Session-Based Recommendations Using Item Embedding

Conference Paper · March 2017

DOI: 10.1145/3025171.3025197

3 authors, including:

Asi Messica
Ben-Gurion University of the Negev

Lior Rokach
Ben-Gurion University of the Negev

Session-Based Recommendations Using Item Embedding

Asnat Greenstein-Messica
Ben-Gurion University of the Negev
Beer-Sheva, Israel
asnatm@post.bgu.ac.il

Lior Rokach
Ben-Gurion University of the Negev
Beer-Sheva, Israel
liorrk@bgu.ac.il

Michael Friedmann
YOOCHOOSE GmbH
Cologne, Germany
michael.friedmann@yoochoose.com

ABSTRACT

Recent methods for learning vector space representations of words, word embedding, such as GloVe [7] and Word2Vec [5] have succeeded in capturing fine-grained semantic and syntactic regularities. We analyzed the effectiveness of these methods for e-commerce recommender systems by transferring the sequence of items generated by users' browsing journey in an e-commerce website into a sentence of words. We examined the prediction of fine-grained item similarity (such as item most similar to iPhone 6 64GB smart phone) and item analogy (such as iPhone 5 is to iPhone 6 as Samsung S5 is to Samsung S6) using real life users' browsing history of an online European department store. Our results reveal that such methods outperform related models such as singular value decomposition (SVD) with respect to item similarity and analogy tasks across different product categories. Furthermore, these methods produce a highly condensed item vector space representation, item embedding, with behavioral meaning sub-structure. These vectors can be used as features in a variety of recommender system applications. In particular, we used these vectors as features in a neural network based models for anonymous user recommendation based on session's first few clicks. It is found that recurrent neural network that preserves the order of user's clicks outperforms standard neural network, item-to-item similarity and SVD (recall@10 value of 42% based on first three clicks) for this task.

Author Keywords

Session-based recommender system; e-commerce; item embedding; recurrent neural network; deep learning; word embedding; word2vec; GloVe.

ACM Classification Keywords

I.2.6 [Artificial Intelligence]: Learning; I.2.1 [Artificial Intelligence]: Applications and Expert Systems.

IUI 2017, March 13-16, 2017, Limassol, Cyprus
© 2017 ACM. ISBN 978-1-4503-4348-0/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3025171.3025197

INTRODUCTION

Semantic vector space models of language represent each word with a real-valued vector in a low-dimensional space relative to the vocabulary size. Recent methods for learning vector space representations of words, word embedding, such as GloVe [7] and Word2Vec [5] have succeeded in capturing fine-grained semantic and syntactic regularities. These methods produced state-of-the-art performance in word analogy such as "king is to queen as man is to woman" and similarity tasks. Furthermore, these vectors, when used as the underlying input representation, have been shown to boost the performance in natural language process (NLP) applications such as information retrieval [3], question answering [10] and document ranking [6].

In this work we analyzed the effectiveness of using these methods in e-commerce applications by transferring the sequence of items generated by users' browsing journey in an e-commerce website into a sentence of words. We then used these methods to generate real-valued vectors in a low-dimensional space relative to the number of available items representing the used items, *item embedding*. We examined the prediction of fine-grained item similarity (such as item most similar to iPhone 6 64GB smart phone) and item analogy (such as iPhone 5 is to iPhone 6 as Samsung S5 is to Samsung S6) using two weeks of real life users' browsing history of an online European department store. Our results reveal that such methods and particularly GloVe outperforms related models such as singular value decomposition (SVD) on item similarity and analogy tasks across different product categories.

We also demonstrated the efficiency of using these meaningful highly condensed vectors (about 15,000 products are represented by a vector dimension of 50) generated from the GloVe model in an e-commerce recommender system. In particular, we used these vectors as features in a neural network based models for session based recommendations. A session-based recommendation [4] where user is anonymous and the recommendations are based on short session available data instead of long user history is relatively hardly dealt in the recommender systems community. However, anonymous user recommendation is common in real-life for occasional user scenario when the user is not logged in or the user-id is not tracked for either technical or privacy reasons. Recurrent neural network (RNN) that preserves the order of user's clicks outperforms standard neural network, item-to-item similarity, item popularity and SVD (recall@10 value of 42% based on first three clicks) for this task.

