Hamido Fujita · Moonis Ali
Ali Selamat · Jun Sasaki
Masaki Kurematsu (Eds.)

Trends in Applied Knowledge-Based Systems and Data Science

29th International Conference
on Industrial Engineering and Other Applications
of Applied Intelligent Systems, IEA/AIE 2016
Morioka, Japan, August 2–4, 2016
Proceedings

Springer

Lecture Notes in Artificial Intelligence 9799

Subseries of Lecture Notes in Computer Science

Editors
Hamido Fujita
Iwate Prefectural University
Iwate
Japan

Jun Sasaki
Iwate Prefectural University
Iwate
Japan

Moonis Ali
Department of Computer Science
Texas State University
San Marcos, Texas
USA

Masaki Kurematsu
Iwate Prefectural University
Iwate
Japan

Ali Selamat
Universiti Teknologi Malaysis (UTM)
Bahru
Malaysia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-42006-6 ISBN 978-3-319-42007-3 (eBook)
DOI 10.1007/978-3-319-42007-3

Library of Congress Control Number: 2016943422

LNCS Sublibrary: SL7 – Artificial Intelligence

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Intelligent Systems in Modeling Phase of Information Mining Development Process

Sebastian Martins[1,3], Patricia Pesado[2], and Ramón García-Martínez[3(✉)]

[1] PhD Program on Computer Science, National University of La Plata, La Plata, Argentina
smartins089@gmail.com
[2] III-LIDI. Computer Sc School, National University of La Plata – CIC Bs as, La Plata, Argentina
ppesado@lidi.info.unlp.edu.ar
[3] Information Systems Research Group, National University of Lanus, Lanús, Argentina
rgm1960@yahoo.com

Abstract. The Information Mining Engineering (IME) understands in processes, methodologies, tasks and techniques used to: organize, control and manage the task of finding knowledge patterns in information bases. A relevant task is selecting the data mining algorithms to use, which it is left to the expertise of the information mining engineer, developing it in a non-structured way. In this paper we propose an Information Mining Project Development Process Model (D-MoProPEI) which provides an integrated view in the selection of Information Mining Processes Based on Intelligent Systems (IMPbIS) within the Modeling Phase of the proposed Process Model through a Systematic Deriving Methodology.

1 Introduction

Information Mining is defined as the sub-discipline of information systems which provides to the Business Intelligence [1, 2] the tools to transform information into knowledge [3, 4]. Information mining based on intelligent systems [5] refers especially to the application of intelligent systems-based methods to discover and enumerate existing patterns in information. Intelligent systems-based methods allow retrieving results about the analysis of information bases that conventional methods fail to achieve [6], such as: TDIDT algorithms (Top Down Induction Decision Trees), Self-Organizing Maps (SOM) and Bayesian networks. TDIDT algorithms allow the development of symbolic descriptions of the data to distinguish between different classes [7]. Self-Organizing Maps can be applied in the construction of information clusters. They have the advantage of being tolerant to noise and the ability to extend the generalization when needing the manipulation of new data [8]. Bayesian networks can be applied to identify discriminative attributes in large information bases and detect behavior patterns in the analysis of temporal series [9].

H. Fujita et al. (Eds.): IEA/AIE 2016, LNAI 9799, pp. 3–15, 2016.
DOI: 10.1007/978-3-319-42007-3_1

An Information Mining Process is defined as a group of logically related tasks [10], which from a set of information with a degree of value for the organization obtains knowledge pieces that generalize the previous information. The Information Mining Engineering understands in processes, methodologies, tasks and techniques used to: organize, control and manage the task of finding knowledge patterns in information bases [11]. A Process Model for Information Mining Engineering is defined as set of phases, where each phase comprehends a set of tasks and each task has well defined inputs and outputs. Each phase of the Process Model is oriented to obtain a partial product. The whole Process Model is oriented to obtain a final product. This final product is derived from the partial products of phases. In case of Information Mining Engineering, the final product is a set of knowledge pieces [12]. There are several process models for information mining engineering: KDD [13], CRISP-DM [14], KDD + Std. IEEE 1074 [16]. Failure rate of this kind of projects is over 60 % [15].

In previous work [12] we have pointed out that the process models mentioned above are deficient in structuring the general tasks, leading to unnecessary iterations that cause delays and increasing costs. Applying data preparation tasks before determining the tools and algorithms, leads to a possible need of looping back to the stage of data preparation, when it is possible to identify those elements in advance. Furthermore, selection of data mining algorithms to use is left to the expertise of the information mining engineers.

