A Survey on Rough Set Theory and Their Extension For Data Mining

Jalaluddin Khan[1a], Ghufran Ahmad Khan[2b], Mohammad Shahid[3c], Jian Ping Li[1d], Asad Malik[2e], Shadma Parveen[1f]

[1]University of Electronic Science and Technology of China (UESTC), [2]Southwest Jiaotong University, [3]National Taiwan University of Science and Technology {jalal4amu@yahoo.com[1a], ghufraan.alig@gmail.com[2b], mdshahid84@gmail.com[3c], jpli2222@uestc.edu.cn[1d], asad@my.swjtu.edu.cn[2e], yourshadma@yahoo.com[1f] }

***Abstract*—Nowadays the amount of data has been huge and to extract useful information is too difficult. By the frequently research of thirty years, a new mathematical or data mining tool, the rough set theory, evolve with vague, imprecise and uncertainty information by the researcher Pawlak. Rough set theory is well known for knowledge discovery and popular for making the good decision with specific data. It is also dealing with the approximation concept for providing the decision such as acceptance and rejection. In this paper I summarized the basic concept of rough set theory, different operation with little example and the extension of rough set theory. By using extension, we can deal any proposed task in the field of Data mining.**

Keywords—Probabilistic rough set, Rough set theory, Variable Precision rough set, Decision theoretic rough set.

I. Introduction

The rough set theory concept was firstly introduced by Pawlak [1][2] to comprise the concept of vague, imprecise and uncertain information because no earlier information about rough set theory and popularity jumped to the sky in different area of research such as data mining [19], machine learning [23] intelligent data analysis [21], granular computing [20], etc. core and Reduct is also very consistent topic of the rough set theory and has depleted attention from many researchers. However, decision cost is absolutely important work in the DTRS model [13]. For noisy data Pawlak theory are so sensitive. Thus, many researches move to the probabilistic approaches into rough set theory by simply introduced the threshold value [17]. Different extension of probabilistic models has to be proposed like as, the decision theoretic rough set (DTRS) model [13][16], the variable precision rough set (VPRS) model [5] and 0.5 probabilistic rough set model [18] and etc.

The Pawlak rough set theory is used only for the categorical data and Rough set theory moves around the core and reduct which is basic concept of RST specially used for data reduction and feature selection. In this paper section 2 referred as related work, section 3 referred to basic concept, section 4 referred to extension of RST and section 4 for the conclusion..

II. Related work

The concept of rough set theory was firstly or initially introduced by Pawlak et al [1][2]. Lakshmi et al. [24] proposed model for clustering of high dimensional data and this method identifies non-redundant and interesting subspace cluster for getting better cluster result because in subspace clustering with huge number of cluster is challenging problem. Wei et al. [25] proposed new idea for combining information from multi sources, multi-modality and multi-scale from the prospective of attribute reduction, object and decision making through rough set theory. Data collected from real world have missing value so handling this issue. [26][27][28][32]proposed method for handling the missing value through rough set theory and he introduced the characteristics relation for focusing the incompletely decision table then for getting indiscernibility relation and used the completely specified table. Su et al. [29] proposed a novel approach for attribute reduction through fish swarm algorithm and rough set theory. Feature selection is another method for reducing the unnecessary features and rough set theory use for defining the importance of features. Greco et al. [33][34] also proposed new method for handling the missing value in dataset and he adopted both classical and new approaches of rough set theory because classical approach is suitable for selection of multi-attribute classification problem and new approach handles multi-attribute sorting problem and integrating both feature by eliminating the missing value in the dataset. Sallam et al. [35] also proposed new model for handling numerical missing value by dividing the data into complete and incomplete information table and use distance function between complete and incomplete information table. Baniya et al. [36] prosed method for classification of musical based emotion through rough set approach and he extracted some feature like rhythm, spectral, harmony etc. and based on these features can easily make decision for attribute selection. Nguyen et al. [37] propose new model for searching the useful document from the web and is challenging problem so using rough set theory can make decision.

