

Data Mining Approach to Predicting Serial Publication Periods and Mobile Gamification Likelihood for Webtoon Contents

Hyun Seok Jang*, Kun Chang Lee**

Abstract

This paper proposes data mining models relevant to the serial publication periods and mobile gamification likelihood of webtoon contents which were either serialized or completed in <Naver Webtoon> platform. The size of the cartoon industry including webtoon takes merely 1% of the total entertainment contents industry in Korea. However, the significance of webtoon business is rapidly growing because its intellectual property can be easily used as an effective OSMU (One Source Multi-Use) vehicle for multiple types of contents such as movie, drama, game, and character-related merchandising. We suggested a set of data mining classifiers that are deemed suitable to provide prediction models for serial publication periods and mobile gamification likelihood for the sake of webtoon contents. As a result, the balanced accuracies are respectively recorded as 85.0% and 59.0%, from the two models.

▶ Keyword: Webtoon, OSMU, Serial Publication Periods, Mobile Gamification Likelihood, Data Mining Models

I. Introduction

2015년 기준 국내 만화산업 매출액은 약 9,194억원을 기록했다. 동 기간 게임산업 매출액 10조7,223억원, 캐릭터산업 10조807억원, 영화산업 5조1,122억원[1]에 비하면 작은 규모이지만 만화가 OSMU(One Source Multi Use)로서 영화, 드라마, 게임, 광고, 캐릭터 상품화 등에 활용되는 부가가치를 감안할 때 콘텐츠 산업 전반의 베이스 캠프(Base Camp)와 같은 역할을 담당하고 있다고 할 수 있다.

만화를 기반으로 한 영화에 대하여 해외의 경우, 마블코믹스의 만화를 영화화한 <어벤저스(The Avengers)>와 <캡틴 아메리카: 시빌 워(Captain America: Civil War)>가 2012년과 2016년 전세계 박스오피스 순위 1위를 차지한 바 있다[2].

국내의 경우, <다음 웹툰>에서 연재된 윤태호 작가의 작품들 가운데 영화화된 2015년 개봉작 <내부자들>이 극장 관객

수 705만명을 기록하였으며, 최중훈 작가의 2013년 개봉작 <은밀하게 위대하게>는 696만명을 기록한 바 있다.

주호민 작가의 <신과 함께>는 순제작비 350억원(1,2편 합산)의 블록버스터로 완성되어 2017년 12월 개봉하여 41일만에 극장 관객 수 1400만명을 돌파하였다[3].

웹툰 원작은 영화 뿐만 아니라 드라마에서도 폭넓게 활용되고 있다. 2014년 드라마화된 윤태호 작가의 <미생>은 비정규직 인턴의 이야기를 다루어 비정규직 문제에 대한 정부 대책이 논의되는 등 사회적인 영향을 가져왔다. 이후 2015년 <송곳(JTBC, 최규석 원작, 네이버)>, <냄새를 보는 소녀(SBS, 서수경 원작, 올레마켓)>, 2016년 <치즈인더트랩(tvN, 순끼 원작, 네이버)>, <마음의 소리(KBS, 조석 원작, 네이버)>, 2017년 <부암동 복수자들(tvN, 사자토끼 원작, 다음)>, <고백부부(KBS, 밤토끼 원작, 네이버)>

• First Author: Hyun Seok Jang, Corresponding Author: Kun Chang Lee

*Hyun Seok Jang (hitmank@naver.com), BigRadar Co. Ltd., Republic of Korea

**Kun Chang Lee (kunchanglee@gmail.com), SKK Business School/SAIHST (Samsung Advanced Institute for Health Sciences & Technology)/Health Mining Research Center, Sungkyunkwan University, Seoul 03063, Republic of Korea

• Received: 2018. 01. 04, Revised: 2018. 01. 30, Accepted: 2018. 02. 26.

• This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP: Ministry of Science, ICT & Future Planning) (No. 2017R1A2B4010956).

등이 제작되어 히트를 기록한 바 있다[4].

모바일 게임의 경우, 2014년 <와라 편의점 for kakao(뉴에프오)>, 2015년 <갓 오브 하이스쿨(YD온라인)>, 2016년 <마음의 소리(네오위즈)> 등이 각각 100만 다운로드 이상을 기록하였으며 이외 다수의 웹툰 원작 게임들이 개발 중에 있다[5].

국내 웹툰들의 해외 진출도 활발하다. 만화 시장 규모가 연간 5조원이 넘는 일본에서 2017년 11월 기준 애플 앱스토어에서 가장 매출을 많이 올린 웹툰 앱 7개 중 4개가 <라인>, <카카오>, <NHN코미코> 등의 국내기반 회사들이었다[6].

특히 <네이버 웹툰>의 글로벌 서비스 플랫폼인 <라인 웹툰>은 라인 메신저를 많이 쓰는 일본과 동남아 등을 중심으로 성장하여 월간 사용자 수가 4,000만명에 이르고 있다[7].

웹툰은 부담없이 즐기는 스낵 컬처(Snack Culture)를 대표하는 콘텐츠이지만 영화, 드라마, 모바일 게임 등 다른 콘텐츠로의 확장이 용이하고, 일본, 동남아 및 중국 등 해외 진출이 가속화되고 있으므로 웹툰 콘텐츠를 심층적으로 분석하여 통찰된 결과를 활용한다면 다양한 부가 가치를 기대할 수 있을 것이다.

