

온라인 쇼핑몰에서 소셜 네트워크 데이터를 고려한 상품 트렌드 분석

Item Trend Analysis Considering Social Network Data in Online Shopping Malls

박수빈*, 최도진**, 유재수**, 복경수***

충북대학교 빅데이터협동과정*, 충북대학교 정보통신공학과**, 원광대학교 SW융합학과***

Soobin Park(vivapark530@chungbuk.ac.kr)*, Dojin Choi(mycdj91@chungbuk.ac.kr)**,
Jaesoo Yoo(yjs@chungbuk.ac.kr)**, Kyoungsoo Bok(ksbok@wku.ac.kr)***

요약

온라인 쇼핑몰의 활성화로 소비자들의 소비 활동이 활발해짐에 따라 기업들은 매출 증대를 위해 소비자의 상품 트렌드 분석을 수행하고 있다. 기존의 상품 트렌드 분석 기법들은 온라인 쇼핑몰 사용자의 활동만을 고려하여 분석하기 때문에 구매 이력이 없거나 새로운 상품에 대한 트렌드를 파악하기 어렵다. 본 논문에서는 쇼핑몰에서 사용자의 트렌드와 잠재적 고객의 트렌드를 분석하기 위해 온라인 쇼핑몰 데이터와 소셜 네트워크 데이터를 결합한 트렌드 분석 기법을 제안한다. 제안하는 기법은 서비스 내 데이터 분석을 위해 사용자의 활동 로그를 분석하고 활동 로그가 없는 잠재적인 사용자들의 관심도를 반영하기 위해 소셜 네트워크 데이터에서 단어 집합 추출을 통해 생성한 핫 토픽을 결합하였다. 최종적으로 상품 지수와 소셜 네트워크에서의 언급수를 활용하여 시간에 따른 상품 트렌드 변화를 탐지한다. 소셜 네트워크 데이터를 활용한 성능 평가를 통해 제안하는 기법의 우수성을 입증한다.

■ 중심어 : | 소셜 네트워크 서비스 | 트렌드 분석 | 온라인 쇼핑몰 | 핫 토픽 | 영향력 분석 |

Abstract

As consumers' consumption activities become more active due to the activation of online shopping malls, companies are conducting item trend analyses to boost sales. The existing item trend analysis methods are analyzed by considering only the activities of users in online shopping mall services, making it difficult to identify trends for new items without purchasing history. In this paper, we propose a trend analysis method that combines data in online shopping mall services and social network data to analyze item trends in users and potential customers in shopping malls. The proposed method uses the user's activity logs for in-service data and utilizes hot topics through word set extraction from social network data set to reflect potential users' interests. Finally, the item trend change is detected over time by utilizing the item index and the number of mentions in the social network. We show the superiority of the proposed method through performance evaluations using social network data.

■ keyword : | Social Network Service | Trend Analysis | Online Shopping Mall | Hot Topic | Influence Analysis |

* 이 논문은 2019학년도 원광대학교의 교비지원에 의해 수행됨

접수일자 : 2019년 11월 14일

수정일자 : 2019년 12월 12일

심사완료일 : 2019년 12월 12일

교신저자 : 복경수, e-mail : ksbok@wku.ac.kr

I. 서론

온라인 쇼핑몰을 활성화로 인해 소비자들은 시간과 장소에 구애받지 않고 인터넷을 통해 원하는 상품을 검색하거나 구매할 수 있다. 온라인 쇼핑몰에서 소비자가 원하는 상품을 적재적소에 배치하여 구매자의 관심을 유도하는 것은 실질적인 구매와 연관이 되기 때문에 중요한 요소이다[1][2]. 따라서 온라인 쇼핑몰에서 소비자들의 관심 트렌드를 분석하여 이를 상품 판매에 활용하고 있다[3-5]. 기존의 초기 상품 트렌드 분석에서는 온라인 쇼핑 환경에서 수집된 클릭 스트림이나 세션 타임과 같은 로드를 분석하였다. V.Gala는 온라인 쇼핑몰 데이터를 활용하여 상품에 대한 사람들의 반응과 연관된 트렌드를 예측하는 기법을 제안하였다[5]. 상품에 대한 정보와 상품을 구매한 사람들의 후기 정보를 이용하여 해당 상품에 대한 감정 분석을 수행하여 특정 상품에 대한 미래의 반응을 예측했다[5]. 하지만 온라인 쇼핑몰에서 수집한 사용자 로그만을 고려한 트렌드 분석은 서비스 내에서 적극적인 활동을 했던 사용자의 특성은 고려할 수 있으나 활동이 적은 잠재적 고객의 특성을 반영하지 못한다. 즉, 잠재적 고객 발굴을 위해서는 적극적인 구매 활동이 없는 사용자에게 트렌드를 분석 결과를 제공하는 것이 필요하다. 특히, 새로운 사용자가 관심을 가질만한 상품의 특성을 반영하는 것이 필요하다.

