

미디어학 석사학위 논문

# 통계와 시각화를 결합한 데이터분석:

## 예측 모형 대한 시각화 검증

Data analysis by Integrating statistics and  
visualization: Visual verification for the prediction  
model

아주대학교 대학원

라이프 미디어 협동과정

문성민

통계와 시각화를 결합한 데이터분석:  
예측 모형에 대한 시각화 검증

지도교수 이 경 원

이 논문을 미디어학 석사학위 논문으로 제출함

2015년 12월

아 주 대 학 교 대 학 원

라 이 프 미 디 어 협 동 과 정

문 성 민

문성민의 미디어학 석사학위 논문을 인준함

심사 위원장               이 경 원               인

심 사 위 원               신 현 준               인

심 사 위 원               유 재 인               인

아 주 대 학 교   대 학 원

2 0 1 5 년   1 2 월 1 5 일

## 요 약

최근 정보통신의 발달과 함께 예측분석의 활용성이 중요해 지고 있으며 이러한 예측분석은 우리의 실생활과 밀접한 관계가 있다. 하지만 예측 분석은 패턴인식 (Pattern recognition) 혹은 기계학습(Machine learning)으로 불리는 확률적 학습 알고리즘을 기반으로 하기 때문에 사용자가 분석과정에 개입하여 더 많은 정보를 얻어내기 위해서는 높은 통계적 지식수준이 요구된다. 또한 사용자는 분석 결과의 다른 정보를 확인 할 수 없고 데이터의 특성 변화와 데이터 하나하나의 특성을 파악하기 힘들다는 단점이 있다. 본 연구는 이러한 예측분석의 단점을 보완하고자 통계적인 데이터 분석 방법과 시각화 분석 방법을 결합하여 데이터 분석을 진행하였으며 통계적인 분석 방법만을 진행 할 경우 발생하는 단점을 보완하고 데이터에서 더 많은 정보를 도출해 내기 위한 방법론을 제시 하고자하였다. 또한 본 연구는 통계적인 분석과 시각화 분석을 결합하여 분석을 진행 할 때 영화의 흥행성을 예측하는 것을 목적으로 하였으며 분석을 통해 도출된 결과는 다음과 같다. 첫째, 의사결정나무분석에서 제시된 분할 기준이 적용될 때 마다 변하는 데이터의 패턴을 파악할 수 있다. 둘째, 제시된 최종 예측 모형에 포함된 데이터들의 특성을 확인 할 수 있다. 본 연구의 시사점은 예측모형의 단점을 보완하고 데이터로부터 더 많은 정보를 추출하기 위해 통계적인 데이터 분석과 시각적인 데이터 분석을 결합하여 시행하였다는 것이다. 통계적인 분석 방법을 통해 각 변수의 관계를 파악하고 높은 영화 흥행성을 예측하기 위한 예측모형을 도출하였으며, 시각화 분석에서는 변수들의 분포를 파악하는 사용자 인터랙션이 가능한 다양한 기능을 제공함으로써 최종적으로 제시된 예측모형을 검증하고 데이터로부터 더 다양한 정보를 도출하기 위한 방법론을 제시하였다.