

게임 리뷰 분석을 통한 긍정적 단어 탐색 및 게임 트렌드 분석

김정환
성균관대학교
e-mail : patrick46@g.skku.edu

Analysis of Game Reviews: Identifying Positive Keywords and Exploring Gaming Trends

Jung-Hwan Kim
Sungkyunkwan University

요약

연구에서는 Steam 게임 리뷰를 분석하여, 긍정적인 평가에 기여하는 주요 요인을 식별하고 게임 트렌드를 탐구하기 위해 두 가지 접근 방식을 적용하였다. 첫 번째 접근 방식에서는 SGD Classifier를 활용한 지도학습을 통해 리뷰의 감성을 예측하고, 긍정적인 리뷰와 관련된 영향력 있는 키워드를 추출하였다. 이 모델은 약 80%의 정확도를 기록하였으며, 긍정적 리뷰에서 자주 등장하는 상위 20개의 키워드를 성공적으로 식별하였다. 두 번째 접근 방식에서는 추출된 키워드를 바탕으로 최근 3년간의 Steam 게임 순위를 분석하여, 유저들이 선호하는 게임 트렌드를 집계하고 분석하였다. 데이터 전처리 과정에서는 결측값 처리, 날짜 형식 변환, 리뷰 점수의 통계적 지표(평균, 중앙값, 표준편차) 계산을 통해 데이터를 정제하였으며, 각 게임의 긍정적 리뷰 비율과 시간에 따른 리뷰 작성 분포를 시각화하여 트렌드를 분석하였다. 또한, 장르별 리뷰 점수와 감성 분석을 통해 게임 리뷰 데이터의 전반적인 품질을 파악하였다. 이 연구는 Steam 리뷰 데이터를 종합적으로 이해하는 데 기여했으며, 이 결과는 게임 개발 및 마케팅 전략 수립에 유용한 인사이트를 제공할 수 있을 것이다.

1. 서론

최근 몇 년간 게임 산업은 전 세계적으로 급격한 성장을 이뤄왔다. 게임 리뷰는 단순한 피드백을 넘어, 유저들의 감정과 게임에 대한 전반적인 평가를 반영하는 중요한 자료이다. 특히, Steam과 같은 플랫폼에서 제공되는 대규모 리뷰 데이터는 게임 개발자와 마케터들이 게임의 성공 요소를 분석하고, 유저들의 선호도를 이해하는 데 중요한 역할을 한다.

이 연구는 Steam에서 인기 있는 게임들의 리뷰 데이터를 수집하고, 이를 통해 게임 리뷰의 감정 분석 및 트렌드 파악을 목적으로 한다. 이를 위해 지도학습을 활용해, 리뷰의 감성을 분석하고, 이를 통해 각 게임에 대한 긍정적 평가를 이끄는 주요 요인들을 식별하고 이를 통해 게임의 트렌드를 파악할 수 있었다.

특히, 본 연구에서는 SGD Classifier를 활용해 리뷰의 감성을 예측하고, 긍정적 평가에 기여한 핵심 단어들을 추출했다. 이러한 접근 방식은 게임 개발 및 마케팅 전략을 수립하

는 데 있어 중요한 통찰을 제공할 것으로 기대된다.

본 논문에서는 먼저 데이터 수집과 전처리 과정을 설명한 후, 감정 분석 결과를 바탕으로 주요 분석 결과를 논의하고, 이를 통해 도출된 게임 트렌드를 소개할 것이다.