온라인 소셜 네트워크가 대중화되기 이전에는 상품에 대한 정보나 트렌드를 파악할 수 있는 채널이 부족했다. 최근 소셜 네트워크 서비스의 발달로 소비 패턴과 정보 확산 방식의 변화가 이루어졌다[6-8]. 소비자들은 특정 상품에 대한 리뷰나 기대치에 대한 의견을 소셜 네트워크를 통해 공유하고 있다. 따라서 소셜 네트워크는 상품의 트렌드를 파악하기 위한 중요한 데이터를 제공하고 있다. 온라인 소셜 네트워크 특성을 고려할 경우 온라인 쇼핑몰에서 사용자 활동 데이터가 없거나 적은 경우를 보완하여 효과적으로 트렌드 분석을 할 수 있다. M. Mathioudakis는 소셜 네트워크 환경에서 키워드 분석을 통한 트렌드를 감지하는 기법을 제안하였다[9]. 하지만 단순하게 키워드 기반으로 트렌드를 감지하여 부정적인 이유로 많이 언급된 키워드에

대해서도 트렌드로 검출될 수 있기 때문에 상품 트렌드 분석에 적합하지 않다. 따라서 온라인 쇼핑몰 사용자의 명시적인 관심 표현인 사용자 활동을 반영하여 상품 트렌드를 분석하는 연구가 필요하다.

본 논문에서는 소셜 네트워크 데이터를 이용하여 시간 속성을 고려한 TF-IDF(Term Frequency-Inverse Document Frequency) 기법[9]을 통해 핫 토픽 지수를 계산한다. 그리고 온라인 쇼핑몰에서 상품 구매나 구매 후기 작성과 같은 상품에 대한 관심을 표현하는 활동이 적거나 없는 사용자에게 적합한 상품을 추천하기 위해 수집된 데이터와 소셜 네트워크 데이터를 결합하여 상품의 트렌드를 분석하는 기법을 제안한다. 또한, 기존 기법에서 현재 시간만을 기준으로 트렌드를 분석했을 때의 문제점을 해결하기 위해 제안하는 기법에서는 시간의 영향을 받아 빠르게 바뀌는 트렌드 특성을 고려하여 시간 변화에 따른 분석 기법을 사용한다.

본 논문의 이어지는 구성은 다음과 같다. II장은 관련 연구에 대하여 기술한다. III장에서는 제안하는 개인화 추천 기법의 구조 및 처리 과정에 대해서 설명하고 IV장에서는 제안하는 기법의 성능 평가에 대해 기술한다. 마지막으로 V장은 본 논문의 결론을 제시한다.

II. 관련 연구

[9]에서는 트위터와 같은 소셜 네트워크 환경에서의 트렌드 분석 방법을 제안하였다. 실시간으로 수집되는 트위터를 이용하여 트렌드를 분석하기 위해 트렌드 감지와 트렌드 분석 과정을 수행한다. 트렌드 감지에서는 이전 시간에 많이 언급 되지 않은 단어들을 분석하고 이를 Bursty 키워드로 분류한다. 해당 Bursty 키워드는 유사한 의미를 가지는 단어들과 그룹핑을 수행하여 Bursty 그룹을 생성한다. 트렌드 분석에서는 확률 모델을 기반으로 해당 Bursty 그룹이 스펜인지, 우연히 많이 발생한 단어인지, 일상적으로 자주 나오는 단어였는지를 판별하여 최종 트렌드를 추출한다. 그러나 [9]에서는 문서상에서의 어조를 분별하지 못하고 단순 Bursty 키워드를 기반 하기 때문에 실제로 Bursty 하지 않아도 중요한 트렌드로 분류할 수 있는 단어들을 인식하지 못

한다. 상품 트렌드에서 어조를 분별하지 못한다면 비판적인 글이 많은 상품에 대해서 트렌디한 상품으로 인식하는 오류가 발생할 수 있다.

[11]에서는 사이버 위험 이벤트를 감지하기 위한 트위터 기반 트렌드 분석 방법을 제안하였다. [11]에서는 트렌드 분석의 결과를 트렌드 지수에 따라 랭킹하고, 소셜 네트워크의 특성을 반영하기 위해서 사용자의 영향력 지수(팔로워 수, 사이버 관련 트윗 수)를 트윗에 반영한다. DBSCAN[12]과 같은 클러스터링 기법을 통해서 단어를 그룹핑하고, Novel 한 키워드인지, Novel and Trendy한 키워드인지에 따라 단어를 랭킹한다. 그러나 Trendy 여부를 판별하기 위해서 기존에 언급되었던 단어와의 코사인 유사도 비교를 하는데 임계값을 사용자가 지정해야만 한다. 또한, [11]와 마찬가지로 문서의 긍/부정 값이 반영되지 않는다. 사용자의 영향력 지수를 반영함에 앞서 상품 트렌드 분석에서는 상품 홍보를 위해 특정 상품과 관련된 사용자가 존재하는 경우가 많고 해당 상품에 대한 언급을 주로하기 때문에 영향력 높은 사용자로 잘못 판별 될 확률이 높다.

[13][14]에서는 소셜 네트워크의 데이터를 고려하여 핫 토픽을 예측하는 기법을 제안하였다. [13]에서는 키워드 추출에 많이 사용되는 TF-IDF 알고리즘을 사용하여 특정 단어의 출현 빈도와 역문서 빈도를 기반으로 키워드의 중요도를 구분하였다. 또한 이전 시간의 TF-IDF의 값을 비교함으로써 시간 속성을 고려하여 단어의 급격한 출현에 의한 키워드 추출 기법을 제안하였다. [14]에서는 트위터에서 핫 토픽 예측에 사용자 활동과 관련된 요소인 팔로워 수, 트윗 수, 멘션 수 기준으로 사용자의 영향력을 도출하여 핫 토픽 예측에 반영하여 사용한다.

