World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:1, 2015

Optimal Classifying and Extracting Fuzzy Relationship from Query Using Text Mining Techniques

Faisal Alshuwaier, Ali Areshey

Abstract—Text mining techniques are generally applied for classifying the text, finding fuzzy relations and structures in data sets. This research provides plenty text mining capabilities. One common application is text classification and event extraction, which encompass deducing specific knowledge concerning incidents referred to in texts. The main contribution of this paper is the clarification of a concept graph generation mechanism, which is based on a text classification and optimal fuzzy relationship extraction. Furthermore, the work presented in this paper explains the application of fuzzy relationship extraction and branch and bound (BB) method to simplify the texts.

Keywords—Extraction, Max-Prod, Fuzzy Relations, Text Mining, Memberships, Classification.

I. Introduction

THIS paper aims to use text-mining techniques to classify the text and extract optimal fuzzy relationship from query with reasonably high classification accuracy. In the expression Text Mining, concerned with information learning from pre-processed text. In this research we address a problem to classify the query, extract events, which are also extensively applied within the domain of the query [2], [6], [19]. There is an increasing interest in text mining, text classification and fuzzy relationship extraction strategies applied due to the increasing number of electronically available publications stored in databases [3], [6], [15], [16]. While many of the categories do in fact refer to define keywords that are important words in a document and meant to indicate the topic or the contents of the query [1], [17]. It is also proposed that text mining should make the task easier and less time-consuming [1]-[3]. However, to date, most research in this area has focused on developing objective performance metrics for comparing different text mining systems [4], [18]. In this research, we described initial feedback from the use of text mining within the query and indexer, and report on experiments to evaluate how well our extracting system helps for fuzzy relationship. Our method is divided into six stages which areas follow: query preprocessing, web crawling, keyword classification and selection using tf-idf algorithm, equivalence relations. The most common types of membership functions are triangular, trapezoidal, and Gaussian shapes. The two membership functions used in the proposed system are Distance membership function and Gaussian membership function. The Selected Two Membership Functions (TMFs) are applied and among them the Optimal Membership Function (OMF) is selected based on the minimum Euclidean length measure. Optimal Fuzzy Relations extraction using some of the com-positional operators and graph generation using BB algorithm. The paper is organized as follows. Section II briefly describes the related works. Section III introduces the optimal fuzzy relations extraction architecture. Section IV details the proposed algorithm. Section V shows some obtained experimental results. Finally, Section VI concludes.

Faisal Alshuwaier is with the Department of The National Center for Computation Technology and Applied Mathematics, King Abdulaziz City for Science and Technology, Riyadh City, Kingdom of Saudi Arabia (e-mail: shuwaier@kacst.edu.sa).

Ali Areshey is with the Department of The National Center for Computation Technology and Applied Mathematics, King Abdulaziz City for Science and Technology, Riyadh City, Kingdom of Saudi Arabia (e-mail: aareshey@kacst.edu.sa).

II. Related Works

Pattern literature has been extensively studied in the information extraction. Extensively work is done by [7], who focused precisely on reduction in the time of the biomedical entity interactions when using the IE system, designed to recognize abstracts containing the pattern entity interactions. Reference [8] described a method for processing texts from widely differing domain and format. Reference [9] presented a system to identify components to extract fuzzy relationship in journal articles. Reference [1] proposed a technique extract the pattern relations using soft clustering data mining algorithm to increase the accuracy. Reference [4] presented a method for detecting the presence of fuzzy relationship in the text and dealing with spatial relationship using named entity extraction techniques coupled with self-learning fuzzy logic techniques. Reference [11] illustrated a novel concept graph creation technique which is under- pinned by a fuzzy extraction and text mining method. Reference [12] initiated text mining with a set of data using natural language processing techniques. Reference [13] improved the report and image retrieval by using ranking technique. Reference [14] proposed a novel fuzzy clustering algorithm using text mining. Reference [31] described a fuzzy transitive closure analysis, which is a powerful technique for pattern recognition. Reference [32] proposed system identifies the number of clusters which are equal to the number of implicit classes in multi class data. The best membership function for multi class data is searched in third phase with the minimum Euclidean norm.

III. Architecture

Fig. 1 shows the basic architecture of classification and extracting the optimal fuzzy relations. Individual components

are illustrated into five main stages, including Preprocessing, Web Crawling, Keyword Classification and Selection, Optimal Fuzzy Relation Extraction and Graph Generation. These stages are described as follows.

