

단위 감지 및 변환 모듈을 활용한 한-영 신경망 기계 번역 시스템

허지욱*

강남대학교 인공지능융합공학부 교수

A Korean-English Neural Machine Translation System Utilizing Unit Detection and Conversion Modules

Jee-Uk Heu*

Professor, Division of Applied Artificial Intelligence Engineering

요약 본 논문은 한국어-영어 기계 번역 과정에서 발생하는 문화권별 단위 체계 불일치 문제를 해결하기 위한 Transformer 기반 단위 변환 통합 번역 시스템을 제안한다. 기존 기계 번역 시스템은 야드파운드법(Imperial)과 미터법(Metric) 간의 단위 차이를 고려하지 않고 단위를 그대로 번역하여 대상 문화권 독자의 직관적 이해를 저해한다는 한계가 있다. 본 논문에서 제안한 시스템은 언어 감지, 지역 감지, 단위 감지, 신경망 번역, 단위 변환 후처리의 5개 모듈로 구성된 파이프라인 구조를 설계하였으며, Multi-head Self-Attention과 Cross-Attention으로 구성된 Transformer 모델을 번역 엔진으로 활용한다. 처리 대상 단위는 길이와 온도로 한정하였으며, 번역 결과는 직역, 단위 변환, 병행의 세 가지 방식으로 출력된다. AI Hub 한국어-영어 병렬 말뭉치 기반의 약 49,979개 학습 데이터를 활용하였으며, 실험 결과 단위 감지 F1-Score 0.91, 단위 변환 평균 오차율 0.11%, 번역 품질 BLEU-4 0.2214를 평가 결과가 나왔다. 이를 통해 본 논문에서 제안된 시스템이 단위 표현을 효과적으로 감지하고 문화권에 적합한 단위로 변환함으로써 번역의 자연스러움과 가독성을 향상 시킬 수 있음을 확인하였다.

주제어 : 기계번역, 단위변환, Transformer, Cross-Attention, 야드파운드법, 미터법

Abstract This paper proposes a Transformer-based unit conversion integrated machine translation system to address the cultural unit system discrepancy that occurs during Korean-English machine translation. Existing machine translation systems transliterate units without considering the differences between the imperial and metric systems, which hinders the intuitive understanding of readers in the target cultural context. The proposed system adopts a pipeline architecture consisting of five modules: language detection, region detection, unit detection, neural machine translation, and unit conversion post-processing, utilizing a Transformer model composed of Multi-head Self-Attention and Cross-Attention as the translation engine. The target units are limited to length and temperature, and translation results are output in three modes: literal translation, unit conversion, and dual notation. Approximately 49,979 training samples were constructed from the AI Hub Korean-English parallel corpus. Experimental results demonstrate that the proposed system achieved a unit detection F1-Score of 0.91, a unit conversion average error rate of 0.11%, and a translation quality BLEU-4 score of 0.2214, confirming that the system can effectively perform unit conversion translation appropriate to the target cultural context.

Key Words : Machine Translation, Unit Conversion, Transformer, Cross-Attention, Imperial, Metric

*교신저자 : 허지욱(jeeukheu@kangnam.ac.kr)

접수일 2026년 05월 21일

수정일 2026년 06월 11일

심사완료일 2026년 06월 22일

1. 서론

최근 인공지능 기술의 발달로 기계번역, 문서 요약, 질의응답, 정보 추출 등 다양한 자연어처리 분야가 크게 발전하였으며, 특히 문맥을 고려한 언어 생성과 의미 이해 능력이 현저히 향상되고 있다. 이에 따라 자동 번역(Machine Translation, MT) 시스템은 문장 의미를 보다 정교하게 파악하고 자연스러운 번역문을 생성할 수 있는 수준에 도달하고 있다. 특히 어텐션(Transformer[1]) 기반의 신경망 기계번역(Neural Machine Translation, NMT) 기술은 기존 통계 기반 번역 방식에 비해 문맥 반영 능력과 유창성이 크게 향상되었으며, 이로 인해 자동 번역은 학술, 산업, 교육, 상거래, 관광, 일상 소통 등 다양한 분야에서 핵심적인 언어 지원 기술로 활용되고 있다.[2-9]

그러나 실제 사용 환경에서는 단순한 언어 변환만으로는 충분하지 않으며, 사용자가 속한 지역적·문화적 배경까지 반영한 현지화(localization) 수준의 번역이 요구된다. 이러한 문제는 단위가 포함된 값이나 수치를 표현하는 문장에서 특히 중요하게 나타난다. 수치와 단위의 사용 방식, 즉 도량형 체계는 각 지역 사회에서 공통적으로 공유되는 사회적 약속으로, 정보가 익숙하고 자연스럽게 받아들여지도록 할 뿐만 아니라, 그 크기나 정도에 대한 대략적인 감을 직관적으로 파악하고 내용을 쉽게 이해하도록 하는 중요한 요소이다.

