ConfiLog: An approach to Guide Product Line Configuration Processes based on Configuration Traces

Houssem Chemingui, Ines Gam, Raul Mazo, Camille Salinesi, Henda Ben Ghézala

ConfiLog: An approach to Guide Product Line Configuration Processes based on Configuration Traces

Houssem Chemingui, Ines Gam, Raúl Mazo, Camille Salinesi, Henda Ben Ghezala

Centre de Recherche en Informatique, Paris 1 Panthéon Sorbonne University, Paris, France
RIADI Laboratory, National School for Computer Studies, Manouba, Tunisia
Grupo de investigación GIDITIC, Universidad Eafit, Medellín, Colombia
Lab-STICC, ENSTA Bretagne, Brest, France

Abstract— [Context] Product line engineering is a conception and production paradigm in which the purpose is no longer to develop a single product but to develop a collection of products. They present a strategy that provides to the organizations a more competitive edge by improving productivity and decreasing costs. However, these benefits can be missed if users find difficulties when they are configuring a product from the line. This situation is widespread mainly when they deal with large product lines that contain a plethora of characteristics. **[Problem]** In these cases, the configuration process becomes monotonous and an error prone task. Furthermore, it is crucial to guide stakeholders by providing assistance during such complex process. Consequently, the configuration process envisages recommending the best configuration alternatives to users until leading them to a satisfying experience. **[Contribution]** Our proposal aims to enhance future configuration processes and maximize the user satisfaction based on previous executions. Therefore, we mine and analyze previous product line configurations by logging user's actions. User guidance is orchestrated by a goal-question-metric approach. **[Results]** Our research has shown that guidance solutions can be proposed according to the trace mining. Integrating these solutions throughout the configuration process can assist potential users that they are configuring complex lines. The usability of analyzing traces in order to guide future configurations is demonstrated using an e-shop case study from the software industry.

Keywords— Product line engineering, configuration process, assistance, recommending, user's actions, goal-question-metric, trace mining.

I. INTRODUCTION

Mass customization is a phenomenon in vogue and widely adopted by companies. Major industry clusters work hard to develop personalized offers that meet their customers' distinct preferences [Da Silveira et al., 2001]. At the same time, they work to improve the efficiency of their production processes. Indeed, product lines (PLs) have been a strategy that corresponds to these needs. PLs have become a promising approach for the development of configurable systems that are built from a set of reusable components. PLs aim to explore the similarities and variabilities in a collection of products that share a specific domain [Northrop, 2002]. In practice, the variability and commonality between products are captured by a Product Line Model (PLM) like a Feature Model (FM) [Kang et al., 1990], an Orthogonal Variability Model (OVM) [Pohl et al., 2005] or a Dopler model [Dhungana et al., 2010]. These modelling languages, such as the FM used in this paper allow distinguishing between variable products that is possible to configure from the line. Some PLM features define points of variation and their role is to permit the instantiation of different products by selecting specific PLs functionalities. A fundamental task in PLs is the process of selecting and/or deselecting features from the FM in order to construct a new product configuration. Usually, when the number of features increase in a PLM, the number of potential products increases also. Thus, industrial-sized PLMs with hundreds or thousands of features can make the configuration process impractical. For example, product lines in the automotive industry can hold up to 10^{21} possible configurations [Astesana et al., 2010] that involve multiple stakeholders [Mendonca et al., 2008]. Indeed, the configuration process becomes quickly complex when increases the number of decisions to be made, the fact that these decisions must be taken in a predefined order that respects user preferences on the one hand and the model constraints on the other hand.

Assistance and automatic propagation of some choices are needed to improve the configuration correctness and quality when dealing with many possible combinations.

Thereby, our purpose is to reason on complete or incomplete users' traces of configurations recorded in logs in order to have an idea of : (i) *Which informations can be useful to guide the PLs configuration?* (ii) *How to make use of previous configurations traces and interpret knowledge from it.* (iii) *How to use this knowledge into the future configuration processes?* (iv) According to traces, *what guidelines can be used to assist the configuration?* Several analyses have to be performed in order to answer to the previous questions. In order to do so, we configure PLMs, we import the configuration logs and analyze them by process mining techniques [Aalst, 2011].

The approach presented in this paper uses process mining techniques applied to the logs captured from user's' configuration actions. The primary contribution of this research is to assist users during the product configuration by providing (according to previous traces) them an easier and personalized configuration. Answering to this issue, industries dealing with PLs can improve the configuration experience of their configuration tools and improving the satisfaction of who is configuring new products. Mining traces and analysing then after allow us to capture the occurrence statistics and the timestamp measurement of user activities and their effects which allow us to develop some guidance metrics to improve the configuration process.

However, the identification of metrics it is not a simple task since too many metrics

can be identified for different goals. This problem tends to overwhelm process designers in their interpretation and analysis. Therefore, it becomes necessary to investigate methods of identifying the pertinent metrics that we need to assess and improve our processes. In software engineering, many frameworks were proposed to address this issue. Examples of those are: the Goal Question Metric (GQM) approach [Basili, 2005], Personal Measurement Software (PSM) [Card, 2003-a p738], and the framework developed by [Fenton and Pfleeger, 1997] in Software Metrics a Rigorous and Practical Approach. In this research, we are interested in the GQM framework.

The remainder of this work is organized as follows: Section II surveys a pertinent literature related to PL configurations and Process mining. It elucidates also how it possible to mine PL traces. Section III outlines the challenges of process mining in the context of PL configuration and proposes our GQM guidance approach. Section IV presents preliminary results through our running example. Section V presents related works while the concluding section discusses research perspectives.

II. Theoretical basis and ConfiLog positioning

In this section, we introduce, first, the PL paradigm and its concepts. Then, we highlight the main goals and concepts of process mining. In subsection C, we describe related recommendations methods related to our subject. And finally, we propose an overview of our approach describing the main lines that we handle with in this research.

A. Product line concepts and interests

A PL is "a set of systems that share a common, managed set of features satisfying the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way"[1]. The realization of a PL is based on two distinct processes: domain engineering and application engineering. The domain engineering process is designed to capitalize and document all artefacts (ie. code, software components, libraries, etc.), also called assets, necessary to produce software systems in a clearly identified domain (e.g., enterprise management software and mobile phone operating system). These artifacts are organized in such a way as to determine which are indispensable, interchangeable or dependent on others, within the same architecture. The application engineering process involves reusing existing artifacts to produce a new product that meets specific needs. The process of choosing artifacts to reuse is also called the configuration of a product. A configuration represents a product of the PL, completely or partially defined, and it contains all the elements of the PL that are selected according to the requirements of the users and the constraints prescribed during the specification and modeling of the corresponding PL. A configuration can be considered also as a tuple (V, D, C) where V = {v1, v2, ..., vn} represents the set of finite domain variables of the corresponding PLM, D = {dom(v1), dom(v2), ..., dom(vn)} represents the set of corresponding domains of each variable, and C represents a set of configuration constraints.

B. Process mining concepts

[1] http://www.sei.cmu.edu/. SEI web page

Process mining is a set of techniques that enable to discover, monitor and improve processes through the extraction of information from event logs available in information systems of today [Aalst, 2011]. Process mining was defined by Günther [Günther, 2009] as an area of research that is concerned with posteriori analysis of business processes, based on event logs of their execution. Its purpose is to extract global information and highlighting the different aspects of the process from these event logs. Hence, the process mining is adopted by some organizations that do not have enough informations about what is happening in their activities. Figure 1 recapitulates the general principle of process mining. The starting point for process mining is an information system that records executions traces of processes and store them in event logs. An event log can be seen as a collection of cases and a case can be seen as a sequence of events [Aalst 2011]. Each event contain informations about its execution like the timestamp, the activity that involve, the originator of this event, etc. According to the event logs, process models can be generated with differents concerns and techniques. during the application of process mining techniques. There are three fundamental techniques of process mining [Aalst, 2011]: (i) To mine process models (Discovery); (ii) To identify, explain and quantify process differences when existing process models are compared with the discovered ones (Conformance); and (iii) to apply changes to the existing process models based on the discovered ones (Enhancement). Some techniques aim to go further and use traces to suggest which tasks may follow a current activity [Mobasher et al., 2000], [Schonenberg et al., 2008]. Based on traces and generated models, predictions and recommendations can be introduced on ongoing and potential processes.

