Cross Domain Recommendation Using Vector Space Transfer Learning

Masahiro Kazama
Recruit Technologies Co.,Ltd.
Tokyo, Japan
masahiro_kazama@r.recruit.co.jp

István Varga
Recruit Technologies Co.,Ltd.
Tokyo, Japan
vistvan@r.recruit.co.jp

ABSTRACT

The cold start problem, frequent with recommender systems, addresses the issue in cases where we don't know enough about our users (e.g., the user hasn't rated anything yet, or there are no user activities) in that specific domain. In our paper we present a simple and robust transfer learning approach where we model users' behavior in a source domain, transferring that knowledge to a new, target domain. First, we vectorize the items by using word2vec for each dataset independently. Second, we calculate the transformation matrix that connects the source dataset to the target dataset by using their common users.

Keywords

Recommender Systems, Transfer Learning, Word Embedding

1. INTRODUCTION

The economic advantages of cross selling products between different domains (e.g., movies or books) are appealing and cross-domain recommendation has become a hot topic in recent years. However, while it is not always trivial how to identify patterns that successfully model user behavior, doing so across domains is even more difficult. If one tends to like David Lynch movies, we may be able to recommend "The Elephant Man", but what does this knowledge translates to in another similar yet different domain, such as the books domain? What about in an apparently unrelated domain, such as the restaurants domain?

In this paper we propose a very simple and robust transfer-learning based approach that can successfully model the behavior patterns of users across different domains, showing promising results even with only a small number of common users.

2. RELATED WORKS

In recent years, numerous transfer learning approaches have been proposed. Sahebi et.al. [2] use canonical correlation analysis in a cross-domain recommendation setting, but their method can only recommend items that are rated by the common users. Our work is partly inspired and extends the work presented in Zhang et.al. [3]. They attempt to identify correspondences between words across different
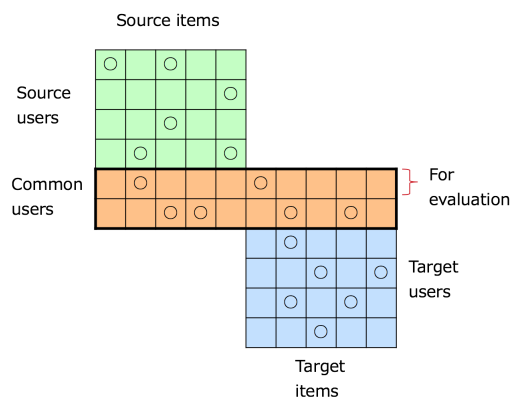
Figure 1: Transfer learning approach overview. Positive ratings are marked with circles.

timeframes by first building two independent vector spaces for each timeframe using word2vec [1]. Next, using anchor words whose meaning does not change significantly over time, they learn a transformation matrix between the two vector spaces, which enables the correspondence between the two timeframes, identifying word pairs that are semantically similar in their respective timeframes such as *Walkman* - *iPod*. We can apply the same methodology to recommender systems to address the cold start problem or to augment for missing information, where ratings from a source domain can be used to leverage for ratings in a new, target domain with the common users acting as anchors.

3. PROPOSED METHOD

Consider the case when we have a source domain and a new, target domain with no common items, but common users. Such cases are very frequent in real-life situations. We propose a cross-domain recommender methodology that can accurately recommend items from the entire inventory of the target domain, leveraging on the common users. We illustrate the scope of our paper in figure 1. Our method captures analogies in user behavior across domains by modeling user preference in the source model, transferring this knowledge to a new, target domain. The prerequisite for our proposal is only the existence of a relatively small group of common users.

The proposed method consists of two steps. First, we generate the vector representation for each item, separately for both the source and target domains. We use word2vec for

this generation process. Word2vec is usually employed for text analysis, but in our process we use it for user rate logs, treating items as words, with each users' positive items organized into paragraphs. Secondly, using the common users, we learn a transformation matrix M between the source and target domains.

$$M = \arg\min_M \sum_{i=1}^{N} \|Mu_i^s - u_i^t\| \qquad (1)$$

where $u_i^s (\in \mathbb{R}^K)$ is the vector of user i in the source vector space, u_i^t in the target vector space. User vectors are calculated as the mean of the item vectors rated by the user as positive. K is the dimension of each vector. N is the number of common users. $M (\in \mathbb{R}^{K \times K})$ is the transformation matrix between source and target spaces. Source dataset leveraged recommendation is performed by calculating the user vector and converting it to the target vector space by using the transformation matrix M. Once this conversion is achieved, the converted user vector can be compared to the item vectors in the new domain and items with most similar vectors can be recommended to this user. In this paper, the similarity measure used is the cosine similarity.

