Knowledge-Based Systems 227 (2021) 107160

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A constrained optimization approach for cross-domain emotion distribution learning

Xiaorui Qin [a], Yufu Chen [a], Yanghui Rao [a], Haoran Xie [b,*], Man Leung Wong [b], Fu Lee Wang [c]

[a] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[b] Department of Computing and Decision Sciences, Lingnan University, New Territories, Hong Kong Special Administrative Region
[c] School of Science and Technology, The Open University of Hong Kong, Kowloon, Hong Kong Special Administrative Region

ARTICLE INFO

Article history:
Received 7 November 2020
Received in revised form 27 February 2021
Accepted 17 May 2021
Available online 27 May 2021

Keywords:
Emotion distribution learning
Non-negative Matrix Tri-Factorization
Domain adaptation
Content-based constraint

ABSTRACT

Emotion distribution learning aims to annotate unlabeled instances with a set of emotion categories and their strengths. Non-negative Matrix Tri-Factorization (NMTF) introduces an association matrix between document clusters and word clusters to help the domain adaptation task in emotion distribution learning. Nevertheless, many prior cross-domain emotion distribution learning methods had two major deficiencies. First, they hypothesize that there is a one-to-one correspondence between document clusters and emotion labels. In their experiments, the number of document clusters depends on the number of labels. Second, the prior work does not endow models with adequate constraints. In the real scenario of cross-domain emotion distribution learning, there are potential constraints that may improve the performance of such models. In order to address these problems, we propose a constrained optimization approach based on NMTF for cross-domain emotion distribution learning. In our model, the relationship between document clusters and emotion labels is not always one-to-one. A novel content-based constraint is also endowed based on the hypothesis that documents belonging to the same clusters must have similar content. We solve the optimization problem by an alternately iterative algorithm and show the proof of convergence. Experiments on 12 real-world cross-domain emotion distribution learning tasks validate the effectiveness of our method.

1. Introduction

Emotion distribution learning [1,2], which aims to predict the intensity values of an instance across a set of emotion categories, is valuable to mine the fine-grained emotion information. Many methods of emotion distribution learning have been developed for a single domain [1,3], but the model trained on a source domain may not be transferred to a target domain successfully [4]. When the distribution of features changes across different domains, models need to be retrained using newly collected training samples. Unfortunately, it is expensive or prohibitive to achieve training data for diverse application areas. Thus, cross-domain emotion distribution learning [5–7] has attracted increasing attention in the research community.

Non-negative Matrix Tri-Factorization (NMTF), which has been widely used for document clustering [8–11], is valuable to capture the hidden associations of unlabeled texts from diverse domains [5]. In recent years, it plays an important role in cross-domain emotion distribution learning. Some existing NMTF based models use high-level concepts (e.g. word clusters) to help alleviate the data distribution difference [12,13]. The basic intuition of these models is that high-level semantic features generated from clusters can be applied to disambiguate different emotion contexts, and thus to boost the performance of learning emotion distributions. Since the cluster-level representation is relatively unambiguous, it is easier for a model to learn and predict the corresponding emotion labels. However, these models have two major deficiencies in the real scenario.

First, they hypothesize that there is a one-to-one correspondence between document clusters and emotion labels [12,13]. In their experiments, the number of document clusters depends on the number of labels. However, this hypothesis is not perfectly reasonable. One cluster of documents may trigger several different emotions while one emotion may be triggered by different document clusters. For example, the document cluster "the fight against Novel Coronavirus" associates with different emotions, e.g., "sad", "moved", and "sympathy". The emotion "sad" can be evoked by many document clusters, e.g., "the fight against Novel Coronavirus" and "the death of a famous person". Thus we assume that the relationship between document clusters and

* Corresponding author.
E-mail addresses: qinxr5@mail2.sysu.edu.cn (X. Qin), chenyf66@mail2.sysu.edu.cn (Y. Chen), raoyangh@mail.sysu.edu.cn (Y. Rao), hrxie2@gmail.com (H. Xie), mlwong@ln.edu.hk (M.L. Wong), pwang@ouhk.edu.hk (F.L. Wang).

https://doi.org/10.1016/j.knosys.2021.107160
0950-7051/

emotion labels is not always one-to-one and the number of document clusters may not equal to that of emotional labels.

Second, the prior work does not endow models with adequate constraints. The goal of these models is to simply minimize the distance between the original matrix and the product of decomposed matrices. For matrix factorization based methods, additional constraints provide valuable information based on human intuition and knowledge which may contribute to various tasks [14–17]. In the real scenario of cross-domain emotion distribution learning, there are potential constraints that may improve such models. For example, based on our intuition, documents within the same cluster would share more similar content than those distributed in different clusters. Accordingly, we can introduce an additional constraint to help the model capture more reasonable relationships between documents.

In order to improve the approach for cross-domain emotion distribution learning based on NMTF in above-mentioned aspects, the main contributions of this study can be summarized as follows:

- We propose that the relationship between document clusters and emotion labels is not always one-to-one and the number of document clusters may not equal to that of labels. On the one hand, one document cluster may evoke different emotions simultaneously, e.g., "joy" and "surprise", "moved" and "sympathy", etc. On the other hand, an emotion can be triggered by different document clusters, e.g., "joy" can be triggered by "win a scholarship" and "meet up with friends".
- We endow the NMTF based method with a novel content-based constraint. The corresponding hypothesis is that documents belonging to the same clusters must have similar content. This constraint aims to clearly depict that documents within the same cluster are more likely to talk about similar content.

Experimental results on 12 cross-domain emotion distribution learning tasks validate that the proposed model outperforms other state-of-the-art baselines. Ablation experiments validate the effectiveness of the two contributions, i.e., the many-to-many relationships between document clusters and labels, and the content-based constraint.

2. Related work

2.1. Cross-domain sentiment classification

Sentiment analysis, which refers to the inference of users' views, positions, and attitudes in their written or spoken documents [18–23], plays an increasingly important role in real-life applications [24–30]. Both lexical and learning-based approaches have been utilized for this task [31,32]. Lexical-based methods detect sentiments by exploiting a predefined list of words or phrases, where each word or phrase is associated with a specific sentiment [33–37]. However, the performance may be limited due to the employed lexicons are often domain-specific and language dependent. Learning-based methods often use labeled data to train supervised algorithms, which could adapt the trained models for specific purposes and contexts [38]. Some supervised algorithms employed and proposed for sentiment analysis include: naïve Bayes, maximum entropy, support vector machines, convolutional neural networks, and attention-based deep models [38–46].

