Int. J. Advanced Networking and Applications 1176
Volume: 03; Issue: 03; Pages:1176-1179 (2011)

Multidimensional Association Rules Extraction in smoking Habits Database

Shashank Swami
MP Bhoj University, Bhopal
Email: Shashank_swami@rediffmail.com
R. S. Thakur
MaNIT, Bhopal, MP
Email: Ramthakur2000@yahoo.com
R. S. Chandel
Geetanjali Girls College, Bhopal, MP

----------**ABSTRACT**----------

Aiming at the smoking habits of youth in India, we collected the data from a survey, on which we form the multidimensional association rule and its model of smoking habits. Multidimensional association rule field of data mining is applied to discover the various habits and circumstances in smoking habits of youth. Based on the above approach we can take some preventive measures to reduce the various habits of smoking in youths.

Key words: smoking habit, data mining, multidimensional association rule.

Date of Submission: August 10, 2011 Date of Acceptance: October 02, 2011

1. Introduction

Smoking, chewing and other kinds of such addictions habits are increases day-by-day by the people of India. Tobacco products such as beedi (cheap form of cigarette), cigarettes, chutta (other kind of handmade cigarettes) and a mixture of tobacco and Gutkha are used by the people living in various regions and it increases the most common types of cancer in India [3][4]. We collected smoking habit data from the various survey reports, of a particular region on which we used to apply the mathematical statistics method to calculate the probability of the smoking and to analyze the causes of the smoking. The proposed method does not reflect all kind of conditional factors to stop the smoking habits in the youth.

It is well known that the smoking habits is not only related to the any single cause but there are many circumstances like family, friends circle, environment etc[3][4]. How to find out the issues and analyzes the cause from a large number of smoking habits data by multidimensional association rule. Association rule mining [5, 7] is a method to discover the association rules or relation in the large amount of data, that is to find out the predicates (attributes) which often occur together, then symbolizes their relations in regular form [1].

The multi-dimensional association rule [6] is an important technique in association rule mining which can decomposes the causes of the smoking habits and gain useful association rule expressions depending on the analysis of frequent degree decreasing the factors of smoking habits. The expression is like as [2] " Cause →smoking habits"

As it turns the single factor theory into systematic cause theory which finds the only cause through mining the conditional factors of the smoking habits. Based on the above facts we can reduce the smoking habits of the youth to prevent them form the various occurrences of diseases from these habits.

II. The Multidimensional data model in Smoking habits

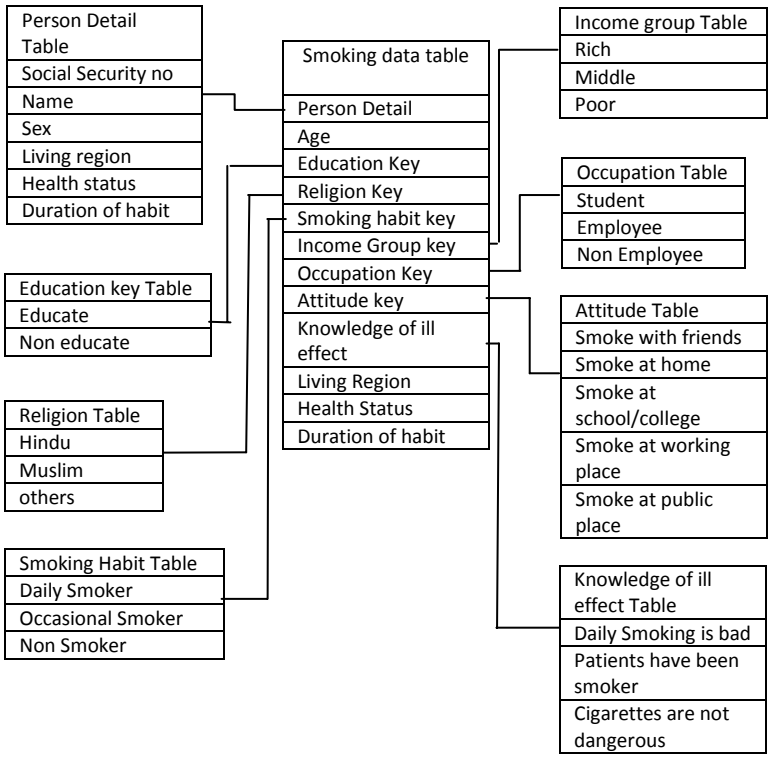
Multidimensional data model is a key element of the multidimensional association rule technique. The data that we are used to define the smoking habit are multidimensional, like other large – scale data. By the help of this multidimensional data model we can describe and analyze the multidimensional character of data, in order to create a multidimensional logic model for by which we can simulate the real life problems.

