

Data Mining for Customer Relationship Management (CRM)

Jaideep Srivastava
srivasta@cs.umn.edu

1 Introduction

Data Mining has enjoyed great popularity in recent years, with advances in both research and commercialization. The first generation of data mining research and development has yielded several commercially available systems, both stand-alone and integrated with database systems; produced scalable versions of algorithms for many classical data mining problems; and introduced novel pattern discovery problems.

In recent years, research has tended to be fragmented into several distinct pockets without a comprehensive framework. Researchers have continued to work largely within the parameters of their parent disciplines, building upon existing and distinct research methodologies. Even when they address a common problem (for example, how to cluster a dataset) they apply different techniques, different perspectives on what the important issues are, and different evaluation criteria. While different approaches can be complementary, and such a diversity is ultimately a strength of the field, better communication across disciplines is required if Data Mining is to forge a distinct identity with a core set of principles, perspectives, and challenges that differentiate it from each of the parent disciplines. Further, while the amount and complexity of data continues to grow rapidly, and the task of distilling useful insight continues to be central, serious concerns have emerged about social implications of data mining. Addressing these concerns will require advances in our theoretical understanding of the principles that underlie Data Mining algorithms, as well as an integrated approach to security and privacy in all phases of data management and analysis.

2 Customer Relationship Management (CRM)

Customer Relationship Management (CRM) emerged in the last decade to reflect the central role of the customer for the strategic positioning of a company. CRM takes a holistic view over customers. It encompasses all measures for understanding the customers and for exploiting this knowledge to design and implement marketing activities, align production and coordinate the supply-chain. CRM puts emphasis on the coordination of such measures, also implying the integration of customer-related data, meta-data and knowledge and the centralized planning and evaluation of measures to increase customer lifetime value. CRM gains in importance for companies that serve multiple groups of customers and exploit different interaction channels for them. This is due to the fact that information about the customers, which can be acquired for each group and across any channel, should be integrated with existing knowledge and exploited in a coordinated fashion.

It should be noted, however, that CRM is a broadly used term, and covers a wide variety of functions, not all of which require data mining.¹ These functions include *marketing automation* (e.g., campaign

¹ Customer Relationship Management (CRM) as a product/market segment was first introduced by companies such as Siebel and Oracle, with many other players, including SAP, PeopleSoft and Microsoft joining subsequently. The initial set of products mostly support easy management of information for customer facing functions, including *contact management*, *sales force automation*, etc. Applying data mining to better understand customers, and its use for relationship management, constitutes a more recent phenomenon.

management, cross- and up-sell, customer segmentation, customer retention), *sales force automation* (e.g., contact management, lead generation, sales analytics, generation of quotes, product configuration), and *contact center management* (e.g., call management, integration of multiple contact channels, problem escalation and resolution, metrics and monitoring, logging interactions and auditing), among others. We focus on how backend data mining and analytics can make these functions more effective.

A Note on Personalization: While personalization and CRM are often spoken of in the same breath, it is important to note that the two are distinct, and each can exist without the other. Specifically, one could have customer relationship management without personalization, i.e., one in which the manner in which relationship management is done is not sensitive to the needs of individual (or groups of) customers. Furthermore, non-personalized CRM does not necessarily indicate bad or unsatisfactory service; it just means that ‘everyone is treated the same’. For example, Domino’s Pizza has (or at least used to have) a ‘30 minute or free’ guarantee on Pizza delivery, regardless of who was placing the order. The service itself was rated very high on customer satisfaction, but it was nonetheless non-personalized.

On the other hand, personalization can be for purposes other than customer relationship management. For example, analysis of an auto insurance applicant’s driving and credit history to increase insurance rates, or even cancel the insurance is an example of personalization; albeit in a rather negative sense. Current application of traveler profiling and analysis by the Transportation Security Administration (TSA) in the United States is another such example.

