World Academy of Science, Engineering and Technology
International Journal of Information and Communication Engineering
Vol:11, No:10, 2017

Application of Granular Computing Paradigm in Knowledge Induction

Iftikhar U. Sikder

***Abstract*—**This paper illustrates an application of granular computing approach, namely rough set theory in data mining. The paper outlines the formalism of granular computing and elucidates the mathematical underpinning of rough set theory, which has been widely used by the data mining and the machine learning community. A real-world application is illustrated, and the classification performance is compared with other contending machine learning algorithms. The predictive performance of the rough set rule induction model shows comparative success with respect to other contending algorithms.

***Keywords*—**Concept approximation, granular computing, reducts, rough set theory, rule induction.

I. Introduction

THERE is a growing interest in data science, particularly in data analytics and its applications in a wide range of disciplines that encompasses business, engineering and medical domains. The emerging surge of data science is mainly triggered by the rapid production of large amount of data generated by numerous computing devices both on and offline. With the overabundance of data along with exponential increase in computing and processing power, there is a growing need to understand the underlying patterns and relationships inherent in data, forcing the research communities to develop novel data mining algorithms to match with the growth of data volume. As an emerging computing paradigm, granular computing [1]-[5] offers a framework that includes data processing methodologies, tools, and techniques which view data not in isolation, but in granules or clumps that are naturally inherent in data. The formation of such granules could occur by virtue of indistinguishability or indiscernibility, spatio-temporal adjacency, similarity in attribute values, e.g. feature space, clustering, etc. The underlying assumption in granular computing framework is that there exist different patterns at different levels of resolutions, hence the machine learning algorithms should take advantage this feature in the development of corresponding methods and techniques reflective of such processes. Rough set theory and fuzzy set are based on this assumption [6]. In particular, in rough set theory, one can generate 'granules' or concepts wherein various subset of elementary sets of observations could exist by virtue of indiscernibility or indistinguishability. These granules can approximate a concept with respect to other granules, thus allowing various feature extraction of rule induction.

Iftikhar U. Sikder is an Associate Professor at the department of Information System in Cleveland State University, USA (phone: 216-687-4758; fax: 216-687-5448; e-mail: i.sikder@ csuohio.edu).

II. Rough Set Theory as Knowledge Induction Method

A. Review of Rough Set Model

A widely used approach of Granular Computing paradigm in Artificial Intelligence is the Rough Set Theory [7]. Information granulation or the concept of indiscernibility or granularity is a fundamental aspect of rough set theory. The concept of a class is characterized in terms of elementary sets in an approximation space of conditioning attributes. In data mining, when the feature space has an intrinsic tendency to form local clumps or granules because of inherent similarity, rough set is an appropriate tool to deal with them. The finer the granulation, the greater is the definability of the concept. Granularity is induced by the partition of data in the attribute space resulting in the indiscernibility relations and the associated equivalence class (i.e., relations are reflexive, symmetric, and transitive). The concept of indiscernibility helps identify the boundary-line cases. If a set is a subset of a concept, then it is completely definable, on the other hand, if the intersection of a set with a class or a concept is not null then it implies that the set approximately belongs to a class or to its complements with respect to the available attribute, and thus resulting in the boundary line cases [8]. Rough set theory is also relevant when the data is vague, incomplete, imprecise and fragmentary [9]-[11]. It also offers a "non-invasive" approach to knowledge discovery. Unlike black box models like Neural Network, or a priori-dependent Bayesian inference, rough set is transparent to modeler because of its data driven approach [12]. It does not require additional assumption about membership function like fuzzy systems or statistical distribution parameter. Rough set has been extensively used in machine learning, data mining giving rise to many applications in predictive modeling, feature reduction in various disciplines[11], [13]-[16].

B. Mathematical Formalism of Rough Set Theory

Rough set allows to characterize a decision class in terms of elementary attribute sets in an approximation space. The decision table is represented $(U, C \cup \{d\})$, where $d \notin C$ is the decision attribute in a given data set and U is the closed universe which consists of non-empty finite set of objects and C is a non-empty finite set of attributes, such that $c : U \to V_c$, for every c $c \in$ C, V_c is a value of attribute c. For $B \subseteq C$, the granule of knowledge about a theme with respect to indiscernibility relation can be represented as:

$$Ind(B) = \{x, x') \in U^2 \mid \forall c \in Bc(x) = c(x')\} \quad (1)$$

where *Ind*(*B*) represents the indiscernibility with respect to attribute subset B. Thus, the object *x* and x' are indiscernible from each other if $(x, x') \in Ind(B)$. The decision class is approximated by lower and upper approximation of decision concept as follows:

