

Transferable Interactive Memory Network for Domain Adaptation in Fine-Grained Opinion Extraction

Wenya Wang, Sinno Jialin Pan

Nanyang Technological University, Singapore
{wangwy, sinnopan}@ntu.edu.sg

Abstract

In fine-grained opinion mining, aspect and opinion terms extraction has become a fundamental task that provides key information for user-generated texts. Despite its importance, a lack of annotated resources in many domains impede the ability to train a precise model. Very few attempts have applied unsupervised domain adaptation methods to transfer fine-grained knowledge (in the word level) from some labeled source domain(s) to any unlabeled target domain. Existing methods depend on the construction of “pivot” knowledge, e.g., common opinion terms or syntactic relations between aspect and opinion words. In this work, we propose an interactive memory network that consists of local and global memory units. The model could exploit both local and global memory interactions to capture intra-correlations among aspect words or opinion words themselves, as well as the interconnections between aspect and opinion words. The source space and the target space are aligned through these domain-invariant interactions by incorporating an auxiliary task and domain adversarial networks. The proposed model does not require any external resources and demonstrates promising results on 3 benchmark datasets.

Introduction

Fine-grained opinion extraction involves the identification of aspect terms (a.k.a opinion targets) and opinion terms. As an illustrating example, given a review sentence “*It’s kinda loud when the computer fan is running*”, the aspect term and the opinion term to be extracted are *computer fan* and *loud*, respectively. Several supervised models have been proposed for fine-grained opinion extraction given sufficient labeled training data. They treated the task as a sequence labeling problem and apply graphical models (Jakob and Gurevych 2010; Li et al. 2010) or deep learning models (Liu, Joty, and Meng 2015; Yin et al. 2016; Wang et al. 2017; Li and Lam 2017) to solve it. However, the success of these supervised models depends on the availability of annotated resources, which are not the case in many real situations. Although unsupervised methods have also been proposed to extract aspect/opinion terms without any labeled data (Hu and Liu 2004; Qiu et al. 2011), they largely relied on manual constructions of rules with pipelined procedures and the performances are not promising.

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

To address these problems, unsupervised domain adaptation methods can be applied to transfer knowledge learned from labeled source domain to any unlabeled target domain. Suppose the source domain contains restaurant reviews, e.g., “*They have nice food*”, and the target domain is laptop, e.g., “*The problem is the noisy fan*”. Li et al. (2012) proposed to use opinion seeds (*nice*, *noisy*) and pre-mined common syntactic relations between aspect and opinion terms (*noisy*→*amod*→*fan*, *nice*→*amod*→*food*) to bridge the gap between different domains. They applied bootstrapping method to iteratively select candidate aspect terms and opinion terms for the target domain. Ding, Yu, and Jiang (2017) proposed to use auxiliary tasks as a supervision that are integrated into a recurrent neural network to learn shared representations for words across domains. The auxiliary tasks are generated from manually-designed rules that could inform common syntactic relations between target words. The above works both depend on prior knowledge to construct the bridges which are fixed and inflexible. Recently, Wang and Pan (2018) applied recursive neural networks with pre-generated dependency tree for each sentence to learn dependency-sensitive representation for each word, where structural correspondences across domains are built according to syntactic structures. This approach requires a dependency parser that is not guaranteed to be accurate especially on user-generated texts, which may lead to negative effect on the extraction performance.

To alleviate the dependency on linguistic resources to construct prior knowledge, we propose a Transferable Interactive Memory Network (TIMN) that could learn shared representations across domains automatically and effectively. To be specific, the model consists of multiple layers of attention modules with both local and global memory units. The local memory includes the hidden representation of each word within an input sentence, whereas the global memory includes an aspect summary vector and an opinion summary vector. At each layer, the attention module produces 2 sets of transformed representation and attention weight for each word to update global aspect memory and opinion memory, respectively, by exploiting the complex interactions between local and global memory units. The attention mechanism selects most probable aspect words to be incorporated in the aspect summary, and opinion words to be incorporated in opinion summary.

Existing works have shown that two kinds of pivot¹ information are crucial to transfer knowledge across different domains: (1) common opinion terms. (2) syntactic relations among aspect and opinion words within a sentence. Motivated by (1), we incorporate a domain adversarial network (DAN) on global opinion memory at each layer to learn similar representations for opinion words across domains. For (2), instead of using syntactic structures that rely on imperfect parsers, we automatically generate 3 types of pivot interactions among target terms. The first type refers to intra-correlations among aspect words, e.g., *French fries* in restaurant domain and *sound output* in laptop domain are both aspect compounds. The second type refers to intra-correlations among opinion words, e.g., *tasty and enjoyable* in restaurant domain and *fast and efficient* in laptop domain both contain opinion conjunctions. The third type relates to inter-relations between aspect and opinion words, e.g., *nice food* and *noisy fan* share the same pattern where the opinion word modifies the aspect word. These 3 interactions can be modeled through global memory interactions, with an auxiliary task to classify the interaction type in order to align two different domains according to target-word interactions.

In summary, our model has the following contributions: 1) We propose to focus on the complex interactions among the target words within a sentence that could be transferred for cross-domain fine-grained opinion extraction and summarize these interactions into 3 different categories. 2) We present a novel memory network with both local and global memory interactions that can model the desired relations with sharing strategies automatically without the reliance on external linguistic resources. 3) We demonstrate the effectiveness of the proposed model with extensive experiments.

