Sequential Transfer Learning: Cross-domain Novelty Seeking Trait Mining for Recommendation

Fuzhen Zhuang[†], Yingmin Zhou[†,‡], Fuzheng Zhang[§], Xiang Ao[†], Xing Xie[§], Qing He[†]
[†]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing.
[‡]University of Chinese Academy of Sciences, Beijing.
[§]Microsoft Research.
{zhuangfz, zhouym, aox, heq}@ics.ict.ac.cn, {fuzzhang, xing.xie}@microsoft.com.

ABSTRACT

Recent studies in psychology suggest that novelty-seeking trait is highly related to consumer behavior, which has a profound business impact on online recommendation. This paper studies the problem of mining novelty seeking trait across domains to improve the recommendation performance in target domain. We propose an efficient model, CDNST, which significantly improves the recommendation performance by transferring the knowledge from auxiliary source domain. We conduct extensive experiments on three domain datasets crawled from Douban (www.douban.com) to demonstrate the effectiveness of the proposed model. Moreover, we find that the property of sequential data affects the performance of CDNST.

Keywords

Recommendation; Novelty-seeking Trait; Transfer Learning

1. INTRODUCTION

In consumer behavior and recommender system research, understanding personality trait is particularly crucial since consumers' attributes are strong indicators of their purchasing behaviors [3]. Unlike most of previous works, there has been some efforts devoted to model an individual's propensity from psychological perspective for recommendation systems in recent years, and Zhang et al. [4] proposed to model novelty-seeking trait in one single domain for personalization recommendation. Behaviors of users are relatively consistent in similar situations [1]. The modeling of novelty-seeking trait in one single domain may not completely characterize each individual's profiles, while the sequential behavioral data of one user from different domains may help to exploit the novelty-seeking trait. We are interested in this problem and thus crawled tags of movies, music, and books from the well known Chinese social media platform Douban[1] commented by users, the viewing time and ID of users. Observing these three domains of sequential behavioral data, we find that users listened some music and then after a period of time they would watch some related movie, e.g., the music is the theme music of the movie; users sometimes watch some movies after they read some related books, from which the movies are derived. Based on these observations, whether the sequential behavioral data of domains of Book and Music can help to model the novelty-seeking trait in the domain of Movie? This issue might be helpful to improve the recommendation performance. On the other hand, transfer learning aims to transfer the knowledge from related auxiliary source domain to target domain. Along this line, we propose a new cross-domain novelty-seeking trait mining model, termed as CDNST, in which the parameters characterizing the novelty-seeking trait are shared across different domains to achieve improvements for recommendation.

[1]https://www.douban.com/

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054229

2. MODEL AND SOLUTION

Action and **Choice**: Denote x^s and x^t the specific observed behavior taken by an individual in the source domain s and target domain t, respectively. Meanwhile, x^s and x^t are separately selected from their optional choices $\mathbb{O}^s$ and $\mathbb{O}^t$, i.e., $x^s \in \mathbb{O}^s = \{o_1^s, \cdots, o_{M_s}^s\}$ and $x^t \in \mathbb{O}^t = \{o_1^t, \cdots, o_{M_t}^t\}$, where M_s and M_t are numbers of choices in the domains of s and t, respectively. The action sequence $\mathbf{x}^{(\cdot)} = (x_1^{(\cdot)}, x_2^{(\cdot)}, \cdots, x_N^{(\cdot)})$ of an individual refers to the actions taken in chronological order in a specific domain, where N is the number of actions. In this paper, $\mathbf{x}^{(\cdot)}$ is the tags of data ranked by the viewing time by users in a specific domain.

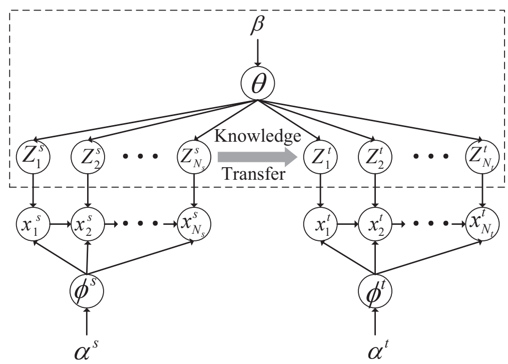

Figure 1: A graphical representation of our general novelty seeking model.