본 연구의 연구 대상으로 선정한 <네이버 웹툰>은 2004년 6월에 오픈하여 2017년 11월 기준 모바일 앱 주간 이용자 수가 약 432만명으로 2위 업체인 <다음 웹툰>의 90만명을 훨씬 상회하는 독점적 플랫폼이다[8]. 이용자 데모그래픽은 남성 55%, 여성 45%이며, 10대 26%, 20대 45%로서 10대~20대가 전체 트래픽의 71%를 차지하고 있다[9]. 이처럼 <네이버 웹툰>은 10대~20대의 문화를 선도하는 ‘젊은’ 콘텐츠로서도 대표성을 확보하고 있으므로 <네이버 웹툰>의 콘텐츠 특징과 이용자들의 소비 특징을 연구한 결과는 국내 웹툰 산업의 발전 뿐만 아니라 10대~20대가 선호하는 콘텐츠의 특징을 이해하는 데에도 도움이 될 것이다.

본 연구는 <네이버 웹툰>의 연계 데이터를 데이터 마이닝 기법으로 분석하기 위해 아래와 같이 두 가지 연구 질문(RQ : Research Question)을 설정하였다.

RQ 1: 연계 초기의 독자 반응을 측정하여 연재의 지속 여부를 예측할 수 있는가?

RQ 2: 과거 모바일 게임화된 웹툰들의 특징을 분석하여 모바일 게임으로 확장하기에 적합한 웹툰을 선별할 수 있는가?

본 연구에 사용된 데이터는 <네이버 웹툰>에서 연재 중이거나 완결, 또는 휴재된 웹툰 가운데 기업의 후원을 받은 브랜드 웹툰 및 이벤트성 기획의 릴레이 웹툰을 제외한 302편을 대상으로 하였다.

본 연구의 구성은 다음과 같다. 2장에서는 웹툰에 관한

기존 선행 연구를 소개하고 3장에서는 변수 및 분석 기법에 대해 서술한다. 4장에서는 구현된 모델을 토대로 구체적인 예측 결과를 제시한다. 5장에서는 본 연구의 시사점과 의의를 서술한다.

II. Preliminaries

웹툰에 관한 선행 연구는 크게 세가지 방향으로 진전되었다. 첫 번째 방향은 웹툰 플랫폼에 대한 산업적 연구이다.

김재필, 성승창(2015)은 국내 웹툰 시장이 1,000억원 규모로 발전한 원인을 웹툰 플랫폼으로부터 찾았다. 네이버와 다음이 서비스하는 웹툰 플랫폼은 광고주와 독자를 연결하여 연재 작가들에게 안정적인 수입원을 제공하였고 음악, 동영상, 게임, 커머스 등 OSMU(One Source Multi Use)로의 확장을 추진하였으며 예비작가를 위한 인큐베이팅 시스템 등을 마련함으로써 전세계 유례없는 한국 고유의 생태계 모델을 구축하였다고 분석했다[10, 11]. 한창완(2016)은 웹툰 플랫폼이 포탈로부터의 분사, 타겟층을 전문화한 프리미엄 웹툰 서비스의 등장 등 스피드 오픈되고 있는 현상에 주목하며 웹툰 생태계가 실시간으로 혁신 및 확장 중에 있으므로 창작자의 기획창작모델과 독자의 소비모델간의 연계 효과 및 진화 모델 연구의 필요성을 제기하였다[12].

두 번째 방향은, 웹툰의 스토리텔링 및 타 매체로의 전환에 대한 콘텐츠적 연구이다.

한창완, 홍난지(2014)는 웹툰 <이끼>의 영화화 사례를 중심으로 웹툰의 영화 매체 전환에 대해 연구하였다. 연구자는 웹툰을 영화 매체로 전환하는 것에 대한 장점을 독자(프로슈머)들의 활발한 참여, 시장 검증력, 영화의 스토리보드를 연상시키는 스크롤 연출에 있다고 보았다. 또한 영화 매체에 맞는 압축적인 전개를 위해 웹툰 <이끼>의 복합적인 구성을 장르적인 특성 위주로 강조하고 캐릭터군을 단순화한 스토리텔링 구조로 변환한 것을 흥행 성공의 요인으로 분석하였다[13].

하주영, 박진완(2017)은 웹툰의 내러티브가 모바일 게임 형태에서 어떻게 적용되는지를 RPG장르 모바일 게임 <신과 함께>의 사례를 중심으로 연구하였다. 연구자들은 유명 웹툰의 모바일 게임화에 대한 장점으로 인지도를 기반으로 한 마케팅과 프로모션의 유리함, 기획 단계의 개발 기간 단축, 스토리 및 캐릭터의 재창조를 통한 흥미 유발 등을 제시했다. 그러나 <신과 함께> 웹툰의 평점이 9.9점(10점 만점)인데 반해 모바일 게임의 평점은 2.8점(5점 만점)에 불과한 점에 주목하여 웹툰이 모바일 게임으로서 새로운 가치를 창출하기 위해서는 게임의 완성도, 서비스의 관리 수준 등이 동반되어야 함을 제언했다[5].

세 번째 방향은 웹툰에 대한 정량적인 연구이다.

