International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Rough Set Theory in Intelligent Information Retrieval: A Comprehensive Survey

[1]*Anil Sharma, [2]Suresh Kumar

ABSTRACT *— Today online resources are the largest pool of information in terms of volume, but still users are struggling to get relevant information. Information retrieval systems suffers mainly due to two reasons: first, information overload problem, and second, vagueness and imprecision prevailing in document representations as well as in the information need description by the users. The ability to handle vague, incomplete and imprecise information laid the foundation for applying rough set theory in various domains related to artificial intelligence including information retrieval. In this survey, we have focused on rough set and its generalization models applied in information retrieval. Some existing surveys tried to comprehend rough set based information retrieval models but were restricted in their scope. This study provides systematic and comprehensive elucidation of different rough set based information retrieval models, their basic approaches, key features, strengths and limitations. A comparison of reviewed frameworks is also included for critical analysis.*

Keywords— Rough Set Theory, Rough Extension models, Information Retrieval, Equivalence Relation, Rough Approximations.

I. INTRODUCTION

Rough Set Theory (RST) is a computational intelligence technique for handling vagueness and uncertainty in information [1]. RST is used to approximate definite and possible sets from given data. The concept of indiscernibility is central to RST. Two objects are considered as indiscernible if one cannot distinguish between two objects based on given set of attributes. This indiscernibility relation is represented by a binary equivalence relation on the universe. RST proposed by Pawlak [1] is not related to degree of belongingness but vagueness. This concept is based in boundary region (BoR) of a set. Rough set (RS) has non-vacant boundary region that signifies partial knowledge about the set.

A rough set is approximated using two precise sets known as lower and upper approximations. These rough approximation sets partition the universe of objects into pairwise disjoint regions [2, 3] namely positive region, boundary region and negative region. An empty BoR of a set indicates a precisely defined set, whereas a non-empty BoR signifies an imprecisely defined set due to insufficient knowledge.

---

[1]*University School of Information, Communication and Technology, Guru Gobind Singh Indraprastha University, Delhi, India,anilsharma@aiactr.ac.in*

[2] *Department of Computer Science & Engineering, Ambedkar Institute of Advanced Communication Technologies & Research, Delhi, India, drsureshpoonia@gmail.com*

With the availability of information systems in every field, Information Retrieval (IR) has emerged as an area of interest for researchers. Despite of decades of research on the IR, no single proposal has emerged as winner. This indicates the criticality of topic to be taken for further investigations. IR systems suffers mainly due to two reasons: first, information overload problem and second, vague and imprecise specification of information needs by the users. The ability to handle incomplete, vague and imprecise information has laid the foundation for application of RST in various domains such as pattern recognition, feature selection, information retrieval, neural computing, data mining, conflict analysis, machine learning and so on [4 -10].

1.1 Motivation of the survey

Since the inception of RST, there has been an extensive research on the application of RST in information retrieval. But the literature shows the publications of only a few surveys [11, 12, 13, 14, 15] on the topic. These surveys contribute considerable knowledge in the field but were not thoroughgoing and seems restricted in some aspect.

[11] focused on fuzzy sets and rough sets in information retrieval. Author further discussed the concept of clustering and hierarchical classifications in the context of IR. But inclusion of very few numbers of frameworks in this survey limited its scope. [12] concentrated on IR models based on rough set theory. Author thrown light upon concept of rough set theory and traditional IR models also. Failure to include comparative analysis of cited frameworks and lack of enough number of frameworks are the limitations of this survey. [13] discussed soft-computing based intelligent IR models. Further, application of probability theory, fuzzy sets, genetic algorithm and artificial neural networks in context of soft web mining were also included. Again, surveys were confined in terms of challenges and research gaps. [14] investigated the application of RST in IR. In this survey, research gaps and key features of cited frameworks were discussed, but limited number of proposals were included that refrained the survey from exhibiting describe state-of-art. In [15] author presented basic notion and features of RST. Further extension models of RS, their applications, RST application tools and key issues in applying RST to different domains were discussed.

It is summarized from above discussion that none of the cited proposals exhibit the thorough coverage of RST in IR. This inspired us to present a thoroughgoing and well-structured survey on rough set based intelligent IR models which includes the description of rough set extension models, their taxonomy, limitations and challenges. Table 1 summarizes and reflect comparison of presented proposal with referred work in the context of attributes such as taxonomy, comparative analysis, tabular representation, graphical representation and research directions, where (✓) shows inclusion and (✗) indicates non-inclusion of above attributes in proposal.

Table 1: Present proposal compared with the cited works

Proposals	Models Taxonomy	Comparative Analysis of Frameworks	Tabular Representation of Results	Graphical Representation of Results	Research Directions
S. Miyamoto [11]	✗	✗	✗	✗	✗

J. Hua [12]	×	×	×	×	×
Ahmed and Ansari [13]	×	✓	×	×	✓
B. Zhou [14]	×	×	×	×	×
Zhang et al. [15]	×	✓	×	×	✓
Present survey	✓	✓	✓	✓	✓

1.2 Scope and organization of the survey

The present proposal includes:

- Discussion on rough set theory and its various extension models.
- Tabular representation of Rough set-based IR models for easy grasping of facts.
- Comparison of our proposal with cited surveys in terms of methodology, approach, key features and limitations.

