IIMB Management Review (2010) 22, 16—24

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/iimb

Attribute selection in marketing: A rough set approach

Sabita Mahapatra [a], Sreekumar [b], S.S. Mahapatra [c,*]

[a] Marketing, Indian Institute of Management Indore, Pigdamber, Rau, Indore 453331, Madhya Pradesh, India
[b] Rourkela Institute of Management Studies, Rourkela 769015, India
[c] Department of Mechanical Engineering, National Institute of Technology, Rourkela 769008, India
Available online 13 April 2010

KEYWORDS
Rough set;
Marketing;
Attributes;
Rule discovery

Abstract Using an illustrative case study on the Indian cosmetic industry, this paper illustrates the advantages of the rough set approach over conventional techniques for the extraction of decision rules from data sets, which can be useful in various marketing applications. The rule generated through the methodology can act as an 'expert', which may be referred to in future strategic decision-making. The approach gives results similar to the results obtained through statistical methods but without making any assumption.

Introduction

The IT revolution has radically changed the way data is collected and generated, facilitating the process of decision making. A huge set of data has no practical relevance unless it can be mined to provide useful information pertaining to the interests of the organisation. The patterns in the data need to be deciphered in order to gain insights about aspects such as customer preferences, market trends, and business performance. Quick responses to the changing market environment are possible only if timely and accurate insights into the business and the market conditions are readily available (Sreekumar & Panda, 2005). The perpetually growing volume of data needs to be reduced into useful information, and this calls for tools that are capable of distinguishing the various properties of the data collected/generated. Such dimensionality reduction would enable companies to remain more focused, and would thereby reduce the labour and communication costs for data collection. Researchers generally tend to apply statistical inferences on the existing data, emphasising efficient use of organisational data through data mining and data warehousing (Ha & Park, 1998). However, the conventional techniques cannot reduce the data dimensions efficiently and the persistent redundant attributes would affect the rule discovery process, leading to highly degraded rules (Zhong, Dong, & Ohsugu, 2001).

All these factors have opened up the scope for some of the newer techniques which have been developed in recent years (Beynon, Curry, & Morgan, 2001). In this paper, the rough set theory (RST) developed by Pawlak (1982) is adopted as an alternative technique for the extraction of decision rules from data sets. The rough set approach has several advantages over the conventional methods (Dimitras, Slowinski, Susmaga, & Zopounidis, 1999; Shen & Loh,

* Corresponding author. Tel.: +91 0661 2462512; fax: +91 0661 2472926.
E-mail addresses: mahapatrass2003@yahoo.com, ssm@nitrkl.ac.in (S.S. Mahapatra).
0970-3896 Peer-review under responsibility of Indian Institute of Management Bangalore.
doi:10.1016/j.iimb.2010.03.001

Production and hosting by Elsevier

2004). The tool is based on the original data, and does not need any external information. It is a tool suitable for analysing quantitative as well as qualitative attributes. This tool discovers important facts hidden in the data set, and expresses these facts in the natural language of decision rules or computational algorithms, and not as mathematical functional forms (Wolfram, 2002). The former is better at pattern recognition than the latter, and so can have better managerial applications. The set of derived decision rules gives a generalised description of the knowledge contained in the database, eliminating any redundancy inherent in the original data. The derived decision rules are based on facts—each decision rule is supported by a set of real examples. The results of rough sets are easy to understand and process, while the results of the other methods usually require an interpretation of the technical parameters with which the user may not be familiar.

This paper uses the basic ideas of RST to show how rule discovery can be made, and to present the relationship between the attributes. The fundamental concepts of the rough set approach are briefly explained in the following section. A case study is undertaken based on the data collected from twenty-three Indian cosmetic companies. The names of the companies have been withheld to maintain confidentiality.

Fundamentals of rough set theory

The rough set theory (RST) was developed by Pawlak (1982) at the Institute of Computer Sciences, Warsaw. It was initially proposed as an alternative data analysis method but subsequently found application in the areas of artificial intelligence, knowledge discovery, decision analysis, and expert systems among others. RST can deal with inexact, uncertain, and vague datasets (Shyng, Wang, Tzeng, & Wu, 2007); it is a new mathematical approach to vagueness. According to Pawlak and Skowron (2007), the rough set philosophy was founded on the assumption that some information (data, knowledge) is associated with every object of the universe of discourse. For example, if the objects under study are patients suffering from a certain disease, the symptoms of the disease form the information about the patients. Objects characterised by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of RST. The rough set theory has been applied in various fields like marketing, banking, engineering, and medicine among others.

