Variational Bayesian Context-aware Representation for Grocery Recommendation

Zaiqiao Meng, Richard McCreadie, Craig Macdonald and Iadh Ounis
University of Glasgow, Scotland, UK
(firstname.lastname)@glasgow.ac.uk

ABSTRACT

Grocery recommendation is an important recommendation use-case, which aims to predict which items a user might choose to buy in the future, based on their shopping history. However, existing methods only represent each user and item by single deterministic points in a low-dimensional continuous space. In addition, most of these methods are trained by maximizing the co-occurrence likelihood with a simple Skip-gram-based formulation, which limits the expressive ability of their embeddings and the resulting recommendation performance. In this paper, we propose the Variational Bayesian Context-Aware Representation (VBCAR) model for grocery recommendation, which is a novel variational Bayesian model that learns the user and item latent vectors by leveraging basket context information from past user-item interactions. We train our VBCAR model based on the Bayesian Skip-gram framework coupled with the amortized variational inference, so that it can learn more expressive latent representations that integrate both the non-linearity and Bayesian behaviour. Experiments conducted on a large real-world grocery recommendation dataset show that our proposed VBCAR model can significantly outperform existing state-of-the-art grocery recommendation methods.

KEYWORDS

Context-Aware, Recommender Systems, Variational Bayesian, Skip-gram, Grocery Recommendation

ACM Reference Format:
Zaiqiao Meng, Richard McCreadie, Craig Macdonald and Iadh Ounis. 2019. Variational Bayesian Context-aware Representation for Grocery Recommendation. In ,. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/xxxxxxx.xxxxxxx

1 INTRODUCTION

Recommender systems that use historical customer-product interactions to provide customers with useful suggestions have been of interest to both academia and industry for many years. Various matrix completion-based methods [4, 12, 13] have been proposed to predict the rating scores of products (or items) for customers (or users). Recently, many grocery recommendation models [3, 15, 16] were proposed that target grocery shopping use-cases. In real grocery shopping platforms, such as Amazon and Instacart, users' interactions with items are sequential, personalized and more complex than those represented by a single rating score matrix. Thus effective recommendation models for this use-case are designed to learn representations of users and items so that contextual information, such as basket context [16] and time context [9], are captured within the learned representations, which results in increased recommendation performance. In the grocery shopping domain, prod2vec [3] and triple2vec [16] are two state-of-the-art models that learn latent representations capturing the basket context, based on the Skip-gram model for grocery recommendation. In these models, both the user's general interest (which items the user likes) and the personalized dependencies between items (what items the user commonly includes in the same basket) are encoded by the embeddings of users and items. Furthermore, when combined with negative sampling approaches [11], these Skip-gram-based models are able to scale to very large shopping datasets. Meanwhile, through the incorporation of basket contextual information during representation learning, significant improvements in grocery recommendation have been observed [3, 16].

However, these representation models still have several defects: (1) they represent each user and item by single deterministic points in a low-dimensional continuous space, which limits the expressive ability of their embeddings and recommendation performances; (2) their models are simply trained by maximizing the likelihood of recovering the purchase history, which is a point estimate solution that is more sensitive to outliers when training [1].

To alleviate the aforementioned problems, we propose a *Variational Bayesian Context-Aware Representation* model, abbreviated as *VBCAR*, which extends the existing Skip-gram based representation models for grocery recommendation in two directions. First, it jointly models the representation of users and items in a Bayesian manner, which represents users and items as (Gaussian) distributions and ensures that these probabilistic representations are similar to their prior distributions (using the variational auto-encoder framework [5]). Second, the model is optimized according to the amortized inference network that learns an efficient mapping from samples to variational distributions [14], which is a method for efficiently approximating maximum likelihood training. Having inferred the representation vectors of users and items, we can calculate the preference scores of items for each user based on these two types of Gaussian embeddings to make recommendations. Our contributions can be summarized as follows:

(1) We propose a variational Bayesian context-aware representation model for grocery recommendation that jointly learns probabilistic user and item representations while the item-user-item triples in the shopping baskets can be reconstructed.
(2) We use the amortized inference neural network to infer the embeddings of both users and items, which can learn more expressive latent representations by integrating both the non-linearity and Bayesian behaviour.

ACM RecSys, Workshop on Context-Aware Recommender Systems

(3) We validate the effectiveness of our proposed model using a real large grocery shopping dataset.

