NEAR SET AN APPROACH AHEAD TO ROUGH SET: AN OVERVIEW

Kavita R Singh, Shivanshu Singh
Department of Computer Technology, YCCE, Nagpur (MS), India.

ABSTRACT

Rough Set Theory is a fairly new concept that has found applications to various soft computing techniques. It offers a set theory approach to manage the uncertainty in data systems. It has been used for the discovery of data dependencies, importance of features, patterns in sample data, feature space dimensionality reduction, and the classification of objects. Objects can be classified by means of their attributes when considered in the context of an approximation space. The Near Sets represent a generalization of Rough Sets. It presents a nearness approach to classifying objects. In this paper we present an overview of basics of rough sets and near sets along with their application to face recognition problem.

KEYWORDS: Rough Sets, Near Sets.

I. INTRODUCTION

Rough set theory [2, 3, 4, 5] introduced by Z. Pawlak in 1991, is one of the new approaches towards soft computing finding a wide application today. Rough Set Theory manages the vagueness in a data system and has been successfully used to formulate the rules. These rules can be used to discover the hidden patterns in data. In addition, Rough Set methods can be used to classify new samples based on what is already known. Unlike other computational intelligence techniques, Rough Set analysis requires no external parameters and uses only the information presented in the given data. Briefly, Pawlak suggested that Rough Set when used as a classifier, objects can be classified by means of their attributes [1]. By way of extension of Pawlak's approach to classification, Near Set is an approach to solving the problem of what it means for objects with common features to be near each other qualitatively but not necessarily spatially.

Near Sets presents a nearness approach to classifying objects. It harkens back to the original 1981 paper by Z. Pawlak, who pointed out that exact classification of object is often impossible [1]. Thus near Sets represent a generalization of the approach to the classification of objects introduced by Z. Pawlak.

From a Rough Sets point-of-view, the main focus is on the approximation of sets with non-empty boundaries. In contrast, in a Near Sets approach to set approximation, the focus is on the discovery of Near Sets in the case where there is either a non-empty or an empty approximation boundary. Object recognition problems, especially in images [10, 11 and 22] using the nearness of objects have motivated the introduction of Near Sets.

In this paper we are providing an overview of a Rough Set and general theory of nearness of objects in a Near Set approach to set approximation.

The paper is organized as follows. Section 2 presents an overview of Rough Set theory. Section 3 presents an overview on the concept of Near Sets. Section 4 describes the use of both Rough Set theory and Near Set in feature selection. Section 5 briefs on the application of set approximation approach from Rough Sets and Near Set to face recognition followed by conclusion.

II. ROUGH SETS

An approach put forth by mathematician Z. Pawlak in the beginning of the eighties, Rough Sets have come up as a mathematical tool to treat the vague and imprecise data. Rough Set Theory is similar to

Fuzzy Set Theory in many aspects. However the uncertainty and imprecision is expressed by the Boundary Region of a set, as opposed to the Partial Membership as in Fuzzy Set Theory. Rough Set concept is generally defined by means of interior and closure topological operations known as Approximations [1].

Fuzzy Sets are defined by employing the Fuzzy Membership Function involving advanced mathematical structures, numbers and functions. Rough Sets are defined by topological operations called Approximations, thus the definition requires some advanced mathematical concepts.

Moreover, like other computational intelligence techniques, Rough Set analysis requires no external parameters and uses only the information presented in the given data. An attractive feature of Rough Set theory is that it can predict whether the data is complete or not, based on the data itself. If the data is incomplete, it suggests more information about the object is required. On the other hand, if the data is complete, Rough Sets can determine whether there are any redundancies in the data and find the minimum data needed for classification. This property of Rough Sets is very important for applications where the domain knowledge is very limited or data collection is expensive or laborious since it makes sure the data collected is just sufficient to build a good classification model without sacrificing the accuracy and without wasting time and effort to gather extra information about the objects [3, 4 and 5].

The uncertainty and imprecision in is expressed by a boundary region of a set. It deals with the approximation of an arbitrary subset of a universe by two definable or observable subsets called Lower and Upper Approximations of a Rough Set. By using the concepts of Lower and Upper Approximations in Rough Set theory, the knowledge hidden in information systems can be explored and correct decisions could be derived.

