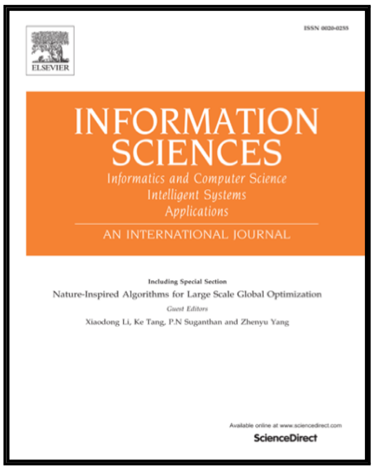
Accepted Manuscript




Reliability Quality Measures for Recommender Systems

Jesus Bobadilla

PII: S0020-0255(16)31005-2
DOI: 10.1016/j.ins.2018.02.030
Reference: INS 13435

To appear in: *Information Sciences*

Received date: 21 September 2016
Revised date: 29 November 2017
Accepted date: 15 February 2018

Please cite this article as: Jesus Bobadilla , Reliability Quality Measures for Recommender Systems, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.02.030

**RELIABILITY QUALITY MEASURES FOR RECOMMENDER SYSTEMS**

Jesus Bobadilla
Universidad Politécnica de Madrid. Computer Science Crta. De Valencia, K 28031 Madrid, Spain

**Abstract**

Users want to know the reliability of the recommendations; they do not accept high predictions if there is no reliability evidence. Recommender systems should provide reliability values associated with the predictions. Research into reliability measures requires the existence of simple, plausible and universal reliability quality measures. Research into recommender system quality measures has focused on accuracy. Moreover, novelty, serendipity and diversity have been studied; nevertheless there is an important lack of research into reliability/confidence quality measures.

This paper proposes a reliability quality prediction measure (*RPI*) and a reliability quality recommendation measure (*RRI*). Both quality measures are based on the hypothesis that the more suitable a reliability measure is, the better accuracy results it will provide when applied. These reliability quality measures show accuracy improvements when appropriated reliability values are associated with their predictions (i.e. high reliability values associated with correct predictions or low reliability values associated with incorrect predictions).

The proposed reliability quality metrics will lead to the design of brand new recommender system reliability measures. These measures could be applied to different matrix factorization techniques and to content-based, context-aware and social recommendation approaches. The recommender system reliability measures designed could be tested, compared and improved using the proposed reliability quality metrics.

## 1. Introduction

In the Recommender Systems (RS) field, confidence or reliability has been defined as [6] “How sure the recommender system is that its recommendation is accurate”. Quality metrics are a key factor for researchers in Collaborative Filtering (CF) RS. By combining quality measures (QM) and open datasets, researchers can improve results from previous works. RS researchers have focused on accuracy QM to test their methods and algorithms. Nevertheless, reliability measures (RM) and reliability quality measures (RQM) did not have the importance of accuracy or novelty research.

We claim RM are very important to RS users, since we know prediction and recommendation values have only a relative meaning. Electronic commerce clients usually look up the number of users that have rated products; we prefer a 4-star rated product based on 50 opinions to a 4.5-star rated product based on 2 opinions. In this case, the client naive quality metric is just the number of opinions. Sometimes we check the mass function of opinions: we prefer a 3-star product based on twenty 3-star opinions to a 3-star product based on 10 one-star and 10 five-star opinions.

Following the above examples, an electronic commerce website could design a really simple RM combining both the number of ratings and the inverse of the rating's standard deviation. Additionally, it could add some other useful information such as the *KNN* number of neighbors involved in the prediction, content-based information, etc. This electronic commerce website could provide each client recommendation with the pair: <number of stars, reliability value>. Clients would understand this information as a set of their friends recommending some films: “we believe you will really love film A, but we are pretty sure you will like film B”; Film A <0.95,0.60>, Film B <0.70,0.92>.

It is necessary to know the difference between a RM and a RQM. The first one assigns reliability values to each <user,item> prediction; the second one applies a testing strategy (such as cross-validation) to obtain the quality of the reliability values, that is the quality of the RM. Figure 1 shows these concepts.

As can be seen in Figure 1, RM or reliability methods provide reliability values (*Lu,i*). These values can be obtained in the same RS stage that returns predictions, or they can be obtained using separate algorithms or methods. Using machine learning techniques (such as matrix factorization), reliability values could be obtained: a) directly from a modified machine learning algorithm, b) from the generated model (factorized matrices in the MF example). As far as we know there are no RS methods to get <prediction, reliability> pairs from machine-learning models only based on rating datasets; this is an important open research field that requires reliability quality metrics (such as the ones proposed) to be explored.