



HAL
open science

Knowledge extraction using a conceptual information system (ExCIS)

Laurent Brisson

► **To cite this version:**

Laurent Brisson. Knowledge extraction using a conceptual information system (ExCIS). ODBIS 2005/2006: 1st and 2nd VLDB Workshops Ontologies-Based Databases and Information Systems, Sep 2005, Séoul, South Korea. pp.119 - 134, 10.1007/978-3-540-75474-9_8 . hal-02281800

HAL Id: hal-02281800

<https://hal.science/hal-02281800>

Submitted on 11 Apr 2023

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.

Knowledge Extraction Using a Conceptual Information System (ExCIS)

Laurent Brisson

Laboratoire I3S - Université de Nice, 06903 Sophia-Antipolis, France
brisson@i3s.unice.fr

Abstract. It is a well known fact that the data mining process can generate thousands of patterns from data. Various measures exist for evaluating and ranking these discovered patterns but often they don't consider user subjective interest. We propose an ontology-based data-mining methodology called ExCIS (Extraction using a Conceptual Information System) for integrating expert prior knowledge in a data-mining process. Its originality is to build a specific Conceptual Information System related to the application domain in order to improve datasets preparation and results interpretation. This paper focus on our ontological choices and an interestingness measure IMAK which evaluates patterns considering expert knowledge.

1 Introduction

One important challenge in data mining is to extract interesting knowledge and useful information for expert users. Numerous works focused on indexes that measure the interestingness of a mined pattern [5,9]. They generally distinguished objective and subjective interest. Silberschatz and Tuzhilin [14] proposed a method to define unexpectedness and actionability via *belief systems* while Liu [9] developed a method that use *user expectations*.

In most data mining projects, prior knowledge is implicit or is not organized as a structured conceptual system. ExCIS is dedicated to data mining situations where the expert knowledge is crucial for the interpretation of mined patterns. In this approach, an application ontology is built by analyzing existing databases with collaboration of expert users who play a central role. The main objective in ExCIS is to propose a framework in which the extraction process makes use of a well-formed conceptual information system (CIS) for improving the quality of mined knowledge. We consider the paradigm of CIS as defined by Stumme [18]: a relationnal database together with conceptual hierarchies. The CIS provides an useful structure for further mining tasks.

An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world [4]. Extracting ontological structures from data is very similar to processes of retrieving a conceptual schema from legacy databases [6]. They are based on the assumption that sufficient knowledge is stored in databases in order to construct the ontology.. They generally apply a matching between ontological

concepts and relational tables such that the ontology extracted is very close to the conceptual database schema.

In ExCIS, the ontology provides a conceptual representation of the application domain by analyzing the existing operational databases. ExCIS main characteristics are:

- Prior knowledge conceptualization: the CIS is specially designed for data mining tasks
- Adaptation of the CRISP-DM [2] methodology with CIS based preparation of data sets to be mined, CIS based post processing of mined knowledge in order to extract surprising and/or actionable knowledge and an incremental evolution of the expert knowledge stored in the CIS.

Our project deals with data from the ‘family’ branch of the French national health care system. The issue we address is to improve relationships between beneficiaries and the CAF organism. In this case study, we had two sources of information: a database storing data on beneficiaries and expert users aware of the business processes, behaviors and habits in the organism.

The topic of this paper is the use of ontologies for data mining. Our goal is to enhance data mining tasks and to extract interesting patterns according user’s knowledge. The novelty of ExCIS methodology lies in the creation of a CIS in order to compare knowledge and extracted patterns. We use an ontology based approach for unexpected and actionable patterns extraction while works on interestingness measures deals with templates [9] or beliefs [14]. Furthermore, using user’s knowledge in actionable patterns extraction differs from Piatetsky-Shapiro [13] or Silberschatz [15] approaches.

The paper is organized as follows. Section 2 presents related works. Section 3 gives an overview of the ExCIS approach. Section 4 describes the underlying conceptual structures of the ontology. In Section 5, we give a detailed description of CIS construction. Section 6 focus on knowledge database construction and interesting patterns extraction. Section 7 presents experiments results. Section 8 concludes the paper.

2 Related Works

2.1 Interestingness Measures

Numerous works focused on indexes that measure the interestingness of a mined pattern [5,9,10]. They generally distinguished objective and subjective interest. Among these indexes there are quantitative measures of objective interestingness such as confidence, coverage, lift, success rate while unexpectedness and actionability are proposed for subjective criteria. Since our work deals with user interestingness, we focus this state of the art on the former. According to the actionability criteria, a model is interesting if the user can start some action depending on it [15]. On the other hand, unexpected models are considered interesting since they contradict user expectations which depend on his beliefs.