Dynamic Choice Novelty(DCN): Given an arbitrary domain, DCN [4] is a $N \times M$ matrix, where N is the length of action sequences and M denotes the number of choices in such domain. Every element in DCN is an integer in $[1, M]$. The DCN is used to present partial orders among M choices at each position. The DCN for a given domain s for example can be computed according to the following principle

$$DCN^s_{x_i^s} \propto \frac{1}{(\sum\limits_{\mathbb{K}_i^s \in x_i^s} \#\mathbb{K}_i^s + 1) \cdot (\sum\limits_{\substack{\mathbb{K}_{i-1}^s \in x_{i-1}^s \\ \mathbb{K}_i^s \in x_i^s}} T_{\mathbb{K}_{i-1}^s \mathbb{K}_i^s} + 1)} \tag{1}$$

where x_i^s and x_i^t respectively denote the observed actions for the i-th position in the domains of s and t, $\mathbb{K}_i^s$ is a keyword of keywords in the tags of x_i^s and $\#\mathbb{K}_i^s$ refers to the frequency of keyword in x_i^s before the i-th position in this individual's sequence. It measures the popularity of keyword at that moment in view of this individual's history behaviors. $T_{\mathbb{K}_{i-1}^s \mathbb{K}_i^s}$ refers to the transition probability of keyword in $x_{i-1}^s \rightarrow x_i^s$ before i-th position in this individual's sequence, which measures the keyword transition popularity at the moment in view of this individual. The notation for the target domain $DCN_{x_i^t}^t$ can be obtained by the similar way.

In the following, we detail the proposed model, which is inspired by the recent NSM model [4]. In CDNST, we extend the framework of NST to multiple domains and give its graphical model as Fig. 1. As shown in Fig. 1, z_i^s is the latent variable that represents the novelty-seeking level at the i-th position in the source domain s. Similarly, z_i^t denotes the corresponding latent variable for the target domain. Both of them are sampled from a *shared* multinomial novelty-seeking distribution θ. In addition, we use latent variables $\phi^s = \{\phi_1^s, \phi_2^s, ..., \phi_{M_s}^s\}$ and $\phi^t = \{\phi_1^t, \phi_2^t, ..., \phi_{M_t}^t\}$ to represent the utility of each choice in the domains of s and t, respectively. They can be interpreted by this individual's preference for each choice in the corresponding domain. Furthermore, α^s, α^t and β are the relevant hyper-parameters to ϕ^s, ϕ^t and θ, respectively. The probability value of x_i^s relies on the novelty-seeking level at i-th position, namely z_i^s, the choice utility distribution ϕ^s, and the previous chosen action which is considered in the $DCN_{x_i^s}^s$. The generation process of x_i^t is similar to x_i^s but relies on the corresponding variables in the target domain.

3. EXPERIMENTS

We crawled data of Movie, Music and Book from Douban website, and extracted the registered users who perform sequential behaviors on at least two domains. Finally we constructed 6 transfer learning recommendation problems (i.e., 3 pairs of data sets). We conduct extensive experiments to demonstrate the effectiveness of the proposed CDNST, and find that the temporal property of sequential data affects the performance of CDNST.

Table 1: Recommendation Performance

		OF(OF_U)	MC(MC_U)	NSM(NSM_U)	CDNST
$A \rightarrow B$	MRR	0.1601(0.1522)	0.2015(0.1779)	0.3128(0.3017)	**0.3623**
	nDCG@15	0.2153(0.2047)	0.2677(0.2299)	0.3821(0.3673)	**0.4363**
	p@3	0.1044(0.0937)	0.1409(0.1203)	0.2822(0.2736)	**0.3325***
$B \rightarrow A$	MRR	0.3982(0.2413)	0.4135(0.2575)	**0.5644**(0.3180)	0.5014
	nDCG@15	0.4998(0.3279)	0.5125(0.3715)	**0.6489**(0.3945)	0.5687
	p@3	0.3373(0.2100)	0.3649(0.2241)	**0.5488**(0.2992)	0.4797

OF (Order by Frequency) gives a recommendation list according to the frequency in the individual's historical behavior sequence, while for OF_U we compute the frequency in both source and target domains. MC (Markon Chain) [2] models sequential behaviors in target domain by learning a transition graph and performing predictions (In our report, the factorization was set at 20 for comparison), NSM (Novel Seeking Model) [4] is a data-driven model to predict the behaviour (In our report, the novelty seeking level was set at 9). MC_U and NSM_U are similarly defined with OF_U.

