APPLICATION OF PREDICTIVE ANALYTICS IN CUSTOMER RELATIONSHIP MANAGEMENT: A LITERATURE REVIEW AND CLASSIFICATION

Tala Mirzaei
University of North Carolina at Greensboro
t_mirzae@uncg.edu

Lakshmi Iyer
University of North Carolina at Greensboro
lsiyer@uncg.edu

ABSTRACT

This study is aimed to provide a comprehensive literature review and a classification scheme for application of predictive analytics and tools in customer relationship management (CRM). The application of predictive analytics in CRM is an emerging trend. PA methods help to analyze and understand customer behaviors and acquire and retain customers and also maximize customer value. Thus it facilitates CRM decisions making and supports development of CRM strategies in a customer-centric economy. This paper is aimed to present a comprehensive review of literature related to application of predictive analytics in CRM published in both academic and practitioner journals between 2003 and 2013.

Keywords

Predictive analytics, Customer relationship management, Classification scheme

INTRODUCTION

Customer relationship management is one the most significant managerial tasks in organizations. (Pan & Lee, 2003). Lind and Yen (2001) defined customer relationship management (CRM) as a combination of systems and techniques that supports building strategic relationships with customers in a long term and profitable fashion. Therefore effective CRM strategies are essentially based on customer data (Padmanabhan & Tuzhilin 2003). Specifically in the era of big data, web servers collect enormous amount of behavioral data on customers automatically. Bose (2002) defined CRM development lifecycle as customer acquisition, analysis and use of knowledge. This lifecycle system is developed to enable organizations with a more efficient approach to develop their customer base and ultimately increase sales.

An efficient CRM system comprises two main components which are operational CRM and analytical CRM (Gefen and Ridings, 2002; Massey et al., 2001, Saeed et al., 2011). Generally the focus of the CRM systems that are implemented to date are at improving operational aspects of CRM such as technologies to receive customer queries and address issues to enhance customer satisfaction and loyalty. Operational CRM deals with collecting customer data through a range of touch points such as contact center, contact management system, mail, fax, sales force and web. In this framework the data is stored in a customer centric database, which is available to all users in an organization to interact with their customers. However Analytical CRM is generally concentrated on customer data mining and analysis. In analytical CRM framework a combination of analytical tools are adopted to analyze customer data that are stored in databases and create their profiles and segmentations and also to identify the similar patterns in customer behaviors and predict the level of satisfaction of the customer. These analytical tools help to identify the profitability of the customers and also to note about the customer that are prone to churn. Therefore, analytical CRM and use of data-based analytical methods for customer analysis, customer interactions, and profitability management gain more importance and popularity. Organizations enabled with analytical CRM tools can incorporate better strategies to use their resources in more efficient way (Ngai et al., 2009). Various technologies are identified in supporting the analytical CRM systems among which “CRM portals”, “data warehouses”, “predictive and analytical engines”, “sequential patterns”, “clustering”, “classification”, and “evaluation of customer value” are discussed by various researchers (Xu & Walton, 2005). Predictive analytics (PA) tools are popular means of analyzing customer data within the analytical CRM framework. As a result of this analysis, customers are more effectively segmented and managers are able to make strategic decisions about their relationship with customers in offering products and services based on customer profiles (Xu & Walton, 2005).

Predictive analytics have been extensively used in telecommunication and banking industries for several years to comprehend customer behavior and profitability. In recent years the use of these techniques are rapidly generalized to other firms as well who are involved in business–to-customer transactions. The notion of big data and the potential of producing actionable information from the existing databases are the main drivers of predictive analytics application (Halper, 2011). Thus, the number of organizations who intend to make partners with PA vendors is on the rise (Barkin, 2011). According to the

"Predictive Analytics Business Application Survey" by Predictive Impact in 2009, the top three reasons for implementing a PA solution are to obtain strategic insights (75%), achieve decision support (57.1%) and enact decision automation (46.4%). This survey also denotes that 90.1% of respondents who have deployed predictive analytics attained a positive return on investment (ROI) from their most successful initiative.

PRIOR RESEARCH

Predictive analytics comprise collection of statistical and empirical models with the goal of creating empirical predictions and further assessing the quality of those predictions in practice. These techniques are applicable in both theory building and theory testing approaches. Hence, Shmueli and Koppius (2011) consider these techniques as necessary components of scientific research in information systems. The notion of predictive analytics is different from causal statistical modeling. Inferential statistics modeling is aimed to evaluate the proposed hypotheses and identify the causal factors that are significantly influence the construct of interest. However the focus of predictive analytics is on forecasting and improving the predictive power of the proposed model.

