用于细粒度情感分类的交互注意力网络

Interactive Attention Networks for Aspect-Level Sentiment Classification

Dehong Ma[1], **Sujian Li**[1,2], **Xiaodong Zhang**[1], **Houfeng Wang**[1,2]
[1]MOE Key Lab of Computational Linguistics, Peking University, Beijing, 100871, China
[2]Collaborative Innovation Center for Language Ability, Xuzhou, Jiangsu, 221009, China
{madehong, lisujian, zxdcs, wanghf}@pku.edu.cn

Abstract

Aspect-level sentiment classification aims at identifying the sentiment polarity of specific target in its context. Previous approaches have realized the importance of targets in sentiment classification and developed various methods with the goal of precisely modeling thier contexts via generating target-specific representations. However, these studies always ignore the separate modeling of targets. In this paper, we argue that both targets and contexts deserve special treatment and need to be learned their own representations via interactive learning. Then, we propose the interactive attention networks (IAN) to interactively learn attentions in the contexts and targets, and generate the representations for targets and contexts separately. With this design, the IAN model can well represent a target and its collocative context, which is helpful to sentiment classification. Experimental results on SemEval 2014 Datasets demonstrate the effectiveness of our model.

作者认为目标和上下文都应该独立对待，需要通过交互学习来得到各自的表示

1 Introduction

Aspect-level sentiment classification is a fine-grained task in sentiment analysis which aims to identify the sentiment polarity of targets in their context [Pang and Lee, 2008; Liu, 2012]. For example, Given the mentioned targets: *staff*, *pizza* and *beef cubes*, and their context sentence *"a group of friendly staff, the pizza is not bad, but the beef cubes are not worth the money!"*, the sentiment polarity for the three targets, *staff*, *pizza* and *beef cubes*, are positive, neutral and negative respectively.

Aspect-level sentiment classification is a fundamental task in natural language processing and catches many researchers' attention. Traditional approaches mainly focus on designing a set of features such as bag-of-words, sentiment lexicon to train a classifier (e.g., SVM) for aspect-level sentiment classification [Jiang *et al.*, 2011; Perez-Rosas *et al.*, 2012]. However, feature engineering is labor intensive and almost reaches its performance bottleneck. With the development of deep learning techniques, some researchers have designed effective neural networks to automatically generate useful low-dimensional representations from targets and their contexts and obtain a promising result on the aspect-level sentiment classification task [Dong *et al.*, 2014; Vo and Zhang, 2015; Wang *et al.*, 2016; Tang *et al.*, 2016b].

传统特征工程已经遇到瓶颈了

利用深度神经网络学习得到相应的低维目标和上下文表示

As Jiang et al. [2011] point out that 40% of sentiment classification errors are caused by not considering targets in sentiment classification, recent work tends to especially strengthen the effect of targets when modeling the contexts. Dong et al. [2014] propose an adaptive recursive neural network (RNN) to propagate the sentiments from context words to specific targets based on syntactic relations on tweet data. Vo and Zhang [2015] separate the whole context into three components, i.e., target, its left context and right context, and then use sentiment lexicon and neural pooling functions to generate the target-dependent features. Tang et al. [2016a] divide the contexts into left part with target and right part with target and use two long short-term memory (LSTM) models to model the two parts respectively. Then the composed target-specific representations from both parts are used for sentiment classification. Wang et al. [2016] design aspect embeddings for targets and concatenate them with word representations to generate the final representations using LSTM networks and attention mechanism.

1.通过一种句法信息将其建立上下文和目标之间的关联

2.通过左右上下文、情感词典和神经池化

3.同理，分左半部分和右半部分，用LSTM来建模

4.训练一个目标嵌入矩阵

The studies above have realized the importance of targets and developed various methods with the goal of precisely modeling contexts via generating target-specific representations. However, they all ignore the separate modeling of targets, especially with the aid of contexts. In our opinion, only the coordination of targets and their contexts can really enhance the performance of sentiment classification. Let us take "The picture quality is clear-cut but the battery life is too short" as an example. When "short" is collocated with "battery life", the sentiment tends to be negative. But when "short" is used with "spoon" in the context "Short fat noodle spoon, relatively deep some curva", the sentiment can be neural. Then, the next problem is how to simultaneously model targets and contexts precisely. First, target and context can determine representations of each other. For example, when we see the target "picture quality", context word "clear-cut" is naturally associated with the target. And it is vice versa - "picture quality" is first connected with "clear-cut". In such cases, we argue that targets and contexts can be modeled separately but learned from their interaction. Second, our common sense is that the context is composed of many words. In fact, targets are also not limited to only one word. No matter