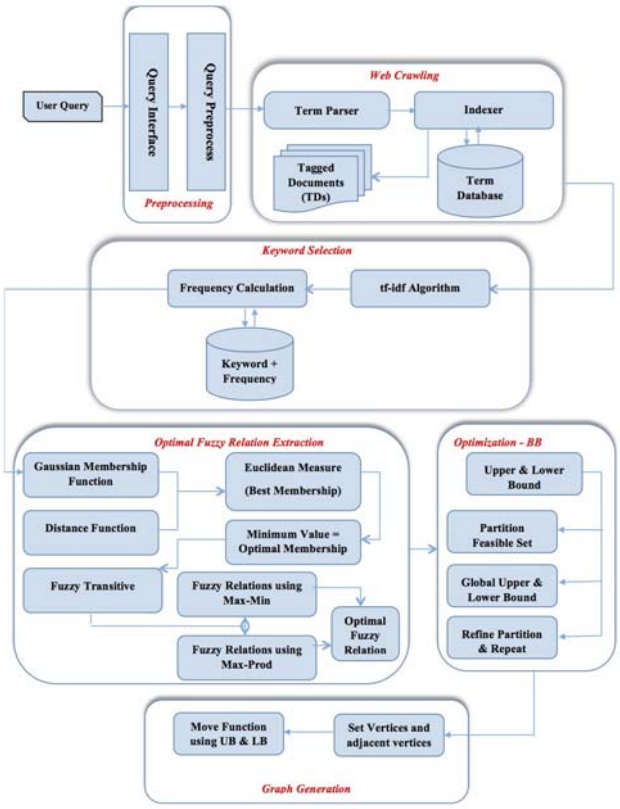

Fig. 1 The Basic Architecture of Optimal Fuzzy Relation Extraction

A. Preprocessing

The system has many components implemented as Web Services. In fact, the actual structure is described in this section.

1) Query Interface: The relevant user and functional queries are guided by the user interface.

2) Query Preprocess: Stop words are frequently used words that appear in the text but hold very little significant information to the text. For example, words the, of, a, and so on appear frequently in the texts [5], [25]. So, a predefined stop word list is applied to remove those words that do not discriminate for the queries.

B. Web Crawling

The web crawling collects the relevant portions of documents from web that must be indexed according to the mapping between the terms so that we only search for matches in those indexes, which fall into the same equivalence class. This is very important for performance of searching as it reduces the search space [22], [25].

1) Indexer: It retrieves the list content (H), which the terms are stored in this list and URLs out of web pages efficiently. It can deal with different types of variety of user query requests. It also works as parser and do not contain any rich information useful for the classifier [23], [25].

2) Term Parser: It tests each term that is stored in list (H) to assign and extract any functional term relation included in them. Relevant information about each useful word of user query collection is stored in fixed-size records (H) [24], [25].

3) Tagged Documents (TDs): In order to get data sources for our method, we indicate some fixed and reliable documents that are some websites is selected based on the results from indexer and term parser.

C. Keyword Classifications and Selection

We can build classifiers that use the term-frequency to represent textual information in the vector space. For query topic identification, we can define a feature for each word, indicating whether the query contains that word. We can then define a feature extractor that simply checks whether each of these words is present in a given query [10], [33]. In our method, the tf-idf weight gives how important is a word to a query in a collection, and that's why tf-idf incorporates local and global frameworks, because it takes in quest not only the isolated term but also the term within the query collection. To overcome this problem, the term frequency of a query on a vector space is usually also normalized.

1) tf-idf Algorithm: In support of classification, several information theory approaches have been employed. The tf-idf is the result of two statistics, such as term frequency (tf), which is the number of times that word 't' occurs in query 'q' and inverse document frequency (idf), which is the total number of the tagged documents where term t appears [20], [34]. tf-idf is word frequent 'x' that inverses document frequency used for measure of the important word 't' in the query and tagged document 'di' [21], [35]. tf-idf is described as:

$$tf(t,q) = n_{t,q} / \sum_{j=1}^{l} n_{j,q} \quad (1)$$

where,
n_t = number of occurrences of term t in q.
n_j = number of occurrences of all terms in q.

$$idf(t,di) = log(M/m_t + 0.01) \quad (2)$$

where,
M = total number of tagged documents in the web crawling.
mt = total number of tagged documents in the web crawling where word t appears.

$$tf - idf(t,q,di) = tf(t,q) \times idf(t,di) \quad (3)$$

The tf-idf term weighting scheme has been used extensively and has become the default choice in text classification. Thus, the keywords are stored in a list (H).

D. Optimal Fuzzy Relation Extraction Model

Fuzzy relations map elements of one universe , say X, to those of another universe, say Y, through the Cartesian product of the two universes. However, the strength of the relation between ordered pairs of the two universe is not measured with the characteristic function, but rather with a membership function expressing various degree of strength of the relation on the unit interval [0,1]. Hence a optimal fuzzy relation R is a mapping from the Cartesian space X*Y to the [0,1], where the strength of the mapping is expressed by the membership function of the relation for ordered pairs from the two universes, or μR (x,y) [27], [36]. The optimal fuzzy relation extraction model consists of the following:

1) Equivalence Memberships or Relations: The equivalence membership (EM) is very important in text mining because the EM on a set of objects determines a set of compatibility levels. The properties of EM can be applied as reflexive and transitive. The relation R can be described by a membership function $\{0, 1\}$, so if the membership value is '1' then the objects are related to each other at the same level [31]. The membership of each object (xi, yj) in the equivalence relation R is defined by xi*yj $\rightarrow$ [0, 1] [28], [34]. The grade of membership μ maps the object to positive real numbers in the interval [0, 1]. The definition of the membership function is given below:

$$\mu : x \rightarrow [0, 1], x \in X \tag{4}$$

where x is a real object value, X refers to the universal set.

2) Membership Functions: A membership function provides a measure of the degree of similarity of the object to a fuzzy set. Each membership function states skilled opinion to get proper numerical values for fuzzy properties [30], [32]. The proposed method is for finding the best membership functions for the input data-set so that the optimal number of cluster can be revealed by the proposed system. The most common types of membership functions are triangular, trapezoidal, and Gaussian shapes. The two membership functions used in the proposed system are Distance membership function and Gaussian membership function [37].

