Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Ontologies, CRM, Data Mining: How to integrate?

M. Gottgtroy[1] & P. Gottgtroy[2]
[1]*Information Systems and Operations Management Department Centre of Digital Enterprise, University of Auckland, New Zealand*
[2]*Knowledge Engineering and Discovery Institute, Auckland University of Technology, New Zealand*

Abstract

Customer Relationship Management (CRM) systems rely in providing better information services to a population of end users/customers. A typical service of a CRM solution is to provide to a consumer access to a FAQ system to get advice on commonly occurring product problems. The CRM solution also typically supports direct communication with users, such as email communication, allowing the users to make requests for information via emails and, in some cases, receiving automatically generated responses to those emails. Providing information to users implies in access, and integration of diverse and distributed sources of information. On the other hand, the information collected from theses sources and from user's interactions should be of great value if integrated in a data mining process to maintain the current users, prospect new ones, and reaching some competitive advantage among competitors.

Ontologies have been used to provide semantic integration and reasoning in decision support systems and knowledge management activities. The ontology captures the intrinsic conceptual structure of a domain. For any given domain, its ontology forms the heart of the knowledge representation.

Our goal in this paper is to analyse some of the main opportunities to build a system that leverages the semantic content of ontologies in order to improve services provided by CRM solutions, and use this rich information environment to apply knowledge discovery techniques to reach a competitive advantage.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

1 Introduction

Customer Relationship Management (CRM) systems rely in providing better information services to a population of end users/customers. Furthermore, customer relationship is a conceptual management problem that should be analyzed as a relation between a company's value proposition, target customer segments, distribution channels and the actual customer interaction [1]. In summary, CRM focus on enhancing the relationship by collecting and analyzing data coming from the diverse range of interactions with the customers. It is well-known that the KDD process plays a very important role in providing a conceptual vision of the business and its relationship with clients, accessing and integrating diverse and distributed sources of information.

Ontologies [2] have been used to allow semantic integration and reasoning in decision support systems and knowledge management activities. The ontology captures the intrinsic conceptual structure of a domain, forming the heart of the knowledge representation, for any given domain. The ontology is a representation of a common vocabulary of an area and defines the meaning of the terms and the relations between them. Ontologies can be constructed using different levels of formality and representing different levels and types of knowledge within a domain.

Our goal in this paper is to analyze some of the opportunities and main requirements to build a system that leverages the semantic content of ontologies in order to improve services provided by CRM solutions, and use this rich information environment to apply knowledge discovery techniques to achieve competitive advantage.

The first part of this paper presents an overview of ontologies and its application on CRM. In the second part we describe some approaches that focus on the integration of ontologies and knowledge discovery from data. After that we present an analysis of the inclusion of an ontology-based approach during in a KDD process considering the evolving CRM scenario. Future directions of our research integrating ontology and different business scenarios in the context of knowledge representation, business analysis and knowledge discovery conclude this work.

2 Ontologies/CRM

Customer relationship refers to the way a firm goes to market, how it actually reaches its customers and how it interacts with them. It includes the people, processes, and technology; questions associated with marketing, sales, and services. In today's competitive world, organizations looking to implement successful CRM strategies need to focus on a *unified view* of the customer using integrated information systems and services that allow the customer to interact via any desired communication channel.

Ontologies in current computer science language are computer based resources that represent agreed domain semantics [3]. Unlike data models, the fundamental asset of ontologies is their relative independence of particular

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

applications, i.e. ontology consists of relatively generic knowledge that can be reused by different kinds of applications/tasks.
In the following paragraphs we present two main scenarios and examples where ontologies are applied in CRM applications: a framework for business analysis and ontology integrating different sources of information.

2.1 Ontology as a business analysis tool

Pigneur and Osterwalder [4] define customer relationship as an integral part of a business analysis tool. In their proposal the business model is an ontological approach that can be broken down into four simple pillars, which are the "what", the "who", the "how" [5] and the "how much" of a firm. In other words, these pillars allow to express *what* a company offers, *who* it targets with this, *how* this can be realized and *how much* can be earned by doing it.

As shown in Figure 1 [1], the proposed CRM ontology covers all customer related aspects.