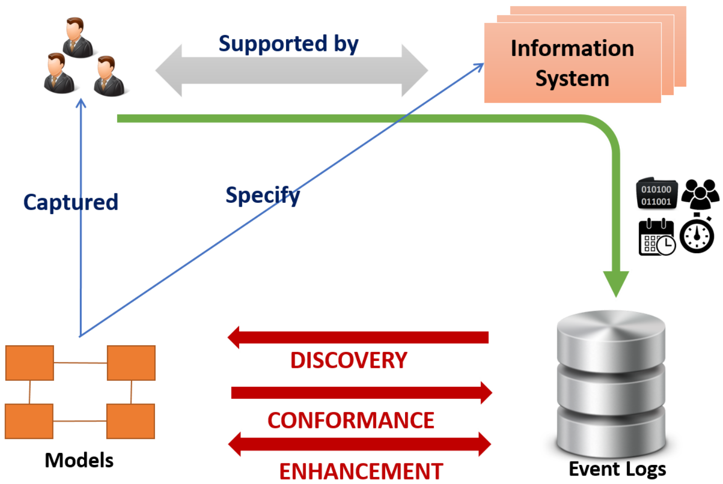

Figure 1- Process mining overview [Aalst, 2011]

C. Guidance based on hybrid recommendations

Different types of recommendations exist in the literature: the collaborative recommendation [Resnick et al., 1994], the content-based recommendation [Pazzani & Billsus, 1997] and the knowledge-based recommendation [Burke, 2000]. Other research [Rabiser, 2011] has demonstrated that a hybrid approach, using more than

one recommendation type could be more effective. In our research we focus on this kind of recommendation.

The collaborative recommendation predicts the usefulness of a product for a particular user based on evaluations of this product made by other users. To recommend a product for a specific user, the system has to calculate the utility of the product estimated by the users who are similar to him. Therefore, to recommend configurations to a user, we have to look for users with similar choices for configurations. The most appreciated configurations by these users will be recommended to the user. The recommendation based on the content makes possible to recommend products similar to those that the user has already appreciated. It allows to predict the utility of a product based on the user-assigned utilities to similar products [Adomavicius & Tuzhilin, 2005]. This similarity is calculated from the product descriptions and the user profile in terms of the characteristics. For each characteristic a weight is associated indicating its degree of importance. Thus, a weight vector is associated to each product. The user profile can be constructed from previous evaluations or actions. In our case, the content-based recommendation is reflected by the fact that we analyze the user actions in order to recommend to him similar alternatives to his preferences. A third type of recommendation is based on knowledge about characteristics, user preferences, and the context. The knowledge-based recommendation rely on a feedback loop in which the user requirements and preferences are gathered. When the user preferences are difficult to identify or gathered at once then the feedback loop could be employed multiple times.

In our research, we are interested in the hybrid recommendation that combines two or more types of recommendations to provide better performance. Several researches such as [Adomavicius & Tuzhilin, 2005] evaluated the performance of the hybrid with the pure collaborative, content-based and knowledge-based methods. Then, they demonstrated that the hybrid methods can provide more accurate recommendations than pure approaches. Generally, Hybrid recommendations are used to overcome some problems if we use a unique method of recommendation [Burke, 2002]. The role of process mining, is to support the decision making by enabling process guidance. Moreover, process mining is interesting in all recommendation methods as it can reveal and visualize the relationship systems, contents and users. Nonetheless, the reasoning processes for developing a recommender strategy is a separate research objective to our current one.

D. When process mining meets product line configurations?

To apply process mining techniques in the context of PL configuration, we propose, ConfiLog: an approach that aims to provide assistance during the configuration of products. Figure 2, shows an overview of this approach that consists of two main

compartments notably a front office process and a back office process. In the front office, a set of choices is made by users during the PL configuration through the tool interface. Before configuring a product, generally users have always a full or a part of the desired features of the product to configure (*“Product Model”* Ⓐ*)* but their are constrained by a *“PL model”* that they have to respect. The back office is a set of processes when happen guiding directives. From the front office, all the users’ traces are captured, extracted and recorded in event logs. Throughout this process, ETL (Extraction, Transformation, and Load) mechanisms cooperate in order to treat configuration traces made by users, transform their structure, load logs and filter pertinent data according to intentions [Rodic et al., 2009]. A continuous learning is planned inside the *“Logs Data Collector”* Ⓑ that take as input, a collection of unfiltered logs resulting from configuration processes. This process aims to save relevant logs which will serve to precise guidance issues.

Mining past PL configuration processes is mainly based on capitalized event logs resulting from Ⓑ. Throughout process mining treatments, a definition of mining criteria and queries is important. Monitoring flows and relationships is also commendable. The choice of process mining algorithms depends on the desired intentions and ends. Transformed traces will be considered as parameters of the process mining algorithms. The outputs of these algorithms are *“Process models”* Ⓒ that produce a repository used in order to interpret models and provide guidance answers. Given answers by process mining, are injected in a “*Goal question metric”* [Basili et al., 1994] repository Ⓓ (cf. section III.A) that capitalize answers for specific guidance questions using specific metrics. Take into account Ⓐ and Ⓓ guidance indicators are provided to make the configuration task more interesting, flexible and guided.

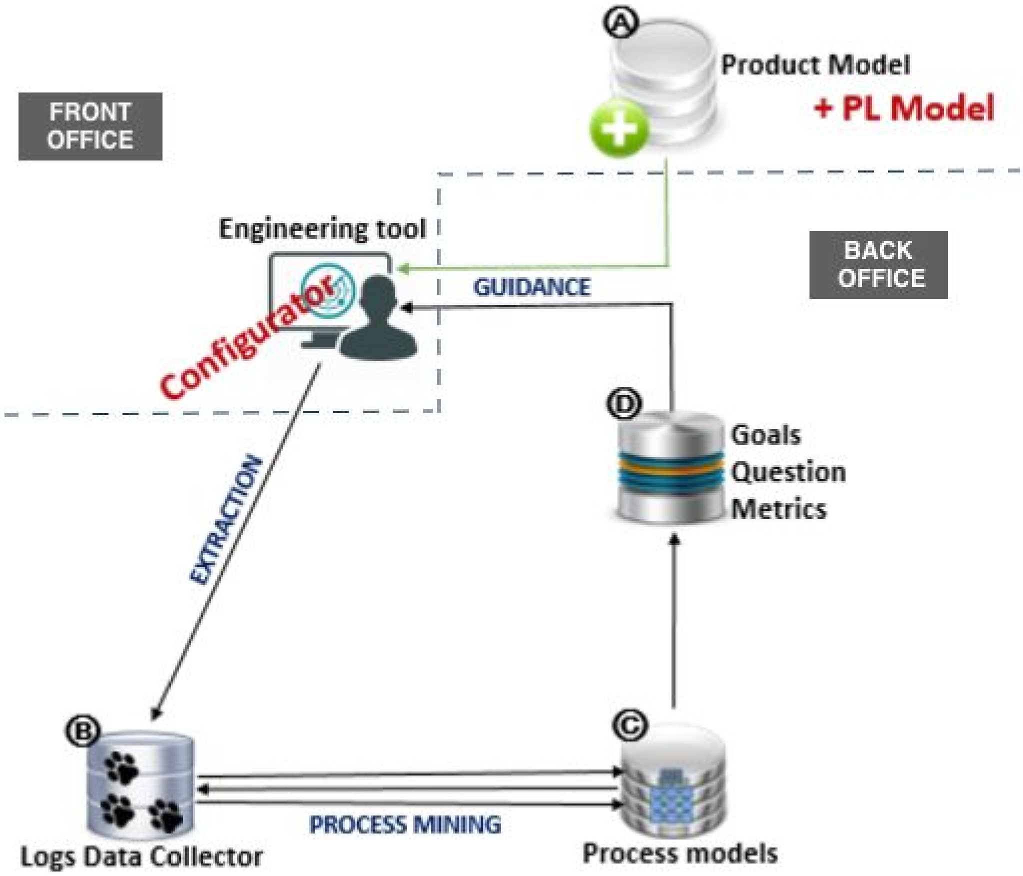

Figure 2. Overview of the ConfiLog approach

We are convinced that this approach is conveyed and portable regardless of the PL configuration context. In this paper we target to provide guidance during the configuration of PLs. Therefore, the tool engineering is considered as a PL configurator and the users are constrained by a PL model. However, the portability and the genericity of this solution seems promising. This approach can be applied to guide complex business processes of different contexts. Based on mining traces of users and correlations between these traces, it is possible to generate process models. Then, using these process models, the guidance will be monitored through the goal-question-metric method to provide specific answers for specific guidance goals.

III. The ConfiLog Approach underpinnings

This section presents the ConfiLog approach to guide customers during the configuration of the PLM according to the traces of previous configurations. The first part of this section introduces the ConfiLog framework. The second part illustrates the process of ConfiLog and the third part shows how to use the proposed approach according to a running example.

A. The ConfiLog *framework*

We propose the following framework to offer a solution for assisting users during a PL configuration based on previous traces. In this sense, we focus on the identification of the relevant data for logging, extracting and mining. In the PL configuration context, we decided to structure the event logs as in Table 1.