The task of training and evaluating models on different data distributions is known as domain adaptation [47,48]. In a preliminary study, [49] proposed the structural corresponding learning (SCL) algorithm, which adapts between the source and target domains by the selected pivot features. When applied to cross-domain sentiment classification, the SCL algorithm constructed a set of tasks to model the relationship between pivot features and non-pivot features, and selected pivots according to their mutual information with the source labels [50]. Unlike the above work, [51] proposed the spectral feature alignment (SFA) algorithm to cross-domain sentiment classification. In this study, the SFA used some domain-independent words as a bridge to construct a bipartite graph for modeling the co-occurrence relationship between domain-specific words and domain-independent words. Then, the SFA employed spectral clustering to group the domain-specific words for domain adaption in sentiment classification. [52] proposed the topic correlation analysis approach which used the probabilistic latent semantic analysis (PLSA) and expectation maximization (EM) algorithm to extract topics, and then clustered the domain-specific features by constructing a topic mapping matrix. [53] proposed a domain adaptation method for sentiment classification by taking both labeling adaptation and instance adaptation into account. In the area of sentiment analysis, this idea is based on the observation that, features with different types of part-of-speech tags have a distinct change in distribution for domain adaptation. Note that the goal of cross-domain sentiment classification is to assign a polarity or a score to documents [48,54,55] while that of cross-domain emotion distribution learning is to identify the emotion category and the emotion intensity in cross-domain tasks [6].

2.2. Cross-domain emotion distribution learning

Emotion Distribution Learning [1], which aims to predict the intensity values of a sentence across a set of emotion categories, has attracted increasing attention since the "SemEval" conference in 2007 [56]. With respect to domain adaptation for emotion distribution learning, [6] proposed to modify the regularization term in the logistic regression model for weighting the documents in the source domain. Although such a kind of document-level strategy can prevent the overfitting issue on the source domain, it may not guarantee a good generalization performance on the target domain.

NMTF [57,58], which is a popular tool for capturing the hidden associations of text information from diverse domains, performs quite well in cross-domain text categorization [5]. By introducing a shared matrix that captures the association between word clusters and document classes, [12] and [13] used high-level concepts (e.g. word clusters) to help alleviate the data distribution difference. They respectively proposed the Triplex Transfer Learning (TriTL) model and Transitive Transfer Learning (TTL) model based on NMTF. The two methods assume that all the data domains have the same set of shared concepts, which are used as the bridge for knowledge transferring. These models all decompose the word–document matrix into the matrix for word clusters, the matrix for the association between word clusters and document clusters, and the matrix for document clusters [5]. However, such methods show two shortcomings when applied to cross-domain emotion distribution learning. First, they unreasonably hypothesize that the number of document clusters must equal to that of emotional labels [5,12,13]. Second, they do not develop adequate constraints. Our work attempts to address such research questions from these two perspectives.

3. Approach

NMTF is a prevalent unsupervised approach for document clustering, which is valuable to capture the hidden associations of unlabeled texts from diverse domains [12,13]. To transfer knowledge from a source domain to target domains, an optimization for two matrix tri-factorizations is conducted on the source and target domains jointly, in which the shared connection between word and document clusters acts as a bridge across different domains. Our model decomposes the document–word matrix into three non-negative matrices, i.e., the document-cluster matrix, the association matrix between document and word clusters, and the word-cluster matrix. We name our model as cNMTF by introducing a content-based constraint and an alternately iterative algorithm for training. Then, the mapping matrix between document clusters and emotion labels is used to predict the emotion distribution for the documents in the target domain.

3.1. Problem formalization

Fig. 1 illustrates the architecture of our cNMTF model, which decomposes the document–word matrix into the document-cluster matrix, the association matrix between document and word clusters, and the word-cluster matrix, in the source and target domains, respectively. A content-based constraint and the hypothesis of the many-to-many correspondence between document clusters and labels are added to help our model learn these matrices better. Given n_s and n_t documents in the source and target domains, we use d_s and d_t to denote the numbers of words in the source and target domains, respectively. For the combination of all documents, let c_n be the number of document clusters. Similarly, c_d represents the number of word clusters for the combination of all features. For the source and target domains, we use X_s and X_t to denote the document–word matrices, F_s and F_t to denote the document-cluster matrices, and G_s and G_t to denote the word-cluster matrices. The sizes of X_s, X_t, F_s, F_t, G_s, and G_t are $n_s\times d_s$, $n_t\times d_t$, $n_s\times c_n$, $n_t\times c_n$, $d_s\times c_d$, and $d_t\times c_d$, respectively. To connect document and word clusters in the source and target domains, a shared matrix S with the size of $c_n\times c_d$ is used to transfer knowledge. Assume that the number of emotions is l, we can get a document-emotion matrix E_s for the source domain, and the size of E_s is $n_s\times l$. The basic objectives of our NMTF-based model are as follows:

$$\begin{aligned} X_s &= F_s S G_s^T, \\ X_t &= F_t S G_s^T. \end{aligned} \tag{1}$$

3.2. Content-based constraint

We borrow the main idea of clustering methods to propose our constraint. Many popular clustering methods, e.g., K-means [59], Spectral Clustering [60,61], and Non-negative Matrix Factorization (NMF) [62], cluster the data based on the similarities along the feature. In text clustering tasks, the feature refers to the content in many circumstances. Clearly, the content of different documents plays the most important role in document clustering. Take the document cluster of "the fight against Novel Coronavirus" as an example, it may include different documents that contain similar words, e.g., "Covid-19", "masks", and "pneumonia". The more similar the content of documents is, the more likely they belong to the same cluster.

Thus, we hypothesize that documents belonging to the same clusters must have similar content. We incorporate this constraint into our objective function through a binary indicator matrix H, s.t., $H_{ij} = 1$ for $sim(X_{s(i)}, X_{s(j)}) \geq \lambda$, and $H_{ij} = 0$ for the other case. In the above, λ is the similarity threshold that can be tuned by cross-validation and $sim(X_{s(i)}, X_{s(j)})$ represents the cosine similarity between vectors of two documents, i.e., $X_{s(i)}$ and $X_{s(j)}$. The content-based constraint is captured through the constraint $\left\|H\circ(F_sF_s^T - X_sX_s^T)\right\|_F^2$, where $\|M\|_F$ is the Frobenius norm of matrix M and $\circ$ is the component wise product or the Hadamard product between two matrices. Accordingly, the following optimization formulation is proposed for the clustering of cross-domain documents:

$$\begin{aligned} \min_{S,F_s,G_s,F_t,G_t} & \left\|X_s - F_sSG_s^T\right\|_F^2 + \alpha\left\|X_t - F_tSG_t^T\right\|_F^2 \\ & + \beta\left\|H\circ(F_sF_s^T - X_sX_s^T)\right\|_F^2, \end{aligned} \tag{2}$$

where $\alpha \geq 0$ and $\beta \geq 0$ are the trade-off factors for two matrix tri-factorizations and the aforementioned content-based constraint, respectively.

