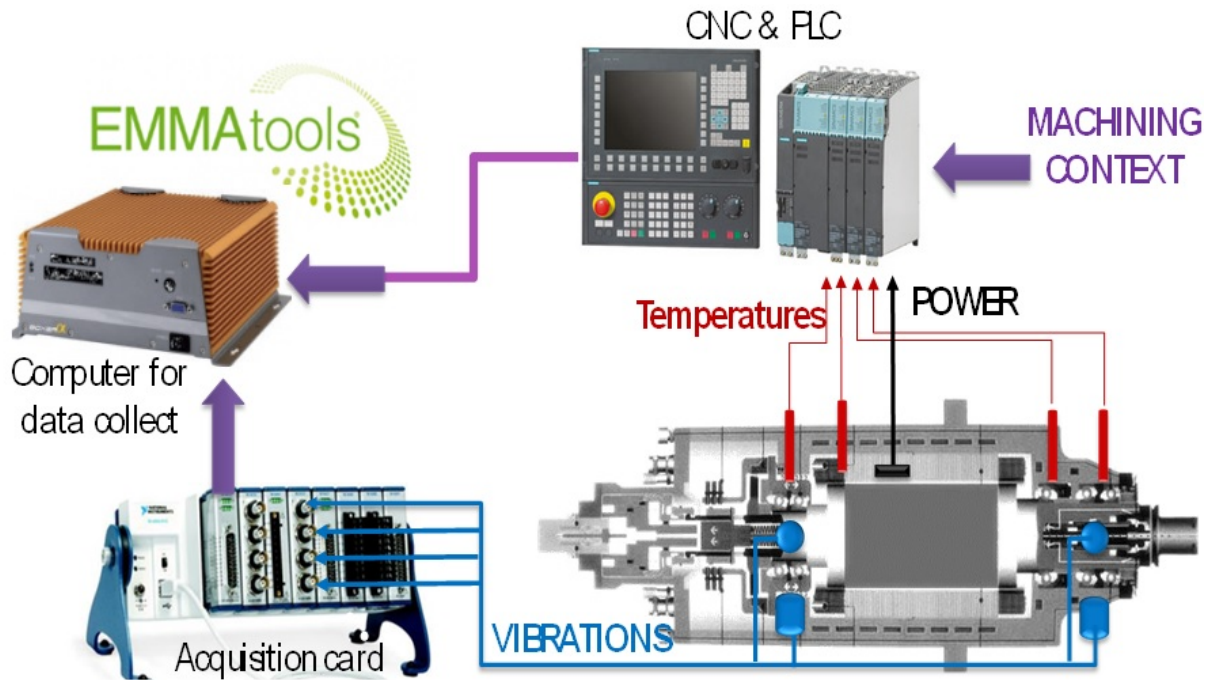


Figure 3.1 – Emmatools for process monitoring and data collection

Emmatools collects data every tenth of a second. The data comes from three information sources: the Computer Numerical Control (CNC) contextual information, the sensors integrated into the machine-tool, and the added vibrations sensors. By field bus, the CNC provides machining context data such as:

- the time scale, including the day, the time index (every 0.1 s);
- the machine mode;
- the program information, including the 'OF' (production order), 'IDprogP' (Principal program), 'IDprogC' (Current program), 'IDligne', 'IDbloc' and 'Numseq' (the three are the current positions in the G-code).
- the Tool, including the 'ID Tool' (tool reference), 'Tool Instance' (the tool number for the same tool reference in the tool shop), and the 'Tool DVP' (tool life).
- potentiometer information, including 'Pot rotation' (for spindle speed N , in %) and 'Pot feedrate' (for axis motion, in %).

The sensors integrated into the machine-tool collect the instantaneous information such as:

- the instantaneous spindle speed N (RPM);
- the instantaneous machine-tool feedrate V_f (m/min);
- the instantaneous spindle power P (kW);
- the Axis position as well as the Axis speed (m/min);
- the lubrication (in L/h) and the temperature ($^{\circ}C$)

The added vibrations sensors are the four accelerometers installed into the spindle (radially at each bearing). It includes the vibration level and the vibration frequency spectrum.

the vibration level:

- $Arms$ (root mean square - RMS of the vibration acceleration values, in m/s^2) for four accelerometers,
- $Vrms$ (vibration velocity root-mean-square, in mm/s) for four accelerometers,
- $Apic$ (maximum of vibration value over a given period, in m/s^2) for four accelerometers,
- $Vpic$ (maximum of vibration velocity in a given period, in mm/s) for four accelerometers,
- resultant of the variables above for the front bearing like Rav_Arms (in m/s^2) and Rav_Vrms (in mm/s), as well as for the rear bearing like Rar_Arms (in m/s^2) and Rar_Vrms (in mm/s),
- chatter criteria for front and rear bearing like $Chatter_av$ (in m/s^2) and $Chatter_ar$ (in m/s^2),
- unbalance value of shaft for front and rear bearing like $Balourd_av$ (in m/s^2) and $Balourd_ar$ (in m/s^2).

the vibration frequency spectrum:

- three harmonic frequencies for four accelerometers (in Hz),
- five non-harmonic frequencies for four accelerometers,
- three amplitudes (in m/s^2) corresponding to the harmonic frequencies for four accelerometers,
- five amplitudes corresponding to the non-harmonic frequencies for four accelerometers.

To illustrate the raw data, there are 864000 rows and 135 columns for just one day's data, representing a volume of 380 Mb. If the whole lifetime of a spindle (around one year) is considered, 140 Gb data can be collected. If several machine-tools are considered, there is an even bigger volume of data to mine. The objectives are to mine the raw data and aggregate it into simplified, smaller, and meaningful data.

As shown in the six steps of the CRISP-DM process [Wirth and Hipp, 2000], Fig. 3.2, the beginning step of the data mining is 'Business understanding' and 'Data understanding'. The 'Business understanding' is explained in chapter 2 as business rules. However, in the 'Data understanding', Emmatools collects the measured data during machining and stores it in a database. Then, the data mining methods will be used to recognize the noise data (e.g., abnormal values of accelerometers) and the missing data (e.g., loss of connection between CNC and sensing system).

CRISP-DM phases

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

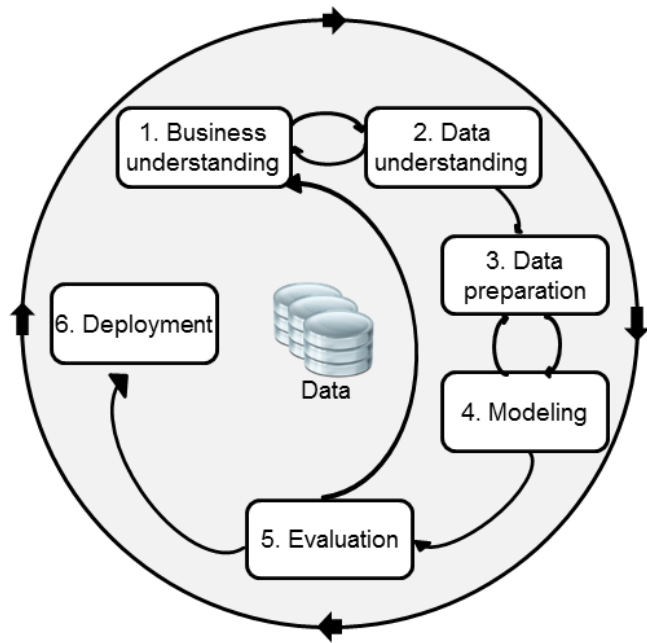


Figure 3.2 – The CRISP-DM life cycle[Wirth and Hipp, 2000]

The noise comes one part from the measurement problem, and another part from disturbing sources. The measurement problem can, for example, come from the wrong installation of accelerometers leading to the false measurement values. They should not be considered. Fig. 3.3 presents an example where the measurement of accelerometer Y in the rear bearing is missing in the machine-tool #1. It has been possible to integrate a fourth accelerometer at the rear of a given machine-tool spindle.

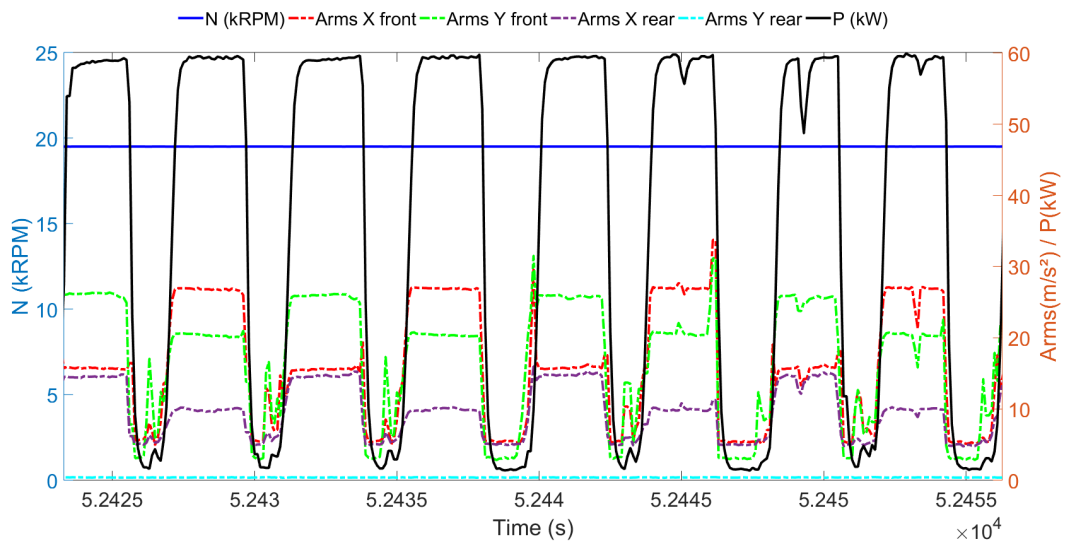


Figure 3.3 – Accelerometer Y in the rear bearing is missing for the machine-tool #1

As mentioned above, there are four accelerometers installed in the machine-tool. In Fig. 3.3, the *Arms* for each accelerometer are presented by different colors: X in the front bearing (red dashed

line), Y in the front bearing (green dashed line), X in the rear bearing (purple dashed line), and Y in the rear bearing (cyan dashed line). The spindle speed N (in blue line) and the spindle power P (in black line) indicate the machine-tool states (*tool cutting or not*). It is clear that there are enough differences between *tool cutting* and *tool not cutting* for the *Arms* of X front bearing, Y front bearing, X rear bearing. However, the *Arms* of Y rear bearing is negligible (around 0) no matter *tool cutting or not*. We can conclude that the accelerometer Y in the rear bearing is not well installed. Therefore, only the accelerometers X in front bearing, Y in front bearing and X in rear bearing will be considered in future data mining.

The noise coming from disturbing sources should be neutralized after the knowledge of its context. Disturbing sources can be classified into two clusters: sources from the principal programs for machining and sources from the preparation programs (not for machining). The preparation programs (such as 'tool change', 'warm up', 'tool measure', 'blowing and probing', etc.) are the programs to prepare the tools and workpieces before machining programs. The noise from these preparation programs is mostly due to normal vibrations. For example, during 'tool change', normal vibrations are due to the contact between the fixture of the tool and the tool shop, which is not a source of collision (between the tool and any other part of the machine or fixture [Godreau, 2017]) for machine-tool. Fig. 3.4 presents the three accelerometers: X front, Y front and X rear (Y rear is not considered due to the missed measurement) as well as the N and P during the program of 'tool change' in machine-tool #1.

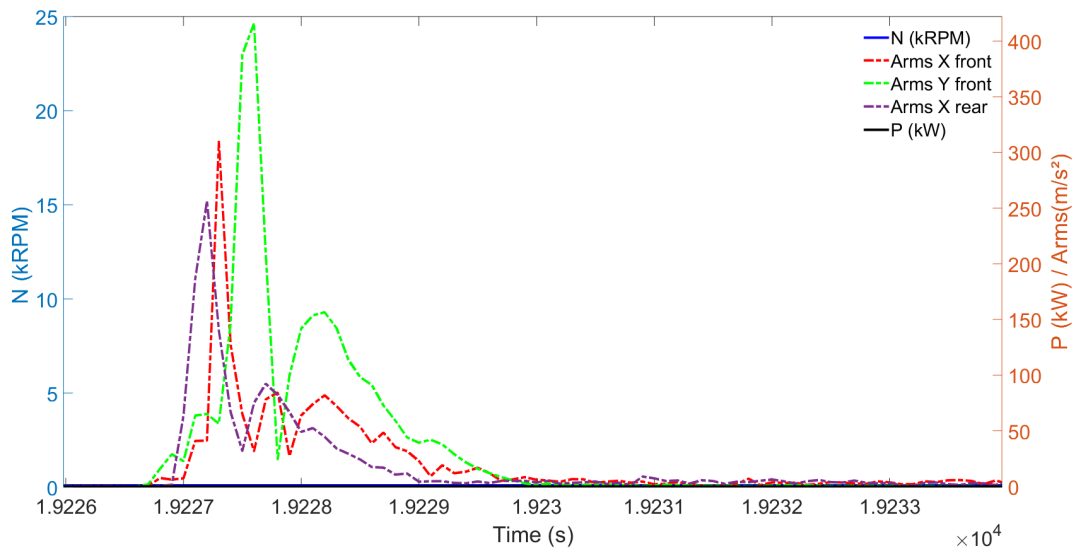


Figure 3.4 – Noise from the preparation program 'tool change' for the machine-tool #1

N (in blue line) and P (in black line) were nulls during the 'tool change'. While the *Arms* X front (in red dashed line), the *Arms* Y front (in green dashed line) and the *Arms* X rear (in purple dashed line) were very high (reaching 300 m/s^2) and appeared the dropout form due to the saturation of the piezoelectric material. These were normal vibrations due to the contact between the fixture of tools and the tool shop during the tool change. It was not a source of collision for machine-tool. This noise could be neutralized.

The noise in the principal programs for machining is mainly due to normal vibrations sources. For

example, the high value of accelerometers generated by lubrication startup (in the period *post spindle startup*) can be neutralized because it is not the source of chatter. Fig. 3.5 presents the high value of the accelerometer of Y front during the beginning of lubrication:

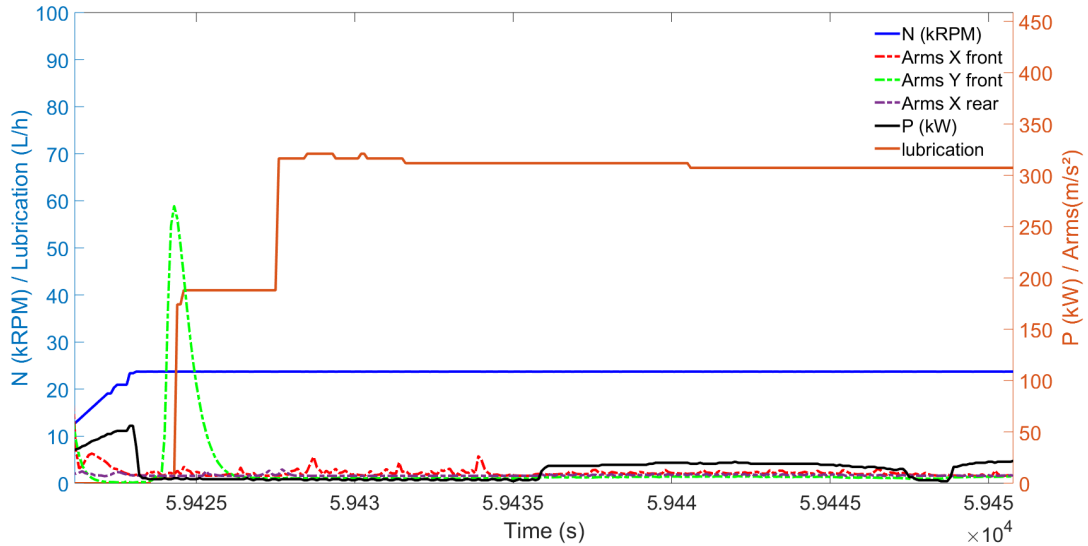


Figure 3.5 – Noise from the lubrication startup (in the period *post spindle startup*) in the principal machining program for the machine-tool #1

During this *post spindle startup*, the machine-tool begins to lubricate (in brown line), leading to high values for accelerometer Y in front bearing. However, this is not a chatter problem or a collision problem for the machine-tool. Therefore, this type of disturbing sources could be neutralized.

The objective of the data cleaning is to identify the wrong measurements and the missing values as well as neutralize the high values, which are not an indication of machining problems (chatter, tool breakage, excessive vibration, or choc). However, the over-cleaning of data should not happen to avoid the missing detection of incidents.

The proposed data cleaning enables better detection of machining incidents and better computation of KPIs.

3.2 Improvement of chatter identification

Chatter is an unstable cutting phenomenon that involves unacceptable quality defect (of waviness) on the workpiece surface [Godreau et al., 2019]. To detect the chatter, a new criterion Nh was developed in the thesis manuscript of Godreau [2017], based on Order Tracking. The criterion Nh is the sum of the amplitudes of the five dominant non-harmonic contributions of the vibration spectrum. To calculate Nh more precisely, the classification of the harmonic contributions in the frequency spectrum is necessary.

3.2.1 Classification of harmonic frequency

In the analysis of the frequency spectrum, there are three kinds of frequency: harmonic frequencies, non-harmonic frequencies (both due to the cut), and frequencies related to the machine vibration or spindle condition (such as f_{BPFO} and f_{BPFI}).

- harmonic frequencies (f_h in Eq. (3.2)) is defined as the integral multiples of the spindle rotation frequency f_b (Hz). The f_b is defined as in Eq. (3.1).
- non-harmonic frequencies (f_{nh}) is defined as the non-integral multiples of the spindle rotation frequency f_b , except the f_{BPFO} and f_{BPFI} . It is defined in Eq. (3.3)
- Frequencies BPFO (f_{BPFO}) and BPFI (f_{BPFI}) are defined as in Eq. (3.4) and in Eq. (3.5) [Randall and Antoni, 2011]. They are good criteria for the spindle condition [De Castelbajac et al., 2014].

$$f_b = \frac{N}{60} \quad (3.1)$$

$$f_h = n.f_b \quad (n \in \mathbb{N}) \quad (3.2)$$

$$f_{nh} = a.f_b, \quad (a \notin \mathbb{N}) \quad (3.3)$$

$$f_{BPFO} = f_b \cdot \frac{B}{2} \left(1 - \frac{\Phi}{D_m} \cos \alpha\right) \quad (3.4)$$

$$f_{BPFI} = f_b \cdot \frac{B}{2} \left(1 + \frac{\Phi}{D_m} \cos \alpha\right) \quad (3.5)$$

where N is the spindle rotation speed (in RPM), B is the number of balls in the bearing, Φ is ball diameter (in mm), D_m is the mean diameter of the bearing (in mm), and the α is the contact angle. It should be noted that, in Eq. (3.3), $f_{nh} \neq f_{BPFO}$ and $f_{nh} \neq f_{BPFI}$.

The equations above are the theoretical definitions. However, in industry, there is always variability for the frequency specific (harmonic, BPFO and BPFI) in the spectrum around the theoretical frequency. It was proposed to classify a frequency as harmonic, based on a given tolerance range around the theoretical value. The deviation from the closest theoretical harmonic Δf is defined as in Eq. (3.6).

$$\Delta f = |f - n.f_b| \quad (n \in \mathbb{N}) \quad (3.6)$$

While the f_{BPFO} and f_{BPFI} are specific for different machine-tools, the machine-tool #2 is taken as an example. Table 3.1 lists the theoretical value of BPFO and BPFI as well as their tolerances up and down for the machine-tool #2.

Table 3.1 – Frequency theoretical of $BPFO$ and $BPFI$ as well as their tolerances [Godreau, 2017]

Criterion	Frequency theoretical	Tolerance down	Tolerance up
f_{BPFI}	$13.7 f_b$	$13.25 f_b$	$13.84 f_b$
f_{BPFO}	$11.3 f_b$	$11.19 f_b$	$11.68 f_b$

The tolerance range was proposed as a constant value ($=12 Hz$) in the thesis manuscript of Godreau [2017], Then The business rule to distinguish the harmonic and non-harmonic is:

- if $\Delta f \leq 12Hz$, then f is assigned to the harmonic frequencies; else, f is assigned to non-harmonic frequencies.

Based on this business rule, the amplitudes according to the order of spindle rotation frequency (f/f_b) for accelerometer X in front bearing in machine-tool #2 during one day can be plotted in Fig. 3.6. Only the maximum rotation speed ($N=23720 RPM$) was plotted to test its performance.

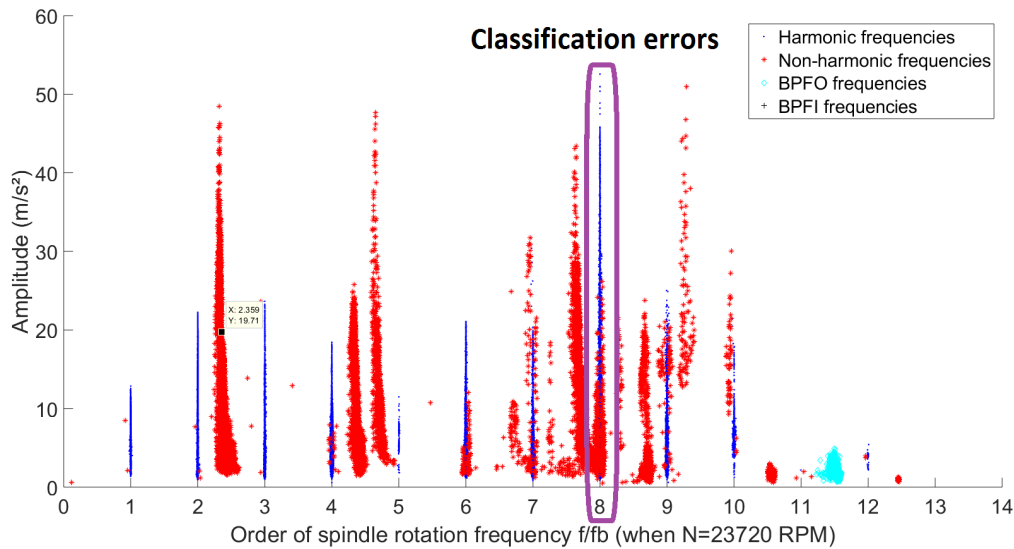


Figure 3.6 – Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for machine-tool #2

The harmonic frequencies (f_h) are plotted with blue points, non-harmonic frequencies (f_{nh}) are plotted with red stars, $BPFO$ frequencies (f_{BPFO}) are plotted with cyan diamonds, while $BPFI$ frequencies (f_{BPFI}) are plotted with black '+'. The tolerance range ($12Hz$) worked well when the orders of spindle rotation frequency were small (from 1 to 3). However, there are many classification errors from 4th order where many points of harmonic frequencies (Actual) were falsely classified into non-harmonic frequencies (Estimated) around the integer order, especially around the 8th order (purple rectangle). It means the tolerance range $12 Hz$ is too small for the higher frequencies.

The tolerance range increases with the spindle rotation frequency (f_b). By testing and manual mining on several machine-tools, a tolerance of 10% of the spindle speed was chosen.

The new business rules to distinguish the harmonic from non-harmonic is proposed:

- if $\Delta f \leq 10\%.f_b$, the f is assigned to the harmonic frequencies; else, the f is assigned to the non-harmonic frequencies.

Based on this business rule, the amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing in machine-tool #2 during the same day can be plotted in Fig. 3.7. The maximum rotation speed was always tested.

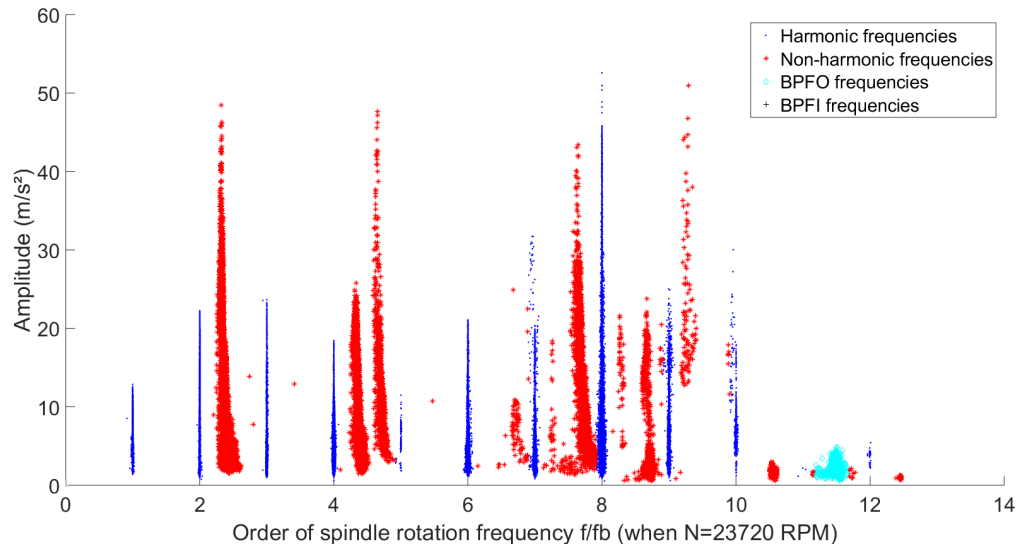


Figure 3.7 – Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for machine-tool #2

There were no classification errors in Fig. 3.7, the harmonic frequencies and the non-harmonic frequencies were well classified.

The business rule to classify the harmonic / non-harmonic frequency was tested in the machine-tool #1 and machine-tool #3, which are also for aluminum machining. Amplitudes via the orders of spindle rotation frequency (f/f_b) for accelerometer X in front bearing are plotted in Fig. 3.8, harmonic frequencies are plotted with blue points while non-harmonic frequencies are plotted with red stars. The harmonic frequencies and the non harmonic frequencies in frequency spectrum of the machine-tool #1 and #3 were well classified.

