Research Article

Deng Pan* and Hyunho Yang

A text-Image feature mapping algorithm based on transfer learning

<https://doi.org/10.1515/phys-2018-0134>

Received Oct 09, 2018; accepted Nov 14, 2018

Abstract: The traditional uniform distribution algorithm does not filter the image data when extracting the approximate features of text-image data under the event, so the similarity between the image data and the text is low, which leads to low accuracy of the algorithm. This paper proposes a text-image feature mapping algorithm based on transfer learning. The existing data is filtered by ‘clustering technology’ to obtain similar data with the target data. The significant text features are calculated through the latent Dirichlet allocation (LDA) model and information gain based on Gibbs sampling. Bag of visual word (BOVW) model and Naive Bayesian method are used to model image data. With the help of the text-image co-occurrence data in the same event, the text feature distribution is mapped to the image feature space, and the feature distribution of image data under the same event is approximated. Experimental results show that the proposed algorithm can obtain the feature distribution of image data under different events, and the average cosine similarity is as high as 92%, the average dispersion is as low as 0.06%, and the accuracy of the algorithm is high.

Keywords: Transfer learning, text-image, feature mapping, clustering, LDA model, BOVW model

PACS: 07.05.Kf, 07.05.Pj, 07.05.Tp

1 Introduction

With the development of online information dissemination technology, the amount of event information accompanied by text-image is increasing. Traditional text mining

technology can no longer satisfy people’s learning needs of multimedia information. However, it is still difficult to develop knowledge models directly in the feature space of multimedia data, especially in the image feature space. Whether the mature text mining technology alongside sufficient text information on the Internet can be used to assist the knowledge learning of image data is a hotspot of current research.

Reference [1] proposes an adaptive control scale-invariant feature transform (SIFT) feature uniform distribution algorithm (called uniform distribution algorithm) based on the characteristics of stop and reverse (SAR) image data. By using local texture features and combining them with optimization screening strategy, the SIFT feature points can be reasonably distributed in image space and scale space by adaptively controlling the distribution of features in different spaces while ensuring the stability and accuracy of feature points. But without considering the timeliness of image data, the accuracy of this algorithm is not high. In Reference [2], a fast connected component labeling algorithm implemented on an field-programmable gate array (FPGA) is proposed. Run-length encoding is used to optimize image annotation, which can reduce the number and length of tags and extract the features of components during the run-length encoding. Due to the complexity of the algorithm, the efficiency of the algorithm is low. In Reference [3], by calculating image features and delays, enhanced scanning (DE-MRI) is used to analyze heterogeneous machine learning, and an uncertainty assessment framework with potential ablation target recognition is constructed. However, due to the relationship between image features and delay, in the analysis of heterogeneous machine learning situation, data is not sufficient, and data analysis efficiency is low, which has certain limitations.

Transfer learning method develops a compact and effective representation from the annotated data of a source domain and a few annotated or unannotated data of the target domain, and then applies the learning feature representation method to the learning task of the target domain. In this method, not only self-annotated data but also unannotated data are used, so it is neither supervised learn-

*Corresponding Author: Deng Pan: College of Computer Information and Communication Engineering, Jiujiang University, Jiujiang, 332005, China, 29882; Email: 4035panda@gmail.com

Hyunho Yang: School of Computer Information and Communication Engineering, Kunsan National University, 54150, South Korea; Email: hhyang@kunsan.ac.kr

ing [4], nor unsupervised learning and semi-supervised learning, but a new machine learning method. During feature migration, even if the data in the source data space and the target data space do not intersect at the instance level, they may be related at the feature level [5]. Data with two feature perspectives can be used to establish a link between two different feature spaces. These data are not necessarily used as training data for knowledge learning, but they can act as a role of dictionary. Taking a subject event as the background, the sufficient text-image information about the event on the Internet is used as a basis for knowledge migration.

Aiming at these problems, in this paper a text-image feature mapping algorithm based on transfer learning is proposed, which uses clustering technology to filter the existing data to find the data very similar to the target data [6]. The significant text features are calculated by LDA model based on Gibbs sampling and information gain. The BOVW model and naive Bayesian method are used to model the subject of image data. With the help of the text-image co-occurrence data [7] in same event, the text feature distribution is mapped to the image feature space, and the feature distribution of image data under the same event is approximated.

