

Disparity-preserved Deep Cross-platform Association for Cross-platform Video Recommendation

Shengze Yu¹, Xin Wang^{1*}, Wenwu Zhu^{1*}, Peng Cui¹ and Jingdong Wang²

¹ Department of Computer Science and Technology, Tsinghua University, China

² Microsoft Research

ysz0323@163.com, {xin_wang, wwzhu, cuip}@tsinghua.edu.cn, jingdw@microsoft.com

Abstract

Cross-platform recommendation aims to improve recommendation accuracy through associating information from different platforms. Existing cross-platform recommendation approaches assume all cross-platform information to be consistent with each other and can be aligned. However, there remain two unsolved challenges: i) there exist inconsistencies in cross-platform association due to platform-specific disparity, and ii) data from distinct platforms may have different semantic granularities. In this paper, we propose a cross-platform association model for cross-platform video recommendation, i.e., Disparity-preserved Deep Cross-platform Association (DCA), taking platform-specific disparity and granularity difference into consideration. The proposed DCA model employs a partially-connected multi-modal autoencoder, which is capable of explicitly capturing platform-specific information, as well as utilizing nonlinear mapping functions to handle granularity differences. We then present a cross-platform video recommendation approach based on the proposed DCA model. Extensive experiments for our cross-platform recommendation framework on real-world dataset demonstrate that the proposed DCA model significantly outperform existing cross-platform recommendation methods in terms of various evaluation metrics.

1 Introduction

Recommender systems are playing an important role in our life. With the emergence of various online services, people are now getting used to engaging on different platforms simultaneously in order to meet their increasing diverse information needs [Chen *et al.*, 2012b]. The complementary information from various platforms jointly reflects user interests and preferences, providing us with a great opportunity to tackle the data sparsity problem and improve the recommendation accuracy through associating information across

platforms. In order to improve user experience and increase user adoption, many online service providers release new features to encourage cross-platform data associations. For example, Google+ encourages users to share their homepages on other platforms, enabling different accounts of the same person to be linked together. To date, quite a number of various platforms tend to associate with each other, which provides chances for cross-platform information transfer and association analysis possible.

Existing cross-platform recommendation works try to improve their cross-platform performances mainly by investigating the ability of cross-platform association. Yan *et al.* propose a user-centric topic association framework to map cross-platform data in a common latent space [2014]. They then introduce a predefined micro-level metric to adaptively weight data while doing the data integration [2015]. Man *et al.* propose an embedding and mapping framework, EM-CDR, to represent and associate users across different platforms [2017]. However, existing works on cross-platform video recommendation ignore the inconsistencies in cross-platform association and differences in semantic granularities, two challenging phenomena discovered in this paper.

The first challenge, inconsistency in cross-platform association, results in a reconsideration for the current assumption that all cross-platform information of the same user is consistent and can be aligned, which will be empirically validated on a real-world cross-platform dataset later. Our discovery demonstrates that the inconsistency is mainly caused by the platform-specific disparity, i.e., in addition to a user's inherent personal preferences, its interests shown on different platforms also contain platform-specific factors due to different focuses of various platforms. The second challenge is that different platforms may have different semantic granularities. Take Twitter and YouTube as an example, tags on Twitter are mainly hot event topics, which are generally finer-grained than video categories on YouTube. As such, a good cross-platform video recommendation model based on the correlated information across platforms should have the ability to reveal characteristics at different semantic granularities.

To address these challenges, we propose the Disparity-preserved Deep Cross-platform Association (DCA) model for cross-platform recommendation, which employs a partially-connected multi-modal autoencoder to explicitly capture and preserve platform-specific disparities in latent representa-

*Corresponding Author

tion. Besides, we also introduce nonlinear mapping function, which is superior to their linear competitors in terms of handling granularity difference. To validate the advantages of our proposed DCA model, we further design a cross-platform video recommendation approach based on the proposed DCA model, which is capable of automatically targeting and transferring useful cross-platform information from a comprehensive semantic level. The contributions of this paper can be summarized as follows:

- We recognize the inconsistency phenomenon, which is caused by platform-specific disparity, as well as the granularity difference problem in cross-platform association.
- We propose a novel **Disparity-preserved Deep Cross-platform Association (DCA)** model for cross-platform recommendation through employing a modified multi-modal autoencoder, which is able to handle the cross-platform inconsistency issue by preserving the platform-specific disparities and solve the granularity difference problem via nonlinear functions.
- We present a cross-platform video recommendation approach based on the proposed DCA model to test its effectiveness compared with other approaches.
- We conduct experiments on a real-world cross-platform dataset on Twitter and YouTube to demonstrate the superiority of our proposed DCA model over several state-of-the-art methods in terms of various evaluation metrics.

