

평점과 리뷰 텍스트 감성분석을 결합한 추천시스템 향상 방안 연구

How to improve the accuracy of recommendation systems : Combining ratings and review texts sentiment scores

저자 (Authors)	현지연, 유상이, 이상용 Jiyeon Hyun, Sangyi Ryu, Sang-Yong Tom Lee
출처 (Source)	지능정보연구 25(1) , 2019.3, 219-239(21 pages) Journal of Intelligence and Information Systems 25(1) , 2019.3, 219-239(21 pages)
발행처 (Publisher)	한국지능정보시스템학회 Korea Intelligent Information Systems Society
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07996729
APA Style	현지연, 유상이, 이상용 (2019). 평점과 리뷰 텍스트 감성분석을 결합한 추천시스템 향상 방안 연구. 지능정보연구, 25(1), 219-239
이용정보 (Accessed)	통계청 125.128.71.*** 2020/07/23 14:14 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

평점과 리뷰 텍스트 감성분석을 결합한 추천시스템 향상 방안 연구

현지연

한양대학교 경영대학
비즈니스인포매틱스학과 석사
(wildus5578@hanyang.ac.kr)

유상이

한양대학교 경영대학
경영학과 석사
(ryussang2@hanyang.ac.kr)

이상용

한양대학교 경영대학
교수
(tomlee@hanyang.ac.kr)

개인에게 맞춤형 서비스를 제공하는 것이 중요해지면서 개인화 추천 시스템 관련 연구들이 끊임없이 이루어지고 있다. 추천 시스템 중 협업 필터링은 학계 및 산업계에서 가장 많이 사용되고 있다. 다만 사용자들의 평점 혹은 사용 여부와 같은 정량적인 정보에 국한하여 추천이 이루어져 정확도가 떨어진다는 문제가 제기되고 있다. 이와 같은 문제를 해결하기 위해 현재까지 많은 연구에서 정량적 정보 외에 다른 정보들을 활용하여 추천 시스템의 성능을 개선하려는 시도가 활발하게 이루어지고 있다. 리뷰를 이용한 감성 분석이 대표적이지만, 기존의 연구에서는 감성 분석의 결과를 추천 시스템에 직접적으로 반영하지 못한다는 한계가 있다. 이에 본 연구는 리뷰에 나타난 감성을 수치화하여 평점에 반영하는 것을 목표로 한다. 즉, 사용자가 직접 작성한 리뷰를 감성 수치화하여 정량적인 정보로 변환해 추천 시스템에 직접 반영할 수 있는 새로운 알고리즘을 제안한다. 이를 위해서는 정성적인 정보인 사용자들의 리뷰를 정량화 시켜야 하므로, 본 연구에서는 텍스트 마이닝의 감성 분석 기법을 통해 감성 수치를 산출하였다. 데이터는 영화 리뷰를 대상으로 하여 도메인 맞춤형 감성 사전을 구축하고, 이를 기반으로 리뷰의 감성점수를 산출한다. 본 논문에서 사용자 리뷰의 감성 수치를 반영한 협업 필터링이 평점만을 고려하는 전통적인 방식의 협업 필터링과 비교하여 우수한 정확도를 나타내는 것을 확인하였다. 이후 제안된 모델이 더 개선된 방식이라고 할 근거를 확보하기 위해 paired t-test 검증을 시도했고, 제안된 모델이 더 우수하다는 결론을 도출하였다. 본 연구에서는 평점만으로 사용자의 감성을 판단한 기존의 선행연구들이 가지는 한계를 극복하고자 리뷰를 수치화하여 기존의 평점 시스템보다 사용자의 의견을 더 정교하게 추천 시스템에 반영시켜 정확도를 향상시켰다. 이를 기반으로 추가적으로 다양한 분석을 시행한다면 추천의 정확도가 더 높아질 것으로 기대된다.

주제어 : 감성 분석, 감성 사전, 추천 시스템, 텍스트 마이닝, 협업 필터링

논문접수일 : 2019년 1월 13일 논문수정일 : 2019년 3월 17일 게재확정일 : 2019년 3월 29일
원고유형 : 학술대회(급행) 교신저자 : 이상용

1. 연구배경 및 목표

전 세계적으로 하루에 생성되는 정보의 양은 2조 5000억 MB 이상이 생산된다고 추산한다. 생성 속도로 보면 매년 60%씩 증가하는 추세이다. 오프라인 저장소에 저장할 수 있는 물리적인 한

계를 벗어나 온라인으로 접근이 가능해지면서 엄청난 양의 데이터를 저장할 수 있게 되었다. 접근 가능한 정보의 양이 증가하면서 날마다 새롭게 생성되는 정보의 홍수 속에 사람들은 극도의 피로감을 느끼고 있다. 모든 정보를 볼 수 없기에 자신의 선호에 맞는 정보를 찾아 선택하는

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066740).

것에 많은 어려움이 따른다. 이러한 배경에서 수많은 정보 가운데 필요 없는 정보를 덜어내 개인의 기호에 맞게 정보를 선택할 수 있는 정보를 제공해주는 추천 시스템의 중요성이 점차 커지고 있다.

개인에게 맞춤형 정보를 제공해주기 위한 기존 추천 시스템 관련 연구들은 사용자들의 구매 여부, 제품에 대한 평점(Rating), 방문 접속 횟수 등과 같이 계량화하기 용이한 정형 데이터를 활용한 분석이 주류를 이뤄왔다(Ricci et al., 2011). 그 중 사용자 선호를 나타내는 가장 대표적인 지표로써 평점 데이터를 추천 시스템에 사용해 왔다.

다만, 최근 사용자들의 선호를 나타내는 지표인 평점에 국한하여 추천하는 것은 정확도가 떨어진다는 문제가 제기되고 있다(Jeon et al., 2015). 사용자 개인의 정교한 선호를 고려하지 않았기 때문에 추천 시스템 개발에 한계가 있다는 문제점이 제기되었다.

<Figure 1>은 영화를 시청한 특정 사용자(gmlx****)의 텍스트 리뷰와 평점이다. 해당 사용자는 영화 ‘쿵푸팬더’에 대해 “와우!! 진짜 재

밌어요. 짱 귀엽구...ㅎㅎ 100점 줘도 더주고 싶은 심정이네요. ㅎㅎ” 라는 리뷰를 남긴 반면, 영화 ‘해운대’에 대해서는 “배우가 조금 아쉽지만 굿 불만하다” 라는 리뷰를 남겼다. 평점은 10점으로 동일하지만, 같은 리뷰 작성자가 상이한 영화에 평점을 부여할 때 동일한 정량적 평가 점수를 부여함에도 불구하고 텍스트 안에 담긴 감성의 강도(Intensity)의 정도는 다르다는 것을 알 수 있다. 이러한 점은 추천 시스템의 정확도를 떨어뜨리는 중요한 요인이 될 수 있다.

평점을 사용한 협업 필터링은 활용도가 높고, 명시적인 정보를 수리적인 체계로 처리하기 쉽다는 장점이 있어서 많이 사용되어왔지만, 평점 정보가 사용자의 선호를 제대로 반영하고 있는지 여부에 대한 의문이 제기되고 있다. 최근 이러한 한계를 완화시키기 위해 사용자 선호를 반영한 협업 필터링 모델을 구축하기 위한 다양한 시도들이 존재해왔다.

대표적으로 많은 연구자들은 텍스트 리뷰 기반의 평점 예측에 관심을 기울이고 있다. 사용자가 특정 아이템에 대해 가지고 있는 선호를 보다 상세하게 알 수 있기 때문에 리뷰는 추천 시스템

Index	Rating	Review	User_ID
2863396	★★★★★ 10	고사:피의 중간고사 이거 보고 가위눌려서 굿	gmlx** 08.08.11
2863309	★★★★★ 10	다크 나이트 이영화는 정말 우리의 상상력을 뛰어넘는다	gmlx** 08.08.11
2681329	★★★★★ 10	쿵푸팬더 와우!! 진짜 재밌어요, 짱귀엽구.. ㅎㅎ 100점 줘도 더주고 싶은 심정이네요. ㅎㅎ	gmlx** 08.06.05
364444	★★★★★ 10	차우 여자가 제일 무섭네 ㄷㄷㄷ 진짜 평점 더주고싶다!! 올려름에 한국영화가 대세	gmlx** 06.07.18
485214	★★★★★ 10	해운대 배우가 조금 아쉽지만 굿 불만하다	gmlx** 09.11.08

<Figure 1> Ratings and reviews for movies rated by user gmlx****

에서 활용도가 높다. 사용자의 텍스트 리뷰가 추천 시스템 구축에 중요한 요소로의 역할을 하고 있다는 것은 Park(2014)과 Deng et al.(2018)의 연구들을 통해 입증되었다.

