LTG-Oslo Hierarchical Multi-task Network: The importance of negation for document-level sentiment in Spanish

Jeremy Barnes[0000−0002−8043−8058]

Language Technology Group
University of Oslo, Norway
jeremycb@ifi.uio.no

Abstract. This paper details LTG-Oslo team's participation in the sentiment track of the NEGES 2019 evaluation campaign. We participated in the task with a hierarchical multi-task network, which used shared lower-layers in a deep BiLSTM to predict negation, while the higher layers were dedicated to predicting document-level sentiment. The multi-task component shows promise as a way to incorporate information on negation into deep neural sentiment classifiers, despite the fact that the absolute results on the test set were relatively low for a binary classification task.

Keywords: Sentiment Analysis · Negation · Multi-task

1 Introduction

Sentiment analysis has improved greatly over the last decade, moving from models trained on hand-engineered features [Pang et al., 2002, Das and Chen, 2007] to neural models that are trained in an end-to-end fashion [Socher et al., 2013]. The success of these neural architectures is often attributed to their ability to capture compositionality effects [Socher et al., 2013, Linzen et al., 2016], of which negation is the most common and influential for sentiment analysis [Wiegand et al., 2010]. However, recent research has shown that these models are still not able to fully resolve the effect that negation has on sentence-level sentiment [Barnes et al., 2019].

Explicit negation detection has proven useful to create features for lexicon-based sentiment models [Councill et al., 2010, Cruz et al., 2016] and machine-learning approaches to sentiment classification [Lapponi et al., 2012]. At the same time, these approaches build upon work on negation detection as its own task [Vincze et al., 2008, Morante and Blanco, 2012].

More recent approaches to sentiment, however, have concentrated on learning the effects of negation in an end-to-end fashion. Current state-of-the-art approaches employ neural networks which implicitly learn to resolve negation, by either directly training on sentiment annotated data [Socher et al., 2013, Tai et al., 2015], or by pre-training the model on a language modeling task [Peters et al., 2018, Devlin et al., 2019]. State-of-the-art neural methods, however, have not attempted to harness explicit negation detection models and annotated negation datasets to improve results. We hypothesize that multi-task learning (MTL) [Caruana, 1993, Collobert et al., 2011] is an appropriate framework to incorporate negation information into neural models.

In this paper, we propose a multi-task learning approach to explicitly incorporate negation annotated data into a neural sentiment model. We show that this approach improves the final result, although our model performs weakly in absolute terms.

2 Related Work

In this section, we briefly review previous work that is relevant to (*i*) attempts to use negation information in sentiment analysis, (*ii*) research on negation detection as a separate task, and (*iii*) multi-task learning.

2.1 Negation informed Sentiment Analysis

Negation is a pervasive linguistic phenomenon which has a direct effect on the sentiment of a text [Wiegand et al., 2010]. Take the following example from the SFU ReviewSP-Neg training data, where the negation cue is shown in **bold** and the scope is underlined.

Example 1.
El hotel está situado en la puerta de toledo, **no** está lejos del centro.

The English translation is “The hotel is located at the *puerta de toledo*, it is not far from the center.” A sentiment classification model must be able to identify the relevant sentiment words (in this case “lejos del centro”), negation cues (“no”), and resolve the scope in order to correctly predict that this sentence expresses negative polarity. Intuitively, a sentiment model that has access to negation scope information should perform better than a non-informed version.

The first approaches to detecting negation scope for sentiment used heuristics, such as assuming all tokens between a negation cue and the next punctuation mark are in scope [Hu and Liu, 2004]. However, this simplification does not work well on noisy text, such as tweets, or texts that use more complex syntax, such as those in the political domain.

Later research showed that using machine-learning techniques to detect the scope of negation could improve both lexicon-based [Councill et al., 2010, Cruz et al., 2016] and machine learning [Lapponi et al., 2012] classification of sentiment.