윤기현외 3인(2015)은 2001년부터 13년간 2014년 1월 1일까지 온라인 매체에서 원고료가 지급된 웹툰 총 1,928개 작품을 대상으로 통계 분석을 실시하였다. 연구자들은 기간 중 웹툰의 연재 기간이 평균 1년 정도이고, 장르는 드라마(478편), 개그(349편), 판타지(245편) 순으로 분포되어 있음을 확인하였다[14].

최근 들어 웹툰에 대한 정량적인 연구에 대한 시도가 증가하고 있고 분석 대상인 웹툰의 표본 크기도 점차 확대되고 있어

데이터 마이닝, 머신 러닝 등 빅데이터에 적합한 분석 기법을 적용한 연구의 필요성이 부각되고 있다.

III. The Proposed Scheme

3.1 Data and Variables

본 연구는 <네이버 웹툰>에서 2018년 1월 현재 연재 중이거나 완결, 혹은 휴재된 웹툰 302편을 대상으로 하였다.

3.1.1 Dependent Variables

연재 지속 여부

웹툰이 연재를 완결하거나 휴재 기간을 가지는 대부분의 경우에는 스토리의 완결, 소재의 고갈, 건강 혹은 개인 신상 등의 이유를 게시한다.

하지만 플랫폼으로부터 주1회 연재기준 월 최소 원고료 100만원~200만원 및 PPS(Page Profit Sharing, 웹툰 페이지에 광고를 게재하고 그 수익을 플랫폼과 작가가 나누는 방법) 프로그램을 통한 수익을 포기하면서 작가 스스로 완결 혹은 휴재를 결정하기는 쉽지 않을 것이다.

네이버 웹툰에 연재하기 위한 아마추어들의 등단 코너인 <도전 만화>에서 활동하는 작가 수는 14만명 정도로 정식 연재를 위한 경쟁 또한 치열하다. 웹툰 플랫폼의 경우, 작가마다 다른 계약조건을 제시하지만 대부분은 트래픽을 기준으로 3개월 단위로 원고료 협상을 하여 계약을 갱신한다[16]. 따라서 연재 종료의 실질적인 이유는 저조한 트래픽으로 인해 계약을 종료하거나 수익 조건이 악화되었기 때문으로 보는 것이 합리적인 추정일 것이다.

그러므로 연재의 지속 여부는 웹툰의 인기도와 밀접한 관계가 있을 것으로 추정하여 첫 번째 종속 변수로 결정하였다.

총 302편의 관찰 데이터에서 연재 지속(Positive) 중인 웹툰은 140편(46.4%)이고, 완결 혹은 휴재된 웹툰은 162편(53.6%)으로서 클래스간 균형적인 구조를 보이고 있다.

모바일 게임화 가능성 여부

웹툰이 영화나 드라마의 형태로 변환될 때에는 기존 웹툰 사용자는 다소 수동적인 독자(reader)나 관람자(viewer)의 지위를 유지한다. 그러나 모바일 게임과 같은 상호작용성이 특화된 매체로 변환될 경우, 이용자(User) 혹은 플레이어(Player)로서 보다 능동적이고 적극적인 참여를 도모하는 지위로 상승한다[17]. 즉, 웹툰의 모바일 게임화에는 영화화, 혹은 드라마화와는 다른 접근법과 내러티브가 필요하며 웹툰 수용자 또한 콘텐츠에 대해 보다 충성적이고 높은 관여 성향을 갖고 있어야 성공적인 협업(Collaboration)을 추구할 수 있다.

관찰 데이터에서 모바일 게임으로 출시되거나, 기존 모바일 게임에 아이템, 혹은 이벤트 형태로 제휴된 웹툰, 또는 모바일 게임으로 제작 중인 웹툰은 총 21편(7.0%)이다. 이와 같이 심

각한 클래스 불균형을 보이는 자료형의 경우 분류기가 가장 빈번한 클래스로만 추정하게 되면 높은 정확도를 쉽게 얻을 수 있기 때문에 이를 보정하는 균형정확도(Balanced Accuracy)나 카파 통계량(Kappa Statistic)의 수치가 특히 중요하다.

3.1.2 Independent Variables

유통화 모델 도입 여부

어떤 웹툰들은 미리 3회차 가량을 제작 완료하여 회차별 유료(300원) 서비스로 운영한다. 이른바 ‘기다리면 무료’라는 수익 모델인데 웹툰을 한 회차 구독한 다음 일정 시간이 지나면 다음 회차를 무료로 이용할 수 있는 서비스다[6]. <네이버 웹툰>에서는 이를 프라임 서비스(Prime Service)라고 한다. 프라임 서비스를 적용한 웹툰은 전체의 34.4%인 104편이다.

작가 경력

작가가 현재까지 네이버에서 연재했던 작품 수이다. 관찰 데이터에서는 264명의 작가가 평균 3.9편의 경력을 보유하고 있다. 모델링시 정규화(Normalization)하여 사용하였다.

서사 구조

웹툰은 스토리, 에피소드, 유니버스로 서사 구조를 분류한다.

- 스토리 : 주인공이 하나의 이야기를 이끄는 드라마 형식
- 에피소드 : 주인공이 여러 이야기에 나오는 시트콤 형식
- 유니버스 : 각 회마다 다른 주제나 인물이 등장하여 이야기를 풀어가는 형식

장르

액션, 코믹, 일상, 드라마, 판타지, SF, 시대물, 공포, 스릴러, 순정, 학원, 스포츠 등 12개 장르로 내용을 구분하였다. 위 서사구조, 장르와 연재 요일 등 명목형의 경우 더미(Dummy)로 변환하여 사용하였다.

