

2018 INFORMS International Conference
Taipei, Taiwan, June 17–20, 2018
© 2018 INFORMS|ISBN 978-0-9906153-1-6
https://doi.org/10.1287/infp.2018.XXXX
pp. 000–000

Sales forecasting from Heterogeneous Data Sources

Hong-Han Shuai
National Chiao Tung University, Hsinchu, 30010, Taiwan, hhshuai@nctu.edu.tw

Jui-Yi Tsai
National Chiao Tung University, Hsinchu, 30010, Taiwan, vincenttsai@iis.sinica.edu.tw

Yi-Chun Chen
Digital Service Innovation Institute, Institute for Information Industry, Taipei City, Taiwan, divienchen@iii.org.tw

Abstract Due to the convenience and low cost for both retailers and customers, E-commerce platforms have been indispensable for our modern lives. One of the key technologies for e-commerce platforms is to predict the next sales of commodities. An accurate prediction helps retailers not only reserve commodities in a smart way but also price the commodities for maximizing the total revenue. In this paper, we propose a fast and effective tensor decomposition model for sales forecasting via learning from heterogeneous data, including the information from omni-channel retailing, cross-channel user behavior series and social media. Abundant information with tensor decomposition facilitates the learning framework to avoid the cold-start problem as well as handles the missing data problem. We show the effectiveness of the proposed method on a real dataset including two e-commerce platforms (Yahoo and Ruten) and an online cosmetics review platform (UrCosme).

Keywords Tensor decomposition, heterogeneous data

1. Introduction

E-commerce platforms have been grown vertically and horizontally in a rapid rate. According to the U.S. Census Bureaumarket, the e-commerce share in Q3 of 2017, as a percentage of all retail sales, has increased to 9.1%, which almost triples the value a decade ago.[1] One of the key techniques is the sales prediction for commodities. An accurate prediction minimizes the cost of overstock, whereas a wrong prediction causes a huge waste. Moreover, an accurate prediction can facilitate different applications in e-commerce, e.g., auto-pricing, product promotion. However, it is difficult to predict the sales of commodities because 1) too many factors may affect the sales, and 2) new commodities suffer from the cold start problem. Although the data collected from different platforms may alleviate the problem, without a good model that integrates heterogeneous data sources, the prediction accuracy suffers from the curse of dimensionality.

Currently, a recent line of study aims at solving prediction problem via machine learning approaches, such as SVM, Decision Tree, KNN and Random Forest, by considering different features, such as sales transactions [1], user website-browsing behaviors [2], and social

[1] https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

media [3]. On the other hand, different multi-source learning algorithms are proposed for heterogeneous data sources, e.g. [4, 5, 6].

In this paper, we propose a tensor co-decomposition framework for extracting effective latent features from multi-source data and predict the next sales at the same time. Such latent features can 1) alleviate cold start problem by using the auxiliary information from other commodities, 2) reduce the dimensionality sent to machine leaning, 3) impute the missing values, which often occur in web data, and 4) directly predict the next sales. The contributions of this paper are summarized as follows.

- We propose a feature extraction framework for sales forecasting from heterogeneous data sources, which facilitate the smart product stocking.
- The e-commerce data usually suffer from missing data, which has been resolved by the proposed matrix factorization.
- The experimental results show the effectiveness and efficiency of the proposed approach.

The rest of this paper is summarized as follows. Section 2 formulates the problem of sales forecasting from heterogeneous data sources and present the proposed algorithm. Experimental results are presented in Section 3, and we conclude this paper in Section 4.

2. Proposed Approach and Experimental Results

In this section, we first introduce the features extracted from products. Afterward, a feature extraction framework that co-factorizing the product transaction data and product features is presented.

2.1. Product features

The factors that influence the decision making (whether to buy or not are listed below.

- **Target customers:** most products have their target customers. For example, most cosmetics' target customers are female; if a product is not compatible with users' requirement, then users will not buy it even they have favorable comments on the product.
- **Item properties:** the properties of item such as brand, product effects or ingredients.
- **The reviews from users:** the use experience such as actual result after using it, price-performance ratio, or the comparison with other similar products.

As such, the following features are extracted: 1) texture, color, and shape of the product photo, 2) price, 3) volume, 4) weight, 5) series, 6) dates of listing, 7) brands, 8) attributes and (9) history records.

