International Arab Conference on Information Technology (ACIT'2016)

An Efficient Hybrid Feature Selection Method based on Rough Set Theory for Short Text Representation

Mohammed Bekkali, Abdelmonaime Lachkar
L.I.S.A, Dept. of Electrical and Computer Engineering, ENSA, USMBA, Fez, Morocco
mohammed.bekkali2@usmba.ac.ma, abdelmonaime_lachkar@yahoo.fr

Abstract— *With the rapid development of Internet and telecommunication industries, various forms of information such as short text which plays an important role in people's daily life. These short texts suffer from curse of dimensionality due to their sparse and noisy nature. Feature selection is a good way to solve this problem. Feature selection is a process that extracts a number of feature subsets which are the most representative of the original feature set; thus it becomes an important step in improving the performance of any Text Mining task.*

In this paper, a hybrid feature selection, based on Rough Set Theory (RST) which is a mathematical tool to deal with vagueness and uncertainty and Latent Semantic Analysis (LSA) which is a theory for extracting and representing the contextual-usage meaning of words, is proposed in order to improve Arabic short text representation. The proposed method has been tested, evaluated and compared using an Arabic short text categorization system in term of the F1-measure. The experimental results show the interest of our proposition.

Keywords—*Arabic Language; short text; feature selection; Rough Set Theory; Latent Semantic Analysis*

1. Introduction

With the increasing growth of the Internet, a large number of short texts have been generated on the web and mobile applications, including search snippets, micro-blog, products review, and short messages. Short text is different from traditional documents in its Shortness and Sparseness [12][20]. As a result, short text tends to be ambiguous without enough contextual information. Most existing approaches try to enrich the representation of a short text using additional semantics. The additional semantics could be from the short text data collection itself or be derived from a much larger external knowledge base like Wikipedia, WordNet and search engines [17][19][25, 26]. The former requires shallow Natural Language Processing (NLP) techniques while the later requires a much larger and appropriate dataset [12].

In the opposite direction of enriching short text representation, the high dimension of the feature vector can deteriorate the performance of machine learning algorithm. Feature selection is a typical step in text categorization, which transforms the data representation into a shorter, more compact, and more predictive one [8]. The new space is easier to handle because of its size. In Feature selection a subset of original features is selected, and only the selected features are used for training and testing the classifiers. The removed features are not used in the computations anymore.

Several feature selection methods for short text have been conducted for English and other European languages, yet very little researches have been done out for Arabic short text. Arabic language is a highly inflected language and it requires a set of pre-processing to be manipulated, it is a Semitic language that has a very complex morphology compared with English or any other language.

In this paper, we propose a hybrid feature selection method for Arabic short text based on Rough Set Attribute Reduction (RSAR) which is a powerful tool for dealing with imprecise or incomplete information, knowledge reduction and decision rule extraction [18]; and Latent Semantic Analysis (LSA) which is an automatic method that transforms the original textual data to a smaller semantic space by taking advantage of some of the implicit higher-order structure in associations of words with text objects [3][5][10]. RSAR feature selection is applied to get optimal feature subset in order to eliminate all features which are likely to be noisy and irrelevant features, starting by building the decision table followed by reducts extraction using

QuickReduct algorithm [1][13]. Further, Reduced feature set is sent to LSA component which is used to reduce the number of rows while preserving the similarity structure among columns.

RST has been introduced by Pawlak in the early 1980s [18], it has been integrated in many Text Mining applications. Certain attributes in an information system may be redundant and can be eliminated without losing essential information. RST provide a method to determine for a given information system the most important attributes. A reduct is the essential part of an information system (related to a subset of attributes) which can discern all objects discernible by the original set of attributes of an information system [23]. On the other hand, LSA was originally proposed as an information retrieval method [5]; it has also been widely used in text categorization as well. In this theory, a statistical analysis is applied to a large corpus of text in order to provide a reduced semantic space through a matrix operation called Singular Value Decomposition (SVD).

