

컴퓨터과학이 여는 세계 과제 4, Exercise 4 “사회현상예측/개인명성가능&위협”
『소셜 네트워크 데이터 분석: 청소년 심리불안 문제에서의 적용』

농경제사회학부 농업자원경제학전공 이동훈

Social Network의 등장과 함께 엄청난 양의 Social Network 데이터가 축적되면서, 예측 불가능하던 사회는 점차 Social Network 데이터를 적절히 분석함에 따라 예측 가능한 공간으로 변모하고 있다. 어떻게 Social Network 데이터를 분석함으로써 사회를 예측할 수 있는지 예를 통해 알아보자. 2008년부터 구글은 Flu Trends라는 서비스를 제공하였다. Flu Trends란 전세계 각지 사람들의 구글 검색어를 분석하여 특정 지역의 독감 확산을 예측하는 것이다. 사람들은 일반적으로 감기 증상이 나타난다고 생각되면 인터넷을 통해 자신이 느끼는 증상이 감기나 기타 호흡기 질병의 증상이 맞는지 확인하고자 한다. 이를테면 ‘인플루엔자’라던가 ‘인플루엔자 증상’과 같은 단어들을 검색하는 것이다. 구글은 이런 감기 관련 단어들의 검색 빈도와 검색 지역 등의 상관관계를 분석하여 독감이 유행할 것이라고 예상되는 지역을 예측하기 시작하였다. 이렇게 시행된 Flu Trend는 2011년부터 2013년까지 약간의 시행착오 기간을 제외하고는 대부분 미국의 질병통제예방센터의 독감 예측과 비슷하게 예측하였다.¹ 구글 Flu Trend는 왜 특정 지역에서 독감이 확산되었는지에 대한 필연적 인과관계를 설명해 주지 않는다. 단지 Social Network 데이터를 분석하고 개연적으로 그러한 결과가 나타날 확률이 높았음을 설명할 뿐이다. 이와 마찬가지로 최근의 사회학자들은 Social Network 데이터를 분석함으로써 사회에 주어지는 input과 그에 따라 나타나는 output 간의 인과관계를 설명하지는 못하지만, 어떤 input에 따라 굉장히 높은 확률로 어떠한 output이 나타날 것이라고 예측하기 시작하였다. 이런 맥락에서 점차 더 넓은 범위에서 예측가능해지는 사회를 Frank Pasquale는 “Black-box society”라고 일컫는다.² 많은 연구자들은 보다 나은 사회를 만들기 위해서, Social Network 데이터 분석을 통한 사회 예측을 적지적소에 활용하고자 노력하고 있다.

보다 나은 사회를 만들기 위한 연구의 일환으로, 우리는 Social Network 데이터를 분석함으로써 최근의 심각한 문제인 청소년 자살 및 청소년 비행문제를 예측하고 이를 미연에 방지할 수 있을 것이다. 다시 말해 청소년 자살 및 청소년 비행문제를 일으킬 가능성이 상대적으로 높은, 심리적으로 불안정한 상황에 놓여있는 청소년들을 그렇지 않은 청소년들과 구분하여 적절한 상담 및 교육을 행함으로써 그들의 자살이나 범죄를 막을 수 있을 것이다. 우리나라의 청소년 자살률은 2001년 10만명당 3.19명에서 2011년 5.58명으로 10년 사이에 57.2%가 증가했다.³ 동기간 내 OECD 31개국 평균 아동청소년의 자살률이 줄어든 것에 비하면 상대적으로 높은 수치이다. 이렇게 우리나라의 청소년자살률이 높은 원인 중 하나는 많은 청소년들이 우울, 분노 등의 부정적인 정서에 휩싸여 있기 때문이다. 이렇게 부정적이며 극도로 불

¹ 과학동아 편집부, ‘국회의원 인맥 찾아내는 소셜 사이언스’, “과학동아”2011년 02월호, 2011, p. 45.

² “How Big Data’s Inaccuracy Hurts People”, 2014.11. (<http://fleishmanhillard.com/2014/11/true/big-datas-inaccuracy-hurts-people/>). 2015.05.26.

³ 통계청, ‘사망원인통계’, 2011.

안정한 심리상태를 지닌 청소년들은 자살을 감행할 가능성이 높을 뿐만 아니라, 청소년 비행 등의 사회 문제를 일으킬 가능성이 높다. 또한 이들의 부정적 정서 상태를 방치해두면, 이들이 성인이 되어 장차 여러 가지 심각한 사회문제를 일으킬 가능성이 크다. 따라서 그들을 다른 청소년들과 구분해서 심층적인 심리치료 및 상담을 행하는 것이 보다 나은 사회를 만드는 데 큰 도움이 될 것이다. 그렇다면 과연 어떻게 Social Network 데이터를 이용하여 심리적으로 건전치 못한 청소년들을 구분해낼 수 있을까?

