Artificial Intelligence Times
인공지능신문
www.aitimes.kr

서울대 국문학과 박진호 교수, 언어인공지능이 언어학 도메인 지식과 만나면 AI 자연어처리 능력은 어떻게 변화될까?

정한영 기자 | 승인 2022.10.11 12:09

> **"현재의 딥러닝 기반 AI나 자연어처리 시스템을 완전히 갈아엎자는 건 아니고 최대한 활용을 하되 언어학의 도메인 지식도 활용해 최선의 결과를 얻자는 것!"**

서울대학교 국어국문학과 박진호 교수

지금의 인공지능은 언어학적 지식 없이도 상당한 수준의 자연어처리(NLP) 역량을 보여주고 있지만, AI 언어 인공지능 모델은 인간처럼 우리를 이해하고, 참여하며, 대화할 수 있는 진정한 지능형 AI 시스템과는 아직 거리가 있다. 실제 환경에 더 적응할 수 있는 모델을 구축하기 위해서 AI는 사람들과 함께 보다 다양하고 광범위한 시각과 언어학적 기반과 조화로 수행되어야 하는 새로운 연구가 지속적으로 이어지고 있다.

이에 본지는 코딩과 AI 개발이 가능한 국내에 독보적인 한국어학자이자 여러 가지 다른 언어의 체계적인 기술과 비교를 통해서, 자연언어의 개별성과 보편성을 명백히 하는 언어유형론(Typology)적 관점에서의 국내 최고의 전문가인 서울대학교 국어국문학과 박진호 교수를 만나 언어 인공지능을 조명하고 언어학의 도메인 지식을 활용하면 자연어처리 능력을 어떻게 고도화 할 수 있는지 알아본다.

이번 인터뷰는 KAIST 전산학 박사로 전 세종대 교수이자 ICT 전문가로 현재, 인공지능 데이터 사업의 기획을 위한 총괄위원회 위원 및 다양한 AI위원회의 위원으로 활동 중인 한상기 박사가 진행하고 정한영 기자가 정리했다.(편집자 주)

Q . *먼저, 교수님 소개 부탁드립니다.*

A . 저는 한국어를 연구하는 언어학자이고요, 서울대 국어국문학과에서 학사 석사 박사를 했습니다. 국문과는 좀 보수적인 동네인데, 저는 순수한 학문뿐 아니라 그것을 실생활에 응용하는 데 관심이 있었고, 문과와 이과의 경계선에 있는 성향이기도 합니다.

대학원 박사과정 수료하고 조교로 있던 1998년부터 문화체육관광부의 21세기 세종 계획이 시작돼서 10년간 지속되었는데, 쿠퍼스(corpus), 전자사전 등 자연어처리(NLP)에 도움이 될 수 있는 데이터를 구축하려는 프로젝트였습니다. 이 프로젝트에 참여하면서 자연어처리에 관심을 갖게 됐습니다.

보통의 인문학 공부는 책이나 논문을 읽는 식으로 공부하면 되는 데 비해, 자연어 처리는 논저를 많이 읽는다고 되는 게 아니라 실전적인 경험이 굉장히 중요하다는 걸 느꼈어요. 실전 경험에 대한 갈증을 해결하기 위해 <언어과학>이라는 회사에 입사했습니다.

기계번역, 음성인식, 음성합성을 주로 하던 회사인데, 거기서 일한 번역기 만드는 일에 투입이 됐습니다. 개발자들과 소통을 원활하게 하기 위해 독학으로 프로그래밍을 공부하기 시작했고, 제가 원래 하던 국어학 주전공도 있지만 자연어 처리 쪽이 부전공처럼 됐습니다.

Q . *최근 2, 3년 동안에 초거대 AI 또는 거대 언어 모델(large language models. 이하, LLM) 등 AI 언어 모델에 대한 발전이 굉장히 두드러진 가운데 언어학자, 국어학자로서 현재 인공지능 자연어 처리 기술 수준을 어떻게 평가하시는지요*

A . '놀랍다, 대단하다. 이런 것까지 되는구나'라는 느낌이 제일 큽니다만, 본질적인 것을 놓치고 있는 것은 아닌가 하는 아쉬운 느낌도 있어요. 자연어처리(NLP)의 역사를 돌이켜보면 몇 십 년 전 기호 기반 AI(symbolic AI)가 유행하던 시절에는 언어 전문가가 언어 처리에 필요한 지식과 규칙을 작성하고 이를 코드로 전환해서 AI를 만들려고 했는데 아시다시피 별로 성과가 좋지 않았죠.

