

Multimodal Implicit Feedback for Recommender Systems

Ladislav Peska

Department of Software Engineering, Faculty of Mathematics and Physics
Charles University, Prague, Czech Republic

Abstract. *In this paper, we present an overview of our work towards utilization of multimodal implicit feedback in recommender systems for small e-commerce enterprises. We focus on deeper understanding of implicit user feedback as a rich source of heterogeneous information. We present a model of implicit feedback for e-commerce, discuss important contextual features affecting its values and describe ways to utilize it in the process of user preference learning and recommendation. We also briefly report on our previous experiments within this scope and describe a publicly available dataset containing such multimodal implicit feedback.*

1 Introduction

Recommender systems belong to the class of automated content-processing tools, aiming to provide users with unknown, surprising, yet relevant objects without the necessity of explicitly query for them. The core of recommender systems are machine learning algorithms applied on the matrix of user to object preferences. In large enterprises, user preference is primarily derived from explicit user rating (also referred as *explicit feedback*) and collaborative-filtering algorithms [11] usually outperforms other approaches [3].

In our research, however, we focus on small or medium-sized e-commerce enterprises. This domain introduces several specific problems and obstacles making the deployment of recommender systems more challenging. Let us briefly list the key challenges:

- High concurrency has a negative impact on user loyalty. Typical sessions are very short, users quickly leave to other vendors, if their early experience is not satisfactory enough. Only a fraction of users ever returns.
- For those single-time visitors, it is not sensible to provide any unnecessary information (e.g., ratings, reviews, registration details).
- Consumption rate is low, users often visit only a handful (0-5) of objects.

All the mentioned factors contribute to the data sparsity problem. Although the total number of users may be relatively large (hundreds or thousands per day), explicit feedback is very scarce. Also the volume of visited objects per user is limited and utilizing popularity-based approaches w.r.t. purchases is questionable at best. Furthermore the identification of unique user is quite challenging.

Despite these obstacles, the potential benefit of recommender systems is considerable, it can contribute towards better user experience, increase user loyalty and consumption and thus also improve vendor's key success metrics.

Our work within this framework aims to bridge the data sparsity problem and the lack of relevant feedback by

modelling, combining and utilizing novel/enhanced sources of information, foremost various *implicit feedback* features, i.e., features based on the observed user behavior.

Contrary to the explicit feedback, usage of implicit feedback [5], [17], [18], [28] requires no additional effort from the users. Monitoring implicit feedback in general varies from simple features like user visits or play counts to more sophisticated ones like scrolling or mouse movement tracking [12], [29]. Due to its effortlessness, data are obtained in much larger quantities for each user. On the other hand, data are inherently noisy, messy and harder to interpret [10]. **Figure 1** depicts a simplified view of human-computer interaction on small e-commerce enterprises with accent on the implicit feedback provided by the user.

Our work lies a bit further from the mainstream of the implicit feedback research. To our best knowledge, the vast majority of researchers focus on interpreting single type of implicit feedback [6], proposing various latent factor models [10], [26], its adjustments [9], [19] or focusing on other aspects of recommendations using implicit feedback based datasets [2], [25]. Also papers using binary implicit feedback derived from explicit user rating are quite common [16], [19].

In contrast to the majority of research trends, we consider implicit feedback as multimodal and context-dependent. As our aim in this direction is a long-term one, we already published some of our findings [17], [18], [20], [22], [23]. In our aim towards improving recommender systems on small e-commerce enterprises, we focused on following aspects of implicit feedback:

- Cover the *multimodality of implicit feedback*.
- Propose relevant *context of collected feedback*.
- Derive models of *negative preference* based on implicit feedback

