

베이지안 선택 모형을 이용한 영화 흥행 예측

Predicting Financial Success of a Movie Using Bayesian Choice Model

이경제*, 장우진**

* 서울대학교 기술경영대학원 (strutjet@tam.snu.ac.kr)

** 서울대학교 산업공학과 (changw@snu.ac.kr)

Abstract

영화는 대표적인 경험재로 가치판단이 주관적이고 제품 수명주기가 매우 짧아 예측의 불확실성이 높기 때문에 이를 정량적인 방법으로 모형화하기는 쉽지 않다. 이러한 한계점에도 불구하고 한 영화의 상업적 성공을 예측하는 것은 영화 제작자나 배급사, 극장 등 모든 주체에게 수익과 직결되는 중요한 문제이기 때문에 지금까지 다양한 통계 모형이 제시되었다. 그러나 이들 모형의 대부분은 영화 흥행에는 영향을 미치나 측정할 수 없는 효과를 반영하지 못한다거나, 추정 모수의 효과가 모든 영화에 대해서 같다는 동일성 가정으로 인해 영화간 이질성을 고려하지 못하고 있다. 따라서, 본 연구에서는 추정 모수의 사전분포를 모호사전분포로 정의함으로써 변수들의 불확실성을 반영할 수 있고, 영화간 이질성을 고려할 수 있는 베이지안 선택 모형을 제안하였다. 모수의 사후분포는 마코프체인 몬테카를로 기법인 깃스 샘플러를 이용하여 추정하였다. 또한, 감독, 배우, 장르 등의 영화 별 속성 변수뿐만 아니라, 입소문에 의한 영화관람 결정 등의 구전효과와 경쟁영화의 개봉으로 인한 효과를 반영할 수 있는 변수를 추가하여 모형의 정확성을 높였다. 2005년과 2006년 상반기에 상영된 영화를 바탕으로 모형을 구축하고 인공신경망 모형과 비교한 결과, 전체적인 예측 정확도에서는 인공신경망 모형과 비슷한 결과를 보이나 상업적으로 성공한 영화를 예측하는 데에는 베이지안 선택모형이 보다 더 우수한 것으로 나타났다. 또한, 개봉 주의 경쟁심화 정도 및 개봉 첫 주의 스크린 수 등이 영화 흥행에 가장 중요한 변수로 나타났으며, 영화 개봉 전 그 영화에 대한 기대치가 높을수록 흥행 성적 또한 좋을 수 있었다. 배우의 힘 및 계절성, 영화 평점 등은 이질성을 고려하지 않은 전체수준에서는 통계적으로 유의하지 않은 것으로 나타났으나, 그룹 간 이질성을 반영한 모형에서는 어느 정도 흥행한 영화를 만들기 위해서는 고려되어야 할 요소로 나타났다.

Key words: 베이지안 선택 모형, MCMC(Markov Chain Monte Carlo), 인공신경망, 이질성, 구전효과, 경쟁효과

1. 서론

영화는 도서, 음반등과 함께 대표적인 문화상품으로 생활수준이 향상되고 문화 소비에 대한 욕구가 높아지면서 이들에 대한 수요는 크게 증가하고 있다. 이들 문화상품의 대부분은 경험재(Experience Good)이기 때문에 상품에 대한 가치판단이 주관적이고, 수명 주기가 일반 상품과 비교해 매우 짧기 때문에 수요예측이 쉽지 않다. 그럼에도 불구하고 끊임없이 새로운 영화를 만들어야 하는 영화 제작사와 수많은 영화중에서 성공가능성이 높은 영화를 선택해야 하는 영화 배급사들의 경우 영화에 대한 정확한 수요예측은 위험 요소의 감소 및 시장에서의 성공을 위한 필수요소라 하겠다(김휴중 1997).

흥행성과에 영향을 미치는 요인에 대한 탐구는 영화를 문화산업으로 분석함에 있어 가장 선행되어야 하는 작업 중 하나이다. 우수한 영화가 시장에서 참패를 하기도 하고 전혀 의외의 작품이 흥행에 성공을 하기도 한다. 영화시장에서 성공을 거두기 위해서는 어떠한 요소들이 영향을 미치는가에 대한 연구는 구조적인 측면에 대한 이해를 통해서 영화 산업을 이해하는 초석이 될 뿐 아니라 스크린 쿼터나 영상산업 진흥책과 같은 영화 관련 정책을 수립하는데 있어서도 기초 자료가 되기도 한다. 텔레비전 프로그램의 시청률에 미치는 영향에 대한 연구가 꾸준히 이루어진데 비해 유사한 문제인 영화의 흥행성과에 미치는 영향에 대한 연구는 상대적으로 적었다. 이는 영화가 텔레비전에 비해 언론매체로서의 영향력이 적다는 점과 철저히 개인의 선택에 의해 소비된다는 인식 때문이며(유현석 2002), 또 그 동안 우리 영화가 그 다지 산업적으로 주목받을 정도가 아니었다는 점에서 기인한다.