3.3. Convex sub-problem

To solve the optimization problem of our model, we use Φ to represent the objective function by:

$$\Phi = \left\|X_s - F_sSG_s^T\right\|_F^2 + \alpha\left\|X_t - F_tSG_t^T\right\|_F^2 + \beta\left\|H\circ(F_sF_s^T - X_sX_s^T)\right\|_F^2. \tag{3}$$

Since Φ is non-convex in F_s, we use the variable substitution technique to derive a convex sub-problem, as follows:

$$\begin{aligned} \Psi = \Phi & - \beta\left\|H\circ(F_sF_s^T - X_sX_s^T)\right\|_F^2 + \beta\left\|H\circ(F_sA^T - X_sX_s^T)\right\|_F^2 \\ & + \beta\|F_s - A\|_F^2. \end{aligned} \tag{4}$$

By introducing the auxiliary matrix A, the new objective function Ψ is convex in all its variables. Note that $\Phi = \Psi(F_s = A)$.

Lemma 3.1. *The problem $\Psi^* = \min\left\{\Psi\right\}$ provides a lower bound to the problem $\Phi^* = \min\left\{\Phi\right\}$.*

Proof. Suppose that F_s^* represents the solution to Φ^*. As a solution to the minimization problem, we have:

$$\Psi^* \leq min\left\{\Psi(F_s = F_s^*)\right\} \leq min\left\{\Psi(F_s = F_s^*, A = F_s^*)\right\} = \Phi^*. \quad \square \tag{5}$$

Solving Ψ ensures a parsimonious fit to Φ because $\Psi^* = \min\left\{\Psi\right\}$ provides a lower bound to the problem $\Phi^* = \min\left\{\Phi\right\}$.

3.4. Alternately iterative algorithm

We propose an alternately iterative algorithm to solve the new optimization problem. Take F_s as an example, we first compute the derivative of Ψ on F_s:

$$\begin{aligned} \frac{\partial\Psi}{\partial F_s} = & -2X_sG_sS^T + 2F_sSG_s^TG_sS^T + 2\beta H\circ(F_sA^T)A \\ & - 2\beta H\circ(X_sX_s^T)A + 2\beta(F_s - A). \end{aligned} \tag{6}$$

Then we can place the negative part of the derivative in the numerator while the positive part in the denominator to construct the update rule for F_s:

$$F_s \leftarrow F_s \circ \frac{X_sG_sS^T + \beta H\circ(X_sX_s^T)A + \beta A}{F_sSG_s^TG_sS^T + \beta H\circ(F_sA^T)A + \beta F_s}. \tag{7}$$

Note that non-negativity is always satisfied because all variables are updated in this multiplication form. A constant ϵ is used to prevent division by zero. ε is used as the tolerance for the optimization for two consecutive iterations. Our total inference method is showed in Algorithm 1.

Theorem 3.1. *Algorithm 1 is guaranteed to converge to a locally-optimal solution.*

First, we prove the convergence of the update rule for F_s in Eq. (7). The auxiliary function which is similar to that used in the EM algorithm is defined.

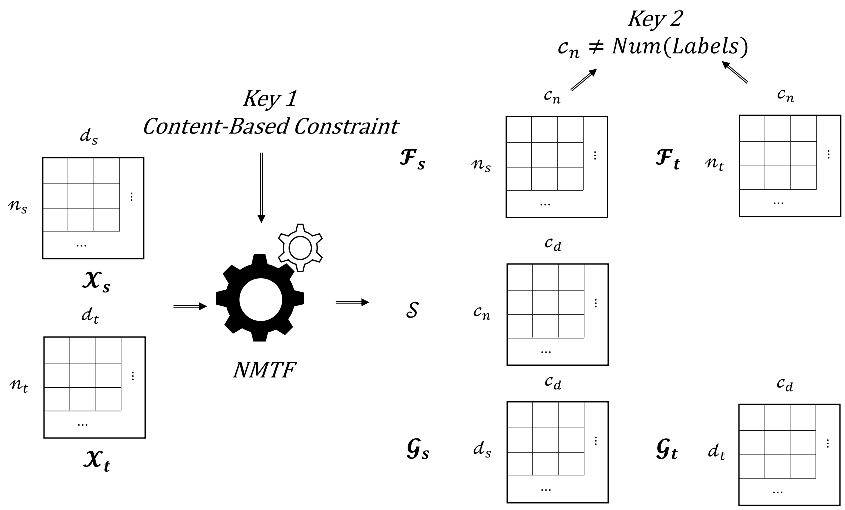

Fig. 1. Structure of cNMTF.

Algorithm 1 Alternately Iterative Algorithm

Input: X_s, X_t, E_s, E_t
Output: F_s, F_t, G_s, G_t, S, A

1. Initialize $F_s^{(0)}$, $F_t^{(0)}$, $G_s^{(0)}$, $G_t^{(0)}$, $S^{(0)}$;
2. $A^{(0)} = F_s^{(0)}$;
3. $t = 0$;
4. **repeat**
5. $t = t + 1$;
6. Compute $F_s^{(t)}$ using the update rule in Eq. (7);
7. Similarly compute $F_t^{(t)}$, $G_s^{(t)}$, $G_t^{(t)}$, $S^{(t)}$, and $A^{(t)}$;
8. **until** $\Psi^{(t-1)} - \Psi^{(t)} \le \varepsilon$ or $t \ge maxIteration$.

Definition 3.1. $G(x, z)$ is an auxiliary function for $F(x)$ if the following conditions are satisfied.

$G(x, z) \ge F(x)$, $G(x, x) = F(x)$.

Lemma 3.2. *If G is an auxiliary function, then F is non-increasing under the following update rule:*

$$x^{t+1} = arg\min_x G(x, x^t). \tag{8}$$

Proof. $F(x^{t+1}) \le G(x^{t+1}, x^t) \le G(x^t, x^t) = F(x^t)$. □

As long as we can prove that the update rule of F_s confirms to Eq. (8) for an appropriate auxiliary function, we can conclude that F_s would converge to a local minima.

Lemma 3.3. *Suppose that $z = F_{s(ij)} > 0$ and $G(x, z)$ is an auxiliary function for $F(z) = \Psi(F_{s(ij)} = z)$.*

$$G(x, z) = F(z) + \frac{\partial F(z)}{\partial z}(x - z) + \frac{(F_sSG_s^TG_sS^T + \beta H \circ (F_sA^T)A + \beta F_s)_{ij}}{z}(x - z)^2. \tag{9}$$

Proof. Clearly, $G(x, x) = F(x)$. Taylor expansion of $F(x)$ is:

$$F(x) = F(z) + \frac{\partial F(z)}{\partial z}(x - z) + \frac{1}{2}\frac{\partial^2 F(z)}{\partial z^2}(x - z)^2, \tag{10}$$

where $\frac{\partial^2 F(z)}{\partial^2 z} = 2(SG_s^TG_sS^T)_{jj} + 2\beta H \circ (A^TA)_{jj} + 2\beta$.

In order to show that G is an auxiliary function, we need to get $G(x, z) \ge F(x)$, i.e.,

$$\frac{(F_sSG_s^TG_sS^T + \beta H \circ (F_sA^T)A + \beta F_s)_{ij}}{z} \ge \frac{1}{2}\frac{\partial^2 F(z)}{\partial z^2}.$$

Lemma 3.4. $\frac{(SM^TM)_{ij}}{S_{ij}} = \frac{1}{S_{ij}}\sum_k S_{ik}(M^TM)_{kj} \ge (M^TM)_{jj}$.