The contribution of this research is twofold: first we show that GloVe is highly efficient for e-commerce item analogy and similarity tasks. Secondly, we show that similar to NLP, it is efficient to use the resulted, item embedding, generated using the GloVe method as features in RNN for session-based recommendation.

RELATED WORK

Word Embedding Models

Word2Vec and GloVe methods and applications attracted a great amount of attention in recent years. The vector representation of words learned by these methods has been shown to carry semantic meanings and is useful in various NLP tasks.

The Word2Vec [5] method uses efficient neural network model for unsupervised learning of word representation, from large-scale text corpora. The GloVe [7] (Global Vectors) method is based on global log-bilinear regression model and combines the advantages of the global matrix factorization and local context window methods. The GloVe method explicitly factorizes the word-context co-occurrence matrix on symmetric word windows across the corpus.

Session-Based Recommendations

Much of the work in the area of recommender systems has focused on models that work when a user identifier is available and a clear user profile can be built. In this setting, matrix factorization methods such as singular value decomposition (SVD) provide state of the art results. In real life session-based recommendation where user is anonymous or not logged in yet and the recommendations are based on short session available data instead of long user history is quite common. In such cases item-to-item recommendation approach [8] is often used. With this method, items which are usually clicked or bought together with the items the user clicked are recommended. Another approach to session-based recommendation are Markov Decision Processes (MDPs) [9] methods which are based on sequential stochastic decision problems. Recently [4], RNN was applied to session-based recommendation with remarkable results.

ITEM ANALOGY AND SIMILARITY

Experiment

The dataset for this experiment was collected from real life users' browsing history of an online European department store over two weeks at the beginning of January 2016. It contains click stream of e-commerce click and purchase events as well as products catalog for this period. The products catalog covers approximately 80% of the products users clicked during this period across wide range of product categories.

To apply the Word2Vec and GloVe methods to predict fine grained item analogy and similarity for e-commerce scenario, we bagged all the daily click events per user sorted by its time stamp, into a sequence of items. For a logged-in user the user id is used to assign the relevant events to the user, where for anonymous user either cookie mechanism or the session id provided by the webserver are used to assign the relevant clicks to the anonymized user. We mapped each item into a word, and each sequence into a sentence. To improve the accuracy of the prediction, the effectiveness of adding metadata words which represent the item category and item price category was evaluated. Item category was extracted from the products catalog. For each product category five price categories were defined by dividing the highest and lowest price of items in the category to five equal bins. A price category was assigned per item. When adding the metadata words, a category id and price category id metadata words were added for each item word. We provided a corpus generated out of all these sessions as an input to the Word2Vec and GloVe methods to generate the item embedded vectors (i.e. *item embedding*). We used the publicly available implementations of Word2Vec and GloVe. As a baseline, we used publicly available implementation of the SVD method.

The following equation gives the local cost function of the GloVe method, which was minimized using gradient descent.

$$J = \sum_{i,j=1}^{V} f(X_{ij}) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log(X_{ij}) \right)^2$$

Here V is the number of items in the catalog, X_{ij} denotes the number of times item j occurs in the context of item i taking into account also the distance between the items within the context window, w_i is the vector representation of item i (i.e. the *item embedding*), $\tilde{w}_j$ is the context item vector, b_i , $\tilde{b}_j$ are bias terms and f is a weighting function that cuts off low co-occurrences, which are usually noisy, as well as avoid overweighting high co-occurrences. Here, w_i $\tilde{w}_j$, b_i , $\tilde{b}_j$ are the parameters to be learnt during training.

For the word2vec model, we used the CBOW model, with a linear combination of a one-hot encoded vector per context item as an input. Fully connected, feed forward neural network with a linear activation function between the input-layer to the hidden-layer, and a linear mapping in conjunction with *softmax* between the hidden layer and the output layer was used. Table 1 presents the optimized training parameters per method. Word2Vec provided better results with a smaller context window than GloVe. A possible reason for that is the GloVe sensitivity to the order of clicks, which may lead to high accuracy even with relatively large context window.

