Cold-Start Aware Deep Memory Network for Multi-Entity Aspect-Based Sentiment Analysis

Kaisong Song\textsuperscript{1}, Wei Gao\textsuperscript{2}, Lujun Zhao\textsuperscript{1}, Jun Lin\textsuperscript{1}, Changlong Sun\textsuperscript{1} and Xiaozhong Liu\textsuperscript{3}

\textsuperscript{1}Alibaba Group, China
\textsuperscript{2}Victoria University of Wellington, New Zealand
\textsuperscript{3}Indiana University Bloomington, USA

kaisong.sks@alibaba-inc.com, wei.gao@vuw.ac.nz, lujun.zlj@alibaba-inc.com, linjun.lj@alibaba-inc.com, changlong.scl@taobao.com, liu237@indiana.edu

Abstract

Various types of target information have been considered in aspect-based sentiment analysis, such as entities and aspects. Existing research has realized the importance of targets and developed methods with the goal of precisely modeling their contexts via generating target-specific representations. However, all these methods ignore that these representations cannot be learned well due to the lack of sufficient human-annotated target-related reviews, which leads to the data sparsity challenge, a.k.a. cold-start problem here. In this paper, we focus on a more general multi-entity aspect-based sentiment analysis (ME-ABSA) task which aims at identifying the sentiment polarity of different aspects of multiple entities in their context. Faced with severe cold-start scenario, we develop a novel and extensible deep memory network framework with cold-start aware computational layers which use frequency-guided attention mechanism to accentuate on the most related targets, and then compose their representations into a complementary vector for enhancing the representations of cold-start entities and aspects. To verify the effectiveness of the framework, we instantiate it with a concrete context encoding method and then apply the model to the ME-ABSA task. Experimental results conducted on two public datasets demonstrate that the proposed approach outperforms state-of-the-art baselines on ME-ABSA task.

1 Introduction

With the rapid growth of popular e-commerce websites such as Taobao.com and review websites such as Yelp.com, users can conveniently express their opinions about various aspects of all sorts of products in real time. The enormous subjective texts generated by consumers have fueled the merchants with torrents of subjective texts. Mining consumers’ sentiment orientations from such vast volume of subjective texts gives insights on consumer needs as well as their product experience. Sentiment analysis has drawn close attention from research communities due to its importance to many practical applications in a wide range of fields.

Aspect-based sentiment analysis (ABSA) has been intensively studied as a fine-grained extension of traditional sentiment analysis task, which aims to detect sentiment polarities (e.g., positive, negative or neutral) of targets in their context [Pang et al., 2002]. The targets can be entities (e.g., product names) or aspects pre-defined as attribute categories (e.g., product features). Existing studies develop various methods with the goal of precisely modeling targets’ contexts via generating target-specific representations. Wang et al. [2016] introduce an aspect-specific attention mechanism to attend to different parts of the sentence. Ma et al. [2017] use an interactive attention mechanism to alternately learn attentions in the review and entity, and generate their representations separately. However, these studies ignore that many expressions are generally related to both entity and aspect for a certain sentiment. For example, in an experience-sharing post about paper diaper brands “I just love Unicharm, very dry, but expensive. Kao not good, but cheap.”, there exist two entities: Unicharm and Kao, and two aspects: drying and price. Given the same aspect “price”, the review expresses opposite sentiment polarities, i.e., “expensive” and “cheap”, towards different entities “Unicharm” and “Kao”, and ABSA has no way to distinguish the sentiment polarities for “price” without differentiating the entities, the same for “drying”. Obviously, ABSA cannot be directly applied in such more general cases.

