Using Conditional Random Fields to Forecast Product Promotions

Ananth Sivasankaran, Aniketh Kulkarni, Sudeep Kurian, Matthew A. Lanham

Purdue University Krannert School of Management

asivasa@purdue.edu; kulkar52@purdue.edu; kurians@purdue.edu; lanhamm@purdue.edu

Abstract

This study examines the efficacy of Conditional Random Fields (CRFs) to predict promotions on various products at different times. The study involves understanding the historical performance of product promotions, product groups and stores that typically go on sale together. The primary goal is to provide a working tool to help Merchants/Category Managers estimate the promotions on various products throughout the year and allow them to plan their inventory accordingly. The data available to us however, is sparse and discontinuous adding to the complexity. The traditional predictive modelling approaches perform poorly due to this and hence we seek to investigate the suitability of sequential modelling techniques such as CRFs

Introduction

Discount sales and promotions in the clothing retail industry are usually decided at a regional level based on several parameters but merchants and store managers often have little insight into when and what SKUs would go on sale and at what level of discounts. The information gap results in poor planning and reduced efficiencies at the store level. This problem could be viewed as a classic time-series forecasting challenge involving trend and seasonality. However, the available data is such that a SKU on discount is not registered unless it is sold on a particular day. The data is hence highly discontinuous and sparse.

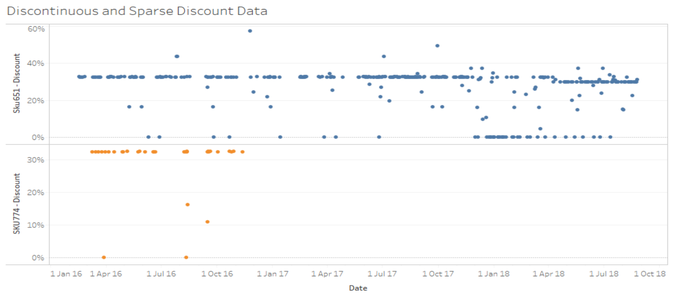

Figure 1. Discontinuity of Data

This complexity motivates the need to explore sequential modeling techniques which can utilize the information latent in the sequential nature of discount sales. Hence, the primary research objective of the study is to investigate the efficacy of CRFs to predict the discounts in a such a scenario. We also seek to detect groups of SKUs behaving similarly.

Literature Review

Our approach for the given problem statement is two pronged:

1. Group the SKUs that follow similar patterns in discount sales through the year.
2. Group the geographies that follow similar discount sale patterns.

There is sufficient literature in clustering methods to achieve the above goals when there is sufficient and continuous data available. However, when it comes to sparse time-series data, there is little research that focuses on the application and efficacy of sequential modeling techniques such as hidden Markov models. Such models have been extensively used in the context of natural language processing systems as explained in the next section.

We build upon the work of researches such as Dhillon et all. to explore the application of CRFs in sparse time-series forecasting. Insights obtained from this literature about the failure of traditional time-series forecasting techniques such as ARIMA, as well as traditional clustering algorithms like k-means form the guidance to this project.

Methodology

Figure 2 outlines our study design, starting from data collection, data cleaning, data pre-processing, feature creation and selection, model/approach selection, cross-validation design, and model assessment/performance measures.

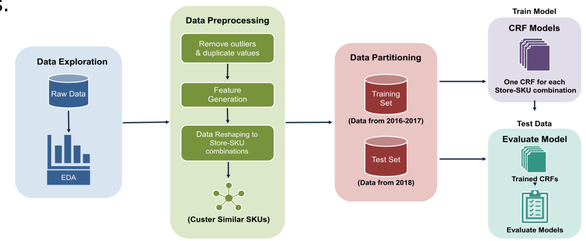

Figure 2. Study Design

Data:

- For our data, we had transactional data from a retail store. The data had the following information.