2.1 Analysis of Data Mining in CRM

Problem Context: The maximization of lifetime values of the (entire) customer base in the context of a company's strategy is a key objective of CRM. Various processes and personnel in an organization must adopt CRM practices that are aligned with corporate goals. For each institution, corporate strategies such as diversification, coverage of market niches or minimization of operative costs are implemented by "measures", such as mass customization, segment-specific product configurations etc. The role of CRM is in supporting customer-related strategic measures.

Customer understanding is the core of CRM. It is the basis for maximizing customer lifetime value, which in turn encompasses customer segmentation and actions to maximize customer conversion, retention, loyalty and profitability. Proper customer understanding and actionability lead to increased customer lifetime value. Incorrect customer understanding can lead to hazardous actions. Similarly, unfocused actions, such as unbounded attempts to access or retain *all* customers, can lead to decrease of customer lifetime value (law of diminishing return). Hence, emphasis should be put on correct customer understanding and concerted actions derived from it.

Figure 1 shows an idealized CRM cycle.

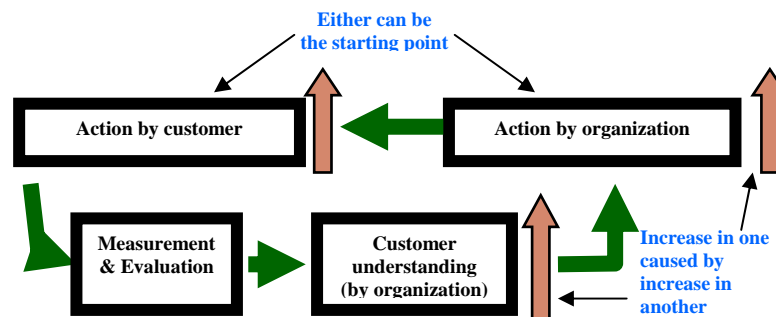


Figure 1. The Basic CRM Cycle.

In this figure, boxes represent actions:

- The customer takes the initiative of contacting the company, e.g. to purchase something, to ask for after sales support, to make a reclamation or a suggestion etc.
- The company takes the initiative of contacting the customer, e.g. by launching a marketing campaign, selling in an electronic store or a brick-and-mortar store etc.
- The company takes the initiative of understanding the customers by analyzing the information available from the other two types of action. The results of this understanding guide the future behaviour of the company towards the customer, both when it contacts the customer and when the customer contacts it.

The reality of CRM, especially in large companies, looks quite different from the central coordination and integration suggested by Figure 1:

- Information about customers flows into the company from many channels, but not all of them are intended for the acquisition of customer-related knowledge.
- Information about customers is actively gathered to support well-planned customer-related actions, such as marketing campaigns and the launching of new products. The knowledge acquired as the result of these actions is not always juxtaposed to the original assumptions, often because the action-taking organizational unit is different from the information-gathering unit. In many cases, neither the original information, nor the derived knowledge are made available outside the borders of the organizational unit(s) involved. Sometimes, not even their existence is known.
- The limited availability of customer-related information and knowledge has several causes. Political reasons, e.g. rivalry among organization units, are known to lead often in data and knowledge hoarding. A frequently expressed concern of data owners is that data, especially in aggregated form, cannot be interpreted properly without an advanced understanding of the collection and aggregation process. Finally, confidentiality constraints, privacy considerations and law restrictions often disallow the transfer of data and derived patterns among departments.
- In general, one must assume that data gathered by an organization unit for a given purpose cannot be exported unconditionally to other units or used for other purpose and that in many cases such an export or usage is not permitted at all.
- Hence, it is not feasible to strive for a solution that integrates all customer-related data into a corporate warehouse. The focus should rather be in mining non-integrated, distributed data while preserving privacy and confidentiality constraints.²

A more realistic picture of current CRM, incorporating the typical flow of information, is shown in Figure 2:

- Data is collected from multiple organizational units, for different purposes, and stored in multiple locations, leading to redundancies, inconsistencies and conflicting beliefs.
 - No organization unit has access to all data and to all derived knowledge.
 - Some data are not analyzed at all.
 - Not all background knowledge is exploited.
 - Data analysis is performed by many units independently.
 - Some analyses do not mount to actions.

