Receipt Embedding and Shopping Purpose Segmentation

Yinxing Li[1]* and Nobuhiko Terui [2]
[1]Graduate School of Economics and Management, Tohoku University, Japan
yinxing.li.a8@tohoku.ac.jp
[2]School of Management, Tokyo University of Science, Japan
terui@rs.tus.ac.jp

1. Introduction

A battery of studies has applied the Word2Vec model by Mikolov et al. (2013) to marketing problem by using large-scale shopping data, such as Prod2Vec (Grbovic et al. 2015), Item2Vec (Barkan and Koenigstein, 2016), and Meta-Prod2Vec (Vasile et al., 2016). They show that the framework of Word2Vec outperforms existing models in the prediction of sales. However, these existing approaches lack the interpretability of the model since the Word2Vec framework cannot evaluate the effect of variables, which may limit its use in the marketing, such as the effective personalization and targeting (Essex, 2009). The extended models, such as Prod2Vec, include various marketing variables such as price and customer demographic data, however, the role of the variables is still not discussed.

Li and Terui (2021) extend these studies to propose a LDA2Vec (Moody, 2016) framework that incorporates marketing variables to have the purposes: (i) improving the precision of forecasts by embedding the marketing environment to the Word2Vec framework with marketing mix variables, (ii) investigating the role of the marketing mix variables, and (iii) distinguishing different types of customers with the hierarchical structure of context vector. Compared to the previous studies, it produces higher forecasting precision by incorporating the marketing environment and customer heterogeneity in the model and it provides richer interpretability with a hierarchical model.

Ruiz et al. (2019) proposed the sequential probabilistic model of shopping basket named Shopper, and it has the utility and economics concept of substitutes and complements of products in the store and proposes the comprehensive choice model based on the consumer preference and by scanning large-scale shopping data. Their model is similar in spirit to Product2Vec and each possible product is associated with latent attributes with vector representations that are learned from the data. Shoppe also contains thinking ahead algorithm for modeling shopping behavior by considering the shopping context.

Sharing the similar spirits with shopper and extending the model of Li and Terui (2021), this study proposes a machine learning model by extending Shopper in several directions: (i) embedding receipt that is characterized as multiple purchase in a shopping trip. Shopper introduces thinking ahead algorithm to represent shopping context, however, we directly embed the receipt to derive vector representation by summing the item vectors in a basket. This dramatically reduces the computational costs, and derives implications by vector representation of receipt, that is, (ii) identifying shopping purpose of customers by clustering receipt vectors, (iii)identifying price and seasonal effects. Shopper only include price variable as covariate and price response is generally changing on the season when each item is demanded and the seasonal term could be a confounder to identify the price effect.

We explain our model in Section 2. Section 3 presents the empirical results for forecasting sales, and demonstrates the model performance and interpretability, and shows a managerial implication by a simulation study. Finally, we conclude our study in Section 4.

2. Model

We denote the vector of pivot item *i* and the vector of target item *j* as $\vec{u}_i$, and $\vec{v}_j$ respectively, then the probability that the customer *h* purchases the target item *j* conditional on the pivot item *i* in the basket is defined as

$$\mathrm{p}\left(\vec{v}_j | \vec{u}_i, h\right) = \sigma\left({\vec{u}_i}^T \vec{v}_j + \varphi_{hj}\right) \prod_{k=1}^{N} \sigma\left(-{\vec{u}_i}^T \vec{v}_k - \varphi_{hk}\right), \tag{1}$$

where

$$\varphi_{hj} = \beta_j + \boldsymbol{\gamma}'_{\boldsymbol{h}} \boldsymbol{Z}_{\boldsymbol{h}} + \epsilon_\varphi, \quad \epsilon_\varphi \sim N\left(0, \Lambda_\varphi\right). \tag{2}$$

