Harmonization and Categorization of Metrics and Criteria for Evaluation of Recommender Systems in Healthcare From Dual Perspectives

Adekunle Oluseyi Afolabi, University of Eastern Finland, School of Computing, Kuopio, Finland
Pekka Toivanen, University of Eastern Finland, School of Computing, Kuopio, Finland

ABSTRACT

Researchers’ choice of metrics and criteria in evaluating recommender systems depends on what the researcher feels is popular among other researchers, or sometimes based on the objective of the research. There is no harmonized set of criteria and metrics that can be referenced when evaluating recommender systems in healthcare. In this article, a set of metrics and criteria are harmonized and categorized as a guide for evaluating recommender systems. By means of an online survey, the opinions of forty-four experienced researchers and other stakeholders from eight countries and four continents were sought on the relevance of identified metrics and criteria. Analysis of the results show speed and timeliness are at the top. Topping the list of criteria is the provision of information that will guide users to useful decisions. The result is presented from two logical perspectives. Four categories are then identified as a useful guide for evaluating recommender systems.

KEYWORDS
Categorization, Criteria, Evaluation, Harmonization, Healthcare, Metrics, Recommender Systems

INTRODUCTION

Recommender systems have played an important role in people’s search for articles of interest, thus making searching an easy and enjoyable exercise. But more importantly, giving personalized recommendations is the beauty of any recommender system. Effectiveness of a recommender system in this regard essentially depends on the quality of recommendations obtained from it. This no doubt, and explains why many authors have placed high degree of importance on determining the quality of recommendations obtained. This can be judged using several metrics and criteria. In addition to using metrics and criteria, effectiveness can also be measured from users and system perspectives (Cremonesi et al., 2013). It is reasonable that, for an effective evaluation of recommenders systems to take place, this must be done from both perspectives. Evaluation of a recommender system is incomplete without evaluating user experiences in using a certain recommender system, no matter how
effective the system might seem from system’s perspective. This perspective will involve considering evaluation metrics. There have been attempts to produce an evaluation framework that combines system-centric and user-centric evaluation methods (Kavu et al., 2017). This seems to be a way of not leaving out any part in trying to get a clear picture of effectiveness or quality of a recommender system. From these perspectives, harmonized metrics and criteria for carrying out such an evaluation is desirable, especially if they can be categorized. It has been affirmed that performing an evaluation before deploying a recommender system is very essential (Fazeli et al., 2017). This therefore requires a set of metrics and criteria that will provide an agreeable result among relevant stakeholders. While our work focuses on healthcare, there is general lack of uniformity in the metrics for evaluation of recommender systems because many of them are in use (Del Olmo & Gaudioso, 2008). A number of these metrics have been used by individuals based on their perceived appropriateness to their work. Therefore, harmonizing and categorizing them will provide a uniform platform for evaluating recommender systems regardless of the objectives of an individual researcher or stakeholder. Valdez et al. (2016) listed evaluation of a recommender system as one of the important steps. It can then be deduced that thinking about developing a recommendation system should go along with thinking about making it meet evaluation criteria. This obviously is better done as part of requirements and design of such a recommender system.

In justifying their choice of metrics and proving their appropriateness, many authors have chosen to evaluate their work using two or three metrics. Accuracy has been the most popular metric for evaluating recommender systems and this sometimes is based on the algorithm used (Del Olmo & Gaudioso, 2008; Sokolova & Lapalme, 2009; Rebouças Filho et al., 2017; Rodrigues et al., 2018). Related closely to this are Precision, recall and sort priority which have also been identified as the common evaluation metrics for evaluation of recommender systems. (Zhong & Li, 2016; Moreira et al., 2018) However, some have argued that there is the need to look beyond accuracy (Vargas & Castells, 2011; Wu et al., 2012; He et al., 2016). It has also been reported that quality of recommendation is an important metric that deserves an attention rather than just the predictive accuracy of algorithms (Ge et al., 2008). While accuracy has been given wider publicity as a metric for determining the effectiveness of a recommender system, some researchers have opined that it is not a good measure of quality perceived by the users (Cremonesi et al., 2011) but other metrics such as serendipity and coverage (Ge et al., 2008) along with confidence (Duan et al., 2011) have more roles to play in satisfying users recommendations desired other than the accuracy. While it may be true to have surprise recommendations that fit the needs of a particular user (serendipity), delivering health recommendations as a surprise to the user should be done with a lot of caution. In an ongoing project, Recommendations Sharing Community for Aged and Chronically Ill People (ReSCAP), concern is more about addressing the specific needs of the individuals within this community in order to reduce time spent in searching, and because of the chronic nature of the ailments of the individuals involved. In this type of project, timeliness as an important factor to be considered can enhance accuracy (Zhang et al., 2017).

In this work, Criteria is defined as a set of standards by which the effectiveness and quality of a recommender system is determined while Metrics are a set of measurable attributes that can be used to determine performance of a recommender system. Evaluation is the process of judging the quality, importance, or value of a recommender system. (Afolabi and Toivainen, 2018). The main objective of this work is to harmonize and categorize metrics and criteria for evaluating recommender systems and not how they are measured. Several authors have dealt with this other objectives of this work includes determination of relevant metrics and criteria for evaluation of recommender systems in healthcare, determination of the degree of importance of each metric or criterion to the stakeholders, and ascertaining the relationship between degree of important of the metrics and criteria and the institution where stakeholders carry out their research activities. After analyzing the result of an online survey, speed, timeliness, accuracy, and usefulness are on the top of metrics on the degree of importance to stakeholders. For criteria however, providing information that will guide users to
Data Visualization with IBM Watson Analytics for Global Cancer Trends Comparison from World Health Organization

Nursing Home
Shuyan Xie, Yang Xiao and Hsiao-Hwa Chen (2010). *Pervasive and Smart Technologies for Healthcare: Ubiquitous Methodologies and Tools* (pp. 39-77).
[www.igi-global.com/chapter/nursing-home/42374?camid=4v1a](www.igi-global.com/chapter/nursing-home/42374?camid=4v1a)