In RST, data is represented through a data table also known as an attribute-value table, an information table or a database. The rows of the table stand for the *objects*, the columns represent the *attributes*, and the entries are called *attribute values*. A database S is a pair represented by $S = \{U, A\}$, where U and A are both finite non empty sets; U is the universal set and A is the set of attributes. The subset of attributes in the database is the cluster of objects having the same attribute values or the same features. Objects that have the same features are indiscernible (similar); these blocks provide the elementary granules of knowledge. These granules are called concepts or elementary sets, and form the elementary building blocks (atoms) of knowledge. Any union of elementary sets is called a crisp (precise) set, and any other set is referred to as a rough set (vague, imprecise). Associated with every set X, there are two crisp sets called the lower and the upper approximations of X. The lower approximation of X is the union of all the elementary sets which are included in X, and the upper approximation of X is the union of all the elementary sets which have a non-empty intersection with X.

Exhibit 1 illustrates these concepts using the data of six objects (companies), the three attributes $\{a_1, a_2, a_3\}$, and the decision state D (Profit); the three attributes used are availability of research and development (R&D) facilities, adoption of state of the art technology, and marketing expenditure. These three variables could be considered to represent the independent variable, and proper utilisation of these resources in the right combination could lead the company to profit. Profit is the result of the decision variable, and it can take two attribute values—yes or no; so Profit can be considered to be the decision state. In principle, there can be more than one decision variable, but in this case only the decision variable Profit has been considered, which is a *distinguished* attribute. The methodology demonstrated here could also be used to discover rules that distinguish small profit-making firms from large loss-making firms.

$C3$ has R&D facilities and state of the art technology, and spends a very high amount for its marketing activities, and the attribute value of the decision variable (Profit) is Yes; i e, the company $C3$ makes profits. The information about $C3$ has the following attribute values: (a_1, Yes), (a_2, *Yes*), (a_3, *Very High*), (D, *Yes*).

Let U denote the set of all cases, A the set of all attributes, and V the set of all attribute values. A table such as the one in Exhibit 1 would define an information function ρ: $U \times A \rightarrow V$. The attribute values can also be written as a function, of the form $\rho(C3, a_1) = Yes$.

Let $a \in A$, $v \in V$, and $t = (a, v)$ be an attribute-value pair. A block of t, denoted by $[t]$, is a set of all cases from U for which the attribute a has the value v. So the information

Exhibit 1 Representation of sample database with six objects and three attributes.

Company	Availability of R&D facilities (a_1)	Adoption of state of the art technology (a_2)	Marketing expenditure (a_3)	Profit (D)
$C1$	No	Yes	High	Yes
$C2$	Yes	No	High	Yes
$C3$	Yes	Yes	Very high	Yes
$C4$	No	Yes	Average	No
$C5$	Yes	No	High	No
$C6$	No	Yes	Very high	Yes

in Exhibit 1 can also be represented as [(R&D, *Yes*)] = {*C*2, *C*3, *C*5}; [(R&D, *No*)] = {*C*1, *C*4, *C*6}; [(*Marketing*, *High*)] = {*C*1, *C*2, *C*5}; [(*Marketing*, *Average*)] = {*C*4}, and so on.

We now introduce the concept of indiscernibility with respect to more than one attribute, and the decision consequence. *C*2, *C*3, and *C*5 are indiscernible with respect to the attribute a_1. Similarly *C*3 and *C*6 are indiscernible with respect to the attributes a_2 and a_3. The indiscernible matrix corresponding to the sample database in Exhibit 1 is represented in Exhibit 2, which can be used to find the lower approximation, the upper approximation, and the boundary cases for the profit making set, which is defined as decision state *D*: (profit, *yes*). (*C*1 has been dropped from the *objects*, and *C*6 from the *attributes* in the indiscernible matrix.).

All the three attributes a_1, a_2, and a_3 are present in the intersection of *C*2 and *C*5, but *D* is absent, which means that even though all the attribute values of these two companies match, their decisions are different. So for *C*2 and *C*5, the decision variable *D* cannot be characterised by the attributes a_1, a_2, and a_3. Hence, *C*2 and *C*5 form the boundary line cases which cannot be accurately classified with the available knowledge. *C*1, *C*3 and *C*6 are profit making companies, and form the lower approximation of the set. *C*2 and *C*5 cannot be excluded from the set of profit making companies with certainty; and *C*4 does not earn profit. So the upper approximation contains *C*1, *C*2, *C*3, *C*5 and *C*6.

