ISSN 2250-0987
Anirban Mitra, UNIASCIT, Vol 2 (2), 2012, 282-285

CLUSTERING ANALYSIS IN SOCIAL NETWORK USING ROUGH SET AND SOFT SET

Anirban Mitra

Dept. Of CSE, M.I.T.S. Rayagada Odisha, INDIA 765017.

Abstract: **Apart from the visual interface for a site that handles the social network data, there are lot many features, which are complex in nature, makes the interface to work smoothly and efficiency. Such characteristics, mainly works from behind of the interface to decrease the accessing time of users, increase reliability and enhance resource access and management. We have broadly classified these activities at preliminary level into three level of analysis: clustering, mathematical modelling and data caching. The problem of classification and clustering can be approach with tools like Rough set and Soft Set. These tools proved efficient even, if the dataset consist of missing or uncertain data. But, for solving the problem using rough set, the problem must holds true for equivalence relation, which is the fundamental property for rough set. Hence, specific problem of social network must be redefine keeping the properties of transitive, symmetric and reflexive before implementing rough set. In our work, we have studied on the concept of Fiksel's societal network for redefining the social network problem in terms of equivalence relationships.**

Our work starts with a brief discussion on social network, followed by the definition of rough set. Next, we had defined Social network in terms of graph theory and mathematical relation. The next section defines the Fiksel's societal network and social network with respect to rough set. Further, we have discussed on the limitation of Rough set and observed that use of soft set as tool can be an alternative. Hence, to keep continuity, we had made a brief discussion on soft set. Lastly, we have concluded discussing about the relationship between rough set and soft set and their application to social network.

***Keyword---* Social network, Societal Network, Rough Set and Soft set.**

I. INTRODUCTION

The set of relationships between individuals, where each individual is a social entity is known as Social network. Social networks represent both a collection of ties between people and strength of those ties. Mathematically, we can interpret a social network as a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency, such as values, visions, ideas, financial exchange, friendship, sexual relationships, kinship, dislike, conflict or trade [2,10].

Apart from the visual interface for a site that handles such type of Network, there are lot many features, which are complex in nature, makes the interface to work smoothly and efficiency. Among the characteristic, that lies behind the interface of such network is to decrease the access time of users to assets, increase reliability and enhance resource access and management. The working of these activities at preliminary level may be broadly classifies into three levels of analysis, i.e. clustering, mathematical modelling and data caching [4,5]. The problem of clustering or classification may be approach with tools like Rough set and Soft Set.

A rough set is represented by a pair of crisp sets, called the lower and upper approximations of the set. The lower approximation of a rough set comprises of those elements of the universe, which can be said as those who belongs definitely with the available knowledge. The upper approximation on the other hand comprises of those elements which are possibly in the set with respect to the available information [8].

Joseph Fiksel [6] defined the term societal network instead of social network to denote a mathematical system having well defined properties, where node represents individual people and arcs represent directed relations. Thus, societal network is defined as labelled directed network (N, A, T) where $'N'$ is a finite set of nodes, $'T'$ is a finite set of relational type, and $A \subseteq N \text{ X } T$. The set $'A'$ represents all the arcs in the network which link pairs of nodes. Thus, if $(x, y; u) \in A$, then there is an arc from node $'x'$ to node $'y'$ of relational type $'u'$.

II. ON ROUGH SET

The concept of rough set is an approach to deal with imperfect knowledge. It was introduced by Z. Pawlak in 1982 [7,8,9]. They produces extension of Crisp Set Theory, for representing incomplete knowledge. The basic concept of the Rough Set is the notion of approximation space, which is an ordered pair

$A = (U, R)$, where U : nonempty set of objects, called Universe and R: equivalence relation on U, called indiscerinibility relations. If $x, y \in U$ and $x R y$ then $'x'$ and $'y'$ are indistinguishable in $'A'$.

