Polibits

ISSN: 1870-9044

polibits@nlp.cic.ipn.mx

Instituto Politécnico Nacional

México

Das, Nibaran; Ghosh, Swarnendu; Gonçalves, Teresa; Quaresma, Paulo

Comparison of Different Graph Distance Metrics for Semantic Text Based Classification

Polibits, vol. 49, 2014, pp. 51-57

Instituto Politécnico Nacional

Distrito Federal, México

Comparison of Different Graph Distance Metrics for Semantic Text Based Classification

Nibaran Das, Swarnendu Ghosh, Teresa Gonçalves, and Paulo Quaresma

***Abstract*—Nowadays semantic information of text is used largely for text classification task instead of bag-of-words approaches. This is due to having some limitations of bag of word approaches to represent text appropriately for certain kind of documents. On the other hand, semantic information can be represented through feature vectors or graphs. Among them, graph is normally better than traditional feature vector due to its powerful data structure. However, very few methodologies exist in the literature for semantic representation of graph. Error tolerant graph matching techniques such as graph similarity measures can be utilised for text classification. However, the techniques like Maximum Common Subgraph (mcs) and Minimum Common Supergraph (MCS) for graph similarity measures are computationally NP-hard problem. In the present paper summarized texts are used during extraction of semantic information to make it computationally faster. The semantic information of texts are represented through the discourse representation structures and later transformed into graphs. Five different graph distance measures based on Maximum Common Subgraph (mcs) and Minimum Common Supergraph (MCS) are used with k-NN classifier to evaluate text classification task. The text documents are taken from Reuters21578 text database distributed over 20 classes. Ten documents of each class for both training and testing purpose are used in the present work. From the results, it has been observed that the techniques have more or less equivalent potential to do text classification and as good as traditional bag-of-words approaches.**

***Index Terms*—Graph distance metrics, maximal common subgraph, minimum common supergraphs, semantic information, text classification.**

I. Introduction

The research on automatic text classification task [1], [2] is one of the interesting area to the Natural Language Processing (NLP) researchers for the last few decades due to having its huge applications. The task becomes still more challenging with the ever increasing volume of complex text information especially through web-based services. State of the art approaches typically represent documents as vectors (bag-of-words) and use a machine learning algorithm, such as k-NN, Naïve Bayes, SVM to create a model and to classify new documents. But these approaches fail to represent the semantic content of the documents which is necessary for certain kind of tasks such as opinion mining, sentiment analysis etc. Therefore, in spite of being able to obtain good results, these approaches are utilized only for limited number of tasks. To overcome the limitations, the researchers are aiming to evaluate and use more complex knowledge representation structures [3], [4].

In this paper, a new approach which integrates a deep linguistic analysis of the documents with graph-based representation has been proposed for the text classification. Discourse representation structures (DRS) [5] are used to represent the semantic content of the texts and are transformed by our system into graph structures. Then, we proposed, applied, and evaluated several graph distance metrics [6] on 20 document classes from Reuters21578 text database taking 10 docs of each class for both training and testing purpose using a k-NN classifier. Later we compared the obtained results with the result obtained by traditional bag-of-words approaches.

The paper is organized as follows. Section 2 briefly describes the theoretical background related to our approach: discourse representation theory, graph representation, k-NN classifiers, and graph metrics. Section 3 presents our system and its modules. Section 4 exposes the performed experiments and discusses the obtained results. In the final section of the paper, conclusions and future work are presented.

II. Theory and Algorithms

A. Brief description of DRS

Extracting information from documents can be carried out in many ways, starting from statistic or probabilistic models to the ones involving deep linguistic structures. Our main goal in this work is to develop a technique which analyses documents in lexical, syntactic as well as semantic level.

Discourse Representation Theory (DRT), proposed by Kamp and Reyle [5] is one of the most advanced form of representing semantic context of a document. In DRT, a sequence of sentences $S_1, S_2, \ldots, S_n$ is passed into an algorithm. It starts with syntactic analysis of the first sentence S_1 and transforms it roughly top down, left to right fashion according to some DRS construction rules. This new DRS K_1 serves as a context for analyzing S_2 which in turn generates $K_{1,2}$ by appending the new semantic content to K_1.