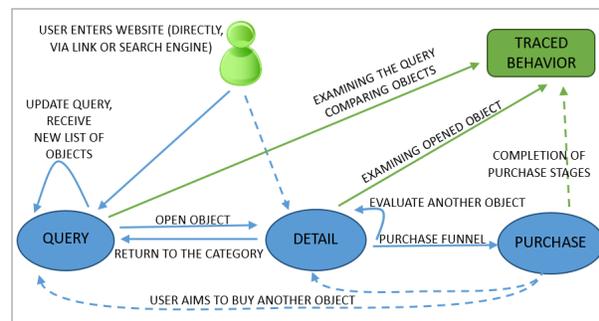


Figure 1: Simplified state diagram of human-computer interaction in e-commerce: User enters the site via *query* or object's *detail*. He/she can navigate through category or search result pages implicitly updating his/her query, or proceed to evaluate *details* of selected objects and eventually execute steps to buy them.

We reserve Section 2.1 to the description of multimodal implicit feedback, Section 2.2 to the contextualization of user feedback and Section 2.3 to the problem of learning negative preference. For each problem, we describe relevant state of the art, current challenges as well as our proposed methods and models. Finally, we remark on the evaluation of proposed methods in Section 3 and conclude in Section 4.

2 Materials and Methods

2.1 Multimodal Implicit Feedback

Despite the large volume of research based on a single implicit feedback feature, we consider implicit feedback to be inherently multimodal. Users utilize various I/O devices (mouse, keyboard) to interact with different webpage’s GUI controls, so there is an abundant amount of potentially interesting user actions. As the complexity of such environment is overwhelming, we imposed some restrictions:

- Limit to the feedback related directly to some specific object, i.e., collect the feedback only from the object’s detail page.
- Aggregate the same types of user actions on per session and per object basis.
- Focus only on user actions which can be numerically aggregated, i.e., the desired feedback features have numerical domain.

See **Figure 2** for an example of feedback features derived from user actions. In the following experiments, we consider these implicit feedback features¹:

- Number of *views* of the page
- *Dwell Time* (i.e., the time spent on the object)
- Total *distance* travelled by the *mouse* cursor.
- Total *mouse* in motion *time*.
- Total *scrolled distance*.
- *Scroll Time* (i.e., the time spent by scrolling)
- *Clicks count* (i.e., the volume of mouse clicks)
- *Purchase* (i.e., binary information whether user bought this object).

Although multimodal implicit feedback is not a mainstream research topic, we were able to trace some research papers. One of the first paper mentioning implicit feedback was Claypool et al. [5], which compared three implicit preference indicators against explicit user rating. More recently Yang et al. [29] analyzed several types of user behavior on YouTube. Authors described both positive and negative implicit indicators of preference and proposed linear model to combine them. Also Lai et al. [12] work on RSS feed recommender utilizing multiple reading-related user actions.

However, the lack of publicly available datasets containing multimodal implicit feedback significantly hinders advance of the area. Some work towards bridging

¹ Please note that the dataset used for the experiments contains also other feedback features such as number of page prints, followed links count, several non-numeric feedback features etc. These features seemed not relevant for the current task, however they may be utilized in the future.

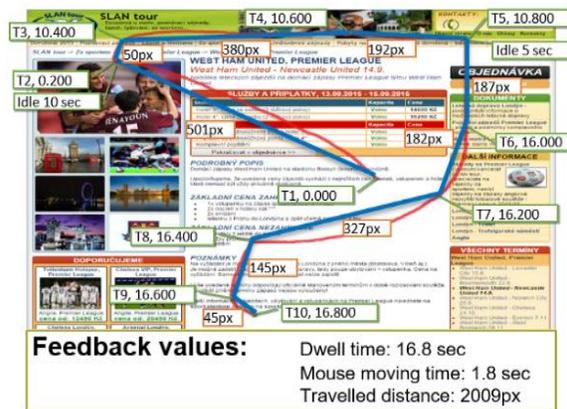


Figure 2: An example of mouse movement-based feedback collection on an e-commerce product detail page. Cursor positions (red line) are sampled periodically. Based on the samples, approximated mouse in-motion time (green boxes) and travelled distance (blue line) are calculated. Cursor motion log is also stored for later reasoning.

this data gap was done quite recently in RecSys Challenges 2016 and 2017². Both challenges’ datasets focused on job recommendation and contained several types of positive and negative user feedback. Although the dataset was not made publicly available, some approaches proposed relevant methods to deal with multimodal implicit feedback, e.g., fixed weighting scheme [31], hierarchical model of features [15] or utilizing features separately [28]. Some authors also mention the probability of re-interaction with objects on some domains [4], [33].

