

## 온라인 쇼핑몰에서 소셜 네트워크 특성을 고려한 상품 트렌드 분석 기법

### Product Trend Analysis Scheme Considering Social Network Features in Online Shopping Malls

박수빈\*, 김이나\*, 최도진\*\*, 박재열\*\*, 유승훈\*\*,  
송재오\*\*\*, 복경수\*\*, 유재수\*\*+  
충북대학교 빅데이터학과\*, 충북대학교 정보통신  
공학과\*\*, (주)케이아이씨티 기술융합연구소\*\*\*

Soobin Park\*, Ina Kim\*, Dojin Choi\*\*,  
Jaeyeol Park\*\*, Seunghun Yoo\*\*, Joo Song\*\*\*,  
Kyoungsoo Bok\*\*, Jaesoo Yoo\*\*  
Chungbuk National University\*, \*\*,  
Research Institute of K-ICT Co.,Ltd\*\*\*

#### 요약

온라인 쇼핑몰에서 소비자들이 원하는 상품을 노출시켜 정보를 제공하기 위해서는 상품의 트렌드 분석에 대한 연구가 필요하다. 본 논문에서는 대량의 SNS 데이터와 서비스 내 사용자 데이터를 결합하여 보다 효율적인 상품 트렌드 분석 기법을 제안한다. 온라인 소셜 네트워크의 대중화로 소비자들은 시공간에 구애받지 않고 상품에 대한 정보를 SNS로 교류할 수 있다. 제안하는 기법은 이 과정에서 발생한 SNS 데이터와 사용자 성향 데이터에 시간 속성을 고려하여 상품 트렌드를 분석한다.

## I. 서론

온라인 쇼핑몰에서 소비자들의 관심 트렌드를 분석하여 이를 상품 판매에 활용하는 연구가 진행되고 있다. 기존의 상품트렌드 분석에서는 온라인 쇼핑몰에서 수집된 클릭스트림이나 세션타임만을 분석하여 고객의 관심변화에 대해 연구하였다[1]. 하지만 기존의 사용자 로그만 고려한 추천이나 트렌드 분석은 서비스 내에서 적극적인 활동을 했던 사용자에게는 효과적이지만, 해당 서비스를 처음 접해보거나 사용자 활동이 적은 잠재적 고객에게는 적용하기 어렵다는 한계를 가지고 있다. 따라서 기존의 기법에서 더 나아가 넓은 범위의 사용자들에게 적용시킬 수 있도록 대중의 관심도와 연관되는 SNS(Social Network Service) 분석이 필요하다. 온라인 소셜 네트워크가 대중화되기 이전에는 상품에 대한 정보나 트렌드를 파악할 수 있는 채널이 부족했다. 그러나 최근 SNS의 발달로 소비 패턴과 정보 확산 방식의 변화가 이루어졌다. 소비자들은 특정 상품에 대한 리뷰나 기대치에 대한 의견을 SNS를 통해 주고받으며 다량의 정보를 생산한다.

따라서 SNS는 상품의 트렌드를 파악하기 위한 중요한 데이터를 제공하고 있다. 이와 같은 온라인 소셜 네트워크 특성을 고려하면 사용자 활동 데이터가 없거나 적은 경우를 보완하여 온라인 쇼핑몰에 맞는 효과적인 트렌드 분석을 할 수 있다. 그러나 SNS 데이터에서 핫 토픽을 검출하는 기존의 기법에서는 시간에 따른 단어 출현의 빈도수 변화를 기준으로 하고 있기 때문에 시간의 변화에 따른 핫 토픽을 검출할 수 없다는 단점이 존재한다[2].

본 논문은 온라인 쇼핑몰에서 수집된 데이터와 SNS 데이터를 결합한 상품 트렌드 분석 기법을 제안한다. 제안하는 기법은 시간의 영향을 받아 빠르게 바뀌는 트렌드 특성에 따라 트렌드 분석에 시간 속성을 고려하였다. 제안하는 기법은 시간 속성을 고려한 변형된 TF-IDF 기법 [3]을 통해 핫 토픽 지수를 계산한다. 온라인 쇼핑몰에서 사용자들이 생성한 세션타임, 행위, 의견을 활용하여 서비스 내 인기 지수를 계산한다. 두 값을 이용하여 최종적인 트렌드 지수를 도출하고, 도출된 값을 이용하여 사용자에게 트렌드 결과를 제시한다.

## II. 제안하는 상품 트렌드 분석 기법

### 1. 특징

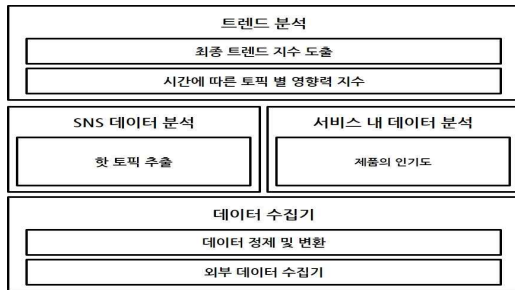
서비스 내 사용자의 활동 로그가 충분하지 않을 경우 해당 고객에 대한 맞춤 트렌드를 제공할 수 없는 기존 기법의 한계를 해결하기 위해, 제안하는 기법은 SNS 데

+ 교신 저자 : yjs@chungbuk.ac.kr

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업(No. NRF-2017M3C4A7069432), 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. 2016R1A2B3007527)과 중소기업청에서 지원하는 2017년도 산학연협력 기술개발사업(No.C0514301)의 지원을 받아 수행된 연구임

이터와 사용자 활동 로그를 결합하여 종합적인 상품 트렌드 결과를 검출한다. 또한 데이터 분석에 시간 속성을 고려하여 시시각각 변하는 트렌드를 빠르게 파악할 수 있다. 변형된 TF-IDF 알고리즘을 이용하여 소셜 핫 토픽을 추출하고, 사용자의 행위 분석을 통하여 사용자의 패턴을 분석한다.