2 A text-image feature mapping algorithm based on transfer learning

2.1 Transfer learning algorithm for clustering text

Although the existing auxiliary data is out of date, there should be some data in the existing data that is very similar to the test data and can be used to help target tasks learning [8]. Therefore, clustering technology is used to find data that is very similar to test data from existing data.

2.1.1 Introduction to clustering

Clustering is an important form of data mining. The purpose of text clustering is to group large-scale text datasets into multiple classes, and make the text in the same class have a high degree of similarity, while the text between different classes is quite different [9]. As a function of data mining, clustering can be used as an independent tool to obtain data distribution, observe the characteristics of

each cluster, and focus on some specific clusters for further analysis. At the same time, clustering technology can also be used as a pre-treatment step of other algorithms to effectively improve the performance of other algorithms [10].

2.1.2 Text representation and text similarity formula

According to the traditional vector space model (VSM) representation, the text content can be expressed as a weighted feature vector. Let D be a text set, d_i is a text in the set, t is a feature word, t_i is the i -th feature word, and w_i is a weight of the i -th feature word.

$$d_i = (t_1, w_1; t_2, w_2; \dots; t_n, w_n) \quad (1)$$

Where, the weight w_i can be represented by the tf-idf weight of each feature. The tf-idf formula is as follows:

$$tf - idf = \sum_{d \in D} tf(d, t) \cdot \log \frac{|D|}{df(t)} \quad (2)$$

Where, $tf(d, t)$ is the word frequency of word t in text d , $df(t)$ is the number of text containing word t in text set D , and $|D|$ represents the number of text contained in text set D .

The similarity between two texts can be calculated by the cosine of the angle α between two vectors. Assuming that two texts are $d_1 = (t_1, w_1; t_2, w_2; \dots; t_n, w_n)$ and $d_2 = (t_1, \sigma_1; t_2, \sigma_2; \dots; t_n, \sigma_n)$, the similarity between d_1 and d_2 is expressed as follows:

$$\sim (d_1 d_2) = \cos \alpha = \frac{\sum_{i=1}^n w_i \times \sigma_i}{\left(\sum_{i=1}^n w_i^2 \times \sum_{i=1}^n \sigma_i^2 \right)^{\frac{1}{2}}} \quad (3)$$

The greater the value of $\sim (d_1 d_2)$ is, the more similar the two texts are, of which w is the weight of the feature and σ is the approximate text weight.

2.1.3 Algorithm principle

Firstly, the auxiliary training data are clustered together with the target training data [11]. The result of clustering is that the intra-cluster similarity is high, and that of the data inter-cluster is different. Therefore, after clustering, no auxiliary data clustered in the same cluster with the target training data is filtered out. All that is left is data with high similarity to the target data, and training them with the target data will greatly improve the performance of the classifier [12]. The definitions of some basic symbols used in the paper are given below.

Definition 2.1.1 * set X_b as the target sample space and X_a as the auxiliary sample space. * set $Y = \{0, 1\}$ as a class space.

Definition 2.1.2 (test data set) $S = \{(x_i^t)\}$, of which $x_i^t \in X_b$, $i = 1, 2, \dots, k$, and k are the number of elements of set S .

Definition 2.1.3 (training data set) The training dataset consists of two parts:

$T_b = \{(x_j^b, c(x_j^b))\}$, where $x_j^b \in X_b$, $j = 1, 2, \dots, m$;

$T_a = \{(x_i^a, c(x_i^a))\}$, where $x_i^a \in X_a$, $i = 1, 2, \dots, n$;

where, $c(x)$ is the real class label of the instance, t is the feature word, i and j are the number of permutations, is the target training data set, T_a is the auxiliary training data set. m and n are the size of target training dataset and auxiliary training dataset respectively.

2.1.4 Algorithm steps

Input: two training datasets T_a and T_b , a test data set S .

Output: classification result $h_t(X_t)$.

Read the training data T_a and T_b .

The training data are classified into N classes according to class labels: T_i ($i = 1, \dots, N$), of which T_i is the instances set of classes labeled i ;

$$\text{For } i \leftarrow 1 \text{ to } N \quad (4)$$

a. call a basic clustering algorithm to cluster T_i and return clustering results.

b. scan T_i , delete auxiliary data from instances that are not clustered with target data.