A complete DRS expression is composed of: (a) a set of referents, which are the entities that have been introduced into

Manuscript received on January 4, 2014; accepted for publication on February 6, 2014.

Nibaran Das (corresponding author) is with the Computer Science and Engineering Department, Jadavpur University, Kolkata-700032, India (phone: +91 332 414 6766; fax: +91 332 414 6766; e-mail: nibaran@ieee.org).

Swarnendu Ghosh is with the Computer Science and Engineering Department, Jadavpur University, Kolkata-700032, India.

Teresa Gonçalves is with the Dept. of Computer Science, School of S&T, University of Évora, Évora, Portugal.

Paulo Quaresma is with the Dept. of Computer Science, School of S&T, University of Évora, Évora, Portugal, and with with L2F – Spoken Language Systems Laboratory, INESC-ID, Lisbon, Portugal.

the context, (b) a set of conditions, which are the relations that exist between the referents.

DRT provides a very logical platform for the representation of semantic structures of sentences including complex predicates like implications, propositions and negations, etc. It is also able to separately localize almost every kind of events and find out their agents and patients.

Here is an example of a DRS representation of the sentence "He drinks water.". Here, *x1*, *x2*, and *x3* are the referents and *male(x1), water(x2), drink(x3), event(x3), agent(x3, x1), patient(x3, x2)* are the conditions

[x1, x2, x3:
male(x1), water(x2), drink(x3),
event(x3), agent(x3, x1), patient(x3, x2)]

B. Brief description of GML

Graph Modeling Language (GML) [7] is a simple and efficient format for representing weighted directed graphs. A GML file is primarily a 7-bit ASCII file. Its simple format allows us to read, parse, and write without much hassle. Moreover, several open source software systems are available for viewing and editing GML files.

Graphs are represented using several keys like "graph", "node", "edge" etc. while nodes have "id" associated with them which are later referenced from the "source" and "target" attributes. Edge weights are represented through "label" attribute associated with an edge key.

C. k-NN classifier

The k-nearest-neighbour is one among the most simple and popular machine learning algorithms. These kinds of classifiers depend solely on the class labels of the training examples that are similar to the test example instead of building explicit class representation. Distance measures such as Euclidean distance, Manhattan distance are generally used to compare the similarity between two examples. In standard k-NN algorithm the majority vote of its neighbours are used to classify a new example. Usually, the number of neighbours (value of k) is determined empirically to obtain best results.

D. Distance metrics for graphs

As we have mentioned before, the goal of our current work is to make a comparative analysis of different kinds of distance metrics for text classification task.

We have taken five different distance metrics from [6], which are used in this work. They are popularly used in object recognition task, but for text categorization they have not been used popularly. For two graph $G_1$and G_2, if $d(G_1, G_2)$ is the dissimilarity/similarity measure, then $d(G_1, G_2)$ would be a distance, if d has the following properties:

(i) $d(G_1, G_2) = 0$ iff $G_1 = G_2$

(ii) $d(G_1, G_2) = d(G_2, G_1)$

(iii) $d(G_1, G_2) + d(G_2, G_3) \geq d(G_1, G_3)$

The measures that are involved in the current work follow the above rules. The corresponding distance metrics for these measures are:

$$d_{mcs}(G_1, G_2) = 1 - \frac{|mcs(G_1, G_2)|}{max(|G_1|, |G_2|)} \quad (1)$$

$$d_{wgu}(G_1, G_2) = 1 - \frac{|mcs(G_1, G_2)|}{|G_1| + |G_2| - |mcs(G_1, G_2)|} \quad (2)$$

$$d_{ugu}(G_1, G_2) = |G_1| + |G_2| - 2|mcs(G_1, G_2)| \quad (3)$$

$$d_{MMCS}(G_1, G_2) = |MCS(G_1, G_2)| - |mcs(G_1, G_2)| \quad (4)$$

$$d_{MMCSN}(G_1, G_2) = 1 - \frac{|mcs(G_1, G_2)|}{|MCS(G_1, G_2)|} \quad (5)$$

In the above equations, $mcs(G_1, G_2)$ and $MCS(G_1, G_2)$ denote maximal common subgraph and minimum common super graphs of two graphs $G_1$and G_2. Theoretically $mcs(G_1, G_2)$ is the largest graph in terms of no. of edges which is isomorphic to a subgraph of both $G_1$and G_2. The $mcs(G_1, G_2)$ has been formally defined in the work of Bunk et al. [8].