One of the managerial concerns in any business is to understand the risks and also the opportunities that the organization is prone to. Predictive models are used to identify these challenges through investigating historical and transactional customer data. The predictive models facilitate understanding the relationships between different pieces of data and identifying the underlying factors of customer behavior and also forecast the potential deficiencies in business in advance to be able to make informed decisions regarding interactions with customers. Thus, predictive analytics are considerably different from traditional business intelligence tools. Traditional BI tools are only useful to depict past trends and performance of organizations based on historical data while predictive analytics are capable of making predictions, inform decisions and forecast future movements of the customers and industries.

The application of predictive analytics in CRM is expedited by IT tools. In order for CRM to become the underlying infrastructure, one definition of CRM emphasizes on IT as an enabling technology for organizations to nurture closer relationships with customers, and enables customer information analysis to provide a coherent view of the customer behavior (Goodhue et al., 2002). To this aim, an integrated information system needs to be in place to provide relevant, real-time and accurate information to all employees in the organization. Such an integrated information system usually requires the integration of several departments in an organization such as marketing, sales and service functions (Pan et al., 2003). The integration of information in the organization usually requires a centralized database or data warehouse to store all the relevant information of customers as well as the data required for the operations within the organization. In this case, relevancy of data gains more importance than quantity of data. There is evidence that wealth of relevant customer data is critical for organizations to become customer-centric (O'Halloran, 2003).

This research presents a comprehensive review of literature related to application of predictive analytics in CRM published in both academic and practitioner journals between 2003 and 2013. A classification framework is provided. This paper is continued by explanation of the research methodology used in the study; presentation of the articles that are used in the research and the results of the classification are reported. Finally the conclusion, implications and limitations of the research are discussed.

RESEARCH METHODOLOGY

To fulfill the objectives of the research, both academic and practitioner journals are considered. Literature shows that technologies for CRM system developments were augmented in years 2003 and 2004 (Campbell, 2003; Rowley, 2004; Minna & Aino, 2005). Therefore the time frame of this study is focused on the publications between 2003 and 2013. The nature of research in CRM and predictive analytics is not restricted to specific discipline and the relevant material are scattered across various journals. Therefore the following online journal databases were searched to provide a comprehensive bibliography on CRM and predictive analytics: Business Source Complete, Emerald fulltext, Elsevier and IEEE Library.

According to Wu and Coggeshall (2012) the fundamental methodologies of predictive analytics comprise: Linear regression, Nonlinear regression, Time series analysis, Model goodness measures, Optimization methods, Multidimensional scaling.

In addition to this list, neural network, social network analysis and decision tree methods have been applied to predict costumer behavior as well in the literature. The literature search for this study was based on multiple keywords "Customer Relationship Management", "CRM", "Predictive analytics", "regression", "time-series", "optimization", "scaling" and "forecasting". This search resulted in approximately 107 articles. Further article selection was focused on the article abstracts to ensure that they are clearly describing the predictive analytics techniques that could be applied and assisted in CRM strategies. Thus, after the initial filtering of the abstract the number of relevant papers reduced to 57. The ultimate goal of this study is to provide a classification framework for PA techniques in CRM.

Table 1 Descriptive summary of databases

Database	Counts	Percent
EBSCO	2	4%
BSC	25	44%
Science Direct	29	51%
Emerald Fulltext	1	2%
Total	57	100%

According to Swift (2001) and Parvatiyar and Sheth (2001) CRM consists of five dimensions: Customer acquisition, Customer attraction, Customer retention, Customer development, and Customer equity growth.

These five dimensions share the common goal of enhancing the customer value to organizations. Customer acquisition is the initial step in customer relationship management life-cycle. This phase involves targeting the population who has the potential to become profitable customers to the organization (Kracklauer et al., 2004). Customer targeting and customer segmentation are the important elements in this phase (Noh et al, 2004)

Customer attraction is the next step after customer acquisition which involves the direct attempts of the organizations to allocate appropriate resources to the target customers that are identified in customer acquisition stage. An element of customer attraction is direct marketing (Chun, 2012).

The next step is customer retention. Customer retention is the most prevalent concern in customer relation management which comprises of customer satisfaction, customer loyalty and churn management. The elements of customer retention include personalization programs, loyalty programs and complaints management and customer churn. Specifically, churn analysis, credit scoring, service quality or satisfaction form part of loyalty programs (Flint, 2011; Netemeyer, 2007; Glady, 2009; Ngai et al., 2009)

Customer development is the next step after customer retention. It encompasses consistent expansion of transaction value and relationships of the existing customers. Elements of customer development include up/cross selling and market basket analysis. (Ha et al., 2007; Chen, 2012)

Customer equity growth ensures the consistent growth of customers' transaction value in terms of the future revenues and the long term profitability of the customers. Elements of customer equity growth are customer profitability and customer lifetime value (Liu, 2005; Kim & Lee, 2007; Chen, 2013) Predictive analytics techniques help to accomplish this goal through the above mentioned methods.

