

#### 4. 사회의 심리를 진단하고 치료하자

사회학과 이하영

사회학자 에밀 뒤르켐의 고전 <자살론>은, 개인적인 사정에 따른 것이라 여겨지는 자살이 실제로는 사회적 요인으로 인해 발생한다고 주장한다. <자살론>에 의하면 사적인 우울과 고난보다도 더 체계적으로 자살에 영향을 미치는 변수들이 있다. 자살률은 사회 통합과 규제의 정도에 따라 달라진다. 예컨대 조직 생활의 구속력이 약하고 신도의 개별적 자유를 더 허용하는 개신교 사회는, 종교적인 결사체에 의해 끈끈하게 통합되는 천주교 사회보다 자살률이 통계적으로 높다. 이외에도 다양한 통계적 분석을 바탕으로 뒤르켐은 자살을 4가지로 유형화한다. 이기적 자살은 사회 통합의 정도가 지나치게 낮을 때, 이타적 자살은 지나치게 높을 때 발생한다. 홀로 지내는 지식인의 자살은 전자의 예시이고, 세계 2차 대전 당시 카미카제 특공대의 자폭은 후자의 예시이다. 아노미적 자살은 사회를 지배하는 규범이 부재할 때, 숙명적 자살은 노예의 인생처럼 규제가 너무 무겁게 이루어질 때 발생한다.

특히 '이기적 자살'의 개념은 개인의 심리적인 고난이 사회적인 조건과 연관되어 있음을 실증한다. 어떤 이의 불안과 정신 병리, 자살에의 취약성은 그 사람만의 미시적인 문제가 아니라, 사회 전체가 직면한 상황과도 복잡하게 얽혀 있는 것이다. 따라서 개별 구성원의 심리 상태들은 전체 사회의 건강을 반영하고 암시한다.

빅 데이터로 개개인의 심리적인 불안 요소를 파악할 수 있다면, 그를 종합하여 사회가 처해 있는 위험을 진단할 수 있지 않을까? 예를 들어 20대의 카카오톡 유저들이 실업과 진로에 대하여 걱정하는 메시지들이 대거 수집된다면 그것을 통해 현재 청년들이 자신의 미래에 대해 불안해하고 있다고 밝혀낼 수 있는 것이다. 집단적인 관념과 고민을 빅 데이터로 알아내자는 의미이다.

요컨대 소셜 네트워크 데이터에 나타나는 사람들의 불안, 강박, 우울을 수집해 현 사회의 구성원들이 직면한 고민들을 분석하고 예측하는 작업이 가능하다. 현황이 파악된다면 해결책도 제시할 수 있을 것이다. 진단과 처방이 효과적일 경우 사람들의 삶의 질에 직접적이고 빠르게 기여할 수 있다.

이 작업은 구체적으로 어떤 과정을 거쳐 실현될 수 있을까? 먼저 심리적인 고난과 관련된 언어 표현들을 인간이 지정해 컴퓨터에 셋팅한다. '힘들다', '죽고 싶다', '불안하다', '슬프다', '우울하다' 등이 그 예시이다. 그 뒤로 소셜 네트워크 데이터를 쭉 훑으며 미리 지정해놓은 부정적인 감정 표현이 포함된 메시지들의 빈도를 측정한다. 이 빈도가 전체 메시지의 양과 비교했을 때 유의미하게 높다고 판단하면(그 비율은 미리 컴퓨터로 설정해야 할 것이다), 컴퓨터는 지정된 표현들이 사용된 맥락, 즉 힘겹다는 언급과 함께 그 전후로 정해진 용량만큼의 메시지를 더 저장해놓는다. 컴퓨터는 이 데이터를 이용하여 부정적 감정 표현들의 앞뒤로 빈번히 사용되는 어휘들을 골라낸 뒤 비슷한 것끼리 묶는다. 결과적으로는 사람들이 '힘들다'라는 표현과 함께 어떤 주제들을 언급하는지 알아낼 수 있다. 예상해보면 대학생들은 '연애', '취업', '과제' 등의 키워드에 대해 고민할 것이고 고등학생들은 '입시', '공부', 등과 부정적 언어를 연루할 것이다.

