Database Marketing Process Supported by Ontologies: A Data Mining System Architecture Proposal

Filipe Mota Pinto[1] and Teresa Guarda[2]

[1]Polytechnic Institute of Leiria,
[2]Superior Institute of Languages and Administration of Leiria,
Portugal

1. Introduction

Marketing departments handles with a great volume of data which are normally task or marketing activity dependent. This requires the use of certain, and perhaps unique, specific knowledge background and framework approach.

Database marketing provides in depth analysis of marketing databases. Knowledge discovery in database techniques is one of the most prominent approaches to support some of the database marketing process phases. However, in many cases, the benefits of these tools are not fully exploited by marketers. Complexity and amount of data constitute two major factors limiting the application of knowledge discovery techniques in marketing activities. Here, ontologies may play an important role in the marketing discipline.

Motivated by its success in the area of artificial intelligence, we propose an ontology-supported database marketing approach. The approach aims to enhance database marketing process supported by a data mining system architecture proposal which provides detailed step-phase specific information.

From a data mining framework, issues raised in this work both respond and contribute to calls for a database marketing process improvement. Our work was evaluated throughout a relationship marketing program database. The findings of this study not only advance the state of database marketing research but also shed light on future research directions using a data mining approach. Therefore we propose a framework supported by ontologies and knowledge extraction from databases techniques. Thus, this paper has two purposes: to integrate the ontological approach into Database Marketing and to make use of a domain ontology - a knowledge base that will enhance the entire process at both levels, marketing and knowledge extraction techniques.

2. Motivation

Knowledge discovery in databases is a well accepted definition for related methods, tasks and approaches for knowledge extraction activities (Brezany et al., 2008) (Nigro et al., 2008). Knowledge extraction or Data Mining (DM) is also referred as a set of procedures that cover all work ranging from data collection to algorithms execution and model evaluation. In each

of the development phases, practitioners employ specific methods and tools that support them in fulfilling their tasks. The development of methods and tasks for the different disciplines have been established and used for a long time (Domingos, 2003) (Cimiano et al., 2004) (Michalewicz et al., 2006). Until recently, there was no need to integrate them in a structured manner (Tudorache, 2006). However, with the wide use of this approach, engineers were faced with a new challenge: They had to deal with a multitude of heterogeneous problems originating from different approaches and had to make sure that in the end all models offered a coherent business domain output. There are no mature processes and tools that enable the exchange of models between the different parallel developments at different contexts (Jarrar, 2005). Indeed, there is a gap in the KDD process knowledge sharing in order to promote its reuse.

The Internet and open connectivity environments created a strong demand for the sharing of data semantics (Jarrar, 2005). Emerging ontologies are increasingly becoming essential for computer science applications. Organizations are beginning to view them as useful machine-processable semantics for many application areas. Hence, ontologies have been developed in artificial intelligence to facilitate knowledge sharing and reuse. They are a popular research topic in various communities, such as knowledge engineering (Borst et al., 1997) (Bellandi et al., 2006), cooperative information systems (Diamantini et al., 2006b), information integration (Bolloju et al., 2002) (Perez-Rey et al., 2006), software agents (Bombardier et al., 2007), and knowledge management (Bernstein et al., 2005) (Cardoso and Lytras, 2009). In general, ontologies provide (Fensel et al., 2000): a shared and common understanding of a domain which can be communicated amongst people and across application systems; and, an explicit conceptualization (i.e., meta information) that describes the semantics of the data.

Nevertheless, ontological development is mainly dedicated to a community (e.g., genetics, cancer or networks) and, therefore, is almost unavailable to others outside it. Indeed the new knowledge produced from reused and shared ontologies is still very limited (Guarino, 1998) (Blanco et al., 2008) (Coulet et al., 2008) (Sharma and Osei-Bryson, 2008) (Cardoso and Lytras, 2009).

To the best of our knowledge, in spite of successful ontology approaches to solve some KDD related problems, such as, algorithms optimization (Kopanas et al., 2002) (Nogueira et al., 2007), data pre-processing tasks definition (Bouquet et al., 2002) (Zairate et al., 2006) or data mining evaluation models (Cannataro and Comito, 2003) (Brezany et al., 2008), the research to the ontological KDD process assistance is sparse and spare. Moreover, mostly of the ontology development focusing the KDD area focuses only a part of the problem, intending only to modulate data tasks (Borges et al., 2009), algorithms (Nigro et al., 2008), or evaluation models (Euler and Scholz, 2004) (Domingues and Rezende, 2005). Also, the use of KDD in marketing field has been largely ignored (with a few exceptions (Zhou et al., 2006) (El-Ansary, 2006) (Cellini et al., 2007)). Indeed, many of these works provide only single specific ontologies that quickly become unmanageable and therefore without the sharable and reusable characteristic. Such research direction may became innocuous, requiring tremendous patience and an expert understanding of the ontology domain, terminology, and semantics.

Contrary to this existing research trend, we feel that since the knowledge extraction techniques are critical to the success of database use procedures, researchers are interested

in addressing the problem of knowledge share and reuse. We must address and emphasize the knowledge conceptualization and specification through ontologies.

Therefore, this research promises interesting results in different levels, such as:

- Regarding information systems and technologies, focusing the introduction and integration of the ontology to assist and improve the DM process, through inference tasks in each phase;

- In the ontology area this investigation represents an initial approach step on the way for real portability and knowledge sharing of the system towards other similar DBM process supported by the DM. It could effectively be employed to address the general problem of model-construction in problems similar to the one of marketing (generalization), on the other side it is possible to instantiate/adapt the ontology to the specific configuration of a DBM case and to automatically assist, suggest and validate specific approaches or models DM process (specification);

- Lastly, for data analyst practitioners this research may improve their ability to develop the DBM process, supported by DM. Since knowledge extraction work depended in large scale on the user background, the proposed methodology may be very useful when dealing with complex marketing database problems. Therefore the introduction of an ontological layer in DBM project allows: more efficient and stable marketing database exploration process through an ontology-guided knowledge extraction process; and, portability and knowledge share among DBM practitioners and computer science researchers.

3. Background

3.1 Database marketing

Much of the advanced practice in Database Marketing (DBM) is performed within private organizations (Zwick and Dholakia, 2004) (Marsh, 2005). This may partly explain the lack of articles published in the academic literature that study DBM issue (Bohling et al., 2006) (Frankland, 2007) (Lin and Hong, 2008).

However, DBM is nowadays an essential part of marketing in many organizations. Indeed, as the main DBM principle, most organizations should communicate as much as possible with their customers on a direct basis (DeTienne and Thompson, 1996). Such objective has contributed to the expressive grown of all DBM discipline. In spite of such evolution and development, DBM has growth without the expected maturity (Fletcher et al., 1996) (Verhoef and Hoekstra, 1999).

In some organizations, DBM systems work only as a system for inserting and updating data, just like a production system (Sen and Tuzhiln, 1998). In others, they are used only as a tool for data analysis (Bean, 1999). In addition, there are corporations that use DBM systems for both operational and analytical purposes (Arndt and Gersten, 2001). Currently DBM is mainly approached by classical statistical inference, which may fail when complex, multi-dimensional, and incomplete data is available (Santos et al., 2005).

One of most cited origins of DBM is the retailers' catalogue based in the USA selling directly to customers. The main means used was direct mail, and mailing of new catalogues usually took place to the whole database of customers (DeTienne and Thompson, 1996). Mailings result analysis has led to the adoption of techniques to improve targeting, such as CHAID (Chi-Squared Automated Interaction Detection) and logistic regression (DeTienne and

Thompson, 1996) (Schoenbachler et al., 1997). Lately, the addition of centralized call centers and the Internet to the DBM mix has introduced the elements of interactivity and personalization. Thereafter, during the 1990s, the data-mining boom popularized such techniques as artificial neural networks, market basket analysis, Bayesian networks and decision trees (Pearce et al., 2002) (Drozdenko and Perry, 2002).

