Cosmetic Products Selection Using Reduct Generation Algorithm

Myat Thu Aung , Aung Cho
laminthawdar09@gmail.com, pothingyan.87@gmail.com
Computer University (Maubin)

Abstract

Data mining is the task of discovering interesting patterns from large amounts of data. These data can be stored in database. Data classification in data mining tasks is the process of building a model from available data called training data set and classifying objects according to their attributes. The rough set theory can be used for classification that offers a viable approach for extracting of decision rules from data set. This system describes to develop the improvement quality for selection attributes of cosmetic products by using reduct generation algorithm under rough set theory. Any features from cosmetic products are produced by using reduct generation algorithm from rough set theory. This system also expresses for improvement quality of cosmetic products from selection of users. In the analysis of make up users, the set of rules are useful in classifying data. Users select rules of make up by their appropriate skin and age.

Keywords: data mining, data classification, rough set theory, cosmetic products, reduct generation algorithm

1. Introduction

Data mining refers to extract knowledge from many large of data. Data mining is discovering knowledge is database. Classification method includes K-nearest neighbor, case-base reasoning, genetic algorithm, rough set and fuzzy set approaches. Data mining also called knowledge – Discovery Database (KDD) or Knowledge-Discovery and Data mining is the process of using tools such as classification, association rule mining and clustering etc. One of these, the rough set theory has been proved to be very useful in practice as clear from the record of many real life application; e.g. in medicine, pharmacology, engineering, banking financial, market analysis and cosmetic selection.

Data mining can be viewed as a result of the natural evolution of information technology. The database system industry has witnessed an evolutionary path in the development of the following functionalities-data collection and database creation, data management. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging.

Many people treat data mining as a synonym for another popularly used term, Knowledge Discovery from Data, or KDD. Data mining involves an integration of techniques from multiple disciplines such as database and data warehouse technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial or temporal data analysis. The discovered knowledge can be applied to decision making, process control, information management, and query processing. In this system, reduct generation algorithm under rough set theory is introduced.

In this system, a cosmetic product may be viewed as a solution to a need. Successful products are those that provide elegant and efficient solutions to customer needs. In order to develop elegant and efficient solutions to customer needs, two different types of information have to be combined: “need information” (what users need) and “solution information” (how products are selected). The value of a cosmetic product is a function of the quality of the solution that it offers and of the relevance of the needs that it solves.

The aim of this paper is to develop the system that selects automatically the quality of cosmetic product from the users’ appropriate skin and age. This paper is structured as follows: in Section (2) we present some related work. In Section (3) we give a brief overview related with rough set approach and discuss reduct generation algorithm by using the condition of user’s skin type and age to apply the system. In Section (4) we initiate the system design of cosmetic product selection. In Section (5) we describe the implementation of the system while we evaluate the system. The paper is finally concluded in Section (6).

2. Related work

A comparative study of the rough sets model versus multivariable discriminate analysis (MDA) can be found in [15]. It was demonstrated that through the use of rough set theory, the prediction of corporate bankruptcy was 97% accurate and improvement over MDA which achieved and accuracy of 96% in prediction of business failure. In the work reported in [5], the problem of how to deduce rules that map the financial indicators at the end of a month to the stock price changes a month later was addressed. This was based on 15 market indicators. From this study, only a satisfactory performance was achieved with many issues still to be tackled, such as data filtration and how to handle missing data.

In [8], research was carried out into rough set reduct analysis and rule construction for forecasting the total index of the Olso stock exchange. This also achieved satisfactory results, with a highest accuracy of 45% in financial investment. Rough set data analysis is also applied to the problem of extracting protein-protein interaction sentences in biomedical literature [4]. Due to the abundance of published information relevant to this area, manual information extraction is a formidable task. This approach develops decision rules of protein names, interaction words, and their mutual positions in sentences.

To increase the set of potential interaction words, a morphological model is developed generating spelling and inflection variants. The performance of the method is evaluated using a hand-tagged dataset containing 1894 sentences, producing a precision recall break even performance of 79% with leave one out cross validation. In data mining methods, there are many other Hybrid Approaches. A hybrid approach to Knowledge Discovery in Database (KDD) combines more than one approach and is also called a multi-paradigmatic approach.

