A new method of location detection with knowledge transfer learning

YuTing Liu, YinJie Ni
Jiangnan Institute of computing technology wuxi 214000 china
spylyt@qq.com

Abstract. This paper propose a new method of location detection with knowledge transfer learning(KTL-LD). Processing Framework was developed and compared with existed experimental results. In this Framework, the label extraction model is used to extract the Weibo user tags, using the existing word embedding model as the source domain, and applying the knowledge transfer learning model to transfer the knowledge to the target domain, which obtain the target domain vector, then the classification model is used to train and predict the user’s location. This method does not require manual tagging data, the training time of this model is effectively reduced, and the ability of generalization is improved as well.

1. Introduction
In the framework of traditional machine learning, the machine learning task is to learn a classification or clustering model based on adequate training data, and then use the model to predict the test data. However, for the data in social media, training sample tags are scarce. As the scale of social networks continues to expand, data labeling for training will cost a lot of effort and resources. Without a large amount of tagging data, many studies and applications related to learning can not be carried out. In addition, traditional machine learning assumes that training data and test data must be in the same feature space and have the same distribution. However, in many real-world applications, this assumption may not satisfied in many cases. Such as the difference between the training data online and the online prediction data, it is often necessary to labeling a large number of training data to meet the training needs. From another point of view, if you have a large amount of training data under different distributions, it is wasteful to discard all of these data completely. So, in order to solve these problems, this paper proposes a text knowledge transfer model based on transfer vector model can use news training data to micro-blog in the field of geographic position detection, the experiments show that this method can greatly reduce the text labeling effort, and has a good prediction effect.

2. Investigation
In recent years, transfer learning algorithms and related theories have attracted much attention. transfer learning is a new machine learning method that uses existing knowledge to solve different but related problems. It relaxes the two basic assumptions of traditional machine learning conditions. According to whether the labeled source domain and target domain, transfer learning is divided into 3 categories: [1] target in the field of a small amount of labeled samples of inductive transfer learning [2], only the source in the field of transductive transfer learning [3] and the source domain and target domain are not labeled samples of unsupervised transfer learning [4][5] labeling samples. There are a large number of labeled samples in the source domain, and a small number of labeled samples in the target domain, the typical model has literature [6][7][8][9][10]. Harel and Mannor[6] using the spectral
method and combining the information of the source data and target data to establish the correspondence between the source domain data again said, then train the classifier to classify the targets in the field of data. Wang et al [7] based on class information and feature information established two similarity matrix and a similarity matrix, and then use the mapping relation between the spectral decomposition method to learn the source domain and target domain to a new low dimensional space. The new space maintains the original class and feature information of the data points. After all data points are represented by a new representation, traditional machine learning methods can be implemented to accomplish clustering tasks. Duan et al [8] use support vector machine theory, combination of learning information, the source domain and target domain in the data are mapped to a new low dimensional space, then, combined with the characteristics of space to form a new feature representation.

Weibo data is usually not standardized, colloquial, including typos and new words. This is a great difficulty for the extraction of the characteristic words, so the model for the analysis of these text information needs to be carefully studied and designed. At present, the research hotspot direction is some unexpected events, such as Zhao Hua and others based on Sina microblogging data research H7N9 infectious disease topic model [11], Ting Hua et al. Based on Twitter data analysis Virginia shooting incident [12]. These studies learn the results of the user's label based on transfer learning theory. Ting et al. proposed a semi-supervised target event detection (STED) method in analyzing target topics such as disease outbreaks, criminal events, or mass emotions [13]. By creating a label extraction model, the news label was migrated to Twitter to produce an initial label Data, the label propagation model generates the feature extension label data, the graph partition model divides a tweets into multiple subsets, and then uses the SVM classification model to identify the tweets associated with the target theme.

In this paper, based on the characteristics of Sina Weibo data and improving the traditional training set construction method, this paper proposes a geographic location detection algorithm based on knowledge migration, extracts user tags through word segmentation model and label production model, The News datasets is used as the source domain model to construct the Weibo data eigenvector by transfer learning, and then the classification model is used to train and forecast. This method is mainly focus on:
1. Construct transfer learning model.
2. Extending user eigenvector through tf*idf model.
3. Using stacking as model ensembling technique.

