A Comprehensive Analysis of using Semantic Information in Text Categorization

Kerem Çelik
Department of Computer Engineering
Boğaziçi University
Istanbul, Turkey
celikerem@gmail.com

Tunga Güngör
Department of Computer Engineering
Boğaziçi University
Istanbul, Turkey
gungort@boun.edu.tr

Abstract—**Traditional text categorization methods only deal with the content of the documents and use some statistic based metrics to represent the documents. The representation is then used by a machine learning approach to determine the document class. In this picture, the meaning of the document is missing. In order to add meaning into the text categorization process, we start with using part-of-speech tagging (POS). As expected, in a document each part-of-speech tag does not contribute the same amount of information to the document meaning. In addition to the POS information, we make use of WordNet to add semantic features such as synonyms, hypernyms, hyponyms, meronyms and topics into classification process. Using WordNet's semantic features introduces ambiguity and not all semantic features are really related to the document content. To overcome this problem, we introduce a new method to eliminate the ambiguity. Various combinations of POS, WordNet and word sense disambiguation are applied and the results show that using semantic features perform better than the traditional, context based methods.**

Keywords-component; text categorization, semantic, wordnet, pos tagging, word sense disambiguation

I. INTRODUCTION

Since the early 1990s, the accessibility and abundance of digital documents make text categorization an important and necessary research field. Today, text categorization is being applied in many contexts in order to organize and manipulate documents. The arrival of machine learning methods in text categorization is one of the essential factors that improve the effectiveness of text categorization.

When we consider all the traditional text categorization techniques, the meaning of text is missing in the picture. By using semantic features for categorization, better results may be obtained. Determining part-of-speech tag of a term is critical information since not all word forms tell same about the document content. For example, nouns are usually more descriptive than adverbs. In addition, including *WordNet* features into document's original terms can be important especially when there are few terms in the document. In this way, by increasing the number of terms in the document, the documents will be represented better. Incorporating those features and information into a document makes it richer in terms of content. Even though the performance may degrade, better categorization accuracy will be taken. Besides the positive effects of using WordNet semantic features, there is the ambiguity problem. Since WordNet cannot know the context of the document, some of the semantic features are incorrectly taken. We introduce a new method to eliminate the ambiguity.

In this work, we build a framework to investigate the effectiveness of using semantic information in text categorization. In the framework all the combinations of POS tags and WordNet relations are examined and the results are compared with traditional methods. In addition, contribution of the word sense disambiguation method is evaluated by using two different relations, hypernyms and topics. The results show that using semantic information yield better performance in terms of accuracy.

II. RELATED WORKS

The arrival of machine learning methods in the text categorization domain is one of the most important factors that accelerate the improvement in this field. Reducing dimensionality is another critical issue in text categorization which can be done by feature selection methods. Feature selection improves the efficiency and accuracy of the classifiers by selecting only more discriminative terms in a dataset as features. In the literature, various feature selection methods have been presented and analyzed [1].

In [6], the comparison of using morphological and ontological information in text categorization is studied. Stemming is used as morphological information, whereas, hypernyms based representation is used for ontological analysis. It is asserted that the morphological method performs better than the other. In a similar study, contextual and conceptual features are incorporated and the results are compared [7]. In contrast to the previous work the authors show that using conceptual information performs better than contextual information.

In the studies [8] and [9], WordNet relations are incorporated and the effectiveness of the relations is evaluated.

978-1-4799-0661-1/13/$31.00 ©2013 IEEE

It is shown that the use of synonym and hypernym relations increase the accuracy. Moreover, in [10], WordNet is used to understand the category names and the work has a clearer representation of topics described in the content of the documents. In addition, [11] tries to build a better representation of documents compared to the bag of words method by incorporating WordNet relations. Similarly, study in [12] attempts to categorize the documents by measuring the similarity between documents and categories. They annotate topics and documents with WordNet relations and obtain more accurate results. Nonetheless, [13] investigates the effectiveness of POS tags and uses only nouns as its semantic term origin. WordNet relation of the term is provided to select terms based on overlapping relations. The more the number of overlapping relations, the more the term is related to the document.