산업계에서도 학계에서도 그런 방식으로 도메인 전문가의 지식을 대량으로 구축을 해서 코드화하여 AI를 만드는 방식은 잊혀지고 사장됐다고 할 수가 있겠고, 머신러닝 특히 딥러닝이 나와서 도메인 전문가의 지식을 명시적으로 표상하지 않더라도 어떤 과업을 수행할 때 필요한 패턴이나 특징을 신경망이 스스로 알아서 다 찾아서 묵시적으로 표상해도 과업을 잘 수행하게 되었죠.

도메인 전문가의 역할은 점점 축소가 됐고, 요즘은 '도메인 지식 별로 필요 없다. 그냥 종단간(end-to-end)으로 데이터만 많이 있고 모델 크기가 충분히 크면 일을 잘 할 수 있다'는 분위기가 팽배해 있습니다.

저 같은 도메인 전문가로서는 서글픈 일입니다만, 이런 상황이 초래된 데는 저 같은 언어학자의 책임이 크다고 생각합니다. 저희가 일을 잘 해서 성과가 잘 나왔다면 도메인 전문가가 여전히 높이 평가될 텐데 그렇게 해서는 잘 안 됐고 오히려 도메인 전문가를 배제하고 했더니 잘 되고 있으니까요.

Q . *언어학자들이 아이들의 언어 습득 과정, 인간이 언어를 배우고 그것을 활용하는 것에 대해서 연구를 많이 하시는데, 그런 입장에서 보면 말을 엄청나게 많이 듣는다고 해서 제대로 된 언어습득이 되는 거라고 볼 수 있을까요?*

A . 아주 중요한 부분을 짚어주셨는데요, 현재의 딥러닝 방식의 자연어처리는 학습을 위해서 엄청나게 많은 데이터를 필요로 하는데요, 인간이 말을 배울 때 그렇게까지 많은 데이터가 필요하지는 않거든요. 훨씬 소량의 데이터 가지고도 언어를 완벽하게 습득합니다.

그리고 지금의 LLM은 그 규모가 엄청나잖아요. 물론 인간의 뇌의 뉴런의 구조도 엄청나게 복잡하지만, 그래도 에너지 소모량 같은 측면에서 보면 인간의 뇌가 훨씬 더 경제적, 효율적으로 학습을 하죠. 현재의 LLM으로 대표하는 주류 접근법은 비효율적인 거죠. 데이터도 너무 많이 필요하고 모델도 너무 크고 에너지 소비도 많고. 추세는 그걸 더 키워서 더 잘 해보겠다는 쪽으로 계속 가고 있는데, 뭔가 아쉬워요. 인간은 훨씬 작고 효율적인 모델 가지고도 잘하고 있는데 말이에요.

Q . *현재의 이런 패러다임으로 가면 언어, 한국어를 전공하는 학생들, 연구자들의 미래 전망은 어떻습니까?*

"한국어에는 이렇게 뭐가 생략되고 축약되는 현상이 많은데, 이런 변형의 유형을 철저하게 찾았더니 한 200가지 유형이 있더라고요. 그래서 이 문제를 변형이 아니라 분류 문제로 풀 수 있게 된 거고, 딥러닝이 분류는 굉장히 잘하거든요. 그래서 기존의 95%를 뛰어넘어 f1 스코어 98%의 성과를 얻었습니다."(사진:최광민 기자)

A . 자연어처리에서 언어학자가 기여할 수 있는 부분을 우리가 찾아서 후학들에게 길을 열어줘야겠죠. 그래서 제가 관심을 갖고 연구해 온 것은, 현재의 딥러닝 기반 AI나 자연어처리 시스템을 완전히 갈아엎자는 건 아니고 최대한 활용을 하되 언어학의 도메인 지식도 활용하여 최선의 결과를 얻자는 겁니다.