User expectations is a method developed by Liu [9]. The first approach neither dealt with unexpectedness nor actionability. User had to specify a set of patterns according to his previous knowledge and intuitive feelings. Patterns had to be expressed in the same way that mined patterns. Then Liu defined a fuzzy algorithm which matches these patterns. In order to find actionable patterns, the user has to specify all actions that he could take. Then, for each action he specifies the situation under which he is likely to run the action. Finally, the system matches each discovered pattern against the patterns specified by the user using a fuzzy matching technique.

Silberschatz and Tuzhilin [14] proposed a method to define unexpectedness via belief systems. In this approach, there are two kinds of beliefs: soft beliefs that the user is willing to change if new patterns are discovered and hard beliefs which are constraints that cannot be changed with new discovered knowledge. Consequently this approach assumes that we can believe in some statements only partially. That's why some degree or confidence factor is assigned to each belief. A pattern is said to be interesting relatively to some belief system if it 'affects' this system, and the more it 'affects' it, the more interesting it is. However, interestingness of a pattern depends also on the kind of belief.

2.2 Databases and Ontologies

Ontologies provide a formal support to express beliefs and prior knowledge on a domain. Domain ontologies are not always available; they have to be built specially by querying expert users or by analyzing existing data. Extracting ontological structures from data is very similar to the process of retrieving a conceptual schema from legacy databases. Different methods [7,6,17] were proposed. They are based on the assumption that sufficient knowledge is stored in databases for producing an intelligent guide for ontology construction. They generally apply a matching between ontological concepts and relational tables such that the ontology extracted is very close to the conceptual database schema.

2.3 Ontologies and Data Mining

For the last ten years, ontologies have been extensively used for knowledge representation and analysis mainly in two domains: Bioinformatics and web content management. Biological knowledge is nowadays most often represented in 'bio-ontologies' that are formal representations of knowledge areas in which the essential terms are combined with structuring rules that describe relationships between the terms. Bio-ontologies are constructed according to textual descriptions of biological activities. One of the most popular bio-ontology is *Gene Ontology*¹ that contains more than 18 thousands terms. It describes the molecular function of a gene product, the biological process in which the gene product participates, and the cellular component where the gene product can be found. Results of data mining processes can then be linked to structured knowledge

¹ <http://www.geneontology.org/>

within bio-ontologies in order to explicit discovered knowledge, for instance to identify biological functions of genes within a cluster. Interesting surveys of ontologies usage for bio-informatics can be found in [1,16]. A successful project of data mining application using bio-ontologies is described in [19].

In the domain of web content management, OWL (Ontology Web Language)² is a Semantic Web standard that provides a framework for the management, the integration, the sharing and the reuse of data on the Web. Semantic Web aims at the sharing and processing of web data by automated tools as well as by people. It can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms, i.e. an ontology. Web ontologies can be used to enrich and explain extracted patterns in many knowledge discovery applications to web such as web usage profiling [3] for instance.

3 Overview of the ExCIS Approach

ExCIS integrates prior knowledge all along the mining process: the first step structures and organizes the knowledge in the CIS and further steps exploit it and enrich it too.

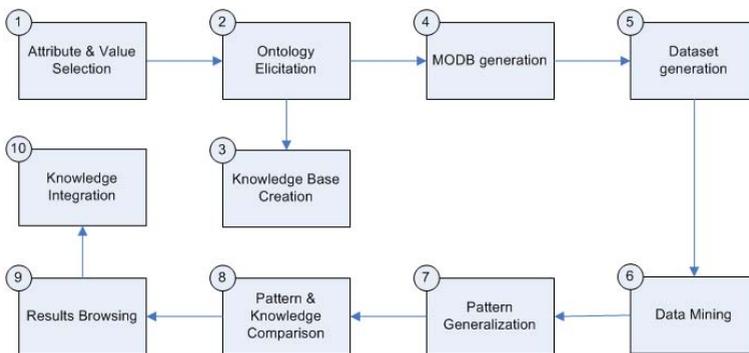


Fig. 1. ExCIS Process

Figure 1 describes ExCIS process from attribute selection to extraction of interesting patterns. Figure 2 describe information flow in ExCIS approach. On each arrow a number refers to a subprocess in figure 1. In this paper, section 4 describe subprocesses 1 to 3 while section 6 describe subprocess 8.

The global ExCIS process presented in these figures shows:

- The CIS construction where:
 - The ontology is extracted by analyzing original databases and by interacting with expert users.

² <http://www.w3.org/TR/owl-ref/>

- The knowledge base, set of factual informations, is obtained in a first step from dialogs with expert users.
 - The new generic Mining Oriented Database (MODB) is built. It contains data cleaned and prepared using domain knowledge.
- The pre-processing step where specific datasets may be built for specific mining tasks.
 - The standard mining step which extracts patterns from these datasets.
 - The post-processing step where discovered patterns may be interpreted and/or filtered according to both prior knowledge stored in the CIS and individual user attempt.