2 RELATED WORK

In this section, we briefly discuss two lines of related work, namely methods for grocery recommendation and deep neural network-based methods for recommendation.

A grocery recommender is a type of recommender system employed in the domain of grocery shopping to support consumers during their shopping process. The most significant difference between the grocery recommendation task and other recommendation tasks, such as video recommendation [13] and movie rating prediction [12], is that the basket contextual information is more common and important in grocery shopping scenarios. However, most existing matrix completion-based methods [4, 12, 13] are unable to incorporate such basket information. Hence, many approaches have been proposed to learn latent representations that incorporate the basket information to enhance the performance of grocery recommendation [3, 6, 16], among which Triple2vec [16] is one of the most effective. Triple2vec [16] is a recent proposed approach, which uses the Skip-gram model to capture the semantics in the users' grocery basket for product representation and purchase prediction. In this paper, we also apply the Skip-gram model to calculate the likelihood of the basket-based purchase history, but we further extend it to the Bayesian framework that represents users and items as Gaussian distributions and optimize them with the Amortized Inference [5, 14].

Besides the Skip-gram-based models, other deep neural network-based recommendation methods have also achieved success due to the highly expressive nature of deep learning techniques [4, 8, 17]. For instance, the Neural Collaborative Filtering [4] model is a general framework that integrates deep learning into matrix factorization approaches using implicit feedback. Meanwhile, Li et al. proposed a collaborative variational auto-encoder [7] that learns deep latent representations from content data in an unsupervised manner and also learns implicit relationships between items and users from both content and ratings. Additionally, to better capture contextual information, Manotumruksa et al. [9] proposed two gating mechanisms, i.e. a Contextual Attention Gate (CAG) and Time- and Spatial-based Gates (TSG), incorporating both time and geographical information for (venue) recommendation. In this work, to further enhance the expressive ability of the learned embeddings for grocery recommendation, we propose to use the variational auto-encoder-based deep neural network [5] to approximately optimize the variational lower bound.

3 METHODOLOGY

In this section, we first briefly introduce the basic notations and the problem that we plan to address (Section 3.1). Next, we briefly review the Skip-gram model as well as a state-of-the-art representation model called Triple2vec [16] tailored to grocery recommendation (Section 3.2). Then, we present our proposed representation learning model, i.e. Variational Bayesian Context-Aware Representation (VBCAR), as well as show how to use the learned embeddings for downstream recommendation tasks.

3.1 Problem Definition and Notations

We use $\mathcal{U} = \{u_1, u_2, \cdots, u_N\}$ to denote the set of users and $\mathcal{I} = \{i_1, i_2, \cdots, i_M\}$ to denote the set of items, where N is the number of users and M is the number of items. Then, in a grocery shopping scenario, the users' purchase history can be represented as $\mathcal{S} = \{(u, i, o) \mid u \in \mathcal{U}, i \in \mathcal{I}, o \in \mathcal{O}\}$ with $\mathcal{O} = \{o_1, o_2, \cdots, o_L\}$ being the set of orders (i.e. baskets). We also use $\mathbf{Z}^u \in \mathbb{R}^{N \times D}$ and $\mathbf{Z}^i \in \mathbb{R}^{M \times D}$ to denote the latent representation matrices for users and items, respectively, where D denotes the dimension of these latent variables.

Given $\mathcal{U}$, $\mathcal{I}$, $\mathcal{O}$ and $\mathcal{S}$, the task we aim to address in our paper is to infer the latent representation matrices of users and items, i.e. $\mathbf{Z}^u$ and $\mathbf{Z}^i$, so that the missing preference scores of items for each user that estimate future user purchase probabilities can be predicted (using these latent representation matrices).

3.2 Skip-gram and Triple2vec

The Skip-gram model was originally designed for estimating word representations that capture co-occurrence relations between a word and its surrounding words in a sentence [11]. It aims to maximize the log-likelihood of a target entity (word) v predicting contextual entities (words) C_v:

$$\log p(C_v \mid v) = \sum_{v' \in C_v} \log P\left(v' | v\right), \tag{1}$$

where $P(v' \mid v)$ is defined by the softmax formulation $P(v'|v) = \frac{\exp(f_v^T g_{v'})}{\sum_{v''} \exp(f_v^T g_{v''})}$ with f_v and $g_{v'}$ being the latent representations of the target entity and its contextual entities, respectively.

