Demographic Prediction from Purchase Data Based on Knowledge-Aware Embedding

Yiwen Jiang[1,2,3](✉), Wei Tang[1,2,3], Neng Gao[1,3], Ji Xiang[3], and Yijun Su[2,3]

[1] State Key Laboratory of Information Security, CAS, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] Institute of Information Engineering, CAS, Beijing, China
{jiangyiwen,tangwei,gaoneng,xiangji,suyijun}@iie.ac.cn

Abstract. Demographic attributes are crucial for characterizing different types of users in developing market strategy. However, in retail scenario, individual demographic information is not often available due to the difficult manual collection process. Several studies focus on inferring users' demographic attribute based on their transaction histories, but there is a common problem. Hardly work has introduced knowledge for purchase data embedding. Specifically, purchase data is informative, full of related knowledge entities and common sense. However, existing methods are unaware of such external knowledge and latent knowledge-level connections among items. To address the above problem, we propose a Knowledge-Aware Embedding (KAE) method that incorporates knowledge graph representation into demographic prediction. The KAE is a multi-channel and item-entity-aligned knowledge-aware convolutional neural network that fuses frequency-level and knowledge-level representations of purchase data. Through extensive experiments on a real world dataset, we demonstrate that KAE achieves substantial gains on state-of-the-art demographic prediction models.

Keywords: Demographic prediction · Convolutional neural networks · Knowledge graph representation

1 Introduction

Knowing users' demographic attributes is significant for many companies and retailers to make market basket analysis [1], adjust marketing strategy [6], and provide personalized recommendations [16,24]. For example, Nike's basketball shoes target mainly a relatively young (age) and male (gender) users with enough purchasing power (income). Additionally, in recommender systems, demographic information have been wildly used to improve the quality of the systems and solve the cold start problem.

T. Gedeon et al. (Eds.): ICONIP 2019, CCIS 1143, pp. 297–308, 2019.
https://doi.org/10.1007/978-3-030-36802-9_32

Generally, the collection of users' demographic information is difficult as users are reluctant to offer their personal information in case of data breaches. Besides, in reality, there are only partial demographic attributes known for most of users and some users have no attributes at all, as Wang et al. [20] pointed out.

There are several works that have made effort on transaction history for demographic prediction. Wang et al. [20] were the first ones to investigate the power of uses' purchase data for demographic prediction, and they proposed a Structured Neural Embedding (SNE) model which uses bag-of-item (BOI) representation as shared embedding. Resheff et al. [14] concatenated sequence embedding of transaction data with relational data to form raw user representation for demographic prediction. Kim et al. [7] proposed an Embedding Transformation Network with Attention (ETNA) model to learn the representations from purchase data with also naive shared representations at the bottom.

Although they proposed several valuable methods and made some meaningful achievements, their works overlooked a significant information for embedding. Existing methods are unaware of purchase data external knowledge and cannot fully discover latent knowledge-level connections among items. Specifically, knowledge of what user buy and what user may buy are informative and full of knowledge entities. For example, a user has a piece of transaction record with the baby clothes bought that contains two knowledge entities: "baby","clothes". In fact, the user may also be interested in dried milk with high probability, which is strongly connected with the previous one in terms of common sense reasoning. However, previous methods are hardly able to discover latent knowledge-level connection. As a result, a user's transaction history embedding for demographic prediction will be narrowed down to a limited representation circle.

To address above problem, we propose a novel method that takes advantage of external knowledge for transaction history embedding, called Knowledge-aware embedding (KAE). Specifically, for a piece of transaction record, we first enrich its information by associating each item with a relevant entity in the knowledge graph. We also search and use the set of contextual entities of each entity (i.e., its immediate neighbors) to provide more complementary and distinguishable information. Then we fuse the item representation and knowledge-level representations of purchase data through convolutional neural networks and generate a knowledge-aware embedding. The intuitive understanding of the superiority of KAE is that it maintains the alignment of multiple representations for an purchase item and explicitly bridges different embedding spaces.

We conduct extensive experiments on a real-world benchmark dataset. The results show that KAE achieves prominent gains over the state-of-the-art methods, and prove that the usage of knowledge for transaction history embedding can bring significant improvement on all the previous models for demographic prediction. Overall, our contributions are as follows:

1. We make the first attempt to introduce the knowledge graph into demographic prediction based on purchase data.

2. We propose a novel KAE method that introduces external knowledge for transaction history embedding, and fuses item and knowledge-level representations in an aligned way through CNN.
3. We conduct extensive experiments on a real-world dataset to demonstrate the effectiveness of the proposed KAE method as compared with state-of-the-art baselines.