1.1 연구의 목적 및 범위

따라서, 본 연구에서는 한국 영화 시장에서 흥행성을 평가하는 요소는 무엇인지 살펴보고, 영화의 상업적 성공을 판단할 수 있는 정량적 모형을 제시하고자 한다. 또한, 흥행에 성공한 영화와 실패한 영화간의 이질성(Heterogeneity)은 존재하는지, 존재한다면 그것을 바탕으로 얻을 수 있는 정책적 함의는 무엇인지를 알아보하고자 한다. 본 연구에서 제시한 베이지안 선택모형과 인공신경망과의 비교를 통해 베이지안 선택모형의 타당성을 검증하고자 한다.

2. 기존연구

2.1 영화 흥행 결정 요인

어떤 영화의 흥행 성과 예측은 쉽지 않음에도 불구하고 과연 어떤 요인이 흥행에 영향을 미치는지에 대해서는 많은 연구가 진행되어 왔다. 주로 헐리우드 영화에 대한 분석이 대부분이지만 이들의 결과는 시사하는 바가 크다. 먼저, 영화를 구성하는 기본적 속성들 즉, 장르나 감독, 배우 등에 대한 고려는 모든 영화 분석에 있어서 필수적으로 포함되고 있다(Levin et al. 1997, Albert 1998, Bagella and Becchetti 1999, Vany and Walls 1999). 이와 함께 마케팅 효과 및 구전 효과, 경쟁 효과 등의 영화 외적 요인들도 영화 흥행을 결정하는 중요한 요소로 받아들여지고 있다(Mahajan et al. 1984, Jedidi et al. 1998, Elberse and Eliashberg 2003).

i) 영화 속성 관련 요인

대부분의 분석에서 감독의 힘이 영화 흥행에 미치는 영향력은 거의 없거나 배우의 힘보다 작은 것으로 나타나고 있지만, Levin et al. (1997)은 설문조사를 통해 영화는 경험재로서 감독이나 배우가 자체가 독립적인 브랜드를 가진다고 하였다. Albert (1998)는 영화흥행의 기본적으로 출연 배우가 전작에서 어떤 인상을 남겼느냐가 현재 상영중인 영화의 흥행패턴을 좌우한다고 가정하고 이를 실증적으로 분석하였다.

반면, Vany and Walls (1999)는 극장 수입의 불확실성을 스타가 줄일 수 있는가라는 의문을 제기하면서 해답은 스타가 아니라 영화 그 자체에 있음을 보였다. 이탈리아 영화 시장을 분석한 Becchetti (1999) 또한 스타와 감독의 힘도 중요하지만 영화 제작사의 능력도 중요하다고 주장하였다. 영화의 장르의 중요한 요소이기는 하나 많은 연구에서 흥행과는 관련이 적은 것으로 나타났다(Albert 1998, Ravid 1999).

ii) 영화 외적 요인

이런 영화의 기본 속성들 외에 마케팅 효과, 구전 효과, 경쟁 효과 등을 영화 외적 요소들을 고려한 연구들이 있다. Ravid (1999)는 마케팅 비용과 영화흥행과는 양의 상관관계가 있음을 보였고, Zufryden (2000)은 웹 사이트를 통해 효과적인 마케팅을 전개하면 영화의 수명주기를 늘릴 수 있다고 하였다. 이는 박형현 (2001)과 손원경 (2005)의 국내 문헌에서도 같은 결과를 볼 수 있으며, 앞으로는 더욱더 마케팅의 효과가 클 것으로 결론짓고 있다. 마케팅 효과와 관련하여 Ainslie et al. (2005)은 거대 영화 배급사일수록 마케팅 효과가 크음을 보였다.

영화 공급자에 의해 진행되는 마케팅의 효과를 제외하고 흥행 성격에 중요한 요소로 고려되고 있는 것은 구전효과이다(Mahajan et al. 1984, Eliashberg and Shugan 1997). Mahajan et al. (1984)은 영화나 음반, 서적 같은 경험재의 경우는 구전효과가 제품의 확산에 중요한 역할을 하는데 특히, 부정적인 구전으로 인해 확산이 지체될 수 있음을 보였다. 영화와 관련해 입으로 전해지는 직접적인 구전효과 말고 대중매체를 통해 전해지는 영화비평 또한 구전의 중요한 수단인 되고 있을 뿐 아니라 결정적인 요인은 아니지만 영화 흥행과 유의한 관계가 있음을 보이고 있다(이소민 2005).

