The Identification of Secular Variation in IoT Based on Transfer Learning

Caidan Zhao[1], Zhibiao Cai[1], Minmin Huang[1], Mingxian Shi[1], Xiaojiang Du[2], Mohsen Guizani[3]

[1]Dept. of Communication Engineering, Xiamen University, Xiamen 361005, China.

[2]Dept. of Computer and Information Science, Temple University, Philadelphia, PA 19122 USA.

[3]Dept. of Electrical and Computer Engineering, University of ldaho, Mosocow, ldaho ID 83844, USA.

Email: zcd@xmu.edu.cn, {1587523315, smxfu, 920893160}@qq.com, dux@temple.edu, mguizani@gmail.com.

***Abstract*—In the Internet of Things(IoT) equipment, the characteristic space of the physical layer has changed slightly due to prolongation of the use time and the change of the environment, which may result to the terrible identification of the new target. To solve the problem, this paper uses transfer learning to update the instance weights and combines the weight with rejection sampling to construct the training set. This method provides a black box for transfer learning and a possibility for building multi-classification transfer learning. Some experimental results show that the rate can increase 10% when the number of target samples is too small to train a new learning model.**

***Index Terms*—Internet of Things, transfer learning, rejection sampling, multi-classification**

I. Introduction

With the development of communication technology, the security of wireless equipment becomes more and more important [1], and there are many researchers studying related subjects [2]. Each transmitter inevitably has its own unique feature which mainly produced by analog devices in physical layer, such as crystal, filter and amplifier, so the attacker can be correctly found with the detection in physical layer, rather than the fraud detection in application layer [3]. However, the expansion of wireless equipment usage time will caused changes in physical layer. In addition, the location and environment of wireless equipment also may cause changes in the received signals [4]. This paper proposes a novel algorithm based on transfer learning to solve the above two problems. We can transfer the knowledge of the old samples to raise the identification rate of the new samples, because there are tiny difference in features between old samples and new samples.

Zhao *et al.* [5] and Carbino *et al.* [6] both discussed the identification of wireless equipment. They extracted features from physical layer and achieved good results on many e-quipments. However, due to the complexity of the wireless channel, and aging of the equipment [7], [8], features of the new samples are different from the old, and the model trained by the old samples has worse identification results of the new samples. So the classifier needs to learn new knowledge by itself because of the small number of new samples. We found that the characteristics of transfer learning can meet the requirements of the practical signals. There are further refinements of transfer learning based on instances or features [9], [10], Pan [11] also studied the problem of homogeneous and heterogeneous transfer learning [12], [13]. Transfer learning provides a new method to transfer knowledge, but its own problem of negative transfer is hard to process. Chattopadhyay *et al.* [14] proposed multiple sources transfer learning based on multi-classification to reduce the influence. To solve this problem, most of the base learners default to the instance of weights, such as libsvm [15] and svmlight [16]. The re-sampling method is used to incorporate importance weights into AdaBoost [17]. And Zadrozny *et al.* [18] proposed a method converting from cost-sensitive learners to rejection sampling.

Contributions of this paper are presented as follows:

- The aging of IoT equipments will lead to changes of the feature. So this experiment adopts transfer learning method to improve precision.
- Most of articles focus on transfer algorithm of knowledge rather than the universal and can not feed weights to classification algorithm. This paper proposes a method that transfer learning based on rejection sampling and multiple sources.
- The analysis of transfer learning is always binary classification. This paper provides a new thought for multi-classification transfer learning.

The reminder of this paper is organized as follows: Section II introduces the framework of the whole system. Section III discusses the algorithm of transfer learning and combines the rejection sampling with transfer learning. Section IV applies transfer learning based on rejection sampling to wireless

This work is supported by the Nature Science Foundation of China No.91638204.