Remainder of the paper is organized as follows: Next section discusses rough set framework. Section 3 throws light on rough set extension models, in which we focused mainly on probabilistic rough set model, decision theoretic rough set model (DTRS), variable precision rough set model (VPRS), Bayesian rough set model (BRSM), fuzzy rough set model (FRSM) and tolerance based rough set model (TRSM). Section 4 explores rough set theory in IR. Section 5 is about discussions and analysis, while Section 6 throws light on conclusions and research directions.

II. ROUGH SET FRAMEWORK

In RST, a non-empty finite set of objects is called universe U. Let R is indiscernibility relation on the universe U and approximation space P_A is represented as (U, R). Essentially, R is equivalence relation that induces partitions of the universe into granules of knowledge knowns as equivalence classes denoted as U/R. The granules containing element x is given by [x] = {y ∈ U | xRy}. The lower and upper approximations of X ⊆ U are given as [1, 2]:

The lower approximations of set X with respect to (wrt) R is the set of all objects which can be for certain (surely) classified as X wrt R.

$$\underline{P}_A(X) = \{x \in U \mid [x] \subseteq X\}$$

The upper approximations of set X with respect to R is the set of all objects which can be possibly (not certainly) classified as X wrt R.

$$\overline{P}_A(X) = \{x \in U \mid [x] \cap X \neq \phi\}$$

The boundary region (BoR_A) of set X with respect to R is the set of all objects which can be classified neither as X nor as ~X (where ~X represents complement of X).

$$BoR_A(X) = \underline{P}_A(X) - \overline{P}_A(X)$$

The pair $(\underline{P}_A(X), \overline{P}_A(X))$ represents rough set. Diagram 1 demonstrate the rough approximations of a set X in classical rough set (RS). Each box in the diagram represents granules of knowledge denoted as E, when relation R induces partitions of universe U. The set X is represented by black color oval, where set in dark green color represents lower approximation of X denote by $\underline{P}_A(X)$. While set in light green color (outside the black oval) signifies upper approximation of X denoted by $\overline{P}_A(X)$ and set in light green color (inside the black oval) shows boundary region denoted by $BoR_A(X)$.

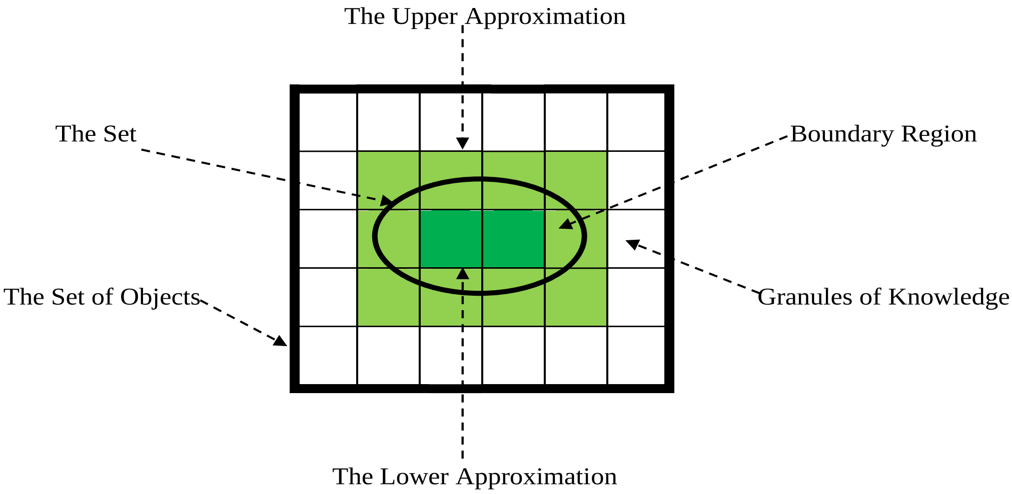


Figure 1: Pawlak's Rough Set

The universe U can be partitioned into three disjoint regions based on rough approximations of X: $POS_A(X)$ the positive region, $BoR_A(X)$ the boundary region, and $NEG_A(X)$ the negative region [3] as:

$$POS_A(X) = \underline{P}_A(X)$$
$$BoR_A(X) = \underline{P}_A(X) - \overline{P}_A(X)$$
$$NEG_A(X) = U - POS_A(X) \cup BoR_A(X) = U - \overline{P}_A(X)$$

If any object is member of target set X, it will be definitely categorized into positive region $POS_A(X)$ of set X. If any object is not a member of set X, it would be definitely categorized into negative region $NEG_A(X)$ of target set X. If we are not certain about belongingness of an object wrt target set X, then it would fall into boundary region $BoR_A(X)$ of target set X.

Limitations: The classical RS was not able to handle certain degree of overlap between set of objects belonging to different equivalence classes. The lack of error tolerance for misclassification of data and a controlled degree of uncertainty remained one of restrictions of classical RS. Further, classical RS depends upon discrete data for approximations and handling of real valued data was outside the realm of this approach, which was another major drawback of classical rough set. To overcome these limitations of classical rough set, many generalization models of classical rough set were developed, which will be in the next section.