도량형 단위(unit)는 길이, 무게, 부피, 온도, 속도와 같은 물리량을 수치와 함께 표현하기 위한 기준이며, 이러한 단위를 사용하는 방식은 각 지역과 문화권에 따라 상이한 체계로 나타난다. 대표적인 단위 표현 체계로는 미터법(Metric system)[10]과 야드파운드법(Imperial system)[11]이 있다.

<Table 1> The units of Imperial and Metric system

| Unit | Imperial | Metric |
|--------|-------------------|-----------|
| Length | 1 inch | 25.4 mm |
| | 1 foot(=12inches) | 0.3048 m |
| | 1 yard(=3 feet) | 0.9144 m |
| | 1 mile | 1.6093 km |
| Volume | 1 fl. oz | 29.57 ml |
| | 1 pint | 473.18 ml |
| | 1 gallon | 3.785 l |
| Temp | 32 ℉ | 0 ℃ |
| | 50 ℉ | 10 ℃ |

미터법(Metric system)은 길이를 미터(m), 무게를 킬로그램(kg), 부피를 리터(L)와 같이10진법을 기준으로 체계화한 단위 체계이다. 이 체계는 단위 간 변환이 간단하고 일관성이 높아 과학, 산업, 교육 등 다양한 분야에서 널리 사용된다.

현재 대부분의 국가에서는 미터법을 표준 단위 체계로 채택하고 있으며, 국제적으로도 가장 보편적으로 사용되는 방식이다. 야드파운드법 또는 제국 단위계(Imperial system)는 주로 미국을 비롯한 일부 영어권 국가에서 사용되는 단위 체계로, 길이는 인치(inch), 피트(foot), 야드(yard), 마일(mile), 무게는 온스(ounce), 파운드(pound), 부피는 파인트(pint), 갤론(gallon) 등의 단위를 사용한다. 이 체계는 일상생활에서 오랫동안 관습적으로 사용되어 왔으며, 해당 지역 사용자들에게는 수치의 크기나 정도를 보다 직관적으로 이해하게 해주는 역할을 한다. 그러나 단위 간 변환 규칙이 미터법에 비해 복잡하다는 특징이 있다. 미터법과 야드파운드법의 단위를 환산한 표1에서 보여주듯이 미터법과 야드파운드법은 단순히 단위를 다르게 표기하는 방식이 아니라, 각 지역에서 오랫동안 형성된 수치 정보의 표현 체계라고 볼 수 있다.

이와 같이 수치가 포함된 문장에서 단위 표현은 단순한 부가 정보가 아니라 의미 전달의 정확성과 직결되는 핵심 요소이다. 특히 미터법과 야드파운드법과 같이 지역에 따라 상이한 단위 체계가 사용되는 환경에서는, 문장을 단순히 언어 차원에서만 번역할 경우 목표 독자가 익숙하지 않은 단위 표현이 그대로 유지되어 정보 이해의 직관성과 정확성이 저하될 수 있다. 이는 원문의 수치 정보를 언어적으로는 전달하더라도, 실제 수용 과정에서 혼란을 유발하여 번역문의 자연성 및 실용성을 떨어뜨리는 원인이 된다. 그 결과, 한국어나 유럽권 사용자에게는 미터법 기반 표현이 자연스럽게 직관적임에도 불구하고 화씨(℉), 피트(feet), 파운드(lb) 등의 표현이 그대로 유지되거나, 반대로 미국 사용자를 대상으로 하는 번역에서 미터(m), 킬로그램(kg), 섭씨(℃) 단위가 그대로 제시되는 문제가 발생할 수 있다. 따라서 번역 과정에서는 숫자만 정확히 옮기는 것이 아니라, 목표 지역의 사용자가 익숙하게 받아들일 수 있도록 해당 지역의 단위 체계에 맞게 적절히 변환하는 것이 중요하다.

