



OPEN

Context aware semantic adaptation network for cross domain implicit sentiment classification

Enguang Zuo¹, Alimjan Aysa^{1,2}, Mahpirat Muhammat¹, Yuxia Zhao^{1,3} & Kurban Ubul^{1,2}✉

Cross-domain sentiment classification could be attributed to two steps. The first step is used to extract the text representation, and the other is to reduce domain discrepancy. Existing methods mostly focus on learning the domain-invariant information, rarely consider using the domain-specific semantic information, which could help cross-domain sentiment classification; traditional adversarial-based models merely focus on aligning the global distribution ignore maximizing the class-specific decision boundaries. To solve these problems, we propose a context-aware semantic adaptation (CASA) network for cross-domain implicit sentiment classification (ISC). CASA can provide more semantic relationships and an accurate understanding of the emotion-changing process for ISC tasks lacking explicit emotion words. (1) To obtain inter- and intrasentence semantic associations, our model builds a context-aware heterogeneous graph (CAHG), which can aggregate the intrasentence dependency information and the intersentence node interaction information, followed by an attention mechanism that remains high-level domain-specific features. (2) Moreover, we conduct a new multigrain discriminator (MGD) to effectively reduce the interdomain distribution discrepancy and improve intradomain class discrimination. Experimental results demonstrate the effectiveness of different modules compared with existing models on the Chinese implicit emotion dataset and four public explicit datasets.

Sentiment analysis is considered one of the fundamental problems in natural language processing (NLP), and with social media developing rapidly, it is widely applied in real scenarios such as comment analysis, food safety monitoring, and public opinion mining. Such tasks are usually defined as identifying the emotional polarity (e.g., positive, negative, or neutral) of a given text, sentence, or aspect.

The expression of emotion can be explicit or implicit. The implicit expression of emotions is defined as 'A language fragment (sentence, clause or phrase) that expresses subjective sentiment but contains no explicit sentiment word'^{1,2}. We exploit the following examples to show the difference of two expressions:

Explicit: 你做的蛋炒饭太好吃了,我很喜欢! (English translation: "The rice fried with eggs is so delicious; I like it very much!" label=positive)

Implicit: 这家蛋炒饭有种妈妈的味道! (English translation: "The rice fried with eggs in this restaurant reminds me of my mother!" label=positive)

Example 1 uses the word '喜欢 (like)' to show a clear positive tendency. To express their views, people also use Example 2 (e.g., metaphor, sarcasm). In this sentence, no explicit emotional words are used and the individual's emotional tendency is embedded in the semantic meaning of the text. This phenomenon creates an exceptional challenge for implicit sentiment classification tasks. Moreover, accompanied by the absence of a large-scale labeled corpus, even with an advanced deep learning model, the classification accuracy of implicit sentiment classification tasks is not ideal³.

One solution to this problem is cross-domain sentiment classification, which aims to exploit the rich labeled data in the source domain, e.g., explicit sentiment corpus, to help the sentiment analysis task in another domain lacking for or even without labeled data, e.g., implicit sentiment corpus. Recently, relevant models on

¹College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China. ²Xinjiang Multilingual Information Technology Key Laboratory, Urumqi 830046, China. ³School of Mathematics and Computer Applications Shangluo University, Shangluo 726000, China. ✉email: Kurbanu@xju.edu.cn

Source domain: 我经常感冒，因为我抵抗力差。(I often get a cold, because I have weak resistance.)

Target domain: 我对你就如此没有抵抗力吗？(Am I so resistless to you?)

Figure 1. Example of domain-invariant and domain-specific. The sentiment expressions marked by red lines are virus domain-specific, while the broken blue lines marked domain-invariant.

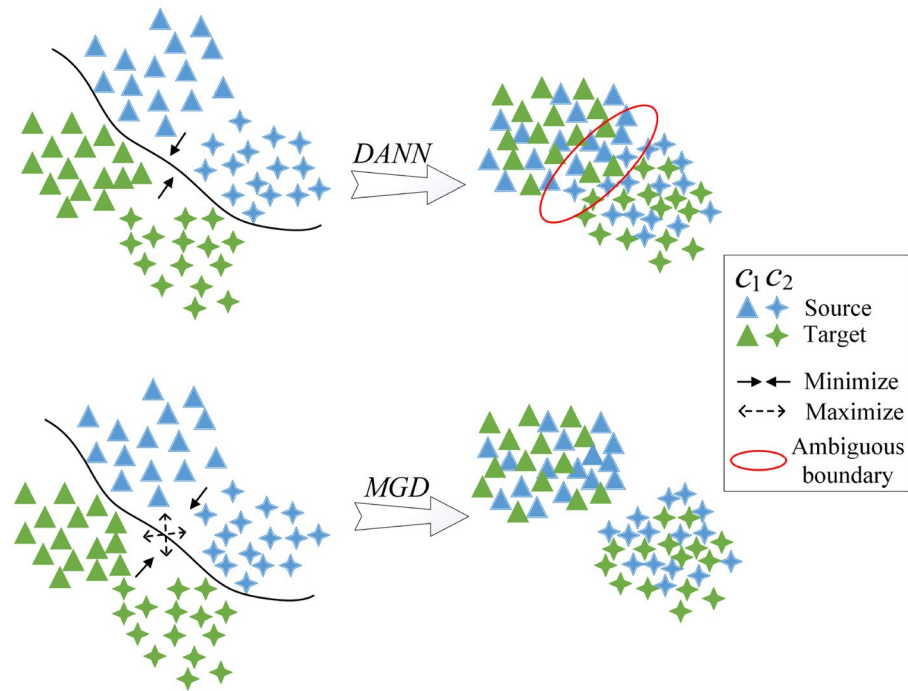


Figure 2. A comparison between the traditional domain discriminator DANN and the proposed MGD, where minimization means that the distribution difference is minimized in different domains and maximization means that the distribution difference is maximized between different classes, which come from different domains.

cross-domain sentiment classification have mainly focused on learning domain-invariant features whose distribution is similar in the source and target domains⁴. These methods attempt to reduce the discrepancy between domain-specific latent feature representations. Inspired by this idea, most existing adversarial-based methods, e.g., domain adversarial neural network (DANN)⁵, reduce feature differences by fooling a domain discriminator and have achieved promising results^{6,7}. However, to achieve explicit-to-implicit positive transfer, these methods still have two major inherent drawbacks that need to be addressed:

- Existing studies mostly focus on learning the domain-invariant information (e.g. ‘喜欢 (like)’, ‘坏 (bad)’, ‘差 (weak)’), rarely consider the usage of domain-specific semantic information (e.g. ‘蛋炒饭 (the rice fried with eggs)’, ‘抵抗力 (resistance)’), which is also helpful for cross-domain implicit sentiment classification^{7,8}. Figure 1 shows that when the source domain-specific words appear in the target domain, the semantic knowledge learned from the source domain helps the target domain classification.
- Traditional adversarial-based models merely minimize the marginal distribution of the two domains and ignore maximizing the class-specific decision boundaries. As shown in Fig. 2 (DANN), the features near the decision boundary may be ambiguous and even tangled together with traditional domain discriminator training, thus blocking adaptation performance.

To tackle the above limitations identified above, we aim to use graph convolutional networks (GCNs). GCNs have a multilayer architecture, with each layer aggregating the information of nodes in the graph structure using features of immediate neighbors. Nevertheless, sequential free texts are unstructured data. Therefore, GCN-based text learning must conduct graph representation learning from the free text before graph convolution. Different from sequence learning models, GCNs can directly represent complex structured data. A GCN has the potential to capture domain-specific semantic information with GCN layers. Recently, GCN models have

gained widespread attention and have been successfully deployed on text-word relationships^{9–11}, and explicit sentiment analysis^{12,13}. However, these graph-based models only considered intrasentence hierarchical dependency relationships and ignored intersentence semantic associations.

Therefore, in this paper, we propose a novel context-aware semantic adaptation network (CASA) for cross-domain implicit sentiment classification via GCNs. To obtain inter- and intrasentence semantic associations, we build a context-aware heterogeneous graph (CAHG). CAHGs build graphs in each document by regarding tokens and sentences as nodes (hence heterogeneous graph). The intrasentence propagation is constrained by the syntactic dependency tree, and intersentence propagation is constrained by the sentence-free sequence and term frequency-inverse document frequency (TF-IDF) local token cooccurrence information. The information propagates inter- and intrasentences via the GCN layers, followed by an attention mechanism that keeps high-level domain-specific features. We also conduct a multigrain discriminator (MGD) which is imposed during domain adaptation to minimize domain distribution and maximize class identification. The domain adaptation layer makes the source domain and target domain inseparable through adversarial training, reduces the representation distribution gap between source and target domain data in a coarse-grained manner. The class adaptation layer utilizes a classifier to judge different domain samples, whether class consistent, to distinguish samples in different classes. Figure 2 illustrates the difference between the traditional domain adversarial method of domain-adversarial training of DANN⁵ and MGD.