To test the efficiency and the effectiveness of our proposed feature selection method for Arabic short text, we built a short text categorization system. Short text categorization tries to find a relation between a set of short texts and a set of categories. Machine Learning (ML) is the tool that allows deciding whether a short text belongs to a set of predefined categories [1].The obtained results illustrate the efficiency of our proposition.

The remainder parts of this paper are organized as follows: we begin with a brief survey on related work about short text feature selection in the next section. Section 3 describes the basics of RSAR and LSA. Section 4 presents our hybrid feature selection method for Arabic shot text. Dataset, experiments and results are discussed in section 5. Finally, section 6 concludes this paper and presents future work and some perspectives.

2. Related Work

Feature selection has been widely adopted for dimensionality reduction of text datasets in the past. Yiming Yang et al. [28] compare and evaluate five of these methods including, document frequency, information gain (IG), mutual information (MI), χ^2-test (CHI) and term strength (TS). They found IG and CHI are the most effective in aggressive term removal without losing categorization accuracy. Peng et al. in [8] study how to select good features according to the maximal statistical dependency criterion based on mutual information. First, they derive an equivalent form, called minimal redundancy-maximal-relevance criterion (mRMR), for first-order incremental feature selection. Then, they present a two-stage feature selection algorithm by combining mRMR and other more sophisticated feature selectors (e.g., wrappers).

Hui et al. introduce an algorithm to cluster Chinese short texts for mining web topics based on Chinese chunks. The algorithm employs N-gram feature extraction to capture Chinese chunks from texts, which reflect the text semantic structure and character dependency [9]. Mahajan et al. In [2] propose a novel feature selection method using Wavelet Packet Transform (WPT) they chooses the most discriminating features by computing inter-class distances in the transformed space.

In contrast, and in comparison with the English language, limited work in the text categorization field has been done for Arabic, and here we survey a significant selection of the recent published work in this area. For instance, [16] implements Support Vector Machines (SVMs) based text classification system for Arabic language articles. This classifier uses CHI square method as a feature selection and they have achieved practically accepted results. Chantar et al [7] propose a BPSO/KNN hybrid method, working on a training set, to output a specific subset of features. Then they evaluate this subset of features on a test set, in which they can use any machine learning/classification method for the evaluation. They conclude that the combination of BPSO and K nearest neighbour performs well in selecting good sets of features. Hawashin et al. [4] propose an improved feature selection method for Arabic text classification based on Chi-Square-based and they conclude that their method outperformed Information Gain, DF, Chi-square, Mean TF.IDF, Wrapper Approach with SVM and Best Search, and Feature Subset Selection with Best Search.

3. Feature Selection

The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. We present in the following subsections two of the most techniques used in this task and which have been used in our research.

3.1 Rough Set Attribute Reduction

Rough Set Theory has been originally developed as a tool for data analysis and classification [11][18]. It has been successfully applied in various tasks, such as

features selection/extraction, rule synthesis and classification.

Suppose that a dataset is viewed as a decision table T where attributes are columns and objects are rows. Let U denote the set of all objects in the dataset and A the set of all attributes such that $a : U \rightarrow V_a$ for every $a \in A$ where V_a is the value set for attribute a. In a decision system, A is decomposed into the set C of conditional attributes and the set D of decision attributes which are mutually exclusive and $C \cup D = A$. For any $P \subseteq A$, there is an equivalence relation I(P) as follows [23]:

$$I(P) = \{(x, y) \in U^2 \mid \forall a \in P\ a(x) = a(y)\} \quad (1)$$

If $(x, y) \in I(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility equivalence relation I(P) are denoted $[x]_P$. Given an equivalence relation I(P) for $P \subseteq C$, the lower and upper approximations $L_P(X)$ $U_p(X)$is defined for any $X \subseteq U$ as follows:

$$L_P(X) = \{x \in U \mid [x]_P \subseteq X\} \quad (2)$$
$$U_P(X) = \{x \in U \mid [x]_P \cap X \neq \Phi\} \quad (3)$$

