Relation between Data Mining and Business Fields in the Four Dimensional CRM Model

Iva Šalov
University of Zadar, Department of Economics, Croatia
isalov@student.unizd.hr

Aleksandra Krajnović
University of Zadar, Department of Economics, Croatia
akrajnov@unizd.hr

Ante Panjkota
University of Zadar, Department of Economics and Maritime Department, Croatia
apanjkot@unizd.hr

Abstract. Various academic sources show how data mining techniques (DM) can be successfully applied in 4-dimensional CRM model. However, there is no known systematic synthesis related to the correlation between DM, CRM, and business fields (e.g., banking, e-commerce, telecommunication, etc.). This paper deals with that problem throughout the review of primary and secondary sources of the application of DM techniques in the four CRM dimensions related to different business fields in the period from 2006 to 2016. The proposed research model is multidimensional where, for the sake of simplified representation, its base consists of three dimensions – DM techniques, CRM dimensions and business fields. Each mentioned dimension is multi-dimensional by itself. This preliminary study indicates the high impact of DM on CRM dimensions in all business fields with high interaction with customers. Besides that, implicitly or explicitly customer’s privacy and security issues arose as a problem in most recent studies on the application of DM techniques in the frame of the contemporary CRM systems. In that sense, we proposed extended CRM model with customers’ security and privacy as root dimension.

Keywords: CRM, Data Mining, Business Fields, Customer Security, Customer Privacy

1 Introduction

Four-dimensional CRM model with customer identification, customer attraction, customer retention and customer development as the primary dimension has been independently proposed by various authors (Swift 2001), (Parvatiyar and Sheth 2001) and (Kracklauer et al. 2003). Many others CRM models can be found in literature among which we are pointing toward the two models. The first one is the five-dimensional CRM model comprised of the following dimensions: CRM activity investments, relative CRM expenses, information generation, information dissemination and responsiveness (Josniassen, Assaf, and Cvelbar 2014, 132). The second one is a four-dimensional model with a dimension of focusing on key customers, the dimension of organizing around CRM, the dimension of incorporating CRM-based technology and finally the dimension of managing knowledge (Yim, Anderson, and Swaminathan 2004, 267).

In this research we decided to use first mentioned four-dimensional model because it is more appropriate for DM techniques usage analysis in CRM than the other dimensional CRM models – same is confirmed in the widely cited paper (Ngai, Xiu, and Chau 2009).

Over the years’ various data mining techniques have been successfully used in the selected 4-dimensional CRM model. Thus, multilayer neural networks are widely used in customer attraction dimension (Femina and Sudheep 2015; Lam et al. 2014; Ramalingam et al. 2006). In the customer retention dimension different decision trees algorithms are utilized regarded its high interpretability.
Association rules can produce cause and effect scenario based on “what-if” strategy, so they are mostly tailored for the customer development dimension (Chiang 2011; Hu and Yeh 2014). In the end, in the customer identification dimension different clustering algorithms have been applied (López et al. 2011; Hiziroglu 2013; Ansari and Riasi 2016).

Although diverse sources commonly provide a bright, extensive and precise description of data mining approaches applied to the particular CRM dimension, there is no systematic synthesis related to the correlation between the DM, CRM dimensions, and business fields (e.g., banking, telecommunication, e-commerce, automobile industry, etc.). Moreover, it is interesting to ask the question: In the era of Big data and extensive irrefutable usage of DM in the CRM can we detect and prove discrepancies between business fields and application of the efficient DM techniques in the single CRM dimension? This paper answers the posed question by reviewing relevant primary and secondary sources from 2006–2016.

Analysing collected academic sources throughout proposed multidimensional methodological framework main result indicates the high impact of DM on CRM dimensions in all business fields with high interaction with customers, or in other words business fields that provide different services to customers. Besides that, our analysis and synthesis provide interesting insight into customers’ security and privacy issues that explicitly or implicitly arises in most recent studies on the application of DM techniques in the frame of the contemporary CRM systems. In that sense, we suggest an extension of the four-dimensional CRM model with root dimension which refers to a customer security and privacy.