Although implementation may be more difficult, hybrid tools are able to combines the strengths of various approaches. Deductive databases and genetic algorithms have also been in Hybrid Approaches. Such methods can be categorized into supervised and unsupervised: supervised learning methods have required knowing the value of class label attributes: although unsupervised learning methods do not require knowing the value of class label attribute. In general, supervised learning techniques enjoy a better success rate as defined in terms of usefulness of discovered knowledge. Classification approach is a supervised learning approach. Thus, classification is one of the most data mining tasks and we should use the classification approach if we have a training data set which include class label attribute.

3. Background Theory

Classification can describe different data classes or concepts for purpose of being able to use model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data. This system is describe the factors of quality cosmetic products selection by using one of the classification method of rough set theory. This theory describe for use in factors from imprecise data. Rough set theory is include with three basic component, granularity of knowledge, approximation of sets and data mining. One aspect of data mining is the finding of all reducts. The first step is rough set approach to generate all reducts is to form a discernibility matrix.

3.1 Discernibility matrix

The difference between the attributes of each pair of objects can be stored into a matrix called discernibility matrix. The reduct is the minimum elements in discernibility matrix. Let T= (U,A,C,D) be decision table, with U = {x_1, x_2, x_3, ..., x_n}. By a discernibility, matrix of T, denoted M (T), we will mean n x n matrix defined as A.

$$m_{ij} = \{a \in C; a(x_i) \neq a(x_j) \wedge (d \in D, d(x_i) \neq d(x_j)\} \quad \text{for } i, j = 1, 2, 3, \ldots\ldots n \tag{1}$$

C = {a, b, c, d} is called attributes
D = {E} is called decision

	a	b	c	d	E
x_1	1	0	2	1	1
x_2	1	0	2	0	1
x_3	1	2	0	0	2
x_4	1	2	2	1	0
x_5	2	1	0	0	2
x_6	2	1	1	0	2
x_7	2	1	2	1	1

The example of discernibility matrix is given as the above description. Using discernibility matrix, the reducts of the decision table can be found in Table 1. The reduct intersects all the elements of the discernibility matrix. The reducts can be obtained by using reduct generation algorithm. Using discernibility matrix, it is now possible to from the discernibility function using “and” operator, “$\wedge$” is called conjunction. We use term “or” operator “$\vee$” and “and” operator “$\wedge$” operator for simplicity.

Table 1: Reduct of Decision Table

	x_1	x_2	x_3	x_4	x_5	x_6
x_2						
x_3	b,c,d	b,c				
x_4	b	b,d	c,d			
x_5	a,b,c,d	a,b,c	-	a,b,c,d		
x_6	a,b,c,d	a,b,c	-	a,b,c,d	-	
x_7	-	-	a,b,c,d	a,b	c,d	c,d

The discernibility function is described by or-ing all the values in one enter in the discernibility matrix and then and -ing all these together. The discernibility function can often be simplicity by the process of absorption. For example suppose one of the disjuncts in the discernibility function is (a v b) while another disjunct is (a v b v c). since attributes a or b is required to satisfy the disjunct, the second disijunct will be satisfied by either a or b and attribute c is not required so this function eliminates the (a v b v c) term.

In order to proceed further it is necessary to know the expansion law, which is used in the algorithm. The expansion law (i) finds the attribute X that occurs most frequently (at least twice), (ii) applies AND of X and all other OR from of elements of the discernibility matrix which do not contain X, (iii) applies the connective AND between OR from of all the elements in which if the elements contains X eliminate X and combines the elements obtained from (a) and (b) by AND.

3.2 Reduct Generation Algorithm

The reduct generation algorithm is an important concept in rough set theory and data reduction is a main application of rough set theory in pattern recognition and data mining. In this system, the aim is to apply reduct generation algorithm under rough set theory to search for improvement quality of cosmetic products. The concepts of reduct generation algorithm are as follows:

Given $f = f1 \wedge f2 \wedge f3 \wedge \ldots \wedge fn$ is the discernibility function.