End for;

Call a basic classification algorithm and get a classification model according to the filtered training data and test data S .

$$h_t : X \rightarrow Y.$$

Test the performance of the classification model on S and output it [13].

2.2 A text-image feature mapping algorithm based on transfer learning

Based on the previous section, the existing data is filtered by clustering technique [14], and the data which is very similar to the target data is obtained. Data with two feature perspectives are used to establish a link, and two different

feature spaces are connected. These data are not necessarily used as training data for knowledge learning, but they can act as a dictionary. Taking a subject event as the background, sufficient text-image information about the event on the internet is used as a basis for knowledge migration.

2.2.1 Text-image co-occurrence data constrained by events

In the heterogeneous spatial learning model, the difficulty of the whole learning process will be greatly reduced if a data with two feature spatial perspectives is used as an aid [15]. The heterogeneous spatial learning model under event constraint provides the possibility. The text-image co-occurrence data under event constraints are given here. E is an event set, event $e \in E$; V is the whole image data set, and the relevant image $\{v\} \in V$ under event e . D is the whole text data set, and the text set under event e is $\{d\} \in D$; U_v is the image feature space, and U_d is the text feature space. Text-image co-occurrence data instances $vd \in S$ and S are co-occurrence data set, f are operation coefficients, and $u_v \in U_v$ and $u_d \in U_d$ are corresponding features of image data instances and text data instances, respectively. Under the constraint of events, the text-image co-occurrence data vd is formally described at the feature level as follows:

$$P(u_v, u_d) = \int_D P(u_v, d) P(u_d | d) dd \quad (4)$$

$$P(u_v, u_d) = \int_V P(v, u_d) P(u_v | v) dv \quad (5)$$

$$P(u_v, u_d) = \iint_{vd} P(v, d) P(u_d | d) P(u_v | v) dv dd \quad (6)$$

where, $P(u_d | d)$ and $P(u_v | v)$ are feature extraction processes.

2.2.2 Text subject modeling

The LDA model based on Gibbs sampling is used to extract subject information from text sets for modeling [16], and the probability model is:

$$w_i | z_i, \phi^{(z_i)} \sim \text{disc}(\phi^{(z_i)}), \phi \sim \text{dir}(\beta) \quad (7)$$

$$z_i | \theta^{(d_i)} \sim \text{disc}(\theta^{(d_i)}), \theta \sim \text{dir}(\alpha)$$

In order to deal with the new text outside the event training text and facilitate parameter reasoning, the symmetric $\text{dir}(\alpha)$ and $\text{dir}(\beta)$ prior probability assumptions are made

for $\theta^{(d)}$ and $\phi^{(z)}$. In order to obtain the probability distribution of text subjects, the posterior probability $P(w|z)$ of lexicon w for text subjects is calculated instead of ϕ and θ , and then ϕ and θ are calculated indirectly by Gibbs sampling. By calculating the most discriminant feature in each subject feature space, the feature which has the more information gain can be as the significant text feature.

2.2.3 Image data modeling

A Naive Bayesian model is used to model the image. Firstly, the speeded up robust features (SURF) are computed and the bag of visual words (BOVW) model is established. Image v is considered as a set of visual words. Each visual word f comes from the visual vocabulary F , $v = \{f | f \in F\}$, and F represents the whole image feature space. According to the feature independence hypothesis, the image classification model is defined as: an event category c determines an image feature distribution $P(f \in F | c)$. Through this model, the maximum posteriori is used to infer the image's classification objective function $h_{NB} : V \rightarrow C$, and the image subject category modeling is completed. For target image v , the subject categories are:

$$h_{NB} = \arg \max_c p(c) \prod_{f \in v} P(f|c) \quad (8)$$

2.2.4 Text-image feature mapping

Both text subject modeling and image subject modeling belong to discrete object models. The feature independence hypothesis [17] can be applied to their features, that is, each feature independently affects the posterior probability of an instance under a given event category. In the process of text-image feature migration, the problem of feature migration can be greatly simplified by separating text features and image features for mapping [18]. Figure 1 is a schematic diagram of text-image feature migration:

