2018 한남대학교 빅데이터 경진대회 최종보고서

소비자 감정으로 기업을 감정하다

: 2014~2018년 twitter를 중심으로

김준호 · 김영진 · 임원기 · 김봉영 · 양원준

한남대학교 경제학과 · 통계학과

---

1. 연구배경

오늘날 SNS(Social Network Service) 등 소셜미디어를 활용한 빅데이터 분석이 더욱 활발히 진행되고 있는 가운데 특정 주제에 대한 소비자의 특성과 시장 트렌드를 파악하는 것까지 진행되고 있다. 특히 트위터에서 쏟아지는 정보는 사용자 수, 데이터 양, 증가 속도 등 모든 면에서 대표적인 빅데이터 중 하나로 자리매김 하였고, 그 가치가 큰 주목을 받고 있는 실정이다. 특히 기업에서는 제품 및 서비스에 대한 평가 혹은 불만사항 등에 대한 다양한 분석 통해 제품이나 서비스의 질을 향상시키고자 한다. 트위터 데이터를 분석하는 방법 중 하나인 오피니언 마이닝은 흔히 감성분석으로도 불려지는 것으로, 텍스트에 나타난 사람들의 의견이나 성향 같은 주관적인 데이터를 분석하는 자연어 처리 기술이라고 할 수 있다. 이러한 분석으로 나타난 감정추세를 파악하여 기업에 대한 소비자들의 여론을 파악할 수 있으며, 더 나아가 미래를 예측할 수 있을 것으로 판단된다. 즉 오피니언 마이닝은 기업의 신속한 의사결정을 보조하는 수단으로 쓰여질 수 있기 때문에, 미래에 발생 가능한 일들에 대한 정보를 제공할 필요가 있다. 이미 오피니언 마이닝을 활용하여 주가를 예측한다거나, 국가경제위기를 예측하는 다양한 시도들이 수행되고 있다.

2. 연구목적

본 연구에서는 2014년 1월부터 2018년 10월까지 현대차 공식계정(@HyundaiUSA)에 나타난 미국 현대차 소비자들의 의견을 통하여 미국에서의 현대차 브랜드 이미지를 파악해보고, 나아가 이미지를 지수화 하여 자동차 판매량의 예측가능성을 살펴보고자 한다. 또한 긍 · 부정 단어의 연관어 분석을 통해 현대차의 강점과 약점을 파악하여 개선방안을 제안하고자 한다.

3. 연구방법

3.1 데이터 수집

본 연구에서는 twitterscraper를 이용하여 2014년 1월부터 2018년 10월 까지 현대차 공식계정(to:hyundaiUSA)으로 보내온 트윗을 수집하였다. 현대차 공식 계정에 대한 반응을 분석한 이유는 좀더 적극적인 반응을 보이는 소비자들을 분석하기 위함이며, 그 결과 중복 데이터를 제외한 총 25297개의 트윗을 수집하였으며 연도별로 보면 2014년 5255개, 2015년 6063개, 2016년 5611개, 2017년 4712개, 2018년 3656개를 수집하였다.

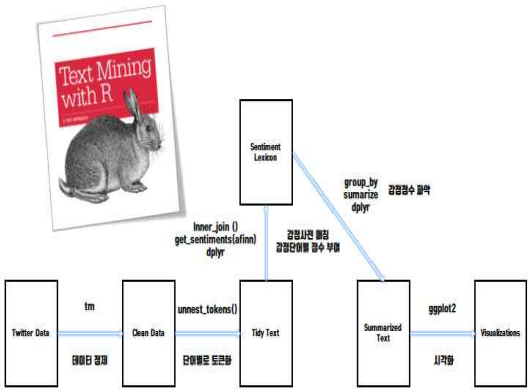
3.2 데이터 전처리

수집된 데이터를 전처리 하기 위해 R 프로그램을 사용하였으며, 텍스트 마이닝에 적합한 tm, tidyverse 패키지를 이용하였다. 우선 동영상 및 사진의 링크주소를 삭제하고 qdap, tm 패키지를 통하여 관사, 전치사, 접속사, 조사와 분석자의 주관적 판단 하에 의미가 없을 것으로 판단되는 단어인 불용어(stop words)를 삭제하고, 의미가 같은 단어 (예를들어 santa cruz -> santacruz, santa fe -> santafe) 는 하나의 통합된 단어로 변경하였다. 또한 영단어의 원형이 되는 어근(steaming)의 형태로 변경하고 tidyverse 패키지의 unnest_tokens 함수를 이용하여 토큰화(단어별로 나누기) 한 후 분석을 진행하였다.

3.3 감정분석

감정분석은 다음과 같이 분석할 수 있다.