The Triple2vec [16] model further extends the Skip-gram model for capturing co-purchase product relationships within users' baskets according to sampled triples from the grocery shopping data. Here each triple reflects two items purchased by the same user in the same basket. Specifically, Triple2vec samples a set of triples $\mathcal{T} = \{(u, i, j) \mid (u, i, o) \in \mathcal{S}, (u, j, o) \in \mathcal{S}\}$ from the purchase history $\mathcal{S}$ as the purchase context for training and assumes that a triple $(u, i, j) \in \mathcal{T}$ is generated by a probability σ calculated by the function of $p((u, i, j) \mid \mathbf{z}_u^u, \mathbf{z}_i^i, \mathbf{z}_j^i)$:

$$\sigma = p((u, i, j) \mid \mathbf{z}_u^u, \mathbf{z}_i^i, \mathbf{z}_j^i) = P(i|j, u)P(j|i, u)P(u|i, j), \tag{2}$$

where $\mathbf{z}_u^u \in \mathbf{Z}^u$ and $\mathbf{z}_i^i, \mathbf{z}_j^i \in \mathbf{Z}^i$ are the latent representations of user u and items i and j, respectively, $P(i \mid j, u) = \frac{\exp\left(\mathbf{z}_i^{iT}(\mathbf{z}_j^i + \mathbf{z}_u^u)\right)}{\sum_{i'} \exp\left(\mathbf{z}_{i'}^{iT}(\mathbf{z}_j^i + \mathbf{z}_u^u)\right)}$ and $P(u \mid i, j) = \frac{\exp\left(\mathbf{z}_u^{uT}(\mathbf{z}_i^i + \mathbf{z}_j^i)\right)}{\sum_{u'} \exp\left(\mathbf{z}_{u'}^{uT}(\mathbf{z}_i^i + \mathbf{z}_j^i)\right)}$. The Skip-gram based models [3, 16] can learn representations for users and items at scale, and with the aid of basket information they have previously been shown to be effective for grocery recommendation. However, these models represent each user and item by single deterministic points in a low-dimensional continuous space, which limits the expressive ability of their embeddings and recommendation performance. To address this problem, we propose a new Bayesian Skip-gram model that represents users and items by Gaussian distributions, as illustrated in Section 3.3.1. Then, we describe how to approximately optimize the Bayesian Skip-gram model with a Variational Auto-encoder and the amortized Inference (Section 3.3.2). We provide an overview of our overall proposed model in Figure 1.

3.3 The Variational Bayesian Context-aware Representation Model

3.3.1 Bayesian Skip-gram Model. Here we present our proposed Variational Bayesian Context-aware Representation model, i.e. VBCAR, which represents the users and items as random variables, i.e. $\mathbf{Z}^u$ and $\mathbf{Z}^i$ that are independently generated according to their priors. Like other probabilistic methods for embedding [10] and recommender systems [4, 8], these priors are assumed to be the standard Gaussian distributions:

$$p\left(\mathbf{Z}^u\right) = \mathcal{N}\left(0, \alpha^2\mathbf{I}\right), \qquad p\left(\mathbf{Z}^i\right) = \mathcal{N}\left(0, \alpha^2\mathbf{I}\right) \tag{3}$$

where α^2 is the same hyperparameter for all the priors - we used the default setting of $\alpha = 1$ in our paper, following [5].

Consider past purchase triples $(u, i, j) \in \mathcal{T}$ that are sampled from historical grocery shopping data [16]. These sampled triples are *positive examples* that should be precisely predicted according to the latent variables of users and items. We use $n+$ to denote the number of times that a given triple is observed in the *total sample* $\mathcal{T}$.Then $n+$ is a *sufficient statistic* of the Skip-gram model, and it contributes to the likelihood $p(n+ \mid \mathbf{z}_u^u, \mathbf{z}_i^i, \mathbf{z}_j^i) = \sigma^{n+}$ [1]. Thus, one needs to also construct an associated rejected triples set (i.e. $n-$, *negative examples*) that are not in the total sample so that we can conduct an efficient *negative sampling* for approximate optimization [11]. We let $n\pm = \{n+, n-\}$ be the combination of both positive and negative examples, then the likelihood of this complete purchase context is obtained by:

$$\log p\left(n\pm|\mathbf{Z}^u, \mathbf{Z}^i\right) = \sum_{(v_i, v_j, u_u) \in \mathcal{T}} \log \sigma + \sum_{(v_i, v_j, u_u) \notin \mathcal{T}} \log(1-\sigma), \tag{4}$$

where σ is calculated by the same function as in Triple2vec [16] (i.e. $p((u, i, j) \mid \mathbf{z}_u^u, \mathbf{z}_i^i, \mathbf{z}_j^i)$ in Equation (2)).