The semantic features can be used in addition to the original terms of document. The studies in [14] and [15] investigate the effectiveness of the using both original and semantic features together and claim that using these data together gives better results. In this study we make use of this methodology named as boosting.

III. Proposed Methodology

This section discusses the contribution of semantic features in text categorization. Various researches have been done to improve the performance of text categorization. In the field of text categorization the majority of studies focused on feature selection metrics and different types of classifiers. In this work we focus on contribution of semantic features rather than feature selection and machine learning techniques.

A. Part-of-Speech Tagging

Every term in the document has a part-of-speech tag such as noun, verb, etc. In this work we only consider noun, verb, adjective and adverb. It is clear that not all the word forms contribute to the meaning of the document in the same amount. For example, adverbs are kind of transition words and do not tell much about the content in the document, whereas nouns tell much. Thus we make a comprehensive analysis related to each word form in text categorization. Moreover, in order to find WordNet relations, it is essential to have POS of a term. Thus POS of a term will be both analyzed and used in the next step. We use a lexicon-based part-of-speech tagger given in [4]. Given a sentence the tagger returns the tokenized terms with part-of-speech information.

B. WordNet Relations

We use WordNet to find the synonym, hypernym, hyponym, meronym and topic information of a given word. WordNet stores terms in synsets, and each synset has relations to other synsets. WordNet has other types of relations but in this study we use these five relations since they carry significant information about the content. The semantic features are included in the term list just as other terms found in the document. One of the problems that can be faced when the semantic features are used is that not all synsets are really related to the context of the document. Thus word sense disambiguation is applied as will be explained in the next subsection.

C. Word Sense Disambiguation

In natural languages, most of the words posses several meaning (synsets in WordNet) and all the synset of a word are not related to the content of a document [5]. There are many synsets for every word and we cannot really say that all the synsets are related to the context of the document. WordNet tries to do the best it can to disambiguate the irrelevant synsets. But it still does not know the context the term is in. We have to find a mechanism to eliminate the non-related synsets for the document.

A term can be represented as

$$t_i = \{s_1, s_2, ..., s_m\}$$

where t_i is the set of synset of the term and s_i is any synset that contains synonyms, hypernyms, hyponyms, meronyms and topics. And a document can be represented as

$$d_i = \{t_1, t_2, ..., t_n\}$$

as a union of synset of each term, document synset can be represented as

$$dS_i = \{s_1, s_2, ..., s_k\}$$

We come up with a scoring metric to calculate a score for all synsets in dS_i and then, apply a threshold to select the synsets that receives the best scores. For score calculation please see the next section.

1) Disambiguation Score Calculation

In this section we explain score calculation of each synsets in the document. We will use either hypernyms or topics for score calculation. Hypernyms denotes root or more general concept of the words. Where as topics tells us topic information of words. We will analyze both of them for disambiguation calculation process. Once we identify the synset we will calculate score of every synset by calculating the similarity of the synset with all other synsets. Thus, we will use the total similarity as score of the synset. After scoring phase, we will simply apply a threshold to select the synsets of documents.

$$Score(s_i) = \sum_{j=0, i \neq j}^{k} Similarity(s_i, s_j)$$

where *Score*(s_i) denotes the scores of s_i in d_i and *Similarity*($s_{i,}$ *sj)* is defined as

$$Similarity(s_i, s_j) = CommonCount(Hypernyms(s_i), Hypernyms(s_j))$$

$$Similarity(s_i, s_j) = CommonCount(Topics(s_i), Topics(s_j))$$

where *Hypernyms*(s_i) and *Topics*(s_i) denotes the hypernym term list and topic term list respectively for synset s_i and *CommonCount(HypernymsS*$_{i,}$ *HypernymsS*$_j$*)* denotes the number of common terms in the given two list.

