A Survey of Transfer Learning and Categories

Masoume Gholizade[1], Hadi Soltanizadeh[2], Mohammad Rahmanimanesh[3]

***Abstract* - In a variety of real-world scenarios, techniques such as machine learning and data mining are applied. Traditional machine learning frameworks suppose that training data and testing data come from the same domain, have the same feature space, and have the same feature space distribution. This assumption, however, is capable of being applied in certain realistic machine learning cases, especially when gathering training data is prohibitively costly or impossible. As a result, high-performance learners must be developed using data that is more conveniently gathered from various domains. Transfer learning is the name given to this method; it is a learning environment based on a person's capacity to extrapolate information through activities to learn more quickly. Transfer learning tries to establish a structure for applying previous knowledge learned skills to tackle new but related issues more swiftly and efficiently. Transfer learning methodologies, in opposition to traditional machine learning technics, use data from auxiliary domains to enhance predictive modelling of distinct data patterns in the present domain. Transfer learning focuses on improving target participants' performance on target domains by passing data or knowledge from numerous but similar source domains. As a result, the reliance on a various number of target-domain available data for building target learners can be minimized. This survey paper explains transfer learning categories based on problems and solutions and explains experiment results and examples of its application and perspective related to transfer learning. Also, it provides a concise overview of the processes and methods of transfer learning, which may aid readers in better understanding the current research state and idea.**

***Keywords* - Transfer Learning, Source Domain, Target Domain, Task, Domain Adaption.**

INTRODUCTION

Traditional machine learning has created a big impact and has been utilized in a variety of practical applications, but it does have certain limitations in real-world settings. Machine learning assumes that both the testing and training instances in dataset come from the same distribution, although, the assumption, does not prove true in several real-world machine learning settings.

In many situations, gathering enough training data is prohibitively costly, taking a long time, or perhaps even impossible [1]. Semi-supervised machine learning can help to resolution this issue in part by removing the need for large amounts of labeled data. A semi-supervised methodology usually uses a vast volume of data that does not have label, to increase learning accuracy and then contains a limited amount of data that has been labeled [2]. However, collecting unlabeled data is often complicated and traditional models may not provide accurate results, for this reason, to solve this type of problem, the use of transfer learning has been proposed to transfer knowledge, which is inspired by human learning.

Learning to transfer has recently emerged as a way to utilize the knowledge gained to solve real-world problems. This learning method utilizes the knowledge accumulated in the auxiliary domains to build prediction models in the target domain with insufficient training data and in fact, inspired by human learning, creates a bridge to solve problems.

Therefore it is a technique for improving a learner from a domain by transmitting knowledge or data from another [3]. Let's assume two learners who would like to learn the piano as an example. One participant has no prior musical knowledge, while the other has substantial musical expertise gained from playing the guitar [4]. In this situation, the person who can be playing the guitar may take knowledge from a subsequently learned task and apply that to a new task in a helpful way.

In recent years, research and studies focused on transfer learning and its application to real-world problems in computational intelligence have increased, and as a possible solution, transfer learning is intended to use aggregation knowledge in situations where sufficiently labeled data is lacking. In an auxiliary domain, new models make predictions much faster and more efficiently. This auxiliary domain is somehow (or to some extent) related to the main domain. Therefore, transfer learning methods have been proposed to minimize divergence and classification learning with a stronger generalization ability for both training and experimental sets. In transitional learning methods, educational sets and experimental sets are extended to more general concepts, referred to as source domains and target domains [2],[5].

As Figure 1 shows, with knowledge transferred from a source domain, transfer learning can reduce the cost of

[1] Ph.D. Candidate, Electrical and Computer Engineering Dept., Semnan University, Semnan, Iran, Email Address: masoume_gholizade@semnan.ac.ir

[2] Faculty of Electrical and Computer Engineering Dept., Semnan University, Semnan, Iran, Email Address: h_soltanizadeh@semnan.ac.ir

[3] Faculty of Electrical and Computer Engineering Dept., Semnan University, Semnan, Iran, Email Address: rahmanimanesh@semnan.ac.ir

achieving the same level of learning performance in a target domain, or in other words, reduce learning performance by paying the same cost compared to machine learning algorithms. In addition to the cost of tagging training data, this cost can also be attributed to the cost of privacy and time spent training the model. The curve shown in Figure 1 shows the effectiveness of transfer learning:

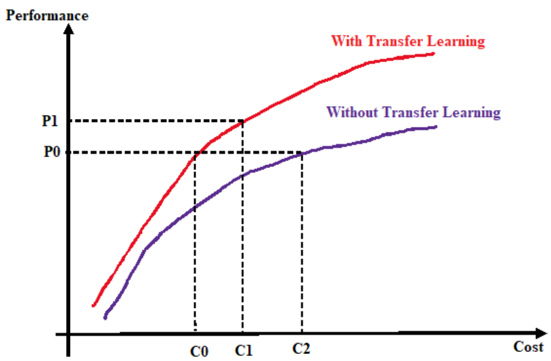

Fig. 1. The use of transfer learning.