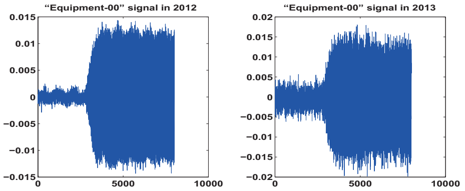

Fig. 1. The "Equipment-00" signal in 2012 and 2013

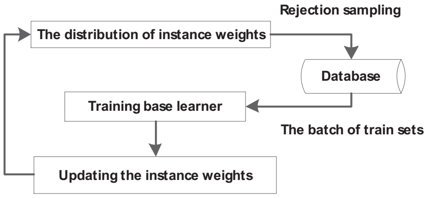

Fig. 3. Rejection sampling used for big database

signals. The conclusion of the present problem and future work are given in Section V.

II. Identification Framework

The identification of wireless signal provides a new solution for the security of IoT equipment. However, the aging of wireless equipments and environmental changes may affect the stability of features in physical layer, and cause a small offset. This shift will sometimes exceed the classification of the edge of the network, and leads to rapid decline in the recognition rate, which result in poor performance of identification.

Fig. 1 is the "Equipment-00" signal in 2012 and 2013, which shows that the signal envelope has changed slightly in complex environment. In addition to different signal to noise ratio, the envelope of the two signals are slightly different, because of the change of environmental, the device amplifier, filter and other physical layer hardware aging. In this case, the recognition rate drops from 90% to about 38% when the model trained by old samples. This paper uses transfer learning to provide a new model for this problem. The machine learning framework based on transfer learning is shown in Fig. 2.

T_b is target domain of training data, and which feature distribution is different from the target instance (a large number of samples). T_a is auxiliary training data, it has a different feature distribution with the target instance (a large number of samples). Transfer learning adapts AdaBoost to increase the weights of samples and the Hedge algorithm [19] to reduce the weights of samples.

The feature distribution is different between 2012 and 2013, but they are very similar. Thus, setting the data in 2012 as the auxiliary data sources X_a and data in 2013 as the target data T_b. Then the knowledge of X_a is transferred to help T_b train a new learning model, the new model can get a better recognition rate, though the number of T_b is very small.

Transfer learning is always based on binary classification, which is called "One vs. One (OvO)" or "One vs. Rest (OvR)", so it is completely unsuitable for multi-classification transfer learning. After transfer learning, the marginal distribution of classifier has changed, which is not suitable for the class that without transfer learning. What's more, to avoid over-fitting, it could takes the other class as additional auxiliary data sources, and training the binary classify model by transfer learning each time. After transfer learning, the instance weights of X_a and T_b is saved to the database. Then all classes are sampled by instance weights to construct a new training set. Finally, the training set is used for training a model of multi-classification.

III. Key Technology

A. Rejection sampling by instance weights

Most classifiers are not cost-sensitive classification, except for libsvm [15] and svmlight [16]. The svmlight allows users to input weight to train model, the libsvm is similar to svmlight and has many documents. But the value of instance weights for libsvm is equal to or greater than zero, and weight normalization treatment is adapted to svmlight. In transfer learning, it needs to input the normalized weight to train model. Although svmlight is suitable for transfer learning, it is hard to integrate into the established machine learning system. The re-sampling [20] offers a new way for this problem, whose cost-proportionate is separated from transfer learning algorithm, and training set is sampled from database.

$$E_{x,y,\hat{P}}[Istance] = \frac{1}{N}E_{x,y,P}[w * Instance] \qquad (1)$$

According to the expected formula, we can get Equation.2 directly from Equation.1.

$$\hat{P}(x,y,w) = \frac{w}{E_{x,y,w\sim P[w]}}P(x,y,w) \qquad (2)$$

$\hat{P}$ is the distribution of instance weights, x is the input sample, y is the sample category, w is the weight of the sample, P is the distribution of feature space, $\frac{w}{E_{x,y,w\sim P[w]}}$ is the probability of the re-sampling.