The C-positive region of D is the set of all objects from the universe U which can be classified with certainty into classes of U/D employing attributes from C, that is,

$$POS_C(D) = \cup_{X \in U/D} L_P(X) \quad (4)$$

An attribute $c \in C$ is dispensable in a decision table T if $POS_{(C-\{c\})}(D) = POS_C(D)$; otherwise attribute c is indispensable in T . A set of attributes $R \subseteq C$ is called a reduct of C if it is a minimal attribute subset preserving the condition: $POS_R(D) = POS_C(D)$. With regard to computational complexity and memory requirements, however, the calculation of all reducts is an NP-hard task [5]. To solve this problem, we use QUICKREDUCT algorithm [14] shown below for feature selection of Web-page classification. The algorithm uses the degree of dependency γ_P (D) as follows:

$$\gamma_P(D) = \frac{\| POS_P(D) \|}{\| U \|} \quad (5)$$

as a criterion for the attribute selection as well as a stop condition. This algorithm does not always generate a minimal reduct since γ_P (D) is not a perfect heuristic. It does result in only one close-to-minimal reduct, though it is useful in greatly reducing dataset dimensionality. The average complexity of QUICKREDUCT algorithm was experimentally determined to be approximately O(n) for a dimensionality of n though the worst-case runtime complexity is O(n!) [22].

QuickReduct(C, D, R)
Input: the set C of all conditional attributes the set D of decision attributes.
Output: the reduct R of C ($R \subseteq C$)
1. $R \leftarrow \varphi$
2. **do**
3. $T \leftarrow R$
4. $\forall x \in (C - R)$
5. **if** $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
6. $T \leftarrow R \cup \{x\}$
7. $R \leftarrow T$
8. **until** $\gamma_R(D) = \gamma_C(D)$
9. **return** R

3.2 Latent Semantic Analysis

Latent Semantic Analysis models the meaning of words and documents by projecting them into a vector space of reduced dimensionality; the reduced vector space is built up by applying Singular Value Decomposition (SVD) [5].

SVD is a well-known dimensionality reduction technique. In the process of SVD, a given rectangular matrix X_{m*n} is decomposed into three matrices of special forms [6][15]:

$$X_{m*n} = U_{m*r} * S_{r*r} * V^T_{r*n} \quad (6)$$

where U and V are orthogonal matrices that contain the left and right singular vectors of X, respectively, S is the diagonal matrix that contains the singular values of X, and the subscript r denotes the number of singular values (i.e. the rank of X). If the singular values are sorted in descending order, SVD can project the data onto a lower, k-dimensional space spanned by their singular vectors corresponding to the k largest singular values [6][15]. The new decomposition becomes:

$$\tilde{X}_{m*n} = \tilde{U}_{m*k} * \tilde{S}_{k*k} * \tilde{V}^T_{k*n} \quad (7)$$

The resulting matrix $\tilde{X}$ is an approximation of X, and is of rank k [6]. It has been shown that $\tilde{X}$ is the best rank-k approximation of the original data in the least-squares sense [6][15].

In the following section, we present our hybrid feature selection for Arabic short text based on RSAR and LSA.

4. Proposed Hybrid Approach for Feature Selection

4.1 Feature Selection Algorithm

In this section, we present in detail our hybrid method for feature selection for Arabic short text. There two main component: RSAR and SVD decomposition, but first we start with the preprocessing step where each short text will be cleaned by removing Arabic stop words, Latin words and special characters like(/, #, $, etc…).

The proposed algorithm (Figure 1) for feature selection works in two steps: RSAR reduce the dimensionality of the data space by removing irrelevant features. The reduction of features is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the original. Further, reduced feature set is sent to the SVD algorithm to get Singular-value decomposition allows the arrangement of the space to reflect the major associative patterns in the data, and ignore the smaller, less important influences by constructing a semantic space and places terms and documents that are highly correlated together.

In order to test the effectiveness of our proposed method for feature selection for Arabic short text, we built an Arabic short text categorization system.

We have been used two of the most popular machine learning algorithms for Text Categorization (TC): Naïve Bayesian (NB) and the Support Vector Machine (SVM).