It is possible to express a proper description of transfer learning problems, Pen et al. [4] identified two concepts to help distinguish different transfer learning strategies in a survey paper on transfer learning: the first is "domain" and the second is "task". So a domain is made up of space consists of features which has a marginal probability distribution (that is, the features belong to data and their distribution which has in the dataset), and a task is made up of space consists of labels and a function which express objective predictive (that is, a collection of labels and a predictive function learned from data as training)[3], [4] Consequently, a transfer learning challenge may be moving information from a source domain to a separate target domain or from a source task to a distinct target task (or it can be a combination of the two scenarios)[6], where Domain and Task are defined as below:

Definition 1: Assume the domain $\mathcal{D}$ is made up of two parts: a $\mathcal{X}$ as feature space which has $n-dimentional$ and P as marginal distribution for instance set X, that $X = \{x_1, x_2 \dots, x_n\} \in \mathcal{X}$ To put it more simply, it can display as $\mathcal{D} = \{\mathcal{X}, P(X)\}$, According to this definition, if the source domain and target domain are different. It means that the feature space or margin distribution of the features in the two domains are different. [2]. Figure 2 shows the marginal distribution difference, $\mathcal{D}_s$denotes Source Domain and $\mathcal{D}_t$denotes target Domain.

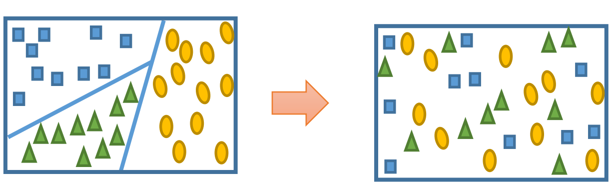

Fig. 2. Marginal distribution difference

Definition 2: Task $\mathcal{T}$ is given for each Domain $\mathcal{D}$. The task $\mathcal{T}$ contains two parts, a label space as $\mathcal{Y}$ and a decision (prediction) function as f, so it can be displayed by $\mathcal{T} = \{\mathcal{Y}, f(x)\}$, where the decision function $f(x)$ predicts label y for instance x.

Definition 3: Assume source domain $\mathcal{D}_s$, learning task $\mathcal{T}_s$, target domain $\mathcal{D}_t$ and learning task $\mathcal{T}_t$, transfer learning tries to enhance the learning of the target prediction function $f(x)_t$in $\mathcal{D}_t$ using the knowledge in $\mathcal{D}_s$and $\mathcal{T}_s$ where $D_s \neq D_t$or $\mathcal{T}_s \neq \mathcal{T}_t$.
The requirement $D_s \neq D_t$ in the above formulation implies that either $\mathcal{X}_s \neq \mathcal{X}_t$ or $P(X)_s \neq P(X)_t$, Similarly, the condition $\mathcal{T}_s \neq \mathcal{T}_t$ implies that either $\mathcal{Y}_s \neq \mathcal{Y}_t$ or $f(x)_t \neq f(x)_s$. it means when two tasks are different then label set or conditional distribution are different. Figure 3 shows the conditional distribution difference.

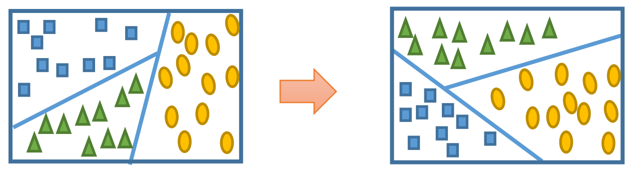

Fig. 3. Conditional distribution difference.

Based on these definitions, we can explain transfer learning as: The source domain $\mathcal{D}_s$, the learning task $\mathcal{T}_s$, the target domain $\mathcal{D}_t$, and the learning task $\mathcal{T}_t$. The purpose of transfer learning is to improve the learning of the prediction function in $\mathcal{D}_t$ using knowledge in $\mathcal{D}_s$, when the domains are different, i.e., $D_s \neq D_t$ or the tasks are different, i.e., $\mathcal{T}_s \neq \mathcal{T}_t$.
In this definition, the difference in the two domains occurs when the feature space or probability marginal distribution are different, i.e., $\mathcal{X}_s \neq \mathcal{X}_t$ or $P(X)_s \neq P(X)_t$. Similarly, two tasks are different when the labeling space is different, i.e., $\mathcal{Y}_s \neq \mathcal{Y}_t$, or the conditional probability distribution is different, i.e., $P(\mathcal{Y}_s|x_s) \neq P(\mathcal{Y}_t|x_t)$ [1]. Table I shows a review of generally used symbols.