2.1.1 Methods Utilizing Multimodal Feedback

Vast majority of the state-of-the-art approaches transforms multimodal implicit feedback into a single numeric output \bar{r} , which can be viewed as a proxy towards user rating. However, these methods mostly use some fixed model of implicit feedback (i.e., predefined weights or hierarchy of feedback features), or perform predictions based on each feedback feature separately [28].

In contrast to the other approaches, we aim on estimating \bar{r} via machine learning methods applied on a purchase prediction task. Our approach is based on the fact that the only measurable implicit feedback with direct interpretation of preference is buying an object. Such events are however too scarce to be used as a sole user preference indicator. However, we can define a classification task to determine, based on the values of other feedback types, whether the object will be purchased by the user. The estimated rating \bar{r} is defined as the probability of the *purchased* class.

We evaluated several machine learning methods, such as decision trees, random forests, boosting, lasso regression and linear regressions. We also evaluated approaches based on *the more feedback the better* heuristics, i.e., the higher value of particular feedback feature implies higher user preference. In order to make the domains of all feedback features comparable, we utilized either standardization of

² [2016]2017.recsyschallenge.com

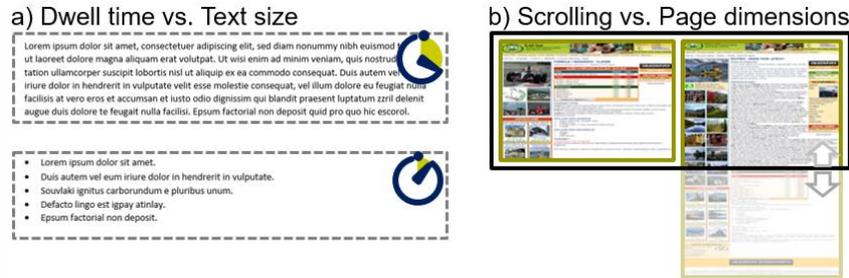


Figure 4: Two examples of relevant presentation context. A: overall length of text (and possibly also amount of other objects, e.g., images) affects the necessary reading time accessed through the *dwell time* feedback feature. B: the difference between page dimensions and device displaying size affects the necessity to scroll the content and thus, e.g., *scrolled distance* feature.

feedback values (denoted as *Heuristical with STD* in results), or used empirical cumulative distribution instead of raw feedback values (*Heuristical with CDF*). The estimated rating \bar{r} is defined as the mean of *STD* or *CDF* values of all feedback features for the respective user and object. For more details on heuristical approaches please refer to [23], for more details on machine learning approaches please refer to [18].

2.2 Context of User Feedback

Although the user feedback may be processed directly, the perceived feedback values are significantly affected by the presentation of the page (i.e., device parameters) and also by the amount of contained information. Both the displaying device and the page's complexity can be described by several numeric parameters, which we generally denote as the *presentation context*. See **Figure 4** for some examples of relevant presentation context.

We can trace some notions of presentation context in the literature. Yi et al. [30], proposed to use dwell time as an indicator of user engagement. Authors discussed the usage of several contextual features, e.g., content type, device type or article length as a baseline dwell time estimators. Furthermore, Radlinski et al. [24] and Fang et al. [8] considered object position as a relevant context for clickstream events.

The *presentation context* differs significantly from more commonly used *user context* [1] in both its definition, methods for feature collection as well as ways to incorporate it into the recommending pipeline. While the nature of *user context* is rather restrictive [32], we interpret *presentation context* as a baseline predictor or an input feature for machine learning process.

