A Novel Approach for Feature Selection using Rough Sets

Reference:

Yadav, N., & Chatterjee, N. (2017, July). A novel approach for feature selection using Rough Sets. In International Conference on Computer, Communications and Electronics (Comptelix), pp. 195-199. IEEE.

Motivation

- Rough Set is a mathematical tool to find patterns hidden in data with uncertainty.
- A major step for reduction of high dimension data, present in various forms, is selection of appropriate features.
- In this work we propose a new indiscernibility relation based on clusters, and compare its effectiveness with that of classical Rough Set based indiscernibility.
- In particular, we study the proposed Rough Set based scheme for feature set reduction.
- Rough-Cluster (RC) based approximate algorithms are proposed.

Motivation

- The major advantage of these algorithms over the classical method is that they work well even without data discretization.
- The accuracy, measured in terms of the proportion of correctly classified data samples, is obtained on various standard data sets.
- The results are found to be on par with those obtained through classical Rough Set based technique for the problem of feature selection.

Feature Selection

- Reduction of high dimension data, present in various forms, is performed by selection of appropriate features.
- Feature Selection is an important pre-processing step towards data mining. Feature Selection [11, 12, 13] is classified into following types [6].

- **Filter methods.** These methods, in general, use basic mean, covariance, gain ratio based approaches.
- **Wrapper methods.** These techniques [4, 9] require the use of classifiers as wrapper to improve feature selection capabilities.
- **Embedded methods.** These methods are intrinsically different as here the feature selection is embedded in the classifier itself. Some relevant examples in this regard are Quadratic Programming [18], SVM [1] among others. An embedded method selects features, and also classifies the objects simultaneously.

FEATURE SELECTION

- Over the last few decades Rough Sets [16] have been used for the problem of feature selection.
- Zhao-Tusang [21] have proposed Rough Sets based feature selection and have experimented on various UCI datasets viz. *tae*, *glass*, *wine*, *bupa* having 5, 10, 13, 6 features, respectively.
- Khan-Revett [16] have worked on *PIMA* dataset. Concern in feature selection is how to handle large sized datasets.
- In this regard, Maji [13] have proposed Rough hypercuboid based method for feature selection. The work focuses on several benchmark datasets of UCI and Kent Ridge Bio-Medical Data Set Repository. The results obtained have been found to be an improvement over classical methods.

Rough Sets

- Originated by Zdzislaw Pawlak in 1980’s.
- The methodology is concerned with the classificatory analysis of imprecise, uncertain and incomplete information expressed in terms of data acquired from experience.

Classical Rough Set based techniques suffer from several problems.

- Firstly, the problem of discretization, data need to be discretized before application of Rough Set based algorithms.
- Secondly, each discretization process brings in some unintended errors. Further, it will have as associated overhead of complexity of discretization.
- Thirdly, the indiscernibility relation may become sparse as the number of attributes increases.

PROPOSED ROUGH SET TECHNIQUE

- Two techniques have been proposed for Rough Set based feature selection, namely,
  - *Rough-Cluster K means* (*RCK*),
  - *Rough-Cluster Hierarchial* (*RCH*).
- The novelty of the approach is that here equivalence classes are formed by clusters instead of attributes as in conventional Rough Set based approach.
- Advantage of using the proposed techniques is that they do not require the overhead of discretization.
- The attributes that improve the aggregate degree of dependency are chosen as they contain higher discerning ability.

PROPOSED ROUGH SET TECHNIQUE

- In this work we propose a new indiscernibility relation that partitions the objects of Universe into classes of similar objects through clustering process.
- Clustering is performed using *K* means and hierarchical clustering approach.
- The advantage is that by specifying the number of clusters the data can be partitioned into equivalence classes without the problem of sparseness.
- Euclidean distance is used for computing the similarity.
- This is in fact a partition of the universe based on cluster knowledge. Below we define the related concepts used in this work.

