

음성인식 개발 프로젝트 위험 요인에 관한 연구

김선교*, 박준호*

*(주)솔루게이트 기업부설연구소

e-mail: {pcrlatjsry, park1058}@solugate.com

A Study on Risk Factors of Speech Recognition Development Project

Sunkyo Kim*, Joonho Park*

*Research and Development Laboratory, Solugate Ltd.

요 약

음성인식은 음향모델, 언어모델, 디코더 등의 기술을 이용한다. 음성인식은 하드웨어와 소프트웨어 구성이 정확하게 설계가 되어야 한다. 음성인식 프로젝트는 인프라 구성과 도입되는 음성인식 엔진도입, 인식률 그리고 시스템과의 연계가 중요하다. 하지만 음성인식 프로젝트는 솔루션 도입으로 인지하고 수행할 경우에는 많은 위험이 발생한다. 이 중 가장 문제가 되는 것이 인식률이다. 본 논문에서 음성인식 개발 프로젝트 수행에 도출되는 인식률을 개선하는 방안을 제시하겠다.

1. 서론

음성인식 프로젝트는 최근 코로나19 이후에 지속적으로 산업이 발전하고 있다. 음성인식 기술은 휴대폰에서 사람의 말로 기능을 동작하는데 많이 이용하고 있다. 또한, 최근에는 콜센터에서 고객과의 대화 내용을 음성인식 기술을 통해 데이터화 하고 있다. 공공기관에서는 민원 콜센터의 데이터 분석과 음성인식 기술을 이용한 민원 처리 업무에 적용하고 있다. 하지만 음성인식 프로젝트를 성공적으로 수행하기 위해서는 학습을 위한 인프라 환경과 음원 데이터 수집 및 전사작업이 진행되어야 한다. 이를 통해 음성인식의 인식률(정확도)을 측정한다. 본 논문에서는 음성인식 프로젝트에서 이슈가 많이 되고 있는 인식률(정확도) 작업을 시스템으로 구축하는 방안을 제시한다.

2. 관련연구

본 논문에서는 공공 프로젝트와 민간에서 수행하는 프로젝트별 위험요인 중 인식률 향상을 위한 방법에 대해 분석한다.

전사작업을 진행하고, 확보된 데이터를 통해 인식률을 측정하고, 최소의 비용과 시간을 통해 인식률을 향상하는 방안에 대해 연구한다.

2.1 음성인식 프로젝트 위험 요인 분석

본 논문에서는 음성인식 프로젝트 위험 요인에 인식률과 학습, 정확도 측정에 대해 제시한다. 인식률은 음성인식 엔진의 인식률을 의미한다. 몇 개의 단어가 틀린 것을 기준으로 계산한다. 일반적인 음성인식 엔진의 경우 평균

정확도는 75% 수준이다. 하지만 음성인식 프로젝트의 인식률 목표는 90 % 이상을 요구하는 것이 위험요소이다. 인식률 향상을 위해서는 음향 모델, 언어 모델을 학습 시켜야 한다. 음향인 경우에는 음성 데이터를 전처리하여 학습 데이터를 만들어야 한다. 언어 모델의 경우에도 실제 데이터를 전처리 하여 학습 데이터셋을 만들어야 한다. 학습을 하기 위해서는 GPU가 있는 서버에서 학습을 진행해야 한다. 하지만 일부 인프라의 경우 GPU가 없는 경우가 많이 있다. 마지막으로 전사 작업이다. 학습 데이터 구축을 위해서는 음성 데이터와 텍스트가 데이터가 하나의 셋으로 구성해야 한다. 이를 위한 작업으로 전사 작업이 진행된다.

2.2 제안하는 방식

본 논문에서는 음성인식 프로젝트의 위험 요인을 제거하고 성공적으로 수행하기 위한 방안을 제안한다. [그림1]

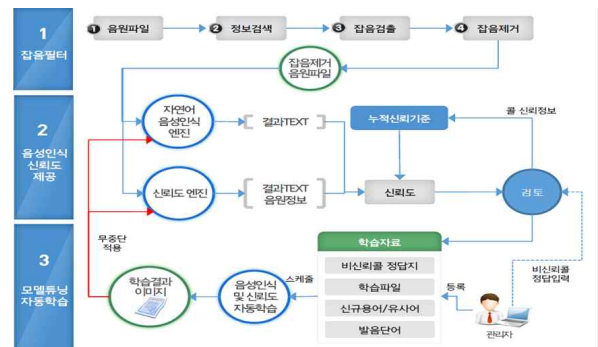


그림 1 구성방안

음성인식 개발 프로젝트를 성공적으로 수행하기 위한 개

발 방법론을 제시하고, 각 단계별 필요한 개발 산출물을 제시한다. 음성인식 개발 프로젝트는 애자일 방법론을 기반으로 위험 요인을 제거한다. 또한 학습 데이터 구축을 위한 데이터베이스 구축 설계에 대한 방안도 제시한다.

3. 음성인식 프로젝트위험 요인 개선 방안

본 장에서는 음성인식 프로젝트에 적용되는 음성인식 엔진은 한국전자통신연구원에서 이전받은 음성인식을 기본으로 사용한다. 프로젝트 개발 방법론은 전자정부 개발 방법론을 사용한다.

