ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

# UNIVERSITY OF PIRAEUS

## DEPARTMENT OF DIGITAL SYSTEMS

## POSTGRADUATE PROGRAMME: INFORMATION SYSTEMS & SERVICES

SPECIALIZATION: ADVANCED INFORMATION SYSTEMS

# TRANSFER LEARNING IN RECOMMENDER SYSTEMS

Christina Zografou (ME1832)

# Table Of Contents

# Abstract

Recommendation systems are becoming more and more important nowadays providing companies with a great advantage. Data sparseness, though, is a major problem for collaborative filtering (CF) techniques in recommender systems, especially for new users and items, making it difficult to make good recommendations. That is where Transfer Learning comes in. In this study, the goal is to improve the quality of recommendations in a sparse dataset by exploiting knowledge from another dataset. More specifically, the goal is to improve the recommendations in the CDs & Vinyl domain by utilizing rating data from the Digital Music domain. Experimental results show that CDs & Vinyl recommendations can indeed benefit from making use of Digital Music recommendations.

Τα συστήματα συστάσεων γίνονται ολοένα και πιο σημαντικά στις μέρες μας προσφέροντας μεγάλο πλεονέκτημα στις επιχειρήσεις. Όμως, η ύπαρξη λίγων και αραιών δεδομένων είναι ένα σημαντικό πρόβλημα για τα συστήματα συστάσεων που κάνουν χρήση τεχνικών collaborative filtering (CF), ειδικά σε περιπτώσεις νέων χρηστών ή νέων αντικειμένων, δυσκολεύοντας τη δημιουργία προτάσεων. Εδώ έγκειται η Μεταφορά Γνώσης (Transfer Learning). Σε αυτή τη μελέτη, ο στόχος είναι να βελτιωθεί η ποιότητα των προτάσεων σε ένα σύνολο δεδομένων χρησιμοποιώντας γνώση από ένα άλλο σύνολο δεδομένων. Πιο συγκεκριμένα, ο στόχος είναι να βελτιωθούν οι συστάσεις στον τομέα CDs & Vinyl χρησιμοποιώντας βαθμολογίες από τον τομέα Digital Music. Τα πειραματικά αποτελέσματα δείχνουν ότι οι προτάσεις στον τομέα CDs & Vinyl μπορούν όντως να επωφεληθούν από τη χρήση γνώσης από τον τομέα Digital Music.

# Chapter 1: Transfer Learning

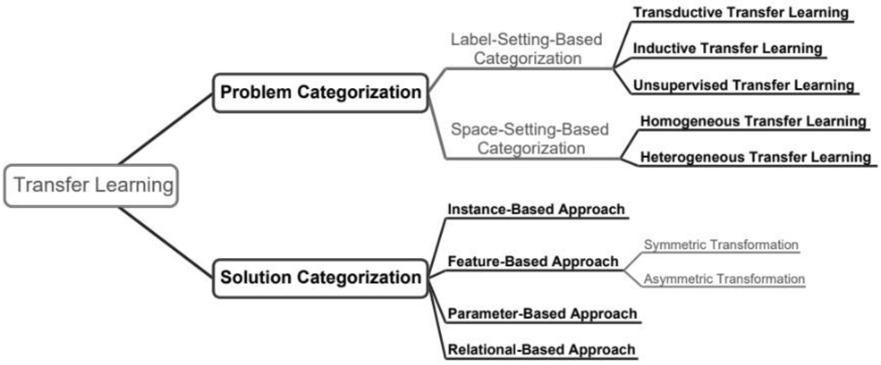
This chapter contains the basic definitions and categorizations of Transfer Learning, some of the most important related studies in each category and a review on the type of applications it has been used for.

## 1.1 Transfer Learning Definition & Categorizations

The fundamental motivation for Transfer learning in the field of machine learning was discussed in a NIPS-95 workshop on "Learning to Learn", which focused on the need for lifelong machine-learning methods that retain and reuse previously learned knowledge. Research on transfer learning has attracted more and more attention since 1995 in different names: learning to learn, lifelong learning, knowledge transfer, inductive transfer, multi-task learning, knowledge consolidation, context-sensitive learning, knowledge-based inductive bias, meta learning, and incremental/cumulative learning. In 2005, the Broad Agency Announcement (BAA) 05-29 of Defense Advanced Research Projects Agency (DARPA)'s Information Processing Technology Office (IPTO) gave a new mission of transfer learning: the ability of a system to recognize and apply knowledge and skills learned in previous tasks to novel tasks. In this definition, transfer learning aims to extract the knowledge from one or more source tasks and applies the knowledge to a target task.

The key motivation is the fact that most models which solve complex problems need a whole lot of data, and getting vast amounts of labeled data for supervised models can be really difficult, considering the time and effort it takes to label data points. Traditional learning is isolated and occurs purely based on specific tasks, datasets and training separate isolated models on them. No knowledge is retained which can be transferred from one model to another. Inspired by human beings' capabilities to transfer knowledge across domains, transfer learning aims to leverage knowledge from a related domain (called source domain) to improve the learning performance or minimize the number of labeled examples required in a target domain.

Before offering the definition of transfer learning, it is necessary to define the notion of a domain and a task. A domain D is composed of two parts: a feature space X and a marginal distribution P(X). In other words, $D = \{X, P(X)\}$. And the symbol X denotes an instance set, which is defined as $X = \{x | x_i \in X, i = 1, \cdots, n\}$. The marginal distribution P(X) is generally an invisible component, and it is hard to obtain its explicit formulation. Given a specific domain, $D = \{X , P(X)\}$, a task consists of two components: a label space Y and an objective predictive function f (denoted by $T = \{Y, f(\cdot)\}$), which is not observed but can be learned from the training data, which consist of pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in Y$. The function $f(\cdot)$ can be used to predict the corresponding label, f(x), of a new instance x. From a probabilistic viewpoint, f(x) can be written as P(y|x), which represents the predicted conditional distributions of instances. Given a source domain $D_S$ and learning task $T_S$, a target domain $D_T$ and learning task $T_T$, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $D_T$ using the knowledge in $D_S$ and $T_S$, where $D_S \neq D_T$, or $T_S \neq T_T$. Since $D = \{X , P(X)\}$, the condition $D_S \neq D_T$ implies that either $X_S \neq X_T$ or $P_S(X) \neq P_T(X)$. Similarly, since $T = \{Y, P(Y|X)\}$, the condition $T_S \neq T_T$ implies that either $Y_S \neq Y_T$ or $P(Y_S|X_S) \neq P(Y_T|X_T)$.




Following a space-setting-based categorization, transfer learning can be divided into two categories, i.e., homogeneous and heterogeneous transfer learning. Homogeneous transfer learning approaches are developed and proposed for handling the situations where the domains are of the same feature space. In homogeneous transfer learning, some studies assume that domains differ only in marginal distributions. Therefore, they adapt the domains by correcting the sample selection bias or covariate shift. However, this assumption does not hold in many cases and so some studies further adapt the conditional distributions. Heterogeneous transfer learning refers to the knowledge transfer process in the situations where the domains have different feature spaces. In addition to distribution adaptation, heterogeneous transfer learning requires feature space adaptation.

Following a label-setting-based categorization, transfer learning can be:

- Inductive, where labeled data are available in the target domain while labeled or no-labeled data are available in the source domain. In this setting, source and target tasks are different.
- Transductive, where labeled data are available only in the source domain. In this setting, source and target domains are different.
- Unsupervised, where no labeled data are available. In this setting, source and target tasks are different but related.

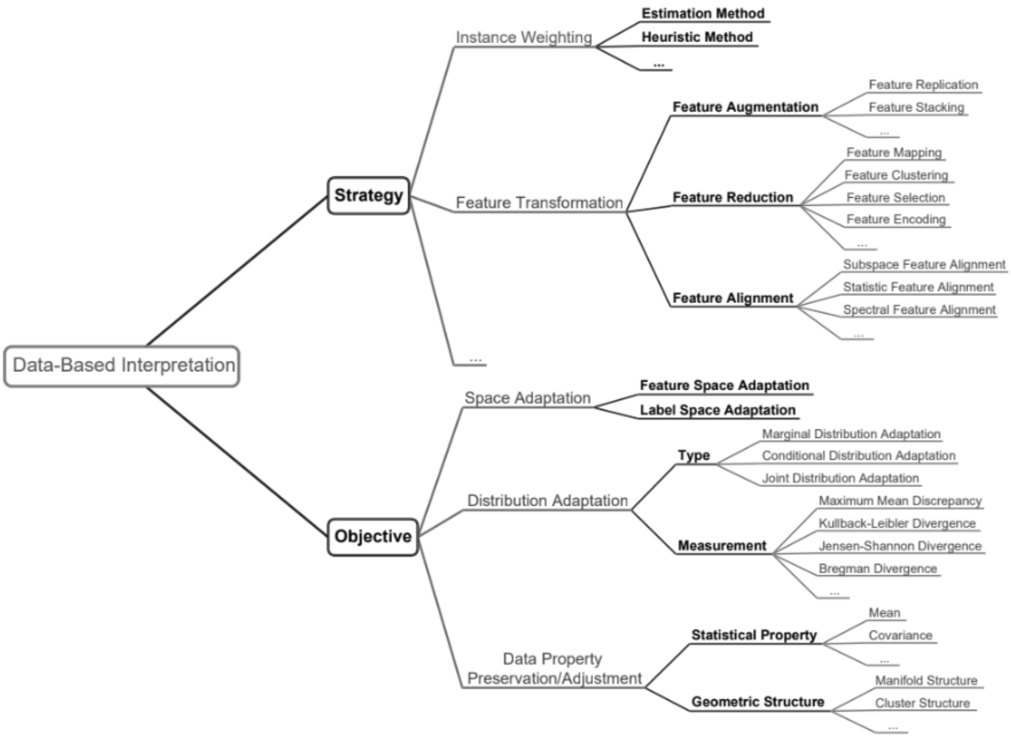
In transfer learning, there are three main research issues: (1) What to transfer; (2) How to transfer; (3) When to transfer. "What to transfer" asks which part of knowledge can be transferred across domains or tasks. Some knowledge is specific for individual domains or tasks, and some knowledge may be common between different domains such that they may help improve performance for the target domain or task. After discovering which knowledge can be transferred, learning algorithms need to be developed to transfer the knowledge, which corresponds to the "how to transfer" issue. "When to transfer" asks in which situations, transferring skills should be done. In some situations, when the source domain and target domain are not related to each other, brute-force transfer may be unsuccessful. In the worst case, it may even hurt the performance of learning in the target domain, a situation which is often referred to as negative transfer. Based on "what to transfer" there are four main transfer learning approaches:

- Instance transfer, where labeled data in the source domain are re-weighted for use in the target domain.
- Feature representation transfer, with the goal of finding a good feature representation that reduces the difference between the source and the target domain. Feature-based

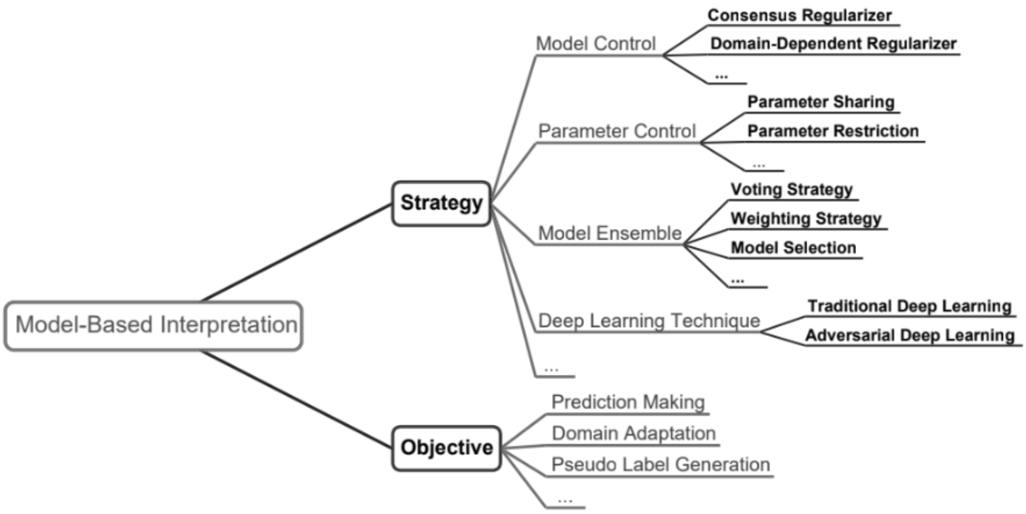
approaches can be asymmetric or symmetric. Asymmetric approaches transform the source features to match the target ones, while symmetric approaches attempt to find a common latent feature space and then transform both the source and the target features into a new feature representation.

- Parameter transfer, with the goal of discovering shared parameters or priors between the source and target domain models.
- Relational knowledge transfer, where a mapping of relational knowledge between the source and the target domain is built. This approach attempts to handle non-IID data, such as data that is not independent and identically distributed.

Transfer learning approaches are interpreted from the data and the model perspectives according to [1]. Data-based approaches focus on transferring the knowledge via the adjustment and the transformation of data, while model-based approaches focus more on the model.




Before listing the related studies that have been conducted in each category, it is important to give some basic definitions of the strategies, starting with the feature transformation ones. In Feature Replication usually a padding with extra zeroes takes place in order to make the dimensions from the two domains the same. Feature Stacking is one of the ensemble learning methods that fuses multiple weak classifiers to get a better performance than any single one. In Feature Mapping instances are mapped to a lower-dimensional feature space. Feature Clustering aims to find a more abstract feature representation of the original features. Feature Selection aims to extract the pivot features, which are the ones that behave in the same way in different domains. Feature augmentation and feature reduction mainly focus on the explicit features in a feature space. Feature Encoding makes use of autoencoders to produce the new feature representation. Autoencoders basically consist of an encoder, which produces a more abstract representation of the input, and a decoder, which maps back the representation and minimizes the reconstruction error. While feature augmentation and feature reduction mainly focus on the explicit features in a feature space, feature alignment also focuses on some implicit features such as the statistic features, the spectral features, and the subspace features. Subspace Feature Alignment makes use of a set of intermediate subspaces to model the shift between the two domains. Statistical Feature Alignment is based on the alignment of statistic features such as the covariance matrices. Spectral Feature Alignment utilizes the spectrum, i.e. eigenvalues, of the similarity matrix to reduce the dimension of the features before clustering.




Model Control Strategies are about directly adding the model-level regularizers to the learner's objective function. In this way, the knowledge contained in the pre-obtained source models can be transferred into the target model during the training process. Parameter Sharing approaches share the parameters of the source learner to the target learner. Some are network-based, while others are matrix-factorization-based. In contrast, Parameter Restriction only requires the parameters of the source and target models to be similar. Model Ensemble strategies combine a number of weak classifiers to make the final predictions. Finally, Deep Learning Techniques make use of artificial neural networks for representation learning.

## 1.2 Transfer Learning Studies

In order to review previous works on transfer learning, the categorizations of the previous section are utilized. Firstly, related works are viewed in reference to the label-based settings and secondly, data-based and model-based interpretations are examined in depth.

In the inductive transfer learning setting, all four types of approaches have been tried in terms of "what to transfer". In the transductive transfer learning setting, only instance and feature representation transfer approaches have been tried. Finally, in the unsupervised transfer learning setting, only feature representation approaches have been tried.