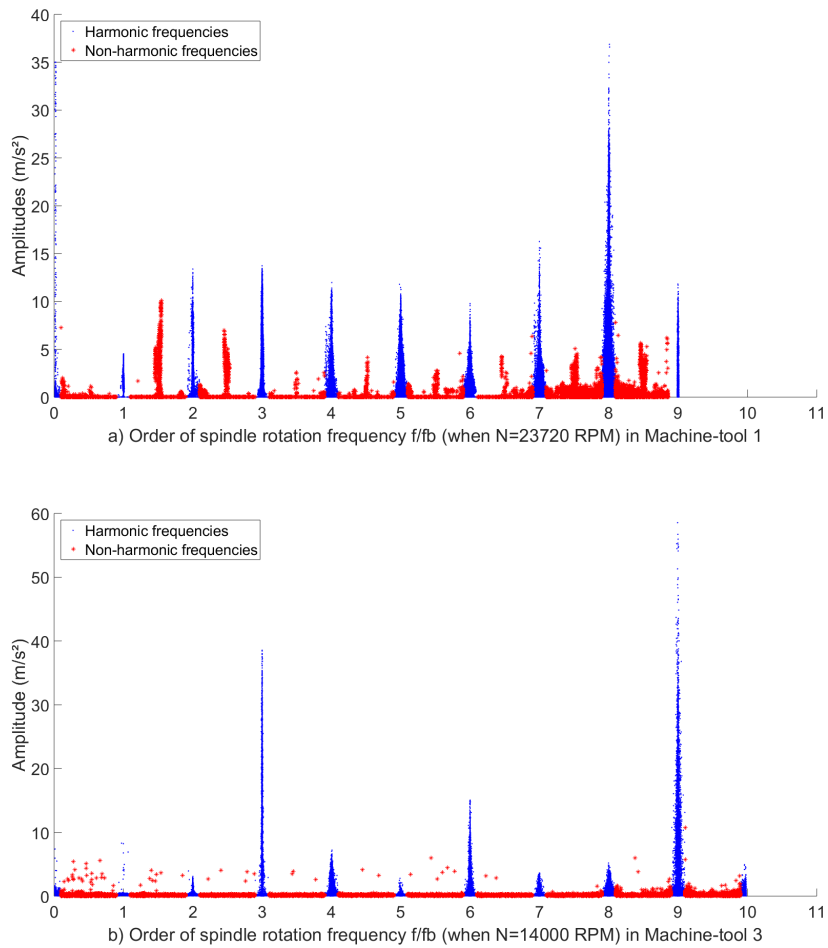


Figure 3.8 – Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for a) machine-tool #1 and for b) machine-tool #3

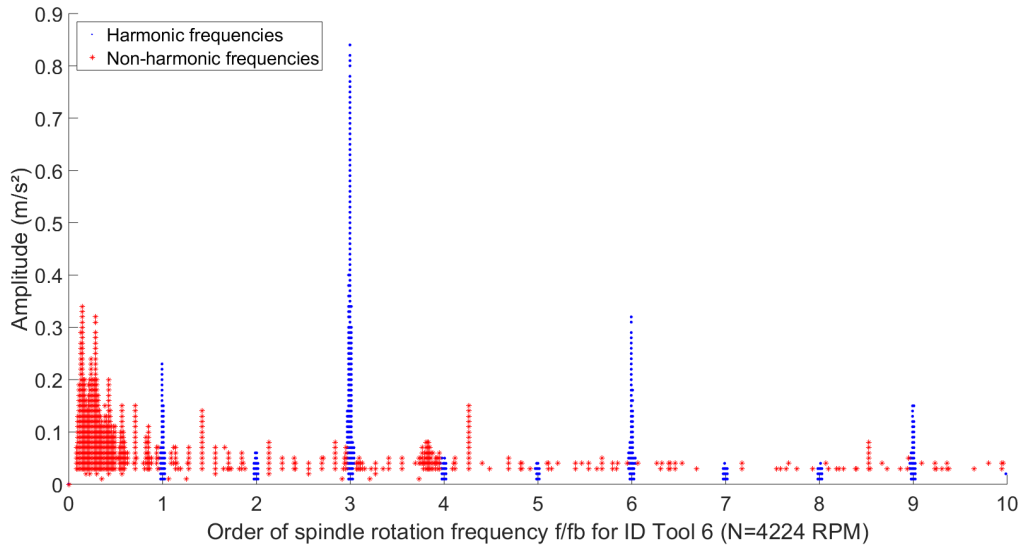


Figure 3.9 – Amplitudes according to the order of spindle rotation frequency for accelerometer X in front bearing for ID Tool 6 in machine-tool #4

However, the machine-tool #4 for titanium machining whose spindle speed N (reach 950 RPM) is much slower than aluminum machining ($N > 10000$ RPM), the threshold to classify harmonic vs. non harmonic is different:

- if $\Delta f \leq 3\%.f_b$, the f is assigned to the harmonic frequencies; else, the f is assigned to the non-harmonic frequencies.

The amplitudes via the order of spindle rotation f/f_b for the ID Tool 6 (cutting rotation speed $N=4224$ RPM) in Machine-tool #4 is also tested in Fig. 3.9.

The harmonic and non-harmonic frequencies are also well classified according to the business rule and the threshold for titanium machining. Therefore, the frequencies, as well as the corresponding amplitudes, can be well classified into harmonic and non-harmonic in the raw data at the beginning of data mining, as listed in section 3.1.

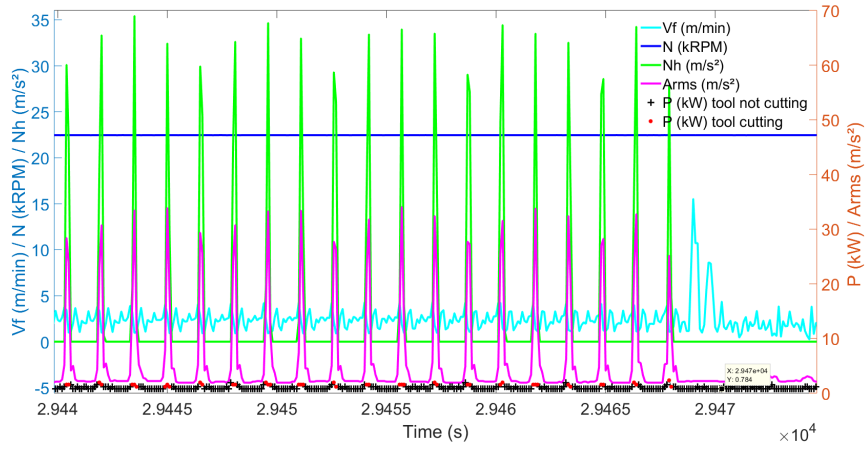
3.2.2 Chatter / non Chatter

Once the frequency was well classified, the criterion of chatter Nh can be computed at each instant (0.1s). Moreover, the chatter Nh should be evaluated when *tool is cutting* (cf. chapter 2), and the chatter occurs only when the criterion Nh exceed the threshold T_{Nh} . Therefore, the business rule to detect the chatter is :

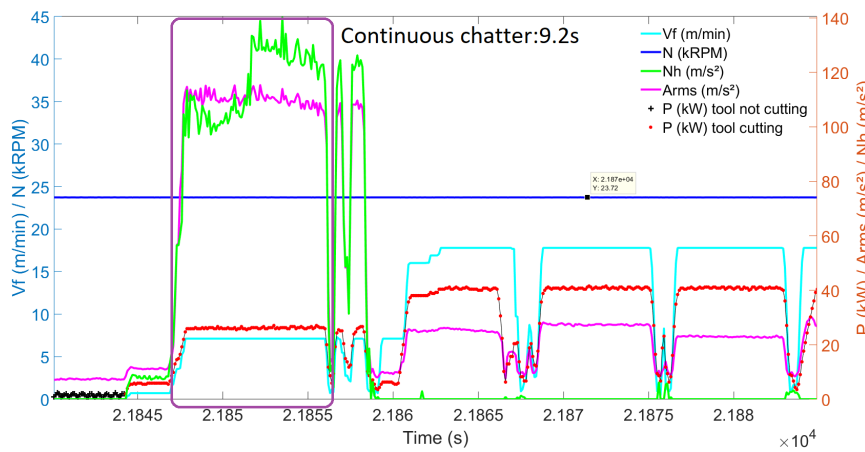
- If $(Nh > T_{Nh})$ and *tool is cutting*, then '*chatter detection*'=1.

Some chatter occurs very punctually (0.1s), while some chatter appears continuously (more than 0.2s). Fig. 3.10 presents these two types of chatter in machine-tool #2, where the spindle speed N is in blue line, feedrate Vf is in cyan line, the vibration signal $Arms$ is in pink line. '*tool not cutting*' is presented

by black '+' in the spindle power P curve, 'tool cutting' is represented by red points in P curve. In Fig. 3.10(a), the tool is cutting only during 0.2s (or 0.3s) each time, the chatter occurs just one instant (0.1s) during this short cutting period, which is called punctual chatter. In Fig. 3.10(b), the cutting is continuous, the chatter sustains several instants (9.2s in the purple rectangle) during the cutting period, which is called continuous chatter.



(a) Punctual chatter



(b) Continuous chatter

Figure 3.10 – Chatter in Machine-tool #2 for a) punctual and for b) continuous

To validate the chatter detection by the business rule, a case study will be held in the following subsection.

3.2.3 Test on a new database

Before detecting the chatter, the determination of the threshold T_{Nh} is necessary. The spindle #1 in the machine-tool #2 is taken as an example. Fig. 3.11 presents a histogram of Nh criterion, considering the whole spindle lifetime (215 days). The histogram is presented in log scale due to the rarity of abnormal events [Godreau, 2017]. The histogram of Nh can be modeled using GMM with three components (Y1 - normal population in green, Y3 - chatter population in black and Y2 - severe chatter in red). The criterion threshold T_{Nh} is then chosen as the abscissa value where the two Gaussians (Gaussian Y1 for the normal population and the Gaussian Y3 for chatter population) cross (to minimize classification errors). The threshold T_{Nh} of 17 m/s^2 is learned in Fig. 3.11.

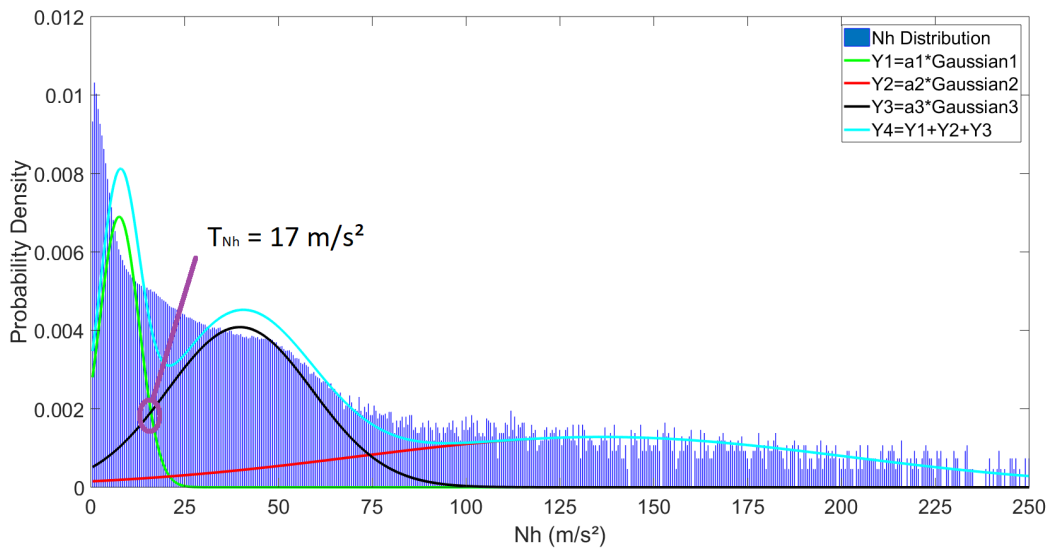


Figure 3.11 – Modeling of Nh by GMM (composed of a normal population in green and of a chatter population in black as well as of a severe chatter population in red) and learning of the detection threshold T_{Nh} for chatter

According to the business rule, once the threshold of T_{Nh} is determined, the chatter can be detected in real-time combining with the context information of 'tool cutting or not'. In chapter 2, 'tool is cutting' was classified into five different machining operations, and also into five different states of tool engagement. Therefore, chatter could be detected in different contexts when 'tool is cutting'. For example, in machine-tool #2, a lot of chatter was detected when 'tool entry materials' for the ID Tool 50. To verify the chatter during this period 'tool entry materials', the spindle speed N and the chatter criterion Nh as well as different states of tool engagement labeled in the spindle power P curve during the 15th day were plotted in Fig. 3.12.

In Fig. 3.12, there are many cutting periods during which *tool entry materials* (pink diamond) for 0.2s, and then *tool exit materials* (dark green diamond), it takes only 0.4s for these cutting periods. These cutting periods were during the ID prog 99 when ID Tool 50 cutting in the 15th day for Machine-tool #2. A lot of chatter during 'tool entry materials' were detected. To verify if the frequencies when this chatter appears were always non-harmonic, the order tracking of the frequency spectrum for this tool (ID Tool=50) and for ID prog 99 was plotted in Fig. 3.13.

The abscissa axis presents the order tracking of spindle rotation frequency ($=f/f_b$) where the spindle

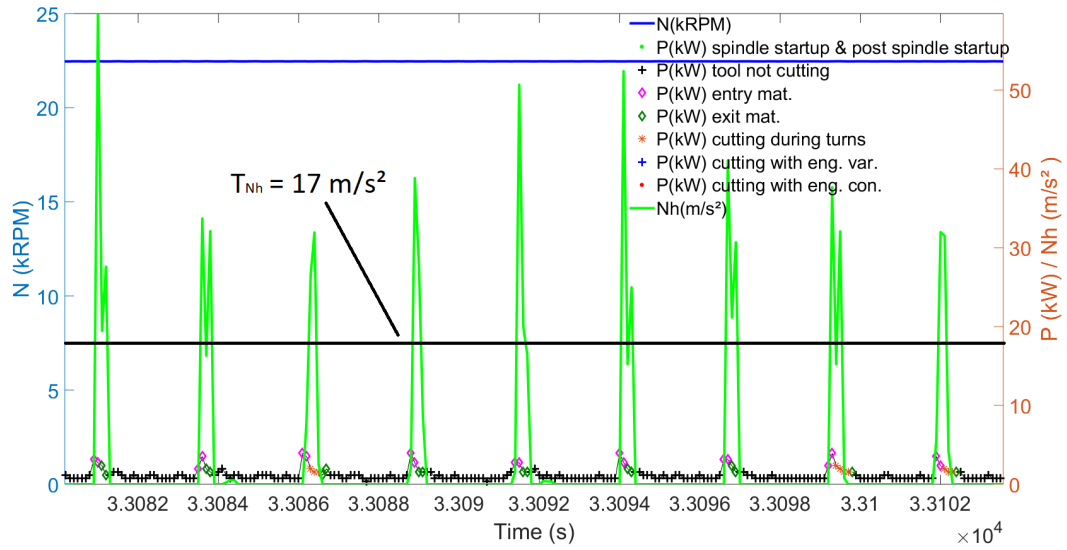


Figure 3.12 – Visualization of chatter criterion Nh (in green) with the classification of different states of tool engagement

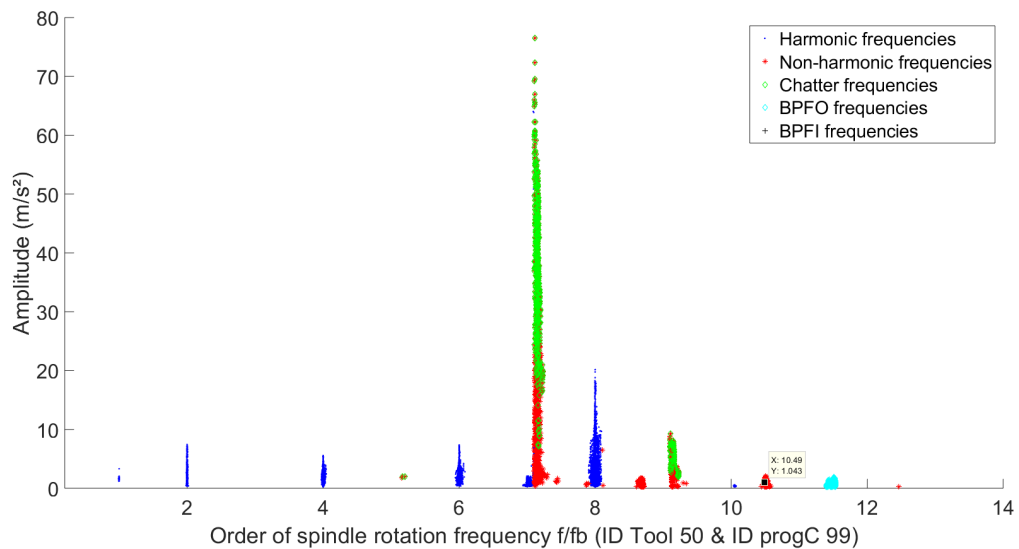


Figure 3.13 – Order tracking of the frequency spectrum for ID Tool 50 and for ID prog 99 in the 15th day for Machine-tool #2

rotation frequency $f_b = N/60 = 374.3$ Hz the vertical axis is the amplitudes corresponding to the frequency f . The harmonic (blue dots) and non-harmonic frequencies (red stars) were well classified according to the business rule above. The frequencies in the frequency spectrum collected by Emmatools when chatter occurred were plotted with green diamonds. They were all in the area of non-harmonic frequencies ($[5.168, 5.199].f_b$, $[7.102, 7.23].f_b$ and $[9.102, 9.23].f_b$). Especially at the frequencies $[7.102, 7.23].f_b$ where the amplitudes were very dangerous (up to 70 m/s^2), these were the main chatter frequencies. The modulation at \pm the tooth passing frequencies around the main

chatter frequencies confirms the occurrence of chatter ([Godreau, 2017]). Here teeth number for this tool is 2, thus the modulation is at +/-2 orders, at $[5.102, 5.23].f_b$ and $[9.102, 9.23].f_b$. The frequencies when this chatter appears during the *tool entry material* for ID Tool 50 were always non-harmonic is validated.

This chatter was particularly challenging to detect due to the proximity of the chatter frequency with its closest harmonic. The example illustrates the noticeable improvement of the chatter detection method, based on better harmonic classification and the threshold learning.

The contribution of this section is listed below:

- Propose a complete business rule combined with GMM for the chatter detection,
- Improve the classification on harmonic / non-harmonic frequencies, this will help calculate the chatter criterion Nh more accurately,
- Decipher the tool engagement context when chatter occurs.

3.3 Tool breakage

The tool breakage detection is an important issue because the machine-tool could continue machining with a broken tool, damaging the workpiece and even the machine-tool. To detect a tool breakage, two criteria of tool breakage are proposed: Ub for delayed detection of light tool breakage and Ub' for instantaneous detection of large tool breakage.

The criterion Ub estimates the unbalancement of the tool. It is defined as the vibration amplitude of the contribution at the spindle rotation frequency (f_b) [Godreau et al., 2019], when *spindle rotating at constant speed* (to prevent from false detection due to acceleration/deceleration of N) and *tool is not cutting* (to prevent from false detection due to cutting forces). It can detect the tool breakage sensitively, even for the light tool breakage. However, there is a delay in detection because the unbalancement of the tool is difficult to estimate when *tool is cutting*.

The criterion Ub' also estimates the unbalancement of shaft. It is defined as the vibration amplitude of the contribution at the spindle rotation frequency (f_b), when *spindle is rotating at constant speed*. When the large tool breakage occurred, this criterion can detect the exact instant of the breakage. However, this criterion is not able to detect light tool breakage due to the compensation between the tool breakage and the cutting forces when *tool is cutting*.

In conclusion, the criterion Ub can detect the light tool breakage well, but there is a delay, and the criterion Ub' can detect the large tool breakage instantaneously but can not detect well the light tool breakage. Therefore, it is necessary to combine these two criteria for tool breakage detection.

3.3.1 Business rule for detection of tool breakage

Based on the contextual classification (cf. Chapter 2) and the classification on harmonic / non-harmonic frequencies (cf. section 3.2), the criteria Ub and Ub' are computed at each instant (0.1s). Moreover, the criteria (Ub or Ub') are supposed to exceed the corresponding thresholds (T_{Ub} or T'_{Ub})

when the tool breakage occurs. Then, two business rules of tool breakage detection based on the monitoring of criteria (Ub or Ub') is listed as below:

- When *spindle rotating at constant speed*, If $Ub > T_{Ub}$ and *tool is not cutting*, then '*tool breakage detection*'=1. This business rule can detect all the tool breakage, but there is a delay for detection.
- When *spindle rotating at constant speed*, If $Ub' > T'_{Ub}$, then '*tool breakage detection*'=1. This business rule can detect the large tool breakage instantaneously but can not detect well the light tool breakage.

The objective is to calculate the thresholds (T_{Ub} and T'_{Ub}) by GMM. The spindle #1 in the machine-tool #2 is taken as an example. Fig. 3.14 presents a histogram of Ub criterion (when *tool is not cutting*) for tool breakage, considering a whole spindle lifetime (215 days). The distribution of Ub (in log scale) was modeled using GMM with two components (normal population in green and tool breakage population in red). The criterion threshold T_{Ub} is then chosen as the abscissa value where the two Gaussian cross (to minimize classification errors). For machine-tool #2, the threshold T_{Ub} of 6.32 m/s^2 is learned in Fig. 3.14.

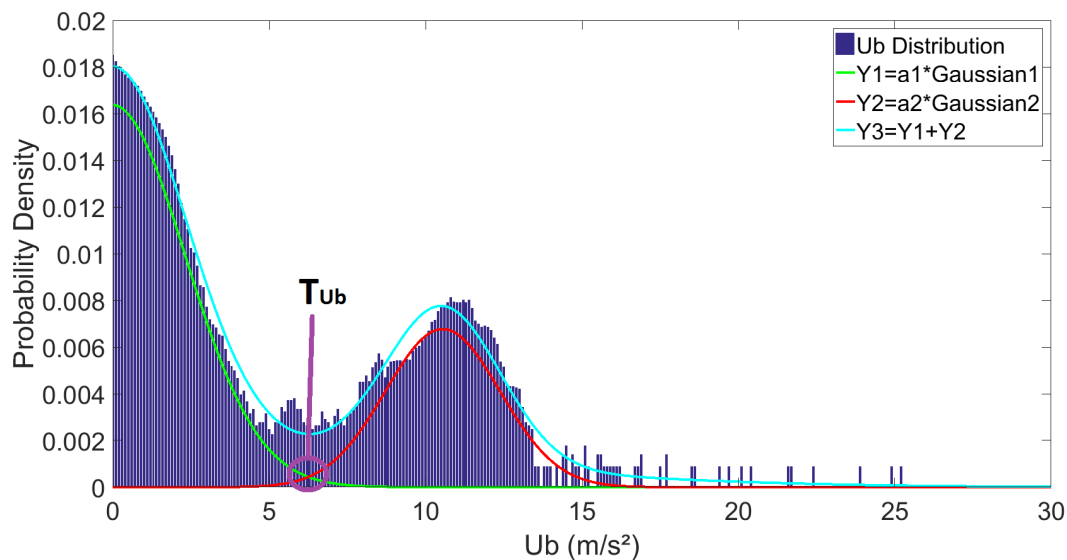


Figure 3.14 – Modeling of Ub distribution by GMM (composed of a normal population in green and of a tool breakage population in red) for Machine-tool #2 and learning of the detection threshold T_{Ub} for tool breakage

The same method, the threshold T'_{Ub} for the criterion Ub' is obtained as 8.1 m/s^2 .

Once the thresholds (T_{Ub} and T'_{Ub}) are determined, the tool breakage can be detected in real-time based on the business rule. The following subsection will illustrate the application to find the tool breakage during the whole lifetime of the spindle for the machine-tool #2 (215 days).

The contribution of this subsection is that the thresholds for the tool breakage criteria T_{Ub} and T'_{Ub} could be computed automatically by GMM, and two business rules for tool breakage detection were set up.

3.3.2 Application

By basic data mining, it can be found that large vibrations (*Arms* in pink) occurred during the 136th day of the production, at the 6th hour in Fig. 3.15. However, without contextual information, machine learning, and knowledge management, it is not possible to determine if it is normal or dangerous, nor to understand what happened that day. That is why the business rules for tool breakage detection are important, as well as deciphering the operational context, thanks to the contextual classification.

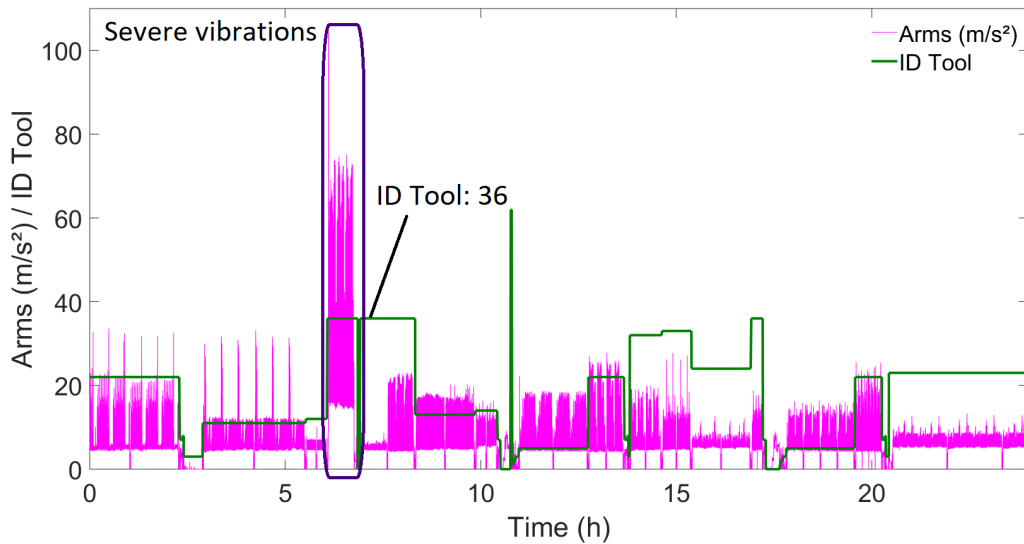


Figure 3.15 – Raw data of vibration level *Arms* during the 136th day of spindle #1 production for machine-tool #2

Fig. 3.16 presents the Ub criterion focusing on ID Tool 36 during the 136th day. The tool instance number of ID Tool 36 is also plotted in green.

