

관용구 기계번역을 위한 한-영 데이터셋 구축 및 평가 방법

최민주[○]

국민대학교 컴퓨터공학과

mjchoi0831@gmail.com

A Data Set and Evaluating Method for Korean-English Idiom Machine Translation

Min-Joo Choi[○]

Dept. of Computer Science, Kookmin University

요 약

관용구는 둘 이상의 단어가 결합하여 특정한 뜻을 생성한 어구로 기계번역 시 종종 오역이 발생한다. 이는 관용구가 지닌 함축적인 의미를 정확하게 번역할 수 없는 기계번역의 한계를 드러낸다. 따라서 신경망 기계 번역(Neural Machine Translation)에서 관용구를 효과적으로 학습하고 번역 결과를 평가하려면 관용구에 특화된 번역 쌍 데이터셋과 평가방법이 필요하다. 본 논문에서는 관용구 기계번역을 위한 한-영 번역 쌍 데이터셋을 구축하는 방법과 관용구 번역 결과를 평가하기 위해 블랙리스트를 구축하는 방법을 제안한다.

1. 서 론

관용구는 둘 이상의 단어가 결합하여 특정한 뜻을 생성한 어구로, 단어 각각의 뜻을 직역하여 조합한 것과는 다른 의미를 가진다. 예를 들어 ‘눈이 높다’라는 관용구의 경우 ‘눈’, ‘높다’라는 2개의 단어가 조합되었지만 실제 관용구의 의미는 ‘좋은 것을 찾다’ 또는 ‘안목이 높다’이므로 관용구를 구성하는 단어들의 뜻과 관용구의 뜻은 의미적인 관련이 없다. 따라서 관용구를 번역할 때 관용구를 구성하는 단어 그대로 번역할 경우 그림 1의 예와 같은 오역이 발생한다.

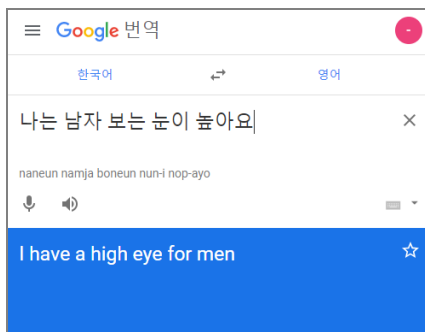


그림 1 관용구 포함 문장 기계번역 결과

이는 함축적인 의미를 지닌 관용구를 정확하게 번역할 수 없는 기계번역의 한계를 드러낸다. 이러한 한계를 극복하기 위해서는 기계번역 모델이 관용구를 효과적으로 학습하고 번역 결과를 평가할 수 있도록 관용구 번역에 특화된 데이터셋과 평가방법이 필요하다.

본 논문에서는 관용구 기계번역을 위한 한-영 번역 쌍 데이터셋 ‘KISS¹’를 구축하는 방법을 제안한다. 이어서 기계번역 모델을 이용한 관용구 번역 결과의 품질을 평가하기

위해 블랙리스트를 이용하는 방법을 소개하고, KISS를 이용하여 블랙리스트를 구축하는 방법을 제안한다.

2. 관련 연구

신경망을 이용하는 기계번역 모델 Neural Machine Translation(이하 NMT)이 개발되기 이전에는 구-대-구 방식으로 관용구의 구조 정보를 이용하여 번역하였다. 특히 사전 기반 기계번역 모델을 이용한 과거에는 문장 내에 관용구 사전에 존재하는 관용구가 포함되어 있으면 관용구로 번역하는 방법을 이용하였다.[1]

그러나 관용구 사전을 이용하는 방법은 구축하는 데에 많은 시간과 노력이 드는 것은 물론 문장 단위로 학습하고 번역하는 NMT에는 적합하지 않다.