The Gaussian membership function (μ_G) is represented according to the following formula:

$$\mu_G(x) = [\exp^{-1/2}(x - b/\sigma)^2] \tag{5}$$

where,
x is the real object value,
b is the center of the membership function and
σ is a constant.

The general formula applied for the distance membership function ($\mu_D ist$) is Minkowski class:

$$D(i, j) = (|xi1 - xj1|^q + |xi2 - xj2|^q + ... + |xip - xjp|^q)^{1/q} \tag{6}$$

where,
i, j, p are indexes of Minkowski distance.
q is a distance function parameter.

Let a relation $\mu_D ist$ defined in terms of an appropriate distance function by the following formula:

$$\mu_D ist(i, j) = 1 - 1/mD(i, j) \tag{7}$$

where,
m = the number of samples in the training data that ensures that $\mu_D ist$(i, j) $\in$ [0,1].

3) Optimal Membership Function (OMF): The Selected Two Membership Functions (TMFs) are applied and among them the best membership function is selected based on their Euclidean length measure. Initially all the membership functions are stored in the file. Then the Euclidean length measure is calculated which drives the conclusion of goodness of fit measure. The membership function with minimum Euclidean length is selected as the Optimal Membership Function (OMF) for the input data-set.

4) Fuzzy Relation Computation: A fuzzy compatibility relation is defined on a set, which is maximum-minimum transitive. A similarity class is a fuzzy set in which the membership grade of any object represents the similarity of that objects to the set. Each equivalence relation is associated to the set of partitions.

5) Fuzzy Relations Using Transitive Closure: In order to achieve an equality fuzzy relation we apply a transitive closure. This strategy depends on the selection of the fuzzy similarity index and allows finding a partition of the universe depending on the level of similarity considered. A very important theorem proves that the partition obtained from the transitive closure is the same that with the hierarchical method of single linkage and the fuzzy connected components of the fuzzy graph defined by the matrix [38]. The transitive closure is determined by simple algorithm that can be calculated as follows:

$$R^+ = \cup_{i \in \{1,2,3,...\}} R^i \tag{8}$$

where,
R^i is the i^{th} power of R, defined inductively by R^i = R, and for i $>$ 0: R^{i+1} = R $\circ$ R^i where, $\circ$ denotes composition of relations [38].

6) Optimal Fuzzy Relations Using Some of the Com-Positional Operators: An optimization model with objective function about max-product composition subject to the fuzzy relations about Max-Product (Max-Prod) and Max-Min composition are provided. In fuzzy relation applications, the Max-Min and Max-Prod com-positional operators are the most commonly and frequently used due to their computational efficiency. The Max-Prod Method of any fuzzy relations R and T is calculated by formula:

$$R \circ T = Max\{\mu R(x, y) \circ \mu T(y, z)\}/(x, z) \tag{9}$$

The Max-Min Method of any fuzzy relations R and T is calculated by formula:

$$R \circ T = Max\{Min\{\mu R(x, y), \mu T(y, z)\}\} \tag{10}$$

E. Graph Generation

The arbitrary graphs are used to formulate and represent the basic structure such as concepts and relations between fuzzy terms and the user query [26]. In our method, we use a branch-and-bound algorithm that consists of a systematic enumeration of candidate solutions by means of state space search for global optimization for non-convex problems. The basic idea is rely on two factors that efficiently compute a lower and an upper bound on the optimal value over a given region · Upper bound can be found by selecting any point in the region, or by a local optimization method · Lower bound can be found from convex relaxation, duality, or other bounds

The set of candidate solutions is thought of as forming a rooted graph with the full set at the root. The algorithm explores links of this graph, which represent subsets of the solution set. Before calculating the candidate solutions of a link, the link is checked against upper and lower estimated bounds on the optimal solution, and is discarded if it cannot produce a better solution than the best one found so far by the algorithm. The calculation is as follows:

$$b(i) = A_i + \sum_{j=1}^{i} B_j + min_{j \neq i} C_j \tag{11}$$

where,
B_j is the summation of the minimum value of the rows and columns and,
A_i is the value of the element.

IV. Algorithm

This section begins with a discussion of the query text, classification, optimal fuzzy relations extraction and graph generation that have been used. We then specify our implementation of the query and extraction algorithm to identify the keywords and optimal fuzzy relations. The section then covers the methods used for evaluating the query extraction algorithm. As described earlier, the algorithm requires a string of the query. It has several parameters that will set up and affect how the text extraction process will be performed. It enhances the accuracy and performance of extraction of optimal fuzzy relations from web crawling. Similarity is a major concept in the representation of vague resource. A common approach for extracting the optimal fuzzy relations is to treat the relevant term to represent the group of the pattern [29].