3.1.1 Definition

DBM refers to the use of database technology for supporting marketing activities (Leary et al., 2004) (Wehmeyer, 2005) (Pinto et al., 2009). Therefore, it is a marketing process driven by information (Coviello et al., 2001) (Brookes et al., 2004) (Coviello et al., 2006) and managed by database technology (Carson et al., 2004) (Drozdenko and Perry, 2002). It allows marketing professionals to develop and to implement better marketing programs and strategies (Shepard, 1998) (Ozimek, 2004).

There are different definitions of DBM with distinct perspectives or approaches denoting some evolution an evolution along the concepts (Zwick and Dholakia, 2004). From the marketing perspective, DBM is an interactive approach to marketing communication. It uses addressable communications media (Drozdenko and Perry, 2002) (Shepard, 1998), or a strategy that is based on the premise that not all customers or prospects are alike. By gathering, maintaining and analyzing detailed information about customers or prospects, marketers can modify their marketing strategies accordingly (Tao and Yeh, 2003). Then, some statistical approaches were introduced and DBM was presented as the application of statistical analysis and modeling techniques to computerized individual level data sets (Sen and Tuzhiln, 1998) (Rebelo et al., 2006) focusing some type of data. Here, DBM simply involves the collection of information about past, current and potential customers to build a database to improve the marketing effort. The information includes: demographic profiles; consumer likes and dislikes; taste; purchase behavior and lifestyle (Seller and Gray, 1999) (Pearce et al., 2002).

As information technologies improved their capabilities such as processing speed, archiving space or, data flow in organizations that have grown exponentially different approaches to DBM have been suggested: generally, it is the art of using data you've already gathered to generate new money-making ideas (Gronroos, 1994) (Pearce et al., 2002); stores this response and adds other customer information (lifestyles, transaction history, etc.) on an electronic database memory and uses it as basis for longer term customer loyalty programs, to facilitate future contacts, and to enable planning of all marketing. (Fletcher et al., 1996) (Frankland, 2007); or, DBM can be defined as gathering, saving and using the maximum amount of useful knowledge about your customers and prospects...to their benefit and organizations' profit. (McClymont and Jocumsen, 2003) (Pearce et al., 2002). Lately some authors has referred DBM as a tool database-driven marketing tool which is increasingly aking centre stage in organizations strategies (Pinto, 2006) (Lin and Hong, 2008).

In common all definition share a main idea: DBM is a process that uses data stored in marketing databases in order to extract relevant information to support marketing decision and activities through customer knowledge, which will allow satisfy their needs and anticipate their desires.

3.1.2 Database marketing process

During the DBM process it is possible to consider three phases (DeTienne and Thompson, 1996) (Shepard, 1998) (Drozdenko and Perry, 2002): data collection, data processing (modeling) and results evaluation.

The Figure 1 presents a simple model of how customer data are collected through internal or external structures that are closer to customers and the market, how customer data is transformed into information and how customer information is used to shape marketing strategies and decisions that later turn into marketing activities. The first, Marketing data, consists in data collection phase, which will conduct to marketing database creation with as much customer information as possible (e.g., behavioral, psychographic or demographic information) and related market data (e.g., share of market or competitors information's). During the next phase, information, the marketing database is analyzed under a marketing information perspective throughout activities such as, information organization (e.g., according organization structure, or campaign or product relative); information codification (e.g., techniques that associates information to a subject) or data summarization (e.g., cross data tabulations). The DBM development process concludes with marketing knowledge, which is the marketer interpretation of marketing information in actionable form. In this phase there has to be relevant information to support marketing activities decision.
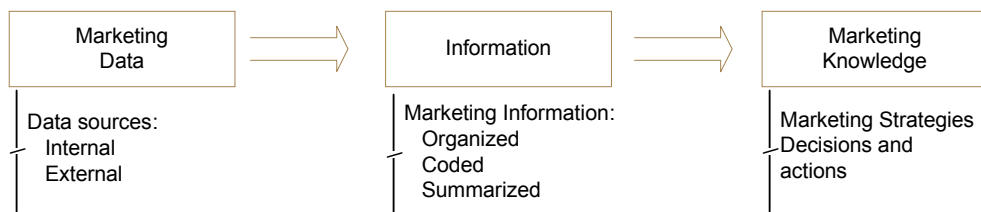

Fig. 1. Database marketing general overall process

Technology based marketing is almost a marketing science imperative (Brookes et al., 2004) (Zineldin and Vasicheva, 2008). As much as marketing research is improving and embracing new challenges its dependence on technology is also growing (Carson et al., 2004). Currently, almost every organization has its own marketing information system, from single customer data records to huge data warehouses (Brito, 2000). Nowadays, DBM is one of the most well succeed marketing technology employment (Frankland, 2007) (Lin and Hong, 2008) (Pinto et al., 2009).

3.1.3 DBM process with KDD

Database marketing is a capacious term related to the way of thinking and acting which contains the application of tools and methods in studies, their structure and internal organization so that they could achieve success on a fluctuating and difficult to predict consumer market (Lixiang, 2001).

For the present purpose we assume that, database marketing can be defined as a method of analyzing customer data to look for hidden, useful and actionable knowledge for marketing purposes. To do so, several different problem specifications may be referred. These include market segmentation (Brito et al., 2004), cross-sell prediction, response modeling, customer valuation (Brito and Hammond, 2007) and market basket analysis (Buckinx and den Poel, 2005) (Burez and Poel, 2007). Building successful solutions for these tasks requires the application of advanced DM and machine learning techniques to obtain relationships and patterns in marketing databases data and using this knowledge to predict each prospect's reaction to future situations.

In literature there are some examples about KDD usage in DBM projects usage for customers' response modeling whereas the goal was to use past transaction data of customers, personal characteristics and their response behavior to determine whether these clients were good or not (Coviello and Brodie, 1998) e.g., for mailing prospects during the next period (Pearce et al., 2002) (den Poel and Buckinx, 2005). At these examples different analytical approaches were used: statistical techniques (e.g., discriminate analysis, logistic regression, CART and CHAID), machine learning methods (e.g., C4.5, SOM) mathematical programming (e.g., linear programming classification) and neural networks to model this customer's response problem.

Other KDD related application in DBM projects is customer retention activities. The retention of its customers is very important for a commercial entity, e.g., a bank or a oil distribution company. Whenever a client decides to change to another company, it usually implies some financial losses for this organization. Therefore, organizations are very interested in identifying some mechanisms behind such decisions and determining which clients are about to leave them. As an example one approach to find such potential customers is to analyze the historical data which describe customer behavior in the past (den Poel and Buckinx, 2005) (Buckinx and den Poel, 2005) (Rebelo et al., 2006) (Burez and Poel, 2007) (Buckinx et al., 2007).

3.2 Ontologies

Currently we live at a web-based information society. Such society has a high-level automatic data processing which requires a machine-understandable of representation of information's semantics. This semantics need is not provided by HTML or XML-based languages themselves. Ontologies fill the gap, providing a sharable structure and semantics of a given domain, and therefore they play a key role in such research areas such as knowledge management, electronic commerce, decision support or agent communication (Ceccaroni, 2001).

Ontologies are used to study the existence of all kinds of entities (abstract or concrete) that constitute the world (Sowa, 2000). Ontologies use the existential quantifier $\exists$ as a notation for asserting that something exists, in contrast to logic vocabulary, which doesn't have vocabulary for describing the things that exist.

They are also used for data-source integration in global information systems and for in-house communication. In recent years, there has been a considerable progress in developing the conceptual bases for building ontologies. They allow reuse and sharing of knowledge components, and are, in general, concerned with static domain-knowledge.

Ontologies can be used as complementary reusable components to construct knowledge-based systems (van Heijst et al., 1997). Moreover, ontologies provide a shared and common understanding of a domain and describe the reasoning process of a knowledge-based system, in a domain and independent implementation fashion.