CLASSIFICATION OF THE ARTICLES

The detailed distribution of all 57 articles that are classified by the proposed framework is presented in table 2. Out of the 57 collected articles 41 papers have presented application of predictive analytics with empirical data and 16 papers adopted qualitative research approach.

Among all five customer relationship dimension, customer retention is the most prevalent that investigated using predictive analytics in the literature (20 out of 41 article, 50%). 80% of the articles in customer retention are focused on customer churn analysis. Customer equity growth forms about one quarter of the total number of article (11 out of 41, 27%) among which 8 out of 11 is allocated to customer lifetime value. Customer attraction and customer development dimensions of CRM forms the least number of articles. Predictive analytics has been applied to these dimensions of CRM limitedly.

Table 2 Distribution of articles according to the proposed classification model

CRM Dimension	CRM Element	Predictive Analytics Method	Analytic tool	Reference
Customer Acquisition	Customer Targeting	Regression forests/ Random forests		Buckinx, Verstraeten (2007)
		Logistic Regression		Noh et al. (2004)
	Customer Segment	Segmentation		Benoit, Poel (2009)
		Data Mining		Hwang, Jung (2004)
		Decision Tree	CART	Han, Lu (2012)
Customer Attraction	Direct Marketing	Monte Carlo Simulation	Excel Solver	Chun (2012)
		Logestic Regression		Soroush et al. (2012)
		Neural Network		
Customer Retention	Loyalty	Model Goodness Measures	LISREL	Flint, Blocker (2011)
		Multiple Regression		Ndubisi, Wah (2007); Laohasirichaikul (2005)
	Complaint Management	Predictive Validity		Netemeyer (2007)
	Customer Churn	Logistic Regression	MATLAB	Glady (2009); Migueis (2012)
		Decision Tree		Galdy (2009); Hung, Yen (2006)
		Sequence Analysis		Migueis (2012)
		Neural Network/ Hybrid Neural Network		Tsai, Lu (2009); Kisioglu, Topcu (2011); Hung, Yen (2006)
		Random Forest/ Regression Forest	RandomForests	Lariviere, Poel (2005); Burez, Poel (2007)
		Quantile Regression		
		Social Network		Benoit, Poel (2012)
		Survival Analysis	MATLAB	Anthes (2004); Larivier, Poel (2004), Burez, Poel (2008)
		Markov Chain	MATLAB	Burez, Poel (2007)
Customer Development	Up/Cross Selling	Cluster Analysis	C5.0	Ha et al. (2007)
		Decision Tree		
	Market Basket	Multiple Kernel Learning		Chen, Fan (2012)
		Support vector Machine		
		Segmentation		Otto (2009)

CRM Dimension	CRM Element	Predictive Analytics Method	Analytic tool	Reference
Customer Equity Growth	Customer Profitability	Decision Tree	Weka	Padmanabhan, Zheng (2006)
		Markov Chain		Rust (2011)
		Optimization		Kim, Lee (2007)
	Customer Lifetime Value	Hybrid Data Mining		Liu (2005); Hwang, Jung (2004)
		Markov Chain	CART, MATLAB	Haenlein (2007); Ma, Li (2008)
		Optimization		Crwoder (2007)
		Analytical Hierarchy Process		Liu (2005)
		Support vector Machine	MATLAB	Chen, Fan (2013)
		Quantile Regression		Benoit, Poel (2009)

The adopted IT tools are not disclosed in the majority of the articles. The IT tools that have been disclosed in the articles are MATLAB, CART, WEKA, C5 and RandomForest. The most prevalent predictive analytic techniques in the context of CRM are identified as decision tree, logistic regression, Markov Chain and survival analysis. The most prevalent dependent variable that has been predicted using predictive analytics in the literature is customer lifetime value followed by customer churn. The most prevalent independent variable that have been considered as predictors are identified as customer previous revenues and transactions followed by customer demographic characteristics such as age and gender and geographical location and also customer recency, frequency and monetary value (RFM).

IMPLICATIONS AND CONCLUSION

The application of predictive analytics is emerging in the context of customer relationship management across various industries. The results of this research shows PA techniques are gaining popularity not only in banking and telecommunication industry but also in different retailer, manufacturing, insurance, healthcare and casinos.

