한국어, 영어 번역 기술

기계번역기술

<https://www.sciencetimes.co.kr/news/%EA%B8%B0%EA%B3%84%EB%B2%88%EC%97%AD-%EC%96%B4%EB%94%94%EA%B9%8C%EC%A7%80-%EC%A7%84%ED%99%94%ED%96%88%EB%82%98/>

자동번역 또는 기계번역이라고 불리는 기술은 크게 규칙기반(RBMT, Rule-Based Machine Translation)과 통계기반(SMT, Statistical Machine Translation), 이를 합한 하이브리드(RBMT+SMT) 기반으로 구분된다. 이는 인공신경망 번역(NMT) 기술로 발전하고 있다.

인공신경망 번역(NMT) 기술 원리와 기존의 통계적 기계번역(SMT) 기술 비교

[[한국어와 인공지능]⑦ 네이버 김준석 연구원 "인공신경망 번역 성능 2배 이상‥'파파고' 연계 서비스 쏟아진다" - Chosunbiz > 테크 > ICT/미디어](https://biz.chosun.com/site/data/html_dir/2016/11/08/2016110802495.html)

기계가 구절 단위로 번역해 문장을 만드는 대신, 문장에 담긴 모든 정보, 네트워크를 통째로 학습하게 된다. 그러면 기계가 한 번도 학습한 적이 없는 문장을 입력하더라도 네트워크를 통과시켜 자동으로 번역해준다.

인공신경망을 이용한 기계번역은 대용량의 학습 데이터와 몇 차원 벡터로 표현할 것인지의 정보, 학습에 필요한 변수 정보만 사람이 지정해주면 기계가 알아서 번역법을 학습한다. 기존의 기계번역 방식이 어순과 어법 등 각종 규칙(특징 정보)를 사람이 지정해주던 것과 비교하면 큰 장점이 있는 셈이다.

인공신경망 기계번역에서는 입력 문장이 들어오면, 이 문장을 가상 공간의 하나의 점을 의미하는 벡터(좌표값)와 같은 숫자로 변환하고, 학습 데이터를 통해 배운 가중치를 더해 번역을 수행한다. 벡터에는 단어와 구절, 어순 정보까지 다 들어 있다. 인공신경망 기계번역 방식이 문맥을 그만큼 많이 본다는 뜻이다. 이 때문에 번역도 훨씬 자연스럽다.

문장 정보를 숫자로 바꾼 벡터에 대한 설명도 포함되어 있다.

<https://www.oss.kr/info_techtip/show/65f88554-1882-4af9-8036-6490d3314c65>

<https://brunch.co.kr/@kakao-it/155>

Kakao Brain <https://www.kakaobrain.com/blog/36>

네이버 파파고 <https://www.ncloud.com/product/aiService/papagoNmt>

Seq2Seq 오픈소스 / 설명

<https://wiserloner.tistory.com/1316>

<https://tutorials.pytorch.kr/intermediate/seq2seq_translation_tutorial.html>

NMT 오픈소스 <https://opennmt.net/>

Tutorial

<https://github.com/Parkchanjun/OpenNMT-Colab-Tutorial?fbclid=IwAR10spuUNAVaVSjSeNuKkRLW396f8KJxObMY6Y6F95E7akTiADXD40avUh8>

Korean-English NMT 오픈소스

<https://github.com/haven-jeon/ko_en_neural_machine_translation>

<https://github.com/dlcjfgmlnasa/NMT-Korean-To-English>

Papago NMT API Reference <https://developers.naver.com/docs/nmt/reference/>

예제코드 <https://developers.naver.com/docs/nmt/examples/>

한국어-영어 번역(병렬) 말뭉치 AI 데이터 <https://aihub.or.kr/aidata/87>