As stated earlier, finding the maximum common subgraph is a NP complete problem and, the algorithm of finding the *mcs*() is actually a brute force method, which first finds all the subgraphs of both the graphs and select the graph of maximum size which is common to both G_1 and G_2 . To increase computational speed of the program, it is modified to an approximate version of actual $mcs(G_1, G_2)$ residing on the fact that the nodes that possess a greater similarity in their local neighborhood of the two graphs have a larger probability of inclusion in the *mcs*. The two stage approach used in the present work to form the approximate $mcs(G_1, G_2)$ is as follows:

1. All the node pairs (one from each graph) are sorted according to the decreasing order of their similarity of local structures. In the present case, the number of self-loops which have equal labels in both the graphs is used for similarity measures.
2. Build the *mcs* by first adding each self-loop vertex pair (starting with the one with the highest no. of matching labels) and considering it as an equivalent vertex, then include the rest of the edges (non-self-loop edges) which satisfy the chosen self-loops in both the graphs.

In this way it can be ensured that the approximation version possesses most of the properties of a *mcs*, while complexity is contained within a polynomial upper bound.

The minimum common supergraph (MCS) [4] is formed using the union of two graphs, i.e. $MCS(G_1, G_2) = G_1 \cup G_2$. The distance metrics of Equations 1, 2, and 5 were used without modification, but those of Equations 3–4 were divided by $(|G_1| + |G_2|)$ and $|MCS(G_1, G_2) + mcs(G_1, G_2)|$, respectively to make them normalized, keeping the value of distance metrics in the range $[0, 1]$.

E. Tools

In order to extract DRS from summarized texts we used "C&C" and "Boxer" [10] [11], which are very popular open source tools available for download at http://svn.ask.it.usyd.edu.au/trac/candc. The tools consist of a combinatory categorical grammar (CCG) [9] parser and outputs the semantic

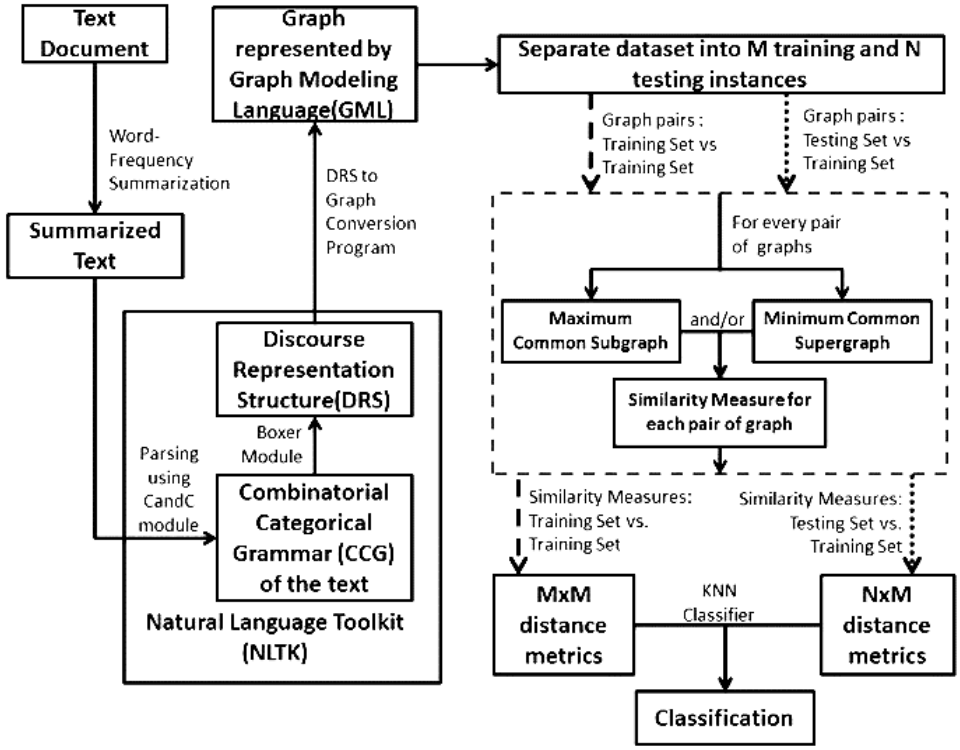

Fig 1. Block Diagram of the system showing major stages like Semantic Information Extraction, Formation of Graphs, Calculation of Distance Metrics, and Classification using k-NN Classifier

representations of texts using discourse representation structures (DRS) which are defined in Discourse Representation Theory (DRT) [5].

