Development of a Business Logic using Simple Application of Rough Set Theory

Sujogya Mishra[1], Shakti Prasad Mohanty[2], Sateesh Kumar Pradhan[3]

[1]Research scholar, Utkal University
Bhubaneswar-751004, India
[2]Department of Mathematics, College of Engineering and Technology
Bhubaneswar-751003, India
[3]Department of Computer Science, Utkal University
Bhubaneswar-751004, India

Abstract

In today's scenario, business establishment usually fails due to improper planning and lack of business knowledge. In this paper, we develop a concept which is precise and reduces the number of data attributes to start any business. For the development of the concept we have used the concept of rough set idea for reduction of number of attributes. Initially, we take a survey of 1000 samples and consider the best 20 samples for our purpose. Using rough set theory, we consider the most essential attributes out of several attributes for a business development.

Key words: Rough Set Theory, Business related data, Granular computing, Data mining.

1. INTRODUCTION

The rising demand for business and wide use of internet for the growth of business resulted in huge data generation in manifold ways. The data so generated not only confuses the mind of the user but also it creates problem to derive the useful data for the application of the user.This has posed an obvious challenge for the researchers to develop methods to reduce the data set and to derive the relevant data for desired application. The application of rough set theory has an important role to play for knowledge discovery in data base(s).The ever growing field of knowledge discovery (KD) helps in extraction of hidden information from large database[3]. Data mining is also considered as essential tool in this knowledge discovery process which uses techniques from different disciplines ranging from machine learning, statistics information sciences, database, visualization ([4]-[12]). Further, prediction of business failure needs a systematic and scientific study. The first approach to predict business failure started in 1995 by Zopounidis([24]-[26]). The methods proposed are the "five C" methods, the "LAPP" method, and the "credit-men" method. Then, financial ratios methodology was developed for business failure prediction problem. This approach gives rise the methods for business failure prediction based on multivariate statistical analysis (Altman ([13]-[15]), Beaver[17], Courtis[18]). Frydman *et al*[19] first employed recursive portioning, while Gupta *et a*l[20] use mathematical programming as an alternative to multivariate discriminant analysis for business failure prediction problem. Other methods used were survival analysis by Luoma, Laitinenl[21] which is a tool for company failure prediction, expert systems by Messier and Hansen[22] , neural network by Altman *et al*[16], multi-factor model by Vermeulen *et al*[23] are also other methods developed for business failure prediction.
This paper presents a methodology for business success by reduction of attributes using rough set theory.

2. PRILIMINARIES

2.1 Rough set

Rough set theory as introduced by Z. Pawlak[2] is an extension of conventional set theory that support approximations in decision making.

2.1.2 Approximation Space:
An Approximation space is a pair (U , R) where U is a non-empty finite set called the universe R is an equivalence relation defined on U.

2.1.3 Information System:
An information system is a pair S = (U , A), where U is the non-empty finite set called the universe, A is the non-empty finite set of attributes

2.1.4 Decision Table:
A decision table is a special case of information systems S= (U , A= C U {d}), where d is not in C.
Attributes in C are called conditional attributes and d is a designated attribute called the decision attribute.

2.1.4 Approximations of Sets:
Let S = (U, R) be an approximation space and X be a subset of U.
The lower approximation of X by R in S is defined as

$$\underline{R}X = \{ e \in U \mid [e] \subseteq X\}$$ and

The upper approximation of X by R in S is defined as

$$\overline{R}X = \{e \in U / [e] \cap X \neq \phi\}$$

where [e] denotes the equivalence class containing e.

A subset X of U is said to be R-definable in S if and only if

$$\overline{R}X = \underline{R}X$$

A set X is rough in S if its boundary set is nonempty.

2.2 Dependency of Attributes

Let C and D be subsets of A. We say that D depends on C in a degree k ($0 \leqq k \leqq 1$) denoted by $C \rightarrow_k D$ if

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

$$k = \gamma(C,D) = \frac{|POS_C(D)|}{|U|}$$

where $POS_C(D) = \bigcup_{X \in U/D} \underline{C}(X)$ *called a positive region of the partition* U/D *with respect to C, which is the set of all elements of* U *that can be uniquely classified to blocks of the partition* U/D

If k = 1 we say that D depends totally on C.
If k < 1 we say that D depends partially (in a degree k) on C.

2.3 Dispensable and Indispensable Attributes

Let S = (U, A = C ∪ D) be a decision table.
Let c be an attribute in C.
Attribute c is dispensable in S if POSC(D)= POS(C-{c})(D) otherwise, c is indispensable.
A decision table S is independent if all attributes in C are indispensable.

