한국컴퓨터정보학회 동계학술대회 논문집 제20권 제1호 (2012. 1)

# 사용자 정보에 따른 오피니언 마이닝 신뢰성 향상 방법

임지연°, 김이준*, 김응모*
°*성균관대학교 컴퓨터공학과
e-mail:{01039374479, uk3080789}@naver.com, umkim@ece.skku,ac.kr

# The way to improve trust ratio of opinion mining by using user information

Jiyeon Lim°, Leejun Kim*, Ungmo Kim*
°*Dept of Computer Engineering, SungKyunKwan University

● 요 약 ●

소셜 네트워크의 부상과 함께 소셜 네트워크를 이용하여 홍보를 하는 소셜 커머스 시장도 커지고 있다. 소셜 커머스의 경우 일정한 인원 이상이 구입을 해야 거래가 성립한다. 그래서 실질적으로 환불이나 반품이 힘들기 때문에 그만큼 상품평이 구매에 미치는 영향이 크다고 볼 수 있다. 하지만 이러한 상품평의 경우에도 개인의 상황이나 취향 등에 따라 상품평이 주는 정보의 방향이 크게 바뀔 수 있다는 단점도 있다. 본 논문에서는 오피니언 마이닝을 이용하여 의미를 추출하고, LIWC를 통해 사용자의 기본 정보 및 심리 등을 파악하여 보다 정확한 고객의 개인별 상황에 맞는 상품 평점을 제시한다.

**키워드:** 오피니언 마이닝(Opinion Mining), LIWC, Social Commerce

## I. 서론

본 논문은 소셜 커머스에서 상품평의 효과적인 오피니언 마이닝 방법과 함께 결과에 신빙성을 줄 수 있는 사용자 정보추출 하는 법에 대해 기술한다. 고객은 본인에게 맞는 상품을 고르기 위해서 많은 이들의 상품평을 참고한다. 이 때, 상품평을 남긴 사용자의 연령대나 성별 등을 고려함으로써 신뢰성 높은 정보를 제공할 수 있다. 우리는 소셜 커머스 상에서 오피니언 마이닝을 했으며, 추출된 정보의 신빙성을 위하여 사용자의 정보를 추출했다.

본 논문의 구성은 다음과 같다. 섹션2에서는 관련 연구에 대해서 알아본다. 섹션 3에서는 케이스 스터디를 적용해보고, 섹션 5에서 결론을 맺는다.

## II. 관련 연구

### 1. 관련연구

#### 1.1 Social Commerce

현실 세계에서의 관계가 가상 세계로 이동하면서 소셜 네트워크의 성장세가 두드러지고, 소셜 네트워크내에서의 입소문을 통한 공동구매 형식의 소셜 커머스도 활기를 띄고 있다. 본 논문에서 신뢰성을 보장하기위해 개인의 상황에 맞는 배경 지식을 고려한 수치화된 데이터를 참고한다. 이로써 아는 사람만 안다는 공동구매 형식에서 벗어나 좀 더 저렴한 가격에 홍보효과를 누릴 수 있다.

#### 1.2 Opinion Mining

오피니언 마이닝은 사용자가 직접 평점을 입력하지 않더라도 글에서 사용자의 의견을 수치적으로 표현하는 기술이다. 이는 상품이나 서비스를 구매하고자 원하는 고객에게 많은 정보를 제공한다. 우리는 오피니언 마이닝 기법 중 사용자의 기본 정보와 심리 파악 등을 통해 자신에게 맞는 의견을 추출하는 Sentiment Classification 기법을 사용한다.

#### 1.3 LIWC (Linguistic Inquiry and Word Count)

LIWC는 단어 수세기 방법을 이용하여 텍스트를 분석하고 사용자의 성별, 나이 등 기본정보와 심리적 상태를 확인할 수 있는 효과적인 방법이다. 노인은 1인칭 단수 대명사 사용이 줄고 긍정적 정서 단어를 많이 사용하며 인과 관계 단어나 통찰 단어, 현재 동사, 미래 동사 그리고 긴 단어를 많이 사용 한다(Stone, & Pennebaker, 2002). 본 논문에서는 이러한 특성들을 이용하여 소셜 커머스 상품평 의미를 판단하는데 도움이 되는 배경지식을 추출한다.