Proof. Note that the elements of M^TM are always greater than or equal to 0, thus the condition of the inequality is that elements of S are always greater than or equal to 0.

$$\begin{aligned}\frac{(SM^TM)_{ij}}{S_{ij}} &= \frac{1}{S_{ij}}\sum_k S_{ik}(M^TM)_{kj} \\ &= \frac{1}{S_{ij}}S_{ij}(M^TM)_{jj} + \frac{1}{S_{ij}}\sum_{k\neq j} S_{ik}(M^TM)_{kj} \ge (M^TM)_{jj}.\end{aligned}$$ □

Given that $F_{s(ij)} = z$, we get following inequalities according to Lemma 3.4:

$$\frac{(\beta H \circ (F_sA^T)A)_{ij}}{z} \ge \beta H \circ (A^TA)_{jj},$$

$$\frac{(F_sSG_s^TG_sS^T + \beta F_s)_{ij}}{z} \ge (SG_s^TG_sS^T)_{jj} + \beta.$$

This establishes that G is an auxiliary function for F. □

Proof. To show that Algorithm 1 converges (i.e., Theorem 3.1), we need to show that the update rule of F_s follows Eq. (7). $\frac{\partial G(x,z)}{\partial x}$ is listed as follows:

$$\frac{\partial G(x, z)}{\partial x} = -2(X_sG_sS^T + \beta H \circ (X_sX_s^T)A + \beta A)_{ij} + 2(F_sSG_s^TG_sS^T + \beta H \circ (F_sA^T)A + \beta F_s)_{ij}\frac{x}{z}. \tag{11}$$

Solving $\frac{\partial G(x,z)}{\partial x} = 0$ for x, we get the update rule as mentioned in Eq. (7). Since G is the auxiliary function for F, the value of F is non-increasing. We can prove the convergence of update rules for F_t, G_s, G_t, S, and A in Algorithm 1 similarly. Thus, Algorithm 1 is guaranteed to converge to a locally-optimal solution. □

In the part, we describe the computational complexity of our method. Take F_s as an example, the time complexity of updating F_s is $O(d_sc_n(3c_d + n_s) + n_s^2(3c_n + d_s + 2) + c_n^2(c_d + n_s) + 2n_sc_n)$ per iteration. Empirically, the average number of iterations required for convergence is less than 100. Thus, our method scales linearly to dataset size.

Table 1
Dataset statistical information.

Domain	#news reports	#ratings	Vocabulary size
Economics	1,554	58,111	4,835
Culture	1,241	39,889	5,974
Law	1,963	2,50,657	7,064
Society	1,602	2,06,260	5,719

Table 2
Hyperparameters search space.

Hyperparameter	Search space
α	[0.1,0.5,1.0]
β	[0.1,0.5,1.0]
λ	[0.1,0.5]
Number of document clusters	[4,8,50]
Number of word clusters	[4,40,100]
Number of iterations	[10,20,50,200]

3.5. Prediction

Since we propose that the number of document clusters may not equal to that of labels, the output of Algorithm 1 cannot directly represent the prediction results. We thus develop a prediction method using the association matrix.

Particularly, we use the mapping matrix M to represent the association between document clusters and emotion labels. The size of M is $c_n \times l$. M is defined as:

$$F_s M = E_s. \tag{12}$$

We can use this association matrix to predict the emotion distribution of X_t as follows:

$$F_t M = F_t(F_s^{-1} E_s). \tag{13}$$

4. Experiments

4.1. Dataset

We evaluate cNMTF on a real-world dataset.[1] for cross-domain emotion distribution learning: 115,000 news reports from ChinaNews website[2] This dataset contains the collection of news documents from 39 specific domains. In our experiments, four domains which contain top numbers of documents are selected: Economics, Culture, Law, and Society. Each document is voted by readers over 8 basic emotions, i.e., "moved", "sympathy", "boring", "anger", "funny", "sad", "delighted", and "not-interested". In these 4 domains, the total number of emotion ratings from readers is 554,917. The documents with the emotion rating number less than 5 are discarded. In order to constitute the emotion distribution, we calculate the proportion that ratings of each emotion account for the total emotion ratings for each document. Note that the sum of such proportions of emotions in each document is 1. To pre-process our dataset, we filter out the non-Chinese characters and stop words in documents. The rare words that occur in 10 or less documents in the 4 selected domains are discarded to restrict the vocabulary.

We show the statistical information of our dataset, i.e., the number of news reports, the number of emotion ratings, and the size of vocabularies for each domain in Table 1.

4.2. Experimental setting

We use $tf - idf$ weighting scheme [63] to initialize the document–word matrices, i.e., X_s and X_t. PLSA [64] is employed to initialize the following matrices. F_s and F_t are initialized as the document-topic distributions while G_s and G_t are initialized as the transpose of word-topic distributions by PLSA. Note that the initialization of the auxiliary matrix A is the same as that of F_s. The shared matrix S is initialized randomly.

We construct 12 cross-domain emotion distribution learning tasks among the 4 domains, e.g., Culture → Economics. In each pair, all documents in the source domain and 10% documents in the target domain are used for training. Another 10% documents and the remaining 80% in the target domain are used as the validation set and the testing set, respectively.

In the validation set, the idea of grid search [65] is borrowed to select the best values of hyperparameters, i.e., α, β, λ, the numbers of document clusters, word clusters, and iterations. Then we use the selected best values of hyperparameters to predict the emotion distribution of each document the testing set. Tables 2 and 3 show the search space of hyperparameters for grid search of our model and the optimal parameter settings for different cross-domain learning tasks, respectively. All the following experiments are conducted on a workstation equipped with an Intel(R) Xeon(R) CPU E5-2680 v3 167 @ 2.50 GHz, 8 cores and 128G memory.

4.3. Baselines

We compare our model with some state-of-the-art baselines: Cross-Domain Text Classification (CDTC) model [5], Triplex Transfer Learning (TriTL) method [12], Cross-Domain Emotion Tagging (CDET) method [6], Transitive Transfer Learning (TTL) method [13], Contextual Sentiment Topic Model (CSTM) [7], Multi-Task Convolutional Neural Network (MTCNN) [1], and Dependency Embedded Recursive Neural Network (DERNN) [3]. All the aforementioned models follow the experimental setting in Section 4.2. The hyperparameters of all these models are selected by grid searching on the validation set.

To measure the similarity or distance between the predicted and real emotion distributions, we choose 6 fine-grained metrics [1,66] for evaluation, i.e., Euclidean, Sϕrensen, Squared$\mathcal{X}^2$, Intersection, Fidelity, and Cosine. By treating emotion distribution learning as a classification problem, we also adopt *Accu*@1 [6] and *NCDG*@1 [67] as coarse-grained metrics.