Rough Set Attribute Reduction (RSAR) provides a filter based tool by which knowledge may be extracted from a domain in a concise way; retaining the information content whilst reducing the amount of knowledge involved.

2.4 Reduct and Core

Let S = (U, A=C U D) be a decision table.
A subset R of C is a reduct of C, if
POS**R**(D) = POS**C**(D) and S' = (U, RUD) is independent, ie., all attributes in R are indispensible in S'.
Core of C is the set of attributes shared by all reducts of C.
CORE(C) = ∩RED(C)
where, RED(C) is the set of all reducts of C.
The reduct is often used in the attribute selection process to eliminate redundant attributes towards decision making.

3. Basic idea

The basic idea for the proposed work is conceived from the general market systems. We initially consider 1000 samples, by considering five conditional attributes such as Location, Combination, Quality, Advertisement and Knowledge (Business Knowledge) and three decision attributes such as failure, partial success and success. Then, by correlation analysis , only 20 samples are selected which we consider as the best for the purpose. Using rough set theory concept we reduced the number of attributes which will be helpful for the business house to start the business taking care of minimum numbers of attributes and neglecting the redundant attributes which has no contribution for the growth of the business. Initially we divide entire business system, according to their types and class. To get a common result for all types of business, we consider some standard attributes as our conditional and decision attributes. which can be applied to all types of business house.

4.Data Reduction

As the volume of data is increasing day by day, it is very difficult to find which attributes are important for a particular application and which are not that important and can be neglected. The aim of data reduction is to find the relevant attributes that have all essential information of the data set. The process is illustrated through the following 20 samples by using the rough set theory.

Table-1: Decision Table

	LOC	COMB	QUA	ADVT	KNOW	DEC
E_1	GOOD	AVE	AVE	AVE	GOOD	PSUCC
E_2	BAD	BAD	BAD	AVE	GOOD	FAIL
E_3	BAD	BAD	BAD	HIGH	GOOD	FAIL
E_4	BAD	GOOD	AVE	HIGH	GOOD	PSUCC
E_5	BAD	BAD	AVE	HIGH	GOOD	PSUCC
E_6	GOOD	BAD	AVE	HIGH	GOOD	PSUCC
E_7	GOOD	GOOD	AVE	HIGH	GOOD	SUCC
E_8	PGOOD	GOOD	AVE	AVE	BAD	PSUCC
E_9	PGOOD	BAD	AVE	POOR	BAD	FAIL
E_{10}	PGOOD	AVE	AVE	AVE	BAD	PSUCC
E_{11}	PGOOD	AVE	AVE	AVE	BAD	SUCC
E_{12}	PGOOD	BAD	GOOD	POOR	BAD	PSUCC
E_{13}	PGOOD	AVE	BAD	POOR	BAD	FAIL
E_{14}	PGOOD	BAD	BAD	POOR	BAD	SUCC
E_{15}	PGOOD	AVE	AVE	AVE	BAD	FAIL
E_{16}	PGOOD	AVE	AVE	POOR	BAD	PSUCC
E_{17}	PGOOD	AVE	BAD	POOR	BAD	FAIL
E_{18}	PGOOD	BAD	GOOD	POOR	BAD	SUCC
E_{19}	PGOOD	AVE	BAD	AVE	BAD	FAIL
E_{20}	PGOOD	BAD	BAD	AVE	BAD	SUCC

(LOC- Location, COMB- Combination, QUA- Quality, ADVT- Advertisement, KNOW- Knowledge, DEC- Decision. PSUCC-partial success, SUCC- success, FAIL-failure, AVE- average, PGOOD-Partially good)

