International Journal of Engineering & Technology, 7 (3.29) (2018) 543-549

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Probability relevance classification approach for service information discovery using semantic domain knowledge

V. Srilakshmi [1] *, Dr. K. Anuradha [2], Dr. C. Shoba Bindu [3]

[1] *Research Scholar, Dept. of CSE, JNTU Anantapur, Ananthapuramu*
[2] *Professor, Dept. of CSE, GRIET, Bachupally, Hyderabad*
[3] *Professor, Dept. of CSE, JNTU Anantapur, Ananthapuramu*
**Corresponding author E-mail: potlurisrilakshmi@gmail.com*

Abstract

The intense growth of information systems and domain services has made it difficult to provide accurate and relevant information in relation to queries and domain service needs. Conventional domain service categorization facilitates searching for related services and helps to determine classification with defined domain service knowledge and taxonomy, but it fails to relate the service which is conceptually related as such. The nonexistence of any automated mechanism for domain knowledge and taxonomy enhancement causing a high number of irrelevant services information discovery for a requested query. This paper proposes a Probability Relevance Classification (PRC) approach to overcoming the constraint of automatic classification and conceptual knowledge enhancement through constructing relevance domain knowledge semantically in support of Domain Ontology Model (DOM). The proposed PRC approach classifies the information in support of a customized Naive Bayes method and Semantic Terms Similarity method in association to DOM constructed. The experimental assessment of the recommended approach shows an improvement in the service classification and achieves better relevance results in related to the service query request. The classification accuracy in comparison with the existing classifiers shows an improvisation of the proposal.

Keywords: *Information Discovery; Classification; Probability Relevance; Domain Knowledge.*

1. Introduction

The emergence of internet and e-business over the web in support of information services is the next level of evolution [1 - 3] which provides the flexibility to integrate applications through standard protocols [4-5]. The classification of services utilizing a diversity of mechanisms generally facilitates the discovery of Information services. They have thousands of communities to make decisions and assign the appropriate category for a service and even the Information services have a large distribution in a different domain. This makes quite challenging for the interpretation of this information relation among this information without having a clear association and the knowledge of the multi-level hierarchical taxonomies of such information distributions [6 - 9]. As a result, a number of services related to a specific customer service request, the service might not be considered during the service discovery. It features a complete categorization of the information domain service in association to necessitate for appropriate expertise and knowledge.

The existing information discovery systems, such as search engines, return a long list of results based on keyword matching. However, users typically view only the top few documents in the long list of results that the search engine returns. To do this, the information retrieval system needs to display the most relevant documents at the top in order to improve user satisfaction with the search engine. However, without knowledge of the user context, it is difficult to do this because the "relevance" of the document depends on the user and the individual query. Even the limitation of automatic classification of information and information services erstwhile exhaustively researched in the recent decade [10-12]. The necessity for maintaining tools and methods is mounting with the constantly increasing amount of information and services accessible in "e-libraries", "web commerce", "blogs" and "forums" [13-15].

On the earlier, it may have an immense information quantity necessitate for explicit categorizing texts as be appropriate to the specific options for a "subject domain" which is moreover an exceptionally significant topic in "information processing technologies for knowledge discovery". This assignment can be made possible by given that resources for "supervising the classification" of the text quantity observed for corresponding the domain services [15 - 17]. The traditional systems are mostly based on keyword input to extract the relevant information. The obtained results have a large list of results which is very difficult to relate the requirement need. However, a document without user's knowledge of the "relevance" of the user and this task because it depends on a separate query, the user must have knowledge. They surf the web anonymously through a proxy, which most users, which is a major drawback and identities are unseen and tough to get. So, these dependencies of user input query over these systems have limitations in terms of providing the accurate information and user satisfaction.

In case of service-oriented computing [11], [18] the process of searching and recommendation has an intense problem due to highly depended on the "UDDI registries" and the search keywords to provide the search result [19], [20]. Similarly, a user's search query is usually based on the information summary needs and is also based on an approximation. Accurate matching is an inadequate way to find correct or even relevant information based on terminology in search queries. Instead, users find that users really want to understand the method and they can use the relationships between these terms. These terms, the user is only a rough approximation of the target information or may be unfamiliar with the search engine's mechanisms, especially due to ambiguity or even inaccuracy for certain types of information.

This paper intends to contribute a Probability Relevance Classification (PRC) approach using Semantic Terms Similarity method and probability measures using modified Naive Bayes method [2], [10]. It utilizes a domain ontology model (DOM) for classification and decision making. So far, ontology has appeared as the means for extracting knowledge from a past available "domains", "application", and "tasks" [13], [16]. It is mainly involved in the various development process of the classification which enhances the necessitate of "semantic-based applications" and it also provides the necessary information for representing the "domain-oriented conceptualization". By means of development of the "Semantic Web" and the occurrence of a large number of implements for "ontology construction", "management and contribution", "ontology regarding different knowledge domains" are previously presented [18]. Their participation in "machine learning-based knowledge discovery" has improved and shows potential outcomes being achieved. The initial ontology knowledge will build using known domain service documentation and then automatically updated later during runtime classification. This approach allows modification the classifier to the background acquaintance for the particular domain automatically and improves classifiers to classify unknown domain service documents more accurately and more precisely.