IV. EXPERIMENTAL SETUP & RESULTS

In this study, we use support vector machine (SVM) as a classifier which outperformed other classification methods in text categorization consistently in previous studies [1, 2]. We perform our experiments on five standard datasets, widely used in text categorization research

TABLE I. PROPERTIES OF DATASETS

Datasets	# of docs	# of terms	# of classes	Skewness
20Newsgroup	18846	90812	20	0.11 (Homogenous)
Classic3	3891	10930	3	0.14 (Homogenous)
7Sectors	3308	56314	7	0.45 (Skew)
WebKB	5396	102285	4	0.81 (High Skew)
Reuters-21578	12902	20308	90	3.32 (High Skew)

A. *Performance Measures*

To evaluate the performance of the contribution of text categorization, we use the commonly used F-measure metric which is equal to the harmonic mean of recall ρ and precision π [3]. Micro-averaged F-measure gives equal weight to each document. On the other hand, macro-averaged F-measure gives equal weight to each category regardless of its frequency and thus it is influenced more by the classifier's performance on rare categories.

B. *Results and Discussions*

1) *Contribution of Part-of-Speech Tagging*

As we explained earlier in the proposed methodology we only consider four main tags: noun, verb, adjectives and adverbs. We investigate the contribution of each tag individually and combination of two, three and four tags.

TABLE II. CONTRIBUTION OF POS (MICRO-F)

Configuration	20 Newsgroup	7 Sector	Web KB	Reuters 21578
Raw	0.8753	**0.5886**	0.7634	0.8467
Noun (N)	0.8620	0.5734	0.7681	0.8415
Verb (V)	0.5668	0.4498	0.7336	0.6573
Adjective (Adj)	0.4845	0.4913	0.6893	0.6593
Adverb (Adv)	0.2183	0.3969	0.5785	0.2029
N+V	0.8644	0.5752	0.7750	0.8419
N+Adj	0.8610	0.5625	0.7599	0.8431
N+V+Adj	0.8648	0.5561	0.7646	0.8460
N+V+Adj+Adv	0.8650	0.5643	0.7715	0.8465
Raw+Noun	**0.8813**	0.5634	0.7795	0.8513
Raw+Verb	0.8755	0.5725	0.7564	0.8428
Raw+Adjective	0.8751	0.5734	0.7576	0.8461
Raw+Adverb	0.8751	0.5851	0.7646	0.8469
Raw+N+V	0.8799	0.5486	**0.7841**	0.8502
Raw+N+Adj	0.8796	0.5597	0.7715	**0.8532**
Raw+N+V+Adj	0.8795	0.5579	0.7738	0.8511
Raw+N+V+Adj+Adv	0.8798	0.5570	0.7750	0.8507

In Tables 2 and 3, the results of all combinations can be seen. Note that we exclude Classic3 dataset results for this section and in later sections since their values are all between 0.99 and 1. The results named with Raw are the base line for this study which does not consider any semantic information at all. First, we note that using POS increases both Micro-F and macro-F measure for all datasets expect 7Sectors. It can be said that raw features alone do give us good results, but using with some specific part-of-speech gives better results. Considering the results we can not come up with a single combination configuration but we can say that nouns usually give better results. In addition, verbs and adjectives increase both Micro-F and Macro-F measures considerably.

TABLE III. CONTRIBUTION OF POS (MACRO-F)

Configuration	20 Newsgroup	7 Sector	Web KB	Reuters 21578
Raw	0.8617	**0.5312**	0.8968	0.3837
Noun (N)	0.8458	0.5142	0.8877	0.3781
Verb (V)	0.5413	0.3527	0.5419	0.1395
Adjective (Adj)	0.4645	0.3808	0.5599	0.1603
Adverb (Adv)	0.2085	0.2245	0.3672	0.0477
N+V	0.8480	0.5099	0.8922	0.3778
N+Adj	0.8463	0.5076	0.8908	0.3894
N+V+Adj	0.8498	0.4871	0.8882	0.3841
N+V+Adj+Adv	0.8501	0.5006	0.8943	0.3838
Raw+Noun	**0.8693**	0.5037	0.8850	0.4161
Raw+Verb	0.8619	0.5138	0.8883	0.3804
Raw+Adjective	0.8623	0.5113	**0.9025**	0.3904
Raw+Adverb	0.8616	0.5289	0.8972	0.3922
Raw+N+V	0.8677	0.4853	0.8863	0.4218
Raw+N+Adj	0.8681	0.5012	0.8953	0.4243
Raw+N+V+Adj	0.8680	0.4948	0.8894	**0.4268**
Raw+N+V+Adj+Adv	0.8683	0.4929	0.8963	0.4227