In this context, this paper proposes a possible solution to reduce the need of human expertise in the selection of information mining algorithms. To deal with this problem, we present in Sect. 2 an Information Mining Project Development Process Model (D-MoProPEI). This Process Model has a phase called "Modeling" in which the Information Mining Processes to be used are selected. The Information Mining Processes based on intelligent systems used in the "Modeling" phase are presented in Sect. 3. A process to derive the Information Mining Processes in "Modeling" phase is presented in Sect. 4. Section 5 provides a concept proof and Sect. 5 presents the conclusions of our work.

2 Information Mining Project Development Process Model (D-MoProPEI)

MoProPEI is composed by two sub-processes: *Development*, focus on the technical activities and *Management* which covers those activities oriented to control and organize the process. Each sub-process is integrated by a set of phases, which group several activities according to their goals. Management sub-process contains five phases: Initiation, Project Planning, Support, Quality and Control and Closure. Development is integrated by six phases: Business Understanding, Data Understanding, Modeling, Data Preparation, Implementation and Assessment and Presentation.

The execution of both sub-process is not sequential, they are applied in parallel. The activities of the Management sub-process provide support to those activities associated with the construction of the final product (pieces of knowledge).

Development sub-process covers those activities associated with the identification of relevant and novel patterns, as well as the analysis and comprehension of the result obtained to generate interesting and valuable pieces of knowledge that bring added value to the organization. Figure 1 shows the internal dependencies between the development activities. Phases and their activities, inputs and outputs (on the left and right side respectively) are presented in the same horizontal line of the picture.

