How to Interpret Implicit User Feedback?

Ladislav Peska, Peter Vojtas
Faculty of Mathematics and Physics, Charles University in Prague
[peska|vojtas]@ksi.mff.cuni.cz

ABSTRACT
Our research is focused on interpreting user preference from his/her implicit behavior. There are many types of relevant behavior e.g. time on page, scrolling, clickstream etc. which we will further denote as *Relevant Behavior Types* (RBT). RBT s varies both in quality and incidence and thus we might need different approaches to process them. In this early work we focus on how to derive user preference from each RBT separately. We selected number of common indicators, design two novel e-commerce specific RBT interpreting methods and conducted series of off-line experiments. After the off-line evaluation an A/B test on the real-world users of a travel agency was conducted comparing best off-line method with simple binary feedback. The experiments, although preliminary, showed importance of considering multiple RBTs together.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval - *Information Filtering*

General Terms
Measurement, Human Factors, Experimentation.

Keywords
Implicit Feedback, Recommender Systems, User preference

1. INTRODUCTION
Recommender Systems have been widely studied in the last two decades. They successfully complement search engines or on-site catalogues on video streaming services, book databases[1], e-commerce[2] etc. Although the recommender systems are relatively widespread nowadays, we focus on yet neglected domain: recommending on small e-commerce websites without dominant position on the market. Among the most sewer challenges of this domain is users' disloyalty and high ratio between number of objects and users. This disqualifies otherwise successful *Collaborative Filtering* (CF) methods as they stuck in persistent cold-start problem [4]. Another related challenge is the scarcity of explicit feedback. Due to users disloyalty and absence of incentives to do so, users generally do not provide explicit feedback in small e-commerce websites. Our only option is to focus on implicit user feedback. Unlike e.g. Hu et al. [2], we focus on multiple *behavior types* relevant for user *preference* (e.g. time on page, scrolling, purchases etc.), which will hopefully provide better user understanding and thus better recommendations than a single type of feedback. In our previous works we focused on deriving negative preference from implicit feedback [6] or various approaches to combine RBTs e.g. [7]. Our current research switched towards correct interpretation of RBT values as the first step towards learning user preference. We can track several similar approaches in the literature e.g. [1] comparing implicit signals with explicit user rating on an open-web user study, [8] categorizing several user activities as positive or negative feedback on an online music service, RSS feed recommender analyzing implicit reading-related user actions [3] or using normalized item level dwell time as relevance measure [9]. However to our best knowledge, there are no approach in the literature focusing on interpreting RBTs in small e-commerce and thus our set of RBTs and methods for their interpretation based on purchasing behavior are rather unique.

[1] Librarything.com

[2] Amazon.com

2. IMPLICIT PREFERENCE INDICATORS
Virtually any observable user behavior can serve as implicit feedback. The majority of user behavior consists of separated user actions (mouse click, typing, scrolling event etc.). Although it is possible to consider these actions as a stream, we opted for aggregating the same type of behavior while user visits particular webpage. Thus the *Relevant Behavior Types* (RBTs) are integer variables containing volumes of each type of action aggregated throughout user's visit of the webpage. So far we considered only several basic types of behavior as shown in Table 1, however we plan to use the full scope of the RBT collecting component [5] in the future work. Note that not all RBTs are triggered for all visits.

Table 1: Considered RBTs. Coverage column describes for how many visits we have also information from this RBT.

RBT	Triggered event	Coverage
Pageview	JavaScript *Load()*	99%
Mouse	JavaScript *MouseOver()*	44%
Scroll	JavaScript Scroll()	49%
Time	Total time spent on page	69%
Purchase	Object was purchased	0.5%

3. PREFERENCE LEARNING METHODS
The key research question of this poster is to learn dependence between values of each RBTs and user preference $\bar{r}$. It is possible to use simple binary model like "all visited objects are equally preferred" or simple numeric model with linear dependence between the value of RBT and user preference [2]. We added two collaborative approaches specific for the e-commerce domain, considering other users purchasing behavior.

Binary user preference is defined as $\bar{r} = 1$ for all visited objects.

Direct preference normalization is a user-wise linear normalization of each indicator into the [0,1] interval. This approach is similar to [2]. For user *u* and type *t*, the preference based on type is:

$$\bar{r_t} = val(RBT) \;/\; argmax_u\big(val(RBT)\big)$$

Purchase-based approaches considers whether other users with similar values of RBTs purchased the object or not and computes purchase rate *PR.* The approaches differ in definition of neighborhood ε for RBT values:

- *KNN*: use K nearest neighbor visits to compute *PR*. K is defined as ε * total number of all visits
- *Distance:* use all visits from interval [(1-ε) * *val(RBT)*, (1+ε) * *val(RBT)*]

The *PR* is then computed as: $PR = \#purch/(\#all\_purch * \varepsilon)$, where *#purch* is volume of purchases from defined ε neighborhood, *#all_purch* is volume of all purchases in the dataset. Intuitively *PR* for *KNN* represents ratio between mass of purchases in current interval and expected one for uniform distribution. Finally we use *PR* in sigmoid function to smoothly normalize user rating into [0,1] interval: $\overline{r_t} = 1/(1 + e^{-PR+1})$.

