Aspect-based Sentiment Analysis using Supervised Restricted Boltzmann Machines

No Author Given

No Institute Given

Abstract. Recent years, many studies have addressed problems in sentiment analysis in different levels, and building aspect-based methods has become a central issue for deep opinion mining. However, to the best of our knowledge, previous studies need to use many modules in order to extract aspect-sentiment word pairs, then predict the sentiment polarity. In this paper, we use Restricted Boltzmann Machines in supervised way to build the unique model which not only extracts aspect terms appeared and classify them into respective categories, but also completes the sentiment polarity prediction task. The experiments show that the method we use in categories classification and polarity prediction tasks achieving accuracy scores of up to xyz% and abc% respectively.

bỏ

two separated

combination with ...

a joined model:

better than other state-of-the-art approaches

Keywords: Aspect-based, Sentiment Analysis, Opinion Mining, Restricted Boltzmann Machine, Supervised Learning

1 Introduction

SA là gì, ví dụ, ứng dụng

Importance of Sentiment Analysis

With the development of opinionated user-generated review sites, many customers can write reviews and express their opinions about the products, services, people or places on the websites. Each review contains many aspects/features about the products. For example, consider some reviews in restaurant domain, we may have features such like view (of the restaurant), cost, dishes, etc. Therefore, new customers may find it difficult to explore the large number of reviews in order to make right decision, since many people have different aims to buy, which base on different aspects of the product. So Sentiment Analysis can play an important role, not only in addressing this issue but also help companies improve their products based on reviews from users.

Aspect-based SA

ABSA gồm ...
ABSA quan trọng ...
ABSA cần 2 modules:
AE và SA

Aspect-based Sentiment Analysis (ABSA) has received much attention in recent years. ABSA system aims to analyze opinionated texts, such as opinions, sentiments by extracting the aspect-term appeared in the sentences, classifying them into two classes (positive or negative). ABSA system has many possible uses in the Natural Language Processing field, there has been an explosive growth of both research in academia and applications in the industry. It helps people make the right decisions thanks to the fine-grained analysis on each aspect of products. In previous works, they proposed combining two or more modules to complete the ABSA model [1–3]. For instance, we can use one for extracting aspect-sentiment word pairs, one for sentiment prediction, another optional module is to summarize the opinions.

name of authors

phần này em nói họ có hạn chế gì nhưng trong paper em có giải quyết các hạn chế đó không?

Some work has been investigated to extract the candidate product feature opinion pairs. Kumar and Raghuveer (2012) [3] use the rules on the typed dependency tree, then they use lexicons to classify and generate a summary of the product. Such approach, however, have failed to eliminate the redundant pairs, which are generated by wrong rules. In Toh and Wang (2014) [4], they build a Conditional Random Field (CRF) based classifier for Aspect Term Extraction and a linear classifier for Aspect Term Polarity Classification, but there is no method to extract the opinion words in the sentences. Unsupervised method such as Latent Dirichlet Allocation (Blei et al., 2003) [5] is also used to extract and group corresponding representative words into categories. However, inaccurate approximations of the distribution over topics can cause low accuracy in performing.

vd chỗ này có thể nói là các pp ABC bị hạn chế là phải dùng 2 modules riêng lẻ, hoặc chỉ giải quyết 1 trong 2 vấn đề

bỏ (2015)

Hence, Wang et al. (2015) [6] adapted the Sentiment-Aspect Extraction based on Restricted Boltzmann Machines (SERBM) to overcome this problem. They use unsupervised version of RBM to group reviews in categories and also classify opinion. Three different types of hidden units are used to represent aspects, sentiments, and background words in this model, respectively. However, much uncertainty still exists about the defined hidden units in hidden layer, unsupervised method can just cluster reviews into categories and can not fix the unit for desired category.

nói rõ họ giải quyết hạn chế trên = cách nào (= cách dùng SERBM)
Tuy nhiên, SERBM có một số hạn chế ...

In this paper, we implement supervised RBM to extract aspect term appeared in sentences. In this two-layer structure model, we do not use hidden layer as ouput layer but the dependencies between the components of observations in visible layer. This RBM model has the ability to jointly modeling aspect and sentiment information together. The results obtained based on the dataset of reviews in restaurant domain, which widely adopted by previous work [7–9] (Ganu et al., 2009; Brody and Elhadad, 2010; Zhao et al., 2010).

Trong paper này chúng tôi đề xuất XYZ để ...
Lí do chúng tôi đề xuất XYZ là ...

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 introduces RBM. Section 4 describes our approach to classify reviews into aspect categories and predict sentiment polarity of the reviews. Experimental results are presented in Section 5. Finally, Section 6 concludes the paper and discusses future work.

