Communication Specific Emitter Identification Based on Transfer Learning

Hui Liu¹*, Pengjiang Hu¹ and Zhen Liu²

¹ National University of Defense Science and technology, Hefei, China
² Air Force Early Warning Academy, Wuhan, China

*Corresponding author email: christ592604@163.com

Abstract. In the application of specific communication emitter identification (CSEI), the emitter slight features will change along with different times, places and conditions, which makes training samples and test samples obey different distribution and makes most machine learning algorithms not work well. To solve this problem, this paper introduced transfer learning to CSEI. Compared with traditional machine learning, transfer learning does not relax the assumption that the data is independent and distributed. It can explore useful knowledge from the data which is different from the target domain but similar to the source domain. It can also learn the classifier when the labeled samples in the same domain are insufficient. Extensive experiments show the effectiveness of the new method.

Keywords: Transfer learning; Specific communication emitter identification.

1. Introduction

In the real battlefield environment, all kinds of communication emitter widely use spread spectrum, power control, and other low interception probability technologies to continuously improve their anti-reconnaissance and anti-jamming capabilities, which greatly increases the difficulty of intercepting and receiving the enemy emitter signal[1][2]. It's difficult to get enough training samples which can be used to identify system training and learning. In order to solve this problem, some scholars have proposed communication emitter recognition method based on semi-supervised learning[3]. The semi-supervised learning can make use of the unlabeled data to assist the learning of classification model, thus making up for the insufficient information caused by the insufficient labeled samples. In practical application, we find that semi-supervised learning classifier is not always "safe". Sometimes, using unlabeled data may cause serious interference to learning. This shows that there are some differences between the structure of unlabeled data and that of labeled data, which leads to the decline of learning performance. Transfer learning does not need the assumption of traditional machine learning. It can effectively exploit and transfer the knowledge from an existing similar but not identical labeled source domain data for target learning. It has been widely concerned and studied. With the deepening of research, transfer learning has been widely applied in many fields, such as text processing, computer vision, image processing, indoor localization, automatic control, network recognition, etc[5][6]. In this paper, transfer learning is introduced into this application to make full use of limited data and improve recognition performance of CSEI.
2. Methodology

The data set of the source domain is \( \{ (x^S_1, y^S_1), (x^S_2, y^S_2), \ldots, (x^S_{N^S}, y^S_{N^S}) \} \), where \( x^S_i \) represents the eigenvector of the \( i \)-th sample in the data set, \( y^S_i \) is the corresponding label, and the number of samples in the source domain is \( N^S \). The target domain data set is \( \{ (x^T_1, x^T_2), \ldots, (x^T_{N^T}) \} \), where \( x^T_i \) is the eigenvector of the \( i \)-th sample in the data set, and the number of samples in the target domain is \( N^T \). The set of target domain data and source domain data is expressed as \( X = X_S \cup X_T \), and the total number of samples is \( N = N^S + N^T \). In this paper, we assume that there are a large number of labeled sample data in the source domain, but only unlabeled sample data in the target domain, and the characteristics and corresponding labels of the source domain and the target domain are different from the joint distribution, that is \( P(x, y) \neq P_T(x, y) \).

In practical application, the joint distribution \( P(x, y) \) of data \( x \) and its corresponding label \( y \) is often different between the source domain and the target domain. It is difficult to guarantee the credibility of the model when it is applied to the target domain.

BRSD (Bias Reduction via Structure Discovery) \(^7\) is a classical algorithm of transfer learning. Based on clustering analysis and resampling, a new set of training samples with small deviation from the target domain is generated to improve the reliability of the learning model. This method not only does not need to estimate the domain distribution directly, but also can correct the differences among different types of domains. Figure 25 shows the basic idea of the migration learning framework through an example of two categories. The circle and square in the figure represent two categories of data, the solid point represents the marked training sample, and the hollow point represents the unmarked test sample. In Figure 1(a), it can be seen that the distribution of training samples is different from the distribution of test samples. Using this classification model with biased training samples learning directly has a high error classification rate for test samples. Figure 1(b) - (d) shows the migration learning process based on clustering analysis and resampling: first, through clustering analysis of the whole data set to find the data structure information (Figure 1(b)); then, from each sub class, according to the same proportion, through a certain resampling strategy, a new set of training samples with small deviation from the target domain data distribution is generated for the target domain learning (Figure1(c) The classification performance of the model learned on the new training sample set in the target domain has been greatly improved (Figure1(d)).

