

Aspect Category Detection in Product Reviews using Contextual Representation

Shiva Ramezani
College of Information and Computer
Sciences, UMass, Amherst
sramezani@cs.umass.edu

Razieh Rahimi
College of Information and Computer
Sciences, UMass, Amherst
rahimi@cs.umass.edu

James Allan
College of Information and Computer
Sciences, UMass, Amherst
allan@cs.umass.edu

ABSTRACT

Aspect category detection (ACD) is one of the challenging sub-tasks in aspect-based sentiment analysis. The goal of this task is to detect implicit or explicit aspect categories from the sentences of user-generated reviews. Since annotation over the aspects is time-consuming, the amount of labeled data is limited for supervised learning. In this paper, we study contextual representations of text segments in the reviews using the BERT model to better extract useful features from them, and train a supervised classifier with a small amount of labeled data for the ACD task. Experimental results obtained on Amazon reviews of six product domains show that our method is effective in some domains.

KEYWORDS

Aspect category detection, contextual representation, attention-based encoder

ACM Reference Format:

Shiva Ramezani, Razieh Rahimi, and James Allan. 2020. Aspect Category Detection in Product Reviews using Contextual Representation. In *Proceedings of ACM SIGIR Workshop on eCommerce (SIGIR eCom'20)*. ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

User-generated reviews in e-commerce websites like Amazon¹ have valuable information for both the users and the producers of products or services. A potential user can make an educated decision for purchasing a product by analyzing experiences of other users mentioned in the reviews. Such data can also help the producers to refine their products or services. However, it is impractical for a user or producer to read a huge amount of reviews and analyze them manually. Therefore, there is an emergent need for systems that can automatically process the huge amount of reviews and provide useful information about reviews in a suitable form. Opinion mining, sentiment analysis, and opinion summarization [2, 12, 19] for online reviews have attracted much attention to automatically analyze a large number of reviews. Aspect extraction is the first and foremost subtask in these problems which aims to extract entities or aspects of entities that people have expressed their opinions about. In general, there are two subtasks in aspect extraction: (1)

aspect term extraction (ATE) which aims to extract all aspect terms appearing in an opinionated text segment and (2) *aspect category detection* (ACD) which aims to identify the predefined categories of aspects discussed in a given text segment where aspects may not be mentioned explicitly. For example, given the sentence “The 32” screen is very suitable for your average living room or a bedroom”, the ATE should extract “32” screen” as an aspect term, and ACD should identify “size” as the aspect category.

Several supervised models have been proposed for the aspect extraction task. These models mainly lie in ATE task. Early works on supervised approaches commonly model aspect extraction as a sequence labeling problem and utilize graphical models like Hidden Markov Model [15] or Conditional Random Field [13, 20]. These classifiers are trained with rich engineered features based on linguistic or syntactic information from annotated data to predict a label for each token in a given text segment. With the development of deep learning techniques, different neural models are proposed to automatically learn features and reduce the feature engineering effort. However, training deep neural networks usually requires a large number of training data for each specific domain which is not the case in many real situations. Several cross-domain models have been proposed to address the problem of lack of training data in target domains [14, 17, 30], but differences in source and target domains made the design of cross-domain models challenging.

Unsupervised models have been proposed for aspect extraction task to avoid reliance on labeled data. These models also do not have the problems of cross-domain models. These models mainly lie in ACD task. Recent unsupervised neural models [2, 11, 23] are trained on large sets of unlabeled data. However, these models do not benefit from in-domain aspect-specific features which can be extracted from in-domain labeled data. According to the results reported by Yu et al. [33], a supervised model with a small amount of labelled in-domain data can outperform a cross-domain model. This means that domain specific features of aspect categories are very important for the aspect extraction task such that even a small amount of in-domain examples can boost the performance by a high margin.

Since preparing a large number of manually-labeled training data is expensive, only a small number of labeled data for the ACD task is available. Recently, a dataset has been released which contains a large number of unlabeled reviews along with a small number of labeled reviews (e.g. 50 reviews per domain) for the ACD task [2]. Using this dataset and motivated by the observation that even a small number of labeled in-domain examples are very useful for the ACD task [33], we study if supervised classifier models can be used to learn domain-specific representations of aspect categories based on a small number of labelled in-domain examples.