The purple curve plots Ub for tool breakage monitoring from the accelerometer signals for ID Tool 36. The black line indicates the learned threshold T_{Ub} (from Fig. 3.14). The business rules are processed, which compares the monitoring criterion Ub to the detection threshold T_{Ub} (that was obtained by unsupervised machine learning GMM, cf. Fig. 3.14) in the context of *tool not cutting* state (to avoid false alarms due to the cutting forces). In this way, the breakage of ID Tool 36 is detected from 31st to the 70th minute. The green line presents the tool instance for the ID Tool 36. It can be observed that the machine-tool changed the cutting tool (from 5th tool instance to 6th tool instance) at the 75th minute. The red line presents the Boolean for tool breakage detection (1 - tool breakage, 0 - good tool condition). The detection of tool breakage for ID Tool 36 in the 136th day is validated.

Without any detection method, this tool breakage was not identified, and the machine-tool has continued machining for 39 min with a broken tool, which is very dangerous.

In this case of tool breakage, the unbalancement of tool Ub is severe (up to 12 m/s^2 compared to the threshold 6.32 m/s^2), it is a large tool breakage. Another criterion Ub' is applied to detect the exact instant of the tool breakage and decipher more precisely the operational context when *tool is cutting*. The threshold for this criterion Ub' has been obtained as 8.1 m/s^2 by GMM method. Therefore, an

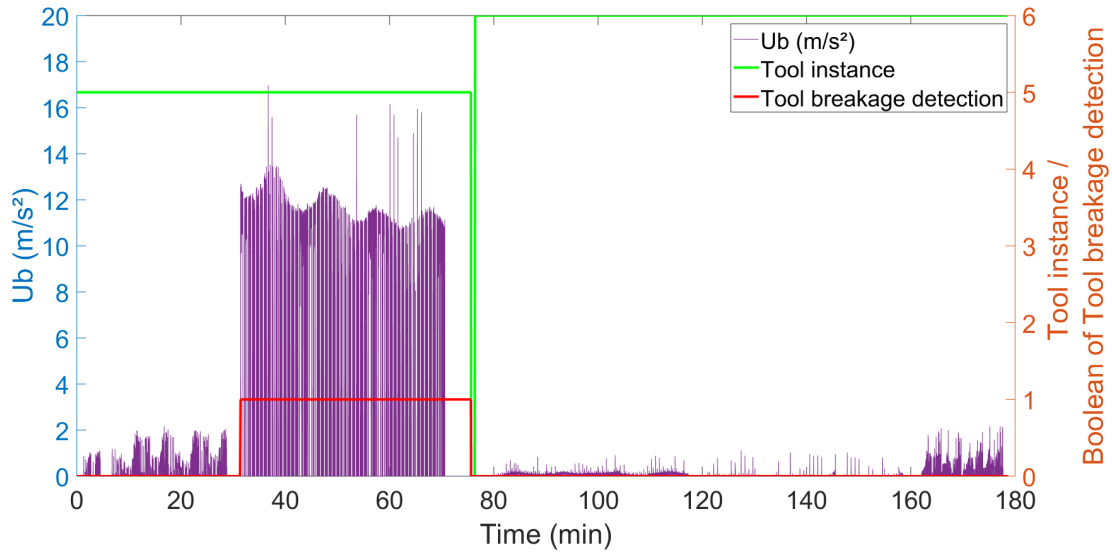


Figure 3.16 – Visualization of U_b criterion (in purple) for ID Tool 36, the index of tool instance (in green) and the Boolean for tool breakage detection (in red)

adaptive threshold for tool breakage can be set up as: T_{Ub} (6.32 m/s^2) when *tool is not cutting* and T'_{Ub} (8.1 m/s^2) when *tool is cutting*. Thanks to the contextual classification, the operational context and the unbalancement of shaft (U_b') as well as the adaptive threshold for tool breakage were plotted in Fig. 3.17:

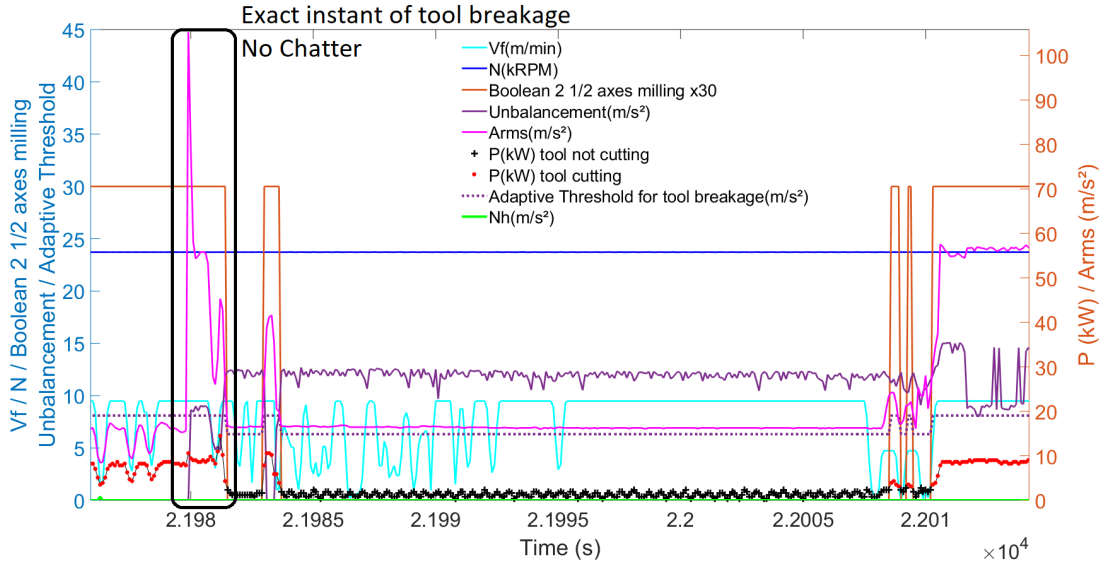


Figure 3.17 – Operational context when tool breakage occurs for ID Tool 36 in ID prog 154 in the 136th day in the machine-tool #2

To highlight the $2\frac{1}{2}$ axes milling operations (brown line), the scale has been multiplied by 30 in Fig. 3.17. In the black rectangle, the vibration $Arms$ (rose line) increased to a dangerous value (100 m/s^2), but there is no chatter (Nh around zero) at this moment. Excessive forced-vibrations have

damaged the tool. At the same moment, the unbalancement of shaft (Ub' when *tool is cutting*) which is in purple line, firstly exceed the adaptive threshold for tool breakage ($T'_{Ub}=8.1 m/s^2$ when *tool is cutting*) which is in purple dashed line. It was confirmed that this moment is the exact instant of the tool breakage. It was found that the tool was in $2\frac{1}{2}$ axes milling, and the tool was cutting with constant tool engagement when the tool breakage occurred due to excessive forced-vibrations.

Based on these business rules for tool breakage detection, other tool breakages can be detected. Two lifetimes of spindles (spindle #1 - 215 days and spindle #2 - 211 days) for machine-tool #2 are analyzed. Compared to the Old tool breakage detection ([Godreau et al., 2019]), new tool breakages were detected and validated by manual mining. It demonstrates that the proposed method has improved accuracy. Fig. 3.18 listed and compared the tool breakage detected by the Old method ([Godreau et al., 2019]) and by the new business rules detailed in this thesis.

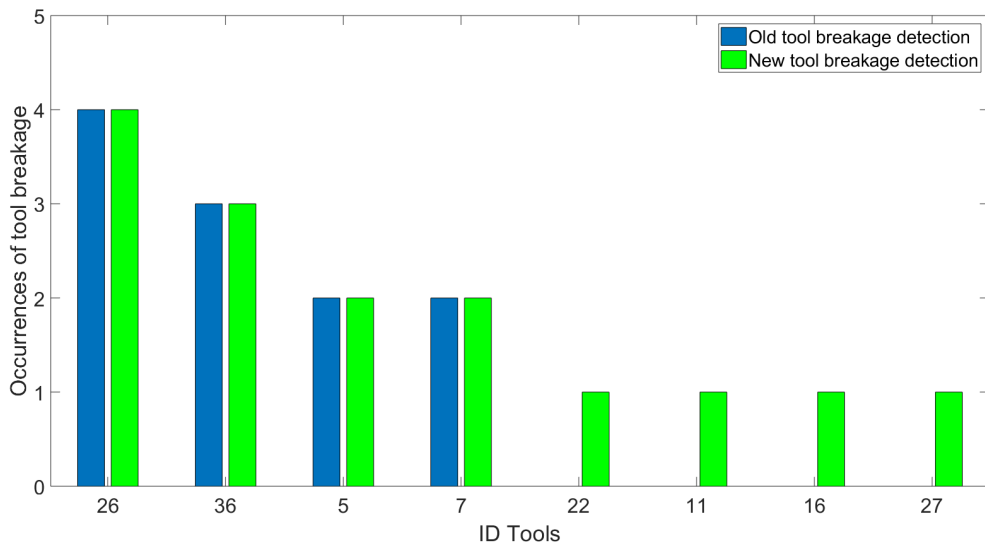


Figure 3.18 – Occurrences of tool breakage via ID Tool during 2 spindle lifetimes for machine-tool #2 based on the Old and New tool breakage detections

The abscissa axis presents the ID Tool for the tools with tool breakage detections. And the vertical axis was the number of occurrences of tool breakage detected for each ID Tool. Compared to the Old tool breakage detection (in blue), 4 new tool breakages (in green) were detected thanks to the new business rules:

- ID Tool 22 in the 18th day during the spindle #1 manufacturing.
- ID Tool 11 in the 54th day during the spindle #2 manufacturing.
- ID Tool 16 in the 54th day during the spindle #2 manufacturing.
- ID Tool 27 in the 152th day during the spindle #2 manufacturing.

Among these new tool breakage detection, ID Tool 22 in the 18th day during the spindle #1 manufacturing is taken as an example. To decipher the operational context at the exact instant of the

tool breakage, the Boolean for $2\frac{1}{2}$ axes milling (x30) and the unbalancement of shaft as well as the adaptive threshold for tool breakage were plotted in Fig. 3.19.

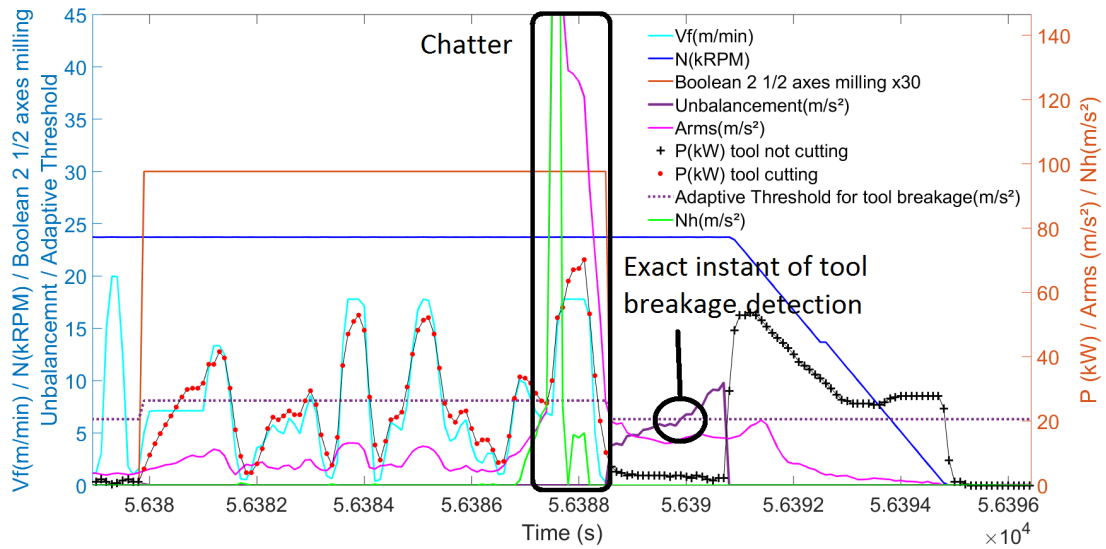


Figure 3.19 – Operational context when tool breakage occurs for ID Tool 22 in ID prog 108 in the 18th day during the spindle #1 manufacturing in the machine-tool #2

After the severe chatter (Nh in green reached a dangerous value - 140 m/s^2) during *tool is cutting* (plotted by red dots), the unbalancement of shaft ($=Ub$ when *tool is not cutting*) exceeded the corresponding threshold (T_{Ub}) which is in purple dashed line. It indicated the tool breakage detection. Just after this tool breakage, the spindle speed N decelerated, and the machining program is stopped. The machine-tool has correctly detected the breakage and changed this Tool instance for this ID Tool 22 just after this tool breakage. The tool breakage of this ID Tool 22 during the $2\frac{1}{2}$ axes milling in this program (ID prog 108) in the 18th day during the spindle #1 was confirmed and explained by a severe chatter occurrence.

This tool breakage is difficult to detect, because it is a light tool breakage (Ub is close to the threshold $T_{Ub}=6.32\text{ m/s}^2$). That is why it was not detected during the cut by the criterion Ub' , which is less sensitive.

Among these detected tool breakages, the tool breakage for ID Tool 26 in the 99th day in spindle #1 and the tool breakage for ID Tool 22 in the 18th day in the spindle #1 as well as the tool breakage for ID Tool 16 in the 54th day in the spindle #2 would be missing if the criterion Ub' with its threshold T'_{Ub} was applied to detect the tool breakage. These were light tool breakages. In other words, the criterion Ub combined with its threshold T_{Ub} (when *tool is not cutting*) can detect all of the tool breakages, but there is a delay in detection.

Beyond the detection, the traceability of the instances of tool breakage detection and their operational context that has been deciphered by the proposed classification can be reused for diagnostic. They can be analyzed, by the inference engine based on another set of business rules, in order to identify the main faulty cutting tools and contexts, to determine the cause of the occurrence and to solve it. [Ladj, Wang, Meski, Belkadi, Ritou, and Da Cunha, 2020].

3.4 Excessive vibrations

Excessive vibrations are defined as severe forced vibrations when the *tool is cutting* and there is no chatter nor tool breakage appear in this period. The criterion based on raw data like $Arms$ (m/s^2), $Vrms$ (mm/s) can be useful to detect excessive vibrations. The business rule to detect the excessive vibrations is defined as below:

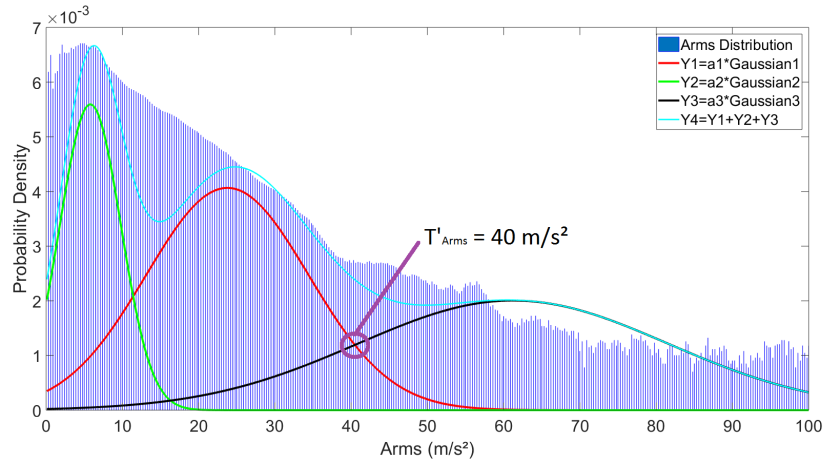
- If ($Arms > T'_{Arms}$ or $Vrms > T_{Vrms}$) and '*tool is cutting*' and '*chatter detection*'=0 and '*tool breakage detection*'=0, then '*excessive detection*'=1'.

where T'_{Arms} (in m/s^2) and T_{Vrms} (in mm/s) are the excessive vibration threshold of the $Arms$, $Vrms$. It should be noted that the T'_{Arms} is much higher than T_{Arms} which is the threshold of $Arms$ to detect '*tool is cutting or not*' (cf. chapter 2).

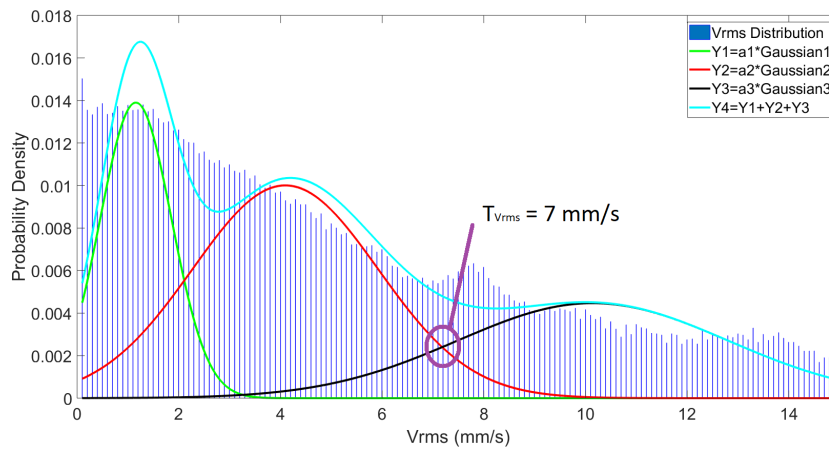
Before computation of these two thresholds T'_{Arms} and T_{Vrms} , it is necessary to recognize and deal with the data noise (coming from data collection cf. section 3.1).

The method is the same as in sections 3.2 and 3.3, the objective is to find the threshold T'_{Arms} and T_{Vrms} by GMM to detect excessive vibrations. The spindle #1 (215 days) in the machine-tool #2 is taken as an example. Fig. 3.20 (a) presents the histogram of $Arms$ (in log scale) criterion for excessive vibrations, considering a whole spindle lifetime (215 days). The distribution of $Arms$ was modeled using GMM with three components (Y2 - normal population when the *tool is not cutting* in green and Y1 - normal population when the *tool is cutting* in red as well as Y3 - the excessive vibration population in black). The criterion threshold T'_{Arms} is then chosen as the abscise value where the two Gaussians (Gaussian Y1 for normal population when the *tool is cutting* and the Gaussian Y3 for the excessive vibration) cross (to minimize classification errors). The threshold T'_{Arms} of $40 m/s^2$ is learned in Fig. 3.20 (a). Applying the same method, the distribution of $Vrms$ was modeled using GMM with three components (Y1 - normal population when *tool is not cutting* in green and Y2 - normal population when *tool is cutting* in red as well as Y3 - the excessive vibration population in black). The criterion threshold T_{Vrms} is then chosen as the abscise value where the two Gaussians (Y2-normal population when the *tool is cutting* and Y3-excessive vibration), $T_{Vrms} = 7$ (mm/s) is learned in Fig. 3.20 (b).

Compared to the experiments, the modelings are not as good as the previous GMM for the detections of context or incident. However, with the proposed interpretation of the meaning of each Gaussian, the classifications between normal or excessive vibration are correct (as it has been verified by manual mining). Moreover, the thresholds found by GMM are consistent with the thresholds recommended by the [ISO 10816, 2009] (see group II in Fig. 3.21). Besides, using a unique learning method for different contexts or incidents is suitable. Therefore, the modeling by 3 GMM has been validated.



(a) Modeling of Arms distribution by GMM with 3 Gaussians



(b) Modeling of Vrms distribution by GMM with 3 Gaussians

Figure 3.20 – Modeling of a) $Arms$ and b) $Vrms$ distributions by GMM (with 3 populations: normal population when *tool is not cutting*, normal population when *tool is cutting* as well as the excessive vibration) for Machine-tool #2 and learning of the detection threshold T'_{Arms} and T_{Vrms} for the detection of excessive vibration

To verify if the business rule combining these thresholds (T'_{Arms} and T_{Vrms}) works well or not, one case-study will be done on the machine-tool #2. Many excessive vibrations were found, from which one excessive vibration is illustrated. Fig. 3.22 presents this excessive vibration when ID Tool 36 works during the ID prog 249 in the 188th day. The spindle rotation speed N (blue line) reaches 23720 *RPM* when *tool is cutting*, *tool cutting* (red dots) and *tool not cutting* (black '+') are presented in the curve of spindle power P .

In Fig. 3.22, it is clear that the criterion of chatter Nh (green line) is smaller than the threshold T_{Nh} ($T_{Nh}=17 \text{ m/s}^2$ according to Fig. 3.11). And the criterion of tool breakage Ub (purple line) is around 0 (smaller than the threshold $T_{Ub}=6.32 \text{ m/s}^2$ according to Fig. 3.14). So there is no chatter in this area, neither the tool breakage. However, there are two periods where the $Arms$ (pink line) exceed the threshold T'_{Arms} (pink dashed line), and where the $Vrms$ (dark green line) also exceed the

Machine		Class I small machines	Class II medium machines	Class III large rigid foundation	Class IV large soft foundation
Vibration Velocity Vrms	0.01	0.28			
	0.02	0.45			
	0.03	0.71		good	
	0.04	1.12			
	0.07	1.80			
	0.11	2.80		satisfactory	
	0.18	4.50			
	0.28	7.10		unsatisfactory	
	0.44	11.2			
	0.70	18.0			
	0.71	28.0		unacceptable	
	1.10	45.0			

Figure 3.21 – thresholds recommended by the ISO 10816 standard [ISO 10816, 2009]

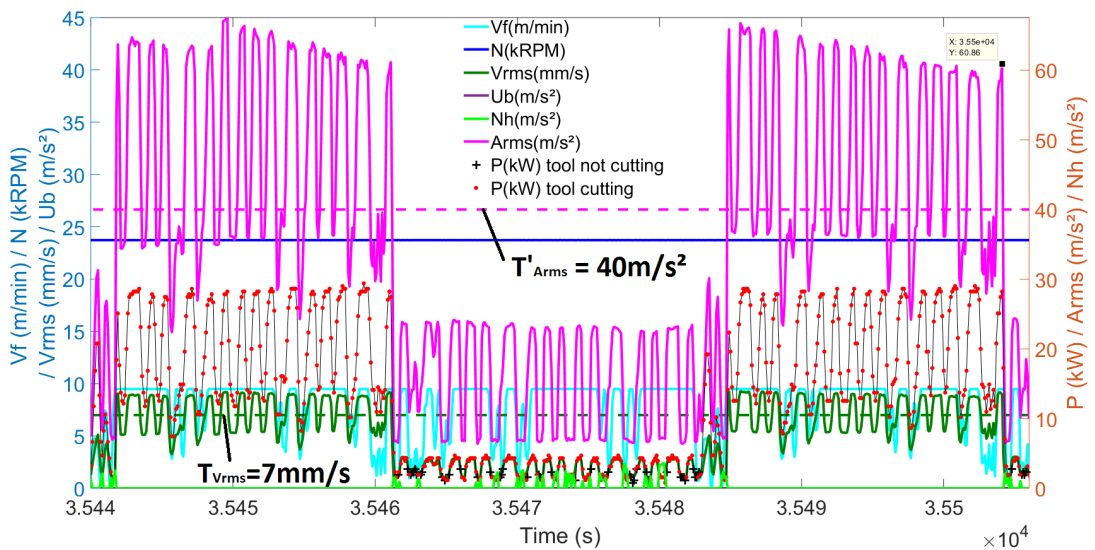


Figure 3.22 – Excessive vibration for the ID Tool 36 during the ID prog 249 in the 188th day for machine-tool #2

threshold T_{Vrms} (dark green dashed line). These periods are while *tool is cutting* and indeed, periods of excessive vibrations.

The contribution of this subsection is that a complete business rule combined with two thresholds corresponding to $Arms$ and $Vrms$ for excessive vibration detection is proposed. It is validated in the use-case of machining in the 188th day for ID Tool 36 during the ID prog 249 on machine-tool #2.

3.5 Conclusion

This chapter presented different business rules for the detection of three kinds of incidents, based on the criteria and their corresponding thresholds: chatter, tool breakage and excessive vibrations.

Before data mining, data understanding is necessary. It is important to identify the abnormal data and noise. The data issued from measurement problem is not considered while the noise coming from the disturbing source will be neutralized after deciphering its context.

The contributions of this chapter are listed below:

- a business rule for chatter detection has been proposed. The classification on harmonic / non-harmonic frequencies was improved to calculate the criterion Nh more accurately. This business rule is validated in the use-case of chatter during the spindle #1 manufacturing on the machine-tool #2.
- the proposal of two business rules combined with the criterion (Ub when *tool is not cutting* or Ub') as well as the adaptive threshold (T_{Ub} or T'_{Ub}) for tool breakage detection. Ub can detect all of the tool breakages, including the light tool breakage, but there is a delay. Ub' can detect the large tool breakage instantaneously, but it can not detect the light tool breakage well. Four new tool breakages were detected for the two spindle lifetimes on machine-tool #2 thanks to the improvement of the business rule for tool breakage detection, compared to the old tool breakage detections by [Godreau et al. \[2019\]](#).
- a business rule combined with the thresholds (T'_{Arms} and T_{Vrms}) for excessive vibration is proposed. This business rule has been validated in the spindle #1 database on the machine-tool #2.
- the learning of thresholds for the detection of chatter, tool breakage, and excessive vibration by GMM automatically;
- the deciphering of the operational context when incidents occur, based on the contextual classification that explains the types of machining and tool engagement.

