ISSN:0975-9646

Anshul Arora et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5241-5244

Segmentation of Mobile Customers for Improving Profitability Using Data Mining Techniques

Anshul Arora[1] ,Dr. Rajan Vohra[2]
[1]Student of M.Tech (C.E) ,[2]Head of Department
[1,2]Department Of Computer Science & Engineering,P.D.M College Of Engineering
Sector 3A, Sarai Aurangabad, Bhadurgarh

Abstract-This work helps in identifying the activities Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d.we can calculate the score with the help of different- different attributes. These attributes are- STD Calls Local Calls, Local SMS, STD SMS, Roaming, Tariff Plan, and Data Plan
Second, Revenue maximization profiling-use of clustering for identifying maximal cluster. With the help of clustering profiling is done for maximization of revenue. This all will be done using data mining tools which will provide data statistics. In second problem Revenue maximization profiling was done by using clustering for identifying maximal cluster. Revenue maximization was done invidually for each cluster.

Key Words:Data base of mobile customers, Data mining, Weka tool, K-mean clustering and Profiling

1. INTRODUCTION

The telecommunications industry was one of the first to adopt data mining technology. This is most likely because telecommunication companies routinely generate and store enormous amounts of high-quality data, have a very large customer base, and operate in a rapidly changing and highly competitive environment. Telecommunication companies utilize data mining to improve their marketing efforts, identify fraud, and better manage their telecommunication networks. However, these companies also face a number of data mining challenges due to the enormous size of their data sets, the sequential and temporal aspects of their data, and the need to predict very rare events—such as customer fraud and network failures—in real-time. The popularity of data mining in the telecommunications industry can be viewed as an extension of the use of expert systems in the telecommunications industry.

Call Detail Records (CDRs) describe customer utilization behaviour. In this paper, clustering of mobile customers is done based on the call detail records and using Self Organizing Maps. SOMs convert high dimensional data to a lower dimension representation scheme (a two-dimensional map) that can be easily visualized and understood.

With the globalization of world economy and market internationalization, the market competition of the domestic mobile communications industry has become more reasonable and increasingly fierce. The fast-growing user group, diversified operations and competition environment have put forward higher requirements for the service quality of the mobile communications industry. The competition to acquire and retain customers among mobile service providers is fierce. The key to survival in this competitive industry lies in knowing the customers better. Different people have different preferences in using mobile telecommunication services and mobile phones. Consider for example popular mobile service SMS. For a given population, majority of subscribers may be sending SMS every day whereas few may be such that they do not use SMS frequently.

Significance Of The Problem

- Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d.We can calculate the score with the help of different- different attributes. These attribute are- STD Calls, Local Calls, Local SMS, STD SMS, Roaming, Tariff Plan, and Data Plan.
- Revenue maximization profiling-use of clustering for identifying maximal cluster.

2. RESEARCH METHODOLOGY

For solving the above two problems some research techniques and methodologies are used for obtaining the desired result. Some tools and algorithms are required for obtaining the result. Main steps under the research methodologies are:-

Review literature or research papers – first of all literatures and research papers were reviewed for getting more information about the problem and knowing which type of work was done by others on this topic and by which method.

Identify tools – then tools required for solving the problem were identified and the best tool was selected from all.

Study database attributes and data structure – attributes and structure of the database was thoroughly studied for finding out useful attributes from the mobile office.

Organize filed visits to mobile office, From there we get information about the network packs, mobile services and customer database.

Determine nature and definition of research problem and work flow of the problem for getting accurate and desired result.

Organize the database with useful attributes and populate it then perform data analysis using suitable tool e.g., WEKA in order to generate the result.

3. Conceptual Framework

Clustering technique of data mining is used to solve the problems of this work. Clustering is a method of unsupervised learning and a well known technique for statistical data analysis. Clustering is a division of data into different groups. Data are grouped into clusters in such a way that data of the same group are similar and those in other groups are dissimilar [7]. Clustering has many applications, including part family formation for group technology, image segmentation, information retrieval, web pages grouping, market segmentation, and scientific and engineering analysis [8].Clustering aim is the objects in a group should be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity within a group and the greater the difference between groups. the better the clustering [9].

K-Means [10] is one of the simplest unsupervised non-hierarchical learning methods among all partitioning based clustering methods. It classifies [11] a given set of n data objects in k clusters, where k is the number of desired clusters and it is required in advance.

