Incorporating Dependency Trees Improve Identification of Pregnant Women on Social Media Platforms

Yi-Jie Huang[1], Chu Hsien Su[2], Yi-Chun Chang[3], Tseng-Hsin Ting[3], Tzu-Yuan Fu[3], Rou-Min Wang[3], Hong-Jie Dai[3*], Yung-Chun Chang[4*], Jitendra Jonnagaddala[5] and Wen-Lian Hsu[6]

[1]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan.
[2]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan.
[3]Department of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan.
[4]Graduate Institute of Data Science, Taipei Medical University, Taipei, Taiwan.
[5]School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia.
[6]Institute of Information Science, Academia Sinica, Taipei, Taiwan.

Abstract

The increasing popularity of social media lead users to share enormous information on the internet. This information has various application like, it can be used to develop models to understand or predict user behavior on social media platforms. For example, few online retailers have studied the shopping patterns to predict shopper's pregnancy stage. Another interesting application is to use the social media platforms to analyze users' health-related information. In this study, we developed a tree kernel-based model to classify tweets conveying pregnancy related information using this corpus. The developed pregnancy classification model achieved an accuracy of 0.847 and an F-score of 0.565. A new corpus from popular social media platform Twitter was developed for the purpose of this study. In future, we would like to improve this corpus by reducing noise such as retweets.

1 Introduction

The web has become a powerful medium for disseminating information about diverse topics, people can share information anytime and anywhere. Real-time user generated information on the web, epitomized by social media and in particular microblogs, are becoming an important data source to complement existing resources for disease surveillance (Brownstein, Freifeld, & Madoff, 2009), behavioral medicine (Ayers, Althouse, & Dredze, 2014), and public health (Dredze, 2012). Studies have shown that 26% of online adults discuss health information using social media(GE-Healthcare, 2012), with approximately 90% women using online media for health-care information, and 60% using pregnancy related apps for support. These statistics suggest that social media sources may contain key information regarding specific cohorts, such as pregnant women, and their drug usage habits. Twitter—a micro-blogging site which is actively used by over 328 million users[1]—is a very popular social network currently being extensively used for public health monitoring tasks (Chandrashekar, Magge, Sarker, & Gonzalez, 2017; Jonnagaddala, Jue, & Dai). It is also an attractive resource for biosurveillance related shared tasks and competitions because it carries health-related knowledge expressed by various cohorts(Adam, Jonnagaddala, Chughtai, & Macintyre, 2017). However, the noisy nature of data on Twitter demands sophisticated models and techniques for mining the knowledge encapsulated.

The primary aim of this study is to detect whether a tweet convey pregnancy or not. This information further downstream can be used to study the safety of drugs in pregnancy are of paramount importance. Typically, pregnant woman in social media are detected using simple regular expressions and rules such as – matching the phrase

* Corresponding author

[1] https://about.twitter.com/company

Proceedings of the International Workshop on Digital Disease Detection using Social Media 2017 (DDDSM-2017), pages 26–32, Taipei, Taiwan, November 27, 2017. ©2017 AFNLP

"i am twenty weeks pregnant" (Chandrashekar et al., 2017; Wang, Paul, & Dredze, 2014). However, employing rule based detection can lead to many false positives since it doesn't consider context or sentiment embedded in the tweet. Often, the tweets recognised by rule based methods seem to be sarcastic. For example, consider the tweet "*I look like I'm 6 months pregnant*". This tweet is a sarcastic tweet and the user actually is not pregnant. Thus, in order to overcome this issue, we propose to use a tree kernel-based approach to detect pregnant woman more effectively. Tree kernel-based approaches have been applied to many different researches, such as relation extraction (Culotta and Sorensen, 2004), question classification (Zhang and Lee, 2003) and protein interaction detection (Miwa et al., 2010). In recent years, tree kernel-based models were used to analyze Twitter data, but most of those studies were focused on opinion mining and sentiment classification (Agarwal et al., 2011; Alicante et al., 2016). In this study, we investigate the effectiveness of applying the approach on the task of determining whether a tweet is posted by a pregnant woman or not.

