Hard Disk Failure Prediction via Transfer Learning

Rui Zhao[1,2], Donghai Guan[1,2](✉), Yuanfeng Jin[3], Hui Xiao[1,2], Weiwei Yuan[1,2], Yaofeng Tu[1], and Asad Masood Khattak[4]

[1] Nanjing University of Aeronautics and Astronautics, Nanjing, China
{zrui821,dhguan,xiaohui822,yuanweiwei}@nuaa.edu.cn
[2] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China
[3] Yanbian University, Yanbian, China
yfkim@ybu.edu.cn
[4] Zayed University, Abu Dhabi, UAE
Asad.Khattak@zu.ac.ae

Abstract. Due to the large-scale growth of data, the storage scale of data centers is getting larger and larger. Hard disk is the main storage medium, once a failure occurs, it will bring huge losses to users and enterprises. In order to improve the reliability of storage systems, many machine learning methods have been widely employed to predict hard disk failure in the past few decades. However, due to the large number of different models of hard disks in the heterogeneous disk system, traditional machine learning methods cannot build a general model. Inspired by a DANN based unsupervised domain adaptation approach for image classification, in this paper, we propose the DFPTL (Disk Failure Prediction via Transfer Learning) approach, which introduce the DANN approach to predict failure in heterogeneous disk systems by reducing the distribution differences between different models of disk datasets. This approach only needs unlabeled data (the target domain) of a specific disk model and the labeled data (the source domain) collected from a different disk model from the same manufacturer. Experimental results on real-world datasets demonstrate that DFPTL can achieve adaptation effect in the presence of domain shifts and outperform traditional machine learning algorithms.

Keywords: Disk failure · Transfer learning · Heterogeneous disk systems

1 Introduction

The development of the Internet has brought about an explosive growth in the amount of data, and the reliability of hard disks as the common and primary storage devices is crucial. Because of the complex structure and huge volume of the storage system, hard disk failures become the norm. However, due to the physical characteristics of the hard disk, once the hard disk fails, it will often cause a relatively large accident, and at the slightest degree, the data service provided by the data center is unavailable, and at the worst, it may cause permanent loss of the stored data, causing huge losses to users and

Y. Tian et al. (Eds.): ICBDS 2020, CCIS 1415, pp. 522–536, 2021.
https://doi.org/10.1007/978-981-16-3150-4_43

enterprises. Microsoft has made statistics on hardware failures in data center [1]. Of these replacements failures, a majority (78%) were for hard disks, followed by a few (5%) due to raid controller and even fewer (3%) due to memory. In a year, approximately 2.7% of hard drives in the data center have been replaced. Furthermore, the hard disk as hardware, the longer it is used, the greater the probability of failure. Therefore, in order to improve the reliability of the storage system, some fault tolerance mechanisms have been adopted, which are mainly divided into passive fault tolerance and active fault tolerance. Compared with the passive fault tolerance mechanism of copying and erasing code, the advantage of active fault tolerance is that it can predict hard disk failure in advance. So that users have sufficient time to take protective measures, which can greatly reduce the loss of enterprises and users.

The original active fault tolerance mechanism of hard disks was implemented by the Self-Monitoring, Analysis and Reporting Technology (SMART) technology [2] built into the hard disk by hard disk manufacturers. Through SMART technology, individual hard disk can be monitored, and the collected information of each state attribute inside the hard disk is compared with the predefined fault threshold, if any attribute value exceeds its threshold, it will raise an alarm. However, this threshold-based detection method can only achieve a failure detection rate of 3%–10% at most (that is, it can predict 3%–10% of failed hard drives) with 0.1% false alarm rate [3, 4]. In order to solve the problem of low accuracy of hard disk failure prediction, many supervised machine learning methods are widely used [5−12]. These methods formulate the hard disk failure prediction problem as a binary classification problem. Specifically, these approaches take SMART attributes as input, and each hard disk is classified either as health or failure by the trained classifier. But these methods usually assume that training data and test data have the same distribution. The real storage systems consist of a large number of different models of hard disks, which are called heterogeneous disk systems [13, 14]. The distributions of hard drives SMART attributes from different models are different, so if the model trained with data from one model of hard drive is directly applied to predict another different model of hard drive, the results are usually inaccurate.