따라서 NMT가 관용구를 학습하기 위해서는 관용구가 포함된 다량의 문장 데이터셋이 필요하므로 관용구 번역 쌍 데이터셋을 구축하기 위한 연구와[2], NMT를 이용해 관용구를 번역한 결과를 효과적으로 평가하기 위해 블랙리스트를 구축하는 연구가 이루어졌다. [3]

본 논문에서는 선행연구를 참고하여 공개된 한-영 번역 쌍 말뭉치로부터 관용구가 포함된 번역 쌍을 추출하여 관용구 학습 데이터셋을 구축하는 방법에 대해 서술한다. 이어서 구축한 데이터셋을 이용하여 관용구가 포함된 문장의 번역 결과를 평가할 수 있도록 블랙리스트를 생성하는 방법을 설명한다.

3. KISS : 관용구 포함 한-영 번역 쌍 데이터셋 구축

NMT는 다량의 문장 번역 쌍을 이용하여 학습하므로 NMT가 관용구를 학습하기 위해서는 관용구가 포함된 다량의 문장 번역 쌍이 필요하다. 3장에서는 공개된 한-영 번역 쌍 말뭉치로부터 관용구가 포함된 번역 쌍을 추출하여 데이터셋을 구축하는 방법에 대해 설명한다.

¹ KISS : Korean-english Idioms in Sentences dataSet
<https://github.com/Judy-Choi/KISS>

3.1 관용구 수집

사전에 수록된 한국어 관용구를 수집하기 위해 표준국어대사전 온라인 사이트²에서 총 3,887개의 관용구 목록을 다운로드하였다.

3.2 관용구를 포함하는 한-영 번역 쌍 수집

수집한 한국어 관용구가 포함되는 한-영 문장 번역 쌍을 말뭉치로부터 추출한다. 말뭉치로 AI Hub³에서 제공하는 한국어-영어 번역(병렬) 말뭉치 AI 데이터를 사용하였으며, 해당 말뭉치는 문어체 한영 번역 110만 문장(뉴스 80만, 정부 웹사이트 콘텐츠 10만, 조례 10만, 한국문화 10만), 구어체 한영 번역 50만 문장(구어체 40만, 대화체 10만)으로 총 160만개의 한영 번역 문장으로 이루어져 있다.

먼저 160만개의 한영 번역 문장으로부터 3,887종류의 한국어 관용구가 포함된 문장 쌍을 형태소 비교 및 수작업을 거쳐 추출 시도한 결과 430개의 한국어 관용구를 포함하는 18,808 개의 문장을 추출하였다.

이어서 관용구의 종류가 고르게 분포한 데이터셋을 구축하기 위해 동일한 관용구가 포함된 문장의 개수를 4개 이상 40개 이하로 제한하여 최종적으로 표1 과 같은 형태로 420개의 한국어 관용구를 포함하는 7,500 개의 한-영 번역 쌍 데이터셋 KISS를 구축하였다.

표 1 관용구를 포함하는 한-영 번역 쌍 예시

관용구	눈이 높다
한국어 원문	나는 여자 보는 <u>눈이 높아요.</u>
영어 번역 쌍	I have <u>high standards</u> for woman.
관용구	마침표를 찍다
한국어 원문	아버지의 명예회복을 위한 김지훈의 기나긴 여정이 마침내 <u>마침표를 찍었다.</u>
영어 번역 쌍	Kim Ji-hoon's long journey to restore his father's honor has finally <u>come to an end.</u>
관용구	총대를 메다
한국어 원문	기획재정부가 혁신성장 관련 규제 완화에 <u>총대를 멘다.</u>
영어 번역 쌍	The Ministry of Economy and Finance <u>takes charge of</u> the deregulation related to innovative growth.
관용구	화촉을 밝히다
한국어 원문	중국 출신 할리우드 스타 장쯔이(章子怡, 28)가 현 남자친구인 미국의 억만장자 비비 네보(42) 와 내년 <u>화촉을 밝힌다.</u>
영어 번역 쌍	Chinese Hollywood star Zhang Ziyi (28) will <u>marry</u> her current boyfriend, U.S. billionaire Aviv Nevo (42), next year.

