http://www.jsebs.org
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.24, No.2, May 2019, pp.1-14
https://doi.org/10.7838/jsebs.2019.24.2.001

단어 임베딩 및 벡터 유사도 기반 게임 리뷰 자동 분류 시스템 개발

Development of An Automatic Classification System for Game Reviews Based on Word Embedding and Vector Similarity

양유정(Yu-Jeong Yang)*, 이보현(Bo-Hyun Lee)**,
김진실(Jin-Sil Kim)***, 이기용(Ki Yong Lee)****

초 록

게임은 소프트웨어 특성상 출시 후 사용자들의 반응을 빠르게 파악하여 개선하는 것이 중요하다. 하지만 구글 플레이 앱 스토어 등 사용자들이 게임을 다운로드하고 리뷰를 올릴 수 있는 대부분의 사이트들은 게임 리뷰에 대한 매우 제한적이고 모호한 분류 기능만을 제공한다. 따라서 본 논문에서는 사용자들이 사이트에 올린 게임 리뷰를 보다 명확하고 운영에 유용한 주제들로 자동 분류하는 시스템을 개발한다. 본 논문에서 개발한 시스템은 리뷰에 포함된 단어들을 대표적인 단어 임베딩 모델인 word2vec을 사용하여 벡터들로 변환하고, 이 벡터들과 각 주제 간 유사도를 측정하여 해당 리뷰를 관련된 주제로 분류한다. 특히 분류 성능에 직접적인 영향을 미치는 벡터 간 유사도 측정 방법을 선택하기 위해 본 연구에서는 대표적인 벡터 간 유사도 측정 방법인 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도의 성능을 실제 데이터를 사용하여 비교하였다. 또한 어떤 리뷰가 둘 이상의 주제에 해당하는 경우를 위해 임계값에 기반한 다중 분류 방법을 사용하였다. 구글 플레이 앱스토어의 실제 데이터를 사용한 실험 결과 본 시스템은 95%까지의 정확도를 보임을 확인하였다.

ABSTRACT

Because of the characteristics of game software, it is important to quickly identify and reflect users' needs into game software after its launch. However, most sites such as the Google Play Store, where users can download games and post reviews, provide only very limited and ambiguous classification categories for game reviews. Therefore, in this paper, we develop an automatic classification system for game reviews that categorizes reviews into categories that are clearer and more useful for game providers. The developed system

이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2018R1D1A1B07045643).

* First Author, Department of Computer Science, Sookmyung Women's University (diddbwjd96@sookmyung.ac.kr)
** Co-Author, Division of Computer Science, Sookmyung Women's University(lbhsos29@gmail.com)
*** Co-Author, Division of Computer Science, Sookmyung Women's University(kjinsil0819@gmail.com)
**** Corresponding Author, Division of Computer Science, Sookmyung Women's University (kiyonglee@sookmyung.ac.kr)

Received: 2019-03-25, Review completed: 2019-04-15, Accepted: 2019-05-13

converts words in reviews into vectors using word2vec, which is a representative word embedding model, and classifies reviews into the most relevant categories by measuring the similarity between those vectors and each category. Especially, in order to choose the best similarity measure that directly affects the classification performance of the system, we have compared the performance of three representative similarity measures, the Euclidean similarity, cosine similarity, and the extended Jaccard similarity, in a real environment. Furthermore, to allow a review to be classified into multiple categories, we use a threshold-based multi-category classification method. Through experiments on real reviews collected from Google Play Store, we have confirmed that the system achieved up to 95% accuracy.

키워드 : 리뷰 분석, 텍스트 분류, Word2vec, 워드 임베딩, 유사도 측정
Review Analysis, Text Classification, Word2vec, Word Embedding, Similarity Measure

1. 서 론

네트워크 환경 발전과 함께 스마트폰의 보급이 확대되면서 모바일 게임 시장의 규모는 점점 커지고 있다. DMC 리포트에 따르면 2018년 모바일 게임 시장의 규모는 703억 달러에 이를 것으로 추정되었으며, 모바일 게임은 전체 게임 시장의 규모 중 51%로 가장 높은 비중을 차지한다[2]. 한국콘텐츠진흥원에 따르면 2018년 국내 모바일 시장 규모도 5조 원을 돌파하며 점점 커지고 있다[6]. 이런 성장 규모에 맞춰 다양한 모바일 게임이 등장하였다.

게임은 소프트웨어 특성상 출시 후, 업데이트가 빈번히 발생한다. 특히 업데이트 직후 사용자들의 반응과 요구사항을 빠르게 파악하여 개선 단계에 반영하는 과정이 반복된다. <Figure 1>은 현재 모바일 게임에 대한 구글 플레이 스토어(Google Play Store)의 리뷰 요약 정보를 나타내는 화면으로, 리뷰에 대한 정보는 각 별점의 분포수와 기능에 대한 세 가지 범주뿐으로 매우 제한적이다. 기능에 대한 평점의 범주는 게임 플레이, 그래픽, 컨트롤로 구성되며 게임 운영 및 시스템에 관한 범주는 미흡하다. 따라서 해당 범주에 대한 점수가 구체적으로 게임의 어떤 부분에 대한 만족도인지 명확하게 알기 어렵다. 또한 리뷰는 보통 줄글로 남기기 때문에 주제를 파악하는데 많이 시간이 소요된다. 긴 줄글의 경우, 줄글의 특성상 한 리뷰 내에서도 다양한 주제를 내포할 수 있고, 오타나 맞춤법 오류로 인하여 정확한 구분이 어려운 경우가 많다.