또한, 영화가 상영될 당시 기존 상영영화와의 경쟁 등 시장상황을 고려한 요소들도 많이 고려되었다(Jedidi et al. 1998, Elberse and Eliashberg 2003). 최대 상영관 수나 개봉 첫 주 상영관 수, 경쟁 영화를 고려한 영화 개봉 시점등과 관련된 변수들을 통해 경쟁 효과 등을 측정하고자 하였다. 이들 변수들 중 개봉 첫 주의 상영관 수는 영화 흥행에 가장 중요한 변수로 나타나고 있다.

2.2 분석 기법

앞 절에서 제시된 요인들을 변수로 하여 영화 흥행과의 인과관계를 살펴보는 주된 방법은 회귀분석이었다. Elberse and Eliashberg (2003)는 스크린 수와 수익을 내생변수 (Endogenous Variable)로 정의하고 감독, 배우, 마케팅 비용 등을 외생변수로 하는 로그-선형 형태의 연립방정식을 제안하였다. 내생변수의 존재로 인해 오차항이 서로 연관되어 있기 때문에 OLS 보다는 2SLS나 3SLS 접근법을 제안하고 있으나 결과를 보면 거의 차이가 없음을 알 수 있다. 이는 모형은 연립방정식 형태이나 실제 추정과정에서는 수익에 대한 항이 두 식에서 서로 다르게 나타나기 때문에 결국 독립된 두 식을 추정하는 것이 되어 동시에 추정하나 각각 따로 추정하나 같은 결과를 얻게 되는 한계점을 지닌다. 무엇보다 영화 개봉 첫 주와 그 다음 주들에 대한 방정식을 따로 수립함으로써 개봉 첫 주의 영향을 효과적으로 반영할 수 없었다. 또한, 흥행에 성공하지 못한 영화 (Sleeper-Type)나 블록버스터 타입 (Blockbuster-Type)의 영화 등 수익과 스크린 수에 있어 매우 다른 패턴을 보이는 영화에 대한 이질성(Heterogeneity)을 고려하지 못했다. 이와 관련해 Jedidi et al. (1998)은 관객수가 지수적으로 감소한다는 가정하에 유한 혼합 모형 (Finite Mixture Model)을 이용해 102편의 영화를 5개의 군집 (Cluster)으로 나누고, 판별분석 (Discriminant Analysis)을 통하여 각 군집 별 특성을 파악하였다. 간단한 방법으로 영화의 이질성을 고려하였으나 수요 예측은 어려운 점이 있다.

영화 관람객 수의 확산 패턴과 관련하여, Ainslie et al. (2005)과 Sawhney and Eliashberg (1996)는 영화 개봉후 관객의 확산 패턴이 감마분포 (Gamma Distribution)를 따를 것이라고 가정하였다. Sawhney and Eliashberg (1996)는 누적 관객 수가 이항분포 (Binomial Distribution)를 따른다고 가정한 후, 각 영화에 대해 확산패턴을 추정하였으나, 영화와 관련된 변수나 동시에 상영된 경쟁 영화에 의한 효과 등을 고려하지 못했다. 반면, Ainslie et al. (2005)은 누적 관객수를 단순히 이항분포를 따른다고 가정하지 않고 예상 시장 점유율을 종속변수로 하는 선택 모형 (Choice Model)의 추정 모수 형태로 재모수화 (Reparameterize) 시켰다. 또한, 영화 외의 다른 제품 선택의 경우를 위해 외부재로 (Outside Good) 인한 가상 효용을 모형화하여 흥행에 중요한 요소가 무엇인지를 알아내고자 하였다. 영화의 이질성이나 Elberse and Eliashberg (2003)와 같이 변수의 내생성 문제를 해결하지는 못했지만, 영화 배급사 이름과 같이 수많은 명목 변수 (Categorical Variable)들을 더미 변수 (Dummy Variable)화 하지 않고, 각 배급사의 효과를 한꺼번에 추정할 수 있는 접근법을 사용한 것은 좋은방법으로 보인다 (Steenburgh et al. 2003). 한편, Vany and Walls (2004)는 영화와 같은 경험재의 수익은 평균에서 높은 정점을 갖고 꼬리가 두터운 분포 형태를 지닌 안정적 파레시안

(Stable Paretian) 분포가 잘 맞음을 보였다. 하지만, 영화 수익에 배우만 영향을 미친다고 가정한 것은 무리가 따르며, 모든 영화의 수익을 합쳐서 봤더니 무한 분산을 갖기 때문에 "Nobody knows"의 결과가 나왔다고 하는 것은 개별 영화의 수익성이나 수요 예측에는 한계점을 지닌다고 할 수 있다. 위의 연구들과 달리 Eliashberg et al. (2000)은 관객수나 수익이 일정 패턴을 지닌다고 가정하지는 않았으나 구전효과 (Word-of-mouth)에 의해 관객수가 마코프 체인 (Markov Chain)을 따른다고 가정하여, Sawhney and Eliashberg (1996)의 결과보다 좋음을 보였다. 그러나 이 연구는 구전효과와 함께 확산에 큰 영향을 미치는 경쟁효과나 시즌효과 (Seasonality)를 고려하지 못했다.