¹<https://www.amazon.com/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR eCom'20, July 30, 2020, Virtual Event, China

© 2020 Copyright held by the owner/author(s).

One important part of text classification models is input representation, where useful features can be extracted from input to detect the text label [18]. Current ACD models only consider a text segment of a review as the input of the classifier. However, we observed that some text segments do not have enough information to extract useful features from. Indeed, the text segment in a review may be dependent on the other parts of the review and thus does not have enough information by itself for detection of aspects. For example, consider this review from the TV domain: “There were a lot of menus and setups involved when we first got it. That was a little daunting.” In this example, there are two sentences that both of them are talking about the feature category “ease of use” in the TV domain. However, the second sentence itself cannot give this information and we need the first sentence to fully understand the second one. To tackle this issue, we propose an in-review contextual representation of text segments to better extract useful features from a given text segment. In our model, we feed the entire review as input and try to generate a representation of a text segment by attending to the entire review.

The rest of the paper is organized as follows. We provide related work on aspect extraction in Section 2. In Section 3, we define the problem and in section 4, we present our proposed model in details. Then we describe experimental settings in section 5 and empirically validate our hypotheses in section 6. Finally, we talk about the future work and conclude the paper in section 7.

2 RELATED WORK

Aspect extraction has gained more attention by emerging of the pioneer work of [12]. In this paper, authors hypothesize that noun and noun phrases of each sentence are most likely to be product features. At the first step, they use part-of-speech tagging techniques to find noun/noun phrases as candidate aspect of item in the review. Then, they use some predefined rules to find frequent features and use feature pruning techniques to remove redundant or uninteresting features. Following this pioneer work, many rule-based models has been designed using frequency co-occurrence or syntactic structure of the sentence [22, 26, 35]. These models need feature engineering to define useful rules and are heavily depend on quality of text parser that constructs the dependency tree. In addition to that, user generated reviews are not always precise enough to be parsed with a text parser and it cause inaccurate dependency tree.

Supervised models mainly consider this task as sequence labeling and try to assign label to each token of given text sentence. These works employ recurrent neural networks [21], use dependency-based embeddings as features in a Conditional Random Field [32], or combine a recursive neural network with CRFs to jointly model aspect and sentiment terms [31]. Considering the fact that opinion and aspect in a text segment are highly correlated, many multi task models have been proposed to use syntactic relation between them to simultaneously extract opinion and aspects [19, 28]. However, the gold data is not always available for this joint learning, and dependency parsers are not always accurate enough to extract the relation between opinion terms and aspect terms. Recently, [17] designed a multi-task cross domain model by proposing a multi-hop Dual Memory Interaction (DMI) mechanism to automatically

capture the latent relations among aspect and opinion words and get rid of need for linguistic resources.

To deal with the problem of training data, researchers tried to design cross domain models [8, 17, 29, 30] to transfer sentiment knowledge from source domain to a target domain that doesn't have any in-domain labelled training data. But some other problem in this setting makes cross domain models challenging for this task: 1) Different aspect spaces. Although some domains have common aspect categories like price and size, each domain has its own specific aspect categories. For example battery life is an aspect category for laptop but not a valid aspect for laptop cover. 2) Different term usage. different domains can have different terms for same aspect category. For example for aspect category “siz”, terms like XL and L are used in clothing domain while numbers are used in shoes domain. 3) Different meaning. Some words have different meaning in different domains. For example the word memory is related to the feature storage in Laptop domain while it has a different meaning in Mattress domain.

