Persona Model Transfer for User Activity Prediction Across Heterogeneous Domains

Takahiro Hara[✉]

Osaka University, 1-5 Yamadaoka, Suita, Osaka 5650871, Japan
hara@ist.osaka-u.ac.jp
http://www-nishio.ist.osaka-u.ac.jp/∼hara/

Abstract. In this keynote talk, we present our project on cross-domain digital marketing where we assume totally different service domains such as Web advertisement domain and E-commerce domain. Cross-domain approaches are useful in situations where some domain does not have enough amount data to develop an accurate prediction model on user activities. Our idea is to transfer persona (user) model from one domain which has richer data to the target domain with less data, i.e., it has a worse prediction model. This project is technically very challenging since we assume totally different domains where the users' activities are different. We present some of recent achievements of our project and also talk about our future plans.

Keywords: Digital marketing · Machine learning · Prediction model

1   Project Overview

This project aims to develop cross-domain approaches based on machine learning and big data processing to provide digital marketing services across different services (domains). While there have been a large number of online services, they generally cannot share raw data such as service usage logs and user IDs due to privacy and right issues, and thus, it often happens that service providers cannot provide sufficient personalized services even if they have a large amount of data in total. Our idea to tackle this problem is developing techniques for persona matching across differently domains and that for transferring prediction models from a source domain to a target domain. With this approach, a domain with less data (i.e., historical data on service usage) can reuse a richer model for user activity prediction built in another domain. Our project is technically very challenging since we assume totally different domains where the users' activities are different. We need to catch some hints to predict user activities in one domain from the model constructed in the other domain in which user activities are totally different. To do so, we develop machine learning and big data processing techniques.

Figure 1 shows the overview of research topics in our project. Below, we briefly present the outline of each topic.
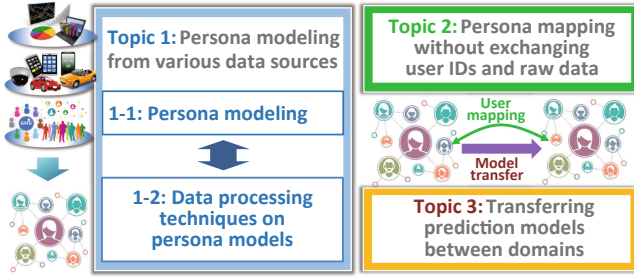
Fig. 1. Overview of our research project

1.1   Topic 1: Persona Modeling from Various Data Sources

Persona modeling is the most fundamental operations in our project. We assume that a persona corresponds to a user, but it can be a virtual user including a representative entity of a group of users. A persona has general attributes such as age, sex, and preferences, and also has activity models, both of which are constructed from data obtained in each domain (e.g., service usage logs). A persona model defines not only each persona but also relationships between personas. It can be represented as a graph where nodes correspond to users and edges corresponds to relationships between users. It also can be represented as embeddings where user vectors represents both the users' characteristics and the relationships with other users simultaneously, e.g., the similarity between two users can be represented as the distance in the vector space.

This topic has a subtopic on data processing on persona models. Because a persona model basically has a very complicated data structure, efficient data processing techniques are essential. Our goal here is to achieve a few decades or hundreds times faster data processing than existing techniques.

1.2   Topic 2: Persona Mapping Without Exchanging User IDs and Raw Data

Identifying same or similar users (user groups) between different domains is useful for effective digital marketing services, e.g. customer transfer across domains. This topic aims to develop matching techniques of same or similar users between domains. Here, we assume that we have bridge users who give us a permission to use their IDs (i.e., ID matching can be made) and their service usage logs in both domains. Therefore, we use data obtained from the bridge users to learn attributive and structural similarity of same or similar users in both domains (in the training phase), and use the findings to identify same or similar (non-bridge) users (in the test phase).

1.3   Topic 3: Transferring Prediction Models Between Domains

This topic addresses the main issue of our project. We assume that there is some latent space which covers all domains. Thus, transferring a prediction model from

one (source) domain to another (target) domain is identical to a task of finding a reverse projection function from the source domain to the latent space and then finding a projection function from the latent space to the target domain.

2   Cross-Domain Digital Marketing: Web Advertisement × E-Commerce

We have conducted our first study on cross-domain digital marketing since 2019, where we obtained data from real services in a Web advertisement domain and an E-commerce domain [2]. We performed persona modeling (topic 1), persona matching (topic 2) and cross-domain recommendation (topic 3) as shown in Fig. 2.
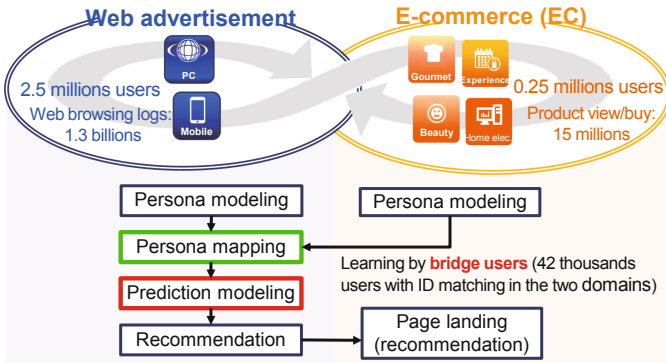

Fig. 2. Web advertisement domain × E-commerce domain

2.1   Persona Modeling

We have developed two different approaches for persona modeling as below.

