Selecting Reliable Instances Based on Evidence Theory for Transfer Learning

Ying Lv[a], Bofeng Zhang[h,g,c], Xiaodong Yue[b,*], Thierry Denœux[d,e,f], Shan Yue[g]

[a]*Shanghai Artificial Intelligence Laboratory, Shanghai, China*
[b]*Artificial Intelligence Institute of Shanghai University, Shanghai University, China*
[c]*School of Computer Engineering and Science, Shanghai University, China*
[d]*Université de technologie de Compiègne, Compiègne, France*
[e]*Sino-European School of Technology, Shanghai University, China*
[f]*Institut universitaire de France, Paris, France*
[g]*School of Computer Science and Technology, Kashi University, China*
[h]*School of Computer and Information Engineering, Shanghai Polytechnic University, China*

Abstract

The aim of transfer learning is to improve the performance of learning models in the target domain by transferring knowledge from the related source domain. However, not all data instances in the source domain are reliable for the learning task in the target domain. Unreliable source–domain data may lead to negative transfer. To address this problem, we propose a novel strategy for selecting reliable data instances from the source domain based on evidence theory. Specifically, a mass function is formulated to measure the degree of ignorance and reliability of the source domain data with respect to the learning task in the target domain. By selecting reliable instances with low degree of ignorance from the source domain, the domain adaptation of the transfer learning models is enhanced. Moreover, the proposed data-selection strategy is independent of specific learning algorithms and can be regarded as a common preprocessing technique for transfer learning. Experiments on both simulated and real-world datasets validated that

---

*Corresponding author.
  Email addresses: `lvying@pjlab.org.cn` (Ying Lv), `bfzhang@sspu.edu.cn` (Bofeng Zhang), `yswantfly@shu.edu.cn` (Xiaodong Yue), `thierry.denoeux@utc.fr` (Thierry Denœux), `a13579246504@outlook.com` (Shan Yue)

the proposed data selection strategy can improve the performance of various types of transfer learning methods.

---

1. Introduction

Although machine learning has achieved excellent performance in many practical applications (Liang et al., 2021), challenges in real-world scenarios remain. The ideal scenario for machine learning is an abundance of labeled training instances that have the same distribution as the test data. However, collecting sufficient labeled training data is often expensive, time-consuming, and unrealistic in many scenarios. Transfer learning (TL) has been proposed to address the limitation of labeled data in transferring knowledge across domains. (Iman et al., 2023; Ajith & Gopakumar, 2023; Jiang et al., 2022). The primary objective of TL is to enhance or expedite learning tasks in the target domain by acquiring knowledge from a related but not entirely identical source domain. TL has been successfully applied in various fields, including visual object recognition (Sohn et al., 2023; Öztürk et al., 2023), text classification (Gupta & Jalal, 2022; Guo et al., 2020), medical image analysis (Yu et al., 2022; Kora et al., 2022) and machine translation (Zhao et al., 2020), etc.

In transfer learning, the source and target domains typically share some data and knowledge. Through this data sharing and knowledge transfer, the model can leverage information from the source domain to better understand and adapt to the data in the target domain. To obtain as much shared knowledge between the source and target domains as possible, the consistency between the distributions of the two domains should be increased. Many transfer learning methods, such as transfer component anaysis (TCA) (Pan et al., 2010), correlation alignment (CORAL) (Sun et al., 2016), geodesic flow kernel (GFK) (Gong et al., 2012), joint distribution adaptation (JDA) (Long et al., 2013b), balanced distribution adaptation (BDA) (Wang et al., 2017) and scatter component analysis (SCA) (Ghifary et al., 2017), adopt distribution-matching metrics to learn a highly consistent common feature space for both the source and target domain data. In addition, in deep transfer learning approaches, fine-tuning pre-trained strategies have been used a means to reduce the gap between the source and target domains(Subramanian et al.,

2022; Zhong et al., 2023).

Most TL methods consider shared feature representations, model parameters, or even relationships between tasks, overlooking the presence of unreliable instances in the source-domain data. In fact, the source domain includes noisy data, as well as data unrelated to the target domain task that arise from domain differences, such as data noise and label noise caused by incomplete data observations, limitations of the underlying measurement devices, and subjectivity in data labeling. These unreliable data have a high degree of ignorance regarding the learning task in the target domain. Thus, knowledge transfer based on unreliable data instances from the source domain may have a negative impacts on TL, referred to as *negative transfer* (Zhang et al., 2023). Specifically, unreliable data in the source domain aggravate the discrepancy with the target domain and influence the feature transformation and the model training in TL (Raghu et al., 2019; Mustafa et al., 2021; Sariyildiz et al., 2023; Yosinski et al., 2014; Yang et al., 2023). For example, in the classification transferred from ImageNet to medical images, some images in ImageNet are incompatible with medical images, leading to negative transfer impacts in the medical image classification (Morid et al., 2020; Shang et al., 2019; Xie et al., 2021). Some researchers have used the transferred AdaBoost strategy to measure the importance of the source domain data instances for domain adaptation (Dai et al., 2007; Yao & Doretto, 2010; Li et al., 2022; Xu et al., 2020), however, integrating this strategy into other TL methods is difficult.

To address this problem, we propose a general strategy based on evidence theory (Dempster, 1967; Shafer, 1976; Denœux et al., 2020) for selecting reliable data instances from the source domain for transfer learning. This can be viewed as a pre-processing technique that can be easily integrated into the TL process. Figure 1 illustrates the strategy of selecting reliable data instances from the source domain in a transferred binary classification task. By comparing Figure 1(a) and Figure 1(b), we observe that by removing the outliers in the source domain, the classifier constructed on the source domain adapts better to the target domain.

As an effective tool for uncertainty analysis, it is natural to use evidence theory to measure the reliability of data instances in TL. Specifically, the mass function in evidence theory is adopted to represent the degree of ignorance and reliability of the source domain data with respect to the learning task in the target domain. By selecting reliable instances from the source domain, the adaptation of the transfer learning models to the target domain
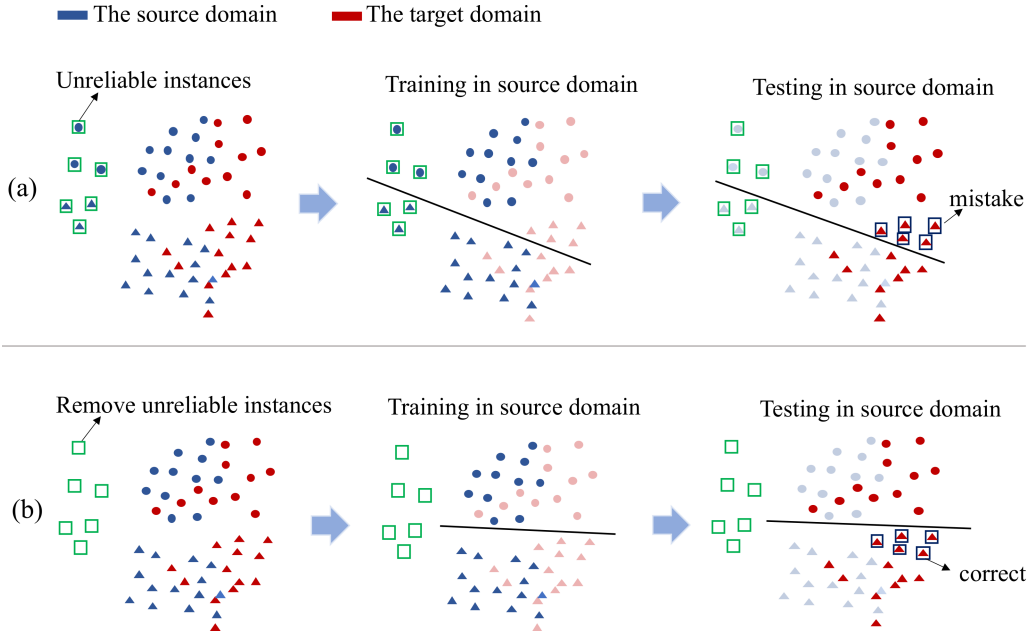
Figure 1: The domain adaption improvement of the transferred classification brought by removing the unreliable data instances from the source domain.

can be enhanced. The proposed instance selection strategy is independent of specific learning algorithms and can improve the performance of various types of transfer learning methods. The contributions of this study are summarized as follows.

- For each source domain data instance, we construct an evidence set according to its neighborhood in the target domain and the evidence set can be decomposed into subsets of classes and individual evidences.

- Through decomposing the evidence set, we formulate the mass function with multilevel evidences to measure the ignorance degree and reliability of source domain data with respect to the target domain.

- With the reliability measure, we implement an algorithm to select reliable data instances from source domain to improve the transferred classification.

The rest of the paper is organized as follows. We review the related work in Section 2. Section 3 introduces the strategy for selecting reliable source do-

main instances for transfer learning. Section 4 presents experimental results with transfer learning methods. The conclusion is given in Section 5.

2. Related Work

In this section, we briefly review the previous works on transfer learning and evidence theory that are related to our work.

2.1. Transfer Learning

According to literature surveys (Zhang & Gao, 2024; Zhuang et al., 2020), most TL methods can be roughly organized into instance-based method, feature-based method, classifier-based method, and deep learning-based method.

Most instance-based methods assign the weights to data instances of source domain in the cost function. The weights can be estimated by feature distribution matching across different domains. Jiang & Zhai (2007) proposed an intuitive instance weighted method, which calculate the distribution difference between source and target instances by four parameters. Dai et al. (2007) proposed a TrAdaBoost to tune instances weights based on Boosting algorithm. Huang et al. (2006) and Chu et al. (2013) utilize the kernel mean matching (KMM) to calculate the weighting for reducing the difference between source domain and target domain. Long et al. (2014) proposed the Transfer Joint Matching (TJM) method by minimizing the maximum mean discrepancy (MMD). Yan et al. (2017) proposed a weighted maximum mean discrepancy (WMMD) for transfer learning.

Feature-based methods transform each original feature into a new feature representation for transfer learning. The objective is to learn a new feature representation with some distribution matching metrics between source and target domains. Pan et al. (2010) firstly proposed a transfer component analysis (TCA) by introducing the marginal maximum mean discrepancy (MMD) with projection as the loss function. The joint distribution adaptation (JDA) proposed by Long et al. (2013b) further introduced the conditional MMD on the basis of TCA, such that the cross-domain distribution alignment becomes more discriminative. Gong et al. (2012) proposed a geodesic flow kernel (GFK), in which the geodesic flow kernel was used to model the domain shift by integrating an infinite number of subspaces. Sun et al. (2016) proposed a simple but efficient correlation alignment (CORAL) by aligning the second-order statistic (i.e. the covariance) between source and target distributions instead of the first-order metric.