Table 1. Definition of the event logs structure

Attribute	Description
IDConfig	A1: The configuration process unique identifier
Originator	A2: The user identifier
IDFeature	A3: The configured feature unique identifier
FeatureName	A4: The configured feature name
FeatureType	A5: The type of the configured feature (Template model, mandatory, grouped or optional)
FeatureValue	A6: The value that the user assign for the given feature? (1 for selected feature, 0 for rejected feature)
FeatureDecisionType	A7: The type of configuration decision (manual or propagated: Automated follows other choices)
FeatureDecisionStep	A8: The order of the decision in the configuration
Timestamp	A9 : The date and time of the decision

We introduce solutions to express where and how guidance strategies can be developed to face configuration difficulties. To highlight our assistance goals, we need to make sure that we make the right answers to the right questions using the right metrics. The GQM (Goal Question Metric) method is one simple way of making sure that the metrics we collect is closely tied to the goals. The GQM was proposed as a framework for defining and interpreting software measurement. It was expanded later with many other concepts and become a result of many years of practical experience and academic search. Figure 3 shows the hierarchy levels of the GQM model: goals, questions and metrics. The upper level presents goals (conceptual level) that have to be defined. Each goal has to be operationalized by generating questions that help to understand how to meet it. Questions (the operational level) try to characterize the object of measurement in a context of a quantified issue. The lower level presents the metrics (the quantitative level) that need to be collected to answer the questions. The some metrics can be used to answer more than one question.

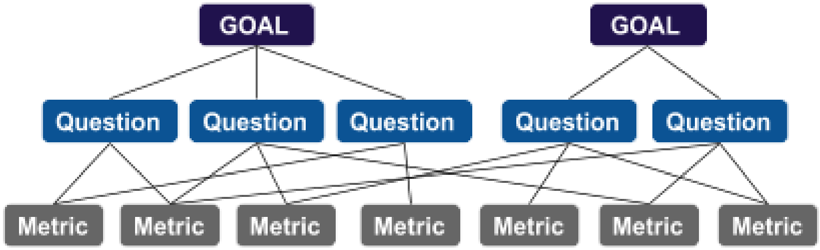

Figure 3- GQM Model structure [Basili et al., 1994]

To proceed with the GQM approach, goals should be identified initially. Considering our current context, the guidance type that we try to provide will match the goal in the GQM model. In the second step, questions have been developed to assess each goal. We carefully created the questions by refining the goal into several questions and ensure the questions we created are measurable. As several questions were developed, it is possible to get a complex guidance support to integrate throughout the configuration process. Hence, we decided to categorize them by combining the goals that will affect others. Finally, we consider that operations that we make, to answer to a given question, are the set of metric that provide the information to answer those questions. In this case, we will refine all the questions into metrics. As a result, five guidance modes act as goals in the GQM model :

Goal 1: Provide user-friendliness of the configuration task from the process miner viewpoint.

Goal 2: Reduce the required time to reach a valid configuration from the process miner viewpoint

Goal 3: Maximize the flexibility and efficiency of configuration alternatives to the lost users from the process miner viewpoint

Goal 4: Provide a maximum of personalization of configuration suggestions from the process miner viewpoint

Goal 5: Provide recommendations of configuration suggestions from the company interests viewpoint

Each goal has a purpose, an issue, an object, a stakeholder viewpoint and at least one question. Each question needs more than one metric to evaluate in order to have an answer. More details about guidance goals, questions and metrics are in the following subsubsections.

GQM Model 1 : Table 2 describes the GQM model for the following goal: *"Provide user-friendliness of the configuration task from the process miner viewpoint".* User-friendliness is considered as as functioning in the best manner with least waste of user effort. In order to achieve this goal we answered to the **Q1,n** questions, with n= 1..8, using the **M1,n,m** metrics.

Table 2: GQM model 1 for "Provide user-friendliness of the configuration task from the process miner viewpoint" goal

Goal 1	**Purpose**	**Provide**

	Issue	**user-friendliness of**
	Object	**the configuration task**
	ViewPoint	**from the process miner viewpoint**
Questions / Metrics		**Q1.1.** What are the not suitable processes that lead to abandonment? ***M1.1.1.*** *Median number of user interaction in a configuration* ***M1.1.2.*** *Number of undo actions in a configuration* ***M1.1.3.*** *Provide/not a final valid product* **Q1.2.** What are the processes that include fashionable products? ***M1.2.1.*** *Context information: Optimized/not optimized fashion interval date* ***M1.2.2.*** *Similarity measurements between configurations.* ***M1.2.3.*** *Provide/not a final valid product* **Q1.3.** What are the processes that include the bestseller products? ***M1.3.1.*** *Similarity measurements between configurations* ***M1.3.2.*** *Number of mostly selected feature* ***M1.3.3.*** *Provide/not a final valid product* **Q1.4.** What are the processes that include the worst-seller products? ***M1.4.1.*** *Similarity measurements between configurations* ***M1.4.2.*** *Number of rarely selected features* ***M1.4.3.*** *Provide/not a final valid product* **Q1.5.** What are the processes that include full option products? ***M1.5.1.*** *Number of selected features in x configuration* ***M1.5.2.*** *Number of undo actions in x configuration* ***M1.5.3.*** *Provide/not a final valid product* **Q1.6.** What are the processes that include minimum option products? ***M1.6.1.*** *Number of rejected features in x configuration* ***M1.6.2.*** *Number of undo actions in x configuration* ***M1.6.3.*** *Provide/not a final valid product* **Q1.7.** What are the processes that have to be interrupted by the system? ***M1.7.1.*** *Time taken without any user interaction* ***M1.7.2.*** *Configuration completeness degree* **Q1.8.** What are the processes that have been interrupted by the user? ***M1.8.1.*** *Number of taken decisions in x configuration* ***M1.8.2.*** *Number of undo actions in x configuration* ***M1.8.3.*** *Configuration completeness degree*

GQM Model 2 : Table 3 describes the GQM model for the following goal: *"Reduce the required time to reach a valid configuration from the process miner viewpoint"*. Required time is considered as as performing in the best manner with least waste of user time. In order to achieve this goal we answered to the **Q2,n** questions, with n= 1..5, using the **M2,n,m** metrics.

Table 3: GQM model 2 for "Reduce the required time to reach a valid configuration from the process miner viewpoint" goal

Goal 2	**Purpose**	**Reduce**
	Issue	**the required time to reach**
	Object	**a valid configuration**
	ViewPoint	**from the process miner viewpoint**
Questions / Metrics		**Q2.1.** What are the features that should be configured first? ***M2.1.1.*** *Number of processes that begin with a given feature* ***M2.1.2.*** *Number of mostly selected features* ***M2.1.3.*** *Provide/not a final valid product* **Q2.2.** What is the average number of steps and/ or required time to execute a process? ***M2.2.1.*** *Number of user interactions occurred while configuring a product.* ***M2.2.2.*** *Time taken to complete n configuration steps* ***M2.2.3.*** *Provide/not a final valid product* **Q2.3.** What is the ideal process order that minimize the solution charge? ***M2.3.1.*** *Number of user interactions occurred while configuring a product.* ***M2.3.2.*** *Time taken to complete n configuration steps* ***M2.3.3.*** *Provide/not a final valid product* **Q2.4.** What are the processes that take a lot of time? ***M2.4.1.*** *Time taken to complete n configuration steps* ***M2.4.2.*** *Tolerance in the face of the configuration duration average* **Q2.5.** What are the processes that were executed quickly? ***M2.5.1.*** *Robot/Real user Configuration* ***M2.5.2.*** *Time taken to complete n configuration steps*

GQM Model 3 : Table 4 describes the GQM model for the following goal: *"Maximize the flexibility and efficiency of configuration alternatives to the lost users from the process miner viewpoint"*. Efficiency and flexibility are vital to provide users an attractive configuration process. Helping lost users during their configuration tasks has to answer to the **Q3,n** questions, with n= 1..6, using the **M3,n,m** metrics.

Table 4: GQM model 3 for "Maximize the flexibility and efficiency of configuration alternatives to the lost users from the process miner viewpoint" goal

Goal 3	**Purpose**	**Maximize**
	Issue	**the flexibility and efficiency of**
	Object	**configuration alternatives to lost users**
	ViewPoint	**from the process miner viewpoint**
Questions / Metrics		***Q3.1.*** *How to know if the user is lost in a configuration process?* ***M3.1.1.*** *Time taken to configure a given feature* ***M3.1.2.*** *Number of undo actions in x configuration* ***Q3.2.*** *What are the features that cause backtracks during the configuration process?* ***M3.2.1.*** *Successful/Unsuccessful undo actions in x configuration* ***M3.2.2.*** *Manual/automatic value assignment of x feature* ***M3.2.3.*** *Number of implied features in an undo actions* ***Q3.3.*** *What are the processes that are executed with a maximum number of manual (user) choices?* ***M3.3.1.*** *Number of Manual/automatic choices in x configuration* ***M3.3.2.*** *Number of undo actions in x configuration* ***Q3.4.*** *What are the processes that are executed with a maximum number of automatic (propagated) choices?* ***M3.4.1.*** *Number of Manual/automatic choices in x configuration* ***M3.4.2.*** *Number of undo actions in x configuration* ***Q3.5.*** *What are the doubtful paths that lead to interrupt the process?* ***M3.5.1.*** *The average of transitions frequencies between feature* ***M3.5.2.*** *Number of undo actions in x configuration* ***M3.5.3.*** *Provide/not a final valid product* ***Q3.6.*** *What are the safest paths that lead to a valid product?* ***M3.6.1.*** *The average of transitions frequencies between feature* ***M3.6.2.*** *Number of undo actions in x configuration* ***M3.6.3.*** *Provide/not a final valid product*

GQM Model 4 : Table 5 describes the GQM model for the following goal: *"Provide a maximum of personalization of configuration suggestions from the process miner viewpoint"*. Capturing user identity, habits and past practices is very important in order to make them loyal. In response to this goal, we answered to the **Q4,n** questions, with n=

1..4, using the **M4,n,m** metrics.