Based on the business rules proposed in this chapter, the incidents can be detected in real-time during manufacturing. It can help the production line to stop manufacturing once the incidents are detected and avoid further losses. These contextual information and incidents detections will be very useful to calculate KPIs in the following chapter.

Proposition and management of KPIs

Contents

4.1 Knowledge Discovery in Database (KDD) procedure	74
4.1.1 Management and aggregation	75
4.1.2 Proposition of new aggregation operators	76
4.2 KPIs for production organization	77
4.2.1 Advanced Overall Equipment Effectiveness (OEE)	77
4.2.2 Workpieces production	79
4.3 Traceability of Workpiece quality	80
4.3.1 Respect of cutting specifications	80
4.3.2 Detection of incidents	84
4.4 KPIs for production monitoring	86
4.4.1 Traceability of duration of machining operations	86
4.4.2 Traceability of cutting aspects	87
4.4.3 Human factor during production	88
4.4.4 Incidents detected during production	89
4.5 KPIs for tool performance	89
4.5.1 Distribution of chatter per tool	89
4.5.2 Distribution of excessive vibration per tool	91
4.5.3 Tool use duration before tool breakage	92
4.6 KPIs for program performance	92
4.6.1 Distribution of chatter per program	92
4.6.2 Distribution of excessive vibration per program	95
4.6.3 Productivity of CAM programs	95
4.6.4 Slow idle motions	96
4.7 KPIs for monitoring of machine-tool condition	97
4.7.1 Spindle condition monitoring	97
4.7.2 Incidents detection	98
4.8 The impact of the classification errors on the evaluation of KPIs .	98
4.9 Conclusion	100

Summary

This chapter presents the proposition and the management of decision-aid indicators (also called Key Performance Indicators - KPIs). It starts from the Knowledge Discovery in Database (KDD) procedure. The data is successively aggregated following a 4-levels approach, from real-time signal, monitoring criteria, Smart Data and KPIs. The collected data by Emmatools were processed to perform the contextual classification (cf. chapter2) and to find the incidents (cf. chapter3). These contextual classification combined with the incidents detected are aggregated into Smart Data. Finally, the KPIs can be evaluated from these meaningful and simplified Smart Data. The new KPIs are proposed from section 4.2 to section 4.7. They are classified and presented based on the scope of the decision: production organization; workpiece quality traceability, production monitoring; performances of tools; performance of programs; and monitoring of machine-tool condition. The last section investigates the impact of potential classification errors on the evaluation of KPIs.

The objective of this thesis is to propose a decision-aid system for intelligent and connected machine-tools. Emmatools, which combine the CNC information and the sensors integrated into the machine-tool (power and accelerometers), collect the raw data in real time. However, one significant challenge is how to exploit these detected incidents and associated contexts for decision aid and operational management perspective. The decision-aid framework has been proposed as a part of the ANR SmartEmma project (Fig. 4.1). The relevant and appropriate KPIs need to be proposed and formalized to help decision aids for operational management. As a part of this thesis, new KPIs have been proposed.

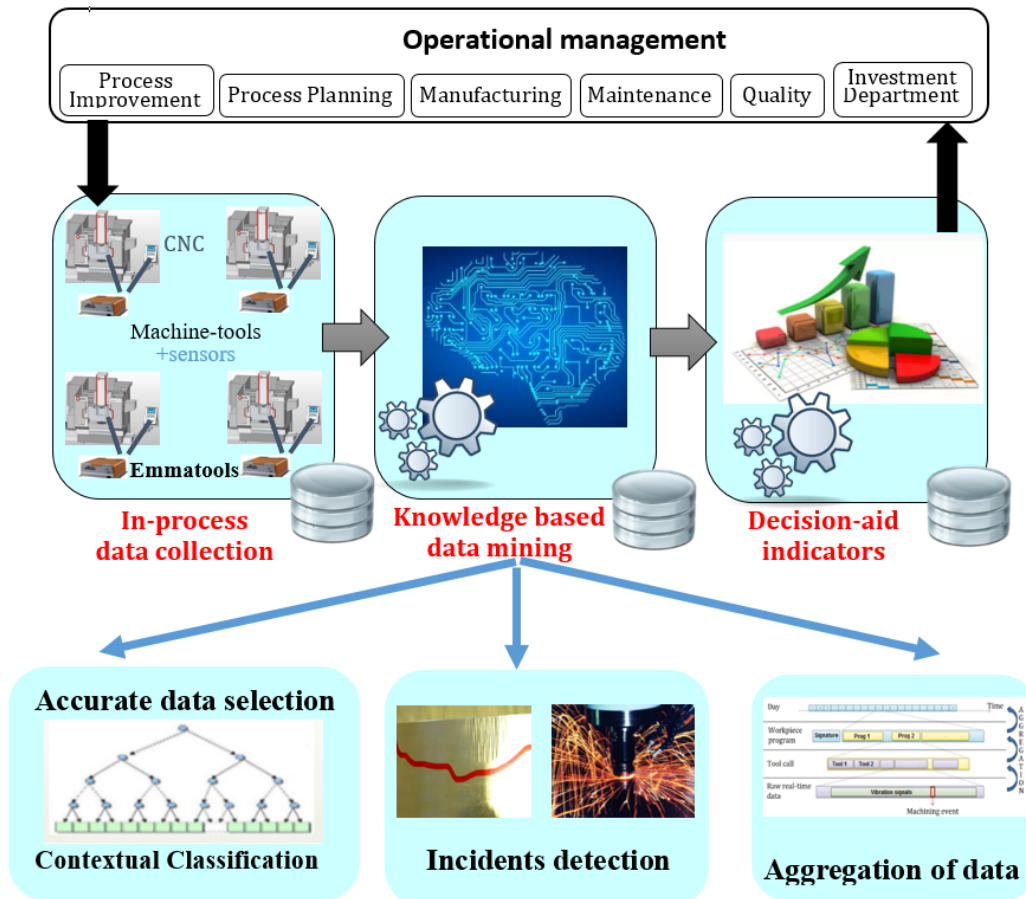


Figure 4.1 – Decision-aid framework for intelligent and connected machines-tools

As shown in Fig. 4.1, three modules are interconnected to support the decision-making process. In the first module, Emmatools perform real-time signal processing to compute the monitoring criteria and collect the in-process heterogeneous data. The following step is a knowledge-based data mining. It uses unsupervised machine learning algorithms to do the contextual classification (cf. chapter 2), and incidents detection (cf. chapter 3), as well as the data aggregation, conducting to Smart Data. Finally, new KPIs are computed and delivered to various departments according to their needs in process management and improvement. A knowledge-engineering process is used to support the extraction, the formalization, and the integration of other manufacturing knowledge.

To be efficient, the decision-makers should obtain the right information at the right moment. To do so,

the decision-aid framework should be able to manage large quantities of data and ensure the flexible circulation of this data throughout the enterprise network. To solve this issue, experts' knowledge can be exploited for data aggregation. Only small quantities of meaningful machining-process information are then used to construct decision-aid indicators according to the experts' intention and the business context. In this way, a knowledge-based and multi-level approach has been proposed (Ritou et al. [2019]).

4.1 Knowledge Discovery in Database (KDD) procedure

Figure 4.2 presents the Knowledge Discovery in Database (KDD) procedure. The data is analyzed through four levels from raw data to the decision-aid indicators (KPIs) [Ritou et al., 2019]:

- level 0: real-time signals measured by Emmatools, at 25kHz for accelerometers or 10Hz otherwise. Then real-time signal processing computes the monitoring criteria collected with the operational context every tenth of a second.
- level 1: through contextual classification and monitoring, the contextual information (ex. *tool cutting or not, spindle rotates at constant or varying speed or is stopped*) as well as the incidents detected are recorded in real-time, and are added into the raw data to obtain a more complete database. It takes around 400 Mb per day and per machine-tool.
- level 2: Smart Data is obtained by aggregating per machining operation from the database complete in level 1. It takes around 500 Kb per machine per day, which compresses further the raw data (400 Mb). More importantly, it contains all the meaningful machining-process information for the calculation of decision-aid indications.
- level 3: finally, the decision-aid indicators are computed based on the Smart Data.

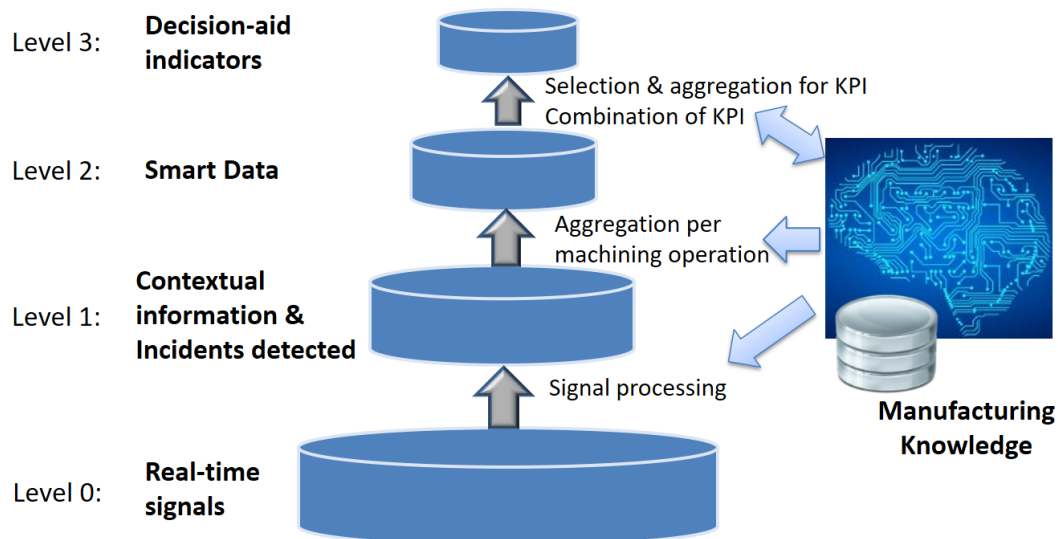


Figure 4.2 – Knowledge-based multi-level data aggregation [Ritou et al., 2019]

4.1.1 Management and aggregation

The contextual classification and the detection of incidents have been presented in chapters 2 and 3; hence this contextual information and detected incidents will be aggregated per machining operation. In the aeronautic industry, a machining operation with a given tool can last for several hours, with a large volume of corresponding data. We chose to aggregate the monitoring data into a few Smart Data, at the time scale corresponding to one usage period of a given cutting tool. Figure 4.3 presents the diagram for the organization of the data and the different potential stages of aggregation: data can be aggregated e.g. by ID Tool or program.

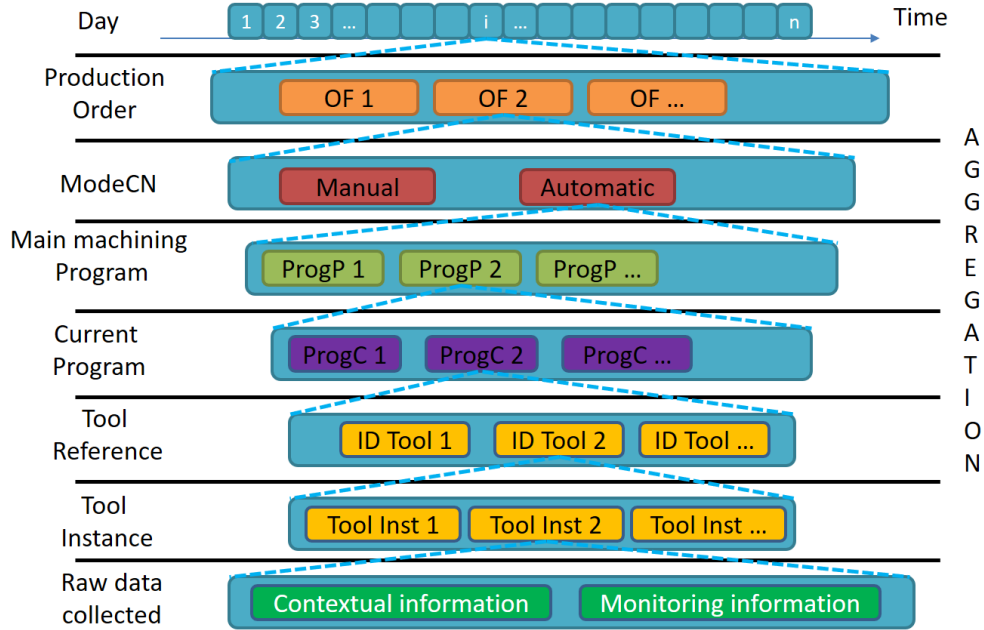


Figure 4.3 – Data organization and different potential stages of aggregation

Operators are necessary for aggregation. It can be statistical ones, such as means, medians, standard deviations, or maximum values. However, these basic operators are not relevant for machining incidents or for the machining vibrations. In fact, machining incidents are very rare (e.g. one severe event for a few seconds per week) and need to be clearly emphasized. That is why a Criticality Operator (CO) has been proposed by [De Castelbajac et al. \[2014\]](#). It quantifies the excessive magnitudes of a monitoring criterion, e.g. chatter (Nh), over a critical threshold T_i . Let $X_i = \{x_i(k), k = 0 \dots n\}$ the time series representing an i^{th} monitoring criterion. The operator is defined in Eq. (4.1):

$$CO[X_i > T_i] = \sum_{k=t_i}^{t_f} y(k) \cdot \Delta t \quad \text{where} \quad \begin{cases} y(k) = x_i(k) - T_i & \text{if } x_i(k) > T_i \\ y(k) = 0 & \text{if } x_i(k) \leq T_i \end{cases} \quad (4.1)$$

with t_i and t_f the initial and final time of the considered period (based on operational context data) and $\Delta t = 0.1s$ the sampling period. A second proposed operator *Time* evaluates the time during which a criterion exceeds a critical threshold. It was proposed by [Godreau et al. \[2019\]](#), see Eq. (4.2). Note that, in the case of vibrations, the monitoring values obtained from each of the four accelerometers are aggregated by a quadratic mean, before the aforementioned aggregation.

The critical thresholds T_i are obtained by unsupervised Machine Learning GMM from the machining database (cf. chapter 3). They become new manufacturing explicit knowledge that can be reused.

$$Time[X_i > T_i] = \sum_{k=t_i}^{t_f} z(k) \cdot \Delta t \quad \text{where} \begin{cases} z(k) = 1 & \text{if } x_i(k) > T_i \\ z(k) = 0 & \text{if } x_i(k) \leq T_i \end{cases} \quad (4.2)$$

The definition of Smart Data and the way to select and aggregate the data are determined based on business rules and some tacit knowledge of manufacturing experts (issued from their experience). For instance, the Smart Data for a given tool usage period can be the duration of chatter ($Time[Nh > T_{Nh}]$), the tool breakage ($CO[Ub > T_{Ub}]$), the critical excessive vibration ($CO[Vrms > T_{Vrms}]$ or $CO[Arms > T_{Arms}]$), or the average cutting power. By this means, meaningful and easy-to-use Smart Data are obtained.

The last step is the aggregation of Smart Data (associated to a given tool-usage period), into Key Performance Indicators (KPIs) for decision aid, dedicated to answering the specific managerial needs of different departments of a machining company. The instantiation of KPI models relies on accurate data selection, which is enabled by operational context data. The definition of a decision-aid indicator is obtained according to four instantiation parameters, which describe the context of the decision:

- the objective of the decision for which decision-aid indicators are defined;
- the decider who exploits the resulting decision-aid indicator according to a given business point of view;
- the scope of analysis in terms of machine, workpiece, program, tool, or other entities;
- the instantiation mode that indicates when a given decision-aid indicator is computed. Three modes are considered: periodic, on-demand, and on-event.

4.1.2 Proposition of new aggregation operators

Sometimes, the criterion is not conform to the target value (e.g. potentiometer for spindle speed $Pot(N)$ or for feedrate $Pot(Vf)$ which are not equal to 100%) due to program errors or human factors. That is why another aggregation operator - Abnormality Operator (AO) - was necessary and has been proposed in this thesis. It records the mean criterion value which is not conform to the target value. Let $X_i = \{x_i(k), k = 0 \dots n\}$ the time series representing an i^{th} monitoring criterion. The operator is defined in Eq. (4.3):

$$AO[X_i \neq C_i] = \frac{\sum_{k=t_i}^{t_f} x_i(k) \cdot e(k)}{\sum_{k=t_i}^{t_f} e(k)} \quad \text{where} \begin{cases} e(k) = 1 & \text{if } x_i(k) \neq C_i \\ e(k) = 0 & \text{if } x_i(k) = C_i \end{cases} \quad (4.3)$$

t_i and t_f are the initial and final time of the considered period (based on operational context data), and C_i represents the target value (=100% in the case of potentiometer). The time during which the criterion does not conform to the target value can be defined in Eq. (4.4):

$$Time[X_i \neq C_i] = \sum_{k=t_i}^{t_f} e(k) \cdot \Delta t \quad \text{where} \begin{cases} e(k) = 1 & \text{if } x_i(k) \neq C_i \\ e(k) = 0 & \text{if } x_i(k) = C_i \end{cases} \quad (4.4)$$

where $\Delta t = 0.1s$ is the sampling period.

Moreover, for traceability in the aeronautic machining industry, the Abnormality Operator (AO) should be able to quantify the mean value when the criterion was below the target value. For example, the lubrication flow rate should be more than the target value of flow rate when *tool is cutting* to avoid the metallurgical issue. In this case, the mean value not conform to the target value is recorded in Eq. (4.5)

$$AO[X_i \leq C_i] = \frac{\sum_{k=t_i}^{t_f} x_i(k).e(k)}{\sum_{k=t_i}^{t_f} e(k)} \quad \text{where} \begin{cases} e(k) = 1 & \text{if } x_i(k) \leq C_i \\ e(k) = 0 & \text{if } x_i(k) > C_i \end{cases} \quad (4.5)$$

where C_i is the target value. The duration when the criterion is below the target value can be recorded in the Eq. (4.6):

$$Time[X_i \leq C_i] = \sum_{k=t_i}^{t_f} e(k).dt \quad \text{where} \begin{cases} e(k) = 1 & \text{if } x_i(k) \leq C_i \\ e(k) = 0 & \text{if } x_i(k) > C_i \end{cases} \quad (4.6)$$

Besides these aggregation operators (CO and AO), the basic statistical operators (e.g. means, medians, standard deviations or maximum values) under specified context (e.g. *tool cutting or not*, *tool entry materials*) are necessary to record and compare the criterion value (e.g. P or $Arms$) in mass production. For example, the mean cutting power (P) under some context (*tool is cutting* or *tool entry material*) is interesting for industry to recognize the criterion value. And it is computed based on Eq. (4.7):

$$means[X_i] = \frac{\sum_{k=t_i}^{t_f} x_i(k).j(k)}{\sum_{k=t_i}^{t_f} j(k)} \quad \text{where} \begin{cases} j(k) = 1 & \text{if } l = 1 \\ j(k) = 0 & \text{if } l = 0 \end{cases} \quad (4.7)$$

Where l represents different context, it can be *tool is cutting* or *tool entry material* etc.. Moreover, other statistical operators under specified context can be computed in the same method.

4.2 KPIs for production organization

4.2.1 Advanced Overall Equipment Effectiveness (OEE)

In the aeronautic machining industry, there are many machine-tools manufacturing in the same facility simultaneously. It is necessary for the process planning to know the load of each machine-tool and organize the production. For this purpose, the advanced Overall Equipment Effectiveness (OEE) is proposed to evaluate the machine-tool use rate.

Overall Equipment Effectiveness (OEE) measures how well a manufacturing resource is utilized (facilities, time, and material) compared to its full potential. It identifies the percentage of manufacturing time that is truly productive. An OEE of 100% means that only good parts are produced (100% quality), at the maximum speed (100% performance), and without interruption (100% availability). In the section, we considered that all the parts are good parts (100% quality) to only measure performance and availability.

In this section, a new KPI called 'advanced OEE' is proposed. It measures the machine-tool use rate during a given period. According to the contextual classification (cf. chapter 2), the whole life time of a machine-tool is composed by the following 3 parts:

- Machine stopped: the machine-tool is stopped (spindle speed $N=0$ or machine-tool feedrate $Vf=0$),
- Motion without cutting: the machine-tool moves ($Vf \neq 0$) and the spindle rotates ($N \neq 0$) and *tool is not cutting*,
- Machine removing material: the machine-tool moves ($Vf \neq 0$) and the spindle rotates ($N \neq 0$) and *tool is cutting*.

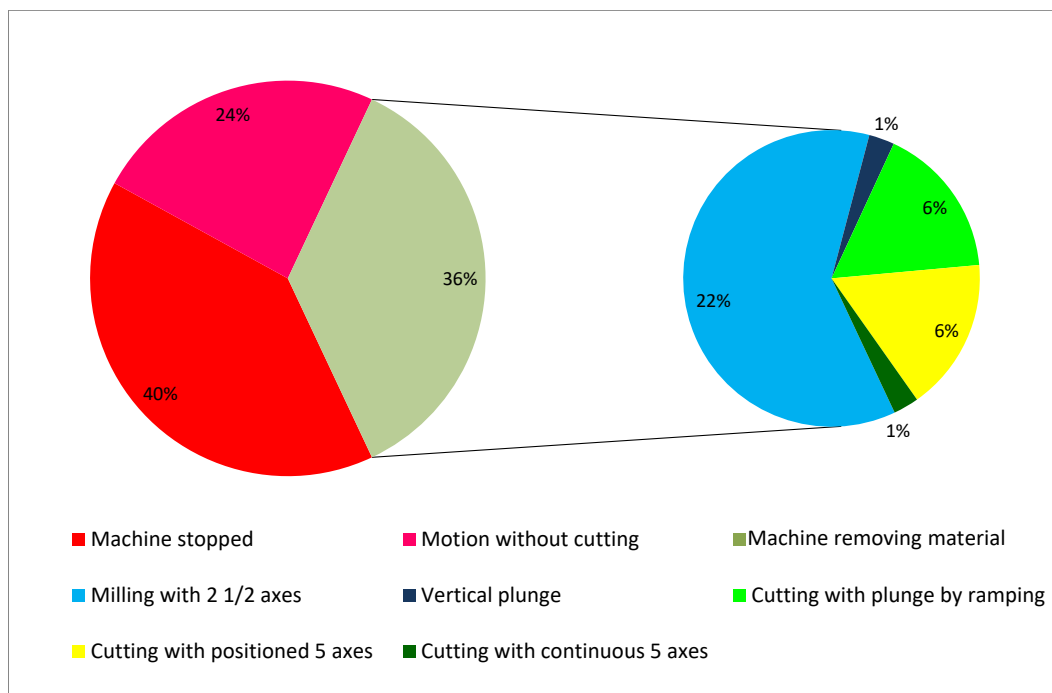


Figure 4.4 – Machine-tool advanced OEE for 17 days database in the machine-tool #1

Moreover, the third part (*tool is cutting*) can also be classified into 5 different operations of machining (cf. Table 2.4, page 35). Then, the machine-tool advanced OEE is composed of the following 7 machining operations:

- the percentage of time when the machine-tool is stopped (in red);
- the percentage of time when the machine-tool moves ($Vf \neq 0$) and the spindle rotates ($N \neq 0$) without cutting (in pink),

- the percentage of time for the *milling with $2\frac{1}{2}$ axes* (in light blue);
- the percentage of time for the *cutting with vertical plunge* (in dark blue);
- the percentage of time for the *cutting with plunge by ramping* (in light green);
- the percentage of time for the *cutting with positioned 5 axes* (in yellow);
- the percentage of time for the *cutting with continuous 5 axes* (in dark green);

The 17 days manufacturing database in the machine-tool #1 is taken as an example. The real machine-tool advanced OEE during the 17 days manufacturing in the machine-tool #1 can be obtained in Fig.4.4. It should be noted that the factory where data were collected works in a three-shift system for 5 days a week and that week-ends are included in the stopped times of Fig.4.4.

This KPI can help the process planning to know the load of each machine-tool. Then it will help the process planning manage many machine-tools in the same facility and help to schedule in the future. For the machine-tool for 5 axes, which is more expensive, a special focus will be made on the percentage of time for *cutting with positioned 5 axes* and *cutting with continuous 5 axes*.

This KPI could also help for future investment to know the state of the factory, and to make a decision to invest a new machine-tool or not, according to this KPI and the customer's demands.

4.2.2 Workpieces production

It is also important for the process planning to visualize how many times a workpiece has been machined. A new KPI called 'Occurrence of machined workpieces' is proposed in Fig. 4.5.

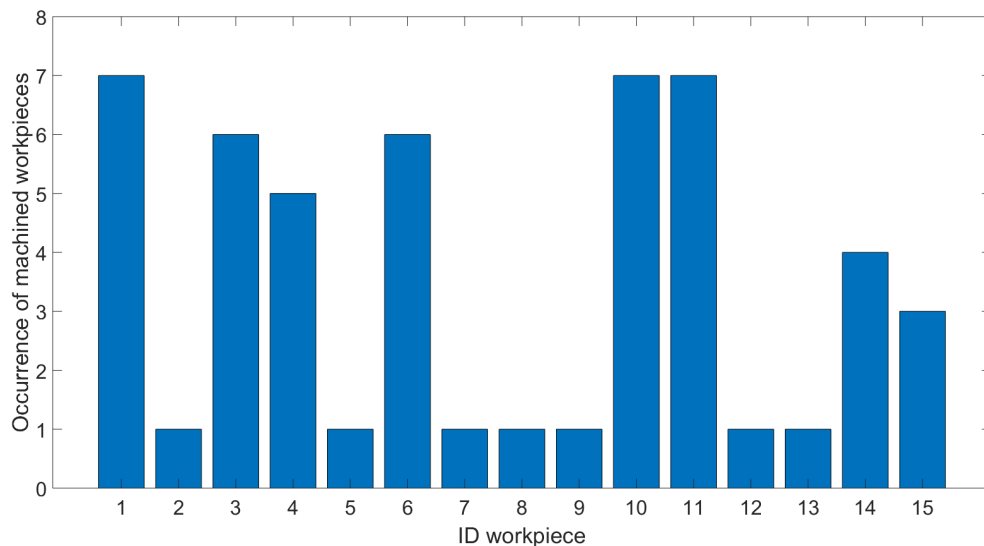


Figure 4.5 – Occurrence of machined workpieces during 17 days on the machine-tool #1

The workpieces N° 1, 3, 4, 6, 10, 11, 14 and 15 have been machined at least 3 times (by a CAM Program). However, workpieces N° 2, 5, 7, 8, 9, 12 and 13 have been machined only once. This production system is thus flexible.