In short, the main contributions of this paper are as follows:

1. We proposed a new transfer learning model CASA for the implicit sentiment classification task via GCNs, which is the first attempt to transfer explicit sentiment information to implicit sentiment.
2. Our model provides a context heterogeneous graph, which can effectively extract inter- and intrasentence semantic information. Moreover, CASA improved the model's generalization ability for implicit classification tasks and increased each sentiment's polarity discrimination in the domain.
3. We evaluate CASA on the Chinese implicit sentiment analysis dataset (SMP-ECISA 2019). CASA outperforms existing models in four different source domains. We also provide a visualization to demonstrate that CAHG can capture domain-specific information, and MGD can make features near decision boundaries more distinguishable.

Related works

Sentiment analysis. The existing sentiment calculation and sentiment analysis methods can be divided into three categories: knowledge-based methods, statistical methods and hybrid methods¹⁴.

Knowledge-based methods are popular because of their simplicity and ease of use, but their effectiveness is limited mainly by the depth and breadth of the established knowledge base. Statistical methods based on machine learning and deep learning have been widely used in Chinese sentiment classification, such as fine-grained sentiment analysis framework¹⁵, multi-label sentiment analysis model¹⁶, aspect-level sentiment analysis research based on Reinforcement Learning¹⁷, etc. On the other hand, more and more scholars have realized the particularity of Chinese characters and tried to model Chinese radicals^{18–22}. The hybrid method aims to describe the rules of emotional expression better and realize the machine's perception of semantics²³. The model CASA in this paper considers contextual semantic perception and introduces cross-domain explicit emotional knowledge.

Attention mechanism shows good performance in sentiment analysis tasks. It improves the interpretability of neural networks by letting people know the location of the focus^{24–27}. They exemplified recent research on attention-based sentiment analysis.

Various sentiment analysis tasks usually focus on realizing binary classification (positive and negative classification), which cannot better describe emotions. In contrast, Wang proposed a multi-level emotion perception method with contradiction processing²⁸. However, these sentence-level sentiment analyses cannot be directly applied to implicit sentiment tasks because, in implicit sentiment tasks, the emotion of the target sentence holds different polarities for different contexts.

Implicit sentiment analysis and GCNs. Models based on RNNs^{29,30} and CNNs³¹ in deep neural networks are widely used in sentiment classification tasks. In RNN-based models, an attention mechanism is usually introduced because each word in the text contributes differently to the classification task^{32–34}. CNN-based models^{35,36} use character-level CNNs to extract semantic information from text. However, these models lack an effective mechanism to capture the information in the dependency tree structure. In works of^{37–39}, structural and semantic information extracted from the tree structure of sentences, such as a dependency tree or grammar tree by LSTM or BiLSTM, was used for the sentiment classification task. Although Tree-LSTM can extract more accurate semantic information from text, it is difficult to perform parallel computing and requires a longer time to train. After that, CNNs were introduced into the tree structure information encoding process by^{40,41}. In their work, a phrase structure tree and syntax-dependent tree were used to encode the semantic information of the target sentence and the context, respectively. However, in the above tree-based convolutional neural network model, sentences are considered to be independent of each other, so information regarding the relationship between sentences is lost. To solve this problem, our model extracts semantic information via GCNs⁹.

GCN models have attracted widespread attention and have been successfully deployed in NLP tasks. Yao et al.¹⁰ builds a corpus-level text graph by word-word co-occurrence and document-word relations for text classification. Zhang et al.⁴² introduces a TreeGCN, where the GCN is used to encode the dependency syntactic structure. Zhang et al.¹² presented an aspect-based GCN to demonstrate that GCNs can achieve long-range word dependency. Zhang et al.⁴³ employ gated graph neural networks document-level graph word interaction.

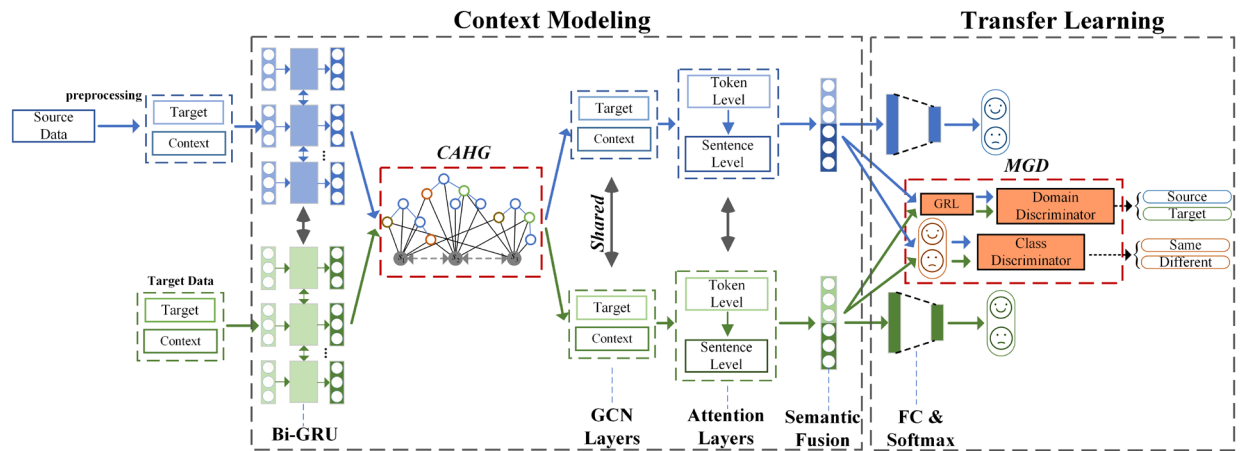


Figure 3. Overview of the context-aware semantic adaptation network. CAHG represents the context background structure in both the source and target domains. Attention layers are hierarchical attention, token-level of target sentence, and sentence-level of context background. There are two layers of GCNs. MGD represents the proposed multigrain discriminator. Details on the CAHG and MGD are discussed in later sections.

In contrast to their works, we regard the tokens and sentences in each document as graph nodes. The graph maintains inter- and intrasentence constraints to capture semantic information. It can obtain more accurate text semantics while increasing the interpretability of the model.

Transfer learning in sentiment analysis. Even with a strong deep learning model, the classification accuracy of implicit emotion classification problems is not ideal in the absence of sufficient labeled data^{34,41}. Yosinski et al.⁴⁴ discusses the application of transfer learning in deep neural networks for single domains. It draws an important conclusion: adding fine-tuning will improve the performance of the deep transfer network. In the cross-domain scenario, the difference between the probability distribution of the source domain's data and the target domain's data is significant. The use of fine-tuning alone may lead to negative transfer. The purpose of domain adaptation, also known as cross-domain learning, is to reduce the difference in distribution representation of the data of the source and target domains. In domain adaptation, a direct way to solve the above problem is to use a certain distribution distance measurement method to measure the distance between distributions and reduce the distance in the model training process. However, the calculation of distance measurements is difficult. Many methods based on domain adversarial models have been proposed^{5,6}.

However, existing domain adversarial models for sentiment analysis focus on explicit sentiment classification or aspect-level sentiment classification^{45,46} without considering implicit situations. Due to data scarcity and the task's value, transfer learning is more urgent for implicit sentiment analysis. To the best of our knowledge, CASA is the first explicit-to-implicit transfer learning model. MGD improved the model's generalization ability to implicit emotion classification and the discrimination of each sentiment polarity.

Methods

Problem definition and approach overview. *Domain definition.* A domain D consists of marginal distribution $P(x)$ and m -dimensional feature space X , namely $D = \{P(x), X\}$.

Task definition. Given the domain D , the task is composed of a classifier $f(x)$ and a class label set Y , namely $T = \{f(x), Y\}$, where $f(x) = Q(y|x)$ represent the conditional probability distribution, $y \in Y$.

Purpose. Given a training data come from source domain $D_s = \{(x_i, y_i)\}_{i=1}^{n_s}$ and target domain $D_t = \{(x_{n_s+i}, y_{n_s+i})\}_{i=1}^{n_t}$, we assume that there is a difference between the probability distribution $P(X^s)$ and $P(X^t)$. under this settings, the purpose is to learn a prediction function $f(x) : X \rightarrow Y$ that classify the target examples correctly during testing step.