4.4. Experimental results

Table 4 shows the average value and the mean of variances of cNMTF and other baselines over the 12 domain adaptation tasks on different metrics with 10 independent runs under the same experimental setting. The best average value on each metric is highlighted by underline. We can conclude that on most metrics, the proposed cNMTF performs much better than other baseline models. One possible reason is that the hypothesis of many-to-many relationships between document clusters and labels and the content-based constraint provide more reasonable information for our model. On average, the result of cNMTF is 2% better than the best-performing baseline (i.e., MTCNN), although MTCNN uses the word-embedding pre-trained on Wikipedia with additional information provided by the external corpus. In addition, MTCNN costs more than 2 h for one cross-domain learning task, which is not adaptive to a large-scaled dataset, while our cNMTF only costs several minutes on average.

To analyze the stability of different models, the mean of variances indicate that our model is more stable than CDET, CSTM, MTCNN, and DERNN. We can observe that matrix factorization

[1] https://github.com/hostnlp/ChinaNews-Data
[2] http://www.chinanews.com

Table 3
Best hyperparameter values in 12 cross-domain learning tasks. E, C, L, and S represent different domains, i.e., Economics, Culture, Law, and Society, respectively.

Task	α	β	λ	#document clusters	#word clusters	#iterations
$E \rightarrow C$	0.1	1.0	0.5	8	4	20
$E \rightarrow L$	0.1	0.1	0.5	8	40	10
$E \rightarrow S$	0.5	0.1	0.1	8	4	10
$C \rightarrow E$	0.5	1.0	0.5	4	40	20
$C \rightarrow L$	0.1	0.5	0.1	8	40	10
$C \rightarrow S$	0.5	1.0	0.1	4	4	20
$L \rightarrow E$	0.1	1.0	0.1	4	4	50
$L \rightarrow C$	0.1	0.5	0.1	8	4	10
$L \rightarrow S$	0.5	0.5	0.1	8	4	50
$S \rightarrow E$	1.0	1.0	0.1	8	4	10
$S \rightarrow C$	0.1	1.0	0.1	8	4	10
$S \rightarrow L$	0.1	0.1	0.1	4	40	50

Table 4
Average performance of different models. Within 10 independent runs under the same setting, two results are separated by "|". The first one is the average value of metrics in 12 tasks, with "•" indicating significance difference of the two-tailed t-tests with 5% significance level between cNMTF and other models. The second result represents the mean of variances of all tasks within 10 runs. For all average values, "↓" after the metric indicates smaller is better while "↑" indicates larger is better. The best performance on each measure is highlighted by underline. Besides, the smaller the mean of variances, the better the model stability is.

Method	Euclidean(↓)	Sϕrensen(↓)	Squared$\mathcal{X}^2$(↓)	Intersection(↑)	Fidelity(↑)	Cosine(↑)	Accu@1(↑)	NCDG@1(↑)
CDTC	0.7847•|0.0002	0.7326•|0.0001	1.2702•|0.0006	0.2674•|0.0001	0.4544•|0.0001	0.3258•|0.0002	0.1490•|0.0002	0.2675•|0.0004
TriTL	0.5545•|0.0000	0.5877•|0.0000	0.8794•|0.0001	0.4123•|0.0000	0.6324•|0.0000	0.4908•|0.0000	0.1197•|0.0001	0.2331•|0.0002
CDET	0.4428•|0.0016	0.4814•|0.0015	0.6812•|0.0015	0.5279•|0.0015	0.7736•|0.0015	0.6857•|0.0017	0.4027•|0.0013	0.5526•|0.0017
TTL	0.5585•|0.0001	0.5911•|0.0001	0.9114•|0.0005	0.4089•|0.0001	0.6288•|0.0001	0.4854•|0.0002	0.1208•|0.0004	0.2342•|0.0007
CSTM	0.4290•|0.0006	0.4425•|0.0008	0.5792•|0.0008	0.5450•|0.0008	0.7332•|0.0007	0.6834•|0.0007	0.4235•|0.0006	0.5911•|0.0007
MTCNN	0.4336•|0.0009	0.4259|0.0007	0.5618•|0.0006	0.5641•|0.0008	0.7784•|0.0007	0.7053•|0.0008	0.4645•|0.0007	0.6433•|0.0007
DERNN	0.4560•|0.0029	0.4729•|0.0032	0.6314•|0.0029	0.5275•|0.0026	0.7606•|0.0028	0.6649•|0.0030	0.3795•|0.0027	0.5300•|0.0032
cNMTF-c	0.4778•|0.0001	0.4782•|0.0001	0.6562•|0.0003	0.5218•|0.0001	0.7463•|0.0000	0.6445•|0.0002	0.3736•|0.0002	0.5408•|0.0003
cNMTF-n	0.4211|0.0001	0.4292|0.0001	0.5521|0.0002	0.5708|0.0001	0.7878|0.0000	0.7184|0.0001	0.4558•|0.0002	0.6336•|0.0003
cNMTF	0.4200|0.0001	0.4287|0.0001	0.5520|0.0002	0.5714|0.0001	0.7880|0.0000	0.7193|0.0001	0.4835|0.0002	0.6610|0.0003

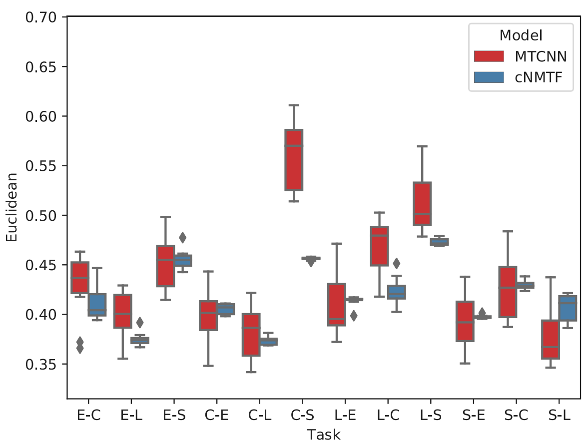

Fig. 2. Euclidean(↓).

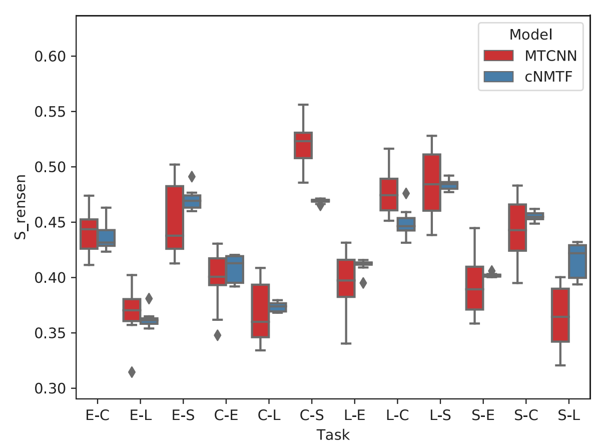

Fig. 3. Sϕrensen(↓).

based models (i.e., CDTC, TriTL, TTL, and cNMTF) have the better stability than other models, and cNMTF performs the best among these models. As a matrix factorization based model with a reasonable hypothesis and the content-based constraint, cNMTF balances well between the model performance and the stability.

