# Transfer Learning from APP Domain to News Domain for Dual Cold-Start Recommendation

Jixiong Liu<sup>†</sup>, Jiakun Shi<sup>†</sup>, Wanling Cai<sup>†</sup>, Bo Liu<sup>‡</sup>, Weike Pan<sup>†</sup> Qiang Yang<sup>\*‡</sup> and Zhong Ming<sup>\*†</sup> <sup>†</sup>College of Computer Science and Software Engineering, Shenzhen University <sup>‡</sup>Department of Computer Science and Engineering, Hong Kong University of Science and Technology {1455606137,1033150729,382970614}@qq.com, {bliuab,qyang}@cse.ust.hk, {panweike,mingz}@szu.edu.cn \*: corresponding author.

# ABSTRACT

News recommendation has been a must-have service for most mobile device users to know what has happened in the world. In this paper, we focus on recommending latest news articles to new users, which consists of the new user coldstart challenge and the new item (i.e., news article) coldstart challenge, and is thus termed as dual cold-start recommendation (DCSR). As a response, we propose a solution called *neighborhood-based transfer learning* (NTL) for this new problem. Specifically, in order to address the new user cold-start challenge, we propose a cross-domain preference assumption, i.e., users with similar app-installation behaviors are likely to have similar tastes in news articles, and then transfer the knowledge of neighborhood of the coldstart users from an APP domain to a news domain. For the new item cold-start challenge, we design a category-level preference to replace the traditional item-level preference because the latter is not applicable for the new items in our problem. We then conduct empirical studies on a real industry data with both users' app-installation behaviors and news-reading behaviors, and find that our NTL is able to deliver the news articles more accurately than other methods on different ranking-oriented evaluation metrics.

## **CCS Concepts**

•Information systems  $\rightarrow$  Personalization; •Humancentered computing  $\rightarrow$  Collaborative filtering;

## Keywords

Transfer Learning; News Recommendation; Cold-Start Recommendation

# 1. INTRODUCTION

Intelligent recommendation systems [5] have been a ubiquitous service in our daily life, which has saved us a lot of

RecSysKTL Workshop @ ACM RecSys '17, August 27, 2017, Como, Italy © 2017 Copyright is held by the author(s)

time in finding proper information such as music, goods and news articles. For instance, personalized news recommendation [1, 2] has been one of the must-have services for most mobile device users, which plays an important role in helping users keep up with the current affairs in the world. In this paper, we focus on recommending latest news articles to new users, i.e., the users are newly registered in a certain news recommendation service and have not read any news articles before, and the news articles have not been read by any users before. We term it as *dual cold-start recommendation* (DCSR), denoting both cold-start users and cold-start items.

For the dual cold-start problem, previous news recommendation methods [1, 2] are not applicable, because they rely on users' historical reading behaviors and news articles' content information that are not available in our case.

We turn to address the cold-start recommendation problem from a transfer learning [3, 4] view. Although there are no users' behaviors about the cold-start users and cold-start items in the news domain, there may be some other related domains with users' behaviors. Specifically, we leverage some knowledge from a related domain, i.e., APP domain, where the users' app-installation behaviors are available. We find that most cold-start users in the news domain have already installed some apps, and that may be helpful in determining his/her preferences in news articles. In particular, we assume that users with similar app-installation behaviors are likely to have similar interests in some news topics. In other words, close neighbors in the APP domain are likely to be close neighbors in the news domain.

With the above cross-domain preference assumption, we propose to take the neighborhood in the APP domain as the knowledge and try to transfer it to the target domain of news articles. Specifically, we design a *neighborhood-based transfer learning* (NTL) solution that transfers knowledge of neighborhood from the APP domain to the news domain, which addresses the new user cold-start challenge. With the neighborhood, some well-studied neighborhood-based recommendation methods are applicable for news recommendation.

We conduct empirical studies on a real industry data in order to verify our cross-domain preference assumption and the effectiveness of our transfer learning solution. Experimental results show that the two domains of apps and news articles are indeed related and can share some knowledge for preference learning.

## 2. OUR SOLUTION

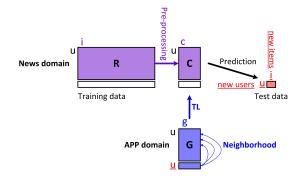


Figure 1: An illustration of *neighborhood-based* transfer learning (NTL) for dual cold-start recommendation (DCSR).

# 2.1 **Problem Definition**

In our studied news recommendation problem, we have two domains, including an APP domain and a news domain.

Firstly, in the APP domain, we have a set of triples, i.e.,  $(u, g, G_{ug})$ , denoting that user u has installed  $G_{ug}$  times of mobile apps belonging to the genre g. The data of the APP domain can then be represented as a user-genre matrix **G** as shown in Figure 1.

