Theoretical Computer Science 940 (2023) 90–107

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Analysis on methods to effectively improve transfer learning performance

Honghui Xu, Wei Li, Zhipeng Cai *

Georgia State University, Department of Computer Science, Atlanta, GA, USA

ARTICLE INFO

Article history:
Received 26 March 2022
Received in revised form 19 July 2022
Accepted 24 September 2022
Available online 14 October 2022

Keywords:
Transfer learning
Collaborative transfer learning
PAC

ABSTRACT

Transfer learning has become a prevailing machine learning technique thanks to its superiority in learning knowledge from limited training data for prediction. In the existing works, collection and collaboration are two major approaches to realize the improvement of transfer learning performance. Even though the effectiveness of these approaches has been validated in extensive experiments, there lacks the support of theoretical analysis. Consequently, how to enhance transfer learning effectively is an open problem. In light of this, in this paper, we thoroughly and deeply study the methods of improving transfer learning performance in order to provide the guidelines for applying transfer learning in real applications. Through our proof process, critical conclusions are drawn to help learn the motivation of implementing collection and collaboration, the performance gap between collection and collaboration, and the impacts of data sharing strategies on transfer learning in collaboration. These conclusions can further build a theoretical foundation for future research on transfer learning.

1. Introduction

With impressive progress of machine learning techniques, they have been explosively applied in many practical applications [1–3]. Traditional machine learning techniques can be well implemented in an ideal scenario, in which there are a considerable amount of labeled training instances, and the test instances have the same distribution as the training samples [4,5]. Unfortunately, it may be impractical or impossible to collect sufficient labeled training data in every application due to expensive collection cost (*e.g.*, collection time), limited collection resources (*e.g.*, user's network bandwidth), or harsh collection environment (*e.g.*, dense forest). To overcome this challenge, transfer learning [6,7], which takes advantage of transferring the shared knowledge between training data and testing data, has been proposed as a promising machine learning methodology to leverage the relatively fewer training instances for expected prediction.

Recently, designing algorithms to improve transfer learning has attracted lots of research attentions. In the literature, there are two typical methods to achieve the improvement of transfer learning performance. One vein of research focuses on collection approaches. Specifically, the performance of transfer learning can be enhanced by collecting more training samples thanks to the increase of training data volume [8–10], and can also be promoted through collecting more attributes owing to the expansion of shared knowledge [11–13]. However, these works have a common limitation that the effectiveness

* Corresponding author.
E-mail addresses: hxu16@student.gsu.edu (H. Xu), wli28@gsu.edu (W. Li), zcai@gsu.edu (Z. Cai).

https://doi.org/10.1016/j.tcs.2022.09.023
0304-3975/

of collection approaches has not been demonstrated via rigorous proof. On the other hand, some works prompt institutions to get involved into collaboration for the improvement of transfer learning performance [14–22]. Although comprehensive experiments have illustrated the usefulness of these collaborative transfer learning models, the above works are still short of analysis on whether collaboration can indeed yield the enhancement of transfer learning. Nonetheless, considering the proliferation of transfer learning models in real applications, it is urgent to propose theoretical analysis to guide entities on how to make decisions on collection or collaboration before applying these models for improving transfer learning performance.

To fill this blank, we conduct a thorough investigation about the methods of enhancing transfer learning effectively. Firstly, we answer the question that how to select a more effective option from two collection approaches for a better transfer learning performance. Then, we discuss whether collaborative transfer learning in a two-party collaboration scenario can improve transfer learning performance. In addition, we compare the effectiveness of collection and collaboration to learn which one is a better choice. Finally, we study the influence of data sharing strategies on transfer learning performance during collaboration. Through our thorough analysis, one can understand the incentive of collection and collaboration, the performance gap between different collection/collaboration approaches, the strategies of sharing data in collaboration, which provides fundamental guidance of transfer learning implementation and application. Our contributions in this paper are addressed in the following.

- To the best of our knowledge, this paper is the first work to systematically investigate the effectiveness of approaches for enhancing transfer learning performance.
- We provide theoretical analysis on the selection of collection approaches for performance enhancement in transfer learning.
- We demonstrate whether collaboration in a two-party scenario can help achieve the promotion of transfer learning performance.
- We illustrate the effectiveness difference between collection and collaboration through rigorous comparison.
- We present deep insight into data sharing strategies in collaborative transfer learning.

We briefly summarize the related works in Section 2. After presenting the problem formulation in Section 3, we elaborate on investigation in Section 4, Section 5, Section 6, and Section 7. Finally, we end up with a conclusion of this paper in Section 8.

2. Related works

Transfer learning [6,7] is one of promising learning models to realize information transformation from learned models to changing or unknown data distributions by leveraging the knowledge shared between the known training data and the unknown testing data. In the state-of-the-arts, there exist four major types of transfer learning models. (i) Instance-based transfer learning models [8–10,23,24] are mainly inspired by the instance weighting strategy. (ii) Feature-based transfer learning models [11–13,25] achieve the mapping from the original features to a new feature space, which can be implemented either asymmetrically or symmetrically. The asymmetric methods [26,11] aim to match the features in the source domain with the features in the target domain through a transformation scheme. The symmetric methods [27,13,25] figure out a common feature space and then map the source and target features into the common feature space. (iii) Parameter-based transfer learning models [28–30] transfer the shared knowledge at the parameter level. (iv) Relational-based transfer learning models [31,32] transfer the logical relationship learned in the source domain to the target domain.

Since the cost of collecting new labeled training samples can be significantly reduced, transfer learning models have been widely implemented in real applications for different purposes, such as disease prediction [33], sign language recognition [34], and target online display advertising [35]. Nevertheless, the theoretical study on the performance improvement of transfer learning has not been clearly addressed in literature yet.

Collaborative transfer learning employs collaboration [36,37] to enhance transfer learning performance via sharing data among different organizations/institutions. In [17,19,38], different algorithms of realizing collaborative transfer learning were proposed. While, [16,18,20,21,15,22,39] studied the application of collaborative transfer learning in specific scenarios. For example, with the help of sharing data among institutions, collaborative transfer learning can be implemented to enhance the accuracy of medical diagnosis (*e.g.*, COVID-19 diagnosis) in the process of healthcare analysis [16,20], be applied to promote the financial marketing forecasting performance during the financial marketing analysis [18], and be used to achieve the improvement of the performance of EEG signal classification [15], urban risk recognition [22], cross-domain prediction for smart manufacturing [21], and disaster classification in social computing networks [39]. But, in the above works, the efficiency of collaborative transfer learning is not theoretically analyzed.

So far, no work has clearly answered how to efficiently enhance transfer learning performance. Motivated by this observation, in this paper, we endeavor to provide theoretical guidance for the improvement of transfer learning by carrying out comprehensive and thorough investigation on the methods of enhancing transfer learning performance.

3. Problem formulation

In this paper, we use Probably Approximately Correct (PAC) learning framework [40] to establish problem formulation for transfer learning.

Transfer learning can be taken as a classifier training process, where the classifier is trained by the training instances attached with known labels to make a prediction for the testing instances attached with the novel/unknown labels [41]. Inspired by the idea of [42], the definition of transfer learning aiming at learning a classifier is described in Definition 1.

Definition 1. There are two mapping functions in a classifier. The first one maps a training instance space $\mathcal{X}$ to an attribute space $\mathcal{T}$ (*i.e.*, $\mathcal{X} \to \mathcal{T}$), and the second one maps $\mathcal{T}$ to a label space $\mathcal{Y}$ (*i.e.*, $\mathcal{T} \to \mathcal{Y}$).

For the mapping $\mathcal{X} \to \mathcal{T}$ with $T \in \mathbb{N}^+$ attributes in $\mathcal{T}$, T binary classifiers should be trained by using the training instances in $\mathcal{X}$; and for the mapping $\mathcal{T} \to \mathcal{Y}$, the T trained binary classifiers are exploited to forecast class labels in $\mathcal{Y}$. The labels of each instance are denoted as a 0-1 binary attribute vector. If there is an attribute in an instance, the element corresponding to the attribute is equal to 1 in the attribute vector; otherwise, the element corresponding to the attribute is equal to 0 in the attribute vector.

We use $L_{\mathcal{X}\to\mathcal{Y}}$ to represent the training loss of the whole learning process $\mathcal{X} \to \mathcal{Y}$ and $\tau \in (0, 1)$ to represent the upper bound of $L_{\mathcal{X}\to\mathcal{Y}}$. The probability that $L_{\mathcal{X}\to\mathcal{Y}}$ is not larger than τ is computed below:

$$\mathbb{P}(L_{\mathcal{X}\to\mathcal{Y}} \leq \tau) = 1 - \gamma, \tag{1}$$

in which $\gamma \in (0, 1)$ represents the error probability.

