

A Dataset for Multi-Target Stance Detection

Parinaz Sobhani¹, Diana Inkpen¹ and Xiaodan Zhu²

¹EECS, University of Ottawa

²National Research Council Canada

{psobh090, diana.inkpen}@uottawa.ca

{xiaodan.zhu}@nrc-cnrc.gc.ca

Abstract

Current models for stance classification often treat each target independently, but in many applications, there exist natural dependencies among targets, e.g., stance towards two or more politicians in an election or towards several brands of the same product. In this paper, we focus on the problem of multi-target stance detection. We present a new dataset that we built for this task. Furthermore, We experiment with several neural models on the dataset and show that they are more effective in jointly modeling the overall position towards two related targets compared to independent predictions and other models of joint learning, such as cascading classification. We make the new dataset publicly available, in order to facilitate further research in multi-target stance classification.

1 Introduction

The subjectivity, for example, sentiments or stances, expressed towards different targets is often considered independently. In a wide range of contexts, however, they are closely related. For example, in an electoral document, the stance toward one candidate may be relevant or even inferrable from tweets about other candidates. This could be true in many other domains, such as product reviews.

Stance detection is the task of automatically determining from the text whether the author of the text is in favor of, against, or neutral towards a proposition or target. The target may be a person, organization, government policy, movement or product.

In this paper, our first goal is to provide a benchmark dataset to jointly learn subjectivities corre-

sponding to related targets. Then, we investigate the problem of jointly predicting the stance expressed towards multiple targets (two at a time), in order to demonstrate the utility of the dataset.

The closest work related to our work is Deng and Wiebe (2015a), where sentiment toward different entities and events is jointly modeled using a rule-based probabilistic soft logic approach. The authors also made their dataset MPQA 3.0 (Deng and Wiebe, 2015b) available. However, this dataset is relatively small (it contains 70 documents) and has a potentially infinite number of targets (the target sets depend on the context), which makes it hard to train a system. Instead, we provide a reasonably large dataset for training and evaluation. Our dataset contains 4,455 tweets manually annotated for stance towards more than one target simultaneously. We will refer to this data as the Multi-Target Stance Dataset. Moreover, we make available a much larger unlabeled dataset providing more choices for users to further investigate the multi-target stance detection problem by learning more knowledge about the relationship between target entities.

We propose a framework that leverages deep neural models to jointly learn the subjectivity toward two target entities, given the text of a tweet. We treat the task as sequence-to-sequence learning, where the entire text of the tweet is mapped to a vector at the encoder side using a bidirectional recurrent neural network (RNN). On the decoder side, another RNN conditioned on the input vectors generates stance labels toward the related entities. By using an attention-based network, the model can focus on different parts of the tweet text to generate each stance label. Because stance labels are generated conditionally dependent on the previously-generated labels toward other entities, the model removes the independence assumption between different targets and specifically focuses

on the dependencies.

2 Dataset

We collected tweets related to the 2016 US election. We selected four presidential candidates: ‘Donald Trump’, ‘Hillary Clinton’, ‘Ted Cruz’, and ‘Bernie Sanders’ as our targets of interest and identified a small set of hashtags (which are not stance-indicative) related to these targets¹. We used the Twitter API to collect more than eleven millions of tweets containing any of these hashtags. For approximately 25% of the tweets, the hashtag of interest appeared at the end. Hashtags at the end of the tweets may not have any contribution to the meaning of the tweets; this means that the targets of opinions may not be the same as the targets of interest and, therefore, an inference is required. This is one of the main differences between our task and aspect-based sentiment analysis. Here is an example from our dataset. None of the targets of interest, ‘Hillary Clinton’ or ‘Bernie Sanders’, are mentioned explicitly, except by the hashtags at the end of the tweet, but humans can infer that the tweeter is likely against both of them:

Tweet: Given a choice to kill 100 ISIS or 100 white American men, leftist scum would choose the latter. #UniteBlue #nomorerefugees #Bernie #Hillary

2.1 Data Annotation

We selected three target pairs for our Multi-Target Stance Dataset: Donald Trump and Hillary Clinton, Donald Trump and Ted Cruz, Hillary Clinton and Bernie Sanders. Further, we filtered the collected tweets by removing short tweets, retweets and those having a URL. We also discarded tweets that do not include at least two hashtags, one for each of the targets of interest. For each of the three selected target pairs, we randomly sampled 2,000 tweets. These tweets were annotated through CrowdFlower². We asked the annotators two questions, one for the stance towards each of the presidential candidates in the target pair of interest. For stance annotation, the same annotation instructions were used as in (Mohammad et al., 2016c).