Secondly, in the news domain, we have a user-item matrix **R** denoting whether a user has read an item. Each item *i* is associated with a level-1 category  $c_1(i)$  and a level-2 category  $c_2(i)$ . We thus have a set of quadruples, i.e.,  $(u, i, c_1(i), c_2(i))$ , denoting that user *u* has read an item *i* belonging to  $c_1(i)$  and  $c_2(i)$ . Finally, we have a user-category matrix **C** after pre-processing, where each entry denotes the number of items belonging to a certain category that a user has read.

Our goal is to recommend a ranked list of new items (i.e., latest news articles) to each new user, who has not read any items before. We can see that it is a new user cold-start and new item cold-start problem, which is thus termed as *dual cold-start recommendation* (DCSR). Note that we only make use of items' category information, but not content information.

We put some notations in Table 1.

#### 2.2 Challenges

The main difficulty of the DCSR problem is the lack of preference data for new users and new items. Specifically, there are two challenges, including (i) the new user cold-start challenge, i.e., the target users (to whom we will provide recommendations) have not read any items before; and (ii) the new item cold-start challenge, i.e., the target items (that we will recommend to the target users) are totally new for all users. Under such a situation, most existing recommendation algorithms are not applicable.

#### 2.3 Neighborhood-based Transfer Learning

In most recommendation methods [5], the user-user (or item-item) similarity is a central concept, because the neighborhood can be constructed for like-minded users' preference aggregation and then for the target user's preference prediction. Mathematically, the preference prediction rule for user u to item i is as follows,

$$\hat{r}_{u,i} = \frac{1}{|\mathcal{N}_u|} \sum_{u' \in \mathcal{N}_u} \hat{r}_{u',i},\tag{1}$$

where  $\mathcal{N}_u$  is a set of nearest neighbors of user u in terms of a certain similarity measurement such as cosine similarity, and  $\hat{r}_{u',i}$  is the estimated preference of user u' (a close neighbor of user u) to item i. The aggregated and normalized score  $\hat{r}_{u,i}$  is taken as the preference of user u to item i, which is further used for item ranking and top-K recommendation.

For our studied *dual cold-start recommendation* problem, we can not build correlations between a cold-start user in the test data and a warm-start user in the training data using the data from the news domain only. The main idea of our transfer learning [3] solution is to leverage the correlations among the users in the APP domain with the assumption that users with similar app-installation behaviors are likely to be similar in news taste. For instance, two users with the installed apps of the same genre *business* may both prefer news articles on topics like *finance*.

With the cross-domain preference assumption, we first calculate the cosine similarity between a cold-start user u and a warm-start user u' in the APP domain as follows,

$$s_{u,u'} = \frac{G_{u}.G_{u'.}^T}{\sqrt{G_{u}.G_{u}^T}\sqrt{G_{u'}.G_{u'.}^T}},$$
(2)

where  $G_{u}$  is a row vector w.r.t. user u from the user-genre matrix **G**. Once we have calculated the cosine similarity, for each cold-start user u, we first remove users with a small similarity value (e.g.,  $s_{u,u'} < 0.1$ ), and then take some (e.g., 100) most similar users to construct a neighborhood  $\mathcal{N}_u$ .

For the item-level preference  $\hat{r}_{u',i}$  in Eq.(1), we are not able to have such a score directly because the item *i* is new for all users, including the warm-start users and the target cold-start user u'. We thus propose to approximate the itemlevel preference using a category-level preference,

$$\hat{r}_{u',i} \approx \hat{r}_{u',c(i)},\tag{3}$$

where c(i) can be the level-1 category or level-2 category. We then have two types of category-level preferences,

$$\hat{r}_{u',c(i)} = \hat{r}_{u',c_1(i)} = N_{u',c_1(i)},$$
(4)

$$\hat{r}_{u',c(i)} = \hat{r}_{u',c_2(i)} = N_{u',c_2(i)},$$
(5)

where  $N_{u',c_1(i)}$  and  $N_{u',c_2(i)}$  denote the number of read items (by user u') belonging to the level-1 category  $c_1(i)$  and the level-2 category  $c_2(i)$ , respectively.

Finally, with the Eqs.(3-5), we can rewrite Eq.(1) as follows,

$$\hat{r}_{u,i} \approx \frac{1}{|\mathcal{N}_u|} \sum_{u' \in \mathcal{N}_u} N_{u',c_1(i)},\tag{6}$$

$$\hat{r}_{u,i} \approx \frac{1}{|\mathcal{N}_u|} \sum_{u' \in \mathcal{N}_u} N_{u',c_2(i)},$$
(7)

which will be used for preference prediction in our empirical studies. Specifically, the neighborhood  $\mathcal{N}_u$  addresses the new user cold-start challenge, and the category-level preference  $N_{u',c_1(i)}$  or  $N_{u',c_2(i)}$  addresses the new item cold-start challenge.

