

- 강의명: CSCI E-89B: 자연어 처리 입문
- 주차: Lecture 13
- 교수명: Dmitry Kurochkin
- 목적: Lecture 13의 핵심 개념 학습

기계 번역(Machine Translation) 심층 분석

자연어 처리(NLP) 입문 - 13주차 강의 통합 요약

December 10, 2025

□ 핵심 요약

문서 개요 (Executive Summary)

- 기계 번역(MT)은 소프트웨어를 통해 인간 수준의 번역 품질을 목표로 하는 기술입니다.
- 초기 규칙 기반(RBMT)에서 통계 기반(SMT)을 거쳐, 현재는 신경망 기반(NMT)이 주류입니다.
- Seq2Seq 모델은 입력과 출력을 연결하는 혁신을 가져왔으나 병목 현상이 있었고, 이를 어텐션(Attention)과 트랜스포머(Transformer)가 해결했습니다.
- 번역 품질 평가는 주로 BLEU 점수를 사용하며, 최근에는 규칙과 신경망을 합친 하이브리드 모델도 특정 분야에서 강세를 보입니다.

Contents

1	기본 개념 및 용어 정리	3
2	기계 번역(MT)의 개요와 발전사	3
2.1	기계 번역이란 무엇인가?	3
2.2	발전의 3단계 흐름	3
3	기계 번역의 난제 (Challenges)	4
4	핵심 기술 1: Seq2Seq 모델	4
4.1	Seq2Seq의 구조	4
4.2	해결책: 어텐션 메커니즘 (Attention Mechanism)	5
5	핵심 기술 2: 트랜스포머 (Transformers)	5
5.1	기존 모델(RNN/LSTM)과의 차이	5
5.2	핵심 구성 요소	5
6	평가 지표: BLEU Score	6
6.1	BLEU (Bilingual Evaluation Understudy)의 개념	6
6.2	계산 방법 (직관적 이해)	6
7	하이브리드 모델 (Hybrid Models)	6
7.1	왜 섞어 쓰는가?	6
7.2	구성 요소 및 장점	6
8	학습 체크리스트 및 FAQ	7
8.1	학습 완료 체크리스트	7
8.2	FAQ: 자주 묻는 질문	7
9	부록: 1페이지 요약 (Cheat Sheet)	8

1 기본 개념 및 용어 정리

기계 번역을 처음 접할 때 반드시 알아야 할 핵심 용어들입니다. 이 용어들은 문서 전체에서 반복되므로 미리 숙지하면 이해가 빠릅니다.

Table 1: 기계 번역 핵심 용어 정의

용어 (약어)	원어	쉬운 설명 및 비교
MT	Machine Translation	컴퓨터가 한 언어를 다른 언어로 번역하는 기술 전반.
RBMT	Rule-Based MT	1950 80년대 방식. 사람이 직접 문법 규칙과 사전을 입력해줌. (예: "I am" → "나는 이다")
SMT	Statistical MT	1990 2010년대 방식. 수많은 번역 데이터에서 '확률'을 계산해 번역. (예: 이 단어 뒤엔 저 단어가 올 확률이 80%)
NMT	Neural MT	2010년 이후 방식. 딥러닝(인공신경망)을 사용해 문장 전체의 맥락을 이해하고 번역. 현재의 주류.
Seq2Seq	Sequence-to-Sequence	입력 문장 (Sequence)을 받아 출력 문장 (Sequence)을 만들어내는 딥러닝 모델 구조.
Attention	Attention Mechanism	긴 문장을 번역할 때, "지금 번역하는 단어와 관련된 입력 단어"에 집중(Attention)하게 만드는 기술.
Transformer	Transformer	RNN을 버리고 'Attention'만으로 구성하여 속도와 성능을 획기적으로 높인 현대 AI의 기반 모델 (BERT, GPT의 조상).
BLEU	Bilingual Evaluation Understudy	기계 번역이 얼마나 잘 됐는지 채점하는 점수. 인간 번역과 얼마나 겹치는지 확인.

