

초기 시청시간 패턴 분석을 통한 대흥행 드라마 예측

남기환

한국과학기술원 경영대학 경영공학부
(namkh@kaist.ac.kr)

성노윤

한국과학기술원 경영대학 경영공학부
(nyseong@kaist.ac.kr)

TV 드라마는 타 장르에 비해 시청률과 채널 홍보 효과가 매우 크며, 한류를 통해 산업적 효과와 문화적 영향력을 확인시켜줬다. 따라서, 이와 같은 드라마의 흥행 여부를 예측하는 일은 방송 관련 산업에서 매우 중요한 부분임은 주지의 사실이다. 이를 위해서 본 연구에서는 2003년부터 2012년까지 10년간, 지상파 채널을 통해 방송된, 총 280개의 TV 미니시리즈 드라마를 분석하였다. 이들 드라마 중 평균 시청률 상위 45개, 하위 시청률 45개를 선정하여 흥행 드라마의 시청시간 분포 (5%~100%, 11-Step) 모형을 만들었다. 이들 기준 모형과 신규 드라마의 시청시간 분포와의 이격 거리를 Euclidean/Correlation으로 측정한 유사도(Similarity)를 통해, 시청자의 초기(1~5회) 시청시간 분포로 신규 드라마의 성패 여부를 예측하는 모형을 만들었다. 또한 총 방송 시간 중 70% 이상 시청한 시청자를 열혈 시청층(이하 열혈층) 으로 분류하고, 상위/하위 드라마의 평균값과 비교하여, 신규 드라마의 흥행여부를 판별할 수 있도록 설계하였다. 연구 결과 드라마의 초반 시청자 충성도(시청시간)는 드라마의 대흥행 여부를 예측하는데 중요한 요소임을 밝혔으며, 최대 75.47%의 확률로 대흥행 드라마의 탄생을 예측할 수 있었다.

주제어 : 유사도, 시청시간 분포, 대흥행 드라마 예측, 충성도, 최근접이웃

논문접수일 : 2018년 5월 20일 논문수정일 : 2018년 11월 6일 게재확정일 : 2018년 11월 26일

원고유형 : 학술대회(급행) 교신저자 : 성노윤

1. 서론

TV 드라마는 다른 방송 장르보다 시청률이 높고, 채널의 홍보 효과가 크며, 뉴미디어 창구를 통한 수익 효과가 높다 (Wittebols, 2004). 대한민국은 2012년 총 1,580억 원의 방송 프로그램을 수출하였으며, 전체 매출의 84.2%에 해당하는 약 1,449억 원을 드라마 수출이 담당하고 있어, TV 매체 중에서 드라마가 가장 높은 매출을 차지하고 있다 (Jeong and Shin, 2013). 또한, 대한민국 드라마의 해외 진출로 한류가 확산되며, 드

라마는 단순한 상품으로서의 가치를 지닐 뿐만 아니라, 한국 문화에 대한 관심과 관광 등 스펀 오버 효과 (spillover effect)를 일으키고 있다 (Ryoo, 2009). 이처럼 드라마가 방송사의 중요한 수입원이 됨에 따라, 드라마를 성공적으로 만드는 것은 중요하며, 방송사는 드라마를 성행시키기 위해 막대한 인적/물적 자원을 투자하고 있다.

TV 프로그램의 성과를 측정하는 방법은 다양한 방식이 존재한다. 프로그램에 대한 시청자 및 전문가의 비평을 의미하는 질적 성과, 흔히 시청률로 평가되는 수용자의 접근 정도를 나타내는

수용자 성과, 그리고 비용 및 비용 대비 수익과 같은 효율성으로 표현되는 경제적 성과의 세 차원으로 구분하는 방법이 있다 (Tebert, 2000). 또한, 질적 성과, 지불 가치, 영향력 등을 계량적으로 측정하는 방식도 있다 (Corporation, 2004). 하지만, 일반적으로는 수용자 성과인 시청률이 드라마의 주된 성과지표로 인식되고 있다 (Ahn et al., 2017).

따라서 시청률은 방송사뿐만 아니라, 광고 산업과 관련된 다양한 이해관계자들 모두에게 중요한 관심사이며, 이를 예측하는 것은 방송 업계에서는 항상 중요한 문제이다. 하지만 이렇게 시청률 예측과 관련한 연구의 중요성이 날로 커짐에도 불구하고 국내의 시청률 예측 관련 연구는 매우 미비하다 (Lee et al., 2012). 특히 기존의 연구들은 연출자, 작가, 출연자 등 주로 인적 요소인 프로그램 속성 변수와 방송 채널, 전 회차 시청률 등 외형적 채널 요소에 대부분 치중하고 있으며, 단순히 이들 변수를 나열한 회귀식으로서 시청률 예측을 시도하여왔다 (Bae, 2005; Kwon et al., 2009; Lee et al., 2012; Yu and Kim, 2010). 또한, 기존 연구의 대부분은 방송 프로그램이 방영되기 전인 사전 예측을 중심으로 연구되어 왔으나 사전 예측의 경우 수없이 다양한 변수 존재하고, 돌발적 변수들을 예측하기 어려워 그간의 논문 결과가 적절한 예측력을 보이지 못한다는 점에서 한계를 가지고 있었다. 하지만, 한국 방송 시장에서는 드라마를 제작할 때, 사전에 전 회차를 촬영하지 않고, 방송 일주일 전에 최종 작업을 거친다 (Ahn et al., 2017). 이와 같은 방송 산업 구조에서는, 첫 방송 방영 전 사전 예측의 의미가 퇴색된다. 따라서, 본 논문에서는 사전 예측의 한계점을 해결하기 위해 사후 예측으로 예측 방법을 전환하였다.

또한, 기존 연구들은 장르와 배우, 연출자 등의 영향을 보이는 것이 일반적이었다. 하지만, 빅데이터 시대에 도입하면서 이용할 수 있는 데이터가 늘어나고 이를 뒷받침하는 방법론들이 등장하면서, 소비자의 행동 경향성을 이용한 시청률 예측 문제가 연구되고 있지만, 아직 연구의 진행은 미비한 상황이다 (Ahn et al., 2017). 특히, TV 프로그램은 특정 소비자층 (열혈 소비자층)의 영향이 크고 이들을 분석하는 것이 중요하다 (Sabavala and Morrison, 1977). 하지만, 현재까지는 빅데이터에 기반한 TV 프로그램에 대한 소비자 행태와 충성도를 정량화하고 이에 대한 분포를 측정하는 연구는 없었다. 본 논문에서는 TV 프로그램의 충성도를 재정의하고 이것의 분포를 정량화하여 측정하였다. 즉, 연구 문제는 시청시간 분포 경향성 분석을 통해 대흥행 드라마 예측 여부이다.

충성도의 분포가 대흥행 드라마와 비흥행 드라마에서 다르기 때문에, 새로운 프로그램이 나왔을 때, 충성도의 분포를 확인하고 기존에 어떤 분포와 유사한지를 확인할 수 있다. 이를 통해 대흥행 드라마인지 비흥행 드라마인지 예측할 수 있다. 분포의 유사성을 확인할 때는 일반적으로 기계 학습의 최근접 이웃 알고리즘이 사용되기 때문에, 본 논문에서는 최근접 이웃 알고리즘을 통해 충성도의 분포를 예측하여 대흥행 여부를 확인하였다. 대흥행 여부는 평균 시청률 20%를 기준으로 하였다.

