

Rule Detection in Mobile Ad Hoc Network using Rough Set and Probability

P. Seethalakshmi
Department of Mathematics,
Kongu Polytechnic College,
Perundurai, 638060, Tamilnadu, India

S. Senthilkumar
Department of Basic Sciences,
Kongu Polytechnic College,
Perundurai, 638060, Tamilnadu, India

ABSTRACT

This paper proposes an innovative approach for rule detection from a decision table. The aim is to apply rough set concepts and probabilistic properties to search for rule discovery. Rough set theory is generally a comparatively new intelligent technique used within the invention of data of knowledge dependencies; it evaluates the importance of attributes, discovers the patterns of information, reduces all redundant objects and attributes, and seeks the minimum subset of attributes. Moreover, it is getting used for the extraction of rules from databases. With every decision rule in decision table, two conditional probabilities, the certainty and the coverage coefficient of the rule are associated. The Probabilistic approach is an extension of the Rough set approach that reveals some probabilistic structure of the data being analyzed. Finally, these techniques will be applied for finding rules in mobile ad hoc network for the selection of best routing path with minimum number of resources.

Keywords

Rough Set; Entropy; Information gain; Certainty and Coverage; Decision rule.

1. INTRODUCTION

The rough set theory is adopted as an alternative technique for the extraction of decision rules from data sets. It is a mathematical approach to manage vague and uncertain data or problems related to information systems, indiscernibility relations and classification, attribute dependence and approximation accuracy, reduct and core attribute sets, and decision rules. In essence, it identifies dependencies in attributes, reduces the dataset by removing the weakest attributes from the information table, leading to sequential building of rules in 'if then' form. An associated decision rules depict the lower approximation of outcome in term of conditions, while uncertain decision rules refer to the upper approximation of outcome.

Two conditional probabilities, called the certainty and the coverage coefficient are associated with every decision rule. The certainty coefficient expresses the conditional probability that an object belongs to the decision class specified by the decision rule, given that it satisfies the conditions of the rule. The coverage coefficient gives the conditional probability with the reasons for a given decision. The aim of this paper is to employ statistical methods which are compatible with the rough set philosophy to rule discovery. And experimental results provide to indicate the relevancy of the planned technique.

2. LITERATURE REVIEW

Rough set theory has been introduced as a mathematical approach to deal with vagueness and uncertainty in data

analysis [1]. This theory deals with illustration, learning and generalization of unsure information [2,3,4,5] and its distillation is to diminish information and turn out crisp decision rules with none previous data on the far side data set to be dealt with. The rough set approach has several advantages over the conventional methods [6,7]. The tool is based on the original data, and does not need any external information. It is a tool suitable for analyzing quantitative as well as qualitative attributes. The ideas of lower and upper approximation employed in rough set theory, knowledge hidden in information could also be unraveled and expressed within the form of decision rules [8,9].

Although main stream research in rough set theory has been dominated by algebraic and non-probabilistic studies; probabilistic approaches have been applied to the theory since its inception [10]. In recent years, researchers, driven by a need to represent information qualitatively, have planned several models to include probabilistic approaches into rough set theory. The proposals include probabilistic rough set models [11], variable precision rough set models, rough membership functions and Bayesian rough set models [12,13,14,15]. How to develop measures to automatically extract and to evaluate interesting, relevant, and novel rules becomes an insistent and practical topic in this area. All these factors have opened up the scope for some of the most recent techniques that have been developed in recent years.

3. ROUGH SET THEORY REVIEW

The rough set philosophy is predicated on the idea that with each object of the universe there's associated a particular quantity of information (data, knowledge), expressed by means of some attributes used for object description [16]. Information system, Lower and Upper approximation and Information gain were the main concepts used to perform the task of rough set theory on the specified data.

3.1 Information system

An information system can be viewed as a table of data, consisting of objects (rows) and attributes (columns). An information system is also extended by the insertion of decision attributes. Such a system is termed as a decision system. Consider two finite and non empty sets U and A where U is called the universe and A , a set of attributes. With attribute $a \in A$, we associate a set V_a (value set) called the domain of a .