III. ROUGH SET EXTENSION MODELS

Motivated by success of classical rough set in data analysis applications, various extensions models [16, 17, 18] of classical rough set were proposed mainly probabilistic rough set model [17, 18], decision theoretic rough set model [19, 20, 21], variable precision rough set model [22, 23, 24], Bayesian rough set model [25, 26, 27], tolerance rough set model [28] and fuzzy rough set model [29,30]. Diagram 2 depicts some of extensions models of classical rough set as shown below.

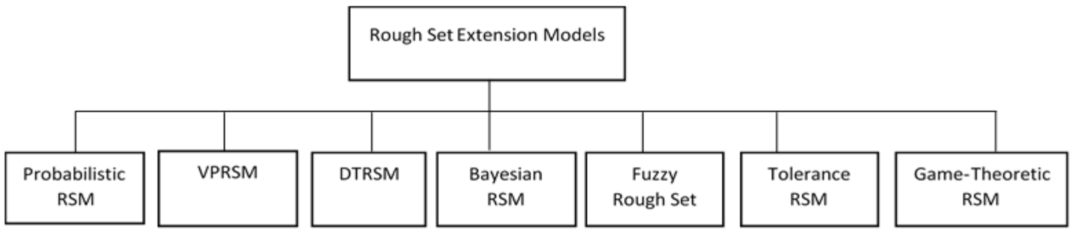


Figure 2: Rough Set Extension Models

3.1 Probabilistic Rough Set Model

To incorporate probabilistic approach in classical rough set, Wong and Ziarko [17,18] introduced probabilistic approximations in RST. Based on rough membership and rough inclusion [31], both approximations for probabilistic RS can be estimated using conditional probability. Let P (X | [x]) be conditional probability of an object belonging to target set X given that the object belongs to equivalence class [x], where P (X | [x]) can be calculated as:

$$P(X \mid [x]) = \frac{|X \cap [x]|}{|[x]|} \text{ where } |X \cap [x]| \text{ denotes cardinality of set.}$$

The three approximation regions of classical rough set: the positive region $POS_A(X)$, the boundary region $BoR_A(X)$, and the negative region $NEG_A(X)$ can be estimating using conditional probability as:

$$POS_A(X) = \{x \in U \mid P(X \mid [x]) = 1\}$$
$$NEG_A(X) = \{x \in U \mid P(X \mid [x]) = 0\}$$
$$BoR_A(X) = \{x \in U \mid 0 < P(X \mid [x]) < 1\}$$

Based on probability approximations and parameters calculations, three extension of probabilistic rough set were developed. These extensions are the variable precision rough set model [22, 23, 24], decision theoretic rough set model [19, 20, 21] and Bayesian rough set model [25, 26, 27].

3.2 Variable Precision Rough Set Model

Data classification was one of application area where RST was applied successfully. Sensitivity towards noisy data and uncertain information was main limitation of classical RS in handling data classification problems. Ziarko [22] proposed VPRSM to overcome these limitations of classical RS. VPRSM is an extension of classical RS that allows partial set inclusion relation, thus allowing objects to be misclassified under a predefined threshold β where $0 \leq \beta < 0.5$. In the light of threshold β, three approximation regions are:

Received: 27 Feb 2019 | Revised: 20 Mar 2019 | Accepted: 30 Apr 2020

$$POS_{A}(X) = \{x \in U \mid P(X \mid [x]) \geq 1-\beta\}$$
$$NEG_{A}(X) = \{x \in U \mid P(X \mid [x]) \leq \beta\}$$
$$BoR_{A}(X) = \{x \in U \mid \beta < P(X \mid [x]) < 1-\beta\}$$

In VPRSM, threshold parameter β is estimated by the user. Absence of efficient technique to estimate threshold parameter remained one of the challenges in VPRSM.

3.3 Decision Theoretic Rough Set Model

The Pawlak's RS was intolerant about any classification error in acceptance and rejection decisions. However, some tolerance of accuracy is allowed in probabilistic rough sets. But problem with VPRSM and other probabilistic RS extension models was lack of systematic procedure to estimate threshold parameters. This motivated Yao [19] to propose another generalization of classical RS model known as DTRS. This DTRS model uses loss function and Bayesian decision method to estimate threshold parameters; and loss function is interpreted in terms of costs and risks.

In DTRSM, a pair of probabilistic thresholds α, β were considered with the following conditions: $0 \leq \beta < \alpha \leq 1$ and $0 \leq \alpha, \beta \leq 1$. Based on these pair of thresholds, three approximation regions are:

$$POS_{(\alpha,\beta)}(X) = \{x \in U \mid P(X \mid [x]) \geq \alpha\}$$
$$NEG_{(\alpha,\beta)}(X) = \{x \in U \mid P(X \mid [x]) \leq \beta\}$$
$$BoR_{(\alpha,\beta)}(X) = \{x \in U \mid \beta < P(X \mid [x]) < \alpha\}$$

For classical RS, we can assume parameters $\alpha = 1$ and $\beta = 0$. In DTRSM, the pair of probabilistic threshold parameters are estimated using Bayesian decision and conditional probability is calculated using naïve Bayesian rough set [26].