III. Preliminaries

A. *Definition 1. Information and Decision System*

Basically, rough set is the extension of conventional set which is based on the theory of approximation for making the decision process. In this section we will define the all the basic of rough set theory. [2]An information system $S = (U; A; V; f)$, where U is a non-empty set of finite objects (the universe) and A is a non-empty finite set of attributes such that $f_a : U \to V_a$ for every $a \in A$. V_a is the set of values that attribute a. The decision systems, $A = \{C \cup D\}$, C is the conditional attribute and D refer to the decision attribute.

B. Definition 2. Indiscernibility Relation

[1][2][3][9] Each attribute a determines a function The indiscernibility relation I(R) can be defined by-

$$I(R) = \{(x, y) \in U \times U : f_a(x) = f_a(y), \forall a \in R\}$$

The equivalence relation is equal to the indiscernibility relation.

C. Definition 3. Rough set Membership

It is necessary to defined the membership degree because uncertainty is closely related to membership degree so rough membership degree can be defined by:

$$\mu_X^R(x) = \frac{|X \cap R(x)|}{|R(x)|}, \ \mu_X^R(x) \in [0,1] \tag{1}$$

D. Definition 4. Core and Reduct

Data collected are noisy so cleaning are required and it is done by the attribute reduction [4][6]. Reduct is nothing but is the minimal set of attributes has same property as the attribute of whole contains [11].

$$RED(A) = \{R : R \subset A, IND(R) = IND(A)\} \tag{2}$$

Core of A is the intersection of all the reduct of A:

$$CORE(A) = \cap RED(A) \tag{3}$$

E. Definition 5. Upper and Lower Approximation

[1][2][3][10][15][16]Given an information system S = (U, A), where U = $\{x_1, x_2, x_3, \dots\dots\dots, x_n\}$ be the non-empty finite set and A = $\{a_1, a_2, a_3, \dots\dots\dots, a_n\}$ is an attribute set, R is an attribute subset, i.e., R⊆A. For any subset X⊆U, upper-approximation set and lower-approximation set of X are given by,

$$\left.\begin{aligned} \overline{R}(X) &= \{x \mid x \in U \wedge [x]R \cap X \neq \Phi\} \\ \underline{R}(X) &= \{x \mid x \in U \wedge [x]R \subseteq X\} \end{aligned}\right\} \tag{4}$$

And we can say that lower approximation follows the certainly belong the object and in contrast upper approximation may refer that belongness of the object depends on the possibility in the set.

F. Definition 6. Positive, Negative and Boundary Region

Given a domain U, an equivalence relation R and a target set X, and again we defined the three-disjoint region namely as positive region, negative region and boundary region.

$POS_R(X) = \underline{R}(X)$ is called the positive region.

$NEG_R(X) = U - \overline{R}(X)$ is called the negative region.

$BND_R(X) = \overline{R}(X) - \underline{R}(X)$ is refer to as boundary region. [8][9] In other words finally, we can say directly that the positive region ($POS_R(X) = \underline{R}(X)$) is the union of all (every) equivalence classes which can definitely belong to the decision class X; and the boundary region is the union of all equivalence classes which can encourage a partial decision of the decision class X; and the negative region is the union of all equivalence classes that definitely cannot belong the decision class X.