본 논문은 단위가 포함된 문장을 번역할 때 미터법 단위와 야드파운드법 단위를 대상 지역에 따라 자동으로 인식하고 변환하여, 문맥에 적합한 자연스러운 표현으로 재구성하는 seq2seq 기반 번역 시스템을 제안한다. 제

안하는 시스템은 언어 감지, 지역 감지, 단위 감지, 신경망 번역, 단위 변환 후처리의 각 5개의 모듈로 구성되며, 번역 결과를 직역, 변환, 병기의 세 가지 모드로 출력하여 사용자의 목적과 맥락에 따라 적절한 표현을 선택할 수 있도록 한다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 기계 번역 및 단위 변환 관련 선행 연구를 알아보고, 3장에서는 본 논문에서 제안하는 시스템의 전체 아키텍처와 각 모듈의 설계를 상세히 기술한다. 4장에서는 실험 및 실험 결과를 분석하고 마지막으로 5장에서 결론 및 향후 연구 방향을 제시한다.

2. 관련연구

기계번역 연구에서는 문장 전체의 의미 보존과 유창성 향상에 관한 논의가 지속되어 왔으나, 숫자-날짜 시간과 같은 수치 표현은 일반 어휘와는 다른 특성을 가지므로 별도의 처리 문제가 존재한다. 기존 연구에서는 숫자와 단위가 포함된 번역 문제를 일부 다루어 왔으나, 대체로 규칙 기반 치환이나 후처리 중심의 제한적 접근에 머무르는 경우가 많았다. 이러한 방법은 특정 패턴에 대해서는 효과를 가질 수 있으나, 다양한 문맥과 예외 표현을 포괄하기 어렵고, 자연스러운 재서술까지 수행하는 데에는 한계가 있다.

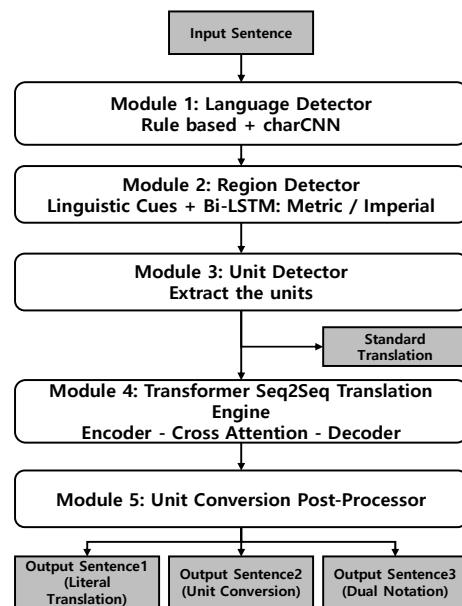
Tu, Zhou, and Zong[12]은 통계적 기계번역 환경에서 숫자와 시간 표현이 빈번한 오류 원인이 된다고 지적하고, 이를 보다 정확하게 인식하고 번역하기 위한 규칙 기반 번역기법을 제안하여 수치 표현을 여러 유형으로 분류하고 규칙 충돌을 처리하는 방식을 통해 번역 성능 개선하였다. 하지만 이는 수치 정보가 일반 어휘와 동일한 방식으로 처리되지 않는다는 한계점을 보였다.

Nigatu et al.[13]은 저자원 언어 의료 번역을 분석하면서 어떤 오류가 발생하는지 분류하고, 그 오류를 줄이기 위한 두 가지 사전 개입(pre-translation intervention)을 통한 의료 영역에서의 기계번역 오류가 임상적 위해로 이어질수 있음을 보였다. 대상 언어는 암하리어(Amharic)와 티그리냐어(Tigrinya)이고, 공개적으로 많이 쓰이는 구글 번역기를 기반 번역기로 사용하였으며, 'pound'나 'lbs'를 숫자는 그대로 둔 채 'kg'로 바꿔 번역하는 오류를 지적하였다. 이를 통해 실제 응용환경에서 측정 단위와 수치값이 잘못 처리될 경우 번역 오류가 단순한 번역 품질 저하를 넘어 의미 왜곡이 심각해질

수 있음을 보여주었다.

