


신경망 기반 기계번역(NMT)


-Papago, Google 번역 등이 좋아진 이유에 관하여

2015312452 이준규

목차



1. 기계번역의 발전



2. 기계번역의 역사



3. NMT의 원리



4. 마치며

0. 주제선정이유



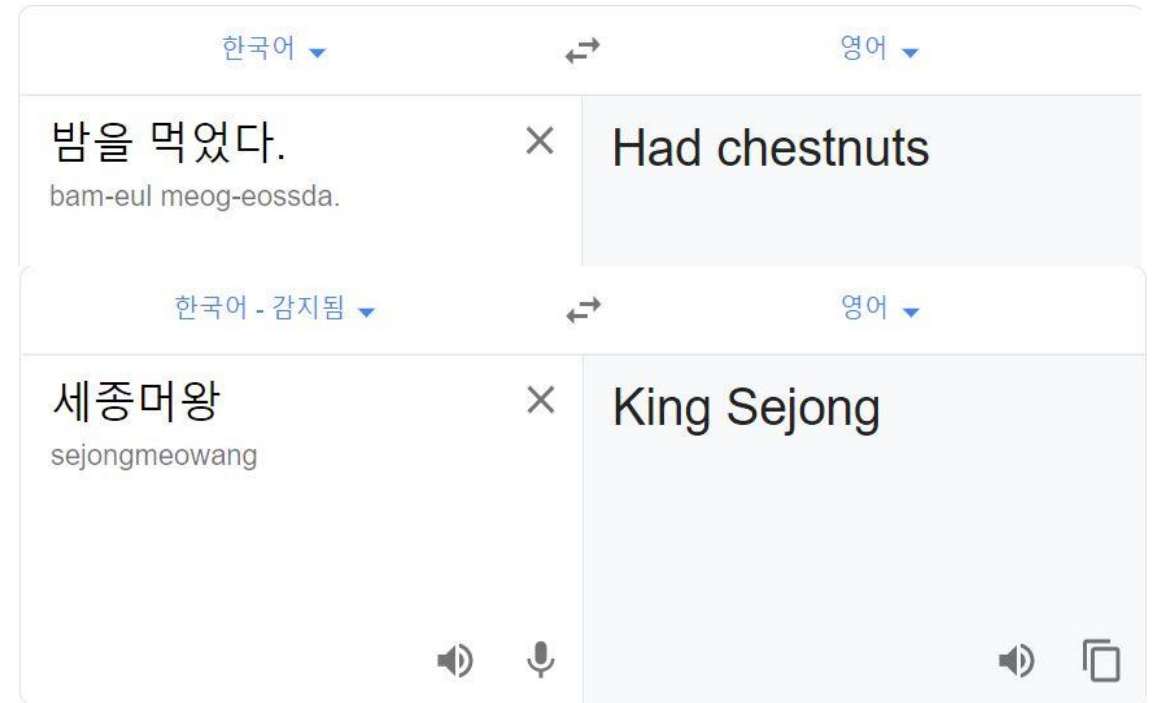
1. 기계번역의 발전

rnasterpiece

○ 2016년 이전



○ 최근



✓ 2016년 말 ~ 2017년부터 번역기의 성능이 비약적으로 상승.

1. 기계번역의 발전

○ Why?

네이버, 모바일 통역앱 파파고에 '인공신경망 번역' 적용

매일경제 - 2016. 10. 21.

네이버, 모바일 통역앱 파파고에 '인공신경망 번역' 적용 ... 이밖에 파파고와 별도

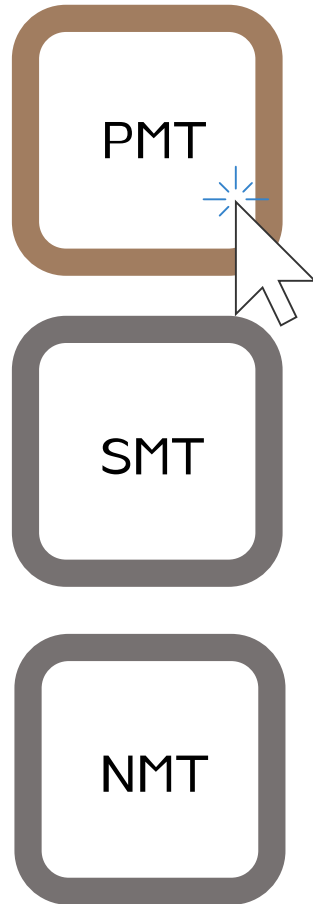
AI기술은 어떻게 '번역 서비스' 확 바꿨나

ZD넷 코리아 - 2017. 3. 15.