TABLE I
The Notations

Symbols	*Explanation*	*Symbols*	*Explanation*
$\mathcal{D}_s$	Source domain	$\mathcal{D}_t$	Target domain
$\mathcal{X}$	Feature space	$\mathcal{Y}$	Label space
$\mathcal{T}$	Predictive learning task	$f(x)$	Objective predictive function
$P(X)$	Marginal distribution	$P(y\|x)$	Conditional distribution

The survey paper supplies a thorough review of the classification of transfer learning strategies based on the problem and a classification based on solution. Therefore, there are some categories based on problem such as inductive, transductive, unsupervised, homogeneous and heterogeneous transfer learning and there are some categories based on solution such as label based, feature based, parameter based and relation based. Accordingly, based on this evidence, the rest of our paper is arranged as follows: Section II consists of a description of the categories of transfer learning techniques depending on problems that are discussed using recent works in that area. In section III, a short survey is conducted in the field of solution categories in transfer learning. In the fourth part, we'll look at some practical transfer learning implementations. Finally, the conclusion corresponds to the final part of the text, which includes an outline of possible works.

TRANSFER LEARNING CATEGORIES BASED ON PROBLEMS

Transfer learning can be classified according to several factors and there is no definitive category of transfer learning. This section organizes transfer learning categories based on the problem by label, space, and learning style, all of which are widely used metrics to assist readers to comprehend transfer learning from a variety of viewpoints. The arrangement of transfer learning categories is depicted in Figure 4.

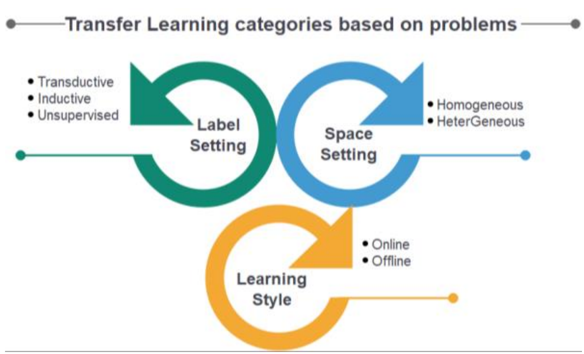

Fig. 4. Transfer Learning categories based on problems.

A. Transfer Learning based on Label-Setting

The problems in this group are put to the determine and evaluated based on the label. Three sub-categories for transfer learning are specified according to the labels defined for the source domain or target. Therefore, in label setting transfer learning, there are three types of transfer learning problems: transductive, inductive and unsupervised transfer learning:

Transductive transfer learning: In this case, there is a lot of labeled data in the source domain but none in the target domain [7], so the domains are different, $(D_s \neq D_t)$, Both source and target tasks are identical. In transductive transfer learning, two further situations occur due to unique circumstances in the source and goal domains. In the first scenario, it is assumed that feature spaces in the source and target domains are distinct, i.e., $\mathcal{X}_s \neq \mathcal{X}_t$, and in the second scenario, it is assumed that feature spaces in the source and target domains are identical but have different marginal probability distribution functions, $P(X_s) \neq P(X_t)$ [8].

Some of the most well-known transductive transfer uses are multi-view object detection, detection of electroencephalogram signals, multi-view object recognition, cross-resolution face recognition, multi-view facial expression recognition, bimodal-vein data mining, and cross-spectral face recognition [9], [10].

Inductive transfer learning: this type of transfer learning applies to scenarios when the source and target tasks are not the same. In inductive transfer learning, the labels of the target domain are known, and the target task and the source task are distinct. In the majority of inductive transfer learning experiments [11], both the source and destination domain labels are known. The variations in the tasks might be explained by a different label space (different categories/classes in the train and test sets) or a different conditional probability distribution $P(Y|x)$. The supervised sample distribution is ascribed to this conditional probability distribution, which is dependent on the quantity of labeled data for each class. inductive transfer learning algorithms aim to improve the estimation of the target probability distribution function $f_t(.)$ in the target domain. The inductive transfer learning strategy emphasizes transferring information from the source task to the target task to achieve excellent results. Multi-task techniques, on the other hand, assign several tasks at the same time, covering both source and target tasks [12].
One of the most well-known forms of inductive transfer learning methods is zero-shot learning [13]–[15], in which The training and testing data's label space (and also the task) are completely disjointed, Multitask learning is a similar paradigm in which the model tries to learn both the source and target tasks using a large amount of data in the source domain. Furthermore, multitask learning makes no assumptions about data scarcity in either the source or target domains. When no source domain data is available for training, self-taught learning is another variation of inductive transfer learning.