III. Methods

Our method involves three primary phases. The first involves extraction of semantic information from summarized documents. The second phase indulges into the conversion of DRS into a graphical structure. Finally the third one focuses on the learning phase where Distance metrics are computed on the basis of graphical structures which are further used for classification using k-NN classifier [2]. The flowchart of the entire system is shown in Fig. 1.

A. Extraction of Semantic Information

Bag-of-Words approach has been one of the most common approaches for text classification. Though the incredible amount of success achieved by this approach yet it fails to actually understand a language. A language is not merely a collection of words placed randomly. A language is defined by its grammar which binds different entities with relations that gives a document a sense of entirety with respect to its contents. Hence we have to move on from bag-of-words approaches to truly understand a language. Hence it is very essential to explore the semantic level analysis of the languages and DRT is such a framework.

However, before using DRS we need to convert it to a more dynamic data structure. We have decided to use graphs as they possess an intrinsic property that makes them suitable to represent DRS. Referents and conditions are easily represented through the nodes and edges of graph. Graphs also ensure faster traversal through semantic networks. Moreover numerous graph similarity metrics exist which can be used to compare two documents and find their similarity. Hence, a robust system may be built which can minutely observe and analyze complex semantics of natural language and efficiently categorize them.

However, we should note that the traditional $mcs()$ and $MCS()$ is a NP complete problem. To minimize the complexity, summarizations of documents are performed. Summarization is done on the basis of frequency of words. The sentences are chosen whose words occur with greatest frequency over a particular class. Throughout this process stop-words are ignored completely. The sentences are ranked in order to be able to easily choose the best ones for summarization. The summarization is done using the tool available from git://github.com/amsqr/NaiveSumm.gitmaster.

The summarized text is then sent to the C&C parser [9] to identify the CCG derivations, POS tags, lemmas and named entity tags which are then used by Boxer [9] to produce the DRSs based on the inherent semantic interpretation of the sentence.

B. Formation of Graphs

The DRS output provided by Boxer is converted to graph structure. For building the graph we used the format of Graph Modeling Language (GML). As mentioned earlier, Boxer is capable of representing various kinds of complex predicates like proposition, implication, negation etc. However, the entire