The rest of this paper is organized as follows. Section 2 summarizes the related works. Section 3 presents the formulation of problem. Section 4 introduces the proposed KAE in details. Section 5 talks about the experimental results. Finally in Sect. 6, we conclude our work.

2 Related Work

2.1 Demographic Prediction

Many works have been devoted to demographic prediction in different scenarios. Early work on demographic prediction are mostly based on the linguistics writing and speaking to predict demographic attributes [13,15]. Later, some works proposed to infer demographic attributes based on users' browsing history [9,17]. Then, with the fast development of online social networks and mobile computing technologies, large scale of user data are accumulated, which make it possible and also valuable to infer users' demographic attributes in network and mobile scenarios [4,12]. Also, location data have been used to predict demographic attributes [25].

Recently, some people make effort to use transaction history for demographic prediction. Wang et al. [20] first proposed to carry out demographic prediction on purchase data and constructed a SNE model to learn the shared representation based on BOI representation. Resheff et al. [14] concatenated sequence embedding of transaction data with structured relational data to form raw user representation, and obtained deep user representation through full-connected neural network. Raehyun et al. [7] proposed an ETNA model for demographic prediction which obtains task-specific representations through linear transformation based on shared BOI representation at the bottom. In this paper, on the basis of the previous work, we introduce knowledge graph for demographic prediction task.

2.2 Knowledge Graph Embedding

The goal of knowledge graph embedding is to learn a low-dimensional representation vector for each entity and relation that preserves the structural information of the original knowledge graph. At present, translation-based knowledge graph embedding methods have received great attention due to their concise models and superior performance. There are many scenarios that successfully apply knowledge graph embedding to achieve better performance, such as machine reading [22], text classification [19], and news recommendation [18]. Moreover, there are many translation-based methods included TransE [2], TransH [21], TransR [10] and TransR [10] used in this paper. Today, we make an attempt to employ knowledge graph embedding for demographic prediction via transaction data.

2.3 CNN for Representation Learning

Previous methods [7,14,20] usually represent purchasing data using an aggregation function or the BOI technique, i.e., taking goods counting statistics as the feature of purchasing history. However, these methods have insufficient representation ability for ignoring some potential information in purchasing history and are vulnerable to the sparsity problem, which leads to poor generalization performance. Inspired by the multi-channel feature of convolutional neural networks (CNN) and the successful application in the filed of computer vision [9] and sentence representation learning [3,18,23], we use CNN to fuse the multiple type of purchasing data representations.

3 Problem Formulation

In this section, we formulate the demographic prediction problem as follows. Let $D = [(x, y)]$ represents a given user, where $x = [I_1, I_2, \ldots]$ denotes a list of transaction items, each item I may be associated with an entity e in the knowledge graph. And $y = y_1, y_2, \ldots, y_K$ denotes the set of attribute label, and K is the number of attributes. According to the defined notations, we follow two types of problem used in [20]:

Partial Label Prediction: For the situation of users with partial demographic attributes. The objective is to learn a function to predict the remaining unknown attributes:

$$f : X, Y^L \to Y^U$$

where Y^L and Y^U denote the observed attributes and that to be predicted over the same set of users X respectively.

New User Prediction: For the situation of new users demographic prediction. The objective function is:

$$f : X^L, Y^L, X^N \to Y^N$$

where X^L and Y^L denote the purchase histories and attributes of labeled users, X^N and Y^N denote the purchase history and the attributes of the new users. Note that here $X^L \cap X^N = \emptyset$.

4 Knowledge-Aware Embedding

In this section, we introduce the proposed KAE method in detail. We first present the overall framework of KAE, then discuss the process of knowledge distillation from a knowledge graph, and finally talk about the representations combination through convolutional neural networks.

4.1 KAE Framework

The framework of KAE is illustrated in Fig. 1. For each user, his/her transaction history are specially processed to BOI representation, relevant entity embedding and contextual entities embedding. Then an extension of traditional CNN that allows flexibility in incorporating symbolic knowledge from a knowledge graph into BOI representation are employed to generate the final shared embedding for the next demographic prediction.