![Figure 1. The basic ideas of transfer learning based on cluster analysis and resampling.](image-url)

It can be proved that the distribution of the new training samples selected through Fig.1(c) is the same as that of the total data set: when each clustering sub class contains less data, it can be assumed that the data distribution within the sub class is approximately the same. Under this assumption, the new training samples are selected from each sub class according to the same proportion and equal probability. Define the variable \( t \). If the data \( (x, y) \) is the training sample, then \( t = 1 \); otherwise, \( t = 0 \), then \( P(t=1) \) represents the probability of the training sample (generally, \( P(t=1) = P \)) represents the proportion of
the number of new training samples to the total number of samples in the data set). \( P(t=1|\mathbf{x},y) \) represents the probability that the data \((\mathbf{x},y)\) is the training sample. Through clustering the data set \( \mathbf{X} \), we can get \( N_c \) subclass \( \bigcup_{i=1}^{N_c} C_i = \mathbf{X} \), and select new training samples from each subclass according to the same proportion \( P_i \) equal probability, then we can get \( P(t=1|C_i,y) = P_i = P(t=1) \). It can be obtained:

According to formula (1), \( P(t=1|\mathbf{X},y) = P(t=1,\mathbf{X},y)/P(\mathbf{X},y) = P(t=1) \) means that the selection of training samples has nothing to do with the data distribution, so it can be considered that the distribution of training samples selected from each subclass according to the same proportion and equal probability is consistent and unbiased with the distribution of the total data set.

\[
P(t=1,\mathbf{X},y) = \sum_{i=1}^{N_c} P(t=1,C_i,y) \]
\[
= \sum_{i=1}^{N_c} P(t=1|C_i,y) P(C_i,y) 
= \sum_{i=1}^{N_c} P(t=1) \times P(C_i,y) 
= P(t=1) \times P(\mathbf{X},y) \tag{1}
\]

In theory, any clustering algorithm with good performance can be used, such as BRSD adopts a clustering analysis algorithm based on DBSCAN (density based spatial clustering of application with noise)[8] can find the irregular shape of the sub class without knowing the number of sub classes in advance, so it has better clustering performance. Its main idea is: starting from a certain core point, it gradually includes all the data in the density reachable area, thus forming a new subclass with only the core point and the boundary point and the density between the two points is reachable. The algorithm flow of DBSCAN is shown in Table 5.

**Table 1.** The basic flow of DBSCAN clustering algorithm.

| Input: Dataset \( \mathbf{X} \), Parameters \( \text{MinPts} \) and \( \text{Eps} \). |
|---|
| 1. Identifies all data in \( \mathbf{X} \) as unclassified. |
| 2. Based on \( \text{MinPts} \) and \( \text{Eps} \), an unclassified core data \( \mathbf{x} \) is found in \( \mathbf{X} \). Then add a new cluster sub class, take \( \mathbf{x} \) as the starting point of the cluster sub class, and mark \( \mathbf{x} \) as the classified state. |
| 3. Find all the unclassified data in the \( \text{Eps} \) neighborhood of \( \mathbf{x} \) to form a temporary data set \( \mathbf{X}_{\text{temp}} \). |
| 4. Take out a data \( \mathbf{x}_i \) in \( \mathbf{X}_{\text{temp}} \), add it to the newly added cluster sub class in step 2), mark \( \mathbf{x}_i \) as classified state, and then remove \( \mathbf{x}_i \) from \( \mathbf{X}_{\text{temp}} \). |
| 5. If \( \mathbf{x}_i \) in step 4) is also a core point, all unclassified data in the \( \text{Eps} \) neighborhood of \( \mathbf{x}_i \) is also added to the temporary data set \( \mathbf{X}_{\text{temp}} \). |
| 6. Repeat steps 4) and 5) until the temporary data set \( \mathbf{X}_{\text{temp}} \) is empty. |
| 7. Repeat steps 2) through 6) until dataset \( \mathbf{X} \) has no core point data. If unclassified data still exists in \( \mathbf{X} \), they are identified as noise points. |