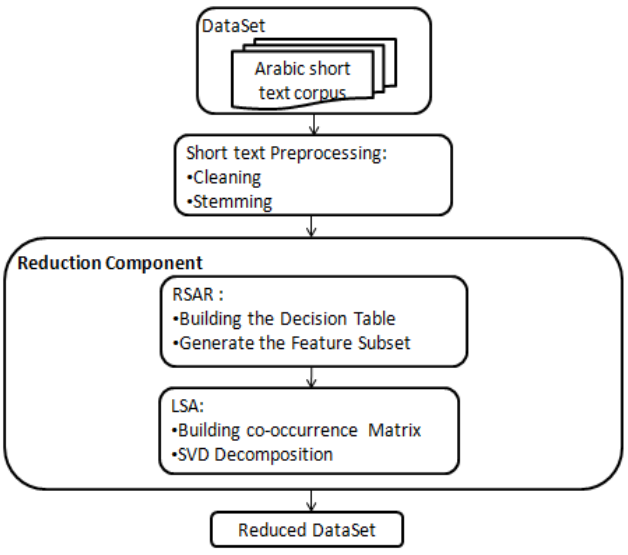


Fig. 1. Proposed system for feature selection

4.2 Machine Learning for Text Categorization

TC is the task of automatically sorting a set of documents into categories from a predefined set. This section covers two algorithms among the most used Machine Learning Algorithms for TC: NB SVM.

4.2.1 Naïve Bayesian Classifier

The NB is a simple probabilistic classifier based on applying Baye's theorem, and its powerful, easy and language independent method [21].

When the NB classifier is applied on the TC problem we use equation (8):

$$p(class \mid doc) = p(class).p(doc \mid class) / p(doc) \quad (8)$$

where:

- p (class | doc): It's the probability that a given document D belongs to a given class C
- p (doc): The probability of a document, it's a constant that can be ignored
- p (class): The probability of a class, it's calculated from the number of documents in the category divided by documents number in all categories
- p (doc | class): it's the probability of document given class, and documents can be represented by a set of words:

$$p\ (doc \mid class) = \prod_i p\ (word_i \mid class) \quad (9)$$

so:

$$p\ (class \mid doc) = p\ (class).\prod_i p\ (word_i \mid class) \quad (10)$$

where:

$p\ (word_i \mid class)$: The probability that a given word occurs in all documents of class *C*, and this can be computed as follows:

$$p(word_i \mid class) = Tct + \lambda / Nc + V \quad (11)$$

where:

- Tct: The number of times that the word occurs in that category C.
- Nc: The number of words in category C.
- V: The size of the vocabulary table.
- λ: The positive constant, usually 1, or 0.5 to avoid zero probability.

4.2.2 Support Vector Machine Classifier

SVM introduced by [24] and has been introduced in TC by [13]. Based on the structural risk minimization principle from the computational learning theory, SVM

seeks a decision surface to separate the training data points into two classes and makes decisions based on the support vectors that are selected as the only effective elements in the training set.

Given a set of N linearly separable points $N = \{x_i \in R^n \mid i = 1, 2\ldots N\}$, each point x_i belongs to one of the two classes, labeled as $y_i \in \{-1, 1\}$. A separating hyper-plane divides S into 2 sides, each side containing points with the same class label only. The separating hyper-plane can be identified by the pair (w, b) that satisfies: $w.x + b = 0$

and:

$$\begin{cases} w.x_i + b \geq +1 \text{ if } y_i = +1 \\ w.x_i + b \leq -1 \text{ if } y_i = -1 \end{cases} \quad (12)$$

For i = 1, 2… N; where the dot product operation (.) is defined by:

$$w.x = \sum w_i.x_i \quad (13)$$

For vectors w and x, thus the goal of the SVM learning is to find the Optimal Separating Hyper plane (OSH) that has the maximal margin to both sides. This can be formularized as: minimize: ½ .w.w
subject to:

$$\begin{cases} w.x_i + b \geq +1 \text{ if } y_i = +1 \text{ for } i = 1, 2,..., N \\ w.x_i + b \leq -1 \text{ if } y_i = -1 \end{cases} \quad (14)$$

Figure 2 shows the optimal separating hyper-plane.