The main technical difference between transfer learning and traditional learning processes is that transfer learning aims at training T binary classifiers rather than a classifier for prediction via training instances. The training loss of each binary classifier during the process $\mathcal{X} \to \mathcal{T}$ is represented as $L_{\mathcal{X}\to\mathcal{T}}$. In order to satisfy the requirement that $L_{\mathcal{X}\to\mathcal{Y}} \leq \tau$, we have $L_{\mathcal{X}\to\mathcal{T}} \leq \frac{\tau}{T}$ with the following probability:

$$\mathbb{P}(L_{\mathcal{X}\to\mathcal{T}} \leq \frac{\tau}{T}) = 1 - \delta, \tag{2}$$

in which $\delta \in (0, 1)$ denotes the error probability. Thus, the probability of satisfying $L_{\mathcal{X}\to\mathcal{T}} \leq \frac{\tau}{T}$ for T classifiers should be $(1-\delta)^T$, and the probability of meeting $L_{\mathcal{X}\to\mathcal{Y}} \leq \tau$ with T classifiers should be $\mathrm{BinoCDF}(\tau; T, \frac{\tau}{T})$ that is the binomial cumulative distribution probability. Then, based on PAC learning framework [40], in order to meet the requirements of Eq. (1) and Eq. (2), at least N training instances are expected to exist in $\mathcal{X}$, and the value of N can be estimated via Eq. (3).

$$N = \frac{T^2}{\tau^2}\lceil \ln(\frac{2}{\delta}) \rceil. \tag{3}$$

By Eq. (3), we can get $\delta = \frac{2}{e^{\frac{N\tau^2}{T^2}}}$.

Therefore, when N training instances and T classifiers are exploited to train transfer learning model, we use the total probability $P(N, T)$ of the event that the training loss is not larger than τ to define the performance of the transfer learning process $\mathcal{X} \to \mathcal{Y}$ as follows.

$$P(N, T) = (1 - \frac{2}{e^{\frac{N\tau^2}{T^2}}})^T \cdot \mathrm{BinoCDF}(\tau; T, \frac{\tau}{T}) \cdot (1 - \gamma). \tag{4}$$

The performance of transfer learning can be promoted through two mainstream ways. (i) **Collection:** collecting more training instances can train better binary classifiers, and collecting more attributes is helpful to predict novel labels more accurately. (ii) **Collaboration:** collaborative transfer learning encourages involved participants to share their collected training instances and attributes for performance improvement.

To answer the problem of how to effectively improve transfer learning performance, we carry out the following studies:

- When considering the benefit and/or the cost of collecting instances and collecting attributes, which collection approach is better?
- When taking into account the benefit and/or the cost of collaborative transfer learning, can collaboration help make an improvement?
- When both collection and collaboration are available, which one is more beneficial?
- When joining a collaboration, how to decide data sharing strategies?

Our rigorous analysis on transfer learning are detailed in Section 4, Section 5, Section 6, and Section 7 in order.

All the mathematical notations used in this paper are listed in Table 1.

Table 1
Description of Mathematical Symbols.

Symbols	Description
$\mathcal{X}$	Training instance space
$\mathcal{Y}$	Label space
$\mathcal{T}$	Attribute space
$L_{\mathcal{X}\to\mathcal{Y}}$	Training loss of the learning process $\mathcal{X}\to\mathcal{Y}$
$L_{\mathcal{X}\to\mathcal{T}}$	Training loss of the learning process $\mathcal{X}\to\mathcal{T}$
τ	Upper bound of $L_{\mathcal{X}\to\mathcal{Y}}$
N	Number of training instances in $\mathcal{X}$
T	Number of attributes in $\mathcal{T}$
γ	Error probability that $L_{\mathcal{X}\to\mathcal{Y}}$ is not larger than τ
δ	Error probability that $L_{\mathcal{X}\to\mathcal{T}}$ is not larger than $\frac{\tau}{T}$
$P(N,T)$	Performance of transfer learning
r_n	Benefit of collecting one more instance
r_t	Benefit of collecting one more attribute
c_n	Cost of collecting one instance
c_t	Cost of collecting one attribute
N_1	Number of training instances collected by party **A**
N_2	Number of training instances collected by party **B**
T_1	Number of attributes collected by party **A**
T_2	Number of attributes collected by party **B**
n	Number of overlapping instances between party **A** and party **B**
t	Number of overlapping attributes between party **A** and party **B**
α_1	Ratio of shared instances for party **A**
α_2	Ratio of shared instances for party **B**
β_1	Ratio of shared attributes for party **A**
β_2	Ratio of shared attributes for party **B**

4. Collecting instances vs. collecting attributes

In this section, we first only compare collection benefit (that purely indicates transfer learning performance) and then extend to compare collection utility (that equals to benefit minus cost) to comprehensively understand which collection method is more effective.

4.1. Benefit comparison

To begin with, in order to simply present $P(N,T)$, we set

$$f(N,T) = (1 - \frac{2}{e^{\frac{N\tau^2}{T^2}}})^T, \tag{5}$$

and

$$g(T) = \mathrm{BinoCDF}(\tau; T, \frac{\tau}{T}). \tag{6}$$

Then, according to these simplified expressions, $P(N,T)$ can be rewritten in Eq. (7).

$$P(N,T) = f(N,T)g(T)(1-\gamma). \tag{7}$$

We can calculate the benefit of collecting one more instance r_n by Eq. (8) and the benefit of collecting one more attribute r_t by Eq. (9).

$$r_n = \Delta P_N(N,T) = P(N,T) - P(N-1,T). \tag{8}$$

$$r_t = \Delta P_T(N,T) = P(N,T) - P(N,T-1). \tag{9}$$

Accordingly, the benefit difference $k(N,T)$ should be:

$$\begin{aligned} k(N,T) = r_t - r_n &= P(N,T) - P(N,T-1) - [P(N,T) - P(N-1,T)] \\ &= P(N-1,T) - P(N,T-1) \\ &= [f(N-1,T)g(T) - f(N,T-1)g(T-1)](1-\gamma). \end{aligned} \tag{10}$$

Taylor Theorem [43] and Newton's forward interpolation formula [44] are exploited to get a bivar linear function $\tilde{k}(N,T)$ for the approximation of $k(N,T)$. Also, to avoid the gradient vanish for further calculation, $k(N,T)$ should be approximated at the point $(3,3)$, presented in Eq. (11).

$$\tilde{k}(N,T) = k(3,3) + (N-3)\Delta k_N(3,3) + (T-3)\Delta k_T(3,3). \tag{11}$$

Using Eq. (10), Eq. (12) can be obtained.

$$k(3,3) = [f(2,3)g(3) - f(3,2)g(2)](1-\gamma). \tag{12}$$

From Eq. (10) and Newton's forward interpolation formula [44], the computation of gradients can be expressed below:

$$\begin{aligned} \Delta k_N(3,3) &= \frac{k(3,3)-k(2,3)}{3-2} = k(3,3) - k(2,3) \\ &= [f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma). \end{aligned} \tag{13}$$

$$\begin{aligned} \Delta k_T(3,3) &= \frac{k(3,3)-k(3,2)}{3-2} = k(3,3) - k(3,2) \\ &= [f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma). \end{aligned} \tag{14}$$

With the substitution of Eq. (12), Eq. (13), and Eq. (14), $\tilde{k}(N,T)$ can be re-expressed in Eq. (15).

$$\begin{aligned} \tilde{k}(N,T) = &[f(2,3)g(3) - f(3,2)g(2)](1-\gamma) \\ &+ (N-3)[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma) \\ &+ (T-3)[f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma). \end{aligned} \tag{15}$$

If $\tilde{k}(N,T) \le 0$, collecting one more training instance can bring more benefit for performance enhancement; otherwise, collecting one more attribute can bring more benefit. Hence, by comparing $\tilde{k}(N,T)$ with 0, we can determine which collection option can bring more benefit for one party to improve transfer learning performance.

Theorem 1. *Suppose that one party has N training instances and T attributes. Collecting one more training instance can bring more benefit than collecting one more attribute for the enhancement of transfer learning performance, when any one of the following two conditions holds:*

(i) *Condition 1: N, T, and τ satisfy Eq. (16) and Eq. (17);*
(ii) *Condition 2: N, T, and τ satisfy Eq. (18) and Eq. (19).*

$$\begin{aligned} N \le &-(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\ &- \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} + 3. \end{aligned} \tag{16}$$

$$f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0. \tag{17}$$

$$\begin{aligned} N \ge &-(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\ &- \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} + 3. \end{aligned} \tag{18}$$

$$f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0. \tag{19}$$

Note that $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,2) = (1 - \frac{2}{e^{\frac{\tau^2}{2}}})^2$, $f(3,1) = (1 - \frac{2}{e^{3\tau^2}})$, $f(1,3) = (1 - \frac{2}{e^{\frac{\tau^2}{9}}})^3$ according to Eq. (5), and $g(1) = \mathrm{BinoCDF}(\tau; 1, \frac{\tau}{T})$, $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T})$ according to Eq. (6).

Proof. Collecting one more instance to get more benefit means that $\tilde{k}(N,T) \le 0$. According to Eq. (15), there should be

$$\begin{aligned} &[f(2,3)g(3) - f(3,2)g(2)](1-\gamma) \\ &+ (N-3)[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma) \\ &+ (T-3)[f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma) \le 0. \end{aligned} \tag{20}$$

Three cases exist when Eq. (20) is resolved:

(i) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0$ (*i.e.*, Eq. (17) is satisfied), then Eq. (16) can be gained;
(ii) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0$ (*i.e.*, Eq. (19) is satisfied), then we can obtain Eq. (18);
and (iii) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) = 0$, it is a senseless case for our investigated problem.