이러한 배경에서 본 논문에서는 정량 데이터인 평점만을 활용해오던 추천 시스템의 정확도를 개선하기 위해, 정성 데이터인 리뷰를 활용하여 리뷰 속에 나타나 있는 사용자의 감정을 수치화하여 평점을 정교하게 보정하는 방법으로 추천 시스템의 성능을 향상시키는 모델을 제안한다. 이를 통해 기존 연구에서 시도하지 않았던 리뷰를 감성점수화 하여 평점에 적용해 추천 시스템의 정확도를 향상시키는 알고리즘을 새롭게 제안한다.

논문의 구성은 다음과 같다. 2장에서는 추천 시스템과 감성 분석 관련 이론적 배경에 대해 서술하였고, 이와 관련한 다양한 선행 연구들을 살펴본다. 그 후 선행연구에 대한 고찰을 통해 본 연구의 연구 방향을 제시한다. 3장에서는 본 연구에서 제안하는 사용자 리뷰의 감성 수치를 반영한 추천 시스템의 연구 모형을 제시하여 각 단계별 수행 절차를 서술하였다. 4장에서는 제안하는 알고리즘의 성능을 검증하기 위한 평가 방법과 결과를 기술한다. 마지막으로 5장에서는 본 연구의 의의를 논한 뒤, 본 연구의 한계 및 향후 연구방향에 대하여 제시한다.

2. 이론적 배경 및 선행 연구

2.1 추천시스템

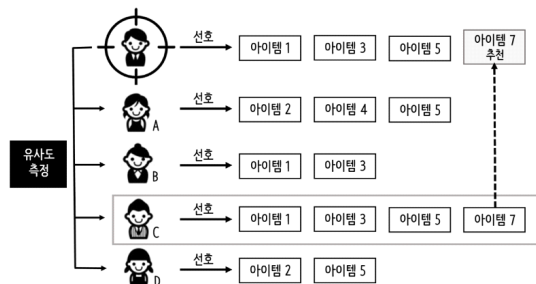
추천 시스템은 고객이 관심을 가질 것으로 예상되는 아이템에 관한 정보나 인구 통계학적 정

보, 과거 구매 패턴 등을 토대로 고객의 선호에 맞는 아이템을 추천해주는 것이다. 넷플릭스(Netflix)와 아마존(Amazon)이 추천시스템으로 성과를 거두면서 사용자의 취향에 맞는 아이템을 추천하기 위한 노력들이 국내에서도 왓차(Whatcha)의 영화 추천 시스템, 네이버 인공지능 에어스(AiRS)를 통한 메인 뉴스 추천, 실시간으로 사용자의 반응을 분석하여 콘텐츠를 추천하는 카카오의 루빅스(RUBICS) 등에서 활발하게 이루어지고 있다. 이렇게 추천 시스템은 우리 주변에서 빈번하게 사용되고 있고, 이와 관련한 연구들은 끊임없이 진행되고 있다. 가장 많이 활용되고 있는 대표적인 추천 알고리즘으로는 협업 필터링 방식이 학계 및 산업계에서 사용되고 있다.

협업 필터링은 Goldberg et al.(1992)에 의하여 개념이 처음으로 소개되어진 방식으로 특정 아이템에 대한 선호도가 유사한 사용자들은 다른 아이템에 대해서도 비슷한 선호도를 나타낼 것이라는 기본 가정을 바탕으로 사용자 혹은 아이템 간 유사도를 기반으로 선호도를 예측하는 방법이다. Breese et al.(1998)는 협업 필터링 방식을 메모리 기반 알고리즘과 모델 기반 알고리즘으로 분류하였다. 메모리 기반 알고리즘은 Neighborhood Model라고도 전체 사용자-아이템 매트릭스를 생성하여 사용자 혹은 아이템 기반으로 유사한 사용자 혹은 아이템을 찾은 뒤 사용자가 부여할 것이라 예상되는 평점을 추론하는 알고리즘이다. 이는 사용자 기반의 협업 필터링(User-based collaborative filtering)과 아이템 기반의 협업 필터링(Item-based collaborative filtering)으로 나누어진다.

사용자 기반의 협업 필터링은 Group-Lens에서 처음으로 소개되었고 Ringo 음악 추천, Facebook

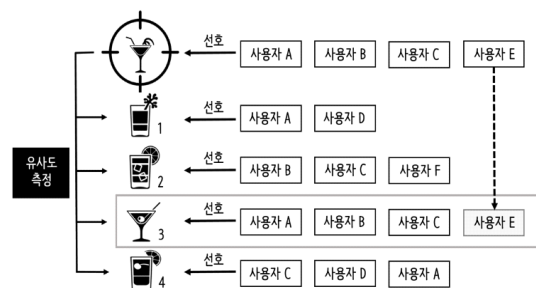
친구 추천 등에 사용된다. 이는 사용자가 부여한 평점 정보를 이용하여 사용자와 유사한 성향을 갖는 이웃 사용자를 선별한 뒤, 선별된 이웃들이 공통적으로 선호하는 아이템을 고객에게 추천해주는 방식이다.



〈Figure 2〉 User-based Collaborative Filtering

기본적인 개념은 <Figure 2>와 같다. 추천 대상 사용자와 취향이 가장 유사한 사용자는 ‘아이템 1’, ‘아이템 2’, ‘아이템 5’를 구매한 ‘사용자 C’이다. ‘사용자 C’의 정보를 기반으로 ‘사용자 C’는 선호하였으나 추천 대상 사용자는 아직 선호하지 않은 ‘아이템 7’을 선정하여 추천 대상 사용자에게 추천하는 것이다.

아이템 기반 협업 필터링은 Sarwar et al.(2001)에 의하여 학계에 소개되어 Youtube와 Netflix의 비디오 추천, 아마존 제품 추천 등에 사용된다.



〈Figure 3〉 Item-based Collaborative Filtering

기본적인 개념은 <Figure 3>와 같다. 추천 대상 아이템을 기준으로 유사도가 가장 높은 아이템을 선택한다. 선택된 ‘아이템 3’은 추천 아이템을 구매하지 않은 ‘사용자 E’에게 추천 대상 아이템으로 추천하는 것이다.

모델 기반 알고리즘은 메모리 기반의 협업 필터링 방식의 과정은 기본적으로 하며 군집화, 분류, 예측의 단계에서 기계학습 또는 데이터 마이닝 기법을 활용하는 것이다(Son et al., 2015). 과거 사용자의 평점을 기반으로 사용자를 모델링하여 사용자가 평점을 주지 않은 새로운 아이템에 부여할 평점을 예측하는 방식으로 Matrix Factorization, Clustering models 등으로 분류할 수 있다. Matrix Factorization 방식은 사용자와 아이템 사이의 직접적인 관계가 아닌 잠재 요인을 통하여 평점을 예측하는 방식이다. 대표적 알고리즘으로 SVD(Singular Value Decomposition), SVD++가 사용된다. Clustering model은 개체간의 유사성을 척도로 유사한 특성을 가진 개체들이 함께 그룹 지어지도록 하는 방법이다. 대표적으로 사용되는 알고리즘으로는 k-means, DBSCAN을 사용한다.

2.2 사용자 리뷰를 반영한 추천 시스템

사용자 리뷰를 반영한 추천 시스템의 연구는 2000년대 중반 학계에 소개되었다. Leung et al.(2006)은 사용자 리뷰의 긍/부정에 대한 감성 방향과 강도를 추정하는 모형을 개발하여 이를 기반으로 협업 필터링 추천 시스템을 제안하였다. Garcia-Cumbreras et al.(2013)는 리뷰를 감성 분석하여 해당 사용자가 낙관주의자인지 비관주의자인지를 분류하고, 각 집단 별로 협업 필터링을 수행하여 사용자 리뷰를 집단 분류의 기준으로

로 사용하여 이전의 연구들과 차별화된 접근 방식을 시도하였다. Choi et al.(2016)의 연구에서는 평점과 연관성이 높은 단어들을 추출하여 리뷰에서 추출된 단어들의 빈도를 기반으로 긍정과 부정 외에도 중립적인 의견과 같은 다 범주 분류에 사용되었다. 이와 같은 연구는 긍정 및 부정의 이분법으로 분류하는 연구와 중립이 포함되기도 하는 다중 감성으로의 분류까지 현재 까지도 많은 연구가 진행되고 있다(Liu, 2015). 나아가 Oh et al.(2015)의 연구에서는 리뷰 속 의견과 평가에 관련된 감성 문장들의 집합을 만들어 감성사전을 구축하였고, 이를 영화평에 적용하여 극성에 따라 1부터 10까지로 분류하여 평점을 추론하는 연구들의 시도도 계속되고 있다. 다만 이러한 연구들은 리뷰 내용을 직접적으로 알고리즘에 반영하지 못하여 정보 손실이 발생한다는 한계가 있다. 최근 감성 분석 연구에서는 감성 분류가 아닌 정도를 세밀하게 수치로 측정하는 감성 점수에 관한 연구가 진행되고 있다. Park et al.(2014)는 네티즌 리뷰 등에 포함된 감정 표현 어휘 등을 분석하게 된다면 평점 중심의 평가 패러다임이 가지고 있는 한계를 극복하고 보다 심층적으로 영화의 속성 및 구조 등을 살펴보는 것이 가능하다고 하였다. 이런 이유로 감성의 정도를 파악하는 연구들은 활발하게 진행되고 있다.