We used CrowdFlower’s gold annotations scheme for quality control, wherein about 10%

¹Our hashtags list includes: #DonaldTrump, #Trumpt, #Trump2016, #TedCruz, #Cruz, #Cruz2016, #TedCruz2016, #HillaryClinton, #Hillary, #Hillary2016, #BernieSanders, #Bernie, #Bernie2016

²<http://www.crowdfunder.com>

Target Pair	# total	# train	# dev	# test
Clinton-Sanders	1366	957	137	272
Clinton-Trump	1722	1240	177	355
Cruz-Trump	1317	922	132	263
Total	4455	3119	446	890

Table 1: Distribution of instances in the Train, Development and Test sets for different target pairs in the Multi-Target Stance Dataset

of the data was annotated internally (by the authors). During crowd annotation, these gold questions were interspersed with other questions, and the annotator was not aware which is which. However, if she got a gold question wrong, she was immediately notified of it. If the accuracy of the annotations on the gold questions falls below 70%, the annotator was refused further annotation. This served as a mechanism to avoid malicious annotations and as a guide to the annotators.

Each tweet was annotated by at least eight annotators. To aggregate stance annotation information from multiple annotators for an instance rather than opting for a simple majority, the instances with less than 50% agreement on any of the candidates in the target pairs were discarded. We refer to this dataset as the Multi-Target Stance Dataset and we make it available online³. The inter-annotator agreement on this dataset is 79.74%. We kept the rest of the tweets that were not used in the annotation process as unlabeled data, which can be used to obtain additional information about stance and relations between relevant entities.

2.2 The Multi-Target Stance Dataset

We partitioned the Multi-Target Stance Dataset into training, development, and test sets, based on the timestamps of the tweets. All annotated tweets were ordered by their timestamps; the first 70% of the tweets formed the training set, the next 10% the development set, and the last 20% formed the test set. Table 1 shows the number of instances in the training, development, and test sets over different target pairs in our Multi-Target Stance Dataset.

Having different US presidential candidates as the targets of interest does not necessarily imply that the tweeters have opposing positions toward them. There are several cases where authors have favorable stances towards both, or similarly, opposing positions towards both of them. In our dataset, approximately 20% of the tweet-

³The dataset is available at: http://www.site.uottawa.ca/~diana/resources/stance_data/

Opinion		Clinton		
Toward		favor	against	neither
Sanders	favor	7.5	33.9	3.7
	against	12.6	12.0	3.8
	neither	2.3	5.6	18.6

Table 2: Distribution across the 9 stance classes for the Hillary Clinton-Bernie Sanders target pair

Opinion		Clinton		
Toward		favor	against	neither
Trump	favor	0.5	52.3	1.2
	against	14.0	9.0	3.5
	neither	0.3	3.9	15.2

Table 3: Distribution across the 9 stance classes for the Donald Trump-Hillary Clinton target pair

ers have the same position towards both entities, 50% of tweeters have opposing positions towards the given targets, and for 17% of the data, the positions towards none of the targets is inferable. The example below shows a tweet that have the same position towards two candidates:

Targets: Donald Trump & Hillary Clinton

Tweet: *Looking at the List of PC's for 2016 is like looking at the McDonalds Menu. You just know that shit is bad for you. #Trump2016 #Hillary2016*

To illustrate more details about the correlation between subjectivities towards targets of interest, the stance distribution across the 9 classes for different target pairs in the Multi-Target Stance Dataset are depicted in tables 2, 3 and 4. We note that the numbers vary between target pairs.

3 Multi-Target Stance Classification

In this section, we propose a framework that leverages recurrent neural models to capture the potentially complicated interaction between subjectivities expressed towards multiple targets. We experimentally show that the attention-based encoder-decoder framework is more effective in jointly modeling the overall position towards two related targets, compared to independent predictions of positions and other popular frameworks for joint learning, such as cascading classification.

