2014 IEEE 14th International Conference on Advanced Learning Technologies

Layered Evaluation for Data Discovery & Recommendation Systems: an Initial Set of Principles

Nikos Manouselis
Agro-Know
Athens, Greece
e-mail: nikosm@ieee.org

Charalampos Karagiannidis
Department of Special Education
University of Thessaly
Volos, Greece
e-mail: karagian@uth.gr

Demetrios G Sampson
Department of Digital Systems
University of Piraeus & CERTH
Piraeus, Greece
e-mail: sampson@iti.gr

***Abstract*—This paper examines how a layered evaluation framework proposed for adaptive systems (AS) can be applied in the case of recommender systems (RecSys). Our analysis indicates that implementing a layered-based evaluation has the potential to facilitate a more detailed and informed evaluation of RecSys, allowing researchers and developers to better understand how to improve them.**

Keywords-recommender systems, layered evaluation, adaptive systems

I. INTRODUCTION AND RELATED WORK

A series of layered evaluation frameworks have been proposed in the AS literature, advocating that each different AS component should be evaluated separately in order to get detailed information on the pros and cons of each part of the system. Reviewing the evolution of the state of the art in 2010, Paramythis et al. grouped together the main approaches and suggested the following main layers of adaptation: (i) collection of input data; (ii) interpretation of the collected data; (iii) modeling the current state of the "world"; (iv) deciding upon adaptation; and (v) applying (or instantiating) adaptation [1].

A similar survey of evaluation issues in RecSys was carried out by Pu et al. [2]. This survey identified a generic interaction model for such systems that includes three crucial components that corresponded to groups of interaction activities between the user and the system: (i) the initial preference elicitation process, (ii) the preference refinement process, and (iii) the presentation of the system's recommendation results. This decomposition is very close to the way that layered evaluation frameworks are decomposing AS in separate components that can be evaluated one by one. Pu et al. suggested that layered evaluation can be used in RecSys research as a powerful technique in identifying areas of a system that require further improvements.

II. ADAPTING THE LAYERED EVALUATION FRAMEWORK

To illustrate how a layered de-composition can serve as a starting point for the development of a more concrete and practical evaluation framework, we elaborate on the mapping of the layers proposed by Karagiannidis & Sampson [3] to the RecSys components proposed by Pu et al. [2], in order to provide some generic principles and guidelines that a RecSys researcher could use. The aim of this exercise is to demonstrate how such a generic framework can be built using the specific evaluation layers (user model and recommendation system), rather than proposing a complete and detailed generic framework. It will help us inform the analysis of a specific case study in the section that follows.

More specifically, we focus on each layer and breakdown the interaction components to distinguishable elements – for example, separating the way that user preferences are elicited from the way that they are revised. Then, using an existing analysis of RecSys to various dimensions, we further analyze the interaction components to more fine-grained sub-components such as the way that the user model is being represented and the way that it is generated. This analysis can be adapted to the level of granularity that the evaluation framework designers believe that is necessary to provide meaningful results to the researchers. In Table 1, this is illustrated with a handful of RecSys dimensions from the many that the analysis framework of Manouselis & Costopoulou identifies [4].

After the specific sub-components have been identified, what is needed is a suggestion of appropriate evaluation methods, protocols, metrics and instruments. There are several RecSys evaluation studies that may be used as a source for this information. In Table 1 we present two specific attributes: the suggestion of an appropriate evaluation method and a corresponding metric that may be used based on Shani & Gunawardana [5]. Such a suggestion indicates that the representation of the user model may be evaluated by:

- running an offline experiment (as we demonstrate later in the case study) that will try to measure e.g. the utility of the engaged user model,
- a focused user trial that e.g. may ask users about their desired way of representing their preferences,
- or an online evaluation where e.g. two different methods for representing the user model may be compared through an A/B test.

In a similar way, the sub-components responsible for updating the user model, for generating the recommendations, or for presenting the recommendations, may be tested using different methods and several possible metrics that Shani & Gunawardana propose.

