Analysing the Behaviour of a Telecom User using Rough Set Theory

Haridas Kataria[1] & Dr. Vikram Singh[2]

[1] Lecturer CSE, Govt. Polytechnic for Women, Sirsa.

[2] Professor, Department of Computer Science, CDLU, Sirsa

Abstract – **Understanding the user's behaviour in the telecommunication is very important towards improving the quality of information and the speed of accessing large-scale telecom data sources. Churn analysis is extremely helpful in developing a sustainable and robust strategy for customer retention in the telecom company. When the companies are aware of the percentage of customers who end their relationship with it in a given time period, they can easily come up with a detailed analysis of the causes for the churn rate. This helps in developing effective customer retention programs for the company. This paper proposes a novel technique for analysing the user's behaviour using the Rough Set Theory.**

Keywords** – **Rough Set, Decision Rules, Clusters, Churn Analysis, Prefetching.

I. INTRODUCTION

Churn analysis is the calculation of the rate of attrition in the customer base of any company. It involves identifying those consumers who are most likely to discontinue using your service or product. Churn analysis is extremely helpful in developing a sustainable and robust strategy for customer retention in a company. When aware of the percentage of customers who end their relationship with the company in a given time period it can easily come up with a detailed analysis of the causes for the churn rate. This helps in developing effective customer retention programs for the company. Churn rate typically applies to many industries chiefly among them are subscription services, such as long-distance phone service or magazines. Churn analysis helps in understanding the behaviour of customers that unsubscribe and move their business to a competitor and predicting the likelihood of this event to occur. Other uses vary from calculating employee attrition in any given company.

Churn rate is the amount of customers or subscribers who cut ties with the service or company during a given time period. These customers have "Churned". Maintaining a handle on the churn rate always help over data – driven decisions over time.

Customer churn refers to when a customer (player, subscriber, user, etc.) ceases his or her relationship with a company. Online businesses typically treat a customer as churned once a particular amount of time has elapsed since the customer's last interaction with the site or service. The full cost of customer churn includes both lost revenue and the marketing costs involved with replacing those customers with new ones. Reducing customer churn is a key business goal of every online business.

Rough set theory is a new mathematical tool for imperfect data analysis. The theory has found applications in many domains such as decision support, engineering, environment, banking, medicine and others. Rough set is an approach to aid decision making in the presence of uncertainty. It classifies imprecise, uncertain or incomplete information expressed in term of data acquired from experience. Each rough set has boundary line element. The rough sets can be considered as uncertain or imprecise. Approximations are the fundamental concepts of rough set theory. Upper and lower approximations are used to identify and utilize the context of each specific object and reveal relationships between objects. The upper approximation includes all objects that possibly belong to the concept while the lower approximation contains all objects that surely belong to the concept.

Rough set based data analysis starts from a data table called a decision table, columns of which are labelled by attributes, rows – by objects of interest and entries of the table are attribute values. Each row in a decision table induces a decision rule, which specifies decision, if some conditions are satisfies.

Therefore, it is necessary to develop techniques such as rough set theory that can discover hidden and useful relationships among data fetched from the different clusters or sessions. This paper is divided into five sections. The Section I give the introduction, Section II represents the framework used, Section III represents the Implementation of the theory, Section IV focuses on the proposed Algorithm, Section V gives the Experimental Design & Analysis, Section VI focuses on the Results & Analysis and finally Section VII concludes the work done.

II. FRAMEWORK

The churn analysis and modelling in telecommunications is a simplified method to predict the customers in any telecom network. Here, six facts concerning six client segments are presented for the process. These are also known as the condition attributes. In the table condition attributes describing client profile are: *In* – incoming calls, *Out* – outgoing calls within the same operator, *Change* – outgoing calls to other mobile operator, the decision attribute describing the consequence is *Churn* and *N* is the number of similar cases.

Each row in the table determines a decision rule. E.g., row 2 determines the following decision rule: *"if the number of incoming calls is high and the number of outgoing calls is high and the number of outgoing calls to the mobile operator is low then these is no churn"*.