Classifier-based methods focus on classifier adaptation that can learn a generic classifier by using labeled source domain and few labeled target domain. Yang et al. (2007) proposed an adaptive support vector machine(ASVM) for classification task of target. Duan et al. (2009) proposed a multiple kernel learning (MKL) for transfer learning. Liu et al. proposed a decision-level combination method for multi-source domain adaptation based on Dempster-Shafer theory. Long et al. (2013a) and Cao et al. (2018) proposed ARTL and DMM that used manifold regularization based structural risk and MMD minimization MMD between domians for classifier training. Wang et al. (2018) proposed a domain-invariant classifier MEDA in Grassmann manifold with structural risk minimization, while performing cross-domain distribution alignment of marginal and conditional distributions with different importances. Gholami & Pavlovic (2017) proposed a probabilistic latent variable model (PUnDA) for unsupervised domain adaptation, by simultaneously learning the classifier in a projected latent space and minimizing the MMD based domain disparity.

For the deep TL methods, fine-tuning pre-trained networks are common strategies (Bengio, 2012)(Donahue et al., 2014), which utilize a well-trained CNN on a large data set (e.g., ImageNet) as the base model and transfer it to the target domain to fine-tune weights. As recently reported in a review by Cheplygina et al. (2019), ImageNet data set is the most commonly used data set for TL-based medical image analysis. However, some studies (Raghu et al., 2019; Mustafa et al., 2021) have found that fine-tuning with the ImageNet data set does not significantly increase the performance of networks on medical tasks. Their some experiments suggest that domain mismatch between ImageNet data set and medical images inhibits domain adaptation.

The above-mentioned methods take into account shared feature representations, model parameters, and even relationships between tasks. However, they overlook the presence of unreliable instances in source domain data. If we can remove these unreliable data, it can further enhance the performance of transfer learning methods.

2.2. Learning with Evidence Theory

Evidence theory can be considered as a generalized probability (Dempster, 1967; Shafer, 1976; Denœux et al., 2020). It is based on the representation of elementary pieces of evidence by belief functions, and on their combination by a specific operator referred to as Dempster's rule. Based on this view, Den}oeux and his collaborators first combined evidence theory

with machine learning and design some supervised and unsupervised algorithms that can solve the problem of imprecise information to improve the robustness of algorithms. Such as, Evidential K-NN classification (Denœux, 1995; Denœux et al., 2019; Gong et al., 2023; Su et al., 2020), Evidential Linear Discriminant Analysis (Quost et al., 2017), Evidential logistic regression (Denœux, 2019, 2024), Evidential Neural Network Classifier (Tong et al., 2021; Zhou et al., 2023) and multiple evidential clustering algorithms (Campagner et al., 2023). In this section, we recall the definitions of mass function and Dempster's rule.

Let $\Omega = \{y_1, y_2, \ldots, y_c\}$ be a finite set that includes all possible answers in a decision problem and the elements of a set are mutually exclusive and exhaustive. In classification tasks, Ω is the label space. We denote the power-set as 2^Ω and the cardinality of power-set of Ω is 2^Ω. The mass function m is a mapping from 2^Ω to the interval [0,1] satisfying the following conditions,

$$\begin{cases} \sum_{A \in 2^\Omega} m(A) = 1 \\ m(\emptyset) = 0 \end{cases} \tag{1}$$

where $m(A)$ represents a share of a unit mass of belief allocated to subset $A \subseteq \Omega$, and which cannot be allocated to any strict subset of A. In classification problems, $m(y_k)$ can be interpreted as the degree of support that instance belongs to class y_k. $m(\Omega)$ can be interpreted as a degree of ignorance about the classification. For example, consider a classification problem in which the classes are colors. The label space is $\Omega = \{red, green, blue\}$. The power-set is $2^\Omega = \{\emptyset, \{red\}, \{green\}, \{bule\}, \{red, green\}, \{red, blue\}, \{green, blue\}, \Omega\}$ and $\Omega = 3$, $2^\Omega = 8$. The quantity $m(green|x; E)$ represents a possibility that instance x belongs to the green class based on evidence E, while $m(\Omega|x; E)$ represents the degree of uncertainty, which represents the probability that the color of instance x cannot be determined based on evidence E.

Dempster's rule is an operator that enables the aggregation of independent pieces of evidence. Suppose we have two mass functions, denoted as m_1 and m_2, which are derived from independent sources of evidence. These mass functions can be effectively combined using Dempster's rule to construct a new mass function, represented as $m_1 \oplus m_2$, defined as follows:

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - \kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \tag{2}$$

7

where $A \subseteq \Omega$ signifies a subset of the finite set of possible answers within Ω, with the stipulation that A is not an empty set. $(m_1 \oplus m_2)(\emptyset) = 0$, ensuring that an empty set receives zero belief. $\oplus$ represents the combination operator of Dempster's rule. κ quantifies the degree of conflict between m_1 and m_2, defined as:

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{3}$$

The κ value reflects the extent to which the two pieces of evidence, m_1 and m_2, are in disagreement or conflict with each other. It can also be thought of as a normalization factor.

To provide a practical illustration related to the previous discussion: Imagine we possess two pieces of evidence concerning the color of an object. Mass function m_1 could represent the belief degree based on one source of evidence, while m_2 might depict the belief degree stemming from another source of evidence. By applying Dempster's rule, we can combine these two belief functions to generate a new one, denoted as $m_1 \oplus m_2$, which conveys our updated belief regarding the object's color. The degree of conflict between these two evidence sources, represented by κ, quantifies the extent to which these sources are in disagreement. This combination of evidence empowers us to make more informed decisions or assessments in situations involving multiple sources of evidence.

3. Selecting Reliable Instances for Transfer Learning

This section begins by introducing the measurement of data reliability based on evidence theory and provides an example of its calculation process using binary classification. Secondly, a data selection strategy is constructed based on this reliability measurement approach. Finally, we demonstrate the effectiveness of the selection strategy.

3.1. Estimating Reliability of Instances Based on Evidence Theory

Evidence theory is a generalized form of Bayesian theory for describing subjective probabilities, and its fundamental probability assignment is completed by the mass function. The mass function $m(\cdot)$ evaluates elements of the power set of the discriminant framework Ω and has the ability to directly express ignorance and uncertainty. In the case of classification tasks, Ω represents the label space, and the mass function is used to assign membership possibilities to subsets of categories. Among these, the mass function $m(\Omega)$

8

assigned to the entire label space indicates that all category labels have the same likelihood, which can be considered as the uncertainty of data in the classification task.

If an instance x^s comes from the source domain D^s, Ω is the category space of the target domain, and the evidence set Φ is constructed from the target domain D^t, then $m(\Omega|x^s; \Phi)$ can be interpreted as the uncertainty of x^s regarding the classification results in the target domain. This uncertainty can measure the reliability of the source domain instance regarding the target domain task in the context of transfer learning. A larger $m(\Omega|x^s; \Phi)$ indicates lower uncertainty about the classification results of x^s in the target domain, indicating higher reliability. Based on this, evidence sets are established for each data instance in the source domain based on the labeled data from the target domain, and the mass function $m(\Omega|x; \Phi)$ is constructed to measure the reliability of source domain instances regarding the target domain task.

Given a source-domain instance x^s, its evidence set $\Phi \subset D^t$ consists of the similar instances from target domain and can be formulated as a neighborhood surrounding x^s. Suppose $\Phi = \{e_1, e_2, \cdots, e_l\}$, the elements in the evidence set are l target-domain instances similar to the source domain instance x^s. To ensure the validity of the evidence set, the discrepancy between a source-domain instance and the elements of its evidence set should be small. Motivated by this, we design the objective function of obtaining an evidence set for a source-domain instance x^s as

$$\Phi = \arg\min_{\Phi'} f(x^s, \Phi'), \tag{4}$$

in which function $f(\cdot)$ measures the discrepancy between x^s of source domain and the evidence set Φ in a reproducing kernel Hilbert Space (RKHS) $\mathcal{H}$. $f(\cdot)$ is defined as

$$f(x^s, \Phi) = \left\| \psi(x^s) - \frac{1}{|\Phi|} \sum_{e \in \Phi} \psi(e) \right\|_{\mathcal{H}}^2, \tag{5}$$

where $\psi : \mathcal{X} \mapsto \mathcal{H}$ is the feature mapping, $|\Phi|$ is the cardinality of evidence set Φ. In this paper, we use the radial basis function kernel $K(\cdot, \cdot)$ for feature mapping to construct the kernel Hilbert space,

$$K(e, x^s) = \psi(e)^T \psi(x^s) = \exp\left(-\gamma \|e - x^s\|^2\right), \tag{6}$$

where $\|e - x^s\|$ is the Euclidean distance between e and x^s and γ is a scaling

parameter. The function $f(\cdot)$ can be rewritten as

$$f\left(x^{s}, \Phi\right)=\left|\frac{1}{|\Phi|^{2}} \sum_{e_{i}, e_{j} \in \Phi} K\left(e_{i}, e_{j}\right)-\frac{2}{|\Phi|} \sum_{e \in \Phi} K\left(x^{s}, e\right)\right|. \qquad (7)$$

The objective function of Equation (4) to obtain the evidence set can be specified as

$$\Phi=\underset{\Phi}{\operatorname{argmin}}\left|\frac{1}{|\Phi|^{2}} \sum_{e_{i}, e_{j} \in \Phi} K\left(e_{i}, e_{j}\right)-\frac{2}{|\Phi|} \sum_{e \in \Phi} K\left(x^{s}, e\right)\right|. \qquad (8)$$

The optimal evidence set Φ in Equation (8) can be solved by greedy search in the labeled target domain.