딥러닝을 갈아엎지 말자는 것은, 과거에 도메인 전문가가 규칙을 만들고 그걸 코드화하는 방식에 비해서 딥러닝이 패턴을 잘 포착하고 더 잘 하는 부분이 분명히 있기 때문입니다. 딥러닝이 더 잘 하는 부분은 더 잘 할 수 있게 하되, 그렇게 모델을 크게 키울 필요는 없고 최소한의 모델로 좋은 결과를 얻자는 겁니다. 딥러닝 기반 언어모델이 잘 못하는 부분들도 있는데 그런 틈새를 찾아서 언어학의 도메인 지식으로 메꿔줘서, 자연어처리 시스템의 성능 향상을 꾀하자는 겁니다.

예를 들어, 제가 실제로 해본 작업 두 가지를 소개하겠습니다. 하나는 한국어 형태소 분석기입니다. 머신러닝이 유행하기 전에 이미 95~96%의 정확도를 보이는 꽤 쓸 만한 한국어 형태소 분석기들이 많이 나와 있었는데, 이 정도면 쓸 만하다고 할 수도 있겠지만, 과업에 따라서는 아쉽다고 할 수 있습니다. 95%라면 20번 중에 한 번은 틀린다는 얘기인데요, 문어체 텍스트에서 하나의 문장이 20개 정도의 어절로 이루어져 있다고 하면 한 문장에 하나는 틀린다는 얘기예요.

그리고 형태소 분석은 그 다음 단계의 여러 가지 자연어 처리 과업으 수행하기 위한 첫 단계 분석인데 여기서 오류가 나오면 그 오류를 떠안고서 그 다음 단계로 넘어가니까 그 오류가 증폭됩니다. 그래서 형태소 분석은 정확도가 더 높아질 필요가 있습니다. 95~96% 정도의 정확도에서 획기적인 돌파구가 나오지 않고 정체된 상태였고, 이건 딥러닝이 유행하게 된 뒤에도 마찬가지였어요.

영어는 LSTM 기반 parts-of-speech(POS) tagger가 나와서 발전이 많이 이루어졌어요. 영어에서 어떤 단어의 품사가 명사도 있고 동사도 있어서 중의성이 있는데 이걸 해소하려면 앞뒤 문맥을 보고서 해야 됩니다. 중의성 해소의 단서가 가까이 있으면 기존 은닉 마르코프(Hidden Markov) 모델 같은 걸로 잘 해결되는데 그 단서가 좀 멀리 떨어져 있는 경우에는 잘 못했어요. 그런데 LSTM은 꽤 원거리에 있는 정보도 잘 추출할 수가 있어서 상당한 성능 향상을 보았는데, 한국어 형태소 분석은 영어하고 다릅니다.

영어는 그냥 띄어쓰기 단위를 하나의 단어로 봐서 처리하면 되는데 한국어는 체언 뒤에 조사가 붙고 용언 어간 뒤에 어미가 붙어서 복잡합니다. "밥을 먹었다."에서 "밥"과 "을"을 떼어내야 되고 "먹었다."도 "먹", "었", "다"를 떼어내야 합니다. 그 다음에 LSTM 모델을 적용해서 각각의 형태소에 품사를 부여하면 됩니다만, 이건 형태소 분리가 잘 되어 있음을 전제로 합니다.

그런데 형태소 분리가 난이도가 있고 어려운 일이었어요. 형태소 분리하는 과제를 딥러닝으로 해결하려고 하면 잘 안 됐던 거죠. "먹었다"가 입력으로 들어오면 "먹"+"었"+"다"를 출력으로 내야 됩니다. 머신러닝이 가장 잘 하는 과제가 분류(classification)와 회귀(regression)인데, 위의 과제는 분류도 아니고 회귀도 아니고 일종의 변형이잖아요. 이 과제를 그냥 변형 과제로 놓고 딥러닝으로 해결하려고 하면 아주 잘하기가 쉽지 않습니다.