So for the decision variable *D*: (profit, *yes*), the lower approximation of the set is {*C*1, *C*3, *C*6}, the upper approximation is {*C*1, *C*2, *C*3, *C*5, *C*6}, and the boundary line cases are *C*2 and *C*5. Similarly *C*4 is not a profit making company, and *C*2 and *C*5 cannot be excluded from the set of non-profit making companies. So for the decision variable *D*: (profit, *no*), the lower approximation of the set is {*C*4}, the upper approximation is {*C*2, *C*4, *C*5}, and the boundary line cases are *C*2 and *C*5.

The concepts of reduct and core set can also be used for rule discovery from the database. In practical applications, it can often be observed that some attributes of an information system may be redundant or superfluous with respect to a specific classification A^* generated by the attributes $A \subseteq Q$, where *Q* is the finite set of attributes. Using the dependency properties of attributes, one can find a reduced set of the attributes by removing the superfluous ones, without loss of the classification power of the reduced information system. The set of all indispensable attributes in the set $A \subseteq Q$ is called the *core* of *A* in *S*, where *S* is any information system. The core contains all the attributes that cannot be removed from the set *A* without changing the original classification $A*$ (Swiniarski, 2001). These concepts of reduct—sufficient information and core (minimum sufficient information)—can be used to generate rules for the discrimination of sets or objects.

Literature review

Mckee (2000) develops a bankruptcy model using the rough set approach. The model is 93% accurate in predicting bankruptcy on a developmental model composed of 100 companies, and 88% accurate in predicting bankruptcy with separate holdout samples of 100 companies. Tseng and Huang (2007) apply RST for feature selection in customer relationship management (CRM); they present both the mathematical formulation and the heuristic algorithm to derive the decision rules from historical data for identifying the features that contribute to CRM. Shen and Loh (2004) use RST to retrieve knowledge that could guide investors on when to buy and sell.

Huang, Liu, Ou, Yao, and Zhong (2003) apply the algorithm of attribute reduction—based on a combination of RST with the boosting algorithm—to the linear model of market value functions, which is a new method of direct marketing. In the direct marketing problems, it is crucial to reduce the attributes in order to deal with the large databases. In another study, the data from the financial statements of 240 such businesses was used to compute financial ratios (Bose, 2006). The rough set technique was used to evaluate whether the financial ratios could predict financial health based on the available data. The most predictive financial ratios were identified, and interesting rules concerning the financial ratios and the financial health of dot-coms were discovered. Rough sets were found to satisfactorily predict financial health, and were considered more suitable than the other contemporary techniques for detecting unhealthy dot-coms. As illustrated in this study the rough set approach helped identify which aspects of financial statements were needed to decide the financial future of the dotcom. It also led to the creation of rules linking the dependent and the independent variables, which is valuable to financial analysts. A classification system should provide an explanation of the decision, and in a rough set analysis this is provided by the rules that are discovered by the system. Another benefit is that the rules are based on the data and are supported by real examples, thereby improving the validity of the results and making them understandable.

Swiniarski (2001) applies the rough set approach and statistical methods to feature reduction and pattern recognition, emphasising the role of rough set reducts in feature selection and data reduction in pattern recognition. The paper also contains a description of the algorithm for feature selection and reduction based on a combination

Exhibit 2 Indiscernible matrix corresponding to the sample database.

	*C*1	*C*2	*C*3	*C*4	*C*5
*C*2	a_3, *D*				
*C*3	a_2, *D*	a_1, *D*			
*C*4	a_1, a_2	—	a_2		
*C*5	a_3	a_1, a_2, a_3	a_1	*D*	
*C*6	a_1, a_2, *D*	*D*	a_2, a_3, *D*	a_1, a_2	—

of the rough sets method with principal component analysis.

Many rough set algorithms are used for feature selection and rule discovery. The fundamental method of finding minimal reducts is to generate all possible reducts, and then to choose any with minimal cardinality; this can be done by constructing a kind of discernibility function from the dataset and simplifying it (Bazan, Nguyen, Nguyen, Synak, & Wróblewski, 2000; Wang, Yang, Teng, Xia, & Jensen, 2007). There are a number of softwares available for rough set applications, like the Rough Set Exploration System (RSES) and the Rough Set Data Explorer (ROSE). RSES can be used to find the reducts, generate decision rules using the reducts, decompose large data into parts that share the same properties, search for patterns in the data etc (see RSES 2.2 User's Guide, 2005). ROSE is an interactive system designed for analysis and knowledge discovery based on RST (see Predki, Słowiński, Stefanowsk, Susmaga, & Wilk, 1998 for a review of this software).