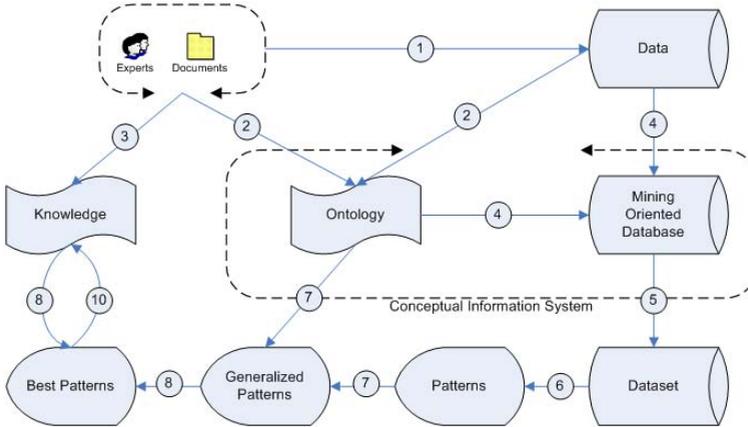


Fig. 2. Information flow

In this paper we call “patterns” a set of itemsets. Technically, we use the CLOSE algorithm [12] to extract association rules (one itemset for antecedent and another one for consequent).

The MODB is said to be generic since it will be used as a kind of basic data repository from which any task-specific dataset may be generated. We call MODB a relational database whose attributes and values are concepts of the ontology we defined. The underlying idea in the CIS is to build structures which will provide more flexibility not only for pre-processing the data to be mined, but for filtering and interpreting discovered patterns in a post-processing step. Hierarchical structures and generalization/specialization links between ontological concepts play a central role to allow reducing the volume of extracted patterns and to provide a tool for interpreting results obtained by clustering algorithms.

For numerical or categorical data, they provide different granularity levels which are useful in the pre-processing and the post-processing steps.

4 Conceptual Structures of the Ontology

4.1 Ontology

In ExCIS the ontology is an essential means both for improving data mining processes and for interpreting data mining results. It's an *application ontology* as defined by Guarino [4], ie. an ontology which describe concepts depending both on a particular domain (family branch of the french health care system in our application) and task (data mining in ExCIS methodology). The ontology is defined by a set of concepts and relationships among them which are discovered by analyzing existing data. It provides support both in the pre-processing step for building the MODB and in the post-processing steps for refining mined results.

Generalization/specialization relationships between ontological concepts provide valuable information since they may be used intensively for reducing and interpreting results. For instance, a set of dependency rules may be reduced by generalization on attributes or by generalization on values. Thus the guidelines in the ontology construction are:

- To distinguish attribute-concept (a data property) and value-concept (a value of a data property).
- To establish matching between source attributes and attribute-concepts and a matching between source values and value-concepts.
- To define concept hierarchies between concepts.

This ontology does not contain any instances since values are organized in hierarchies and considered as concepts. The MODB is a relational database whose role is to store the most fine-grain data elicited from the original database. MODB attributes are those which are identified as relevant for the data mining task and MODB tuples are composed of the most fine-grain values.

4.2 Ontology Relationships

A relationship is an oriented link between two concepts. In ExCIS there are two different kinds of concepts (see figure 3) and we distinguish relationships between concepts of the same hierarchy and concepts of different hierarchies.

Relationships from an attribute-concept toward a value-concept are forbidden since the relationship "is value of" has no meaning in this situation. There exist five different relationships between concepts (see Table 1). Numbers in this table refer to relationships in figure 4(a), 4(b) and 5. Among all the relationships, we can set up 3 different categories:

Relationships between value-concepts. Generalization or specialization relationships between value-concepts (see relationship 5 figure 5) are useful in order to generalize patterns during the post-processing step. Furthermore, relationships between two value concepts of *the same hierarchy* are essential since they allow to select data granularity in datasets generated from the MODB (see relationship 3 figure 4).

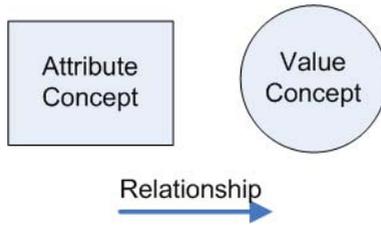


Fig. 3. Representation of concept and relationship

Table 1. Concept relationships

Concept	Within the same hierarchy		Between different hierarchies	
	Attribute	Value	Attribute	Value
Attribute	1 gens			
Value	2 valueOf	3 gens	4 valueOf	5 relationWith

Relationships between attribute-concepts. Generalization or specialization relationships between attribute-concepts are useful in order to generalize models during the post-processing step.

Relationships between two concrete attribute-concepts of the same hierarchy are specific because they have to be checked during datasets generation: indeed these attributes cannot be in the same dataset to avoid redundancy.

ExCIS method forbids relationships between attribute-concepts of different hierarchies because attribute-concepts which are semantically close have to be located together in the same hierarchy (see relationship 1 figures 4,5).

