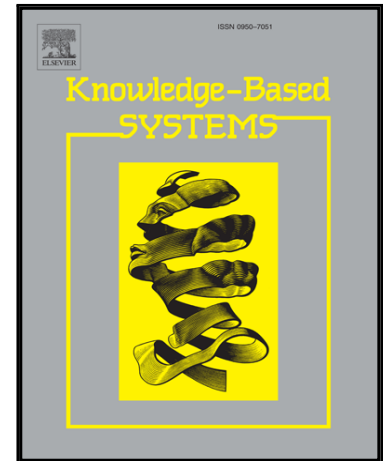


Accepted Manuscript

Efficient Global Correlation Measures for a Collaborative Filtering Dataset

Adrian Satja Kurdija, Marin Silic, Klemo Vladimir, Goran Delac

PII: S0950-7051(18)30065-0
DOI: [10.1016/j.knosys.2018.02.013](https://doi.org/10.1016/j.knosys.2018.02.013)
Reference: KNOSYS 4221



To appear in: *Knowledge-Based Systems*

Received date: 21 July 2017
Revised date: 25 January 2018
Accepted date: 6 February 2018

Please cite this article as: Adrian Satja Kurdija, Marin Silic, Klemo Vladimir, Goran Delac, Efficient Global Correlation Measures for a Collaborative Filtering Dataset, *Knowledge-Based Systems* (2018), doi: [10.1016/j.knosys.2018.02.013](https://doi.org/10.1016/j.knosys.2018.02.013)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Efficient Global Correlation Measures for a Collaborative Filtering Dataset

Adrian Satja Kurdija, Marin Silic, Klemo Vladimir, and Goran Delac

*University of Zagreb, Faculty of Electrical Engineering and Computing,
Consumer Computing Lab, Unska 3, 10000 Zagreb, Croatia*

Abstract

Recommender systems based on collaborative filtering (CF) rely on datasets containing users' taste preferences for various items. Accuracy of various prediction approaches depends on the amount of similarity between users and items in a dataset. As a heuristic estimate of this data quality aspect, which could serve as an indicator of the prediction ability, we define the Global User Correlation Measure (GUCM) and the Global Item Correlation Measure (GICM) of a dataset containing known user-item ratings. The proposed measures range from 0 to 1 and describe the quality of the dataset regarding the user-user and item-item similarities: a higher measure indicates more similar pairs and better prediction ability. The experiments show a correlation between the proposed measures and the accuracy of standard prediction models. The measures can be used to quickly estimate whether a dataset is suitable for collaborative filtering and whether we can expect high prediction accuracy of user-based or item-based CF approaches.

Keywords: collaborative filtering, dataset quality, global correlation, user

Download English Version:

<https://daneshyari.com/en/article/6861559>

Download Persian Version:

<https://daneshyari.com/article/6861559>

[Daneshyari.com](https://daneshyari.com)