〈Figure 1〉 Review Summary in the Google Play Store

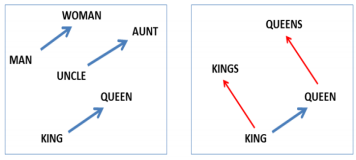
따라서 본 논문에서는 게임 리뷰를 보다 명확하고 운영에 유용한 주제들로 자동 분류하는 시스템을 개발한다. 본 논문에서 수집한 리뷰 데이터는 해당 리뷰가 어느 카테고리인지에 대한 명시적인 정답이 주어지지 않은 데이터이다. 따라서 정답이 주어진 라벨링 된 데이터(labeled data)를 활용한 기계학습 기법은 사용할 수 없으며, 리뷰에 구성된 단어만을 활용하여 카테고리를 판단하여야 한다. 본 논문에서 개발한 시스템은 게임 리뷰에 포함된 단어들을 대표적인 단어 임베딩 모델인 word2vec을 사용하여 벡터들로 변환하고, 이 벡터들과 각 주제간 유사도를 측정하여 가장 가까운 주제로 분류한다. 또한 리뷰가 하나 이상의 주제를 내포하는 경우를 고려하여 두 가지 카테고리로도 배정될 수 있도록 하였다. 본 논문에서는 특히 카테고리 분류 성능에 직접적인 영향을 미치는 벡터 간 유사도 측정 방법을 선택하기 위하여 대표적인 벡터 간 유사도 측정 방법인 유클리디안(Euclidean) 유사도, 코사인(cosine) 유사도, 확장된 자카드(Jaccard) 유사도의 성능을 실제 데이터를 사용하여 비교하였다. 실제 데이터를 사용한 다양한 실험을 통해, 본 논문에서는 95%의 분류 정확률로 가장 좋은 정확도를 보이는 확장된 자카드 유사도를 시스템 구현에 채택하였다. 또한 리뷰와 거리가 제일 가까운 카테고리와 그 다음으로 제일 가까운 카테고리 간의 차이가 임계값 이하일 때는 두 가지 카테고리로 분류하도록 하였으며, 이 때 최적의 임계값을 결정하기 위해 실제 데이터를 사용하여 가장 좋은 성능을 보이는 임계값을 선택하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 논문에서 사용한 단어 임베딩의 개념을 간략히 설명하고, 온라인 사용자 리뷰 분석을 살펴본다. 제3장에서는 본 논문에서 개발한 시스템의 개요를 소개하고, 구현 기술을 자세히 설명한다. 제4장에서는 다양한 유사도 측정 방법들을 본 시스템에 적용했을 때의 성능 평가 결과와 다중 분류 시 임계값에 따른 성능 변화 결과를 보이며, 제5장에서는 결론을 맺는다.

2. 관련 연구

2.1 단어 임베딩(Embedding)

Word2vec은 한 문장 내에서 단어의 등장 빈도 데이터를 이용하여 단어 자체가 가지는 의미를 다차원 벡터 공간에 임베딩하는 모델이다. NNLM(Neural Net Language Model)을 기반으로 하여 대량의 문서 말뭉치(corpus)를 빠르게 학습하여 처리할 수 있다[10]. 학습 방식은 주변 단어를 이용하여 중간 단어를 예측하는 방법인 CBOW(Continuous Bag of Word)와 중간에 있는 단어를 이용하여 주변 단어를 예측하는 Skip-Gram 방식이 있다.



〈Figure 2〉 Example of Word2vec Embedding

<Figure 2>는 word2vec을 사용하여 단어를 2차원 벡터로 변환한 예이다. 그림에서 'Man'과 'Woman'의 거리는 'King'과 'Queen'의 거리와 비슷한 것을 볼 수 있다. 이처럼 학습을 통해 단어의 문맥적 의미를 수치적으로 보존하기

때문에 이를 이용하여 각 단어들 간의 유사도를 측정하거나 수치적으로 쉽게 다룰 수 있다[8, 11].

2.2 온라인 사용자 리뷰 분석

인터넷 환경의 발달과 뉴미디어의 등장으로 다양한 형식의 리뷰들이 등장하고 있다. 리뷰를 통해 서비스에 대한 사용자 혹은 구매자의 생각과 의견을 알 수 있으며, 이는 다른 소비자의 구매 결정에 영향을 주는 중요한 정보 원천이다[7]. 기업은 리뷰를 분석한 결과를 의사 결정 지원에 활용하거나 서비스에 대한 개선점이나 매출 증대와 같은 마케팅에 전략적으로 활용할 수 있다[5, 16].

또한 게임과 같이 직접 사용해보지 않으면 해당 서비스의 품질 및 만족도를 알 수 없는 경험재의 경우, 리뷰는 다른 사용자에게 더 큰 영향을 미친다[18]. 온라인 리뷰는 온라인 구전(word-of-mouth)의 가장 대표적인 형태이다. 다양한 분야의 리뷰를 통한 온라인 구전 연구는 온라인 구전이 기업의 매출에 미치는 영향을 분석하여 온라인 구전 효과에 대한 중요성을 강조하고 있다[1, 3].

단어 임베딩을 위해 필요한 빠른 학습이 가능해 지면서 다량의 문서들 간의 유사도를 계산하여 유사한 문서들 별로 군집화하거나 분류할 수 있게 되었다[9, 15, 17]. 또한 상품을 벡터로 임베딩 시키는 것도 가능해지면서 다양한 추천 시스템이 등장하였다. 이러한 추천 시스템으로는 리뷰 데이터를 학습하여 음식점 예약 서비스를 제공하는“OpenTable”[13]과 사용자 플레이리스트의 노래를 벡터화하여 추천에 활용하는 스트리밍 음원 제공 서비스인 “Spotify”[12] 등이 있다.

