Browsing2purchase: Online Customer Model for Sales Forecasting in an E-Commerce Site*

Jinyoung Yeo[†], Sungchul Kim[‡], Eunyee Koh[‡], Seung-won Hwang[§], Nedim Lipka[‡]
[†]Pohang University of Science and Technology (POSTECH), Pohang, South Korea
[‡]Adobe Research, San Jose, CA, United States
[§]Yonsei University, Seoul, South Korea
jinyeo@postech.ac.kr, {sukim, eunyee, lipka}@adobe.com, seungwonh@yonsei.ac.kr

ABSTRACT

This paper covers a sales forecasting problem on e-commerce sites. To predict product sales, we need to understand customers' browsing behavior and identify whether it is for purchase purpose or not. For this goal, we propose a new customer model, *B2P*, of aggregating predictive features extracted from customers' browsing history. We perform experiments on a real world e-commerce site and show that sales predictions by our model are consistently more accurate than those by existing state-of-the-art baselines.

Keywords

Sales forecasting; Sales prediction; Customer model; E-commerce

1. INTRODUCTION

Sales forecasting [1, 2] has been a vital task to establish management and marketing strategies for business success. This paper covers the sales forecasting problem in e-commerce sites. For that, customers' browsing behavior on the website is important evidence to predict online purchases. However, browsing history aggregates behaviors of very different intentions which need to be distinguished. For example, over 92% of customers in our dataset browsed product pages but did not purchase, and we call them "window-shoppers" in this paper. To overcome this challenge, we propose a new customer model, *Browsing2purchase* (*B2P*), that considers not only customers' browsing history but also their browsing intention. Because individual's browsing intention is indeed hidden, our proposed model estimates whether it is for purchase purpose or not, by mining purchasing and window-shopping patterns of individual/collective online customers.

2. E-COMMERCE SALES PREDICTION

Given an e-commerce site, let $\mathcal{C}$ be a set of customers and let $\mathcal{P}$ be a set of products. We call (c,p) a purchase candidate pair (shortly pair) where $c \in \mathcal{C}$ and $p \in \mathcal{P}$. Inspired by [1], we define a pair's transitional states as $\{\text{ACTIVE}, \text{ADOPT}\}$: (c,p) becomes ACTIVE (c is interested in p) since c has visited p's page (*e.g.*, product information page), then becomes ADOPT (c purchases p) since c has made an online order for p. After ADOPT state, the pair (c,p) returns to ACTIVE when c visits p's page again. Given $p \in \mathcal{P}$ and a set $\mathcal{C}_{\text{ACTIVE}(p,t)}$ of its ACTIVE customers at current time t, future sales $\sigma_{[t:t+\phi]}(p)$ can be defined as the cardinality of the customers reaching ADOPT state within a certain time period ϕ after t, *i.e.*, $\mathcal{C}_{\text{ADOPT}(p,[t:t+\phi])} \subseteq \mathcal{C}_{\text{ACTIVE}(p,t)}$. A straightforward way to compute $\sigma_{[t:t+\phi]}(p)$ is using Monte-Carlo simulations: For each iteration, the pair (c,p) becomes ADOPT if the probability $Pr(c,p)$ of a customer c adopting a given product p is larger than a threshold $\theta_{c,p}$ assigned uniformly at random from 0 to 1. After k iterations, the expected value of $\sigma_{[t:t+\phi]}(p)$ is defined as estimated sales. In this work, because social influence between customers is not considered, $\sigma_{[t:t+\phi]}(p)$ converges to $\sum_{c\in\mathcal{C}_{\text{ACTIVE}(p,t)}} Pr(c,p)$ when $k \rightarrow \infty$.

*This work was done when the first author was at Adobe Research.

3. BROWSING2PURCHASE MODEL

In this paper, we generalize the probability $Pr(c,p)$ to consider browsing history $\mathcal{H}_c^*$ (a set of page visit events of c and their metadata such as URL and visit time). Formally, we define $Pr(c,p)$ as $Pr(\text{ADOPT}|\mathcal{H}_c^*)$ quantifying adoption probability given $\mathcal{H}_c^*$ (The term 'ADOPT' represents that c purchases p). In particular, we identify two predictive factors $\mathcal{H}_{c,p}^1$ (*Browsing Ratio*) and $\mathcal{H}_{c,p}^2$ (*Browsing Duration*) for predicting the intention of $\mathcal{H}_c^*$. Fig. 1 illustrates how *adopter* and *window-shopper* can be distinguished using $\mathcal{H}_{c,p}^1$ and $\mathcal{H}_{c,p}^2$. More formally, we redefine $Pr(c,p)$ as:

$$Pr(c,p) = Pr(\text{ADOPT}|\mathcal{H}_c^*) \approx Pr(\text{ADOPT}|\mathcal{H}_{c,p}^1, \mathcal{H}_{c,p}^2) \quad (1)$$

We now present the datails of *Browsing Ratio* and *Browsing Duration* as follows.