Table 5: GQM model 4 for "Provide a maximum of personalization of configuration suggestions from the process miner viewpoint" goal

Goal 4	**Purpose**	**Provide**
	Issue	**a maximum of personalization**
	Object	**of configuration suggestions**
	ViewPoint	**from the process miner viewpoint**
Questions / Metrics		***Q4.1.*** *What are the processes that were configured by serious users?* ***M4.1.1.*** *Average time taken by an x user to complete a configuration step* ***M4.1.2.*** *Number of successful tasks in given time* ***M4.1.3.*** *Number of selected/rejected features comparing to the average* ***Q4.2.*** *What are the processes of a specific user based on his history?* ***M4.2.1.*** *Features selections and rejections of a given user* ***M4.2.2.*** *Similarity measurements between configurations of an x user* ***M4.2.3.*** *Timestamp of a given user* ***Q4.3.*** *What are the users that have similar configuration intentions?* ***M4.3.1.*** *Average of Features selections and rejections of n users* ***M4.3.2.*** *Similarity measurements between given configurations* ***Q4.4.*** ** What are the processes that include the best-seller products for each category of users (Gender, Age, Country)/ or for a specific period (for example summer holidays)* ***M4.4.1.*** *Similarity measurements between configurations* ***M4.4.2.*** *Include/not common processes executions* ***M4.4.3.*** *Provide/not a final valid product* ***M4.4.4.*** *Context informations: Measurements of differences and similarities between users* ***M4.4.5.*** *Context informations : Given period of filter*

* *: A question that need domain informations*

GQM Model 5 : Table 6 describes the GQM model for the following goal: *"Provide recommendations of configuration suggestions from the company interests viewpoint"*. Selling maximum of stock and/or minimizing production costs, etc. are interesting goals for the company. In response, we answered to the **Q5 question**, using the **M5,m** metrics.

Table 6: GQM model 5 for "Provide recommendations of configuration suggestions from the company interests viewpoint" goal

Goal 5	**Purpose**	**Provide**

	Issue	**Recommendations**
	Object	**of configuration suggestions**
	ViewPoint	**from the company interests viewpoint**
Question / Metrics		***Q5.*** ** What are the processes that lead to decrease the product cost and/or to increase the profitability for the company and/or to sell as much stock and/or to minimize the human resources cost* ***M5.1.*** *Similarity measurements between configurations.* ***M5.2.*** *Include/not common processes executions* ***M5.3.*** *Provide/not a final valid product* ***M5.4.*** *Context informations: Development cost of each feature* ***M5.5.*** *Context informations : Existent Stock of each feature* ***M5.6.*** *Context informations : Human resource needs of each feature*

** : A question that need domain informations*

B. The CONFILOG *process*

The CONFILOG process prescribes a sequence of phases to produce user assistance and improve product line configurations. Nevertheless, in our analysis, we were faced to different users concerns and contexts. Several guidance answers are possible from process mining algorithms. Therefore, we organized all these answers in the GQM framework in order to achieve our challenges. Each guidance scenario has a goal, questions and metrics that are extracted from logs. Figure 4 shows the phases of this guidance process. We are convinced that this process is portable also and can be applied in other process guidance contexts regardless of the product line configuration process.

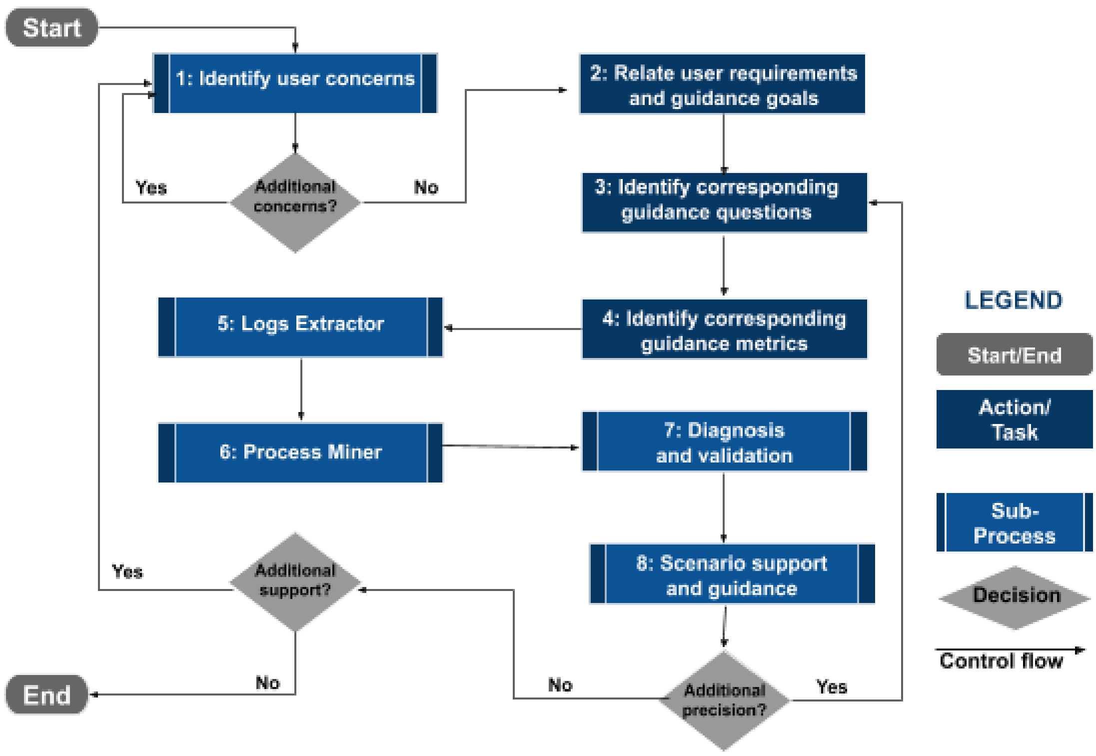

Figure 4. General flowchart for the ConfiLog process

Phase 1: "Identify users concerns" This phase identifies the user concerns. According to the PL context and other domain parameters, we identify what are the user configuration preferences and what they are searching about. Create the appropriate concern and relate specific concern to general ones is important. A return to Phase 1 to add new user concerns may be required.

Phase 2: "Relate user requirements and guidance goals" This phase relates all the user concerns with the corresponding guidance goals. First, identify an instance of the guidance goal. Second, refine the goal in a set of specific goals that satisfies the main goal individually or all together. (cf. section III.A)

Phase 3: "Identify corresponding guidance questions" This phase identifies from the ConfiLog framework, what are the most corresponding questions (cf. section III.A) that they meet the user concerns and the selected guidance goal.

Phase 4: "Identify corresponding guidance metrics" This phase identifies from the ConfiLog framework, what are the metrics (cf. section III.A) that have to be exploited in order to answer to the question selected in the third phase.

Phase 5: "Logs extractor" This phase implied the extraction and integration of data from the configurator. For each user action several attributes have to be extracted (cf. section II.C). Furthermore, knowledge is transferred tacitly during this phase through data filtering and comprehension. This guided us further in understanding how we could generate configuration support.

Phase 6: "Process Miner" This phase consists in applying process mining and analytics techniques (cf. section II.C) to the event logs. The aim of this stage is to provide answers for questions, asked in phase 3, using metrics, identified in phase 4, in order to achieve guidance goals, identified in phase 2.

Phase 7: "Diagnosis and Validation" This phase targets the users of our configurator. Our process should pass by the diagnosis and validation of the findings. Thus, the objectives of the diagnosis phase are : to interpret the mining results correctly, identify interesting or particular aspects, and propose other guidance questions for further iterations. The objective of the validation is to compare the findings to the expectations of the process stakeholder before providing him the guidance.

Phase 8: "Scenario support and guidance" This phase aims to provide the users what they are waiting for. This subprocess, is a transitions from statistics and mining results to real proposed features, configuration paths and/or ready configured products.

Once support is provided to users, the system's analyst review the precision and usability of the result. If additional precisions are required we go back to phase 3. Else, if others user concerns are detected or added we go back to phase 1.

