A cross-domain recommender system through information transfer for medical diagnosis

Wenjun Chang[a,b,*], Qian Zhang[b,*], Chao Fu[a], Weiyong Liu[c], Guangquan Zhang[b], Jie Lu[b,†]

aSchool of Management, Hefei University of Technology, P.R. China

bCenter for Artificial Intelligence, University of Technology Sydney, Australia

cDepartment of Ultrasound, The First Affiliated Hospital of USTC, P.R. China

E-mail address: Wenjun.Chang@uts.edu.au; Qian.Zhang-1@uts.edu.au; chaofu@hfut.edu.cn;

lwy_ultras@126.com;Guangquan.Zhang@uts.edu.au; Jie.Lu@uts.edu.au.

Abstract

The electronic diagnostic records of patients, primarily collected by hospitals, comprise valuable data for the development of recommender systems to support physicians in predicting the risks associated with various diseases. For some diseases, the diagnostic record data are not sufficient to train a prediction model to generate recommendations; this is referred to as the data sparsity problem. Cross-domain recommender systems offer a solution to this problem by transferring knowledge from a similar domain (source domain) with sufficient data for modeling to facilitate prediction in the current domain (target domain). However, building a cross-domain recommender system for medical diagnosis presents two challenges: (1) uncertain representations, such as the symptoms characterized by interval numbers, are often used in medical records, and (2) given two different diseases, the feature spaces of the two diagnostic domains are often disparate because the diseases are only likely to share a few symptoms. This study addresses these challenges by proposing a cross-domain recommender system, named information transfer for medical diagnosis (ITMD), to provide physicians with personalized recommendations for disease risks. In ITMD, a novel dissimilarity measurement was performed for diagnosis, represented as interval numbers. The space alignment technique eliminated the feature space divergence caused by different symptoms between two diseases, and the development of collective matrix factorization enabled knowledge transfer between the source and target domains. Experiments and a case study using real-world data demonstrated that ITMD outperforms four baselines and improves the accuracy of recommendations for disease risks in patients to support physicians in determining a final medical diagnosis.

---

*These authors contributed equally to this work.
† This is the corresponding author.

1. Introduction

In clinical medicine, medical diagnosis is often the first step in predicting and preventing the possible onset of a disease [1]. Early diagnosis helps in preventing further deterioration and increases the possibility of complete healing, especially for cancer. For example, breast tumors are a common malignant neoplasm in women, and over 60% of patients with breast tumors are incurable [2]. However, breast cancer is preventable and curable if the tumor is diagnosed in its early stages. Electronic medical records comprise significant information that could help physicians in diagnosing several diseases accurately in the early stage. Because the symptoms and situations vary among patients, personalized recommendations are required for medical diagnosis by physicians. Recommender systems are an effective tool to provide personalized decision support for medical diagnosis, such as chronic disease [3], heart disease [4], and right heart catheterization [5]. In a recommender system for medical diagnosis, a user is represented by a patient, an item is represented by the level of disease risks, and the probability of a patient to a disease risk is represented by a rating of the user to an item given by the physician or a prediction model.

Although recommender systems can provide medical diagnosis support, they encounter the data sparsity problem [8–10]. Data sparsity in medical diagnosis occurs owing to insufficient medical records of some diseases because the incidence of these diseases is low or physicians are inexperienced; it can also be attributed to the limited access to medical records. Decision support with sparse data in the medical field may generate inaccurate results for disease diagnosis. For example, tetanus is a rare but life-threatening disease. It is difficult to use medical records to support physicians in providing more accurate diagnoses when the available data are insufficient and outdated for tetanus. The use of such insufficient data to estimate the relations between symptoms (e.g., age, pattern, severity, and interventions) in the diagnosis of tetanus is inconclusive; thus, the deductions generated with insufficient data are not permitted to support the diagnosis of this disease [11]. Recommender systems built in domains without sufficient data fail to provide accurate recommendations to physicians.

To alleviate the problem of data sparsity, a feasible strategy involves the adoption of transfer learning. Transfer learning has been widely used to manage data sparsity in medical diagnosis [12–14]. In recommender systems for medical diagnosis, transfer learning can be combined with collaborative filtering to extract knowledge from a source domain with sufficient data and transfer it to the target domain, known as cross-domain recommender systems [15].

In medical diagnosis, two types of disease diagnoses are considered to be of two similar domains when there are three perspectives of similarity between them [16–18]; that is, (a) the organs in which the two diseases occur belong to one type, such as the gland, abdomen, and superficial organs; (b) in general, the cytopathy and pathogenesis of the two diseases are similar; and (c) the two diseases are diagnosed by similar standards. Owing to the three perspectives of similarity, two similar diseases will lead to similar relations between their symptoms, which are considered as the knowledge transferred from the source disease to the target disease. To transfer the relations between symptoms across two similar diseases, the overlapping symptoms serve as a bridge to align these relations. The extracted relations between symptoms are used to generate recommendations of the target disease.

A cross-domain recommender system for medical diagnosis uses the knowledge extracted from relatively sufficient symptom observation data of one disease (source domain) to support physicians in making decisions regarding the diagnosis of another disease with sparse symptom data (target domain). The establishment of such recommender system generally faces two challenges. (1) Medical records contain uncertain representations of symptoms and disease risks. For example, five diagnostic categories of disease risks are represented as interval numbers in the diagnosis of thyroid cancer, as {TIRADS 3, TIRADS 4A, TIRADS 4B, TIRADS 4C, TIRADS 5} = {[0, 0.03], (0.03, 0.25], (0.25, 0.75], (0.75, 0.95], (0.95, 1]}. TI-RADS 5 means the physician evaluates the thyroid legions to have a 95–100% chance of being malignant, which strongly indicates thyroid cancer, while TI-RADS 4A indicates a 3–25% chance of malignancy. The prediction of diagnoses as a disease risk with an interval number and measuring the differences between diagnosis predictions with interval numbers between two domains are urgent problems that require solutions. (2) The feature (symptom) spaces of two different diseases in the source and target domains are different and mismatched. Considering breast cancer as the source domain and tophus as the target domain, the five symptoms of a breast lump suggest the presence of breast cancer, such as the margin, contour, echogenicity, calcification, and vascularity. In contrast, tophus is assessed against six symptoms, i.e., the number, size, contour, echogenicity, calcification, and halo. Thus, the two diagnoses only share three symptoms. Additionally, the severity values related to the shared symptoms may not be in the same distribution, which also needs to be considered before knowledge extraction.

To address these two challenges, this study develops a new cross-domain recommendation method, called information transfer for medical diagnosis (ITMD), to provide personalized recommendations for the risk level of

particular diseases to support physicians in diagnosing patients with insufficient medical records. In the ITMD method, after normalization, two transformation matrices are learned to align the symptom space of the source domain with that of the target domain. A novel dissimilarity is proposed to measure the difference in disease risks indicated in the diagnostic categories. The symptom relations between two domains are then learned as shared knowledge by minimizing the dissimilarities between the diagnoses and their predictions in the two domains simultaneously. In summary, the main contributions of this study are as follows:

(1) A new dissimilarity measurement is developed to characterize the discrepancies between diagnoses with interval numbers corresponding to challenge (1). This measurement is more suitable for diagnostic tasks than distance measurement because it fully reflects the risk intervals associated with each diagnostic category.

(2) A space alignment method is established to deal with the mismatch of two different symptom spaces corresponding to challenge (2). This method eliminates the differences in the number of symptoms and their data distributions in the symptom spaces of the source and target domains.

(3) The ITMD method is proposed, which is a new cross-domain recommender system to recommend suitable medical diagnostics; it can effectively predict the risk level of patients for particular diseases. With constraints of condition number on collective matrix factorization, this method transfers knowledge from the source domain for support decision making in the target domain. The proposed dissimilarity measurement and space alignment method are used in the recommender system.

(4) A set of comprehensive experiments and a real-world case study (with thyroid cancer and breast cancer as the source and target domains, respectively) are performed, which demonstrated the accuracy and effectiveness of the proposed ITMD method for recommending disease risks in comparison to four baselines. The results of the case study particularly indicate the effectiveness considering decision support for real-world diagnosis for four physicians.

The remainder of this paper is organized as follows. Section 2 presents the work related to this study. Section 3 introduces the medical diagnosis problems based on the interval number in one domain and formally defines the problem being examined in this study. In Section 4, the ITMD method is proposed. Section 5 details the experiments conducted in two cases: one where the symptom spaces are the same, and the other where they are different. The results show that our method performs better than the four baseline methods. Section 6 illustrates how the proposed method works in a real-world scenario as a case study. The conclusions and directions for further research are provided

in Section 7.

2. Related work

This section first presents a review of the relevant work on recommender systems and their applications in the medical domain, followed by a summary of the current state-of-the-art cross-domain recommender systems.

2.1. Recommender systems and their applications in medical domain

Because a large number of electronic health records have been collected by hospitals, the information extracted from these data is used to develop recommender systems that can support the decision-making process of physicians for diagnosing diseases. For example, a recommendation system framework was developed to support physicians with personalized prescriptions to improve their efficiency and reduce the risk of making errors in daily clinical consultation with patients [19]. Another example involves the use of a collaborative filtering technique with clustering to provide medical advice to cardiac patients [20]. Measurement between treatments and a unified recommendation method were proposed in [21] to recommend treatments to new patients according to their demographic information and disease severity.

In addition, recommender systems have been developed to support diagnosis in different domains of disease diagnosis based on collaborative filtering. Hussein et al. [3] proposed a chronic disease diagnosis recommender system approach based on a hybrid method using multiple classifications and unified collaborative filtering to provide high-accuracy disease risk prediction and medical recommendation. Hassan and Syed [4] built a collaborative filtering framework to achieve high accuracy in predicting sudden cardiac death and recurrent myocardial infarction by concurrently matching new cases to historical records as well as patient demographics to adverse outcomes. Davis et al. [22] presented a recommender system that used collaborative filtering to predict the potential disease risks of a patient based on sufficient medical history and that of similar patients.

The aforementioned recommender systems and their key characteristics are summarized in Table 1. The existing recommender systems in the medical domain, for example, [3–4] and [19–21], cannot manage the challenges of interval data and data sparsity, which commonly occur in practice.

2.2. Cross-domain recommender systems

Cross-domain recommender systems can be divided into two broad categories, i.e., user-item matrix-based methods and additional information-based methods.

(1) Cross-domain recommender systems with user-item rating matrix

Cross-domain recommender systems with a user-item rating matrix can have non-overlapping, partially overlapping, or fully overlapping users/items. Non-overlapping users/items implies that all users and items in the source and target domains are different. Here, users and items are usually clustered into groups, and shared knowledge is extracted at the group level. Li et al. [23] provided a generative rating matrix model by transferring a shared cluster-level rating matrix across domains. Gao et al. [24] created a novel cross-domain recommendation model that transferred rating patterns at a common cluster level. Zhang et al. [25] extracted the shared group-level knowledge based on consistent user/item group information, adjusted by a domain adaptation technique.

Partially overlapping users/items refer to situations where some of the users or items are common to both domains. These methods usually adopt matrix factorization models and implement shared knowledge transfer based on overlapping users or items. Pan et al. [26] established a coordinate system transfer method based on sparse matrix tri-factorization to reduce the effect of data sparsity, while Pan and Yang [27] constructed a transfer framework based on collective factorization with sparse data by transferring explicit binary rating information from the source to the target domain. Zhang et al. [28] proposed a cross-domain recommender system that used kernel-induced knowledge transfer to manage partially overlapping entities. The entity correlations of the overlapping entities were determined via domain adaptation and diffusion kernel completion methods, which also served as constraints to generate a new matrix factorization technique.

(2) Cross-domain recommender systems with additional information

Some cross-domain recommender systems connect two domains through additional information, instead of user-item ratings [29], such as social network information, tagging information, review information, and metadata. Chen et al. [30] established a cross-domain recommender system that fused social network information and cross-domain rating data to improve recommendations. The cluster-level tensor was considered as shared information in the source and target domains. Jiang et al. [31] proposed a hybrid random walk method to transfer knowledge from auxiliary item domains to the target domain using social network data to create a star-structured hybrid graph. Shi et al. [32] used shared tags as bridges to connect the source and target domains. To study the correlations between tags from different domains, Fang et al. [33] constructed a tag matrix transfer model to obtain rating patterns by transferring the shared tag co-occurrence matrix information in multiple domains. The user-tag-item relation was factorized into user-

tag and item-tag relations. Hao et al. [34] regularized joint matrix factorization with the constraints of inter-domain and intra-domain correlations, built based on the tagging information. Xin et al. [35] proposed a nonlinear cross-domain recommendation framework with review information. Joint tensor factorization [36] was modeled to transfer the review information in cross-domain recommendations. Fernández-Tobías et al. [37] developed cross-domain hybrid matrix factorization models that exploited metadata as a bridge between items across domains.

These cross-domain recommender systems and their key characteristics are summarized in Table 1. The existing cross-domain recommender systems, for example, [22–28] and [30–37], are not specifically designed for medical diagnosis problems; thus, they cannot be applied directly to the problem being investigated in this study. Moreover, the interval data that are commonly observed in medical diagnosis cannot be addressed by any existing study.