The length of the session has an influence on the item embedding process and sessions which are shorter than two

Method	Parameters
Word2Vec (W2V) CBOW	Vocabulary minimum word count 10 Context window size 8 # of iterations 15 Item embedding vector size 50
GloVe	Vocabulary minimum word count 10 Context window size 15 # of iterations 15 X_max 10 Item embedding vector size 50
SVD	Items minimum event count 10 Users minimum event count 3 Item embedding vector size 100

Table 1. Item embedding optimized parameters

clicks have no contribution. After filtering one click sessions, we had 999,444 sessions and 16,014 items which were used as input. 64% of the filtered sessions had three clicks or more". An item embedded vector of size 50 provided the best results for item analogy and similarity tasks with this corpus.

Figure 1 presents Word2Vec neural network input layer and desired output layer for a session with five item clicks and a context window of eight, where metadata words representing product category id and price category id per item were included. In our experiment, I the number of available items in the corpus was 16,014, T the number of product categories was 1,045. P the number of price categories was 5, and N the size of the *item embedding* vector was 50.

Results and Analysis

We conduct experiments on the item analogy and similarity

6692 > 5794 > 12050 > 1668 > 14750
Input (t-2): AEG Double socket, product id 6692, product category id 86, price category id 0
Input (t-1): Apple Smartphone iPhone 6, product id 5794, product category id 906, price category id 4
Output (t): Samsung Smartphone S6, product id 12050, product category id 906, price category id 3
Input (t+1): Canon Multi-function printer, product id 1668, product category id 969, price category id 0
Input(t+2): Dell Notebook, product id 14750 , product category id 1002 , price category id 0

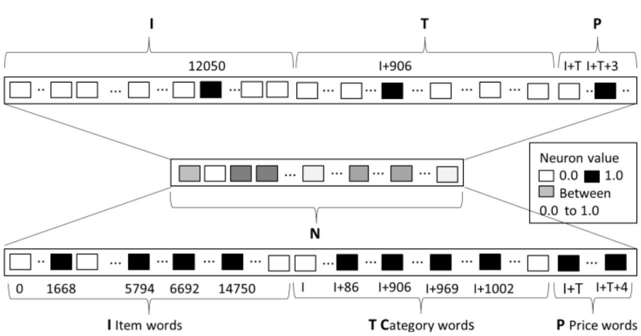

Figure 1. Word2Vec neural network - five items session

tasks across different product categories. We determined the answers for the item analogy test as in [5] using cosine similarity, selecting the item whose vector maximizes the formula: $\arg max_{b^* \in V} \left(cos(b^*, b - a + a^*) \right)$

Where b^* is the item to be guessed (e.g., "Samsung S6"), b is its known paired item (e.g. "Samsung S5"), and a, a^* are the known pair (e.g. "iPhone 5", "iPhone6").

For the similarity tests we selected the word whose vector maximizes the cosine similarity value. Table 2 illustrates three results of analogy tests using the different methods. Table 3 illustrates three results for similarity tests. These results show that both GloVe and Word2Vec can be effectively applied for fine granularity e-commerce item analogy and similarity tasks using the suggested methodology. GloVe outperforms both Word2Vec and the SVD baseline method (the latter, fails to provide similar level of accuracy). The addition of metadata words has a small impact on the results accuracy. These results are consistent over a wide variety of items we tested.