This KPI can give the process planning a global view of the production flow.

4.3 Traceability of Workpiece quality

In machining, the quality of the workpieces is impacted by many factors. Firstly, machine-tools should respect the cutting specifications during machining. Secondly, incidents (e.g. chatter or tool breakage) would directly generate poor workpiece quality.

4.3.1 Respect of cutting specifications

To ensure a good quality of the workpieces, the cutting parameters for each tool (such as spindle rotation speed N , machine-tool feedrate Vf , depth of cut ap (mm), width of cut ae (mm)) should respect the cutting specifications determined by the Manufacturing department; the potentiometers (Pot(N) or Pot(Vf)) should always be at 100% during machining; the lubrication flow rate should be enough during cutting.

Therefore, several KPIs should be considered together to check if machine-tools respect the cutting specifications. The 17 days' database of industrial production on machine-tool #1 is taken as an example. To gain the traceability of workpiece quality, different KPIs for each workpiece (identified by their NumOF, the ID of the Production Order) will be introduced and computed as follows. It should be noted that NumOF starts from 13 in this 17 days' database.

Potentiometer (Pot(N) or Pot(Vf)) should be at 100% during machining. However, the operators could turn the potentiometer to change N or Vf if they meet some urgent situations (e.g. collision or severe vibration). Hence, the KPI of the time during which the potentiometer is not 100% is proposed (cf. Fig. 4.6)

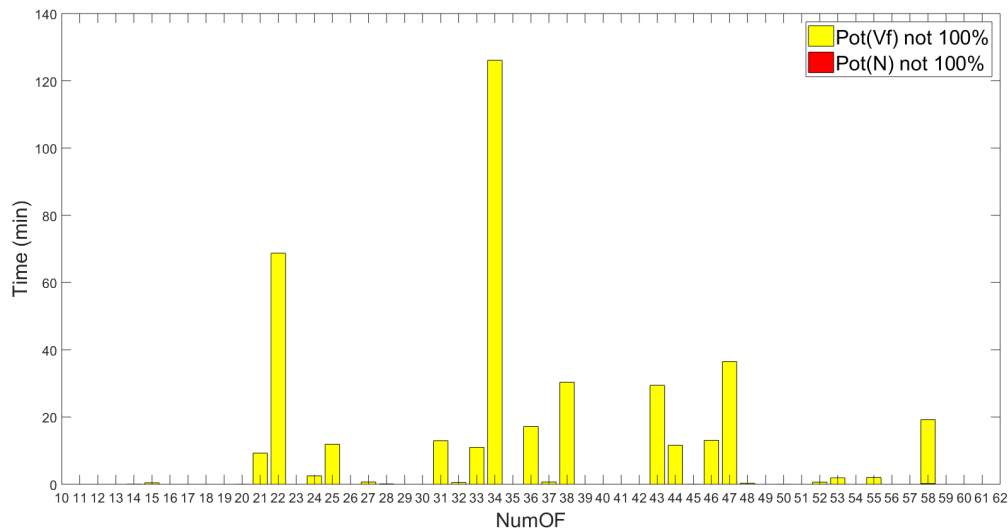


Figure 4.6 – Time when potentiometers differs from 100%, per production order NumOF on the machine-tool #1

Thanks to the aggregation operator - Abnormality Operator (AO) which was introduced in equations (4.3) and (4.4), this KPI can be computed for a given workpiece (NumOF) based on the Smart Data. Figure 4.6 plots the time when a potentiometer differs from 100% for each workpiece. The NumOF 34 takes the longest of time (126min) with $Pot(Vf)$ different from 100%, and the mean $Pot(Vf)$ equals to 2.32% during this period. However, after analysis by manual data mining, machine was stopped ($N=0$ and $Vf=0$) when the $Pot(Vf)$ is not 100% during this NumOF 34, there was no production at that moment.

But when the $Pot(Vf)$ (in yellow dashed line) is not 100% in NumOF 58, there is a period during which *tool is cutting* (in red dots) (cf. Fig. 4.7). It decreases productivity and might produce a poor surface quality.

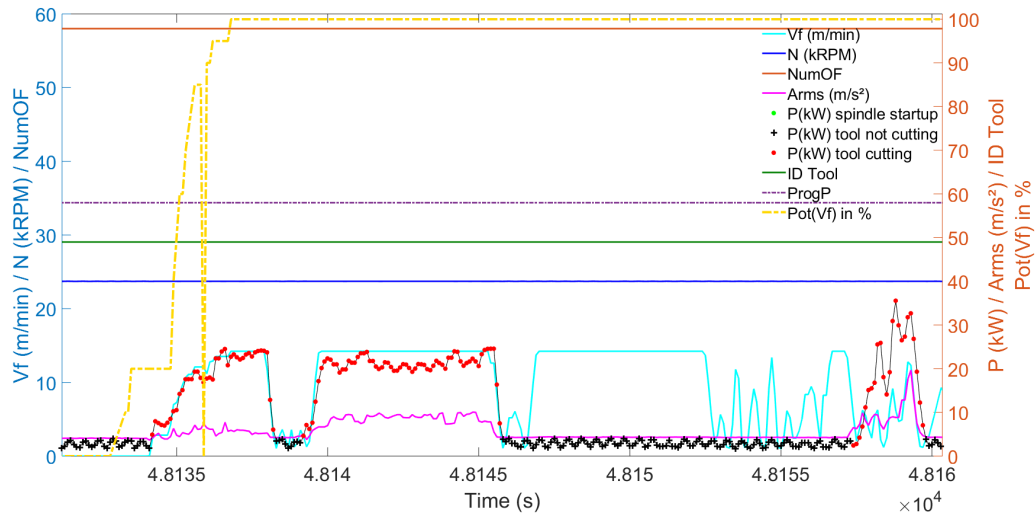


Figure 4.7 – Visualization of the period when $Pot(Vf)$ is not 100% in NumOF 58 on the machine-tool #1

There is only during NumOF 58 that $Pot(N)$ differs from 100% (cf. Fig. 4.8). The N (in blue line) decelerated when the $Pot(N)$ (in red dashed line) changed from 100% to 90%. The spindle power P varied due to deceleration of N , but *tool is not cutting* (in black '+'). However, after the $Pot(N)$ maintained 90% and the N was constant, *tool is cutting* (in red dots) during 1.2s. It is also dangerous for the tool and the spindle.

The spindle rotation speed N is defined by the Manufacturing (based on tap-test and optimization to avoid chatter) and the recommend spindle speed should be respected in every CAM program. However, there may be an error in a given program. To verify if N respects the recommend spindle speed during machining, the KPI such as the time when N didn't respect the recommended spindle speed during cutting for all tools per NumOF is proposed (cf. Fig. 4.9). However, it should exclude the case when $Pot(N)$ not 100% because it has been discussed above. In Figure 4.9, NumOF=17, 18, 19, 20, 21, 22, 23, 32, 33, 35 whose N didn't respect the specification, because the spindle speed N for ID Tool 6 didn't respect the recommend value (19500 RPM). Industry reduced the N to 17550 RPM for this tool due to the chips' problem.

The feed per tooth fz (in mm/tooth) should exceed a given minimal threshold (T_{fz}) when the *tool is*

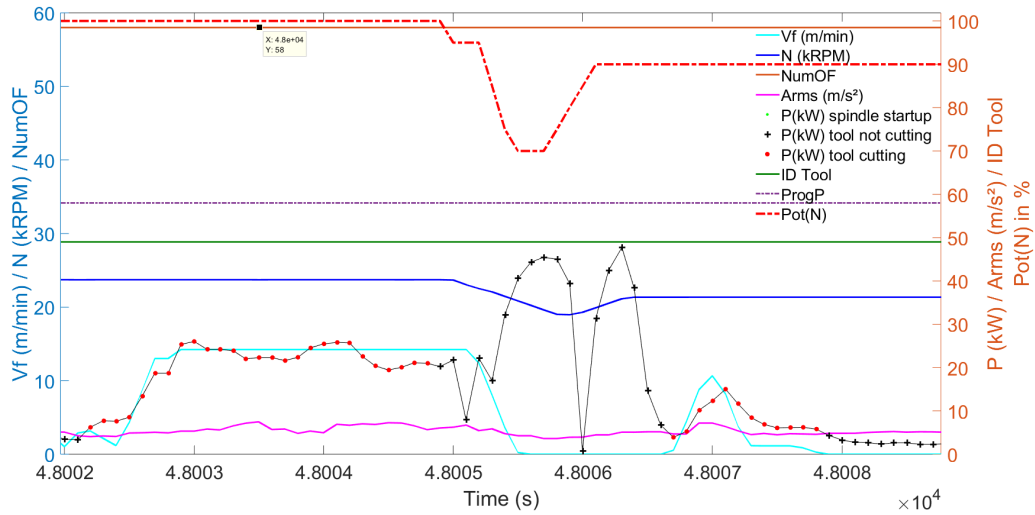


Figure 4.8 – Visualization of the period when $Pot(N)$ is not 100% in NumOF 58 on the machine-tool #1

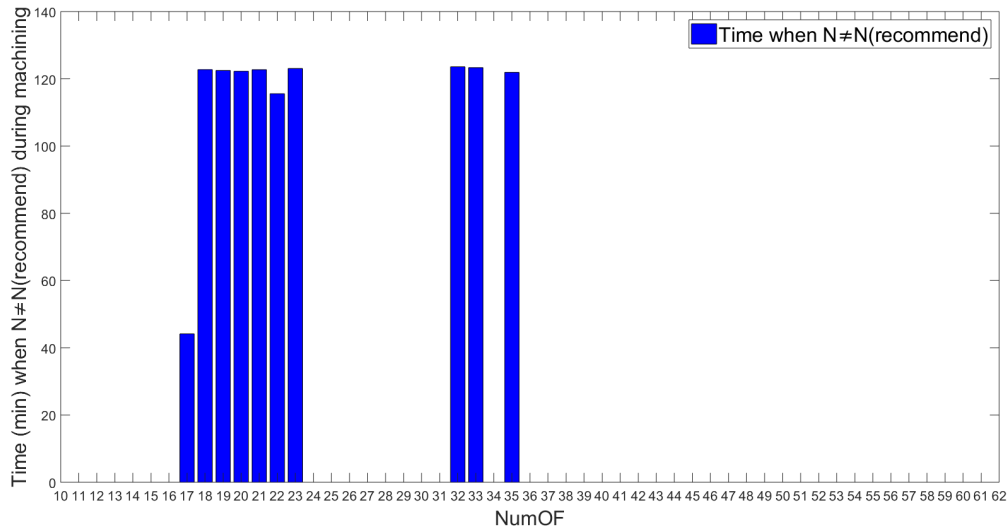


Figure 4.9 – Time when the spindle speed N didn't respect the recommend speed (except $Pot(N)$ not 100%) for all tools per production order NumOF on the machine-tool #1

cutting with constant tool engagement. Then, the KPI corresponding to the time $T[fz < T_{fz} | eng.cst]$ when fz is less than the minimal threshold and *tool is cutting with constant tool engagement* is proposed (cf. Fig. 4.10). The feed per tooth fz (mm/tooth) is calculated as in the Eq. (4.8):

$$fz = \frac{1000.Vf}{Z.N} \quad (4.8)$$

where Vf is the feedrate of machine-tool (in m/min), Z is the number of tooth for the tool used, N is the spindle speed rotation speed (in RPM). According to the Figure 4.10, the NumOF 31 is the longest in which fz is lower than the threshold, it takes 2.5 min. And it was measured at a mean value of 0.0039 mm/tooth.

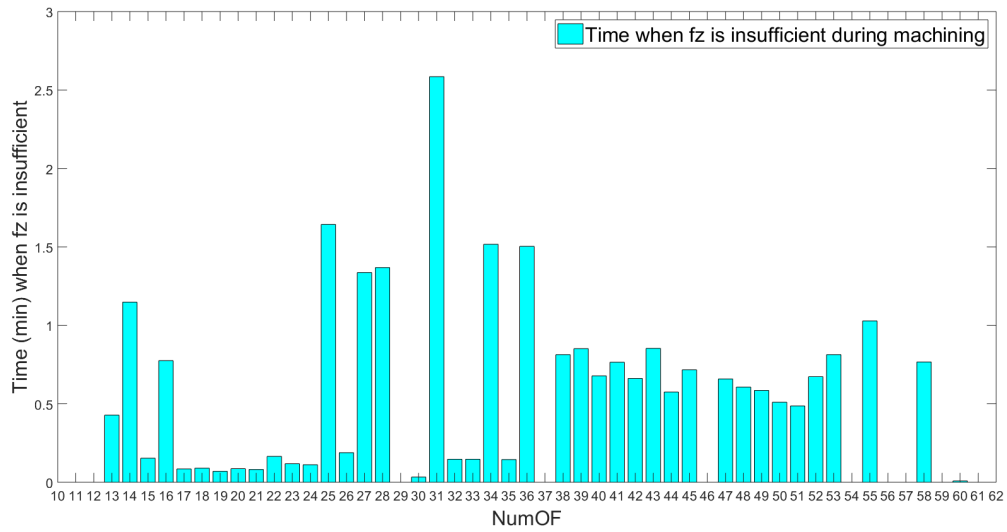


Figure 4.10 – The time $T[fz < T_{fz} | eng.cst]$ per production order NumOF on the machine-tool #1

To visualize the period when fz is insufficient during machining, Fig. 4.11 presents the period in the NumOF 31 when ID Tool 19 is *milling with $2\frac{1}{2}$ axes* (with $N=23720$ RPM). According the Eq. (4.8), the feedrate Vf should exceed a given threshold also. The purple rectangle presents the periods where the Vf is insufficient (fz is insufficient) during machining. It can be observed that it is during turns.

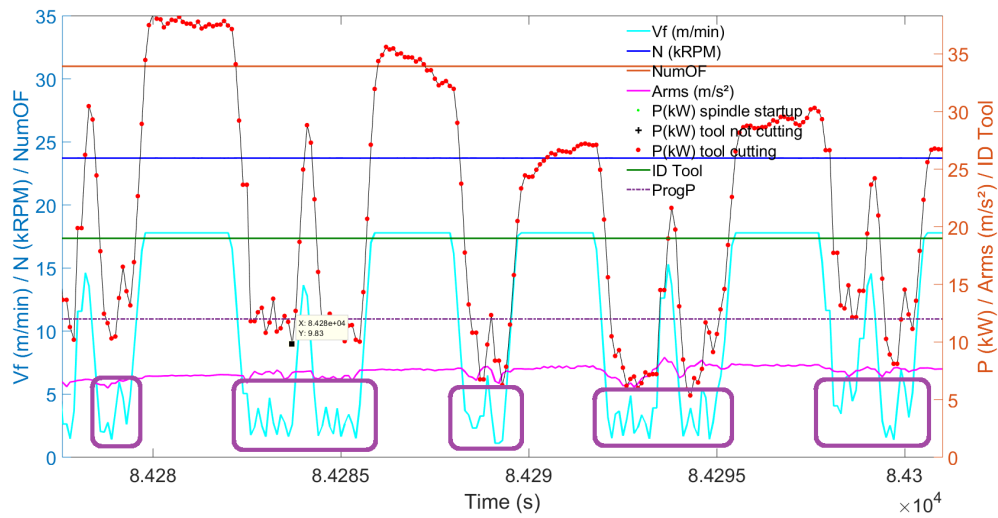


Figure 4.11 – Visualization of the period during which fz is insufficient when the ID Tool 19 is *milling with $2\frac{1}{2}$ axes* in NumOF 31 on the machine-tool #1

The lubrication flow rate QL should exceed a given minimal threshold when *tool is cutting*. Then, the proposed KPI is the time (in orange bar) when QL is lower than the threshold during cutting: $T[QL < T_{QL} | cutting]$. According to Figure 4.12, NumOF 51 is the longest one for which the lubrication flow QL is insufficient and its mean measured value is 21 L/h.

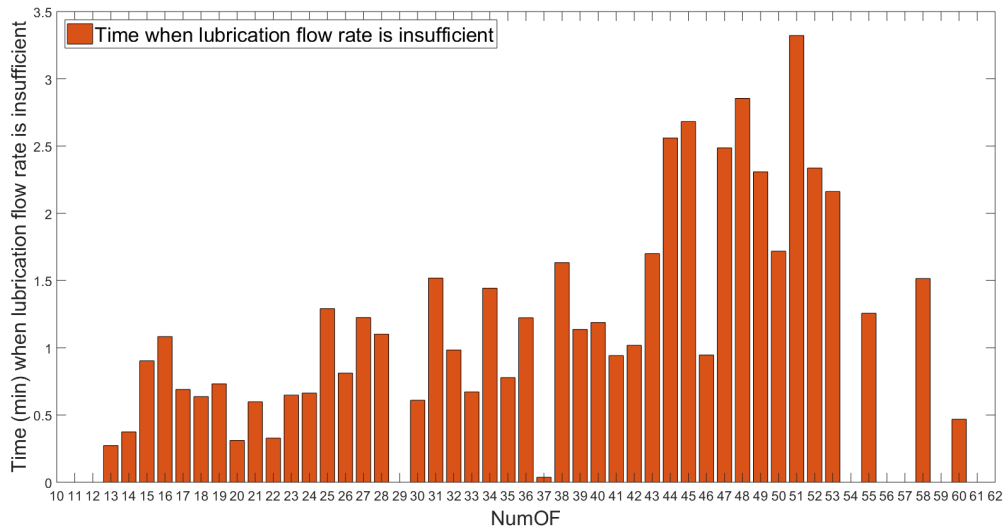


Figure 4.12 – The time $T[QL < T_{QL} | cutting]$ per production order NumOF on the machine-tool #1

Figure 4.13 presents the period in NumOF 51 during which QL (in orange dashed line) is insufficient (in the black rectangle).

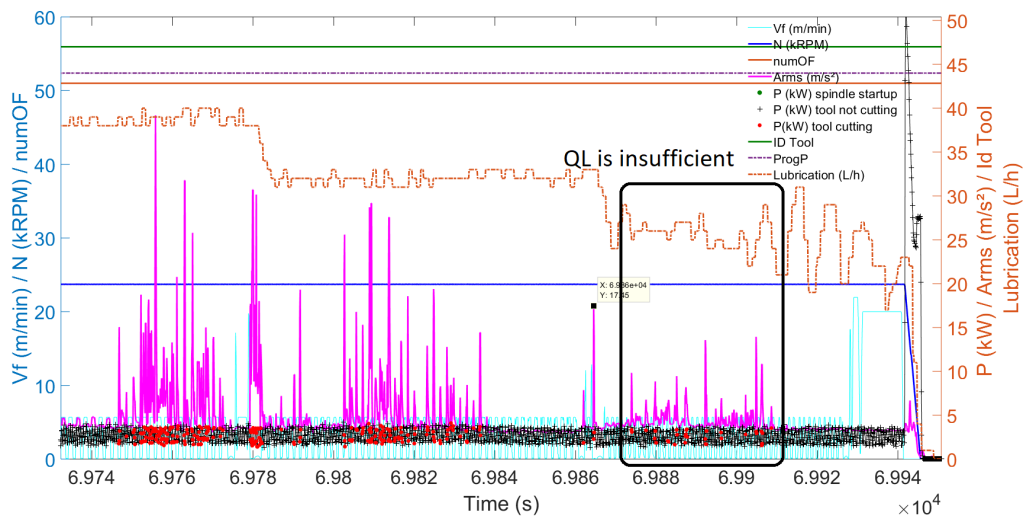


Figure 4.13 – Visualization of the period during which QL is insufficient when the ID Tool 47 is cutting in NumOF 51 on the machine-tool #1

4.3.2 Detection of incidents

To ensure the workpiece quality, there should be neither chatter, nor tool breakage during machining. The proposed KPIs are the chatter time $T[Nh > T_{Nh} | cutting]$ (in green) for each NumOF and the occurrence of tool breakage (in purple) for each NumOF. According to the Figure 4.14, the chatter lasts 0.5s for NumOF 16, and 19.1s for NumOF 31. However, there is no tool breakage in these

database.

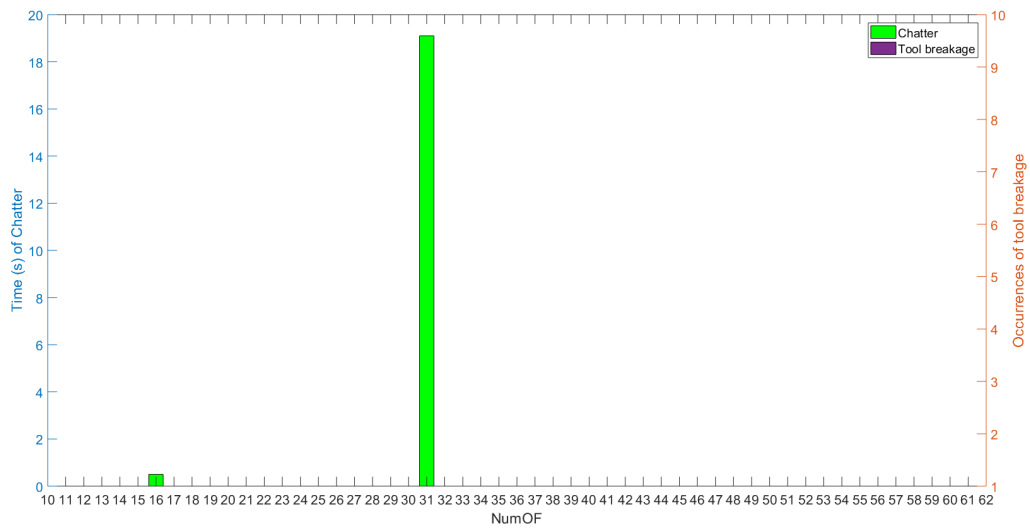


Figure 4.14 – Chatter time $T[Nh > T_{Nh} \mid cutting]$ as well as occurrences of tool breakage per production order NumOF on the machine-tool #1

Figure 4.15 presents the chatter during 19.1s in NumOF 31. The chatter criterion Nh (the green

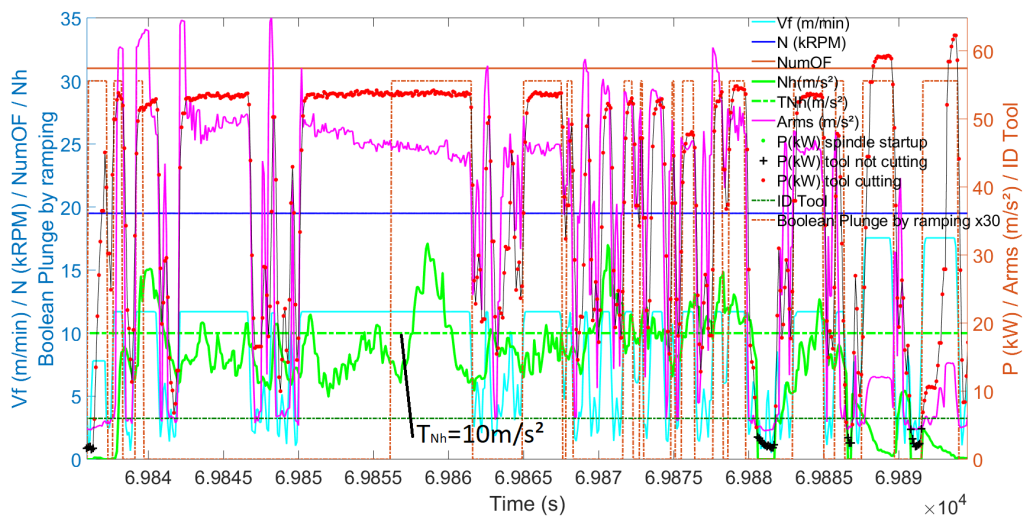


Figure 4.15 – Visualization of the chatter when the ID Tool 6 is cutting in NumOF 31 on the machine-tool #1

line) exceeds the threshold T_{Nh} (in light dashed green line) which was learned as 10 m/s^2 by GMM. To decipher the operational context at the moment when the chatter occurred, the scale of boolean of *Plunge by ramping* operations (brown dashed line) has been multiplied by 30 in the Fig. 4.15 (0-milling with $\frac{1}{2}$ axes and 30-cutting with *plunge by ramping*). By contextual analysis, we identify that, the ID Tool 6 is *cutting with plunge by ramping*, then *milling with $\frac{1}{2}$ axes* alternately during this

chatter area.

All of these KPIs should be considered together to check if the criteria respect the corresponding specifications. Fig. 4.16 summarizes all of these KPIs (the respect of cutting specifications as well as the incidents detected) during the 17 days' production in machine-tool #1. Any workpiece that did not respect the specifications (or incidents occurred) can be pointed out by the Quality department thanks to these new KPIs.