Table 1. Summary of related works

	Data			Domain			User/item overlap		Field	
	User-item rating		Additional information	Single	Two	Multiple	Non-overlap	Overlap	Medical	Non-medical
	Discrete value	Interval number								
[19]	×			×			—	—	×	
[20]	×			×			—	—	×	
[21]	×			×			—	—	×	
[3]	×			×			—	—	×	
[4]	×			×			—	—	×	
[22]	×			×			—	—	×	
[23]	×					×	×			×
[24]	×					×	×			×
[25]	×				×		×			×
[26]	×				×			×		×
[27]	×				×			×		×
[28]	×				×			×		×
[30]	×		×		×		×			×
[31]	×		×		×			×		×
[32]	×		×			×	×			×
[33]	×		×			×	×			×
[34]	×		×		×		×			×
[35]	×		×		×			×		×
[36]	×		×		×			×		×
[37]	×		×		×			×		×
ITMD		×	×		×		×		×	

3. Problem definition

This section first presents the problem formulation for medical diagnosis with interval number in one domain. Then, we formally describe the cross-domain medical diagnosis problem under study.

3.1. Medical diagnosis with interval number in one domain

In medical diagnosis, physicians diagnose the likely cause of the conditions of a patient based on their symptoms. A formal definition of this medical diagnosis problem for a group of patients is presented below.

Consider three lists, $P = \{P_1, \ldots, P_M\}$, $S = \{S_1, \ldots, S_L\}$, and $C = \{C_1, \ldots, C_N\}$, where P is a list of patients, S is a list of symptoms, and C is a list of diagnostic categories. Each symptom, S_l, has N_l levels of severity, as $L(S_l)=\{L_n(S_l)$, $n=1, \ldots, N_l\}=\{[L_1^-(S_l), L_1^+(S_l)], (L_2^-(S_l), L_2^+(S_l)], \ldots, (L_{N_l}^-(S_l), L_{N_l}^+(S_l)]\}$, where $L_1^-(S_l)=0$, $L_{N_l}^+(S_l)=1$, $L_n^+(S_l)= L_{n+1}^-(S_l)$, $n = 1, \ldots, N_l-1$, and $\bigcup_{n=1}^{N_l} L_n(S_l) = [0, 1]$. The relations between patients, symptoms, and diagnostic categories were recorded in three matrices. $U=(u_{ml})_{M \times L}$ denotes the patient-symptom matrix, where $u_{ml} \in L(S_l)$ is the level of a symptom, S_l, in patient P_m. $V = (v_{ln})_{L \times N}$ denotes the symptom-category matrix, where $v_{ln} \in [0, 1]$ and $\sum_{n=1}^N v_{ln} = 1$. $R = (r_{mn})_{M \times N}$ is the patient-category matrix, where $r_{mn} \in [0, 1]$ and $\sum_{n=1}^N r_{mn} = 1$. These matrices correspond to the reasoning followed by a physician when making a diagnosis; i.e., relate a patient to some symptoms, relate those symptoms to a diagnostic category, and then relate the patient to that diagnostic category. Thus, the diagnosis problem involves the determination of matrix R from matrices U and V.

A recommender system is used to assist physicians to provide the patient-category matrix R, which can be simply formulated as $\hat{R} = f(U, V)$. The following example can help in understanding the details of this problem.

Example 1. Suppose a physician needs to diagnose whether eight patients P_m ($m = 1, \ldots, 8$) have breast cancer. This requires assessing the suspected breast nodules against five symptoms S_l ($l = 1, \ldots, 5$), i.e., margin, contour, echogenicity, calcification, and vascularity. Here, the same levels $L(S_l) = \{[0, 0.03], (0.03, 0.25], (0.25, 0.75], (0.75, 0.95], (0.95, 1]\}$ are applied to symptoms. A patient-symptom matrix, $U=(u_{ml})_{8 \times 5}$, is presented in Table 2.

Table 2. Patient-symptom matrix for eight patients with breast nodules.

	S_1	S_2	S_3	S_4	S_5
P_1	(0.03, 0.25]	(0.95, 1]	(0.25, 0.75]	(0.95, 1]	(0.03, 0.25]
P_2	[0, 0.03]	[0, 0.03]	[0, 0.03]	[0, 0.03]	(0.03, 0.25]
P_3	(0.25, 0.75]	(0.95, 1]	(0.25, 0.75]	(0.25, 0.75]	(0.95, 1]
P_4	(0.03, 0.25]	(0.95, 1]	(0.25, 0.75]	(0.25, 0.75]	(0.95, 1]
P_5	(0.25, 0.75]	(0.25, 0.75]	(0.75, 0.95]	(0.95, 1]	[0, 0.03]
P_6	(0.03, 0.25]	(0.03, 0.25]	[0, 0.03]	[0, 0.03]	(0.03, 0.25]
P_7	[0, 0.03]	[0, 0.03]	(0.25, 0.75]	(0.95, 1]	(0.03, 0.25]
P_8	(0.03, 0.25]	(0.25, 0.75]	(0.25, 0.75]	(0.03, 0.25]	(0.95, 1]

The physician then relates these symptoms to the five diagnostic categories {TI-RADS 3, 4A, 4B, 4C, and 5} denoted as C_n ($n = 1, \ldots, 5$); this produces the symptom-category matrix, $V=(v_{ln})_{5 \times 5}$, as depicted in Table 3.

Table 3. Symptom-category matrix of the five symptoms and five diagnostic categories.

	C_1	C_2	C_3	C_4	C_5
S_1	0.2	0.15	0.35	0.15	0.15
S_2	0.3	0.3	0.25	0.05	0.1
S_3	0.25	0.3	0.2	0.15	0.1

S_4	0.1	0.15	0.3	0.25	0.2
S_5	0.3	0.25	0.2	0.2	0.05

Matrices U and V now provide a basis for relating the eight patients to the five diagnostic categories; i.e., matrix R $= (r_{mn})_{8 \times 5}$, as presented in Table 4.

Table 4. Patient-category matrix showing the correspondences of the eight patients to the five diagnostic categories.

	C_1	C_2	C_3	C_4	C_5
P_1	0	0	0	1	0
P_2	1	0	0	0	0
P_3	0	0	1	0	0
P_4	0	0	0	1	0
P_5	0	0	1	0	0
P_6	0	1	0	0	0
P_7	0	0	1	0	0
P_8	0	0	1	0	0

Thus, the diagnosis of breast cancer is the process of determining matrix R from matrices U and V.

3.2. Cross-domain recommender systems for medical diagnosis

To formally define this medical diagnosis problem as it relates to our cross-domain setting, we assume that there are two related but different disease domains. One domain, the target domain, does not contain enough records to provide adequate diagnostic support; however, the other, the source domain, has an abundance of records. The problem under study, therefore, involves determining the usage of records from the source domain to improve diagnostic support in the target domain. First, a symptom space must be defined with interval numbers to describe the problem.

Suppose that $S = \{S_1, \ldots, S_N\}$ represents a list of symptoms in a medical diagnosis problem with interval numbers. The symptom space for the diagnosis is generated by the associations between symptoms S_l and S_k, where $l, k \in \{1, \ldots, N\}$, which describe the influences of S_l and S_k on the disease risks. When the symptoms in the source and target domains have the same number and distribution, the symptom spaces are the same in the two domains; otherwise, the symptom spaces are different. In this study, we focus on the latter, which is more challenging. Accordingly, the medical diagnosis problem under study is formally described.

Suppose that $(U_s = (u^s_{ml})_{M_s \times L_s}$, $V_s = (v^s_{ln})_{L_s \times N_s}$, $R_s = (r^s_{mn})_{M_s \times N_s})$ and $(U_t = (u^t_{ml})_{M_t \times L_t}$, $V_t = (v^t_{ln})_{L_t \times N_t}$, $R_t = (r^t_{mn})_{M_t \times N_t})$ represent the patient-symptom matrix, symptom-category matrix, and patient-category matrix in the source and target domains, respectively. Here, M_s and M_t denote the number of patients, L_s and L_t are the number of symptoms, and N_s and N_t are the number of symptoms in the two domains. When the symptom spaces in the two domains are different,

a cross-domain recommender system extracts knowledge from the source domain and predicts medical diagnoses in the target domain by $\hat{R}_t = g(R_s, U_s, V_s, U_t, V_t)$.

4. Cross-domain recommender system through information transfer for medical diagnosis

In this section, we present the proposed ITMD recommendation method. The ITMD method transfers the shared information contained in the symptom spaces from the source domain to the target domain when the symptom spaces in the two domains are different. The procedure of the ITMD method is presented in Fig. 1, which consists of five steps. (1) The patient-symptom matrices in the source and target domains are normalized. (2) Symptom space alignment is designed to generate consistent symptom spaces in the two domains. (3) A new dissimilarity measurement is used to calculate the dissimilarities between the diagnoses and their predictions. (4) Information transfer is conducted based on the constraints on the symptom relation matrix. (5) Medical diagnosis support in the target domain is enabled. Next, we further demonstrate the ITMD method in detail.
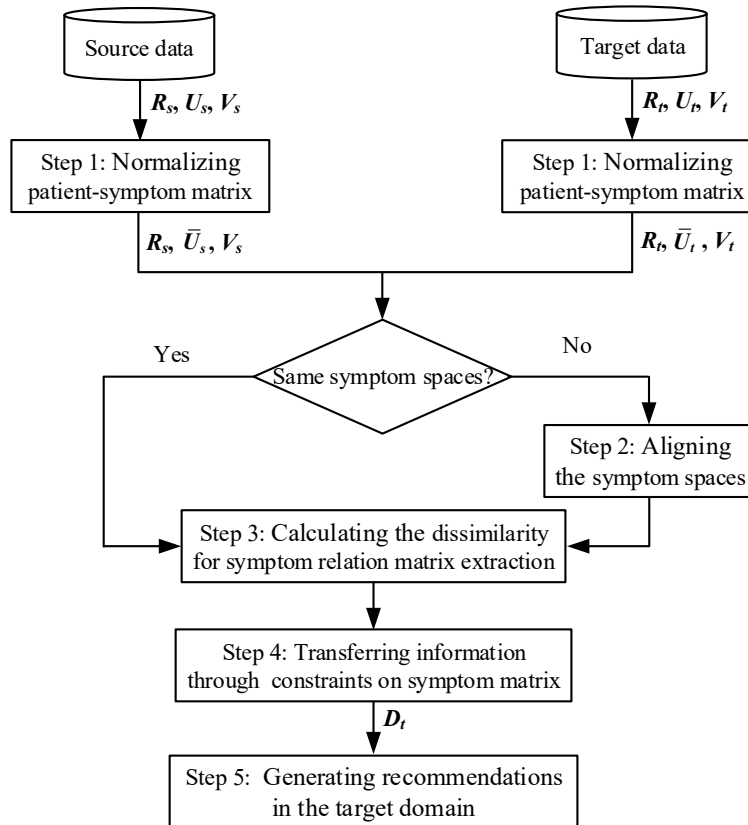

Fig. 1. Five steps of the ITMD framework.

4.1. Step 1: Normalizing the patient-symptom matrix

To normalize the patient-symptom matrices U_s and U_t, the interval-valued matrices must be transformed into numerical matrices. Because all elements in U_s and U_t are drawn from the interval sets $L(S_l)$, U_s and U_t can be converted

into averaged matrices $\bar{U}_s = (\bar{u}^s_{ml})_{M_s \times L_s}$ and $\bar{U}_t = (\bar{u}^t_{ml})_{M_t \times L_t}$, respectively. Here, $\bar{u}^s_{ml} = \dfrac{u^{s-}_{ml} + u^{s+}_{ml}}{2}$, where u^{s-}_{ml} and u^{s+}_{ml}

denote the lower and upper bounds of u^s_{ml}, respectively. Likewise, $\bar{u}^t_{ml} = \dfrac{u^{t-}_{ml} + u^{t+}_{ml}}{2}$, where u^{t-}_{ml} and u^{t+}_{ml} denote the

lower and upper bounds of u^t_{ml}, respectively.

Demonstrating the rationality of converting u^s_{ml} and u^t_{ml} into $\bar{u}^s_{ml}$ and $\bar{u}^t_{ml}$, respectively, involves generalizing the

relations between them as two propositions, as presented below.

Proposition 1. Consider a list of intervals I_n, $n = 1, \ldots, N$ that satisfy

$$
\begin{aligned}
&I_1 = [I_1^-, I_1^+] = [0, I_1^+], \\
&I_n = (I_n^-, I_n^+], \ n = 1, \ldots, N-1, \\
&I_N = (I_N^-, I_N^+] = (I_N^-, 1], \\
&0 \le I_n^- < I_n^+ \le 1, \ n = 1, \ldots, N, \\
&I_n^+ = I_{n+1}^-, \ n = 1, \ldots, N-1, \text{ and} \\
&\bigcup_{n=1}^{N} I_n = [0, 1],
\end{aligned}
\tag{1}
$$

and suppose that $\bar{I}_n$ denotes the average of I_n ($n = 1, \ldots, N$); it is calculated by $\bar{I}_n = \dfrac{I_n^- + I_n^+}{2}$, $n = 1, \ldots, N$. Then, it

follows that $I_m = I_n$ iff $\bar{I}_m = \bar{I}_n$, $m, n = 1, \ldots, N$.