Fig 1. 감정분석 흐름도



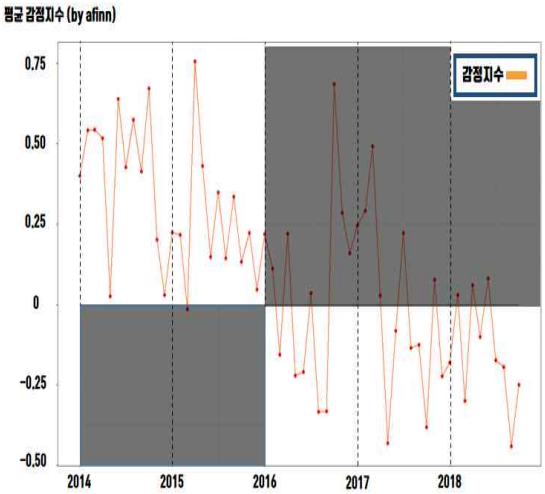
우선 원 데이터를 unnest_tokens 패키지를 이용하여 토큰화한 후 dplyr, qdap, tm 패키지를 이용하여 데이터를 원하는 날짜를 필터링하고 데이터를 정제한다. 그 다음 깔끔하게 정제된 데이터와 사용하고자 하는 감정사전(nrc,bing,afinn 방법 등이 있으며 본 연구에서는 afinn 방법을 이용함)과 매치하여 감정에 해당하는 단어에 -5점에서 5점 사이의 점수를 부여하여 계산하였다.

3. 연구결과

3.1 월별 감정분석

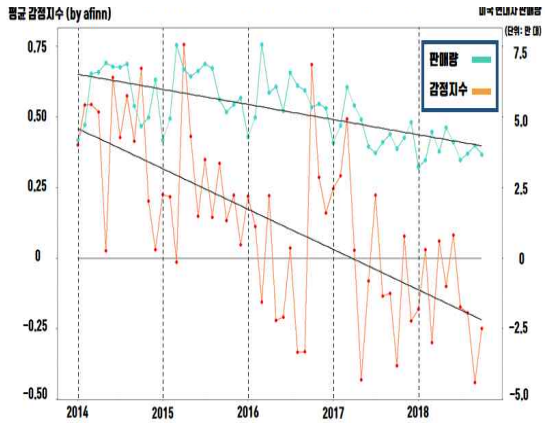
최근 5년간 현대차에 대한 소비자들의 인식변화를 살펴보고자 한다. 여기서 감정지수 그래프는

Fig 2. 월별 현대차 평균 감정지수



감정단어의 개수가 아닌, 긍·부정 감정단어에 +5점에서 -5점까지 점수를 메긴 후 월별로 평균을 내어 시계열로 표현한 그래프이다. 2014~15년까지의 월 감정지수는 긍정(+)적인 반면 2016년 이후로는 부정(-)적인 감정이 많이 나타남을 알 수 있다. 또한 2016년을 기점으로 이러한 감정의 반전이 나타난 것으로 보아 2016년 현대차에 대한 부정적 여론이 형성되었음을 파악할 수

Fig 3. 월별 현대차 판매량과 평균 감정지수

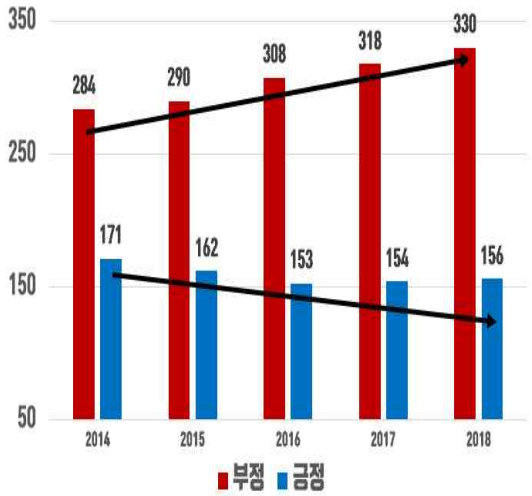


있다. 또한 현대차 감정지수와 판매량의 비교 그래프를 보면 2016년도를 기점으로 감정지수의 부정(-) 전환과 함께 판매량이 감소하고, 월별 판매량과 월별 감정지수가 동시에 하락하는 패턴을 알 수 있다. 이러한 결과들을 종합해 볼 때 기업에 대한 소비자의 감정이 판매량에 영향을 미칠 수 있을 것이라는 가능성을 제시할 수 있다.

3.2 긍 · 부정 빈도분석

연도별 긍정단어와 부정단어를 통해 소비자들이 현대차에 대해 느끼는 감정이 어떠한지를 파악할 수 있다. 우선 긍정단어와 부정단어가 나타나는 빈도수를 보면 부정단어는 시간이 지남에 따라 지속적으로 증가하는 반면 긍정단어는 일정수준을 유지하는 것을 볼 수 있다.