별점 점수

독자들은 각 회별로 1점부터 10점까지 별점을 매길 수 있다. 네이버 웹툰 독자들의 별점은 후한 편으로 평균 별점은 9.7점이다.

별점 참여 수

각 회별 별점에 참여한 인원 수이다.

하트 참여 수

독자들은 각 회별로 ‘하트’를 줄 수 있다. 페이스북의 ‘좋아요’와 유사한 파라미터이다.

댓글 수

독자들은 500자 이내로 댓글을 남길 수 있다. 댓글 가운데 독자들의 ‘하트’ 지지를 많이 받은 댓글은 ‘베스트 댓글(베댓)’로 선정되어 첫 페이지에 최대 15개까지 등록된다.

별점 점수, 별점 참여 수, 하트 참여 수, 댓글 수 등은 모델링시 정규화하였다.

매주마다 업데이트되는 연재의 특성상 첫 회의 반응과 함께 독자들의 지속성 여부도 반영이 필요하다. 이를 위해 4회차, 8회차 및 12회차와 1회차간 변화율을 측정하여 변수로 사용하였다.

3.1.3 Moderator Variables

연재 기간

관찰 데이터로부터 실제 예측을 위한 예측 데이터 셋 (Prediction Set)을 분리하기 위해 연재 기간을 매개 변수로 활용하였다.

연재 기간의 경우, 연재 시작 3개월마다 계약 갱신이 가능함을 고려하여 24주차를 예측 셋 분류 기준으로 하였다.

모바일 게임화 가능성 여부는 기존 웹툰의 모바일게임화 사례를 분석하여 240주차 이상을 예측 셋 분류 기준으로 하였다.

관찰 데이터의 평균값은 118주이며 중간값은 86주이다. 최소값은 10주이며, 최대값은 593주의 범위에 있다.

Table 1. Variable Descriptions and Dependencies

No	Independent Variables	Dependent Variables	
		Serialization	Gamification
1	Prime Service	0.75	0.10
2	Writer's Career	-0.07	0.05
3	Narrative_Story	0.18	-0.04
4	Narrative_Episode	-0.09	0.03
5	Narrative_Omnibus	-0.16	0.01
6	Genre_Action	0.17	0.12
7	Genre_Comic	-0.06	0.15
8	Genre_Daily	-0.12	-0.08
9	Genre_Drama	-0.11	-0.12
10	Genre_Fantasy	0.09	0.15
11	Genre_Historic	-0.07	-0.04
12	Genre_Horror	0.01	-0.04
13	Genre_Romance	0.09	-0.08
14	Genre_School	0.07	-0.05
15	Genre_Sf	0.03	0.03
16	Genre_Sports	-0.08	-0.04
17	Genre_Thriller	-0.03	-0.06
18	Weekday_sun	0.06	-0.06
19	Weekday_mon	0.01	0.01
20	Weekday_tue	0.01	0.06
21	Weekday_wed	-0.07	-0.01
22	Weekday_thu	-0.04	-0.04
23	Weekday_fri	0.09	0.03
24	Weekday_sat	-0.05	0.01
25	Ep1_Score	0.07	-0.10
26	Ep1_ScoreJoin	0.29	0.20
27	Ep1_HeartJoin	0.49	0.13
28	Ep1_OpJoin	0.10	0.29
29	Ep4_Score_Rate	0.05	0.09
30	Ep4_ScoreJoin_Rate	-0.14	0.01
31	Ep4_HeartJoin_Rate	-0.08	-0.01
32	Ep4_OpJoin_Rate	-0.10	-0.03
33	Ep8_Score_Rate	0.04	0.14
34	Ep8_ScoreJoin_Rate	-0.09	0.07
35	Ep8_HeartJoin_Rate	-0.14	0.01
36	Ep8_OpJoin_Rate	-0.02	0.02
37	Ep12_Score_Rate	0.03	0.12
38	Ep12_ScoreJoin_Rate	-0.08	0.05
39	Ep12_HeartJoin_Rate	-0.12	-0.02
40	Ep12_OpJoin_Rate	-0.04	0.01

Table 1은 독립 변수와 종속 변수 간의 상관 관계를 나타낸 표이다. 연재 지속 여부에 밀접하게 관여하는 변수는 유료화 모델 도입 여부로서 0.75의 높은 상관값을 보였다. 이외에도 첫 회의 유저 반응, 스토리 서사 구조, 액션 장르 등의 변수가 연재

지속에 긍정적인 영향을 주고 있다. 움니버스 서사 구조의 경우에는 연재 지속에 약한 부정 영향을 주는 것으로 나타났다.

모바일 게임화 가능성 여부의 경우, 첫 회의 유저 반응이 일정한 긍정 영향을 주고 있고, 코믹과 판타지 장르도 약한 긍정 영향을 주고 있다.

3.2 Classifier

모델 탐색에는 아래와 같은 분류기들을 사용하였으며 해당 분류기들의 파라미터 값을 조정하는 작업은 통계프로그램 R에서 각 모델의 자동 튜닝을 지원하는 caret 패키지를 이용하였다[18].