2 기계 번역(MT)의 개요와 발전사

2.1 기계 번역이란 무엇인가?

기계 번역은 단순히 단어를 1:1로 바꾸는 것이 아닙니다. 목표는 "*"의 의미와 문화적 뉘앙스를 보존하면서 문법적으로 정확한 문장을 만드는 것**입니다.

- **목표:** 인간 번역가의 품질을 복제하는 것.
- **용도:** 국제 비즈니스, 외교 문서, 실시간 회의 통역, 관용구(예: "hit the nail on the head" → "정곡을 찌르다")의 적절한 의역 등.

2.2 발전의 3단계 흐름

기계 번역은 크게 세 가지 시대를 거쳐 발전했습니다. "왜 이전 기술이 도태되었는가?"를 이해하는 것이 중요합니다.

1. 규칙 기반 (RBMT, 1950s–1980s)

- **방식:** 언어학자가 문법 규칙과 사전을 일일이 코딩했습니다.
 - **장점:** 낱씨 예보나 법률 문서처럼 형식이 정해진 문서는 아주 정확합니다.
 - **단점:** 예외가 너무 많아 확장이 어렵고, 자연스러운 구어체 번역에 실패합니다.
 - **사례:** 1954년 조지타운-IBM 실험 (러시아어 → 영어).
2. 통계 기반 (SMT, 1990s–2010s)
- **방식:** 수많은 이중 언어 데이터(예: UN 회의록)를 통계적으로 분석해 "이 단어 옆엔 저 단어가 온다"는 확률을 학습합니다.
 - **장점:** 규칙을 일일이 짤 필요가 없고, 데이터가 많으면 성능이 좋아집니다.
 - **단점:** 문맥보다는 '구(Phrase)' 단위로 쪼개서 번역하므로 문장이 뚝뚝 끊기는 느낌이 듭니다.
3. 신경망 기반 (NMT, 2010s–Present)
- **방식:** 딥러닝(Deep Learning)을 사용해 문장 전체를 하나의 벡터(숫자 덩어리)로 이해하고 번역합니다.
 - **장점:** ***"End-to-End" 학습** (입력에서 출력까지 한 번에 학습). 문맥을 파악하여 훨씬 유창하고 자연스럽습니다.
 - **현재:** Seq2Seq와 Transformer 모델이 이 시대를 이끌고 있습니다.

3 기계 번역의 난제 (Challenges)

번역은 왜 어려울까요? 컴퓨터 입장에서 가장 헷갈리는 문제들입니다.

1. 언어의 모호성 (Ambiguity)
 - 단어 하나가 여러 뜻을 가집니다.
 - 예시: 영어 "Bank"는 '은행'일 수도, '강둑'일 수도 있습니다. 문맥 없이 "I went to the bank"만 보고는 알 수 없습니다.
2. 구조적 차이 (Structural Differences)
 - 어순이 다릅니다.
 - 예시: 영어는 SVO (주어-동사-목적어)인 반면, 한국어는 SOV (주어-목적어-동사)입니다. 이를 맞추려면 문장 전체를 다 듣고 재배열해야 합니다.
3. 관용구 (Idiomatic Expressions)
 - 직역하면 의미가 통하지 않습니다.
 - 예시: "Shoot the breeze" → (직역) 산들바람을 쏜다? → (의역) 수다를 떨다.
4. 문맥 보존 (Context Preservation)
 - 긴 문단에서 앞 내용을 기억해야 합니다. 대명사(그것, 그 사람)가 무엇을 가리키는지 파악하는 것은 매우 어렵습니다.

4 핵심 기술 1: Seq2Seq 모델

4.1 Seq2Seq의 구조

NMT(신경망 번역)의 시초가 된 모델입니다. 크게 두 부분으로 나뉩니다.

- **인코더 (Encoder):** 원문을 읽고 그 의미를 압축하여 하나의 **컨텍스트 벡터 (Context Vector)**로 만듭니다. (독해 담당)
- **디코더 (Decoder):** 컨텍스트 벡터를 받아서 도착 언어로 문장을 생성합니다. (작문 담당)

주의사항

Seq2Seq의 치명적 단점: 병목 현상 (Bottleneck) 초기 Seq2Seq는 긴 문장을 아주 작은 벡터 하나에 억지로 구겨 넣어야 했습니다.