Variable	Type	Description
Date	Date	The Date of the transaction
location_ordinal	Numeric	Location (Masked due to confidentiality reasons)
SKU	Categorical	The Item (SKU) ID
price	Numeric	The price for the given SKU
discount	Numeric	The discount given for an SKU for a transaction
color_desc	Text	Color Description of an SKU
style_desc	Text	Style Description of an SKU
pid_desc	Text	Style Description for the Product (Redundant)
district_ordinal	Numeric	District (Masked due to confidentiality reasons)
state _ordinal	Numeric	State (Masked due to confidentiality reasons)

Table 1. Data Dictionary

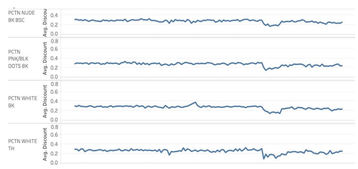

Figure 3. Generating Clusters

Data Cleaning and pre-processing:

- Removing outliers and Duplicate values: It was found that the data contained a few duplicate entries. Also, there were a few outliers, like some SKUs having negative discount values. These duplicate values and outliers were removed
- Generating Store ID: A feature called Store ID was created by combining the ordinal variables for State, District and Location

Feature Selection:

- Generating Clusters: We generated cluster values for SKUs based on an analysis of the product style and color description. There were 130 unique style descriptions. 10 Unique clusters were generated for the 130 style descriptions (Figure 3)
- For our CRF model, we needed a latent variable as an input. For a given SKU, we used the average discount in the cluster the SKU belonged to (except the SKU) on a sequential level as our predictor latent variable.

Model Design:

- We generated Store-level, SKU-level data. Each Store-SKU would have its own CRF model. Because our data was sparse, we aggregated the data to a weekly level.
- We partitioned data into 80 weeks–20+ weeks training and test set because we needed enough data to generate an appropriate transition matrix out of given data.

Methodology Selection:

- **Conditional random fields** is a statistical modelling method often applied in pattern recognition and structured prediction. They use contextual information from previous labels, thus increasing the amount of information the model has to make a good prediction.
- In our case, we try to predict the promotion offer that was offered at past date based on the sale offers of other similar products.

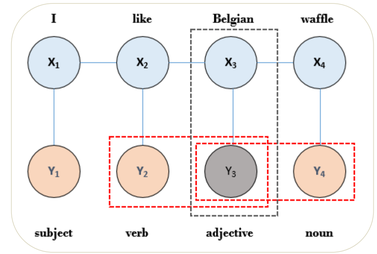

Figure 3. Illustration of CRF in Parts-of-speech tagging

- We expect the promotion to be a function of four factors: type of the item, promotion provided on that item, the previous day, the demographics of the store and offers on other items in the same category.
- We particularly chose CRF because they provide a discriminative model for sequential data. It also assumes that the sequence is Markovian.

Results

To measure the efficacy of our model, we used EMR (Exact Match Ratio) as our metric. For these test scenarios, we use our CRF model to find the expected discount bucket and compare it with the actual discount buckets. Table 2 shows how many entries are classified in each category and what their original classification distribution was. The diagonal elements are correctly classified.

Obs\Pred	A	B	C	D	E	F	G
A	0.7%	0.0%	68.8%	30.3%	0.2%	0.0%	0.0%
B	0.0%	0.0%	65.9%	34.1%	0.0%	0.0%	0.0%
C	0.0%	0.0%	88.8%	11.2%	0.0%	0.0%	0.0%
D	0.0%	0.0%	40.2%	59.8%	0.0%	0.0%	0.0%
E	0.0%	0.0%	15.8%	78.8%	5.4%	0.0%	0.0%
F	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
G	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%

Obs\Pred	A	B	C	D	E	F	G
A	75.0%	0.0%	0.0%	25.0%	0.0%	0.0%	0.0%
B	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
C	2.8%	0.2%	69.4%	27.3%	0.2%	0.1%	0.0%
D	2.1%	0.3%	18.6%	76.4%	2.5%	0.0%	0.0%
E	9.1%	0.0%	18.2%	0.0%	72.7%	0.0%	0.0%
F	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
G	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 2. Confusion Matrix for Train (left) and Test (right) datasets

Method	Train Accuracy	Test Accuracy	Comments
Mean Imputation	37%	38%	Baseline Accuracy
Random Forest	97%	69%	Overfitted Model
Conditional Random Fields	62%	71%	Best of the lot

Table 3. Benchmarking with other imputation methods

$$Exact\ Match\ Ratio = \frac{1}{n}\sum_{i=1}^{n} I(Y_i = Z_i)$$

The most appropriate option among the three seem to be predicting promotions through conditional random fields.

Conclusions

Identifying the amount of promotion is very crucial for any company in retail sector. Modelling with CRF is a effective way of predicting the effect of promotion. It is a much robust method for missing value imputation in comparison to the conventional methods. This also helps us predict the realistic value of the markdown price.

This paper provides a method through which Probabilistic Graphical Models can be used to predict promotions effectively.

Acknowledgements

We thank Professor Matthew Lanham for constant guidance on this project.