To statistically evaluate the differences between cNMTF and other models, we also performed the two-tailed t-tests with 5% significance level on paired models in terms of different metrics. For the pair of each baseline model and cNMTF, we computed the average value of each metric on 12 cross-domain learning tasks for each independent run and conducted t-tests. The result indicates that the proposed cNMTF outperforms most baselines significantly.

In order to validate the effectiveness of our two main contributions, i.e., the content-based constraint and the flexible relationship between document clusters and emotion labels, we conduct ablation experiments in these 12 tasks. The proposed cNMTF-c follows the idea that the relationship between document clusters

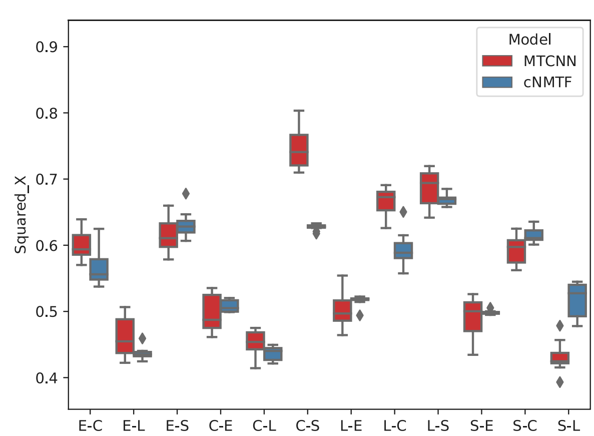

Fig. 4. Squared$\mathcal{X}^2$ ↓.

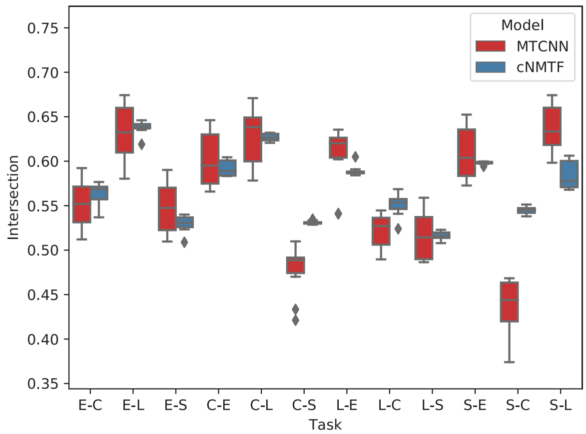

Fig. 5. Intersection(↑).

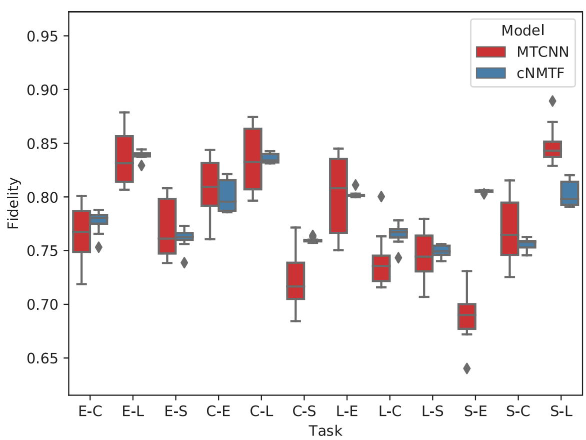

Fig. 6. Fidelity(↑).

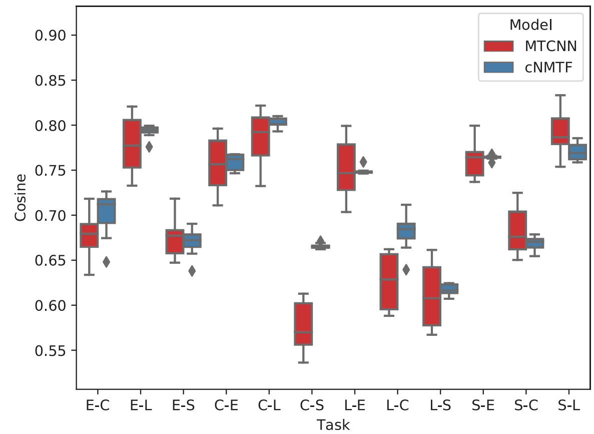

Fig. 7. Cosine(↑).

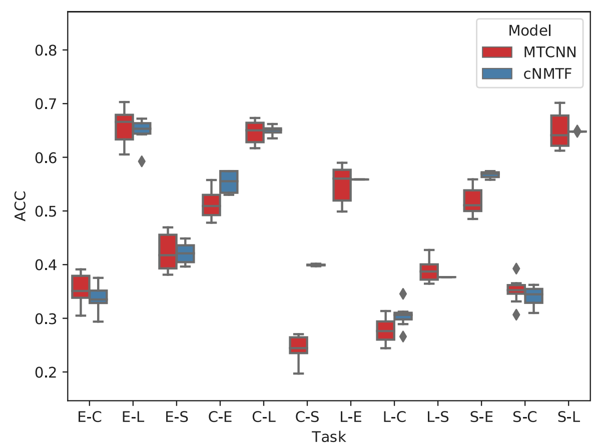

Fig. 8. Accu@1(↑).

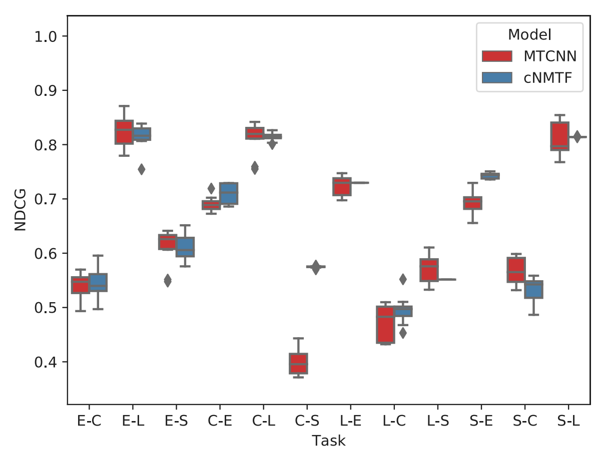

Fig. 9. NDCG@1(↑).

and emotion labels is not always one-to-one without the content-based constraint. For cNMTF-n, the number of document clusters always equal to that of labels while the content-based constraint is added. Compared to cNMTF-c and cNMTF-n, cNMTF performs better on all metrics. These results indicate that both two contributions we developed for NMTF are beneficial to the performance of cross-domain emotion distribution learning.

