

Aspect Extraction from Reviews using Convolutional Neural Networks and Embeddings*

Peiman Barnaghi, Georgios Kontonatsios, Nik Bessis, and Yannis Korkontzelos

Edge Hill University, Liverpool, United Kingdom
{barnaghp, Georgios.Kontonatsios, Nik.Bessis,
Yannis.Korkontzelos}@edgehill.ac.uk

Abstract. Aspect-based sentiment analysis is an important natural language processing task that allows to extract the sentiment expressed in a review for parts or aspects of a product or service. Extracting all aspects for a domain without manual rules or annotations is a major challenge. In this paper, we propose a method for this task based on a Convolutional Neural Network (CNN) and two embedding layers. We address shortcomings of state-of-the-art methods by combining a CNN with an embedding layer trained on the general domain and one trained the specific domain of the reviews to be analysed. We evaluated our system on two SemEval datasets and compared against state-of-the-art methods that have been evaluated on the same data. The results indicate that our system performs comparably well or better than more complex systems that may take longer to train.

Keywords: Aspect-based sentiment analysis · Aspect extraction · Convolutional Neural Networks · Deep learning · NLP

1 Introduction

Currently immense volumes of text-based reviews are available, in a great variety of diverse domains. Consumers can share their experience on services and products. Natural Language Processing (NLP) methods can be used to extract meaningful information from this data. Quantifying sentiment expressed for various aspects of a product or service can help producers and consumers to monitor, assess and make decisions. Significant volume of research has focused on Extensive research has focussed on analysing online reviews for a variety of topics or products, e.g. movies, restaurants, mobile applications and software projects.

Aspect-based sentiment analysis is a variation of sentiment analysis that considers different aspects of the object of a text-based review and classifies the comments for each aspect as positive, negative or neutral. For example, in the comment “the food is great but expensive and service is slow” three

* The final authenticated version is available online at https://doi.org/10.1007/978-3-030-23281-8_37.

aspects are mentioned, i.e. quality, price and service. Lately, neural networks have been shown to perform very well in sentiment analysis when combined with word embeddings. Word embeddings are vector representations of textual vocabularies, useful for finding similar words. Each word is mapped to a vector that captures its context in different sentences. Embeddings retain syntactic and semantic similarities and relations among words. Most neural network based systems for text analysis have employed Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

Given the success of CNN [7] on aspect extraction, we propose a CNN-based system for extracting aspects from reviews and we combine it with different embedding layers. CNN models are less complex than RNN models. Experiments in the literature show that CNN models train faster than RNN models and tuning their hyper-parameters is simpler. State-of-the-art aspect extraction systems combine neural networks with word embeddings. The contribution of this paper is the combination of a CNN with two word embeddings concurrently: one trained on the general domain and one trained on the domain of the reviews. The model performs comparably or better than methods that integrate more complex architectures, such as RNNs. It outperforms a CNN-based model that uses either general domain embeddings or domain-specific ones, only.

2 Related Work

Aspect based sentiment analysis identifies sentiment expressed for each aspect of a product or service. It was introduced for summarising customer reviews and was addressed by a rule-based model [5]. Since then, a variety of systems have been proposed and several competition tasks have been organised in the SemEval (Semantic Evaluation) series. Task 4A in SemEval 2014 focussed on the extraction of aspects in reviews. Liu [9] discussed four approaches for aspect identification: frequent terms, opinion and target relations, supervised classification and topic modelling algorithms. Conditional Random Fields (CRF) have been employed to consider long term dependencies when extracting aspects [6], and performed better than other supervised models for feature extraction [23]. Toh and Wang [22] used a tagging model with linguistic features that consider resources, such as WordNet, for aspect extraction and polarity classification. Brun et al. [1] combined word features, parsing and a sentiment lexicon to train a Support Vector Machine (SVM) for aspect-based sentiment classification.

In SemEval 2016, the best performing system used CRFs for sequential labelling, i.e. aspect extraction, and a single-layer feed-forward neural network for classification [18]. A CNN-based aspect extraction method tagged each word in subjective text [16]. The CNN tags each word as aspect or not, in different layers. The model performed better than state-of-the-art approaches. CNNs, as non-linear models, fit the data better than linear models, such CRFs.

In summary, latest research uses deep learning to improve aspect extraction and aspect-based sentiment classification performance, as it has been very successful in supervised and unsupervised settings. Shortcomings of these models

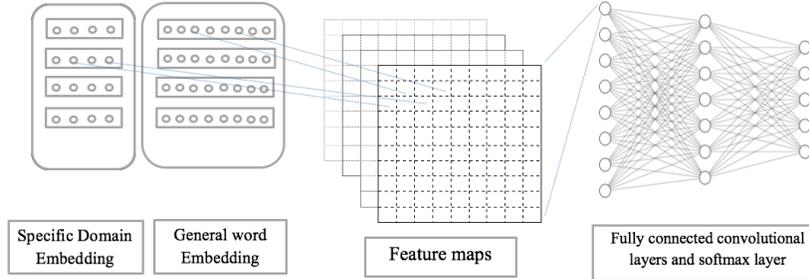


Fig. 1. Proposed CNN model with domain and general word embedding.

for extracting hidden aspects concern long distance dependencies and domain-specific expressions. In this paper, we address the latter shortcoming, by combining general and domain-specific embeddings.