The category label of each text in D under event constraint is the same as that of the image target category c . Text d is represented by a subject feature word bag as $d = \{t | t \in T\}$. Thematic feature dictionary T is the subject vocabulary in a text feature space. At the same time, there is a $S = \{(v, d)\}$ set of text-image co-occurrence data under corresponding event. To infer the image feature distribution $P(f|c)$ under event category c , the most significant text features in text set D are first computed, and then the most significant text features are mapped to the image feature space by means of text-image co-occurrence data set S . Distribution of image features under the target category

Figure 1: Text image feature migration under event constraint

is inferred from text significant features and text-image co-occurrence data in event text sets.

$$P(f|c) = N_c \sum_{w \in W(c)} P(f|w, c, S) P(w|c, D) \quad (9)$$

Where, $W(c)$ is the most prominent text feature set in the text set D under event category c , N_c is normalization coefficient, $P(w|c, D)$ is text feature distribution under event category c , and $P(f|w, c, S)$ is image's feature conditional distribution probability on text-image co-occurrence data.

Eq. (9) shows that the probability of a particular image feature appearing is proportional to the probability of it appearing in the text-image co-occurrence data associated with each significant text feature if the event category c is given [19]. At the same time, the probability of specific image features is related to the importance of each significant text feature for the target concept [20]. Next, the calculations of $P(f|w, c, S)$ and $P(w|c, D)$ are elaborated.

Firstly, the text feature distribution $P(w|c, D)$ is computed for each event category concept $c \in C$, and the most significant event text feature set $W(c)$ is calculated, and n is the operation coefficient. The LDA model is used to model the event text set, and Laplace smoothing is used to solve the sparse problem of text subject features.

$$P(w|c, D) = [1 + n(w, c, D)] / [|W| + n + (c, D)] \quad (10)$$

$$n(w, c, D) = \sum_{d \in D} n(w|d) P(c|d) \quad (11)$$

$$n(c, D) = \sum_{d \in D} n(n|d) P(c|d) \quad (12)$$

Then, the image feature conditional distribution $P(f|w, c, S)$ in the text-image co-occurrence dataset is computed, and Laplacian smoothing is still used.

$$P(f|w, c, S) = \quad (13)$$

$$[1 + n(f, w, c, S)] / [|F| + n(w, c, S)]$$

$$n(f, w, c, S) = \sum_{(v,d) \in S} n(f|v)P(w, c|d) \quad (14)$$

$$n(w, c, S) = \sum_{(v,d) \in S} n(v)P(w, c|d) \quad (15)$$

2.2.5 Evaluation Criteria

The goal of the text-to-image feature mapping algorithm is to estimate the feature distribution of image information under event categories [21]. According to the feature independence hypothesis of the BOVW model, image features are regarded as random variables which appear independently. Image feature distribution can be represented as a vector with the same size as the bag of character words.

$$P(f|c) = \langle P_i \geq 0 \rangle_{i=0}^{|F|-1}, \sum_i P_i = 1 \quad (16)$$

Cosine similarity and K-L (Kullback-Leibler) divergence dispersion is used as the performance evaluation scale [22], it is assumed that p probability distribution is the datum distribution, and the other probability distribution q is the approximation of distribution p . The greater the cosine similarity of the two approximations is, the closer the two feature distributions are and the higher the approximation degree is. The formula for cosine similarity is as follows:

$$CS(P, q) = \sum_i P_i q_i / \left(\sqrt{\sum_i P_i^2} + \sqrt{\sum_i q_i^2} \right) \quad (17)$$

K-L dispersion is an asymmetric measure to evaluate the difference between two probability distributions. Its value reflects the approximation of distribution q to distribution p . In the determination of the characteristic distribution of the reference image data, the K-L dispersion is defined as:

$$KL(P||q) = \sum_i P_i \ln \frac{P_i}{q_i} \quad (18)$$

Based on the above methods, 15 categories of video security incidents on the Internet are analyzed as data sets [23], corresponding categories are: E1: Sanlu milk powder incidents; E2: red-cored duck egg incidents; E3: Turbot incidents; E4: Jinhao tea oil incidents; E5: Maile chicken incidents; E6: Plasticizer incidents; E7: Clenbuterol incidents; E8: paraffin wax in hot pot incidents; E9: Gutter oil event; E10: Crayfish event; E11: Fushou snail incidents; E12: Poisonous steamed bread incidents; E13: Maggot citrus incidents; E14: Bursting watermelon incidents; E15: Poisonous

bird's nest incidents. Depending on the duration of the incidents [24], the number of related text downloads ranged from 800 to 2000, with text-image accompanying text accounting for about one-third to one-second. A text-image accompanying sample is regarded as a co-occurrence data instance, but in the case of multiple images in a sample, it is considered that one image corresponds to the same accompanying text, and the number of co-occurrence data instances is calculated according to the number of images. Based on artificially collecting, the image data of each food safety event from the Internet search engine and related web pages are searched. For each event, 200~400 images are collected separately. The BOVW model is used to represent each image in a bag of visual word, and the histogram vector expression of each image is obtained.