Recent work has defined a new task called Multi-Entity Aspect-Based Sentiment Analysis (ME-ABSA) [Yang et al., 2018]; “Given entities as well as aspects mentioned in the text, the goal is to predict sentiment polarity towards each entity-aspect combination”. In the paper diaper example above, ME-ABSA will predict the results for all the combinations in the post: [(Unicharm, drying), positive], [(Unicharm, price), negative], [(Kao, drying), negative] and [(Kao, price), positive]. Obviously, ABSA is a special case of ME-ABSA with the number of entities limited to one. In their work, Yang et al. [2018] proposed a memory network approach CEA, which updates the entity-specific vector and aspect-specific vector based on simple addition with the context memory, and achieves state-of-the-art performance on the ME-ABSA task. However, similar to ABSA methods, they ignore the existing cold-start problem, which makes CEA more problematic than helpful in reality. According to our statistics, 95.23% of (entity, aspect) combinations appear less than 6 times and occupy 39.45% reviews from SemEval-16 [Pontiki et al., 2016].
Neural networks usually require a large amount of well-annotated training data for producing reasonably good results. Unfortunately, the human-annotated data with target-level labels are scare and costly to obtain, which makes target representations cannot be learned well due to the lack of sufficient target-related reviews as training set [Wang et al., 2018]. This is a data sparsity challenge, a.k.a. cold-start issue here. Existing sentiment analysis approaches have no way to learn the representations of the cold-start targets well in the settings of ABSA and ME-ABSA. In recommender systems [Zhang et al., 2014], introducing external contexts can help mitigate such data sparsity or cold-start problem, however such kind of context is not always available and the resulting models are hard to generalize. With the data at hand, we observe that similar targets (e.g., brands “Unicharm” and “Kao”) usually share the similar characteristics, such as contextual word usages, target attributes and target names, which will lead to similar target representations. Thus, similar targets can be considered as helpful complementary information, which provides the possibility to mitigate cold-start problem.

To deal with the issue, we propose a novel and extensible Cold-start Aware Deep Memory Network (CADMN) framework for the more general ME-ABSA task. We motivate the basic idea of our cold-start solution based on the hypothesis that if the model does not have enough information to create a good representation of entity/aspect, then we try to enhance it with a representation derived from other entities/aspects which are mostly related to the cold-start targets. The specific challenge is how to automatically find and concentrate on the most related targets that can contribute more than others to the representation of the cold-start targets, for which we adopt a target attention mechanism tailored for ME-ABSA task. Overall, our contributions are summarized as follows:

- Considering the cold-start targets in ME-ABSA task, we propose a novel CADMN framework with cold-start aware computational layers, which use frequency-guided attention mechanism to focus on the most related targets, and then compose their representations into a complementary vector for improving the representations of cold-start entities and aspects.
- We instantiate the CADMN framework as a sentiment analysis model by introducing a concrete context modeling method, and then apply it for the ME-ABSA task.
- Extensive experimental results on a benchmark English review dataset and a public Chinese review dataset show that our CADMN-based model outperforms state-of-the-art comparative methods for the ME-ABSA task.

2 Related Work

Aspect-based sentiment analysis (ABSA) is an important sub-task of sentiment analysis [Pang and Lee, 2005; Song et al., 2015; Tang et al., 2016a; Feng et al., 2018]. ABSA aims to detect sentiments at a more fine-grained level which is more useful for many practical applications in a wide range of fields, such as finance [Jangid et al., 2018], politics [Mohammad et al., 2016] and businesses [Ma et al., 2017].

ABSA-related studies can be roughly divided into two variations: entity-based classification [Ma et al., 2017; Tang et al., 2016b; 2016a] and aspect-based classification [Wang et al., 2016]. The former aims at identifying the sentiment polarity of specific target given its context, while the latter is focused on the polarity of the aspect categories which are usually predefined and may not be explicitly provided in review texts. In this work, we call the specific target as entity and aspect category as aspect for convenience. Ma et al. [2017] proposed the interactive attention networks (IAN) to interactively learn attentions in the contexts and targets, and generate the representations for targets and contexts separately. Tang et al. [2016b] proposed the deep memory network to explicitly capture the importance of each context word when inferring the sentiment polarity of an aspect. Wang et al. [2016] proposed the ATAE-LSTM by modeling the context via LSTM and weighing the hidden vectors with attention mechanism to produce context representation. There are also some other methods on attention modification [Liu et al., 2018a; Tay et al., 2018; Liu et al., 2018b] and incorporating prior knowledge [Ma et al., 2018; He et al., 2018]. However, these methods cannot be directly applied to the ME-ABSA task. Recently, Yang et al. [2018] addressed ME-ABSA task by modeling context memory, entity memory and aspect memory together, which is the most related work to ours. However, the target-specific representations learned by all these methods are heavily dependent on sufficient target-related training instances, which can easily result in cold-start problem.