그림 1은 제안하는 기법의 시스템 구조도를 나타낸다. 데이터 수집기를 사용하여 포털사이트의 데이터와 SNS 데이터를 수집하고 정제하여 DB에 저장한다. SNS 데이터 분석을 통해 핫 토픽을 추출하고 서비스 내 데이터 분석을 통해 제품의 인기를 분석한다. 시간에 따른 토픽별 영향력 지수를 계산하여 최종 트렌드 지수를 도출한다.



▶▶ 그림 1. 제안하는 기법의 시스템 구조

## 2. 소셜 핫 토픽 추출

트렌드는 대중들의 인기의 흐름을 의미한다. 그렇기 때문에 그 당시 대중들이 가장 선호하거나 언급하는 단어를 추출해야하기 때문에 핫 토픽 추출 과정을 거치게 된다. 제안하는 소셜 핫 토픽 추출 기법은 변형된 TF-IDF를 사용한다. TF는 불린 빈도 방식을 이용하여 산출한다. IDF를 산출 할 때는 시간 속성을 적용하여 시간에 따른 변화량을 측정한 뒤 도출하여 사용한다. 식 (2)에서  $df$ 는 단어가 포함된 문서 수를 의미하며,  $idf$ 는  $df$ 의 역수를 나타낸다.

$$TF = \begin{cases} 1 & (\text{frequency} > 0) \\ 0 & (\text{frequency} = 0) \end{cases} \quad (1) \quad IDF = \frac{idf_i}{idf_{0,i-1}} \quad (2)$$

## 3. 쇼핑몰 사용자 패턴

서비스를 사용하는 사용자의 상품에 대한 관심을 표출하는 행위인 좋아요와 평점, 클릭 수를 통해 특정 상품의 인기를 측정할 수 있다. 식 (3)과 식 (4)는 제품 영향력 지수를 도출하기 위한 수식이다. 식 (3)은 특정 제품 A에 대한 SNS 언급량을 나타낸다. 식 (4)는 서비스 내 사용자의 활동 로그를 이용한 특정 제품 A의 서비스 내 인기를 나타낸다. 온라인 쇼핑몰을 사용하는 사용자들의 관심도를 나타낼 수 있는 요소로써 좋아요 수(Like), 평균 평점(Rating), 일정 시간 이상 머무른 사용자들의

수(Num<sub>stay</sub>)를 기준으로 하여 해당 제품의 영향력 지수를 측정한다.

$$I_{P_A} = \frac{\sum mTF - IDF}{\alpha} \quad (3) \quad I_{P_A} = \frac{(\gamma \cdot Like + \delta \cdot Num_{stay}) \times Avg(Rating)}{\beta} \quad (4)$$

식 (5)는 최종적인 제품별 영향력 지수를 계산하기 위한 수식이다. 계산된 SNS 언급량과 제품 영향력 지수를 더하여 최종적인 제품별 영향력 지수를 계산한다.

$$I_{P_A} = \log(I_{P_A}^{sns}) + \log(I_{P_A}^{in}) \quad (5)$$

## 4. 최종 트렌드 분석

제안하는 기법은 각 단어 별 트렌드 예측 지수와 시간에 따른 트렌드 지수의 비율 값을 고려한다. 식 (6)은 각 단어별 빈도수를 바탕으로 트렌드 예측 지수를 계산한다. 식 (7)은 시간에 따른 트렌드 지수의 비율 값을 나타낸다. 과거에 비해 최근의 변화량이 큰 단어 또는 제품에 대한 트렌드를 예측할 수 있는 척도가 된다.

$$I_t^P = \sum Freq \times I_{P_A} \quad (6) \quad R_t^P = \frac{I_t^P - I_{t-1}^P}{I_t^P + I_{t-1}^P} \quad (7)$$

## III. 결론

본 논문에서는 온라인 쇼핑몰에서 소셜 네트워크의 특성을 고려한 상품 트렌드 분석 기법을 제안하였다. 제안하는 기법에서는 변형된 TF-IDF를 사용하여 시간에 따른 소셜 핫 토픽을 고려하였고 서비스 내 제품의 인기를 고려하여 상품이나 특정 관심 분야에 대한 트렌드 분석 결과를 도출하였다. 향후 연구에서는 용어 클러스터링을 활용하여 크롤러를 통해 소셜 핫 토픽 정확도를 향상시킬 예정이다.

## ■ 참고 문헌 ■

- [1] 김재경, 안도현, 조윤호 “인터넷 쇼핑몰을 위한 데이터마케팅 기반 개인별 상품추천방법론의 개발”, 한국지능정보시스템학회논문지, 제9권, 제3호, pp.177-191, 2003
- [2] 김남규, 이동훈, 최호창, William, X. S. W “텍스트 분석 기술 및 활용 동향”, 한국통신학회논문지, 제42권, 제2호, pp.471-492, 2017
- [3] 노연우, 임종태, 복경수, 유재수 “소셜 네트워크에서 사용자의 영향력과 전문성을 고려한 핫 토픽 예측 기법”, 한국정보과학회 학술발표논문집, pp.314-316, 2016