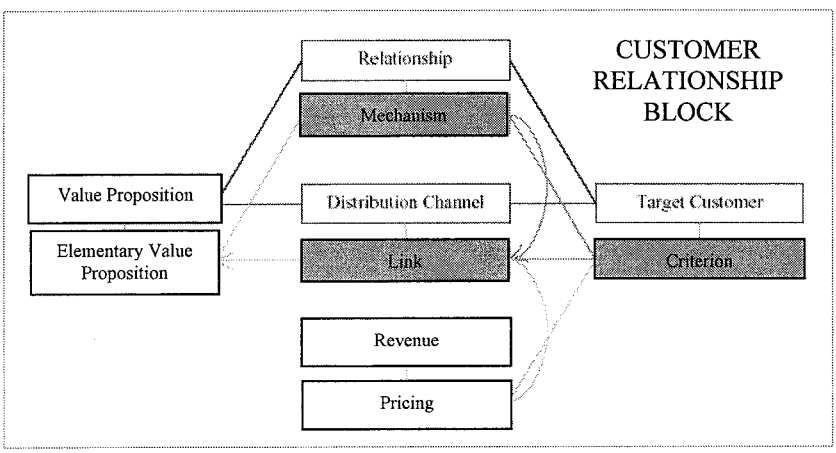

Figure 1: CRM ontology model.

The Customer Relationship Ontology helps managers formally define which customer segments they want to offer value, through which channels they want to do this and by establishing what relationship. By using this ontological approach, customer relationships become communicable, comparable, analyzable and easily modifiable because of the building-block- like structure. On the one hand this should improve communication between managers, middle managers, process modellers and Information Systems staff. On the other hand this should foster innovation in customer relationship management, because modellers can easily "play around" and navigate trough the ontology representation.

2.2 Ontology as an integration tool

CRM brings together processes, people, and business perspectives, among other from diverse source of information. Furthermore the amount of data produced daily by these information systems grows at a staggering rate. Thus the main challenge for CRM applications is "integration" of these diverse and dynamic sources of data.

Product Ontologies [6] are a good example of this ontology integration approach. Catalog of products can be represented in various formats and levels of detail. Each department has its own set of interest in product information as well as contributes with different information to be provided for customers, i.e., financial department is more interested in price, orders, than any other department.

We can add another level of complexity to this scenario considering that a company has a supply chain for its products, probably with more than one supplier, and this chain should be integrated in a transparent way for the customers. For example, in the IT service industry scenario where each individual participant potentially can use his own format to represent the products (Figure 2) in his catalogue, and these catalogues must be integrated to help customers to find one product or configure it.

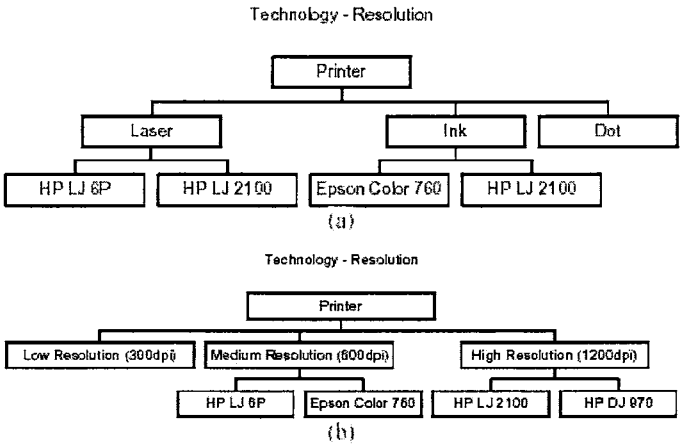

Figure 2: Product ontology [6].

It is already well accepted that customer relationship systems can help managers to optimize channels, improve customer acquisition, retention and add-on selling. However, the advantage of an ontology approach for customer relationship are beyond of providing better services and a better understanding of the way a company interacts with its clients. A unified conceptual representation brings an integrated vision of the business model which allows us to express and share the business logics and its interaction among different departments and stakeholders. Further, this sharable knowledge representation permits a transparent integration among different sources of information.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

3 Ontology/Data Mining

The current interest in ontologies is the latest version of Artificial Intelligence's alternation of focus between content theories and mechanism theories [7]. Sometimes, the Artificial Intelligence community gets excited by some mechanism such as rule systems, frame languages, neural nets, fuzzy logic, constraint propagation, or unification. The mechanisms are proposed as the secret of making intelligent machines. At other times, we realize that, however wonderful the mechanism is, it cannot do much without a good content theory of the domain on which it is to work. Moreover, we often recognize that once good content theory is available, many different mechanism might be used equally well to implement effective systems, all using essentially the same content [8].