Unsupervised transfer learning: It occurs where the label information for both the source and target domains is unclear. It means when there is no labeled data in both the source and target domains, unsupervised transfer learning techniques are used to train an effective model for the target domain. Clustering, dimensionality reduction, and other unsupervised learning problems are examples [16]. Unsupervised transfer learning is identical to inductive transfer learning, with the exception that labeled data is missing from both the source and target domains, Unsupervised transfer learning, on the other hand, focuses on addressing unsupervised activities like Spoken Human Language [17].
This classification is dependent on label space continuity between the source and target [4]. Domain adaptation-based approaches attempt to train a model suited for the target task where $D_s \neq D_t$ and $\mathcal{T}_s = \mathcal{T}_t$, given the source and target domains, D_s and D_t, source and target and tasks $\mathcal{T}_s$ *and* $\mathcal{T}_t$. A

few labelled or unlabeled target domain samples $\mathcal{X}_t$ may be provided in some situations [18]. Table 2 Shows the dissimilarities and similarities in label setting transfer learning.

TABLE II
Dissimilarities and Similarities in Label Setting Transfer Learning

label setting Transfer Learning	**source and target domain**	**Source Domain Label**	**Target Domain Label**	**source and target task**
Inductive	same	Available or unavailable	available	Different but related
transductive	Different but related	available	Un available	same
unsupervised	Different but related	Un-available	Un-available	Different but related

B. Transfer Learning based on Space-Setting

The last categorization for the problems in transfer learning is in the similarity and dissimilarity space of the source and target domain. Based on the category transfer learning may be split into two sub-categories. homogeneous and heterogeneous transfer learning:

Homogeneous transfer learning: Homogeneous domain matching is a subset of transfer learning in which the feature space is the same, i.e., $\mathcal{X}_s = \mathcal{X}_t$ but the corresponding probability distribution (marginal probability distribution) is different, i.e., $P(X)_s \neq P(X)_t$. If we refer an example, two categories of comments are to be considered: comments related to an electronic device such as a digital camera and comments related to a book. Since the dictionary of comment words is the same for both products, they have the same feature space. However, the frequency of words in these two categories of comments is significantly different. For example, users have used the phrase "work well" to express positive comments for the digital camera, while users who have read the book have used the phrase "is good" to express their positive comments. The probabilities for these two domains are different.
Homogeneous domain adaption focuses more on unsupervised homogeneous domain matching , where the source and target domains have the same feature space but the samples in the target domain are not labeled.[3];
Heterogeneous transfer learning: Heterogeneous domain adaption is a subset of transfer learning in which the feature space is not the same, i.e., $\mathcal{X}_s \neq \mathcal{X}_t$, and the corresponding probabilistic distribution is also different .,$P(X)_s \neq P(X)_t$. If we refer to the example before, two categories of comments are to be considered: comments related to an electronic device such as a digital camera and comments related to a book. If the comments are from two different languages, for example, the comments about the digital camera are written in English and also the comments about the book are written in French, these words are used to express that the comments are completely different and therefore have a different feature space. However, the frequency of words in these two categories of comments is significantly different. For example, users have used the phrase "work well" to express positive comments for the digital camera, while users who have read the book have used the phrase "is good" to express their positive comments. The probabilities for these two domains are different.
There are three types of heterogeneous domain matching methods: Heterogeneous Supervised Domain Adaptation (HeSDA), Heterogeneous Semi-Supervised Domain Adaptation (HeSSDA), and Heterogeneous Unsupervised Domain Adaptation (HeUDA).
The majority of works on heterogeneous transfer learning have been conducted in the last years, indicating that it is still a relatively recent field of research. in the papers [19], [20] heterogeneous transfer learning is used to image recognition, in [21]–[23] is used for text classification, and in the paper [24] software defect classification.

C. Transfer Learning based on Learning Style

There are 2 types of learning based on learning styles, the first style is online transfer learning and the second style is offline transfer learning.

Online transfer learning: As new data is created; the model is updated and the model developed through this learning technique is more flexible. Because the data in the target domain is transmitted dynamically and analyzed in real-time, the distribution difference is difficult to evaluate, making offline transfer learning more difficult. Online transfer learning methods are used to explore some specialized issues, such as the regression problem of detecting driver sleepiness [2].

Offline transfer learning: Based on definition in the style, The source domain and the target domain both are resolved in offline transfer learning, therefore the goal of learning is to perform only one knowledge transfer to accomplish the model adaptation [2]. Several signal analysis papers, utilized experimental data from the public P300 database. Due to the lack of online updates, the model may underperform on subsequent data sets after being trained with this previous data via offline transfer learning [2].