2 Related Work

Most of the studies in mining Twitter are focused on drug safety domain, e.g. drug abuse and adverse drug reaction (Dai, Touray, Wang, Jonnagaddala, & Syed-Abdul, 2016; Sarker et al., 2016). However, this information can also be used for health surveillance of pregnant women. The study most related to ours is presented by (Chandrashekar et al., 2017) in which they annotated 1,200 tweets with pregnancy announcements to allow the identification of pregnancy trimesters. Klein et. al, constructed an annotated corpus from Twitter focusing on personal medication intake (Klein, Sarker, Rouhizadeh, O'Connor, & Gonzalez, 2017). In an another related study an integrated corpus composing of 2,000 sentences from Twitter and PubMed called TwiMed was presented (Alvaro, Miyao, & Collier, 2017). The corpus contains the annotations of diseases, symptoms and drugs, and their relations.

Tree kernel-based approaches have been widely applied to text classification tasks. Zhang and Lee (2003) utilized tree kernel to question classification and demonstrated that syntactic structures is useful for questions classification. However, the space of tree fragments is too large to compute their inner products. In order to recursively and efficiently compute the common substructures similarity between two trees, Moschitti (2006) proposed convolution tree kernel and developed a toolkit for public. Wang et al. (2009) adopted convolution tree kernel to find out similar questions from Yahoo Answers dataset. They observed that the convolution tree kernel function can effectively utilize the syntactic structure of a sentence. On the other hand, Agarwal (2011) applied partial tree kernel (PTK) to classify sentiment polarity of twitter data and achieved remarkable performance. The kernel provides the ability to analyze additional semantic information by considering the contribution of shared subsequences containing all children of nodes. PTK compares words by the order of alphabets to determine word similarity for ameliorating the weak point of convolution tree kernel. Later on Croce et al. (2011) proposed smoothing partial tree kernel (SPTK) and improved the calculation method of word similarity by using singular value decomposition (SVD) to transform all words into vectors for determining the cosine similarity between them.

In this paper, we built models based on SPTK and three kinds of tree structures. Besides part-of-speech (POS) tags, the new tree structures incorporate dependency tree and grammatical relations for extracting more useful features. Furthermore, we used word embedding to substitute the vectors generated by SVD. Many studies have demonstrated that word embedding is really useful in many natural language processing tasks. With this in mind we also investigated the effectiveness of combining word embedding with SPTK and different tree structures on the task of identifying pregnancy women on Twitter.

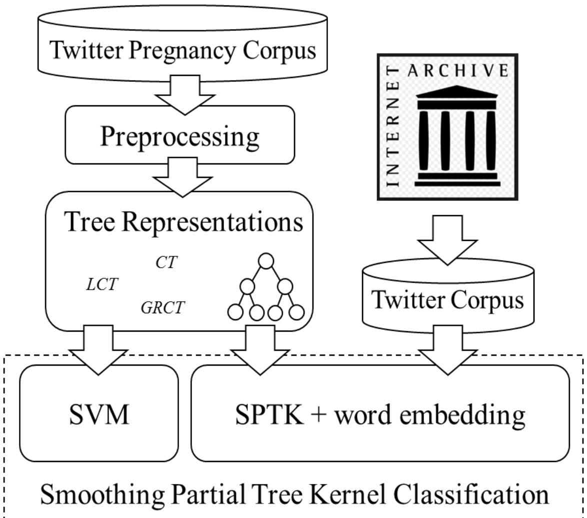

Figure 1: System architecture

3 Methods

The Figure 1 shows the architecture of the proposed pregnancy detection method. The architecture comprises three key components: preprocessing, tree representations, and smoothing partial tree kernel classification. Firstly, the preprocessing component processes a set of tweets that may convey pregnancy (called *candidate tweets* hereafter) through heuristic rules. Then, each candidate tweet is represented by three kinds of tree structures for capturing the information of syntactic, content, and semantic of the tweet. Finally, the smoothing partial tree-kernel classification component measures the similarity between tweet in terms of their tree structures, and the tree kernel is incorporated into support vector machine (SVM) for learning a classifier. We elaborate each component in the following sub-sections.

3.1 Preprocessing

Given a tweet, we first apply the Stanford parser[2] to generate the output of parse tree and PoS tagging. We further remove stop words and URLs from tweets. In addition, we observe that retweet is impossible to convey pregnancy since Twitter's retweet feature is to help users quickly share that tweet with all of users' followers. Therefore, we filter out tweets if the text start with "RT" (i.e. retweet). By filtering out retweets, the rest are the candidate tweets which may convey pregnancy.