The following paper presents a background study in section-2, in section-3 a detail explanation of the proposed probability relevance classification approach is discussed, in section-4 discuss the experiment evaluation and results and section-5 conclude the conclusion of the paper.

2. Background study

In the past, various researchers presented the enhancement to the "role of context" in asserting different circumstance of domains in different areas such as, "artificial intelligence", "context awareness applications", and "information retrieval". There is numerous description that can put into the description of the user circumstance, but here it believes three important factors that participate an important responsibility in "personalized Web information extraction". These three self-determining but interrelated aspect will be the "short-term information requirements", such as queries of current activities or localized context, the semantic familiarity of the domain beneath exploration, and user profiles that capture long-term awareness. Each of these elements is considered a source of important evidence that is part of the knowledge that sustains the user's context ambiguity about accessing information.

The reconsider of associated work in literature concern for the accomplishments in the "domain service classification" with ontology involvedness are discussed in [21 - 23]. The argument definitions for syntactic or semantic concepts phrase varies according to the document classification methods for the classification of services [24], [25].

Sergeja Vogrincic et al. [6] proposed an approach for "multi-label classification" of domain text. The "semi-automatic ontology" is implemented so that the discovered documents can detect the most noteworthy concepts in the domain, and subsequently, these conceptions are utilized as possible topics of previously unknown text. A "Multi-label classification" operation is converted to one or more "single-label operations". The conversion method can be utilized through any classifier that can approximate the probability allocation for the class such as, "Support Vector Machines", "Decision Tree", "Naive Bayes", and "k-nearest neighbors" have been implemented to evaluate the conversion method. It has taken into explaining the relationship among dissimilar labels by directly handling the multi-label data. The "Multi-label neural network" and "multi-label k-nearest neighbor algorithms" are being utilized.

Lina Yao et al. [26] proposed an innovative approach which unifies the recommendations of "collaborative filtering" and "content-based information" services. A systematic approach that combines classic "collaborative filtering" and "content-based recommendation" employ a "three-way aspect model". It considered the similarities of semantic information services simultaneously with user ratings and presents a hybrid approach. This approach mostly depends on the semantically defined information services for the recommendation which limiting the capacity of the service's recommendation.

Hai Dong et al. [27] considered three major issues in online crawling as, precisely and efficiently service discovery, information formatting, and indexing information over the Internet with the intention of mining the service. The effort proposes a novel self-adaptive SASF crawler in the account of the above issue for the solution. Without the framework of the Web environment, in order to keep up the concert of the crawler, crawling on the semantic ontology learning methods. The existence of unsupervised learning vocabulary based on this research lies in drafting, and semantically related concepts and metadata for matching the discoveries of a hybrid algorithm.

Qinming He et al. [28], presented a framework for text classification with ontological support to structure and categorize text documents. The organization of ontology provides an inter-relation of items structure knowledge which helps to extract the required information to express the relation with the specific domains. The ontology supports organization aids to generate well-constructed outcome documents that outline a "vector spaces" with characterizing semantically. In general, ontological support is provided as an inference rule to avoid ambiguity in meaning. Two "machine learning algorithms" are executed in the designing for construction of classifiers and classifying unidentified text articles.

Jian Ma et al. [29], proposes an Ontology-based Text Mining (OTMM) method to cluster research proposals in a research funding agency. Ontology in the domain of research is created to categorize different research areas. The proposals are then classified into different disciplines and a text clustering algorithm, Self-organized Mapping, is applied to cluster the research proposals on the basis of similarity. After the grouping, the proposals are assigned to the reviewers. Hence this approach reduces the time of grouping the research proposals and assigning to the intended reviewers and promotes efficiency in proposal grouping process.

E. Alan Calvillo et al. [30] proposes a method to cluster research papers by using text clustering. The K-Means algorithm is used to implement the semi-supervised learning clusters to identify and approximate search using as per defined pattern. The limitation of this work is that it applies semi-automatic learning from a knowledge base. An automatic learning process can be applied that can enhance the search for the manipulated texts. Such techniques can be applied to the database knowledge with the help of filter, wrapper, and even ontology.

S. C. Punitha and M. Punithavalli [31], studied two approaches for text clustering and compared them. The first method is based on pattern recognition with semantic driven methods for clustering text documents. The second method is an ontology-based text clustering approach. Both algorithms are analyzed in terms of efficiency and speed of clustering. Experiments proved that both techniques were efficient in clustering process, but the performance of ontology-based approach was better in terms clustering quality, but a relatively slow speed because of more computations.

M. M. Gowthul Alam and S. Baulkani [32] proposed a web document clustering algorithm name as WDC-KABC. The process of the WDC-KABC algorithm is based on the features of "K-means" and "Artificial Bee Colony (ABC)". The proposed algorithm main objective is to cluster all relevant documents. It utilizes the ABC algorithm as global clustering optimizer and k-means as local

clustering optimizer. The analytical study of the WDC-KABC over different datasets and with the existing algorithms it shows an improvisation in the performance measures.
The classification techniques [29], [33] have shown promising performance in current research development. The information data collected for the classification of web services mostly consists of web documents hyperlinks and their actions on the web-pages elements which run the background processes. These information knowledge are quite limited for the classification of web services. In this aspect, a domain ontology knowledge of the semantic category will be very useful support. We enhance this development with ontology support in this paper aiming different service domain ontology with automatic ontology updating.