The overview of the CASA framework is described in Fig. 3. As shown in Fig. 3, the proposed approach is divided into two steps: context modeling and transfer learning. Specifically, we first built CAHG in the context modeling phase to extract the semantic relationships unrelated to explicit and implicit sentiment domains. In the transfer learning step, MGD realizes fine-grained adaptation through a domain discriminator and label consistency discriminator so that the model has stronger generalization ability. We present the details of different components as well as the training process in the following section.

Bi-GRU encoder. First, we use the bidirectional gated recurrent unit (Bi-GRU)⁴⁷ to encode the text input from the source or target domain to obtain contextualized word-level representation $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_i, \dots, \mathbf{h}_m]$, where m denotes the vector dimension and $\mathbf{h}_i \in R^m$ is the hidden layer state vector at time i . The reason for

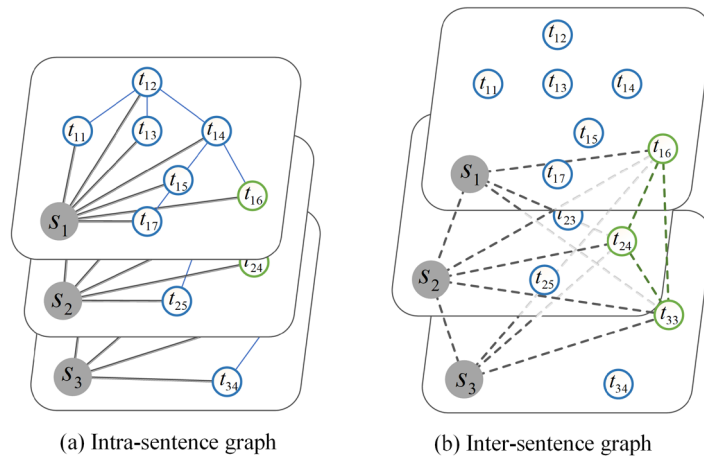


Figure 4. A tony demo of the CAHG. S_i represents for the i -th sentence, t_{ij} represents for the j -th token in the i -th sentence, the intrasentence edges between the tokens are dependency relations, and the green color represents these tokens at least part of the Chinese characters are the same.

employing this layer is to correct the information of the syntactic dependency tree, which was built by HANLP. We use HANLP for this: <https://www.hanlp.com>.

GCNs over the context-aware heterogeneous graph. To express the text’s more valuable information, we construct a document-level heterogeneous graph, which retains intrasentence dependency information, and has intersentence relationship representation. We define $G = (V, E)$ as a text graph, where $|V| = n$ is the number of nodes and V, E represents the node set and the edge set of graph G , respectively. We construct G based on the token-token dependency relationship, TF-IDF, and sentence order. The token’s TF-IDF determines the edge weight between the sentence node and the token node in the sentence. We define TF as the number of words that appear in the sentence, and IDF as the logarithm of the overall sentence number in the text to the sentence number, which contains the token. The formal definition is as follows formula (1–3).

$$TF - IDF = TF \times IDF \tag{1}$$

$$TF_w = \frac{S_w}{S} \tag{2}$$

$$IDF_w = \log\left(\frac{Doc}{Doc_w + 1}\right) \tag{3}$$

Among them, S_w and S represent the number of token occurrences in the input sentence and the sentence token’s total number, respectively. Doc_w and Doc represent the number of sentences in the input text with the token and the total number of sentences, respectively.

Meanwhile, we introduce the sentence sequential features to represent the sentence-sentence relationship. The matrix $\mathbf{H} \in R^{m \times n}$ consists of the feature vector $\mathbf{h}_i \in R^m$ of n nodes, and the adjacency matrix $A \in R^{m \times n}$ is used to represent the weights between nodes in graph G . The relationships of the edges between nodes p and q are formally defined as function (4–5):

$$A_{p,q} = \begin{cases} D(p, q) & p, q \text{ all tokens} \\ TF - IDF_{p,q} & p \text{ is sentence, } q \text{ is token} \\ Order(p, p + 1) & p \text{ is sentence} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$D_{p,q} = \begin{cases} Tree(p, q) & p, q \text{ in same sentence} \\ \frac{N_{syntactic}}{N_{total}} & p, q \text{ in different sentence} \end{cases} \tag{5}$$

$Order(p, p + 1)$ denotes sentence $p \rightarrow p + 1$ ’s natural reading order in the text. $Tree(p, q)$ is the relationship between token nodes p and q in the dependency tree. $N_{syntactic}$ represents the number of times that the tokens p and q have a relationship in the current text (the relationship means that p and q in at least part of the Chinese characters are the same), and N_{total} is the number of times that p and q appear throughout the dataset. To facilitate the description of the process of information transfer in the graph, we divide CAHG into two subgraphs in Fig. 4. One of the subgraphs describes the transfer of information within sentences, and the other describes the transfer of information across sentences.

The information propagates inter- and intrasentences via the GCN layers, followed by a hierarchical attention layer that keeps high-level domain-specific features. The final text representation is defined as follows:

$$\alpha_{t,i} = \frac{\exp(\omega_t^\top \tanh(\mathbf{W}_t \mathbf{H}_{t,i}^G + \mathbf{b}_t))}{\sum_{i'} \exp(\omega_t^\top \tanh(\mathbf{W}_t \mathbf{H}_{t,i'}^G + \mathbf{b}_t))} \quad (6)$$

$$\mathbf{O}_t = \sum_i \alpha_{t,i} \mathbf{H}_{t,i}^G \quad (7)$$

$$\mathbf{r} = (\mathbf{O}_t \oplus \mathbf{O}_c) \quad (8)$$

In formula (6–8), \mathbf{r} is the final representation of the text, which is composed of the target sentence \mathbf{O}_t and the relevant context \mathbf{O}_c . \oplus represents the splice operation. \mathbf{H}_t^G represents the output from two-layer GCNs⁵. Then \mathbf{H}_t^G through a one-layer multilayer perceptron. Finally, \mathbf{O}_t is obtained by weighted summation of the MLP output. The sentence-level attention mechanism \mathbf{O}_c is a mirror of the word-level attention mechanism.

Multi-grain discriminator. Ben-David and Ganin proved that to perform domain adaptation to reduce the target domain's prediction error, maximizing the discriminator error between the source and target is necessary^{5,48,49}. Therefore, adversarial-based transfer learning is widely used to solve the domain adaptation problem. Although it has considerable advantages, we observed that the domain discriminator could only reduce the marginal distribution distance.

However, the relationship between marginal distributions and conditional distributions is uncertain. As indicated in⁵⁰, minimizing the difference between conditional distributions is critical to the robustness of distribution adaptation. We propose MGD, which consists of a domain discriminator T and an emotional polarity discriminator D . T makes the source domain and target domain not separable through domain confrontation. D maximizes the difference between different labels through label consistency identification. When the class discriminator D can accurately identify whether the sample labels from the source domain and the target domain are the same, the model can learn the class invariant features from the two domains. The formula is derived as follows:

$$G_T = \text{softmax}(\mathbf{W}_T \mathbf{r} + \mathbf{b}_T), \cdot \in (s, t) \quad (9)$$

$$L_{T_i} = -Y_i^D \log(G_T(\mathbf{r}_i)) - (1 - Y_i^D) \log(G_T(\mathbf{r}_i)) \quad (10)$$

where G_T denotes the domain discriminator, $G_T(\cdot)$ is the output prediction labels, L_{T_i} is the classification error, $Y^D = [s, t]$ is a set of domain labels, and s and t represent the source and target domains, respectively.

To maximize the extraction of domain invariant features, we hope to maximize the discrimination error G_t . As suggested in⁵, this min-max game is implemented by a gradient reversal layer. Specifically, when the network is undergoing the gradient back-propagation process, we will change ∇L_T into $-\eta \nabla L_T$. $\eta > 0$ is a controllable hyper-parameter.

$$G_D = \text{softmax}(\mathbf{W}_D (\mathbf{r}_s \oplus \mathbf{r}_t) + \mathbf{b}_D) \quad (11)$$

$$L_{D_i} = -I \log(G_D(\mathbf{r}_i)) - (1 - I) \log(G_D(\mathbf{r}_i)) \quad (12)$$

where I denote an indicator function:

$$I = \begin{cases} 1 & \text{label}(\mathbf{r}_s) = \text{label}(\mathbf{r}_t) \\ 0 & \text{label}(\mathbf{r}_s) \neq \text{label}(\mathbf{r}_t) \end{cases} \quad (13)$$

Here, G_D denotes the class-consistency discriminator, $G_D(\cdot)$ is the output prediction labels, and L_{D_i} is the classification error. Note that as L_D drops, the distribution difference also decreases in the same sentiment polarity between different domains. Finally, G_D enhances the identifiability of each sentiment polarity.