2. 이론적 배경

2.1 게임 리뷰와 소비자 평가

디지털 시대에 들어서며, 소비자들은 제품 구매 시 더 많은 정보를 탐색하여 구매 결정을 내리기 위해 다른 소비자들의 사용 후기와 같은 정보를 적극적으로 활용하게 되었다. 이는 온라인 구전(Online Word-of-Mouth, WOM)이라는 개념으로, 소비자 간의 제품 정보나 사용 경험, 추천 등의 정보 교환을 의미한다 [1]. 특히 온라인 리뷰는 이러한 온라인 구전의 대표적인 형태로서, 기업이 제공하는 공식 정보보다 높은 신뢰성을 가지며, 소비자의 구매 결정과 기업 이미지 형성에 중요한 역할을 한다. 게임 산업에서도 이러한 온라인 리뷰는 게

임의 성공과 판매량에 큰 영향을 미치며, 잠재적 구매자에게 신뢰할 수 있는 정보를 제공한다. 따라서, 온라인 리뷰의 내용과 감정 분석은 게임 개발자와 마케터들에게 중요한 연구 주제가 되고 있다.

2.2 감정 분석(Sentiment Analysis)

감정 분석(Sentiment Analysis)은 텍스트 데이터에서 긍정적, 부정적, 중립적 감정을 식별하고 평가하는 자연어 처리(NLP) 기술이다. 이 기술은 주로 고객 리뷰, 소셜 미디어 게시물, 설문 조사 응답 등의 데이터를 분석하는 데 사용되며, 사용자들이 특정 제품이나 서비스에 대해 어떤 감정을 느끼고 있는지 파악하는 데 유용하다. 감정 분석은 머신러닝 모델, 특히 지도학습(Supervised Learning) 모델을 활용해 수행된다. 텍스트 데이터를 레이블링하여 학습된 모델은 이후 새로운 데이터의 감정을 자동으로 예측할 수 있다. 본 연구에서는 SGD Classifier를 사용하여 감정 분석을 수행했으며, 이는 확률적 경사 하강법(Stochastic Gradient Descent)을 활용한 빠르고 효율적인 선형 모델이다.

2.3 텍스트 데이터의 전처리

텍스트 데이터는 일반적으로 비정형 데이터를 포함하며, 이를 분석하기 위해서는 적절한 전처리 과정이 필수적이다. 전처리 과정에는 불필요한 단어 제거, 어간 추출(Stemming) 또는 표제어 추출(Lemmatization), 토큰화(Tokenization), 단어 임베딩(Word Embedding) 등이 포함된다. 이러한 과정을 통해 텍스트 데이터는 분석에 적합한 형태로 변환되며, 머신러닝 모델이 이를 보다 효율적으로 학습할 수 있게 된다.

Document D1	The child makes the dog happy the: 2, dog: 1, makes: 1, child: 1, happy: 1
Document D2	The dog makes the child happy the: 2, child: 1, makes: 1, dog: 1, happy: 1



	child	dog	happy	makes	the	BoW Vector representations
D1	1	1	1	1	2	[1,1,1,1,2]
D2	1	1	1	1	2	[1,1,1,1,2]

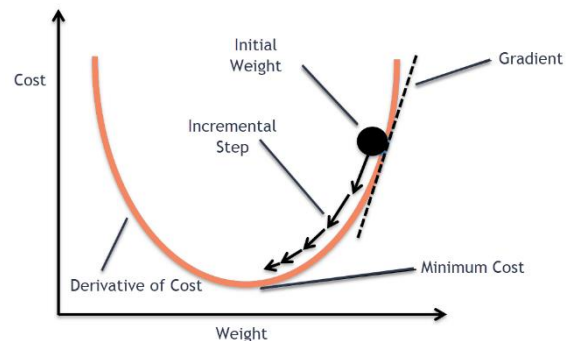
(그림 1) 텍스트 벡터화 방법 중 Bag of Words (BoW)

본 연구에서 리뷰 텍스트를 분석하기 전에 데이터 전처리를 수행했으며, 이를 통해 모델의 성능을 향상시킬 수 있었다. 텍스트 전처리는 데이터의 품질을 좌우하는 중요한 요소이다.