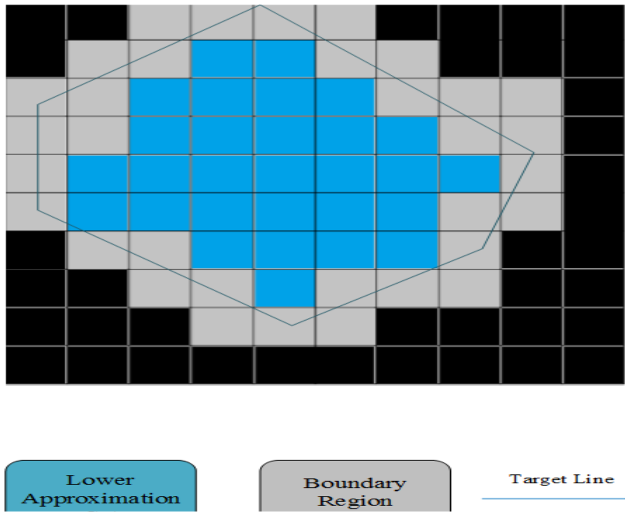

Figure 1 Approximation concept of rough set theory

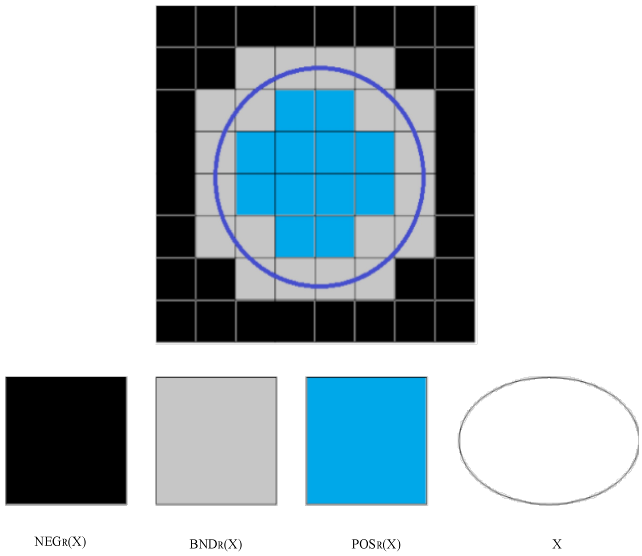

Figure 2: Positive, Negative and Boundary Region

TABLE 1: Example of Approximation (Lower and Upper)

Objects	*Ages*	*Lems*	*Walk*
y_1	10-25	40	Yes
y_2	10-25	0	No
y_3	26-40	1-20	No
y_4	26-40	1-20	Yes
y_5	41-55	21-45	No
y_6	10-25	21-45	Yes
y_7	41-55	21-45	No

Where Information System: (*U*, *A*)
U: non-empty finite set of objects
A: non-empty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$

Decision System: T = $(U, A \cup \{d\})$ and $d \notin A$ & the element of A refers to condition attributes.
Attributes = {Ages, Lems}
Objects = {*$y_1, y_2, y_3, y_4, y_5, y_6, y_7$*}
Indiscernibility:

$$IND(B) = \{(x, x^{'}) \in U^2 \mid \nabla a \in B, a(x) = a(x^{'})\}$$

The non-empty subsets of the condition attributes are *{Age}, {Lems}, {Age,Lems}*
***IND ({Age})** = {{y1,y2,y6}, {y3,y4}, {y5,y7}}*
***IND ({Lems})** = {{y1},{y2},{y3,y4},{y5,y6,y7}}*
***IND ({Age, Lems})** = {{y1}, {y2}, {y3,y4},{y5,y7}, {y6}}*

B-Lower and B-Upper approximation:

$$\underline{B}X = \{x \mid [x]_B \subseteq X\}$$

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \Phi\}$$

Let W = {x | Walk(x) = yes}

$$\underline{A}W = \{y1, y6\}$$

$$\overline{A}W = \{y1, y3, y4, y6\}$$

$$BN_A(W) = \{y3, y4\}$$

$$U - \overline{A}W = \{y2, y5, y7\}$$

Based on the above discussion the decision class, Walk is rough due to boundary region is not empty

IV. EXTENSION OF ROUGH SET MODEL

In a simple word we can say that the inclusion relation is the main point for extension of rough set model. One of the main problem with rough set model is categorization analysis so see these problem day to day problem is resolved.

A. 0.5-probabilistic rough set

Yao et. al. [16] given an information system S = (U, A) and used the threshold α 0.5, where U = $\{x_1, x_2, x_3, \ldots\ldots\ldots, x_n\}$ is a nonempty finite set, and A = $\{a_1, a_2, a_3, \ldots\ldots\ldots, a_n\}$ is an attribute set, and R is an attribute subset, i.e., R⊆A. For any subset X⊆U, the lower approximation set and Upper -approximation set of X are given by,

$$\left.\begin{aligned} \underline{R}^{(\alpha,1-\alpha)}(X) &= \{x \in U \mid P(X \mid [x]) > \alpha\} \\ \overline{R}^{(\alpha,1-\alpha)}(X) &= \{x \in U \mid P(X \mid [x]) \geq \alpha\} \end{aligned}\right\} \quad (5)$$