2) *Contribution of WordNet features*

In this section the results of using WordNet features in text categorization will be given and discussed. We will also measure the contribution of using or not using WordNet features. The system is built on top of POS and one of the configurations (Raw, Noun, Verb and Adjective) that give good results (The results can be seen in the previous section) is taken. And then contribution of using WordNet features is investigated. In Tables 4 and 5, the results show that incorporating those features increases the performance of the classification accuracy. We can say that using synonyms and hypernyms always increases both Micro-F and Macro-F measures, and usually using all of the features gives the best results.

TABLE IV. CONTRIBUTION OF WORDNET (MICRO-F)

Configuration	20 Newsgroup	7 Sector	Web KB	Reuters 21578
Synonyms	0.8860	0.5161	0.7795	0.8429
Hypernyms	0.9063	0.5402	0.7974	**0.8516**
Hyponyms	0.8886	0.5003	0.7818	0.8332
Meronyms	0.8834	0.5616	0.7773	0.8454
Topics	0.8950	0.5383	0.7874	0.8499
Syn+Hype	0.8924	0.5467	0.7852	0.8436
Hype+Top	0.8975	0.5268	0.7897	0.8504
Syn+Hype+Top	0.8834	0.5335	0.7773	0.8444
All Features	**0.9088**	**0.5652**	**0.7996**	0.8263
No Sem. Features	0.8795	0.5579	0.7738	0.8511

3) *Contribution of Word Sense Disambiguation*

In this section the contribution of applying word sense disambiguation will be investigated. Disambiguation process is controlled by a threshold value that represents the elimination amount. It can be seen in the results that there are three selected thresholds: 100% (No disambiguation), 70%, 50% and 30%.

TABLE V. CONTRIBUTION OF WORDNET (MACRO-F)

Configuration	20 Newsgroup	7 Sector	Web KB	Reuters 21578
Synonyms	0.8686	0.4267	0.8901	**0.4499**
Hypernyms	0.8840	0.4481	0.9058	0.4350
Hyponyms	0.8794	0.4547	0.9012	0.4143
Meronyms	0.8842	0.4730	0.9060	0.4238
Topics	0.8719	0.4611	0.8935	0.4201
Syn+Hype	0.8779	0.4572	0.8996	0.4272
Hype+Top	0.8780	0.4303	0.8997	0.4348
Syn+Hype+Top	0.8756	0.4402	0.8972	0.4321
All Features	**0.8942**	**0.5036**	**0.9164**	0.3937
No Sem. Features	0.8680	0.4948	0.8894	0.4268

TABLE VI. 20 NEWSGROUP WSD WITH TOPICS (MICRO-F)

Configuration	No Disambig	70%	50%	30%
Synonyms (Syn)	0.8860	0.8922	**0.8924**	0.8913
Hypernyms (Hype)	0.9063	0.9067	0.9072	**0.9089**
Hyponyms (Hypo)	0.8886	0.9051	0.9062	**0.9062**
Meronyms (Mero)	0.8834	0.8888	**0.8899**	0.8886
Topics (Top)	0.8950	0.8954	**0.8954**	0.8955
Syn+Hype	0.8924	0.8962	0.8984	**0.9016**
Hype+Top	0.8975	0.8970	0.8973	**0.8979**
Syn+Hype+Top	0.8834	0.8854	0.8876	**0.8887**
All	0.9088	0.9346	0.9354	**0.9377**