- 1R 규칙 학습자(OneR Rule, 이하 1R)
- 최근접 이웃 분류(k-nearest neighbor, 이하 KNN)
- C5.0 의사결정 트리(C5.0 Decision Tree, 이하 C50)
- 나이브 베이즈 분류(Naïve Bayes, 이하 NB)
- 인공신경망(Neural Network, 이하 NN)
- 서포트 벡터 머신(Support Vector Machine, 이하 SVM)
- 랜덤 포레스트(Random Forest, 이하 RF)

분류기의 세부 설명은 참고 문헌 [22], [23]으로 대체한다.

3.3 Evaluation Criteria

분류기들에 대한 성능 평가는 2x2 혼동 행렬(Confusion Matrix)로부터 도출할 수 있는 아래의 척도들을 활용하였다.

- 정확도(Accuracy)
- 민감도(Sensitivity)
- 특이도(Specificity)
- 정밀도(Precision)
- F-통계량(F-Measure)
- 균형 정확도(Balanced Accuracy)
- 카파 통계량(Kappa Statistic)

본 연구에서 알고자 하는 모바일게임화 가능성 여부 변수는 긍정과 부정 클래스간의 심각한 불균형성이 존재하고 있다. 따라서 이를 보정하는 적도인 균형 정확도와 카파 통계량이 주요한 해석 기준이 되었다.

2x2 혼동 행렬에서 균형 정확도의 수식은 (1)과 같이 긍정(positive, event) 분류율과 부정(negative, non-event) 분류율을 각각 계산하여 평균을 낸 값이므로 긍정의 개수와 부정의 개수가 불균형적이더라도 보정된 값을 산출한다.

$$Balanced Accuracy = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right) \quad \dots(1)$$

카파 통계량(Kappa Statistic)은 수식 (2)와 같이 실제 일치 확률인 Pr(a)와 무작위 선택 가정 하에서의 예상 일치 확률인 Pr(e)간의 관계식으로 분류기의 정확도를 수치로 보여준다. 심각한 클래스간 불균형이 있는 데이터의 경우 분류기가 가장 빈번한 클래스로만 추정하게 되면 높은 정확도를 쉽게 얻을 수 있기 때문에 랜덤 추정 확률을 고려함으로써 분류기만의 성능을 보다 명확히 해석할 수 있도록 한다.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad \dots(2)$$

카과 통계량에 대한 일반적인 해석은 다음과 같다[18].
 거의 일치하지 않음(Poor Agreement) = 0.20보다 작음
 어느 정도 일치(Fair Agreement) = 0.20 ~ 0.40
 보통 일치(Moderate Agreement) = 0.40 ~ 0.60
 좋은 일치(Good Agreement) = 0.60 ~ 0.80
 매우 좋은 일치(Very Good Agreement) = 0.80 ~ 1.00

IV. Results

4.1 Serial Publication Periods Prediction Model

4.1.1 Modelling

수집한 변수를 활용하여 연재 기간 예측 모델을 탐색하였다.

Table 2. Data Set Split for Serialization Model

Dataset	Observations	Condition
Train Set	200 obs. (about 70%) Serialization(1): 44.0%	Random sampling excluding Prediction Set
Test Set	87 obs. (about 30%) Serialization(1): 42.0%	
Prediction Set	15 obs. Not Determined	Serial Periods are under 24 weeks

Table 2와 같이 302편의 웹툰 데이터를 훈련용(Train Set), 테스트용(Test Set) 및 예측용(Prediction Set)으로 분리하였

Table 3. Classifiers evaluation between Train and Test set

Measures	1R	KNN	C50	NB	NN	SVM	RF
Accuracy	0.86	0.82	0.8	0.81	0.85	0.82	0.83
Sensitivity	0.73	0.68	0.78	0.84	0.84	0.57	0.78
Specificity	0.96	0.92	0.8	0.78	0.86	1	0.86
Precision	0.93	0.86	0.74	0.74	0.82	1	0.81
F-Measure	0.82	0.76	0.79	0.79	0.83	0.73	0.79
Balanced Accuracy	0.85	0.8	0.81	0.81	0.85	0.78	0.82
Kappa Statistic	0.71	0.62	0.61	0.61	0.7	0.6	0.65

Table 4. Implementation to the Prediction set

No.	Title Name	1R	KNN	C50	NB	NN	SVM	RF
1	New Undercover Inspector	1	1	1	1	1	1	1
2	She from the Random Chatting	1	1	1	1	1	0	1
3	Your side, My side	1	1	1	1	1	1	1
4	28th October	1	1	1	1	1	0	1
5	Tickling Tickling	1	0	1	0	1	0	1
6	Martial Arts Class 3	1	1	1	1	1	0	1
7	The Abyss sky Season 5	1	1	1	0	1	0	1
8	Troll Trap	1	1	1	1	1	1	1
9	Sweet Home	1	1	1	1	1	1	1
10	30 minutes on foot	0	1	1	1	1	1	1
11	My love is an idol star	1	1	1	1	1	1	1
12	School of reborn animals	1	0	1	1	1	1	1
13	Doctor Hound	1	1	1	1	1	0	1
14	Mystery Horror Subway	1	1	1	1	1	0	1
15	My favorite part	1	1	1	1	1	1	1

다. 예측용은 현업에서 연재 기간 계약 갱신이 1회이상 발생하
는 6개월(24주)를 분리 기준으로 하였다.