$$B\underline{X} = \{x \in U \mid Ind(x) \subseteq X \quad (2)$$

$$B\overline{X} = \{x \in U \mid Ind(x) \cap X \neq \varnothing\} \quad (3)$$

The lower approximation $B\underline{X}$ is the set of he objects in which objects can be classified with certainty on basis of knowledge in *B*, while, the objects in upper approximation $B\overline{X}$ can be only classified as the possible occurrence of a given class. The boundary region $B\overline{X} - B\underline{X}$ represents the uncertainty in decisive classification. Given, this partition, rough set allows to develop a methodology to identify the critical attributes or minimal subset of attributes that maintains the same partition. By eliminating redundant attributes, it provides a method for feature reduction or knowledge compression methods in data. Such minimal subset of attributes is called *reduct.* Thus, a reduct represents an attribute subset $B \subseteq C$ of an information system such that, after removal of an attribute/s from an equivalence class, it preserves indiscernibility relation. There are various ways that one can compute reduct; however, a classical approach is to rely on a discernibility function expressed as follows:

$$f_A(a_1,.....,a_m) = \wedge\{c_{ij} : 1 \leq i < j \leq n \,\&\, c_{ij} \neq \varnothing\} \quad (4)$$

It represents the prime implicants of candidate attributes $(a_{i1} \wedge .... \wedge a_m)$. Although computing prime implicants is NP-Hard, there are numerous heuristic methods available to minimize the search space to a computationally efficient way to generate minimal subset. There could be many reducts in an information system that preserves the partition induced by the equivalence classes. In such cases, the intersection of all reducts is called the *Core*. The removal of an attribute from the core changes the partition induced by equivalence class structure. Thus, attributes belonging to core are indispensable.

It is also possible to discover the degree of dependency of the attributes and their inclusiveness. In rough set, rules are derived from reducts in the form of "*If..then*", where the antecedents are the conjunctions of attribute conditions and the consequents are the decision class or its disjunction [17], [18]. For a set of attributes $(B_1, B_2, B_3, \ldots, B_n)$ and decision *d*, rules can be induced in the form:

$$(B_i = a) \ \wedge (B_j = b) \wedge \ \ldots \ (B_k = b) \rightarrow (d = d_1) \quad (5)$$

III. Research Method

The rough set knowledge induction process has been implemented to derive a set of rules. In particular, we use MODLEM [19], [20] algorithm which heuristically generates unordered minimal set of rules for every observed class using sequential covering scheme. The classification strategy involves using nearest rules and class approximation. The class conditional entropy measure is used to choose the elementary 'best' condition of rule (i.e., the first candidate for the condition part). It is possible to generate of certain rules from lower approximation and possible rules from upper approximation of rough set. The research design includes:

1) Selection: This involves generation of decision table in the specification of rough set. We used a dataset Wisconsin Breast cancer data from UCI Machine learning website [21]. This dataset consists of 10 features and a class variable. These data are the accumulated data of different groups. Each group of data is collected at different times.

1. Id number
2. Clump Thickness
3. Uniformity of Cell Size
4. Uniformity of Cell Shape
5. Marginal Adhesion
6. Single Epithelial Cell Size
7. Bare Nuclei
8. Bland Chromatin
9. Normal Nucleoli
10. Mitoses
11. Class (Benign or Malignant)

All the features are in a uniform scale from 1 to 10. Every feature is considered as a risk factor such that the risk will increase as the value of the features are increased.

2) Preprocessing: The initial preprocessing step involves data cleaning. The attribute representing the id number was removed as it does not contribute any knowledge. There are 16 observations out of 699 which had missing values. These datasets were removed. Next, few observations were further removed by detecting outliers and extreme values based on interquartile ranges. The sampling scheme involves removal of class imbalance. Since the minority class "malignant" accounts for only 35% of the observation, the imbalance needs to be removed to avoid any potential biased estimate of the model prediction. We used Synthetic Minority Oversampling Technique (SMOTE) [22], [23] to remove the class imbalance by generating synthetic samples based on nearest neighbor distribution pattern. The dataset was partitioned in 10-fold for cross-validation where each fold is used as testing set to predict the class value based on the remaining nine folds treated as training set. The process is repeated for each fold, and the predictive performance of the weighted average is reported.

