

## 제목: 데이터 마이닝 기법

-의사결정나무 분석에서 분류 기준법의 비교

컴퓨터의 발달과 자료의 축적으로 인하여 많은 양의 자료를 분석하여 우리가 알지 못했던 사실들을 찾아내는 기법이 데이터마이닝이다. 자료를 분석하는 여러 가지 데이터마이닝 기법 중 비통계학자들에게도 이해하기 쉬운 것이 의사결정나무 방법 (Decision Tree Method) 이다. 의사결정나무 방법은 반응변수의 성질에 따라 회귀 의사결정나무(Regression Trees)와 분류의사결정나무(Classification Trees) 로 구분된다.

본 강연에서는 먼저 의사결정나무 분석방법의 간단한 소개와 분석방법을 알아 본 후, 의사 결정나무를 만드는데 필요한 새로운 분류기준법을 제시하여 기존의 방법들과 비교하고자 한다. 전통적인 분류기준법은 왼쪽과 오른쪽의 자식노드들을 그들의 노드숫자를 가중치로 한 가중평균으로 하여 분류변수와 기준값을 선택하였다. 그러나 제시하는 새로운 방법은 가중평균을 하는 대신 우리가 관심있는 노드를 효과적으로 선택하고 나머지는 무시해 버림으로써 우리가 원하는 부분을 가능한 한 빨리 선택하고자 하는 것이 목적이다. 이렇게 해서 만들어진 의사 결정나무는 비록 정확성은 기존의 방법에 비해 다소 떨어질진 모르나 매우 설명력이 뛰어나고 이해하기가 쉽다. 정확성과 설명력 사이에는 상호 조정 관계 (Trade-Off) 에 있기 때문에 어느 한 쪽을 강조하면 다른 쪽은 잃어버릴 수 있다. 대부분의 기존 방법들은 정확성을 높이는 데에 중점을 두기 때문에 설명력에는 큰 관심이 없었다. 반면 여기에서는 설명력에 더 중점을 두었다.

마지막으로, 실제자료를 이용하여 회귀의사결정나무와 분류의사결정나무에 적용한 결과를 토대로 하여 기존의 방법과 비교 분석하고자 한다.