2.3 감성 분석을 활용한 추천 시스템

소셜 미디어, 소셜 네트워크 서비스(SNS)가 활성화되면서 대중들은 정보를 소비하는 주체가 되는 동시에 생산하는 주체로 변화하기 시작하였다. 대중들은 SNS 뿐만 아니라 각종 웹사이트의 게시판 참여, 댓글 작성 등의 형태로 자신들

의 의견이나 감정을 표출한다. 이에 따라 리뷰 데이터가 급증하면서 감성 분석에 대한 관심이 높아졌다.

감성 분석은 빅데이터 분석의 중요한 분야 중 하나인 텍스트 마이닝에 기반하여 사람들의 주관적 태도나 감성을 추출해 내는 분석 기법이다. 감성 분석을 활용한 추천 시스템과 관련 연구를 보면 Park et al.(2014)은 영화 추천 시스템에서 개인에게 맞춤형 영화를 추천하기 위하여 사용자들의 리뷰나 코멘트로부터 감정 단어를 추출하였다. 감성 사전인 SentiWordNet을 활용하여 추출된 감정 단어에 대한 감정 정도를 분석해 특정 영화에 대한 여러 사용자의 감정을 분석하여 다른 사용자에게 적절한 영화 추천이 가능함을 보였다. Kim et al.(2015)은 평점과 리뷰의 감정 추론을 위해 영화에 대한 감성 사전인 리뷰 온톨로지를 정의하며 이를 기반으로 영화 리뷰의 감정 단어에 대한 감정 정도를 분석하여 협업 필터링 방법과 상황 기반 기법을 함께 사용해 사용자에게 맞춤형 된 영화를 추천한다. Bhojne et al. (2017)의 연구에서는 레스토랑의 사용자들이 남긴 리뷰를 SentiWordNet을 사용해 긍정 및 부정의 방향성과 감성 수치를 산출하여 사용자의 속성에 기반한 추천 시스템을 구축하는데 활용하였다. Jakob et al.(2009)과 Wang et al.(2012)의 연구에서는 평점 기반으로 이루어졌던 전통적인 추천 알고리즘에 사용자 리뷰 마이닝을 반영하면 성능이 더 개선되는 것을 보여주었다.

감성 분석을 사용한 추천 시스템의 선행연구들을 보면 SentiWordNet과 같은 감성 사전을 사용하는 연구들이 주를 이룬다. 국외의 경우에는 영어 텍스트를 기반으로 한 감성 분석의 경우 AFFIN, SentiWordNet, EmoLex와 같은 사전을 활용한 연구들이 활발하게 이루어졌다. 반면 한

국어에 대한 감성 분석 연구는 영어에 비해 상대적으로 부족한 실정이다. 이는 한국어의 언어적 특성상의 문제라고 볼 수 있다. 감성 분석을 하기 위한 전단계로 자연어를 처리하는 과정에서 문장 안의 명사, 동사, 형용사 등 품사를 태깅(tagging)하는 과정을 거쳐야 한다. 영어의 경우는 굴절어라서 띄어쓰기 단위로 품사가 끊어지는 경향이 있어 띄어쓰기를 기준으로 품사를 태깅하기 용이하다. 이에 반해 한국어는 교착어이기 때문에 띄어쓰기 단위로 품사가 구별되지 않는 경우가 많다. 이러한 이유로 한국어인 텍스트의 감성 분석 연구는 활발하게 진행되지 못하였다. Ahn(2014)과 Jung et al.(2017)는 국내에서는 외국에 비해 감성 사전이 부족한 실정이며 더 활발한 연구가 진행될 필요가 있다고 한다.

2.4 사전 구축 기반의 감성 분석

사전 기반의 분석 방법은 수집된 리뷰 데이터를 전처리 과정을 거친 후, 미리 구축된 감성 사전에 매칭하여 리뷰를 정량화하는 방법론이다. 사전을 활용한다면 감성 정도를 파악하기 용이하지만, 범용 감성사전을 사용하여 감성 분석을 시행한다면 같은 단어가 도메인에 따라 정 반대의 극성을 가질 수 있기 때문에 정확성이 떨어지는 문제가 발생할 수 있다. 그렇기 때문에 Kim et al.(2015)은 감성 분석을 제대로 수행하기 위해서 적용 분야별 도메인의 특성을 살린 전문 사전을 구축할 필요가 있다고 하였다. Lee et al.(2016)도 동일한 어휘라도 분석 대상 영역의 특성에 따라 서로 다른 의미로 사용될 수 있기 때문에 범용 감성 사전을 사용하는 것보다 각 도메인의 특성에 맞게 감성 사전을 구축하는 것이 더 높은 분류 성능을 보인다고 한다. 이와 관련하여 도메인

에 맞는 감성사전을 구축하여 감성 분석을 시행하는 연구들이 활발하게 진행되었다. Kim et al.(2013)은 의료기관 홈페이지에서 수집한 4,300여건의 VOC(voice of customer) 데이터를 이용하여 병원에 특화된 감성 어휘와 감성 극성 값을 도출하여 감성 사전을 구축해 분류 예측 정확도를 높였다. Yu et al.(2013)의 연구에서도 경제 뉴스 데이터로부터 주식 도메인에 특화된 감성 사전을 구축한 것이 주가지수의 상승과 하락을 예측하는데 정확도 향상에 도움이 된다고 하였다. Song et al.(2010)은 주제에 특화된 감성사전을 사용하는 것이 범용적으로 사용하는 감성 사전을 활용하여 감성 분석을 수행하는 것보다 정확성이 향상되는 것을 입증하였다.

앞서 선행 연구들을 정리하면 다음과 같다. 학계나 산업계에서 가장 많이 사용되는 협업 필터링을 수행하기 위해 기존에는 평점, 구매 여부, 방문 횟수와 같은 정량적인 데이터를 사용해왔다. 그 중 가장 대표적으로 사용되는 평점은 최근 들어 추천 시스템의 정확도를 떨어뜨리는 중대한 원인 중 하나가 될 수 있다는 점이 제기되었다. 이러한 배경에 추천 시스템의 정확도를 향상시키기 위한 수단으로 다양한 방법들이 제기되었다. 대표적으로 사용자 리뷰를 추천 시스템에 활용하는 방식이 제기되었지만 리뷰를 긍정, 중립, 부정 등 분류 문제에 그쳐 리뷰 속에 내포되어있는 사용자의 자세한 만족도까지 추천 시스템에 반영하지 못하였다. 최근 감성의 정도를 파악하는 연구들이 진행되면서 추천 시스템에 적용하는 연구들이 활발하게 진행되고 있다. 하지만 한국어의 경우 공개되어 있는 감성 사전이 많지 않다는 점과 존재하는 사전은 범용 사전이기에 추천의 정확도가 떨어진다는 한계가 있다.

이러한 배경에서 선행연구들을 기반으로 본

논문에서는 리뷰의 감성 수치를 산출하여 평점과 결합한 방식으로 추천 시스템의 정확도를 향상시키는 모델을 제시한다. 구체적으로 리뷰의 감성 수치를 산출하기 위해 도메인 맞춤형 감성 사전을 구축하고, 이를 기반으로 사용자가 직접 작성한 리뷰 데이터의 감성 수치를 산출하여 추천 시스템에 반영한다.

