

머신러닝을 이용한 공연문화예술 개인화 장르 추천 시스템

A Personalized Recommendation System Using Machine Learning for Performing Arts Genre

김 형 수 (Hyung Su Kim) 한성대학교 스마트경영공학부, 교신저자
박 예 린 (Yerin Bak) 한성대학교 무역학과
이 정 민 (Jeongmin Lee) 한성대학교 지식정보학부

요 약

공연문화예술 시장의 확대에도 불구하고, 중소규모 공연장은 소비자의 정보 접근성이 좋지 않아 어려움을 겪고 있다. 본 연구는 중소규모 공연장의 마케팅 역량을 강화할 수 있는 하나의 대안으로써 머신러닝 기반의 장르 추천 시스템을 제시하고자 한다. 국내 한 공연장의 고객 마스터 DB와 거래이력 DB를 활용하여 고객당 3개의 장르를 추천하는 5개의 추천 시스템을 개발하였다. 추천시점 이후 1년 동안의 실제 공연구매 이력을 바탕으로 추천 시스템의 성능을 비교하여 최적의 추천시스템을 제안하였다. 분석 결과, 단일 예측모형보다는 앙상블 모형 기반의 추천시스템이 우수한 성능을 보이는 것으로 나타났다. 본 연구는 공연문화예술 분야에는 일천했던 개인화 추천 기법을 적용했고, 분석 결과 공연문화예술 분야에서도 충분히 활용할 만한 가치가 있음을 시사하고 있다.

키워드 : 공연문화예술, 머신러닝, 추천 시스템, 고객관계관리, 연관성 분석, 협업 필터링, 앙상블

I. 서 론

최근 들어 뮤지컬, 연극, 클래식 등 다양한 공연이 제작되고 흥행하면서, 우리나라의 공연문화예술시장의 규모는 급속히 확대되고 있다(유은영, 진현정, 2014). 공연 매출액을 보면, 2008년 2,322억 원에서 꾸준한 성장세를 보이며 2017년 4,600억 원대로 진입하였고, 지난 10년간 99% 성장한 것으로 나타났다(문화체육관광부, 2018). 향후에도 소득수준이 높아지고 여가 생활에 대한 중요성

이 커지면서 공연문화예술시장은 지속적으로 확대될 것으로 보인다.

그러나, 공연문화예술시장 내의 공연장 규모에 따라 실적의 차이는 크게 벌어지고 있다. 2018 공연예술실태조사 결과에 따르면, 2017년 국내 공연시장 규모는 전년 대비 8.7% 증가했지만, 같은 기간 중소 공연장과 공연단체의 실적은 오히려 감소한 것으로 나타났으며, 전체 중 92.8%를 차지하는 공공 민간 공연단체의 공연 실적은 7.2%에 불과한 대형 기획사의 공연 실적보다 현저히 낮은 수준이

다(문화체육관광부, 2018). 이와 같은 현상은 공연 규모에 따른 마케팅 방식의 차이에서 비롯된다고 볼 수 있다. 대형 극장 작품의 경우 대중 매체를 활용한 전폭적인 광고를 제공함으로써 회마다 90% 이상의 관객 점유율을 기록하고 있지만(손예지, 2019), 소극장 공연의 경우 단순 지역홍보나 초대권 배부를 통한 마케팅 활동이 전부이기 때문에(김제관, 2013) 관객모집의 어려움을 겪고 있다. 이러한 현상은 예술경영지원센터(2018)의 연구결과에 따르면, 대규모 공연장에 비해 중소규모 공연장의 경우 주로 구전과 같은 제한된 방식으로 공연상품 정보를 접하고 있기 때문에 정보 접근성이 상대적으로 떨어진다는 점을 지적하였고, 김선영(2018)의 연구와 문화체육관광부(2016)의 보고서에서는 이러한 정보 접근성과 공연장의 접근성이 공연을 자주 보지 못하는 주요 요인으로 제시하였다.

그럼에도 불구하고, 현재 대부분의 중소 공연단체나 공연장의 경우 모든 고객에게 동일한 공연 정보를 안내하거나, 일정한 규칙에 따라 공연중인 상품을 무작위로 선정하여 홍보하는 이른바 랜덤 추천 방식으로 마케팅을 운영하고 있는 실정이다. 그러나, 이러한 중소 규모 공연장 역시 온/오프라인 회원가입을 통한 고객정보를 확보하고 있으므로 마케팅적 노력에 따라 개인화 상품추천 방법론에 의해 이러한 정보 접근성의 한계를 상당히 극복할 수 있다. 개인화 상품 추천 방법론은 일종의 1:1 마케팅 기술로서 자사의 상품 중 고객의 니즈에 따라 차별화된 상품정보를 제공할 수 있는 효과적인 전략으로 많은 산업영역에서 활용되어 왔으나, 공연문화예술 분야에서는 아직 연구와 활용이 매우 미비한 실정이다. 본 연구에서는 공연문화예술 상품에 대한 소비자 정보 접근성의 한계를 극복하고, 중소 규모 공연장의 공연 실적 개선을 도모하기 위한 하나의 방안으로써 개인화 상품추천 모형을 제안하고자 한다. 개인화 상품추천은 도메인 전문가에 의해 창출된 일련의 추천규칙 집합에 의해 상품이 추천되는 이른바 전문가 시스템(Expert System)에 의한 개인화 추천방식도 존재하

지만, 본 연구에서는 머신러닝 알고리즘에 기반한 개인화 추천 방법론을 제시하고자 한다. 협업 필터링이나 연관규칙과 같은 머신러닝 모형에 기반한 개인화 추천모형은 고객들의 구매행태를 학습하여 결정된 추천규칙으로 운영되기 때문에 기존의 추천방법론에 비해 더욱 현실적이고, 추천성과가 우수한 것으로 알려져 왔다(e.g., 안현철 등, 2006; 박지희, 남기환, 2017).

이를 위해 경기도 소재의 한 중견 공연장의 고객 마스터 데이터와 거래이력 데이터를 활용하여 5개의 추천 모형을 설계하고, 각 추천 모형의 성능을 비교함으로써 공연문화예술분야에 가장 적합한 추천 모형을 확인하고자 한다. 개인화된 장르 추천 시스템의 성능이 검증될 경우 중소 규모 공연장은 인터넷 및 모바일 환경의 홈페이지를 통해 접속하는 고객에게 개인의 취향에 맞는 문화공연 상품을 안내해줄 수 있을 뿐만 아니라, 회원들에게 맞춤형 아웃바운드 커뮤니케이션을 제공하게 됨으로써 문화공연 상품에 대한 정보 접근성의 문제를 어느 정도 해소할 수 있을 것으로 기대한다.

II. 관련 연구

2.1 공연문화예술 마케팅

공연문화의 확대와 문화소비의 증가가 진행되면서 공연문화예술 마케팅에 관해 비교적 짧은 시간 동안 많은 연구들이 축적되었다. 공연문화예술 마케팅에 관한 기존 연구들은 크게 고객 관점, 시장 관점, 전략 관점으로 구분하여 고찰할 수 있다.