3 Method

To extract aspects of reviews, we use a CNN [7] with fully connected layers combined with two independent embedding layers, as shown in figure 1. The input is a sentence of any size that mentions zero or more aspects. Each word of the sentence is looked up in both embeddings and the two resulting vectors are concatenated together. The general embedding is a pre-trained Global Vectors for word representation (GloVe) model [13], trained on 840 billion tokens. The vocabulary size is 2.2 million vectors of dimension 300. We selected this model due to the size of the data it was trained on and its popularity for aspect extraction. The domain-specific embeddings are trained either on *Yelp* [11], a restaurant review dataset, or on *Amazon reviews* for laptops [4]. Reviews in both datasets come labeled with aspect terms.

The joined vector is the input of the multi-layer CNN. Each layer uses a convolutional filter of fixed window width and kernel size. For example, with kernel size $k = 5$, two words on the left and right of the current one are kept. Each filter represents each word and its nearby words. An activation function is used to choose the maximum value of each features node, and a dropout is applied to prevent overfitting during training. Finally, a Softmax layer is applied to a fully connected layer to select the sequence with the highest position weight and assign a label to each word, accordingly.

4 Experiments

Datasets: For evaluation, we use two benchmark SemEval datasets: the laptop review dataset in SemEval-2014 Task 4 and the restaurant review dataset in SemEval-2016 Task 5. Table 1 shows statistics of the two datasets.

Tuning Network Hyper-Parameters: 100 randomly-selected data instances were excluded from the training data, to be used as validation data for tuning. A

popular technique for avoiding underfitting, is to evaluate the model for various layer sizes, parameters and learning rates. If validation accuracy is higher than training accuracy then the model is underfitting, or otherwise it is overfitting. Each CNN layer consists of 256 filters of kernel size 3. Processing continues to the end of the vector and feature weights are computed. We used common parameter values for the dropout and learning rates: 0.5 and 10^{-4} , respectively.

Evaluation: Following common practice, we use F-score (F1), the geometric mean of precision and recall. For evaluation, we used the SemEval script. We compare our proposed model with all methods, for which results on the SemEval 2014 and 2016 datasets have been made available [8].

IHS_RD [2] was the best system on laptop reviews in task 4 of SemEval 2014. It used conditional random fields for cross-domain feature extraction [15].

NLANGP [18] was the best system in aspect extraction on restaurant reviews in task 5 of SemEval 2016. It is also based on neural networks [14].

AUEB [20] a CRF-based method for sequence labeling that uses hand-crafted features and embeddings. It was ranked among the top systems in SemEval 2016.

CRF [12] a CRF-based method using general embeddings and basic features.

Semi-Markov CRF (Semi-CRF) [17] uses features in Cuong et al. [3].

DLIREC [22] a CRF-based classifier that uses semantic features and clustering on unlabeled data. It was ranked second in both SemEval tasks.

WDEmb [21] a CRF that uses linear and dependency context information.

RNCRF [19] a CRF and RNN combination for aspect extraction.

LSTM [10] uses an RNN and general pre-trained word embeddings.

MIN [8] uses two LSTM models for aspect extraction and one for sentiment classification.

We used two baselines: (1) the proposed multilayer CNN model with general word embeddings, only; and (2) the proposed model with in-domain word embeddings, only. The performance gap between the baselines and the proposed system shall highlight the impact of combining the two word embeddings.

5 Discussion

The experimental results in Table 2 show that the proposed method, (PM-G&D), performs better than state-of-the-art systems for aspect extraction. It outperforms the two baselines that use the same model with either general (PM-G) or in-domain embeddings (PM-D), only. This result stresses the contribution of the combination of the two embeddings, since all other settings are kept the same.

The performance difference between the two datasets indicates that the in-domain embedding is more effective in laptop reviews. This is probably because

Table 1. Dataset statistics

Dataset	Training instances	Aspect terms	Test instances	Aspect terms
Restaurant	2000	1743	676	622
Laptop	3,045	2358	800	654

Table 2. F-score results on the restaurant (R) and laptop (L) review dataset. PM, G and D stand for our proposed model, general and in-domain embedding, respectively.

Data	IHS_RD	NLANGP	AUEB	CRF	Semi-CRF	DLIREC	WDEmb	RNCRF	LSTM	MIN	PM-G	PM-D	PM-G&D
R	-	72.34	70.44	69.56	66.35	-	-	69.74	71.26	73.44	69.80	68.23	73.81
L	74.55	-	-	74.01	68.75	73.78	75.16	77.26	75.25	77.58	73.19	73.37	78.26

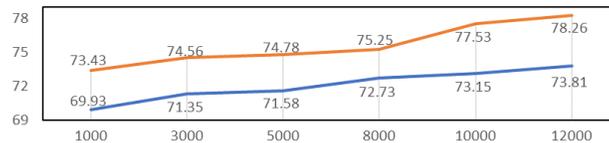
this domain has more keywords than restaurant reviews, which mainly contains general words possibly available in general embeddings. Our model performs better than CRF-based models that specialise on label dependencies, since both datasets mostly contain single-word aspects.