OK

2 Related work

For the supervised learning approach, to determine the dependency and the other relevant information in the sentence, Wei and Gulla (2010) propose a hierarchical classification model. Jiang et al. (2011), Boiy and Moens (2009) use dependency parser to generate a set of aspect dependent features for classification, which weighs each feature based on the position of the feature relative to the target aspect in the parse tree. Other supervised learning models have been widely used as Hidden Markov Models, SVMs, Conditional Random Fields to extract aspects which appear in sentences (Jin et al., 2009; Choi and Cardie, 2010; Jakob and Gurevych, 2010; Kobayashi et al., 2007).

Supervised learning is dependent on the training data and has difficulty to scale up to a large number of application domains. A model or classifier trained from labeled data in one domain often performs poorly in another domain [10]. Hence, unsupervised methods are often adopted to avoid this issue. Several recent studies investigate statistical topic models which are unsupervised learning methods that assumes each document consists of a mixture of topics. Specifically, Latent Dirichlet Allocation (LDA), Multi-Grain LDA model (Titov and McDonald, 2008) are used to model and extract topics from document collections. The two-step approach to detect aspect-specific opinion words (Brody and Elhadad, 2010) includes identifying aspects using topic models and identifying aspect-specific sentiment words by considering adjectives only. The joint sentiment/topic model (JST) is also used to separate aspect words and sentiment words (Lin and He, 2009). The topic-sentiment mixture model (Mei et al., 2007) based on three separated models with the help of some external training data is proposed to extract aspect and sentiment words. The SERBM model (Wang et al., 2015) also jointly address these two tasks in an unsupervised setting.

Other methods also use lexicon-based approach have been proposed (Hiu and Liu, 2004; Ding et al., 2008) which has been shown to perform quite well in a large number of domains. They use a sentiment lexicon, expressions, rules of opinions, and the sentence parse tree to help classify the sentiment orientation on each aspect appeared in a review. They also consider sentiment shifters (i.e. not, none, nobody, etc). However, these rule-based methods have a shortcoming in processing complex documents where the aspect is hidden in sentence, and failing to group extracted aspect terms into categories.

3 Restricted Boltzmann Machines

phần này chỉ là 1 đoạn ngắn tóm tắt về RBM, gom chung vào trong phần sau luôn (để là 3.1)

3.1 Energy-based models

3. Proposed method

3.1 RBM

Energy-based models associate a scalar energy to each configuration of the variables of interest. Learning phrase modifies that energy function so that it will has desirable properties. Energy-based probabilistic models define a probability distribution through an energy function, as follows:

$$p(x) = \frac{e^{-E(x)}}{Z}. \tag{1}$$

where $E(x)$ is the energy function and Z is called the partition function defined by

$$Z = \sum_{x} e^{-E(x)} \tag{2}$$

Stochastic gradient descent can be used to perform learning phrase for energy-based model.

3.2 Restricted Boltzmann Machines - Two-layer network

Boltzmann Machines (BMs) are a particular form of log-linear Markov Random Field (MRF), i.e., for which the energy function is linear in its free parameters. RBM further restrict BMs to those without visible-visible and hidden-hidden connections, or there is no connection between neurons in the same layer. A graphical depiction of an RBM is shown below.

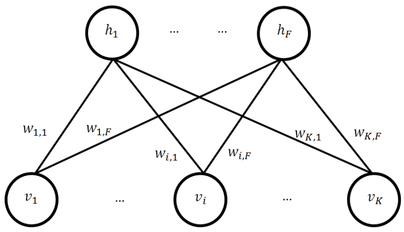

Fig. 1. The network graph of an RBM model with K visible and F hidden units

As shown in Figure 1, RBM model is a two-layer neural network which contains one visible layer and one hidden layer. Visible layer is constituted by visible units correspond to the components of an observation (e.g., one visible unit for each word in a input document). The hidden layer composed of hidden units which model dependencies between the components of observations (e.g., dependencies between the words in the document). To make them powerful enough to represent complicated distributions, we can increase the numbers of hidden variables (also called hidden units) in order to increase the modeling capacity of the RBM.

In this context, the visible layer $\mathbf{v}$ is represented as a $D \times K$ matrix, where D is the document length, or the number of sentences in reviews, and K is the dictionary size. More precisely, RBM model takes input as sentences in review, each visible unit correspond to each word in sentence. If a sentence i-th has word k-th in dictionary, we set $\mathbf{v}_i^k = 1$. The hidden layer can be expressed as a vector $h \in 0, 1^F$, where F is the dimension of the vector, or number of hidden units. Hidden units of RBM model corresponds to the relation between visible units.