Equation.2 shows that the distribution of feature spaces with instance weights can be re-sampled by weighting the original sample. Thus, it presents a theory of the re-sampling based on instance weights, which is also called black box. Black box separates the cost-sensitive from the machine learning algorithm, which can also be used for big database model. However, it is impossible to directly put all samples into training machine learning model. So black box of re-sampling can draw samples from big database directly and construct the training set, as shown in Fig. 3. This paper uses the rejection sampling, it obtains some batches of training set each time. Each sample in batch of training set is different, it can reduces the effects of over-fitting.

B. Transfer learning based on rejection sampling

Transfer learning has been successfully applied in many fields. And this paper researches supervised learning model and focus on the same tasks. X_a and T_b both are labeled data, and the number of T_b is always very small in practice, which

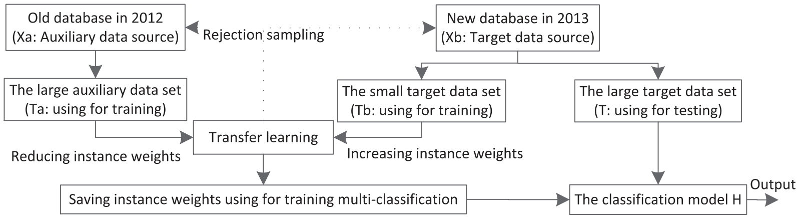
Old database in 2012
(Xa: Auxiliary data source)
Rejection sampling
New database in 2013
(Xb: Target data source)
The large auxiliary data set
(Ta: using for training)
The small target data set
(Tb: using for training)
The large target data set
(T: using for testing)
Transfer learning
Reducing instance weights
Increasing instance weights
Saving instance weights using for training multi-classification
The classification model H
Output

Fig. 2. The system of multi-classification

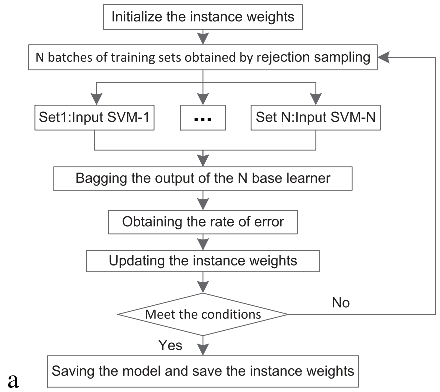
Initialize the instance weights
N batches of training sets obtained by rejection sampling
Set1:Input SVM-1
...
Set N:Input SVM-N
Bagging the output of the N base learner
Obtaining the rate of error
Updating the instance weights
Meet the conditions
No
Yes
Saving the model and save the instance weights
a

Fig. 4. Flowchart of transfer learning by rejection sampling

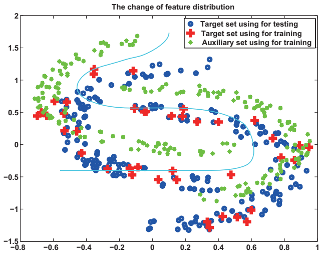

Fig. 5. The changes of the feature distribution

is hard to train a new classifier. There is a large number of X_a data, which can help to increase the recognize rate.

$Y = 0, 1$ is the target domain of binary classification. T_a is the training sample, which is part of X_a. T_b is the other training sample, which is part of X_b. T is the test set.

The multi-classification transfer learning is based on rejection sampling, and the base learner is SVM, which can be used as a multi-classification by using OVR forms. The program flow is shown in Fig. 4.

The wireless signals collected in 2012, 2013 are adopted in the experiment. This paper uses signals collected in 2012 as X_a, and signals collected in 2013 as X_b. Every base classifier has different effect, we use SVM as the base learner, which is more intuitive and structural risk minimization. N indicates the iterative times, then the algorithm is as follows:

- Initialize the weights as $w = (w_1, ..., w_{a+b})$, a and b are the number of T_a and T_b.
- Set $\beta = \frac{1}{1+\sqrt{2lna/N}}$ [21].
- Repeat: for $i = 1 : N$
  - Normalize the weights as $p^i = \frac{w^i}{\sum_{j=1}^{a+b} w^i}$
  - With the distribution P of weights, taking t times of rejection sampling to construct the training set. Then the process is as follows:
  - Repeat: for $k = 1 : t$
    - $*$ Define $T^{'}$ as a training subset, which is selected from T by rejection sampling with P distribution, and then put $T^{'}$ into the base learner h_k.
    - $*$ Output is as follows: $h(x) = sign(\sum_{k=1}^{t} h_k(x))$
  - Calculate the error rate $\varepsilon_i < 0.5$. $\varepsilon_i = \sum_{j=a+1}^{a+b} \frac{w_j^i |h_i(x_j) - y(x_j)|}{\sum_{j=a+1}^{a+b} w_j^i}$ [21].
  - Set $\beta_i = \frac{\varepsilon_i}{1-\varepsilon_i}$, and update the weights as:

$$w_j^{i+1} = \begin{cases} w_j^i \beta^{|h_i(x_j) - y(x_j)|} & j = 1, ..., a \\ w_j^i \beta^{-|h_i(x_j) - y(x_j)|} & j = a+1, ..., a+b \end{cases}$$

- The average loss tends to zero when it has been trained for $N/2 \sim N$ times [21].

$$f(x) = \begin{cases} 0 & \sum_{i=N/2}^{N} len(\frac{1}{\beta_i}) h_i(x) \geq \frac{1}{2} \sum_{i=N/2}^{N} len(\frac{1}{\beta_i}) \\ 1 & \text{other} \end{cases}$$

Fig. 5 shows the changes of feature distribution, the green hollow circle points represent the distribution of the auxiliary set(X_a). The red cross shape and the blue solid circle points both represent the distribution of the new target X_b, which has bias compared to the distribution of X_a.

It needs a long time if N is too big. The model in the previous iteration has been close to the stability, and the model will too concerned about the wrong sample too. So we set $N = 50$ for transfer learning, the red cross shape points are the T_b, and the blue solid circle points are the test data T. SVM is the base learner. The training result is shown in Fig. 6.

In Fig. 6, 1 to 50 scale in axis of X means that transfer

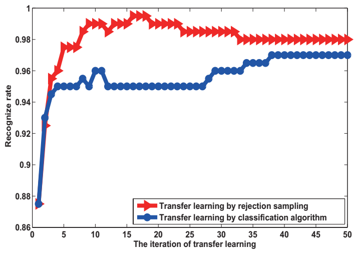

Fig. 6. Recognition rate based on transfer learning

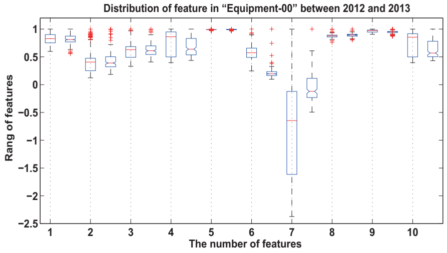

Fig. 7. The ”Equipment-00” feature distribution in 2012 and 2013

learning updates the instance weights 50 times. The blue line is transfer learning trained by classification algorithm, whose recognition rate is nearly 88% when training SVM with X_a and T_a directly rather than transfer learning. After updating the instance weights 50 times by transfer learning, the recognition rate increases to nearly 97%. The red line is transfer learning by rejection sampling with a slight improvement in performance. In a word, transfer learning based on rejection sampling is reasonable and practicable.

IV. Results analysis

Although the feature distribution is changed because of the wireless environment and aging of analog circuits, some remains stable, such as Fig. 7.

Fig. 7 shows the feature distribution of the terminal sensor nodes ”Equipment-00” in 2012 and 2013. The ten boxes plotted with dotted line present the feature distribution in 2012, and the notched box plotted without dotted line is 2013. There is some feature offset, as the number of 6, 7, 10 in Fig. 7. The feature offset is not the same as different signals, but there are still some features can remain stable, it can help to transfer the knowledge to the target source.

Fig. 8 shows the Euclidean distance of ”Equipment-00” in 2012 and 2013. The feature space changes little, so transfer learning based on instances is suitable for the situation of Fig. 8. A different situation is that when feature space has large offset, as is shown in Fig. 9, transfer learning based on instances has worse performance in this case. The solution to the problem in Fig. 9 requires the use of feature-based transfer learning, but this is irrelevant to this discussion.