Related Work

Fine-grained opinion extraction has been actively investigated using supervised methods, including graphical models (Jin and Ho 2009; Li et al. 2010) and deep learning models (Liu, Joty, and Meng 2015; Yin et al. 2016; Wang et al. 2016; Xu et al. 2018; Zhang, Zhang, and Vo 2015; Li and Lam 2017). Syntactic structure has been proven to be useful for the extraction task (Yin et al. 2016; Wang et al. 2016), however, they are prone to parsing errors. Recent works have applied attention mechanism (Mnih et al. 2014) and memory network (Weston, Chopra, and Bordes 2015) to automatically model aspect-opinion correlations without the reliance on linguistic parsers (Wang et al. 2017; Li and Lam 2017), but they have not made these interactions precise, neither did they explore the possibility for knowledge transfer. On the other hand, unsupervised methods were long since been proposed including rule-based mining (Hu and Liu 2004; Qiu et al. 2011; Zhuang, Jing, and Zhu 2006; Popescu and Etzioni 2005), topic modeling (Titov and McDonald 2008; Mei et al. 2007; Lu, Zhai, and Sundaresan 2009; Zhang et al. 2010; He et al. 2017) and word alignment (Liu et al. 2013). However, prior knowledge is required to achieve reasonable perfor-

¹Pivot information refers to common knowledge shared across domains as a bridge to associate other uncommon information.

mances and they can not make use of labeled data in the cross-domain setting.

Many domain adaptation methods have been proposed for cross-domain sentence-level or document-level sentiment classification. Some of them aimed to utilize pivot information that are shared across domains as a bridge to align different feature spaces (Blitzer, Dredze, and Pereira 2007; Pan et al. 2010; Bollegala, Maehara, and ichi Kawarabayashi 2015; Yu and Jiang 2016). Another group of works directly learned shared spaces across different domains through projection via auto-encoders (Glorot, Bordes, and Bengio 2011; Chen et al. 2012; Zhou et al. 2016) or domain adversarial network (Li et al. 2017). Few works have addressed the problem of cross-domain fine-grained opinion extraction. Jakob and Gurevych(2010) proposed cross-domain CRF that was built on non-lexical invariant feature engineering. Li et al.(2012) and Ding, Yu, and Jiang(2017) explored pivot knowledge on opinion lexicon and syntactic relations to build correspondences across domains. Wang and Pan(2018) constructed dependency-based recursive neural networks that incorporated syntactic correspondences into the deep model learning process. However, all of these methods depend on external linguistic tools and thus are prone to knowledge errors. To alleviate this problem, we propose an end-to-end memory network without additional resources.

Problem Statement & Motivation

Formally, the task of aspect and opinion terms extraction is modeled as a sequence labeling problem. The input is a sequence of tokens denoted as $w = \{w_1, w_2, \dots, w_n\}$, with the corresponding word embeddings as $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$. The output is a label sequence that assigns a label for each word $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ with $y_i \in \{\text{BA, IA, BO, IO, N}\}$. Here we use “BIO” encoding scheme to denote “BA” (“BO”) as beginning of an aspect (opinion) term, “IA” (“IO”) as inside of an aspect (opinion) term, and “N” as none of the above. For cross-domain extraction, we are given a set of labeled data $\mathcal{D}_S = \{(\mathbf{x}_i^S, \mathbf{y}_i^S)\}_{i=1}^{N_S}$ from source domain and a set of unlabeled data $\mathcal{D}_T = \{(\mathbf{x}_j^T)\}_{j=1}^{N_T}$ from target domain. We aim to transfer knowledge from \mathcal{D}_S to \mathcal{D}_T in order to produce label sequence \mathbf{y}^T for target domain.

Motivation & Overview

It has been shown that aspect words and opinion words within a sentence are usually correlated through certain syntactic relations which are crucial for the extraction by propagating label information. However, explicit relations require an external parser to obtain and are not guaranteed to be accurate especially for user-generated texts. On the other hand, memory networks have proven to be capable of modeling implicit interactions through attention mechanism and memory updates. With this intuition, we propose to use a memory network with multi-layer attentions to iteratively and interactively select most relevant aspect words and opinion words within each sentence. The selection is achieved by exploiting 3 types of interactions: intra-correlations within aspect words, intra-correlations within opinion words and inter-connections between aspect and opinion words, each

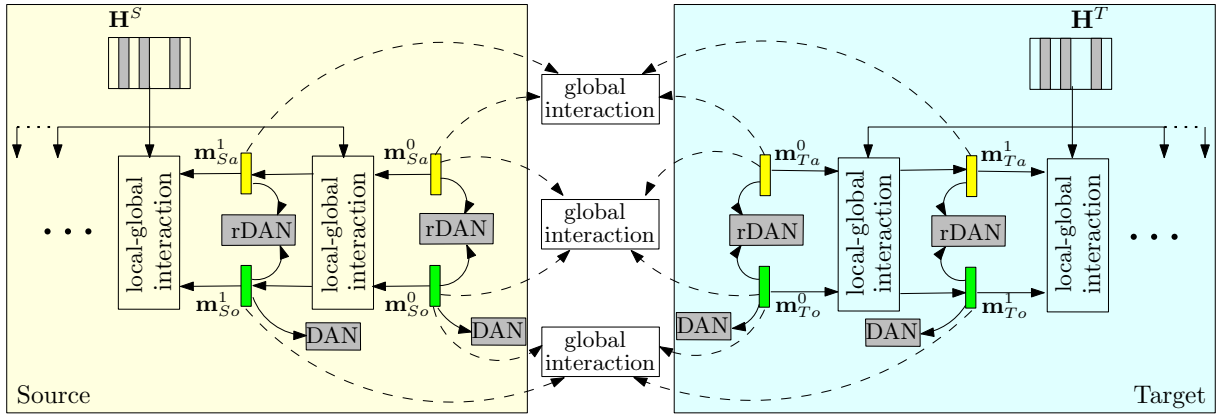


Figure 1: The overall architecture of TIMN.

of which is measured by a 3-dimensional tensor operation. Specifically, we design an interactive memory network that consists of local memories with hidden representation for each word, a global aspect memory and a global opinion memory that integrates aspect candidates and opinion candidates, respectively. At each layer, the attention module computes the interactions between each local memory and global memories. In that case, if a word has both active intra-correlations with aspect memory and inter-connections with opinion memory, it will be attended as an aspect candidate.