고객 관점에서의 공연문화예술 마케팅에 관한 대표적인 연구로서 김범석, 전인오(2015)는 문예회관 이용객을 대상으로 고객 만족도와 재관람 의도에 유의미한 변수를 찾기 위해 회귀분석을 진행하였고, 분석 결과 품성적 특성 변수(관심도, 작품에 대한 사전 지식 등), 상황적 특성 변수(출연진 등), 정보적 특성 변수(공연 후기 등)가 고객 만족도를 높이고 재방문 유도를 할 수 있다는 것을 제시

하였다. 또한, 유은영, 진현정(2014)은 공연 소비자의 인구통계학적 변수와 인지 및 행동적 특성에 해당하는 관여도, 지식수준, 정보획득 수단을 기준변수로 소비자그룹을 분류하였고, 관여도와 지식수준이 인구통계학적 변수보다 공연 소비에 더 큰 영향을 미치는 것을 입증하였다. 한편, Andreasen(1990)의 연구에서는 ‘공연예술 수용 과정’ 모델을 통해 관객 유형을 무관심관객, 관심관객, 열성관객 3가지로 구분하고, 공연 소비자들은 다양한 동기를 통해 단계적으로 상위의 소비자 유형으로 발전한다는 것을 제시함으로써 공연관람 확대를 위한 시사점을 제공하였으며, 이러한 Andreasen(1990)의 연구를 바탕으로 정순규(2013)는 군집분석을 활용하여 관객 유형을 열성관객, 관심관객 그리고 무관심관객으로 분류하고, 유형에 따라 공연장 서비스 품질, 공연작품 품질, 공연 관람 동기

등이 구매의사에 영향을 미치는 정도의 차이가 있다는 것을 실증분석을 통해 검증하였다. 이상의 선행 연구들에서 보여주고 있듯이 고객 관점에서의 공연문화예술 마케팅 연구는 관객 유형 분류와 분류된 관객 유형에 따른 공연 소비자의 공연 관람 구매 요인에 관한 분석이 주로 진행되었다는 것을 알 수 있다.

시장 관점에서의 연구는 시장 세분화와 관객 개발 전략에 관한 연구가 주를 이루고 있다. 조남규(2010)는 전용극장 관객개발과 마케팅의 성공적인 사례 분석을 통해 공연예술시장의 마케팅 필요성을 주장하였고, Kangun *et al.*(1992)은 심포니 관객의 구매의도를 통해 4개의 세분시장으로 분류하고, 각 세분시장별 티켓 구매에 영향을 미치는 구매 태도, 공연에 대한 관심 등의 구매 요인 변수들에 대해 파악하였다(김소영, 신혜선, 2006).

〈표 1〉 공연문화예술 마케팅 관련 연구

	저자	내용
고객 관점	김범석, 전인오 (2015)	공연예술 수용자의 관점에서 통계적 기법을 통해 작품적 선택 특성과 서비스 품질요인이 사후행동에 미치는 영향을 제시했다.
	유은영, 진현정 (2014)	개인특성변수를 사용하고 의사결정나무 기법을 활용해 공연 마케팅 전략에 활용될 기초적인 정보를 제시했다.
	정순규(2013)	군집 분석을 활용해 관객 유형을 분류한 후 관객 유형에 따라 공연 관람 구매 의사 요인에 대한 분석을 실시했다.
	Andreasen(1990)	공연 소비자를 ‘공연예술 수용과정’모형을 기반으로 3가지 유형으로 분류하고 각 유형별 소비 동기를 제시했다.
시장 관점	조남규(2010)	공연예술시장의 관객개발과 마케팅 전략 연구의 필요성에 대해 서술하며 국내 공연예술의 발전방안과 전망에 대해 제시했다.
	김소영, 신혜선 (2006)	뮤지컬 시장의 대학생 관객층을 mixture model을 통해 4개의 세분시장으로 구분하여 세분시장 간 유의한 세분화 변수들을 제시했다.
	Kangun(1992)	오케스트라 시장의 대학생 관객층을 대상으로 시장 세분화를 하고 세분화 유형별 구매 의사 요인에 대한 분석을 진행했다.
전략 관점	김현정(2015)	인구통계학적 특성에 따른 무용 공연 관람여부와 관람 결정에 영향을 미치는 요인을 분석을 실시해 고객 개발을 위한 마케팅 방안 제공했다.
	지숙영 등(2011)	실제 온라인 구매 기록을 기반으로 소비 행위를 측정된 결과를 토대로 공연예술문화 분야의 정책 수립의 근거를 제시했다.
	Kotler and Armstrong(2012)	공연문화예술 시장에서 관객 분류는 소비자 지향적 마케팅 전략 수립을 위한 기초적 정보임을 주장했다.
	Kolb(2000)	공연문화예술 마케팅에서의 관객개발을 위한 마케팅믹스 및 인구통계학적 분석의 중요성을 제시했다.

이와 유사하게 김소영, 신혜선(2006)은 뮤지컬 대학생 시장을 Mixture Model로 뮤지컬 선호형, 뮤지컬 매니아형, 뮤지컬 중간형, 뮤지컬 무관심형의 4개 세분시장으로 구분하였고, 각 세분시장에서의 관람 행동, 라이프스타일, 인구통계학적 특성에서 유의한 차이가 있음을 파악해 이를 토대로 각 세분시장별 마케팅 전략을 제시하였다.

한편, 전략 관점에서의 연구로서 김현정(2015)은 인구통계학적 변수에 따른 무용공연 관람여부와 관람 결정에 영향을 미치는 요인을 분석하여 마케팅 전략 수립 시 필요한 기초자료를 제공하였다. 또한, 지숙영 등(2011)은 실제 온라인 구매 기록을 통해 누가, 언제, 어떻게 문화상품을 소비했는지 한국의 문화소비자형을 객관적으로 조망하고 공연문화예술 분야의 정책 수립의 근거를 제시하였다. 또한, Kotler and Armstrong(2012)은 개인특성요인 분석을 활용한 관객 분류가 소비자 지향적 마케팅 전략을 설계하기 위한 기초적인 정보라는 것을 주장하였고(유은영, 진현정, 2014), Kolb(2000)는 성공적인 관객 개발을 위해 공연예술 관객의 6단계의 수용과정 중 ‘도입기’ 단계에서 인구통계학적 분석과 마케팅믹스가 필요하다는 것을 주장했으며, 신규 관객개발을 위해 관람동기를 파악할 수 있는 정성적 연구의 필요성을 제시하였다. 이와 같이 전략 관점에서의 공연문화예술 마케팅에 관한 연구는 공연문화예술 마케팅 전략 수립 시 필요한 기초적인 정보와 근거를 제시하는 방향으로 진행되었다.

이상의 선행연구에서 보듯이 공연문화예술 마케팅 연구는 관객의 구매요인 분석이나 시장 세분화, 마케팅 믹스 등에 연구가 주를 이루고 있으며, 최근 4차 산업혁명의 구성기술로 확대되고 있는 머신러닝 기반의 디지털마케팅 기술에 관한 연구는 일천한 상황이라고 할 수 있다.