3.1 Window-Based Classification

One popular approach to detect subjectivity towards different targets, as is used in aspect-based sentiment classification (Brychcín et al., 2014), is to consider a context window of size n in both directions around the target terms and to extract features for that target’s classifier based on its context. This approach is based on the assumption

Opinion		Cruz		
Toward		favor	against	neither
Trump	favor	18.7	22.5	2.8
	against	10.3	17.4	4.8
	neither	3.3	2.3	18.0

Table 4: Distribution across the 9 stance classes for the Ted Cruz-Donald Trump target pair

that the words outside the context window do not have an influence on the target. We will first include such a baseline for our task.

3.2 Cascading Classifiers

To capture dependencies between stance labels of related targets, one possibility is to use the predicted class toward one target as an extra feature in other targets’ models. This framework is based on cascade classification, where several classifiers of related tasks are combined to improve the overall system performance (Heitz et al., 2009). we adopted this framework for multi-target stance classification by starting from an independent classifier to predict stance toward the first target based on the text representation and exploit its prediction as an extra feature for other classifiers.

The major restriction of this framework is that the classification algorithm should have a mechanism to add new features based on other learners’ outputs. Most of the machine learning algorithms for text classification that rely on hand-crafted features extracted from text to represent it, provide such mechanism, but, for the state-of-the-art deep neural models, where the feature vectors for the text representation are learned with the classification model during training, adding new features to the model is not trivial.

3.3 Sequence-to-Sequence Model to Capture Dependencies in Output Space

Encoder-decoder sequence-to-sequence models (Sutskever et al., 2014; Cho et al., 2014b) were originally used for machine translation, where a recurrent neural network is trained to learn the representation for the source language and generate the translation in the target language. Later, it was proven to be effective for many different tasks such as speech recognition (Hannun et al., 2014) and question answering (Hermann et al., 2015). Bahdanau et al. (2014) extended the encoder-decoder architecture by an attention-based mechanism where the model is capable of automatically searching for more relevant regions in the input when handling different output targets.

We propose to use the attention-based encoder-decoder for multi-target stance classification. Specifically, we regard the given tweet as the input, and the model is trained to generate the stance labels for targets. This model can naturally capture the dependencies among the target stance labels when searching the best label sequence, based on automatically-learned input features. The attention mechanism has the potential of dynamically focusing on different words of the input text to generate stance labels for each target of interest. As such, the attention-based encoder-decoder is expected to have the strengths of both the window-based classification, by dynamically customizing the feature vector to predict each target stance label, and the cascading classification, by conditioning each label generation on the other labels without inheriting the limitations of these models. The model automatically learns the features and regions of the input that should be paid attention to.

4 Experiments

We evaluate the effectiveness of our models on the multi-target stance dataset described earlier, where two stance labels are predicted for each tweet. Note that all the models can be easily extended to predict more than two labels as well. For all methods, the tweets were tokenized with the CMU Twitter NLP tool (Gimpel et al., 2011). All the models we proposed were implemented in Python.

As the evaluation measure for each target, we use the average of the F1-scores (the harmonic mean of precision and recall) for the two main classes, Favor and Against. A similar metric was used for stance detection—SemEval 2016 Task 4 (Mohammad et al., 2016a). For multiple targets (in our dataset, target pairs) the average over all the targets is calculated. To report a single number for all three target pairs, we take the average of three values returned for each target pair and we refer to it as macro-averaged F-score. All the models are evaluated on the test sets.

As mentioned before, we used encoder-decoder attention-based deep models for multi-target stance detection. We followed (Bahdanau et al., 2014; Luong et al., 2015) to train our models using the minibatch stochastic gradient descent (SGD) algorithm with adaptive learning rate (Adadelta (Zeiler, 2012)). As RNN unit, we used a Gated Recurrent Unit (Cho et al., 2014a) with 128 cells. The word vectors at the embedding layer have 100

dimensions. All the parameters are initialized randomly, but the word vectors are pretrained using related unlabeled tweets (11,873,771 tweets) that we collected in the same time period. As training algorithm, we employed the Word2Vec Skip-gram model (Mikolov et al., 2013).