Sentiment classification and training. The vector \mathbf{r} obtained by the feature extractor is sent to the cascade of the fully connected layer and softmax layer to generate class distributions. The formula is as follows:

$$P = \text{softmax}(\mathbf{W}_p \mathbf{r} + \mathbf{b}_p) \quad (14)$$

where $P \in R^C$ represents the predicted soft distribution, C is the number of classifications, and $\mathbf{W}_p \in R^{c \times m}$ and b_p represent training weights and offsets. The cross-entropy has a loss function, given as:

$$L_C = \sum_{i \in D} \sum_{j=1}^C Y_{ij} \log P_{ij} \quad (15)$$

D is the document's index with the label, and Y is the real label matrix. Therefore, we can obtain the classification loss of the source domain L_{C_s} and the classification loss of the target domain L_{C_t} .

Dataset		Documents	AvgLength	T2TEdges	S2SEdges	S2TEdges	Classes
Explicit	Virus	7146	35.46	63.47	2.11	44.67	2
	Usual	22019	39.37	64.8	1.85	45.62	2
	Weibo	119999	40.99	72.32	2.83	51.34	2
	Hotel	9998	88.78	187.03	5.02	129.7	2
Implicit	SMP_2019	9733	51.81	111.51	4.44	78.95	2

Table 1. Statistics of the target domain and source domain datasets. The AvgLength denotes the average number of tokens in each text; the T2TEdges, the S2SEdges, and the S2TEdges indicate the average number of times token-token, sentence-sentence, sentence-token edges occurrences in each text, respectively.

$$D_{KL}(P_s \parallel P_t) = \sum_{i=1}^{N_s} \sum_c^C p_s^c(\mathbf{r}_i) \log \left(\frac{p_s^c(\mathbf{r}_i)}{p_t^c(\mathbf{r}_i)} \right) \quad (16)$$

Before the training stage, we were motivated by⁴², who proposed mutual learning in supervised single-domain tasks. The Kullback Leibler divergence of the predicted source class distribution and predicted target class distribution is calculated and vice versa. The two KL divergences measure the similarity of the two distributions. Finally, the overall loss function in CASA consists of both source and target loss, which are given as follows: $L_s = L_{C_s} - \eta_T L_{T_s} + \eta_D L_{D_s} + \lambda D_{KL}(P_s \parallel P_t)$ and $L_t = L_{C_t} - \eta_T L_{T_t} + \eta_D L_{D_t} + \lambda D_{KL}(P_t \parallel P_s)$. Here, η_T , η_D and λ are hyperparameters.

Experiments

Datasets and evaluation indicator. Our target domain data set is The Evaluation of Chinese Implicit Sentiment Analysis task in SMP2019 (one of the top academic conferences on social media processing in China). The dataset contains two types of content in each document: context and target sentence. We chose four benchmark datasets of explicit sentiments as the source domain. They are the Weibo-60000 dataset, the hotel review dataset, the SMP-2020's virus Weibo dataset, and the SMP-2020's general Weibo dataset. The data is available at: <http://biendata.com/competition/smpecisa2019/>, <http://www.pudn.com/Download/item/id/3993718.html>, <http://www.searchforum.org.cn/tansongbo/corpus1.php>, <https://smp2020.aconf.cn/>.

To clean the dataset, we performed some preprocessing. The construction of a heterogeneous graph requires contextual sentence nodes and a dependency tree structure. We performed the following preprocessing, and Table 1 summarizes the statistics:

- (1) To keep intact the dependent syntax structure, filter out sentences, which have no subject-predicate structure.
- (2) As suggested by^{41,51}, sentiment polarity consistency exists between the context semantic background and the target sentence. To match the target domain data granularity, we randomly select a sentence in each source domain document as the target sentence and the rest as the context.

We compute every model classification accuracy and F1 score in test dataset as evaluation indicator. The F1 score calculation and accuracy are shown as follows:

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i} \quad (17)$$

$$\text{Accuracy} = \frac{|P(x) = Y(x) \subseteq x : x \in D_t|}{|x : x \in D_t|} \quad (18)$$

where P_i and R_i mean the precision and recall of i -th sentiment polarity. After the above equation, we can through calculate metrics for each label, and find their unweighted mean getting macro-F1 score, i.e. macro-F1 = $\frac{1}{N} \sum_{i \in N_i} F1_i$. $P(x)$ is the predicted label and $Y(x)$ is the actual label of sample x , respectively.

Models for comparison. To fully verify and understand CASA, we divide the baseline models into two groups for comparison:

Non-transfer. To demonstrate the benefits from heterogeneous graphs, we compare with the following methods without transfer:

- *TextCNN*³¹, *TextRNN*⁵² and *BiLSTM+Att*⁵³: These are the basic deep neural networks in sentiment classification.

Systems		P	R	macro-F1	Acc
Non-Transfer	TextCNN	0.7543	0.7544	0.7543	0.7544
	TextRNN	0.7248	0.7237	0.7238	0.7243
	BiLSTM+Att	0.7625	0.7626	0.7625	0.7626
	TreeLSTM	0.7364	0.7279	0.7278	0.7317
	TreeGCN	0.8119	0.8045	0.8043	0.8057
	CASA-T-D	0.8266	0.8282	0.8256 [◇]	0.8259 [◇]

Table 2. Model comparison results. The state-of-the-art result of each evaluation indicator is bolded. The marker [◇] refers to p-value < 0.05 when compared with TreeGCN in the paired t-test. All models run over five times with random initializations and are report average precision, recall, macro-F1, and accuracy.

- *TreeLSTM*⁵⁴: An LSTM network based on a tree structure, which solves the problem of the emotional classification of nonlinear systems such as dependent trees.
- *TreeGCN*³⁹: BiLSTM is used to encode the input word vector to obtain the hidden state with context information. It then uses a GCN convolution to obtain the neighboring node information, which enhances the GCN's robustness.
- *CASA-T-D*: The CASA feature extractor part for examining the ability of CAHG to express text information.

Transfer. To investigate the effectiveness of each part in the CASA, we also compare the following frameworks for experiments. For a fair comparison, we use CASA-T-D as a feature extractor in other methods.

- *Fine-tuning*: Initialize the CASA-T-D parameters randomly, then train on the source domain dataset, and finally fix the parameters and fine-tune the model in the target domain dataset.
- *DANN*⁺⁵: The model adopts the idea of domain adversarial, which has a feature extractor and a domain discriminator.
- *CCSA*⁵⁵: A unified framework for supervised domain adaptation is created.
- *d-SNE*⁵⁶: d-SNE is a novel technique method based on the distance metric that has achieved great transfer results on the image benchmark datasets.
- *DAS*⁺⁴⁵: This model employs two regularizations, entropy minimization and self-ensemble bootstrapping to refine its classifier while minimizing the domain divergence.
- *DAAN*⁺⁵⁷: DAAN is an adaptation network with dynamic adversarial.
- *SAFN*⁺⁵⁸: SAFN is the state-of-the-arts across many visual domain adaptation benchmarks.
- *ML*⁴²: We apply the standard mutual learning in our task directly. ML can make the source domain and target domain collaborative and teach each other throughout the training process.

The original DANN, DAS, DAAN and SAFN are unsupervised domain adaptation models. As suggested in⁵⁹, we use the source code of DANN, DAS, DAAN and SAFN and extend them to DANN⁺, DAS⁺, DAAN⁺ and SAFN⁺, which utilize target supervised information. They all have improved performances, respectively.

Implementation details. The word embeddings dimension is initialized with 200-dimensions. The GRU hidden layer size is 64 dimensions. The batch size is 64. The GCN hidden layers size is 128 dimensions, and the initial learning rate is 0.001. η_T is not constant and η_D and λ are set in to be 0.1 and 0.9, respectively. Details on the hyperparameters are discussed in later sections. For a fair comparison of experimental results, feature extractors for all transfer learning models are set to CASA-T-D. Adam optimizer⁶⁰ is used to train up to 30 epochs on the dataset, and the loss value is output every 100 batches. The training is stopped when the verification loss does not decrease for ten consecutive times.