2.2 추천 시스템

추천 시스템은 사용자의 구매행태에 기반하여

상품의 정보를 능동적으로 제시할 수 있는 기술로써 아마존과 넷플릭스 등을 비롯하여 이미 상당히 많은 상거래 영역에서 활용되고 있다(손지은 등, 2015). 추천 시스템의 핵심은 추천 알고리즘이라고 할 수 있는데, 대표적인 추천 알고리즘은 연관성 분석, 협업 필터링, 앙상블 기법 등이 있다(손지은 등, 2015). 우선 연관성 분석이란 동일한 거래에서 동시에 출현한 상품의 패턴을 연관 규칙의 형태로 표현한 것이다(Agrawal *et al.*, 1993; 이동원, 2017). 연관성 분석을 활용한 추천 시스템 연구로서 이동원(2017)은 회귀분석 모형과 결합한 선호도 추출 모형의 성능이 우수함을 주장하였으며, Aggarwal *et al.*(1998)은 고객 프로필 기반의 연관 규칙 알고리즘을 제시하였다. 이와 유사하게 안현철 등(2006)은 고객 프로필 기반의 연관규칙 알고리즘을 조정 에이전트와 결합한 추천 모형을 제안함으로써 유용성을 입증하였다.

협업 필터링은 아이템이나 사용자의 유사성을 분석한 결과를 바탕으로 선호도를 추론하는 방법으로써 사용자 기반 협업 필터링과 아이템 기반 협업 필터링으로 구분할 수 있다(정영진, 조운호, 2017). 사용자 기반 협업 필터링은 대상 사용자의 행위를 분석하고 사용자 사이의 유사도를 측정하여 유사한 사용자를 찾고 유사 사용자의 아이템 평가 정보를 기반으로 대상 사용자의 추천 목록을 생성하고 이를 추천하는 기법으로 다양한 상거래 영역에서 활용되어 왔다(Kim *et al.*, 2010; 박두순, 2013). 이와 관련된 연구로서 박지희, 남기환(2017)은 오프라인 매장 데이터에 사용자 기반 협업 필터링 알고리즘을 적용하여 추천 수량과 재 추천을 고려하는 범용적 추천 시스템을 제시하였다. 한편, 아이템 기반 협업 필터링은 대상 사용자가 사용한 아이템과 유사도가 높은 다른 아이템을 추천하는 방식으로써(강호운, 옥창수, 2015), 사용자별 유사성과 아이템별 연관성을 동시에 고려한 혼합형 협업 필터링(강호운, 옥창수, 2015)이나 협업 필터링의 희박성, 확장성 문제를 해결하기 위한 장르별 협업 필터링(이재식, 박석두, 2007) 등 다양한 과생

기법이 연구되어 왔으며(Vozalis and Margritis, 2004), 대표적 연구로서 Kim *et al.*(2010)은 새로운 아이템의 추천을 위하여 선호 바운더리 알고리즘을 결합한 혼합형 협업 필터링을 제시하고, 기존의 협업 필터링과 비교함으로써 성능의 우수성을 입증하였다.

한편, 앙상블 기법은 분류의 정확성 개선을 위해 여러 개의 분류기를 결합하여 사용하는 추천 알고리즘으로써(박철용, 2016), 대표적으로 Bagging과 Bumping, 랜덤 포레스트(Random Forest) 등의 알고리즘이 있다(이연정, 김경재, 2013; 이택승, 2018). 의사결정나무 분석을 심화시킨 랜덤 포레스트는

하나의 나무가 아닌 여러 개의 나무로 확장해 기계 학습을 기반으로 예측하는 모형으로써 취업교육 사이트와 같이 사용자에게 콘텐츠를 추천하는 용도로 사용될 수 있음이 확인되었다(이택승, 2018). Bagging은 기계학습기법에서 분류나 예측의 성과와 안정성 제고를 위해 여러 개의 기계학습기법의 결과를 조합하는 기법이며, Bumping은 분류기의 가장 낮은 오차 결과값을 기반으로 분류기를 결합하는 기법으로써 마찬가지로 일반 상품뿐만 아니라 다양한 형태의 서비스 추천에도 사용될 수 있음이 확인되었다(이연정, 김경재, 2013).

〈표 2〉 추천 시스템 관련 연구

	저자	내용
연관성 분석	이동원(2017)	추천 상품의 순위를 결정하기 위해 연관성 분석 기법에 회귀분석모형을 적용하여 선호도를 도출하는 방안을 제시했다.
	안현철 등(2006)	연관성 분석과 분류 기법을 조정 에이전트를 통해 결합한 새로운 추천시스템 모형을 제안했다.
	Aggarwal <i>et al.</i> (1998)	고객 프로필 정보와 구매 행태 간의 관계를 연관성 분석을 통해 연관규칙 알고리즘을 제안했다.
사용자 기반 협업 필터링	박지희, 남기환 (2017)	오프라인 매장 POS 데이터를 활용한 사용자 기반 협업 필터링을 통해 매장 재고관리에 용이한 추천 시스템을 제안했다.
	박두순(2013)	협업 필터링의 희박성 문제를 해결하기 위해 가중치를 가진 개인 성향을 협업 필터링에 활용하는 추천시스템을 제안했다.
	Kim <i>et al.</i> (2010)	새로운 아이템의 추천을 위한 협업 필터링을 접목시킨 하이브리드 알고리즘을 제시하였다.
아이템 기반 협업 필터링	강호윤, 옥창수(2015)	협업 필터링의 한계점을 개선하기 위해 아이템 간 연관성을 고려한 혼합형 추천시스템을 제안했다.
	이재식, 박석두(2007)	아이템을 최종으로 추천하기 전 상위 카테고리를 활용하는 장르별 협업 필터링을 제안했다.
	Vozalis and Margritis(2004)	인구통계학 변수와 아이템 유사도를 활용한 하이브리드형 협업 필터링을 제안했다.
앙상블	이택승(2017)	교육 사이트 실제 데이터를 활용해 기존 협업 필터링보다 성능 우수성이 높은 랜덤 포레스트 기반의 하이브리드 추천시스템을 제시했다.
	김민정, 조운호 (2015)	고객 행태를 표현할 수 있는 5종의 다중 프로파일을 3가지 앙상블 기법을 활용해 협업 필터링의 이웃 탐색 과정에 적용했다.
	최예림 등(2015)	모바일 디바이스 사용 로그를 분석하여 잠재 관심 콘텐츠를 추천하는 통계적 학습과 규칙 기반의 앙상블 기법을 제안했다.
	이연정, 김경재 (2013)	기호 반응을 한 데이터마이닝 기법과 다중모형조합기법을 결합한 상품추천시스템 모형을 제안했다.
	Zhang and Min(2016) Pazzani(1999)	3가지 랜덤 포레스트 모형을 결합해 고객 선호도 기반의 추천시스템을 제안했다. 다중프로파일 데이터에 앙상블 기법을 적용하는 새로운 추천시스템을 제시했다.