Table 3에 따르면 7개 분류기의 정확도가 모두 0.8 ~ 0.86
의 유의한 수준을 보이고 있다. 이외 척도들도 평균적으로 유의
한 수치를 나타내고 있어 어떤 분류기의 성능이 상대적으로 우
월하다고 판단하기는 어려워 보인다.

4.1.2 Predicting

현재까지 연재 기간이 24주 이하이면서 연재 종료로 되지
않은 예측용 셋에 분류기 모델을 적용하여 Table 4와 같이 결
과를 확인하였다.

가장 관련이 높은 특징 하나만 사용하는 1R 분류기는 15개
관측치 중에 유일하게 유료화 서비스를 하지 않고 있는 <걸어
서 30분(30 minutes on foot)>을 연재 종료로 분류하였다.

SVM 분류기는 15개 중 7개 웹툰을 연재 종료로 분류하여
종속변수의 클래스 분포와 유사할 정도로 많은 관측치를 분류
했다. SVM는 민감도(Sensitivity)는 0.57로 가장 낮은 반면, 정
밀도(Precision)는 1을 기록하였다. 이는 긍정 클래스인 연재
지속에 대한 예측은 정확하다는 것을 의미하므로, 부정 클래스
인 연재 종료로 분류된 관측치 결과에 주목하기보다 긍정으로
분류한 관측치의 결과를 보다 더 신뢰하는 쪽으로 해석하는 것
이 합리적일 것이다.

15개 웹툰 가운데 <간질간질(Tickling Tickling)>이 KNN,
NB, SVM등 3개의 분류기로부터 연재 종료로 분류되어 가장
많은 지지를 받았다.

<간질간질>은 관측치 중에 첫회 별점 점수가 가장 낮았고, 4
회, 8회, 12회 등 연재가 지속되면서 독자 참여 변화율도 현저

히 낮아지는 경향을 보이고 있다.

4.2 Mobile Gamification Likelihood Model

4.2.1 Modelling

모바일게임화 예측 모델의 데이터셋 홀드 아웃을 위해 <네이버 웹툰> 연재작 가운데 실제로 모바일게임화 되었거나 IP 계약이 완료되어 개발 중인 웹툰들의 연재 시작일로부터 게임 출시일까지 소요된 기간을 Table 5와 같이 조사하였다.

Table 5. Mobile Gamification Cases of Naver Webtoons

Title	Initial Publication	Mobile Gamification	Duration (weeks)
Electronic Game Guard	2014-10-08	2018-01-08	170
Natural Fools	2012-06-13	2017-12-28	289
Hive	2014-01-20	2017-11-09	198
Hold on the Spirit	2009-08-27	2017-08-24	417
Lookism	2014-11-20	2017-04-27	127
Nobleless	2007-12-30	2017-04-25	486
Denma	2010-01-08	2016-12-15	362
Gaus Company	2011-06-06	2016-09-12	275
Hero Maker	2006-11-27	2016-08-30	509
Sound of Heart	2006-09-08	2016-04-25	502
Americano Exodus	2014-05-16	2016-02-25	93
Tower of God	2010-06-30	2016-02-18	294
God of Highschool	2011-04-08	2015-05-25	215
Come on CVS	2008-07-02	2014-02-25	295
Mean			302
Median			292

Table 7. Classifiers evaluation between Train and Test set

Measures	1R	KNN	C50	NB	NN	SVM	RF
Accuracy	0.89	0.94	0.93	0.9	0.9	0.94	0.94
Sensitivity	0	0	0.2	0.2	0	0	0
Specificity	0.95	1	0.97	0.95	0.96	1	1
Precision	0	-	0.33	0.2	0	-	-
F-Measure	-	-	0.25	0.20	-	-	-
Balanced Accuracy	0.47	0.5	0.59	0.57	0.5	0.5	0.5
Kappa Statistic	-0.06	0	0.21	0.15	-0.05	0	0

Table 8. Implementation to the Prediction set

No.	Title Name	1R	KNN	C50	NB	NN	SVM	RF
1	Weird Weird	0	0	0	0	0	0	0
2	Trump	1	0	1	0	0	0	0
3	Zombie Chronicles of Joseon	0	0	0	0	0	0	0
4	Only You don't know!	0	0	0	0	0	0	0
5	Aimer	0	0	0	0	0	0	0
6	Wanted	0	0	0	0	0	0	0
7	Devil is bored	0	0	0	0	0	0	0
8	Evergreen Story	0	0	0	0	0	0	0
9	How to live in the fantasy	0	0	0	0	0	0	0
10	TAL	0	0	0	0	0	0	0
11	Real Zoo Zoo	0	0	0	0	0	0	0
12	Panda Dog	0	0	0	0	0	0	0
13	Penguin loves Mev	0	0	0	0	0	0	0
14	Supernatural Investigation Team	0	0	0	0	0	0	0
15	One Last God, Kubera	0	0	0	0	0	0	0
16	Story of Narb	0	0	0	0	0	0	0
17	Meddling with Life	0	0	0	0	0	0	0
18	The Magician	0	0	0	0	0	0	0
19	Mr. Tiger	0	0	0	0	1	0	0

Table 5에 따르면 모바일 게임으로 제작된 웹툰의 경우, 연재 시작일로부터 게임 출시일까지 평균 302주, 중앙값으로는 292주가 소요되었다. 계약 협상부터 개발 완료까지 게임 출시에 평균 1년(52주)을 예상한다면, 약 240주 정도가 모바일 게임화 가능성을 구분하는 적절한 분리 기준으로 추정된다.