978-1-4799-4038-7/14 $31.00 © 2014 IEEE
DOI 10.1109/ICALT.2014.152

TABLE I EVALUATION LAYERS/COMPONENTS AND EVALUATION EXAMPLES/GUIDELINES

Evaluation layers *Karagiannidis & Sampson, 2000 [1]*	**Interaction components** *Pu et al., 2012 [2]*	**RecSys dimension** *Manouselis & Costopoulou, 2007 [4]*	**Evaluation Method** *Shani & Gunawardana, 2011 [5]*	**Evaluation Metric** *Shani & Gunawardana, 2011 [5]*
interaction assessment	*Elicit user preferences*	User Model Representation	Offline experiment, User study, Online evaluation	User Preference, Utility
		User Model Generation	Offline experiment, User study	User Preference, Utility
	Revise user preferences	User Model Update	Offline experiment, User study	Adaptivity
adaptation decision making	*Display recommendations*	Personalization algorithm	Offline experiment	Prediction Accuracy, Coverage, Robustness, Scalability
		Personalization output	User study, Online evaluation	User Preference, Trust, Novelty, Serendipity

III. CONCLUSIONS AND FUTURE WORK

In this paper we discuss the application of a layered evaluation framework for RecSys, revisiting past evaluation results through a layered evaluation view. Our analysis indicates that implementing a layered-based RecSys evaluation has the potential to facilitate a more detailed and informed evaluation of such systems, allowing researchers and developers to better understand how to improve them. Building upon this analysis we summarize below some directions of future work in the direction of developing more concrete layered evaluation approaches for RecSys.

Current layered frameworks offer suggestions of useful methods and instruments, but they are not specific enough to guide practical applications. It would be useful to have an out-of-the-box approach with pre-defined protocols and suggested experiments to carry out for each type of recommender system. Any additional support (such as decision trees) to help adopters chose the most appropriate tool for their system and setting would be of high value.

The five layers of Paramythis et al. [1] need to be better connected to the components identified by information filtering and RecSys studies, such as Hanani et al. [6] and Manouselis & Costopoulou [4]. This will help the layers become more specific in terms of RecSys aspects and dimensions that they focus on, also allowing the suggestion of various techniques for each dimension and type.

Even if a successful layered approach is applied to evaluate different components of a RecSys, the way that the results can be combined into a set of (possibly measurable) indicators has not been yet provided. Such a set of indicators could help the RecSys researcher or developer to decide on the trade-offs of devoting resources to improve one dimension over the other, and to have an overall and comparable view of the outcomes of an evaluation compared to a similar one.

There are several existing frameworks in AS that suggest relevant methods and techniques for the evaluation of each layer. In a similar sense, there are also surveys and studies of relevant evaluation approaches for recommender systems, including practical guidelines and recommendations. This existing body of knowledge should be carefully studied and combined in order to equip new frameworks with suggested methods, tools and instruments that would fit each component.

Overall, further work towards the standardization of the layered evaluation frameworks applied to RecSys should be expected to make it possible to facilitate the comparison and generalization of research results, and their reuse across different application domains.

ACKNOWLEDGEMENTS

The work of Nikos Manouselis has been funded with the support by European Commission, and more specifically the FP7 project SemaGrow "Data Intensive Techniques to Boost the Real-Time Performance of Global Agricultural Data Infrastructures" (http://semagrow.eu). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

BIBLIOGRAPHY

[1] A. Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods", User Modeling and User-Adapted Interaction, vol. 20, pp. 383-453, 2010.

[2] P. Pu, L. Chen, and R. Hu, "Evaluating recommender systems from the user's perspective: survey of the state of the art", User Modeling and User-Adapted Interaction, vol. 22, pp. 317-355, 2012.

[3] P. Brusilovsky, C. Karagiannidis, and D. Sampson, "Layered Evaluation of Adaptive Learning Systems", International Journal of Continuous Engineering Education and Lifelong Learning, vol. 14, pp. 402-421, 2004.

[4] N. Manouselis, and C. Costopoulou, "Analysis and Classification of Multi-Criteria Recommender Systems", World Wide Web, vol. 10, pp. 415-441, 2007.

[5] G. Shani, and A. Gunawardana, "Evaluating Recommendation Systems", in Recommender Systems Handbook, F. Ricci, L. Rokach, and B. Shapira, Eds. Springer, 2011, pp. 257-297.

[6] U. Hanani, B. Shapira, and P. Shoval, "Information Filtering: Overview of Issues, Research and Systems", User Modeling and User Adapted Interaction, vol. 11, pp. 203-259. 2001.