Purchase Forecast of the Promotion Day Based on Model Fusion and Migration Learning
Hao WU and Shu-fang LI*
Beijing Key Laboratory of Network System Architecture and Convergence
Beijing Laboratory of Advanced Information Network
Beijing University of Posts and Telecommunications Beijing, China
*Corresponding author

Keyword: Promotion day, Purchase forecast, Model fusion, Migration learning.

Abstract. For the violent data fluctuation of the promotion day data and the characteristics of limited data size, combined with the idea of migration learning, two modeling ideas of parallel model and serial model are proposed. The algorithm avoids the difference in data distribution between promotion day and non-promotion day, and makes full use of non-promotion day data to improve the accuracy of model prediction. Experiments conducted by real promotion day data show that migration learning can effectively reduce prediction errors.

Introduction
In recent years, the rise of the shopping festival represented by Double Eleven has brought huge economic benefits to e-commerce platforms and shops. It is increasingly important to accurately predict the sales of the shopping festival and develop a more rational marketing strategy. However, during the shopping festival, merchant and platform promotions can cause traffic distribution to change dramatically, and models trained on normal traffic do not match these special traffic well. How to make better use of massive transaction data to efficiently and accurately predict the user's purchase intention is a technical problem that artificial intelligence and big data need to continue to solve in the e-commerce scenario.

In the purchase forecast, the network purchase behavior prediction based on feature selection and model fusion is a common method[1]. It design the feature engineering which can capture the information of user and commodity, and apply the gradient promotion decision tree (GBDT) as the Training model[2]; Some scholars have learned a separate classifier for each customer, and have made very fine-grained predictions on the products purchased in a particular purchase of a particular user. The results show that the shopper's shopping list can be predicted to some extent[3];

Some scholars also use the Bayesian method to predict online advertising from the perspective of statistics and probability[4]; The model using Bagging integrated learning method will reduce the generalization error because of the certain independence between different models, so that it will have better prediction effect than a single model[5]; Some articles build a predictive model based on a series of user behavior sequence information such as user browsing, collection, shopping cart and purchase evaluation, such as Yan Xianyu's potential factor selection model for user purchase prediction[6]; There are also BP neural network[7], convolutional neural network (CNN)[8] to carry out deep mining of e-commerce data, using neural network algorithm can effectively avoid the manual construction of complex feature engineering, can achieve automatic feature extraction, However, such algorithms need to be lacking in accuracy, and have high requirements on data size and training time.

In the existing research, there are few researches on the sales promotion forecast. This paper takes the promotion day commodity purchase forecast as the research object, constructs a series of characteristics for the promotion factor, and uses parallel model and serial model to conduct sales promotion forecast with combining the idea of migration learning and model fusion. The remainder of this paper is organized as follows. In Section II, we introduce experimental data. In Section III,
we discuss the feature building ideas and briefly explain the features built. In Section VI, we discuss the method of model building. In Section V, we discuss the results of the experiment and conclude this paper.

Experimental Data

The experimental data is the massive real transaction data of the Taobao E-commerce platform, with a total of about 10 million pieces of data. The data is obtained from the user's sample of the eight consecutive days of browsing log data of a certain type of product advertisement, wherein the data of the first seven days is usual data (recorded as \(D_u\)) and the data of the eighth day if promotion data (recorded as \(D_p\)). \(D_p\)’s CVR (Advertiseing Conversion Rate) fluctuations have changed drastically compared to \(D_u\). We will use \(D_u\) and the part of \(D_p\) (the data before 12 o’clock) to forecast the another part of \(D_p\) (the data after 12 o’clock).

Feature Building

The features are constructed from three perspectives: user perspective, store and commodity perspective, and attribute perspective.

User perspective: It is difficult to mine the consumption habits of a single user through the user's browsing records of certain types of goods, but the users' common consumption habits can be used to evaluate the user's purchase probability for a certain item. For example, if a user browses a certain item multiple times, the possibility of generating consumption will be greater. Store and commodity perspective: the scale of data related to a certain store or a certain product is very considerable, so it can be tapped out which types of users are more attractive to shops and goods, or which types of shops or goods are more attractive to users; Attribute perspective: the range or the number of attributes' values is limited (such as gender and age), but based on the data of sufficient scale, the attributes are combined to construct a lot of joint features. And we can mine and analyze the potential connections between attributes.

Model Building

Problems with Conventional Modeling Methods

<table>
<thead>
<tr>
<th>data set</th>
<th>Coverage days</th>
<th>Total purchases</th>
<th>Total browsing</th>
<th>Purchase rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D_u)</td>
<td>7</td>
<td>101564</td>
<td>9354861</td>
<td>0.0108</td>
</tr>
<tr>
<td>(D_p)</td>
<td>1</td>
<td>49646</td>
<td>1077175</td>
<td>0.0460</td>
</tr>
</tbody>
</table>

\(D_u\) has a stable data distribution and a sufficient data size. In contrast, \(D_p\) is more volatile and its data size is limited. This results in a model based on \(D_p\) with lower training error but higher generalization error, and the model based on \(D_u\) will have two types of errors.

Modeling Basis

Although the number of browse, volume, and transaction success rate on the promotion day is much higher than usual due to the promotion activity, this does not mean that there is no similarity between \(D_n\) and \(D_s\): from the customer's point of view, during the promotion day and non-promotion day, the influence of gender, age, occupation and rank on the shopping tendency of customers is similar; from the perspective of shops and commodities, the popularity, favorable rate, and attraction of a certain commodity or a certain store is inherent and don’t change much because of the arrival of the promotion day.
As shown in the figure below, due to the promotion factors, the page views and purchases of a certain product vary greatly, but the influence of gender factors on the sales of goods has not changed significantly.