In our work, we considered following *presentation context* features:

- *Volumes of text, images and links* on the page.
- *Page dimensions* (width, height).
- *Browser's visible area dimensions* (width, height).
- *Visible area ratio*.
- *Hand-held device indicator*.

We evaluated several approaches utilizing biases for feedback feature values based on the current presentation context. Such approaches, however, were not very successful and in particular often did not improve over the baselines without any context at all. On the other hand, we

significantly improve over the baseline methods while using presentation context features as additional input of the machine learning methods described in Section 2.1.1.

2.3 Negative Implicit Feedback and Preferential Relations

One of the open problems in implicit feedback utilization is learning negative preference from implicit feedback. Several approaches were proposed for this task including uniform negative preference of all unvisited objects [10], considering low volume of feedback as negative preference [20] or defining special negative feedback feature [13], [29], [31].

We propose to utilize negative preference as relations among less and more preferred objects, i.e. to model a partial ordering $o_1 <_p o_2$. This model is based on the



Figure 3: An example of preference relations based on the feedback on a category page. Initially, objects $o_1 - o_4$ are visible. After some time, user scrolls and also object o_5 gets to the visible window. However, objects o_6 and o_7 remains outside of the visible area. If user clicks on object o_3 , his/her behavior induces negative feedback on objects $o_{1,2,4,5}$ and we collect relations $o_{1,2,4,5} <_p o_3$. However, the intensity of $o_5 <_p o_3$ is smaller than for the other objects, because o_5 was visible only for a short period of time and thus it is more probable that user did not notice it.

work of Eckhardt et al. [7] proposing to consider user ratings (implicit feedback in our scenario) in the context of other objects available on the current page. Implicit preferential relations can be naturally obtained from implicit feedback collected on category pages, search results or similar pages. In such cases, user usually selects one (or more) objects out of the list of available objects for further inspection. By this behavior, user also implicitly provides negative feedback on the ignored choices and thus induce a preferential relation $ignored \prec_p selected$. However, we need to approach to such negative feedback with caution as some of the options might not be visible for the user at all, or only for a very short time. This is quite serious problem, because, in average, only 47% of the catalogue page content was visible in the browser window in our dataset. Thus, we also introduce intensity of the relation \prec_p based on the visibility of the ignored object.

Figure 3 illustrates this.

We incorporate preferential relations into the recommendation pipeline by extending collected relations along the content-based similarity of both *ignored* and *selected* objects (decreasing level of similarity effectively decreases also the intensity of the relation). Afterwards, we apply re-ranking approach taking output of some baseline recommender and re-order the objects so that the relations with higher intensity holds. Re-ranking algorithm considers the relations according to the increasing intensity and corrects the ordering induced by the relation. Thus, more intense relations should be preferred in case of conflicts. Details of the re-ranking algorithm can be found in [22].

3 Evaluation

In this section, we would like to report on the experiments conducted to evaluate models and methods utilizing multimodal implicit feedback. However, let us first briefly describe the dataset and evaluation procedures.

3.1 Evaluation Procedure

The dataset of multimodal user feedback (including presentation context) was collected by observing real visitors of a mid-sized Czech travel agency. The dataset was collected by the IPIget tool [17] over the period of more than one year, contains over 560K records and is available for research purposes³. In addition to the feedback features, the dataset also contains several content-based attributes of objects and thus enables usage of content-based recommender systems as well.

In evaluation of the methods, we considered following tasks:

- *Purchase prediction* based on the other feedback available for particular user-object pair. This scenario provides preliminary results for the methods aiming to estimate user rating \bar{r} .
- *Recommending purchased objects*. In this scenario, we employ leave-one-out cross-validation protocol on purchased objects (i.e., for each purchased object, all other feedback is used as a train set and we aim to recommend object, which was actually purchased by the user).
- *Recommending “future” user actions*. In this scenario, we use older user feedback (usually 2/3 of available feedback per user) as a train set. During the recommendation phase, we recommend top-k objects to each user, while the objects from the test set visited by the user should appear on top of the list.

