**연구노트**

**(수학과)이 재 봉**  2017.05.10 (수)

|  |  |  |
| --- | --- | --- |
|  | | |
| **연 구**  **주 제** | **특허분석을 통한 미래기술 예측** | |
| **데이터 마이닝 기술 조사** | |
| **구 분** | **연구내용** | **진행률** |
| **연 구**  **내 용** | ■ **데이터 마이닝**  [정의]  **데이터 마이닝(Data Mining)**이란 의미 있는 패턴과 규칙을 발견하기 위해서 자동화되거나 반자동화된 도구를 이용하여 대량의 데이터를 탐색하고 분석하는 과정이다(Berry and Linoff, 1997, 2000).  [적용분야]  데이터마이닝은 다양한 분야에서 활용되고 있다. 군사분야에서는 미사일의 정확도에 영향을 주는 요인들이 어떠한 작용을 하는지를 알아내기 위해 데이터 마이닝을 활용하고 있으며, 국가정보기관은 엄청난 양으로 도청되는 통신들 가운데 특히 중요성이 높은 통신을 찾아내기 활용하기도 한다. 한편 보안전문가들은 패킷별로 네트워크에 위협요인을 갖고 있는지를 판단하기 위해, 그리고 의학연구자들은 암의 재발가능성을 예측하기 위해 데이터 마이닝을 사용한다.  [주요개념]  ㆍ분류(Classification): 분류결과가 알려진 유사 데이터를 사용하여 규칙들을 찾아낸 다음, 그 규칙들을 분류결과가 알려지지 않은 해당 데이터에 적용하는 것이다.  ㆍ추정(Estimation): 연속형 변수의 값을 예측한다.  ㆍ예측(Predication): 집단의 값을 예측한다는 점을 제외하고는 분류와 추정 문제와 유사하다.  ㆍ연관성규칙(Association Rules): 대량의 고객거래 데이터베이스는 구매항목들 간의 연관성, 즉 어떤 항목이 어떤 항목과 관련되는지에 대한 분석에 알맞다.  ㆍ군집화(Clustering): 군집화는 이질적인 집단을 몇 개의 동질적인 소집단을 세분화하는 작업이다. 군집화가 분류와 구분되는 가장 큰 특징은 사전에 정의된 집단을 가지고 있지 않다는 것이다.  [참고문헌]  Galit Shmueli 《Data Mining for Business Intelligence : Concepts, Techniques, and Applications》 |  |
| **차 주**  **계 획** | ■ 계획  - 텍스트 마이닝 기술 조사 | |
| **이 슈** | ■ **해당사항 없음** | |

|  |  |
| --- | --- |
|  |  |