추천 시스템 알고리즘은 다양한 업종에서 사용되어 왔으나, 일반적으로 단일 모형보다는 앙상블 모형에 의한 개인화 추천 성능이 더 우수하다고 알려져 왔다(Pazzani, 1999; 이원철, 정석봉, 2017; 최예림 등, 2015). 최예림 등(2015)은 콘텐츠 소비 로그 패턴을 분석하여 잠재 관심 콘텐츠를 추천하기 위해 단일 분류모형과 앙상블 모형을 비교했으며, Pazzani(1999)는 레스토랑 고객의 선호 정보와 해당 고객의 인구통계학적 정보, 레스토랑 상품 정보 등 서로 다른 유형의 프로파일 데이터에 대해 협업 필터링과 내용 기반 필터링 기법을 이 두 가지 모형을 결합한 앙상블 기법과 비교하여 앙상블 모형이 더 우수하다는 것을 증명하였다(이원철, 정석봉, 2017). 마찬가지로 김민정, 조운호(2015)는 선호사이트, 인구통계학 정보, 인터넷 사용행태, 검색 키워드, rating 등을 활용한 앙상블 개인화 추천 모형을 제시하였고, 이러한 앙상블이 단일 예측 모형보다 성능이 우수하다는 것을 주장하였다. 그러나, 앙상블 모형의 우수성은 아직 공연문화예술 산업 영역에서 검증된 연구가 없으므로 본 연구의 주요 목적 중 하나는 이러한 개인화 추천을 위한 앙상블 모형이 공연문화예술 산업에서도 상대적으로 우수한 성능을 보이는 지를 검증하는 일일 것이다.

Ⅲ. 연구 모형 및 설계

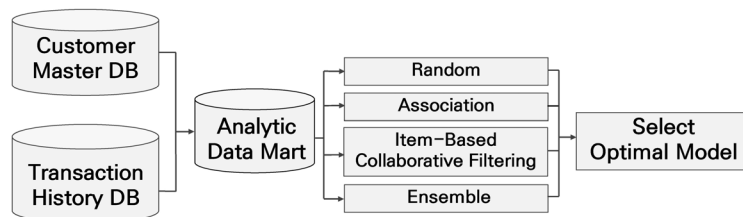
3.1 연구 모형

본 연구에서는 공연문화예술 마케팅에서 소비

자에게 맞춤형 정보를 제공하기 위한 방안으로서 공연장의 실제 데이터를 기반으로 5가지의 가능한 추천 시스템을 적용한 후 최적의 추천 시스템을 파악하여 고객별 3개의 장르를 추천하는 방안을 제시하고자 한다(<그림 1> 참조).

데이터분석을 위해 고객마스터 DB의 고객 프로파일 정보와 거래이력 DB의 거래정보를 결합하여 분석용 데이터마트를 생성하고, 무작위 추천 모형, 연관성 분석 모형, 아이템 기반 협업 필터링 모형, 그리고 앙상블 추천 모형(랜덤 포레스트와 그래디언트 부스팅)을 학습하여 생성된 추천 규칙에 따라 고객별 총 3가지의 공연정보를 추천한다. 추천 시스템의 다양성을 고려한 성능비교 연구를 위해서는 각 고객에게 복수 개의 상품을 추천하는 방법이 일반적이다(Adomavicius and Kwon, 2012; 손지은 등, 2015). 본 연구에서는 각 고객에게 추천되는 상품 목록의 다양성을 높이고, 지나치게 많은 아이템을 동시에 추천함으로써 발생할 수 있는 고객의 반감을 방지하기 위해 최대 3순위까지의 장르를 추천한다.

연구모형에서 랜덤 추천 시스템은 머신러닝 추천 모형을 사용하지 않은 채 무작위로 장르를 추천하는 모델로써 다른 머신러닝 추천 시스템과의 성능비교를 위해 투입되었다. 연관성 분석 기반 추천 시스템은 고객의 거래이력 데이터를 기반으로 구매상품 간 연관 규칙을 도출하고, 이를 기반으로 장르를 추천하게 되며, 아이템 기반 협업 필터링 시스템은 장르 간 아이템의 유사도를 계산하여 고객별 추천 장르를 도출하게 된다. 한편, 앙상블 모형 기반의 추천 시스템은 랜덤 포레스트와



<그림 1> 연구 모형도

그래디언트 부스팅 알고리즘을 기반으로 상품들의 구매 확률값을 추정하여 고객별 장르를 추천한다. 연구모형의 마지막 단계인 최적 모형 선정 단계에서는 이전 단계에서 학습된 각 추천 시스템의 성능을 평가하여 가장 우수한 성능의 추천 시스템을 선정하는 단계이다. 5가지 학습된 추천 시스템을 별개의 검증용 데이터에 적용함으로써 실제 구매 여부를 바탕으로 추천 적중률(hit-rate)을 계산하고 이를 상호 비교하였다. 여기에서 각 추천 시스템은 고객별 3순위의 장르까지 추천이 가능하도록 설정하였기 때문에 각 추천 시스템의 적중률은 추천된 3순위 장르 중 구매에 성공한 경우를 판단하여 계산하였다.

3.2 데이터 수집 및 전처리

본 연구에서 사용한 데이터는 국내 경기도 소재의 A공연장의 실제 데이터로서 2015년 1월 1일부터 2017년 11월 10일까지 수집된 고객 프로파일 데이터 62,714건과 거래이력 데이터 112,938건으로 구성된다. 이 두 가지 DB를 결합하여 분석용 데이터마트를 생성하였으며, 오류 데이터나 이상치를 갖는 데이터를 삭제하여 최종적으로 62,701명의 고객 데이터와 112,928건의 거래이력 데이터를 사용하였다. <표 3>은 거래 이력 데이터의 구매현황에 대한 기초통계자료이다. 총 16개 장르 중 클래식이나 뮤지컬, 대중음악 등 음악관련 장

