

## LSTM을 활용한 수어 단어 인식

정의손<sup>o</sup>, 조동휘\*, 박세희\*, 강현아\*, 박승보\*

<sup>o</sup>인하대학교 소프트웨어융합공학과,

\*인하대학교 소프트웨어융합공학과

e-mail: {justicehand1999<sup>o</sup>, junejoe97\*, SW\_ParkSehee\*}@inha.edu,  
ha.adastra@gmail.com\*, molaal@inha.ac.kr\*

## Recognition of Sign Language using LSTM

Eui-Son Jung<sup>o</sup>, Dong-Hee Jo\*, Se-Hee Park\*, Hyeon-Ah Gang\*, Seung-Bo Park\*

<sup>o</sup>Dept. of Software Convergence Engineering, INHA University,

\*Dept. of Software Convergence Engineering, INHA University

### ● 요약 ●

기존 수어 통역 서비스의 문제점들을 보완하고 대체할 수 있는 수단이 필요하다. 본 논문에서는 LSTM을 활용한 수어 단어 번역을 제안한다. 데이터는 총 단어의 개수 56개를 사람 16명으로, 그리고 관측 방향을 5개로 나누어 총 4480개의 단어별 영상을 사용하였다. 학습 횟수 250회 검증 정확도는 78%에 도달하였다.

**키워드:** 수어, 청각장애, 미디어 파이프, 골격 추출, 딥러닝

### I. Introduction

농인과 청인이 주로 사용하는 의사소통 방법으로는 수어가 55.9% 구화가 21.4%, 기타 11.7%, 필담 6.0% 스마트 기기가 0.4%로 거의 모든 농인은 수화를 사용한다[1]. 농인과 청인의 의사소통은 수어 통역 서비스를 받을 수 있지만, 수어 통역사의 공급이 제한적이기 때문에 대기시간이 길고 혹은 부담스러운 수어 통역비로 인해 수어 통역 서비스를 일상생활에서 손쉽게 이용할 수 없다. 수어 통역 서비스를 이용하지 않는 경우, 주로 필담을 이용하여 의사소통이 이루어 지지만 농인과 청인 모두에게 익숙하지 않은 방법이므로 이를 대체할 수 있는 수단이 필요하다[2]. 수어 통역사를 대체하기 위한 서비스를 구성하기 위해서는 선제적으로 수어 문장에 대한 인식이 필요하고, 더 앞서서 단어의 정확한 인식이 필요하다. 그리고 수어는 의사소통이며 시계열 데이터이다. 따라서 본 논문은 시계열 데이터처리에 적합한 RNN 기반의 모델 중 가장 높은 정확도를 보인 LSTM을 활용한 수어 단어 번역을 제안한다.

### II. Preliminaries

#### 1. Mediapipe Skeleton Detection

Mediapipe의 Skeleton Detection은 구글에서 제공하는 서비스로 동영상 속 대상의 골격 추출을 진행하여 각 골격의 관절 별 위치를 제공한다[4]. 본 논문에서는 이를 이용하여 대상의 배경 및 환경적 요소 차이를 줄였다.

#### 2. LSTM (Long Short-Term Memory)

RNN은 시계열 데이터를 효과적으로 처리하기 위해서 만들어진 모델이며, LSTM은 기존 RNN 모델이 과거 데이터를 현재에 연결이 어려운, 즉 장기 의존성 개선을 위해 메모리 셀을 개선한 모델이다.

### III. System Scheme of Sign Language Recognition

#### 1. System Overview

전체 영상에서 단어 부분의 영상을 골격 추출 후 벡터를 모델에 넣는다. 모델은 LSTM 2층 레이어로 구성되었으며 그 후 두 번의 Fully Connected를 통과한다. 활성화 함수로는 SoftMax 함수를 사용하였다.

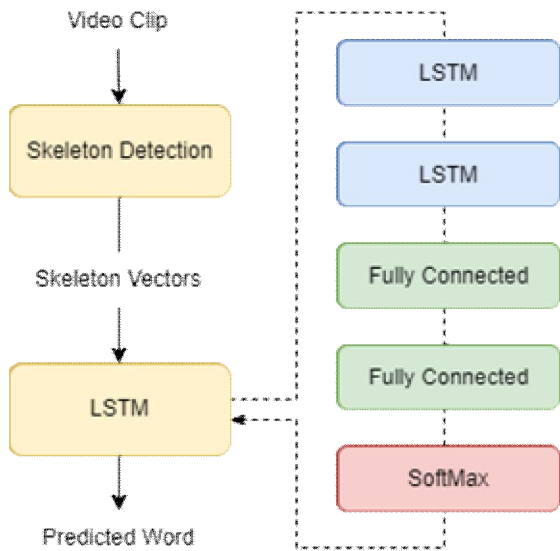


Fig. 1. System Architecture

## 2. Data Before & After

Table 1. Changed Data Count

	초기 데이터	보강 데이터	차이 점
단어의 개수	56	56	0
단어표현 사람	4	16	12
관측 방향	1	5	4
총 계수	224	4480	4256

초기 데이터의 구성은 총 단어 56개를 사람 4명 관측 방향 1방향으로 224개의 데이터를 사용하였지만, 정확도 40%가 나왔다. 그래서 데이터 보강작업을 진행하였고 보강 내용은 총 단어의 개수 56개를 사람 16명으로, 관측 방향을 5개로 나누어 총 4480개의 영상으로 증가하였다.

## 3. Experimental Result

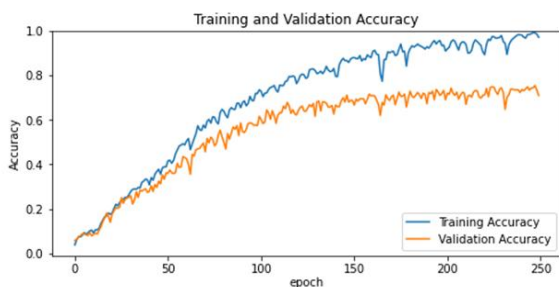


Fig. 2. Experimental Result

변경된 데이터 기준으로 학습을 진행하였을 때 학습 횟수 250회 검증 정확도는 78%에 도달하였다. 22%의 오차값이 발생한 이유는 데이터의 부족 현상으로 사료 되며 추후 보강할 필요가 있다.

## IV. Conclusions

본 논문은 수어 인식을 위해 LSTM 기반의 수어 자동 인식 방법을 제안하고 구현하였다. 이 모델의 정확도는 78%로 나타났다. 정확도가 22% 부족한 이유는 학습으로 사용된 데이터의 부족으로 추정되며, 향후 데이터를 보강하여 정확도를 향상할 필요가 있다. 또한, 번역된 단어를 기반으로 문장화시키는 학습 모델을 추가 구현할 예정이다.

## ACKNOWLEDGEMENT

이 논문은 2021년도 정보(교육부)의 재원으로 한국연구재단 4단계 두뇌한국(BK)21 사업 대학원 혁신지원을 받아 수행된 연구임

## REFERENCES

- [1] National Institute of the Korean Language, 2020 Korean Sign Language Utilization Survey, 2020-12, Seoul: Ministry of Culture, Sports and Tourism, 1984, p62, 11-1371028-000851-01.
- [2] Ibid., p85.
- [3] Tao Liu, 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 2871-2875.
- [4] Mediapipe Pose, <https://google.github.io/mediapipe/solutions/pose.html>