

# 컴퓨터 과학이 여는 세계

027.013 Computational Civilization

**HW2 - 휴먼 컴퓨테이션 (Human Computation)**

인문대학 독어독문학과  
2010- 박진형

# 휴먼 컴퓨테이션, 인간과 컴퓨터가 함께 그려내는 미래

## 1. 서론

하나 둘 씩 기계가 발명될 때의 인간은 계산기 따위의 기계가 모방할 수 없는 뛰어난 존재였다. 그때까지의 기계는 인간이 할 수는 있지만 번거롭거나 위험한 일을 대신해주는 보조적인 도구였다. 그러 증기 기관의 발명으로 산업 혁명을 거치고, 컴퓨터와 인터넷의 발명으로 정보 혁명을 거치면서, 일각에서는 기계가 모든 것을 해결해줄 것이라는 기술만능주의가 나타날 정도로 기계 특히 컴퓨터의 능력이 급속도로 부상하게 되었다. 한 때 인공지능에 대한 열망을 품고 다양한 분야의 과학자들이 인간과 같은 기계를 만들기 위해 연구하는 것이 대대적인 유행일 정도였다. 그러나 점차 프로젝트가 진행되면서 인간에게는 너무나도 쉬운 일을 컴퓨터로 처리하기 힘들다는 것이 밝혀졌고, 서서히 기계의 한계와 인간의 능력에 대한 논의가 진행되기 시작했다. 이러한 변증법적 과정을 거쳐 오늘날에는 컴퓨터가 잘 하는 부분과 인간이 잘 하는 부분을 적절히 조합시킨 휴먼 컴퓨테이션이 등장하기에 이르렀다. 이 분야의 선구자인 루이스 폰 안(Luis von Ahn)과 EteRNA의 예를 살펴본 후 논의를 더 확장시켜 보겠다.

### a) 루이스 폰 안 - 휴먼 컴퓨테이션의 선구자

루이스 폰 안은 인간과 컴퓨터가 협력할 수 있다는 것을 처음으로 인식한 몇 안 되는 과학자 중 하나였다. 그의 업적 중 가장 유명한 것 두 가지가 바로 ESP Game과 reCAPTCHA다. ESP Game은 어떤 이미지를 인식하는 능력에 있어서는 인간이 컴퓨터 보다 월등하다는 점에 착안한 것으로, 과학적인 목적을 가지고 설계된 첫 번째 게임이다. 전체적인 구성을 보면, 먼저 두 명의 플레이어가 하나의 이미지를 보고 제한 시간 안에 각각 단어를 쓴다. 그리고 참가자들이 쓴 단어가 일치하면 인센티브를 주는 식이다. 두 명의 참가자가 필요하도록 게임을 설계한 것은 매우 절묘한 방법이다. 컴퓨터가 그것이 맞는지 틀린지 알 수 없는 상황에서 또 한 명의 참가자를 등장시켜 결과적으로 컴퓨터가 둘의 답안이 일치하는지만 검사하도록 만들었기 때문이다. 이렇게 수집된 이미지를 묘사하는 정보는 컴퓨터에 저장되고, 패턴화 된다. 구글 이미지 검색의 정확도를 높이는 데 사용되었으며, 이제는 구글에 의해 운영되고 있다. reCAPTCHA 역시 인간의 패턴 인식 능력과 컴퓨터의 데이터 처리 능력을 이용한 것이다. 자동가입방지를 위해 개발되었던 CAPTCHA에서 한 단계 발전한 것으로, 컴퓨터가 왜곡된 형태의 단어 두 개를 보여주면 사람이 다시 입력하는 방식이다. 앞서 ESP Game에서는 두 명의 참가자를 등장시켜 진위 여부를 가렸다면, reCAPTCHA에서는 컴퓨터가 알고 있는 단어 하나 모르는 단어 하나를 제시함으로써 컴퓨터가 온전히 진위 여부를 가린다. 각기 다른 사람이 입력한 열 개의 동일한 경우를 두고 비교해보기 때문에 새로 입력된 정보가 틀릴 일은 거의 없다. 수집된 정보는 고문서의 내용을 컴퓨터가 읽는 데 쓰인다<sup>1)</sup>. 근래에 폰 안이 주도하는 프로젝트는 Duolingo라는 번역 프로그램이다. ESP Game과 같이 이중적인 목적을 가진 프로그램으로 사용자는 번역을 하며 무료로 외국어를 배울 수 있고, 수집된 자료는 정확한 자동 번역을 실행하는 데 쓰인다. 구조적으로 초급, 중급, 고급의 다양한 레벨로 나누어져 있고, 모든 단어의 뜻이 제시되어 있으므로 굉장히 효율적이다. 실제로 전문 번역가의 번역과 참가자의 번역이 크게 다르지 않을