In order to eliminate the training–testing mismatch in heterogeneous disk systems, we utilize deep transfer learning technology to build an accurate and effective hard disk failure prediction model. Our work is inspired by a DANN based unsupervised domain adaptation approach for image classification [15]. This approach conducts domain adaptation based on the idea of adversary. Specifically, the classifier, the feature extractor, and the domain discriminator are learned at the same time. By minimizing the classifier error and maximizing the discriminator error, the learned feature representation has cross-domain invariance. Then in this feature representation space, the discriminative model learned from source domain features can also be applied to target domain features. In addition, this approach only needs the labeled training data from the source domain and unlabeled data from the target domain. In this study, we introduce the DANN approach to predict failure in heterogeneous disk systems by reducing the distribution differences between different models of hard disk datasets. However, applying DANN to predict disk failure is not straightforward and trivial. A critical challenge is one SMART record of the hard disk is collected at a specific time point, so it is one-dimensional, which is different from two-dimensional image features. To tackle this challenge, in this paper, we study how to construct the original 1D SMART data into 2D. The constructed 2D SMART attributes can directly deploy the deep transfer learning algorithm of DANN. We call the proposed approach DFPTL (Disk Failure Prediction via Transfer Learning).

We have conducted experiments on public datasets to verify the effectiveness of our method. The experimental results show that DFPTL achieves better performance than traditional unsupervised and supervised approaches when performing disk failure prediction on target dataset using the model learned from source dataset. The main contributions of our paper are summarized as follows:

- To the best of our knowledge, we pioneer the use of the transfer learning method based on unsupervised domain adaptation to predict disk failure, which can transfer knowledge of labeled disk data in the source domain to predict disk failure in the target domain.
- To deploy the deep transfer learning algorithm of DANN, we reconstruct the original 1D SMART data into 2D SMART data.
- We evaluate our approach on real-world datasets, and the experimental results demonstrate the effectiveness of the method.

The remainder of this paper is structured as follows: We first survey the related work in Sect. 2. Section 3 describes the details of the proposed method. Section 4 discusses experimental settings and results. Finally, conclusions are drawn in Sect. 5.

2 Related Work

Nowadays, most manufacturers equip hard disks with SMART technology to monitor and analyze the health status of the hard disk. A SMART record of the hard disk contains at most 30 meaningful attributes, describing the operating status of the hard disk from various aspects. Each SMART attribute contains four values, Raw value, Normalized Value, Threshold Value and Worst Value.

- ID: The unique identifier assigned to the SMART attribute.
- Raw: Measured value of each attribute when the hard disk is running, such as Celsius degree, Power-On Hours, etc.
- Normalized: Calculated by a specific algorithm built into the hard disk using its raw value.
- Threshold: The reliable attribute value specified by the hard disk manufacturer is calculated by a specific formula. If an attribute value is lower than the corresponding threshold, it means that the hard disk will become unreliable.
- Worst: The largest abnormal value that has ever occurred in the operation of the hard disk.

Because the raw value and normalized value can best reflect the current health status of the hard disk, in our paper, we use them as the characteristic value for building the hard disk failure prediction model.

There have proposed many machine learning algorithms for disk failure prediction models based on SMART data. Hughes et al. [16] proposed two statistical hypothesis test methods to improve performance of the detection method based on the SMART threshold. They used Wilcoxon rank-sum test and OR-ed single variate test and achieved 60% failure detection rate (FDR) and 0.5% false alarm rate (FAR). Hamerly et al. [17] studied two Bayesian methods named Naive Bayes clusters trained using expectation-maximization (NBEM) and naive Bayes classifier, and conducted experiment on a dataset from Quantum Inc., which contains 1927 hard disks, but only 9 failed hard disks. Experimental results show that under the condition of a FAR of 1%, the FDR of naive Bayes classifier can reach 55%, and the FDR of NBEM is 35–45%. Wang et al. [18] proposed a disk failure prediction model based on Mahalanobis Distance (MD), which converts multivariate SMART data into a single variable representing the health degree of disks. The health degree represents the change in the health of the hard disk. Finally, a specific health degree is used to analyze the abnormal changes of the hard disk's health state. When there are enough abnormal changes in a certain period of time, it means that the hard disk is about to fail. They achieved a 68% FDR with 0% FAR. Zhao Y et al. [9] proposed to employ Hidden Semi-Markov Models (HSMMs) and Hidden Markov Models (HMMs) to predict disk failure. They believe that there is a connection between the continuously collected SMART attribute values and the hard disk health status. The proposed model uses the connection of the same SMART attribute at different time points to represent the health status of the hard disk, which has the advantage that it does not require expensive parameter searching. Experimental results show that when using the best SMART attributes, HSMMs can achieve a failure detection rate of 30% and a false alarm rate of 0%, while HMMs model achieves a failure detection rate of 46% and a false alarm rate of 0%.