Table 1: Some notations and explanations. Note that when a user u has read the same article more than one time, we only count it one in calculating  $N_{u,c_1}$  and  $N_{u,c_2}$ .

u	user id
i	item (i.e., news article) id
g	genre id of the apps
$\mathcal{C}_1$	a set of level-1 categories, $c_1 \in \mathcal{C}_1$
$\mathcal{C}_2$	a set of level-2 categories, $c_2 \in \mathcal{C}_2$
$N_{u,c_1}$	the number of read items (by user $u$ ) belonging to a level-1 category $c_1$
$N_{u,c_2}$	the number of read items (by user $u$ ) belonging to a level-2 category $c_2$
$N_{c_1} = \sum_u N_{u,c_1}$	the number of read items (by all users) belonging to a level-1 category $c_1$
$N_{c_2} = \overline{\sum}_{u=1}^{u} N_{u,c_2}$	the number of read items (by all users) belonging to a level-2 category $c_2$
$N_{c_{2}}^{-1} = \frac{\sum_{u}^{u} N_{u,c_{2}}^{u,r_{1}}}{\sum_{c_{1}} \sum_{c_{1}' \in \mathcal{C}_{1}} N_{c_{1}'}}$ $p_{c_{2}} = \frac{\sum_{c_{2}' \in \mathcal{C}_{2}} N_{c_{1}'}}{\sum_{c_{2}' \in \mathcal{C}_{2}} N_{c_{2}'}}$	the popularity of the level-1 category $c_1$ among the users
$p_{c_2} = \frac{N_{c_2}}{\sum_{c_2' \in \mathcal{C}_2} N_{c_2'}}$	the popularity of the level-2 category $c_2$ among the users
$\mathcal{N}_u$	a set of neighbors of user $u$

# 3. EXPERIMENTAL RESULTS

=

## **3.1 Dataset and Evaluation Metrics**

In our empirical studies, we use a real industry data, which consists of an APP domain and a news domain.

**APP Domain** In the auxiliary domain, i.e., APP domain, we have 827,949 users and 53 description terms (i.e., genres) of the users' installed mobile apps, where the genres are from Google Play. Considering our target task of news recommendation, we removed 14 undiscriminating or irrelevant genres such as *tools, communication, social, entertainment, productivity, weather, dating,* etc. Finally, we have a matrix **G** with 827,949 users (or rows) and 39 genres (or columns), where each entry represents the number of times that a user has installed apps belonging to a genre.

**News Domain** In the target domain, i.e., news domain, we have two sets of data, including a training data and a test data. The training data spans from 10 January 2017 to 30 January 2017, and contains 806,167 users, 747,643 items (i.e., news articles), and 16,199,385 unique (user, item) pairs. We can see that a user has read about 16199385/806167 = 20.09 articles on average from 10 January 2017 to 30 January 2017. The test data are from the data on 31 January 2017, which contains 3,597 new users, 28,504 new items (i.e., news articles), and 4,813 unique (user, item) pairs. We can see that a cold-start user read about 4813/3597 = 1.34 articles on 31 January 2017. Note that we have  $|C_1| = 26$  level-1 categories and  $|C_2| = 222$  level-2 categories about the items in the news domain.

For performance evaluation, we adopt some commonly used evaluation metrics in ranking-oriented recommendation such as precision, recall, F1, NDCG and 1-call. Specifically, we study the average performance of the top-15 recommended list generated for each cold-start user in the test data.

## 3.2 Baselines and Parameter Settings

We compare our proposed transfer learning solution with a random method and two popularity-based methods using category information.

• Random recommendation (Random). In Random, we randomly select K = 15 items in the test data for each cold-start user.

- Popularity-based ranking via level-1 category (PopRank-C1). In PopRank-C1, we first calculate the popularity  $p_{c_1}$  of each level-1 category  $c_1 \in C_1$  in the training data, and then use  $\hat{r}_i = p_{c_1(i)}$  in Table 1 as the score to rank each item (i.e., article) i in the test data. For the most popular level-1 category, there may be more than K = 15 items (i.e., articles) in the test data, we then randomly take K items (i.e., articles) from that level-1 category for recommendation.
- Popularity-based ranking via level-2 category (PopRank-C2). In PopRank-C2, we use  $\hat{r}_i = p_{c_2(i)}$  in Table 1 as the prediction rule similar to that of PopRank-C1.

