Knowledge-Based Systems 171 (2019) 81–92

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Semi-supervised Aspect-level Sentiment Classification Model based on Variational Autoencoder

Xianghua Fu [a], Yanzhi Wei [b], Fan Xu [c,*], Ting Wang [b], Yu Lu [a], Jianqiang Li [b], Joshua Zhexue Huang [b]

[a] Faculty of Arts and Sciences, Shenzhen Technology University, China
[b] College of Computer Science and Software Engineering, Shenzhen University, China
[c] College of Creative Design, Shenzhen Technology University, China

ARTICLE INFO

Article history:
Received 18 October 2018
Received in revised form 28 December 2018
Accepted 6 February 2019
Available online 12 February 2019

Keywords:
Aspect-level
Variational Autoencoder
Word embeddings
Semi-supervised
Sentiment classification

ABSTRACT

Aspect-level sentiment classification aims to predict the sentiment of a text in different aspects and it is a fine-grained sentiment analysis task. Recent work exploits an Attention-based Long Short-Term Memory Network to perform aspect-level sentiment classification. Most previous work are based on supervised learning that needs a large number of labeled samples, but the problem is that only a limited subset of data samples are labeled in practical applications. To solve this problem, we propose a novel Semi-supervised Aspect Level Sentiment Classification Model based on Variational Autoencoder (AL-SSVAE) for semi-supervised learning in the aspect-level sentiment classification. The AL-SSVAE model inputs a given aspect to an encoder a decoder based on a variational autoencoder (VAE), and it also has an aspect level sentiment classifier. It enables the attention mechanism to deal with different parts of a text when different aspects are taken as input as previous methods. Due to that the sentiment polarity of a word is usually sensitive to the given aspect, a single vector for a word is problematic. Therefore, we propose the aspect-specific word embedding learning from a topical word embeddings model to express a word and also append the corresponding sentiment vector into the word input vector. We compare our AL-SSVAE model with several recent aspect-level sentiment classification models on the SemEval 2016 dataset. The experimental results indicate that the proposed model is able to capture more accurate semantics and sentiment for the given aspect and obtain better performance on the task of the aspect level sentiment classification. Moreover, the AL-SSVAE model is able to learn with the semi-supervised mode in the aspect level sentiment classification, which enables it to learn efficiently using less labeled data.

1. Introduction

With the arrival of new technological advancement, the Internet has witnessed a tremendous change. Along this change, an large number of Internet users have began to make comments on the Internet more frequently. These comments can reflect users' attitudes on a product or service [1]. Sentiment classification of these reviews can not only make businesses easier to learn the needs of users, but also help users to decide whether a product is worth buying or not. Therefore, it is very meaningful to mine and analyze the sentiments of reviews [2].

Traditional approaches to classifying sentiment usually detect the sentiment of the entire text, without considering some different aspects. Because many businesses and users often prefer to know these sentiments of the product in different aspects, the sentiment of the entire text is too rough. Therefore, a fine-grained aspect level sentiment classification needs to be proposed, which analyzes different sentiments of different aspects of the same product. For example, "The vegetables here are delicious, but their service is very bad". The sentiment of aspect "food" is positive, while the sentiment of the given aspect "service" is negative.

Aspect-level sentiment classification [3] is a fine-grained subtask in the field of sentiment analysis. It reports detailed information to businesses and users as a reference to improve products sales. Currently, the aspect-level sentiment classification methods typically build an aspect-level sentiment classifier in a supervised mode with representative feature-based SVM [4] or neural network models [5]. Attention-based LSTM with Aspect Embedding models (ATAE-LSTM) [6] have been proposed to perform the aspect level sentiment classification by using attention mechanism [7] in LSTM models [8] to capture the key part of text with respect to a given aspect. However, the existing supervised learning methods need a large number of labeled data and cause high cost of

* Corresponding author.
E-mail addresses: fuxianghua@sztu.edu.cn (X. Fu), weiyanzhi2017@email.szu.edu.cn (Y. Wei), xufan@sztu.edu.cn (F. Xu), wangtingwtc@163.com (T. Wang), lvyu@sztu.edu.cn (Y. Lu), lijq@szu.edu.cn (J. Li), zx.huang@szu.edu.cn (J.Z. Huang).

https://doi.org/10.1016/j.knosys.2019.02.008
0950-7051/

human annotation. Therefore, how to make use of both labeled and unlabeled data for semi-supervised aspect-level sentiment classification becomes critical. Existing semi-supervised methods based on generative models is simple and intuitive, but the performance of the models will degrade [9] if the generative models are inconsistent with data distribution. As a result, such methods lack scalability and generality. Thus, there is a large potential to apply semi-supervised learning method in our scenario [10].

Recently, variational autoencoder (VAE) [11] has been proposed for semi-supervised learning [12] on image classification task. Based on the standard autoencoder, VAE adds a constraint on the encoder so that the latent variable follows a normal distribution. Since the variational autoencoder has been verified that it can extract global features of the text (such as sentiment, topic, and style) [13], it has also been successfully applied to semi-supervised text classification tasks [14]. Weidi Xu et al. have recently proposed Semi-supervised Sequential Variational Autoencoder (SSVAE), which applied the variational autoencoder to semi-supervised text classification successfully by feeding labels into its LSTM decoder at each time-step, and it has a satisfactory performance.

For the aspect-level sentiment classification, we normally use the word embedding (also known as word vector) to represent a word. Word embeddings learning techniques, such as Continuous Bag-of-Words (CBOW) and Skip-Gram [15], can learn continuous vector representations of words. The word vectors can capture both syntactic and semantic information of a word by modeling surrounding contexts. However, a single vector for a word cannot distinguish different sentiments corresponding to different aspects, because the sentiment polarity of a word is usually sensitive to a given aspect. For example, we hope the laptop has high performance, but not expect high price. In this case, researchers proposed a number of methods to improve these word vectors, such as a topical word embeddings model (TWE) [16], which considers both topics and contexts of words in the Skip-Gram model to learn topic-specific word embeddings so that it can detect different topics in different contexts and capture different semantics.

In this paper, we propose a novel Semi-supervised Aspect Level Sentiment Classification Model based on Variational Autoencoder (AL-SSVAE). The AL-SSVAE model is based on the variational autoencoder framework, introduces the given aspect of text into the encoder and decoder, and also adds an aspect level sentiment classifier for semi-supervised learning in the aspect level sentiment classification. First, we apply the original ATAE-LSTM model as the classifier to obtain the text label, and the input text is also encoded by it. Then we append the input aspect embedding into each word input vector in the decoder correspondingly. In addition, for the problem that a single word vector ignores its sentiment information and the discussed aspects, the topical word embedding model TWE is used to learn the aspect-specific word embedding. In particular, an aspect is considered as a topic. Instead of using the probabilistic topic model LDA [17] to obtain word-topic assignments in the TWE model, aspect and sentiment assignments of words are obtained by using a joint sentiment-topic model (JST) [18]. We learn aspect-specific word embeddings by treating word-aspect pairs as words in the Skip-Gram model and also decide a one-hot sentiment vector for each word according to the word-sentiment assignment. In the AL-SSVAE model, we put the learned aspect-specific word embeddings as input and also feed the corresponding sentiment vector at each time step. Therefore, the AL-SSVAE model can detect the discussed aspect and sentiment information of each word in the text, and model the interdependence between the aspects and sentiments of words and the input aspect. It also can be effective at extracting global semantics and sentiment features, and then achieves strong performance on the semi-supervised aspect level sentiment classification.

2. Related work

In this section, related works on aspect-level sentiment classification are first comprehensively reviewed. Then, a number of key related researches on variational autoencoder and word embeddings are introduced and discussed.

2.1. Aspect-level sentiment classification

Aspect-level sentiment classification is a fine-grained sentiment classification task, and it aims to predict the sentiment polarity of a text corresponding to the given aspect. Analyzing the sentiments of a person towards certain aspects is a fundamental task of natural language understanding [19].

Traditional methods are based on machine learning. These methods mainly focus on building classifier with a set of features like sentiment lexicons features and bag-of-word features to train a sentiment classifier [20–22]. These methods achieve satisfactory performance, but the classification results highly depend on the effectiveness of the feature extraction.

In recent years, different deep neural networks are used for sentiment classification [23–25]. However, these models do not consider the targets or aspects of texts. Jiang et al. [20] first argue that 40% of sentiment classification errors are due to that targets are not considered, and also point out that targets are critical in sentiment classification.