Algorithm 1: Pseudo Code of Stop Word Removal;
Input: Query q= w1,w2,...wn
Output: Query without stop word
Function: Convert StopWord()
1: Read the query from user using ScanQuery(System.in)
2: Declare the dictionary of stop words
3: Split parameter into words
4: Allocate new dictionary to store found words
5: Store results in this String Builder
6: Loop through all words using InputNextLine()
7: Convert to lowercase
8: IF this is a usable word, THEN add it
9: Return (string with words removed)
10: Display query without stop words using Convert(Query)

Algorithm 2: Pseudo Code of Web Crawler;
Input: Sentence from the Query
Output: Document Tagging
Function: $Web_Crawler$(Keyword)
1: S = An empty list (Tagged Documents)
2: Initialize TERM = String
3: Initialize n = Integer
4: r = number of records in Database
5: While n not equal r
6: Search TERM in the Database
7: IF TERM is matched the term in the Database THEN
8: S = Collect the Document in the List
9: n = n + 1
10: ELSE
11: n = n + 1
12: UNTIL r = n;
13: Return (Term, S)

Algorithm 3: Pseudo Code of Query Classification;
Input: Term, S
Output: Keyword and the category of the query
Function: $Query_Classification$(line1 to lineN)
1: Use Convert StopWord()
2: Get vector for each sentence
3: For defining term frequency (tf) as (1)
4: Define inverse document frequency (idf) as explained in (2)
5: Calculate tf-idf as explained in (3)
6: Select the maximum value of tf-idf (Max-Val)
7: Keyword = Get the word with the maximum value of tf-idf
8: Return (Keyword, Max-Val)

Algorithm 4: Pseudo Code of Gaussian Membership Function;
Input: q
Output: Relations using Gaussian Function
Function: $Gaussian_Membership$(H, Keyword, q)
1: Initialize sigma σ = 1
2: Initialize cent, Gaussian (μG) = 0
3: initialize x = words from q
4: For all x Do
5: Calculate σ for all inputs
6: Cent = center of μ_G
7: Calculate μ_G as explained in (5)
8: Until Finish of processing all the variables
9: Return (μ_G)

Algorithm 5: Pseudo Code of Distance Membership Function;
Input: Keyword, Oi (Keyword)
Output: Relations using Distance Metrics
Function: $Distance_Membership$(Keyword)
1: Use Minkowski class for estimating the proximity of terms (μ_Dist) as explained in (6,7)
2: Return (μ_Dist)

Algorithm 6: Pseudo Code of Optimal Membership Function;
Input: μ_G, $\mu_D ist$, Min
Output: OMF
Function: $Optimal_M embership$($\mu_D ist$, $\mu_D ist$, Min)
1: Call μ_G
2: Call $\mu_D ist$
3: OMF = Minimum Euclidean length (μ_G, $\mu_D ist$)
4: Return(OMF)

Algorithm 7: Pseudo Code of Optimal Fuzzy Relations Extraction;
Input: OMF
Output: Fuzzy Relation Extraction (FRE)
Function: FRE(μ_G, $\mu_D ist$, OMF)
1: Set M = a constant that ensure the relation $\in$ [0, 1]
2: Calculate the Equivalence Relation as explained in (7)
3: Use a transitive closure as explained in (8)
4: Use Max-Prod Method to choose Optimal Fuzzy Relations as explained in (9)
5: Use Max-Min Method to choose Optimal Fuzzy Relations as explained in (10)
6: FRE elements = Select the Minimum Value
7: Return (FRE elements)

Algorithm 8: Pseudo Code of Bound and Bound Graph;
Input: FRE elements
Output: BB Graph Function
Function: BBGraph(FRE elements)
1: Use BB Method as explained in (11)
2: Split the solution into groups
3: Each splitting incurs a lower bound

V. Implementation

Implementation is the process of executing a plan or design to achieve some output. In this paper, the implementation encompasses keyword classification, the extraction of optimal fuzzy relations in the query, fetching them in the system to go through the query preprocessing techniques, forwarding the preprocessed documents to the web crawling to obtain the tagged documents, then sending them to the keyword selection system. Then the best membership function is searched to generate the optimal membership grades. The membership function with minimum Euclidean length is selected as the Optimal Membership Function (OMF) for the input data-set. The information contained in Table I is used to generate the required document classification data for applying fuzzy equivalence relation between these words from the query and the extracted words economy, operation, fund, financial, medical, machine. Since it is not directly possible; so first determine the keyword from the information contained in Table I. Then, we already computed the optimal fuzzy relations extraction using some of the compositional operators. Finally, the BB graph is efficiently generated using BB algorithm.

To illustrate the method based on classification and fuzzy equivalence relation, let us take an example. In this example there are six words as shown below In Table I.

TABLE I: Input Query for the Experimental Analysis

Input Query (Word)	Equivalence
Economy	W1
Operation	W2
Fund	W3
Financial	W4
Medical	W5
Machine	W6

By applying the tf-idf in (3), the keywords are extracted by the traditional tf-idf method as shown in Table II.

TABLE II: Extracted Keywords by tf-idf

Word	tf	mt	M	idf	tf-idf	Keyword
W1	3	695M	6825M	0.99	2.97	$\surd$
W2	1	842M	6825M	0.90	0.90	$\times$
W3	2	538M	6825M	1.1	2.2	$\times$
W4	1	1,450M	6825M	0.67	0.67	$\times$
W5	1	1,830M	6825M	0.57	0.57	$\times$
W6	1	1,470M	6825M	0.66	0.66	$\times$

Then we will choose a particular channel and use the tagged documents for the idf value to calculate the word frequency as shown in Table III.

TABLE III: Word Frequency for the Tagged Documents

Tagged Documents	Word Frequency					
	W1	W2	W3	W4	W5	W6
TD1	4	3	3	2	1	1
TD2	3	2	2	3	0	0
TD3	4	3	4	2	0	1
TD4	3	2	2	2	0	0
TD5	2	1	1	1	0	0

By calculating the word frequency in Table IV, which the truthfulness degree that the word frequency belongs to specific tagged document and word ID.