In the mean time, recent progress in data mining research has led to the development of numerous efficient and scalable methods for mining interesting patterns and knowledge in large databases [9, 10]. Artificial Neural Network (ANN) [11] is popular in the field of knowledge discovery in databases (KDD). However, an impediment to the more widespread acceptance of ANN's is the absence of a capability to explain to the user, in a human-comprehensible form, how the network arrives at a particular decision or which relationships can be reused in a different context [12].

In order to keep the knowledge domain up to date, sharable, and reusable for different applications, hybrid approaches putting together the state of art of the AI methods for knowledge discovery in large databases (KDD) and the ontology engineering are being investigated.

In the following items we present two approaches that show initial results of integrating both mechanisms and content theories to improve the KDD process.

3.1 Ontologies in recommender systems

Middleton et al [13] explore the acquisition of user profiles by unobtrusive monitoring of browsing behaviour and application of supervised learning techniques coupled with an ontological representation to extract user preferences. They use ontology to investigate how domain knowledge can help in the acquisition of user preferences and evaluate it in a real work setting, measuring the effectiveness of using a hierarchical topic ontology compared with an extendable flat list. The results suggest how using an ontology in the profiling process results in superior performance over using a flat list of topics.

Additionally, visualizing their preferences trough the ontology the users could build a better conceptual model (understanding) of the system, helping to engender a feeling of control and eventually trust in the system.

Although these preliminary results need to be extended so as to enable the application of more rigorous analysis, the trend shows the utility of ontologies in recommender systems.

3.2 Ontologies in biomedical discovery

Kasabov et al [14] investigate the explosion of biomedical data and the growing number of disparate biomedical data sources building a hybrid system that aims to solve the challenge of how to acquire, maintain and share knowledge from the large and distributed databases in the context of this rapidly evolving research.
In addition to the "Conceptual Biology" challenge, to build a knowledge repository capable of moving the current data collection era into one of hypothesis–driven, experimental research, the authors suggest an approach that adds to the literature fact knowledge acquired from a KDD process.

Their work focus on the use of ontologies in the KDD's initial and final stages[15]. The data preparation phase is preceded by an ontology analysis of the domain and the data mining phase is followed by an update of the new patterns and the discovered knowledge into the ontology.

Infogene Map [14] is the first result of this approach that aims to build a multi-dimensional ontology able to share knowledge from different experiments undertaken across aligned research communities in order to connect areas of science seemingly unrelated to the area of immediate interest.

4 Ontologies, CRM, Data Mining

The previous examples show some of the on going projects that use ontology approach in order to enhance semantically the results of a KDD process and CRM systems. In this session we are using the steps of a KDD process, shown in the Figure 4, to analyze the inclusion of an ontology-based approach in the evolving CRM scenario. As in [16] we do not presume that ontology is complete at the time a new data mining application is started. In contrary, we believe that new domains will bring new types of variables and knowledge about them. However, we also believe that data mining is not simply the one-time application of a program to a new database. In our own work, a data mining process frequently starts with small pilot studies and manual bias space search, including feature construction. With preliminary confirmation that the programs can find some interesting relationships, more data and greater expectations are introduced.

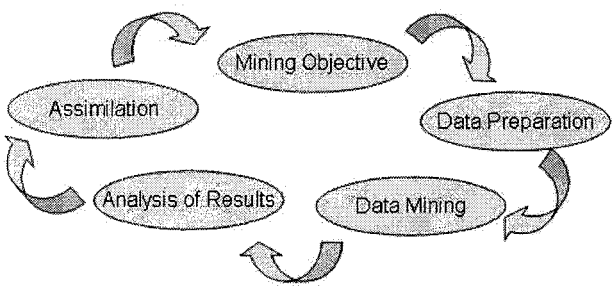

Figure 3: Steps of a KDD process.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

4.1 Mining objective determination

Goal - understanding of the application domain & formulating the goals.
Ontologies can be used as shown in section 2.1 to enhance communication between managers, middle managers, process modellers and Information Systems staff in order to build better business scenarios and draw new strategies to archive competitive advantage among competitors and provide better services to customers. Being able to link the different business scenarios with a more general and formal business model allows better and more productive goal-facts selection.

