India-Norway Workshop on Web Concepts and Technologies, Trondheim, Norway, October 3rd, 2011

Graph-based Representations for Text Classification

Kjetil Valle, Pinar Öztürk
kjetil.valle@bekk.no, pinar@idi.ntnu.no
Department of Computer Sciences
Norwegian University of Science and Technology
Trondheim, Norway

Abstract—**This paper presents a graph-based method for document representation intended for text classification with the vector space model. Terms are weighted by their centrality in networks constructed from the text. We evaluated a wide range of centrality measures, and experimented with two graph representations — co-occurrence networks and dependency networks. We compared the graph-based representations to the classical term frequency (TF) and term frequency-inverse document frequency (TF-IDF) based representations for classification. The graph-based representations performed better than the frequency-based measures on two datasets with different characteristics. We also found that representations considering only information local to each document, analogous of the TF measure, outperformed those including global information about the entire document collection similar to TF-IDF.**

1. Introduction

This paper investigates graph-based term weighting methods as an alternative to the classical TF and TF-IDF term weighting measures predominantly used in Vector Space Models (VSM) in information retrieval. TF and TF-IDF are examples of the so called *bag of words* (BoW) representations. In BoW approaches, a document is typically represented as an unordered set of terms and the frequency of the terms in the document. Hence, a term independence assumption underlies the BoW models. It has been maintained that there is more information in a text than only term occurrences or frequencies and that this should be taken into consideration. The impact of the term order has been a popular issue in this respect and relationships between the terms in general is claimed to play an important role in text processing [26], [10], [20]. The work presented in this paper focuses on the use of graph-based representations to capture various term relations. More specifically, co-occurrence information and term dependencies have been used in the assessment of node centrality of terms, as a method for discovering important terms. A large number of centrality measures, based on degree, closeness, betweenness, and eigenvectors [9], [8], are investigated for two types of networks. One captures co-occurrence of terms within sentences while the other type of network is constructed using a dependency grammar.

Graphs are used to measure the influence of terms in documents using the notion of centrality. The centrality values are then used to represent the documents as vectors, similar to the classical VSM. Centrality has been assessed on both local and global basis in order to weight terms. The local measure is defined for a single document, while the global measure incorporates information about a term's overall importance in the entire document collection.

In this work, we have used the cosine metric to measure the distance between document vectors.

Text classification has been used as the application task in the experiments, which has been conducted using two different data sets. The results show that graph-based term weighting methods may perform better than classical frequency-based models. We find that local representations, both TF and local graph-based representations, perform considerably better than their global counterparts. Although our experiments are not entirely conclusive, as we have so far experimented with only two datasets, the preliminary results are informative.

The paper starts with a brief description of related work in the next section. Our representation is next described in Section 3. Section 4 presents our experiments, while the results are discussed in Section 5. Section 6 finally sums up the main conclusions and points to some possible future research directions.

2. Related Work

We are not the first to look into the idea of graph-based representations for text.

Mihalcea and Tarau study graph-based representations in their TextRank system [15], [14], where they also employ the use of centrality measures. They do not focus on text classification, however, and their representation of nodes and edges consequently differ from ours.

TextRank is mainly used as a system for text summarization based on sentence extraction. This is done by using sentences as nodes, and extracting the most central sentences in the document as the summary. Edges in the sentence-networks reflects overlapping terms between sentences. Of the three measures tested — PageRank, HITS and the Positional Power Function [11] — they select PageRank as their centrality measure.

It is indicated that the approach could be used for other text processing tasks as well, by representing other textual units than sentences as nodes in the networks. As an example of this, they demonstrate unsupervised keyword extraction using TextRank. In [16] they describe a similar approach to the problem of word sense disambiguation based on use of PageRank on networks created from semantic relations in text.

The TextRank system is in turn influenced by the LexRank system by Erkan and Radev [7]. LexRank is basically the same as TextRank when applied to sentence extraction/summarization, except that the edges are created in a different way. In LexRank, the presence and strength of an edge between two sentences is determined using cosine of TF-IDF vectors representing each sentence.