Main result analysis. *Comparison with non-transfer.* We compared the model in the context modeling stage with the current nontransfer model to explore the heterogeneous graph's presentation ability in CASA. The results are shown in Table 2.

We note that (1) the CASA-T-D results are far better than those of the other models. Despite the same tree structure, the TreeGCN results are 7.4% higher than that of TreeLSTM, which shows that GCN captures depth features more effectively. (2) CASA-T-D and TreeGCN are both GCN convolutions but different in representation learning. In comparison, the result of CASA-T-D is 2.02% higher than that of TreeGCN. (3) The BiLSTM + Att has a higher performance than TextRNN. One possible reason is that the attention mechanisms in text classification plays an important role. Thus, it could be more convincing that the CASA-T-D model has superior performance mainly due to the CAHG, which gathers rich semantic information.

Comparison with transfer. It can be observed from the experimental results in Table 3 that (1) for the popular technology "fine-tuning", after adding Virus, Usual and Weibo source domain data, the target data accuracy rates dropped by 0.68%, 1.4%, and 2.94%, respectively. This resembles our predicted results because the feature distribution gaps in the source and target domains are too large. Fixed source domain parameters cannot be corrected

Systems		Virus→Target		Usual→Target		Weibo→Target		Hotel→Target	
		F1	Acc	F1	Acc	F1	Acc	F1	Acc
Transfer	Fine-tuning	0.8192	0.8191	0.8093	0.8119	0.7966	0.7965	0.8272	0.8284
	DANN ⁺	0.8340	0.8345	0.8306	0.8314	0.8276	0.8284	0.8371	0.8386
	CCSA	0.8224	0.8232	0.8358	0.8376	0.8293	0.8304	0.8265	0.8273
	d-SNE	0.8189	0.8191	0.8349	0.8356	0.8300	0.8301	0.8202	0.8201
	DAS ⁺	0.8467	0.8469	0.8397	0.8396	0.8336	0.8335	0.7942	0.7955
	DAAN ⁺	0.8217	0.8099	0.8210	0.8109	0.8180	0.8078	0.8171	0.8068
	SAFN ⁺	0.8278	0.8284	0.8461	0.8469	0.8384	0.8397	0.8404	0.8407
	ML	0.8469	0.8479	0.8351	0.8366	0.8315	0.8325	0.8370	0.8376
	CASA	0.8541 [†]	0.8540 [†]	0.8433 [†]	0.8448 [†]	0.8398 [‡]	0.8407 [‡]	0.8475 [‡]	0.8489 [‡]

Table 3. Model comparison results with domain adaptation. The state-of-the-art result of each dataset is bolded. The marker † refers to p-value < 0.05 by comparing with ML in paired t-test, while the marker ‡ refers to p-value < 0.05 by comparing with SAFN⁺ in paired t-test. All models run over five times with random initializations and report the mean results.

Systems	Accuracy
Best	CASA 0.8541
w/o CAHG	CASA* 0.8397
w/o Class adaptation	CASA-D 0.8376
w/o Domain adaptation	CASA-T 0.8335
w/o MGD	CASA-T-D 0.8259
w/o MGD&CAHG	CASA-T-D* 0.8181

Table 4. Ablation study results. For all transfer learning models, the source domain is virus. w/o CAHG means removing the sentence nodes and intergraph token-token links and turning them into a homogeneous graph structure for training. Accuracy is reported as the average result over 5 runs with random initialization.

sufficiently during fine-tuning; finally, negative migration occurs. (2) We try to apply advanced models in the visual DA to this task, but most of the results are not very outstanding. CCSA, d-SNE and DAAN⁺ even showed negative transfer on individual datasets. For example, the accuracy of CCSA and d-SNE on Virus Target dropped by 0.27 and 0.68, respectively. This result may be caused by the diversity between the image and the natural language processing field. (3) The target domain experimental performances are improved with domain adaptation models CASA, DANN⁺, DAS⁺, SAFN⁺ and ML for all source domains. This proves that the knowledge learned from explicit sentiment is helpful to the recognition of implicit sentiment.

(4) SAFN⁺ has the outperformance result in the transfer of Usual → Target, which shows that many commonalities between visual domain adaptation and NLP domain adaptation could be mined. (5) In addition, the contribution of different source domains to implicit sentiment recognition is different. From Table 1, we know that the virus and hotel datasets are characterized by a small amount of data but a single topic of content. In contrast, the datasets Usual and Weibo have a large amount of data but on various topics. For CASA, the effect of single-topic transfer (Virus→Target, Hotel→Target) is better than that of multitopic transfer (Usual→Target, Weibo→Target). The Virus→Target, with a more similar structure, has a better transfer effect than Hotel→Target.

Ablation study. To further compare each component CASA's contribution, we sorted out the data of the ablation study part of the main experiment and plotted it into a line chart, as shown in Table 4. From Table 4, we can intuitively find the following information.

(1) Removing the CAHG, class discriminator, and domain discriminator of CASA separately, the experimental results drop by 1.44%, 1.65%, and 2.06% and are still higher than those of the nontransfer model CASA-T-D (82.59%). This shows that the fine-grained adjustment contributes to this transfer task, and the coarse-grained adjustment has greater performance than the heterogeneous structure. (2) When CASA-T-D removes CAHG and CASA* removes MGD, the accuracy drops by 0.78% and 2.16%, respectively, indicating that the CAHG and MGD we proposed can have a great impact on the experimental results.

Hyperparameter study. In this section, we will present how to choose the value of hyperparameters η_T , η_D and λ .

Hyperparameter η_T . Inspired by⁵, η_T is not a constant, but changes from 0 to 1, namely $\eta_T = \frac{2}{1 + \exp(-\alpha \cdot p)} - 1$. Wherein, the hyperparameter $-\alpha$ in this paper is set to 10 as in⁵; The relative value p of the iterative process, that is, the current number of training steps / the total number of training steps, changes from 0 to 1 with the progress of training. The above formula means that at the beginning, $\eta_T = 0$, the domain classification loss will

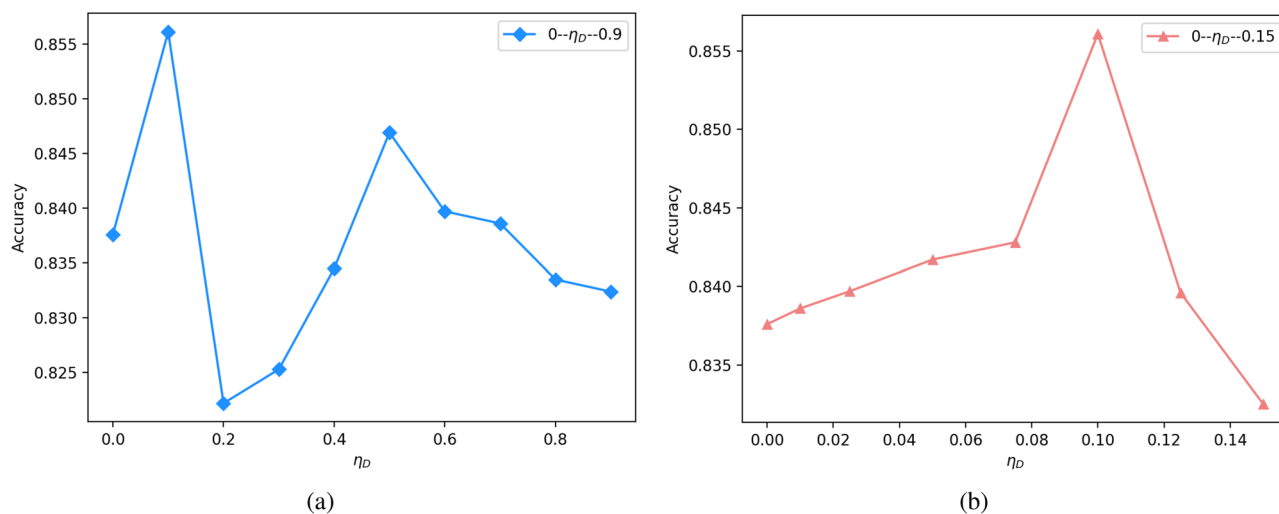


Figure 5. Hyperparameter study. For all η_D values, the source domain is virus. Accuracy is reported as the average result over 3 runs with random initialization.

not be passed back to the feature extractor network, and only the domain classifier is trained; As the training progresses, η_T gradually increases, and the feature extractor is trained and begins to gradually generate features that can confuse the domain classifier.