2 Related Work

Existing cross-platform recommendation works can be categorized in two ways.

Categorization by what to associate. One option is to take advantage of different platform characteristics towards collaborative applications. Particularly, Qian et al. propose a generic cross-domain collaborative learning framework based on nonparametric Bayesian dictionary learning model for cross-domain data analysis [2015]. Min et al. propose a cross-platform multi-modal topic model which is capable of differentiating topics and aligning modalities [2015]. An alternative option is to associate in a user-centric way, which focuses on integrating multiple sources of overlapped user’s activities. In particular, XPTrans [Jiang et al., 2016] optimally bridges different platforms through exploiting the information from a small number of overlapped crowds. Yan et al. propose an overlapped user-centric topic association framework based on latent attribute sparse coding, and prove that bridging information from different platforms in common latent space outperforms explicit matrix-oriented transfer [2014]. Man et al. propose an embedding and mapping framework, EMCDR, where user representations on different platforms are first learnt through matrix factorization and then mapped via multi-layer perceptron [2017]. Although the user-centric works above are based on different premises, they share the same core idea that all cross-platform information is consistent and should be aligned. However, several works have pointed out the data inconsistency phenomenon in cross-platform data association, and attempted to solve this

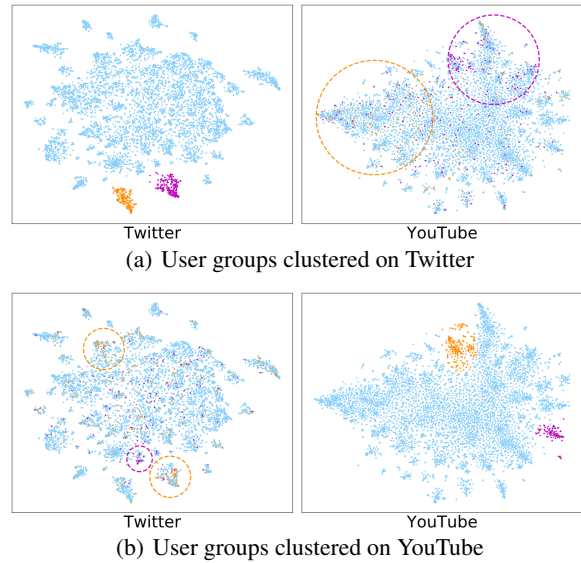


Figure 1: Visualization of users on Twitter and YouTube in the CASIA-crossOSN subset. Orange and purple dots represent two groups of aligned users across platforms. Concentrated areas are circled with dashed lines.

problem by data selection. To be concrete, Lu et al. find that selecting consistent auxiliary data is important for cross-domain collaborative filtering [2013]. They propose a novel criterion to assess the degree of consistency, and embed it into a boosting framework to selectively transfer knowledge. Yan et al. divide users into three groups and introduce a pre-defined micro-level user-specific metric to adaptively weight data while integrating information across platforms [2015]. Our proposed DCA model associates information from different platforms in the user-centric way.

Categorization by entire model structure. One group of methods [Chen et al., 2012a; Jiang et al., 2016; Jing et al., 2014; Qian et al., 2015] build the model in a unified framework, where the former two works adopt matrix factorization and the latter two employ the probabilistic models. Another group of works [Yan et al., 2014; Yan et al., 2015; Man et al., 2017] adopt a two-step framework. These works first map users from different platforms into corresponding latent spaces for representations, and then associate these representations. Our proposed model in this work utilizes the two-step structure in a topic-based way.