Step (1) Absorption law is applied to eliminate all disjunctive expressions which are supersets of another disjunctive expression.

Step (2) Each set is replaced to strongly equivalent attribute by dummy variable.

Step (3) The attribute is selected which belongs to the large number of conjunctive sets, numbering at least two and apply the expansion law.

Step (4) Thus, step 1 to 3 repeats until the expansion law cannot be applied for each component.

Step (5) All strongly equivalent classes are substituted for their corresponding attributes.

Step (6) The reduct is calculated in each component.

Step (7) The Integrated reduct is written in the system.

The calculation of reduct generation algorithm is shown in Table 2. In this table, the attributes of F1, F2, ….. F11, and F12 are expressed the following facts as user inputs. (F1: Age - 5-75, F2: Sweat Resistance, F3: Composition - Moisturizer, F4: Composition - Vitamin, F5: Composition – Milk Fluid, F6: Composition – Oil Free, F7: Expire time, F8: Spots, F9: Brightness or Lighting, F10: Resistance Time, F11: Smooth Touch, and F12: Natural Look)

Table 2: Calculation of Reduct Generation Algorithm

	x_1	x_2	x_3	x_4
x_2	F1,F5,F7, F8, F10, F11			
x_3		F2,F3,F4,F7, F8,F10,F12		
x_4	F2,F3,F4,F7, F8, F10		F1,F5,F7, F8,F10, F11, F12	

The above algorithm is illustrated by the following example.

Given:

F= {F1vF5vF7vF8vF10vF11} $\wedge$ {F2vF3vF4 vF7v F8vF10} $\wedge$ {F2vF3vF4vF7vF8F10vF12} $\wedge$ {F1 vF5vF7vF8vF10vF11vF12}

as user' s inputs to apply data set of discernibility Matrix in the system. Subsequently, this system applies absorption in law, as

{F2vF3vF4vF7F8vF10} $\subseteq$ {F2vF3vF4vF7vF8vF10vF12}

Therefore, we have, similar

{F2vF3vF4vF7vF8vF10} $\wedge$ {F2vF3vF4vF7vF8vF10vF12} = {F2vF3vF4vF7vF8vF10}

Hence, the discernibility relation becomes

F = {F1vF5vF7vF8vF10vF11} $\wedge$ {F2vF3vF4v F7vF8vF10} $\wedge$ {F1vF5vF7vF8vF10vF11v F12}

This system repeats absorption law,

{F1vF5vF7vF8vF10vF11} $\subseteq$ {F1vF5vF7vF8vF10v F11vF12}

Consequently, the discerniblity relation becomes

F = {F1vF5vF7vF8vF10vF11} $\wedge$ {F2vF3vF4vF7v F8vF10}

Also, this system is denoted F8vF10 = M.

Hence, the discerniblity becomes

F = {F1vF5vF7vMvF11} $\wedge$ {F2vF3v F4vF7vM}

The attribute "F7"appears most frequently. Using it applies expansion law:

F = [{F7}] $\wedge$ [{F1vF5vMvF11} $\wedge$ {F2vF3vF4 vM}]

Now, all the components are in simple form.

On replacing M by F8vF10, we have, F =[{F7}] $\wedge$ [{F1vF5vF8vF10vF11} $\wedge$ {F2vF3vF4 vF8vF10}]

The attribute "F8" appears most frequently.

[{F8}] $\wedge$ [{F1vF5vF10vF11} $\wedge$ {F2vF3vF4vF10}]

The attribute "F10" appears most frequently.

[{F10}] $\wedge$ [{F1vF5vF11} $\wedge$ {F2vF3vF4}]

The reduct of the first component is {F7}, {F8}, {F10} and the reduct of the second component is {F1,F2}, {F1,F3}, {F1,F4}, {F2,F5}, {F3,F5}, {F4,F5}, {F2,F11}, {F3,F11}, {F4,F11}

Hence, the intergrated reduct is {F7}, {F8}, {F10}, {F1,F2}, {F1,F3}, {F1,F4}, {F2,F5}, {F3,F5}, {F4,F5}, {F2,F11}, {F3,F11}, {F4,F11}

4. Design of the System

The design of the cosmetic product selection system is described detail specification of all system elements in Figure 1.

