
스마트워치 SNS 리뷰 데이터와 오피니언 마이닝을 통한 감성 분석 처리에 대한 연구

신동현* · 최용락**

*숭실대학교 소프트웨어특성화대학원

A Study on Smartwatch review data of SNS and sentiment analytical using
opinion mining

Donghyun Shin* · YongLak Choi**

*Graduate School of Software, Soongsil University

E-mail : kylesoshin@gmail.com

요 약

IoT(Internet of Things)에 대한 관심과 함께 웨어러블 디바이스 또한 차세대 융합 기술의 핵심으로 그 관심이 증가하고 있다. 특히, 초기 단계인 스마트워치 시장의 선점을 위하여 여러 기업들이 경쟁하고 있으며, 사용자들은 이러한 경쟁 속에서 각 기기에 대한 의견을 SNS를 통하여 공유하며 그에 대한 선호도를 표출하고 있다.

따라서 본 논문에서는 스마트워치에 관련된 속성과 감성단어들에 대한 감성사전을 먼저 구축한 뒤 이를 토대로 의견 데이터 모델을 통하여 수집된 SNS의 데이터를 속성별로 분류한다. 이후 수집된 데이터를 자연언어 처리 기법을 이용하여 전반적 극성 및 속성별 극성을 판단하고 이를 통하여 각 스마트워치 리뷰에 대한 분석을 수행하고자 한다. 그리고 수집된 자료 분석을 통하여 사용자들이 선호하는 스마트워치의 속성을 파악할 수 있도록 하고 이를 통해 각 기기별 발전방향을 판단하는데 기여하도록 한다.

ABSTRACT

Wearable device, along with IoT(Internet of Things), is considered the core of upcoming generation's convergence technology. Companies are intensely competing one another for prior occupation in the smartwatch market. Consumers that use smartwatch express their preferences by sharing their opinions through SNS(Social Networking Service). Through this study, emotions dictionary is built, which consists of attributes and emotional words related to smartwatch. Based on the emotions dictionary, SNS data has been categorized according to the attributes through opinion data model. Afterwards, overall polarity and attribute polarity of collected data are distinguished through natural language parsing, followed by an analysis of smartwatch reviews. This study will contribute to determination of which attributes of smartwatch to be improved, to arise consumer's interest for individual smartwatch.

키워드

스마트워치, SNS, 오피니언 마이닝, 감성 분석

1. 서 론

미국의 정보 기술 연구 및 자문 회사인 가트너(Gartner, Inc.)는 2015년 10대 전략 기술 동향을 발표하면서 모든 사물에 인터넷을 연결해 사물끼리 정보를 공유하는 환경을 의미하는 사물인터넷(IoT)을 두 번째 순위로 게재하였다.[1] 사물인터넷에 대한 관심이 증가하면서 이를 일상생활에서 더욱 쉽게 제어할 수 있는 웨어러블 디바이스 또한 차세대 융합 기술의 핵심으로 관심을 받고 있다. 실제로 많은 시장조사 기관에서는 웨어러블 디바이스의 시장이 지속적으로 성장할 것으로 판단하고 있으며[2], 웨어러블 디바이스 중에서도 특히 애플의 애플워치, 삼성의 갤럭시 기어, LG의 LG 어베인, 샤오미의 샤오미워치 등 아직 초기 단계인 스마트워치 시장의 선점을 위하여 많은 기업들이 경쟁하고 있다. 하지만 기업은 웨어러블 디바이스와 같은 자사제품을 개발하면서 소비자들의 반응을 예측하지만 실제 반응이 이와 일치하는지는 확신할 수 없다. 때문에 제품에 대한 소비자들의 반응이 기대와 일치하는지 확인할 수 있는 방법은 이들의 반응을 살펴보는 것이고 이는 웹을 통해서 활발히 이루어지고 있기 때문에 웹을 통해서 살펴볼 수 있다. 소비자들은 SNS(Social Network Service)를 통해서 제품에 대한 자신들의 의견을 공유하고 이를 통하여 다른 소비자들은 구매 전 그 제품에 대한 정보를 얻는다. 이러한 의견들에는 기업에서도 예상하지 못했던 소비자들의 반응과 제품에 대한 의견이 나와 있기 때문에 기업에서는 이를 이용하여 홍보 방향을 정한다든가 혹은 추후 제품에 소비자들의 의견을 반영하여 발전시킬 수 있다.