4. EXPERIMENTS

In our experiments we used the ratings from the yelp dataset[1]. User ratings range from 1 to 5, but we only consider positive ratings (4 and 5), ignoring the rest. Items are tagged by various labels (e.g. Restaurants, Shopping).

We validated our proposal on the 3 most frequent categories in terms of item count: Restaurants, Shopping and Food. For each source-target domain, we removed the items which were labeled in both categories. For example, in the Restaurants-Shopping pairing, after removing the common items, we retained 131,825 and 10,899 users respectively, among them, 8,642 were common users. Table 1 summarizes the statistics of the above categories.

Table 1: Statistics of the source-target pairs used in our experiments

Dataset A	Restaurants	Restaurants	Shopping
Dataset B	Shopping	Food	Food
Users in A	131,825	122,419	10,230
Users in B	10,899	15,555	27,687
Common users	8,642	13,234	5,374
Items in A	23,424	21,398	8,757
Items in B	9,369	6,373	7,811
Ratings in A	728,115	650,486	43,249
Ratings in B	47,140	67,273	122,189

As test data, we used 20% of the common users in all source-target pairing configurations.(e.g., 1728 users in the Restaurant-Shopping settings). The rest of data is used for training. In order to observe the effect of the number of common users, we incrementally increased the number of common users in our training data, starting from 10% of the training data and increasing by 10% at each step. For the evaluation metric we used recall@100, as we are interested in correctly identifying the items the users actually rate as positive in target domain, using the data only from the source domain. Besides recall@100, we experimented with other

---

[1]https://www.yelp.com/dataset_challenge

metrics such as recall@10 and @50, precision@10, @50 and @100. However, the same tendencies were observed in all experiments. As baseline, we treat the source and target domains as a single dataset, generating one common vector space model using word2vec in a similar way with our proposal, with the recommended items being the ones whose vector is closest to the user's vector.

Table 2: Recall@100 for the most frequent 3 categories with all source → target configurations. Horizontal axis represents the training data size ratio.
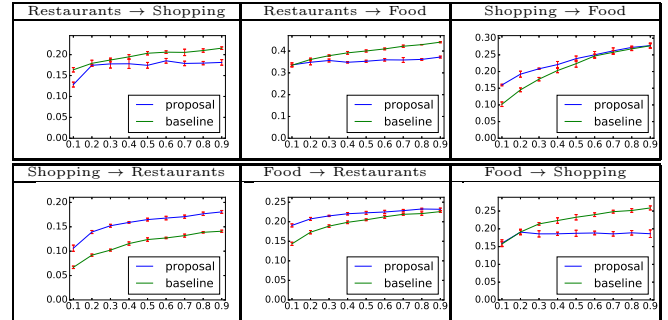

Table 2 shows the results for our selected categories. Our method is most effective when the number of users in the source domain is smaller than the number of the users in the target domain, outperforming the baseline. We also observed that our method is most efficient when the training size ratio is small, i.e., the number of common users in source-target domains is small. This is observed for the Food → Restaurants case, where our method outperformed the baseline method by 4.71 points, for a training size ratio of 10%. Both of the above tendencies were observed with source-target combinations unrestricted to the most frequent categories.

5. CONCLUSIONS AND FUTURE WORK

We proposed a simple, but robust transfer learning approach that can bridge the gap across differing domains. We found that our method is most efficient (1) when the number of users in the source domain is smaller than the number of the users in the target domain and (2) when the number of common users is small across the two domains.

We are investigating the efficiency of other item vector representation methods besides word2vec, such as matrix factorization within our transfer learning framework.

6. REFERENCES

[1] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS 2013*, pages 3111–3119, 2013.

[2] S. Sahebi and P. Brusilovsky. It takes two to tango: An exploration of domain pairs for cross-domain collaborative filtering. In *Proceedings of RecSys 2015*, pages 131–138, 2015.

[3] Y. Zhang, A. Jatowt, S. S. Bhowmick, and K. Tanaka. Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time. In *Proceedings of ACL 2015*, pages 645–655, 2015.