According to [1]: *"One of the main problem that have to be solved by marketing departments of wireless operators is to find the way of convincing current clients that they continue to use the services. In solving this problems can help churn modeling. Churn model in telecommunications industry predicts customers who are going to leave the cur-rent operator"*.

Table 1: Client segments

Segment	*In*	*Out*	*Change*	*Churn*	*N*
1	medium	medium	low	no	200
2	high	high	low	no	100
3	low	low	low	no	300
4	low	low	high	yes	150
5	medium	medium	low	yes	220
6	medium	low	low	yes	30

In other words we want to explain churn in terms of clients profile, i.e., to describe market segments f4, 5, 6g (or f1, 2, 3g) in terms of condition attributes *In*, *Out* and *Change*.

The problem cannot be solved uniquely because the data set is *inconsistent*, i.e., segments 1 and 5 have the same profile but different consequences.
Let us observe that:

– segments 2 and 3 (4 and 6) can be classified as sets of clients who *certainly* do not churn (churn),

– segments 1, 2, 3 and 5 (1, 4, 5 and 6) can be classified as sets of clients who *possibly* do not churn (churn),

– segments 1 and 5 are *undecidable* sets of clients.

This leads us to the following notions:

– the set f2,3g (f4,6g) is the *lower approximation* of the set f1,2,3g(f4,5,6g),

– the set f1,2,3,5g (f1,4,5,6g) is *the lower approximation* of the set f1,2,3g (f4,5,6g),

– the set f1,5g is the *boundary region* of the set f1,2,3g(f4,5,6g),

which will be discussed in the next paragraph more exactly.

III. IMPLEMENTATION OF RST

In this section we will examine approximations more exactly. First we define a data set, called an information system.

An *information system* is a pair $S = (U; A)$, where U and A, are finite, nonempty sets called the *universe*, and the set of *attributes*, respectively. With every attribute a 2 A we associate a set V_a, of its *values*, called the *domain* of a. Any subset B of A determines a binary relation $I(B)$ on U , which will be called an *indiscernibility relation*, and de-fined as follows: $(x; y)$ 2 $I(B)$ if and only if $a(x) = a(y)$ for every a 2 A, where $a(x)$ denotes the value of attribute a for element x. Obviously $I(B)$ is an equivalence relation. The family of all equivalence classes of $I(B)$, i.e., a partition determined by B, will be denoted by $U = I(B)$, or simply by $U = B$; an equivalence class of $I(B)$, i.e., block of the partition $U = B$, containing x will be denoted by $B(x)$. If $(x; y)$ belongs to $I(B)$ we will say that x and y are *B-indiscernible (indiscernible with respect to B)*. Equivalence classes of the relation $I(B)$ (or blocks of the partition $U = B$) are referred to as *B-elementary sets* or *B-granules*.
Suppose we are given an information system $S = (U; A)$; X U , and B A. Let us define two operations assign-ing to every X U two sets $B(X)$ and $B(X)$, called the *B-lower* and the *B-upper approximation* of X , respectively, and defined as follows:

$$B(X) = \bigcup_{x2U} B(x) : B(x) \quad X \;;$$

$$B(X) = \bigcup_{x2U} B(x) : B(x) \setminus X \; 6= \emptyset \;:$$

Hence, the *B-lower approximation* of a set is the union of all *B-granules* that are included in the set, whereas the *B-upper approximation* of a set is the union of all *B-granules* that have a nonempty intersection with the set. The set

$$BN_{B(}X) = B(X) \;\; B(X)$$

will be referred to as the *B-boundary region* of X .
If the boundary region of X is the empty set, i.e., $BN_{B(}X) = \emptyset$, then X is *crisp (exact)* with respect to B; in the opposite case, i.e., if $BN_{B(}X) = 6\ \emptyset$; X is referred to as *rough (inexact)* with respect to B.