3.4 Bayesian Rough Set Model

An extension of VRRSM [22] was developed by Slezak and Ziarko [25] that do not require threshold parameters, rather it exploit probability of occurrence of target event to approximate a set. This model is known as Bayesian rough set model (BRSM). BRSM brings RST and Bayesian reasoning together for approximating a set.

In BRSM framework, the positive region of a set X signifies the area of universe where probability of set X is higher than its prior probability. The negative region of a set X defines the area of universe where probability of set X is lower than its prior probability. The BRS boundary region of a set X defines the area of universe where probability of set X is equal to its prior probability.

$$POS_{BRS}(X) = \cup\{E: P(X \mid [x]) > P(X)\}$$
$$NEG_{BRS}(X) = \cup\{E: P(X \mid [x]) < P(X)\}$$
$$BoR_{BRS}(X) = \cup\{E: P(X \mid [x]) = P(X)\}$$

Received: 27 Feb 2019 | Revised: 20 Mar 2019 | Accepted: 30 Apr 2020

Where E represents equivalence class. Non-parametric nature of BRSM make it suitable choice for certain decision making applications that seek certainty gain based on available information.

3.5 Fuzzy Rough Set Model

Classical S was not suitable for handling data analysis on real valued data sets. Amalgamation of fuzzy set theory with RST proposed a generalization of classical RS known as fuzzy rough set model (FRSM). The hybridization of RST with fuzzy set theory resulted in two approaches [33]: first, constructive approach that uses fuzzy equivalence classes to estimate both approximation regions of a set X and second, axiomatic approach that elaborated on mathematical properties of fuzzy RS [29, 30]. FRSM approximations are given as [29]:

$$\underline{P}_A(X)(x) = \min\{\max(1 - R(x, y), X(y)) : y \in U\}$$

$$\overline{P}_A(X)(x) = \max\{\min(R(x, y), X(y)) : y \in U\}$$

Where R is a fuzzy equivalence relation and X is target set. The pair $(\underline{P}_A(X), \overline{P}_A(X))$ is called fuzzy RS.

3.6 Tolerance Rough Set Model (TRSM)

The basic granules of knowledge in classical RS are equivalence relation, which is reflexive, symmetric and transitive in nature. But in TRSM, another relation called tolerance relation (reflexive, symmetric but not transitive) is employed which produces overlapping tolerance classes [28]. Both approximations of a set X in TRSM are estimated as:

$$\underline{P}_{TRSM}(X) = \{x \in U \mid T_R(x) \subseteq X\}$$

$$\overline{P}_{TRSM}(X) = \{x \in U \mid T_R(x) \cap X \neq \phi\}$$

Where $T_R(x)$ is tolerance class of element x. The pair $(\underline{P}_T(X), \overline{P}_T(X))$ denotes tolerance RS.

IV. ROUGH SET THEORY IN INFORMATION RETRIEVAL

Dynamic nature of web makes web-based IR systems different from traditional IR systems in terms of knowledge representation, indexing, query expansion and interpretation, retrieving relevant documents, raking and presentation of resultant web pages. Information retrieval models are broadly divided into two categories. First, traditional IR models were based on keyword search and were mainly dependent upon syntactics of search terms. These systems suffered mainly due to two reasons: first, problem of synonyms and polysemy; and second, lack of standards for information representation. Semantics of search terms were ignored in traditional search methods as they focused on syntactic properties of search words.

To handle the issue of vagueness and imprecision, number of proposals for incorporating soft computing techniques in IR were made. Rough set theory backed by robust theoretical foundation is a successful approach to imperfect knowledge. The ability to handle uncertainty in data, capability to model query and documents, estimation of threshold parameters from given data are some of reason behind application of rough set theory in IR. The following selected framework throw light on rough set based IR models in literature.

The classical RS is successfully applied in IR by many researchers [35, 36, 37, 38, 39, 40, 41, 42]. These proposals were based on rough set approximations, rough relations and indiscernibility (equivalence) relation. Additionally, user's relevance feedback was employed to enhance the effectiveness of these IR models [35, 36, 38]. Some authors also introduced formal concept analysis and ontology with rough set approximations in IR proposals. There have been some proposals of introducing generalization of RS models in IR [43, 44, 45, 46]. These proposals incorporated probabilistic approach in classical RS by allowing partial set inclusion relation, thus permitting a control degree of misclassification of data. These proposals were based on conditional probability and threshold parameters; and users were expected to provide threshold parameters which were critical for performance of IR systems.

Some authors proposed fuzzy logic and rough set based IR models [47, 48, 49, 50] by assigning fuzzy weights to documents/queries. Rough set approximations were performed on these weighted documents/queries using fuzzy relations. Tolerance rough set allow overlapping of classes by replacing equivalence relation in classical RS with tolerance relation. In some applications like IR, the transitivity property is not always satisfied by a relation. Literature shows proposals applying tolerance rough set in information retrieval [51, 52, 53, 54, 55]. A detailed analysis and comparison of these cited proposals are discussed in next section.