Relationships between value-concepts and attribute-concepts. These relationships are essential in order to build data or to provide different semantic views during the post-processing step. For instance, “98001” is both a “Home Location” and a “Zip Code” (see relationships 2,4 figure 5). If concepts are semantically close they must be in the same hierarchy and if they are slightly different they can be into two different hierarchies.

5 Conceptual Information System Construction

ExCIS differs from CRISP-DM mainly in the data preparation step. In this step CRISP-DM describes 5 tasks: select, clean, construct, integrate and format data. Selection and format are identical in both methods but in ExCIS cleaning, construction and integration are improved in order to elicitate the ontological concepts and to build the MODB.

Let A the set of source database attributes, C the set of ontology concepts and C_z the set of concepts associated to an attribute $z \in A$. C is defined by $\bigcup_{z \in A} C_z$.

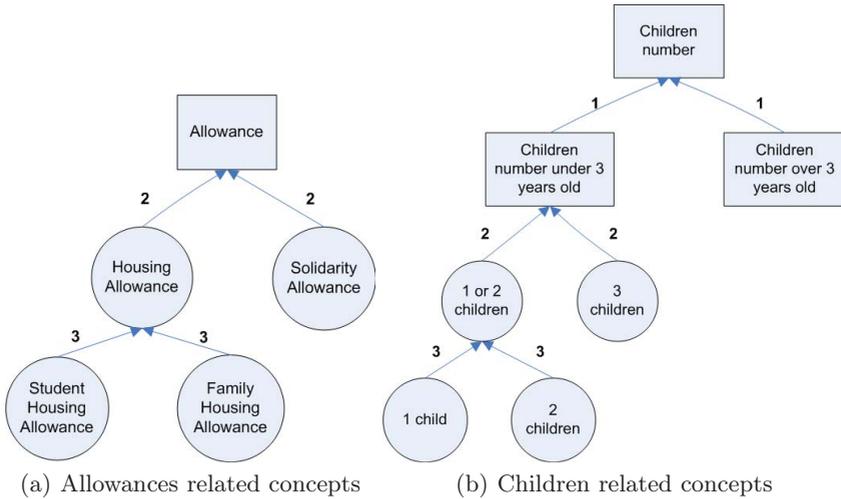


Fig. 4. Concept Hierarchies

5.1 Scope Definition and Source Attribute Selection

First steps of ExCIS method are related to the Business Understanding and the Data Understanding steps of CRISP DM method. They need an important interaction with expert users.

1. Determine objectives: in our case study, objectives are to improve "relationships with beneficiaries".
2. Define themes: analysis of data allow to gather them into semantic sets called themes. For example we create 3 themes: Allowance beneficiaries profiles, contacts (by phone, by mail, in the agency, ...) and events (holidays, school starts, birth, wedding, ...).
3. For each theme select a set of source attributes with experts users.

5.2 Data Analysis and Attribute-Concept Elicitation

4. For each selected attribute z :
 5. Examine name and description in order to:
 - Associate n concepts to the attribute.
 - Into C , clean homonyms (different concepts with same name), synonyms (same concepts with different names like age and date of birth) and useless attributes according the objectives.
 6. Examine values (distribution, missing values, duplicates values, ...) in order to:
 - Refine C_z (add or delete concepts) according to information collected in step 6.

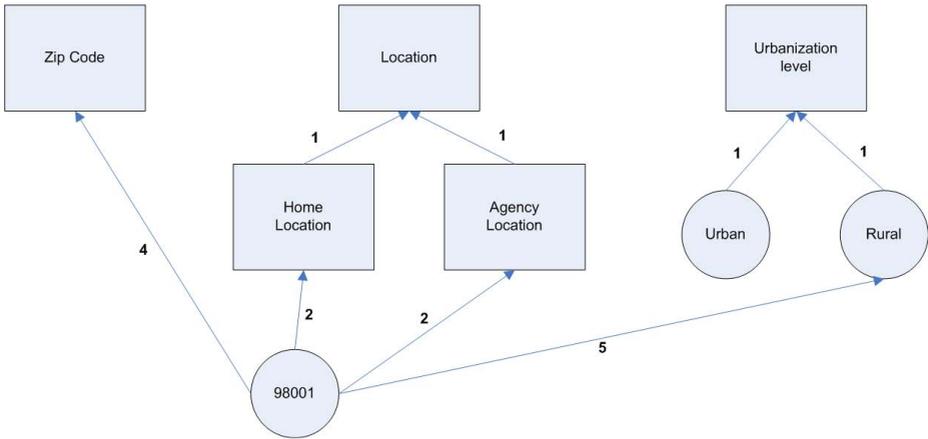


Fig. 5. Location related concepts

- Clean again homonyms, synonyms and useless attributes. For example by analyzing values we realized that ‘allowances’ was in fact 2 homonyms concepts. Thus we created the ‘allowance amount’ concept and the ‘allowance beneficiary’ concept.

