Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction[*]

Xin Li and **Wai Lam**

Key Laboratory on High Confidence Software Technologies (Sub-Lab, CUHK),
Ministry of Education, and
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Hong Kong
{lixin, wlam}@se.cuhk.edu.hk

Abstract

We propose a novel LSTM-based deep multi-task learning framework for aspect term extraction from user review sentences. Two LSTMs equipped with extended memories and neural memory operations are designed for jointly handling the extraction tasks of aspects and opinions via memory interactions. Sentimental sentence constraint is also added for more accurate prediction via another LSTM. Experiment results over two benchmark datasets demonstrate the effectiveness of our framework.

1 Introduction

The aspect-based sentiment analysis (ABSA) task is to identify opinions expressed towards specific entities such as *laptop* or attributes of entities such as *price* (Liu, 2012a). This task involves three subtasks: Aspect Term Extraction (ATE), Aspect Polarity Detection and Aspect Category Detection. As a fundamental subtask in ABSA, the goal of the ATE task is to identify opinionated aspect expressions. One of most important characteristics is that opinion words can provide indicative clues for aspect detection since opinion words should co-occur with aspect words. Most publicly available datasets contain the gold standard annotations for opinionated aspects, but the ground truth of the corresponding opinion words is not commonly provided. Some works tackling the ATE task ignore the consideration of opinion words and just focus on aspect term modeling and learning (Jin

et al., 2009; Jakob and Gurevych, 2010; Toh and Wang, 2014; Chernyshevich, 2014; Manek et al., 2017; San Vicente et al., 2015; Liu et al., 2015; Poria et al., 2016; Toh and Su, 2016; Yin et al., 2016). They fail to leverage opinion information which is supposed to be useful clues.

Some works tackling the ATE task consider opinion information (Hu and Liu, 2004a,b; Popescu and Etzioni, 2005; Zhuang et al., 2006; Qiu et al., 2011; Liu et al., 2012b, 2013a,b, 2014) in an unsupervised or partially supervised manner. Qiu et al. (2011) proposed Double Propagation (DP) to collectively extract aspect terms and opinion words based on information propagation over a dependency graph. One drawback is that it heavily relies on the dependency parser, which is prone to generate mistakes when applying on informal online reviews. Liu et al. (2014) modeled relation between aspects and opinions by constructing a bipartite heterogenous graph. It cannot perform well without a high-quality phrase chunker and POS tagger reducing its flexibility. As unsupervised or partially supervised frameworks cannot take the full advantages of aspect annotations commonly found in the training data, the above methods lead to deficiency in leveraging the data. Recently, Wang et al. (2016) considered relation between opinion words and aspect words in a supervised model named RNCRF. However, RNCRF tends to suffer from parsing errors since the structure of the recursive network hinges on the dependency parse tree. CMLA (Wang et al., 2017a) used a multi-layer neural model where each layer consists of aspect attention and opinion attention. However CMLA merely employs standard GRU without extended memories.

We propose MIN (Memory Interaction Network), a novel LSTM-based deep multi-task learning framework for the ATE task. Two LSTMs with extended memory are designed for handling

the extraction tasks of aspects and opinions. The aspect-opinion relationship is established based on neural memory interactions between aspect extraction and opinion extraction where the global indicator score of opinion terms and local positional relevance between aspects and opinions are considered. To ensure that aspects are from sentimental sentences, MIN employs a third LSTM for sentimental sentence classification facilitating more accurate aspect term extraction. Experiment results over two benchmark datasets show that our framework achieves superior performance.

2 Model

2.1 Overview

Let an input review sentence with T word tokens and the corresponding distributed representations be w $= \{w_1, ..., w_T\}$ and x $= \{x_1, ..., x_T\}$ respectively. The ATE task is treated as a sequence labeling task with BIO tagging scheme and the set of aspect tags for the word w_t is $y_t^A \in \{B, I, O\}$, where B, I, O represent beginning of, inside and outside of the aspect span respectively. Commonly found training data contains gold annotations for aspect terms and opinionated sentences, but the gold standard of opinion words are usually not available.

In our multi-task learning framework, three tasks are involved: (1) aspect term extraction (ATE), (2) opinion word extraction and (3) sentimental sentence classification. We design a task-specific LSTM, namely, A-LSTM, O-LSTM and S-LSTM, for tackling each of the above tasks respectively. The first component of our proposed framework consists of A-LSTM and O-LSTM where we equip LSTMs with extended operational memories and some operations are defined over the memories for task-level memory interactions. The second component is to determine if a review sentence is sentimental. This is achieved by employing a vanilla LSTM, namely, S-LSTM.