Ideally, the CRM cycle should encompass:

- the exploitation of all data and background knowledge
- the coordination of analyses, and resulting actions, in different parts of the organization

² As these observations indicate, CRM solutions address a mix of technical and organizational issues. Some of the problems listed above can be solved by making top level management directly involved in the CRM initiatives. We focus our discussion on the technical aspects, while noting that significant organizational challenges must also be addressed in any real CRM undertaking.

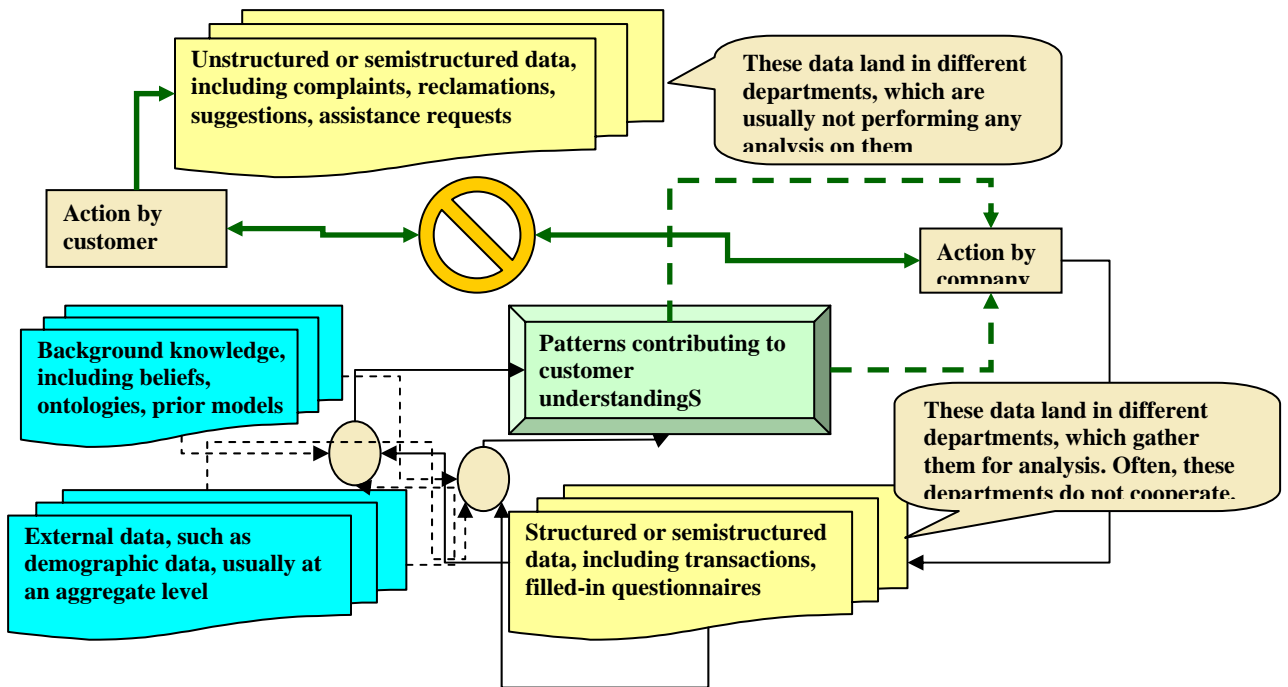


Figure 2. Expanded CRM Cycle – Current State.