7. For each concept associated to z create the method which generates value-concepts.

In the step 7, if the attribute-concept doesn’t exist we have to create a 4 fields record table. These fields are the attribute associated to the concept, the name of the attribute table in source database, the attribute domain value and the reference to the procedure which may generate value-concepts. There is only one procedure for record in the table. A domain value can be a distinct value or a regular expression and is the input of the procedure. Procedure output provides references to value-concepts. The procedure might be an SQL request (SELECT or specific computation) or an external program (script, shell, C, ...). However, if the attribute-concept already exists we just have to add a record in the table and create a new procedure.

5.3 Value-Concept Elicitation

At this point, all of the methods to generate value-concepts are created.

8. Give a name to each value-concept.
9. Clean homonyms and synonyms among value-concepts.

5.4 Ontology Structuration

10. Identify generalization relationships among value-concepts (see figure 4(a)).

11. If necessary, add new concepts to structure the ontology. For instance ‘Location’ concept in figure 5.
12. Create relationships between value-concepts of different hierarchies (see relation 4 figure 5).

5.5 Generation of the Mining Oriented Database

13. Generate the database by using procedures defined in step 7.

In this final step a program reads the tables created for each attribute-concept and calls the procedures in order to generate the MODB.

6 Interesting Patterns Extraction According Prior Knowledge

6.1 Knowledge Properties

We chose to express knowledge like “if ... then ...” rules in order to simplify comparison with extracted association rules. Each knowledge have some essential properties to select the most interesting association rules:

- Source: user defined knowledge or association rule selected as “new knowledge”
- Confidence level: 5 different values are available to describe knowledge confidence according a domain expert. These values are range of confidence value: 0-20%, 20-40%, 40-60%, 60-80% and 80-100%. We call confidence the probability the consequence of a knowledge occurs when the condition holds.
- Certainty:
 - Triviality: cannot be contradicted
 - Standard knowledge: domain knowledge usually true
 - Hypothesis: knowledge the user want to check

This is an example of knowledge:

KNOWLEDGE 1

Objective=‘To be paid’ \wedge Allowance=‘Housing Allowance’ \wedge Distance=‘0km’ \rightarrow Contact=‘At the agency’

- Source: user-defined
- Confidence level: 60-80%
- Certainty: Hypothesis

6.2 Ruled-Based Knowledge Base

Knowledge, like interestingness, differs from people and changes over time. That’s why our knowledge base is divided into several sets. A main set defines high level and consensual knowledge while subsets allow to define user’s knowledge. Since knowledge may differs between experts, the main objective of this knowledge base is to provide the expert some personalized models according his current knowledge.

6.3 Interesting Patterns Extraction

Interestingness definitions. In [14] Silberschatz presents a classification of measures of interestingness and identifies two major reasons why a pattern is interesting from the subjective (user-oriented) point of view:

- Unexpectedness: a pattern is interesting if it is surprising to the user
- actionability: a pattern is interesting if the user can do something with it to his or her advantage

Therefore a pattern can be said to be interesting if it is both unexpected and actionable. This is clearly a highly subjective view of the patterns as actionability is dependent not only on the problem domain but also on the user's objectives at a given point in time [11].

Although unexpected patterns are interesting it's necessary to consider actionable expected patterns. In our approach we deal with actionability using knowledge certainty property:

- If a pattern match a trivial knowledge it isn't actionable since actions concerning trivial knowledge are most likely known
- Since user knowledge define his main points of interest, a pattern matching standard knowledge could be actionable
- If a pattern matches a hypothesis, it is highly actionable

Patterns and knowledge comparison. In this paper we compare patterns and knowledge considering generalization relationships between them. In future works, we will introduce a distance measure which will consider both differences and generalization relation between patterns and knowledge. Liu introduce such a distance measure in [9] to deal with *general impressions*.

We propose an interestingness measure *IMAK* (Interestingness Measure According Knowledge) which consider:

- actionnality, using certainty knowledge property
- unexpectedness, using generalization relationships between patterns and knowledge.

At the moment, we don't use a distance measure so we cannot consider patterns that differ partially from knowledge if there is no generalization relationship between them. However these patterns are interesting and need further treatments. Patterns that are totally different from knowledge can't be evaluated by *IMAK* measure but they could be interesting since they're unexpected.

Consequently *IMAK* is useful in order to evaluate interestingness of patterns which are comparable to prior knowledge. This measure describe four levels of interest:

- none: uninteresting information
- low: confirmation of standard knowledge

Table 2. IMAK values when pattern and knowledge have similar confidence level

Knowledge Certainty → ↓ Pattern is ...	Triviality	Standard knowledge	Hypothesis
more general	low	medium	high
identic	none	low	medium
more specific	none	low	medium

Table 3. IMAK values when a pattern have the best confidence level

Knowledge Certainty → ↓ Pattern is ...	Triviality	Standard knowledge	Hypothesis
more general	medium	high	high
identic	none	low	medium
more specific	none	medium	high

- medium: new information about a standard knowledge/confirmation of a hypothesis
- high: new information about a hypothesis

Tables 2, 3 and 4 show *IMAK* value according generalization relationship between a pattern and a knowledge, certainty of the knowledge and comparison of confidence level between pattern and knowledge.