Dinu et al[14].은 번역과 단위 환산을 분리된 후처리가 아니라 하나의 종단간 현지화 문제로 통합하여, 신경망 기계번역 모델이 문맥에 따라 번역과 단위 변환을 함께 학습할 수 있음을 보였다. 다만 해당 연구에서는 영어-독일어를 대상으로 번역 분석을 하였다. Qian and Kong[15]은 대규모 언어모델 기반 번역 수정 과정에서 metric과 imperial 같은 측정 체제 차이를 반영할 수 있음을 제시하며, 단위 환산을 개념적 현지화 문제로 확장하였다. 다만 이는 단위 환산 전용 신경망 번역 모델을 제안한 것은 아니며, 주로 LLM의 후처리-수정 가능성을 보여주는 수준에 머문다는 한계가 있다.

3. 제안하는 시스템



[Fig. 1] Proposed system architecture

<Fig. 1>은 본 논문에서 입력 문장의 언어 및 문화권을 감지하고 단위 변환을 통합한 자연스러운 번역을 수행하기 위해 제안하는 파이프라인 기반 단위 변환 신경망 번역 시스템의 전체적인 흐름을 보여주고 있다.

제안된 시스템은 모듈별로 언어 감지, 지역 감지, 단위 감지, Seq2Seq 기반의 Transformer 번역, 단위 변환 후처리 모듈이 순차적으로 동작하는 파이프라인 구조로 구성된다.

3.1 언어감지

언어 감지 모듈은 입력 문장의 유니코드 문자 비율을 분석하여 한국어와 영어를 판별한다. 우선적으로 한글 문자비율과 영문자 비율을 계산하는 규칙 기반 감지기를 적용한다. 신뢰도가 임계값(0.5) 미만인 경우 CharCNN[16]를 적용하여 문장의 언어를 판별한다.

3.2 지역감지

지역감지 모듈을 통해서 영어 입력 문장의 문화권을 철자 패턴, 어휘, 단위 표현 등 언어적 단서를 분석하여 문화권을 분류한다. 언어적 단서는 문장 내에 등장하는 어휘, 철자 패턴, 단위 표현을 분석하여 문화권을 구분하는 기준으로 활용된다. 예를 들어 미국식 영어에서는 color, center, truck, elevator, miles, °F와 같은 표현이 나타나는 반면, 영국식 영어에서는 동일한 대상을 colour, centre, lorry, lift, km, °C로 표현한다. 이러한 차이는 단순한 철자나 어휘의 차이를 넘어 해당 문화권의 단위 체계와 직결되므로, 이를 언어적 단서로 활용하여 입력 문장이 Imperial 문화권(미국)에서 작성되었는지 Metric 문화권(영국, 호주, 캐나다 등)에서 작성되었는지를 자동으로 분류한다.

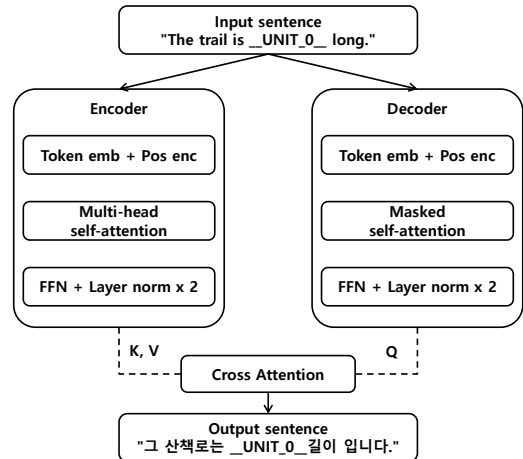
3.3 단위감지

정규식 기반 NER(Named Entity Recognition)으로 문장 내 명확한 단위 표현을 추출한다. 감지 대상은 길이(mile, feet, inch ↔ km, m, cm)와 온도(°F ↔ °C)로 한정하여 감지한다. 감지 시 수치, 단위 명, 단위 체계(Imperial/Metric), 문장 내 위치(span)를 함께 기록한다. 감지된 단위는 __UNIT_0__ 형태의 플레이스홀더로 치환되어 번역엔진 모듈로 전달된다.