Cold-start is a challenging issue which is rarely studied in natural language processing, especially for sentiment analysis. Song et al. [2017] proposed a text-driven latent factor model for review rating prediction, meanwhile they introduced a frequency-guided cold-start optimization strategy. Amplayo et al. [2018] alleviated the cold-start problem among users and products by using a shared vector and a frequency-guided selective gate, in addition to the original distinct vector. However, there is no study considering cold-start in the ABSA and ME-ABSA settings which have unique characteristics, that is, similar entities or aspects share the same characteristics, such as contextual word usage, target attributes or similar names. In this work, we use frequency-guided attention mechanism to concentrate on the most related (or similar) targets which can contribute more than others to the representation of cold-start targets, so that the target representations can be updated recursively based on a linear combination of a target-specific context vector learned from target-related reviews and a complementary vector learned from the similar targets.

3 Background: DMN and Cold-Start Problem

In this section, we first give an overview of the Deep Memory Network (DMN) and then discuss the issues of the cold-start problem, which paves way for proposing our approach.

3.1 Deep Memory Network (DMN)

DMN models have been successfully applied to ABSA task and the more general ME-ABSA task, which usually consists of three modules: context encoding, target updating and sentiment classification [Tang et al., 2016b; Yang et al., 2018].
Context Encoding. The model input contains pre-trained word vector of each word $w_s \in \mathbb{R}^d$ in the review, the entity vector $v_e \in \mathbb{R}^d$ and the aspect vector $v_a \in \mathbb{R}^d$, where $d$ is the vector size. The target vector $v_e$ ($v_a$) can be initialized by averaging word vectors of each contained word in the entity (aspect) term. For example, aspect vector for “food quality” is the average of the word vector of “food” and that of “quality”. Previous sentiment analysis studies mainly develop various methods with the goal of precisely modeling contexts of targets by stacking a series of well-designed neural layers that model interactions between target and its context. Context modeling has been well studied (see [Yang et al., 2018] for detail), which is not our focus. For clarity and simplicity, we represent the general process of context encoding as:

$$c = \text{ContextEncoding}(v_e, v_a, \{w_s\})$$  

(1)

Target Updating. In order to learn multiple levels of abstractive representations for entities and aspects, a multi-hop updating strategy is adopted by stacking multiple computational layers. For the $l$-th hop ($l \in [1, L]$), entity vector $v_e^{(l)}$ and aspect vector $v_a^{(l)}$ can be learned recursively by a shared context vector $c^{(l-1)}$ of the review as below:

$$v_e^{(l)} = v_e^{(l-1)} + c^{(l-1)}, \quad v_a^{(l)} = v_a^{(l-1)} + c^{(l-1)}$$  

(2)

Sentiment Prediction. The entity vector $v_e^{(L)}$ and aspect vector $v_a^{(L)}$ in the final layer (i.e., $l = L$) are then concatenated as the inputs of a final output layer, which consists of a linear transformation layer and then a softmax function to predict the sentiment polarity as the final output.

3.2 Cold-Start Problem

The lack of target-related reviews makes the learning difficult on target representation (i.e., $v_e$ and $v_a$). It is challenging to improve the classification quality for the cold-start targets. Although the problem is rarely studied in sentiment analysis, we are inspired by some basic ideas from existing works in recommender systems. For example, Zhang et al. [2014] resort to additional context information such as product attributes, user’s social relation, etc. for product recommendation. But such kind of context is not always available in ME-ABSA and the resulting models are hard to generalize. In our work, we attempt to improve the target representations based on the intuitive idea that other most related targets with good representations provide the possibility to profile the characteristics of the cold-start targets.

4 Cold-Start Aware DMN (CADMN)

In this section, we propose a CADMN framework (see Figure 1) to mitigate the cold-start problem in ME-ABSA task with a novel updating strategy in each computational layer. This paves way for proposing our final approach by instantiating CADMN with a concrete context encoding module.

4.1 Entity Memory and Aspect Memory

The framework introduces two target memories, namely entity memory and aspect memory. Specifically, all the entity vectors $\{v_e\}$ are stacked as the entity memory $E \in \mathbb{R}^{N \times d}$ and all the aspect vectors $\{v_a\}$ are stacked as the aspect memory $A \in \mathbb{R}^{M \times d}$, where $N$ and $M$ denote the number of entities and aspects appearing in the dataset, respectively.

