

감성 분석 기반의 제품 평판 마이닝

송인환*, 한진주^o, 온병원*

프럼데이타웨어, 군산대학교, 군산대학교

songh@prumdataware.com, jinjuhan7@kunsan.ac.kr, bwon@kunsan.ac.kr

Product reputation mining based on sentiment analysis

In-Hwan Song, Jinju Han^o, Byung-Won On

Prumdataware, Kunsan University, Kunsan University

요 약

스마트폰 보급의 확산으로 제품 구매 시 웹 사이트 및 SNS를 이용하여 제품 리뷰를 참고하는 소비자들이 증가하고 있다. 전자 상거래 사이트의 제품 리뷰는 구매 예정자들에게 유용한 정보로 활용되고 있다. 하지만 구매 예정자가 직접 제품에 대한 리뷰 데이터를 찾아 전체 내용을 일일이 읽고 분석해야하기 때문에 시간이 오래 걸릴뿐만 아니라 가공되지 않은 데이터가 줄 수 있는 정보는 한정적이다. 또한 이러한 리뷰들은 상품의 특징을 파악하기에도 어려움이 있다. 본 논문에서는 제품의 주요 이슈를 추출하고 주요 이슈에 대한 감성 분석과 감성 요약을 통해 제품 분석 및 평가를 제공하는 시스템을 설계 및 구현하였다. 이를 휴대폰 제품에 적용하여 구축한 시스템을 통해 소비자가 방대한 양의 제품의 리뷰 데이터를 분석할 필요 없이 제품의 주요 이슈와 가공된 분석 결과를 시각적으로 빠르게 제공할 수 있음을 보였다.

주제어: 텍스트마이닝, 선호도 조사, 감성 분석, 토픽 추출

1. 서론

스마트폰 보급의 확산으로 제품에 대한 다양한 의견의 교류가 확대되고 있으며, 제품을 구매하기 직전 웹사이트 및 SNS를 통해 제품 리뷰를 이용하는 소비자들이 증가하고 있다. 소비자들은 제품을 구매하기 전에 구매에 확신을 갖고자 다양한 정보들을 탐색하여 제품 구매 시 내재된 위험을 줄이기 위해 앞선 구매자들의 구매 후기에 의존해 참고하여 구매 여부를 판단한다. 이러한 소비자들을 위해 기존에는 전자 상거래 사이트 및 웹 게시판을 통해 구매 예정자들에게 제품에 대한 정보 및 개인의 리뷰를 제공하였으며, 더 나아가 개인의 블로그 및 SNS를 통해 불특정 다수에게 리뷰를 공개하는 것으로 확대되었다. 하지만 기존의 방법들은 구매 예정자가 직접 제품에 대한 리뷰 데이터를 찾아서 전체 내용을 일일이 읽고 분석해야하기 때문에 시간이 오래 걸릴 뿐만 아니라 가공되지 않은 리뷰 데이터가 줄 수 있는 정보는 한정적이기 때문에 리뷰들을 통해 상품의 특징을 파악하기에는 어려움이 있다. 또한 상품에 대한 만족도를 리뷰로 표현하는 사용자들이 늘고 있어 리뷰의 개수는 수없이 많으며, 앞으로는 더더욱 증가할 것이다. 그러므로 리뷰를 일일이 읽으면서 상품에 대한 특징을 파악하고 구매 결정을 내리는 것은 점점 더 어려운 일이 될 것이다. 따라서 제품에 대한 리뷰 데이터를 수집하고 리뷰 문장의 감성 분석과 이를 가공하여 효과적으로 제공하는 시스템에 대한 고찰은 의미가 있다. 이와 관련하여 최근 제품 리뷰 데이터를 분석 하여 평가 결과를 제공하는 시스템들과 관련 연구가 활발하게 진행되고 있다. 하지만 기존의 리뷰 분석 시스템들은 제품에 대한 평점을 제공하는 수준이거나, 리뷰 문장 분석에 있어서 제품의 속성을 파악할 뿐 어떤 이야기들이 사람들에게 많이 언급되는지 이슈를 파악하는 것이 부족하다.