There are 10 dimensions of Garbo features of wireless equipment, which are not the same with 2 dimensions of features in Fig. 5. Labeling the ”Equipment-00” samples to class 0, and another to class 1. The number of class 0 and class 1 samples are the same. The number of T_a samples in 2012 is set to 100 and the number of T_b samples in 2013 is set to 10. T is another 100 samples extracted for testing model. The result is shown in Fig. 10.

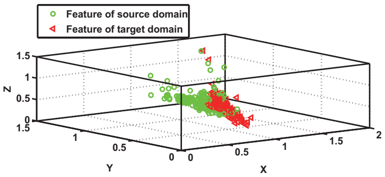

Fig. 8. Small changes in feature space

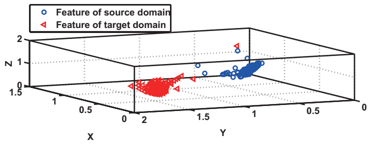

Fig. 9. Large offsets in feature space

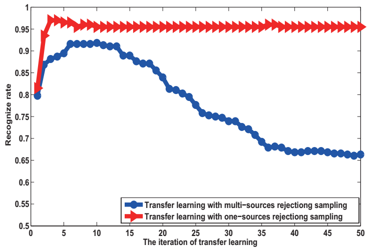

Fig. 10. Transfer learning of multiple sources by rejection sampling

In Fig. 10, the blue line improves quickly, and then decreases slowly, which is called negative transfer. This paper uses the multiple sources as auxiliary data sources, the multiple sources can reduce the degree of steep and increase the number of selected samples. The performance is great as shown in red line. So, multiple sources can better solve the problem of negative transfer.

Three learning models are used to compare the performance of transfer learning, where T_b and T_a are changed and other parameters are the same with Fig. 10. In the first model, The number of T_b is between the range of 1 and 100. In the second model, T_b is set to 0, and only the T_a is used to train the SVM. In the third model, T_b is between the range of 1 and 100, and it is combined with the T_a to train SVM. The result is shown in Fig. 11.

The green line in Fig. 11 is the SVM trained with T_b set to 0, its recognition rate is about 70%. The blue line is the SVM trained with T_b which is between the range of 1 and 100, its recognition rate is improved slowly. The red line is that combining T_a with T_b for training transfer learning, in

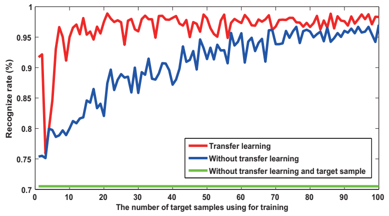

Fig. 11. The comparison of binary classification

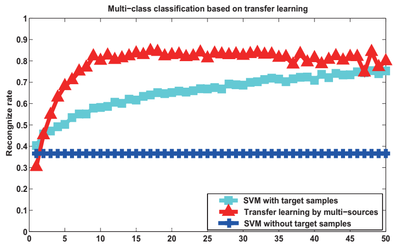

Fig. 12. The comparison of multi-classification

which T_a is 100 and T_b is between the range of 1 and 100, its recognition rate is improved rapidly. However, it appears the negative transfer when T_b is 2. Comparing the three curves, it can be inferred that the performance of transfer learning is good for overcoming the problem of the secular variation in wireless equipment. However, the advantage gradually become weakly when there is a large number of T_b. In this case, transfer learning only can transfer a little knowledge to the target source, due to the fact that the number of T_b is enough to train a good classification model.

The task of this experiment is to classify five class of wireless equipment, and the process of multi-classification transfer learning can be separated into four steps. Firstly, putting one target class, auxiliary training data and one negative class into train model. Secondly, saving the instance weights of training result. Thirdly, putting the instance weights with database into rejection sampling program and then output the selected samples. Finally, using the selected samples to train the SVM model. The result is shown in Fig. 12.

