

RECOMMENDER SYSTEMS - AN INTRODUCTION

CORRIGENDA LIST

DIETMAR JANNACH

ABSTRACT. Despite our best efforts when proof-reading the book, there are a few errors and imprecisions in the book.

1. CORRIGENDA

1.1. **Chapter 2, Section 2.1.1, page 15.** We use the symbol \bar{r}_a both in Equation 2.1 and Equation 2.3 to denote a user's average rating. When computing the similarity of two users in Equation 2.1, only the co-rated items should be used to determine the averages (and the similarity) of the respective users. In Equation 2.3, in contrast, *all known ratings* of the target user are meant by \bar{r}_a as we are interested in the user's global rating bias.

1.2. **Chapter 2, Section 2.1.2, page 17.** When describing the principle of case amplification we say that *"values close to +1 and -1 are emphasized by multiplying the original weights by a constant factor ρ "*.

Instead, according to [1], the amplified weight $\omega'_{a,i}$ is calculated as $\omega_{a,i}^\rho$ when $\omega_{a,i} \geq 0$ and as $-(\omega_{a,i}^\rho)$ otherwise.

1.3. **Chapter 2, Section 2.2.1, page 20.** Equation 2.9 for calculating the prediction in item-based collaborative filtering is printed as

$$pred(u, p) = \frac{\sum_{i \in ratedItems(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItems(\mathbf{a})} sim(i, p)}$$

but should be

$$pred(u, p) = \frac{\sum_{i \in ratedItems(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItems(\mathbf{u})} sim(i, p)}$$

(In the denominator, $ratedItems(a)$ has to be replaced by $ratedItems(u)$)

1.4. **Chapter 2, Section 2.2.2, page 21.** In the second paragraph, we say:

"At run time, a prediction for a product p and user u is made by determining the items that are most similar to i ...".

Instead, it should be:

"At run time, a prediction for a product p and user u is made by determining the items that are most similar to p ...".

Thanks to Fatih Gedikli and others for helping us to improve the book.

1.5. **Chapter 2, Section 2.4.1, page 28.** “In the example, we calculate U , V , and Σ (with the help of some linear algebra software) but retain only the two most important features by taking only the first two columns of U and \mathbf{V}^T , see Table 2.5.”

Actually, the text should be “... but retain but retain only the two most important features by taking only the first two columns of U and \mathbf{V} ...”.

1.6. **Chapter 2, Section 2.5.2, page 46.** When explaining Google’s news personalization algorithm, we say that “*A co-visit means that an article has been visited by the same user within a defined period of time.*”.

This is unfortunately a misleading formulation. Co-visitation actually means that two different articles have been visited by the same user in a defined period of time. Here is the original quote from [2]: “*Our item based technique for generating recommendations makes use of covisitation instances, where covisitation is defined as an event in which two stories are clicked by the same user within a certain time interval (typically set to a few hours)*”.

REFERENCES

1. John S. Breese, David Heckerman, and Carl Myers Kadie, *Empirical analysis of predictive algorithms for collaborative filtering*, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (Madison, Wisconsin, USA) (Gregory F. Cooper and Serafin Moral, eds.), Morgan Kaufmann, 1998, pp. 43–52.
2. Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram, *Google news personalization: scalable online collaborative filtering*, Proceedings of the 16th International Conference on World Wide Web (WWW’07) (New York, NY, USA), ACM Press, 2007, pp. 271–280.

TU DORTMUND

E-mail address: dietmar.jannach@tu-dortmund.de