3. Proposed probability relevance classification approach

The proposed Probability Relevance Classification (PRC) approach of domain services performed categorizing in two phases as shown in Figure-1. In the first phase, we discuss the construction of domain ontology model (DOM) is performed, and in the second phase, we discuss the mechanism of classification using probability relevance using modified Naive Bayes method.

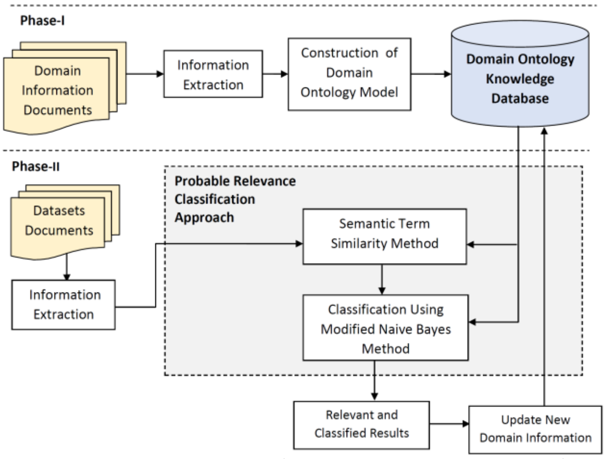

Fig. 1: Proposed Architecture of Probability Relevance Classification Approach.

The approach performs the mapping of service to a domain service ontology in a different level of dimension by using a semantic similarity and further estimates the probability of relevance for a decision of service information relevancy and classification.

3.1. Construction of domain ontology model (DOM)

It is often implemented in a logic-based ontology domain knowledge to use, and accurate classification for the existing facts [34] from the argument of semantics to web pages. Ning Zhong [35] discussed different methods of data mining, natural language understanding, and learning policy and proposed a domain ontology. Roberto Navigli et al. [36] understand the terms and relationships from the web to discover the OntoLearn development documents. Web content mining methods exist for learning the meaning of the domain-specific knowledge from text documents used by Jiang and Tan [17]. In this section, a model for construction domain ontology for Information services is introduced.

3.1.1. Information extraction

To construct organized service ontology it is important to know different domain service knowledge. This needs to cover the required subject and topic information to build a knowledge structure. To do so we collected few known information services documents categorically and extracts the available metadata information. The obtained metadata undergoes a preprocessing mechanism to remove the standard and non-relevant terms to generate the associated Information Service Terms (IST) collection. The IST collection set is the references for two or more domain service subjects which are related in some manner with each other in an ontology hierarchy.

3.1.2. Construction of DOM

The constructed DOM is performed using the IST generated from the information service documents. To initiate the construction we first identify the Parent Service (PS) category. The PS category is to be the main head of a service ontology under which Child Service (CS) category are linked, and finally, to each CS, multiple Service Terms (ST) are being linked to construct the DOM structure. The model continues the linkage of CS to PS using the IST to construct the service ontology model for the classification reference.
The construction of the DOM initiated for an Information Service (IS) category which will consist of a two dimension vector for Parent Service category as V_PS and a vector for child Service category as V_CS having key and value pair combination. These can be represented as,

$$V\_PS = [('p', 'PS\_Category_p'), ('p+1', 'PS\_Category_{(p+1)}'), \ldots, ('p+n', 'PS\_Category_{(p+n)}')x], \text{ and} \tag{1}$$

$$V\_CS(p) = [('c', 'CS\_Category_c'), ('c+1', 'CS\_Category_{(c+1)}'), \ldots, ('c+k', 'CS\_Category_{(c+k)}')x]. \tag{2}$$

The vectors are interlinked through a unique id generated for each parent service information node, to which the child information is bonded. Each child service node is linked to its associated information service terms (IST) in a two-dimensional vector as V_ST, and it can be represented as,

$$V\_ST(p, c) = [('t', 'ST\_Term_t'), ('t+1', 'ST\_Term_{(t+1)}'), \ldots, ('t+s', 'ST\_Term_{(t+s)}')x] \tag{3}$$

The above linkage of vector V_PS, V_CS and V_ST are an illustration for information service DOM structure is shown in Fig.2.

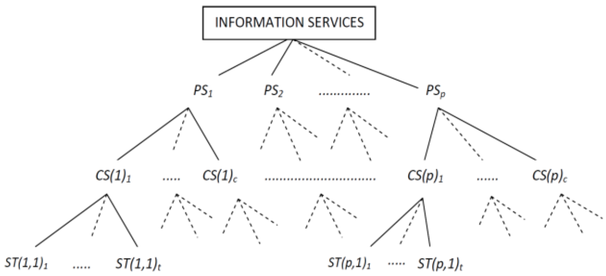

Fig. 2: An Illustration DOM Structure.

3.2. Probability relevance classification (PRC)

The Probability Relevance Classification approach implements two methods to classify the information document as, 1). Semantic Term Similarity Method, and 2). Probable Measure using Modified Naive Bayes Method. The integration model will be improvising the accuracy of classification using the defined service ontology model.

3.2.1. Semantic term similarity method

This Semantic Term Similarity Method identified the documents which are semantically related to the constructed DOM knowledge. The document which is above the relevant threshold is being considered for the service discovery classification. The "Semantic-based classification" generally performs relevance association computation [1 - 4] to relate the terms semantic relevancy. The document of a service of information, the semantic-based classification of metadata and identity of each service will compute the similarity between the terms information. To perform the association of the discovered document we derived a Semantic Term Similarity method [37]. The information of such words and phrases or metadata are considered to finds its relevance. Sometimes it may also be associated with other domain elements it may belong to.

