Acquisition of Domain-Specific Senses and its Extrinsic Evaluation Through Text Categorization

Attaporn Wangpoonsarp, Kazuya Shimura, and Fumiyo Fukumoto

Graduate Faculty of Interdisciplinary Research, University of Yamanashi 4-3-11, Takeda, Kofu, Yamanashi, 400-8510 Japan
{g16dhl01,g17tk008,fukumoto}@yamanashi.ac.jp

Abstract. This paper focuses on domain-specific senses and proposes a method for detecting predominant sense depending on each domain. We applied a simple Markov Random Walk (MRW) model to rank senses for each domain. It decides the importance of a vertex (senses) within a graph by using the similarity of senses. The similarity of senses is obtained by using distributed representations of words from gloss texts in the thesaurus. It captures large semantic context and thus does not require manual annotation of sense-tagged data. In order to evaluate the method, we applied the results of domain-specific senses to text categorization. The performance achieved in our test set WordNet3.1 and the Reuters corpus demonstrates applicability for the text categorization task.

Keywords: Domain-specific senses · Word mover distance · Text categorization.

1 Introduction

Detection of domain-specific senses is crucial information for many NLP tasks such as word sense disambiguation (WSD), machine translation, and QA, and thus has attracted the attention of NLP researchers since the earliest days of corpus-based NLP. The simplest way is so-called the first sense heuristic (FS) that is a method to choose the first or predominant sense of a word and it is often used as a baseline for supervised WSD systems [30, 12]. Because it is powerful, especially for words with highly skewed sense distributions [36, 12]. However, the drawback in the FS applied to WordNet is a small amount of SemCor corpus which causes data sparseness problem, *i.e.*, we cannot apply the FS to the senses that do not appear in SemCor. Furthermore, the FS is not based on the domain but instead on the simple frequency counts of SemCor data. Consider the noun word, "ball". There are twelve noun senses of "ball" in the WordNet. The first sense of "ball" is "round object that is hit or thrown or kicked in games", and it is often used in the "sports" domain rather than the "military" domain. In contrast, the second sense of "ball", *i.e.*, "a solid projectile that is shot by a musket" is more likely to be used in the "military" domain.

In this paper, we focus on domain-specific senses of nouns and propose a method for detecting predominant sense in each domain/category. Our model employs distributed representations of words learned by using Word2Vec [22] and thus does not require manual annotation of sense-tagged data. We applied a simple Markov Random Walk (MRW) model to rank senses for each domain. The similarity of senses which is used to decide the importance of senses is obtained by using distributed representations of words from gloss texts in the thesaurus. To examine the effectiveness of our detection method, we applied the results to text categorization on the dataset collected from the Reuters corpus. The result shows that our model gains great improvement over the single-channel Convolutional Neural Network (CNN). The main contributions of our work can be summarized: (1) we propose a method for identifying domain-specific noun senses which leverage distributed representations of words and thus does not require manual annotation of sense-tagged data, while some of the existing work required a considerable amount of hand-labeling. (2) From the perspective of robustness, the method is automated and required only documents from the given domains/categories such as the Reuters corpus, and thesaurus with gloss texts such as WordNet. The method is easily applicable to a new domain or sense inventory, given sufficient documents. (3) We empirically evaluate our model and show that the result of domain-specific noun senses is effective for the text categorization task.

2 Acquisition of Domain-Specific Senses

The first sense heuristic is a very powerful heuristic and often used as a baseline of sense disambiguation systems. However, there are at least two major problems to use it as a sense disambiguation heuristic [20]. The first is that the predominant sense of a word varies according to the source of the document belonging to the domain. The second problem with obtaining predominant sense information is that it needs manual annotation of the corpus which causes a relatively small amount of resources such as SemCor. A methodology to find domain-specific senses without requiring manual annotation of data is needed.