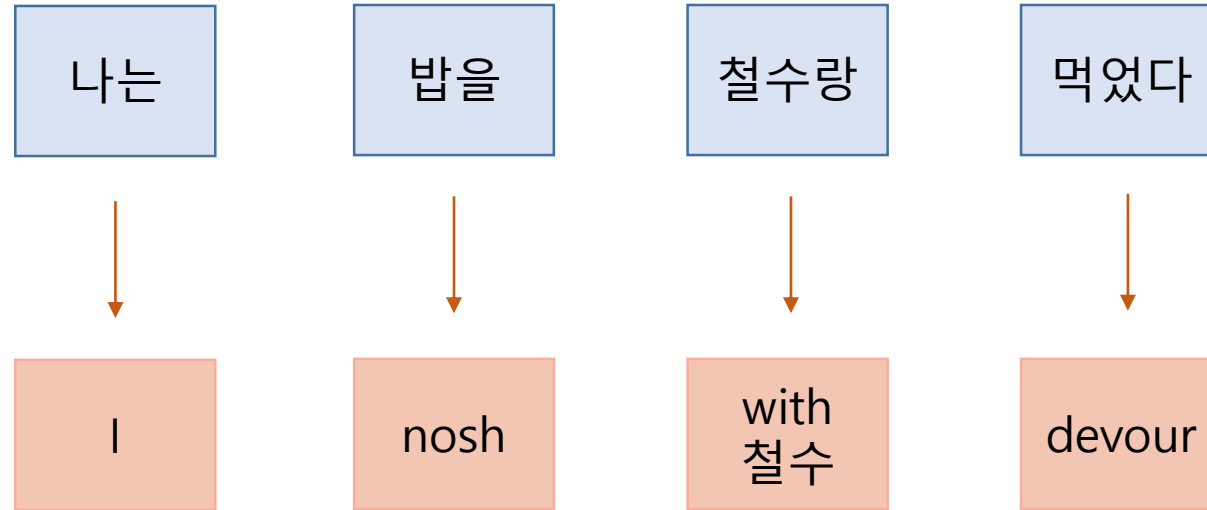
이와 함께 글로벌 관련 최신 기술 동향과 함께 앞으로 파파고 번역 서비스 활용 ... AI를 활용한 '인공신경망 기계번역(NMT)' 기술 덕분이다. ... 번역 커뮤니티를 개설하고, 경

✓ 인공신경망 번역(NMT)을 적용했기 때문.

2. 기계번역의 역사

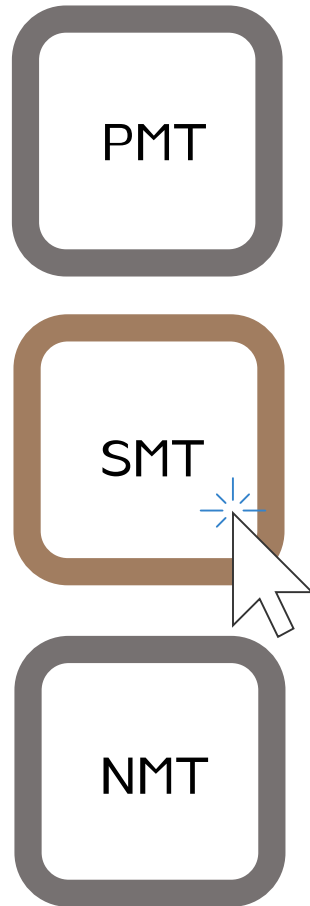


○ 어구 기반 기계 번역 (Phrase Machine Translation)

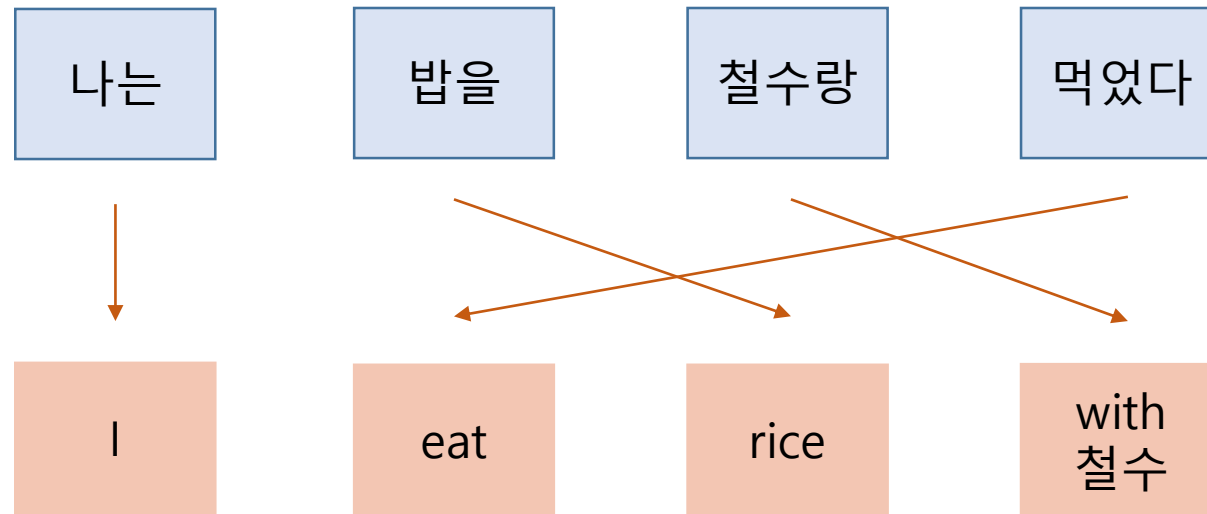


- ✓ 1950년대 ~ 1980년대
- ✓ 외국어를 잘 모르는 사용자가 사전만 펴놓고 번역을 진행한 식.

