

# 생각을 들여다 본다, 오피니언 마이닝

김준래 칼럼니스트

미국의 44대 대통령인 버락 오바마(Barack Obama) 전 대통령이 재선을 위해 애를 쓰던 시기인 2012년 5월의 어느 날, 그를 지지하는 할리우드의 미남 배우 조지 클루니(George Clooney)가 재선 자금 모금을 위해 성대한 저녁 만찬을 개최했다.

만찬에 참가하기 위한 조건은 단 하나. 지갑을 열어 티켓을 사기만 하면 되는 일이었다. 하지만 가격이 문제였다. 티켓 가격이 일반인들의 상상을 초월할 만큼 비싼 1인당 4만 달러였다.

높은 금액에도 불구하고 40대 여성 지지자들이 대거 참여해서 이 만찬은 무려 1,500만 달러라는 막대한 기부금을 모으는 데 성공했다.

대부분의 사람은 만찬 시작부터 끝까지의 모든 과정이 주최자인 클루니의 생각에서 나온 것이라 생각했지만, 사실은 전혀 아니었다.

감독은 따로 있었다. 바로 오바마 전 대통령의 선거캠프였다. 캠프에서 빅데이터를 담당했던 분석팀은 재선 2년 전부터 법이 허용하는 범위 안에서 각종 데이터를 수집해서 선거캠프는 정치 자금 모금 만찬을 개최했을 시 참여 가능성이 높은 대상과 이들에게 호감을 줄 수 있는 최적의 주최자 등을 분석했다.

그 결과 만찬 참여 가능성이 높은 대상으로 40대 여성층이 선택됐고, 이들의 지갑을 열 수 있는 주인공으로 클루니가 꼽혔다. '오피니언 마이닝'의 힘이었다.

## 여론과 의견을 유용한 정보로 재가공

오피니언 마이닝(opinion mining)이란 인터넷 웹사이트와 소셜미디어에 나타난 여론과 의견을 분석해서 유용한 정보로 재가공하는 기술이다. 일종의 '신개념 여론 분석 기술'이라 할 수 있다.

따라서 오피니언 마이닝을 활용하면 사람들이 특정 제품 및 서비스에 대해 긍정적인지 부정적인지에 대해 분석할 수 있고, 여기서 한 발 더 나아가 그 원인까지 도출할 수 있다.

특히 정치나 경제, 또는 사회적으로 특정한 사안들이 발생했을 때, 여론이나 대중의 관심도가 실시간으로 어떻게 변하는지를 확인할 수 있는 것은 오피니언 마이닝이 가진 강점이라 할 수 있다.

오피니언 마이닝은 오픈마켓이나 소셜커머스 같은 전자상거래 시스템과 함께 발전하기 시작했다. 상품을 직접 눈으로 보거나 손으로 만져볼 수 없는 전자상거래의 특성상 해당 상품을 먼저 구입했던 소비자들의



구매 후기가 커다란 영향을 미치면서, 오피니언 마이닝 기법에 대한 연구가 본격적으로 시작된 것이다.

소비자 입장에서 보면 수많은 구매 후기들 중에 실제로 상품을 구입하여 사용해 본 소비자들의 후기가 긍정적인지 또는 부정적인지를 구별할 필요가 있다. 반면에 판매자의 입장에서는 방대한 양의 구매 후기들 중에서 상품에 대한 소비자들의 반응을 아는 것이 필요하다.

이런 이유로 소비자들이나 판매자들 모두 자연어 처리 및 텍스트 마이닝 같은 기법을 활용하여 대량의 데이터 속에서 유용한 정보를 찾아내는 방법을 찾기 시작했고, 결과적으로 오피니언 마이닝이라는 기법이 탄생하게 된 계기가 됐다.

### 생각과 의견을 통계화하고 수치화

오피니언 마이닝 기법을 활용하면 네티즌들이 올리는 댓글이나 포스팅 등의 텍스트에서 긍정적 성향과 부정적 성향을 분류하여 더 객관적이고도 정확하게 상품이나 서비스의 평판을 파악할 수 있다. 그러나 사람들이 일상적으로 사용하는 말이나 글을 의미 있는 정보로 변환한다는 것이 결코 쉬운 일은 아니다.

예를 들어 '작다'라는 어휘가 세탁기를 평가하는 사용 후기에서 사용될 경우에는 '용량이 작다'라는 의미이기 때문에 부정적인 의미를 갖는다. 반면 똑같이 '작다'라는 표현이지만 휴대폰의 사용 후기에서는 소지하기 간편하다거나 귀엽다라는 긍정의 의미로 해석될 수 있다. 따라서 상황 정보에 대한 고려 없이 단순히 '작다'라고만 표현된다면 잘못된 결과가 나타날 수도 있는 것이다.

이 같은 오류를 피하고자 오피니언 마이닝은 텍스트 속에 포함된 진정한 의미를 찾기 위해 단계별 분석 기법을 도입한다.

첫 번째 기법은 글이 주관적인지 또는 객관적인지를 결정하는 '주관성 분석'이다. 주어진 텍스트에 나타난 글쓴이의 태도가 주관적인지 혹은 객관적인지를 판단하는 분석 기법이다. 여기서 주관적이라고 하는 것은 긍정적 또는 부정적 의견을 포함하는 것이고, 객관적이라고 하는 것은 감정적으로 중립 상태를 표시한다.

두 번째인 '극성 분석'은 주어진 텍스트가 주관적인 의견을 갖고 있을 경우 긍정인지 혹은 부정인지를 분류하는 기법이다. 이 기법을 적용하기 위해서는 분석하려는 글의 핵심 어휘나 신조어들을 미리 정의해 둘 필요가 있다. 특히 온라인상에서는 특정한 시기마다 유행하는 신조어가 올라오기 때문에 신속하게 업데이트하는 것이 중요하다.

마지막으로 '극성의 정도 분석'은 주관적으로 쓰여진 글에 대해 긍정적인 정도와 부정적인 정도를 측정하는 기법이다. 이 기법을 잘 활용해야만 오피니언 마이닝을 통해 제대로 된 정보를 확보할 수 있다.

이처럼 오피니언 마이닝은 까다롭고 복잡하지만, 모래알처럼 흩어져 있던 온라인상의 텍스트 데이터를 좀 더 가치 있는 정보로 바꿔준다.

온라인과 소셜미디어의 여론관리에 대한 중요성이 그 어느 때보다 부각되고 있는 상황에서 오피니언 마이닝 기법은 기업이나 단체의 의사결정에 있어 필수적인 솔루션이 될 것으로 전망된다. 