Despite the increasing trend of application of predictive analytics techniques in customer relationship management, there is a lack of comprehensive literature review and a classification scheme for it. This research provided a classification framework to fulfill this gap in the literature and guide future research. The five dimension of CRM are identified as customer acquisition, customer attraction, customer retention, customer development and customer equity growth. Predictive analytics is mostly applied in customer retention to foresee the customer churn challenges in organizations and to make informed decision to avoid these problems. Also PA techniques are practiced in customer equity growth dimension specifically to predict customer lifetime value. Little attention is paid to customer acquisition and customer attraction in the literature in terms of target marketing and direct marketing. More studies need to be conducted in these dimensions of CRM in order to empower the organizations in terms of customer target analysis and prediction of potential customers. The most prevalent predictive techniques that have been applied are logistic regression, decision tree, segmentation and Markov chain. Multiple logistic regressions is identified as a better model in terms of interpretability of the parameters in the literature compare to decision tree and neural networks. The use of IT tools is not disclosed in majority of the article. MATLAB is used for programming among the articles that denote the applied tool. CHIAD and C5 algorithm are suggested for visual presentation and understanding of the data for instance in regression tree and semantic rules to facilitate the interpretation of model parameters.

REFERENCES

1. Barkin, E. (May 01, 2011). CRM + Predictive Analytics: Why It All Adds Up. *CRM Magazine, 15,* 5.)
2. Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a

non-contractual FMCG retail setting. *European Journal of Operational Research, 164*(1), 252-268. doi: 10.1016/j.ejor.2003.12.010

3. Buckinx, W., Verstraeten, G., & Van den Poel, D. (2007). Predicting customer loyalty using the internal transactional database. *Expert Systems with Applications, 32*(1), 125-134. doi: 10.1016/j.eswa.2005.11.004
4. Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, *32*(2), 277-288. doi: 10.1016/j.eswa.2005.11.037
5. Burez, J., & Vandenpoel, D. (2008). Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications, 35*(1-2), 497-514. doi: 10.1016/j.eswa.2007.07.036.
6. Butler, S. (2000) "Customer Relationships: Changing the Game: CRM in the e-World", Journal of Business Strategy, Vol. 21 Iss: 2, pp.13 – 14
7. Chang, C.-W., Lin, C.-T., & Wang, L.-Q. (2009). Mining the text information to optimizing the customer relationship management. *Expert Systems with Applications, 36*(2), 1433-1443. doi: 10.1016/j.eswa.2007.11.027
8. Chen, Z.-Y., & Fan, Z.-P. (2012). Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM approach. *Knowledge-Based Systems, 35*, 111-119. doi: 10.1016/j.knosys.2012.04.023
9. Gurau, C. (2003). Tailoring e-service quality through CRM. *Managing Service Quality, 13*(6), 520-531. doi: 10.1108/09604520310506577
10. Goodhue, D. L., Wixom, B. H., & Watson, H. J. (January 01, 2002). Realizing business benefits through CRM: hitting the right target in the right way. *Mis Quarterly, 26,* 2.)
11. Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research, 34*(10), 2902-2917. doi: 10.1016/j.cor.2005.11.007
12. Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*(7), 5259-5264. doi: 10.1016/j.eswa.2009.12.070
13. Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications, 31*(3), 515-524. doi: 10.1016/j.eswa.2005.09.080
14. Kim, Y. (2006). Toward a successful CRM: variable selection, sampling, and ensemble. *Decision Support Systems, 41*(2), 542-553. doi: 10.1016/j.dss.2004.09.008
15. King, S. F. (2007). Citizens as customers: Exploring the future of CRM in UK local government. *Government Information Quarterly, 24*(1), 47-63. doi: 10.1016/j.giq.2006.02.012
16. Larivière, B., & Van den Poel, D. (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. *Expert Systems with Applications, 27*(2), 277-285. doi: 10.1016/j.eswa.2004.02.002
17. Ling, R., & Yen, D. C. (January 01, 2001). Customer Relationship Management: An Analysis Framework and Implementation Strategies. *The Journal of Computer Information Systems, 41,* 82-97.
18. Miguéis, V. L., Poel, D., Camanho, A. S., & Falcão e Cunha, J. (2012). Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences. *Advances in Data Analysis and Classification, 6*(4), 337-353. doi: 10.1007/s11634-012-0121-3
19. Miguéis, V. L., Van den Poel, D., Camanho, A. S., & Falcão e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, *39*(12), 11250-11256. doi: 10.1016/j.eswa.2012.03.073
20. Morgan, R. M., Beatty, S. E., & Vivek, S. D. (2012). Customer Engagement: Exploring Customer Relationships Beyond Purchase. *The Journal of Marketing Theory and Practice*, *20*(2), 122-146. doi: 10.2753/mtp1069-6679200201
21. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, *36*(2), 2592-2602. doi:

10.1016/j.eswa.2008.02.021

22. Padmanabhan, B., Zheng, Z., & Kimbrough, S. O. (June 01, 2006). An Empirical Analysis of the Value of Complete Information for eCRM Models. *Mis Quarterly, 30,* 2, 247-267.
23. Pan, S. L., & Lee, J.-N. (January 01, 2003). Using e-CRM for a unified view of the customer.(electronic Customer Relationship Management). *Communications of the Acm, 46,* 4.)
24. Shmueli, G., & Koppius, O. R. (September 01, 2011). Predictive analytics in information systems research. *Mis Quarterly, 35,* 3.)
25. Wu, J., Coggeshall, S.,(2012) "Foundations of predictive analytics", Boca Raton, FL: CRC Press
26. Xu, M., & Walton, J. (July 01, 2005). Gaining customer knowledge through analytical CRM. *Industrial Management & Data Systems, 105,* 7, 955-971.
27. Audzeyeva, A., Summers, B., & Schenk-Hoppé, K. R. (2012). Forecasting customer behaviour in a multi-service financial organisation: A profitability perspective. International Journal of Forecasting, 28(2), 507-518. doi: 10.1016/j.ijforecast.2011.05.005
28. Benoit, D. F., & Van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. Expert Systems with Applications, 36(7), 10475-10484. doi: 10.1016/j.eswa.2009.01.031
29. Benoit, D. F., & Van den Poel, D. (2012). Improving customer retention in financial services using kinship network information. Expert Systems with Applications, 39(13), 11435-11442. doi: 10.1016/j.eswa.2012.04.016
30. Brown, S. P., & Lam, S. K. (2008). A Meta-Analysis of Relationships Linking Employee Satisfaction to Customer Responses. Journal of Retailing, 84(3), 243-255. doi: 10.1016/j.jretai.2008.06.001
31. Chan, S. L., Ip, W. H., & Cho, V. (2010). A model for predicting customer value from perspectives of product attractiveness and marketing strategy. Expert Systems with Applications, 37(2), 1207-1215. doi: 10.1016/j.eswa.2009.06.030
32. Chen, Z.-Y., & Fan, Z.-P. (2013). Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach. Knowledge-Based Systems, 43, 123-134. doi: 10.1016/j.knosys.2013.01.022
33. Chew, M. Y. L., Tan, S. S., & Kang, K. H. (2005). Contribution Analysis of Maintainability Factors For Cladding Facades. Architectural Science Review, 48(3), 215-227. doi: 10.3763/asre.2005.4828
34. Chijoriga, M. M. (2011). Application of multiple discriminant analysis (MDA) as a credit scoring and risk assessment model. International Journal of Emerging Markets, 6(2), 132-147. doi: 10.1108/17468801111119498
35. Chun, Y. H. (2012). Monte Carlo analysis of estimation methods for the prediction of customer response patterns in direct marketing. European Journal of Operational Research, 217(3), 673-678. doi: 10.1016/j.ejor.2011.10.008
36. Crowder, M., Hand, D. J., & Krzanowski, W. (2007). On optimal intervention for customer lifetime value. European Journal of Operational Research, 183(3), 1550-1559. doi: 10.1016/j.ejor.2006.08.062
37. Dev, C. S., Buschman, J. D., & Bowen, J. T. (2010). Hospitality Marketing: A Retrospective Analysis (1960-2010) and Predictions (2010-2020). Cornell Hospitality Quarterly, 51(4), 459-469. doi: 10.1177/1938965510376353
38. Fader, P. S., & Hardie, B. G. S. (2005). The value of simple models in new product forecasting and customer-base analysis. Applied Stochastic Models in Business and Industry, 21(4-5), 461-473. doi: 10.1002/asmb.592
39. Fader, P. S., & Hardie, B. G. S. (2009). Probability Models for Customer-Base Analysis. Journal of Interactive Marketing, 23(1), 61-69. doi: 10.1016/j.intmar.2008.11.003
40. Flint, D. J., Blocker, C. P., & Boutin, P. J. (2011). Customer value anticipation, customer satisfaction and loyalty: An empirical examination. Industrial Marketing Management, 40(2), 219-230. doi: 10.1016/j.indmarman.2010.06.034.