Data set

Data from twenty-three Indian cosmetics companies was considered for our analysis. Information pertaining to the expenditure of the companies on five parameters—marketing, advertising, distribution, miscellaneous, and research and development—was collected over a three-year period, from 2003 to 2005. These five parameters form the attribute set for our analysis. We also considered the sales figures of these companies over the same period, which became our decision variable. In this context, the marketing expenditure (abbreviated as Mkt, and notated as a_1) includes all the expenditure incurred for corporate promotion, which includes event marketing, sales promotion, direct marketing etc. The advertising expenditure (Advt, a_2) includes promotional activities through various media like television, newspaper, Internet etc. The distribution cost (Dist, a_3) includes the expenses incurred for logistics, supply chain etc. The miscellaneous expenditure (Misc, a_4) is mainly incurred through activities like corporate social responsibility. The investments made on the development of new products, and other research activities constitute the R&D expenditure (R&D, a_5). The sales represent the total sales made by the company (Sales, notated as D). The average of the data collected is considered to be the representative figure, and is tabulated in Exhibit 3. We use the notation C_i (where $i = 1, 2, \ldots, 23$) instead of the actual names of the companies.

Elementary exploratory data analysis

The analysis of the data from the twenty-three cosmetics companies is detailed below.

Descriptive data analysis

Fundamental descriptive statistical analysis tools were applied to the original data collected, and the results are represented in Exhibit 4.

The positive skewness results show many values at the low end, and a few at the high end. The negative skewness

Exhibit 3 Expenditure on five parameters and total sales of 23 Indian cosmetics companies.

Company	Mkt (a_1)	Advt (a_2)	Dist (a_3)	Misc (a_4)	R&D (a_5)	Sales (D)
$C1$	2.093333	1.516666667	0.373333	1.81	0	15.40333333
$C2$	18.27667	162.2366667	30.23667	72.14667	9.156667	1220.586667
$C3$	0.823333	0	0	2.84	0.973333	50.40666667
$C4$	2.076667	5.393333333	6.793333	8.29	0.383333	215.7666667
$C5$	0.496667	1.33	0.433333	2.733333	0.393333	42.59333333
$C6$	0.94	0.06	0.666667	5.89	1.243333	166.41
$C7$	27.33333	38.66	16.49667	24.34333	1.523333	561.6966667
$C8$	6.166667	0	7.046667	5.55	0	195.2466667
$C9$	45.48667	0	7.313333	4.25	0	197.45
$C10$	7.033333	866.9166667	508.6767	637.53	38.96333	11449.56
$C11$	0.026667	0.043333333	0	0.47	0	19.23
$C12$	4.323333	4.173333333	1.753333	3.176667	0.003333	60.89
$C13$	38.51667	40.04666667	3.126667	8.026667	0.056667	303.57
$C14$	13.13333	5.896666667	2.456667	4.086667	0.15	110.9266667
$C15$	19.37333	32.68	17.96333	28.62	0	2416.386667
$C16$	0.603333	0.036666667	0.393333	0.613333	0.016667	20.52333333
$C17$	0.43	0.183333333	0.293333	0.433333	0	21.88666667
$C18$	2.49	0	0	1.37	0.086667	80.08
$C19$	14.62333	47.74666667	20.68	31.23667	0.24	627.0433333
$C20$	0.746667	0	8.896667	1.28	0	325.79
$C21$	1.476667	0.046666667	0.12	0.286667	0	17.8
$C22$	2.08	1.453333333	0.56	0.693333	0	11.94333333
$C23$	0.466667	0.46	0.993333	3.803333	0.053333	62.83666667

All non-ratio figures are ten million INR.
Source: CMIE-PROWESS database.

results show many values at the high end, and few at the low end. It can be observed from the descriptive statistics that all the skewness values are positive, and the skewness and the kurtosis values are significant.

Correlation analysis

To find the degree of association between the attributes and the decision variables, we applied a correlation analysis. The correlation coefficients calculated between the various parameters are shown in Exhibit 5.

There is a very high degree of significant positive correlation in all the columns except in the first column, i e, for marketing expenditure. The correlation value of all the other variables with marketing is very low, and is not significant. Referring to the descriptive statistics to throw some light on this situation, we find that companies spend heavily on advertising, but a proportional amount is not spent on marketing. No negative correlation is observed in the correlation matrix.