We find the equivalence class as (Attributes)\(Decision). We consider three basic attributes such as Success, Partial Success and Failure . The equivalence class corresponding to SUCCESS = $\{E_7,E_{11},E_{14},E_{18},E_{20}\}$, the equivalence class corresponding to PSUCESS = $\{E_1,E_4,E_5,E_8,E_{10},E_{12},E_{16}\}$, the equivalence class corresponding to FAILURE =$\{E_2, E_3, E_9, E_{13},E_{15},E_{17},E_{19}\}$.In the process we find the lower approximation and lower approximation. Lower approximation leads to success and upper approximation consideration of both success and partial success as the basic attributes. E\(CONDITON) is the equivalence classes to start with. Initially, we consider single attribute as decision (Table-2), next we consider two attributes (Table-3) and finally all three attributes are considered for decision making(Table-4).
From the decision table (Table-1), we find
E(LOCATION)$_{GOOD}$ = $\{E_1,E_6,E_7\}$
E(LOCATION)$_{BAD}$ = $\{E_2,E_3,E_4,E_5\}$
E(LOCATION)$_{PGOOD}$ = $\{E_8,E_9,E_{10},E_{11},E_{12},E_{13},E_{14},E_{15},E_{16},E_{17}, E_{18},E_{20}\}$
E(COMBINATION)$_{AVE}$=$\{E_1,E_{10},E_{13},E_{15},E_{16},E_{17},E_{19}\}$
E(COMBINATION)$_{BAD}$ = $\{E_2,E_3,E_5,E_6,E_9,E_{12},E_{14},E_{18},E_{20}\}$
E(QUALITY)$_{AVE}$=$\{E_2,E_3,E_5,E_6,E_9,E_{12},E_{14},E_{18},E_{20}\}$
E(QUALITY)$_{AVE}$=$\{E_1,E_4,E_5,E_6,E_7,E_8,E_9,E_{10},E_{11},E_{15},E_{16}\}$
E(QUALITY)$_{BAD}$=$\{E_2,E_3,E_{13},E_{14},E_{17},E_{19},E_{20}\}$
E(QUALITY)$_{GOOD}$=$\{E_{12},E_{18}\}$
E(ADVERTISEMENT)$_{HIGH}$=$\{E_3,E_4,E_5,E_6,E_7\}$
E(ADVERTISEMENT)$_{AVE}$=$\{E_1,E_2,E_{10},E_{11},E_{15},E_{19},E_{20}\}$
E(ADVERTISEMENT)$_{POOR}$=$\{E_9,E_{12},E_{13},E_{14},E_{17},E_{18}\}$
E(KNOWEDGE)$_{GOOD}$=$\{E_1,E_2,E_3,E_4,E_5,E_6,E_7\}$
E(KNOWEDGE)$_{BAD}$=$\{E_8,E_9,E_{10},E_{11},E_{12},E_{13},E_{14},E_{15}, E_{16}, E_{17}, E_{18},E_{19},E_{20}\}$

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Next, we find the combination of two attributes each to generate the reduct.
E(location, combination)$_{good}$={E_7},
E(location, combination)$_{ave}$ = null,
E(location, combination)$_{bad}$=null
E(combination, Quality)$_{good}$=null
E(combination, Quality)$_{ave}$ ={E_1,E_{10},E_{15}},
E(combination, Quality)$_{BAD}$={E_2,E_3,E_{14}}
E(Location, Quality)$_{good}$={E_6},
E(Location, Quality)$_{ave}$=null, ,
E(Location, Quality)$_{bad}$={E_2,E_7}
E(Location, advertisement)$_{good}$={E_6,E_7}
E(Location, advertisement)$_{ave}$=null,
E(Location ,advertisement)$_{bad}$=null,
E(Location ,knowledge)$_{good}$={E_1,E_7},

As location combination provides the same attribute ie E_7 so we drop combination attribute from the table. In similar way we can safely remove the advertisement from the table as this attribute as it provides null result when combined with location, in both average and bad cases.
Both advertisement and location give rise the same conclusion. Now, attribute list falls to three from five which are location, quality and business knowledge which is depicted in Table-2.

Table-2: Reduced Decision Table-1

E	LOCATION	QUALITY	KNOWEDGE	DECESION
E_1	GOOD	AVERAGE	GOOD	PSUCC
E_2	BAD	BAD	GOOD	PSUCC
E_3	BAD	BAD	GOOD	PSUCC
E_4	BAD	AVE	GOOD	PSUCC
E_5	BAD	AVE	GOOD	PSUCC
E_6	GOOD	AVE	GOOD	PSUCC
E_7	GOOD	AVE	GOOD	SUCC
E_8	PGOOD	AVE	BAD	PSUCC
E_9	PGOOD	AVE	BAD	FAIL
E_{10}	PGOOD	AVE	BAD	PSUCC
E_{11}	PGOOD	AVE	BAD	SUCC
E_{12}	PGOOD	GOOD	BAD	PSUCC
E_{13}	PGOOD	BAD	BAD	FAIL
E_{14}	PGOOD	BAD	BAD	SUCC
E_{15}	PGOOD	AVE	BAD	FAIL
E_{16}	PGOOD	AVE	BAD	PSUCC
E_{17}	PGOOD	BAD	BAD	FAIL
E_{18}	PGOOD	GOOD	BAD	SUCC
E_{19}	PGOOD	BAD	BAD	FAIL
E_{20}	PGOOD	BAD	BAD	SUCC