Content Based Approach. The first one is a word2vec (or content) based approach where documents in Web pages which were browsed by users (i.e., ads were shown on the pages) are used to generate the users' embeddings and documents in product descriptions are used to generate the products' embeddings. For both user and product embeddings, a word2vec technique is used where each dimension in the embeddings corresponds to a same word for both embeddings.

This content based approach aims to tackle to a cold-start problem for both users and products in the e-commerce domain. In most e-commerce services, products on sale quickly change (e.g., in almost every 2 weeks) and most users registered to the services have only a few times or no purchase experiences, i.e.,

most products and users have no interactions. Thus, it is difficult or almost impossible to effectively model new products and new users from the purchase history (i.e., interactions in the e-commerce domain) and predict such users' purchase activities.

Our idea to solve this cold-start problem is utilizing a cross-domain approach. More specifically, since most people often browse Web pages (i.e., they have enough historical data to model themselves), we try to transfer a rich persona model constructed in the Web advertisement domain to the E-commerce domain. Our hypothesis here is that while these two domains have totally different characteristics, there are some hints in Web browsing pattern to predict user (purchase) activities in the E-commerce domain.

Meta-pass Based Approach. The second approach is a meta-pass (interaction) based approach where the information on interactions between users and Web pages (i.e. Web browsing) and that between users and products (i.e. purchase) are used for persona modeling. The basic idea of using rich information in the Web advertisement domain is the same as that of the first approach, but it is totally different because it does not use any texts in persona modeling.

This approach is motivated by the fact that the first approach (i.e. content based approach) suffers from information losses on user modeling which are caused by blocked accesses, missing links, and meaningless contents. Since the second approach does not use any textual information on user modeling, it does not suffer from such information loses. In addition, the meta-pass based approach has another advantage that it can distinguished two cases in which users browse similar Web pages such as Yahoo! news and Google news. This is because even if two Web pages are similar, these have different URLs. In many cases, such differences in choice of services well represent differences in user preferences.

On the other hand, except for information loss cases, contents generally have richer information than interaction data. Therefore, in total, it depends on situations whether the content based approach or the meta-pass based approach works well.

2.2   Cross-Domain Product Recommendation

After generating user and product embeddings, we apply DMF [3] and NeuMF [1] methods (the original methods have been extended to fit to our problem) to build a prediction model of user purchase activities.

Figure 3 shows a result of performance studies. We compare the top-k hit ratios of our word2vec based methods (denoted by DMF and NeuMF) with some comparison methods including random recommendation and cosine-similarity based method. As a result, we found that our methods significantly outperform the comparison methods. In particular, NeuMF achieved about 26% hit ratio by recommending 10 products (i.e. $k = 10$) among 1500 products on sale during the test period, which is surprisingly high.
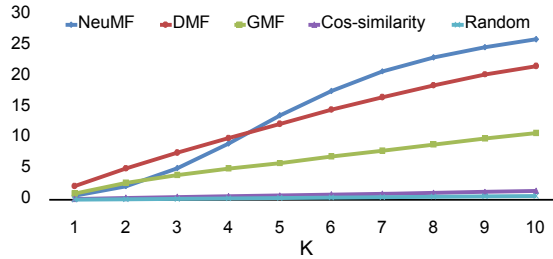
Fig. 3. Performance study

3   Future Plans

We have just started to work on cross-domain approaches for user activity prediction using other domains' data such as public WiFi. We have also worked on environmental modeling using SNS data to catch the trend and user preferences, which can be used as a bias for user activity prediction.

We also plan to investigate the impact of unusual situations such as COVID-19 issue on user activities. It is obvious that user activities significantly changed in such situations, however it is not easy to know how user activity prediction models can adjust to the changes. Therefore, we will work on research of prediction model transfer from ordinary situations to unusual situations.

References

1. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.-S.: Neural collaborative filtering. In: Proceedings of the International Conference on World Wide Web, pp. 173–182 (2017)
2. Wang, H., et al.: A DNN-based cross-domain recommender system for alleviating cold-start problem. IEEE Open J. Ind. Electron. Soc. **1**, 194–206 (2020)
3. Xue, H.-J., Dai, X.-Y., Zhang, J., Huang, S., Chen, J.: Deep matrix factorization models for recommender systems. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 3203–3209 (2017)