

시간단위 전력사용량 시계열 패턴의 군집 및 분류분석

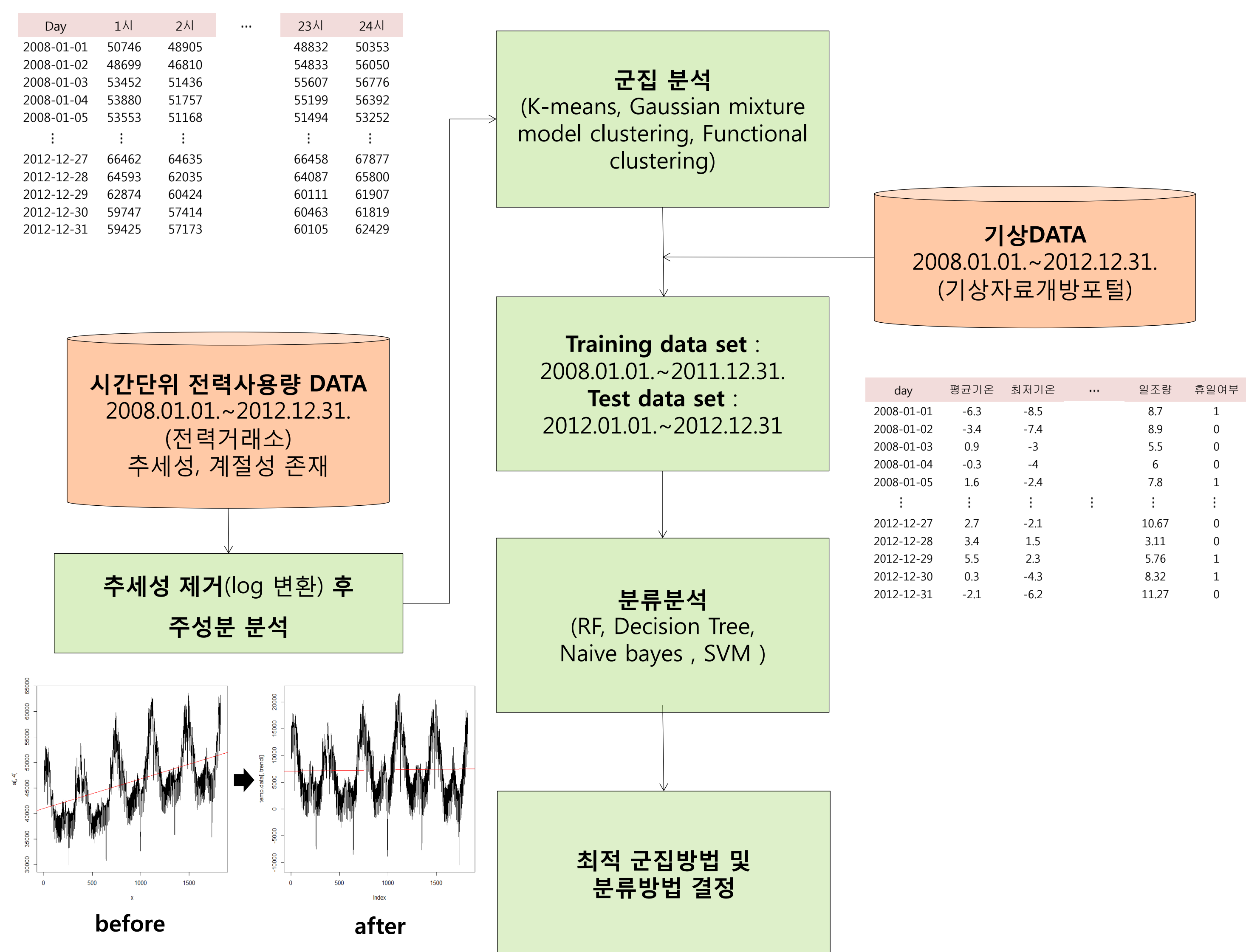
박다인 (석사과정, 대구대학교 통계학과)
윤상후 (조교수, 대구대학교 전산통계학과)



Purpose

전기는 저장이 불가능하기 때문에 전력거래소의 전력수요 예측은 **전력시스템**의 **안정적**이고 **경제적**인 운전에 필수적이다. 본 연구에서는 시간단위 전력 사용량 **시계열 데이터를 군집하고 분류 하는 방법**에 대해 살펴보고자 한다.