Therefore, Theorem 1 is proved.

Theorem 2. *Suppose that one party has N training instances and T attributes. Collecting one more attribute can bring more benefit than collecting one more training instance for enhancing transfer learning performance, when any one of the following two conditions holds:*

(i) *Condition 1: N, T, and τ satisfy Eq. (21) and Eq. (22);*
(ii) *Condition 2: N, T, and τ satisfy Eq. (23) and Eq. (24).*

$$\begin{aligned} N > &-(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\ &- \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} + 3. \end{aligned} \tag{21}$$

$$f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0. \tag{22}$$

$$\begin{aligned} N < &-(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\ &- \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} + 3. \end{aligned} \tag{23}$$

$$f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0. \tag{24}$$

Note that $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,2) = (1 - \frac{2}{e^{\frac{\tau^2}{2}}})^2$, $f(3,1) = (1 - \frac{2}{e^{3\tau^2}})$, $f(1,3) = (1 - \frac{2}{e^{\frac{\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(1) = \mathrm{BinoCDF}(\tau; 1, \frac{\tau}{T})$, $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T})$ *according to Eq. (6).*

Proof. Collecting one more attribute to obtain more benefit indicates that $\tilde{k}(N,T) > 0$. Based on Eq. (15), there is

$$\begin{aligned} &[f(2,3)g(3) - f(3,2)g(2)](1-\gamma) \\ &+ (N-3)[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma) \\ &+ (T-3)[f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma) > 0. \end{aligned} \tag{25}$$

Considering Eq. (25), we should discuss the following cases:

(i) $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0$ (*i.e.*, Eq. (22) is met), then we have Eq. (21);
(ii) $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0$ (*i.e.*, Eq. (24) is satisfied), then we get Eq. (23);
and (iii) $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) = 0$, it is a senseless case for our investigated problem.

Hence, Theorem 2 is proved.

4.2. Utility comparison

When collection cost is taken into account, utility is used for comparison. Assume that the cost of collecting one instance is c_n [45,46], and the cost of collecting one attribute is c_t [47,48]. Accordingly, the utility of collecting one more instance u_n and the utility of collecting one more attribute u_t can be computed by Eq. (26) and Eq. (27), respectively.

$$u_n = \Delta P_N(N,T) - c_n = P(N,T) - P(N-1,T) - c_n. \tag{26}$$

$$u_t = \Delta P_T(N,T) - c_t = P(N,T) - P(N,T-1) - c_t. \tag{27}$$

Then, we can obtain the approximated utility difference function $\tilde{k}(N,T)$ in Eq. (28) by following the same calculation process in Section 4.1.

$$
\begin{aligned}
\tilde{k}(N,T) = & u_t - u_n = \\
= & [f(2,3)g(3) - f(3,2)g(2)]\,(1-\gamma) \\
& + (N-3)[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma) \\
& + (T-3)[f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma) \\
& + c_n - c_t.
\end{aligned} \tag{28}
$$

Also, similar to the analysis in Section 4.1, we compare $\tilde{k}(N,T)$ with 0 to study which collection option brings a higher utility increase in transfer learning for one party.

Theorem 3. *Suppose that one party owns N training instances and T attributes. Compared with collecting one more attribute, collecting one more instance can generate a higher utility increase in transfer learning when any one of the following two conditions holds:*

(i) *Condition 1: N, T, c_n, c_t, γ and τ satisfy Eq. (29) and Eq. (30);*
(ii) *Condition 2: N, T, c_n, c_t, γ and τ satisfy Eq. (31) and Eq. (32).*

$$
\begin{aligned}
N \le & -(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\
& - \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\
& + \frac{c_t - c_n}{[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma)} + 3.
\end{aligned} \tag{29}
$$

$$
f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0. \tag{30}
$$

$$
\begin{aligned}
N \ge & -(T-3)\frac{f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\
& - \frac{f(2,3)g(3) - f(3,2)g(2)}{f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)} \\
& + \frac{c_t - c_n}{[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma)} + 3.
\end{aligned} \tag{31}
$$

$$
f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0. \tag{32}
$$

Note that $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,2) = (1 - \frac{2}{e^{\frac{\tau^2}{2}}})^2$, $f(3,1) = (1 - \frac{2}{e^{3\tau^2}})$, $f(1,3) = (1 - \frac{2}{e^{\frac{\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(1) = \mathrm{BinoCDF}(\tau; 1, \frac{\tau}{T})$, $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T})$ *according to Eq. (6).*

Proof. If collecting one more instance can increase learning utility even more, there should be $\tilde{k}(N,T) \le 0$. In accordance with Eq. (28), we have

$$
\begin{aligned}
& [f(2,3)g(3) - f(3,2)g(2)]\,(1-\gamma) \\
& + (N-3)[f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2)](1-\gamma) \\
& + (T-3)[f(2,3)g(3) - f(3,2)g(2) - f(2,2)g(2) + f(3,1)g(1)](1-\gamma) \\
& + c_n - c_t \le 0.
\end{aligned} \tag{33}
$$

When Eq. (33) holds, we should consider three cases:

(i) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) > 0$ (*i.e.*, Eq. (30) is satisfied), then we can get Eq. (29);
(ii) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) < 0$ (*i.e.*, Eq. (32) is satisfied), then Eq. (31) is obtained;
and (iii) if $f(2,3)g(3) - f(3,2)g(2) - f(1,3)g(3) + f(2,2)g(2) = 0$, this is meaningless for our investigated problem.

Thus, Theorem 3 is proved.

Theorem 4. *Suppose that one party owns N training instances and T attributes. Compared with collecting one more training instance, collecting one more attribute can generate a higher utility increase in transfer learning when any one of the following two conditions holds:*

(i) *Condition 1: N, T, c_n, c_t, γ and τ satisfy Eq. (34) and Eq. (35);*

(ii) *Condition 2: N, T, c_n, c_t, γ and τ satisfy Eq. (36) and Eq. (37).*

$$\begin{aligned} N > &-(T-3)\frac{f(2,3)g(3)-f(3,2)g(2)-f(2,2)g(2)+f(3,1)g(1)}{f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)} \\ &-\frac{f(2,3)g(3)-f(3,2)g(2)}{f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)} \\ &+\frac{c_t-c_n}{[f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)](1-\gamma)}+3. \end{aligned} \tag{34}$$

$$f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)>0. \tag{35}$$

$$\begin{aligned} N < &-(T-3)\frac{f(2,3)g(3)-f(3,2)g(2)-f(2,2)g(2)+f(3,1)g(1)}{f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)} \\ &-\frac{f(2,3)g(3)-f(3,2)g(2)}{f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)} \\ &+\frac{c_t-c_n}{[f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)](1-\gamma)}+3. \end{aligned} \tag{36}$$

$$f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)<0. \tag{37}$$

Note that $f(2,3)=(1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$, $f(3,2)=(1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,2)=(1-\frac{2}{e^{\frac{\tau^2}{2}}})^2$, $f(3,1)=(1-\frac{2}{e^{3\tau^2}})$, $f(1,3)=(1-\frac{2}{e^{\frac{\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(1)=\mathrm{BinoCDF}(\tau;1,\frac{\tau}{T})$, $g(2)=\mathrm{BinoCDF}(\tau;2,\frac{\tau}{T})$, $g(3)=\mathrm{BinoCDF}(\tau;3,\frac{\tau}{T})$ *according to Eq. (6).*

Proof. Collecting one more attribute is better, which indicates $\tilde{k}(N,T)>0$. From Eq. (28), we have

$$\begin{aligned} &[f(2,3)g(3)-f(3,2)g(2)](1-\gamma) \\ &+(N-3)[f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)](1-\gamma) \\ &+(T-3)[f(2,3)g(3)-f(3,2)g(2)-f(2,2)g(2)+f(3,1)g(1)](1-\gamma) \\ &+c_n-c_t>0. \end{aligned} \tag{38}$$

When Eq. (38) holds, we should take into account three cases:

(i) if $f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)>0$ (*i.e.*, Eq. (35) is satisfied), then we can get Eq. (34);
(ii) if $f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)<0$ (*i.e.*, Eq. (37) is satisfied), then Eq. (36) is obtained;
and (iii) if $f(2,3)g(3)-f(3,2)g(2)-f(1,3)g(3)+f(2,2)g(2)=0$, this is meaningless for our investigated problem.

Therefore, Theorem 4 is proved.

5. Whether to collaboration

Similar to Section 4, both the benefit comparison (that purely implies transfer learning performance) and utility comparison (that considers both benefit and cost) are conducted to analyze whether collaboration is beneficial to improve transfer learning.