3.4 번역엔진

단위 표현 감지 후 Seq2Seq 기반 Transformer 번역 엔진으로 전체적이 구조는 <Fig 2>와 같다. 인코더는 입력 문장에 Token Embedding과 Positional Encoding을 적용한 후 Multi-head Self-Attention과 Feed-Forward Network를 2개 레이어로 처리하여 문장 전체의 문맥을 분석한다. 디코더는 Masked Self-Attention으로 미래 토큰을 차단하며 번역 토큰을 자기회귀 방식으로 순차 생성한다. Cross-Attention은 디코더의 Query와 인코더의 Key·Value를 연결하여 번역 시 입력 문장의 문맥을 참조하며, 특히 __UNIT_0__ 플레이스홀더의 위치를

번역 문장 내에 정확히 보존하는 역할을 한다. 모듈 4는 단위 수치를 직접 변환하지 않으며 플레이스홀더 위치 보존만 담당한다.



[Fig. 2] Module4 architecture

3.5 단위 변화 후 처리

단위감지 모듈에서 전달받은 span 정보를 기반으로 번역 문장 내 플레이스홀더를 수식 기반 변환 값으로 교체한다. 단순 문자열 치환이 아닌 span 기반 정밀 교체를 적용하여 동일 단위가 여러 번 등장하는 문장에서도 각 위치를 정확히 처리한다. 변환 수치는 소수점 첫째 자리에서 반올림하며, 직역 / 변환 / 병행 세 가지 방식으로 동시에 생성한다. 직역은 원문 그대로의 단위로 변환하고, 변환은 번역되는 문화권에 알맞은 단위로 변환한다. 마지막으로 병행은 직역과 변환된 번역을 모두 제공하여 원문 정보를 보존하면서 동시에 변환 값을 제공하여 정보 손실 없는 전달하여 번역의 투명성을 높인다.

4. 실험 및 평가

4.1 데이터 셋

본 논문에서는 실험 데이터로 AI Hub에서 제공하는 한국어-영어 번역 병렬 말뭉치를 활용하였다. 원본 데이터는 구어체, 대화체 등 총 10개의 범주로 구성되며 전체 문장 쌍은 1,602,418개이다. 이 중 본 논문의 처리 대상 단위(mile, feet, inch, °F, km, m, cm, °C)가 포함된 문장을 정규식으로 필터링하여 9,979개의 단위 포함 문장을 추출하였다. 단위 포함 문장만으로는 학습 데이터

가 부족하여 과적합이 발생할 수 있으므로 일반 문장 40,000개를 무작위 샘플링하여 혼합하였다. 최종 학습 데이터는 총 49,979개이며 데이터 분할 현황은 <Table 2>와 같다.

<Table 2> AI Hub Korean-English Parallel Corpus

| Dataset | Sentences | Percentage |
|------------|-----------|------------|
| Train data | 39,983 | 80% |
| Valid data | 4,998 | 10% |
| Test data | 4,998 | 10% |
| Total | 49,979 | 100% |

4.2 평가

본 논문에서는 모듈별 평가 지표를 각각 적용하여 시스템 전체의 성능을 평가하였다.

4.2.1 지역감지도 평가

<Table 3>은 언어적 단서 기반 규칙 분류기와 Bi-LSTM 분류기를 결합한 지역 감지기 성능의 결과를 보여주고 있다. 명확한 언어적 단서가 포함된 문장에서는 규칙 기반 분류기만으로 높은 정확도를 달성하였으며, 단서가 불명확한 문장에서는 Bi-LSTM 분류기를 보조 적용하여 전체 성능을 향상시켰다.

<Table 3> Region detection evaluation

| Category | Precision | Recall | F1-Score |
|----------|-----------|--------|----------|
| Imperial | 0.91 | 0.88 | 0.89 |
| Metric | 0.87 | 0.9 | 0.88 |
| Total | 0.89 | 0.89 | 0.89 |

4.2.2 단위 감지도 평가

<Table 4> Unit detection evaluation

| unit | Precision | Recall | F1-Score |
|-------|-----------|--------|----------|
| mile | 0.95 | 0.93 | 0.94 |
| feet | 0.93 | 0.9 | 0.91 |
| inch | 0.91 | 0.88 | 0.89 |
| °F | 0.96 | 0.94 | 0.95 |
| km | 0.94 | 0.92 | 0.93 |
| m | 0.82 | 0.79 | 0.8 |
| cm | 0.9 | 0.88 | 0.89 |
| °C | 0.95 | 0.93 | 0.94 |
| Total | 0.92 | 0.9 | 0.91 |

<Table 4>는 정규식 기반 단위 감지기의 성능을 수동 레이블링 된 평가 데이터 기준으로 측정된 결과를 보여 주고 있다. 맥락상 애매한 표현을 감지 대상에서 사전 배제한 설계로 오탐(False Positive)을 최소화하였으며, 명확한 단위 표현에 대해 높은 감지 정확도를 달성하였다.