3) Transformation: This involves converting the numeric attributes into nominal type using a supervised algorithm. We used minimum description length method [24] into ranges of values that takes in account the distribution of

the predictor class.

4) Mining: The mining step involves generating reducts, core, ranking of attributes and rule induction. The "if-then" rules are mined in a two-stage process. This process produces minimal attribute subsets, attribute dependency followed by patterns and rules which are generated from these subsets.
5) Evaluation: Mined patterns or rules are applied to test dataset (K-fold cross validation) to classify new instances. Individual patterns or rules are measured or manually inspected. Classificatory performance is compared with other algorithm using ROC curve, precision recall, F-measure, kappa coefficient.

IV. Research Results

The rough set approximation of the decision classes for upper and lower approximation is generated by using Rough set software ROSE2 (rough set data explorer) [18]. Reducts were generated by using heuristic method. The generated reducts are:

Reduct 1: {Clump Thickness, Uniformity of Cell Shape, Single Epithelial Cell Size, Bare Nuclei}
Reduct 2: {Uniformity of Cell Shape, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin}
Reduct 3: {Clump Thickness, Uniformity of Cell Shape, Bare Nuclei, Normal Nucleoli}
Reduct 4: {Clump Thickness, Uniformity of Cell Size, Bare Nuclei, Bland Chromatin}
Reduct 5: {Clump Thickness, Uniformity of Cell Size, Bare Nuclei, Normal Nucleoli}
Reduct 6: {Clump Thickness, Uniformity of Cell Size, Single Epithelial Cell Size, Bare Nuclei, Normal Nucleoli}
Reduct 7: {Uniformity of Cell Size, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Normal Nucleoli}

The intersection of all the reducts i.e., a core is *Bare Nuclei* which occur in all the reducts making it an indispensable attribute. On the other hand, the least occurrence is *Marginal Adhesion* which has just 14.29% frequency. To compare the result we use a quick reduct algorithm to evaluate the subsets using rough set dependency and return a subset giving only the rough set positive region [25]. The worth of an attribute was also estimated by measuring the information gain with respect to the class. The result is shown in Table I.

TABLE I
Attribute Subsets and Ranking

Attributes	Information Gain Score
Uniformity of Cell Size	0.684
Uniformity of Cell Shape	0.661
Bare Nuclei	0.596
Bland Chromatin	0.537
Single Epithelial Cell Size	0.525
Normal Nucleoli	0.48
Marginal Adhesion	0.446
Clump Thickness	0.435
Mitoses	0.199

Although *Bare Nuclei* is indispensable from rough set point of view; however, from information gain point of view, the *Uniformity of Cell Size* appears as the most significant. The rule induction algorithm MODLEM generates rough set rules. In total, 32 rules were generated of which 17 rules are have decision class *malignant.* Below, few significant rules are shown:

If Bare Nuclei > 2.5. & Normal Nucleoli < 2.5 & Clump_Thickness > 6.5 => (Class = malignant) (28/28, 11.72%)

(Normal Nucleoli between (2.5-8.5) & (Marginal Adhesion > 3.5) & (Uniformity of Cell Shape >4.5}) => (Class = 4) (56/56, 23.43%)

TABLE II
Comparative Measures of Predictive Performances

Classifier	Precision	Recall	F-Measure	ROC Area	Kappa
Rough Set (MODLEM)	0.972	0.972	0.972	0.971	0.9424
Decision Tree (J48)	0.964	0.964	0.964	0.967	0.9253
Random Forest	0.977	0.977	0.977	0.993	0.9512
Naïve Bayes	0.977	0.977	0.977	0.995	0.9513

Finally, we compare the predictive performance of the rough set (MODLEM) classifier with other classifier such as decision tree (J48), Random forest, Naïve Bayes methods using the same 10-fold cross validation technique. Table II shows comparative measures of the predictive performances. The overall accuracy of MODLEM is around 95.6%. The kappa coefficient indicates a model's predictive performance beyond random assignment of class labels. The rough set model scored higher kappa coefficient than decision tree. The recall measure of MODLEM is 97.2% which indicates very high true positive rate. The receiver operating curve (ROC) indicates very high predictive performance for both the positive (*malignant*) and negative (*benign*) classes. The area under the ROC for MODLEM is comparable to that of other classifier and outperforms decision tree.