Any subset B of A determines a binary relation $INDS(B)$ on U which will be called an indiscernibility relation [17]:

$$INDS(B) = \{(x, y) \in U / \forall a \in B, a(x) = a(y)\}, \quad (1)$$

where $INDS(B)$ is equivalence relation and is termed B-indiscernibility relation.

3.2 Lower and Upper Approximation

Suppose $B \subseteq A$ and $X \subseteq U$. X can approximate by using only the information contained in B by constructing lower approximation (2) and upper approximation (3) of x in the following way:

$$\underline{B}(x) = \{ x \in U: B(x) \subseteq x \} \quad (2) \text{ and}$$

$$\overline{B}(x) = \{ x \in U: B(x) \cap X \neq \emptyset \} \quad (3)$$

Equivalence categories contained among X belongs to the lower approximation whereas equivalence categories among X and on its border outline the upper approximation. Let P and Q be sets of attributes together with equivalence relation over U . Then the positive region is defined as

$$POS_P(Q) = \bigcup_{x \in U/Q} PX \quad (4)$$

where $POS_P(Q)$ comprises all objects of U that can be classified to classes U/Q using the information contained within attributes P .

3.3 Information Gain

Attribute selection in ID3 and C4.5 algorithms are based on minimizing an information entropy measure applied to the examples at a node [18]. Entropy has widely applied to many fields. The entropy measure is employed to decide on the attributes providing the very best information gain. Quinlan's ID 3 decision tree algorithm grasps the entropy concept for attribute selection [19]. A data set with some discrete valued condition attributes and one discrete valued decision attribute can be presented in the form of knowledge representation system $J=(U, CU D)$, where $U=\{u_1, u_2, \dots, u_s\}$ is the set of data sample, $C=\{c_1, c_2, \dots, c_n\}$ is the set of condition attributes and $D=\{d\}$ is the one-elemental set with the decision attribute or class label attribute. Suppose this class label attribute has m distinct values defining m distinct classes, d_i (for $i=1, 2, \dots, m$), and let s_i be the number of samples of U in class d_i . The entropy for a subset is given by

$$\text{Entropy}(S) = - \sum_{i=1}^m p_i \log_2 p_i \quad (5)$$

where p_i is the probability that an object is in class i . $G(S, A)$, an information gain of example set S on attribute A is defined as,

$$G(S, A) = \text{Entropy}(S) - \sum_v (|S_v| / |S|) * \text{Entropy}(S_v) \quad (6)$$

where Σ is every significance value v of all probable values of attribute A , S_v is subset of S for which attribute A has value v , $|S_v|$ indicates the number of elements in S_v and $|S|$ represents the number of elements in S .

4. PROBABILISTIC REVIEW

Probabilistic properties of decision tables were studied and well-established of churn modeling in telecommunications [20]. A probabilistic approach for reducing dimensions and extracting rules of information systems using expert systems is also proposed [21].

4.1 Decision tables and decision rule

Let $S = (U, C, D)$ where C and D are disjoint sets of condition attributes and decision attributes respectively. Every $x \in U$

determines a sequence $c_1(x), c_2(x), \dots, c_n(x)$; $d_1(x), d_2(x), \dots, d_m(x)$, where $\{c_1, c_2, \dots, c_n\} = C$ and $\{d_1, d_2, \dots, d_m\} = D$. If $c_1(x), c_2(x), \dots, c_n(x) \rightarrow d_1(x), d_2(x), \dots, d_m(x)$, then the sequence is termed as decision rule induced by x (in S) or in short $C \rightarrow_x D$.