Gap Analysis: The grand KDD challenges in CRM arise from the objectives of exploiting all information and coordinating analysis and actions. These objectives require methods to deal with several specific challenges, which we discuss in turn:

- *Cold start:* In CRM, one tries to influence customer behavior on the basis of prior knowledge. Often, there is no (reliable) such prior knowledge.
- *Correct vs. incorrect customer understanding:* CRM is about understanding the customer. It's about time to elaborate on the impact of *misunderstanding* the customer, and to fold this into our analyses and action workflows.
- *Data sovereignty:* There is no such thing as a CRM data warehouse; we are faced with multiple data sources. If and when problems of semantic disparity are solved, we will still face legislative and the political hurdles.³ We need solutions where the data owner is allowed to specify what data she wants to deliver, at what level of abstraction, and with what meta-information.
- *Data quality:* Some Customer-Company-Interaction channels deliver very good data. Others deliver notoriously poor quality data; web server logs belong to the latter category. The issue of data sovereignty impedes integration of raw data. Despite these odds, data must be enhanced to ensure that the results of the analysis are reliable.
- *Deeper understanding:* Profiling is based on some rudimentary behavioral data and some preferences, carefully extracted from questionnaires or learned from the data using data mining methods. Integration of cultural and psychological data is at its infancy. The experts in those domains come from outside of the KDD community (marketing researchers, practitioners, etc.) and we should establish collaborative relationships with them.
- *Questioning prior knowledge:* Everything associated with prior knowledge is assumed correct. We need mechanisms that capture and align prior knowledge in the face of conflicting information.
- *Actionability:* Pattern discovery should lead to actions. In some cases, this is straightforward, e.g., site customization and personalization, but this step is often omitted. We need mechanisms that

³ These are organizational and legal challenges that can be resolved in one way or another. However, this is a separate (non-technical) dimension of CRM.

incorporate patterns into the action-taking processes in a seamless way. We also need an understanding of the action-taking processes and their influence on what patterns are most valuable.

3 Data Mining Challenges & Opportunities in CRM

In this section, we build upon our discussion of CRM and Life Sciences to identify key data mining challenges and opportunities in these application domains. The following is a list of challenges for CRM:

- **Non-trivial results almost always need a combination of DM techniques.** Chaining/composition of DM, and more generally data analysis, operations is important. In order to analyze CRM data, one needs to explore the data from different angles and look at its different aspects. This should require application of different *types* of DM techniques and their application to different “*slices*” of data in an interactive and iterative fashion. Hence, the need to use various DM operators and combine (chain) them into a single “exploration plan”.
- **There is a strong requirement for data integration before data mining.** In both cases, data comes from multiple sources. For example in CRM, data needed may come from different departments of an organization. Since many interesting patterns span multiple data sources, there is a need to integrate these data before an actual data mining exploration can start.
- **Diverse data types are often encountered,** which requires the integrated mining of diverse and heterogeneous data. In CRM, while dealing with this issue is not critical, it is nonetheless important. Customer data comes in the form of structured records of different data types (e.g., demographic data), temporal data (e.g., weblogs), text (e.g., emails, consumer reviews, blogs and chat-room data), (sometimes) audio (e.g., recorded phone conversations of service reps with customers).
- **Highly and unavoidably noisy data must be dealt with.** In CRM, weblog data has a lot of “noise” (due to crawlers, missed hits because of the caching problem, etc.). Other data pertaining to customer “touchpoints” has the usual cleaning problems seen in any business-related data.
- **Privacy and confidentiality considerations for data and analysis results** are a major issue. In CRM, lots of demographic data is highly confidential, as are email and phone logs. Concern about inference capabilities makes other forms of data sensitive as well—e.g., someone can recover personally-identifiable information (PII) from web logs.
- **Legal considerations influence what data is available for mining and what actions are permissible.** In some countries it is not allowed to combine data from different sources or to use it for purposes different from those for which they have been collected. For instance, it may be allowed to use an external rating about credit worthiness of a customer for credit risk evaluation but not for other purposes. Ownership of data can be unclear, depending on the details of how and why it was collected, and whether the collecting organization changes hands.
- **Real-world validation of results is essential for acceptance.** In CRM, as in many DM applications, discovered patterns are often treated as hypotheses that need to be tested on new data using rigorous statistical tests for the actual acceptance of the results. This is even more so for taking or recommending actions, especially in such high-risk applications as in the financial and medical domains. Example: recommending investments to customers (it is actually illegal in the US to let software give investment advice).
- **Developing deeper models of customer behavior:** One of the key issues in CRM is how to understand customers. Current models of customers mainly built based on their purchase patterns and click patterns at web sites. Such models are very shallow and do not have a deep understanding of customers and their individual circumstances. Thus, many predictions and actions about customers are wrong. It is suggested that information from all customer touch-points be considered in building customer models. Marketing and psychology researchers should also be involved in this effort. Two specific issues need to be considered here. First, what level should the customer model be built at, namely at the aggregate level, the segment level, or at the individual level? The deciding factor is how personalized the CRM effort needs to be. Second is the issue of the dimensions to be considered in the customer profile. These include demographic, psychographic, macro-behavior (buying, etc.), and micro-behavior (detailed actions in a store, e.g. individual clicks in an online store) features.