하지만 단어들을 임베딩한 경우에도 각 카테고리를 특정 단어 하나만으로 나타내는 경우, 리뷰와 각 카테고리 간의 유사도가 정확하게 계산되지 않을 수 있다. 또한 어떤 유사도 측정 방법을 사용하느냐에 따라 분류 성능이 달라질 수 있다. 따라서 본 논문에서는 word2vec의 결과로 나온 단어 벡터들을 사용하여 리뷰와 각 카테고리 간의 유사도를 효과적으로 계산하고, 그 결과로 리뷰를 하나 또는 두 개의 카테고리로 분류하는 시스템을 제안한다.

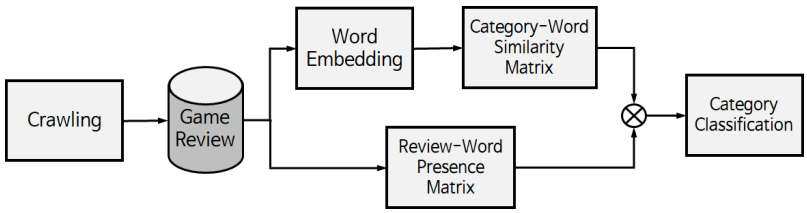
3. 게임 리뷰 자동 분류 시스템

본 장에서는 본 논문에서 제안하는 게임 리뷰 자동 분류 시스템에 대해 상세히 설명한다.

3.1 개요

본 논문에서 제안하는 게임 리뷰 자동 분류 시스템은 입력받은 리뷰를 게임의 운영 및 시스템에 관한 8가지 주제(카테고리)로 자동으로 분류한다. <Figure 5>의 왼쪽 열은 본 시스템이 리뷰를 분류하는 8개의 카테고리를 나타낸다. 본 시스템을 활용하면 게임 운영자는 수많은 리뷰를 하나씩 읽어보지 않고도 리뷰들을 자동으로 분류하여 시스템의 개선에 효과적으로 사용할 수 있다.

<Figure 3>은 본 시스템의 수행 흐름을 나타내는 순서도이다. 본 시스템은 먼저 게임에 대한 사용자들의 리뷰 데이터를 크롤링(crawling)하여 수집한다. 이후 수집된 리뷰 데이터에서 전처리 과정을 통해 의미 있는 단어들만을 추출하여 이들을 word2vec을 통해 벡터로 임베딩한다.



〈Figure 3〉 The Execution Flow of the System

임베딩이 끝나면 각 카테고리를 나타내는 대표 단어들과 리뷰에 나타나는 각 단어들 간의 벡터 간 유사도를 계산하여 그 결과로 카테고리-단어 간 유사도 행렬을 생성한다. 또한 각 리뷰에 대해 어떤 단어들이 포함되어 있는지를 0과 1로 나타내는 리뷰-단어 존재 행렬을 구축한다. 이렇게 두 개의 행렬이 구축되면 마지막으로 이 두 행렬을 사용하여 각 리뷰별로 그에 포함된 단어들과 각 카테고리를 나타내는 단어들 간의 유사도를 합산하여 그 총합으로 해당 리뷰를 스코어가 가장 큰 카테고리로 분류한다. 다음은 각 단계를 상세히 설명한다.

3.2 데이터 수집 및 전처리 단계

본 논문에서는 Python 기반의 Selenium 프레임워크와 BeautifulSoup 라이브러리를 사용하여 모바일 애플리케이션 구매 사이트인 구글 플레이 스토어에 올라온 게임 리뷰들을 크롤링하였다. Selenium은 웹 애플리케이션을 위한 테스팅 프레임워크로 웹드라이버(webdriver) API를 이용하여 크롤링하고자 하는 해당 페이지의 스크롤을 자동으로 제어한다. 웹드라이버를 통해 렌더링 된 페이지의 요소(element)는 모두 BeautifulSoup 라이브러리의 객체로 전달된다. BeutifulSoup으로 생성된 객체는 이를 바탕으로 HTML 파일의 여러 요소에 접근하여 원하는 부분만 자동으로 크롤링한다. 본 논문에서는 113개의 게임에 대한 리뷰를 수집하였으며, 수집한 리뷰는 약 총 7만 개이다. 수집된 리뷰는 작성자 이름, 내용, 평점, 유용한 정도, 날짜, 게임 카테고리로 구성된다.

리뷰가 수집되고 나면 Konlpy의 Twitter 라이브러리를 통해 형태소를 분석한 뒤, 분석에 필요하다고 판단되는 명사, 형용사, 동사만 선별하였다. Twitter 라이브러리는 스칼라(Scala)로 쓰인 한국어 처리기로 정규화, 토큰화, 어근화, 어구 추출 과정을 통해 주어진 리뷰에 형태소를 태깅(tagging)한다. 리뷰가 너무 짧은 경우 명확한 분류가 어려워 형태소 분석 결과로 나온 토큰(token)이 6개 이상인 리뷰만 분석 대상으로 사용하였다. <Figure 4>는 수집된 데이터를 전처리한 예시로, 학습에 적합한 형태로 바꾸기 위해 리뷰에 포함된 각 단어들에 형태소를 태깅하고 불필요한 품사는 제거하였다.