For a source-domain instance x^{s}, based on its evidence set Φ from target domain, we can further decompose Φ and refine the mass function to implement the transferred classification of x^{s}. Given c classes, we decompose the evidence set Φ into different classes,

$$\Phi=\left\{\Phi_{1}, \ldots \Phi_{k} \ldots, \Phi_{c}\right\}, \qquad (9)$$

$\Phi_{k}=\left\{e_{k1}, \ldots e_{kl}\right\}$ is the evidence subset in which all the target domain instances have the class label y_{k}, e_{kl} is the lth element in the evidence subset.
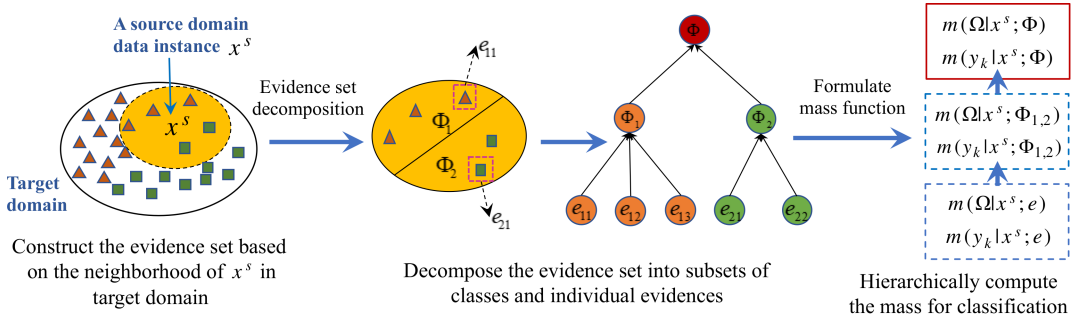

Figure 2: Formulating mass function based on evidence set

In Figure 2, the decomposition of the evidence set Φ is presented by a tree, the middle nodes of tree represent the evidence subsets of classes and the

leaf nodes denote the elements in the evidence subset. Through decomposing the evidence set, we can adopt Dempster's rule to refine the mass function $m(\cdot|x; \Phi)$ with multilevel evidences for the transferred classification as

$$m(\cdot|x; \Phi) = \bigoplus_{\Phi_k \subseteq \Phi} m\left(\cdot \mid x^s; \Phi_k\right) = \bigoplus_{\Phi_k \subseteq \Phi} \left(\bigoplus_{e \subseteq \Phi_k} m\left(\cdot \mid x^s; e\right)\right). \quad (10)$$

In the mass function, we focus on the mass of x^s belonging to a class or ignorance, which can be hierarchically calculated based on evidence set decomposition. At the fine-grained evidence level, given an element e in a evidence subset, we compute the mass of x^s belonging to the class y_k and its ignorance mass as

$$m\left(y_k|x^s; e\right) = \exp\left(-d\left(x^s, e\right)\right), \quad (11)$$
$$m\left(\Omega|x^s; e\right) = 1 - \exp\left(-d\left(x^s, e\right)\right), \quad (12)$$

where $m\left(\Omega|x^s; e\right)$ reflects the ignorance degree of x^s according to the evidence e, $m\left(y_k|x^s; e\right)$ represents the possibility that x^s belongs to class y_k, $d(\cdot)$ is a distance metric that is defined by

$$d\left(x^s, e\right) = K\left(x^s, x^s\right) - 2K\left(x^s, e\right) + K\left(e, e\right). \quad (13)$$

Using Dempster's rule to combine the masses under single evidence $e \in \Phi_k$, we can obtain the masses of x^s under the evidence subset Φ_k.

$$m\left(y_k|x^s; \Phi_k\right) = \bigoplus_{e \in \Phi_k} m\left(y_k|x^s; e\right) = 1 - \prod_{e \in \Phi_k} m\left(\Omega|x^s; e\right), \quad (14)$$
$$m\left(\Omega|x^s; \Phi_k\right) = \bigoplus_{e \in \Phi_k} m\left(\Omega|x^s; e\right) = \prod_{e \in \Phi_k} m\left(\Omega|x^s; e\right). \quad (15)$$

At the top evidence level, we calculate the masses of x^s under the entire evidence set Φ through accumulating the masses under evidence subsets,

$$m\left(\Omega|x^s; \Phi\right) = \bigoplus_{\Phi_k \subseteq \Phi} m\left(\Omega|x^s; \Phi_k\right) = \frac{1}{\kappa}\prod_{k=1}^{n} m\left(\Omega|x^s; \Phi_k\right), \quad (16)$$
$$m\left(y_k|x^s; \Phi\right) = \bigoplus_{\Phi_k \subseteq \Phi} m\left(y_k|x^s; \Phi_k\right) = \frac{1}{\kappa}m\left(y_k|x^s; \Phi_k\right)\prod_{j \neq k} m\left(\Omega|x^s; \Phi_j\right), \quad (17)$$

11

where κ is a normalizing factor,

$$\kappa = \sum_{k=1}^{n}(m\left(y_k|x^s;\Phi_k\right)\prod_{j\neq k}m\left(\Omega|x^s;\Phi_k\right)) + \prod_{k=1}^{n}m\left(\Omega|x^s;\Phi_k\right). \qquad (18)$$

Based on the above calculations, the mass function $m(\Omega|x^s,\Phi)$ is obtained, which represents the reliability of the source domain instance x^s with respect to the target domain task under the support of evidence set Φ.

As an example, let's provide a detailed explanation of the design and calculation process of the mass function for cross-domain binary classification tasks. Suppose the category label space is $\Omega = \{1,2\}$, and the evidence set Φ for the source domain instance x^s has been constructed based on the target domain. Φ is defined as Φ_1, Φ_2,

$$\begin{aligned} \Phi_1 &= \{(e_{11}, y_{11} = 1), (e_{12}, y_{12} = 1), (e_{13}, y_{13} = 1)\}, \\ \Phi_2 &= \{(e_{21}, y_{21} = 2), (e_{22}, y_{22} = 2)\}. \end{aligned} \qquad (19)$$

We first calculate the masses of x^s under the fine-grained evidence according to Equations 11 and 12,

$$\begin{aligned} m\left(y_k|x^s;e_{kj}\right) &= \alpha_{kj}, \\ m\left(\Omega|x^s;e_{kj}\right) &= 1 - \alpha_{kj}, \end{aligned} \qquad (20)$$

The computation of the mass function is shown in Table 1. Through combining the masses under fine-grained evidences, we can obtain the masses of x^s under the evidence subsets,

$$m\left(y_k|x^s;\Phi_k\right) = \bigoplus_{x\in\Phi_k} m\left(y_k|x^s;e\right) = 1 - \prod_{j=1}^{|\Phi_k|}\left(1 - \alpha_{kj}\right), \qquad (21)$$

$$m\left(\Omega|x^s;\Phi_k\right) = \bigoplus_{x\in\Phi_k} m\left(\Omega|x^s;e\right) = \prod_{j=1}^{|\Phi_k|}\left(1 - \alpha_{kj}\right). \qquad (22)$$

Table 2 presents the computation of the mass function under the evidence subsets Φ_1, Φ_2 .

Finally, we calculate the mass function $m(\cdot|x^s;\Phi)$ under the entire evidence set as

$$\begin{aligned} m(y = 1|x^s;\Phi) &= (1 - A_1)^{-1} * A_2, \\ m(y = 2|x^s;\Phi) &= (1 - A_1)^{-1} * A_3, \\ m(\Omega|x^s;\Phi) &= (1 - A_1)^{-1} * A_4, \end{aligned} \qquad (23)$$

12

Table 1: Computation of mass function under individual evidences.

Evidences	$m\left(y_k\middle|x^s; e_{kj}\right)$	$m\left(\Omega\middle|x^s; e_{kj}\right)$
e_{11}	α_{11}	$1 - \alpha_{11}$
e_{12}	α_{12}	$1 - \alpha_{12}$
e_{13}	α_{13}	$1 - \alpha_{13}$
e_{21}	α_{21}	$1 - \alpha_{21}$
e_{22}	α_{22}	$1 - \alpha_{22}$

Table 2: Computation of mass function under evidence subsets.

Evidence subsets	$m\left(y_k\middle|x^s; \Phi_k\right)$	$m\left(\Omega\middle|x^s; \Phi_k\right)$
Φ_1	$1 - \prod_{j=1}^{\lvert\Phi_1\rvert}\left(1 - \alpha_{1j}\right)$	$\prod_{j=1}^{\lvert\Phi_1\rvert}\left(1 - \alpha_{1j}\right)$
Φ_2	$1 - \prod_{j=1}^{\lvert\Phi_2\rvert}\left(1 - \alpha_{2j}\right)$	$\prod_{j=1}^{\lvert\Phi_2\rvert}\left(1 - \alpha_{2j}\right)$

in which

$$
\begin{aligned}
A_1 &= m(y = 1|x^s; \Phi_1) * m(y = 2|x^s; \Phi_2), \\
A_2 &= m(y = 1|x^s; \Phi_1) * m(\Omega|x^s; \Phi_2), \\
A_3 &= m(\Omega|x^s; \Phi_1) * m(y = 2|x^s; \Phi_2), \\
A_4 &= m(\Omega|x^s; \Phi_1) * m(\Omega|x^s; \Phi_2).
\end{aligned}
\tag{24}
$$

We have

$$
m(\cdot|x^s; \Phi) = \begin{cases} m(y = 1|x^s; \Phi), \\ m(y = 2|x^s; \Phi), \\ m(\Omega|x^s; \Phi). \end{cases}
\tag{25}
$$

So far, a detailed explanation has been provided for the design and calculation process of the mass function to measure the reliability of transferred data.

3.2. Constructing the Reliable Instance Selection Strategy

Our objective is to select reliable instances from the source domain to improve the transfer learning performance. Suppose D^s is the original source

domain, D^r is the filtered source domain, D^t is the target domain containing limited labeled data and Φ is the evidence set constructed on the target domain. We design the mass function $m(\Omega|x^s; \Phi)$ to measure the instance reliability and thereby define the filtered source domain D^r that consists of the selected reliable instances with low ignorance as

$$D^r = \{x^s \in D^s | m(\Omega|x^s; \Phi) \leq \beta\}, \tag{26}$$

where Φ is the evidence set of the labeled data instances in the target domain surrounding x^s, $\beta \in [0, 1]$ is the threshold parameter.

Figure 3 illustrates the framework and Algorithm 1 lists the main steps of selecting reliable instances from the source domain.
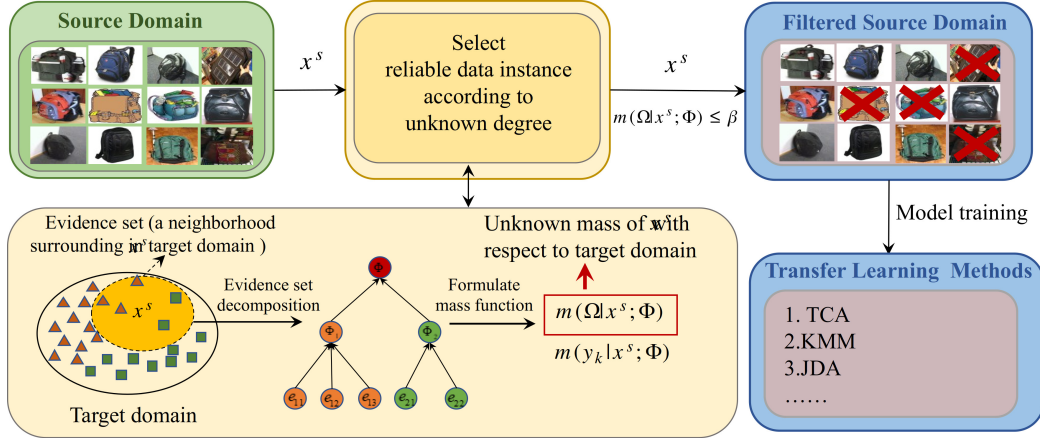
Figure 3: Framework of selecting reliable data instances from source domain

---
Algorithm 1 Selecting Reliable Instances from Source Domain

Input: Source domain D^s, target domain D^t with limited labeled data;
Output: Filtered source domain D^r;

1: **for** each x^s in D^s **do**
2:    Construct an evidence set Φ for x^s based on its neighborhood of the labeled data instances in target domain D^t;
3:    Decompose the evidence set Φ and formulate the mass function $m(\cdot|x^s; \Phi)$ with multilevel evidences;
4:    Hierarchically compute the ignorance mass $m(\Omega|x^s; \Phi)$ to measure the reliability of x^s for transfer learning;
5:    **if** $m(\Omega|x^s; \Phi) \leq \beta$ **then**
6:       Select the instance x^s and add x^s into D^r;
7:    **end if**
8: **end for**
9: **return** D^r.