Hyperparameters η_D and λ . Which are selected through the validation set. First we have to judge whether η_D and λ is the same order of magnitude, and the order of magnitude after verification is -1 . Therefore, set $\eta_D + \lambda = 1$, When the verification set loss is minimum, checking the test accuracy.

Figure 5 shows the test set accuracy under different η_D values. From (a), it shows that the optimal range of η_D value is 0 to 0.2. Thus we did further experiments, as shown in (b). From (b), we can see that the optimal η_D value is 0.10, so λ is 0.90.

Effectiveness verification

Case study. We want to explore what domain-specific information, which has been learned from the source domain, would enhance implicit emotion classification. In Table 5, we visualized the hierarchical attention layer in the nontransfer model CASA-T-D and the full model CASA. Using the CASA-T-D model as a benchmark, we compare the differences in attention weight distribution of different models under the same text.

The first two samples contain only one context and the target sentence. Although the two models' attention scores are different, the critical points of emotional judgment are noted, such as '差 (poor)' and '伤害 (harm)'. Obviously, these tokens are domain-independent words, which are usually used in many domains. In the third example, the CASA model learned the virus domain-specific information '抵抗力 (resistless)', but CASA-T-D could not.

Then, we examine the impact of long contextual text on the two models. As shown in the last example, CASA-T-D focuses most of its attention on the token '不到 (not)' in the target sentence and does not notice the token '不到 (good)' in the context. In contrast, the CASA attention score is relatively scattered, conducive to the judgment of emotional polarity, which may benefit from source domain knowledge.

Feature visualization. To better illustrate how the CASA works, we used t-SNE⁶¹ to reduce the dimensionality of the feature to two and visualize the data distributions after domain adaptation Virus \rightarrow Target.

Figure 6 shows that the baseline models (a), (b) and (c) all virtually guarantee the source and target domain data fusion. On the other hand, the distance between the classes in the domain is still very close.

In contrast, benefitting from our proposed MGD, CASA's intraclass boundary is significantly more significant, conducive to intraclass identification.

Document	CASA-T-D	CASA	Label
context	S ₀ 这是 哪家 ？ ？ Which shop is this ? ?	S ₀ 这是 哪家 ？ ？ Which shop is this ? ?	
target	在 徐家汇 五番 街 的 拉面 玩家 ， 服务态度 真的 差 ✓ The Ramen Player on Wufan Street in Xujiahui, which has a poor service attitude.	在 徐家汇 五番 街 的 拉面 玩家 ， 服务态度 真的 差 ✓ The Ramen Player on Wufan Street in Xujiahui, which has a poor service attitude.	Negative
context	S ₀ 你 伤害 了 人家 人家 发了 四个 啊 ！ You hurt him, and he sent a message with four "ah" !	S ₀ 你 伤害 了 人家 人家 发了 四个 啊 ！ You hurt him, and he sent a message with four "ah" !	
target	不是 这个 ， 这个 像 路 人 ， 是 叫 他 去 看 必淑录 节目 那个 ✓ It is not this person. This is a passer-by, the one who asked him to watch the show of Bi Shulu.	不是 这个 ， 这个 像 路 人 ， 是 叫 他 去 看 必淑录 节目 那个 ✓ It is not this person. This is a passer-by, the one who asked him to watch the show of Bi Shulu.	Negative
context	S ₀ 我 看到 的 时候 已经 两点 多 了 你 好 不容易 早 睡 不 愿意 吵 你 It's already past two o'clock when I saw it. You finally went to bed early, and I did not want to bother you.	S ₀ 我 看到 的 时候 已经 两点 多 了 你 好 不容易 早 睡 不 愿意 吵 你 It's already past two o'clock when I saw it. You finally went to bed early, and I did not want to bother you.	
target	你 貌似 把 我 传染 了 我 对 你 就 如此 没有 抵抗力 嘛 ✗ I seem to be infected by you. Am I so resistless to you?	你 貌似 把 我 传染 了 我 对 你 就 如此 没有 抵抗力 嘛 ✓ I seem to be infected by me. Am I so resistless to you?	Negative
context ₀	S ₀ 在 拉市海 ， 骑马 是 不错 的 选择 。 Horse riding is a good choice in Lashihai.	S ₀ 在 拉市海 ， 骑马 是 不错 的 选择 。 Horse riding is a good choice in Lashihai.	
context ₁	S ₁ 每个 师傅 带 着 两匹马 - Each qualified worker will bring two horses -	S ₁ 每个 师傅 带 着 两匹马 。 Each qualified worker will bring two horses .	
context ₂	S ₂ 骑马 一 定 要 起 高高大大 的 马 。 You better ride a tall horse.	S ₂ 骑马 一 定 要 起 高高大大 的 马 。 You better ride a tall horse.	
target	穿行 在 山 中 的 林道 中间 ， 那 种 感觉 你 不 骑马 是 体会 不到 的 。 ✗ Walking in the middle of the forest road in the mountains, you will not experience that feeling without riding a horse.	穿行 在 山 中 的 林道 中间 ， 那 种 感觉 你 不 骑马 是 体会 不到 的 。 ✓ Walking in the middle of the forest road in the mountains, you will not experience that feeling without riding a horse.	Positive

Table 5. Case study. Visualization of attention weight distribution from CASA-T-D and CASA on testing examples. For CASA, the source domain is Virus. Marker ✓ signifies correct prediction, while marker ✗ signifies incorrect prediction. Blue denotes the sentence weight, and red denotes the word weight.

Conclusions and future work

This paper proposes a CASA network via graph convolution for the cross-domain implicit sentiment classification problem, first building a relation between explicit and implicit sentiment. Existing studies either rarely consider using domain-specific semantic information or ignore maximizing class-specific decision boundaries. We aim to address the above two drawbacks. First, CASA provides a CAHG, effectively extracting domain-specific semantical information for both sources and targets. Hence, CASA improved the model's generalization ability for implicit classification tasks. The case study shows that our model can effectively capture high-level domain-specific features. Second, CASA conducts an MGD to adapt the domain distribution, enhancing class distinction in each sentiment polarity decision boundary during domain adaptation. The feature visualization results show that CASA clarifies samples' boundaries from different classes while adapting to the domain.

Moreover, there are several worthy challenges in cross-domain implicit sentiment tasks, such as transferring between the different single-domain topics, fine-grain sentiment transferring, ambivalence handling, and transferring explicit-to-implicit topics where the target domain tags are not given. We believe that all these factors can help us comprehend the link between explicit and implicit sentiment and that implicit sentiment analysis will be solved more effectively in the future.