3.2.1 Ontologies definition

From the philosophy perspective, ontology is the theory or study of being, i.e., of the basic characteristics of all reality. Though the term was first coined in the 17th century, ontology is synonymous with metaphysics or first philosophy as defined by Aristotle in the 4th century BC (Guarino, 1995). Ontology is a part of metaphysics (Newell and level, 1982): it is the science of the existence which investigates the structure of being in general, rather than analyzing the characteristics of particular beings.

To answer the question "*but what is being?*" it was proposed a famous criterion but which did not say anything about what actually exists:"*To be is to be the value of a quantified variable*"(Quine, 1992). Those who object to it would prefer some guidelines for the kinds of legal statements. In general, furt her analysis is necessary to give the knowledge engineer some guidelines about what to say and how to say it.

From artificial intelligence literature there is a wide range of different definitions of the term ontology. Each community seems to adopt its own interpretation according to the use and purposes that the ontologies are intended to serve within that community. The following list enumerates some of the most important contributions:

- One of the early definitions is: 'An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary.' (Neches et al., 1991);
- A widely used definition is: 'An ontology is an explicit specification of a conceptualization´ (Gruber, 1993);
- An analysis of a number of interpretations of the word ontology (as an informal conceptual system, as a formal semantic account, as a specification of a conceptualization, as a representation of a conceptual system via a logical theory, as the vocabulary used by a logical theory and as a specification of a logical theory) and a clarification of the terminology used by several other authors is in Guarino and Giaretta work (Guarino, 1995).
- From Gruber's definition and more elaborated is: 'Ontologies are defined as a formal specification of a shared conceptualization.'(Borst et al., 1997);
- 'An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.' (Swartout et al., 1996);
- A definition with an explanation of the terms also used in early definitions, states: 'conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared refers to the notion that an ontology captures consensual knowledge, that is, it is not primitive to some individual, but accepted by a group (Staab and Studer, 2004);
- An interesting working definition is: Ontology may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definitions and explicitly designates how concepts are interrelated which collectively impose a structure on the domain and constrain the possible interpretations of terms. Moreover, ontology is virtually always the manifestation of a shared understanding of a domain that is agreed between communities. Such agreement facilitates accurate and effective communication of meaning, which in turn, leads to other benefits such as inter-operability, reuse and sharing. (Jasper and Uschold, 1999);
- More recently, a broad definition has been given: 'ontologies to be domain theories that specify a domain-specific vocabulary of entities, classes, properties, predicates, and functions, and to be a set of relationships that necessarily hold among those vocabulary terms. Ontologies provide a vocabulary for representing knowledge about a domain and for de scribing specific situations in a domain.' (Farquhar et al., 1997) (Smith and Farquhar, 2008).

For this research, we have adopted as ontology definition: *A formal and explicit specification of a shared conceptualization, which is usable by a system in actionable forms*. Conceptualization refers to an abstract model of some phenomenon in some world, obtained by the identification of the relevant concepts of that phenomenon. Shared reflects the fact that an ontology captures consensual knowledge and is accepted by a relevant part of the scientific community. Formal refers to the fact that ontology is an abstract, theoretical organization of terms and relationships that is used as a tool for the analysis of the concepts of a domain. Explicit refers to the type of concepts used and the constraints on their use (Gruber, 1993) (Jurisica et al., 1999). Therefore, ontology provides a set of well-founded constructs that can be leveraged to build meaningful higher level knowledge. Hence, we consider that ontology is usable through systems in order to accomplish our objective: assistance work throughout actionable forms.

3.2.2 Reasons to use ontologies

Ontology building deals with modeling the world with shareable knowledge structures (Gruber, 1993). With the emergence of the Semantic Web, the development of ontologies and ontology integration has become very important (Fox and Gruninger, 1997) (Guarino, 1998) (Berners-Lee et al., 2001). The SemanticWeb is a vision, for a next generation Web and is described in a Figure 7 called the "layer cake" of the Semantic Web (Berners-Lee, 2003) and presented in the Ontology languages section.

The current Web has shown that string matching by itself is often not sufficient for finding specific concepts. Rather, special programs are needed to search the Web for the concepts specified by a user. Such programs, which are activated once and traverse the Web without further supervision, are called agent programs (Zhou et al., 2006). Successful agent programs will search for concepts as opposed to words. Due to the well known homonym and synonym problems, it is difficult to select from among different concepts expressed by the same word (e.g., Jaguar the animal, or Jaguar the car). However, having additional information about a concept, such as which concepts are related to it, makes it easier to solve this matching problem. For example, if that Jaguar IS-A car is desired, then the agent knows which of the meanings to look for.

Ontologies provide a repository of this kind of relationship information. To make the creation of the Semantic Web easier, Web page authors will derive the terms of their pages from existing ontologies, or develop new ontologies for the Semantic Web.

Many technical problems remain for ontology developers, e.g. scalability. Yet, it is obvious that the Semantic Web will never become a reality if ontologies cannot be developed to the point of functionality, availability and reliability comparable to the existing components of the Web (Blanco et al., 2008) (Cardoso and Lytras, 2009).

Some ontologies are used to represent the general world or word knowledge. Other ontologies have been used in a number of specialized areas, such as, medicine (Jurisica et al., 1999) (CeSpivova et al., 2004) (Perez-Rey et al., 2006) (Kasabov et al., 2007), engineering (Tudorache, 2006) (Weng and Chang, 2008), knowledge management (Welty and Murdock, 2006), or business (Borges et al., 2009) (Cheng et al., 2009).

Ontologies have been playing an important role in knowledge sharing and reuse and are useful for (Noy and McGuinness, 2003):

- *Sharing common understanding* of the structure of information among people or software agents is one of the more common goals in developing ontologies (Gruber, 1993), e.g., when several different Web sites contain marketing information o r provide tools and

techniques for marketing activities. If these Web sites share and publish the same underlying ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications;

- *Enabling reuse of domain knowledge* was one of the driving forces behind recent surge in ontology research, e.g., models for many different domains need to represent the value. This representation includes social classes, income scales among others. If one group of researchers develops such an ontology in detail, others can simply reuse it for their domains. Additionally, if we need to build a large ontology, we can integrate several existing ontologies describing portions of the large domain;

- Making *explicit domain assumptions* underlying an implementation makes it possible to change these programming-language codes making these assumptions not only hard to find and understand but also hard to change, in particular for someone without programming expertise. In addition, explicit specifications of domain knowledge are useful for new users who must learn what terms in the domain mean;

- *Separating the domain knowledge from the operational knowledge* is another common use of ontologies, e.g., regarding computers hardware components, it is possible to describe a task of configuring a product from its components according to a required specification and implement a program that does this configuration independent of the products and components themselves. Then, it is possible develop an ontology of *PCcomponents* and apply the algorithm to configure made-to-order PCs. We can also use the same algorithm to configure elevators if we "feed" it an elevator component ontology (Rothenfluh et al., 1996);

- *Analyzing domain knowledge* is possible once a declarative specification of the terms is available. Formal analysis of terms is extremely valuable when both attempting to reuse existing ontologies and extending them.

Often ontology of the domain is not a goal in itself. Developing an ontology is akin to defining a set of data and their structure for other programs to use. Problem-solving methods, domain-independent applications, and software agents use ontologies and knowledge bases built from ontologies as data (van Heijst et al., 1997) (Gottgtroy et al., 2004). Within this work we have develop an DBM ontology and appropriate KDD combinations of tasks and tools with expected marketing results. This ontology can then be used as a basis for some applications in a suite of marketing-managing tools: One application could create marketing activities suggestions for data analyst or answer queries of the marketing practitioners. Another application could analyze an inventory list of a data used and suggest which marketing activities could be developed with such available resource.