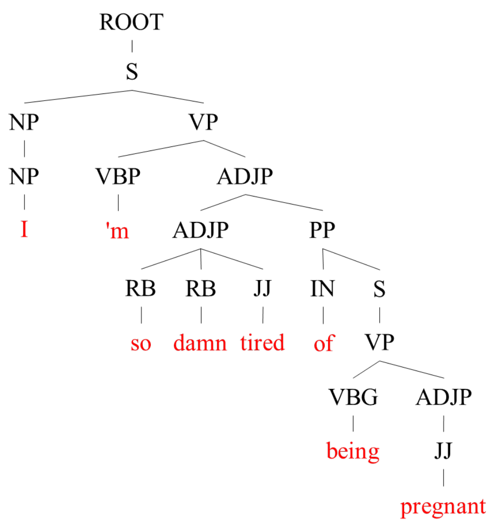

Figure 2: Constituency tree (CT)

3.2 Tree Representations

Different tree representations in tree kernel-based approach may lead to modeling more effective syntactic or semantic feature spaces. In this paper, three kinds tree structure are used to represent tweet, they are constituency tree (CT), lexical centered tree (LCT), and grammatical relation centered tree (GRCT). To facilitate comprehension of the different tree representations, we take a pregnant woman's Tweet "*I'm so damn tired of being pregnant*" as an example.

Figure 2 is CT, which is the basic tree representation generated by Stanford Parser. The parser works out the grammatical structure of sentences by grouping words together as phrases that could represent the subject or object of a verb. However, CT only contain information of the grammatical structure. Croce et al. (2011) proposed GRCT and LCT to complement CT. GRCT and LCT involve grammatical relations (GR), PoS tags and dependencies. GRCT adds tags of grammatical relations and lexical information as new nodes in CT to emphasize grammatical relationship information while LCT enhance the lexical information by adding grammatical relations and PoS-tags as the rightmost children. Figure 3 and Figure 4 show the same example sentence for GRCT and LCT.

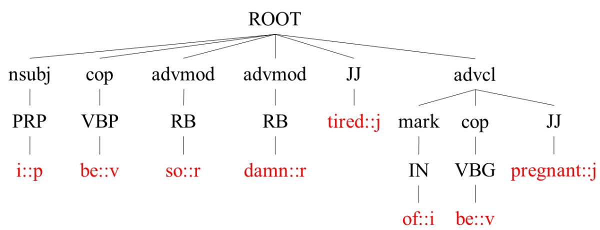

Figure 3: Grammatical relation centered tree (GRCT)

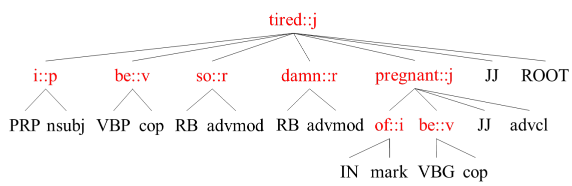

Figure 4: Lexical centered tree (LCT)

3.3 Smoothing Partial Tree Kernel Classification

In SVMs, a kernel function is employed to cleverly compute the similarity between two instances

[2] http://nlp.stanford.edu/software/lex-parser.shtml

without requiring the identification of the entire feature space. In the case of tree kernel, it represents tree in terms of their substructures and evaluates the number of common tree fragments between two trees T_1 and T_2 through the following equation:

$$K(T_1,T_2)=\sum_{n_1\in N_{T_1}}\sum_{n_2\in N_{T_2}}\Delta(n_1,n_2) \qquad (1)$$

where N_{T_1} and N_{T_2} denote the sets of nodes in T_1 and T_2, respectively. The function $\Delta(n_1, n_2)$ is equal to the number of common fragments rooted in the n_1 and n_2 nodes. Since the number of different sub-trees is exponential with the parse tree size, it is computationally infeasible to directly use the feature vector.

In recent years, multiple tree kernels have been proposed for resolving this computation issue, such as syntactic tree kernel (Collins and Duffy, 2002), partial tree kernel (Moschitti, 2006), and lexical semantic kernel (Basili et al., 2005). However, the lexical in these tree kernels must belong to the leaf nodes of exactly the same structures limits its applications. Trivially, it cannot work on dependency trees. Croce et al. (2011) proposed a much more general smoothed tree kernel (i.e. smoothing partial tree kernel, SPTK) that can be applied to any tree and exploit any combination of lexical similarities, respecting the syntax enforced by the tree. Therefore, we adopt SPTK to capture the syntactic similarity between the above high dimensional vectors implicitly, as well as partial lexical similarity of trees. The $\Delta_{SPTK}(n_1, n_2)$ can be defined as follows:

(1) If nodes n_1 and n_2 are leaves, then $\Delta_{SPTK}(n_1, n_2) = \mu\lambda\sigma(n_1, n_2)$
(2) Otherwise, calculate $\Delta_{SPTK}(n_1, n_2)$ recursively as:

$$\Delta_\sigma(n_1,n_2)= \mu\sigma(n_1,n_2)\times(\lambda^2+\sum_{\vec{I}_1,\vec{I}_2,l(\vec{I}_1)=l(\vec{I}_2)}\lambda^{d(\vec{I}_1)+d(\vec{I}_2)}\times \prod_{j=1}^{l(\vec{I}_1)}\Delta_\sigma(c_{n1}(\vec{I}_{1j}),c_{n2}(\vec{I}_{2j}))), \qquad (2)$$

where σ is any similarity between nodes, μ, $\lambda\in[0,1]$ are two decay factors, $\vec{I}_1$ and $\vec{I}_2$ are two sequence of indices, which index subsequences of children u, $\vec{I}=(i_1,\ldots,i_{|u|})$, in sequences of children s, $1\le i_1<\ldots<i_{|u|}\le|s|$, i.e., such that $u=s_{i1}\ldots s_{i|u|}$, and $d(\vec{I})=i_{|u|}-i_1+1$ is the distance between the first and last child. c is one of the children of the node n , also in indexed by $\vec{I}$. This provides an advantage that tree fragments can be matched by applying word embedding similarity σ. Even those tree fragments are not identical but are semantically related.

3.4 Dataset

To the best of our knowledge, there is no openly available corpus for pregnant woman detection. Therefore, we compiled a dataset for the performance evaluation. We employed Tweetinvi[3] to collect tweets mentioning pregnancy written in English from May 1, 2017 to May 29, 2017. To retrieve the tweets, we used a list of pregnancy-related query terms to search tweets online. For all 14,824 collected raw tweets, we pre-processed them by removing emoticons, line feeds, extra spaces and dots based on regular expression. The collected tweets contain duplications owing to the same tweets retrieved by different queries and the retweets shared by different users. We removed duplicated tweets or tweets contain similar descriptions by calculating Levenshtein distance among the collected tweets. If the similarity score of two tweets is larger than 70%, we discard the shorter one and reserve the longer one. Finally, we obtained 7,984 tweet sentences.

We randomly selected 3,000 tweets from the collected dataset to build an initial corpus. Five annotators were recruited to annotate the corpus by using MAE (Multi-document Annotation Environment) (Rim, 2016). They determined whether the tweet authors are pregnant or not based on the context information and gave "Yes" or "No" annotations indicating positive or negative cases. A preliminary consistency test was conducted on 500 tweets by having the first two of the annotators annotate the data, while the last one checked their annotations for consistency. The Fleiss' kappa coefficient value for the initial consistency test is 0.42 (moderate agreements). After examining the consistency, all annotators adjusted their annotations and re-annotated the entire data set. After finishing the annotation process, a voting method was employed to determine the

[3] https://github.com/linvi/tweetinvi

final annotation for each tweet resulting in a corpus containing 642 positive and 2,358 negative annotations.

3.5 Experimental Setting

We use the KeLP package (Simone Filice, 2015) to implement SPTK classification component, and develop three kinds tree representations. To derive credible evaluation results, we utilize the 10-folds cross validation method (Manning and Schütze, 1999). The evaluation metrics used to determine relative effectiveness of the compared methods include the precision, recall, F_1-score, and accuracy (Manning and Schütze, 1999).

For computing lexical similarity, we collected approximating 15.8M tweets from Internet Archive[4] instead of using existing pre-trained word embeddings. In the collected dataset, we removed non-English tweets and transformed each word to its lemma for learning word embeddings (300-dimension) through continuous bag of words with default settings of the word2vec[5] toolkit (Mikolov et al., 2013).

4. Results and Discussion

The performance comparison of our methods with other methods is provided in the Table 1. In order to show complexity of pregnancy detection, the convolution tree kernel with constituency tree (TK_{CT}) was used as the baseline method. We also compared the performance with other two tree structures that included grammatical functions and dependency (TK_{GRCT} and TK_{LCT}) respectively. Moreover, smoothing partial tree kernel with both dependency tree structures were also compared ($SPTK_{GRCT}$ and $SPTK_{LCT}$).