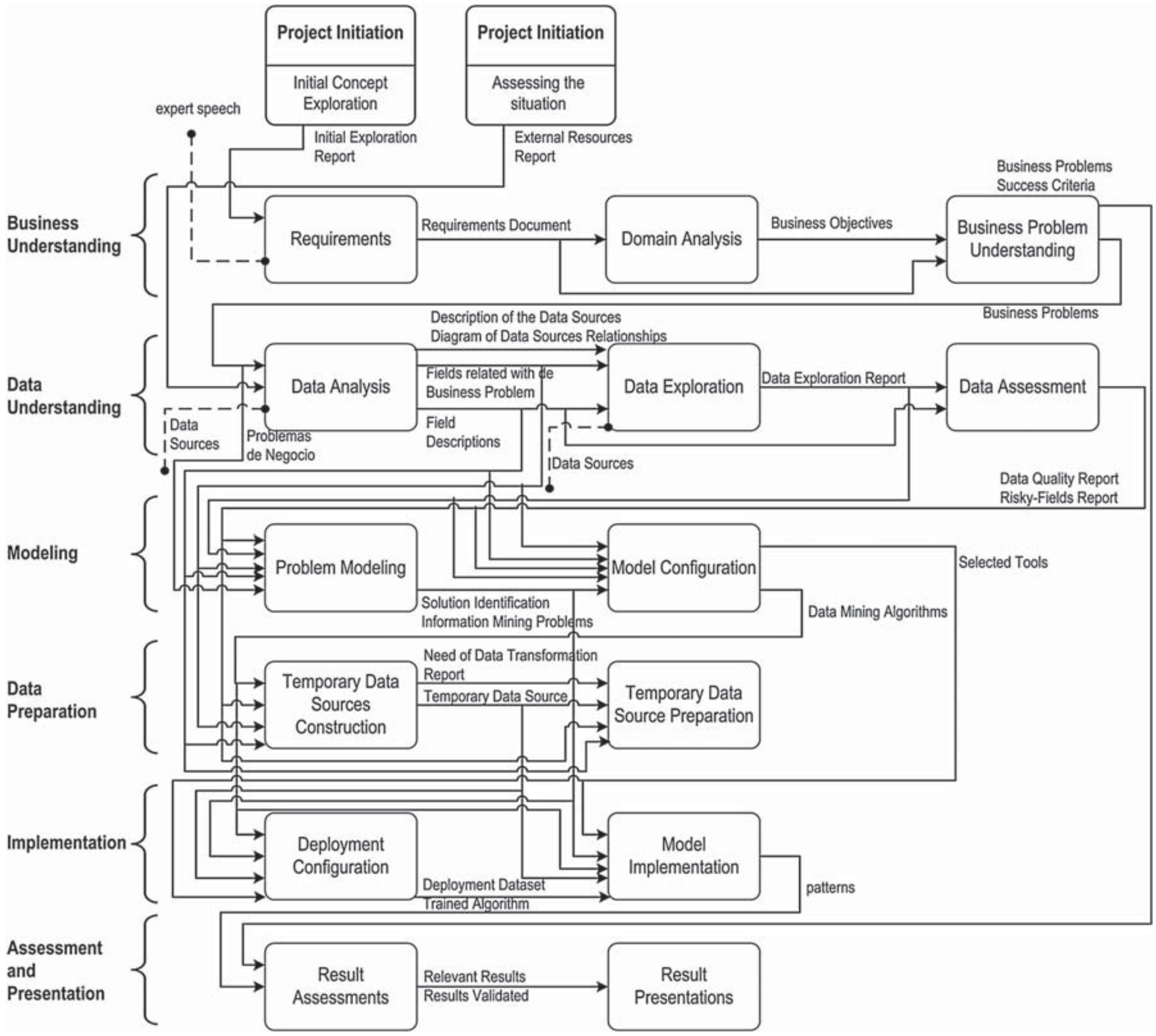

Fig. 1. Development sub-process structure

Phase 1 - Business Understanding, consist of 3 activities: *Requirements* whose inputs are expert speech (external) and Initial Exploration Report created as output in the Initial Concept Exploration phase from the Management sub-process and produces as output the Requirements Document; *Domain Analysis* whose input is the output of the previous activity and identifies the Business Objectives; and *Business Problem Understanding* whose inputs are the outputs of the two previous activities and produces the Business Problems and their Success Criteria.

Phase 2 - Data Understanding, compound by 3 activities: *Data Analysis*, whose inputs are Business Problem, External Resources Report (from the Management sub-process) and Data Sources and its outputs are Description of the Data Sources, Diagram of Data Sources Relationships, Fields related with de Business Problem, and Field Descriptions; *Data Exploration* whose inputs are the elements generated in the first activity of the phase and Data Sources and generates as output Data Exploration Report; and *Data Assessment* whose inputs are Fields related with de Business Problem and Data Exploration Report and produces as outputs Data Quality Report and Risky-Fields Report.

Phase 3 – Modeling, integrated by 2 activities: *Problem Modeling,* whose inputs are the outputs of Data Assessment, Data Exploration Report, Fields related with de Business Problem, Field Descriptions and Business Problem, and generates as output Solution Identification and Information Mining Problems; and *Model Configuration* whose inputs are Data Quality Report, Risky-Fields Report, Data Exploration Report, Fields related with de Business Problem, Field Descriptions and the outputs of the previous activity, and its outputs are Selected Tools and Data Mining Algorithms.

Phase 4 – Data Preparation, conformed by 2 activities: *Temporary Data Sources Construction* whose inputs are Data Mining Algorithms, Data Quality and Risky-Fields Reports, Fields related with de Business Problem, Field Descriptions and Business Problem, and generates as output Temporary Data Source and Need of Data Transformation Report; and *Temporary Data Source Preparation* whose inputs are the outputs of the previous activity, Data Quality and Risky-Fields Reports and transforms the Temporary Data Source.

Phase 5 – Implementation, integrated by 2 activities: *Deployment Configuration* whose inputs are Selected Tools and Data Mining Algorithms, Solution Identification and Information Mining Problems and Temporary Data Source and its outputs are Deployment Dataset and Trained Algorithm; and *Model Implementation* whose inputs are the outputs of the previous activity, Solution Identification and Information Mining Problems, Selected Tools, Data Mining Algorithms and Temporary Data Source and it generates as output the knowledge patterns.

Phase 6 – Assessment and Presentation, conformed by 2 activities: *Result Assessments* whose inputs are the results produced from the implementation model and Business Problems Success Criteria and its outputs are Relevant Results and Results Validated; and *Result Presentations* whose inputs are the outputs of the previous activity and produces as output the Final Project Report and the results are presented to the client. As result of implementing the assessment activity, belonging to the current phase, new questions about the considered problem may raise as well as the need of analyzing it in depth, being necessary to iterate over the previous stages.

Please note that activities shown Fig. 1 provide an abstraction layer which groups sets of tasks according to their goals, presenting only the results that are dependents of other activities.

3 Information Mining Processes Based on Intelligent Systems (IMPbIS)

In this section, the information-mining processes based on Intelligent Systems proposed in [12] are presented: simple processes (Sect. 3.1) and compound processes (Sect. 3.2).