Let's consider a domain page, Đ which consists of a set of terms belong a domain page as, Đ = {t_1, t_2, ... , i_n }. To find the semantic association we calculate the relation between the terms and the metadata of the domain page semantically using a frequent of term count procedure as, freq(t) for understanding the similarity dependency of document terms to the terms by using the equation (4).

$$freq(t) = \frac{T_n}{Z}, T \in M_k \quad (4)$$

where, T is the terms of domain page related to the terms of domain ontology M_k , and Z is the total number of extracted terms from information service documents.

In order to compute the "Semantic Term Similarity", we measure the probable similarity relevancy as P(t|s), with the information service DOM category term represent as M, against each discovered document as d, and its terms as t to classify the document service category using the equation (5).

$$sim(d{:}M_i) = Prob\left(\frac{(t \cap M_i)}{|M_i|}\right) \quad (5)$$

Let's assume an information service document, d have a phrase likes "Laptop and Mobile Sales". Using the equation(5) we can compute the probable "Semantic Term Similarity" in related to DOM, M_i as,

$$sim(d{:}M_i) = Prob\left(\frac{(Laptop, Mobile, Sales) \cap M_i)}{|M_i|}\right) \quad (6)$$

And, the computed $sim(d{:}M_i)$ are processed in a $T \times M$ similarity measure computation, where T is the terms of the document and M is the DOM for each domain as shown in Table-1.

Table 1: $T \times M$ Similarity Measure

	M_1	M_2	...	M_k	$\sum values$
t_1	sim(t_1,M_1)	sim(t_1,M_2)	...	sim(t_1,M_k)	$\sum_{n=1}^{T} sim(t_1, M_k)$
t_2	sim(t_2, M_1)	sim(t_2,M_2)	...	sim(t_1,M_k)	$\sum_{n=1}^{T} sim(t_2, M_k)$
...	...	...	...	...	...
t_n	sim(t_n, M_1)	sim(t_n M_2)	...	sim(t_n,M_k)	$\sum_{n=1}^{T} sim(t_n, M_k)$

Based on the total $\sum Value$ of the document we decided either the document need to considered for the service classification or not depend on the relevance threshold value.

3.2.2. Classification using modified naive bayes method

The functionality of a "Naive Bayes Classifiers" is completely independent of the class attribute values variations, which generally being termed as "Condition-Independence". This approach is constructed based on the assumption for the simplification and generalization of class information in relating to the "Naïve Bayes" probability assumptions. The constructed class information symbolize the character of the class attribute relation which supports classifier for the accuracy classification. It was experienced that "Bayes approach" is used in assured circumstances and it's extremely reliant on the hypothesis of the target classification information for the proficient outcomes. Because of high dependency, a little divergence in the supposition hypothesis constructs a set of inaccuracy in recognition. In the proposed classification method we modified the Naive Bayes to compute the probability of relevance in compare to the designed ontology model M as shown in Figure-3.

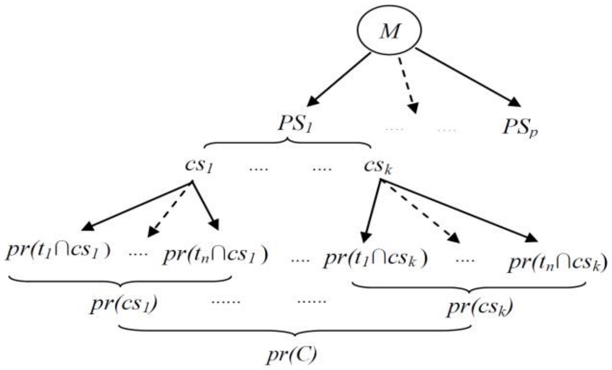

Fig. 3: Probable Relevance Classification Using Naive Bayes.

In the Figure-3, M consists of p number of PS and each PS consist of c number of CS. The t represents the terms in the service documents. The probability of relevance of terms t with CS is defined as, pr($t_{(1\rightarrow n)}$ ∩ $cs_{(1\rightarrow k)}$). Based on this pr ($cs_{1\text{-}k}$) obtained for each CS, we select the highest probable relevance CS as, pr(C) to classify the service document. The methodology of the modified Naïve Bayes is to compute the Bayes probability similarity, β relevant to DOM terms for efficiently classifying the data sets as provided in Algorithm-1.

Algorithm-1: Probable Measure using Modified Naïve Bayes Method
Input:
T [ ] → Set of Information Service Terms
Service_ontologySet[ ][ ] → n-dimensional vector of service ontology[ ]
Service_Terms[ ] → Set of Sevice Terms from Service_ontologySet[ ][ ]
For each term t_i of T []
For of each Service_ontologySet [c_i] [ ]
Service_Terms[ ] ← Service_ontologySet[x][y]
For each service term of s_i in Service_Terms [k]
Compute the Bayes probability similarity, β of
s_i in Service_Terms [k]
End for
End for
IF B >=1 THEN
$T_{I,}$ CLASSIFIED AS → C_I
END IF
END for

To evaluate the above proposal we perform an extensive analysis of more than 100 different information services pages collected from a different domain.