4.2 Data preparation

Ontology brings its best features to this KDD phase in CRM applications. We should use ontology to investigate how domain knowledge can help in the acquisition of data for the data mining phase.

4.2.1 Collection & integration of data
Goal - combining multiple data sources in a common source.
Ontology is well accepted as a powerful tool for data integration and mediation. Combining this feature with the main CRM goal - integration of diverse, distributed and dynamic sources of data – is a natural path.

4.2.2 Data cleaning
Goal - removing noise data and irrelevant data from the collection.

4.2.3 Data selection & transformation
Goal - find useful features, dimensionality/variable reduction ...
The domain ontology – structural knowledge of the domain could be used in both previous steps to guide the processes and prepare the data for the next data mining phase.

There is not much information presented in the literature that proves the benefits of ontologies in these phases. Besides that, some old questions as in [7], and the experiments described in section 3.1 and 3.2, show a new area for application of ontologies in KDD. The statistical results presented in the recommender system, section 3.1, aligned with the qualitative results need for the biomedical discovery research represented in the preliminarily results of Infogene Map, section 3.2., show the feasibility of integrating these three areas CRM/Ontologies/KDD. And in our opinion the applicability of an Ontology/KDD approach to CRM is well suited to the data preparation phase and can bring answers to many current KDD questions as well as face the researchers to new challenges.

4.3 Data mining

Goal - choose data mining techniques and search for patterns of interest.
Ontologies can be used in this phase to support the decision makers in which technique is more appropriate to a specific domain or problem. A similar

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

approach for problem solving methods is presented in [17] and is one of the core functionalities of the Protégé knowledge-base systems development environment[18].

4.4 Analysis of results

Goal - visualization, transformation, removing redundant patterns.
Ontology visualisation allows a combination of user customization interface with a deep semantic analysis to the data visualization allowing a multi-dimensional analysis of the patterns and discovered results.

4.5 Assimilation

Goal - use of discovered knowledge.
From a philosophical point of view, discovery can be defined as "the act of becoming aware of something previously existing but unknown"[19]. This broad definition includes both kinds of scientific discovery: factual and conceptual. The former typically happens during the investigation of current "known" facts or data. The latter emerges from different points of view of "unknown" facts or data, and frequently finishes with a paradigm shift. Thus, for scientific discoveries "imagination" is essential as well as reasoning. An Ontology-based approach can be used to facilitate both forms of scientific discovery by providing a common framework for several systems and problem solving methods; it allows experimentation, supporting imagination exercise.

5 Conclusions

We departed from the point of view that the key to successfully managing customer relationships lies in the ability to integrate existing sources of information and processes in order to provide better services, optimize channels, improve customer acquisition, retention and add-on selling. Additionally that KDD is an essential tool to achieve this success. Based on that, we presented ontology as an essential and viable tool for data integration and mediation.

Besides of its well known potential in integrating diverse and distributed sources of information, an ontology representing the business model of the organization can be a useful and a handing tool for managers. An ontology-based model can guide the visualization of the relationships among the different concepts and views that compose their business. It can guide the usage of knowledge discovery and other important supportive tools that allows good use of the available information to achieve profits and competitiveness. An important add-in is the possibility of not only analysing the information already available, but to decide what information is missing and how to acquire it - which are the main channels that could provide the required information to the organization.

Furthermore, in the section Ontology/KDD, we presented some on going projects facing the challenge of how to integrate ontologies and knowledge discovery from data in order to use this rich information environment to apply knowledge discovery techniques to reach a competitive advantage.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

The analysis in the section 4 follows one KDD life cycle tracing in each phase the potential use of ontologies in a CRM context. This analysis is a basis to guide further investigation, discussion and development in how to sieve these three rich fields, customer relationship management, knowledge discovery in databases, and ontology engineering to support successful business applications. Our on going research aims to analyze the application of ontology engineering in different real business scenarios, making use of the available business intelligence strategies to face CRM challenges. As a machine learning powerful resource our goal is to integrate KDD, ontology engineering and software agents using a business ontology-model as a communication platform and a common framework used by software agents to integrate different business applications, including those that bring additional e-commerce and decision support systems requirements.