3.3.2 The Variational Evidence Lower Bound and Amortized Inference. Since we assume that both $\mathbf{Z}^u$ and $\mathbf{Z}^i$ are random variables, the exact inference of their posterior density is intractable due to the non-differentiable marginal likelihood $p\left(n^{\pm}\right)$ [5]. Variational Bayes resolves this issue by constructing a tractable lower bound of the logarithm marginal likelihood and maximizing the lower bound instead [2]. Following the Variational Autoencoding framework [5], we also solve this problem by introducing the two variational distributions to formulate a tractable lower bound and optimize the lower bound by the Amortized Inference [14]. To infer the users' and items' embedding, we start by formulating the logarithm marginal likelihood of $n^{\pm}$:

$$\begin{aligned}
\log p\left(n^{\pm}\right) &= \log \mathbb{E}_{q_\phi(\mathbf{Z}^u, \mathbf{Z}^i)}\left[\frac{p\left(n^{\pm}, \mathbf{Z}^u, \mathbf{Z}^i\right)}{q_\phi\left(\mathbf{Z}^u, \mathbf{Z}^i\right)}\right] \\
&\geq \mathbb{E}_{q_\phi(\mathbf{Z}^u, \mathbf{Z}^i)}\left[\log \frac{p\left(n^{\pm}, \mathbf{Z}^u, \mathbf{Z}^i\right)}{q_\phi\left(\mathbf{Z}^u, \mathbf{Z}^i\right)}\right] \\
&= \mathbb{E}_{q_\phi(\mathbf{Z}^u, \mathbf{Z}^i)}\left[\log p\left(n^{\pm} \mid \mathbf{Z}^u, \mathbf{Z}^i\right)\right] \\
&\quad - KL\left(q_\phi\left(\mathbf{Z}^u, \mathbf{Z}^i\right) \| p(\mathbf{Z}^u, \mathbf{Z}^i)\right) \\
&\stackrel{\text{def}}{=} \mathcal{L},
\end{aligned} \tag{5}$$

where the inequation of the second line is derived from the Jensen's inequality; $\mathcal{L}$ is called the Evidence Lower BOund (ELBO) of the observed triple context [5]; $KL(\cdot\|\cdot)$ is the Kullback-Leibler (KL) divergence and $q_\phi(\mathbf{Z}^u, \mathbf{Z}^i)$ is the variational distribution, which can be factorized in a mean-field form:

$$q_\phi(\mathbf{Z}^u, \mathbf{Z}^i) = q_{\phi_1}(\mathbf{Z}^u) q_{\phi_1}(\mathbf{Z}^i), \tag{6}$$

where ϕ_1 and ϕ_2 are the trainable parameters of the inference models (encoders). In order to get more expressive latent factors of users and items, we consider that the variational distributions $q_{\phi_1}(\mathbf{Z}^u)$ and $q_{\phi_2}(\mathbf{Z}^i)$ are Gaussian distributions and are encoded from the identity codes of users and items such that we have:

$$q_{\phi_1}\left(\mathbf{Z}^u \mid \mathbf{F}^u\right) = \mathcal{N}\left(\boldsymbol{\mu}^u, \boldsymbol{\sigma}^{u2}\mathbf{I}\right), \tag{7}$$

$$q_{\phi_2}\left(\mathbf{Z}^i \mid \mathbf{F}^i\right) = \mathcal{N}\left(\boldsymbol{\mu}^i, \boldsymbol{\sigma}^{i2}\mathbf{I}\right), \tag{8}$$