본 논문의 결과는 다음과 같다. 먼저, 본 논문에서는 사전 조사를 통해 비흥행 드라마와 대흥행 드라마의 충성도 분포가 다름을 확인할 수 있었다. 또한, 2003년부터 2010년까지 드라마를 훈련 데이터로, 2011년부터 2012년까지 드라마를 예측 데이터로 하여 최근접이웃 알고리즘으로

예측을 하였을 때, 약 75%의 예측 정확도를 보이며, 대흥행 드라마의 충성도 분포와 비흥행 드라마의 충성도 분포가 유의미하게 다르며, 이를 이용하여 흥행 여부를 예측하는 것은 유의미하다는 것을 밝혀냈다.

본 연구의 기여는 다음과 같다. 첫 번째, 시청률 예측은 방송사와 광고 대행사 등 많은 이해관계자에게 매우 중요한 문제임에도 불구하고 연구가 많이 되지 않았는데 이 연구의 간격을 본 논문에서 메운다. 두 번째, 기존 연구에서 사용되던 시청률 사전 예측의 한계점을 해결하고자 사후 예측으로 전환하여 뛰어난 예측력을 보여주며, 한국 시장에 특화된 예측 모델의 개발 필요성을 입증하였다. 세 번째, 기존 연구들은 시청률을 예측할 때, 드라마의 특성을 가지고 예측을 하였지만, 본 논문은 단순 특성이 아닌 소비자의 행태를 살펴본다 열혈 시청자층이 있다는 것을 밝혀낸 첫 번째 논문이며, 이를 이용해 시청률을 예측한 첫 시도를 하며, 본 논문은 시청률 예측에서도 단순히 드라마 특성을 가지고 예측을 하는 것이 아닌, 행동 경제 이론에 기반한 예측이 가능하다는 것을 입증하였다.

본 논문은 2장에서 문헌 연구를 소개하며, 3장에서 연구 방법을 소개하며, 그 결과를 4장에서 설명한다. 5장에서는 본 연구의 의의에 관해 설명하며, 마지막으로 6장에서 본 논문을 결론지으며, 한계점과 후속 연구에 대해 토의한다.

2. 문헌 연구 및 가설 설립

2.1 기존 연구 동향

앞서 언급한 바와 같이, 기존 시청률 예측과

관련된 연구는 구체적인 예측 모형과 실증적인 검증 결과를 명확히 제시하지 못하였으며, 이에 따라 방송사 등 실제 현업에서 활용되었던 사례는 없었다. 시청률의 예측과 관련된 주요 연구는 다음과 같다.

Bae (2005)에서는 2000-2003년 동안 방송된 드라마를 대상으로 지상파 드라마의 시청률 성과에 미치는 요인을 드라마 자체의 속성 및 수용자 요인의 차원에서 살펴본 바 있다. 분석 결과, 외주제작 여부는 유의미한 관련성이 없으며, 스타출연, 채널, 드라마 유형에 따라 시청률 성과가 달라지는 것으로 판단하였다. 즉, 스타가 출연할 경우 시청률이 높아지고(남자 스타보다는 여자 스타가 더 높게 나타남), 채널 별로는 KBS1TV가 가장 높았고, 드라마 유형에서는 주말드라마(저녁)와 수목드라마의 성과가 높았으며, 초기 반복 시청률이 높을수록 시청자 구성이 동질적이지 않을수록 드라마 시청률이 높은 것으로 분석하였다.

Kim (2008)에서는 지상파 방송사에서 방영한 드라마 225개를 대상으로 하여 실증 분석한 결과, 장르 친숙성, 주연배우 지명도, 인터넷 기사 등은 드라마 시청률에 긍정적인 영향을 미치는 것으로 나타났으며, 장르 친숙성과 인터넷 기사, 주연배우 지명도와 인터넷 기사 등의 상호작용 효과를 분석한 결과는 주연배우 지명도와 인터넷 기사 상호 작용 항만 드라마 시청률에 유의한 영향이 있는 것으로 나타났다. 구체적으로, 주연배우 지명도가 드라마 시청률에 미치는 긍정적인 영향은 인터넷 기사 수가 많을수록 강화되는 것으로 나타난 것으로 분석하였다.

Yu and Kim (2010)에서는 외주제작 드라마를 대상으로 드라마의 시청률에 영향을 미치는 요인을 제작 요인과 인적 요인을 중심으로 분석하

였다. 2004년부터 2009년 9월까지 지상파 방송 3사에서 방영된 324편의 드라마 중 외주제작사가 제작에 참여한 드라마 204편을 대상으로 하였으며, 제작 요인은 드라마의 제작 형태(단독제작/공동제작), 제작사의 제작 경험, 제작사의 성격(매니지먼트 기능 수행 여부), 드라마의 제작 방영 편수, 인적 요인으로는 드라마의 주요 생산요소인 PD, 작가, 배우를 중심으로 이들 간의 연결성, 친밀성, 스타성을 살펴보았다. 이에 대한 분석 결과로는 방영 편수가 길수록, 제작사가 매니지먼트 기능을 수행할 경우, 인적 요인 간 연결성이 많을 경우, 스타 작가일 경우 시청률이 더 높은 것으로 나타났으며 이들 요인 중 드라마의 시청률에 가장 영향을 미치는 요인을 살펴보기 위해 회귀 분석한 결과 제작 방영 편수, 제작사가 매니지먼트의 기능을 보유한 경우, PD-작가의 친밀성이 시청률에 긍정적인 영향을 미치는 것으로 분석하였다.

Lee et al. (2012)에서는 편성되는 요일과 시간, 타 방송사프로그램의 경쟁력/프로그램 제작에 투입된 외부 지출 비용(제작비) 및 내부 자원 사용량/ 프로그램을 제작하는 과정에서 수행되는 프로그램 평가 및 제작에 참여하는 인력, 외주 제작 등의 조건/KBS 드라마 프로그램의 방영 전에 수주된 해당 프로그램에 대한 광고량 또는 금액 등 열거한 조건들이 시청률에 영향을 미칠 것이라는 가설하에 회귀분석을 실시한 결과, 방영 요일의 토요일 여부, 23시 시간대 여부, 타사시청률 합계(경쟁 강도), 방영 당시 시청인구 비율, 광고 판매금액, 프로그램 완성도 점수, 편집실 사용 시간 등 최종 7개 항목으로 예측방정식을 완성하였다.