4. Experiment evaluation

4.1. Datasets

We are using the information services metadata to generate a synthetic data for the experimental investigation of the service ontology structure as shown in Figure-4 and also collected 100 different information services pages from different domains for test database. The domain information is collected from well-known service related to four domains as HealthCare, E-Commerce, Travelling and Manufacturing from Open Directory Project (http://www.dmoz.org/). The constructed ontology model will support service classification.

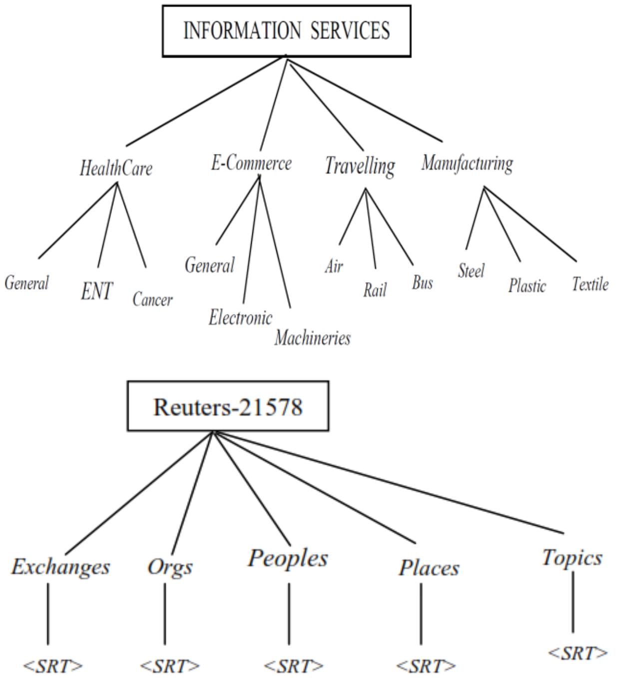

Fig. 4: DOM of Information Services and Reuters-21578 Dataset.

To extend the evaluation of a data set collection we choose a Reuters-21578 Datasets. It consists of 22 SGML-DTD data file and each file consisting of 1000 records, and each records describing data format related to six categories which are used to index the data for text categorization. The datasets provide a classification task with challenging properties having multiple categories with overlapping and not exhaustive relationships among the categories [37]. To evaluate and measure the system we implement a series of experiments to evaluate the PRC approach with synthetic and Reuters-21578 Datasets and perform the Precision and Recall calculations for the evaluation.

4.2. Results evaluation

4.2.1. Results using synthetic data

The analysis made using synthetic data using different domain queries with a variation of association threshold value from 10 to 100. The related classified information service list is shown in Table-2.

Table 2: Classified Information Services in Different Services

Parent Service Terms (PS)	Child Service Terms (CS)	Classified Information Services Results
HealthCare	General Care	http://www.carehospitals.com/
HealthCare	General Care	http://www.owaisihospital.com/
HealthCare	Cancer care	http://www.apollohospitals.com/
E-Commerce	General	http://www.quikr.com/
E-Commerce	General	http://www.ebay.in/
E-Commerce	General	http://www.jabong.com/
Travel	Air	http://www.airindia.com/
Travel	Air	http://www.jetairways.com/
Travel	Rail	http://www.indianrail.gov.in/
Travel	Bus	http://www.ksrtc.in/
Travel	Bus	http://www.redbus.in/
Manufacturing	Textiles	http://www.vtiltowels.com/
Manufacturing	Textiles	http://www.yarnsandfibers.com/
Manufacturing	Plastic	http://www.calcuttaplastics.in/

Figure-5 and 6 describe the precision and recall performance against the relevancy threshold of PRC approach. The ontology-based classification shows an improvisation precision with a low recall against different relevancy threshold. Relevancy threshold is a value which controls the filter of the relevant document for a query. The higher the threshold, the higher is the precision, which proves the improvisation of the proposal.

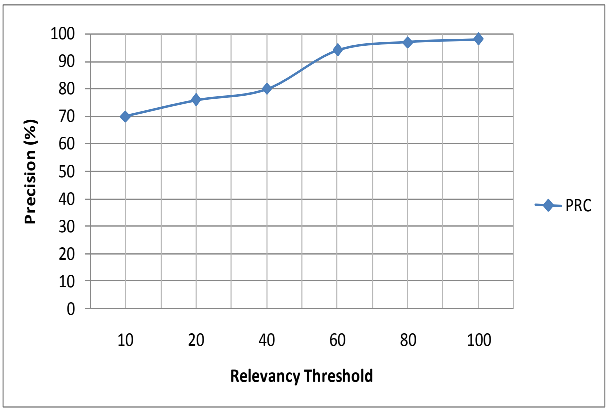

Fig. 5: Precision vs Relevancy Threshold

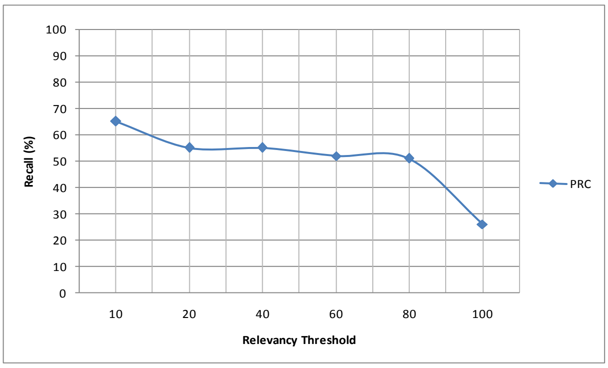

Fig. 6: Recall vs Relevancy Threshold.