Later, Tang et al. [26] developed two target-dependent long short-term memory (LSTM) and then they also developed deep memory networks [27] to take the target information into account. Zhang et al. [28] proposed Gated Neural Networks for targeted sentiment analysis, and Ma et al. [29] proposed Interactive Attention Networks for aspect-level sentiment classification. However, these networks are all proposed for aspect-term level sentiment classification. There are two different aspect-level sentiment classification tasks: the aspect-term level sentiment classification and the aspect category-level sentiment classification. For example, given a sentence "The hamburger is delicious but expensive". The aspect-term level task is to choose a word that has to exist in the sentence as the target aspect, such as "hamburger", and then classify sentiment. However, the aspect-category level task is to choose a category word as target aspect, such as "food", and then classify sentiment. But the category word "food" does not necessarily exist in the sentence [6]. In this paper, our aspect level sentiment classification task only considers the aspect-category level task. So aspect level sentiment classification is actually referred to the aspect-category level task in this paper unless otherwise stated.

The aspect-term level sentiment classification models aim to infer the sentiment polarity of a text in terms of a given target word mentioned in the text. A target is an explicit aspect expression, so these target-dependent sentiment classifications cannot deal with implicit aspect expressions. An aspect, which is an implicit aspect expression, is a category of similar parts or attributes of the product. In particular, Wang et al. [6] propose an attention-based LSTM model with aspect embedding (ATAE-LSTM) to perform aspect level sentiment classification. In order to capture important information in response to a given target, they design an attention mechanism to concentrate on different parts of a sentence when different aspects are taken as input.

Despite the ATAE-LSTM model shows strong performance on aspect level sentiment classification, it still has two drawbacks. On one hand, ATAE-LSTM is a supervised learning method, which needs a large amount of labeled reviews and is a time consuming, labor and resources demanding technique. On the other hand, the ATAE-LSTM model uses a single word vector to represent the word, so it is impossible for the model to distinguish the different aspects of the word discussing in the text and the different sentiment expressing.

2.2. Variational autoencoder

In 2013, variational autoencoder (VAE) [11] was proposed as a generative model, which exhibits impressive performance in image related generation tasks. Based on the standard autoencoder, the VAE imposes some restrictions on the encoding process. It makes the latent vector to roughly follow a standard normal distribution, which is the most difference between variational autoencoder and the standard autoencoder. In this way, it is only necessary to generate a random latent vector from a standard normal distribution. And then we can generate any desired data by the decoder without having to encode the original data at first. Thus, VAE has attracted much more attention and has been widely used in data generation, visualization of high-dimensional data, and the filling of missing data.

In 2014, variational autoencoder was applied in semi-supervised learning [12] and achieved the best results in several image semi-supervised learning tasks. It has shown better performance in image task, but it has been problematic to sequential text for a long time. Then, a recurrent neural network was introduced [13], which is based on variational autoencoder generative model that incorporates distributed latent representations of entire texts. The model can be effective to capture global features from texts and generate diverse and well-formed texts through simple deterministic decoding. Until recently, TVAE model was proposed [30], which use variational autoencoder for review sentiment classification. These works show that variational autoencoder can be applied in Natural language processing successfully. However, the task of MTVAE is different from our task, which classifies the sentiment polarities in document level with multi-task learning.

There was a variational autoencoder proposed [14] that works in image semi-supervised classification task. But it fails in text semi-supervised classification if the decoder is vanilla LSTM. The work emphasizes the decoder's capability to distinguish between different labels, and it built a Semi-supervised Sequential Variational Autoencoder (SSVAE) with two specific decoder structures by feeding labels into its decoder at each time step to increase the capability. Therefore, variational autoencoder is also promising in the semi-supervised text classification. In this paper, we consider to introduce variational autoencoder to solve the semi-supervised learning problem in aspect-level sentiment classification.

2.3. Word embeddings

In natural language processing, word embeddings are usually used to represent a word. Word embeddings represent the word as a dense, low-dimensional and real-valued vector. Each dimension of vector contains a certain amount of semantic information. After training, words that have similar meaning are projected onto a similar point space. In this space, the distance between vectors can determine the similarity in their meanings and semantics. This is a basically simple and efficient word vector representation. Most of the latter research work of the word vector representation are based on this way.

Current word embeddings learning methods are usually based on the neural network language model. Bengio et al. firstly proposed neural network language model [31] in 2003 to learn word embeddings. The method uses a four-layer structure of statistical language models, including input layer, linear projection layer, nonlinear hidden layer and output layer, to automatically learn word embedding features representation that contains certain words meaning. However, the model results in enormous calculation time and high complexity due to the influence from nonlinear hidden layer. For this disadvantage, Word2vec model [15] is developed by Tomas Mikolov team of Google studio and it is one of the most widely used word embeddings learning tools currently. Word2vec model improves training efficiency and learns high-quality word feature representation by simplifying the internal structure of the neural network language model. The model removes hidden layer which is complicated and time-consuming. And its projection and output layer are also optimized in the training process, which makes the model train the data more flexibly and efficiently.

At present, each word is represented as a single vector in most word embeddings methods. But some words may have multiple senses. Later, some researchers proposed multi-prototype word embeddings models [32,33] and these works are mainly extended into two aspects. On one hand, the similar meaning words are clustered together firstly and then the target training word is extended in word meanings based on the results of the clustering. On the other hand, a single word semantic is expanded by using the outputs of the probabilistic model. Liu et al. [16] proposed topical word embeddings (TWE) based on the probabilistic topic model LDA [17]. The TWE model uses the final word-topic assignments obtained from LDA model as the input of the Skip-Gram model of Word2vec to train topical word embeddings. The model takes the topic information of the target word and the location context information into account together so that the learned topical word embeddings can detect different topic of words in the different contexts. Recent aspect-level sentiment classification models also use a single word vector for a word and ignore the discussed aspects and the sentiment information. However, the sentiment polarity of a word is usually sensitive to the target aspects, so whether these models can distinguish the discussed aspect and sentiment of a word in text is important for aspect-level sentiment classification. In this paper, we attempt to take advantage of the TWE model to learn aspect-specific word embeddings and thus to improve the performance of aspect-level sentiment classification with corresponding sentiment vector.

3. Models

3.1. Topical word embeddings model

Recent most word embeddings learning methods usually use a single word vector to represent a word. But in practice, a word may discuss different topics and has different semantics under different topics. To address this problem, Yang Liu et al. propose Topic Word Embedding model (TWE) [16] to identify different topics for words. The TWE model firstly applies latent topic model LDA [17] to obtain the final word-topic assignments. Then it considers the topic information and contexts of the target word simultaneously in the Skip-Gram model [15], and thus introduces topic information into word vector representations to generate topical word embeddings. There are three TWE models: TWE-1, TWE-2 and TWE-3. The TWE-1 model considers each topic as a pseudo word, and the word embeddings and topic embeddings are learned separately. The TWE-2 model considers each topic-word pair as a pseudo word, and gets topical word embeddings directly. The TWE-3 model concatenates each word embedding and their corresponding topic embedding to predict surrounding words in each slide window. Among these three models, TWE-2 model can get topical word embeddings directly, while the left two models have to concatenating the embedding of word and topic. And TWE-2 fully considers the inner interaction of a word-topic pair. Therefore, we attempt to take advantage of TWE-2 to learn aspect-specific word embeddings in this paper, and the TWE model in the rest of our paper is TWE-2. The architecture of the TWE model is illustrated in Fig. 1.

