PRIVACY PRESERVING DATA MINING FOR CUSTOMER RELATIONSHIP MANAGEMENT

[1]MAJIDREZA ARASTEH, [2]MOHAMMAD BABA, [3]SHAHAB REZAIAN

[1,2]Master student of e-commerce, IT department, Faculty of engineering, Islamic Azad University E-Campus – Tehran - Iran
[2,3]Lecturer at IT department, Faculty of engineering, Islamic Azad University E-Campus – Tehran - Iran
[3]Ph.D. student at Advanced Informatics School- UTM- KL- Malaysia
E-mail: [1]Madjidreza@gmail.com, [2]Mohammad.Baba@gmail.com, [3]srezaian@outlook.com

Abstract- Customer Relationship Management (CRM) not only is a competitive advantage, but also maintaining customers and explaining specific strategies for each client is essential in any business. Substantial growth in hardware and software technology, large amounts of data are generated and stored. As a result, the main problem in CRM is not the lack of information about the customer, but the problem emerges when processing the huge volume of information and translating them into useful knowledge with the aim of improving CRM and creating a long-term and profitable relationship with customers. The huge volume of informationhave provided an excellent bed for the use of data mining techniques so that by applying the knowledge discovery models on this information to create very good knowledge, especially in CRM. No data mining technique can operate without violating the privacy of the data owners. Given that ETL is one of the data mining steps, the focus of preserving privacy is more in data mining algorithms and less attention in ETL is paid to this issue. Thispaper aims at examining solutions proposed for preserving privacy in data mining,especially in the techniques of association rules, which has many applicationsin CRM. In addition, by using the best proposed method and simultaneous use of it and techniques used in preserving privacy in ETL,we have attempted to find a better balance between privacy and the economic benefits.

Keywords- Data mining, CRM, Privacy Preserving, ETL, Association Rules.

INTRODUCTION

Today, companies try to gain and keep more valuable customers have increased and most volume of studies is done on customer retention [1]. The focus has changed from production-oriented to customer-oriented and a focus is made on the collection and storage of data and the use of data mining methods for taking better decisions [2].

Data Mining:

Data mining defines as a sophisticated data search capability that uses statistical algorithms to discover patterns and correlations among the data [3]. Today, due to the massive amounts of data in the database,computers and powerful algorithms are replaced by manual analysis to extract information and knowledge from existing data. Data mining is the process of discovering significant relationships, patterns and trends, which examine and deal with large volume of information stored in the database using pattern recognition technology [4].

CRM:

Customer Relationship Management (CRM)defines based on four elements from a simple framework: knowledge, objective, sales and service [5]. Examining feedback of customer's information to organizations provides organizationswith chance of obtaining useful information, which can be used to reduce the weaknesses of their business performance and as a result to improve satisfaction, loyalty and maintaining profitable long-term relationship with their customers [6]. CRM receives customer's data and simultaneously interact with them and for each of the different groups of customer pursue specific strategiesto have better communication with profitable customers, to find new profitable customers and to find solutions for customers who are not profitable enough [7]. According to studies, 70 percent of CRM projects fail and the cost of setting up and implementation of these projects is high [8].

The process of data mining in CRM:

Data mining is one of the elements of CRM and facilitates tomove,companies towards being consumers-oriented. The process of data mining in CRM deals with Extract, Transform and Load (ETL) on raw data and convert them to data warehouse and conduct the process of preparing data and getting prepared data through different algorithms for building model and through evaluating model, we assess the model and will get a hypothesis [9].