With the development of neural network technology, researchers have gradually turned their attention to the field of neural networks. A neural network includes an input layer, several hidden layers, and an output layer. Neuron nodes of different levels are connected by specific network weight values. The neural network-based hard disk failure prediction method uses historical hard disk SMART data as the input of the input layer to adjust and optimize the network weights of nodes at different layers in the network, and complete the training of the neural network model; when performing hard disk failure prediction, the real-time hard disk SMART data is input and processed by the entire neural network model to obtain the predicted hard disk operating state. Zhu et al. [11] implemented a backward propagation (BP) neural network model and an improved support vector machine (SVM) model. Both models were tested on a dataset from Baidu Inc., including 22962 good drives and 433 failed drives, and achieved much higher prediction accuracy compared with previous studies. Because SMART attributes gradually deteriorates over time, the methods mentioned above do not take into account time series characteristics. Xu et al. [19] introduced a new method based on Recurrent Neural Network (RNN) to assess the health of hard drives, making full use of the timing of SMART data. Experiments results show that this method can not only reasonably explain the health status of the hard disk, but also achieve better prediction results. But the standard RNN algorithm has the problem of gradient explosion or gradient disappearance. When the number of neurons in the loop layer is large, the early input historical data will be invalid due to the disappearance of the gradient. Lima et al. [33] made improvements on this basis and proposed a variant algorithm of recurrent neural network LSTM (Long Short-Term Memory) for long-term prediction of hard disk failures. Compared with the traditional recurrent neural network algorithm, this algorithm can achieve similar results in short-term prediction, and has a significant improvement effect in long-term prediction.

The above methods are all supervised algorithms, based on sufficient labeled data, and only train a model using SMART data from one disk model is not applicable to other different models even from the same manufacturer. When existing data set that are relevant but not identical to the target domain are available, transfer learning becomes an effective solution. Transfer learning has a wide range of applications, including but not limited to computer vision [20], text classification [21], behavior recognition [22], medical health [23] and so on. In recent years, transfer learning has been applied to the field of hard disk failure prediction. Botezatu et al. [10] adopted an instance-based transfer learning method to eliminate the sample selection bias between source data and target data, so as to apply a model trained on a specific hard drive model to a new one from the same manufacturer. Its main idea is to train a classifier to indicate the probability that a hard disk belongs to a certain model, and then utilize the classifier to sample the labeled hard disk model to make it obey the same distribution as the target hard disk dataset. Pereira et al. [13] proposed a new source building method called clustering-based information source and groups them according to their similarity to build a novel information source for transfer learning. Zhang et al. [24] explored an iterative transfer learning approach to solve the failure prediction problem of minority disks lacking sufficient training data. Specifically, the weights of instances are adjusted in each iteration, and larger weights are assigned to the instances of the majority disks that are similar to the instances of the minority disks, otherwise, smaller weights are assigned. They also proposed a method to select appropriate disk models based on the KLD value. These instance-based transfer learning methods mainly reduce the differences in the distribution of the source domain and the target domain in two ways: by adjusting the weight of the source domain instance and selecting the source domain instance according to the similarity with the target domain instance, they make it possible to transfer health status information from one disk model with enough data available to another disk model with insufficient data. Different from these methods, the transfer learning component in our method is a feature-based transfer learning method that learns a common feature representation space on the source domain and the target domain to realize shared-classifier.

3 Proposed Approach

In this section, we will introduce the details of our proposed approach, called DFPTL (Disk Failure Prediction via Transfer Learning), which use transfer learning technology for disk failure prediction. Figure 1 shows the overall workflow of DFPTL.

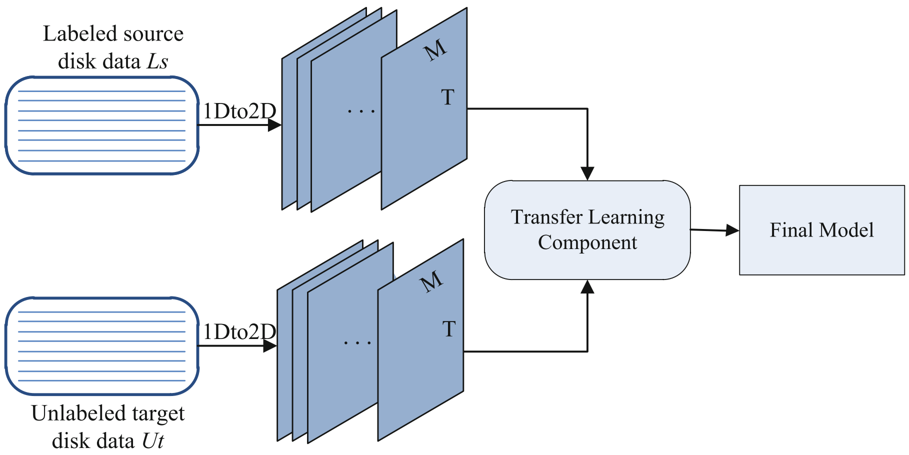

Fig. 1. The overall workflow of DFPTL. For each individual disk, we first use 1D to 2D technology to construct SMART records from two different domains L_s and U_t as 2D-SMART attributes of size M * T, then input them into the transfer learning component to train the base model.