Method	***P.***	***R.***	***F₁***	***Acc.***
TK_{CT}	**74.61**	37.60	50.00	83.90
TK_{GRCT}	63.90	40.81	49.81	82.40
TK_{LCT}	70.82	44.24	54.46	84.17
$SPTK_{GRCT}$	65.50	40.81	50.29	82.73
$SPTK_{LCT}$	71.88	**46.57**	**56.52**	**84.67**

Table 1: The pregnant woman recognition results of compared methods.

As shown in Table 1, the TK_{CT} with the basic tree kernel method can only achieve a mediocre performance. On the contrary, since GRCT and LCT encode the dependency information and grammatical relation in the tree structures, using both dependency tree can benefit the performance of pregnancy detection. It is worth mentioning that TK_{LCT} outperforms TK_{GRCT}, this indicates utilizing lexical features as central node is effective in representing pregnancy information in a tweet. Finally, the SPTK considers lexical similarities on the tree structure to extract features of pregnant woman posts. Therefore, $SPTK_{LCT}$ achieves the best detection performance with an F_1-score of 56.52%.

Figure 5 illustrates a word cloud consisting of the top 100 words frequently occurring in the positive and negative tweets of the corpus. The pink words refer to the positive ones and the blue refer to the negative ones. We excluded mention tags (e.g. @John), numeric numbers, punctuations, and words listed in the stop word list[6]. We observed that among the top 100 words, 55% of them appeared in both positive and negative cases. We also checked the words that are only included in the positive and negative word frequency, and we did not find anything special except the positive words "father" and "husband" and negative word "abortion". It seems that the use of words between positive and negative corpus are highly consistent. The reasons caused this condition we thought that we used the simple and similar queries to search twitter and the negative corpus is larger than the positive corpus.

Figure 5: Word cloud of the compiled corpus. Colors and their corresponding category: Pink – Positive and Blue - Negative.

5 Conclusion

We presented a tree kernel-based model to identify tweets posted by pregnant women. Unlike traditional approaches which detect using regular expression patterns or rules, we employ different tree

[4] A non-profit library of millions of free books, movies, software, music, websites.

[5] https://code.google.com/archive/p/word2vec/

[6] http://www.webconfs.com/stop-words.php

structures together with two tree kernels that incorporate the dependency and grammatical relation information represented in the form of tree structures. We evaluated our models on manually annotated Twitter corpus specifically developed for the purpose of this study. The results demonstrate that employing dependency tree can improve the performance of pregnancy detection. We also observe that lexical features as central node is effective in representing pregnancy information in a tweet. The best performed model had an F-score of 0.5652 on our corpus. The SPTK model considers lexical similarities on the tree structure.

In future, we would like to explore different ways to integrate deeper semantics into tree structure for pregnancy detection on social media platforms. Moreover, we also would like to improve the corpus by ignoring retweets, avoid possible noise and obtain more meta data such as timestamp details. In addition, the pregnancy related tweets may reveal the other health behaviour related information of the of the pregnant authors like drug usage, food intake, any disease symptoms, and emotion. We would like to use this information in future to conduct syndromic surveillance on pregnant women using social media.

Acknowledgments

We are grateful for the constructive comments from three anonymous reviewers. This work was supported by grant MOST106-3114-E-001- 002 and MOST105-2221-E-001-008-MY3 from the Ministry of Science and Technology, Taiwan.

Reference

Alfred. V. Aho and Jeffrey D. Ullman. 1972. The Theory of Parsing, Translation and Compiling, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

American Psychological Association. 1983. Publications Manual. American Psychological Association, Washington, DC.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. Journal of the Association for Computing Machinery, 28(1):114-133.

Association for Computing Machinery. 1983. Computing Reviews, 24(11):503-512.

Dan Gusfield. 1997. Algorithms on Strings, Trees and Sequences. Cambridge University Press, Cambridge, UK.

Adam, D., Jonnagaddala, J., Chughtai, A. A., & Macintyre, C. R. (2017). *ZikaHack 2016: A digital disease detection competition*. Paper presented at the Proceedings of the International Workshop on Digital Disease Detection using Social Media 2017, Taipei, Taiwan.

Alvaro, N., Miyao, Y., & Collier, N. (2017). TwiMed: Twitter and PubMed Comparable Corpus of Drugs, Diseases, Symptoms, and Their Relations. *JMIR Public Health and Surveillance, 3*(2).