5.1. Benefit comparison

We can get a bivar linear function $\tilde{P}(N,T)$ for approximating $P(N,T)$ by employing Taylor Theorem [43] and Newton's forward interpolation formula [44]. In addition, the approximated $P(N,T)$ at the point $(3,3)$, where the gradient vanish can be avoided for $\tilde{P}(N,T)$, can be obtained in Eq. (39).

$$\tilde{P}(N,T)=P(3,3)+(N-3)\Delta P_N(3,3)+(T-3)\Delta P_T(3,3). \tag{39}$$

By Eq. (7) and Newton's forward interpolation formula [44], we have

$$\begin{aligned} \Delta P_N(3,3) &= \frac{P(3,3)-P(2,3)}{3-2}=P(3,3)-P(2,3) \\ &= f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma), \end{aligned} \tag{40}$$

and

$$\begin{aligned}\Delta P_T(3,3) &= \frac{P(3,3)-P(3,2)}{3-2} = P(3,3)-P(3,2)\\ &= f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma).\end{aligned} \tag{41}$$

By substituting Eq. (40) and Eq. (41), we mathematically re-describe $\tilde{P}(N,T)$ in Eq. (42).

$$\begin{aligned}\tilde{P}(N,T) =& P(3,3) + (N-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\\ &+ (T-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{42}$$

We investigate a two-party collaboration scenario, consisting of party **A** and party **B**, to study whether collaboration can bring the improvement of transfer learning performance. Without loss of generality, we assume that N_1 training instances and T_1 attributes have been collected by party **A**, and N_2 training instances and T_2 attributes have been collected by party **B**. Without collaboration, based on Eq. (42), party **A**'s transfer learning performance $\tilde{P}(N_1,T_1)$ in Eq. (43) can be estimated via its collected training instances and attributes.

$$\begin{aligned}\tilde{P}(N_1,T_1) =& P(3,3) + (N_1-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\\ &+ (T_1-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{43}$$

Suppose that the number of overlapping instances between party **A** and party **B** is n and the number of overlapping attributes between party **A** and party **B** is t. Accordingly, with collaboration, based on Eq. (42), party **A** should achieve the transfer learning performance $\tilde{P}(N_1+N_2-n, T_1+T_2-t)$:

$$\begin{aligned}&\tilde{P}(N_1+N_2-n, T_1+T_2-t)\\ =& P(3,3) + (N_1+N_2-n-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\\ &+ (T_1+T_2-t-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{44}$$

If $\tilde{P}(N_1+N_2-n, T_1+T_2-t) \geq \tilde{P}(N_1,T_1)$, we can say that party **A** can obtain benefit for enhancing transfer learning performance through collaboration; otherwise, party **A** cannot get the increased benefit of transfer learning performance by participating collaboration.

Theorem 5. *In a two-party collaboration, party* **A** *owns N_1 training instances and T_1 attributes, and party* **B** *owns N_2 training instances and T_2 attributes. Party* **A** *can increase transfer learning performance through collaboration, when N_2, T_2, n, t, and τ satisfy Eq. (45).*

$$N_2 \geq -(T_2-t)\frac{f(3,3)g(3)-f(3,2)g(2)}{f(3,3)g(3)-f(2,3)g(3)} + n, \tag{45}$$

where $f(3,3) = (1-\frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ *according to Eq. (6).*

Proof. For party **A**, collaboration is beneficial, which means $\tilde{P}(N_1+N_2-n, T_1+T_2-t) \geq \tilde{P}(N_1,T_1)$. Eq. (46) can be obtained with the substitution of Eq. (43) and Eq. (44),

$$\begin{aligned}&P(3,3) + (N_1+N_2-n-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\\ &+ (T_1+T_2-t-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \geq\\ &P(3,3) + (N_1-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\\ &+ (T_1-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{46}$$

One can easily see that $[f(3,3)g(3) - f(2,3)g(3)] > 0$ as $f(3,3) > f(2,3)$ and $g(3) > 0$. Then, Eq. (45) is the solution of Eq. (46). Therefore, Theorem 5 is proved.

Theorem 6. *In a two-party collaboration, party* **A** *owns N_1 training instances and T_1 attributes, and party* **B** *owns N_2 training instances and T_2 attributes. For party* **A***, collaborative transfer learning cannot help improve the transfer learning performance if N_2, T_2, n, t and τ satisfy Eq. (47).*

$$N_2 < -(T_2-t)\frac{f(3,3)g(3)-f(3,2)g(2)}{f(3,3)g(3)-f(2,3)g(3)} + n, \tag{47}$$

where $f(3,3) = (1-\frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ *according to Eq. (6).*

Proof. With respect to party **A**, collaborative transfer learning fails to benefit transfer learning performance, so $\tilde{P}(N_1+N_2-n, T_1+T_2-t) < \tilde{P}(N_1, T_1)$. Accordingly, we can obtain Eq. (48) based on Eq. (43) and Eq. (44).

$$\begin{aligned}
&P(3,3)+(N_1+N_2-n-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_1+T_2-t-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]<\\
&P(3,3)+(N_1-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_1-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)].
\end{aligned} \tag{48}$$

By observing Eq. (48), it is obvious that $[f(3,3)g(3)-f(2,3)g(3)]>0$ due to $f(3,3)>f(2,3)$ and $g(3)>0$. Then, by solving Eq. (48), we can get the solution as shown in Eq. (47); that is, Theorem 6 is proved.

5.2. Utility comparison

Although collaborative transfer learning can help improve transfer learning performance through sharing data, there still exists a threat of privacy leakage during collaboration since the shared data can be used to mine various sensitive/private information. Thus, with the dilemma between the benefit of collaboration and the cost of privacy leakage, it is rational for two parties to decide data sharing strategies. For party **A**, the ratio of shared instances is denoted as $\alpha_1 \in [0,1]$, and the ratio of shared attributes as $\beta_1 \in [0,1]$; similarly, for party **B**, the ratio of shared instances and the ratio of shared attributes are represented by $\alpha_2 \in [0,1]$ and $\beta_2 \in [0,1]$, respectively.

As a matter of fact, if party **A** collaborates with party **B**, party **A** suffers not only the collection cost but also the privacy cost. (i) **Collection Cost:** Considering the cost of collecting one instance c_n and the cost of collecting one attribute is c_t, party **A**'s collection cost is $(c_n\alpha_1N_1+c_t\beta_1T_1)$ in collaboration. (ii) **Privacy Cost:** With party **A**'s shared data, party **B** can launch inference attack to reveal party **A**'s private information through their shared instances and attributes [49,50]. Thus, the inference performance can be used to measure party **A**'s privacy cost that depends on the number of instances and attributes. More specifically, the privacy cost yielded by the instances is defined as $c_{np}N^q$ $(0<q<1)$ [51,52], and the privacy cost yielded by the attributes is defined to be $c_{tp}T$ [53,54]. Additionally, we suppose that in the two-party collaboration, the number of the overlapping instances between two parties during sharing is n_α and the number of the overlapping attributes between two parties during sharing is t_β. Accordingly, for party **A**, the privacy cost caused by the shared instances is $c_{np}(\alpha_1N_1+\alpha_2N_2-n_\alpha)^q$, and the privacy cost caused by the shared attributes is $c_{tp}(\beta_1T_1+\beta_2T_2-t_\beta)$.

Then, we can define the utility of collaboration $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)$ for party **A** as

$$\begin{aligned}
&U(\alpha_1N_1+\alpha_2N_2-n_\alpha,\beta_1T_1+\beta_2T_2-t_\beta)\\
=&\tilde{P}(\alpha_1N_1+\alpha_2N_2-n_\alpha,\beta_1T_1+\beta_2T_2-t_\beta)-(c_n\alpha_1N_1+c_t\beta_1T_1)\\
&-c_{np}(\alpha_1N_1+\alpha_2N_2-n_\alpha)^q-c_{tp}(\beta_1T_1+\beta_2T_2-t_\beta).
\end{aligned} \tag{49}$$

According to Eq. (42), we can further rewrite $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)$ as

$$\begin{aligned}
&U(\alpha_1N_1+\alpha_2N_2-n_\alpha,\beta_1T_1+\beta_2T_2-t_\beta)\\
=&P(3,3)+(\alpha_1N_1+\alpha_2N_2-n_\alpha-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(\beta_1T_1+\beta_2T_2-t_\beta-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]\\
&-(c_n\alpha_1N_1+c_t\beta_1T_1)-c_{np}(\alpha_1N_1+\alpha_2N_2-n_\alpha)^q-c_{tp}(\beta_1T_1+\beta_2T_2-t_\beta).
\end{aligned} \tag{50}$$

On the other hand, if party **A** does not collaborate with party **B**, the sharing strategy does not need to be considered. Thus, for party **A**, the collection cost is $(c_nN_1+c_tT_1)$, the privacy cost caused by party **B**'s instances without collaboration is $c_{np}N_2^q$, and the privacy cost caused by party **B**'s attributes without collaboration is $c_{tp}T_2$. Then, without the two-party collaboration, we can define party **A**'s utility $U(N_1,T_1)$ in Eq. (51).

$$U(N_1,T_1)=\tilde{P}(N_1,T_1)-(c_nN_1+c_tT_1)-(c_{np}N_2^q+c_{tp}T_2). \tag{51}$$

With the substitution of Eq. (43), we can rewrite $U(N_1,T_1)$ in Eq. (52).

$$\begin{aligned}
U(N_1,T_1)=&P(3,3)+(N_1-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_1-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]\\
&-(c_nN_1+c_tT_1)-(c_{np}N_2^q+c_{tp}T_2).
\end{aligned} \tag{52}$$

If $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta) \geq U(N_1,T_1)$, collaboration can help party **A** improve transfer learning performance; otherwise, collaboration cannot make any improvement. In order to check whether collaboration is beneficial for party **A** to enhance transfer learning performance, we compare $U(N_1,T_1)$ with $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)$.