<표 3> 공연상품 구매 데이터 요약

	대중 음악	디지털 퍼포먼스	마술쇼 서커스	무용	문화예술 교육	뮤지컬	발레	아동극
10대~20대	0.93% (577)	0.03% (17)	0.35% (215)	0.26% (159)	0.16% (100)	2.34% (1464)	0.84% (523)	1.22% (763)
30대	2.07% (1296)	0.04% (25)	0.29% (178)	0.34% (210)	0.15% (93)	3.61% (2261)	0.72% (450)	0.21% (129)
40대	3.96% (2480)	0.04% (22)	1.13% (705)	0.36% (224)	0.41% (252)	6.02% (3770)	2.26% (1416)	5.15% (3223)
50대	3.61% (2262)	0.02% (12)	0.97% (602)	0.38% (237)	0.47% (294)	3.32% (2081)	1.77% (1104)	1.41% (880)
60대 이상	2.4% (1502)	0.01% (5)	0.14% (85)	0.15% (89)	0.28% (171)	1.04% (646)	0.56% (351)	0.11% (68)
총합계	12.95% (8117)	0.13% (81)	2.85% (1785)	1.47% (919)	1.46% (910)	16.31% (10222)	6.14% (3844)	8.08% (5063)
	아이스쇼	연극	오페라	월드뮤직	재즈	전시	크로스오버	클래식
10대~20대	0.33% (203)	1.04% (648)	0.69% (431)	0.02% (8)	0.22% (133)	0.16% (96)	0.18% (111)	6.14% (3846)
30대	0.1% (57)	1.27% (795)	0.52% (323)	0.05% (29)	0.5% (309)	0.46% (287)	0.15% (93)	3.01% (1886)
40대	0.39% (244)	1.71% (1069)	0.83% (517)	0.08% (50)	0.5% (308)	2.42% (1512)	0.26% (162)	7.13% (4470)
50대	0.9% (561)	1.33% (831)	1.02% (635)	0.14% (82)	0.32% (195)	0.72% (446)	0.24% (147)	12.02% (7534)
60대 이상	0.28% (170)	1.26% (784)	0.52% (326)	0.05% (31)	0.18% (108)	0.15% (88)	0.12% (70)	3.46% (2165)
총합계	16.31% (10222)	6.14% (3844)	8.08% (5063)	1.97% (1235)	6.59% (4127)	3.56% (2232)	0.32% (200)	31.74% (19901)

르가 다른 장르에 비해 선호도가 높게 나타나고 있지만, 특정 장르가 특정 연령대에서만 독점적으로 구매되고 있는 데이터의 편이는 존재하지 않으므로 모델링에 큰 문제가 없다고 판단하였다.

IV. 실험 및 평가

4.1 연구 방법

최종 분석에 사용된 분석용 데이터 셋은 <표 4>와 같이 고객 프로파일의 14개 변수와 거래이력의 4개 변수 등 총 18개의 변수가 사용되었다. 그중 CP와 RP는 원천 데이터베이스에 존재하지 않는 일종의 전략적 파생변수(Strategic Derived Variables)

ables)로써 고객의 행태예측을 더욱 효과적으로 수행하기 위해 개발되었다. CP(Cross Purchase index)와 교차구매 지수로써 각 고객이 구매했던 장르의 다양성을 의미하는 변수이고, RP(Relationship Prematurity)는 관계조속도로써 나이에 비해 해당 공연장과의 거래기간의 비율을 나타내는 변수이다.

$$RP = \frac{Longevity}{Age}$$

본 연구는 아래 <그림 2>의 연구 절차도와 같은 프로세스로 진행되었으며, 전체적인 데이터 전처리와 모형 학습 등의 데이터 분석을 위해 Python ver.3.6.5(The Python Programming Language, 2018)를 이용하였다.

<표 4> 분석용 데이터의 사용 변수

원천 DB	변수명	의미	모형	사용 변수	의미
고객프로파일 DB	ID	회원 아이디	연관성 분석	ID	회원 아이디
	AGE	나이		genre	공연 장르
	RESI_DAY	회원가입일자	협업 필터링	ID	회원 아이디
	longevity	가입기간		genre	공연 장르
	Purchase	구매금액	그라디언트 부스팅	AGE	나이
	Purchase_1y	최근 1년간 구매금액		longevity	가입기간
	Purchase_ave	연간 평균 구매금액		count_all	총 구매횟수
	day_last	마지막 구매일자		RP	관계조속도
	count_ave	연간 평균 구매횟수		R	최근성
	count_all	총 구매횟수		CP	교차구매지수
	RP	관계조속도	랜덤 포레스트	AGE	나이
	R	최근성		longevity	가입기간
	CP	교차구매지수		count_all	총 구매횟수
	Favorite_genre	선호장르		RP	관계조속도
거래이력 DB	ID	회원 아이디		R	최근성
	program	공연 프로그램명		CP	교차구매지수
	place	장소			
	genre	공연장르			



<그림 2> 연구 절차도

데이터 전처리에는 오류 데이터 및 이상치 제거, 데이터 형태의 변환 등을 처리한 후 모형학습과 평가를 위해 학습용 데이터셋과 검증용 데이터셋으로 분리하는 과정을 거쳤다. 고객 프로파일 데이터의 마지막 구매일자를 기준으로 데이터를 분할하여 최근 1년의 데이터 26,722건은 추천 결과의 검증용 데이터셋으로, 그 이전의 데이터 35,979건은 모형 학습용으로 활용하였다. 고객 프로파일과 거래 데이터의 통합을 위해 Pandas 라이브러리의 crosstab 모듈을 활용하여 고객 ID를 기반으로 병합한 후 각 고객의 장르별 구매 여부 및 구매 횟수를 도출하였다. 그 결과로 도출되는 ID-장르 매트릭스는 <그림 3>과 같다.

모형 개발과 학습과정에서 랜덤 추천 시스템은 Random 패키지의 choice 모듈을 활용하여 검증용 데이터 집합의 각 고객에게 16개의 장르 중 3개 장르를 무작위로 선정하여 추천하였고, 연관성 분석 기반 추천 시스템 학습을 위해 ID-장르 매트릭스를 활용하여 mlxtend 패키지의 frequent_patterns 모듈을 통해 연관성 분석을 수행하고 3개의 추천 장르가 선정될 수 있도록 연관 규칙을 추출하였다. 추출된 연관 규칙은 지지도가 높은 순서에 따라 장르별 추천 상품의 우선순위를 결정하여 최종적으로 8개의 추천 규칙이 도출되었다. 한편, 아이템 유사도 기반 추천 시스템은 아이템 기반 협업 필터링을 학습용 데이터의 ID-장르 매트릭스에 적용하여 아이템 유사도를 추출하고, 추출된 아이템 유사도를 기준으로 3개의 추천 장르를 선정하는 방식을 사용하였다. 아이템 유사도는 대표적인 유사도 측정

척도인 코사인 유사도를 활용한 아이템 기반 협업 필터링을 시행하여 계산하였으며(손지은 등, 2015), 결과적으로 도출된 장르 간 아이템 유사도를 기반으로 기준 장르 별 Top 3 추천 장르를 선점함으로써 최종적으로 16개의 규칙을 도출하였다.