Kwon et al. (2009)에서는 국내 드라마의 성공 방정식을 추출하지 못하였으나, 국내 드라마의

성공 요인을 다양한 방법으로 접근하면서 의미 있는 결과를 도출하였다. 드라마의 핵심적인 성공 요인은 작품의 완성도이며, 작품의 완성도는 매력적인 등장인물의 창조와 그 등장인물의 개성과 참신하고 탄탄한 이야기 구조가 함께 어우러지는 것을 의미하지만, 드라마의 완성도, 매력적이거나 참신한 등장인물의 창조, 탄탄한 이야기 구조 등을 계량적으로 분석할 수 없었기 때문에 이 요소가 가장 중요하다는 것을 실증적으로 보여줄 수 없다고 설명하였다. 또한 연구의 분석과정에서 얻어진 의미 있는 결과로 드라마의 내용을 구성하는 개별 요소가 드라마의 시청률에 미치는 영향은 상당히 제한적이며 드라마의 시청률에는 편성시간, 드라마 형태(일일연속극, 미니시리즈 등), 방영 횟수, 방영 채널과 같은 편성요인의 영향을 많이 받는 것으로 기술하였다. 특이한 점은 기존 연구 결과와는 다르게, 드라마의 시청률은 스타의 출연 여부에 영향을 받지 않는 것으로 분석하였으며, 시청률 상승세에 기여하는 주요 요인은 편성 대진, 도입부에서의 시청층 유입 정도, 주인공의 시련이나 위기 배치와 정도 등이고, 저조한 시청률을 보인 드라마의 경우는 낮은 소재와 형식을 이용한 경우, 극적 긴장감을 유발하는 갈등이 부재한 경우, 그리고 드라마의 중심인물 혹은 주인공을 확실히 자리매김하지 못한 경우도 많은 것으로 보고하였다.

이처럼 기존 연구들은 성공적인 드라마를 정의하고 그 요소들을 찾는 노력을 주로 하였다. 하지만, 기존 연구에는 소비자의 특정 행동(충성도)을 확인하려는 시도는 없었으며, 내적, 외적 요소를 모두 고려하는 것이 아닌 외적 요소만을 고려하였다는 한계를 가지기에 본 논문에서는 기존 연구들에서 확인하지 못한 내적 요소, 소비자의 행태를 통한 성공적인 드라마 예측 문

제를 해결한다.

2.2 충성도

우리가 아는 한, 대흥행 드라마와 시청 시간과의 관계에 대해서는 아직 미디어 업계에서 구체적으로 언급한 사례를 찾지 못하였다. 본 논문에서는 방송업과 유사한 신문업에 비추어 현황을 바라보며, 이를 경제학에서 주로 사용되는 충성도 개념으로 해석하여 가설을 도출한다.

먼저, 시청률의 개념과 유사한 구독률과 시청 시간으로 대응시킬 수 있는 열독률과의 관계에 대비시켜 본다. (Park and Jeon, 2006)에서는 독자의 기사 평가는 일차적으로 열독으로 나타나며, 열독이 절대기준은 아니지만, 기사 평가가 전제 조건이며 신문의 만족도 향상과 독자 증가의 선행조건임을 부인하기 어렵다고 언급하여, 열독률과 구독률과의 연관성을 언급한 바 있다. 이는 결국 특정 신문에 대한 고객의 만족도 향상이 열독으로 이어지고, 최종 구독률 향상으로 귀결되어 열독률과 구독률과는 양의 상관관계가 성립됨을 알 수 있다.

이를 경제학적 관점에서 해석해 보면 시청 시간과 열독률은 그 매체에 대한 충성도를 의미하며, 시청률과 구독률은 높은 충성도에 따른 재사용 여부를 의미한다 (Sabavala and Morrison, 1977). 따라서, 본 논문에서는 시청률 및 시청 시간을 충성도의 관점에서 해석한다.

고객 충성도는 특정 브랜드 혹은 기업을 꾸준히 이용하거나 재구매를 하는 등 기여를 많이하는 것을 의미한다. 충성도가 높은 고객의 대표적인 행위로는 상품의 지속적 구매와 우호적 구전 효과 등이 있다 (Dick and Basu, 1994).

일반적으로 충성도는 사용률이 얼마나 높은지

에 따라 결정된다. 자연계의 분포는 일반적으로 상위 20%의 사용자가 80%의 사용을 담당한다는 파레토의 법칙에 따라 (Pareto and Page, 1971), 충성도도 특정한 분포를 가진다고 가정하여 일각에서는 충성도가 높은 사람은 상위 20%의 사용률을 가진 사람을 정의한다 (Reinartz and Kumar, 2003). 또한, 충성도는 브랜드에 얼마나 기여하였느냐로 정의가 되며, 이 또한 특정한 분포를 가진다고 가정하여 4단계로 나누어 정의한다 (Kotler et al., 2016). 즉, 충성도에 대한 조작적 정의는 분야별로 상이함을 알 수 있으나, 충성도는 특정한 분포를 가진다 (Singh et al., 2008).

(Sabavala and Morrison, 1977)에 따르면, 시청 시간 또한 충성도로 베타 분포를 따른다. 저자들은 시청 시간 분포의 모양에 따라 충성도 지수를 선정하는 방법을 제시하고, 충성도가 높은 고객들과 충성도가 낮은 고객들이 다른 시청 시간 분포를 가진다고 가정하였으며, 상관관계 분석을 통해 방법의 타당성을 입증하였다. 즉, 시청 시간도 충성도의 한 종류로 해석을 할 수 있으며, 특정한 분포를 따른다고 할 수 있다. 따라서 우리는 다음과 같이 가정한다.

연구 가설 1: 시청률이 높은 드라마의 시청자 시청 시간(충성도) 분포는 특정한 패턴을 나타낼 것이다.

연구 가설 2: 시청률이 낮은 드라마의 시청자 시청 시간(충성도) 분포는 특정한 패턴을 나타낼 것이다.

3. 연구 방법

3.1 데이터

본 연구는 드라마의 충성도(시청시간)가 시청률에 미치는 영향을 분석하는 것을 목적으로 한다. 본 연구에서는 시청시간의 비중을 충성도로 정의한다. 본 연구의 목적달성을 위한 분석에 사용된 데이터는 세계 시장 조사 기관인 AGB닐슨의 시청률 조사/분석 프로그램인 아리아나(Ariana)를 통해서 추출한 국내 시청률 조사 자료이다. 본 데이터는 2003년부터 2012년까지 10년간 총 280개의 월/화, 수/목 국내 3개 공중파 방송사 3개 채널(KBS2, MBC, SBS) 미니시리즈에 대한 데이터로, 비교 대상의 동일성을 중시하기 위해 공중파 방송사 4개 채널(KBS1, KBS2, MBC, SBS)로 한정하였다. 총 280개의 원천 데이터 중에서 2003년부터 2010년의 드라마를 훈련 데이터(Training Data)로 227개, 2011년부터 2012년의 드라마를 예측 데이터(Test Data)로 53개를 사용하였다.

3.2 흥행 드라마 측정

먼저, 종속 변수는 시청률로 선정하였다. 시청률은 동적 데이터로 예측을 하기 위해서는 기준이 필요하다. 일반적으로 평균 시청률을 사용하지만, 동적 데이터는 평균만으로 그 의미를 파악하기 힘들기 때문에, 본 논문에서는 최고 시청률, 평균 시청률을 종속 변수로 설정하였다. 즉 모델의 건장성을 확인하기 위해, 단순히 평균 시청률에 대한 기준으로 흥행드라마를 예측 하는 것이 아닌, 최고 시청률 기준으로도 흥행 예측을 시도해 보고 평균 시청률로도 예측을 시행해 보았다.