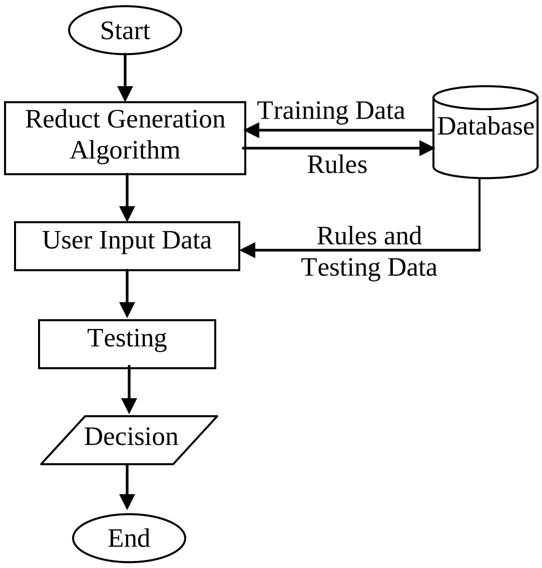

Figure1 : Design of System Architecture

This system describes the suitable cosmetic product selection by using reduct generation algorithm based on rough set theory. The cosmetic database consists of huge amount records in all cases: facts, decision with respect to users' inputs. The training data set has attributes, values and class labels. Initially, this system uses the training data to generate the minimum set rules with reduct generation algorithm. The set of rules are stored in the database and will be used for classifying the new user data. Also, these rules are compared and checked with user input data.

In this system, the user can choose the facts dealing with users' conditions based on skin and age. Then, this system tests the user input and can view the result to the user. Thus, to obtain the accurate decision for improvement quality of cosmetic products selection, this system decides that the cosmetic product is 'good' and 'poor'. In this system, cosmetic products selection data sets are 12 attributes and 2 target classes based on users' conditions with skin and age as shown in Table 3.

Table 3: Cosmetic Product of Make up Attributes Set

Attributes	Facts	Description	Value
F1	Age	User's age	5-75
F2	Sweat Resistance	Do you sweat resistance of make up?	1(yes) 0 (no)
F3	Composition	Moisturizer	1(yes) 0 (no)
F4	Composition	Vitamin	1(yes) 0 (no)
F5	Composition	Milk Fluid	1(yes) 0 (no)
F6	Composition	Oil Free	1(yes) 0 (no)
F7	Expire Time	Do you make up of expire time? (month)	6-12
F8	Spots	Do you loss of spots on face?	1(yes) 0 (no)
F9	Brightness or Lighting	Do you make up of lighting?	1(yes) 0 (no)
F10	Resistance Time	Do you make up of resistance time?	hr$\geq$ 6
F11	Smooth Touch	Do you make up smooth touch?	1(yes) 0 (no)
F12	Natural Look	Do you natural look?	1(yes) 0 (no)

As well in this system, the classes are good and poor.

5 Implementation of the System

This system is intended to improve the quality of cosmetic products selection system by using one of the reduct generation algorithms of rough set theory. To make rule sets for the system, the user collects the training data in the database. After that, these data are executed by using reduct generation algorithm as shown in Figure 2. Consequently, the set of rules are used for classing new user input data. Thus, discovered rules have been successfully evaluated by expert beautician in the system.

Then, the user is going to select 12 attributes options for user's condition and clicks 'Calculate' button to find out the decision result. According to testing process, this system implements the decision result as 'good' or 'poor' for best quality of cosmetic products.