The number $\text{supp}_x(C, D) = |A(x)| = |C(x) \cap D(x)|$ is called as a support of the decision rule $C \rightarrow_x D$ and the number $\sigma_x(C, D) = \frac{\text{supp}_x(C, D)}{|U|}$, is referred to as the strength of the decision rule $C \rightarrow_x D$ wherever $|X|$ denotes the cardinality of X . With every decision rule, denoted $C \rightarrow_x D$ there associates the certainty factor of the decision rule, denoted $\text{cer}_x(C, D)$ and defined as follows:

$$\text{cer}_x(C, D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{\text{supp}_x(C, D)}{|C(x)|} = \frac{\sigma_x(C, D)}{\pi(C(x))}, \text{ where}$$

$$\pi(C(x)) = \frac{|C(x)|}{|U|}.$$

The certainty factor may be interpreted as a conditional probability that y belongs to $D(x)$ given y belongs to $C(x)$, symbolically $\pi_x(D/C)$.

If $\text{cer}_x(C, D) = 1$, then $C \rightarrow_x D$ is referred as a certain decision rule whereas if $0 < \text{cer}_x(C, D) < 1$, then decision rule will be referred as an uncertain decision rule.

In addition, coverage factor of the decision rule, denoted by $\text{cov}_x(C, D)$ is defined as

$$\text{cov}_x(C, D) = \frac{|C(x) \cap D(x)|}{|D(x)|} = \frac{\text{supp}_x(C, D)}{|D(x)|} = \frac{\sigma_x(C, D)}{\pi(D(x))}, \text{ where}$$

$$\pi(D(x)) = \frac{|D(x)|}{|U|}.$$

Similarly

$$\text{cov}_x(C, D) = \pi_x(C/D).$$

If $C \rightarrow_x D$ is a decision rule, then $D \rightarrow_x C$ is called as an inverse decision rule. The inverse decision rules are used to give explanations (reasons) for a decision.

4.2 Probabilistic properties of decision rule

Decision tables that include significant probabilistic properties are given as follows:

Let $C \rightarrow_x D$ be a decision rule, and then the following properties are valid:

$$\sum_{y \in C(x)} \text{cer}_x(C, D) = 1 \quad (7)$$

$$\sum_{y \in D(x)} \text{cov}_x(C, D) = 1 \quad (8)$$

$$\begin{aligned} \pi(D(x)) &= \sum_{y \in D(x)} \text{cery}(C, D) \cdot \pi(C(y)) \\ &= \sum_{y \in D(x)} \sigma_y(C, D). \end{aligned} \quad (9)$$

$$\begin{aligned} \pi(C(x)) &= \sum_{y \in C(x)} \text{cov}_y(C, D) \cdot \pi(D(y)) \\ &= \sum_{y \in C(x)} \sigma_y(C, D). \end{aligned} \quad (10)$$

$$\text{cer}_x(C, D) = \frac{\text{cov}_x(C, D) \cdot \pi(D(x))}{\sum_{y \in C(x)} \text{cov}_y(C, D) \cdot \pi(D(y))} = \frac{\sigma_x(C, D)}{\pi(C(x))} \quad (11)$$

$$\text{cov}_x(C, D) = \frac{\text{cer}_x(C, D) \cdot \pi(C(x))}{\sum_{y \in D(x)} \text{cery}(C, D) \cdot \pi(C(y))} = \frac{\sigma_x(C, D)}{\pi(D(x))} \quad (12)$$

Every decision table satisfies all the above properties and it is observed that property (9) and property (10) refers the well known total probability theorem, whereas property (11) and property (12) refers the Bayes' theorem. Thus in order to compute the certainty and coverage factors of decision rules according to property (11) and property (12) it is enough to know the strength (support) of all decision rules. The strength of decision rules can be computed from data or can be a

subjective assessment.

5. METHODOLOGY AND MODELING

An alternative approach based on probabilistic properties in rough set methodology is described in this work. Through the combination of Rough set theory and Probability, the effectiveness of decision rule extraction can be improved for evaluating the best routing path. The selection is based on Mobile ad hoc network in consideration of six resource constraints such as bandwidth, computer efficiency, power consumption, traffic load, number of internodes and total vector cost [22].