² <https://stdict.korean.go.kr/>

³ <http://www.aihub.or.kr/>

4. 블랙리스트 평가 방법

3장에서 구축한 KISS를 이용해 관용구 번역 블랙리스트를 생성하여 NMT 번역 결과를 평가할 수 있다. 4장에서는 번역 품질 평가 방법 중 블랙리스트를 이용하는 방법을 소개하고, KISS를 이용하여 블랙리스트를 구축하는 방법에 대해 설명한다.

4.1 관용구 블랙리스트를 이용한 번역 평가

관용구의 오역 여부를 판별하고 번역 품질을 평가하기 위해 Shao Yutong, et al[3] 이 블랙리스트를 이용하는 방법을 제안하였다. 블랙리스트 평가 방법은 직역으로 인한 번역 오류를 탐지할 수 있도록 고안되었으며, 단어-대-단어 구조의 관용구를 번역한 결과로부터 오역 여부를 판별할 수 있다.

표2 는 관용구의 한-영 번역 결과를 평가하기 위한 블랙리스트의 예시로, 관용구를 한-영 번역한 결과에 블랙리스트 단어가 1개 이상 포함되어 있으면 오역으로 간주한다. 예를 들어 표3 과 같이 ‘눈이 높다’ 라는 관용구를 영어로 번역한 문장에 블랙리스트 단어 ‘eye’ 가 존재하면 해당 문장을 오역된 문장으로 간주한다.

표 2 한-영 관용구 블랙리스트 예시

관용구	블랙리스트
꼬집어 말하다	nip pinch twitch
눈 높다	eye
운을 떴다	lucky
유명을 달리하다	famous

표 3 블랙리스트를 이용한 오역 탐지

관용구	눈이 높다
블랙리스트	eye
한국어 원문	나는 여자 보는 <u>눈이 높아요.</u>
영어 번역 쌍	I have <u>high standards</u> for woman.

4.2 블랙리스트 구축

관용구를 구성하는 각 단어의 의미와 KISS 의 오역 결과를 참고하여 표4 및 표5 와 같이 블랙리스트에 단어를 추가한다. 하나의 관용구 당 블랙리스트 단어는 1개 이상 최대 5개로, 총 420개의 관용구에 대한 블랙리스트를 구축하였다.

표 4 단어의 의미를 이용한 블랙리스트 단어 추출

관용구	<u>꼬집어</u> 말하다
블랙리스트	nip pinch twitch
관용구	<u>눈</u> 높다
블랙리스트	eye

표 5 오역결과를 이용한 블랙리스트 단어 추출

관용구	운을 <u>떴다</u>
한국어 원문	정부가 노인 연령 기준을 높이는 방안에 대해 <u>운을</u> <u>떴다.</u>

영어 번역 쌍	The government has been lucky about ways to raise the criteria of age for senior citizens.
블랙리스트	lucky
관용구	<i>유명을 달리하다</i>
한국어 원문	오키나와 현지사인 오나가 다케시가 지난 8일 췌장암으로 유명을 달리했다 .
영어 번역 쌍	Takeshi Onaga, an incumbent Governor of Okinawa, became famous for pancreatic cancer on the 8th.
블랙리스트	famous

5. 평가

5.1 데이터셋 오역 평가

3절에서 구축한 한-영 번역 쌍 데이터셋 KISS에는 다수의 오역 결과가 포함되어 있으므로 블랙리스트 평가 방법을 이용하여 최종적으로 오역 없이 275개 관용구를 포함하는 3,461개 번역 쌍으로 이루어진 데이터셋을 추출하였다.

5.2 기계번역 서비스의 번역 품질 평가

기계번역 서비스의 번역 품질을 평가하기 위해 5.1절에서 추출한 데이터셋의 한국어 문장 3,461개를 국내 주요 기계번역 서비스를 이용하여 영어로 번역하고, 번역 결과에 블랙리스트 평가 방법을 적용한다.