1.目标和上下文的对应关系

搭配

2.同时精准建立目标和上下文

理所应然

首先作者认为，可以分别建模但是要同时学习他们的互相影响

然后作者认为上下文和目标中均需使用注意力权重，以便更好地抓取信息

targets or contexts, different words may have different contributions to the final representation. For example, it is easy to know that "picture" plays a more important role in the representation of the target "picture quality" which is described by "clear-cut". Thus, we first propose that both targets and contexts should be computed their attention weights to capture their important information respectively.

Based on the two points analyzed above, we propose an interactive attention network (IAN) model which is based on long-short term memory networks (LSTM) and attention mechanism. IAN utilizes the attention mechanism associated with a target to get important information from the context and compute context representation for sentiment classification. Further, IAN makes use of the interactive information from context to supervise the modeling of the target which is helpful to judging sentiment. Finally, with both target representation and context representation concatenated, IAN predicts the sentiment polarity for the target within its context. Experiments on SemEval 2014 demonstrate that IAN can precisely model both targets and contexts, and achieve the state-of-the-art performance.

1.利用与目标相关联的注意力机制生成更加好的上下文表示

2.使用交互式的学习

1.1 Interactive Attention Networks

The IAN model is composed of two parts which model the target and context interactively. With word embeddings as input, we employ LSTM networks to obtain hidden states of words on the word level for a target and its context respectively. We use the average value of the target's hidden states and the context's hidden states to supervise the generation of attention vectors with which the attention mechanism is adopted to capture the important information in the context and target. With this design, the target and context can influence the generation of their representations interactively. Finally, target representation and context representation are concatenated as final representation which is fed to a softmax function for aspect-level sentiment classification.

词嵌入作为输入
LSTM得到抽象特征

平均值进行连接权重生成

最终将目标和上下文进行连接，然后输入到softmax层

2 Model

In this section, we first introduce the architecture of interactive attention networks (IAN) model for aspect-level sentiment classification. Next, we show the training details of IAN. The overall architecture of IAN model is shown in Figure 1.

Specifically, let us first formalize the notation. We suppose that a context consists of n words $[w_c^1, w_c^2, ..., w_c^n]$ and a target has m words $[w_t^1, w_t^2, ..., w_t^m]$. w denotes a specific word. To represent a word, we embed each word into a low-dimensional real-value vector, called word embedding [Bengio *et al.*, 2003]. Then, we can get $w^k \in R^d$ from $M^{v \times d}$, where k is the word index in the context or target, d means the embedding dimension and v gives the vocabulary size. Word embeddings can be regarded as parameters of neural networks or pre-trained from proper corpus via language model [Collobert and Weston, 2008; Mikolov *et al.*, 2010; Huang *et al.*, 2012; Pennington *et al.*, 2014]. In our model, we choose the latter strategy.

此处输入的词向量既可以是当做参数进行训练，
也可以是预先训练好的词向量

Next, we use the LSTM networks to learn the hidden word semantics, since words in a sentence have strong dependence

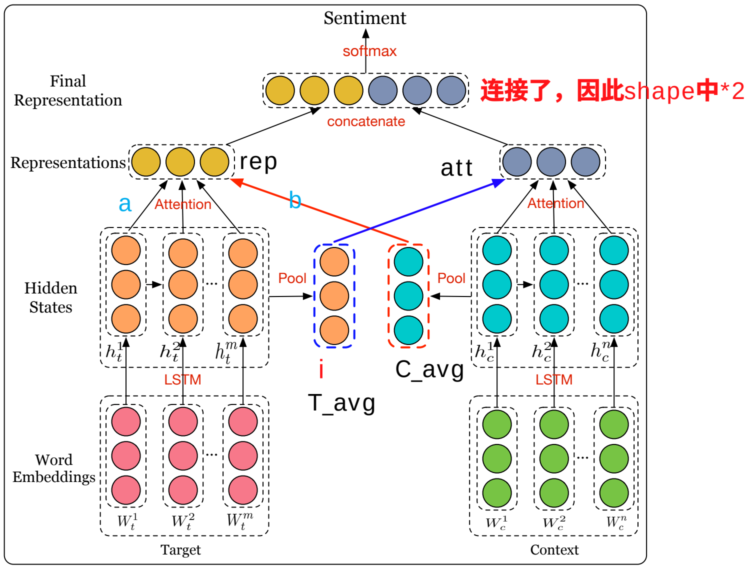


Figure 1: The overall architecture of IAN.