The goal of our model is to identify predominant sense distributions of a word depending on the domain. We applied a simple graph centrality algorithm, Markov Random Walk (MRW) to the extracted noun words from the documents assigned to a specific domain/category and identify domain-specific senses for the domain. The idea of MRW model is that of *voting*. When one vertex links to another one, it is basically casting a vote for that other vertex. The ranking is conducted by two metrics. One is a metric that the larger the number of votes that are cast for a vertex, the higher the importance of the vertex. Another is a metric that how important the vote itself is. We applied the algorithm to detect the domain-specific-sense of words.

The input of the MRW model is a graph consisting of vertices, *i.e.*, each possible noun sense appeared in a specific domain and edges with similarity value between vertices. We represent each noun sense as its gloss text extracted from the thesaurus, WordNet. We calculated sense similarity by using Word Mover Distance (WMD) [13]. WMD measures the dissimilarity between two sentences as the minimum amount of distance that the embedded words of one sentence need to *travel* to reach the embedded words of another sentence. The word embedding is learned by using Word2Vec [22] with Continuous Bag-of-Words (CBOW).

Let $\mathbf{X} \in R^{d \times n}$ be a Word2Vec embedding matrix for vocabulary size of n words. The i^{th} column, $\mathbf{x_i} \in R^d$ indicates the embedding of the i^{th} word in d-dimensional space. We represent gloss text of each sense as normalized Bag-Of-Words (nBOW) vector, $g \in R^n$. The objective of the model is to minimize cumulative cost C of moving the gloss text g to g':

$$C = \sum_{i,j=1}^{n} \mathbf{T}_{i,j} \ \ || \ \mathbf{x}_i - \mathbf{x}_j \ ||_2,$$
$$\text{subject to:} \sum_{j=1}^{n} \mathbf{T}_{ij} = g_i, \ \forall i \in \{1, \cdots, n\},$$
$$\sum_{i=1}^{n} \mathbf{T}_{ij} = g'_j. \ \forall j \in \{1, \cdots, n\}. \quad (1)$$

$\sum_{j=1}^{n} \mathbf{T}_{ij} = g_i$ in Eq. (1) shows that outgoing flow from word i equals g_i. Similarly, $\sum_{i=1}^{n} \mathbf{T}_{ij} = g'_j$ indicates that incoming flow to word j matches g'_j. Each score of the sense in a specific domain is obtained by the principal eigenvector of the matrix. We applied the algorithm for each domain. We note that the matrix M is a high-dimensional space. Therefore, we used a ScaLAPACK, a library of high-performance linear algebra routines for distributed memory MIMD parallel computing [24], which includes routines for solving systems of linear equations, least squares, eigenvalue problems.

We selected the topmost $K\%$ words (senses) according to rank score for each domain and make a sense-domain list. For each word w in a document, find the sense s that has the highest score within the list. If a domain with the highest score of the sense s and a domain in a document appeared in the word w match, s is regarded as a domain-specific sense of the word w.

3 Application to Text Categorization

Our hypothesis about text categorization is that document assigned to a specific category includes predominant word sense related to the category. We combined the knowledge of domain-specific senses with the embedding of documents. For the words and senses in the documents, we propose a model which has two components shown in Figure 1: one channel is for the word embedding of the document, and another channel is for sense embedding, *i.e.*, each word which is disambiguated its sense in the document is replaced to its gloss text. Both of them are the input of Convolutional Neural Network (CNN). With this model, we can learn rich features from both the word level and the sense level, respectively.

Similar to other CNN [8, 14], our model which is shown in Figure 1 is based on [11]. Let $\mathbf{x}_i \in \mathbb{R}^k$ be the k-dimensional word vector with the i-th word in the input of CNN obtained by applying skip-gram model provided in Word2Vec. The input with length n is represented as $\mathbf{x}_{1:n} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{nk}$. A convolution filter $\mathbf{w} \in \mathbb{R}^{hk}$ is applied to a window size of h words to produce a new feature, $c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b)$ where $b \in \mathbb{R}$ indicates a bias term and f refers to a non-linear activation function. We applied this convolution filter to each possible window size in the input and obtained a feature map, $m \in \mathbb{R}^{n-h+1}$. As shown in Figure 1, we then apply a max pooling operation over the feature map and obtain the maximum value $\hat{m}$ as a feature of this filter. We obtained