표6은 블랙리스트 평가 결과로, 관용어 블랙리스트 단어가 1개 이상 문장 번역 결과에 존재하는 경우 해당 문장을 ‘블랙리스트 탐지’로 분류한다. 번역 정확도는 전체 문장 중 블랙리스트 미탐지된 문장의 비율로, 국내 주요 기계번역 서비스의 번역 정확도 및 BLEU 점수를 측정하였다.

표 6 기계번역 서비스 번역 품질 평가

	Google 번역 ⁴	Naver Papago ⁵	Kakao i 번역 ⁶
블랙리스트 탐지	1,179	1,093	1,049
블랙리스트 미탐지	2,282	2,368	2,412
번역 정확도 (%)	65.93	68.41	69.69
평균 BLEU 점수	30.04	13.47	33.83

Naver Papago의 경우 평균 BLEU(Bilingual Evaluation Understudy)[4] 점수가 현저히 낮는데, 번역 정확도는 다른 번역 서비스의 정확도와 비슷한 수치이므로 번역이 잘못된 것으로 단정지을 수는 없다. 예를 들어 표7과 같이 기계번역이 데이터셋의 영어 번역 쌍보다 한국어 원문을 더욱 충실하게 번역하였을 경우 BLEU 점수가 낮게 나타날 수 있다.

표 7 번역 품질과 BLEU 점수가 비례하지 않는 예시

관용구	<i>가닥을 잡다</i>
한국어 원문	당초 강 위원장은 국정감사에서 이 문제가 불거지자 “감사원 감사를 받겠다”며 버텼지만 즉각 사퇴로 가닥을 잡았다 .
영어 번역 쌍	At the beginning, Representative Kang said, “I will be audited by the auditor,” when the matter was raised in the state audit.
Naver Papago 번역 결과	Initially, Kang endured the issue during a parliamentary audit, saying he would undergo an audit by the Board of Audit and Inspection, but he immediately decided to step down.
BLEU 점수	4.7930e-76

6. 결론

본 논문에서는 NMT가 관용구 번역을 학습할 수 있도록 관용구가 포함된 한-영 문장 번역 쌍 데이터셋 KISS를 구축하는 방법과, NMT 번역 결과 품질을 평가하기 위해 KISS를 이용하여 블랙리스트를 구축하는 방법을 제안하였다.

제안한 방법을 이용하여 420개의 한국어 관용구를 포함하는 7,500개의 한-영 번역 쌍 데이터셋 KISS를 구축하고, 블랙리스트를 이용해 오역을 제외한 결과 최종적으로 275개 관용구를 포함하는 3,461개 번역 쌍으로 이루어진 데이터셋을 추출할 수 있었다.

또한 구축한 블랙리스트를 이용하여 기계번역 서비스의 번역 정확도를 계산한 결과 BLEU 점수보다 정확하게 번역 품질을 평가할 수 있었다.

향후 본 연구는 NMT에서 관용구뿐만 아니라 속어, 속담과 같이 함축적인 의미를 지닌 어휘를 학습할 수 있도록 데이터셋을 구축하는 연구에 응용될 수 있으며, NMT 번역 품질을 평가하는 연구에 응용될 수 있다.

특히 본 연구에서 제안한 방법은 한-영 번역 쌍뿐만 아니라 한국어-일본어, 한국어-프랑스어와 같은 다른 종류의 언어 쌍에도 적용될 수 있으므로 폭넓게 응용될 수 있다.

참고문헌

- [1] 이호석, 김영택. 영어-한국어 기계번역을 위한 언어와 속어 트랜스퍼 사전. (구) 정보과학회논문지. 20.7: 976-987. 1993.
- [2] Fadaee Marzieh, et al. Examining the tip of the iceberg: A data set for idiom translation. arXiv preprint arXiv:1802.04681. 2018.
- [3] Shao Yutong, et al. Evaluating machine translation performance on Chinese idioms with a blacklist method. arXiv preprint arXiv:1711.07646. 2017.
- [4] Papineni Kishore, et al. BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics. 2002.

⁴ <https://translate.google.com/>

⁵ <https://papago.naver.com/>

⁶ <https://translate.kakao.com/>