음성 인터페이스의 사용자 관심 정보와 RNN-LSTM을 기반으로 한 예측 서비스

이지훈, 문남미
호서대학교 컴퓨터소프트웨어학
monaminino@gmail.com

Predict Service Based On RNN-LSTM And User Interest Information Of Voice Interface

Ji-Hoon Lee, Nammee Moon
Dept of Computer Software, Hoseo University

요약

음성 인터페이스는 Google Cloud Platform에서 제공하는 STT(Speech To Text)와 TTS(Text To Speech) API를 이용하여 구현한다. 사용자 정보는 프로그램 사용을 위한 회원가입 시 제공받는 정보를 기반으로 검색어를 추천한다. 본 논문에서는 각 사용자의 관심정보에 따라 필요로하는 예측 서비스를 제공하며, 그 예시로 관심정보가 경제 또는 주식일 경우, RNN(Recurrent Neural Networks)의 LSTM(Long Short-Term Memory models)을 활용하여 주가를 예측하는 서비스를 제공한다.

1. 서론

전자기기의 사용량이 많아지고 1인당 보유하고 있는 기기의 개수가 증가하는 현대에 사람들은 본인이 원하는 데이터와 본인에게 딱 맞는 정보를 전달받거나 추천받는 것을 기대한다. 이는 인터넷의 보급을 통해 정보가 많아짐으로 사용자가 본인에게 불필요한 정보를 함께 열람할 수 밖에 없는 환경이 원인으로 작용한다.

모바일 기술의 발전으로 인해 데스크탑 PC가 문서 작성 등 사무용으로 사용된다는 점을 고려하여 PC사용자가 키보드, 마우스 등 기존의 인터페이스를 사용함과 동시에 음성으로 명령을 전달할 수 있는 음성 인터페이스를 제안한다. 그 중 검색 기능에서 이루어지는 데이터 필터링은 음성 인터페이스 사용을 위한 회원가입이 이루어 질 때 제공받는 성별, 나이, 거주지 등의 정보를 토대로 크롤링 정보의 필터링과 추천된 검색어가 제공된다.

본 논문에서는 Google Cloud Platform에서 제공하는 STT와 TTS API를 활용한 음성 인터페이스의 설계 및 구현 사항을 설명하며, 사용자 관심 정보가 경제 항목일 경우, 예시로 RNN의 일종인 LSTM을 활용하여 Apple사의 주가를 예측해본다. 2장에서는 음성 인터페이스와 RNN의 LSTM에 대해 살펴보며 3장에서 음성 인터페이스의 설계 및 구현과 Apple사의 주가 예측 테스트 과정을 확인한다. 4장에서는 위 시스템의 테스트 결과 및 기대효과에 대해 다룬다.

2. 관련 연구

사람이 의사소통을 위해 사용하는 가장 일반적이며 효과적인 수단은 음성이다. 인터페이스는 서로 다른 두 시스템에서 서로 정보나 신호를 주고받는 경우의 접점이며 서로를 연결하는 장치이다. 이 음성 인터페이스는 사람이 다른 시스템과의 의사소통에 있어 가장 효과적인 수단이 될 수 있다.

현재 음성 인식으로 대화형 서비스를 진행중인 기기로는 마이크로소프트사의 '코타나'와 애플의 '시리', 카카오 '미니'와 삼성의 '빅스비' 등이 있다. 마이크로소프트사의 '코타나'를 제외한 대부분의 기기들은 모바일의 음성 인식 비서 시스템으로 서비스중이며, '코타나'는 현재 한국을 제외한 해외에서 윈도우10을 기반으로 서비스중이다. 위 서비스들의 공통 핵심 기술로는 STT(Speech To Text)와 음성 합성 (Speech To Text) 기술이 있다.