TABLE IV: Word ID and Total Frequency

Words	Word ID	Total Frequency
Economy (W1)	1	16
Operation (W2)	2	11
Fund (W3)	3	12
Financial (W4)	4	10
Medical (W5)	5	1
Machine (W6)	6	2

Table V shows the Gaussian membership based on (5).

TABLE V: Gaussian Membership

Gaussian Membership	W1	W2	W3	W4	W5	W6
W1	1	0.71	0.80	0.60	0.04	0.065
W2	0.71	1	0.98	0.98	0.25	0.32
W3	0.80	0.98	1	0.95	0.20	0.25
W4	0.60	0.98	0.95	1	0.33	0.40
W5	0.04	0.25	0.20	0.33	1	0.98
W6	0.065	0.32	0.25	0.40	0.98	1

In this implementation, there are six tagged documents and six words as shown in Fig. 2.

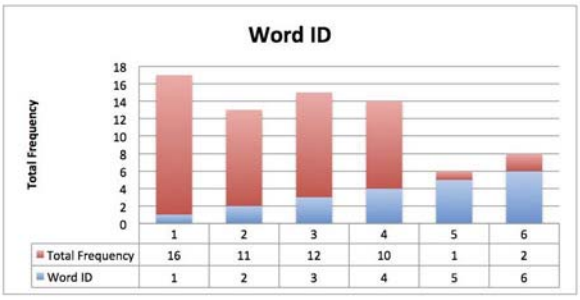

Fig. 2 Chart of Six Tagged Documents and Six Words

TABLE VI: DATA T AND INDEXES

T = Word ID	1	2	3	4	5	6
WT = Total Frequency	16	11	12	10	1	2
XT = Index	0	1	2	3	4	5

The data T and Indexes for each word are shown in Table VI.

The pairs of total frequency and Index are shown in Table VII.

TABLE VII: TOTAL FREQUENCY AND INDEXES

Data Points	Pairs (WT, XT)
W1	(16,0)
W2	(11,1)
W3	(12,2)
W4	(10,3)
W5	(1,4)
W6	(2,5)

The largest Euclidean distance from (6) is 16 Then sigma = 1/16 = 0.0625, when q = 2.

Table VIII shows the Distance membership based on (7).

TABLE VIII: DISTANCE MEMBERSHIP

Distance Membership	W1	W2	W3	W4	W5	W6
W1	1	0.70	0.72	0.58	0.03	0.07
W2	0.70	1	0.91	0.86	0.34	0.38
W3	0.72	0.91	1	0.86	0.30	0.35
W4	0.58	0.86	0.86	1	0.43	0.49
W5	0.03	0.34	0.30	0.43	1	0.91
W6	0.07	0.38	0.35	0.49	0.91	1

The best membership function is selected based on their Euclidean length measure. Initially all the membership functions are stored in the file. Then the Euclidean length measure is calculated which drives the conclusion of goodness of fit measure. The membership function with minimum Euclidean length is selected as the Optimal Membership Function (OMF) for the input data-set. Table IX shows the Euclidean distance for the Distance membership elements.

Table X shows the Euclidean distance for the Gaussian membership elements.

The Optimal Membership Function (OMF) is presented in the Table XI, which OMF is calculated based on the minimum vale of the Euclidean distance of the membership grades.

Given a relation R(X,X), the element if R^1 is the Maximum-Prod (xrj, xjs) with j varying from 1 to k. The

TABLE IX: EUCLIDEAN LENGTH OF DISTANCE MEMBERSHIP

Euclidean Distance (ED) of $\mu_D ist$	Value
ED (W1, W2)	0.68
ED (W1, W3)	0.66
ED (W1, W4)	0.76
ED (W1, W5)	1.80
ED (W1, W6)	1.66
ED (W2, W3)	0.14
ED (W2, W4)	0.28
ED (W2, W5)	1.61
ED (W2, W6)	1.40
ED (W3, W4)	0.31
ED (W3, W5)	1.51
ED (W3, W6)	1.43
ED (W4, W5)	1.31
ED (W4, W6)	1.23
ED (W5, W6)	0.16
Total	14.94

TABLE X: EUCLIDEAN LENGTH OF GAUSSIAN MEMBERSHIP

Euclidean Distance (ED) of μ_G	Value
ED (W1, W2)	0.67
ED (W1, W3)	0.58
ED (W1, W4)	0.78
ED (W1, W5)	1.82
ED (W1, W6)	1.77
ED (W2, W3)	0.13
ED (W2, W4)	0.16
ED (W2, W5)	1.74
ED (W2, W6)	1.66
ED (W3, W4)	0.30
ED (W3, W5)	1.82
ED (W3, W6)	1.74
ED (W4, W5)	1.63
ED (W4, W6)	1.53
ED (W5, W6)	0.12
Total	16.45

TABLE XI: OPTIMAL MEMBERSHIP FUNCTION

Two Membership Function (TMF)	Value
Euclidean Distance (ED) of μ_G	16.45
Euclidean Distance (ED) of $\mu_D ist$	14.94
Optimal Membership Function (OMF)	14.94

Fuzzy Transitive Closure (FTC) can be determined by (8). This is continued until no new relationship is generated. The relation $R^2 = R^1 \circ R^1$. Take the elements in the first row and first column in R^1.