where $\mathbf{F}^u \in \mathbb{R}^{N \times F_1}$ and $\mathbf{F}^i \in \mathbb{R}^{M \times F_2}$ are the identity representation (can be one-hot or binary encoded) of users and items respectively, with F_1 and F_2 being the dimension of their identity representation respectively, and $\boldsymbol{\mu}^u$, $\boldsymbol{\sigma}^{u2}$, $\boldsymbol{\mu}^i$ and $\boldsymbol{\sigma}^{i2}$ are inferred by the encoder networks. Specifically, the parameters of these Gaussian embeddings are encoded from their identity codes, i.e. $\mathbf{F}^u$ and $\mathbf{F}^i$, according to two two-layer fully-connected neural networks:

$$[\boldsymbol{\mu}^u, \boldsymbol{\sigma}^{u2}\mathbf{I}] = \mathbf{W}_2^u \mathbf{tanh}(\mathbf{W}_1^u \mathbf{F}^u + \mathbf{b}_1^u) + \mathbf{b}_2^u, \tag{9}$$

$$[\boldsymbol{\mu}^i, \boldsymbol{\sigma}^{i2}\mathbf{I}] = \mathbf{W}_2^i \mathbf{tanh}(\mathbf{W}_2^i \mathbf{F}^i + \mathbf{b}_2^i) + \mathbf{b}_2^i, \tag{10}$$

where **tanh** is the non-linearity activation function, and $\mathbf{W}_1^u$, $\mathbf{W}_2^u$, $\mathbf{b}_1^u$, $\mathbf{b}_2^u$, $\mathbf{W}_1^i$, $\mathbf{W}_2^i$, $\mathbf{b}_1^i$ & $\mathbf{b}_2^i$ are trainable parameters of the neural networks.

Since we assume the priors and the variational posteriors are Gaussian distributions, the KL-divergence terms in Equation (5) have analytical forms. By using the Stochastic Gradient Variational Bayes (SGVB) estimator and the reparameterization trick [5], we can directly optimize ELBO by sampling deterministic and differentiable embedding samples from the inferred variational distributions:

$$\begin{aligned}
\mathbf{Z}^u &= \boldsymbol{\mu}^u + \boldsymbol{\sigma}^{u2} \odot \boldsymbol{\epsilon}^{(l)}, \boldsymbol{\epsilon}^{(l)} \sim \mathcal{N}(0, \mathbf{I}), \\
\mathbf{Z}^i &= \boldsymbol{\mu}^i + \boldsymbol{\sigma}^{i2} \odot \boldsymbol{\epsilon}^{(l)}, \boldsymbol{\epsilon}^{(l)} \sim \mathcal{N}(0, \mathbf{I}),
\end{aligned} \tag{11}$$

to approximate and regularize maximum likelihood training, which is also referred to as **amortized inference** [14].

3.4 Recommendation Tasks

Our model infers the embeddings of both users and items according to the variational auto-encoder and represents them by means of their variational Gaussian distributions. Since we have taken advantage of the basket information, having obtained the embedding

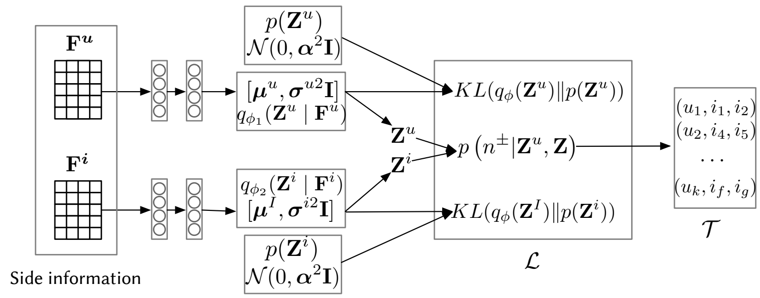


Figure 1: The architecture of our proposed VBCAR model. The model takes the user and item one-hot identity representation, i.e. F^u and F^i, as input and outputs Gaussian distributions with means and variances as latent embeddings for all users and items. The model then uses the deterministic variables Z^u and Z^i, reparameterized from their Gaussian distributions, to predict the sampled triples.

of users and items in our VBCAR model, we can follow a similar approach to [16] tackling both next-basket product recommendation and within-basket product recommendation:

(1) Next-basket product recommendation: Recommending a given user u with products for the next basket, we can obtain a preference score $s_{ui} = \textbf{dot}\left(\mathbf{z}_u^u, \mathbf{z}_i^i\right)$[1] for each item i, then return the top-K items with the highest preference scores.
(2) Within-basket product recommendation: If the products in the current basket b are given, we can first compute a preference score of item i for user u by: $s_{ui} = \textbf{dot}(\mathbf{z}_u^u + \sum_{i' \in b} \mathbf{z}_i^{i'}, \mathbf{z}_i^i)$, then return the top-$K$ preference score items as recommendations.