그러나 기존 연구의 대부분은 영화 흥행에 미치는 결정적 요인을 파악하는데 그 목적이 있었기 때문에 신규 영화에 대한 수요 예측과 관련된 연구는 많지 않았다. 이는 서론에서도 기술한 바와 같이 개인의 선호도에 따라 다양한 양상을 보이는 영화 관람 패턴을 예측하기란 쉽지 않기 때문이다. 이 주제와 관련된 가장 최근의 연구로 Sharda and Delen (2006)의 인공신경망(Artificial Neural Network) 모형이 있다. 영화 흥행성적에 따라 9 단계로 이산화 시킨 후 두 개의 숨은 층(Hidden layer)과 9개의 목표(Target) 노드를 갖는 다층 퍼셉트론(Multi-layer Perceptron)을 이용하여 로지스틱 회귀분석이나 관별분석, 의사결정나무(Decision Tree) 등의 기법보다 예측력이 우수함을 보였다. 인공신경망 기법은 데이터의 비선형성이 높을 때에도 우수한 예측력을 보이는 것은 사실이나 회귀분석처럼 변수 간 인과관계를 설명하기 어렵다거나 국부 최대화(Local Maxima) 문제나 때때로 과도추정으로 인한 예측력의 저하가 발생하는 문제점이 존재한다. 영화의 경우 흥행성적만 고려하면 한쪽으로 매우 치우친 분포를 보이는 것이 일반적이다. 즉, 흥행에 성공하는 영화는 극히 일부에 지나지 않고 대부분의 영화는 흥행성적이 저조하기 때문에 많은 개체가 속한 그룹을 과도 추정할 경우, 상대적으로 개체수가 적은 그룹에 대한 예측력은 떨어질 수 있다. 따라서 본 연구에서는 Sharda and Delen (2006)의 제시한 방법처럼 영화 수입을 이산화 하되, 베이지안 선택 모형을 사용하여 각 그룹간의 이질성을 반영을 통한 인공신경망 기법에서 나타나는 문제점을 해결하고자 하였다.

3. 베이지안 선택 모형

베이지안 선택 모형은 마케팅 분야에서 고객의 선택 문제를 다루면서 많이 쓰이는 모형으로 이 모형이 갖는 장점으로 고객의 선호도에 대한 이질성을 고려할 수 있다는 점과 추정 모수를 확률분포로 나타냄으로써 회귀분석의 오차와 같은 불확실성을 줄일 수 있다는 점이다(Andrews et al. 2002, Solgaard and Hansen 2003, Suarez et al. 2004).

선택 모형은 효용을 최대화(Utility Maximization)시키는 대안을 선택한다는 경제이론에서부터 출발한다. 임의의 소비자 i 가 대안 j 에 할당된 랜덤 효용은 다음 식과 같다.

$$U_{ij} = V_{ij} + e_{ij} = X_{ij}\beta_j + e_{ij} \quad (1)$$

여기서, V_{ij} 는 측정 가능한 변수 X 에 의한 확정적 효용(Deterministic)을 말하며, e_{ij} 는 측정 불가능한 랜덤 오차를 말한다. 이제 소비자 i 가 대안 j 를 선택할 확률은 다음과 같다.

$$p_{ij} = Pr(U_{ij} > U_{ik}) \quad (2)$$

식 (2)에서 k 는 모든 선택 대안 J 중에서 j 를 제외한 나머지 대안들을 의미한다. 즉, 다른 대안들보다 효용이 높다면 그 대안을 선택하게 되는 것이다. 이 선택 확률은 로짓 모형이나 프로빗 모형을 통해 쉽게 추정할 수 있다. 본 연구에서는 식 (3)과 같이 다항 로짓 모형을 사용하여 선택 확률을 추정하였다.

$$P_i(j | \beta_j) = \frac{\exp(X_{ij}\beta_j)}{\sum_k \exp(X_{ij}\beta_k)} \quad (3)$$

Sharda and Delen (2006)의 연구에서처럼 흥행 성적을 이산화하여 그룹핑할 경우, 이는 순서(Rank)를 갖는 이산변수가 된다. 따라서, 본 연구에서는 순서형 다항 로짓 모형을 사용하여 모수를 추정하였다. 또한, 추정모수 베타의 분포를 모호사전분포로 가정함으로써 베타의 불확실성을 최대한 반영하고자 하였다. 인공신경망은 상용 패키지인 NeuroShell을 이용하였으며, 베이지안 선택 모형은 WinBugs 소프트웨어를 이용하여 추정하였다.