### *Instance Weighting*

Instance-weighting has been used in many approaches, both estimation and heuristic. Huang et al. [2] proposed Kernel Mean Matching (KMM) which matches the means between the source and the target domain instances in a Reproducing Kernel Hilbert Space (RKHS). Sugiyama et al. [3] proposed the Kullback-Leibner Importance Estimation Procedure (KLIEP), which depends on the minimization of the Kullback-Leibner divergence and incorporates a built-in model selection procedure. Sun et al. [4] proposed the 2-Stage Weighting Framework for Multi-Source Domain Adaptation (2S-MDA). Dai et al. [5] proposed TrAdaBoost. In addition to the estimation approaches mentioned above, heuristic approaches have also been tried e.g. the work of Jiang and Zhai [6]. Their target classifier is constructed with the use of labeled target domain instances, unlabeled target domain instances and labeled source domain instances. After the training, source domain instances are ranked and finally they are heuristically weighted, the weights of the top-k instances with wrong predictions are set to 0 and the other weights to 1.

### *Feature Transformation*

The feature transformation strategy has been tested for both supervised and unsupervised feature construction. Diving deeper into the feature transformation categories mentioned in the previous section, Daum et al. [7] proposed the Feature Augmentation Method (FAM) which transforms the original features by simple feature replication. The new feature representation consists of general features, source-specific features and target-specific features. They later extended (2010) FAM to also make use of unlabeled instances. Another feature augmentation approach was that of Li et al. [8] [9]. They proposed the Heterogeneous Feature Augmentation (HFA), which maps the original features into a common feature space and then performs a feature stacking operation.

### *Feature Mapping*

In the category of feature mapping, Pan et al. [10] proposed the Transfer Component Analysis (TCA). TCA adopts MMD to measure the marginal distribution difference and enforces the scatter matrix as the constraint. It just needs to learn a linear mapping from an empirical kernel feature space to a low-dimensional feature space. Long et al. [11] proposed the Joint Distribution Adaptation (JDA). JDA attempts to find a transformation matrix that maps the instances to a low-dimensional space where both the marginal and the conditional distribution differences are minimized. To realize it, the MMD metric and the pseudo-label strategy are adopted. Wang et al. [12] proposed the Balanced Distribution Adaptation (BDA), an extension of JDA. BDA attempts to balance the marginal and the conditional distribution adaptation. The authors also proposed the Weighted BDA (WBDA), where the conditional distribution difference is measured by a weighted version of MMD to solve the class imbalance problem. Long et al. [13] also proposed the Adaptation Regularization Based Transfer Learning (ARTL). The goals of ARTL are to learn the adaptive classifier, to minimize the structural risk, to jointly reduce the marginal and the conditional distribution difference, and to maximize the manifold consistency between the data structure and the predictive structure. Duan et al. [14] proposed the Domain Transfer Multiple Kernel Learning (DTMKL), where the kernel function is assumed to be a linear combination of a group of base kernels. DTMKL aims to minimize the distribution difference, the classification error etc. simultaneously.

### *Feature Clustering*

As far as feature clustering is concerned, Dai et al. [15] proposed the Co-Clustering Based Classification (CoCC) which is used for document classification. In CoCC both the source and the target document-to-word matrices are co-clustered. The source document-to-word matrix is co-clustered to generate the word clusters based on the known label information and these word clusters are used as constraints during the co-clustering process of the target domain data. Dai et al. [16] also proposed an unsupervised clustering approach, the Self-Taught Clustering (STC), which aims to simultaneously co-cluster the source domain and target domain instances with the assumption that these two domains share the same feature clusters in their common feature space. Different from the above-mentioned co-clustering-based ones, some approaches extract the original features into concepts (or topics). In the document classification problem, the concepts represent the high-level abstractness of the words (e.g. word clusters). LSA [17] aims at mapping the document-to-word matrix to a low-dimensional space. It attempts to find the true meanings of the words. To realize this, the SVD technique is used. PLSA [18], based on a statistical view of LSA, assumes that there is a latent class variable which reflects the concept, associating the document and the word. PLSA constructs a Bayesian network and the parameters are estimated by using the Expectation-Maximization (EM) algorithms. Dual-PLSA [19] assumes there are two latent variables associating the documents and the words. TPLSA (Topic-Bridged PLSA), by Xue et al. [20], assumes that the source domain and target domain instances share the same mixing concepts of the words. Collaborative Dual PLSA (CD-PLSA), by Zhuang et al. [21] [22], is proposed for multi-domain text classification. The authors adopt the EM algorithm to find the parameters. Zhuang et al. [23] also proposed the Homogeneous-Identical-Distinct-Concept-Model (HIDC). HIDC is composed of three generative models i.e. the models of the identical-concept, which is the directly transferable one, the homogeneous-concept, which is the domain-specific one that may have different effects at different domains, and the distinct-concept, which is the non-transferable domain specific one.

### *Feature Selection*

In the category of feature selection, Blitzer et al. [24] proposed the Structural Correspondence Learning (SCL). The obtained pivot features are utilized to find a low-dimensional common latent feature space and a new feature representation is constructed by stacking the original features with the obtained low-dimensional features.

### *Feature Encoding*

In reference to feature encoding, Glorot et al. [25] proposed the Stacked Denoising Autoencoder (SDA). The denoising autoencoder contains a randomly corrupting mechanism that adds noise to the input before mapping. Chen et al. [26] [27] proposed the Marginalized Stacked Linear Denoising Autoencoder (mSLDA), which adopts linear autoencoders, and marginalizes the randomly corrupting step in a closed form. That way the training time is shortened.

### *Feature Alignment*

In the category of feature alignment, Fernando et al. [28] proposed the Subspace Alignment (SA), where the subspaces are generated by performing PCA. Sun and Saenko [29] proposed the Sunspace Distribution Alignment between Two Subspaces (SDA-TS), that aligns both the subspaces and the the distributions. Gong et al. [30] proposed the Geodesic Flow Kernel (GFK), closely related to GFS (Geodesic Flow Subspaces) [31]. GFK integrates an infinite number of subspaces located on the geodesic curve from the source to the target subspace, in contrast with GFS where a finite number of interpolated subspaces is generated between the two subspaces. Sun et al. [32] proposed the Co-Relation Alignment (CORAL), which constructs the transformation features by aligning the second-order statistic features i.e. the covariance matrices. Pan et al. [33] proposed the Spectral Feature Alignment (SFA) for sentiment classifications, while Ling et al. [34] proposed the Cross-Domain Spectral Classifier (CDSC).

### *Model Control*

Under the model control strategy, Duan et al. [35] proposed the Domain Adaptation Machine (DAM), which is designed for multi-source transfer learning. Luo et al. [36] [37] proposed the Consensus Regularization Framework (CRF) for multi-source transfer learning with no labeled target domain instances. The classifiers of each source domain need to reach mutual consensuses on the domain target. This framework is implemented based on logistic regression. In light of the manifold assumption and the graph-based regularizer, Fast-DAM, proposed by Duan et al. [38], designs a domain-dependent regularizer. Univer-DAM's [39] objective function contains an additional regularizer. This regularizer usually utilizes an additional dataset termed Universum where the instances do not belong to either the positive or the negative class. The source domain instances are treated by the authors as the Universum for the target domain.

### *Parameter Sharing*

In the category of parameter sharing, Zhuang et al. [40] proposed the Matrix Tri-Factorization Based Classification Framework (MTrick) for text classification. MTrick attempts to find the connections between the document classes and the concepts conveyed by the word clusters through matrix tri-factorization. The main idea is to decompose a document-to-word matrix into three matrices: document-to-word, connection, cluster-to-word. MTrick assumes that the domains share similar concepts behind their word clusters. An extension of MTrick, TriTL [41] assumes that the concepts of the domains can be divided into three types: domain-independent, transferable domain-specific and non-transferable domain-specific. Wang et al. proposed a transfer learning framework for image classification [42], as well as an approach that integrates two matrix tri-factorizations into a joint framework [43]. Do et al. [44] utilized matrix tri-factorization to discover both the implicit and the explicit similarities for cross domain recommendation.

### *Parameter Restriction*

In the category of parameter restriction, Tommasi et al. [45] proposed the Single-Model Knowledge Transfer (SMKL), which is based on Least-Squares SVM (LS-SVM) for category-learning. SMKL selects one of the pre-obtained binary decision functions and transfers the knowledge contained in its parameters. The authors further extended SMKL by utilizing all the pre-obtained decision functions in an approach termed Multi-Model Knowledge Transfer (MMKL) [46].

### *Model Ensemble*

Under the model ensemble strategy, TrAdaboost, which was previously mentioned in the instance weighting strategy, ensembles weak classifiers via voting and weighting. TaskTrAdaBoost [47] is an extension of TrAdaBoost for handling multi-source scenarios. The difference between the original AdaBoost and the second stage of TaskTrAdaBoost is that in each iteration the former constructs a new candidate classifier on the weighted target domain instances, while the latter selects one pre-obtained candidate classifier which has the minimal classification error on the weighted target domain instances. Gao et al. [48] proposed the Locally Weighted ensemble (LWE), which focuses on the ensemble process of various learners, that could be constructed on different source domains or be built by performing different learning algorithms on a single source domain. In LWE a learner is usually assigned with different weights when classifying different target domain instances. The weights are estimated using a graph approach based on the manifold assumption (if two instances are close to each other in a high-density region, they often have similar labels). Zhuang et al. [49] proposed the Ensemble Framework of Anchor Adapters (ENCHOR), which constructs a group of weak learners via using different representations of the instances produced by anchors (specific instances). The higher the similarity between a certain instance and an anchor, the more likely the feature of that instance remains unchanged relative to the anchor.

### *Deep Learning Techniques*

As far as deep learning techniques are concerned, both traditional and adversarial approaches have been tested. SDA and mSDA, which were already mentioned above, make use of Deep Learning techniques. Zhuang et al. [50] [51] proposed Transfer Learning with Deep Autoencoders (TLDA), where two autoencoders are adopted for source domain and target domain respectively. These two autoencoders share the same parameters. The encoder and the decoder both have two layers with activation functions. Long et al. [52] [53] performed multi-layer adaptation and utilized a multi-kernel technique proposing Deep Adaptations Networks (DAN). DAN is initialized by a pre-obtained AlexNet, and was later extended to include an additional regularizer and to be generalized both in terms of architecture (e.g. applied to GoogleNet, ResNet) and in terms of measurement. Long et al. [54] also constructed residual transfer networks for domain adaptation. Besides, they proposed the Joint Adaptation Network (JAN), which adapts the joint distribution difference of multiple layers. Sun and Saenko [55] extended CORAL for deep domain adaptation and proposed an approach termed Deep CORAL (DCORAL), in which the CORAL loss is added to minimize the feature covariance. Pan et al. [56] constructed three prototypical networks, corresponding to $D_S$, $D_T$, and $D_S \cup D_T$, and incorporated the thought of multi-model consensus. They also adopted the pseudo-label strategy and adapted both the instance-level and the class-level discrepancy. Kang et al. [57] proposed the Contrastive Adaptation Network (CAN), which is based on the discrepancy metric termed contrastive domain discrepancy. Zhu et al. [58] aimed to adapt the extracted multiple feature representations, and proposed the Multi-Representation Adaptation Network (MRAN). Zhu et al. [59] also proposed the Multiple Feature Spaces Adaptation Network (MFSAN).

In the category of adversarial deep learning, the original GAN [60] is composed of two models, a generator G and a discriminator D. G produces the counterfeits of the true data for the purpose of confusing the D and making him produce wrong detection. D is fed with the mixture of the true data and the counterfeits, and it aims to detect whether a data is the true one or the fake one. These two models actually play a two-player minimax game. Ganin et al. [61] [62] proposed the Domain-Adversarial Neural Network (DANN) for domain adaptation. DANN assumes that there is no labeled target domain instance to work with. Its architecture consists of a feature extractor, a label predictor, and a domain classifier. The feature extractor is like the generator aiming to produce the domain-independent feature representation for confusing the domain classifier. The domain classifier plays the role of the discriminator attempting to detect whether the extracted features come from the source or the target domain. The label predictor produces the label prediction of the instances, which is trained on the extracted features of the labeled source domain instances. Long et al. [63] proposed the Conditional Domain Adversarial Network (CDAN), which utilizes a conditional domain discriminator to assist adversarial adaptation. Zhang et al. [64] proposed the Importance Weighted Adversarial Nets-Based Domain Adaptation (IWANDA) for partial domain adaptation. Partial transfer learning refers to the scenario when the target domain classes are less than the source domain classes. In this case source domain instances with different labels may have different importance for domain adaptation. IWANDA uses two domain-specific feature extractors for the source and the target domain respectively, two domain classifiers and one label predictor. The first classifier is optimized to minimize the domain classification error. The second classifier plays a minmax game with the target feature extractor. This classifier aims to detect whether an instance is the instance from the target domain or the weighted instance from the source domain, and to reduce the uncertainty of the label prediction. The target feature extractor aims to confuse the second classifier. The work by Cao et al. [65] also focuses on partial transfer learning by constructing a selective adversarial network. Last but not least, Chen et al. [66] investigated the transferability and the discriminability in the adversarial domain adaptation and proposed a spectral penalization approach to boost the existing adversarial transfer learning methods.

*Relational Approaches*

It is also important to underline some relational approaches. Mihalkova et al. [67] proposed the algorithm TAMAR that transfers relational knowledge with Markov Logic Networks (MLNs) across relational domains. Davis et al. [68] also proposed an approach based on a form of second-order Markov logic.

## 1.3 Transfer Learning Applications

Transfer learning has been used in various applications. There have been text-related applications, such as text classification and sentiment classification (CoCC, TPLSA, CD-PLSA, HIDC, MTrick, TriTL, SDA, SPA etc), and image-related applications (DAN, DCORAL, DANN).

Transfer learning has also been utilized for biomedical imaging analysis. For example, Shin et al. [69] fine-tuned the pre-trained deep learning neural network to help diagnose. Similarly, Maqsood et al. [70] fine-tuned the AlexNet for the detection of Alzheimer. Byra et al. [71] utilized the transfer learning technology to help assess knee osteoarthritis. In addition to imaging analysis, transfer learning has other applications in the medical area. For example, the work by Tang et al. [72] combines the active learning and the domain adaptation technologies for the classification of various medical data. Zeng et al. [73] utilized transfer learning for automatically encoding ICD-9 codes that are used to describe a patient’s diagnosis.

Transfer learning has also been tested in the field of bioinformatics. For example, Mei et al. [74] proposed a gene-ontology-based model for the predictions of protein subcellular localization, where the gene-ontology terms are transferred, and a kernel-based classifier is trained. Another work by Mei [75] proposes an approach to predict the subcellular multi-locations of plant proteins via knowledge transfer. Besides, transfer learning can also be utilized to construct models for predicting the associations between genes and phenotypes [76].

Communication applications, such as Wi-Fi localization tasks and wireless-network applications have been widely tested. Apart from that, Bastug et al. [77] proposed a caching mechanism. The knowledge contained in contextual information, which is extracted from the interactions between devices, is transferred to the target domain. Li et al. [78] proposed an energy saving scheme for cellular radio access networks. Zhao and Grace [79] applied transfer learning to topology management for reducing energy consumption.