3. 연구 설계 및 프로세스

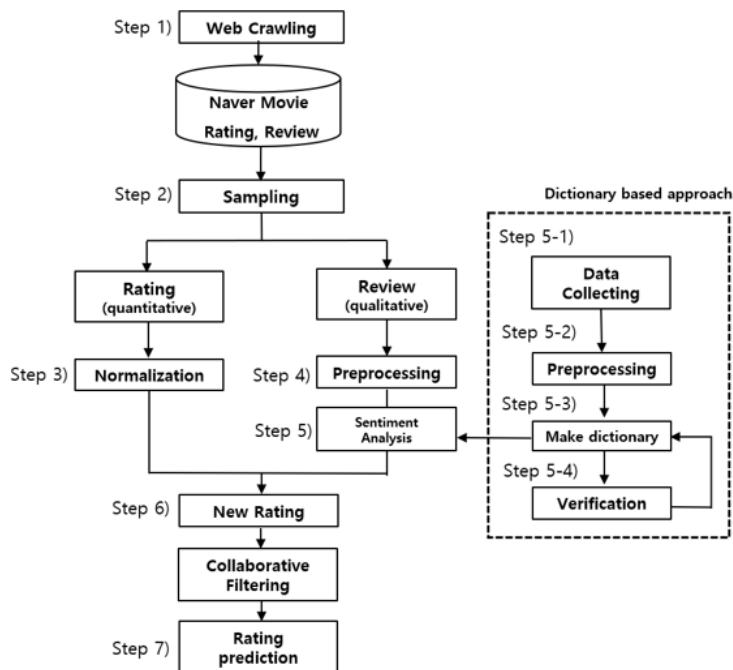
본 연구는 사용자가 작성한 정성적 데이터인 리뷰를 토대로 도메인 맞춤형 사전을 구축한다. 이를 기반으로 텍스트의 감성 수치를 도출하여 정량적 데이터인 평점에 반영해 추천 시스템의

정확성을 향상시킬 수 있는 새로운 추천 알고리즘을 제안한다.

제안 알고리즘은 <Figure 4>과 같은 단계로 구현된다.

1단계. 데이터 수집(Web Crawling)

데이터는 국내 최대 포털 사이트 네이버에서 운영하는 ‘네이버 영화(movie.naver.com)’에서 수집하였다. 웹 크롤러를 통해 사용자에게 접근한 뒤 사용자가 남긴 영화, 평점, 리뷰를 수집하였다. 최종적으로 수집된 데이터는 3856명의 사용자와, 32486개의 서로 다른 영화로 구성되어 있고, 총 점수를 매긴 영화와 리뷰의 수는 100230개이다. 평점은 1점부터 10점까지 1점 단위로 이루어져 있다.



〈Figure 4〉 Overview of the proposed framework

2단계. 데이터 표본 추출(Sampling)

사용자는 평점과 리뷰를 남긴 전체의 영화 중 평점을 부여한 영화보다 평점을 부여하지 않은 영화의 개수가 많아 데이터가 희소한 문제가 발생한다. 이는 평가한 영화의 수가 적을 경우 추천 과정 중 유사한 사용자나 아이템을 발견하기 위해 유사도를 구하는 과정에서 잘못된 유사도를 구할 수 있다. 데이터가 희소한 문제를 줄이기 위하여 적어도 10개 이상의 영화에 대한 평점과 리뷰를 남긴 사용자만을 본 연구의 대상으로 선정했다.

본 연구에서 선정된 사용자(User)는 537명이고, 전체의 영화 개수(Movie)는 4211개이다.

3단계. 평점(Rating) - 평점 정규화(Normalization)

사용자들의 평점 데이터는 사용자의 성향에 따라 편향되어 있다. 사용자들이 아이템에 평점을 주는 기준은 사람마다 각기 다르기 때문에 평점을 높게 주는 경향이 있는 사용자와 평점을 낮게 주는 경향의 사용자가 있다. 전자의 경우의 5점은 재미없는 영화에 속하지만, 후자의 5점은 재미있는 영화일 수 있는 것처럼 두 사람의 평점을 같게 보는 것은 서로 다른 평가 성향을 반영하지 못하여 이를 기반으로 예측을 한다면 정확도가 떨어지는 문제가 발생할 수 있다. 외부 요인에 의한 편향을 줄여주기 위해 사용자들의 개인의 평가 성향을 바탕으로 데이터를 정규화하였다. 정규화를 하면 사용자 간의 유사도 및 사용자의 영화 선호도를 더욱 정확하게 평가할 수 있다.

4단계. 리뷰(Review) - 전처리(Preprocessing)

리뷰 데이터를 보다 정확하게 분석하기 위해 형태소 분석하기 전에 의미를 알 수 없는 불용어

인 단어나 특수문자 및 문장부호, 영문자, 숫자 등을 제거한다. 그 후 리뷰의 필요한 품사를 추출하기 위한 형태소 분석을 시행한다. 본 연구에서는 여러 종류의 형태소 분석기 중 한글 형태소 분석기인 라이노(RHINO) 라이브러리를 활용하여 감성 분석에서 가장 많이 사용되는 명사, 동사, 형용사의 품사만을 선별하여 추출하였다.

5단계. 리뷰(Review)

- 감성 분석(Sentiment Analysis)

5-1단계. 리뷰 데이터 수집(Data Collecting)

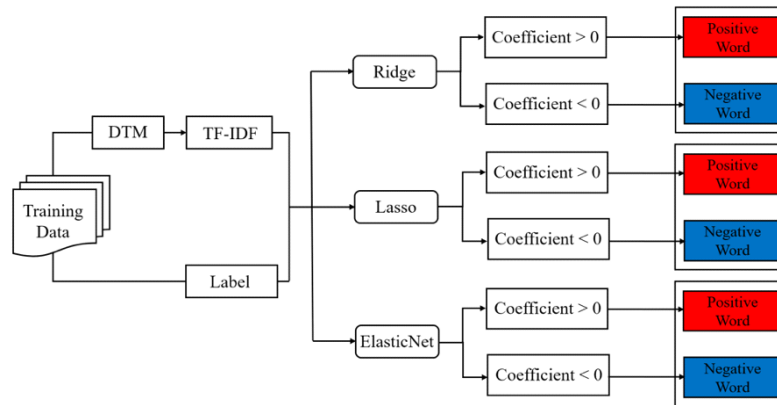
사전을 구축하기 위해 ‘네이버 랩(<http://github.com/e9t/nsmc>)’에서 제공하는 데이터를 추가로 사용하였다. 총 20만 개의 리뷰 데이터로 평점은 1점부터 10점 사이의 정수 값을 가진다. 평점이 1점에서 3점의 경우인 리뷰는 0(부정)으로, 9점에서 10점은 1(긍정)로 레이블(Label)을 부여한다. 이 중 부정, 긍정의 극성 비율을 동일하게 추출하여 10만 건의 영화 중 7만 5천 건의 영화평을 사전 구축용, 2만 5천 건의 영화평은 사전의 정확도 검증용으로 사용한다.

5-2단계. 리뷰 데이터 전처리(Preprocessing)

4단계의 전처리 과정과 같은 방식으로 형태소 분석을 한다. 라이노(RHINO) 라이브러리를 활용하여 형태소 분석 후 리뷰 속의 명사, 동사, 형용사의 품사만을 선별하여 추출한다.

5-3단계. 사전 구축(Make a dictionary)

전처리 과정을 거친 리뷰 데이터를 DTM(Document Term Matrix) 행렬로 만든다. 이때 텍스트를 벡터화하기 위하여 특정 단어가 문서 내에서 얼마나 자주 등장했는지 단어의 중요성 정도를 나타내는 값인 TF-IDF(Term Frequency,



〈Figure 5〉 Building sentiment dictionary

Inverse Document Frequency) 방식을 사용한다. 독립변수는 리뷰-단어의 TF-IDF 값으로 이루어진 매트릭스이고, 종속변수는 각각의 리뷰가 가지고 있던 0과 1로 이루어진 Label 값이다.

사전을 구축하는 방식으로는 회귀분석(Regression Analysis)을 시행한다. 각 단어의 회귀계수를 구하여 0보다 큰 경우에는 긍정 사전에 추가하고, 0보다 작은 경우에는 부정 사전에 추가하여 감성사전을 구축한다. 다만 데이터 특성상 텍스트 데이터는 구조가 희소하고, 차원의 수가 크기 때문에 회귀분석을 실시할 시 변수의 선택 및 추출 과정이 분석 성과를 높이기 위해 중요하다. 따라서 회귀분석 방법 중 릿지 회귀(Ridge Regression), 라쏘 회귀(Lasso Regression), 엘라스틱넷(ElasticNet) 방법을 사용한다.

릿지 회귀분석은 Hoerl과 Kennard(1971)이 제안한 방식으로 회귀계수의 크기에 패널티(Penalty)를 줌으로써 회귀계수를 축소하는 방법이다. Ridge의 추정치는 다음의 식 (1)과 같다.

$$\hat{\beta}_{\text{Ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (1)$$

식 (1)의 $\lambda (\lambda \geq 0)$ 는 회귀계수 크기의 축소 양을 결정한다. λ 의 값이 커질수록 축소되는 양이 많아지며 회귀계수 값이 0에 가까워진다.