Rough set analysis and results

The rough set approach operates on a data set or information table (as shown in Exhibit 3) which contains data about the universe *U*, the attributes, and the decision variable. The objective is to derive rules, which would be useful in finding how the decision variable depends on the condition attributes. We derive the rule by partitioning the universe *U* into a finite number of blocks called equivalence classes. We first normalised the data in the database given in Exhibit 3, and formed a classifying rule to categorise the attribute values into Low (normalised value < 0.3), Average (normalised value $0.3 - 0.7$), and High (normalised value > 0.7). (Expert opinion or managerial consensus may be used to find the different cut-off points for classifying the variables into the Low, Medium, and High categories.) The predictive model using regression analysis—a usual method of model building in marketing research—causes loss of information, often leading to misclassification. This drawback can be easily eliminated using the rough set approach. The modified information table is shown in Exhibit 6.

Based on the value of the decision variable, Exhibit 6 can be broken down into three blocks as shown in Exhibit 7. *C*4, *C*8, *C*9, *C*13, and *C*20 may be categorised into $D =$ Low as well as $D =$ Average because they are boundary line cases based on the value of the decision variable. In fact, they are the members of a rough set because they do not precisely belong to either the lower approximation set or the upper approximation set.

Case 1: $D =$ High

If a_1 and a_2 are Average and the other three attributes are Low, then *D* is High. If a_1 and a_2 are Low and the other three attributes are High, then *D* is High. So the rule can be formulated as: *if the attribute values a_1 and a_2 are Average and a_3, a_4, a_5, are Low, or if a_1 and a_2 are Low and a_3, a_4, a_5, are High, then the decision variable D is High.*

Case 2: $D =$ Average

Exhibit 7 shows some mixed cases. There are three cases—*C*4, *C*8, and *C*20—in which all the attribute values are Low but the decision variable is Average. There are two cases—*C*7 and *C*19—in which a_1 *and* a_2 *are* Average and a_3, a_4, *and* a_5 are Low, and the decision variable is Average. There are two cases—*C*9 and *C*13—in which a_1 *and* a_2 are High and a_3, a_4, *and* a_5 are Low, and the decision variable is Average. So discarding the case where all the attribute values are Low, a rule can be formulated as: *if the attribute values a_1 and a_2 are Average and a_3, a_4, and a_5 are Low, or if a_1 and a_2 are High and a_3, a_4, and a_5 are Low then the decision variable D is Average.*

Exhibit 4 Results of descriptive statistical analysis.

	Mkt (a_1)	Advt (a_2)	Dist (a_3)	Misc (a_4)	R&D (a_5)	Sales (D)
Mean	9.088	52.560	27.621	36.934	2.315	791.045
Median	2.093	1.330	1.753	3.803	0.053	110.927
Std Error	2.677	37.744	21.932	27.510	1.713	497.265
Std Deviation	12.839	181.015	105.182	131.934	8.213	2384.800
Skewness	1.765*	4.525*	4.750*	4.682*	4.434*	4.438*
Kurtosis	2.459**	21.030*	22.685*	22.203*	20.264*	20.401*

* indicates *p* value significant at 0.001. ** indicates significant at 0.05.

Exhibit 5 Correlation matrix corresponding to the data set of 23 Indian cosmetics companies.

Correlation	Mkt (a_1)	Advt (a_2)	Dist (a_3)	Misc (a_4)	R&D (a_5)	Sales (D)
Mkt (a_1)	1.000					
Advt (a_2)	0.041	1.000				
Dist (a_3)	0.001	0.991*	1.000			
Misc (a_4)	0.011	0.996*	0.998*	1.000		
R&D (a_5)	0.004	0.995*	0.982*	0.990*	1.000	
Sales (D)	0.050	0.981*	0.984*	0.986*	0.967*	1.000

* indicates *p* value significant at 0.001.

Exhibit 6 Database with categorised attribute values.