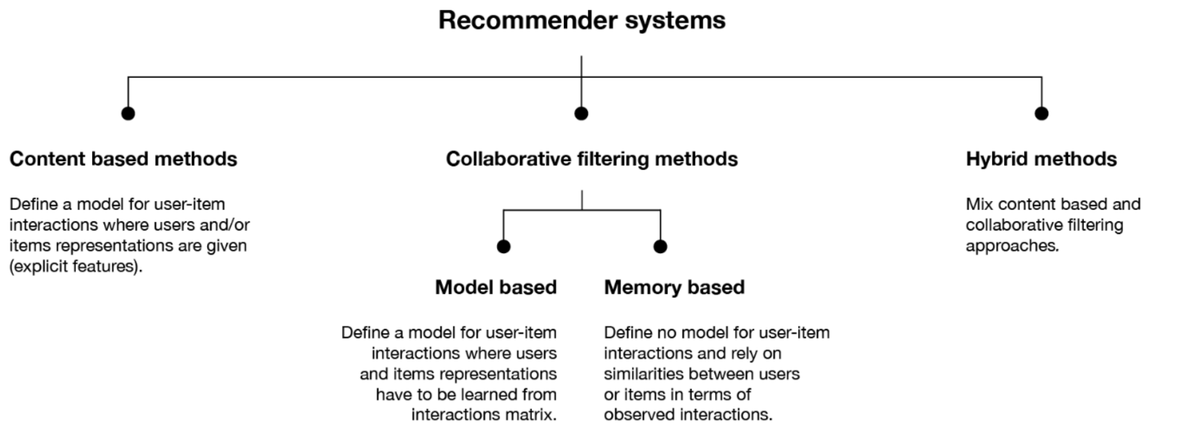
Transfer learning has also been tested in the field of transportation for anomaly detection, traffic sign recognition, vehicle classification etc. For example, Gopalakrishnan et al. [80] made use of a pre-trained deep neural network to construct a model for automatic detection of pavement cracks.

Other transfer learning applications include hand gesture recognition, face recognition, activity recognition, and speech emotion recognition. Arnold et al. [81] focused on name-entity recognition problems. Zhuo et al. [82] studied how to transfer domain knowledge to learn relational action models across domains in automated planning. Raykar et al. [83] proposed an algorithm for training classifiers for computer aided design (CAD). Kuhlmann and Stone [84] proposed a graph-based method for identifying previously encountered games and applied this technique to automate domain mapping for value function transfer and speed up reinforcement learning on variants of previously played games. Transfer Learning has also been tested for Knowledge Graph completion in a study of Piao and Breslin [85].

# Chapter 2: Recommendation Systems

This chapter starts with the basic definitions and categorizations concerning Recommendations Systems and continues with the use of Transfer Learning for Recommendation Systems. To be more specific, there is an overview of the building blocks of the studies concerning Transfer Learning for Recommendation Systems, a categorization of the cross-domain recommendation techniques, and a discussion about the challenges and open issues.

## 2.1 Recommendation Systems Definition & Categorizations




Summary of the different types of recommender systems algorithms.

A recommender system, or a recommendation system, is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems are utilized in a variety of areas and are most commonly recognized as playlist generators for video and music services like Netflix, YouTube and Spotify, product recommenders for services such as Amazon, or content recommenders for social media platforms such as Facebook and Twitter. These systems can operate using a single input, like music, or multiple inputs within and across platforms like news, books, and search queries. Recommender systems usually make use of either or both collaborative filtering and content-based filtering (also known as the personality-based approach), as well as other systems such as knowledge-based systems.

Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past. The system generates recommendations using only information about rating profiles for different users or items. Collaborative filtering methods are classified as memory-based and model-based. A memory-based approach conducts certain forms of nearest neighbour search in order to predict the rating. On the other hand, a model-based approach uses the observed user-item ratings to train a compact model that explains the hidden pattern of the given data.

In memory based collaborative methods, no latent model is assumed. The algorithms directly work with the user-item interactions: for example, users are represented by their interactions with items and a nearest neighbours search on these representations is used to produce suggestions. There are two types of approaches, user-user and item-item. The user-user method roughly tries to identify users with the most similar “interaction profile” (nearest neighbours) in order to suggest items that are the most popular among these neighbours. This method is said to be “user-centred” as it represents users based on their interactions with items and evaluate distances between users. On the other hand, the idea of item-item method is to find items similar to the ones the user already “positively” interacted with. Two items are considered to be similar if most of the users that have interacted with both of them did it in a similar way. This method is said to be “item-centred” as it represents items based on interactions users had with them and evaluate distances between those items. The user-user method is based on the search of similar users in terms of interactions with items. Conversely, the item-item method is based on the search of similar items in terms of user-item interactions. Thus, this approach is less personalized than the user-user approach but more robust.

In model based collaborative methods, some latent interaction model is assumed. Model based collaborative approaches only rely on user-item interactions information and assume a latent model supposed to explain these interactions. New suggestions can then be done based on this model. For example, matrix factorization algorithms consist in decomposing the huge and sparse user-item interaction matrix into a product of two smaller and dense matrices: a user-factor matrix (containing user representations) that multiplies a factor-item matrix (containing item representations). The main assumption behind matrix factorization is that there exists a pretty low dimensional latent space of features in which we can represent both users and items and such that the interaction between a user and an item can be obtained by computing the dot product of corresponding dense vectors in that space. The consequence of such factorization is that close users in terms of preferences as well as close items in terms of characteristics end up having close representations in the latent space.

Content-based filtering methods are based on a description of the item and a profile of the user's preferences. These methods are best suited to situations where there is known data on an item (name, location, description, etc.), but not on the user. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on product features. In this system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past or is examining in the present. Thus, in content-based methods some latent interaction model is also assumed. However, here, the model is provided with content that defines the representation of users and/or items.

If the classification (or regression) is based on user features, we say the approach is item-centred: modelling, optimizations and computations can be done “by item”. This model leads, in general, to pretty robust models as a lot of users have interacted with the item. However, it can be considered as being less personalised (more biased) than the user-centred method thereafter. If the classification (or regression) is based on item features, the method is then user-centred: modelling, optimizations and computations can be done “by user”. The model obtained is more personalised than its item-centred counterpart as it only takes into account interactions from the considered user. However, most of the time a user has interacted with relatively few items and, so, the model we obtain is a far less robust than an item-centred one.

Finally, let's mention that content based methods can also be neither user nor item centred: information about both user and item can be used for the models, for example by stacking the two features vectors and making them go through a neural network architecture.

Collaborative filtering approaches often suffer from cold start, scalability, and sparsity problems. At the same time, in content-based filtering the system is limited to recommending content of the same type as the user is already using, and so the value from the recommendation system is significantly less than when other content types from other services can be recommended. This is why most recommender systems now use a hybrid approach, combining collaborative filtering, content-based filtering, and other approaches. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem, as well as the knowledge engineering bottleneck in knowledge-based approaches.

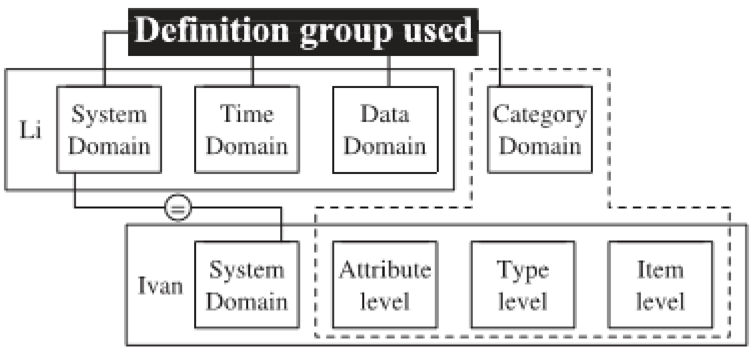
## 2.2 Transfer Learning In Recommender Systems

### 2.2.1 Summary Of Literature

The study of Khan & Ibrahim [86] did a particularly good job at summarizing the literature in Cross Domain Recommendation Systems. The objective of this study was to identify the widely used CDRS building-block definition and to classify and visualize current CDRS research in the frame of identified building-block definitions. The three building blocks that every primary study tries to identify itself with are domain difference, user-item overlap scenario, and recommendation tasks.

*Domain Difference*

As far as domains are concerned, Khan & Ibrahim concluded in the categorization below, that makes use of previous definitions by Li et al. [87] and Cantador et al. [88].



System domain refers to situations when data in the target recommender system rating matrix (e.g. MovieLens) are sparse as compared to some related recommender system (e.g., Netflix), and therefore each recommender system is considered as a distinct domain. In recommender systems, users' interactions with items can be stored in the form of numeric ratings (1–5) or binary ratings, for example, item likes or dislikes. These multi-dimensional data for similar interaction is considered to be a different data domain. A time domain is formed when a rating matrix having timestamps is split into different time slices and each slice is then considered as a separate temporal domain. Finally, recommender systems items can be grouped with respect to item attributes or types as far as they reside inside a single system domain. Hence, each attribute or type can be described as a different category. When knowledge is transferred between different categories for recommendation generation, it is considered as category domain transfer. For example, Hu et al. [89] and Loni et al. [90] transferred knowledge among AMAZON items having type book, Music CD, DVD, and VHS, Tang et al. [91] and Shapira et al. [92] transferred knowledge among Facebook items having type Music, Movie, TV, and books, while Berkovsky et al. [93] and Nakatsuji et al. [94] transferred knowledge between EachMovies items having different genres.

### *User-Item Overlap*

The most prominent study related to user-item overlap was conducted by Cremonesi et al. [95] where they identified four user-item overlap scenarios. In the first scenario, no user and no item are found to be common between participating domains (NU-NI). Ratings of both domains are analyzed to identify similarity of users and items. In the second scenario, users are found to be common between participating domains hence assisting in recommendation generation (U-NI). In the third scenario, items are found to be common between participating domains; hence assisting in recommendation generation (NU-I). Finally, in the fourth scenario, users and items are found to be common between participating domains; hence both assist in recommendation generation (U-I). For example, Gao et al. [96] and Li et al. [97] transferred knowledge from EachMovies to the MovieLens dataset based on similar items; Berkovsky et al. [93] and Zhang et al. [98] transferred knowledge between different movie genres among the same users; Pan et al. [99] [100] transferred knowledge from binary interactions to numerical ratings between same users and items; while Huang et al. [101] and Xin et al. [102] transferred knowledge between rating matrices having different timestamps between same users and items.

### *Recommendation Tasks*

As far as recommendation tasks are concerned, there are two main factors involved: the scope of recommended items and the scope of target users. Recommended items can come from both source or target domains or from only one of the two domains. Similarly, target users can reside in either one or both of the two domains. Khan & Ibrahim proposed the following categorizations:

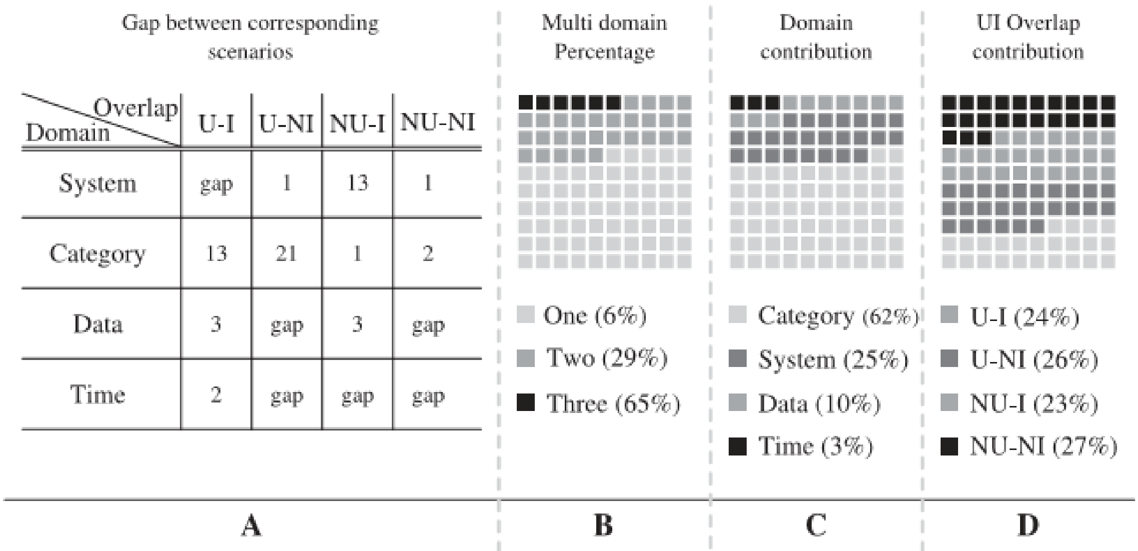
C1: Single target domain recommendation. Items in one domain are recommended to users of the respective domain based on knowledge acquired from another domain.

C2: Combined Recommendation. Items from both domains contribute to the generated recommendation that is then presented to users of either one or both of the two domains.

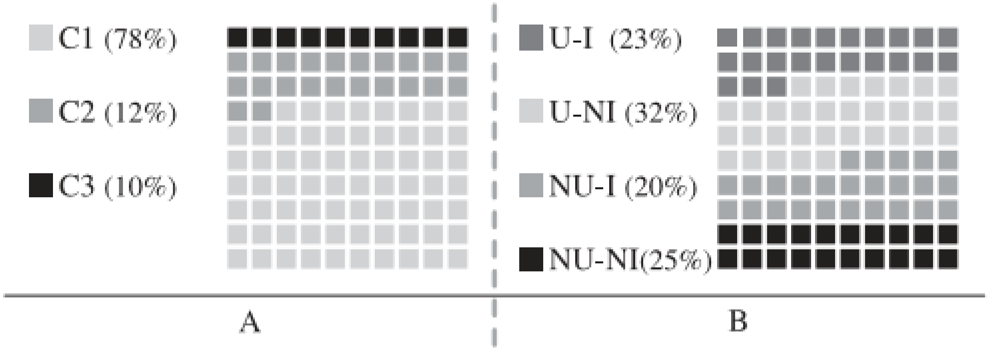
C3: Cross domain recommendation. Items residing in one domain are recommended to other domain users based on knowledge learned from users and items of both domains.

For example, Gao et al. [96] and Li et al. [97] generated recommendation for users in target domain, while Zhang et al. [98] generated recommendation for users of both domains.

Khan & Ibrahim further continued their study to find out the combinations of the above blocks that have been tested. Considering the case of a single source domain, no study was found for scenarios between the following: system domain vs. user-item overlap; data domain vs. user-no item overlap, and no user-no item overlap; time domain vs. user-no item; no user-item, and no user-no item overlap. The reason highlighted by Fernández-Tobías et al. [103] for such a trend is non availability of the appropriate dataset. Considering the case of multiple source domains, some studies were found that transferred knowledge with respect to two domains, whereas only a few transferred knowledge with respect to three domains. One reason found for the decline in studies with respect to increasing domain was related to algorithm complexity. Single domain scenarios are abundant, whereas three domain overlap scenarios are the least common. The category domain obtained a maximum contribution of 37 studies, whereas the time domain received input from only two studies. In the case of a user-item overlap comparison, all of the scenarios were found to be participating nearly as equals.




Among all recommendation tasks, the single target domain recommendation (C1) provides the majority contribution followed by combined recommendation (C2) and, finally, cross domain recommendation (C3), respectively. Based on the gathered primary studies, the research gap was found at the intersection of the cross domain recommendation (C3), User-Item overlap and combined recommendation (C2), No User-Item overlap. These gaps are due to the definitions of the categories and the lack of proper datasets. For recommendation tasks vs. user-item overlap scenarios, user-no item overlap scenario was found to have maximum contribution whereas the no user - item overlap had a minimum contribution of 20%.




### *Algorithms*

Algorithms that enable cross domain recommender systems research can be grouped into seven categories, specifically: clustering, semantics, graph-based, probability distribution, factorization, tag-based association, and others. The last one refers to studies that proposed techniques for transfer learning between participating domains, which were application specific.