There are two sets of input data. The unlabeled disk data U_t to be predicted serves as the target domain, and the labeled disk data L_s from the same manufacturer but different model as the target domain hard disk serves as the source domain. Every disk is classified either as "health" or "failure". DFPTL comprises three main components: (1) data processing, (2) Construction of 2D-SMART attributes, (3) transfer learning. We will describe more details about each component in the following sections.

3.1 Data Processing

In our disk dataset, each sample contains up to 30 SMART attributes, but some attributes are useless for failure prediction since they keep unchanged during operation, so we get rid of these attributes, while some attributes change significantly over time, and the values on healthy and failed hard disks are obviously different, so we keep these attributes. We select 15 attributes using principal component analysis (PCA), the selected SMART attributes are shown in Table 1.

Table 1. The 15 selected SMART attributes.

Smart ID	SMART attribute name	Attribute type
1	Raw read error rate	Normalized
3	Spin-up time	Normalized
5	Reallocated sectors count	Raw
7	Seek error rate	Normalized

(*continued*)

Table 1. (*continued*)

Smart ID	SMART attribute name	Attribute type
9	Power-on hours	Normalized
187	Reported uncorrectable errors	Raw
188	Command timeout	Raw
190	Airflow temperature	Raw
193	Load/unload cycle count	Raw
194	Temperature	Normalized
197	Current pending sector count	Raw
198	Offline uncorrectable sector count	Raw
240	Head flying hours	Raw
241	Total LBAs written	Raw
242	Total LBAs read	Raw

The 15 attributes above have different value intervals. The large difference in the value interval makes the model difficult or even unable to converge. Therefore, normalization is very necessary. We use min-max scaling [5, 12] to normalize the range of selected SMART attributes.

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x_i is the original value of i-th SMART attribute, x_{max} and x_{min} are the maximum and minimum value of the attribute, respectively. After normalization, all attribute values are mapped to the range [0, 1].

3.2 Construction of 2D-SMART Attributes

The DANN algorithm [15] was first proposed for image classification and achieved good performance in experiments. Correspondingly, the feature extractor in the DANN model architecture is composed of convolutional neural networks (CNN) to automatically extract image features. After the steps in Sect. 3.1, the feature of each sample is one-dimensional, referred to as 1D-SMART attribute. In order to support the deployment of DANN for transfer learning, the 1D-SMART attribute needs to be transformed into a two-dimensional feature similar to an image. In [25], in order to use the GAN-based model for disk failure prediction, Jiang et al. convert 1D-SMART attributes into 2D attributes chunks, this technology is called 1Dto2D. Inspired by [25], we employ 1Dto2D to reconstruct the 1D-SMART attributes into 2D-SMART attributes, which is regarded as the input of the feature extractor. As shown in Fig. 2, 1D-SMART attribute represents a SMART record of a hard disk, including M SMART attributes after feature selection. Then stack continual 1D-SMART attributes and segment data with time window of size T. The 2D-SMART attributes constructed in this way is conducive to deploying the deep transfer learning algorithm of DANN, and can take advantage of the automatic feature extraction of the CNN-based feature extractor.

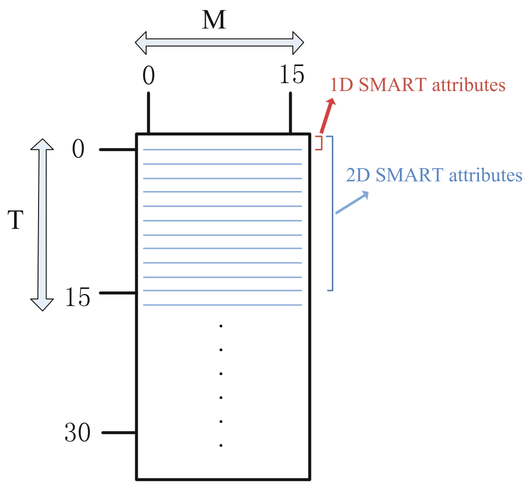

Fig. 2. Construction of 2D-SMART attributes. Each row of records (blue lines) represents selected 15 SMART attributes of a hard disk, and then stack the continuous SMART attributes. Because SMART attributes are collected on a daily basis, T represents hard disk data for T consecutive days. We call the constructed SMART attributes of size M * T as 2D-SMART attributes.