In this paper, we only evaluate the performance of our model based on the next-basket product recommendation and leave the evaluation of within-basket product recommendation for future work.

4 EXPERIMENTS

In the following, we first introduce the research questions we aim to answer in this paper (Section 4.1). Next, we describe our experimental setup (Section 4.2), followed by our results and analysis (Section 4.3).

4.1 Research Questions

In this paper, we aim to answer the following two research questions:

(**RQ**1) Can our proposed model outperform the Triple2vec model for grocery recommendation?
(**RQ**2) Can our Bayesian model learn more expressive representations of users and items than Triple2vec?

4.2 Experimental Setup

Dataset. We evaluate our model using the **Instacart** [16] dataset, which is a public large grocery shopping dataset from the Instacart Online Grocery Shopping Website[2]. This dataset contains over 3 million grocery orders and 33.8 million interactions from 0.2 million users and 50 thousand items. We first clean the dataset by filtering users and items using a number of thresholds. In particular, users that have less than 7 orders or less than 30 items, as well as items that were purchased by less than 16 users in the purchase history were removed. Next, we uniformly sample different percentages of users and items to construct different sizes of evaluation dataset. For model evaluation, we split all the sampled datasets into training (80%) and testing (20%) sets according to the temporal order of baskets. Table 1 shows the statistics of these datasets.

Table 1: Statistics of the datasets in used in our experiments.

Percentage	#Users	#Items	#Orders	#Interactions
5%	47,207	5,679	1,441	354,946
10%	154,285	11,888	3,124	1,103,361
25%	527,431	46,850	9,174	4,010,904
50%	1,186,957	59,549	16,121	12,217,555
100%	2,741,332	119,098	32,243	29,598,689

Baseline and Evaluation Metrics. To provide a fair comparison, we use Triple2vec [16] as a state-of-the-art baseline, since Triple2vec incorporates basket information in a similar way to our proposed VBCAR model. We evaluate the effectiveness of our model for next-basket (grocery) product recommendation, where we evaluate the top-K items recommended by each model. We report the standard recommendation evaluation metrics Recall@K and NDCG@K [4, 16, 17] to evaluate the preference ranking performance. We report results for $K = 10$ in our subsequent experiments, however we observed similar results when testing other values of K (e.g. 5 and 20).

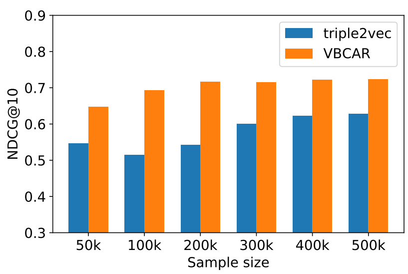
4.3 Results and Analysis

To answer **RQ1**, we evaluate our model as well as the triple2vec baseline on the task of item recommendation with the same size of triple samples (i.e. 1 million). Table 2 shows the overall performance of our proposed model as well as that of the baseline method. For both the Triple2vec and our proposed VBCAR approach, we empirically set the embedding size to be 64 and train both models, with a batch size of 512 and a RMSprop optimizer. From Table 2, we can clearly see that our VBCAR model performs better than Triple2vec on all the datasets. This result suggests that our model can learn more expressive latent representations by integrating both non-linearity and a Bayesian behaviour.

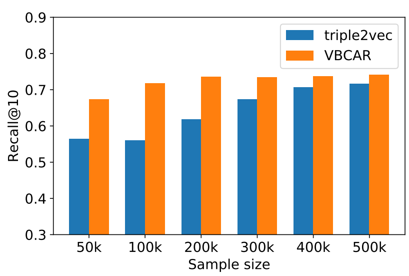
[1] **dot** is the dot product for two vectors.
[2] https://www.instacart.com/datasets/grocery-shopping-2017

Table 2: Overall performance on item recommendation. The best performing result is highlighted in bold; and * denotes a significant difference compared to the baseline result, according to the paired t-test p < 0.01.