3. Framework and method
As shown in Figure 1, the structure of KTL-LD can be divided into three parts: the word tokenizer and tag extraction model, TF * IDF model is used to extract the label and produce the word frequency matrix; By introducing the news word embedding model as the knowledge base, get Weibo text feature vector through knowledge transfer model, combined with the word frequency matrix as other features to get training set and test set; Finally, stacked model is used for training and test.
3.1. Label extraction model

The label extraction model is composed of word extraction models, which are used to extract words that have meaning in microblogging. These words can reflect the connection between users and extend the user tag set.

The commonly used word weight extraction model is the TF * IDF model, where TF represents the number of occurrences of words in a piece of text, representing the frequency of words, DF is the number of times the word appears in different texts, representing the frequency of the document. IDF is the reciprocal of the document frequency, then TF * IDF is used to measure the importance of a word in the text.

Obviously, the weight value of a word in the TF * IDF model is proportional to the word frequency of the word, inversely proportional to the frequency of the document. There are obvious shortcomings, TF * IDF will be a collection of text as a whole, regardless of the words in different users in the distribution of microblogging. So, the weight of some words is high but no sense, some words and tasks related to the weight is very low.

This short message text of Weibo, social contact is reflected in thumbs up, comment, forwarding, or symbols such as @ (mentioned someone), usually with the same words microblogging it is likely to be talking about the same topic, and it is also possible in the same area, so these connection words in the text classification has a very important role.

Through the TF * IDF model, microblogging text is expressed as a binary feature of the eigenvector:

\[
d = \{ (t_1, w_1), (t_2, w_2), \ldots, (t_n, w_n) \}
\]

(1)

Where \( n \) represent the number of labels, \( t_i \) is the characteristic term, \( w_i \) is the weight corresponding to \( t_i \), that is, \( t_i \) * idf value, \( 1 \leq i \leq n \). The feature space is constant, the formula can be expressed as:

\[
d = \{ w_1, w_2, \ldots, w_n \}
\]

(2)

So as to get the word frequency matrix \( A \) of the Weibo datasets, the matrix element \( w_{ij} \) denote the word frequency of the word \( j \) in the user’s text \( i \). Word frequency matrix can reflect whether different users are concerned about the same theme, the use of word frequency matrix to expand the user characteristics to achieve good results in this paper.

3.2. Knowledge transfer learning model

Domain Definition: Field \( D \) consists of feature space \( X \) and edge probability distribution \( P(X) \), where \( X = \{ x_1, \ldots, x_n \} \in X \). \( X \) is the space of all word vectors, \( x_i \) is the \( i \)-th word vector, corresponding to some text, and \( X \) is a specific learning sample. In general, if the two domains are different, they may have different feature spaces or different edge probability distributions.

For a given field \( D \), a classification task \( T \) consists of the category space \( Y \) and the target prediction function \( f(\cdot) \). The target prediction function can not be obtained by observing, and it can be obtained the learning process. The training data is composed of \( \{ x_i, y_i \} \), where \( x_i \in X \) and \( y_i \in Y \).

Transfer Learning Definition: Given the source domain \( D_S \) and the learning task \( T_S \), the target field \( D_T \) and the learning task \( D_T \), the migration learning is intended to use the knowledge in \( D_S \) and \( T_S \) to help improve the learning of the prediction function \( f_T(\cdot) \) in the target domain \( D_T \), \( D_S \neq D_T \) or \( T_S \neq T_T \).

Our task is to get the user's location information through the microblogging feature vector. Considering that the current classification of geography information is not common, we may turn this problem into how to get the word vector from a more comprehensive knowledge base, The word vector can reflect the semantic information of the words in different sentences and paragraphs, so the source domain and the target domain can not need to label the information, so that the migration learning task can be defined as unsupervised migration learning.