4. 실증분석

4.1 데이터 수집 및 변수선정

본 연구에서 사용된 데이터는 2005년 1월부터 2006년 3월까지 국적을 불문하고 한국 극장에서 상영된 영화 170편을 대상으로 하였으며, 개봉 첫 주의 스크린 수, 개봉 첫 2주의 총 매출액 대비 비율, 누적 관객 수에 대한 자료는 영화진흥위원회에서 제공하는 입장권 통합 전산망 자료를 이용하였다. 그리고 각 영화의 감독, 배우, 장르, 제작사 및 배급사, 영화 평점 등의 상세정보는 엠파스(www.empas.com), 네이버(www.naver.com), 다음(www.daum.net) 등의 포털 사이트와 씨네21(www.cine21.co.kr), 필름2.0(www.film20.co.kr), 무비위크(http://www.movieweek.co.kr/) 등의 영화 전문 잡지를 통해 각 영화 리뷰 정보를 수집하였다. 본 연구에서 사용된 독립변수는 다음과 같다.

i) 영화 속성 관련 변수

- 국적: 더미 변수(한국:1, 외화:0)
- 감독, 배우: 이산 변수(상: 직전 작품이 300만명 이상의 흥행성적을 올린 영화, 중: 100만명 이상, 하:100만명 이하)
- 장르: 더미 변수(드라마, 코메디, 액션, 호러/스릴러, 멜로/로맨스, 애니메이션)

ii) 구전효과 관련 변수

- 잡지 리뷰 수: 3대 영화 잡지에 실린 리뷰 수의 자연로그 값
- 영화 평점: 포털 사이트에 일반인이 영화 기대치에 대해 평가한 점수의 평균(10점)

- 만점)의 자연로그 값
 ○ 평가자 수: 평점을 매긴 사람의 수

iii) 경쟁효과 관련 변수

- 스크린 수: 개봉 첫 주 스크린 수의 자연로그 값
- 계절성: 전년도 총 관객 수 대비 각 월별 관객 수의 비율
- 경쟁: 한 영화의 개봉 당시 전 주에 흥행 순위 상위 5위안에 든 영화들의 현재 관객 점유율
- 배급사: 전년도의 각 배급사별 시장 점유율

한편, 종속변수는 다음 표 1과 같은 기준으로 5개의 그룹으로 나누어 각 영화들을 할당하였다.

표 1. 흥행 성적에 따른 그룹화

그룹	1	2	3	4	5
설명	흥행에 실패한 영화	기대에 못미친 영화	손익분기점은 넘긴 영화	장르별 흥행 순위는 상위에 속하는 영화	흥행에 성공한 영화
기준 영화 수	50만이하	100만이하	300만이하	500만이하	500만이상
영화 수	93	36	28	9	4

4.2 결과

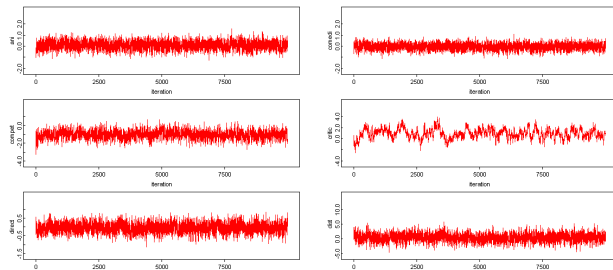


그림 1. 추정 모수들이 시계열 분포

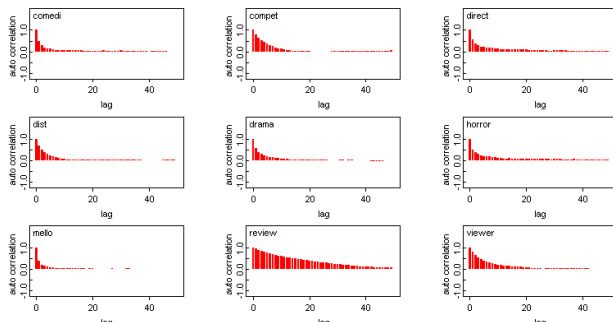


그림 2. 추정 모수들이 자동 상관도

모수들의 사후분포 추정에 앞서 10000번의 반복계산을 통해(Burning period) 그림 1과 같이 모수들이 정상분포를 보임으로써 수렴성을 확인하였고, 그림 2에서 보여 지는 것처럼 자기 상관도 또한 낮기 때문에 사후분포에서 모수 값들을 샘플링하는데 문제가 없음을 확인하였다. 그런 다음 또다시 10000번의 반복계산을 통해, 표 2와 3의 결과를 얻었다.