The conducted study has been limited only toward detection of relationships between DM and business fields in the four-dimensional CRM model and as such can be seen only as a preliminary study that gives pointers for further research efforts on finding effective DM approaches that can be applied to develop CRM in different business fields.

The rest of this paper is organized as follows. Section 2 describes the methodological model, section 3 presents results of this study, while section 4 emphasize significant implications of obtained results. Finally, section 5 highlights main conclusions of the research.

2 Methodological model

Proposed multidimensional methodological model directly emanates from central research question in which we are interested in observing the relationship between particular CRM dimension from the 4-dimensional CRM model, data mining technique that has been used in it and business field of interest. In the first section references to the chosen 4-dimensional model are given, and here we bring some necessary explanations from the standpoint of the possible use of the DM techniques in every dimension.

- Customer identification – refers to a DM approaches of finding groups of customers with similar characteristics. That can be potential customers or most profitable customers (Panniello, Hill, and Gorgoglione 2016; Hamka et al. 2014) or some segment of medical patients for appropriate allocation of healthcare service resources (Y. S. Chen et al. 2012).
- Customer attraction – after customers identification business organizations can direct their resources and efforts toward targeting customer segments. Best strategies for that process can be found by using many DM approaches (Tsai, Hu, and Lu 2015).
- Customer retention – DM can be useful in finding best or adequate description of customers degree of realignments between its expectations and the obtained quality of products or services in the massive amount of collected data from websites, email, operational databases, social networks, and many others sources (Cheng and Sun 2012). Also, enormous DM efforts in this dimension are directed to customer churn predictions (Verbeke et al. 2011).
• Customer development – how to increase customer lifetime value is the one of the core interest in this dimension. DM can reveal regularities in customer service usage or purchasing behavior so that can produce opportunities for uplifting customers short time and total value (Thanuja, Venkateswarlu, and Anjaneyulu 2011). Besides that, DM possess tools for predicting customers’ buying affinities so they can produce highly probable items or services related to the specific customer which results with a new added value for the company (Gomez-Uribe and Hunt 2015).

Explanations of the four dimensions in the 4-dimensional CRM model that is more oriented toward business and managerial perspective can be found in the paper (Kumar, B Santhosh, Sivaparthipan, C.B., Kalaikumaran, T., Karthik 2013).

What needs to be taken into consideration when choosing a suitable DM techniques in CRM processes is reviewed in the paper (Rygielski, Wang, and Yen 2002). An exhaustive review on using data mining techniques on four CRM dimensions (customer identification, customer attraction, customer retention and customer development) in performing the essential DM tasks: association, classification, clustering, forecasting, regression, sequence discovery and visualization is elaborated in the paper (Ngai, Xiu, and Chau 2009).

In the proposed methodological model, we do not make distinctions based on putting DM algorithms in groups, e.g., group of clustering algorithms, group of decision trees, group of neural networks, etc. Preferably, in this research, we observed individual algorithm usage, in particular, CRM dimension and corresponding business field.

Business field is self-explanatory and represents a specific business area with all its characteristics (e.g., banking, telecommunication, manufacturing, tourism, etc.). Based on all of this proposed methodological model can be simplified as is shown in Figure 1.
Each of three primary dimensions (DM techniques, CRM dimensions of the 4-dimensional CRM model and business fields) is comprised of many dimensions – Figure 1. Theoretically, with this representation, every possible business field can be observed in every dimension of the 4-dimensional CRM model related to different DM algorithms that can be used.

3 Results

Based on the methodological model described in previous section 43 papers are selected and analyzed. In this section, main results are summarized and presented in the tabular and graphical form.