2. 기계번역의 역사



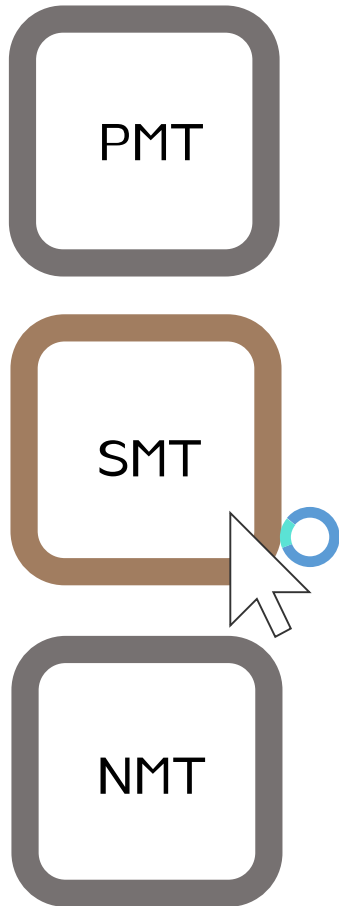
○ 통계 기반 기계 번역 (Statistical Machine Translation)



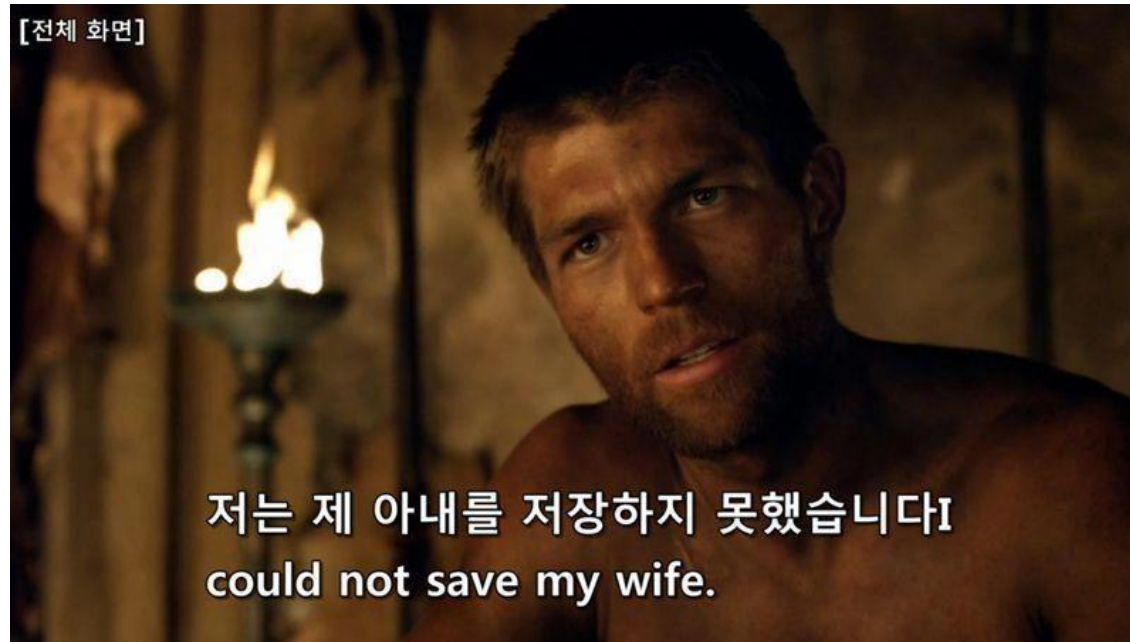
✓ 1988년 ~ 2016년

✓ 수많은 번역 데이터를 수집해 통계를 낸 후, 이 가운데 가장 널리 이용되는 번역 방식을 활용하는 기술

2. 기계번역의 역사

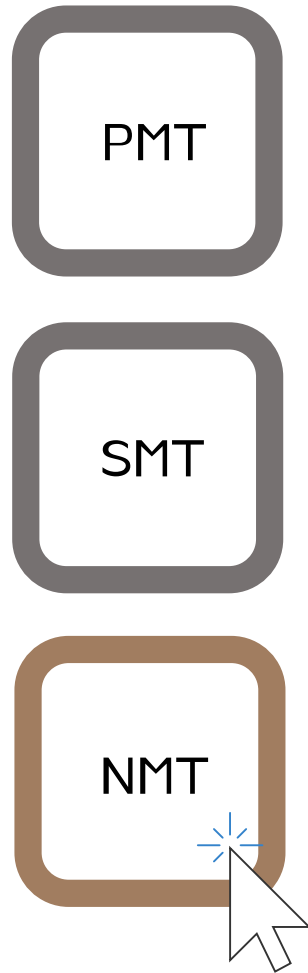


- 통계 기반 기계 번역 (Statistical Machine Translation)