Dataset	Triple2vec		VBCAR	
	NDCG@10	Recall@10	NDCG@10	Recall@10
5%	0.557	0.708	**0.731***	**0.748***
10%	0.558	0.664	**0.723***	**0.720***
25%	0.626	0.608	**0.686***	**0.628***
50%	0.708	0.525	**0.719***	**0.643***
100%	0.726	0.660	**0.768***	**0.742***



(a) NDCG@10 performance by different triple size



(b) Recall@10 performance by different triple size

Figure 2: Performance comparison for the various sample sizes on 5% of the Instacart data.

To further validate this argument (**RQ2**), we also compare the recommendation performance of our VBCAR model with Triple2vec using a different number of triple samples. Figure 2 shows the NDCG@10 and Recall@10 performances for triple sample sizes ranging from 50k to 500k. In these experiments, we set the embedding dimension to 64, while the other parameters for both models are tuned to be optimal except the fixed triple sample size. Again, we can clearly observe that our VBCAR model outperforms Triple2vec in terms of both metrics and on all triple sample sizes. Moreover, the gap between the performance of VBCAR model and Triple2vec is larger on small sample sizes. This result validates our hypothesis that our VBCAR model can learn more expressive latent representations with limited input samples.

5 CONCLUSIONS

In this paper, we have proposed the VBCAR model, a variational Bayesian context-aware representation model for grocery recommendation. Our model was built based on the variational Bayesian Skip-gram framework coupled with the amortized inference. Experimental results on the Instacart dataset show that our VBCAR model can learn more expressive representations of users and items than Triple2vec and does significantly outperform Triple2vec under both the NDCG and Recall metrics. Indeed, we observe up to a 31% increase in recommendation effectiveness over Triple2vec (under NDCG@10). For future work, we plan to extend our model to infer latent representations for new users and new items by taking the side information about users and items into account.

Acknowledgements

The research leading to these results has received funding from the European Community's Horizon 2020 research and innovation programme under grant agreement n° 779747.

REFERENCES

[1] Oren Barkan. 2017. Bayesian neural word embedding. In *Thirty-First AAAI Conference on Artificial Intelligence*. 3135–3143.
[2] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational Inference: A Review for Statisticians. *J. Amer. Statist. Assoc.* 112, 518 (2017), 859–877.
[3] Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Jaikit Savla, Varun Bhagwan, and Doug Sharp. 2015. E-commerce in Your Inbox: Product Recommendations at Scale. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1809–1818.
[4] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. 173–182.
[5] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014*. 1–14.
[6] Duc Trong Le, Hady W Lauw, and Yuan Fang. 2017. Basket-sensitive Personalized Item Recommendation. In *Proceedings of the 26th International Joint Conference on Artifical Intelligence*. 2060–2066.
[7] Xiaopeng Li and James She. 2017. Collaborative Variational Autoencoder for Recommender Systems. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 305–314.
[8] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. 689–698.
[9] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. 2018. A Contextual Attention Recurrent Architecture for Context-aware Venue recommendation. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 555–564.
[10] Zaiqiao Meng, Shangsong Liang, Hongyan Bao, and Xiangliang Zhang. 2019. Co-embedding Attributed Networks. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. 393–401.
[11] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.
[12] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic Matrix Factorization. In *Advances in Neural Information Processing Systems*. 1257–1264.
[13] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th conference on uncertainty in artificial intelligence*. 452–461.
[14] Rui Shu, Hung H Bui, Shengjia Zhao, Mykel J Kochenderfer, and Stefano Ermon. 2018. Amortized Inference Regularization. In *Advances in Neural Information Processing Systems*. 4393–4402.
[15] Mengting Wan, Di Wang, Matt Goldman, Matt Taddy, Justin Rao, Jie Liu, Dimitrios Lymberopoulos, and Julian McAuley. 2017. Modeling Consumer Preferences and Price sensitivities From Large-scale Grocery Shopping Transaction Logs. In *Proceedings of the 26th International Conference on World Wide Web*. 1103–1112.
[16] Mengting Wan, Di Wang, Jie Liu, Paul Bennett, and Julian McAuley. 2018. Representing and Recommending Shopping Baskets with Complementarity, Compatibility and Loyalty. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1133–1142.
[17] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*.