C. *How to use the* CONFILOG *approach through a running example*

To illustrate CONFILOG, we provide real guidance scenarios in the online commercialisation of software as a service domain. Figure 5, shows the feature model [Kang et al., 1990] of the provided website product line [Mendonca et al., 2008] which is composed of 25 interrelated features. For example, the "Performance" Feature is a mandatory feature but the feature "Additional Services" is optional. Adding to that, if the feature "Reports" is selected in a configured product, the "File" and "Milliseconds" features must to be selected also. The selection of the feature "Milliseconds" excludes the selection of the feature "Https". The goal is to offer personalized solutions to enterprises that are looking to configure websites in different market segments.

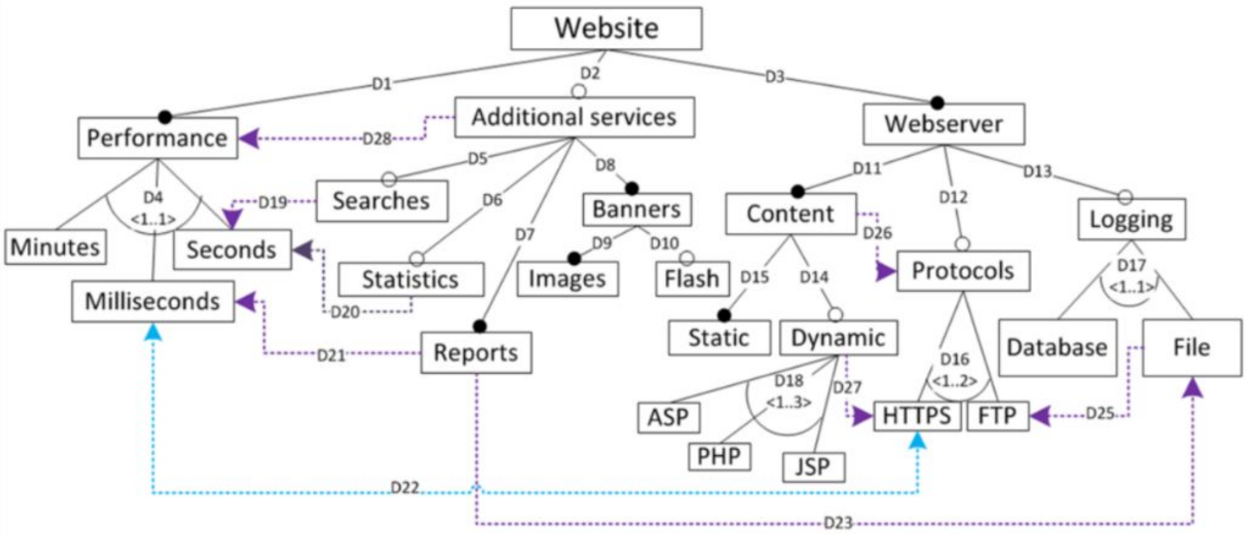

Figure 5. Adapted version of the website Feature Model [Mendonca et al., 2008]

To configure a website from this line, the customer interacts with an interface that allows him to select the website characteristics according to his preferences. In fact, If the user is looking for a premium website, the performance have to be powerful (on “seconds” or “milliseconds”) by automatic propagation of choices. On the other hand, several decisions have to be made by the user such as, his preferences for the “Content” branch, etc.

In a such e-commerce domain, it is common to offer products that attract new customers by identifying personalisation scenarios. Let us consider a configurator that is able to identify the best-selling websites in 2016 for textile companies in France, the best selling websites for particulars or the full options configured websites. According to traces, it is possible to identify also the historic of a specific user and his behavior with a link to the detailed product description. Suggest configuration processes and paths based on similarity measures between previous configuration is another guidance way that consists in answering the question. In our GQM model, it consists in answering the appropriate question using the adequate metrics (cf. section III.A). In our running example, we used the ConfiLog process in order to answer to all these configuration guidance requirements. Mining the old configurations makes possible to improve the next configurations and guiding the user’s choices.

Phase 1 in action: We are convinced that providing *an easy configuration process* and *maximize the personnalisation* in the same time, are the two important requirements that users are looking for when they are configuring a website from this line.

Phase 2 in action: To answer for these two user concerns, we have to identify what are the most oriented goals that match these guidance needs. Therefore, we selected from our GQM framework *Goal 1: “Provide user-friendliness of the configuration task from the process miner viewpoint”* and *Goal 4: “Provide a maximum of personalization of configuration suggestions from the process miner viewpoint”.*

Phase 3 in action: During this phase, the system analyst has to select the most pertinent guidance questions that correspond to the user requirements and the context. In our running example, we selected *Q1.2. What are the processes that include fashionable products?, Q1.3. What are the processes that include the bestseller products?, Q1.4. What are the processes that include the worst-seller products?, Q1.5. What are the processes that include full option products?, Q1.6. What are the processes that include minimum option products?, Q4.2. What are the processes of a specific user based on his history?* and *Q4.3. What are the users that have similar configuration intentions?*

Phase 4 in action: In order to answer to our selected guidance questions (in phase 3), we have to evaluate the following metrics using traces extracted from previous configuration processes : *M1.2.1. Context information: Optimized/not optimized fashion interval date, M1.2.2. Similarity measurements between configurations, M1.2.3. Provide/not a final valid product, M1.3.1. Similarity measurements between configurations, M1.3.2. Number of mostly selected feature, M1.3.3. Provide/not a final valid product, M1.4.1. Similarity measurements between configurations, M1.4.2. Number of rarely selected features, M1.4.3. Provide/not a final valid product, M1.5.1. Number of selected features in x configuration, M1.5.2. Number of undo actions in x configuration, M1.5.3. Provide/not a final valid product, M1.6.1. Number of rejected features in x configuration, M1.6.2. Number of undo actions in x configuration, M1.6.3. Provide/not a final valid product, M4.2.1. Features selections and rejections of a given user, M4.2.2. Similarity measurements between configurations of an x user, M4.2.3. Timestamp of a given user, M4.3.1. Average of Features selections and rejections of n users* and *M4.3.2. Similarity measurements* *between n configurations.*

Phase 5 in action: Evaluating metrics selected in phase 4 is based on traces (cf. section III.A). In our running example, we captured unfiltered traces from several users. Nevertheless, ETL measures were applied in order to select only pertinent traces that correspond to complete configurations. As attributes we selected all the informations that we planified (cf. table 1): *IDConfig, Originator, FeatureName, FeatureType, FeatureDecisionType, FeatureDecisionStep and Timestamp.* We are convinced that these attributes are useful and permit to evaluate the metrics above.

Phase 6 in action: After extracting traces about previous configuration, we used a process mining tool. We imported all these traces in a required format. The tool allow us to have frequency and performance statistics about configuration processes and it become possible to identify: *fashionable websites, bestseller websites, worst-seller websites, full option websites, minimum option websites, common last processes of x, users that have similar configuration intentions to x*. How to get these answers? Further details about mining steps, queries and comments are in section IV.

Phase 7 in action: In the one hand, we have descriptive statistics of previous configurations (Phase 6). In the other hand, we have guidance expectations that we have to meet. This stage is done by the user analyst that validates and diagnoses the usefulness and the sufficiency of process mining outputs in the current process. A combination and a synchronisation of all results are important.

Phase 8 in action: In this stage, the system analyst translates results of phase 6 to real guidance assets. For example, according to analysis, the feature *Additional Services* was selected in more than 70% processes. The manner and timing of recommending and displaying this mining result for next users is the aim of this phase.

IV. Preliminary Evaluations and Results

To perform our experiments, several steps have to be taken and they have to be in a certain order. Thus, a process for how to perform our experiments is needed. Figure 6 illustrates our experiment process which is divided into the following main activities. *Scoping* is the first step, where we scope the experiment in terms of goals. *Planning* comes next, where hypothesis, context, variables of the experiment is determined and the instrumentation is considered. *Operation* of the experiment follows from the design. In the operational activity, measures are collected which then are analyzed and evaluated in *analysis and interpretation*.

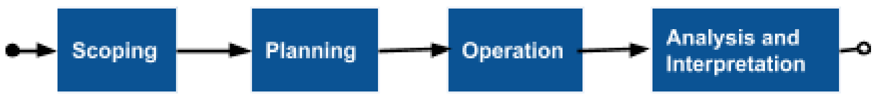

Figure 6. Experiment process inspired from [Wholin et al 2012]

A. *Experiment Scoping (Why?)*

The starting point is insight, and the idea that would be a possible way for evaluating whatever we are interested in. We planified the following experiment goal with a structure of the GQM method to determine goals:

"Provide solutions to users For the purpose of guidance With respect to the PL constraints In the context of PL configuration."

B. Experiment Planning (How?)

The planning phase determines the foundation for the experiment and how the it was conducted. The input of the planning phase is the goal definition. The planning experiment can be divided into 4 steps:

Context Selection: Our context of experiments is a Website configuration from a PL described in section III.C. Hence the experiment is conducted by customers that they are configuring a website from the line.. The experiments addresses a real problem, ie. the differences in performance and understanding of the domain, user differences profiles, full model constraints, etc.

Hypothesis formulation: An important aspect is to know and to formalize state clear what is going to be evaluated. This leads to the formulation of hypothesis. Here, it has been chosen to focus on two hypothesis : a PLM that we have to respect its constraints and a product model that contain the product that the user have in mind.