There are two main parts:-

1. Profiling Of Customers

- **Segmentation of mobile customer improving profitability using mining techniques .**

The clusters obtained from ESOM need to be analyzed for the values depicted for the 16 features chosen .Component planes for each individual component could be drawn. The mean value for each of the 16 features for each cluster would be computed. Further derived measures for each cluster such as total number of Voice calls as a percentage of total outgoing calls, total number of SMS calls as a percentage of total outgoing calls etc. can be computed. Distinguishable marketing strategies can be designed for each cluster depending on the value depicted by the cluster for each of the derived measure. Following are some examples: Design suitable short message price policy for customers of cluster which has highest short messaging index value. Marketing managements need to encourage customers having low consuming ability to use more mobile service. It can be known which group of mobile customers often travels out and suitable roaming policy could be designed for customers belonging to this cluster.

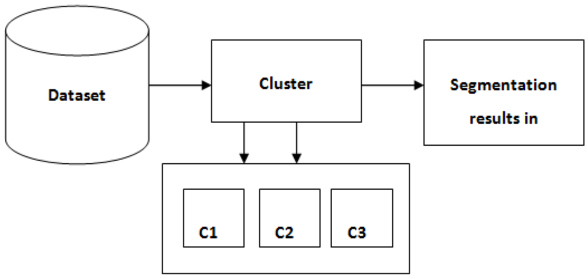


Fig. Profiling of customers

Score Calculation

Score calculation is the first problem of my work. For this one database was used. Calculation of score was done using different types of packages.

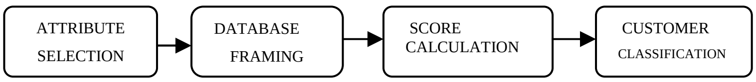


Fig- Framework to solve the problem.

Calculating Total Score –

Total score was calculated by summing the STD calls local calls local SMS, STD SMS, roaming, tariff plan, and data plan. Then we get total score for every entry package for mobile customer

Then total score is computed simply getting sum of ALL PACKAGES.

TOTAL SCORE = (LOCAL CALL) + (STD CALL)
+ (LOCALSMS) + (STD SMS)
+ (ROAMING) + (TARIFF PLAN)
+ (DATA PLAN)

	A	B	C	D	E	F	G	H	I	J	K
1	Name	Phone no.	Network Provider	Local Calls	STD calls	Local SMS	STD SMS	Roaming	Tariff Plan	Data Plan	TOTAL
2	Mohan lal	9818322445	A	125	50	100	80	20	11	1	507
3	Bhagwan das	9466091234	B	135	70	300	100	88	5	2	882
4	Shayam lal	8950134790	B	280	30	230	310	78	99	0	1197
5	Daleep kumar	9812356230	C	130	90	120	200	45	0	4	859
6	Manju sharma	8912569234	C	200	45	280	130	0	28	0	770.5
7	Garima	9466109010	B	88	23	389	400	100	11	3	1371.5
8	Seema Arora	8923134781	A	128	60	200	90	55	27	1	691
9	Munish Kumar	8912567980	A	180	28	340	26	89	0	0	690
10	Prem lal	9212511251	E	219	130	23	6	0	49	6	552
11	Sunder dass	9255448107	E	350	250	88	7	23	5	7	956.5
12	Sonia	9868673455	A	200	100	125	50	30	11	1	647
13	Deepika Singh	8967453623	E	231	300	135	70	90	49	6	1117
14	Minakshi	8956445566	C	45	230	280	30	45	39	4	928
15	Mukesh	9868957348	B	150	120	130	90	23	0	2	717
16	Priyanka	9868957325	C	29	280	200	45	60	28	5	873.5

Fig. Attributes of all networks.

Note: all the data has simulated. It is not correct information of customers.

2. Revenue maximization profiling-use of clustering for identifying maximal cluster

In second problem Revenue maximization profiling was done by using clustering for identifying maximal cluster. Revenue has been calculated manually and the formula is:-