And the U are to be divide into three disjoint regions and represented as,

$$\left.\begin{aligned} POS_R^{(\alpha,1-\alpha)}(X) &= \{x \in U \mid P(X \| [x])\} > \alpha = \underline{R}(\alpha,1-\alpha)(X) \\ BND_R^{(\alpha,1-\alpha)}(X) &= \{x \in U \mid P(X \| [x])\} = 0.5 = \overline{R}^{(\alpha,1-\alpha)}(X) - \underline{R}(\alpha,1-\alpha)(X) \\ NEG_R^{(\alpha,1-\alpha)}(X) &= \{x \in U \mid P(X \| [x])\} < 1-\alpha = U - \overline{R}(\alpha,1-\alpha)(X) \end{aligned}\right\} \quad (6)$$

B. Probabilistic rough set

[15] Given an information system S = (U, A) with a pair of thresholds α, β $(0 \leq \beta < \alpha \leq 1)$, where U = $\{x_1, x_2, x_3, \ldots\ldots\ldots, x_n\}$ is a nonempty finite set, A = $\{a_1, a_2, a_3, \ldots\ldots\ldots, a_n\}$ is an attribute set, and R is an attribute subset (R⊆A). For any target set X⊆U, the upper-approximation set and lower-approximation set of the probabilistic rough set are defined as follow.

$$\left.\begin{aligned} \underline{R}^{(\alpha,\beta)}(X) &= \{x \in U \mid P(X \mid [x]) \geq \alpha\} \\ \overline{R}^{(\alpha,\beta)}(X) &= \{x \in U \mid P(X \mid [x]) > \beta\} \end{aligned}\right\} \quad (7)$$

And Domain U can be divided into three disjoint regions as-

$$\left.\begin{aligned} POS_R^{(\alpha,\beta)}(X) &= \{x \in U \mid P(X \| [x]) \geq \alpha\} = \underline{R}^{(\alpha,\beta)}(X) \\ BND_R^{(\alpha,\beta)}(X) &= \{x \in U \mid \beta < P(X \| [x]) < \alpha\} = \overline{R}^{(\alpha,\beta)}(X) - \underline{R}^{(\alpha,\beta)}(X) \\ NEG_R^{(\alpha,\beta)}(X) &= \{x \in U \mid P(X \| [x]) \leq \beta\} = U - \overline{R}^{(\alpha,\beta)}(X) \end{aligned}\right\} \quad (8)$$

C. Variable Precision Rough set

[5][6] The approximation space is to be divided as the prior information of Pawlak theory S = (U, R) where U is the non-empty finite universe and R is the equivalence relation on U. the equivalence relation is referred to as indiscernibility relation. And partitioning the U into set of equivalence class or elementary set as $R^* = \{E_1, E_2, \ldots\ldots\ldots E_n\}$ .So he derived the β-lower and β-upper approximation

$$\left.\begin{aligned} \underline{R}_\beta X &= \cup\{E \in R^* : c(E, X) \leq \beta\} \\ \overline{R}_\beta X &= \cup\{E \in R^* : c(E, X) < 1 - \beta\} \end{aligned}\right\} \quad (9)$$

And also derived the three regions with β threshold as-

$$\left.\begin{aligned} POSR_\beta X &= \cup\{E \in R^* : c(E, X) \leq \beta\} \\ BNR_\beta X &= \cup\{E \in R^* : \beta < c(E, X) < 1 - \beta\} \\ NEGR_\beta X &= \cup\{E \in R^* : c(E, X) \geq 1 - \beta\} \end{aligned}\right\} \quad (10)$$

D. Three-way decisions rough set

[12][13][16] DTRS Model used the Bayesian decision procedure for the probabilistic approximations. Let *Ω* = $\{\omega_1, \omega_2, \omega_3, \ldots\ldots\ldots, \omega_n\}$ be a finite set of states, and *A* = $\{a_1, a_2, a_3, \ldots\ldots\ldots, a_m\}$ a finite set of possible actions and Let $\lambda(a_i \mid \omega_j)$ define the loss for taking the a_i when the state is ω_j. The Bayesian procedure based on the conditional

Identify applicable funding agency here. If none, delete this text box.

A.

probability so Let $P(\omega_j \mid x)$ be the conditional probability of an object x in the state ω_j. The defining the expected loss accrued during the action a_i as –

$$R(a_i \mid x) = \sum_{j=1}^{s} \lambda(a_i \mid \omega_j) P(\omega_j \mid x) \quad (11)$$

In DTRSM Ω = {A, A^C} defining the set of state that object in A or not in A respectively.".