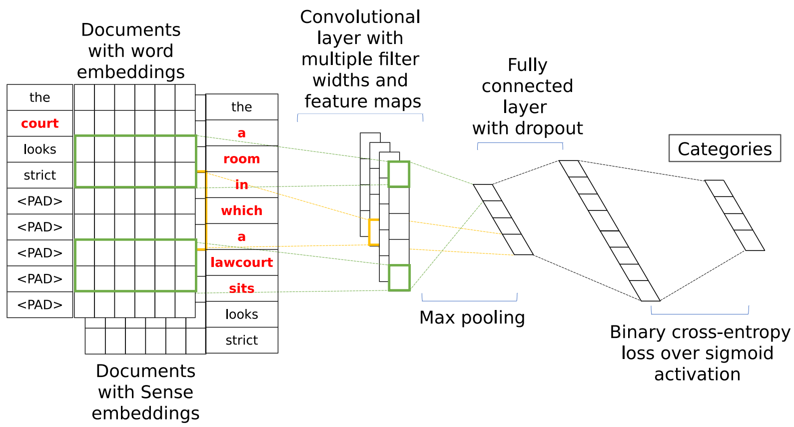

Fig. 1. Two channels CNN

multiple filters by varying window sizes and multiple features. These features form a pooling layer and are passed to a fully connected layer. In the fully connected layer, we applied dropout [7]. The dropout randomly sets values in the layer to 0. Finally, we obtained the probability distribution over categories. The network is trained with the objective that minimizes the binary cross-entropy (BCE) of the predicted distributions and the actual distributions by performing stochastic gradient descent.

4 Experiments

We selected Reuters corpus and WordNet 3.1 thesaurus to evaluate our method.

4.1 Acquisition of Senses

The Reuters corpus consists of 806,791 documents organized into 126 categories. There are no existing sense-tagged data for domains that we can utilize for evaluation. Therefore, we used the Subject Field Codes (SFC) resource which semi-automatically annotates WordNet 2.0 synsets with domain labels [16]. The SFC consists of 115,424 words assigning 168 domain labels which include some of the Reuter's categories. We manually assigned Reuter's categories to SFC labels which are shown in Table 1. "The # of doc" in Table 1 refers to the number of documents in each category.

We applied POS tagger, lemmatizer and Named Entity Recognition of the Stanford CoreNLP [18] to Reuters corpus. We removed stopwords and words whose frequency is less than five. We extracted noun words and named entities, "person", "organization", and "location". We set the number of dimensions to 100, and the window size to 5 in the Word2Vec. For each domain, we collected the topmost 50% noun words. The total number of words is 1,181 and senses is 1,313. We randomly divided these nouns into two: training and test data. The training data is used to estimate $K\%$ words (senses) according to rank score, and test data is used to test the method using the estimated

Table 1. Correspondences between the Reuters and SFC categories

Reuters	SFC	The # of doc
Sports	Sports	35,225
War	Military	32,580
Legal/Judicial	Law	32,194
Economics	Economy	117,501
Politics	Politics	56,834

value K. We manually evaluated a sense-domain list. As a result, we set K to 20%. We built an individual model for each category. The results are shown in Table 2.

Table 2. The results of sense assignments

Category	Sense	DSS	SFC	Correct	F-score	IRS	P_IRS
Sports	297	59	63	50	0.820	4.40	4.66
War	548	110	115	85	0.756	4.95	5.28
Legal/Judicial	740	148	153	101	0.671	4.06	5.58
Economics	577	115	120	77	0.655	4.63	5.33
Politics	815	163	174	107	0.635	4.92	5.67
Average	509	102	109	72	**0.737**	**4.37**	**4.98**

Table 2 shows the results obtained by the topmost 20% senses according to rank score. “Sense” indicates the total number of senses which should be assigned to each category. “DSS” and “SFC” refer to the number of senses obtained by our method and appeared in the SFC resource, respectively. “Correct” shows the number of senses appearing in both of our method and SFC. “F-score” indicates F-measure. “IRS” refers to Inverse Rank Score and the higher the IRS value, the better the system performance. “P_IRS” denotes indicates the perfect correct value of IRS.