Unsupervised models in the other hand, doesn't have these problems. These approaches have been started by LDA based topic models [4, 6, 24, 34] that try to provide word distributions or rankings for each aspect category. Zhao et. al. [34] proposed MaxEnt-LDA to jointly extract aspect and opinion words. Chen et. al. [6] proposed to discover aspects by automatically learning prior knowledge from a large amount of online data. Wang et. al. [27] proposed a modified restricted Boltzmann machine (RBM), which jointly learn aspects and sentiment of text by using prior knowledge. Recently, unsupervised neural model attempted to learn aspect representatives from unlabeled corpus by reconstructing the input sentence. An unsupervised neural model named Aspect Based Auto Encoder (ABAE) [11] proposed an autoencoder model based on attention mechanism to train a latent representation that indicates probability distribution of input text segment over different aspect categories. The autoencoder part of this model attempts to reconstruct the input segment's encoding as a linear combination of aspect embeddings where aspect embeddings are learned by minimizing the segment reconstruction error. Then a weakly supervised extension of ABAE model named Multi-seed Aspect Extractor (MATE) [2] has been proposed that utilizes small amount of labelled data to extract seed words for each aspect category, and utilizes weighted sum of these seed words for each aspect category to create embedding of that aspect. This model then initializes aspect matrix of ABAE model with these values and fix them during training. Aspect Extraction with Sememe Attentions [23] is a hierarchical model similar to ABAE that in addition to word vectors and aspect vectors, this model also considers sense and sememe [5] vectors in computing the attention distribution.

3 PROBLEM DEFINITION

Aspect Category Detection is a classification task where a text segment in a review should be classified according to a subset of predefined aspect labels. For each product domain d , e.g., keyboards or vacuums, we have a corpus containing set of reviews $R_d = \{re_1, \dots, re_n\}$ where each review is split into text segments (s_1, s_2, \dots, s_k) . A text segment in the ACD task can be a sentence, phrase or Elementary Discourse Unit (EDU) [9] which corresponds

to a clause-level text segment. In our setting we are using EDU as text segment since in our dataset, aspect labels are available in EDU level [2]. The reason they are preferred is that EDU level segmentation have been shown to facilitate other related tasks like summarization, single document opinion extraction and document level sentiment analysis [1, 3, 16]. The ACD aims to classify a text segment $s = \{w_1, \dots, w_z\}$ into an aspect category $y \in \mathbb{A}_d = \{y_1, \dots, y_m\}$ where \mathbb{A}_d is set of predefined aspect categories in domain d . Each \mathbb{A}_d contains a general aspect which is assigned to segments that do not discuss any specific aspects.

4 METHODOLOGY

We hypothesize that the representation of each text segment is correlated with other text segments in a review. Considering this hypothesis, we propose a supervised model that can benefit from a better representation built by using the entire review. Figure 1 shows the architecture of our model. The network consists of two main components: 1) a representation layer containing a BERT component that tries to transform a given text segment to a rich vector representation, and 2) a multi-layer perceptron (MLP) classifier that takes the vector representation and returns probability distribution of aspect categories.

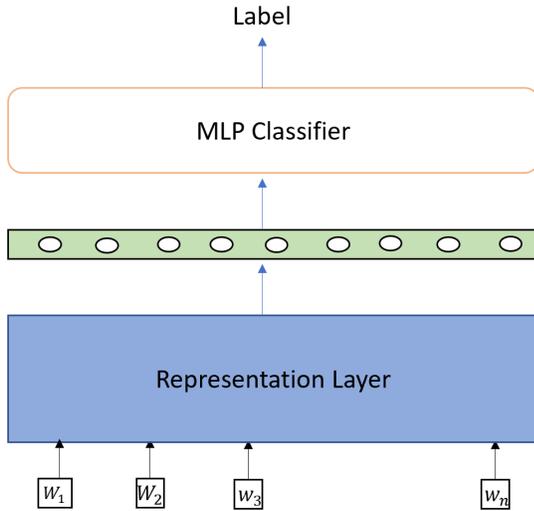


Figure 1: Architecture of our model

4.1 Representation Layer

The representation layer transforms raw text to a fixed-size vector representation encoding useful information of the input text. We use BERT, a strong feature extraction model for this layer. However, if the input text segment does not have enough information, we cannot extract useful information even with a good feature extraction model. Since EDU segments are clause level discourse and are roughly short, they may not have enough information by themselves. Thus, having the entire review as input can help to better represent a given text segment. To do so, we feed the entire review to the representation layer and update the representation of each

word in the target text segment by paying attention to the other parts of the review. Figure 2 shows the representation layer of the proposed model. The input text to this layer is built by adding a [CLS] token at the beginning and a [SEP] token at the end of each text segment in a review. To discriminate the target text segment from others, we also add specific tokens to the beginning and end of the target text segment in addition to [CLS] and [SEP] tokens. Then we consider the average of token representations in the target text segment as its representation. This manner, the target text segment will have a representation that is aware of other related parts in the review.