- **Browsing Ratio:** Visited pages indicate which products a customer has interests in. When visiting the pages of p frequently, the customer is more likely to adopt p. Otherwise, the customer will visit other pages more often. We thus define the ratio $\mathcal{H}_{c,p}^1$ as the estimation of adopt probability given a sequence of pages visited by c as:

$$Pr(\text{ADOPT}|\mathcal{H}_{c,p}^1) \approx \mathcal{H}_{c,p}^1 = \frac{V_{positive}}{V_{positive} + V_{negative}} \quad (2)$$

  where $V_{positive}$ is times c visits p's pages (*i.e.*, visit frequency) and $V_{negative}$ is times c visits other pages, while (c,p) is on ACTIVE state.

- **Browsing Duration:** While people tend to take from a few hours to a few days in browsing p's pages until adopting p, a large portion of window-shoppers lose their interests on p

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889394.

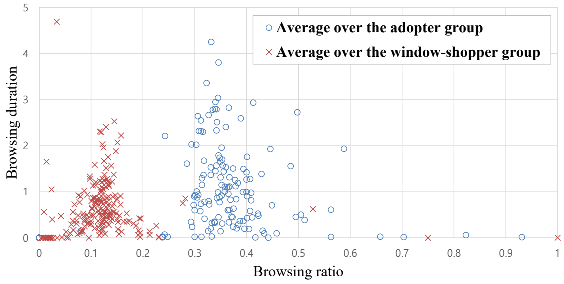

Figure 1: Plot of browsing ratio and browsing duration. A point represents a product (phone) and its values are the average over adopters (blue circle) or window-shoppers (red x).

soon. Using such a gap, we first divide customers reaching ACTIVE before t into two groups, $\mathcal{C}_{\text{ADOPT}(p,<t)}$ of adopting p before t (namely *adopter*) and otherwise $\mathcal{C}_{\text{ACTIVE}(p,<t)}$ (namely *window-shopper*). Then, we generate two probability density functions X_{ADOPT} and X_{ACTIVE} of "browsing duration" (random variable), which is defined as a time period from reaching ACTIVE state to reaching ADOPT state (for *adopter*) or to last visiting p's pages (for *window-shopper*), respectively. Based on two customer groups and probability density functions of browsing duration, we estimate adoption probability given a browsing duration $\mathcal{H}^2_{c,p}$ of (c, p) as:

$$Pr(\text{ADOPT}|\mathcal{H}^2_{c,p}) \approx \frac{C_{positive}}{C_{positive} + C_{negative}} \quad (3)$$

where $C_{positive} = |\mathcal{C}_{\text{ADOPT}(p,<t)}| \cdot Pr(X_{\text{ADOPT}} = \mathcal{H}^2_{c,p})$ and $C_{negative} = |\mathcal{C}_{\text{ACTIVE}(p,<t)}| \cdot Pr(X_{\text{ACTIVE}} = \mathcal{H}^2_{c,p})$.

As shown in Fig. 1, browsing ratio and duration are not strongly dependent. Thus, we approximately decompose the adoption probability below to simplify the estimation of probability values.

$$\begin{aligned} Pr(\text{ADOPT}|\mathcal{H}^1_{c,p}, \mathcal{H}^2_{c,p}) &= \frac{Pr(\mathcal{H}^1_{c,p}, \mathcal{H}^2_{c,p}|\text{ADOPT})Pr(\text{ADOPT})}{Pr(\mathcal{H}^1_{c,p}, \mathcal{H}^2_{c,p})} \\ &= \frac{Pr(\mathcal{H}^1_{c,p}|\text{ADOPT})Pr(\mathcal{H}^2_{c,p}|\mathcal{H}^1_{c,p}, \text{ADOPT})Pr(\text{ADOPT})}{Pr(\mathcal{H}^1_{c,p}, \mathcal{H}^2_{c,p})} \\ &\approx \frac{Pr(\mathcal{H}^1_{c,p}|\text{ADOPT})Pr(\mathcal{H}^2_{c,p}|\text{ADOPT})Pr(\text{ADOPT})}{Pr(\mathcal{H}^1_{c,p})Pr(\mathcal{H}^2_{c,p})} \\ &= Pr(\text{ADOPT}|\mathcal{H}^1_{c,p})Pr(\mathcal{H}^2_{c,p}|\text{ADOPT})/Pr(\mathcal{H}^2_{c,p}) \\ &= Pr(\text{ADOPT}|\mathcal{H}^1_{c,p})Pr(\text{ADOPT}|\mathcal{H}^2_{c,p})/Pr(\text{ADOPT}) \end{aligned} \quad (4)$$

To reduce the gap between actual sales and estimated sales, we approximate $Pr(\text{ADOPT})$ into a constant value, which can be trained by matching actual sales and estimated sales from time $t - \phi$ to t.