먼저 STT(Speech To Text)는 과거 미리 저장해 둔 음성 패턴과 비교하여 개인 인증 등의 용도로 사용하는 '화자 인식'을 이용한 음성 인식 서비스가 주를 이루었다면, 현재의 음성인식은 대표적으로 HMM(Hidden Markov Model)[1]을 활용하여 여러 사용자들이 낸 음성을 통계적으로 모델링하여 음향모델을 구성한다. 음성 인식 기술의 기본적 원리는 마이크로 받은 음성 신호를 받은 후 해당 음성신호의 특징을 추출한다. 그 후 음성모델을 이용한 거리 계산을 하는데, 음성모델 학습을 통한 Database를 거리계산을 위한 정보로 사용한다.[2]

이러한 인식 기술을 토대로 키보드 대신 문자를 입력하는 방식이며 STT API(Application Programming Interface)는 제공 업체별로 차이가 있다. 제공하는 업체에 따라 각각 API의 장단점과 특화된 기능이 다르다.

Open API(Application Programming Interface)란 누구나 사용할 수 있도록 공개된 API이다.[2] Open API를 사용하여 개발자는 공개된 API를 토대로 프로그램에 필요한 기능들을 선택하여 구현하는 것으로 개발시간을 단축할 수 있다. Open API를 제공하는 업체는 대표적으로 Google과 Naver, Kakao, Watson이 있으며 아래의 [표 1]에서는 이 중 Google의 Google Cloud Speech API, 네이버의 클로바, Watson의 Watson Speech to Text API를 비교했다.

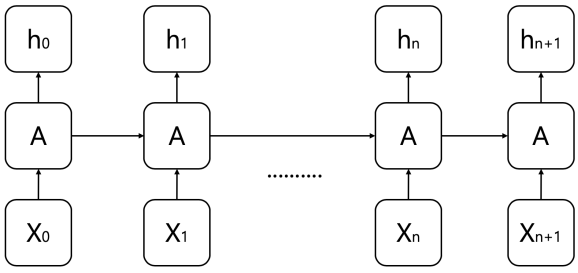
[표1] STT Open API 차이

	Google Cloud Speech API[3]	카카오 뉴톤 [4]	Watson Speech to Text API[5]

지원 언어	한국어 포함 80여개 언어	한국어	한국어 포함 10개 언어 (베타 미포함)
기술 지원범위	SDK 제공, API 문서 제공, 설치 및 설정 가이드, 설정 최적화 예시, 샘플 어플리케이션 제공 등		
무료 서비스 기간	1시간/1개월 무료	일일 제공량 20,000건 (이후 초과 사용량 제휴를 통해 증량)	Lite Plan 기준 500분/1개월 무료
활용 분야	스마트폰, PC, 태플릿, IoT기기	스마트폰	웹

두 번째로 TTS(Text To Speech)로 불리는 음성 합성 기술은 음성을 기계가 음파로써 만들어 내는 기술이며 사람의 말소리를 녹음하여 음성의 분절음을 합성한다. 이 분절음을 합성할 때에는 각 분절음의 경계를 토대로 앞 음성의 뒷부분과 뒤 음성의 앞부분을 함께 저장하는 diphone 처리를 한다. TTS API는 위의 STT API에서 제공하는 업체에서 함께 제공한다. 본 인터페이스에서는 PC환경에서 구현하기 때문에 Google Cloud Speech API를 사용하였다.

RNN(Recurrent Neural Networks)는 Hidden Node가 방향을 가진 Edge로 연결되어 순환구조를 이루는 인공신경망의 한 종류이다. RNN은 기존 신경망 모델에서 반복성의 특징을 가지게 되며, 과거의 데이터가 미래에 영향을 줄 수 있는 구조를 가지고 있다.



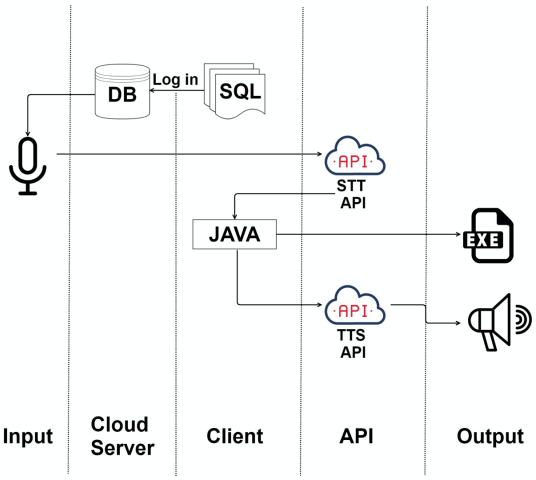
[그림1] RNN의 장기 의존성

위 [그림1]에서 보이는 것처럼, RNN은 장기 의존성 문제를 가지고 있다. RNN은 매 단계마다 위 과정을 반복하며 역전파 시 더 많은 연산에 따른 경사 감소로 뒤의 노드까지 영향을 끼치지 못한다.