3.1 Customer identification dimension

Table 1 Main results from 12 papers on using DM techniques in customer identification dimension

<table>
<thead>
<tr>
<th>Business Field</th>
<th>DM technique</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance</td>
<td>Not specified</td>
<td>(Bhatnagar and Ranjan 2011)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Clustering and association rules</td>
<td>(Cheng and Sun 2012)</td>
</tr>
<tr>
<td>Tourism</td>
<td>K-means; SOM</td>
<td>(Dursun and Caber 2016)</td>
</tr>
<tr>
<td>Retailing</td>
<td>K-means</td>
<td>(Gupta, Aggarwal, and Rani 2016)</td>
</tr>
<tr>
<td>Service industry</td>
<td>SOM; C5.0</td>
<td>(Ha 2007)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Decision tree</td>
<td>(Han, Lu, and Leung 2012)</td>
</tr>
<tr>
<td>Curtain manufacturing</td>
<td>GA based k-means</td>
<td>(Ho et al. 2012)</td>
</tr>
<tr>
<td>Banking</td>
<td>K-means; ARIMA</td>
<td>(Khajvand and Tarokh 2011)</td>
</tr>
<tr>
<td>Banking</td>
<td>Ensemble SVM</td>
<td>(Lai et al. 2007)</td>
</tr>
<tr>
<td>Textile manufacturing</td>
<td>K-means</td>
<td>(D. C. Li, Dai, and Tseng 2011)</td>
</tr>
<tr>
<td>Retailing</td>
<td>VARCLUS</td>
<td>(Miguéis, Camanho, and Falcão E Cunha 2012)</td>
</tr>
<tr>
<td>E-commerce</td>
<td>DT; NN; Logistic regression and Bagging</td>
<td>(Shim, Choi, and Suh 2012)</td>
</tr>
</tbody>
</table>

Data from Table 1 which gives details on using DM techniques in customer identification dimension on different business fields can be summarized as it is shown in Figures 2 and 3. Also, Table 1 emphasize that many authors have been used multiple DM techniques in their work. Figure 2 shows a cumulative number of same DM techniques used in customer identification dimension. On the other hand, Figure 3 depicts papers distribution in the task of customer identification per business fields. In 6 papers of 12, clustering algorithm has been applied for customer identification – Figure 2. From the same Figure 2, it is evident that SCM and ensembles algorithms are mentioned in two papers and decision trees in three papers. All others algorithms are only ones represented in that task. In one paper authors did not specify DM algorithm. Based on 12 analyzed papers in the dimension of customer identification telecommunication, retailing and banking are a business field of interest in 2
papers each while tourism, insurance, service industry, textile manufacturing, manufacturing curtains, and e-commerce are mentioned in one article each – Figure 3.

![Cumulative usage of DM techniques in customer identification dimension from 12 articles](image1)

Figure 2 Cumulative usage of DM techniques in customer identification dimension from 12 articles

![Business fields distribution related to customer identification dimension from 12 articles](image2)

Figure 3 Business fields distribution related to customer identification dimension from 12 articles

### 3.2 Customer attraction dimension

From total eight papers, we analyze the relation between DM techniques, customer attraction dimension, and business fields. Main results of that analysis are shown in Table 2, while summarized data are depicted in Figures 4 and 5. Same as in the previous subsection graphical visualization is directed toward detection of the DM techniques usage in CRM dimension of interest and in finding the
Table 2 Main results from 8 papers on using DM techniques in customer attraction dimension

<table>
<thead>
<tr>
<th>Business Field</th>
<th>DM technique</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retailing</td>
<td>Decision trees</td>
<td>(Anderson, Jolly, and Fairhurst 2007)</td>
</tr>
<tr>
<td>Banking</td>
<td>MLP and Naive Bayes</td>
<td>(Femina and Sudheep 2015)</td>
</tr>
<tr>
<td>Retailing</td>
<td>Fuzzy clustering; C4.5; Hidden Markov Model</td>
<td>(Deepa, Dhanabal, and Kaliappan 2014)</td>
</tr>
<tr>
<td>Tourism</td>
<td>Apriori algorithm; k-means; Association rules</td>
<td>(Liao, Chen, and Deng 2010)</td>
</tr>
<tr>
<td>Tourism</td>
<td>Factor analysis and Variable Consistency</td>
<td>(Liou 2009)</td>
</tr>
<tr>
<td>Air transportation</td>
<td>Consistency Dominance-based Rough Set Approach</td>
<td>(Liou 2009)</td>
</tr>
<tr>
<td>E-commerce</td>
<td>Self-adaptive GA</td>
<td>(Mahdavi, Movahednejad, and Adbesh 2011)</td>
</tr>
<tr>
<td>Automobile industry</td>
<td>Neuro-Fuzzy clustering</td>
<td>(Mahdavi et al. 2008)</td>
</tr>
<tr>
<td>Tourism</td>
<td>Predictive visual analytics + generic DM</td>
<td>(Mastorakis et al. 2015)</td>
</tr>
</tbody>
</table>