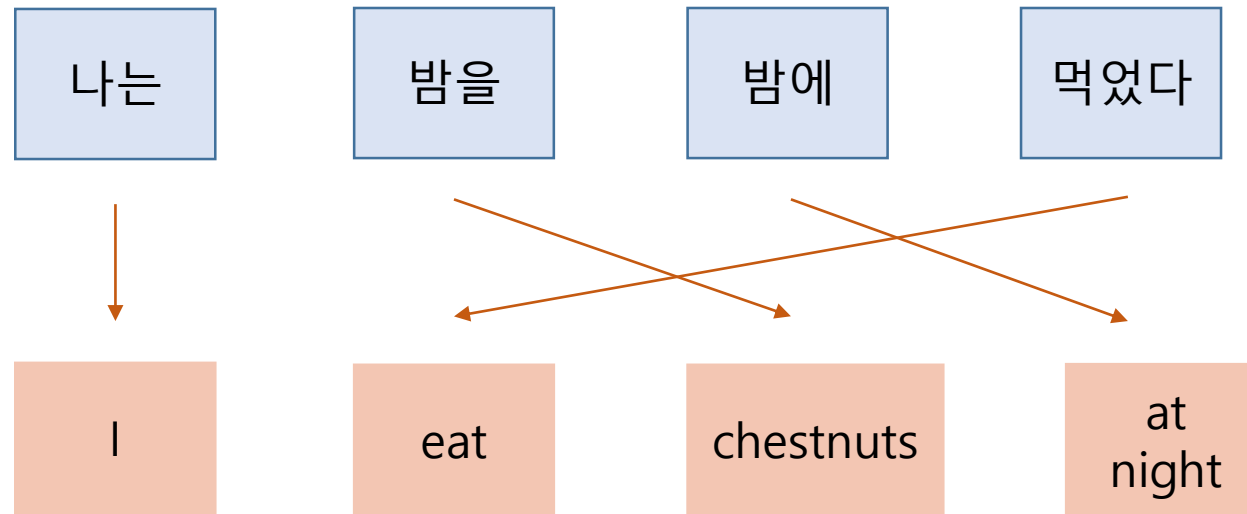


- ✓ 단점 : 가장 널리 이용했다고 그것이 맞는 것은 아님.
특히 어순이 다른 경우 (한글<->영어) 성능이 매우 떨어짐
사용자가 적어 번역 데이터가 부족한 언어는 제대로 번역할 수 없음.
ex) ५ नुलुग लसश

2. 기계번역의 역사



○ 신경망 기반 기계 번역 (Neural Machine Translation)



- ✓ 2016년 말 ~ 현재
- ✓ 사람처럼 문장을 인식해서 번역. 사람은 단어와 어구를 일일이 구별해가며 읽지 않음. 특정 언어를 번역할 수 있는 모델을 사전에 훈련시키고 문장 전체의 흐름 속에서 번역을 실시.

3. NMT의 원리

존은 11년 동안 아이들을 가르쳤다.

John has been teaching children for 11 years

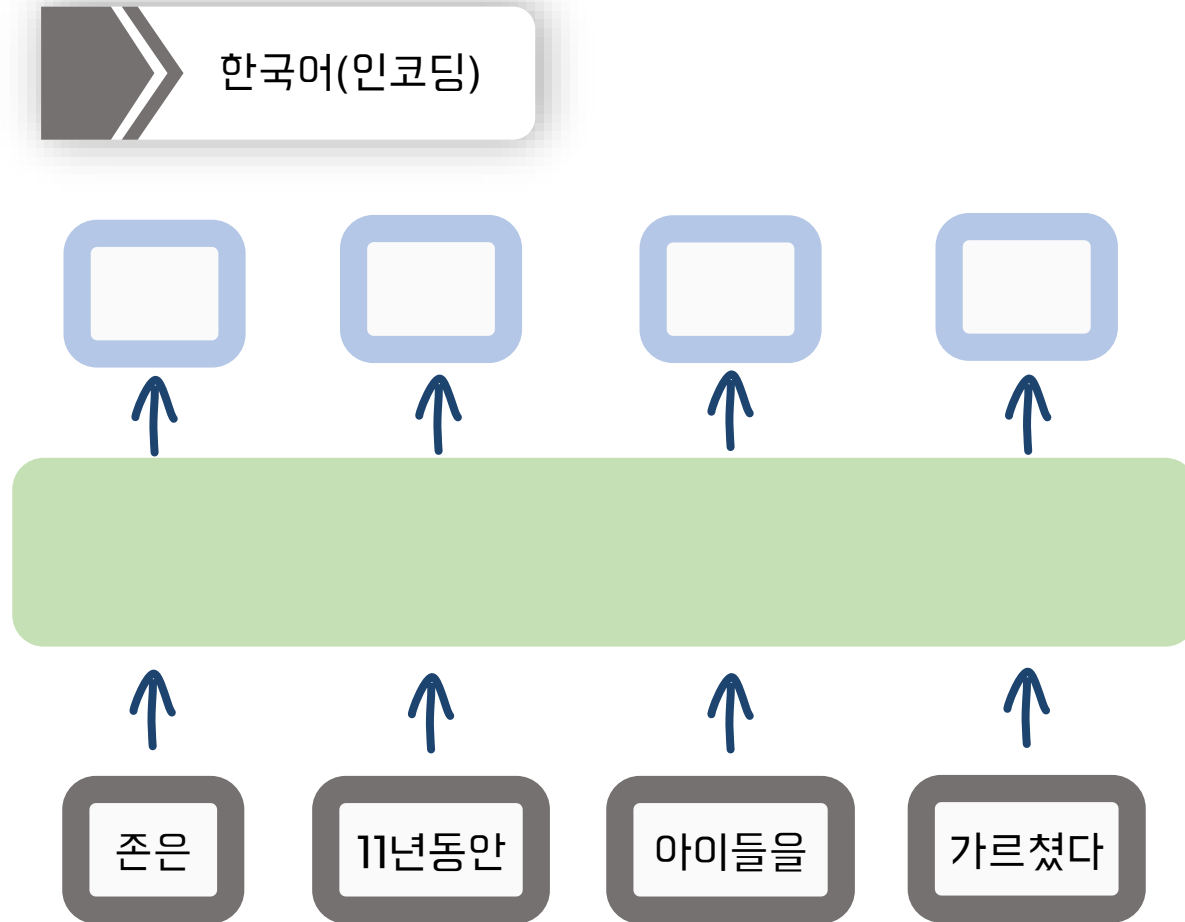
한국어(인코딩)