To transfer knowledge across domains, we utilize 2 types of pivot information to bridge the gap between the source domain and the target domain. Firstly, the two domains share some opinion terms, which makes it easier to align two domains through invariant opinion representations. This could be achieved through a DAN on top of opinion memory vector for each layer of the memory network. Secondly, the interactions among target words, which are summarized into 3 different categories as mentioned above, are also invariant across domains. To learn interaction-invariant representations, we create an auxiliary task for both domains to predict the relation types on global memory interactions. Specifically, we regard the interactions between global aspect (opinion) memories in consecutive layers as intra-correlations and the interactions between aspect and opinion memories in the same layer as inter-connections as the auxiliary labels for each domain. Furthermore, we use another relational DAN (named as rDAN) to discriminate the concatenation of aspect memory and opinion memory at each layer across domains to emphasize invariant aspect-opinion relations. It can be deemed as aligning aspect representations conditioned on their corresponding opinion memories.

The overall architecture of TIMN is shown in Figure 1. For illustration, we present networks of both the source and target domains. For the source domain, TIMN consists of two global memory vectors, namely aspect memory \mathbf{m}_{S_a} and opinion memory \mathbf{m}_{S_o} , and a local memory matrix \mathbf{H}^S , where each column \mathbf{h}_i^S is the hidden representation for w_i . We denote $\mathbf{m}_{S_a}^l$ and $\mathbf{m}_{S_o}^l$ as global memory representations at the l -th layer. For each domain, the local memory interacts with two global memory vectors at each layer to update the

global memories at the next layer. At the same time, domain adaptations are incurred through the following strategies (we use $P \in \{S, T\}$ to indicate source or target domain): 1) $\mathbf{m}_{P_o}^l$ at each layer is taken as input to a DAN with domain labels. 2) Both $\mathbf{m}_{P_a}^l$ and $\mathbf{m}_{P_o}^l$ are fed into rDAN to align aspect-opinion relations. 3) $(\mathbf{m}_{P_a}^l, \mathbf{m}_{P_a}^{l+1})$, $(\mathbf{m}_{P_o}^l, \mathbf{m}_{P_o}^{l+1})$ and $(\mathbf{m}_{P_a}^l, \mathbf{m}_{P_o}^l)$ are fed into the global interaction module to predict interaction type as an auxiliary task.

Transferable Interactive Memory Network

The proposed TIMN focuses on two forms of memory interactions: domain-specific local-global memory interactions and cross-domain global memory correlations. Specifically, local-global memory interactions are firstly exploited within each domain to select domain-specific aspect and opinion words and produce global memory representations for each domain, which are then employed for knowledge transfer.

Local-Global Memory Interactions

For each domain, the task of aspect and opinion terms extraction relies on the modeling of complex relations among aspect words and opinion words. Without any explicit prior knowledge, we summarize these relations into 3 categories and apply tensor operations to exploit multi-level bilinear interactions among the related entities. Figure 2 provides a specific structure for domain-specific local-global memory interaction module for the l -th layer. Here domain subscript is removed for ease of illustration. We use a 3-dimensional tensor \mathbf{T}_a to exploit intra-correlations among aspect words, \mathbf{T}_o to exploit intra-correlations among opinion words, and \mathbf{T}_{ao} to model inter-connections between aspect words and opinion words. At l -th layer, the model reads a global aspect memory \mathbf{m}_a^l , a global opinion memory \mathbf{m}_o^l and a local memory matrix \mathbf{H} . To incorporate sequential information of a sentence, \mathbf{H} is obtained by applying a GRU (Gated Recurrent Unit) network on pre-trained word embeddings.

The attention module for each layer computes two sets of outputs, one for aspect memory and the other for opinion memory. As a first step, the local memory $\mathbf{h}_i \in \mathbb{R}^d$ for each word interacts with the global memories to produce two

transformed correlation vectors via the following formula:

$$\mathbf{z}_{a,i}^l = [\mathbf{m}_a^{l\top} \mathbf{T}_a \mathbf{h}_i : \mathbf{m}_o^{l\top} \mathbf{T}_{ao} \mathbf{h}_i], \quad (1)$$

$$\mathbf{z}_{o,i}^l = [\mathbf{m}_o^{l\top} \mathbf{T}_o \mathbf{h}_i : \mathbf{m}_a^{l\top} \mathbf{T}_{ao} \mathbf{h}_i], \quad (2)$$

where $[\cdot]$ indicates concatenation of vectors and $\mathbf{T}_a, \mathbf{T}_o, \mathbf{T}_{ao} \in \mathbb{R}^{k \times d \times d}$. $\mathbf{z}_{a,i}^l$ and $\mathbf{z}_{o,i}^l$ can be regarded as correlation vectors for aspect module and opinion module, respectively, that measure the extent of interactions between \mathbf{h}_i and global memories. For example, if \mathbf{h}_i has high intra-correlations with aspect memory \mathbf{m}_a^l and inter-connections with opinion memory \mathbf{m}_o^l , it is more likely to be an aspect word. The final transformation vectors at l -th layer are produced through another GRU network to incorporate contextual attention influence:

$$\bar{\mathbf{z}}_{a,i}^l = f_{\text{GRU}}(\mathbf{z}_{a,i}^l, \bar{\mathbf{z}}_{a,i-1}^l; \theta_{\text{GRU}}^a), \quad (3)$$

$$\bar{\mathbf{z}}_{o,i}^l = f_{\text{GRU}}(\mathbf{z}_{o,i}^l, \bar{\mathbf{z}}_{o,i-1}^l; \theta_{\text{GRU}}^o), \quad (4)$$