하지만 스마트폰의 대중화와 함께 SNS를 통한 소비자들의 제품에 대한 리뷰데이터는 대량으로 생산되고 있지만 방대한 데이터 속에서 원하는 정보를 찾는 작업은 점점 어려워지고 있다. 특히나 소비자들의 의견과 같은 정보는 데이터를 모은 후 그 의견이 어떠한 감성을 지니고 있는지 파악해야 하기 때문에 그 어려움이 더해졌다. 이러한 불편함 때문에 SNS로부터 생성되는 수많은 데이터의 정보를 빠르게 분석하고 필요한 정보를 추출해 내는 오피니언 마이닝(opinion mining) 기술이 더욱 중요해지고 있다.

이에 본 논문에서는 오피니언마이닝의 개념 및 기존 연구를 2장에서 알아본다. 이후 3장에서는 스마트워치에 대한 데이터를 어떻게 수집하고 분석하는 지에 대해서 살펴본다. 4장에서는 이를 통해 임의의 데이터를 적용하여 간단한 실험결과를 분석한다. 마지막으로 5장에서는 전반적인 요약과 함께 결론을 맺는다.

II. 연구 배경

2.1 오피니언 마이닝

텍스트 마이닝(Text mining)이 텍스트 정보에서

문맥을 파악하고 텍스트간 연계를 분석하는 비정형화된 문서에서 정보를 얻어내는 방법인 반면, 오피니언 마이닝은 텍스트 마이닝에서 발전된 분석기법으로, SNS 등에서 얻어낸 대량의 데이터의 여론 향방을 긍정(Positive), 부정(Negative), 중립(Neutral)으로 구분하여 선호도를 판별하는 기술이다. 이렇듯 오피니언 마이닝은 텍스트 마이닝에 근거를 둔 분석기법으로써 주로 SNS에 많이 쓰이는 이유는 감성분석을 활용하여 특정 텍스트의 어조와 감정을 파악하기 때문이다.[1]

일반적인 오피니언 마이닝은 (1) 얻고자 하는 대량의 텍스트 데이터를 얻고 (2) 얻은 텍스트 데이터에서 긍정 및 부정을 나타내는 단어 정보를 추출한다. (3)이후 긍정 및 부정 표현을 추출하여 사용자에게 보여주는 절차를 갖는다. 이러한 오피니언 마이닝을 분류할 경우 두 가지로 분류가 가능한데 첫 번째는 문서 단위 오피니언 마이닝이고 두 번째는 속성 기반 오피니언 마이닝 및 요약(Feature-based Opinion Mining and summarization)이다.[2] 감성적 분리라고 하는 방법은 문서 단위 오피니언 마이닝 연구로써 PMI-IR 기법을 이용하여 문서 전체를 대상으로 긍정 또는 부정으로 극성을 판단하는 이분법적인 기법이다. 문서 전체를 대상으로 긍정 또는 부정으로 분류하는 기법이다. 이는 특정한 문장의 패턴을 만족하는 구문들의 Semantic Orientation(SO)을 계산하고 SO의 총합이 양수일 경우 긍정으로, 음수일 경우 부정으로 분류하는 기법이다.[3] 하지만 이러한 결과만 가지고는 전반적인 의견의 극성이 총합의 극성과 같다고 판단하기는 어렵다. 전반적인 내용이 부정일 수는 있으나 대상에 대한 결과에 대해서는 긍정의 내용을 내포할 수 있기 때문이다. 속성 기반 오피니언 마이닝은 앞서 언급한 문서 단위 오피니언 마이닝과 같은 이분법적인 방법이 아닌 문장 전체에서 각 속성별로 세부적으로 추출하여 긍정, 부정, 중립으로 판별하는 방법이다.