Liu et al. [13] further discuss the idea of using centrality on text networks to perform language processing tasks, including classification. They focus on nodes representing terms, and discuss three such networks: co-occurrence, dependency and semantic networks. PageRank is used for term weighting also here. Our work differs from theirs in that we make a far more thorough evaluation based on a wide range of centrality measures and two types of network representations for classification.

Wang et al. [27] also use co-occurrence graphs for text classification. They do not use the same VSM methods for classification, but instead define their own similarity measures. One of these measures is based on correlation between PageRank centralities, another on distances between terms in the networks.

Other tasks similar graph-based representations have been applied to include information retrieval [22], [23] and word sense disambiguation [17], [24].

To summarize, graph-based approaches are beginning to be taken into use for a wide range of text processing tasks. Especially sentence-centrality has proven useful for text summarization. The preliminary studies done on graph-based approaches for classification so far also seems promising, although only co-occurrence networks and PageRank have been studied until now.

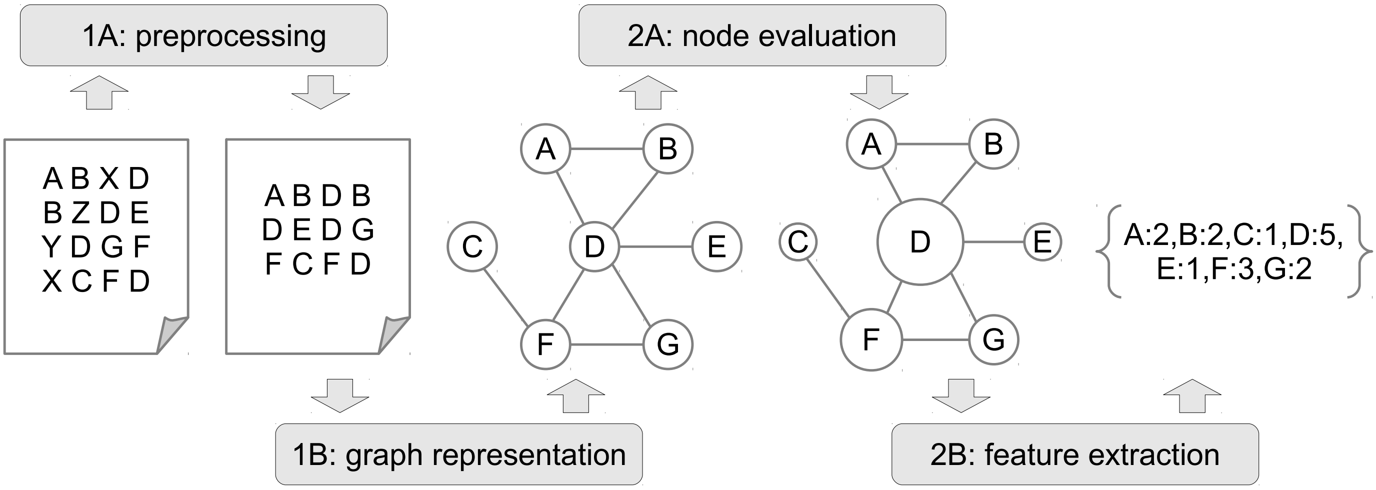

Fig. 1. The main steps of the network representation process.

3. Representation

Fig. 1 illustrates the main stages of the process which transforms the free text representations to vector representations of the document. In step 1, the text is preprocessed and represented as a network structure. Subsequently, in step 2, the graph representation is converted to a vector space representation which can then be used in different application tasks such as classification, word sense disambiguation, summarization and information retrieval. In the initial stages, i.e. during network construction, we have used two types of information: co-occurrence and term dependency. In the final stages, i.e., translation of the network to a vector representation, we have tested different centrality measures to find out which one(s) best captures the term weights.

3.1. *Network Construction*

Co-occurrence and dependency networks have much in common, but they also differ in crucial ways. Nodes in both networks correspond to terms in the documents. The edges of the networks are intended to capture some information about the relationship between the terms they link together. The main difference between the two network types is the type of information represented by the edges. While co-occurrence networks include edges between all terms co-occurring within some context, the dependency network representation is interested only in those co-occurrences that meet certain requirements of syntactic dependency.

