Background & Motivation

Rossman uploaded the data online. The data comes in two sets:

1. Historical sales data for 1,115 Rossman stores from 2013/1/1 to 2015/6/30. Store number, date, day of week, whether there's a promotion, whether it's a holiday and sales on that day are provided in each sample.

2. Stores' individual characteristics. This includes each store's store type, assortment level, nearest competitor's distance and date when the competitor opened, whether and when this store is on a consecutive promotion. Although we don't have to collect, yet the data comes in a messy way, and the data is also very large. So it took us a while to preprocess the data.

Data Interpretation

Although we don’t have to collect, yet the data comes in a messy way, and the data is also very large. So it took us a while to preprocess the data.

70%/30% and k-fold cross validations are used in this problem for training and testing, and we used Root Mean Square Percentage Error (RMSPE) to measure accuracy, which is defined as below:

\[ \text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2} \]

Single Store Prediction

The first step for us to analyze each store separately, regardless of each store's individual characteristics, so only consider the first dataset with historical sales for 1,115 Rossman stores.

First we apply Linear Regression to Store 1, this gave us a RMSPE of 23.7%. Then we figured out that this problem can actually be kernelized (SVR). Here consider that case of applying MAP estimate for B to avoid overfitting, which result in the following primal problem:

\[ \hat{y} = \text{arg} \min_{y} \left( \frac{1}{2} \| y \|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i) \right) \]

For Gaussian Kernel, as RMSPE varies with \( \lambda \) and \( \sigma \), we draw a figure to find the minimum value of \( \lambda \) and \( \sigma \).

3 different side views of this diagram is shown above. Figure (a) shows the overall shape of the surface, (b) shows that when \( \lambda =45 \) RMSPE reaches it's minimum, (c) shows that TMSPE decreases when \( \sigma \), however when \( \sigma > 120 \), RMSPE becomes relatively stable, so we choose \( \sigma \) as 140. From the parameters above chosen, Gaussian gives us a RMSPE of 13.5%, which is better than linear regression.

Our next model for this project is Random Forest Regression. We tried this model because it's fast and can accommodate categorical data. RF first picked a certain amount of data from the dataset randomly and then picked a certain amount of features out of the total features randomly to build decision trees. The final result for each test data is average of results obtained by all these decision trees. We used k=10 folds cross validation for this step for RF.

In consideration of generality, we try to apply the algorithms to all stores. For Gaussian Kernel, we randomly chose several stores and tried to find their optimal \( \lambda \) and \( \sigma \) pair. We found that for all tested stores, \( \lambda =45 \) and \( \sigma = 140 \), may not be the best result, but provide a rather close result to optimal result. So we choose \( \lambda =45 \) and \( \sigma = 140 \) and apply Gaussian Kernel to all stores. The following histogram shows RMSPE of Gaussian Kernel and RF for each fact.

Comparison & Result

- **Linear Regression for Store 1**: 23.7%
- **SVR with Polynomial Kernel for Store 1**: 28.4% for all d=1 and all \( \lambda \)
- **SVR with Gaussian Kernel for Store 1**: 13.5% \( \lambda =45 \) and \( \sigma = 140 \)
- **SVR with Gaussian Kernel for all stores**: 22.4% \( \lambda =5 \) and \( \sigma = 60 \)
- **Random Forest for all stores**: 24.2% Number of trees: 20

Future Work

**As you can see, linear regression, SVR with Gaussian/Polynomial Kernels and RF all have their own limitations and need further improvements. What we plan to do in the next few days is to 1. apply SVR with radial basis function kernel and probably more kernels; 2. combine SVR with RF and this may eliminate some outliers for using them separately; 3. remove some really noisy store and check the results.**

Model For All Data

After trying models for each individual store, our following step is to build a common model for all stores. And of course, we need to take each individual store's characteristics into consideration, so consider both datasets. After repeating the same method for Store 1, we try to apply SVR with Gaussian Kernel to all stores. A pair of \( \lambda=5 \) and \( \sigma = 60 \) gives us a RMSPE of 22.4%, which is not good enough. The biggest issue is that the data set is too large, that only a short time period could be operated on, and also it’s impractical to try as many \( \lambda \) and \( \sigma \) pairs as possible.

We also applied RF after merging all data. We used dummy variables for categorical data. We also got a plot of RMSPE vs. number of trees as shown below. From this plot, we can choose the relatively smallest value, \( k=20 \) and result is 24.2%.

Contact

Yang Li  Chenghao Wang
MS @ Dept. Electrical Engineering  MS @ Dept. Mechanical Engineering
Stanford University  Stanford University
Email: yangli@stanford.edu  Email: wccheng@umich.edu