4.2 Classification Layer

The classifier aims to estimate a distribution of aspect probability $q(y|r_s)$, where r_s is a feature representation of a text segment provided by the representation layer and y is a vector with a size of number of aspect categories. We use a simple Softmax classifier on top of two layer feed forward neural network for this layer. The goal of this layer is to capture aspect related features of the text segment and represent it as a probability distribution over aspect categories.

5 EXPERIMENTAL SETTING

5.1 Dataset

We use the OpoSum dataset [2] for training and evaluating of our model. This dataset contains reviews from 6 product domains from Amazon: Laptop Bags, Bluetooth Headsets, Boots, Keyboards, Televisions, and Vacuums. Reviews are down sampled from the Amazon Product Dataset [10] and are segmented into Elementary Discourse Units (EDUs [9]). For each product domain, the OpoSum dataset contains 100 reviews from 10 different products with around 1000 EDU-level aspect annotations that are equally divided to the development and test sets (50 reviews and 500 EDUs in each set). We use the development set for training and the test set for evaluation.

For pre-processing the dataset, we follow the experimental settings of previous work [2, 11] to make our results comparable to the previously reported results. The dataset is pre-processed by lemmatization, removing punctuation symbols, and removing EDUs with less than two words. Multi-labelled sentences are also removed to avoid ambiguity. Table 1 shows the statistics of pre-processed data for each product domain.

5.2 Evaluation Metric

To evaluate and compare different models, we use macro-F1 score. We calculate the F1 score of each aspect category and then average them to calculate the macro-F1 score of a given domain. We also calculate micro-F1 score of baselines to compare them with previous state-of-the-art models since their result is reported in micro-F1 score.

5.3 Baselines

In order to show the effectiveness of our proposed representation, we compare it with multiple baselines for each product domain. For the baselines, the inputs are target text segments without the knowledge of other text segments in the review, while for our

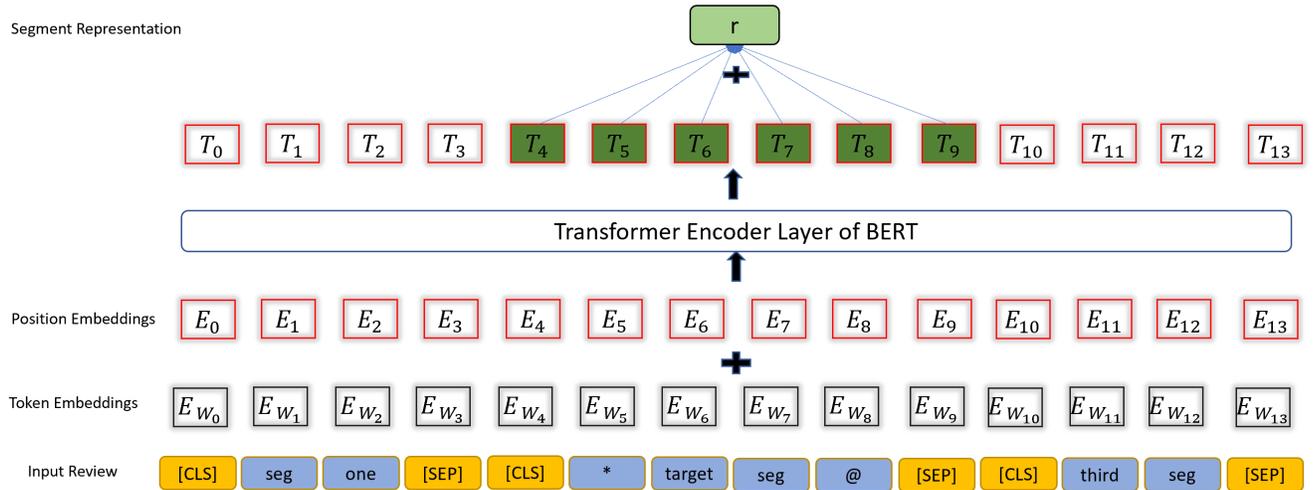


Figure 2: Representation layer

Table 1: The number of EDU segments in each set of the dataset.