III. 연구 방향

본 장에서는 오피니언 마이닝 기술을 활용하는 과정을 과정별로 설명하고자 한다. 이를 위해 가장 먼저 스마트워치의 분석 가능한 평가요소를 결정한다. 하나의 리뷰 안에서도 각 스마트워치에 대한 의견이 드러날 수 있고, 또는 하나의 스마트워치에 대한 속성별 긍정적·부정적 의견이 나타날 수 있기 때문이다. 기본적인 스마트워치의 속성에 대한 요인은 비용합리성, 기능성, 심미성, 착용성, 유용성과 같이 5가지로 정하였고 이에 대한 감성사전을 사전에 구축한다.[4]

3.1 데이터 대상

본 연구에서 수집하고자 하는 제품에 대한 리뷰 데이터는 SNS의 하나인 트위터를 통해서 수집하고자 한다. 트위터는 웹상에서 사용자가 제한

된 텍스트 용량 내에서 자신의 의견을 표출하기 때문에 자신의 의견 중 핵심적인 단어만을 사용한다. 이는 패턴 길이를 최대 3으로 하는 단순한 문장만을 대상으로 연구를 진행하는 본 논문의 취지와 부합하기에 트위터를 통해 진행한다. 단순한 문장은 속성명인 f 를 기준으로 ± 2 의 거리에 있는 단어(w)를 가지고 해당 속성에 대한 극성을 판단할 수 있는 문장을 의미한다. 그 예는 다음과 같다.[5]

1. 디자인(f)이 정말(w_1) 예뻐요(w_2).
2. 기능(f)이 다양(w_1)해서 좋아요(w_2).
3. 애플워치(f) 이정도면 만족(w_1)

이와 같은 단순한 문장은 다음과 같은 품사를 갖는다

1. 명사(디자인), 부사(정말), 형용사(예뻐요)
2. 명사(기능), 형용사(다양), 형용사(좋아요)
3. 명사(애플워치), 형용사(만족)

이렇듯 분류 대상의 문장이 이와 같은 패턴일 경우 문장은 속성에 대한 극성을 판단할 수 있는 감성어를 갖고 있다고 볼 수 있다. 이러한 이유로 트위터가 다양한 SNS 중 연구를 진행하기에 가장 적합한 SNS로 판단되어 트위터의 리뷰 데이터를 얻고자 한다.

3.2 데이터 수집방법

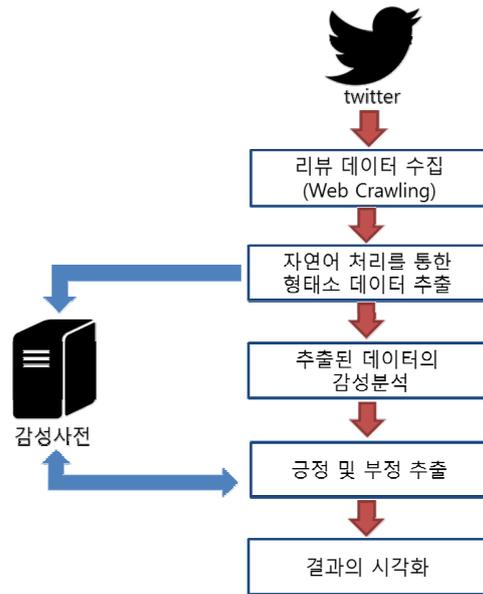
트위터 데이터를 얻기 위한 웹크롤링(Web crawling)은 트위터의 공개 API를 통해서 진행된다. 트위터의 스트리밍(Streaming) API에는 총 3가지의 수집 방식이 있는데 이 중 특정 주제와 관련하여 데이터를 수집하기에 가장 적합한 공개 스트림을 사용한다. 공개 스트림은 GET statuses/sample 방식과 POST statuses/filter 두 개의 방식으로 나뉜다. GET statuses/sample 방식은 공개 스트림 중 무작위로 선정된 샘플값만을 제공하는 방식이고, POST statuses/filter 방식은 하나 혹은 그 이상의 지정된 필터링에 의해서 샘플값을 제공하는 방식이다. 본 연구는 ‘스마트워치’ 라는 특정한 단어를 포함하는 리뷰 데이터를 수집해야 하기 때문에 POST statuses/filter 방식을 사용할 것이다.