클릭 스트림 데이터는 사용자가 웹 서비스를 이용할 때 발생하는 검색 이동 경로를 알 수 있는 데이터이다 [15]. 이를 이용하면 사용자의 방문 목적이나 관심 있는 페이지에 대한 접근이 가능하다. 그러나 온라인 쇼핑물에서 사용자 데이터는 데이터 희소성을 가지고 있다. 특히, 서비스를 사용한지 얼마 되지 않았거나 처음 접하는 사용자들에게는 충분한 데이터를 얻을 수가 없다. 서비스 내의 데이터만을 사용한다면 상품의 트렌드 분석에 있어서 원하는 결과를 얻기 힘들다. 따라서 온라인

인 쇼핑물 서비스에서 데이터 희소성 문제를 보완하고 대중의 이슈를 반영하는 상품의 트렌드를 파악할 수 있는 기법의 연구가 필요하다.

III. 제안하는 트렌드 분석 기법

1. 특징

소셜 네트워크 환경에서 트렌드 분석을 수행하기 위해서는 소셜 네트워크 데이터를 수집하여 정제하는 과정이 필요하다. 그리고 상품에 대한 최신 트렌드 판별하기 위해 온라인 쇼핑물 내에서 발생한 데이터 뿐만 아니라 사용자의 개인 의견이 표출되는 소셜 네트워크 데이터를 분석하는 것이 필요하다. 또한, 트렌드 변화를 파악하기 위해 단어들을 선별할 때는 일상적인 단어와 구분할 수 있도록 시간을 고려하여 이전 시간보다 얼마나 많이 등장했는지에 대한 계산이 필요하다.

본 논문에서는 온라인 쇼핑물에서 비적극적인 사용자의 의견을 반영하기 위해 온라인 쇼핑물에서 발생한 활동 로그와 소셜 네트워크 데이터를 결합한 상품 트렌드 분석 기법을 제안한다. 기존 트렌드 분석 기법은 온라인 쇼핑물에 가입한지 얼마 되지 않았거나 아직 가입하지 않은 잠재적 고객에 있어서 사용자의 활동 로그가 충분하지 않기 때문에 데이터 희소성 문제를 가지고 있다. 이렇게 높은 데이터 희소성을 가진 정보를 사용하면 해당 고객에 대한 맞춤형 트렌드를 제공할 수 없는 한계를 가지고 있다. 제안하는 기법은 데이터 분석 과정에서 이전 시간의 트렌드와 비교를 하여 결과에 반영함으로써 시시각각 빠르게 변하는 트렌드를 잘 파악할 수 있다. 시간별 트렌드 분석을 위하여 타임윈도우를 사용하여 해당 시간에 발생한 소셜 네트워크 문서를 반영하여 트렌드 지수를 계산하기 때문에 Cold Start 문제가 발생하지 않는다. 제안하는 기법은 시간 속성을 고려한 변형된 TF-IDF 알고리즘을 이용하여 소셜 핫 토픽을 추출하고, 온라인 쇼핑물에서 사용자의 의사표현 방식인 째하기, 조회 수, 평점부여하기 등의 행위 분석을 통하여 사용자의 패턴을 분석한다.

[그림 1]은 제안하는 기법의 전체 처리 절차를 나타낸다. 데이터 수집기, SNS 데이터 분석, 서비스 내 데이

터 분석 그리고 최종 트렌드 분석을 수행한다. 데이터 수집기는 포털 사이트를 비롯한 소셜 네트워크 데이터와 온라인 쇼핑몰 데이터를 수집하고 정제하여 데이터 베이스에 저장한다. SNS 데이터 분석 단계에서 소셜 네트워크 데이터에서 일정 시간대별 핫 토픽을 추출하고 온라인 쇼핑몰 사용자 패턴을 분석하는 과정을 수행한다. 이후 앞선 단계들의 결과를 이용하여 최종 트렌드 지수를 계산하여 상품 트렌드를 분석한다.

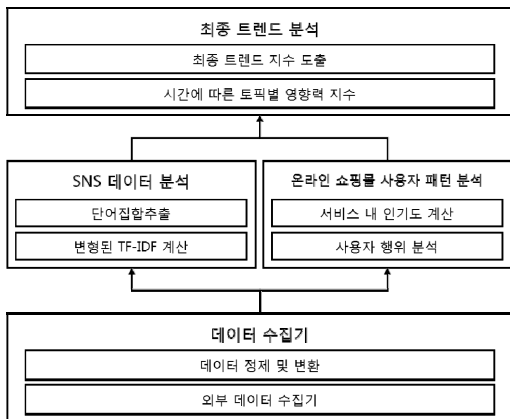


그림 1. 제안하는 기법의 처리절차

2. SNS 데이터 분석

트렌드는 대중들의 인기의 흐름을 의미한다. 소셜 네트워크 서비스를 통해 사람들은 자신과 비슷한 관심사를 가지고 있는 사용자들과 관계 형성을 하며 다양한 정보와 콘텐츠를 공유한다. 따라서 사용자들은 자연스럽게 자신이 흥미를 가지고 있거나 많은 사람들이 관심을 가지고 있는 이슈에 대해서 포스팅을 통해서 표현할 수 있다. 그렇기 때문에 그 당시 대중들이 가장 선호하거나 많이 언급하는 단어를 파악하는 과정은 트렌드 분석에 중요한 역할을 한다. 본 논문에서는 이와 같이 다른 단어들과 비교하여 상대적으로 많이 언급 된 단어를 핫 토픽이라고 정의한다. 트렌드는 시간의 흐름과 대중의 관심도와 직결된다. 따라서 소셜 네트워크 데이터로부터 핫 토픽을 추출하여 트렌드 분석에 사용하기 위해서는 시간에 따른 단어의 출현 빈도의 변화를 모두 고려해야 한다.