**Total** 8

Figure 4 Cumulative usage of DM techniques in customer attraction dimension from 8 articles

Figure 4 shows that decision trees and fuzzy clustering are used in two articles each in the task of customer attraction, while all other algorithms are only found in one article each. In one article authors did not specify DM approach.
Retailing and tourism are found in two articles each as a business field in which various DM techniques are applied for customer attraction – bar chart from Figure 5. From the same chart arise that banking, air transportation e-commerce, and automobile industry are the business subject in customer attraction dimension in only one article each.

3.3 Customer retention dimensions

Same approach as in previous two dimensions (subsections) we apply in this subsection. First, among 43 papers all papers related to the subject of customer retention are analyzed, and corresponding detailed results are tabularly represented in Table 3. After that aggregate data are graphically depicted in the Figures 6 and 7.
Decision trees (C4.5, J4.8, CART) are dominant in customer retention task as can be seen from Figure 6. These DM techniques are addressed to customer retention problem in 7 of 11 papers. In one paper authors omit DM technique specification.

Table 3 Main results from 11 papers on using DM techniques in customer retention dimension

<table>
<thead>
<tr>
<th>Business Field</th>
<th>DM technique</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banking</td>
<td>Logistic regression and J4.8</td>
<td>(Gür Ali and Aritürk 2014)</td>
</tr>
<tr>
<td>E-commerce</td>
<td>Hierarchical SVM</td>
<td>(Z. Y. Chen, Fan, and Sun 2012)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>C5.0 and Hierarchical SOM</td>
<td>(Chu, Tsai, and Ho 2007)</td>
</tr>
<tr>
<td>Newspaper publishing</td>
<td>LSI and SVD; Logistic regression</td>
<td>(Coussement and Van den Poel 2008)</td>
</tr>
<tr>
<td>Automobile industry</td>
<td>K-means</td>
<td>(Hosseini, Maleki, and Gholamian 2010)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Logistic regression: linear classifier; EA; Naïve Bayes; DT; SVM; MLP</td>
<td>(Huang, Kechadi, and Buckley 2012)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>K-means with C5.0; NN with C5.0</td>
<td>(Hung, Yen, and Wang 2006)</td>
</tr>
<tr>
<td>Banking</td>
<td>Not specified</td>
<td>(Y. Li, Li, and Gao 2008)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Hybrid NN</td>
<td>(Tsai and Lu 2009)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Partial focus feature reduction and classification ensemble (J4.8, Naïve Bayes, LibSVM)</td>
<td>(Tu and Yang 2013)</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Miner and ALBA</td>
<td>(Verbeke et al. 2011)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>

It is evident from Table 3 that authors in observed 11 papers of customer retention issues on an average used a combination of more than one DM algorithm. More precisely, only in three of eleven papers researchers are used only one DM algorithm to provide answers to customer retention problem. In comparison to e-commerce, newspaper publishing and automobile industry with only one paper each, and banking with only two papers, telecommunication are found in 6 of 11 papers as a business field in which DM techniques are applied to customer retention problem – Figure 7.
3.4 Customer development dimension

Table 4 Main results from 12 papers on using DM techniques in customer development dimension

