Dynamic aggregation for time series forecasting

Sam Iosevich,
Georgiy Arutyunyants,
Zhuang Hou,
Prognos, an Antuit Company

Abstract—The proliferation of data and emerging of Big Data technologies have led to demand sensing and shaping at the most granular product and geography levels. This has led to a need to optimize tens and many times hundreds of millions of geography product treatments on a weekly basis. The amount of data has overwhelmed the ability to monitor individual recommendations, even by exception. In this scenario, it is imperative that the underlying demand modeling process be as stable as it is highly accurate. The methodology is geared towards automated forecasting systems with large amounts of time series inputs of varying volume and volatility. These systems are often encountered in Retail and CPG applications such as replenishment and pricing. This paper outlines a dynamic modeling approach that produces stable and highly accurate demand forecasts.

Index Terms—Big Data, Time Series, Retail, CPG, Dynamic Modeling.

1 Introduction

We start from defining a group of time series,

$$\{TS_1, \cdots, TS_N\}, \tag{1}$$

where each time series is

$$TS_i = \{a_1^i, \cdots, a_K^i\}. \tag{2}$$

We assume that the system of time series described in (1) can be grouped in geographic/product hierarchies. An example of such system of hierarchies can be seen in Fig. 1, where each combination of nodes on the most granular level of geographic/product hierarchy can be viewed as an element in (1).

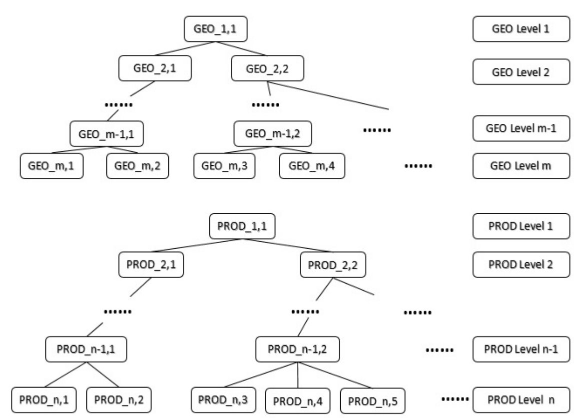

Fig. 1. Geo/Prod hierarchy tree graph

Such a system is common in Retail & CPG applications. Every time series in

(1) can be viewed as a historical sales of a given product in a given geography (store, distribution center, etc). In this paper, we are going to compare two different methodologies of producing the forecast on different levels (consumption levels) on geographic/product hierarchies.

We define

$$FS_i = \{f_1^i, \cdots, f_K^i, \cdots, f_{K+L}^i\} \quad (3)$$

as the forecast of time series i, where L is the forecast horizon.

The following two concepts, aggregation and dis-aggregation, play a critical role in this paper. We define aggregation to a given level of geographic/product hierarchies as the summation of all time series for a given time point under the same parent nodes on those levels.

Fig. 2 represents the geographic/product hierarchy coordinate system. The horizontal axis is the geographic hierarchy levels, and the vertical axis is the product hierarchy levels. Every point of the grid in Fig. 2 can be considered as aggregation level of our system 1. For example, Point P represents the most granular level, whereas point H represents the highest level.

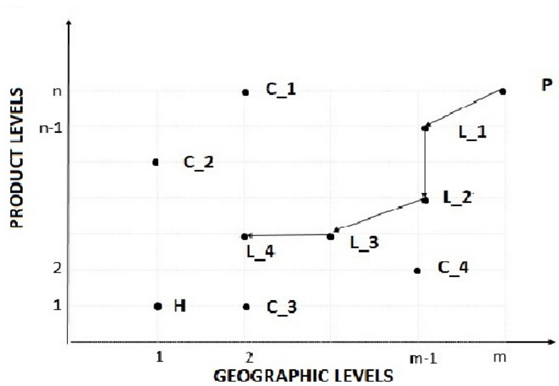

Fig. 2. Bucket Levels

To define dis-aggregation, we need to define non-negative weights of every child for a given parent in such a way that sum of the weights of all children for a given parent is 1. In this paper, we are using sales to define those weights, and we call it ”sales-weighted dis-aggregation”. Once weights are defined, dis-aggregation is a process of proportionally (based on weights) propagating the forecast from the higher level to lower level.

2 Multi-echelon versus Dynamic Aggregation

In this section, we compare the Multi-Echelon forecasting approach to a forecasting approach that utilizes. Dynamic Aggregation. The Multi-Echelon approach aggregates all of the data to multiple levels and a forecast is produced at every level. The forecasting approach that utilized Dynamic Aggregation aggregates each geography product combination to one and only one level. The two approaches are diagrammed in Fig. 3.