TABLE VII. 20 NEWSGROUP WSD WITH TOPICS (MACRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	**0.8686**	0.8200	0.8177	0.8164
Hypernyms (Hype)	**0.8840**	0.8646	0.8569	0.8718
Hyponyms (Hypo)	0.8794	0.8869	**0.8953**	0.8909
Meronyms (Mero)	0.8842	0.8924	**0.8970**	0.8950
Topics (Top)	0.8719	0.8856	**0.8941**	0.8929
Syn+Hype	0.8779	0.8729	0.8768	**0.8786**
Hype+Top	**0.8780**	0.8396	0.8383	0.8446
Syn+Hype+Top	**0.8756**	0.8714	0.8664	0.8642
All	0.8942	0.9132	**0.9274**	0.9256

In table 6 and 7 the results of using topics disambiguation for 20 Newsgroup dataset are given. Results show that applying disambiguation with topics information always increases Micro-F measure. But when it comes to Macro-F, we cannot say the same. But still more than half of the options show that Macro-F measure is increased.

TABLE VIII. 7 SECTOR WSD WITH HYPERNYMS (MICRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	0.5161	**0.5449**	**0.5449**	0.5258
Hypernyms (Hype)	0.5402	0.5373	**0.5449**	0.5392
Hyponyms (Hypo)	0.5003	**0.5542**	0.5458	0.5420
Meronyms (Mero)	**0.5616**	0.5514	0.5326	0.5316
Topics (Top)	0.5383	0.5449	**0.5570**	0.5458
Syn+Hype	0.5467	0.5354	0.5551	**0.5597**
Hype+Top	0.5268	0.5402	0.5439	**0.5542**
Syn+Hype+Top	0.5335	0.5326	**0.5542**	0.5268
All	**0.5652**	0.5383	0.5326	0.5306

TABLE IX. 7 SECTOR WSD WITH HYPERNYMS (MACRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	0.4267	0.4707	**0.4840**	0.4440
Hypernyms (Hype)	0.4481	0.4624	**0.4729**	0.4620
Hyponyms (Hypo)	0.4547	0.4613	0.4459	**0.4789**
Meronyms (Mero)	0.4730	**0.4864**	0.4545	0.4523
Topics (Top)	0.4611	0.4767	**0.4946**	0.4440
Syn+Hype	0.4572	0.4353	0.4813	**0.4969**
Hype+Top	0.4303	0.4627	0.4741	**0.4927**
Syn+Hype+Top	0.4402	0.4267	**0.4816**	0.4497
All	**0.5036**	0.4582	0.4411	0.4307

In Tables 8 and 9, the results of using hypernyms disambiguation for 7 Sectors dataset are given. Results show that applying disambiguation with hypernyms information always increases both Micro-F and Macro-F measures expect the last option. In addition, it can be said that 50% disambiguation gives better results than others.

TABLE X. WEBKB WSD WITH HYPERNYMS (MICRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	0.7807	0.7963	**0.7996**	0.7829
Hypernyms (Hype)	0.7985	**0.8105**	0.8062	0.7952
Hyponyms (Hypo)	0.7908	**0.8018**	0.7919	0.7841
Meronyms (Mero)	0.7795	**0.7852**	0.7773	0.7818
Topics (Top)	0.7738	**0.7795**	0.7681	0.7738
Syn+Hype	0.7996	0.8084	**0.8127**	0.8029
Hype+Top	0.8105	0.8180	0.8127	0.7886
Syn+Hype+Top	0.7974	0.8062	**0.8094**	0.7941
All	**0.8224**	0.8191	0.8073	0.8084

TABLE XI. 7 WEBKB WSD WITH HYPERNYMS (MACRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	0.8850	**0.9083**	0.9049	0.8884
Hypernyms (Hype)	0.8915	0.9118	**0.9140**	0.9036
Hyponyms (Hypo)	0.8995	**0.9045**	0.9035	0.8914
Meronyms (Mero)	0.8979	**0.8996**	0.8927	0.8947
Topics (Top)	0.8801	**0.8966**	0.8893	0.8910
Syn+Hype	0.9076	**0.9138**	0.9115	0.9031
Hype+Top	0.8938	0.9112	**0.9124**	0.9006
Syn+Hype+Top	0.9069	0.9041	**0.9078**	0.9005
All	0.8947	**0.9096**	0.9044	0.8960

In Tables 10 and 11, the results of using hypernyms disambiguation for WebKB dataset are given. The results are very similar to results of 7 Sectors dataset, but for this dataset, 70% disambiguation is better than the others.