제안하는 소셜 핫 토픽 추출 기법은 변형된

TF-IDF[4]를 사용한다. 정보 검색과 텍스트 마이닝에서 사용하는 가중치인 TF-IDF를 이용하면 여러 문서로 이루어진 문서군에서 어떤 단어의 중요도를 알 수 있다. 따라서 문서의 키워드를 추출하거나 문서들 간의 유사도를 계산할 때 사용한다. 단어의 빈도수를 TF에 반영하는 기존의 TF-IDF 기법과 달리 제안하는 기법에서는 소셜 네트워크 서비스에서 순간적으로 나타난 단어를 고려한 분석을 위해 TF에 불린 빈도 방식을 적용하였고, 시간에 따른 트렌드의 변화를 반영하기 위해 서 역문서 빈도인 IDF에 시간 속성을 반영해주었다.

TF_i 는 SNS 데이터 정제 과정을 거쳐 수집된 모든 단어 i 에 대한 출현 유무를 수식(1)과 같이 정의하며 불린 빈도 방식을 이용한다. 최종 트렌드 지수를 도출하는 마지막 단계에서 단어의 출현 빈도를 이용하여 가중치를 부여해주기 때문에 핫 토픽 추출 단계에서는 단어의 출현 유무만을 이용하여 계산한다. 소셜 네트워크 서비스 k 의 데이터에서 하나의 포스팅에 대해서 해당 토픽 i 의 출현 빈도인 frequency가 0보다 크면 1, 한 번도 언급이 되지 않았으면 0으로 산출한다.

$$TF_{i,k} = \begin{cases} 1 & (frequency > 0) \\ 0 & (frequency = 0) \end{cases} \quad (1)$$

일상적인 단어와 핫 토픽을 구별하기 위해서는 해당 하는 단어가 이전에는 많이 출현하지 않다가 현재 시간에 얼마나 많이 출현했는지의 비율이 중요하다. 시간에 따른 변화량을 고려하지 않고 출현 빈도의 크기만가지고 핫 토픽을 추출한다면 일상적으로 많이 언급되는 단어가 포함이 될 수도 있다. 따라서, 시간 구간을 설정하고 타임윈도우 단위로 입력된 단어들의 출현 빈도를 계산한다. 예를 들어, '날씨' 라는 단어는 소셜 네트워크 서비스에서 시간대와 상관없이 출현 빈도가 높지만 일상적인 단어의 성격을 가지고 있기 때문에 핫 토픽으로 볼 수 없다. 그러나 '소나기'라는 단어는 언급량이 갑자기 증가한다. 따라서 변형된 TF-IDF를 사용하면 단어 '소나기'와 같이 이전에는 잘 언급 되지 않았지만 갑자기 언급량이 증가한 단어를 핫 토픽으로 추출할 수 있다. 수식 (2)는 소셜 네트워크 서비스 매체 k 에서 토픽 i 에 대하여 시간 속성을 적용한 IDF 계산이며 이전 시간 대비 현재 시간의 변화량을 측정한 뒤 도출하여 사

용한다. 이때, df (document frequency)는 단어가 포함된 문서의 수를 의미하며 idf 는 df 의 역수를 나타낸다. 다시 말해, 소셜 네트워크 데이터에서 idf 는 해당 단어가 포함된 블로그 포스트의 수 또는 소셜 네트워크 서비스에 올리는 게시글의 역수이다. 제안하는 기법은 시간 단위로 트렌드 분석을 한다. 따라서 idf_t 는 현재 시간 t 를 기준으로 한 시간의 타임 윈도우 내의 idf 값을 나타내고, $idf_{0,t-1}$ 은 0부터 $t-1$ 시간 동안의 idf 값을 의미한다.

$$IDF_{i,k} = \frac{idf_t}{idf_{0,t-1}} \quad (2)$$

잠재 고객의 트렌드를 반영하기 위해 온라인 쇼핑물에 등록되어 있는 상품이 실제 SNS에서 얼마나 영향력이 있는지를 판별한다. 식 (3)은 SNS에서 영향력을 지수 T_i 를 계산하기 위한 것으로 전체 토픽 중 특정 상품 i 에 대한 블로그, 페이스북, 트위터와 같은 k 개의 소셜 네트워크 매체들에서의 영향력 합을 나타낸다. 이때, 가중치 α 는 시스템에서 부여한 상수로 최종 트렌드 분석에 미치는 T_i 에 중요도를 부여한다.

$$T_i = \frac{\sum_{k=1}^n TF_{i,k} \times IDF_{i,k}}{\alpha} \quad (3)$$

3. 온라인 쇼핑물 사용자 패턴 분석

상품의 트렌드를 분석 할 때 소셜 네트워크 데이터만을 사용한다면 상품이 긍정적인 이유로 인기가 있는지, 부정적인 이유인지 알기 힘들기 때문에 사용자의 명시적인 관심을 나타내는 서비스 내 데이터 활용이 필요하다. 온라인 쇼핑 서비스를 사용하는 사용자들은 상품에 대해서 다양한 방법으로 관심을 표출할 수 있다. 실제 상품을 구매하고 점수를 통해서 평가를 내릴 수도 있고, 관심 상품 찜하기와 클릭 수는 특정 상품에 대한 인기를 측정할 수 있는 요소가 된다. 따라서 이러한 요소를 반영하여 상품 트렌드 분석을 해야 한다.

