

# Framework for Decisional Business Modeling and Requirements Modeling in Data Mining Projects

José Gallardo<sup>1</sup>, Gloria Giacaman<sup>1</sup>, Claudio Meneses<sup>1</sup>, and Óscar Marbán<sup>2</sup>

<sup>1</sup> Depto. Ingeniería de Sistemas y Computación, Universidad Católica del Norte.

Av. Angamos 0610, Antofagasta, Chile

{jgallardo,ggiacama,cmeneses}@ucn.cl

<sup>2</sup> Facultad de Informática, Universidad Politécnica de Madrid.

Campus de Montegancedo s/n, Boadilla del Monte, Madrid, España

omarban@fi.upm.es

**Abstract.** This article presents a framework oriented to guide the methodological construction of the decision-making process model of an organization, and the subsequent derivation of a requirements model for the development of Data Mining projects. The decisional model, is defined as a conceptual tool made up of a set of related concepts which represent the decision-making processes of an organization. The importance of a decision-making model is that with a clear definition of the requirements which arise from the organization's goals and which generate the decisional processes; the development is permitted of a Data Mining project which is effectively aligned with the achievement of the organization strategic goals.

**Keywords:** Data Mining, Business Modeling, Goals Strategies, Decision-Making, Requirements Engineering.

## 1 Introduction

Many publications exist in the field of business process modeling, with varying aims such as an improved understanding of the organization, analysis and innovation, or reengineering ([16], [9], [18], [5]). Much research has been carried out in the area of elicitation and requirements representation for the design and implementation of software systems ([3], [4], [6]), and there is broad consensus that the requirements capture phase is one of the critical phases of the software production process.

However, although it is clear that a key factor to optimize the requirements elicitation process, is an adequate knowledge of the organization [8], less research has been carried out ([13],[17]) integrating the organizational model with Requirements Engineering. Moreover, in the case of Data Mining Systems, a major drawback is that the decisional processes at a strategic level of an organization are unstructured and therefore complex to model.

Furthermore, there are many studies ([12], [11], [2]) which report a significant amount of Data Mining projects failing due to aspects clearly related to an inadequate requirements specification phase [7]. In this paper a methodological process is presented, intended to guide the construction of a business decision-making process

model, from which the requirements for the development of a Data Mining project are then derived.

## 2 A Framework for the Business Decision-Making Model and Requirements Modeling

The proposed framework (Figure 1) consists of a three phase compound iterative process. In the first phase, the elicitation of the needed information is carried out, allowing a proper understanding of the business dominion. Subsequently, based on this information, the model is built in the second phase. In the third phase, from the decisional business model obtained, the requirements model to be used in the development of a Data Mining project is derived.

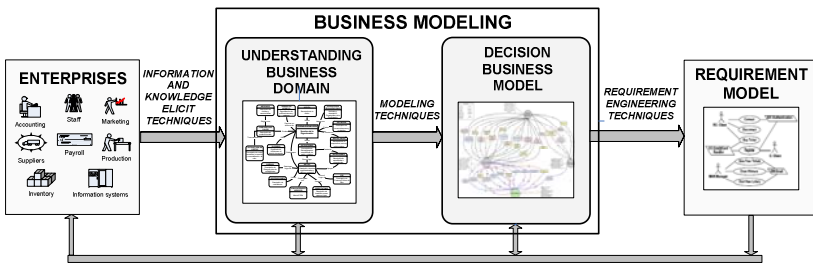


Fig. 1. Framework for modeling decision-making processes and requirements

### 2.1 Understanding the Business Domain Phase

The business domain of an organization is usually, quite complex and should be fully understood before undertaking any project development. The success of a Data Mining project will largely depend on the understanding, from a business or institutional point of view, of the objectives and requirements of the project [2].

