유아의 발화를 통한 딥러닝 기반 유아 나이 예측을 통한 발달 단계 모니터링 시스템

Deep learning based children age recognition with speech signals

이지호
AI.소프트웨어학부
가천대학교
성남대로 1342(성남시,대한민국)
vc54@naver.com

함신아
AI.소프트웨어학부
가천대학교
성남대로 1342(성남시,대한민국)
lihena8@gachon.ac.kr

김윤정
AI.소프트웨어학부
가천대학교
성남대로 1342(성남시,대한민국)
yjung203@naver.com

신민서
AI.소프트웨어학부
가천대학교
성남대로 1342(성남시,대한민국)
minseo3347@gachon.ac.kr

최아영*
AI.소프트웨어학부
가천대학교
성남대로 1342(성남시,대한민국)
aychoi@gachon.ac.kr

요 약

최근 영유아의 발달 단계 모니터링을 위한 연구가 진행이 되고 있다. 본 논문에서는 유아의 발화를 통한 딥러닝 기반 유아 나이 예측을 기반으로 유아의 발달 단계 모니터링 시스템을 제안한다. 모니터링 시스템은 안드로이드 기반 앱으로 구현되었으며, 유아의 음성신호를 입력으로 하여 나이를 예측하고 이를 시간에 따라 모니터링이 가능하도록 하였다. 나이 예측 알고리즘은 음성 데이터를 문장으로 변환하고 불용어 처리, 토큰화, 단어 임베딩, 센텐스 임베딩, 딥러닝 기반 인식 단계를 거쳐서 나이를 예측하도록 하였다. 제안한 방법은 향후 영유아의 발달 단계를 자동으로 모니터링 하고 기록하는 서비스에 응용이 가능하다.

키워드 : 영유아 발달단계, 나이 예측, 자연어 처리, 딥러닝

1. 서론

최근 유아 발달과 관련한 연구가 진행이 되고 있다. 최근 유아 발달과 관련한 연구의 필요성 증가로 인해 많은 연구가 진행되고 있으며 국가에서는 영유아 검사와 관련된 지원을 하고 있다. 영유아의 발달 평가를 위해서는 2014 년에 개발한 K-DST 발달 선별 도구를 주로 사용하고 있다. 이 도구는 행동, 인지, 언어, 사회성 지표를 통해서 유아의 행동 발달을 추정한다. 하지만 문진표의 응답은 아예 하지 못함과 잘함을 0 에서부터 3 의 정도로 분류하여 체크하는 형식으로 이뤄져 있다. 이는 모든 아이에 대해 수치로 기준이 정해진 것이 아닌 부모의 주관적인 관점에서 평가가 이뤄지기 때문에 아이의 발달 단계를 객관적으로 측정하기가 어렵다. 또한 영유아 검진은 병원 측과 수검자의 부모의 어려움으로 인해 미 수검률이 존재하고 있다. 특히 COVID-19 로 인해 마스크의 착용이 의무화되면서 발화 시, 사람들의 입 모양을 볼 수 없게 되고 미디어 매체의 접근이 증가하면서 아이의 언어 발달 속도가 늦어지고 있다. 차세대컴퓨팅의 발달로 인하여 사람의 발화를 텍스트로 변환해 인식하여 구분하는 것이 가능해지며 이 기술들을 활용해 문제를 해결하면 보다 객관적인 발달 평가가 가능하다.

2. 관련연구

최근 많은 연구에서 영유아의 발달을 모니터링하고 평가하기 위한 연구가 진행되고 있다. 신손문 등은 현행 영유아 검진에 대한 문제점을 분석하는 연구를 수행하였다[1]. 2019 년 건강보험 가입자 기준 영유아 검진의 미 수검률은 22.5%로 결과 상담이 충분하지 않고, 검진 항목의 과다한 이유 등으로 미수검율이 발생한다고 보고하였다. 따라서 이의 해결을 위해서는 객관적이고 쉬운 검진 방법에 대한 도입이 필요하다. 김영태 등은 한국 2~4 세 아동 발화 길이에 대한 기초 연구를 통해 나이에 따른 발화 길이에 차이가 있음을 보고하였다 [2]. 성미영 등은 유아의 성별과 놀이 상황 유형별 평균 발화 길이와 어휘가 다양하며 평균 발화 길이나 최장 발화 길이는 연령에 따라 민감한 언어 발달의 척도가 될 수 있음을 나타냈다 [3]. 이를 기반으로 고려하면 영유아의 발화 길이, 사용하는 어휘 등을 기반으로 나이 등의 발달단계를 예측하는 것이 가능하다.