대흥행 드라마란 기존 논문에서 대흥행 드라

마를 시청률 30% 이상인 드라마, 비 흥행 드라마를 시청률 20% 미만인 드라마로 정의하였으며(Kim, 2009), TV 드라마 시청률 요인분석에서는 스타 작가를 20% 이상 시청률을 경험한 작가로 정의한 바 있다(Kwon et al., 2009). 이렇듯 성공한 드라마의 정의가 업계와 학계, 시대의 변화, 방송환경의 변화에 따라 상이하므로 명확한 기준은 찾기 어렵다. 또한 최근 종합편성 채널의 방송시장 진입과 대기업 계열회사의 방송시장 투자로 그 어느 때 보다 시청률 경쟁이 치열하여, 과거의 시청률 잣대를 현재 시점에 적용하는 것은 어렵다고 판단하여, 본 논문에서는 10년간 상위 45개 프로그램의 평균시청률이 20%이상인 드라마를 대상으로 연구모형을 만들고, 평균 시청률이 20% 이상인 드라마로 대흥행 드라마로 기준을 낮추어 정의하고자 한다.

즉, 대흥행 드라마를 정의하기 위해 본 논문에서는 2003년부터 2012년까지 평균 시청률 상위 45개 프로그램의 기준인 시청률이 20% 이상인 드라마들을 대흥행 드라마로 구분하였다.

3.3 충성도 측정

시청자들은 드라마를 시청할 때 여러 행태를 보이게 된다. 본 연구에서는 드라마 충성도에 대한 값을 5%~100%의 값을 추출하였다. 5%에 대한 데이터가 의미하는 바는 전체 드라마 시간 중에 5%의 시간을 시청한 시청자의 누적 퍼센트를 말하고 100%는 드라마 시간의 전체를 시청한 시청자의 누적 퍼센트를 나타낸다. 충성도에 대한 데이터 형태는 아래 그림과 같다.

이렇게 추출된 충성도 데이터를 기준으로 누적 아닌 각 충성도에 따른 실제 값들을 계산을 통하여 각 충성도에 대한 값으로 하였다. 충성도

에 대한 값을 추출은 구하고자 하는 I 번째 누적 충성도 값에서 i-1 번째 누적 충성도 값을 빼서 구하였다. 이는 (1) 과 같이 정의된다.

$$R_i = \{RC_i - RC_{i-1} | RC_0 = 0, 1 \leq i \leq n \text{ where } n = \text{percentile of loyalty}\} \quad (1)$$

3.4 예측 방법

훈련 데이터에서 앞에서 정의한 내용을 기반으로 대흥행으로 분류된 드라마들을 추출하였다. 이 데이터들의 누적 시청자 충성도에 대해 계산된 시청자 충성도 값을 이용하여 각각 대흥행 전체 시청자 충성도 평균의 값을 계산한다. 여기에 앞에서 언급한 이론인 열혈도의 개념을 적용한다. 열혈도를 적용해 시청자 충성도이 70%이상의 비율이 80%를 넘는 드라마들에 대해

서만 대흥행을 구분 짓는 표준 시청자 충성도 분포 틀을 만드는데 적용한다. 이렇게 선정된 대흥행 드라마의 1회차, 5회차 누적, 전체의 시청자 충성도에 대한 분포를 확정한다. 이렇게 계산된 값은 아래 그림과 같다.

하지만, 본 논문이 충성도 기반 예측을 하는 첫 논문이기에 훈련 데이터와 예측 데이터의 구성을 이론적으로 참조할 수 없기 때문에, 데이터 마이닝 분야에서 효율적인 결과를 토출하는 방법 중의 하나인 그리드 서치 방식 (Grid search) 을 사용하여 다양한 구성에 대해서 실험을 하여 의미 있는 결과를 도출한다. 훈련 데이터는 분포를 1회차의 평균, 5회차의 평균, 전체의 평균 3가지로 확인한다. 전체 평균에 대한 값은 예측을 위해서는 사용될 수 없지만, 검증과 전체적인 패턴을 인지하기 위해 확인하는 용도로 사용된다. 즉, 예측을 위해 1회차의 평균과 5회차의 평균 2

	시청시간 충성도										70%이상 비율	80%이상 비율		
	100	90	80	70	60	50	40	30	20	10			5	
대장금 1회	15.27%	37.92%	8.84%	6.76%	10.00%	9.83%	2.99%	1.88%	1.56%	3.32%	1.07%	99.44%	68.79%	62.03%
대장금 5회	11.03%	43.32%	15.24%	7.57%	6.53%	4.54%	2.28%	2.48%	2.39%	2.59%	1.04%	99.01%	77.16%	69.59%
대장금 평균	12.48%	48.20%	21.43%	6.09%	3.57%	2.61%	1.59%	1.38%	1.30%	1.11%	0.43%	100.20%	88.21%	82.12%
주유 1회	3.60%	35.84%	16.24%	9.65%	7.02%	5.08%	5.04%	4.13%	6.70%	5.23%	1.13%	99.66%	65.33%	55.68%
주유 5회	7.04%	40.44%	20.72%	8.41%	6.13%	3.91%	3.25%	2.31%	3.21%	2.77%	1.36%	99.54%	76.60%	68.19%
주유 평균	11.97%	53.46%	17.83%	5.61%	3.41%	2.18%	1.65%	1.24%	1.96%	1.10%	0.47%	100.29%	88.88%	83.26%
내이름은 김상순 1회	8.08%	28.58%	18.81%	12.98%	9.04%	5.91%	5.34%	3.12%	1.93%	3.44%	1.36%	98.59%	68.45%	55.47%
내이름은 김상순 5회	10.55%	42.74%	17.08%	9.14%	6.16%	3.80%	3.02%	2.32%	1.65%	2.50%	0.86%	99.82%	79.51%	70.37%
내이름은 김상순 평균	12.54%	49.54%	15.88%	7.50%	4.51%	2.83%	2.21%	1.62%	1.53%	1.49%	0.57%	100.21%	85.46%	77.96%
출인 1회	8.38%	35.57%	25.53%	11.16%	3.74%	4.21%	3.61%	1.64%	1.74%	2.39%	1.05%	99.02%	80.64%	69.48%
출인 5회	11.75%	41.97%	21.25%	7.61%	3.39%	3.41%	3.04%	1.91%	2.26%	1.78%	1.25%	99.61%	82.57%	74.96%
출인 평균	13.57%	48.53%	18.21%	6.13%	3.81%	2.73%	2.08%	1.46%	1.70%	1.27%	0.72%	100.20%	86.44%	80.31%
제발달 김탁구 1회	5.41%	28.07%	19.41%	10.92%	7.99%	5.88%	5.11%	6.18%	3.73%	3.87%	2.12%	98.69%	63.81%	52.89%
제발달 김탁구 5회	6.22%	32.29%	21.15%	12.73%	7.20%	6.55%	3.70%	4.25%	2.19%	2.40%	1.19%	99.88%	72.39%	59.66%
제발달 김탁구 평균	9.85%	46.68%	17.62%	8.87%	5.00%	3.65%	2.70%	2.32%	1.72%	1.31%	0.63%	100.36%	83.03%	74.16%
베를린은 달 1회	8.37%	30.99%	16.00%	11.80%	8.02%	6.17%	5.09%	4.03%	3.88%	4.32%	2.13%	100.80%	67.16%	55.36%
베를린은 달 5회	9.68%	40.16%	15.94%	9.30%	6.86%	4.69%	3.97%	3.34%	3.04%	2.22%	1.07%	100.28%	75.09%	65.79%
베를린은 달 평균	10.62%	48.43%	16.72%	7.71%	5.02%	3.54%	2.60%	2.01%	1.84%	1.27%	0.54%	100.30%	83.47%	75.77%
선덕여왕 1회	6.91%	24.69%	12.62%	7.88%	6.42%	8.83%	6.99%	13.98%	4.65%	4.13%	1.86%	98.96%	52.10%	44.22%
선덕여왕 5회	6.97%	34.98%	17.49%	8.73%	6.54%	6.57%	5.41%	5.22%	3.15%	2.72%	1.31%	99.09%	68.17%	59.44%
선덕여왕 평균	9.27%	42.15%	23.29%	7.93%	5.25%	3.70%	2.96%	2.07%	1.70%	1.25%	0.51%	100.06%	82.64%	74.71%
전국의 계단 1회	2.69%	33.35%	20.55%	12.99%	6.73%	6.24%	5.48%	3.53%	2.86%	2.22%	1.83%	98.47%	69.58%	56.59%
전국의 계단 5회	9.88%	40.69%	18.94%	8.41%	6.14%	3.70%	3.88%	2.36%	2.12%	1.90%	1.46%	99.49%	77.91%	69.50%
전국의 계단 평균	12.82%	47.49%	17.83%	6.94%	4.25%	2.88%	2.20%	1.69%	1.43%	1.35%	0.89%	99.77%	85.08%	78.13%