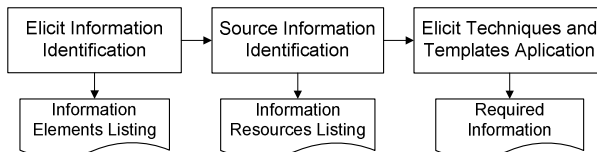


Fig. 2. Information about the business domain

The development of this first phase consists of three activities (figure 2): the identification of the relevant information to be elicited (based on the description of the basic components of the business model proposed in [15], [10]) which will allow an overview of the business, identification of the information sources and the application of requirement eliciting techniques and tools.

## 2.2 Decision-Making Business Model

After completing the business domain understanding phase and identifying the organizational goals, the decision-making process model is constructed. This is structured in a sequence of six stages (figure 3). The description of each stage is as follows:

### 2.2.1 Definition of the Originating Goal of the Decision-Making Process

In this first stage, the strategic goal that generates the decisional process to be modeled must be identified from the organizational goals defined in the business domain understanding phase. This goal will later be the highest level objective to be reached through the achievement of a series of lower-level targets which, in turn, may be split into various tasks performed by the organizational actors.

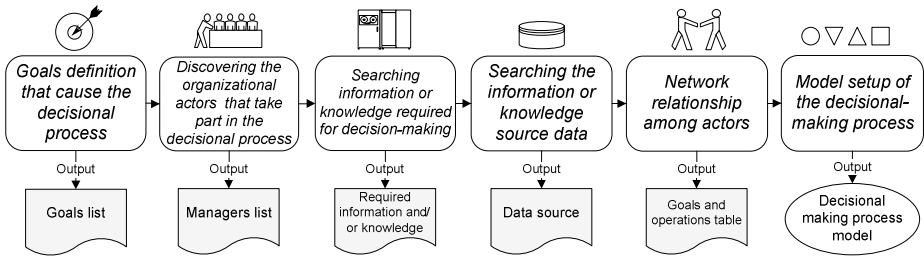


Fig. 3. Stages of the modeling process

### 2.2.2 Definition of the Organizational Actors That Participate in the Decision-Making Process

Within an organization, there are several actors involved in the decision-making processes. At the operational and knowledge levels decisions are structured, that is, decision processes rely on procedures established within the organization. Progressing towards the strategic levels, the decision processes become non-structured or at least semi-structured and include a high level of uncertainty and affect the entire organization. Therefore at this stage, only the actors involved in strategic decision making ("primary actors") should be identified and additionally those actors involved in the decision making process but who do not take decisions ("secondary actors").

### 2.2.3 Elicitation of Information or Knowledge Needed for Decision-Making

In the third stage, the necessary information or knowledge acquired or assimilated by the primary organizational actors to make their decisions is determined. The definition of this information or knowledge constitutes, within the model, the lowest level objectives to be achieved to accomplish the general goal which originates the decisional process. At this stage, an additional challenge for modelers and knowledge engineers is the determination, where ever possible, of factors not explicitly defined but which the decision-makers might unconsciously include in their thought processes.

### 2.2.4 Determination of the Data, That Serves as an Information or Knowledge Source

During this stage all the necessary data sources from which information or knowledge needed by decision makers are determined.

### 2.2.5 Definition of the Dependency Networks between the Various Organizational Actors

This stage aims to determine how the dependency networks are established between the various actors involved in the decision-making process for the achievement of the objectives related to the strategic goal which generates the decisional process. From this information, the goals refinement tree is built where the highest level goal is broken down into achievement goals, associated operations and involved actors. From this information, the graphical representation of the decision-making process is built.