In RST, information about the real world is expressed in the form of an information table. An information table can be represented as a pair$\mathcal{A} = (U, A)$, where, U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that information function $f_a: u \rightarrow V_a$, for every $a \in A$. The set V_a is called the value set of a. Furthermore, a decision system is any information table of the form $\mathcal{A} = (U, A \cup \{d\})$, where $d \notin A$ is a decision attribute. For every set of attributes $B \subseteq A$, an indiscernibility relation $IND(B)$ is defined in the following way: two objects, x_i and x_j, are indiscernible by the set of attributes $B \subseteq A$, if $b(x_i) = b(x_j)$for every $b \subseteq B$. The equivalence class of $IND(B)$ is called elementary set in b because it represents the smallest discernible groups of objects. For any element x_i of u, the equivalence class of x_i in relation $IND(B)$ is represented as$[x_i]_{IND(B)}$. The notation $[x]_B$ denotes equivalence classes. Thus the family of all equivalence classes, partition the universe for all b will be denoted by $U/_B$. This partitions induced by an equivalence relation can be used to build new subsets of the universe. The construction of equivalence classes is the first step in classification with Rough Sets.

Rough Membership Function

Rough sets can also be defined by Rough Membership Functions instead of Approximation. A rough membership function (*rmf*) makes it possible to measure the degree that any specified object with a given attribute values belongs to a given decision set x. let, $B \subseteq A$ and let x be a set of observations of interest. The degree of overlap between x and $[x]_B$ containing x can be quantified with an *rmf given* by:

$$\mu_X^B: \rightarrow [0,1] \qquad (1)$$

$$\mu_X^B(x) = \frac{|[x]_B \cap X|}{|[x]_B|} \qquad (2)$$

where, $|\cdot|$ denotes the cardinality of a set. The rough membership value μ_X^B may be interpreted as the conditional probability that an arbitrary element x belongs to X given B. The decision set x is called a generating set of the rough membership μ_X^B. Thus Rough Membership Function quantifies the degree of relative overlap between the decision set x and the equivalence class to which x belongs.

III. NEAR SET

Near set is a special theory about Nearness of objects. It was first presented by James Peter in the year of 2006 and was formally defined in 2007. It represents a generalization of the approach to the classification of objects introduced by Z. Pawlak during the early 1980s. Like Fuzzy Sets and Rough Sets which instead of contracting complement each other, Near Sets and Rough Sets are also like two sides of the same coin. The various different domains where the Near Set has been successfully applied are: feature selection [14], object recognition in images [11 and 24], image processing [10], granular computing [13 and 19] face recognition [20 and 21] and in various forms of machine learning [1, 12, 13, 16, 15, 17 and 18].

In Near Sets theory, each object is described by a list of feature values. The word feature corresponds to an observable property of physical objects in our environment. For instance, for a feature like the nose of a human face, the feature values would be nose length or nose width. Comparing this list of feature values, similarity between the objects can be determined and can be grouped together in a set, called as Near Sets. Thus Near Set theory provides a formal basis for the observation, comparison and recognition/classification of objects. The nearness of objects can be approximated using Near Sets. Approximation can be considered in the context of information granules (neighbour hoods). Any approximation space is a tuple given in equation (3)

$$AS = (U, \mathcal{F}, v) \qquad (3)$$

where $\mathcal{F}$ is a covering of finite universe of object U, i.e., $\bigcup \mathcal{F} = U$ and $v: P(U) \times P(U) \rightarrow [0,1]$

maps a pair of set to a number in $[0,1]$ representing the degree of overlap between the sets and $P(U)$ is a power set of U [4]. For a given approximation space $AS = (U, \mathcal{F}, v)$, we define a binary link relations $link_{\mathcal{F}} \subseteq U$.

For any, $X \subseteq U, \mathcal{F}$-lower approximation of X, and $\mathcal{F}$-upper approximation of X is defined respectively by (4) and (5).

$$\mathcal{F}_{*}X = \bigcup\{Y \in \mathcal{F} | v(X, Y) = 1\}, \qquad (4)$$

$$\mathcal{F}^{*}X = \bigcup\{Y \in \mathcal{F} | v(X, Y) > 0\}, \qquad (5)$$

The *lower approximation* of a set *X* is the set of all objects, which can be for *certain* classified as *X*.
The *upper approximation* of a set *X* is the set of all objects which can be *possibly* classified as *X*.
The lower and upper approximations of a set lead naturally to the notion of a boundary region of an approximation. Thus, the lower- and upper- approximations result in an increase in the number of neighbourhoods used to assess the nearness of a classification [2].

Overlap Function

Earlier we have seen the concept of rough membership function in the context of Rough Set, used to measure the degree of overlap. In Near Set, it is now possible to formulate a basis for measuring average; the degree of overlap between Near Sets. Let *X, Y* defined in terms of a family of neighborhoods $N_r(B)$. There are two forms of the overlap function.

$$v_{N_r(B)}(X, Y) = \begin{cases} \dfrac{|X \cap Y|}{|Y|}, if\, |Y| \neq \phi \\ 1, otherwise \end{cases} \qquad (6)$$

$$\nu_{N_r(B)}(X,Y) = \begin{cases} \dfrac{|X \cap Y|}{|X|}, if |X| \neq \phi \\ \dfrac{}{1, otherwise} \end{cases} \quad (7)$$

Coverage $\nu_{N_r(B)}(X,Y)$ is used in case where it is known that $|X| \leq |Y|$. For example coverage can be used to measure the degree that a class $[x]_{B_r}$ is covered by the lower approximation $N_r(B)_* X$ in

$$\nu_{Nr}(B)([x]_{Br}, N_r(B)_* X) = \frac{|[x]_{B_r} \cap N_r(B)_* X|}{|N_r(B)_* X|} \quad (8)$$

is called lower coverage.