Theorem 7. *In a two-party collaboration, party **A** has N_1 training instances and T_1 attributes, and party **B** has N_2 training instances and T_2 attributes. When N_1, N_2, T_1, T_2, n_α, t_β, α_1, α_2, β_1, β_2, γ, q and τ satisfy Eq. (53), collaboration can help party **A** increase the utility of transfer learning.*

$$\begin{aligned}
N_2 \geq &-(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_1)\frac{f(3,3)g(3) - f(3,2)g(2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)]} \\
&+ \frac{c_n \alpha_1 N_1 + c_t \beta_1 T_1 - c_n N_1 - c_t T_1}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{c_{np}\left[(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - N_2^q\right] + c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{N_1 - \alpha_1 N_1 + n_\alpha}{\alpha_2}.
\end{aligned} \tag{53}$$

Notice that $f(3,3) = (1 - \frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ *according to Eq. (6).*

Proof. Collaboration is better for party **A**, which indicates $U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta) \geq U(N_1, T_1)$. By replacing Eq. (52) and Eq. (50), we gain Eq. (54).

$$\begin{aligned}
&P(3,3) + (\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (\beta_1 T_1 + \beta_2 T_2 - t_\beta - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n \alpha_1 N_1 + c_t \beta_1 T_1) - c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta) \geq \\
&P(3,3) + (N_1 - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (T_1 - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n N_1 + c_t T_1) - (c_{np} N_2^q + c_{tp} T_2).
\end{aligned} \tag{54}$$

Since $f(3,3) > f(2,3)$, $g(3) > 0$ and $(1-\gamma) > 0$, we have $[f(3,3)g(3) - f(2,3)g(3)] > 0$. Then, Eq. (53) is obtained by solving Eq. (54). Therefore, Theorem 7 is proved.

Theorem 8. *In a two-party collaboration, party **A** has N_1 training instances and T_1 attributes, and party **B** has N_2 training instances and T_2 attributes. If N_1, N_2, T_1, T_2, n_α, t_β, α_1, α_2, β_1, β_2, γ, q and τ satisfy Eq. (55), collaboration cannot enhance the utility of transfer learning for party **A**.*

$$\begin{aligned}
N_2 < &-(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_1)\frac{f(3,3)g(3) - f(3,2)g(2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)]} \\
&+ \frac{c_n \alpha_1 N_1 + c_t \beta_1 T_1 - c_n N_1 - c_t T_1}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{c_{np}\left[(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - N_2^q\right] + c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{N_1 - \alpha_1 N_1 + n_\alpha}{\alpha_2}.
\end{aligned} \tag{55}$$

Notice that $f(3,3) = (1 - \frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$ *according to Eq. (5), and* $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ *according to Eq. (6).*

Proof. If collaboration fails to increase party **A**'s utility, there is $U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta) < U(N_1, T_1)$. Accordingly, Eq. (56) can be gained using Eq. (52) and Eq. (50).

$$\begin{aligned}
&P(3,3) + (\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (\beta_1 T_1 + \beta_2 T_2 - t_\beta - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n \alpha_1 N_1 + c_t \beta_1 T_1) - c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta) < \\
&P(3,3) + (N_1 - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (T_1 - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n N_1 + c_t T_1) - (c_{np} N_2^q + c_{tp} T_2).
\end{aligned} \tag{56}$$

In Eq. (56), $[f(3,3)g(3) - f(2,3)g(3)](1-\gamma) > 0$ because $f(3,3) > f(2,3)$, $g(3) > 0$ and $(1-\gamma) > 0$. Then, the solution to Eq. (56) is Eq. (55). Hence, Theorem 8 is proved.

6. Collection vs. collaboration

In this section, to study the effectiveness of collection and collaboration, our comparison starts from only considering collection/collaboration benefit (that directly reflects transfer learning performance) and then considers collection/collaboration utility (that comprehensively reflects the benefit and the cost of transfer learning).

6.1. Benefit comparison

Still in a two-party collaboration scenario, N_1 training instances and T_1 attributes are owned by party **A**, and N_2 training instances and T_2 attributes are owned by party **B**. On the one hand, with collaboration with party **B**, the transfer learning performance will become $\tilde{P}(N_1+N_2-n, T_1+T_2-t)$ for party **A**. On the other hand, for party **A**, it is reasonable for us to make an assumption that the training instances can be increased from N_1 to N_{NT} and the attributes can be increased from T_1 to T_{NT} with the help of collection. As a result, for party **A**, the transfer learning performance can be increased to $\tilde{P}(N_{NT}, T_{NT})$ as presented in Eq. (57).

$$\begin{aligned}\tilde{P}(N_{NT}, T_{NT}) =& P(3,3) + (N_{NT}-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\ &+ (T_{NT}-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{57}$$

If $\tilde{P}(N_1+N_2-n, T_1+T_2-t) \geq \tilde{P}(N_{NT}, T_{NT})$, collaboration can bring a larger improvement of transfer learning performance than collection for party **A**; otherwise, collection can obtain a higher enhancement of transfer learning performance.

Theorem 9. *In a two-party collaboration, party **A** has N_1 training instances and T_1 attributes, and party **B** has N_2 training instances and T_2 attributes. Assume that party **A** can increase N_1 to N_{NT} and increase T_1 to T_{NT} via collection. Collaboration can bring a higher increase of transfer learning performance than collection for party **A**, when N_1, N_2, N_{NT}, T_1, T_2, T_{NT}, n, t and τ satisfy Eq. (58).*

$$N_2 \geq -(T_1+T_2-t-T_{NT})\frac{f(3,3)g(3)-f(3,2)g(2)}{f(3,3)g(3)-f(2,3)g(3)} + n + N_{NT} - N_1, \tag{58}$$

where $f(3,3) = (1-\frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, and $f(2,3) = (1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$ according to Eq. (5), and $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ according to Eq. (6).

Proof. If collaboration is better than collection, there should be

$$\tilde{P}(N_1+N_2-n, T_1+T_2-t) \geq \tilde{P}(N_{NT}, T_{NT}),$$

which can be rewritten as Eq. (59) according to Eq. (44) and Eq. (57).

$$\begin{aligned}&P(3,3) + (N_1+N_2-n-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\ &+ (T_1+T_2-t-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \geq \\ &P(3,3) + (N_{NT}-3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\ &+ (T_{NT}-3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)].\end{aligned} \tag{59}$$

Notice that $[f(3,3)g(3) - f(2,3)g(3)] > 0$ because $f(3,3) > f(2,3)$ and $g(3) > 0$. Then, we can get Eq. (58) when solving Eq. (59). Thus, Theorem 9 is proved.

Theorem 10. *In a two-party collaboration, party **A** has N_1 training instances and T_1 attributes, and party **B** has N_2 training instances and T_2 attributes. Assume that party **A** can increase N_1 to N_{NT} and increase T_1 to T_{NT} via collection. Collection can help enhance transfer learning performance more than collaboration for party **A**, when N_1, N_2, N_{NT}, T_1, T_2, T_{NT}, n, t and τ satisfy Eq. (60).*

$$N_2 < -(T_1+T_2-t-T_{NT})\frac{f(3,3)g(3)-f(3,2)g(2)}{f(3,3)g(3)-f(2,3)g(3)} + n + N_{NT} - N_1, \tag{60}$$

where $f(3,3) = (1-\frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$ according to Eq. (5), and $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ according to Eq. (6).

Proof. When collection becomes more beneficial for party **A**, we have $\tilde{P}(N_1+N_2-n, T_1+T_2-t) < \tilde{P}(N_{NT}, T_{NT})$. Based on Eq. (44) and Eq. (57), the inequality $\tilde{P}(N_1+N_2-n, T_1+T_2-t) < \tilde{P}(N_{NT}, T_{NT})$ can be equivalently presented as:

$$\begin{aligned}
&P(3,3)+(N_1+N_2-n-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_1+T_2-t-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]<\\
&P(3,3)+(N_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)].
\end{aligned} \tag{61}$$

In Eq. (61), it can be found $[f(3,3)g(3)-f(2,3)g(3)]>0$ due to $f(3,3)>f(2,3)$ and $g(3)>0$. Then, we have Eq. (60) by solving Eq. (61) and thus prove Theorem 10.