Firstly, the feature distribution of the reference image is constructed. Using all the images under each event category c , an image feature distribution is obtained as the base feature distribution by the Naive Bayesian classifier. Theoretically, the Naive Bayesian classifier can calculate the real image feature distribution under the target category when the training data is sufficient. The two intuitive methods are compared with the text-image feature mapping algorithm. The first method is the uniform distribution algorithm, assuming that each image feature appears randomly under the concept of each event target with the same probability. The second method is the tagged query algorithm, which uses the name of category c as the query keyword, searches in the Internet search engine [25], and uses the returned K image to train the Naive Bayesian model to get the image feature distribution. The K value of the experiment is 50 based on experience [26–31].

3 Results

The K value of this experiment is 50 according to experience. The comparison between the three algorithms under cosine similarity is shown in Figure 2, 3, and 4.

Analyzing the results of the three algorithms in the cosine similarity, we can see that the maximum value of the uniform distribution algorithm is 0.94%, the minimum value is 0.74%, the maximum value of the label query algorithm is 0.97%, the minimum value is 0.76%, and the maximum value of the proposed algorithm is 0.99% and the minimum value is 0.76%.

Through data comparison, the cosine similarity of the proposed algorithm is always higher than that of the uniform distribution algorithm and the label query algorithm.

Figure 2: The effect of uniform distribution algorithm on estimation of distribution under cosine similarity

Figure 3: Estimation of distribution effect under tag cosine similarity algorithm

The larger the similarity value is, the more accurate the approximate feature extraction is. Therefore, the accuracy of the proposed algorithm is higher than the other two algorithms.

As shown in Figure 5, 6, and 7, the prediction results of the three algorithms under K-L dispersion are compared:

Analysis of Figure 5, 6, and 7 shows that the results of the three algorithms are comparable under K-L discrete values. The maximum value of the uniform distribution algorithm is 0.27%, the minimum value is 0.05%, the maximum value of the tagged query algorithm is 0.30%, the minimum value is 0.04%, and the maximum value of the

Figure 4: Algorithm for estimating distribution effect under cosine similarity

Figure 5: Uniform distribution algorithm under K-L discrete value to estimate distribution effect diagram

proposed algorithm is 0.17% and the minimum value is 0.03%.

Through data comparison, the K-L dispersion of the proposed algorithm is always lower than that of the uniform distribution algorithm and the label query algorithm. Discreteness is an asymmetric metric measure to evaluate the difference of two probability distributions. Its value reflects the approximations. The smaller the dispersion is, the smaller the difference is. Therefore, the difference of the algorithm in this paper is lower than that of the other two algorithms.

From the comparison of the effects of different algorithms on estimating the distribution under the above different metrics, it can be seen that the image feature distribution generated by the text-image feature mapping algorithm is the closest to its benchmark distribution under most event categories, while the uniform distribution algorithm is only close to the results of other algorithms under

Figure 6: K-L discrete valued mark-up query algorithm to estimate the distribution effect map

Figure 7: The algorithm is used to estimate the distribution effect under K-L discrete values

one category (E6). By checking the data under this category, it is found that this is due to the large differences in the image data between them. The label query algorithm under three categories (E1, E9, E11) is equivalent to the text-image feature mapping algorithm proposed by the author. By checking the data, for these event categories, the event category name is directly input from the search engine as the query keyword, and the resulting images are closely related to the event category, so the similar distribution effect of the label query algorithm is better.