Let's consider the knowledge rule 1, and the two following extracted rules:

EXTRACTED RULE 1

Objective='To be paid' \wedge Allowance='Housing Allowance' \rightarrow Contact='At the agency' [confidence=20%]

EXTRACTED RULE 2

Objective='To be paid' \wedge Allowance='Housing Allowance' \wedge Distance='Less Than30km' \rightarrow Contact='At the agency' [confidence=95%]

Rule 1 is a generalisation of the knowledge (see Section 1). But its confidence is lower than knowledge confidence level. Consequently *IMAK* value is "low" since the knowledge certainty is "hypothesis" (ref table 4 column 3 line 1). Rule 2 is also a generalisation of the knowledge. Its confidence is better than than knowledge confidence level. Consequently *IMAK* value is "high" since the knowledge certainty is "hypothesis" (ref table 3 column 3 line 1).

Table 4. IMAK values when knowledge have the best confidence level

Knowledge Certainty → ↓ Pattern is ...	Triviality	Standard knowledge	Hypothesis
more general	none	none	low
identic	none	low	medium
more specific	none	none	low

Now let's consider the rule:

EXTRACTED RULE 3

Objective='To be paid' \wedge Allowance='Student Housing Allowance' \wedge Distance='0km' \rightarrow Contact='At the agency' [confidence=75%]

Rule 3 is more specific than knowledge and its confidence is similar. Consequently *IMAK* value is "medium" since the knowledge certainty is "hypothesis" (ref table 2 column 3 line 3).

7 Experiments Results

Our approach is based on a toolset, called KEOPS, which allows to manage data preparation process, mining tasks and visualization step. KEOPS main feature is to use expert's prior knowledge all along the data mining process. Furthermore, KEOPS simplifies several complex tasks during the knowledge extraction process. We applied this approach on data of the 'family' branch of the french national health care system.

In order to evaluate our results we compare extracted models to prior knowledge according to their support, confidence and lift values. That's why we define *confidence gain*, *support gain* and *lift gain* in order to visualize statistical information on figures 6, 7 and 8 :

Definition 1 (Measure Gain). *Let R be an extracted rules and let C be a knowledge. We call a measure gain, the difference between the measure evaluation on R and C :*

$$MeasureGain(R, C) = measure(R) - measure(C)$$

IMAK measure allows us to select the most interesting extracted rules according knowledge. Figure 6 shows relative confidence of these rules. On X-axis there are 68 knowledge rules expressed by experts and on Y-axis, for each rule, there is a vertical bar where :

- the upper point shows relative confidence maximum value for rules compared with the knowledge
- the medium point shows relative confidence mean value of all rules compared with the knowledge
- the bottom point shows relative confidence minimum value for rules compared with the knowledge

We may observe that generally for each knowledge there exists an extracted rule with a better confidence value. Furthermore, extracted rules confidence mean value is often better than knowledge confidence value.

Figure 7 shows relative support of the most interesting extracted rules according *IMAK* measure. Legend of figure 7 is similar to figure 6 one. We may observe that generally relative support value is lower than 0. Consequently, extracted