최근에 영유아의 발달 단계를 모니터링 하고 이를 인식하는 연구들이 수행 되었다. Tarun Gupta 등은 나이 및 키 추정을 위한 바이인코더 트랜스포머 혼합 모델을 제안하였다. 실험 결과 남성과 여성 모두 5.54 년과 6.49 년의 RMSE(Root Mean Squared Error)를 보였다. 유아의 발달 평가를 위해서 Ruben Zazo 등은 LSTM-RNN 기반의 연령 추정 시스템을 제안하였다. 이 시스템은 짧은 발화(3~10 초)를 입력으로 하여 예측을 수행하였고 남성의 경우 평균 오차는 입력 시간을 3 초, 5 초, 10 초로 실험하였을 때 남성은 약 7.72 년~8.79 년, 여성은 약 6.97 년~7.95 년의 오차가 나타났다. 7 년 이상의 예측 오차가 발생하기 때문에 3 세~12 세 사이의 영유아 대상의 나이 예측 관련한 연구는 활발히 진행되지 않았다.

3. 실험방법

3.1. 데이터셋

본 연구에 사용된 데이터는 다음과 같다. 본 연구에 사용된 데이터는 다음과 같다. AI Hub 자유대화 음성(소아남여, 유아 등 혼합)의 데이터를 사용하였으며, 총 1000 명 (남아 약 500, 여아 약 500) 연령별 (3-6 세 약 200 명 7-10 세 약 800 명)의 일상생활에서 발생하거나 데이터 수집을 위한 스튜디오 녹음을 통해 대화가 수집되었다. 소아남여, 유아 등 혼합의 남녀 비율은 1:1 이며 비율의 차이는 10% 미만이다. 3~6 세, 7~10 세와 같이 2 개 클래스 구분하였을 때 클래스의 비율은 1:1 이며 비율의 차이는 10% 미만이다. 또한 지역에 따른 억양과 사투리 및 단어 사투리를 반영하기 위해 다양한 지역을 대상으로 대화를 수집하였다. 수집 방법에 따라 AI 챗봇, 음성수집 도구, 스튜디오 분포를 확인하였다. 각 데이터는 JSON 파일 형식으로 저장이 되어 있으며 본 연구에서는 연령, 성별, 발화 문장을 사용하였다. 본 연구에서 사용한 음성 데이터 형태는 그림 1 과 같다.


```
{
  "발화정보" : {
    "stt" : "이번에 한 번 쉬어야 하네 ",
    "scriptId" : "소아남녀-00101",
    "fileNm" : "소아남여_소아남여01_F_1609501296-1_9_수도권_실내_00101.wav",
    "recrdTime" : "2.730",
    "recrdQuality" : "16K",
    "recrdDt" : "2021-01-27 17:40:43",
    "scriptSetNo" : "T_소아남녀_1"
  },
  "대화정보" : {
    "recrdEnvrn" : "실내",
    "colctUnitCode" : "AI 챗봇",
    "cityCode" : "수도권",
    "recrdUnit" : "AndroidOS",
    "convrsThema" : "보드 게임"
  },
  "녹음자정보" : {
    "gender" : "여",
    "recorderId" : "1609501296-1",
    "age" : 9
  }
}
```


그림 1. 음성 데이터 형태

3.2. 실험환경

본 연구에서는 두 가지 방법으로 유아의 언어 발달 나이를 예측하는 방법을 제안하였다. 제안하는 방법의 절차는 다음과 같다. 공통적인 과정으로 음성 신호를 수집한 데이터로부터 획득하고, STT(Signal to Text) 프로세스를 적용하여 음성을 문자로 변환한다. Python 의 Gensim Word2Vec

을 이용하여 불용어 제거와 각 문장의 토큰화를 진행하였다. 첫 번째 방안인, 딥러닝을 활용한 방안은 추가로 토큰화를 수행해 획득한 단어 사전의 단어들을 100 차원의 임베딩 벡터로 임베딩을 진행하여 각 단어의 벡터를 획득한다. 그렇게 획득한 단어들의 벡터를 이용하여 문장을 임베딩시켜 최종적으로는 문장을 벡터의 배열로 표현한 데이터를 획득하였다. 본 실험은 위에서 획득한 데이터를 다음과 같이 실험하여 제안 방법을 검증하였다.