with $\theta_{\text{GRU}}^a, \theta_{\text{GRU}}^o$ representing all parameters within a GRU network. The attentions then compute normalized relevance scores $\alpha_{a,i}^l$ and $\alpha_{o,i}^l$ that indicate the possibility of each word being an aspect word or opinion word, respectively:

$$\alpha_{a,i}^l = \frac{\exp(e_{a,i}^l)}{\sum_{j=1}^n \exp(e_{a,j}^l)}, \quad e_{a,i}^l = \mathbf{v}_a^\top \bar{\mathbf{z}}_{a,i}^l, \quad (5)$$

$$\alpha_{o,i}^l = \frac{\exp(e_{o,i}^l)}{\sum_{j=1}^n \exp(e_{o,j}^l)}, \quad e_{o,i}^l = \mathbf{v}_o^\top \bar{\mathbf{z}}_{o,i}^l, \quad (6)$$

where \mathbf{v}_a and \mathbf{v}_o are one-dimensional vectors to be learned during training. These attentions scores, together with local memories, will generate aspect- or opinion-related representations for the input sentence to update the global memories for the next layer:

$$\mathbf{m}_a^{l+1} = \tanh(\mathbf{M}_a \mathbf{m}_a^l) + \sum_{i=1}^n \alpha_{a,i}^l \mathbf{h}_i, \quad (7)$$

$$\mathbf{m}_o^{l+1} = \tanh(\mathbf{M}_o \mathbf{m}_o^l) + \sum_{i=1}^n \alpha_{o,i}^l \mathbf{h}_i. \quad (8)$$

\mathbf{M}_a and \mathbf{M}_o are transformation matrices for global memories. The updates iteratively refine the global memory representations to incorporate more relevant candidates according to attention mechanism. Without knowledge transfer, this module is similar to (Wang et al. 2017).

Global Memory Correlations

With global aspect and opinion memories generated for each domain, shared pivot knowledge could be exploited to bridge the gap between domains. The first type of pivot knowledge lies in common opinion terms across domains. From (8), the global opinion memory at each layer summarizes most probable opinion words in the input sentence. Then sharing of opinion representations could be achieved through a domain adversarial network on top of the global opinion memories. As shown in Figure 1, we apply DAN for \mathbf{m}_o^l at each layer by integrating a domain discriminator that generates a probability distribution $P(D|\mathbf{m}_o^l)$ over

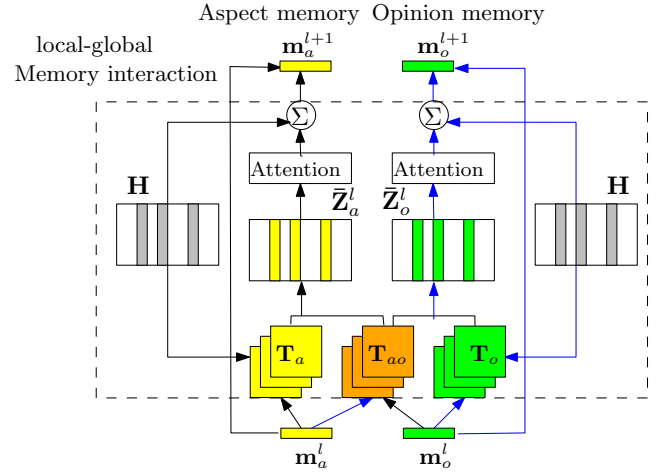


Figure 2: Local-global memory interaction at layer l .

2 domains, where $D = 1$ indicates the source domain. The prediction is produced via a fully-connected layer:

$$\bar{\mathbf{m}}_o^l = \tanh(\mathbf{W}_o^d \mathbf{m}_o^l + \mathbf{b}_o^d), \quad (9)$$

$$\bar{y}_o^d = P(D|\mathbf{m}_o^l) = \text{softmax}(\mathbf{V}_o^d \bar{\mathbf{m}}_o^l + \mathbf{c}_o^d). \quad (10)$$

Since aspect terms are usually disjoint across domains, simply aligning them is difficult. We propose to use rDAN that learns invariant aspect-opinion relations, thus aligning aspect memories conditioned on their corresponding opinion representations. With similar procedures, we obtain

$$\bar{\mathbf{m}}_a^l = \tanh(\mathbf{W}_a^d [\mathbf{m}_a^l : \mathbf{m}_o^l] + \mathbf{b}_a^d), \quad (11)$$

$$\bar{y}_a^d = P(D|\mathbf{m}_a^l, \mathbf{m}_o^l) = \text{softmax}(\mathbf{V}_a^d \bar{\mathbf{m}}_a^l + \mathbf{c}_a^d). \quad (12)$$

The second type of pivot information is domain-invariant correlations among target words. As mentioned previously, there exists 3 types of interactions including aspect-aspect, opinion-opinion and aspect-opinion relations. We formulate these relations with different forms of global memory interactions. As the global aspect memory \mathbf{m}_a^l at each layer contains attended aspect candidates at the current timestamp, the aspect-aspect interactions could be modeled by applying tensor operator \mathbf{T}_a between consecutive layers \mathbf{m}_a^l and \mathbf{m}_a^{l+1} . Similar strategy applies for opinion memories. We use aspect and opinion memories at the same layer to compute inter-connections.

$$\mathbf{r}_a = \mathbf{m}_a^{l\top} \mathbf{T}_a \mathbf{m}_a^{l+1}, \quad (13)$$

$$\mathbf{r}_o = \mathbf{m}_o^{l\top} \mathbf{T}_o \mathbf{m}_o^{l+1}, \quad (14)$$

$$\mathbf{r}_{ao} = \mathbf{m}_a^{l\top} \mathbf{T}_{ao} \mathbf{m}_o^l. \quad (15)$$