Problem description:

Customer Relationship Management (CRM) not only is a competitive advantage, but maintaining customers and explaining specific strategies for each client is essential in any business [10]. In addition, the main problem of companies is not the lack of information, but the problem emerges when processing the huge volume of customers' information[7]. The concern of using data mining is mostly from the social dimension and is often overlooked, which led to issue of privacy[11]. Therefore, the largest challenge is not due to technology, but it is due to people and their behavior, e.g. ethical and legal problems in using knowledge,

specifying the knowledge owner and preserving privacy in the use of knowledge[12]. Accordingly, Privacy Preserving Data Mining (PPDM) as a new field of research, which is of great importance, has emerged [13][13][13]. Making balance between privacy and economic interests is the necessity of each business(Mehdi Basiri: Master of IT, Company of the analysis and design management systems: Gahan, 2007) and achievement to this balance point requires trade-off between obtaining information and maintaining anonymous [14]. So finding this balance point with minimal invasion of privacy and obtaining necessary knowledge in line with economic interests to improve customer relationships and create a profitable long-term relationship is the question in this study we attempt to answer.

Related Work:

Anthony, et al., in 2002provided a comprehensive understanding of data mining with testing market data mining applications and exploredsome technical concerns arising from the use of this application, which are mostly due to social dimension and often are overlooked. He referred to the issue of privacy and examinedproblems in this area, especially in connection with terrorist threats in the use of this information. Hesuggestedconsidering ethical sensitivities relating to customer privacy in data mining, but heproposed nosolution to this problem[11].

Rygielski, et al., in 2002aimed at receiving a variety of data mining, which helped achieving CRM's objectives. In this regard, two techniques fromtwo studies were examined and finally gaining a point of balance between privacy and economic interests were suggested as important issues and no solution was proposed[15].

Oppong, et al., in 2005 argued that the biggest challenge was not due to technology, but it was due to people and their behavior. Afterwards, he examinedthe growth trend of knowledge management in e-commerce and addressedidentifying used strategies. However, heonly referred to the problem of privacy and proposes no solution in this regard[12].

Hung, et al., in 2006 compared different models of data mining, which wereable score based on "propensity-to-churn"' periodically to their mobile subscribers. Heonly referenced to privacy ofdata collection and to solve this problem, they have used random selection[16].

Basiri in 2007has generally defined the data warehouse, data mining and CRM and then has examined the interaction between these concepts and their benefits with taking into consideration the privacy of customer and has briefly referenced to building of nickname structure to solve this problem, which issued in the discussion of data warehouse. Finally, hereferred to the necessity of making balance between privacy and economic interests[17].

Benjamin, et al., in 2007 onlyexaminedthe k-Anonymizationmethod in clustering and attemptedto find a balance point in preserving privacy and obtaining the necessary information for extractingthe required knowledge. Heargued that reaching to this balance point requiredtrade-off between obtaining information and remaining anonymous[14].

Ngai, et al., in 2008 reviewedall academic writings between 2000 and 2006, which included 900 articles in 24 journals on the relation between data mining and CRM,and provided researchers with a good reference for determining future research in this area. This shows the importance of the issue. However, in the review, no reference wasmade to the issue of preserving privacy[1].

Kiran, et al., 2012 examineda developed model in ETL processing in data miningto preserve privacy that was obtained of applying data mining algorithms and wasbeing prepared. The study hadno referred to CRM. Finally, hesuggestedthatprevious researches have not addressed this aspect of subject and hewas needed to develop a standard for future research in this regard[18].

Dorgalale, et al., in 2012 proposed a design for using K-Anonymity on distributed data with the aim of preserving privacy, which had applications in data mining and its application in CRM was not investigated. In addition, he suggested that the move towards distributed data is one of the future challenges in the field and proposes to create a standard way for privacy preservation in this environment[19].

Natarajan, et al., in 2012 introducedPrivacy Preserving Data Mining (PPDM) as a new field of research that is of great importance, which is related to issue of Privacy Preserving Data Mining. He suggested that the importance of the topic was increasing and discussed the different techniques in the field. Finally, referring to the fact that each of these methods approximately preservesthe privacy encourages for more research and creating techniques that are more efficient. Thisarticle did not address the subject from the aspect of CRM too[13].

Umamaheswari & Dr. Janakiraman in 2014 examined the different data mining techniques and their impact on CRM and referredto confidence on not interfering customer privacy[20].