3.2.3 Ontologies main concepts

Here we use ontologies to provide the shared and common domain structures which are required for semantic integration of information sources. Even if it is still difficult to find consensus among ontology developers and users, some agreement about protocols, languages and frameworks exists. In this section we clarify the terminology which we will use throughout the thesis:

- *Axioms* are the elements which permit the detailed modeling of the domain. There are two kinds of axioms that are important for this thesis: defining axioms and related

axioms. Defining axioms are defined as relations multi valued (as opposed to a function) that maps any object in the domain of discourse to sentence related to that object. A defining axiom for a constant (e.g., a symbol) is a sentence that helps defining the constant. An object is not necessarily a symbol. It is usually a class, or relation or instance of a class. If not otherwise specified, with the term axiom we refer to a related axiom;

- A *class* or *type* is a set of objects. Each one of the objects in a class is said to be an instance of the class. In some frameworks an object can be an instance of multiple classes. A class can be an instance of another class. A class which has instances that are themselves classes is called a meta-class. The top classes employed by a well developed ontology derive from the root class object, or thing, and they themselves are objects, or things. Each of them corresponds to the traditional concept of being or entity. A class, or concept in description logic, can be defined intentionally in terms of descriptions that specify the properties that objects must satisfy to belong to the class. These descriptions are expressed using a language that allows the construction of composite descriptions, including restrictions on the binary relationships connecting objects. A class can also be defined extensionally by enumerating its instances. Classes are the basis of knowledge representation in ontologies. Class hierarchies might be represented by a tree: branches represent classes and the leaves represent individuals.

- *Individuals*: objects that are not classes. Thus, the domain of discourse consists of individuals and classes, which are generically referred to as objects. Individuals are objects which cannot be divided without losing their structural and functional characteristics. They are grouped into classes and have slots. Even concepts like group or process can be individuals of some class.

- *Inheritance* through the class hierarchy means that the value of a slot for an individual or class can be inherited from its super class.

- *Unique identifier*: every class and every individual has a unique identifier, or name. The name may be a string or an integer and is not intended to be human readable. Following the assumption of anti-atomicity, objects, or entities are always complex objects. This assumption entails a number of important consequences. The only one concerning this thesis is that every object is a whole with parts (both as components and as functional parts). Additionally, because whatever exists in space-time has temporal and spatial extension, processes and objects are equivalent.

- *Relationships:* relations that operate among the various objects populating an ontology. In fact, it could be said that the glue of any articulated ontology is provided by the network of dependency of relations among its objects. The class-membership relation that holds between an instance and a class is a binary relation that maps objects to classes. The *type-of* relation is defined as the inverse of *instance-of* relation. If *A* is an *instance-of B*, then *B* is a *type-of A*. The *subclass-of* (or *is-a*) relation for classes is defined in terms of the relation *instance-of*, as follows: a class *C* is a *subclass-of* class *T* if and only if all instances of *C* are also instances of *T*. The *superclass-of* relation is defined as the inverse of the *subclass-of* relation.

- *Role*: different users or any single user may define multiple ontologies within a single domain, representing different aspects of the domain or different tasks that might be carried out within it. Each of these ontologies is known as a role. In our approach we do not need to use roles since we only deal with a single ontology. Roles can be shared, or

they can be represented separately in approaches without integration facilities. Moreover, roles can overlap in the sense that the same individuals can be classified in many different roles, but the class membership of an individual, its inherited slots and the values of those slots may vary from role to role. A representation of the similarities and differences between two or more roles is known as a comparison.

- *Slots* (values that properties can assume). Objects have associated with them a set of own slots and each own slot of an object has associated with it a set of objects called slot values. Slots can hold many different kinds of values and can hold many at the same time. They are used to store information, such as name and description, which uniquely define a class or an individual. Classes have associated with them a collection of template slots that describe own slot values considered to hold for each instance of the class. The values of template slots are said to inherit to the subclasses and to the instances of a class. The values of a template slot are inherited to subclasses as values of the same template slot and to instances as values of the corresponding own slot. For example, the assertion that the gender of all *female* persons is *female* could be represented by the template slot *Gender* of class *Female-Person* having the value *Female*. If we create an instance of *Female-Person* called *Linda*, then *Female* would be the value of the own slot *Gender* of *Linda*. Own slots of an object have associated with them a set of own facets, and each own facet of a slot of a frame has associated with it a set of objects called facet values, e.g., the assertion that *Francisco* favorite foods must be *sweet food* can be represented by the facet Value-Type of the *Favorite-Food* slot of the *Francisco* frame having the value *Sweet-Food*. Template slots of a class have associated with them a collection of template facets that describe own facet values considered to hold for the corresponding own slot of each instance of the class. As with the values of template slots, the values of template facets are said to inherit to the subclasses and instances of a class. Thus, the values of a template facet are inherited to subclasses as values of the same template facet and to instances as values of the corresponding own facet.
- A *taxonomy* is a set of concepts, which are arranged hierarchically. A taxonomy does not define attributes of these concepts. It usually defines only the is-a relationship between the concepts. In addition to the basic is-a relation, the part-of relation may also be used;
- A *type* is an ontological category in artificial intelligence (in which it is synonymous of class) and in logic;
- A *vocabulary* is a language dependent set of words with explanations/documentation. It seeks universality and formality in a local context (for example a marketing domain).

Focusing on ontology reuse capability (one of the most important aspect in many research projects), we attain to assist the end user in new DBM and KDD projects through knowledge base instantiation and inference.

4. Research approach

Through an exhaustive literature review we have achieve a set of domain concepts and relations between them to describe KDD process.

Follo wing METHONTOLOGY (Lopez et al., 1999) we had constructed our ontology in terms of process assistance role. This methodology for ontology construction has five (Gomez-Perez et al., 2004) main steps: specification, conceptualization, formalization, implementation and maintenance (Figure 2).
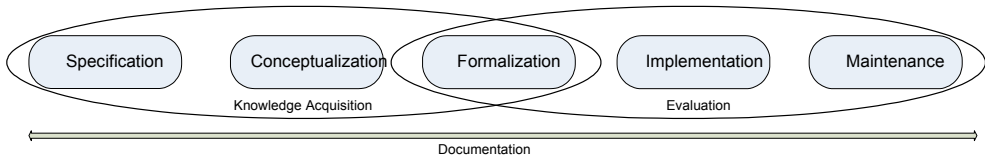
Fig. 2. Methontology framework (adapted from (Lopez *et al.*1999))

Nevertheless, domain concepts and relations were introduced according some literature directives (Smith and Farquhar2008). Moreover, in order to formalize all related knowledge we have used some relevant scientific KDD (Quinlan1986) (Fayyad *et al.*1996) and ontologies (Phillips and Buchanan2001)(Nigro *et al.*2008) published works. However, whenever some vocabulary is missing it is possible to develop a research method in order to achieve such a domain knowledge thesaurus.

At the end of the first step of methontology methodology we have identified the following main classes (Figure 3):

Fig. 3. KDD ontology class taxonomy (partial view)

Our KDD ontology has three major classes: *Resource*, *ProcessPhase* and *ResultModel*. ProcessPhase is the central class which uses resources (Resource class) and has some results (ResultModel class). The former Resource class relates all resources needed to carry the extraction process, namely algorithms and data.

The *ResultModel* has in charge to relate all KDD instance process describing all resources used, all tasks performed and results achieved in terms of model evaluation and domain evaluation. This class is use to ensure the KDD knowledge share and reuse.

Regarding KDD process we have considered four main concepts below the *ProcessPhase* concept (OWL class):

Data Understand focuses all data understanding work from simple acknowledge attribute mean to exhaustive attribute data description or even translation, to more natural language;

Data Preprocessing: concerns all data pre-processing tasks like data transformation, new attribute derivation or missing values processing;

Modeling: Modeling phase has in charge to produce models. It is frequent to appear as data mining phase (DM), since it is the most well known KDD phase. Discovery systems produce models that are valuable for prediction or description, but also they produce models that have been stated in some declarative format, that can be communicated clearly and precisely in order to become useful. Modeling holds all DM work from KDD process. Here we consider all subjects regarding the DM tasks, e.g., algorithm selection or concerns relations between algorithm and data used (data selection). In order to optimize efforts we have introduced some tested concepts from other data mining ontology (DMO) [Nigro *et al.*2008], which has similar knowledge base taxonomy. Here we take advantage of an explicit ontology of data mining and standards using the OWL concepts to describe an abstract semantic service for DM and its main operations. Settings are built through enumeration of algorithm properties and characterization of their input parameters. Based on the concrete Java interfaces, as presented in the Weka software API (Witten and Frank2000) and Protégé OWL, it was constructed a set of OWL classes and their instances that handle input parameters of the algorithms. All these concepts are not strictly separated but are rather used in conjunction forming a consistent ontology;

Evaluation and Deployment phase refers all concepts and operations (relations) performed to evaluate resulting DM model and KDD knowledge respectively.