on each other, and LSTM is good at learning long-term dependencies and can avoid gradient vanishing and expansion problems. Formally, given the input word embedding w^k, previous cell state c^{k-1} and previous hidden state h^{k-1}, the current cell state c^k and current hidden state h^k in the LSTM networks are updated as:

$$\mathbf{i}^k = \sigma(W_{\mathbf{i}}^w \cdot w^k + W_{\mathbf{i}}^h \cdot h^{k-1} + b_{\mathbf{i}}) \tag{1}$$

$$\mathbf{f}^k = \sigma(W_{\mathbf{f}}^w \cdot w^k + W_{\mathbf{f}}^h \cdot h^{k-1} + b_{\mathbf{f}}) \tag{2}$$

$$\mathbf{o}^k = \sigma(W_{\mathbf{o}}^w \cdot w^k + W_{\mathbf{o}}^h \cdot h^{k-1} + b_{\mathbf{o}}) \tag{3}$$

$$\hat{c}^k = tanh(W_c^w \cdot w^k + W_c^h \cdot h^{k-1} + b_{\mathbf{c}}) \tag{4}$$

$$c^k = \mathbf{f}^k \odot c^{k-1} + \mathbf{i}^k \odot \hat{c}^k \tag{5}$$

$$h^k = \mathbf{o}^k \odot tanh(c^k) \tag{6}$$

where **i**, **f** and **o** are input gate, forget gate and output gate respectively, which serve to model the interactions between memory cells and their environments. σ means a sigmoid function. W and b denote weight matrices and biases respectively. The symbol $\cdot$ stands for matrix multiplication, and $\odot$ is element-wise multiplication. Then, we get the hidden states $[h_c^1, h_c^2, ..., h_c^n]$ as the final word representations for context. To better model target's meaning, we also use LSTM networks to obtain the target's hidden states $[h_t^1, h_t^2, ..., h_t^m]$.

Then, we can get the initial representations of context and target (i.e., c_{avg} and t_{avg}) by averaging the hidden states.

所有词求平均值

$$c_{avg} = \sum_{i=1}^{n} h_c^i / n \tag{7}$$

$$t_{avg} = \sum_{i=1}^{m} h_t^i / m \tag{8}$$

是取平均后的

With the initial representations of context and target as input, we adopt the attention mechanism to select important information contributing to judging sentiment polarity. As Section 1 stated, we consider the influence on the context from the target and the influence on the target from context, which can provide more clues to pay attention to those related sentiment features.

We take a pair of context and target to describe the attention process, as shown in Figure 1. With context word representations $[h_c^1, h_c^2, ..., h_c^n]$, the attention mechanism generates the attention vector α_i using target representation t_{avg} for context by:

$$\alpha_i = \frac{exp(\gamma(h_c^i, t_{avg}))}{\sum_{j=1}^{n} exp(\gamma(h_c^j, t_{avg}))} \tag{9}$$

where γ is a score function that calculates the importance of h_c^i in the context. The score function γ is defined as:

通过这条式子建立起
上下文和目标词的关系

$$\boxed{\gamma(h_c^i, t_{avg}) = tanh(h_c^i \cdot W_a \cdot t_{avg}{}^T + b_a)} \tag{10}$$

where W_a and b_a are weight matrix and bias respectively, $tanh$ is a non-linear function and $t_{avg}{}^T$ is the transpose of the t_{avg}.

Similarly, for the target, we calculate its attention vector β_i using context representation c_{avg} by:

不变量

$$\beta_i = \frac{exp(\gamma(h_t^i, c_{avg}))}{\sum_{j=1}^{m} exp(\gamma(h_t^j, c_{avg}))} \tag{11}$$

where γ is the same as in Eq. 10.