각 그룹 간 이질성을 고려하지 않고 전체 수준에서 변수의 유의성을 검증한 결과는 표 2와 같다. 국적, 개봉 주의 경쟁 정도, 개봉 첫 주의 스크린 수 및 영화 개봉 전 사전 평가자 수 등의 변수가 통계적으로 유의하게 나타났으며, 기대하는 부호

표 2. 모수의 사후분포평균 및 95% HPD

Variable	Mean	Std	Val2.5pc	Median	Val97.5pc
aatioa	0.2987	0.2748	0.2367	0.3007	0.8348
actor	0.274	0.1734	-0.06287	0.2747	0.617
compet	-1.008	0.4743	-1.941	-1.012	-0.04672
critic	1.052	0.8414	-0.6307	1.081	2.666
direct	-0.04242	0.2701	-0.5811	-0.03993	0.4695
dist	0.3517	1.325	-2.282	0.3687	2.923
review	0.0266	0.1837	-0.3446	0.02487	0.3811
screea	1.507	0.3003	0.9792	1.507	2.129
season	0.1426	1.405	-2.671	0.1536	2.937
viewer	0.6398	0.1352	0.3791	0.638	0.9123
drama	-0.3041	0.2792	-0.8609	-0.2981	0.2274
comedi	-0.02757	0.2732	-0.5586	-0.0276	0.5064
actioa	-0.2678	0.2682	-0.7966	-0.2681	-0.1626
horror	-0.5824	0.2971	-1.165	-0.5836	7.41E-04
mello	-0.1765	0.293	-0.7457	-0.1786	0.3909
ani	0.1089	0.3659	-0.6209	0.1099	0.8045

표 3. 그룹 이질성을 고려한 사후분포 추정 결과

Variable	Mean	Std	Val2.5pc	Median	Val97.5pc
compet[2]	-0.240	0.127	-0.478	-0.238	0.011
compet[3]	-0.259	0.150	-0.558	-0.254	0.022
compet[4]	-1.671	0.488	-2.659	-1.651	-0.745
compet[5]	-1.553	0.733	-3.050	-1.533	-0.184
screen[2]	0.307	0.461	-0.574	0.310	1.254
screen[3]	0.816	0.503	-0.181	0.804	1.956
screen[4]	1.068	1.055	0.629	1.079	2.103
screen[5]	4.497	1.304	4.221	4.635	5.131
actor[4]	0.387	0.186	0.093	0.366	0.799
season[3]	0.132	0.043	0.052	0.128	0.179
critic[3]	0.5375	0.2406	0.1115	0.5431	1.054

를 가짐을 알 수 있다. 특히, 개봉 첫 주의 스크린 수가 영화 흥행에 미치는 영향력이 큰 것으로 나타났다. 또한, 평가자 수 변수가 유의하게 나타남으로써 영화 흥행에 입소문 효과도 어느 정도 있는 것으로 나타났다. 그리고 특이한 점은 영화 흥행에 액션 장르는 오히려 부정적 영향을 미치는 것으로 나타났다. 이는 조사기간 동안 흥행에 성공한 영화들의 장르는 다양했으나 액션 장르는 적었기 때문이고, 영화 '태풍'과 같이 큰 기대치에 비해 흥행성고가 미미한 액션 영화들이 많았기 때문으로 생각된다.

그러나 표 2의 결과는 각 변수들이 영화 흥행에 미치는 영향력이 모든 그룹에 동일하다는 가정하에서 도출된 결과이기 때문에 상업적으로 매우 성공한 영화에 대한 정밀한 분석이 불가능하다. 따라서 본 연구에서는 그룹 간 이질성을 반영한 베이저안 선택 모형을 추가로 검증하였으며 결과는 표 3과 같다. 베이저안 모형의 타당성을 나타내는 DIC(Deviation Information Criterion, Spiegelhalter 2002)³⁾ 값 또한 전체 수준(DIC=286.3) 보다 이질성을 반영한 모형(DIC=264.0)이 더 작았다. 각 모수의 값은 흥행에 실패한 1 그룹과의 상대적인 차이를 나타내며 숫자는 각 그룹을 나타낸다. 그룹 간 이질성을 반영하지 않은 전체 수준에서 통계적으로 유의하게 나타난 스크린 수와 경쟁 정도를 나타내는 변수는 2 그룹과 3 그룹의 경우 1 그룹과 차이가 있는 것으로 나타났다. 즉, 관객 수가 300만이 안되는 영화는 스크린 수에 크게 영향을 받지 않음을 의미한다. 이는 반대로 상업적으로 성공한 영화를 만들고자 한다면 개봉 첫 주의 스크린 수와 기존 영화와의 경쟁 정도를 면밀히 주시해야 한다는 뜻으로도 볼 수 있다. 그리고 특이한 점은 전체수

표 4. 인공지능경망과 선택 모형간의 예측률 및 제곱합 오차 비교

Group	SET 1			SET 2			SET 3			SET 4			SET 5																	
	TRUE	ANN	MNP	TRUE	ANN	MNP	TRUE	ANN	MNP	TRUE	ANN	MNP	TRUE	ANN	MNP															
1	42	40	38	41	37	38	54	52	52	51	49	48	42	38	40															
2	20	18	18	16	14	11	18	16	12	15	11	10	16	15	14															
3	14	8	7	10	6	7	11	8	9	15	10	11	12	9	9															
4	4	2	2	5	3	3	2	1	1	6	3	4	3	1	2															
5	2	0	1	1	1	0	2	0	2	1	0	1	1	1	1															
Total	82	68	66	73	61	59	87	77	76	88	73	74	74	64	66															
Hit Ratio(%)	82.9			80.5			83.6			80.8			88.5			87.4			83.0			84.1			86.5			89.2		
Mean	ANN: 84.9%						MNP: 84.4%																							
SSE	25.3			22.7			12.8			14.9			16.0			11.3			12.7			10.1			8.5			6.4		
Mean	ANN: 15.1						MNP: 13.3																							