Company	Mkt (a_1)	Advt (a_2)	Dist (a_3)	Misc (a_4)	R&D (a_5)	Sales (D)
$C1$	Low	Low	Low	Low	Low	Low
$C2$	Average	Average	Low	Low	Low	High
$C3$	Low	Low	Low	Low	Low	Low
$C4$	Low	Low	Low	Low	Low	Low
$C5$	Low	Low	Low	Low	Low	Low
$C6$	Low	Low	Low	Low	Low	Low
$C7$	Average	Average	Low	Low	Low	Average
$C8$	Low	Low	Low	Low	Low	Low
$C9$	High	High	Low	Low	Low	Low
$C10$	Low	Low	High	High	High	High
$C11$	Low	Low	Low	Low	Low	Low
$C12$	Low	Low	Low	Low	Low	Low
$C13$	High	High	Low	Low	Low	Low
$C14$	Low	Low	Low	Low	Low	Low
$C15$	Average	Average	Low	Low	Low	High
$C16$	Low	Low	Low	Low	Low	Low
$C17$	Low	Low	Low	Low	Low	Low
$C18$	Low	Low	Low	Low	Low	Low
$C19$	Average	Average	Low	Low	Low	Average
$C20$	Low	Low	Low	Low	Low	Low
$C21$	Low	Low	Low	Low	Low	Low
$C22$	Low	Low	Low	Low	Low	Low
$C23$	Low	Low	Low	Low	Low	Low

Case 3: D = Low

If the values of all the attributes are Low, then the value of the decision variable is also Low. So the rule can be formulated as: *if all attribute values: a_i $\forall$ $i = 1, 2, \ldots, 5$ are Low, then the decision variable D is Low.*

The three cases can be consolidated into a table as in Exhibit 8.

The first part of the rules for Case 1 and Case 2 are the same, but they lead to different decisions. These are the boundary line cases. The conflict can be resolved by taking the decision in favour of Case 2. The new set of rules can be:

Rule 1: If $a_1$and a_2 are Low and a_3, a_4, a_5 are High, then D is High.
Rule 2: If $a_1$and a_2 are Average and a_3, a_4, a_5 are Low, then D is Average.
Rule 3: If all the attribute values are low, then the decision variable D is low.

The inexactness of a set is due to the existence of a border line region. The greater the border line region of a set, the lower would be the accuracy of the set. This may be expressed in terms of an accuracy measure: $\alpha_R(X)=\frac{\text{Card } R\_}{\text{Card } R^-}$, $X \neq \phi$. The accuracy measure $\alpha_R(X)$ is intended to capture the degree of completeness of our knowledge about the set X. $R\_$ and R^- are the R-lower and the R-upper approximations of X respectively. The R-roughness of X, which is the degree of incompleteness of the knowledge R about the set can be represented as $\rho_R(X) = 1-\alpha_R(X)$ (Pawlak, 1991). The three rules stated above take into consideration the border line cases, and make predictions with much more accuracy.

Comparison of rough set approach with alternate traditional approaches

Regression analysis

We use regression analysis in order to gain a better understanding of the relationship between the overall effects of the attributes on the decision variable. The set of attributes is taken as the independent variable, and the decision variable is taken as the dependent variable.The statistical representation of the regression equation can be written as follows:

$$D = b_0 + b_1a_1 + b_2a_2 + b_3a_3 + b_4a_4 + b_5a_5 + \hat{u} \qquad (1)$$

where b_0 = constant (the value of the dependent variable when the value of the independent variable is zero), also called the intercept as it determines where the regression line meets the y-axis; $b_1, b_2, \ldots, b_5$ = regression coefficients which represent the estimated change in the mean value of the dependent variable for each unit change in each of the five independent variables. $\hat{u}$ is treated as the error term.

Now considering the values from Exhibit 9, the regression equation will be of the form:

$$D = 6.708 + 6.924a_1 - 11.725a_2 - 54.491a_3 + 91.54a_4 - 232.49a_5 + \hat{u} \qquad (2)$$

The relationship between the decision variable (Sales) and the attributes a_3, a_4, and a_5 are statistically significant at $p < 0.05$. *Adj* $R^2 = 0.984$ shows that the relationship is statistically significant. Linear regression uses the original data as given, assumes a given linear functional form, and

Exhibit 7 Distribution of data based on value of decision variable.