SESSION-BASED RECOMMENDATION

Similar to NLP where word embedding vectors are used as underlying inputs representation to boost the performance of various applications [3][10], we evaluated the efficiency of using the *item embedding* vectors generated by the GloVe method as an input representation of neural networks based models for session based recommendation. Both standard feed forward fully connected neural network (ANN) and a GRU based RNN [2] were evaluated for this task. One hidden layer and a *softmax* activation function in the output layer were used for both models. For the ANN a sigmoid activation function in the hidden layer was used, where for the RNN a GRU hidden layer was used. The same dataset described in section 3 was used, except that the data was split into train and test data sets according to the session's date. I*tem embedding* vectors of size 50 were generated using the GloVe method with the same methodology as before leveraging only the train data set. 945,806 sessions and 15,725 products were used in the train data set and 53,638 sessions were used in the test data set. The input of the neural network in both cases were the item embedding vectors, representing items which were clicked within a context window of size 3 of an item to be predicted. The output layer was one-hot encoded item representation of the desired output. The predicted item was determined as in section 3. For training, a context window of size 3 before each item in the session was used. For testing the first three item clicks were used per session as input to predict the next item which will be clicked. After filtering the sessions which included less than four clicks 16,592 sessions were left. For both models a hidden layer of size 200 was used. The evaluation metrics [4] for measuring the performance of the models were recall@k, $precision@k$ and $MRR@k$ (Mean Reciprocal Rank) which takes into account the rank of the item. As a baseline the item-2-item similarity, item popularity and the SVD models

Method	Test	Result	
GloVe	Apple iPhone 5 is to Apple iPhone 6 as Samsung S5 is to "?" Expected result: Samsung S6	Samsung S6	√
Glove Metadata		Samsung S6	√
W2V		Samsung S6	√
W2V Metadata		Apple iPhone 6 16G	×
SVD		Huawei G7	×
GloVe	Apple iPhone 5 is to Apple iPhone 6 as Brand A coffee machine € 99.99 is to "?" Expected result: More expensive coffee machine	Brand B coffee machine €319.0	√
Glove Metadata		Brand A coffee machine €129.0	√
W2V		Brand A coffee machine €129.0	√
W2V Metadata		Brand A coffee machine €129.0	√
SVD		Brand A coffee machine €34.99 ×	×
GloVe	Brand C Banana mash is to Brand C Rice mash as Brand D pasta tomato sauce €0.79 is to "?" Expected result: Pasta sauce different flavor similar price	Brand D pasta tomato sauce different flavor €0.79	√
Glove Metadata		Brand D pasta tomato sauce different flavor €0.79	√
W2V		Brand D tomato pasta sauce different flavor €1.29	√
W2V Metadata		Brand E pesto pasta sauce €1.19	√
SVD		Popcorn machine €59.99	×

Table 2. Examples of item analogy tests results

Method	Test	Result	
GloVe	Most similar item to Apple iPhone 6 64G Expected result: Apple iPhone6 different model or configuration	Apple iPhone 6 16G	√
Glove Metadata		Apple iPhone 6 64G + free starter package	√
W2V		Samsung S6	×
W2V Metadata		Apple iPhone 6 16G	√
SVD		Sport treadmill	×
GloVe	Most similar item to Brand F dishwasher €329.0 Expected result: Dishwasher similar price different model	Brand F dish washer different model €329	√
Glove Metadata		Brand F dish washer €349	√
W2V		Brand F dish washer €259	√
W2V Metadata		Brand F dish washer different model €329	√
SVD		Microwave €94.99	×
GloVe	Most similar item to Brand G woman sport trousers €5.99 Expected result: Woman sport trousers same price different model	Brand G woman sport trousers €5.99 (different model)	√
Glove Metadata		Brand G woman sport trousers €5.99 (different model)	√
W2V		Brand G woman sport trousers €6.99	×
W2V Metadata		Brand G woman sport trousers €5.99 (different model)	√
SVD		Brand G woman sport trousers €5.99 (different model)	√

Table 3. Examples of item similarity tests results

were used. Figure 2 presents the *recall@k* recommendation prediction results for each model. Table 4 presents *recall@10*, *precision@10* and *MRR@10* prediction results for each model. The results show that leveraging the item embedding as an input layer for a neural network recommender system models outperform all baseline models for this use case. The RNN GRU model, which preserves the order of the click events, provides better results than the standard ANN.

CONCLUSIONS AND FUTURE WORK

In this paper we evaluated the effectiveness of using state of the art word embedding methods GloVe and Word2Vec for e-commerce. We showed that *item embedding* using GloVe is most effective for fine grained e-commerce item analogy and similarity tasks. Furthermore, we claim that using *item embedding* vectors, as a compressed meaningful input representation will be efficient for various recommender system applications. In particular, we demonstrated its efficiency for session based recommendation comparing to state of the art item-to-item similarity model using ANN and RNN model with item embedding as input layer. RNN which preserves the order of user's clicks outperforms for this task.