TABLE XII. REUTERS 21578 WSD WITH TOPICS (MICRO-F)

Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	0.8436	0.8495	**0.8497**	0.8486
Hypernyms (Hype)	0.8504	0.8507	0.8513	**0.8529**
Hyponyms (Hypo)	0.8332	0.8486	**0.8497**	**0.8497**
Meronyms (Mero)	0.8449	0.8500	**0.8511**	0.8499
Topics (Top)	0.8504	0.8507	0.8507	**0.8509**
Syn+Hype	0.8440	0.8476	0.8497	**0.8527**
Hype+Top	0.8516	0.8511	0.8514	**0.8520**
Syn+Hype+Top	0.8458	0.8477	0.8499	**0.8509**
All	0.8233	0.8467	0.8474	**0.8495**

TABLE XIII. REUTERS 21578 WSD WITH TOPICS (MACRO-F)

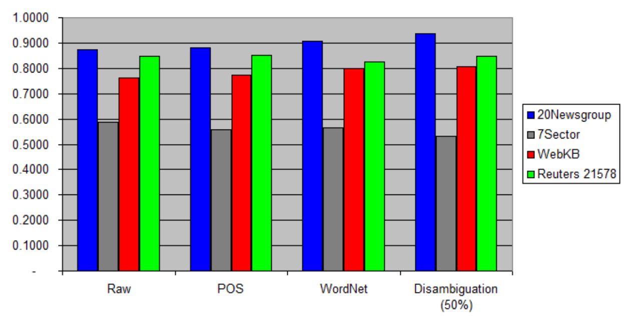
Configuration	No Disambig.	70%	50%	30%
Synonyms (Syn)	**0.4502**	0.4250	0.4238	0.4231
Hypernyms (Hype)	**0.4382**	0.4286	0.4248	0.4322
Hyponyms (Hypo)	0.4173	0.4208	**0.4248**	0.4228
Meronyms (Mero)	0.4197	0.4236	**0.4258**	0.4248
Topics (Top)	0.4174	0.4240	**0.4280**	0.4275
Syn+Hype	0.4287	0.4262	0.4281	**0.4290**
Hype+Top	**0.4462**	0.4267	0.4260	0.4292
Syn+Hype+Top	**0.4314**	0.4293	0.4269	0.4258
All	0.4123	0.4211	**0.4276**	0.4268

In Tables 12 and 13, applying disambiguation by using topics for Reuters-21578 dataset is given. For this configuration the results are different than other three above. It can be clearly said that applying disambiguation increases Micro-F measure with more than 50% disambiguation. But we cannot say the same for Macro-F measure. One of the possible reasons for this might be the skewness of the dataset.

C. Summary of Results

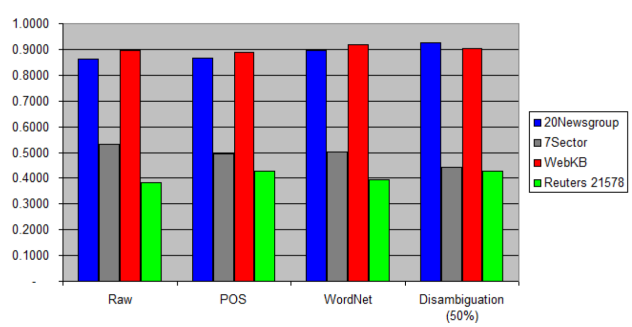
In contribution of part-of-speech tagging section, the results show that using POS of term increases both Micro and Macro F measures. And it can be concluded that using POS with raw terms gives better results than not using them. In addition, using noun, adjective and verb result in better results and using adverb decreases the categorization performance.