In our experiments neither representation use directed edges, since we empirically detected [25] that undirected networks generally performed better than directed ones.

Before the networks are constructed, the texts are preprocessed. Preprocessing includes stemming and case folding. In the case of co-occurrence networks, we also perform stop-word removal to exclude the least meaningful terms. This is not done in the dependency networks, because stop-words are important for the discovery of many dependencies. Dependency networks are sparser than co-occurrence networks, and stop-words typically act as hubs in these network. Consequently, their removal would severely impact the connectedness of the graphs.

We have previously studied certain graph properties of the two representations and found that the networks generally had high clustering coefficients and low path lengths, which are the characteristics of Small-World networks [28]. We also expected the network degree distributions to be scale-free, another known property of this type of networks, but were surprisingly unable to detect this. This is not the focus in this paper, however, and we refer to [25] for further details.

We describe next the different properties of co-occurrence and dependency based networks.

3.1.1. Co-occurrence Networks

Co-occurrence networks capture structure by linking together terms that occur together in the same textual context. There are two types of contexts commonly used with such networks: sentences and sliding n-word windows. When n-word windows were used, smaller contexts generally performed better than the larger ones, with 2 or 3 being good choices for n. We evaluated both sentence and window contexts, and found that sentence contexts performed better than window contexts for classification on our two data sets (see [25] for details). We therefore selected sentence contexts for our representation.

It is puzzling that the best results are obtained by either small contexts with windows of 2–3 words, or the much larger sentence contexts. We cannot explain this fully, but believe that the larger context windows may be penalized for including many cross-sentence co-occurrences, which may hold little or no relevant information.

3.1.2. Dependency Networks

Dependency networks also encode connections between terms co-occurring within sentences. However, a dependency grammar is used to extract only pairs of terms that fulfill predefined syntactic dependency relations in the sentence. Hence, edges in the dependency networks are a subset of the sentence-based co-occurrence representation.

Fig. 2 shows an example network constructed from the single sentence *"Immediately after the second touchdown, the pilot decided to perform a go-around"*. The edge labels indicate the dependency types that define the relationships between terms. A corresponding co-occurrence network using sentences as contexts would be fully connected.

We used the Stanford Lexical Parser[1] to identify the dependencies. The parser is based on a grammar called the Stanford Dependencies [6], which defines 52 different dependency relations.

Our representation makes use of only 49 of the defined dependencies, since we, through empirical evaluations [25], found that many of the dependencies did not contribute much to the performance. Particularly three dependency types (*agent*, *advcl*, and *parataxis*) seemed to decrease classification accuracy, and have therefore been excluded them from the experiments presented in this paper.

3.2. Local and Global Term Weights

We have tested two approaches to term weighting based on centrality — one local and one global. The local measure, *Term Centrality* (TC), is based solely on information drawn from the document itself and simply assigns the node centrality of a term as its weight. The global measure, which we call *Term Centrality-Inverse Corpus Centrality* (TC-ICC), weights the terms by their relative centrality in the document network, compared to their overall centrality in the entire corpus. That is, while TC is measured on a network representing a single document, TC-ICC takes into consideration also the term's centrality in the network representing the entire document collection. We define TC-ICC as

$$\text{TC-ICC}_{t,d} = \frac{\text{TC}_{t,d}}{\text{CC}_t + 1} \quad (1)$$

where $\text{TC}_{t,d}$ is the centrality of term t in document d, and CC_t is the centrality of t in the network constructed from the whole corpus. The $+1$ in the denominator serves to keep the resulting value in the range $[0, 1]$, and avoid division with zero when CC_t is 0.