<table>
<thead>
<tr>
<th>Business Field</th>
<th>DM technique</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-commerce</td>
<td>C4.5</td>
<td>(Chang, Lin, and Wang 2009)</td>
</tr>
<tr>
<td>Retailing</td>
<td>Multiple Kernel Support Vector Regression</td>
<td>(Z.-Y. Chen and Fan 2013)</td>
</tr>
<tr>
<td>E-commerce</td>
<td>Fuzzy clustering; Apriori and Association rules</td>
<td>(Chiang 2011)</td>
</tr>
<tr>
<td>Banking</td>
<td>DM generic Factor Analysis and Rough Set Theory; Decision Flow Graph</td>
<td>(B. Fang and Ma 2009)</td>
</tr>
<tr>
<td>Banking</td>
<td></td>
<td>(S. K. Fang et al. 2012)</td>
</tr>
<tr>
<td>Retailing</td>
<td>Association rules</td>
<td>(Hu and Yeh 2014)</td>
</tr>
<tr>
<td>Health and beauty manufacturing</td>
<td>K-means</td>
<td>(Khajvand and Tarokh 2011)</td>
</tr>
<tr>
<td>Automobile industry</td>
<td>K-means and SOM; DT</td>
<td>(Khajvand et al. 2011)</td>
</tr>
<tr>
<td>Electricity production and distribution</td>
<td>Hopfield K-means</td>
<td>(López et al. 2011)</td>
</tr>
<tr>
<td>Retailing</td>
<td>MNL</td>
<td>(Prinzie and Van Den Poel 2006)</td>
</tr>
<tr>
<td>Retailing</td>
<td>Hierarchical heterogenous Ant Colony Optimisation based Action Rule Mining</td>
<td>(Sreeja and Sankar 2016)</td>
</tr>
<tr>
<td>Retailing</td>
<td>Bayesian multivariate Poisson Regression Model</td>
<td>((Deanna) Wang, Kalwani, and Akçura 2007)</td>
</tr>
</tbody>
</table>

Total 12
Like in previous dimensions authors in customer development dimensions usually combine two or more DM algorithms – see Table 4.

Figure 8 shows that decision trees algorithms, association rules, and k-means clustering can be found as main DM approach in two papers each. From same Figure 8, it can be seen that all other DM approaches (SVM kernel regression, fuzzy clustering, factor analysis and Rough sets, SOM, MNL, hierarchical heterogenous Ant Colony Optimisation based Action Rule Mining and Bayesian multivariate Poisson Regression model) are reported in one article each. Also, in one article authors talks on the generic usage of data mining, so that paper is classified with not specified DM approach.

Retailing is registered in five of twelve papers as a business field of interest in applying DM techniques on customer development problem – Figure 9. Banking and e-commerce are founded in two papers each, while health and beauty manufacturing, automobile industry and electricity production and distribution are business fields of concerns in only one paper each as it is depicted on the bar chart in Figure 9.

Figure 8 Aggregate data of DM techniques usage in customer development dimension from 12 articles
4 Discussion

Customer identification dimension points out clustering algorithms as dominant DM approach while any business field cannot be found with the significant implementation of DM techniques compared to any other. First observation implies that clustering algorithms may be better suited for customer identification tasks than the others DM approaches. Knowing general principles of clustering algorithms with their high capability of grouping objects with similar characteristics and with the ability to provide a smooth mechanism of explanations for obtained groups they can be proper research direction in any business field for customer identification. Moreover, there is need to investigate performances of the broad family of clustering algorithms for that task – e.g., partitioning, hierarchical, model-based, grid-based, density based or hybrid clustering.

In customer attraction dimension, according to this research, there are no dominant DM techniques and also no business field with dominant implementation in observed CRM dimension. Although this research did not give any traces for preferred algorithms in this CRM dimension, possible candidates are discerning in the family of decision trees because of their excellent comprehensibility. This preliminary study shows that in customer retention dimension frequently used DM techniques are various decision tree algorithms while telecommunication can be seen as the business field with the dominant usage of DM techniques in this CRM dimension. The explanation for telecommunication as a dominant business field in this dimension can be found in relatively large customer base and high competitive level in that sector, so churning is more likely to happen than in other fields.

As this study gives, in customer development dimension, retailing has a significant implementation of DM techniques while it cannot be pointed out any particular DM approach for that dimension. Again, possible candidates for this dimension can be decision trees and association rules due to easy interpretability and plausible implementation of “what if” analysis. A significant amount of customer buying data stored in data warehouses or data marts combined with relatively frequent interactions with buyers characterized retailing as business field suitable for DM usage for rising customer lifetime value.