본 논문에서는 제품 평판 마이닝을 위한 웹 기반의 인터페이스를 설계하고, 제품에서 특정하게 많이 이야기되고 있는 이슈를 찾아 리뷰 문장 감성 분석 알고리즘과 주요 이슈의 감성 요약 알고리즘을 제안한다. 여기서 주요 이슈란 제품에서 주로 언급되고 있는 화제이다. 예로 주요 이슈에는 “디자인”, “기능”, “성능”을 말한다. 알고리즘은 리뷰의 주요 이슈의 감성을 평가하여 선호도를 조사한다. 선호도 조사는 긍정, 부정을 판단하여 대중의 제품에 대한 선호도를 주요 이슈 별로 쉽게 알 수 있다. 또한 주요 이슈에서 평가된 선호도의 근거를 뒷받침해주기 위해 감성 요약을 진행한다. 감성 요약은 실제 리뷰에서 주요 이슈와 관련된 내용을 보여줌으로써 대중들의 제품에 대한 선호도의 이유를 알 수 있다. 또한 설계한 내용을 바탕으로 휴대폰 제품에 대한 리뷰를 적용하여 리뷰 분석을 웹으로 제공하는 시스템을 구현하였다.

2. 관련 연구

소비자들은 제품을 구매하기 전에 구매에 확신을 갖고자 다양한 정보들을 탐색하여 제품 구매 시 내재된 위험을 줄이기 위해 앞선 구매자들의 구매 후기에 의존해 참고하여 구매 여부를 판단하는 것은 보편화 되어있다.[1] 그러나 소비자가 기존 구매자의 리뷰를 모두 읽고 제품에 대한 정보를 파악하기에는 어려움이 있다. 게다가 각 제품에 대해 부여된 평점이 제품의 어떤 특징에 대해 부여된 것인지 알 수 없다는 문제점이 있다.

이와 관련하여 제품 리뷰 데이터를 분석하고 평가 결과를 제공하는 시스템[2-3] 및 관련 연구[4]가 활발하게 진행되고 있다. 이러한 연구들은 리뷰에 대한 평점으로 긍정과 부정을 평가하는 방법을 제안하는데 그쳐, 본 논

문에서 제안하는 것처럼 제품의 특정 주제나 화제에 대한 긍정과 부정을 평가하고 선호도의 평가의 증거를 제시하는 연구는 없다.

Wagh et al.[5] 연구에서는 트위터 데이터를 NLTK (Natural Language Toolkit)를 분석하여 기계 학습 기법을 통한 트위터 사용자의 감성을 자동으로 예측하는 연구이나 특정 도메인에서 학습 한 분류 기준을 다른 도메인에 직접 적용하면 도메인 간의 차이로 인해 결과가 좋지 않다. Hur et al.[6]의 연구는 하둡을 이용하여 트위터 데이터와 같은 빅데이터를 저장 및 분석하도록 하는 LiveAD라는 시스템을 설계 및 구축하여 트위터를 대상으로 TV광고 분석을 제안한다. 이 연구는 리뷰 데이터를 이용하지만 제품의 과급력만을 분석할 뿐, 구매 예정자에게 가공된 정보를 제공하기에는 부족한 방법이다.

Kim et al.[7] 연구에서는 R을 이용한 텍스트 마이닝 기법을 사용하여 리뷰에서 중요한 키워드를 추출하고 목적에 맞는 맞춤형 리뷰 정보를 도출시키는 방안을 제안한다. Tribhuvan et al.[8]의 연구는 선택한 제품에 대한 관련된 리뷰를 요약하여 특징별로 평가된 점수로 표현하는 제품 리뷰 요약 시스템을 제안한다. 하지만 이들의 리뷰 분석 시스템들은 해당 사이트에 속하는 제품에 대해서 전체 평점을 제공하는 수준이다. 즉, 리뷰 분석에 있어서 제품의 화제나 주제가 무엇인지 분석하는 것이 부족하고 소비자가 중요하게 생각하는 제품의 주요 이슈(디자인, 가격, 기능, 성능)에 대한 분석들이 필요하다.

한편, 'LDA를 사용하여 두 개의 경쟁 제품에 대한 온라인 리뷰의 주제 분석'에 대한 연구[9]에서는 텍스트 마이닝 접근 방식인 LDA(Latent Dirichlet Allocation)를 통해 두 가지 특정 경쟁 제품에 대한 온라인 리뷰의 주요 주제를 추출하여 선호도를 판별하는 방법을 제안한다. 하지만 제품의 주요 주제별로 평점을 제공할 뿐 주제별로 제품의 평점이 어떻게 측정되었는지 뒷받침 해줄 수 있는 근거는 부족하다.