### 2.2.6 Construction of the Decision-Making Model

The construction of the decision-making process model, is the last stage of the modeling process and involves the application of the “framework i\*” [19], [20], the chosen technique to graphically represent the information captured in the earlier stages, using two complementary models, *the strategic dependency model and the strategic rationale model* (figure 4):

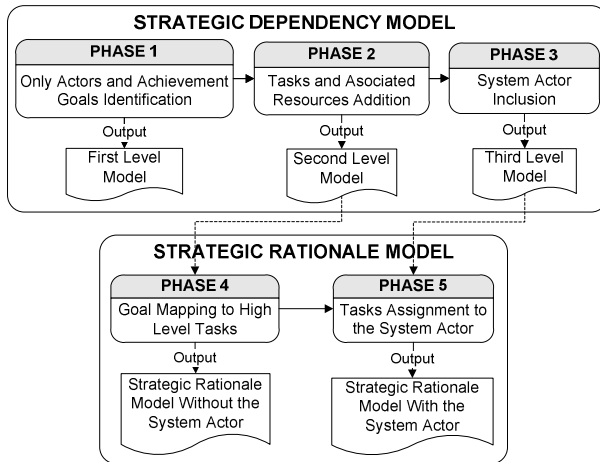


Fig. 4. Phases in the i\* modeling process

#### a) Strategic Dependency Model

This model aims to represent the actors in the decision-making process and their dependence links, such as achievement goals, soft goals, tasks, and/or required or generated resources for the realization of the goals identified in the first phase of the process (business domain understanding). It is proposed to develop this model with an incremental three phase modality:

**Phase 1.** A very basic initial model is developed (*first level model*) with a high abstraction level so all Stakeholders, in the most simple and intuitive way, understand what the model represents at the goal achievement level. In this first model only the actors that participate in the process and their goal dependency network are represented.

**Phase 2.** The second model (*second level model*) has a greater level of detail. In this model, the tasks arising from each achievement goal, the resources required for its concretion, and the resources that are generated from the development of some task or operation of the process are shown.

**Phase 3.** In this phase (*third level model*), the actor "System" is incorporated. The highest level goals are divided into more granular tasks. Then the tasks that can be automated are identified to derive them to the "System" (Data Mining system).

## b) Strategic Rationale Model

This second model, aims to represent more explicitly the resources (*material or informational*) and the sequence of granular events that trigger the activities required for meeting the achievement goals. These can also be carried out incrementally for an improved understanding of the process. The phases are:

**Phase 4.** To simplify the analysis, modeling is initially carried out from the second level of the strategic dependency model. During this phase, the achievement goals of the strategic dependency model are mapped into high level tasks. These tasks, depending on their level of complexity, can be divided by the constructor "*task decomposition*" into less complex tasks that can be developed by a particular actor. If more than one alternative to achieve a goal or task exists, the constructor "*means-end*" is used for its representation.

**Phase 5.** To complete the modeling process, the actor "System" is incorporated, thereby basing the construction of this model on the third level strategic dependency model (phase 3). All tasks are assigned to the actor "System" that need to be automated and that were achievement goals in the strategic dependency model. Then, in the actor System field of action, the achievement goals of the third level strategic dependency model are mapped into high level tasks and then divided into more specific tasks depending on the type of resources available and the level of specialization of each problem.

## 2.3 Requirements Derivation

There are various techniques to obtain a software requirements model from a business model; some are presented in [14], [1] or [17]. The acquisition of the requirements model for Data Mining, is based on the approach proposed in [17]. Consider, therefore, that an organizational model has been identified in which the actors take part in the decisional process and the main achievement goals are part of a broader strategic goal. From this a *use case* model is derived, using as principal inputs the strategic dependencies and strategic rationale models.

The proposal is developed in a sequence of three steps (figure 5). In steps one and two the *use case* actors and the *use cases* are identified respectively, using as input the

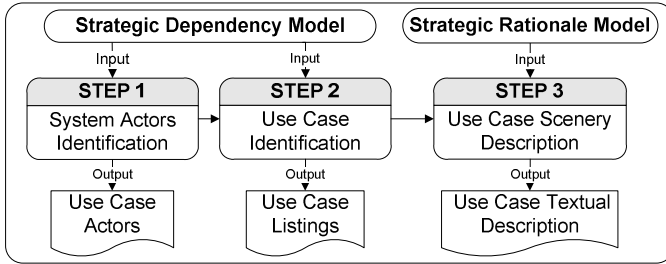


Fig. 5. Activities in the requirements modeling ([17])

strategic dependencies model. In the third step the scenarios of the *use cases* are described, based on information from the strategic rationale model.