3. 논문 본문

본 연구에서는 Steam 플랫폼에서 인기 있는 상위 20개의 게임을 선정하여 해당 게임에 대한 사용자 리뷰 데이터를 수집했다. 웹 크롤링 기법을 활용하여 각 게임의 리뷰를 수집하였으며, 이를 CSV 파일로 저장하여 데이터 분석에 활용하였다. 수집된 데이터에는 리뷰 내용, 작성일, 점수 등의 정보가 포함되어 있으며, 분석에 앞서 데이터 전처리 과정을 거쳤다.

데이터 전처리 과정에서는 텍스트 데이터에서 불필요한 숫자, 특수 문자, URL, 그리고 감정 분석에 크게 기여하지 않는 불용어(stop words)를 제거했다. 이를 통해 텍스트의 핵심 의미를 보존하면서도 분석에 방해가 되는 요소들을 최소화할 수 있었다. 이후, 텍스트를 단어 빈도에 따라 벡터화하여, 모델이 리뷰 텍스트를 수치적으로 이해할 수 있도록 하였다. 전처리된 데이터를 기반으로 확률적 경사 하강법(Stochastic Gradient Descent, SGD)을 이용한 텍스트 분류 모델을 구현하였다. 확률적 경사하강법은 매개변수를 갱신하기 위하여 무작위로 샘플링된 한 개의 데이터 샘플에 대해 손실 함수의 기울기를 계산하여 매개변수를 갱신하는 방법이다 [2].

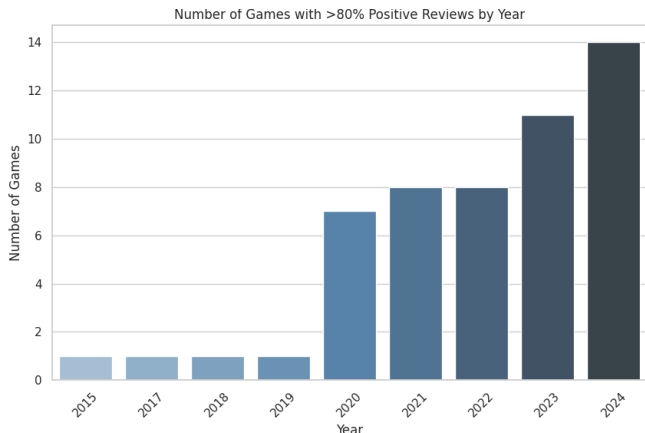


(그림 2) 확률적 경사하강법 과정의 예시

모델 학습 과정에서는 교차 검증 기법을 통해 성능을 검증하고, 최종적으로는 전체 학습 데이터를 사용해 모델을 학습시켰다. 학습이 완료된 후에는 테스트 데이터를 통해 모델의 예측 성능을 평가하고, 정확도와 F1 점수와 같은 성능 지표를 관찰했다. 이후, 모델이 학습한 내용을 바탕으로 리뷰에서 긍정적인 감정을 나타내는 주요 단어 20개를 추출하여 트렌드 분석에 활용하였다.

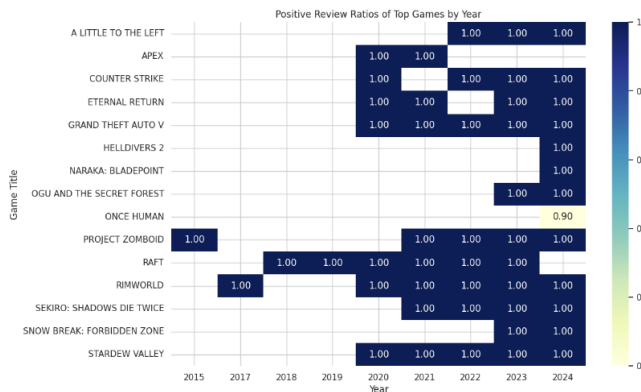
또한, 각 게임에 대한 리뷰 점수의 분포를 분석하고, 이를 바탕으로 게임 간 비교를 시도하였다. 먼저 연도별로 80% 이상의 긍정적 리뷰를 받은 게임의 수를 분석해 시각화했다. 이를 통해 특정 기간 동안 긍정적 평가를 받은 게임 수의 변화를 확인할 수 있었으며, 2020년 이후의 긍정적 평가를 받은 게임 수가 많은 것을 발견했다. 이는 선정된 게임 중 70%가

2020년도 이후에 발매된 게임이었기 때문에 분석되었다.