In addition to the above direct method, the approximation degree of a similar image feature distribution to the benchmark distribution can be measured from different training data scales. Each time, from each category of the collected event image data set, N images are randomly selected, and the Naive Bayesian model is trained for 100

times. The feature distribution and reference distribution of each image are compared. Finally, the results of all repeated rounds are arithmetically averaged. The number of images randomly selected for each event category is 20, 40, 60, 80, 100, 120, 140, 160 in turn. The approximate results of uniform distribution algorithm, label query algorithm and feature mapping algorithm under each category are averaged, and then compared with the above method. Figures 8 and 9 show the average difference between the image feature distribution and the reference distribution obtained by these approximate methods under the two measurement scales.

Figure 8: Comparison of different algorithms for estimating distribution under cosine similarity

As can be seen from Figures 8 and 9, the approximate results of the uniform distribution algorithm, the label query algorithm and the feature mapping algorithm under each category are averaged, and then compared with the above method. The text-image feature mapping algorithm is similar to the feature distribution obtained by training 100 labeled images, and the cosine similarity of the proposed algorithm is 92%; that of the uniform distribution algorithm is 76%; and the label query algorithm is 84%. The average value of cosine similarity of the proposed algorithm is the largest in every category. The discrete degree of the proposed algorithm is 0.06%; that of the uniform distribution algorithm is 0.17%; the label query algorithm is 0.09%; the average value of the discrete results of the proposed algorithm is the smallest in each category. The above data show that the proposed text-image feature mapping algorithm based on transfer learning can effectively learn the image feature distribution under the target

Figure 9: Comparison of different algorithms for estimating distribution under K-L discrete values

event category from the text data of related events and text-image co-occurrence data.

Under the 100 events, the similarity distribution of text-image data is simulated. The proposed algorithm is simulated with the uniform distribution algorithm and label query algorithm. The average optimal fitness and average operation time of the two algorithms are obtained. The detailed results are described in Table 1. From the analy-

Table 1: Simulation results of approximate distribution of image data distribution under 100 events

Type of algorithm	Average optimum fitness /%	Mean operation time /s
Uniform distribution algorithm	7.51	54.09
Tagged query algorithm	8.22	53.69
Algorithm in this paper	9.85	34.72

sis of Table 1, we can see that the optimal fitness of the proposed algorithm is 9.85%, the uniform distribution algorithm is 7.51%, the label query algorithm is 8.22%, and the fitness of the proposed algorithm is the highest. In the comparison of the average operation time, the proposed algorithm takes 34.72 seconds, the uniform distribution algorithm takes 54.09s, and the label query algorithm takes 53.69s, indicating that the proposed algorithm takes the shortest time and has the highest efficiency. This algorithm can quickly and effectively extract the approximate feature distribution of text-image data under 100 events. This al-

gorithm not only extracts the approximate feature distribution of text-image data effectively, but also consumes less time and has high efficiency.

4 Discussion

In the traditional machine learning framework, the task of learning is to learn a classification model based on given sufficient training data, and then use this learning model to classify and predict test documents. However, we see that machine learning algorithms have a key problem in the current Internet mining research: a large amount of training data in some emerging areas is difficult to obtain. It can be seen that the development of internet applications is very fast. A large number of new areas are emerging, from traditional news, to web pages, pictures, blogs, podcasts and so on. Traditional machine learning needs to calibrate a large amount of training data in each field, which will consume manpower and material resources. Without a large amount of annotated data, a lot of learning related researches and applications can't be carried out. Secondly, traditional machine learning assumes that training data and test data obey the same data distribution. However, in many cases, the same distribution hypothesis is not satisfied. In addition, training data is often out of date. This often requires re-labelling a large volume of training data to meet our training needs, but labeling new data is expensive and requires manpower and material resources. On the other hand, if we have a lot of training data with different distributions, it would be wasteful to discard them completely. How to make rational use of this data is the aim of transfer learning.

Main problems are solved. Transfer learning can transfer knowledge from existing data to help future learning. The goal of transfer learning is to use the knowledge learned from one environment to assist learning tasks in the new environment. Therefore, transfer learning will not assume the same distribution assumption as traditional machine learning. At present, the work on transfer learning can be divided into two parts: case-based transfer learning in isomorphic space and feature-based transfer learning in isomorphic space. It is pointed out that case-based transfer learning has stronger knowledge transfer ability, while feature-based transfer learning has wider knowledge transfer ability. These two methods have their own merits. Transfer learning is a relatively new research direction in machine learning. The current research mainly focuses on data mining, natural language processing, information retrieval and image classification. Ma-

chine learning has provided extensive research findings and results, but research into transfer learning is minimal. Features and samples are two important aspects of text categorization. It is important to consider these two factors comprehensively. Sample based transfer learning is another method to solve the problem of transfer learning. Traditional methods also use feature-based or sample-based transfer learning methods, but there is a lack of comprehensive use of these two methods. The algorithm proposed in this paper can find the data very similar to the test data from the existing data and improve the accuracy of the model.