## III. 본론

### 3.1 소셜 커머스 이용횟수 정의

소셜 커머스의 상품평을 마이닝하기 위해서 우선 사용자의 소셜커머스 이용횟수를 정의한다. 우리는 실제 사용자의 정보를 추출하는 범위를 사용횟수 3~10회인 사용자로 제한했다.

I ordered a small and I have received what seems to be a male-small. It doesn't fit at all. How can I exchange this for the shirt I ordered? Thanks so much for taking care of this problem so quickly! A replacement shirt was shipped right away and I'm told it should get in tomorrow. Excellent service! :)

[그림 1] 소셜커머스내 상품평

### 3.2 LIWC 적용

LIWC는 사람의 글에서 심리를 연구한 프로그램으로 단어의 수에 따른 심리 연구를 통해 얻어낸 데이터에 따라 몇 개의 분류로 나누고, 연령이나 성별과 관련해서 나타나는 글쓰기 특징들을 추출 하여 사용자의 정보에 반영한다.

[표 1] LIWC를 이용한 상품평 분석 결과

| LIWC dimension | Your data | Personal texts | Formal texts |
|---|---|---|---|
| Self-references (I, me, my) | 8.47 | 11.4 | 4.2 |
| Social words | 3.39 | 9.5 | 8.0 |
| Positive emotions | 5.08 | 2.7 | 2.6 |
| Negative emotions | 0.00 | 2.6 | 1.6 |
| Overall cognitive words | 8.47 | 7.8 | 5.4 |
| Articles (a, an, the) | 6.78 | 5.0 | 7.2 |
| Big words (〉 6 letters) | 20.34 | 13.1 | 19.6 |

### 3.3 컨텐츠 의미 추출

오피니언 마이닝을 사용하여 우리는 사용하여 익명 사용자가 글을 통해 드러내는 긍정과 부정의 감정들을 분석하여 익명 사용자의 기호나 생각을 추출한다. 여기서 추출된 사용자의 기호나 생각들에 LIWC를 통한 분석 단계에서 추출한 심리학적 데이터들을 적용한다.

[표 2] 상품평 Sentiment 분석 결과

| No | Positive words | Negative words | Sentiment |
|---|---|---|---|
| 1 | | not | -1 |
| 2 | quickly | | +1 |
| 3 | excellent | | +1 |
| total | | | +1 |

## Ⅳ. 결론

우리는 소셜커머스 내 상품평의 오피니언 마이닝 신뢰도를 높이기 위해 LIWC를 적용시켜 보았다. 실험은 신뢰도를 주기 충분했지만, 사용자 특성별 관심사 반영 비율에 더욱 세부적인 실험이 필요할 것으로 보인다. 그리고 우리는 이 실험을 통해 소셜 네트워크 상에서 신상 관련 정보가 기록되어 있지 않아도 대략적으로 파악할 수 있음을 밝혀냈다. 이것은 포렌식, 마케팅 등 여러 분야에서 유용하게 사용 가능하다. 또한 소셜커머스의 올바른 발전을 위하여 다른 방식의 신뢰도 계산법에 대해 연구해야 할 것이다.

## 참고문헌

[1] Kyung Soo Cho, Jae Yoel Yoon, Iee Joon Kim, Ji Yeon Lim, Ung-Mo Kim, “Mining information of anonymous user on a social network service”, 2011.

[2] Pennebaker, J.W. and Booth, R.J. and Francis, M.E., “Linguistic inquiry and word count (LIWC2007)” Austin, TX: LIWC (www. liwc. net), 2007

[3] Kim, S.M. and Hovy, E., “Determining the sentiment of opinions”, Proceedings of the 20th international conference on Computational Linguistics, pp. 1367, 2004.

[4] Pang, B. and Lee, L., “Opinion mining and sentiment analysis”, Foundations and Trends in Information Retrieval, vol. 2, no 1-2, pp. 1-135, 2008.