Similar to Mikolov et al. (2013), we employ negative sampling method for our main equation, where $\sigma(\cdot)$ stands for the sigmoid function, φ_{hj} represents the utility when customer *h* purchase the item *j*, and *N* is the number of negative samples in the equation (1). β_j in the equation (2) means the intercept of item *j*, $\boldsymbol{Z}_{\boldsymbol{h}}$ is the demographic data of customer *h*, and $\boldsymbol{\gamma}_{\boldsymbol{h}}$ is the coefficients vector. Next, we define the receipt vector S_{hm} as

$$S_{ht} = \sum_{i=1}^{M_{ht}} \vec{u}_i \,. \tag{3}$$

That is, the receipt vector is calculated by summing all the item vectors in the t^{th} receipt of customer *h*. Considering the receipt in each trip for customer *h*, we design the prior structure for the receipt vector in the form of generalized state space model as

$$\begin{cases} \boldsymbol{S}_{\boldsymbol{ht}} = \boldsymbol{\mu}_{\boldsymbol{t}} + \boldsymbol{F}_{\boldsymbol{t}} \boldsymbol{Q}_{\boldsymbol{t}} + \boldsymbol{\epsilon}_{\boldsymbol{t}} \\ \boldsymbol{\mu}_{\boldsymbol{t}} = \boldsymbol{\mu}_{\boldsymbol{t-1}} + \boldsymbol{w}_{\boldsymbol{t}} \\ \boldsymbol{F}_{\boldsymbol{t}} = \boldsymbol{F}_{\boldsymbol{t-1}} + \boldsymbol{v}_{\boldsymbol{t}} \end{cases}, \tag{4}$$

where $\boldsymbol{\mu}_{\boldsymbol{t}}$ is intercept vector, $\boldsymbol{Q}_{\boldsymbol{t}}$ is marketing variable, including seasonal dummy, discount and promotion variables, $\boldsymbol{F}_{\boldsymbol{t}}$ means time varying coefficients vector. $\boldsymbol{\epsilon}_{\boldsymbol{t}}, \boldsymbol{w}_{\boldsymbol{t}}$ and $\boldsymbol{v}_{\boldsymbol{t}}$ are error terms in the model (4) assuming to follow Gaussian distribution with corresponding dimensions, respectively. This structure allows us to investigate the influence of the marketing environment on the preference shift during trips.

3. Data and Empirical Results

3.1 Data

We applied our model to daily scanner sales data from a store in Japan. The data were recorded between January 2, 2000, and December 5, 2001. There were 56,630 receipts generated by 1,476 unique customers. The dataset included 11,983 unique items, and the mean number of items in each receipt was 8.83. We used binary factors “discount,” “promotion,” and “weekday” as marketing environment variables, and 12 customer demographic variables, including age, family members, job, etc.. The details of the demographic data are presented in Table 1.

Table 1. Demographic variables

Variables	Type	Description
age	numeric	The age of the customer.
family	numeric	The number of family members of the customer.
time	numeric	The time cost for arriving at the store.
walk	dummy	If the customer walks to the store.
bike/bicycle	dummy	If the customer uses a bike or bicycle.
Car (no drive)	dummy	If the customer reaches the store by car, but not as a driver.
car(drive)	dummy	If the customer drive to the store.
parttime	dummy	If the customer has a part-time job.
fulltime	dummy	If the customer has a full-time job.
unknown	dummy	If the job of the customer is unknown.
housework	dummy	If the customer is a homemaker.
Work home	dummy	If the customer works at home.

3.2 Results

For the model evaluation, we left last one receipt for each customer, and calculate Hit Rate@k as the hold-out samples. We used Item2Vec as the benchmark model in this study. Besides, we also compared proposed RE (Receipt Embedding) with and without prior structure with marketing environment. Table 2 shows the results for the hold-out samples. We note that we empirically define the embedding dimension as 10 after comparing the performance of proposed models with various alternative dimensions in this study.

Table 2. Model Evaluation

Model	Hit Rate@5
Item2Vec	13.281%
RE without Prior	14.918%
RE with Prior	18.129%

According to the results, our proposed model, especially with the prior structure with marketing environment outperforms other models, and it implies the dynamic structure for the receipt vector plays an important role when forecast for the next trip.

Besides, we investigate the effects of marketing environment on the receipt vector. Figure 1 shows the average discount effect by each season. Season 1 in the figure means the period from January to March, and the period from April to June is represented by Season 2, and the same principle applies to the rest of the seasons. The x axis stands for the 10 embedding dimensions, which represent the different preference of the item. The average discount effect is defined as the proportion of absolute value of discount coefficient among those of all dimensions. The result shows that the influence of the discount varies according to season and customer's purchasing purpose represented by embedding dimensions.