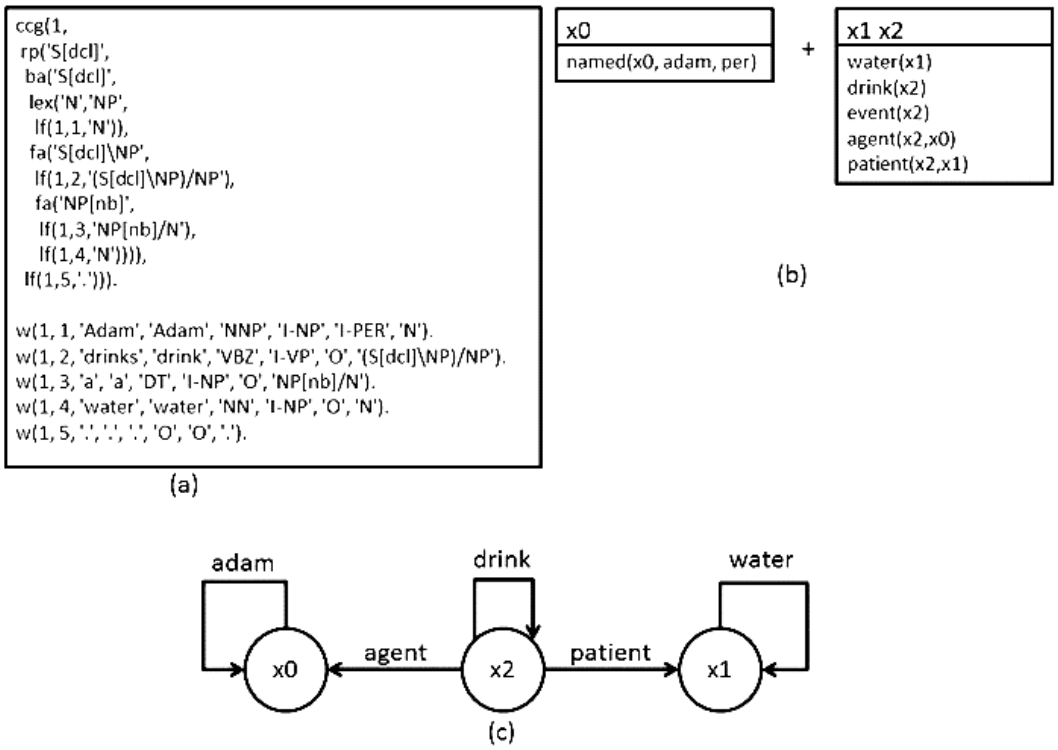

Fig 2. The transformations for the sentence "Adam drinks water": (a) C&C output, (b) Boxer output and (c) the corresponding graph

DRS structure can be broadly broken into referents or entities and conditions or relations. In the graph referents are treated as nodes and the conditions as edges. While assigning referents to vertices all equality cases are resolved beforehand. Conditions are represented as directed edges. The direction assigned is from the first referent to the second referent. In case of conditions with single referents like *male(x1)*, a self-loop is added at the vertex. Special conditions like propositions, implications are handled as conditions in the DRS and hence represented as edges between the concerned referents. Condition names are used as labels for the edges. Agent and patient are also treated as conditions of discourse, hence represented by the edge values of two referents. An example of a sentence and its transformations (syntactic, semantic and graph representation) is shown in the Fig. 2.

To measure the distance between two graphs, the approximate $mcs(G_1, G_2)$ is constructed based on the steps described in Section 2.4, it is then for the creation of $MCS(G_1, G_2) = G_1 + G_2 - mcs(G_1, G_2)$ to make it computationally faster. Fig. 3 shows the mcs and MCS of two graph sentences.

C. Classification using the different distance metrics and the k-NN classifier

It has already been mentioned that the different distance metrics (see Equations 1-5) are calculated based on the $mcs()$ and $MCS()$. The values of $mcs()$ and $MCS()$ are represented by the number of similar vertices or the number of similar edges. Thus, ten different distances are calculated based on Equations 1-5.

During the classification phase two matrices are generated for each of the above ten distance metrics. The training set is an $M \times M$ matrix formed by pairing each training data with all other training data and calculating their distance values. The testing set is a $N \times M$ matrix formed by pairing each testing data with all training data. The feature vector is hence represented as an M dimensional vector which comprises of similarity scores for each of the M training documents. The results obtained were used to evaluate the performance of each distance on the dataset (shown in Table 1 and 2).

IV. Experimental Results and Discussion

Reuters-21578 is one of the most popular text corpora that have been used for text classification. We have used a subset of this dataset. The selected dataset comprises of 20 documents (10 training and 10 testing) belonging to each of 20 selected classes, viz. *acq*, *alum*, *barley*, *bop*, *carcass*, *cocoa*, *coffee*, *copper*, *corn*, *cotton*, *cpi*, *crude*, *dlr*, *earn*, *fuel*, *gas*, *gold*, *grain*, *interest* and *ipi*. As shown in Fig 1, a summarization technique based on word frequencies is used to generate two and three sentences summarization of the entire text.

The summarized texts are then passed into the NLTK toolkit [11] where semantic information is extracted by Boxer. Then an algorithm converts DRS to graph using the format of graph modeling language. Then five different distance metrics (see Equations 1–5) are calculated on pairs of graphs, which is later used for classification using *k*-NN classifiers. The accuracies observed for the test dataset for 3, 5 and 7 nearest neighbours (*k* value) are shown in Table 1 and 2 along with a