From observed 43 articles more than 70% of them have data pre-processing steps, and their authors put some remarks on that phase which usually includes: data filtering, coping with missing data, dimensionality reduction (attribute selection or attribute extraction) and attribute engineering. Another problem mentioned in some of the collected papers is algorithm scalability related to Big Data.
paradigm. Algorithm robustness to noise and outliers are another critical issue. Apparently, there is an urge for cooperation between business domain experts, CRM experts and DM researchers and practitioners in using DM techniques at any CRM dimension and in any business field. We are entirely aware that the research core of 43 articles is main limitations of this study and, due to that reason, we see this research as a preliminary study. In that sense, obtained results and observations present directions of future research rather than the basis for firm conclusions.

Figure 10 Proposed 5-dimensional CRM model as extension of 4-dimensional CRM model with added customer privacy and security as root dimension

Big Data paradigm, social networks, mobile and cloud computing leads us to a critical problem in contemporary CRM systems which is customer privacy and security. Customer privacy and security are closely related with perceived customers’ trust (Katsikas, Lopez, and Pernul 2005). Transactional information systems provide a large amount of customers’ data for CRM system, and DM tools can reveal connections among seemingly unrelated data. In that context, all analyzed papers implicitly or explicitly bring the question of customers privacy and security into research focus. Moreover, customers have more and more awareness about possible intrusion on their privacy, and at the same time, they are concerned about security issues of modern technologies (Pearson 2013; Carlos Roca, José García, and José de la Vega 2009). All mentioned is present in today’s CRM systems (Nguyen, Sherif, and Newby 2007; Virupaksha, Sahoo, and Vasudevan 2014) so we suggest to extend known 4-dimensional CRM model with root dimension related to customer privacy and security as it is shown in Figure 10. This root dimension permeates all four dimensions in the classical 4D CRM model. Proposed, extended four-dimensional CRM model with the root customer privacy dimension demands detailed investigation which will show or refute justification of its use.

5 Conclusion

The most important findings of this paper indicate the high impact of DM on CRM dimensions in all business fields with high interaction with customers, or in other words business fields that provide different services to customers (e.g., telecommunication, banking, tourism).
Further, this research highlighted clustering algorithms as most promising DM tool for customer identification dimension. Also, a group of decision trees algorithms could be efficient in customer retention dimension, while group of association rules and once more decision trees algorithms are prospective DM approaches for coping with problems in customer development dimension. Meanwhile, a limited number of analyzed papers brought no candidates of DM algorithms for dealing with problems in customer attraction dimension. Same reason of the relatively small number of analyzed papers (43) is a major limitation of this research, and for more firm conclusions literature list need to be expanded. In that future research, there is urge of grouping DM algorithms to secure tractable and meaningful research environment. Even this preliminary study claimed to group DM algorithms that belong to the same family of algorithms – a group of clustering algorithms, a group of decision tree algorithms, a group of association rules algorithms, and a group of ensemble algorithms. Besides these most significant findings, this research has brought into the focus one more additional problem related to the nowadays CRM systems. In such a way DM usage in any CRM dimension is related to the massive amount of customers’ collected and stored data. Besides that, coupled CRM and DM technologies are accompanied with high capacity of DM algorithms to give meaning to scattered customers’ data. All of this, along with ever-present security issues related to the information technologies that support DM and CRM, rise question of customer privacy and security in all CRM dimensions. Traces of concern with this problem can be found in analyzed papers directly or indirectly. In that sense, we proposed an extension of the used 4-dimensional model with the core dimension of customer security and privacy. Extended four-dimensional CRM model has to be put under appropriate verification in the future research.

References


of the Airline Market.” 


Shim, Beomsoo, Keanho Choi, and Yongmoo Suh. 2012. “CRM Strategies for a Small-Sized Online Shopping Mall Based on Association Rules and Sequential Patterns.” 


Tsai, Chih Fong, Ya Han Hu, and Yu Hsin Lu. 2015. “Customer Segmentation Issues and Strategies