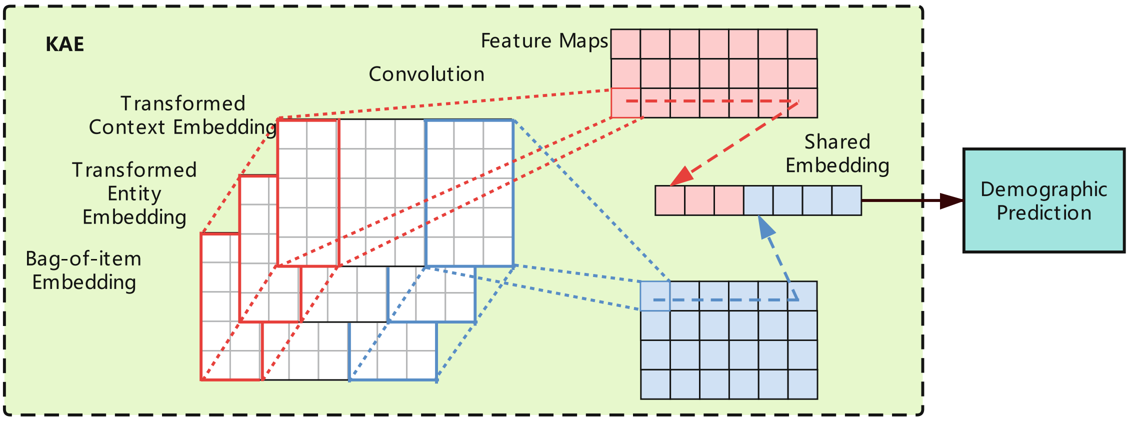

Fig. 1. Illustration of the KAE framework.

4.2 Knowledge Distillation

The process of knowledge distillation is showed in Fig. 2, which mainly consists of four steps. First, to distinguish knowledge entities in transaction history, we employ the technique of entity linking in [11,17] to disambiguate included in items by associating them with predefined entities in a knowledge graph. Based on these identified entities, we construct a sub-graph and extract all relational links among them from the original knowledge graph. There are several knowledge graphs available for academic and commercial, such as NELL[1], DBpedia[2], Google Knowledge Graph[3] and Microsoft Satori[4]. Note that we expand the knowledge sub-graph to all entities within one hop of identified ones in case of sparse relations and lacking diversity. Given the extracted knowledge graph, we apply some knowledge graph embedding methods, such as TransE [2], TransH [21], TransR [10], and TransD [5] for entity representation learning. Learned entity embedding are taken as the input for CNN.

Given that the collected raw transaction data may be insufficient and the information of learned embedding for only the identified entity in dataset is

[1] http://rtw.ml.cmu.edu/rtw/.
[2] https://wiki.dbpedia.org/.
[3] https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html.
[4] https://searchengineland.com/library/bing/bing-satori.

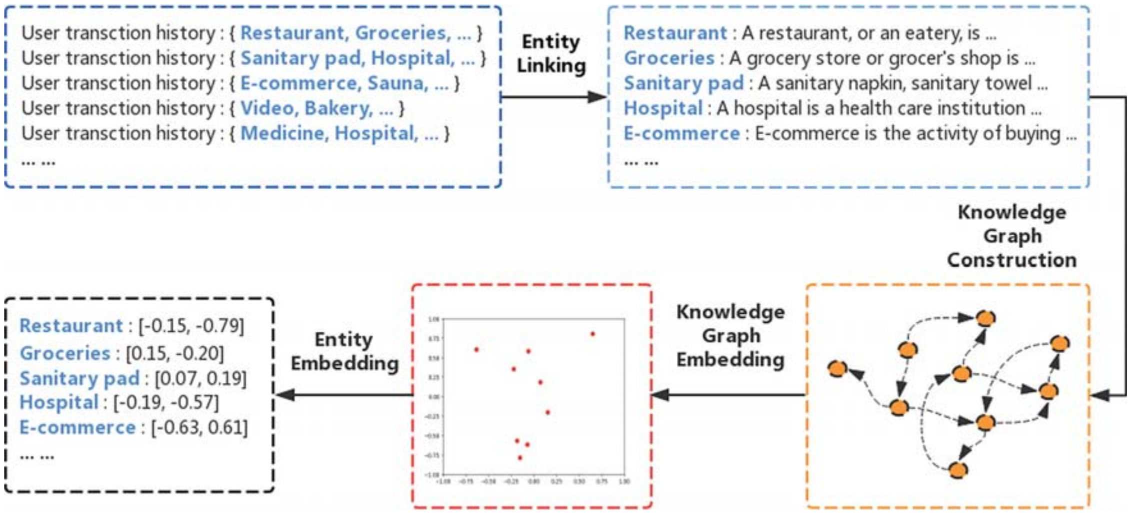

Fig. 2. Process of knowledge distillation.

