Aspect-Level Sentiment Classification

Shamnas TV, Vishal Peter
Department of Computer Science and Automation
Indian Institute of Science, Bangalore
shamnast@iisc.ac.in, vishalpeter@iisc.ac.in

Abstract

In this project, we independently implement two recent models for Aspect Level Sentiment Classification(ALSC) which has displayed state-of-the-art performance. We also propose novel models by incorporating ideas from these two baseline models. Experimental results show that our novel model is able to improve upon the baselines to give better performance on multiple standard datasets. We make the data and source code publicly available at https://github.com/androidKunjappan/asc.

1 Introduction

Aspect level sentiment classification (ALSC) is a fundamental task in sentiment analysis, where the objective is to infer the sentiment polarity towards specific aspects. Compared to document-level or sentence-level sentiment classification, the main challenge here is to distinguish the different sentiments toward each aspect term when there are multiple aspects in a sentence.

From the recent methods published for the ALSC task, we have considered two methods which has displayed state-of-the-art performance.

1) *TNet-ATT(+AS) : Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis* (Tang et al. (2019)) : Proposes a semantic-based progressive approach for improving upon existing neural ASLC models, which automatically mines useful attention supervision information from a training corpus to refine attention mechanisms.

2) *RepWalk: Replicate, Walk, and Stop on Syntax An Effective Neural Network Model for Aspect-Level Sentiment Classification* (Yaowei et al. (2020)) : Proposes a syntactic-based approach that constructs a syntax tree for sentences and performs a replicated random walk on the tree to focus on the informative contextual words effectively.

We propose a new architecture incorporating the progressive attention learning scheme and a Context-Preserving Transformation layer(as proposed in TNet (Li et al. (2018))) into the Repwalk model to improve the weights learned for nodes(words) in the syntax graph.

In summary, our contributions in this project are as follows:

- Re-implement the model TNet-ATT(+AS) independently.
- Re-implement the model RepWalk independently.
- Propose a new architecture which incorporates into the Repwalk model
  - Progressive Attention Learning Scheme
  - Context-Preserving Transformation layer

Submitted towards credit requirement for course E0 334

2 Related Work

Aspect level sentiment classification is a branch of sentiment analysis in which the goal is to identify the sentiment polarity of an aspect target within a given context sentence. Several strategies have recently been suggested to determine sentiment polarities of aspect words. These works can be divided into three categories: the rule-based methods, the semantic-based methods, and the syntactic-based methods.

Rule-based methods are based on handcrafted features derived by imposing set of rules. Kiritchenko et al. (2014) proposes a rule-based method that extracts aspect term by imposing a set of rules. It finds the polarity of aspect term using an SVM classifier. It also finds the category of aspect terms using SVM. Another rule-based method Jiang et al. (2011) proposes to improve target-dependent sentiment classification of tweets by using both target-dependent and context-aware approaches. Here target-dependent approach refers to incorporating syntactic features generated using words syntactically connected with the given target in the tweet to decide whether or not the sentiment is about the given target. In both methods, features are derived from handcrafted rules and may suffer from error propagation due to human ingenuity.

Semantic-based methods automatically learn representations for sentences by taking advantage of the high expressive power of neural network models. With the rapid development of deep learning, dominant neural network based ASC models have evolved such as Wang et al. (2016), Tang et al. (2016), Tang et al. (2015), Ma et al. (2017), Chen et al. (2017), Li et al. (2018), Wang et al. (2018), which can automatically capture the representations of sentences and sentiment polarities of aspects with better accuracy. Usually, these models are equipped with attention mechanisms that help identify the relevance of different context words towards each aspect. Wang, Lu (2018) propose a segmentation attention-based LSTM model with a linear-chain conditional random field (CRF) layer, which tries to simulate the humans' process of sentiment inference.

Syntactic-based methods integrate syntactic context information to learn target words' relationship with context words. Chen, Manning (2014) proposes a neural network-based model which incorporates the dependency tree of sentences. Dong et al. (2014) propagate the sentiment of words to the target using recursive neural networks through sentences' syntactic structure. He et al. (2018) proposes a syntax-based model assuming that informative contextual words are close to the aspect term in the dependency tree. Here it uses the attention mechanism to get the final representation. Hence, it defines an attention window to focus on such words.

3 Proposed Approach

We have built upon the baseline models TNet-ATT(+AS) and RepWalk to create a novel model which has improved the overall classification performance on multiple benchmark datasets. In this section, we explain the baseline models and our novel contributions.