LSTM(Long Short-Term Memory models)는 위의 장기 의존성 문제를 해결하기 위해 Forget Gate, Input Gate, Output Gate를 통해 Cell State에 정보를 반영한다.[6]

3. 설계 및 구현

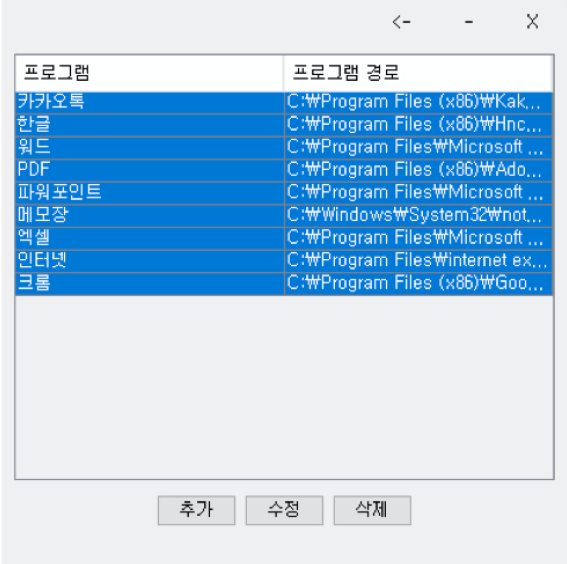
음성 인터페이스는 Google Cloud Platform의 Speech API를 사용하였다.



[그림2] 음성 인터페이스 구성도

사용자는 음성 인터페이스를 사용하기 위해 회원가입을 해야하며, STT API를 통해 명령어를 입력하고 인터페이스는 명령어를 통해 실행파일을 실행하거나, TTS API를 통해 음성으로 명령어를 수행한다.

명령어는 크게 5가지로 구성되며, 읽기, 쓰기, 실행, 종료, 검색의 기능이 있다. 먼저, 읽기 명령어는 사용자가 읽고자 하는 내용을 TTS API를 통해 피드백하며, 쓰기 명령어는 메모장을 기반으로 사용자의 음성 데이터를 "저장"이라는 명령어가 나오기 전까지 입력한다. 실행 명령어는 사용자가 가지고있는 실행 파일의 경로를 저장하여 해당 실행파일이 존재하는 경우 실행시키는 명령어다. 각 사용자마다 가지고 있는 실행파일의 경로는 모두 다르기 때문에 이는 아래 [그림3]과 같이 각자 경로를 지정해 주어야 한다.



[그림3] 실행 파일 경로 지정 GUI

검색 명령어는 사용자가 피드백 받고자 하는 내용에 따라 사용 라이브러리가 달라진다. 사용자가 실시간으로 변하는 정보나 직접적인 웹페이지를 보고자 하는 "보여줘" 또는 "띄워줘"와 같은 명령어는 웹 드라이버 의존성이 높은 Selenium 라이브러리를 통해 크롤링 결과창을 팝업하며, "읽어줘", "들려줘"와 같은 즉각적인 음성 피드백을 필요로 할 경우 BeautifulSoup 라이브러리의 자바 버전인 Jsoup을 사용하여 크롤링된 문장을 TTS API로 피드백하였다. [그림4]는 음성 인터페이스인 울트

론을 실행하여 사용하였을 때의 로그이다.