Our motivation of using TWE lies in that the Topic Word Embedding model (TWE) can identify different topics for words. The performance improvement of AL-SSVAE are on the basis of two

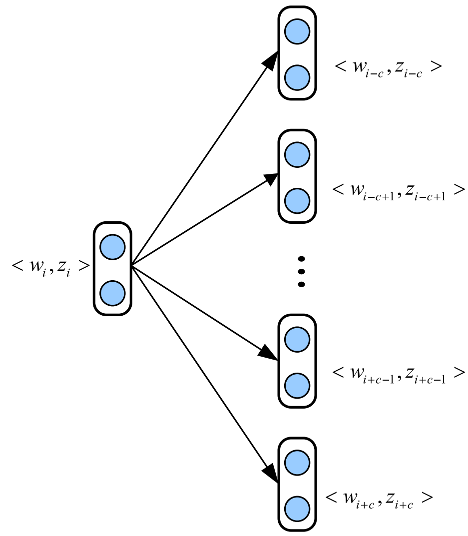


Fig. 1. The architecture of the TWE model.

aspects: the semi-supervised variational autoencoder structure and the aspect-specific word embeddings. In order to obtain the aspect-specific word embeddings, the TWE has been used. The existing aspect level sentiment classification model typically use a single word embedding vector to represent a word, so a word has only a single vector. However, a single vector for a word cannot distinguish different sentiments corresponding to different aspects, because the sentiment polarity of a word may be different for different given aspects. Application of the aspect-specific word embeddings can model the semantic and analysis the sentiment more accurately, while effectively improving the accuracy of sentiment classification on aspect level.

In the TWE model, each word has a labeled topic which is inferred from LDA. And each word-topic pair can be regarded as a pseudo word that contains some topic information from training dataset. The training process of TWE is the same as the Skip-Gram model. The TWE model predicts context words given a target word in a sliding window of size c, and then directly produces topical word embeddings. Given a word-topic pair sequence $\{\langle w_1, z_1\rangle, \ldots, \langle w_M, z_M\rangle\}$, the objective function of TWE is to maximize the following average log probability.

$$L_{TWE} = \frac{1}{M}\sum_{i=1}^{M}\sum_{-c\le k\le c, k\ne 0} \log p(\langle w_{i+k}, z_{i+k}\rangle | \langle w_i, z_i\rangle) \tag{1}$$

where c is the context size of a target word. TWE model formulates the probability $p(\langle w_c, z_c\rangle | \langle w_i, z_i\rangle)$ using a softmax function as follows.

$$p(\langle w_c, z_c\rangle | \langle w_i, z_i\rangle) = \frac{\exp(w_c^{z_c} \cdot w_i^{z_i})}{\sum_{\langle w_c, z_c\rangle \in \langle W, T\rangle} \exp(w_c^{z_c} \cdot w_i^{z_i})} \tag{2}$$

$w_c^{z_c}$ and $w_i^{z_i}$ are respectively the topical word embedding of target word $\langle w_i, z_i\rangle$ and context word$\langle w_c, z_c\rangle$. W is the word vocabulary and T is the number of topics. The TWE model initializes the vector of each word-topic pair with the corresponding word vector from Skip-Gram, and learn TWE model. The TWE model uses the similar stochastic gradient descent (SGD) for optimization as that of Skip-Gram [15], and gradients are calculated using the back-propagation algorithm.

3.2. Attention-based LSTM with aspect embedding

Wang et al. [6] introduced the aspect information into the LSTM [8] and explore the potential correlation of aspect and sentiment polarity in aspect-level sentiment classification. They build three models for aspect-level sentiment classification. The first model concatenates the aspect embedding and the word embedding to make the best use of aspect information. Due to that the standard LSTM is unable to detect which is the important part for aspect-level sentiment classification, for the second model an attention mechanism is developed that can capture the key part of a text given the aspect. In order to better take advantage of aspect information as the first model, they proposed an attention-based LSTM with aspect embedding model (ATAE-LSTM) [6]. It adds the aspect embedding into each word vector based on the second model and then the hidden vector can get the information from the aspect so that the model can obtain the interdependence between words and the aspect. The architecture of the ATAE-LSTM model is illustrated in Fig. 2.

In ATAE-LSTM model, $\{w_1, w_2, \ldots, w_N\}$ represent the word vector in a sentence and its length is N , v_a is the aspect embedding, α is a vector consisting of attention weights, $\{h_1, h_2, \ldots, h_N\}$ are hidden vectors, $H \in R^{d\times N}$ is the matrix consisting of hidden vectors, d is the size of hidden layer, d_a is the dimension of aspect embedding, $e_N \in R^N$ is a N dimensional vector and all of its elements are 1, r is a weighted representation of sentence with given aspect. They satisfy the following Eqs. (3)–(5).

$$M = \tanh\left(\begin{bmatrix} W_h H \\ W_v v_a \otimes e_N \end{bmatrix}\right) \tag{3}$$

$$\alpha = \mathrm{softmax}(\omega^T M) \tag{4}$$

$$r = H\alpha^T \tag{5}$$

The following h^* in Eq. (6) is the final sentence representation given the input aspect.

$$h^* = tanh(W_p r + W_x h_N) \tag{6}$$

where W_h, W_v, ω, W_p, W_x are projection parameters to be learned during training. In addition, the model uses a linear layer to convert the sentence vector to a real-valued vector whose length is the class number, and then adds a softmax layer to transform the vector to conditional probability distribution as Eq. (7), where W_s, b_s are the parameters for softmax layer.

$$y = \mathrm{softmax}(W_s h^* + b_s) \tag{7}$$

The ATAE-LSTM model is trained in an end-to-end way by back-propagation method. The objective function is the cross-entropy loss as $loss = -\sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2$, where y is the target

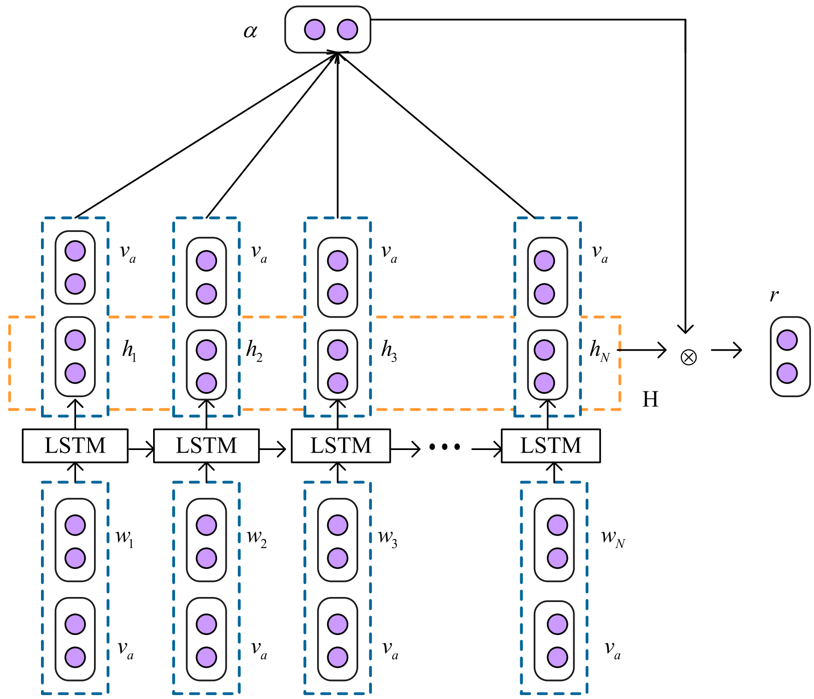


Fig. 2. The architecture of the ATAE-LSTM model.

sentiment distribution, $\hat{y}$ is the predicted sentiment distribution, i is the subscript of the i th sentence and j is the subscript of the j th category, and θ is the parameters. The training process aims to minimize the cross-entropy error between y and $\hat{y}$ for all sentences.

The ATAE-LSTM model is simple, intuitive and effective, and it is often used for aspect-level sentiment classification currently. However, the model uses a single word vector to represent a word, without using the discussed aspect and sentiment information. In fact, many aspects may be considered by a word in a text, and this word may have different sentiments in different given aspects. Therefore, the aspect and the sentiment information of each word are essential for aspect-level sentiment classification more accurately. In addition, the ATAE-LSTM model is a supervised model that needs a large number of labeled data and consumes a lot of manpower and material resources to label the data.

3.3. Semi-supervised sequential variational autoencoder

Recently, variational autoencoder (VAE) [11] have drawn a lot of attentions in text generation due to its impressive results reported in [13]. In addition, semi-supervised learning using VAE can help improving the discriminative results by using unlabeled data [12]. Xu et al. propose that it will fail in text classification task if the decoder is vanilla LSTM. They verify that it is significant for the model whether the decoder can fully distinguish between different categories labels from a perspective of reinforcement learning. Therefore, Semi-supervised Sequential Variational Autoencoder (SSVAE) [14] is proposed, which constructs two specific decoders respectively to add labels into its decoder at each step in order to achieve semi-supervised sequential text classification based on the variational autoencoder.