The result of multi-classification is the same with binary classification. The blue line is a base line, whose model is trained with wireless signals of 2012 and tested with wireless signals of 2013. The base line is to show the performance of the original SVM. The red line is based on transfer learning, and the green line is based on on-line learning. It is clearly that the influences of negative transfer is reduced.

V. Conclusion

This paper adopts transfer learning based on rejection sampling to solve the problem that the individual feature in physical layer has changed, which is caused by wireless equipment aging and interference of environment. In this paper, an idea of constructing multi-classification of transfer learning is put forward, and the result shows that the problem caused by the change of features in physical layer can be efficiently solved by transfer learning. However, some feature distribution are still keeping stable, and this experiment has not been discussed. What's more, the databases without labeled are no used for transfer learning, so the semi-supervised transfer learning also needs to be further discussed.

References

[1] X. Yao, X. Han, X. Du, and X. Zhou, "A lightweight multicast authentication mechanism for small scale IoT applications," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3693–3701, 2013.

[2] X. Du and H. H. Chen, "Security in wireless sensor networks," *Wireless Communications IEEE*, vol. 15, no. 4, pp. 60–66, 2008.

[3] R. Doherty, "Fraud detection system," Apr. 24 2012. US Patent 8,165,563.

[4] Y. Xiao, H. H. Chen, X. Du, and M. Guizani, "Stream-based cipher feedback mode in wireless error channel," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 622–626, 2009.

[5] C. Zhao, X. Wu, L. Huang, Y. Yao, and Y.-C. Chang, "Compressed sensing based fingerprint identification for wireless transmitters," *The Scientific World Journal*, vol. 2014, 2014.

[6] T. J. Carbino, M. A. Temple, and J. Lopez Jr, "A comparison of PHY-Based fingerprinting methods used to enhance network access control," in *IFIP International Information Security Conference*, pp. 204–217, Springer, 2015.

[7] X. Du, Y. Xiao, M. Guizani, and H. H. Chen, "An effective key management scheme for heterogeneous sensor networks," *Ad Hoc Networks*, vol. 5, no. 1, pp. 24–34, 2007.

[8] X. Du, M. Guizani, Y. Xiao, and H. H. Chen, "A routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1223–1229, 2009.

[9] Y. Wu and Q. Ji, "Constrained deep transfer feature learning and its applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5101–5109, 2016.

[10] T. Galanti, L. Wolf, and T. Hazan, "A theoretical framework for deep transfer learning," *Information and Inference*, p. iaw008, 2016.

[11] J. Pan, *Feature-based transfer learning with real-world applications*. PhD thesis, The Hong Kong University of Science and Technology, 2010.

[12] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G.-R. Xue, Y. Yu, and Q. Yang, "Heterogeneous transfer learning for image classification.," in *AAAI*, 2011.

[13] A. Argyriou, A. Maurer, and M. Pontil, "An algorithm for transfer learning in a heterogeneous environment," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 71–85, Springer, 2008.

[14] R. Chattopadhyay, Q. Sun, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, "Multisource domain adaptation and its application to early detection of fatigue," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 4, p. 18, 2012.

[15] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[16] T. Joachims, "Svmlight: Support vector machine," *SVM-Light Support Vector Machine http://svmlight. joachims. org/, University of Dortmund*, vol. 19, no. 4, 1999.

[17] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*, pp. 23–37, Springer, 1995.

[18] B. Zadrozny, J. Langford, and N. Abe, "Cost-sensitive learning by cost-proportionate example weighting," in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pp. 435–442, IEEE, 2003.

[19] W. W. C. R. E. Schapire and Y. Singer, "Learning to order things," *Advances in Neural Information Processing Systems*, vol. 10, no. 451, p. 24, 1998.

[20] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo methods in practice*, pp. 3–14, Springer, 2001.

[21] W. Dai, *Instance-based and Feature-based transfer learning [D]*. PhD thesis, Shanghai Jiaotong University, 2009.