비유: 책 한 권을 읽고 내용을 단 한 문장으로 요약한 뒤, 그 요약본만 보고 다시 책 전체를 다른 언어로 써내라는 것과 같습니다. 정보 손실이 발생하여 긴 문장 번역 품질이 떨어집니다.

4.2 해결책: 어텐션 메커니즘 (Attention Mechanism)

어텐션은 **“다 기억하려 하지 말고, 필요할 때 원문을 다시 컨닝하자!”**는 아이디어입니다.

- 디코더가 번역할 때, 문장의 모든 부분을 동일하게 보는 게 아니라 **관련된 단어에 가중치(집중)**를 둡니다.
- 병목 현상을 해결하고 긴 문장 번역 성능을 획기적으로 높였습니다.

5 핵심 기술 2: 트랜스포머 (Transformers)

2017년 논문 *“Attention is All You Need”*에서 발표된 모델로, 현대 AI(GPT, BERT)의 조상입니다.

5.1 기존 모델(RNN/LSTM)과의 차이

- **기존:** 단어를 순서대로 하나씩 처리(순차적). 시간이 오래 걸리고 앞부분 내용을 까먹기 쉽습니다.
- **트랜스포머:** 문장 전체를 한 번에 처리(병렬 처리). 속도가 매우 빠르고 문장 내 단어 간의 관계를 한눈에 파악합니다.

5.2 핵심 구성 요소

1. **Self-Attention (셀프 어텐션):** 문장 내의 단어들이 서로 어떤 관계가 있는지 계산합니다. (예: “그것”이 앞의 “사과”를 가리키는지, “바나나”를 가리키는지 파악)
2. **Positional Encoding (위치 인코딩):** 한꺼번에 입력받으므로 단어 순서 정보를 인위적으로 추가해 줍니다.
3. **Encoder-Decoder 구조:** 여전히 번역을 위해 인코더와 디코더 구조를 유지하지만, 내부는 전부 어텐션으로 채워져 있습니다.

□ 핵심 요약

트랜스포머의 충격적인 효과

- **확장성:** 병렬 처리가 가능해져서 엄청나게 큰 데이터로 학습이 가능해졌습니다.
- **범용성:** 번역뿐만 아니라 요약, 챗봇, 심지어 이미지 분석(Vision Transformer)이나 단백질 구조 예측에도 쓰입니다.

6 평가 지표: BLEU Score

기계 번역이 잘 됐는지 사람이 일일이 채점할 수 없으므로, ****자동화된 채점 방식****이 필요합니다.

6.1 BLEU (Bilingual Evaluation Understudy)의 개념

기계가 번역한 문장이 ****사람(참조, Reference)**이 번역한 문장과 얼마나 단어가 많이 겹치는가**를 측정합니다.

6.2 계산 방법 (직관적 이해)

1. **N-gram 정밀도 (Precision):** 단어(1-gram), 두 단어 짝(2-gram), 세 단어 짝(3-gram)... 이 얼마나 일치하는지 봅니다.

$$\text{Precision}_n = \frac{\sum \text{Matches (일치하는 n-gram 수)}}{\sum \text{Total (기계 번역의 총 n-gram 수)}} \quad (1)$$

2. **간결성 페널티 (Brevity Penalty, BP):** 기계가 틀릴까 봐 단어 하나만 내뱉는 꼴수를 부리지 못하게, 길이가 참조 문장보다 짧으면 점수를 깎습니다.
3. **최종 점수:** N-gram 정밀도의 평균에 BP를 곱해서 계산합니다. (1에 가까울수록 완벽, 0에 가까울수록 엉망)

주의사항

BLEU 점수의 한계

- **의미 파악 불가:** "아름답다"를 "예쁘다"로 번역하면, 뜻은 통하지만 단어가 달라서 점수가 깎일 수 있습니다. (동의어 처리 미흡)
- **어순 유연성 부족:** 한국어처럼 어순이 자유로운 언어에서는 점수가 부정확할 수 있습니다.