준에서는 통계적으로 유의하지 않으나 그룹 간 차이를 보면 통계적으로 유의한 변수가 있다는 점이다. 표 3에서 보여 지는 바와 같이 4 그룹의 배우와 3 그룹의 계절성 및 영화 평점의 변수가 1 그룹과 유의한 차이가 있는 것으로 나타났다. 배우의 경우 영화 흥행과 완전히 무관한 것이 아니라 어느 정도 상업적으로 성공한 영화라고 지칭될 수 있는 영화에는 배우의 힘이 작용하고 있음을 알 수 있다. 또한, 영화 흥행에는 결정적 요인이 되지 못하나 백만 관객을 동원하는 데는 계절성이나 영화 평점이 유의하게 나타났다. 따라서 영화 제작사나 배급사 입장에서는 이들 그룹 간의 관련성을 바탕으로 영화 개봉 시기 및 배우 등에 대해서 전략적 선택을 할 수 있다.

다음으로 베이지안 선택 모형을 이용하여 신규 영화에 대한 수요예측에 어떻게 활용될 수 있는지를 보고자 한다. 이를 위해 모형을 만드는 계산 표본(Calibration sample)과 예측 정확도를 검증하는 표본(Holdout sample)을 랜덤 추출하여 5개의 데이터 셋을 만든 후 예측률 및 오차4)를 계산하여 인공지능경망 모형과 비교하였다. 결과는 표 4에 나타나 있는 바와 같이, 인공지능경망과 선택 모형 모두 예측률이 우수한 것으로 나타났다. 5개 데이터 셋의 전체 평균을 구한 결과 인공지능경망이 모형이 선택 모형보다 근소한 차로 예측력이 우수한 것으로 나타났으나, 흥행에 성공한 영화(4, 5 그룹)를 예측하는 데에는 베이지안 선택모형이 앞서는 것으로 나타났다. 이는 실제 그룹을 정확히 맞지지는 못하더라도 얼마나 근접하게 예측했는지를 알아보는 제곱합 오차의 계산 결과에서도 확인할 수 있다. 만약 실제로 5 그룹인 영화를 4 그룹인 영화가 아니라 2 그룹이나 1 그룹으로 분류했다면 예측 오차는 그룹 간 거리의 제곱에 비례해 커진다. 따라서 전체적인 예측률은 인공지능경망이 앞서나 예측의 근접성까지 고려하면 베이지안 선택 모형이 낫다고 말할 수 있다. 실제로 영화 흥행을 예측한다는 것은 1, 2 그룹보다는 4, 5 그룹에 속할 가능성을 알아보는 것이기 때문에 이들 그룹에 대한 예측 정확도 및 예측 근접성이 중요하다고 볼 수 있다.

5. 결론

본 연구에서는 베이지안 선택 모형을 이용하여 영화 흥행을 결정하는 요인 탐색 및 영화 흥행성과를 예측해 보았다. 흥행하는 영화일수록 인공지능경망 모형과 비교해 예측력이 더 우수하게 나타났으며, 인공지능경망 모형에서는 알 수 없는 각 그룹별 변수의 유의성 또한 명확하게 파악할 수 있었다. 또한, 그룹 이질성을 반영한 모형을 통해 전체 수준에서는 통계적으로 유의하지 않은 변수라도 각 그룹 수준에서는 통계적으로 유의한 차이가 있는 그룹이 있음을 알 수 있다. 따라서 본 연구에서 제시한 모형을 활용하면 영화 제작자나 배급사, 극장주는 영화 개봉에 대한 전략적인 선택을 할 수 있으리라 기대된다.

하지만, 본 연구에서는 각 그룹별 이질성만 고려했을 뿐, 영화 간 이질성은 고려하지 못한 단점이 있다. 향후 이에 대한 확장이 필요하며, 요즘 영화 흥행에 중요한 요소로 부각되고 있는 마케팅에 관한 효과를 반영할 수 있도록 각 영화의 마케팅 비용 등의 변수 추가 및 마케팅 효과와 구전효과간의 상관관계를 고려한 모형의 수립이 필요하다고 하겠다.