3.3 Transfer Learning Component

The main idea of transfer learning is to transfer related but different domain knowledge to complete or improve the learning effect of the target domain, which is suitable for situation where the source domain and target domain have different distributions. According to the analysis of [10, 24], we know that different models of hard drives exhibit different SMART value distributions, even from the same manufacturer, we refer to this phenomenon as covariate shift [27]. Therefore, the failure prediction model trained on disk data of one model can't be directly transferred to other models of hard disks, otherwise the prediction results will be inaccurate. The solution is to use the transfer learning algorithm to train the model with a large amount of labeled SMART data from source domain and a large amount of unlabeled SMART data from target domain. Inspired by the DANN based unsupervised domain adaptation approach for image classification [15], we adopt DANN to predict disk failure.

Figure 3 shows how to use the DANN approach in disk failure prediction. For the unlabeled disk data U_t in the target domain that needs to be predicted, and the labeled disk data L_s of other models in the source domain, first of all, we adopt the 1Dto2D method to construct each 1D-SMART attribute of U_t and L_s into image-like 2D-SMART attributes, which is then used as the input of the DANN-based transfer learning model. The network architecture is based on a standard feedforward neural network, which consists of three parts: a deep feature extractor G_f used to extract features, generally composed of convolutional layers and pooling layers; A label predictor G_y, which consists of fully connected layers and a logistic classifier, has an output of 0 (health) or 1 (failure); As well as a domain classifier G_d, which forms the adversarial network framework with G_f, is composed of fully connected layers and a cross-entropy classifier. We denote the source domain with 1 and the target domain with 0. The loss function of DANN is defined as:

$$\mathrm{E}\left(\theta_f, \theta_y, \theta_d\right) = \sum_{\substack{i=1..N \\ d_i=[0,1]}} L_y^i\left(\theta_f, \theta_y\right) - \lambda \sum_{i=1..N} L_d^i\left(\theta_f, \theta_d\right) \tag{2}$$

Where θ_f, θ_y, θ_d are the parameters of the network G_f, G_y, G_d respectively. For the i-th training sample, L_y^i and L_d^i represent label prediction loss and domain classification loss respectively. The parameter λ is introduced to trade off two losses during learning. Then we utilize the standard stochastic gradient (SGD) approach to seek the optimized parameters.

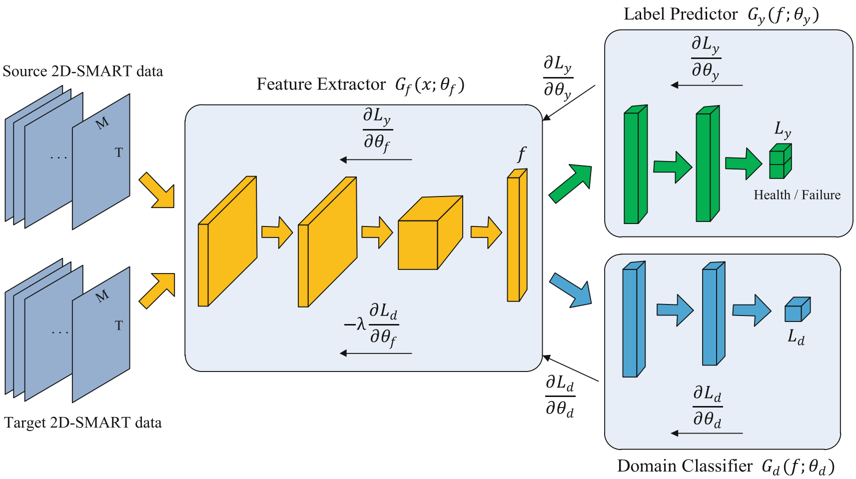

Fig. 3. The DANN approach used for disk failure prediction

4 Experimental Results

4.1 Dataset

We used BackBlaze's public dataset to evaluate our proposed method. From the datasets, we select disk data of different models from two manufacturers. Seagate's ST4000DM000 as the source domain dataset and Seagate's ST12000NM0007 as the target domain dataset respectively. And two models of hard drives from HGST manufacturer, HDS722020ALA330 and HDS5C3030ALA630 as source domain dataset and target domain dataset, respectively. Each hard disk is classified either as "health" or "failure", and each hard disk has many SMART records. Table 2 lists the selected datasets. In order to alleviate the problem that there are much more health disk samples than failure disk samples, we under-sample the health samples to balance the dataset [30]. We have chosen a 1:5 ratio of failure disks to health disks.

