Attention Based GRU Network for Domain Adaptation in Sentiment Classification

Mingfeng Pu and Yi Ge

ABSTRACT

Domain adaptation in sentiment classification, as a worthwhile research area, aims to use the available labeled data of a source domain to classify the unlabeled data in the target domain. This approach saves costs and efforts and also improves the efficiency of sentiment classification. This paper proposes the attention based GRU network for domain adaptation in sentiment classification (AGN). This algorithm learns the semantic meanings of the text based on GRU or Bi-GRU, and makes hierarchical mapping from words to sentences and then to the document. The Attention layer is added to capture the core emotional words and sentences. Then, the adversarial idea is brought in, that is, auxiliary domain classification task and sentiment classification task are performed to jointly learn feature representations. Review datasets of five domains on the Amazon online platform are used, and an experiment is performed with 20 pairs of data. The result indicates that the proposed algorithms are able to improve the accuracy of sentiment classification. Finally, the AGN algorithms are interpreted by visualizing the Attention layer.

1. INTRODUCTION

Sentiment classification is a subset of sentiment analysis, thus a binary classification task. It is highly sensitive to the change of domains. Regardless of the language of the text, as in Chinese or in English, different domains have specific sentiment descriptors. Machine learning-related technologies have been
sophisticated enough to solve classification problems, but traditional machine learning algorithms require training data and testing data to meet independent and identical distribution (IID). However, this condition is difficult to be satisfied in the actual application scenario of sentiment classification. Besides, against the background of big data, while massive amounts of data are easily accessible, high-quality labeled data are far from enough. Labeling data is time-consuming and laborious. More importantly, fresh unlabeled domain data keeps emerging. Therefore, it is necessary to use the available labeled data in one domain to do sentiment classification in another domain, thereby saving costs and efforts. Then there arises a problem—domain adaptation in sentiment classification.


Based on the GRU network, attention mechanism and adversarial idea in deep learning, we attempt to utilize neural network to learn the “sharing” in and “difference” between the source domain and the target domain, to obtain the representations of hidden features. Thereupon, a domain adaptation algorithm is proposed, namely, Attention based GRU Network for domain adaptation in sentiment classification (AGN).

2. METHODS AND DATA

2.1 The Method

The AGN or ABGN algorithm structure is shown in Figure 1, which three components are feature mapping, sentiment classifier and domain classifier. The feature mapping is formulated as $G_f(x; \theta_f)$. Firstly, the AGN algorithm uses the GRU structure to capture the contextual information, while the ABGN attempts to locate the contextual information through the bidirectional GRU (Bi-GRU) structure. Another feature of the AGN and ABGN algorithms is to hierarchically map word representations to sentence representations, and then, to document representations. Hierarchical mapping is able to produce a high-level document representation accurately and in a detailed manner. Secondly, the essence of
sentiment classification task is to obtain the sentiment value of the document by aligning sentiment words in a non-linear manner. In the alignment, each sentence, and even each word, makes different contributions to the overall sentiment of the document. Therefore, Attention mechanism in AGN and ABGN is to address the different contributions of words and sentences, capture core words and sentences, and enhance GRU and Bi-GRU’s mapping ability of the document features. Finally, the model in the feature mapping stage \( G_f(x; \theta_f) \rightarrow v_d \) is produced.

Based on the feature mapping, the sentiment classifier yields the positive and negative sentiment probability by using a Softmax layer, which can be regarded as normalization operation. The function of auxiliary domain classification, grounded on the feature mapping, is to obtain classification probability of the domain through the GRL layer [7] and the Softmax layer. The purpose of introducing GRL layer is to train feature representations to be applicable to two domains, and the AGN and ABGN algorithms incorporates the adversarial idea, that is, auxiliary domain classification tasks are used to render the trained feature representations unable to decide whether the sample is from the source domain or the target domain.

The encounter of sentiment classification task and domain classification task results in a minimized sum of loss functions, and in a feature mapping space \( G_f \) most suited for the sentiment classification of two domains. The sum of the loss functions of sentiment classification and domain classification is shown in Eq.1. The loss function of the sentiment classification is one of cross entropy, as is shown in Eq.2, where \( X_{s,l}^i \) stands for the labeled text data of the source domain, \( y_i \) for the predicted sentiment inclination, and \( y_i' \) for the true label of the text sentiment. The loss function of the domain classification is also cross entropy, as is shown in Eq.3, where \( X_{s,t} \) symbolizes the text data of both source domain and target domain.

\[
\mathcal{L} = \mathcal{L}_s (G_f(X_{s,l}; \theta_f)) + \mathcal{L}_d (G_f(X_{s,t}; \theta_f))
\]

\[
\mathcal{L}_s \left( G_f(X_{s,l}; \theta_f) \right) = -\frac{1}{N_s^l} \sum_{i=0}^{N_s^l} (y_i \ln y_i' + (1 - y_i) \ln(1 - y_i'))
\]

\[
\mathcal{L}_d \left( G_f(X_{s,t}; \theta_f) \right) = -\frac{1}{N_s + N_t} \sum_{i=0}^{N_s + N_t} (d_i \ln d_i' + (1 - d_i) \ln(1 - d_i'))
\]
2.2 Data Set Description

