Enhanced user-user collaborative filtering recommendation algorithm based on semantic ratings

Wen Zhang*, Raja Chiky ** Manuel Pozo**

*Stanford University
zhangwen@cs.stanford.edu,
**ISEP Paris- LISITE Lab
firstname.lastname@isep.fr

1 Introduction

This paper presents a collaborative filtering recommendation algorithm that borrows ideas from content-based models by taking into account both the ratings and user preferences for item attributes. This is achieved by replacing each rating with its corresponding "semantic rating", which combines the original score and the user's historical preference level for the item's attributes (Pozo et al., 2016). This algorithm can achieve better results than a pure collaborative filtering counterpart oblivious to the intrinsic properties of each individual item.

We base our work on a previous experience (Pozo et al., 2016), which proposes the "semantic equation" as a transformation from an original rating to its corresponding semantic rating, thereby capturing a user's preference for an item's attributes. Since it is only a transformation of ratings, this technique can be used in different kinds of collaborative filtering algorithms. (Pozo et al., 2016) applied the semantic equation on top of a matrix factorization algorithm and found that the semantic approaches yielded better results in terms of precision, recall, and intra-list diversity. In this paper, we apply the same idea to a simple neighborhood method based on Pearson correlations. We validate (Pozo et al., 2016) and deeper shows the effect of these "semantic transfomation" by analyzing the "semantic amendments" only.

2 Semantic algorithms

The semantic algorithms that we propose are all based on the semantic equation, which turns a rating $r_{u,i}$ into its amended semantic rating $\mathfrak{r}_{u,i}$ (Pozo et al., 2016). For a user u and an item i, we define their "semantic amendment", which captures the user's historical preference for the item's attributes, as follows:

$$\Delta_{u,i} = \bar{r}_u \cdot \frac{\left| \sum_{a \in A(i)} C_{u,a} W_{t(a)} \right|}{|S(u)|}, \tag{1}$$

where $C_{u,a} = |\{i \in S(u) : a \in A(i)\}|$ denotes the number of times attribute a appears in all items rated by user u, and $W_{t(a)}$ denotes the "weight" of the attribute a's type. Note that $\Delta_{u,i}$