IV. FEATURE SELECTION

Practical outcomes of the family of soft computing tools are feature selection. In Rough Sets the task of feature selection requires choosing the smallest subset of conditional features so that the resulting reduced dataset remains consistent with respect to the decision feature. The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capacity of the decision feature as the original. A *reduct* is defined as a subset of minimal cardinality R_{min} of the conditional attribute set such that

$$R=\{X:X\subseteq C,\gamma_X(D)=\gamma_C(D)\} \quad (9)$$

$$R_{min} = \{X : X \in R, \forall Y \in R, |X| \leq |Y|\} \quad (10)$$

$$CORE(R_{min}) = \bigcap R_{min} \quad (11)$$

The intersection of all the sets in R_{min} is called the core, the element of which are those attributes that cannot be eliminated from the set without changing the original classification to the dataset. Clearly each object can be uniquely classified according to the according to the attribute values remaining.
Feature selection is also one of the important aspects Near Set approach. Here each partition $\xi_{\mathcal{O},} B_r$ contains classes defined by the relation $\sim B_r$. The classes in each $\xi \mathcal{O}, B_r \in N_r(B)$ with information content greater than or equal to some threshold *th* are of interest. The basic idea here is to identify probe functions that lead to partitions with the highest information content, which occurs in partitions with high numbers of classes. In effect, as the number of classes in a partition increases, there is a corresponding increase in the information content of the partition.

V. FACE RECOGNITION WITH ROUGH SET AND NEAR SET

Rough Set theory has been employed by K. Singh *et. al.,* [20] for face recognition using only geometrical features. The ADNN rough neural network [20] employed is built from approximation and decider neuron using the concept of rough sets.
Literature cites that Rough Sets have been successfully used with other theories to build up a hybrid system. Yun *et al*. [6] used rough-support vector machine integration and developed the Improved Support Vector Machine (ISVM) algorithm to classify digital mammography images, where Rough Sets are applied to reduce the original feature sets and the support vector machine is used classify the reduced information.

Based on geometric feature and appearance feature, there are a few works been done on facial expression recognition using Rough Set and support vector machine. Chen *et al.* [7] proposed a novel approach based on Rough Set theory and SVM by considering only geometric features.
Later, S. Gupta *et. al*.[21], extended ADNN [20] for face recognition, with Near Set for facial feature selection. The algorithm used to find partition selection and then to select the best features which can be fed to the SVM classifier. Using near set author has presented how the chosen features can affect the accuracy of face recognition system. Results shows that number of support vectors and margin are maximum when the feature with largest average near coverage ($\bar{v}$) is chosen for face recognition. It has also been shown that better recognition accuracy can be achieved with nose width as selected feature [21].

VI. CONCLUSION

An overview of different approaches to deal with uncertainties has been provided in this paper. While Rough Sets provide a powerful tool to objects classification by means of their attributes, Near Sets present a nearness approach to classifying objects. We have also seen how feature selection can be achieved with these two approaches. Both theories have found rapidly increasing applications in many areas. We explored the implementation of the two approaches in a face recognition system. Both approaches will find more applications to intelligent systems. On this basis we present a study for the reader to understand and differentiate clearly between the two approaches.