2.2 Negation detection

Approaches to negation analysis often decompose the task into two sub-tasks, performing (i) negation cue detection, followed by (ii) scope detection.

Much work was done within the biomedical domain [Morante et al., 2008, Morante and Daelemans, 2009, Velldal et al., 2012] due largely to the availability of the BioScope corpus [Vincze et al., 2008], which is annotated for negation cues and scopes. The *SEM shared task [Morante and Blanco, 2012] instead focused on detection of negation cues and scopes in a corpus of sentences taken from the works of Aurthur Conan Doyle.

Traditional approaches to the task of negation detection have typically employed a wide range of hand-crafted features describing a number of both lexical, morphosyntactic and even semantic properties of the text [Read et al., 2012, Packard et al., 2014, Lapponi et al., 2012, White, 2012, Enger et al., 2017]. More recently, research has moved towards using neural models such as CNNs [Qian et al., 2016], feed-forward networks, or LSTMs [Fancellu et al., 2016], finding that these architectures often outperform the previous methods, while requiring less hand-crafting of features.

2.3 Multi-task learning

Multi-task learning (MTL) is an approach to machine learning where a single model is trained simultaneously on two tasks. By restricting the search space of possible representations to those that are predictive for both tasks, we attempt to give the model a useful inductive bias [Caruana, 1993].

Hard parameter sharing [Caruana, 1993], which assumes that all layers are shared between tasks except for the final predictive layer, is the simplest way to implement a multi-task model. When the main task and auxiliary task are closely related, this approach has been shown to be an effective way to improve model performance [Collobert et al., 2011, Peng and Dredze, 2017, Martínez Alonso and Plank, 2017, Augenstein et al., 2018]. On the other hand, [Søgaard and Goldberg, 2016] find that it is better to make predictions for low-level auxiliary tasks at lower layers of a multi-layer MTL setup. They also suggest that under the hard-parameter framework auxiliary tasks need to be sufficiently similar to the main task for MTL to improve over the single-task baseline.

In this work, we implement a multi-task learning where the lower layers of a deep neural network are shared for the main and auxiliary tasks (in our case sentiment classification and negation detection, respectively), while higher layers are allowed to adapt to the main task.

3 Model

We propose a *hierarchical multi-task model* (see Figure 1) which relies on a BiLSTM to create a representation for each sentence in a document, and a second BiLSTM to aggregate these sentence representations into a full document representation. In this section, we first describe the negation submodel, then the sentiment submodel, and finally the multi-task model.

3.1 Negation Model

In previous work on negation detection, it is common to model negation scope as a two step process, where first the negation cues are identified, and then