In the current stage of our research we are investigating CRM, and Supply Chain businesses aspects and how to represent and integrate these scenarios in the context of e-business using the semantic web cyberspace.

References

[1] Osterwalder, A. & Pigneur Y. (2003). Modelling Customer Relationships in e-Business. In 16th Bled Electronic Commerce Conference eTransformation. 2003. Bled, Slovenia.

[2] Gruber, T. (2002). What is an Ontology? Retrievable from the internet 28/01/2002 at: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

[3] Sowa, J.F. (2001). Ontology, Metadata, and Semiotics. Retrievable from the internet 08/07/2003 at: http://users.bestweb.net/~sowa/peirce/ontometa.htm.

[4] Pigneur, Y. (2002). A framework for defining e-business models. In J. Bruel et al. (Eds) Advances in Object-Oriented Information Systems (OOIS 2002) Montpellier, Lectures Notes on Computer Sciences 2426, September 2003.

[5] Markides, C. (1999). All the Right Moves. Boston, Harvard Business School Press.

[6] Omelayenko, B. (2000). Integration of product ontologies for B2B marketplaces: a preview. ACM SIGecom Exchanges, 2000. **2**(1): p. 19-25.

[7] Chandrasekaran, B.J., et al. (1999). What are ontologies, and why do we need them? IEEE Intelligent Systems, 1999, 14(1094-7167): 20-26.

[8] Chandrasekaran, B. (1994). AI, knowledge, and the quest for smart systems. Expert. IEEE Intelligent Systems, 1994, 9(0885-9000): 2-5.

[9] Han, J. (2002). How can data mining help bio-data analysis? In Proceedings of the SIGKDD02 Conference - BIOKDD02: Workshop on Data Mining in Bioinformatics. Retrievable from the internet 28/01/2002 at: http//:www.cs.rpi.edu/~zaki/BIOKDD02/01-han.pdf.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

[10] Kasabov, N. (2002). Evolving connectionist systems for adaptive learning and knowledge discovery: methods, tools, applications. IEEE Intelligent Systems, 2002, 24-28.

[11] Kasabov, N.K. (1996). Foundations of neural networks, fuzzy systems, and knowledge engineering. Cambridge, Mass.: MIT Press. xvi, 550.

[12] Mitra, S., & Hayashi Y. (2000). Neuro-fuzzy rule generation: survey in soft computing framework. IEEE Transactions on Neural Networks, 2000, 11(3): 748-768.

[13] Middleton, S. et all (2001). Capturing knowledge of user preferences: ontologies in recommender systems. In Proceedings of the international conference on Knowledge capture. Victoria, British Columbia, Canada ISBN:1-58113-380-4.p 100 – 107.

[14] Kasabov, N., Gottgtroy, P.C.M. & MacDonell, S. (2003). Infogene Map - Bridging the gap between biomedical data and data mining. Submitted to III Brazilian Symposium of Mathematical and Computational Biology.

[15] Gottgtroy, P.C.M., Kasabov, N. & MacDonell, S. (2003). An ontology engineering approach for knowledge discovery from data in evolving domains. In Proceedings of Data Mining 2003. Rio de Janeiro, Brazil. In press.

[16] Phillips, J. et all. (2001) Ontology-guided knowledge discovery in databases. In Proceedings of the international Conference on Knowledge Capture, Victoria, British Columbia, Canada.

[17] Crubézy, M. and Musen, M.A. (In press). Ontologies in Support of Problem Solving. Handbook on Ontologies in Information Systems. S. Staab and R. Studer, Springer. Retrievable from the internet 05/09/2003 at: http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-2003-0952.html.

[18] Protégé. http://protege.stanford.edu.

[19] Noé, K. (2002) The Structure of Scientific Discovery: from a Philosophical Progress in Discovery Science 2001, Berlin, Springer-Verlag, p 31-39.