2.2 Model Description

The first component of our framework MIN is composed of A-LSTM and O-LSTM. Both LSTMs have extended memories for task-level memory interactions. A-LSTM involves a large aspect memory $H_t^A \in \mathbb{R}^{n_m \times dim_h^A}$ and an opinion summary vector $m_t^O \in \mathbb{R}^{dim_h^O}$ where H_t^A contains n_m pieces of aspect hidden states of dimension dim_h^A and m_t^O is distilled from H_t^O. As for

O-LSTM, similarly, an opinion memory $H_t^O \in \mathbb{R}^{n_m \times dim_h^O}$ and an aspect-specific summary vector $m_t^A \in \mathbb{R}^{dim_h^A}$ are included.

We use the aspect term annotations in the training data for training A-LSTM. As there is no ground truth available for opinion words in the training data, sentiment lexicon and high-precision dependency rules are introduced to find potential opinion words. Commonly used opinion words can be found in some general sentiment lexicons. To find opinion words, not in sentiment lexicons, in a sentence, we build a small rule set $\mathcal{R}$ composed of dependency relations with high confidence, e.g., $amod$, $nsubj$, and determine if w_t directly depends on the gold aspect word through the dependencies in $\mathcal{R}$. If so, w_t will be treated as a potential opinion word. Then such opinion words are used as training data for O-LSTM.

In the memory-enhanced A-LSTM and O-LSTM, we manually design three kinds of operations: (1) **READ** to select n_m pieces of aspect (opinion) hidden states from the past memories and build H_t^A (H_t^O); (2) **DIGEST** to distill an aspect (opinion)-specific summary m_t^A (m_t^O) from H_t^A (H_t^O) where influences of opinion terms and relative positions of inputs are considered; (3) **INTERACT** to perform interaction between A-LSTM and O-LSTM using the task specific summaries (i.e., m_t^A and m_t^O).

Consider the work flow of A-LSTM for aspect term extraction. Since opinion words and aspect terms should co-occur, the goal of A-LSTM participating in memory interactions is to acquire opinion summaries from O-LSTM (i.e., m_t^O) for better aspect prediction. First of all, MIN will **READ** n_m pieces of opinion memories which are most related to w_t from O-LSTM. Syntax structure could be used but syntactic parsers are not effective for processing short informal review sentences. Therefore, MIN selects memory segments temporally related to w_t. Precisely, the opinion memory at the time step t is $H_t^O = [h_{t-1}^O; ...; h_{t-n_m}^O]$ where h_{t-i}^O is the $(t-i)$-th hidden state from O-LSTM. Since the linear context contains most of the parent nodes and the child nodes of w_t on the dependency parse tree, treating the corresponding memory segments as relevant segments to w_t is reasonable.

Then MIN will **DIGEST** the collected opinion memories H_t^O in the A-LSTM. As different memory segments are not of equal importance for the

current decision and the same segment in different memories (i.e., different H_t^O) also makes a difference, MIN leverages two kind of weights to summarize the collected content. The first weight is the indicator score of being opinion terms denoted as $v^I \in \mathbb{R}^{n_m}$, which is used to measure how much opinion information the word w_{t-i} ($i = 1, .., n_m$) holds. We adopt Euclidean distance between distributed representations of w_{t-i} and opinion words. It is obvious that computing the distance between x_{t-i} and each opinion word is expensive. Thus, we run an off-the-shelf clustering algorithm over opinion words in the training set and then use the produced n_c centroids to estimate the indicator score v_i^I of w_{t-i} being an opinion word:

$$v_i^I = \sum_{j=1}^{n_c} \frac{1}{||x_{t-i} - c_j||_2} \quad (1)$$

where x_{t-i} is the distributed representation of w_{t-i} and c_j is the centroid vector representation of j-th cluster. This weighting scheme ensures that w_{t-i} is assigned a high score as long as x_{t-i} is close to a particular centroid. The aspect decision of w_t is also affected by relative position between w_{t-i} and w_t. Thus, MIN employs the second weight v^P to explicitly model their positional relevance and the initial weight for the i-th segment v_i^P is calculated as below:

$$v_i^P = \frac{n_m - i + 1}{\sum_{k=1}^{n_m} k} \quad (2)$$

where n_m is the number of hidden state in H_t^O. This position-aware weight enables that the closer the word w_{t-i} is to the current input, the more the corresponding memory segment will contribute to the current decision. To better capture the local positional relevance, we make the initialized v^P as learnable parameters. Combining the above two weights helps to utilize each active memory segment according to the importance for prediction and m_t^O, the summary of H_t^O is generated:

$$m_t^O = (H_t^O)^\top (\frac{v^I \odot v^P}{||v^I||_2}) \quad (3)$$

where $\odot$ denotes element-wise multiplication and $|| * ||_2$ is Euclidean norm of vectors. From Equation 3, m_t^O is dominated by the associated memory segment of w_{t-i} that obtains the high combined weights.