From the table {E_1,E_6,E_7} ,{E_{13},E_{14},E_{19},E_{20},},{E_4,E_5} has same attribute values so for that
{E_1,E_6,E_7}→ E_1, ,{E_{13},E_{14},E_{19},E_{20},}→ E_{13}, {E_4,E_5}→ E_4, now the table reduces to the following form (Table-3)

Table-3: Reduced Decision Table-2

E	LOCATION	QUALITY	KNOWEDGE	DECESION
E_1	GOOD	AVE	GOOD	PSUCC
E_2	BAD	BAD	GOOD	FAIL
E_3	BAD	BAD	GOOD	FAIL
E_4	BAD	AVE	GOOD	PSUCC
E_5	BAD	AVE	GOOD	PSUCC
E_8	PGOOD	AVE	BAD	PSUCC
E_9	PGOOD	AVE	BAD	FAIL
E_{10}	PGOOD	AVE	BAD	PSUCC
E_{11}	PGOOD	AVE	BAD	SUCC
E_{12}	PGOOD	GOOD	BAD	PSUCC
E_{13}	PGOOD	BAD	BAD	FAIL
E_{15}	PGOOD	AVE	BAD	FAIL
E_{16}	PGOOD	AVE	BAD	PSUCC
E_{18}	PGOOD	GOOD	BAD	SUCC

Similarly E_{10},E_{11} ambiguous so for that we remove both from the table now the new table appears to be {E_1,E_6,E_7}→ E_1, ,{E_{13},E_{14},E_{19},E_{20},}→ E_{13}, {E_4,E_5}→ E_4, now the table reduces further as (Table-4)

Table-4: Reduced Decision Table-3

E	LOCATION	QUALITY	KNOWEDGE	DECESION
E_1	GOOD	AVE	GOOD	PSUCC
E_2	BAD	BAD	GOOD	FAIL
E_3	BAD	BAD	GOOD	FAIL
E_4	BAD	AVE	GOOD	PSUCC
E_5	BAD	AVE	GOOD	PSUCC
E_8	PGOOD	AVE	BAD	PSUCC
E_9	PGOOD	AVE	BAD	FAIL
E_{12}	PGOOD	GOOD	BAD	PSUCC
E_{13}	PGOOD	BAD	BAD	FAIL
E_{15}	PGOOD	AVE	BAD	FAIL
E_{16}	PGOOD	AVE	BAD	PSUCC
E_{18}	PGOOD	GOOD	BAD	SUCC

From the above table, we get the conclusion that the lower approximation is the set {E_1,E_6,E_{11}} and upper approximation in this case is {E_1,E_4,E_6,E_{11},E_8,E_{10},E_{12},E_{18}}
Quality of approximation is 3/8, and the boundary region in this case upper approx\ lower approx = {E_4,E_8,E_{10},E_{12},E_{18}} which cannot be classified further so we remove these entities from the table safely so the new table reuces further as (Table-5)

Table-5: Final Reduce Decision Table

LOCATION	QUALITY	KNOWEDGE	DECESION
GOOD	AVE	GOOD	PSUCC
BAD	BAD	GOOD	FAIL
BAD	BAD	GOOD	FAIL
BAD	AVE	GOOD	PSUCC
PGOOD	AVE	BAD	FAIL
PGOOD	BAD	BAD	FAIL
PGOOD	AVE	BAD	FAIL
PGOOD	AVE	BAD	PSUCC

From the above table we draw the inference that, to make the business successful the essential attributes required are Location, Quality and Knowledge. The other attributes are not of much use.

5. Conclusion

So we conclude with this idea that to become successful entrepreneur, essential attributes are Location, Quality, and Business knowledge.
This work can be extended further by prioritizing the essential attributes. Besides correlation analysis, other methods are to be developed for selection of few samples from the large collection of samples. The procedure can be used to generate decision based on the student feedback.

REFERENCES

[1] S.K. Pal, A. Skowron, Rough Fuzzy Hybridization: A new trend in decision making, Berlin, Springer-Verlag, 1999

[2] Z. Pawlak, “Rough sets”, International Journal of Computer and Computer and Information Sciences, Vol. 11, 1982, pp.341–356

[3] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving, Vol. 9, The Netherlands, Kluwer - Academic Publishers, Dordrecht, 1991

4] Han, Jiawei, Kamber, Micheline, Data Mining: Concepts and Techniques. San Franciso CA, USA, Morgan Kaufmann Publishers, 2001

[5] Ramakrishnan, Naren and Grama, Y. Ananth, “Data Mining: From Serendipity to Science”, IEEE Computer, 1999, pp. 34-37.