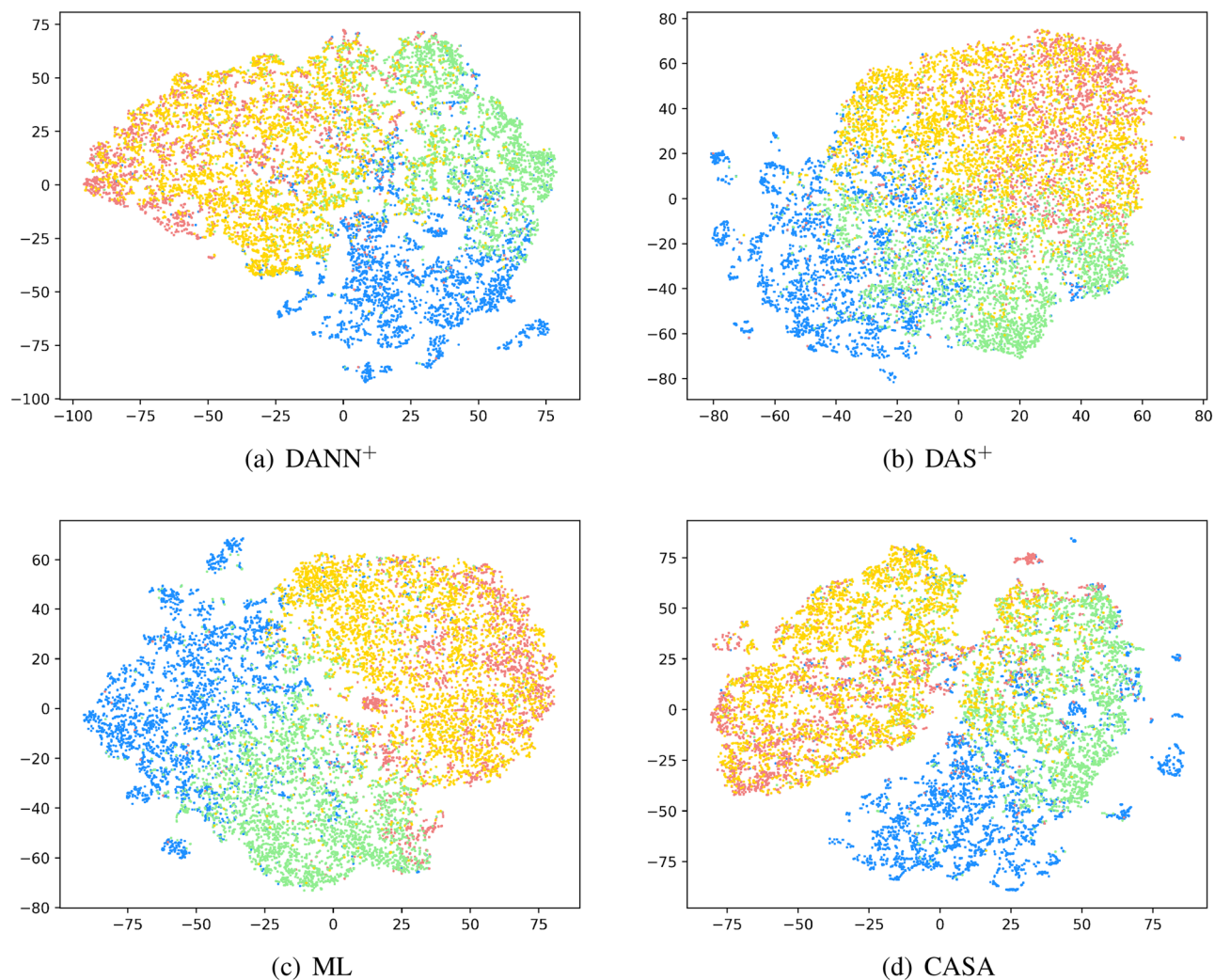


Figure 6. t-SNE visualization of the distribution of the features after different domain adaptation methods. Different colors indicate different sentiment polarities. (source domain: positive = dodger blue, negative = light coral; target domain: positive = light green, negative = gold).

Received: 30 May 2021; Accepted: 20 October 2021
Published online: 11 November 2021

References

- Zhang, L. & Liu, B. Sentiment analysis and opinion mining. *Encycl. Mach. Learn. Data Min.* 1152–1161. https://doi.org/10.1007/978-1-4899-7687-1_907 (2017).
- Liu, B. Sentiment analysis and opinion mining. *Synthesis Lect. Hum. Lang. Technol.* 5, 1–184. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016> (2012).
- Zuo, E., Zhao, H., Chen, B. & Chen, Q. Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. *IEEE Access* 8, 37967–37975. <https://doi.org/10.1109/ACCESS.2020.2975244> (2020).
- Zellinger, W., Lughofer, E., Saminger-Platz, S., Grubinger, T. & Natschläger, T. Central moment discrepancy (CMD) for domain-invariant representation learning. *arXiv:arXiv:1702.08811* (2017).
- Ganin, Y. *et al.* Domain-adversarial training of neural networks. *Adv. Comput. Vis. Pattern Recognit.* 189–209. https://doi.org/10.1007/978-3-319-58347-1_10 (2017). *arXiv:1505.07818*.
- Li, Z., Zhang, Y., Wei, Y., Wu, Y. & Yang, Q. End-to-end adversarial memory network for cross-domain sentiment classification. *IJCAI Int. Joint Conf. Artif. Intell.* 2237–2243. <https://doi.org/10.24963/ijcai.2017/311> (2017).
- Hu, M. *et al.* Domain-invariant feature distillation for cross-domain sentiment classification. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference* 5559–5568. <https://doi.org/10.18653/v1/d19-1558> (2020). *arXiv:1908.09122*.
- Peng, M., Zhang, Q., Jiang, Y. G. & Huang, X. Cross-domain sentiment classification with target domain specific information. in *ACL 2018—56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*. Vol. 1. 2505–2513. <https://doi.org/10.18653/v1/p18-1233> (2018).
- Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. in *5th International Conference on Learning Representations, ICLR 2017—Conference Track Proceedings*. *arXiv:1609.02907* (2019).
- Yao, L., Mao, C. & Luo, Y. Graph convolutional networks for text classification. *arXiv* (2018).
- Liu, X., You, X., Zhang, X., Wu, J. & Lv, P. Tensor graph convolutional networks for text classification. *arXiv* (2020).
- Zhang, C., Li, Q. & Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv* (2019).

