**Next Basket Recommendation Engine – Case study**

Online grocery stores have a unique opportunity of personalizing the experience for their customers. Recommending a next basket based on previous purchases will delight the customers as they no longer need to maintain groceries list.

**Our Solution**   
 The goal of the project is to capture the underlying purchase patterns of customers. Based on the analysis, a recommendation engine is built which gives a personalized list of groceries every time a customer comes back. The recommendations are interpretable so that a sales manager can fine tune the engine to suit his needs.

**Our Approach**:

The preliminary analysis of data resulted in patterns which can be broadly classified into three categories

* User specific patterns
* Population specific patterns
* Cluster specific patterns

After extensive experimentation, best model is fitted for each category of patterns. The end model is an ensemble.

**Impact**

* The model could recommend the next list of grocery items to individual customers with an accuracy of 80%
* Based on recommendation generated, overall demand can be forecasted as well.
* A sales manager can identify the opportunities to cross sell based on the recommendations generated from the engine.

**Description of the Model**

The data consists of user’s transaction history. Each transaction consists of all products ordered with time stamp. A product catalog with aisle and department tagged for each is also available.

The objective of the recommendation model is to capture the buying patterns of every customer. The algorithm we developed is based on Temporal Annotated Sequence based Predictor – TARS [1] which is a published work. TARS are sequences generated for each user data based on FP growth algorithm. Each sequence is tagged with features like inter time, intra time and periodicity and recurrence. Based on these parameters, potential TARS are filtered.

**Anatomy of a sequence:**

S = Milk Curd

Milk = Precedent, Curd is Antecedent

Inter times = [a1, a2...an ] ; Intra times = [I1, I2...In] ; Qmed = Median number of occurrences,

Pperiods =[P1,P2,….Pn] ; Recurrence = |P|

Customers tend to buy items in pairs/sets which can be termed as co-occurrence. Hence the first step of the algorithm uses periodic FP growth algorithm to mine frequent co-occurring items.

Groceries are purchased sequentially i.e one item after other in sequence. Hence the next step for our model is to calculate intra-time which incorporates sequentiality into the TARS Sequences.

A customer makes specific purchases only at specific periods of time in a year which can be modeled as periodicity. We calculate inter time for each sequence and find periods which are tagged.

And it is observed that many of periodic patterns recur frequently. A recurrence parameter is also calculated for each sequence.

At the end we have a set of potential recurrent sequences with each sequence tagged with inter and intra times, periodicity and recurrence.

**Predictive Modeling :**

From the potential sequences, less recurring are filtered out. For each of remaining sequence, unique predictor value is calculated for antecedent of each sequence based on three factors.

1. Support of antecedent among sequences
2. Difference of last period occurrences and median number of occurrences among all periods
3. Difference of days since last precedent occurrence and median intra time

A predictor value is assigned for all the sequence - antecedents and top-k items are selected and given as next basket recommendation.

**Interpretability and flexibility of Model**

TARS based predictor Model has several parameters which can be understood by marketing managers and can be tweaked to improve the performance of recommendation based on their vast retail domain.

**Results, Interpretation and future work**

Our end model was an ensemble of all the above models and gave an F1 score of 0.375. We could reach a maximum F1 score of 0.7 - 0.8 for a good number of customers.

**Reference**

1. “Riccardo Guidotti, Giulio Rossetti, Luca Pappalardo, Fosca, Giannotti, and Dino Pedreschi”, Next Basket Prediction using Recurring Sequential Patterns.