Social Network, 그 중에서도 Facebook에서 직접적 의사표현방법인 '답벼락 게시물'과 간접적 의사표현방법인 '좋아요' 버튼을 분석함으로써 심리적으로 불안정한 청소년들을 찾아낼 수 있을 것이다. 정서적으로 우울하거나 분노에 차 있는 청소년들은 자신의 감정이 또래집단의 다른 사람에게 공감되기를 바라는 경향이 있으므로 자신의 감정을 담은 글들을 Social Network 공간에 올리는 경향이 있다. 이를테면, 밝혀지지 않은 대상에 대한 욕설과 비방이 담긴 글이나 '자살하고 싶다' 등의 우울한 감정이 표출된 글이 바로 그런 것들이다. 이때, 게시물에 위로나 걱정의 댓글이 많이 달린 글보다 댓글이 하나도 안 달린 글을 쓴 청소년이 보다 더 부정적인 정서를 가지고 있을 확률이 높다. 한편, 우울이나 분노에 가득 차 있는 청소년들은 자신의 심정을 대변해주는 듯한 글에 '좋아요'를 누르는 경향이 크다. 따라서 정리하자면, 다음 두 가지 조건 '1) 우울이나 분노의 감정과 관련된 키워드들이 많은 글을 자주 올림과 동시에 쓴 글에 항상 적은 댓글이 달리는 청소년', '2) 우울이나 분노의 감정과 관련된 키워드들이 많은 글에 빈번하게 좋아요를 누르는 청소년'을 유의적으로 만족하는 청소년들이 바로 심리적으로 불안정한 청소년들이다. 이들은 곧 잠재적으로 자살을 시도하거나 범죄를 저지를 가능성이 상대적으로 높은 청소년들이므로, 데이터 분석을 통해 자살이나 범죄 등의 사회문제를 미리 예측하여 이들에 대해 심층적인 심리치료 및 상담이 이루어진다면 보다 나은 사회로 나아갈 수 있을 것이다.

하지만 이는 자칫하면 정서적으로 불안정한 청소년들이 사회적으로 우울증 환자로 낙인찍힐 수 있다는 점에서 위험성을 가진다. Pasquale의 의견처럼, 만약 시간이 흘러 회사의 고용인이 청소년기에는 불안정한 정서를 가졌지만 심리치료를 통해 성인이 되어서 건전한 정서를 가지게 된 피고용인을 평가한다고 하자. Social Network 데이터는 그 특성상 상대적으로 쉽게 거래되고 이용할 수 있기 때문에, 모든 고용인들이 그가 이전에 불안정한 시기를 보냈음을 확인하고 그를 고용하기를 망설일 수 있다. 이와 같이 우리가 모르는 사이에 우리의 사생활은 지금 이 순간에도 Social Network 데이터의 일종으로 기록되고 있으며, 상대적으로 쉽게 데이터들이 다른 사람들에 의해 열람될 수 있다는 점에서 사생활 침해의 위험성이 제기될 수 있다.

본인이 제시한 Social Network 데이터를 통한 청소년 자살 및 범죄 방지 문제의 경우, 사생활 침해 문제를 예방하기 위해 1차적으로 Facebook이나 Twitter 등의 소셜 네트워크가 보다 폐쇄적인 공간이 되어야 한다. 소셜 네트워크를 제공하는 기업 측에서는, 지인이 아닌 제 3자가 쉽게 다른 사람의 글이나 좋아요 등을 열람할 수 없도록 보다 신경 써야 한다. 또한 소셜 네트워크 이용자의 정보가 쉽게 유출되지 않도록 최선을 다해야 한다. 이와 동시에, 소셜 네트워크 이용자들은 자신의 게시물이 전체 공개 대상인지 친구에게만 공개되는 것인지 잘 확인하는 등 보다 조심스럽게 소셜 네트워크를 이용하려는 노력을 해야 할 것이다.

2차적으로는 정서적으로 문제를 가지는 청소년을 상담하고 교육시키는 해당 기관에서 그들에 대한

정보를 데이터화할 때, 그것이 다른 사람에게 쉽게 열람되지 못하도록 정보를 암호화하는 시스템을 도입해야 할 것이다. Pasquale는 비단 청소년 문제뿐만 아니라, 일반인의 의료 기록이 인터넷을 쉽게 열람될 수 있다면 엄청난 사생활 피해가 될 것이라고 주장한다. 이에 대한 적절한 대응이 이루어져 심리적 환자인 청소년들을 비롯한 많은 환자들이 사생활 침해에 따른 이중적 고통을 받지 않도록 해야 할 것이다.