To make these relations invariant across domains, an auxiliary task is applied on top of $\mathbf{r}_p \in \mathbb{R}^k$ ($p \in \{a, o, ao\}$) to predict the relation type, which is a 3-class classification task:

$$\bar{y}_p^u = \text{softmax}(\mathbf{V}^u \mathbf{r}_p + \mathbf{c}^u). \quad (16)$$

Training

For token-level prediction, we take transformed hidden representations $\bar{\mathbf{z}}_{a,i}^l$ and $\bar{\mathbf{z}}_{o,i}^l$ from (3), (4) of each word at each

layer as the input and apply a softmax function to classify the ‘‘BIO’’ position of the word corresponding to aspect terms or opinion terms:

$$\bar{y}_{a,i} = \text{softmax}(\mathbf{V}_a \sum_{l=1}^L \bar{\mathbf{z}}_{a,i}^l + \mathbf{c}_a), \quad (17)$$

$$\bar{y}_{o,i} = \text{softmax}(\mathbf{V}_o \sum_{l=1}^L \bar{\mathbf{z}}_{o,i}^l + \mathbf{c}_o), \quad (18)$$

where $\bar{y}_{a,i}, \bar{y}_{o,i} \in \mathbb{R}^3$ with each entry indicating the probability of position ‘‘B’’, ‘‘I’’ or ‘‘O’’ for aspect or opinion predictions, respectively. The final prediction for each word is generated by comparing $\bar{y}_{a,i}$ with $\bar{y}_{o,i}$. If both of them agree on ‘‘O’’, the final prediction then follows. If only one of them is ‘‘O’’, we abide by the other prediction. When neither of them belong to ‘‘O’’, we compare the maximum entry for each vector and choose the largest as the final prediction.

In our unsupervised domain adaptation setting, the aspect and opinion labels are only available for the source domain, which we denote as $\mathcal{D}_S = \{(\mathbf{x}_n^S, \mathbf{y}_{a,n}^S, \mathbf{y}_{o,n}^S)\}_{n=1}^{N_S}$. For DAN and rDAN, we generate domain-labeled data by combining source and target domain as $\mathcal{D}_d^o = \{(\mathbf{m}_{o,n}^l, \mathbf{y}_n^d)\}_{l=1, n=1}^{L, N_S+N_T}$ and $\mathcal{D}_d^a = \{(\mathbf{m}_{p,n}^l, \mathbf{m}_{p,n}^{l+1}, \mathbf{y}_{p,n}^u)\}_{l=1, n=1, p \in \{a, o, ao\}}^{L, N_S+N_T}$. Furthermore, the auxiliary task for interaction prediction involves both source and target domains with corresponding relation labels as $\mathcal{D}_u = \{(\mathbf{m}_{p,n}^l, \mathbf{m}_{p,n}^{l+1}, \mathbf{y}_{p,n}^u)\}_{l=1, n=1, p \in \{a, o, ao\}}^{L-1, N_S+N_T}$. During training, the following loss functions are adopted:

$$\mathcal{L}_S = \sum_{\mathcal{D}_S} \sum_{p \in \{a, o\}} \ell(\mathbf{y}_{p,n}^S, \bar{\mathbf{y}}_{p,n}^S), \quad (19)$$

$$\mathcal{L}_d = \sum_{\mathcal{D}_d^o} \sum_{l=1}^L \ell(\mathbf{y}_n^d, \bar{\mathbf{y}}_{a,l,n}^d) + \sum_{\mathcal{D}_d^a} \sum_{l=1}^L \ell(\mathbf{y}_n^d, \bar{\mathbf{y}}_{o,l,n}^d), \quad (20)$$

$$\mathcal{L}_u = \sum_{\mathcal{D}_u} \sum_{l=1}^L \sum_{p \in \{a, o, ao\}} \ell(\mathbf{y}_{p,n}^u, \bar{\mathbf{y}}_{p,l,n}^u), \quad (21)$$

where ℓ is the cross-entropy loss. The final objective is

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_u - \beta \mathcal{L}_d, \quad (22)$$

where λ and β are trade-off parameters to control the impact of auxiliary task and adversarial networks. Let $\theta_f, \theta_S, \theta_d, \theta_u$ denotes the parameter set for feature learning, word prediction, domain prediction and auxiliary relation prediction, respectively. We adopt the Gradient Reversal Layer (GRL) (Ganin and Lempitsky 2015) for training DAN and rDAN. The updates can be carried in the following rules:

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial \mathcal{L}_S}{\partial \theta_f} + \lambda \frac{\partial \mathcal{L}_u}{\partial \theta_f} - \beta \frac{\partial \mathcal{L}_d}{\partial \theta_f} \right), \quad (23)$$

$$\theta_S \leftarrow \theta_S - \mu \left(\frac{\partial \mathcal{L}_S}{\partial \theta_S} \right), \quad (24)$$

$$\theta_u \leftarrow \theta_u - \mu \left(\lambda \frac{\partial \mathcal{L}_u}{\partial \theta_u} \right), \quad (25)$$

$$\theta_d \leftarrow \theta_d - \mu \left(\beta \frac{\partial \mathcal{L}_d}{\partial \theta_d} \right). \quad (26)$$

Dataset	Description	Sentences	Training	Testing
RES	Restaurant	5,841	4,381	1,460
LAP	Laptop	3,845	2,884	961
DEV	Digital Device	3,836	2,877	959

Table 1: Data statistics from each domain.

Here μ represents the learning rate and all parameters are shared across domains.