라쏘(Least absolute shrinkage and selection operator, Lasso) 회귀분석은 Tibshirani(1996)에 의해 고안된 방법으로, 릿지 회귀와 같은 방식으로 회귀계수의 크기에 패널티를 부여함으로써 회귀계수의 값을 축소하는 방법이다. 이는 영향력 없는 변수의 회귀계수 값을 0으로 만들어서 변수 선택을 가능하게 하는 변수 선택에 사용되는 추정 방법이다. Lasso의 추정치는 식 (2)와 같다.

$$\hat{\beta}_{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

식 (2)의 $\lambda (\lambda \geq 0)$ 의 값이 커질수록 회귀계수 값이 0이 된다.

릿지 회귀와 라쏘 회귀의 두 모델을 비교하자면 릿지는 계수의 제곱, 라쏘는 절대 값을 사용한다는 점이 있다. 릿지는 패널티가 크다고 하더라도 각각의 독립변수의 계수가 0에 가까워지는

것이지 0은 아니기 때문에 모든 독립 변수를 포함한다. 반면 라쏘는 패널티가 클수록 일부 변수들의 계수를 0으로 만들기 때문에 변수에서 제거되어 일부 변수만 포함된다.

엘라스틱넷은 릿지와 라쏘의 두 가지 모형을 합쳐서 나온 알고리즘이다. ElasticNet의 추정치는 다음의 식 (3)과 같다.

$$\hat{\beta}_{ElasticNet} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \left[\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right] \right\} \quad (3)$$

엘라스틱넷은 릿지와 라쏘의 패널티 식을 추가하고 λ 로 조절하여 최적화된 모델을 찾는 것이다. 다만, α 라는 하나의 파라미터를 추가적으로 넣어 둘 사이의 관계를 지정한다. λ 만으로 조절하던 릿지나 라쏘와는 달리 α 가 들어가며, α 값이 커질수록 라쏘의 효과가 강해지고, α 값이 작아질수록 릿지의 효과가 강해진다.

릿지, 라쏘, 엘라스틱넷의 방법을 수행할 때 조절 모수인 λ 를 추정하기 위하여 교차타당성(cross-validation)을 사용한다. 5-fold cross validation을 거쳐 오차가 가장 작을 때의 최적의 λ 를 찾아, 해당 λ 일 때의 각 단어의 회귀계수 값이 0보다 큰 값을 가지는 단어는 긍정 사전으로, 0보다 작은 값을 가지는 단어는 부정 사전으로 분류하여 긍정과 부정 사전을 각각 구축한다. 각각의 구축된 사전의 단어들은 직접 확인하여 불필요한 단어의 경우 제외한다. 예를 들어 명사의 경우 배우 이름, 장소 등의 감정과 관련이 없는 단어가 사전에 포함된 경우 해당 단어를 제외하였다.

5-4단계. 사전 정확성 검증(Verification)

2만 5천 개의 테스트 데이터로 정확도를 평가

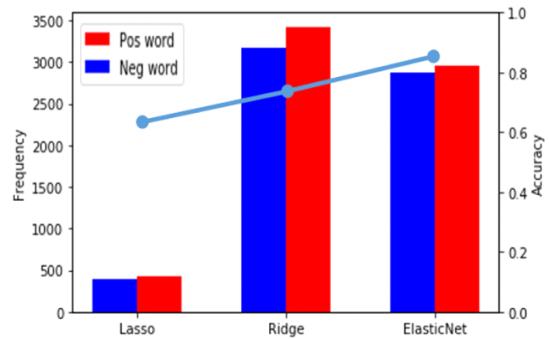
하였다. 사전을 기반으로 리뷰의 감성 점수를 산출하여 감성 점수가 0보다 큰 경우에는 긍정으로, 0보다 작은 경우에는 부정으로 평가한다. 이때의 감성 점수는 식 (4)와 같이 얻어진다.

$$\text{Sentiment Score} = \frac{\text{Pos word} - \text{Neg word}}{\text{Pos word} + \text{Neg word}} \quad (4)$$

긍정과 부정을 예측한 결과의 정확도(Accuracy)를 측정하여 성능이 높은 사전을 분석에 사용한다. 정확도는 <Table 1>의 Confusion Matrix에서 $\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$ 로 정의된다.

<Table 1> Confusion Matrix

Predicted \ Actual	Positive	Negative
True	True Positive	False Negative
False	False Positive	True Negative



<Figure 6> Dictionary Accuracy and Pos/Neg Word Frequency

<Figure 6>는 각 사전에 사용된 긍정 어휘와 부정 어휘의 빈도 수 및 정확도이다. 라쏘 기반의 사전에서 사용되는 긍정 어휘의 수는 398개이고, 부정 어휘의 수는 421개이며 정확도는

70%이다. 릿지 기반의 사전에서 사용되는 긍정 어휘 수는 3164개, 부정 어휘 수는 3425개이며 정확도는 79%이다. 엘라스틱넷에서 사전을 구축하기 위해 α 값을 조절하여 정확도가 가장 높았을 때인 0.3일때의 사전을 구축하였다. 긍정 어휘 수는 2875개, 부정 어휘 수는 2954개로 정확도는 83%이다. 따라서 본 연구에서 리뷰의 감성 수치를 산출하기 위해 사용되는 긍/부정 사전으로 정확도가 가장 높은 엘라스틱넷 기반의 사전을 채택하였다.

6단계. 감성 수치를 반영한 평점(New Rating)

감성 사전을 바탕으로 텍스트 내에 존재하는 긍정, 부정 단어를 통해 텍스트 전체의 감성을 수치로 나타낸다. 감성점수는 앞서 식 (4)와 같이 얻어진다. 감성 분석을 통해 산출된 감성 수치를 평점에 반영하여 새로운 평점(Proposed new rating)을 생성한다. 새로운 평점의 예시는 <Table 2>의 우측 표와 같다. 기존의 평점은 1점부터 10점까지의 정수 값을 가졌지만, 제안된 방식에서의 평점은 실수 값을 가지게 된다.

7단계. 평점 예측(Rating prediction)

평점 예측을 위해 메모리 기반의 사용자 기반 협업 필터링, 아이템 기반 협업 필터링과 모델 기반의 Matrix Factorization의 대표 알고리즘 SVD, SVD++를 사용하였다.

사용자 기반 협업 필터링(이하 UBCF)은 사용자가 입력한 평점을 이용하여 고객과 유사한 성향을 갖는 이웃 사용자를 선정된 뒤, 선정된 이웃들이 공통적으로 선호하는 아이템을 고객에게 추천해주는 방식이다. UBCF로 평점을 예측하기 위한 가장 중요한 단계는 사용자 간의 유사도를 계산하는 것이다.

두 사용자 a와 b의 유사도 $similarity_{a,b}$ 는 식 (5)과 같다.

$$similarity_{a,b} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{b,i} - \bar{r}_b)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in I} (r_{b,i} - \bar{r}_b)^2}} \quad (5)$$

I는 아이템의 전체 집합이며 $r_{a,i}$ 는 사용자 a가 특정 아이템 i를 평가한 점수, \bar{r}_a 는 사용자 a가 전체 아이템에 준 평균 점수이다. 유사도 측정을 통해 유사한 사용자를 선정하면 이들의 과거 구매 이력을 바탕으로 가중합 방식을 사용하여 평점을 예측한다. 고객 a가 아이템 i에 대해 부여할 것이라 예측되는 평점은 식 (6)과 같다.

$$\hat{r}_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u \in U} |w_{a,u}|} \quad (6)$$

\bar{r}_a 는 추천 대상 사용자가 모든 아이템에 준 점수의 평균이고, \bar{r}_u 는 다른 사용자가 모든 아이템에 준 평균 점수이다. $w_{a,u}$ 는 추천 대상 사용자 a와 사용자 u사이의 유사도에 따른 가중치로 유

<Table 2> Original ratings and Proposed new ratings (Example)

	Item 1	Item 2	Item 3	...	ItemM		Item 1	Item 2	Item 3	...	ItemM
User 1	2		4		9	User 1	2.24		4.48		9.45
User 2		8	6			User 2		7.75	6.21		
User 3	5				3	User 3	4.84				2.98
...						...					
User N		5			4	User N		5.12			4.34

사도가 높을수록 큰 가중치를 준다.