We can see from Table 2 that the overall performance depends on the categories. The best performance was “Sports”. In contrast, the results of “Politics”, “Economics” and “Legal” were $0.635 \sim 0.671$. One reason is that these three domains are semantically similar to each other compared to “Sports” domain. Our method depends on the size of gloss text in the WordNet. Efficacy can be improved if we can assign gloss texts from another thesaurus e.g., Roget’s by using corpus statistics. This is a rich space for further exploration.

We note that some senses of words that were obtained correctly by our method did not appear in the SFC resource because of the difference in WordNet version, *i.e.*, we used WordNet 3.1, while SFC was based on WordNet 2.0.

Table 3 illustrates some examples obtained by our method but that did not appear in the SFC. Table 3 gives an example for each domain. For example, military sense of the word “Redoubt” and the act of meting out the justice of “Administration” which were correctly obtained by our method but did not occur in the SFC resource. This clearly supports the usefulness of our automated method.

Table 3. Sense example identified by our method

Category	Word	Sense
Sports	Jerk	Raising a weight from shoulder height to above the head by straightening the arms.
War	Redoubt	(Military) A temporary or supplementary fortification; typically square or polygonal without flanking defenses.
Legal/Judical	Administration	The act of meting out justice according to the law.
Economics	Spending	Money paid out; an amount spent.
Politics	Labour party	A political party formed in Great Britain in 1900; characterized by the promotion of labor's interests and formerly the socialization of key industries.

4.2 Text Categorization

We applied all the results of domain-specific senses to text categorization to examine how the results obtained by our method affect categorization performance. For each category, we divided all the documents into two folds: 80% for training and 20% for test data. We further divided the training data into two folds: 80% for training data and 20% for validation data. Our model setting for CNN is shown in Table 4. Dropout rate1 in Table 4 shows dropout immediately after embedding layer, and Dropout rate2 denotes dropout in a fully connected layer.

The categorization using CNN is as follows. For the target category, we replaced each word in the test document with its gloss. If the category assigned to the test document by CNN model and the target category match, the test document is judged to classify into the target category. The procedure is applied to each test document and the target category. The results are summarized in Table 5.

Table 4. CNN model settings

Description	Values	Description	Values
Input word vectors	Word2Vec	Filter region size	(2,3,4)
Stride size	1	Feature maps (m)	128
Filters	128×3	Activation function	ReLu
Pooling	1-max pooling	Dropout	Randomly selected
Dropout rate1	0.25	Dropout rate2	0.5
Hidden layers	1,024	Batch sizes	100
Learning rate	Predicted by Adam	Epoch	40 with early stopping
Loss function	BCE loss over sigmoid activation	Threshold value for BSF and MSF	0.5

Table 5. Categorization performance (Topmost 20%)

Category	1ch	2ch-DSS	2ch-SFC
Sports	0.926	0.926 (+.000)	0.928 (+.002)
War	0.680	**0.720** (+.040)	**0.740** (+.020)
Legal/Judical	0.673	**0.698** (+.025)	**0.717** (+.019)
Economics	0.903	**0.916** (+.013)	0.922 (+.006)
Politics	0.725	0.731 (+.006)	**0.787** (+.056)
Macro F-score	0.781	**0.798** (+.017)	**0.819** (+.021)

Table 5 shows categories, categorization performance (F-score) with and without domain-specific senses. “1ch” refers to without domain-specific senses and “2ch-DSS” shows the results obtained by our method. Moreover, we obtained the results (2ch-SFC) by using gold-standard SFC codes. Bold font in 2ch-DSS shows that the results obtained by 2ch-DSS are statistically significant compared to those obtained by 1ch. Similarly, bold font ins 2ch-SFC indicates that the results by 2ch-SFC are statistically significant compared to those by 2ch-DSS. We used a t-test, p-value < 0.05.