온라인 쇼핑환경을 사용하는 사용자들은 관심 있는 물건에 대해 다양한 행동을 할 수 있는데 이는 관심도

를 나타낼 수 있는 명시적인 사용자 패턴으로 간주할 수 있다. 대표적인 사용자 행위로부터 얻을 수 있는 온라인 쇼핑물 데이터로 특정 상품 i 에 대한 찜하기 수 ($Like_i$), 일정 시간 이상 상품 페이지에 머무른 사용자들의 수($Nstay_i$), 평점 평균($Rating_i$)이 있다. 온라인 쇼핑물에 존재하는 상품의 인기를 분석하기 위해서는 사용자의 명시적인 행위들을 활용하여 계산할 수 있다. 수식 (4)의 P_i 는 서비스 내 사용자의 활동 로그를 이용한 특정 상품 i 의 서비스 내 인기도이다. 본 논문에서는 10초 이상 5분 이내 머무른 사용자들을 의미 있는 클릭으로 계산하였다. 사용자의 클릭 수, 즉 상품을 조회하는 횟수가 적극적인 의사표현 방식인 찜하기 수에 비해서 많기 때문에 가중치를 이용하여 값을 조정해 주었다. 그리고 두 개를 더하여 평점 평균과 곱한 값을 사용하였다.

$$P_i = \frac{(\gamma \cdot Like_i + \delta \cdot Nstay_i) \times Avg(Rating_i)}{1 - \alpha} \quad (4)$$

4. 최종 트렌드 분석

본 논문에서 제안하는 기법의 최종 트렌드 분석 모듈에서는 각 단어 별 트렌드 지수를 이용하여 시간에 따른 트렌드 지수의 비율 값을 고려한다. 소셜 네트워크 데이터를 이용한 영향력 T_i 와 온라인 쇼핑물 사용자 패턴을 고려한 인기도 P_i 값을 가중치 α 로 중요도를 조절할 수 있다. 제안하는 기법은 특정 시간에 발생한 문서를 이용하기 때문에 최종 트렌드 분석 결과를 통해서 시간별 트렌드 상품 및 키워드를 도출할 수 있다.

최종 상품별 트렌드 지수는 소셜 네트워크에서 해당 상품의 영향력과 온라인 쇼핑물 사용자들로부터 얻은 인기도를 반영한다. 소셜 네트워크에서는 현재 많은 언급이 되고 있는 상품 i 와 관련된 단어들의 영향력 지수가 T_i 로 계산되어 인기를 알 수 있고, 온라인 쇼핑물 사용자들의 활동 로그를 통해 P_i 로 계산된 상품 i 의 인기도 지수를 알 수 있다. 예를 들어 현재 방영되고 있는 인기 드라마에서 주인공이 들고 나온 가방이 화제가 되었다면 제안하는 기법을 통한 SNS 데이터 분석에서는 드라마의 제목, 배우의 이름, 가방의 정보 등이 키워드로 선정 될 수 있고 관심 있는 소비자들은 온라인 쇼핑

물에서 상품을 검색하고 찜하기를 누르는 등의 사용자 활동을 하기 때문에 해당 상품에 대한 트렌드 지수가 커진다.

제안하는 기법에서는 SNS 데이터와 온라인 쇼핑몰 데이터를 결합한 상품의 트렌드를 분석한다. 수식 (5)는 상품의 트렌드 지수를 나타낸 것으로 SNS에서 상품의 영향력 지수 T_i 와 서비스 내 인기도 P_i 를 결합하여 계산한다. 이때, T_i 또는 P_i 의 일부 변화가 전체적인 트렌드 지수의 결과를 급격하게 변화시키는 것을 방지시키기 위해 T_i 또는 P_i 의 로그 값을 이용한다.

$$I_i = \log(T_i) + \log(P_i) \quad (5)$$

수식 (6)은 각 단어별 빈도수를 바탕으로 트렌드 지수를 계산한다. 2절의 SNS 데이터 분석 과정에서 변형된 TF-IDF 값을 사용했는데 이때 불린 빈도 방식을 사용하여 단어의 출현 유무만 고려했기 때문에 단어의 빈도에 대한 가중치가 고려되지 않았다. 그러나 상품의 인기도와 사람들의 상품 언급 횟수 $Freq_i$ 는 밀접한 상관관계를 가지고 있기 때문에 이를 반영할 필요가 있다. 따라서, SNS 언급량과 상품의 트렌드 지수를 더한 값인 I_i 에 해당 단어가 소셜 네트워크 데이터에서 출현한 빈도수를 곱하여 계산 할 수 있다.