저는 국어학자로서 한국어에 대한 지식이 있어서 이 과제를 분류 과제로 재정의했습니다. "먹었다"는 비교적 쉽지만 “흘렀다"는 불규칙 활용을 하니까 "흐르"+"었"+"다"로 해야 됩니다,

"저 사람은 학생이다."는 "학생"+"이"+"다"로 하면 되니까 쉬운데 "저 사람은 내 친구다."에서 "친구다"는 "친구"+"다"로 분석하면 안 되고, 그 뒤에 구문 분석이나 의미 분석을 제대로 하려면 "친구"+"이"+"다"로 분석해야 돼요. "이"가 생략된 거죠.

한국어에는 이렇게 뭐가 생략되고 축약되는 현상이 많은데, 이런 변형의 유형을 철저하게 찾았더니 한 200가지 유형이 있더라고요. 그래서 이 문제를 변형이 아니라 분류 문제로 풀 수 있게 된 거고, 딥러닝이 분류는 굉장히 잘하거든요. 그래서 기존의 95%를 뛰어넘어 f1 스코어 98%의 성과를 얻었습니다.

Q . *요즘은 형태소 분석을 그렇게 중요하게 안 보는 거 아닌가요? 형태소 분석을 안 해도 다운스트림 과업을 잘 하면 된다는 생각 같은데요?*

A . 맞습니다. 형태소 분석기가 한국어 자연어처리에서 갖는 중요도가 예전보다 많이 낮아졌어요. 그래서 요즘 딥러닝 기반 자연어처리에서는 굳이 한국어 형태소 분석기를 쓰지 않고도 웬만큼 잘 해내고 있어요. 그래도 한국어 형태소 분석기가 여전히 한국어의 언어 공학, 언어 산업의 생태계에서 여전히 중요한 역할을 하고 있는데, 그중 하나는 검색 엔진입니다.

네이버 같은 검색 엔진에서 사용자가 쿼리를 던졌을 때 적절한 문서를 뽑아오려면, 엄청나게 많은 한국어 웹문서를 인덱싱을 해야 하는데 그 전에 형태소 분석을 해야 돼요. 구글에서 만든 BPE라든지 Wordpiece 같은 범용의 분석기로는 한국어를 제대로 분석하지 못하기 때문에 그걸 쓸 수는 없습니다. 그래서 한국의 산업계에서 형태소 분석기는 여전히 중요합니다.

Q . *국어학, 언어학에서 AI 기술을 활용한 개선한 사례와 언어학자와 공학자 사이에 교류는 어떤지요*

A . 열망은 대단합니다. 온 세계에서 AI를 떠드니까 귀가 솔깃하고 '나도 좀 알아야 될 것 같아', '내가 하는 일에 이용할 수 없을까?' 하는 생각을 당연히 하죠. 그런데 인공지능 기술을 자기 연구에 써먹으려면 코딩도 좀 하고 인공지능 기술에 대한 이해가 어느 정도는 있어야 되는데, 그 정도 수준에 올라 있는 사람은 아주 적은 편이에요.

제가 느끼기에 옛날 GOFAI(good old fashioned AI) 시대보다 교류가 줄어든 것 같아요. 언어 공학 쪽에서 언어학 도메인 지식이 중요하지 않다는 생각이 확산돼 있기 때문에, 언어학자들한테 같이 하자고 손을 내미는 일이 없어진 것 같아요. 언어학자들은 손을 내밀고 같이 하고 싶은데, 지금 AI 쪽 사람들은 사회적으로 수요가 많고 바빠서 잘 안 되는 것 같아요.

Q . *BERT 같은 모델은 인풋 문장을 잘 이해하고, GPT 같은 모델은 그럴 듯하게 문장을 만들어내고, 블렌더봇이나 구글의 람다 같은 건 놀라울 정도로 대화가 자연스럽게 이루어지고 있습니다. 대화 시스템에 대한 지금의 AI 접근 방식을 평가하신다면?*

A . 아까 얘기의 연장선인데요, 대화 시스템이 안에 가지고 있는 거는 엄청나게 큰 신경망이고 그 안에 가중치들이, 즉 숫자들이 엄청나게 들어 있지만, "미국의 수도는 워싱턴DC다"라는 라든지 "임진왜란은 1592년에 일어났다" 같은 지식, 상식이 명시적인 형태로 표상되어 있는 건 아니잖아요.