Variable Selection: The variable of our experiments is the features of our PLM and the traces captured after each user interaction.

Tooling instruments: Several tools like VariaMos[2] [Mazo et al., 2015] and SPLOT[3] [Mendonca et al., 2009] are proposed to configure PLs. These tools offer the possibility of configuring PLMs interactively with the user and recording latter the traces of a given configuration but not in the suitable format that we need. In process mining, two tools are reported to be extensively used: ProM[4] [Dongen et al., 2005] and DISCO[5] [Gunther 2012]. ProM contains a variety of mining algorithms but is less suited for handling very big event logs. On the contrary, DISCO is a commercial tool that can handle big data. Considering that, we opt to use DISCO in the current experimentation. With this tool it is also possible to create insightful process maps and statistics automatically.

C. Experiment Operation

Our experiment has been planned and in this step it must be carried out in order to collect the data that should be analyzed.

Preparation: The users were not aware of what aspects were going to be studied. They configured products that goes with their preferences.

Execution: The experiments was executed over 4 days, during which the product line model was configured 46 times. After 46 website configurations, we perceive that 1257 events (Feature selection or rejection) have been produced and recorded into the CSV (Comma-Separated Values) file. Each configuration contain more than an event. Every

[2] https://variamos.com/home/
[3] http://www.splot-research.org/
[4] http://www.promtools.org/doku.php
[5] https://fluxicon.com/disco/

event corresponds to the selection or the rejection of a given feature.

Data validation: Data was collected from 36 users. Many users configured more than one the PLM because we treated 46 configuration in our logs. We treated all traces, even traces coming from incomplete configurations. We are interested to listen to all user actions.

D. Experiment Analysis and Interpretations

Several treatments and interpretations can be held depending essentially on the intentions behind mining. Answers to the guidance questions are based on frequency and performance analysis. In this section, we focused on the questions of the first and the fourth goal of our approach. Further experiments details and answers to all the questions mentioned in section III are treated in the technical document below[6]. We created also a video demonstration[7] of our approach showing all tools, methods and procedures to do process mining in the context of PL configuration. In this demonstration we used the running example (cf. section III.C) in order to guide users.

First analysis discussions : According to the traces, a map model of the 46 website configurations processes is generated automatically using DISCO. Figure 6 shows a part of this map model that is highlighting frequent decisions: The start of the process is illustrated by the triangle symbol at the top of the process map. Similarly, the end of the process is illustrated by the stop symbol. Activities are represented by boxes (selection: 1 or rejection: 0 of a given feature) and the process flow between two activities is represented by an arrow. Dashed arrows point to activities that occurred at the very beginning or at the very end of the process. The absolute frequencies are displayed in the number at the arcs and the activities. The thickness of the arrows and the coloring of the activities visually support these numbers.

[6] https://docs.google.com/document/d/1M1sJ6nIxOvMuy_IoKo2fj_aCFIqGZjCTBCcI7Wj6Q2M/edit?usp=sharing
[7]https://www.youtube.com/watch?v=TdPe8Felgk8

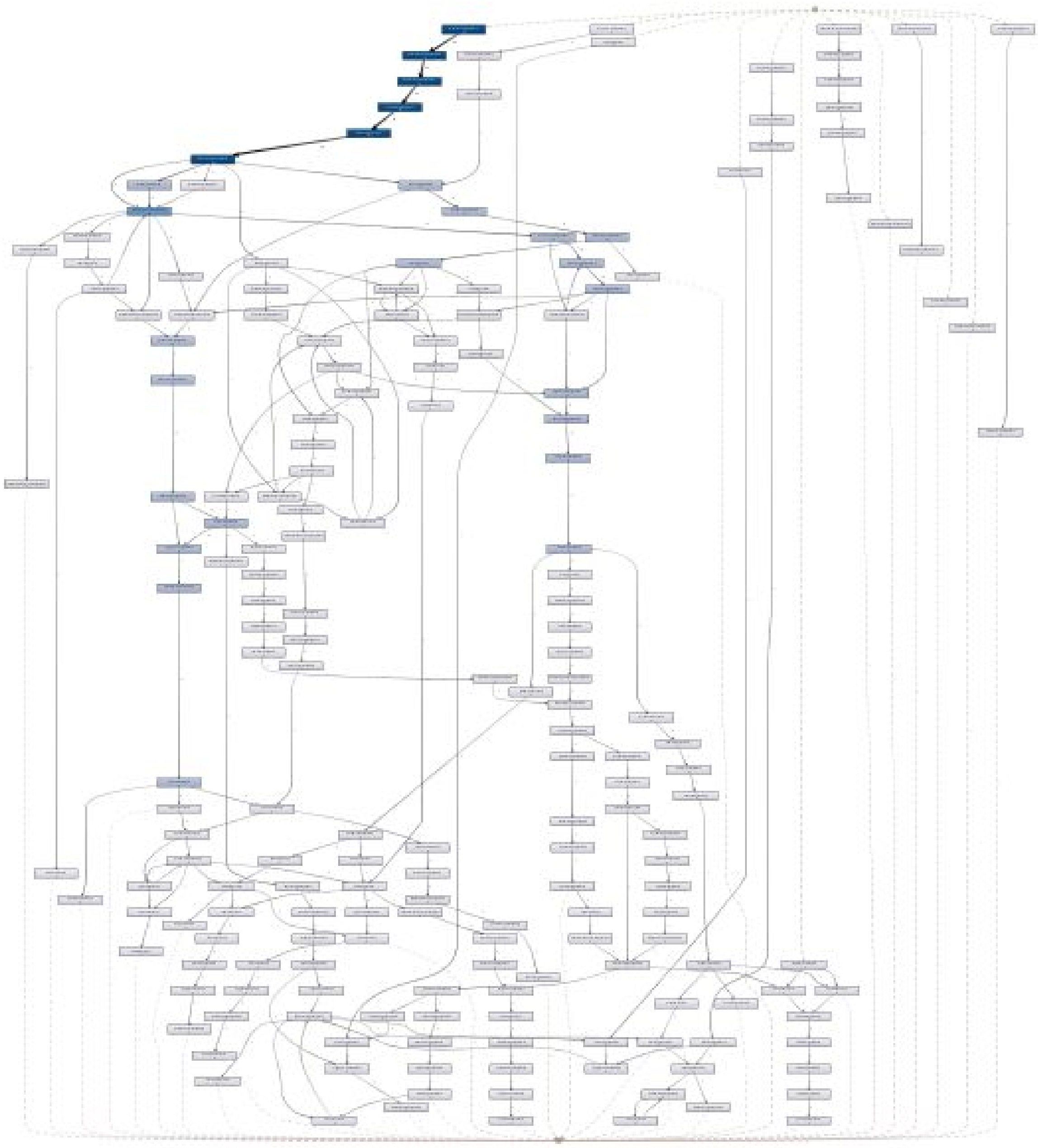

Figure 6- Map model of the frequent decisions made during configurations of the Website PL

From the map model, it’s validated that 46 configurations took place. The color of a feature varies according to its frequency in the traces. It is possible to recommend the dark and ignore the light colors. We can calculate the percentages of passage from one feature to another. From that point, metrics can be proposed, for example a threshold of 20%. Let’s agree, that less than 20%, we can judge that the given feature is not too relevant. And vice versa, for example + 80% a feature can be judged to be relevant. Every PLM feature is presented in the map model by two nodes. If a feature has a single node that means that it is mandatory in the model. And it will appear in each configuration for example

"Website", "Performance", "Webserver", "Content", "Static"; "Protocol".

In our running example, the mean frequency of activities is equal to 50.28. In fact, in a global context, activities with more than this threshold have to be recommended. In addition to that, there is 36 resources originators of traces. This reflects that from 46 configurations there are users who have done more than one configuration. The mean frequency of events according to users is equal to 29.23. We can judge if a user has produced more than 29 events, it is an active user. Less than 29 events, he is a simple user. In 1257 events, 678 features were selected (53.94%), 517 were rejected (41,13%) and 62 features redefined by backtracks (4,93%).

The performance analysis involves the time measurement criteria. The first configured feature was "Minutes" in 20 processes, and "File" in 11 processes. This reflects that the feature "Minutes" is important and numerous users tried to start their configurations with this feature. The last configured features were "Database", "Reports", "Banners", "Image" and "File". According to the captured traces, the average configuration duration is 22.6 minutes. The capture of traces is during the period from 25/05/2016 at 14h11 until 29/05/2016 at 11h58. A configuration that lasts more than 22 minutes can be considered as a long configuration. In our study, the longest configuration lasted 1 hour and 39 minutes, the shortest lasted 3 minutes. Answers to the question asked from the beginning (cf. Section III/A) become available thanks to the frequency and performance analysis.