Let A= $\{a_1, a_2, a_3\}$ bet the set of action where a_1, a_2, a_3 used for classify the object for deciding the POS(A), NEG(A) and BND(A) and the probabilities $P(A \mid [x])$ and $P(A^c \mid [x])$ represent the object in it equivalent class belong to A and A^c . The expected loss $R(a_i \mid [x])$ for taking the corresponding action can be define as:

$$\left.\begin{aligned} R(a_1 \mid [x]) &= \lambda_{11} P(A \mid [x]) + \lambda_{12} P(A^c \mid [x]) \\ R(a_2 \mid [x]) &= \lambda_{21} P(A \mid [x]) + \lambda_{22} P(A^c \mid [x]) \\ R(a_3 \mid [x]) &= \lambda_{31} P(A \mid [x]) + \lambda_{32} P(A^c \mid [x]) \end{aligned}\right\} \quad (12)$$

The Bayesian Decision Procedure defined the minimum risk decision rule as follow

$$\text{(P) If} \quad \left.\begin{aligned} R(a_1 \mid [x]) &\le R(a_2 \mid [x]) \\ &\& \\ R(a_1 \mid [x]) &\le R(a_3 \mid [x]) \end{aligned}\right\} \quad \textit{Decide POS(A)}$$

$$\text{(N) If} \quad \left.\begin{aligned} R(a_2 \mid [x]) &\le R(a_1 \mid [x]) \\ &\& \\ R(a_2 \mid [x]) &\le R(a_3 \mid [x]) \end{aligned}\right\} \quad \textit{Decide NEG(A)}$$

$$\text{(B) If} \quad \left.\begin{aligned} R(a_3 \mid [x]) &\le R(a_1 \mid [x]) \\ &\& \\ R(a_3 \mid [x]) &\le R(a_2 \mid [x]) \end{aligned}\right\} \quad \textit{Decide BND(A)}$$

V. CONCLUSION

The rough set theory had trivial impacts in the area of data analysis in the worldwide and by the extension it's remains valuable in all ages for the extraction of data in terms of result produced by the proposed work of data mining. The credit goes to Pawlak who introduced first and after that rapid research is going on to improve the result. We just explain in this paper as a basic concept of the rough set, different operation can possible with the examples and showed extension of the rough set theory in a very clean manner. In which we can see Rough Set Theory is a flexible, effectively useable technique in the approximation sets, positive negative boundary region, probabilistic rough set, variable precision rough set and three-way decisions rough set to represent a vague concept.

For the future direction we would like to comparison all the attributes with the available software of rough set theory and analyze which one is better showing extracted result of the particular data.

VI. REFERENCES:

[1.] Z. Pawlak, "Rough set approach to knowledge-based decision support," Eur. J. Oper. Res., vol. 99, no. 1, pp. 48–57, 1997.

[2.] Q. Shen and R. Jensen, "Rough Sets , their Extensions and Applications," vol. 4, no. January, pp. 100–106, 2007.

[3.] Q. Zhang, Q. Xie, and G. Wang, "A survey on rough set theory and its applications," vol. 1, pp. 323–333, 2016.

[4.] X. Jia, W. Liao, et al. , Minimum cost reduction in decision-theoretic rough set models, Inf. Sci.219 (2013) 151-167

[5.] W. Ziarko, "Variable Precision Rough," 1993.

[6.] Z. Gong, Z. Shi, and H. Yao, "Variable Precision Rough Set Model for Incomplete Information Systems and Its B -Reducts," vol. 31, pp. 1385–1399, 2012.

[7.] J.P. Herbert, J.T. Yao, Game-theoretic rough sets, Fundam. Inform. 108 (3–4) (2011) 267–286.

[8.] Y.Y. Yao, B. Zhou, Naive Bayesian rough sets, in: Proceedings of RSKT, 2010, in: LNAI, vol. 6401, 2010, pp. 719–726

[9.] Y. Yao and B. Zhou, "Two Bayesian approaches to rough sets," Eur. J. Oper. Res., vol. 251, no. 3, pp. 904–917, 2016.