그래도 사용자가 "임진왜란은 몇 년에 일어났어?"라고 물어보면 웹 서치를 한다거나 학습 할 때 사용했던 코퍼스(corpus)에 포함된 관련 문장들을 바탕으로 신경망이 작동을 해서 그럴 듯한 답을 내놓기는 합니다. 하지만 이렇게 명시적인 지식 표상 없이 대화를 흉내내는 시스템은 옛날의 ELIZA와 비슷합니다.

ELIZA보다 훨씬 정교하고 복잡한 시스템이지만 기본적인 성격은 비슷합니다. 그저 흉내 내는 것이고 눈속임일 뿐이라는 거죠.

Q . *예를 들어, '게리 마커스(Gary Marcus)' 뉴욕대학교(New York University) 심리학·신경과학 교수의 비판도, 그리고 Q&A 시스템이 바보 같은 엉터리 대답을 내놓는 경우도, 이런 문제점을 고려할 때, 현재의 방법대로 갔을 때 우리가 정말 원하는 일반적인 대화 시스템을 만들 수 있는지요*

A . 잘 안 될 겁니다. 지금은 인간이 가진 지식 체계를 정면으로 맞닥뜨려서 기계로 표상하려는 시도를 하는 게 아니라 편법을 써서, 그런 표상이 전혀 없는데도 마치 가지고 있는 것처럼 눈속임하는 식이고, 그런 편법을 점점 더 정교하게 만드는 쪽으로 기술이 발전하고 있다고 생각됩니다.

Q . *제프 호킨스의 <천개의 뇌>라는 책에서, 우리 뇌 안에 있는 피질 기둥이라든가 소기둥 같은 것들이 세계 모델을 만들어내고 그들이 투표를 하면서 정보 처리를 하고 있는데, 하드웨어까지 사람의 뇌처럼 그렇게 만들지 않고서는 절대로 지능의 문제를 풀지 못할 것이라고 주장합니다. 지금의 방식으로는 절대로 지능을 못 만들어낸다고 강력하게 비판하는 거죠. 과거의 GOFAI에서는 콘셉트, 규칙, 프레임, 스키마 등 여러 가지 지식 표현의 방법을 동원했는데, 그걸 접목시킬 수는 없을까요?*

A . 엔지니어링 접근법 중에 자연으로부터 배우자는 방법이 있는데, 자연을 똑같이 모사할 필요까지는 없습니다. 하지만 핵심적인 부분은 참고할 필요가 있죠. 비행기를 만들 때 새처럼 날개를 퍼덕거리는 방법을 굳이 택할 필요는 없지만 그래도 날개 같은 게 있긴 있어야 부력을 받아가지고 뜰 수 있을 테니까요. 비행기는 새를 똑같이 모사한 건 아니지만 그래도 새의 중요한 특징을 받아들인 거잖아요.

자연을 참고하여 엔지니어링을 할 때, 너무 하드웨어 디테일 구석구석까지 다 따라서 만들려고 하는 한쪽 극단적인 입장이 있다면, 반대쪽 극단에는 '자연계의 작동 방식을 너무 신경 쓸 필요 없다. 완전히 다른 방식으로 해도 된다'는 반대쪽 극단이 있는데, 그 사이에서 적절한 중용의 지점을 찾아야 될 것 같아요.

인간이 언어를 처리할 때 세계에 대한 방대한 지식 체계 등을 활용하는데, 자연어처리 시스템이 이걸 어떻게 배우고 어느 정도까지 모사할 것인가가 문제죠.

우리 뇌에 '미국의 수도는 워싱턴 DC다' 같은 명제가 어디에 명시적으로 있는 건 아니죠. 뇌의 하드웨어를 보면 아무런 의미가 없는 것 같은 신경 세포들 사이의 여러 전기적인 상호작용일 뿐이잖아요.

그렇지만 그 하드웨어의 로우 레벨에서 어떻게든 상위 레벨로 올라가서 지식 같은 게 창발(emerge)하는 거잖아요. 그걸 창발하게 만들려면, 기계도 인간의 뇌세포와 똑같이, 하드웨어 디테일까지 다 똑같이 만들어야 된다고 생각하는 건 전자의 극단적인 입장입니다.