아이템 기반 협업 필터링(이하 IBCF)은 특정 아이템이 기준이 되어, 사용자들에 의해 평가된 점수가 유사한 아이템을 이웃 아이템으로 선정한다. 이웃 아이템을 평가한 점수를 바탕으로 추천 대상 사용자가 특정 아이템에 대해 갖게 될 평점을 예측한다. 두 아이템 i 와 j 의 유사도 $similarity_{i,j}$ 는 식 (7)과 같다.

$$similarity_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (7)$$

U 는 아이템 i 와 j 에 모두 평점을 준 사용자 전체 집합이고, $r_{u,i}$ 는 사용자 u 가 아이템 i 에 준 점수, \bar{r}_i 는 전체 사용자가 아이템 i 에 준 평균 점수다. IBCF에서는 단순가중평균 방식을 통하여 점수를 예측하며 식 (8)와 같다.

$$\hat{r}_{u,i} = \frac{\sum_{n \in N} (r_{a,n} \times w_{i,n})}{\sum_{n \in N} |w_{i,n}|} \quad (8)$$

\bar{r}_a 는 추천 대상 사용자가 모든 아이템에 준 점수의 평균, \bar{r}_u 는 다른 사용자가 모든 아이템에 준 평균 점수이다. $w_{i,n}$ 은 예측하고자 하는 아이템과 다른 아이템과의 유사도를 가중치로 주어서 예측하고자 하는 아이템과 유사한 아이템의 평점을 크게 반영하여 예측 값을 계산한다.

모델 기반의 Matrix Factorization 중 SVD와 SVD++는 협업 필터링에서 대표적으로 사용되는 방식이다. SVD(Singular Value Decomposition)는 하나의 행렬을 여러 행렬의 곱으로 분해하는 방법이다. 모든 사용자와 아이템에 대한 $m \times n$ 크기의 매트릭스 M 을 특이값 분해하면 $M = U \Sigma V^T$ 와 같이 세 개의 행렬의 곱으로 나타낼 수 있다. $U_{m \times m}$ 은 사용자 매트릭스를 나타내고, $\Sigma_{m \times n}$ 은

특이값을 대각항으로 가지는 대각 매트릭스, $V_{n \times n}^T$ 는 영화 매트릭스를 나타낸다. 다만, 매트릭스 M 은 희소 행렬(Sparse Matrix)이기 때문에 사용자가 부여하지 않은 평점인 결측값(Missing Value)이 많아 SVD가 정의되지 않을 수 있다. 이를 위해 식 (9)의 정규화 된 모델을 사용하여 사용자가 부여한 평점을 중심으로 오차 함수를 최소화하는 요인 벡터를 찾아내는 방식으로써 평점을 예측한다.

$$\min \sum_{(u,i) \in M} (r_{ui} - \hat{r}_{ui})^2 + \lambda (\|q_i\|^2 + \|p_u\|^2) \quad (9)$$

최소화하는 방법으로 SGD(Stochastic Gradient Descent)를 사용하여 예측오차를 계산 후 매개변수를 수정하며 식 (10), (11)를 산출하여 $\hat{r}_{u,i}$ 를 예측할 수 있다.

$$q_i \leftarrow q_i + \gamma ((r_{u,i} - q_i^T) p_u - \lambda q_i) \quad (10)$$

$$p_u \leftarrow p_u + \gamma ((r_{u,i} - q_i^T) q_i - \lambda p_u) \quad (11)$$

SVD++는 기존 SVD 방식이 명시적인 피드백 정보(Explicit feedback information)를 고려하였다면, 나아가 암시적 피드백 정보(Implicit feedback information)까지 동시에 고려하는 방식이다. SVD를 기반으로 하여 각 아이템에 대해 평점을 준 것과 관계없이 아이템에 대한 특징이 반영된다. SVD++ 방식을 사용하여 평점을 예측하는 형식은 아래 식 (12)와 같다.

$$\hat{r}_{u,i} = \mu + b_u + b_i + q_i^T \left(p_u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j \right) \quad (12)$$

평점에 대한 예측치 $\hat{r}_{u,i}$ 는 데이터 전체의 평점 평균 μ , 사용자와 아이템 각각에 대한 개별적 편향치 b_u, b_i 의 합으로 계산할 수 있다. 사용

자와 아이템 간의 연관성을 추가로 고려하기 위해, 명시적 평점 데이터 행렬과 암시적 평점 데이터 행렬을 SVD를 기반으로 각각 분해하여 사용자와 아이템 데이터를 공통으로 표현하는 저차원의 은닉 공간을 탐색하고, 은닉 공간 상에 표현되는 아이템 벡터 q_i 와 사용자 벡터 p_u 를 구하는 것이다. $R(u)$ 는 아이템에 대해 선호를 보여준 사용자는 벡터로 특징화된다. y_u 는 사용자 u 를 설명하는 속성이다.

4. 분석 결과

4.1 평가 방법

평점만을 고려한 방식과 감성 점수가 결합된 평점으로 추천 시스템의 평점 예측을 평가하기 위하여 MAE(Mean Absolute Error)와 RMSE(Root Mean Square Error)를 사용한다. 두 측정 지표는 협업 필터링을 이용한 추천 시스템에서 가장 많이 사용되는 지표로 사용자의 예측 평점과 실제 사용자가 입력한 평점 간의 차이를 나타낸다. MAE는 절대평균오차로 실측치와 예측치 사이의 오차를 절대 값으로 나타내고, 이를 합하여 예측치의 개수만큼 나누는 방식이다. RMSE는 평균제곱근오차로 실제 값과 예측 값의 차이를

제공한 값을 합하여 예측치의 개수로 나눈 값에 제곱근을 취한 것이다. MAE와 RMSE가 작을수록 추천 시스템의 예측 정확성이 우수함을 나타낸다.

4.2 실험 결과

앞에서 제시한 MAE와 RMSE를 사용하여 예측 성능을 평가해보았다.

‘Original’은 평점 만을 고려한 기본 성능을 나타내며, ‘Proposed’는 본 연구에서 제안된 방법인 감성 수치가 반영된 평점으로 평점을 예측하는데 적용했을 때의 성능을 나타낸다. 평가 지표가 MAE일 때 UBCF에서는 0.059, IBCF는 0.0862, SVD는 0.1012, SVD++은 0.188만큼 개선되었다. 평가 지표가 RMSE일 때는 UBCF에서는 0.0431, IBCF는 0.0882, SVD는 0.1103, SVD++에서는 0.1756 만큼 향상되었다. 분석 결과 본 논문에서 제안된 감성 수치를 반영한 평점의 예측 성능이 기존의 평점만을 고려한 방식보다 전체적으로 예측력이 우수한 것을 알 수 있다.

나아가 제안하는 모델이 더 나은 방식이라고 할 근거가 충분한지 검정하기 위해 Paired t-test를 시행하여 평점만을 고려한 방식과 감성 수치를 반영한 방식의 평균에 차이가 있는지 비교한다.

〈Table 3〉 Experimental Results of Proposed Model with Test data

		MAE		RMSE	
		Original	Proposed	Original	Proposed
Memory based CF	UBCF	2.3013	2.2417	3.0867	3.0426
	IBCF	2.0984	2.0122	2.8946	2.8064
Model based CF	SVD	1.9754	1.8742	2.6146	2.5043
	SVD++	1.9625	1.7745	2.6228	2.4532

〈Table 4〉 Paired t-test (MAE)

	Paired Differences					t	df	Sig. 2-tailed
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Ori_UBCF – Pro_UBCF	.0568	.0629	.0114	.0333	.0804	4.948	29	.000
Pair 2 Ori_IBCF – Pro_IBCF	.0838	.0743	.0135	.0561	.1116	6.183	29	.000
Pair 3 Ori_SVD – Pro_SVD	.1031	.0306	.0055	.0916	.1145	18.43	29	.000
Pair 4 Ori_SVDpp – Pro_SVDpp	.1892	.0834	.0152	.1581	.2204	12.42	29	.000

〈Table 5〉 Paired t-test (RMSE)

	Paired Differences					t	df	Sig. 2-tailed
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Ori_UBCF – Pro_UBCF	.0406	.0904	.0165	.0068	.0743	2.461	29	.020
Pair 2 Ori_IBCF – Pro_IBCF	.0886	.0575	.0105	.0671	.1101	8.437	29	.000
Pair 3 Ori_SVD – Pro_SVD	.1214	.0287	.0052	.1107	.1321	23.16	29	.000
Pair 4 Ori_SVDpp – Pro_SVDpp	.1664	.1247	.0227	.1198	.2130	7.306	29	.000

MAE의 대응표본 t-검정 결과, 모든 방식에서 유의확률 p값이 95% 신뢰수준 하에서 유의한 것 (UBCF t=4.948, p-value=0.00, IBCF t=6.183, p-value=0.00, SVD t=18.432, p-value=0.00, SVD++ t=12.428, p-value=0.00)으로 나타난다.