본 연구에서는 제품의 리뷰에서 주요 이슈를 추출하고 선호도 조사를 통해 주요 이슈 별로 선호도 평가와 감성 분석을 통한 감성 요약을 제공하는 인터페이스를 제안한다.

3. 제품 평판 마이닝 시스템

그림 1은 리뷰를 수집하여 제품에서 많은 사람들에게 언급되고 있는 이슈를 추출하고 주요 이슈 별로 감성평가와 감성요약을 실시하여 분석 결과를 웹 인터페이스를 통해 제품의 리뷰에 대한 가공된 정보를 시각적으로 제공하는 시스템이다.

시스템은 데이터 입력, 제품 평가 분석, 웹 인터페이스로 구성되며 데이터 입력은 크롤러를 통해 웹상의 제품에 대한 사용자 리뷰 데이터를 수집한다. 사용자 리뷰 데이터는 휴대폰 리뷰 데이터를 사용하였다. 제품 평가 분석 단계에서는 사용자 리뷰 데이터에서 토픽 모델을 적용하여 주요 이슈를 추출한다.

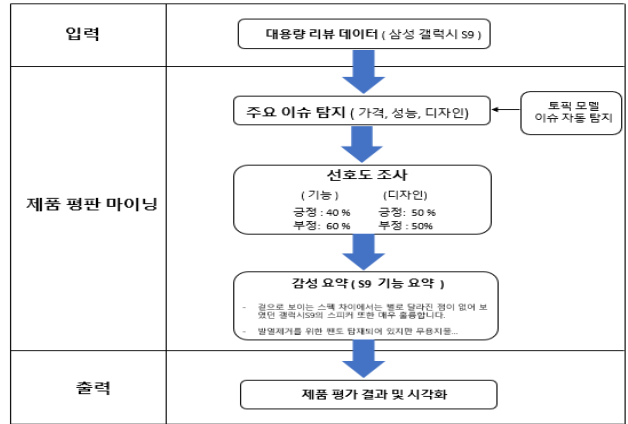


그림 1 제품 평판 마이닝 시스템

여기서 휴대폰 제품 리뷰 데이터의 주요 이슈는 '디자인', '기능', '가격', '성능' 이 있으며 이슈는 어떤 주제나 화제가 표현되는지 알 수 있다. 추출된 이슈는 구축된 감성 사전을 이용하여 알고리즘에 따라 선호도 조사를 진행하게 된다. 선호도 조사는 긍정, 부정의 비율을 평가하고 추출된 이슈를 평가하는 기준이 된다. 다음은 선호도 조사의 근거를 뒷받침해줄 수 있는 감성 요약을 진행한다. 감성 요약은 주요 이슈의 핵심 문장들을 추려 감성 평가를 한 뒤 긍정인 문장 Top-3와 부정인 문장 Top-3를 추출한다. 휴대폰 제품의 주요 이슈 중 '디자인' 에서 60%의 선호도 평가가 되었다면 디자인 측면에서 선호의 이유를 뒷받침 해주는 실제 리뷰 요약을 제공한다. 즉 주요 이슈에서 긍정 내용과 부정 내용을 요약하여 보여줌으로써 대중들이 주요 이슈에서 무엇을 긍정으로 보고 있는지, 무엇을 부정적으로 여기는지를 일목요연하게 알 수 있다. 마지막으로 제품에 대한 리뷰 데이터 분석 결과를 웹 인터페이스를 통해 사용자에게 시각적으로 제공한다.

3.1 데이터 입력

사용자 리뷰 수집을 위해 휴대폰 리뷰 사이트인 'gsmarena' [10]에서 '삼성', '애플', 'LG' 15종의 휴대폰 영문 리뷰 데이터 19,302건을 수집하였다.

데이터 기반의 제품 평판 마이닝 시스템을 사용하고자 하는 사용자는 특정 제품에 대한 정보만을 원할 수 있다. 따라서 수집된 리뷰 데이터에서 특정 제품의 정보만을 보여주는 방안을 설계하였다. 수집된 19,302건의 리뷰 데이터에서 사용자에게 키워드를 입력받아 관련 키워드의 리뷰 데이터만을 추출한다.