First, the model has two objective functions for labeled and unlabeled data. Given a labeled data(x, y), the evidence lower bound with corresponding latent variable z is as follows.

$$\log p_\theta(x, y) \geq E_{q_\phi(z|x,y)}[\log p_\theta(x|y, z)] + \log p_\theta(y) - \mathrm{D}_{KL}(q_\phi(z|x, y) \parallel p(z)) = -L(x, y) \quad (8)$$

The first term is the expectation of the conditional log-likelihood on latent variable z. The last term is the Kullback–Leibler divergence between the prior distribution $p(z)$ and the learned latent posterior$q_\phi(z|x, y)$.

Given a labeled data x, the unknown label y is predicted from the learnable classifier $q_\phi(y|x)$. The evidence lower bound is as follows.

$$\log p_\theta(x) \geq E_{q_\phi(y,z|x)}[\log p_\theta(x|y, z) + \log p_\theta(y) + \log p(z) - \log(q_\phi(y, z|x))] = \sum_y q_\phi(y|x)(-L(x, y)) + H(q_\phi(y|x)) = -U(x) \quad (9)$$

Actually, the classification loss also has to be added to the final objective function, so that the classifier $q_\phi(y|x)$ can be learned from labeled dataset. Hence the final objective function is as follows.

$$J^\alpha = \sum_{(x,y)\in S_l} L(x, y) + \sum_{(x)\in S_u} U(x) + \alpha \cdot E_{(x,y)\in S_l}[-\log q_\phi(y|x)] \quad (10)$$

S_l and S_u are labeled and unlabeled data set respectively, α is a hyper parameter of additional classification loss of labeled data.

The SSVAE model is consisted of three main parts: an encoder, a decoder and a text classifier, corresponding to $q_\phi(z|x, y)$, $p_\theta(x|y, z)$ and $q_\phi(y|x)$ respectively. In the encoder, each data pair (x, y) is encoded into a latent variable z, which is parameterized by a Gaussian distribution $q_\phi(z|x, y)$. The decoder is a generative model that estimates the probability to generate x with a given latent variable z and a label y. The encoder, the decoder and the classifier $q_\phi(y|x)$ can be implemented by various neural networks.

In order to enable the decoder to distinguish between different labels, the SSVAE model uses the label in each time step. Wang et al. studies two decoder structures. The first one concatenates word embedding and the label vector simply at each time-step. The second one directly passes the label information to the memory cell without the process of four gates in LSTM. It is defined by the following equations.

$$i_t = \sigma(W_{wi}w_t + W_{hi}h_{t-1}) \quad (11)$$

$$f_t = \sigma(W_{wf}w_t + W_{hf}h_{t-1}) \quad (12)$$

$$o_t = \sigma(W_{wo}w_t + W_{ho}h_{t-1}) \tag{13}$$

$$\hat{c}_t = \tanh(W_{wc}w_t + W_{hc}h_{t-1}) \tag{14}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \hat{c}_t + \tanh(W_{yc}y) \tag{15}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{16}$$

The SSVAE model applies variational autoencoder to the semi-supervised learning in the field of natural language processing at the first time. Compared with the best method of the previous generation, the SSVAE model has achieved some improvement in the sentiment classification tasks. Therefore, our paper aims to apply the variational autoencoder to perform semi-supervised learning in aspect-level sentiment classification based on the SSVAE model, which is simple, intuitive, and effective in semi-supervised text classification.

3.4. Semi-supervised aspect level sentiment classification based on variational autoencoder

To the best of our knowledge, the current aspect-level sentiment classifications are based on supervised learning, relying on a large number of labeled samples. Thus, it is necessary to turn to semi-supervised aspect-level emotional classification in order to make better use of unlabeled samples. The variational autoencoder generative model introduced in Section 2 is simple and intuitive, it can be also effectively used for semi-supervised text sentiment classification. Thanks to these advantages, we are the first attempt to combine the variational autoencoder with the aspect-level sentiment classification model for semi-supervised learning on aspect level sentiment classification at the first time.

Based on the TWE, ATAE-LSTM and SSVAE models introduced in Sections 2 and 3, we propose a Semi-supervised Aspect Level Sentiment Classification Model based on Variational Autoencoder, which we call AL-SSVAE model. The model is based on the following three aspects: (1) The TWE model [16] integrates topics into basic word embedding representation and allows the resulting topical word embeddings to model different meanings of a word under different context, which is consistent with the actual text expression. (2) The ATAE-LSTM model [6] developed an attention mechanism to obtain the key part for a given aspect in the text. It also uses the hidden vector to calculate the attention weight, thus the interdependence between words and the given aspect can be modeled, so it is more accurate in aspect-level sentiment classification. (3) The SSVAE model [14] used a conditional LSTM as the decoder of VAE [11], which introduces the text label into the decoder at each step. Therefore, it has successfully applied the variational autoencoder to semi-supervised learning for texts, and achieved a certain improvement in semi-supervised text sentiment classification.

Based on the variational autoencoder framework, the AL-SSVAE model uses the aspect information in the encoder and the decoder, and adds an aspect level emotion classifier. We use the ATAE-LSTM model as a classifier for AL-SSVAE and also use it to encode text to get the feature representation of a text given an input aspect. Then we append the input aspect embedding and the label into each word input vector to correspond to the encoder and the decoder. In addition, we consider the topic as the aspect. First, we analyze the aspect and sentiment of each word in the text using joint sentiment-topic model (JST) [18] to obtain final aspect and sentiment assignments of all words. Second, we input the word-aspect pair into the TWE model to learn aspect-specific word embeddings and decide a one-hot sentiment vector for each word according the sentiment assignment of each word. Then, we also consider the discussed aspect and sentiment information of each word in the text by introducing the aspect-specific word embedding to represent the word and corresponding sentiment vector in the AL-SSVAE model. Therefore, the AL-SSVAE model is able to detect the aspect and the sentiment that each word expresses in the training process. The richer semantic information and sentiment features can be obtained more accurately so that successfully utilizes the variational autoencoder for semi-supervised aspect level sentiment classification. The architecture of the AL-SSVAE model is illustrated in Fig. 3.

In our AL-SSVAE model, the JST model gets final aspect and sentiment assignments of all words at first, where$\langle w'_t, a_t\rangle$ is the current word-aspect pair, and$\langle w'_{t-c}, a_{t-c}\rangle, \dots, \langle w'_{t+c}, a_{t+c}\rangle$ are the context word-aspect pair. Then the TWE model inputs word-aspect pairs to learn aspect-specific word embeddings which can capture the aspect information of each word in different contexts, where$\{w_1, w_2, \dots, w_N\}$ represent aspect-specific word embeddings and the length of the sentence is N , $\{s_1, s_2, \dots, s_N\}$ are the corresponding sentiment vectors of aspect-specific words,v_a is the given aspect embedding,α is a vector consisting of attention weights, and $\{h_1, h_2, \dots, h_N\}$ are hidden vectors.

In the encoder of AL-SSVAE model, each data pair (x, y, v_a) is encoded into a latent space, where $x = \{[w_1, s_1], [w_2, s_2], \dots, [w_N, s_N]\}$, the $[w_t, s_t]$ is connection of an aspect-specific word embedding and its corresponding sentiment, and it represents aspect-specific sentimental word. The encoder introduces the aspect embedding to get the final sentence representation $\hat{x}$, and then use $\hat{x}$ and the label y to parameterize the posterior probability $q_\phi(z|x, y, v_a)$. The distribution of latent variable z is parameterized by a Gaussian distribution $q_\phi(z|x, y, v_a)$. The equations of the encoder are as follows.

$$\hat{x} = f_{enc}(x, v_a) \tag{17}$$

$$q_\phi(z|x, y, v_a) = N(\mu(\hat{x}, y), diag(\sigma^2(\hat{x}, y))) \tag{18}$$

$$z \sim q_\phi(z|x, y, v_a) \tag{19}$$

The decoder is a conditional generative model that uses the sampled latent variable to reconstruct the input, and introduces aspect information by concatenating each word with the aspect embedding. The decoder estimates the probability of a generating x with the given latent variable z, the label y, and the given aspect v_a.