The proof of Proposition 1 can be easily obtained. From Proposition 1, it is evident that there is a one-to-one

correspondence between I_n and $\bar{I}_n$. However, is the difference in size between I_m and I_n consistent with that between

$\bar{I}_m$ and $\bar{I}_n$? This is determined by measuring the possibility degree of $\bar{I}_m$ being superior to I_n [38] as follows:

$$
p(I_m > I_n) = \begin{cases}
1, & I_m^- \ge I_n^+ \\[2ex]
\dfrac{(I_m^+ - I_n^-)^2 - (I_m^+ - I_n^+)^2 - (I_m^- - I_n^-)^2}{2 \cdot (I_m^+ - I_m^-) \cdot (I_n^+ - I_n^-)}, & I_n^- \le I_m^- < I_n^+ \le I_m^+ \\[2ex]
0.5 \cdot \dfrac{(I_m^+ - I_n^-) \cdot (I_m^+ - I_n^-)}{(I_m^+ - I_m^-) \cdot (I_n^+ - I_n^-)}, & I_m^- \le I_n^- < I_m^+ \le I_n^+ \\[2ex]
\dfrac{I_m^+ - I_n^+}{I_m^+ - I_m^-} + 0.5 \cdot \dfrac{I_n^+ - I_n^-}{I_m^+ - I_m^-}, & I_m^- \le I_n^- < I_n^+ \le I_m^+ \\[2ex]
\dfrac{I_m^- - I_n^-}{I_n^+ - I_n^-} + 0.5 \cdot \dfrac{I_m^+ - I_m^-}{I_n^+ - I_n^-}, & I_n^- \le I_m^- < I_m^+ \le I_n^+ \\[2ex]
0, & I_n^- \ge I_m^+
\end{cases}
\tag{2}
$$

From Eq. (1), there are three possible correlations between I_m and I_n: 1) $I_n^- < I_n^+ \leq I_m^- < I_m^+$; 2) $I_n^- = I_m^- < I_m^+ = I_n^+$; and

3) $I_m^- < I_m^+ \leq I_n^- < I_n^+$. Accordingly, Eq. (2) can be transformed into:

$$p(I_m > I_n) = \begin{cases} 1, & I_n^- < I_n^+ \leq I_m^- \leq I_m^+ \\ 0.5, & I_n^- = I_m^- < I_m^+ = I_n^+ \\ 0, & I_m^- < I_m^+ \leq I_n^- < I_n^+ \end{cases}.$$

Proposition 2. Given a list of intervals $\{I_n, n = 1, \ldots, N\}$ that satisfy Eq. (1), suppose $\bar{I}_n$ is the average of I_n. Then, it

follows that

(1) $p(I_m > I_n) = 1$ iff $\bar{I}_m > \bar{I}_n$;

(2) $p(I_m > I_n) = 0.5$ iff $\bar{I}_m = \bar{I}_n$; and

(3) $p(I_m > I_n) = 0$ iff $\bar{I}_m < \bar{I}_n$.

The proof of Proposition 2 can be easily obtained. From Proposition 2, it can be deduced that the difference in size

between I_m and I_n is consistent with that between $\bar{I}_m$ and $\bar{I}_n$.

From Propositions 1 and 2, it is reasonable to transform $\boldsymbol{U}_s$ and $\boldsymbol{U}_t$ into $\bar{\boldsymbol{U}}_s$ and $\bar{\boldsymbol{U}}_t$, respectively.

4.2. Step 2: Aligning the symptom spaces

Once the patient-symptom matrix is normalized, the patient-category matrices in the two domains, $\boldsymbol{R}_s$ and $\boldsymbol{R}_t$ can be

factorized. However, first, a symptom matrix for a single domain must be defined to demonstrate the relation between

symptoms.

Definition 1 (symptom matrix). Given the medical diagnosis problem presented in Section 3, the symptom matrix $\boldsymbol{D}$

$= (d_{ij})_{L \times L}$ is defined to represent the degree of influence of symptom S_i on the disease risks of symptom S_j, resulting

in the diagnostic categories, where d_{ij} satisfies $d_{ij} \in [0, 1]$, $d_{ij} = d_{ji}$, and $\sum_{j=1}^{L} d_{ij} = 1$.

According to Definition 1, $\boldsymbol{R}_s$ and $\boldsymbol{R}_t$ can be factorized as $\boldsymbol{R}_s = \bar{\boldsymbol{U}}_s \boldsymbol{D}_s \boldsymbol{V}_s$ and $\boldsymbol{R}_t = \bar{\boldsymbol{U}}_t \boldsymbol{D}_t \boldsymbol{V}_t$, respectively, where $\boldsymbol{D}_s =$

$(d_{ij}^s)_{L_s \times L_s}$ denotes the symptom matrix in the source domain and $\boldsymbol{D}_t = (d_{ij}^t)_{L_t \times L_t}$ in the target domain. It should be noted

that the patient-symptom matrices $\boldsymbol{U}_s$ and $\boldsymbol{U}_t$, the symptom-category matrices $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$, and the patient-category

matrices $\boldsymbol{R}_s$ and $\boldsymbol{R}_t$ in the source and target domains can be obtained from medical records. If the medical records in

the target domain are insufficient, the symptom relation matrix D_s learned from $R_s = \bar{U}_s D_s V_s$ may be imprecise. To obtain more precise D_s, sufficient records in the source domain can be used to help learn D_s using the connection of the symptom relation matrices D_s and D_t. When both symptom spaces in the source and target domains are the same, any or all information in the source domain D_s can be transferred to the target domain D_t. Otherwise, the symptom spaces in the two domains must be aligned first.

Aligning the symptom spaces involves the process of factorizing the symptom matrix in the source domain D_s into three matrices, $D_s = M_1 D_s^1 M_2$, where $M_1 = (m_{ij}^1)_{L_s \times L_t}$ with $m_{ij}^1 \in [0,1]$ and $\sum_{j=1}^{L_s} m_{ij}^1 = 1$, $M_2 = (m_{ij}^2)_{L_t \times L_s}$ with $m_{ij}^2 \in [0,1]$ and $\sum_{j=1}^{L_s} m_{ij}^2 = 1$ are the two transformation matrices, and $D_s^1 = (d_{ij}^{s1})_{L_t \times L_t}$ is the new symptom matrix. Matrices D_s^1 and D_t share the same symptom space. Thus, matrix R_s is factorized into five matrices, i.e., $R_s = \bar{U}_s M_1 D_s^1 M_2 V_s$.

4.3. Step 3: Calculating the dissimilarity for symptom relation matrix extraction

Information cannot be transferred from the source domain to the target domain without knowing the dissimilarity between R_s and $\hat{R}_s$ and that between R_t and $\hat{R}_t$. To obtain the dissimilarities, the disease risks indicated in the diagnostic categories must be considered. For a list of diagnostic categories $C = \{C_1, \ldots, C_N\}$, we denote their risk intervals as $\{p(C_n), n=1, \ldots, N\} = \{[\,p^-(C_1),\ p^+(C_1)\,],\ (\,p^-(C_2),\ p^+(C_2)\,],\ \ldots,\ (\,p^-(C_N),\ p^+(C_N)\,]\}$ such that $p^-(C_1)=0,\ p^+(C_N)=1,\ 0 \le p^-(C_n) < p^+(C_n) \le 1, n = 1, \ldots, N,\ p^+(C_n) = p^-(C_{n+1}), n = 1, \ldots, N-1$, and $\bigcup_{n=1}^N p(C_n) = $ [0, 1]. Following Propositions 1 and 2, $p(C_n)$, $n = 1, \ldots, N$ can be transformed into their averages, $\bar{p}(C_n) = \dfrac{p^-(C_n) + p^+(C_n)}{2}$, and $n = 1, \ldots, N$; accordingly, a new dissimilarity measurement is constructed that describes the dissimilarity between R and its prediction $\hat{R}$, which is formally defined as follows.

Definition 2 (dissimilarity of diagnoses with risk intervals). Suppose that $p(C_n)$, where $n = 1, \ldots, N$ denotes the risk intervals of the diagnostic categories $C_n, n = 1, \ldots, N$. A dissimilarity measurement between a diagnostic matrix $R = (r_{mn})_{M \times N}$ and its predicted matrix $\hat{R} = (\hat{r}_{mn})_{M \times N}$ is defined as

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}) = \frac{\sum_{m=1}^{M}\sqrt{\sum_{l=1}^{N-1}\sum_{k=l+1}^{N}(r_{ml}-\hat{r}_{ml})^2\cdot(r_{mk}-\hat{r}_{mk})^2\cdot(\overline{p}(C_k)-\overline{p}(C_l))^2}}{M\cdot(\overline{p}(C_N)-\overline{p}(C_1))},\tag{3}$$

where $\overline{p}(C_n)$ denotes the average of $p(C_n)$.

The dissimilarity measurement between $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$ constructed in Definition 2 is distinct from other distance measures, such as the Euclidean distance [39], Manhattan distance [40], or Chebyshev distance [41], because it considers $\overline{p}(C_n)$ ($n=1,\ \ldots,\ N$) such that $0\le\overline{p}(C_1)<\overline{p}(C_2)<\ldots<\overline{p}(C_N)\le 1$. Hence, using $\overline{p}(C_k)-\overline{p}(C_l)$ ($l=1,\ \ldots,$ N, $k=l+1,\ \ldots,\ N$) distinguishes the dissimilarities between $(r_{ml},\ r_{mk})$ and $(\hat{r}_{ml},\ \hat{r}_{mk})$. The following example can help in explaining the construction of this dissimilarity measurement.

Example 2. Suppose that there are five diagnostic categories, C_n, $n=1,\ \ldots,\ 5$, in a medical diagnosis problem. Given the risk intervals defined in the diagnostic categories $\{p(C_n), n=1,\ \ldots,\ 5\} = \{[0, 0.03], (0.03, 0.25], (0.25, 0.75], (0.75, 0.95], (0.95, 1]\}$, assume that $\boldsymbol{R} = (1\ 0\ 0\ 0\ 0)$ and the two predictions are $\hat{\boldsymbol{R}}_1 = (0\ 1\ 0\ 0\ 0)$ and $\hat{\boldsymbol{R}}_2 = (0\ 0\ 0\ 0\ 1)$.

The averages of $\{p(C_n), n=1,\ \ldots,\ 5\}$ are $\{\overline{p}(C_n), n=1,\ \ldots,\ 5\} = \{0.015, 0.14, 0.5, 0.85, 0.975\}$. From Eq. (3), $d(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1)$ and $d(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2)$ are calculated as

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = \frac{\sum_{m=1}^{1}\sqrt{\sum_{l=1}^{4}\sum_{k=l+1}^{5}(r_{ml}-\hat{r}_{ml}^1)^2\cdot(r_{mk}-\hat{r}_{mk}^1)^2\cdot(\overline{p}(C_k)-\overline{p}(C_l))^2}}{(\overline{p}(C_5)-\overline{p}(C_1))} = 0.397;\ \text{and}$$

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2) = \frac{\sum_{m=1}^{1}\sqrt{\sum_{l=1}^{4}\sum_{k=l+1}^{5}(r_{ml}-\hat{r}_{ml}^2)^2\cdot(r_{mk}-\hat{r}_{mk}^2)^2\cdot(\overline{p}(C_k)-\overline{p}(C_l))^2}}{(\overline{p}(C_5)-\overline{p}(C_1))} = 1.$$

To ascertain whether our dissimilarity measurement (Eq. (3)) is more reasonable than the Euclidean, Manhattan, or Chebyshev distance, we compared all three distances between $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}_1$ or $\hat{\boldsymbol{R}}_2$. Each distance was calculated as follows:

Euclidean distance: $d_E(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = \sqrt{2}$ and $d_E(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2) = \sqrt{2}$;

Manhattan distance: $d_M(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = 2$ and $d_M(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2) = 2$; and

Chebyshev distance: $d_C(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = 1$ and $d_C(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2) = 1$.

Thus, we conclude that $d_E(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = d_E(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2)$, $d_M(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = d_M(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2)$, and $d_C(\boldsymbol{R},\ \hat{\boldsymbol{R}}_1) = d_C(\boldsymbol{R},\ \hat{\boldsymbol{R}}_2)$. Because $0 <$ $\bar{p}(C_1)< \bar{p}(C_2)< \ldots < \bar{p}(C_5)< 1$, $\hat{\boldsymbol{R}}_1$ is closer to $\boldsymbol{R}$ than $\hat{\boldsymbol{R}}_2$, which supports the use of our dissimilarity measurement.

The properties of the dissimilarity measurement between $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$ are summarized below.

Property 1. Suppose that $d(\boldsymbol{R},\ \hat{\boldsymbol{R}})$ represents the dissimilarity between two matrices $\boldsymbol{R}=(r_{mn})_{M\times N}$ and $\hat{\boldsymbol{R}}=(\hat{r}_{mn})_{M\times N}$ provided in Definition 2. Then, it follows that

$$0 \leq d(\boldsymbol{R},\ \hat{\boldsymbol{R}}) \leq 1, \tag{4}$$

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}) = d(\hat{\boldsymbol{R}},\ \boldsymbol{R}), \tag{5}$$

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}) = 0 \text{ iff } \boldsymbol{R} = \hat{\boldsymbol{R}}, \text{ and} \tag{6}$$

$$d(\boldsymbol{R},\ \hat{\boldsymbol{R}}) = 1 \text{ iff } r_{m1}{=}1 \text{ and } \hat{r}_{mN}{=}1 \text{ or } r_{mN}{=}1 \text{ and } \hat{r}_{m1}{=}1, m = 1, \ldots, M. \tag{7}$$

The proof of Property 1 is given in Appendix A. Using the dissimilarity measurement shown in Eq. (3), the dissimilarity between $\boldsymbol{R}_s$ and $\hat{\boldsymbol{R}}_s$ and that between $\boldsymbol{R}_t$ and $\hat{\boldsymbol{R}}_t$ can be obtained by $d(\boldsymbol{R}_s, \hat{\boldsymbol{R}}_s)$ and $d(\boldsymbol{R}_t, \hat{\boldsymbol{R}}_t)$, respectively.