The relational array $R^2 = R^1 \circ R^1$ is obtained as follows: Take the elements in the first row and first column in R^1:

W1 (First Row):	1	0.70	0.72	0.58	0.03	0.07
W1 (First Column):	1	0.70	0.72	0.58	0.03	0.07
Prod W1 ∘ W1:	1	0.49	0.52	0.34	0.0009	0.0049
Max:	1	-	-	-	-	-

Take the elements in the first row and second column in R^1:

W1 (First Row):	1	0.70	0.72	0.58	0.03	0.07
W2 (Second Column):	0.7	1	0.91	0.86	0.34	0.38
Prod W1 ∘ W2:	0.7	0.7	0.66	0.50	0.01	0.02
Max:	0.7	0.7	-	-	-	-

Similarly, all elements in xij are found in R^2. Now, the Fuzzy Transitive Closure using Max-Prod technique based on (9) is shown in Table XII and Fig. 3.

TABLE XII: FTC USING MAX-PROD

FTC (Max-Prod)	W1	W2	W3	W4	W5	W6
W1	1	0.70	0.72	0.62	0.25	0.28
W2	0.70	1	0.91	0.86	0.37	0.42
W3	0.72	0.91	1	0.86	0.37	0.42
W4	0.62	0.86	0.86	1	0.45	0.49
W5	0.25	0.42	0.37	0.45	1	0.91
W6	0.28	0.42	0.42	0.49	0.91	1

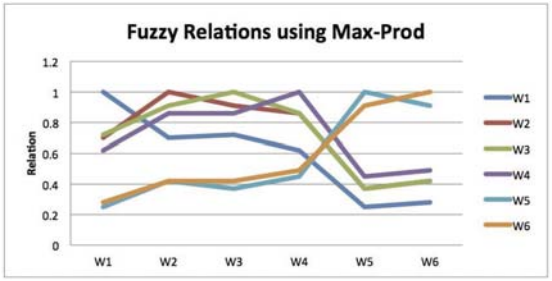

Fig. 3 Fuzzy Relation using Max-Prod

Now, the Fuzzy Transitive Closure using Max-Min technique based on (10) is shown in Table XIII and Fig. 4.

TABLE XIII: FTC USING MAX-MIN

FTC (Max-Min)	W1	W2	W3	W4	W5	W6
W1	1	0.72	0.72	0.72	0.43	0.49
W2	0.70	1	0.91	0.86	0.43	0.49
W3	0.72	0.91	1	0.86	0.43	0.49
W4	0.72	0.86	0.86	1	0.49	0.49
W5	0.43	0.43	0.43	0.49	1	0.91
W6	0.49	0.49	0.49	0.49	0.91	1

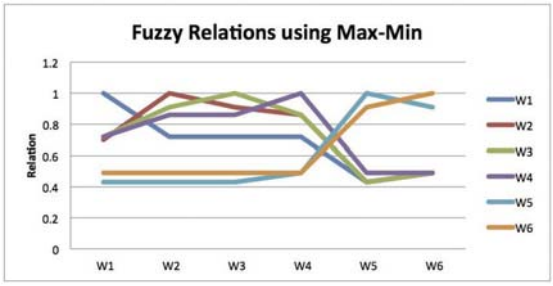

Fig. 4 Fuzzy Relation using Max-Min

The reduced matrix is to get the lower bound of the path starting at node 1: W1 as shown in Table XIV.

TABLE XIV: BRANCH AND BOUND FOR NODE 1: W1

BB Node	W1	W2	W3	W4	W5	W6	Reduced
W1	∞	0.31	0.35	0.17	0	0	0.25
W2	0.33	∞	0.42	0.29	0	0.02	0.37
W3	0.35	0.4	∞	0.29	0	0.02	0.37
W4	0.17	0.27	0.29	∞	0	0.01	0.45
W5	0	0.03	0	0	∞	0.63	0.25
W6	0	0	0.02	0.01	0.63	∞	0.28
Reduced	-	0.14	0.12	0.20	-	0.03	-

The Reduced Cost for node 1 (W1) is {0.25 + 0.37 + 0.37 + 0.45 + 0.25 + 0.28} + { 0.14 + 0.12 + 0.20 + 0.03} = 2.46. Finally, the reduced cost of the last node (W5) is cost of Node W3 + A(5,1) + lower bound = 3.31 + 0 + 0 = 3.31. The BB graph is presented in Fig. 5.

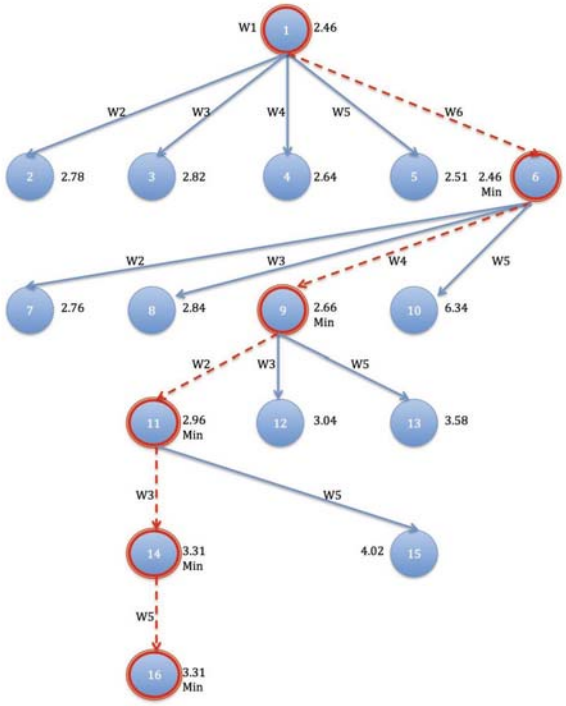

Fig. 5 Branch and Bound Graph

VI. CONCLUSION

Text mining, fuzzy and classification techniques are really powerful. This paper study was completely based on these techniques. The system was created for classifying and extracting the optimal fuzzy relation from the query. Query data has fuzzy characteristics; so extracting and computing fuzzy is sometimes better suitable for text mining in comparison with conventional classification.