REFERENCES

[1]. Pawlak, Z.: Classification of objects by means of attributes, Research Report PAS 429, Institute of Computer Science, Polish Academy of Sciences, ISSN 138-0648, January (1981).
[2]. Z. Pawlak, "Rough sets", *International J. Comp. Inform. Science*, vol. 11, pp.341-356, 1982.
[3]. Z. Pawlak,"Rough sets – Theoretical aspects of reasoning about data". Kluwer, 1991.
[4]. Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, "Rough Sets". *Communications of the ACM*, vol.38, no.11, pp.88-95, 1995.
[5]. L. Polkowski, "Rough Sets: Mathematical Foundations". Physica-Verlag, 2003.
[6]. Y. Jiang, Z. Li, L. Zhang, P. Sun, "An Improved SVM Classifier for Medical Image Classification", in M. Kryszkiewicz et al., Eds., *Int. Conf. on Rough Sets and Emerging Intelligent Systems Paradigms*, LNAI, vol. 4585, pp. 764-773, 2007.
[7]. P. Chen, G. Wang, Y. Yang and J. Zhou, "Facial Expression Recognition Based on Rough Set Theory and SVM" Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Rough Sets and Knowledge Technology, Volume 4062/2006, pp.772-777, 2006
[8]. S.K Pal; J. F. Peters; L. Polkowski; A .Skowron: "*Rough Neural Computing. An Introduction*". In Pal et al
[9]. J.F.Peters and Marcin S. Szczuka: "*Rough Neurocomputing: A Survey of Basic Models of Neurocomputation*" J. J. Alpigni et al. (Eds): RSCTC 2002.
[10]. M. Borkowski, J.F. Peters, Matching 2D image segments with genetic algorithms and approximation spaces, *Transactions on Rough Sets*, **V**, LNCS 4100 (2006), 63-101.
[11]. C. Henry, J.F. Peters, Image Pattern Recognition Using Approximation Spaces and Near Sets, In: *Proceedings of Eleventh International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* (RSFDGrC 2007), *Joint Rough Set Symposium* (JRS 2007), *Lecture Notes in Artificial Intelligence*, 4482 (2007), 475-482.
[12]. D. Lockery, J.F. Peters, Robotic target tracking with approximation space-based feedback during reinforcement learning, Springer Best Paper Award, In: *Proceedings of Eleventh International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* (RSFDGrC 2007), *Joint Rough Set Symposium* (JRS 2007), *Lecture Notes in Artificial Intelligence*, 2007, 483-490.
[13]. J.F. Peters, Perceptual granulation in ethology-based reinforcement learning. In: Pedrycz, W., Skowron, A., Kreinovich, V. (Eds.), *Handbook on Granular Computing*, Wiley, NY, 2007.
[14]. J.F. Peters, S. Ramanna, Feature Selection: Near Set Approach. In: Z.W. Ras, S. Tsumoto, D.A. Zighed (Eds.), 3*rd Int. Workshop on Mining Complex Data* (MCD'08), ECML/PKDD-2007, LNAI, Springer (2007), *in press*.
[15]. J.F. Peters, M. Borkowski, C. Henry, D. Lockery, D.S. Gunderson, Line-Crawling Bots that Inspect Electric Power Transmission Line Equipment. In: *Proc. Third Int. Conference on Autonomous Robots and Agents* (ICARA 2006), Palmerston North, New Zealand (2006), 39-44.

[16]. J.F. Peters, M. Borkowski, C. Henry, D. Lockery; "Monocular vision system that learns with approximation spaces," In: Ella, A., Lingras, P., Slezak, D., Suraj, Z. (Eds.), *Rough Set Computing: Toward Perception Based Computing*, Idea Group Publishing, Hershey, 1-22,2006.

[17]. J.F. Peters, C. Henry; "Reinforcement learning with approximation spaces," *Fundamenta Informaticae*, 71, nos. 2-3,323-349, 2006.

[18]. J.F. Peters, S. Shahfar, S. Ramanna, T. Szturm; " Biologically-inspired adaptive learning: A Near Set approach," In: *Proc. Frontiers in the Convergence of Bioscience and Information Technologies* (FBIT07), IEEE, NJ, 11 October 2007, *in press*.

[19]. A. Skowron, J.F. Peters; "Rough granular computing," In: Pedrycz, W.,Skowron, A., Kreinovich, V. (Eds.), *Handbook on Granular Computing*, Wiley, NY, 2007.

[20]. K. R. Singh and M.M. Raghuwanshi; "Face Recognition with Rough-Neural Network: A Rule Based Approach", *International workshop on Machine Intelligence Research*, pp 123-129, 24th Jan 2009.

[21]. S. Gupta, K.S.Patnaik; "Enhancing performance of face recognition system by using Near Set approach for selecting facial features" *Journal of Theoretical and Applied Information Technology,* pp.433-441, 2008.

[22]. J.F.Peters;"Near Sets. General Theory about Nearness of Objects" Applied Mathematical Sciences, Vol. 11, no.53, 2609-2629, 2007.

Author's Biography.

Kavita R Singh received the B.E degree in Computer Technology in 2000 from RGCERT, Nagpur University, Nagpur, India, the M.Tech. degree in Computer Science and Engineering from Birla Institute of Technology, Ranchi, India, in 2007, and now she is pursuing her Ph. D. degree in Computer Science from the Sardar Vallabhbhai National Institute of Technology, Surat, India. She is currently a Lecturer in Computer Technology Department, YCCE, Nagpur. Her current research interests include the area of Data structures, Database, Rough sets, Image Processing, pattern Recognition and Near Sets.

Shivanshu Singh is a final year student in Computer Technology of YCCE, Nagpur University, Nagpur, India. He did his higher education from St. Xavier's High School, Ranchi, and Jharkhand in 2006. His interests include the area of programming, rough sets and Near Sets.