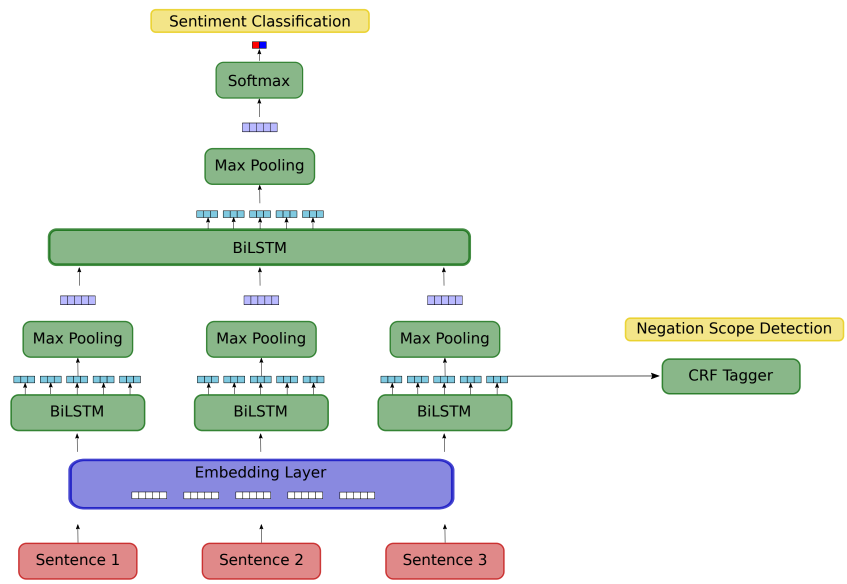

Fig. 1: Hierarchical multi-task model. The lower BiLSTM is used both to perform sequence-tagging of negation, as well as creating sentence-level features. These features are then aggregated using a second BiLSTM layer and used for predicting the sentiment at document-level.

negation scope is determined. However, we hypothesize that within a multi-task framework, it is more beneficial for a network to learn to both identify cues and resolve scope jointly. Therefore, we model negation as a *sequence labeling task* with BIO tags. In the cases where there are more than one negation scope in a sentence that overlap, we flatten these multiple representations, as shown in Figure 2. The negation model, therefore, attempts to identify all cues and all scopes in a sentence at the same time. Note that scopes can also begin before the negation cue and also be discontinuous. While this is an oversimplification of the full negation scope task, we argue that in order to classify sentiment, it is enough for a model to know which tokens are negated.

The negation model is comprised of an embedding layer which embeds the tokens for each sentence. The embeddings pass to a bidirectional Long Short-Term Memory module (BiLSTM), which creates contextualized representations of each word. A linear chain conditional random field (CRF) uses the output of the BiLSTM layer as features. We use Viterbi decoding and minimize the negative log likelihood of CRF predictions.

	Es	que	no	nos	ayudó	,	y	luego	ni	siquera	llamó
negation labels:	O	O	CUE^1	N^1	N^1	O	O	O	CUE^2	CUE^2	N^2
BIO labels:	O	O	B_cue	B_neg	I_neg	O	O	O	B_cue	I_cue	B_neg

Fig. 2: An example of the negation which has been converted to BIO labels. Although the example here shows two negation structures where the cue is at the beginning of the scope, there are also examples where the scope begins before the cue or is discontinuous.

3.2 Sentiment Model

As mentioned above, the sentiment model uses a hierarchical approach. For each sentence in a document, we first extract features with a BiLSTM. We take the max of the BiLSTM output as a representation for the sentence. This is then passed to a second BiLSTM layer, after which we again take the max. We use a softmax layer to compute the sentiment predictions for each document and minimize the cross entropy loss. As a baseline, we train a single-task sentiment model (STL) on the available sentiment data.

3.3 Multi-task Model

For the hierarchical multi-task model (MTL), we train both tasks simultaneously by sequentially training the negation classification model for one full epoch and then training the sentiment model. We use Adam as an optimizer, and a dropout layer (0.3) after the embedding layer to regularize the model, as this is common for both the main and auxiliary tasks.

4 Experimental Setup

Given that neural models are sensitive to random initialization, we perform five runs for each model on the development data with different random seeds and report both mean accuracy and standard deviation across the five runs. As the final submission required a single prediction for each document, we take a majority vote of the five learned classifiers in order to provide an ensemble prediction.

Besides the proposed STL and MTL models, we also compare with a baseline (BOW) which uses an L2 regularized logistic regression classifier trained on a bag-of-words representation of the documents. We choose the optimal C parameter on the development data.

4.1 Dataset

The SFU ReviewSP-NEG dataset [Jiménez-Zafra et al., 2018b] provided in the shared task contains 400 Spanish-language reviews from eight domains (books, cars, cellphones, computers, hotels, movies, music, and washing machines) which also contain annotations for negation cues, negation scope, and relevance of the negation to sentiment. The participants were provided with the train and dev splits, while the test split was kept from participants until after the final results were posted. Table 1 shows the statistics of the dataset.