As all systems in Table 2 are using general embeddings, the results show that combining large general and small in-domain word embeddings can improve aspect extraction performance. Figure 2 shows the effect of increasing the training data size on performance. Results improve mildly as the size of the training data for the in-domain embeddings increases.

Embeddings can be combined to improve the aspect extraction performance in other domains and probably other languages. Although recent work shows that RNN models are state-of-the-art, we have achieved comparable results using a much simpler model, which is faster to train, by adding an extra learning layer.

6 Conclusion & Future Work

We proposed a new model for aspect extraction from text-based reviews. It uses a convolutional neural network and two word embedding layers: a general and a domain-specific one, trained on data of the specific domain of the reviews. Evaluation on two benchmark SemEval datasets, containing restaurant and laptop reviews, shows that the model performs comparably or better than more complex neural network architectures, that take longer to train. In the future, we plan to comparatively evaluate more aspect extraction methods, deep learning architectures and embedding types on diverse domains.

**Fig. 2.** Performance for increasing size of the domain-specific embedding layer for restaurants (blue line) and laptops (red line)

Acknowledgment

This research work is part of the TYPHON Project, which has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 780251.

References

1. Brun, C., Popa, D.N., Roux, C.: XRCE: hybrid classification for aspect-based sentiment analysis. In: Proc. SemEval. pp. 838–842 (2014)
2. Chernyshevich, M.: Ihs r&d belarus: Cross-domain extraction of product features using crf. In: Proc. SemEval. pp. 309–313 (2014)
3. Cuong, N.V., Ye, N., Lee, W.S., et al.: Conditional random field with high-order dependencies for sequence labeling and segmentation. *The Journal of Machine Learning Research* **15**(1), 981–1009 (2014)
4. He, R., McAuley, J.: Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In: Proc. WWW. pp. 507–517 (2016)
5. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proc. SIGKDD. pp. 168–177. ACM (2004)
6. Jakob, N., Gurevych, I.: Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In: Proc. EMLNP. pp. 1035–1045 (2010)
7. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint:1408.5882 (2014)
8. Li, X., Lam, W.: Deep multi-task learning for aspect term extraction with memory interaction. In: Proc. EMNLP. pp. 2886–2892 (2017)
9. Liu, B.: Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* **5**(1), 1–167 (2012)
10. Liu, P., Joty, S., Meng, H.: Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Proc. EMNLP. pp. 1433–1443 (2015)
11. Mikolov, T., Chen, K., Corrado, G., et al.: Efficient estimation of word representations in vector space. arXiv preprint:1301.3781 (2013)
12. Okazaki, N.: Crfsuite: a fast implementation of conditional random fields (crfs) (2007), www.chokkan.org/software/crfsuite
13. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proc. EMNLP. pp. 1532–1543 (2014)
14. Pontiki, M., Galanis, D., Papageorgiou, H., et al.: Semeval-2016 task 5: Aspect based sentiment analysis. In: Proc. SemEval. pp. 19–30 (2016)
15. Pontiki, M., Galanis, D., Pavlopoulos, J., et al.: Semeval-2014 task 4: Aspect based sentiment analysis. In: Proc. SemEval. pp. 27–35 (2014)
16. Poria, S., Cambria, E., Gelbukh, A.: Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems* **108**, 42–49 (2016)
17. Sarawagi, S., Cohen, W.: Semi-markov conditional random fields for information extraction. In: Proc. NIPS. pp. 1185–1192 (2005)
18. Toh, Z., Su, J.: Nlangp at semeval-2016 task 5: Improving aspect based sentiment analysis using neural network features. In: Proc. SemEval. pp. 282–288 (2016)
19. Wang, W., Pan, S.J., Dahlmeier, D., et al.: Recursive neural conditional random fields for aspect-based sentiment analysis. arXiv preprint:1603.06679 (2016)
20. Xenos, D., Theodorakakos, P., Pavlopoulos, J., et al.: Aueb-absa at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In: Proc. SemEval. pp. 312–317 (2016)

21. Yin, Y., Wei, F., Dong, L., et al.: Unsupervised word and dependency path embeddings for aspect term extraction. arXiv preprint:1605.07843 (2016)
22. Zhiqiang, T., Wenting, W.: Dlirec: Aspect term extraction and term polarity classification system. Proc. the 8th International Workshop on Semantic Evaluation (SemEval 2014) pp. 235–240 (2014)
23. Zhuang, L., Jing, F., Zhu, X.Y.: Movie review mining and summarization. In: Proc. CIKM. pp. 43–50. ACM (2006)