5 Conclusions

In this paper, a text-image feature mapping algorithm based on transfer learning is proposed. Firstly, clustering technology is used to filter the existing data and find the data which is similar to the target data to help the learning of the target task and improve the performance of the classifier. Then, the event text data is modeled by the potential Dirichlet assignment method, and the most prominent text features are selected by calculating the information gain of the topic features; the event images are modeled using the visual word bag model and the naive Bayesian method; The approximate extraction of the image feature distribution is realized by the text data feature distribution and the text-image co-occurrence data feature distribution under the same event. Compared with the traditional uniform distribution algorithm and labeled query algorithm, the average cosine similarity of the proposed algorithm is 92%, that of the uniform distribution algorithm is 76%, and the labeled query algorithm is 84%. The average dispersion of the proposed algorithm is 0.06%, that of the uniform distribution algorithm is 0.17%, and the labeled query algorithm is 0.09%. The experimental data shows that the proposed algorithm has the advantage of high cosine similarity and low dispersion.

References

- [1] Wang F., Youh J., Fux Y., Auto-Adaptive Well-Distributed Scale-Invariant Feature for SAR Images Registration, *Geomat. Inform. Sci. Wuhan Univ.*, 2015, 40(2), 159-163.
- [2] Wang K., Shil Z., Design and Implementation of Fast Connected Component Labeling Algorithm based on FPGA, *Comp. Eng. Appl.*, 2016, 52(18), 192-198.
- [3] Lozoya R.C., Berte B., Cochet H., Model-based Feature Augmentation for Cardiac Ablation Target Learning from Images, *IEEE Trans. Biomed. Eng.* PP, 2018, (99), 1-1.
- [4] Cazade P.A., Zheng W., Pradagracia D., A Comparative Analysis of Clustering Algorithms: O2 Migration in Truncated Hemoglobin I from Transition Networks, *J. Chem. Phys.*, 2015, 142(2), 025-103.
- [5] Wan S., Niu Z., A Learner Oriented Learning Recommendation Approach based on Mixed Concept Mapping and Immune Algorithm, *Knowledge-Based Syst.*, 2016, 103(C), 28-40.
- [6] Han X.H., Xiong X., Duan F., A New Method for Image Segmentation based on BP Neural Network and Gravitational Search Algorithm Enhanced by Cat Chaotic Mapping, *Appl. Intel.*, 2015, 43(4), 855-873.
- [7] Zhou T., Hu W., Ning J., An Efficient Local Operator-based Qcompensated Reverse Time Migration Algorithm with Multistage Optimization, *Geophys.*, 2018, 83(3), S249-S259.
- [8] Gorodnitskiy E., Perel M., Geng Y., Depth Migration with Gaussian Wave Packets based on Poincaré Wavelets, *Geophys. J. Int.*, 2016, 205(1), 301-318.
- [9] Rastogi R., Srivastava A., Khonde K., An Efficient Parallel Algorithm: Poststack and Prestack Kirchhoff 3D Depth Migration Using Flexi-depth Iterations, *Comp. Geosci.*, 2015, 80, 1-8.
- [10] Tosun S., Ozturk O., Ozkan E., Application Mapping Algorithms for Mesh-based Network-on-chip Architectures, *J. Supercomp.*, 2015, 71(3), 995-1017.
- [11] Kalantar B., Mansor S.B., Sameen M.I., Drone-based Land-cover Mapping Using a Fuzzy Unordered Rule Induction Algorithm Integrated into Object-based Image Analysis, *Int. J. Remote Sens.*, 2017, 38(8-10), 2535-2556.
- [12] Mackenzie C., Pichara K., Protopapas P., Clustering Based Feature Learning on Variable Stars, *Astrophys. J.*, 2016, 820(2), 138.
- [13] Li H., Zhu G., Cui C., Energy-efficient Migration and Consolidation Algorithm of Virtual Machines in Data Centers for Cloud Computing, *Comput.*, 2016, 98(3), 303-317.
- [14] Xiang T., Yan L., Gao R., A Fusion Algorithm for Infrared and Visible Images based on Adaptive Dual-channel Unit-linking PCNN in NSCT Domain, *Infrared Phys. Technol.*, 2015, 69, 53-61.
- [15] Dong J., Xiao X., Menarguez M.A., Mapping Paddy Rice Planting Area in Northeastern Asia with Landsat 8 Images, Phenology based Algorithm and Google Earth Engine, *Remote Sens. Envir.*, 2016, 185, 142-154.
- [16] Li Q., Zhou H., Zhang Q., Efficient Reverse Time Migration based on Fractional Laplacian Viscoacoustic Wave Equation, *Geophys. J. Int.*, 2016, 204(1), 488-504.
- [17] Medrano E.A., Wiel B.J.H.V.D., Uittenbogaard R.E., Simulations of the Diurnal Migration of *Microcystis Aeruginosa*, based on a Scaling Model for Physical-biological Interactions, *Ecolog. Mod.*, 2016, 337, 200-210.
- [18] Matsubayashi A., Asymptotically Optimal Online Page Migration on Three Points, *Algorithmica*, 2015, 71(4), 1035-1064.
- [19] Yap W. S., Phan C.W., Yau W.C., Cryptanalysis of a New Image Alternate Encryption Algorithm based on Chaotic map, *Nonlin. Dyn.*, 2015, 80(3), 1483-1491.
- [20] Rastogi R., Londhe A., Srivastava A., 3D Kirchhoff Depth Migration Algorithm, *Comp. Geosci.*, 2017, 100(C), 67-75.
- [21] Thierry P., Lambaré G., Podvin P., 3-D Preserved Amplitude Prestack Depth Migration on a Workstation, *Geophys.*, 2015, 64(1), 222-229.
- [22] Zheng X.W., Lu D.J., Wang X.G.A., Cooperative Coevolutionary Biogeography-based optimizer, *Appl. Intel.*, 2015, 43(1), 1-17.