13. Bai, X., Liu, P. & Zhang, Y. Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network. *IEEE/ACM Trans. Audio Speech Lang. Process.* **29**, 503–514. <https://doi.org/10.1109/TASLP.2020.3042009> (2021). [arXiv:2002.09685](https://arxiv.org/abs/2002.09685).
14. Cambria, E. Affective computing and sentiment analysis. *IEEE Intell. Syst.* **31**, 102–107. <https://doi.org/10.1109/MIS.2016.31> (2016).
15. Chen, Z. *et al.* Joint model for subsentence-level sentiment analysis with Markov logic. *J. Assoc. Inf. Sci. Technol.* **66**, 1913–1922. <https://doi.org/10.1002/asi.23301> (2015).
16. Liu, S. M. & Chen, J. H. A multi-label classification based approach for sentiment classification. *Expert Syst. Appl.* **42**, 1083–1093. <https://doi.org/10.1016/j.eswa.2014.08.036> (2015).
17. Peng, H., Ma, Y., Poria, S., Li, Y. & Cambria, E. Phonetic-enriched text representation for Chinese sentiment analysis with reinforcement learning. *Inf. Fusion* **70**, 88–99. <https://doi.org/10.1016/j.inffus.2021.01.005> (2021). [arXiv:1901.07880](https://arxiv.org/abs/1901.07880).
18. Sun, Y., Lin, L., Yang, N., Ji, Z. & Wang, X. Radical-enhanced Chinese character embedding. *Lect. Notes Comput. Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **8835**, 279–286. https://doi.org/10.1007/978-3-319-12640-1_34 (2014). [arXiv:1404.4714](https://arxiv.org/abs/1404.4714).
19. Chen, X., Xu, L., Liu, Z., Sun, M. & Luan, H. Joint learning of character and word embeddings. *in IJCAI International Joint Conference on Artificial Intelligence 2015—January*. 1236–1242 (2015).
20. Yu, J., Jian, X., Xin, H. & Song, Y. Joint embeddings of Chinese words, characters, and fine-grained subcharacter components. in *EMNLP 2017—Conference on Empirical Methods in Natural Language Processing, Proceedings*. 286–291. <https://doi.org/10.18653/v1/d17-1027> (2017).
21. Peng, H., Cambria, E. & Zou, X. Radical-based hierarchical embeddings for Chinese sentiment analysis at sentence level. in *FLAIRS 2017—Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*. 347–352 (2017).
22. Tao, H. *et al.* A radical-aware attention-based model for Chinese text classification. in *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*. 5125–5132. <https://doi.org/10.1609/aaai.v33i01.33015125> (2019).
23. Sorini, C. & Falcone, M. Jumping NLP curves: A review of natural language processing research. *IEEE Comput. Intell. Mag.* **9**, 48–57 (2014).
24. Yinxia, L. O., Zhang, Y., Fei, L. I., Qian, T. & Donghong, J. I. Emoji-based sentiment analysis using attention networks. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **19**. <https://doi.org/10.1145/3389035> (2020).
25. Usama, M. *et al.* Attention-based sentiment analysis using convolutional and recurrent neural network. *Future Gener. Comput. Syst.* **113**, 571–578. <https://doi.org/10.1016/j.future.2020.07.022> (2020).
26. Xi, C., Lu, G. & Yan, J. Multimodal sentiment analysis based on multi-head attention mechanism. *ACM Int. Conf. Proc. Ser.* 34–39. <https://doi.org/10.1145/3380688.3380693> (2020).
27. Yadav, R. K., Jiao, L., Goodwin, M. & Granmo, O. C. Positionless aspect based sentiment analysis using attention mechanism [formula presented]. *Knowl.-Based Syst.* **226**. <https://doi.org/10.1016/j.knosys.2021.107136> (2021).
28. Wang, Z., Ho, S. B. & Cambria, E. Multi-level fine-scaled sentiment sensing with ambivalence handling. *Int. J. Uncertain. Fuzz. Knowl.-Based Syst.* **28**, 683–697. <https://doi.org/10.1142/S0218488520500294> (2020).
29. Hochreiter, S. & Unger Schmidhuber, J. Long shortterm memory. *Neural Comput.* **9**, 17351780 (1997).
30. Dong, L. *et al.* Adaptive recursive neural network for target-dependent Twitter sentiment classification. in *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014—Proceedings of the Conference*. Vol. 2. 49–54. <https://doi.org/10.3115/v1/p14-2009> (2014).
31. Kim, Y. Convolutional neural networks for sentence classification. in *EMNLP 2014—2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. 1746–1751. <https://doi.org/10.3115/v1/d14-1181>. [arXiv:1408.5882](https://arxiv.org/abs/1408.5882) (2014).
32. Yang, Z. *et al.* Hierarchical attention networks for document classification. in *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016—Proceedings of the Conference* 1480–1489. <https://doi.org/10.18653/v1/n16-1174> (2016).
33. Ma, D., Li, S., Zhang, X. & Wang, H. Interactive attention networks for aspect-level sentiment classification. in *IJCAI International Joint Conference on Artificial Intelligence*. 4068–4074. <https://doi.org/10.24963/ijcai.2017/568> (2017).
34. Wei, J., Liao, J., Yang, Z., Wang, S. & Zhao, Q. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neurocomputing* **383**, 165–173. <https://doi.org/10.1016/j.neucom.2019.11.054> (2020).
35. Zhang, X., Zhao, J. & Lecun, Y. Character-level convolutional networks for text classification. *Adv. Neural Inf. Process. Syst.* **2015—January**, 649–657 (2015). [arXiv:1509.01626](https://arxiv.org/abs/1509.01626).
36. Conneau, A., Schwenk, H., Cun, Y. L. & Barrault, L. Very deep convolutional networks for text classification. in *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017—Proceedings of Conference*. Vol. 1. 1107–1116. <https://doi.org/10.18653/v1/e17-1104> (2017).
37. Paulus, R., Socher, R. & Manning, C. D. Global belief recursive neural networks. *Adv. Neural Inf. Process. Syst.* **4**, 2888–2896 (2014).
38. Teng, Z. & Zhang, Y. Head-lexicalized bidirectional tree LSTMs. *Trans. Assoc. Comput. Linguist.* **5**, 163–177. https://doi.org/10.1162/tacl_a_00053 (2017).
39. Zhang, Y., Qi, P. & Manning, C. D. Graph convolution over pruned dependency trees improves relation extraction. in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*. 2205–2215. <https://doi.org/10.18653/v1/d18-1244>. [arXiv:1809.10185](https://arxiv.org/abs/1809.10185) (2020).
40. Mou, L., Li, G., Zhang, L., Wang, T. & Jin, Z. Convolutional neural networks over tree structures for programming language processing. in *30th AAAI Conference on Artificial Intelligence, AAAI 2016*. 1287–1293. [arXiv:1409.5718](https://arxiv.org/abs/1409.5718) (2016).
41. Liao, J., Wang, S. & Li, D. Identification of fact-implied implicit sentiment based on multi-level semantic fused representation. *Knowl.-Based Syst.* **165**, 197–207. <https://doi.org/10.1016/j.knosys.2018.11.023> (2019).
42. Zhang, Y., Xiang, T., Hospedales, T. M. & Lu, H. Deep mutual learning. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 4320–4328. <https://doi.org/10.1109/CVPR.2018.00454>. [arXiv:1706.00384](https://arxiv.org/abs/1706.00384) (2018).
43. Zhang, Y. *et al.* Every document owns its structure: Inductive text classification via graph neural networks. *arXiv*. <https://doi.org/10.18653/v1/2020.acl-main.31>. [arXiv:2004.13826](https://arxiv.org/abs/2004.13826) (2020).
44. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks?. *Adv. Neural Inf. Process. Syst.* **4**, 3320–3328 (2014). [arXiv:1411.1792](https://arxiv.org/abs/1411.1792).
45. He, R., Lee, W. S., Ng, H. T. & Dahlmeier, D. Adaptive semi-supervised learning for cross-domain sentiment classification. in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*. 3467–3476. <https://doi.org/10.18653/v1/d18-1383>. [arXiv:1809.00530](https://arxiv.org/abs/1809.00530) (2020).
46. Qu, X., Zou, Z., Cheng, Y., Yang, Y. & Zhou, P. Adversarial category alignment network for cross-domain sentiment classification. in *NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference*. Vol. 1. 2496–2508. <https://doi.org/10.18653/v1/n19-1258> (2019).
47. Bahdanau, D., Cho, K. H. & Bengio, Y. Neural machine translation by jointly learning to align and translate. in *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*. 1–15. [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) (2015).
48. Ben-David, S. *et al.* A theory of learning from different domains. *Mach. Learn.* **79**, 151–175. <https://doi.org/10.1007/s10994-009-5152-4> (2010).

49. Ben-David, S., Blitzer, J., Crammer, K. & Pereira, F. Analysis of representations for domain adaptation. *Adv. Neural Inf. Process. Syst.* 137–144. <https://doi.org/10.7551/mitpress/7503.003.0022> (2007).
50. Qian-Sun. A two-stage weighting framework for multi-source domain adaptation. *Adv. Neural Inf. Process. Syst.* 505–513 (2011).
51. Chen, H. Y. & Chen, H. H. Implicit polarity and implicit aspect recognition in opinion mining. in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016—Short Papers*. 20–25. <https://doi.org/10.18653/v1/p16-2004> (2016).
52. Liu, P., Qiu, X. & Xuanjing, H. Recurrent neural network for text classification with multi-task learning. in *IJCAI International Joint Conference on Artificial Intelligence 2016—January*. 2873–2879. [arXiv:1605.05101](https://arxiv.org/abs/1605.05101) (2016).
53. Zhou, P. *et al.* Attention-based bidirectional long short-term memory networks for relation classification. in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016—Short Papers*. 207–212. <https://doi.org/10.18653/v1/p16-2034> (2016).
54. Tai, K. S., Socher, R. & Manning, C. D. Improved semantic representations from tree-structured long short-term memory networks. in *ACL-IJCNLP 2015—53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*. Vol. 1. 1556–1566 (2015).
55. Motiian, S., Piccirilli, M., Adjeroh, D. A. & Doretto, G. Unified deep supervised domain adaptation and generalization. *arXiv* (2017).
56. Xu, X., Zhou, X., Venkatesan, R., Swaminathan, G. & Majumder, O. D-SNE: Domain adaptation using stochastic neighborhood embedding. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2019–June**, 2492–2501. <https://doi.org/10.1109/CVPR.2019.00260> (2019). [arXiv:1905.12775](https://arxiv.org/abs/1905.12775).
57. Yu, C., Wang, J., Chen, Y. & Huang, M. Transfer learning with dynamic adversarial adaptation network. in *Proceedings—IEEE International Conference on Data Mining, ICDM 2019—November*. 778–786. <https://doi.org/10.1109/ICDM.2019.00088>. [arXiv:1909.08184](https://arxiv.org/abs/1909.08184) (2019).
58. Xu, R., Li, G., Yang, J. & Lin, L. Larger norm more transferable: an adaptive feature norm approach for unsupervised domain adaptation. in *Proceedings of the IEEE International Conference on Computer Vision 2019—October*. 1426–1435. <https://doi.org/10.1109/ICCV.2019.00151>. [arXiv:1811.07456](https://arxiv.org/abs/1811.07456) (2019).
59. Li, Z., Wei, Y., Zhang, Y., Zhang, X. & Li, X. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. in *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 4253–4260. <https://doi.org/10.1609/aaai.v33i01.33014253>. [arXiv:1811.10999](https://arxiv.org/abs/1811.10999) (2019).
60. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. in *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2015).
61. Van Der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2625 (2008).

Acknowledgements

The work is supported in part by the National Key Research and Development Program of China (2019YFC1606100 and sub-program 2019YFC1606104), the Major science and technology projects of Xinjiang Uygur Autonomous Region (2020A03001 and sub-program 2020A03001-3), the National Natural Science Foundation of China (61862061, 61563052 and 61363064), in part by the Scientific Research Initiate Program of Doctors of Xinjiang University (BS180268), Shaanxi Provincial Natural Science Foundation (NO.2020GY-093), and Shangluo City Science and Technology Program Fund Project (NO.SK2019-83).

Author contributions

E.Z., A.A., M.M., Y.Z. and K.U. conceived and designed the study. E.Z., A.A. and M.M. collected and preprocessed the data. All authors contributed to the manuscript. Y.Z. and K.U. reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.U.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021