(Figure 1) Distribution of viewing time of the top 10 programs with average ratings.

가지를 활용하고, 예측 데이터에서는 드라마 종영 전에 예측하는 것이 의미가 있으므로, 전체 평균은 예측에는 활용하지 않는다. 즉, 본 논문에서는 훈련데이터와 예측데이터에 따라 총 6가지 경우의 수에 대해 비교분석을 시행한다. 이렇게 계산된 충성도의 분포를 기반으로 예측을 위한 기준을 만들고, 새로운 드라마의 충성도 분포를 기준과 비교 하여 흥행 드라마 여부를 예측하는 방법을 적용한다.

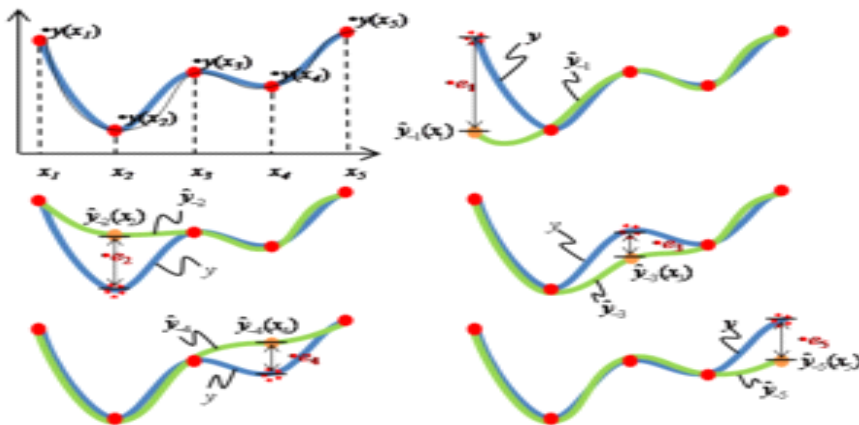
이후, 검증 데이터에 대해 예측을 하는 방식은 실시간 시계열 데이터를 예측하는 데 많이 사용되고 (Pedro and Coimbra, 2015; Taehoon and Hyunchul, 2015; Yakowitz, 1987; Yu et al., 2016), 특정 분포를 예측하는 데 주로 사용되는 (González et al., 2016; McNames, 1998) 최근접 이웃 알고리즘 (Nearest-Neighbour Methods)을 사용한다. 이를 활용해서 기준이 되는 흥행 드라마의 분포와 새로 예측을 해야 하는 드라마 분포의 거리를 계산하여 의사결정을 하는 데 활용 된다. 이러한 방식을 그림으로 표현하면 아래 방법과 같다.

최근접 이웃 알고리즘은 새로운 데이터가 들어올 때, 훈련 데이터에 있는 데이터와의 거리를 모두 구해, 가장 가까운 거리를 가지고 있는 점과 같은 예 값을 가지게 한다. 거리를 구하는 알고리즘으로는 일반적으로 유클리디언 (Euclidean) 과 상관계수 (Pearson Correlation)이 주로 사용되기 때문에 (Adeniyi et al., 2016; Hu et al., 2016), 본 논문에서는 유클리디언과 상관계수를 사용하였다. 두 개의 데이터를 p와 q라고 할 때, 유클리디언 거리는 (2)와 같이, 상관계수는 (3)과 같이 정의된다.

$$d(p, q) = \sqrt{\sum (p_i - q_i)^2} \quad (2)$$

$$r_{pq} = \frac{\sum p_i q_i - n \bar{p} \bar{q}}{(n-1) s_p s_q} \quad (3)$$

즉, 본 논문에서는 두 가지 거리 방법에 따라 총 12가지 경우의 수에 대해 비교한다. 또한 본 거리를 구하는 알고리즘 이외에 다양한 방법에 대해서도 유의미한 결과를 유지하는지 파악하기



〈Figure 2〉 Cross validation

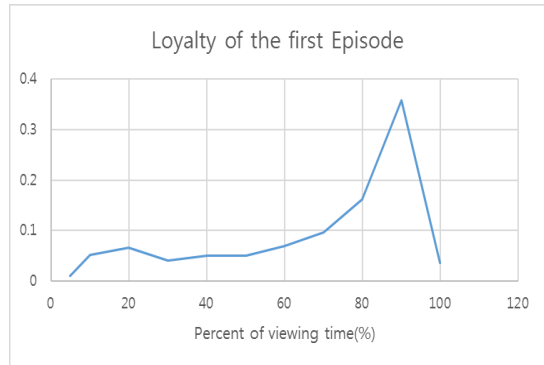
위해 Cosine 계수법, Jaccard 지수의 방법도 적용하여 결과의 건장성을 확인하였다.

4. 연구 결과

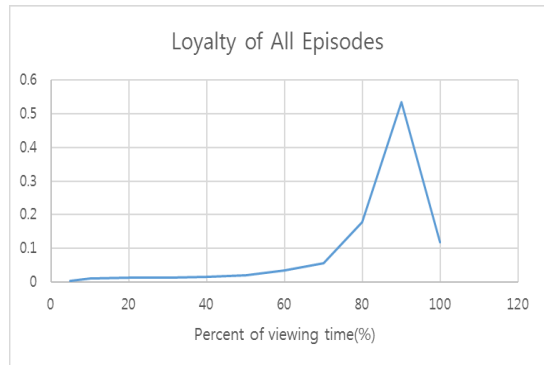
4.1 사전 분석

본 연구에서 분석을 시행하기에 앞서 사전 분석을 통해 데이터의 특성을 살펴보았다. 본 연구에서 사용될 중요한 요소는 시청자 충성도 데이터이기 때문에 이와 관련된 사전 분석을 시행해 본다. 위에서 계산된 시청률 상위 드라마의 시청자 충성도에 분포를 살펴보고 하위 드라마의 시청자 충성도에 분포를 살펴보며 경향성을 비교해 본다. 아래 그림에서 X축은 시청자의 시청 비율인 충성도를 구간별로 나타낸 것이고, Y축은 해당 구간에 포함된 시청자들의 비율을 나타낸다. <Figure 3>과 <Figure 4>는 대흥행 드라마 주몽의 1회차 충성도 분포와 전체 회차 충성도 분포를 나타내며, <Figure 5>는 비흥행 드라마 가시나무새의 1회차 충성도 분포를 나타낸다.