RMSE의 대응표본 t-검정 결과도 MAE와 마찬가지로 모든 방식에서 유의확률 p값이 95% 신뢰

수준 하에서 RMSE의 차이가 서로 유의한 것 (UBCF t=2.416, p-value=0.02, IBCF t=8.437, p-value=0.00, SVD t=23.166, p-value=0.00, SVD++ t=7.306, p-value=0.00)으로 나타난다. 이처럼 평점만을 고려한 추천 시스템과 리뷰를 반영한 추천 시스템의 모델 간의 성능에 유의한 차이가 있다고 할 수 있다. 결과적으로 기존의

MAE/RMSE 값보다 제안된 방식에서 오차 값이 낮아 감성수치를 반영한 평점이 평점만을 고려한 방식과 비교해 더 우수한 예측 정확도를 나타낸다.

5. 결론

본 연구는 정량 데이터인 평점만을 활용하여 추천 결과를 생성하는 기존의 협업 필터링 방식의 정확도를 개선하기 위해, 정성 데이터인 사용자 리뷰를 반영하여 협업 필터링의 성능을 높일 수 있는 새로운 추천 알고리즘을 제안하였다. 도메인 맞춤형 기반의 사전을 구축하여 이를 기반으로 리뷰의 감성 정도를 수치화한 것을 평점과 결합하여 감성 수치가 반영된 새로운 평점을 생성하였다. 그 결과 평점만을 고려한 기존의 방식보다 감성 수치가 반영된 평점의 경우의 알고리즘의 종류와 상관없이 예측 정확도가 높은 것을 확인할 수 있다.

본 연구는 학술적, 실무적으로 다음과 같은 의의를 갖는다. 대다수의 추천 시스템 연구에서 사용된 평점 데이터에만 의존하지 않고, 리뷰 데이터를 활용하여 추천 정확도의 성능을 향상시키는 새로운 방식을 제시한 점은 기존 연구와 본 연구가 차별화되는 부분이다. 특히 기존 연구에서 시도하지 않았던 리뷰를 감성점수화 하여 평점에 적용하는 방법론을 새롭게 제안하여 기존의 추천 시스템보다 예측 정확도를 향상시킨 점은 본 연구가 갖는 중요한 학술적 의의라 생각한다.

실무적 의의로는 첫 번째로 리뷰 데이터 정보 손실 문제를 해결하였다. 기존의 연구에서는 리뷰 내용을 긍정/부정으로 파악하는 연구에 그쳐 직

접적으로 알고리즘에 반영하지 못하여 정보 손실 문제가 발생하였으나, 리뷰를 정량화하여 추천 시스템에 반영함으로써 문제를 해결할 수 있다. 이처럼 비정형 데이터를 정형데이터화 함으로써 기존의 대용량의 리뷰 데이터의 정보는 잃지 않을 수 있다.

두 번째로 사용자의 의견을 수치화하여 추천 시스템 알고리즘에 직접 반영하였다. 평점만으로 사용자의 선호 정도를 판단한 기존의 선행연구들이 가지는 한계를 극복하고자 리뷰를 수치화시켜 평점에 반영했다는 점에서 사용자의 선호를 정교하게 추천 시스템에 반영시킬 수 있다.

세 번째로 비정형 데이터를 정형 데이터화 함으로써 정형 데이터만이 가능하던 다양한 분석 방법이 가능하다. 비정형 데이터인 텍스트의 경우에는 기존의 양적 데이터의 분석 방법론을 직접적으로 사용할 수 없었다. 본 연구에서는 정형 데이터의 형태로 변환했기 때문에 다양한 양적 분석 방법론들을 적용해 볼 수 있게 된다. 예를 들어 본 연구를 통해 새롭게 산출한 평점을 토대로 추천 시스템에 활용한다면 사용자 간의 유사도를 더욱 정밀하게 나타낼 수 있다. 한 예로 사용자 간의 유사도를 바탕으로 모든 항목에 대해 예측 값을 계산하여 높은 예측 값을 갖는 상위 N개의 추천 리스트를 생성하는 KNN(K-Nearest Neighbors)을 활용하여 새로운 평점을 예측할 때 사용자에게 대한 선호를 더욱 정확하게 나타낼 수 있다. 또한, 소셜 네트워크 분석 결과를 반영한 추천 알고리즘에서도 사용할 수 있다. 사용자의 평점 외에 사용자 간의 신뢰 관계를 중심 위치를 나타내는 척도인 내향 및 외향 중심성을 활용하여 사용자 간의 유사도를 산출하여 더 상세하게 추천이 이루어 질 수 있다. 나아가 기존에는 텍스트 마이닝 기반으로 감성 분석 연구에 초점을

맞춤으로써 다양한 변수를 고려한 연구가 이루어지지 못하였는데, 생성된 새로운 평점을 기초로 하여 평점에 영향을 미치는 다른 변수들을 추가로 통합함으로써 더 좋은 성능을 보이는 추천 시스템을 구축할 수 있다.

마지막으로 평점과 리뷰 데이터를 같이 사용하는 다양한 분야에서 활용이 가능하다. 자신의 의견을 표할 수 있는 리뷰를 남길 수 있는 플랫폼이 늘어나고 있는 만큼 리뷰에 나타난 감성을 파악하는 연구는 끊임없이 진행되고 있다. 본 연구에서는 영화 도메인에 맞춤형 사전을 구축하여 평점과 리뷰를 같이 사용하였지만, 평점과 리뷰 데이터를 같이 사용하는 분야는 다양하게 존재한다. 이를 활용한다면 리뷰 속의 감성 정도를 빠르게 파악할 수 있고, 리뷰를 수치화하기에 좋을 것으로 기대된다.

본 연구의 한계점으로는 본 연구에서 제안하는 알고리즘은 영화 도메인에서 평점 예측을 위해 구축된 사전이므로 다른 도메인에서 사용하여 평점을 예측하는데 부적절하다. 사전은 도메인 의존성이 높기 때문에 같은 단어라도 분야마다 다른 감성을 가질 수 있다. 본 연구에서 구축한 사전은 영화 도메인에 한정하여 사용해야 한다는 점이다.

또한 구축된 사전은 지속적으로 관리하여 업데이트 해야 한다. 리뷰에는 신조어, 축약어 등 당시의 트렌드를 반영하는 경우가 많다. 이는 사용자의 감성을 이해하는데 있어서 필수적인 요소이기에 구축된 사전을 지속적으로 관리하며 새로운 단어를 추가해야 한다.

아울러 본 논문에서 제안하는 추천 알고리즘의 성능을 보다 개선하기 위한 향후 연구로는 부사를 반영한 사전을 추가적으로 구축할 것이다. 본 논문에서는 사전을 구축할 때 명사, 동사, 형

용사만을 사용하였었다. 감정 표현 단어의 의미 강도를 나타내기 좋은 품사인 부사를 고려하지 못하였다. 감정 정도를 더욱 세밀하게 하기위해 부사어 사전을 추가로 만들어 부사어 단어의 점수를 가중치를 주어 반영한다면, 감성수치가 더욱 정교화되어 추천 시스템의 정확도가 더욱 향상될 것으로 예상된다.

참고문헌(References)

- Ahn, J. K. and H. W. Kim , “Building a Korean Sentiment Dictionary and Applications of Natural Language Processing”, *J Intell Inform Syst*, (2014), 177~182.
- Bhojne, N. G., Deore, S., Jagtap, R., Jain, G. and Kalal, C., “Collaborative Approach based Restaurant Recommender System using Naive Bayes”, *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 6, No.4(2017).
- Breese, J. S., Heckerman, D., and Kadie, C., “Empirical analysis of predictive algorithms for collaborative filtering”, *In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, (1998), 43-52.
- Choi, D. J., H. S. Choi and C. Y. Park., “Classification of ratings in online reviews”, *Journal of the Korean Data & Information Science Society*, (2016), 845~854.
- Deng, D., Jing, L., Yu, J., Sun, S. and Zhou, H., “Neural gaussian mixture model for review-based rating prediction”, *RecSys '18 Proceedings of the 12th ACM Conference on Recommender Systems*, (2018).
- Garcis-Cumbreras, M. A., A. Montejo-Raez, and