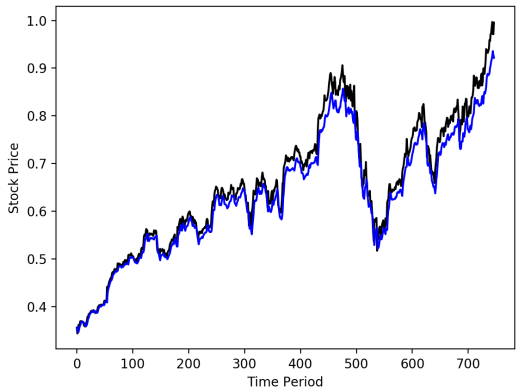
```
울트론 시작
나 : 울트론 인터넷 켜 줘
나 : 울트론 인터넷 꺼 줘
인터넷을 종료합니다.
나 : 울트론 메모장 켜 줘
나 : 울트론 메모장 꺼 줘
메모장을 종료합니다.
```

[그림3] 음성 인터페이스 로그

주가 예측 방법은 Yahoo Finance에서 제공하는 Apple의 주식 정보를 기반으로 진행하였다. 데이터는 2010년부터 2019년 10월 31일까지의 데이터를 기반으로 학습하였으며, 예측한 주가는 명일인 2019년 11월 1일의 데이터를 예측해냈다. 예측에 앞서 Min-Max Scaling을 통해 데이터를 정규화하였으며, 동일한 방법으로 역정규화하였다. 이후 Pandas를 통해 크롤링한 주식 CSV 파일을 불러온다. 손실함수로는 평균제곱오차를 사용하며 학습용 데이터와 테스트용 데이터의 오류를 계속해서 기록한다. 이를 기반으로 matplotlib을 통해 epoch와 예측 결과를 시각화하였다.

4. 결과 및 기대 효과

예측된 주식 정보는 사용자가 필요로 하다면 아래 [그림4]와 같은 그래프를 팝업시켜주며, 불필요한 경우 단지 예측된 주가를 TTS를 통해 음성 피드백 해준다.



[그림4] 예측한 Apple사의 주가 그래프

```
elapsed_time: 0:01:12.083399
elapsed_time per epoch: 0:00:00.072083
```

[그림5] 학습 경과 시간

위 [그림4]에서 검은색은 주가의 실제 데이터이며 파란색은 RNN/LSTM을 활용하여 예측해낸 주가 그래프이다. 그림에서 나타나듯 유사한 모양을 띄고 있으며, [그림5]에서는 전체 학습 시간인 elapsed_time과 각 epoch 별 학습에 경과된 시간을 확인 할 수 있으며, 예측된 값은 아래 [그림6]에서 확인할 수 있듯이 203.4583이다.

```
test_predict [0.79536676]
Predicted stock price [203.4583]
```

[그림6] 학습 결과 및 실제 주가 정보

위 주식 정보는 사용자가 회원가입 시 제공하는 관심사항 정보에 따른 주가 예측 서비스이며, 사용자가 지정하는 관심항목에 따라 각각 다른 추천 시스템을 적용해 볼 수 있다. 예를 들어 뉴스 분야로는 거짓 뉴스 파악이 있을 수 있고, Tech와 같은 과학기술정보로는 메일 API를 추가하여 사용자가 관심있어하는 항목에 새로 업데이트된 내용이 업로드 될경우 메일로 피드백 해주는 기능을 추가할 수 있다.

기존의 키보드나 마우스와 같은 PC 인터페이스를 제외하고 오로지 마이크와 스피커만으로 PC에 명령을 전달할 수 있다는 점과 Crawling 및 RNN/LSTM을 활용하여 정보를 전달할 수 있다는 점을 토대로 이는 비장애인 사용자 뿐만이 아닌 시각 장애를 가지고 있는 사용자에게도 편리함을 제공해 줄 수 있을 것으로 기대된다.

5. 참고문헌

[1] 한국멀티미디어학회지: “음성인식 기술”, 김희린, (2003), 제7권 특집
[2] Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology: “Comparison Analysis of Speech Recognition Open APIs’ Accuracy, 최승주, 김종배, August, (2017), vol.7, No.8, pp. 411-418
[3] https://cloud.google.com/speech-to-text/docs/?hl=ko, API 문서
[4] https://developers.kakao.com/docs/android/speech, API 문서
[5] https://www.ibm.com/kr-ko/cloud/watson-speech-to-text, API 문서
[6] 한국정보과학회: 김영균, “A comparison study on the prediction of the House market price using Simple RNN and RNN/LSTM”, 2154-2156, (2018.6)