**Step 1: Identification of the system actors.** They are all actors that are linked with the System actor (directly or indirectly) by some form of dependency. Those actors, without any type of dependency relationship to the System actor, cannot be regarded as actors of a *use case*.

**Step 2: Identification of use cases.** All the goal dependencies are identified (*dependum*) for which the actor identified in the first step plays the role of *dependee* (on whom it is depended) in the dependency relationship. Also identified are the goal dependencies in which the actor carries out a *dependor* roll (dependent) in the relationship.

- When the actor acts as a "*dependee*" for a goal dependency, it is assumed that the system, in order to generate an output, requires that the actor provides some kind of informational resource and the dependency object (*dependum*), making the latter a *use case*.
- The case in which the actor acts as a "*dependor*" in the dependencies league with the system, it represents the use of the knowledge or information that the system provides the actor with. In this case, again the object of the dependency (*dependum*) becomes a *use case*.

**Step 3: Description of the Use Case Scenario.** The third step is the description of the *use case* scenario associated with a particular actor. That is, the description of the sequence of internal events that are triggered during the task execution and the acquisition of the resources associated with the fulfillment of a goal. This sequence of events is described graphically in the strategic rationale model. The description of the scenarios associated with the *use cases* is the textual description associated with each *use case* in terms of the involvement of each actor in the achievement of a goal, task, or resource represented in the strategic rationale model. Finally the collection of use cases identified is part of the contract document which establishes the project sponsor and developers.

### 3 Evaluation of the Proposed Framework

From the point of view of comparing the proposed framework with other systems with similar purposes, currently there is not any similar framework in the literature, to

**Table 1.** Comparing two data mining projects to report benefits of the ER-DM

Criterion	Project A (without ER-DM)	Project B (with ER-DM)
Goal achievement	The objectives were established informally and therefore when finalizing the project was difficult to establish if the business objectives were fulfilled.	The objectives of the project were well-established and indeed were aligned with the business objectives.
User participation	The degree of user participation decayed significantly during the project development time.	The project was characterized by the active participation of the user in the diverse stages of the project.
Development time	The time of development surpassed largely the planned initially. (more than 30% of time increment)	The development time was adjusted to the established initially.
Effort	Greater efforts were dedicated than the initially considered (more than 40% of HH estimated initially).	The development time was adjusted to the established initially.

guide the construction of the decision-making process model and derive a requisite model for Data Mining projects, limiting the comparison base. Given this, and in order to establish a base of evaluation and estimation of the degree of benefit of the proposed Framework (ER-DM), this one was applied to the development of a Data Mining project. Previously, another project had been developed in the same business dominion, but without the application of an appropriate methodological procedure to derive its requirements. The project leaders in both cases (project A and project B) were the same ones, and the corresponding project objectives were similar but with different users and data analysts. The comparison is made on the basis of criteria of degree of fulfillment of the business objectives, time of development, effort of dedication and participation of the user, and it is summarized in Table 1.

## 4 Conclusions

The proposal described in this paper was applied during the development of a practical case to define the requirements of a Data Mining Project. This test verified that the proposed framework brings together a series of relevant aspects to be considered at the beginning of the development of a Data Mining project. These aspects relate to the project requirements, scope, necessary resources, risks and restrictions, and in general all the significant elements that emerge from the business domain understanding for the development of a Data Mining project.

Additionally it has been found that the framework addresses many of the problems that typically occur during the development of Data Mining projects [2].

The decisional model construction process, developed in conjunction with business experts, allows a clearer definition of which activities of the strategic decision-making process the Data Mining projects could support.