3 Measurement and Observation

To validate the existence of the inconsistency, we first visualize users in CASIA-crossOSN, a cross-platform dataset linking user accounts between YouTube and Twitter (whose detailed description will be given in Section 6), following three steps as follows. 1) For users who engage in both Twitter and YouTube, we extract their interest representations by Latent Dirichlet Allocation (LDA). 2) We adopt the interest representations as inputs to the visualization tool t-SNE [Maaten and Hinton, 2008] so that users are represented as two-dimensional vectors and visualized as dots — the

Platform	Average Euclidean distance from center of user groups		
	Random	Clustering on Twitter	Clustering on YouTube
Twitter	0.346 (1)	0.244 (0.705)	0.341 (0.985)
YouTube	0.224 (1)	0.223 (0.995)	0.198 (0.883)

Table 1: Average Euclidean distance from the center of user groups clustered on Twitter (resp. YouTube) vs. random. User groups clustered on Twitter (resp. YouTube) no longer concentrate on YouTube (resp. Twitter), indicating cross-platform user interest disparity.

closer two dots are in distance, the more similar their corresponding users are in topic space. 3) We cluster users into different groups by their Twitter (resp. YouTube) interest representations, select two groups, and mark them with orange and purple respectively in Figure 1 such that users who share a same color are regarded as similar.

We observe from Figure 1 that only a part of users who are similar on Twitter (resp. YouTube) still stay similar on YouTube (resp. Twitter) (circled with dashed line) and the rest actually becomes quite different. Moreover, in Figure 1(b), orange user group clustered on YouTube splits into two subset on Twitter, indicating that two platforms may have different semantic granularities.

In addition to visualizing the interest representations, Table 1 further confirms the existence of the inconsistency phenomenon through a measurement study on CASIA-crossOSN. i) For each clustered user group on Twitter and YouTube, we calculate the average Euclidean distance from the group center to group members. ii) We do the same thing for randomly sampled groups on both platforms and use the calculated average Euclidean distances as normalizers to obtain the ‘‘concentration ratio’’ such that a larger concentration ratio indicates a less clustered pattern.

It can be observed from Table 1 that when we cluster users by their interest representations on Twitter (resp. YouTube), the concentration ratio is small on the origin platform while becomes quite large (close to 1) on the other platform, indicating that users having similar interests on one platform tend to behave quite similarly to randomly sampled users on the other — they no longer share similar interests.

We close this section by giving a conclusion on the existence of inconsistency in cross-platform association: users’ interests on different platforms may be diverse and inconsistent.

4 DCA: Disparity-preserved Deep Cross-platform Association

In this section, we employ multi-modal autoencoder to present our disparity-preserved deep cross-platform association (DCA) model.

4.1 Problem Formulation

We first introduce the problem of user-centric cross-platform association. Given a set of aligned users U , for each user $u \in U$, his or her representations on different platforms (\mathbf{u}^T on Twitter and \mathbf{u}^Y on YouTube in the experiment) are known.

The goal is to find the association among these representations (direct mapping function or unified latent representation), so that cross-platform applications can be carried out (specifically inferring \mathbf{u}^Y for video recommendation in the experiment).

4.2 Disparity-preserved Deep Cross-platform Association

Multi-modal autoencoder is an extension to basic autoencoder. It is trained to reconstruct multiple inputs from a unified hidden layer when given multiple modalities of the same entity, thus is able to discover correlations across modalities. It has been widely applied to various multi-modal tasks since being proposed [Ngiam *et al.*, 2011; Zhang *et al.*, 2014; Wang *et al.*, 2015]. Hong *et al.* extend the concept of ‘‘multi-modal’’ to ‘‘multi-source’’ [2015]. In this paper, we treat representations of the same user on different platforms as multiple modalities of his or her unified latent representation, and modify the structure of multi-modal autoencoder to conduct disparity-preserved cross-platform association. The detailed structure of multi-modal autoencoder is formulated as:

$$\mathbf{h} = g \left[\left(\sum_i \mathbf{W}_1^i \mathbf{x}^i \right) + \mathbf{b}_1 \right], \quad \hat{\mathbf{x}}^i = g \left(\mathbf{W}_2^i \mathbf{h} + \mathbf{b}_2^i \right), \quad (1)$$

where $i \in \{T, Y\}$ denotes different platforms (e.g., Twitter and YouTube). \mathbf{x}^i is the ‘‘multi-modal’’ input layer, specifically the representations \mathbf{u}^i of the user u on different platforms. \mathbf{h} is the hidden layer, i.e., the derived unified user latent representation. $\hat{\mathbf{x}}^i$ is the output layer, i.e., the reconstructions of inputs. Weight matrices \mathbf{W} and bias units \mathbf{b} serve as parameters of the multi-modal autoencoder, denoted as θ together. $g(\cdot)$ is the activation function. The loss function of multi-modal autoencoder is defined as follows:

$$L(\mathbf{x}^i; \theta) = \sum_i \left\| \hat{\mathbf{x}}^i - \mathbf{x}^i \right\|_2^2 + \lambda \sum_{\mathbf{W} \in \theta} \|\mathbf{W}\|_F^2 + \mu \|\mathbf{h}\|_1, \quad (2)$$

where the first term is reconstruction error. The second term is regularizer to avoid overfitting. The third term is a sparsity constraint to encourage sparse latent representations. We minimize the loss function and train the parameters through backpropagation.

Multi-modal autoencoder takes representations of the same user on different platforms as inputs, mapping them to a unified hidden layer \mathbf{h} and expecting accurate reconstructions. It learns mapping functions from representations on different platforms to the unified representation and then vice versa, which automatically completes the process of cross-platform association. Moreover, it is able to handle the granularity difference through introduction of nonlinearity.

We note that all multi-modal inputs should be given for the multi-modal autoencoder to work properly. However, in many cross-platform applications, we need to infer unknown data on other platforms, which is extremely difficult. One straightforward solution is to train several networks with a certain combination of several platforms as inputs and output the results for all platforms, setting every decoding weight

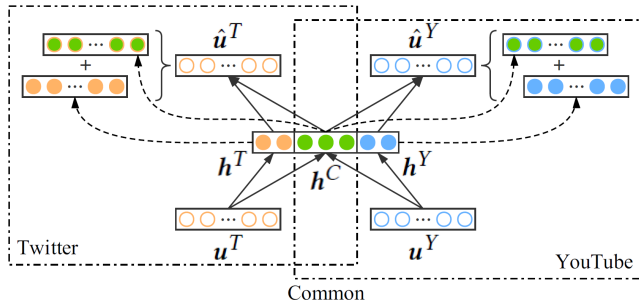


Figure 2: Disparity-preserved Deep Cross-platform Association model. u^T and u^Y are representations of a same user on Twitter and YouTube respectively. In latent representations, h^T and h^Y are platform-specific parts preserving the disparities, and h^C is the common part associating different platforms. The estimated representations \hat{u}^T and \hat{u}^Y are derived from both common and platform-specific parts.

equally. However, such an approach does not scale well as we will need an exponential number of models. Ngiam et al. propose to train bi-modal autoencoder using augmented but noisy data with additional examples that have only one single modality as inputs [2011]. In this paper, we augment data with average values of different platforms, and still expect the network to “reconstruct” real representations for all modalities. To make the model become more robust to absent inputs and being capable of inferring unknown data in cross-platform association task, we split the training data as follows: one-third of the training data includes user representations on Twitter and average user representation on YouTube, one-third consists of average user representation on Twitter and user representations on YouTube, the remaining one-third contains groundtruth representations.

As aforementioned, platform-specific disparity may cause inconsistencies in cross-platform association. We need to envisage it and avoid messing up the association. To achieve this goal, we divide the hidden layer into three parts and cut off certain links, through which the shared and platform-specific parts are explicitly captured and preserved. Specifically, the hidden layer is divided into $h = [h^T, h^C, h^Y]$, where h^T and h^Y are Twitter and YouTube platform-specific parts and h^C is the common part. As is shown in Figure 2, we assume that Twitter representations are derived from h^T and h^C without h^Y while YouTube representations are derived from h^Y and h^C without h^T . The fully-connected structure is modified through cutting off links between units: from x^T to h^Y , from x^Y to h^T , from h^T to \hat{x}^Y and from h^Y to \hat{x}^T . The model is able to automatically map representations of the same user on different platforms to a unified representation while preserving the information in platform-specific parts at the same time. It does not require cross-platform data to be thoroughly consistent by allowing the existences of inconsistencies. Thanks to the usage of shared common structure, cross-platform association becomes clearer and tighter.

In addition, most traditional linear methods associate representations on different platforms by utilizing a linear transfer matrix, which is formulated in Eq. (3) and Eq. (4). The linear combination limits the power in modeling complex re-

lations across different semantic granularities. Our modified multi-modal autoencoder naturally introduces nonlinear activation function $g(\cdot)$, possessing great advantages in handling the challenging granularity differences over linear methods. We choose Sigmoid as the activation function in this paper and other activation functions such as ReLU or tanh can also be used.