Thus, the set of elements is rough (inexact) if it cannot be defined in terms of the data, i.e. it has some elements that can be classified neither as member of the set nor its complement in view of the data.

IV. PROPOSED WORK

Any decision table induces a set of *"if ... then"* decision rules.
Any set of mutually, exclusive and exhaustive decision rules, that covers all facts in S and preserves the indiscernibility relation included by S will be called a decision algorithm in S.
An example of decision algorithm in the decision Table 1 is given below:

1) *if (In, high) then (Churn, no)* 1.00

2) *if (In, low) and (Change, low) then (Churn, no)* 1.00

3) *if (In, med.) and (Out, med.) then (Churn, no)* 0.48

4) *if (Change, high) then (Churn, yes)* 1.00

5) *if (In, med.) and (Out, low) then (Churn, yes)* 1.00

6) *if (In, med.) and (Out, med.) then (Churn, yes)* 0.52

Finding a minimal decision with a given decision table is rather complex. Many methods have been proposed to solve this problem, but we will not consider this problem here.

If we are interested in *explanation* of decisions in terms of conditions we need an *inverse* decision algorithm which is obtained by replacing mutually conditions and decisions in every decision rule in the decision algorithm.

For example, the following inverse decision algorithm can be understood as explanation of churn (no churn) in terms of client profile:

1’) *if (Churn, no) then (In, high) and (Out, med.)*

2’) *if (Churn, no) then (In, high)*

3’) *if (Churn, no) then (In, low) and (Change, low)*

4’) *if (Churn, yes) then (Change, yes)*

5’) *if (Churn, yes) then (In, med.) and (Out, med.)* 6’) *if (Churn, yes) then (In, med.) and (Out, low)*

Observe that certainty factor for inverse decision rules are coverage factors for the original decision rules.

V. EXPERIMENTAL DESIGN & ANALYSIS

If we distinguish in an information system two disjoint classes of attributes, called *condition* and *decision attributes*, respectively, then the system will be called a *decision table* and will be denoted by $S = (U; C; D)$, where C and D are disjoint sets of condition and decision attributes, respectively.

Let $S = (U; C; D)$ be a decision table. Every $x \in U$ determines a sequence $c_1(x), \ldots, c_n(x), d_1(x), \ldots, d_m(x)$, where $\{c_1, \ldots, c_n\} = C$ and $\{d_1, \ldots, d_m\} = D$.

The sequence will be called a *decision rule induced by x* (in S) and will be denoted by $c_1(x), \ldots, c_n(x) \rightarrow d_1(x), \ldots, d_m(x)$ or in short $C \rightarrow_x D$.

The number $supp_x(C; D) = |A(x)| = |C(x) \cap D(x)|$ will be called a *support* of the decision rule $C \rightarrow_x D$ and the number

$$\sigma_x(C; D) = \frac{supp_x(C; D)}{|U|}$$

will be referred to as the *strength* of the decision rule $C \rightarrow_x D$, where $|X|$ denotes the cardinality of X.

With every decision rule $C \rightarrow_x D$ we associate the *certainty factor* of the decision rule, denoted $cer_x(C; D)$ and defined as follows:

$$cer_x(C; D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{supp_x(C; D)}{|C(x)|} = \frac{\sigma_x(C; D)}{\pi(C(x))};$$

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 $cer_x(C; D) = 1$, then $C \rightarrow_x D$ will be called a *certain decision* rule; if $0 < cer_x(C; D) < 1$ the decision rule will be referred to as an *uncertain decision rule*.

Besides, we will also use a *coverage factor* of the decision rule, denoted $cov_x(C; D)$ and defined as

$$cov_x(C; D) = \frac{|C(x) \cap D(x)|}{|D(x)|} = \frac{supp_x(C; D)}{|D(x)|} = \frac{\sigma_x(C; D)}{\pi(D(x))};$$

$$cov_x(C; D) = \pi_x(C|D):$$

If $C \rightarrow_x D$ is a decision rule then $D \rightarrow_x C$ will be called an *inverse decision rule*. The inverse decision rules can be used to give *explanations (reasons)* for a decision.