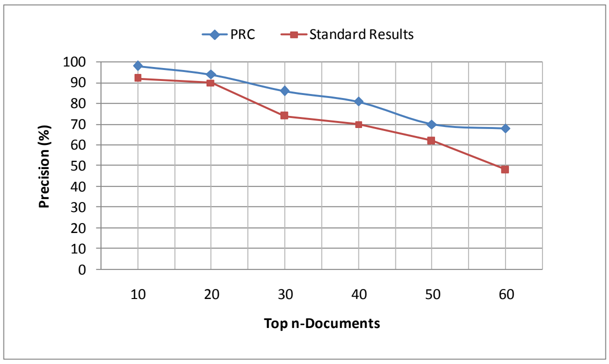

Fig. 7: Precision vs Top n-documents.

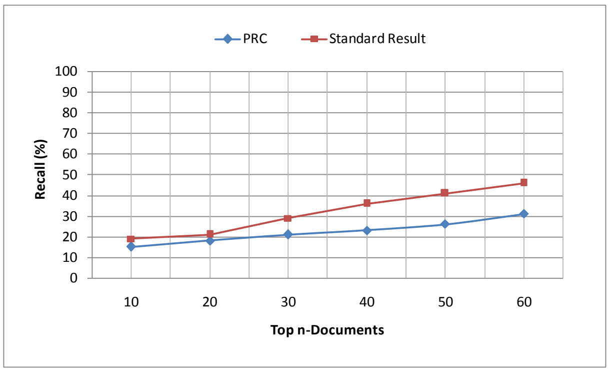

Fig. 8: Recall vs Top N-Documents.

Figure-7 and 8 describe the precision and recall measures of PRC in compare with the standard search result [30]. The standard search result means the normal result is retrieved from the search engine for the query. It shows result comparison with top n-documents retrieved for the query for the number of results obtained from 10 to 60. It shows a better precision against standard search result due to the support of DOM knowledge and also runtime updating also improvise it further.

4.2.2. Results using reuters-21578 datasets

To evaluate the extensive performance of PRC we compare the classification with existing DT Tree and NB Tree classifiers [3], [10], using Reuters-21578 Datasets. The experiment evaluation measure "True positive", "False positive", "True negative" and "False negative" to compute the performance of the classifiers as shown in Figure-9. It shows an average of 25% of improvisation in compare to the existing classifiers.

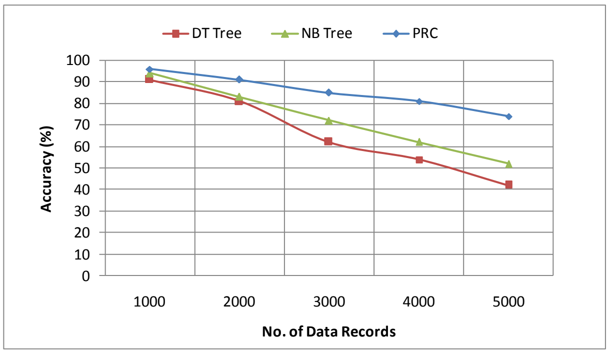

Fig.9: Classification Accuracy Comparison vs Reuters Dataset Records.

Fig-9 illustrates the classification accuracy performance at a different scale of data records. The result shows an improvement in accuracy rate in comparison to the existing classifiers due to the correctness of the class prediction due to DOM knowledge support. With the increasing number of data records, the proposed PRC construct more DOM knowledge which makes it achieve an accuracy level of 25% higher in comparing the existing classifiers suggesting the clear impact of the DOM knowledge in information classification and in the search improvisation.

To have a further comparative analysis with an existing proposal with WDC-KABC proposed in [32] to measure the classification of the relevant document. It is based on "K-means and Artificial Bee Colony (ABC)" mechanism to perform relevant document clusters. In the evaluation of Precision, Recall and Accuracy measures using Reuters-21578 Datasets it shows a satisfactory improvisation.

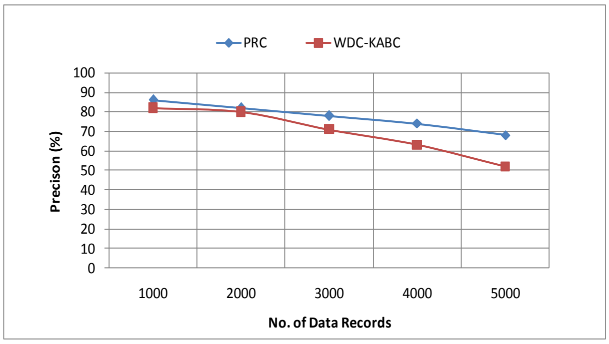

Fig. 10: Precision Comparison.

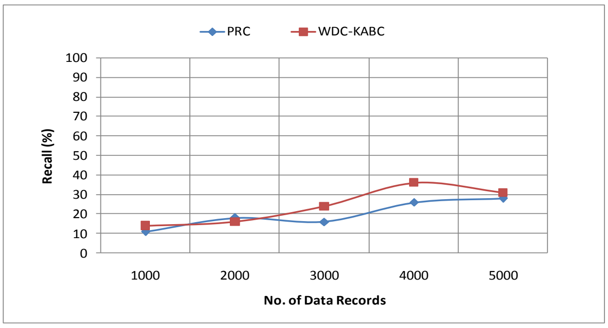

Fig. 11: Recall Comparison.