The time complexity is $O(MNk(n_1 + n_2))$ for training and $O(k(n_1 + n_2))$ for testing. M is the number of training iterations, N is the number of users for training, k is the hidden layer dimension, n_1 and n_2 are the input dimensions of two platforms. Our proposed DCA model, whose structure is presented in Figure 2, automatically captures common information and preserves platform-specific disparities through its partially-connected structure, and meanwhile naturally adapts to granularity difference via nonlinear functions.

5 Cross-platform Video Recommendation

In this section, we present the cross-platform video recommendation approach based on the proposed DCA model.

5.1 Problem Formulation

We first formulate the cross-platform video recommendation problem. Given a set of aligned users U , for each user $u \in U$, his or her behaviors on social platform are observed (i.e., tweet text t in collection T_u on Twitter). The goal of cross-platform video recommendation is to recommend videos for those aligned users on video platform (i.e., YouTube) through making use of the cross-platform information.

5.2 DCA-based Cross-platform Video Recommendation

As is shown in Figure 3, our proposed cross-platform video recommendation method includes preprocessing and matching stages in addition to the DCA model. The capability of DCA model in tackling the cross-platform inconsistency issue through explicitly capturing shared information and preserving platform disparities enables the cross-platform recommendation method to automatically concentrate on useful information and discard useless one. Moreover, nonlinear mapping functions in our proposed approach enhance the method’s ability to conduct association across platforms at a deeper and more comprehensive semantic level.

In order to apply our proposed DCA model for the associations between Twitter and YouTube, we first obtain user representations by Latent Dirichlet Allocation (LDA) [Blei et al., 2003]. On Twitter, we regard each user’s tweeting data together as a document in Latent Dirichlet Allocation (LDA). Then each user $u \in U$ can be represented by a topic distribution u^T . On YouTube, as we focus on the relevance between videos and users in the recommendation problem, we extract representations for both users and videos by two steps: (1) treat the surrounding textual information (title, category, tags and description) of each video together as a document and to get video representation v through online LDA [Hoffman et al., 2010], (2) take the average v of the interacted videos consumed by each user to get u^Y . We remark that v and u^Y are in the same YouTube topic space, which enables similarity matching for recommendation.

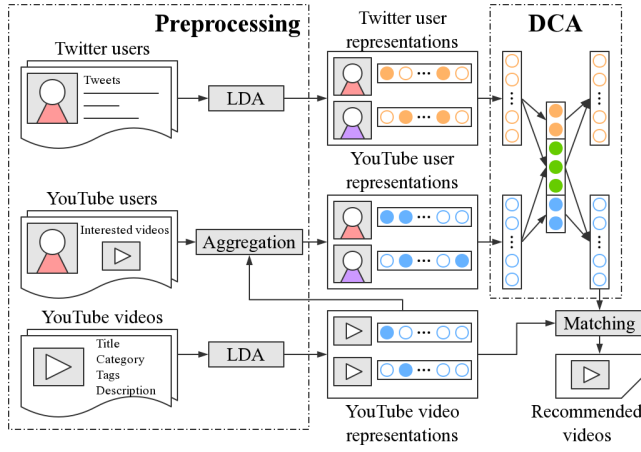


Figure 3: Cross-platform video recommendation framework based on the proposed Disparity-preserved Deep Cross-platform Association model

Then the representation pairs of the same user as well as the augmented examples are fed to the DCA model to learn cross-platform association. Because of the input augmenting, the trained model is able to predict user representations on YouTube when only information from Twitter is available. By setting the input from YouTube to its average value, the corresponding output can be used to predict user representations \hat{u}^Y .

We finally adopt Euclidean similarities between the representations of the predicted user and candidate videos as the properness measure. The recommendation results can be obtained according to the rank of candidate videos.

6 Experiments

In this section, we carry out extensive experiments on a real-world cross-platform dataset and compare our proposed method with several state-of-the-art algorithms to show the advantages of the proposed approach.