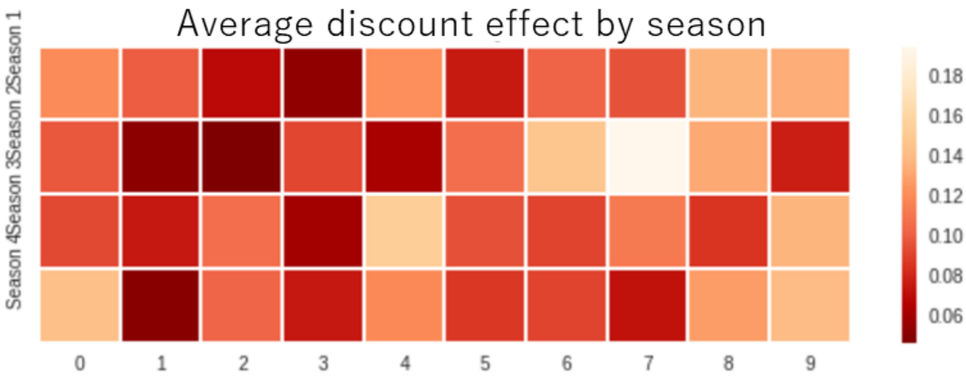

Figure 1. Average discount effect by season

4. Concluding Remarks

We proposed a novel approach in this study by involving the concept of receipt vector into the Item2Vec framework, as well as the prior structure which represents the dynamic preference shift of the receipt in terms of generalized state space model.

Our study highlighted the importance of the marketing environment when forecast the market basket for the future trips. The results of empirical results help managers understand the purchasing patterns and preference shift for a certain customer in different marketing environment. Several issues remain. The interpretation of each dimension for the item and receipt are still very challenging. Next, we use full Bayesian approach for the proposed model, and we improve computation efficiency by employing the parallel computing. However, the computation cost still highly rely on embedding dimension, and the computation cost significantly increase when we increase embedding dimension with big-scale data. We leave these issues for the future work.

Reference

Barkan, O. and Koenigstein, N. (2016). Item2vec: Neural Item Embedding for Collaborative Filtering. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing, 1-6.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003). Latent Dirichlet Allocation. The Journal of Machine Learning Research, 3, 993–1022.

Caselles-Dupré, H., Lesaint, F. and Royo-Letelier J. (2018). Word2Vec applied to Recommendation: Hyperparameters Matter. Proceedings of the 12th ACM Conference on Recommender Systems, 352-356.

Essex, D. (2009). Matchmaker, matchmaker. Communications of the ACM, 52(5),16–17.

Grbovic, M., Radosavljevic, V., Djuric, N., Bhamidipati, N., Savla, J., Bhagwan, V. and Sharp. (2015). D. E-commerce in your inbox: Product recommendations at scale. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’15, 1809–1818.

Koren Y, Bell R, Volinsky C. (2009). Matrix factorization techniques for recommender systems. Computer, 1(8), 30-7.

Levy, O. and Goldberg, Y. (2014). Dependency-Based Word Embeddings. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, (2), 302–308.

Li, Y. and Terui, N. (2021). Product Embedding for Large-Scale Disaggregated Sales Data. In Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. 1, 69-75.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Proceedings of the International Conference on Learning Representations,2013, 1-13.

Moody, C. (2016). Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec.

Naik, P., Wedel M., Bacon, L. and Bodapati, A. (2008). Challenges and opportunities in high-dimensional choice data analyses. Marketing Letters, 19(3), 201-213.

Paquet, U. and Koenigstein, N. (2013). One-class collaborative filtering with random graphs. In Proceedings of the 22nd international conference on World Wide Web, 999-1008.

Pennington, J., Socher, R. and Manning, C. (2014). Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language, 1532–1543.

Francisco J. R. Ruiz, Susan Athey, David M. Blei (2019) “SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements,” Papers 1711.03560, arXiv.org, revised Jun 2019.

Salakhutdinov R. and Mnih A. (2008). Bayesian Probabilistic Matrix Factorization using Markov chain Monte Carlo. In Proceedings ICML, 880-887.

Terui, N. and Li, Y. (2019). Measuring large-scale market responses and forecasting aggregated sales: Regression for sparse high-dimensional data. Journal of Forecasting, 38, 440-458.

Vasile, F., Smirnova, E. and Conneau, A. (2016). Meta-prod2vec: Product embeddings using side-information for recommendation. In Proceedings of the 11th ACM Conference on Recommender Systems. New York: ACM Press, 225–232.