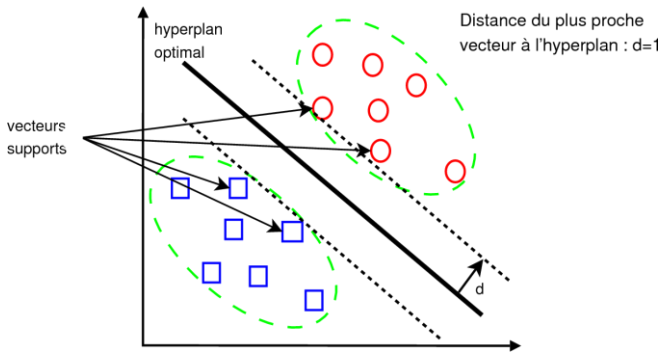


Fig 2. The optimal separating hyper-plane

The small crosses and circles represent positive and negative training examples, respectively, whereas lines represent decision surfaces. Decision surface σ_i (indicated by the thicker line) is, among those shown, the best possible one, as it is the middle element of the widest set of parallel decision surfaces (i.e., its minimum distance to any training example is maximum). Small boxes indicate the Support Vectors.

During classification, SVM makes decision based on the OSH instead of the whole training set. It simply finds out on which side of the OSH the test pattern is located. This property makes SVM highly competitive, compared with other traditional pattern recognition methods, in terms of computational efficiency and predictive accuracy [27].

The method described is applicable also to the case in which the positives and the negatives are not linearly separable. Yang and Liu experimentally compared the linear case (namely, when the assumption is made that the categories are linearly separable) with the nonlinear case on a standard benchmark, and obtained slightly better results in the former case [27].

5. Experiments Results

In order to illustrate that our proposed hybrid feature selection method can improve short text categorization. In this section a series of experiments has been conducted (Figure 3).

The dataset used is a subpart of the ODP (Open Directory Project[1]) directory which is the most widely distributed database of Web content classified by humans. The dataset is classified into seven categories (Table 1). These categories are: (commerce, study, family, international, news, health and sport.

The effectiveness of our system has been evaluated and compared in term of the F1-measure using the NB and the SVM classifiers in our Arabic short text categorization system. The chosen technique for validation is cross validation over 10 fold.

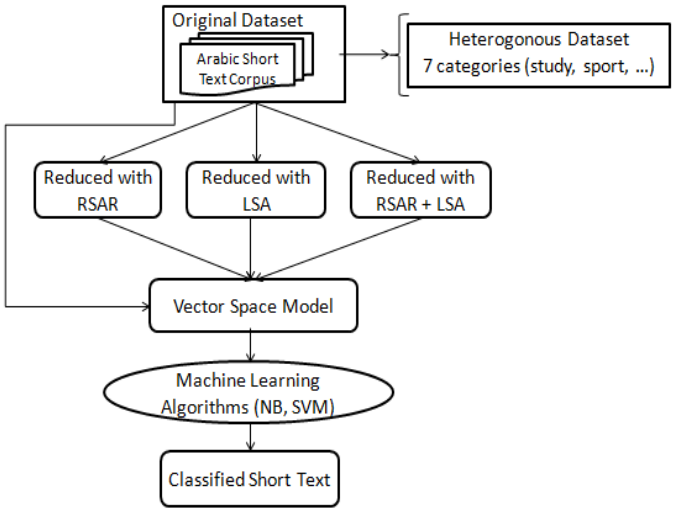


Fig 3. Proposed system for feature selection

[1] www.dmoz.org

Table 1. Dataset Description

Category	Number of texts
Commerce	242
Family	69
Health	145
International	241
News	252
Sport	105
Study	153
Total	**1207**

F1-measure (F1) can be calculated using Precision (P) and Recall (R) measures as follow:

$$F1 = 2*(P * R) / P + R \quad (15)$$

Precision and Recall both are defined as follows:

$$P = TP / TP + FP \quad (16)$$
$$R = TP / TP + FN \quad (17)$$

where:

- True Positive (TP) refers to the set of short texts which are correctly assigned to the given category.
- False Positive (FP) refers to the set of short texts which are incorrectly assigned to the category.
- False Negative (FN) refers to the set of short texts which are incorrectly not assigned to the category.

