

비지도학습 - 기본

목차

1-1 비지도학습

1-2 군집(clustering)

1-3 변환

1-4 연관

1-1 비지도 학습

- ▶ 군집화(clustering)
- ▶ 변환(transform)
- ▶ 연관(association)

1-1 비지도 학습

▶ 군집화(clustering)

- 비슷한 것들을 찾아서 그룹을 만드는 것.

▶ 변환(transform)

- 비지도 변환은 데이터의 원래 표현과 비교하여 인간이나 다른 머신러닝 알고리즘이 더 쉽게 이해할 수 있는 데이터의 새로운 표현을 생성하는 알고리즘

▶ 연관(association)

- 서로 연관된 특징을 찾아내는 것. 일명 장바구니 분석이라 한다.
- 서로 관련 있는 특성(열)을 찾아주는 머신러닝 기법

1-1 비지도 학습

▶ 사용되는 분야

- (1) Market segmentation – 시장 점유율 분석
- (2) Social Network 상에서의 활동하는 사람들이 연결성 분석
- (3) 구글이나 Facebook에서의 추천 서비스 활용

1-2 군집화(Clustering)

▶ 군집화(clustering) 기본 이해

(1) 비지도학습의 대표적인 기술. X 에 대한 레이블이 없는 데이터를 그룹핑하는 분석 알고리즘

(2) 데이터 집단(클러스터)를 정의. 클러스터란 비슷한 특성을 가진 데이터들의 집단

1-2 군집화(Clustering)

▶ 군집화(clustering) 알고리즘의 종류

(1) K-means

(2) Mean Shift

(3) Gaussian Mixture Model(GMM)

(4) DBSCAN

(5) Agglomerative Hierarchical Clustering(통합 계층적 클러스터링)

1-3 변환

▶ 변환

(1) 기존의 데이터를 더 쉽게 해석할 수 있도록 새롭게 표현하는 알고리즘

1-3 변환

▶ 변환의 예

- (1) 차원 축소(Dimensionality reduction)
- (2) 비음수 행렬 분해(Non-negative matrix factorization)
- (3) t-SNE(t-distributed stochastic neighbor embedding)

1-3 변환

▶ 변환의 예

- (1) 차원 축소(Dimensionality reduction)
- (2) 비음수 행렬 분해(Non-negative matrix factorization)
- (3) t-SNE(t-distributed stochastic neighbor embedding)

1-4 association(연관)

▶ association(연관)

(1) 데이터 간의 관계(relationship)혹은 패턴(pattern)을 학습

(2) 특징(feature)을 그룹핑.

가) 라면을 구매한 사람들은 김치를 구매 -> 라면과 김치의 상관관계 찾기

(3) 상품을 예로 들어 설명한다면 연관성이 높은 상품을 추천하는 것.

(4) 연관규칙 알고리즘은 진화를 거듭하여 지금은 개인화 추천/협업 필터링으로 발전.

1-4 association(연관)

▶ association(연관) 알고리즘

(1) Apriori 알고리즘

(2) FP-growth 알고리즘

1-4 association(연관)

▶ association(연관) 알고리즘

(1) Apriori 알고리즘

가. 가장 빈번하게 발생하는 아이템 세트를 찾기 위한 목적으로 제안된 최초 알고리즘.

나. 후보 항목 집합을 구성한 후, 사전 지식(priori knowledge)을 이용하여 패턴 아이템 집합을 생성.

(2) FP-growth 알고리즘

가. 빈출 패턴 성장(FP-Growth : frequent Pattern Growth algorithm) 알고리즘.

나. Apriori 알고리즘 단점으로 생각되었던 DB 스캔(검색) 횟수와 후보 집합 생성 과정을 효율적으로 개선한 알고리즘