영어(디코딩)

- ① 한국어와 영어로 된 한 쌍의 트레이닝 데이터 충분히 확보 후 딥러닝 알고리즘에 대입.
- ② 한국어 문장을 통째로 코드화 (인코딩) 후 다시 코드를 풀어 (디코딩) 영어 문장을 생성.
- ③ 생성된 영어 문장과 정답 사이의 '오류 거리'를 측정.
- ④ 수많은 영어 문장을 반복하며 오류 거리를 줄이다 보면 알고리즘이 스스로 규칙 파악 및 학습.

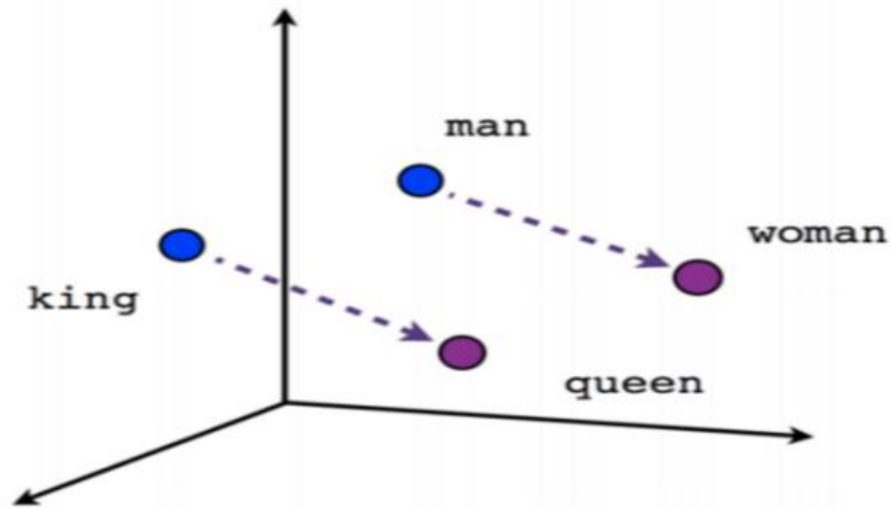
3. NMT의 원리



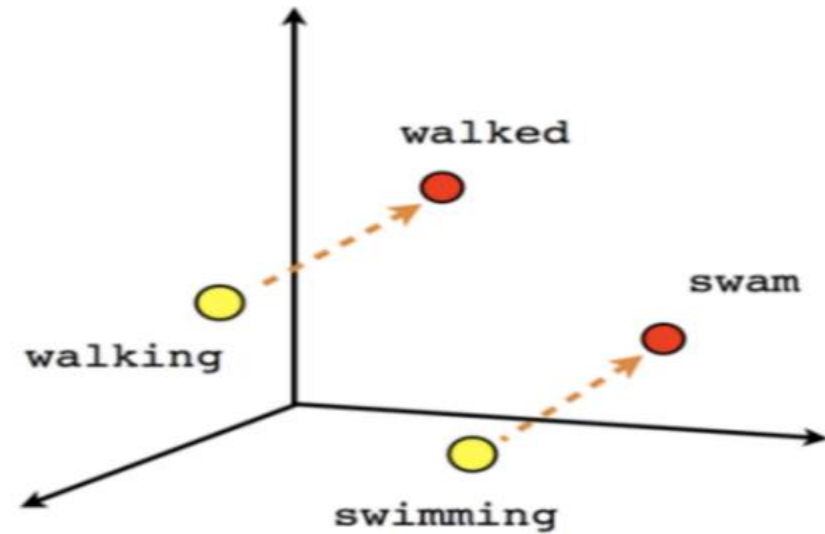
- ✓ 입력 문장을 단어 단위로 끊어
각 단어를 벡터에 인코딩
< 사람의 언어 -> 컴퓨터의 언어 >
- ✓ 각 단어마다 벡터를 하나씩 만들고
그 단어에 관련된 정보를 저장
- ✓ 벡터 값은 그 단어를 1000차원으로
보냈을 때의 값. 이 값을 통해
Cat과 kitten 사이의 거리와
Dog와 puppy 사이의 거리가
매우 비슷함을 알 수 있음.