1) cf. Wikipedia, "Luis von Ahn." [http://en.wikipedia.org/wiki/Luis\\_von\\_Ahn](http://en.wikipedia.org/wiki/Luis_von_Ahn)

수 있는 이유다. 위키피디아에 있는 영어 자료를 스페인어로 번역하는 일을 예로들어 통계적으로 비교해보자면, 전문 번역가들을 고용하면 오천만 달러의 예산이 들고 천문학적인 시간이 걸리는 것을 만 명의 참가자들을 가지고 무료로 하루만에 처리할 수 있는 정도로 강력한 것이다. 또한 Duolingo는 부가적으로 가난한 사람들에게도 외국어를 배울 수 있는 기회를 제공한다는 공공의 가치를 지니고 있기도 하다<sup>2)</sup>.

#### b) EteRNA - RNA구조를 밝혀내는 획기적인 비디오 게임

RNA는 단백질을 구성하고 유전자를 규제하기 위해 필요한 생명체의 필수요소다. RNA를 연구하고 그 비밀을 풀어내게 되면, 신약 개발과 질병 치료에 획기적인 발전을 이루어낼 수 있는 것이다. 그러나 그동안 과학자들이 RNA를 과소평가해왔기 때문에 RNA 구조에 대해 축적된 연구 결과가 부족해, RNA에 관한 심화된 연구는 난관에 봉착했다. 이 때 Adrien Treuille 팀이 처음 개발한 것이 단백질을 변형하는 게임 Rosetta@home이다. 이것이 시초가 되어 EteRNA게임이 만들어졌다. 게임 플레이어들 사이에 랭킹이 있고, 뛰어난 플레이어들은 특별한 미션을 부여받는데, 정기적으로 우수한 RNA구조모델이 선택되어 스탠포드 실험실에서 배양된다. EteRNA 전에도 RNA를 모델링하는 알고리즘은 있었다. 그러나 빈약한 연구 결과를 가지고 설계된 알고리즘이니만큼 문제점이 많았다. 빈약한 정보를 가지고 설계된 알고리즘을 가지고 전혀 다른 새로운 패턴의 구조를 찾아내기는 무리였던 것이다. 컴퓨터를 가지고 상당 부분이 미지로 남아있는 문제를 해결하는 것은 역부족이었다. 그러나 인간은 고작 수 십 번의 실패를 겪고 적절한 결합 구조를 찾아낼 수 있는 직관을 가지고 있었다. 게임을 설계하고 만들어낸 과학자들조차 놀랄 만한 플레이어들이 등장했고, 그들은 그들이 가지고 있던 직관을 이내 법칙으로 만들어내 RNA연구에 획기적인 발전을 가져왔다. 플레이어가 너무 뛰어난 나머지 전문적으로 트레이닝을 받은 과학자들의 지위를 위협하는 정도라고 하니 그들의 기여가 얼마나 엄청난 것인지 추측해볼 수 있다. 흥미로운 점은 인간이 밝혀낸 새로운 패턴을 알고리즘에 입력시켜 이제 인간보다 나은 알고리즘이 탄생했다는 것이다. 인간이 미지로 남아있던 부분의 많은 것을 찾아줌에 따라 컴퓨터의 정보 검색공간이 획기적으로 줄고 알고리즘의 정확성과 효율이 올라간 것이다. 휴먼 컴퓨테이션의 대표적인 예라고 할 수 있다<sup>3)</sup>.