$$p_\theta(x|y, z, v_a) = D(x|f_{dec}(y, z, v_a)) \tag{20}$$

where $f_{dec}(y, z, v_a)$ is used to parameterize a distribution D, typically a Bernoulli or Gaussian distribution. Based on the decoder of SSVAE model, the AL-SSVAE model also feeds labels at each time step and appends the given aspect embedding into each aspect-specific word vectors. The LSTM decoder network $f_{dec}(y, z, v_a)$ is defined by the following equations.

$$i_t = \sigma(W_{wi}w_t + W_{si}s_t + W_{v_a i}v_a + W_{hi}h_{t-1}) \tag{21}$$

$$f_t = \sigma(W_{wf}w_t + W_{sf}s_t + W_{v_a f}v_a + W_{hf}h_{t-1}) \tag{22}$$

$$o_t = \sigma(W_{wo}w_t + W_{so}s_t + W_{v_a o}v_a + W_{ho}h_{t-1}) \tag{23}$$

$$\hat{c}_t = \tanh(W_{wc}w_t + W_{sc}s_t + W_{v_a c}v_a + W_{hc}h_{t-1}) \tag{24}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \hat{c}_t + \tanh(W_{yc}y) \tag{25}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{26}$$

For the AL-SSVAE model, we have two cases to consider for the labeled data and the unlabeled data. In the first case, given the labeled data (x, y, v_a) and the latent variable z, the variational lower bound is as follows.

$$\begin{aligned}\log p_\theta(x, y, v_a) \geq{}& E_{q_\phi(z|x,y,v_a)}[\log p_\theta(x|y, v_a, z)] \\ &+ \log p_\theta(y) + \log p_\theta(v_a) \\ &- D_{KL}(q_\phi(z|x, y, v_a) \parallel p_\theta(z)) = -L(x, y, v_a)\end{aligned} \tag{27}$$

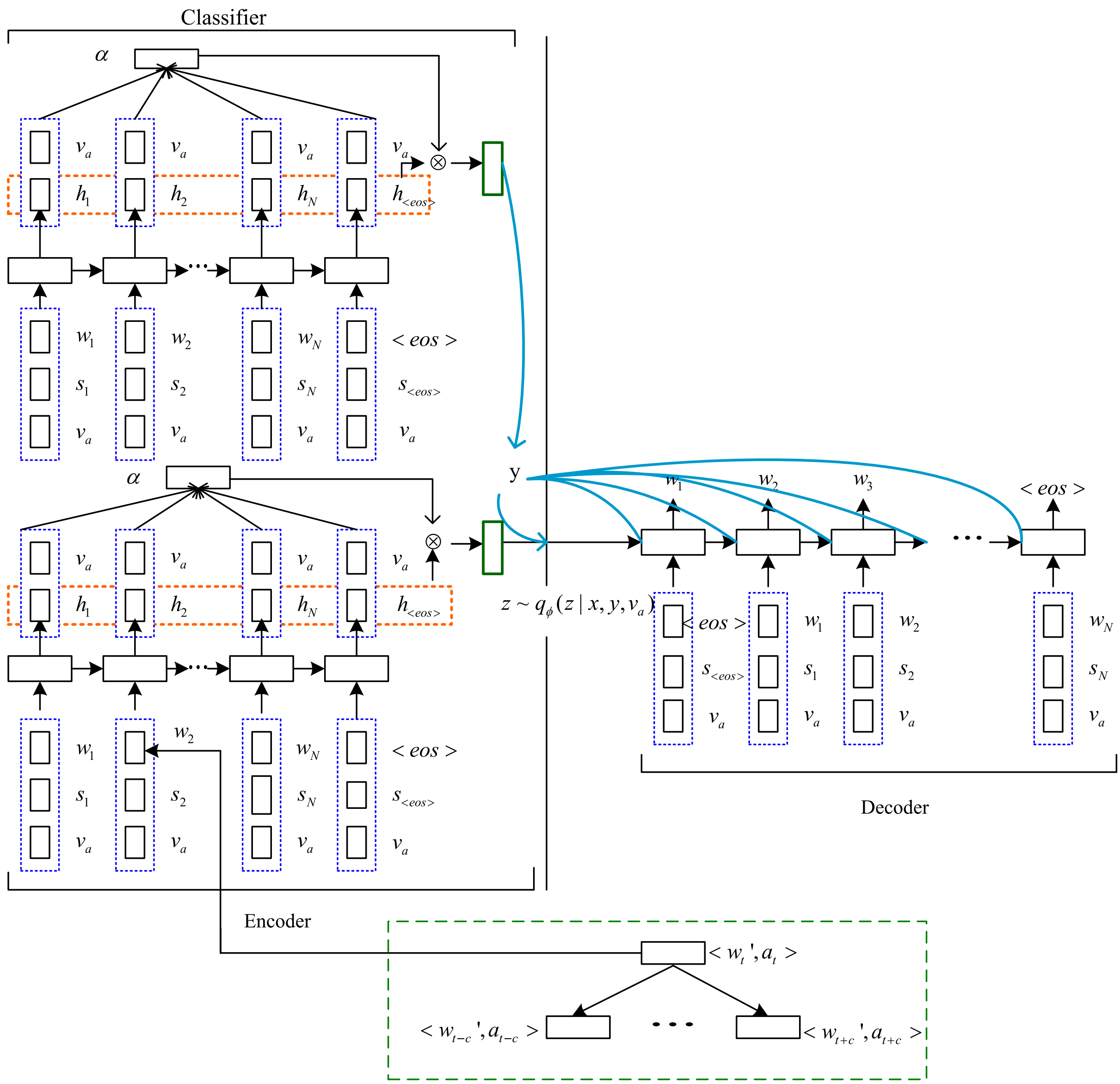


Fig. 3. The architecture of the AL-SSVAE model.

where the first term is the expectation of the conditional log-likelihood on latent variable z, and the last term is the Kullback–Leibler divergence between the prior distribution $p_\theta(z)$ and the learned latent posterior $q_\phi(z|x, y, v_a)$.

In the second case of given the unlabeled data (x, v_a), the unknown label y is predicted from the learnable classifier $q_\phi(y|x, v_a)$. The evidence lower bound is as follows.

$$\log p_\theta(x, v_a) \geq E_{q_\phi(y,z|x,v_a)}[\log p_\theta(x|y, v_a, z) + \log p_\theta(v_a) + \log p_\theta(y) + \log p_\theta(z) - \log q_\phi(y, z|x, v_a)] \tag{28}$$

We simplify the above formula as follows.

$$\log p_\theta(x, v_a) \geq \sum_y q_\phi(y|x, v_a)(-L(x, y, v_a)) + H(q_\phi(y|x, v_a)) = -U(x, v_a) \tag{29}$$

Therefore, the objective function of the entire data set is as follows.

$$J = \sum_{(x,y,v_a)\in S_l} L(x, y, v_a) + \sum_{(x,v_a)\in S_u} U(x, v_a) \tag{30}$$

In fact, we also add the classification loss to the above objective function so that the classifier $q_\phi(y|x, v_a)$ can be learned from labeled dataset. Hence the final objective function is as follows.

$$J^\alpha = J + \alpha \cdot E_{(x,y,v_a)\in S_l}[-\log q_\phi(y|x, v_a)] \tag{31}$$

where S_l and S_u are labeled and unlabeled data set respectively, and α is a hyper parameter of additional classification loss of labeled data.

4. Experiments

To investigate the performance of our proposed AL-SSVAE model on semi-supervised aspect-level sentiment classification, we compare it with existing aspect-level sentiment classification models. To the best of our knowledge, we have not found any related models for semi-supervised aspect-level sentiment classification based on a recurrent neural network using a single word vector for a word. Therefore, we compare the AL-SSVAE model with LSTM, AE-LSTM, AT-LSTM, ATAE-LSTM models in order to demonstrate that the proposed model has better performance in the semi-supervised learning on aspect-level sentiment classification. The model is evaluated from three aspects: (1) to calculate accuracy of sentiment classification; (2) to visualize the attention weights and analyze that how attention focuses on words with the influence of a given aspect; (3) to illustrate the effectiveness of the model from the sentiment classification results of some typical examples.