Experiments

Experimental Setup

Datasets Our experiments are conducted on three benchmark datasets with customer reviews including restaurant reviews, laptop reviews and reviews from digital device. Follow the setup from (Wang and Pan 2018), we combine restaurant reviews from SemEval 2014 task 4 subtask 1 (Pontiki et al. 2014) and SemEval 2015 task 12 subtask 1 (Pontiki et al. 2015) to form our restaurant domain. The laptop reviews consist of challenge dataset from the laptop domain in SemEval 2014 task 4 subtask 1. The device domain is a combination of the reviews from 5 different digital products provided by (Hu and Liu 2004). Detailed statistics are shown in Table 1. We use three different splits following (Wang and Pan 2018), where each split consists of 3/4 of the original data for training, and the rest constitutes the test set. The number of sentences for training and testing for each split is shown in Table 1. For precise comparison, we take the average results of 3 splits as the final performance.

Settings The input word embeddings are pre-trained using *word2vec* (Mikolov et al. 2013), with the training corpus being a combination of Yelp Challenge dataset² and electronics dataset from Amazon reviews³. The dimension of word vectors is set to be 100. As mentioned earlier, the input local memory for each word is obtained through a GRU network on its word embedding. We set the dimension of GRU-transformed vector to be 50. For 3-dimensional tensor parameters, we set the number of bilinear interactions as $k = 20$. For our experiments, the transferable interactive memory network consists of 3 layers. The trade-off parameters for adversarial loss and auxiliary loss are $\beta = 0.5$ and $\lambda = 1.0$. For training, we apply stochastic gradient descent to update all the parameters with 0.1 as the learning rate. To avoid over-fitting, partial dropout strategy is adopted with dropout rate set as 0.5 for non-recurrent GRU parameters as well as 3-dimensional tensors. All the experiments are conducted with 20 epochs and the best performance for each split is being recorded. Under the setting of unsupervised domain adaptation, no labels are available for the target domain. Hence, we only use the labeled training data from source domain and unlabeled training data from target domain to train all the models. And we investigate two different settings: transductive setting that tests the model on training data from the target domain and inductive setting

²<http://www.yelp.com/dataset/challenge>

³<http://jmcauley.ucsd.edu/data/amazon/links.html>

Models	RES→LAP		RES→DEV		LAP→RES		LAP→DEV		DEV→RES		DEV→LAP	
	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
TCRF	19.72	59.20	21.07	52.05	28.19	65.52	29.96	56.17	6.59	39.38	24.22	46.67
RAP	25.92	62.72	22.63	54.44	46.90	67.98	34.54	54.25	45.44	60.67	28.22	59.79
ARNN	33.66	-	33.20	-	48.10	-	31.25	-	47.97	-	34.74	-
RNSCN (r)	39.27	59.41	33.42	57.24	45.79	69.96	38.21	59.12	45.36	72.84	50.45	68.05
RNSCN	40.43	65.85	35.10	60.17	52.91	72.51	40.42	61.15	48.36	73.75	51.14	71.18
IMN	38.88	66.89	31.31	59.12	40.27	69.62	36.92	61.78	45.62	74.05	50.60	65.94
TIMN	43.68	68.44	35.45	59.05	54.12	73.69	38.63	62.22	53.82	76.52	52.46	69.32

Table 2: Comparisons in F1 scores with different baselines.

that tests the model on the test data in the target domain. The evaluations are conducted using F1 scores that only assign a true positive prediction to an exact match for any multi-word aspect or opinion term. We use Intel(R) Xeon(R) CPU E5-1650 v2 @ 3.50GHz for the experiments. Specifically, training 1 epoch with around 6000 sentences takes around 30 mins. The training usually converges at about 5 epochs.

Baselines To demonstrate the effectiveness of TIMN, we compare its performance with several baseline models for unsupervised domain adaptation, listed in the following:

- **TCRF**: Transferable CRF model proposed by Jakob and Gurevych (2010). It implements a linear-chain CRF for sequence prediction that focuses on non-lexical features, e.g., POS tags and dependency relations, that are shared across different domains.
- **RAP**: A cross-domain boosting method called Relational Adaptive bootstrapping (Li et al. 2012) that iteratively expand target lexicon according to pivot opinion words and syntactic relations.
- **ARNN**: A recurrent neural network with auxiliary tasks which are formed from manually-designed rules that inform the relations among aspect and opinion words (Ding, Yu, and Jiang 2017).
- **RNSCN**: The recently proposed recursive neural network by Wang and Pan (2018). It incorporates syntactic structure into the deep model and an auto-encoder to denoise relation labels generated from the parser.
- **RNSCN (r)**: Using the same model as RNSCN, but changing the dataset by randomly replacing dependency relations in the source domain that connect to an aspect or opinion word with 50% probability.
- **IMN**: Single-domain Interactive Memory Network. We directly use the model trained on labeled source domain to make predictions in the target domain.
- **TIMN**: The proposed Transferable Interactive Memory Network which learns a shared space through memory representations and memory interactions.

Experimental Results

The comparison results with various baselines are shown in Table 2 for inductive setting with average F1 scores over 3 splits. Clearly, traditional feature-engineering-based and boosting-based models TCRF and RAP are not good enough

to extract necessary information to be adapted in the target domain, compared to deep learning models. ARNN relies on the construction of auxiliary tasks, and hence is not flexible to adapt to specific dataset, especially for user-generated texts. Although the most recent method RNSCN proposed by (Wang and Pan 2018) addresses the limitation of manual-designed rules by incorporating syntactic information into a dependency-tree-based recursive neural network, the performance still relies on the accuracy of the dependency parser. This can be revealed when observing the performance of RNSCN (r). By randomly replacing some of the dependency relations obtained from the parser, it simulates the situation when the parser does not perform well, which is common for informal texts. In this case, the performance for RNSCN (r) is inferior than RNSCN using the same model, with large degradation for opinion predictions and aspect prediction for LAP→RES (7.12% drop). Without any transfer strategy, IMN is already able to achieve comparable performances for most of the experiments by simply sharing all the parameters learned from the source domain with the target domain. This indicates that the attention mechanism is transferable to some extent in the way it models interactions among aspect and opinion words through local-global memory interactions. To further exploit these shared information, we propose TIMN that explicitly project the interactions from both domains into a shared space. Without any external resources, TIMN achieves the state-of-the-art performances on most experiments. Specifically, it shows 3.25% and 2.59% improvements on aspect and opinion term predictions for RES→LAP, 5.46% and 2.77% improvements on aspect and opinion term predictions for DEV→RES.