마지막으로 앙상블 모형 기반 추천 시스템을 위해 임의추출 방식으로 데이터를 분할하여 70%는 학습용, 30%는 검증용으로 사용하였다. 모형의 독립변수 선정 시 실무에서 모형을 적용할 때에 발생할 수 있는 다중공선성 문제를 해결하기 위해 (Guyon and Elisseeff, 2003; 김상엽 등, 2018) 피어슨 상관계수를 활용하여 변수 간 상관관계를 계산하고, 변수들 간의 상관관계가 높은 경우를 제외함으로써 나이, 가입기간, 관계조속도, 최근성, 교차구매지수, 총 구매 횟수 등이 독립변수로 사용되었다. 앙상블 추천 시스템의 랜덤 포레스트 모형은 학습용 데이터 중 고객 프로파일과 ID-장르 매트릭스의 6개의 변수를 독립변수로, 16개의 장르를 각 모형의 종속변수로 사용하는 분류기로서 지니계수(gini)를 이용하여 트리의 최대 깊이를 15로 제한하였다. 그래디언트 부스팅 모형 역시 16개의 그래디언트 부스팅 모형이 결합된 앙상블 모형으로서 각 모형은 랜덤 포레스트와 동일한 독립변수를 사용하고, 16개의 각 장르를 종속변수로 하는 모형을 학습하였다. 모형의 과적합 문제를 해결하기 위해 학습률과 최대 깊이 등의 옵션 값을 조정하였으며, 모형의 적합성 검증은 랜덤 포레스트와 동일하게 진행하였다. 두 가지 앙상블 모형 기반 추천 시스템의 분석 결과는 각 장르별

ID	대중음악	디지털 퍼포먼스	마술쇼 서커스	전시	크로스 오버	클래식
user_1	1	0	0	1	0	1
user_2	0	0	1	0	1	1
user_3	0	0	0	0	0	0
user_4	0	0	0	1	0	0
user_5	0	1	0	0	1	1
user_6	1	0	0	0	0	0
user_7	0	0	1	0	0	1

<그림 3> ID-장르 매트릭스 예시

<표 5> 추천 시스템 별 도출 규칙수

추천 시스템	도출규칙 수
랜덤 추천 시스템	N/A
연관성 분석 기반 추천 시스템	8
아이템 기반 협업 필터링 추천 시스템	16
랜덤 포레스트 모형 기반 추천 시스템	16
그래디언트 부스팅 모형 기반 추천 시스템	16



적용 추천 시스템	적용 규칙	추천 장르		
		1순위	2순위	3순위
랜덤 추천 시스템	무작위 추천	전시	월드뮤직	문화예술교육
연관성 분석 기반 추천 시스템	{연극}→{클래식, 뮤지컬, 아동극}	클래식	뮤지컬	연극
아이템 기반 협업 필터링 추천 시스템	{연극}→{뮤지컬, 아동극, 대중음악}	뮤지컬	아동극	대중음악
랜덤 포레스트 모형 기반 추천 시스템	장르별 구매 확률값을 통해 도출	뮤지컬 (0.4543)	대중음악 (0.1776)	발레 (0.0982)
그래디언트 부스팅 모형 기반 추천 시스템	장르별 구매 확률값을 통해 도출	뮤지컬 (0.8432)	재즈 (0.6127)	발레 (0.6079)

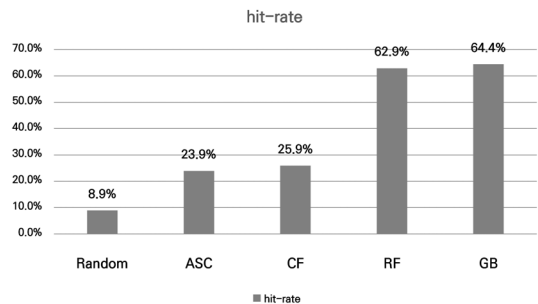
* ()는 도출된 확률값을 뜻함

<그림 4> 추천 시스템 적용 결과

구매 확률 값으로 도출된다. <표 5>는 유형별 각 추천 시스템의 추천 규칙 개수를 나타낸다.

마지막으로 추천 시스템의 성능을 평가하기 위해 5개 추천 시스템으로부터 도출된 규칙을 통해 추천된 상위 3개 추천 장르가 검증용 데이터 셋에 실제로 구매되었는지를 판단하여 추천 적중률(hit-rate)을 계산하여 비교하였다. <그림 4>는 유형별 추천 시스템을 특정 고객에게 실제로 적용했을 때 사용된 추천규칙과 그에 따라 도출된 상위 3개 추천 장르를 예시하고 있다.

시스템을 의미한다.



<그림 5> 추천 적중률

4.2 실험 결과 비교 및 분석

공연문화예술 업종을 위한 공연장르 개인화 추천을 위해 5개의 추천 시스템을 적용한 추천 적중률은 <그림 5>와 같다. 그림에서 Random은 랜덤 추천 시스템, ASC는 연관성 분석 기반 추천 시스템, CF는 아이템 기반 협업 필터링 추천 시스템, RF는 랜덤 포레스트 모형 기반 추천 시스템, 그리고 GB는 그래디언트 부스팅 모형 기반의 추천시

분석 결과, 랜덤 추천 시스템은 8.9%, 연관성 분석 기반 추천 시스템은 23.9%, 아이템 기반 협업 필터링 추천 시스템은 25.9%, 랜덤 포레스트 모형 기반 추천 시스템은 62.9%, 그리고, 그래디언트 부스팅 모형 기반의 추천시스템은 64.4%의 추천 적중률을 보였다. 즉, 5가지 추천 시스템 상황에서 랜덤 추천 시스템의 성능이 가장 떨어지고, 두 가지 단일 추천 모형인 연관성 분석 모형과 협업 필

터링의 성능이 조금 더 우수하고, 두 가지 앙상블 모형인 랜덤 포레스트와 그래디언트 부스팅의 추천 성능이 가장 우수한 것으로 나타났다.

이러한 분석결과는 공연문화예술 마케팅 환경하에서도 무작위 혹은 편의적 추천방식 보다는 머신러닝 기법에 의한 추천 방식이 보다 효율적인 마케팅 성과를 기대할 수 있음을 의미하고, 다른 산업에 적용되었던 기존 연구들(Pazzani, 1999; 이원철, 정석봉, 2017; 최예립 등, 2015)과 마찬가지로 공연문화예술 분야에서도 앙상블 기반의 추천 시스템이 가장 효과적일 수 있다는 의미를 내포하고 있다.

V. 결 론

공연문화예술 시장의 확대에도 불구하고, 마케팅 역량이 충분하지 않은 중소규모의 공연장의 경우 오히려 판매실적이 악화되고 있는 실정이다. 그럼에도 불구하고 문화예술 분야의 마케팅 연구는 전통적인 시장 세분화나 소비자 분석, 또는 구매 요인 분석 등에만 집중되어 중소규모 공연장의 마케팅 실무에 즉각적인 활용이 요원한 상태이다. 본 연구는 이러한 문제점을 해소할 수 있는 하나의 방안으로 공연 데이터의 소비자 정보 접근성을 개선함으로써 소비자의 공연관람 성과를 높이고자 머신러닝 기반의 장르 추천 시스템을 제시하고자 하였다. 이를 위해 경기도 소재의 실제 공연장의 고객 마스터 DB와 거래이력 DB를 활용하여 5개의 추천 시스템을 적용하고, 각 추천 시스템으로부터 도출된 규칙을 기반으로 고객에게 최대 3개의 공연 장르를 추천한 뒤 1년 동안의 실제 공연구매 이력을 바탕으로 추천 모형의 성능을 검증하였다. 연구결과 공연문화예술 분야에도 머신러닝 기반의 추천 시스템 활용 효과가 충분하다는 것을 검증하였고, 분석결과 다양한 추천 시스템 중 다른 산업과 유사하게 앙상블 모형이, 특히 그래디언트 부스팅 모형 기반의 추천 시스템이 가장 우수한 성능을 보이는 것을 확인하였다.