limited for demographic prediction, we propose to extract additional contextual information for each entity, and definition of the "context" of entity e is the set of its immediate neighbors in the knowledge graph, ie.,

$$context(e) = \{e_i | (e, r, e_i) \in \mathcal{G} \quad or \quad (e_i, r, e) \in \mathcal{G}\}$$

where r is the relation between e and e_i, and $\mathcal{G}$ is the knowledge graph. Since the defined contextual entities closely associate with the current entity in semantics and logic, the usage of context could provide more complementary information and assist in improving the identifiability of entities. For a given entity e, the context embedding is calculated as the average of its contextual entities:

$$\bar{\mathrm{e}} = \frac{1}{|context(e)|} \sum_{e_i \in context(e)} \mathrm{e_i}$$

where $\mathrm{e_i}$ and $\bar{\mathrm{e}}$ are the embedding of entity e_i and its context learned by knowledge graph embedding respectively.

4.3 CNN-Based Embedding Combination

As the notations given in Sect. 3, we use $x = [I_1, I_2, \ldots, I_n]$ to denote the sequence of a user's n purchased items, and use $X = [\mathrm{I_1 I_2 \ldots I_n}] \in \mathbb{R}^{d \times n}$ to denote the item embedding matrix generated through BOI representation [20]. After the knowledge distillation introduced above, each item I_i may also be associated with an entity embedding $\mathrm{e_i} \in \mathbb{R}^{k \times 1}$ and the corresponding context embedding $\bar{\mathrm{e}}_i \in \mathbb{R}^{k \times 1}$, where k is the dimension of entity embedding.

Based on the three input given above, a straightforward approach to combine the different embedding is concatenating them to the new sequence as shared embedding for the further process. However, there are some limitations for this simple concatenating strategy: (1) It breaks up the connection between items and associated entities and is unaware of their alignment. (2) Item and entity embedding are learned by different methods, which means it is not suitable to deal with them together in a single vector space. (3) The concatenating strategy implicitly forces item embedding and entity embedding to share the same dimension, which may not be optimal in practical settings since the optimal dimensions for item and entity embedding may different.

In view of the above-mentioned limitations, we propose a CNN-based method for embedding combination, which combine items frequency and knowledge information in an aligned way. As showed in Fig. 1, we use item embedding $X = [\mathrm{I_1 I_2} \ldots \mathrm{I_n}]$, transformed entity embedding $\mathrm{g(e)} = [\mathrm{g(e_1)g(e_2)} \ldots \mathrm{g(e_n)}]$ and transformed context embedding $\mathrm{g(\bar{e})} = [\mathrm{g(\bar{e}_1)g(\bar{e}_2)} \ldots \mathrm{g}(\bar{\mathrm{e}}_n)]$ as input. Where $\mathrm{e_i}$ and $\bar{\mathrm{e}}_i$ are set to be zero if $\mathrm{I_i}$ has no corresponding entity. And g is the non-linear transformation function:

$$\mathrm{g(e)} = tanh(M\mathrm{e} + b)$$

where $M \in \mathbb{R}^{d\times k}$ is the trainable transformation matrix and $b \in \mathbb{R}^{d\times 1}$ is the trainable bias. Since the transformation function is continuous, the entity embedding and context embedding can be mapped from the entity space to the item space as well as retaining original spatial relationship. Thus, entity embedding g(e) and context embedding $\mathrm{g(\bar{e})}$ have the same size with item embedding X and can be served as the multiple channels analogous to colored images. We therefore align and stack the three embedding matrices as follow:

$$\mathrm{X} = [[\mathrm{I_1 g(e_1) g(\bar{e}_1)}], [\mathrm{I_2 g(e_2) g(\bar{e}_2)}] \ldots [\mathrm{I_n g(e_n) g}(\bar{\mathrm{e}}_n)]] \in \mathbb{R}^{d\times n\times 3}$$