III. TRANSFER LEARNING CATEGORIES BASED ON SOLUTIONS

In the field of transfer learning solutions, several approaches have been suggested, the main focus of which is to reduce the distributional distinctions between the source and target domains. In general, the proposed methods in the field of transfer learning are separated into three common categories: the methods as instance-based, the methods as model-based, and the methods as feature-based, based on the survey in [4]. Figure 5 shows Transfer Learning categories based on solutions.

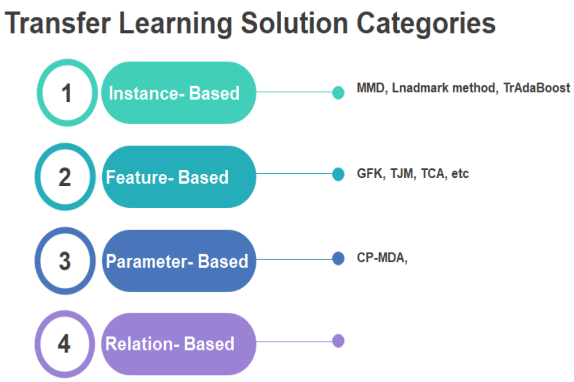

Fig. 5. Transfer learning category based on solution.

Instance-based approach: In an instance-based approach, the goal is to re-weight or select samples from the source that minimize the difference in distribution between the source and target domains [25]. The landmark selection method [25] is one of the famous instance-based methods that referred instances of the source domain that are most similar to the samples of the target domain in terms of distribution. In this method, the *maximum mean discrepancy* (MMD) method is used to calculate the maximum distribution difference between the means of the source domain samples with the target domain samples.

$$MMD(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \varphi(x_s^i) - \frac{1}{n_s} \sum_{j=1}^{n_t} \varphi(x_t^j) \right\|_H^2$$

Based on kernel tricks, MMD can be calculated quickly, in Reproducing kernel Hilbert Space (RKHS). Note that the aforementioned determines the instance weight by minimizing the MMD distance between domains, Then, more weight is assigned to the samples of the source domain that have the least distribution difference with the samples of the target domain. The main problem with the method of selecting landmarks is that in selecting landmarks, some characteristics may be limited to the domain of the source or purpose, which is not considered in the selection of landmarks.

One of the most important methods in this category, proposed by Yao [26], is to transfer knowledge from multiple source domains based on a boosting algorithm in an attempt to reduce the effects of negative transfer from a single unrelated source domain. The boosting process needs some amount of labeled data in target domain, also [27] introduced Gap-Boost, a new multi-source boost method for transfer learning based on instance.

Feature-based methods: Changes the property space to create a flexible representation of the source and target domains. Then In the new space, a standard classifier is trained on the source domain data and applied to the target domain [28], so the original features are transformed into a new feature expression of feature-based approaches. some feature-based methods as TCA, GFK, TJM defines as bellow:

In the Transfer component analysis (TCA) method [29], the variance of the mapped data increases as the marginal distribution difference decreases. This aim preserves the main input data structure. Based on these two goals, a new display of source and target domains is created in a new space which is common between domains [28].

The Geodesic flow kernel method as GFK [30] is another marginal reduction method that maps resource and target data to a new subspace where the distribution of source and target ranges are close together. In this method, due to the reduction of dimension to find the new subspace, the main data is not correctly displayed in the displayed subspace.

The main idea of the GFK method is to embed the source and target data in a manifold space so that the marginal distribution difference between the domains is reduced. This kernel-based approach consists of three steps, which include: determine the optimal dimensionality of the subspaces to embed domains, construct the geodesic curve and compute the geodesic flow kernel.

The transfer joint matching (TJM) [31] method is a combination of instance-based methods and feature-based methods that are proposed for problems that have a high distribution of distribution. The TJM method is a hybrid method using two methods of feature matching and sample selection that simultaneously pursues the following two objectives: 1. A common subspace that has a minimum marginal distribution and creates the maximum conserved variance 2. Uses the $\ell_{2,1}$ norm method to re-weight training domain data. In this method, more weight is assigned to the samples of the source domain that have the most similarity in terms of distribution to the samples of the target domain. The TJM has a fairly complex Kernel and the goal is to solve its objective function

The method in [32] uses a transfer learning technique to improve a target classifier learned to forecast targeted online display advertisement outcomes by using the weighted outputs of multiple source classifiers.

Parameter-based methods: They are built on the idea that features from source domains and the target domain have common parameters. As a result, these strategies are ineffective in circumstances where the domain change is considerable. Parameter-based approaches can be easily extracted from multi-task learning methods in this context. Multi-task learning, on the other hand, is usually required to learn all of the tasks at once, whereas parameter-based transfer learning is solely concerned with maximizing the target task [33].

In multi-task learning, the loss functions for all tasks are the same, but the loss function in the target domain has higher weights in transfer learning.