Proposed method:

To propose for the area of preserving privacy on discovered patterns of data mining, first,the needed patterns of data mining on sensitive information of customer should be examined and then the sensitive information should be removed from it. Therefore, outsourcing data mining projects given access to confidential information with a third party company is not rational, especially that customers avoid the sharing of their confidential information. Therefore, to solve this problem, it suggests that in various parts of data mining projects, including data collection and

ETL and before running data mining algorithms and in particular the association rules algorithm, which has special uses in CRM, as possible to use techniques relating to privacy preserving,

A. Attribute and Concepts
B. Transformation, Constraints and Notes
C. Provider Relationship and Serial Composition
D. Part-of and Candidate Relationship
E. ECC and DD Transform

In this regard, we can by using techniques of creating alias structure and encryption to reduce sensitivity of data so that companies and consumers could easier to give their information to data miners. In addition, by using PPDM techniques, we can further reduce sensitivity. According to the following table[21], in this stage, various methods can be used, which among the association rule algorithms the best one is Perturbation Algorithms. Overall, the use of algorithms is not sufficient, but rather we suggest the use of hybrid methods, which combine two or more methods.

Table 1: Privacy Preserving Data Mining techniques comparisons &survey results

No	Method name	Concept	Results
1	k-anonymity and l-diversity	Reduce the granularity of the data representation using generalization. This granularity is reduced sufficiently that any given record maps onto at least *k* other records in the data. The *l*-diversity model was designed to handle some weaknesses in the *k*-anonymity model.	K-anonymity is able to preserve the privacy but l-diversity is much more secure.
2	Association Rule	Distributing the data vertically into equivalent segments	Association rules are generated which are distribution based for data mining and hence privacy preserving.
3	Data Perturbation	Adding a random noise to the original data to preserve the patterns to be accurately estimated	Random matrices are generated to keep data private.
4	Decision Tree	A tree structure is generated to identify the best attribute and the tree grows and apply the privacy over them.	A graph-based framework to prevent sensitive information from attacks.
5	Generalization and suppression	Replacing the attribute values with some character , special character, number etc. to unidentified the original record	The resultant dataset will have protected values so that no intruder can overcome the original information.
6	Classification framework	Uses the anonymization and perturbation approaches and their sub-methods to modify the original data	Modifies the data using the particular technique and generates secure datasets.
7	Cryptographic technique	Use any encryption/decryption algorithm so as to keep the data in a secure manner.	The information provided is encrypted with the help of some key or certifications that provides original data only upon the appropriate decryption method.
8	Condensation approach	Just the original data are modified which better helps in the privacy and works for pseudo-data.	The pseudo-data is non-significant as it has the same format as the original data.
9	Blocking based technique	The sensitive values of the attributes are replaced with unknown values and the final dataset is no longer mined.	Unknown values preserve privacy but reconstruction of original data becomes difficult.
10	Value class membership and distorted value	The data is split into various classes of similar type and the values are distorted with the help of some class functions	The original data are generated in the form of intervals through discretization and adding random perturbations.
11	Distributed privacy preservation	Data sets are partitioned either horizontally or vertically across the individual entities which wish to derive aggregate results.	In horizontally partitioned data sets, the individual records are spread out across multiple entities, each of which have the same set of attributes. In vertical partitioning, the individual entities may have different attribute of the same set of records.
12	Hybrid approach	Combining the methods and as a result the privacy can be maintained.	It is more accurate and it can keep data secure than any other technique.

CONCLUSIONS

Given the importance of data mining in CRM, especially in the banking field, access to information for data mining, given the privacy of customers' information, is very difficult and sometimes impossible. Finding appropriate strategies for preserving customer's privacy in mining of the customers' data warehouse and the use of it can encourage senior managers to issue permits for access to this information and motivate customers to provide some of their useful information for data mining. In this context, the impact of selecting appropriate strategies for each client or groups of customers and creating profitable long-term relationships with customers can be considered as an incentive for them to continue the process.

However, the remaining questions are that how far can this trend survive?Where is the balance point between privacy and economic interests?Can we expect a border for it?

