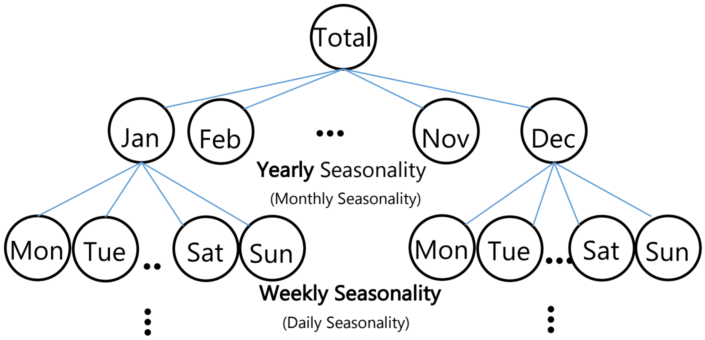
Bayesian Hierarchical Approach to Time Series with Multiple Seasonalities

2018/08
Hyunji(Angie) Moon,
SNU University

Hypothesis

Time series with mutiple seasonality can be analyzed with bayesian hierarchical model by discretizing their seasonal cycle. A seasonal cycle (e.g. Year or Week) is discretized into individual components (e.g. February or Monday) which constitute one layer of the hierarchicty. Parameters for trends, the other seasonalities, and holiday effects are **learned through partial pooling**. One set of time series data can be represented in various hierarchical structures.



New Attempts

This approach has its novelty in two aspects: time series are modeled **not with cross-sectional hierarchy, but with time unit hierarchy**. Also it **models time series with multiple seasonality in a discrete domain**, unlike the majority of the previous researches where gaussian process regression(GPR) was applied.

1. Time Unit hierarchy

Most Hierarchical modelings are based on **cross-sectional hierarchy.**

e.g. Hierarchical Modeling in Bayesian Inference:
- *Scores in 8 schools (Gelman),Tadpole mortality (McElreath)*

Hierarchical Modeling in Time Series Forecasting:
- *Hierarchical tourist forecasting based on geography (Hyndman)**
- *Hts R package: Hierarchical, grouped time series*

Other Studies with Similar Concepts:
*Alg. Predcition of Healthcare Costs (Bertsimas, D.)*** Clustering patients based on their cost, risk, acuteness, build prediction models for each cluster.

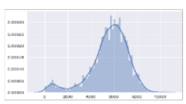
→ If an observation from one cluster improves inference for other clusters, hierarchical modeling should be applied. Updating sales result for Monday facilitates Tuesday's sales forecasting, as Monday and Tuesday share the popoultiation. Time has its own hierarchical structure which makes a 'time unit hierarchy' model natural and effective.

2. Limitation of Gaussian Process Regression

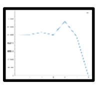
Time series data are divided into discrete and continuous domain. GPR lets us **pool the data on continous domain**. Time series with multiple seasoanlity are commonly addressed with GPR from continuous domain. But in reality, time series with multiple seasoanlity posses certain discrete characteristics. This is because, time series data are originally generated from discrete events that occur at certain point, and affect other points that correspond to certain hour, day, week, month, and quarter. The discrete nature of data generating processes justifies the discrete apporach as part of the generative modeling movement.

Theoretical and Realistic Grounds

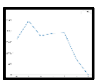
Immense differences exist between components in one seasonality: both in **data generating process** and their **statistical results (variance, mean)**. The left two figure shows the average sales in one week, and the right figure is the average daily accomplishments in one month. PA11 demand shows highest demand for Friday, whereas P94 displays highest demand on Tuesday. In both cases both mean and variance is more than two time bigger for the max compared to the min, which is Sunday. The second example's pattern is generated by 'evaluation event' which happens at the last day of every month. This illudes that the data generating process of each seasonal components are inherently different and approaching it as a discrete domain might be more effective than continuos GPR.