In our experiments, the final aspect and sentiment assignments for each word in the datasets need to be obtained using the joint sentiment-topic model (JST) [18]. Hyper-parameters of the JST model α, β, γ are set to 0.1, 0.01 and 0.1. The number of aspects is determined according to the given dataset. Then we utilize the topical word embedding model TWE [16] to obtain aspect-specific word embeddings. The initial learning rate is 0.025 and the window size is five. The dimension of word embeddings is 300 and we use the hierarchical softmax techniques during learning the TWE model. In addition, the dimension of aspect embeddings and the size of hidden layer are 300. The length of attention weights is the same as the length of sentence. We train the AL-SSVAE model using the ADAM [34] optimizer with learning rate of 4e-3. Refer to Bowman's work [13], the cost annealing trick was adopted to smooth the training by gradually increasing the weight of KL cost from zero to one and Word dropout technique is also utilized whose rate was scaled from 0.25 to 0.5. Hyper-parameter α was scaled from 1 to 2. The dimension of latent variable is 50. And we tune parameters on the validation set.

4.1. Datasets

Our experiments use the dataset of SemEval 2016 Task 5 [35]. The dataset consists of user reviews and provides 19 training sets and 20 test sets in 8 languages and 7 fields. It can be used in many tasks such as the detection of aspect categories, extraction of viewpoints target expressions, and sentiment classification. Our proposed AL-SSVAE model is to predict the sentiment polarity of a review on a given aspect, so we extract the datasets that meet our experimental requirements in four areas. The experiment is conducted in both English and Chinese languages. The Chinese dataset includes camera reviews (CAME) and mobile phone reviews (PHNS). The English datasets includes restaurant reviews (REST) and laptop computer reviews (LAPT). Each review contains a list of aspect categories and corresponding sentiment polarities. Each aspect category is defined by "entity E# attribute A". Entity E and attribute A are selected from entity types (e.g. camera, display, battery, system) and Attributes (e.g. performance, design) respectively. Each determined aspect category is assigned a sentiment polarity. We choose three sentiment labels representing positive, neutral, and negative. The data statistics is presented in Table 1. In particular, when training a joint sentiment-topic model, we employ the prior information during the initialization of posterior distribution. It indicates that we assign a sentiment label to each word in advance. In this paper, we use the MPQA[1] as an English sentiment lexicon and HowNet[2] as a Chinese sentiment lexicon to incorporate prior information. And we select Wikipedia, the largest online knowledge base, to pre-train the word vectors by using the Google Word2Vec toolkit [15].

Before training the AL-SSVAE model, the dataset must be preprocessed. For the Chinese datasets, we use the jieba[3] tool to segment words, and also remove stop words according to the HIT's stop wordlist[4] and some special characters. For the English datasets, we first convert uppercase letters to lowercase letters and then remove punctuation, numbers and other non-alphabet characters. We also remove stop words[5] based on the stop word list provided by University of Glasgow's information retrieval language tool.

Table 1
The experimental datasets.

Datasets	Train	Test	Total
CAME	1259	481	1740
PHNS	1333	529	1862
REST	2507	859	3366
LAPT	2909	801	3710

Table 2
The division of each experimental dataset.

<table>
<tr><td colspan="4">Each dataset of SemEval 2016 Task 5</td></tr>
<tr><td colspan="3">Original training set</td><td rowspan="3">Test set</td></tr>
<tr><td colspan="2">Training set</td><td rowspan="2">Validation set</td></tr>
<tr><td>Part A: labeled data</td><td>Part B: unlabeled data</td></tr>
</table>

Our data are all from the four datasets of SemEval 2016 Task 5, which are CAME, PHNS, REST and LAPT respectively. Each dataset consists of training data and test data. For each dataset, test data are not modified in our experiment. But we divide the original training data into an actual training set and a validation set. Then, the training set is further divided into two parts: part A and part B. For part A, we remain the labels to become our actual labeled training data. For part B, we remove the labels to become our unlabeled data. The division of our labeled and unlabeled data is illustrated in Table 2.

4.2. Aspect level sentiment classification

The performance indicator of our proposed AL-SSVAE model on aspect level sentiment classification is accuracy. The accuracy rate is between 0 and 1. Higher accuracy indicates that the result of aspect-level classification is more accurate. To illustrate the performance of the AL-SSVAE model in aspect-level sentiment classification, experiments were performed on CAME, PHNS, REST, and LAPT datasets. The AL-SSVAE model was compared with the other four models mentioned above. Table 3 shows accuracy on the aspect-level classification.

From the results of the semi-supervised aspect-level sentiment classification on different datasets in Table 3, we can find that the AL-SSVAE model has the highest accuracy compared with the other four models for all datasets. On the CAME dataset, the accuracy of the AL-SSVAE model is 79.72% while the accuracy of the ATAE-LSTM model is 76.66%. On the PHNS dataset, the accuracy of the AL-SSVAE model is 80.66%, which is about 3% higher than that of the ATAE-LSTM model. On the English datasets REST and LAPT, the accuracy of the AL-SSVAE model is up to 86.72% and 88.98% respectively. Compared with the ATAE-LSTM model, the accuracy was improved by more than 4%.

Because of the variational autoencoder structure, the AL-SSVAE model can effectively capture the global semantic information and sentiment characteristics of the text. In addition, the model also introduces aspect-specific word embeddings and sentiment vectors to make it possible to identify the discussed aspect and sentiment information of words in the text during training so that it can capture more accurate semantic and sentiment. Thus, the AL-SSVAE model can improve the accuracy on aspect-level sentiment classification. In general, the accuracy of sentiment classification on English datasets is higher than that on Chinese datasets. Maybe

1 http://www.cs.pitt.edu/mpqa/.
2 http://www.keenage.com/html/c_bulletin_2007.htm.
3 https://github.com/fxsjy/jieba.
4 https://github.com/dongxiexidian/Chinese/blob/master/stopwords.dat.
5 http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words.

Table 3
Accuracy on aspect-level sentiment classification.

Dataset	Model	Accuracy
CAME	LSTM	74.85%
	AE-LSTM	75.44%
	AT-LSTM	76.06%
	ATAE-LSTM	76.66%
	AL-SSVAE	79.72%
PHNS	LSTM	75.24%
	AE-LSTM	75.93%
	AT-LSTM	76.68%
	ATAE-LSTM	77.40%
	AL-SSVAE	80.66%
REST	LSTM	81.30%
	AE-LSTM	82.12%
	AT-LSTM	83.01%
	ATAE-LSTM	82.70%
	AL-SSVAE	86.72%
LAPT	LSTM	82.11%
	AE-LSTM	83.03%
	AT-LSTM	84.13%
	ATAE-LSTM	84.56%
	AL-SSVAE	88.98%

the small-scale Chinese datasets CAME and PHNS results in insufficient semantics or the Chinese datasets are less normal than the English datasets in their language representations and contain more noise information.

4.3. Visualization of attentions

Our proposed AL-SSVAE model utilizes an attention mechanism to capture key words in the text in response to a given aspect which is important for aspect-level sentiment classification. Because we introduce specific-aspect word embeddings and sentiment vectors based on the original model ATAE-LSTM [6], the AL-SSVAE model can model the interdependence between the aspects and sentiments of words and the given aspect. Thus, the AL-SSVAE model can capture more accurate key part of text to improve the performance of aspect-level sentiment classification. Therefore, it is necessary to analyze which words decide the sentiment polarity of the text given an aspect. We can obtain the attention weight in Eq. (4) and visualize the attention weights accordingly. Table 4 lists some examples of sentiment attention results of the ATAE-LSTM and AL-SSVAE model on REST dataset. The higher weight, the more important it is for sentiment classification in response to a given aspect. Here we color the words with top-5 scores.

Example (1) and (2) are the same sentence but has two different aspects respectively. We find that the ATAE-LSTM is confused to catch sentiment words for the given aspect. Given aspect "AMBIENCE#GENERAL", "fantastic" which expresses another aspect is assigned to high attention weight. And given aspect "FOOD#QUALITY", both "small" and "cramped" obtain high scores. However, "place is small" and "cramped" get higher attention weights than other words in the text when given aspect "AMBIENCE#GENERAL" and the expression "the food is fantastic" captures top weights in response to the aspect "FOOD#QUALITY" in the AL-SSVAE model. Thus, we find out that AL-SSVAE model can more accurately catch related sentiment words to an aspect comparing with the ATAE-LSTM model. Example (3) does not contain any sentiment words and example (4) is a question, but both ATAE-LSTM and AL-SSVAE model is able to catch important words. This is because such sentences are relatively simple and often appear in text datasets, and the model can learn the regularization to capture semantic meanings of the text. But, Table 4 shows that the "again" in the example (3) and the "want" in the example (4) get higher attention weights when we perform the aspect-level sentiment classification in the AL-SSVAE model. Compared with the ATAE-LSTM model, the AL-SSVAE model can more accurately assign weights to related words for a given aspect and the variational autoencoder structure enables the model to obtain richer global semantic information and sentiment characteristics so as to perform semi-supervised aspect-level sentiment classification better.