One of the primitive algorithms for cluster-based cross domain recommender system research was proposed by Moreno et al. [104]. They designed a method to cluster ratings in a source domain based on users and items having similar rating patterns. This cluster was then transferred to a target domain and expanded according to similar user and items. Other studies based on clustering algorithm are those by Chen et al. [105], Wang et al. [106], Gao et al. [96], Yi et al. [107], Li et al. [97], Berkovsky et al. [93], Tang et al. [108], Li et al. [109], and Li et al. [110].

Semantic-based approaches find their root in knowledge engineering and ontology. The main idea behind semantic-based approaches is to generate a knowledge map using information available in the source domain and then transferring this knowledge map to the target domain for appropriate classification of items according to generated ratings. This approach was used by Moe and Aung [111] and Kumar et al. [112].

The graph-based approach attempts to identify the connection between users and items in the source domain to generate a connection between similar users and items in the target domain. Studies that carry out a graph based approach are those by Jiang et al. [113], Shapira et al. [92], Iwata and Takeuchi [114], Biadsy et al. [115], Guo and Chen [116], and Nakatsuji et al. [94].

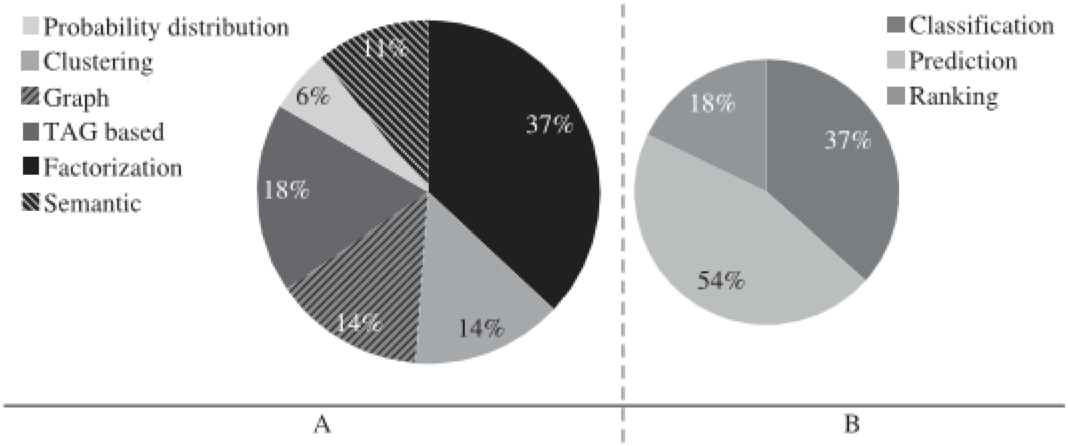
Probability distribution works for similar items identified in both domains. It attempts to learn the probability for each item with respect to all users of the source domain to find a probable recommendation score. Once learned, knowledge is transferred to the target domain for recommendation purposes. Studies utilizing probability distribution are Aizenberg et al. [117], Ren et al. [118], and Lu et al. [119].

Factorization techniques attempt to factorize a source rating matrix into a couple of feature matrices that are further combined with a target rating matrix to generate missing ratings. Papers covering factorization techniques are Shi et al. [120], Hu et al. [89], Huang et al. [101], Gao et al. [96], Xin et al. [102], Zhao et al. [121], Loni et al. [90], Shi et al. [122], Pan et al. [123], Pan and Yang [124], Shi et al. [125], Jing et al. [126], and Pan et al. [99] [100].

Tags-based association refers first to TAG-based association first group source users and items with respect to their assigned TAGs. Second, on the association between source and target domain TAGs being identified, a rating pattern can be shared. Dong and Zhao [127], Yang et al. [128], Guo and Chen [129], and Moe and Aung [130] utilized TAGs-based association to transfer knowledge from source to target domain.

### *Evaluation Metrics*

Evaluation techniques are required to measure performance of proposed algorithms and to identify evaluation techniques used for CDRS research. Based on the study of Khan & Ibrahim, three groups were formed to identify evaluation metrics, that is, classification metrics, prediction metrics,and ranking metrics. Classification metrics are used to measure an algorithm's ability to identify true positives, true negatives, false positives, and false negatives with respect to an external judgment. Prediction metrics are similar to classification metrics and are usually used for algorithms that tend to improve with each iteration. They find the amount of error between the algorithms generated values and the actual values. Ranking metrics are usually used for measuring the degree of similarity between two ranked lists of items. MAE and RMSE are the most widely used metrics for evaluating collaborative filtering approaches, but ranking-based metrics such as Precision, Success, Mean Reciprocal Rank, and F-measure have been used as well. The following image displays the popularity of each algorithm category and each evaluation metric category that has been used in the literature.




The majority of research concentrates on factorization, graph and clustering based approaches while in fact the most often used evaluation metrics are prediction followed by classification and ranking metrics, respectively.

### *Datasets*

In terms of used datasets, it was observed that the MovieLens dataset gained the maximum contribution from 23 studies at 22%, followed by Netflix, which was used by 11 studies. Moreover, many publicly accessible sources were used as datasets, in which 4 studies generated datasets of their own. In conclusion, 29 datasets were used in shortlisted primary studies, whereas the majority of the researchers focused on the popular datasets only.

### 2.2.2 Cross-Domain Recommendation Techniques

According to Iván Cantador and Paolo Cremonesi [88], cross-domain recommendation techniques can be categorized into two main classes: aggregating knowledge, or linking and transferring knowledge. The aggregated knowledge can be obtained at any stage of the recommendation process. In particular, it can be obtained from user preferences acquired at the user modeling stage, from intermediate user modeling data utilized at the item relevance estimation stage, or from item relevance estimations used at the recommendation generation stage. The knowledge linkage and transfer can be done explicitly – e.g., via common item attributes, semantic networks, association rules, and inter-domain user preference similarities –, implicitly by means of latent features shared by domains, or by means of rating patterns transferred between domains. Aggregating and linking techniques are discussed deeper below.

*Aggregating approach – Merging User Preferences*

One of the aggregating approaches is merging user preferences e.g., ratings, tags, transaction logs, and click-through data. In this case there is a need for having a significant amount of user preferences in multiple domains, and methods for accessing and merging the user profiles from different systems, which may have distinct types and/or representations of user preferences. The most favorable scenario for aggregation-based methods implies that different systems share user preferences of the same type and representation. This scenario was addressed by Berkovsky et al. [93] [131] with a mediation strategy for cross-domain CF. The authors considered a domain-distributed setting where a global rating matrix R is split, so that single-domain recommenders store local rating matrices $R_d$ having the same structure. In this setting, a target domain recommender imports rating matrices $R_d$ from the source domains, integrates the local and remote rating data into the unified rating matrix R, and applies CF to R.

Beyond numeric ratings and unary/binary data, other types of user preferences have also been aggregated for cross-domain recommendations. In particular, several studies have focused on aggregating user profiles composed of social tags and semantic concepts. In this context, there is no need for user or item overlap between domains, since tags and concepts are used as a common representation to merge user preferences from multiple domains. Szomszor et al. were among the first to correlate tag-based user profiles from multiple systems. They [132] presented an architecture that transforms a set of raw tags into a set of filtered tags aligned between folksonomies in different domains. Crossing social-tag based profiles from the Delicious and Flickr folksonomies, the authors showed that filtered tags increase the overlap between domains, and allows discovering prominent user interests, locations, and events. In a follow-up work, Szomszor et al. [133] extended their framework to map social tags to Wikipedia concepts, and build cross-domain user profiles composed of Wikipedia categories. An evaluation showed that new concepts of interest were learnt when expanding a user tag cloud with an external repository. Related to these works, Abel et al. [134] investigated the aggregation of a user's tag clouds from multiple systems. They evaluated a number of methods for semantic enrichment of tag overlap between domains, via tag similarities and via association rules deduced from the tagging data across systems. Aiming to analyze commonalities and differences among tag-based profiles, Abel et al. [135]

later mapped tags to WordNet categories and DBpedia concepts. They used the mapped tags to build category-based user profiles, which revealed significantly more information about the users than the profiles available in specific systems. Also in the context of tag-based user profile aggregation, Fernández-Tobías et al. [136] presented an approach that maps tags to emotional categories, under the assumption that emotions evoked by items in an entertainment domain can be represented through tags of folksonomies in which the items are annotated. Hence, emotions assigned to preferred items would be the bridge to merge user profiles across domains. Regarding the use of semantic concepts as user preferences, Loizou [137] presented an approach that builds a graph where the nodes are associated with Wikipedia concepts describing items liked by the users, and the edges encode the semantic relationships between those concepts, obtained by integrating user ratings and Wikipedia hyperlinks. Using such a graph, a Markov chain model was used to produce recommendations by assessing the probability of traversing the graph towards a particular item, using the nodes in the user's profile as starting points. A related approach was studied by Fernández-Tobías et al. [138] and by Kaminskas et al. [139]. The authors presented a knowledge-based framework of semantic networks that link concepts from different domains. These networks are weighted graphs, in which nodes with no incoming edges represent concepts belonging to the source domain, and nodes with no outgoing nodes represent concepts belonging to the target domain. The framework provides an algorithm that propagates the node weights, in order to identify target concepts that are most related to the source concepts. Implemented on top of DBpedia, the framework was evaluated for recommending music suited to places of interest, which were related through concepts from several domains and contextual dimensions of location and time.

Instead of aggregating user preferences directly, several researches have focused on directed weighted graphs that link user preferences from multiple domains. Nakatsuji et al. [94] presented an approach that builds domain-specific user graphs whose nodes are associated with users, and whose edges reflect rating-based user similarities. Domain graphs are connected via users who either rated items from several domains or shared social connections, to create a cross-domain user graph. Over this graph, a random walk algorithm retrieves items most liked by the users associated with the extracted nodes. Cremonesi et al. [95] built a graph whose nodes are associated with items and whose edges reflect rating-based item similarities. In this case, the inter-domain connections are the edges between pairs of items in different domains. The authors also proposed to enhance inter-domain edges by discovering new edges and strengthening existing ones, through strategies based on the transitive closure. Using the built multi-domain graph, several neighborhood- and latent factor-based CF techniques were evaluated. Tiroshi et al. [140] collected a dataset containing user preferences in multiple domains extracted from social network profiles. The data was merged into a bipartite user-item graph, and various statistical and graph-based features of users and items were extracted from the graph. These features were exploited by a machine learning algorithm that addressed the recommendation problem as a binary classification problem.

The last type of cross-domain recommendation based on user preference aggregation is formed by the approaches that map user preferences from multiple domains to domain-independent features, and use the mapped feature-based profiles to build machine learning models that predict a user's preferences in the target domain. Although not conducting evaluations, González et al. [141] proposed an approach for unifying single-domain user models by interoperability and coordination of several agents. In addition to user tastes and interests, the unified model is composed of the user's socio-demographic and emotional features. Focusing on user personality features, Cantador et al. [142] studied the relations that exist between personality types and user preferences in multiple entertainment domains, namely movies, TV, music, and books. They analyzed a large number of Facebook user profiles composed of both Big Five personality trait scores and explicit preferences for 16 genres in each of the above domains. As a result, the authors inferred similarities between personality-based user stereotypes in different domains. Finally, Loni et al. [90] presented an approach that encodes rating matrices from multiple domains as real-valued feature vectors. With these vectors, an algorithm based on factorization machines finds patterns between features from the source and target domains, and outputs preference estimations associated with the input vectors.

Aggregating ratings from several CF systems is the simplest method, but requires access to user profiles, and a significant rating overlap between domains, which may not be achievable in real situations. Thus, most aggregation-based methods transform user preferences from multiple domains into a common representation, independent of the domains of interest, and usable for establishing inter-domain data relations and overlaps. For this purpose, social tags and semantic concepts serve as the main types of user preferences. More recent methods focus on aggregating several sources of user preferences from multiple domains into a single graph.

### *Aggregating approach – Mediating User Modeling Data*

Another aggregating approach is mediating user modeling data e.g., user similarities and user neighborhoods. The central idea behind user model mediation is that importing any user modeling data from source recommenders may benefit a target recommender – the mediation can enrich the user models of the target recommender, and yield more accurate recommendations. The simplest scenario includes importing the user models, whereas more complex scenarios include mediating specific recommendation data. For example, in a CF system, cross-domain mediation may import the list of nearest neighbors. This is underpinned by two assumptions: (i) there is overlap of users between domains, and (ii) user similarity spans across domains, i.e., if two users are similar in a source domain, they are similar also in the target domain. This idea was leveraged in the heuristic variant of cross-domain mediation developed by Berkovsky et al. [131]. There, it was shown that importing nearest neighbors, and computing their similarity with the target domain data only, can produce more accurate recommendations than single-domain recommendations. A similar idea was formulated by Shapira et al. [92] as the k nearest neighbors (k-NN) source aggregation. They used multi-domain Facebook data to produce the set of candidate nearest neighbors, and compute their local similarity degree in the source domain. This allowed overcoming the new user problem and the lack of ratings in the target domain. Another attempt to use multi-domain Facebook data was done by Tiroshi and Kuflik [143]. They applied random walks to identify source domain-specific neighbor sets, which were used to generate recommendations in the target domain. Aggregating the lists of nearest neighbors relies on their data in the target domain only, which may be too sparse and result in noisy recommendations. Thus, one could consider importing and aggregating also the degree of their similarity in the source domain. This approach was referred to as cross-domain mediation. A content-based and a statistical variant of domain distance metrics were evaluated by Berkovsky et al. [144], producing comparable results and outperforming single-domain recommendations. The weighted k-NN aggregation was further enhanced by Shapira et al. [92]. The authors compared several weighting schemes, the performance of which was consistent across several metrics and recommendation tasks. The above scenarios of cross-domain mediation assume an overlap in the sets of users. An analogous scenario refers to a setting where items overlap between the source and target domains, which opens the opportunity for further mediation. One of them, involving only the music domain, but two systems (for tagging and for blogging) was studied by Stewart et al. [145]. The authors leveraged the tags assigned by similar users on Last.fm in order to recommend tags on Blogger.

Moving from CF to latent factor-based methods, we highlight two works compatible with the user modeling data mediation pattern. Low et al. [146] developed a hierarchical probabilistic model that combines user information across multiple domains, and facilitates personalization in domains with no prior user interactions. The model is underpinned by a global user profile based on a latent vector, and a set of domain-specific latent factors that eliminate the need for common items or features. Pan et al. [123] dealt with transferring uncertain ratings, i.e., expected rating range or distribution derived from behavioral logs, using latent features of both users and items. The uncertain ratings were transferred from the source to the target domain, and leveraged there as constraints for the matrix factorization model.

*Aggregating approach – Combining Recommendations*

Another way of aggregating knowledge is combining recommendations e.g., rating estimations and rating probability distributions. Contrarily to the mediation-based cross-domain recommendation scenarios, the predicted recommendations from the source domain may inform on their own to the target domain recommender. Hence, the central question in combining single-domain recommendation refers to the weights assigned to recommendations coming from the source domains, which reflect their importance for the target domain. These weights may be computed through various factors, such as the reliability of each recommender, distance between the domains, and so forth. The idea of combining single-domain recommendations was referred to as remote-average mediation. In the studies of Berkovsky et al. [93] [131], movie ratings were partitioned into domains according to the genres of the movies. Since movies combine elements from multiple genres, and users watch movies from various genres, the user- and item- overlap are both present. This allows computing stand-alone recommendations in the source domains, and aggregating them for the target domain. Weighted aggregation of single-domain recommendations also was studied by Givon and Lavrenko [147]. The authors focused on the book recommendation task, accomplished using two different methods. Standard CF recommendations were complemented by relevance model-based recommendations, relying on the similarity of a book and the user's model, both consisting of book contents and tags assigned to the book. The two were combined in a weighted manner, such that the relative importance of the CF recommendations increased with the number of ratings available. A relevant approach for cross-domain consensus regularization, although applied to classification problems and not to recommender systems, was proposed by Zhuang et al. [37]. The central contribution of that work is a framework for learning from multiple source domains, and reconciling discrepancies between the classifiers using the local data of the target domain. One of the advantages of the framework is that it does not require overlaps in either the user or item sets.