Collected review
"캐릭터 너무 이쁘고 진짜 재미있어요~ 스토리 굿굿굿!"
Classification of morpheme
캐릭터/Noun, 너무/Adverb, 이쁘다/Adjective, 진짜/Adverb, 재미있다/Adjective, 스토리/Noun, 굿굿굿/Noun
Removal of part-of-speech
캐릭터/Noun, 이쁘다/Adjective, 재미있다/Adjective, 스토리/Noun, 굿굿굿/Noun

〈Figure 4〉 Example of Data Preprocessing

3.3 카테고리-단어 간 유사도 행렬 생성 단계

리뷰로부터 형태소가 태깅된 단어들을 추출하고 나면, 이 단어들을 word2vec의 Skip-gram 방식을 사용하여 벡터들로 임베딩한다. Skip-gram 방식은 중심 단어로 주변 단어를 예측하는 방법으로 본 논문에서는 한 번에 학습할 단어의 개수인 윈도우(window)를 2로 지정하여 모델을 생성하였다. 임베딩 과정에서 정확도를 향상시키기 위해 출현 빈도가 20번 미만인 단어는 분석에서 제외하였으며, 임베딩 차원 수는 100으로 지정하여 각 단어를 100차원의 벡터로 변환하였다.

Category	Representative workds
결제	결제, 구입, 구매, 현질, 환불
계정	계정, 아이디, 연동, 구글, 로그인
서버	연결, 서버, 접속, 로딩, 렉
구성	맵, 이벤트, 퀘스트, 스테이지, 미션
연출	배경, 그래픽, 퀄리티, 사운드, 디자인
캐릭터	스킬, 너프, 영웅, 캐릭터, 신캐
시스템	용량, 다운, 앱, 실행, 설치
기타	광고, 신고, 채팅, 욕, 처벌

〈Figure 5〉 Representative Words for Each Category

<Figure 5>는 본 시스템에서 게임 리뷰가 분류되는 8개의 카테고리를 나타낸다. 이 8개의 카테고리는 결제, 계정, 서버, 구성, 연출, 캐릭터, 시스템, 기타 등 게임 운영, 구성, 시스템에 관한 범주들로 구성된다. 이 카테고리들은 모바일 게임 회사의 문의 사항 처리 사이트를 바탕으로 구성되었다. 인기 순위에 있는 다양한 모바일 게임 회사의 문의 사항 처리 사이트에서 가장 많이 언급된 문의들을 바탕으로 카테고리를 선정하였으며, 각 카테고리 당 각 카테고리를 나타내는 5개의 대표 단어들을 추가로 추출하였다. 이 대표 단어들은 단어 임베딩 결과 각 카테고리를 나타내는 단어와 가장 유사한 5개의 단어들로 구성되며, 각 카테고리에 대해 여러 개의 단어를 사용함으로써 리뷰 분류에 대한 정확도를 높인다.

단어 임베딩을 통해 각 단어가 100차원의 벡터로 변환되면, 본 논문에서는 카테고리들을 대표하는 각 단어들과 리뷰에 포함된 각 단어들 간의 유사도를 나타내는 카테고리-단어 간 유사도 행렬을 생성한다. 카테고리-단어 간 유사도 행렬의 각 행은 8개 카테고리들을 대표하는 각 단어들을 나타내며, 각 열은 리뷰에서 최소 출현 빈도 조건을 만족하는 단어들을 나타낸다. 카테고리-단어 간 유사도 행렬의 각 원소는 카테고리를 대표하는 어떤 단어 X와 리뷰에 나타나는 어떤 단어 Y 간의 유사도를 나타낸다. 100차원 벡터로 표현되는 X와 Y 간의 대표적인 유사도 측정 방법으로는 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도 등이 있으며, 이들은 주어진 데이터의 특징에 따라 다른 성능을 보인다[4]. 본 논문에서는 상기 3개의 유사도 측정 방법을 사용하여 카테고리-단어 간 유사도 행렬을 각각 계산한 뒤, 실제 데이터에 대해 최고의 분류 정확도를 보이는 방법을 시스템 구현에 채택하였다. 제4.2절에서는 이들에 대한 성능 평가 결과를 보인다.

본 절에서는 본 논문에서 사용한 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도를 설명한다. n차원 공간상의 두 벡터를 각각 $X=(x_1, x_2, x_3, ..., x_n), Y=(y_1, y_2, y_3, ..., y_n)$라 하자. 이들에 대한 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도는 각각 다음과 같이 정의된다.

3.3.1 유클리디안(Euclidean) 유사도

유클리디안 유사도는 두 벡터 간의 유클리디안 거리를 사용하여 유사도를 측정한다. 유클리디안 거리는 두 벡터 간의 직선거리를 나타내며, 두 벡터를 구성하는 각각의 원소들 간의 차이를 모두 고려하고자할 때 많이 사용되는 척도이다. 두 벡터 X와 Y 간의 유클리디안 유사도 $Sim_E(X, Y)$는 다음과 같이 정의된다. 여기서 MAX는 X와 Y가 가질 수 있는 유클리디안 거리의 최대값을 나타낸다. 유클리디안 유사도는 유클리디안 거리가 커질수록 작은 값을 가지며 0에서 1 사이의 값을 가진다.

$$S_E(X, Y) = 1 - \frac{\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{MAX}$$

3.3.2 코사인(cosine) 유사도

코사인 유사도는 두 벡터가 이루는 내각의 크기로 유사도를 측정한다. 코사인 유사도는 문서들을 각각 해당 문서에 나타나는 단어들의 빈도수들로 구성된 벡터로 표현했을 때, 이들 간의 유사도를 구하는 데 많이 사용된다. 이 경우 코사인 유사도는 유사한 의미나 주제를 가지는 문서를 찾아내는데 매우 효과적인 것으로 알려져 있으며, 데이터 마이닝이나 정보 검색(information retrieval) 분야에서 많이 사용된다. 두 문서를 나타내는 벡터 간의 각도가 0°로 완전히 동일한 경우 코사인 유사도는 1의 값을 가지며, 90°인 경우에는 0의 값을 가진다. 벡터의 모든 원소가 0 이상의 값을 가지는 경우 코사인 유사도는 0에서 1 사이의 값을 가지며, 값이 1에 가까울수록 유사도가 높다. 코사인 유사도는 벡터의 크기는 고려하지 않고 두 벡터가 이루는 각도만을 이용하여 유사도를 구하기 때문에 비교하고자 하는 두 문서의 길이가 많이 차이 나는 경우에도 좋은 성능을 보인다. 두 벡터 X와 Y 간의 코사인 유사도 $S_C(X, Y)$는 다음과 같이 정의된다.