본 연구의 학술적 의의는 첫째 개인화 추천 기

법 연구에 많이 적용되었던 연관규칙 기반 추천 시스템, 협업 필터링 기반 추천 시스템, 앙상블 기반 추천 시스템 등 현재 가용한 머신러닝 기반 개인화 추천 모형을 종합적으로 분석하여 비교하였고, 그 동안 관련분야 연구가 일천했던 공연문화예술 분야에 적용하여 실용적 가치를 증명했다는 점이다. 따라서, 공연상품에 대한 정보 접근성 문제를 가지고 있는 중소 규모의 공연장에서는 머신러닝 기반의 개인화 추천 시스템 모듈을 자사의 웹/모바일 환경에 탑재함으로써 수동적인 정보제공 환경에서 벗어나 능동적인 1:1 마케팅을 실현할 수 있는 시스템적, 전략적 체도를 적극적으로 수행할 필요가 있을 것이다. 둘째, 개인의 니즈에 맞는 공연장르를 추천하기 위해 관계조속도와 같은 제3의 변수를 개발하였고, 추천 모형에서 유의미한 독립변수 역할을 하고 있다는 것을 발견했다는 점이다. 이러한 전략적 파생변수의 개발은 일반적으로 중소 규모의 공연장에서 획득할 수 있는 고객 정보의 한계를 고려했을 때 새로운 고객 통찰력을 확보할 수 있는 중요한 실무적 방안으로 고려해볼 수 있을 것이다.

그러나, 본 연구의 한계점도 존재한다. 우선, 추천 시스템의 모형에서 다양한 인구통계적 변수가 투입되지 못했고, 소비자의 실제 구매가 이루어진 후 수집되는 구매패턴 정보들 위주로 모형이 생성되었다는 한계가 있다. 실제 공연장의 회원가입 시 고객이 기재하는 고객정보가 단순히 이름과 나이, 연락처 정도만을 요구하기 때문에 현실적으로 불가피한 상황이지만, 이 경우 신규고객에 대한 장르 추천은 불가능하거나 무리한 추정을 해야 하는 한계를 갖는다. 따라서, 추후 연구에는 온라인 로그정보를 부가적으로 활용하여 신규고객에 대해서도 적용할 수 있는 개인화 추천모델이 개발되는 것이 바람직하다. 둘째, 본 개인화 추천 시스템의 추천 대상물은 개별 공연상품이 아닌 장르로 일반화된 카테고리 수준이라는 점이다. 대부분의 추천 연구가 품목 단위로 개인화 추천을 수행하지만, 이럴 경우 실무적으로 활용할 때 추천결과에 대해

다시 상세한 상품정보를 매핑해야 하는 부가적인 작업이 필요하다는 것을 의미한다. 향후에는 구체적인 공연상품에 대한 속성 정보를 활용하여 조합적(Constructive) 방식에 의한 실제 상품추천이 가능한 개인화 추천 모형이 개발되기를 기대한다.

참고문헌

- [1] 강호윤, 옥창수, “아이템 연관성을 고려한 협력적 필터링 기반 추천시스템 개발”, *Entrue Journal of Information Technology*, 제14권, 제1호, 2015, pp. 135-144.
- [2] 김민정, 조윤희, “빅데이터 기반 추천시스템 구현을 위한 다중 프로파일 앙상블 기법”, *지능정보연구*, 제21권, 제4호, 2015, pp. 93-110.
- [3] 김범석, 전인오, “공연예술 선택 특성과 서비스 품질이 사후 행동에 미치는 영향”, *한국콘텐츠학회논문지*, 제15권, 제12호, 2015, pp. 106-120.
- [4] 김상엽, 박경섭, 이상민, 허병문, 류근호, “머신러닝 기반의 온실 제어를 위한 예측모델 개발”, *Journal of Digital Contents*, 제19권, 제4호, 2018, pp. 749-756.
- [5] 김선영, “빅데이터, 공연예술 관객 개발의 가능성 연다”, *예술경영지원센터*, 2018, Available at http://www.gokams.or.kr/webzine/wNew/column/column_view.asp?idx=1992&page=3&c_idx=48&searchString=&c_idx_2=.
- [6] 김소영, 신혜선, “뮤지컬 관객의 시장세분화 연구”, *한국콘텐츠학회논문지*, 제6권, 제5호, 2006, pp. 131-144.
- [7] 김제관, “[view point] 뮤지컬 시장 공급 과잉 심각”, *매일경제*, 2013.08.06., Available at <http://www.mk.co.kr/news/culture/view/2013/08/682092/>.
- [8] 김현정, “무용공연 활성화를 위한 공연문화 관람객의 소비성향분석”, *한국사회체육학회지*, 제60권, 2015, pp. 261-274.
- [9] 문화체육관광부, 2018 *문화향수실태조사*, 문화체육관광부, 2018.
- [10] 문화체육관광부, *예술관객 확대를 위한 비관객 세분화 전략*, 문화체육관광부, 2016.
- [11] 박두순, “개인성향과 협업 필터링을 이용한 개선된 영화 추천 시스템”, *정보처리학회논문지 컴퓨터 및 통신시스템*, 제2권, 제11호, 2013, pp. 475-482.
- [12] 박지희, 남기환, “추천 수량과 재 추천을 고려한 사용자 기반 협업 필터링 추천 시스템”, *Information Systems Review*, 제19권, 제2호, 2017, pp. 71-94.
- [13] 박철용, “랜덤포레스트의 크기 결정을 위한 간편진단통계량”, *한국데이터정보과학회지*, 제27권, 제4호, 2016, pp. 855-863.
- [14] 손예지, “[공연계의 오늘] ① ‘K-공연’ 선도하는 대극장”, *헤럴드경제*, 2019.02.22., Available at http://biz.heraldcorp.com/culture/view.php?ud=201902221303096415290_1.
- [15] 손지은, 김성범, 김현중, 조성준, “추천 시스템 기법 연구동향 분석”, *Journal of the Korean Institute of Industrial Engineers*, 제41권, 제2호, 2015, pp. 185-208.
- [16] 안현철, 한인구, 김경재, “연관규칙기법과 분류모형을 결합한 상품 추천 시스템: G 인터넷 쇼핑몰의 사례”, *Information Systems Review*, 제8권, 제1호, 2006, pp. 181-201.
- [17] 예술경영지원센터, 2018 *공연예술트렌드 조사 보고서*, 예술경영지원센터, 2018.
- [18] 유은영, 진현정, “의사결정나무모형을 이용한 공연예술시장의 소비자그룹 분석”, *소비자학연구*, 제25권, 제6호, 2014, pp. 65-91.
- [19] 이동원, “연관상품 추천을 위한 회귀분석모형 기반 연관 규칙 척도 결합기법”, *지능정보연구*, 제23권, 제1호, 2017, pp. 127-141.
- [20] 이연정, 김경재, “다중모형조합기법을 이용한 상품추천시스템”, *지능정보연구*, 제19권, 제2호, 2013, pp. 39-54.