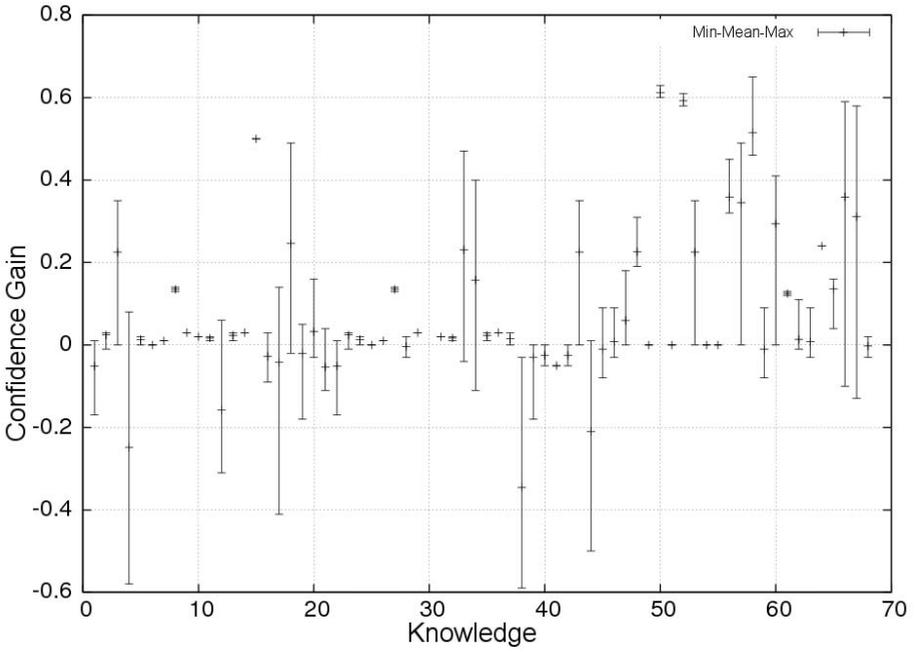


Fig. 6. Relative confidence between knowledge and associated extracted rules

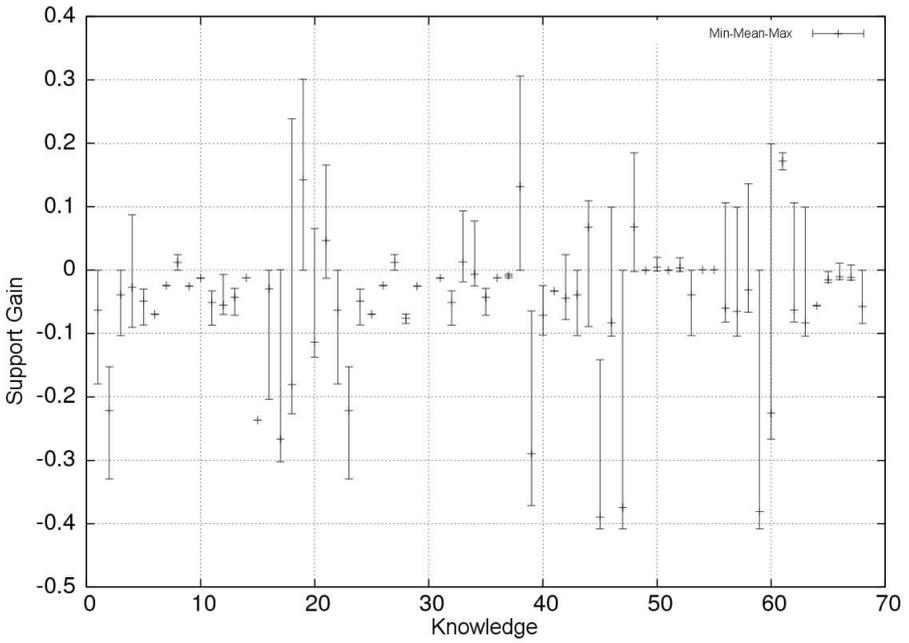


Fig. 7. Relative support between knowledge and associated extracted rules

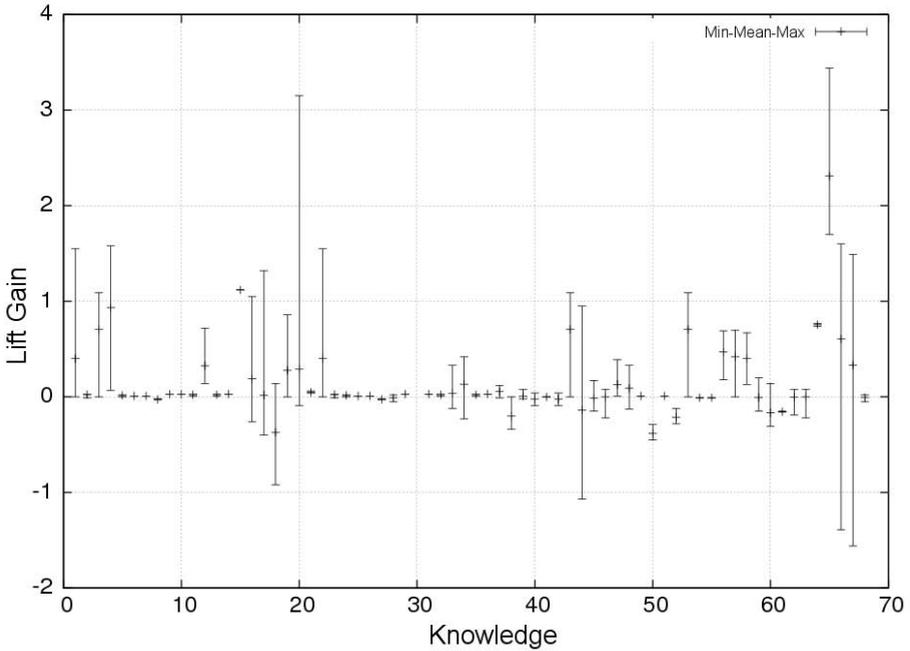


Fig. 8. Niveaux de lift relatif entre les connaissances et leurs règles associées

rules support is lower than knowledge support: these rules are infrequent cases. While some of them have also good confidence and lift, IMAK method catches some rare cases.

Figure 7 shows relative lift of the most interesting extracted rules according IMAK measure. Legend of figure 6 is similar to figure 6 one. We may observe that lift value of knowledge and extracted rules is similar. It's important to notice that expert knowledge has always a lift value greater than 1.