For this example we use the **star schema model** for representing the multidimensional databases. The data in the model is regarded as n-dimensional data cube, which consists of **"Dimension and Fact"** [1]. The word dimension represents the view of human beings observing the problems, and it corresponds with a dimension form, while the term fact is the measure around the main subject which corresponds with a fact form. The fields in the fact form contain mainly the name and the measure of the fact, and the key value of each related dimension form. As to the general data of smoking habits, the dimension is the character of how the smoking habits causes distributes in **Person detail, Age, Sex, Education, Religion, Smoking**

Volume: 03; Issue: 03; Pages:1176-1179 (2011)

habits, Income Group, Occupation, attitude, Knowledge of ill effect, Health status, Living region, Duration of habit, Smoking brands etc. where fields are taken from various surveys. And the fact is the information which is used to describe and measure the smoking habits. Thus the multidimensional data model in smoking habit built. According to the [1] when using the 2 – dimensional form to express the multidimensional concept, we can divide the multidimensional structure into two kinds of forms: one is fact from, which is used to storage the measures value of the fact and code value of each dimension; the other is dimension form. To each dimension, at least a form will be used to save the metadata which is the characterizing information of the dimension, including the level of the dimension, the division of the members, etc. In this model, each dimension of the smoking habit data has no level, there is only one row of the dimension form, in which holds all legal values of the dimension. In the related fact form, these values can derive the column of the dimension. This kind of structure, in which fact form associate with the dimension forms by the value of each dimension is called **Star schema** [1]. Fig. 1 shows the star schema data structure of surveyed smoking habit data. This structure can be described by relational database, and be implemented by defining the main external-key key relative relation between facts and dimension forms.

Fig. 1 Star Schema of a data warehouse for Smoking habits



In Fig. 1 we express the multidimensional concept by 2-dimension relation. The above star schema can be used to inquire data and can be implementing in the relational database. For the above multidimensional model during the mining of the multidimensional rule first we search for frequent itemset by using the general concept of algorithm which is discussed below:

Step 1: Input multidimensional data, *minimum_support* and *minimum_confidence*

Step 2: initialize the k = 1, L = ϕ{set of frequent itemsets}

Step 3: Produce 1 candidate itemset (C1) and frequent itemset (L1)

Step 4: do while (L_{p-1} ! = ϕ)

Step 5 : {

Step 6: Produce k – itemsets (C_k) from frequent itemsets (L_{k-1}) using *multidimensional_joining_process*

Step 7: Produce frequent itemsets (L_k) from k – itemsets (C_k)

Step 8: k ++

Step 9: }

Step 10: Produce frequent itemsets (L)

Step 11: Produce association rule (R) using association generation formula.

III. Experimental result:

To explore the behavior of above algorithm and for finding the multidimensional association rule we have choose the 4 fields (age, sex, education, habit) for the simplicity of explanation from the above database. Collected samples of data for these fields into the MS-Excel sheet shown by Fig.2 and use SAS 9.2 is used on the collected surveyed values of the 4 fields of database.

All the experiments for rules are performed on a Pentium 3 Ghz machine and 2 GB Ram, running Microsoft windows 2007. The different associations are formed on various minimum support values 2%, 5%, 10%, 15%, 20%, 25% and the threshold values are 10, 20, 40, 60, 80, and 100. The total generated rules shown by Fig.3, Fig.4, Fig.5, Fig.6, Fig.7, Fig.8 in the form of graphs for the above minimum support and minimum threshold values.

