Evaluating Context-aware Recommendation Systems Using Supposed Situation

Juraj VIŠŇOVSKÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
visnovsky.j@gmail.com

In the age of information overload we witness an increase the popularity of personalized systems. Among many solutions coping with content adaptations to users' needs, recommendation systems seem to stand above all. The main purpose of recommendation system is to deliver relevant information to the user and thus simplify his navigation in data overflow.

User's actions and decision, while using software system, are influenced by many factors. These factors, also known as contexts, describe the situation and the environment of the user. Personalized systems may collect a wide range of different types of context. Such may be user's current mood, health condition, activity, location, etc. Including contexts in the recommender system, makes it possible to generate more specific recommendations. This may lead to recommendation quality enhancement.

When evaluating or choosing one of several recommendation algorithms, we may employ one of three possible approaches: offline evaluation, online evaluation and user study [3]. Offline evaluation is easy and cheap to conduct, but its outputs are less reliable than outputs from online evaluation and user study [4]. Although, both user study and online evaluation are more accurate, they have one major drawback, which is their costly conduct [4]. In this paper, we focus on improving user study evaluation approach for context-aware recommendation systems.

To reduce the drawbacks of the user study approach (setting up environment according to contexts and gathering satisfying set of users) we propose using supposed situations. Our goal is to maximize the amount of evaluated recommendations.

Proposed evaluation approach requires input data of context-aware recommendations and user study participants. Firstly, we analyze recommendations data and identify its frequent patterns. At the same time user study participants are being modelled. Participant's model reflects test his ability to evaluate recommendations using supposed situations. We identified and analyzed two approaches to measure participants' reliability. A general approach is based on

---

*   Supervisor: Dušan Zeleník, Institute of Informatics and Software Engineering

measuring social perspective taking skills, which reflects the participant's tendency to adopt psychological perspectives of another person. Another, more specific approach is based on experiences of experiment participant.

We conducted a pilot experiment to evaluate reliability of user study participants. In this experiment we let six participants fill out a form defining their short- and long-term contexts. Then we addressed them Interpersonal reactivity index inquiry. Finally, we let the participants to rate twenty recommended items.

In the following step, we let a subset of the original set of participants to evaluate a random set of supposed situations, based on explicit feedback from participants acquired in the previous phase. Experiment participants evaluated a total of 87 supposed situations. We observed that good results acquired by inquiry measuring social perspective taking skills did not affect the ability of participants to correctly evaluate supposed situations. Then we examined effects of the rate of experience with presented contexts on evaluation accuracy. We observed that there is no correlation between these two variables.

In our next experiment we will introduce two approaches to measure experiment participant's reliability. We will adopt Gehlbach's inquiry [2], which outputs should be more reliable than the outputs of Interpersonal reactivity index inquiry [1], used in pilot experiment, as it is harder for the participants to sanitize its results. We also assume that higher number of participants will be beneficial as it will bring diversity.

Then, we will carry second experiment to evaluate proposed user study evaluation approach. As our goal is to improve user study coverage, we will examine the attributes of common user study approach with proposed user study experiment using supposed situations. Our attention will be paid mostly on the number of recommendations covered by each approach, but we will measure accuracy and F-measure as well. We expect our approach's accuracy to be slightly inferior to common user study experiment, but we believe in increase of F-measure.

References

[1] Davis, M.: A multidimensional approach to individual differences in empathy. In: *JSAS Catalog of Selected Documents in Psychology*, American Psychological Association, 1980, vol. 10, no. 4, pp .85.

[2] Gehlbach, H.: A New Perspective on Perspective Taking: A Multidimensional Approach to Conceptualizing an Aptitude. In: *Educational Psychology Review*, Kluwer Academic Publishers, 2004, vol. 16, no. 3, pp. 207–234.

[3] Shani, G., Gunawardana, A.: Evaluating recommendation systems. In: *Recommender systems handbook*, 2011, pp. 257–297.

[4] Shani, G., Gunawardana, A.: A survey of accuracy evaluation metrics of recommendation tasks. In: *The Journal of Machine Learning Research*, 2009, vol. 10, pp. 2935–2962.