ID No	Age	Sweat resistant	Moisturizer	Vitamin	Milk fluid	Oil free	Expire time	Loss of spots	Brightness	Resistance time	Smooth touch	Natural look	Decision
1	20	Yes	No	No	Yes	Yes	4	Yes	No	4	No	No	Poor
2	32	Yes	Yes	No	Yes	Yes	6	No	No	8	No	Yes	Good
3	45	No	Yes	No	Yes	Yes	8	No	No	8	Yes	Yes	Good
4	48	No	Yes	No	Yes	No	4	Yes	Yes	4	Yes	No	Poor
5	17	Yes	Yes	Yes	Yes	Yes	6	No	Yes	8	No	Yes	Good
6	70	Yes	No	Yes	No	Yes	4	Yes	No	6	No	Yes	Good
7	28	No	No	Yes	No	Yes	4	No	Yes	4	No	No	Poor
8	52	Yes	No	No .	No	Yes	6	No	Yes	6	Yes	No	Poor

Figure 2: Generate Rule Sets by using Reduct Generation Algorithm

6. Conclusion

Rough set turned out to be a very useful method for generation decision rule from cosmetic of make up data. . Rough set can also be used for feature reduction and relevance analysis. It evaluates significance of data and can be helpful in generating sets of decision rules from data. Reduct generation Algorithm is evaluates significant of data and can help for decision from user's data.

It helps and guides in order to quick and correct cosmetic product selection. The decision table is huge in size with respect to number of records but this system can reduce the set of rules by using reduct generation algorithm. This system can have correct and accurate decisions for cosmetic products selection by using reduct generation algorithm of the rough set theory. It can get quick and correct choice whether user is selecting of cosmetic products or not.

REFERENCES

[1] A.I. Dimitras, R. Slowinski, R. Susmaga, C.Zopounidis. Business failure prediction using rough sets. European Journal of Operational Research, vol. 114, pp.263-280, 1999.

[2] A. Skowron and C. rauszer, "The discernibility matrices and functions in information systems", Fundamental Informaticae 15 (2), pp. 331-362, 1991.

[3] D. Slezak. Various approaches to reasoning with frequency-based decision reducts: a survey.

[4] F.Ginter, T.Pahikkala, S.Pyysalo, J.Boberg, J.J"arvinen, T.Salakoski. Extracting Protein-Protein Interaction Sentences by Applying Rough Set Data Analysis. In : Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence 3066, pp.780-785, 2004.

[5] J.Bazan, A.Skowron, P.Synak. Dynamic reducts as a tool for extracting laws from decision tables. In : Z.W. Ras, M. Zemankova (Eds.), Proceedings of the 8th Symposium on Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence 869, Springer-Verlag, pp.346-355, 1994.

[6] J.F.Peters, Z.Suraj, S.Shan. S.Ramanna, W.Pedrycz, N.J.Pizzi. Classification of meteorological volumetric redar data using rough set methods. Pattern Recognition Letters, vol. 24, no.6, pp. 911-920, 2003.

[7] J.Han and M.Kamber, "Data Mining concepts and Techniques", ISBAN 1-55860-489, Morgon Kaufmann.

[8] J.K.Baltzersen. An attempt to predict stock market data: a rough sets approach. Diploma Thesis, Knowledge Systems Groups, Department of Computer Systems and Telematics, The Norwegian Institute of Technology, University of Trondheim, 1996.

[9] K.Farion, W.Michalowski, R.Slowinski, S.Wilk, S.Rubin. Rough Set Methodology in Clinical Practice: Controlled Hospital Trial of the MET System. In: Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence 3066, pp.805-814, 2004.

[10] K.Mannar and D.Ceglarek. Continuous Failure Diagnosis for Assembly Systems using Rough Set Approach. Annals of the CIRP, vol.53, pp.39-42, 2004.

[11] M.P.Hippe. Towards the Classification of Musical Works: A Rough Set Approach. In: Proceedings of the 2002 International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence 2475, pp. 546-553, 2002.

[12] L.Lazareck, S.Ramanna. Classification of Swallowing Sound Signals: A Rough Set Approach. In: Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence 3066, pp.679-684, 2004.

[13] L.Shen, F.E.H. Tay, L.Qu and Y. Shen. Fault diagnosis using Rough Sets Theory. Computers in Industry, vol. 43, pp.61-72, 2000.

[14] Pawlak Z:Rough Sets. Theoretical Aspect of Reasoning about Data, Kluwer Academic Publisher, Dordrecht, 1991.

[15] A. Szladow, D. Mills. Tapping financial databases. Business Credit, vol.95, no. 7, p.8, 1993.