3.1 Simple Processes

There have been defined three types of simple processes: [a] the **process for discovery of behavioral rules** which applies when it is necessary to identify which are the conditions to get a specific outcome in the problem domain. The following problems are examples among others that require this process: identification of the characteristics for the most visited commercial office by customers, identification of the factors that increase the sales of a specific product, definition of the characteristics or traits of customers with high degree of brand loyalty, definition of demographic and psychographic attributes that distinguish the visitors to a website. For the discovery of behavioral rules from classes attributes in a problem domain that represents the available information base, it is proposed the usage of TDIDT induction algorithms [16] to discover the rules of behavior for each class attribute; [b] the **process of discovery of groups** which applies when it is necessary to identify a partition on the available information base of the problem domain. The following problems are examples among others that require this process: identification of the customers segments for banks and financial institutions, identification of type of calls of customer in telecommunications companies, identification of social groups with the same characteristics, identification of students groups with homogeneous characteristics. For the discovery of groups [17, 18] in information bases of the problem domain for which there is no available "a priori" criteria for grouping, it is proposed the usage of Kohonen's Self-Organizing Maps or SOM [19–21]. The use of this technology intends to find if there is any group that allows the generation of a representative partition for the problem domain which can be defined from available information bases; and [c] the **Process of Discovery of Significant Attributes** which applies when it is necessary to identify which are the factors with the highest incidence (or occurrence frequency) for a certain outcome of the problem. The following problems are examples among others that require this process: factors with incidence on the sales, distinctive features of customers with high degree of brand loyalty, key-attributes that characterize a product as marketable, key-features of visitors to a website. Bayesian Networks [22] allows seeing how variations in the values of attributes, impact on the variations in the value of class attribute. The use of this process seeks to identify whether there is any interdependence among the attributes that model the problem domain which is represented by the available information base.

3.2 Compound Processes

There have been also defined two complex processes: [a] the **process of discovery of group membership rules** applies when it is necessary to identify which are the conditions of membership to each of the classes of an unknown partition "a priori", but existing in the available information bases of the problem domain. The following problems are examples among others that require this process: types of customer's profiles and the characterization of each type, distribution and structure of data of a web site, segmentation by age of students and the behavior of each segment, classes of telephone calls in a region and the characterization of each class. For running the process of discovery of group-membership rules it is proposed to use of self-organizing maps (SOM) for finding groups and once the groups are identified, the usage of induction algorithms (TDIDT) for defining each group behavior rules [21, 23, 24]; and [b] the **Process of Weighting of Behavior or Group-membership Rules** that first required that all sources of information (databases, files, others) are identified, and then they are integrated together as a single source of information which will be called integrated data base. Based on the integrated data base, the Self-Organizing Maps (SOM) is applied. As a result of the application of SOM, a partition of the set of records in different groups is achieved which is called identified groups. The associated files for each identified group are generated. This set of files is called "ordered groups". The "group" attribute of each ordered group is identified as the class attribute of that group, establishing it in a file with the identified class attribute (GR). Then TDIDT is applied to the class attribute of each "GR group" and the set of rules that define the behavior of each group is achieved.

4 Deriving IMPbIS in Modeling Phase of Proposed Information Mining Development Process

We have developed a methodology [25] (shown in Fig. 2) to derive the processes of information mining from frames and semantic nets. The methodology has three phases: "Analysis of Business Domain", "Analysis of the Problem of Information Mining", and "Analysis of the Process of Information Mining".

The phase "Analysis of Business Domain" develops three tasks: "Identification of the Elements and Structure of the Business Domain", "Identification of Relationships