Besides, we also use a visualization tool HemI [36] to visualize the sentences. The sentiment attention results of ATAE-LSTM and AL-SSVAE for a sentence with two different aspects are shown in Fig. 4. The color depth expresses the attention weight of the word in the text.

Comparing with Fig. 4(a) and (b), it can be seen that the AL-SSVAE model has higher accuracy than the ATAE-LSTM model when matches the sentiment words to more than one aspect in a sentence. In the ATAE-LSTM model, given aspect "FOOD#QUALITY", the model not only assigns high weights to "food", "not", and "worth", but also "like" and "atmosphere" get some weights which are actually used to express aspect "AMBIENCE#GENERAL" in the sentence. Given aspect "AMBIENCE #GENERAL", the model also assigns some weights to "not" and "worth". However, in the AL-SSVAE model, it is obvious that more accurate weights are assigned to each word in a sentence. Given aspect "AMBIENCE#GENERAL", "like" and "atmosphere very much" have higher weights than other words. The expression "the food was not worth the price" is obviously higher given aspect "FOOD#QUALITY". Therefore, our proposed AL-SSVAE model can find the corresponding sentiment words correctly in aspect-level sentiment classification.

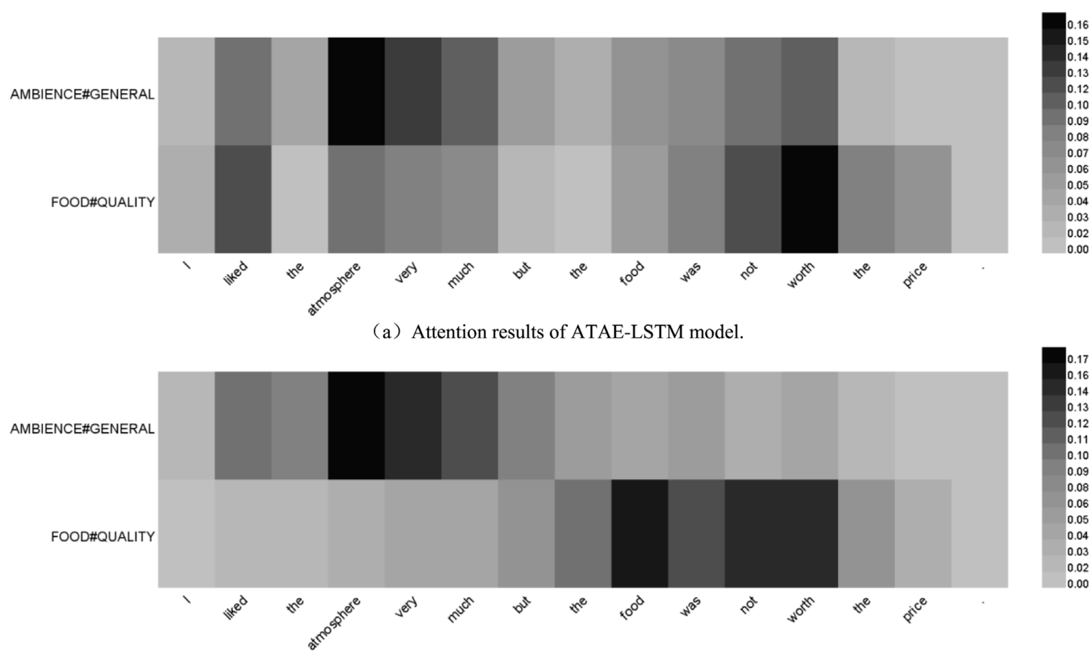
4.4. Examples study

Based on the experimental results, our proposed AL-SSVAE model has obtained better performance on aspect-level sentiment classification. We also can list some typical examples on LAPT dataset to show the advantages of the model in Table 5.

Example (1) and Example (2) are the same sentence corresponds to two aspects: "LAPTOP#GENERAL" and "MOUSE #OPERATION_PERFORMANCE". The results show that the AL-SSVAE model can distinguish different sentiments in response to different aspects in a sentence. Both example (3) and example (4) contain the same sentiment word "high", but the word has different sentiment polarities. "High" is positive for the performance of laptop. As for the price of laptop, it is negative. It can be seen that our proposed AL-SSVAE model can distinguish the different sentiment expressed by the same word for different aspect in the text. The results show that considering both the discussed aspect and sentiment information of each word in the AL-SSVAE model is important to accurately discriminate different sentiment polarities of words for different aspects. Both the example (5) and example (6) do not contain any sentiment words, but the AL-SSVAE model also can correctly predict the sentiment polarities because the two sentences are relatively simple, and some similar sentences often appear in the dataset so that our model can continuously learn the sentiment. Example (7) is a question that is generally difficult to predict the sentiment, but the AL-SSVAE model assigns higher weight to "how could" and "not love" and correctly predicts the sentiment polarity given the input aspect. The result shows that our model can capture very rich semantic information and sentiment characteristics of texts. Examples (8) to example (11) are the same sentence with four different aspects. The sentence is very long and complex, but the AL-SSVAE model can use the attention mechanism to accurately obtain the different key parts for different aspects in the text, and thus correctly predict the sentiment polarities.

Table 4
Examples of sentiment attention results of ATAE-LSTM and AL-SSVAE on REST dataset.

Model	Example	Aspect	Attention
ATAE-LSTM	(1)	AMBIENCE#GENERAL	The place is small and cramped but the food is fantastic.
	(2)	FOOD#QUALITY	The place is small and cramped but the food is fantastic.
	(3)	RESTAURANT#GENERAL	We would go back again.
	(4)	RESTAURANT#GENERAL	What more could you want?
AL-SSVAE	(1)	AMBIENCE#GENERAL	The place is small and cramped but the food is fantastic.
	(2)	FOOD#QUALITY	The place is small and cramped but the food is fantastic.
	(3)	RESTAURANT#GENERAL	We would go back again.
	(4)	RESTAURANT#GENERAL	What more could you want?



(a) Attention results of ATAE-LSTM model.
(b) Attention results of AL-SSVAE model.

Fig. 4. Sentiment attention results of ATAE-LSTM model and AL-SSVAE model for a sentence.

4.5. Model analysis

The performance improvement of AL-SSVAE comes from two aspects: the semi-supervised variational autoencoder structure and the aspect-specific word embeddings.

For the purpose of exploring the influence of SSVAE (semi-supervised variational autoencoder structure), we remove the SS-VAE component in our model to experiment. The accuracy decreased on our four datasets as shown in Table 6. The results show the effectiveness of the SSVAE, and illustrate that SSVAE contributes to the improvement.

In order to explore the influence of the aspect-specific word embeddings, we use conventional word embeddings to experiment. We find out the accuracy also decreased on our four datasets as shown in Table 6. The results can show the effectiveness of the aspect-specific word embeddings, and illustrate that aspect-specific word embeddings contribute to the improvement. In addition, we notice that the accuracy of AL-SSVAE without aspect-specific word embeddings is lower than the standard ATAE-LSTM on our four datasets. This may be due to that ATAE-LSTM is supervised learning method and AL-SSVAE is semi-supervised learning method. For AL-SSVAE, we divide part of samples from the training set as unlabeled data. This results in a reduction in the number of labeled training samples to AL-SSVAE model. For ATAE-LSTM, it has more labeled training data than ATAE-LSTM. As a result, the standard ATAE-LSTM have higher accuracy.

5. Conclusions and future works

In this paper, we have proposed the AL-SSVAE model for semi-supervised learning in the aspect level sentiment classification. On the basis of that variational autoencoder (VAE) can be applied to

Table 5
Examples of aspect-level sentiment classification results on LAPT dataset.

