



## CPS Paper

### Receipt Embedding and Shopping Purpose Segmentation

**Author:** Yinxing Li

**Coauthors:** Nobuhiko Terui

**Submission ID:** 508

**Reference Number:** 508

#### Presentation File

[abstracts/ottawa-2023\\_e90c6d65276a37e24f7974bd1fd89a4.pdf](https://www.isi2023.org/abstracts/ottawa-2023_e90c6d65276a37e24f7974bd1fd89a4.pdf)

#### Brief Description

Marketing data are expanding in several modes nowadays, as the number of variables explaining customer behavior has greatly increased, and the automated data collection in the store has also led to the recording of customer choice decisions which generate large scale samples.

Thus, high-dimensional models have recently gained considerable importance in several areas, including marketing.

Some distributed representation models based on product embedding such as Prod2Vec, for instance by Ruiz, Athey and Blei (2019), involve various marketing variables such as price and customer demographic data, but the role of these variables in forecasting and marketing decisions have never been well discussed.

Our study not only aims to propose a model with better forecasting precision but also to reveal how firm's marketing and customer demographics affect customer behavior and uncover the shopping purpose hidden in the receipt by extending product embedding approach.

#### Abstract

Marketing data are expanding in several modes nowadays, as the number of variables explaining customer behavior has greatly increased, and the automated data collection in the store has also led to the recording of customer choice decisions which generate large scale samples. Thus, high-dimensional models have recently gained considerable importance in several areas, including marketing. Some distributed representation models based on product embedding such as Prod2Vec, for instance by Ruiz, Athey and Blei (2019), involve various marketing variables such as price and customer demographic data, but the role of these variables in forecasting and marketing decisions have never been well discussed. Our study not only aims to propose a model with better forecasting precision but also to reveal how firm's marketing and customer demographics affect customer behavior and uncover the shopping purpose hidden in the receipt by extending product embedding approach.

First, based on Bayesian Word2Vec model, we assume that the purchasing probability of a certain product conditional on an existing market basket is determined by the following two factors: 1) the compatibility of products among the market basket, which is represented by the inner product of product vectors, 2) customer utility for the product, where we incorporate hierarchical structure for the utility connected to customer's demographic information. Then, we constitute the receipt vector from embedded product vectors of the purchased items. We assume that the customer's receipt represents the shopping purpose since the customer will consider the whole shopping context at each shopping before the choice of a product. Then we apply a state space model for the receipt vectors through the whole shopping trips for each customer, where marketing and seasonality variables are incorporated as covariates. We also show that the segmentation of large number of receipt vectors as well as the effectiveness of covariates could identify customer's shopping purpose at daily, weekly, and seasonal occasions for the efficient store management.

As for the model estimation, we implement fully MCMC Bayesian inference through the joint use of efficient algorithm of Polya-gamma sampling and parallel computing. This reduces the computational time drastically to a feasible time horizon, and then we precisely derive HPD region of parameters for testing significance and predictive distribution of forecasts.

Our proposed model produces not only higher precision of forecasts by incorporating marketing variables and customer heterogeneity into the model, but also provide better interpretability by our structural modeling. In the empirical analysis, we use ID-POS receipt data from a retail store, containing many kinds of customer's demographics, and store's marketing variables such as price, display and features. We show the effectiveness of marketing variables on the forecasting by using the measure Hit Rate@K for the hold-out sample comparing to several benchmark models, and the usefulness of interpretable model structure.