To analyze the specific performance of all domain adaptation tasks on different metrics, we choose the best performing baseline (i.e., MTCNN) and compare it with our cNMTF. Figs. 2 to 9 show the value of different metrics within 10 independent runs on 12 cross-domain learning tasks of the two models, in which E, C, L, and S represent different domains, i.e., Economics, Culture, Law, and Society, respectively. It can be easily observed that in most learning scenarios, cNMTF balances well between the model performance and the stability. We can also conclude that the domain adaptation tasks on different source domains and target domains have significantly different performance. One possible reason is that domains with different feature distributions have different latent relationships. For example, Society → Economics performs better than Society → Culture. Clearly, there are some common words in Society and Economics domains. On the other hand, words in Economics are less used in Culture.

5. Conclusions

In this paper, we propose a constrained optimization approach based on NMTF for cross-domain emotion distribution learning (cNMTF). We introduce the association matrix between document clusters and word clusters to help the domain adaptation task. Note that we propose that the relationship between document clusters and emotion labels is not always one-to-one. We also endow cNMTF with a novel content-based constraint based on the hypothesis that documents belonging to the same clusters must have similar content. Experiments on 12 real-world cross-domain emotion distribution learning tasks indicate that our method consistently outperforms state-of-the-art baselines.

CRediT authorship contribution statement

Xiaorui Qin: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Yufu Chen:** Software, Validation, Formal analysis, Writing - review & editing, Visualization, Data curation. **Yanghui Rao:** Software, Writing - original draft, Formal Analysis, Writing - review & editing, Validation, Formal analysis, Data curation, Supervision. **Haoran Xie:** Formal analysis, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Man Leung Wong:** Supervision, Project administration, Funding acquisition. **Fu Lee Wang:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research described in this article has been supported by HKIBS Research Program, China Grant (HCRG-201-002), Faculty Research, China Grant (DB21A9) and LEO Dr David P. Chan Institute of Data Science of Lingnan University, Hong Kong, and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS16/E01/19).

References

[1] Y. Zhang, J. Fu, D. She, Y. Zhang, S. Wang, J. Yang, Text emotion distribution learning via multi-task convolutional neural network, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI, 2018, pp. 4595–4601.
[2] Z. Zhao, X. Ma, Text emotion distribution learning from small sample: A meta-learning approach, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, 2019, pp. 3955–3965.
[3] C. Wang, B. Wang, W. Xiang, M. Xu, Encoding syntactic dependency and topical information for social emotion classification, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2019, pp. 881–884.
[4] M. Long, H. Zhu, J. Wang, M.I. Jordan, Unsupervised domain adaptation with residual transfer networks, in: Proceedings of the Advances in Neural Information Processing Systems, NIPS, 2016, pp. 136–144.
[5] F. Zhuang, P. Luo, H. Xiong, Q. He, Y. Xiong, Z. Shi, Exploiting associations between word clusters and document classes for cross-domain text categorization, in: Proceedings of the SIAM International Conference on Data Mining, SDM, 2010, pp. 13–24.
[6] Y. Zhang, N. Zhang, L. Si, Y. Lu, Q. Wang, X. Yuan, Cross-domain and cross-category emotion tagging for comments of online news, in: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2014, pp. 627–636.
[7] Y. Rao, Contextual sentiment topic model for adaptive social emotion classification, IEEE Intell. Syst. 31 (1) (2016) 41–47.
[8] A. Salah, M. Ailem, M. Nadif, Word co-occurrence regularized non-negative matrix tri-factorization for text data co-clustering, in: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI, 2018, pp. 3992–3999.
[9] T. Li, V. Sindhwani, C.H.Q. Ding, Y. Zhang, Knowledge transformation for cross-domain sentiment classification, in: Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2009, pp. 716–717.
[10] F. Wang, T. Li, C. Zhang, Semi-supervised clustering via matrix factorization, in: Proceedings of the SIAM International Conference on Data Mining, SDM, 2008, pp. 1–12.
[11] T. Li, C.H.Q. Ding, Y. Zhang, B. Shao, Knowledge transformation from word space to document space, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2008, pp. 187–194.
[12] F. Zhuang, P. Luo, C. Du, Q. He, Z. Shi, Triplex transfer learning: exploiting both shared and distinct concepts for text classification, in: Proceedings of the 6th ACM International Conference on Web Search and Data Mining, WSDM, 2013, pp. 425–434.
[13] B. Tan, Y. Song, E. Zhong, Q. Yang, Transitive transfer learning, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, 2015, pp. 1155–1164.
[14] H. Ito, T. Amagasa, An optimization scheme for non-negative matrix factorization under probability constraints, in: Proceedings of the IEEE International Conference on Big Data and Smart Computing, BigComp, 2019, pp. 1–8.
[15] Y. Chen, J. Wu, J. Lin, R. Liu, H. Zhang, Z. Ye, Affinity regularized non-negative matrix factorization for lifelong topic modeling, IEEE Trans. Knowl. Data Eng. 32 (7) (2020) 1249–1262.
[16] M. Tian, C. Leng, H. Wu, A. Basu, Total variation constrained graph-regularized convex non-negative matrix factorization for data representation, IEEE Signal Process. Lett. 28 (2021) 126–130.
[17] Y. Wang, A. Pal, Detecting emotions in social media: A constrained optimization approach, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI, 2015, pp. 996–1002.
[18] G. Katz, N. Ofek, B. Shapira, Consent: Context-based sentiment analysis, Knowl.-Based Syst. 84 (2015) 162–178.
[19] M.S. Akhtar, A. Ekbal, E. Cambria, How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble [application notes], IEEE Comput. Intell. Mag. 15 (1) (2020) 64–75.
[20] E. Cambria, S. Poria, A.F. Gelbukh, M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intell. Syst. 32 (6) (2017) 74–80.
[21] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, E. Cambria, Learning word representations for sentiment analysis, Cogn. Comput. 9 (6) (2017) 843–851.
[22] E. Cambria, S. Poria, D. Hazarika, K. Kwok, SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings, in: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI, 2018, pp. 1795–1802.
[23] J. Feng, Y. Rao, H. Xie, F.L. Wang, Q. Li, User group based emotion detection and topic discovery over short text, World Wide Web 23 (3) (2020) 1553–1587.
[24] A. Khatua, A. Khatua, E. Cambria, A tale of two epidemics: Contextual word2vec for classifying twitter streams during outbreaks, Inf. Process. Manage. 56 (1) (2019) 247–257.
[25] Y. Li, S. Wang, Y. Ma, Q. Pan, E. Cambria, Popularity prediction on vacation rental websites, Neurocomputing 412 (2020) 372–380.
[26] A. Valdivia, E. Martínez-Cámara, I. Chaturvedi, M.V. Luzón, E. Cambria, Y. Ong, F. Herrera, What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules, J. Ambient Intell. Humaniz. Comput. 11 (1) (2020) 39–52.
[27] A. Khatua, A. Khatua, E. Cambria, Predicting political sentiments of voters from twitter in multi-party contexts, Appl. Soft Comput. (2020) 106743.
[28] I. Chaturvedi, E. Cambria, D. Vilares, Lyapunov filtering of objectivity for Spanish Sentiment Model, in: Proceedings of 2016 International Joint Conference on Neural Networks, IJCNN, 2016 pp. 4474–4481.
[29] E. Cambria, Y. Li, F.Z. Xing, S. Poria, K. Kwok, SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis, in: Proceedings of the 29th ACM International Conference on Information and Knowledge Management, CIKM, 2020, pp. 105–114.
[30] Q. Yang, Y. Rao, H. Xie, J. Wang, F.L. Wang, W.H. Chan, Segment-level joint topic-sentiment model for online review analysis, IEEE Intell. Syst. 34 (1) (2019) 43–50.
[31] F. Bravo-Marquez, M. Mendoza, B. Poblete, Meta-level sentiment models for big social data analysis, Knowl.-Based Syst. 69 (2014) 86–99.
[32] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowl.-Based Syst. 69 (2014) 78–85.
[33] T.L. Griffiths, M. Steyvers, D.M. Blei, J.B. Tenenbaum, Integrating topics and syntax, in: Proceedings of the Advances in Neural Information Processing Systems, NIPS, 2004, pp. 537–544.
[34] W. Li, L. Zhu, Y. Shi, K. Guo, E. Cambria, User reviews: Sentiment analysis using lexicon integrated two-channel CNN-LSTM family models, Appl. Soft Comput. 94 (2020) 106435.
[35] E. Cambria, B.W. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intell. Syst. 28 (2) (2013) 15–21.
[36] J. Xu, L. Zhao, H. Yan, Q. Zeng, Y. Liang, X. Sun, LexicalAT: Lexical-based adversarial reinforcement training for robust sentiment classification, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, 2019, pp. 5517–5526.
[37] M. Huang, H. Xie, Y. Rao, Y. Liu, L.K.M. Poon, F.L. Wang, Lexicon-based sentiment convolutional neural networks for online review analysis, IEEE Trans. Affect. Comput. (2020).