6.2. Utility comparison

When party **A** increases N_1 instances to N_{NT} instances and increases T_1 attributes to T_{NT} attributes via collection, the corresponding utility $U(N_{NT}, T_{NT})$ can be defined in Eq. (62).

$$U(N_{NT},T_{NT})=\tilde{P}(N_{NT},T_{NT})-(c_nN_{NT}+c_tT_{NT})-(c_{np}N_2^q+c_{tp}T_2). \tag{62}$$

Then, with the substitution of Eq. (57), we can rewrite $U(N_{NT}, T_{NT})$ in Eq. (63).

$$\begin{aligned}
U(N_{NT},T_{NT})=&P(3,3)+(N_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]\\
&-(c_nN_{NT}+c_tT_{NT})-(c_{np}N_2^q+c_{tp}T_2).
\end{aligned} \tag{63}$$

On the other hand, for party **A**, the utility of transfer learning in collaboration is $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)$. Thus, by comparing $U(N_{NT}, T_{NT})$ and $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)$, we can determine which option between collection and collaboration is a better way to increase the utility of transfer learning.

Theorem 11. *In a two-party collaboration, party **A** has N_1 training instances and T_1 attributes, and party **B** has N_2 training instances and T_2 attributes. Assume that party **A** can increase N_1 to N_{NT} and increase T_1 to T_{NT} via collection. If N_1, N_2, N_{NT}, T_1, T_2, T_{NT}, α_1, α_2, β_1, β_2, n_α, t_β, γ, q and τ satisfy Eq. (64), compared with collection, collaboration can bring a larger utility increase in transfer learning for party **A**.*

$$\begin{aligned}
N_2\geq&-(\beta_1T_1+\beta_2T_2-t_\beta-T_{NT})\frac{f(3,3)g(3)-f(3,2)g(2)}{\alpha_2[f(3,3)g(3)-f(2,3)g(3)]}\\
&+\frac{c_n\alpha_1N_1+c_t\beta_1T_1-c_nN_{NT}-c_tT_{NT}}{\alpha_2[f(3,3)g(3)-f(2,3)g(3)](1-\gamma)}\\
&+\frac{c_{np}\left[(\alpha_1N_1+\alpha_2N_2-n_\alpha)^q-N_2^q\right]+c_{tp}(\beta_1T_1+\beta_2T_2-t_\beta-T_2)}{\alpha_2[f(3,3)g(3)-f(2,3)g(3)](1-\gamma)}\\
&+\frac{N_{TN}-\alpha_1N_1+n_\alpha}{\alpha_2}.
\end{aligned} \tag{64}$$

Note that $f(3,3)=(1-\frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2)=(1-\frac{2}{e^{\frac{3\tau^2}{4}}})^2$, and $f(2,3)=(1-\frac{2}{e^{\frac{2\tau^2}{9}}})^3$ according to Eq. (5), and $g(2)=\mathrm{BinoCDF}(\tau;2,\frac{\tau}{T_1})$, $g(3)=\mathrm{BinoCDF}(\tau;3,\frac{\tau}{T_1})$ according to Eq. (6).

Proof. The requirement of collaboration to be more beneficial is $U(\alpha_1N_1+\alpha_2N_2-n_\alpha, \beta_1T_1+\beta_2T_2-t_\beta)\geq U(N_{NT}, T_{NT})$, which can be expressed by Eq. (65) via substituting Eq. (50) and Eq. (63).

$$\begin{aligned}
&P(3,3)+(\alpha_1N_1+\alpha_2N_2-n_\alpha-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(\beta_1T_1+\beta_2T_2-t_\beta-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]\\
&-(c_n\alpha_1N_1+c_t\beta_1T_1)-c_{np}(\alpha_1N_1+\alpha_2N_2-n_\alpha)^q-c_{tp}(\beta_1T_1+\beta_2T_2-t_\beta)\geq\\
&P(3,3)+(N_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)]\\
&+(T_{NT}-3)[f(3,3)g(3)(1-\gamma)-f(3,2)g(2)(1-\gamma)]\\
&-(c_nN_{NT}+c_tT_{NT})-(c_{np}N_2^q+c_{tp}T_2).
\end{aligned} \tag{65}$$

We know that $[f(3,3)g(3)-f(2,3)g(3)](1-\gamma)>0$ as $f(3,3)>f(2,3)$, $g(3)>0$ and $(1-\gamma)>0$. Then, the solution to Eq. (65) is Eq. (64); that is, Theorem 11 is proved.

Theorem 12. *In a two-party collaboration, party* **A** *has N_1 training instances and T_1 attributes, and party* **B** *has N_2 training instances and T_2 attributes. Assume that party* **A** *can increase N_1 to N_{NT} and increase T_1 to T_{NT} via collection. If N_1, N_2, N_{NT}, T_1, T_2, T_{NT}, α_1, α_2, β_1, β_2, n_α, t_β, γ, q and τ satisfy Eq. (66), compared with collaboration, collection is better for party* **A** *to enhance the utility of transfer learning.*

$$
\begin{aligned}
N_2 <& -(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_{NT}) \frac{f(3,3)g(3) - f(3,2)g(2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)]} \\
&+ \frac{c_n \alpha_1 N_1 + c_t \beta_1 T_1 - c_n N_{NT} - c_t T_{NT}}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{c_{np}\left[(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - N_2^q\right] + c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta - T_2)}{\alpha_2[f(3,3)g(3) - f(2,3)g(3)](1-\gamma)} \\
&+ \frac{N_{TN} - \alpha_1 N_1 + n_\alpha}{\alpha_2}.
\end{aligned} \tag{66}
$$

Note that $f(3,3) = (1 - \frac{2}{e^{\frac{\tau^2}{3}}})^3$, $f(3,2) = (1 - \frac{2}{e^{\frac{3\tau^2}{4}}})^2$, $f(2,3) = (1 - \frac{2}{e^{\frac{2\tau^2}{9}}})^3$ according to Eq. (5), and $g(2) = \mathrm{BinoCDF}(\tau; 2, \frac{\tau}{T_1})$, $g(3) = \mathrm{BinoCDF}(\tau; 3, \frac{\tau}{T_1})$ according to Eq. (6).

Proof.

$$
U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta) < U(N_{NT}, T_{NT})
$$

is the condition of collection to be more beneficial. By substituting Eq. (50) and Eq. (63), we obtain Eq. (67) as follows.

$$
\begin{aligned}
&P(3,3) + (\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (\beta_1 T_1 + \beta_2 T_2 - t_\beta - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n \alpha_1 N_1 + c_t \beta_1 T_1) - c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q - c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta) < \\
&P(3,3) + (N_{NT} - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (T_{NT} - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n N_{NT} + c_t T_{NT}) - (c_{np} N_2^q + c_{tp} T_2).
\end{aligned} \tag{67}
$$

Due to $f(3,3) > f(2,3)$, $g(3) > 0$ and $(1-\gamma) > 0$, we have $[f(3,3)g(3) - f(2,3)g(3)](1-\gamma) > 0$. Then, for Eq. (67), the solution is Eq. (66). Therefore, Theorem 12 is proved.

7. Collaboration strategy analysis

In this section, to comprehensively understand how to determine the data sharing strategy for party **A** to increase the utility of transfer learning in a two-party collaboration, the instance sharing strategy (*i.e.*, α_1 and α_2) and the attribute sharing strategy (*i.e.*, β_1 and β_2) are studied. That is, we analyze the influence of α_1, α_2, β_1, and β_2 on $U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta)$. For presentation simplicity in the following, we use U_{clb} to represent $U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta)$ for short as shown in Eq. (68).

$$
\begin{aligned}
U_{clb} =& U(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha, \beta_1 T_1 + \beta_2 T_2 - t_\beta) \\
=& (\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\
&+ (\beta_1 T_1 + \beta_2 T_2 - t_\beta - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\
&- (c_n \alpha_1 N_1 + c_t \beta_1 T_1) - c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^q \\
&- c_{tp}(\beta_1 T_1 + \beta_2 T_2 - t_\beta).
\end{aligned} \tag{68}
$$

7.1. Influence of instance sharing strategy

With respect to α_1, we calculate its first-order partial derivative $\frac{\partial U_{clb}}{\partial \alpha_1}$ in Eq. (69) and its second-order partial derivative $\frac{\partial^2 U_{clb}}{\partial^2 \alpha_1}$ in Eq. (70).

$$
\begin{aligned}
\frac{\partial U_{clb}}{\partial \alpha_1} =& [f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]N_1 \\
&- c_n N_1 - q c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^{q-1} N_1.
\end{aligned} \tag{69}
$$

$$\frac{\partial^2 U_{clb}}{\partial^2 \alpha_1} = -q(q-1)c_{np}N_1^2(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha)^{q-2}. \tag{70}$$

Since $0 < q < 1$ and $(\alpha_1 N_1 + \alpha_2 N_2 - n_\alpha) \geq 0$, $\frac{\partial^2 U_{clb}}{\partial^2 \alpha_1} \geq 0$ can always hold, which means that U_{clb} has a minimum value. According to Eq. (69), we obtain the minimum value point α_1^* in Eq. (71) by solving $\frac{\partial U_{clb}}{\partial \alpha_1} = 0$.

$$\alpha_1^* = \frac{\left(\frac{[f(3,3)g(3)(1-\gamma)-f(2,3)g(3)(1-\gamma)-c_n]}{qc_{np}}\right)^{\frac{1}{q-1}} + n_\alpha - \alpha_2 N_2}{N_1}. \tag{71}$$

Thus, in a two-party collaboration, the instance sharing ratio α_1^* is the strategy providing a lower bound of party **A**'s received utility, which indicates that collaboration can guarantee one party's lowest received transfer learning utility. Furthermore, we can reach two conclusions as follows.