[10.] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341-356

[11.] R. Vashist, "An Algorithm for Finding the Reduct and Core of the Consistent Dataset," 2015.

[12.] B.Q. Hu , Three-way decisions space and three- way decisions, Inf. Sci. 281 (2014) 21–52 .

[13.] Y. Yao and J. Yao, "A Decision-Theoretic Rough Set Model," pp. 1–33, 2010.

[14.] W. Ziarko, "Probabilistic approach to rough sets," vol. 49, pp. 272–284, 2008.

[15.] Y. Y. Y. Yao, "Probabilistic Rough Set Approximations," pp. 42–46.

[16.] L. I. U. Dun, "Two Decades ' Research on Decision-theoretic Rough Sets," no. 1.

[17.] J.P. Herbert, J.T. Yao, Game-theoretic rough sets, Fundam. Inform. 108 (2011) 267–286.

[18.] Z. Pawlak, S.K.M. Wong, W. Ziarko, Rough sets: probabilistic versus deterministic approach, Inform. Sci. 29 (1988) 81–95.

[19.] Y. Peng, G. Kou, Y. Shi, Z.X. Chen, A descriptive framework for the field of data mining and knowledge discovery, Int. J. Inform. Technol. Decis. Mak. 7 (2008) 639–682.

[20.] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Systems, CRC Press, Francis Taylor, Boca Raton, 2013

[21.] T. Herawan, M.M. Deris, J.H. Abawajy, A rough set approach for selecting clustering attribute, Knowl.-Based Syst. 23 (2010) 220–231.

[22.] B. Zhou, Y.Y. Yao, J.G. Luo, A three-way decision approach to email spam filtering, in: Proceedings of the

23th Canadian Conference on Artificial Intelligence, LNAI 6085, 2010, pp. 514–522.

[23.] J.H. Zhang, Y.Y. Wang, A rough margin-based support vector machine, Inform. Sci. 178 (2008) 2204–2214.

[24.] A rough set-based subspace clustering technique for high dimensional Data

[25.] Wei, W., & Liang, J. (2019). Information fusion in rough set theory: An overview. Information Fusion, 48(January 2018), 107–118.

[26.] Grzymala-Busse, J. W. (2006). Rough set strategies to data with missing attribute values. Studies in Computational Intelligence, 9, 197–212.

[27.] Slimani, T. (2013). Application of Rough Set Theory in Data Mining. International Journal of Computer Science & Network Solutions, 1(3), 1–13.

[28.] Devi, G. L., Sujatha, M., & Jayasuma, G. (2015). A Quick Overview of Operations and Applications in Rough Set Theory. International Conference on Intelligent Systems, Control & Manufacturing Technology, 14–19.

[29.] Zhou, J., Yang, Y., Chen, P., & Wang, G. (2006). Speech Emotilon Ruecognition Based.

[30.] Su, Y., & Guo, J. (2017). A novel strategy for minimum attribute reduction based on rough set theory and fish swarm algorithm. Computational Intelligence and Neuroscience, 2017.

[31.] Lingras, P., & Peters, G. (2012). Rough Sets: Selected Methods and Applications in Management and Engineering. https://doi.org/10.1007/978-1-4471-2760-4

[32.] Zhang, M., & Yao, J. T. (2004). A rough sets-based approach to feature selection. IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS ’04., 434–439 Vol.1.

[33.] Greco, S., Matarazzo, B., & Slowinski, R. (1999). Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1711(January), 146–157.

[34.] Panda, B. S., Gantayat, S. S., & Misra, A. (2016). Retrieving the Missing Information from Information Systems Using Rough Set , Covering Based Rough Set and Soft Set, *7*(3), 1403–1407

[35.] Sallam, E., Medhat, T., Ghanem, A., & Ali, M. E. (2017). Handling Numerical Missing Values Via Rough Sets, (April), 22–36.

[36.] Baniya, B. K., & Lee, J. (2017). Rough set-based approach for automatic emotion classification of music. Journal of Information Processing Systems, 13(2), 400–416.

[37.] Nguyen, H. S., & Ho, T. B. (2008). Rough Document Clustering and the Internet. *Handbook of Granular Computing*, 987–1003.