By a knowledge base, we express the relational system $A = (U, R)$, for any subset $P(\neq \phi) \subset R$, the intersection of all equivalence relations in $'P'$ is denoted by $IND(P)$ and is called the indiscernilibity relation over $'P'$. For any $X \subseteq U$ and $R \in IND(A)$, we derive two subsets $\underline{R}X = \bigcup\{Y \in U / R : Y \subseteq X\}$ and $\bar{R}X = \bigcup\{Y \in U / R : Y \bigcap X \neq \phi\}$ are called the R-lower and R-upper approximations of X respectively. The R-boundary of X is denoted by $BN_R(X)$ and is given by $BN_R(X) = \bar{R}X - \underline{R}X$. We say that $'X'$ is rough with respect to $'R'$ if and only if $\underline{R}X \neq \bar{R}X$, equivalently $BN_R(X) \neq \phi$.

III. DEFINING SOCIAL NETWORK IN TERMS OF GRAPH AND MATHEMATICAL RELATIONS

Since long, mathematical notations and relations have been used to describe social network and its properties [14, 15].

A mathematical relation focuses on the ordered pair of actors in a network. Network is the tie or the relationship between the actors. Thus considering a set of objects, $N = \{n_1, n_2, n_3, ..., n_g\}$.
Social network describe these objects as the actors. In case of graph, there are known as nodes, and the edges shows the relationship link between the pairs of actors.

Fiksel's societal network is based on equivalence relationship. A relation is said to be an equivalence relation if it holds true for reflexive, symmetric and transitive.

Definition for Reflexivity: A graph consisting of all the loops such $< n_i, n_i >$ should holds true, then the graph represents a reflexive relation. For example, mother loves her baby boy and the baby also loves his mother.

Definition for Symmetry: A relation is symmetric if, ***i*** chooses ***j*** , then ***j*** also chooses ***i***, thus ***iRj*** if and only if ***jRi.*** A non directional relation represented by a graph is symmetric. In a directed graph, symmetry holds true when the arc $l_k =< n_i, n_i >$ is the set of lines ρ , the arc $l_m =< n_i, n_i >$ is also in ρ .

Definition for Transitivity: The property of transitivity is the patterns of triples of actor in a network or triples of nodes in a graph. A relation is said to be transitive if every time that ***iRj*** and ***jRk***, then ***iRk*** must holds true. By discussing a relation like "is brother of" is transitive whenever ***i*** is the brother of ***j*** and ***j*** is a brother of ***k***, which implies to ***i*** is brother of ***k***.

IV. FIKSEL'S SOCIETAL NETWORK AND ROUGH SET

A societal network is defined to be a finite directed graph in which individuals are represented by nodes, and relations between individuals by labeled arcs. Each individual undergoes state transition at discrete instant of time, so that the societal network may be thought of as a deterministic dynamic process [6]. It was shown by Fiksel that that the individuals of such a network may be divided into equivalence classes, so that the original network may be represented by a reduced network containing one node for each equivalence class. In other words, the concept of structural equivalence and class structure can be used to reduce a societal network to a network of equivalence classes [6].

Definition: In a societal network, a pair of nodes x and y are said to be structurally equivalent when,

(i) $\vee(x, y) = \wedge(x, y)$

(ii) For any $\nu \in \wedge(x, y), R(x, \nu) = R(y, \nu)$

(iii) If $y \in G(x), R(x, y) = R(y, x)$

Where $G(x)$ is the set of nodes adjacent to x .

Definition: A class structure of a societal network $G = (N, A, T)$ is a partition of the set of nodes $'N'$ into disjoint sets $'C_1, C_2, C_3, ...C_k'$, called classes, when satisfies the following two conditions : for any two nodes x, y of a class $'C_i'$ there exist a mapping $M : G(x) \rightarrow G(y)$ such that for $z \in G(x)$:

(i) $R(x, z) = R(y, M(z))$ and

(ii) $'z'$ is in the same class as $M(z)$.

Here, $'M'$ is the mapping and it may be one-to-many or many-to-one.