V. DISCUSSION AND ANALYSIS

This survey presents a comparison of rough set models applied in intelligent IR and their limitations in the context of web search. A comparative analysis of framework surveyed is presented in Table 2. In the survey, we observed how different techniques are employed by these models. The Diagram 4 (a) shows that some of features employed in rough set based IR models are equivalence relations, tolerance relations, classical rough set, rough set extensions, user's relevance feedback and fuzzy hybridization. Furthermore, it can be noticed from Diagram 4 (b) that 29% surveyed models employed equivalence relations whereas 10% used tolerance relations. We observed that 25% preferred rough set extension models while 14% frameworks utilized classical rough set. User's feedback and fuzzy hybridization was used by 10% and 12% frameworks respectively.

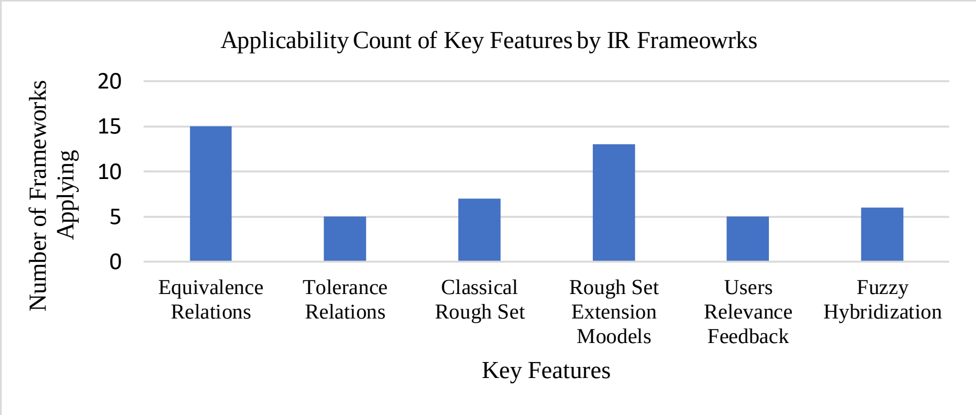
Table 2: Comparative Analysis of Semantic Web-Based Information Retrieval Models

S. No.	Model	Techniques	Strength	Limitations
1	Machine learning approach to information retrieval [35]	Probabilistic classification, RSA, Adaptive learning algorithm, user's relevance feedback	Problem of term independencies is removed in proposed model	Model does not support degree of relevance in term weights
2	Non-pattern matching intelligent IR	RSA, equivalence classes, Rough relations (equality,	Rough comparison between query and web	few rank levels for searched results were produced, intolerance to

	method based on RSA [36]	inclusion and overlap), relevance feedback	documents for IR where exact match is not found	misclassification of index terms
3	Information retrieval method based on classical rough set [37]	Rough relations, hierarchical retrieval strategies (trivial strategies, direct comparison strategies, upward completion strategies	Inexact match between user's query and web documents (in the absence of exact match) based on rough approximations	Lack of effective method to distinguish between web documents with identical rank level in result set
4	Self-adaptive search engine based on rough sets [38]	RSA, users feedback information, thesaurus relations	Customized search results based on user's feedback	Storage space for feedback information repository and search speed are main challenges
5	Rough set based manufacturing process document retrieval [39]	Rough set based classification rules, premise terms, VSM document-term weight	Enriched documents representation that capture notion of 'premise terms'	The assumption that all key terms in query are equally important may not be true in real world applications
6	Rough set based reasoning and sequential pattern mining IF model [40]	User profile learning method, threshold parameter, rough set decision theory, probability distribution of objects	Issues related to low frequent pattern and computations efficiency were also solved	The implicit feedback and negative feedback were not considered for user's profile learning
7	Semantic web search based on RS and FFCA [41]	Fuzzy formal concept analysis, RSA, Concept Lattice, automatic construction of ontology	Proposal that facilitates the user with maximum flexibility in selecting preferred answer using RST approximations and FFCA	Information content provides useful information regarding search concepts, that may be augmented with this approach to make system more effective
8	Semantic search method based on FCA, RST, and Wikipedia [42]	FCA, RST, concept similarity, Wikipedia	Limitations of ontology based IC computation approaches were removed by this approach	Model works well for general domains but considered less suitable for specialized domains

9	Information filtering system based DTRS [43]	DTRS, hierarchical filtering system, category level user profiles, document level queries	Query is formulated on their user's concept space rather than on the space of web resources	Lack of systematic method for automatically calculating loss function from data itself
10	Probabilistic rough set approach in IR [44]	Probabilistic rough set, conditional probability, equivalence classes, threshold parameters (ℓ, u)	Total inclusion relation required by classical rough set was replaced by partial inclusion relation, thus increasing effectiveness of document ranking	Lack of systematic method to calculate threshold parameters (ℓ, u)
11	Variable precision rough set model based IR [45]	VPRSM, rough sets and fuzzy sets, conditional probability, Cosine similarity, Rough relations	Rough match allowed with 50% query and index terms in common (partial set inclusion)	The lack of systematic method to calculate automatically values of threshold parameters (ℓ, u)
12	Personalized web retrieval model based on rough-fuzzy sets [46]	VPRSM, search engine, user preferences, rough similarity measures	Highly personalized search results based on user's preferences	Re-ranking of results may cause performance issue
13	Vocabulary mining framework for IR [47]	Fuzzy sets, rough sets, weighted fuzzy terms, RSA, asymmetric similarity measure using lower approximations	Weighted description of documents and query, proposal utilizes term relations other than synonymy	Calculating optimal value for threshold parameter γ for different types of vocabulary relations is a challenging task
14	Client-side document filtering system based on rough-fuzzy reasoning [48]	Rough-fuzzy reasoning scheme, user's feedback, backend search engine	Proposed technique is more powerful than computing term frequency and inverse document frequency techniques	Automatically document classification using RS may be proposed
15	Query refinement scheme based on fuzzy RS [49]	Fuzzy RS, lower approximation of upper approximation of user	Scheme works on query level rather than individual term in the query	May have performance issues with thesaurus consisting of thousands of links between terms

