Do Users Matter? The Contribution of User-Driven Feature Weights to Open Dataset Recommendations

Anusuriya Devaraju
CSIRO Mineral Resources
Kensington, Western Australia 6151
anusuriya.devaraju@csiro.au

Shlomo Berkovsky
CSIRO Data61
Eveleigh, New South Wales 2015
shlomo.berkovsky@csiro.au

ABSTRACT

The vast volumes of open data pose a challenge for users in finding relevant datasets. To address this, we developed a hybrid dataset recommendation model that combines content-based similarity with item-to-item co-occurrence. The features used by the recommender include dataset properties and usage statistics. In this paper, we focus on fine-tuning the weights of these features. We experimentally compare two feature weighting approaches: a uniform one with predefined weights and a user-driven one, where the weights are informed by the opinions of system users. We evaluated the two approaches in a study, involving the users of a real-life data portal. The results suggest that user-driven feature weights can improve dataset recommendations, although not at all levels of data relevance, and highlight the importance of incorporating target users in the design of recommender systems.

KEYWORDS

feature weighting, hybrid recommender system, open data.

1 INTRODUCTION

The adoption of open data policies by research institutions and government agencies has led to a dramatic increase in the volume of open data. Although open data brings numerous benefits, the proliferation and the diversity of data make it difficult for users to find relevant datasets. Current data repositories primarily support keyword and faceted search modes. These may benefit users, who can precisely express their needs and are familiar with the data repository, but may pose a challenge otherwise. In addition, the search may return a long list of loosely related results, which may aggravate the dataset discovery task. All this raises the issue of delivering personalized dataset recommendations to users. Recommender systems were applied in the past to assist the discovery of scholars, articles, and citations [1]. To the best of our knowledge, recommending open datasets has not been thoroughly investigated yet. Singhal et al. [4] developed a context-based search for research datasets, which deployed similarity-based ranking based on topic, abstract, and authors of datasets [4]. In our previous work, we developed a hybrid dataset recommendation model that identified relevant datasets by using both content-based and statistical features, including dataset metadata and observable usage patterns [2]. The features were combined in a linear manner into a single dataset-to-dataset similarity score.

In this paper, we focus on the feature weights. We deploy and evaluate two weighting models. The first uses fixed uniform weights, which are defined heuristically by the system designers. The second utilizes the weights derived from a survey among target users of the system. Our evaluation aims to uncover whether *user-driven feature weights lead to better recommendations than the uniform weights*. The results indicate that the user-driven weights can improve dataset recommendations, although this observation is mainly valid at certain level of data relevance. This finding highlights the importance of considering the opinions of target users when designing a dataset recommender system.

RecSys '17 Poster Proceedings, Como, Italy
2017.

2 OPEN DATA RECOMMENDATION MODEL

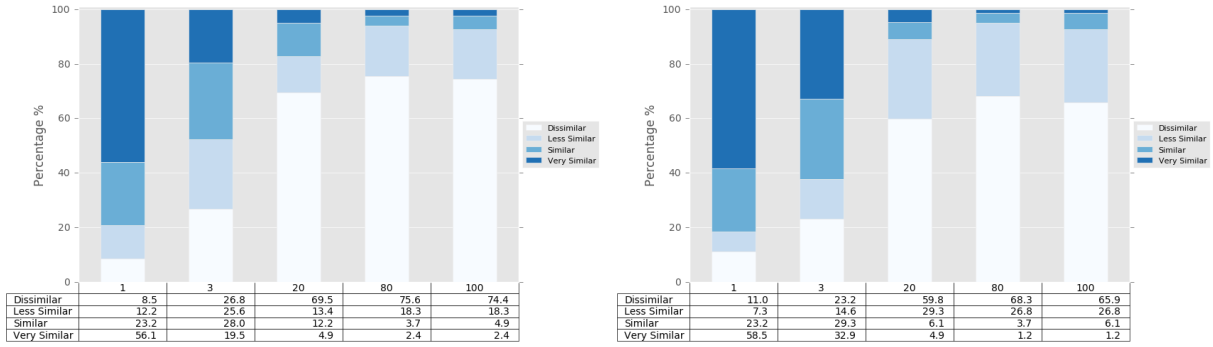
Given a target dataset d examined by a user, we recommend its n most relevant datasets $(d_1, \ldots, d_n)$ that are ranked according to their similarity to d. The similarity between of d and d_i is:

$$overall\_sim(d, d_i) = \sum_{i=1}^{n} (\omega_i \cdot sim_i (d, d_i)), \quad (1)$$

where ω_i is the weight associated with a feature i and $sim_i(d, d_i)$ is the similarity of d and d_i with respect to i. In total, we consider ten features: *title, description, keyword, activity, research field, creator, contributor, spatial, search,* and *download* [2]. We deploy *content-based similarity* and *item-to-item co-occurrence* [3] to compute the similarity of datasets. For the first eight features, the content-based similarity is used to identify similar datasets based on their metadata. For example, we use TF-IDF term weighting with Cosine Similarity for text-based features like *title* and *description*, and Jaccard's coefficient for categorical features like *research field* and *creator*. The *item-to-item co-occurrence* quantifies the similarity of datasets by comparing their statistical co-occurrence based on their joint appearance in *search* results and joint *download* by users. The underlying assumption is that two datasets are related if they are returned in response to similar queries or are downloaded in the same session.