The approach proceeds with four stages: first, information gain is utilized for distinguishing importance among the attributes. Second, a decision table can be reduced by removing redundant attributes without any information loss. In the third stage, decision rules can be extracted from the equivalence classes. Finally, with every decision rule in a decision table, three coefficients specifically the strength, the certainty and the coverage factors of the rule are associated. This gives a new look into the interpretation of Bayes' theorem and offers a new technique employed in statistical reasoning. The overall processes are illustrated by a simple example of path selection in Mobile ad hoc network.

6. ILLUSTRATIVE EXAMPLE

A data set of resources allotted to five paths is given to select efficient path as in Table 1. Consider five condition attributes: Band width, Computer efficiency, Power consumption, Traffic load and Number of internodes and a decision attribute: Total vector cost that represents the minimum cost for the selection of best path.

6.1 Information Gain

The data is supposed to classify using rough set theory to analyze the information. Therefore, each conditional attribute is provided with three classes low, medium and high conditions as L, M and H respectively. The arrangement of all attributes has been undertaken to define the specified level and assigning a system to each specified attribute as in Table 2.

ID-3 uses an information theoretic approach aimed at minimizing the expected number of tests to classify an object. Using (5) and (6), each attributes information gain with G (Band width) = 0.24, G (Computer efficiency) = 0.42, G (Power consumption) = 0.44, G (Traffic load) = 0.94 and G (Number of internodes) = 0.54 is calculated. Since, power consumption, traffic load and number of internodes have the highest information gain among the five attributes, band width and computer efficiency may be excluded due to their less importance. The data set is shown in Table 3.

The decision attribute (Total vector cost) have two values, Good and Poor. Each value may be classified into its partition. For example path 2, 3 and 5 belong to partition X_G and path 1 and 4 belong to partition X_P . $X_G = \{2, 3, 5\}$, $X_P = \{1, 4\}$.

The information in table 3 can also be represented as [(Power cons, High)]= {P1,P2}; [(Power cons, Medium)]= {P3,P4,P5}; [(Traffic load, Low)]= {P2,P5}; [(Traffic load, Medium)]= {P3,P4}; [(Traffic load, High)]= {P1}, and so on. The problem cannot be solved uniquely because the data set is inconsistent.

Let us observe that for the decision variable D: (Satisfaction, Good), the lower approximation of the set is {P2, P5}, the upper approximation is {P2, P3, P4, and P5}, and the

boundary line case is P3 and P4. Hence P1 is not a satisfied vector cost, and P3 and P4 cannot be excluded from the set of non-satisfied vector cost. Similarly for the decision variable D: (satisfaction, Poor), the lower approximation of the set is {P1}, the upper approximation is {P1, P3, P4}, and the boundary line cases are P3 and P4. Hence, path P2 and P5 are considered as the best routing path.

Besides the following decision rules gives a clear in sight in the decision structure imposed by the decision table.

Table 1. Data set of Resources allotted to five paths

Path	Band width	Computer efficiency	Power Consumption	Traffic load	No. of Inter nodes	Total Vector cost
1	0.85	0.888	0.28	1	0	3.018
2	0.571	0.777	0.42	0.25	0.5	2.518
3	0.42	0.444	0.21	0.5	1	2.574
4	0.28	1	0	0.75	0.77	2.8
5	1	0.222	0.14	0	0.166	1.528

Table 2. Data set with categorized attribute values

Path	Band width	Computer efficiency	Power Consumption	Traffic load	No. of Inter nodes	Total Vector cost
1	H	H	H	H	L	3.018
2	H	H	H	L	M	2.518
3	L	L	M	M	H	2.574
4	M	H	M	M	H	2.8
5	H	L	M	L	M	1.528

Table 3. Data set after the calculation of information gain

Path	Condition Attribute			Decision Attribute	
	Power Consumption	Traffic load	No. of Inter nodes	Satisfaction	Total Vector cost
P1	H	H	L	Poor	3.018
P2	H	L	M	Good	2.518
P3	M	M	H	Good	2.574
P4	M	M	H	Poor	2.8
P5	M	L	M	Good	1.528

Table 4. Parameters of the decision rule

Decision rule	Strength	Certainty factor	Coverage
1	0.24	1.00	0.52
2	0.20	1.00	0.38
3	0.21	0.48	0.39
4	0.23	0.52	0.48
5	0.12	1.00	0.23