[6] Williams, J. Graham, Simoff, J. Simeon, Data Mining Theory, Methodology, Techniques, and Applications (Lecture Notes in Computer Science/ Lecture Notes in Artificial Intelligence), Springer, 2006.

[7] D.J. Hand, H. Mannila, P. Smyth, Principles of Data Mining. Cambridge, MA: MIT Press, 2001

[8] D.J. Hand, G.Blunt, M.G. Kelly, N.M.Adams, “Data mining for fun and profit”, Statistical Science, Vol.15, 2000, pp.111-131.

[9] C. Glymour, D. Madigan, D. Pregibon, P.Smyth, “Statistical inference and data mining”, Communications of the ACM, Vol. 39, No.11,1996, pp.35-41.

[10] T.Hastie, R.Tibshirani, J.H. Friedman, Elements of statistical learning: data mining, inference and prediction, New York: Springer Verlag, 2001

[11] H.Lee, H. Ong, “Visualization support for data Mining”, IEEE Expert, Vol. 11, No. 5, 1996, pp. 69-75.

[12] H. Lu, R. Setiono, H. Liu,“Effective data Mining using neural networks”, IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, 1996, pp. 957-961.

[13] E.I Altman, “Financial ratios, discriminants analysis and prediction of corporate bankruptcy”, The journal of finance, Vol. 23 , 1968, pp.589-609

[14] E.I.Altman, R.Avery, R.Eisenbeis, J. Stnkey, “Application of classification techniques in business, banking and finance. Contemporary studies in Economic and Financial Analysis”, vol.3, Greenwich, JAI Press,1981.

[15] E.I Altman, “The success of business failure prediction models: An international surveys”, Journal of Banking and Finance Vol. 8, no.2, 1984, pp.171-198

[16] E.I Altman, G. Marco, F. Varetto, “Corporate distress diagnosis: Comparison using discriminant analysis and neural networks”, Journal of Banking and Finance, Vol. 18, 1994, pp. 505-529

[17] W.H Beaver, “Financial ratios as predictors of failure. Empirical Research in accounting : Selected studies”, Journal of Accounting Research Supplement to Vol 4, 1966, pp.71-111

[18] J.K Courtis, “Modelling a financial ratios categoric frame Work”, Journal of Business Finance and Accounting, Vol. 5, No.4, 1978, pp71-111

[19] H.Frydman, E.I Altman ,D-lKao, “Introducing recursive partitioning for financial classification: the case of financial distress”, The Journal of Finance, Vol.40, No. 1, 1985, pp. 269-291.

[20] Y.P.Gupta, R.P.Rao, P.K. , Linear Goal programming as an alternative to multivariate discriminant analysis a note journal of business fiancé and accounting Vol.17, No.4, 1990, pp. 593-598

[21] M. Louma, E, K. Laitinen, “Survival analysis as a tool for company failure prediction”. Omega, Vol.19, No.6, 1991, pp. 673-678

[22] W.F. Messier, J.V. Hanseen, “Including rules for expert system development: an example using default and bankruptcy data”, Management Science, Vol. 34, No.12, 1988, pp.1403-1415

[23] E.M. Vermeulen, J. Spronk, N. Van der Wijst., The application of Multifactor Model in the analysis of corporate failure. In: Zopounidis,C.(Ed), Operational corporate Tools in the Management of financial Risks, Kluwer Academic Publishers, Dordrecht, 1998, pp. 59-73

[24] C. Zopounidis, A.I. Dimitras, L. Le Rudulier, A multicriteria approach for the analysis and prediction of business failure in Greece. Cahier du LAMSADE, No. 132, Universite de Paris Dauphine, 1995.

[25] C. Zopounidis, N.F. Matsatsinis, M. Doumpos, “Developing a multicriteria knowledge-based decision support system for the assessment of corporate performance and viability: The FINEVA system, “Fuzzy Economic Review, Vol. 1, No. 2, 1996, pp. 35-53.

[26] C. Zopounidis, M. Doumpos, N.F. Matsatsinis, “Application of the FINEVA multicriteria knowledge-decision support systems to the assessment of corporate failure risk”, Foundations of Computing and Decision Sciences, Vol. 21, No. 4, 1996, pp. 233-251