The above definition of societal network clearly mention on its properties of equivalence classes. Hence, it is now easy to relate the societal network in terms if reflexive, transitive and symmetric. Equivalence relation is the fundamental property for rough set based classification. Therefore, rough set can be used for clustering process, once the societal network is defined in terms of equivalence classes. Further, we have mentioned the definition of equivalence relation and equivalence classes from classical set theory which matches with the expression for lower approximation and upper approximation of the rough set.

Definition for equivalence relation: A relation $'R'$ defined on a set $'A'$ is said to be an equivalence relation in $'A'$ if and only if $'R'$ is reflexive, symmetric and transitive.

Definition for equivalence classes: For any $'A'$ be a non empty set, $'R'$ be an equivalence relation in $'A'$. For each $x \in A,$ the sets $[x]$ are called equivalence classes of $'A'$ given by the relation $'R'$ defined as $[x] = \{y \in A \mid y\,R\,x\}.$

V. LIMITATION FOR USING ROUGH SET TECHNIQUES

There exist some limitations of rough sets which restricts its suitability in some practice. According to Hu, Lin and Han [13], Rough sets theory uses the strict set inclusion definition to define the lower approximation, which does not consider the statistical distribution/noise of the data in the equivalence class. This drawback of the original rough set model has limited its applications in domains where data tends to be noisy or dirty. One other drawback of rough sets theory is the inefficiency in computation, which limits its suitability for large data sets in real-world applications. In order to find the reducts, core and dispensable attributes, the rough sets model needs to construct all the equivalent classes based on the attribute values of the condition and decision attributes. This process is very time-consuming, and thus the model is very inefficient and infeasible, and doesn't scale for large data set, which is very common in data mining applications.

In the recent years, original concept of rough sets has been extended in many different directions. The classical rough set theory is based on equivalence relations, but in many situations, equivalence relations found to be not suitable for clustering with granularity. One such extension is to relax the requirements for the basic relations to be equivalence relations. To be precise it is achieved by dropping the requirement for the base relations to be symmetric or transitive (or both) [1,11,12]. These relations are more abundant in real life situations. These are significant generalizations of the primary notion of rough sets and enhance its applicability. However, soft set cannot be termed as an extension of rough set. It was proposed by Molodtsov (1999) as a mathematical tool for dealing with uncertainties. But, there is a significance relationship between soft set and rough set [1,11]. Moreover, due to relaxation of equivalence relations, soft set in many instances was observed as a better alternative.

VI. ON SOFT SET

Soft Set was proposed by Molodtsov in 1999 [3] as a mathematical tool for dealing with uncertainties. The absence of any restrictions on the approximate description in soft set theory makes it very convenient and applicable.

According to Molodtsov [3], Let $'U'$ be an initial universe set and let E be a set of parameters. $P(U)$ is the power set of $'U'$ and a pair *(F,A)* is called a soft set over U where F is a mapping given by $F: A \to P(U).$ For $\varepsilon \in A,\, F(\varepsilon)$ may be considered as the set of $\varepsilon-$ approximation elements of the soft set.

A. Relationship between Rough Set and Soft Set

Atkas and Cagman in 2007 [1] proved that every rough set may be considered as soft set. The proof is as follows:

Proof: Let *R(X)* be a rough set of '*X*' in the universe '*U*', with respect to the equivalence relation '*R*'. The rough set of '*X*' is defined by an R - upper approximation $\overline{R}(X),$ an R - lower approximation $\underline{R}(X),$ and the equivalence relation '*R*', so all these entities are known. Considering the predicates $p_1(x),$ which stands for "$[x]_R \subseteq X$", and $p_2(x),$ which stands for "$[x]_R \cap X \neq \varnothing$". The conditions $p_1(x)$ and $p_2(x)$ can be treated as elements of a parameter set; that is $E = \{p_1(x), p_2(x)\}.$ Then we can write the function as

$$F: E \to P(U),$$
$$F(p_i(x)) = \{x \in U : p_i(x) \text{ is true}\}, \quad i = 1, 2.$$