3.1 Baseline Models

3.1.1 TNet-ATT(+AS)

TNet-ATT(+AS) proposes a semantic-based progressive approach for improving upon existing neural ASLC models, which automatically mines useful attention supervision information from a training corpus to refine attention mechanisms. Typically, semantic-based models that perform ALSC have some attention mechanism that tries to learn how a context word is important to the given aspect. It leads to a major drawback in existing attention-based models; specifically, there is a higher chance that the attention mechanism gives more attention (max attention weight) to frequently occurring words while training and less attention to less frequent words with sentiment polarities. As a result, the model may overfit towards frequent words and lead to insufficient learning of other context words, especially low-frequency ones with sentiment polarities. In the paper, the authors discuss a novel approach that tries to resolve this issue. In the initial stage of training, the algorithm finds the most influential word of each training sentence(if there exists an influencing word). It masks it for further iterations to discover other low-frequency words with sentiment polarities. Then it assigns weights for each masked word in each sentence. Later during the final stage of training, the model will try to match the word's weight with weights obtained during the initial stage of training.

3.1.2 Repwalk

Repwalk proposes a model that can sufficiently leverage the syntactic structure of sentences. The authors of RepWalk hypothesize that the contextual features of the informative words(towards an aspect) on the subtree of a sentence's syntax tree may significantly identify the sentiment expressed on the specific aspect in the sentence. To find such informative context words in a sentence, the authors describe a neural network model that performs a replicated-random walk on the sentence's syntax tree.

The model first creates a dependency tree for every sentence in the data corpus by utilizing recently advanced parsing technology (Chen, Manning (2014)). Then, for each aspect term in a sentence, Repwalk transforms the sentence's syntax tree to one rooted in the aspect term. The model then generates representation for every node(words in the sentence) as well as edges in the tree. Let $q_j \in \mathbb{R}^{d_q}$ denote the representation for j^{th} node in the sentence(generated using glove vectors passed through a BiGRU layer). For any given edge $e = (u, r, v)$ in the syntax tree, which connects node u to node v with a distinct edge label $r \in \mathbb{R}$, where the embedding for r is $\theta_r \in \mathbb{R}^{d_r}$, we define a function $p(e)$ mapping each edge e to a probability value given as:

$$p(e) = \sigma(\begin{bmatrix} q_u \\ q_v \end{bmatrix} W_p \theta_r + b_p) \tag{1}$$

where σ denotes the sigmoid activation function, and $W_p \in R^{2d_q X d_r}$ and $b_p \in \mathbb{R}$ are learned parameters.

Then, for each word(say word w_i) in the sentence, a weight is calculated by taking the product of the edges' scores in the path from that word's node to the tree's root(aspect term) as

$$\alpha_i = \prod_{e \in \mathcal{E}_{w_i}} p(e) \tag{2}$$

where $\mathcal{E}_{w_i}$ denotes the unique path from word w_i to aspect node(root) in syntax tree.

Finally, sentence representation is calculated as the weighted average of word representations.

$$o = \sum_{i=1}^{n} \alpha_i q_i \tag{3}$$

The sentence representation thus obtained is passed through a fully connected layer to output labels for classification.

3.2 Proposed Novel Model

3.2.1 Context-Preserving Transformation

Inspired by TNet, we integrate a Context-Preserving Transformation layer that enables learning target-specific word representations in a deep neural architecture. CPT mainly consists of three components. (1) Target-Specific Transformation(TST) to consolidate word representation and target representation. TST contains a Bi-LSTM to generate target word representations with semantic context; it is then used to calculate word representations via an attention mechanism. (2) Context-Preserving Mechanism(CPM) to take advantage of the context information. (3) position-aware scaling, used to encode the positional relevance between a word and a target; by doing this, the words close to the target will be highlighted, and those far away will be downgraded. The CPT can capture the semantic similarities between the context words and aspect words and actual positional relevance. We calculate the representation of sentence from CPT layer using attention-based weghted average of context word representations and add it to the representation derived from RepWalk model to get final representation. We call this model Repwalk(CPT).

3.2.2 Pogressive Attention Supervision

In Repwalk, the final sentence representation is obtained by taking a weighted average of word representations(q_i) in the sentence, using the syntax tree structure to get this word weight α_i. These weights are analogous to attention weights in models such as TNet-ATT(+AS), and hence the

Method	Rest14		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
TNet-ATT(+AS)(reimplementation)	81.48	73.16	76.21	71.76	79.19	78.31
TNet-ATT(+AS)(paper)	81.53	72.9	77.62	73.84	78.61	77.72
Repwalk(reimplementation)	83.60	76.60	76.50	72.30	73.30	70.80
Repwalk(paper)	83.80	76.90	78.20	74.30	74.40	72.60

Table 1: Experimental results of Re-Implemented models on various datasets

Method	Rest14		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
TNet-ATT(+AS)(reimplementation)	81.48	73.16	76.21	71.76	**79.19**	**78.31**
Repwalk(reimplementation)	**83.60**	**76.60**	76.50	72.30	73.30	70.80
Repwalk(AttS)	82.95	74.87	**79.15**	**76.25**	73.98	72.07
Repwalk(CPT)	83.48	75.65	77.12	73	74.28	72.93
Repwalk(AttS+CPT)	83.30	75.45	78.21	73.82	72.98	71.33

Table 2: Comparison of our new models proposed with the re-implemented models.

progressive attention learning scheme can be incorporated into the Repwalk to regularize these node weights. Applying the progressive attention learning scheme requires the word weights in a sentence to be normalized to add up to 1. Hence We add a normalization step given as

$$\alpha_i' = \frac{\alpha_i}{\sum_{j=1}^{n} \alpha_j} \tag{4}$$

and use the new weight α_i' for sentence representation and learn it progressively. We call this model Repwalk(AttS) to denote the **Att**ention **S**upervision scheme incorporated into Repwalk.