Then, we have represented above concept hierarchy in OWL language, using protégé OWL software.

```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xml:base="http://www.semanticweb.org/ontologies/2009/5/DBMiPhDfpinto.owl">
  <owl:Class rdf:ID="InformationType">
   <rdfs:subClassOf>
    <owl:Class rdf:ID="Data"/>
   </rdfs:subClassOf>
  <owl:Class rdf:ID="Personal">
   <rdfs:subClassOf>
    <owl:Class rdf:ID="InformationType"/>
```

```
    </rdfs:subClassOf>
  </owl:Class>
</owl:Class>
<owl:Class rdf:ID="Demographics">
 <rdfs:subClassOf>
   <owl:Class rdf:ID="Personal"/>
 </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#InformationType">
 <rdfs:subClassOf rdf:resource="#Data"/>
</owl:Class>
```

Following Methontology, the next step is to create domain-specific core ontology, focusing knowledge acquisition. To this end we had performed some data processing tasks, data mining operations and also performed some models evaluations.

Each class belongs to a hierarchy (Figure 4). Moreover, each class may have relations between other classes (e.g., *PersonalType* is-a *InformationType* subclass). In order to formalize such schema we have defined OWL properties in regarding class' relationships, generally represented as:

Modeling^ has Algorithm (algorithm)
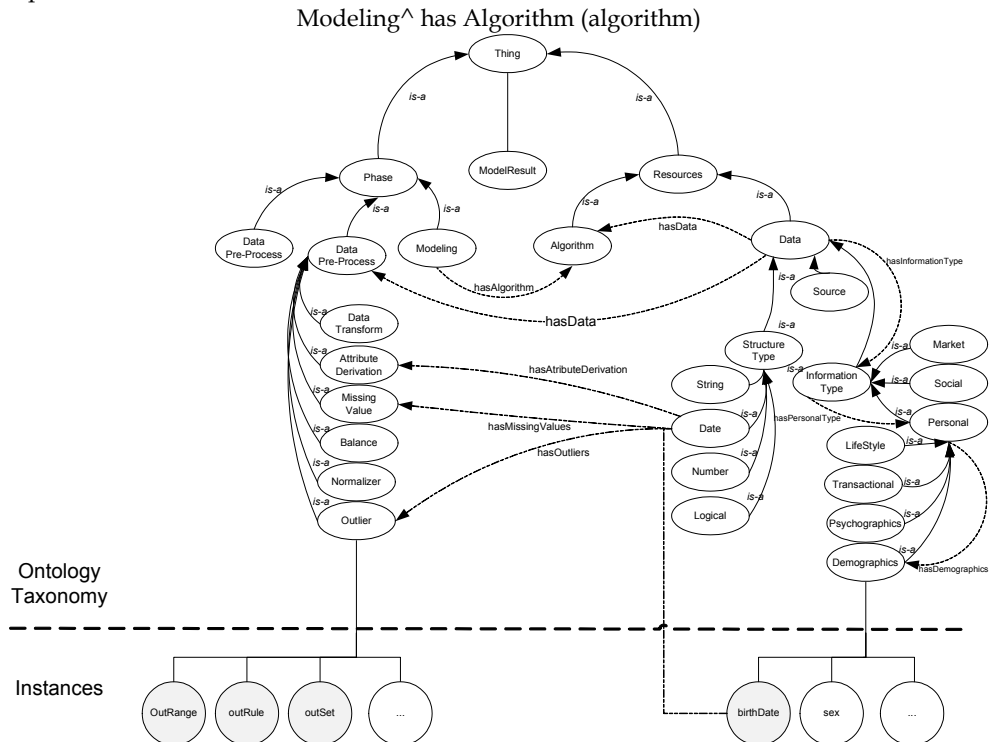
Fig. 4. KDD class/property/instance relation example ilustration

In OWL code:

```
<owl:Class rdf:ID="AlgorithmSelection">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom rdf:resource="#Algorithms"/>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="hasAlgorithm"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Modeling"/>
  </rdfs:subClassOf>
</owl:Class>
```

The ontology knowledge acquisition, firstly, happens through direct classes, relationships and instances load. Then through the KDD instantiation, the ontology acts according to the semantic structure.

Each new attribute is presented to the ontology, it is evaluated in terms of attribute class hierarchy, and related properties that acts according it.

In our ontology Attribute is defined by a set of three descriptive items: *Information Type*, *Structure Type* and allocated *Source*. Therefore it is possible to infer that, Attribute is a subclass of *Thing* and is described as a union of *InformationType*, *StructureType* and *Source*.

At other level, considering that, data property links a class to another class (subclass) or links a class with an individual, we have in our ontology the example:

> *StructureType(Date)*
> → *hasMissingValueTask*
> → *hasOutliersTask*
> → *hasAttributeDerive*

Attribute InformationType (Personal) & Attribute PersonalType(Demographics)
> → hasCheckConsistency

As example, considering the *birthDate* attribute, ontology will act as:

    ? Attribute hasDataSource
            attribute hasDataSource (CustomerTable).
    ? Attribute hasInformationType:
            attribute hasInformationType (Personal) then:
            attribute hasPersonalType(Demographics)
    ? Attribute hasStructureType
            attribute hasStructureType (Date).
            : attribute hasStructureType(Date) AND
    PersonalType(Demographics) then:
            : attribute (Demographics; Date) hasDataPreparation
            : attribute (Demographics; Date) hasDataPreProcessing
    AND Check missing values
    AND Check outliers
    AND Check consistency
    AND deriveNewAttribute

In above example, the inference process is executed on reasoner for description logic (Pellet). It acts along both class hierarchy (e.g., *Personal or Demographics*) and defined data properties

(e.g., *hasStructureType* or *hasDataPreparation*). In above example the attribute belongs at two classes: *Date* and *Demographics*. Through class membership, the *birthDate*, attribute inherits related data properties, such as *hasDataPreparation* or *hasDataPre-Processing*

5. Ontology leaning cycle

Ontology assistance to KDD aims the improvement of the process allowing both better performance and extracted knowledge results.Since KDD process is the core competency of database use, it is the centre focus of our work.
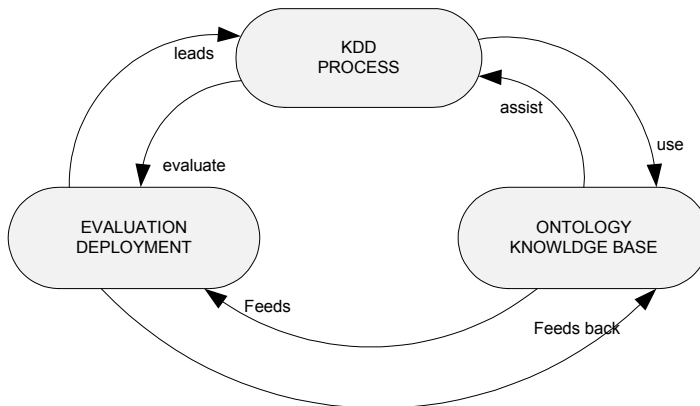
Fig. 5. Ontology learning cycle

As depicted in Figure 5, KDD process is located at the centre of our system. Therefore, data analyst uses knowledge during the process execution; knowledge feeds performance for higher achievement, and performance leads measures performance through evaluation and deployment methods; performance feeds back knowledge (ontology update) for later use of that knowledge. Also knowledge drives the process to improve further operations.