Fig. 1 The Receiver Operating Curve (ROC) *malignant* class

V. Conclusion

In this paper, we present an approach to discover knowledge under based on rough set theory rule induction as illustration of granular computing paradigm. The comparison

of predictive performance shows that rough set rule induction performance scores are comparable with those baseline classifiers. However, rough set rule induction method slightly outperforms the decision tree (J48) in terms precision, recall, kappa coefficient and the area under the ROC. The overall predictive performance is significantly high. The higher performance is due to fact that rough set rules were derived based on the context of nearest neighbors as well as approximation space.

REFERENCES

[1] A. Bargiela and W. Pedrycz, *Granular computing: an introduction* vol. 717: Springer Science & Business Media, 2012.
[2] T. Y. Lin, Y. Y. Yao, and L. A. Zadeh, *Data mining, rough sets and granular computing* vol. 95: Physica, 2013.
[3] Y. Yao, "A triarchic theory of granular computing," *Granular Computing,* vol. 1, pp. 145-157, 2016.
[4] L. Livi and A. Sadeghian, "Granular computing, computational intelligence, and the analysis of non-geometric input spaces," *Granular Computing,* vol. 1, pp. 13-20, 2016.
[5] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: perspectives and challenges," *IEEE Transactions on Cybernetics,* vol. 43, pp. 1977-1989, 2013.
[6] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information sciences,* vol. 177, pp. 3-27, 2007.
[7] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information sciences,* vol. 177, pp. 28-40, 2007.
[8] I. U. Sikder, "Rough Sets and Granular Computing in Geospatial Information," in *Handbook of Research on Geoinformatics*, ed: IGI Global, 2009, pp. 154-159.
[9] C. H. Aldridge, "A Theoretical Foundation for Geographic Knowledge Discovery in Databases," in *First International Conference on Geographic Information Science*, Georgia, USA, 2000.
[10] I. U. Sikder and A. Gangopadhyay, "Distributed Data Warehouse for Go-spatial Services," in *ERP & Data Warehouse in Oganizations: Issues and Challenges*, G. Grant, Ed., ed: IRM Press, 2003, pp. 132-145.
[11] I. U. Sikder, "A variable precision rough set approach to knowledge discovery in land cover classification," *International Journal of Digital Earth,* vol. 9, pp. 1206-1223, 2016.
[12] I. Düntsch and G. Gediga, "Rough set data analysis--A road to non-invasive knowledge discovery," 2000.
[13] J. T. Yao and N. Azam, "Web-Based Medical Decision Support Systems for Three-Way Medical Decision Making With Game-Theoretic Rough Sets," *Ieee Transactions on Fuzzy Systems,* vol. 23, pp. 3-15, Feb 2015.
[14] I. U. Sikder and A. Gangopadhyay, "Managing uncertainty in location services using rough set and evidence theory," *Expert Systems with Applications,* vol. 32, pp. 386-396, Feb 2007.
[15] I. U. Sikder and T. Munakata, "Application of rough set and decision tree for characterization of premonitory factors of low seismic activity," *Expert Systems with Applications,* vol. 36, pp. 102-110, 2009.
[16] S. Mal-Sarkar, I. U. Sikder, and V. K. Konangi, "Spatio-temporal pattern discovery in sensor data: a multivalued decision systems approach," *Knowledge-Based Systems,* vol. 109, pp. 137-146, 2016.
[17] J. W. Grzymala-Busse and S. Siddhaye, "Rough set approaches to rule induction from incomplete data," in *Proceedings of the IPMU*, 2004, pp. 923-930.
[18] B. Prędki and S. Wilk, "Rough set based data exploration using ROSE system," in *Foundations of Intelligent Systems: 11th International Symposium, ISMIS'99 Warsaw, Poland, June 8–11, 1999 Proceedings*, Z. W. Raś and A. Skowron, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 172-180.
[19] J. Stefanowski and S. Wilk, "Improving rule based classifiers induced by MODLEM by selective pre-processing of imbalanced data," in *Proc. of the RSKD Workshop at ECML/PKDD, Warsaw*, 2007, pp. 54-65.
[20] J. Stefanowski, "The rough set based rule induction technique for classification problems," in *In Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT*, 1998.
[21] "Breast Cancer Wisconsin (Original) Data Set," W. H. Wolberg, Ed., ed. UCI Machine Learning Repository, 1992.
[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.
[23] T. Zhu, Y. Lin, and Y. Liu, "Synthetic minority oversampling technique for multiclass imbalance problems," *Pattern Recognition,* 2017.
[24] U. Fayyad and K. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," 1993.
[25] A. Chouchoulas and Q. Shen, "Rough set-aided keyword reduction for text categorization," *Applied Artificial Intelligence,* vol. 15, pp. 843-873, 2001.