Company	Mkt (a_1)	Advt (a_2)	Dist (a_3)	Misc (a_4)	R&D (a_5)	Sales (D)
Decision variable with value = high						
*C*2	Average	Average	Low	Low	Low	High
*C*10	Low	Low	High	High	High	High
*C*15	Average	Average	Low	Low	Low	High
Decision variable with value = average						
*C*4	Low	Low	Low	Low	Low	Average
*C*7	Average	Average	Low	Low	Low	Average
*C*8	Low	Low	Low	Low	Low	Average
*C*9	High	High	Low	Low	Low	Average
*C*13	High	High	Low	Low	Low	Average
*C*19	Average	Average	Low	Low	Low	Average
*C*20	Low	Low	Low	Low	Low	Average
Decision variable with value = low						
*C*1	Low	Low	Low	Low	Low	Low
*C*3	Low	Low	Low	Low	Low	Low
*C*4	Low	Low	Low	Low	Low	Low
*C*5	Low	Low	Low	Low	Low	Low
*C*6	Low	Low	Low	Low	Low	Low
*C*8	Low	Low	Low	Low	Low	Low
*C*11	Low	Low	Low	Low	Low	Low
*C*12	Low	Low	Low	Low	Low	Low
*C*14	Low	Low	Low	Low	Low	Low
*C*16	Low	Low	Low	Low	Low	Low
*C*17	Low	Low	Low	Low	Low	Low
*C*18	Low	Low	Low	Low	Low	Low
*C*21	Low	Low	Low	Low	Low	Low
*C*22	Low	Low	Low	Low	Low	Low
*C*23	Low	Low	Low	Low	Low	Low

the model explains 98.4% of the variation in the sales performance.

Discriminant analysis

Discriminant analysis can be very useful when objects have to be classified into two or more groups based on the knowledge of some set of variables related to them. This analysis is the appropriate statistical technique when the dependent variables are categorical (nominal or non metric) variables, and the independent variables are metric variables (Hair, Black, Babin, Anderson, & Tatham, 2006). The discrimination is achieved by calculating the variate's weight for each independent variable to maximise the difference between the groups. The discriminant function is given by,

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \ldots + W_nX_{nk}$$

The analysis is done using SPSS 16.0; to run the model, the decision state is coded as Low = 1, Average = 2 and High = 3, so that the dependent variable becomes categorical. The predictions using discriminant analysis are given in Exhibit 10.

Exhibit 8 Consolidated distribution of data based on value of decision variable.

	a_1	a_2	a_3	a_4	a_5	Supporting Cases
Case 1: D = High	Average	Average	Low	Low	Low	*C*2, *C*15
	Low	Low	High	High	High	*C*10
Case 2: D = Average	Low	Low	Low	Low	Low	*C*4, *C*8,*C*20
	Average	Average	Low	Low	Low	*C*7, *C*19
	High	High	Low	Low	Low	*C*9, *C*13
Case 3: D = Low	Low	Low	Low	Low	Low	*C*1, *C*3, *C*4, *C*5, *C*6, *C*8, *C*11, *C*12, *C*14, *C*16, *C*17, *C*18, *C*21, *C*22, *C*23

Exhibit 9 Results of regression analysis.

Independent variables	Coefficients	Std error coefficients	t	P	Summary
Constant	6.708	100.222	0.067	0.947	*Coefficient of determination*: $R^2 = 0.984$
Mkt (a_1)	6.924	7.075	0.979	0.341	*Adj* $R^2 = 0.984$
Advt (a_2)	−11.725	10.836	−1.082	0.294	
Dist (a_3)	−54.491	23.537	−2.315	0.033	
Misc (a_4)	91.540	30.592	2.992	0.008	
R&D (a_5)	−232.490	107.492	−2.163	0.045	

The discriminant model is able to classify 91.3% of the original cases correctly. Wilk's Lambda is 0.146, and $p = 0.000$. The low value of Lambda indicates high significance.

Conclusion

The paper presents an application of the rough set theory (RST) as a methodology for rule derivation, which can be useful in various marketing applications. The rule generated through the methodology can act as an *expert*, to which reference can be made for future strategic decision-making. This could be achieved through a plug-and-play software based on this methodology, where the attributes are plugged in through a simulated exercise to see 'what if' scenarios to take business decisions. We observe that the sales of a company could be high, if a high level of investment is made towards distribution, R&D and miscellaneous expenditure, and could be low, if the level of investment made towards marketing and advertising expenditure is low. If the level of investment made in all the parameters is low, then the sales level becomes low. If the level of investment made on marketing and advertising is average, then the sales level remains at an average level even if the expenditure on the other attributes is low.

The results of our statistical inferences indicate that for the Indian cosmetics industry, the distribution, R&D and miscellaneous expenditure attributes play an important role. The statistical analysis shows a low degree of insignificant correlation value of the marketing attribute with all the other attributes. Moreover, the regression coefficient of the marketing and advertising attributes is not significant. The rules derived in the paper are based on a small database, but the method can be extended to larger databases with better results.