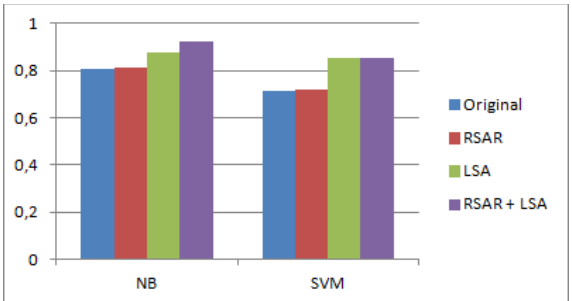
The obtained results for precision, recall, and F1-measure are presented in table 2. In this table, RSDR and LSA mean the feature selection methods which are used in our experiments, and whether they were applied or not, is denoted respectively by the symbols ✓ or ×. P, R, F1 denote precision, recall and F1-measure respectively. We can note the following remarks:

Table 2. Categorization Performance

Classifier	Performance				
	RSAR	LSA	P	R	F1
NB	×	×	0,807	0,804	0,805
	✓	×	0,816	0,809	0,812
	×	✓	0,900	0,856	0,877
	✓	✓	**0,926**	**0,917**	**0,921**
SVM	×	×	0,727	0,706	0,716
	✓	×	0,730	0,707	0,718
	×	✓	0,886	0,824	0,853
	✓	✓	**0,887**	**0,825**	**0,854**

The performance shown in table 2 indicates the following results: applying any feature selection method such as RSDR or LSA can improve the categorization performance because NB and SVM both have higher F1 values when these feature selection methods were applied than they did when no feature selection was applied. Furthermore, for each classifier, the LSA method is more effective than the RSAR method because the F1 value obtained when the LSA method was applied is higher than the highest F1 value obtained when the RSAR method was applied. But when we use RSAR combined with LSA we obtain the best results either with NB or SVM classifier (Figure 4).

In addition, our proposed method has been evaluated using NB classifier. The obtained results for precision, recall, and F1-measure presented in table 2 demonstrate the interest of our contribution.



F1-measure	Original	RSAR	LSA	RSAR + LSA
NB	0,805	0,812	0,877	0,921
SVM	0,716	0,718	0,853	0,854

Fig 4. F1-measure results

6. Conclusion and Future Work

In this paper, a hybrid feature selection method for Arabic short text is proposed based on RSAR and LSA. It is capable of compressing short text feature representation and reducing noise. The obtained results show the interest of our contribution. This technique can prove useful to a number of social media data analysis applications. In our future work, we try to explore other techniques for topic modeling such as Probabilistic Latent Semantic Analysis (PLSA) or Latent Dirichlet Allocation (LDA) in order to overcome the gaps of LSA.

References

[1] A. Chouchoulas and Q. Shen, Rough set-aided keyword reduction for text categorization, Applied Artificial Intelligence 15 (2001), 843–873.

[2] Anuj Mahajan, Sharmistha, Shourya Roy. Feature Selection for Short Text Classification usingWavelet Packet Transform. Proceedings of the 19th Conference on Computational Language Learning, pages 321–326, Beijing, China, July 30-31, 2015.

[3] Berry, M. W., Dumais, S. T., & O'Brien, G. W, Using Linear Algebra for Intelligent Information Retrieval, SIAM Review, 37:573-595, 1995

[4] Bilal Hawashin, Ayman M Mansour, Shadi Aljawarneh , "An Efficient Feature Selection Method for Arabic Text Classification", International Journal of Computer Applications (0975 – 8887) Volume 83 – No.17, December 2013

[5] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R. (1990) Indexing by Latent Semantic Analysis, Journal of the American Society of Information Science, 41(6): 391-407.

[6] G. H. Golub, C. F. Van Loan. Matrix Computations. Johns Hopkins Press, Baltimore, MD, 1989.