동일한 세대는 사회 구조에 급격한 변화가 있지 않는 이상 유사한 불안과 우울을 앓으므로, 현재 혹은 과거의 데이터들은 미래의 병리를 예측하는 데에 근거 자료가 될 수 있다. 사회 구성원들이 어떤 시기에 어떤 문제들로 고민하고 심리적으로 괴로워할지 예상하는 일이 가능해진다. 이는 장기적인 문제 해결의 기반이 되며 정책적인 제안으로 이어질 수도 있다

하지만 심리 분석을 위한 데이터 수집은 사생활의 극심한 침해를 초래한다. 고민들은 대개 내

밀한 개인사와 직간접적인 관련이 있다. 예컨대, 고민들을 포괄적으로 유형화해서 모두 '연애'의 카테고리 안에 일반화시킨다 해도 저마다의 만남과 이별의 사정은 지극히 개인적일 것이다. 그런데 컴퓨터가 진단 작업을 진행하기 위해서는 전체 데이터를 읽어야 하므로 사적인 사연들도 빠짐 없이 수집된다. 컴퓨터에게 비밀을 털어놓는 셈이다. 만약 컴퓨터가 소장한 소셜 네트워크 데이터가 유출되기라도 한다면, 주민등록번호나 전화번호 유출보다도 더욱 심각한 개인사의 유출이 발생할 수도 있다.

둘째로 수집된 데이터를 바탕으로 제도를 실행하거나 정책을 수립하는 과정에서 일반화의 오류를 범할 수 있다. 사람마다 SNS나 카톡을 이용하는 정도는 매우 상이하다. SNS의 활발한 이용자라고 해도, 자신의 감정을 진지하게 드러내는 사람들은 그 중에서 일부이다. 그런데 소셜 네트워크 데이터를 통해 고민을 수집하게 되면 적극적으로 의사를 표시하거나 SNS를 많이 사용하는 사람들의 감정은 과잉대표되는 반면, 그렇지 못한 사람들의 감정은 과소대표된다. 따라서 전자의 의견을 위주로 수집된 자료가 컴퓨터를 통과하여 결과로써 내놓는 의제들은, 모집단이 실제로 고민하는 문제들과 다를 수 있는 것이다. 데이터의 수집이 소규모로 이루어질 경우 이 문제는 더욱 부각된다. 예컨대 서울대학교 학생들을 대상으로 데이터를 입수하여 고민을 분석했다고 하자. 학생회는 컴퓨터가 출력한 고민-의제들을 중심으로 제도를 구상하고 실행한다. 그런데 수집된 정보들은 SNS 상에서 활동적인 서울대생의 의견을 토대로 만들어졌으므로, 소셜 네트워크를 잘 이용하지 않는 서울대생의 고민은 일반화 아래 무시될뿐더러 새로운 정책에서 소외될 것이다.

사생활을 보호하기 위해서는, 컴퓨터가 데이터를 수집하는 과정에서 고민과 관련된 용어들을 걸러낼 때 결과로 출력할 어휘들을 제외한 메시지들은 자동으로 삭제해야 한다. 예를 들어 "동생이 곧 대학에 들어가서 등록금을 마련해야 하는데, 요즘 취직이 어려워져서 힘들어."라는 메시지가 수집되었다고 하자. 컴퓨터는 이 메시지를 읽고 다른 메시지들과 비교하면서, 빈번히 사용되는 '등록금', '취직', '힘들어' 등을 추출한다. 이 때 메시지의 나머지 부분, 즉 "동생이 곧 대학에 들어간다"는 사적인 정보이므로 컴퓨터는 이를 데이터베이스에서 삭제하는 것이다. 하지만 무엇이 수집될 정보이고 무엇이 삭제될 정보인지를 판단하는 기준을 고안해야 한다. 한 가지 가능한 해결책은 명사만 수집하는 방식이다. 동사 없이는 사정의 맥락을 파악할 수 없기 때문이다. 하지만 동시에 결과로 출력된 의제들을 해석하기 어렵다는 단점도 있다.

일반화의 오류를 범하지 않기 위해서는, 컴퓨터가 내놓은 결과값을 공유한 뒤 그것이 사회적으로 수용된 이후에만 실제 정책이나 제도에 이용하는 방법이 있다. 요즘 ~세대는 ~한 고민들에 빠져 있으며 ~와 같은 심리적인 문제들을 경험한다고 컴퓨터가 공표하면 그에 대한 사람들의 동의나 반대를 살핀다. 만약 대다수의 경향성이 합의에 가까우면 해당 출력은 일반적으로 의미가 있고, 만약 반대가 우세하면 해당 출력은 편향된 것이다. 즉 결과의 공개와 결과를 반영한 제도의 실행 사이에 여론을 수렴하는 절차를 투입함으로써 일반적 수용 여부를 판단할 수 있다.

이러한 심리 진단 메커니즘은 문제에서 제시된 바와 같이 모든 사람들의 소셜 네트워크 데이터를 수집할 수 있다는 가정 하에 제시 가능하다. 현실에서는 사람들의 고민을 분석할 수 있을 정도로 양질의 데이터를 수집할 수 없을지 모른다.

그럼에도 불구하고 이러한 상상은 SNS 상에 표출되는 감정들을 무의미한 외침으로 치부하지 않고 사람들의 삶의 질에 실질적으로 기여하기 위해 이용한다는 점에서 그 의의가 있다.