Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks

Binxuan Huang
binxuanh@cs.cmu.edu

Kathleen M. Carley
kathleen.carley@cs.cmu.edu

Carnegie Mellon
institute for SOFTWARE RESEARCH

Outlines

- Introduction
- Previous work
- Method
- Experiments
- Conclusion

CASOS

Introduction

- Sentence-level sentiment classification
  - Identify the overall sentiment polarity of a sentence as positive, negative, or neutral.
  - Many times a document is mixed with different aspects and opinions.
    - "The food in this restaurant is excellent, but the service is not good."
  - Jiang et al. examined that 40% of sentiment classification errors come from not considering aspects[1].

Introduction

- Aspect level sentiment classification
  - Identify the sentiment polarity for each aspect in one document.
  - "The food in this restaurant is excellent, but the service is not good."

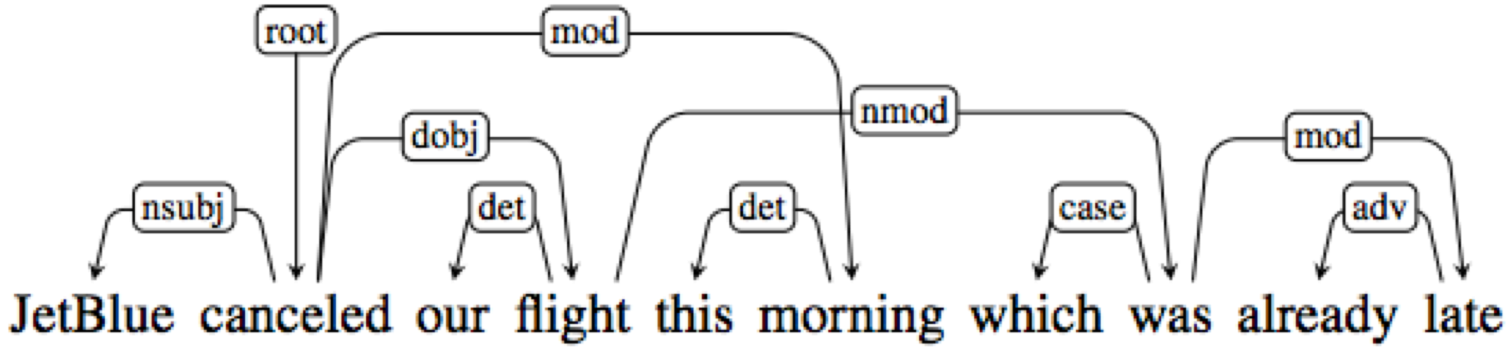
Introduction

- Formal problem definition of Aspect level sentiment classification
  - Given a sentence $s = [w_1, w_2, \dots, w_i \dots, w_j, \dots w_n]$ and an aspect target $t = [w_i, \dots, w_j]$, the goal is to classify the sentiment as positive, negative, or neutral.

Sentences	Aspects	Sentiment
The food in this restaurant is excellent, but the service is not good.	food	+1
The food in this restaurant is excellent, but the service is not good.	service	-1
Boot time is super fast, around anywhere from 35 seconds to 1 minutes.	Boot time	+1

Introduction

- Motivation
  - Using syntactic structure of the sentence is helpful for identifying sentiment features directly related to the aspect target.
    - eg. “The food, though served with bad service, is actually great”
  - It is also helpful to resolve potential ambiguity in a word sequence.
    - eg. “Good food bad service”, “a bad sushi lover”



Picture credit: https://web.stanford.edu/~jurafsky/slp3/13.pdf

Previous work

- Most previous work treat a sentence as a word sequence and use LSTM or CNN to extract aspect related features.
  - **Feature-based SVM** utilizes n-gram features, parse features and lexicon features for aspect-level sentiment classification. [1]
  - **TD-LSTM** uses two LSTM networks to model the preceding and following contexts surrounding the aspect term. The last hidden states of these two LSTM networks are concatenated for predicting the sentiment polarity. [2]
  - **AT-LSTM** first models the sentence via a LSTM model. Then it combines the hidden states from the LSTM with the aspect term embedding to generate the attention vector. [3]
  - **AOA-LSTM** introduces an attention-over-attention (AOA) based network to model aspects and sentences in a joint way and explicitly capture the interaction between aspects and context sentences. [4]
  - **PG-CNN** is a CNN based model where aspect features are used as gates to control the feature extraction on sentences. [5]

Method

- Model components
  - Text representation
  - Graph attention network
  - Target-dependent graph attention network
  - Classification