16	User specific IR model using fuzzy RS and Wikipedia [50]	Fuzzy RS, Wikipedia, Discretization, Fuzzy clustering	Problem with creating equivalence classes with WordNet was removed by this approach	This model needs domain ontology to be applied in specialized domain
17	IR framework using tolerance rough set model [51]	Tolerance relation, tolerance class, Overlapping classes	Calculate semantic relations between user's query and web documents using upper approximations even if web document does not share common terms	Original work on TRSM based IR model does not address term weighting that has an important role in enriching documents in text processing
18	TRSM based search results (snippets) clustering algorithm [52]	Tolerance rough set based clustering, TF-IDF weighting scheme, k-means algorithm	Tackles the problem related to poor vector representation of snippets in web search results clustering algorithms	Lack of automatically estimating value of threshold parameter θ to control word relatedness in tolerance class
19	Web search results clustering based on TRSM [53]	Rough tolerance relation, normalized goggle distance (NGD), cluster content similarity, cluster overlap techniques	The coverage of each cluster is maximized using search results clustering	Using search results from one source search engine seems limitation of this method, proposal should consider full documents rather than snippets for clustering process
20	Architecture of semantic IR system based on tolerance rough set [54]	Ontologies and Compound Analytics based search, TRSM	Existing search engine's functionality enhancement for document search	Management of hierarchy of computational tasks in response to a compound queries seems one of the challenges
21	TRSM based Semantic text retrieval for Indonesian corpus [55]	TRSM, automatically generates tolerance value and documents are represented with lexicon based method	Proposed automatic tolerance value generator algorithm mechanism for determination of threshold from a set of documents	Proposed framework should include several formats and types of documents



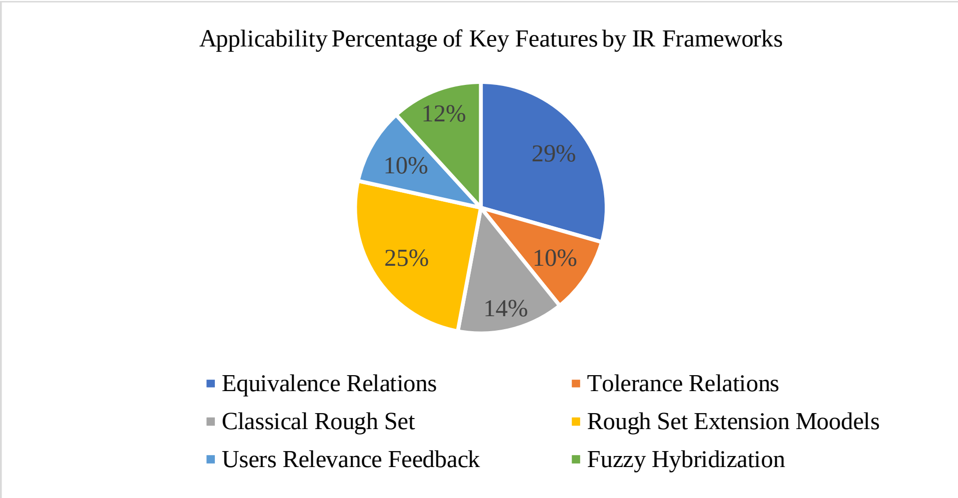


figure 4: (a) Applicability count and (b) Applicability percentage of key features by IR models

VI. CONCLUSION AND RESEARCH DIRECTIONS

IR systems struggle with issues like dealing with imprecise and vague information, lack of standards for knowledge representation and improper utilization of semantic knowledge encoded in webpages etc. This paper presents significant research work on rough set based information retrieval. In particular, the basic concepts of indiscernibility, equivalence classes, approximation space, rough approximations, and rough regions (positive, boundary and negative) were discussed. The paper also provides outline of some extension models of classical rough set including probabilistic rough set model, variable precision rough set model, decision theoretic rough set model, Bayesian rough set model, fuzzy rough set model and tolerance rough set model. These extension models enhanced the capabilities of classical rough set making it suitable for application in various branches of artificial intelligence such feature selection, conflict analysis, expert systems, image processing, information retrieval and data mining to name a few. The purpose of this study is to highlight the limitations of these rough set models and techniques employed for information retrieval. The aim of this survey includes focus on the challenges of rough set based information retrieval and identifications of related issues, those were not addressed in the existing surveys of this topic.

From the perspective of dealing with vagueness and imprecise knowledge rough set theory has emerged as winner. With present survey we are exploring capabilities of rough set in IR. From present survey it is observed that researchers are more inclined towards employing rough set extension models for IR tasks. Also, survey shows trends of using tolerance relations in place of equivalence relation and user's feedback for creating an effective IR system. Application of fuzzy hybridization for including term weights for documents and query is also observed in present survey.