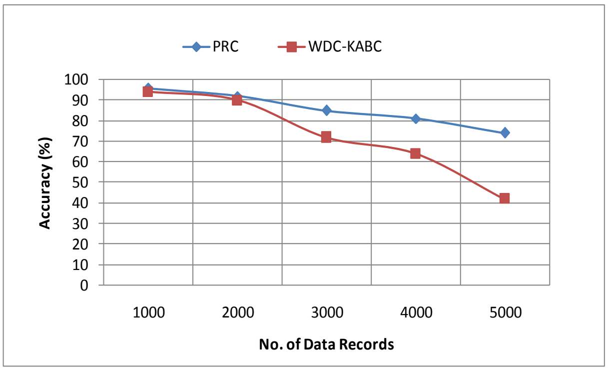

Fig. 12: Accuracy Comparison.

Fig. 10, 11 and 12 show a comparative evaluation of Precision, Recall and Accuracy measures. It depicts that with increasing number of data record the precision rate is low in case of WDC-KABC and recall rate is also higher in comparison to the PRC. The utilization of DOM knowledge helps to classifies the documents more precise and improves the precision and show low recall rate in comparison. Even in comparison to the accuracy, it shows an average of 15% higher accuracy results in comparison to the existing WDC-KABC. As the number of data records increases the number of classification feature also increases which generate ambiguity for WDC-KABC for identifying the correct class causing the reduction of accuracy in compared to PRC, as it takes the advantage of the feature in form DOM knowledge to classify document more precise and in such improve the accuracy.

5. Conclusion

This paper proposed a Probability Relevance Classification (PRC) approach to overcome the challenges and limitation of automatic classification and conceptual knowledge construction for various domains. The PRC approach classifies the information based on “Semantic Terms Similarity method" and classification depending on modified "Naive Bayes method" using Domain Ontology Model (DOM).The Semantic classification performs relevance association computation to relate the semantic relevancy. This approach accounts to find the association among the information terms of metadata and ontology service. The experimental evaluation is performed using synthetic and Reuter-21578 datasets to measure the precision, recall and accuracy rate. In comparison to the existing classification approaches , the proposed PRC outperforms with 25% higher accuracy. The contribution of DOM knowledge construction for accurate classification satisfies the required need for the future search techniques. Even the obtained precision and recall rate is satisfactory for relevancy and top-n document have shown a satisfactory improvisation. In future, it can be integrated with the various domain systems in big data framework to improve the effectiveness.

References

[1] J. Li, N. Zaman, A. Rayes, E. Custodio, "Semantics - enhanced Online Intellectual Capital Mining Service for Enterprise Customer Centers", *IEEE Transactions On Services Computing*, Vol. 10, pp. 436 - 447, 2017.

[2] M. Shirakawa, K. Nakayama, T. Hara, S. Nishio, "Wikipedia-Based Semantic Similarity Measurements for Noisy Short Texts Using Extended Naive Bayes", *IEEE Emerging Topics in Computing*, Vol. 3, No. 2, 2015.

[3] K. Belhajjame, S. M. Embury, and N. W. Paton "Verification of Semantic Web Service Annotations Using Ontology-Based Partitioning", *IEEE Transactions On Services Computing*, Vol. 7, No. 3, July-September 2014.

[4] G. X. Xu, C.-xhi Wang, "Semantic classification method for web network", *Springer - Cluster computation conference*, DOI 10.1007/s10586-017- 0742-6, 19, 2017.

[5] P. Borkowski, K. Ciesielski, "Semantic classifier approach to document classification", *CSIR Poland*, 2017.

[6] S. Vogrincic, Z. Bosnic , "Ontology-based multi-label classification of economic articles", *Computer Science and Information Systems*, Vol.8(1), pp.101-19, 2011.

[7] X. Jiang and A.-H. Tan, "Mining Ontological Knowledge from Domain-Specific Text Documents", *Proc. Fifth IEEE International Conference Data Mining (ICDM '05)*, pp. 665-668, 2005.

[8] S. Sekine and H. Suzuki, "Acquiring Ontological Knowledge from Query Logs", *Proc. 16th International Conference World Wide Web*, pp. 1223-1224, 2007.

[9] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing", *Web Intelligence and Agent Systems*, vol. 1, nos. 3/4, pp. 219-234, 2003.

[10] B. Tang, H. He, P. M. Baggenstoss, S. Kay "A Bayesian Classification Approach Using Class-Specific Features for Text Categorization", *IEEE Transactions on Knowledge and Data Engineering*, Volume: 28, Issue: six Pages: 1602 - 1606, 2016.

[11] E. Al-Masri and Q.H. Mahmoud, "Investigating Web Services on the World Wide Web", *Proc. 17th International Conference World Wide Web (WWW '08)*, Apr. 2008.

[12] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web", *Proc. 11th International Conference World Wide Web*, pp. 662-673, 2002.

[13] F.-Dombeu, J., Huisman, M., "Combining ontology development methodologies and semantic web platforms for e-government domain ontology development", *International Journal of Web & Semantic Technology*, Vol.2 (2), pp.12-25, 2011.