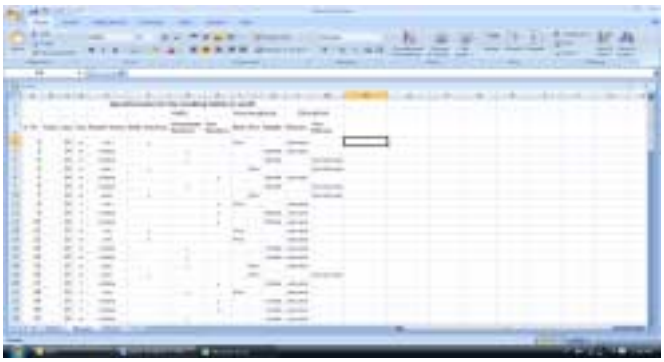


Fig2. Ms- Excel chart for the data smoking habits in youth

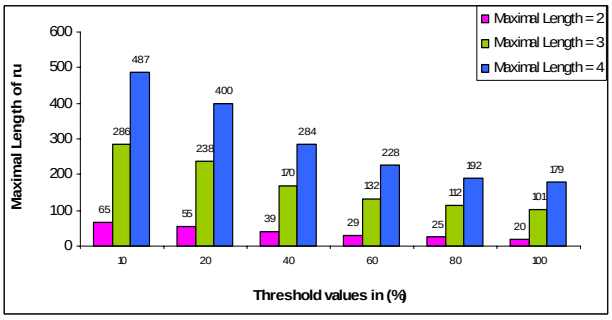


Fig3. For association rule when minimum support = 2%

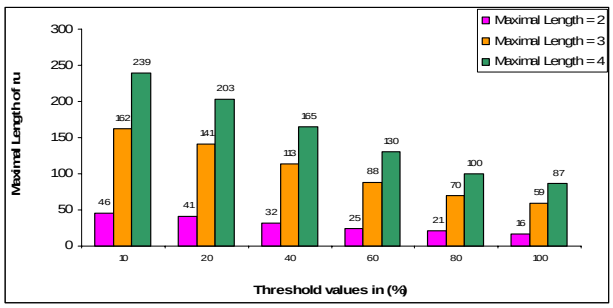


Fig4. For association rule when minimum support = 5%

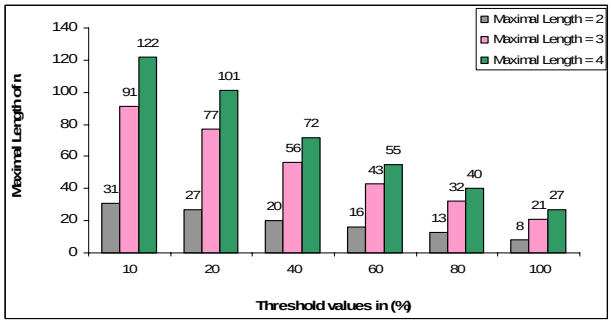


Fig5. For association rule when minimum support =10%

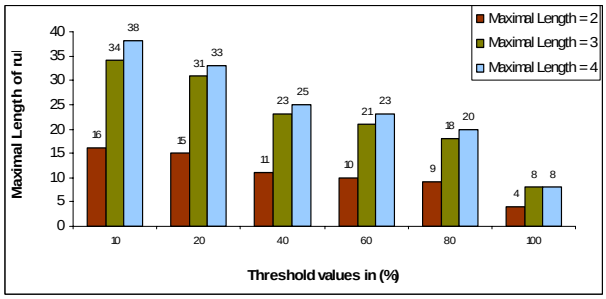


Fig6. For association rule when minimum support =15%

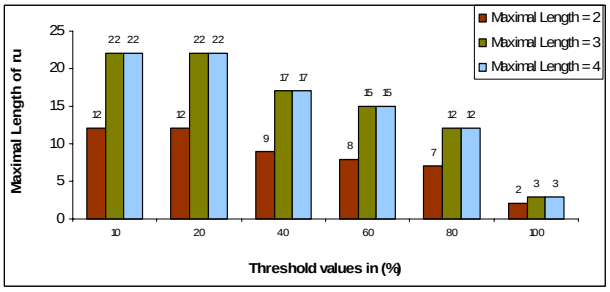


Fig7. For association rule when minimum support =20%

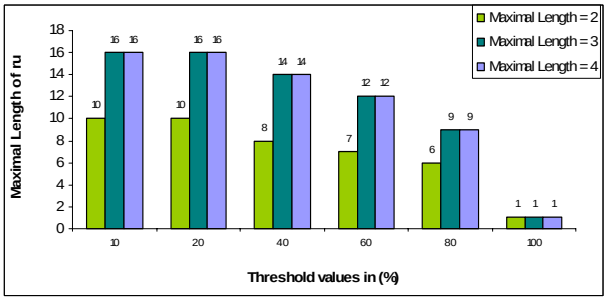


Fig8. For association rule when minimum support =25%

IV. Conclusion

Based on the analysis of smoking habits it is clear that if we use the multidimensional association rule technique the field of data mining we can deal with complex and multidimensional smoking habits data more accurately and also mine the conditional factors of the various types of smoking habits. Performance shows that method is the well improved method to implement well on this example and to explain the multidimensional association rule simply by this example. By this method it has following advantages.

1. By this we have to do less effort for the calculation of various relations.
2. By this method we can easily understand the various conditions responsible for the various smoking habits in the youth.

3. The method is a great help at some extant to stop the youth to prevent them from such addiction.

Thus we can provide an additional support to the prevention and to take preventive decision of the various smoking habits. This method prevent and analysis the various prevention method for the youth to prevent them from the various smoking habits. The technique also shows the use of multidimensional association rule on the data to find the very fast, efficient and strong analysis of results form the data.

Acknowledgement
This work is also supported by research grant from MANIT, Bhopal, India under grants in Aid scheme 2010-11, No Dean(R&C) /2010/63 dated 31/08/2010.

References

[1]. J. han and M. Kamber, data mining : concepts and techniques 2rd ed., Beijing: higher education Press, 2006, pp. 255 – 242.

[2]. H. Song – bai, W. Ya – sun, S. Yue-kun, G. Wen-wei, C. Qiang an ya-qin: research of multidimensional association rule in traffic accidents. IEEE Xplore, 2008.

[3]. S. Gavarasana, V.P. Doddi, V. S. N. R. Prasad, A. Allam, S. R. Murthy Jpn J. Cancer. Res, 82, 142-145, 1991.

[4]. R. P. Dikshit and S. Kanhere Tobacco habits and risk of lung, orphayngeal & oral cavity cancer: a population based case control study in Bhopal India. International Journal of epidemiology 2000, 29, 609 – 614.

[5]. H. Sug, Discovery of Multidimensional Association Rules Focusing on Instances in Specific Class: International Journal Of Mathematics And Computers In Simulation, Issue 3, Volume 5, 2011.

[6]. R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216.

[7]. R. S. Thakur, R. C. Jain, K. R. Pardasani, Graph theoretic Based Algorithm for Mining Frequent patterns in Proc. IEEE World Congress on Computational Intelligence, HonkKong, pages 629-633 June 2008.