4.4. Step 4: Transferring information through constraints on the symptom matrix

The process of transferring information differs depending on whether the symptom spaces in the source and target domains are the same. We first consider the more challenging of the two cases, where the symptom spaces are different.

As discussed in Step 2, when the symptom spaces are different, the patient-category matrices $\boldsymbol{R}_s$ and $\boldsymbol{R}_t$ must be factorized into $\boldsymbol{R}_s = \bar{U}_s M_1 D_s^1 M_2 V_s$ and $\boldsymbol{R}_t = \bar{U}_t D_t V_t$, respectively, and the symptom relation matrices D_s^1 and D_t share one symptom space. To further characterize the symptom space, we introduce the definition of the condition number of a matrix with the Frobenius norm, which describes the influence of the input variations in a matrix function on the output variations.

Definition 3 (condition number of matrix) [42]. Given matrix $A =(a_{ij})_{L_1\times L_2}$, $L_1, L_2 \in \bullet^{+}$, the condition number of A with the Frobenius norm is defined as $cond(A) = \|A\|_F \cdot \|A^{+}\|_F$, where $\|A\|_F = \sqrt{\sum_{i=1}^{L_1}\sum_{j=1}^{L_2} a_{ij}^{2}}$ and A^{+} is the generalized inverse matrix of A.

It is evident that $cond(A) \geq 1$. The condition numbers of $\boldsymbol{D}_t$ and $\boldsymbol{D}_s^1$ are denoted by $cond(\boldsymbol{D}_t)$ and $cond(\boldsymbol{D}_s^1)$, respectively. Now, we must analyze the influence of $cond(\boldsymbol{D}_t)$ on the matrix factorization of $\boldsymbol{R}_t = \bar{\boldsymbol{U}}_t \boldsymbol{D}_t \boldsymbol{V}_t$ and that of $cond(\boldsymbol{D}_s^1)$ on $\boldsymbol{R}_s = \bar{\boldsymbol{U}}_s \boldsymbol{M}_1 \boldsymbol{D}_s^1 \boldsymbol{M}_2 \boldsymbol{V}_s$. The influence of $cond(\boldsymbol{D}_t)$ on the sensitivity of the relative error of $\boldsymbol{R}_t$ to that of $\bar{\boldsymbol{U}}_t$ or $\boldsymbol{V}_t$ in the target domain is demonstrated below.

Theorem 1. If $\boldsymbol{R}_t = \bar{\boldsymbol{U}}_t \boldsymbol{D}_t \boldsymbol{V}_t$, the following two conclusions are satisfied.

(1) If $\Delta \bar{\boldsymbol{U}}_t$ and $\Delta \boldsymbol{R}_t$ satisfy $\boldsymbol{R}_t + \Delta \boldsymbol{R}_t = (\bar{\boldsymbol{U}}_t + \Delta \bar{\boldsymbol{U}}_t) \boldsymbol{D}_t \boldsymbol{V}_t$, then

$$\frac{1}{cond(\boldsymbol{D}_t) \cdot cond(\boldsymbol{V}_t)} \cdot \frac{\|\Delta \bar{\boldsymbol{U}}_t\|_F}{\|\bar{\boldsymbol{U}}_t\|_F} \leq \frac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F} \leq cond(\boldsymbol{D}_t) \cdot cond(\boldsymbol{V}_t) \cdot \frac{\|\Delta \bar{\boldsymbol{U}}_t\|_F}{\|\bar{\boldsymbol{U}}_t\|_F}, \tag{8}$$

where $\frac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F}$ and $\frac{\|\Delta \bar{\boldsymbol{U}}_t\|_F}{\|\bar{\boldsymbol{U}}_t\|_F}$ represent the relative error of $\boldsymbol{R}_t$ and $\bar{\boldsymbol{U}}_t$, respectively.

(2) If $\Delta \boldsymbol{V}_t$ and $\Delta \boldsymbol{R}_t$ satisfy $\boldsymbol{R}_t + \Delta \boldsymbol{R}_t = \bar{\boldsymbol{U}}_t \boldsymbol{D}_t (\boldsymbol{V}_t + \Delta \boldsymbol{V}_t)$, then

$$\frac{1}{cond(\boldsymbol{D}_t) \cdot cond(\bar{\boldsymbol{U}}_t)} \cdot \frac{\|\Delta \boldsymbol{V}_t\|_F}{\|\boldsymbol{V}_t\|_F} \leq \frac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F} \leq cond(\boldsymbol{D}_t) \cdot cond(\bar{\boldsymbol{U}}_t) \cdot \frac{\|\Delta \boldsymbol{V}_t\|_F}{\|\boldsymbol{V}_t\|_F}. \tag{9}$$

where $\frac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F}$ and $\frac{\|\Delta \boldsymbol{V}_t\|_F}{\|\boldsymbol{V}_t\|_F}$ represent the relative error of $\boldsymbol{R}_t$ and $\boldsymbol{V}_t$, respectively.

The proof of Theorem 1 is given in Appendix B. From Eqs. (8) and (9), we can conclude that the smaller the value of $cond(\boldsymbol{D}_t)$, the narrower the range of the relative error of $\boldsymbol{R}_t$, which indicates that the relative error of $\boldsymbol{R}_t$ is more controllable. In the source domain, the influence of $cond(\boldsymbol{D}_s^1)$ on the sensitivity of the relative error of $\boldsymbol{R}_s$ to that of $\bar{\boldsymbol{U}}_s \boldsymbol{M}_1$ or $\boldsymbol{M}_2 \boldsymbol{V}_s$ can similarly be deduced from Theorem 1. When $\boldsymbol{D}_s^1$ and $\boldsymbol{D}_t$ share the same symptom space, we have $cond(\boldsymbol{D}_s^1) = cond(\boldsymbol{D}_t)$. $\boldsymbol{D}_s^1$, $\boldsymbol{D}_t$, $\boldsymbol{M}_1$, and $\boldsymbol{M}_2$ can be learned by minimizing $d(\boldsymbol{R}_s, \bar{\boldsymbol{U}}_s \boldsymbol{M}_1 \boldsymbol{D}_s^1 \boldsymbol{M}_2 \boldsymbol{V}_s)$ and $d(\boldsymbol{R}_t, \bar{\boldsymbol{U}}_t \boldsymbol{D}_t \boldsymbol{V}_t)$. Minimizing the difference between $\boldsymbol{D}_s^1$ and $\boldsymbol{D}_t$ also needs to be considered. Thus, the cost function is constructed as follows:

$$J = d(\boldsymbol{R}_s, \hat{\boldsymbol{R}}_s) + d(\boldsymbol{R}_t, \hat{\boldsymbol{R}}_t) + \frac{\lambda}{2L^2} \cdot \sqrt{\sum_{i=1}^{L_t} \sum_{j=1}^{L_t} (d_{ij}^t - d_{ij}^{s1})^2}, \tag{10}$$

where $\hat{R}_s = (\hat{r}^s_{mn})_{M_s \times N_s} = \bar{U}_s M_1 D_s^1 M_2 V_s$, $\hat{R}_t = (\hat{r}^t_{mn})_{M_t \times N_t} = \bar{U}_t D_t V_t$, $\bar{U}_s$ and $\bar{U}_t$ are the average matrices of U_s and U_t, respectively, and λ is a parameter such that $0 \leq \lambda \leq 1$. Based on Eq. (10), D_s^1, D_t, M_1, and M_2 are learned by the following optimization problem:

$min\ J$

s.t. $cond(D_s^1) = cond(D_t)$

$$0 \leq D_s^1, D_t \leq 1, \ D_s^1 = (D_s^1)^T, \ D_t = D_t^T, \ \sum_{j=1}^{L_s}(D_s^1)_j = 1, \ \sum_{j=1}^{L_t}(D_t)_j = 1,$$

$$0 \leq M_1, M_2 \leq 1, \ \sum_{j=1}^{L_t}(M_1)_j = 1, \ \sum_{j=1}^{L_t}(M_2)_j = 1,$$

where $(D_s^1)_j$, $(M_1)_j$, and $(M_2)_j$ denote each column of D_s^1, M_1, and M_2, respectively.

This is a sequential quadratic programming problem [43] that can be solved by several existing solvers. The information transfer between the source and target domains is summarized in Algorithm 1.

Algorithm 1: Information transfer

Input: R_s, U_s, V_s, patient-category matrix, patient-symptom matrix, and symptom-category matrix of source domain
 R_t, U_t, V_t, patient-category matrix, patient-symptom matrix, and symptom-category matrix of target domain

Output: D_s^1, aligned symptom matrix of source domain
 D_t, symptom matrix of target domain
 M_1, M_2, transformation matrices for symptom space alignment

1 **INITIALIZE** $D_s^1, D_t \in \sim^{L_t \times L_t}$, $M_1 \in \sim^{L_s \times L_t}$, $M_2 \in \sim^{L_t \times L_s}$, $J_{min} \leftarrow 0$, $J \leftarrow 0$

2 **WHILE** $J = 0$ **OR** $J - J_{min} > \varepsilon$ **DO**

3   **FOR** D_s^1, D_t, M_1, M_2 in each iteration

4   **UPDATE** D_s^1, D_t, M_1, M_2 using the sequential quadratic programming method

5   **ENDFOR**

6   **UPDATE** J as in Eq. (10)

7   **IF** $J_{min} > J$

8         $J_{min} = J$

9   **ENDIF**

10 **ENDWHILE**

11 **RETURN** D_s^1, D_t, M_1, M_2

The ITMD method can also be used to support medical diagnosis with the same symptom spaces. In this situation, suppose that $D_s = (d_{ij}^s)_{L \times L}$ and $D_t = (d_{ij}^t)_{L \times L}$. Because D_s and D_t are aligned, R_s is factorized as $R_s = \bar{U}_s D_s V_s$; that is, M_1 and M_2 are not required. Because the optimization problem of solving D_s and D_t is similar to Eq. (10), the details are omitted.

4.5. Step 5: Generating recommendations in the target domain

Diagnostic support in the target domain is generated by $\hat{R}_t = \bar{U}_t D_t V_t$, where $\hat{R}_t$ is the predicted diagnostic category of patients; $\bar{U}_t$ is the average matrix of U_t, which denotes the relations between the patients and symptoms; V_t represents the relations between the symptoms and diagnostic categories; and D_t is the symptom matrix in the target domain. The highest prediction of diagnosis is the recommendation for physicians.

The ITMD method is presented in Algorithm 2.

Algorithm 2: ITMD method

Input: R_s, U_s, V_s, patient-category matrix, patient-symptom matrix, and symptom-category matrix of source domain
$\quad\quad$ R_t, U_t, V_t, patient-category matrix, patient-symptom matrix, and symptom-category matrix of target domain

Output: C_m, recommendation of target domain for patient p_m

1 Normalize U_s and U_t with $\bar{U}_s$ and $\bar{U}_t$

2 IF $L_s = L_t$

3 $\quad$ **INITIALIZE** $D_s \in \sim^{L_s \times L_s}$, $D_t \in \sim^{L_t \times L_t}$

4 $\quad$ $\hat{R}_s \leftarrow \bar{U}_s D_s V_s$

5 $\quad$ $\hat{R}_t \leftarrow \bar{U}_t D_t V_t$

6 $\quad$ **CALCULATE** $d(R_s, \hat{R}_s)$ and $d(R_t, \hat{R}_t)$ as in Eq. (3)

7 $\quad$ **CALCULATE J** as in Eq. (10) with $\hat{R}_s = \bar{U}_s D_s V_s$

8 $\quad$ **LEARN** D_s, D_t using the sequential quadratic programming method

9 ELSE IF

10 $\quad$ **INITIALIZE** $D_s^1, D_t \in \sim^{L_t \times L_t}$, $M_1 \in \sim^{L_s \times L_t}$, $M_2 \in \sim^{L_t \times L_s}$

11 $\quad$ $\hat{R}_s \leftarrow \bar{U}_s M_1 D_s^1 M_2 V_s$

12 $\quad$ $\hat{R}_t \leftarrow \bar{U}_t D_t V_t$

13 $\quad$ **LEARN** D_s^1, D_t, M_1, M_2 using Algorithm 1

14 ENDIF

15 $\hat{R}_t \leftarrow \bar{U}_t D_t V_t$

16 RETURN C_m for patient p_m using $C_m = \max\limits_{i=1,\dots,N_t} \left\{ \hat{r}_{mn}^t \right\}$

---

5. Experiments and analyses

Owing to the privacy and inaccessibility of diagnostic data for most diseases, it is necessary to generate synthetic data to investigate the effectiveness of the ITMD method. The generation of synthetic data should adequately reflect the characteristics of medical diagnosis at the risk intervals. We conducted two sets of experiments where the symptom spaces were either the same in the two domains or different. First, we explain the process of generating synthetic data, followed by the evaluation metrics and the baselines for comparison. Finally, the results are presented.

