

# Symbolic Data Analysis Tools for Recommendation Systems

Byron L. D. Bezerra<sup>1</sup>, Francisco de A.T. de Carvalho<sup>2</sup>

Departamento de Sistemas e Computação, Universidade de Pernambuco  
Rua Benfica, 455 - Madalena - CEP 50720-001 - Recife (PE) - Brazil

**Abstract** Recommendation Systems have become an important tool to cope with the information overload problem by acquiring data about the user behavior. After tracing the user behavior, through actions or rates, Computational Recommendation Systems use information filtering techniques to recommend items. In order to recommend new items, one of the three approaches has been mainly adopted: Content Based Filtering, Collaborative Filtering or hybrid filtering methods. This paper presents three information filtering methods, each of them based on one of the previous approaches. In our methods, the user profile is designed through Symbolic Data Structures and the user and item correlations are computed through distance functions adapted from the Symbolic Data Analysis domain. The usage of Symbolic Data Analysis tools have improved the performance of Recommendation Systems, specially when there is little information about the user.

**Keywords:** Symbolic Data Analysis, Recommender Systems, Information Filtering, Information Retrieval.

## 1 Introduction

Recommendation systems allow e-commerce websites to suggest products to their costumers, providing relevant information to help them in shopping tasks [1], [2], [3]. Additionally, most often this sort of system has increased their importance in entertainment domains [4], [5]. For instance, some interesting features are personalized TV guides in digital televisions and music recommendation in on-line stations.

In order to suggest items, Recommender Systems need user preferences to build suggestions. The process of collecting user preferences may be made implicitly (listening to some music in a CD store or, even better, buying a CD) or explicitly (evaluating some article with a grade in a on-line magazine) [2]. Independently of the acquiring approach (implicit or explicit), as much preferences are collected from user, better suggestions will be provided. But, in fact, a relevant problem remains in this process: the user has not enough time to giving information about him/her. So, it is necessary to learn about user with as little information as possible. This problem is even more difficult to cope with in the first system usage, when there is no information about user. In this case, it is interesting a suitable strategy to acquire user preferences.

The next step is filtering in relevant information in order to present it to the user through his/her profile previously acquired. The proposed solutions for this subject can be classified in two main groups concerning the kind of Information Filtering approach, e. g., Content-based Filtering (which is based on the correlation between the user profile and items content) or Collaborative Filtering (which is based on the users profiles correlation)

[3]. These techniques have inherent limitations, such as impossibility to codify some information in the first approach and latency (or cold-start problem) in the second one. Therefore, several works have exploited hybrid recommenders to overcome the drawbacks of each [2].

In this paper we describe information filtering techniques which the user profile and item information are modeled by Modal Symbolic (MS) data. This kind of structure was firstly defined in the Symbolic Data Analysis (SDA) field. SDA provides suitable tools for managing aggregated data described by multi-valued variables, where data table entries are sets of categories, intervals, or probability distributions [6].

Based on SDA data structures and tools, we develop three methods. The first one is an evolution of the Content-based approach presented in [7]. The second method may be classified as a Collaborative Filtering approach. Finally, we propose a Hybrid Filtering approach supported by SDA tools.

A deep experimental analysis was carried, taking into account the following issues: (i) the methodology of acquiring preferences, (ii) the algorithm used to learn user preferences, (iii) the amount of information known about the user, (iv) the database size, e. g., the number of users in the community, and (v) the metrics used to evaluate the proposed methods. The experiments were conducted in the movie recommendation domain, where the user profile is formed by way of a list of items that the user either preferred or disliked in the past, along with their respective grades.

The usage of Symbolic Data Analysis tools have improved the performance of Recommendation Systems, specially when there is little information about the user.

## References

- [1] F. Ricci and H. Werthner, Introduction to the Special Issue: Recommender Systems, *International Journal of Electronic Commerce* **11** (2006) 5–9.
- [2] K. Wei, J. Huang and S. Fu, A Survey of E-Commerce Recommender Systems, *International Conference on Service Systems and Service Management* (2007) 1–5.
- [3] Z. Huang, D. Zeng and H. Chen, A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce, *IEEE Intelligent Systems*, **Vol. 22, N. 5** (IEEE, 2007) 68–78.
- [4] Y. Blanco-Fernandez, J.J. Pazos-Arias, M. Lopez-Nores, A. Gil-Solla and M. Ramos-Cabrer, AVATAR: An Improved Solution for Personalized TV based on Semantic Inference, *IEEE Transactions on Consumer Electronics* **52** (2006) 223–231.
- [5] J. Salter and N. Antonopoulos, CinemaScreen Recommender Agent: Combining Collaborative and Content-Based Filtering, *IEEE Intelligent Systems* **Vol. 21, N. 1** (2006) 35–41.
- [6] H.H. Bock, E. Diday, *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data* (Springer-Verlag, Heidelberg, 2000).
- [7] B.L.D. Bezerra, F.A.T. de Carvalho, A Symbolic Approach for Content-based Information Filtering *Information Processing Letters* **92** (2004) 45–52.