Finally it is important to note that, from the requirements model derived from the organization's business model, it is not only Data Mining *use cases* that can be identified, but also, other functionalities that can be met by traditional software systems, which are complemented with the Data Mining to achieve the organizational goals.

## References

1. Alencar, F., et al.: From Early Requirements Modeled by The i\* Technique to Later Requirements Modeled in Precise UML. In: III Workshop de Engenharia de Requisitos, WER 2000 (2000)
2. Britos, P., Dieste, O., García-Martínez, R.: Requirements Elicitation in Data Mining for Business Intelligence Projects. In: Avison, D., Kasper, G., Pernici, B. (eds.) IFIP International Federation for Information Processing. Advances in Information Systems Research, Education and Practice, vol. 274, pp. 139–150. Springer, Boston (2008)
3. Cysneiros, L., Sampaio, J.: Nonfunctional Requirements: From Elicitation to Conceptual Models. *IEEE Transactions on Software Engineering* 30(5) (2004)
4. Gacitúa, R.: Identification of Requirements: A Focus Based on a Verb Taxonomy. *Theoria* 10, 67–78 (2001)
5. Gordijn, J., Akkermans, H.: Business Models for Distributed Energy Resources in a Liberalized Market Environment. *The Electric Power Systems Research Journal* 77(9), 1178–1188 (2007)
6. Gorschek, T., Wohlin, C.: Requirements Abstraction Model. *Requirements Eng.* 11, 79–101 (2006)
7. Kelley, C., Adelman, S.: Where Can I Find Sources About Failed Data Mining Projects and The Reason for Their Failure? *DM Review Online* (April 2003)
8. Kotonya, G., Sommerville, I.: *Requirements Engineering. Processes and Techniques*. J. Wiley, USA (1998)
9. Koubarakis, M., Plexousakis, D.: Business Process, Modelling and Design: A Formal Model and Methodology. *BT Technology Journal* 17(4) (1999)
10. Lagha, B., Osterwalder, A., Pigneur, Y.: An Ontology for E-Business Models. In: Wendy, C. (ed.) *Value Creation from E-Business Models*. Butterworth-Heinemann, Butterworths (2004)
11. Marbán, O.: *Modelo Matemático Paramétrico de Estimación para Proyectos de Data Mining (DMCOMO)*, Tesis Doctoral, Facultad de Informática, U.P.M. (2003)
12. Meta Group Research: The Top 5 Global 3000 Data Mining Trends for 2003/04, Enterprise Analytics Strategies, Application Delivery Strategies, Meta Group Research-Delta Summary, 2061, March 26 (2003)
13. Mylopoulos, J., Kolp, M., Castro, J.: Towards Requirements-Driven Information Systems Engineering: The Tropos Project. *Information Systems* 27(6), 365–389 (2002)
14. Ortín, M., et al.: De los Procesos de Negocio a los Casos de Uso. In: *JISBD 2000*, Valladolid, España, Fecha (November 2000)
15. Osterwalder, A., Pigneur, Y., Tucci, C.: Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of AIS* 15, 751–775 (2005)
16. Ould, M.: *Business Processes-Modelling and Analysis for Re-engineering and Improvement*. John Wiley & Sons, Chichester (1995)
17. Santander, V., Castro, J.: Deriving Use Cases from Organizational Modeling. In: *Proceedings of the IEEE Joint International Conference on Requirements Engineering (RE 2002)*. IEEE, Los Alamitos (2002)
18. Vérosle, J., Stap, R., Salden, A.: A Generic Model for WLAN, Hostpots- A Roaming, Business Case in The Netherlands. In: *WMASH 2003*, San Diego CA, USA, September 19 (2003)
19. Yu, E.: *Modelling Strategic Relationships for Process Reengineering*, Ph.D. thesis, Department of Computer Science, University of Toronto (1995)
20. Yu, E., Mylopoulos, J.: AI Models for Business Process Reengineering. *IEEE Expert Intelligent Systems and Their Applications* 11(4), 16–23 (1996)