$$I_t^i = \sum Freq_i \times I_i \quad (6)$$

수식 (7)은 수식 (6)에서 계산한 I_t^i 에 대해서 시간에 따른 트렌드 지수의 비율을 반영한 값을 나타낸다. 시간 t 에 대한 트렌드 지수와 이전 시간인 $t-1$ 에 대한 트렌드 지수의 합과 차를 이용한 비율로 계산한다. 트렌드는 시간의 흐름에 따라 사람들의 관심이 변하며 생기는 현상이다. 트렌드를 분석하기 위해서는 시간대에 따른 상품의 관심도를 반영할 필요가 있다. 따라서 과거에 비해 최근의 변화량이 큰 단어 또는 상품에 대한 트렌드를 분석 할 수 있는 척도가 되는 계산식을 이용하였다.

$$R_t^i = \frac{I_t^i - I_{t-1}^i}{I_t^i + I_{t-1}^i} \quad (7)$$

[표 1]은 제안하는 기법의 시간에 따른 최종 트렌드 지수를 계산하는 예시를 나타낸다. 시간에 따른 키워드가 '미세먼지'와 '유기농'으로 추출되고 각 단어가 한 번씩 출현했다고 하면 단어별 트렌드 지수는 출현빈도와 영향력 지수를 곱한 값이다. 단어별 트렌드 지수를 수식(7)에 대입하여 각 단어의 시간에 따른 트렌드 지수를 계산할 수 있다. [표 1]에서 단어 '미세먼지'에 대한 트렌드 지수 값이 '유기농'에 대한 값보다 크기 때문에 현재 트렌드를 단어 '미세먼지'라고 할 수 있다.

표 1. 최종 트렌드 지수 예시

시간	핫 토픽	영향력 지수 (T_i)	트렌드 지수 (I_t^i)	시간에 따른 트렌드 지수 (R_t^i)
13시	'미세먼지'	0.4	$1 * 0.4 = 0.4$	'미세먼지' $= \frac{1.2 - 0.4}{1.2 + 0.4}$ $= 0.5$
	'유기농'	0.6	$1 * 0.6 = 0.6$	
14시	'미세먼지'	1.2	$1 * 1.2 = 1.2$	'유기농' $= \frac{0.2 - 0.6}{0.2 + 0.6}$ $= -0.5$
	'유기농'	0.2	$1 * 0.2 = 0.2$	

IV. 성능 평가

제안하는 기법의 우수성을 검증하기 위하여 성능 평가를 수행하였다. 성능 평가는 Intel core i7-6700 CPU 4.00GHz, 64GB 메모리를 가지는 시스템에서 Apache Spark 환경에서 Scala 언어를 이용하여 구현하였다. 성능 평가를 위해 2018년 03월 1일부터 3월 26일까지 한 달 동안 화장품에 대한 블로그 데이터와 사용자 로그를 수집한 샘플 데이터를 사용했다. 성능 평가 환경은 [표 2]와 같다.

표 2. 성능 평가 환경

항목	값
CPU	Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz, 4 Core
Memory	64GB
OS	CentOS Linux release 7.4
사용 플랫폼 및 DB	Apache Spark 2.2.1, MongoDB 3.6.3
사용 데이터 셋	2018.03.01. ~ 2018.03.26. 화장품에 대한 블로그 데이터와 사용자 로그

[표 3]은 성능 평가 기간인 2018년 3월 한 달간 1주 단위로 블로그 데이터와 사용자 로그를 함께 분석하여 나온 트렌드 상위 10개 단어 키워드에 대한 결과이다. 화장품에 대한 데이터를 사용했지만 상품명에 한정되어 있지 않고 해당 화장품을 사용한 유명 연예인 이름이나 드라마 속의 화장품이 화제가 된 경우 트렌드 키워드로 드라마 제목도 함께 검출된다. 화장품 트렌드의 경우 상품 주 소비자층의 관심도나 성향에 많이 영향 받는 것을 확인할 수 있다. 결과로 제시된 키워드들을 보면 2주차에 유명 연예인의 화보가 공개되어 대중들의 인기를 끌어 이를 반영한 결과가 나타나고, 3주차에는 인기 드라마 속 배우의 메이크업이 화제가 되어 키워드의 변화가 있는 것을 볼 수 있다. 기존의 서비스 내 사용자 데이터만 고려한다면 상품의 인기 이유를 예측하기 힘든 문제점이 있다. 하지만 많은 사람들이 사용하는 소셜 네트워크 데이터와 결합하여 트렌드 분석을 수행하면 결과 도출된 키워드를 통해 트렌드 흐름을 쉽게 유추할 수 있다.

표 3. 상위 10개 단어 키워드에 대한 결과

1주차 (03.05 ~ 03.11)	2주차 (03.12 ~ 03.18)	3주차 (03.19~03.26)
자연발효 에너지 오일	력키박스	리뉴얼
유자 수분씨 화이트닝 앰플 인 크림	한란 슬리핑 마스크	플레이 컬러 아이즈 와인파티
유기농	캔메이크	루주뷔르꾸뛰르 베르니 아 레브르틴트
천연성분	UV 아쿠아 리치 워터리 에센스	포토에이지앰플
엘리자베스아덴	루주뷔르꾸뛰르 베르니 아 레브르틴트	아토팜MLE 수딩 젤 로션
미세먼지	대용량	신생아
맥 립스틱	지속력	추리의여왕
벨벳 루즈 힐	홍조	AOA
포토에이지앰플	김고은	컬러 인 리퀴드립스 무스
주름관리	랑콤 홀리데이 컬렉션	벨벳 루즈 힐

소셜 네트워크 서비스 데이터를 이용함으로써 해당 화장품의 인기도 뿐만 아니라 관련된 이슈 그리고 트렌드 키워드로 검출된 이유를 추측할 수 있다. 예를 들어, 트렌드 키워드에서 상품명과 유명 연예인의 이름이 같이 검출된 경우 해당 연예인이 화보에서 사용한 화장

품이 이슈가 되었다는 것을 실제 블로그 글의 내용 예시를 통해 파악할 수 있다. [표 4]는 트렌드 분석 결과 검출된 키워드와 실제 사용자들이 작성한 글 내용과의 연관관계를 나타내고 있다.