*Linking approach – Linking Domains by Common Knowledge*

Considering the non-aggregating techniques, the first approach is linking domains by a common knowledge, e.g., item attributes, association rules, semantic networks, and inter-domain correlations. A natural approach to address the heterogeneity of several domains is to identify correspondences between their characteristics. In general, such inter-domain correspondences may be established directly using some kind of common knowledge between domains, e.g., item attributes, semantic networks, association rules, and inter-domain preference-based similarities or correlations. A recommender system could identify potentially relevant items in the target domain by selecting those that are related to others in the source domains, and for which the user has expressed a preference in the past. Besides, inter-domain similarities and correlations can be exploited to adapt or combine knowledge transferred from different domains. One of the earliest approaches for linking domains was explored by Chung et al. [148]. Aiming to support the decision making process in recommendation, they proposed a framework for designing personalized filtering strategies. In the framework, relevant items in the target domain are selected according to the attributes they have in common with items in the source domain the user is interested in. That is, the inter-domain links are established through the overlap of item attributes, and no user or item overlap between the domains is required. Conversely in a realistic setting, items are highly heterogeneous, and often no common attributes between domains can be found.

To address this situation, we may establish more complex, likely indirect relations between items in different domains. Hence, when suitable knowledge repositories are available, concepts from several domains can be connected by the means of semantic properties, forming semantic networks that explicitly link the domains of interest. Along these lines, Loizou [137] proposed to use Wikipedia as a universal vocabulary to express and relate user preferences across multiple domains. The author presented an approach that builds a graph, the nodes of which represent concepts (Wikipedia pages) describing items liked by the users, and edges encode the semantic relationships between those concepts, obtained by integrating user ratings and Wikipedia hyperlinks. Using such a graph, a Markov chain model produces recommendations by assessing the probability of traversing the graph from the nodes in the user's profile as a starting point toward the recommendable items. A major difficulty of the above approaches is the well known knowledge acquisition problem, that is, building the above mentioned knowledge repositories. To address this problem, information has to be extracted and stored in a formal and structured representation that can be exploited by a recommender. Fernández-Tobías et al. [138] and Kaminskas et al. [139] envisioned Linked Data as a solution to the problem. Specifically, they proposed a framework for extracting a multi-domain semantic network from the DBpedia ontology, which links items and concepts in the source and target domains. Over the extracted network, a constrained spreading algorithm computes semantic similarities to rank and filter items in the target domain.

Inter-domain association rules have also been explored as an alternative to relate various types of items. In this direction, Azak [149] presented a framework for cross-domain recommendation in which knowledge-based rules defined by domain experts facilitate mapping between attributes in distinct domains, e.g., "people who like romance drama movies also like dramatic poetry books." These rules are then used to enhance CB and CF recommendations, adjusting the predicted ratings whenever rule conditions hold. Cantador et al. [142] related user personality types with domain-dependent preferences by means of automatically generated association rules. The authors also extracted personality stereotypes for sets of domain genres. Based on these stereotypes, inter-domain similarities were computed between genres, which may be used to support knowledge transfer between domains.

Instead of linking domains by mapping attributes, an alternative way to transfer knowledge is to compute similarities or correlations between domains based on user preference or item content analysis. In an early work, Berkovsky et al. [144] explored this idea aiming to identify related domains, from which user data would be imported and utilized to enrich the user model in the target domain. The proposed approach makes use of web directories to identify websites that characterize the domains of interest. Then, the approach establishes domain similarities by computing the cosine similarity between the TF-IDF term vectors of the domains' websites. We note that this method requires no overlap of users or items, but rather an external source of representative documents classified to several domains.

Another way of exploiting inter-domain similarities for cross-domain recommendation consists of integrating them into the matrix factorization method [150]. Specifically, such similarities are imposed as constraints over user or item latent factors when jointly factorizing rating matrices. For instance, Cao et al. [151] proposed an approach in which inter-domain similarities are implicitly learnt from data, as model parameters in a non-parametric Bayesian framework. Since user feedback is used to estimate the similarities, user overlap between the domains is required. Addressing the sparsity problem, Zhang et al. [98] adapted the probabilistic matrix factorization method to include a probability distribution of user latent factors that encodes inter-domain correlations. One strength of this approach is that user latent factors shared across domains are not needed, allowing more flexibility in capturing the heterogeneity of domains. Instead of automatically learning implicit correlations in the data, Shi et al. [120] argued that explicit common information is more effective, and relied on shared social tags to compute cross-domain user-to-user and item-to-item similarities. Similarly to previous approaches, rating matrices from the source and target domains are jointly factorized; but in this case user and item latent factors from each domain are restricted, so that their product is consistent with the tag-based similarities.

*Linking approach – Sharing Latent Features*

A second linking approaching is sharing latent features. Latent factor models are among the most popular CF techniques. In these models, user preferences and item attributes, which are typically very sparse, are characterized through a reduced set of latent factors discovered from data, to obtain a denser representation. The assumption is that using the new representation, latent user preferences and item attributes are better captured and matched. Related to the what to transfer aspect of transfer learning, latent factors shared between domains can be exploited to support cross-domain recommendations. Two types of approaches have been studied to address the how to transfer aspect; namely, adaptive and collective models. In the former, latent factors are learnt in the source domain, and are integrated into a recommendation model in the target domain, while in the latter, latent factors are learnt simultaneously optimizing an objective function that involves both domains.

Pan et al. [152] addressed the sparsity problem in the target domain following the adaptive approach, proposing to exploit user and item information from auxiliary domains where user feedback may be represented differently. In particular, they studied the case in which users express binary like/dislike preferences in the source domain, and utilize 1-5 ratings in the target domain. Their approach performs singular value decomposition (SVD) in each auxiliary domain, in order to separately compute user and item latent factors, which are then shared with the target domain. Specifically, transferred factors are integrated into the factorization of the rating matrix in the target domain and added as regularization terms so that specific characteristics of the target domain can be captured.

Latent factors can also be shared in a collective way, as studied by Pan et al. [153]. In this case, instead of learning latent features from the source domains and transferring them to the target domain, the authors proposed to learn the latent features simultaneously in all the domains. Both user and item factors are assumed to generate the observed ratings in every domain, and, thus, their corresponding random variables are shared between the probabilistic factorization models of each rating matrix. Moreover, the factorization method is further extended by incorporating another set of factors that capture domain-dependent information, resulting in a tri-factorization scheme. A limitation of the proposed approach is that the users and items from the source and target domains have to be identical.

Instead of focusing on sharing latent factors, Enrich et al. [154], and Fernández-Tobías and Cantador [155] studied the influence of social tags on rating prediction, as a knowledge transfer approach for cross-domain recommendations. The authors presented a number of models based on the SVD++ algorithm [156] to incorporate the effect of tag assignments into rating estimation. The underlying hypothesis is that information about item annotation in a source domain can be exploited to improve rating prediction in a target domain, as long as a set of common tags between the domains exists. In the proposed models, tag factors are added to the latent item vectors, and are combined with user latent features to compute rating estimations. The difference between these models is in the set of tags considered for rating prediction. In all the models knowledge transfer is performed through the shared tag factors in a collective way, since these are computed jointly for the source and the target domains. Hu et al. [89] presented a more complex approach that takes domain factors into account. There, the authors argue that user-item dyadic data cannot fully capture the heterogeneity of items, and that modeling domain-specific information is essential to make accurate predictions in a setting, where users typically express their preferences in a single domain. They proposed a tensor factorization algorithm to exploit the triadic user-item-domain data. In that method, rating matrices from several domains are simultaneously decomposed into shared user, item, and domain latent factors, and genetic algorithm automatically estimates optimal weights of the domains.

*Linking approach – Transferring Rating Patterns*

Finally, the linking of two domains can be done by transferring rating patterns, explicit or implicit. Rather than sharing user or item latent factors for knowledge transfer, these approaches analyze the structure of rating data at the community level. These methods are based on the hypothesis that even when their users and items are different, close domains are likely to have user preferences sampled with the same population. Therefore, latent correlations may exist between preferences of groups of users for groups of items, which are referred to as rating patterns. In this context, rating patterns can act as a bridge that relates the domains, such that knowledge transfer can be performed in either adaptive or collective manners. In the adaptive setting, rating patterns are extracted from a dense source domain. In the collective setting, data from all the domains are pulled together and jointly exploited, even though users and items do not overlap across domains.

Lee et al. [157] proposed one of the first approaches to exploit rating patterns for cross-domain recommendation. Similarly to the cross-domain mediation proposed by Berkovsky et al. [93], global nearest neighbors are identified by adding the similarity scores from each domain. Then, patterns of items commonly rated together by a set of neighbors are discovered using association rules. Finally, in the recommendation stage, rating predictions are computed with the standard user-based CF algorithm, but enhanced with the user's rules that contain the target items. Li et al. [97] proposed an adaptive method based on simultaneously co-clustering users and items in the source domain, to extract rating patterns. Clustering is performed using a tri-factorization of the source rating matrix. Then, knowledge is transferred through a codebook, a compact cluster-level matrix computed in the source domain taking the average rating of each user-item cluster. In the target domain, missing ratings are predicted using the codebook.

Moreno et al. [104] extended the codebook idea to a scenario in which various source domains contribute to the target domain proposing the model TALMUD. The approach is based on a linear combination of codebooks, where the weights are learnt by minimizing the prediction error in the target domain. Another approach concerning transfer learning for multiple domains is that of Zhuang et al. [158]. They proposed a transfer collaborative filtering method with consensus regularization termed TRACER, which in contrast to TALMUD learns and transfers knowledge simultaneously. The consensus regularization forces the predicted result to be the value of the majority of the results of the sources. In addition, the two methods have different mechanisms of integrating predicted results. TALMUD assigns the same weight to the predicted ratings learnt from the same source domain, thus there are s different weights if existing s source domains. While TRACER can be regarded as a locally majority voting method, that is, we perform majority voting on each (user, item) pair in the target domain.

Li et al. [159] further extended the codebook idea to a collective approach using a probabilistic framework. Instead of relying on an dense source domain data to build the codebook, all rating matrices are pulled together to extract the shared patterns. Furthermore, rather than having each user/item belonging to a single cluster, a probability distribution is introduced to allow users and items belong to multiple clusters, with distinct membership degrees. In the same fashion, the ratings associated with each user-item cluster are also given by a conditional probability distribution. In this way, a generative rating model is obtained, since the ratings of each domain can be recovered by drawing users and items from the shared cluster-level model, and then drawing the expected rating conditioned to the user-item cluster.

A strength of both approaches is that neither overlap of users nor of items is required. However, Cremonesi and Quadrana [160] partially disproved it, showing that the codebook does not transfer knowledge when source and target domains do not overlap. They provided an alternative explanation to the accuracy increase using a codebook that does not involve knowledge transfer between domains. Finally, Gao et al. [161] followed the idea of extracting rating patterns by co-clustering rating matrices, and addressed two limitations of previous methods. First, they argued that some domains are more related to the target domain than others, and this cannot be captured using identical rating patterns. Second, they hypothesized that performance may suffer when the domains are diverse, and do not share common rating patterns. To overcome these limitations, the authors proposed a model capable of controlling the amount of knowledge transferred from each domain. Specifically, they used a co-clustering algorithm of but split the extracted rating patterns into a shared part and a domain-specific part. Optimization is performed in a collective way, since the shared part of the rating patterns is learnt simultaneously from all the domains.

## 2.3 Challenges & Open Issues

Apart from the gaps already mentioned in the above section, the study of Khan & Ibrahim [86] also pointed out future research directions for cross domain recommender systems. Those can be arranged into five groups, namely domain similarity enhancement, algorithm improvement, using of “big data” as a source domain, conventional recommender systems problems, and dataset extensions.

Domain similarity enhancement can be achieved by analyzing heterogeneous data, by analyzing user interest drift, by including more related domains (of the same or of different primary type), or by context enhancement. User interest can change over time, therefore user interactions can be analyzed with respect to time. Moreover, Fernández-Tobías et al. [103], pointed out the importance of exploiting contextual features, such as the time (e.g. movies and music compositions usually consumed on Christmas) and the users' mood and sentiments (e.g. movies and music compositions that usually yield nostalgic feelings). Hybrid cross-domain recommendation approaches have been barely investigated. This may be due to the lack of public datasets with both content-based and collaborative filtering information about several domains. Thus, a hybrid cross-domain recommender system could exploit relations between user preferences and/or item attributes based on content, collaborative filtering, and contextual information, across multiple domains.

As far as big data compatibility is concerned, using big data as the source domain leads to three future directions. First of all, cross domain recommender systems can utilize big data services to tune target recommendation. Big data are enriched with demographic and other statistical information that can help in personalized recommendation. Secondly, cross domain recommendation can be improved by proposing distributed algorithms that can be scaled as per the requirement. Thirdly, social media contains valuable user interactions related to different items, and the majority of these interactions are publicly available on Facebook, twitter, LinkedIn, and so on. Thus, they can be a potential source of improving target recommendations.

In terms of conventional recommender system problems, risk, adaptivity, robustness, novelty and privacy are concerned. As described in Ricci et al. [162], the risk is associated with loss of customers as a result of the wrong or inappropriate recommendation. CDRS can assist in avoiding risk by utilizing the user review sentiments available in other domains, hence reducing recommendation risk. Adaptivity is related to user interest drift over time. CDRS can assist with adaptivity by transferring knowledge from the source domain within a recent time frame. Robustness is related to avoiding recommendation based on fake ratings. CDRS can assist with robustness by transferring knowledge from more than one related domain, hence reducing the probability of fake ratings. Novelty is related to the items that the user did not know about and found interesting when recommended by recommender systems. CDRS can assist by transferring source items related to items that were already rated by the user but are not exactly the same. Privacy is associated with revealing the identity of people who like similar items or are connected to target users. CDRS recommendation does not face a privacy problem for recommendation across the system domain because no system declares user similarity for other systems.

Focusing more on evaluation metrics, according to Winoto and Tang [163], we could expect that cross-domain recommendations, generated with a certain amount of user preferences, are less precise than those based on the same amount of preferences but expressed in the target domain. The advantage of cross-domain recommendation may not be improved accuracy, but added novelty and more diverse recommendations, and thus may offer higher satisfaction and utility to the user. In this context recently proposed novelty and diversity metrics could be taken into consideration.

There's also the issue of compatible datasets which refers to the fact that existing datasets were created for conventional recommender systems and sometimes researchers use them in scenarios for which they were not made. Although transformed data can run a developed algorithm, it cannot, however, be related to a real-world scenario. One way of reducing dataset incompatibility is by creating new datasets that classify their association with the appropriate domain, recommendation scenario, and recommendation tasks. Although this is possible, it may take considerable time. The other option is to use existing datasets for CDRS and standardize conditions and limits that should be considered when using existing datasets for scenarios having a specific domain, recommendation scenarios, and recommendation tasks.