그런 지식을 명시적으로 표상하는 일 없이 그냥 신경망의 숫자놀음만 가지고 흉내내는 방식에 만족하는 사람은 후자의 극단적인 입장입니다. 저는 게리 마커스처럼 중용의 지점이 좋다고 생각해요. 옛날의 GOFAI 시절에 했던 방식 그대로 하자는 얘기는 아니고, 딥러닝이 잘 하는 부분은 딥러닝에게 맡기고 잘 못하는 부분은 GOFAI의 지식 표현 등을 활용해서 접목하자는 겁니다.

Q . *정부가 한 3~4년 동안 예산을 많이 들여서 굉장히 많은 데이터를 생성해 왔죠. 특히 언어 데이터세트 가운데 부족한 점이 있다면?*

A . 인공지능을 위한 데이터세트 구축 사업이 이렇게 대규모로 벌어지고 있는 거는 정말 바람직한 일이고 좋은 일인데요. 아쉬운 점은, 어떤 데이터가 필요하고 그 데이터에 어떤 정보가 들어가야 되는지는 데이터를 사용하는 엔진 그리고 그 엔진에 대한 기술의 발전과 함께 가는 거잖아요.

인간의 방대한 지식 체계를 단지 흉내 내는 게 아니라 그걸 좀 더 정면 대결을 해서 제대로 표상 할 수 있는 방법을 모색해야 되는데, 인공지능 엔진이 그걸 더 잘 처리할 수 있도록 하려면 현재 엔진의 아키텍처 그대로가 아니라 상당히 큰 변화가 있어야 될 겁니다.

현재의 인공지능 데이터 구축 사업은 인간의 방대한 지식 체계가 딥러닝에 결합된 걸 전제로 한다기보다, 그냥 현재의 숫자 놀음인 신경망만 있다고 전제하고 있습니다. 이 숫자밖에 없는 이 신경망한테 집어넣어 줘서 결과가 잘 나오게끔 하기 위해 데이터를 더 다변화하고 잘 만들자는 생각인 거죠.

현재의 단순한 숫자 놀음인 신경망에 그치는 게 아니라 거기에 방대한 지식 체계를 결합시키는 기술에 돌파구가 생겨서 지금과는 사뭇 다른 인공지능 엔진이 나올 텐데, 그런 인공지능 엔진이 필요로 하는 데이터의 성격은 지금과는 사뭇 달라질 수도 있을 겁니다.

Q . *이 인공지능 데이터세트 과제는 현재의 기술을 전제로 할 수밖에 없겠죠. 그럼에도 불구하고 딥러닝을 위한 한국형 데이터세트를 만들 때 현재 빠진 범주는 없나요?*

A . KorQuAD나 KLUE 같은 게 나오면서 인공지능 언어 모델의 성능을 평가하기 위한 평가 세트가 좀 생겨서 갈증이 약간 해소되긴 했지만, 평가 세트는 여전히 부족하다는 게 많은 사람들의 공통된 인식입니다.

그리고 자연어 처리의 다운스트림 과업이 굉장히 다양한데 그것들을 모두 다 빠짐없이 다 망라하고 있는 건 아니고요. 자연어 처리의 다운스트림 과업도 더 세분화되고 특화되는 경향이 있으니까, 그런 수요를 충족시켜 나가야겠죠.

Q . *과학기술정보통부가 주관하고 한국지능정보사회진흥원(NIA)이 추진하는 '2022년도 인공지능 학습 데이터 구축사업'으로 메타의 AI 챗봇 블렌더봇(BlenderBot)의 한국형 블랜더봇 데이터세트를 만들기 위한 과제의 자문위원으로 참여하고 계시는데 어떤 자문을 하시는지요*

A . 블렌더봇의 특징이 메모리 내지 히스토리, 과거에 대화했던 내용들을 기억하는 거, 사용자의 기분, 감정 상태를 잘 파악해서 거기에 맞게 대화하는 거, 블렌더봇 자신의 페르소나, 정체성을 일관성 있게 유지하는 거죠. 이중 사용자의 감정 파악 및 대처에 관심이 많이 있습니다. 자연어처리에서도 감정분석 연구들이 있습니다만 긍정-부정의 바이너리한 분류에 너무 많이 치우쳐 온 것 같아요.