when obtaining the multi-channel input X, we use multiple filter $h \in \mathbb{R}^{d\times l\times 3}$ with varying window sizes l to extract specific local patterns in the purchase history. The local activation of sub-matrix $\mathrm{X}_{i:i+l-1}$ about h can be written as:

$$c_i^h = f(\mathbf{h} * \mathrm{X}_{i:i+l-1} + b)$$

and a max-over-time pooling operation is used on the output feature to choose the largest feature:

$$\tilde{c}^h = max\{c_1^h, c_2^h, \ldots, c_{n-l+1}^h\}$$

we get the final shared embedding $e(i)$ of the i-th user by concatenating all feature $\tilde{c}^{h_i}$ together as follow:

$$e(i) = [\tilde{c}^{h_1}\tilde{c}^{h_2} \ldots \tilde{c}^{h_m}]$$

where m is the number of filters.

5 Experiments

5.1 Dataset

We use the transaction dataset came from previous work [7][5] and is collected by a Korean multi vendor loyalty program provider. The dataset consists of

[5] https://github.com/dmis-lab/demographic-prediction.

purchasing histories of 56,028 users with the gender, age, and marital status labels. The transaction records contain the information of user ID, company ID, type of purchased items and purchased items. The statistics of our dataset are summarized in Table 1.

Table 1. Distribution of users' attributes in dataset.

Attributes	Value	Distribution
Gender	Male	37%
	Female	63%
Age	Young	22.3%
	Adult	54.1%
	Middle age	14.3%
	Old	9.4%
Marital status	Married	19.9%
	Single	80.1%

As described in Sect. 3, we have two tasks in this paper. For the partial label prediction task, the goal is to predict the unknown attributes of users while the model is trained with the observed attributes. However, in our dataset, all demographic attributes are known for all the users. As a result, for partial prediction problem setting, we randomly set certain attributes as observed for training. For experiments on new user prediction task, we split our dataset into non overlapping sets, and the training, validation and testing split ratio are set to be 8:1:1.

5.2 Evaluation Metrics

We employ F-measure to evaluate our model. F-measure is a widely used measure method as a complement for accuracy, and it is the most popular evaluation metrics for demographic prediction. F1 score is calculated as the harmonic mean of precision and recall. The precision and recall are formulated differently depending on the following types of F1 score: micro, macro, and weighted. In our experiment, we use the weighted precision (wP), recall (wR) and F1 ($wF1$) score as the evaluation metrics since we consider all classes to be equal important. The weighted F1 is calculated as follows:

$$wP = \sum_{y \in Y} \left(\frac{\sum_{i=1}^{u} I(y_i^* = \hat{y}_i \& y = \hat{y}_i)}{\sum_{i=1}^{u} I(y = \hat{y}_i)} * weight \right)$$

$$wR = \sum_{y \in Y} \left(\frac{\sum_{i=1}^{u} I(y_i^* = \hat{y}_i \& y = \hat{y}_i)}{\sum_{i=1}^{u} I(y = y_i^*)} * weight \right)$$

$$wF1 = 2 * \frac{wP * wR}{wP + wR}$$

where $I(\cdot)$ is an indicator function, u denotes the total number of new users, Y is the set of all label combinations to be predicted, y_i^* denotes the ground truth of attributes for the i-th new user, $\hat{y}_i$ denotes the predicted attributes, and $weight = \frac{1}{u}\sum_{i=1}^{u} I(y = y_i)$. The weighted F1 assigns a high weight to the large classes to account for label imbalance.

5.3 Baseline Models

We compare our models with state-of-the-art case studies on demographic prediction:

SNE. Structured Neural Embedding [7,20] maps users' transaction histories into a shared embedding. This embedding is processed by average pooling and then fed into a log-bilinear model for structured predictions.

FMTD. In [14], the sequences embedding of transactions and structured relational vectors are concatenated as user representation, and the user representation is fed into a full connected layer to get a deep representation for prediction.

ETNA. Embedding Transformation Network with Attention [7] model uses a shared embedding at the bottom and feeds it into an embedding transformation layer to obtain the transformed representation. Then put the transformed representation into a task-specific attention layer to take into account the importance of each element in user profile for better prediction.

5.4 Experimental Settings

We implement the baseline models as described in [7] and FMTD [14]. We search all occurred entities in the dataset as well as the ones within one hop in the Microsoft Satori knowledge graph with confidence greater than 0.8 for all edges. The dimension of both item embedding and entity embedding are set as 128. The number of filters are set as 64 for window sizes 2 and 3. We use the Adam [8] with a mini-batch size of 64 and a learning rate of $1e^{-3}$. For partial label prediction, the observed attribute ratio is set as 50%.