Research flow



Methodology

◆ clustering

- K-means clustering:** 각 클러스터의 중심(μ_i)과 그룹 내 자료간 거리 차이의 분산을 최소화하는 방식

$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} ||x - \mu_i||^2$$

- Gaussian mixture model:** 혼합 가우시안 분포기반 집단의 중심점 거리에 있는 객체들을 동일 군집으로 분류(타원형자료)

$$\phi_k(x | \mu_k, \sum_k) = (2\pi)^{-\frac{p}{2}} |\sum_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^t \sum_k^{-1} (x_i - \mu_k)\right)$$

- Functional clustering:** 2개의 주성분 자료를 하나의 함수로 간주하여 함수간 거리에 따른 군집 분석

$$Y_{ij} = \mu_k(t_{ij}) + b_j + \epsilon_{ij}, \quad y_j \sim N(\mu_k, \sum_k)$$

◆ classification

• Decision Tree

트리 구조를 통한 노드 분할 방법으로 각 마디에서 적절한 최적 분류 규칙을 찾는 **성장단계**, 오차를 크게 할 위험이 높거나 부적절한 추론규칙을 제거하는 **가지치기 단계**, 이익도표, 위험도표 등을 통한 **평가**를 통해 구축

• Random forest

여러 개의 결정 트리들을 임의적으로 학습하는 방식의 앙상블 방법

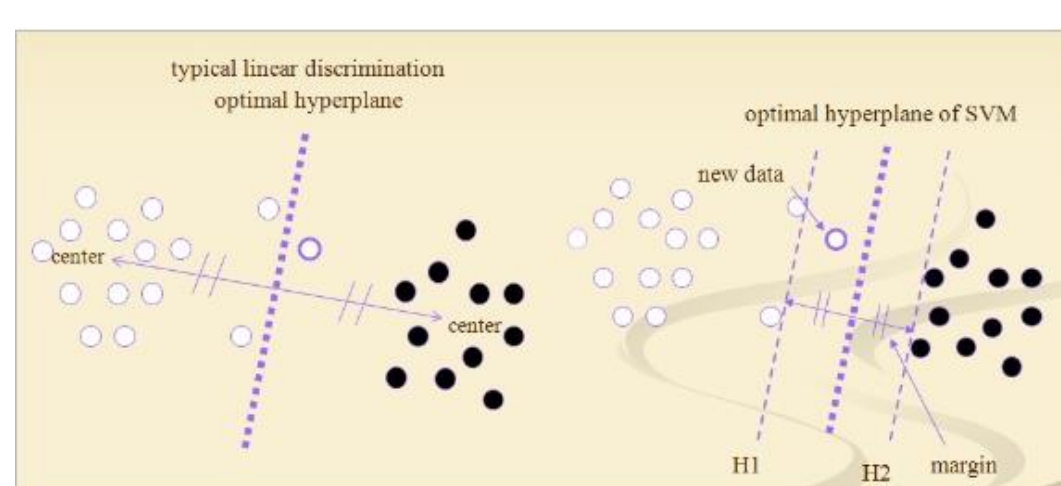
• Naïve Bayes

베이즈 정리를 이용하여 분류하고자 하는 대산의 각 분류별 확률을 토대로 분류

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}.$$

• SVM (Support Vector Machine)