The conditional probability based multi - source domain adaptation (CP-MDA) technique, which is a procedure based on rectifying conditional distribution discrepancies between the source and target domains, is one of the parameter-based approaches. The CP-MDA method is based on the assumption that only a small quantity of labelled target data is available. The basic concept is to identify the unlabeled target data using a mix of source domain classifiers [34].

Relation-based methods Relationship-based transfer learning is not as well-known as other topics. Relation-based transfer learning strategies, unlike some of the other three categories of learning methods, do not presume that the source and target data are independent and identically distributed [35]. The main focus of these methods is on semi-supervised transfer learning.

As a result, relation-based solutions are far more adaptable and durable than traditional methods. However, in recent years, there haven't been many studies on this subject. Furthermore, the majority of these algorithms are based on statistical learning approaches. Similar relations exist in other domains, which is the basis for relation-based transfer learning. For example, photographs of a teacher giving lectures to students can be found in the source domain, while images of a manager making a speech to staff can be found in the target domain. Despite the fact that the two sets of photos depict different items, they share the same relationship.

IV. THE APPLICATIONS OF TRANSFER LEARNING

The papers cited in the current survey paper show that the transfer learning methodology has been used in a variety of real-world situations. There are a variety of natural language processing technology examples available, including emotion classification, document classification, ham and spam email identification, and classification of texts as multiple languages. Classification of Images and video description classification are two other well-represented transfer learning applications. WiFi localization classification, classification of muscle exhaustion, medication effectiveness classification, human function classification, classification of heart arrhythmias, and device defect classification are some of the applications that have been more specifically discussed in previous articles. Using transfer learning methodologies, for The above issue has been explored in detail and reported in articles [4], [19], [36]–[38], also in The following there are some examples of how transfer learning is used to solve the problems:

Nature language processing: The goal of natural language processing (NLP), often known as the study of human languages, is for computers to be able to understand natural language processing. Sub-learning problems in NLP areas include text-based learning difficulties (e.g., text classification or non-topical text analysis), linguistic knowledge comprehension, and so on. The unlabeled test data and the labelled training samples have different distributions in this situation.

Behbood et al. [39] developed a domain adaptation method based on Fuzzy logic named as FDA technique for text categorization and investigated its applicability. Moreover, emotion categorization, which is a major issue in the field, transfer learning approach, as well as non-topical text analysis by completely leveraging the information from both, such as applying nave Bayes to domain adaptation for sentiment analysis the unlabeled new-domain data set and the old-domain data set.

In addition, the transfer learning technique may be utilized to solve difficulties with linguistic knowledge comprehension. Swietojanski et al. [40] used combines various acoustic data to the target languages in a deep neural network on the unsupervised cross-lingual transfer of knowledge for speech recognition. Huang et al. [41] used a shared-hidden-layer multi-lingual deep neural network to deal with cross-language knowledge transfer learning tasks.

Computer vision and image processing: In these and other situations, transfer learning can save time that would otherwise be spent categorizing the target data by humans. Providing ground truth in some situations, such as picture semantic segmentation, is time-consuming. Transfer learning has been used autonomous navigation [42], for semantic segmentation [43], in a medical context for chest X-ray segmentation [44], and MRI segmentation [45]. Another use of transfer learning in computer vision is knowledge transfer between multiple handwritten character recognition tasks [46].

Also, Ma's paper [47] uses transfer learning to improve global climate simulations by classifying atmospheric dust aerosol particles. Transfer learning is used to improve disease prediction in the paper by [48]. A rule-based learning approach for modelling various types of gene expression data utilizing abstract source domain data is presented in this solution.

In paper [49] For biomedical picture categorization, transfer learning is used in convolutional neural networks, It extracts standard image features from nature image datasets, with a portion of these features being found in minor datasets. Also, the transfer learning methodology is used in [50] and [51]to diagnose and categorize breast cancer and diagnostic imaging, respectively. For breast cancer classification, three related frameworks were examined: GoogLeNet, ResNet, and VGGNet. Even paper [52] proposes a novel method for converting knowledge between the source domain and the target domain. It uses partial samples of the target domain as kernels to start the source knowledge transfer. So, transfer learning is being used to further forecast targeted ads in the growing industry of online display site advertising.

V. CONCLUSION

From the data and model viewpoints, we have described the processes and methods of transfer learning. The survey provides simple descriptions of transfer learning and helps to identify a vast range of representative transfer learning methods and associated works using a single symbol language.

After analyzing various current transfer learning environments, this survey paper paints an image of transfer learning. It can be classified based on problems into five types: inductive, transductive, unsupervised TL, homogeneous and heterogeneous, and also the paper explains the methods of solution rely on some papers which described transfer learning methodologies.

Transfer learning algorithms focused on optimizing different distributions between the source and target domains. Though, in many implementations, we may want to move data across tasks or domains that contain different function spaces, as well as from a variety of such source domains.