Domain	Unlabelled dataset	Training set	Test set
Laptop Bags	0.586M	566	613
Bluetooth Headsets	1.424M	634	637
Boots	0.966M	507	578
Keyboards	0.605M	649	667
Televisions	1.424M	670	701
Vacuums	1.459M	687	687

proposed model the input is the entire review. We also compare our baselines with state-of-the-art unsupervised and weak supervised models to study if a supervised model trained only on a small number of labelled in-domain data can outperform unsupervised or weakly supervised models trained on large unlabeled in-domain data. The baseline methods are as follows:

5.3.1 ABAE_{init} [11]. This baseline is an unsupervised neural network approach that uses Neural Bag of Words (NBOW) as sentence representation and learns aspect embeddings in a reconstruction process, where an attention mechanism is used to filter non-aspect words. This model contains an aspect embedding matrix which will be trained on a large amount of unlabeled data.

5.3.2 MATE + MT [2]. This model is a weakly supervised auto-encoder extension of the ABAE model that initializes the aspect embedding matrix using seed words for each aspect category and fix them during training. These seed words are extracted from a small amount of in-domain labeled data.

5.3.3 Avg. This is a baseline neural model for text classification task that considers the average of embedding of words in a text segment as its representation. For word embeddings, we use embeddings trained on large unlabeled in-domain data.

5.3.4 Attention. In this model, we utilize embeddings of aspect categories to generate different aspect-dependent representations of the given text segment by attending to the embedding of each aspect category, and then use concatenation of these aspect-dependent representations as the representation of the text segment. Figure 3 shows the architecture of this model in detail. The first step in this model is generating the embedding matrix of aspect categories. To do so, we follow the approach proposed by Angelidis and Lapata (2018b) to obtain a ranked list of terms and their scores using a variant of a clarity scoring function [7] which measures the probability of observing word w in the subset of segments that discuss aspect a . The weighted sum of embeddings of the top ranked words (e.g. top 30 words) for each aspect category is considered as the embedding representation of the aspect category. In the next step, we compute different representations of a text segment by attending to each aspect category and provide aspect-dependent representations. For each aspect category, we calculate aspect-dependent representations of the text segment by computing the weighted sum of word embeddings in the text segment. The weights are cosine similarities between the embedding of the current word and word embeddings of the aspect. We then concatenate these aspect-dependent representations and feed them to a MLP to classify the input text segment.

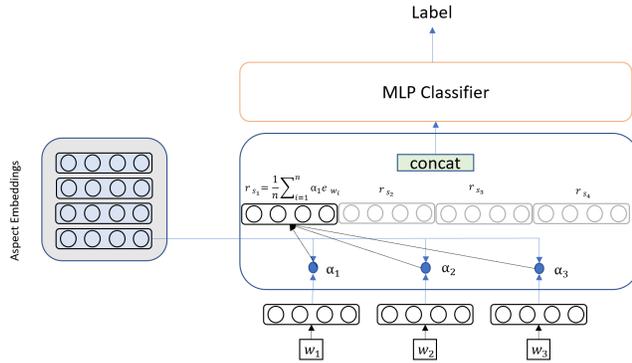


Figure 3: Architecture of attention model

5.3.5 *BERT – CLS*. In this baseline model, a text segment is fed to the base-uncased BERT model and the representation of the [CLS] token is considered as the representation of input. We then feed this representation to the classifier layer to classify the text segment. Indeed, this model is a base model in BERT model for sentence classification. We consider this baseline to study if other parts of the review that used in our model are useful to categorize a given text segment.