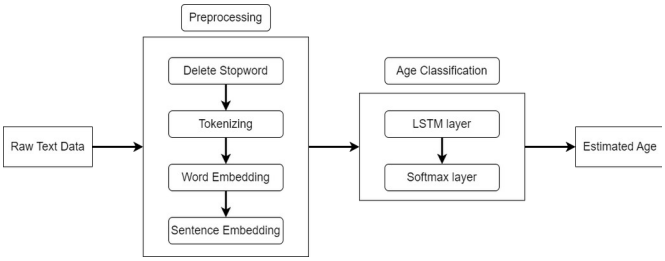

그림 3. 데이터 처리 및 분석 프로세스

딥러닝 모델은 은닉층으로 LSTM layer 와 drop out 을 적용하였고, fully connected layer 를 통해 최종적으로 나이를 인식하도록 하였다. LSTM layer 는 1 개부터 3 개까지 변경하여 실험을 수행하였다. Adagrad optimizer 를 옵티마이저 함수로 Cross Entropy 를 loss 함수로써 사용하였다.

다중 회귀 모델을 이용한 모델은 3~4 세, 5~7 세 남녀 각각 219 명의 데이터를 활용해 관련 논문을 토대로 남녀를 분리하고 아이들의 평균 형태소의 길이, 최장 형태소의 길이, 최단 형태소의 길이를 통해 각각 관계를 파악했으며 최장 형태소의 길이, 평균 형태소의 길이가 성별과 연령대마다 차이가 있는 것을 확인할 수 있었다. 이를 바탕으로 형태소와 단어를 기준으로 다중회귀 모델을 사용하여 아이들의 나이를 예측했다.

4. 시스템 구현 및 실험 결과

제안한 알고리즘을 기반으로 유아 발달단계 모니터링을 위한 시스템을 구축하였다. 앱은 안드로이드 기반으로 구성하였으며, 로그인 기능, 언어발달을 모니터링하는 기능, 행동 발달을 모니터링 하는 기능과 시간에 따른 발달 상태를 파악할 수 있는 그래프로 구성이 되었다.

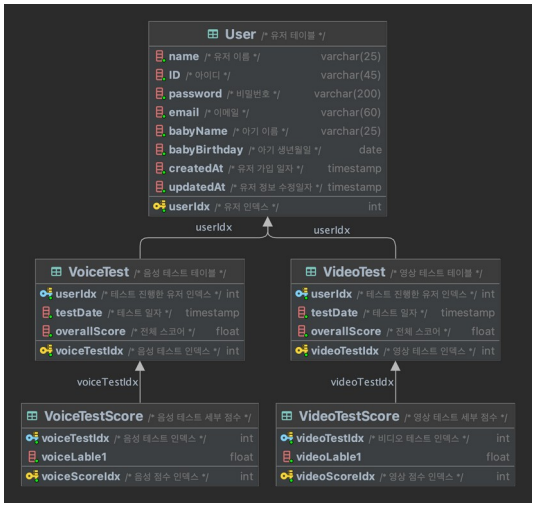

그림 4. 데이터베이스 구축 현황

그림 5. 앱 개발 결과

제안한 알고리즘에 대한 성능을 분석한 결과 LSTM layer and 1 softmax layer 를 사용하여 분석한 결과 정확도는 약 24% 정도 획득이 되었다. LSTM layer 3 개와 fully connected layer 을 적용한 정확도는 31.6%로 획득이 되었다. 다중 회귀 모델의 경우 정확도는 약 26% 정도 획득되었다. 분석 결과 실험 데이터가 균등하지 않아 높은 정확도를 보이지 않음을 확인할 수 있었다. 3~6 세 아동은 200 명이 실험에 참여하였으며, 7~10 세 아동은 800 명 참여하여 전체 데이터 중 80%의 데이터가 7~10 세 데이터로