6.1 Dataset

CASIA-crossOSN¹ is a cross-network user dataset with account linkages between YouTube and Twitter, created by Institute of Automation, Chinese Academy of Sciences. It contains 11,687 aligned users across platforms and 2,280,129 YouTube videos. On YouTube, three kinds of user behaviors (i.e., uploading, favoring and adding to playlist) as well as videos' rich metadata are collected. On Twitter, only users' representations extracted from their tweets via standard topic modeling (LDA) process [Blei *et al.*, 2003] are provided because of privacy concerns. We filter aligned users and YouTube videos by keeping users who interacted with at least 3 videos and videos which were consumed by at least 3 users. We finally obtain a subset containing 7,338 users and 34,144 videos, and randomly select 80% users for training, and the rest 20% for testing.

¹CASIA-crossOSN dataset: <http://www.nlpr.ia.ac.cn/mmc/homepage/myan/dataset.html>

6.2 Experimental Settings

On Twitter, CASIA-crossOSN offers a 60-dimensional topical distribution for each user. On YouTube, we resort to perplexity to determine the number of topics. We select 80, which leads to the smallest perplexity on a 5% held-out YouTube video subset, in our experiments.

To evaluate the effectiveness of the proposed DCA model, we implement the following baselines:

Linear Regression-based Association (LR)

LR treats cross-platform association as a linear transfer problem, and pursues an explicit transfer matrix based on regression. The objective function is:

$$\min_{\mathbf{W}} \left\| \mathbf{W}\mathbf{U}^T - \mathbf{U}^Y \right\|_F^2 + \lambda \|\mathbf{W}\|_2, \quad (3)$$

Latent Attribute-based Association (LA)

Instead of pursuing hard transfer, Yan *et al.* introduce another method by discovering the shared latent structure behind two topic spaces [2014], whose objective function is:

$$\begin{aligned} \min_{\mathbf{D}^T, \mathbf{D}^Y, \mathbf{S}} & \left\| \mathbf{U}^T - \mathbf{D}^T \mathbf{S} \right\|_F^2 + \left\| \mathbf{U}^Y - \mathbf{D}^Y \mathbf{S} \right\|_F^2 + \lambda \|\mathbf{S}\|_1, \\ \text{s.t.} & \left\| \mathbf{d}_i^Y \right\|_2 \leq 1, \left\| \mathbf{d}_j^T \right\|_2 \leq 1, \quad \forall i, j, \end{aligned} \quad (4)$$

The problem can be efficiently solved by sparse coding algorithm [Lee *et al.*, 2007] after a few transformations.

MLP-based Nonlinear Mapping (MLP)

Man *et al.* employ MLP to cope with the latent space matching problem [2017]. MLP is a flexible nonlinear transformation. The optimization problem can be formulated as:

$$\min_{\theta} \sum_{\mathbf{u} \in \mathbf{U}} \left\| f_{mlp}(\mathbf{u}^T; \theta) - \mathbf{u}^Y \right\|_2^2, \quad (5)$$

To train the above neural network-based model and our DCA model, we adopt Adam optimizer to minimize the loss functions.

6.3 Evaluation of Cross-platform Association

As for the proposed model in this paper, multi-modal autoencoder with \mathcal{L}_1 -norm but without disparity-preserved structure (MA) and the DCA model are examined.

We perform evaluations in two scenarios: (1) Infer YouTube from Twitter. Given users with their Twitter representation \mathbf{u}^T , to estimate their YouTube representation \mathbf{u}^Y . (2) Infer Twitter from YouTube in turn. As is discussed before, by setting the unknown input to zeros, we can get the corresponding output as the predicted representation $\hat{\mathbf{u}}$. In both scenarios, we utilize Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as evaluation metric:

$$\begin{aligned} \text{MAE}^i &= \frac{1}{|U_{test}|} \sum_{U_{test}} \frac{1}{K^i} \left\| \hat{\mathbf{u}}^i - \mathbf{u}^i \right\|_1, \\ \text{RMSE}^i &= \frac{1}{|U_{test}|} \sum_{U_{test}} \sqrt{\frac{1}{K^i} \left\| \hat{\mathbf{u}}^i - \mathbf{u}^i \right\|_F^2} \end{aligned} \quad (6)$$

where $i \in \{T, Y\}$ denotes platform, and K^i is the dimension of representation. \mathbf{u}^i and $\hat{\mathbf{u}}^i$ are the real and predicted user representation.