Analysis interpretation and GQM recommendations: After analyzing previous website configurations, we detail answers to the guidance questions selected after phase 2 of our process (cf. section III.C). Answers to all the questions are available in our tutorial cited in reference 7. Table 7 contains 6 columns: The first column defines goals. In our processes we selected : Goal 1 and Goal 4. The second columns presents the question selected in order to achieve a given goal. Column 3 corresponds to the evaluated metrics that allow us to have a guidance answer. The fourth column presents the exploited traces taken into account to evaluate the metric, then Ai corresponds to the i attribute number (cf. table 1). Column 5 illustrates some comments about queries and filtering treatments that our process miner used to answer to the question. Recommendation results are shown in the last column.

Table 7 : GQM based on mining website line configuration processes using DISCO

Goal	Question	Metric	Attributes focus									Mining Queries & comments	Recommendation Results
			A1	A2	A3	A4	A5	A6	A7	A8	A9		
G1	**Q1.2**	**M1.2.1, M1.2.2, M1.2.3**	x								x	According to the context : **1**/Select processes that have recent Date and time < 2 days **2**/Select Variant with Max similar processes	Feature Nar · Feature Val · Feature Nar · Feature Val Website 1 · JSP 1 Performa... 1 · PHP 0 Webserver 1 · ASP 0 Content 1 · Additional... 1 Static 1 · Reports 1 Protocols 1 · Banners 1 FTP 1 · Images 1 Logging 1 · Flash 1 Database 0 · Minutes 0 File 1 · Secondes 1 HTTPS 1 · Searches 1 Milliseco... 0 · Statistics 0 Dynamic 1

	Q1.3	**M1.3.1, M1.3.2, M1.3.3**	x				x	x				Select Processes that have max similarity of events Feature Value with other events from other processes	Feature Na, Feature Val, Feature Na, Feature Val Website 1, Reports 0 Perform... 1, Banners 0 Webserver 1, Images 0 Content 1, Flash 0 Static 1, Dynamic 1 Protocols 1, HTTPS 1 Minutes 1, ASP 0 Milliseco... 0, PHP 1 Secondes 0, JSP 0 Searches 0, FTP 1 Statistics 0, Logging 0 Addition... 0, Database 0 File 0
	Q1.4	**M1.4.1, M1.4.2, M1.4.3**	x				x	x				Select Processes that have max difference of events Feature Value with other events from other processes	Processes with Id Config: 17, 18, 19, 21.
	Q1.5	**M1.5.1, M1.5.2, M1.5.3**	x		x	x	x					**1**/Select Processes that have max of events Feature Value = 1 **2**/Select processes without Backtracks.	Feature Name, Feature Value Website 1 Performance 1 Webserver 1 Content 1 Static 1 Protocols 1 FTP 1 Logging 1 File 1 HTTPS 1 Dynamic 1 JSP 1 Additional Service 1 Reports 1 Banners 1 Images 1 Flash 1 Secondes 1
	Q1.6	**M1.6.1, M1.6.2, M1.6.3**	x		x	x	x					**1**/Select Processes that have max of Feature Value = 0 **2**/Select processes without Backtracks.	Feature Name, Feature Value Database 0 Millisecondes 0 PHP 0 ASP 0 Minutes 0 Statistics 0

G4	**Q4.2**	**M4.2.1, M4.2.2, M4.2.3**	x	x									Select Processes executed by User = X (Sonia for example)	Resource · Feature Name · Feature Value Sonia · Website · 1 Sonia · Performance · 1 Sonia · Minutes · 1 Sonia · Millisecondes · 0 Sonia · Secondes · 0 Sonia · Additional Service · 0 Sonia · Searches · 0 Sonia · Statistics · 0 Sonia · Reports · 0 Sonia · Banners · 0 Sonia · Images · 0 Sonia · Flash · 0 Sonia · Webserver · 1 Sonia · Content · 1 Sonia · Static · 1 Sonia · Dynamic · 0 Sonia · ASP · 0 Sonia · PHP · 0 Sonia · JSP · 0 Sonia · Protocols · 1 Sonia · HTTPS · 1 Sonia · FTP · 0 Sonia · Logging · 1 Sonia · Database · 1 Sonia · File · 0
	Q4.3	**M4.3.1, M4.3.2**			x	x	x	x					**1**/ Select Similar processes that selected the same features **2**/ Identify their users	Processes with Id Config: 29 and 39 Users : Henri + Colette -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Processes with Id Config: 12 and 38 Users : Ines + Stephane -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Processes with Id Config: 20 and 26 Users : Pascal + Ines -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- Processes with Id Config: 36 and 37 Users : Ali + Melanie

From previous configuration of websites, assistance guidelines become available thanks to to the mining treatments. In this section we focused only in two goals : *Goal 1: “Provide user-friendliness of the configuration task from the process miner viewpoint”* and *Goal 4: “Provide a maximum of personalization of configuration suggestions from the process miner viewpoint”*. Combining these results enhance and assist users during their future experience and offer to them an easier configuration task. For example, from *Q1.2, Q1.3, Q1.4, Q1.5, Q1.6* the website provider can offers customers a specific discount for fashionable websites or highlights from the beginning the best sellers and/or premium configured websites. Allowing users to follow their last configurations processes (*Q4.2*) is also a kind of guidance by showing them their last choices and preferences. It is possible also thanks to the *Q5.3* to suggest paths based on similarity measures between users.

V. Related work and Discussion

In PL configurations that offer a large number of possible combinations it is difficult and time-consuming to produce a coherent configuration that respect the PLM restrictions and meet the user preferences in the same time. A solution is to help the user through an interactive process that reduces the number of decision that have to be done in each step.

As far as we know, there are no studies that explicitly attempt to apply process mining on PL configurations. Whereas many other works tried to capture configuration traces and exploit them in order to enhance the PL processes. We consider this kind of research as an implicit proposals of mining configuration processes. In this sense, [Chastel et al., 2015] proposed to suggest to users the next action when they are building a configuration. The authors used the result of clustering to suggest actions. They construct, later, the trace of the actions carried out and then by classification, using the same measures as for the construction of the clusters, they determine the closest cluster. Other works [Triki et al., 2013] proposed a recommendation approach combined with the product line configuration. The idea is to combine knowledge-based recommendation and the content-based recommendation. In other words, it is a combination between the dynamic recommendation and the similarity measure. The recommendation is based on textual data of previous configurations.

Moreover, it is possible to guide the configuration task by using heuristics that aim to reduce the number of user actions and/or minimize impact of decisions [Mazo et al, 2014]. To deal with complex configurations and also to integrate intelligent configuration processes, the notions of views and workflows have been defined [Abbasi et al., 2011] [Hubaux et al., 2009].

Our research aims also to provide recommendations during the configuration process and to suggest to the user next actions. We are simply handling with user traces but we are interested with techniques of process mining. Analyzing process traces and interpreting results aims to recommend personalized decisions for users and make easier their configuration task. This axis corresponds to the subject of our current and future research.

Nonetheless, several scientific questions are arisen concerning the the consequences of this approach for bsidered at design time. In our research, in order to provide guidance, we capitalized traces coming from the PLMs. If the PLM evolve, we come across an incompatibility between what is capitalized in trace and the new evolved PLM. The guidance and recommendation based on last PLM before changing become incoherent with the new PLM configuration requirements. What planifications we have to think about? What is the degree of dependence between the structure of logs in face of evolution, what artefacts can be reused in spite of new changes? In this context another promising question arises: what dependency and correlations exists between the configuration process and the PLM? What is the dependence degree between a configuration process and the PLM?

On the other side, a potential problem of ConfiLog is the “cold start” phenomenon because it is evolving an automatic extraction of traces. In the beginning, our approach cannot draw any inferences for users or items about which it has not yet gathered sufficient information. Indeed, the approach always starts with an absence of background. There are no logs of activities from which we will found relevant configuration guidelines and recommendations. ConfiLog is trace-dependent, since we extract traces we are reinforcing its performance. One of the effective solutions of the “cold start” problem is to adopt machine learning techniques that can improve the performance of ConfiLog by predicting the usefulness of configuration solutions.

For analyzing and eliciting user goals when they are configuring a PL, several approaches may be exploited. We tried to think about the representative ones to improve the ConfiLog feedbacks. The timing and the manner of recommendation can take into account the configuration context and the events order. Focusing on the an evenementiel approach synchronized with a context approach seems promising. We judge that goal models ie. KAOS (Knowledge Acquisition and AutOmated Specification) proposed by [Dardenne et al. 1993] may lead to relevant solutions. Under KAOS, we handle with a conceptual elicitation and modeling of functional and nonfunctional requirements, acquisition strategies to requirements modeling and an automated assistant of the processes.

Exploring the same context of prediction and anticipation of the user goals and intentions, we can also adopt the Bayesian networks; Especially as we have goals that depend on situations. The use of a markov approach seems also promising when we are dealing with stochastic processes : the predictions depend on the current state of the decisions.

Other future directions including the validation of the performance, the accuracy, and the scalability of the proposed approach are needed.