After computing the word attention weights, we can get context and target representations c_r and t_r based on the attention vectors α_i and β_i by:

$$\boxed{\begin{aligned} c_r &= \sum_{i=1}^{n} \alpha_i h_c^i & (12)\\ t_r &= \sum_{i=1}^{m} \beta_i h_t^i & (13) \end{aligned}}$$

Finally, the target representation t_r and context representation c_r are concatenated as a vector $\mathbf{d}$ for a classifier. Here, we use a non-linear layer to project $\mathbf{d}$ into the space of the targeted C classes. That is,

$$\boxed{x = tanh(W_l \cdot \mathbf{d} + b_l)} \tag{14}$$

where W_l and b_l are the weight matrix and bias respectively. The probability of labeling document with sentiment polarity $i (i \in [1, C])$ is computed by:

$$\boxed{y_i = \frac{exp(x_i)}{\sum_{i=1}^{C} exp(x_i)}} \tag{15}$$

The label with the highest probability is set as the final result.

2.1 Model Training

In IAN, we need to optimize all the parameters notated as Θ which are from LSTM networks: $[W_{\mathbf{i}}^{w}, W_{\mathbf{f}}^{w}, W_{\mathbf{o}}^{w}, W_{\mathbf{c}}^{w}, W_{\mathbf{i}}^{h}, W_{\mathbf{f}}^{h}, W_{\mathbf{o}}^{h}, W_{\mathbf{c}}^{h}, b_{\mathbf{i}}, b_{\mathbf{f}}, b_{\mathbf{o}}, b_{\mathbf{c}}]$, the attention layers: $[W_a, b_a]$, the softmax layer: $[W_l, b_l]$ and the word embeddings. Cross entropy with L_2 regularization is used as the loss function, which is defined as:

yi是分类计算的概率

$$J = -\sum_{i=1}^{C} g_i \log(y_i) + \lambda_r(\sum_{\theta \in \Theta} \theta^2) \tag{16}$$

参数的平方和

where $g_i \in R^C$ denotes the ground truth, represented by one-hot vector; $y_i \in R^C$ is the estimated probability for each class, computed as in Eq. (15). λ_r is the coefficient for L_2 regularization.

We use the backpropagation method to compute the gradients and update all the parameters Θ by:

$$\Theta = \Theta - \lambda_l \frac{\partial J(\Theta)}{\partial \Theta} \tag{17}$$

这是上下文句子的得分，目标达到得分需要换

where λ_l is the learning rate.

In order to avoid overfitting, we use dropout strategy to randomly omit half of the feature detectors on each training case. After learning Θ, we test the instance by feeding the target with its contexts into the IAN model, and the label with the highest probability stands for the predicted sentiment polarity of the target.

3 Experiments

3.1 Experiment Preparation

Dataset

We conduct experiments on SemEval 2014 Task 4[1] to validate the effectiveness of our model. The SemEval 2014 dataset is composed of reviews in two categories: *Restaurant* and *Laptop*. The reviews are labeled with three sentiment polarities: *positive*, *neutral* and *negative*.

Evaluation Metric

To evaluate the performance of aspect-level sentiment classification. we adopt the *Accuracy* metric, which is defined as:

$$Acc = \frac{T}{N} \tag{18}$$

where T is the number of correctly predicted samples, N is the total number of samples. *Accuracy* measures the percentage of correct predicted samples in all samples. Generally, a well performed system has a higher accuracy.

Table 1 shows the training and test instance numbers in each category.

3：1的黄金比例

Dataset	Positive		Neural		Negative		
	Train	Test	Train	Test	Train	Test	
Restaurant	2164	728	637	196	807	196	4728
Laptop	994	341	464	169	870	128	2966

Table 1: Statistics of SemEval 2014 Dataset.

[1] The detail introduction of this task can be seen at: http://alt.qcri.org/semeval2014/task4/

Dataset	Restaurant	Laptop
Majority	0.535	0.650
LSTM	0.743	0.665
TD-LSTM	0.756	0.681
AE-LSTM	0.762	0.689
ATAE-LSTM	0.772	0.687
IAN	**0.786**	**0.721**

Table 2: Comparison with baselines. Accuracy on 3-class prediction about SemEval 2014 Task 4 which includes restaurants and laptops. Best performances are in **bold**.