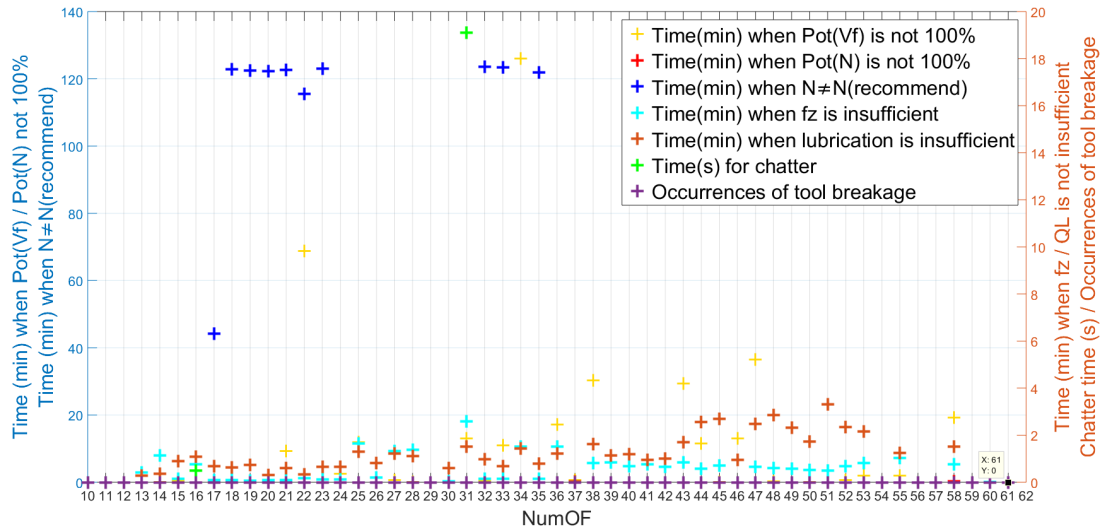


Figure 4.16 – Summary of the traceability of workpiece quality per Production Order NumOF on the machine-tool #1

4.4 KPIs for production monitoring

This section focuses on the KPIs for production monitoring. They will help the Manufacturing department to check: the traceability of duration of machining operations; the traceability of cutting aspects; if the potentiometer (for V_f and N) are always 100% during production; if there are any incidents (chatter, tool breakage or excessive vibration) during production.

4.4.1 Traceability of duration of machining operations

In the case of large batch production, it is important for the Production department to record the distribution duration of machining operations. It is estimated by the use duration of each tool, for the machining of a given part (more precisely by the time during which a tool is in spindle). If one tool is used longer than usual, it should be pointed out.

The 10 days' manufacturing database in machine-tool #4 is taken as an example. The distribution of tool use duration during these days is plotted in Fig.4.17.

During this period, 114 Titanium parts were machined. The machine-tool always used the same CAM program to obtain the same part. And during this CAM program, it always uses the same

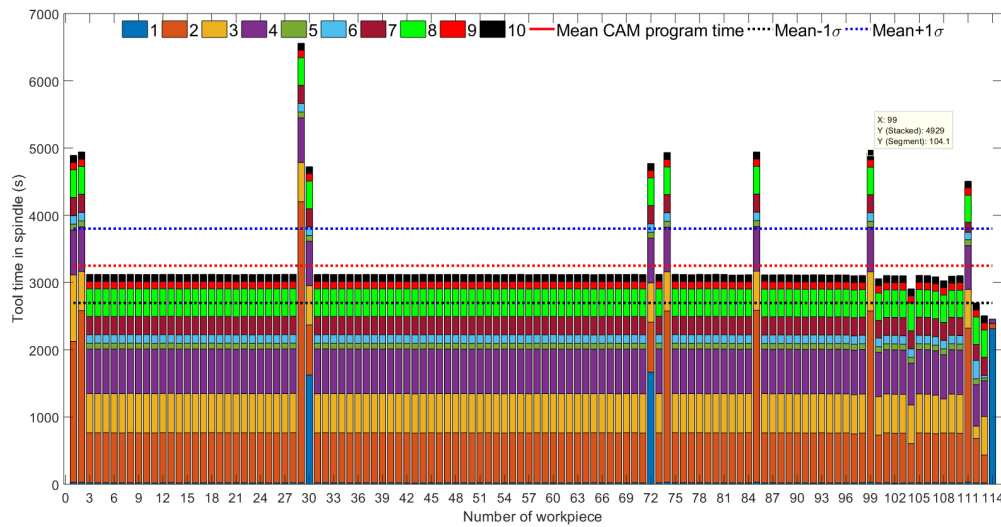


Figure 4.17 – Distribution of tool use time in spindle via workpiece during 10 days’ production in the machine-tool #4

series of tools with the same order (from ID Tool 1 to ID Tool 10). The duration for each ID Tool is plotted in different colors in the figure. Usually, the tool use duration for each ID Tool should not change. However, for some workpieces, some tools were used longer (or shorter) than for the other workpieces. The mean value of this CAM program for all these 114 workpieces is calculated and plotted in red dashed line. Around this mean value of duration, thresholds were set at $\pm 1\sigma$. Outside these thresholds, machining is considered as abnormal. The Production department should pay more attention to the workpieces 1, 2, 29, 30, 72, 74, 85, 99 and 111, whose CAM program times are much longer, and also the workpieces 113 and 114, whose CAM program time are much shorter. For the workpieces 1, 2, 29, 74, 85, 99 and 111, the longer CAM program time can be explained by the fact that ID Tool 2 was used one time more than normal workpieces. For workpieces 30 and 72, it is because they use ID Tool 1 (not a cutting tool) one time more than normal workpieces. And for the workpiece 113, the much shorter CAM program times are due to ID Tool 2, which use is much shorter than normal workpieces. For workpiece 14, which is the final workpiece, Emmatools just record one part of data, data is not complete.

This KPI can help the Production department to find the abnormal duration of machining of workpieces, during large batch production.

4.4.2 Traceability of cutting aspects

Usually, the cutting power P consumed by the same tools in the same machining program (same cutting parameters) should not change a lot. However, when the tool is worn, the power consumed by this tool will increase, compared to the tool in good condition. Therefore, the traceability of the power consumed by tools for each workpiece is proposed.

Figure 4.18 presents the power P_{cut} consumed by the ID Tool 3 during 10 days’ database in machine-tool #4 that produces Titanium parts. The days are in cyan line. There are only 10 days’ production

data available. The P_{cut} is equal to the mean spindle power P subtracting the power during idle motion P_{idle} . The P_{idle} (in black line) for this ID Tool 3 is constant (0.12 kW), which is calculated as the mean spindle power when the machine-tool is in idle motion with this tool.

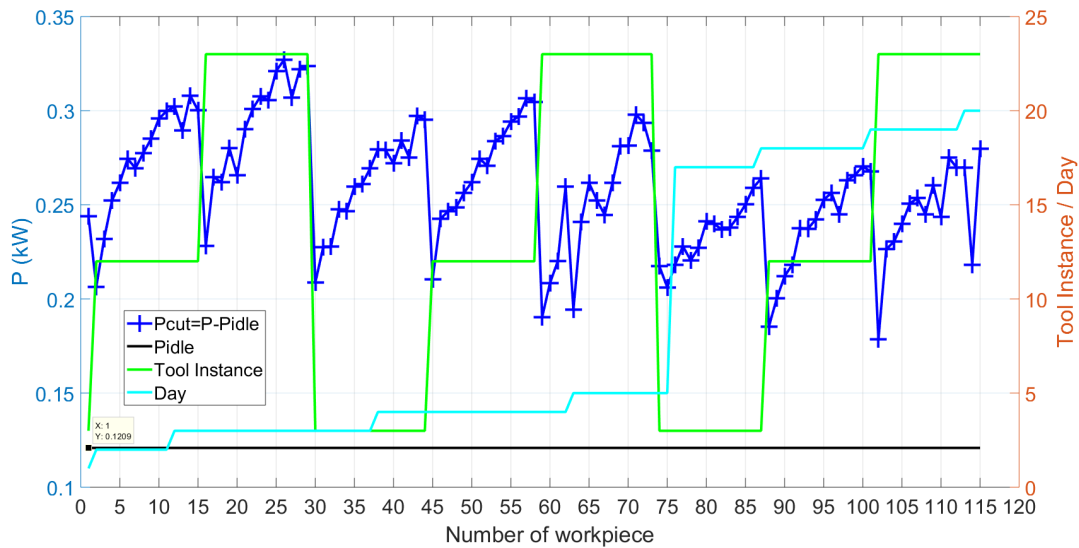


Figure 4.18 – P_{cut} as well as the Tool Instance (in green) for ID Tool 3 during 10 days in the machine-tool #4

According to Fig. 4.18, the cutting power consumed increases regularly with tool wear. Each time the Tool Instance (in green line) is changed, the power P_{cut} comes back to a lower value. That shows the relevance of P_{cut} as a criterion of tool worn for hard-to-cut materials. In other words, if the power P_{cut} is much higher than the normal value, the tool could be worn.

This KPI can help the Manufacturing department to monitor the tool condition.

It is interesting for the Manufacturing department to have the information like mean power (P) consumed by the cutting tool as well as the time when severe vibration occurred ($T[V_{rms} > 8 \text{ mm/s}]$ according to [ISO 10816, 2009]). This information is particularly interesting for large batch production. Therefore, KPIs such as the mean P (in black) during cutting per workpiece (NumOF) as well as the time $T[V_{rms} > 8 \text{ mm/s}]$ (in dark green) per NumOF is proposed (cf. Fig.4.19). Pay attention to NumOF=29, 37, 54, 56, 57, 59, 61 for which the mean value of cutting power P is zero. It means that these workpieces were not machined. After manual data mining, the NumOF=37, 54 and 61 didn't appear in the database due to the loss of connection between machine-tool and Emmatools. Moreover, for NumOF=29, 56, 57 and 59, the machine was stopped ($N=0$ and $Vf=0$) during some seconds. Operators loaded the workpiece (NumOF), but nothing was machined, and they immediately moved on to the next workpiece.

4.4.3 Human factor during production

During production, an operator may turn the potentiometer in case of danger or after an incident that he/she has detected (cf. section 4.3). It is interesting for the Production department to know if the production of a workpiece was interrupted or not. This KPI can be computed from potentiometer

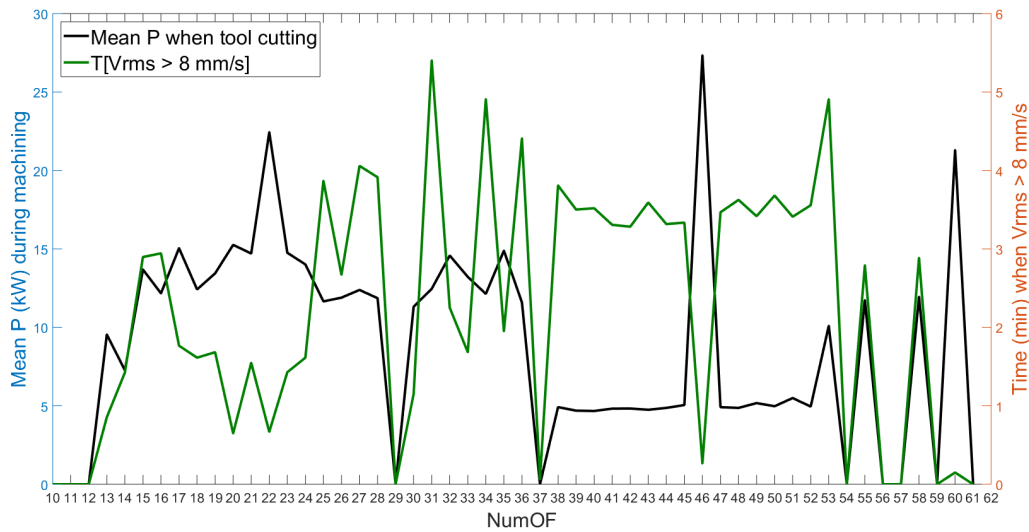


Figure 4.19 – Mean P and the time $T[V_{rms}>8 \text{ mm/s}]$ during cutting per NumOF during 17 days' production in the machine-tool #1

values, similarly to section 4.3.1 and Fig. 4.6 concerning the respect of cutting specifications.

4.4.4 Incidents detected during production

It is also important for the Production department to know the workpieces for which incidents occurred (e.g. chatter, tool breakage, collisions, or excessive vibration). These KPIs can help the Manufacturing department to identify the workpieces with poor surface quality and to ask for a polishing operation to solve the issue for these workpieces. These KPIs can be computed by the same method as in Figure 4.14.

4.5 KPIs for tool performance

This section focuses on the KPIs for the performances of tools, including the distribution of chatter per tool; distribution of excessive vibration per tool, and tool use duration before tool breakage.

4.5.1 Distribution of chatter per tool

For Manufacturing engineers, it is interesting to know the distribution of chatter during a given period. Then, a KPI such as the distribution of chatter per kind of machining operations for the 10 most problematic tools is proposed. The chatter includes punctual chatter (PC) (chatter time $< 0.2\text{s}$) and continuous chatter (CC) (chatter time $\geq 0.2\text{s}$) (cf. section 3.2). Moreover, the chatter can be impacted by the spindle's condition ([Ritou et al., 2018]). Firstly, a period of database during which the spindle is in good condition is considered. The spindle #1 database (215 days) in machine-tool #2 is taken as an example. Figure 4.20 presents the distribution of chatter per kind of machining operations for the 10 most problematic tools when the spindle is in good condition in machine-tool #2, based on the contextual classification in chapter 2.

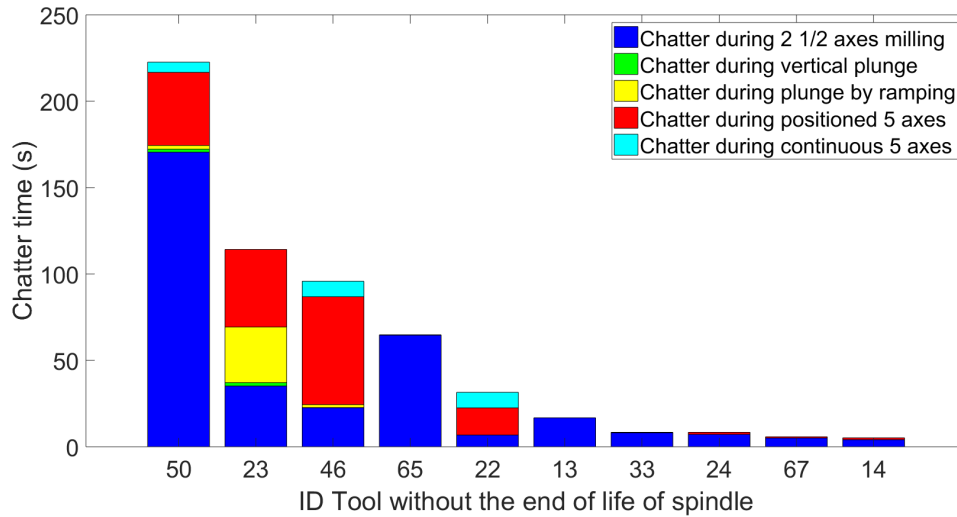


Figure 4.20 – Distribution of chatter per kind of machining operations for the 10 most problematic tools (when spindle #1 is in good condition in machine-tool #2).

Table 4.1 – Chatter duration (PC in green + CC in yellow) per machining operation per kind of tool engagement for ID Tool 23 in machine-tool #2.

Chatter time [s]	2½ axes milling		Vertical plunge		Plunge by ramping		Positioned 5 axes		Continuous 5 axes	
	PC	CC	PC	CC	PC	CC	PC	CC	PC	CC
Entry materials	0.8	4.7	0	0	0.1	4	1.4	2.5	0	0
Exit materials	0.5	0	1.1	0	0.1	0	1.2	0.9	0	0
Turns	2.1	1.8	0.8	0	0	15.9	2.4	22.4	0	0
Static tool eng.	1.6	19.6	0	0	0	7.1	0.8	5.2	0	0
Var. tool eng.	0	4.1	0	0	0.1	4.9	0.3	7.8	0	0

When the spindle was in good condition, chatter occurred during $2\frac{1}{2}$ axes milling (in blue), during positioned 5 axes (in red), during continuous 5 axes (in cyan) and during plunge by ramping (in yellow) for ID Tool 50. However, chatter also occurred during positioned 5 axes for ID Tool 23, 46 and 22. And chatter also occurred during plunge by ramping for ID Tool 23.

To decipher the context in which chatter appeared, the distribution of chatter time per kind of tool engagement for each operation and for each tool (with chatter problem) can be calculated. Table 4.1 summarizes the chatter duration (including punctual chatter and continuous chatter) per machining operation and per kind of tool engagement, for ID Tool 23. Most of time, 3 types of machining are concerned: the chatter during turns in plunge by ramping (15.9s for CC), the chatter during turns in positioned 5 axes (2.4s for PC and 22.4s for CC) as well as the chatter during static tool engagement in $2\frac{1}{2}$ axes milling (1.6s for PC and 19.6s for CC).

If the period during which the spindle is worn is also considered (the monitoring of the spindle

condition is discussed in [De Castelbajac et al., 2014]), the distribution of chatter per kind of machining operations for the 10 most problematic tools is changed, as can be observed in Fig. 4.21:

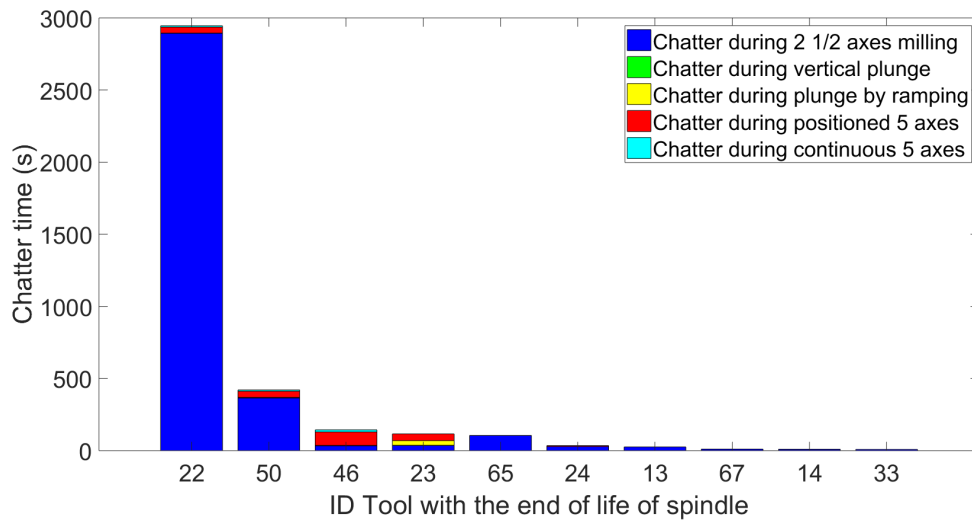


Figure 4.21 – Distribution of chatter per kind of machining operations for the 10 most problematic tools (with the end of life of spindle #1 in machine-tool #2)

The sum of chatter duration is much longer than the sum without the end of life of spindle. That is because a worn spindle can generate more chatter. The chatter mainly occurred during $2\frac{1}{2}$ axes milling (in blue) when *tool is cutting during turns* and only with ID Tool 22. Also, chatter slightly occurred during *positioned 5 axes* (in red) for ID Tool 22, 50, 46, and 23. Moreover, the chatter occurs during *plunge by ramping* (in yellow) for ID Tool 23.

The same method, the distribution of chatter per operation and per kind of tool engagement for each problematic tool can be obtained by the same method as in Table 4.1.

This KPI can help the Manufacturing department to identify when chatter occurred for each tool, for which operations and which tool engagement. The Manufacturing department could optimize the cutting parameters for the ID Tool to solve the chatter problem.

4.5.2 Distribution of excessive vibration per tool

The excessive vibration could be detected by the business rules listed in the section 3.4:

- If $(Arms > T'_{Arms}$ or $Vrms > T_{Vrms})$ and '*tool is cutting*' and 'Chatter detection=0' and 'Tool breakage detection=0', then 'excessive detection=1'.

Based on this rule, excessive vibrations can be detected in real-time. Then, the KPI of the distribution of excessive vibration per tool is here proposed. The spindle #1 database (215 days) in machine-tool #2 is taken as an example. Fig.4.22 presents the distribution of excessive vibration per tool during the whole lifetime (215 days) of spindle #1.

The total duration of excessive vibration is 25 min. The ID Tool 36 (in red) and ID Tool 26 (in purple)

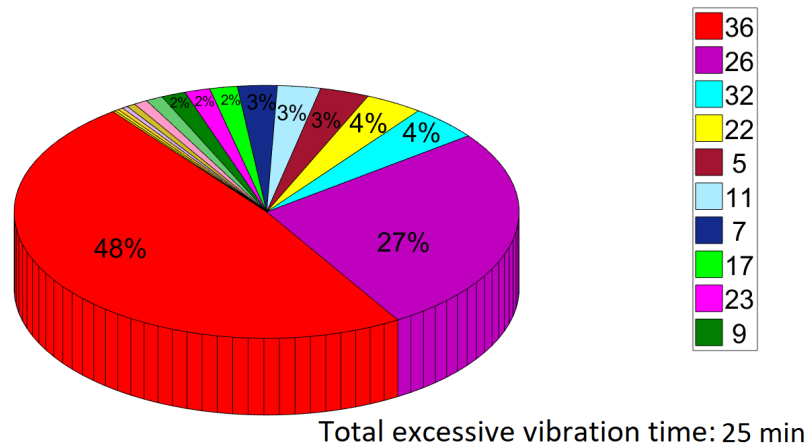


Figure 4.22 – Distribution of excessive vibration per tool (during the whole spindle #1 lifetime of machine-tool #2)

concern most of the vibrations (ID Tool 36 - 48% and ID Tool 26 - 27%). Attention should be paid to these tools.

This KPI could help the Manufacturing department to identify which ID Tools generate the most excessive vibration and to deal with them.

4.5.3 Tool use duration before tool breakage

In machining, the cutting tool life is impacted by the cutting parameters (e.g. spindle rotation speed N , feedrate V_f , depth of cut ap or width of cut ae). Taylor [1906] firstly defined an empirical relationship between tool lifetime and cutting speed, for turning. Due to tool wear, a given ID Tool is regularly changed (and replaced by a new Tool instance).

Each ID Tool has a tool lifetime (DVP) defined by the Manufacturing department. Tool breakages were detected for the spindle #1 and #2 database in machine-tool #2 (Fig. 3.18, page 65). It is interesting to know when a tool breakage occurred, in relation to its theoretical lifetime. It can be computed as presented in Fig.4.23.

This KPI can help the Manufacturing department to adjust the recommended lifetimes for the tools.

4.6 KPIs for program performance

This section focuses on the KPIs for the CAM programs, including the distribution of chatter per program; distribution of excessive vibration per program; the productivity of CAM programs and slow idle motion.

4.6.1 Distribution of chatter per program

Similarly to the distribution of chatter per tool presented in section 4.5, the distribution of chatter per program can be computed by the same method, based on the contextual classifications in chapter

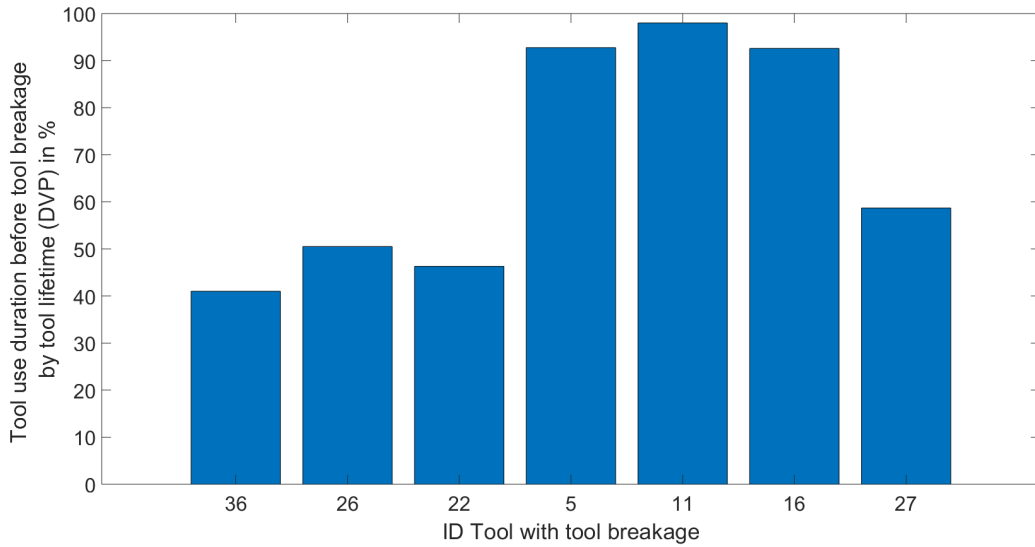


Figure 4.23 – Tool use duration before tool breakage by tool lifetime in % (in blue) for the problematic ID Tools during the spindle #1 and #2 in the machine-tool #2

2. The KPI of the distribution of chatter per kind of machining operations and per kind of tool engagement for the 10 most problematic programs is proposed. Firstly, a period of database with a good condition of spindle is analyzed. The spindle #1 database of machine-tool #2 is taken as an example. Figure 4.24 presents the distribution of chatter per kind of machining operations for the 10 most problematic programs when the spindle is in good condition.