Since the KDD process generates as output models, it was considered useful to represent them in a computable way. Such representation works as a general description of all options taken during the process. Based on PMML descriptive DM model we have introduced an OWL class in our ontology named *ResultModel* which holds instances with general form:

> *ResultModel {*
> > *domain Objective Type;*
> > *algorithm;*
> > *algorithmTasks;*
> > *algorithmParameters;*
> > *workingAlgorithmDataSet;*
> > *EvaluationValue;*
> > *DeploymentValue*
> *}*

Moreover, our ontology has the learning capability mutually assigned to aforementioned model the ontology structure. Then it is possible both: so suggest (e.g., algorithm) and rank each suggestion (e.g., accuracy). Such approach may lead in a future to the development of an automatic learning capability and is depicted in figure 6.
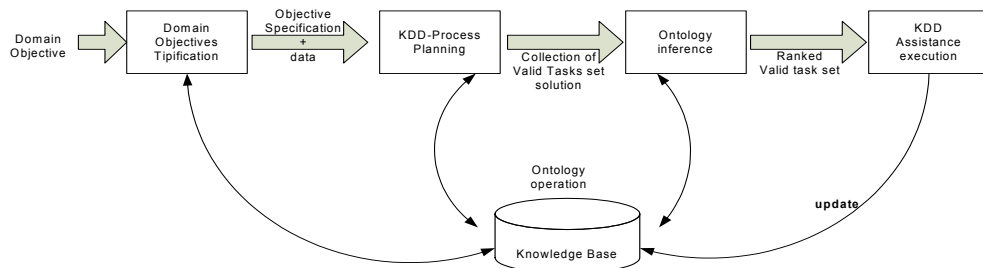
Fig. 6. Database Marketing ontology knowledge base operations

Data analyst is guided through the entire process supported by knowledge base. Such support is carried by domain objectives specification, KDD process planning, ontology inference or KDD assistant execution.

6. Results

As results we have achieved an explicit KDD ontology which integrates background and practical knowledge (Figure 7).

The KDD structure has two main distinct classes: resources and phase, as depicted in Figure 7. The former, holds and refers to all assets used at KDD process, like data repositories or algorithms; the latter, refers to the practical development of KDD process phases, like data preparation or modeling. Each super class has its own subclass hierarchy. Moreover, there are relationships between each class (e.g., *hasData* or *hasAlgorithm*).
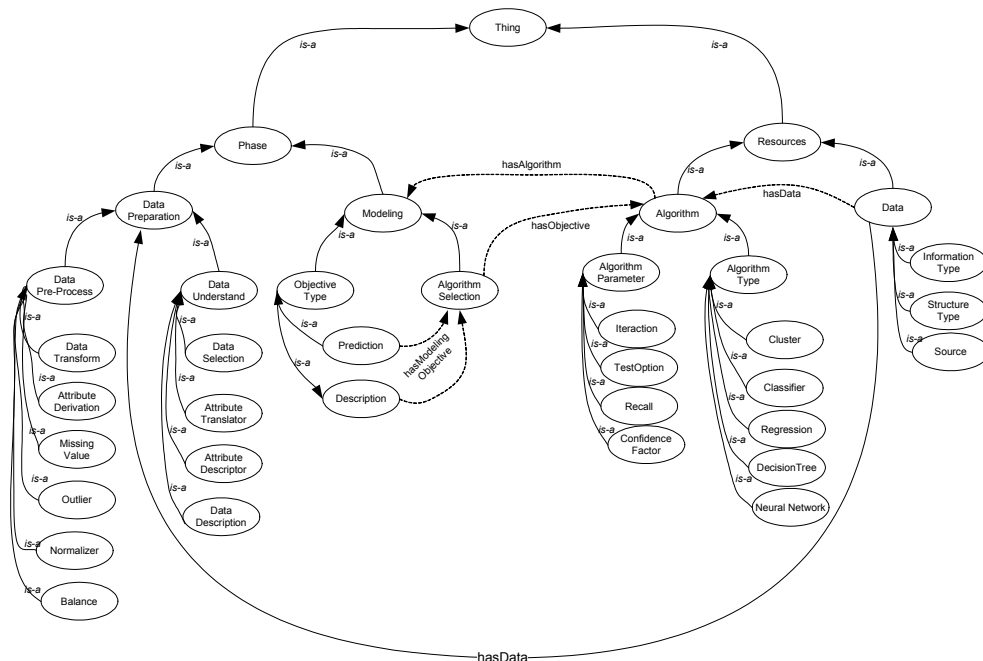

Fig. 7. KDD ontology class-properties hierarchy general view

7. Conclusions

During this work we have introduced process oriented ontology for database marketing knowledge based on Data Mining system architecture. Instead of imposing a fixed order for the DBM process, we have proposed a solution based on the ontologies and the knowledge extraction process. This approach is useful since it is used for end user assistance in the entire process development.

The proposed architecture defines, at different levels, a connection between ontology engineering and KDD process. It also defines a hybrid life cycle for the DBM process, based on both approaches. This life cycle that effectively assists the end-user, is composed by the knowledge extraction process phases and other specific marketing domain activities. Each phase is divided in tasks, directly or indirectly, related to ontology engineering, marketing and KDD.

This ontology is meant to be a subcomponent in the overall KDD process. Its usage of knowledge obtained from prior examples makes it applicable when several related databases are used.

Further work can be done in a variety of ways: this can be used for more specific knowledge extraction process or for more business oriented objectives. We believe that this approach convincingly addresses a pressing KDD need.

8. References

Arndt, D. and Gersten, W. (2001). Data management in analytical customer relationship management. In ECML/PKDD 2001 Workshop Proceedings. Data Mining For Marketing Applications.

Bean, R. (1999). Building a foundation for database marketing success. Technical report, DM Review Magazine.

Bellandi, A., Furletti, B., Grossi, V., and Romei, A. (2006). Ontology-driven association rule extraction: A case study. Contexts and Ontologies: Representation and reasoning in 16th European Conference on Artificial Intelligence, 1:1–10.

Berners-Lee, T., H., and J., Lassila, O. (2001). The semantic web. Scientific American, 284:34–43.

Berners-Lee, T. (2003). Www past and future. Technical report, W3C.

Bernstein, A., Provost, F., and Hill, S. (2005). Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification. IEEE Transactions on knowledge and data engineering, 17(4).

Blanco, I. J., Vila, M. A., and Martinez-Cruz, C. (2008). The use of ontologies for representing database schemas of fuzzy information. International Journal of intelligent Systems, 23:419–445.

Borges, A. M., Corniel, M., Gil, R., Contreras, L., and Borges, R. (2009). Towards a study opportunities recommender system in ontological principles-based on semantic web environment. WSEAS Transactions on Computers, 8(2):279–291.

Borst, P., Akkermans, H., and Top, J. (1997). Engineering ontologies. The International Journal of Human Computer Studies. In Special Issue: Using explicit ontologies in knowledge-based system development, Vol. 2/3 pp.365-406., 46(2-3):365–406.

Bolloju, N., Khalifa, M., and Turban, E. (2002). Integrating knowledge management into enterprise environments for the next generation decision support. Journal of Decision Support Systems, 33(2):163–176.

Bombardier, V., Mazaud, C., Lhoste, P., and Vogrig, R. (2007). Contribution of fuzzy reasoning method to knowledge integration in a defect recognition system. Journal of computers in industry, 58(4):355–366.

Bohling, T., Bowman, D., LaValle, S., Mittal, V., Narayandas, D., Ramani, G., and Varadarajan, R. (2006). Crm implementation: Effectiveness issues and insights. Journal of Service Research, 9(2):184–194.

Brezany, P., Janciak, I., and Tjoa, A. M. (2008). Data Mining with Ontologies: Implementations, Findings, and Frameworks, chapter Ontology-Based Construction of Grid Data Mining Workflows, pages 182–210. Information Science Reference - IGI Global.

Brito, P. Q. (2000). Como Fazer Promocao de Vendas. Mc Graw-Hill.

Brito, P. Q. and Hammond, K. (2007). Strategic versus tactical nature of sales promotions. Journal of Marketing Communications, 13(2):131–148.