The energy function of the model can be defined as

$$E(\mathbf{v}, h) = -\sum_{i=1}^{D}\sum_{j=1}^{F}\sum_{k=1}^{K} W_{ij}^k h_j v_i^k - \sum_{i=1}^{D}\sum_{k=1}^{K} v_i^k b_i^k - \sum_{j=1}^{F} h_j a_j. \tag{3}$$

where W_{ij}^k is the connection weight between i-th visible unit of value k to the j-th hidden node, b_i^k specifies a bias of visible unit v_i^k, and a_j specifies to a bias of h_j. The conditional probabilities from hidden to visible layer and from visible to the hidden layer we can obtain from (1) and (3) are probabilistic version of the usual neuron activation function.

$$P(v_i^k = 1|h) = sigm(b_i^k + \sum_{j=1}^{F} h_j W_{ij}^k) \quad (4)$$

$$P(h_j = 1|\mathbf{v}) = sigm(a_j + \sum_{i=1}^{D} \sum_{k=1}^{K} v_i^k W_{ij}^k) \quad (5)$$

where $sigm(x) = \frac{1}{(1+e^{-x})}$ is the logistic function.

Phần này mô tả em áp dụng RBM cụ thể như thế nào

4 Our Aspect-based Sentiment Analysis model

RBM model as a generative stochastic artificial neural network is treated like a model in the field of deep learning. RBM have found applications in dimensionality reduction, classification, collaborative filtering, feature learning, especially applications in the field of image processing. Recently, RBM is developed and applied in the field of Natural Language Processing, including topic modeling, sentiment analysis. It can be trained in supervised or unsupervised ways, which depends on the task.

4.1 Latent topics

For the customer restaurant reviews analyzing task, every word in a review may mention a certain aspect (e.g. “delicious” corresponds to *food* aspect), or a certain opinion (e.g. “good” is about positive sentiment, while “bad” is about negative sentiment). In addition, there are other words do not mention about aspect or sentiment, they can be removed during preprocessing phrase. The latent topics in a review are considered as the factors which generate aspect and sentiment words within that review. To technically illustrate this, hidden layer which includes hidden units in RBM would contain information of the latent topics. In the generation process, they produce the values of visible units, which are also the information of the aspect and sentiment words in reviews.

4.2 Structure

In previous approach, they have used unsupervised learning to build model of RBM [6]. Although this model can help to overcome the limitations of handlabeled training data, but there still exist certain shortcomings. First, this unsupervised RBM model fixes the units to serve a function in the hidden layer. In particular, they fix hidden units 0-6 to represent the target aspects *Food*, *Staff*, *Ambience*, *Price*, *Ambience*, *Miscellaneous*, and *Other Aspects*, respectively. This may reduce the accuracy of the classification, since unsupervised learning model does not fix the categories that it will cluster into the output units before the classification phrase is accomplished. There is no information in the visible layers for the connected hidden units to get which unit for which aspect. Second, each document which put into learning model is transformed into a $K \times D$ matrix $\mathbf{v}$, where K is the dictionary size, and D is the document length. If visible unit i in $\mathbf{v}$ takes the k-th value, v_i^k is set to 1. This makes the dictionary size will be very large if training reviews become abundant. From that, the feature vectors which are fed into the model will be sparse, leading to the decrease of the model's accuracy .

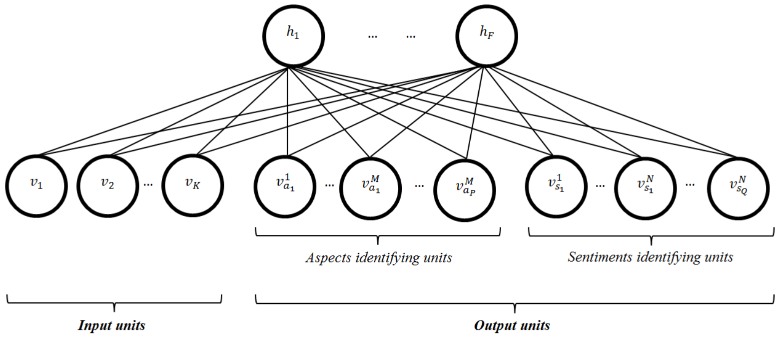

Fig. 2. The network graph of our Supervised Sentiment-Aspect Extraction RBM model

To overcome these problems, we propose using supervised RBM model which is illustrated in Figure 2. The ***Output units*** including the units represent the aspects and sentiment orientations of the reviews, which called *Aspects identifying units* and *Sentiments identifying units* respectively. These units will be put together with ***Input units*** in the visible layers, instead of being placed in hidden layers. Meanwhile, the units in the hidden layer will take the role of the relationship between the units in the visible layers. Each input unit in the visible layer is a component of the feature vector, created by the word representation model instead of using the bag of words model as previous approach. This helps reduce the dimensionality of the input matrix while keeping documents' semantics.