The core idea of Algorithm 1 is to assess the reliability of each source domain instance in transfer learning by calculating the uncertainty measure $m(\Omega|x^s; \Phi)$ and selecting those instances with low uncertainty. This process forms the filtered source domain D^r, aiming to enhance the performance of transfer learning.

Furthermore, we further discuss the computational complexity of the method as follows:

- Complexity of constructing the evidence set Φ: Constructing the evidence set involves considering the neighborhood relationships between source domain instances x^s and labeled data instances in the target domain. This entails distance calculations or similarity measurements between data points, and thus, its computational complexity may depend on the size and dimensionality of the dataset.

- Calculation of the ignorance mass $m(\Omega|x^s; \Phi)$: Formulating the mass function involves considering multi-level evidence extraction and hierarchical computations of the mass function. This could potentially lead to an increase in computational complexity, especially when dealing with large-scale datasets. In summary, the proposed method may exhibit significant computational complexity, particularly when handling large-scale datasets. Therefore, in practical applications, consideration

of computational resources and efficiency is essential to ensure the feasibility of the method.

3.3. Proof of the Effectiveness of the Selection Strategy

The goal of transfer learning is to minimize the differences between the source and target domains as much as possible, allowing the model to adapt better to the data in the target domain, thereby enhancing transfer performance. To elucidate how the selection strategy can enhance the transferability of the source domain data and improve the domain adaptability of transfer learning models, maximum mean discrepancy (MMD) is introduced in this section to demonstrate that the selected source domain data exhibit smaller differences from those in the target domain. This proof demonstrates the effectiveness of the selection strategy.

Let D^s denote the original source domain, and D^r represent the selected source domain. Based on MMD, the distances from the original source domain and the selected source domain to the target domain, denoted as d_1 and d_2 respectively, are defined as

$$
\begin{aligned}
d_1 &= \left\| \frac{1}{n_s} \sum_{x^s \in D^s} \varphi\left(x^s\right) - \frac{1}{n_t} \sum_{x^t \in D^t} \varphi\left(x^t\right) \right\|_H^2, \\
d_2 &= \left\| \frac{1}{r_s} \sum_{x^s \in D^r} \varphi\left(x^s\right) - \frac{1}{n_t} \sum_{x^t \in D^t} \varphi\left(x^t\right) \right\|_H^2.
\end{aligned}
\tag{27}
$$

Here, $\varphi(\cdot)$ is the feature mapping function, and n_s, r_s, n_t represent the numbers of instances in the source domain, selected source domain, and target domain respectively. By proving that $d_1 \geq d_2$, we establish that the selection strategy reduces domain differences, thereby enhancing the credibility of the source domain data.

Assuming that the evidence set Φ contains only a single element $e \in D^t$, the instances in the selected source domain D^r satisfy the condition

$$
m\left(\Omega \mid x^s \in D^r; \Phi\right) = 1 - \exp\left(\varphi\left(x^s\right)\varphi(e)\right) \leq \beta.
\tag{28}
$$

With some inequalities transformation, we get

$$
\varphi\left(x^s\right)\varphi(e) \leq -\ln(1 - \beta).
\tag{29}
$$

16

Based on the description in Section 3.1, where e is a target domain instance x^t, it follows that selected source domain instance x^s and target domain instance x^t satisfy

$$\varphi\left(x^s\right)\varphi\left(x^t\right) \leq -\ln(1-\beta). \tag{30}$$

Let $s(x^s, x^t) = \varphi(x^s)\varphi(x^t)$, and expand both d_1 and d_2, resulting in

$$d_1 = \frac{1}{n_s\left(n_s-1\right)}\sum_{i\neq j}^{n_s} s\left(x_i^s, x_j^s\right) + \frac{1}{n_t\left(n_t-1\right)}\sum_{i\neq j}^{n_t} s\left(x_i^t, x_j^t\right) - \frac{2}{n_s n_t}\sum_{i,j=1}^{n_s,n_t} s\left(x_i^s, x_j^t\right),$$

$$d_2 = \frac{1}{r_s\left(r_s-1\right)}\sum_{i\neq j}^{r} s\left(x_i^s, x_j^s\right) + \frac{1}{n_t\left(n_t-1\right)}\sum_{i\neq j}^{n_t} s\left(x_i^t, x_j^t\right) - \frac{2}{r_s n_t}\sum_{i,j=1}^{r_s,n_t} s\left(x_i^s, x_j^t\right). \tag{31}$$

Within Equation (31), parts (1) and (2) satisfy

$$(1) \leq \frac{2}{n_s n_t}\left(-r_s \ln(1-\beta) + \sum_{x^s \notin D^r} s\left(x^s, x^t\right)\right),$$

$$(2) \leq -\frac{2}{r_s n_t} r_s \ln(1-\beta). \tag{32}$$

By padding D^r with zeros to ensure $r = n$, the inequality $d_1 - d_2 \geq 0$ is derived.

The above analysis demonstrates that $d_1 \geq d_2$, indicating that the selected source domain exhibits smaller differences from the target domain. Therefore, the selection strategy is effective in improving the quality of source domain data, enhancing data transferability.

4. Experiments

In the experiments, to evaluate the proposed instance selection strategy for transfer learning, we perform our method to implement cross-domain classification on various kinds of data including texts and images. The descriptions of the data sets and the experimental setting are listed below.

Text data The cross-domain text data are generated from the data set of Amazon product reviews, which is the benchmark text corpora widely used for transfer learning evaluation (Blitzer et al., 2007). The reviews are about four product domains: books (denoted by domain B), dvds

17

(domain D), electronics (domain E) and kitchen appliances (domain K). Each review is assigned to a sentiment label of -1 (negative review) or +1 (positive review) based on the rating score given by the review author. In each domain, there are 1,000 positive reviews and 1,000 negative ones. For this data set, we can construct 12 cross-domain sentiment classification tasks: $B \to D$, $B \to E$, $B \to K$, $D \to B$, $D \to E$, $D \to K$, $E \to B$, $E \to D$, $E \to K$, $K \to B$, $K \to D$, $B \to E$, where the capital letter before the arrow denotes the source domain and the letter after the arrow denotes the target domain.

Image data The cross-domain image data are generated from the data sets Office and Caltech-256, which are widely adopted for visual domain adaptation evaluation. Office data set consists of 4563 images with 31 categories and Caltech-256 data set consists of 30607 images with 256 categories. In the experiments, we mix Office and Caltech data sets as in (Gong et al., 2012) for cross-domain visual object recognition. It includes four domains: Amazon (denoted by domain A, images downloaded from online merchants), Webcam (domain W, low-resolution images by a web camera), DSLR (domain D , high-resolution images by a digital SLR camera), and Caltech-256 (domain C). The data set includes 10 classes: BACKPACK, TOURING-BIKE, CALCULATOR, HEAD-Caltech, PHONES, COMPUTER-KEYBOARD, LAPTOP-COMPUTER, COMPUTER-MONITOR, COMPUTER-MOUSE, COFFEE-MUG and VIDEO PROJECTOR. There are 8 to 151 images in each category for one domain. For this data set, we can construct 9 cross-domain classification tasks: $A \to C$, $A \to W$, $C \to A$, $C \to W$, $D \to A$, $D \to C$, $D \to W$, $W \to A$, $W \to C$.

Medical image data The data contain three domains: X-Ray image data set (denoted by domain X) (Kermany et al., 2018), ultrasound image data set (domain U) (Al-Dhabyani et al., 2020), tomography image data set (domain T) (Kermany et al., 2018). X-Ray image data set contains 5856 X-ray images including 3883 pneumonia cases and 1349 normal cases. Ultrasound image data set contains 780 breast ultrasound images of three classes (normal, benign, and malignant). Tomography image data set contains 109309 optical coherence tomography (OCT) images for retinal diseases (37206 with choroidal neovascularization, 11349 with diabetic macular edema, 8617 with drusen, and 51140 nor-

mal). Adopting ImageNet data set (Deng et al., 2009) as the source domain, we construct three transferred classification tasks from ImageNet to the medical image data sets: $ImageNet \rightarrow T$, $ImageNet \rightarrow X$ and $ImageNet \rightarrow U$.

In the experiment, for each cross-domain classification task, we adopt the source domain data set to train the classifier and select 10% of the data in target domain to construct the evidence set to measure the reliability of data instances in transfer learning. We use the classification accuracy of the target domain data instances as the evaluation criterion. Suppose $\mathcal{X}$ is the target domain data set,

$$\text{Accuracy} = \frac{|\{x : x \in \mathcal{X} \wedge g(x) = y\}|}{|\{x : x \in \mathcal{X}\}|}, \tag{33}$$

where y is the ground truth label of x, $g(x)$ is the label predicted by the classifier.