Ayers, J. W., Althouse, B. M., & Dredze, M. (2014). Could behavioral medicine lead the web data revolution? *Jama, 311*(14), 1399-1400.

Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—harnessing the Web for public health surveillance. *New England Journal of Medicine, 360*(21), 2153-2157.

Chandrashekar, P. B., Magge, A., Sarker, A., & Gonzalez, G. (2017). Social media mining for identification and exploration of health-related information from pregnant women. *arXiv preprint arXiv:1702.02261*.

Dai, H.-J., Touray, M., Wang, C.-K., Jonnagaddala, J., & Syed-Abdul, S. (2016). Feature Engineering for Recognizing Adverse Drug Reactions from Twitter Posts. *Information*.

Dredze, M. (2012). How social media will change public health. *IEEE Intelligent Systems, 27*(4), 81-84.

GE-Healthcare. (2012). Twenty six percent of online adults discuss health information online; privacy cited as the biggest barrier to entry. Retrieved from http://www.businesswire.com/news/home/20121120005872/en/Twenty-percent-online-adults-discuss-health-information

Jonnagaddala, J., Jue, T. R., & Dai, H. *Binary classification of Twitter posts for adverse drug reactions*.

Klein, A. Z., Sarker, A., Rouhizadeh, M., O'Connor, K., & Gonzalez, G. (2017). Detecting Personal Medication Intake in Twitter: An Annotated Corpus and Baseline Classification System. *BioNLP 2017*, 136.

Rim, K. (2016). *MAE2: Portable Annotation Tool for General Natural Language Use.* Paper presented at the In Proceedings of the 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation, Portorož, Slovenia, May 28, 2016.

Sarker, A., O'Connor, K., Ginn, R., Scotch, M., Smith, K., Malone, D., & Gonzalez, G. (2016). Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. *Drug Safety, 39*(3), 231-240. doi:10.1007/s40264-015-0379-4

Wang, S., Paul, M. J., & Dredze, M. (2014). *Exploring health topics in Chinese social media: An analysis of Sina Weibo*. Paper presented at the

AAAI Workshop on the World Wide Web and Public Health Intelligence.

Simone Filice, Giuseppe Castellucci, Danilo Croce and Roberto Basili. 2015. KeLP: a Kernel-based Learning Platform for Natural Language Processing. *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pages 19–24.

Michael Collins and Nigel Duffy. 2002. New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron. In *Proceedings of ACL'02.*

Danilo Croce, Alessandro Moschitti and Robert Basili. 2011. Structured lexical similarity via convolution kernels on dependency trees. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1034–1046, Edinburgh, UK.

Alessandro Moschitti. 2004. A Study on Convolution Kernels for Shallow Semantic Parsing. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.* Association for Computational Linguistics, Stroudsburg, PA, USA, ACL'04.

Thorsten Joachims. 1999. Making Large-scale Support Vector Machine Learning Practical. In Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola, editors, *Advances in Kernel Methods*, MIT Press, Cambridge, MA, USA, pages 169–184. http://dl.acm.org/citation.cfm?id=299094.299104.

Aron Culotta and Jeffrey Sorensen. 2004. Dependency tree kernels for relation extraction. In *Proceedings of ACL'04.*

Dell Zhang and Wee Sun Lee. 2003. Question classification using support vector machines. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, New York, NY, pp. 26–32. ACM Press

Makoto Miwa, Rune Sætre, Yusuke Miyao and Jun'ichi Tsujii. 2009. Protein-protein interaction extraction by leveraging multiple kernels and parsers. *International Journal of Medical Informatics*, 78(12):e39–e46.

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In: *Proceedings of ACL'11 Workshop on Languages in Social Media*. pp. 30–38.

Anita Alicante, Anna Corazza and Antonio Piront. 2016. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016).*

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of ICLR'13 Workshop*.

Christopher D. Manning and Hinrich Schütze. 1999. Foundations of statistical natural language processing. *MIT Press. Cambridge, MA*.

Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In *Proceedings of The 17th European Conference on Machine Learning*, pages 318–329, Berlin, Germany.

Kai Wang and Zhaoyan Ming and Tat-Seng Chua. 2009. A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services. In Proceedings of SIGIR, pages 187-194.

Roberto Basili, Marco Cammisa, and Alessandro Moschitti. 2005. Effective use of WordNet semantics via kernel-based learning. In *Proceedings of CoNLL-2005*, pages 1–8, Ann Arbor, Michigan. Association for Computational Linguistics.