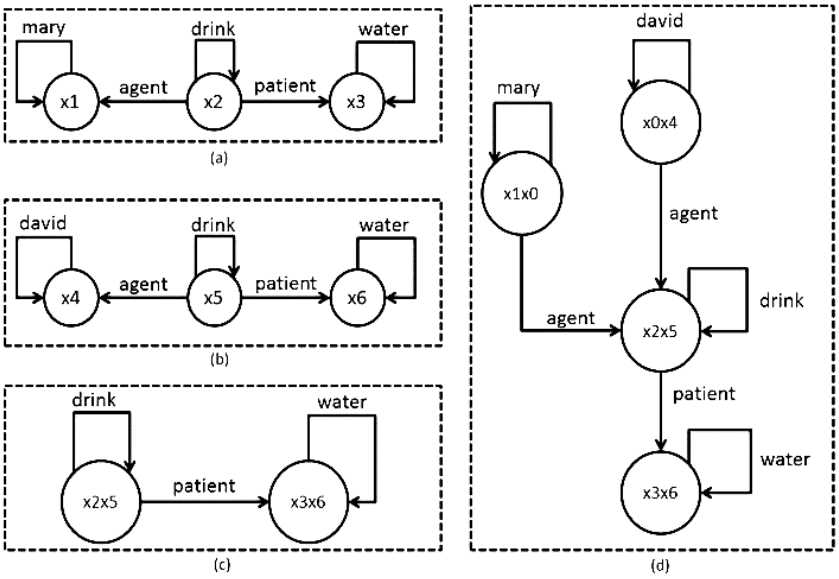

Fig 3. Graphical overview of mcs and MCS: (a), (b) graph representation of sentences meaning "Mary drinks water" and "David drinks water" (c) maximum common subgraph (d) minimum common supergraph

result for the traditional bag-of-words approach.

From Tables 1 and 2 and Fig. 4 it could be observed that all the edge based distance metrics perform better than their vertex equivalent.

Therefore, the DRS conditions or relations, which are represented by edge values, play an important role in the classification job. Since the edge values are the main indicators of the underlying semantics it can be concluded that semantic information is essential for text categorization. From Table 1 and 2 it can also be observed that average recognition accuracies of two sentences are lower than that of the three sentence summarization techniques. This can be easily visualized in Fig. 5.

The maximum accuracy observed in the present work is 51.50% for edge based experiment is for Equation 5 with 3 sentence summarization for k = 7. The minimum accuracy observed is 35.50% for edge based experiment for Equation 4 with 2 sentence summarization for k = 2. The average of 3 sentence summarization accuracy over k, observed for the five different distances with edge based calculations are 49.00 ± 1.00%, 49.50 ± 1.41%, 49.00 ± 0.87%, 49.83 ± 0.85% and 49.83 ± 2.01%. From the result it is observed distance metrics 4 and 5 provide the same average accuracy on the test dataset.

The overall average accuracy of the five types of distance metrics on 3 sentence summarized texts and calculated using edge based formulae averaged over 'k' is 49.43 ± 0.38%, which denotes that the five distances are more or less comparable based on the observed recognition accuracies.

To analyze the result further, precision, recall and F1 measures were calculated for the bag-of-words and the best graph distance (E5):

TABLE 1.
K-NN CLASSIFICATION ACCURACIES FOR TWO-SENTENCE SUMMARIZATION

Distance metric		Value of K		
		3	5	7
$d_{mcs(\,)}$	V1	**37.50%**	37.50%	37.50%
	E1	45.00%	45.50%	**49.50%**
$d_{wgu(\,)}$	V2	40.00%	40.50%	**41.50%**
	E2	45.00%	**51.00%**	50.50%
$d_{ugu(\,)}$ / $(\lvert G_1\rvert + \lvert G_2\rvert)$	V3	37.00%	38.50%	**40.00%**
	E3	44.00%	48.00%	**50.00%**
$d_{MMCS(\,)}$ / $\lvert MCS(G_1,G_2) + mcs(G_1,G_2)\rvert$	V4	35.50%	**39.50%**	39.50%
	E4	42.50%	46.50%	**49.50%**
$d_{MMCSN(\,)}$	V5	39.50%	**42.00%**	42.00%
	E5	45.00%	49.00%	**49.50%**
bow		**50.50%**	50.00%	48.50%