REFERENCES

1. Pawlak, Z. (1982). Rough Sets. International Journal of Computer and Information Science, 11(5), 341-356.
2. Pawlak, Z. (1991). Rough Sets: Theoretical Aspects of Reasoning about Data. System Theory, Knowledge Engineering and Problem Solving, vol. 9, Kluwer Academic Publisher, Dordrecht.
3. Pawlak, Z., Skowron, A. (2007). Rudiments of rough sets. Information Sciences, 177, 3–27.
4. Chen, Y. S., Cheng, C. H.(2010). Forecasting PGR of the Financial Industry Using a Rough Sets Classifier Based on Attribute-Granularity. Knowledge and Information Systems, 25(1), 57-79.
5. Deogun, J. S., Raghavan, V. V., Sarkar, A., Sever, H. (1997). Data mining: trends in research and development. In: T.Y. Lin and Cercone, N. (Eds.) Rough Sets and Data Mining, Kluwer Academic Publishers, Boston, 9-45.
6. Qiu, G. F., Zhang, W. X., Wu, W. Z. (2005). Characterizations of attributes in generalized approximation representation spaces, LNAI, 3641, 84-93.
7. Li, Y., Zhang, C., Swan, J. R. (1999). Rough set-based model in information retrieval and filtering. In: Proceeding of the 5th International Conference on Information Systems Analysis and Synthesis, 398-403.
8. Srinivasan, P., Ruiz, M. E., Kraft, D. H., Chen, J. (2001). Vocabulary mining for information retrieval: rough sets and fuzzy sets, Information Processing and Management, 37, 15-38.
9. Yao, J. T., Zhang, M. (2005). Feature selection with adjustable criteria, LNAI, 3641, 204-213.
10. Yao, J. T., Herbert, J. P. (2007). Web-based Support Systems based on Rough Set Analysis. In: Kryszkiewicz M., Peters J. F., Rybinski H., Skowron A. (eds) Rough Sets and Intelligent Systems Paradigm, LNCS, 4585, 360-370.
11. Miyamoto, S. (1998). Application of rough sets to information retrieval. Journal of the American Society for Information Science, 49(3), 195-205.
12. Hua, J. (2009). Study on Information Retrieval Model Based on Rough Set Theory. International Symposium on Intelligent Ubiquitous Computing and Education, IEEE, 440-444.
13. Ahmed, M. W., Ansari, M. A. (2012). A Survey: Soft computing in Intelligent Information Retrieval Systems. International Conference on Computational Science and Its Applications, IEEE, 26-34.
14. Zhou, B. (2013). Applying rough set theory to information retrieval. 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). doi: https://doi.org/10.1109/ccece.2013.6567836
15. Zhang, Q., Xie, Q., Wang, G. (2016). A survey on rough set theory and its applications. CAAI Transactions on Intelligence Technology, 1, 323-333.
16. Pawlak, Z., Skowron, A. (2007). Rough sets: Some extensions. Information Sciences, 177, 28–40.
17. Yao, Y. Y. (2003). Probabilistic approaches to rough sets, Expert Systems, 20, 287–297.

18. Ziarko, W. (2008). Probabilistic approach to rough sets, International Journal of Approximate Reasoning, 49, 272–284.
19. Yao, Y.Y., Wong, S. K. M. (1992). A decision theoretic framework for approximating concepts. International Journal of Man–machine Studies, 37, 793–809.
20. Yao, Y.Y. (2003). Information granulation and approximation in a decision-theoretical model of rough sets. In: Pal, S. K., Polkowski, L., Skowron, A. (Eds.), Rough-neural Computing: Techniques for Computing with Words, Springer, Berlin, 491–518.
21. Yao, Y. Y., Wong, S. K. M., Lingras, P. (1990). A decision-theoretic rough set model. In: Ras, Z.W., Zemankova, M., Emrich, M.L. (Eds.) Methodologies for Intelligent Systems, 5, 17–24.
22. Ziarko, W. (1993).Variable precision rough set model, Journal of Computer and System Science, 46, 39–59.
23. Katzberg, J. D., Ziarko, W. (1994). Variable precision rough sets with asymmetric bounds. In: Ziarko, W. (Ed.) Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer, London, 167–177.
24. Katzberg, J. D., Ziarko, W. (1996). Variable Precision Extension of Rough Sets. Fundamenta Informaticae, 27(2-3), 155-168.
25. Slezak, D., Ziarko, W. (2002). Bayesian rough set model. In: Proceeding of international workshop on foundation of data mining (FDM'2002), Japan, 131–135.
26. Slezak, D. (2005). Rough sets and Bayes factor. In: Peter J. F., Skowron, A. (eds) Transactions on Rough Sets III, LNCS, 3400, 202–229.
27. Slezak, D., Ziarko, W. (2005). The investigation of the Bayesian rough set model. International Journal of Approximate Reasoning, 40(1-2), 81–91.
28. Skowron, A., Stepaniuk, J. (1996). Tolerance approximation spaces. Fundamenta Informaticae, 27, 245–253.
29. Dubois, D., Prade, H. (1992). Putting Rough Sets and Fuzzy Sets Together. Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht, 203–232.
30. Yao, Y.Y. (1997). Combination of Rough and Fuzzy Sets Based on α-Level Sets. In Rough Sets and Data Mining: Analysis of Imprecise Data, T. Y. Lin, N. Cereone (eds.), Kluwer Academic Publishers, 301–321.
31. Pawlak, Z., Skowron, A. (1994). Rough Membership Functions. In Advances in the Dempster-Shafer Theory of Evidence, R. Yager, M. Fedrizzi, J. Kacprzyk (eds.), Wiley, New York, 251–271.
32. Yao, Y. Y., Zhou, B. (2010). Naive Bayesian rough sets. In: Rough Set and Knowledge Technology-International Conference, RSKT 2010, Beijing, China, 719-726.
33. Yeung, D. S., Chen, D. G., Tsang, E. C. C., Lee, J. W. T., Wang, X. Z. (2005). On the Generalization of Fuzzy Rough Sets. IEEE Transactions on Fuzzy Systems, 13(3), 343–361.
34. Wu, W. Z., Zhang, W. X. (2004). Constructive and Axiomatic Approaches of Fuzzy Approximation Operators. Information Sciences, 159(3-4), 233–254.
35. Wong, S. K. M., Ziarko, W. (1986). A machine learning approach to information retrieval. In: Proceeding of ACM SIGIR conference, Italy, 228–233.
36. Srinivasan, P. (1989). Intelligent information retrieval using rough sets approximations. Information Processing & Management, 25(4), 347–361.