4.1. Test of Instance Selection Strategy

In order to verify the effectiveness of the proposed instance selection strategy in transfer learning, we visualize the selected reliable instances in the transferred classification on a synthetic data set. The data set contains two domains, in which the source domain and target domain consist of two dimensional data points of two classes and each class has 500 data points. The data points in each class are generated from a Gaussian distribution $x \sim \mathcal{N}(\mu, \sigma)$. Changing the mean and deviation μ, σ of Gaussian distribution, we generate the data sets of source domain and target domain. As shown in Figure 4, the source domain is marked by green color and target domain is marked by yellow color. In each domain, the data points of class 1 are marked by circle and the points of class 2 are marked by triangle.
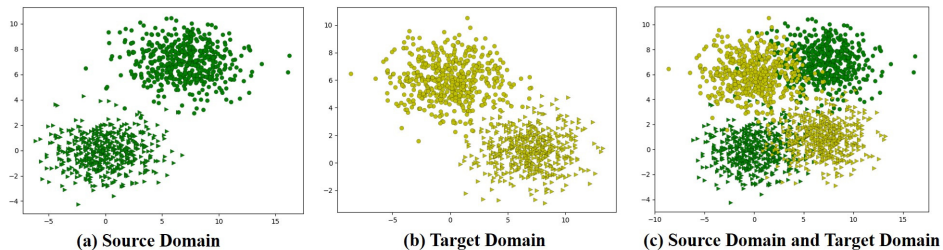
(a) Source Domain **(b) Target Domain** **(c) Source Domain and Target Domain**

Figure 4: Synthetic data set
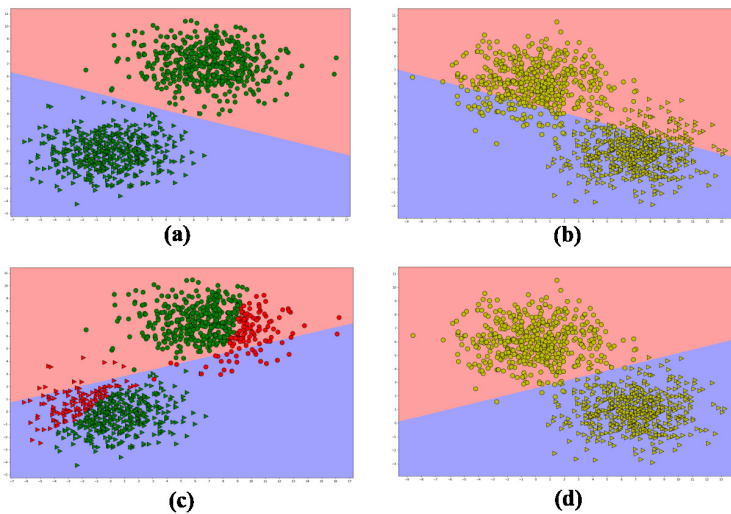

(a) **(b)**

(c) **(d)**

Figure 5: (a) Classification hyperplane of f_{D^s} on the original source domain data. (b) classification hyperplane of f_{D^s} on the target domain data. (c) classification hyperplane of f_{D^r} generated by the reliable source domain data. (d) classification hyperplane of f_{D^r} on the target domain data.

We first train a linear classifier f_{D^s} based on the original source domain data set D^s. Figure 5(a) shows the classification hyperplane on the source domain data. Using f_{D^s} to classify the data in target domain, because the data distribution of target domain is different from the source domain, the classifier cannot accurately distinguish the data and the accuracy is 78.10%. As shows in Figure 5(b), we can see that the hyperplane is not suitable to separate the data of different classes in target domain.

In contrast, using the mass function to measure the reliability of data instances in the source domain, we can select the reliable data instances

for transfer learning. As shown in Figure 5(c), the data instances with low reliability are marked by red color. Removing the unreliable instances, we generate the filtered source domain data set D^r and obtain the classifier f_{D^r}. Using the classifier $f_{D^r}(\cdot)$ to classify the data in target domain, as shown in Figure 5(d), we can see that the classification hyperplane can accurately separate different classes of target domain and the accuracy is improved by 97.9%. Comparing Figure 5(b) with Figure 5(d), we find that the strategy of selecting reliable data instances in source domain is helpful to adjust the classifier built on source domain to suit the target domain data.

4.2. Comparative Studies

In the second experiment, to further validate our method, we integrate the proposed reliable instance selection strategy into the typical transfer learning methods and compare the classification results with and without the instance selection. The comparative transfer learning methods include: Kernel Mean Matching (KMM) (Huang et al., 2006), Transfer Component Anaysis (TCA) (Pan et al., 2010), Geodesic Flow Kernel (GFK) (Gong et al., 2012), Joint Distribution Adaptation (JDA)(Long et al., 2013b), Correlation ALignment (CORAL) (Sun et al., 2016), Scatter Component Analysis (SCA) (Ghifary et al., 2017), Balanced Distribution Adaptation (BDA) (Wang et al., 2017), Manifold Embedded Distribution Alignment (MEDA) (Wang et al., 2018), Practically Easy Transfer Learning (EasyTL) (Wang et al., 2019), Manifold Dynamic Distribution Adaptation (MDDA)(Zhu et al., 2021), Discriminative Manifold Propagation (DMP) (Luo et al., 2020), Conditional Kernel Bures (CKB) (Luo & Ren, 2021). We integrate the selection strategy into all the transfer learning methods and evaluate their classification performances on the Amazon product reviews and Office+Caltech data sets. In addition, to verify that our method is effective for the fine-tuning pre-trained neural network, we integrate the selection strategy as a preprocessing technique into five deep convolutional neural networks: MNASNet (Tan et al., 2019) , MobileNet (Howard et al., 2017), ShuffleNet (Zhang et al., 2018), SqueezeNet (Iandola et al., 2016) and DenseNet121 (Huang et al., 2017) and evaluate their performances in the transferred classification from ImageNet data set to the medical image data set. For any transfer learning method *, we briefly denote the method integrated with reliable instance selection strategy as $F - *$.
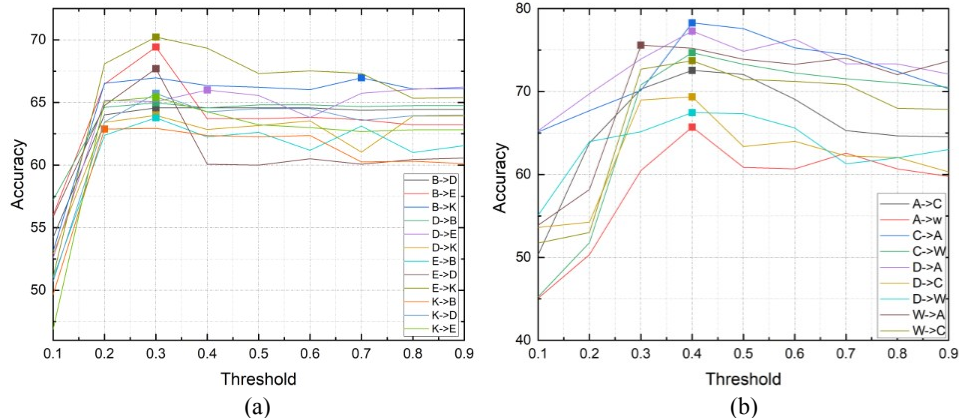
Figure 6: (a) Accuracies of cross-domain sentiment classification on Amazon product reviews with varying threshold β. (b) Cross-domain classification accuracies on Office+Caltech image data with varying threshold β.

4.2.1. Parameter Setting

In our experiments, we empirically set the scaling parameter γ in radial basis function as 0.7. The threshold parameter $\beta \in [0,1]$ in Equation (26) is used to select reliable data instances. A small value of β can ensure that the selected source domain data are reliable for target domain but may filter out too much source domain data that are useful in learning task. To determine the optimal threshold β, we performed the data selection method against varying β values on the text and image data sets.

We constructed 12 cross-domain sentiment classification tasks on the text data of Amazon product reviews and 9 cross-domain visual object classification tasks on the image data. We considered the values of β from 0.1 to 0.9 with step length 0.1. As shown in Figure 6(a) and Table 3, we can see that $\beta = 0.3$ produces the most accurate classification results on Amazon product reviews, thus we set $\beta = 0.3$ as default on the text data. Similarly, Figure 6(b) and Table 4 indicate that $\beta = 0.4$ leads to the best classification performance. Thus we set $\beta = 0.4$ on the image data.

4.2.2. Test of Text Classification

As introduced above, we constructed 12 cross-domain sentiment classification tasks on the text data of Amazon product reviews, i.e., $B \rightarrow D$,

22

Table 3: Accuracies of cross-domain sentiment classification on Amazon product reviews with varying threshold β.

$Task\beta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$B \to D$	54.22	64.01	**64.57**	64.57	64.57	64.57	64.38	64.44	64.44
$B \to E$	56.01	66.48	**69.43**	63.7	63.7	63.82	63.63	63.21	63.21
$B \to K$	53.12	66.53	**66.97**	66.35	66.22	66.03	66.97	66.09	66.09
$D \to B$	57.22	64.62	**64.98**	64.57	64.81	64.81	64.68	64.75	64.75
$D \to E$	52.57	65.19	65.07	66.00	65.56	63.81	65.74	66.06	**66.24**
$D \to K$	52.9	63.42	**63.97**	62.85	63.16	63.54	61.03	63.91	63.91
$E \to B$	50.98	62.38	**63.78**	62.25	62.62	61.18	63.12	61.00	61.56
$E \to D$	55.85	64.77	**67.7**	60.07	60.00	60.51	60.07	60.44	60.57
$E \to K$	51.10	68.10	**70.23**	69.34	67.31	67.52	67.33	65.33	65.40
$K \to B$	49.66	62.88	**62.94**	62.37	62.25	62.37	60.25	60.31	60.12
$K \to D$	50.71	63.45	**65.72**	64.25	64.51	64.51	63.57	63.94	64.00
$K \to E$	46.88	65.1	**65.43**	64.25	63.22	62.99	62.68	62.81	62.81
Average	52.60	64.74	**65.89**	64.21	63.99	63.80	63.62	63.52	63.59

Table 4: Cross-domain classification accuracies on Office+Caltech image data with varying threshold β.

$Task/\beta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$A \to C$	50.30	63.80	70.29	**72.55**	72.06	69.07	65.30	64.66	64.55
$A \to W$	45.07	50.33	60.46	**65.72**	60.85	60.68	62.58	60.67	59.77
$C \to A$	65.14	67.69	70.13	**78.28**	77.58	75.28	74.43	72.43	70.33
$C \to W$	45.25	51.77	70.81	**74.68**	73.29	72.25	71.55	71.03	70.52
$D \to A$	65.28	69.73	73.92	**77.28**	74.84	76.32	73.35	73.33	72.10
$D \to C$	53.62	54.25	68.99	**69.36**	63.37	64.01	62.23	62.05	60.31
$D \to W$	55.08	63.95	65.15	**67.46**	67.33	65.6	61.29	62.02	63.01
$W \to A$	53.89	58.19	**75.60**	75.22	73.89	73.29	74.04	72.04	73.67
$W \to C$	51.74	53.00	72.7	**73.72**	71.48	71.23	70.85	67.98	67.85
Average	53.93	59.19	69.78	**72.70**	70.52	69.74	68.40	67.35	66.90

$B \rightarrow E$, $B \rightarrow K$, $D \rightarrow B$, $D \rightarrow E$, $D \rightarrow K$, $E \rightarrow B$, $E \rightarrow D$, $E \rightarrow K$, $K \rightarrow B$, $K \rightarrow D$, $B \rightarrow E$. In each cross-domain classification task, we used the word2vec tool to extract the features of the review texts and performed the transfer learning methods with and without reliable instance selection to generate the sentiment classification results respectively. Logistic regression was used in this experiment. The classification accuracies of the comparative study are listed in Table 5.