$$S_C(X, Y) = \frac{X \bullet Y}{|X| \times |Y|}$$

3.3.3 확장된 자카드(Jaccard) 유사도

확장된 자카드 유사도는 타니모토 계수(Tanimoto coefficient)[14]로도 알려져 있는 확장된 자카드 계수(extended Jaccard coefficient)를 사용하여 유사도를 측정한다. 확장된 자카드 계수는 두 집합 간의 유사도를 나타내며, 기존의 자카드 계수를 연속적인 수에 적용 가능하도록 확장한 것이다. 두 벡터의 각도와 상대적인 거리를 모두 고려하며, 벡터를 구성하는 원소들 간의 합집합과 교집합 간의 비율을 나타낸다. 확장된 자카드 계수는 0에서 1사이의 값을 가지며 1에 가까울수록 유사하다. 두 벡터 X와 Y 간의 확장된 자카드 유사도 $S_J(X, Y)$는 다음과 같이 정의된다.

$$S_J(X, Y) = \frac{X \bullet Y}{|X|^2 + |Y|^2 - X \bullet Y}$$

3.3.4 카테고리-단어 간 유사도 행렬 생성 예

앞 절에서 설명한 유사도 측정 방법을 사용하여 카테고리들을 각 단어들과 리뷰 내의 각 단어들 간의 유사도를 나타내는 행렬을 생성한다. 이 행렬의 원소 s_{ij}는 카테고리들을 나타내는 단어들 중 i번째 단어와 리뷰에 나타나는 단어들 중 j번째 단어 간의 거리를 나타내며 0에서 1 사이의 값을 가진다. 카테고리를 나타내는 단어와

리뷰에 나타나는 단어 간의 유사도가 높을수록 1에 가까운 값을 가지며, 유사도가 낮을수록 0에 가까운 값을 가진다. <Figure 6>은 카테고리-리뷰 단어 간 유사도 행렬의 예로서, 카테고리들을 나타내는 단어 중 '결제'와 리뷰에 나타나는 단어 중 '가격'과 '현질'은 관련이 높기 때문에 1에 가까운 유사도를 가지지만, 상대적으로 관련이 낮은 '채팅', '처벌'과 같은 단어는 낮은 유사도를 가진다.

	게임	가격	가끔	…	채팅	처벌	현질
결제	0.75	1	0.45	…	0.23	0.25	0.98
구입	0.67	0.97	0.53	…	0.35	0.16	0.93
…	…	…	…	…	…	…	…
욕	0.34	0.18	0.36	…	0.82	0.89	0.23
처벌	0.36	0.27	0.28	…	0.71	0.86	0.17

〈Figure 6〉 Category-Word Similarity Matrix

지금까지 설명한 카테고리-단어 간 유사도 행렬 외에도 각 리뷰에 어떤 단어들이 포함되어 있는지를 나타내는 리뷰-단어 존재 행렬을 구축한다. 리뷰-단어 존재 행렬의 각 행은 리뷰를 나타내고 각 열은 단어를 나타내며, 해당 리뷰에 해당 단어의 등장 여부에 따라 0 또는 1의 값을 가진다. <Figure 7>은 원 리뷰 데이터를 리뷰-단어 행렬로 변환한 예로서, 리뷰들을 단어 등장 여부에 따라 0과 1로 구성된 행렬로 바꾼 결과를 나타낸다.

	게임	가격	가끔	…	채팅	처벌	현질
Review 1	1	0	1	…	0	0	0
Review 2	0	1	0	…	0	1	1
Review 3	1	1	0	…	0	1	0
Review 4	0	1	1	…	1	0	1
Review 5	0	0	1	…	0	1	0
…	…	…	…	…	…	…	…

〈Figure 7〉 Review-Word Presence Matrix

3.4 리뷰 별 카테고리 분류 단계

카테고리-단어 간 유사도 행렬(<Figure 6>)과 리뷰-단어 존재 행렬(<Figure 7>)이 구축되면 마지막으로 리뷰 별로 각 카테고리에 대한 스코어를 구한다. 어떤 리뷰의 각 카테고리에 대한 스코어를 구하기 위해서는 먼저 리뷰-단어 존재 행렬에서 해당 리뷰를 나타내는 행을 가져온다. 그 후 해당 행을 카테고리-단어 간 유사도 행렬의 각 행들과 내적(inner product)하여 해당 리뷰와 각 카테고리 단어와의 총 유사도를 구한다. 마지막으로 각 카테고리를 나타내는 모든 단어들에 대한 해당 리뷰의 총 유사도를 모두 더하여 해당 리뷰의 해당 카테고리에 대한 최종 스코어를 구한다. 이 때 리뷰는 가장 큰 최종 스코어를 가지는 카테고리로 분류된다.

하지만 어떤 리뷰는 둘 이상의 주제를 내포하고 있을 수도 있다. 이 경우를 처리하기 위해 가장 높은 최종 스코어 값과 그 다음 순위의 최종 스코어 값의 차이가 일정 임계값보다 작은 경우, 그 범주가 모호하다고 판단하여 두 번째 카테고리로도 분류한다.