Eq. (1) is our initial attempt at capturing local and global centrality information into a single measure. We have not tested many variations

[1]The parser is freely available from http://nlp.stanford.edu/software/lex-parser.shtml

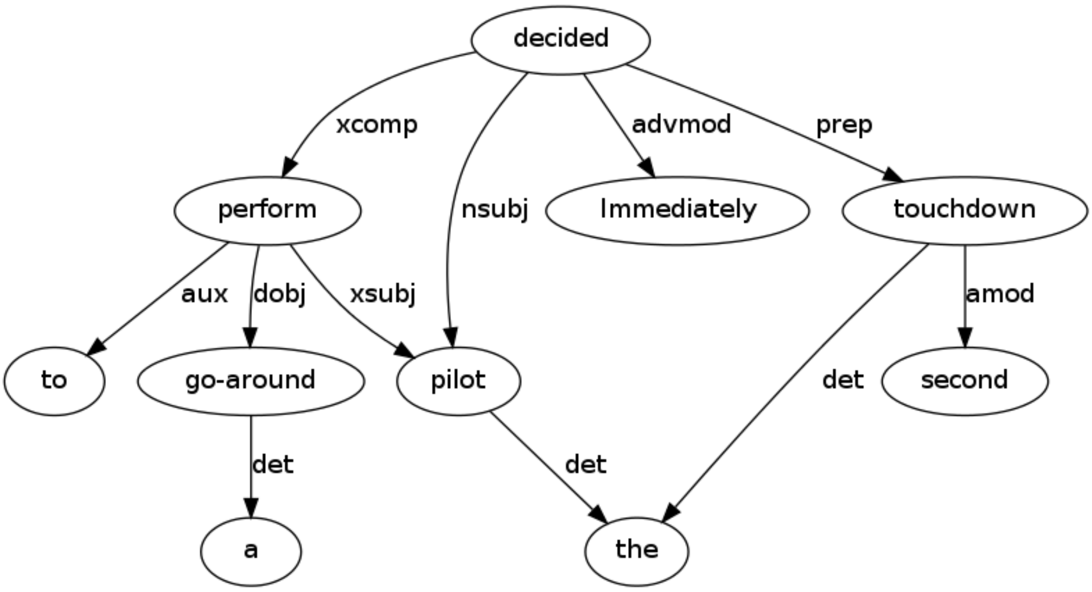

Fig. 2. Dependency network structure as constructed by the Stanford Lexical Parser from the sentence "*Immediately after the second touchdown, the pilot decided to perform a go-around*".

of this yet, and there is thus a good chance that further experiments might reveal better ways to combine TC and CC.

TC and TC-ICC are comparable to TF and TF-IDF, respectively, in that the former considers only information from the document itself, while the latter takes into account information about the rest of the corpus as well.

3.3. Centrality for Term Vector Representation

There exists a wide range of different node centrality algorithms. We have evaluated the most common ones, in order to find the one best suited to our representation.

Three main groups of centrality measures have long been prominent [9]. They consist of measures based on *degree*, *closeness*, and *betweenness* of terms, respectively. An additional fourth group, based on *eigenvectors*, has emerged more recently. We have evaluated several measures from each of these groups.

The degree-based measures are simplest. These define node centrality in terms of the number and strengths of connections between a node and its neighbors. We have tested weighted and unweighted versions of *degree* centrality.

Both closeness and betweenness centrality measures are based on paths through the network. Closeness is defined in terms of the lengths of the shortest paths from a node to the rest of the nodes in the network. Our experiments included standard *closeness centrality* and a variation called *current-flow closeness* [4]. Betweenness centrality describes whether a node is part of the shortest paths between other nodes in the network. Here we evaluated standard *betweenness centrality*, *current-flow betweenness* [18], [4], and *load centrality* [3].

The last group operates on eigenvectors of the adjacency matrix of the graph. These measures capture not only the number of neighbors a node has, but also take into account the importance of each neighbor. From this group we have evaluated the standard *eigenvector centrality* [2], Google's *PageRank* [19], [5], and the *HITS measure* [12].