REFERENCES

[1] E. W. T. Ngai, L. Xiu and D. C. K. Chau, "Application of data mining techniques in customer relationship

management: A literature review and classification," Elsevier, Expert Systems with Applications, vol. 36, pp. 2592-2602, 2008.

[2] Wanghualin, "Data Mining and Its Applications in CRM," in IEEE Second International ConferenceComputer Research and Development, Kuala Lumpur, 2010.

[3] H. Newton, Newton's Telecom Dictionary, CMP Books, 2011.

[4] M. Kantardzic, Data Mining: Concepts, models, methods, and algorithms, New Jersey: Wiley Inter Science, 2011.

[5] M. R. Mohlabeng and V. D. Walt, "The role of Data Mining in Customer Relationship Management," in International Conference on Emerging Trends of Computer & Information Technology, India, 2012.

[6] B. A. Galitsky, M. P. Gonzalez and C. I. Chesnevar, "A novel approach for classifying customer complaints through graphs similarities in argumentative dialogues: Decision Support Systems," Elsevier, Decision Support Systems, vol. 46, no. 3, pp. 717-729, 2009.

[7] V. Kumar and W. Reinartz, Customer Relationship Management: Concept, Strategy, and Tools, Springer, 2012.

[8] D. P. Goyal and S. Sharma, "Evaluating effectiveness of data mining software for CRM systems," in IEEE 6th International Conference Advanced Information Management and Service, Seoul, 2010.

[9] P. Shah Samandi, "Data mining in CRM," Tadbir Magazine, vol. 156, 2008.

[10] M. J. Dr. Tarohk and K. Sharifian, "Data mining is used to improve customer relationship management," Motaleat-e modiriat-e sanati, vol. 6, no. 17, pp. 153-181, 2010.

[11] D. Anthony, H. Oscar and J. Gandy, "All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining," Journal of Business Ethics , vol. 40, pp. 373-386, 2002.

[12] S. A. Oppong, D. C. Yen and J. W. Merhout, "A new strategy for harnessing knowledge management in e-commerce," Elsevier, vol. 27, pp. 413-435, 2005.

[13] R. Natarajan, R. Dr. Sugumar, M. Mahendran and K. Anbazhagan, "A survey on Privacy Preserving Data Mining," International Journal of Advanced Research in Computer and Communication Engineering, vol. 1, no. 1, pp. 103-112, 2012.

[14] C. M. Benjamin, F. K. Wang, S. Philip and Y. Fellow, "Anonymizing Classification Data for Privacy Preservation," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, pp. 1-14, 2007.

[15] C. Rygielski, J. C. Wang and D. C. Yen, "Data mining techniques for customer relationship management," Elsevier, Pergamon, Technology in Society , vol. 24, pp. 483-502, 2002.

[16] S. Y. Hung, D. C. Yen and H. Y. Wang, "Applying data mining to telecom churn management," Elsevier, vol. 31, pp. 515-524, 2006.

[17] M. Basiri, "Application of Data mining techniques in CRM," in The 4th e-Commerce Conference, Tehran, 2007.

[18] P. Kiran, S. Kumar and N. P. Kavya, "An Extended Conceptual Modeling for ETL Processes in Privacy Preserving Data Mining," International Journal of Future Computer and Communication, vol. 1, no. 3, pp. 264-266, 2012.

[19] N. Dorgalale, M. Parmouze, M. Honarmand and F. Sepehri, "Provide data on the distribution plan for the use of K-Anonymity Method for the privacy of persons," in National Conferences New Idea in Electrical Engineering, Isfahan, 2012.

[20] K. Umamaheswari and S. Dr. Janakiraman, "Role of Data Mining in CRM," COMPUSOFT, An i nternational journal of advanced computer technology, vol. 3, no. 2, pp. 75-78, 2014.

[21] S. Parmar, P. Gupta and P. Sharma, "A Comparative Study and Literature Survey on Privacy Preserving Data Mining Techniques," International Journal of Computer Science and Mobile Computing , vol. 4, no. 4, pp. 480-486, 2015.

★★★