3.4.1 리뷰 별 카테고리 분류의 예

<Figure 8>은 앞에서 설명한 방법에 따라 각 리뷰에 대해 각 카테고리에 대한 최종 스코어를 계산한 예이다. 리뷰 1의 경우 '시스템' 카테고리에 대한 최종 스코어가 0.7로 가장 높은 값을 가지며, 그 다음으로 최종 스코어가 높은 카테고리는 '서버' 카테고리이다. 한편 리뷰 2의 경우 '연출' 카테고리가 0.75로 가장 높은 최종 스코어를 가지며, 그 다음으로 최종 스코어가 높은 카테고리는 '구성' 카테고리이다.

	결제	계정	구성	서버	시스템	연출	캐릭터	기타
Review 1	0.17	0.04	0.05	0.32	**0.7**	0.2	0.11	0.02
Review 2	0.03	0.17	**0.44**	0.05	0.33	**0.75**	0.02	0.01
Review 3	**0.9**	0.03	0.04	0.2	0.46	0.34	0.01	0.03
Review 4	0.21	0.14	**0.58**	**0.77**	0.1	0.09	0.18	0.27
Review 5	0.07	**0.79**	0.21	0.3	0.01	0.15	0.24	0.07
...	...	...	...	...	...	...	...	...

〈Figure 8〉 Final Scores for Each Category

3.4.2 다중 카테고리 분류 기준

앞서 설명한 바와 같이 본 시스템은 가장 높은 최종 스코어 값과 그 다음 순위의 최종 스코어 값의 차이가 일정 임계값보다 작은 경우 해당 리뷰를 두 카테고리 모두로 분류한다. 예를 들어 임계값이 0.35인 경우, <Figure 8>의 예에서 리뷰 1은 가장 높은 최종 스코어인 0.7과 그 다음 순위의 최종 스코어인 0.32의 차이가 0.38로 임계값보다 크므로 해당 리뷰는 '시스템' 카테고리로만 분류된다. 이에 비해 리뷰 2의 경우 가장 높은 최종 스코어를 가지는 '구성'과 '연출'카테고리의 최종 스코어 값의 차이가 0.31(0.75-0.44)로 임계값보다 작으므로 해당 리뷰는 '구성'과 '연출' 두 카테고리로 모두 분류된다.

본 논문에서는 리뷰 분류의 정확도를 높이기 위하여 이 임계값을 0에서 시작하여 점차 높여가며 성능을 측정하여 최적의 임계값을 선택하였다. 제4.2절에서는 임계값 변화에 따른 성능 평가 결과를 보인다.

4. 실험 결과

본 장에서는 본 논문에서 개발한 게임 리뷰 자동 분류 시스템의 성능을 최대화하기 위해, 분류 성능에 영향을 미치는 주요 요소 2가지를 변화시켜가며 성능을 측정한 결과를 보인다. 첫 번째 요소는 벡터 간 유사도 측정 방법으로서, 제3.3.1절에서 설명한 대표적인 유사도 측정 방법인 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도를 사용했을 때의 분류 성능을 비교한 결과를 보인다. 두 번째 요소는 주어진 리뷰를 둘 이상의 카테고리로 분류할 때 사용되는 임계값의 크기로서, 임계값을 변화시켜가며 분류 성능의 변화를 측정하였다. 본 논문에서는 실험 결과 가장 좋은 성능을 보이는 유사도 측정 방법과 임계값을 실제 시스템 구현에 채택하였다.

4.1 실험 환경 및 방법

본 논문에서 구현한 리뷰 자동 분류 시스템은 Gensim 라이브러리와 Python 3.7을 이용하여 구현되었다. 시스템의 분류 정확도는 수집한 리뷰 데이터 중 1,000개를 임의로 추출하였으며, 형태소 분석 결과 6개 이상의 유의미한 단어를 가진 리뷰만을 대상으로 정확도를 측정하였다. 추출된 리뷰는 먼저 가장 가깝다고 판단되는 카테고리를 부여하여 정답으로 지정한 후, 시스템에서 판정한 카테고리와 몇 개나 동일한지 확인하였다. 추출된 리뷰에 따라 시스템의 성능이 달라지기 때문에 이 과정을 5번 반복한 후 평균값을 취하여 사용하였다.

4.2 실험 결과

본 절에서는 다양한 벡터 간 유사도 측정 방법에 따른 성능 비교 결과와 다중 분류 시 사용되는 임계값 변화에 따른 성능 측정 결과를 보인다.

<Figure 9>는 실제 리뷰에 대해 카테고리 분류를 수행한 실제 실험 결과의 일부를 보여준다. 다중 분류를 위해 사용된 임계값에 따라 각 리뷰를 하나 또는 두 개의 카테고리로 분류했음을 볼 수 있다.

Review	Category
용량 너무 커서 설치 안돼요 ㅡ.ㅡ	시스템
시작하고 업데이트 파일 설치할 때 한 줄 치면 무한로딩;;;;	서버, 시스템
신캐 사기캐인듯ㅋㅋㅋㅋㅋ 개좋아ㅠㅠ	캐릭터
결제한거 빨리 환불해주세요ㅡㅡ 그리고 렉걸리는 것도 빨리 고쳐줘요!!	결제, 서버
이번 퀘스트 너무 어려워요ㅠㅠ 난이도 조절 좀	구성
겜 퀄리티는 좋은데 채팅 창 욕설 좀 관리해주세요	연출, 기타