37. Green, J., Horne, N., Ortowska, E., Siemens, P. (1996). A rough set model of information retrieval. Fundamenta Informaticae, 28(3-4), 273-296.
38. Xu, B., Zhang, W., Yang, H., Chu, W.C. (2001). A rough set based self-adaptive web search engine, in: International Computer Software and Applications Conference (COMPSAC), 377–382.
39. Huang, C. C., Tseng, T. L., Chuang, H. F., Liang, H. F. (2006). Rough-set-based approach to manufacturing process document retrieval. International Journal of Production Research, 44(14), 2889–2911.
40. Zhou, X. (2008). Rough Set-based Reasoning and Pattern Mining for Information Filtering (Doctoral Dissertation). Queensland University of Technology, Brisbane, Australia.
41. Formica, A. (2012). Semantic web search based on rough sets and fuzzy formal concept analysis. Knowledge-Based Systems, 26, 40-47.
42. Jiang, Y., Yang, M. (2018). Semantic Search Exploiting Formal Concept Analysis, Rough Sets, and Wikipedia. International Journal on Semantic Web and Information Systems, 14(3), 99-109.
43. Li, Y., Zhang, C., Swan, J. R. (2000). An information filtering model on the web and its applications in JobAgent. Knowledge-Based Systems, 13, 285–296.
44. Ziarko, W., Fei, X. (2002). VPRSM approach to WEB searching. In: Proceedings of third International conference on rough sets and current trends in computing, RSCTC'02, LNCS, 2475, 514–521.
45. He, M., Feng, B. (2005). Intelligent information retrieval based on variable precision rough set model and fuzzy sets. In: Rough sets, fuzzy sets, data mining, and granular computing. RSFDGrC, LNCS, 3642, 184-192.
46. Duan, Q., Miao, D., Zhang, H., Zheng, J. (2007). Personalized Web Retrieval based on Rough-Fuzzy Method. Journal of Computational Information Systems, 3(2), 203-208.
47. Srinivasan, P., Ruiz, M. E., Kraft, D. H., Chen, J., Kundu, S. (2001). Vocabulary mining for information retrieval: rough sets and fuzzy sets. Information Processing and Management, 37(1), 15-38.
48. Singh, S., Dey, L. (2003). A rough–fuzzy document grading system for customized text information retrieval. Information Processing and Management, 25(4), 347–361.
49. De Cock, M., Cornelis, C. (2005). Fuzzy rough set-based web query expansion. In: Proceedings of rough sets and soft computing in intelligent agent and web technology. International workshop at WIIAT, 9-16.
50. Yadav, N., Chatterjee, N. (2018). Fuzzy rough set based technique for user specific information retrieval: a case study on Wikipedia data. International Journal of Rough sets and Data Analysis, 5(4), 1-16.
51. Ho, T. B., Funakoshi, K. (1990). Information retrieval using rough sets. Journal of Japanese Society for AI, 23(102), 424-433.
52. Ngo, C. L., Nguyen, H. S. (2005). A method of web search result clustering based on rough sets. In: International Conference on Web Intelligence (WI'05), 673–679.
53. Meng, X., Chen, Q. C., Wong, L. (2009). A tolerance rough set based semantic clustering method for web search results. Information Technology Journal, 8(4), 453-464.
54. Nguyen, H. S. (2013). Tolerance Rough Set Model and its Applications in Web Intelligence. In: International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), IEEE, 237-244.
55. Virginia, G., Nguyen, H. S. (2015). A Semantic Text Retrieval for Indonesian Using Tolerance Rough Sets Models. In: Peters J., Skowron, A., Slezak, D., Nguyen, H., Bazan, J. (Eds.): Transactions on Rough Sets XIX, LNCS, 8988, 138–224.