In the last operation **INTERACT**, A-LSTM communicates with O-LSTM by acquiring m_t^O from O-LSTM and incorporating the summary into the memory update. The update process is as follows:

$$\begin{aligned}
i_t^A &= \sigma(W_i^A x_t + U_i^A[H_t^A[1] : m_t^O]) + b_i^A) \\
f_t^A &= \sigma(W_f^A x_t + U_f^A[H_t^A[1] : m_t^O]) + b_f^A) \\
\hat{c}_t^A &= \tanh(W_c^A x_t + U_c^A[H_t^A[1] : m_t^O]) + b_c^A) \\
o_t^A &= \sigma(W_o^A x_t + U_o^A[H_t^A[1] : m_t^O]) + b_o^A) \\
c_t^A &= i_t^A \odot \hat{c}_t^A + c_{t-1}^A \odot f_t^A \\
h_t^A &= \tanh(c_t^A) \odot o_t^A
\end{aligned} \quad (4)$$

where W_*^A, U_*^A and b_*^A are weight parameters of the A-LSTM and σ is the *sigmoid* activation function. $[:]$ denotes vector concatenation operation. m_t^O can be seen as the summary of the opinion indicator in the left context of w_t and $H_t^A[1]$ is the most immediate hidden memory of A-LSTM. MIN blends the opinion summary from O-LSTM with the memory from A-LSTM. The co-occurrence relation between aspects and opinion words is modeled by such "memory fusion" strategy. Since opinion words can appear on both sides of w_t, memory segments corresponding to the right context (i.e., "future" memory) should be included. Hence, we conduct bi-directional training for A-LSTM.

The work flow of memory interaction and the update process of the internal memories in O-LSTM are kept same with those in A-LSTM except the **DIGEST** operation. Specifically, we set m_t^A, the task-specific summary of A-LSTM, as h_t^A.

The second component of MIN is a generic LSTM called S-LSTM for discriminating sentimental sentences and non-sentimental sentences. The design and the process of the memory update in this component are similar to that in Jozefowicz et al. (2015). In sentences not conveying any sentimental meanings, some words like *food*, *service* tend to be misclassified as aspect terms since they are commonly used in user reviews. To avoid this kind of error, we add a constraint that an aspect term should come from sentimental sentence. Specifically, S-LSTM learns the sentimental representation h_T^S of the sentence and then feeds it in aspect prediction as a soft constraint:

$$P(y_t^A|x_t) = \text{softmax}(W_{fc}^A([h_t^A : h_T^S])) \quad (5)$$

where W_{fc}^A denotes the weight matrix of the fully-connected softmax layer.

On the whole, our proposed MIN framework has three LSTMs and each of them is differentiable. Thus, our MIN framework can be efficiently trained with gradient descent. For A-LSTM and O-LSTM, we use the token-level cross-entropy error between the predicted distribution $P(y_t^{\mathcal{T}}|x_t)$ and the gold standard distribution $P(y_t^{\mathcal{T},g}|x_t)$ as the loss function ($\mathcal{T} \in \{A, O\}$):

$$Loss(\mathcal{T}) = -\frac{1}{N*T} \sum_{i=1}^{N} \sum_{t=1}^{T} P(Y_{i,t}^{\mathcal{T},g}|X_{i,t})\odot \quad (6)$$
$$\log[P(Y_{i,t}^{\mathcal{T}}|X_{i,t})]$$

For S-LSTM, sentence-level cross entropy error are employed to calculate the corresponding loss:

$$Loss(S) = -\frac{1}{N} \sum_{i=1}^{N} P(Y_i^{S,g}|X_i) \odot \log[P(Y_i^{S}|X_i)]$$
$$(7)$$

Then, losses from different LSTMs are combined to form the training objective of the MIN framework:

$$J(\theta) = Loss(A) + Loss(O) + Loss(S) \quad (8)$$

.

	#TRAIN/#TEST Sentences	#TRAIN/#TEST Aspects
D_1	3045/800	2358/654
D_2	2000/676	1743/622

Table 1: Statistics of datasets.