(i) When $\alpha_1 \in [0, \alpha_1^*)$, $\frac{\partial U_{clb}}{\partial \alpha_1} < 0$, which indicates that the utility of transfer learning decreases with the increase of instances shared from party **A**. That is, if party **A** shares a relatively smaller portion of instances, the increase in the benefit of transfer learning is smaller than the increase in the cost of transfer learning.
(ii) When $\alpha_1 \in [\alpha_1^*, 1]$, $\frac{\partial U_{clb}}{\partial \alpha_1} \geq 0$, which implies that the utility of transfer learning is enhanced with the increase of instances shared by party **A**. In other words, if party **A** shares a relatively larger portion of instances, the increase in the benefit of transfer learning is larger than the increase in the cost of transfer learning.

By observing Eq. (68), we can see that α_1 and α_2 are symmetrical in the mathematical expression with respect to U_{clb}. Thus, with the same analysis, we can also obtain the same conclusions about the effect of α_2 on party **B**'s utility.

7.2. Influence of attribute sharing strategy

Regarding β_1, we calculate the first-order partial derivative $\frac{\partial U_{clb}}{\partial \beta_1}$ in Eq. (72).

$$\frac{\partial U_{clb}}{\partial \beta_1} = [f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma) - c_t - c_{tp}]T_1. \tag{72}$$

Correspondingly, we obtain the following critical findings in a two-party collaboration.

(i) When $[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma) - c_t - c_{tp}] < 0$, we can calculate that $\frac{\partial U_{clb}}{\partial \beta_1} < 0$. For party **A**, the increase of shared attributes (*i.e.*, the increase of $\beta_1 \in [0, 1]$) leads the decrease in the utility of transfer learning. The reason is that the increase in the benefit of transfer learning is slower than the increase in the cost of transfer learning.
(ii) When $[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma) - c_t - c_{tp}] \geq 0$, $\frac{\partial U_{clb}}{\partial \beta_1} \geq 0$. The increase of shared attributes (*i.e.*, the increase of $\beta_1 \in [0, 1]$) can promote the utility of transfer learning for party **A**, because the increase in the benefit of transfer learning is faster than the increase in the cost of transfer learning.

Similarly, β_1 and β_2 are symmetrical with respect to U_{clb} in Eq. (68), we therefore can draw the same conclusions about how β_2 affects the utility of transfer learning for party **B**.

7.3. Influence of overlapping data

Let $o_\alpha = \frac{n_\alpha}{\alpha_1 N_1}$ and $o_\beta = \frac{t_\beta}{\beta_1 T_1}$ ($o_\alpha, o_\beta \in [0, 1]$), which are the ratios of overlapping data volume to party **A**'s shared data volume in terms of instance and attribute, respectively. Then, by replacing n_α by $o_\alpha \alpha_1 N_1$ and replacing t_β by $o_\beta \beta_1 T_1$, we can rewrite U_{clb} as below:

$$\begin{aligned} U_{clb} =& (\alpha_1 N_1 + \alpha_2 N_2 - o_\alpha \alpha_1 N_1 - 3)[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)] \\ &+ (\beta_1 T_1 + \beta_2 T_2 - o_\beta \beta_1 T_1 - 3)[f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \\ &- (c_n \alpha_1 N_1 + c_t \beta_1 T_1) - c_{np}(\alpha_1 N_1 + \alpha_2 N_2 - o_\alpha \alpha_1 N_1)^q \\ &- c_{tp}(\beta_1 T_1 + \beta_2 T_2 - o_\beta \beta_1 T_1). \end{aligned} \tag{73}$$

In order to study the influence of overlapping shared instances on the utility of transfer learning in a two-party collaboration, we compute o_α's first-order partial derivative $\frac{\partial U_{clb}}{\partial o_\alpha}$ and second-order partial derivative $\frac{\partial^2 U_{clb}}{\partial^2 o_\alpha}$ in Eq. (74) and Eq. (75), respectively.

$$\begin{aligned} \frac{\partial U_{clb}}{\partial o_\alpha} =& -[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]\alpha_1 N_1 \\ &+ qc_{np}\alpha_1 N_1(\alpha_1 N_1 + \alpha_2 N_2 - o_\alpha \alpha_1 N_1)^{q-1}. \end{aligned} \tag{74}$$

$$\frac{\partial^2 U_{clb}}{\partial^2 o_\alpha} = -q(q-1)c_{np}\alpha_1^2 N_1^2(\alpha_1 N_1 + \alpha_2 N_2 - o_\alpha \alpha_1 N_1)^{q-2}. \tag{75}$$

Clearly, $\frac{\partial^2 U_{clb}}{\partial^2 o_\alpha} \geq 0$ can always be satisfied as $0 < q < 1$ and $(\alpha_1 N_1 + \alpha_2 N_2 - o_\alpha \alpha_1 N_1) \geq 0$, which implies that there is the minimum value of U_{clb}. And the minimum value point o_α^* should be the solution of $\frac{\partial U_{clb}}{\partial o_\alpha} = 0$, as presented in Eq. (76).

$$o_\alpha^* = \frac{\alpha_1 N_1 + \alpha_2 N_2 - (\frac{[f(3,3)g(3)(1-\gamma) - f(2,3)g(3)(1-\gamma)]}{qc_{np}})^{\frac{1}{q-1}}}{\alpha_1 N_1}. \tag{76}$$

The existence of o_α^* means in a two-party collaboration, there is a lower bound of party **A**'s utility; that is, collaboration can ensure one party's received transfer learning utility. Besides, two significant conclusions can be reached.

(i) If $o_\alpha \in [0, o_\alpha^*)$, then $\frac{\partial U_{clb}}{\partial o_\alpha} < 0$, which suggests that the increase of overlapping instances can cause the decrease of the utility of transfer learning. For party **A**, when there is a relatively smaller overlap between two shared instance datasets, the increase in the benefit of transfer learning does not exceed the increase in the cost of transfer learning.
(ii) If $o_\alpha \in [o_\alpha^*, 1]$, then $\frac{\partial U_{clb}}{\partial o_\alpha} \geq 0$, which shows that the increasing ratio can bring the increase of the utility of transfer learning. That is to say, for party **A**, when there exists a relatively larger overlap between two shared instance datasets, the increase in the benefit of transfer learning surpasses the increase in the cost of transfer learning.

Moreover, for studying how the overlapping attributes affect the utility of transfer learning in a two-party collaboration, o_β's first-order partial derivative $\frac{\partial U_{clb}}{\partial o_\beta}$ can be calculated as follows:

$$\frac{\partial U_{clb}}{\partial o_\beta} = \beta_1 T_1[c_{tp} - f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)]. \tag{77}$$

Accordingly, we can figure out two crucial discoveries:

(i) When $[c_{tp} - f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] < 0$, we have $\frac{\partial U_{clb}}{\partial o_\beta} < 0$. The increase of the ratio of overlapping attributes (*i.e.*, the increase of $o_\beta \in [0, 1]$) results in the decrease of the utility of transfer learning for party **A**, because the increase in the benefit of transfer learning falls behind the increase in the cost of transfer learning.
(ii) When $[c_{tp} - f(3,3)g(3)(1-\gamma) - f(3,2)g(2)(1-\gamma)] \geq 0$, $\frac{\partial U_{clb}}{\partial o_\beta} \geq 0$ for party **A**. That is, the increase of the ratio of overlapping attributes (*i.e.*, the increase of $o_\beta \in [0, 1]$) can increase the utility of transfer learning for party **A**. The main reason is that the increase in the benefit of transfer learning exceeds the increase in the cost of transfer learning.

8. Conclusion

With the proliferation of transfer learning in real applications, it is imperative to investigate how to promote transfer learning effectively. For this purpose, this paper theoretically analyzes the methods of improving transfer learning performance. Specifically, by comparing two collection approaches, we present theorems to help choose a better collection option. Then, we propose theorems to determine whether collaboration is beneficial for a party to improve transfer learning. What's more, via the comparison between collection and collaboration, theorems are provided to check which one can bring more increase in transfer learning performance. Finally, we analyze how to strategically share data for more improvement in collaborative transfer learning. More importantly, our proposed theorems and conclusions in the analysis process can be exploited as practical decision-making guidance to enhance transfer learning performance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.W. Libbrecht, W.S. Noble, Machine learning applications in genetics and genomics, Nat. Rev. Genet. 16 (6) (2015) 321–332.
[2] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13 (2015) 8–17.
[3] B. Abbasi, D.M. Goldenholz, Machine learning applications in epilepsy, Epilepsia 60 (10) (2019) 2037–2047.
[4] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260.
[5] G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, A. Lopez Garcia, I. Heredia, P. Malík, L. Hluchỳ, Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey, Artif. Intell. Rev. 52 (1) (2019) 77–124.
[6] U. Blanke, B. Schiele, Remember and transfer what you have learned-recognizing composite activities based on activity spotting, in: International Symposium on Wearable Computers, IEEE, 2010, pp. 1–8.