3. NMT의 원리

○ Word2vec



Male-Female

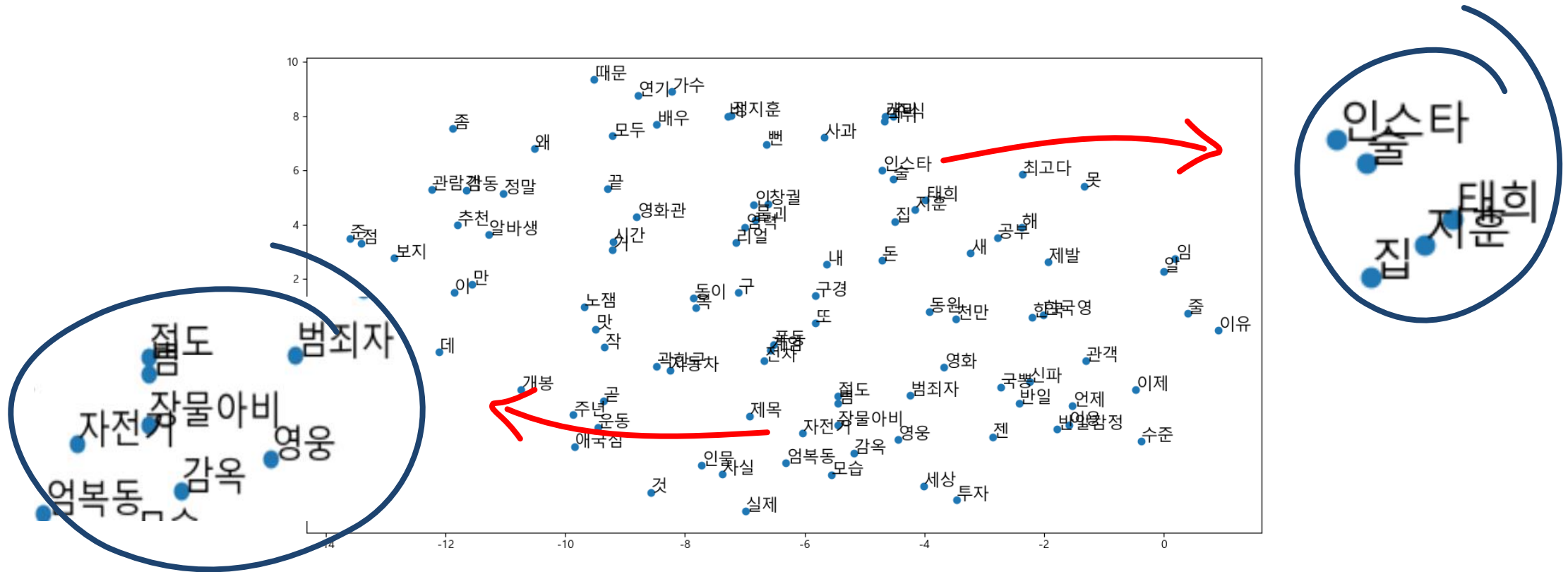


Verb tense

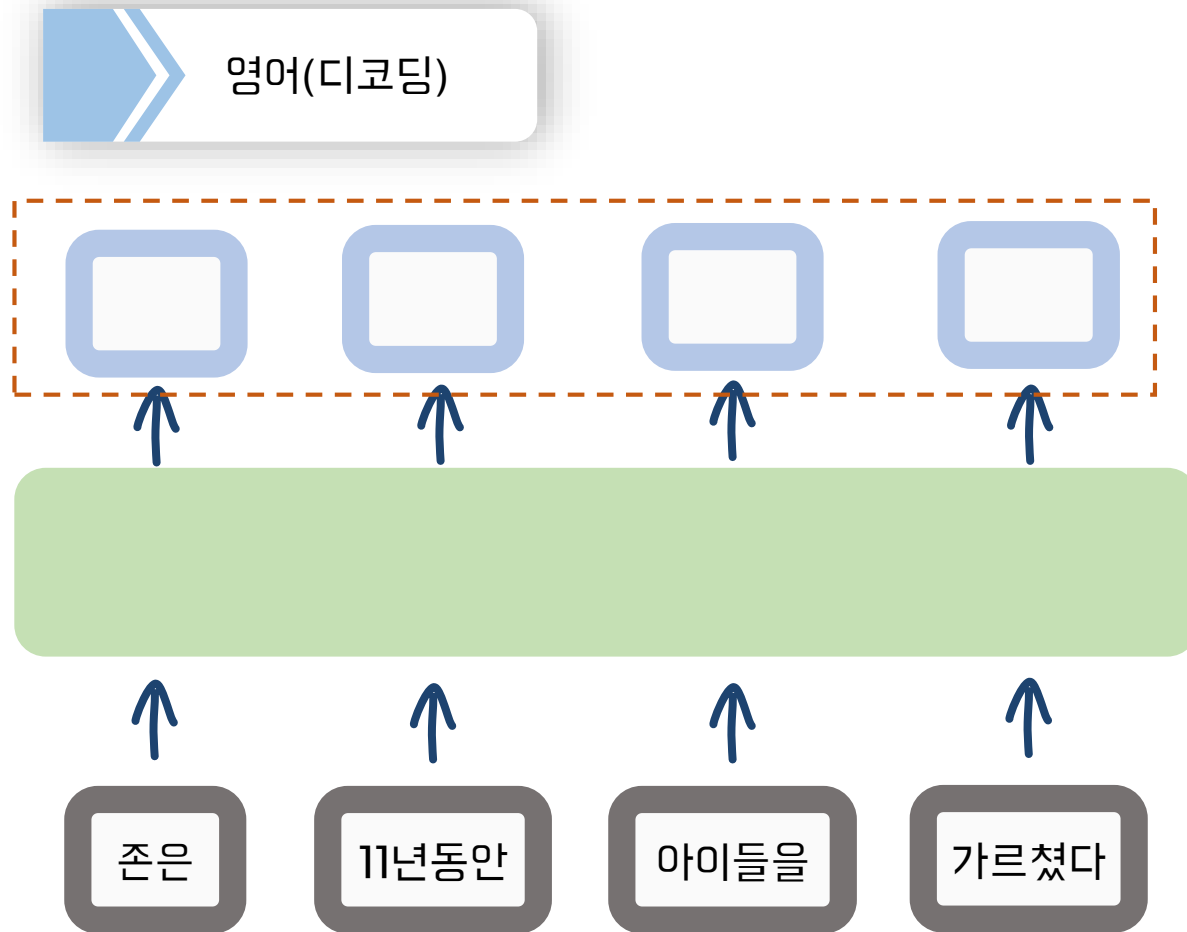
- ✓ King 과 queen 사이의 거리가 Man 과 Woman 사이의 거리와 같음.
- ✓ 품사에 따라 적용할 수 있음.