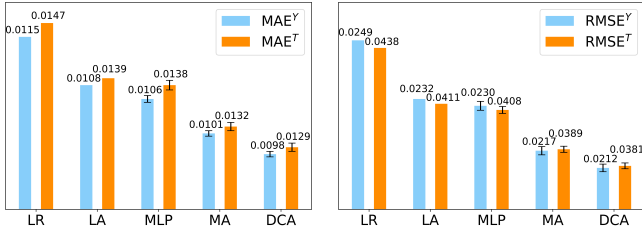


Figure 4: MAE^T , MAE^Y and $RMSE^T$, $RMSE^Y$ of the examined models in two testing scenarios. The results show that both disparity preserving and nonlinearity contribute to better cross-platform association performance.

For model parameters, in LR model as Eq. 3, λ is selected by line search. In LA model as Eq. 4, λ and the number of latent attributes m are selected by two-dimensional grid search. In the DCA model as Eq. 2, parameters consist of λ , μ and m , which further divides into $[m^T, m^C, m^Y]$. We first tune the MA model (equivalent to set $m^T = m^Y = 0, m = m^C$ in the DCA model) by multi-dimensional grid search, and get an optimal setting for λ , μ and m . We then carefully fine-tune $[m^T, m^C, m^Y]$ by a combined line-search strategy. Specifically for all randomly initialized autoencoder model, we conduct 6 experiments and take the average result.

Performances are shown in Figure 4. Several observations can be made: (1) LA and MLP achieve better result than LR, indicating that the shared latent structure and non-linear mapping function both benefit the association. (2) MA outperforms LA and MLP through combining the two strategies. (3) Compared with MA, DCA further improves the performance, demonstrating the contribution of the disparity-preserved structure. We also notice that compared with non-linear models (MLP, MA and DCA), the optimal parameters of the linear models vary greatly between two scenarios. We think it is the granularity differences that cause the parameter gaps. Non-linear mappings can naturally tackle this issue, and have more advantages over linear ones. In total, the proposed DCA model outperforms LA by reducing relatively 9.3% on MAE^Y , 8.6% on $RMSE^Y$, 7.2% on MAE^T and 7.3% on $RMSE^T$, and outperforms MLP by reducing relatively 7.5% on MAE^Y , 7.8% on $RMSE^Y$, 6.5% on MAE^T and 6.6% on $RMSE^T$. It greatly improves the performance by preserving the disparities and introducing nonlinearity, validating that envisaging the inconsistency caused by platform-specific disparity and granularity difference jointly contributes to better cross-platform association performance.

6.4 Evaluation of Cross-platform Video Recommendation

Then we evaluate the proposed DCA-based cross-platform video recommendation method for new YouTube users, following the above ‘‘infer YouTube from Twitter’’ scenario.

For each test user u , we randomly select as many as groundtruth interacted videos as additional candidates. We perform a top- k recommendation task: to recommend the top k YouTube videos with the highest topic-based Euclidean similarity, and adopt top- k precision, recall and F-score as

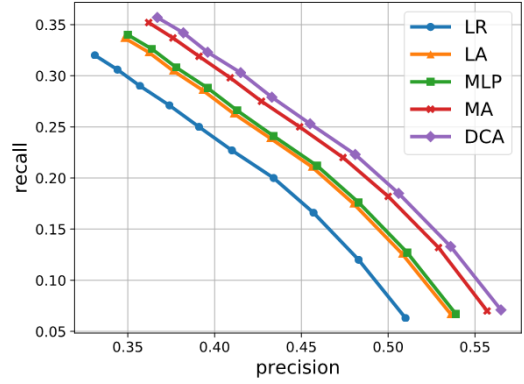


Figure 5: Precision@ k vs. Recall@ k , k from 1 to 10 of examined methods.

Model	Precision	Recall	F-score
LR	0.331	0.320	0.268
LA	0.348	0.337	0.279
MLP	0.350	0.340	0.281
MA	0.362	0.352	0.290
DCA	0.367	0.357	0.296

Table 2: Top-10 Precision, Recall and F-score of examined methods. The results show that the proposed DCA-based cross-platform video recommendation method outperforms other approaches.

evaluation metrics [Herlocker *et al.*, 2004]. The evaluation metrics are calculated by examining whether the recommended videos are included in u ’s groundtruth video set V_u . Final results are averaged over all test users.