The evaluation was performed with both TC and TC-ICC on the TASA900 data set, which is described in Section 4.1.1. The results are summarized in Fig. 3. As seen in the figure, despite some differences between co-occurrence and

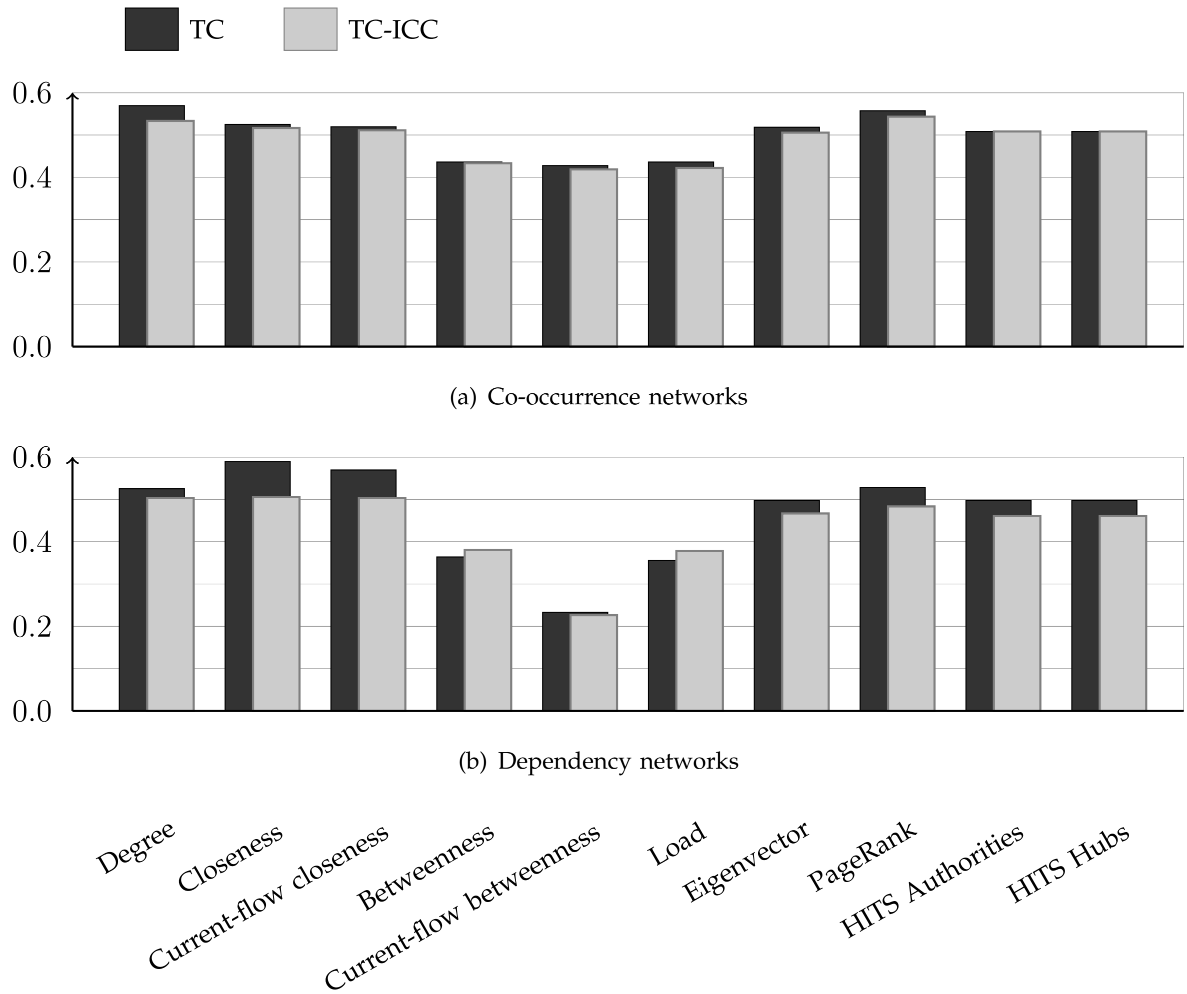

Fig. 3. Classification accuracy with the various centrality measures for TASA900.

dependency networks, some patterns clearly emerge for both representations. The performance of betweenness centralities is poor on both network types. This seems to indicate that being part of the shortest paths between other nodes is not a good indicator of term importance, making betweenness-based centralities unsuited for term weighting.

3.3.1. Centrality in Co-occurrence Networks

The measures that stood out as the best performers for co-occurrence networks were the degree centrality and PageRank. Degree centrality performed slightly better than PageRank, and was therefore selected for use with co-occurrence networks.