5.1. Synthetic datasets

To ensure that the synthetic data can reflect the structure of real medical records, they are generated based on the essential information for disease diagnosis. They include a set of patients $P = \{P_1, \ldots, P_M\}$, a set of symptoms $S = \{S_1, \ldots, S_L\}$, a set of diagnostic categories $C = \{C_1, \ldots, C_N\}$ with risk intervals $p(C_n)$ ($n=1, \ldots, N$), and the levels of severity of symptoms $L(S_l)$ ($l = 1, \ldots, L$).

For the first group of experiments, thyroid cancer and breast cancer were considered as the source and target domains, respectively. Both domains contain five symptoms, i.e., $S_s = S_t = \{S_1, \ldots, S_5\}$ = {margin, contour, echogenicity, calcification, vascularity}, and five diagnostic categories denoted by $C_s = C_t = \{C_1, \ldots, C_5\}$ [44]. The risk intervals of the diagnostic categories and the levels of severity of symptoms in the two domains are $\{p(C_n), n=1, \ldots, 5\} = \{L(S_l),$ $l = 1, \ldots, 5\}$ = {[0, 0.03], (0.03, 0.25], (0.25, 0.75], (0.75, 0.95], (0.95, 1]} [45]. Suppose that the diagnoses of patients on the symptoms in the two domains $U_s = (u_{ml}^s)_{M_s \times L_s}$ and $U_t = (u_{ml}^t)_{M_t \times L_t}$ follow two different probability distributions, and without loss of generality, the two probability distributions are assumed to be $(p(u_{ml}^s = L(S_1)), \ldots, p(u_{ml}^s = L(S_5)))$ = (0.3, 0.2, 0.3, 0.1, 0.1) and $(p(u_{ml}^t = L(S_1)), \ldots, p(u_{ml}^t = L(S_5)))$ = (0.4, 0.1, 0.1, 0.2, 0.2). Based on the essential information and possibility distributions in the two domains, 40000 and 10000 diagnostic records (patient-symptom ratings) were generated for the two domains, respectively. The parameter λ was set to 0.5. The number of records with the same symptom space in this group of experiments is listed in Table 5.

For the second group of experiments, gastric cancer and liver cancer were considered as the source and target domains, respectively. The diagnosis of gastric cancer involves six symptoms, $S_s = \{S_1^s, \ldots, S_6^s\}$ = {margin, contour, echogenicity, depth, lymph node metastases, size} [46], and eight diagnostic categories denoted by $Cs = \{C_1^s, \ldots, C_8^s\}$. The diagnosis of liver cancer includes five symptoms, $S_t = \{S_1^t, \ldots, S_5^t\}$ = {margin, echogenicity, contour, vascularity, size} [47], and five diagnostic categories denoted by $Ct = \{C_1^t, \ldots, C_5^t\}$. The TNM system of the American Joint Committee on Cancer/International Union Against Cancer Classification (AJCC/UICC) is used to represent the risk intervals of the diagnostic categories and the severity of symptoms in the diagnosis of gastric cancer; that is, $\{p(C_n^s), n=1, \ldots, 8\} = \{L(S_l^s), l = 1, \ldots, 8\}$ = {[0, 0.261], (0.261, 0.413], (0.413, 0.584], (0.584, 0.717], (0.717, 0.84], (0.84, 0.884], (0.884, 0.951], (0.951, 1]} [48]. Correspondingly, the risk intervals of the diagnostic categories and the

levels of severity of symptoms in the diagnosis of liver cancer are described by the risk stages of laparoscopic ultrasonography in the evaluation of liver nodules, which are $\{p(C_n^t), n=1, \ldots, 5\} = \{L(S_l^t), l = 1, \ldots, 5\} = \{[0, 0.18], (0.18, 0.24], (0.24, 0.3], (0.3, 0.84], (0.84, 0.1]\}$ [49]. Suppose that $U_s = (u_{ml}^s)_{M_s \times L_s}$ and $U_t = (u_{ml}^t)_{M_t \times L_t}$ follow two different probability distributions, and without loss of generality, they were assumed to be $(p(u_{ml}^s = L(S_1^s)), \ldots, p(u_{ml}^s = L(S_8^s))) = (0.1, 0.15, 0.1, 0.1, 0.15, 0.1, 0.1, 0.2)$ and $(p(u_{ml}^t = L(S_1^t)), \ldots, p(u_{ml}^t = L(S_5^t))) = (0.4, 0.1, 0.1, 0.2, 0.2)$. Based on the essential information and distributions in the two domains, 40000 and 10000 diagnostic records (patient-symptom ratings) of the two domains were generated. The parameter λ was set to 0.5. The number of records with the same symptom space in this group of experiments is listed in Table 5.

Table 5. Number of records for five groups of records with same and different symptom spaces.

Symptom spaces	Group ID	1:1		2:1		3:1		4:1	
		Source	Target	Source	Target	Source	Target	Source	Target
Same	G_1	2000	2000	4000	2000	6000	2000	8000	2000
	G_2	4000	4000	8000	4000	12000	4000	16000	4000
	G_3	6000	6000	12000	6000	18000	6000	24000	6000
	G_4	8000	8000	16000	8000	24000	8000	32000	8000
	G_5	10000	10000	20000	10000	30000	10000	40000	10000
Different	G_6	2000	2000	4000	2000	6000	2000	8000	2000
	G_7	4000	4000	8000	4000	12000	4000	16000	4000
	G_8	6000	6000	12000	6000	18000	6000	24000	6000
	G_9	8000	8000	16000	8000	24000	8000	32000	8000
	G_{10}	10000	10000	20000	10000	30000	10000	40000	10000

In all experiments, the test set was obtained from the target domain; then, we performed five-fold cross-validation to generate the training and test data, as demonstrated in [50–51]. Five-fold cross-validation was applied to the dataset in the target domain, and the entire dataset in the source domain was used for training.

5.2. Evaluation metrics and baselines

We selected the prediction accuracy as our evaluation metric, which is calculated as $AR_t = 1 - d(R_{|Y|}, \hat{R}_{|Y|})$, where Y is the test set in the target domain, $|Y|$ is the number of test sets, and $R_{|Y|}$ and $\hat{R}_{|Y|}$ represent the true values and predictions of diagnoses in the test set, respectively. The greater the prediction accuracy, the better the performance.

We selected four non-transfer learning methods as the baselines for comparison, named B_1 [52], B_2 [45], B_3 [53], and B_4 for simplicity. For baselines B_1 and B_2, the diagnoses were generated using a linear combination of the relations between patients and symptoms (denoted by $u_{ml} = [u_{ml}^-, u_{ml}^+], m = 1, \ldots, M, l = 1, \ldots, L$, where M and L are the number of patients and symptoms, respectively) and the relative weights of symptoms (denoted by $w_l, l = 1, \ldots, L$). Thus, the

risk intervals for patients denoted by $p_m = [\, p_m^-,\ p_m^+\,]$ can be predicted by $\hat{p}_m = [\, \hat{p}_m^-,\ \hat{p}_m^+\,] = \sum_{l=1}^{L} u_{ml} \cdot w_l$. The distances between p_m and $\hat{p}_m$ in the two baselines were respectively calculated with

$$d_1(p_m,\ \hat{p}_m) = \sqrt{(\bar{p}_m - \bar{\hat{p}}_m)^2 + \frac{1}{12} \cdot ((l(p_m))^2 + (l(\hat{p}_m))^2) - \frac{1}{6} \cdot (l(d_m))^2} \quad \text{and}$$

$$d_2(p_m,\ \hat{p}_m) = \left| \bar{p}_m \cdot (1 - \frac{1}{\sqrt{12}} \cdot l(p_m)) - \bar{\hat{p}}_m \cdot (1 - \frac{1}{\sqrt{12}} \cdot l(\hat{p}_m)) \right|,$$

where $\quad \bar{p}_m = 0.5 \cdot (p_m^- + p_m^+) \quad, \quad \bar{\hat{p}}_m = 0.5 \cdot (\hat{p}_m^- + \hat{p}_m^+) \quad, \quad l(p_m) = p_m^+ - p_m^- \quad, \quad l(\hat{p}_m) = \hat{p}_m^+ - \hat{p}_m^- \quad,$ and

$l(d_m) = (p_m \cap \hat{p}_m)^+ - (p_m \cap \hat{p}_m)^-$. For baseline B_3, the relation between patients and symptoms, u_{ml}, was replaced by their averages $\bar{u}_{ml}$; they were subsequently used to generate the diagnostic predictions by linearly combining the relative weights of symptoms w_l, $l = 1, \ldots, L$. That is, the average risk intervals for patients denoted by q_m, $m = 1, \ldots,$ M can be predicted by $\hat{q}_m = \sum_{l=1}^{L} \bar{u}_{ml} \cdot w_l$. The distance between q_m and $\hat{q}_m$ was calculated using $d_1(q_m,\ \hat{q}_m) = |q_m - \hat{q}_m|$. B_4 is a non-transfer learning variant of the ITMD method. The prediction model is the matrix factorization of $\boldsymbol{R} = \bar{\boldsymbol{U}}\boldsymbol{D}\boldsymbol{V}$. All methods were evaluated using the same test set. The averages and standard deviations of the experimental results from 20 random initializations are reported.

5.3. Experimental results

A comparison of the two sets of test results for the five different ratios is presented in Table 5 and Figs. 2 and 3. The experimental results indicate that the ITMD method exhibited the best performance for each group of records. In addition, the performance of the ITMD method improved as the source: target domain data ratio increased from 1:1 to 4:1. These results provide solid support for the effectiveness of knowledge transfer, both when the symptom spaces are the same and when they are not. The following four highlights can be derived from Tables 6 and 7.

(1) Baseline B_4, the non-transfer learning variant of the ITMD method, performed better than the other three baselines, which did not incorporate knowledge transfer. However, compared with the ITMD method, the performance of the four baselines was poor, indicating that the transfer of shared information from the source domain to the target domain significantly improves the final result.

21

(2) The standard deviations for B_1, B_2, and B_3 were significantly larger than those for B_4 and the ITMD method, which means that B_1, B_2, and B_3 lost some stability when the target domain had insufficient records, whereas B_4 and the ITMD method did not.

(3) In the experiments for all five groups of records, the ITMD method exhibited superior performance than the baselines. Further, the larger the ratio of data between the source and target domains, the better the ITMD method performed. The largest improvements in the ITMD method in comparison to the baselines are presented in Table 7. The largest improvements in the prediction accuracy of the ITMD method with the same and different symptom spaces were 0.0973 and 0.0827, respectively.

(4) As the number of records in the target domain increased, the performance of the ITMD method gradually decreased. Table 7 demonstrates that the maximum improvement rate for the ITMD method decreased from 13.43% to 8.15%, with the same symptom spaces, and from 11.3% to 8.85%, with different symptom spaces as the number of records increased from 2000 to 10000. This is reasonable because, with more data, the ITMD method does not need to provide as much support.

Table 6. Prediction accuracy for five groups of records with same and different symptom spaces.

Group ID	B_1	B_2	B_3	B_4	ITMD			
					1:1	2:1	3:1	4:1
G_1	0.7243±0.0215	0.7361±0.0236	0.7436±0.0217	0.7756±0.0108	0.7926±0.0069	0.8022±0.0083	0.8163±0.0113	**0.8216±0.0102**
G_2	0.7357±0.0233	0.7457±0.0214	0.7538±0.0233	0.7791±0.0096	0.7873±0.0088	0.7976±0.0073	0.8052±0.0082	**0.8231±0.0066**
G_3	0.7536±0.0218	0.7413±0.0243	0.7563±0.0211	0.7828±0.0045	0.7996±0.0077	0.8066±0.0089	0.8127±0.0061	**0.8202±0.0073**
G_4	0.7499±0.0238	0.7468±0.0211	0.7456±0.0178	0.7903±0.0079	0.8044±0.0041	0.8067±0.0063	0.8111±0.0046	**0.8179±0.0087**
G_5	0.7547±0.0213	0.7501±0.0137	0.7589±0.0205	0.8018±0.0063	0.8041±0.0046	0.8062±0.0073	0.8091±0.0059	**0.8112±0.0074**
G_6	0.7348±0.0176	0.7401±0.0201	0.7319±0.0219	0.7786±0.0043	0.7952±0.0066	0.8027±0.0052	0.8067±0.0069	**0.8146±0.0023**
G_7	0.7403±0.0198	0.7429±0.0211	0.7504±0.0168	0.7902±0.0062	0.8041±0.0046	0.8087±0.0035	0.8101±0.0042	**0.8144±0.0064**
G_8	0.7476±0.0161	0.7394±0.0143	0.7490±0.0157	0.7961±0.0069	0.8015±0.0037	0.8043±0.0054	0.8082±0.0041	**0.8151±0.0022**
G_9	0.7510±0.0214	0.7433±0.0153	0.7469±0.0188	0.7989±0.0073	0.8038±0.0033	0.8069±0.0025	0.8110±0.0051	**0.8155±0.0048**
G_{10}	0.7488±0.0139	0.7505±0.0123	0.7444±0.0149	0.8017±0.0047	0.8043±0.0026	0.8073±0.0035	0.8088±0.0019	**0.8103±0.0036**

Table 7. Improvement of the ITMD method for five groups of records with same and different symptom spaces.