$$\textbf{Revenue of Network Provider}(S) = \sum_{k=0}^{n} \textbf{Total}$$

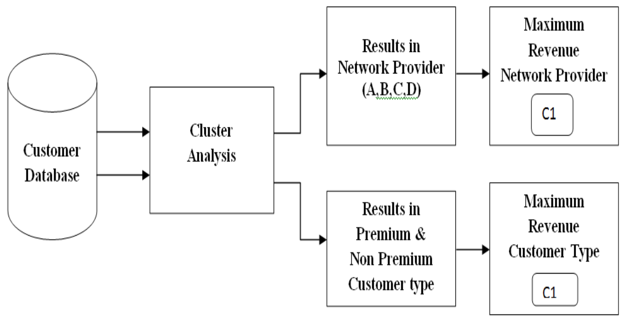


Fig-Revenue maximization

4. Research Background

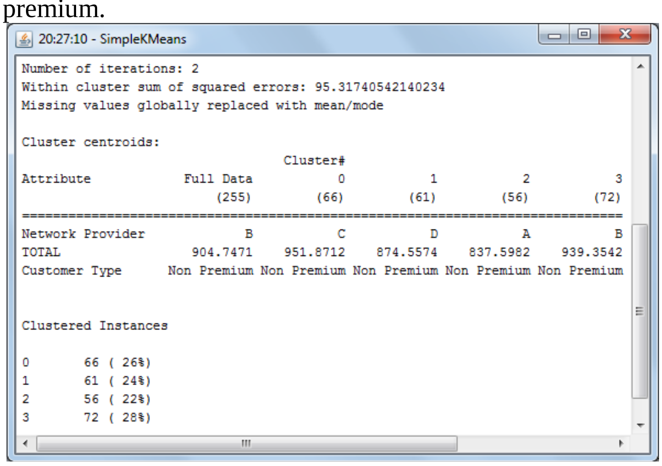
Gary M. Weiss, Fordham University, USA, in 2001 proposed that the telecommunications industry has been one of the early adopters of data mining and has deployed numerous data mining applications. The primary applications relate to marketing and network monitoring. Data mining in the telecommunications industry faces several challenges, due to the size of the data sets, the sequential and temporal nature of the data, and the real-time requirements of many of the applications. New methods have been developed and existing methods have been enhanced to respond to these challenges. The competitive and changing nature of the industry, combined with the fact that the industry generates enormous amounts of data, ensures that data mining will play an important role in the future of the telecommunications industry.

According to a Winter Corporation survey (2003), the three largest databases all belong to telecommunication companies, with France Telecom, AT&T, and SBC having databases with 29, 26, and 25 Terabytes, respectively. Thus, the scalability of data mining methods is a key concern. A second issue is that telecommunication data is often in the form of transactions/events and is not at the proper semantic level for data mining. For example, one typically wants to mine call detail data at the customer (i.e., phone-line) level but the raw data represents individual phone calls. Thus it is often necessary to aggregate data to the appropriate semantic level (Sasisekharan, Seshadri & Weiss, 1996) before mining the data. An alternative is to utilize a data mining method that can operate on the transactional data directly and extract sequential or temporal patterns (Klemettinen, Mannila & Toivonen, 1999.

5. Results And Discussion

Profiling of customers

In first problem that is profiling of customers. In this segmentation of mobile customer, customers are segregated into groups under the categorization of network provider, was done using different datasets. From these datasets we can calculate score .then according to the score we can classify the customer as Premium and Non premium.



There are FOUR types of clusters- cluster0, cluster1,CLUSTER2,CLUSTER3

Cluster 0 -> NETWORK C

Cluster 1 -> NETWORK D

Cluster 2 -> NETWORK A

Cluster 3 -> NETWORK B

1.1. CLUSTER 0 –

Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE

Total Entries = 66 (26%)

1.2. CLUSTER 1 –

Cluster 1 HAVING NON PREIMUM CUSTOMER TYPE

Total Entries = 61 (24%)

1.3. CLUSTER 2–

Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE

Total Entries = 56 (22%)

1.4. CLUSTER 3–

Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE

Total Entries = 72 (28%)

Revenue maximization:-

Revenue/ Network Provider	Cluster No.	Premium (Rs.)	Non Premium (Rs.)	Total (Rs.)	Max. Revenue
A	2	17,790	28,935.5	46,726	Non Premium
B	3	36,516	31,117.5	67,634	Premium
C	0	32,763	30,060.5	62,824	Premium
D	1	21,709	31,639	53,348	Non Premium
Total		1,08,958	1,21,752.5	2,30,532	**Non Premium**
Dominant		B	D	**B**	

A, B, C & D are different –different type of network provider. There are two categories of customers in different-different networks Premium and Non-premium.Acoording to revenue maximization A and D are Non premium type customers and B and C are premium type customers. In the premium and non premium type-B and D both are dominant respectively.Accroding to usage pattern total of all customer-B is the dominant.

*6.*CONCLUSION &FUTURE SCOPE

Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d.We can calculate the score with the help of different- different attributes. These attribute are- STD Calls, Local Calls, Local SMS, STD SMS, Roaming, Tariff Plan, and Data Plan. Revenue maximization profiling-use of clustering for identifying maximal cluster. . Revenue has been calculated manually

FUTURE SCOPE

There is always a scope of improvement in any research and so is with this work also. This work used simulated database and not real one, so it is a kind of algorithm using which real world problem can be solved in future if real world data becomes available.

In future we will match the entries between all four databases using template matching or any other matching algorithm; this will reduce the human work. We will also use different algorithms for clustering and compare the result among them.

ACKNOWLEDGEMENT

Authors would like to thanks to their head Dr. Rajan Vohra, HOD of CSE & I.T department, PDMCE, Bahadurgarh, for his valuable support and help.