Table 6. Data Set Split for Gamification Model

Dataset	Observations	Condition
Train Set	200 obs. (about 71%) Gamification(1): 8.0%	Random sampling excluding Prediction Set
Test Set	83 obs. (about 29%) Gamification(1): 6.0%	
Prediction Set	19 obs. Not Determined	Serial Periods are over 240 weeks

Table 6과 같이 302편의 웹툰 데이터를 훈련용(Train Set), 테스트용(Test Set) 및 예측용(Prediction Set)으로 분리하였다. 예측용 셋은 연재 기간이 240주 이상이면서 모바일 게임으로 제작되거나 계약되지 않은 조건의 웹툰 20편으로 구성되었다.

Table 7에 따르면 7개 분류기의 정확도가 0.89 ~ 0.94 범위에 분포하고 있어 유의한 수준으로 보이지만, 모바일게임화 여부를 판단하는 종속 변수(Gamification)의 심각한 클래스간 불균형성을 감안하면 신뢰할 수 없다. 분류기가 가장 빈번한 클래스인 93.0%의 부정 클래스로만 판정해도 0.93의 정확도를 확보할 수 있기 때문이다. 이와 같은 경우, 긍정 클래스와 부정

클래스의 분류율을 따로 산출하여 평균을 낸 균형 정확도와 랜덤 추정 확률을 고려한 카과 통계량이 기준 척도로 활용될 수 있음을 이미 3.3에서 언급한 바 있다.

Table 7의 결과에서 볼 수 있듯이, C50가 균형 정확도와 카과 통계량에서 어느 정도의 성능을 보였다. C50의 카과 통계량 0.21은 7개 분류기 가운데 어느 정도 일치(Fair Agreement)로 간주할 수 있는 구간인 0.20 ~ 0.40 에 유일하게 포함되어 있는 수치다.

4.2.2 Predicting

연재 기간이 240주 이상이면서 모바일 게임화가 되지 않은 예측용 셋에 구성된 19개 관측치에 분류기 모델을 적용하여 Table 8과 같이 결과를 확인하였다.

어느 정도 신뢰할 수 있는 C50 분류기가 <트럼프>를 모바일 게임화 가능성이 있는 웹툰으로 분류하였다.

<트럼프>는 세상을 떠난 아내와 기억을 지운 아이 사이에서 갈등하는 남자를 주인공으로 0.8초 후의 시공간을 배경으로 한 판타지물이다. 2012년 12월에 연재를 시작하여 2018년 1월 26일까지 첫 시작일로부터 265주차이며 총 220회가 연재 중에 있다. 장르가 모바일 게임이 선호하는 판타지인 점, 첫 회 별점 참여수 및 댓글 수가 예측 셋에서 1위를 기록한 것이 주 선정 요인으로 작용하였다.

IV. Conclusions

웹툰은 다른 콘텐츠와 비교해 볼 때, 저비용으로도 제작과 유통이 가능한 저위험-저수익(Low Risk Low Return) 콘텐츠로 인식되었다. 그러나 네이버, 다음 등 웹툰 플랫폼들이 콘텐츠에 대한 체계적인 유통과 관리를 담당하면서 블록버스터 영화, 드라마, 게임으로 발전하는 사례들이 속속 생겨나는 등 잠재적인 부가 가치가 주목 받고 있다.

웹툰은 콘텐츠로 볼 때는 프로토타이핑(Prototyping)이며, 제품으로 볼 때는 최소 기능 제품(Minimum Viable Product, MVP)이라고 할 수 있다. 특히 빠르게 아이디어를 시장에 선보여 가치를 검증받고, 유저 평가와 인지도를 바탕으로 IP의 확장을 도모할 수 있다는 점에서 웹툰과 MVP는 공통점이 많다.

게임업계에는 ‘잘 키운 IP 하나 열 게임 안 부럽다’는 속설이 있다. 이미 국내 모바일 게임 시장에서는 <리니지 레볼루션>, <리니지M>, <테라 M>, <열혈강호> 등 인기 온라인 게임이나 인기 원작 만화 등의 세계관과 캐릭터로 모바일 게임을 만드는 것이 트렌드로 자리 잡았다. 모바일 게임의 역사가 10년이 되어가면서 모바일의 특징을 참신하게 활용하거나 개발 기술을 차별화하는 것만으로는 한계를 보이고 있기 때문이다. 이를 보완하기 위해 기존 IP를 활용하거나 만화 등 다른 미디어 콘텐츠와의 제휴를 모색하고 있는 것이 현재 국내 모바일 게임 산업의 추세이다.