It is obvious that in all the cross-domain text classification tasks, selecting the reliable instances from source domain can improve the classification accuracy of every adopted transfer learning method. Specifically, for 9 comparative transfer learning methods, using the instance selection strategy improves the accuracies of the cross-domain sentiment classifications by 3.74%, 2.7%, 3.51%, 3.44%, 6.19%, 4.11%, 1.94%, and 2.85%, 2.50%, respectively. Moreover, we find that the transfer learning method KMM achieves the largest performance improvement 6.19% by using the reliable instance selection strategy. This is because that the KMM method is sensitive to domain difference and the selection strategy removes the unreliable source domain instances that are inconsistent with the target domain data. The transfer learning methods TCA, JDA, and BDA also achieved good performance improvements by using the instance selection strategy. All these methods use the maximum mean discrepancy (MMD) as the distance metric to minimize the domain difference. Selecting reliable instances from source domain is helpful to further decrease the domain gap.

4.2.3. Test of Image Classification

In this test, we further validate the proposed instance selection strategy on Office+Caltech image data sets. We constructed 9 cross-domain visual object classification tasks on the image data, i.e., $A \rightarrow C$, $A \rightarrow W$, $C \rightarrow A$, $C \rightarrow W$, $D \rightarrow A$, $D \rightarrow C$, $D \rightarrow W$, $W \rightarrow A$, $W \rightarrow C$. Because there are a few images in the data set DSLR (domain D), we do not construct any cross-domain classification tasks in which D is set as the target domain. We utilize deep convolutional activation features (DeCAF6 features) (Donahue et al., 2014) to represent all the images, in which the outputs from the 6th layer in the deep convolutional neural network are transformed to 4,096 dimensional features. In each cross-domain classification task, we applied the transfer learning methods with and without reliable instance selection to generate the image classification results respectively. Multinomial logistic regression was used in this experiment. The classification accuracies of the comparative

Table 5: Cross-domain sentiment classification accuracies of Amazon product reviews generated by the transfer learning methods with and without reliable instance selection.

Methods	$B \to D$	$B \to E$	$B \to K$	$D \to B$	$D \to E$	$D \to K$	$E \to B$	$E \to D$	$E \to K$	$K \to B$	$K \to D$	$K \to E$	Ave acc
TCA	77.76	75.54	78.74	76.05	76.38	79.34	73.35	73.66	79.74	73.05	77.26	78.74	76.63
F-TCA	**80.31**	**82.01**	**84.43**	**80.65**	**81.37**	**83.73**	**75.75**	**76.91**	**82.09**	**77.6**	**78.66**	**80.87**	**80.37**
CORAL	70.76	66.21	70.00	73.05	68.7	71.96	69.9	65.71	72.35	67.45	68.61	75.68	70.03
F-CORAL	**74.31**	**71.80**	**72.75**	**75.55**	**72.10**	**73.05**	**73.75**	**66.97**	**74.06**	**72.05**	**69.01**	**77.41**	**72.73**
GFK	75.76	72.00	73.50	71.85	68.96	75.70	72.60	71.11	76.20	73.75	74.21	76.58	73.52
F-GFK	**77.96**	**77.13**	**77.79**	**79.85**	**75.88**	**78.99**	**77.00**	**73.36**	**78.29**	**74.95**	**74.56**	**78.57**	**77.03**
JDA	77.26	75.93	78.09	77.65	76.03	78.29	72.65	72.16	80.14	75.05	77.56	80.32	76.76
F-JDA	**80.06**	**81.54**	**84.13**	**80.15**	**81.64**	**82.39**	**76.70**	**75.26**	**82.34**	**78.20**	**78.56**	**81.42**	**80.20**
KMM	83.76	79.02	75.9	80.5	68.51	76.45	73.7	77.86	80.39	74.25	75.96	85.00	77.61
F-KMM	**85.01**	**81.96**	**85.73**	**85.20**	**82.90**	**85.78**	**77.75**	**83.11**	**88.52**	**80.75**	**83.86**	**85.00**	**83.80**
BDA	75.01	73.04	78.09	74.55	71.8	75.3	71.40	71.61	78.49	71.15	71.66	76.93	73.89
F-BDA	**79.01**	**77.83**	**81.94**	**79.15**	**78.28**	**77.89**	**75.25**	**74.21**	**81.94**	**76.30**	**74.36**	**79.42**	**78.00**
SCA	79.41	78.82	78.14	74.95	77.53	76.00	73.15	71.86	79.79	73.70	75.71	82.41	76.79
F-SCA	**83.26**	**79.68**	**78.84**	**81.65**	**77.63**	**76.00**	**75.45**	**74.76**	**81.63**	**74.60**	**78.46**	**82.81**	**78.73**
EasyTL	76.76	76.08	82.14	83.9	80.91	85.22	76.4	72.61	86.37	73.25	70.86	75.93	78.37
F-EasyTL	**79.31**	**83.61**	**83.23**	**84.30**	**85.15**	**85.58**	**75.35**	**77.71**	**86.78**	**77.10**	**77.41**	**78.13**	**81.22**
MDDA	77.80	80.00	79.90	74.70	80.40	81.00	63.80	62.50	84.40	63.50	72.20	80.70	75.10
F-MDDA	**79.50**	**82.20**	**81.00**	**78.60**	**81.20**	**82.40**	**67.90**	**67.10**	**85.50**	**67.40**	**75.70**	**82.40**	**77.60**

study are listed in Table 6.

As shown in Table 6, the average classification accuracies of the transfer learning methods with the instance selection strategy on 9 tasks are 86.56%, 84.14%, 84.84%, 85.88%, 81.79%, 86.37%, 91.22%, 81.89%, 90.90% and 91.60%, respectively. Comparing with the original transfer learning methods without instance selection, selecting reliable data instances from the source domain improves the accuracies of different cross-domain classification tasks by 5.34%, 1.52%, 4.51%, 4.84%, 1.45%, 2.27%, 1.41% , 2.76%, 0.80% and 1.30%. As the experiments on the review text data, the instance selection strategy achieves the good classification improvements on the transfer learning methods of TCA, JDA, and BDA in which the MMD metric is adopted to minimize the differences between source domain and target domain. This indicates that when the source domain data are sufficient, selecting reliable data instances from the source domain facilitates to reduce the domain differences in cross-domain image classification tasks.

4.2.4. Test of Medical Image Classification

Besides the traditional transfer learning methods, we also verified that the proposed reliable instance selection strategy is effective to improve the cross-domain pre-trained deep neural networks for medical image classification. We pre-trained five deep convolutional neural networks including MNASNet (Tan et al., 2019) , MobileNet (Howard et al., 2017), ShuffleNet (Zhang et al., 2018), SqueezeNet (Iandola et al., 2016) and DenseNet121 (Huang et al., 2017) on ImageNet data and transfered these deep neural networks to the medical image data sets for classification. As introduced above, we constructed three cross-domain medical image classification tasks: $ImageNet \rightarrow T$, $ImageNet \rightarrow X$ and $ImageNet \rightarrow U$. In each cross-domain classification task, we pre-trained all the deep neural networks with and without reliable instance selection to perform the medical image classifications and compared the classification results. The classification accuracy of comparative study is illustrated in Table 7.

Table 6: Cross-domain classification accuracies on Office+Caltech image data sets generated by the transfer learning methods with and without reliable instance selection.

Methods	$A \rightarrow C$	$A \rightarrow W$	$C \rightarrow A$	$C \rightarrow W$	$D \rightarrow A$	$D \rightarrow C$	$D \rightarrow W$	$W \rightarrow A$	$W \rightarrow C$	Ave acc
TCA	75.69	75.59	89.77	74.92	89.24	73.46	**98.30**	80.38	73.64	81.22
F-TCA	**83.44**	**82.37**	**91.54**	**89.15**	**90.71**	**79.43**	96.95	**86.01**	**79.52**	**86.56**
CORAL	83.7	74.58	89.98	78.64	85.70	79.16	**99.66**	77.14	74.98	82.62
F-CORAL	**84.15**	**80.34**	**90.40**	**81.69**	**86.12**	**79.52**	97.29	**82.15**	**75.60**	**84.14**
GFK	76.85	68.47	88.41	80.68	85.80	74.09	**98.64**	75.26	74.8	80.33
F-GFK	**81.66**	**80.68**	**90.71**	**86.44**	**88.10**	**79.34**	94.58	**84.34**	**77.74**	**84.84**
JDA	75.07	70.85	89.67	80.00	88.31	73.91	**98.31**	80.27	72.93	81.04
F-JDA	**82.99**	**80.00**	**91.65**	**87.46**	**89.87**	**78.09**	96.95	**87.16**	**78.72**	**85.88**
KMM	83.08	74.24	91.23	80.34	84.34	71.86	**98.98**	71.81	67.14	80.34
F-KMM	**84.86**	**78.31**	**92.07**	80.00	**84.66**	**77.47**	94.57	**80.06**	**71.68**	**81.79**
BDA	83.79	74.92	89.46	82.03	88.83	81.30	**99.31**	80.85	76.49	84.10
F-BDA	**87.09**	**77.63**	**91.96**	**87.46**	88.83	**82.72**	96.95	**85.70**	**78.98**	**86.37**
MEDA	87.71	85.76	91.07	84.07	92.90	**87.89**	**98.98**	93.21	86.73	89.81
F-MEDA	**87.89**	**89.15**	**92.59**	**93.90**	**93.42**	87.34	95.93	**93.52**	**87.27**	**91.22**
EasyTL	81.3	72.88	90.50	74.91	83.00	73.64	**93.22**	74.53	67.31	79.03
F-EasyTL	**82.1**	**79.66**	**90.60**	**80.00**	**84.86**	**78.90**	90.51	**78.60**	**71.77**	**81.89**
DMP	86.60	91.30	92.80	88.50	91.40	85.30	97.70	91.90	85.60	90.10
F-DMP	**87.10**	**92.10**	**93.50**	**89.10**	**92.30**	**87.10**	**97.80**	**92.90**	**86.20**	**90.90**
CKB	87.00	90.20	93.40	90.80	92.70	83.50	98.80	92.40	84.30	90.30
F-CKB	**88.40**	**92.00**	**93.60**	**92.20**	**94.10**	**85.70**	98.80	**93.30**	**86.30**	**91.60**

Table 7: Classification accuracies of medical image data generated by the deep neural networks pre-trained on ImageNet with and without reliable instance selection.