REFERENCES

1. Adriaans, P., & Zantinge, D., (1996). Data Mining. Addison Wesley, Harlowe, EngIand.
2. Aggarwal, C. (Ed.). (2007). Data Streams: Models and Algorithms. New York: Springer.
3. Alves, R., Ferreira, P., Belo, O., Lopes, J., Ribeiro, J., Cortesao, L., & Martins, F. (2006).Discovering telecom fraud situations through mining anomalous behaviour Patterns. Proceedings of the ACM SIGKDD Workshop on Data Mining for Business Applications
4. Berry, M. J.A., & Linoff, G. S., (1997). Data mining techniques: for marketing, sales, and customer. John Wiley & Sons, Inc.
5. Berson, A., & Smith, S. J., (1997). Data Warehousing, Data Mining, and OLAP. McGraw-Hill, New York, NY.
6. Mattison, R., (1997). Data Warehousing and Data Mining for Telecommunications. Artech House, Norwood, MA.
7. Dr. Sankar Rajagopal, "Customer Data Clustering Using Data Mining Technique" International Journal of Database Management Systems (IJDMS) Vol.3, No.4, November 2011.
8. Pham, D.T. and Afify, A.A. (2006) "Clustering techniques and their applications inengineering". Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science.
9. Er. Arpit Gupta, Er.Ankit Gupta,Er. Amit Mishra," Research Paper on Cluster Techniques of Data Variations", International Journal of Advance Technology & Engineering Research (IJATER).
10. Hartigan, J., A. and Wong, M., A. 1979, "A K-Means Clustering Algorithm", Applied Statistics, Vol. 28, No. 1,.
11. Selim, S., Z. and Ismail, M., A. 1984, "K-Means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality", IEEE Trans. Pattern Anal. Mach. Intel., Vol. 6, No. 1,.Jiawei Han, Micheline Kamber 'Data Mining: Concepts and Techniques'
12. Fawcett, T., & Provost, F. (2002). Fraud Detection. In W. Klosgen & J. Zytkow (Eds.), Handbook of Data Mining and Knowledge Discovery.
13. Freeman, E., & Melli, G. (2006). Championing of an LTV model at LTC. SIGKDD Explorations.
14. Mani, D., Drew, J., Betz, A., & Datta, P (1999). Statistics and data mining techniques for lifetime value modeling. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM Press.
15. Rosset, S., Murad, U., Neumann, E., Idan, Y., & Gadi, P. (1999). Discovery of fraud rules for telecommu- nications—challenges and solutions. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 4. New York: ACM Press.)
16. Rosset, S., Neumann, E., Eick, U., & Vatnik (2003). Customer lifetime value models for decision support. Data Mining and Knowledge Discovery.
17. Sasisekharan, R., Seshadri, V., & Weiss, S (1996). Data mining and forecasting in large-scale telecommunica- tion networks. IEEE Expert
18. Provost, F., & Volinsky, C. (2006). Network- based marketing: Identifying likely adopters via con- sumer networks. Statistical Science
19. Kaplan, H., Strauss, M., & Szegedy, M. (1999). Just the fax—differentiating voice and fax phone lines us- ing call billing data. Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms
20. Philadelphia, PA: Society for Industrial and Applied Mathematics. Klemettinen, M., Mannila, H., & Toivonen, H. (1999). Rule discovery in telecommunication alarm data. Journal of Network and Systems Management.
21. Krikke, J. (2006). Intelligent surveillance empowers security analysts. IEEE Intelligent Systems.
22. I. O. Folasade, "Computational Intelligence in Data Mining and Prospects in Telecommunication Industry", Journal of Emerging Trends in Engineering and Applied Sciences
23. G. M. Weiss, "Data Mining in the Telecommunications Industry", IGI Global – Section: Service, 2009.
24. J. K. Pal, "Usefulness and applications of data mining in extracting information from different perspective", Annals of Library and Information Studies .
25. K. Tulankar, M. Kshirsagar, R. Wajgi, "Clustering Telecom Customers using Emergent Self Organizing Maps for Business Profitability", International Journal of Computer Science and Technology ,
26. Adriaans, P., & Zantinge, D., (1996). Data Mining. Addison Wesley, Harlowe, EngIand.
27. Berry, M. J.A., & Linoff, G. S., (1997). Data mining techniques: for marketing, sales, and customer. John Wiley & Sons, Inc.
28. Berson, A., & Smith, S. J., (1997). Data Warehousing, Data Mining, and OLAP. McGraw-Hill, New York, NY.
29. Mattison, R., (1997). Data Warehousing and Data Mining for Telecommunications. Artech House, Norwood, MA.
30. Russell, S., (1996). Neural Networks for Business Systems in Data Warehousing. McGraw-Hill, New York, NY.