ACKNOWLEDGMENT

The authors would like to acknowledge all the reviewers for their valuable suggestions, which contributed to the clarity of the paper and in particular for their comments for assigning broad semantic types to extract the optimal fuzzy relations.

REFERENCES

[1] S. Vashishtha, and Y. Kumar, "Efficient Retrieval of Text for Biomedical Domain using Expectation Maximization Algorithm". *IJCSI International Journal of Computer Science*. Issues, Vol. 8, Issue 6, No 1, 2011.

[2] F. Hogenboom, F. Frasincar, U. Kaymak, and F. Jong F, "An Overview of Event Extraction from Text. Proceedings of Detection". *Representation and Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011) at Tenth International Semantic Web Conference (ISWC 2011)*, Volume 779, pp. 48–57, CEUR-WS.org, Bonn, 2011.

[3] B. Alex, C. Grover, B. Haddow, M. Kabadjov, E. Klein, M. Matthews, S. Roebuck, R. Tobin, And X. Wang, "Assisted Curation: Does Text Mining Really Help?". *Pacific Symposium on Biocomputing 13*, pp. 556–567. PSB, Island of Hawaii, 2008.

[4] V. Kanagavalli, and K. Raja, "Detecting and resolving spatial ambiguity in text using named entity extraction and self learning fuzzy logic techniques". *National Conference on Recent Trends in Data Mining and Distributed Systems*, SBN 978-81-909042-5-4, pp. 71–76. NCT2DS, 2011.

[5] D. MacKinnon, L. Goldberg, G. Clarke, D. Elliot, J. Cheong, A. Lapin, E. Moe, and J. Krull, "Mediating Mechanisms in a Program to Reduce Intentions to Use Anabolic Steroids and Improve Exercise Self-Efficacy and Dietary Behavior". P*revention Science*, Vol. 2, No. 1, pp. 15–28. PS Press, 2001.

[6] Y. Garten, "Text Mining of the Scientific Literature to Identify Pharmacogenomic Interactions". *Stanford University*, 2010.

[7] I Donaldson, J. Martin, B. de Bruijn, C. Wolting, V. Lay, B. Tuekam, S. Zhang, B. Baskin, G.D. Bader, k. Michalickova, T. Pawson, and C.W.V. Hogue, "PreBIND and Textomy - mining the biomedical literature for protein-protein interactions using a support vector machine". B*MC Bioinformatics*, Vol. 4, pp. 11. Springer, 2003.

[8] N. Karamanis, I. Lewin, R. Seal, R. Drysdale, and E. Briscoe, "Integrating natural language processing with FlyBase curation". *In Proceedings of PSB 2007*, pp 245-256. PSB Press, Maui, Hawaii, 2007.

[9] T. Murphy, T. McIntosh, and J. Curran, "*In Australian Language Technology Workshop*", pp. 59-66. ALTW Press, 2006.

[10] S. Samarawickrama, L. Jayaratne, "Focused Web Crawling Using Named Entity Recognition For Narrow Domains". *IJRET*, http://www.ijret.org/ 2012.

[11] R. Lau, D. Song, Y. Li, T. Cheung, and J. Hao, "Toward a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning". *IEEE Transactions On Knowledge And Data Engineering*, VOL. 21, NO. 6, pp. 800–813. IEEE, 2009.

[12] M. Abulaish, and L. Dey, "Biological relation extraction and query answering from MEDLINE abstracts using ontology-based text mining". *Data & Knowledge Engineering Journal*, Volume 61, Issue 2, pp. 228-262, 2007.

[13] A. Prasad, S. Ramakrishna, D. Kumar, and B. Padmaja, "Extraction of Radiology Reports using Text mining". *(IJCSE) International Journal on Computer Science and Engineering*, Vol. 02, No. 05, pp. 1558–1562. IJCSE Press, 2010.

[14] M. Rodrigues, and L. Sacks, "A Scalable Hierarchical Fuzzy Clustering Algorithm for Text Mining". *In Proceedings of the 5th International Conference on Recent Advances in Soft Computing*. pp. 269–274. Nottingham, 2004.

[15] S. Jusoh, and H. Alfawareh H, "Techniques, Applications and Challenging Issue in Text Mining". *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 6, No 2. IJCSI Press, 2012.

[16] S. Ghosh, S. Roy, and S. Bandyopadhyay, "A tutorial review on Text Mining Algorithms". *International Journal of Advanced research in Computer and Communication Engineering,* ISSN 2278-1021, Vol. 1, Issue 4. IJARCCE Press, 2012.