The degree centrality value is calculated as

$$C_d(i) = \frac{1}{N-1} \sum_{j}^{N} A_{ij} \qquad (2)$$

where A is the adjacency matrix, and $A_{ij} = 1$ if an edge exists between i and j, and 0 otherwise. N is the number of nodes in the graph.

3.3.2. Centrality in Dependency Networks

Fig. 3(b) shows evaluation results for dependency networks. Similarly to the co-occurrence networks, PageRank and degree centrality performed good here as well. However, the measures that performed best on this network type were the closeness centralities. Since standard closeness had the highest classification accuracy and also is by far the simplest of the closeness measures, we select it for use with dependency networks.

Closeness centrality is defined as

$$C_c(i) = \frac{N-1}{\sum_{j=1}^{N} d(i,j)} \tag{3}$$

where $d(i,j)$ denotes the length of the geodesic path between nodes i and j, i.e. the shortest path in terms of number of edges.

A problem with this formulation is that it is only meaningful for connected graphs. If the graph is unconnected, every node has at least one other node which cannot be reached from it, and thus

$$\sum_{j=1}^{N} d(i,j) = \infty$$

for all i. We handled this by treating each connected subgraph separately, using the following formula.

$$C_c(i) = \frac{M-1}{N-1} \frac{1}{\sum_{j=1}^{N} d(i,j)} \tag{4}$$

M is here the number of nodes in the connected subgraph where node i resides. Normalization is adapted to reflect the node's closeness in the overall graph, based on the size of its subgraph.

4. Experiments

We tested the representations empirically on two data sets for the document classification task, as described below.

In the experiments, a classifier is trained on parts of the data set, and used subsequently to classify the documents in the remainder of the data. The classification accuracy, i.e. the portion of the test documents correctly classified, is the outcome of the experiment.

We use a k-Nearest Neighbours (k-NN) classifier for the classification, with $k = 5$. Each test document is labeled with the label most common among the 5 most similar training documents. Document similarity is measured using the cosine similarity measure.

4.1. Data sets

We perform classification experiments on the following two data sets.

4.1.1. TASA900

The TASA data set is a corpus containing text sampled from curriculum used in American high schools. The collection consist of 37 600 documents arranged into nine categories, *Business*, *HomeEconomics*, *LanguageArts*, *Science*, *Unspecified*, *Health*, *IndustrialArts*, *Miscellaneous*, and *SocialStudies*, totalling approximately 10 million tokens of text.

TASA is a diverse corpus with documents from a wide range of topics, both within and across categories, which makes it a challenging data set to classify.

We have created a subset of the corpus by selecting the first 100 documents from each category. This forms our *TASA900* data set. The data set is split into two parts, of which 60% is used for training and the remaining 40% for testing. The split is made randomly, but so that that the 60/40 split is maintained also within each category.

4.1.2. Reuters

Reuters-21578 is a collection of news articles that appeared on the Reuters newswire in 1987. It has been manually labeled by Reuters personnel, and is widely used for text classification tasks. The set consists of 21578 documents, some of which are unlabeled and some labeled with one or more of the 672 different categories.

While the data set originally is available[2] in SGML format, our copy is downloaded from a version maintained[3] by Alessandro Moschitti at the University of Trento, who has done a great job of structuring and processing the data into a more user friendly format.

The documents in the *Reuters* data set are generally shorter than those in TASA, but tend to have longer sentences. The documents are

[2]http://www.daviddlewis.com/resources/testcollections/reuters21578/

[3]http://disi.unitn.it/moschitti/corpora.htm

also less topically diverse, both within and across categories.

The documents are separated into training and testing sets according to the ModApté split, which is described in the README-file accompanying the distribution of the data set. Our collection consist of documents from 90 different categories — all categories including at least 1 training document and 1 test document. There are a total of 9598 training documents and 3744 testing documents.

4.2. Results

Table I lists the classification accuracies on the two data sets for both the co-occurrence and the dependency based representations. We observe that the best results are achieved using graph-based TC weighting, and that local representations (TC and TF) generally outperform their corresponding global counterparts (TC-ICC and TF-IDF, respectively).