패턴의 분포를 가장 잘 분류할 수 있는 기준선을 찾아내는 알고리즘



Result

◆ 주성분 분석 결과

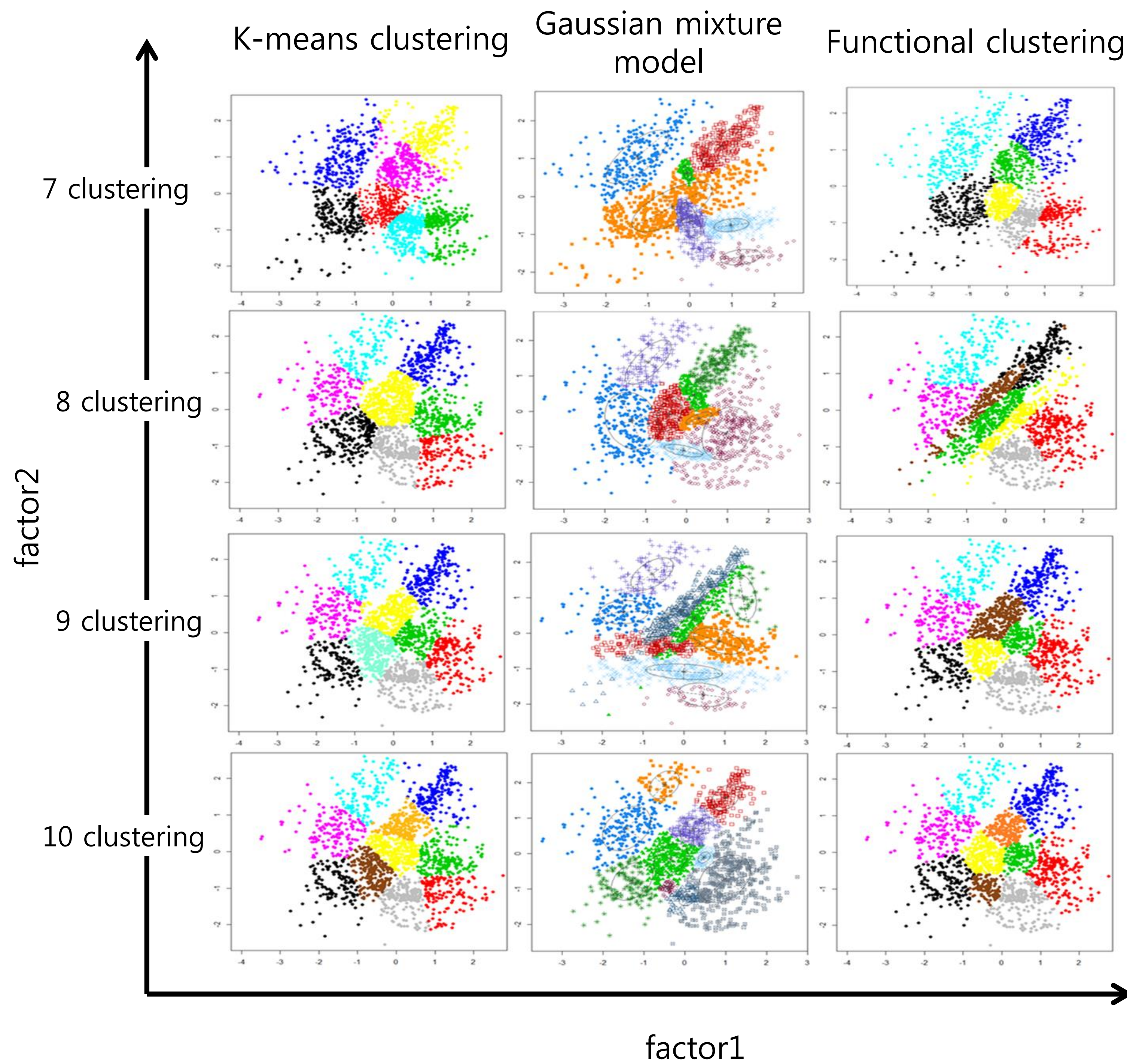
Factor1: Day time
Factor2: Night time

	Factor1	Factor2
SS loadings	13.826	9.551
Proportion Var	0.576	0.398
Cumulative Var	0.576	0.974