Table 2. The selected disk models

Manufacturer	Disk model	Health	Failure
Seagate	ST4000DM000	8230	1646
	ST12000NM0007	7650	1530
HGST	HDS722020ALA330	4580	229
	HDS5C3030ALA630	4020	134

4.2 Evaluation Metric

We evaluate the effectiveness of our proposed approach using Precision, Recall, F1-Score and AUC metrics, which have been widely used to evaluate the capability of a classification model in machine learning [29]. Precision is defined as the proportion of predicted failed disks that are predicted accurately:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Recall represents the proportion of true failed disks that are correctly predicted as failed:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

Where TP, FP and FN denote true positive, false positive and false negative, respectively. F1-Score is the balance between Precision and Recall. The higher the F1-Score, the better the model:

$$\text{F1} - \text{Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

AUC represents the Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) curve, it considers the classifier's ability to classify positive and negative samples at the same time. In the case of imbalanced classes, it can still make a reasonable evaluation of the classifier.

4.3 Results

The Effectiveness of Transfer Learning Component. In order to evaluate the effectiveness of the transfer learning component in our method, the transfer learning model trained with the labeled disk data of the source domain and the unlabeled disk data of the target domain will be compared with the SOURCE-ONLY model, which is trained without consideration for target-domain data (no domain classifier branch included into the network). In addition, The TRAIN-ON-TARGET model serves as an upper performance bound, which is trained and tested on the target domain, assuming labels are available.

Table 3. Experimental results of different methods.

Disk model	Methods	Precision	Recall	F1-Score	AUC
ST-A $\rightarrow$ ST-B	DFPTL	0.4309	0.8030	**0.5608**	**0.8332**
	SOURCE-ONLY	0.2204 0.2204	0.9303	0.3564	0.7533
	TRAIN-ON-TARGET	0.8011	0.8909	0.8436	0.8816
HDS-A $\rightarrow$ HDS-B	DFPTL	0.6923	0.6338	**0.6618**	**0.8101**
	SOURCE-ONLY	0.4742	0.6479	0.5476	0.8066
	TRAIN-ON-TARGET	0.8425	0.7571	0.7975	0.8471

Table 3 shows the Precision, Recall, F1-Score and AUC for disk failure prediction for different source and target domains. The experiments are conducted on two pairs of datasets, including ST-A $\rightarrow$ ST-B, and HDS-A $\rightarrow$ HDS-B. ST-A and ST-B represent hard drive models ST4000DM000 and ST12000NM0007 from the Seagate manufacturer, respectively. HDS-A and HDS-B represent hard drive models HDS722020ALA330 and HDS5C3030ALA630 from the HGST manufacturer, respectively. The right side of the arrow denotes the unlabeled disk dataset to be predicted and the left side of the arrow denotes the labeled auxiliary disk dataset.

As we can see, our DFPTL shows higher F1-Score and AUC than SOURCE-ONLY. The reason is that when the SMART attributes distributions of the source domain disk dataset and the target domain disk dataset are different, if the model trained with the disk dataset of the source domain directly predicts the disk failure of the target domain, the result is often inaccurate. The SOURCE-ONLY model only capture the feature distribution of hard disk data in the source domain, and the trained label predictor can only classify the source domain hard disk. Our DFPTL's transfer learning component can map the features of the two domains to the common feature space through the adversarial network framework, reducing the difference in the distribution of SMART attributes, so that the classifier trained based on the labeled data of the source domain can be applied to the target domain. The experimental results verify the effectiveness of the transfer learning component.

The Effectiveness of 1Dto2D. In this section, we evaluate the effectiveness of proposed 1Dto2D approach. We use sliding windows with lengths of 1, 5, 10, and 15 to segment continual 1D-SMART attributes stacked along with time, respectively. Note that each hard drive collects one SMART record a day, so T = 1 represents the 1D-SMART attribute of a certain day, T = 5 represents the SMART records collected for 5 consecutive days, the size of the corresponding 2D-SMART attribute is 15 * 5, and so on, for T = 10, 15, the size of corresponding 2D-SMART attribute are 15 * 10, 15 * 15, respectively. Figure 4 and Fig. 5 represent the prediction results on two target domains, respectively. We calculate the F1-Score of the model of different percentage of labeled data. Therefore, the horizontal axis represents the percentage of labeled data in the target domain. As shown in these figures, when T = 1, the prediction effect of the trained model was the worst, because 1D-SMART is not converted to 2D-SMART attributes, and the advantages of CNN's automatic feature extraction cannot be utilized well. The performance of the model at T = 5, 10, 15 is better than that at T = 1, indicating the effectiveness of 1Dto2D approach. Besides, we observe that when T = 15, the model achieved the highest F1-Score. Therefore, we set T = 15 in the experiments in this paper.