Overall, the results showed that domain-specific senses improved text categorization performance. The best improvement was “War” (+0.040), and the poorest was “Sports” (+0.000). The observation is similar to “2ch-SFC”, *i.e.*, the improvement compared to our method is 0.020 for “War” and 0.002 for “Sports”. The text categorization used here is very simple, *i.e.*, CNN with two channels. There are lots of text categorization techniques applicable to the small number of training documents, [34, 32] and it will be worthwhile examining these with our model.

5 Related Work

Semantic-oriented applications such as Question Answering and Machine Translation systems need not only fine-grained and large-scale semantic knowledge but also tune the sense of the word heuristic depending on the domain in which the word is used. Magnini et al. presented a lexical resource where WordNet 2.0 synsets were annotated with Subject Field Codes (SFC) by a procedure that exploits WordNet structure [16, 17]. They annotated 96% of WordNet synsets of the noun hierarchy, while mapping domain labels for word senses were semi-automated and required hand-labeling.

Several authors addressed the problem and have attempted to use specific domain knowledge and show that WSD using specific domain knowledge outperforms generic supervised WSD [2, 6, 31]. Abualhaija et al. proposed D-bee algorithm to find out the best domain for sense in context based on hive and bee agents concept [1]. Lopez-Arevalo el al. proposed their method using the auxiliary corpus that is generated from web information. They test on sports and finance category that obtained the best precision and recall 66.4 and 65.7 respectively and they also test with BNC corpus and the precision and recall is 31.6 and 31.0 [15]. McCarthy et al. proposed an automated method for assigning predominant noun senses[19]. They find words with a similar distribution to the target word from parsed data. The motivation for their work was similar to ours, *i.e.*, to capture changes in the ranking of senses for documents from different domains. They tested 38 words containing two domains of Sports and Finance from the Reuters corpus [29], while we tested five domains with 536 senses in all. Moreover, we applied the results to the text categorization task to evaluate the results quantitatively.

In the context of similarity metric, there have been many attempts which project document dimensional space into lower dimensional space, e.g., Latent Semantic Indexing (LSI) [5] and Latent Dirichlet Allocation (LDA) [4]. Mikolov et al. presented Word2Vec model that was a well known shallow model for training text and generate word embedding [22]. Pagliardini et al. proposed Sent2Vec is a novel sentence embedding algorithm.The principle is built document embeddings by averaging the word embeddings [26]. Kusner et al. presented WMD to compute the similarity between two sentences [13]. It is based on Word2Vec embeddings. It measures the dissimilarity between two sentences as the minimum amount of distance that the embedded words of one sentence reach the embedded words of another sentence.

Some of the earliest attempts to exploit graph-based ranking method for link analysis are in the field of NLP and its application such as unsupervised WSD [23] and document summarization [21]. The basic idea is that of "voting" between nodes. Reddy attempted to use the Personalized PageRank algorithm [3] over a graph representing WordNet to disambiguate ambiguous words [28]. They combine sense distribution scores and keyword ranking scores into the graph to personalize the graph for the given domain. The results showed that exploiting domain-specific information within the graph based methods produce better results than when this information is used individually. However, sense distribution scores are based on the frequency of neighbors of the target word from the thesaurus which is difficult to capture the distance between individual words. Perozzi et al. presented DeepWalk to learn latent representations of vertices in a network [27]. They used local information obtained from truncated random walks to learn latent representations by treating walks as the equivalent of sentences. They applied DeepWalk to several multi-label network classification tasks including Flickr and Youtube and showed that it outperforms baseline methods.