3 Experiment

3.1 Dataset

We conduct experiments on two benchmark datasets from SemEval ABSA challenge (Pontiki et al., 2014, 2016) as shown in Table 1. D_1 (SemEval 2014) contains reviews from the laptop domain and D_2 (SemEval 2016) contains reviews from the restaurant domain. In these datasets, aspect terms have been labeled and sentences containing at least one golden truth aspect are regarded as sentimental sentences. As gold standard annotations for opinion words are not provided, we select words with strong subjectivity from MPQA[1] as potential opinion words. Apart from the common opinion words in the sentiment lexicon, we also treat words, which directly depend on gold standard aspect terms through high-precision dependency rules, as opinion words.

3.2 Experiment Design

To evaluate the proposed MIN framework, we perform comparison with the following two groups of methods:
(1) *CRF* based methods:

- **CRF**: Conditional Random Fields with basic feature templates[2] and word embeddings.

- **Semi-CRF**: First-order semi-Markov conditional random fields (Sarawagi et al., 2004) and the feature template in Cuong et al. (2014) is adopted.

- **IHS_RD** (Chernyshevich, 2014), **NLANGP** (Toh and Su, 2016): Best systems in ATE subtask in SemEval ABSA challenges (Pontiki et al., 2014, 2016).

- **DLIREC** (Toh and Wang, 2014), **AUEB** (Xenos et al., 2016): Top-ranked CRF-based systems in ATE subtask in SemEval ABSA challenges (Pontiki et al., 2014, 2016).

- **WDEmb** (Yin et al., 2016): Enhanced CRF with word embeddings, linear context embeddings and dependency path embeddings.

(2) *Neural Network* based methods

- **LSTM**: Vanilla bi-directional LSTM with pre-trained word embeddings[3].

- **RNCRF** (Wang et al., 2016): Dependency Tree based Recursive Neural Network with CRF extractor[4].

For datasets in the restaurant domain, we train word embeddings of dimension 200 with word2vec (Mikolov et al., 2013) on Yelp reviews[5]. For those in laptop domain, we use pre-trained $glove.840B.300d$[6].

---

[1] http://mpqa.cs.pitt.edu/

[2] http://sklearn-crfsuite.readthedocs.io/en/latest/

[3] As we use our own implementation of LSTM, the reported results are different from those in (Liu et al., 2015)

[4] Specifically, we list the result of RNCRF over D_1 without opinion annotations for fair comparison. As no result is provided for RNCRF-no-opinion over D_2, we report the corresponding performance of the full model. See their following works (Wang et al., 2017a,b). Also, CMLA (Wang et al., 2017a) reports better results than RNCRF but we do not compare with it. The reason is that CMLA introduces the gold standard opinion labels in the training data while such labels are not available for our experiments

[5] https://www.yelp.com/dataset_challenge

[6] https://nlp.stanford.edu/projects/glove/

	D_1	D_2
CRF	74.01%	69.56%
Semi-CRF	68.75%	66.35%
IHS_RD	74.55%	-
DLIREC	73.78%	-
NLANGP	-	72.34%
AUEB	-	70.44%
WDEmb	75.16%	-
LSTM	75.25%	71.26%
RNCRF	77.26%	69.74%
Our Work	**77.58%**	**73.44%**

Table 2: Experiment results

The hyper-parameters are selected via ten-fold cross validation. The dimension of hidden representations are 100, 20, 40 for A-LSTM, O-LSTM and S-LSTM respectively. The dropout rate for O-LSTM and S-LSTM is 0.4. The size of the aspect (opinion) memory n_m is 4. The batch size is set to 32. As for initialization of network parameters, we adopt the strategy that the initial weights are sampled from the uniform distribution (Glorot and Bengio, 2010). We employ ADAM (Kingma and Ba, 2014) as optimizer and the default settings of ADAM are used.

To better reveal the capability of the proposed MIN, we train 5 models with the same group of hyper-parameters and report the average F_1 score over the testing set.

3.3 Results and Analysis

Table 2 depicts experiment results. Compared to the best systems in SemEval challenge, MIN achieves 3.0% and 1.1% absolute gains on D_1 and D_2 respectively. Besides, our MIN outperforms WDEmb, a strong CRF-based system benefiting from several kinds of useful word embeddings, by 2.1% on D_1. With memory interactions and consideration of sentimental sentence, our MIN boosts the performance of vanilla bi-directional LSTM (+2.0% and +1.7% respectively). It validates the effectiveness of the manually designed memory operations and the proposed memory interaction mechanism. MIN also outperforms the state-of-the-art RNCRF on each dataset suggesting that memory interactions can be an alternative strategy instead of syntactic parsing. To further study the impact of each element in MIN, we conduct ablation experiments. As shown in Table 3, removing bi-directionality decreases the extraction performances (-2.0% and -1.0%). The soft

sentimental constraint proves to be useful since MIN is 1.5% and 1.0% superior than the framework without S-LSTM on D_1 and D_2 respectively. O-LSTM brings in the largest performance gains on D_2 compared with ablated framework (i.e., MIN without O-LSTM), verifying our postulation that aspect-opinion "interaction" is more effective than only considering aspect terms. We also observe that the contribution of O-LSTM is less significant than that of bi-directionality on D_1 (+1.6% vs +2.0%). This is reasonable since using opinion words as adjective modifiers placed after the aspects is common in English.