[7] A. Torralba, K.P. Murphy, W.T. Freeman, Sharing visual features for multiclass and multiview object detection, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 854–869.
[8] J. Jiang, C. Zhai, Instance weighting for domain adaptation in nlp, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, ACL, 2007, pp. 264–271.
[9] X. Liao, Y. Xue, L. Carin, Logistic regression with an auxiliary data source, in: Proceedings of the 22nd International Conference on Machine Learning, ACM, 2005, pp. 505–512.
[10] P. Wu, T.G. Dietterich, Improving svm accuracy by training on auxiliary data sources, in: Proceedings of the 21st International Conference on Machine Learning, ACM, 2004, p. 110.
[11] S.-I. Lee, V. Chatalbashev, D. Vickrey, D. Koller, Learning a meta-level prior for feature relevance from multiple related tasks, in: Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 489–496.
[12] T. Jebara, Multi-task feature and kernel selection for svms, in: Proceedings of the 21st International Conference on Machine Learning, ACM, 2004, p. 55.
[13] C. Wang, S. Mahadevan, Manifold alignment using procrustes analysis, in: Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, pp. 1120–1127.
[14] H. Yang, H. He, W. Zhang, X. Cao, Fedsteg: a federated transfer learning framework for secure image steganalysis, IEEE Trans. Netw. Sci. Eng. 14 (2018) 78–88.
[15] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu, C. Guan, Federated transfer learning for eeg signal classification, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 3040–3045.
[16] Y. Chen, X. Qin, J. Wang, C. Yu, W. Gao, Fedhealth: a federated transfer learning framework for wearable healthcare, IEEE Intell. Syst. (2020) 84–94.
[17] H. Daga, P.K. Nicholson, A. Gavrilovska, D. Lugones, Cartel: a system for collaborative transfer learning at the edge, in: Proceedings of the ACM Symposium on Cloud Computing, ACM, 2019, pp. 25–37.
[18] Z. Ma, J. Ma, Y. Miao, K.-K.R. Choo, X. Liu, X. Wang, T. Yang, Pmkt: privacy-preserving multi-party knowledge transfer for financial market forecasting, Future Gener. Comput. Syst. 106 (2020) 545–558.
[19] S. Sharma, C. Xing, Y. Liu, Y. Kang, Secure and efficient federated transfer learning, in: 2019 IEEE International Conference on Big Data, IEEE, 2019, pp. 2569–2576.
[20] A. Qayyum, K. Ahmad, M.A. Ahsan, A.I. Al-Fuqaha, J. Qadir, Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge, CoRR, arXiv:2101.07511 [abs], 2021, arXiv:2101.07511, https://arxiv.org/abs/2101.07511.
[21] I. Kevin, K. Wang, X. Zhou, W. Liang, Z. Yan, J. She, Federated transfer learning based cross-domain prediction for smart manufacturing, IEEE Trans. Ind. Inform. 16 (2021) 4088–4096.
[22] H. Gao, Y. Liu, Fedurr: a federated transfer learning framework for multi-department collaborative urban risk recognition, in: Proceedings of the 7th ACM SIGSPATIAL International Workshop Emergency Management Using GIS, ACM, 2021, pp. 1–5.
[23] L. Xun, J. Zhang, F. Yao, D. Cao, Improved identification of cotton cultivated areas by applying instance-based transfer learning on the time series of modis ndvi, Catena 213 (2022).
[24] T.Q. Lam, D.D. Nguyen, D.T. Nguyen, H.K. Lam, T.H. Cai, S.N. Hoang, H.D. Do, Instance-based transfer learning approach for Vietnamese speech synthesis with very low resource, in: Future of Information and Communication Conference, Springer, 2022, pp. 148–164.
[25] K. Khalil, U. Asgher, Y. Ayaz, Novel fnirs study on homogeneous symmetric feature-based transfer learning for brain–computer interface, Sci. Rep. 12 (1) (2022) 1–12.
[26] A. Argyriou, T. Evgeniou, M. Pontil, Convex multi-task feature learning, Mach. Learn. 73 (2008) 243–272.
[27] R. Raina, A. Battle, H. Lee, B. Packer, A.Y. Ng, Self-taught learning: transfer learning from unlabeled data, in: Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 759–766.
[28] N.D. Lawrence, J.C. Platt, Learning to learn with the informative vector machine, in: Proceedings of the 21st International Conference on Machine Learning, ACM, 2004, p. 65.
[29] A. Schwaighofer, V. Tresp, K. Yu, Learning Gaussian process kernels via hierarchical Bayes, in: Advances in Neural Information Processing Systems, MIT Press, 2005, pp. 1209–1216.
[30] T. Evgeniou, M. Pontil, Regularized multi–task learning, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 109–117.
[31] L. Mihalkova, T. Huynh, R.J. Mooney, Mapping and Revising Markov Logic Networks for Transfer Learning, AAAI, vol. 7, AAAI, 2007, pp. 608–614.
[32] J. Davis, P. Domingos, Deep transfer via second-order Markov logic, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 217–224.
[33] H.A. Ogoe, S. Visweswaran, X. Lu, V. Gopalakrishnan, Knowledge transfer via classification rules using functional mapping for integrative modeling of gene expression data, BMC Bioinform. 16 (2015) 1–15.
[34] A. Farhadi, D. Forsyth, R. White, Transfer learning in sign language, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[35] C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, F. Provost, Machine learning for targeted display advertising: transfer learning in action, Mach. Learn. 95 (2014) 103–127.
[36] K. Hsieh, A. Harlap, N. Vijaykumar, D. Konomis, G.R. Ganger, P.B. Gibbons, O. Mutlu, Gaia: geo-distributed machine learning approaching {LAN} speeds, in: 14th {USENIX} Symposium on Networked Systems Design and Implementation, USENIX, 2017, pp. 629–647.
[37] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, L. Tang, Neurosurgeon: collaborative intelligence between the cloud and mobile edge, Comput. Archit. News 45 (2017) 615–629.
[38] Y. Liu, Y. Kang, C. Xing, T. Chen, Q. Yang, A secure federated transfer learning framework, IEEE Intell. Syst. 35 (4) (2020) 70–82.
[39] Z. Zhang, N. He, D. Li, H. Gao, T. Gao, C. Zhou, Federated transfer learning for disaster classification in social computing networks, J. Saf. Sci. Resil. 3 (1) (2022) 15–23.
[40] M.J. Kearns, U.V. Vazirani, U. Vazirani, An Introduction to Computational Learning Theory, MIT Press, 1994.
[41] W. Wang, V.W. Zheng, H. Yu, C. Miao, A survey of zero-shot learning: settings, methods, and applications, ACM Trans. Intell. Syst. Technol. 10 (2019) 1–37.
[42] M. Palatucci, D. Pomerleau, G.E. Hinton, T.M. Mitchell, Zero-shot learning with semantic output codes, in: Advances in Neural Information Processing Systems, MIT Press, 2009, pp. 1410–1418.
[43] A. Rababah, Taylor theorem for planar curves, Proc. Am. Math. Soc. 119 (1993) 803–810.
[44] O.A. AL-Sammarraie, M.A. Bashir, Generalization of Newton's forward interpolation formula, Int. J. Sci. Res. Publ. (2015).
[45] J. Magtoto, A. Roque, Real-time traffic data collection and dissemination from an Android smartphone using proportional computation and freesim as a practical transportation system in metro manila, in: TENCON 2012 IEEE Region 10 Conference, IEEE, 2012, pp. 1–5.
[46] M. Lin, A. Wierman, L.L. Andrew, E. Thereska, Dynamic right-sizing for power-proportional data centers, IEEE/ACM Trans. Netw. 21 (5) (2012) 1378–1391.
[47] C. Vondrick, D. Patterson, D. Ramanan, Efficiently scaling up crowdsourced video annotation, Int. J. Comput. Vis. 101 (1) (2013) 184–204.
[48] J. Carroll, R. Haertel, P. McClanahan, E.K. Ringger, K. Seppi, Assessing the costs of sampling methods in active learning for annotation, in: Proceedings of ACL-08: HLT, Short Papers, ACL Press, 2008, pp. 65–68.

[49] S.P. Liew, T. Takahashi, Faceleaks: inference attacks against transfer learning models via black-box queries, CoRR, arXiv:2010.14023 [abs], 2020, arXiv: 2010.14023, https://arxiv.org/abs/2010.14023.
[50] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P.S. Yu, X. Zhang, Membership inference attacks on machine learning: a survey, ACM Comput. Surv. (2021).
[51] S. Truex, L. Liu, M.E. Gursoy, L. Yu, W. Wei, Demystifying membership inference attacks in machine learning as a service, IEEE Trans. Serv. Comput. (2019).
[52] S. Rahimian, T. Orekondy, M. Fritz, Sampling attacks: amplification of membership inference attacks by repeated queries, arXiv preprint arXiv:2009.00395, 2020.
[53] C.A. Choquette-Choo, F. Tramer, N. Carlini, N. Papernot, Label-only membership inference attacks, in: International Conference on Machine Learning, PMLR, 2021, pp. 1964–1974.
[54] Z. Li, Y. Zhang, Membership leakage in label-only exposures, in: Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2021, pp. 880–895.