[14] M. Janik, K. Kochut, "Wikipedia in action: Ontological knowledge in text categorization", *IEEE International Conference on Semantic Computing*, pp.268-75, 2008.

[15] R. Gligorov, W. ten Kate, Z. Aleksovski, and F. van Harmelen, "Using Google Distance to Weight Approximate Ontology Matches", *Proc. 16th International Conference World Wide Web*, pp. 767-776, 2007.

[16] A. Rozeva, "Classification of text documents supervised by domain ontologies", *Applied Technologies & Innovations*, Vol. 8, pp.1-12, 2012.

[17] X. Jiang and A.H. Tan, "Mining Ontological Knowledge from Domain-Specific Text Documents", *Proc. Fifth IEEE International Conference Data Mining (ICDM '05)*, pp. 665-668, 2005.

[18] R. Mohanty, V. Ravi, M. Patra, "Classification of Web Services Using Bayesian Network", *In Journal of Software Engineering and Applications*, pp. 291-296, 2012.

[19] A. Sieg, B. Mobasher , R. Burke, "Web Search Personalization with Ontological User Profiles", *ACM CIKM'07, isboa, Portugal*, pp. 6-8, 2007.

[20] A. Singh and K. Nakata, "Hierarchical classification of web search results using personalized ontologies", *Proc. 3rd International Conference on Universal Access in Human-Computer Interaction*, Las Vegas, NV, 2005.

[21] X. Tao, Y. Li, N. Zhong, and R. Nayak, "Ontology Mining for Personalized Web Information Gathering", *Proc. IEEE/WIC/ACM International Conference Web Intelligence*, pp. 351-358, 2007.

[22] T. Tran, P. Cimiano, S. Rudolph, and R. Studer, "Ontology-Based Interpretation of Keywords for Semantic Search", *Proc. Sixth International Semantic Web and Second Asian Semantic Web Conference (ISWC '07/ ASWC '07)*, pp. 523-536, 2007.

[23] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs", *IEEE Trans. Knowledge and Data Engineering*, Vol. 18, no. 4, pp. 554-568, Apr. 2006.

[24] Z. Duo, L. Zi, and X. Bin, "Web service annotation using ontology mapping", *In IEEE International Workshop on Service-Oriented System Engineering*, pp. 235-242, 2005.

[25] M. Corella, P. Castells, "Semi-automatic semantic-based Web service classification", *In Springer Business Process Management Workshops*, Vol. 4103, pp. 459-470, 2006.

[26] L. Yao, Q. Z. Sheng, A. H.H. Ngu, J. Yu, and A. Segev, "Unified Collaborative and Content-Based Web Service Recommendation", *IEEE Transactions On Services Computing*, Vol. 8, No. 3, May/June 2015.

[27] H. Dong and F. K. Hussain, "Self-Adaptive Semantic Focused Crawler for Mining Services Information Discovery", *IEEE Transactions on Industrial Informatics*, Vol. 10, No. 2, May 2014.

[28] Q. He, L. Qiu, G. Zhao, S. Wang, "Text categorization based on domain ontology", *In Springer Web Information Systems - WISE 2004*, Vol.3306, pp.319-24, 2004.

[29] J. Ma, W. Xu, Y.-h.Sun, E. Turban, S. Wang, O. Liu, "An Ontology-Based Text - Mining Method to Cluster Proposals for Research Project Selection", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Vol. 42, Iss. 3, pp. 784 - 790, 2012.

[30] E. A. Calvillo, A. Padilla, J. Munoz, J. Ponce, J. T. Fernandez, "Searching research papers using clustering and text mining", *Proc. for IEEE International Conference on Electronics, Communications and Computing*, pp. 78-81, 2013.

[31] S. C. Punitha, M. Punithavalli, "Performance Evaluation of Semantic-Based and Ontology-Based Text Document Clustering Techniques", *Proc. for Elsevier International Conference on Communication Technology and System Design, Procedia Engineering, Science Direct*, pp. 100-106, 2012.

[32] M. M. Gowthul Alam, S. Baulkani, "A Hybrid Approach for Web Document Clustering Using K-means and Artificial Bee Colony Algorithm", *International Journal of Intelligent Engineering and Systems*, Vol. 9, No.4, 2016.

[33] N. Oldham, C. Thomas, A. Sheth, and K. Verma, "METEOR-S Web Service Annotation Framework with Machine Learning Classification", *In Semantic Web Services and Web Process Composition*, Vol. 3387, pp. 137-146, Jan. 2005.

[34] A. Harth, M. Janik, and S. Staab, "Semantic Web architecture", *In Springer Handbook of Semantic Web Technologies*, pp. 43-75, 2011.

[35] N. Zhong, "Representation and Construction of Ontologies for Web Intelligence", *International Journal of Foundation of Computer Science*, vol. 13, no. 4, pp. 555-570, 2002.

[36] R. Navigli, P. Velardi, and A. Gangemi, "Ontology Learning and Its Application to Automated Terminology Translation", *IEEE Intelligent Systems*, Vol. 18, No. 1, pp. 22-31, 2003.

[37] W. Zang, F. Xiao, "Using SVD on Clusters to Improve Precision of Inter-document Similarity Measure", *In Computational Intelligence and Neuroscience*, Vol. 10, 2016.