3. NMT의 원리

○ Word2vec



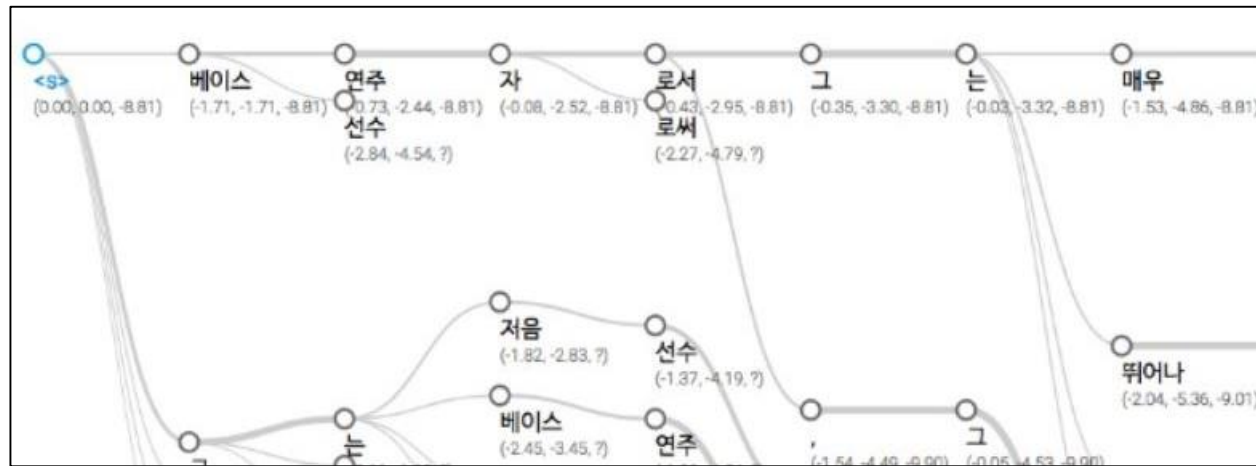
3. NMT의 원리



- ✓ Encoding 결과물을 Decoding 하여 문장을 생성.
- ✓ 이 때, 문장을 하나만 생성하는 것이 아니라 수많은 문장을 동시에 생성.
- ✓ 문장의 각 단어에 확률이 부여되고, 문장 전체 확률이 높을수록 좋은 번역문이 됨.

