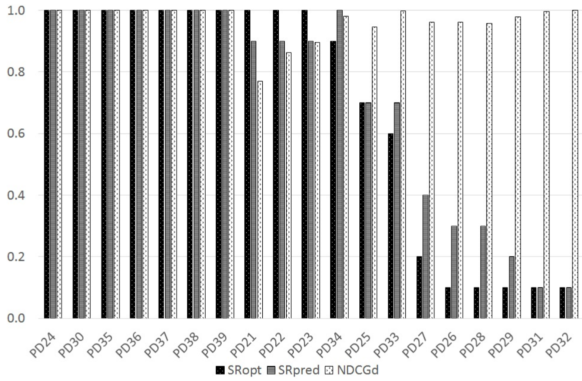

Figure 3: Sampling rate predictions of the 19 datasets.

We used the $NDCG$ metric [9], where the gain of each item was proportional to its rank in the complete predictions, to quantify the predictions generated by the sampled models[4]. We considered the complete model and the sampled model to be sufficiently similar, as long as the $NDCG$ computed by the sampled model for the fixed test set was greater than $\Delta = 0.95$. Thus, we decreased the sample rate SR by steps of 0.1 as long as we managed to obtaine $NDCG \geq 0.95$. The minimal SR, for which this $NDCG$ had been obtained, was considered the optimal sampling rate SR_d^{opt}.

On top of this, we applied the model in Equation 2 to predict the sampling rate SR_d^{pred} for each d. Since the optimal sampling rate SR_d^{opt} was an approximation found by search with steps of 0.1, also SR_d^{pred} was rounded to the closest 0.1 mark. Having set the sampling rate of d to SR_d^{pred}, we created the sampled dataset, then built the sampled MF recommendation model, and generated predictions for the fixed test set. Finally, we evaluated the performance, $NDCG_d^{pred}$, of the recommendation model built using the predicted sampling rate SR_d^{pred}.

Figure 3 presents the results of the sampling rate predictions for the 19 datasets. Each dataset is represented by three bars: namely, SR_d^{opt}, SR_d^{pred}, and $NDCG_d^{pred}$. The datasets are sorted in a decreasing order of SR_d^{opt}. As can be seen, the values of SR_d^{opt} vary across the datasets from 1 (no sampling is needed, all the users are necessary) to 0.1 (only 10% of users are necessary). For 10 datasets out of the 19 we observe $SR_d^{opt} = 1$, which aligns with the established sparsity problem in recommender systems. However, for 6 datasets we observe $SR_d^{opt} \leq 0.2$, indicating that some datasets have high degree of redundancy in the data.

Overall, the predicted sampling rates produced by the model in Equation 2 are close to the optimal ones. We observe that SR_d^{opt} and SR_d^{pred} are identical for 10 datasets out of the 19 (note that for 7 datasets, we observe $SR_d^{opt} = SR_d^{pred} = 1$, i.e., no sampling needed), for 6 datasets the difference is 0.1 (3 over-sampled and 3 under-sampled), and for 6 datasets the difference is 0.2 (SR_d^{pred} over-samples). The average difference between SR_d^{opt} and SR_d^{pred} across the 19 datasets is 0.063. Note that when $SR_d^{opt} \neq SR_d^{pred}$, we prefer to over-sample, i.e., $SR_d^{pred} > SR_d^{opt}$, as in this case, despite keeping unnecessary users, the recommendation model still achieves the desired degree of similarity to the complete model.

We also observe high $NDCG_d^{pred}$ scores, such that for 15 datasets we achieve $NDCG_d^{pred} \geq 0.95$. These include the 7 datasets with $SR_d^{opt} = SR_d^{pred} = 1$, where no sampling is performed and we obviously achieve $NDCG_d^{pred} = 1$. The average $NDCG_d^{pred}$ obtained across the 19 datasets stands at 0.964. Finally, we observe negative correlation of -0.382 between the obtained $NDCG_d^{pred}$ scores and the absolute value of the difference between the predicted and optimal sampling rate, $|SR_d^{opt} - SR_d^{pred}|$. This result is not surprising, since the accuracy of the recommendation models built using the sampled data deteriorates with the error in the sampling rate predictions generated by the model in Equation 2.

[4]As $NDCG$ combines ranking and predictive accuracy metrics, we deem it to be a reliable model performance indicator.

5. CONCLUSIONS

Our work was driven by the need to instantiate data quality models for recommender systems. To this end, we addressed two practical considerations of large-scale recommenders: sparsity of user ratings and redundancy of users in the datasets. We developed two models for predicting the data cleansing parameters and demonstrated their validity using a large collection of datasets. Notably, these models capitalize only on the parameters of the datasets and do not require the costly recommendation model building.

This work paves the way for future works on data quality in recommender systems. First, the proposed predictive models for data cleansing parameters were evaluated using the MF recommendation model. However, our models should be evaluated with other recommendation techniques, as, for instance, the item threshold may depend on the underlying recommendation model. Second, the impact of data cleansing on other performance metrics. The filtering of cold users/items and the sampling of users can affect the coverage and the diversity of the generated recommendations. Hence, there is a need to strike the balance between data quality assurance and these metrics. Third, we will consider the ways to incorporate content features of the items and demographic features of the users in the proposed predictive models.

6. REFERENCES

[1] X. Amatriain, A. Jaimes, N. Oliver, and J. M. Pujol. Data mining methods for recommender systems. In *Recommender Syst. Handbook*. 2011.

[2] X. Amatriain, J. M. Pujol, N. Tintarev, and N. Oliver. Rate it again: increasing recommendation accuracy by user re-rating. In *Recommender Syst. Conf.*, pages 173–180, 2009.

[3] S. Berkovsky, T. Kuflik, and F. Ricci. The impact of data obfuscation on the accuracy of collaborative filtering. *Expert Syst. Appl.*, 39(5):5033–5042, 2012.

[4] B. M. Marlin and R. S. Zemel. Collaborative prediction and ranking with non-random missing data. In *Recommender Syst. Conf.*, pages 5–12, 2009.

[5] U. Paquet and N. Koenigstein. One-class collaborative filtering with random graphs. In *Int. World Wide Web Conf.*, pages 999–1008, 2013.

[6] L. L. Pipino, Y. W. Lee, and R. Y. Wang. Data quality assessment. *Commun. of the ACM*, 45(4):211–218, 2002.

[7] R. Ronen, N. Koenigstein, E. Ziklik, M. Sitruk, R. Yaari, and N. Haiby-Weiss. Sage: recommender engine as a cloud service. In *Recommender Syst. Conf.*, pages 475–476, 2013.

[8] A. Said, B. J. Jain, S. Narr, and T. Plumbaum. Users and noise: The magic barrier of recommender systems. In *Int. Conf. on User Modeling, Adaptation, and Personalization*, pages 237–248, 2012.

[9] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender Syst. Handbook*. 2011.

[10] R. Y. Wang and D. M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, pages 5–33, 1996.