3.2.3 Combined model

We also propose a model combining both the Context-Preserving Transformation 3.2.1 and Progressive Attention Learning3.2.2, named as Repwalk(AttS+CPT).

4 Experimental Results

4.1 Reimplementation

The models TNet-ATT(+AS) and Repwalk were re-implemented and the performance was compared against the values reported in the papers. We observe that accuracy and F1 scores observed were similar to the numbers reported in the papers. The results are detailed in Table 1.

4.2 Novel Models

Comparative results of our new models proposed with the re-implemented models on various datasets are given in Table 2.

Repwalk(AttS) We observe that incorporating the progressive attention scheme gives significant performance improvements for Laptop dataset(+2.65% accuracy) and minor improvements for Twitter dataset (+0.68% accuracy) over the vanilla Repwalk model. However a slight performance reduction was seen for the Restaurant dataset.

Repwalk(CPT) Incorporating Context Preserving transformation too gives performance improvements for Laptop(+0.62%) and Twitter(+0.98%) datasets. However, the improvement was not observed for Restaurant dataset.

Repwalk(CPT) The combined model also improves over vanilla Repwalk model for Laptop dataset but not for the other two datasets. However we believe this could be improved with extensive hyperparameter tuning on this model.

References

Chen Danqi, Manning Christopher D. A fast and accurate dependency parser using neural networks // Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014. 740–750.

Chen Peng, Sun Zhongqian, Bing Lidong, Yang Wei. Recurrent attention network on memory for aspect sentiment analysis // Proceedings of the 2017 conference on empirical methods in natural language processing. 2017. 452–461.

Dong Li, Wei Furu, Tan Chuanqi, Tang Duyu, Zhou Ming, Xu Ke. Adaptive recursive neural network for target-dependent twitter sentiment classification // Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers). 2014. 49–54.

He Ruidan, Lee Wee Sun, Ng Hwee Tou, Dahlmeier Daniel. Effective attention modeling for aspect-level sentiment classification // Proceedings of the 27th international conference on computational linguistics. 2018. 1121–1131.

Jiang Long, Yu Mo, Zhou Ming, Liu Xiaohua, Zhao Tiejun. Target-dependent Twitter Sentiment Classification // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, Oregon, USA: Association for Computational Linguistics, VI 2011. 151–160.

Kiritchenko Svetlana, Zhu Xiaodan, Cherry Colin, Mohammad Saif. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews // Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: Association for Computational Linguistics, VIII 2014. 437–442.

Li Xin, Bing Lidong, Lam Wai, Shi Bei. Transformation Networks for Target-Oriented Sentiment Classification // ACL. 2018. 946–956.

Ma Dehong, Li Sujian, Zhang Xiaodong, Wang Houfeng. Interactive attention networks for aspect-level sentiment classification // arXiv preprint arXiv:1709.00893. 2017.

Tang Duyu, Qin Bing, Feng Xiaocheng, Liu Ting. Effective LSTMs for target-dependent sentiment classification // arXiv preprint arXiv:1512.01100. 2015.

Tang Duyu, Qin Bing, Liu Ting. Aspect level sentiment classification with deep memory network // arXiv preprint arXiv:1605.08900. 2016.

Tang Jialong, Lu Ziyao, Su Jinsong, Ge Yubin, Song Linfeng, Sun Le, Luo Jiebo. Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. VII 2019.

Wang Bailin, Lu Wei. Learning latent opinions for aspect-level sentiment classification // Proceedings of the AAAI Conference on Artificial Intelligence. 32, 1. 2018.

Wang Shuai, Mazumder Sahisnu, Liu Bing, Zhou Mianwei, Chang Yi. Target-sensitive memory networks for aspect sentiment classification // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018. 957–967.

Wang Yequan, Huang Minlie, Zhu Xiaoyan, Zhao Li. Attention-based LSTM for aspect-level sentiment classification // Proceedings of the 2016 conference on empirical methods in natural language processing. 2016. 606–615.

Yaowei Zheng, Richong Zhang, Samuel Mensah, Yongyi Mao. Replicate, Walk, and Stop on Syntax: an Effective Neural Network Model for Aspect-Level Sentiment Classification // AAAI. 2020.