4.2 Target Updating

The basic idea of our cold-start strategy is to enhance the target representation based on a complementary vector derived from most related targets which share the same characteristics. Previous studies have proven that the amount of targets appearing in the dataset (i.e., occurring frequency) can be used as helpful prior information to optimize the estimation of target-specific parameters [Song et al., 2017], indicating that high frequency usually determines a good target representation due to sufficient volume of training instances. In this work, we resort to frequency information to guide the selection of memory slices based on a target attention mechanism for composing the attended target-related vectors as the complementary vector.

Frequency Attention. Compared to absolute frequency, relative frequency is better which is not sensitive to data scale. Specifically, the relative frequency can be defined by dividing the absolute values of target-specific frequencies $f(e)$ and $f(a)$ by the values of average frequencies $Avg\{f(e)\}$ and $Avg\{f(a)\}$, respectively. The larger the relative frequency, the more attention should be paid on the memory slice. For each memory slice in $E$ and $A$, the corresponding attention weight $p_e$ and $p_a$ for entity $e$ and aspect $a$ can be formulated as:

$$p_e = 1 - 1/e^{AVG(f(e))} \quad p_a = 1 - 1/e^{AVG(f(a))}$$  

(3)

Target Attention. Similar targets share the same characteristics, which reflects similar contextual word usages, target attributes, etc. However, different targets contribute differently to the representation of the cold-start targets. This layer aims to automatically concentrate on the most related targets via target attention mechanism. Specifically, we feed the concatenation of an entity representation $v_e^{(l)}$ in the $l$-th computational layer and the output of frequency attention layer
\( p_{e_i} \cdot E_i \) through a one-layer MLP to get a hidden representation \( u_i \), where \( p_{e_i} \) denotes the frequency of the \( i \)-th entity (i.e., \( E_i \) or \( v_e \)) in \( E \). Then we measure the importance of each memory slice as the similarity of \( u_i \) with a trainable context vector \( z \) and get a normalized importance weight \( \beta_i \) through a softmax function as below:

\[
\beta_i = \text{softmax}(z^T u_i), \quad \beta_i \in [0, 1]
\]

\[
u^{(l)}_{e} = \sum_{i \in [1,N]} \beta_i E_i
\]

\[\nu^{(l)}_{v} = z^T u_i \]

where \( W_0 \in \mathbb{R}^{d \times 2d} \) and \( b_0 \in \mathbb{R}^{2d} \) are parameters of the MLP. \( z \in \mathbb{R}^{2d} \) can be seen as a high level representation of a fixed query “what is the related target”, \( \beta_i \) is attention weight of memory slice \( E_i \) and \( \nu^{(l)}_{v} \in \mathbb{R}^d \) is the entity complementary vector in the \( l \)-th hop. We can also compute attention weight \( \gamma_j \) of memory slice \( A_j \) (or \( v_a \)) and finally derive the aspect complementary vector \( v_{ca} = \sum_{j \in [1,N]} \gamma_j A_j \).

Fusion Gate. Finally, individual target representation can be represented as a real combination of a context vector learned from target-related reviews (i.e., \( v^{(l-1)}_{v} + c^{(l-1)} \) in formula 2) and a complementary vector derived from most related targets (i.e., \( v^{(l)}_{v} \) in formula 4). Instead of simple element-wise addition, we use a flexible and self-adapting combination strategy called fusion gate which performs better and is formulated as:

\[
q_e = \sigma(W_{qe}\left[v^{(l)}_{v}; v^{(l-1)}_{v} + c^{(l-1)}\right])
\]

where \( W_{qe} \in \mathbb{R}^{1 \times 2d} \) is the trainable weight matrix, \( q_e \in [0, 1] \) is the fusion gate, \( \sigma(\cdot) \) is the sigmoid function.