Brito, P. Q., Jorge, A., and McGoldrick, P. J. (2004). The relationship between stores and shopping centers: Artificial intelligence and multivariate approach assessement. In Proceedings of Annual Conference of IABE 2004, Las Vegas, USA.

Buckinx, W. and den Poel, D. V. (2005). Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual fmcg retail setting. European Journal of Operational Research, 164 (1):252–268.

Buckinx, W., Verstraeten, G., and den Poel, D. V. (2007). Predicting customer loyalty using the internal transactional database. Expert Systems with Applications, 32:125–134.

Brookes, R. W., Brodie, R. J., Coviello, N. E., and Palmer, R. A. (2004). How managers perceive the impacts of information technologies on contemporary marketing practices: Reinforcing, enhancing or transforming? Journal of Relationship Marketing, 3(4):7–26.

Cannataro, M. and Comito, C. (2003). A data mining ontology for grid programming. In First International Workshop on Semantics in Peer-to-Peer and Grid Computing, in conjunction with WWW2003, pages 113–134.

Cardoso, J. and Lytras, M. (2009). Semantic Web Engineering in the Knowledge Society. Information Science Reference - IGI Global, New York.

Carson, D., Gilmore, A., and Walsh, S. (2004). Balancing transaction and relationship marketing in retail banking. Journal of Marketing Management, 20:431–455.

Ceccaroni, L. (2001). Ontoweeds - An ontology-based environmental decision-support system for the management of wastewater treatment plants. PhD thesis, Universitat Politecnica de Catalunya, Barcelona.

CeSpivova, H., Rauch, J., Svatek, V., and Kejkula, M. (2004). Roles of medical ontology in association mining crisp-dm cycle. In Knowledge Discovery and Ontologies.

Cellini, J., Diamantini, C., and Potena, D. (2007). Kddbroker: Description and discovery of kdd services. In 15th Italian symposium on Advanced Database Systems.

Cheng, H., Lu, Y.-C., and Sheu, C. (2009). Automated optimal equity portfolios discovery in a financial knowledge management system. Expert Systems with Applications, 36(2):3614–3622.

Cimiano, P., Hotho, A., Stumme, G., and Tane, J. (2004). Conceptual knowledge processing with formal concept analysis and ontologies. In ICFCA - Second International Conference on Formal Concept Analysis.

Coviello, N., Milley, R., and Marcolin, B. (2001). Understanding it-enabled interactivity in contemporary marketing. Journal of Interactive Marketing, 15(4):18–33.

Coviello, N., Winklhofer, H., and Hamilton, K. (2006). Marketing practices and performance of small service firms - an examination in the tourism accommodation sector. Journal of Service Research, 9(1):38–58.

Coviello, N. and Brodie, J. (1998). From transaction to relationship marketing: an investigation of managerial perceptions and practices. Journal of Strategic Marketing, 6(3):171–186.

Coulet, A., Smail-Tabbone, M., Benlian, P., Napoli, A., and Devignes, M.-D. (2008). Ontology-guided data preparation for discovering genotype-phenotype relationships. Journal of BMC Bioinformatics, 9:1–9.

den Poel, D. V. and Buckinx, W. (2005). Predicting online-purchasing behaviour. European Journal of Operational Research Dirk Van den Poel and Wouter Buckinx, 166(2):557–575.

DeTienne, K. B. and Thompson, J. A. (1996). Database marketing and organizational learning theory: toward a research agenda. Journal of Consumer Marketing, 13(5):12–34.

Diamantini, C., Panti, M., and Potena, D. (2004). Services for knowledge discovery in databases. In In Int. Symposium of Santa Caterina on Challanges in the Internet and Interdisciplinary Research SSCCII-04, volume 1.

Diamantini, C., Potena, D., and Cellini, J. (2006a). Uddi registry for knowledge discovery in databases services. In Proc. of AAAI Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition, pages 94–97, Arlington, VA, USA.

Diamantini, C., Potena, D., and Smari, W. (2006b). Collaborative knowledge discovery in databases: A knowledge exchange perspective. In Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition, pages 24–31, Arlington, VA, USA. AAAI, AAAI.

Domingos, P. (2003). Prospects and challenges for multi-relational data mining. SIGKDD Explorer Newsletter, 5(1):80–83.

Drozdenko, R. and Perry, D. (2002). Optimal Database Marketing. SAGE Publications, Thousand Oaks, USA.

El-Ansary, A. I. (2006). Marketing strategy: taxonomy and frameworks. European Business Review, 18:266–293.

Euler, T. and Scholz, M. (2004). Using ontologies in a kdd workbench. In ECAI-2004 Workshop on Ontology Learning and Population.

Farquhar, A., Fikes, R., and Rice, J. (1997). Tools for assembling modular ontologies in ontolingua. In In Proceedings of Association for the Advancement of Artificial Intelligence 97, pages 436–441. AAAI Press.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. In Magazine, A., editor, AI Magazine, volume 17, pages 37–54, Univ Calif Irvine, Dept Comp & Informat Sci, Irvine, Ca, 92717 Gte Labs Inc, Knowledge Discovery Databases Kdd Project, Tech Staff, Waltham, Ma, 02254. American Association for Artificial Intelligence.

Fensel, D., Horrocks, Harmelen, V., Decker, S., Erdmann, M., and Klein1, M. (2000). Oil in a nutshell. Proceedings of the 12th European Workshop on Knowledge Acquisition,

Modeling, and Management - EKAW'00 Lecture Notes in Artificial Intelligence, 1937:1–16.

Fletcher, K., Wright, G., and Desai, C. (1996). The role of organizational factors in the adoption and sophistication of database marketing in the uk financial services industry. Journal of Direct Marketing, 10:10–21.

Fox, M. and Gruninger, M. (1997). On ontologies and enterprise modelling. In Springer-Verlag, editor, Proceedings of International Conference on Enterprise Integration Modelling Technology. Springer- Verlag.

Frankland, D. (2007). How firms use database marketing services. Technical report with user interview data, ForresterResearch Inc.

Gronroos, C. (1994). From marketing mix to relationship marketing:towards a paradigm shift in marketing. Management Decision, 32(2):4–20.

Gomez-Perez, A., Fernandez-Lopez, M., and Corcho, O. (2004). Ontological engineering. Springer, 2nd edition.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5:199–220.

Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. International Journal of Human and Computer Studies, 43:625–640.

Guarino, N. (1998). Formal ontology and information systems. In Guarino, N., editor, FOIS'98 Formal Ontology in Information systems, pages 3–15, Amsterdam. IOS Press.

Jarrar, M. (2005). Towards Methodological Principles for Ontology Engineering. PhD thesis, Vrije Universiteit Brussel, Faculty of science.

Jasper, R. and Uschold, M. (1999). A framework for understanding and classifying ontology applications. In IJCAI 99 Ontology Workshop, pages 16–21.

Jurisica, I., Mylopoulos, J., and Yu, E. (1999). Ontologies for knowledge management: An information systems perspective. In for Information Sciences, A. S., editor, Proceedings of the Annual Conference of the American Society for Information Sciences (ASIS'99). American Society for Information Sciences.

Kasabov, N., Jain, V., Gottgtroy, P., Benuskova, L., Wysoski, S., and Joseph, F. (2007). Evolving brain-gene ontology system (ebgos): Towards integrating bioinformatics and neuroinformatics data to facilitate discoveries. In IEEE, editor, International Joint Conference on Neural Networks, Orlando - US.

Kopanas, I., Avouris, N. M., and Daskalaki, S. (2002). The Role of Domain Knowledge in a Large Scale Data Mining Project, volume 2308 of Lecture Notes in Computer Science, chapter Methods and Applications of Artificial Intelligence, pages 288–299. Springer Berlin / Heidelberg.

Leary, C. O., Rao, S., and Perry, C. (2004). Improving customer relationship management through database/internet marketing. European Journal of Marketing, 38(3/4):338–354.

Lin, C. and Hong, C. (2008). Using customer knowledge in designing electronic catalog. Expert systems with Applications, 34:119–127.