두 가지 예 모두 인간의 패턴 인식 능력을 활용하고 있다. 형태가 왜곡되어 있거나, 정의되어 있지 않은 미지의 것일 때 인간의 인지 능력의 진가가 드러난다. 인간은 어떤 말로 표현하기 힘든 직관을 가지고 있는데, 이러한 직관이 법칙으로 구현되고 말로 표현되면 기계가 배울 수 있게 되는 것이다. 또한 Duolingo의 예에서 볼 수 있는 것처럼, 인간은 컴퓨터와 비교하여 의미를 이해하고 조직화하는데 탁월한 능력을 보인다. 반면, 컴퓨터는 월등한 데이터 처리 능력을 보유하고 있다. 매일 수집되는 천문학적인 양의 데이터들이 빠른 시간에 분류되고 처리된다. 그리고 인간과의 상호작용으로 업그레이드 된 알고리즘은 금세 인간의 능력을 추월하고 있다. 마지막으로 컴퓨터는 플랫폼으로서의 역할을 한다. 두 가지 예

2) cf.Luis von Ahn TED Talk.

[http://www.ted.com/talks/luis\\_von\\_ahn\\_massive\\_scale\\_online\\_collaboration.html](http://www.ted.com/talks/luis_von_ahn_massive_scale_online_collaboration.html)

3) Brendan I. Koerner, "New Videogame Lets Amateur Researchers Mess With RNA", *Wired Magazine*, 07.05.2012. : [www.wired.com/wiredscience/2012/07/ff\\_rnagame/all](http://www.wired.com/wiredscience/2012/07/ff_rnagame/all)

모두 세계 각지의 수많은 사람들이 동시 다발적으로 온라인에 접속하기 때문에 더욱 효과적으로 진행될 수 있었다.

## 2. 연장선상에서 - 휴먼 컴퓨테이션과 문화인류학

단기간에 수많은 데이터를 처리하는 컴퓨터와 의미있는 정보의 패턴을 선별해내는 인간의 콜라보레이션은 꼭 자연과학 분야가 아니더라도 광범위하게 적용될 수 있을 것이라는 생각이 든다. 이미 사회과학 분야에서도 휴먼 컴퓨테이션의 움직임이 나타나고 있다. 내가 공부하고 있는 문화인류학 분야에서도 컴퓨터의 빠르고 효율적인 정보처리 능력은 획기적인 발견을 촉진시킬 것으로 기대된다. 컴퓨터가 기여할 수 있는 바를 보여주는 단적인 예가 프란츠 보아스(Franz Boas)의 전설적인 인체 측정 작업이다. 이 연구는 문화인류학에서 가장 고전적인 연구 중의 하나로 남아있다. 1911년 보아스는 1만 7821명의 이주자를 대상으로 한 대규모 두개골 연구 결과를 발표하고, 자료를 정밀하게 통계 분석하였다. 보아스는 이주해온 부모와 미국에서 출생한 자녀들의 두개골 형태가 상당히 다르다는 사실에 의거해, 그것이 전혀 영구적이지 않다는 점을 보여주었다. 이 결과는 상당히 파격적인 것이었는데, 당시 두개골 측정에 기반하여 인종적 변이를 정의하고 인종차별을 정당화하는 것이 당연시되었기 때문이다. 이 선구적인 연구의 결과로 보아스는 고정되어 있다고 믿어진 특징, 유전적으로 전해진 특징이 사실은 환경에 의해 바뀐다는 것을 증명했다. 또한 두개골의 형태처럼 변하지 않을 것 같은 인종적 특징이 환경의 영향을 받는다면, 모든 인종적 분류와 특징은 의심해보아야 한다는 점을 인식시켰다<sup>4)</sup>. 당시에는 그럴 수 없었지만, 이 모든 과정이 컴퓨터와의 협업으로 진행되었다면 얼마나 빨리 자료를 분류하고, 통계를 내고, 또 숨겨진 의미를 읽어낼 수 있었을까하는 생각이 든다. 더 나아가서 컴퓨터의 정보처리 능력은 문화인류학의 오랜 믿음 가운데 하나인 문화의 패턴을 찾아내는데 지대한 공헌을 할 수 있을 것이다. 문화의 패턴은 루스 베네딕트의 저작 『문화의 패턴』으로부터 유명해진 개념인데, 루스 베네딕트는 문화의 패턴화를 이렇게 설명하고 있다.