5.4 Model Configuration

For each product domain, we fine-tune the hyperparameters on 20% of training data and then train the model on the rest of training data for 300 epochs. Test data is used for evaluation. In the classifier, number of neurons of feed forward layer is 64. We use learning rate of 1e-4 and batch size of 16 for training the models. We use 300-dimensional word embeddings pre-trained on unlabeled data of each product domain using skip-gram [25], and fix it during training phase of the classifier model. Out of vocabulary words are replaced by the unknown (UNK) token. For aspect embedding matrix, used in MATE+MT and attention model, we use the top 30 seeds of each aspect category and calculate the weighted sum of seed words as the aspect embedding.

6 RESULTS

Table 2 shows the results of the state-of-the art unsupervised and weakly supervised models trained on large unlabeled data, supervised models trained on small labeled data, and our proposed model trained on small labeled data. Bold numbers are the best performance and underlined numbers are the second best performance. First of all we can observe that result of worst supervised model is better than performance of state-of-the-art unsupervised and weakly supervised models. This shows that supervised models trained only on a small amount of labeled data can perform much better, and in-domain aspect-specific features are very important in this task. By comparing the results of the Avg model and Attention model, we can realize that information about aspect categories can boost the performance by a high margin. The results in the table also show that the best baseline is the BERT-CLS model which shows feature extraction from text segments is an important part

in the text classification. BERT language model which is pretrained on large amount of data can be transferred to a new task by small amount of task-specific labeled data and can provide a better text representation. At the end, the result shows that our model is performing better than other models on average. The performance of our model is the second best performance in some domains. This shows that there are some text segments that their representation is dependent on the other parts of the review. However, we can see the performance decreases in Boots and Vacuums domains in comparison to the BERT-CLS model which might show that our proposed representation can also add noise in some cases.

7 CONCLUSION AND FUTURE WORK

Aspect category detection is a crucial sub-task in fine-grained sentiment analysis and related problems. Since providing a large labeled training data to train deep learning models for this task is expensive, very small number of product domains have labelled data for this task. Researchers proposed several cross domain and unsupervised models to tackle this problem, however considering the fact that each product domain has its own specific features and language, designing a cross-domain model is challenging for this task. On the other hand, unsupervised models cannot learn semantic features of the domain very well. One important part in text classification tasks is providing a rich and useful representation of the text. According to our observation, some text segments do not have enough information by themselves, and their meanings are highly correlated with other parts of their reviews. Considering this observation, we proposed a new contextual in-review representation in which a segment representation is generated by attending to other parts of the review and can be enriched by informative parts of the review. Experimental results show that our model can slightly improve the performance. In addition, we observed that not only representations of two nearby segments, but also their aspect categories are correlated. In the future, we would like to design a semi supervised model that can use large unlabeled data to learn language-specific features of each domain and use small in-domain labeled data to extract aspect-dependent semantic features of the domain. We are also interested in developing models for joint aspect-category prediction of nearby text segments in a review following our observations in this study.

ACKNOWLEDGEMENT

This work was supported in part by the Center for Intelligent Information Retrieval and in part by NSF grant #1813662. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

- [1] Stefanos Angelidis and Mirella Lapata. 2018. Multiple instance learning networks for fine-grained sentiment analysis. *Transactions of the Association for Computational Linguistics* 6 (2018), 17–31.
- [2] Stefanos Angelidis and Mirella Lapata. 2018. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 3675–3686.
- [3] Parminder Bhatia, Yangfeng Ji, and Jacob Eisenstein. 2015. Better document-level sentiment analysis from rst discourse parsing. *arXiv preprint arXiv:1509.01599* (2015).

Table 2: Comparison of different models. The numbers in the upper part are micro-f1 and the numbers in the lower part are macro-f1.