PROPOSED ROUGH SET TECHNIQUE

- We define the *Rough-Cluster indiscernibility* relation for a set of attributes A as: $\boldsymbol{IND^{Cluster}(A)} = \{(x,y) \in U \times U: x \in Cluster_A(i) , y \in Cluster_A(i), \text{for some } 1 \leq i \leq K\}$
- $Cluster_A(i)$ = i^{th} cluster formed by using attribute subset A of the original features. $U/^{Cluster}A$ denotes the classes of $IND^{Cluster}(A)$.
- *K* denotes the number of clusters.
- We define the Rough-Cluster based positive region mathematically as:
- $\boldsymbol{POS^{Cluster}{}_A(Q)} = \cup \{\underline{A}X: X \in U/^{Cluster}Q\}$.
- The *degree of dependency* of an attribute set A measures the importance of A in classifying objects of classes of $U/^{Cluster}Q$. We define the degree of dependency mathematically as:
- $\boldsymbol{\gamma^{Cluster}{}_A(Q)} = |POS^{Cluster}{}_A(Q)| / |U|$.

Reduct

A subset Q of P is said to be a Reduct of P if

Q is independent and IND(Q)=IND(P)

There may be more than one Reduct.

The set of all indispensable relations in P is called the Core of P

Note: CORE(P) = ∩ RED(P)

where RED(P) is the family of all reducts of P

QUICKREDUCT ALGORITHM

The QUICKREDUCT algorithm yields a reduct for a dataset without forming all subsets of C, the set of conditional attributes.

QUICKREDUCT(C,D)

Input: C, the set of all conditional attributes; D, the set of decision attributes.

Output: R, the attribute subset reduct, $R \subseteq C$

(1) $R \leftarrow \{\}$

(2) do

(3) $T \leftarrow R$

(4) for each $x \in (C - R)$

(5) if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$

(6) $T \leftarrow R \cup \{x\}$

(7) $R \leftarrow T$

(8) until $\gamma_R(D) = \gamma_C(D)$

(9) return R

QUICK_CLUSTER_REDUCT(C,D)

Input: C, the set of all conditional attributes; D, the set of decision attributes.

Output: R, the attribute subset reduct, $R \subseteq C$

(1) $R \leftarrow \{\}$

(2) do

(3) $T \leftarrow R$

(4) for each $x \in (C - R)$

(5) if $\gamma^{Cluster}_{R \cup \{x\}}(D) > \gamma^{Cluster}_{T}(D)$

(6) $T \leftarrow R \cup \{x\}$

(7) $R\_Old \leftarrow R$;

(8) $R \leftarrow T$

(9) until $\gamma^{Clustrer}_{R}(D) = \gamma^{Cluster}_{R\_OLD}(D)$

(10) return R

Rough Cluster Reduct

- The proposed algorithms for feature selection are evaluated on standard datasets.
- Accuracies are computed using SVM and KNN [2].
- The datasets used for evaluation are:
  - ***colon***,
  - ***svmguide3***,
  - ***sonar***,
  - ***leukemia***,
  - ***duke data***,
  - ***mushroom***,
  - ***Ionosphere***,
  - ***satimage*** **from UCI and LibSVM.**

Rough Cluster Reduct Experiments

- Experiments were performed with varying values of *K*, the number of clusters in the Rough-Cluster reduct algorithm.
- All experiments were done on random samples of the dataset with 5 fold cross validation.

Data set

Data Set	Number of Objects	Number of Features	Number of classes
Leukemia	72	7129	2
Colon	62	2000	2
Satimage	6435	36	6
Ionosphere	351	34	2
Wine	178	13	3
Sonar	208	60	2
Duke Data	44	7129	2
Svmguide3	1284	21	2
Lung Cancer	32	56	2

Results with SVM

DataSets	QR	RCK K=5	RCH K=5	RCK K=7	RCH K=7	RCK K=10	RCH K=10	RCK K=11	RCH K=11
Svmguide3	83.47	81.25	81.48	**83.94**	81.13	80.85	80.62	82.73	80.85
Leukemia	**86.43**	83.57	80.00	82.14	85.71	80.71	81.42	81.42	83.57
Colon	78.33	79.16	80.83	82.50	81.66	76.66	83.33	**85.83**	84.16
Satimage	89.91	**90.17**	85.93	88.31	88.56	89.31	88.01	89.98	88.09
Ionosphere	**91.71**	75.14	91.11	75.14	88.00	75.14	89.42	75.14	88.85
Wine	95.43	94.41	92.00	94.12	**96.57**	93.71	94.57	94.28	93.42
Iris	96.00	97.33	92.33	97.00	94.00	**97.66**	96.33	97.33	96.33
Lung Cancer	71.66	70.00	71.66	73.33	68.33	68.33	73.33	68.33	**78.33**
Duke	**76.25**	73.75	65.00	70.00	68.75	72.50	71.25	**76.25**	70.00