This tendency was also evident for most other centrality measures in the evaluation from Fig. 3, where TC with only few exceptions performed better than TC-ICC. The differences between the two were more pronounced with dependency networks than with co-occurrence networks.

TABLE I
CLASSIFICATION RESULTS

	Representation	Reuters	TASA900
local	TF	0.6693	0.5678
	Co-occurrence (TC)	**0.6880**	0.5694
	Dependency (TC)	0.6827	**0.5889**
global	TF-IDF	0.6375	0.5655
	Co-occurrence (TC-ICC)	0.6875	0.5333
	Dependency (TC-ICC)	0.6763	0.5056

5. Discussion

From our experiment results, summarized in Table I, we see that the graph-based TC representations show a clear improvement over both TF and TF-IDF. The TC-ICC representations perform well on Reuters, but is outperformed on the TASA900 data set. It is not clear from the results whether networks based on co-occurrences or dependencies are best, as each perform better on one of the data sets. Both versions of TC are, however, better than both TF and TF-IDF on both data sets.

We cannot say for sure why the graph-based representations perform better. An intuitive explanation is that the performance is increased because the representations are able to retain information about term order. The representation also captures a number of different terms with which each term interacts, and the importance of the neighbors.

For co-occurrence networks, we found the degree centrality to perform better than the other measures, closely followed by PageRank which is more commonly used in this type of systems (e.g. [15], [7], [13]). The good performance of degree centrality is surprising, since it is the simplest of the measures. It is also the centrality measure that best mimics the frequency based term weighting. The difference between degree centrality and term frequency is that while TF reflects how many times a term is used, degree centrality score higher those terms that co-occur with a large set of other distinct terms.

Closeness centrality, which proved to be the best centrality measure for dependency networks, is a more interesting measure. Closeness considers more of the network structure, and tends to favour terms that are well connected with all the other terms, although not necessarily directly.

Another interesting aspect of the results is that the local measures TC and TF perform better than the global TC-ICC and TF-IDF, respectively. The differences between the local and global versions of each measure vary, but the local version performs better in all cases.

TF-IDF, originally used in IR, has also been adopted in text classification research. IDF weights rare terms higher, which is appropriate for IR. It is not clear that this is the best approach in the text classification task, however. Several researchers have suggested that the use of IDF is indeed inappropriate for text classification, advocating the use of supervised term weighting methods instead [21], [1].

Our results show that the preference for rare terms in TF-IDF is indeed harmful to classification performance, and that the same trend is present in the TC-ICC measures. Analogous to the way IDF favors rare terms, ICC weights the terms that have overall low centrality in the document collection higher. We therefore speculate that for unsupervised weighting methods it may be better to use the information local to each document, rather than the global information about the term in the entire corpus.

That the local information is sufficient is a good news when it comes to computational costs. With TC-ICC, centrality values must be computed over networks spanning the entire corpus. For some centrality measures, especially the eigenvector-based and the betweenness centralities, this will be slow or even unmanageable for large corpora.

Graph-based methods are of course more complex than TF and TF-IDF, also without the use of corpus centrality. The computation of term centrality is, however, fortunately only an initial cost required once for each document.

6. Conclusions

We have studied term weighting methods using co-occurrence and dependency-based graph representations. Effectiveness of various centrality measures in representing term weights have been investigated. We have analyzed both document level and corpus level information about terms. The text classification task has been used as the application task in the experiments. In our experiments conducted on two datasets centrality measures performed better than the widely used TF and TF-IDF representations. In general, the local measures outperformed the global ones.

As yet the graph-based representations are fairly simple. In dependency networks, terms of all types, except stop-words, are included equally as nodes, and the dependency types are all treated the same way. A more detailed study of which dependency types and which combination of dependency types performs best is needed. Another possible improvement is to use various types of textual networks together. By combining the term-networks presented here with the type of sentence networks used in TextRank, terms could be treated differently based on the importance of the sentences where they were used. It would also be interesting to investigate the usefulness of semantic relations for document representation. By utilizing information from resources such as WordNet, new types of edges could be introduced in the networks.