Similarly, we can also obtain a fusion gate \( q_a \in [0, 1] \). Finally, the entity/aspect representation can be represented as below:

\[
\nu^{(l)}_{e} = q_e \cdot \nu^{(l)}_{v} + (1 - q_e) \cdot \left( v^{(l-1)}_{v} + c^{(l-1)} \right)
\]

\[
\nu^{(l)}_{a} = q_a \cdot \nu^{(l)}_{v} + (1 - q_a) \cdot \left( v^{(l-1)}_{v} + c^{(l-1)} \right)
\]

4.3 Sentiment Prediction

The final entity vector \( \nu^{(L)}_e \) and aspect vector \( \nu^{(L)}_a \) are then concatenated as the inputs of a final output layer, which consists of a linear transformation layer and then a softmax function to produce sentiment probability distribution \( y \) as below:

\[
x = W_2 \left( W_1 [\nu^{(L)}_e; \nu^{(L)}_a] + b_1 \right) + b_2
\]

\[
y_i = \text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j \in [1,C]} e^{x_j}}
\]

where \( C \) is the number of classes, \( y_i \in \mathbb{R} \) is the predicted probability for the \( i \)-th class (\( i \in [1,C] \)), \( W_1 \in \mathbb{R}^{d \times 2d} \) and \( W_2 \in \mathbb{R}^{C \times d} \) are trainable weight matrices, \( b_1 \in \mathbb{R}^d \) and \( b_2 \in \mathbb{R}^C \) are trainable bias vectors. We output \( y \in \mathbb{R}^C \) which is the probability distribution of sentiment classes.

4.4 Optimization and Training

The models are trained by minimizing the cross entropy error of sentiment classification with L2 regularization loss as:

\[
\mathcal{L}(\Theta) = - \sum_{i \in [1,C]} \sum_{j \in [1,T]} g_i^j \log(y_i^j) + \lambda \sum_{\theta \in \Theta} \theta^2
\]

where \( T \) is the number of training instances, \( g_i^j \) is 1 or 0 indicating whether the \( i \)-th class is a correct answer for the \( j \)-th training instance, \( y_i^j \) is the corresponding probability predicted, \( \lambda \) is the empirically fixed regularization coefficient and \( \theta \) any trainable parameter in the parameter set \( \Theta \). After learning \( \Theta \), we feed each test instance into the final model, and the label with the highest probability stands for the predicted sentiment polarity. We use back propagation to calculate the gradients of all the model parameters, and update them with Adam [Kingma and Ba, 2014].

5 Context Representation

In this section, we instantiate the CADMN with a concrete context encoding approach as that used in [Yang et al., 2018]. Note that our memory scheme is not confined by this specific context encoding method which can be easily replaced.

Interaction Layer. This layer aims to take entity and aspect information into consideration when encoding context memory, which produces concatenated word vector by:

\[ f(w_s, v_c, v_a) = [w_s; w_c \odot v_c; w_a \odot v_a] \]

where \( \odot \) denotes element-wise multiplication.

Position Attention Layer. Based on the idea that sentiment towards the entity and aspect is more likely to be expressed by the words near them. This layer pays attention to each word \( w_s \) and the attention weight is based on the position attention function \( g_e \) for the entity and \( g_a \) for the aspect:

\[
g_e(p_s, p_e, S) = 1 - |p_s - p_e|/S
\]

\[
g_a(p_s, p_a, S) = 1 - |p_s - p_a|/S
\]

where \( p_s \) is the position of the \( s \)-th word \( w_s, p_e \) is the position of the entity (aspect) term nearest to \( w_s \), and \( S \) is the length of the review. The output after the position attention layer is calculated by:

\[
o_s = f(w_s, v_c, v_a) \odot g_e(p_s, p_e, S) \odot g_a(p_s, p_a, S)
\]

LSTM Layer. The output is then fed into a Long Short-Term Memory (LSTM) network [Hochreiter and Schmidhuber, 1997] which produces the hidden states \( \{h_s; s \in [1, S]\} \) to capture context information [Chen et al., 2017], where each \( d \)-dimensional hidden state \( h_s = LSTM(o_s) \).