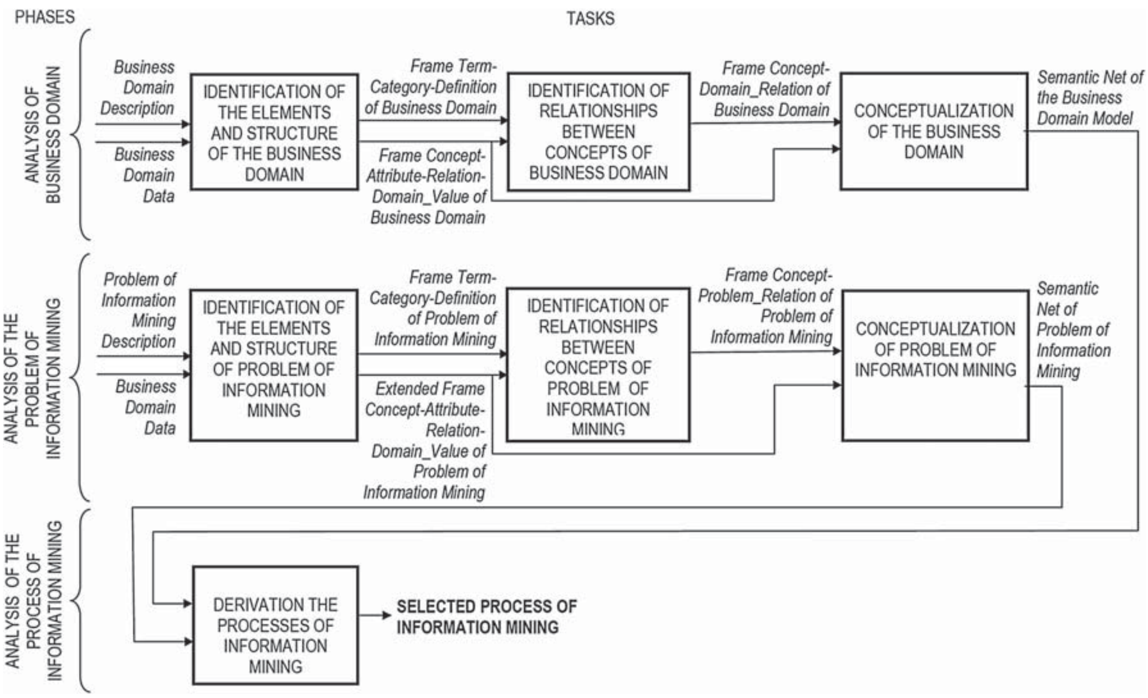

Fig. 2. Methodology to derive the processes of information mining

between Concepts of Business Domain", and "Conceptualization of the Business Domain". The task "Identification of the Elements and Structure of the Business Domain" has as input the "Business Domain Description" and the "Business Domain Data"; and produces as output the "Frame Term-Category-Definition of Business Domain" and the "Frame Concept-Attribute-Relation-Domain_Value of Business Domain". The task "Identification of Relationships between Concepts of Business Domain" has as input the "Frame Term-Category-Definition of Business Domain" and the "Frame Concept-Attribute-Relation-Domain_Value of Business Domain"; and produces as output the "Frame Concept-Domain_Relation of Business Domain". The task "Conceptualization of the Business Domain" has as input "Frame Concept-Domain_Relation of Business Domain" and the "Frame Concept-Attribute-Relation-Domain_Value of Business Domain"; and produces as output the "Semantic Net of the Business Domain Model".

The phase "Analysis of the Problem of Information Mining" develops three tasks: "Identification of the Elements and Structure of Problem of Information Mining", "Identification of Relationships between Concepts of Problem of Information Mining" and "Conceptualization of Problem of Information Mining". The task "Identification of the Elements and Structure of Problem of Information Mining" has as input the "Problem of Information Mining Description" and the "Business Domain Data"; and produces as output the "Frame Term-Category-Definition of Problem of Information Mining" and the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining". The task "Identification of Relationships between Concepts of Problem of Information Mining" has as input the "Frame Term-Category-Definition of Problem of Information Mining" and the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining"; and produces as output the "Frame Concept- Problem_Relation of Problem of Information Mining". The task "Conceptualization of Problem of Information Mining" has as input the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining" and the "Frame Concept- Problem_Relation of Problem of Information Mining"; and produces as output the "Semantic Net of Problem of Information Mining".

The phase "Analysis of the Process of Information Mining" develops one task: "Derivation the Processes of Information Mining" which has as input the "Semantic Net of the Business Domain Model" and the "Semantic Net of Problem of Information Mining"; and produces the "Selected Process of Information Mining".

5 Concept Proof

In this section, it presents a case of study solved by applying MoProPEI. In the following paragraphs, we present some of the documents generated by implementing the tasks that compound each phase of the Development sub-process.

Phase 1 - Business Understanding. Interactions with the experts of the different business areas are performed to comprehend the characteristics, goals and business problems of the organization. Identifying requirements, resources (either information documents/sources, as person experts in the various areas of interest), describing the domain terminology, goals/problems, risks/contingencies and success criteria. In the following paragraph the most relevant contents obtained by interviews with the experts are presented:

> "*…The purpose of the project is to facilitate the appropriation of knowledge in college/university education in massive contexts, in this particular case, focusing on the subject informatics. Providing information for proper design of public policies in college/university education, contributing to a better ownership of knowledge. In this direction, a relevant dimension is associated with student characteristics, the main actor in this complex scenario. It is desired to understand the features that provide clues about the difficulties of a student in the fulfillment of the career curricula, being able to identify and act, in an early stage, providing to the student with tools which allow him/her to overcome as far as possible obstacles that may arise. It is highlighted the interest of comprehend how the socioeconomics features are related with the academic performance of the students in massive contexts…*"

In Table 1, the business problem is presented which is related to a Business Objective previously identified. Then, experts, risk/contingencies and success criteria related with that problem are identified.