[17] T. Nogueira, H. Camargo, and S. Rezende, "Fuzzy-DDE: a fuzzy method for the extraction of document cluster descriptors". *International Journal of Computer Information Systems and Industrial Management Applications*, ISSN 2150-7988, Vol. 5. pp. 472–479. IJCISIM Press, 2013.

[18] K. Oh, C. Lim, S. Kim, and H. Choi, "Research Trend Analysis using Word Similarities and Clusters". *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 8, No. 1. IJMUE Press, Tasmania, 2013.

[19] N. Uramoto, H. Matsuzawa, T. Nagano, A. Murakami, H. Takeuchi, and K. Takeda, "A text-mining system for knowledge discovery from biomedical documents". *IBM Systems Journal*, Vol. 43, No. 3. IEEE Press, Japan, 2004.

[20] L. Magdalena, M. Ojeda-Aciego, and J. Verdegay (eds), "Extracting topics in texts: Towards a fuzzy logic approach". *Proceedings of IPMU'08*. pp. 1733–1740. Torremolinos, 2008.

[21] M. Kathuria, N. Duhan, and C. Nagpal, "Application Of Fuzzy Logic In Web Mining Domain: A Survey". *International Journal of Advanced Research in IT and Engineering*. ISSN: 2278-6244, Vol. 1, No. 3. Tamilnadu, 2012.

[22] T. Martin, and M. Azmi-Murad, "An Incremental Algorithm to find Asymmetric Word Similarities for Fuzzy Text Mining". *Soft Computing as Transdisciplinary Science and Technology Advances in Soft Computing*, Vol. 29. pp. 838–847. Springer, 2005.

[23] A. Kaladevi, S. Padmavathy, and S. Theetcchenya, "Augmentation of Knowledge Reuse Employing Fuzzy Ontology Based Approach". *International Journal of Engineering and Management Research*, ISSN: 2250-0758, Vol. 3, Issue 2. pp. 17-21. IJMER Press, India, 2013.

[24] B. Anjali, and G. Bamnote, "Web Document Clustering Using Fuzzy Equivalence Relations". *Journal of Emerging Trends in Computing and Information Sciences.* ISSN: 2079-8407, Vol. 2, Special Issue. pp. 22–27. CIS Journal, 2011.

[25] F. Alshuwaier, W. Almutairi, and A. Areshey, "Smart Search Tools Using Named Entity Recognition". *Proceeding in Information Technology and Applications (ITA)*. ISBN: 978-1-4799-2876-7. pp. 304–311. IEEE, Chengdu, 2013.

[26] A. Muhammad, and L. Dey, "Biological ontology enhancement with fuzzy relations: a text-mining framework". ISBN: 0-7695-2415-X. pp. 379–385. IEEE, Canada, 2005.

[27] R. Lau, D. Song, Y. Li, T. Cheung, and X. Jin Hao, "Towards A Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning". *CiteSeerX Scientific Literature Digital Library and Search Engine*. CiteSeerX, 2013.

[28] J. Yen, and R. Langari, "Fuzzy logic: intelligence, control, and information". *Prentice-Hall*, Inc. 1998.

[29] M. Guelpeli, and A. Garcia, "An Analysis of Constructed Categories for Textual Classification Using Fuzzy Similarity and Agglomerative Hierarchical Methods". *Third International IEEE Conference on Signal and Image Technologies and Internet-Based System*. IEEE, Lisboa Reitoria, 2008.

[30] D. Georgiou, T. Karakasidis, J. Nieto, and A. Torres A, "Use of Fuzzy Clustering Technique and Matrices to Classify Amino Acids and Its Impact to Chou's Pseudo Amino Acid Composition". *Journal of Theoretical Biology*, Reference: YJTBI 5356. JTB Press, 2008.

[31] M. Sridharan, "Fuzzy mathematical model for the analysis of geomagnetic field data". *The Society of Geomagnetism and Earth, Planetary and Space Sciences (SGEPSS); The Seismological Society of Japan*; Earth Planets Space, 61, pp. 1169-1177, Japan 2009.

[32] N. Archana, P. Girish, and P. Sandip P, "Improved Membership Function for Multiclass Clustering with Fuzzy Rule Based Clustering Approach". *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Volume 3, Issue 5, 2014.

[33] B. Steven, K. Ewan, and L. Edward, "Chapter from Natural Language Processing with Python". Version 3.0, USA, 2014.

[34] A. Kaladevi, and S. Padmavathy, "Ontology Extraction for E-Learning - A Fuzzy Based Approach". *International Conference on Computer Communication and Informatics (ICCCI -2013)*, INDIA, 2013.

[35] C. Rakhi, "Domain Keyword Extraction Technique: A New Weighting Method Based On Frequency Analysis". *ACER 2013*, pp. 109-118, CS & IT-CSCP, 2013.

[36] B. SankaraSubramanian, R. Vasanth Kumar Mehta, "Contradiction Analysis in Text Mining: A Fuzzy Logic Approach". *SCSVMV University*, kanchipuram, INDIA, 2009.

[37] M. Foley, "The Application of Fuzzy Logic in Determining Linguistic Rules and Associative Membership Functions for the Control of a Manufacturing Process". *Masters Dissertation*. Dublin Institute of Technology, 2011.

[38] F. Bertran, N. Clara, and J. Ferrer, "Extending The Roughness Of The Data Via Transitive Closures Of Similarity Indexes". Vol. XII, No. 2, pp. 75-84, 2007.