Lixiang, S. (2001). Data mining thechniques based on rough set theory. PhD thesis, National University of Singapore.

Lopez, M. F., Gomez-Perez, A., Sierra, J. P., and Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. IEEE Intelligent Systems Journal, 1:37–46.

Marsh, R. (2005). Drowning in dirty data. Database Marketing & Customer Strategy Management, 12(2):105–112.

McClymont, H. and Jocumsen, G. (2003). How to implement marketing strategies using database approaches. The Journal of Database Marketing & Customer Strategy Management, 11(2):135–148.

Michalewicz, Z., Schmidt, M., Michalewicz, M., and Chiriac, C. (2006). Adaptive Business Intelligence. Springer.

(Newell and level, 1982) Newell, A. and level, T. (1982). The knowledge level. Artificial Intelligence, 18:87–127.

(Neches et al., 1991) Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. R. (1991). Enabling technology for knowledge sharing. Artificial Intelligence Magazine, 12(3):36–56.

(Nigro et al., 2008) Nigro, H. O., Cisaro, S. G., and Xodo, D. (2008). Data Mining with Ontologies: Implementations, Findings and Frameworks. Information Science Reference. Information Science Reference - IGI Global, London, igi global edition.

Noy and McGuinness, 2003) Noy, N. F. and McGuinness, D. L. (2003). Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University.

Nogueira, B. M., Santos, T. R. A., and ZÃ¡rate, L. E. (2007). Comparison of classifiers efficiency on missing values recovering: Application in a marketing database with massive missing data. In Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007).

Ozimek, J. (2004). Case studies: The 2003 information management project awards. Journal of Database Marketing & Customer Strategy Management, 12(1):55.

Pearce, J. E., Webb, G. I., Shaw, R. N., and Garner, B. (2002). A systemic approach to the database marketing process. ANZMAC Conference Proceedings, 1:2941–2948.

Perez-Rey, D., Anguita, A., and Crespo, J. (2006). Biological and Medical Data Analysis, volume 4345/2006 of Lecture Notes in Computer Science, chapter OntoDataClean: Ontology-Based Integration and Preprocessing of Distributed Data, pages 262–272. Springer Berlin / Heidelberg.

Phillips, J. and Buchanan, B. G. (2001). Ontology-guided knowledge discovery in databases. In ACM, editor, International Conference On Knowledge Capture 1st international conference on Knowledge capture, pages 123–130. International Conference On Knowledge Capture

Pinto, F. (2006). A descoberta de conhecimento em bases de dados como suporte a actividades de business intelligence - aplicaÃ§Ã£o na Ã rea do database marketing. Master's thesis, Universidade do Minho.

Pinto, F., Santos, M. F., and Marques, A. (2009). WSEAS Transactions on Business and Economics, volume 6, chapter Database marketing intelligence supported by ontologies, pages 135–146. World Scientific and Engineering Academy and Society.

Quine, W. V. (1992). Theories and Things (Revised ed.). Harvard University Press.

Quinlan, R. (1986). Induction of decision trees. Machine Learning, 1(1):81–106.

Rebelo, C., Brito, P. Q., Soares, C., and Jorge, A. (2006). Factor analysis to support the visualization and interpretation of clusters of portal users. In WI 2006: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pages 987–990, Washington, DC, USA. IEEE Computer Society.

Rothenfluh, T. E., Gennari, J. H., Eriksson, H., Puerta, A. R., Tu, S. W., and Musen, M. A. (1996). Reusable ontologies, knowledge-acquisition tools, and performance systems: Protege-ii solutionsto sisyphus-2. International Journal of Human-Computer Studies 44: 303-332., 44:303–332.

Santos, M. F., Cortez, P., Quintela, H., and Pinto, F. (2005). A clustering approach for knowledge discovery in database marketing. Data Mining VI: Data Mining, Text Mining and their Business Applications, 35:399–407.

Seller, M. and Gray, P. (1999). A survey database marketing. Technical report, Center for Research on information Technology and Organizations.

Sen, S. and Tuzhiln, A. (1998). Making sense of marketing data: Some mis perspectives on the analysis of large data sets. Journal of Market Focused Management, 3:91–111.

Sharma, S. and Osei-Bryson, K.-M. (2008). Framework for formal implementation of the business understanding phase of data mining projects. Expert Systems with Applications, page in press.

Shepard, D. (1998). The New Direct Marketing: How to Implement A Profit-Driven Database Marketing Strategy. David Shepard Ass, 3rd edition.

Schoenbachler, D. D., Gordon, G. L., Foley, D., and Spellman, L. (1997). Understanding consumer database marketing. Journal of Consumer Marketing, 14(1):5–19.

Smith, R. G. and Farquhar, A. (2008). The road ahead for knowledge management: An ai perspective. American Association for Artificial Intelligence, 1:17–40.

Sowa, J. F. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co.

Staab, S. and Studer, R. (2004). Handbook on ontologies. International handbooks on information systems. Springer-Verlag.

Swartout, B., Patil, R., Knight, K., and Russ, T. (1996). Ontosaurus: A tool for browsing and editing ontologies. Knowledge Acquisition Workshops, 1.

Tao, Y.-H. and Yeh, C.-C. R. (2003). Simple database marketing tools in customer analysis and retention. International Journal of Information Management, 23:291–301.

Tudorache, T. (2006). Employing Ontologies for an Improved Development Process in Collaborative Engineering. PhD thesis, University of Berlim, Germany.

van Heijst, G., Schreiber, A. T., and Wielinga, B. J. (1997). Using explicit ontologies in kbs development. International Journal of Human-Computer Studies, 46(2):183–292.

Verhoef, P. and Hoekstra, J. (1999). Status of database marketing in the dutch fast moving consumer goods industry. Journal of Market Focused Management, 3:313â€"331.

Wehmeyer, K. (2005). Aligning it and marketing - the impact of database marketing and crm. Journal of Database Marketing & Customer Strategy Management, 12(2):243.

Welty, C. and Murdock, J. W. (2006). Towards knowledge acquisition from information extraction. In Springer, editor, In Proceedings of ISWC-2006., Athens.

Weng, S.-S. and Chang, H.-L. (2008). Using ontologies network analysis for research document recommendation. Expert Systems with Applications, 34:1857–1869.

Witten, I. H. and Frank, E. (2000). Data Mining: Practical Machine Learning Tools and Technique. The Morgan Kaufmann Series in Data Management Systems, 2nd edition.

Zairate, L. E., Nogueira, B. M., Santos, T. R. A., and Song, M. A. J. (2006). Techniques for missing value recovering in imbalanced databases: Application in a marketing database with massive missing data. In IEEE International Conference on Systems, Man, and Cybernetics, pages 2658–2664, Taiwan. IEEE.

Zhou, X., Geller, J., and Halper, Y. P. M. (2006). An Application Intersection Marketing Ontology, chapter Theoretical Computer Science, pages 143–163. Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Zineldin, M. and Vasicheva, V. (2008). Cybernization management in the cyber world: a new management perspective. Problems and Perspectives in Management, Volume 6, Issue 1, 2008, 1:113–126.

Zwick, D. and Dholakia, N. (2004). Whose identity is it anyway? consumer representation in the age of database marketing. Journal of Macromarketing, 24(1):31–43.
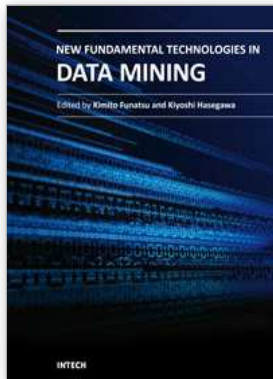
New Fundamental Technologies in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-547-1

Hard cover, 584 pages

Publisher InTech

Published online 21, January, 2011

Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Filipe Mota Pinto and Teresa Guarda (2011). Database Marketing Process Supported by Ontologies: a Data Mining System Architecture Proposal, New Fundamental Technologies in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, InTech, Available from: http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/database-marketing-process-supported-by-ontologies-a-data-mining-system-architecture-proposal

INTECH
open science | open minds