4.1 Results and Discussion

Table 5 presents the macro-averaged F-scores of different models on the Multi-Target Stance dataset. Row i. shows the result obtained by a random classifier and row ii. shows the result obtained by the majority classifier. When we have multiple targets to predict overall positions towards them, one possibility is to have a single learners per target that are independently trained. Row a. shows the result of having two independent linear Support Vector Machine (SVM) classifiers whose parameters are tuned using the development datasets. We used the implementation provided in the Scikit-learn Machine Learning library (Pedregosa et al., 2011). Row b. is the result of applying Window-based SVM on our Multi-Target Stance Dataset. Because we collected our data based on hashtags related to the targets, those hashtags can be considered as target terms and we place a context window around them. We used the development set to find the best value for the window size. The main limitation of this approach on this dataset is that for the majority of the tweets, the contexts windows have significant overlaps, as the two hashtags appeared in the close vicinity of each other. Row c. presents the results of the Cascading SVMs; this model shows improvement over the baseline of independent SVMs.

Another possibility when there is more than one output to predict is to combine all the outputs and train a single model. For our task of predicting stance toward a target pair, where each can take one of the three possible labels: “Favor”, “Against” and “None”, combining the two prediction results in a 9-class learning problem. Row A. shows the result of this classifier. The main limitation of combining outputs is that the number of classes can grow substantially, while there is a fixed number of labeled instances which results in a drop in performance. Another issue is that some of the classes might not have enough representative instances and this can lead to a highly imbalanced classification problem. Row B. shows the results of applying the attention-based encoder-

Classifier	F-macro
<i>Baselines</i>	
i. random	34.26
ii. majority	32.11
<i>One Classifier per Target</i>	
a. Independent SVMs	51.37
b. Window-based SVMs	48.32
c. Cascading SVMs	52.05
<i>Single Model</i>	
A. 9-Class SVM	50.63
B. Seq2Seq	54.81

Table 5: Macro-averaged F-scores of different models on the Multi-Target Stance dataset

decoder deep neural model on our dataset. This model has both the advantages of windows-based and cascading classification, and it has the best performance compared to all other models and baselines. By applying paired t-test on these results, we concluded that the differences between sequence-to-sequence model and all other models are statistically significant.

5 Related Work

Stance Detection Over the last decade, there has been active research in modeling stance (Thomas et al., 2006; Somasundaran and Wiebe, 2009; Anand et al., 2011; Sobhani et al., 2015; Walker et al., 2012a; Hasan and Ng, 2013; Sobhani et al., 2016). However, all of these previous works treat each target independently, ignoring the potential dependencies that could exist among related targets. Stance detection was one of the tasks in the SemEval-2016 shared task competition (Mohammad et al., 2016a). Out of 19 participant teams, most used standard text classification features such as n -grams and word embedding vectors, as well as standard sentiment analysis features, while others used deep neural models such as RNNs and convolutional neural nets.

Most of the existing datasets for stance detection were created from online debate forums like 4forums.com and createdebates.com (Somasundaran and Wiebe, 2010; Walker et al., 2012b; Hasan and Ng, 2013). The majority of these debates are two-sided, and the data labels are often provided by the authors of the posts. Recently, Mohammad et al. (2016b) created a dataset of tweets labeled for both stance and sentiment. None of the prior work has created a dataset annotated for more than one target simultaneously, neither has explored the dependencies and relationships between targets when predicting overall

positions towards them.

Deep Recurrent Neural Models Different structures of deep RNNs have recently shown to be very effective in a wide range of sequence modeling problems, particularly for opinion mining and sentiment analysis (Zhu et al., 2015a; Socher et al., 2013; Zhu et al., 2015b; Irsoy and Cardie, 2014; Zhu et al., 2016). These neural models were extended for tasks with variable input and output sequence length including: end-to-end neural machine translation (Sutskever et al., 2014; Cho et al., 2014b), image-to-text conversion (Vinyals et al., 2015b), syntactic constituency parsing (Vinyals et al., 2015a) and question answering (Hermann et al., 2015). Subsequently, the attention mechanism allowed the models to learn alignments between different parts of the source and the target such as between speech frames and the text in speech recognition (Chorowski et al., 2014) or between image frames and the agent’s actions in dynamic control problems (Mnih et al., 2014). We are the first to adopt these techniques for the task of multi-target stance classification.