Entity-Aspect Attention Layer. This layer calculates the context memory by using entity-aspect attention mechanism:

\[
att_s = W_3 \tanh(W_4 h_s; \odot v_c; h_s \odot v_a) + b_3
\]

\[
o_s = \text{softmax}(att_s) = \frac{e^{att_s}}{\sum_{s \in [1,S]} e^{att_s}}
\]

where \( W_3 \in \mathbb{R}^{1 \times 3d} \) and \( W_4 \in \mathbb{R}^{3d \times 3d} \) are trainable weight matrices, \( b_3 \in \mathbb{R}^d \) is a bias vector, and probability \( o_s \) is the weight for \( h_s \). At last, the context memory \( c \) is calculated by:

\[
c = \sum_s o_s h_s
\]
6 Experiments and Results

6.1 Experimental Setting

Our experiments are conducted based on two datasets: a English restaurant review dataset RES from SemEval 2016 Task1, and a public Chinese babycare review dataset BBC from [Yang et al., 2018]. RES is initially designed for ABSA task on sentence-level, and each sentence has been annotated with aspect term “sushi” and its associated category “Food#Quality”. We consider the aspect term as entity and its category as aspect. Later, we set aside 10% from the training set as the validation set for parameter adjustment. BBC is a document-level ME-ABSA dataset which has been split into training, validation and test sets. A summary of statistics for both datasets are displayed in Table 1.

The word vectors are initialized by two embedding resources. The English resource is built by ourselves with SSWE [Tang et al., 2014] which encodes sentiment information into the word embeddings. In our implementation, the corpus is obtained from Yelp dataset (i.e., yelp.com/dataset/challenge) which provides 5.2M rated restaurant reviews (i.e., with 1-5 stars). The vocabulary size is about 1.3M words and the dimensionality of word embeddings is 300. The Chinese one is also a 300-dimensional embedding resource built by Yang et al. [2018] with Glove [Pennington et al., 2014]. Other trainable parameters are given initial values by sampling from uniform distribution $U(-0.01, 0.01)$. The size of hidden state is set as 300. The other hyper-parameters are tuned on the validation set. Specifically, the initial learning rate is fixed as $10^{-3}$, the regularization weight $\lambda$ is set to $10^{-3}$, the dropout rate is set as 0.5, the batch size is 25, the number of hops is set as 3 and the number of epochs is 10. The performance is evaluated using standard Accuracy and Macro $F_1$.

6.2 Comparison on Different Methods

In our experiments, we compare our proposed models with the following state-of-the-art baseline methods:

1. LSTM: The method uses a LSTM to model the review. After that, the average value of all the hidden states is fed into a linear layer and a softmax layer to estimate the probability of each sentiment polarity. LSTM+ is an enhanced version of LSTM by concatenating entity vector, aspect vector and the last hidden state as the inputs of the next layers.

2. ATAE-LSTM: The model incorporates the aspect vector into the attention mechanism. As such, the model learns to attend to different parts of the sentence based on the aspect [Wang et al., 2016]. ATAE-LSTM+ improves ATAE-LSTM by adding entity in the same manner as the aspect.

3. IAN: The method uses an interactive attention mechanism to alternately learn attentions in the review and entity, and generate their representations separately. Finally, both review and target vectors are concatenated and fed into a softmax layer for classification [Ma et al., 2017]. IAN+ is implemented by adding aspect part in the same manner as entity.

4. MemNet: The deep memory network on ABSA task proposed by [Tang et al., 2016b]. This work only updates aspect memory without updating context memory. MemNet+ adds entity part in the same manner as aspect.

5. CEA: The state-of-the-art ME-ABSA model which improves MemNet by modeling review text memory, entity memory and aspect memory [Yang et al., 2018]. We run their provided source code on the datasets and tune the parameters on the validation set. CEA ignores the cold-start problem.

6. CADMN: Our approach which only considers the target attention layer; CADMN$_{freq}$: Our approach which only considers the frequency attention layer.

From Table 2, we can observe that LSTM has low performance because they ignore both entity and aspect information. ME-ABSA models LSTM+, IAN+ and ATAE-LSTM+ perform better than their corresponding basic models LSTM, IAN and ATAE-LSTM respectively, because expressions are generally related to both entity and aspect. MemNet+ is faster but weaker than IAN+ for ignoring the word order which leads to the difficulty in context understanding. CEA is not stronger than other baseline methods because of severe data-sparsity problem in RES dataset. Our method outperforms all the baseline methods on both the datasets, which verifies the effectiveness of our approaches on alleviating cold-start problem. Compared with CEA, our method achieves the averaged improvements of 2.58% of Accuracy and 7.49% of Macro $F_1$ on the RES dataset, as well as