[7] Hamouda K.Chanta, David W. Corne, "Feature Subset Selection for Arabic Document Categorization using BPSO-KNN", Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on

[8] Hanchuan Peng, Fuhui Long, and Chris Ding. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27:1226–1238.

[9] He Hui, Chen Bo, Xu Weiran. Short Text Feature Extraction and Clustering for Web Topic Mining. 3rd International Conference on [5] J.J. Verbeek. "Supervised Feature Extraction for Text Categorization". Tenth Belgian-Dutch Conference on Machine Learning, (2000).

[10] H. Froud, A. Lachkar, S. A. Ouatik "A Comparative Study of Root-Based and Stem-Based Approaches for measuring the Similarity between Arabic words For Arabic Text Mining Applications", Advanced Computing: An International Journal (ACIJ), Vol.3, No.6, November 2012

[11] Jan Komorowski, Lech Polkowski, Andrzej Skowron. Rough Sets: A Tutorial. (1998)

[12] Jiliang TANG, Xufei WANG, Huiji GAO, Xia HU and Huan LIU.Enriching short text representation in microblog for clustering Front. Comput. Sci., 6(1) DOI 10.1007/s11704-009-0000-0. (2012)

[13] Joachims T. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In Proceedings of the European Conference on Machine Learning (ECML), pp.173-142. (1998)

[14] J. Katzberg and W. Ziarko, Variable precision extension of rough sets,W. Ziarko, ed., Fundamenta Informaticae, Special Issue on Rough Sets 27(2–3) (1996), 155–168.

[15] J. Lin and D. Gunopulos. Dimensionality reduction by random projection and latent semantic indexing. In Proceedings of the Text Mining Workshop, at the 3rd SIAM International Conference on Data Mining, May 2003.

[16] MESLEH A., (2007). Chi Square Feature Extraction Based SVMs Arabic Language Text Categorization System. Journal of Computer Science 3 (6): 430-435

[17] M. Bekkali, I. Sahmoudi, A. Lachkar. Enriching Arabic Tweets Representation based on Web Search Engine and the Rough Set Theory. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 Pages 1573-1574

[18] M. Chen, X. Jin, and D. Shen. Short text classification improved by learning multigranularity topics. Proc. 22nd International Joint Conference on Artificial Intelligence, pp. 1776-1781. (2011)

[19] Pawlak, Z. Rough sets: Theoretical aspects of reasoning about data, Kluwer Dordrecht. (1991)

[20] Peng Wang, Heng Zhang, Bo Xu, Chenglin Liu, and Hongwei Hao. Short text feature enrichment using link analysis on topic-keyword graph. In Natural Language Processing and Chinese Computing, pages 79–90. Springer, 2014

[21] Saleh Alsaleem. Automated Arabic Text Categorization Using SVM and NB, International Arab Journal of e-Technology, Vol. 2, No.2. (2011)

[22] Toshiko Wakaki. Hiroyuki Itakura, Masaki Tamura, "Rough Set-Aided Feature Selection for Automatic Web-Page Classification", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)

[23] U. Qamar "A Rough-Set Feature Selection Model for Classification and Knowledge Discovery". IEEE International Conference on Systems, Man, and Cybernetics. 2013

[24] Vapnik V. The Nature of Statistical Learning Theory, chapter 5. Springer-Verlag, New York. (1995)

[25] X. Hu, N. Sun, C. Zhang, and T.-S. Chua. Exploiting internal and external semantics for the clustering of short texts using world knowledge. Proc. 18th ACM Conference on Information and Knowledge Management, pp. 919-928. (2009)

[26] X.-H. Phan, L.-M.Nguyen, and S. Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. Proc. 17th International Conference on World Wide Web, pp. 91-100. (2008)

[27] Yang Y. and X. Liu. A re-examination of text categorization methods. Procin 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99), pp. 42– 49. (1999)

[28] Yiming Yang and Jan O. Pedersen. 1997. A comparative study on feature selection in text categorization. In Proceedings of the Fourteenth International Conference on Machine Learning, ICML '97, pages 412–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. Semantics, Knowledge, and Grid(SKG 2007), pp. 382-385.