

Time Unit Clustering (TUC) Model for Forecasting

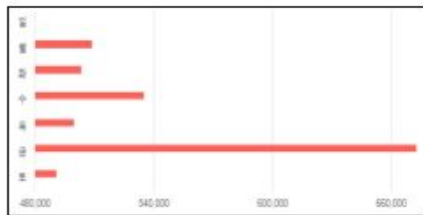
요약 자료

2018.10.12 문현지 작성 (NextOpt 대표)

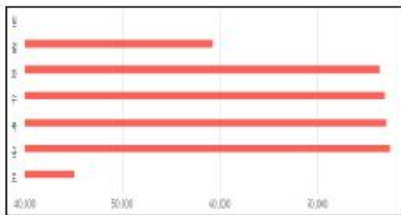
목차

1. TUC Model Background
2. TUC Model Contribution
3. TUC Model Flow Diagram
4. TUC Time Series Structure
5. TUC Application Example
6. Experimental Results

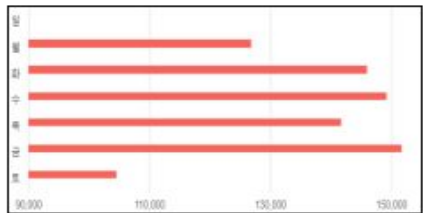
‘시간단위별 이산화’ 필요성



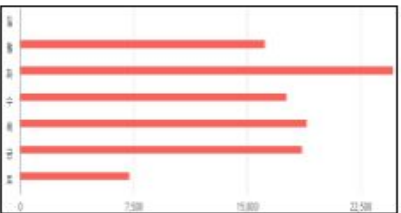
(a) Pallet 1375 week-day pattern



(b) Pallet 1057 week-day pattern



(c) Pallet 1091 week-day pattern



(d) Pallet 1817 week-day pattern

Figure 3. 네 가지 유형의 파렛트 이동량의 요일별 패턴

각 **group**: 파렛트 유형

Time Unit: 요일(week-day)

실제 관찰 결과 파렛트 유형에 따라 요일별 이동량 패턴 상이.

이 group의 고유 패턴을 반영하는 예측모델이 필요.

Time Unit Clustering(TUC) 모델 제안.

1. TUC Model Background

1. 시계열 예측 세분화

- a. ex. 30개 품목 취급 기업이 [과거: 총 판매량(30개 합)만 예측 -> 현재: 30개 품목별 예측]
- b. 예측대상이 '총 합'에서 하위 '유형, 계약처, 품목, 지역' 등으로 세분화되면, 각 group특성이 분명해지고 따라서 예측정확도 측면에서 group특성 반영 과정이 중요

2. 데이터 수집 빈도 증가에 따른 계절성 복잡화(complex seasonality)

3. 데이터 종류 다양화에 따른 설명변수 후보 풍부화

빨간색: 주요 contribution

2. TUC Model Contribution

1. 시계열 예측 세분화

- a. 기존 모델1(arima, ets) 예측의 각 group의 여러 특성 반영 어려웠던 문제점 개선
- b. 기존 모델2(prophet**) 예측의 각 group의 'Time Unit별 패턴' 특성 반영 어려웠던 문제점 개선