TABLE 2.
K-NN CLASSIFICATION ACCURACIES FOR THREE-SENTENCE SUMMARIZATION

Distance metric		Value of K		
		3	5	7
$d_{mcs(\,)}$	V1	45.50%	45.50%	**48.00%**
	E1	47.50%	49.50%	**50.00%**
$d_{wgu(\,)}$	V2	48.00%	**48.50%**	48.50%
	E2	47.50%	**50.50%**	50.50%
$d_{ugu(\,)}$ / $(\lvert G_1\rvert + \lvert G_2\rvert)$	V3	45.00%	**46.00%**	45.50%
	E3	48.00%	**49.50%**	49.50%
$d_{MMCS(\,)}$ / $\lvert MCS(G_1,G_2) + mcs(G_1,G_2)\rvert$	V4	**46.00%**	46.00%	45.00%
	E4	49.00%	**51.00%**	49.50%
$d_{MMCSN(\,)}$	V5	47.50%	48.00%	**49.00%**
	E5	47.00%	51.00%	**51.50%**
bow		**49.50%**	49.50%	49.00%

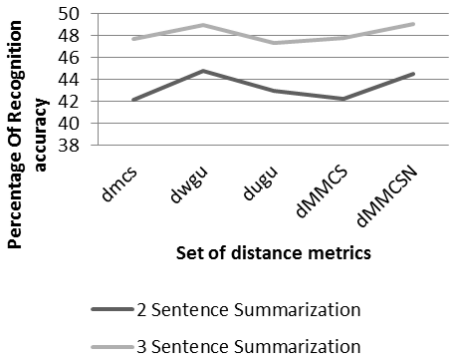

Fig 4. Recognition accuracies for vertex vs. edge based techniques

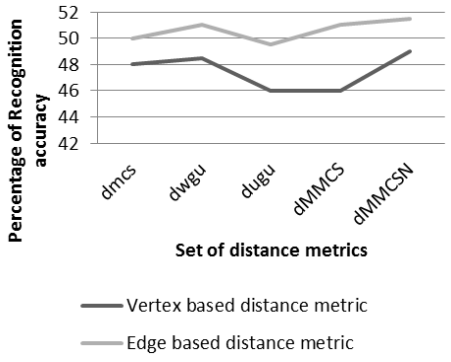

Fig 5. Recognition accuracies for 2 sentence vs. 3 sentence summarization

$$Recall = \frac{TP}{TP + FN},$$
$$Precision = \frac{TP}{TP + FP},$$
$$F\text{-}Measure = 2\frac{Recall \times Precision}{Recall + Precision}.$$

The comparative assessment of the two approaches is shown in Table 3. There it can be observed that sometimes the graph distance provides significantly better results than bag-of-words approach.

In the case of the *carcass* class the bag-of-word approach provides very satisfactory result due to having simple words like "beef" or "pork" which are enough to uniquely identify the category. On the other hand, the gold class shares some common words with other classes. The word "gold" itself can be found in copper and alum classes.

TABLE 3.
RECALL, PRECISION AND F-MEASURE FOR BAG-OF-WORDS AND GRAPH BASED SEMANTIC APPROACHES

Class	***Recall***		***Precision***		***F-Measure***	
	BOW	**Graph**	**Bow**	**Graph**	**Bow**	**Graph**
acq	0.10	0.60	0.20	0.60	0.13	0.60
alum	0.50	0.30	1.00	0.38	0.67	0.33
barley	0.60	0.50	0.46	0.50	0.52	0.50
bop	0.50	0.80	0.46	0.67	0.48	0.73
carcass	0.90	0.30	1.00	0.33	0.95	0.32
cocoa	0.60	0.60	0.86	0.86	0.71	0.71
coffee	0.50	0.50	0.46	0.46	0.48	0.48
copper	0.30	0.50	0.60	0.31	0.40	0.39
corn	0.40	0.20	0.80	0.40	0.53	0.27
cotton	0.80	0.40	0.47	0.27	0.59	0.32
cpi	0.80	0.60	0.38	0.60	0.52	0.60
crude	0.70	0.60	0.54	0.43	0.61	0.50
dlr	0.30	0.80	0.75	0.62	0.43	0.70
earn	0.40	0.80	1.00	0.89	0.57	0.84
fuel	0.30	0.50	0.19	0.39	0.23	0.44
gas	0.30	0.30	1.00	0.33	0.46	0.32
gold	0.20	0.60	0.33	0.86	0.25	0.71
grain	0.60	0.30	0.40	0.43	0.48	0.35
interest	0.50	0.30	0.28	0.60	0.36	0.40
ipi	0.90	0.80	0.75	0.80	0.82	0.80
Average	**0.51**	**0.52**	**0.60**	**0.54**	**0.51**	**0.51**

Other common words like "reserves" occur in many other classes. Moreover, there are words like "ounces" or "carat" which are overlooked in the bag-of-words approach due to their comparatively low no. of occurrences.