Last but not least, although there have been a lot of studies on transfer learning involving various applications, there are not many available source codes online. The problem that has been mainly studied is Image Classification that conducts Transfer Learning with the use of pre-trained models such as VGG16 or InceptionV3. This problem setting is either checked on the “Dogs VS Cats” dataset or on randomly created blobs. Apart from the above-mentioned problem, there has been an implementation that deals with sentiment classification using MXNet [164]. The IMDB movie reviews dataset is used to train the base network from which learning is transferred to a Hotel reviews dataset. Finally, there is an implementation for an image-based recommendation system [165]. Basically, the model is trained on movie posters and it extracts features from a pre-trained Convolutional Neural Network (ConvNet) trained on ImageNet. It is obvious that there are a lot of problem settings uncovered and the unavailability of open source code is a factor that certainly affects the wider use of Transfer Learning in more applications.

# Chapter 3: Transfer Learning Via Aggregating Knowledge: An Experimental Study

After examining the literature on Transfer Learning and its effects in Recommendation Systems, a simple recommender is created which takes advantage of transfer learning and tests these effects in practice. The approach followed falls under the category of transfer learning via aggregating knowledge. The main idea is that users have similar rating patterns in different domains and thus user ratings in a domain B provide useful knowledge that can benefit predictions about his ratings in a domain A. This concept requires a common set of users between the two domains.

### *The Algorithm*

The recommender that was created for this approach makes use of SVD to make predictions. To be exact, the SVD algorithm used comes from the python library called Surprise, which provides the SVD algorithm as popularized by Simon Funk during the Netflix Prize.

Singular Value Decomposition (SVD) is a method from linear algebra that has been generally used as a dimensionality reduction technique in machine learning. SVD is a matrix factorization technique, which reduces the number of features of a dataset by reducing the space dimension from N-dimension to K-dimension (where K<N). In the context of the recommender system, the SVD is used as a collaborative filtering technique. It uses a matrix structure where each row represents a user, and each column represents an item. The elements of this matrix are the ratings that are given to items by users. SVD is a method of decomposing a matrix into three other matrices: $A = USV^T$, where A is a m x n utility matrix, U is a m x r orthogonal left singular matrix, which represents the relationship between users and latent factors, S is a r x r diagonal matrix, which describes the strength of each latent factor, and V is a r x n diagonal right singular matrix, which indicates the similarity between items and latent factors. SVD decreases the dimension of the utility matrix A by extracting its latent factors. It maps each user and each item into a r-dimensional latent space. This mapping facilitates a clear representation of relationships between users and items.

The prediction is set as:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u$$
,

where $p_u$ represents the user factors, $q_i$ represents the item factors, $b_u$ represents the user biases and $b_i$ represents the item biases.

If user $u$ is unknown, then the bias $b_u$ and the factors $p_u$ are assumed to be zero. The same applies for item i with $b_i$ and $q_i$.

To estimate all the unknown, the following regularized squared error is minimized:

$$\sum_{r_{ui} \in R_{train}} (r_{ui} - \hat{r}_{ui})^2 + \lambda \left( b_i^2 + b_u^2 + ||q_i||^2 + ||p_u||^2 \right)$$

The minimization is performed by a very straightforward stochastic gradient descent:

$$\begin{aligned} b_u &\leftarrow b_u + \gamma(e_{ui} - \lambda b_u) \\ b_i &\leftarrow b_i + \gamma(e_{ui} - \lambda b_i) \\ p_u &\leftarrow p_u + \gamma(e_{ui} \cdot q_i - \lambda p_u) \\ q_i &\leftarrow q_i + \gamma(e_{ui} \cdot p_u - \lambda q_i) \end{aligned}$$

where $e_{ui} = r_{ui} - \hat{r}_{ui}$.

These steps are performed over all the ratings of the trainset and repeated n_epochs times. Baselines are initialized to 0. User and item factors are randomly initialized according to a normal distribution, which can be tuned. The library also allows you to have control over the learning rate γ and the regularization term λ.

### *Tuning parameters*

To make the best use of the SVD algorithm for the target dataset, the best values for the algorithm parameters had to be chosen. For this step, the parameters to be tested were the number of epochs, the learning rate and the regularization term. To find the best values of these parameters the GridSearchCV class of the Surprise library was used. The GridSearchCV class computes accuracy metrics for an algorithm on various combinations of parameters, over a cross-validation procedure. In this experiment the average RMSE was evaluated over a 3-fold cross-validation procedure to examine which values yield the best results. The following table shows the parameter values that were tested. The values with the best RMSE score are those in bold.

| **GridSearchCV** | | | |
|---|---|---|---|
| *n_epochs* | 10 | 15 | **20** |
| *lr_all* | 0,002 | 0,005 | **0,01** |
| *reg_all* | 0,02 | **0,1** | 0,4 |

Following the results of the GridSearchCV, the model was trained for 20 epochs, with a learning rate of 0,01 and a regularization term of 0,1.

### *Transfer Learning*

As mentioned above, the approach followed belongs to the category of transfer learning via aggregating knowledge and requires user overlap between domains. In order to explain the approach let's refer to $D_A$ as the original domain, the domain to which knowledge is transferred, and $D_B$ as the auxiliary domain, the domain from which knowledge is transferred. $D_A$ has less or sparser rating data which makes the recommendation of items difficult, especially in cases of new items or users. On the other hand, $D_B$ contains more or denser rating data, and thus its users can get better item recommendations. Since the domains share some users it would be useful to harness knowledge about those users in $D_B$, where information is plenty, and use it to enhance recommendations in $D_A$. This is all based on the assumption that users share similar rating patterns in different domains. Sometimes there is even a connection between these domains that reflects on the rating patterns. For example, movies and books are connected through their genres. If a user likes romantic movies, it is likely that he will like romance novels. In the same logic, someone's preferences in digital music can reflect to his preferences in CDs.

This is what this approach intents to demonstrate with the use of Amazon datasets from the domains of Digital Music and CDs & Vinyl. More specifically the goal is to improve recommendation accuracy in the CDs & Vinyl domain by making use of ratings from the Digital Music domain. To achieve this, the first step is to find the shared users between the two domains. Then, the Digital Music ratings referring to the shared users are extracted and concatenated with the CDs & Vinyl ratings. After that, the concatenated data are passed in the SVD algorithm. More details about the data and the conducted experiments are given below.

### *Experiment Details*

Since Transfer Learning works best in sparse datasets, according to previous studies, a sample of the CDs & Vinyl dataset is selected. To be exact, from the initial 3749004 ratings from the CDs & Vinyl dataset, 187450 are used (5%), while all 836006 ratings from the Digital Music dataset are used. 45069 common users were found, that are involved in 166569 Digital Music ratings, creating a sum of 354019 ratings.

The goal is to improve recommendations for the users of the CDs & Vinyl dataset therefore the algorithm needs to be tested for ratings coming from the CDs & Vinyl dataset. To achieve that, the train-test split is not shuffled, as it is usually done. The test set consists of the last 25% of the data, which is where the CDs & Vinyl ratings are located. The other 75% of the data is used in the training process. The test set is furtherly divided into two sets. The first test set contains ratings of users that are common between the two domains, while the second test set consists of ratings that belong to users of the CDs & Vinyl domain only. The reason behind this is to retrieve deeper knowledge about how well the transferring of knowledge can work under different conditions. Test set 1 consists of 38803 ratings, while test set 2 consists of 49702 ratings.

To evaluate the model, the RMSE metric is used, along with precision and recall at k=10 recommendations. The RMSE serves to aggregate the magnitudes of the errors between predicted and true ratings, precision at k is the proportion of recommended items in the top-k set that are relevant, and recall at k is the proportion of relevant items found in the top-k recommendations. Precision and recall are binary metrics used to evaluate models with binary output. Therefore, in this case it is assumed that any true rating above 3.5 corresponds to a relevant item and any true rating below 3.5 is irrelevant. A relevant item for a specific user-item pair means that this item is a good recommendation for the user in question. The RMSE is calculated with the use of the recmetrics library, while precision and recall are calculated via a custom-made method.

### *Results*

To prove that this approach is improving recommendations of CDs & Vinyl the program is tried 5 times and the average of all the metrics is used for comparison. As mentioned above, the algorithm is trained for 20 epochs, with the learning rate of 0,01 and a regularization term of 0,1. The values of the metrics of the standalone CDs & Vinyl recommender are compared with the values of the metrics of the recommender that makes use of transfer learning. Table 1 shows the results of the standalone model, Table 2 shows the results of the transfer learning model in the full test set, Table 3 shows the results of the transfer learning model in test set 1, and Table 4 shows the results of the transfer learning model in test set 2.

| **SVD** | | | |
|---|---|---|---|
| | **RMSE** | **Precision** | **Recall** |
| 1rst Run | 1,0136332 | 0,8528492 | 0,8600689 |
| 2nd Run | 1,0251841 | 0,8507202 | 0,8586232 |
| 3rd Run | 1,0192861 | 0,8513214 | 0,8590060 |
| 4rth Run | 1,0314805 | 0,8504997 | 0,8581115 |
| 5th Run | 1,0281140 | 0,8499402 | 0,8580891 |
| | | | |
| **Average:** | 1,0235396 | 0,8510662 | 0,8587797 |

***Table 1***

| **TL (full test set)** | | | |
|---|---|---|---|
| | **RMSE** | **Precision** | **Recall** |
| 1rst Run | 0,9899211 | 0,8521497 | 0,8610529 |
| 2nd Run | 0,9897234 | 0,8521799 | 0,8611257 |
| 3rd Run | 0,9905304 | 0,8519561 | 0,8609187 |
| 4rth Run | 0,9901670 | 0,8520003 | 0,8609765 |
| 5th Run | 0,9900486 | 0,8519621 | 0,8609635 |
| | | | |
| **Average:** | 0,9900781 | 0,8520496 | 0,8610075 |

***Table 2***

| **TL (test set 1)** | | | |
|---|---|---|---|
| | **RMSE** | **Precision** | **Recall** |
| 1rst Run | 0,9039397 | 0,8406368 | 0,8595070 |
| 2nd Run | 0,9033497 | 0,8404326 | 0,8593598 |
| 3rd Run | 0,9051672 | 0,8399771 | 0,8589996 |
| 4rth Run | 0,9048292 | 0,8401185 | 0,8591469 |
| 5th Run | 0,9039774 | 0,8400941 | 0,8592086 |
| | | | |
| **Average:** | 0,9042526 | 0,8402518 | 0,8592444 |

***Table 3***

| TL (test set 2) | | | |
|---|---|---|---|
| | RMSE | Precision | Recall |
| 1rst Run | 1,0521752 | 0,8582306 | 0,8618694 |
| 2nd Run | 1,0522396 | 0,8583846 | 0,8620584 |
| 3rd Run | 1,0523723 | 0,8582831 | 0,861932 |
| 4rth Run | 1,0519903 | 0,8582761 | 0,8619429 |
| 5th Run | 1,0523634 | 0,8582306 | 0,8618904 |
| | | | |
| **Average:** | 1,0522281 | 0,8582810 | 0,8619387 |

***Table 4***

According to the results, the algorithm using transfer learning generally improves recommendations (Table 1-Table 2). The standalone recommender has an RMSE of 1,0235396, a precision of 0,8510662 and a recall of 0,8587797, while the recommender utilizing transfer learning has an RMSE of 0,9900486, a precision of 0,8520496 and a recall of 0,8610075.

Diving deeper into the test sets, for shared users the RMSE gets smaller with the use of transfer learning, recall slightly improves, but precision worsens (Table 1-Table 3). For users of solely the CDs & Vinyl domain, transfer learning slightly worsens the RMSE, while it improves precision and recall (Table 1-Table 4).

All in all, it is demonstrated that transfer learning can indeed benefit recommendations for the CDs & Vinyl domain.

# Conclusion & Future Work

In this study, a simple Transfer Learning approach is proposed based on SVD to improve recommendations and solve the data sparseness problem in collaborative filtering tasks. The experimental studies clearly demonstrate that the CDs & Vinyl recommender can benefit from exploiting ratings from the Digital Music domain in cases of sparse user data.

Future work could dive deeper into the reasons behind the different results in the different test sets and find ways to improve all the metrics (RMSE, precision, recall) both for shared and not-shared users. In addition, fine tuning the model parameters could also enhance the model efficacy. For this approach only a few values were tested so there is room for more research. Experiments could also be conducted with different sizes of samples and different data densities. This could also provide more insight about transfer learning. Finally, in terms of the transfer learning process, this is an instance-based approach that belongs to the aggregating category, since ratings from the Digital Music are added to the ratings of CDs & Vinyl. An approach following the linking strategy could add more to the model, while the selection of instances could be additionally filtered, or different weights could be used to the instances of both domains.

# References


1. A Comprehensive Survey on Transfer Learning by *Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He*, *arXiv:1911.02685v2, December 2019*
2. Correcting sample selection bias by unlabeled data by *J. Huang, A.J. Smola, A. Gretton, K.M. Borgwardt, and B. Scholkopf*, *in Proc. 20th Annual Conference on Neural Information Processing Systems, Vancouver, Dec. 2006*
3. Direct importance estimation for covariate shift adaptation by *M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. Bnau, and M. Kawanabe*, *in Ann. Inst. Stat. Math., vol. 60, no. 4, pp. 699–746, Dec.2008*
4. A twostage weighting framework for multi-source domain adaptation by *Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye*, *in Proc. 25th Annual Conference on Neural Information Processing Systems, Granada, Dec. 2011, pp. 505–513*
5. Boosting for transfer learning by *W. Dai, Q. Yang, G. Xue, and Y. Yu*, *in Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, Jun. 2010, pp. 1855–1862*
6. Instance weighting for domain adaptation in NLP by *J. Jiang and C. Zhai*, *in Proc. 45th Annual Meeting of the Association of Computational Linguistics, Prague, Jun. 2007, pp. 264–271*
7. Frustratingly easy domain adaptation by *H. Daum III*, *in Proc. 45th Annual Meeting of the Association for Computational Linguistics, Prague, Jun. 2007, pp. 256–263*
8. Learning with augmented features for heterogeneous domain adaptation by *L. Duan, D. Xu, and I.W. Tsang*, *IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 6, pp. 1134–1148, Jun. 2014*
9. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation by *W. Li, L. Duan, D. Xu, and I.W. Tsang*, *in Proc. 29th International Conference on Machine Learning, Edinburgh, Jun. 2012, pp. 1–8*
10. Domain adaptation via transfer component analysis by *S.J. Pan, I.W. Tsang, J.T. Kwok, and Q. Yang*, *IEEE Trans. Neural Netw., vol. 22, no. 2, pp. 199–210, Feb. 2011*
11. Transfer feature learning with joint distribution adaptation by *M. Long, J. Wang, G. Ding, J. Sun, and P.S. Yu*, in Proc. *IEEE International Conference on Computer Vision, Sydney, Dec. 2013, pp. 2200–2207*
12. Balanced distribution adaptation for transfer learning by *J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen*, *in Proc. 17th IEEE International Conference on Data Mining, New Orleans, Nov. 2017, pp. 1129–1134*
13. Adaptation regularization: A general framework for transfer learning by *M. Long, J. Wang, G. Ding, S.J. Pan, and P.S. Yu*, *IEEE Trans. Knowl. Data Eng., vol. 26, no. 5, pp. 1076-1089, May 2014*
14. Domain transfer multiple kernel learning by *L. Duan, I.W. Tsang, and D. Xu*, *IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 465–479, Mar. 2012*
15. Co-clustering based classification for out-of-domain documents by *W. Dai, G.-R. Xue, Q. Yang, and Y. Yu*, *in Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, Aug. 2007, pp. 210–219*