2. 데이터 수집 빈도 증가에 따른 계절성 복잡화

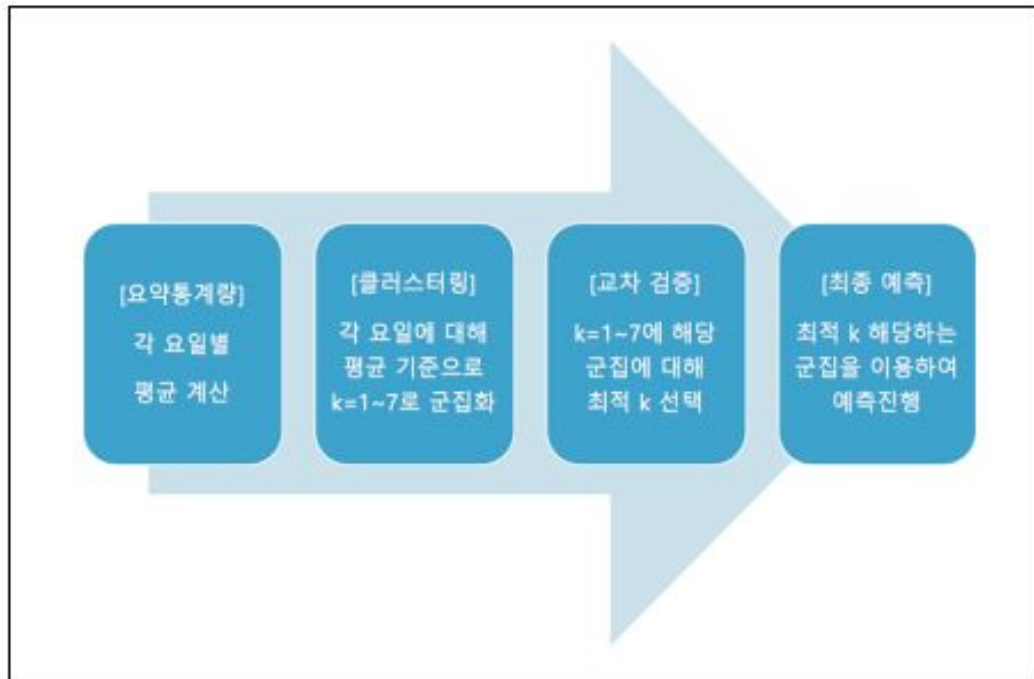
Fourier Series 이용 복잡계절성 문제 개선

3. 데이터 종류 다양화에 따른 설명변수 후보 풍부화

Multilevel(Hierarchical) Model 이용 해석가능성, 예측정확도 문제 개선

+ 데이터 변화를 위의 3가지 특성의 관점으로 각각을 맞는 방식으로 모델링 (3개 방식의 결합)

3. TUC Model Flow Diagram



1. summary statistics for each time unit group
2. clustering based on the summary statistics
3. cross validation for the number of cluster (k)
4. forecasting

Figure 4. 시간단위 군집화 모델의 진행단계

4. TUC Time Series Structure

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

일반 time series decomposition을 따라

1. growth
2. seasonality 성분

+ holiday 성분 (demand forecasting에 중요*)

growth	$g(t) = (k + \mathbf{a}(t)^\top \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\top \boldsymbol{\gamma})$	$\delta_j = 0 \text{ w.p. } \frac{T-s}{T},$ $\delta_j \sim \text{Laplace}(0, \lambda) \text{ w.p. } \frac{s}{T}.$
seasonality	$s(t) = X(t)\boldsymbol{\beta}$	$\boldsymbol{\beta} \sim \text{Normal}(0, \sigma^2)$
holiday	$h(t) = Z(t)\boldsymbol{\kappa}$	$\boldsymbol{\kappa} \sim \text{Normal}(0, \nu^2)$

각 성분별로 다음 함수 이용

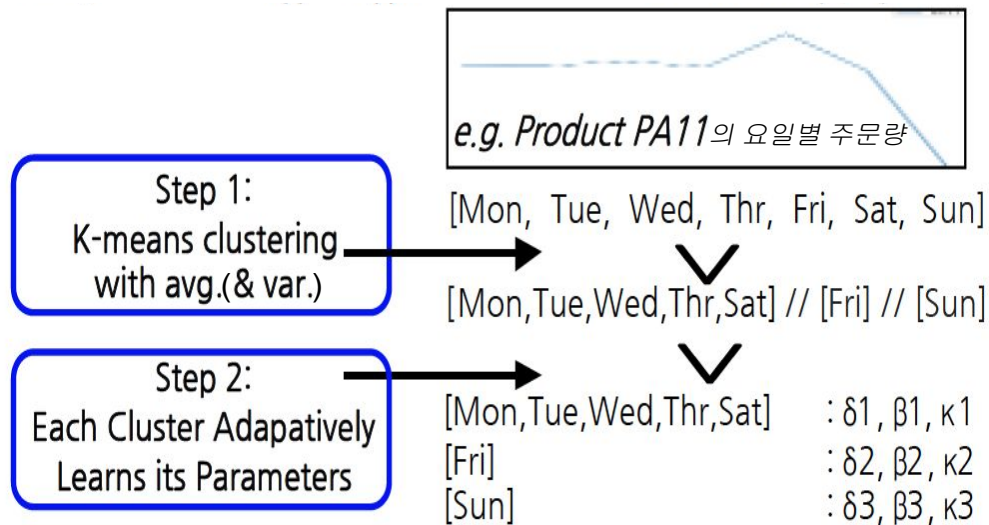
trend: piecewise linear function

seasonality: fourier series

holiday: indicator function

* Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45.

5. TUC Application Example



→ time unit별 clustering을 진행하고,

→ 각 cluster마다 독립적으로 모수 학습

(cluster-wise prediction)

6. Experimental Results

Table 4. 선택된 군집화 집합과 기존 프로펫 모델의 예측오차(MAE) 비교

파렛트 종류	선택된 군집화 집합	프로펫 모델	시간단위 군집화 모델
1815	'0', '1234', '5', '6'	5,281	3,597
1041	'012345', '6'	1,056	1,054
1627	'05', '134', '2', '6'	503	453

기존 모델2(prophet)과 비교해 TUC 모델이 평균 10~20% 예측정확도 향상

(기존 모델1과 비교한 기존 모델2의 예측정확도 향상은 선행연구에서 증명되었음)