예를 들어 사용자는 삼성의 갤럭시 S9와 애플의 iPhone X의 분석 정보를 원한다면 사용자는 갤럭시 S9 또는 iPhone X를 검색하여 해당 리뷰만을 검색하여 리뷰 분석을 받을 수 있다.

3.2 제품 평판 마이닝

리뷰 데이터는 의견을 표현하는 글이다. 의견을 표현하는 글은 문어적이기보다는 구어적이며, 완벽하게 규범

지향적이기보다는 자유분방하고 탈규범적인 경향이 있다. 또한 어떤 상황에 관한 느낌을 표현할 때 쓰이는 전형적인 어휘와 구문들이 있다. ‘좋다’, ‘싫다’, ‘나쁘다’, ‘그저 그렇다’, ‘지저분하다’, ‘깔끔하다’, ‘재미있다’, ‘흥미롭다’, ‘싫지만은 않다’ 등의 예에서 볼 수 있듯이 자주 사용하는 표현은 표현 자체에 어떤 주제를 함축하고 있기도 하다. 또한 구문이나 패턴 중에는 기능적 범주로, 어떤 대상에 대한 느낌을 다른 대상과 비교하거나 강조하는 구문도 있다. 따라서 특정 대상에 대한 감성 분석에서 여러 제품을 동시에 비교하거나, 극성의 정도성을 판별할 때 중요한 단서를 제공한다. 나아가 최근 감성 분석에서는 ‘좋다’나 ‘나쁘다’와 같은 단순한 극성 범주 결정보다는 어떤 측면에서 좋고 나쁜지, ‘어떤 주제나 화제’가 표현되는지가 중요하다.

3.2.1 대용량의 리뷰 문서에서 주로 언급되는 이슈 추출

제안하는 제품 평판 마이닝 시스템에서는 리뷰 데이터의 주요 이슈를 추출하여 어떤 측면에서 좋고 나쁜지 제품에 대해 대중들이 자주 언급하고 있는 것들이 무엇인지를 찾기 위해 토픽 모델인 LDA(Latent Dirichlet Allocation)[11]를 사용하였다. 이후, 추출된 토픽 내에 있는 연관어들의 토픽일관성(Topic coherence)를 측정하여 토픽일관성이 높은 것을 주요 이슈라고 명명하였다.

3.2.2 이슈에 대한 선호도 조사

사용자에게 주요 이슈들의 선호도 조사 결과를 제공한다. 선호도 조사는 주요 이슈를 긍정, 부정으로 판단하여 대중의 제품에 대한 선호도를 주요 이슈 별로 쉽게 알 수 있다. 예를 들어 삼성 갤럭시 S9의 주요 이슈 ‘기능’, ‘디자인’, ‘가격’, ‘성능’의 선호도를 판별하여 선호도 비율을 평가 할 수 있다.

리뷰 데이터의 주요 이슈들의 선호도를 조사하기 위한 알고리즘으로 알고리즘의 기본 실행 단위는 주요 이슈들의 연관어들이다. 감성사전은 NLTK(Natural Language Toolkit) 패키지 중 NLTK 3.3 버전[12]을 사용하였다.

Algorithm 1. Measure Aspect preferences

Input: a Review Aspect Word

Output: Positive Score S_p , Negative Score S_n , pos, neg

```

1: for  $w \in T = \{w_1, w_2, \dots, w_n\}$ 
2:   if  $w \in D_p$ 
3:      $S_p = P(w) * degree(w)$ 
4:   else if  $w \in D_n$ 
5:      $S_n = P(w) * degree(w)$ 
6:    $pos = \frac{S_p}{S_p + S_n}$ 
7:    $neg = \frac{S_n}{S_p + S_n}$ 

```

Algorithm 1에서 입력된 이슈 연관어들에 대한 알고리즘의 수행 결과는 이슈의 긍정 언급도와 부정 언급도 및 전체 선호도 점수이며 값의 초기 값은 0이다.

먼저, 긍정 언급도(pos)의 경우 이슈의 연관어가 긍정인 단어(D_p)가 감성 사전에 포함되어있는지를 판단하게 된다(line 2). 이때 이슈에 긍정인 단어(D_p)가 존재하는 경우 긍정 언급도(S_p)에 이슈의 연관어가 나올 확률 분포 값($P(w)$)과 사전의 감성강도인 $degree(w)$ 를 곱한다. 마찬가지로 부정인 단어(D_n)가 존재할 경우 부정 언급도(S_n)에 확률분포 값과 $degree(w)$ 를 곱한다. 이후, 긍정 언급도와 부정 언급도의 수식을 계산하여 전체 선호도 점수를 평가한다(line 6-7).