16. Self-taught clustering by *W. Dai, Q. Yang, G. Xue, and Y. Yu, in Proc. 25th International Conference of Machine Learning, Helsinki, Jul. 2008, pp. 200–207*
17. Indexing by latent semantic analysis by *S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, J. Am. Soc. Inf. Sci., vol. 41, pp. 391–407, Sep. 1990*
18. Probabilistic latent semantic analysis by *T. Hofmann, in Proc. 15th Conference on Uncertainty in Artificial Intelligence, Stockholm, Jul. 1999, pp. 289–296*
19. Probabilistic matrix tri-factorization by *J. Yoo and S. Choi, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Apr. 2009, pp. 1553–1556*
20. Topic-bridged PLSA for cross-domain text classification by *G.-R. Xue, W. Dai, Q. Yang, and Y. Yu, in Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, Jul. 2008, pp. 627–634*
21. Collaborative Dual-PLSA: Mining distinction and commonality across multiple domains for text classification by *F. Zhuang, P. Luo, Z. Shen, Q. He, Y. Xiong, Z. Shi, and H. Xiong, in Proc. 19th ACM International Conference on Information and Knowledge Management, Toronto, Oct. 2010, pp. 359–368*
22. Mining distinction and commonality across multiple domains using generative model for text classification by *F. Zhuang, P. Luo, Z. Shen, Q. He, Y. Xiong, Z. Shi, and H. Xiong, IEEE Trans. Knowl.Data Eng., vol. 24, no. 11, pp. 2025–2039, Nov. 2012*
23. Concept learning for cross-domain text classification: A general probabilistic framework by *F. Zhuang, P. Luo, P. Yin, Q. He, and Z. Shi, in Proc. 23rd International Joint Conference on Artificial Intelligence, Beijing, Aug. 2013, pp. 1960–1966*
24. Domain adaptation with structural correspondence learning by *J. Blitzer, R. McDonald, and F. Pereira, in Proc. Conference on Empirical Methods in Natural Language Processing, Sydney, Jul. 2006, pp. 120–128*
25. Domain adaptation for large-scale sentiment classification: A deep learning approach by *X. Glorot, A. Bordes, and Y. Bengio, in Proc. 28th International Conference on Machine Learning, Bellevue, Jun.2011, pp. 513–520*
26. Marginalized denoising autoencoders for domain adaptation by *M. Chen, Z. Xu, K. Weinberger, and F. Sha, " in Proc. 29th International Conference on Machine Learning, Edinburgh, Jun. 2012, pp. 767–774*
27. Marginalizing stacked linear denoising autoencoders by *M. Chen, K.Q. Weinberger, Z. Xu, and F. Sha, J. Mach. Learn. Res., vol. 16, no. 1, pp. 3849–3875, Jan. 2015*
28. Unsupervised visual domain adaptation using subspace alignment by *B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, in Proc. IEEE International Conference on Computer Vision, Sydney, Dec. 2013, pp. 2960–2967*
29. Subspace distribution alignment for unsupervised domain adaptation by *B. Sun and K. Saenko, in Proc. British Machine Vision Conference, Swansea, Sep. 2015, pp. 24.1–24.10*
30. Geodesic flow kernel for unsupervised domain adaptation by *B. Gong, Y. Shi, F. Sha, and K. Grauman, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Providence, Jun. 2012, pp. 2066–2073*
31. Domain adaptation for object recognition: An unsupervised approach by *R. Gopalan, Ruonan Li, and R. Chellappa, in Proc. IEEE International Conference on Computer Vision, Barcelona, Jun. 2011, pp. 999-1006*

32. Return of frustratingly easy domain adaptation by *B. Sun, J. Feng, and K. Saenko, in Proc. 30th AAAI Conference on Artificial Intelligence, Phoenix, Feb. 2016, pp. 2058–2065*
33. Cross-domain sentiment classification via spectral feature alignment by *S.J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, in Proc. 19th International Conference on World Wide Web, Raleigh, Apr. 2010, pp. 751–760*
34. Spectral domain transfer learning by *X. Ling, W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, in Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Aug. 2008, pp. 488–496*
35. Domain adaptation from multiple sources via auxiliary classifiers by *L. Duan, I.W. Tsang, D. Xu, and T.-S. Chua, in Proc. 26th International Conference on Machine Learning, Montreal, Jun. 2009, pp. 289–296*
36. Transfer learning from multiple source domains via consensus regularization by *P. Luo, F. Zhuang, H. Xiong, Y. Xiong, and Q. He, in Proc. 17th ACM Conference on Information and Knowledge Management, Napa Valley, Oct. 2008, pp. 103–112*
37. Crossdomain learning from multiple sources: A consensus regularization perspective by *F. Zhuang, P. Luo, H. Xiong, Y. Xiong, Q. He, and Z. Shi, IEEE Trans. Knowl. Data Eng., vol. 22, no. 12, pp. 1664–1678, Dec. 2010*
38. Domain adaptation from multiple sources: A domain-dependent regularization approach by *L. Duan, D. Xu, and I.W. Tsang, IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 3, pp. 504–518, Mar. 2012*
39. Inference with the universum by *J. Weston, R. Collobert, F. Sinz, L. Bottou, and V. Vapnik, in Proc. 23rd International Conference on Machine Learning, Pittsburgh, Jun. 2006, pp. 1009–1016*
40. Exploiting associations between word clusters and document classes for cross-domain text categorization by *F. Zhuang, P. Luo, H. Xiong, Q. He, Y. Xiong, and Z. Shi, Stat. Anal. Data Min., vol. 4, no. 1, pp. 100–114, Feb. 2011*
41. Triplex transfer learning: Exploiting both shared and distinct concepts for text classification by *F. Zhuang, P. Luo, C. Du, Q. He, Z. Shi, and H. Xiong, IEEE T. Cybern., vol. 44, no. 7, pp. 1191–1203, Jul. 2014*
42. Dyadic transfer learning for cross-domain image classification by *H. Wang, F. Nie, H. Huang, and C. Ding, in Proc. International Conference on Computer Vision, Barcelona, Nov. 2011, pp. 551–556*
43. Softly associative transfer learning for cross-domain classification by *D. Wang, C. Lu, J. Wu, H. Liu, W. Zhang, F. Zhuang, and H. Zhang, IEEE T. Cybern., to be published. doi: 10.1109/TCYB.2019.2891577*
44. Unveiling hidden implicit similarities for cross-domain recommendation by *Q. Do, W. Liu, J. Fan, and D. Tao, IEEE Trans. Knowl.Data Eng., to be published. doi: 10.1109/TKDE.2019.2923904*
45. The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories by *T. Tommasi and B. Caputo, in Proc. British Machine Vision Conference, London, Sep. 2009, pp. 80.1–80.11*
46. Safety in numbers: Learning categories from few examples with multi model knowledge transfer by *T. Tommasi, F. Orabona, and B. Caputo, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, Jun. 2010, pp. 3081–3088*

47. Boosting for transfer learning with multiple sources by *Y. Yao and G. Doretto, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, Jun. 2010, pp. 1855–1862*
48. Knowledge transfer via multiple model local structure mapping by *J. Gao, W. Fan, J. Jiang, and J. Han, in Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Aug. 2008, pp. 283–291*
49. Ensemble of anchor adapters for transfer learning by *F. Zhuang, P. Luo, S.J. Pan, H. Xiong, and Q. He, in Proc. 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, Oct. 2016, pp. 2335–2340*
50. Supervised representation learning: Transfer learning with deep autoencoders by *F. Zhuang, X. Cheng, P. Luo, S.J. Pan, and Q. He, in Proc. 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Jul. 2015, pp. 4119–4125*
51. Supervised representation learning with double encoding-layer autoencoder for transfer learning by *F. Zhuang, X. Cheng, P. Luo, S.J. Pan, and Q. He, ACM Trans. Intell. Syst. Technol., vol. 9, no. 2, pp. 1–17, Jan. 2018*
52. Learning transferable features with deep adaptation networks by *M. Long, Y. Cao, J. Wang, and M.I. Jordan, in Proc. 32nd International Conference on Machine Learning, Lille, Jul. 2015, pp. 97–105*
53. Transferable representation learning with deep adaptation networks by *M. Long, Y. Cao, Z. Cao, J. Wang, and M.I. Jordan, IEEE Trans. Pattern Anal. Mach. Intell., to be published. doi: 10.1109/TPAMI.2018.2868685*
54. Unsupervised domain adaptation with residual transfer networks by *M. Long, H. Zhu, J. Wang, and M.I. Jordan, in Proc. 30th Annual Conference on Neural Information Processing Systems, Barcelona, Dec. 2016, pp. 136–144*
55. Deep CORAL: Correlation alignment for deep domain adaptation by *B. Sun and K. Saenko, in Proc. European Conference on Computer Vision Workshops, Amsterdam, Oct. 2016, pp. 443–450*
56. Transferrable prototypical networks for unsupervised domain adaptation by *Y. Pan, T. Yao, Y. Li, Y. Wang, C.-W. Ngo, and T. Mei, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun. 2019, pp. 2239–2247*
57. Contrastive adaptation network for unsupervised domain adaptation by *G. Kang, L. Jiang, Y. Yang, and A.G. Hauptmann, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun. 2019, pp. 4893–4902*
58. Multi-representation adaptation network for cross-domain image classification by *Y. Zhu, F. Zhuang, J. Wang, J. Chen, Z. Shi, W. Wu, and Q. He, Neural Netw., vol. 119. pp. 214–221, Nov. 2019*
59. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources by *Y. Zhu, F. Zhuang, and D. Wang, in Proc. 33rd AAAI Conference on Artificial Intelligence, Honolulu, Jan. 2019, pp. 5989–5996*
60. Generative adversarial nets by *I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, in Proc. 28th Annual Conference on Neural Information Processing Systems, Montreal, Dec. 2014, pp. 2672–2680*
61. Unsupervised domain adaptation by backpropagation by *Y. Ganin and V. Lempitsky, in Proc. 32nd International Conference on Machine Learning, Lille, Jul. 2015, pp. 1180–1189*
62. Domain-adversarial training of neural networks by *Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F.Laviolette, M. Marchand, and V. Lempitsky, J. Mach. Learn. Res., vol. 17, pp. 1–35, Apr. 2016*

63. Conditional adversarial domain adaptation by *M. Long, Z. Cao, J. Wang, and M.I. Jordan, in Proc. 32nd Annual Conference on Neural Information Processing Systems, Montreal, Dec. 2018, pp. 1640–1650*
64. Importance weighted adversarial nets for partial domain adaptation by *J. Zhang, Z. Ding, W. Li, and P. Ogunbona, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun. 2018, pp. 8156–8163*
65. Partial transfer learning with selective adversarial networks by *Z. Cao, M. Long, J. Wang, and M.I. Jordan, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun. 2018, pp. 2724–2732*
66. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation by *X. Chen, S. Wang, M. Long, and J. Wang, in Proc. 36th International Conference on Machine Learning, Long Beach, Jun. 2019, pp. 1081–1090*
67. Mapping and revising markov logic networks for transfer learning by *L. Mihalkova, T. Huynh, and R. J. Mooney, in Proceedings of the 22nd AAAI Conference on Artificial Intelligence, Vancouver, British Columbia, Canada, July 2007, pp. 608–614*
68. Deep transfer via second-order markov logic by *J. Davis and P. Domingos, in Proc. 26th International Conference on Machine Learning, Montreal, Jun. 2009, pp. 217–224*
69. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer Learning by *H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R.M. Summers, IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, May 2016*
70. Transfer learning assisted classification and detection of Alzheimer’s disease stages using 3D MRI scans by *M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O. Song, Sensors, vol. 19, no. 11, pp. 1–19, Jun. 2019*
71. Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention UNet with transfer learning by *M. Byra, M. Wu, X. Zhang, H. Jang, Y.-J. Ma, E.Y. Chang, S. Shah, and Jiang Du, Magn. Reson. Med., Sep. 2019, doi: 10.1002/mrm.27969*
72. On combining active and transfer learning for medical data classification by *X. Tang, B. Du, J. Huang, Z. Wang, and L. Zhang, IET Comput. Vis., vol. 13, no. 2, pp. 194–205, Feb. 2019*
73. Automatic ICD-9 coding via deep transfer learning by *M. Zeng, M. Li, Z. Fei, Y. Yu, Y. Pan, and J. Wang, Neurocomputing, vol. 324, pp. 43–50, Jan. 2019*
74. Gene ontology based transfer learning for protein subcellular localization by *S. Mei, W. Fei, and S. Zhou, BMC Bioinformatics, vol. 12, no. 1, pp. 1–12, Feb. 2011*
75. Predicting plant protein subcellular multi-localization by Chou’s PseAAC formulation based multi-label homolog knowledge transfer learning by *S. Mei, J. Theor. Biol., vol. 310, pp. 80–87, Oct. 2012*
76. Transfer learning across ontologies for phenome-genome association prediction by *R. Petegrosso, S. Park, T.H. Hwang, and R. Kuang, Bioinformatics, vol. 33, no. 4, pp. 529–536, Feb. 2017*
77. A transfer learning approach for cache-enabled wireless networks by *E. Bastug, M. Bennis, and M. Debbah, in Proc. 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Mumbai, May 2015, pp. 161–166*

78. TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks by *R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, IEEE Trans. Wirel. Commun., vol. 13, no. 4, pp. 2000–2011, Apr. 2014*
79. Transfer learning for QoS aware topology management in energy efficient 5G cognitive radio networks by *Q. Zhao and D. Grace, in Proc. 1st International Conference on 5G for Ubiquitous Connectivity, Akaslompolo, Nov. 2014, pp. 152–157*
80. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection by *K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, and A. Agrawal, Constr. Build. Mater., vol. 157, pp. 322–330, Dec. 2017*
81. A comparative study of methods for transductive transfer learning by *A. Arnold, R. Nallapati, and W. W. Cohen, in Proceedings of the 7th IEEE International Conference on Data Mining Workshops. Washington, DC, USA: IEEE Computer Society, 2007, pp. 77–82*
82. Transferring knowledge from another domain for learning action models by *H. Zhuo, Q. Yang, D. H. Hu, and L. Li, in Proceedings of 10th Pacific Rim International Conference on Artificial Intelligence, December 2008*
83. Bayesian multiple instance learning: automatic feature selection and inductive transfer by *V. C. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. B. Rao, in Proceedings of the 25th International Conference on Machine learning. Helsinki, Finland: ACM, July 2008, pp. 808–815*
84. Graph-based domain mapping for transfer learning in general games by *G. Kuhlmann and P. Stone, in 18th European Conference on Machine Learning, ser. Lecture Notes in Computer Science. Warsaw, Poland: Springer, September 2007, pp. 188–200*
85. Transfer Learning for Item Recommendations and Knowledge Graph Completion in Item Related Domains via a Co-Factorization Model by *Guangyuan Piao and John G. Breslin, chapter in Lecture Notes in Computer Science · June 2018*
86. Cross Domain Recommender Systems: A Systematic Literature Review by *Muhammad Murad Khan and Roliana Ibrahim, ACM Comput. Surv. 50, 3, Article 36 (June 2017), 34 pages. DOI: http://dx.doi.org/10.1145/3073565*
87. Cross-domain collaborative filtering: A brief survey by *Bin Li, Computational Intelligence and Neuroscience Volume 2018, Article ID 1425365, 12 pages https://doi.org/10.1155/2018/1425365*
88. Cross-domain Recommender Systems by *Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, Paolo Cremonesi, in book: Recommender Systems Handbook (pp.919-959) Chapter: 27, January 2015*
89. Personalized recommendation via cross-domain triadic factorization by *Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu, In Proceedings of the 22nd International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 595–606*
90. Cross-domain collaborative filtering with factorization machines by *Babak Loni, Yue Shi, Martha Larson, and Alan Hanjalic, In Advances in Information Retrieval. Springer, 656–661*
91. Cross-domain collaboration recommendation by *Jie Tang, Sen Wu, Jimeng Sun, and Hang Su, In Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data Mining. ACM, 1285–1293*