〈Figure 9〉 Examples of Review Classification

4.2.1 유사도 측정 방법 비교 실험

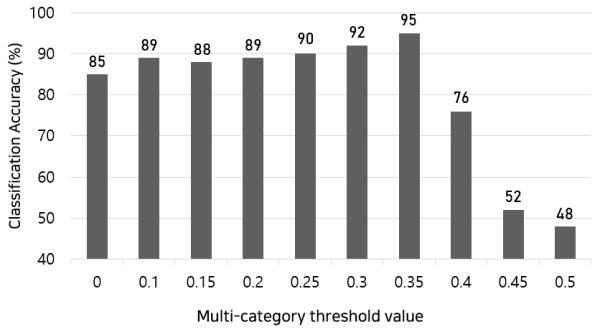
<Figure 10>은 카테고리-단어 간 유사도 행렬 계산에 사용된 유사도 측정 방법별 분류 정확도를 나타낸다. 1,000개의 리뷰 중 형태소 분석 결과 의미 있는 단어를 6개 이상 포함하는 리뷰는 686개로 해당 리뷰에 대해서만 정확도를 측정하였다. 실험 결과 유클리디안 유사도의 경우, 다른 두 방법에 비해 좋지 않은 성능을 보이며 리뷰 수가 상대적으로 '구성'과 '서버'카테고리에만 집중되어 있음을 볼 수 있다. 코사인 유사도와 확장된 자카드 유사도는 둘 다 높은 정확도를 보이며, 이 중에서도 확장된 자카드 유사도를 사용했을 때 평균 95%로 가장 높은 정확도를 보였다. 실험 결과 본 응용에서는 두 벡터의 유사도는 각 원소 값의 차이보다는 두 벡터의 방향성의 차이가 더 큰 영향을 미치는 것으로 판단된다. 하지만 두 벡터의 방향성만을 고려하는 코사인 유사도와는 달리, 확장된 자카드 유사도는 두 벡터 간의 방향성과 크기 차이를 모두 고려하기 때문에 본 응용에서 가장 좋은 성능을 보인 것으로 판단된다.

	Euclidean similarity	Cosine similarity	Extended Jaccard similarity
결제	53	41	40
계정	13	80	62
구성	232	90	103
서버	189	141	155
시스템	31	74	77
연출	89	145	118
캐릭터	79	105	121
기타	0	10	10
Accuracy	85%	92%	95%

〈Figure 10〉 Evaluation Results for Different Similarity Measures

4.2.2 임계값 변화 실험

본 실험에서는 리뷰의 다중 분류에 사용되는 임계값을 변화시켜가며 분류 성능의 변화를 측정하였다. 이 때 유사도 측정 방법으로는 앞 실험



〈Figure 11〉 Evaluation Result for Various Threshold Values

에서 가장 높은 분류 정확도를 보이는 확장된 자카드 유사도를 사용하였다. 다중 분류를 위한 임계값은 0을 시작으로 0.05 크기씩 증가시켜가며 성능 변화를 측정하였다. <Figure 11>은 임계값 변화에 따른 실험 결과를 나타낸다. <Figure 11>에서 볼 수 있듯이 0을 시작으로 임계값이 0.35까지 증가할 때는 점차 높은 정확도를 보이며, 임계치가 0.35일 때 95%의 가장 높은 정확도를 보인다. 이는 임계값이 적절한 크기를 유지하는 경우, 임계값의 사용이 리뷰에 담긴 둘 이상의 주제를 파악하는데 효과적인 방법임을 나타낸다. 하지만 임계값이 0.35를 넘어가면 분류 정확도가 점차 낮아지며, 0.4일 때를 기점으로 급감한다. 이는 임계값이 너무 커지면 리뷰가 최종 스코어가 매우 낮은 카테고리로도 분류되기 때문이다.

4.2.3 시스템 구현

지금까지 설명한 유사도 측정 방법 및 임계값 변화에 따른 성능 비교 실험 결과, 확장된 자카드 유사도가 95%의 정확도로 유클리디안 유사도와 코사인 유사도에 비하여 높은 성능을 보였다. 또한 확장된 자카드 유사도를 사용했을 때, 다중 분류를 위한 임계값이 0.35일 때 가장 높은 정확도를 보였다. 이에 따라 본 논문에서는 임베딩된 단어 간의 유사도를 계산할 때는 확장된 자카드 유사도를 사용하였으며, 다중 분류를 위한 임계값은 0.35로 설정하여 게임 리뷰 자동 분류 시스템을 구현하였다.

5. 결 론

본 논문에서는 모바일 게임 리뷰를 수집하고, 수집된 리뷰들을 word2vec 단어 임베딩 모델을 활용하여 관련된 카테고리들로 자동 분류하는 시스템을 개발하였다. 최신 트렌드 분석과 신속한 사용자 요구사항 반영이 중요한 게임 소프트웨어에서는 다양한 주제들을 포함하고 있는 리뷰들을 주제별로 빠르게 파악하는 것이 중요하다. 따라서 입력된 리뷰들이 각각 어떤 카테고리에 해당되는 것인지, 리뷰의 입력과 동시에 시스템이 자동적으로 파악 및 분류해 줄 수 있다면 게임의 신속한 개선에 큰 도움이 될 것이다.

본 논문에서는 리뷰 분류의 정확도를 최대화하기 위해 다양한 벡터 간 유사도 측정 방법들을 고려하였다. 이를 위해 현재 대표적인 유사도 측정 방법인 유클리디안 유사도, 코사인 유사도, 확장된 자카드 유사도를 사용했을 때의 성능을 각각 측정하였으며, 이들 중 최고의 정확도를 보이는 확장된 자카드 유사도를 시스템 구현에 사용하였다. 또한 리뷰의 다중 분류에 사용되는 임계값을 다양하게 변화시키며 가장 좋은 성능을 보이는 값을 시스템 구현에 사용하였다. 본 논문에서 수집한 리뷰 데이터는 명시적인 정답이 주어지지 않은 데이터로, 리뷰의 구성 단어만을 활용하여 리뷰의 카테고리를 분류하는 시스템이다. 실험을 통하여 확장된 자카드 유사도와 최적의 임계값을 사용했을 때, 본 논문에서 개발한 시스템은 최대 95%까지의 분류 정확도를 보임을 확인하였다.