**Modeling Method**

Based on the above analysis, combined with the idea of migration learning, two modeling ideas are proposed: parallel model and serial model.

a) The so-called parallel model refers to dividing the training set multiple times on the original data set in the form of inclusion, and obtaining multiple models (recorded as $model_{p1}, model_{p2}, model_{p3}$ and $model_{p4}$) based on different training sets, and then weighting the prediction results of the multiple models to obtain the final model (recorded as $model_{pf}$).

The data set is divided into 4 parts: 8\textsuperscript{th} day data set, 6\textsuperscript{th}-8\textsuperscript{th} day data set, 4\textsuperscript{th}-8\textsuperscript{th} day data set and 1\textsuperscript{st}-8\textsuperscript{th} day data set. We build the model on the four data sets respectively, and forecast the data on the afternoon of the 8\textsuperscript{th} day with the four models. We obtained four predictions and weighed the four results in a certain proportion to get the final result.

There is a progressive inclusion relationship between the four data sets divided, and the closer the data is near the forecast day, the more times it’s used. Among them, the data of the 8\textsuperscript{th} day is used most. On the one hand, this modeling method makes full use of the data of the previous seven days on the basis of retaining the data distribution of the 8\textsuperscript{th} day data as much as possible. On the other hand, the fusion of multiple models also helps to reduce the generalization error.

b) A large-scale normal day data is used to obtain a model (recorded as $model_{imp}$) that can stably predict the normal day purchase situation, and we obtain an in accurate forecast result by predicting the promotion day with $model_{imp}$. Then add the prediction result as a new feature to the promotion date data. Finally, the final model is obtained from the promotion day data, which is the serial model (recorded as $model_{s}$). Although the forecast result made by $model_{imp}$ is inaccurate, it can reflect to a certain degree of influence towards commodity heat, user group and other factors on whether or not purchase occurs, that helps $model_{s}$ to make more accurate predictions.
Analysis of Results

Metrics

In the experiment, logarithmic loss (recorded as logloss) was used to evaluate the model effect. The formula is as follows:

\[
\text{logloss} = \frac{1}{N} \sum_{i=1}^{N} (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))
\]

Where \(N\) is the number of samples in the test set, \(y_i\) is the actual label of the \(i^{th}\) sample in the test set, and \(p_i\) is the estimated conversion rate of the \(i^{th}\) sample.

Verification Set

Since the competition platform does not disclose the data after 12 o'clock on the promotion day, so we use the data form 9:00 to 12:00 on the promotion day as the verification set.

Parallel Model

Four models were obtained from the four data sets. First, we forecast verification set with the four models separately and obtain four forecast results. Then, the four forecast results were merged according to the ratio of 4:3:2:1. The results are shown in the following table:

<table>
<thead>
<tr>
<th>model</th>
<th>model(_{p1})</th>
<th>model(_{p2})</th>
<th>model(_{p3})</th>
<th>model(_{p4})</th>
<th>weighted fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>logloss</td>
<td>0.169723</td>
<td>0.170156</td>
<td>0.171002</td>
<td>0.173524</td>
<td>0.168991</td>
</tr>
</tbody>
</table>

From \(model_{p1}\) to \(model_{p4}\), the proportion of non-promotional day data in the training data is increasing, making the data distribution of training data more and more inclined to normal day data, so the model's ability to fit the promotion day data is worse, which can be seen from the prediction error from \(m1\) to \(m4\) gradually increasing. After weighted fusion of the four prediction results, the prediction error is lower than the \(model_{p1}\) that only uses the promotion day data as the training data.

Serial Model

As shown in Table 3, \(model_s\) is the model obtained by applying migration learning, and \(model_{p1}\) just uses the promotion day data as the training data. It can be seen from the above table that when the time coverage of the verification set is short, the application of migration learning will lead to poor performance of the model. But with the expansion of the verification set, the effect of migration learning is getting better and better. This is because as the coverage duration of the verification set increases, the overall similarity between the verification set data and the training set data gradually decreases. It is more unreliable to rely solely on the training set data to predict the verification set, which is consistent with common sense: the longer the prediction interval is, the more difficult it is to predict. We are able to mine intrinsic factors such as commodity popularity, user common shopping habits, and shop popularity from a large number of data of usual days, and these factors are less sensitive to the duration of the prediction interval, so when the verification set coverage time increases, we can get better forecast results.

<table>
<thead>
<tr>
<th>Verification set</th>
<th>model(_{p1})</th>
<th>model(_s)</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>11h</td>
<td>0.160905</td>
<td>0.160998</td>
<td>-9.3*10^{-3}</td>
</tr>
<tr>
<td>10h-11h</td>
<td>0.163068</td>
<td>0.162977</td>
<td>9.1*10^{-3}</td>
</tr>
<tr>
<td>9h-11h</td>
<td>0.166828</td>
<td>0.166641</td>
<td>18.7*10^{-3}</td>
</tr>
<tr>
<td>8h-11h</td>
<td>0.169723</td>
<td>0.169412</td>
<td>31.1*10^{-3}</td>
</tr>
</tbody>
</table>

Conclusion

This paper focuses on how to more effectively use the data of usual days to forecast purchases of
promotion day, constructs features from three perspectives of Store and commodity, users, and attributes, and builds the model with multi-dataset model fusion and migration learning. To a certain extent, it avoids the difference in data distribution between the data of usual days and the data of promotion day, and applies information such as commodity heat, user common consumption habits and shop popularity in the daily data to the promotion day forecast to reduces prediction error and improve model’s generalization ability.

References