우리 인간이 대화할 때 대화 상대방의 지금 기분 상태가 어떤지를 어떻게 파악하는지를 생각을 해보면, 얼굴 표정도 보고 말투, 말의 음향적인 특징도 보고, 그래도 확신하지 못할 때가 많이 있어서 상대방의 기분 상태를 탐지하기 위한 탐색 질문도 던지잖아요? 블렌더봇도 그런 탐색 질문을 기술적으로 잘 하면 좋을 것 같아요.

Q . *사람의 감성 상태와 공감하는 수준의 챗봇을 만든다는 거잖아요. 거기에서 언어학자의 역할이 있나요? 심리학자, 인지심리학자 같은 사람들의 역할이 더 큰 거 아닌가요?*

A . 그런 부분은 언어학과 심리학의 경계가 좀 모호한데요. 학제적인 심리언어학(psycholinguistics) 같은 분야도 있고, 담화 분석, 대화 분석 같은 분야도 있어요. 사회학자들도 참여합니다. 우리 인간들끼리 대화를 할 때 아주 엄격한 문장 내의 문법 규칙 같은 것은 아니지만 그렇다고 해서 랜덤한 것도 아니고 상당히 느슨하긴 하지만 대화의 규칙이랄지 경향성 같은 게 있거든요.

처음 대화를 시작할 때는 어떤 문장이 나올 확률이 높고 어떤 대답을 할 확률이 높고, 그런 경향성 말이죠. 그래서 담화 문법(grammar of discourse), 대화 규칙(conversational rule) 같은 것들에 대한 연구들이 있습니다.

챗봇과 인간이 대화할 때 인간과 인간의 대화하고 똑같지는 않겠지만 그래도 상당히 비슷하게 모델링을 해야 될 거 아니에요. 블렌더봇은 사용자 상대방의 감정에 공감하는 챗봇을 만들려고 하는데 공감하기 전에 상대방의 지금 감정 상태가 어떤지를 우선 정확하게 파악해야 공감도 정확하게 할 것입니다.

그러나 상대방의 감정을 정확하게 파악하는 일이 한 번에 되는 게 아니라, 우리 인간의 경우에도 계속 탐색 질문을 던지고 하면서 조금씩 정제해 나가는 과정이 상당히 있다는 겁니다.

더불어 담화 분석, 대화 분석 쪽의 연구 중에 재밌는 연구들이 많은데 특히, 젠더 간의 차이도 그중 하나입니다. 여러 사람이 참여하는 대화에서 다른 사람이 말하고 있고 아직 끝나지 않았는데 중간에 말을 가로채는 현상이 여성보다 남성이 훨씬 높고, 누가 뭐를 말했을 때 특별한 정보적 가치는 적지만 "아, 그래?", "아", "아이고" 이런 식의 맞장구 쳐주는 건 여성들이 훨씬 높다고 합니다.

이런 담화 분석, 대화 분석의 연구 성과를 참고하면 좋겠죠? 블렌더봇이 여성의 페르소나를 가지고 말할 때는 상대방의 말에 맞장구도 많이 쳐줘야겠죠.

Q . *담화의 법칙, 대화의 법칙에 대해 여러 분야에서 학제적으로 연구를 많이 해 왔습니다만, 그 많은 결과들을 블렌더봇에 어떻게 반영해야 할까요? 알려준다고 해서 기계가 그걸 잘할 수 있을지*

A . 대화 데이터를 수집할 때 전략이 있으면 더 좋겠다는 거죠. 그러니까 상대방의 감정 상태를 파악하기 위한 탐색 질문을 던지는 부분이 들어가 있는 대화, 그렇게 일단 어느 정도 파악한 다음에 공감하고 맞장구 쳐주는 문장이 많이 들어간 대화, 이런 걸 신경 써서 많이 수집을 해서 구축을 하면 나중에 도움이 되죠.