[38] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2002, pp. 79–86.
[39] S. Poria, E. Cambria, A.F. Gelbukh, Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2015, pp. 2539–2544.
[40] A. Mishra, K. Dey, P. Bhattacharyya, Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL, 2017, pp. 377–387.
[41] B. Pang, L. Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, in: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, ACL, 2005, pp. 115–124.
[42] C.N. dos Santos, M. Gatti, Deep convolutional neural networks for sentiment analysis of short texts, in: Proceedings of the 25th International Conference on Computational Linguistics, COLING, 2014, pp. 69–78.
[43] S. Poria, I. Chaturvedi, E. Cambria, A. Hussain, Convolutional MKL based multimodal emotion recognition and sentiment analysis, in: Proceedings of 16th International Conference on Data Mining, ICDM, 2016, pp. 439–448.
[44] M. Huang, H. Xie, Y. Rao, J. Feng, F.L. Wang, Sentiment strength detection with a context-dependent lexicon-based convolutional neural network, Inform. Sci. 520 (2020) 389–399.
[45] M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, U.R. Acharya, ABCDM: An attention-based bidirectional CNN-rnn deep model for sentiment analysis, Future Gener. Comput. Syst. 115 (2021) 279–294.
[46] Y. Li, K. Zhang, J. Wang, X. Gao, A cognitive brain model for multimodal sentiment analysis based on attention neural networks, Neurocomputing 430 (2021) 159–173.
[47] H.D. III, D. Marcu, Domain adaptation for statistical classifiers, 2011, CoRR, arXiv:abs/1109.6341.
[48] F.Z. Xing, F. Pallucchini, E. Cambria, Cognitive-inspired domain adaptation of sentiment lexicons, Inf. Process. Manage. 56 (3) (2019) 554–564.
[49] J. Blitzer, R.T. McDonald, F. Pereira, Domain adaptation with structural correspondence learning, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2006, pp. 120–128.
[50] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification, in: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, ACL, 2007, pp. 440–447.
[51] S.J. Pan, X. Ni, J. Sun, Q. Yang, Z. Chen, Cross-domain sentiment classification via spectral feature alignment, in: Proceedings of the 19th International Conference on World Wide Web, WWW, 2010, pp. 751–760.
[52] L. Li, X. Jin, M. Long, Topic correlation analysis for cross-domain text classification, in: Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI), 2012, pp. 998–1004.
[53] R. Xia, C. Zong, X. Hu, E. Cambria, Feature ensemble plus sample selection: Domain adaptation for sentiment classification, IEEE Intell. Syst. 28 (3) (2013) 10–18.
[54] S.L. Lo, E. Cambria, R. Chiong, D. Cornforth, A multilingual semi-supervised approach in deriving singlish sentic patterns for polarity detection, Knowl. Based Syst. 105 (2016) 236–247.
[55] J. Pang, Y. Rao, H. Xie, X. Wang, F.L. Wang, T. Wong, Q. Li, Fast supervised topic models for short text emotion detection, IEEE Trans. Cybern. (2019) 1–14.
[56] C. Strapparava, R. Mihalcea, SemEval-2007 Task 14: Affective text, in: Proceedings of the 4th International Workshop on Semantic Evaluations (ACL), 2007, pp. 70–74.
[57] T. Li, Y. Zhang, V. Sindhwani, A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge, in: Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, ACL/IJCNLP, 2009, pp. 244–252.
[58] J. Pan, G. Xue, Y. Yu, Y. Wang, Cross-lingual sentiment classification via bi-view non-negative matrix tri-factorization, in: Proceedings of Advances in Knowledge Discovery and Data Mining - 15th Pacific-Asia Conference, PAKDD, in: Lecture Notes in Computer Science, vol. 6634, 2011, pp. 289–300.
[59] E. Sherkat, J. Velcin, E.E. Milios, Fast and simple deterministic seeding of kmeans for text document clustering, in: Proceedings of the 9th International Conference of the CLEF Association, CLEF, 2018, vol. 1018, pp. 76–88.
[60] R. Janani, S. Vijayarani, Text document clustering using spectral clustering algorithm with particle swarm optimization, Expert Syst. Appl. 134 (2019) 192–200.
[61] Z. Xu, Y. Ke, Effective and efficient spectral clustering on text and link data, in: Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM, 2016, pp. 357–366.
[62] P. Chalise, Y. Ni, B.L. Fridley, Network-based integrative clustering of multiple types of genomic data using non-negative matrix factorization, Comput. Biol. Med. 118 (2020) 103625.
[63] C.W. Schmidt, Improving a tf-idf weighted document vector embedding, 2019, CoRR arXiv:abs/1902.09875.
[64] T. Hofmann, Probabilistic latent semantic indexing, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 1999, pp. 50–57.
[65] H.A. Fayed, A.F. Atiya, Speed up grid-search for parameter selection of support vector machines, Appl. Soft Comput. 80 (2019) 202–210.
[66] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, X. Geng, Emotion distribution learning from texts, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2016, pp. 638–647.
[67] K.H. Lin, H. Chen, Ranking reader emotions using pairwise loss minimization and emotional distribution regression, in: Proceedings of 2008 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2008, pp. 136–144.