(그림3) 연도별로 80% 이상의 긍정적인 리뷰를 받은 게임수

각 게임마다 연도별 긍정적 리뷰 비율을 비교하기 위해 히트맵을 생성했다. 이 히트맵은 각 게임 제목과 연도별 긍정적 리뷰 비율을 색상으로 표현하여, 시간이 지남에 따라 게임의 긍정적인 리뷰 비율이 어떻게 변화했는지를 한눈에 파악할 수 있도록 했다.



(그림4) 연도별 상위 게임의 긍정적인 리뷰 비율 히트맵

이 연구에서는 감정 분석 모델을 활용해 게임 리뷰에서 긍정적인 단어 20개를 식별했다. 분석 결과, 상위 20개의 긍정적 단어는 다음과 같았다: 'games', 'players', 'better', 'got', 'game', 'time', 'playing', 'early', 'access', 'content', 'system', 'play', 'nt', 'lot', 'money', 'want', 'characters', 'way', 'player', 'buy'. 이 단어들은 게임 리뷰에서 자주 긍정적인 반응을 나타내는 표현들로, 주로 게임의 콘텐츠와 사용자 경험을 강조하고 있다. 특히 'content'와 같은 단어는 게임의 풍부한 콘텐츠와 관련된 긍정적인 피드백을 반영한다. 이를 바탕으로, 최근 3년간 Steam 인기 순위의 게임을 통해 각 게임의 콘텐츠를 분석하여 게임의 트렌드를 파악하고자 하였다.

게임의 콘텐츠를 분석하기 위해 다시 Steam 게임 페이지에서 게임 설명을 수집했다. Steam은 게임의 특성, 스토리, 그리고 게임 플레이에 관한 정보를 포괄적으로 제공하는 플랫폼

이다. 웹 스크래핑 기법을 통해 게임 설명의 텍스트 데이터를 수집하여 분석에 필요한 원자료를 확보했다. 이때 2021년도부터 2023년도까지 각 연도마다 게임의 설명을 수집했으며, 분석에는 <https://store.steampowered.com/charts>에서 각 연도별로 Platinum부터 Silver 부문까지의 인기 게임들을 포함했다.

게임 설명이 수집된 후, 텍스트 전처리 과정을 통해 불필요한 기호를 제거하고, 텍스트를 분석에 적합한 형태로 변환했다. 이후 Latent Dirichlet Allocation (LDA) 기법을 사용해서 주요 주제를 추출했다. LDA는 문서 집합에서 잠재적인 주제 구조를 발견하는 데 사용되는 확률적 모델로, 각 문서는 여러 잠재적 주제를 혼합하여 표현된다고 가정한다. 이런 가정 하에 각 문서와 주제의 확률 분포를 추론해 각 문서의 잠재적인 주제 구성을 파악함으로써, 다양한 분야에서 문서의 잠재적 정보를 탐색하는 데 활용되고 있다 [3]. 이를 통해 게임 설명에서 주제를 나타내는 주요 키워드를 얻을 수 있었다.

추출된 주제는 연도별로 집계해 데이터 프레임으로 정리했으며, 각 주제의 빈도를 연도별로 기록하여, 시간에 따른 주제의 변화를 추적했다. 이 과정에서 빈도가 1인 주제는 필터링하여 보다 의미 있는 데이터만을 남기고, 연도가 증가함에 따라 출현 횟수가 증가한 주제와 감소하는 주제를 파악할 수 있었다.