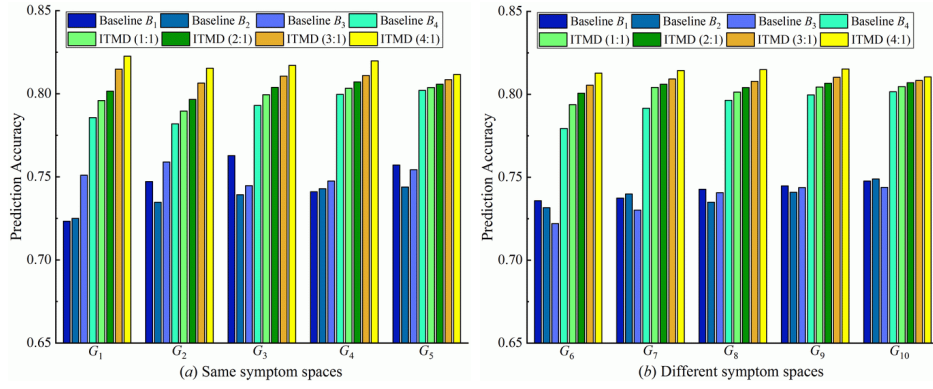| Group ID	1:1	2:1	3:1	4:1	Max improvement rate
G_1	0.0683	0.0779	0.0920	**0.0973**	**13.43%**
G_2	0.0516	0.0619	0.0695	0.0874	11.88%
G_3	0.0583	0.0653	0.0714	0.0789	10.64%
G_4	0.0588	0.0611	0.0655	0.0723	9.7%
G_5	0.0540	0.0561	0.0590	0.0611	8.15%
G_6	0.0633	0.0708	0.0748	**0.0827**	**11.3%**
G_7	0.0688	0.0734	0.0748	0.0791	10.76%
G_8	0.0621	0.0649	0.0688	0.0757	10.24%
G_9	0.0605	0.0636	0.0677	0.0722	9.71%
G_{10}	0.0599	0.0629	0.0644	0.0659	8.85%

Fig. 2. Prediction accuracy of all methods for G_1-G_{10} with same and different symptom spaces.
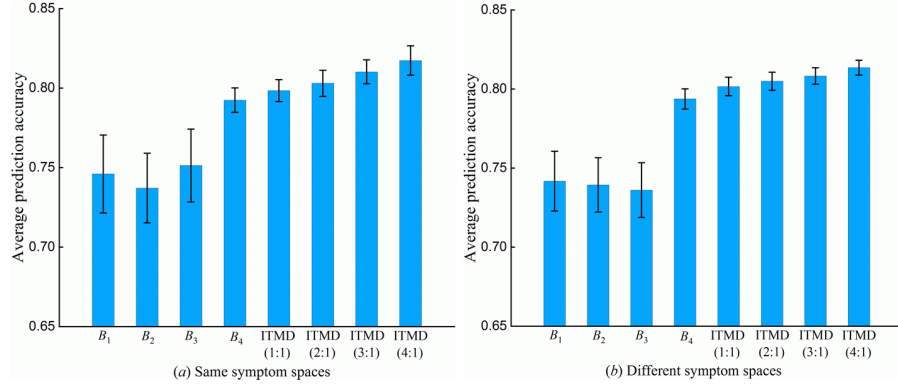

Fig. 3. Average prediction accuracy of all methods with same and different symptom spaces.

6. Real-world case study

This section presents a case study to demonstrate the operation and performance of the proposed ITMD method in practical situations. Thyroid cancer and breast cancer were considered as the source and target domains, respectively; additionally, real-world datasets of medical diagnoses were used for each domain.

6.1. Description of real-world case

This case was obtained from a tertiary-level hospital in China. It is famous for the diagnosis of thyroid cancer; however, it is inadequate in the diagnosis of breast cancer. Thus, the data related to the diagnosis of breast cancer are significantly less than those related to the diagnosis of thyroid cancer. Thyroid and breast cancers both have a high incidence rate, and both use ultrasound as a common means of early diagnosis. The ultrasonic diagnoses of thyroid and breast cancers mainly have three perspectives of similarity, as presented below.

(1) Both thyroid and breasts are superficial organs. Based on this, a high-frequency linear ultrasound probe with the same frequency of L14-5WU was used to check the structure, echogenicity, and vascularity of the two organs [54].

(2) Both thyroid and breasts are endocrine glands. In the two glands, a malignant tumor is generally created due to lesions and necrocytosis of the endocrine cells. The external features of the lesions include the burr of margin, calcification, and definition of contour [55].

(3) The criteria for assessing nodules and stratifying risks in thyroid and breast cancers are similar. BI-RADS was published by the American College of Radiology to help physicians in making uncertain diagnosis of breast cancer. Based on BIRADS, TIRADS was developed for the diagnosis of thyroid cancer [45].

From the first and second perspectives of similarity, the diagnostic process generally starts with detecting the thyroid and breast nodules by ultrasound and examining them according to five characteristics (symptoms): margin, contour, echogenicity, calcification, and vascularity. For the third perspective of similarity, both TI-RADS and BI-RADS include eight diagnostic categories, five of which are uncertain interval-valued symptoms[‡]. The five diagnostic categories are generally used by physicians in hospitals; that is, TI/BI-RADS 3, 4A, 4B, 4C, and 5. Details regarding TI-RADS and BI-RADS used in the ultrasonic department of the hospital are presented in Table 8. Physicians use the aforementioned five diagnostic categories to evaluate the tumor on each symptom. By using TI/BI-RADS to express the evaluations of symptoms and the diagnoses of cancer risks, the relations between symptoms are consistent in the diagnoses of thyroid and breast cancers.

In clinical practice, physicians transform the observations on the five symptoms into TIRADS categories and further the risk intervals {[0, 0.03], (0.03, 0.25], (0.25, 0.75], (0.75, 0.95], (0.95, 1]}, which are used to describe the relations between patients and symptoms. Five diagnostic categories may be recommended to patients, and their cancer risk intervals are used to determine the relations between patients and diagnostic categories. For example, when a radiologist evaluates a nodule of thyroid and breast on the five symptoms (TI/BI-RADS 3, TI/BI-RADS 4A, TI/BI-RADS 4B, TI/BI-RADS 3, TI/BI-RADS 4A), the overall diagnosis of the thyroid and breast nodules will be TI/BI-RADS 4A, which is generated by synthetically considering the evaluation of symptoms. Thus, the high similarity between the diagnoses of thyroid and breasts results in consistent relations between the symptoms of the two diseases, and such relations can be used as the shared knowledge transferred across the two diseases.

Our datasets comprise the diagnostic assessments of ultrasonic examination collected by four physicians from 2017 to 2019. The statistical information is presented in Table 9. Similar to Section 5, both records in the source and target domains are used to train the proposed model discussed in Section 4.4. The dataset in the target domain is divided into two parts using five-fold cross-validation.

Table 8. TI/BI-RADS used in the hospital.

[‡] The other three diagnostic categories are TI/BI-RADS 1: negative (0% risk of malignancy); TI/BI-RADS 2: benign (0% malignancy), and TI/BI-RADS 6: biopsy proven malignancy.

Categories	Findings	Cancer risk
3	Probably benign	[0, 0.03)
4A	Undetermined	(0.03, 0.25]
4B	Suspicious	(0.25, 0.75]
4C	High suspicious	(0.75, 0.95]
5	Probably malignant	(0.95, 1]

Table 9. Statistics for the diagnostic records of four physicians

Physician ID	Serving periods	Number of patients		1:1		2:1		3:1		4:1	
		Thyroid	Breast	Source	Target	Source	Target	Source	Target	Source	Target
F_1	2017-2019	627	153	143	143	286	143	429	143	572	143
F_2	2017-2019	632	166	153	153	306	153	459	153	612	153
F_3	2017-2019	685	138	125	125	250	125	375	125	500	125
F_4	2017-2019	475	107	91	91	182	91	273	91	364	91

The four methods, B_1, B_2, B_3, and B_4, mentioned in Section 5 were selected as the baselines for comparison with the ITMD method. The experiments for the ITMD method were conducted with four ratios of records between the source and target domains, which were set to 1:1, 2:1, 3:1, and 4:1, respectively. The regularization factor was set to 0.5. The experiments for all compared methods were conducted on the same test sets. The results reported denote the averages with standard deviations from 20 random initializations.

6.2. Case study results and analysis

The results are presented in Table 11 and a visual comparison is shown in Fig. 4, which leads to the following observations:

(1) The ITMD method performed better than the four baselines with the datasets of all four physicians. Table 12 indicates that in comparison to the four baselines, the largest improvement in prediction accuracy using the ITMD method was 0.1221 and its improvement rate was 16.69%.

(2) B_1, B_2, and B_3 performed so poorly that they would not be acceptable for diagnosing breast cancer without more data. Moreover, with such few data, none of the three methods could maintain stability.

(3) The ITMD method became more accurate as the ratio of records increased. From Table 12, when the source: target ratio increased from 1:1 to 4:1 for physician F_1, the improvement of prediction accuracy with the ITMD method increased from 0.0608 to 0.1095 in comparison to the four baselines. Similar results were observed for the other three physicians.

Table 11. Prediction accuracy for the datasets of four physicians.

Physician ID	B_1	B_2	B_3	B_4	ITMD			
					1:1	2:1	3:1	4:1
F_1	0.7384±0.0069	0.7262±0.0058	0.7307±0.0045	0.7767±0.0030	0.7870±0.0082	0.7999±0.0059	0.8179±0.0073	**0.8357±0.0014**
F_2	0.7028±0.0306	0.6943±0.0401	0.6897±0.0346	0.7384±0.0018	0.7464±0.0085	0.7582±0.0063	0.7720±0.0023	**0.7777±0.0042**
F_3	0.7180±0.0067	0.7003±0.0095	0.7185±0.0257	0.7766±0.0006	0.7837±0.0016	0.7945±0.0029	0.8016±0.0027	**0.8076±0.0028**
F_4	0.7381±0.0093	0.7317±0.0024	0.7423±0.0157	0.8273±0.0018	0.8356±0.0011	0.8399±0.0018	0.8451±0.0011	**0.8538±0.0019**

Table 12. Improvement of the ITMD method for five groups of records with same and different symptom spaces.

Physician ID	1:1	2:1	3:1	4:1	Max improvement rate
F_1	0.0608	0.0737	0.0917	0.1095	15.08%

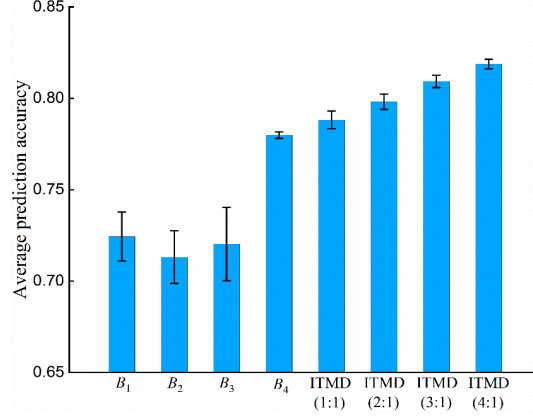|					
F_2	0.0567	0.0685	0.0823	0.0880	12.76%
F_3	0.0834	0.0942	0.1013	0.1073	15.32%
F_4	0.1039	0.1082	0.1134	**0.1221**	**16.69%**

Fig. 4. Average prediction accuracy for all methods.

To further demonstrate the performance of the ITMD method, we conducted experiments when the symptom relation matrix was directly learned from the dataset in the source domain without adjustments. Table 13 presents the prediction results only with the datasets in the source domain.

Table 13. Prediction accuracy for the datasets of four physicians in the source domain

Physician ID	Prediction accuracy			
	1:1	2:1	3:1	4:1
F_1	**0.7816**± 0.0018	**0.7852**± 0.0028	**0.7895**± 0.0027	**0.7919**± 0.0039
F_2	**0.7246**+ 0.0027	**0.7332**± 0.0041	**0.7350**± 0.0014	**0.7432**± 0.0030
F_3	**0.7790**± 0.0003	**0.7797**± 0.0010	**0.7815**± 0.0011	**0.7868**± 0.0010
F_4	**0.8258**± 0.0020	**0.8262**± 0.0012	**0.8296**± 0.0010	**0.8314**± 0.0014

By comparing the results in Tables 11 and 13, we found that the use of the dataset in the source domain for generating recommendations will have minimal effect on the improvement of the prediction accuracy in the target domain. Meanwhile, although there are only a few labels in the target domain, the results indicate that they can significantly improve the prediction accuracy.

In addition to the prediction accuracy of the ITMD method, we wanted to assess the effectiveness of our new dissimilarity measurement. Table 14 presents the different distance measurements applied to the B_4 baseline on the four sets of medical records. Our dissimilarity measurement clearly led to more accurate predictions than other distance measures.

Table 14. Prediction accuracy on the target domain with different distance measurements.