Hyperparameters Setting

In our experiments, all word embeddings from context and target are initialized by GloVe[2] [Pennington *et al.*, 2014], and all out-of-vocabulary words are initialized by sampling from the uniform distribution $U(-0.1, 0.1)$. All weight matrices are given their initial values by sampling from uniform distribution $U(-0.1, 0.1)$, and all biases are set to zeros. The dimensions of word embeddings, attention vectors and LSTM hidden states are set to 300 as in [Wang *et al.*, 2016]. To train the parameters of IAN, we employ the Momentum [Qian, 1999], which adds a fraction γ of the update vector in the prior step to the current update vector. The coefficient of L_2 normalization in the objective function is set to 10^{-5}, and the dropout rate is set to 0.5.

1.上下文和目标向量都是通过GloVe初始化

2.词典以外的词都是通过U分布采样得到

3.权重矩阵W的初始化都是通过U分布得到

4.偏置B初始化设置为0

5.词向量、注意力向量LSTM隐藏层状态都设置为300

6.L2正则的系数设置10的-5

7.dropout率设置为0.5

3.2 Model Comparisons

In order to comprehensively evaluate the performance of IAN, we list some baseline approaches for comparison. The baselines are introduced as follows.

- **Majority** is a basic baseline method, which assigns the largest sentiment polarity in the training set to each sample in the test set.
- **LSTM** only uses one LSTM network to model the context and get the hidden state of each word. After that, the average value of all the hidden states is regarded as final representation and fed to a softmax function to estimate the probability of each sentiment label [Wang *et al.*, 2016].
- **TD-LSTM** adopts two long short-term memory (LSTM) networks to model the left context with target and the right context with target respectively. The left and right target-dependent representations are concatenated for predicting the sentiment polarity of the target [Tang *et al.*, 2016a].
- **AE-LSTM** represents targets with aspect embeddings. First this method models the context words via LSTM networks and then combine the word hidden states with aspect embeddings to supervise the generation of attention vectors, which are in turn used to produce the final representation for aspect-level sentiment classification [Wang *et al.*, 2016].
- **ATAE-LSTM** is developed based on AE-LSTM. ATAE-LSTM further strengthens the effect of aspect embeddings and appends the aspect embeddings with each word embedding vector to represent the context. The other design of ATAE-LSTM is the same as AE-LSTM [Wang *et al.*, 2016].

Table 2 shows the performance comparison of IAN with other baselines. From Table 2, we can observe that, the Majority method is the worst, meaning the majority sentiment polarity occupies 53.5% and 65.0% of all samples in the Restaurant and Laptop categories respectively. All the other methods are based on LSTM models and better than the Majority method, showing that LSTM has potentials in automatically generating representations and can all bring performance improvement for sentiment classification.

The LSTM method gets the worst performance of all the neural network baseline methods, because it treats targets equally with other context words and does not make full use of the target information. This also verifies the work of [Jiang *et al.*, 2011] which points out the importance of targets.

TD-LSTM outperforms LSTM over 1 percent and 2 percent on the Restaurant and Laptop category respectively, since it develops from the standard LSTM and processes the left and right contexts with targets. As we know, targets are twice represented and in some sense are specifically focused in the final representation.

Further, both AE-LSTM and ATAE-LSTM stably exceed the TD-LSTM method because of the introduction of attention mechanism. For AE-LSTM and ATAE-LSTM, they capture important information in the context with the supervision of target and generate more reasonable representations for aspect-level sentiment classification. We can also see that AE-LSTM and ATAE-LSTM further emphasize the modeling of targets via the addition of the aspect embedding, which is also the reason of performance improvement. Compared with AE-LSTM, ATAE-LSTM especially enhance the interaction between the context words and target and thus has a better performance than AE-LSTM.

Our IAN model takes a further step towards emphasizing the importance of targets through learning target and context representation interactively. We can see that IAN achieves the best performance among all baselines. Compared with ATAE-LSTM model, IAN improves the performance about 1.4% and 3.2% on the *Restaurant* and *Laptop* categories respectively. As we know, it is difficult to boost 1 percent of accuracy on sentiment classification. The main reason may be that IAN models the target and context using two connected attention networks which can influence each other. With this design, we can well learn the representations of targets and contexts whose collocation contributes to aspect-level sentiment classification. From Table 2, targets are progressively emphasized among these methods. The more attentions are paid to targets, the higher accuracy the system achieves. This also inspires our future work to further research the modeling of targets.