In our analysis, we have assumed that all the attributes are of equal importance, but this need not always be the case. Some of the attributes may be more important than the others, and this needs to be taken into consideration during the analysis. This paper attempts to understand the basic concepts of RST and its possible applications. Statistical methods such as discriminant analysis and regression analysis make certain assumptions regarding the mathematical or statistical properties of the data whose quality is suspect. The regression model that we used explains 98.4% of the variation in sales, and the discriminant model classifies 91% of the objects correctly even with only a small data set. This paper demonstrates that almost similar accuracy can be achieved without making any mathematical or statistical assumptions regarding the data, even when the quality of data is suspect, with reliability only in the ranking of observations and not in the actual magnitudes.

Exhibit 10 Classification results of discriminant analysis.

	Class	Predicted group membership			Total
		1	2	3	
Original Count	1	15	0	0	15
	2	1	4	0	5
	3	0	1	2	3
Original %	1	100.0	0	0	100.0
	2	20.0	80.0	0	100.0
	3	0	33.3	66.7	100.0

References

Bazan, J., Nguyen, H. S., Nguyen, S. H., Synak, P., & Wróblewski, J. (2000). Rough set algorithms in classification problem. In L. Polkowski, S. Tsumoto, & T. Y. Lin (Eds.), *Rough set methods and applications: New developments in knowledge discovery in information systems* (pp. 49—88). Heidelberg: Physica-Verlag GmbH.

Beynon, M., Curry, B., & Morgan, P. (2001). 'Knowledge discovery in marketing: An approach through rough set theory'. *European Journal of Marketing, 35*(7/8), 915—937.

Bose, I. (2006). 'Deciding the financial health of dot-coms using rough sets'. *Information & Management, 43*(7), 835—846.

Dimitras, A. I., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). 'Business failure prediction using rough sets'. *European Journal of Operational Research, 114*(2), 263—280.

Ha, S. H., & Park, S. C. (1998). 'Application of data mining tools to hotel data mart on the intranet for database marketing'. *Expert Systems with Applications, 15*(1), 1—31.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). New Jersey: Pearson Prentice Hall.

Huang, J., Liu, C., Ou, C., Yao, Y. Y., & Zhong, N. (2003). 'Attribute Reduction of Rough Sets in Mining Market Value Functions'. In *Proceedings of the IEEE/WIC International Conference on Web Intelligence*. Washington, DC, USA: IEEE Computer Society.

Mckee, T. E. (2000). Developing a bankruptcy prediction model via rough sets theory. *International Journal of Intelligent Systems in Accounting, Finance and Management, 9*(3), 159—173.

Pawlak, Z. (1982). 'Rough sets'. *International Journal of Information and Computer Sciences, 11*(5), 341—356.

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Norwell, MA: Kluwer Academic Publishers.

Pawlak, Z., & Skowron, A. (2007). 'Rudiments of rough sets'. *Information Sciences, 177*(1), 3—27.

Predki, B., Słowiński, R., Stefanowsk, J., Susmaga, R., & Wilk, S. (1998). ROSE—Software implementation of the rough set theory. In L. Polkowski, & A. Skowron (Eds.), *Rough sets and current trends in computing, lecture notes in artificial intelligence, Vol. 1424* (pp. 605—608). Springer.

RSES 2.2 User's Guide. (2005). Warsaw University, Jan 19. http://logic.mimuw.edu.pl/~rses/RSES_doc_eng.pdf (Last accessed 14.07.09).

Shen, L., & Loh, H. T. (2004). 'Applying rough sets to market timing decisions'. *Decision Support Systems, 37*(4), 583—597.

Shyng, J., Wang, F., Tzeng, G., & Wu, K. (2007). 'Rough set theory in analyzing the attributes of combination values for the insurance market'. *Expert Systems with Applications, 32*(1), 56—64.

Sreekumar, & Panda, B. (2005). 'Business intelligence: An overview'. *Journal of IPM, 5*(2), 28—32.

Swiniarski, R. W. (2001). 'Rough sets methods in feature reduction and classification'. *International Journal of Applied Mathematics and Computer Science, 11*(3), 565—582.

Tseng, T., & Huang, C. (2007). 'Rough set-based approach to feature selection in customer relationship management'. *Omega, 35*(4), 365—383.

Wang, X., Yang, J., Teng, X., Xia, W., & Jensen, R. (2007). 'Feature selection based on rough sets and particle swarm optimization'. *Pattern Recognition Letters, 28*(4), 459—471.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.

Zhong, N., Dong, J., & Ohsugu, S. (2001). 'Using rough sets with heuristics for feature selection'. *Journal of Intelligent Information Systems, 16*(3), 199—214.