Table 1. Business problem

<table>
<tr><td colspan="3">Superior Educational Project – Cordoba National University</td><td>Project ID:</td><td>ES.UNC</td></tr>
<tr><td colspan="3" rowspan="2">Business Problem</td><td>Document ID:</td><td>D.1.3.1</td></tr>
<tr><td>Version:</td><td>1.0</td></tr>
<tr><td>ID</td><td>ID# Business Objectives</td><td>Problem Description</td><td colspan="2">Comments</td></tr>
<tr><td>PN.1</td><td>ON.1</td><td>Is it possible to find aspects from students to comprehend and identify in an early stage those who could present difficulties to fulfill the career curricula in the expected way?</td><td colspan="2"></td></tr>
<tr><td colspan="5">General Comments:</td></tr>
</table>

Phase 2 - Data Understanding. From the resources previously identified, a detailed analysis of the organization's information sources (digital and non-digital) is performed and then it proceeds to analyze the characteristics of the data in relation to their applicability to business problems. To accomplished that a detailed description of the data sources and existence fields is performed (making use of existent documents such as ER model), selecting in conjunction with the expert those relevant fields (existent or to be generated from existing fields) to solve the business problems. Furthermore, a quality and risk analysis is performed pointing out possible risky fields. In Table 2, we present the data dictionary (showing only the fields related with the problem) and list the set of relevant data, explaining how to create them (Table 3).

Phase 3 – Modeling. The aim of the phase is to identify the set of tools, techniques and data mining algorithms to find pieces of knowledge to support the decision making process associated with the business problem. To reach that goal, we define the information mining problem (a technical and detailed description of the business process) and apply the methodology to derive the processes of information mining (describe in Sect. 4) identifying the set of algorithms to implement. In Table 4, it is

Table 2. Fields description

Superior Educational Project – Cordoba National University			**Project ID:**	ES.UNC
Fields Description			**ID:**	D.2.1.3
			Version:	1.0
Data Source:	RM.1			
ID	**Field**	**Type**	**Description**	
4	courses Informatics	Integer	Year in which the student courses Informatics	
5	Year entered	Integer	Year that the student entered to the career	
6	Sex	Boolean	sex	
8	Country	String	Country of origin	
9	Province	String	Province of origin	
10	Department	String	Department of origin	
...	...	...	...	
17	Pay studies	String	The way the student pays his/her studies	
19	Last studies father group	String	Last studies reach by the father	
22	Last studies Mother group	String	Last studies reach by the mother	
24	Quantity of subjects coursed	Integer	How many subjects course the student the first half	
25	Approved informatics Date	Date	Approved informatics	
27	subjects approved	Integer	How many subjects approved	
General Comments:				

pointed out the Information Mining Problem (IMP), and in Fig. 3 we present the conceptualization model derived from IMP getting as result the **Process of Discovery of Group-Membership Rules**.

Phase 4 – Data Preparation. After identifying the problem, tools and algorithms to use, we have all the necessary information to understand the format requirements of each of the relevant fields. At this stage the database to be used to implement the algorithms is generated and the fields are formatted and cleaned, ready to knowledge extraction.

Phase 5 – Implementation. In this phase the registers that would be used in the information mining process are defined (training and/or testing dataset). Additionally, we determine the configuration and optimization strategy that will be adopted to obtain the best results (results' understandability and success rate), according to the client's objectives. At the end of the phase, the knowledge patterns are discovered.

Phase 6 – Assessment and Presentation. The aim of this phase is identifying and filtering those patterns which provide novel and interesting knowledge through verifying the results against the success criteria and then validating them against the area expert's opinion.

Once analyzed and covered the various aspects required by the customer, it proceeds to carry out and presents the report in which the results are transferred to the customer. For example, rules that describe students who present worse academic performance are identified: [a] one group that work and are delayed to attend informatics (experts explaining that the fact of working is a strong causal of delay, being able of correcting the situation through scholarships) and [b] one group which do not work (his income comes from their families) and are delayed to attend informatics

Table 3. Fields related with the business problem

Superior Educational Project – Cordoba National University		**Project ID:**	ES.UNC
Fields Related with the Business Problem		**Document ID:**	D.2.1.4
		Version:	1.0
Business Problem	PN.1		

#IDBD.#ID Field	To build	Name	Comments
RT.1.1	x	work	Generated from the **RM.1.17** variable, indicating 1 if one of the value of the multivalued field is "with their own work," and 0 otherwise.
RT.1.2	x	family	Generated from the **RM.1.17** variable, indicating 1 if one of the value of the multivalued field is "with family's support" and 0 otherwise.
RT.1.3	x	scholarship	Generated from the **RM.1.17** variable, indicating 1 if one of the value of the multivalued field is "with scholarship" and 0 otherwise.
RM.1.19		Last studies father group	
RM.1.22		Last studies Mother group	
RT.1.4	x	Approved Informatics before year	Indicates whether the student passed the subject until a year of having attended it (1) or not (0), from the **RM.1.25** variable.