Method

- Text representation

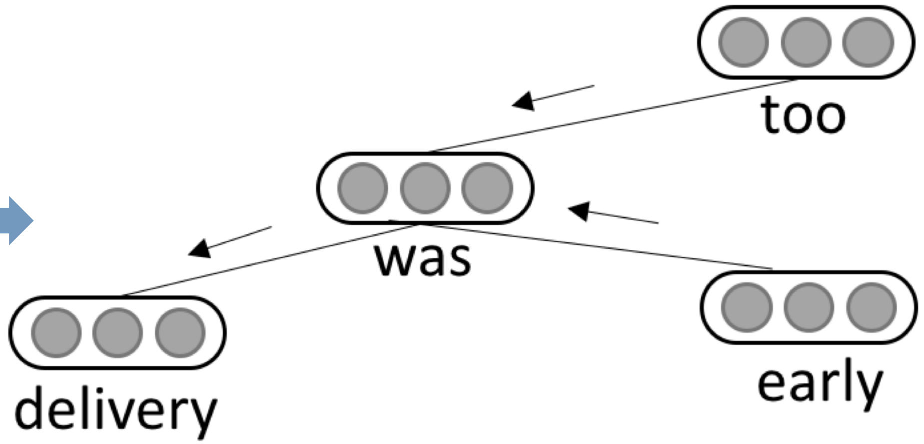
Sentence:

delivery was early too

Aspect:

delivery

Dependency parser



Each node i in the dependency graph A is associated with a GloVe word embedding vector or a contextual BERT representation, denoted as x_i

Method

- Graph attention network (GAT)
  - GAT is a variant of graph neural networks, which propagate features and learn node representations on a graph.
  - At each layer, GAT updates one node's representation by aggregating its neighborhood's representations using multihead attention.

$$H_{l+1} = GAT(H_l, A)$$

$$h_{l+1}^{i} = multihead\left(\left\{h_l^j; j \in nei[i]\right\}\right)$$
$$= ||_{k=1}^{K} \sigma\left(\sum_{j\in nei[i]} \alpha_{lk}^{ij} W_{lk} h_l^j\right)$$

$$\alpha_{lk}^{ij} = \frac{\exp\left(LeakyReLU\left(a_{lk}^T\left[W_{lk}h_l^i || W_{lk}h_l^j\right]\right)\right)}{\sum_{u\in nei[i]} \exp\left(LeakyReLU\left(a_{lk}^T\left[W_{lk}h_l^i || W_{lk}h_l^u\right]\right)\right)}$$

Method

- Target-dependent graph attention network (TD-GAT)
  - In a vanilla GAT network, the aspect information is not explicitly modeled.
  - We use an LSTM network to model the dependency for the aspect terms across layers.

$$\widehat{H_{l+1}} = GAT(H_l, A)$$

$$H_{l+1}, C_{l+1} = LSTM\left(\widehat{H_{l+1}}, (H_l, C_l)\right)$$

$$H_0, C_0 = LSTM\left(XW + b, (0,0)\right)$$

Method

- Classification
  - The probability for each sentiment class is computed by a softmax function after a linear projection layer.

$$P(y=c)=\frac{\exp(Wh_L^t+b)_c}{\sum_{i\in C}\exp(Wh_L^t+b)_i}$$

  - We minimize the cross entropy loss with L_2 regularization to train our model

$$loss=-\sum_{c\in C} I(y=c)\cdot\log\big(P(y=c)\big)+\lambda\big||\Theta|\big|^2$$