Acknowledgements

We are grateful to Touchstone Applied Science Associates Inc for providing us with the TASA dataset. We thank Olav Bjørkøy for providing feedback on an early draft of the paper. Thanks also to the anonymous reviewers for their useful suggestions for improvements to the manuscrip.

References

[1] I. Batal and M. Hauskrecht. Boosting KNN Text Classification Accuracy by using Supervised Term Weighting Schemes. In *Proceedings of the 18th ACM conference on Information and knowledge management - CIKM '09*, pages 2041–2044. ACM Press, 2009.

[2] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.

[3] U. Brandes. On Variants of Shortest-Path Betweenness Centrality and their Generic Computation. *Social Networks*, 30(2):136–145, May 2008.

[4] U. Brandes and D. Fleischer. Centrality Measures Based on Current Flow. *STACS 2005*, pages 533–544, 2005.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, Apr. 1998.

[6] M. De Marneffe and C. Manning. The Stanford typed dependencies representation. *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation - CrossParser '08*, (ii):1–8, 2008.

[7] G. Erkan and D. R. Radev. Lexrank: Graph-based Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research*, 22, 2004.

[8] K. Faust and S. Wasserman. Centrality and prestige: A review and synthesis. *Journal of Quantitative Anthropology*, 4(1):23–78, 1992.

[9] L. C. Freeman. Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3):215–239, 1978.

[10] J. Gao, J.-Y. Nie, G. Wu, and G. Cao. Dependence Language Model for Information Retrieval. *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*, pages 170–177, 2004.

[11] P. J.-J. Herings, G. V. D. Laan, and D. Talman. The positional power of nodes in digraphs. *Social Choice and Welfare*, 24(3):439–454, June 2005.

[12] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, Sept. 1999.

[13] J. Liu, J. Wang, and C. Wang. A Text Network Representation Model. *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, pages 150–154, Oct. 2008.

[14] R. Mihalcea. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, number 4, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

[15] R. Mihalcea and P. Tarau. TextRank: Bringing order into texts. In *Proceedings of EMNLP*, pages 404–411. Barcelona: ACL, 2004.

[16] R. Mihalcea, P. Tarau, and E. Figa. PageRank on semantic networks, with application to word sense disambiguation. In *Proceedings of the 20th international conference on Computational Linguistics - COLING '04*, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

[17] R. Navigli and M. Lapata. Graph connectivity measures for unsupervised word sense disambiguation. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1683–1688, 2007.

[18] M. Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.

[19] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *World Wide Web Internet And Web Information Systems*, pages 1–17, 1998.

[20] M. Sahlgren. An introduction to random indexing. *and Applications of Semantic Indexing Workshop at the*, pages 1–9, 2005.

[21] P. Soucy and G. Mineau. Beyond TFIDF Weighting for Text Categorization in the Vector Space Model. In *International Joint Conference on Artificial Intelligence*, volume 19, pages 1130–1135, 2005.

[22] J. Tomita and G. Kikui. Interactive Web search by graphical query refinement. *Poster Proceedings of the 10th international World Wide Web conference (WWW10)*, Oct. 2001.

[23] J. Tomita, H. Nakawatase, and M. Ishii. Graph-based text database for knowledge discovery. *Poster Proceedings of the 13th International World Wide Web Conference (WWW2004)*, pages 454–455, 2004.

[24] G. Tsatsaronis, I. Varlamis, and K. Nørvåg. An Experimental Study on Unsupervised Graph-based Word Sense Disambiguation. *Computational Linguistics and Intelligent Text Processing*, pages 184–198, 2010.

[25] K. Valle. Graph-based Representations for Textual CBR. Master Thesis, June 2011.

[26] C. van Rijsbergen. A Theoretical Basis for the Use of Co-occurrence Data in Information Retrieval. *Journal of Documentation*, 33(2):106–119, 1977.

[27] W. Wang, D. Do, and X. Lin. Term Graph Model for Text Classification. *Lecture notes in computer science*, 3584:19, 2005.

[28] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, June 1998.