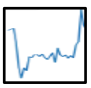
Bimodal profile implies different data generating process



PA11 demand



P94 demand)
(week)



accomplishment
(month)

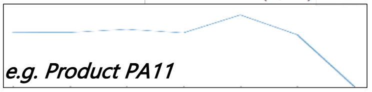
Evidence of the Improvement

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

growth $g(t) = (k + \mathbf{a}(t)^\intercal \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\intercal \boldsymbol{\gamma})$ $\quad \delta_j = 0$ w.p. $\frac{T-S}{T}$, $\delta_j \sim \text{Laplace}(0, \lambda)$ w.p. $\frac{S}{T}$

seasonality $s(t) = X(t)\boldsymbol{\beta}$ $\quad \boldsymbol{\beta} \sim \text{Normal}(0, \sigma^2)$

holiday $h(t) = Z(t)\boldsymbol{\kappa}$ $\quad \boldsymbol{\kappa} \sim \text{Normal}(0, \nu^2)$



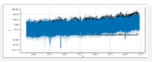
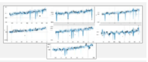
e.g. Product PA11

**Step 1:
K-means clustering
with avg. & var.**

[Mon, Tue, Wed, Thr, Fri, Sat, Sun]
V
[Mon,Tue,Wed,Thr,Sat] // [Fri] // [Sun]
V

**Step 2:
Each Cluster Adaptively
Learns its Parameters**

[Mon,Tue,Wed,Thr,Sat] : δ1, β1, κ1
[Fri] : δ2, β2, κ2
[Sun] : δ3, β3, κ3

	Complete Pooling	No Pooling	Partial Pooling
Concept Graph			
Parameters	Shared parameters (1 set)	Separate parameters (7 sets)	Parameters for each cluster (3 sets)
MAPE for PA11	19.752	14.235	13.217
MAPE for L14	10.081	6.717	5.368

Advantage of this Approach

Learning trend, seasonality, holiday, and other effects from partial pooling **balances between underfitting(complete pooling) and overfitting(no pooling)**. The following describes the additional advantage of modeling time serise with multiple seasonality with bayesian hierarchical models.

+ Flexible Addition of Seasonality as Layers

As Bayesian hierarchical model can *easily add layer as a hyperprior*, seasonality can be added as a new layer to the hierarchical structure

+ Clustering by Seasonal Component is possible

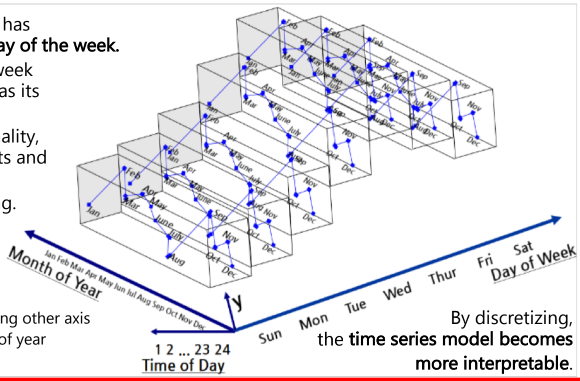
As Bayesian is *robust to missing data*, unlike classical time series approach, clustering by weekdays (or any other seasaonly components) is possible

+ Flexible Addition of Explanatory Variables

As Bayesian hierarchical model can *easily add new explanatory variables as new parameters*, explanatory Variables can be added flexibly + Overfitting is of less concern with reasonable prior distributions in Bayesian which makes it a better model than ARIMAX (General Transfer Function, explanatory var. lagged regression ARIMA)

Multidimensional Diagram of Time series with Multiple Seasonality

This diagram has **discretized day of the week.** Each day of week (Sun ~ Sat) has its own prior for trend, seasonality, holiday effects and learned by partial pooling.



Discretizing along other axis such as month of year is also possible.

By discretizing, the **time series model becomes more interpretable.**

Further Research

- Full Bayesian Hierarchy Modeling of Time Series
- Incorporating explanatory var. to the model with adaptive hyperprior
- MAP to MCMC

*Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis
***Bertsimas, D., Bjarnadóttir, M. V., Kane, M. A., Kryder, J. C., Pandey, R., Vempala, S., & Wang, G. (2008). Algorithmic prediction of health-care costs. *Operations Research*