We select the news datasets as the source domain \( D_S \), the amount of news text data is much larger than the amount of microblogging text data, the sentence of news text is more standardized, rarely
appear wrong words, the vector of words learned from $T_S$ can reflect the relationship between words. The implementation of learning task $T_S$ can refer to [14].

The Weibo text as the target domain $D_T$, through the learning task $T_T$ get the word vector, define the transfer model as:

$$x_{Ti} = \|x_{Si}\|_1$$  \hfill (3)

$$a_{Tk} = \frac{1}{N} \sum \beta_i \|x_{Ti}\|_1$$  \hfill (4)

Among them, $x_{Si} \in X$ is the source domain word vector, $x_{Ti} \in X$ is the target domain word vector, and $\beta_i$ is the weight adjustment coefficient. $a_{Tk}$ is the weighted average vector of the target domain user $k$, and the number of microblogging user tags is $N$.

3.3. Machine learning model based on stacking

Stacking (also called meta ensembling) is a model ensembling technique used to combine information from multiple predictive models to generate a new model. Often times the stacked model (also called 2nd-level model) will outperform each of the individual models due its smoothing nature and ability to highlight each base model where it performs best and discredit each base model where it performs poorly. For this reason, stacking is most effective when the base models are significantly different. Here I provide a simple example and guide on how stacking is most often implemented in practice.

Training set $Tr = \{(U_1, y_1), (U_2, y_2), ..., (U_n, y_n)\}$, Where $U_i$ is the eigenvector of the i-th user, $y_i$ is the geographic value of the i-th user.

$Y_i \in \{‘South China’, ‘North Chine’, ‘Central China’, ‘Northwest China’, ‘Southwest China’, ‘Southeast China’, ‘Northeast China’, ‘Overseas’\}$

The data set is divided into five subsets, the training set is $Tr_i \subset Tr$, the validation set is $Valid_i \subset Tr$, $1 \leq i \leq 5$, the data cross is divided as shown in Figure 2.

![Figure 2 Data partitioning, each classifier trained and validated by a partition](image)

Through the training step, we get the basic classification model set $C = \{C_1, C_2, C_3, C_4, C_5\}$, and have 5 different classifiers. We get 5 predictive fragments $P_{Valid_i}, 1 \leq i \leq 5$, $1 \leq i \leq 5$, merge into the prediction of the whole training set, and then use the output of these classifiers as input to train the regression model on the whole training set. The whole process shown in Figure 3.
Test set is $T_e = \{(U_1), (U_2), ..., (U_n)\}$, Where $U_i$ is the eigenvector of the i-th user. During data fold step, we use the training set $T_r$ as a training section, replacing the test set of each chunk with the test set $T_e$.

1. Through the training step, we get the basic classification model set $C = \{C_1, C_2, C_3, C_4, C_5\}$, and have 5 different classifiers. We get 5 prediction of $T_e$ from the 5 basic classification models.
2. The regression model predicts five predictive fragments, and get the final result.

4. Experimental data and result analysis

4.1. Experimental data
The experimental data from Sina Weibo, training set contains 3200 Weibo users, 237801 microbloggings, test set contains 1248 Weibo users, 74565 microbloggings.

4.2. Experimental result analysis
After the word segmentation, the general need is to filter the label, remove the punctuation and stop words, because the symbols (@, #) reflect social connection, so the word with these symbols has not been segmented, part of the social contact words are shown in Table 1. Through the label extraction model, 61372 labels were extracted. The location of the user's geographical location is shown in Table 1.

| Social connection words                  |
|-----------------------------------------|
| @ShangHai                              |
| @TianJin                               |
| @ZhejiangTV                            |
| @BeijingTV                             |
| @Hero                                  |
| @Tencent                               |
| @BaiDuCloud                            |
| @YouTuber                              |
| @Beijing                               |
| @TuDou                                 |
| @SinaNEWS                              |
| @iQIYi                                 |
Using the formula (4) to get Weibo user’s text feature vector \( w_i \in \mathbb{R}^{300} \). The training set contains 3200 samples. The test set contains 1248 samples. After training in stacking mode, the training set has an accuracy rate of 30%. Through the label extraction model to further expand the use’s characteristics, Weibo user’s text feature vector up to \( w_i \in \mathbb{R}^{61672} \), the accuracy achieve 41%.