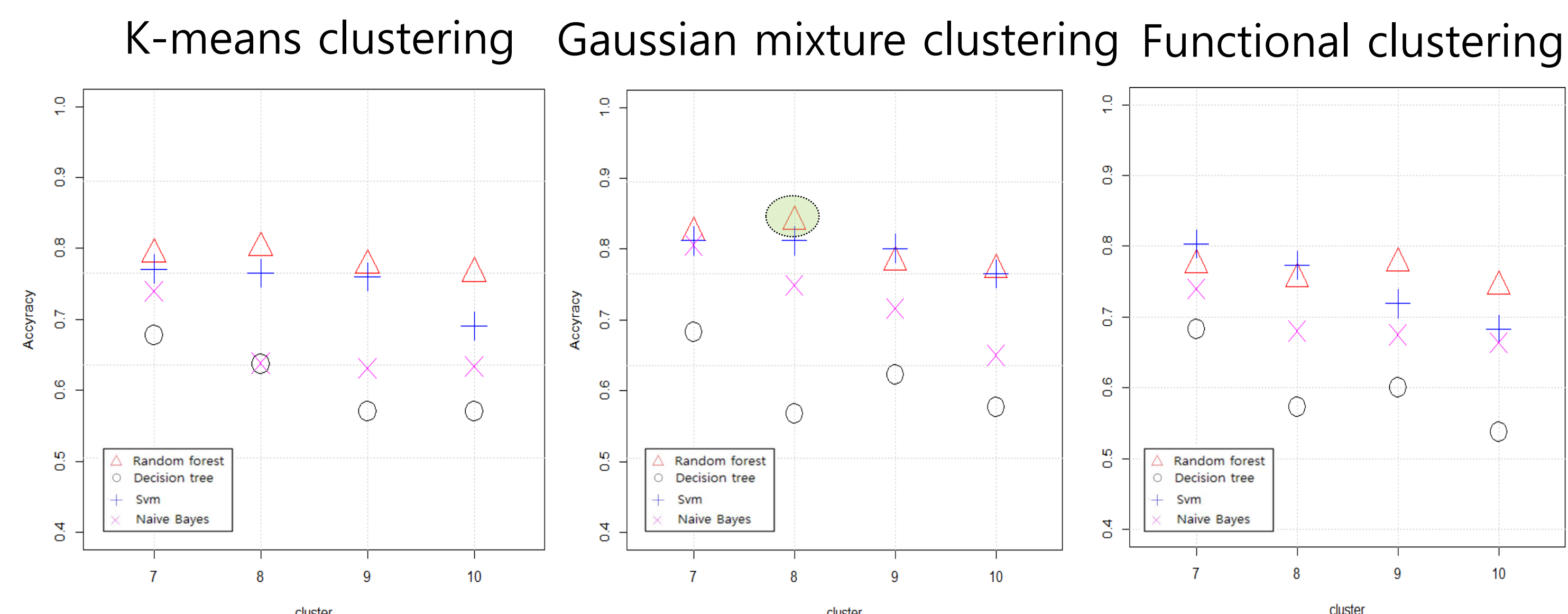
두개의 주성분으로 전체 분산의 약 **97.4%**

	Factor1	Factor2
MW01	0.230	0.961
MW02	0.229	0.959
MW03	0.233	0.960
MW04	0.233	0.966
MW05	0.225	0.974
MW06	0.235	0.967
MW07	0.363	0.916
MW08	0.610	0.768
MW09	0.814	0.545
MW10	0.896	0.410
⋮	⋮	⋮
MW22	0.896	0.399
MW23	0.693	0.680
MW24	0.556	0.791