Hence, every rough set *R(x)* of *'X'* can be considered as a soft set with representation

$$(F,E) = \{(p_1(x), \underline{R}(X)), (p_2(x), \overline{R}(X))\}.$$

Thus, every rough set R(X) of X may also be considered a soft set with the representation $(F,E) = \{(p_1(x), \underline{R}(x)), (p_2(x), \overline{R}(x))\}.$

VII. OBSERVATION

Societal network can be defined using its properties of equivalence classes. Hence, it is easy to relate the societal network with rough set. But in many situations, equivalence relation (approach using rough set) is not suitable for clustering with granularity. In some of such cases, soft set as a tool can be used for clustering. Further, we have observed that every rough set is a soft set but the reverse is not true in general. This part is important, as it gives us an opportunity to solve more similar type of problems in the same area. Apart from equivalence property, absence of any restriction on the approximate description makes this theory convenient and practical. Thus, analyzing societal network using the soft set tool proved to be better alternative.

VIII. CONCLUSION

In this paper, we have discussed on specific challenges of social network. We have broadly classified these activities at preliminary stage into three level of analysis: clustering, mathematical modelling and data caching. Our work deals with the problem of classification and clustering which can be approach with tools like Rough set and Soft Set. These tools proved to be more efficient even, if the dataset consist of missing or uncertain data. In our paper, apart from general definitions and literature review, we had discussed about Social network with respect to Rough Set. For describing Social network in terms of rough set, we have discussed the Fiksel's societal network, which is based upon the concept of equivalence relation. Next, we have discussed on the limitation of Rough set and observed that use of soft set as tool can be an alternative. Lastly, we have concluded discussing about the relationship between rough set and soft set and their application to social network.

REFERENCES

[1] Aktas, H. and Cagman, N., "Soft set and Soft groups", Information Sciences 177, Pp. 2726–2735, 2007.

[2] Backstrom L., Huttenlocher D., Kleinberg J. & Lan X., "Group formation in large social networks: Membership, growth, and evolution", Proceedings of 12th International Conference on Knowledge Discovery in Data Mining, New York: ACM Press, Pp. 44 – 54, 2006.

[3] D. Molodtsov, "Soft Set Theory – First result", Journal of Comput. Math. Application", 37, Pp. 19 – 31, 1999.

[4] Danah Boyd and Jeff Potter, "Social Network Fragments: An Interactive Tool for Exploring Digital Social Connections" Sketch at SIGGRAPH 2003. San Diego, California: ACM, July 27-31, 2003.

[5] Danah Boyd & Ellison N. B. "Social network sites: Definition, history, and scholarship" Journal of Computer-Mediated Communication, Vol. 13(1), 2007, Article 11-15, 2007

[6] Joseph Fiksel, "Dynamic Evolution in Social Networks," Journal of Mathematical Sociology, 7(1980): 27-46, 1980

[7] Pawlak, Z., Rough sets, Int. jour. of Info. And Comp. Sc. II, (1982), 341-356, 1982

[8] Pawlak, Z.: Rough sets, Theoretical aspects of reasoning about data. Kluwer academic publishers, London, 1991.

[9] Tripathy B K & Mitra A, "Some topological Properties of rough set and their applications", Int. J. Granular Computing, Rough Set and Intelligent System, Vol. 1, 4, Pp. 355-360, 2010.

[10] Wasserman, S. and K. Faust (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.

[11] Maji, P.K., Biswas, R., and Roy, A.R. "Soft set theory". Computers and Mathematics with Applications. 45, 555–562, 2003.

[12] Tutut Herawan and Mustsfa Mat Deris, "A Direct Proof of Every Rough Set is a Soft Set", Third Asia International Conference on Modeling and Simulation, (IEEE Comp. Society), 119 – 124, 2009.

[13] Xiaohua Hu C, T. Y. Lin and J Han, "A New Rough Sets Model Based on Database Systems", Fundamenta Informaticae, IOS Press, 1–18, 2004.

[14] Hage P and Harary F., "Structural Models in Anthropology, Cambridge University Press, Cambridge, 1983.

[15] Wasserman s and Faust K., "Social network analysis: methods and Applications", Cambridge University Press, 1994.