3.2.3 이슈에 대한 긍정 및 부정 리뷰 요약

감성 요약은 앞선 선호도 평가를 뒷받침 해주는 증거가 된다. 주요 이슈에 대한 선호도 평가의 이유를 알 수 있는 과정이다. 감성 요약은 NLTK(Natural Language Toolkit)의 Sentiment Analysis[13]를 사용하였다. Sentiment Analysis는 텍스트의 긍정에 대한 점수 [1, 5], 부정에 대한 점수 [-1, -5]로 분류되며, 감성 요약은 리뷰 문장에서 주요 이슈 연관어의 확률분포 값을 연산한다. 연산된 결과 중 상위 15개의 문장을 추출하게 되는데 이슈와 관련된 문장을 매칭 시켜주기 위함이다.

Algorithm 2. Sentiment summary

Input: a Review Sentence

Output: Score Sentence

```

1: for  $s \in S = \{s_1, s_2, \dots, s_n\}$ 
2:   for  $w_i \in S$ 
3:     for  $w_j \in T = \{w_1, w_2, \dots, w_n\}$ 
4:       if  $w_i == w_j$ 
5:         Weight +=  $p(w)$ 
6:          $score_s = \frac{Weight}{s}$ 
7:   Sort(score_s)

```

Algorithm 2는 주요 이슈 감성 요약을 위해 주요 이슈의 연관어와 관련된 문장을 찾는 알고리즘이다. 실제 리뷰 데이터에서 주요 이슈의 연관어와 관련 있는 문장을 매칭 한다. 리뷰 데이터의 문장을 s 라고 지칭하고 주요 이슈의 연관어를 w_j 에 저장한다(line 3). s 에 포함된 단어 w_i 가 주요 이슈의 연관어 w_j 와 일치하면(line 4) w_j 의 확률분포($P(w)$) 값을 더한다(line 5). 이후 모두 더한 확률분포 스코어를 출현한 연관어의 수로 나눠 준 뒤 정렬하여 주요 이슈와 관련된 문장을 상위 15개로 지정한다.

매칭 된 문장들은 Sentiment Analysis 패키지를 이용하여 스코어 점수를 출력한다. 감정 점수는 가장 높은 점수의 긍정적인 문장을 기반으로 지정되며, 문장 내에서 가장 높은 점수를 받은 단어에 기초하여 문장에 대한



그림 2 제품 평판 마이닝 웹 시스템

양수 점수가 부여 된다. 예로 ‘I love you but hate the current political climate’ 라는 문장이 있다면 Sentiment Analysis의 분석 결과는 ‘I love[3] you but hate[-4] the current political’ 로 -1 (sentence : 3, -4) 이라는 스코어 값이 나오게 된다. 추출된 문장은 NLTK의 Sentiment Analysis의 입력으로 들어가게 된다. 따라서 입력 부분은 앞서 주요 이슈의 연관어와 매칭된 문장이며 출력 부분은 문장들의 감성 스코어 값이다.

따라서 출력된 문장의 스코어 값으로 정렬을 하게 된다. 정렬된 문장들은 주요 이슈에 가장 관련이 있는 문장들이라 판단하여 정렬된 문장 Top-3를 추출하여 각각의 이슈의 긍정인 문장 3개, 부정인 문장 3개를 출력하였다.

3.3 자동화 및 시각화를 위한 웹 기반의 인터페이스 개발

본 논문에서 제안하는 데이터 기반의 제품 평가 시스템을 휴대폰 제품에 적용하여 웹 인터페이스를 구현하였다. 본 시스템은 수집된 19,302건의 리뷰 데이터를 저장하고 주요 이슈 추출과 이슈의 선호도 조사, 이슈 감성 요약물을 시각화하여 해당 기능을 웹으로 제공한다. 사용자는 간단한 방법으로 대용량의 리뷰 데이터에서 제품의 분석 결과를 신속하게 제공 받을 수 있다.

4. 제품 평판 마이닝 시스템 웹 인터페이스

웹 인터페이스의 구성은 크게 리뷰 데이터 필터링, 주요 이슈 추출, 선호도 조사, 감성 요약으로 이루어져 있으며, 서버에서 사용자의 요청을 받아 해당 기능을 실행하고 시각적으로 보여주었다.