RT.1.5	x	Fulfill career curricula	If **RS.1.1**> 5 => 0,
			else:
			If **RS.1.1** < 5:
			If (**RS.1.1** * 9 – 3 > **RM.1.27**) => 3,
			else => 4
			If ((**RS.1.1** – 1) * 9 +7 > **RM.1.27**) => 1
			else => 2

RT.1.6	x	Delay in Studying	Scale indicating the time it takes to study informatics, from the difference between the **RM.1.5** and **RM.1.4** variables: If attend the same year => 0, If attends within 2 years later => 1, else 2.
RM.1.6		Sex	
RT.1.7	x	Initial rhythm	Scale indicating the student's progress in the first half of the career, taking into account the **RM.1.24** and **RT.1.6** variable: If assist to 5 subjects (including informatics) => 3, If assist to 3 or more subjects (including informatics) => 2, If assist to 2 or 1 subjects (including informatics) => 1, else 0.
RT.1.8	x	Argentine	if **RM.1.8** is "Argentine" => 1,else 0
RT.1.9	x	Cordoba	if **RM.1.9** is "Cordoba" => 1,else 0
RT.1.10	x	Main Town	if **RM.1.10** is "Main Town" => 1, else 0
General Comments: Fields RT.1.[8,9,10] are dependent variables (**location**)			

Table 4. Information Mining Problem

Superior Educational Project – Cordoba National University			**Project ID:**	ES.UNC
Business Problem			**Document ID:**	D.3.1.1
			Version:	1.0
ID	**#ID PN**	**Problem Description**	**Comments**	
PEI.1	PN.1	How academic student response is characterized, in relation to their similarity among socio-economic variables?		
General Comments:				

(experts pointed out that their failure of the career curricula may not being related with socioeconomics aspects, but with their personal interest, being harder to intervene).

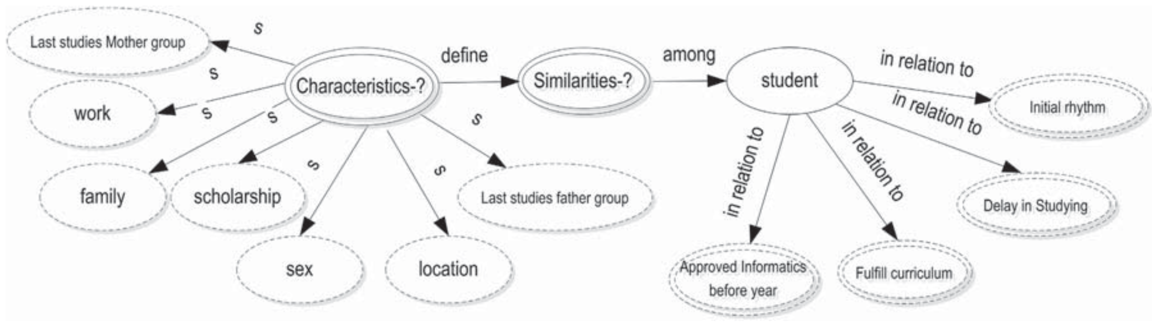

Fig. 3. Information mining problem conceptualization

6 Conclusions

Intelligent Systems has been the corner stone of data mining from the early stages. Data mining has evolved to Information Mining, and in the last decade academic movements towards an information mining engineering are increasing.

There has been pointed a weakness of current process models for information mining engineering related to the selection of information mining algorithm. This weakness lies in dependency on expertise of information mining engineer to select de information mining algorithm. Additionally, an unnecessary iteration that cause delays and increases the project' costs is highlighted.

In order to cope with this problem we propose an Information Mining Project Development Process Model (D-MoProPEI). This Process Model has a phase called "Modeling" in which the Information Mining Processes based on Intelligent Systems (IMPbIS) to be used are selected by a systematic deriving methodology and restructure the sub-process, switching the Modeling and Data Preparation phases.

We are working in an Information Mining Project Management Process Model (M-MoProPEI). Next step is to integrate both process Models in a single one to study its effectiveness to carry out information mining projects.

Besides, there are in the literature many papers and results about the convenience of the usage of certain data mining algorithms compared to others, but it is rarely raised the information mining process associated to these algorithms or the convenience of the usage of one algorithm compared to other for that process. In this context, it is an interesting open problem the identification of the relationship between the data mining algorithm and the process of information mining.

Acknowledgments. The research reported in this paper was partially funded by Project ME-SPU-PROMINF-UNLa-2015-2017 of the Argentinean Ministry of Education and Project UNLa-33A205 of the Secretary of Science and Technology of National University of Lanus (Argentina).