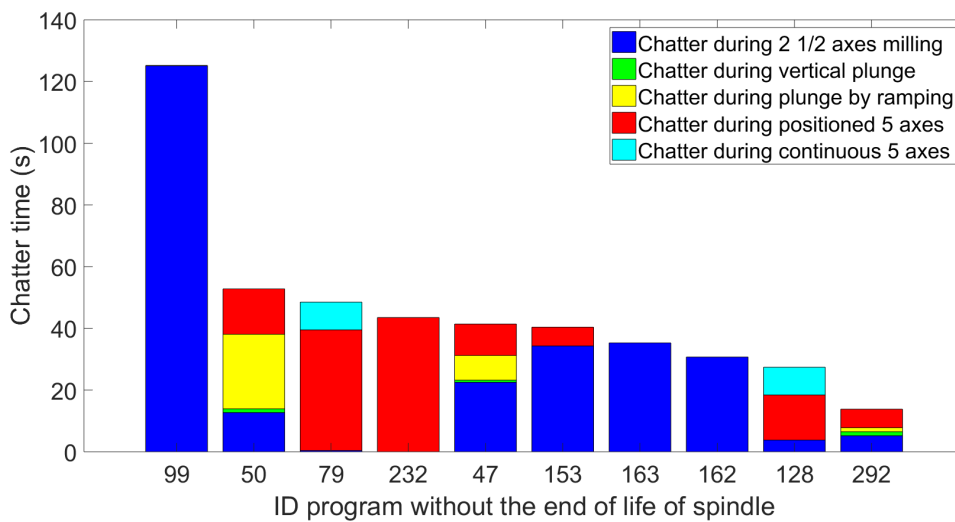


Figure 4.24 – Distribution of chatter per kind of machining operations for the 10 most problematic programs (when the spindle is in good condition for machine-tool #2)

Chatter only occurs during $2\frac{1}{2}$ axes milling (in blue) for ID Program 99. However, the chatter also occurs during *positioned 5 axes* (in red) for ID Program 50, 79, 232, 47, 153, 128 and 292. And the chatter also occurs during *plunge by ramping* (in yellow) for ID Program 50, 47 and 128.

Table 4.2 – Chatter duration (PC in green + CC in yellow) per machining operation per kind of tool engagement for ID program 50, and for ID Tool 23 in machine-tool #2

Chatter time [s]	2½ axes milling		Vertical plunge		Plunge by ramping		Positioned 5 axes		Continuous 5 axes	
	PC	CC	PC	CC	PC	CC	PC	CC	PC	CC
Entry materials	0.5	1.9	0	0	0	3.1	0.2	0	0	0
Exit materials	0.3	0	0.7	0	0.1	0	0.3	0	0	0
Turns	0.9	1	0.5	0	0	11.5	0.4	7.4	0	0
Static tool eng.	1.2	5.8	0	0	0	5.1	0	2.8	0	0
Var. tool eng.	0	1.1	0	0	0	4.4	0	3.6	0	0

To decipher the context when the chatter appears, the distribution of chatter duration per kind of tool engagement for each operation and for each ID program (with chatter problem) can be calculated. Table 4.2 summarizes the chatter duration per machining operation and per kind of tool engagement, for ID program 50 and for ID Tool 23. Among them, the chatter during *turns* in *plunge by ramping* (11.5s for CC) and the chatter during *turns* in *positioned 5 axes* (0.4s in PC and 7.4s in CC) as well as the chatter during *static tool engagement* in *2½ axes milling* (1.2s in PC and 5.8s in CC) concern most of the chatter time.

If the period when the spindle is worn is considered also, the distribution of chatter per kind of machining operations for the 10 most problematic programs can be obtained in the 4.25:

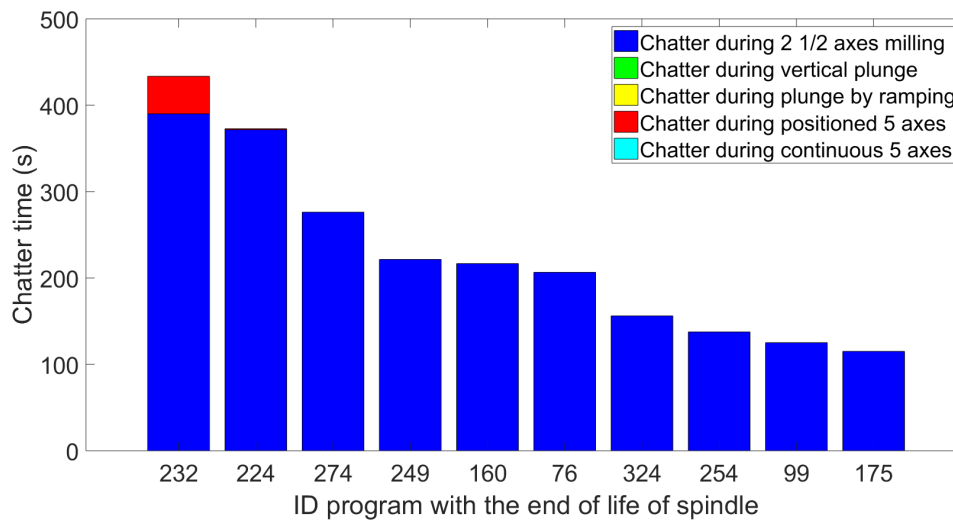


Figure 4.25 – Distribution of chatter per kind of machining operations for the 10 most problematic programs (with the end of life of spindle for machine-tool #2)

The sum of chatter time is much greater than the sum without the end of life of spindle. That is because a worn spindle can generate more chatter. The chatter mainly occurs in ID Program 232 during *2½ axes milling*, compared with the most problematic ID Program 99 without the end of life

of spindle (cf. 4.24). However, the chatter duration in ID Program 99 is in the 9th most problematic program with chatter.

The same method, the distribution of chatter per kind of tool engagement for each operation and for each ID program (with chatter problem) with the end of life of spindle can be calculated by the same method as for Table 4.2.

This KPI can help the CAM programming department to identify where is the chatter for ID programs, and in which operations and in which state of tool engagement when *tool is cutting*.

4.6.2 Distribution of excessive vibration per program

As for the distribution of excessive vibration per tool presented in section 4.5, the KPI of the distribution of excessive vibration per program is proposed. The spindle #1 database in machine-tool #2 is taken as an example. Figure 4.26 presents the distribution of excessive vibration per program during the whole lifetime (215 days) of spindle #1.

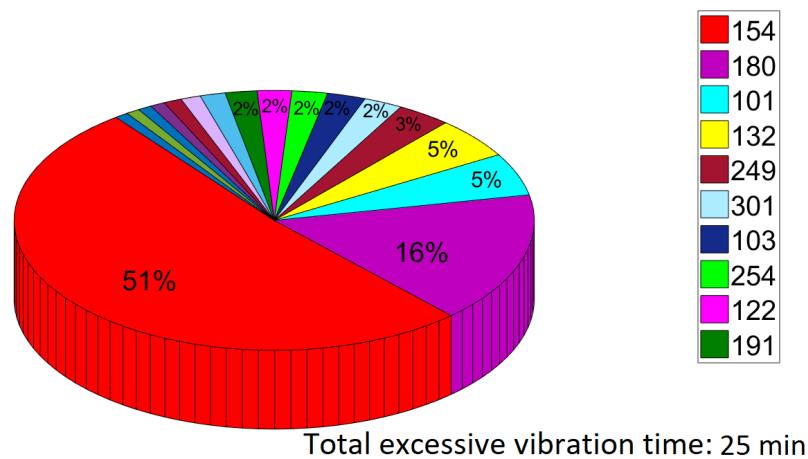


Figure 4.26 – Distribution of excessive vibration per program (during the whole spindle #1 life time's production in the machine-tool #2)

The total excessive vibration time was 25 min (as in Fig.4.22). The ID Program 154 (in red) and ID Program 180 (in purple) concern most of the vibration (Program 154 - 51% and Program 180 - 16%).

This KPI could help CAM programming department to identify which ID Program holds the most excessive vibration. Excessive vibrations are forced vibrations (because chatter is excluded). CAM programming department could improve the cutting condition to avoid excessive vibration.

4.6.3 Productivity of CAM programs

Material Removal Rate (MRR in cm^3/min) is the most commonly used optimization criterion of production rate in machining process [Yan and Li, 2013]. It can be computed by the expression in the Eq. (4.9):

$$MRR = \frac{P_{cut}}{Esp} \quad (4.9)$$

where, P_{cut} is the spindle cutting power and Esp (in $W.min/cm^3$) the specific cutting energy of which the value is provided by the Manufacturing department. The MRR for each tool in each CAM program can be calculated from the collected data. The spindle #1 database in the machine-tool #2 is taken as an example. Figure 4.27 presents the MRR for each ID Program when the ID Tool 5 (a tool mainly used for rough milling) is used in machining as well as the mean MRR of ID Tool 5. The diameter of this tool D is 40 mm, and the number of cutting teeth Z is 4.

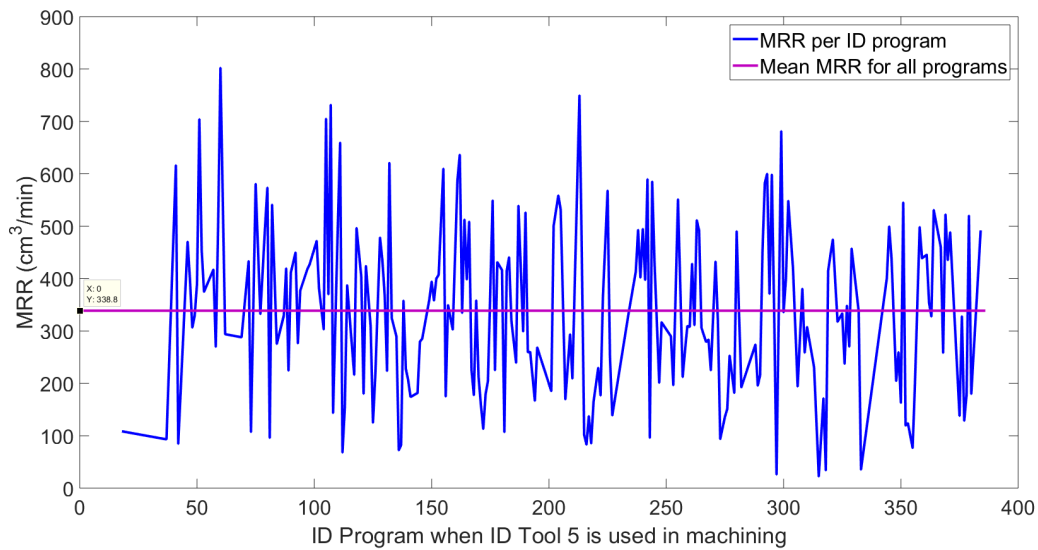


Figure 4.27 – MRR for each ID Program when the ID Tool 5 is used in machining as well as the mean MRR of ID Tool 5 during the whole spindle #1 lifetime in the machine-tool #2

The abscissa shows the ID Program when ID Tool 5 is used in machining, the ordinate is the MRR (in cm^3/min). The MRR for each ID Program is presented in blue line, and the mean MRR for ID Tool 5 is plotted in purple line. The maximum of MRR ($802 cm^3/min$) was in the ID Program 60. And the minimum of MRR ($26 cm^3/min$) was in ID Program 297 with a very low productivity.

This KPI could help CAM programming department to evaluate the production rate for each program and for each tool and improve its programming in the future.

4.6.4 Slow idle motions

During production, when the *tool is not cutting*, it is better than the machine-tool moves (Vf) at a higher speed to make the production more efficient. Therefore, it is important to set up a KPI to evaluate the motion of machine-tool when *tool is not cutting*. However, it is interesting to calculate the motion of machine-tool when the spindle (N) is stopped or the spindle rotates at varying speed, because $N = 0$ or N varying is during the preparing program, not the main machining program. Then, the Vf when the *tool is not cutting*, and N rotates at constant speed is constructed. The database during the whole spindle #1 lifetime (215 days) for machine-tool #2 is analyzed. 346 main

machining programs (ID Program) were operated during these 215 days. The mean V_f when the *tool is not cutting* for each ID Program is computed and results are presented in Fig.4.28.

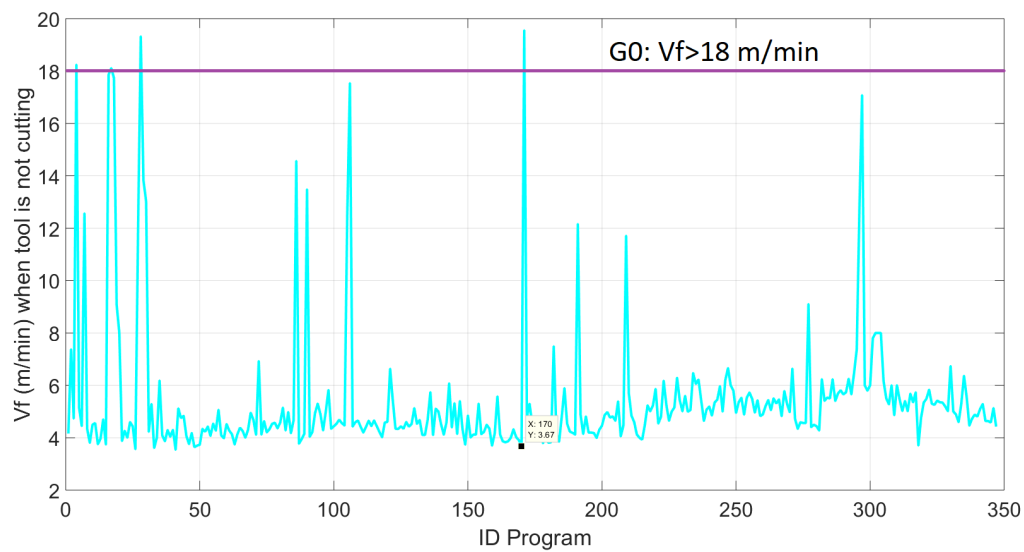


Figure 4.28 – Mean V_f when the *tool is not cutting* per ID Program for the machine-tool #2

The rapid motion G0 is defined as a state of the machine-tool when the *tool is not cutting* and the machine-tool moves at a higher feedrate (V_f). Theoretically, the more G0 is used when the *tool is not cutting*, the more efficient the machine-tool will be. However, it depends on the optimization of the CAM program and on the complexity of the produced piece. Moreover, G0 is machine-tool specific. For machine-tool #2, G0 is defined as $V_f > 18$ m/min.

In Figure 4.28, only the ID Program 4, 17, 28 and 171 have their mean V_f (when the *tool is not cutting*) in the G0 area. It means the other CAM programs can be improved (closer to G0).

This KPI is interesting for the CAM programming department to improve the programs.

4.7 KPIs for monitoring of machine-tool condition

To help the Maintenance department to monitor the machine-tool condition, KPIs such as spindle condition monitoring and incidents detection are also important.

4.7.1 Spindle condition monitoring

In the analyses of the distribution of chatter per tool and the distribution of chatter per program, chatter occurs more at the end of life of spindle. It means that the spindle condition has decreased the stability limit of some tools ([Godreau et al., 2019]). That is why the KPIs for spindle condition monitoring is important for industry. A vibration signature is performed once a day (during idle rotations), during which vibrations are measured to evaluate the spindle condition. The spindle condition is evaluated through the bearing fault vibrations (BPFO, BPFI and FTF) that appear at specific frequencies in the spectrum when an element of the bearing is damaged. It is a classical

technique for condition-based maintenance ([De Castelbajac et al., 2014]).

4.7.2 Incidents detection

The incidents, including chatter, tool breakage, collision and excessive vibration, can damage the machine-tool condition. Then the KPI such as the incidents detected (chatter, tool breakage, collision and excessive vibration) per day are proposed. Fig.4.29 plots the chatter time (in green) and the occurrences of tool breakage (in purple) as well as the spindle condition criterion BPFO (in orange) during the whole spindle #1 lifetime in the machine-tool #2.

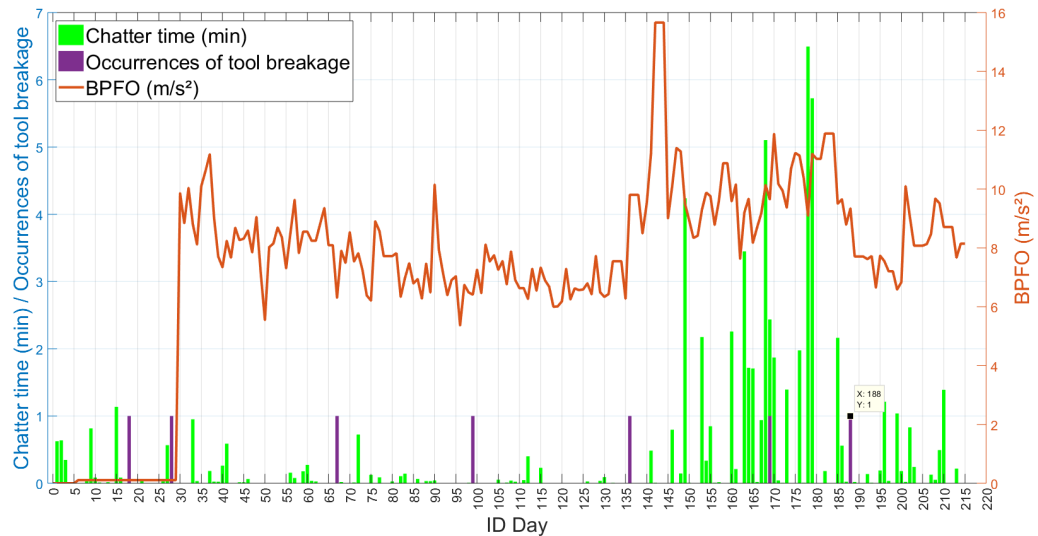


Figure 4.29 – Chatter time and the occurrences of tool breakage as well as the spindle condition criterion BPFO during the whole spindle #1 lifetime in the machine-tool #2

According to Fig. 4.29, chatter mainly occurred at the end of life of spindle. Godreau [2017] has established correlations between the spindle condition and the incidents.

4.8 The impact of the classification errors on the evaluation of KPIs

The computation of KPIs can be impacted by potential errors of contextual classification. Moreover, the contextual classification (for data selection before aggregation) is used not only for the calculation of Smart Data and KPIs, but also for the incidents' detection. The contextual classification is therefore the most basic and the most important at the beginning of data mining, enabling an accurate data selection or not.

To illustrate the impact of the classification errors on the evaluation of KPIs, the advanced Overall Equipment Effectiveness (OEE) (cf. section 4.2) is computed, the spindle #1 manufacturing database of machine-tool #2 is taken as an example. The advanced OEE during 4 hours in which the accuracy of contextual classification on *tool cutting or not* was evaluated (cf. Table 2.6, page 42) has been

calculated based on 3 different classes:

- The predicted classes based on the 'Old contextual classification', which was proposed in the thesis manuscript by Godreau [2017]: If $1 < P < 70\text{kW}$ and *spindle rotating at constant speed* and $Vf > 0$, then *tool is cutting*;
- The predicted classes based on 'New contextual classification', which is proposed in chapter 2: if $(P > P_{idle}$ or $Arms > T_{Arms})$ and $Vf > 0$ and *spindle rotating at constant speed*, except the subclass *spindle startup* and the subclass *post spindle startup*, the *tool is cutting material*;
- The Actual classes relying on the 'Ground truth' with 100% true label on *tool cutting or not* established in chapter 2.

The accuracy of the 'New contextual classification' by GMM during 4 hours is 99.23%, which was good enough (cf. Table 2.6, page 42). The accuracy of the 'Old contextual classification' during the same 4 hours can be estimated by comparing the predicted classes and the 'Actual classes'. The confusion matrix for this 'Old contextual classification' during the same 4 hours is listed in Table 4.3

Table 4.3 – Confusion matrix for 4 hours data from the machine-tool #2 based on 'Old contextual classification'

Confusion Matrix		Actual Classes	
		Tool Cutting	Tool Not Cutting
Predicted Classes	Tool Cutting	TP=51 338	FP=32 525
	Tool Not Cutting	FN=4 288	TN=55 849

The accuracy of this 'Old contextual classification' is $ACC_{\#2_Old}=74.44\%$, which is worse than the 'New contextual classification'.

Based on these 2 different predicted classes obtained by 2 different methods of classification as well as the 'Ground truth', the advanced OEE during this period (4 hours) can be evaluated. The period is divided into 3 contexts:

- Machine stopped: the machine-tool is stopped ($N=0$ or $Vf=0$),
- Motion without cutting: the machine-tool moves ($Vf \neq 0$) and the spindle rotates ($N \neq 0$) and *tool is not cutting*,
- Machine removing material: the machine-tool moves ($Vf \neq 0$) and the spindle rotates ($N \neq 0$) and *tool is cutting*.

Hence, the advanced OEE during this period is composed of 3 parts. The advanced OEE during the same 4 hours based on 2 different predicted classes and the 'Ground truth' are evaluated and listed in Figure 4.30.

In general, the proportion of each part obtained by the 'New contextual classification' is nearly the same as the 'Ground truth'. The tiny difference (38.63% for Ground truth compared to 38.22% for New when *machine removing material*) can be neglected. Then, the proportion of *machine stopped*

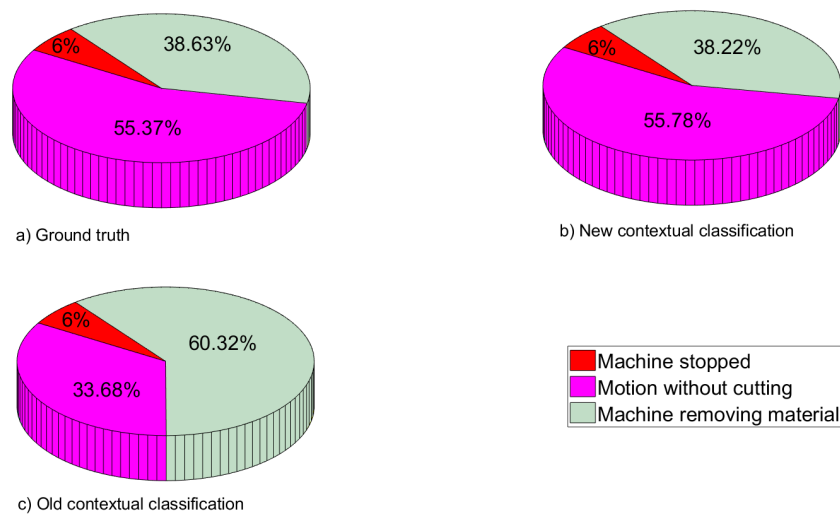


Figure 4.30 – Machine-tool OEE during 4 hours in the machine-tool #2 based on 3 different contextual classifications: a) Ground truth; and predicted classes obtained by b) New; c) Old contextual classification

(in red) is the same (6%) for 3 different classes. Because it was controlled by N or V_f equal to zero or not.

However, the proportion of *machine removing material* (in gray) based on the 'Old contextual classification' (60.32%) is much higher than the 'Ground truth' (38.63%). It is because the threshold of power P for *tool cutting or not* in 'Old contextual classification' (1kW) is too small. It involves that a lot of data are falsely considered as *tool is cutting*, but the *tool is not cutting* actually.

This comparison demonstrates that the 'New contextual classification' (by GMM) permits obtaining more accurate and precise KPIs. It proves that the contextual classification has a significant impact on the KPI assessment. Besides, the 'New contextual classification' is an essential improvement of the previous method.

4.9 Conclusion

This chapter presented the proposition of new Key Performance Indicators (KPIs) for decision aid, dedicated to answering the specific managerial needs of the different departments of a machining company. The data is successively aggregated to compute the KPIs. Some new aggregation operators were proposed, including the Abnormal Operator (AO, which recorded the mean value of a criterion when it differs from a target value); as well as the operator *Time* during which it differs. During the aggregation, the contextual classification is important for data selection. The 'New contextual classification' (chapter 2) has been compared with the 'Old contextual classification' from [Godreau, 2017] and with the 'Ground truth' in the computation of the advanced OEE. It has shown that a good contextual classification is a key point for the KPI assessment and has proved that 'New contextual classification' is a significant improvement of the method.

New sets of KPIs have been proposed for six scopes of analysis: production organization; workpiece quality; production monitoring; performances of tools; performances of programs and monitoring of machine-tool condition.

The contributions of this section are listed below:

- To help the Process planning for organizing the production, two KPIs: advanced OEE and occurrence of machined workpieces are proposed. These two KPIs are computed based on the 17 days' database in the machine tool #1.
- To help the Quality department ensure a good quality of the workpiece, a series of KPIs are proposed. They are divided into two parts: respect of cutting specifications and detection of incidents. The idea is to verify the respect of the cutting specifications (correct cutting conditions and sufficient cooling and feed) for each workpiece; and verify if there is any incidents (chatter or tool breakage) during production on machine-tool #1 data base. They revealed some interesting anomalies during the production. Then manual data mining confirmed and explained these anomalies.
- To help the Manufacturing department monitor the production, four KPIs are proposed: traceability of duration of machining operation; traceability of cutting aspects (mean cutting power P and duration of severe vibration); human factor during production (potentiometers not 100%) and incidents detected during production. These KPIs are calculated based on the machine-tool #4 and machine-tool #1 database.
- To help the Manufacturing evaluate the performances of tools, three KPIs are proposed: distribution of chatter per tool; distribution of excessive vibration per tool and tool use duration before tool breakage. They were all computed on machine-tool #2 data base. The 10 most problematic tools with chatter are revealed. The context information, like in which kind of operation and in which kind of tool engagement chatter occurred can also be deciphered. Moreover, the kind of chatter: punctual chatter or continuous chatter can also be gained according to the KPIs.
- To help the CAM programming to improve the programs, four KPIs are proposed: distribution of chatter per program; distribution of excessive vibration per program; productivity of CAM programs and slow idle motion. They were all evaluated based on the spindle #1 database on the machine-tool #2. The most problematic programs with chatter can be revealed. The contextual information can also be deciphered. The MRR can evaluate the productivity of CAM programs. The Vf during idle motion can be improved through adding more rapid motion into programs reasonably.
- To help the Maintenance department monitor the machine-tool condition, two KPIs are proposed: spindle condition monitoring and incidents detection. The chatter time and the occurrences of tool breakage as well as the spindle condition criterion BPFO were computed each day based on the spindle #1 database on the machine-tool #2.
- The impact of potential classification errors on the evaluation of KPI (advanced OEE) has been investigated. It has shown that a good contextual classification is a key point for the KPI

assessment and has proved that 'New contextual classification' is a significant improvement of the previous method.

Conclusions and Perspectives

Contents

5.1	Conclusions	104
5.2	Perspectives	107

Summary

This chapter summarizes the research work conducted in the framework of this PhD Thesis and draws the conclusions from the results obtained from the previous chapters. Some perspectives for future work will also be addressed.