The Dynamic Aggregation approach is motivated by the observation that each time series has an optimal aggregation point at which the signal to noise ratio is optimized. The higher the data is aggregated the more information loss is introduced, at lower levels data sparsity and white noise makes applying time series algorithms difficult.

Dynamic aggregation consists the following steps:

1) Firstly, the set of levels is defined on which the forecast is produced. On Fig. 2, that set can be viewed as directed graph with the start at point P and end at point L_4. In this paper, we are not discussing

how this directed graph is selected, but it is worth mentioning that we did not find strong dependency between consumption levels $C_1 \cdots C_4$ and vertices of directed graph.

2) Secondly, if we assume that we are at vertex L_j, then data is aggregated to the next level, namely L_{j+1}. For each time series, we run selection criteria. There are many types of selection criteria, for instance, coefficient of variance analysis or average demand. In the next section, average demand is used. All time series that pass selection criteria on level L_{j+1} will be forecast on that level. We repeat the second step until we exhaust all system (1).

3 Comparison Results

This section outlines an experiment that was conducted to understand the differences in run time, accuracy and bias associated with the two aggregation methods. The data was taken from a retail client and is representative of the data that is seen in the field. There are 81,621 records all having at least a two year sales history and the total sales volume is 7,014,139 units.

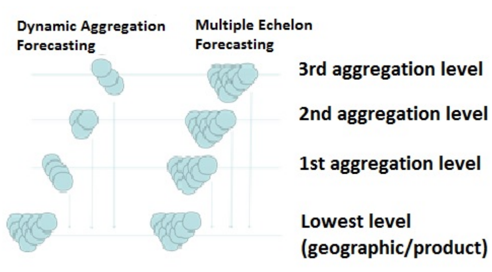

Fig. 3. Aggregation Approach Diagram

TABLE 1
DATA INFORMATION

Consumption Level	Initial Records	Records after aggregation	Consumption Level	
			geo	prod
1	81,621	81,621	geo1 geo2 geo3	prod1 prod2 prod3 prod4 prod5 prod6 prod7 prod8 prod9 prod10
2	81,621	52,330	geo1 geo2 geo3	prod1 prod2 prod3 prod4 prod5 prod6 prod7 prod8 prod9
3	81,621	26,043	geo1 geo2	prod1 prod2 prod3 prod4 prod5 prod6 prod7 prod8 prod9
4	81,621	3,839	geo1 geo2	prod1 prod2 prod3 prod4 prod5 prod6 prod7 prod8

The most granular level consists of three geography hierarchies and seven product hierarchies. In this experiment, we use SAS [1] and SAS High-Performance Forecasting [2]. For the purpose of this paper, we did not use exogenous variables, events, or regressions in the modeling process. As the table below shows, on average not only does dynamic aggregation run much more efficiently than multi-echelon, but also has improvement in both accuracy and absolute bias.

4 Conclusion

It is common to work with Big Data with many time series in a lot of retail and CPG applications. The run time savings are even more important in such ap-

TABLE 2
ACCURACY/ABS. BIAS COMPARISON

Accuracy/Absolute Bias					
Consumption Level	Type	Sales Weighted Accuracy	Sales Weighted Abs. Bias	Percentage Difference	
				Improvements in Accuracy	Improvements,in Abs. Bias
1	Multi-Echelon	37.31%	8.43%	2.84%	4.97%
	Dynamic Aggregation	40.15%	3.46%		
2	Multi-Echelon	49.41%	6.81%	3.25%	3.35%
	Dynamic Aggregation	52.66%	3.46%		
3	Multi-Echelon	62.76%	5.12%	2.87%	1.66%
	Dynamic Aggregation	65.63%	3.46%		
4	Multi-Echelon	83.45%	4.03%	0.85%	0.57%
	Dynamic Aggregation	84.30%	3.46%		

TABLE 3
EFFICIENCY COMPARISON

Efficiency		
Type	Weekly Average Time (hours/minutes/seconds)	Weekly Average Time Difference
Multi-Echelon	1 hour 41 minutes 29 seconds	1 hour 35 minutes 55 seconds
Dynamic Aggregation	5 minutes 34 seconds	

plications. Dynamic Aggregation is a stable, both highly accurate and highly fast methodology for forecasting systems with large amounts of time series inputs of varying volume and volatility.

References

[1] SAS, version 9.4. Cary, NC: SAS Institute Inc, 2013.
[2] SAS High-Performance Forecasting, version 12.3. Cary, NC: SAS Institute Inc, 2013.