- M. C. Diaz-Galiano, "Pessimists and optimists: Improving collaborative filtering through sentiment analysis", *Expert Systems with Applications*, Vol. 40, No.17(2013), 6758~6765.
- Goldberg, D., Nichols, D., Oki, B. M., and Terry, D., "Using collaborative filtering to weave an information tapestry", *Communications of the ACM*, (1992), 61~70.
- Hoerl, E. and Kennard, R.W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems", *Technometrics*, Vol.12, No.1 (1970), 55~67.
- Jakob, "Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations", *TSA '09 Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, (2009), 57~64.
- Jeon, B. K., H. C. Ahn, "A Collaborative Filtering System Combined with Users Review Mining Application to the Recommendation of Smartphone-Apps", *J Intell Inform Syst*, Vol.21, No.2(2015), 1~18.
- Jung, C. H., J. H. Kim, Y. J. Jeon, H. J. Jung, "Korean Sentiment Dictionary based on the Reliability of Review data", *Korean Institute of Information Scientists and Engineers, 2017 Korea Software Congress*, (2017), 1965~1967.
- Kim, D. H. and S. H. Choi, "A Domain Adaptive Sentiment Dictionary Construction Method for Domain Sentiment Analysis", *The Korean Society of Computer and Information, Proceedings of the Korean Society of Computer Information Conference*, (2015), 15~18.
- Kim, S. B., S. J. Kwon. and J. T. Kim, "Building Sentiment Dictionary and Polarity Classification of Blog Review by Using Elastic Net", *Korean Information Science Society*, (2015), 639~641.
- Kim, Y. S. and S. R. Jeong, "Intelligent VOC Analyzing System Using Opinion Mining", *Korea Intelligent Information Systems Society, Journal of Intelligence and Information Systems* 19(3), (2013), 113~125.
- Lee, S. H., J. Choi and J. W. Kim, "Analysis of movie review emotion through customized emotional dictionary construction by domain", *J Intell Inform Syst*, Vol.22, No.2 (2016), 97~113.
- Leung, C. W., Chan, S. C. and Chung, F., "Integrating Collaborative Filtering and Sentiment Analysis: A Rating Inference Approach", *ECAI 2006 Workshop on Recommender Systems*, (2006), 62~68.
- Liu, S. M., Chan, J. H., "A multi-label classification based approach for sentiment classification", *Expert Systems with Application*, Vol. 42, (2005), 1083~1093.
- Oh, Y. J. and S. H. Choi, "Movie Rating Inference by Construction of Movie Sentiment Sentence using Movie comments and ratings", *Journal of Korean Society for Internet Information*, Vol.16, No.2(2015), 41~48.
- Park, J. Y. and B. S. Jeon, "A structural Analysis of the Movie Reviews", *The Journal of the Korea Contents Association*, Vol. 14, No. 5(2014).
- Ricci, F., Rokach, L., Shapira, B. and Kantor, P., *Recommender systems handbook*, 2011.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., "Item-based collaborative filtering recommendation algorithms", *In Proceedings of the 10th international conference on World Wide Web*, (2001), 285~295.

- Son. J. E., S. B. H., H. J. Kim and S. J. Cho., "Review and Analysis of Recommender Systems, *Journal of the Korean Institute of Industrial Engineers*, Vol. 41, No. 2(2015), 185~208.
- Song. J. S., J. B. Back and S. W. Lee, "Automatic Construction of Positive/Negative Dictionary to Improve Performance of Product Review Classification", *Korean Institute of Information Scientists and Engineers, 2010 Korea Computer Congress*, Vol. 37, No.1 (2010), 136~137.
- Tibshirani, R., "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society*, Vol. 58, No. 1(1996), 267~288.
- Wang, Y., Liu, Y. and Yu, X., "Collaborative Filtering with Aspect-Based Opinion Mining: A Tensor Factorization Approach", *Proceedings of 2012 IEEE 12th International Conference on Data Mining (ICDM)*, (2012), 1152~1157.
- Yu, E., Y. Kim, N. Kim, and S. R. Jeong, "Predicting the direction of the stock index by using a domain-specific sentiment dictionary," *KIIS Journal of Intelligence and Information Systems*, Vol. 19, No. 1(2013), 95~110.

Abstract

How to improve the accuracy of recommendation systems: Combining ratings and review texts sentiment scores

Jiyeon Hyun* · Sangyi Ryu** · Sang-Yong Tom Lee***

As the importance of providing customized services to individuals becomes important, researches on personalized recommendation systems are constantly being carried out. Collaborative filtering is one of the most popular systems in academia and industry. However, there exists limitation in a sense that recommendations were mostly based on quantitative information such as users' ratings, which made the accuracy be lowered. To solve these problems, many studies have been actively attempted to improve the performance of the recommendation system by using other information besides the quantitative information. Good examples are the usages of the sentiment analysis on customer review text data. Nevertheless, the existing research has not directly combined the results of the sentiment analysis and quantitative rating scores in the recommendation system. Therefore, this study aims to reflect the sentiments shown in the reviews into the rating scores. In other words, we propose a new algorithm that can directly convert the user 's own review into the empirically quantitative information and reflect it directly to the recommendation system. To do this, we needed to quantify users' reviews, which were originally qualitative information.

In this study, sentiment score was calculated through sentiment analysis technique of text mining. The data was targeted for movie review. Based on the data, a domain specific sentiment dictionary is constructed for the movie reviews. Regression analysis was used as a method to construct sentiment dictionary. Each positive / negative dictionary was constructed using Lasso regression, Ridge regression, and ElasticNet methods. Based on this constructed sentiment dictionary, the accuracy was verified through confusion matrix. The accuracy of the Lasso based dictionary was 70%, the accuracy of the Ridge based dictionary was 79%, and that of the ElasticNet ($\alpha = 0.3$) was 83%. Therefore, in this study, the sentiment

* Business Informatics, School of Business, Hanyang University

** School of Business, Hanyang University

*** Corresponding Author: Sang-Yong Tom Lee

School of Business, Hanyang University

222 Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

Tel: +82-2-2220-1814, Fax: +82-2-2220-1169, E-mail: tomlee@hanyang.ac.kr

score of the review is calculated based on the dictionary of the ElasticNet method. It was combined with a rating to create a new rating.

In this paper, we show that the collaborative filtering that reflects sentiment scores of user review is superior to the traditional method that only considers the existing rating. In order to show that the proposed algorithm is based on memory-based user collaboration filtering, item-based collaborative filtering and model based matrix factorization SVD, and SVD ++. Based on the above algorithm, the mean absolute error (MAE) and the root mean square error (RMSE) are calculated to evaluate the recommendation system with a score that combines sentiment scores with a system that only considers scores. When the evaluation index was MAE, it was improved by 0.059 for UBCF, 0.0862 for IBCF, 0.1012 for SVD and 0.188 for SVD ++. When the evaluation index is RMSE, UBCF is 0.0431, IBCF is 0.0882, SVD is 0.1103, and SVD ++ is 0.1756. As a result, it can be seen that the prediction performance of the evaluation point reflecting the sentiment score proposed in this paper is superior to that of the conventional evaluation method.

In other words, in this paper, it is confirmed that the collaborative filtering that reflects the sentiment score of the user review shows superior accuracy as compared with the conventional type of collaborative filtering that only considers the quantitative score. We then attempted paired t-test validation to ensure that the proposed model was a better approach and concluded that the proposed model is better. In this study, to overcome limitations of previous researches that judge user's sentiment only by quantitative rating score, the review was numerically calculated and a user's opinion was more refined and considered into the recommendation system to improve the accuracy. The findings of this study have managerial implications to recommendation system developers who need to consider both quantitative information and qualitative information it is expect. The way of constructing the combined system in this paper might be directly used by the developers.

Key Words : collaborative filtering, recommendation system, sentiment analysis, sentiment dictionary, text mining

Received : January 13, 2019 Revised : March 17, 2019 Accepted : March 29, 2019

Publication Type : Conference(Fast-track) Corresponding Author : Sang-Yong Tom Lee

저 자 소개



현 지 연

한양대학교 일반대학원 비즈니스인포매틱스학과에서 석사학위를 취득하였다. 주요 관심분야는 Big Data Analysis, Data Mining, Recommendation Systems, Social Media, Text Mining 등이다.



유 상 이

한양대학교 일반대학원 경영학과에서 MIS(Management Information Systems) 전공으로 석사학위를 취득하였다. 주요 관심분야는 Big Data Analysis, Business Strategy, Digital Business, Social Media, Text Mining 등이다.



이 상 용

현재 한양대학교 경영대학 교수로 재직 중이다. 서울대학교 경제학과를 졸업하고, Texas A&M University에서 박사학위를 취득하였다. 주요 관심분야는 정보경제, 개인정보보호 (privacy) 및 보안, 소셜미디어, 정보통신정책, 기술경영 등이다. 관련 연구들을 MIS Quarterly, Management Science, Journal of Management Information Systems를 비롯한 다수의 저널에 관련 논문을 게재하고 있다.