- **Acquiring data for deeper understanding in a non-intrusive, low-cost, high accuracy manner:** In many industrial settings, collecting data for CRM is still a problem. Some methods are intrusive and costly. Datasets collected are very noisy and in different formats and reside in different departments of an organization. Solving these pre-requisite problems is essential for data mining applications.
- **Managing the “cold start/bootstrap” problem:** At the beginning of the customer life cycle little is known, but the list of customers and the amount of information known for each customer increases over time. In most cases, a minimum amount of information is required for achieving acceptable results (for instance, product recommendations computed through collaborative filtering require a purchasing history of the customer). Being able to deal with cases where less than this required minimum is known is a therefore a major challenge.
- **Evaluation framework for distinguishing between correct/incorrect customer understanding:** Apart from the difficulty of building customer models, evaluating them is also a major task. There is still no satisfactory metric that can tell whether one model is better than another and whether a model really reflects customer behaviors. Although there are *some* metrics for measuring quality of customer models (e.g., there are several metrics for measuring the quality of recommendations), they are quite rudimentary, and there is a strong need to work on better measures. Specifically, the recommender systems community has explored this area.
- **Good actioning mechanisms:** Once data mining has been conducted with promising results, how to use them in the daily performance task is critical and it requires significant research effort. It is common that after some data results are obtained, the domain users do not know how to use them in their daily work. This research may require the participation of business and marketing researchers. Another way to accommodate actioning mechanisms is to integrate them into the knowledge discovery process by focusing on the discoveries of actionable patterns in customer data. This would make easier for the marketers or other domain experts to determine which actions should be taken once the customer patterns are discovered.
- **Incorporating prior knowledge:** This has always been a problem in practice. Data mining tends to find many pieces of patterns that are already known or redundant. Incorporating prior domain knowledge can help to solve these problems, and also to discover something novel. However, the difficulties of incorporating domain knowledge result in little progress in the past. There are a number of reasons for this. First of all, knowledge acquisition from domain experts is very hard. This is well documented in AI research, especially in the literature of expert systems building. Domain experts may know a lot but are unable to tell. Also, many times, domain experts are not sure what the relevant domain knowledge is, which can be very wide, although the data mining application itself is very narrow. Only after domain experts have seen some discovered patterns then they remember some domain knowledge. The second reason is the algorithmic issue. Many existing methods have difficulty to incorporate sophisticated domain knowledge in the mining algorithm. Also, once the new patterns are discovered, it is important to develop methods that integrate the newly discovered knowledge with the previous knowledge thus enhancing the overall knowledge base. Although there is some general work on knowledge enhancement, much more needs to be done to advance this area and adapt it to CRM problems. Also, integration of these methods with existing and novel Knowledge Management approaches constitutes a fruitful area of research.