2.2. Product Extraction

Moreover, since the history sales are highly correlated to the next sales, we model the transaction data as a tensor. Given the sales of N items on D platforms corresponding to K time points, we stack the item-platform matrices for each time snapshot $\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_K}$. The next sales prediction is formulated as imputing the matrix X_{K+1} from $\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_K}$. Therefore, given a tensor $\boldsymbol{T} \in R^{N \times D \times K+1}$, $\boldsymbol{T}$ is decomposed as follows:

$$\boldsymbol{T} \approx [\![\lambda; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!] = \sum_r \lambda_r \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r},$$

where $1 \leqslant r \leqslant R$ and R is the given parameter. When $R = 1$, the above equation can be seen as rank-1 decomposition. $\boldsymbol{a_r}$, $\boldsymbol{b_r}$ and $\boldsymbol{c_r}$ are the feature vectors of each dimension corresponding to rank r. Moreover, the embedding matrix $\boldsymbol{E} \in R^{N \times W}$ of item reviews from social media can be extracted via word2vec with W dimensionality. We minimize the following decomposition loss.

$$\min_{A,B,C} \|T - [\![\lambda; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]\|^2_{F*} + \lambda_e \|\boldsymbol{E} - \boldsymbol{AG}\| + \lambda_\alpha \|\boldsymbol{A}\| + \lambda_\beta \|\boldsymbol{B}\|$$

TABLE 1. Comparisons of sale prediction with different learning techniques.

Technique	AVG	RF	DL	Proposed
MAPE	46.35%	28.94%	14.57%	9.13%

where $\boldsymbol{A} \in R^{N \times R}$, $\boldsymbol{B} \in R^{N \times R}$, $\boldsymbol{C} \in R^{N \times R}$ are the latent feature matrices of $\boldsymbol{T}$ and $\|\cdot\|$ is $l_1 - norm$. Moreover, $\|\cdot\|^2_{F*}$ is the Frobenius norm only on the known values. $\boldsymbol{E}$ is decomposed as $\boldsymbol{A}$ and $\boldsymbol{G}$, where $\boldsymbol{A}$ is the same as in the tensor model (co-decomposition). To solve the above optimization problem, we adopt Stochastic Gradient Descent (SGD) and stop the optimization when it converges.

3. Experimental Results

We evaluate our prediction framework on a real dataset containing data from three sources, i.e., Yahoo! Auctions[2], UrCosme[3], and Ruten[4]. Moreover, 1,200 commodities in UrCosme can be matched to 11,955 items in different stores in Yahoo! Auctions and 13,711 items in different stores in Ruten. Our target prediction is the number of "buy" for each commodity in UrCosme. We use 5-fold cross validation, i.e., take 4 folds for training and 1 fold for testing, to evaluate the performance of the proposed features using the proposed learning framework. We use two baselines: 1) average (AVG), 2) random forest (RF), and 3) deep learning (DL). The number of trees for random forest is 300 with maximum tree depth as 5.

To evaluate the results, we adopt Mean Absolute Percentage Error (MAPE) as the metric. Given n testing data, MAPE is calculated as follows.

$$MAPE = \frac{100}{n} \frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)}{y_t}, \tag{1}$$

where $\hat{y}_t$ and y_t are the predicted sale amount and actual sale amount of the t-th commodity, respectively.

Table 1 compares different approaches in terms of MAPE. The results manifest that the proposed feature selection framework can reduce the MAPE of DL by 37.3%, while significantly outperforms AVG and RF.

4. Conclusions

In this paper, we propose a feature selection framework for sales forecasting from heterogeneous data sources. The proposed framework can 1) alleviate cold start problem by using the auxiliary information from other commodities, 2) reduce the dimensionality sent to machine leaning, 3) impute the missing values, which often occur in web data, and 4) directly predict the next sales. The results indicate that the proposed method can reduce the MAPE by 37.3%. As for the next step, we plan to exploit the power of deep learning and propose an attention model to extract features with a data-driven approach.

References

[1] M. Hlsmann, D. Borscheid, C. M. Friedrich, and D. Reith. General Sales Forecast Models for Automobile Markets Based on Time Series Analysis and Data Mining Techniques. ICDM, 2011.

[2] https://tw.bid.yahoo.com/

[3] https://www.urcosme.com/

[4] http://www.ruten.com.tw/

[2] J. Yeo, S. Kim, E. Koh, S. Hwang, and N. Lipka. Browsing2purchase: Online Customer Model for Sales Forecasting in an E-Commerce Site. WWW, 2016.
[3] A. Guille, H. Hacid, C. Favre, and D. A. Zighed. Information diffusion in online social networks: A survey. SIGMOD, 2013.
[4] M. Ghassemi, M. A.F. Pimentel, T. Naumann, T. Brennan, D. A. Clifton, P. Szolovits, and M. Feng. A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU with Sparse, Heterogeneous Clinical Data. AAAI, 2015.
[5] F. Zhuang, X. Cheng, S. J. Pan, W. Yu, Q. He, and Z. Shi, "Transfer learning with multiple sources via consensus regularized autoencoders," *ECML*, 2014.
[6] J. Choo, S. Bohn, G. C. Nakamura, A. M. White, and H. Park, "Heterogeneous data fusion via space alignment using nonmetric multidimensional scaling," *SDM*, 2012.