3.3 Analysis of IAN Model

In this section, we design a series of models to verify the effectiveness of our IAN model. First, we ignore the modeling of targets and design a No-Target model that just uses the context representation. Here we adopt only one LSTM networks with attention mechanism to model the context, where the attention vectors are computed through the averaged value of the target word embeddings. Then, we implement the second model No-Interaction which uses two LSTM networks

[2] Pre-trained word embedding of GloVe can download from http://nlp.stanford.edu/projects/glove/

实验设置很合理

1.没有引入目标的模型

2.没有引入交互的模型

3.有target和content
但是target部分没有引入注意力机制

Dataset	Restaurants	Laptops
No-Target	0.772	0.708
No-Interaction	0.769	0.706
Target2Content	0.775	0.712
IAN	**0.786**	**0.721**

Table 3: Analysis of Interactive attention Networks.

to learn the representations of target and context via their own local attentions without any interaction. Next, we design Target2Content which also employs two LSTM networks to learn target and context representations, but only considers to attention context words via target representations. The only difference between Target2Content and IAN is that IAN also adopts attention mechanism when modeling targets. Table 3 shows the performances of all these models. From Table 3, we can see that the No-Target model achieves worse performances than the IAN model. The results verify that target should be separately modeled and target representations can make contribution to judging the sentiment polarity of a target.

For the No-Interaction model, it gets the worst result among all the approaches. Compared with Target2Content and IAN, there is no interaction between target and context. Therefore, the interaction between target and content plays a great role in generating better representation for predicting target sentiment. Its results are also worse than No-Target, which may be because the effect of target representation, generated by local attention, is less important than the influence of target information for supervising learning content representation via attention.

As for Target2Content, it outperforms No-Interaction and is worse than IAN. Compare with IAN, it just learns the target representation via LSTM networks without supervision of context. As stated in Section 1, the collocated context and target can influence each other. That means that the interaction between target and content is critical to classifying target sentiment polarity and unidirectional attentions are not enough to the final representation.

As we expect, IAN achieves the best performance among all the methods. This is because the IAN model fully considers the effect of the target and the interaction between target and context which makes contributions to the target sentiment classification.

餐厅评论比笔记本
的精确率提升慢

会去分析结果可能
出现的原因，然后
然后引到自己的工作

Furthermore, from Table 2, we can see that, the improvements on *Restaurant* category is much less than those on *Laptop* category. To explain this phenomenon, we analyze these two categories and show the number and ratio of targets distributing in the two categories with respect to the target length (i.e., the word number included in a target), as in Table 4. From Table 4, we can see that, the number of instances with 1-word target in *Restaurant* category is 9% more than that in the *Laptop* category. This means that *Laptop* category has more multi-words targets than *Restaurant* category. In IAN, we model the targets by LSTM networks and interactive attentions. LSTM networks and interactive attention are more effective on modeling long targets than short targets. Conversely, average/max pooling, used in other methods, usually lose more information in modeling long targets compared with shorter target. This demonstrates the effectiveness of separately modeling the targets via LSTM networks and interactive attention.

3.4 Case Study

In this section, we use a review context "***the fish is fresh but the variety of fish is noting out of ordinary.***" and two targets "***fish***" and "***variety of fish***" from Restaurant category as a case study. We apply IAN to model the *context* and *target*, and get the correct sentiment polarity: *negative* and *positive* for two targets respectively. Figure 2 visualizes the attention weights on the *context* and *targets* computed by IAN. The left figure (a) gives the attention weights for two targets, and the right figure (b) shows the attention weights of the context. Each line presents the corresponding target and context pairs. The deeper color means the higher weight.

From Figure 2, we can observe that the common words *"the"* and *"of"* and punctuation "." are paid little attention by IAN in the context. This verifies our intuition that some common words and punctuations makes little contributions to judging target sentiment polarity. The meaning of the context in the case study should be that the quality of *fish* is good but the *variety of fish* has nothing special. Obviously, the words *fresh*, *nothing, out, ordinary* play a great role in the sentiment classification of *fish* and *variety of fish*, and our model pays much attention on those words as we expect. In addition, IAN also gives attention to the evaluated object: *fish* and *variety of fish*. Furthermore, when IAN is applied to the *fish*, the "attention" of the model is mainly paid to the corresponding target and its real collocation, and little attention is given to the *variety of fish* and its collocation. The situation is in contrast when judging the target *variety of fish*.