6 Conclusions and Future Work

We presented the first multi-target stance dataset of a reasonable size from social media, to help further exploration of this task. Each tweet is annotated for position toward more than one target. By making this dataset available, more work on joint learning of subjectivities corresponding to related targets is encouraged. In addition, we presented a framework that relieves the independence assumption by jointly modeling the subjectivity expressed towards multiple targets. We experimentally showed that the attention-based encoder-decoder model is more effective in jointly modeling the overall position toward two related targets, compared to independent predictions of positions and other popular frameworks for joint learning, such as cascading classification.

Directions of future work include annotating a similar dataset for other domains, for example, several brands of the same product, and exploring transfer learning where a model trained for a target pair can be transferred to other related target pairs.

Acknowledgments

The first author of this paper was supported by the Natural Sciences and Engineering Research Council of Canada under the CREATE program.

References

- Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 1–9, Portland, Oregon, June. Association for Computational Linguistics.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Tomáš Brychcín, Michal Konkol, and Josef Steinberger. 2014. Uwb: Machine learning approach to aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 817–822, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014a. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014b. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Jan Chorowski, Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. End-to-end continuous speech recognition using attention-based recurrent nn: first results. *arXiv preprint arXiv:1412.1602*.
- Lingjia Deng and Janyce Wiebe. 2015a. Joint prediction for entity/event-level sentiment analysis using probabilistic soft logic models. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 179–189, Lisbon, Portugal, September. Association for Computational Linguistics.
- Lingjia Deng and Janyce Wiebe. 2015b. Mpqa 3.0: An entity/event-level sentiment corpus. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1323–1328, Denver, Colorado, May–June. Association for Computational Linguistics.
- Kevin Gimpel, Nathan Schneider, Brendan O’Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 42–47, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, et al. 2014. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.
- Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356, Nagoya, Japan, October. Asian Federation of Natural Language Processing.
- Jeremy Heitz, Stephen Gould, Ashutosh Saxena, and Daphne Koller. 2009. Cascaded classification models: Combining models for holistic scene understanding. In *Advances in Neural Information Processing Systems*, pages 641–648.
- Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems*, pages 1693–1701.
- Ozan Irsoy and Claire Cardie. 2014. Opinion mining with deep recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 720–728, Doha, Qatar, October. Association for Computational Linguistics.
- Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.
- Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. 2014. Recurrent models of visual attention. In *Advances in Neural Information Processing Systems*, pages 2204–2212.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016a. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California, June. Association for Computational Linguistics.
- Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016b. A dataset for detecting stance in tweets. In *Proceedings of the Language Resources and Evaluation Conference*, Portorož, Slovenia.

- Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2016c. Stance and sentiment in tweets. *Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media*, In Press.
- Fabian Pedregosa, Gaël Varoquaux, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Parinaz Sobhani, Diana Inkpen, and Stan Matwin. 2015. From argumentation mining to stance classification. In *Proceedings of the Workshop on Argumentation Mining*, pages 67–77, Denver, Colorado, USA.
- Parinaz Sobhani, Saif M. Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (*Sem)*, Edinburgh, Scotland.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, Washington, USA, October. Association for Computational Linguistics.
- Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 226–234, Suntec, Singapore.
- Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, Los Angeles, CA, June. Association for Computational Linguistics.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327–335, Sydney, Australia, July. Association for Computational Linguistics.
- Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. 2015a. Grammar as a foreign language. In *Advances in Neural Information Processing Systems*, pages 2773–2781.
- Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015b. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164.
- Marilyn Walker, Pranav Anand, Rob Abbott, and Ricky Grant. 2012a. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 592–596, Montréal, Canada, June. Association for Computational Linguistics.
- Marilyn Walker, Grace Lin, and Jennifer Sawyer. 2012b. An annotated corpus of film dialogue for learning and characterizing character style. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 1373–1378, Istanbul, Turkey, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1657.
- Matthew D. Zeiler. 2012. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.
- Xiaodan Zhu, Hongyu Guo, and Parinaz Sobhani. 2015a. Neural networks for integrating compositional and non-compositional sentiment in sentiment composition. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 1–9, Denver, Colorado, June. Association for Computational Linguistics.
- Xiaodan Zhu, Parinaz Sobhani, and Hongyu Guo. 2015b. Long short-term memory over recursive structures. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1604–1612.
- Xiaodan Zhu, Parinaz Sobhani, and Hongyu Guo. 2016. Dag-structured long short-term memory for semantic compositionality. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 917–926, San Diego, California, June. Association for Computational Linguistics.