

COVER LETTER OF THE KDMiLe EXTENDED ARTICLE FOR PUBLISHING IN THE JOURNAL JIDM

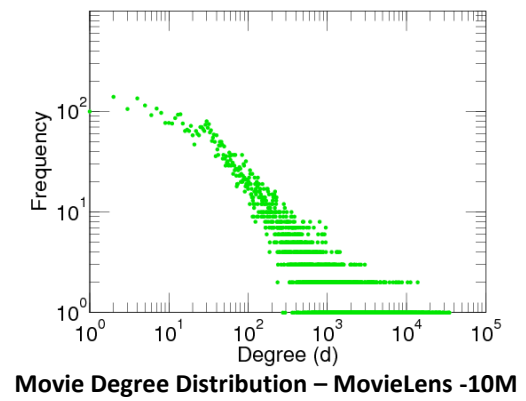
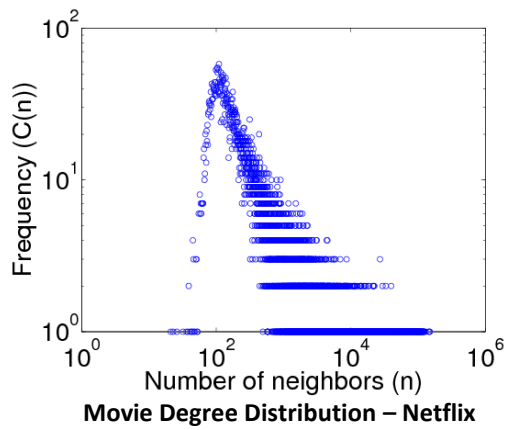
Campina Grande - PB, March 20th 2014.

TO THE EDITORIAL COMMITTEE,

Dear Editors,

In response to the JIDM reviewers' requests and suggestions, the extended version of our article published at KDMiLe'13 entitled "*Predicting the Learning Rate of Gradient Descent for Accelerating Matrix Factorization*" includes the new material listed below. Each item below represents the main suggestions by each reviewer.

- Reviewer A pointed out that the main idea and the parameters of the techniques used as baselines need more clarification. To address this request, we have improved the text adding more details and the intuitions about each one (Sections 2 and 5). We also stress their differences to our approach and explain the meaning of their respective parameters. Basically, these parameters control the importance of past iterations in the update rule of the learning rate.
- In response to Reviewer B, we emphasized why the same initial value of learning rate is used across all algorithms. Actually, this value comes from the literature, i.e., most of the papers we reviewed report this very same value. Also, adopting the same value for all algorithms allow us to compare them in a common ground. Another point raised by the reviewer concerns the influence of the number of latent factors on our method. To answer this, we added Figure 3, which shows that the linear pattern still holds regardless of the different numbers of latents factors used. A new baseline (the Gradient Cosine Adaption strategy) was added in order to stress the reasonable amount of new material added to the paper. Finally, we still don't have definitive answers for the questions "How can we use a general criterion to select (or not) a good greater initial value for the learning rate of the first interaction?", although we already have some clues. The assumption is that recommender datasets sharing common characteristics share similar learning behavior w.r.t. matrix factorization. Although Netflix indeed share many characteristics with the other datasets (in special the MovieLens), in some it differs. For example, the movie degree distribution of Netflix, which indicates how many users watched a movie, is slightly different than the one observed for MovieLens (see plots below). We still don't know the factors that caused the discrepancy of Netflix compared to the others, but we suspect that this might be one of them.
- Reviewer's C main comments concern the normalization of ratings: why and how we did it. We added this explanation to the text. We adopted the same methodology of [Lu et al. 2013] under the argument that this allows to compare the algorithms in a common ground. Please note that the dating dataset is actually the only one that has different a rating scale, all the others datasets have ratings in the usual [1,5] scale. The normalization of this dataset, as now explained in the text, consisted in mapping the ratings 1 and 2 to 1, 3 and 4 to 2, 5 and 6 to 3, 7 and 8 to 4 and 9 and 10 to 5.



We will remain at your disposal for any additional information you may need.

Best regards,

Caio Nóbrega
Leandro Marinho