In several of our previous works (see, e.g., [21] or the results of Matrix Factorization [11] in Table 3) we have shown that purely collaborative methods are not very suitable for small e-commerce enterprises due to the ongoing cold-start problem. Thus, we mostly focus on the content-based and hybrid recommending techniques. More specifically, we utilized Vector Space Model (VSM) [14],

Table 1: Results (nDCG) of *purchase prediction task* based on multimodal implicit feedback and presentation context. The task was considered as ranking (i.e., purchased objects should appear on top of the list of all visited objects).

Method	Dwell Time	Multimodal feedback	Feedback + Context	Feedback + aggregated BP	Feedback + individual BP
Heuristical with CDF	0.663	0.712	0.780	0.696	0.690
Heuristical with STD	0.747	0.703	0.856	0.695	0.704
Linear Regression	0.747	0.789	0.917	0.804	0.925
J48 decision tree	0.663	0.722	0.908	0.839	0.876

Table 2: Results (nDCG) of *recommending purchased objects task*. Combination of most-popular and VSM recommender was used to derive the list of objects. *Aggregated BP* denotes baseline predictors aggregated for a particular feedback feature over all available context, *individual BP* introduces a baseline predictor for each pair of contextual and feedback feature, *Feedback + Context* treats contextual features as additional input of the methods estimating \bar{r} .

Method	Binary Feedback	Dwell Time	Multimodal feedback	Feedback + Context	Feedback + aggregated BP	Feedback + individual BP
Heuristics with CDF		0.255	0.257	0.253	0.258	0.257
Heuristics with STD		0.208	0.174	0.196	0.161	0.158
Linear Regression	0.255	0.256	0.254	0.176	0.252	0.251
J48 decision tree		0.238	0.256	0.273	0.240	0.248

³<http://bit.ly/2tWtRg2>

Table 3: Results (nDCG) of *recommending future interactions* task with re-ranking based on the preferential relations.

Method	nDCG
VSM + Preferential Relations	0.4381
VSM	0.4376
Popular SimCat + Preferential Relations	0.3982
Popular SimCat	0.3962
Matrix Factorization + Preferential Relations	0.220
Matrix Factorization	0.138

its combination with the most popular recommendations and a hybrid algorithm proposing most popular objects from the categories similar (based on collaborative filtering) to the visited ones [22]. As we consider recommending problem as a ranking optimization, all methods were evaluated w.r.t. normalized discounted cumulative gain (nDCG).

3.2 Results and Discussion

Results of several methods aiming to learn estimated rating \bar{r} based on various feature sets are displayed in Table 1 (purchase prediction task) and Table 2 (recommending purchased objects task). As we can see in Table 1, multimodal feedback significantly improves purchase prediction capability across all methods. Usage of presentation context can further improve the results, if used as additional input feature (*Feedback + Context*). However, if the contextual features are used as baseline predictors, the results across all methods are inferior to the results of *Feedback + Context* with just one exception. In several cases, the results are worse than using multimodal feedback alone. This observation indicates that some more complex dependence exists between implicit feedback, presentation context and user preference. Although it seems that the examined machine learning methods can partially discover this relation, another option to try is to hand-pick only several relevant contextual scenarios instead of the global model applied so far. Results of recommendation task also revealed a potential problem of overfitting on the purchase prediction task. *Linear regression*, although it performed the best in purchase prediction scenario, did not improve over binary feedback baseline. On the other hand, we can conclude that if a suitable rating prediction is selected, multimodal implicit feedback together with presentation context can improve the list of recommended objects.