4. MODEL EVALUATION

We perform sales prediction to evaluate our model. As dataset, we use a popular e-commerce site, and for purchase candidate pairs, we collect 276 phone products, 4.6M customers having visited the products' pages, and 0.6B page visit events from Feb 13th to Sep 2nd in 2014. Among these customers, only 0.3M (6.5%) customers purchased at least one product. As experiment parameters, we set a *prediction period* $\phi \in \{1month, 2months\}$ and, for each ϕ, perform sales predictions with 20 different *current time* t (changing with a weekly cycle) and measure their average RMSE between actual sales and estimated sales.

We compare our model *B2P* with two baselines. The first baseline, CTREND [1], learns $Pr(c, p)$ as maximum likelihood estimate which is the fraction of times a customer became ADOPT over the times the customer became ACTIVE. Inspired by the fact [2] that the total sales of an e-commerce site have low variability (a little fluctuating but recovered soon) according to different time, the second baseline, PTREND, learns $\sigma_{[t:t+\phi]}(p)$ as the number of adopters from $t - \phi$ to t, *i.e.*, $\sigma_{[t-\phi:t]}(p) = |\mathcal{C}_{\text{ADOPT}(p,[t-\phi:t])}|$ where $\mathcal{C}_{\text{ADOPT}(p,[t-\phi:t])} \subseteq \mathcal{C}_{\text{ACTIVE}(p,t-\phi)}$.

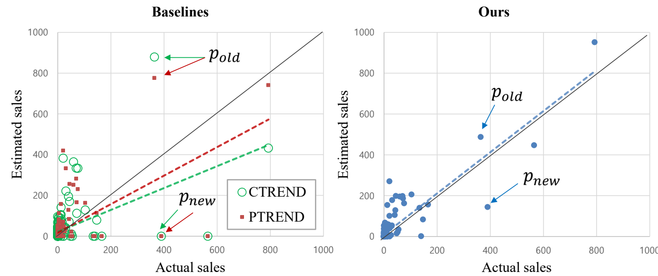

Figure 2: Baselines vs. Ours (*B2P*) on Apr-4th-2014. A point represents a product's sales prediction result. The $x = y$ line is the ideal prediction. Dotted lines are prediction trend lines.

Table 1: Average RMSE for different prediction models

	Prediction period		
	1 month	2 months	Gap
CTREND	39.7 (±19.2)	99.8 (±20.9)	60.1
PTREND	56.1 (±19.0)	80.7 (±18.3)	24.6
B2P model	**32.7 (±10.6)**	**54.5 (±12.0)**	**21.8**

As a result, Tab. 1 shows that the RMSE of our proposed model is lower than that of baselines in both 1 month and 2 months prediction. Also, the standard deviation of RMSE and the performance gap between two prediction periods are lowest in *B2P* model. Lastly, Fig. 2 contrasts how sales of newly added product are predicted in baselines and our model. To illustrate, a new product (p_{new}) and an existing product (p_{old}) are marked. In baselines, the adoption probability of p_{new}, or $\sigma(p_{new})$, is underestimated to near-zero, while that of p_{old} is overestimated. Such gap is more reasonable in our *B2P* model, which shows that prediction results of all products (including p_{new} and p_{old}) are more correlated with the trend line.

5. CONCLUSION

In this paper, we have presented our proposed approach to predict online sales using customer's browsing history. By aggregating two predictive factors of purchasing and window-shopping behaviors, our proposed customer model, *B2P*, was effective in reducing prediction errors of state-of-the-art baselines. More sophisticated analysis for different prediction period is promising future work.

Acknowledgement

This work was supported by a grant of the Institute for Information & Communications Technology Promotion (IITP) funded by the Korea government (MSIP) (No. 10041244, SmartTV 2.0 Software Platform).

6. REFERENCES

[1] S. Bhagat, A. Goyal, and L. V. Lakshmanan. Maximizing product adoption in social networks. In *WSDM*, 2012.

[2] J. W. Byers, M. Mitzenmacher, and G. Zervas. Daily deals: Prediction, social diffusion, and reputational ramifications. In *WSDM*, 2012.