To demonstrate the effect of each component in TIMN, we conduct ablation test for each transfer experiment. The results are shown in Table 3. The first row represents the model IMN where no transfer strategies are used except sharing parameters. The next three rows present the performances for each transfer component. At most situations, DAN is able to achieve better results for opinion term predictions, compared to the other two constituents, which proves that domain adversarial network is capable of aligning the opinion words from two domains for more accurate extraction. On the other hand, rDAN is more advantageous for identifying aspect terms by conditioning the information on corresponding opinions. The auxiliary task shows intermediate performance compared to DAN and rDAN. In general, the models combining any two transfer components outper-

Models	RES→LAP		RES→DEV		LAP→RES		LAP→DEV		DEV→RES		DEV→LAP	
	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
None	38.88	66.89	31.31	59.12	40.27	69.62	36.92	61.78	45.62	74.05	50.60	65.94
DAN	39.70	67.41	32.33	59.13	51.98	72.73	38.27	60.55	51.18	75.72	49.74	69.58
rDAN	42.43	67.72	32.90	58.12	52.33	71.19	38.08	61.30	53.49	76.15	49.79	69.97
aux	41.60	65.25	32.54	59.08	51.20	73.29	37.26	58.60	53.04	75.51	50.47	68.04
DAN+rDAN	42.82	67.91	33.34	57.66	52.46	72.75	37.91	59.79	51.29	74.72	50.58	69.76
DAN+aux	43.11	68.51	33.33	58.71	51.57	72.85	37.95	59.84	51.79	75.56	51.37	70.40
rDAN+aux	42.25	68.05	34.64	58.46	52.96	72.62	37.71	58.72	52.01	75.69	49.81	67.10
TIMN	43.68	68.44	35.45	59.05	54.12	73.69	38.63	62.22	53.82	76.52	52.46	69.32

Table 3: Performances for different combinations of each transfer component in terms of F1 scores.

		RES→LAP		RES→DEV		LAP→RES		LAP→DEV		DEV→RES		DEV→LAP	
		AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
OUT	1 layer	36.69	65.64	32.98	59.55	46.28	71.95	37.88	61.87	49.21	76.03	51.14	68.75
	2 layers	42.65	67.73	31.28	57.20	50.26	72.11	38.75	61.49	51.86	76.09	51.81	68.02
	3 layers	43.68	68.44	35.45	59.05	54.12	73.69	38.63	62.22	53.82	76.52	52.46	69.32
	4 layers	43.38	68.10	35.04	58.44	52.28	73.45	38.27	59.64	53.63	75.97	52.34	69.05
IN	1 layer	34.37	64.89	31.29	58.89	47.32	72.14	39.22	61.27	47.80	76.89	50.33	68.40
	2 layers	39.27	67.42	32.07	57.23	51.69	71.60	40.30	62.13	51.58	76.69	50.34	69.81
	3 layers	40.28	67.64	34.45	59.36	52.24	72.85	38.85	60.78	53.94	76.25	52.21	70.13
	4 layers	39.49	67.05	34.39	59.01	51.77	72.73	38.62	61.70	53.11	76.43	51.87	70.05

Table 4: Comparisons with different number of layers and settings in terms of F1 scores.

form those independent ones, which shows each constituent actively affect the learning process of the other. For example, if DAN helps to align opinion words across domains, rDAN could in turn make more accurate extraction for aspect terms according to the interactions between aspect memory and opinion memory. Without doubt, the joint model that combines all 3 components attains the best performance.

The proposed TIMN model consists of multiple layers to iteratively refine the global memory representations and the transformed hidden vectors for each word. The multi-layer architecture is also crucial for modeling intra-correlations among aspect or opinion words across domains. To prove that, we conduct experiments with different number of layers as shown in Table 4. In most cases, increasing the number of layers from 1 to 3 demonstrates gradual improvements for extraction with the best performance attained using 3 layers. We conjecture that the reason for degradation with 4 layers might be caused by over-extraction when the previous 3 layers are already sufficient to exploit all necessary target information. Note that the model with only 1 layer does not involve global memory interactions for auxiliary task, because there does not exist two consecutive layers to compute the interaction. Hence, it perform much worse compared to other results in terms of aspect extraction. Table 4 also presents results for both inductive (OUT) and transductive (IN) settings. We shall observe little difference in the performances between these two settings, which shows the model’s robustness to guarantee comparable results when the data for evaluation is not presented during training.

The robustness of the proposed model is illustrated in Figure 3. Specifically, we conduct experiments varying the number of bi-linear interactions (k) in tensor operation for

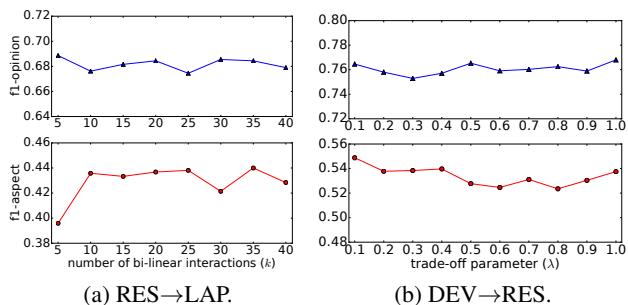


Figure 3: Sensitivity studies for hyper-parameters.