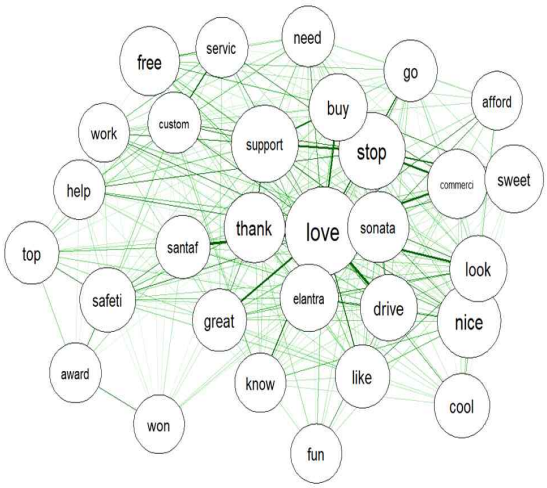
Fig. 4 연도별 현대차 긍 · 부정 단어 빈도수



3.3 연관어 분석

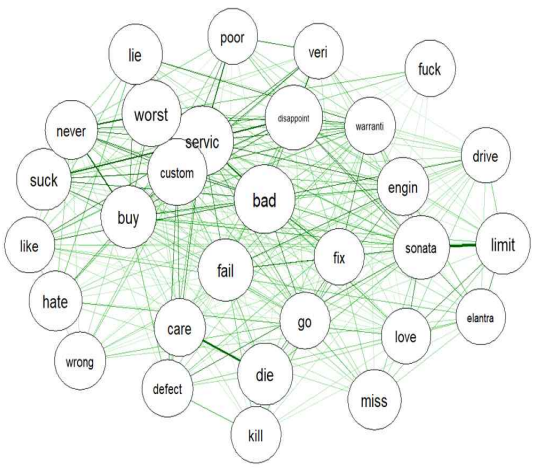
2014~2018년도의 긍정적인 감정을 나타내는 단어들과 연관된 단어를 분석한 결과 중심긍정단어(love)의 감정은 구매(buy)와 연관성이 높게 나타나며 관련된 차종으로는 sonata와 elantre, santafe로 나타난다. 또한 외관, 즉 자동차 디자인에 대한 긍정적 반응(look, nice, cool)이 주로 나타남을 알 수 있다. 따라서 현대차는 디자인이 소비자들에게 있어 매력적으로 인식됨을 알 수 있다.

Fig. 7 2014~2018년 긍정 단어



한편 부정적인 감정을 나타내는 단어들과 연관된 단어들을 분석한 결과 중심부정단어(bad)와 service가 자주 나타남을 볼 수 있으며, service와 custom, worst, lie, fail, never ,buy, suck 등이 있으며, 또한 engine과 관련하여 sonata, fix, limit 등이 나타나는 것으로 보아 소비자 서비스 및 엔진결함에 대한 불만이 많이 나타나는 것을 알 수 있다. 이러한 부정적 영향이 크게 나타나고 있으며 나아가 판매량에 악영향을 끼치는 것으로 분석된다.

Fig. 6 2014~2018년 부정 단어



4. 결론 및 한계점

본 연구는 SNS 소셜 빅데이터를 이용하여 미국에서 현대차 소비자들의 브랜드 인식을 살펴보고 나아가 이를 지수화 하여 자동차 판매량의 예측 가능성을 살펴보고자 한다. 또한 긍·부정 단어의 연관어 분석을 통해 현대차의 강점과 약점을 파악하여 성공적인 제품 및 브랜드 전략을 구축하는데 기초자료를 제공하고자 하였다. 이를 위하여 현대차 공식계정(@HyundaiUSA)에 나타난 2014~2018년 트윗 25,297개를 수집하였다. 비정형데이터 분석을 위하여 수집한 데이터를 분석하기 용이하게 자료를 처리하였으며, 분석 결과 첫째, 현대차에 대한 소비자 인식은 시간이 지날수록 하락하는 경향을 보였으며, 현대차 판매량 도한 같은 추세를 보이는 것을 볼 수 있었다. 둘째, 연관어 분석결과 미국 현대차 소비자들은 주로 디자인과 관련하여 긍정적인 반응을 나타내었고 반면 고객 서비스와 엔진 부문에 있어서 부정적인 반응을 나타내었다. 따라서 현대차는 긍정적인 반응이 나타난 디자인을 더욱 강조하고 부정적인 반응이 나타난 고객 서비스와 엔진 결함 부문에 대한 품질 향상 노력이 필요하다. 다만 본 연구에서는 세부적인 긍정요소와 부정요소를 나타내지 못하였다. 또한 감정지수를 계산하는 데 있어서 단어별로 점수를 부여하였으나, 실제로 사람들이 주고받는 의견에는 문맥에 따른 의미 변화가 많다. 따라서 이후 연구에서는 문장별로 의미를 파악하고 감정점수를 부여하는 방법이 필요하다. 또한 이러한 과정을 발전시켜 실시간으로 반응을 파악할 수 있는 시스템을 갖추고 나아가 현대차 주가 또한 예측할 수 있을 것이라는 가능성을 제시하고자 하였다.

5. 참고자료

taspinar - twitterscraper

Text Mining with R : a tidy approach Julia Silge, David Robinson

xwMOOC 텍스트 처리 - software carpentry

R을 이용한 텍스트 마이닝 - 김진성

R을 홀용한 텍스트 마이닝 입문 - 서진수

SNS 소셜 빅데이터를 통한 아웃도어 의류 소비자 특성과 주요 아웃도어 의류 브랜드 현황분석 - 정혜정, 오경희

소셜 빅데이터 분석을 통한 소비자 가치 인식 연규 - 신규 스마트폰을 중심으로