Previous work [Jiménez-Zafra et al., 2018a] reported Macro F1 score of 75.89 when using a Bayesian logistic regression classifier trained with bag-of-words features plus negation features that indicate that negation changes the polarity of the negated phrase. However, these results are not comparable to those obtained in the shared task, as the authors evaluated their model using 10-fold cross-validation and not on the test set provided by the organizers. Additionally, they had access to negation information in the test set, which participants in the shared task do not.

Task	Train	Dev	Test
Document-level Sentiment	264	56	80
Negative Structures	2,733	645	949

Table 1: Statistics of the document-level sentiment (number of documents) and negation (number of negation structures) data provided by the organizers of the shared task.

4.2 Model performance

As we only had access to the gold labels on the development set, we report the mean and average accuracy of all three models (BOW, STL, MTL) in Table 2. Additionally, we show the official accuracy score of the MTL model on the test set[1]. BOW and STL achieve the same performance, with 71.4 accuracy on the dev set. MTL improves 1.1 percentage points over the other two models on the dev set, and reaches 66.2 accuracy on the test set. In absolute terms, the performance of all models is weak for a binary document-level classification task. This is likely due to the small number of training examples available, as well as the number of domains, which has been shown to be more problematic for machine-learning approaches than lexicon-based approaches [Taboada et al., 2011].

4.3 Error Analysis

Given that the classification task is performed at document-level, it is often difficult to determine what exactly was the cause of a change in prediction from

[1] Note that we do not have access to the gold sentiment or negation labels on the test set, so we cannot perform multiple runs, but must rely on the organizers evaluation.

Model	Dev	Test
BOW	71.4	–
STL	71.4 (5.2)	–
MTL	72.5 (1.8)	66.2

Table 2: Accuracy of the models on the development and test data. Neural models also report mean accuracy and standard deviation on the development data over five runs with different random seeds.

one model to another. Instead, Figure 3 shows a relative confusion matrix of the development results, where positive numbers (dark purple) indicate that the MTL model made more predictions in that square than the STL model and negative numbers (white) indicate fewer predictions. On the development data, the MTL model tends to help with the negative class, while adding little to the positive class. The number of negation structures per class (shown in Table 3) shows that there are more negation structures in documents labeled with negative sentiment in the development set, which seems to corroborate the idea that the MTL model is able to use negation information to improve the results on the negative class.

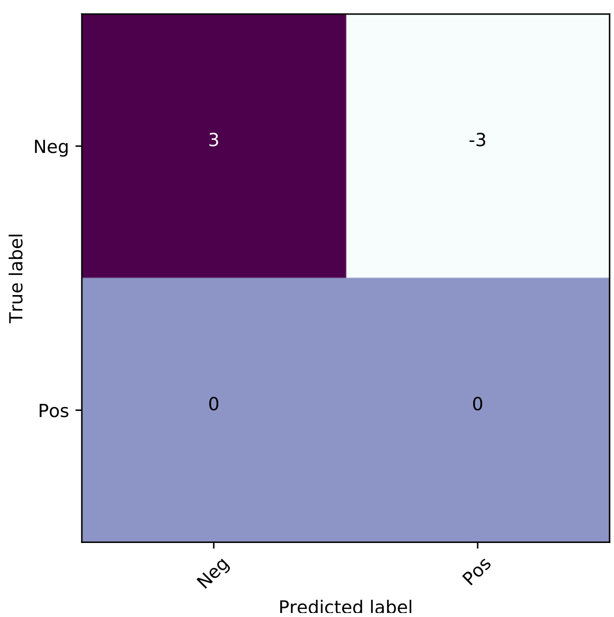

Fig. 3: A relative confusion matrix, where positive numbers (dark purple) indicate that the MTL model made more predictions in that square that the STL model and negative numbers (white) indicate fewer predictions.

	Train	Dev
Positive	1,421	303
Negative	1,312	342

Table 3: Number of negation structures per sentiment class found in the training and development data.