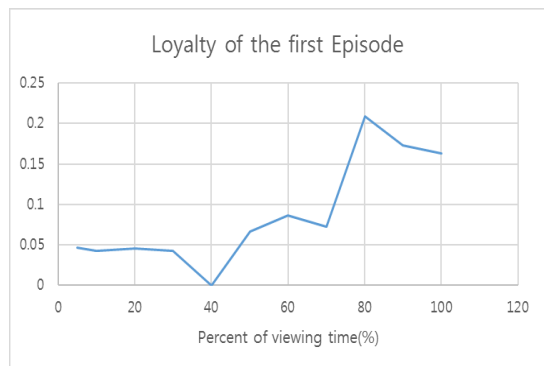
<Figure 3>과 <Figure 4>를 비교하면, 초기 회차와 전체 회차의 충성도 분포는 차이가 있음을 확인할 수 있다. 즉, 예측할 때, 초기 회차와 전체 회차 중 어떤 것을 기반으로 알고리즘을 훈련하는 예측하는 지는 중요한 문제이다. 또한, <Figure 3>과 <Figure 4>를 비교하면, 대흥행 드라마와 비흥행 드라마는 전반적으로 초기부터 충성도 분포에 차이가 있음을 알 수 있다. 따라서, 정량적으로는 아니지만 충성도 분포를 통해 대흥행 드라마를 예측하는 것이 유의미함을 알 수 있다.



<Figure 3> Loyalty distribution of Jumong



<Figure 4> Loyalty distribution of Jumong



<Figure 5> Loyalty distribution of Thornbird

4.2 예측 결과

최근접 이웃 방법으로 예측한 흥행 드라마 여부는 <Table 1>에 나와 있다. 먼저, 유클리디언 방식과 상관계수 방식을 비교하면, 유클리디언 방식은 총 평균 52.20%의 정확도를 보여주며, 상관계수 방식은 17.92%의 정확도를 보여준다. T-검증 결과 $p < 0.01$ 로 유클리디언 방식이 상관계수 방식보다 유의미하게 좋은 예측률을 보여준다는 것을 알 수 있다.

두 번째로, 유클리디언 방식 내에서 훈련 데이터를 변경해가며 확인해보면, 첫 회의 평균만을 가지고 비교하는 것은 평균 예측률이 36.80%가 나오며, 5회의 평균으로 예측을 하는 것은 평균 예측률이 52.83%가 나오고, 전체 회의 평균으로 예측을 하는 것은 평균 예측률이 66.98%가 나온다. ANOVA 검증 결과 전체 회의 평균으로 훈련 데이터를 구성하는 것이 통계적으로 유의하게 가장 좋은 결과를 나타내는 것을 알 수 있다.

마지막으로 예측 데이터를 1회까지 사용하는 것과 5회까지 사용하는 것을 결과표에서는 ‘~Episode 1’ (1회)과 ‘~Episode 5’ (5회)로 정의하였다. 이때, 1회와 5회는 각각 50.94%와 53.46%의 평균 예측률을 가지며, 큰 차이를 보이지는 않는다. 하지만, 그리드 서치의 개념에서 보았을 때, 5회에서 최종적으로 75.47%의 예측률을 보이며, 상당한 예측률을 보이는 것을 확인할 수 있었다. 대표적으로 활용되는 유클리디언 (Euclidean)과 상관계수 (Pearson Correlation)를 활용하여 결과를 도출하였지만, 모델의 건장성을 평가하기 위해 Cosine과 자카드 방법도 적용하여 분석을 시행하였다. 결과는 3% 이내의 정확도 차이가 나타나 통계적으로 유의미한 차이가 없는 것으로 판단되었다. 즉 본 모델은 거리 측정 방법에 흔들리지 않는 안정된 방법이라는 것을 확인하였다.

훈련 데이터에서 전체적인 데이터의 평균과

<Table 1> Results of the analysis

Distance	Train Data	Test Data	Accuracy
Euclidean	~ Episode 1	~ Episode 1	43.40%
	Average of Episode 1	~Episode 5	30.19%
	Average of Episode 5	~ Episode 1	50.94%
	Average of Episode 5	~Episode 5	54.72%
	Average of All Episodes	~ Episode 1	58.49%
	Average of All Episodes	~Episode 5	75.47%
Correlation	Average of Episode 1	~ Episode 1	18.87%
	Average of Episode 1	~Episode 5	22.64%
	Average of Episode 5	~ Episode 1	16.98%
	Average of Episode 5	~Episode 5	15.09%
	Average of All Episodes	~ Episode 1	16.98%
	Average of All Episodes	~Episode 5	16.98%

예측 데이터의 5회까지 데이터가 75.47%의 예측률을 보인다는 것은, 두 충성도 분포가 비슷하다는 것을 의미하며, 이는 5회까지의 시청 행태가 드라마 후반까지와 비슷한 특징을 가진다는 것을 의미하며, 5회까지 시청한 많은 사람이 후반까지 비슷하게 시청을 한다는 것을 의미한다. 이 결과는 충성도를 통한 시청률 예측이 충분히 의미가 있는 것임을 제시한다.

최종 분석 결과, 사전 분석의 결과를 기반으로 사후 분석이 진행되는 만큼, 예측에 의미가 있을지에 대한 사전 분석의 중요성은 강조되어야 한다. 이는 드라마 부분에서 적용해 보았지만 다른 카테고리에서의 예측에서는 다른 결과를 초래할 수 있다. 즉 분야별로 예측의 가능성 유무를 사전 분석을 통해서 분석해 본 후, 더욱 효율적인 분석을 진행한다면 보다 좋으리라 판단되는 부분이다.

5. 연구 의의

먼저, 본 논문은 기존과는 다른 시청률 예측 방식을 제시하면서 높은 예측률을 보여주었다. 현재 방송사들은 소위 스타시스템으로 불리는 몇몇 유명배우들에 의존함으로써, 방송 프로그램의 제작비 상승과 장기적인 경기 침체, 종합편성 채널 및 CJ 등 대기업 계열의 공격적인 투자로 그 어느 때 보다 극심한 경쟁 상황에 놓여 있어 프로그램 제작자와 방송사 모두가 재정적으로 어려운 상황이다. 이러한 방송사들의 기본적인 수익모델은 광고이며, 광고의 집행은 시청률을 기본적인 지표로 삼고 있다. 드라마는 방송사의 다양한 콘텐츠 중에서 성공 시 재정 기여도가 높지만, 드라마 시장은 상품의 특성상 투입에 따

른 수요를 예측하기 힘든 불확실성이 존재하기 때문에 드라마 제작 시에는 생산 과정에서 발생할 수 있는 수많은 불확실성을 예측하고 시장의 실패를 초래하는 위험 요소를 최소화하기 위해 노력해야 한다. 이처럼 초회 시청 시간 분포 분석을 통해 드라마의 흥행여부를 판단함으로써, 관련 기업의 대응 전략 (편성/마케팅/스토리변경 등) 수립에 실질적인 도움이 될 수 있다.