표 4. 트렌드 키워드와 실제 내용 연관 예

결과검출일	뷰티 트렌드 키워드	블로그 글 내용 예시
03.05 ~ 03.11	자연발효 에너지 오일, 주름관리	주름관리합니다. 이니스프리 자연발효 에너지 오일 추천해요
03.05 ~ 03.11	엘리자베스아덴, 미세먼지	미세먼지로부터 피부를 보호하는 엘리자베스 아덴 '프리베이지 시티 스마트'
03.12 ~ 03.18	력키박스, 한란 슬리핑 마스크	이니스프리 력키박스 스누피 한란을 씁니다. 한란 슬리핑마스크로 원하던 거 딱 골라왔어요
03.12 ~ 03.18	김고은, 랑콤 홀리데이 컬렉션	화보 속 김고은 메이크업은 랑콤 홀리데이 컬렉션 심쿵 아이템들과 화보사진까지
03.19 ~ 03.26	리뉴얼, 마몽드 크리미티 컬러 밤	마몽드 크리미티 컬러 밤 리뉴얼 출시되었어요! 제품은 라이트 10종,
03.19 ~ 03.26	아토팜MLE 수딩 젤 로션, 신생아	아기 수딩젤 아토팜MLE 수딩 젤 로션으로 신생아 태열 관리하고 있어요
03.19 ~ 03.26	벨벳 루즈 힐, 추리의여왕, AOA	추리의여왕2 AOA민아 립스틱! 클리오 루즈힐 벨벳발라볼

[그림 2]는 상품 클릭 수, 상품 평점과 같은 사용자 행동만을 반영한 서비스 데이터만 고려한 키워드에 대한 트렌드 지수와 소셜 데이터를 결합하여 같이 분석했을 때의 결과를 각각 나타낸다. 서비스 내의 데이터만 고려하게 된다면 [그림 2]의 왼쪽과 같이 상품명에 대한 순위 정보만 알 수 있다. 하지만 서비스 내의 사용자의 데이터와 소셜 네트워크에서 언급되는 데이터를 같이 분석한 결과 상품명 '플레이 컬러 아이즈 와인파티'의 경우 소셜 네트워크 서비스에서 사람들의 인기를 반영하여 순위의 변화가 있는 것을 확인할 수 있다. 따라서 제안하는 기법을 사용하면 소셜 네트워크 서비스에서 상품과 관련하여 이슈가 되어 있는 키워드의 순위가 반영되어 실제 대중들에게 인기 있는 트렌드에 대한 분석 결과를 얻을 수 있다. 사용자는 온라인 쇼핑물에서 [그림 2]의 오른쪽 결과와 같은 데이터를 보고 현재 이슈가 되고 있는 키워드와 상품을 참고하여 구매 결정에 도움을 줄 수 있다.

순위	서비스데이터*	순위	서비스데이터+ 소셜 데이터**	순위 변동
1	닥터제임 모이스트 얼션	1	리뉴얼	NEW
2	루루웨르푸워르	2	플레이 컬러 아이즈 와인파티	▲5
3	베르니 아 맥브린트	3	루루웨르푸워르	NEW
4	프로에이지얼름	4	베르니 아 맥브린트	▼1
5	UV아쿠아 리치 워터리 에센스	5	프로에이지얼름	NEW
6	이니스프리 한민 스킨링 마스크	6	아르람MIE 수딩 젤 로션	▲1
7	아르람MIE 수딩 젤 로션	7	산쟁이	NEW
8	플레이 컬러 아이즈 와인파티	8	추리의여왕	NEW
9	비지오겔DMT 페이스 로션	9	AOA	NEW
10	유자수분케어미닝 애플 인 크림	10	컬러 인 리퀴드스 무스	NEW
	벨벳 푸즈 힐	10	벨벳 푸즈 힐	-

* 온라인 쇼핑몰 서비스 내 데이터를 이용하여 사용자 패턴 분석만을 통한 상품 트렌드 분석 결과
 ** 서비스 내 데이터와 SNS 데이터를 이용하여 제안하는 기법의 트렌드 분석 결과

그림 2. 기간별 최종 트렌드 지수 변화

V. 결론

본 논문에서는 온라인 쇼핑몰에서 소셜 네트워크의 특성을 고려한 상품 트렌드 분석 기법을 제안하였다. 제안하는 기법은 변형된 TF-IDF를 사용하여 시간에 따른 소셜 핫 토픽을 고려하였고 상품의 인기도를 고려하여 상품이나 특정 관심 분야에 대한 트렌드 분석 결과를 도출하였다. 제안하는 기법은 소셜 네트워크 서비스 데이터를 활용한 성능 평가를 통해 실제 화제가 되고 있는 트렌드와 유사한 결과를 얻었다. 이를 통해 온라인 쇼핑몰을 방문하는 사용자들에게 이슈가 되고 있는 상품들의 트렌드를 제공하여 상품 구매 시 최신 트렌드를 반영한 결정을 할 수 있도록 활용할 수 있다. 그러나 제안하는 트렌드 분석 기법을 통해서만 해당 키워드나 상품이 긍정적이거나 부정적인 이유로 이슈가 된 것인지 알 수 없다. 따라서 향후 연구에서는 소셜 네트워크 데이터를 사용 할 때 게시글의 긍부정 분석을 반영할 예정이다. 또한, 트렌드 분석에 대한 계산 비용을 감소시키기 위해 시간 복잡도를 분석하여 제안하는 기법을 실제 시스템에 적용하여 상품 추천에도 활용할 예정이다.