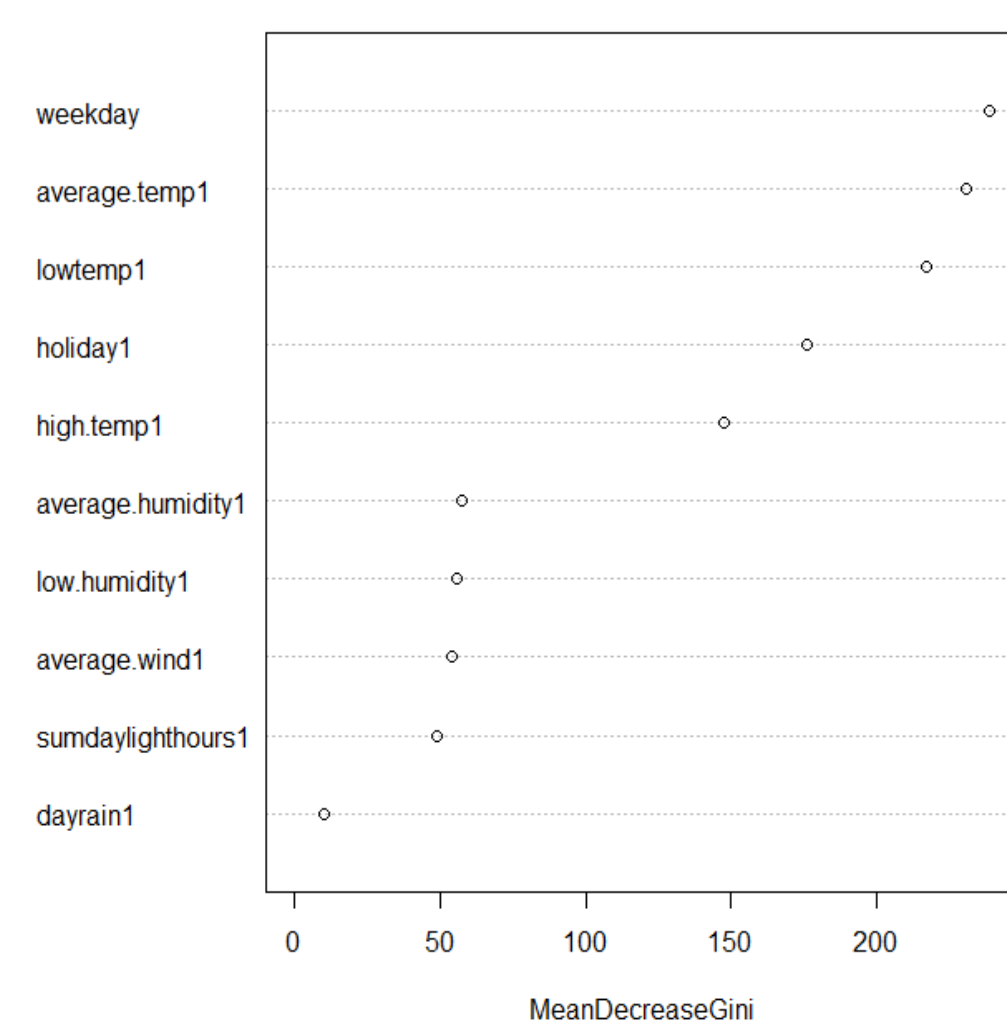
◆ 군집 분석 결과



◆ 분류 분석 결과



Gaussian mixture clustering (8 clustering)



- 분류 분석 결과 **Gaussian mixture model clustering의 8clustering을 Random forest**로 예측하였을 때 **정확도가 약 85%**로 가장 높았으며, 각 군집에 대한 분류분석 방법에서도 대체적으로 Random forest 방법이 가장 좋았다.
- Gaussian mixture model clustering에서 8군집의 변수 중요도는 **요일 - 평균기온 - 최저기온 - 휴일 - 최고기온**의 순으로 나타났다.
- 향후 연구는 군집별 평균 패턴을 추정하여 군집의 특성을 기상요소와 요일요소로 설명하는 것이다. 이를 위해 선정된 8개의 군집별 특성을 구체적으로 살펴볼 예정이다.