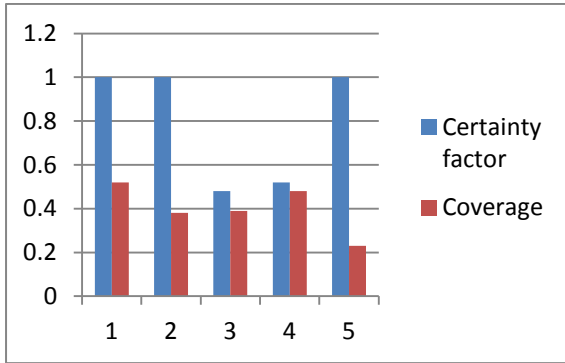


Fig.1: Comparison of certainty factor and coverage

6.2 Decision Rules

It is observed that in execution with property (11) and property (12), the certainty and coverage factors may be computed using solely the strength of decision rules. Strength, certainty and coverage factors for the information table are portrayed in table 4 and figure 1.

It is observed from table 4 and figure 1 that, decision rules 1, 2 and 5 are certain, despite the fact that the remaining decision rules are uncertain. And the decision table induces a set of “if ... then” decision rules. Hence, decision rules are extracted as follows:

Certainty

- If Power cons=high, Traf load= high,
No. of int = low, then Total vec cost = Poor. 1.00
- If Power cons and Traf load= medium,
No. of int = high, then Total vec cost = Poor. 0.52
- If Power cons and Traf load= low,
No. of int = medium, then Total vec cost = Good. 1.00
- If Power cons and Traf load= medium,
No. of int = high, then Total vec cost = Good. 0.48
- If Power cons and No. of int = medium,
Traf load = low, then Total vec cost = Good. 1.00

Finding a minimal decision algorithmic rule associated to a given decision table is quite advanced. It is essential to associate interest in explanation of decisions in terms of conditions, and hence inverse decision algorithmic rule has been applied that is obtained by replacing mutual conditions and choices in each decision rule with coverage factor in the decision algorithm. Hence, the subsequent inverse decision algorithm may be implicit as explanation of good or poor in

terms of vector cost:

Coverage

- If Total vec cost = Poor, then Power cons=high,
Traf load= high, No. of int = low. 0.52
- If Total vec cost = Poor, then Power cons and
Traf load = medium, No. of int = high. 0.48
- If Total vec cost = Good, then Power cons and
Traf load = low, No. of int = medium. 0.38
- If Total vec cost = Good, then Power cons and
Traf load = medium, No. of int = high. 0.39
- If Total vec cost = Good, then Power cons and
No. of int = medium, Traf load= low. 0.23

Observe that certainty factor for inverse decision rules are coverage factors for the actual decision rules. The above property of decision table provides an easy methodology of drawing conclusions from the information and providing explanation of obtained results.

From the decision algorithm and the certainty factors, decision rules were framed as follows:

- Poor total vector cost is implied with certainty by:
–High power consumption and traffic load and low number of internodes (path P1)
- Good satisfaction is implied with certainty by:
–High power consumption, low traffic load and medium number of internodes (path P2)
–Medium power consumption and number of internodes and low traffic load (path P5)

Hence, total vector cost with medium power consumption and traffic load and high number of internodes within the same operator are undecided (Poor vector cost, cer. = 0.52; Good, cer. =0.48) as in path P3 and P4.

Similarly, from the inverse decision algorithm and the coverage factors, decision rules were framed as follows:

- The most probable reason for poor vector cost is high power consumption and traffic load.
- The most probable reason for good satisfaction is low and medium power consumption, traffic load and number of internodes within the same operator.

Hence, path P2 and P5 are confirmed to be the most effective routing path.