4.2.3 번역 품질

번역 엔진의 성능 평가를 위해 200개 문장을 무작위 샘플링하여 BLEU(Bilingual Evaluation Understudy) Score를 측정하였다. BLEU Score는 기계 번역 분야에서 널리 사용되는 자동 평가 지표로, 번역 결과와 정답 번역 간의 n-gram 일치율을 기반으로 번역 품질을 정량적으로 측정하며 식은 아래와 같다.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

BP(Brevity Penalty)는 번역문이 기준 번역문보다 지나치게 짧을 때 점수를 낮추기 위한 값이다. p_n 은 n-gram 정밀도를 의미하며, 번역 결과에 포함된 연속된 단어 묶음이 기준 번역문과 얼마나 일치하는지를 나타낸다. w_n 은 각 n-gram 정밀도에 부여되는 가중치다.

<Table 5> Translation Quality Evaluation

| BLEU-N | Score |
|--------|--------|
| BLEU-1 | 0.5124 |
| BLEU-2 | 0.3847 |
| BLEU-3 | 0.2935 |
| BLEU-4 | 0.2214 |

<Table 5>는 n-gram 단위별 번역 품질 평가 결과를 보여주고 있다. 단일 단어 일치율을 나타내는 BLEU-1은 0.5124로 번역된 단어의 절반 이상이 정답 번역과 일치함을 확인하였다. 연속된 두 단어의 일치율을 나타내는 BLEU-2는 0.3847, 세 단어 연속 일치율인 BLEU-3는 0.2935로 측정되었으며, 기계 번역 평가의 표준 지표로 활용되는 BLEU-4는 0.2214로 측정되었다. 일반적으로 BLEU-4 기준 0.2 이상은 번역 결과가 어느 정도 이해 가능한 수준임을 나타낸다. 또한, 번역이 진행되면서 단위가 변환되는 변환 공식을 기반으로 변환 결과와 실제 정답 값 간의 평균 오차율 0.11%로 측정되었다. 본 논문에서 제안한 시스템은 범용 번역이 아닌 단위 표현이 포함된 문장의 번역과 단위 변환을 목적으로 한 시스템으로, 약 40,000개의 제한된 학습 데이터 규모를 고려할 때 해당 수치는 단위 변환 번역에서 일정 수준의 번역 품질을 보여준다고 할 수 있다.

5. 결론

본 논문은 기계 번역 과정에서 발생하는 문화권별 단위 체계 불일치 문제를 해결하기 위해 5개 모듈로 구성된 Transformer 기반 단위 변환 통합 번역 시스템을 제안하였다. 제안 시스템은 언어 감지, 지역 감지, 단위 감지, 신경망 번역, 단위 변환 후처리의 파이프라인 구조를 통해 입력 문장의 문화적 맥락을 파악하고 대상 문화권에 적합한 단위로 자동 변환된 번역 결과를 제공한다. 본 연구의 주요 기여점은 세 가지이다. 첫째, 단위 감지 모듈(모듈 3)을 독립적으로 설계하여 정규식 기반의 높은 감지 정확도(F1-Score 0.91)와 수식 기반의 정밀한 단위 변환(평균 오차율 0.11%)을 달성하였다. 둘째, 번역 결과를 직역, 변환, 병행의 세 가지 모드로 동시에 생성하여 사용 목적과 문서 유형에 따라 유연하게 활용할 수 있는 구조를 마련하였다. 셋째, 플레이스홀더 치환 방식을 적용하여 Transformer 번역 엔진이 단위 위치를 보존하면서 자연스러운 번역을 생성할 수 있도록 하였다.

향후 연구로는 본 논문에서 길이와 온도 단위로 한정된 처리 범위를 무게(pound ↔ kg), 부피(gallon ↔ L), 넓이(acre ↔ m²) 등으로 확장하여 보다 범용적인 단위 변환 번역 시스템으로 발전시킬 계획이며, BERT, mBART 등 사전 학습 모델을 활용한 Fine-tuning 방식을 적용하여 현재 학습 데이터 규모의 한계로 인한 번역 품질 문제를 개선할 예정이다. 또한 단위 감지를 독립 모듈로 분리하는 본 연구의 방식과 Transformer가 단위를 자체적으로 학습하는 End-to-End 방식 간의 성능 비교 실험을 수행하여 두 접근 방식의 장단점을 분석하여 번역의 품질을 향상시킬 예정이다.