- [23] Wang M., Study on Operation Reliability of Transfer System of Urban Transportation Hub based on Reliability Theory, *Automat. Instrument.*, 2016, (1), 418-534.
- [24] Cong S., Gao M.Y., Cao G., Ultrafast Manipulation of a Double Quantum-Dot Charge Qubit Using Lyapunov-Based Control Method. *IEEE J. Quant. Electr.*, 2015, 51(8), 1-8.
- [25] Yan X., Yang S., Hong H.E., Load Adaptive Control Based on Frequency Bifurcation Boundary for Wireless Power Transfer System, *J. Pow. Supp.*, 2017, 43(4), 1025-1084.
- [26] Lokesha V., Deepika T., Ranjini P.S., Cangul I.N. Operations of Nanostructures Via Sdd, Abc4 and Ga5 Indices. *Applied Mathematics and Nonlinear Sciences*, 2017, 2(1), 173-180.
- [27] Molinos-Senante M., Guzman C., Benchmarking Energy Efficiency in Drinking Water Treatment Plants: Quantification of Potential Savings, *J. Clean. Prod.*, 2018, 176, 417-425.
- [28] Gao W., Farahani M.R., Aslam A., Hosamani S. Distance Learning Techniques for Ontology Similarity Measuring and Ontology Mapping, *Cluster Computing - The J. Net. Soft. Tools Appl.*, 2017, 20(2SI), 959-968.
- [29] Ge S.B., Ma J.J., Jiang S.C., Liu Z., Peng W.X., Potential Use of Different Kinds of Carbon in Production of Decayed Wood Plastic Composite, *Arabian J. Chem.*, 2018, 11(6), 838-843.
- [30] Singh K., Gupta N., Dhingra M., Effect of Temperature Regimes, Seed Priming and Priming Duration On Germination and Seedling Growth On American Cotton, *J. Envir. Biol.*, 2018, 39(1), 83-91.
- [31] Hosamani S.M., Correlation of Domination Parameters with Physicochemical Properties of Octane Isomers, *Appl. Math. Nonlin. Sci.*, 2016, 1(2), 345-352.