92. Facebook single and cross domain data for recommendation systems by *Bracha Shapira, Lior Rokach, and Shirley Freilikhman, User Model. User-Adapt. Interact. 23, 2–3 (2013), 211–247*
93. Cross-domain mediation in collaborative filtering by *Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci, In User Modeling 2007. Springer, 355–359*
94. Recommendations over domain specific user graphs by *Makoto Nakatsuji, Yasuhiro Fujiwara, Akimichi Tanaka, Tadasu Uchiyama, and Toru Ishida, In ECAI. 607–612*
95. Cross-domain recommender systems by *Paolo Cremonesi, Antonio Tripodi, and Roberto Turrin, In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW'11). IEEE, 496–503*
96. A cross-domain recommendation model for cyber-physical systems by *Sheng Gao, Hao Luo, Da Chen, Shantao Li, Patrick Gallinari, Zhanyu Ma, and Jun Guo, IEEE Trans. Emerg. Top. Comput. 1, 2 (2013), 384–393*
97. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction by *Bin Li, Qiang Yang, and Xiangyang Xue, In IJCAI, Vol. 9. 2052–2057*
98. Multi-domain collaborative filtering by *Yu Zhang, Bin Cao, and Dit-Yan Yeung, ArXiv Preprint ArXiv:1203.3535 (2012)*
99. Compressed knowledge transfer via factorization machine for heterogeneous collaborative recommendation by *Weike Pan, Zhuode Liu, Zhong Ming, Hao Zhong, Xin Wang, and Congfu Xu, Knowledge-Based Systems (2015)*
100. Adaptive bayesian personalized ranking for heterogeneous implicit feedbacks by *Weike Pan, Hao Zhong, Congfu Xu, and Zhong Ming, Knowl.-Based Syst. 73 (2015), 173–180*
101. Constrained collective matrix factorization by *Yu-Jia Huang, Evan Wei Xiang, and Rong Pan, In Proceedings of the Sixth ACM Conference on Recommender Systems. ACM, 237–240*
102. A nonlinear cross-site transfer learning approach for recommender systems by *Xin Xin, Zhirun Liu, and Heyan Huang, In Neural Information Processing. Springer, 495–502*
103. Cross-domain recommender systems: A survey of the state of the art by *Fernández-Tobías, Ivan Cantador, Marius Kaminskas, and Francesco Ricci, n Proceedings of the Spanish Conference on Information Retrieval*
104. Talmud: transfer learning for multiple domains by *Orly Moreno, Bracha Shapira, Lior Rokach, and Guy Shani, In Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, 425–434*
105. Making recommendations from multiple domains by *Wei Chen, Wynne Hsu, and Mong Li Lee, In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 892–900*
106. User-based collaborative filtering on cross domain by tag transfer learning by *Weiqing Wang, Zhenyu Chen, Jia Liu, Qi Qi, and Zhihong Zhao, In Proceedings of the 1st International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining. ACM, 10–17*
107. Auxiliary domain selection in cross-domain collaborative filtering by *Chang Yi, Ming-Sheng Shang, and Qian-Ming Zhang, Appl. Math 9, 3 (2015), 1375–1381*
108. The application of transfer learning on e-commerce recommender systems by *Jiuhong Tang, Zhihong Zhao, Jia Bei, and Weiqing Wang, In Proceedings of the 2013 10th Web Information System and Application Conference (WISA'13). IEEE, 479–482*

109. Cross-domain collaborative filtering over time by *Bin Li, Xingquan Zhu, Ruijiang Li, Chengqi Zhang, Xiangyang Xue, and Xindong Wu, In Proceedings of the 22nd International Joint Conference on Artificial Intelligence-Volume Volume Three. AAAI Press, 2293–2298*
110. Recommendation algorithm based on Bispectral clustering and rating-matrix transfer learning by *Xiang Li, Ronglin Hu, Quanyin Zhu, and Haiyan Zhang, Journal of Computational and Theoretical Nanoscience 13, 3 (2016), 1971–1978*
111. Building ontologies for cross-domain recommendation on facial skin problem and related cosmetics by *Hla Hla Moe and Win Thanda Aung, In Proceedings of the International Journal of Information Technology and Computer Science (IJITCS'14) 6, 6 (2014), 33*
112. Semantic clustering-based cross-domain recommendation by *Ajit Kumar, Narendra Kumar, Mutawarra Hussain, Santanu Chaudhury, and Sankalp Agarwal, In Proceedings of the 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM'14). IEEE, 137–141*
113. Social recommendation across multiple relational domains by *Meng Jiang, Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu, and Shiqiang Yang, In Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, 1422–1431*
114. Cross-domain recommendation without shared users or items by sharing latent vector distributions by *Tomoharu Iwata and Koh Takeuchi, In Proceedings of the 18th International Conference on Artificial Intelligence and Statistics. 379–387*
115. Transfer Learning for Content-Based Recommender Systems Using Tree Matching by *Naseem Biadsy, Lior Rokach, and Armin Shmilovici, In Availability, Reliability, and Security in Information Systems and HCI. Springer, 387–399*
116. Cross-domain scientific collaborations prediction with citation information by *Ying Guo and Xi Chen, In Proceedings of the 2014 IEEE 38th International Computer Software and Applications Conference Workshops (COMPSACW'14). IEEE, 229–233*
117. Build your own music recommender by modeling internet radio streams by *Natalie Aizenberg, Yehuda Koren, and Oren Somekh, In Proceedings of the 21st International Conference on World Wide Web. ACM, 1–10*
118. Improving cross-domain recommendation through probabilistic cluster-level latent factor model by *Siting Ren, Sheng Gao, Jianxin Liao, and Jun Guo, n Proceedings of the 29th AAAI Conference on Artificial Intelligence*
119. Selective transfer learning for cross domain recommendation by *Zhongqi Lu, Weike Pan, Evan Wei Xiang, Qiang Yang, Lili Zhao, and ErHeng Zhong, In SDM. SIAM, 641–649*
120. Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering by *Yue Shi, Martha Larson, and Alan Hanjalic, In User Modeling, Adaption and Personalization. Springer, 305–316*
121. Active transfer learning for cross-system recommendation by *Lili Zhao, Sinno Jialin Pan, Evan Wei Xiang, Erheng Zhong, Zhongqi Lu, and Qiang Yang, In AAAI. Citeseer*
122. Exploiting social tags for cross-domain collaborative filtering by *Yue Shi, Martha Larson, and Alan Hanjalic, ArXiv Preprint arXiv:1302.4888 (2013)*
123. Transfer learning in collaborative filtering with uncertain ratings by *Weike Pan, Evan Wei Xiang, and Qiang Yang, in AAAI, 2012*

124. Transfer learning in heterogeneous collaborative filtering domains by *Weike Pan and Qiang Yang, Artif. Intell. 197 (2013), 39–55*
125. Twin bridge transfer learning for sparse collaborative filtering by *Jiangfeng Shi, Mingsheng Long, Qiang Liu, Guiguang Ding, and Jianmin Wang, in Advances in Knowledge Discovery and Data Mining. Springer, 496–507*
126. A transfer probabilistic collective factorization model to handle sparse data in collaborative filtering by *How Jing, An-Chun Liang, Shou-De Lin, and Yu Tsao, In Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM'14). IEEE, 250–259*
127. Experimental analysis on cross domain preferences association and rating prediction by *Zhenhua Dong and Qian Zhao, in Proceedings of the 1st International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining. ACM, 26–31*
128. Tag propagation based recommendation across diverse social media by *Deqing Yang, Yanghua Xiao, Yangqiu Song, Junjun Zhang, Kezun Zhang, and Wei Wang, in Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 407–408*
129. A framework for cross-domain recommendation in folksonomies by *Ying Guo and Xi Chen, J. Autom. Contr. Eng. 1, 4 (2013)*
130. Context aware cross-domain based recommendation by *Hla Hla Moe and Win Thanda Aung, in Proceedings of the International Conference on Advances in Engineering and Technology, International Institute of Engineers, (2014)*
131. Distributed Collaborative Filtering with Domain Specialization by *Berkovsky, S., Kuflik, T., Ricci, F, 1st ACM Conference on Recommender Systems, pp. 33–40 (2007)*
132. . Correlating User Profiles from Multiple Folksonomies by *Szomszor, M. N., Cantador, I., Alani, H., 19th ACM Conference on Hypertext and Hypermedia, pp. 33–42 (2008)*
133. Semantic Modelling of User Interests Based on Cross-Folksonomy Analysis by *Szomszor, M. N., Alani, H., Cantador, I., O'Hara, K., Shadbolt, N., 7th International Semantic Web Conference, pp. 632–648 (2008)*
134. Analyzing Cross-system User Modeling on the ´ Social Web by *Abel, F., Araujo, S., Gao, Q., Houben, G.-J., 11th International Conference on Web Engineering, pp. 28–43 (2011)*
135. Cross-system User Modeling and Personalization on the Social Web by *Abel, F., Helder, E., Houben, G.-J., Henze, N., Krause, D., User Modeling and User-Adapted Interaction 23(2-3), pp. 169–209 (2013)*
136. An Emotion Dimensional Model Based on Social Tags: Crossing Folksonomies and Enhancing Recommendations by *Fernández-Tobías, I., Cantador, I., Plaza, L., 14th International Conference on E-Commerce and Web Technologies, pp. 88–100 (2013)*
137. How to Recommend Music to Film Buffs: Enabling the Provision of Recommendations from Multiple Domains by *Loizou, A., PhD thesis, University of Southampton (2009)*
138. . A Generic Semantic-based Framework for Cross-domain Recommendation by *Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F., 2nd International*

*Workshop on Information Heterogeneity and Fusion in Recommender Systems, pp. 25–32 (2011)*
139. Ontology-based Identification of Music for Places by *Kaminskas, M., Fernández-Tobías, I., Ricci, F., Cantador, I., 13th International Conference on Information and Communication Technologies in Tourism, pp. 436–447 (2013)*
140. Cross Social Networks Interests Predictions Based on Graph Features by *Tiroshi, A., Berkovsky, S., Kaafar, M. A., Chen, T., Kuflik, T., 7th ACM Conference on Recommender Systems pp. 319–322 (2013)*
141. A Multi-agent Smart User Model for Cross- domain Recommender Systems by *González, G., López, B., de la Rosa, J. LL., in: Beyond Personalization 2005 - The Next Stage of Recommender Systems Research, pp. 93–94 (2005)*
142. Relating Personality Types with User Preferences in Multiple Entertainment Domains by *Cantador, I., Fernández-Tobías, I., Bellogín, A., Kosinski, M., Stillwell, D.*, 1st Workshop on Emotions and Personality in Personalized Services, CEUR workshop Proceedings, vol. 997 (2013)
143. Domain Ranking for Cross Domain Collaborative Filtering by *Tiroshi, A., Kuflik, T., 20th International Conference on User Modeling, Adaptation, and Personalization, pp. 328–333 (2012)*
144. Identifying Inter-domain Similarities through Content-based Analysis of Hierarchical Web-Directories by *Berkovsky, S., Goldwasser, D., Kuflik, T., Ricci, F., 17th European Conference on Artificial Intelligence, pp. 789–790 (2006)*
145. Cross-tagging for Personalized Open Social Networking by *Stewart, A., Diaz-Aviles, E., Nejdl, W., Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., 20th ACM Conference on Hypertext and Hypermedia, pp. 271–278 (2009)*
146. Multiple Domain User Personalization by *Low, Y., Agarwal, D., Smola, A. J., 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 123–131 (2011)*
147. Predicting Social-tags for Cold Start Book Recommendations by *Givon, S., Lavrenko, V., 3rd ACM Conference on Recommender Systems, pp. 333–336 (2009)*
148. Integrated Personal Recommender Systems by *Chung, R., Sundaram, D., Srinivasan, A., 9th International Conference on Electronic Commerce, pp. 65–74 (2007)*
149. Crossing: A Framework to Develop Knowledge-based Recommenders in Cross Domains by *Azak, M., MSc thesis, Middle East Technical University (2010)*
150. Probabilistic Matrix Factorization by *Salakhutdinov, R., Mnih, A., Advances in Neural Information Processing Systems 20, pp. 1257–1264 (2008)*
151. Transfer Learning for Collective Link Prediction in Multiple Heterogeneous Domains by *Cao, B., Liu, N.N., Yang, Q., 27th International Conference on Machine Learning, pp.159–166 (2010)*
152. Transfer Learning in Collaborative Filtering for Sparsity Reduction by *Pan, W., Xiang, E.W., Liu, N.N., Yang, Q., 24th AAAI Conference on Artificial Intelligence, pp.210–235 (2010)*
153. Transfer Learning to Predict Missing Ratings via Heterogeneous User Feedbacks by *Pan, W., Liu, N. N., Xiang, E. W., Yang, Q., 22nd International Joint Conference on Artificial Intelligence, pp. 2318–2323 (2011)*
154. Cold-Start Management with Cross-Domain Collaborative Filtering and Tags by *Enrich, M., Braunhofer, M., Ricci, F., 14th International Conference on E-*

*Commerce and Web Technologies, pp. 101–112 (2013)*

155. Exploiting Social Tags in Matrix Factorization Models for Cross-domain Collaborative Filtering by *Fernández-Tobías, I., Cantador, I., 1st International Workshop on New Trends in Contentbased Recommender Systems (2013)*
156. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model by *Koren, Y., 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 426–434 (2008)*
157. Web Personalization Expert with Combining Collaborative Filtering and Association Rule Mining Technique by *Lee, C. H., Kim, Y. H., Rhee, P. K., Expert Systems with Applications 21(3), pp. 131–137 (2001)*
158. Transfer collaborative filtering from multiple sources via consensus regularization by *Fuzhen Zhuang, Jing Zheng, Jingwu Chen, Xiangliang Zhang, Chuan Shi, Qing He, Neural Netw., vol. 108, pp. 287–295, Dec. 2018*
159. Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model by *Bin Li, Qiang Yang, Xiangyang Xue, 26th International Conference on Machine Learning, pp. 617–624 (2009)*
160. Cross-domain recommendations without overlapping data: myth or reality? by *Cremonesi, P., Quadrana, M., 8th ACM Conference on Recommender Systems (2014)*
161. Cross-Domain Recommendation via Cluster-Level Latent Factor Mode by *Gao, S., Luo, H., Chen, D., Li, S., Gallinari, P., Guo, J., 17th and 24th European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 161–176 (2013)*
162. Introduction to Recommender Systems Handbook by *Francesco Ricci, Lior Rokach, and Bracha Shapira*
163. If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations by *Pinata Winoto and Tiffany Tang, New Generation Computing 26, pp. 209–225 (2008)*
164. Transfer Learning using MXNet by *Prince Grover*, Published in Heartbeat, March 2020
165. Creating a movie recommender using Convolutional Neural Networks by *Kasper Knol,* Published in Towards Data Science, February 2018