참 고 문 헌

[1] W. Yanyan, "Empirical Analysis of Factors Influencing Consumers' Satisfaction in Online Shopping Agricultural Products in China,"

Journal of Electronic Commerce in Organizations, Vol.16, No.3, pp.64-77, 2018.

[2] C. Ju, J. Wang, and G. Zhou, "The commodity recommendation method for online shopping based on data mining," Multimedia Tools and Applications, Vol.78, No.21, pp.30097-30110, 2019.

[3] R. V. Karthik, S. Ganapathy, and A. Kannan, "A Recommendation System for Online Purchase Using Feature and Product Ranking," Proc. International Conference on Contemporary Computing, pp.1-6, 2018.

[4] I. T. Afolabi, O. S. Makinde, and O. O. Oladipupo, "Semantic Web mining for Content-Based Online Shopping Recommender Systems," International Journal of Intelligent Information Technologies, Vol.15, No.4, pp.41-56, 2019.

[5] V. Gala, V. Deshpande, I. Ferwana, and M. Milanova, "Product Sentiment Trend Prediction," Proc. International Conference on Social Computing and Social Media, pp.274-283, 2018.

[6] <https://www.twitter.com>, 2019.10.10.

[7] <https://www.facebook.com>, 2019.10.10.

[8] <https://www.section.blog.naver.com>, 2019.10.10.

[9] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," Proc. ACM SIGMOD International Conference on Management of Data, pp.1155-1158, 2010.

[10] I. Yahav, O. Shehory, and D. Schwartz, "Comments Mining With TF-IDF: The Inherent Bias and Its Removal," IEEE Transactions on Knowledge and Data Engineering, Vol.31, No.3, pp.437-450, 2018.

[11] S. A. Gavhane, S. B. Bhadave, and K. Vengatesan, "Review on Latest Trending Topic Detection in Twitter With Stream Processing (Using Fission Pattern)," International Journal of Applied Evolutionary Computation, Vol.10, No.2, pp.43-47, 2019.

[12] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering

clusters in large spatial databases with noise,” Proc. International Conference on Knowledge Discovery and Data Mining, pp.226-231, 1996.

- [13] 노연우, 김대운, 한지은, 육미선, 임종태, 복경수, 유재수, “소셜 네트워크에서 사용자의 영향력을 고려한 핫 토픽 예측 기법,” 한국콘텐츠학회논문지, 제15권, 제8호, pp.24-36, 2015.
- [14] 노연우, 임종태, 복경수, 유재수, “소셜 네트워크 환경에서 변형된 TF-IDF를 이용한 핫 토픽 예측 기법,” 정보과학회 컴퓨팅의 실제 논문지, 제23권, 제4호, pp.217-225, 2017.
- [15] 이홍주, “클릭스트림 데이터를 활용한 전자상거래에서 상품추천이 고객 행동에 미치는 영향 분석,” 한국경영과학회지, 제33권, 제3호, pp.59-76, 2008.

저 자 소 개

박 수 빈(Soo-Bin Park) 준회원



- 2018년 2월 : 한남대학교 산업경영 공학과, 컴퓨터공학과(공학사)
- 2018년 3월 ~ 현재 : 충북대학교 빅데이터협동과정 석사과정

〈관심분야〉 : 빅데이터, 분산 처리, 소셜 네트워크, 상품 추천 등

최 도 진(Do-Jin Choi) 정회원



- 2014년 2월 : 한국교통대학교 컴퓨터공학과(공학사)
- 2016년 2월 : 한국교통대학교 컴퓨터공학과(공학석사)
- 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과 박사과정

〈관심분야〉 : 연속 질의 처리, 그래프 스트림, 빅데이터 처리 등

유 재 수(Jae-Soo Yoo) 중신회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술원 전산학과(공학석사)
- 1995년 2월 : 한국과학기술원 전산학과(공학박사)
- 1995년 2월 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ~ 현재 : 충북대학교 전자정보대학 정교수
〈관심분야〉 : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오 인포메틱스, 빅데이터 처리 등

복 경 수(Kyoung-Soo Bok) 중신회원



- 1998년 2월 : 충북대학교 수학과(이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신공학과(공학박사)
- 2005년 3월 ~ 2008년 2월 : 한국과학기술원 정보전자연구소 Postdoc

- 2008년 3월 ~ 2011년 2월 : 가인정보기술 연구소 차장
- 2011년 3월 ~ 2019년 8월 : 충북대학교 전자정보대학 정보통신공학부 초빙교수
- 2011년 9월 ~ 현재 : 원광대학교 SW융합학과 조교수
〈관심분야〉 : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터 처리 등