5.1 Conclusions

The research work of this PhD thesis is carried out in the framework of the ANR SmartEmma Project whose objective is to develop innovative smart and connected machine-tools for High-Speed Machining. In this context, the devices Emmatools are installed on four machine-tools (three for aluminum alloy machining and one for titanium alloy machining) in the aeronautics industry. It collects a large amount of raw measurements during machining and then stores them in a database. The main objective of this PhD thesis is to extract and discover useful information from these big databases using data mining and machine learning techniques. Then useful information are processed by aggregation steps to be further transformed into comprehensive structures, such as KPIs, for targeted users. These KPIs not only improve machine-tool performance but also support the decision-making of operational management.

Firstly, a literature review concerning the state of the art in the corresponding domains of this research work is done in Chapter 1 in order to identify the scientific barriers as well as the specific requirements in the industrial manufacturing. They include: machining monitoring, data-mining and machine learning techniques, methods of defining KPIs and the existing decision-aid systems.

Taking into consideration all the existing technologies and the problematic of the targeted applications, the necessity of a contextual classification at the beginning stage of the data mining was highlighted. The main advantages of contextual classification is improving the efficiency of the classification system by applying business rules defined according to the nature of the applicable domain. In other words, the use of highly specific domain knowledge can reduce significantly the classification errors, improve the quality of KPI assessment, and reduce computational cost especially when dealing with big dataset.

Therefore, a contextual classification method was proposed for the machining data in the Chapter 2 to know the different states of the machine-tool and of the process at each instant. According to specific knowledge of machining, a business rule was defined for the contextual classification, which is combined with a unsupervised machine learning algorithm of Gaussian Mixture Model (GMM) to learn the detection thresholds. GMM is more suitable than K-means in that case because GMM classifies better the data with different sizes and different densities. The number of classes for GMM was determined by the Bayesian Information Criterion (BIC). The method, to determine if the *tool is cutting* and the type of machining operation, was validated on real data from four different machine-tools in the aerospace industry. Manual data mining and cross-validation by k-fold confirmed that the proposed method provides good results. Based on the contextual classification of material removing, the different machining operations and the different states of tool engagement can hence be detected at each instant.

After that, the business rules for incidents detection based on the contextual classification were proposed in Chapter 3 for the detection of chatter, tool breakage, and excessive vibrations. The operational context was deciphered when incidents occur, based on the contextual classification that explains the types of machining and tool engagement. The classification of harmonic / non-harmonic frequencies was significantly improved, enabling more accurate computation of Nh chatter criterion. Two business rules as well as an adaptive threshold, were proposed for tool breakage detection. Ub can detect all of the tool breakages including light tool breakages, but only when the tool is not cutting, involving a detection delay. Ub' can detect the large tool breakage instantaneously, but it can not

detect light tool breakage. A business rule was proposed for excessive vibration detection. All these proposed business rules were tested and validated with real data collected on the machine-tool #2.

The procedure of the data pre-processing, i.e., data cleaning, realized in this PhD thesis is also addressed in Chapter 3. Data cleaning is an essential step before data mining, whose objectives are to withdraw measurement outliers, to discard the missing values as well as to neutralize the high values, which are not indication of machining problems (chatter, tool breakage, excessive vibrations, or collision). Some improvements were proposed in this thesis (e.g. noise in vibration signals due to the lubrication startup is removed).

Meanwhile, new aggregation operators were proposed, including the Abnormal Operator (AO, which recorded the mean value of a criterion when it differs from a target value); as well as the operator *Time* during which it differs. These new aggregation operators were applied to the computation of KPIs, which prove to be highly efficient.

Finally, several new KPIs for aiding decision making were proposed in Chapter 4 based on the contextual classification combined with the detection of incidents. These new KPIs, such as advanced OEE; traceability of cutting aspects; respect of cutting specifications; detection of incidents; distribution of chatter per tool; MRR etc., are designed to answer to the specific managerial needs of the different departments of machining companies. These new sets of KPIs were applied to real industrial database and provided interesting results. Moreover, the impact of the classification errors on KPIs was evaluated for the advanced OEE by comparing the contextual classifications proposed in this PhD thesis, with a previous method proposed by [Godreau, 2017]. The results showed that the quality of classification results have a significant impact on KPI assessment. The proposed contextual classification in this PhD thesis provides more accurate KPIs than the previous methods.

The work presented in this thesis led to several scientific publications. They are listed below:

Articles in journals:

- **WANG Zhiqiang**, RITOU Mathieu, DA CUNHA Catherine & FURET Benoît. Contextual classification for smart machining based on unsupervised machine learning by Gaussian mixture model, *International Journal of Computer Integrated Manufacturing*, 2020, p. 1-13. <https://doi.org/10.1080/0951192X.2020.1775302>
- LADJ Asma, **WANG Zhiqiang**, MESKI Oussama, BELKADI Farouk, RITOU Mathieu and DA CUNHA Catherine. A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective. *Journal of Manufacturing Systems*, 2020. <https://doi.org/10.1016/j.jmsy.2020.07.018>

Proceeding in international conference:

- **Zhiqiang Wang**, Catherine Da Cunha, Mathieu Ritou, and Benoît Furet (2019). Comparison of k-means and GMM methods for contextual clustering in HSM. *Proceedings of the Changeable, Agile, Reconfigurable and Virtual Production 2018 (CARV2018)*, October 8-10, 2018, Nantes France. (Published in the journal *Procedia Manufacturing*, 28, 154-159). <https://doi.org/10.1016/j.promfg.2018.12.025>

Proceeding in national conference:

- **Wang, Z.**, Ritou, M., Da Cunha, C., & Furet, B. (2019, April). Classification contextuelle pour système d'aide à la décision pour machines-outils. In Colloque National S-mart/AIP-PRIMECA.
- **Wang, Z.**, Ritou, M., Da Cunha, C., & Furet, B. (2019). Classification contextuelle pour les fouilles de données issues de machines-outils. *24ème Congrès Français de Mécanique à Brest (CFM2019)*, August 26-30, 2019, Brest France.
- **Wang, Z.**, Ritou, M., Da Cunha, C., & Furet, B. (2021). Identification du contexte d'usinage basée sur le modèle de mélange gaussien : application au broutement. *Conférence Manufacturing 21*, June, 2021, Angers France. Submitted in March 14, 2020.

5.2 Perspectives

The work carried out during this thesis has led to concrete advances for industrial applications. They also provided ideas for extending the research or new perspectives to deepen the various themes discussed here.

Firstly, in the contextual classification, some labels with 100% true classes were established on a period of several hours to evaluate the accuracy of the prediction with the unsupervised machine learning by GMM. Next step, the supervised machine learning (e.g. ANN, SVM or Bayesian network) can be tested to predict future labels.

Secondly, in the contextual classification, the thresholds for *materials removal or not* (T_{Arms} and P_{idle}) may slightly evolve over time. It is because the machine condition will change (e.g. wear of the spindle). Then, some deep learning methods can be combined with the unsupervised machine learning (GMM) to conduct the contextual classification more accurately and automatically. For example, Long Short-Term Memory (LSTM) is an artificial Recurrent Neural Network (RNN) architecture used in the field of deep learning [Hochreiter and Schmidhuber, 1997]. LSTM networks are well-suited to classifying, processing and making predictions based on time series data.

Thirdly, it is important to realize the automation of the link between machining simulation (e.g. NC-simul) and the decision-aid system. Systematic synchronisation between NC-simul and EMMAtools is important in order to benefit all information in real-time. This requires that each work-piece should be simulated with NC-simul before real production, that the simulation software should be installed on the machine-tools where EMMAtools are present, and that the communication protocol between the two software programs is automated. This will allow an increase in the richness of the databases on larger time scales allowing the capitalization of new information.

Fourthly, extended work on the proposition and the generation of new KPIs will also be an important task to be continued. There is still a lot of unprocessed information (e.g. temperature) in the databases. Their processing will allow us to go further in the use of the acquired data for decision-aid system.

Finally, it is also interesting to apply the contextual classification in other industries, different from the aeronautic or the machining. In the general context of industry 4.0, large volumes of manufacturing data are available on the instrumented factory. It is necessary to conduct the contextual classification before data mining. To extract data from databases more efficiently and accurately. The contextual classifications will be based on other criteria and other business rules depending on the industry knowledge. This will allow to understand the status of each manufacturing equipment and to realize a more relevant calculation of KPIs.

References

- Eberhard Abele and Dennis Korff. Avoidance of collision-caused spindle damages - challenges, methods and solutions for high dynamic machine tools. *CIRP annals*, 60(1):425–428, 2011.
- Jose Vicente Abellan-Nebot and Fernando Romero Subirón. A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology*, 47(1-4):237–257, 2010.
- Faleh A Al-Sulaiman, M Abdul Baseer, and Anwar K Sheikh. Use of electrical power for online monitoring of tool condition. *Journal of Materials Processing Technology*, 166(3):364–371, 2005.
- Y Altıntaş and Erhan Budak. Analytical prediction of stability lobes in milling. *CIRP annals*, 44(1):357–362, 1995.
- Y Altintas and M Weck. Chatter stability of metal cutting and grinding. *CIRP annals*, 53(2):619–642, 2004.
- Amaury B Andre, Eduardo Beltrame, and Jacques Wainer. A combination of support vector machine and k-nearest neighbors for machine fault detection. *Applied Artificial Intelligence*, 27(1):36–49, 2013.
- Nabil Anwer, Paul J Scott, and Vijay Srinivasan. Toward a classification of partitioning operations for standardization of geometrical product specifications and verification. *Procedia CIRP*, 75:325–330, 2018.
- Ana Isabel Rojão Lourenço Azevedo and Manuel Filipe Santos. Kdd, semma and crisp-dm: a parallel overview. *IADS-DM*, 2008.
- Alain Bernard and Nicolas Perry. Fundamental concepts of product/technology/process informational integration for process modelling and process planning. *International Journal of Computer Integrated Manufacturing*, 16(7-8):557–565, 2003.
- P Bhattacharyya, D Sengupta, and S Mukhopadhyay. Cutting force-based real-time estimation of tool wear in face milling using a combination of signal processing techniques. *Mechanical Systems and Signal Processing*, 21(6):2665–2683, 2007.
- Raunak Bhinge, Nishant Biswas, David Dornfeld, Jinkyoo Park, Kincho H Law, Moneer Helu, and Sudarsan Rachuri. An intelligent machine monitoring system for energy prediction using a gaussian process regression. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 978–986. IEEE, 2014.
- Sultan Binsaeid, Shihab Asfour, Sohyung Cho, and Arzu Onar. Machine ensemble approach for simultaneous detection of transient and gradual abnormalities in end milling using multisensor fusion. *Journal of Materials Processing Technology*, 209(10):4728–4738, 2009.
- Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

- Alessandra Caggiano. Cloud-based manufacturing process monitoring for smart diagnosis services. *International Journal of Computer Integrated Manufacturing*, 31(7):612–623, 2018.
- Bruno Castanier, Antoine Grall, and Christophe Béranger. A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering & System Safety*, 87(1):109–120, 2005.
- Mu-Chen Chen. Configuration of cellular manufacturing systems using association rule induction. *International Journal of Production Research*, 41(2):381–395, 2003.
- Xingzheng Chen, Congbo Li, Ying Tang, and Qinge Xiao. An internet of things based energy efficiency monitoring and management system for machining workshop. *Journal of cleaner production*, 199: 957–968, 2018.
- Vladimir Cherkassky and Filip M Mulier. *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.
- Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.
- Côme De Castelbajac, Mathieu Ritou, S Laporte, and Benoît Furet. Monitoring of distributed defects on hsm spindle bearings. *Applied acoustics*, 77:159–168, 2014.
- Story Dey and JA Stori. A bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools and Manufacture*, 45(1):75–91, 2005.
- Hazim El-Mounayri and Haiyan Deng. A generic and innovative approach for integrated simulation and optimisation of end milling using solid modelling and neural network. *International Journal of Computer Integrated Manufacturing*, 23(1):40–60, 2010.
- TI El-Wardany, D Gao, and MA Elbestawi. Tool condition monitoring in drilling using vibration signature analysis. *International Journal of Machine Tools and Manufacture*, 36(6):687–711, 1996.
- Soner Emec, Jörg Krüger, and Günther Seliger. Online fault-monitoring in machine tools based on energy consumption analysis and non-invasive data acquisition for improved resource-efficiency. *Procedia CIRP*, 40:236–243, 2016.
- Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37, 1996.
- Guodong Feng, Chunyan Lai, and Narayan C Kar. Expectation-maximization particle-filter-and kalman-filter-based permanent magnet temperature estimation for pmsm condition monitoring using high-frequency signal injection. *IEEE Transactions on Industrial Informatics*, 13(3):1261–1270, 2016.
- Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.
- Robert Gao, Lihui Wang, Roberto Teti, David Dornfeld, Soundar Kumara, Masahiko Mori, and Moneer Helu. Cloud-enabled prognosis for manufacturing. *CIRP annals*, 64(2):749–772, 2015.

- Victor Godreau. *Extraction des connaissances à partir des données de la surveillance de l'usinage*. PhD thesis, Nantes, 2017.
- Victor Godreau, Mathieu Ritou, Etienne Chové, Benoit Furet, and Didier Dumur. Continuous improvement of hsm process by data mining. *Journal of Intelligent Manufacturing*, 30(7):2781–2788, 2019.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Norden Eh Huang. *Hilbert-Huang transform and its applications*, volume 16. World Scientific, 2014.
- Yonghong Huang, Kevin B Englehart, Bernard Hudgins, and Adrian DC Chan. A gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Transactions on Biomedical Engineering*, 52(11):1801–1811, 2005.
- DIn ISO 10816. 10816-3: 2009 vibrations mécaniques - Évaluation des vibrations des machines par mesurages sur les parties non tournantes - partie 3: Machines industrielles de puissance nominale supérieure à 15 kw et de vitesse nominale de fonctionnement entre 120 r/min et 15 000 r/min, lorsqu'elles sont mesurées in situ (iso 10816-3: 2009). *V.*, Berlin, 2009.
- Julien Jacques. Fouille de données. *Data mining Université Lumière Lyon*, 2:12, 2001.
- Olivier Janssens, Viktor Slavkovikj, Bram Vervisch, Kurt Stockman, Mia Loccufier, Steven Verstockt, Rik Van de Walle, and Sofie Van Hoecke. Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration*, 377:331–345, 2016.
- SV Kamarthi and S Pittner. Fourier and wavelet transform for flank wear estimation - a comparison. *Mechanical Systems and Signal Processing*, 11(6):791–809, 1997.
- Bernard Kamsu-Foguem, Fabien Rigal, and Félix Mauget. Mining association rules for the quality improvement of the production process. *Expert systems with applications*, 40(4):1034–1045, 2013.
- Deogratias Kibira, Michael P Brundage, Shaw Feng, and KC Morris. Procedure for selecting key performance indicators for sustainable manufacturing. *Journal of Manufacturing Science and Engineering*, 140(1), 2018.
- Dong-Hyeon Kim, Thomas JY Kim, Xinlin Wang, Mincheol Kim, Ying-Jun Quan, Jin Woo Oh, Soo-Hong Min, Hyungjung Kim, Binayak Bhandari, Insoon Yang, et al. Smart machining process using machine learning: A review and perspective on machining industry. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 5(4):555–568, 2018.
- Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- Gülser Köksal, İnci Batmaz, and Murat Caner Testik. A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with Applications*, 38(10):13448–13467, 2011.
- K Krishna and M Narasimha Murty. Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3):433–439, 1999.

- P Krishnakumar, K Rameshkumar, and KI Ramachandran. Tool wear condition prediction using vibration signals in high speed machining (hsm) of titanium (ti-6al-4 v) alloy. *Procedia Computer Science*, 50:270–275, 2015.
- Asma Ladj, Zhiqiang Wang, Oussama Meski, Farouk Belkadi, Mathieu Ritou, and Catherine Da Cunha. A knowledge-based digital shadow for machining industry in a digital twin perspective. *Journal of Manufacturing Systems*, 2020.
- Julien Le Duigou, Alain Bernard, Nicolas Perry, and Jean-Charles Delplace. Generic plm system for smes: Application to an equipment manufacturer. *International Journal of Product Lifecycle Management* 7, 6(1):51–64, 2012.
- WK Lee, MM Ratnam, and ZA Ahmad. Detection of chipping in ceramic cutting inserts from work-piece profile during turning using fast fourier transform (fft) and continuous wavelet transform (cwt). *Precision Engineering*, 47:406–423, 2017.
- Pirmin Lemberger, Marc Batty, Médéric Morel, and Jean-Luc Raffaëlli. *Big Data et Machine Learning-2e éd.* Dunod, 2016.
- Juergen Lenz, Thorsten Wuest, and Engelbert Westkämper. Holistic approach to machine tool data analytics. *Journal of manufacturing systems*, 48:180–191, 2018.
- Qing Li, Hongzhen Jiang, Qianlin Tang, Yaotang Chen, Jun Li, and Jian Zhou. Smart manufacturing standardization: reference model and standards framework. In *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, pages 16–25. Springer, 2016.
- Xiaoli Li. A brief review: acoustic emission method for tool wear monitoring during turning. *International Journal of Machine Tools and Manufacture*, 42(2):157–165, 2002.
- Steven Y Liang, Rogelio L Hecker, and Robert G Landers. Machining process monitoring and control: the state-of-the-art. *J. Manuf. Sci. Eng.*, 126(2):297–310, 2004.
- Chao Liu, Hrishikesh Vengayil, Ray Y Zhong, and Xun Xu. A systematic development method for cyber-physical machine tools. *Journal of manufacturing systems*, 48:13–24, 2018.
- Yang Lu. Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 6:1–10, 2017.
- Atsushi Matsubara and Soichi Ibaraki. Monitoring and control of cutting forces in machining processes: a review. *International Journal of Automation Technology*, 3(4):445–456, 2009.
- Cathy Maugis, Gilles Celeux, and Marie-Laure Martin-Magniette. Variable selection for clustering with gaussian mixture models. *Biometrics*, 65(3):701–709, 2009.
- Jeff Morgan and Garret E O’Donnell. Cyber physical process monitoring systems. *Journal of Intelligent Manufacturing*, 29(6):1317–1328, 2018.
- David Parmenter. *Key performance indicators: developing, implementing, and using winning KPIs.* John Wiley & Sons, 2015.

- E García Plaza and PJ Núñez López. Surface roughness monitoring by singular spectrum analysis of vibration signals. *Mechanical Systems and Signal Processing*, 84:516–530, 2017.
- Guillem Quintana and Joaquim Ciurana. Chatter in machining processes: A review. *International Journal of Machine Tools and Manufacture*, 51(5):363–376, 2011.
- Mustafizur Rahman, Zhi-Gang Wang, and Yoke-San Wong. A review on high-speed machining of titanium alloys. *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, 49(1):11–20, 2006.
- Robert B. Randall and Jerome Antoni. Rolling element bearing diagnostics - a tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.
- Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741, 2009.
- Keith Ridgway, Chris W Clegg, DJ Williams, P Hourd, M Robinson, L Bolton, K Cichomska, and J Baldwin. The factory of the future. *Government Office for Science, Evidence Paper*, 29, 2013.
- Mathieu Ritou, Clément Rabréau, Sébastien Le Loch, Benoit Furet, and Didier Dumur. Influence of spindle condition on the dynamic behavior. *CIRP Annals*, 67(1):419–422, 2018.
- Mathieu Ritou, Farouk Belkadi, Zakaria Yahouni, Catherine Da Cunha, Florent Laroche, and Benoit Furet. Knowledge-based multi-level aggregation for decision aid in the machining industry. *CIRP Annals*, 68(1):475–478, 2019.
- Juan J Rodríguez, Guillem Quintana, Andres Bustillo, and Joaquim Ciurana. A decision-making tool based on decision trees for roughness prediction in face milling. *International Journal of Computer Integrated Manufacturing*, 30(9):943–957, 2017.
- Lior Rokach and Oded Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.
- Thomas Rudolf, Christian Brecher, and Frank Possel-Dölken. Contact-based collision detection - a new approach to avoid hard collisions in machine tools. In *International Conference on Smart Machining Systems*, pages 1–4, 2007.
- DR Salgado and FJ Alonso. Tool wear detection in turning operations using singular spectrum analysis. *Journal of Materials Processing Technology*, 171(3):451–458, 2006.
- Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.
- PY Sevilla-Camacho, JB Robles-Ocampo, J Muñoz-Soria, and F Lee-Orantes. Tool failure detection method for high-speed milling using vibration signal and reconfigurable bandpass digital filtering. *The International Journal of Advanced Manufacturing Technology*, 81(5-8):1187–1194, 2015.
- Bernhard Sick. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mechanical systems and signal processing*, 16(4):487–546, 2002.

- Stephen V Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1):77–89, 1997.
- Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9-12):3563–3576, 2018.
- Frederick Winslow Taylor. *On the Art of Cutting Metals...*, volume 23. American society of mechanical engineers, 1906.
- Roberto Teti, Krzysztof Jemielniak, Garret O’Donnell, and David Dornfeld. Advanced monitoring of machining operations. *CIRP annals*, 59(2):717–739, 2010.
- Jinjiang Wang, Yulin Ma, Laibin Zhang, Robert X Gao, and Dazhong Wu. Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48:144–156, 2018.
- Lidong Wang and Guanghui Wang. Big data in cyber-physical systems, digital manufacturing and industry 4.0. *International Journal of Engineering and Manufacturing (IJEM)*, 6(4):1–8, 2016.
- Zhiqiang Wang, Catherine Da Cunha, Mathieu Ritou, and Benoît Furet. Comparison of k-means and gmm methods for contextual clustering in hsm. *Procedia Manufacturing*, 28:154–159, 2019.
- Zhiqiang Wang, Mathieu Ritou, Catherine Da Cunha, and Benoît Furet. Contextual classification for smart machining based on unsupervised machine learning by gaussian mixture model. *International Journal of Computer Integrated Manufacturing*, pages 1–13, 2020.
- Marco Wiering and Martijn Van Otterlo. *Reinforcement learning*, volume 12. Springer, 2012.
- Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Springer-Verlag London, UK, 2000.
- David H Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.
- Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, S Yu Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.
- Kanglin Xing, JRR Mayer, and Sofiane Achiche. Machine tool volumetric error features extraction and classification using principal component analysis and k-means. *Journal of Manufacturing and Materials Processing*, 2(3):60, 2018.
- Jihong Yan and Lin Li. Multi-objective optimization of milling parameters—the trade-offs between energy, production rate and cutting quality. *Journal of Cleaner Production*, 52:462–471, 2013.
- Chen Zhang and Haiyan Zhang. Modelling and prediction of tool wear using ls-svm in milling operation. *International Journal of Computer Integrated Manufacturing*, 29(1):76–91, 2016.
- Yuqing Zhou and Wei Xue. Review of tool condition monitoring methods in milling processes. *The International Journal of Advanced Manufacturing Technology*, 96(5-8):2509–2523, 2018.

-
- Kunpeng Zhu, Yoke San Wong, and Geok Soon Hong. Multi-category micro-milling tool wear monitoring with continuous hidden markov models. *Mechanical Systems and Signal Processing*, 23(2): 547–560, 2009.

Titre : Aide à la décision en usinage basée sur des règles métier et apprentissages non supervisés

Mots clés : Industrie 4.0, machines-outils, fouilles de données, GMM, classification contextuelle, KPIs

Résumé : Dans le contexte général de l'industrie 4.0, une entreprise de fabrication moderne dispose de nombreuses données numériques qui pourraient être utilisées pour rendre les machines-outils plus intelligentes et faciliter la prise de décision en matière de gestion opérationnelle. Cette thèse vise à proposer un système d'aide à la décision pour les machines-outils intelligentes et connectées par l'exploration des données.

L'une des premières étapes de l'approche d'exploration de données est la sélection précise de données pertinentes. Pour ce faire, les données brutes doivent être classées dans différents groupes de contexte. Cette thèse propose un algorithme d'apprentissage automatique non-supervisé, par mélanges gaussiens (GMM), pour la classification contextuelle.

Basé sur les informations contextuelles, différents incidents d'usinage peuvent être détectés en temps réel. Il s'agit notamment de broutement, de bris d'outil et de sur-vibration. Cette thèse présente un ensemble de règles métiers pour la détection du broutement, de bris d'outil et de sur-vibration. Le contexte opérationnel a été décrypté lorsque des incidents se produisent, sur la base de la classification contextuelle qui explique les types d'usinage et d'engagement des outils.

Ensuite, les nouveaux indicateurs clés de performance (KPIs) pertinents et appropriés peuvent être proposés sur la base de ces informations contextuelles combinés avec les incidents détectés afin de soutenir la prise de décision pour la gestion opérationnelle. Cette thèse propose de nouveaux opérateurs d'agrégation et de nouveaux KPIs pertinents pour les différents besoins des départements.

Title : Decision-aid in machining based on business rules and unsupervised machine learning

Keywords : Industry 4.0, machine-tools, Data mining, GMM, contextual classification, KPIs.

Abstract : In the general context of Industry 4.0, large volumes of manufacturing data are available on instrumented machine-tools. They are interesting to exploit not only to improve machine-tool performances but also to support the decision making for the operational management. This thesis aims at proposing a decision-aid system for intelligent and connected machine-tools through Data mining.

The first step in a data mining approach is the selection of relevant data. Raw data must, therefore, be classified into different groups of contexts. This thesis proposes a contextual classification procedure in machining based on unsupervised machine learning by Gaussian mixture model.

Based on this contextual classification information, different machining incidents can be detected in real-time. They include chatter, tool breakage and excessive vibration. This thesis introduces a set of business rules for incidents detection. The operational context has been deciphering when incidents occur, based on the contextual classification that explains the types of machining and tool engagement.

Then, the nouveaux relevant and appropriate Key Performance Indicators (KPIs) can be proposed based on these contextual information and the incidents detected to support decision making for the operational management.