Experiments

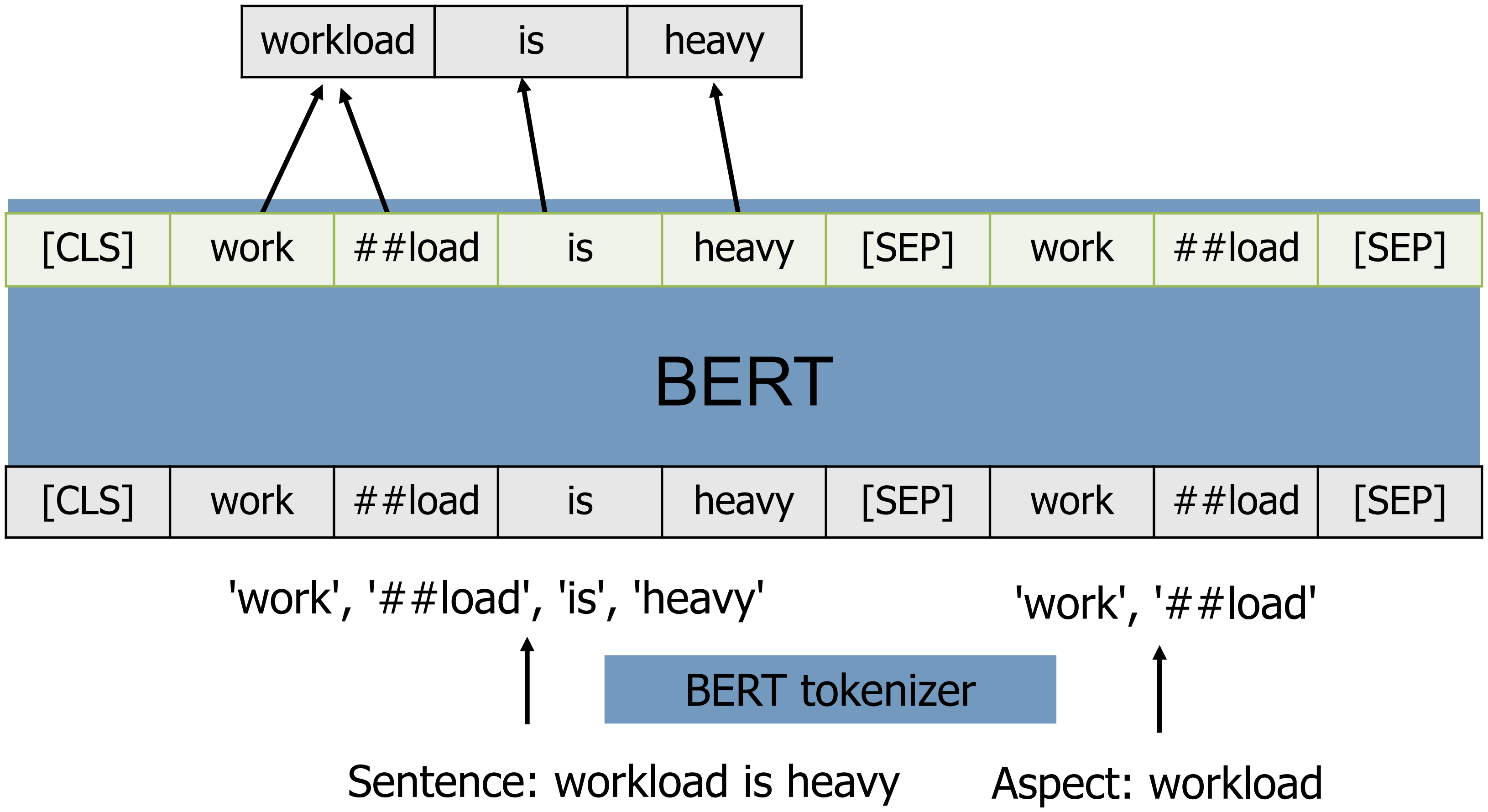
- Datasets

Dataset	Positive	Neutral	Negative
Laptop-Train	767	373	673
Laptop-Dev	220	87	193
Laptop-Test	341	169	128
Restaurant-Train	1886	531	685
Restaurant-Dev	278	102	120
Restaurant-Test	728	196	196

Table 1: Statistics of the datasets.

Experiments

- Using BERT representation



Experiments

- Baseline comparisons
  - We compare our method with various baselines.
  - We report the performance of our model with different number of layers.

	Laptop	Restaurant
Feature+SVM	70.5	80.2
TD-LSTM	68.1	75.6
AT-LSTM	68.9	76.2
MemNet	72.4	80.3
IAN	72.1	78.6
PG-CNN	69.1	78.9
AOA-LSTM	72.6	79.7
TD-GAT-GloVe (3)	73.7	81.1
TD-GAT-GloVe (4)	**74.0**	80.6
TD-GAT-GloVe (5)	73.4	**81.2**
BERT-AVG	76.5	78.7
BERT-CLS	77.1	81.2
TD-GAT-BERT (3)	79.3	82.9
TD-GAT-BERT (4)	79.8	**83.0**
TD-GAT-BERT (5)	**80.1**	82.8

Table 2: Comparison results of different methods on laptop and restaurant datasets. Numbers in parentheses indicate number of layers in our model.

Experiments

- Effects of target information
  - To examine the effects of explicitly modeling target information, we remove the LSTM unit in our model and compare it with TD-GAT
  - As shown in the table, explicitly capturing aspect target information consistently improves the performance of the TD-GAT-GloVe over the GAT-GloVe model.

Dataset	Laptop			Restaurant		
layer	3	4	5	3	4	5
GAT-GloVe	73.0	72.1	72.4	79.6	80.0	79.7
TD-GAT-GloVe	73.7	74.0	73.4	81.1	80.6	81.2
GAT-BERT	78.1	78.5	78.5	82.6	82.2	82.3
TG-GAT-BERT	79.3	79.8	80.1	82.9	83.0	82.8