In the context of text categorization, many authors have attempted to apply deep learning techniques including CNN [33], the attention based CNN [35], bag-of-words based CNN [8], and the combination of CNN and recurrent neural network [37] to text categorization. Most of them demonstrated that neural network models are powerful for learning features from texts, while they focused on single-label or a few labels problem. Several efforts have been made to multi-labels [9]. Liu et al. explored a family of new CNN models which are tailored for extreme multi-label classification [14]. They used a dynamic max pooling scheme, a binary cross-entropy loss, and a hidden bottleneck layer to improve the overall performance. The results by using six benchmark datasets where the label-set sizes are up to 670K showed that their method attained at the best or second best in comparison with seven state-of-the-art methods including FastText [10] based CNN. However, all of these attempts aimed at utilizing a large volume of data. Nooralahzadeh et al. proposed Domain-specific Word embeddings using oil and gas corpus and evaluate them with CNN model and obtained the effective results [25]. Wang et al. proposed a method for short text classification which combines explicit and implicit representations [32]. They conceptualize a short text as a set of relevant concepts using a large taxonomy knowledge base called Probase, and then obtain the embedding of a short text by coalescing the words and relevant concepts on top of pre-trained word vectors. Moreover, they incorporated character level features into their CNN model. Wang et al's attempt are similar to our work, while their method used fine-grained and large-scale semantic knowledge that needs to tune the sense of the word heuristic depending on the domain in which the word is used.

6 Conclusion

We proposed a method for detecting domain-specific noun senses which leverages distributed representations of words and thus does not require manual annotation of sense-tagged data. The results attained at 0.737 F-score for 536 senses. Moreover, the results applying text categorization improved categorization accuracy as the Macro F-score without and with domain-specific senses are 0.781 and 0.798, respectively, and the improvement is 0.017. Future work will include: (i) applying the method to other domains for quantitative evaluation, (ii) comparing the method to the state-of-the-art text categorization techniques [32, 14], (iii) incorporating neural network modeling such as DeepWalk into our current method to identify domain-specific senses, and (iv) extending our method to multi-task learning, identification of domain-specific senses and text categorization tasks.

References

1. Abualhaija, S., Tahmasebi, N., Forin, D., Zimmermann, K.: Parameter Transfer across Dmains for Word Sense Disambiguation. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017. pp. 1–8 (2017)
2. Agirre, E., Lacalle, O.L.D., Soroa, A.: Knowledge-based WSD on Specific Domains: Performing better than Generic Supervised WSD. In: Proc. of the 21st International Joint Conference on Artificial Intelligence. pp. 1501–1506 (2009)
3. Agirre, E., Soroa, A.: Personalizing Pagerank for Word Sense Disambiguation. In: Proc. of the 12th Conference of the European Chapter of the Association for Computational Linguistics. pp. 33–41 (2009)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research **3**, 993–1022 (2003)
5. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science **41**(6), 391–407 (1990)
6. Faralli, S., Navigli, R.: A New Minimally-Supervised Framework for Domain Word Sense Disambiguation. In: Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 1411–1422 (2012)
7. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, H., Salakhutdinov, R.R.: Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. In: arXiv preprint arXiv:1207.0580 (2012)
8. Johnson, R., Zhang, T.: Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. In: Proc of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 103–112 (2015)