	Euclidean distance	Manhattan distance	Chebyshev distance	Dissimilarity measurement
F_1	0.5917±0.0057	0.6428±0.0043	0.6435±0.0029	**0.7767**±0.0030
F_2	0.6076±0.0054	0.6174±0.0050	0.6116±0.0065	**0.7384**±0.0018
F_3	0.6349±0.0015	0.6463±0.0036	0.6450±0.0042	**0.7766**±0.0006
F_4	0.6140±0.0042	0.6083±0.0010	0.6072±0.0010	**0.8273**±0.0018
$\overline{F}$	0.6120±0.0042	0.6287±0.0035	0.6268±0.0036	**0.7798**±0.0018

We also examined the influence of λ with a record ratio of 4:1; the results are shown in Fig. 5. As illustrated, the prediction accuracy monotonically increases when λ varies from 0 to 0.5, with no significant change in accuracy from 0.5 to 1.
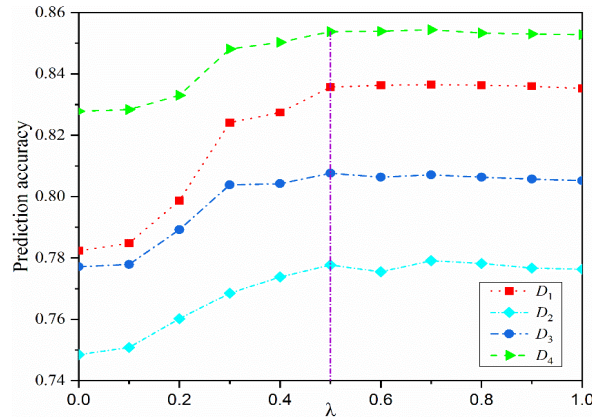

Fig. 5. Prediction accuracy for the four physicians with changes in λ.

7. Conclusion and future study

In this study, we proposed a new cross-domain recommender system called ITMD to provide personalized recommendations to physicians to determine the disease risks of patients for various diseases. The ITMD method was developed by solving the two challenges of uncertainty in medical diagnosis data and mismatched symptom spaces. A new dissimilarity measurement was developed for diagnosis with interval numbers to depict the dissimilarities between the diagnoses and their predictions to manage the first challenge. A space alignment technique was adopted to align the different symptom spaces of the two domains to manage the second challenge. A collective matrix factorization technique was constructed using the ITMD method based on the new measurement and space alignment technique to implement shared information transfer from the source domain to the target domain. The proposed ITMD method can provide effective decision support in the domain of disease diagnosis with insufficient records. A set of experiments were conducted on synthetic data and a case study was conducted with real-world data for the diagnoses of thyroid cancer and breast cancer. The results suggested that the ITMD method exhibits superior performance and can provide personalized recommendations to support physicians with diagnosing the risks associated with various diseases. The significance of the results in this paper lies in providing a novel recommendation method in both theoretical and practical way. 1) Theoretically, this paper solves the data sparsity problem in medical diagnosis with the solution of a cross-domain recommendation method that is able to effectively transfer knowledge from a similar domain. The proposed method is able to deal with two theoretical issues: uncertain and heterogeneous

representations. 2) Practically, the problem is derived from a real-world scenario occurs in hospital and our solution can be directly used to support the radiologists with insufficient experiences and help improve their diagnostic accuracy.

In future study, we intend to focus on extending the principles of the ITMD method to more complex settings, such as the transfer of heterogeneous diagnostic data across domains. We will also consider other salient aspects of medical diagnoses, such as patient privacy, the attitudes of physicians toward risks, and the gold standards in medical diagnosis.

Acknowledgments

Appendix A. Proof of Property 1

Proof. Eqs. (4)-(7) are verified in *A.1-4*.

A.1. Because $(r_{ml} - \hat{r}_{ml})^2 \cdot (r_{mk} - \hat{r}_{mk})^2 \cdot (\bar{p}(C_k) - \bar{p}(C_l))^2 \geq 0$ and $\bar{p}(C_N) - \bar{p}(C_1) > 0$ in Eq. (3), it follows that $d(\boldsymbol{R},$

$\hat{\boldsymbol{R}}) \geq 0$. In addition, $\sum_{l=1}^{N-1}\sum_{k=l+1}^{N} (r_{ml} - \hat{r}_{ml})^2 \cdot (r_{mk} - \hat{r}_{mk})^2 \cdot (\bar{p}(C_k) - \bar{p}(C_l))^2 \leq (\bar{p}(C_N) - \bar{p}(C_1))^2$.

Thus, we can deduce that

$$\frac{\sum_{m=1}^{M}\sqrt{\sum_{l=1}^{N-1}\sum_{k=l+1}^{N}(r_{ml} - \hat{r}_{ml})^2 \cdot (r_{mk} - \hat{r}_{mk})^2 \cdot (\bar{p}(C_k) - \bar{p}(C_l))^2}}{M \cdot (\bar{p}(C_N) - \bar{p}(C_1))} \leq \frac{\sum_{m=1}^{M}\bar{p}(C_N) - \bar{p}(C_1)}{M \cdot (\bar{p}(C_N) - \bar{p}(C_1))} = 1.$$

Correspondingly, we have $0 \leq d(\boldsymbol{R}, \hat{\boldsymbol{R}}) \leq 1$, which verifies Eq. (4).

A.2. Eq. (5) is easily proved.

A.3. Two things must be verified in Eq. (6):

(1) If $\boldsymbol{R} = \hat{\boldsymbol{R}}$, we have $r_{mn} = \hat{r}_{mn}$, m=1, …, M, n=1, …, N. By substituting this into Eq. (3), we obtain $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) = 0$.

(2) According to Eq. (3), we deduced from $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) = 0$ that $(r_{ml} - \hat{r}_{ml})^2 \cdot (r_{mk} - \hat{r}_{mk})^2 \cdot (\bar{p}(C_k) - \bar{p}(C_l))^2 = 0$, m=1, …, M, l=1, …, N-1, k=l+1, …, N. Because $\bar{p}(C_k) - \bar{p}(C_l) > 0$, l=1, …, N-1, k=l+1, …, N, we can conclude that $r_{mn} = \hat{r}_{mn}$, m=1, …, M, n = 1, …, N, i.e., $\boldsymbol{R} = \hat{\boldsymbol{R}}$. Thus, Eq. (6) is verified.

A.4. Two things must be verified in Eq. (7):

(1) Under the condition that $r_{m1}=1$ and $\hat{r}_{mN}=1$ ($m = 1, \ldots, M$), we can infer from $\sum_{n=1}^{N} r_{mn} = 1$ and $\sum_{n=1}^{N} \hat{r}_{mn} = 1$ that

$r_{mn}=0$, $n = 2, \ldots, N$ and $\hat{r}_{mn} = 0$, $n = 1, \ldots, N-1$. Then, $d(\boldsymbol{R}, \hat{\boldsymbol{R}})$ can be calculated as $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) =$

$$\frac{\sum_{m=1}^{M} \sqrt{(r_{m1} - \hat{r}_{m1})^2 \cdot (r_{mN} - \hat{r}_{mN})^2 \cdot (\bar{p}(C_N) - \bar{p}(C_1))^2}}{M \cdot (\bar{p}(C_N) - \bar{p}(C_1))} = 1.$$ Conditioned on $r_{mN}=1$ and $\hat{r}_{m1}=1$ ($m = 1, \ldots, M$), $d(\boldsymbol{R}, \hat{\boldsymbol{R}})=1$

can be similarly verified.

(2) Because $\bar{p}(C_N) - \bar{p}(C_1)$ is the maximum value of $\bar{p}(C_k) - \bar{p}(C_l)$, $l = 1, \ldots, N-1$, $k = l+1, \ldots, N$, we prioritize

the assignment of r_{ml} and $\hat{r}_{ml}$ ($l = 1, \ldots, L$) to $(r_{m1} - \hat{r}_{m1})^2 \cdot (r_{mN} - \hat{r}_{mN})^2$ for achieving the maximum value of $d(\boldsymbol{R}, \hat{\boldsymbol{R}})$,

i.e., 1. This implies that when $d(\boldsymbol{R}, \hat{\boldsymbol{R}}) = 1$, $(r_{m1} - \hat{r}_{m1})^2 \cdot (r_{mN} - \hat{r}_{mN})^2 = 1$ is satisfied, which indicates that $(r_{m1} - \hat{r}_{m1})^2 = 1$

and $(r_{mN} - \hat{r}_{mN})^2 = 1$. Then, we have ($r_{m1}=1$, $\hat{r}_{mN}=1$) or ($r_{mN}=1$, $\hat{r}_{m1}=1$). Thus, Eq. (7) is verified. Accordingly, Property

1 is verified.

Appendix B. Proof of Theorem 1

Proof. Two things must be verified for Eq. (8):

(1) $\dfrac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F} \geq \dfrac{1}{cond(\boldsymbol{D}_t) \cdot cond(\boldsymbol{V}_t)} \cdot \dfrac{\|\Delta \bar{\boldsymbol{U}}_t\|_F}{\|\bar{\boldsymbol{U}}_t\|_F}$ and (2) $\dfrac{\|\Delta \boldsymbol{R}_t\|_F}{\|\boldsymbol{R}_t\|_F} \leq cond(\boldsymbol{D}_t) \cdot cond(\boldsymbol{V}_t) \cdot \dfrac{\|\Delta \bar{\boldsymbol{U}}_t\|_F}{\|\bar{\boldsymbol{U}}_t\|_F}$.

(1) Suppose that $\Delta \boldsymbol{D}_t$ has a small value and satisfies $\left\| \Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+} \right\|_F < 1$, where the symbol "+" represents the

generalized inverse of a matrix, and

$$\boldsymbol{R}_t + \Delta \boldsymbol{R}_t = (\bar{\boldsymbol{U}}_t + \Delta \bar{\boldsymbol{U}}_t)(\boldsymbol{D}_t + \Delta \boldsymbol{D}_t)\boldsymbol{V}_t, \tag{B.1}$$

From Eq. (B.1), we deduce that

$$\Delta \bar{\boldsymbol{U}}_t (\boldsymbol{D}_t + \Delta \boldsymbol{D}_t)\boldsymbol{V}_t = \Delta \boldsymbol{R}_t - \bar{\boldsymbol{U}}_t \Delta \boldsymbol{D}_t \boldsymbol{V}_t \Rightarrow \Delta \bar{\boldsymbol{U}}_t (\boldsymbol{D}_t + \Delta \boldsymbol{D}_t)\boldsymbol{V}_t \boldsymbol{V}_t^{+} \boldsymbol{D}_t^{+} = (\Delta \boldsymbol{R}_t - \bar{\boldsymbol{U}}_t \Delta \boldsymbol{D}_t \boldsymbol{V}_t)\boldsymbol{V}_t^{+} \boldsymbol{D}_t^{+}$$

$$\Rightarrow \Delta \bar{\boldsymbol{U}}_t = (\Delta \boldsymbol{R}_t \boldsymbol{V}_t^{+} - \bar{\boldsymbol{U}} \Delta \boldsymbol{D}_t)\boldsymbol{D}_t^{+}(\boldsymbol{I} + \Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+})^{+}. \tag{B.2}$$

Because $\left\| -\Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+} \right\|_F = \left\| \Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+} \right\|_F < 1$, we have

$$\left\| (\boldsymbol{I} + \Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+})^{+} \right\|_F \leq \frac{1}{1 - \left\| \Delta \boldsymbol{D}_t \boldsymbol{D}_t^{+} \right\|_F} \leq \frac{1}{1 - \left\| \Delta \boldsymbol{D}_t \right\|_F \left\| \boldsymbol{D}_t^{+} \right\|_F}. \tag{B.3}$$

According to Eqs. (B.2) and (B.3), we can conclude that

$$\left\|\Delta\bar{U}_t\right\|_F = \left\|(\Delta R_t V_t^+ - \bar{U}_t \Delta D_t)D_t^+(I + \Delta D_t D_t^+)^+\right\|_F \leq (\frac{\left\|\Delta R_t\right\|_F \left\|V_t^+\right\|_F}{\left\|\bar{U}_t\right\|_F \left\|D_t\right\|_F} + \frac{\left\|\Delta D_t\right\|_F}{\left\|D_t\right\|_F})\frac{\left\|\bar{U}_t\right\|_F \left\|D_t\right\|_F \left\|D_t^+\right\|_F}{1 - \left\|\Delta D_t\right\|_F \left\|D_t^+\right\|_F}. \tag{B.4}$$

Because $R_t = \bar{U}_t D_t V_t$, we have

$$\left\|R_t\right\|_F \leq \left\|\bar{U}_t\right\|_F \left\|D_t\right\|_F \left\|V_t\right\|_F. \tag{B.5}$$

From Eqs. (B.4) and (B.5), we can deduce that

$$\frac{\left\|\Delta\bar{U}_t\right\|_F}{\left\|\bar{U}_t\right\|_F} \leq (\frac{\left\|\Delta R_t\right\|_F \left\|V_t\right\|_F \left\|V_t^+\right\|_F}{\left\|R_t\right\|_F} + \frac{\left\|\Delta D_t\right\|_F}{\left\|D_t\right\|_F}) \cdot \frac{\left\|D_t\right\|_F \left\|D_t^+\right\|_F}{1 - \left\|\Delta D_t\right\|_F \left\|D_t^+\right\|_F} = (\frac{\left\|\Delta R_t\right\|_F}{\left\|R_t\right\|_F} \cdot cond(V_t) + \frac{\left\|\Delta D_t\right\|_F}{\left\|D_t\right\|_F}) \cdot \frac{cond(D_t)}{1 - \left\|\Delta D_t\right\|_F \left\|D_t^+\right\|_F}.$$