In the coming years, transfer learning techniques may be widely used to solve a variety of new and exciting problems. Healthcare, facial recognition, autotuning, human-computer interaction, context perception, and other technologies have all used transfer learning approaches. Future studies in the field of transfer learning will go in some directions. To begin, transfer learning strategies can be investigated further and extended to a broader range of applications. In more complex situations, novel methods are needed to address information transfer issues. In real-world environments, for example, user-relevant source-domain data can come from a different organization.

VI. REFERENCES

[1] F. Zhuang *et al.*, “A Comprehensive Survey on Transfer Learning,” *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[2] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, “A review on transfer learning in EEG signal analysis,” *Neurocomputing*, vol. 421, pp. 1–14, 2021.

[3] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, “A survey of transfer learning,” *J. Big Data*, vol. 3, no. 1, 2016.

[4] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[5] N. Agarwal, A. Sondhi, K. Chopra, and G. Singh, “Transfer Learning: Survey and Classification,” Springer, Singapore, 2021, pp. 145–155.

[6] G. Wilson and D. J. Cook, “A Survey of Unsupervised Deep Domain Adaptation,” *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, 2020.

[7] “Annoyed Realm Outlook Taxonomy Using Twin Transfer Learning,” *Int. J. Pure Appl. Math.*, 2017.

[8] L. Xie, Z. Deng, P. Xu, K. S. Choi, and S. Wang, “Generalized Hidden-Mapping Transductive Transfer Learning for Recognition of Epileptic Electroencephalogram Signals,” *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, 2019.

[9] J. Wang, G. Wang, and M. Zhou, “Bimodal vein data mining via cross-selected-domain knowledge transfer,” *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 3, pp. 733–744, 2018.

[10] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha, “Improving cross-resolution face matching using ensemble-based co-transfer learning,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5654–5669, 2014.

[11] Z. Deng, Y. Jiang, H. Ishibuchi, K.-S. Choi, and S. Wang, “Enhanced Knowledge-Leverage-Based TSK Fuzzy System Modeling for Inductive Transfer Learning,” *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 1, pp. 1–21, Oct. 2016.

[12] R. S. Simões, V. G. Maltarollo, P. R. Oliveira, and K. M. Honorio, “Transfer and Multi-task Learning in QSAR Modeling: Advances and Challenges,” *Front. Pharmacol.*, vol. 9, Feb. 2018.

[13] R. Socher, M. Ganjoo, H. Sridhar, O. Bastani, C. D. Manning, and A. Y. Ng, “Zero-shot learning through cross-modal transfer,” *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, 2013.

[14] M. Palatucci, D. Pomerleau, G. Hinton, and T. M. Mitchell, “Zero-shot learning with semantic output codes,” *Adv. Neural Inf. Process. Syst. 22 - Proc. 2009 Conf.*, pp. 1410–1418, 2009.

[15] B. Romera-Paredes and P. H. S. Torr, “An embarrassingly simple approach to zero-shot learning,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 3, pp. 2142–2151, 2015.

[16] B. Yang, A. J. Ma, and P. C. Yuen, “Learning domain-shared group-sparse representation for unsupervised domain adaptation,” *Pattern Recognit.*, vol. 81, pp. 615–632, 2018.

[17] A. Siddhant, A. Goyal, and A. Metallinou, “Unsupervised transfer learning for spoken language understanding in intelligent agents,” *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 4959–4966, 2019.

[18] S. Ghosh, R. Singh, M. Vatsa, V. M. Patel, and N. Ratha, “Domain adaptation for visual understanding,” *Domain Adapt. Vis. Underst.*, pp. 1–15, 2020.

[19] B. Cao, N. N. Liu, and Q. Yang, “Transfer learning for collective link prediction in multiple heterogenous domains,” *ICML 2010 - Proceedings, 27th Int. Conf. Mach. Learn.*, pp. 159–166, 2010.

[20] Q. Zhu, Yin and Chen, Yuqiang and Lu, Zhongqi and Pan, Sinno Jialin and Xue, Gui-Rong and Yu, Yong and Yang, “Heterogeneous Transfer Learning for Image Classification,” *Proc. Twenty-Fifth AAAI Conf. Artif. Intell.*, pp. 1304–1309, 2011.

[21] C. Wang and S. Mahadevan, “Heterogeneous domain adaptation using manifold alignment,” *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 1541–1546, 2011.

[22] G. J. Qi, C. Aggarwal, and T. Huang, “Towards semantic knowledge propagation from text corpus to web images,” *Proc. 20th Int. Conf. World Wide Web, WWW 2011*, pp. 297–306, 2011.

[23] P. Prettenhofer and B. Stein, “Cross-language text classification using structural correspondence learning,” *ACL 2010 - 48th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 1118–1127, 2010.