대용량의 사용자 리뷰 문서에서 주로 언급되는 이슈 추출은 우선 데이터 필터링 과정을 진행하게 된다. 데이터 필터링 과정은 Java Script 언어로 클라이언트에서 사용자가 입력한 단어를 서버에서 요청 받아 모든 텍스트에 읽어 입력 단어가 포함된 문서만을 하나의 텍스트로 합병하게 된다.

합병된 사용자 리뷰 문서들을 LDA를 통해 토픽을 추출한다. 추출된 토픽은 연관어들의 토픽일관성(Topic coherence)를 측정하여 토픽일관성이 높은 것을 자동 토픽 레이블링 과정을 거쳐 주요 이슈로 제공한다. 제품 평판 마이닝을 위한 웹 인터페이스에서는 “이슈 탐지” 버튼을 클릭하면 해당 과정이 자동으로 진행된다. 주요 이슈는 히트맵으로 표현하여 히트맵에서 가장 큰 면적이 가장 중요하게 생각하는 이슈가 된다. 이슈는 5개를 기본 값으로 지정하였고 사용자는 최대 8개의 이슈를 볼 수 있다. 그림 2는 주요 이슈를 히트맵으로 시각화한 것이다. 해당 그림에서는 “big” 이라는 이슈가 주요 이슈로 도출되었으며 “batteries” 라는 이슈가 두 번째 이슈로 표현되었다. 5개의 이슈가 기본 값으로 출현되었으며 사용자의 편의에 따라 8개까지 주요 이슈를 볼 수 있다.

주요 이슈에 대한 선호도 조사는 앞서 설명한 이슈 선호도 조사 알고리즘을 통해 이슈에 대한 선호도를 평가한다. 히트맵으로 표현된 이슈에 대한 히트맵 여역을 클릭하면 선호도 결과를 시각화하여 제공한다. 주요 내용은 이슈에 포함된 연관어들의 워드 클라우드를 제공하고 이슈에 대한 긍정과 부정에 대한 비율을 제공한다.

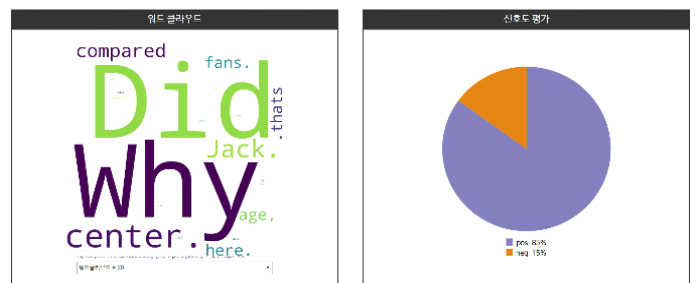


그림 3 주요 이슈 선호도 조사

그림 3는 이슈에 대한 선호도 조사 결과이며 워드 클라우드는 기본 값 20으로 표현되며 최대 50개의 연관어 워드 클라우드를 볼 수 있으며 선호도 조사는 도표로 긍정, 부정 비율을 표시한다. 예제의 도표는 긍정이 85%이며 부정이 15%로 대부분 해당 이슈를 긍정으로 생각하고

있다는 것을 뜻한다.

감성 요약에서는 주요 이슈의 선호도에 대한 증거를 제시할 수 있다. 감성 요약은 주요 이슈의 리뷰 데이터에서 가장 관련이 높은 문장 30개를 추출한다. 이후 NLTK의 Sentiment Analysis 패키지를 사용하여 감성 점수를 척도로 정렬을 한다. 가장 값이 높은 3개의 문장을 긍정적인 문장으로 값이 가장 낮은 3개의 문장을 부정적인 문장으로 측정하였다. 그림 4는 주요 이슈에 대한 긍정 및 부정 요약이며 이슈에 관련된 긍정 리뷰 문장 3개 부정 리뷰 문장 3개를 출력하여 사용자에게 제공한다. 선호도요약(긍정)에서 상위 리뷰 문장이 가장 긍정적인 문장이며, 선호도요약(부정)에서 상위 리뷰 문장이 가장 부정적인 문장이다.