5. Conclusion
In this paper, we propose a method to carry out the knowledge transfer from the existing news domain and obtain the eigenvector of the microbloggings, so as to carry out the method of geo-location detection. The method is characterized by the fact that the microblogging data is rapidly changing and the artificial annotation is time consuming and laborious. The character of our method includes: Using TF * IDF to get the word frequency matrix to extend the microblogging vector; training and forecasting using stacked model; Use Knowledge Transfer Learning to respond to rapid data changes, put forward a better solution.

This method still has some problems in the application process. For example, Weibo users usually make colloquial words and often appear new words and typos. For this kind of problem, we do not propose a reasonable solution. In general, machine learning tasks need specific treatment, so there is still a need for a lot of time to analyze the relationship between data labels and location. In the application of transfer learning theory, the situation of "negative transfer" is very easy to appear. The starting point of knowledge transfer in this paper is that the expression of words is similar in different statements, but there is also the case of polysemy, which is not tested in our method. These problems are the next steps that need to be studied.

References
[1] Zhuang FZ, Ping L, He Q. Research progress of learning transfer [J]. Journal of Software, 2015.26(1):26-39
[2] Wei FM, Zhang JP, Chu Y, Yang J. FSFP: Transfer learning from long texts to the short. Applied Mathematics & Information Sciences, 2014.8(4):2033–2044. [doi: 10.12785/amis/080462]
[3] Dai WY, Xue GR, Yang Q, Yu Y. Co-Clustering based classification for out-of-domain documents. In: Proc. of the 13th ACM Int’l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2007. 210–219. [doi: 10.1145/1281192.1281218]
[4] Dai WY, Yang Q, Xue GR, Yu Y. Self-Taught clustering. In: Proc. of the 24th Int’l Conf. on Machine Learning. San Francisco:Morgan Kaufmann Publishers, 2008. 200–207. [doi: 10.1145/1390156.1390182]
[5] Samanta S, Tirumarai SA, Das S. Cross-Domain clustering performed by transfer of knowledge across domains. In: Proc. of the 2013 IEEE 4th National Conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). 2013. 1–4.[doi: 10.1109/NCVPRIPG.2013.6776213]
[6] Harel M, Mannor S. Learning from multiple outlooks. In: Proc. of the Int’l Conf. on Machine Learning. 2011. 401–408.

[7] Wang C, Mahadevan S. Heterogeneous domain adaptation using manifold alignment. In: Proc. of Int’l Joint Conf. on Artificial Intelligence. 2011. 1541–1546. [doi: 10.5591/978-1-57735-516-8/IJCAI11-259]

[8] Duan LX, Xu D, Tsang IW. Learning with augmented features for heterogeneous domain adaptation. In: Proc. of the Int’l Conf. on Machine Learning. 2012. 711–718.

[9] Li W, Duan LX, Xu D, Tsang IW. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014,36(6):1134–1148. [doi: 10.1109/TPAMI.2013.167]

[10] Zhou JT, Tsang IW, Pan SJ, Tan MK. Heterogeneous domain adaptation for multiple classes. In: Proc. of the Int’l Conf. on Artificial Intelligence and Statistics. 2014. 1095–1103.

[11] Zhao H, Zhang CZ. A contrastive study of topic evolution in Chinese and English emergencies: a case study of H7N9 micro-blog [J]. Intelligence information work, 2016(3)

[12] J. Weng and B. Lee. Event detection in twitter. In ICWSM’11, volume 3, 2011

[13] Hua. T, Chen. F, Zhao. L, et al. STED: Semi-Supervised Targeted-Interest Event Detection in Twitter[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.2013:1466-1469

[14] Goldberg Y, Levy O. word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method[J], Eprint Arxiv, 2014.