Results of re-ranking approach based on preferential relations are depict in **Table 3**. Re-ranking based on preferential relations improved results of all evaluated recommending algorithms, although the improvement was rather modest in case of VSM. During evaluation, we observed that in case of VSM, only the relations with highest intensity should be applied to improve the results. For matrix factorization approach, on the other hand, also relations with very low intensities should be incorporated. Another point is that the offline evaluation is naturally focused on mere learning past user behavior and both VSM and Popular SimCat are largely biased towards exploitation in exploration vs. exploitation problem [27]. Hence, there

may be further benefits of using preferential relations in online scenarios.

4 Conclusions

In this paper, we describe our work in progress towards utilizing multimodal implicit feedback in small e-commerce enterprises. Specifically, we focused on three related tasks. Integrate multiple types of feedback collected on the detail of an object into an estimated user rating \bar{r} , incorporate presentation context into the previous model and utilize negative implicit feedback collected on category pages. We propose models and methods for each of the task and also provide evaluation w.r.t. top-k ranking.

Although the proposed methods statistically significantly improved over the baselines, the relative improvement is not too large, so our work is not finished yet. One of the important future tasks is to perform online evaluation as the offline evaluation was focused on the exploitation only. Further tasks are to propose context incorporation models specific for some context-feedback feature pairs, explore other possibilities to incorporate negative feedback and also to evaluate unified approach integrating all presented methods.

Acknowledgment

The work on this project was supported by Czech grant P46. Source codes and datasets for incorporation of presentation context can be obtained from <http://bit.ly/2rjZzg3>, source codes of preferential relations approach can be obtained from <http://bit.ly/2symm17> and raw dataset can be obtained from <http://bit.ly/2tWtRg2>.

References

- [1] Adomavicius, G. & Tuzhilin, A. Context-Aware Recommender Systems. *Recommender Systems Handbook, Springer US*, **2015**, 191-226
- [2] Baltrunas, L. & Amatriain, X.: Towards time-dependant recommendation based on implicit feedback. *In CARS 2009 (RecSys)*.
- [3] de Campos, L. M.; Fernandez-Luna, J. M.; Huete, J. F. & Rueda-Morales, M. A. Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, **2010**, *51*, 785 – 799
- [4] Carpi, T.; Edemanti, M.; Kamberoski, E.; Sacchi, E.; Cremonesi, P.; Pagano, R. & Quadrana, M. Multi-stack Ensemble for Job Recommendation. *In proceedings of the Recommender Systems Challenge, ACM*, **2016**, 8:1-8:4
- [5] Claypool, M.; Le, P.; Wased, M. & Brown, D.: Implicit interest indicators. *In IUI '01, ACM*, **2001**, 33-40.
- [6] Cremonesi, P.; Garzotto, F.; Turrin, R.: User-Centric vs. System-Centric Evaluation of Recommender Systems. *In INTERACT 2013, Springer LNCS 8119*, **2013**, 334-351
- [7] Eckhardt, A.; Horvath, T. & Vojtas, P. PHASES: A User Profile Learning Approach for Web Search. *In WI-IAT '07, IEEE Computer Society*, **2007**, 780-783
- [8] Fang, Y. & Si, L. A Latent Pairwise Preference Learning Approach for Recommendation from Implicit Feedback. *In CIKM 2012, ACM*, **2012**, 2567-2570