- [21] 이원철, 정석봉, “다중프로파일 앙상블기법을 이용한 금융상품 추천에 관한 연구”, *한국경영학회 통합학술발표논문집*, 2017, pp. 681-696.
- [22] 이재식, 박석두, “장르별 협업필터링을 이용한 영화추천시스템의 성능 향상”, *지능정보연구*, 제13권, 제4호, 2007, pp. 65-78.
- [23] 이택승, *취업사교육 콘텐츠 시장에 효율적인 추천시스템 구축연구* (석사학위논문), 연세대학교 정보대학원, 2018.
- [24] 정순규, *관객 유형에 따른 공연 관람 구매 의사와 관객 개발전략* (석사학위논문), 중앙대학교 대학원, 2013.
- [25] 정영진, 조윤호, “온라인 구매 행태를 고려한 토픽 모델링 기반 도서 추천”, *지식경영연구*, 제18권, 제4호, 2017, pp. 97-118.
- [26] 조남규, “한국 공연예술시장의 현황과 전망”, *한국무용기록학회지*, 제18권, 2010, pp. 79-97.
- [27] 지숙영, 이중식, 김은미, 백영민, “한국의 공연 문화 소비 지형의 기술”, *예술경영연구*, 제20권, 2011, pp. 209-231.
- [28] 최예림, 박종현, 신동완, “모바일 사용자의 잠재 관심 추론을 위한 앙상블 기법”, *정보과학회 컴퓨팅의 실제 논문지*, 제21권, 제11호, 2015, pp. 706-712.
- [29] Adomavicius, G. and Y. Kwon, “Improving aggregate recommendation diversity using ranking-based techniques, knowledge and data engineering”, *IEEE Transactions on*, Vol.24, No.5, 2012, pp. 896-911.
- [30] Aggarwal, C. C., Z. Sun, and P. S. Yu, “Online algorithms for finding profile association rules”, *The Seventh International Conference on Information and Knowledge Management*, 1998, pp. 86-95.
- [31] Agrawal, R., T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases”, *ACM SIGMOD International Conference on Management of Data*, Vol.22, No.2, 1993, pp. 207-216.
- [32] Andreassen, A. R., *Expanding the Audience for Performing Arts*, Seven Lock, Washington, D.C. 1990.
- [33] Guyon, I. and A. Elisseeff, “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, 2003, Vol.3, pp. 1157-1182.
- [34] Kangun, N., G. Otto, and D. C. Randall, “Marketing strategies for increasing symphony season ticket purchases among University students”, *Journal of Cultural Economics*, Vol.16, No.1, 1992, pp. 25-40.
- [35] Kim, H. K., M. K. Jang, J. K. Kim, and Y. H. Cho, “A new item recommendation procedure using preference boundary”, *Asia Pacific Journal of Information Systems*, Vol.20, No.1, 2010, pp. 81-99.
- [36] Kolb, B. M., *Marketing for Cultural Organizations: New Strategies for Attracting to Classical Music, Dance, Museums, Theater and Opera*, Dublin: Oak Tree Press, Dublin, 2000.
- [37] Kotler, P. and G. Armstrong, *Principles of marketing* (14th ed.), Prentice Hall, New Jersey, 2012.
- [38] Pazzani, M. J., “A framework for collaborative, content-based, and demographic filtering”, *Artificial Intelligence Review*, Vol.13, No.5-6, 1999, pp. 393-408.
- [39] The Python Programming Language, 2018, Available at <https://www.python.org>.
- [40] Vozalis, M. and K. G. Margaritis, “Collaborative filtering enhanced by demographic correlation”, *The AIAI Symposium on Professional Practice in AI, Part of the 18th World Computer Congress*, 2004, pp. 393-402.
- [41] Zhang, H. R. and F. Min, “Three-way recommender systems based on random forests”, *Knowledge-Based Systems*, Vol.91, 2016, pp. 275-286.

A Personalized Recommendation System Using Machine Learning for Performing Arts Genre

Hyung Su Kim* · Yerin Bak** · Jeongmin Lee***

Abstract

Despite the expansion of the market of performing arts and culture, small and medium size theaters are still experiencing difficulties due to poor accessibility of information by consumers. This study proposes a machine learning based genre recommendation system as an alternative to enhance the marketing capability of small and medium sized theaters. We developed five recommendation systems that recommend three genres per customer using customer master DB and transaction history DB of domestic venues. We propose an optimal recommendation system by comparing performances of recommendation system. As a result, the recommendation system based on the ensemble model showed better performance than the single predictive model. This study applied the personalized recommendation technique which was scarce in the field of performing arts and culture, and suggests that it is worthy enough to use it in the field of performing arts and culture.

Keywords: *Performing Arts, Machine Learning, Recommendation System, CRM, Association, Collaborative Filtering, Ensemble*

* Corresponding Author, Division of Smart Management Engineering, Hansung University

** Department of International Trade, Hansung University

*** Division of Knowledge and Information, Hansung University

◎ 저 자 소 개 ◎



김 형 수 (hskim@hansung.ac.kr)

현재 한성대학교 스마트경영공학부에서 부교수로 재직 중이며, 산학협력 CRM 연구 및 컨설팅 기관인 고객경영기술원의 원장을 겸임하고 있다. 한국과학기술원에서 CRM으로 경영공학 박사학위를 받았으며, 한국CRM협회 회장과 한국CRM 학회의 부회장을 역임한 바 있다. 주요 관심분야는 CRM, 멤버십 프로그램, 빅데이터분석, 비즈니스 머신러닝 모델 개발 등이다.



박 예 린 (yyam1020@gmail.com)

한성대학교 무역학과에서 경영학사 학위, CRM·디지털마케팅 연계전공에서 CRM 학사 학위를 취득하였다. 주요 관심분야는 머신러닝, 고객관계관리(CRM), 빅데이터, 공연문화예술 마케팅, 데이터마이닝 등이다. 빅데이터와 머신러닝을 활용한 마케팅 기획과 고객관계관리(CRM) 전략 수립에 관심을 가지고 있다.



이 정 민 (jenna4177@naver.com)

한성대학교 지식정보학부에서 도서관학사 학위, 경제학과에서 경제학사 학위, CRM·디지털마케팅 연계전공에서 CRM 학사 학위를 취득하였다. 주요 관심 분야는 빅데이터 분석, 데이터마이닝, 고객관계관리 마케팅, 추천 시스템 등이다. 데이터마이닝을 이용해 여러 문제를 해결하고, 효과적인 마케팅 및 경영전략을 수립하는데 관심을 가지고 있다.

논문접수일 : 2019년 07월 08일

게재확정일 : 2019년 08월 22일

1차 수정일 : 2019년 08월 05일