선호도요약(긍정)	선호도요약(부정)
based personal use iPhone X citing links/links We dumb HahaHn" "name 04 Jun 2018i think consider samsung galaxy s9+ wait samsung galaxy note 9 comes ... moreiPhone X dont stand chance android manufacturers copying design big chin bottom slower processor LolHn"	"I noticed something so-called bezeless phones v30 s9 s8 iPhone x etc The side bezels wider Samsung S3 I hate side bezels They like Sony XA2 thicker 5 years old phonesHn" "best phone note 8 s9+Hn" "Anonymous 20 Apr 2018iPhone x s9 plus production cost

그림 4 주요 이슈에 대한 긍정 및 부정 요약

5. 결론

본 논문에서는 제품 사용자 리뷰 문서를 수집하여 추출한 데이터로부터 주요 이슈를 추출하여 선호도 판별과 감성요약을 통한 제품을 분석 및 평가를 시각적으로 제공하는 인터페이스를 설계하였다. 제안한 시스템은 이슈 추출, 선호도 조사, 감성 요약 및 사용자 인터페이스를 제공하며 이슈 추출은 리뷰 데이터에서 어떤 측면에서 좋고 나쁜지, 어떤 주제나 화제가 표현되는지를 파악하고 선호도 평가는 추출된 이슈의 긍정이나 부정을 측정하고 선호도의 비율을 분석하며 감성 요약은 주요 이슈에 관련된 긍정적인 문장과 부정적인 문장을 추출하여 선호도 조사에서의 선호도 비율에 증거를 제시하여 제품의 분석 및 평가를 제공한다.

또한 본 논문에서는 제안하는 시스템을 휴대폰 상품에 적용하여 웹 인터페이스로 구현하였고, 사용자에게 분석 결과를 시각적으로 표현 하였다. 구현된 시스템을 사용하는 소비자들은 방대한 양의 리뷰 데이터를 일일이 읽을 필요 없이 제품의 분석 결과를 빠르게 제공받을 수 있다. 구현된 시스템을 이용한 사용자들의 설문 분석한 결과, 구매하고자 하는 제품의 리뷰를 다른 평가 척도보다 중요시 하며 제품평가 시스템의 분석 결과들을 시각적으로 제공함으로써 구매 의사 결정에 도움이 된다고 평가하였다.

참고문헌

[1] Ho Geun Lee, Hyun Kwak, "Investigation of Factors Affecting the Effects of Online

ConsumerReviews," Journal of National Information Society Agency, Vol. 20, No. 3, pp. 3-17, 2013.

[2] Extreme Movie [Online]. Available: <http://www.extmovie.maxmovie.com>

[3] Skyscanner [Online]. Available: <http://www.skyscanner.co.kr>

[4] Areum Lee, Jiseon Bang, Yoonhee Kim, "A Design of TV Advertisement Effectiveness Analysis System Using SNS Big-data," Journal of KIISE Transactions on Computing Practices, Vol. 21, No. 9, pp. 579- 586, 2017.

[5] Bhagyashri Wagh, J. V. Shinde, "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques," Journal of Emerging Research in Management and Technology, Vol. 6, No. 12, pp. 49-57, 2017.

[6] Seoyeon Hur, Yoonhee Kim, "A Design of Analysis System on TV Advertising Effect of Social Networking Using Hadoop," Journal of Korean Society for Internet Information, Vol. 14, No. 6, pp. 49-57, 2013.

[7] JooYoung Kim, Dongsoo Kim, "A Study on the Method for Extracting the Purpose-Specific Customized," Journal of Society for e-Business Studies, Vol. 21, No. 2, pp. 151-161, 2016.

[8] Padmapani Tribhuvan, "A Peer Review of Feature Based Opinion Mining and Summarization," Journal of Computer Science and Information Technologies, Vol. 5, pp. 247-250, 2017.

[9] Wenxin Wang, Wenqiang Dai, "Topic analysis of online reviews for two competitive products using latent Dirichlet allocation," Journal of Electronic Commerce Research and Applications, Vol. 29, No. 10, 2018.

[10] Phone Review [Online]. Available: <http://www.gsmarena.com/>

[11] Ggibblda [Online]. Available: <http://jgibblda.sourceforge.net/>

[12] NLTK [Online]. Available: <https://www.nltk.org/api/nltk.sentiment/>

[13] NLTK Sentiment Analysis Api [Online]. Available: <http://www.nltk.org/howto/sentiment>

[14] Sentiment Analysis Demo [Online]. Available: <https://text-processing.com/demo/sentiment/>