[24] J. Nam and S. Kim, “Heterogeneous defect prediction,” *2015 10th Jt. Meet. Eur. Softw. Eng. Conf. ACM SIGSOFT Symp. Found. Softw. Eng. ESEC/FSE 2015 - Proc.*, pp. 508–519, 2015.

[25] B. Gong, K. Grauman, and F. Sha, “Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation,” *30th Int. Conf. Mach. Learn. ICML 2013*, no. PART 1, pp. 222–230, 2013.

[26] Y. Yao and G. Doretto, “Boosting for transfer learning with multiple sources,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1855–1862, 2010.

[27] B. Wang, J. A. Mendez, M. B. Cai, and E. Eaton, “Transfer learning via minimizing the performance gap between domains,” *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[28] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain Adaptation via Transfer Component Analysis,” *IEEE Trans. Neural Networks*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[29] Si Si, Dacheng Tao, and Bo Geng, “Bregman Divergence-Based Regularization for Transfer Subspace Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.

[30] Boqing Gong, Yuan Shi, Fei Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 2066–2073.

[31] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer joint matching for unsupervised domain adaptation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1410–1417, 2014.

[32] C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, and F. Provost, “Machine learning for targeted display advertising: Transfer learning in action,” *Mach. Learn.*, vol. 95, no. 1, pp. 103–127, 2014.

[33] H. Li, Y. Shi, Y. Liu, A. G. Hauptmann, and Z. Xiong, “Cross-domain video concept detection: A joint discriminative and generative active learning approach,” *Expert Syst. Appl.*, vol. 39, no. 15, pp. 12220–12228, 2012.

[34] 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 Trans. Knowl. Discov. Data*, vol. 6, no. 4, 2012.

[35] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, “Adaptation Regularization: A General Framework for Transfer Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.

[36] L. Bin, Y. Qiang, and X. Xiangyang, “Transfer learning for collaborative filtering via a rating-matrix generative model,” *ACM Int. Conf. Proceeding Ser.*, vol. 382, 2009.

[37] B. Li, Q. Yang, and X. Xue, “Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction,” *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 2052–2057, 2009.

[38] M. Jiang, P. Cui, F. Wang, Q. Yang, W. Zhu, and S. Yang, “Social recommendation across multiple relational domains,” *ACM Int. Conf. Proceeding Ser.*, pp. 1422–1431, 2012.

[39] V. Behbood, J. Lu, and G. Zhang, “Text categorization by fuzzy domain adaptation,” *IEEE Int. Conf. Fuzzy Syst.*, 2013.

[40] P. Swietojanski, A. Ghoshal, and S. Renals, “Unsupervised cross-lingual knowledge transfer in DNN-based LVCSR,” *2012 IEEE Work. Spok. Lang. Technol. SLT 2012 - Proc.*, pp. 246–251, 2012,.

[41] J. T. Huang, J. Li, D. Yu, L. Deng, and Y. Gong, “Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 7304–7308, 2013.

[42] J. Yoo, Y. Hong, Y. Noh, and S. Yoon, “Domain Adaptation Using Adversarial Learning for Autonomous Navigation,” 2017.

[43] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, “Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2502–2511, 2019.

[44] C. Chen, Q. Dou, H. Chen, and P. A. Heng, “Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest X-ray segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11046 LNCS, pp. 143–151, 2018.

[45] C. S. Perone, P. Ballester, R. C. Barros, and J. Cohen-Adad, “Unsupervised domain adaptation for medical imaging segmentation with self-ensembling,” *Neuroimage*, vol. 194, pp. 1–11, 2019.

[46] D. C. Ciresan, U. Meier, and J. Schmidhuber, “Transfer learning for Latin and Chinese characters with deep neural networks,” *Proc. Int. Jt. Conf. Neural Networks*, 2012.

[47] Y. Ma, W. Gong, and F. Mao, “Transfer learning used to analyze the dynamic evolution of the dust aerosol,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 153, pp. 119–130, Mar. 2015.

[48] H. A. Ogoe, S. Visweswaran, X. Lu, and V. Gopalakrishnan, “Knowledge transfer via classification rules using functional mapping for integrative modeling of gene expression data,” *BMC Bioinformatics*, vol. 16, no. 1, 2015.

[49] L. D. Nguyen, D. Lin, Z. Lin, and J. Cao, “Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation,” *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2018-May, 2018, .

[50] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[51] S. U. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, “A novel deep learning based framework for the detection and classification of breast cancer using transfer learning,” *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, 2019.

[52] S. M. Salaken, A. Khosravi, T. Nguyen, and S. Nahavandi, “Seeded transfer learning for regression problems with deep learning,” *Expert Syst. Appl.*, vol. 115, pp. 565–577, 2019.