For Table 1 we have the certainty and coverage factors are as shown in Table 2.

Table 2: Parameters of the decision rules

Decision rule	Strength	Certainty	Coverage
1	0.20	0.48	0.33
2	0.10	1.00	0.17
3	0.30	1.00	0.50
4	0.15	1.00	0.38
5	0.22	0.52	0.55
6	0.03	1.00	0.07

Let us observe that if $C \rightarrow_x D$ is a decision rule then

$$\bigcup_{y \in D(x)} \{C(y) : C(y) \subseteq D(x)\}$$

is the lower approximation of the decision class $D(x)$, by condition classes $C(y)$, whereas the set

$$\bigcup_{y \in D(x)} C(y) : C(y) \cap D(x) \neq \emptyset$$

is the upper approximation of the decision class by condition classes $C(y)$.

Approximations and decision rules are two different methods to express properties of data. Approximations suit better to express topological properties of data, whereas decision rules describe in a simple way hidden patterns in data.

VI. RESLUTS AND ANALYSIS

The above properties of decision tables (algorithms) give a simple method of drawing *conclusions* from the data and giving *explanation* of obtained results.

From the decision algorithm and the certainty factors we can draw the following conclusions.

No churn is implied with *certainty* by:

– high number of incoming calls,
– low number of incoming calls and low number of outgoing calls to other mobile operator.

Churn is implied with *certainty* by:

– high number of outgoing calls to other mobile operator,
– medium number of incoming calls and low number of outgoing calls.

Clients with medium number of incoming calls and low number of outgoing calls within the same operator are *undecided* (no churn, cer. = 0.48; churn, cer. = 0.52).

From the inverse decision algorithm and the coverage factors we get the following explanations:

– the *most probable* reason for no churn is low general activity of a client,
– the *most probable* reason for churn is medium number of incoming calls and medium number of outgoing calls within the same operator, whereas the next cluster represents five no. of sessions (fig 5.4).

VII. CONCLUSION

In this paper the basic concepts of rough set theory and its application to drawing conclusions from data are discussed. For the sake of illustration an example of churn modeling in telecommunications is presented

REFERENCES

[1]. Zhang, T., Ramakrishnan, R., & Livny, M. (1996). “BIRCH: an efficient data clustering method for very large databases”. In ACM SIGMOD, pages 103–114.
[2]. Liu, H. & Setiono, R. (1996). “Feature selection and classification – a probabilistic wrapper approach”. Proceedings of the 9th International Conference on Industrial and Engineering Applications of AI and ES, 419-424.
[3]. Lin, T.Y., Cercone, N., (1997). “Rough sets and Data mining: Analysis of Imprecise Data”. Kluwer Academic Publishers.
[4]. Shahabi, C., Zarkesh, A., & Shah, V., (1997). “Knowledge discovery from users web-page navigation”. In workshop on Research Issues in Data Engineering, England.
[5]. Mobasher, B., Cooly, R., & Srivastava, J. (1999). “Automatic personalization based on web usage mining”. TR99-010, Department of Computer Science, Depaul University.
[6]. Guha, S., Rastogi, R., & Shim, K., (1999). “ROCK: a robust clustering algorithm for categorical attributes”. In ICDE.
[7]. Shen, Q. & Chouchoulas A., (2002). “A rough-fuzzy approach for generating classification rules”. *Pattern Recognition*, 35:2425-2438.
[8]. Jensen, R., (2005). “Combining Rough and Fuzzy Sets for Feature Selection”, Ph. D thesis, School of Informatics, University of Edinburgh.
[9]. Premalatha, K., & Natarajan, A. M. (2010). “A Literature Review on Document Clustering” Information Technology Journal, 9 : 993 – 1002.
[10]. Jones,P., Pearce, C., & Salgueiro, G., (2013). “End-to-End Session Identification in IP-Based Multimedia Communication Networks”, Cisco Systems.