5 Conclusion and Future Work

In this paper, we have detailed our participation in the 2019 Neges shared task. Our approach, the hierarchical multi-task negation model, did not give a strong performance in absolute numbers on the test set (66% accuracy), but does indicate that multi-task learning is an appropriate framework for incorporating negation information into sentiment models, improving from 71.4 to 72.5 accuracy on the development set.

The hierarchical RNN model used in this participation is similar to strong performing approaches at sentence-level. However, it is not clear that it is the most adequate model for document-level classification. Convolutional neural networks [Kim, 2014] or self-attention networks [Ambartsoumian and Popowich, 2018] have shown good performance for text classification and may be better models for document-level sentiment tasks.

Additionally, the small training set size for the sentiment task (271 documents) and number of domains (8) complicates the use of deep neural architectures. Lexicon-based and linear machine-learning approaches have shown to perform quite well under these circumstances [Taboada et al., 2011, Cruz et al., 2016]. In the future, it would be interesting to use distant supervision [Tang et al., 2014, Felbo et al., 2017] to augment the sentiment signal, or cross-lingual approaches [Chen et al., 2016, Barnes et al., 2018] to improve the results.

In this work we have only explored using a sequence-labeling approach to negation scope. It would be interesting to incorporate state-of-the-art negation scope models [Fancellu et al., 2017] into a multi-task setup.

Finally, the SFU ReviewSP-NEG dataset has several additional levels of annotation, *i.e.* if a negation structure changes the polarity of the tokens in scope or the final polarity after negation. Future work should explore the use of this information further.

Acknowledgements

This work has been carried out as part of the SANT project, funded by the Research Council of Norway (grant number 270908).

References

Ambartsoumian and Popowich, 2018. Ambartsoumian, A. and Popowich, F. (2018). Self-attention: A better building block for sentiment analysis neural network classifiers. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 130–139. Association for Computational Linguistics.

Augenstein et al., 2018. Augenstein, I., Ruder, S., and Søgaard, A. (2018). Multi-task learning of pairwise sequence classification tasks over disparate label spaces. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1896–1906. Association for Computational Linguistics.

Barnes et al., 2018. Barnes, J., Klinger, R., and Schulte im Walde, S. (2018). Bilingual sentiment embeddings: Joint projection of sentiment across languages. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2483–2493. Association for Computational Linguistics.

Barnes et al., 2019. Barnes, J., Øvrelid, L., and Velldal, E. (2019). Sentiment analysis is not solved!: Assessing and probing sentiment classifiers. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, page to appear, Florence, Italy. Association for Computational Linguistics.

Caruana, 1993. Caruana, R. (1993). Multitask learning: A knowledge-based source of inductive bias. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 41–48. Morgan Kaufmann.

Chen et al., 2016. Chen, X., Athiwaratkun, B., Sun, Y., Weinberger, K. Q., and Cardie, C. (2016). Adversarial deep averaging networks for cross-lingual sentiment classification. *CoRR*, abs/1606.01614.

Collobert et al., 2011. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

Councill et al., 2010. Councill, I., McDonald, R., and Velikovich, L. (2010). What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, Uppsala, Sweden. University of Antwerp.

Cruz et al., 2016. Cruz, N. P., Taboada, M., and Mitkov, R. (2016). A machine-learning approach to negation and speculation detection for sentiment analysis. *Journal of the Association for Information Science and Technology*, 67(9):2118–2136.

Das and Chen, 2007. Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388.

Devlin et al., 2019. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Enger et al., 2017. Enger, M., Velldal, E., and Øvrelid, L. (2017). An open-source tool for negation detection: a maximum-margin approach. In *Proceedings of the EACL workshop on Computational Semantics Beyond Events and Roles (SemBEaR)*, pages 64–69, Valencia, Spain.

Fancellu et al., 2016. Fancellu, F., Lopez, A., and Webber, B. (2016). Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the*

Association for Computational Linguistics (Volume 1: Long Papers), pages 495–504, Berlin, Germany. Association for Computational Linguistics.