9. Johnson, R., Zhang, T.: Semi-Supervised Convolutional Neural Networks for Text Categorization vis Region Embedding. Advances in Neural Information Processing Systems **28**, 919–927 (2015)
10. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of Tricks for Efficient Text Classification. In: Proc. of the 15th Conference of the European Chapter of the Association for Conputational Linguistics. pp. 427–431 (2017)
11. Kim, Y.: Convolutional Neural Networks for Sentence Classification. In: Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing. pp. 1746–1751 (2014)
12. Koeling, R., McCarth, D., Carroll, J.: Domain-Specific Sense Distributions and Predominant Sense Acquisition. In: Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing(HLT/EMNLP). pp. 419–426 (2005)
13. Kusner, M.J., Sun, Y., Kolkin, N.L., Weinberger, K.Q.: From Word Embeddings to Document Distances. In: Proc. of the 32nd International Conference on Machine Learning. pp. 957–966 (2015)
14. Liu, J., Chang, W.C., Wu, Y., Yang, Y.: Deep Learning for Extreme Multi-Label Text Classification. In: Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 115–124 (2017)
15. Lopez-Arevalo, I., Sosa-Sosa, V.J., Rojas-Lopez, F., Tello-Leal, E.: Impoving Selection of Synsets from WordNet for Domain-specific Word Sense Disambiguation. Computer Speech and Language **41**(C), 128–145 (2017)
16. Magnini, B., Cavaglia, G.: Integrating Subject Field Codes into WordNet. In: Proc. of the International Conference on Language Resources and Evaluation 2000. pp. 1413–1418 (2000)
17. Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A.: The Role of Domain Information in Word Sense Disambiguation. Natural Language Engineering **8**, 359–373 (2002)
18. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proc. of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60 (2014)
19. McCarthy, D., Koeling, R., Weeds, J., Carroll, J.: Finding Predominant Word Senses in Untagged Text. In: Proc. of the 42nd Annual Meeting on Association for Computational Linguistics. pp. 279–286 (2004)
20. McCarthy, D., Koeling, R., Weeds, J., Carroll, J.: Unsupervised Acquisition of Predominant Word Senses. Computational Linguistics **4**(33), 553–590 (2007)
21. Mihalcea, R.: Language Independent Extractive Summarization. In: Proc. of the ACL Interactive Poster and Demonstration Sessions. pp. 49–52 (2005)
22. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. Computing Research Repository **abs/1301.3781** (2013)
23. Navigli, R., Lapata, M.: An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(4), 678–692 (2010)
24. Netlib: http://www.netlib.org/scalapack/index.html. In: Netlib Repository at UTK and ORNL (2007)
25. Nooralahzadeh, F., vrelid, L., Lnning, J.T.: Evaluation of Domain-specific Word Embeddings using Knowledge Resources. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
26. Pagliardini, M., Gupta, P., Jaggi, M.: Unsupervised Learning of Sentence Embeddings using Compositional n-Gram Features. Computing Research Repository **abs/1703.02507** (2017)
27. Perozzi, B., AI-Rfou, R., Skiena, S.: DeepWalk: Online Learning of Social Representations. In: Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 701–010 (2014)

28. Reddy, S., Inumella, A., McCarthy, D., Stevenson, M.: IIITH: Domain Specific Word Sense Disambiguation. In: Proc. of the 5th International Workshop on Semantic Evaluation. pp. 387–391 (2010)
29. Rose, T., Stevenson, M., Whitehead, M.: The Reuters Corpus Volume 1 - from Yesterday's News to Tomorrow's Language Resources. In: Proc. of Language Resources and Evaluation-2002 (2002)
30. Snyder, B., Palmer, M.: The English All-Words Task. In: Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. pp. 41–43 (2004)
31. Taghipour, K., Ng, H.T.: Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains. In: Proc. of the 2015 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 314–323 (2015)
32. Wang, J., Wang, Z., Zhang, D., Yan, J.: Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. In: Proc. of the 26th International Joint Conference on Artificial Intelligence. pp. 2915–2921 (2017)
33. Wang, Y., Xia, Y., Zhao, L., Bian, J., Qin, T., Liu, G., Liu, T.Y.: Dual Transfer Learning for Neural Machine Translation with Marginal Distribution Regularization. In: Proc. of the 32nd AAAI Conference on Artificial Intelligence). pp. – (2018)
34. Wu, W., Li, H., Wang, H., Zhu, K.Q.: A Probabilistic Taxonomy for Text Understanding. In: Proc. of the 2012 ACM SIGMOD International Conference on Management of Data. pp. 481–492 (2012)
35. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical Attention Networks for Document Classification. In: Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies. pp. 1480–1489 (2016)
36. Yarowsky, D., Florian, R.: Evaluating Sense Disambiguation Performance Across Diverse Parameter Spaces. Natural Language Engineering **8**(4), 293–310 (2002)
37. Zhang, R., Lee, H., Radev, D.: Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents. In: Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies. pp. 1512–1521 (2016)