마음을 읽는 기술, 여론 마이닝

김원배 정보통신용어표준화위원회(WORDSTD) 위원, 전자신문 부장

여론 마이닝이란 웹사이트 혹은 소셜네트워크 서비스(SNS) 등에 나타난 여론과 의견 등 데이터를 분석, 유용한 정보로 재가공하고 활용하는 기술이다. 즉, 텍스트를 분석해 네티즌의 감성과 의견을 통계·수치화해 객관적 정보로 바꾸는 기술이다.

여론 마이닝은 오픈마켓이나 소셜커머스 같은 전자상거래 시스템과 더불어 발전했다. 상품을 눈으로 보거나 손으로 만져볼 수 없는 전자상거래의 특성상 상품을 앞서 구입한 소비자의 구매 후기가 미치는 영향이 커짐에 따라 여론 마이닝에 대한 연구가 시작됐다.

소비자는 수많은 구매 후기 중 실제로 상품을 구입·사용한 소비자의 후기가 긍정적인지 또는 부정적인지를 구별할 필요가 있다. 반면 판매자는 방대한 양의 구매 후기를 통해 상품에 대한 소비자 반응을 아는 게 필요하다. 여론 마이닝 기법은 소비자와 판매자 모두가 자연어 처리·텍스트 마이닝 같은 기법을 활용해 대량의 데이터에서 유용한 정보를 찾아내는 방법을 찾아낼 필요성을 충족하려는 과정에서 탄생했다.

여론 마이닝은 감성 분석(sentiment analysis)의 한 범주로, 특정 대상에 대한 생각이나 감성이 포함된 데이터를 유의미한 정보로 가공한다. 통상적으로 '감성 분석'을 '여론 마이닝'과 거의 동일한 의미로 사용한다. 다만 약간의 차이가 있는데, 감성 분석은 텍스트 데이터 뿐만 아니라 음성(voice) 데이터에서 발견되는 톤, 템포 등의 정보를 분석하는 것을 포함한다. 감성 분석을 활용하면 분석 가능한 데이터 범위가 넓어질 수 있다. 이에 비해 여론 마이닝은 텍스트, 이미지, 영상 등으로 공유되는 비정형 데이터를 주로 분석한다. 긍정적인지 부정적인지를 분석하고, 그 원인을 도출하는 것을 목적으로 한다.

여론 마이닝은 트렌드 파악뿐만 아니라 제품과 서비스 평가, 미래 예측 등 다양한 분야에서 활용될 수 있다. 정치·경제·사회 특정 사안이 발생했을 때에는 여론이나 대중의 관심도가 실시간으로 어떻게 변화하는 지 파악할 수 있다.

인터넷 댓글, 인터넷 카페와 블로그에 작성된 글 등 텍스트 형태의 비정형 데이터는 단어에 따라 의미가 달라지는 경우가 많다. 또, 상대적 표현이나 개인의 감정이 드러나는 등 정제되지 않은 표현이 많아 분석이 쉽지 않다. 이를 객관적으로 분석하기 위해서는 어휘를 수치화하고, 통계 작업을 거쳐야 한다. 긍정과 부정 어휘를 극성화해 긍정과 관련된 단어를 +1, 부정과 관련된 단어를 -1로 각각 수치화해 긍정도를 계산하는 방법 등이다.

여론 마이닝의 일반적 절차는 다음과 같다. 우선, 분석하고자 하는 텍스트를 수집한다. 대량의 텍스트가 확보되면 여론 마이닝에서 활용될 감성과 관련된 문장을 분리하고, 감성과 관련 없는 부분을 제외한다. 다음 단계에서 문장을 분해, 단어를 도출하고 단어의 극성(긍정·부정)을 분석한다. 텍스트에 포함된 단어가 긍정적인지, 부정적인지를 도출한다. 마지막 단계에서 단어의 극성을 바탕으로 텍스트 극성을 탐지한다.

여론 마이닝을 이용하면 구매 후기, 상품평처럼 정량화하기 어려운 텍스트 데이터에서 유용한 객관적 정보를 추출할 수 있다. 상품 만족도를 표현하는 단순한 방법으로 '평점 시스템'이 있지만 이러한 시스템은 고객들의 다양한 관심사나 취향을 온전하게 반영지 못한다. 그러나 사람이 자유롭게 작성한 방대한 텍스트를 데이터로 모아서 분석하면 '좋았어요'와 같은 단어의 사용빈도로 만족도를 측정할 수 있을 뿐 아니라 상품의 장·단점, 연관 상품, 고객의 관심사 등 다양한

정보를 파악할 수 있다.

여론 마이닝의 진가는 2012년 미국의 대통령 선거에서 드러났다. 당시 오바마 대통령의 재선을 위해 노력했던 선거 캠프는 오피니언 마이닝을 통해 막대한 선거자금을 모금했다. 오바마 대통령 캠프는 유권자가 다양한 방식으로 적은 텍스트를 분석했다. 기부금 모금행사에서 기부금을 낼 가능성이 높은 그룹이 여성 유권자이고, 이들에게 어필할 수 있는 배우가 조지 클루니라는 사실까지 도출하여 성공적으로 기부금을 모금할 수 있었다.

미국 식품회사는 소셜미디어에서 사람들이 생성하는 코멘트를 실시간으로 분석해 소비자 정서를 파악하고, 그 결과를 초콜릿 바 가격에 실시간으로 반영하는 알고리즘을 만들었다. 이에 흥미를 느낀 고객이 늘어 캠페인 기간 매출이 67% 이상, 브랜드 관련 언급량이나 웹사이트 방문 고객 수가 1,000% 이상 증가했다.

여론 마이닝으로 텍스트에 내재된 가치를 구체적으로 분석할 수 있다. 여론 마이닝을 이용하면 모래알처럼 흩어져 있던 온라인 텍스트 데이터가 보다 가치 있는 정보로 바뀐다.

온라인과 소셜미디어의 여론 관리에 대한 중요성이 어느 때보다 막중한 시대다. 여론 마이닝에 대한 수요는 지속 증가할 전망이다.