We presented a method which allows to select interesting rules according prior knowledge. Moreover we demonstrate that we extract rare rules (low support values) with good confidence and lift value. Consequently, our approach optimizes statistic criteria and provides some new and interesting knowledge.

8 Conclusion

We gave a global presentation of the new methodology ExCIS for the integration of prior knowledge in a data mining process. This paper shows how a Conceptual Information System (CIS) can improve data-mining results. We presented CIS ontological structures, and we discussed choices for identifying ontology concepts and relations by analyzing existing operational data. Finally we presented IMAK, an interestingness measure, which evaluate an extracted pattern according to user knowledge. In further works, we'll improve IMAK with distance measure between pattern and knowledge, and we'll add to ExCIS approach mechanisms

in order to generalize patterns (process 7 figure 1) before comparison with knowledge and to browse results after this comparison (process 9 figure 1).

References

1. Bard, J.B., Rhee, S.Y.: Ontologies in Biology: Design, Applications and Future Challenges. *Nature Review Genetics* 5(3), 213–222 (2004)
2. Chapman, P., al.: CRISP-DM - Step by step data mining guide Cross Industry Standard Process for Data Mining, <http://www.crisp-dm.org/>
3. Dai, H., Mobasher, B.: Using Ontologies to Discover Domain-level Web Usage Profiles. In: Proceedings 2nd ECML/PKDD Semantic Web Mining workshop (August 2002)
4. Guarino, N.: Formal Ontology and Information Systems. In: Proceedings of FOIS'98, pp. 3–15 (June 1998)
5. Hilderman, R.J., Hamilton, H.J.: Evaluation of Interestingness Measures for Ranking Discovered Knowledge. In: Cheung, D., Williams, G.J., Li, Q. (eds.) PAKDD 2001. LNCS (LNAI), vol. 2035, pp. 247–259. Springer, Heidelberg (2001)
6. Johannesson, P.: A Method for Transforming Relational Schemas into Conceptual Schemas. In: Rusinkiewicz, M. (ed.) Proceedings 10th ICDE conference, pp. 115–122. IEEE Press, New York (1994)
7. Kashyap, V.: Design and Creation of Ontologies for Environmental Information Retrieval. In: Proceedings 12th workshop on Knowledge Acquisition, Modelling and Management (October 1999)
8. Liu, B., Hsu, W., Chen, S.: Using General Impressions to Analyze Discovered Classification Rules. In: Proceedings 3rd KDD conference, pp. 31–36 (August 1997)
9. Liu, B., Hsu, W., Mun, L.-F., Lee, H.-Y.: Finding Interesting Patterns using User Expectations. *Knowledge and Data Engineering* 11(6), 817–832 (1999)
10. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. *ACM Comput. Surv.* 38(3) (2006)
11. McGarry, K.: A Survey of Interestingness Measures for Knowledge Discovery. *The knowledge engineering review*, 1–24 (2005)
12. Pasquier, N., Taouil, R., Bastide, Y., Stumme, G., Lakhal, L.: Generating a Condensed Representation for Association Rules. *Journal of Intelligent Information Systems*. In: Kerschberg, L., Ras, Z., Zemankova, M. (eds.) Kluwer Academic Publishers
13. Piatetsky-Shapiro, G., Matheus, C.: The Interestingness of Deviations. In: Proceedings of the AAAI-94 workshop on Knowledge Discovery in Databases (1994)
14. Silberschatz, A., Tuzhilin, A.: On Subjective Measures of Interestingness in Knowledge Discovery. In: Proceedings 1st KDD conference, pp. 275–281 (August 1995)
15. Silberschatz, A., Tuzhilin, A.: What Makes Patterns Interesting in Knowledge Discovery Systems. *IEEE Transaction On Knowledge And Data Engineering* 8(6), 970–974 (1996)
16. Stevens, R., Goble, C.A., Bechhofer, S.: Ontology-based Knowledge Representation for Bioinformatics. *Brief Bioinformatics* 1(4), 398–414 (2000)
17. Stojanovic, L., Stojanovic, N., Volz, R.: Migrating Data-intensive Web Sites into the Semantic Web. In: Proceedings 17th ACM Symposium on Applied Computing, pp. 1100–1107. ACM Press, New York (2002)
18. Stumme, G.: Conceptual On-Line Analytical Processing. In: Tanaka, K., Ghandeharizadeh, S., Kambayashi, Y. (eds.) *Information Organization and Databases*, ch. 14, pp. 191–203. Kluwer Academic Publishers, Dordrecht (2000)
19. Tiffin, N., Kelso, J.F., Powell, A.R., Pan, H., Bajic, V.B., Hide, W.A.: Integration of Text- and Data-Mining using Ontologies Successfully Selects Disease Gene Candidates. *Nucleic Acids Research* 33(5), 1544–1552 (2005)