For the target *variety of fish*, *variety* is the head word, and other words are used to modify the head word. Therefore, *"variety"* is more important for expressing the target than the other two words *"of"* and *"fish"*. From Figure 2, we can see that the IAN pays more attention to the word *"variety"* than to the other words. This means that in our model, a target can provide useful information for its context to tune its attentions, and the context also plays an important role in supervising a target to get its focus. Thus, through the IAN, we can well model targets and contexts separately and interactively, and the concatenated representation of target and context are helpful for the aspect-level sentiment classification.

4 Related Work

分为

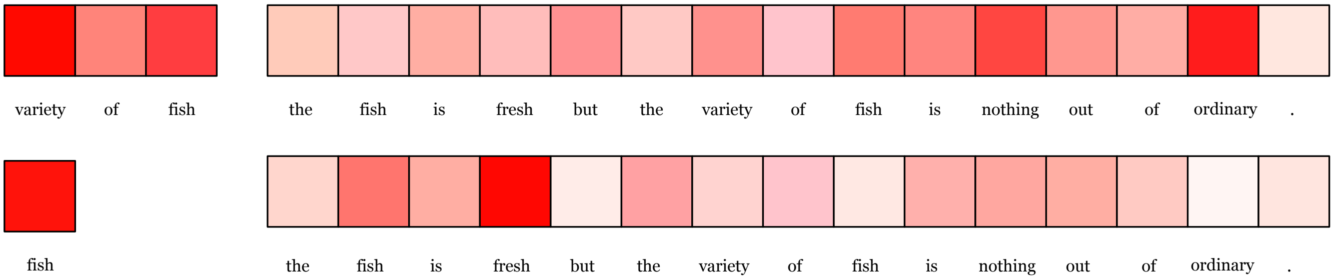
1.传统机器学习
2.神经网络方法

Aspect-level sentiment classification is a branch of sentiment analysis, and its research approaches can be split into two directions: traditional machine learning methods and neural networks methods.

Aspect-level sentiment classification is typically regarded as text classification problem. Accordingly, text classification methods, such as SVM [Pang *et al.*, 2002], can be applied to solve the aspect-level sentiment classification task without consideration of the mentioned target or aspect. Traditional machine learning methods mainly focus on extracting a set of features like sentiment lexicons features and bag-of-word

Dataset	1	2	3	4	5	>5
Restaurants-Train	2233/0.7515	506/0.1703	136/0.0458	47/0.0158	27/0.0091	22/0.0075
Restaurants-Test	677/0.7326	169/0.1829	44/0.0476	14/0.0152	7/0.0076	13/0.0141
Laptops-Train	1227/0.6582	495/0.2656	97/0.0520	33/0.0177	8/0.0043	4/0.0022
Laptops-Test	285/0.6077	147/0.3133	24/0.0512	7/0.0149	4/0.0086	2/0.0043

Table 4: Statistics of target length on SemEval 2014.



(a) weight for target

(b) weight for context

Figure 2: Case Study: Illustration of Attention Weights for *Context* and *Target*.

features to train a sentiment classifier [Rao and Ravichandran, 2009; Kaji and Kitsuregawa, 2007; Jiang *et al.*, 2011; Perez-Rosas *et al.*, 2012; Mohammad *et al.*, 2013]. Although these methods achieve a comparable performance, their results highly depend on the effectiveness of the handcraft features which are labor intensive.

传统ML高度依赖人工提取到特征的质量

Recently, kinds of neural networks methods, such as Recursive Neural networks [Socher *et al.*, 2011; Dong *et al.*, 2014; Qian *et al.*, 2015], Recursive Neural Tensor Networks [Socher *et al.*, 2013], Recurrent Neural networks [Mikolov *et al.*, 2010; Tang *et al.*, 2016a], Tree-LSTMs [Tai *et al.*, 2015] and Hierarchical LSTMs [Ruder *et al.*, 2016], have achieved a promising result on sentiment analysis. However, the neural network based approaches just make use of the contexts without consideration of targets which also make great contributions to judging the sentiment polarity of target.