Methods	$ImageNet \rightarrow X$	$ImageNet \rightarrow U$	$ImageNet \rightarrow T$
DenseNet121	61.40	38.96	39.75
F-DenseNet121	**65.07**	**45.53**	**46.87**
MNASNet	55.52	55.38	37.98
F-MNASNet	**68.20**	**68.65**	**45.43**
MobileNet	49.26	51.11	30.82
F-MobileNet	**60.63**	**71.01**	**47.22**
ShuffleNet	56.18	57.42	30.63
F-ShuffleNet	**66.89**	**63.61**	**46.57**
SqueezeNet	56.35	52.94	34.36
F-SqueezeNet	**63.59**	**66.38**	**47.02**
Ave without instance selection	55.74	51.16	34.71
Ave with instance selection	**64.88**	**63.04**	**46.62**

As shown in Table 7, using the reliable instance selection strategy, the five pre-trained deep neural networks pre-trained on ImageNet (source domain) achieve the average classification accuracies 64.88%, 63.04%, and 46.62% on X, U and T medical image sets (target domains) respectively. In contrast to transferring the pre-trained models that are directly built up on ImageNet, pre-training the deep neural networks based on the selected source domain data gain the significant performance improvements of 9.14%, 11.88%, and 11.91%. The experimental results reveal that selecting reliable instance from source domain can improve the domain adaptation of the pre-trained deep neural networks from ImageNet to various kinds of medical image sets.

4.3. Case Study

In order to further explain the effectiveness of the proposed instance selection strategy, we present a case study to illustrate the reliable instance selection in transfer learning. We exemplify the cross-domain visual object classification task $C \rightarrow W$, in which the images of the source domain C come from Caltech-256 data set and the images of the target domain W come from Webcam data set. There are 10 classes in the cross-domain image classification task and we choose two classes BACKPACK and LAPTOP-COMPUTER for the case study.

BACKPACK

(a)

LAPTOP COMPUTER

BACKPACK

(b)

LAPTOP COMPUTER

Figure 7: (a) Images of classes BACKPACK and LAPTOP-COMPUTER in the source domain. (b) Images of classes BACKPACK and LAPTOP-COMPUTER in the target domain.

BACKPACK

$m(\Omega\,|\,x^s;\Phi) =$ 0.2171   0.1069   0.0978   0.0288   0.1438

(a)

LAPTOP COMPUTER

$m(\Omega\,|\,x^s;\Phi) =$ 0.0545   0.0725   0.0730   0.0692   0.0896

BACKPACK

(b) $m(\Omega\,|\,x^s;\Phi) =$ 0.8420   0.8591   0.7739   0.9124   0.9013   0.9428

LAPTOP COMPUTER

$m(\Omega\,|\,x^s;\Phi) =$ 0.8113   0.7330   0.7967   0.8559   0.7164

Figure 8: (a) Reliable images of BACKPACK and LAPTOP-COMPUTER selected from the source domain for transferred classification. (b) Unreliable images of BACKPACK and LAPTOP-COMPUTER selected from the source domain.

Figure 7(a) and Figure 7(b) show images of backpacks and laptops in the source domain C and target domain W, respectively. It can be observed that in some images from both the source and target domains, objects of the same class exhibit significant differences in appearance and layout, contributing to the domain gap in cross-domain image classification. Figure 8(a) showcases the reliable images selected from the source domain for transfer learning, while Figure 8(b) illustrates the removed unreliable images. It is evident that the objects in the selected source domain images closely resemble the object features in the target domain, with smaller $m(\Omega|x^s; \Phi)$ indicating higher reliability. On the other hand, images of cartoon backpacks and those featuring computers with complex backgrounds exhibit substantial differences in features compared to the target domain, resulting in higher M-values and lower reliability. Inconsistent features may increase domain disparities and lead to negative transfer effects. By eliminating unreliable images from the source domain, data reliability can be enhanced, mitigating negative transfer and improving cross-domain image classification performance.

Additionally, we provide typical examples of medical image classification to illustrate the role of the selection strategy in medical transfer tasks. Specifically, by pretraining DenseNet121 on both the original ImageNet and the ImageNet selected by our proposed strategy, a set of heatmaps for chest X-rays in the I-X pneumonia classification task was obtained using the gradient-weighted class activation mapping (Grad-CAM) technique.
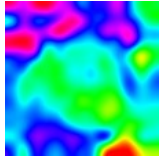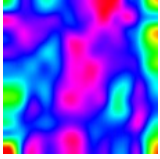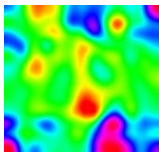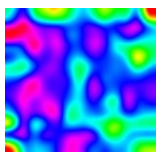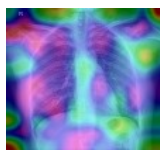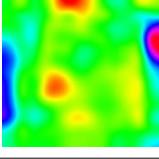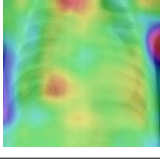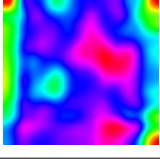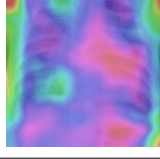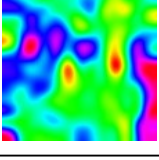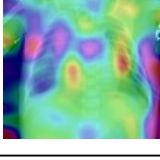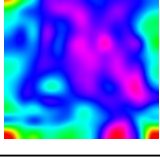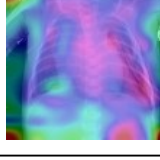
chest X-ray	Trained Network with ImageNet		Trained Network with selected ImageNet	
	CAM heatmap	CAM on image	CAM heatmap	CAM on image

Figure 9: Generated a set of heatmaps of chest X-Ray on the pneumonia classification task based on Grad-CAM

As shown in Figure 9, the darker the color, the more attention the corresponding region receives during the classification process. By comparing the heat maps of pneumonia classification obtained using pretrained networks on ImageNet and the ImageNet selected by our proposed strategy, it is observed that the network trained using the ImageNet selected by the proposed strategy focuses more on the lung region. This is better suited for pneumonia classification tasks, indicating that the relevant regions for pneumonia classification are primarily concentrated in the lungs, particularly in areas with lesions. By contrast, the network trained on the original ImageNet is insensitive to the lung region, focusing on regions unrelated to the lungs for pneumonia classification. This suggests that the original ImageNet contains too many images unrelated to medical imaging, leading to nega-

tive transfer during network training. Therefore, by filtering ImageNet using the proposed data selection strategy, the credibility of ImageNet in the field of medical imaging can be effectively enhanced, thereby increasing network transferability.

5. Conclusion

To remove noisy data from the source domain and data unrelated to the target domain task, we proposed a reliable data-selection strategy based on evidence theory. This strategy utilizes a mass function to measure the reliability of the source domain data with respect to the target domain task, selecting trustworthy source domain data that are more relevant to the task of training the transfer models. Specifically, a mass function was formulated to measure the degree of ignorance and reliability of the source domain data with respect to the learning task in the target domain. By selecting reliable data with a low degree of ignorance from the source domain, the domain adaptation of transfer learning models can be enhanced. The proposed strategy is a general preprocessing technique that can be integrated into most transfer learning methods to improve performance. Experiments on text and medical image data validated the effectiveness of the proposed selection strategy in improving the performance of various types of transfer learning methods.However, when the amount of data in the source domain is relatively small, the improvement effect of this strategy is not significant. This indicates that in cases of limited data, the selection strategy may need further optimization or integration with other techniques to enhance the effect. In the training and fine-tuning of a large language model (LLM), high-quality training data play a crucial role in model performance while saving training resources. In the future, we will apply a selection strategy to the process of building LLM during both pre-training and fine-tuning. For example, in prompt tuning, when generating prompts through LLM, a selection strategy can be incorporated to generate high-quality prompts. (Liu et al., 2023). Also, it is very worth studying to apply the association among features from reliable instances in explicit and interpretable manner like (Liang et al., 2022) to feature-based TL methods.

References

Ajith, A., & Gopakumar, G. (2023). Domain adaptation: A survey. In *Proceedings of Computer Vision and Machine Intelligence* (pp. 591–602). `https://doi.org/10.1007/978-981-19-7867-8_47`.

Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, *28*, 2352–3409. doi:`https://doi.org/10.1016/j.dib.2019.104863`.

Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 17–36).

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 440–447).

Campagner, A., Ciucci, D., & Denœux, T. (2023). A distributional framework for evaluation, comparison and uncertainty quantification in soft clustering. *International Journal of Approximate Reasoning*, (p. 109008). doi:`https://doi.org/10.1016/j.ijar.2023.109008`.

Cao, Y., Long, M., & Wang, J. (2018). Unsupervised domain adaptation with distribution matching machines. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 2795–2802). doi:`https://doi.org/10.1609/aaai.v32i1.11792`.

Cheplygina, V., de Bruijne, M., & Pluim, J. P. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, *54*, 280–296. doi:`https://doi.org/10.1016/j.media.2019.03.009`.

Chu, W.-S., De la Torre, F., & Cohn, J. F. (2013). Selective transfer machine for personalized facial action unit detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3515–3522). doi:`https://doi.org/10.1109/CVPR.2013.451`.

Dai, W., Yang, Q., Xue, G.-R., & Yu, Y. (2007). Boosting for transfer learning. In *Proceedings of the 24th International Conference on Machine Learning* (pp. 193–200).

Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, *38*, 325–339. doi:`https://doi.org/10.1007/978-3-540-44792-4_3`.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE. doi:`https://doi.org/110.1109/CVPR.2009.5206848`.

Denœux, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, *25*, 804–813.

Denœux, T. (2019). Logistic regression, neural networks and Dempster-Shafer theory: A new perspective. *Knowledge-Based Systems*, *176*, 54–67.

Denœux, T. (2024). Uncertainty quantification in logistic regression using random fuzzy sets and belief functions. *International Journal of Approximate Reasoning*, *168*, 109159.

Denœux, T., Dubois, D., & Prade, H. (2020). Representations of uncertainty in artificial intelligence: Beyond probability and possibility. In P. Marquis, O. Papini, & H. Prade (Eds.), *A Guided Tour of Artificial Intelligence Research* chapter 4. (pp. 119–150). Springer Verlag volume 1.

Denœux, T., Kanjanatarakul, O., & Sriboonchitta, S. (2019). A new evidential k-nearest neighbor rule based on contextual discounting with partially supervised learning. *International Journal of Approximate Reasoning*, *113*, 287–302.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). Decaf: A deep convolutional activation feature for

generic visual recognition. In *International conference on machine learning* (pp. 647–655). PMLR.

Duan, L., Tsang, I. W., Xu, D., & Maybank, S. J. (2009). Domain transfer svm for video concept detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1375–1381). doi:`https://doi.org/10.1109/CVPR.2009.5206747`.

Ghifary, M., Balduzzi, D., Kleijn, W. B., & Zhang, M. (2017). Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 1414–1430. doi:`https://doi.org/10.1109/TPAMI.2016.2599532`.

Gholami, B., & Pavlovic, V. (2017). Punda: Probabilistic unsupervised domain adaptation for knowledge transfer across visual categories. In *Proceedings of the IEEE international conference on computer vision* (pp. 3581–3590). doi:`https://doi.org/10.1109/ICCV.2017.387`.