As is shown in Figure 5 and Table 2, the proposed DCA-based cross-platform video recommendation method outperforms LA-based method by enhancing relatively 6.1% on F-score@10, and outperforms MLP-based method by enhancing relatively 5.3% on F-score@10. The proposed DCA-based method improves over existing cross-platform association-based recommendation approaches.

7 Conclusion

In this paper, we discover the existence of inconsistency in cross-platform recommendation. We propose the DCA model, which tackles the inconsistency issue, as well as the granularity difference problem. We further present a cross-platform video recommendation method based on the proposed DCA model. Extensive experiments demonstrate the superiority of the DCA model and the DCA-based cross-platform recommendation approach over several state-of-the-art methods.

Acknowledgments

This research is supported by National Program on Key Basic Research Project No. 2015CB352300, National Natural Science Foundation of China Major Project No. U1611461, China Postdoctoral Science Foundation No. BX201700136 and Shenzhen Nanshan District Ling-Hang Team Grant under No.LHTD20170005.

References

- [Blei *et al.*, 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [Chen *et al.*, 2012a] Leihui Chen, Jianbing Zheng, Ming Gao, Aoying Zhou, Wei Zeng, and Hui Chen. Tlrec: transfer learning for cross-domain recommendation. *Computer Science*, pages 167–172, 2012.
- [Chen *et al.*, 2012b] Terence Chen, Mohamed Ali Kaafar, Arik Friedman, and Rokhsana Boreli. Is more always merrier?: a deep dive into online social footprints. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pages 67–72. ACM, 2012.
- [Herlocker *et al.*, 2004] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.
- [Hoffman *et al.*, 2010] Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *advances in neural information processing systems*, pages 856–864, 2010.
- [Hong *et al.*, 2015] Chaoqun Hong, Jun Yu, Jian Wan, Dacheng Tao, and Meng Wang. Multimodal deep autoencoder for human pose recovery. *IEEE Transactions on Image Processing*, 24(12):5659–5670, 2015.
- [Jiang *et al.*, 2016] Meng Jiang, Peng Cui, Nicholas Jing Yuan, Xing Xie, and Shiqiang Yang. Little is much: Bridging cross-platform behaviors through overlapped crowds. In *AAAI*, pages 13–19, 2016.
- [Jing *et al.*, 2014] How Jing, An-Chun Liang, Shou-De Lin, and Yu Tsao. A transfer probabilistic collective factorization model to handle sparse data in collaborative filtering. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pages 250–259. IEEE, 2014.
- [Lee *et al.*, 2007] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2007.
- [Lu *et al.*, 2013] Zhongqi Lu, Erheng Zhong, Lili Zhao, Evan Wei Xiang, Weike Pan, and Qiang Yang. Selective transfer learning for cross domain recommendation. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 641–649. SIAM, 2013.
- [Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [Man *et al.*, 2017] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. Cross-domain recommendation: An embedding and mapping approach. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 2464–2470, 2017.
- [Min *et al.*, 2015] Weiqing Min, Bing Kun Bao, Changsheng Xu, and M. Shamim Hossain. Cross-platform multi-modal topic modeling for personalized inter-platform recommendation. *IEEE Transactions on Multimedia*, 17(10):1787–1801, 2015.
- [Ngiam *et al.*, 2011] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.
- [Qian *et al.*, 2015] Shengsheng Qian, Tianzhu Zhang, Richang Hong, and Changsheng Xu. Cross-domain collaborative learning in social multimedia. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 99–108. ACM, 2015.
- [Wang *et al.*, 2015] Daixin Wang, Peng Cui, Mingdong Ou, and Wenwu Zhu. Deep multimodal hashing with orthogonal regularization. In *IJCAI*, pages 2291–2297, 2015.
- [Yan *et al.*, 2014] Ming Yan, Jitao Sang, and Changsheng Xu. Mining cross-network association for youtube video promotion. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 557–566. ACM, 2014.
- [Yan *et al.*, 2015] Ming Yan, Jitao Sang, and Changsheng Xu. Unified youtube video recommendation via cross-network collaboration. In *Proceedings of the 5th ACM International Conference on Multimedia Retrieval*, pages 19–26. ACM, 2015.
- [Zhang *et al.*, 2014] Hanwang Zhang, Yang Yang, Huanbo Luan, Shuicheng Yang, and Tat-Seng Chua. Start from scratch: Towards automatically identifying, modeling, and naming visual attributes. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 187–196. ACM, 2014.