문화적 행동은 국지적, 인공적, 아주 가변적인 것이다, 라고 이해한 것만으로는 그 행동의 의미를 완벽하게 파악했다고 할 수 없다. 그런 행동은 거기서 한 걸음 더 나아가 통합되는 경향을 보이기 때문이다. 문화는 개인과 마찬가지로 생각과 행동의 패턴에 어떤 지속적 일관성을 지니고 있다. 각 문화 내에서 발생하는 어떤 특징적 목적들이 있는데 그것들이 반드시 다른 타입의 사회에 의해 공유되는 것은 아니다. ...중략... 생활, 혼인, 전쟁, 신의 예배 등과 관련된 사소한 행동들이, 그 문화 내에서 발달된 선택의 무의식적 기준에 힘입어, 어떤 지속적 패턴으로 굳어지는 것이다<sup>5)</sup>.

문화인류학에서는 줄곧 특정 문화에는 그 문화의 특징적인 일관적이고도 지속적인 어떤 패턴이 나타난다고 생각해왔지만, 포괄적이고도 총체적인 문화의 범위 때문에 그것을 과학적으로 입증해내기가 어려웠다. 그러나 문화인류학자들이 인간의 생활에서 의미를 부여할 만한 수량적 정보들을 찾아내고, 또 그런 정보를 처리할만한 알고리즘을 만들어낸다면, 휴먼 컴퓨테이션을 통해 문화의 패턴과 관련한 획기적인 발견을 할 수 있을 것으로 생각된다. 문화의 의미론적 요소들을 수량화하는 데는 어려움이 있겠지만 할 수 없는 것은 아니다. 패턴을 찾아내기 위해 상징과 의미에 양적 가치를 부여하는 작업은 이미 시작되었다. SNS의 발달로 급부상하고 있는 소셜 네트워크 분석이 그러한 예다. 페이스북은 업로드 된 모든 사

4) 제리 무어, 김우영 (역), 2002, 『인류학의 거장들』, 한길사, p.78.

5) 루스 베네딕트, 이종인 (역), 2008, 『문화의 패턴』, 연암서가, pp.89-90.

용자 정보를 영구적으로 소장한다는 것을 명시하고 있는데, 하루에서 상상을 뛰어넘는 양의 데이터들이 수집된다. 페이스북이 하는 일은 수집된 데이터가 의미 있는 인간관계의 구조를 나타낼 수 있도록 다양한 알고리즘을 적용해 패턴과 패턴이 함축하는 의미를 찾아내는 것이다. 같은 종류의 예로 Aral과 Alstynne의 연구를 들 수 있다. 이들은 같은 회사에서 일하는 참가자들의 10개월 동안의 이메일 저장내역을 분석했다. 이들이 수집한 메일은 십이 만 개를 넘는 것이었으며, 이밖에도 설문조사 자료와 참가자들의 개인정보(나이, 교육수준, 동종업계 근무 경험, 정보 조사 습관 등)가 수집되어 분석되었다. 수집된 정보가 다방면으로 분류되고 통계화 되었는데, 그 중 하나가 키워드 선별을 통해 받은 편지함과 보낸 편지함 모두를 포함하여 이메일의 주제 분포와 다양성을 분류한 것이었다<sup>6)</sup>.

컴퓨터 과학의 발전은 실로 인간의 계산능력을 확장시키는 것이었다. 위에 언급한 것과 같은 방대한 양의 데이터 처리는 전무후무한 일이다. 문화인류학자들이 어떠한 데이터를 수집할 것인지 정하고, 어떻게 분류할 것인지만 알아도 데이터 스스로 패턴을 드러내는 놀라운 일이 생길지도 모른다. 물론, 의미있는 패턴을 찾아내고 함축된 의미를 분석해내는 것은 여전히 사람의 일이다. 그러나 신용카드사용 내역이나, 행동반경과 같은 일상생활과 밀접하게 관련 있는 데이터들을 천문학적으로 수집하고 문화권 별로 비교하는 것 만으로도, 상당 부분의 패턴이 드러나게 되리라고 본다.