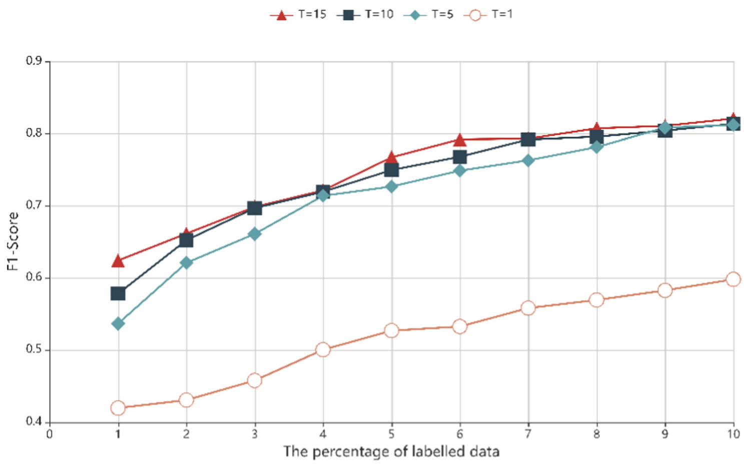

Fig. 4. F1-Scores under different time range T on dataset ST-B.

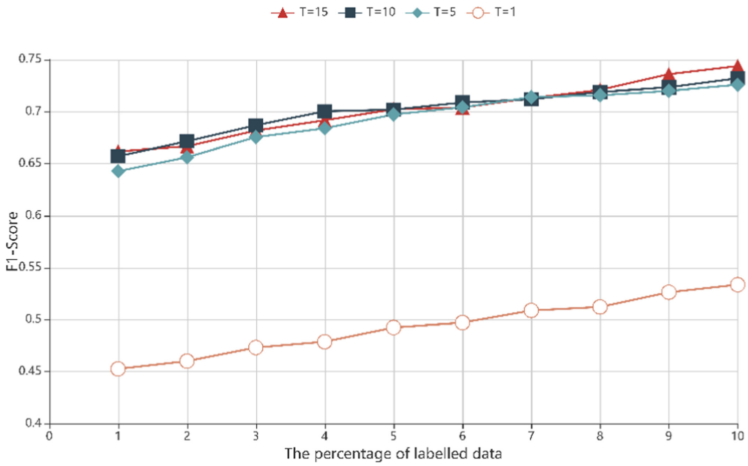

Fig. 5. F1-Scores under different time range T on dataset HDS-B

5 Conclusion

In this paper, we propose an approach called DFPTL for disk failure prediction, which can address the training–testing mismatch problem in heterogeneous disk system. Transfer learning can eliminate the differences of data distribution between the source and the target domains. Moreover, to deploy DANN-based deep transfer learning approach, we stack continuous hard disk SMART data into 2D image-like SMART data. The use of CNN gets rid of manually extracting features. Experimental results with real-world datasets have confirmed that DFPTL can achieve higher detection accuracy, compared to traditional machine learning methods.

Acknowledgements. This research is supported by the key research and development program of Jiangsu province(BE2019012).

References

1. Vishwanath, K.V., Nagappan, N.: Characterizing cloud computing hardware reliability. In: Proceedings of the 1st ACM symposium on Cloud computing, pp. 193–204 (2010)
2. Allen, B.: Monitoring hard disks with smart. Linux J. **117**, 74–77 (2004)
3. Eckart, B., Chen, X., He X., et al.: Failure prediction models for proactive fault tolerance within storage systems. In: 2008 IEEE International Symposium on Modeling, Analysis and Simulation of Computers and Telecommunication Systems, pp. 1–8. IEEE (2008)
4. Murray, J.F., Hughes, G.F., Kreutz-Delgado, K.: Machine learning methods for predicting failures in hard drives: a multiple-instance application. J. Mach. Learn. Res. **6**(May), 783–816 (2005)
5. Xiao, J., Xiong, Z., Wu, S., et al.: Disk failure prediction in data centers via online learning. In: Proceedings of the 47th International Conference on Parallel Processing, pp. 1–10 (2018)