Model	Laptop Bags	Bluetooth Headsets	Boots	Keyboards	Televisions	Vacuums	AVG
ABAE_init	41.6	48.5	41.2	41.3	45.7	40.6	43.2
MATE+MT	48.6	54.5	46.4	45.3	51.8	47.7	49.1
Avg	<u>64.6</u>	62.8	<u>56.6</u>	61.8	62.1	58.7	61.1
Attention	62.3	<u>64.2</u>	56.4	<u>63.0</u>	<u>66.6</u>	<u>60.6</u>	<u>62.2</u>
BERT-CLS	66.7	71.3	67.3	66.6	71.3	67.8	68.5
Avg	55.5	47.0	48.2	43.3	48.3	47.4	48.5
Attention	52.6	<u>50.2</u>	46.6	50.3	52.4	49.4	50.2
BERT-CLS	<u>56.1</u>	54.4	55.6	<u>58.6</u>	<u>54.2</u>	54.0	<u>55.5</u>
NewBERT (Ours)	57.7	54.4	<u>55.3</u>	60.6	57.8	<u>53.2</u>	56.5

- [4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [5] Leonard Bloomfield. 1926. A set of postulates for the science of language. *Language* 2, 3 (1926), 153–164.
- [6] Zhiyuan Chen, Arjun Mukherjee, and Bing Liu. 2014. Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 347–358.
- [7] Steve Cronen-Townsend, Yun Zhou, and W Bruce Croft. 2002. Predicting query performance. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. 299–306.
- [8] Ying Ding, Jianfei Yu, and Jing Jiang. 2017. Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [9] Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 60–68.
- [10] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.
- [11] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 388–397.
- [12] Mingqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168–177.
- [13] Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1035–1045.
- [14] Zhuoren Jiang, Zheng Gao, Jinjong Lan, Hongxia Yang, Yao Lu, and Xiaozhong Liu. 2020. Task-Oriented Genetic Activation for Large-Scale Complex Heterogeneous Graph Embedding. In *Proceedings of The Web Conference 2020*. 1581–1591.
- [15] Wei Jin and Hung Hay Ho. 2009. A novel lexicalized HMM-based learning framework for web opinion mining. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 465–472.
- [16] Junyi Jessy Li, Kapil Thadani, and Amanda Stent. 2016. The role of discourse units in near-extractive summarization. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 137–147.
- [17] Zheng Li, Xin Li, Ying Wei, Lidong Bing, Yu Zhang, and Qiang Yang. 2019. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. *arXiv preprint arXiv:1910.14192* (2019).
- [18] Hong Liang, Xiao Sun, Yunlei Sun, and Yuan Gao. 2017. Text feature extraction based on deep learning: a review. *EURASIP journal on wireless communications and networking* 2017, 1 (2017), 1–12.
- [19] Bing Liu. 2012. *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- [20] Kang Liu, Heng Li Xu, Yang Liu, and Jun Zhao. 2013. Opinion target extraction using partially-supervised word alignment model. In *Twenty-Third International Joint Conference on Artificial Intelligence*.
- [21] Pengfei Liu, Shafiq Joty, and Helen Meng. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1433–1443.
- [22] Huaishao Luo, Tianrui Li, Bing Liu, Bin Wang, and Herwig Unger. 2019. Improving aspect term extraction with bidirectional dependency tree representation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27, 7 (2019), 1201–1212.
- [23] Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019. Unsupervised neural aspect extraction with sememes. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 5123–5129.
- [24] Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers-volume 1*. Association for Computational Linguistics, 339–348.
- [25] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
- [26] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics* 37, 1 (2011), 9–27.
- [27] Linlin Wang, Kang Liu, Zhu Cao, Jun Zhao, and Gerard De Melo. 2015. Sentiment-aspect extraction based on restricted boltzmann machines. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 616–625.
- [28] Wenya Wang and Sinno Jialin Pan. 2018. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2171–2181.
- [29] Wenya Wang and Sinno Jialin Pan. 2019. Transferable Interactive Memory Network for Domain Adaptation in Fine-Grained Opinion Extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7192–7199.
- [30] Wenya Wang and Sinno Jialin Pan. 2020. Syntactically Meaningful and Transferable Recursive Neural Networks for Aspect and Opinion Extraction. *Computational Linguistics* 45, 4 (2020), 705–736.
- [31] Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. *arXiv preprint arXiv:1603.06679* (2016).
- [32] Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. *arXiv preprint arXiv:1605.07843* (2016).
- [33] Jianfei Yu, Jing Jiang, and Rui Xia. 2018. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27, 1 (2018), 168–177.
- [34] Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. 2010. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 56–65.
- [35] Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. 43–50.