음성인식 프로젝트 위험 요인은 2장에서 제시한 사항으로 각 개발 단계별 위험 요인을 제거하는 방법을 제시한다.

3.1 개발 산출물

본 논문에서는 음성인식 프로젝트를 수행하기 위한 기본적인 개발 산출물 리스트[표1]를 제시한다. 일반적인 개발 프로젝트와 유사하지만 인식률 측정 계획서와 결과서가 제시되어 있다. 또한 음성인식 엔진 학습을 위한 전사 작업과 학습 데이터 셋 구성이 포함되어 있다. 전사 작업이 포함되어 있지 않으면 각 도메인에 맞는 음성인식 인식률을 확보하기 어렵다.

표 1. 개발 산출물 리스트

| 단계 | 활동 | 작업 | 개발 산출물 |
|-------|------------|---------------|-------------------|
| 분석/설계 | 요구사항 정의 | 현행 시스템 분석 | 현행 시스템 분석서 |
| | | 요구분석 | 인터뷰 결과서 |
| | | 요구사항 분석서 | 요구사항 분석서 |
| | 아키텍처 설계 | 소프트웨어 아키텍처 정의 | 아키텍처 정의서 |
| | 데이터 베이스 설계 | 데이터 모델링 | ERD |
| | | | 엔티티 정의서/목록 |
| | | | 테이블 정의서/목록 |
| | 테스트 | 테스트 계획 수립 | 테스트 계획서 |
| 개발 | 구현 | 전사 작업 | 학습 데이터셋 구축 |
| | | 모델 학습 | 모델 학습 (음향모델 언어모델) |
| | | 테스트 | 단위테스트 시나리오/결과서 |
| | | | 통합 테스트 시나리오/결과서 |
| | | | 인식률 측정 결과서 |
| | | 인식률 측정 계획서 | 인식률 측정 계획서 |
| 이행 | 시스템 전개 | 교육 계획 수립 | 교육 계획 |
| | | 구현 계획 수립 | 구현 계획서 |

3.2 음성인식 학습 방법 및 측정 방법

본 논문에서 제시하는 음성인식 학습 방법은 음향 모델과 언어 모델에 대해 제시한다. 음성인식 개발 프로젝트에

서 학습은 언어 모델을 중심으로 진행된다. 음향 모델은 언어 모델에 비해 상대적으로 학습 빈도가 적다. 음향 모델은 학습을 위해 음성 데이터와 텍스트 데이터가 함께 구축이 되어야 하며, GPU를 통해 학습이 진행되어야 한다. CPU를 통해 학습도 가능하지만 GPU에 비해 많은 시간이 걸린다.

본 논문에서는 음성인식 프로젝트의 성공을 위해서는 GPU 인프라 구성을 통해 음향 모델을 수행하는 방안을 제시한다. 언어 모델 학습을 통해 개선된 인식률보다 음향 모델 학습을 통해 개선되는 인식률이 더 높게 나타난다. 음성인식 인식률 측정은 정답지가 필요하다. 보통 학습한 데이터 이외의 데이터를 통해 측정하는 경우가 많다. 정확한 측정을 위해서는 학습한 데이터 중 정답지를 별도로 구축하여 측정하는 방안을 제시한다.

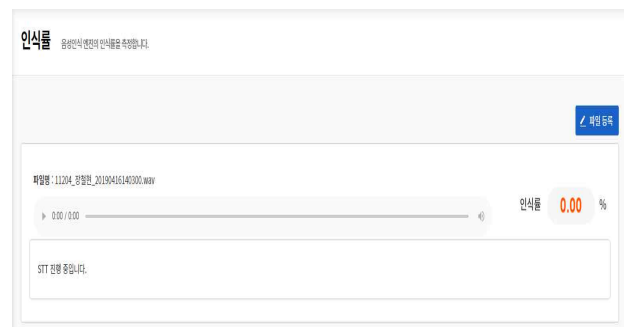


그림 2 인식률 측정 화면

4. 결론

본 논문에서는 음성인식 개발 프로젝트 위험 요인에 대해 제시하였다. 음성인식의 가장 중요한 것은 인식률이다. 하지만 음성인식 인식률은 상용 서비스에는 90% 이상이면 아주 좋은 인식률이라 한다.

향후 본 논문에서는 음성인식 프로젝트의 실증적인 사례 분석과 음성인식 알고리즘별 인식률 향상을 위한 필요한 음성 데이터와 인프라 환경에 대해 제시하고자 한다. 또한 기술이전과 오픈소스 기반의 기술을 활용할 경우 베이스라인 모델 구축을 위한 최소한의 데이터 확보에 대해 제시하고자 한다.

Acknowledgement

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2020년도 문화기술연구개발 지원사업으로 수행되었음

참고문헌

[1] Park,J.H, Lee,N.Y, Park,J.H “An Empirical study of Risk Factors for small-size IT projects”, Lecture Notes in Electrical Engineering, v.421, pp.505 - 510 ,2017