**Output:** cluster structure of \( \mathbf{X} \).

It can be seen that how to find the core point in the whole data set is very important to the effect of DBSCAN clustering. Generally, data \( \mathbf{x}_i \) with the number of nearest neighbors greater than the threshold value is defined as the core point, that is, data points satisfying formula (2)
Among them, the nearest neighbor cardinality $Card(x_i)$ is used to measure the size of the set of $x_i$ nearest neighbors. The threshold $MinPts$ can be estimated by the average of the nearest neighbor cardinality of all data points. $M(x_i, x_j)$ represents the nearest neighbor degree of $x_i$ to $x_j$. The definition of proximity $M(x_i, x_j)$ is as follows

$$M(x_i, x_j) = \begin{cases} 
1 & d(x_i, x_j) \leq Eps \\
0 & \text{otherwise}
\end{cases}$$

Where, $d(x_i, x_j)$ represents the distance between data $x_i$ and $x_j$, generally Euclidean distance, and $Eps$ is the radius of the nearest neighbor, which can be estimated approximately by using the marker data. If $x_j$ is in the neighborhood of $x_i$, then $y$, otherwise $M(x_i, x_j) = 0$. The neighborhood ownership degree $M(x_i, x_j)$ defined by formula (3) is equal to 1 or 0 only according to whether the data is in the neighborhood, so it is also called the discrete neighborhood ownership degree, as shown in Figure 2.

![Figure 2. Crisp neighborhood membership degree.](image)

### 3. Experiment and Discussions

In the experiment, we collected 10 sets of data of different frequency, different speakers and different propagation environments. The model and batch of the radio station are the same, the working frequency is 160MHz or 410MHz, and the transmission signal bandwidth is 25KHz. Three different speakers are used to form baseband voice modulation, which is received in the propagation environment of short distance direct wave and long distance direct wave. The channel bandwidth of the receiver is 100KHz, and the sampling frequency is 204.8KHz. After collecting the measured data of radiation source, the features of communication emitter including Envelope Box dimension, Information dimension, Lempel Ziv complexity, High-order R feature, High-order J feature and Hilbert time-frequency energy parameter[1][3].

The transfer learning experiment data is shown in table 2.

| DATA Set | Frequency | Speaker | Receiving distance |
|----------|-----------|---------|--------------------|
| Target   | 160MHz    | Speaker 1 | long distance     |
| Source 1 | 410MHz    | Speaker 1 | long distance     |
| Source 2 | 160MHz    | Speaker 2 | long distance     |
| Source 3 | 160MHz    | Speaker 3 | long distance     |
| Source 4 | 160MHz    | Speaker 1 | close quarter     |

Randomly select 100 samples from each station data of the target domain data set "target" for the target domain test set. For the source domain data set, first select a source domain from "source 1" to "source 4",
then randomly select $N_s = r \times 100$ data from each radio station in the selected source domain as the source domain data for experiment, where $r$ is the ratio of the sample number in the source domain to the sample number in the target domain, and the value range of $r$ in the experiment is $[0.1, 0.3, 0.5, 0.7, 0.9]$. In order to verify that new algorithm can improve the performance of different base classifiers, five base classifiers are used for experiments: C4.5 decision tree and naive Bayes. BRSD are compared with baseline algorithm. Table 3-4 shows the average results of recognition accuracy of 20 repeated independent experiments with different base classifiers and different source domain data. The maximum classification accuracy between different methods is indicated by bold underline font.

Table 3. The classification accuracy with C4.5.

| $r$ | Source 1 | Source 2 | Source 3 | Source 4 |
|-----|----------|----------|----------|----------|
|     | Baseline | BRSD     | Baseline | BRSD     | Baseline | BRSD     |
| 0.1 | 0.4355   | 0.5740   | 0.6048   | 0.7113   | 0.6600   | 0.6895   | 0.4753   | 0.5055   |
| 0.3 | 0.4250   | 0.7980   | 0.6263   | 0.700    | 0.7045   | 0.7788   | 0.5048   | 0.5283   |
| 0.5 | 0.4160   | 0.6685   | 0.5523   | 0.6188   | 0.6838   | 0.7780   | 0.5248   | 0.5250   |
| 0.7 | 0.3035   | 0.5990   | 0.6053   | 0.6063   | 0.6795   | 0.7485   | 0.5000   | 0.5000   |
| 0.9 | 0.4615   | 0.5295   | 0.5470   | 0.6153   | 0.7220   | 0.7785   | 0.5000   | 0.5003   |
| Ave. | 0.4261   | 0.6285   | 0.5901   | 0.6728   | 0.6750   | 0.7407   | 0.5055   | 0.5096   |