본 논문에서 시스템 구현 시 단어 임베딩 모델로 word2vec을 사용하였다. 추후 연구에는 다른 임베딩 모델들과의 성능 비교를 통하여 단어 임베딩 성능 향상과 고속화 방법에 대하여 연구할 계획이다.

References

[1] Chevalier, J. A. and Mayzlin, D., “The Effect of Word of Mouth on Sales: Online Book Reviews,” Journal of Marketing Research, Vol. 43, No. 3, pp. 345-354, 2006.

[2] DMC Report, “2018 Mobile Game and Mobile Game Advertising Market Size and Status,” https://www.dmcreport.co.kr/content/ReportView.php?type=Market&id=13368&gid=3.

[3] Duan, W., Gu, B., and Whinston, A. B., “The dynamics of online word-of-mouth and product sales－An empirical investigation of the movie industry,” Journal of Retailing, Vol. 84, No. 2, pp. 233-242, 2008.

[4] Huang, A., “Similarity measures for text document clustering,” Proceedings of the 6th New Zealand Computer Science Research Student Conference, pp. 49-56, 2008.

[5] Kim, J., Byeon, H., and Lee, S. H., “Enhancement of User Understanding and Service Value Using Online Reviews,” The Journal of Information Systems, Vol. 20, No. 2, pp. 21-36, 2011.

[6] Korea Creative Content Agency, “2018 Korea Game White Paper,” http://www.kocca.kr/cop/bbs/view/B0000146/1837580.do.

[7] Kostyra, D. S., Reiner, J., Natter, M., and Klapper, D., “Decomposing the Effects of Online Customer Reviews on Brand, Price, and Product Attributes,” International Journal of Research in Marketing, Vol. 33, No. 1, pp. 11-26, 2015.

[8] Lee, D. H. and Kim, K. H., “Web Site Keyword Selection Method by Considering Semantic Similarity Based on Word2Vec,” The Journal of Information Systems, Vol. 23, No. 2, pp. 83-96, 2018.

[9] Lilleberg, J., Zhu, Y., and Zhang, Y., “Support vector machines and Word2vec for text classification with semantic features,” IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), 2015.

[10] Mikolov, T., Chen, K., Corrado, G., and Dean, J., “Efficient Estimation of Word Representations in Vector Space,” ICLR Workshop Paper, 2013.

[11] Mikolov, T., Yih, W., and Zweig, G., “Linguistic Regularities in Continuous Space Word Representations,” Proceedings of NAACL-HLT, 2013.

[12] Setty, V., Kreitz, G., Vitenberg, R., van Steen, M., Urdaneta, G., and Gimåker, S., “The hidden pub/sub of Spotify,” Proceedings of the 7th ACM International Conference on Distributed Eventbased Systems, pp. 231-240, 2013.

[13] Sudeep Das, “Making meaningful restaurant recommendations at opentable,” https://de.slideshare.net/SudeepDasPhD/recsys-2015-making-meaningfulrestaurant-recommendations-at-opentable, 2015.

[14] Tanimoto, T. T., “An elementary mathematical theory of classification and prediction,” IBM Report (November, 1958), cited in: G. Salton, Automatic Information Organization and Retrieval, p. 238, 1968.

[15] Wensen, L., Zewen, C., Jun, W., and Xiaoyi, W., “Short text classification based on Wikipedia and Word2vec,” 2016 2nd IEEE International Conference on Computer and Communications (ICCC), 2016.

[16] Yeon, J. H., Lee, D. J., Shim, J. H., and Lee, S. G., “Product Review Data and Sentiment Analytical Processing Modeling,” The Journal of Society for e-Business Studies, Vol. 16, No. 4, pp. 125-137, 2011.

[17] Zhang, D., Xu, H., Su, Z., and Xu, Y., “Chinese comments sentiment classification based on word2vec and SVMperf,” Expert System with Applications, Vol. 42, No. 4, pp. 1857-1863, 2015.

[18] Zhu, F. and Zhang, X. M., “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” Journal of Marketing, Vol. 74, No. 2, pp. 133-148, 2010.

저 자 소 개

양유정 (E-mail: diddbwjd96@sookmyung.ac.kr)
2019년 숙명여자대학교 소프트웨어학부 (학사)
2019년～현재 숙명여자대학교 컴퓨터과학과 석사과정
관심분야 데이터마이닝

이보현 (E-mail: lbhsos29@gmail.com)
2015년～현재 숙명여자대학교 소프트웨어학부 학사과정
관심분야 데이터마이닝, 데이터베이스

김진실 (E-mail: kjinsil0819@gmail.com)
2019년 숙명여자대학교 소프트웨어학부 (학사)
관심분야 빅데이터

이기용 (E-mail: kiyonglee@sookmyung.ac.kr)
1998년 KAIST 전산학과 (학사)
2000년 KAIST 전산학과 (석사)
2006년 KAIST 전산학과 (박사)
2006년～2008년 삼성전자 소프트웨어연구소 책임연구원
2008년～2010년 KAIST 전산학과 연구조교수
2010년～현재 숙명여자대학교 소프트웨어학부 부교수
관심분야 데이터베이스, 데이터마이닝, 빅데이터, 데이터스트림