The hypothesis behind purchase-based approaches is that purchase is the only RBT with "guaranteed" effect on user preference, so if users evaluate other objects similarly, then although they did not purchase them, they still probably like them. Another reason for this approach is that although we can expect that higher value of each RBT implicates higher preference, the exact dependence is unknown. Purchasee-based approaches allow us to derive non-linear parametric dependence between the value of RBT and expected user preference $\bar{r}$. On the other hand in this approach we neglect different behavior patterns for different users as well as various cognitive demands to evaluate different objects. We would like to perform user clustering of more loyal users with enough feedback in the future work.

4. EVALUATION

4.1 Off-line Evaluation

In the first phase of evaluation we compared various RBT interpretation methods on a travel agency dataset. As we did not consider any specific method for aggregating RBTs, we opted for pairwise comparison of purchased and non-purchased objects for each user. For each (strictly rated) pair and each indicator preference $\overline{r_t}$ we state that this pair is correctly ordered, if the indicator preference of purchased object $\overline{r_{t,p}}$ is greater than preference of non-purchased object $\overline{r_{t,n}}$. Incorrect and equal are defined likewise. Let *Corr*/*Inc*/*Eq* are sums of all correctly/incorrectly/equally ordered pairs. We can now define *paired error metric* as follows:

$$pair_{err} = \alpha * Corr/All + (1 - \alpha) * (Corr + Eq)/All$$

Note that we consider *Eq* as an error too, however its significance is lower than *Inc.* We use α=0.5 in the evaluation.

The evaluation dataset contains 9 months usage data from a travel agency. For the purpose of the experiment, the dataset was restricted to only users with at least one purchase and at least two visited objects (some outliers were also removed) leaving over 8400 pairs of objects from 380 users with 450 purchases.

Table 2: Off-line results of $\boldsymbol{pair_{err}}$ for various values of ε.

RBT	Direct	Dist, 0.2	Dist, 0.9	KNN, 0.01	KNN, 0.7
Pageview	0.797	0.695	**0.850**	0.753	0.825
Mouse	0.772	0.561	0.799	0.695	**0.822**
Scroll	0.569	0.555	0.578	**0.582**	0.573
Time	**0.791**	0.502	0.589	0.632	0.649

According to the off-line evaluation, each RBT needs to be treated differently. As for the *Time*, the best method was *direct normalization* for *Mouse* it was *KNN* with larger ε, for *Pageview* was optimal *Distance* with large ε and *Scrolling* on the other hand requires either KNN with small ε. For *Distance* method, we can clearly see grading improvements with increasing ε for all RBTs, The *KNN* has peak performance around ε=0.7 for all RBTs except scrolling.

4.2 A/B testing

After the off-line evaluation we selected 3 methods for on-line testing: *Binary user preference* as baseline, average of *direct normalization* of all RBTs and average of *best* resulting methods for each RBT according to Table 2. The recommendations were computed via VSM algorithm and we opted for the number of click throughs (CT) as target metric. The evaluation was carried out in June 2015 with in total over 2900 users randomly assigned to one of the preference learning methods.

Table 3: Results of on-line evaluation

	Binary	Direct norm.	Best according to Table 2
CT	208	**232**	213
Users	971	979	976

The on-line experiments are not conclusive yet, but it seems that direct normalization outperforms other methods. We will further experiment with other settings and definitions of purchase-based approaches in the future work. Our current working hypothesis is to use exact values instead of ε expressions.

5. CONCLUSIONS AND FUTURE WORK

In this poster, our aim was to design novel methods to infer user preference from relevant behavior types and to determine optimal approaches to handle different RBTs. The purchase-based methods succeeded in off-line experiments, however further tuning and enhanced on-line evaluation is necessary. Also incorporation of various aggregation methods is in our future work.

Acknowledgements: The work on this paper was supported by the grant SVV-2015-260222, GAUK-126313 and P46.

REFERENCES

[1] Claypool, M.; Le, P.; Wased, M. & Brown, D.: Implicit interest indicators. *In IUI 2001. ACM*, **2001**, 33-40.

[2] Hu, Y.; Koren, Y.; & Volinsky, Ch.: Collaborative Filtering for Implicit Feedback Datasets. *In ICDM '08*. IEEE, 263-272.

[3] Lai, Y., Xu, X., Yang, Z., Liu, Z. User interest prediction based on behaviors analysis. *Int. Jour. of Digital Content Technology and its Applications*, 6 (13), **2012**, 192-204

[4] Peska, L. & Vojtás, P.: Recommending for Disloyal Customers with Low Consumption Rate. *In SOFSEM 2014, Springer, LNCS 8327*, **2014**, 455-465

[5] Peska, L.: IPIget – The Component for Collecting Implicit User Preference Indicators. *In ITAT 2014, Ustav informatiky AV CR*, **2014**, 22-26, http://itat.ics.upjs.sk/workshops.pdf .

[6] Peska, L. & Vojtas, P.: Negative Implicit Feedback in E-commerce Recommender Systems. *In WIMS 2013, ACM,* **2013**, 45:1-45:4

[7] Peska, L. & Vojtas, P.: Evaluating Various Implicit Factors in E-commerce. *In RUE 2012, CEUR,* **2012**, *910*, 51-55

[8] Yang, B.; Lee, S.; Park, S. & Lee, S.: Exploiting Various Implicit Feedback for Collaborative Filtering. *In WWW 2012,* ACM, **2012**, 639-640.

[9] Yi, X.; Hong, L.; Zhong, E.; Liu, N.; & Rajan. S.: Beyond Clicks: Dwell Time for Personalization. *In RecSys '14. ACM*, **2014**, 113-120