4. 결론

(표1) 게임 산업에서 최근 증가하는 주제와 감소하는 주제

범주	증가주제	감소주제
게임 특성	RPG, fast-paced, immersive, challenge, story, gameplay, Warhammer, classic, universe, worlds, explore, craft	shooter, first-person, battles, weapons, land, greatest, escape, hunt, online
기술및개발	technology, engine, development, hypermotion2, new, rise, ground, distinct, creatures, city, build, free-to-play, fast-paced, distinct	awards, award-winning, franchise, free
플레이어 경험	players, teams, friends, women's, men's, distinct	thrilling, experience, play, culture, iconic, team, use,

	world, explore, creatures, fast-paced	equipment
기타	Warhammer, classic, universe, city, fast-paced	culture, year

본 연구에서는 Steam 플랫폼에서 최근 3년간 인기 있었던 상위 게임들을 대상으로 게임의 설명을 분석해 게임 산업의 최신 트렌드를 파악하고자 했다. 그 결과, 게임 몰입감의 증가, 개인화 및 커스터마이징의 강화, 그리고 기술 혁신이라는 세 가지 주요 트렌드를 확인할 수 있었다.

첫째, 게임의 몰입감을 강화하는 요소들이 주목받고 있음을 발견했다. 특히, 스토리라인, 세계관, RPG 요소와 같은 핵심 요소들이 플레이어에게 더욱 깊이 있는 경험을 제공하는 데 중요한 역할을 하고 있다. 이러한 요소들은 플레이어가 가상 세계에 몰입할 수 있도록 돕고, 게임을 단순한 오락을 넘어서는 경험으로 승화시키고 있다.

둘째, 기술 혁신은 게임 개발에 있어 중요한 촉매제가 되고 있다. 새로운 게임 엔진과 발전된 개발 기술은 더욱 정교하고 현실감 넘치는 게임 환경을 조성하고 있으며, 이를 통해 플레이어는 더욱 몰입할 수 있는 경험을 하게 된다. 이는 기술이 게임 몰입감뿐만 아니라 전반적인 사용자 경험을 향상시키는 데 큰 기여를 하고 있음을 시사한다.

셋째, 게임에서의 개인화와 커스터마이징이 점점 더 중요해지고 있다. 플레이어들은 자신만의 세계를 구축하거나 독특한 경험을 할 수 있는 기회를 중시하며, 게임 개발자들은 이에 맞춰 더 많은 선택권과 커스터마이징 옵션을 제공하고 있다. 이는 게임이 단순한 놀이에서 개인의 창의성을 표현하는 도구로 발전하고 있음을 보여준다.

반면, 전통적인 게임 요소와 장르, 예를 들어 슈팅 게임이나 프랜차이즈 게임 등의 인기는 감소하고 있는 것으로 나타났다. 이는 게임 산업이 보다 새로운 형태의 경험과 기술 중심으로 재편되고 있음을 보여주며, 미래의 게임 산업이 보다 혁신적이고 다양화된 방향으로 나아갈 것임을 예고한다.

물론, 본 연구에서 분석된 게임 리뷰의 양과 연도별 게임 데이터의 수가 충분하지 않을 수 있어 결과의 정확성에 한계가 있을 수 있다. 그럼에도 불구하고, 이 분석을 통해 게임 산업의 트렌드를 어느 정도 파악할 수 있었으며, 향후 연구에서 더 많은 데이터를 기반으로 한 심층적인 분석이 이루어진다면 더욱 명확한 결론을 도출할 수 있을 것이다.

참고문헌

- [1] 강하나, 용혜련, 황현석, "게임 유용성 평가에 미치는 요인에 관한 연구: 스팀(STEAM) 게임 리뷰데이터 분석", 한국게임학회 논문지, 17(3), 33-44, 2017.
- [2] 강민제, "딥러닝을 위한 경사하강법 비교", 한국산학기술학회 논문지, 21(2), 189-194, 2020.
- [3] 홍태호, 니우한잉, 임강, 박지영, "LDA를 이용한 온라인 리뷰의 다중 토픽별 감성분석 - TripAdvisor 사례를 중심으로 -", 정보시스템연구, 제27권 1호, 89-110쪽, 2018년.