References

1. Thomsen, E.: BI's promised land. Intell. Enterp. **6**(4), 21–25 (2003)
2. Negash, S., Gray, P.: Business intelligence. In: Bursteiny, F., Holsapple, C. (eds.) Handbook on Decision Support Systems 2. IHIS, pp. 175–193. Springer, Heidelberg (2008)
3. Langseth, J., Vivatrat, N.: Why proactive business intelligence is a hallmark of the real-time enterprise: outward bound. Intell. Enterp. **5**(18), 34–41 (2003)
4. Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., Shan, M.: Business process intelligence. Comput. Ind. **53**(3), 321–343 (2004)
5. Michalski, R., Bratko, I., Kubat, M.: Machine Learning and Data Mining, Methods and Applications. Wiley, New York (1998)
6. Michalski, R.: A theory and methodology of inductive learning. Artif. Intell. **20**, 111–161 (1983)
7. Quinlan, J.: Learning logic definitions from relations. Mach. Learn. **5**, 239–266 (1990)
8. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (1995)
9. Heckerman, D., Chickering, M., Geiger, D.: Learning bayesian networks, the combination of knowledge and statistical data. Mach. Learn. **20**, 197–243 (1995)
10. García-Martínez, R., Britos, P., Rodríguez, D.: Information Mining Processes Based on Intelligent Systems. In: Ali, M., Bosse, T., Hindriks, K.V., Hoogendoorn, M., Jonker, C.M., Treur, Jan (eds.) IEA/AIE 2013. LNCS, vol. 7906, pp. 402–410. Springer, Heidelberg (2013)
11. García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell. J.: Towards an information mining engineering. In: Software Engineering, Methods, Modeling and Teaching, pp. 83–99. Medellín University Press. ISBN: 978-958-8692-32-6 (2011)
12. Martins, S., Pesado, P., García-Martínez, R. (2014). Process Mining Proposal for Information Mining Engineering: MoProPEI (in spanish). Latin-American Journal of Software Engineering, 2(5): 313–332. http://dx.doi.org/10.18294/relais.2014.313-332. ISSN: 2314-2642
13. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**(3), 37–54 (1996)
14. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step by step BI guide. Edited by SPSS (2000)
15. Marbán, Ó., Mariscal, G., Menasalvas, E., Segovia, J.: An engineering approach to data mining projects. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) IDEAL 2007. LNCS, vol. 4881, pp. 578–588. Springer, Heidelberg (2007)
16. Britos, P., Jiménez Rey, E., García-Martínez, E.: Work in progress: programming misunderstandings discovering process based on intelligent data mining tools. In: Proceedings 38th ASEE/IEEE Frontiers in Education Conference (2008)
17. Kaufmann, L., Rousseeuw, P.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York (1990)
18. Grabmeier, J., Rudolph, A.: Techniques of Cluster Algorithms in Data Mining. Data Min. Knowl. Disc. **6**(4), 303–360 (2002)
19. Ferrero, P., Britos, P., García-Martínez, R.: Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks. In: Debenham, J. (ed.) IEA/AIE 2008. IFIP, vol. 218, pp. 1–10. Springer, Boston (2008)
20. Britos, P., Cataldi, Z., Sierra, E., García-Martínez, R.: Pedagogical protocols selection automatic assistance. In: Nguyen, N.T., Borzemski, L., Grzech, A., Ali, M. (eds.) IEA/AIE 2008. LNCS (LNAI), vol. 5027, pp. 331–336. Springer, Heidelberg (2008)

21. Britos, P., Grosser, H., Rodríguez, D., García-Martínez, R.: Detecting unusual changes of users consumption. In: Bramer, M. (ed.) IEA/AIE 2008. IFIP, vol. 276, pp. 297–306. Springer, Boston (2008)
22. Britos, P., Felgaer, P., García-Martínez, R.: Bayesian networks optimization based on induction learning techniques. In: Bramer, M. (ed.) IEA/AIE 2008. IFIP, vol. 276, pp. 439–443. Springer, Boston (2008)
23. Britos, P., Abasolo, M., García-Martínez, R., Perales, F.: Identification of MPEG-4 patterns in human faces using data mining techniques. In: Proceedings 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2005, pp. 9–10 (2005)
24. Cogliati, M., Britos, P., García-Martínez, R.: Patterns in temporal series of meteorological variables using SOM & TDIDT. In: Bramer, M. (ed.) AITP. IFIP, vol. 217, pp. 305–314. Springer, Boston (2006)
25. Martins, S., Rodríguez, D., García-Martínez, R.: Deriving processes of information mining based on semantic nets and frames. In: Ali, M., Pan, J.-S., Chen, S.-M., Horng, M.-F. (eds.) IEA/AIE 2014, Part II. LNCS, vol. 8482, pp. 150–159. Springer, Heidelberg (2014)