We plan to expand our approach to additional recommender system use cases where the order of events is important such as predicting whether a session is a buying session [1] for promotion optimization.

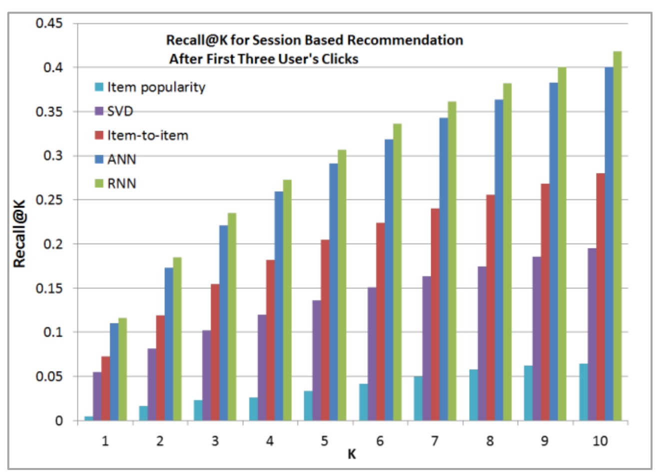

Figure 2. Recall@K for session based recommendation

Method	*Recall@10*	*Precision@10*	*MRR@10*
Item popularity	0.064	0.065	0.065
SVD	0.195	0.198	0.234
Item-to-item	0.280	0.285	0.281
ANN	0.387	0.391	0.331
RNN	0.418	0.422	0.352

Table 4. Session based recommendation evaluation

REFERENCES

1. Bogina, V., Kuflik, T. and Mokryn, O., 2016, March. Learning Item Temporal Dynamics for Predicting Buying Sessions. In Proceedings of the 21st International Conference on Intelligent User Interfaces (pp. 251-255). ACM.
2. Cho, K., van Merriënboer, B., Bahdanau, D. and Bengio, Y., 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.
3. Ganguly, D., Roy, D., Mitra, M. and Jones, G.J., 2015, August. Word Embedding based Generalized Language Model for Information Retrieval. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 795-798). ACM.
4. Hidasi, B., Karatzoglou, A., Baltrunas, L. and Tikk, D., 2015. Session-based Recommendations with Recurrent Neural Networks. arXiv preprint arXiv:1511.06939.
5. Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
6. Nalisnick, E., Mitra, B., Craswell, N. and Caruana, R., 2016. Improving document ranking with dual word embeddings. In Proc. WWW. International World Wide Web Conferences Steering Committee, to appear.
7. Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global Vectors for Word Representation. In EMNLP (Vol. 14, pp. 1532-1543).
8. Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., 2001, April. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295). ACM.
9. Shani, G., Brafman, R.I. and Heckerman, D., 2002, August. An MDP-based recommender system. In Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence (pp. 453-460). Morgan Kaufmann Publishers Inc.
10. Zhou, G., He, T., Zhao, J. and Hu, P., 2015. Learning continuous word embedding with metadata for question retrieval in community question answering. In Proceedings of ACL (pp. 250-259)
11. Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review 10*, 2 (1992), 453-469.
12. How to Classify Works Using ACM's Computing Classification System. http://www.acm.org/class/how_to_use.html.
13. Klemmer, R.S., Thomsen, M., Phelps-Goodman, E., Lee, R. and Landay, J.A. Where do web sites come from? Capturing and interacting with design history. In *Proc. CHI 2002*, ACM Press (2002), 1-8.
14. Mather, B.D. Making up titles for conference papers. *Ext. Abstracts CHI 2000*, ACM Press (2000), 1-2.
15. Schwartz, M. *Guidelines for Bias-Free Writing*. Indiana University Press, Bloomington, IN, USA, 1995.
16. Zellweger, P.T., Bouvin, N.O., Jehøj, H., and Mackinlay, J.D. Fluid Annotations in an Open World. *Proc. Hypertext 2001*, ACM Press (2001), 9-18.