획득되었다. 따라서 모델이 편향되어 적은 수의 데이터 구간에서 정확도가 매우 낮은 문제점이 있었다. Ruben Zazo 등의 연구에서 나타난 실험 결과는 남성은 7.72 년~8.79 년, 여성은 약 6.97 년~7.95 년의 평균 예측 오차를 고려할 때 기존 연구 대비 큰 성능의 저하는 보이지 않는 결과로 판단된다. 성능 개선을 위하여 나이를 반영할 수 있는 특징을 추출하고 이를 입력으로 활용하는 것이 필요하다. 특히, 형태소와 최장 길이의 형태소에서는 연령별로 차이를 보이기 때문에 이러한 지표를 고려하는 것이 필요하다. 또한, 형태소, 단어와 같은 언어적 요소뿐만 아니라 인지적, 사회적 요소들도 같이 파악을 통해 좀 더 정확한 추정이 가능할 것으로 보인다. 데이터의 확보와 환경에 대한 변인(익숙한 상황, 낯선 상황), 용언의 활용과 같은 요소들을 추가하고 다른 모델들을 더 추가한다면 더 좋은 결과를 기대할 수 있을 것이다.

5. 결론

본 논문에서는 유아의 발화를 통한 딥러닝 기반 유아 나이 예측을 기반으로 유아의 발달 단계 모니터링 시스템을 제안한다. 모니터링 시스템은 안드로이드 기반 앱으로 구현되었으며, 유아의 음성신호를 입력으로 하여 나이를 예측하고 이를 시간에 따라 모니터링이 가능하도록 하였다. 나이 예측 알고리즘은 음성 데이터를 문장으로 변환하고 불용어 처리, 토큰화, 단어 임베딩, 센텐스 임베딩 단계를 거쳐서 나이를 예측하도록 하였다. LSTM 알고리즘과 다중 회귀 분석 알고리즘을 적용하여 예측한 결과 각각 24%, 26%의 정확도를 보였다. 분석 결과의 정확도가 낮은 문제가 있는데 적은 수의 국내 아이들의 데이터 셋을 통한 딥러닝과 머신러닝은 다양한 변인이 존재하는 언어 발달 단계를 예측하는데 높은 정확도를 보이지 않은 것으로 판단된다. 또한 인지적, 사회적 요소들도 같이 파악되어야 하는데 이는 데이터로 잘 나타나지 않아 변인의 요소가 부족하여 정확한 추정을 하는데 부족함이 있는 것으로 보인다. 데이터의 확보와 환경에 대한 변인(익숙한 상황, 낯선 상황), 용언의 활용과 같은 요소들을 추가하고 다른 모델들을 더 추가한다면 더 좋은 결과를 기대할 수 있을 것으로 보인다.

* 교신저자

Acknowledgement

이 논문은 2020 년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구이며 (2020S1A5A2A03041734), 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2021R1F1A1062181).

참 고 문 헌

[1] 신손문, 최병민,최지은, “현행 영유아건강검진의 문제점 분석 및 개선 방안 연구, 2017 보건복지부 연구보고서

[2] 김영태, “한국 2- 4 세 아동의 발화길이에 관한 기초연구,” 말-언어장애연구, 제 2 권, pp.5-26, 1997.

[2] 성미영, 장문수, “유아의 성별과 놀이 상황 유형별 평균 발화 길이와 어휘 다양도,” 한국보육지원학회지, 제 10 권 제 6 호, pp. 43-56, 2014.

[3] Gupta, Tarun, Duc-Tuan Truong, Tran The Anh, and Chng Eng Siong. "Estimation of speaker age and height from speech signal using bi-encoder transformer mixture model." arXiv preprint arXiv:2203.11774 (2022).

[4] Zazo, Ruben, Phani Sankar Nidadavolu, Nanxin Chen, Joaquin Gonzalez-Rodriguez, and Najim Dehak. "Age estimation in short speech utterances based on LSTM recurrent neural networks." IEEE Access 6 (2018): 22524-22530.