Figure 1. Over All Results (Micro-F)



In contribution of WordNet features section, it can be said that, for all dataset, using WordNet features increases both Micro-F and Macro-F measures. There are five different features in the results: synonyms, hypernyms, hyponyms, meronyms and topics. We can say that using all of them together gives better results than using individually or combination of them. But if we have to give the order of importance, when all the results are evaluated, it is: hypernyms, synonyms, topics, hyponyms and meronyms. In contribution of word sense disambiguation section we gave results of applying disambiguation compared with no disambiguation. The results show that applying disambiguation increases the both Micro-F and Macro-F measure.

Figure 2. Over All Results (Macro-F)



V. Conclusion And Future Works

In this study, we present a comprehensive analysis of using semantic features in text categorization. Firstly, we analyze using POS Tag of words with and without raw terms and evaluate the performance of classification. The results show that using POS tagging without raw features rarely gives better results, but with raw features achieve best results. In addition, we analyze the use of WordNet features; synonyms, hypernyms, hyponyms, meronyms and topics. And results show that using synonyms, hypernyms, hyponyms and topics gives better results, but we cannot say that using meronyms increases the metrics. Finally, to eliminate the ambiguity, we propose a disambiguation method that gains better results, especially in Micro-F measure, when compared to no disambiguation option. As future work, we will make use of other WordNet's features such as holonyms, troponym, entailment, etc. In addition, greedy based disambiguation algorithm that calculates the score of synsets by measuring the similarity can be improved. Moreover, more datasets can be used to show success of the proposed methodologies.

References

[1] G. Forman, “An extensive empirical study of feature selection metrics for text classification.” Journal of Machine Learning Res. 3, March 2003, pp.1289-1305.

[2] T. Joachims, “Text Categorization with Suport Vector Machines: Learning with Many Relevant Features”., Proceedings of the 10th European Conference on Machine Learning, ECML’98 pp.13-142, London, 1998.

[3] C. D. Manning, P. Raghavan, and H. Schtze, Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA, 2008.

[4] M. Watson, 2012, http://www.markwatson.com/opensource.

[5] Wikipedia, 2012, http://en.wikipedia.org/wiki/Wikipedia:Disambiguation.

[6] C. Koirala and K. Rasheed, “Comparison of the effects of morphological and ontological information on text categorization”. IEEE, In Machine Learning and Applications, 2008. ICMLA’08.Seventh International Conference, pp.783–786, 2008.

[7] L. S. Jensen and T. Martinez, “Improving text classification by using conceptual and contextual features”, PhD thesis, Citeseer, 2000.

[8] K. Zhang, J. Sun, and B. Wang. "A WordNet-based approach to feature selection in text categorization", Intelligent Information Processing II. Chapter, Springer-Verlag, London, UK, pp.475–484, 2005.

[9] J. Li, Y. Zhao, and B. Liu, "Fully automatic text categorization by exploiting wordnet.", In Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology, AIRS ’09, pp.1–12, Berlin, Heidelberg, 2009. Springer-Verlag.

[10] T. N. Mansuy and R. J. Hilderman. "Evaluating WordNet Features in Text Classification Models.", In The Florida AI Research Society Conference, pp.568–573, 2006.

[11] Z. Elberrichi, A. Rahmoun, and M. A. Bentaalah, "Using wordnet for text categorization.", The International Arab Journal of Information Technology, 5(1):16–24, 2008.

[12] A. Bawakid and M. Oussalah, "A semantic-based text classification system.", In Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference, pp.1 –6, 2010.

[13] S. Chua and N. Kulathuramaiyer, "Semantic feature selection using wordnet.", In Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence, WI’04, pp.166–172, Washington, DC, USA, 2004. IEEE Computer Society.

[14] S. Bloehdorn and A. Hotho, "Boosting for text classification with semantic features.", In Proceedings of the 6th international conference on Knowledge Discovery on the Web: advances in Web Mining and Web Usage Analysis, WebKDD’04, pp.149–166, Berlin, Heidelberg, 2006. Springer-Verlag.

[15] E. Ferretti, M. Errecalde, and P. Rosso, "The influence of semantics in text categorisation: A comparative study using the k nearest neighbours method.", In IICAI, pp.749–768, 2005.