<table>
<thead>
<tr>
<th>Dataset</th>
<th>positive</th>
<th>negative</th>
<th>neutral</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RES</td>
<td>1,657</td>
<td>749</td>
<td>101</td>
<td>2,507</td>
</tr>
<tr>
<td>BBC</td>
<td>15,090</td>
<td>10,877</td>
<td>3,387</td>
<td>29,354</td>
</tr>
<tr>
<td></td>
<td>1,858</td>
<td>1,442</td>
<td>384</td>
<td>3,672</td>
</tr>
<tr>
<td></td>
<td>1,858</td>
<td>1,394</td>
<td>425</td>
<td>3,677</td>
</tr>
</tbody>
</table>

Table 1: The statistics of our experimental datasets. Notations $p_e$ and $p_a$ denote whether position information is provided.

<table>
<thead>
<tr>
<th>Methods</th>
<th>RES</th>
<th>BBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Macro-F1</td>
</tr>
<tr>
<td>Majority</td>
<td>71.12</td>
<td>27.70</td>
</tr>
<tr>
<td>LSTM</td>
<td>81.14</td>
<td>51.72</td>
</tr>
<tr>
<td>IAN</td>
<td>84.40</td>
<td>63.11</td>
</tr>
<tr>
<td>ATAE-LSTM</td>
<td>84.51</td>
<td>65.94</td>
</tr>
<tr>
<td>MemNet+</td>
<td>83.70</td>
<td>62.89</td>
</tr>
<tr>
<td>LSTM+</td>
<td>81.95</td>
<td>52.94</td>
</tr>
<tr>
<td>IAN+</td>
<td>86.57</td>
<td>63.66</td>
</tr>
<tr>
<td>ATAE-LSTM+</td>
<td>86.03</td>
<td>63.80</td>
</tr>
<tr>
<td>MemNet+</td>
<td>85.15</td>
<td>65.07</td>
</tr>
<tr>
<td>CEA</td>
<td>85.68</td>
<td>65.12</td>
</tr>
</tbody>
</table>

Table 2: Comparison among different methods for ME-ABSA.
Table 3: Comparison among different methods on sparse dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Ratio</th>
<th>RES</th>
<th>BBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td>Macro-F1</td>
</tr>
<tr>
<td>CEA</td>
<td>80%</td>
<td>85.44</td>
<td>63.81</td>
</tr>
<tr>
<td>CADMN</td>
<td>85.11</td>
<td>63.81</td>
<td>75.25</td>
</tr>
<tr>
<td>CEA</td>
<td>80%</td>
<td>85.44</td>
<td>63.81</td>
</tr>
<tr>
<td>CADMN</td>
<td>87.31</td>
<td>67.85</td>
<td>80.63</td>
</tr>
<tr>
<td>CEA</td>
<td>50%</td>
<td>84.05</td>
<td>59.11</td>
</tr>
<tr>
<td>CADMN</td>
<td>86.03</td>
<td>62.94</td>
<td>79.41</td>
</tr>
<tr>
<td>CEA</td>
<td>20%</td>
<td>82.30</td>
<td>53.04</td>
</tr>
<tr>
<td>CADMN</td>
<td>82.88</td>
<td>53.89</td>
<td>76.96</td>
</tr>
</tbody>
</table>

Figure 2: Visualization of content attention weight $\alpha$ and target attention weight $\beta$, $\gamma$. The color depth indicates different importance.

Figure 3: The Count of reviews containing cold-start targets (entity-aspect combinations) and their Accuracy in the test set against their frequency in the training set. The horizontal axis is the frequency of targets in the training set. The curves are related to the left vertical axis: Accuracy (%). The histogram is related to the right vertical axis: Count of the test reviews containing the targets.

7 Conclusion and Future Direction

In this paper, we propose a novel model CADMN to address the cold-start problem in the ME-ABSA task. We use attention mechanism to concentrate on most related targets which provide helpful complementary information to enhance the representation of cold-start targets. A concrete model is introduced by instantiating the framework and then validated on two public datasets. Extensive experimental results indicate the effectiveness of our approach. In the future, we will study to alleviate the cold-start problem by utilizing additional product description information.
References