7. CONCLUSION

This paper has presented an overview of Rough set theory and Probabilistic approach along with an example. The Probabilistic approach is an extension of the Rough set approach that reveals some probabilistic structure of the data being analyzed in order to further develop its potential field of applicability. Based on this idea, the proposed probabilistic rough set approach was applied to discover the set of all possible decision rules from the information table. The satisfactory set of rules gives as a result that have the average strength of rules about twice higher than other set of rules. The experimental results show that the proposed approach provides a better performance.

8. REFERENCES

- [1] Pawlak, Z. 1982. "Rough sets". *International Journal of Computer and Information Science*, 11, 345- 356.
- [2] Huanglin, Z. 1996. *Rough Set Theory and its Applications*. Chongqing University Press, Chongqing.
- [3] Jia-Cheng, W., Hua-Ping, D., and You-Xian, S. 2003. A novel approach to computing rule confidence in rough set theory. *Proceedings of the 2nd International Conference on Machine Learning and Cybernetics*, 1523-1526.
- [4] Jianguo, T. 2002. "Reliability analysis to deal with imperfect information by rough set theory". *Control Decisions*, 17(2), 255-256.
- [5] Pawlak, Z. 1991. *Rough Set-Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht.
- [6] Zhang, J., Li, T., and Chen, H. 2014. "Composite rough sets for dynamic data mining". *An International Journal of Information Sciences*, 257, 81–100.
- [7] Huang, Y., Li, T., Luo, C., Fujita, H., and Horng, S.J. 2017. "Matrix-based dynamic updating rough fuzzy approximations for data mining". *Knowledge-Based Systems*, 119, 273–283.
- [8] Gu ShM., Gao, J., and Tian XQ. 2007. "A fuzzy measure based on variable precision rough sets". In: Cao BY (ed) *Fuzzy information and engineering (ICFIE)*, ASC, 40, 798–807.
- [9] Wang L., and Yang, G. 2009. Mining Valid Association Rules in Incomplete Information Systems. *First International Workshop on Education Technology and Computer Science*, Wuhan, 633-636.
- [10] Duntsch, I., Gediga, and Roughian, G. 2001. "Rough information analysis". *International Journal of Intelligent Systems*, 16, 121-147.
- [11] Wong, S.K.M., and Ziarko, W. 1987. "Comparison of the probabilistic approximate classification and the fuzzy set model". *Fuzzy Sets and Systems*, 21, 357-362.
- [12] Greco, S., Matarazzo, B., and Slowinski, R. Rough membership and Bayesian confirmation measures for parameterized rough sets. 2005. *LNAI 3641*, 314-324.
- [13] Greco, S., Pawlak, Z., and Slowinski, R. 2004. "Can Bayesian confirmation measures be useful for rough set decision rules". *Engineering Applications of Artificial Intelligence*, 17(4), 345-361.
- [14] Slezak, D., and Ziarko, W. 2003. "Attribute reduction in the Bayesian version of variable precision rough set model". *Electronic Notes in Theoretical Computer Science*, 82, 263-273.
- [15] Slezak, D. 2005. Rough sets and Bayes factor. *LNAI 3400*, 202-229.
- [16] Pawlak, Z., and Skowron, A. 2007. "Rudiments of rough sets". *Information Sciences*, 177(1), 3-27.
- [17] Zimmermann, H.J. 1991. *Fuzzy Set Theory and Its Applications*. Kluwer Academic Publishers.
- [18] Mehmed Kantardzic, 1993. *Data Mining, Concept, Models, Methods, and Algorithms*. Wiley Publishers.
- [19] Wang, T.C., and Lee, H.D. 2006. "Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy". *WSEAS Transactions on Information Science and Applications*, 3(8), 1547–1552.
- [20] Pawlak, Z. 2012. *Rough set theory and its applications*. *Journal of Telecommunications and Information Technology*.
- [21] Hossam A Nabwey. 2011. "A Probabilistic Rough Set Approach to Rule Discovery". *International Journal of Advanced Science and Technology*, 30(May, 2011).
- [22] NagaRaju, Ramchandran, S. 2008. "Fuzzy Cost Based Multipath Routing for Mobile Ad-Hoc Networks". *Journal of Theoretical and Applied Information Technology*, 319-326.