참고문헌

- Ainslie, A., X. Dreze, and F. Zufryden (2005), Modeling Movie Life Cycles and Market Share, *Marketing Science*, 24(3), 508-517
- Allbert, S., (1998), Movie Stars and the Distribution of Financially Successful Films in the Motion Picture Industry, *Journal of Cultural Economics*, 22, 249-270
- Andrews, R. L., A. Ainslie, and I. S. Currim, (200), An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity, *Journal of Marketing Research*, 34, 479-487
- Bagella, M. and L. Becchetti, (1999), The Determinants of Motion Picture Box Office Performance: Evidence from Movies Produced in Italy, *Journal of Cultural Economics*, 23, 237-256
- De Vany, A. S. and W. D. Walls (2004), Motion Picture Profit, the Stable Paretian Hypothesis, and the Curse of the Superstar, *Journal of Economic Dynamics & Control*, 28, 1035-1057
- _____ (1999), Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office?, *Journal of Cultural Economics*, 23, 285-318
- Elberse, A. and J. Eliashberg (2003), Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures, *Marketing Science*, 22(3), 329-354
- Eliashberg, J., J.-J. Jonker, M. S. Sawhney, and B. Wierenga (2000), MOVIEMOD: An Implementable Decision-Support System for Prerelease Market Evaluation of Motion Pictures, *Marketing Science*, 19(3), 226-243
- _____ and S. M. Shugan, (1997), Film Critics: Influencers or Predictors?, *Journal of Marketing*, 16(April), 68-78
- Jedidi, K., R. E. Krider, and C. B. Weinberg (1998), Clustering at the Movies, *Marketing Letters*, 9(4), 393-405
- Levin, A. M., I. P. Levin, and C. E. Heath, (1997), Movie Stars and Authors as Brand Names: Measuring Brand Equity in Experiential Products, *Advances in Consumer Research*, 24, 175-181
- Mahajan, V., E. Muller, and R. A. Kerin, (1984), Introduction Strategy For New Products With Positive And Negative Word-of-Mouth, *Management Science*, 30(12), 1389-1404
- Ravid, S. A., (1999), Information, Blockbusters, and Stars: A Study of the Film Industry, *Journal of Business*, 72(4), 463-492
- Sawhney, M. S. and J. Eliashberg (1996), A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures, *Marketing Science*, 15(2), 113-131
- Sharda, R. and D. Delen, (2006), Predicting Box-Office Success of Motion Pictures with Neural Networks, Expert Systems with Applications, 30, 243-254
- Solgaard, H. S. and T. Hansen, (2003), A Hierarchical Bayes Model of Choice Between Supermarket Formats, *Journal of Retailing and Consumer Services*, 10, 169-180
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde, (2002), Bayesian Measures of Model Complexity and Fit (with discussion), *Journal of Royal Statistics*, 64, 583-640
- Suarez, A., I. R. del Bosque, J. M. Rodriguez-Poo, and I. Moral, (2004), Accounting for Heterogeneity in Shopping Centre Choice Models, *Journal of Retailing and Consumer Services*, 11, 119-129
- Zufryden, F., (2000), New Film Website Promotion and Box-Office Performance, *Journal of Advertising Research*, January/April, 55-64
- 김휴중 (1997), 한국 영화스타의 스타파워 분석, 삼성경제연구소
- 양지훈 (1998), 엔터테인먼트 상품의 품질인지과정에 있어서의 매니아 요소의 역할에 관한 연구, 서울대학교 대학원 석사학위논문
- 유현석 (2002), 한국영화의 흥행요인에 관한 연구: 제작 관련 변수를 중심으로, 한국언론학보, 46(3), 183-213
- 박형현, 박찬수, (2001), 영화 평론과 흥행 성과 간의 관계_ 인터넷 시대에도 유효한가, 마케팅 연구, 16(4), 71-85
- 손원경, (2005), 한국 영화 산업의 효율적 마케팅을 위한 비용 모델링에 관한 연구, 서울대학교 대학원 석사학위논문
- 이소민, (2005), 영화평이 영화 흥행에 미치는 영향에 관한 연구, 서울대학교 대학원 석사학위논문

1) 한 영화는 여러 장르에 속할 수 있으므로 더미 변수의 합이 1이 되지 않는다. 예) 웰컴 투 동막골: 드라마, 액션

2) 베이저안 추정에서 모수의 통계적 유의성은 사후분포의 평균을 포함하는 95% HPD(High Probability Distribution) 구간 안에 0을 포함하고 있는지의 여부로 결정한다. 95% HPD 구간이 0을 포함하고 있다는 것은 모수의 값이 0 즉, 무의미한 값을 가질 수 있기 때문이다. 표에서 굵게 표시된 변수가 통계적으로 유의한 변수이다.

3) $DIC = Dhat(-2\log(p(y|\bar{\theta}))) + 2 * pD(\text{effective number of parameters})$
따라서, 작은 값을 가질수록 모형의 타당도가 높아진다.

$$4) SSE = \sum_{i=1}^N (True - Predict)^2$$