5.5 Comparison of Different Models

The results of comparison of different models are shown in Table 2, where the KAE contains all three embeddings with TransD. We have some findings based on the experimental results.

1. The usage of KAE for embedding could boost the performance of all the baselines for both tasks. For example, in terms of ETNA, the wF1 increases from 56.9% to 64.8% in partial label prediction task and the wF1 increases from 36.0% to 43.8% in new user prediction task.

Table 2. The experimental results of partial label prediction and new user prediction.

Model	Partial label (50%)			New user		
	wP	wR	wF1	wP	wR	wF1
SNE	52.1%	56.3%	54.2%	29.5%	35.1%	32.1%
FMTD	43.4%	45.9%	44.6%	23.7%	29.1%	26.1%
ETNA	55.4%	58.4%	56.9%	33.9 %	38.2%	36.0%
KAE+SNE	59.8%	63.2%	61.5%	35.6%	38.3%	36.9%
KAE+FMTD	50.1%	52.4%	51.2%	27.1%	33.9%	30.1%
KAE+ETNA	63.9%	65.7%	64.8%	40.5%	47.6%	43.8%

"+" denotes the combination of KAE method with previous model.

2. After combining with KAE, the ETNA achieves the maximum gains among the three previous models. This is because ETNA integrates attention mechanism, when provides more knowledge, it has the ability to select valuable ones and abandon the noise for working better.
3. Comparing the wF1 gains between two tasks, we observe that the partial label prediction task obtains a little more improvement, the reason for which might be that observed partial attributes can be a good signal for CNN learning.

5.6 Comparison Among KAE Variants

We compare the variants of KAE on two aspects to demonstrate the efficacy of the design of the KAE framework: the usage of knowledge, and the choice of knowledge graph embedding method, where we combine with ETNA model for prediction. The results are shown in Table 3, from which we can conclude that:

Table 3. Comparison among KAE variants.

Variants	Partial label (50%)			New user		
	wP	wR	wF1	wP	wR	wF1
ETNA with I	55.4%	58.4%	56.9%	33.9%	38.2%	36.0%
KAE+ETNA with I	58.2%	61.7%	59.9%	36.8%	41.7%	39.1%
KAE+ETNA with I&E	62.7%	64.5%	63.6%	39.4%	45.9%	42.4%
KAE+ETNA with I&C	61.6%	62.8%	62.2%	38.1%	44.3%	41.0%
KAE+ETNA with I&E&C	63.9%	65.7%	64.8%	40.5%	47.6%	43.8%
KAE+ETNA with TransE	62.3%	64.3%	63.3%	39.3%	46.4%	42.6%
KAE+ETNA with TransH	62.7%	64.6%	63.6%	39.6%	46.7%	42.9%
KAE+ETNA with TransR	63.1%	65.1%	64.1%	40.0%	47.1%	43.3%
KAE+ETNA with TransD	63.9%	65.7%	64.8%	40.5%	47.6%	43.8%

"I", "E" and "C" denotes item, entity and context embedding respectively.

1. The use of CNN can improve wF1 by about 3% for both tasks. Moreover, the usage of entity embedding and contextual embedding can improve wF1 by 3.7% and 2.3% for partial label prediction as well as 3.3% and 1.9% for new user prediction. And we can attain better performance by combining them together.
2. TransD method used in KAE works better. Probably because TransD is the most complicated model among the four embedding methods, which is able to better capture non-linear relationships among the knowledge graph for demographic prediction.

6 Conclusion

In this paper, we propose a knowledge-aware embedding method that takes advantage of knowledge graph representation in demographic prediction. KAE breaks the traditional naive embedding methods on transaction history, and employs CNN to combine introduced entity and context embedding with item embedding. Experimental results demonstrate that KAE can boost performance of all the baselines.