7 하이브리드 모델 (Hybrid Models)

최근에는 하나의 방식만 고집하지 않고 섞어서 씁니다.

7.1 왜 섞어 쓰는가?

- **NMT(신경망)의 약점:** 가끔 엉뚱한 번역을 하거나 전문 용어를 틀립니다.
- **RBMT/SMT(규칙/통계)의 장점:** 문법이 정확하고 특정 분야 용어 사전을 강제로 적용하기 좋습니다.

7.2 구성 요소 및 장점

- **신경망 + 규칙:** 전체적인 번역은 신경망이 유창하게 하고, 법률 용어나 문법 교정은 규칙 기반 시스템이 마무리합니다.
- **장점:** 유창함(Fluency)과 정확성(Accuracy)을 동시에 잡을 수 있습니다.
- **사례:** SYSTRAN, Microsoft Translator, IBM Watson (의료, 법률 등 특수 분야용).

8 학습 체크리스트 및 FAQ

8.1 학습 완료 체크리스트

RBMT, SMT, NMT의 차이점과 발전 순서를 설명할 수 있는가?

기계 번역의 4가지 난제(모호성, 구조, 관용구, 문맥)를 예시와 함께 말할 수 있는가?

Seq2Seq의 '병목 현상'이 무엇이며, '어텐션'이 이를 어떻게 해결했는지 이해했는가?

트랜스포머가 RNN보다 빠른 이유(병렬 처리)를 아는가?

BLEU 점수의 계산 원리(N-gram 일치)와 한계점(의미 파악 불가)을 아는가?

8.2 FAQ: 자주 묻는 질문

Q. 트랜스포머는 번역에만 쓰이나요? A. 아니요. 트랜스포머는 현재 AI의 표준입니다. 챗GPT 같은 대화형 AI, 요약, 문장 생성, 심지어 코딩 작성까지 모든 자연어 처리에 쓰입니다.

Q. BLEU 점수가 높으면 무조건 좋은 번역인가요? A. 꼭 그렇지는 않습니다. 점수는 높아도 사람이 읽기에 어색할 수 있고, 점수는 낮아도 의미가 완벽하게 통할 수 있습니다. 그래서 사람에 의한 평가도 여전히 중요합니다.

Q. 요즘도 규칙 기반(RBMT)을 쓰나요? A. 순수 RBMT는 거의 안 쓰지만, 하이브리드 모델의 일부로써 문법 교정이나 특정 용어 강제 적용을 위해 여전히 중요한 역할을 합니다.

9 부록: 1페이지 요약 (Cheat Sheet)

Machine Translation 한 장 정리

1. 역사적 흐름 (Evolution)

- **Rule-Based (50s-80s):** 규칙+사전. 정확하지만 확장 불가. (비유: 문법책 보고 작문)
- **Statistical (90s-10s):** 데이터 확률 기반. 유연하지만 문맥 부족. (비유: 단어 맞추기 퍼즐)
- **Neural (10s-Present):** 딥러닝 (Seq2Seq, Transformer). 유창하고 문맥 파악. (비유: 전문 통역사)

2. 핵심 모델 (Models)

- **Seq2Seq:** 인코더 → 컨텍스트 벡터 → 디코더. (단점: 정보 병목)
- **Attention:** "필요한 부분만 집중해서 보자". 병목 해결, 긴 문장 OK.
- **Transformer:** "Attention is All You Need". 병렬 처리로 초고속 학습, 현대 AI의 기반.

3. 평가 (Evaluation - BLEU)

- **원리:** 기계 번역과 사람 번역의 N-gram(단어 뭉치) 일치도 측정.
- **특징:** 1.0 만점. 짧으면 페널티(BP). 동의어/의미 파악은 못함.

4. 하이브리드 (Hybrid)

- 신경망의 유창함 + 통계/규칙의 정확성 결합.
- 법률, 의료 등 전문 분야에서 선호됨.