REFERENCES

- [1] A.Vaswani, N.Shazeer, N.Parmar, J.Uszkoreit, L.Jones, A.N.Gomez, L.Kaiser and I.Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 30, 2017.
- [2] A.Araabi, and C.Monz, "Optimizing transformer for low-resource neural machine translation," In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 3429-3435, 2020.
- [3] E.Michon, J.M.Crego, and J.Senellart, "Integrating domain terminology into neural machine translation," In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 3925-3937, 2020.
- [4] J.Tracz, P.I.Wójcik, K.Jasinska-Kobus, R.Belluzzo, R.Mroczkowski, and I.Gawlik, "BERT-based similarity learning for product matching," In *Proceedings of Workshop on Natural Language Processing in E-Commerce*, pp. 66-75, 2020.
- [5] L.Tan, M.Y.Li and S.Kok, "E-commerce product categorization via machine translation," *ACM Transactions on Management Information Systems (TMIS)*, 11, 3, pp. 1-14, 2020.
- [6] B.Zhang, T.Nakatani, and S.Walter, "Enhancing e-commerce product title translation with retrieval-augmented generation and large language models," *arXiv preprint arXiv:2409.12880*, 2024.
- [7] R.Appicharla, A.Ekbal, and P.Bhattacharyya, "EduMT: Developing machine translation system for educational content in Indian languages," In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, pp. 35-43, 2021.
- [8] L.Han, S.Gladkoff, G.Erofeev, I.Sorokina, B.Galiano, and G.Nenadic, "Neural machine translation of clinical text: an empirical investigation into multilingual pre-trained language models and transfer-learning," *Frontiers in Digital Health*, Vol. 6, 1211564, 2024.
- [9] E.Rancati, I.Abashidze, and A.D'AGATA, "AI-based Machine Translations in Tourism and Hospitality: An Exploratory Study of Travelers' Perceptions and Future Challenges," *European Scientific Journal*, Vol. 21(7), pp. 1-13, 2025.
- [10] International Bureau of Weights and Measures, D.B.Newell and E.Tiesinga, "The international system of units (SI)," *Physical Measurement Laboratory National Institute of Standards and Technology*, pp. 1-138, 2019.
- [11] N.Edward, "Men and measures: a history of weights and measures, ancient and modern," *Smith, Elder & Company*, 1912.
- [12] M.Tu, Y.Zhou, and C.Zong, "A universal approach to translating Numerical and Time Expressions," In *Proceedings of the 9th International Workshop on Spoken Language Translation: Papers*, pp. 209-216, 2012.
- [13] H.H.Nigatu, N.Mehandru, N.H.Abadi, B.Gebremeskel, A.Alaa, and M.Choudhury, "Viability of Machine Translation for Healthcare in Low-Resourced Languages," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 10595-10609, 2025.
- [14] G.Dinu, P.Mathur, M.Federico, S.Lauly and Y.A.Onaizan, "Joint Translation and Unit Conversion for End-to-end Localization," In *Proceedings of the 17th International Conference on Spoken Language Translation*, pp. 265-271, 2020.
- [15] M.Qian and Y.Kong, "Exploring the advantages and challenges of a concept-guided approach in large language model aided machine translation: integrating generative AI

and human-like cognition," Proceedings of the 16th Conference of the Association for Machine Translation in the Americas, Vol. 1: Research Track, pp. 55-72, 2024.

- [16] X.Zhang, J.Zhao, Y.LeCun, "Character-level Convolutional Networks for Text Classification," Advances in neural information processing systems, 28, 2015.

허 지 욱(Jee-Uk Heu)

[정회원]



- 2016년 2월 : 한양대학교
컴퓨터공학과 (공학박사)
- 2017년 2월 ~ 2017년 12월 :
Stanford Center for
Professional Development
연구원

- 2021년 3월 ~ 2025년 2월 : 강남대학교 참인재대학
교양교수부 조교수
- 2025년 3월 ~ 현재 : 강남대학교 인공지능융합공학부
조교수

<관심분야>

빅데이터, 문서요약, 인공지능, 자연어처리, 영상처리