	D_1	D_2
MIN without bi-directionality	75.59%	71.87%
MIN without S-LSTM	76.04%	72.55%
MIN without O-LSTM	75.97%	71.80%
MIN	**77.58%**	**73.44%**

Table 3: Ablation experiment results.

4 Conclusions

We propose Memory Interaction Network (MIN), a multi-task learning framework, to detect aspect terms from the online user reviews. Compared with previous studies, our MIN has following features:

- Co-occurrence pattern between aspects and opinions is captured via memory interactions, where the neural memory operations are designed to summarize task-level information and perform interactions.

- A novel LSTM unit with extended memories is developed for memory interactions.

References

Maryna Chernyshevich. 2014. Ihs r&d belarus: Cross-domain extraction of product features using crf. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 309–313.

Nguyen Viet Cuong, Nan Ye, Wee Sun Lee, and Hai Leong Chieu. 2014. Conditional random field with high-order dependencies for sequence labeling and segmentation. *Journal of Machine Learning Research*, 15(1):981–1009.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of AISTATS*, pages 249–256.

Minqing Hu and Bing Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of KDD*, pages 168–177.

Minqing Hu and Bing Liu. 2004b. Mining opinion features in customer reviews. In *Proceedings of AAAI*, pages 755–760.

Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of EMNLP*, pages 1035–1045.

Wei Jin, Hung Hay Ho, and Rohini K Srihari. 2009. A novel lexicalized hmm-based learning framework for web opinion mining. In *Proceedings of ICML*, pages 465–472.

Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. An empirical exploration of recurrent network architectures. In *Proceedings of ICML*, pages 2342–2350.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *Proceedings of ICLR*.

Bing Liu. 2012a. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Kang Liu, Heng Li Xu, Yang Liu, and Jun Zhao. 2013a. Opinion target extraction using partially-supervised word alignment model. In *Proceedings of IJCAI*, pages 2134–2140.

Kang Liu, Liheng Xu, and Jun Zhao. 2012b. Opinion target extraction using word-based translation model. In *Proceedings of EMNLP/CoNLL*, pages 1346–1356.

Kang Liu, Liheng Xu, and Jun Zhao. 2013b. Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In *Proceedings of ACL*, pages 1754–1763. Association for Computational Linguistics.

Kang Liu, Liheng Xu, and Jun Zhao. 2014. Extracting opinion targets and opinion words from online reviews with graph co-ranking. In *Proceedings of ACL*, pages 314–324. Association for Computational Linguistics.

Pengfei Liu, Shafiq Joty, and Helen Meng. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of EMNLP*, pages 1433–1443.

Asha S Manek, P Deepa Shenoy, M Chandra Mohan, and KR Venugopal. 2017. Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier. *World Wide Web Journal*, 20(2):135–154.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS*, pages 3111–3119.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of EMNLP*, pages 339–346.

Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.

Iñaki San Vicente, Xabier Saralegi, and Rodrigo Agerri. 2015. Elixa: A modular and flexible absa platform. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 748–752.

Sunita Sarawagi, William W Cohen, et al. 2004. Semi-markov conditional random fields for information extraction. In *Proceedings of NIPS*, pages 1185–1192.

Zhiqiang Toh and Jian Su. 2016. Nlangp at semeval-2016 task 5: Improving aspect based sentiment analysis using neural network features. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 282–288.

Zhiqiang Toh and Wenting Wang. 2014. Dlirec: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 235–240.

Wenya Wang, Sinno Jialin Pan, and Daniel Dahlmeier. 2017b. Multi-task coupled attentions for category-specific aspect and opinion terms co-extraction. *arXiv preprint arXiv:1702.01776*.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of EMNLP*, pages 616–626. Association for Computational Linguistics.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017a. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of AAAI*, pages 3316–3322.

Dionysios Xenos, Panagiotis Theodorakakos, John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos. 2016. Aueb-absa at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 312–317.

Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of IJCAI*, pages 2979–2985.

Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of CIKM*, pages 43–50.