RES→LAP as well as changing the trade-off parameter for auxiliary loss (λ) for DEV→RES. The plots demonstrate stableness of TIMN over different hyper-parameters.

Conclusion

This paper presents a novel memory network that emphasizes the complex correlations among aspect words and opinion words which are summarized into 3 categories. The model not only exploits local-global memory interactions for attending target terms at each domain, but also associates global memories and their interactions for aligning source and target spaces. The proposed model could achieve promising results without the dependency over any external resources compared with existing methods.

Acknowledgements

This work is supported by NTU Singapore Nanyang Assistant Professorship (NAP) grant M4081532.020, Singapore

References

- Blitzer, J.; Dredze, M.; and Pereira, F. 2007. Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *ACL*, 187–205.
- Bollegala, D.; Maehara, T.; and ichi Kawarabayashi, K. 2015. Unsupervised cross-domain word representation learning. In *ACL*, 730–740.
- Chen, M.; Xu, Z.; Weinberger, K. Q.; and Sha, F. 2012. Marginalized denoising autoencoders for domain adaptation. In *ICML*, 1627–1634.
- Ding, Y.; Yu, J.; and Jiang, J. 2017. Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. In *AAAI*, 3436–3442.
- Ganin, Y., and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *ICML*, volume 37, 1180–1189.
- Glorot, X.; Bordes, A.; and Bengio, Y. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*, 97–110.
- He, R.; Lee, W. S.; Ng, H. T.; and Dahlmeier, D. 2017. An unsupervised neural attention model for aspect extraction. In *ACL*, 388–397.
- Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *KDD*, 168–177.
- Jakob, N., and Gurevych, I. 2010. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *EMNLP*, 1035–1045.
- Jin, W., and Ho, H. H. 2009. A novel lexicalized hmm-based learning framework for web opinion mining. In *ICML*, 465–472.
- Li, X., and Lam, W. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *EMNLP*, 2886–2892.
- Li, F.; Han, C.; Huang, M.; Zhu, X.; Xia, Y.-J.; Zhang, S.; and Yu, H. 2010. Structure-aware review mining and summarization. In *COLING*, 653–661.
- Li, F.; Pan, S. J.; Jin, O.; Yang, Q.; and Zhu, X. 2012. Cross-domain co-extraction of sentiment and topic lexicons. In *ACL*, 410–419.
- Li, Z.; Zhang, Y.; Wei, Y.; Wu, Y.; and Yang, Q. 2017. End-to-end adversarial memory network for cross-domain sentiment classification. In *IJCAI*, 2237–2243.
- Liu, K.; Xu, L.; Liu, Y.; and Zhao, J. 2013. Opinion target extraction using partially-supervised word alignment model. In *IJCAI*, 2134–2140.
- Liu, P.; Joty, S.; and Meng, H. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *EMNLP*, 1433–1443.
- Lu, Y.; Zhai, C.; and Sundaresan, N. 2009. Rated aspect summarization of short comments. In *WWW*, 131–140.
- Mei, Q.; Ling, X.; Wondra, M.; Su, H.; and Zhai, C. 2007. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *WWW*, 171–180.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*, 3111–3119.
- Mnih, V.; Heess, N.; Graves, A.; and Kavukcuoglu, K. 2014. Recurrent models of visual attention. In *NIPS*, 2204–2212.
- Pan, S. J.; Ni, X.; Sun, J.-T.; Yang, Q.; and Chen, Z. 2010. Cross-domain sentiment classification via spectral feature alignment. In *WWW*, 751–760.
- Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutopoulos, I.; and Manandhar, S. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *SemEval*, 27–35.
- Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; and Androutopoulos, I. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *SemEval 2015*, 486–495.
- Popescu, A.-M., and Etzioni, O. 2005. Extracting product features and opinions from reviews. In *EMNLP*, 339–346.
- Qiu, G.; Liu, B.; Bu, J.; and Chen, C. 2011. Opinion word expansion and target extraction through double propagation. *Comput. Linguist.* 37(1):9–27.
- Titov, I., and McDonald, R. 2008. Modeling online reviews with multi-grain topic models. In *WWW*, 111–120.
- Wang, W., and Pan, S. J. 2018. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In *ACL*, 2171–2181.
- Wang, W.; Pan, S. J.; Dahlmeier, D.; and Xiao, X. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. In *EMNLP*, 616–626.
- Wang, W.; Pan, S. J.; Dahlmeier, D.; and Xiao, X. 2017. Coupled multi-layer tensor network for co-extraction of aspect and opinion terms. In *AAAI*, 3316–3322.
- Weston, J.; Chopra, S.; and Bordes, A. 2015. Memory networks. In *ICLR*.
- Xu, H.; Liu, B.; Shu, L.; and Yu, P. S. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. In *ACL*, 592–598.
- Yin, Y.; Wei, F.; Dong, L.; Xu, K.; Zhang, M.; and Zhou, M. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *IJCAI*, 2979–2985.
- Yu, J., and Jiang, J. 2016. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. In *EMNLP*, 236–246.
- Zhang, L.; Liu, B.; Lim, S. H.; and O’Brien-Strain, E. 2010. Extracting and ranking product features in opinion documents. In *COLING*, 1462–1470.
- Zhang, M.; Zhang, Y.; and Vo, D. T. 2015. Neural networks for open domain targeted sentiment. In *EMNLP*.
- Zhou, G.; Xie, Z.; Huang, J. X.; and He, T. 2016. Bi-transferring deep neural networks for domain adaptation. In *ACL*, 322–332.
- Zhuang, L.; Jing, F.; and Zhu, X.-Y. 2006. Movie review mining and summarization. In *CIKM*, 43–50.