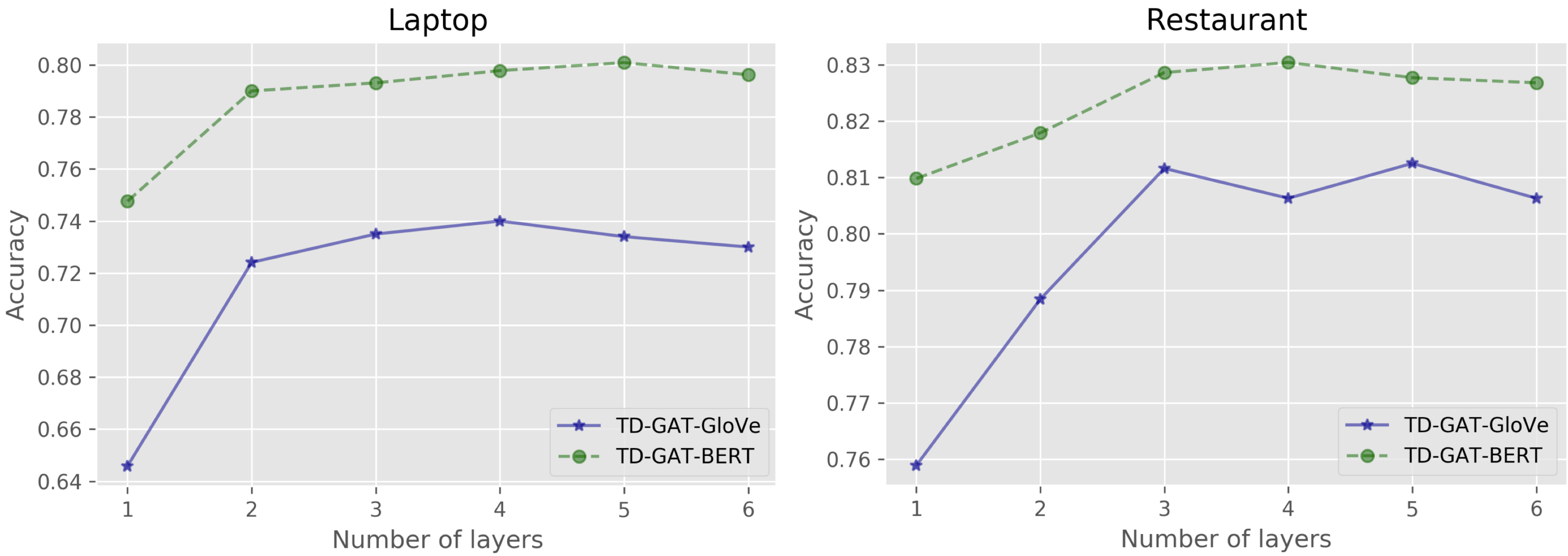
Table 3: An ablation study shows the effect of explicit target information.

Carnegie Mellon

isr institute for SOFTWARE RESEARCH

Experiments

- Effects of model depth



Experiments

- Model size
  - Using the same dimension of hidden states, our TD-GAT-GloVe has a lower model size compared to these LSTM-based methods.
  - When we switch from GloVe embeddings to BERT representations, the training time for a three-layer TD-GAT model on the restaurant dataset only increases from 1.12 seconds/epoch to 1.15 seconds/epoch.

Models	Model size ($\times 10^6$)
TD-LSTM	1.45
MemNet (3)	**0.36**
IAN	2.17
AOA-LSTM	2.89
TD-GAT-GloVe (3)	1.00
TD-GAT-GloVe (4)	1.09
TD-GAT-GloVe (5)	1.18
BERT-CLS	335.14
TD-GAT-BERT (3)	**1.30**
TD-GAT-BERT (4)	1.39
TD-GAT-BERT (5)	1.49

Table 4: The model size (number of parameters) of our model as well as baselines.

Conclusion

- In this paper, we propose a novel target-dependent graph attention neural network for aspect level sentiment classification.
- Using GloVe embeddings, our approach TD-GAT-GloVe outperforms various baseline models.
- After switching to BERT representations, we show that TD-GAT-BERT achieves much better performance.
- It is lightweight and requires fewer computational resources and less training time than fine-tuning the original BERT model.

Future Direction

- Future work could consider using an attention mechanism to focus on important words in the aspect.
- Since this work only uses the dependency graph and ignores various types of relations in the graph, we plan to incorporate dependency relation types into our model and take part-of-speech tagging into consideration as well in the future.
- We would also like to combine such a graph-based model with a sequence-based model to avoid potential noise from dependency parsing errors.

- [1] Kiritchenko, Svetlana, et al. "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews." *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014.
- [2] Tang, Duyu, et al. "Effective LSTMs for target-dependent sentiment classification." *arXiv preprint arXiv:1512.01100*(2015).
- [3] Wang, Yequan, Minlie Huang, and Li Zhao. "Attention-based LSTM for aspect-level sentiment classification." *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016.
- [4] Huang, Binxuan, Yanglan Ou, and Kathleen M. Carley. "Aspect level sentiment classification with attention-over-attention neural networks." *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, Cham, 2018.
- [5] Huang, Binxuan, and Kathleen Carley. "Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification." *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.