3.3 의견 데이터 추출 및 극성 판단

본격적인 데이터 수집에 앞서 일상적으로 많이 사용하는 스마트워치에 대한 속성별 감성단어들을 정의하고 각 극성을 분류한다. SNS에는 은어나 비속어도 많이 쓰이기 때문에 주로 사용되는 은어나 비속어들도 고려하여 감성단어와 의미가 유사한 형태소들도 추출하여 포함한다. [그림 1]의 과정은 다음과 같다.

3.3.1 리뷰데이터 수집

파이썬을 이용하여 트위터 데이터를 얻기 위해 웹크롤링을 진행하는 단계이다. 애플워치, 갤럭시기어 등 각 스마트워치에 대한 의견 데이터를 추출해 낸다.



[그림 1] SNS 리뷰 데이터의 분석 과정

3.3.2 형태소 데이터 추출

추출된 트위터 데이터를 분석하기 위해 형태소별로 데이터를 추출하여 각 스마트워치별 속성을 추출해 낸다. 속성별로 극성을 나타내는 의견데이터 또한 상이하기 때문에 이 과정에서 속성을 추출한다.

3.3.3 추출된 데이터의 감성분석 및 긍정, 부정 추출

각 속성별로 나타난 의견 데이터를 통해 속성별 극성을 판단한다. 이때 기존에 감성사전에 등록되어 있지 않은 속성별 의견이라면 감성사전에 의견 새로운 의견을 추가하고 사전에 극성을 판단하여 감성사전에 이에 대한 극성을 추가한다.

3.3.4 결과의 시각화

위의 과정을 통해 얻어진 데이터를 통해 극성을 판단하고 이를 사용자가 쉽게 확인할 수 있도록 속성별 극성을 표와 같이 시각화한다.

IV. 실험 및 결과

트위터 데이터는 2015년 8월 3일에서 8월 24일 까지 수집된 데이터를 사용하였다. 기준이 된 스마트워치는 애플워치와 갤럭시기어 두 가지로 진행하였으며 수집된 각 데이터 중 의미있는 데이터는 애플워치 86건, 갤럭시기어 49건이 수집되었다. 이 중 디자인에 연관된 의견은 애플워치 42건, 갤럭시기어 17건으로 속성 중 가장 많이 수집되었으며, 그 다음으로는 기능성에 관련된 의견이 애플워치 21건, 갤럭시기어 11건이 수집되었다. 유용성에 대한 의견은 애플워치 13건, 갤럭시기어

8건으로 애플워치의 경우 애플사의 타 기기와의 연동에 관련된 의견이 많았다. 그 다음으로는 착용성에 관련된 의견이 많았는데 이는 애플워치 6건, 갤럭시기어 7건으로 나타났으며, 비용합리성에 관련된 의견은 광고를 제외하고는 거의 수집되지 않아 애플워치 4건, 갤럭시기어 6건으로 나타났다.