Results with KNN

DataSets	QR	RCK K=5	RCH K=5	RCK K=7	RCH K=7	RCK K=10	RCH K=10	RCK K=11	RCH K=11
Svmguide3	82.61	80.11	81.36	**84.37**	80.74	80.85	79.68	**84.37**	79.68
Leukemia	**85.71**	80.71	76.42	80.00	77.14	77.85	80.00	80.14	76.42
Colon	78.83	76.66	77.50	76.67	75.83	78.33	**82.50**	78.33	80.00
Satimage	**90.58**	88.52	86.27	88.95	86.82	86.52	87.04	87.99	87.35
Ionosphere	88.84	84.57	**89.28**	84.57	86.57	84.57	88.57	84.57	87.57
Wine	87.29	**94.00**	90.28	**94.00**	**94.00**	92.85	92.85	93.14	89.42
Iris	96.00	97.33	93.66	96.66	95.67	**97.66**	96.00	97.33	96.00
Lung Cancer	66.66	70.00	76.66	75.00	65.00	70.00	75.00	73.33	**78.33**
Duke	71.25	**78.75**	66.25	75.00	66.25	66.25	65.00	76.25	66.25

Results: Features Selected

DataSets	QR	RCK K=5	RCH K=5	RCK K=7	RCH K=7	RCK K=10	RCH K=10	RCK K=11	RCH K=11
Svmguide3	17.3	4.0	4.5	3.8	4.4	3.6	4.3	4.0	4.7
Leukemia	2.4	22.9	24.9	21.2	23.5	21.8	23.2	21.0	23.1
Colon	3.1	15.1	16.3	14.0	15.1	15.8	17.8	16.3	18.2
Satimage	36	9.3	10.0	12.8	14.0	8.6	9.5	11.0	11.7
Ionosphere	10.5	2.8	3.0	3.6	3.8	2.2	2.4	2.3	2.5
Wine	5.3	4.9	5.3	6.1	6.6	5.8	6.2	6.2	6.7
Iris	3.9	3.8	3.4	3.9	3.4	3.9	3.9	3.9	3.9
Lung Cancer	3.8	3.8	4.2	5.0	5.4	4.7	4.9	5.1	5.3
Duke	2.4	15.6	17.2	15.7	17.0	12.6	13.9	13.3	15.0

Results

- The following points briefly describe the major observations from the experiments:
- Satimage data has 36 features and QR selects all 36 features while RCK and RCH on an average selects 8 to 14 features. The accuracies vary marginally for quick reduct (89.91%) and RCK algorithms (90.17%) for SVM. While for KNN QR the accuracy obtained is 90.58% and for RCK with K=7 the obtained accuracy is 88.95%.
- Ionosphere data set has 34 features. QR selects on an average 10.5 features. However, for RCK and RCH the average number of selected features lie in the range of 2.2 to 4. The KNN classification accuracy of QR i.e. 88.84 is less than the best given by RCH method viz. 89.28. For SVM, a marginal difference in accuracy between QR (91.71) and RCH method (91.11).

Results

- For svmguide3 RCK (83.94%) performs better than QR (83.47%) for SVM classifier with a vast difference in the number of features selected.
- We observe that for lung cancer QR achieves 71.66% accuracy for SVM, while the accuracy is 78.33% for RCH, K=11. The KNN classifier accuracy is 78.33% for RCH, K=11 which is better than QR accuracy.
- The Colon data has 2000 features. The RCK for K=11 is performing best for the SVM classifier giving 85.83% accuracy while QR gives 78.33%. For KNN classifier RCH for K=10 is performing best with 82.50% accuracy and QR giving 78.83% accuracy.
- It can be seen in the results obtained are at par with the results obtained with QR method barring a few cases.

Conclusion

- In this paper we have shown that the proposed Rough Set based feature selection techniques RCK, RCH result in good accuracies when tested on standard classifiers of SVM and KNN.
- We have experimented on various cluster sizes for several data sets. The method has an advantage that it does not require discretization.
- However, one important area that needs exploration is to determine the how many clusters should be used for a particular dataset.
- At present with the proposed algorithm the number has to be given as an input.
- However, in a fully automated system we aim at removing this human intervention.