또한, 본 논문에서는 열혈 시청자들의 행태가 프로그램의 성공에 결정적인 영향을 준다는 사실을 밝혀내었다. 열혈 시청자는 열혈도가 높은 시청자로 정의가 되는데, 열혈도는 TV 시청을 얼마나 열성적으로 지속해서 시청하였는가의 척도로 본 논문에서 정의하였다. 열혈도로 고객의 충성도를 계산하여, 프로그램의 흥행을 성공적으로 예측할 수 있었다. 이처럼 충성도를 계산하는 방식은 다양한 플랫폼 (Platform)에 대한 충성도를 계산하는 방식으로 사용될 수 있다. 또한, 하이라이트, 대본 미리 보기, 메이킹 필름 (Making Film), 캐릭터, 게임 등 프로그램 연계 사업 기획, 기타 마케팅에도 유용하게 쓰일 수 있다.

6. 결론

최근에 지상파를 비롯한 국내 방송사들의 수익성은 좋지 않은 상황이다. 가장 근본적인 원인은 방송사들의 주 수입원인 광고시장의 정체될 수 있으며, 이에 따른 시청률 경쟁의 심화, 시청률을 담보할 수 있는 스타 배우를 캐스팅하기 위한 비용 등 제작비용 상승의 악순환이 계속되고 있다. 이와 같은 상황을 극복하기 위해 방송사는 프로그램과 연계한 사업 기획 및 개발, 마

캐스팅 등의 역할을 수행하는 프로그램 비즈니스 담당자를 만들어 프로그램 제작비 재원확보, 프로그램 제작비 외부 펀딩, 프로그램과 연계한 2차 콘텐츠(하이라이트, 대본 미리 보기, 메이킹 필름, 캐릭터, 게임 등) 판매로 새로운 수익원 만들기에 주력하고 있다. 이와 같은 기획과 판매에는 프로그램 인기의 척도인 시청률이 중요하며, 시청률 예측을 통해 대흥행 드라마의 탄생을 미리 알고, 충분한 시간과 투자를 통해 부가 상품을 준비할 수 있다면 그 경제적 효과는 배가될 것임이 자명하다. 따라서 시청률 예측을 하는 방식을 제시하는 것은 산업적으로 매우 중요한 문제이다.

이러한 중요성에 따라 많은 연구가 진행되었다. 하지만, 기존 연구들은 배우, 작가 등 요인들 기반 시청률을 예측하는 모형이 연구하였다. 하지만 이들은 방송은 소비자의 행태가 중요한 분야임에도 불구하고, 이를 기반으로 한 연구는 존재하지 않았다. 본 연구에서 시청자의 충성도 행태를 기반으로 추후 드라마의 흥행 여부를 평가할 수 있다는 것을 발견하였으며, 본 연구 결과 드라마의 초반 시청자 충성도는 드라마의 대흥행 여부를 예측하는데 중요한 요소임을 밝혔다. 이는 초반 시청률과는 관계없이 초반 시청자 충성도를 분석하면 결국 그 드라마의 성공 여부를 사전에 파악해 볼 수 있다는 큰 시사점을 지니고 있다. 초반 시청자 충성도는 기존 열혈도 라는 관점에서 해석될 수 있으며, 또한 고객 이탈이 최소화되는 고객의 집중도를 높여주는 것이 추후 드라마의 성공 여부를 결정짓는다는 것이다.

본 연구에서는 시청률의 분포를 계산하고 최근접 이웃 알고리즘을 이용해 흥행 여부를 예측하였다. 열혈도는 1회 평균, 5회 누적 평균, 전체 누적 평균 순으로 표본 데이터가 많아질수록 보

다 높은 예측 정확도를 나타내었다. 전체 드라마의 70% 이상을 시청한 시청자의 분포만으로도 최대 65.45%의 확률로 대흥행 드라마 여부를 알 수 있었으며, 시청 시간의 전체 분포 패턴을 분석한 유사도 검증을 통해서 신작 드라마의 초기 시청률(1회 또는 5회 이내)만으로 최대 75.47%의 확률로 대흥행 드라마의 탄생을 예측할 수 있었다.

이와 같은 현장에서의 필요성에 의해서, 본 연구에서는 시청시간(충성도)만으로도 75% 이상의 대흥행 드라마는 예측할 수 있는 모형을 만들었으나, 이는 드라마가 방송되기 전의 사전 예측이 아닌, 1회에서 최대 5회까지의 방송을 한 이후의 사후 예측이라는 점에서 본 연구의 한계가 있다고 할 수 있다. 또한 75% 보다 높은 90% 이상의 높은 정확성이 필요하며, 이 부분은 추후 연구 데이터의 확장과 보완을 통해서 정확도를 높일 수 있을 것으로 생각된다.

참고문헌(References)

- Adeniyi, D., Z. Wei, and Y. Yongquan, "Automated Web Usage Data Mining and Recommendation System Using K-Nearest Neighbor (Knn) Classification Method," *Applied Computing and Informatics*, Vol.12, No.1(2016), 90~108.
- Ahn, J., K. Ma, O. Lee, and S. Sura, "Do Big Data Support Tv Viewing Rate Forecasting? A Case Study of a Korean Tv Drama," *Information Systems Frontiers*, Vol.19, No.2(2017), 411~420.
- Bae, J., "An Analysis on the Factors in Drama Ratings - Focusing on the Drama Attributes

- and Audience Factors," *Korean Association for Broadcasting & Telecommunication Studies*, Vol.19, No.2(2005), 270~309.
- Corporation, B. B., *Building Public Value: Renewing the Bbc for a Digital World*. BBC, 2004.
- Dick, A. S., and K. Basu, "Customer Loyalty: Toward an Integrated Conceptual Framework," *Journal of the academy of marketing science*, Vol.22, No.2(1994), 99~113.
- González, M., C. Bergmeir, I. Triguero, Y. Rodríguez, and J. M. Benítez, "On the Stopping Criteria for K-Nearest Neighbor in Positive Unlabeled Time Series Classification Problems," *Information Sciences*, Vol.328 (2016), 42~59.
- Hu, L.-Y., M.-W. Huang, S.-W. Ke, and C.-F. Tsai, "The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets," *SpringerPlus*, Vol.5, No.1(2016), 1304~1312.
- Jeong, Y., and J. Shin, "Communication Industry Report." *Korea Communications Commission*, 2013.
- Kotler, P., K. L. Keller, M. Brady, M. Goodman, and T. Hanser, *Marketing Management*, Harlow: Pearson Education, 2016.
- Kim, K., "In Search of Blockbuster Drama: The Effects of Genre Familiarity, Star Power, and Internet Accounts on Rating Performance," *Journal of Strategic Management*, Vol.12, No.1 (2009): 25-54.
- Kwon, H., W. Park, S. Yoon, and S. Kim, "Analysis of Success Factor on Tv Drama," *KOCCA Research Report*, 2009.
- Lee, W., N. Lee, and J. Kim, "An Empirical Study on Forecasting Model of Popularity Rating for Drama Programs," *Journal of Digital Contents Society*, Vol.13, No.3(2012), 325~334.
- McNames, J., "A Nearest Trajectory Strategy for Time Series Prediction," *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*: KU Leuven Belgium, 1998.
- Pareto, V., and A. N. Page, "Manual of political economy," *AM Kelley*, 1971.
- Park, J., and H. Jeon, "What Did They Read in the Newspapers? - a New Method of Measuring Readership," *Korean Association For Communication and Information Studies*, Vol.35(2006), 211~249.
- Pedro, H. T., and C. F. Coimbra, "Nearest-Neighbor Methodology for Prediction of Intra-Hour Global Horizontal and Direct Normal Irradiances," *Renewable Energy*, Vol.80(2015), 770~782.
- Reinartz, W. J., and V. Kumar, "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of marketing*, Vol.67, No.1(2003), 77~99.
- Ryoo, W., "Globalization, or the Logic of Cultural Hybridization: The Case of the Korean Wave," *Asian Journal of Communication*, Vol.19, No.2(2009), 137~151.
- Sabavala, D. J., and D. G. Morrison, "A Model of Tv Show Loyalty," *Journal of Advertising Research*, Vol.17, No.6(1977), 35~43.
- Singh, S. S., D. C. Jain, and T. V. Krishnan, "Research Note—Customer Loyalty Programs: Are They Profitable?," *Management science*, Vol.54, No.6(2008), 1205~1211.