For the number of neighbors in our neighborhood-based transfer learning method, we first fix it as 100, and then change it to 50 and 150 in order to study its impact. We denote our transfer learning solution with level-1 category as NTL-C1 and that with level-2 category as NTL-C2, where their prediction rules are shown in Eq.(6) and Eq.(7), respectively. Note that for Random, PopRank-C1, PopRank-C2, and NTL with randomly selected neighbors, we repeat the experiments for 10 times, and report the average results.

#### 3.3 Results

We report the main results in Table 2. From Table 2, we can have the following observations:

- The overall performance ordering is PopRank-C1, Random, PopRank-C2  $\ll$  NTL-C2 < NTL-C1, which clearly shows the effectiveness of our proposed transfer learning solution to the challenging *dual cold-start recommendation* problem.
- The performance of PopRank-C2 and PopRank-C1 are rather poor in comparison with our proposed solution. The reason is that popularity-based methods are nonpersonalized methods and will simply select one most popular level-1 category or level-2 category for all users, which ignores the difference in users' news reading preferences.
- For the relative performance of NTL-C1 and NTL-C2, we can see that NTL-C1 performs better as expected because the level-1 category may introduce more smoothing effect for the cold-start problem.

Table 2: Recommendation performance of random recommendation, popularity-based ranking, and our NTL on cold-start users.

Method	Prec@15	Rec@15	F1@15	NDCG@15	1-call@15
Random	5.56E-05	5.84E-04	9.78E-05	2.27E-04	8.34E-04
PopRank-C1	5.00E-05	5.59E-04	9.02E-05	2.38E-04	7.51E-04
PopRank-C2	1.46E-04	1.74E-03	2.65E-04	6.48E-04	2.20E-03
NTL-C1	0.0053	0.0645	0.0095	0.0255	0.0734
NTL-C2	0.0040	0.0501	0.0073	0.0206	0.0567

We further study the impact of the neighborhood size. The results of our NTL-C1 using 50, 100 and 150 neighbors are shown in Figure 2. We can see that the results are relatively stable with different numbers of neighbors, and configuring it as 100 usually produces the best performance.

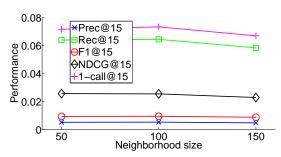


Figure 2: Recommendation performance of our NTL with level-1 category (NTL-C1) using different neighborhood sizes.

In order to gain some deep understanding about the transferred neighborhood, we study the performance of randomly choosing the same number (i.e., 100) of neighbors in our NTL-C1. We report the results in Figure 3, from which we can have the observations:

- The neighborhood constructed using the app-installation behaviors is better than that of the random counterpart, which shows that the two domains are related and can indeed transfer knowledge from one domain to the other.
- The difference between the two types of neighborhood is not as large as that between popularity-based methods and our NTL in Table 2, which can be explained by the fact that a large portion of users' preferences or tastes in news articles are similar.

# 4. CONCLUSIONS AND FUTURE WORK

In this paper, we study an important and challenging news recommendation problem called *dual cold-start recommendation* (DCSR), which aims to recommend latest news articles (cold-start items) to newly registered users (cold-start users). Specifically, we propose a *neighborhood-based transfer learning* (NTL) solution, which is able to address the new user cold-start challenge and the new item cold-start challenge by the transferred neighborhood from the APP domain and the category-level preferences in the news domain, respectively. Empirical results on a real industry data show

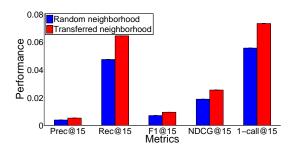


Figure 3: Recommendation performance of our NTL with level-1 category (NTL-C1) using random neighborhood and transferred neighborhood.

that our NTL performs significantly more accurate than the few applicable methods, i.e., popularity-based ranking using category information.

For future works, we are interested in selecting some representative genres and categories in two domains and building a mapping between them, which will be further used to study the neighborhood of the items.

## 5. ACKNOWLEDGMENT

We thank the support of Natural Science Foundation of China (NSFC) Nos. 61502307 and 61672358, China National 973 project No. 2014CB340304, and Hong Kong CERG projects Nos. 16211214, 16209715 and 16244616.

## 6. **REFERENCES**

- A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 271–280, 2007.
- [2] J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In Proceedings of the 15th International Conference on Intelligent User Interfaces, IUE '10, pages 31–40, 2010.
- [3] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- [4] W. Pan. A survey of transfer learning for collaborative recommendation with auxiliary data. *Neurocomputing*, 177:447–453, 2016.
- [5] F. Ricci, L. Rokach, and B. Shapira. Recommender Systems Handbook (Second Edition). Springer, 2015.