3 EXPERIMENT AND RESULTS

As shown in Equation 1, feature-based similarity scores $sim_i(d, d_i)$ are aggregated linearly by using feature weights ω_i. However, how should these weights be set? Will different weighting models affect the quality of the recommendations? We consider two weighting models. The first uses a fixed set of weights, which are defined heuristically by the system designers. For the sake of simplicity, no domain knowledge is applied, and the weights of all ten features are set to $\omega_i = 0.1$. We refer to this as the *uniform* weighting model. The second weighting model is a *user-driven* one, as it is informed by the feature importance perceptions of the target system users. We conducted a survey, which involved 151 users of a real data repository. These users were shown the 8 eight features in the



	1	3	20	80	100
Dissimilar	8.5	26.8	69.5	75.6	74.4
Less Similar	12.2	25.6	13.4	18.3	18.3
Similar	23.2	28.0	12.2	3.7	4.9
Very Similar	56.1	19.5	4.9	2.4	2.4

	1	3	20	80	100
Dissimilar	11.0	23.2	59.8	68.3	65.9
Less Similar	7.3	14.6	29.3	26.8	26.8
Similar	23.2	29.3	6.1	3.7	6.1
Very Similar	58.5	32.9	4.9	1.2	1.2

Figure 1: Distribution of relevance judgments: (left) uniform feature weights, (right) user-driven feature weights.

above list and asked to rate their importance on a 5-Likert scale. The survey revealed that *title, description,* and *keywords* were more important features, while *creators* and *contributors* were deemed less important. These importance scores were mapped onto the feature weights, e.g., $\omega_{title} = 0.123$ and $\omega_{creators} = 0.086$.

Experimental setup: We evaluated the uniform and user-driven weighting models in two intra-group user studies that were approximately 2 months apart. In both studies, we showed to users a target dataset d they were familiar with, as well as with a list of 5 recommended datasets, at fixed ranks $i = 1, 3, 20, 80, 100$ in the list of datasets most similar to d. We showed the recommended datasets in a random order and asked the users to rate their relevance to d on a 4-Likert scale, ranging from 'very similar' to 'dissimilar'. We obtained the judgments of 50 users who participated in both studies and jointly rated 82 target datasets. Thus, our results are based on 410 judgments obtained in each study. Note that in both studies every user judged recommendations by referring to the same target dataset d. That said, the 5 recommended datasets might have changed due to the different feature weighting model.

Results: Figures 1-left and 1-right depict the distribution of the users' relevance judgments assigned to the recommendations produced by the uniform and user-driven weighting models, respectively. The horizontal axis represents the rank i of the recommended dataset and the vertical axis indicates the distribution of the judgments. Since the results of the two studies are similar, we also include the exact judgment distributions below the plots. It can be observed that the user-driven weighting achieves a slight improvement for datasets at rank 1. Here, 81.7% of the datasets were judged 'highly similar' or 'similar', compared to the 79.3% obtained for the uniform weighting. The differences are more pronounced at rank 3 where the user-driven weighting was judged 'highly similar' or 'similar' in 62.2% of cases, while the uniform weighting resulted in 47.5%. The obtained judgments at ranks 20, 80, and 100 are predominantly negative, so these datasets cannot be recommended and are excluded from the analysis. We compared the user judgments obtained across the two studies by using a pairwise t-test for means. We observed statistically significant differences at rank 3, $p < 0.001$ while at rank 1 the differences were not significant.

Discussion: Although the results of both studies were comparable at most ranks, our findings suggest that the user-driven feature weighting improves the quality of the recommendations at ranks 1 and 3. To acquire a better understanding of this, we plot in Figure 2 the average similarity of the recommended datasets at various ranks. This similarity exhibits a long-tail distribution. We believe that the recommended datasets at rank 1 were related regardless of the fine-tuned weights, as the strong user support of about 80% suggests. Hence, the improvement was insignificant. However, at rank 3, the average similarity is about 10% lower than at rank 1, as reflected by the lower user support dropping to the 50-60% mark. Hence, the improvement introduced by the user-driven weighting was found to be strongly significant. We conclude, therefore, that user-driven feature weights turn out to be particularly critical in the borderline areas where the relevance of the datasets is unclear. We believe that this finding reflects the importance of the target system users' opinions.

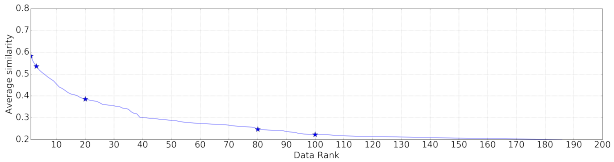


Figure 2: Average similarity of top-200 of 1000 datasets.

4 CONCLUSION

In this paper, we studied the importance of user-driven feature weights in producing open data recommendations. We compared their performance against the baseline of heuristically set uniform weights. The results have showed that user-driven feature weights have a positive effect on user judgments, although this finding may not necessarily be applicable at all ranks. We consider this work to provide an important argument in favor of incorporating target users in the early stage of designing a data recommender system.

REFERENCES

[1] J. Beel, B. Gipp, S. Langer, and C. Breitinger. 2016. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries* 17, 4 (2016), 305–338.

[2] A. Devaraju and S. Berkovsky. 2017. A Hybrid Recommendation Approach for Open Research Datasets. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management (under review).*

[3] L. Leydesdorff and L. Vaughan. 2006. Co-occurrence Matrices and Their Applications in Information Science: Extending ACA to the Web Environment. *J. Am. Soc. Inf. Sci. Technol.* 57, 12 (Oct. 2006), 1616–1628.

[4] A. Singhal, R. Kasturi, V. Sivakumar, and J. Srivastava. 2013. Leveraging Web Intelligence for Finding Interesting Research Datasets. In *International Conferences on Web Intelligence (WI).* 321–328.