For all compared methods, they give a list of recommended choices with prediction probabilities, according to which we sort the candidate choices in descending order. In our experiments, the widely used evaluation metrics of nDCG, MRR and Precision are adopted to evaluate the performance of all algorithms.

These three pairs of data sets are divided into two groups, i.e., the pair of $A \rightarrow B$ and $B \rightarrow A$ are put into different groups (A and B represent two domains), and all the evaluation results are shown in Table 1. Here $A \rightarrow B$ contains three different groups, i.e., *music* $\rightarrow$ *book*, *music* $\rightarrow$ *movie*, *book* $\rightarrow$ *movie*. While $B \rightarrow A$ is a mirror reflection of $A \rightarrow B$ which means it reverses each domain in $A \rightarrow B$. In Table 1, each value is the mean of corresponding three different groups, and the best methods are marked in boldface. From these results, we have several insightful observations.

For the first group namely $A \rightarrow B$, our model incorporates the information from auxiliary source domain and achieves significant improvements compared with all baselines. Other approaches, however, cannot benefit from the data from source domain. That is OF_U, MC_U and NSM_U have no improvements compared with OF, MC and NSM. For baselines, NSM performs better than MC, and MC outperforms OF.

For the second group $B \rightarrow A$, we surprisedly find that all compared methods cannot obtain improvements by incorporating the auxiliary information from source domain. More worse the performances of them degrade dramatically. After analyzing the data, we conjecture that the sequential property of auxiliary domain data may affect the performance.

Overall, the results in Table 1, $A \rightarrow B$ implies that the domains of Music and Book can help learn the model on Movie domain, and Music can help the learning of Book. To intuitively show the temporal property of auxiliary domain data may affect the performances of all algorithms, we carefully investigate the characteristics of data set "Music $\rightarrow$ Movie", and find some interesting phenomenons. For example, 1) One user first listened a song of the Chemical Brothers[2] at time 2012/03/21 in the source domain, then later he would watch the movie of "The Chemical Brothers: Don't Think (2012)" at time 2012/04/04 in the target domain; 2) The user first listened the theme of "An Inaccurate Memoir" composed by Pong Nan at time 2012/04/04, then he would watched the movie of the theme at time 2012/05/04; 3) The user listened to the music of Leslie Cheung at time 2012/04/30, then later he would watch the movie "Farewell My Concubine"[3] with the player Leslie Cheung at time 2012/05/30.

These examples may imply that given the source domain data Music, we can transfer the information to give better recommendation on target domain Movie. However in reverse, if we use Movie as source domain, which may not provide useful information for the recommendation on Music, since the related behaviors in Movie occur after the related ones in Music. Even worse, Movie may become noise data, which leads to the performance degrading. The experiments showed that our model is more accurate, when the source and target domain data satisfy the sequential property, i.e., the related behaviors in source domain occur before the related ones in target domains. Since the sequential property of auxiliary data affect the transfer performance, thus we call the studied problem as sequential transfer learning. In the future, we will also aim to propose new transfer recommendation model to address this problem.

4. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61473273, 61602438, 91546122, 61573335), Guangdong provincial science and technology plan projects (No. 2015 B010109005) and 2015 Microsoft Research Asia Collaborative Research Program.

5. REFERENCES

[1] R. M. Furr and D. C. Funder. Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches âĭ. *Journal of Research in Personality*, 2004.

[2] A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic Probabilistic Systems*, 1971.

[3] H. Stern. The significance of impulse buying today. *The Journal of Marketing*, 1962.

[4] F. Zhang, N. J. Yuan, D. Lian, and X. Xie. Mining novelty-seeking trait across heterogeneous domains. In *WWW*, 2014.

[2] https://en.wikipedia.org/wiki/The_Chemical_Brothers. Electro and Techno are members of the Chemical Brothers.

[3] https://en.wikipedia.org/wiki/Farewell_My_Concubine_(film)