Gong, B., Shi, Y., Sha, F., & Grauman, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2066–2073). doi:`https://doi.org/10.1109/CVPR.2012.6247911`.

Gong, C., Su, Z.-G., Wang, P.-H., Wang, Q., & You, Y. (2023). A sparse reconstructive evidential k-nearest neighbor classifier for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, *35*, 5563–5576. doi:`https://doi.org/10.1109/TII.2023.3241587`.

Guo, H., Pasunuru, R., & Bansal, M. (2020). Multi-source domain adaptation for text classification via distancenet-bandits. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 7830–7838). doi:`https://doi.org/10.1609/aaai.v34i05.6288`.

Gupta, N., & Jalal, A. S. (2022). Traditional to transfer learning progression on scene text detection and recognition: a survey. *Artificial Intelligence Review*, (pp. 1–46). doi:`https://doi.org/10.1007/s10462-021-10091-3`.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional

neural networks for mobile vision applications, . doi:`https://doi.org/10.48550/arXiv.1704.04861`.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708). doi:`https://doi.org/10.1109/CVPR.2017.243`.

Huang, J., Gretton, A., Borgwardt, K. M., Scholkopf, B., & Smola, A. J. (2006). Correcting sample selection bias by unlabeled data. In *Advances in Neural Information Processing Systems 19* (pp. 601–608).

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5mb model size, . doi:`https://doi.org/10.48550/arXiv.1602.07360`.

Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, *11*, 40. `https://doi.org/10.3390/technologies11020040`.

Jiang, J., Shu, Y., Wang, J., & Long, M. (2022). Transferability in deep learning: A survey. `http://arxiv.org/abs/2201.05867`.

Jiang, J., & Zhai, C. (2007). Instance weighting for domain adaptation in nlp. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 264–271).

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F. et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*, 1122–1131. doi:`https://doi.org/10.1016/j.cell.2018.02.010`.

Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., Meenakshi, K., Swaraja, K., Plawiak, P., & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, *42*, 79–107. doi:`https://doi.org/10.1016/j.bbe.2021.11.004`.

Li, J., Li, D., Xiong, C., & Hoi, S. (2022). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning* (pp. 12888–12900).

Liang, X., Guo, Q., Qian, Y., Ding, W., & Zhang, Q. (2021). Evolutionary deep fusion method and its application in chemical structure recognition. *IEEE Transactions on Evolutionary Computation*, *25*, 883–893. doi:`https://doi.org/10.1109/TEVC.2021.3064943`.

Liang, X., Qian, Y., Guo, Q., Cheng, H., & Liang, J. (2022). AF: An association-based fusion method for multi-modal classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*, 9236–9254. doi:`https://doi.org/10.1109/TPAMI.2021.3125995`.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*, 1–35. doi:`https://doi.org/10.1145/3560815`.

Liu, Z. G., Huang, L. Q., Zhou, K., & Denœux, T. (), .

Long, M., Wang, J., Ding, G., Pan, S. J., & Philip, S. Y. (2013a). Adaptation regularization: A general framework for transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *26*, 1076–1089. doi:`https://doi.org/10.1109/TKDE.2013.111`.

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013b). Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2200–2207). doi:`https://doi.org/10.1109/ICCV.2013.274`.

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2014). Transfer joint matching for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1410–1417). doi:`https://doi.org/10.1109/CVPR.2014.183`.

Luo, Y.-W., & Ren, C.-X. (2021). Conditional bures metric for domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13989–13998). doi:`https://doi.org/10.1109/CVPR46437.2021.01377`.

Luo, Y.-W., Ren, C.-X., Dai, D.-Q., & Yan, H. (2020). Unsupervised domain adaptation via discriminative manifold propagation. *IEEE transactions on pattern analysis and machine intelligence*, *44*, 1653–1669. doi:`https://doi.org/10.1109/TPAMI.2020.3014218`.

Morid, M. A., Borjali, A., & Del Fiol, G. (2020). A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in Biology and Medicine*, (p. 104115). doi:`https://doi.org/10.1016/j.compbiomed.2020.104115`.

Mustafa, B., Loh, A., Freyberg, J., MacWilliams, P., Wilson, M., McKinney, S. M., Sieniek, M., Winkens, J., Liu, Y., Bui, P., Prabhakara, S., Telang, U., Karthikesalingam, A., Houlsby, N., & Natarajan, V. (2021). Supervised transfer learning at scale for medical imaging. doi:`https://doi.org/10.48550/arXiv.2101.05913`.

Öztürk, C., Taşyürek, M., & Türkdamar, M. U. (2023). Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the cnn-based deep learning models. *Concurrency and Computation: Practice and Experience*, *35*, 315–319. doi:`https://doi.org/10.1109/ECICE52819.2021.9645649`.

Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, *22*, 199–210. doi:`https://doi.org/10.1109/TNN.2010.2091281`.

Quost, B., Denœux, T., & Li, S. (2017). Parametric classification with soft labels using the evidential EM algorithm: linear discriminant analysis versus logistic regression. *Advances in Data Analysis and Classification*, *11*, 659–690. doi:`https://doi.org/10.1007/s11634-017-0301-2`.

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In *Advances in neural information processing systems* (pp. 3347–3357).

Sariyildiz, M. B., Alahari, K., Larlus, D., & Kalantidis, Y. (2023). Fake it till you make it: Learning transferable representations from synthetic imagenet clones. In *CVPR 2023-IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1–11). doi:`https://doi.org/10.1109/CVPR52729.2023.00774`.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press.

Shang, H., Sun, Z., Yang, W., Fu, X., Zheng, H., Chang, J., & Huang, J. (2019). Leveraging other datasets for medical imaging classification: evaluation of transfer, multi-task and semi-supervised learning. In *International conference on medical image computing and computer-assisted intervention* (pp. 431–439). Springer. doi:`https://doi.org/10.1007/978-3-030-32254-0_48`.

Sohn, K., Chang, H., Lezama, J., Polania, L., Zhang, H., Hao, Y., Essa, I., & Jiang, L. (2023). Visual prompt tuning for generative transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 19840–19851). `https://doi.org/10.1109/CVPR52729.2023.01900`.

Su, Z., Hu, Q., & Denoeux, T. (2020). A distributed rough evidential k-nn classifier: Integrating feature reduction and classification. *IEEE Transactions on Fuzzy Systems*, *PP*, 1–1. doi:`https://doi.org/10.1109/TFUZZ.2020.2998502`.

Subramanian, M., Shanmugavadivel, K., & Nandhini, P. (2022). On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Computing and Applications*, *34*, 13951–13968. doi:`https://doi.org/10.1007/s00521-022-07246-w`.

Sun, B., Feng, J., & Saenko, K. (2016). Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (p. 2058–2065).

Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q. V. (2019). Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2820–2828). doi:`https://doi.org/10.1109/CVPR.2019.00293`.

Tong, Z., Xu, P., & Denœux, T. (2021). An evidential classifier based on Dempster-Shafer theory and deep learning. *Neurocomputing*, *450*, 275–293.

Wang, J., Chen, Y., Hao, S., Feng, W., & Shen, Z. (2017). Balanced distribution adaptation for transfer learning. In *2017 IEEE international conference on data mining (ICDM)* (pp. 1129–1134). doi:`https://doi.org/10.1109/ICDM.2017.150`.

Wang, J., Chen, Y., Yu, H., Huang, M., & Yang, Q. (2019). Easy transfer learning by exploiting intra-domain structures. In *2019 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1210–1215). doi:`https://doi.org/10.1109/ICME.2019.00211`.

Wang, J., Feng, W., Chen, Y., Yu, H., Huang, M., & Yu, P. S. (2018). Visual domain adaptation with manifold embedded distribution alignment. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 402–410). doi:`https://doi.org/10.1145/3240508.3240512`.

Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis*, (p. 101985). doi:`https://doi.org/10.1016/j.media.2021.101985`.

Xu, Y., Yu, H., Yan, Y., & Liu, Y. (2020). Multi-component transfer metric learning for handling unrelated source domain samples. *Knowledge-Based Systems*, *203*, 106132. doi:`https://doi.org/10.1016/j.knosys.2020.106132`.

Yan, H., Ding, Y., Li, P., Wang, Q., Xu, Y., & Zuo, W. (2017). Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2272–2281). doi:`https://doi.org/10.1109/CVPR.2017.107`.

Yang, J., Liu, J., Xu, N., & Huang, J. (2023). Tvt: Transferable vision transformer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 520–530). doi:`https://doi.org/10.1109/WACV56688.2023.00059`.

Yang, J., Yan, R., & Hauptmann, A. G. (2007). Cross-domain video concept detection using adaptive svms. In *Proceedings of the 15th ACM International Conference on Multimedia* (pp. 188–197). doi:`https://doi.org/10.1145/1291233.1291276`.

Yao, Y., & Doretto, G. (2010). Boosting for transfer learning with multiple sources. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1855–1862). IEEE. doi:https://doi.org/10.1109/cvpr.2010.5539857.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Proceedings of the 27th International Conference on Neural Information Processing Systems* (pp. 3320–3328).

Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., & Zhang, Y.-D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing*, *489*, 230–254. https://doi.org/10.1016/j.neucom.2021.08.159.

Zhang, L., & Gao, X. (2024). Transfer adaptation learning: A decade survey. *IEEE Transactions on Neural Networks and Learning Systems*, *35*, 23–44. doi:https://doi.org/10.1109/TNNLS.2022.3183326.

Zhang, W., Deng, L., Zhang, L., & Wu, D. (2023). A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica*, *10*, 305–329. doi:https://doi.org/10.1109/JAS.2022.106004.

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6848–6856). doi:https://doi.org/10.1109/CVPR.2018.00716.

Zhao, H., Hu, J., & Risteski, A. (2020). On learning language-invariant representations for universal machine translation. In *International Conference on Machine Learning* (pp. 11352–11364).

Zhong, H., Yu, S., Trinh, H., Lv, Y., Yuan, R., & Wang, Y. (2023). Finetuning transfer learning based on dcgan integrated with self-attention and spectral normalization for bearing fault diagnosis. *Measurement*, *210*, 112421. doi:https://doi.org/10.1016/j.measurement.2022.112421.

Zhou, H., Chen, W., Cheng, L., Liu, J., & Xia, M. (2023). Trustworthy fault diagnosis with uncertainty estimation through evidential convolutional neural networks. *IEEE Transactions on Industrial Informatics*, . doi:https://doi.org/10.1109/TII.2023.3241587.

Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., & He, Q. (2021). Deep subdomain adaptation network for image classification. *IEEE transactions on neural networks and learning systems*, *32*, 1713–1722. doi:`https://doi.org/10.1109/TNNLS.2020.2988928`.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*, 43–76. doi:`https://doi.org/10.1109/JPROC.2020.3004555`.