Examples	Aspects	Sentences	Sentiment polarities
(1)	LAPTOP#GENERAL	Received this item and everything is great except that my touchpad is reversed.	Positive
(2)	MOUSE#OPERATION_PERFORMANCE	Received this item and everything is great except that my touchpad is reversed.	Negative
(3)	LAPTOP#OPERATION_PERFORMANCE	This is very fast, high performance computer.	Positive
(4)	LAPTOP#PRICE	While the price may seem high, it is worth every penny with its many perks.	Negative
(5)	LAPTOP#GENERAL	I use it everyday.	Positive
(6)	SUPPORT#QUALITY	So I am going back to Genius Bar for visit.	Negative
(7)	LAPTOP#GENERAL	How could you not love a Mac?	Positive
(8)	LAPTOP#OPERATION_PERFORMANCE	The MacBook Pro 13.3 in. laptop is for those who want to use a Mac, but value performance and upgradability over weight and screen resolution.	Positive
(9)	LAPTOP#USABILITY	The MacBook Pro 13.3 in. laptop is for those who want to use a Mac, but value performance and upgradability over weight and screen resolution.	Positive
(10)	LAPTOP#DESIGN_FEATURES	The MacBook Pro 13.3 in. laptop is for those who want to use a Mac, but value performance and upgradability over weight and screen resolution.	Negative
(11)	DISPLAY#QUALITY	The MacBook Pro 13.3 in. laptop is for those who want to use a Mac, but value performance and upgradability over weight and screen resolution.	Negative

Table 6
Contribution analysis of AL-SSVAE.

Dataset	Model	Accuracy
CAME	AL-SSVAE	79.72%
	AL-SSVAE(without SSVAE)	78.10%
	AL-SSVAE(without aspect-specific word embeddings)	74.15%
PHNS	AL-SSVAE	80.66%
	AL-SSVAE(without SSVAE)	78.92%
	AL-SSVAE(without aspect-specific word embeddings)	73.94%
REST	AL-SSVAE	86.72%
	AL-SSVAE(without SSVAE)	84.85%
	AL-SSVAE(without aspect-specific word embeddings)	80.76%
LAPT	AL-SSVAE	88.98%
	AL-SSVAE(without SSVAE)	86.90%
	AL-SSVAE(without aspect-specific word embeddings)	81.43%

the semi-supervised text classification and the current aspect level sentiment classification is supervised based on a large number of labeled data, we have built a model of semi-supervised aspect level sentiment classification network based on the variational autoencoder to make use of the unlabeled data. The aspect information are inputted into the encoder and decoder, and we also add an aspect level sentiment classifier. We have applied the attention-based LSTM with aspect-embedding (ATAE-LSTM) as the classifier and encode the text. Then we append the aspect embedding into each word to generate the aspect information. Further, we have considered both discussed aspect and sentiment information of each word by input sentiment vector and aspect-specific word embeddings learned from the TWE model. Experimental results show that the proposed model obtains better performance over other model and achieves semi-supervised aspect level sentiment classification.

As the Internet becomes pervasive for worldwide users, the customer reviews have been increasing rapidly and significantly. Thus, there is a need for more powerful and effective techniques that are able to perform aspect level sentiment classification. While the existing aspect level sentiment classification models are difficult to meet the requirements in practice. Therefore, how to mine the aspect level sentiment quickly and efficiently is still a challenging task. Parallel processing is a direction for future research, which has the capability of tackling with the problem of large-scale sentiment mining and improving the existing model.

Although our proposed model improves the accuracy of aspect level sentiment classification and achieves the semi-supervised learning in aspect level sentiment classification, the model is unable to predict the aspect level sentiments of some special texts that have comparative sentiment words, long texts without sentiment words, etc. For future work, it is vital to develop specific neural networks to solve the problem. In addition, our model inputs different aspects separately. For the actual situation that a text contains multiple aspects, using the attention mechanism to predict the sentiment polarities of more than one aspect simultaneously is promising.

Acknowledgments

This research is supported by the National Nature Science Foundation of China under Grants 61472258.

References

[1] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
[2] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2016) 102–107.
[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, SemEval-2014 task 4: Aspect based sentiment analysis, Publishing (2014) 27–35.
[4] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster, L. Tounsi, DCU: Aspect-based polarity classification for SemEval task 4, Publishing (2014) 223–229.
[5] T.H. Nguyen, K. Shirai, PhraseRNN: Phrase recursive neural network for aspect-based sentiment analysis, Publishing (2015) 2509–2514.
[6] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, Publishing (2016) 606–615.
[7] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, Recurrent models of visual attention, Publishing (2014) 2204–2212.
[8] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[9] F.G. Cozman, I. Cohen, Unlabeled data Can degrade classification performance of generative classifiers, Publishing (2002) 327–331.
[10] A. Hussain, E. Cambria, Semi-supervised learning for big social data analysis, Neurocomputing 275 (2018) 1662–1673.

[11] D.P. Kingma, M. Welling, 2013. Auto-Encoding Variational Bayes. CoRR, abs/1312.6114.
[12] D.P. Kingma, S. Mohamed, D.J. Rezende, M. Welling, Semi-supervised learning with deep generative models, Publishing (2014) 3581–3589.
[13] S.R. Bowman, L. Vilnis, O. Vinyals, A.M. Dai, R. Józefowicz, S. Bengio, Generating sentences from a continuous space, Publishing (2016) 10–21.
[14] W. Xu, H. Sun, C. Deng, Y. Tan, Variational autoencoder for semi-supervised text classification, in: National Conference on Artificial Intelligence, 2017, pp. 3358–3364.
[15] T. Mikolov, K. Chen, G. Corrado, J. Dean, 2013. Efficient Estimation of Word Representations in Vector Space. CoRR, abs/1301.3781.
[16] Y. Liu, Z. Liu, T.-S. Chua, M. Sun, Topical word embeddings, Publishing (2015) 2418–2424.
[17] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[18] C. Lin, Y. He, Joint sentiment/topic model for sentiment analysis, Publishing (2009) 375–384.
[19] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, Publishing (2018) 5876–5883.
[20] L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao, Target-dependent twitter sentiment classification, Publishing (2011) 151–160.
[21] N. Kaji, M. Kitsuregawa, Building lexicon for sentiment analysis from massive collection of HTML documents, Publishing (2007) 1075–1083.
[22] S. Mohammad, S. Kiritchenko, X. Zhu, NRC-CaNada: Building the state-of-the-art in sentiment analysis of tweets, Publishing 321-327 (2013).
[23] Q. Qian, B. Tian, M. Huang, Y. Liu, X. Zhu, X. Zhu, Learning tag embeddings and tag-specific composition functions in recursive neural network, Publishing (2015) 1365–1374.
[24] R. Socher, J. Pennington, E.H. Huang, A.Y. Ng, C.D. Manning, Semi-supervised recursive autoencoders for predicting sentiment distributions, Publishing (2011) 151–161.
[25] K.S. Tai, R. Socher, C.D. Manning, Improved semantic representations from tree-structured long short-term memory networks, Publishing (2015) 1556–1566.
[26] D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for target-dependent sentiment classification, Publishing (2016) 3298–3307.
[27] D. Tang, B. Qin, T. Liu, Aspect level sentiment classification with deep memory network, Publishing (2016) 214–224.
[28] M. Zhang, Y. Zhang, D.-T. Vo, Gated neural networks for targeted sentiment analysis, Publishing (2016) 3087–3093.
[29] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, Publishing (2017) 4068–4074.
[30] G. Lu, X. Zhao, Y. Jian, W. Yang, L. Bo, Multi-task learning using variational auto-encoder for sentiment classification, Pattern Recognit. Lett. (2018).
[31] Y. Bengio, R. Ducharme, P. Vincent, C. Janvin, A neural probabilistic language model, J. Mach. Learn. Res. 3 (2003) 1137–1155.
[32] E.H. Huang, R. Socher, C.D. Manning, A.Y. Ng, Improving word representations via global context and multiple word prototypes, Publishing (2012) 873–882.
[33] J. Reisinger, R.J. Mooney, Multi-prototype vector-space models of word meaning, Publishing (2010) 109–117.
[34] D.P. Kingma, J. Ba, 2014. Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.
[35] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O.D. Clercq, V. Hoste, M. Apidianaki, X. Tannier, N.V. Loukachevitch, E. Kotelnikov, N. Bel, S.M.J. Zafra, G. Eryigit, Semeval-2016 task 5: Aspect based sentiment analysis, Publishing (2016) 19–30.
[36] W. Deng, Y. Wang, Z. Liu, H. Cheng, Y. Xue, Hemi: A toolkit for illustrating heatmaps, PLOS ONE 9 (2014).