Suppose that $\Delta D = O$, then it can be deduced into $\dfrac{\left\|\Delta R_t\right\|_F}{\left\|R_t\right\|_F} \geq \dfrac{1}{cond(D_t) \cdot cond(V_t)} \cdot \dfrac{\left\|\Delta\bar{U}_t\right\|_F}{\left\|\bar{U}_t\right\|_F}.$

(2) Because $R_t + \Delta R_t = (\bar{U}_t + \Delta\bar{U}_t)D_t V_t$ and $R_t = \bar{U}_t D_t V_t$, we have $\Delta R_t = \Delta\bar{U}_t D_t V_t$. Thus, $\left\|\Delta R_t\right\|_F \leq$

$\left\|\Delta\bar{U}_t\right\|_F \left\|D_t\right\|_F \left\|V_t\right\|_F$, which indicates that $\dfrac{\left\|\Delta\bar{U}_t\right\|_F}{\left\|\bar{U}_t\right\|_F} \geq \dfrac{\left\|\Delta R_t\right\|_F}{\left\|D_t\right\|_F \left\|V_t\right\|_F \left\|\bar{U}_t\right\|_F}.$

From $R_t = \bar{U}_t D_t V_t$, we have $\bar{U}_t = R_t V_t^+ D_t^+$, which infers $\left\|\bar{U}_t\right\|_F \leq \left\|R_s\right\|_F \left\|V_s^+\right\|_F \left\|D_s^+\right\|_F$. Combining the two

inequations, we have $\dfrac{\left\|\Delta\bar{U}_t\right\|_F}{\left\|\bar{U}_t\right\|_F} \geq \dfrac{\left\|\Delta R_t\right\|_F}{\left\|D_t\right\|_F \left\|V_t\right\|_F \left\|R_t\right\|_F \left\|V_t^+\right\|_F \left\|D_t^+\right\|_F} = \dfrac{\left\|\Delta R_t\right\|_F}{cond(D_t)cond(V_t)\left\|R_t\right\|_F}$; that is,

$$\frac{\left\|\Delta R_t\right\|_F}{\left\|R_t\right\|_F} \leq cond(D_t) \cdot cond(V_t) \cdot \frac{\left\|\Delta\bar{U}_t\right\|_F}{\left\|\bar{U}_t\right\|_F}.$$

Thus, Eq. (8) is verified. Similarly, Eq. (9) is proved and, correspondingly, Theorem 1 is verified.

References

[1] A.H. Shahid, M.P. Singh, Computational intelligence techniques for medical diagnosis and prognosis: Problems and current developments, Biocybern. Biomed. Eng. 39 (2019) 638-672.

[2] Y. Gao, P.G. Church, Improving molecular cancer class discovery through sparse non-negative matrix factorization, Bioinformatics 21 (21) (2005) 3970-3975.

[3] A.S. Hussein, W.M. Omar, X. Li, M. Ati, Efficient chronic disease diagnosis prediction and recommendation system, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE, 2012, pp. 17-19.

[4] S. Hassan, Z. Syed, From Netflix to heart attacks: collaborative filtering in medical datasets, Proceedings of the 1st ACM International Health Informatics Symposium, ACM, 2010, pp. 128-134.

[5] N.T. Thong, L.H. Son, HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis, Expert Syst. Appl. 42 (2015) 3682-3701.

[6] T.C.K. Huang, Y.L. Chen, M.C. Chen, A novel recommendation model with Google similarity, Decis. Support Syst. 89 (2016) 17-27.

[7] M. Mao, J. Lu, G. Zhang, J. Zhang, Multirelational social recommendations via multigraph ranking, IEEE Trans. Cybern. 47 (12) (2017) 4049-4061.

[8] J. Son, S.B. Kim, Academic paper recommender system using multilevel simultaneous citation networks, Decis. Support Syst. 105 (2018) 24-33.

[9] S. Bag, S. Kumar, A. Awasthi, M.K. Tiwari, A noise correction-based approach to support a recommender system in a highly sparse rating environment, Decis. Support Syst. 118 (2019) 46-57.

[10] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: a survey, Decis. Support Syst. 74 (2015) 12-32.

[11] C.A.N. Okoromah, A.F.E. Lesi, 2004. Diazepam for treating tetanus. Cochrane Database Syst. Rev. 19, CD003954.

[12] H. Younis, M.H. Bhatti, M. Azeem, Classification of skin cancer dermoscopy images using transfer learning, 2019 15th International Conference on Emerging Technologies (ICET), IEEE, 2019, pp. 1-4.

[13] A. Ramcharan, K. Baranowski, P. Mccloskey, B. Ahmed, J.P. Legg, D.P. Hughes, Deep learning for image-based cassava disease detection, Front. Plant Sci. 8 (2017) 1852.

[14] K. Zhou, W. He, Y. Xu, G. Xiong, J. Cai, Feature selection and transfer learning for Alzheimer's disease clinical diagnosis, Appl. Sci. 8 (8) (2018) 1372.

[15] Q. Do, W. Liu, J. Fan, D. Tao, Unveiling hidden implicit similarities for cross-domain recommendation, IEEE Trans. Knowledge Data Eng. (2019) 33 (1) (2021) 302-315.

[16] S.W. Dubrey, K. Cha, M. Skinner, M.P. Lavalley, R.H. Falk, Familial and primary (AL) cardiac amyloidosis: echocardiographically similar diseases with distinctly different clinical outcomes, Heart 78 (1) (1997) 74-82.

[17] T. Imasawa, M. Tanaka, N. Maruyama, T. Kawaguchi, Y. Yamaguchi, R. Rossignol, H. Kitamura, M. Nishimura, 2014. Pathological similarities between low birth weight-related nephropathy and nephropathy associated with mitochondrial cytopathy. Diagn. Pathol. 9, 181.

[18] C.Y. Ok, K.P. Patel, G. Garcia-Manero, M.J. Routbort, J. Peng, G. Tang, M. Goswami, K.H. Young, R. Singh, L.J. Medeiros, H.M. Kantarjian, R. Luthra, S.A. Wang, 2015. TP53 mutation characteristics in therapy-related myelodysplastic syndromes and acute myeloid leukemia is similar to de novo diseases, J. Hem. Oncol. 8, 45.

[19] Q. Zhang, G. Zhang, J. Lu, D. Wu, A framework of hybrid recommender system for personalized clinical prescription, 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), IEEE, 2015, pp. 189-195.

[20] A. Mustaqeem, S.M. Anwar, M. Majid, 2020, A modular cluster based collaborative recommender system for cardiac patients, Artif. Intell. Med. 102, 101761.

[21] L. Sun, C. Liu, C. Guo, H. Xiong, Y. Xie, Data-driven automatic treatment regimen development and recommendation, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 1865-1874.

[22] D.A. Davis, N.V. Chawla, N. Blumm, N. Christakis, A.L. Barabasi, Predicting individual disease risk based on medical history, Proceedings of the 17th ACM Conference on Information and Knowledge Management, ACM, 2008, pp. 769-778.

[23] B. Li, Q. Yang, X. Xue, Transfer learning for collaborative filtering via a rating-matrix generative model, Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 617-624.

[24] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, Z. Ma, J. Guo, A cross-domain recommendation model for cyber physical systems, IEEE Trans. Emerging Top. Comput. 1 (2) (2013) 384-393.

[25] Q. Zhang, D. Wu, J. Lu, F. Liu, G. Zhang, A cross-domain recommender system with consistent information transfer, Decis. Support. Syst. 104 (2017) 49-63.

[26] W. Pan, E.W. Xiang, N.N. Liu, Q. Yang, Transfer learning in collaborative filtering for sparsity reduction, The 24th AAAI Conference on Artificial Intelligence, ACM, 2010, pp. 230-235.

[27] W. Pan, Q. Yang, Transfer learning in heterogeneous collaborative filtering domains, Artif. Intell. 197 (2013) 39-55.

[28] Q. Zhang, J. Lu, D. Wu, G. Zhang, A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities, IEEE Trans. Neural Netw. Learn. Syst. 30 (7) (2019) 1998-2012.

[29] Y. Shi, M. Larson, A. Hanjalic, Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges, ACM Comput. Surv. 47 (1) (2014) 1-45.

[30] W. Chen, W. Hsu, M.L. Lee, Making recommendations from multiple domains, Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 892-900.

[31] M. Jiang, P. Cui, X. Chen, F. Wang, W. Zhu, S. Yang, Social recommendation with cross-domain transferable knowledge, IEEE Trans. Knowledge Data Eng. 27 (11) (2015) 3084-3097.

[32] Y. Shi, M. Larson, A. Hanjalic, Tags as bridges between domains: improving recommendation with tag-induced cross-domain collaborative filtering, Proceedings of the 19th International Conference on User Modeling, Adaptation, and Personalization, ACM, 2011, pp. 305-316.

[33] Z. Fang, S. Gao, B. Li, J. Li, J. Liao, Cross-domain recommendation via tag matrix transfer, IEEE International Conference on Data Mining Workshop, IEEE, 2015, pp. 1235-1240.

[34] P. Hao, G. Zhang, L. Martinez, J. Lu, Regularizing knowledge transfer in recommendation with tag-inferred correlation, IEEE Trans. Cybern. 49 (1) (2019) 83-96.

[35] X. Xin, Z. Liu, C.Y. Lin, H. Huang, X. Wei, P. Guo, Cross-domain collaborative filtering with review text, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, ACM, 2015, pp. 1827-1833.

[36] T. Song, Z. Peng, S. Wang, W. Fu, X. Hong, P.S. Yu, Review-based cross-domain recommendation through joint tensor factorization, International Conference on Database Systems for Advanced Applications, Springer, 2017, pp. 525-540.

[37] I. Fernández-Tobías, I. Cantador, P. Tomeo, V.W. Anelli, T. Di Noia, Addressing the user cold start with cross-domain collaborative filtering: exploiting item metadata in matrix factorization, User Model. User-Adapt. Interact. 29 (2019) 443-486.

[38] C. Jiang, X. Han, G.R. Liu, G.P. Liu, A nonlinear interval number programming method for uncertain optimization problems, Eur. J. Oper. Res. 188 (1) (2008) 1-13.

[39] D.D. Lee, H.S. Sebastian, Algorithms for non-negative matrix factorization, Proceedings of the 13th International Conference on Neural Information Processing Systems, ACM, 2000, pp. 535-541.

[40] W. Wang, G. Zhang, J. Lu, Member contribution-based group recommender system, Decis. Support Syst. 87 (2016) 80-93.

[41] G. Adomavicius, Y.O. Kwon, New recommendation techniques for multicriteria rating systems, IEEE Intel. Syst. 22 (3) (2007) 48-55.

[42] A.K. Cline, C.B. Moler, G.W. Stewart, J.H. Wilkinson, An Estimate for the Condition Number of a Matrix, SIAM J. Numer. Anal. 16 (2) (1979) 368-375.

[43] P.T. Boggs, J.W. Tolle, Sequential Quadratic Programming, Acta Numer. (1995) 199-242.

[44] B.K. Chan, T.S. Desser, I.R. McDougall, R.J. Weigel, R.B. Jeffrey, Common and uncommon sonographic features of papillary thyroid carcinoma, J. Ultra. Med. 22 (10) (2003) 1083-1090.

[45] C. Fu, W. Chang, W. Liu, S. Yang, 2019, Data-driven group decision making for diagnosis of thyroid nodule, Sci. China Info. Sci. 62 (11), 212205.

[46] K. Akahoshi, Y. Chijiwa, S. Hamada, I. Sasaki, H. Nawata, T. Kabemura, D. Yasuda, H. Okabe, Pretreatment staging of endoscopically early gastric cancer with a 15 MHz ultrasound catheter probe, Gastrointest. Endosc. 48 (1998) 470-476.

[47] T. Yoshida, H. Matsue, N. Okazaki, M. Yoshino, Ultrasonographic differentiation of hepatocellular carcinoma from metastatic liver cancer, J. Clin. Ultrasound 15 (1987) 431-437.

[48] H.S. Ahn, H.J. Lee, S. Hahn, W.H. Kim, K.U. Lee, T. Sano, S.B. Edge, H.K. Yang, Evaluation of the seventh American Joint Committee on Cancer/International Union Against Cancer classification of gastric adenocarcinoma in comparison with the sixth classification, Cancer 116 (2010) 5592-5598.

[49] G. Torzilli, M. Makuuchi, Intraoperative ultrasonography in liver cancer, Surg. Oncol. Clin. N. Am. 12 (2003) 91-103.

[50] P. Memar, F. Faradji, A novel multi-class EEG-based sleep stage classification system, IEEE Trans. Knowledge Data Eng. 32 (8) (2020) 1586-1594.

[51] I. Cantador, I. Fernández-Tobías, S. Berkovsky, P. Cremonesi, Cross-domain recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender systems handbook, Springer, Boston, MA, 2015, pp. 919-959.

[52] C. Fu, W. Chang, D. Xu, S. Yang, An evidential reasoning approach based on criterion reliability and solution reliability, Comput. Ind. Eng. 128 (2019) 402-417.

[53] A. Sengupta, T.K. Pal, Fuzzy preference ordering of interval numbers in decision problems, Springer, New York, 2009.

[54] X. Liang, J. Yu, J. Liao, Z. Chen, 2020, Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging, BioMed Res. Int. 1763803.

[55] D. Schottenfeld, The relationship of breast cancer to thyroid disease, J. Chron. Dis. 21 (5) (1968) 303-313.