- Kim, T., and H. Ahn, "A Hybrid under-Sampling Approach for Better Bankruptcy Prediction," *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 173~190.
- Tebert, M., "Erfolg Durch Qualität. Programmcontrolling Beim Wdr Fernsehen," *Media Perspektiven*, Vol.2, No.2000(2000), 85~93.
- Wittebols, J. H, *The Soap Opera Paradigm: Television Programming and Corporate Priorities*. Rowman & Littlefield, 2004.
- Yakowitz, S., "Nearest-Neighbour Methods for Time Series Analysis," *Journal of time series analysis* Vol.8, No.2(1987), 235~247.
- Yu, B., X. Song, F. Guan, Z. Yang, and B. Yao, "K-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition," *Journal of Transportation Engineering*, Vol.142, No.6(2016), 04016018 1~10.
- Yu, S., and S. Kim, "An Analysis of Factors Which Affect the Rating of Drama ; Focusing on the Production Factors of Outsourcing Drama," *Media economy and culture*, Vol.8, No.3(2010), 7~48.

Abstract

Prediction of a hit drama with a pattern analysis on early viewing ratings

Kihwan Nam* · Nohyoon Seong**

The impact of TV Drama success on TV Rating and the channel promotion effectiveness is very high. The cultural and business impact has been also demonstrated through the Korean Wave. Therefore, the early prediction of the blockbuster success of TV Drama is very important from the strategic perspective of the media industry.

Previous studies have tried to predict the audience ratings and success of drama based on various methods. However, most of the studies have made simple predictions using intuitive methods such as the main actor and time zone. These studies have limitations in predicting. In this study, we propose a model for predicting the popularity of drama by analyzing the customer's viewing pattern based on various theories. This is not only a theoretical contribution but also has a contribution from the practical point of view that can be used in actual broadcasting companies.

In this study, we collected data of 280 TV mini-series dramas, broadcasted over the terrestrial channels for 10 years from 2003 to 2012. From the data, we selected the most highly ranked and the least highly ranked 45 TV drama and analyzed the viewing patterns of them by 11-step.

The various assumptions and conditions for modeling are based on existing studies, or by the opinions of actual broadcasters and by data mining techniques. Then, we developed a prediction model by measuring the viewing-time distance (difference) using Euclidean and Correlation method, which is termed in our study similarity (the sum of distance). Through the similarity measure, we predicted the success of dramas from the viewer's initial viewing-time pattern distribution using 1~5 episodes. In order to confirm that the model is shaken according to the measurement method, various distance measurement methods were applied and the model was checked for its dryness. And when the model was established, we could make a more predictive model using a grid search

* College of Business, Korea Advanced Institute of Science and Technology (KAIST)

** Corresponding Author: Seong NohYoon

College of Business, Korea Advanced Institute of Science and Technology (KAIST)

85 Hoegi-ro, Dongdaemun-gu, Seoul 130-722, Republic of Korea

Tel: +82-10-4542-4525, E-mail: nyseong@kaist.ac.kr

Furthermore, we classified the viewers who had watched TV drama more than 70% of the total airtime as the “passionate viewer” when a new drama is broadcasted. Then we compared the drama’s passionate viewer percentage the most highly ranked and the least highly ranked dramas. So that we can determine the possibility of blockbuster TV mini-series. We find that the initial viewing-time pattern is the key factor for the prediction of blockbuster dramas. From our model, block-buster dramas were correctly classified with the 75.47% accuracy with the initial viewing-time pattern analysis.

This paper shows high prediction rate while suggesting audience rating method different from existing ones. Currently, broadcasters rely heavily on some famous actors called so-called star systems, so they are in more severe competition than ever due to rising production costs of broadcasting programs, long-term recession, aggressive investment in comprehensive programming channels and large corporations. Everyone is in a financially difficult situation. The basic revenue model of these broadcasters is advertising, and the execution of advertising is based on audience rating as a basic index. In the drama, there is uncertainty in the drama market that it is difficult to forecast the demand due to the nature of the commodity, while the drama market has a high financial contribution in the success of various contents of the broadcasting company. Therefore, to minimize the risk of failure. Thus, by analyzing the distribution of the first-time viewing time, it can be a practical help to establish a response strategy (organization/marketing/story change, etc.) of the related company. Also, in this paper, we found that the behavior of the audience is crucial to the success of the program. In this paper, we define TV viewing as a measure of how enthusiastically watching TV is watched. We can predict the success of the program successfully by calculating the loyalty of the customer with the hot blood. This way of calculating loyalty can also be used to calculate loyalty to various platforms. It can also be used for marketing programs such as highlights, script previews, making movies, characters, games, and other marketing projects.

Key Words : Similarity, Viewing-time pattern, prediction of blockbuster drama, Nearest Neighbor

Received : May 20, 2018 Revised : November 6, 2018 Accepted : November 26, 2018

Publication Type : Conference(Fast-track) Corresponding Author : Nohyoon Seong

저 자 소개



남기 환

KAIST 경영대학원 경영공학부에서 MIS 박사학위를 취득하였다. 현재 KAIST 경영대학원 경영공학부, UNIST 경영학부 겸임교수로 재직 중이다. 주요 관심분야는 Business Analytics & Business Intelligence, Big Data Analytics, Data Mining, Statistical Analysis, Recommender Systems, Econometrics Models, Machine Learning, Deep Learning 등 이다. 관련 연구들은 *Decision Support Systems*, *Data Mining and Knowledge Discovery* 등에 논문이 게재되었다. 학문적인 연구뿐만 아니라 이론을 바탕으로 실제 기업에서 다양한 프로젝트를 성공적으로 진행함으로써 학계와 산업계 모두에 실증적인 기여를 하고 있다.



성 노 윤

KAIST에서 물리학 학사 학위를 취득하였다. 현재 KAIST 경영대학원 경영공학부 MIS 박사 과정에 재학중이다. 주요 관심분야는 자연어 처리, 머신러닝, 빅데이터 분석, 계량 경제학, 경제물리학 등이다. 기존 경제학 이론에 머신러닝을 접목하여 사회 전반적인 문제를 해결하는 데에 관심을 가지고 있다.