References

1. Andrews, R.L., Currim, I.S.: Identifying segments with identical choice behaviors across product categories: an intercategory logit mixture model. Int. J. Res. Mark. **19**(1), 65–79 (2002)
2. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: 27th Annual Conference on NIPS 2013, United States, pp. 2787–2795 (2013)
3. Conneau, A., Schwenk, H., Barrault, L., LeCun, Y.: Very deep convolutional networks for natural language processing. CoRR abs/1606.01781 (2016)
4. Dong, Y., Yang, Y., Tang, J., Yang, Y., Chawla, N.V.: Inferring user demographics and social strategies in mobile social networks. In: The 20th ACM SIGKDD, KDD 2014, New York, NY, USA, pp. 15–24 (2014)
5. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: ACL 2015, Beijing, China, 26–31 July 2015, Volume 1: Long Papers, pp. 687–696 (2015)
6. Kalyanam, K., Putler, D.S.: Incorporating demographic variables in brand choice models: an indivisible alternatives framework. Mark. Sci. **16**(2), 166–181 (1997)
7. Kim, R., Kim, H., Lee, J., Kang, J.: Predicting multiple demographic attributes with task specific embedding transformation and attention network. CoRR abs/1903.10144 (2019)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: 3rd ICLR 2015 Conference Track Proceedings, San Diego, CA, USA (2015)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: 2012 26th Annual Conference on Neural Information Processing Systems, United States, pp. 1106–1114 (2012)
10. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the Twenty-Ninth AAAI, USA, pp. 2181–2187 (2015)

11. Milne, D.N., Witten, I.H.: Learning to link with Wikipedia. In: Proceedings of the 17th ACM, CIKM 2008, Napa Valley, California, USA, 26–30 October 2008, pp. 509–518 (2008)
12. Mislove, A., Viswanath, B., Gummadi, P.K., Druschel, P.: You are who you know: inferring user profiles in online social networks. In: Proceedings of the Third International Conference on WSDM
13. Otterbacher, J.: Inferring gender of movie reviewers: exploiting writing style, content and metadata. In: Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, pp. 369–378 (2010)
14. Resheff, Y.S., Shahar, M.: Fusing multifaceted transaction data for user modeling and demographic prediction. CoRR abs/1712.07230 (2017)
15. Schler, J., Koppel, M., Argamon, S., Pennebaker, J.W.: Effects of age and gender on blogging. In: AAAI Spring Symposium, Technical Report SS-06-03, Stanford, California, USA, 27–29 March 2006, pp. 199–205 (2006)
16. Sedhain, S., Sanner, S., Braziunas, D., Xie, L., Christensen, J.: Social collaborative filtering for cold-start recommendations. In: Proceedings of the 8th ACM Conference on Recommender Systems. ACM (2014)
17. Sil, A., Yates, A.: Re-ranking for joint named-entity recognition and linking. In: 22nd ACM, CIKM 2013, San Francisco, CA, USA, 27 October–1 November 2013, pp. 2369–2374 (2013)
18. Wang, H., Zhang, F., Xie, X., Guo, M.: DKN: deep knowledge-aware network for news recommendation. In: Proceedings of the 2018 WWW Conference on World Wide Web, WWW 2018, Lyon, France, 23–27 April 2018, pp. 1835–1844 (2018)
19. Wang, J., Wang, Z., Zhang, D., Yan, J.: Combining knowledge with deep convolutional neural networks for short text classification. In: 2017 Proceedings of the Twenty-Sixth IJCAI 2017, Melbourne, Australia, pp. 2915–2921 (2017)
20. Wang, P., Guo, J., Lan, Y., Xu, J., Cheng, X.: Your cart tells you: inferring demographic attributes from purchase data. In: WSDM, pp. 173–182 (2016)
21. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the Twenty-Eighth AAAI, Canada, pp. 1112–1119 (2014)
22. Yang, B., Mitchell, T.M.: Leveraging knowledge bases in LSTMs for improving machine reading. In: ACL 2017, Vancouver, Canada, 30 July–4 August 2017, Volume 1: Long Papers, pp. 1436–1446 (2017)
23. Zhang, X., Zhao, J.J., LeCun, Y.: Character-level convolutional networks for text classification. In: 2015 Annual Conference on Neural Information Processing Systems, 7–12 December 2015, Montreal, Quebec, Canada, pp. 649–657 (2015)
24. Zhao, X.W., Guo, Y., He, Y., Jiang, H., Wu, Y: We know what you want to buy: a demographic-based system for product recommendation on microblogs. In: Proceedings of the 20th ACM SIGKDD. ACM (2014)
25. Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F., Xie, X.: You are where you go: Inferring demographic attributes from location check-ins. In: 2015 Proceedings of the Eighth ACM, WSDM 2015, China, pp. 295–304 (2015)