6. Li, J., Ji, X., Jia, Y., et al.: Hard drive failure prediction using classification and regression trees. In: 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 383–394. IEEE (2014)
7. Murray, J.F., Hughes, G.F., Kreutz-Delgado, K.: Hard drive failure prediction using non-parametric statistical methods. In: Proceedings of ICANN/ICONIP (2003)
8. Yang, W., Hu, D., Liu, Y., et al.: Hard drive failure prediction using big data. In: 2015 IEEE 34th Symposium on Reliable Distributed Systems Workshop (SRDSW), pp. 13–18. IEEE (2015)
9. Zhao, Y., Liu, X., Gan, S., et al.: Predicting disk failures with HMM-and HSMM-based approaches. In: Industrial Conference on Data Mining, pp. 390–404. Springer, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14400-4_30
10. Botezatu, M.M., Giurgiu, I., Bogojeska, J., et al.: Predicting disk replacement towards reliable data centers. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 39–48 (2016)
11. Zhu, B., Wang, G., Liu, X., et al.: Proactive drive failure prediction for large scale storage systems. In: 2013 IEEE 29th symposium on mass storage systems and technologies (MSST), pp. 1–5. IEEE (2013)
12. Xu, Y., Sui, K., Yao, R., et al.: Improving service availability of cloud systems by predicting disk error. In: 2018 {USENIX} Annual Technical Conference ({USENIX}{ATC} 18), pp. 481–494 (2018)
13. Pereira, F.L.F., dos Santos Lima, F D., de Moura Leite, L G., et al.: Transfer learning for Bayesian networks with application on hard disk drives failure prediction. In: 2017 Brazilian Conference on Intelligent Systems (BRACIS), pp. 228–233. IEEE (2017)
14. Jiang, W., Hu, C., Zhou, Y., et al.: Are disks the dominant contributor for storage failures? a comprehensive study of storage subsystem failure characteristics. ACM Trans. Storage (TOS). **4**(3), 1–25 (2008)
15. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning, pp. 1180–1189 (2015)
16. Hughes, G.F., Murray, J.F., Kreutz-Delgado, K., et al.: Improved disk-drive failure warnings. IEEE Trans. Reliab. **51**(3), 350–357 (2002)
17. Hamerly, G., Elkan, C.: Bayesian approaches to failure prediction for disk drives. ICML **1**, 202–209 (2001)
18. Wang, Y., Ma, E.W.M., Chow, T.W.S., et al.: A two-step parametric method for failure prediction in hard disk drives. IEEE Trans. Ind. Inf. **10**(1), 419–430 (2013)
19. Xu, C., Wang, G., Liu, X., et al.: Health status assessment and failure prediction for hard drives with recurrent neural networks. IEEE Trans. Comput. **65**(11), 3502–3508 (2016)
20. Zhu, Y., Zhuang, F., Wang, J., et al.: Multi-representation adaptation network for cross-domain image classification. Neural Netw. **119**, 214–221 (2019)
21. Prettenhofer, P., Stein, B.: Cross-language text classification using structural correspondence learning. In: Proceedings of the 48th annual meeting of the association for computational linguistics, pp. 1118–1127 (2010)
22. Wang, J., Zheng, V.W., Chen, Y., et al.: Deep transfer learning for cross-domain activity recognition. In: proceedings of the 3rd International Conference on Crowd Science and Engineering, pp. 1–8 (2018)
23. Kermany, D.S., Goldbaum, M., Cai, W., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–1131 (2018)
24. Zhang, J., Zhou, K., Huang, P., et al.: Transfer learning based failure prediction for minority disks in large data centers of heterogeneous disk systems. In: Proceedings of the 48th International Conference on Parallel Processing, pp.1–10 (2019)

25. Jiang, T., Zeng, J., Zhou, K., et al.: Lifelong disk failure prediction via GAN-Based anomaly detection. In: 2019 IEEE 37th International Conference on Computer Design (ICCD), pp. 199–207. IEEE (2019)
26. Zhang, X., Kim, J., Lin, Q., et al.: Cross-dataset time series anomaly detection for cloud systems. In: 2019 {USENIX} Annual Technical Conference ({USENIX}{ATC} 19), pp. 1063–1076 (2019)
27. Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Stat. Plan. Infer. **90**(2), 227–244 (2000)
28. Yang, Y., Loog, M.: Active learning using uncertainty information. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2646–2651. IEEE (2016)
29. Powers, D.M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2011)
30. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
31. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. IEEE (2008)
32. Breunig, M.M., Kriegel, H.P., Ng, R.T., et al.: LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93–104 (2000)
33. dos Santos Lima, F.D., Amaral, G.M.R., de Moura Leite, L.G., et al.: Predicting failures in hard drives with lstm networks. In: 2017 Brazilian Conference on Intelligent Systems (BRACIS), pp. 222–227. IEEE (2017)