The dataset of this experiment is from Amazon's real reviews [9], which has been extensively used in the domain adaptation of sentiment classification. This experiment covers review texts in five domains: DVD (D), books (B), electronics (E), kitchen (K), and video (V). In order to verify the validity of the domain adaptation algorithms, the five domains are paired among them and 20 pairs are obtained for the experiment. Each domain has 6,000 sentiment-labeled sample reviews, including 3000 positive and 3000 negative, and also unlabeled reviews (see Table 1).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Training set</th>
<th>Test set</th>
<th>Unlabeled data</th>
<th>Positive ratio in training set</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVD</td>
<td>5600</td>
<td>400</td>
<td>11843</td>
<td>50%</td>
</tr>
<tr>
<td>Books</td>
<td>5600</td>
<td>400</td>
<td>9750</td>
<td>50%</td>
</tr>
<tr>
<td>Electronics</td>
<td>5600</td>
<td>400</td>
<td>17009</td>
<td>50%</td>
</tr>
<tr>
<td>Kitchen</td>
<td>5600</td>
<td>400</td>
<td>13856</td>
<td>50%</td>
</tr>
<tr>
<td>Video</td>
<td>5600</td>
<td>400</td>
<td>15000</td>
<td>50%</td>
</tr>
</tbody>
</table>

3. ANALYSIS OF EXPERIMENTAL RESULTS

3.1 Comparison Experiments and The Results of Different Methods

The experiment compares the proposed algorithms with three other methods—Source-only, SFA and DANN—to evaluate the validity of the proposed algorithm. The Source-only method trains with data only from the source domain. As this paper aims at solving the problem of domain adaptation in sentiment classification, the Source-only is a necessary reference, for it carries out sentiment classification by using the GBDT model, and its input is the one-hot code produced when the source domain data have been cleansed and pretreated. The SFA algorithm [5] is also our consideration, for it embodies traditional machine learning and graph theory. Because AGN and ABGN algorithms are based on deep learning and the GRL layer of the DANN, this experiment takes the DANN [7] as the third algorithm to be compared with. The accuracy of each method is shown in Table 2.
Accuracy is the evaluation indicator in this experiment. The average accuracy values of SFA, and DANN, AGN and ABGN all score higher than that of Source-only, with AGN and ABGN 12% and 12.7% higher respectively, which testifies the necessity of domain adaptation in sentiment classification.

The accuracy values of deep learning algorithms are higher than that of SFA, which proves that the feature abstract expression ability of deep learning is relatively better. AGN and ABGN produce average accuracy values 3.7% and 4.4% higher respectively than that of DANN; besides, each of the 20 pairs score higher accuracy values by AGN and ABGN. This demonstrates that the network model structure designed to extract text sentiment features is effective. AGN produces higher accuracy than ABGN in just 4 pairs (D → B, B → E, B → K, V → D), with the difference being a few thousandths, an evidence that Bi-GRU can transmit and retain the context information better than the GRU. It can be seen that the transfer from the domain E to the domain K has achieved the best accuracy – 0.8865. Conversely, the transfer accuracy from the domain K to the domain E is also desirable. It is our experience that kitchen wares and electronics products have many things in common, and that is why the experiment can produce satisfactory results. Likewise, DVD and video products also have a lot in common, and the transfer accuracy from the domain D to the domain V is 0.8775, ranking the second. One of the most undesirable accuracy produced by Source-only is that of E → V, whereas satisfactory results—0.8113 and 0.8257—are obtained by the AGN and ABGN.
becomes clear that in two largely different domains, our algorithms can still find hidden feature space and yield better sentiment classification results.

3.2 Visualization of ABGN Attention layer

The following experiment is to demonstrate the interpretability of the AGN and ABGN by visualizing the weights acquired at the Attention layer. Based on ABGN, as is shown in Figure 2, examples (two positive and two negative sentiment texts) are taken from $K \rightarrow B$. The different shades of yellow on the left margins indicate the importance of the sentence in the document. The red and blue in the subgraph indicate how important a word is in conveying positive or negative sentiment. As is shown on the upper left quarter, the first sentence is highlighted with the deepest yellow on the margin, indicating that it has the biggest Attention weight; or rather, it is the most important sentence in the document. Actually, the first sentence, whose first word "fantastic" is underlined with the deepest red, presents the author’s opinion. The visualization of Attention weight makes it clear that "fantastic" exerts an impressive impact on the sentiment inclination of the whole document. The other descriptors like “good”, "very" and “surprisingly” also carry positively powerful sentiment, hence the effectiveness of the proposed algorithms in locating core sentiment words and sentences.

![Figure 2. Visualization of Attention layer weights in the domain adaptation $K \rightarrow B$.](image_url)

The penultimate sentence of the negative sentiment example in the target domain B shows that when there is repeated information, the Attention weight
gradually decreases, another proof that the Attention mechanism focuses on the feature of certain information. An overall analysis of the document reveals that AGN and ABGN models also highlight adversatives and negators, such as "but", "not", "don't", and "isn't" that have a big importance weight. Actually, it is our daily experience that such words do affect the sentiment inclination of the document. In a nutshell, AGN and ABGN are capable of locating core sentiment words and sentences by the Attention layer when learning the sentiment of document.

4. CONCLUSIONS

In conclusion, by setting parameters for the proposed AGN and ABGN and making comparison between the results of these two algorithms and others (Source-only, SFA and DANN), this paper has demonstrated the efficacy of AGN and ABGN. AGN and ABGN have advantages over DANN, for they yield satisfactory results when there are huge differences between domains. By visualizing the weight of the Attention layer, AGN and ABGN can locate the core sentiment words and sentences through the Attention layer. It is true that the accuracy of AGN and ABGN is superior to other algorithms in the 20 experiment pairs, but AGN and ABGN algorithms also have some drawbacks, such as slow training and the necessity of hardware support like GPU.

REFERENCES