Table 4. The classification accuracy with Naive Bayes.

| $r$ | Source 1 | Source 2 | Source 3 | Source 4 |
|-----|----------|----------|----------|----------|
|     | Baseline | BRSD     | Baseline | BRSD     | Baseline | BRSD     |
| 0.1 | 0.4680   | 0.6515   | 0.5468   | 0.7310   | 0.6240   | 0.7565   | 0.4605   | 0.5143   |
| 0.3 | 0.4730   | 0.8050   | 0.5760   | 0.6703   | 0.6343   | 0.7055   | 0.4780   | 0.4795   |
| 0.5 | 0.4185   | 0.6270   | 0.5733   | 0.5955   | 0.6915   | 0.7755   | 0.4718   | 0.4555   |
| 0.7 | 0.4515   | 0.6920   | 0.5753   | 0.6355   | 0.6905   | 0.7250   | 0.4793   | 0.4998   |
| 0.9 | 0.4615   | 0.5225   | 0.5720   | 0.6128   | 0.7675   | 0.7430   | 0.4698   | 0.4730   |
| Ave. | 0.4588   | 0.6487   | 0.5807   | 0.6748   | 0.6712   | 0.733    | 0.4768   | 0.4913   |

From table 3 to table 4, it can be seen that BRSD can effectively improve the recognition rate in most cases, which can make full use of limited data. It also proves the effectiveness of transfer learning in the application of speaker emitter recognition. In addition, it can be found that the recognition rate based on frequency change and speaker change (source 1, source 2 and source 3) is higher than that based on propagation environment change (source 4). That may be due to the large difference between the characteristics of signals received in short distance and those received in long distance resulting in less common knowledge available for transferring.

4. Conclusion

This paper introduced transfer learning to CSEI. Transfer learning does not relax the assumption that the data is independent and distributed. It can mine useful knowledge from the data. With the better explored data structure information, the new method can transfer more useful knowledge from source domain to target domain. The experiments result proved the new method had better performance.

Acknowledgement

Both authors would like to acknowledge the support of Independent Scientific Research Program of National University of Defense Science and Technology(NO.ZK18-03-14) and Anhui Provincial Natural Science Foundation (NO.1908085MF202).
References

[1] Y M Chen, C M Lin, C S Hsueh. Emitter Identification of Electronic Intelligence System using Type-2 Fuzzy Classifier. Systems Science & Control Engineering, 2014, 2(1): 389-397.

[2] L Y Bing, G Juan, L Yun, et al. Radar Emitter Signal Recognition based on Multi-scale Wavelet Entropy and Feature Weighting. Journal of Central South University of Technology, 2014, 21(11): 4254-4260.

[3] Chun Zhang, Shafei Wang, Dongsheng Li. Semi-supervised behavioral learning and its application. Optik 127 (2016) 376–382.

[4] Sun S L, Shi H L, Wu Y B. A Survey of Multi-Source Domain Adaptation. Information Fusion, 2014, 24: 84-92.

[5] Shao L, Zhu F, Li X L. Transfer Learning for Visual Categorization: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(5): 1019-1034.

[6] Cheng B, Liu M, Suk H I, et al. Multimodal Manifold-Regularized Transfer Learning for MCI Conversion Prediction. Brain Imaging and Behavior, 2015, 9(4): 1-14.

[7] J Ren, X Shi, W Fan, et al. Type-independent Correction of Sample Selection Bias via Structural Discovery and Re-balancing. Proceedings of the 8th SIAM International Conference on Data Mining, Auckland: ACM, 2008: 565-576.

[8] Gözde Ulutagay, Efendi Nasibov. Fuzzy and Crisp Clustering Methods based on The Neighborhood Concept: A comprehensive Review. Journal of Intelligent & Fuzzy Systems, 2012, 23(6): 271-281.