Future work

- In future we plan to improve on this shortcoming of the proposed techniques using dynamic clustering algorithms (Dynocs Algorithm).
- Moreover, in order to establish the efficacy of our algorithm the testing needs to be extended for other data sets.
- For our future experiments we target large data sets such as kdd2010 (algebra and bridge to algebra) which contains large number of features.
- Another important aspect of the future direction is to work with large number of decision classes which is missing from our current experiments.

References..

[1] Bradley PS, Mangasarian OL (1998) Feature Selection via Concave Minimization and Support Vector Machines, Machine Learning Proceedings of the Fifteenth International Conference(ICML).

[2] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST),2(3), 27.

[3] Chen D, Cui DW, Wang CX, Wang ZR (2006) A Rough Set-Based Hierarchical Clustering Algorithm for Categorical Data, International Journal of Information Technology 12:3.

[4] Grosan C, Abraham A, Chis M (2006) Swarm Intelligence in Data Mining, Swarm Intelligence in Data Mining, Studies in Computational Intelligence (SCI) 34.

[5] Höppner F, Klawonn F (2008) Clustering with Size Constraints. Computational Intelligence Paradigms:167-180.

[6] Karegowda AG, Manjunath AS, Jayaram MA (2010) Comparative study of attribute selection using gain ratio and correlation based feature selection, International Journal of Information Technology and Knowledge Management 2:271-277.

[7] Li, M., Deng, S., Feng, S., & Fan, J. (2013). Quick attribute reduction based on approximation dependency degree. Journal of Computers, 8(4), 920-928.

[8] Li, Q., Li, J. H., Liu, G. S., & Li, S. H. (2004, August). A rough set-based hybrid feature selection method for topic-specific text filtering. In Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on (Vol. 3, pp. 1464-1468). IEEE.

[9] Li, S., Wu, X., & Hu, X. (2008). Gene selection using genetic algorithm and support vectors machines. Soft computing, 12(7), 693-698.

[10] Lingras, P., & Peters, G. (2012). Applying rough set concepts to clustering. In Rough Sets: Selected Methods and Applications in Management and Engineering (pp. 23-37). Springer London.

[11] Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on knowledge and data engineering, 17(4), 491-502.

[12] Lu, Z., Qin, Z., Zhang, Y., & Fang, J. (2014). A fast feature selection approach based on rough set boundary regions. Pattern Recognition Letters,36, 81-88.

[13] Maji, P. (2014). A rough hypercuboid approach for feature selection in approximation spaces. IEEE Transactions on Knowledge and Data Engineering, 26(1), 16-29.

[14] Maji, P., & Pal, S. *K*. (2007). Rough Set Based Generalized Fuzzy-Means Algorithm and Quantitative Indices. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37(6), 1529-1540.

[15] Patra, B. *K*., & Nandi, S. (2011). Tolerance rough set theory based data summarization for clustering large datasets. In Transactions on rough sets XIV (pp. 139-158). Springer Berlin Heidelberg.

[16] Pawlak, Z. (2012). Rough sets: Theoretical aspects of reasoning about data(Vol. 9). Springer Science & Business Media.

[17] Rezaee, M. R., Goedhart, B., Lelieveldt, B. P., & Reiber, J. H. (1999). Fuzzy feature selection. Pattern Recognition, 32(12), 2011-2019.

[18] Rodriguez-Lujan, I., Huerta, R., Elkan, C., & Cruz, C. S. (2010). Quadratic programming feature selection. Journal of Machine Learning Research,11(Apr), 1491-1516.

[19] Tay, F. E., & Shen, L. (2002). Economic and financial prediction using rough sets model. European Journal of Operational Research, 141(3), 641-659.

[20] Zhao, S., & Tsang, E. C. (2008). On fuzzy approximation operators in attribute reduction with fuzzy rough sets. Information Sciences, 178(16), 3163-3176.

[21] Zhou, B., Chen, L., & Jia, X. (2014). Information Retrieval Using Rough Set Approximations. In ICTs and the Millennium Development Goals (pp. 185-197). Springer US.

Thank You