The use of the semantic approach enables the binding of words like "gold", "reserves", "carat" and "ounce" in such a way that they are highly unique for the gold class, giving better results. Hence, it can be strongly established that the graph distance based approach provides a much better recognition rate for textual data with semantically coherent information.

V. CONCLUSIONS AND FUTURE WORK

In the present work, we have proposed a comparative study of different graph metrics for text classification using sematic information. Our approach combines deep linguistic analysis and graph based classification techniques.

The former part of our work includes extracting discourse information from documents followed by a comprehensive similarity analysis using existent graph based distance metrics. During the calculation of the distance metrics, we have proposed an approximate version for the traditionally NP-Complete problem of finding the maximum common subgraph that is not only computationally faster but also more suited to textual similarity extraction.

Finally, we combined the graph-drs structures and the proposed distance metrics for the text classification task using a k-NN classifier. The obtained results clearly depict that the performance of most of the graph similarity metrics using our approach are more likely same. The obtained results also signify that the proposed approach is nearly equivalent to the standard bag-of-words approach. Even in some cases, it was able to outperform the approach. This result is also a good indicator of the adequacy of using semantic information to represent texts and text content.

Our future work will emphasize to analyze the impact of the summarization module in text classification task. In addition to that different machine learning algorithms, such as multi-layer perceptron, support vector machines using a graph kernel can also be applied to our proposed methodology for obtaining better results.

ACKNOWLEDGMENTS

This work was funded by Emma in the framework of the EU Erasmus Mundus Action 2.

REFERENCES

[1] S. Bleik, "Text Categorization of Biomedical Data Sets Using Graph Kernels and a Controlled Vocabulary," *EEE/ACM Trans. Comput. Biol. BioinformaticsI*, vol. 99, p. 1, Mar. 2013.

[2] L. Zhang, Y. Li, C. Sun, and W. Nadee, "Rough Set Based Approach to Text Classification," *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 3, 2013, pp. 245–252.

[3] Z. Wang and Z. Liu, "Graph-based KNN text classification," *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, vol. 5, 2010, pp. 2363–2366.

[4] R. Angelova and G. Weikum, "Graph-based Text Classification: Learn from Your Neighbors," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006, pp. 485–492.

[5] H. Kamp and U. Reyle, *From Discourse to Logic: An Introduction to Model Theo-retic Semantics of Natural Language, Formal Logic and Discourse Rep-resentation Theory*. Kluwer, Dordrecht: D. Reidel, 1993, p. 717.

[6] "Graph Matching," in *Graph Classification and Clustering Based on Vector Space Embedding*, vol. Volume 77, WORLD SCIENTIFIC, 2010, pp. 15–34.

[7] M. Himsolt and G. Iversität Passau, 94030 Passau, "GML: A portable Graph File Format," 1996.

[8] H. Bunke, P. Foggia, C. Guidobaldi, C. Sansone, and M. Vento, "A Comparison of Algorithms for Maximum Common Subgraph on Randomly Connected Graphs," in *Structural, Syntactic, and Statistical Pattern Recognition SE – 12*, vol. 2396, T. Caelli, A. Amin, R. W. Duin, D. Ridder, and M. Kamel, Eds. Springer Berlin Heidelberg, 2002, pp. 123–132.

[9] J. Curran, S. Clark, and J. Bos, "Linguistically Motivated Large-Scale NLP with C&C and Boxer," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, 2007, pp. 33–36.

[10] J. Bos, "Wide-Coverage Semantic Analysis with Boxer," in *Semantics in Text Processing. STEP 2008 Conference Proceedings*, 2008, pp. 277–286.

[11] E. L. Steven Bird, Ewan Klein, *Natural Language Processing with Python*. O'Reilly Media, 2009, p. 504.