본 연구는 웹툰 콘텐츠의 특징과 서비스 방식을 조사하여 연재 초기 데이터로부터 웹툰의 잠재적 가능성을 확인하고(Serial Publication Periods Prediction Model, 연재 기간 예측 모델), 모바일 게임으로 확장이 용이한 웹툰을 발굴하려는(Mobile Gamification Likelihood Model, 모바일 게임화 가능성 예측 모델) 현업의 실무적 문제를 데이터 마이닝 기법으로 접근하였다는 점에 의의가 있다. 연재 기간 예측 모델은 균형 정확도 85.0%를 기록해 신뢰 할만한 결과를 확보하였다. 모바일 게임화 가능성 예측 모델은 중속 변수의 클래스 불균형성의 존재로 정확도는 낮지만 어느 정도 검토할 만한 결과를 도출하였다. 향후 웹툰에 대한 양질의 데이터가 쌓이고, 콘텐츠 특성에 대한 연구가 보다 활발해진다면 웹툰 작가에게는 연재할 기회가 더욱 많아질 것이고, 게임, 영화, 드라마 등 주요 콘텐츠 산업에도 양질의 IP를 공급하는 화수분 역할을 확대함으로써 웹툰 콘텐츠의 위상은 더욱 높아질 것이다. 이번 연구가 웹툰에 대해 학문적인 접근과 문화 산업적인 관심을 기울이는데 보탬이 될 것으로 기대한다.

REFERENCES

- [1] KOCCA, "2016 Contents Industry Whitepaper," pp. 66-67, Sep. 2017.
- [2] IMDb.com, Inc., www.boxofficemojo.com
- [3] Korean Film Council, www.kobis.or.kr
- [4] Lee, S.J., "A Study on Evolution/Expansion Model of Cartoon OSMU : Focusing on 'Misaeng' by Yun, Tae Ho," The Korean Journal of animation 11(2), pp. 70-84, June 2015.
- [5] Ha, J.Y. and Park, J.W., "A Case Study of Mobile Game Using Webtoon as Original Content," Journal of The Korean Society for Computer Game 30(3), pp. 35-43, Sep. 2017.
- [6] Joongang Ilbo, "Kakao ranked the 2nd place in Japanese Webtoon market," Nov. 2017.
- [7] Joseon Ilbo, "Uprising Webtoon Platforms, Targeting the Global Market," Sep. 2017.
- [8] Ideaware, Inc., www.wiseapp.co.kr, Period from Nov. 20 to Nov. 26, 2017.
- [9] Naver Corp., "Naver Webtoon 10th Anniversary Campaign Page," June 2014.
- [10] Kim, J.P., Seong, S.Ch. and Hong, W.K., "The Evolution of Webtoon Platform and the Future of Korean Webtoon," KT DigiEco, 2014.
- [11] Kim, J.P. and Seong, S.Ch., "Webtoon, Dreaming 1 Billion Market," KT DigiEco, 2015.
- [12] Han, Ch-W., "The Theory about Functional Evolution

of the Business Model in the Webtoon Platform and the Strategy to Spin Off," *Cartoon & Animation Studies* 45, pp 127-146, Dec. 2016.

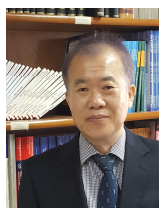
- [13] Han,Ch-W. and Hong,N.J., "The Research of Storytelling for Converting Webtoon into Movie," *Journal of the Korea Contents Association* 11(2), pp. 186-194, Feb. 2011
- [14] Yoon,K-H., Jung,K-H., Choi,I-S. and Choi,H-S., "Features of Korean Webtoons through the Statistical Analysis," *Cartoon & Animation Studies*, pp 177-194, Mar. 2015.
- [15] Blotter, "Profit Creation by Webtoon, Naver reveals PPS Program", Mar. 2013.
- [16] Mediaus, "Webtoon Creator is just 3-month part-time job," Apr. 2015.
- [17] Nam,J.E. and Kim,H.K., "A Study on the Type of Transmedia Contents," *Academic association f global cultural contents*, pp. 23-49, Aug. 2015.
- [18] Lantz,B. and Yoon,S.J., "Machine Learning with R - 2/e," Acorn Publishing Co., Sep. 2017.
- [19] Liu, S.M. and Chen, J.-H. "A multi-label classification based approach for sentiment classification". *Expert Systems with Applications*, Volume 42, Issue 3, pp.1083-1093, Feb. 2015.
- [20] Tsoumakas,G., Katakis,I., and Vlahavas,I., "Mining Multi-label Data," *Data Mining and Knowledge Discovery Handbook*, 2/e, pp. 667 - 685, Springer Science+Business Media, LLC., 2010.
- [21] Park,B-W. and Lee,K.C., "An Exploratory Analysis on Adoption of Potential Customers in Transmedia Stroytelling: Emphasis on Korean TV drama and Movie," *Korean Management Science Review* 27(2), pp. 81-95, July 2010.
- [22] Lee,E.M. and Lee,K.C., "Data mining approach to predicting user's past location," *Journal of the Korea Society of Computer and Information* 22(11), pp. 97-104, Nov. 2017.
- [23] Eo,K.S. and Lee,K.C., "Predicting stock price direction by using data mining methods : Emphasis on comparing single classifiers and ensemble classifier" *Journal of the Korea Society of Computer and Information* 22(11), pp. 111-116, Nov. 2017.

Authors



Hyun Seok Jang received the B.S. in Statistics from Korea University and Executive Master of Business Administration (EMBA) from Sungkyunkwan University. Jang works in the startup company specialized in contents data mining and

mobile game development.



Kun Chang Lee is a full professor of MIS in SKK Business School at Sungkyunkwan University. He is now in charge of Creativity Science Research Institute (CSRI) and Health Mining Research Center (HMRC) as well, Sungkyunkwan University. His

recent research interests lie in data mining, health informatics, creativity science, Human-Robot Interaction (HRI), and artificial intelligence techniques in decision making analysis.