- [9] Hidasi, B. & Tikk, D.: Initializing Matrix Factorization Methods on Implicit Feedback Databases. *J. UCS*, **2013**, 19, 1834-1853
- [10] Hu, Y.; Koren, Y. & Volinsky, C.: Collaborative Filtering for Implicit Feedback Datasets. *In ICDM 2008, IEEE* **2008**, 263-272
- [11] Koren, Y.; Bell, R. & Volinsky, C. Matrix Factorization Techniques for Recommender Systems. *Computer, IEEE Computer Society Press*, **2009**, 42, 30-37
- [12] Lai, Y., Xu, X., Yang, Z., Liu, Z. User interest prediction based on behaviors analysis. *Int. Journal of Digital Content Technology and its Applications*, 6 (13), **2012**, 192-204
- [13] Lee, D. H. and Brusilovsky, P.: Reinforcing Recommendation Using Implicit Negative Feedback. *In UMAP 2009, Springer, LNCS*, **2009**, 5535, 422-427
- [14] Lops, P.; de Gemmis, M. & Semeraro, G. Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook, Springer*, **2011**, 73-105
- [15] Mishra, S. K. & Reddy, M. A Bottom-up Approach to Job Recommendation System. *In proceedings of the Recommender Systems Challenge, ACM*, **2016**, 4:1-4:4
- [16] Ostuni, V. C.; Di Noia, T.; Di Sciascio, E. & Mirizzi, R.: Top-N recommendations from implicit feedback leveraging linked open data. *In RecSys 2013, ACM*, **2013**, 85-92.
- [17] 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>
- [18] Peska, L.: Using the Context of User Feedback in Recommender Systems. *MEMICS 2016, EPTSC 233*, **2016**, 1-12.
- [19] Peska, L.: Linking Content Information with Bayesian Personalized Ranking via Multiple Content Alignments. *To appear in HT 2017, ACM*, **2017**
- [20] Peska, L. & Vojtas, P.: Negative implicit feedback in e-commerce recommender systems *In WIMS 2013, ACM*, **2013**, 45:1-45:4
- [21] Peska, L. & Vojtas, P.: Recommending for Disloyal Customers with Low Consumption Rate. *In SOFSEM 2014, Springer, LNCS 8327*, **2014**, 455-465
- [22] Peska, L. & Vojtas, P. Using Implicit Preference Relations to Improve Recommender System. *J Data Semant*, 6(1), **2017**, 15-30.
- [23] Peska, L. & Vojtas, P. Towards Complex User Feedback and Presentation Context in Recommender Systems. *BTW 2017 Workshops, GI-Edition, LNI, P-266*, **2017**, 203-207
- [24] Radlinski, F. & Joachims, T. Query chains: learning to rank from implicit feedback. *In ACM SIGKDD 2005, ACM*, **2005**, 239-248
- [25] Raman, K.; Shivaswamy, P. & Joachims, T.: Online Learning to Diversify from Implicit Feedback. *In KDD 2012, ACM*, **2012**, 705-713
- [26] Rendle, S.; Freudenthaler, C.; Gantner, Z. & Schmidt-Thieme, L. BPR: Bayesian Personalized Ranking from Implicit Feedback. *In UAI 2009, AUAI Press*, **2009**, 452-461.
- [27] Rubens, N.; Kaplan, D. & Sugiyama, M. Active Learning in Recommender Systems. *In Recommender Systems Handbook, Springer US*, **2011**, 735-767
- [28] Xiao, W.; Xu, X.; Liang, K.; Mao, J. & Wang, J. Job Recommendation with Hawkes Process: An Effective Solution for RecSys Challenge 2016. *In proceedings of the Recommender Systems Challenge, ACM*, **2016**, 11:1-11:4
- [29] Yang, B.; Lee, S.; Park, S. & Lee, S.: Exploiting Various Implicit Feedback for Collaborative Filtering. *In WWW 2012, ACM*, **2012**, 639-640.
- [30] Yi, X.; Hong, L.; Zhong, E.; Liu, N. N. & Rajan, S. Beyond Clicks: Dwell Time for Personalization. *In RecSys 2014, ACM*, **2014**, 113-120.
- [31] Zhang, C. & Cheng, X. An Ensemble Method for Job Recommender Systems. *In proceedings of the Recommender Systems Challenge, ACM*, **2016**, 2:1-2:4
- [32] Zheng, Y.; Burke, R. & Mobasher, B. Recommendation with Differential Context Weighting. *In UMAP 2013, Springer*, **2013**, 152-164
- [33] Zibriczky, D. A Combination of Simple Models by Forward Predictor Selection for Job Recommendation. *In proceedings of the Recommender Systems Challenge, ACM*, **2016**, 9:1-9:4