속성	의견수	극성	극성별 의견수
심미성	42	긍정	20
		부정	22
기능성	21	긍정	16
		부정	5
유용성	13	긍정	9
		부정	4
착용성	6	긍정	5
		부정	1
비용합리성	4	긍정	2
		부정	2
총계	86	-	86

[표 1] 애플워치 속성별 의견결과

속성	의견수	극성	극성별 의견수
심미성	17	긍정	11
		부정	6
기능성	11	긍정	8
		부정	3
유용성	8	긍정	5
		부정	3
착용성	5	긍정	4
		부정	1
비용합리성	7	긍정	2
		부정	5
총계	48	-	48

[표 2] 갤럭시기어 속성별 의견결과

각 속성별 극성을 살펴보면 애플워치의 경우 심미성의 극성은 긍정 20건, 부정 22건이었다. 기능성은 긍정 16건, 부정 5건이었고 유용성은 긍정 9건, 부정 4건이었다. 착용성은 긍정 5건, 부정 1건이었고, 비용합리성은 긍정 2건, 부정 2건이었다. 갤럭시기어의 경우 심미성의 긍정 극성은 11건, 부정 6건이었으며 기능성은 긍정 8건, 부정 3건이었다. 유용성은 긍정 5건, 부정 3건이었으며 착용성은 긍정 4건, 부정 1건이었다. 비용합리성은 긍정 2건, 부정 5건이었다.

한글을 대상으로 스마트워치의 의견을 수집해 서인지 많은 의견이 애플워치와 갤럭시기어의 비교를 나타내고 있었다. 특히 디자인과 유용성면에서 이 둘의 비교가 많았고 디자인에서는 갤럭시

기어에 만족한다는 의견이 더 많았으나 기능 및 유용성면에서 애플워치를 선호한다는 의견이 더 많았다.

V. 결 론

처음 스마트워치 속성별 극성 판단을 위한 정보 수집을 트위터로 선택한 것은 트위터 이용자수와 다양한 이용자를 통한 정보 파생에 대한 기대감 때문이었다. 하지만 트위터를 통한 스마트워치의 속성에 대한 의견 데이터와 속성 비교를 위한 정보의 양은 기대에 미치지 못했다. 트위터의 실제 스마트워치에 대한 정보는 스마트워치의 속성별 개인의견보다는 “애플워치 OS 업데이트 실시”, “갤럭시기어 애플워치 대항마?”와 같은 스마트워치 자체에 대한 광고 및 단순의견이 많았으며, 이 또한 극성을 판단할 수 있는 언어보다는 ‘구매했다’, ‘샀다’ 혹은 ‘발표’, ‘설치’와 같은 등의 동사 및 명사만으로 구성된 것들이 더 많았다.

추후 연구에서는 트위터를 통한 스마트워치 의견데이터 외에도 실제 각 스마트워치를 판매하고 있는 사이트의 리뷰데이터를 추가하여 연구를 진행한다면 더욱 의미있는 결과가 도출될 것으로 판단한다. 또한 지속적인 스마트워치에 대한 연구를 통해 스마트워치 구매에 영향을 미치는 속성을 연구하여 스마트워치 마케팅에 필요한 속성을 구분하고 이에 대한 극성 판단을 한다면 추후 스마트워치 발전방향에 정확성을 더할 것으로 기대한다.

참고문헌

- [1] 박대현, 송동현, “비정형 데이터 활성화의 정치, 경제, 문화적 함의”, KISA Report, 2014.02
- [2] Hu, M. and Liu, B. “Mining and summarizing customer reviews”, In proc. of the 10th ACM SIGKDD Conf., pp.168-177, 2004
- [3] P. Turney, "Thumbs Up or Thumbs Down Semantic Orientation Applied to Unsupervised Classification of Reviews," In Proceedings of the Meeting of the Association for Computational Linguistics, pp.417-424, 2002
- [4] 이정우, 최재현, 박제원, “개인과 기기특성이 스마트 워치 사용의도에 미치는 영향에 관한 실증적 연구, 한국정보기술학회, Vol 12. No. 11, PP.201-214
- [5] 강한훈, 유성준, 한동일, “k-Structure를 이용한 한국어 상품평 단어”, 정보과학회, 제 37권 제 6호, pp.470-479, 2010.06