Compared to standard RBMs, the first difference of this model is the visible layer. Apart from the input units which are the feature vector made from reviews, there are also aspect and sentiment identifying units that represent the output component of the model. Suppose in the training set, there are reviews talking about P aspects and have Q sentiment orientations in total. For detail, set of aspects is $A = \{a_1, a_2, ..., a_P\}$, set of sentiment orientations is $S = \{s_1, s_2, ..., s_Q\}$.

For each aspect, the model will set M units to express the nature of the aspect. For example, if the review mentioned aspect a_i of A, the units from $v_{a_i}^1$ to $v_{a_i}^M$ would be set to 1, and 0 otherwise. The sentiment orientation of the reviews bears comparison with aspect. The model will set N units to express the nature of the sentiment. If the sentiment polarity is s_j included in S, the units from $v_{s_j}^1$ to $v_{s_j}^N$ would be set to 1, and 0 otherwise.

This structure which builds up the strength of the model can solve two tasks simultaneously, including aspect identification and sentiment classification. They are reflected in the aspect and sentiment identifying units which play important roles in the sampling process of RBM, while weight values of connecting edges between the units contain semantic information of the review. In addition, the fixed dimension of the feature vectors which do not depend on the number of vocabulary not only helps the model prevent the occurrence of decreasing speed and accuracy, but also solves the semantic problem in the review.

4.3 Training

Contrastive Divergence (CD), also called Approximate Gradient Descent, is the way that help RBM learns the connection weights in the network. CD has two phrases, which are Positive phrase and Negative phrase, in each phrase we compute the positive and negative value by the multiplication of visible and hidden units, then update the connection weight based on the subtraction from positive value of negative value. For detail, each epoch of CD can be expressed in several steps below:

Step 1. Update the states of the hidden units using the logistic activation rule described in (5). For the j-th hidden unit, compute its activation energy and set its state to 1 with corresponding probability.

Step 2. For each connection edge e_{ij}, get the value from positive phrase by computing $pos(e_{ij}) = v_i h_j$

Step 3. Reconstruct the visible units in a similar manner by using the logistic activation rule described in (4). For the i-th visible unit, compute its activation energy and set its state to 1 with corresponding probability. Then do ***Step 1*** to update the hidden units again.

Step 4. For each connection edge e_{ij}, get the value from negative phrase by computing $neg(e_{ij}) = v_i h_j$

Step 5. Update the connection weights W_{ij} by the equation

$$W_{ij} = W_{ij} + \mathbf{lr}(pos(e_{ij}) - neg(e_{ij})) \tag{6}$$

where **lr** is a learning rate.

After m steps of transfer between visible and hidden layers in a CD-m run of the above steps, values of the hidden units will reflect the relationship between the visible units in the model. The connection weight matrix between the two layers helps hidden layer generate visible units which include input units, aspect identifying units and also sentiment identifying units.

5 Experiments

In this section, we present sequence of experiments to evaluate the performance of our model on the aspect identification and sentiment classification tasks.

5.1 Data

Mô tả data, mô tả cách đánh giá, ...

To evaluate our model performance, we used review restaurant dataset which is used widely in previous work [7–9]. Data contain reviews about the restaurant with 1,644,923 tokens and 52.574 documents in total. Documents in this dataset are annotated with one or more labels from a gold standard label set S = {Food, Staff, Ambience, Price, Anecdote, Miscellaneous}.

5.2 Aspect Extraction

Experimental Setup

Evaluation

5.3 Sentiment Classification

Experimental Setup

Evaluation

6 Conclusion and future work

References

1. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of KDD*, 2004, pp. 168–177.
2. S. Blair-goldensohn, T. Neylon, K. Hannan, G. A. Reis, R. Mcdonald, and J. Reynar, "Building a sentiment summarizer for local service reviews," in *WWW Workshop on NLP in the Information Explosion Era*, 2008.
3. R. kumar V and K. Raghuveer, "Web user opinion analysis for product features extraction and opinion summarization," *International Journal of Web and Semantic Technology*, pp. 69–82, 2012.
4. Z. Toh and W. Wang, "Dlirec: Aspect term extraction and term polarity classification system," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, 2014, pp. 235–240.
5. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research 3*, pp. 993–1022, 2003.
6. L. Wang, K. Liu, Z. Cao, J. Zhao, and G. de Melo, "Sentiment-aspect extraction based on Restricted Boltzmann Machines," in *Proceedings of the 53rd Annual Meeting of the ACL*, 2015, pp. 616–625.
7. G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: Improving rating predictions using review text content," in *Proceedings of WebDB 2009*, 2009, pp. 1–6.
8. S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," in *Proceedings of NAACL-HLT 2010*, 2010, pp. 804–812.
9. W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in *Proceedings of EMNLP 2010*, 2010, pp. 56–65.
10. B. Liu, *Sentiment Analysis and Opinion Mining.* Morgan and Claypool, 2012.