VI. Conclusion

Process mining has proved a powerful way for improving many processes in many domains, such as education, catering, airports, pharmacy, etc. Our proposal is mining previous PL configurations and listen to users actions when they are configuring a product from a line in order to enhance their future configuration processes. We believe that our approach is original as it is an explicit attempt to marry the process mining and the PL engineering contexts. We are convinced also that our approach framework and process is generic and can be applied in other context to support complex business processes. However, some further work is required in order to extend this proposal. Future work will focus on the events logs extraction and propose other informations that have to be captured in the same time, analysis must be based on a larger sample of configurations and from different user profiles. We are convinced that if we develop our own logs extractor the mining task will be more profitable and promising. A next step is to translate the process mining outputs into referral recommendation techniques that guide users and provide them a better experience. Recommendation guidelines have to be automatized according to the process mining analysis.

Acknowledgements

This work was supported by the French–Tunisian CMCU project 16G/1416 (CONFIGURE). All the authors of this paper are grateful and thank the committee members of this research program for their assistance and financial support.

REFERENCES

Aalst, W. M. P. (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st ed. Springer Publishing Company, Incorporated.

Abbasi E. K., Hubaux A., Heymans P. (2011). A toolset for feature-based configuration workflows. In

Proceedings - 15th international software product line conference, splc 2011, p. 65–69.

Batory, D. (2005). Feature models, grammars, and propositional formulas. In Proceedings of the 9th international conference on Software Product Lines (SPLC'05), Henk Obbink and Klaus Pohl (Eds.). Springer-Verlag, Berlin, Heidelberg, 7-20. DOI=http://dx.doi.org/10.1007/11554844_3.

Basili, V.R., Caldiera, C., and Rombach, H.D.(1994) Goal/Question/Metric Paradigm," Encyclopaedia of Software Engineering, Vol. 1,J.J. Marciniak, ed., John Wiley and Sons, New York, pp. 528-532.

Benavides, D., Trinidad, T., Ruiz-Cortés, A. (2005). Automated Reasoning on Feature Models. Springer Publishing Company, Incorporated, Berlin, Heidelberg.

Clements, P., & L. Northrop. (2001). Software Product Lines: Practices and Patterns, 1st ed. Addison-Wesley Professional.

Da Silveira, G., Borenstein, D., & Fogliatto, F. S. (2001). Mass customization: Literature review and research directions. International journal of production economics, 72(1), 1-13.

Dhungana, D., Grünbacher, P and Rabiser, R., (2010). The DOPLER Meta-Tool for Decision Oriented Variability Modeling: A Multiple Case Study. Automated Software Engineering (in press; doi: 10.1007/s10515-010-0076-6).

Djebbi, O., Salinesi, C & Fanmuy, G. (2007). Industry Survey of Product Lines Management Tools: Requirements, Qualities and Open Issues. Proc. of the International Conference on Requirement Engineering (RE), IEEE Computer Society, New Delhi, India.

Dongen, V. B., de Medeiros, A., Verbeek, H., Weijters, A., and van der Aalst, W., The prom framework: A new era in process mining tool support, in Applications and Theory of Petri Nets 2005, ser. Lecture Notes in Computer Science, Ciardo, G., and Darondeau, P., Eds. Springer Berlin Heidelberg, 2005, vol. 3536, pp. 4.

Falkner, A., felfernig A., haag, A. (2011). Recommendation Technologies for Configurable Products". AI Magazine pp. 99-108.

Günther, C. W. (2009). Process Mining in Flexible Environment. PhD thesis, Technische Universiteit Eindhoven.

Gunther, C. W. and Rozinat, A. (2012) Disco: Discover your processes. In BPM.

Hadzic, T., Andersen, H.R. (2004). An introduction to solving interactive configuration problems. Technical Report TR-2004-49, The IT University of Copenhagen.

Hadzic, T., Subbarayan, S., Jensen, R.M., Andersen, H.R., Mller., Hulgaard, J. H.(2004). Fast backtrack-free product configuration using a precompiled solution space representation. Small 10(1), 3.

Hubaux A., Classen A., Heymans P. (2009). Formal Modelling of Feature Configuration Workflows. In Proceedings of the 13th international software product line conference (splc'09), p. 221–230.

Junior, C. M., Cirilo, E., and Lucena, C. (2011). Assisted user-guidance in collaborative and dynamic software product line configuration.

Kang, K ., Cohen, S., Hess, J., Novak, W., & Peterson, S. (1990). Feature-Oriented Domain Analysis (FODA)Feasibility Study. Technical Report CMU/SEI-90-TR-21, Software Engineering Institute, Carnegie Mellon University, USA.

Mazo, R., Dumitrescu, C., Salinesi, C., Diaz, D. (2014). Recommendation Heuristics for Improving Product Line Configuration Processes. In: Recommendation Systems in Software Engineering. Robillard M, Maalej W., Walker R. and Zimmermann T. (Eds.), ISBN 978-3- 642-45135-5, DOI 10.1007/978-3-642-45135-5__19, Springer-Verlag Berlin Heidelberg.

Mazo, R., Muñoz-Fernández, J., Rincón, L., Salinesi, C., Tamura, G., (2015) VariaMos: an extensible tool for engineering (dynamic) product lines. In the XIX International Software Product Line Conference (SPLC), Nashville-USA, July 20-24, 2015.

Mazo, R., Salinesi, C., Diaz, D., (2011). Abstract Constraints: A General Framework for Solver-Independent Reasoning on Product Line Models. Accepted on INSIGHT - Journal of International Council on Systems Engineering (INCOSE).

Mendonca, M., Branco, M., S.p.l.o.t. : software product lines online tools. (2009). In Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications, pages 761–762. ACM.

Mendonca, M., Bartolomei, T.T., and Cowan, D. (2008). Decision-making coordination in collaborative product configuration. In Proceedings of the 2008 ACM symposium on Applied computing, SAC ’08, pages 108–113, New York, NY, USA. ACM.

Mobasher, B., Cooley, R., Srivastava, J. (2000). Automatic personalization based on Web usage mining. Commun. ACM 43, 8 (August 2000), 142-151. DOI=http://dx.doi.org/10.1145/345124.345169

Pohl, K., Böckle, G., et Linden, & F. V. D. (2005). Software Product Line Engineering, Foundations, Principles, and Techniques. Springer Publishing Company, Incorporated.

Quinton C., Romero D., Duchien L. (2014). Automated selection and configuration of cloud environments using software product lines principles. In 2014 IEEE 7th international conference on cloud computing, anchorage, ak, usa, june 27 - july 2, 2014, p. 144–151. IEEE. visited on http://dx.doi.org/10.1109/CLOUD.2014.29

Robillard, M., Maalej, W., Walker R., and Zimmermann, T. (2014). Recommendation Systems in Software Engineering, Berlin Heidelberg.

Rodic, J., and Baranovic, M. (2009). Generating data quality rules and integration into ETL process. In Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP (DOLAP '09). ACM, New York, NY, USA, 65-72. DOI=http://dx.doi.org/10.1145/1651291.1651303

Salinesi, C., Mazo, R., Djebbi, O., Diaz, D., Lora-Michiels, A. (2011). Constraints: the Core of Product Line Engineering. Fifth IEEE International Conference on Research Challenges in Information Science (RCIS), IEEE Press, Guadeloupe-French West Indies, France.

Schonenberg, H., Weber, B., Dongen, B., and Aalst, W. (2008). Supporting Flexible Processes through Recommendations Based on History. In Proceedings of the 6th International Conference on Business Process Management (BPM '08), Marlon Dumas, Manfred Reichert, and Ming-Chien Shan (Eds.). Springer-Verlag, Berlin, Heidelberg, 51-66. DOI=http://dx.doi.org/10.1007/978-3-540-85758-7_7

Tiwari, A., Turner, C., and Majeed, B. (2008). A review of business process mining: state-of-Art and future trends,” Business Process Management Journal, vol. 14, no. 1, pp. 5–22.

Tartler, R., Lohmann, D., Sincero, J., and Schröder-Preikschat, W. (2011). Feature consistency in compile-time-configurable system software: facing the linux 10,000 feature problem. In Proceedings of the sixth conference on Computer systems (EuroSys '11). ACM, New York, NY, USA, 47-60. DOI=http://dx.doi.org/10.1145/1966445.1966451

Triki R., Mazo R., Salinesi C., (2014), Combining configuration and recommendation to enable an interactive guidance of product line configuration, Gérald Kembellec, Ghislaine Chartron and Imad Saleh. Les systèmes et moteurs de recommandation, Hermès, pp.141 pages-160 pages, 2014 pages.

Urli S., Blay-fornarino M., Collet P. (2014). Handling Complex Configurations in Software Product Lines : a Tooled Approach. In ACM (Ed.), Proceedings of the 18th international software product line conference (splc’14), p. 112–121. Florence, Italy.

Urli S. (2015). Processus Flexible de Configuration pour Lignes de Produits Logiciels Complexes. Phd thesis, Université Nice Sophia Antipolis. Visited on sur https://tel.archives -ouvertes.fr/tel-01134191v1.