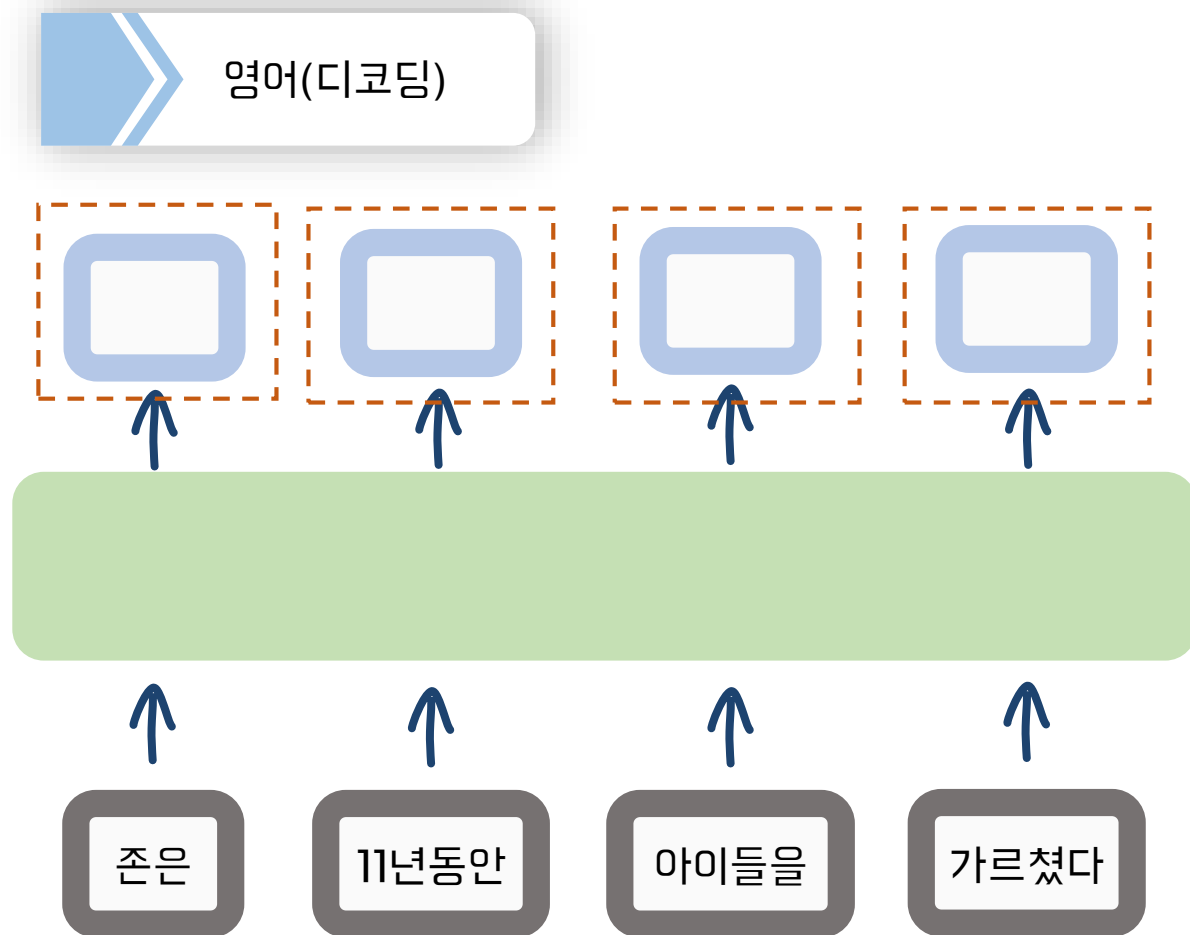
3. NMT의 원리

영어(디코딩)



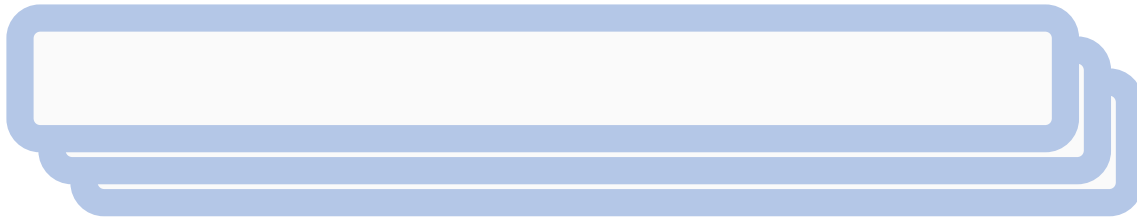
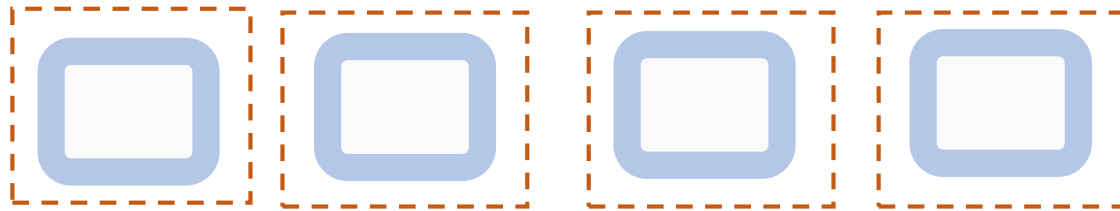
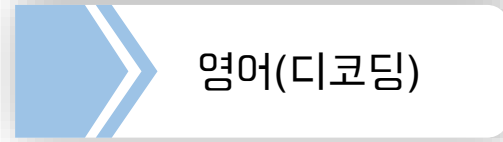
- ✓ 문장이 굉장히 많기 때문에, 실제로는 휴리스틱 트리 서치를 사용하여 생성 후보 제한
- ✓ 휴리스틱 트리서치란 모든 경우의 수를 해보기에 시간과 비용이 많이 들기 때문에 가장 그럴듯한 경우의 수들만 채택.

3. NMT의 원리



- ✓ 각 단어들을 만들 때에는 원문의 단어와 생성할 단어 간의 연관성을 계산.
- ✓ 즉 단어를 생각할 때 전체 입력 문장을 고려하지만, 모든 단어를 동일한 비율로 참고하는 것이 아닌, 해당 시점에서 예측해야할 단어와 연관이 있는 단어 부분을 좀 더 집중(attention)해서 봄.

3. NMT의 원리



VS

John has been teaching children for 11 years



이렇게 나온 문장 후보들과 실제 문장을 비교, 오차율 산출 및 반복.

4. 마치며

○ Papago



○ GooGle



- ✓ Papago는 N2MT, Google은 GNMT 라는 그들의 독자적인 NMT 모델을 사용.
- ✓ 알고리즘 상의 큰 차이는 없으나 Papago의 한국어 실력이 더 준수.

4. 마치며

- ✓ 실제 Papago나 google의 NMT 모델은 훨씬 복잡.
 - > 기존의 NMT 모델에서 LSTM 등과 같은 기술들을 추가적으로 사용
- ✓ 최근에 자연어 처리에 관한 머신 러닝이 눈부시게 발전.
 - > Bert, XLNet 등
 - > 단순 기계 번역을 넘어 질문에 답까지 찾아낼 수 있는 기법.

Q&A