문화 간 비교 연구가 휴먼 컴퓨테이션을 통해 이렇게 광범위하게 이루어질 수 있게 되면, 문화가 인간의 사고에 미치는 영향에 대해 연구하는 인지인류학 분야에도 많은 영향을 끼칠 것이다. 어쩌면 19세기 유럽의 중심적인 민족학자였던 아돌프 바스티안(Adolf Bastian)이 주장했던 원질 사고(Elementargedanken)와 문화적 사고(Voelkergedanken)에 대해 방대한 양의 과학적인 근거를 제시할 수 있게 될 수도 있다. 여기서 원질 사고란, “가장 기초적인 생각들”로서 인간이라면 공유하는 내적 경향을 말한다. 바스티안은 이러한 내적 경향에 근거하여 독특한 사고의 패턴이 나타난다고 보았다. 문화적 사고는 인간이 환경에 적응하고 다른 집단과 상호작용하며 형성되는 것이다. 문화적 사고는 다양한 집단을 각각 특징짓는다. 바스티안은 원질 사고는 인류의 문화적 다양성 뒤에 숨겨져 있다고 보았다<sup>7)</sup>. 즉, 인간과 컴퓨터의 대대적인 콜라보레이션을 통해 인간이 공유하는 본질적인 사고의 패턴은 어떤지, 문화로부터 영향을 받은 사고의 패턴은 어떻게 나타나는지 거시적으로나마 알 수 있게 된다는 것이다.

문화 간 연구가 중요한 것은 다양한 문화들 간의 차이점을 드러내면서 상호 커뮤니케이션의 폭이 넓어진다는 데 있다. 문화의 차이를 이해하면 오해와 무지에서 비롯된 수많은 갈등을 해소할 수 있다. 휴먼 컴퓨테이션은 적절히 활용된다면 문화인류학 분야에서 세계 평화에 기여하는 촉진제 역할을 하게 될 것이다.

### 3. 결론

휴먼 컴퓨테이션은 이제 막 시도되고 있는 분야다. 즉, 그만큼 미숙하지만 또 그만큼 무궁무진한 가능성을 내포하고 있다는 것이다. 컴퓨터가 어디까지 우리의 사고를 확장시켜줄지 모르기에 휴먼 컴퓨테이션의 발전 가능성도 그만큼 무한하다고 할 수 있다. 본문에서는

6) Sinan Aral et al., 2006, "Network Structure & Information Advantage", *Workshop on Information Systems Economics*.

7) c.f. H. Glenn Penny et al., ed. 2003, *Worldly Provincialism: German Anthropology in the Age of Empire*. Ann Arbor: University of Michigan Press.

주로 문화인류학 분야에 국한하여 적용 가능성을 검토해 보았다. 문화인류학은 문화의 다양성을 연구하는 학문으로 어찌면 컴퓨터와는 가장 거리가 멀게 느껴질 수도 있다. 그러나 앞서 살펴본 것처럼, 문화인류학 분야에서도 휴먼 컴퓨테이션은 중요한 의미를 가지고 광범위하게 사용될 수 있다. 이것은 반대로 문화인류학보다 수치에, 통계에 가까운 분야에서 휴먼 컴퓨테이션이 얼마나 크게 기여할 수 있는가를 반증하는 것이기도 하다. 문화인류학에서 이 정도인데, 자연과학 분야에서는 말할 것도 없을 것이다. 그러나 휴먼 컴퓨테이션이 다양한 분야에 널리 보급되기 위해서는 과학자들의 컴퓨터에 대한 이해가 선행되어야 할 것이다. 사회과학자들이 컴퓨터의 기본 원리를 이해하고 프로그래머들과 협업해야 보다 더 생산적인 연구 결과를 도출해낼 수 있을 것이다. 휴먼 컴퓨테이션이 더욱 효과적으로 광범위하게 적용되기 위해서는 아직 여러 가지 과제가 남아있지만, 벌써부터 인간과 컴퓨터가 그려낼 미래가 기다려진다.