Fancellu et al., 2017. Fancellu, F., Lopez, A., Webber, B., and He, H. (2017). Detecting negation scope is easy, except when it isn't. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 58–63, Valencia, Spain. Association for Computational Linguistics.

Felbo et al., 2017. Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625, Copenhagen, Denmark. Association for Computational Linguistics.

Hu and Liu, 2004. Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, pages 168–177.

Jiménez-Zafra et al., 2018a. Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Molina-González, M. D., and Ureña-López, L. A. (2018a). Relevance of the SFU ReviewSP-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish. *Information Processing & Management*, 54(2):240 – 251.

Jiménez-Zafra et al., 2018b. Jiménez-Zafra, S. M., Taulé, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Martí, M. A. (2018b). SFU ReviewSP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation*, 52(2):533–569.

Kim, 2014. Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Lapponi et al., 2012. Lapponi, E., Read, J., and Øvrelid, L. (2012). Representing and resolving negation for sentiment analysis. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops*, ICDMW '12, pages 687–692, Washington, DC, USA. IEEE Computer Society.

Linzen et al., 2016. Linzen, T., Dupoux, E., and Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.

Martínez Alonso and Plank, 2017. Martínez Alonso, H. and Plank, B. (2017). When is multitask learning effective? semantic sequence prediction under varying data conditions. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 44–53. Association for Computational Linguistics.

Morante and Blanco, 2012. Morante, R. and Blanco, E. (2012). *SEM 2012 shared task: Resolving the scope and focus of negation. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 265–274, Montréal, Canada. Association for Computational Linguistics.

Morante and Daelemans, 2009. Morante, R. and Daelemans, W. (2009). A metalearning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL)*.

Morante et al., 2008. Morante, R., Liekens, A., and Daelemans, W. (2008). Learning the scope of negation in biomedical texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Packard et al., 2014. Packard, W., Bender, E. M., Read, J., Oepen, S., and Dridan, R. (2014). Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.

Pang et al., 2002. Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.

Peng and Dredze, 2017. Peng, N. and Dredze, M. (2017). Multi-task domain adaptation for sequence tagging. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 91–100, Vancouver, Canada. Association for Computational Linguistics.

Peters et al., 2018. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237. Association for Computational Linguistics.

Qian et al., 2016. Qian, Z., Li, P., Zhu, Q., Zhou, G., Luo, Z., and Luo, W. (2016). Speculation and negation scope detection via convolutional neural networks. In *The 2016 Conference on Empirical Methods in Natural Language Processing*.

Read et al., 2012. Read, J., Velldal, E., Øvrelid, L., and Oepen, S. (2012). UiO1: Constituent-based discriminative ranking for negation resolution. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)*, Montreal.

Socher et al., 2013. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the EMNLP 2013*, pages 1631–1642.

Søgaard and Goldberg, 2016. Søgaard, A. and Goldberg, Y. (2016). Deep multi-task learning with low level tasks supervised at lower layers. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 231–235. Association for Computational Linguistics.

Taboada et al., 2011. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Tai et al., 2015. Tai, K. S., Socher, R., and Manning, C. D. (2015). Improved semantic representations From tree-structured long short-term memory networks. *Association for Computational Linguistics 2015 Conference*, pages 1556–1566.

Tang et al., 2014. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1555–1565.

Velldal et al., 2012. Velldal, E., Øvrelid, L., Read, J., and Oepen, S. (2012). Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics*, 38(2):369–410.

Vincze et al., 2008. Vincze, V., Szarvas, G., Farkas, R., Móra, G., and Csirik, J. (2008). The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, Suppl 11(Suppl 11).

White, 2012. White, J. (2012). UWashington: Negation resolution using machine learning methods. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)*, Montreal.

Wiegand et al., 2010. Wiegand, M., Balahur, A., Roth, B., Klakow, D., and Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68.