对于神经网络算法，没有考虑到目标信息

To the best of our knowledge, Jiang et al. [2011] first present the importance of targets in sentiment classification and argue that 40% of sentiment classification errors are caused by not considering targets. Later, in order to incorporate targets into a model, Tang et al. [2016a] develop two target-dependent long short-term memory (LSTM) to model the left and right contexts with target, where the target information is automatically taken into account. In addition, Tang et al. [2016b] designs deep memory networks which consist of multiple computational layers to integrate the target information. Each layer is a context- and location- based attention model, which first learns the importance/weight of each context word and then utilizes the information to calculate context representation. To take target into account, Wang et al. [2016] explore the potential correlation of targets and sentiment polarities in aspect-level sentiment classification. In order to capture important information in response to a given target, they design an attention-based LSTM to concentrate on different parts of a sentence when different targets are taken as input.

As described in Section 1, target can help to attention the closely related part in the context. Therefore, we build an interactive attention network (IAN) model which respectively utilizes the target and context to compute the attention vector and learn the target and context representations. In this way, IAN can well acquire the appropriate final representations for aspect-level sentiment classification compared with other methods.

5 Conclusion

In this paper, we design an interactive attention networks (IAN) model for aspect-level sentiment classification. The main idea of IAN is to use two attention networks to model the target and content interactively. The IAN model can pay close attention to the important parts in the target and context and well generate the representations of target and context. Then, IAN benefits from the target representation which is always ignored in other methods. Experiments on SemEval 2014 verify that IAN can learn effective features for target and content and provide enough information for judging the target sentiment polarity. The case study also shows that IAN can reasonably pay attention to those words which are important to judging the sentiment polarity of targets.

Acknowledgements

Our work is supported by Major National Social Science Fund of China (No.12&ZD227) and National Natural Science Foundation of China (No.61370117 & No.61572049). The corresponding authors of this paper are Houfeng Wang & Sujian Li.

References

[Bengio *et al.*, 2003] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *journal of machine learning research*, 3(Feb):1137–1155, 2003.

[Collobert and Weston, 2008] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.

[Dong *et al.*, 2014] Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, pages 49–54, 2014.

[Huang *et al.*, 2012] Eric H Huang, Richard Socher, Christopher D Manning, and Andrew Y Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 873–882, 2012.

[Jiang *et al.*, 2011] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 151–160, 2011.

[Kaji and Kitsuregawa, 2007] Nobuhiro Kaji and Masaru Kitsuregawa. Building lexicon for sentiment analysis from massive collection of html documents. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1075–1083, 2007.

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

[Mikolov *et al.*, 2010] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Interspeech*, pages 1045–1048, 2010.

[Mohammad *et al.*, 2013] Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*, 2013.

[Pang and Lee, 2008] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(12):1–135, 2008.

[Pang *et al.*, 2002] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical methods in natural language processing*, pages 79–86, 2002.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1532–1543, 2014.

[Perez-Rosas *et al.*, 2012] Veronica Perez-Rosas, Carmen Banea, and Rada Mihalcea. Learning sentiment lexicons in spanish. In *LREC*, pages 3077–3081, 2012.

[Qian *et al.*, 2015] Qiao Qian, Bo Tian, Minlie Huang, Yang Liu, Xuan Zhu, and Xiaoyan Zhu. Learning tag embeddings and tag-specific composition functions in recursive neural network. In *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics*, pages 1365–1374, 2015.

[Qian, 1999] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[Rao and Ravichandran, 2009] Delip Rao and Deepak Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682, 2009.

[Ruder *et al.*, 2016] Sebastian Ruder, Parsa Ghaffari, and John G Breslin. A hierarchical model of reviews for aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02745*, 2016.

[Socher *et al.*, 2011] Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the conference on empirical methods in natural language processing*, pages 151–161, 2011.

[Socher *et al.*, 2013] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, Christopher Potts, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

[Tai *et al.*, 2015] Kai Sheng Tai, Richard Socher, and Christopher D Manning. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*, 2015.

[Tang *et al.*, 2016a] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. Effective lstms for target-dependent sentiment classification. In *International Conference on Computational Linguistics*, pages 3298–3307, 2016.

[Tang *et al.*, 2016b] Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. In *Proceedings of the conference on empirical methods in natural language processing*, pages 214–224, 2016.

[Vo and Zhang, 2015] Duy-Tin Vo and Yue Zhang. Target-dependent twitter sentiment classification with rich automatic features. In *IJCAI*, pages 1347–1353, 2015.

[Wang *et al.*, 2016] Yequan Wang, Minlie Huang, Li Zhao, and Xiaoyan Zhu. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the conference on empirical methods in natural language processing*, pages 606–615, 2016.