Q . *이루다가 '이루다2.0'으로 업그레이드돼서 다시 나왔는데, 첫 번째 등장할 때만큼의 이슈는 없는 것 같아요. 그리고 케어콜이라고 노인들과 대화하는 시스템이 있죠. 둘 다 굉장히 특정 도메인에서의 챗봇이라면 보다 넓은 일상 대화 챗봇을 만들려면 특정 도메인의 대화 시스템과 일반 대화 시스템은 많이 다르겠죠?*

A . 이루다의 경우 완전히 오픈 도메인으로 모든 영역을 다 커버한다기보다 트레이닝 데이터로 삼은 것 중에 남녀 간의 연애나 썸타기 대화가 핵심 데이터세트였고 그걸로 트레이닝을 했기 때문에 그런 대화에 능숙하고 젊은이들의 관심을 끌었던 것도 바로 그런 점에서의 재미였다고 생각합니다.

완전히 열려 있는, 모든 영역을 다 커버할 수 있는 챗봇을 만들려면 아까 얘기로 돌아가서 세상에 대한 지식 기반과 현재의 딥러닝이 결합되어야 가능할 거고, 그게 없이는 계속 더 영리한 ELIZA 정도라는 한계는 계속 지닐 것 같고요.

연애에 특화된, 헬스케어에 특화된, 특허나 의료나 법률에 특화된, 이런 식의 도메인에 한정된 챗봇을 더 빨리 만들 수 있고 퍼포먼스도 잘 나오고 산업적으로 현실성 있는 것 같고, 좀 더 범용의 챗봇은 장기적인 관점에서 지식 기반과 결합을 지금보다 더 중시해야 되지 않을까 생각합니다.

Q . *마지막으로 계산 언어학(computational linguistics) 같은 분야에서 데이터를 통한 언어 분석, 언어 연구가 꾸준히 진행되고 있습니다. 사회과학도 그쪽으로 많이들 하고 있는데 언어학, 국어학을 선택한 후배들한테 앞으로의 연구를 위해서, 어떤 조언을 해주신다면?*

A . 다른 학문 분야도 그렇겠지만 국어학도 역사와 전통이 꽤 쌓여서, 학생 입장에서는 축적된 과거 선배 학자들의 연구 성과에 짓눌리는 압력이 꽤 큰 것 같아요. 그 것을 내가 제대로 학습을 해서 섭렵을 해야 조금이라도 더 보탬이 될 수 있는 연구를 할 수 있다는 거죠. 너무 성실하고 착한 모범생 같은 태도이긴 한데 그게 너무 지나치면 문제의식이나 연구 질문도 선배 학자들이 던졌던 것과 똑같거나 거의 비슷한 것만 자꾸 던지고 엇비슷한 논문만 자꾸 쓰는 경향이 있습니다.

선배 학자들의 연구를 너무 무시하고 해도 문제겠지만 거기에 너무 짓눌려도 문제인데 제가 느끼는 한국의 인문학의 전반적인 경향은 너무 짓눌리는 쪽인 것 같아요. 학문도 태평성대 같은 시대가 있고 격변의 시대가 있고 한데, 격변의 시대에는 선배 학자들의 연구 성과에 짓눌리기 보다는 발랄하게, 선배 학자들과 다른 연구 질문을 던지고 새로운 모색을 많이 하는 게 시대의 분위기에도 맞고, 자기가 학계에 기여할 수 있는 부분도 더 많다고 생각합니다.

저하고 다른 태도의 교수님들도 많이 있어요. 기본기를 강조하고 논문에서 선행 연구 정리하고 성실하게 정리하는 것을 중요시하는 교수님도 있는데, 저는 좀 반대 성향으로 "선행 연구 좀 몰라도 괜찮다. 너 스스로 재밌다고 생각하는 것, 선배 학자들이 던지지 않은 새로운 연구 질문을 던져라"라고 주로 얘기를 많이 합니다. 국어학에서 과거에 많이 하지 않았던 새로운 연구 질문을 탐구할 때는 계산학적인 스킬이 있으면 더 연구를 잘할 수 있는 부분이 많습니다.

최근에 대규모 코퍼스도 있으니까 코딩을 조금만 하면 거기서 자기가 원하는 정보들을 추출할 수 있습니다. 자기가 코딩이나 AI 쪽에서 잘할 수 있다고 생각되면 아예 자기 연구의 중심이 이쪽으로 더 올 수도 있는 거고요. 어떤 사람은 그냥 그 중간쯤에서 다양하게 학제적인 연구를 할 수도 있겠죠.

정한영 기자 hyjung@aitimes.kr