1. 과제명: STT, SVM을 이용한 청각장애인용 통화 SW개발

2. 팀원: 25-1 최환종(PL), 25-1 안중환, 25-1 정다비치

3. 아이디어 제안

청각장애인 또는 듣기가 불편한 고령자들은 전화통화에 어려움을 겪는다. 이런 문제에 도움을 주기 위해 첫째로 상대방의 음성을 Text로 변환하여 전화중인 청각장애인에게 보여준다. 두 번째로 듣지 못하는 청각장애인은 텍스트만 보고는 상대방이 어떤 감정으로 말을 하였는지를 알기 어렵다. 우리는 상대방의 음성을 이용해 감정을 분석하여 청각장애인에게 감정까지 보여준다.

4. 개발 방법

1. STT를 이용한 텍스트 변환  
   통화음성을 Android에서 묵음단위로 녹음하여 Core Server로 보낸 후 서버에서 Google STT Server에 Http 요청으로 JSON Object를 얻어오도록 하였습니다. 또한 이 음성을 감정결과를 얻기 위한 원형 Signals로 사용합니다. 또한 실시간 처리와 Connection 관리를 위해서는 GCM방식은 적합하지 않아, Connection Server에 메시지 Stream을 두는 방식으로 끊김에 상관없고 데이터의 손실이 없게 처리하였다.
2. 에너지의 통계적 분포(PLP)를 이용한 지도학습:

* 음성 전처리 및 SVM 지도학습:

EMO DB를 사용하여 0~2700Hz구간에서 6개의 감정자질들을 추출하기 위해 Wav Signal을 FFT이후 피처추출에 적합하도록 정규화하고, PLP 일부 과정을 처리한 후 신호의 세기에 무관하도록 Scaling하였다. 주파수영역으로 바뀐 신호를 6등분하고 각각의 파워를 구하여 특징으로 사용하였다. 분류정확도를 향상시키기 위해 한국드라마를 추가로 Training Set으로 사용하였다. SVM은 RBF커널을 사용하였다.

* Android byte버퍼 전처리:

FFT시간을 줄이기 위한 작업으로 묵음구간과 잡음을 최대한 제거하여 Signal을 줄이고 16Khz에서 5333hz로 샘플링주파수를 재조정하여(1~2700hz까지 획득) 처리시간을 향상시킨 후 전처리 작업을 하도록 하여 3초이내에 감정을 얻어낼 수 있게 하였다. 또한 저주파Index에서 같은 패턴을 보여 저주파의 에너지 값을 조정하여 고주파를 강조하는 방법으로 패턴을 구별하였다.

5. 프로젝트 기여도 평가 (기여도의 총합 = 팀원 수 X 100%, 각 팀원의 기여도 최대치 130%)

|  |  |  |  |
| --- | --- | --- | --- |
| **팀원** | **기여도(%)** | **역할** | **근거** |
| 최환종 | 80 | 감정자질추출을 위한 음성전처리 | 감정 별 주파수영역의 특징을 분석하여 최적특징을 음성신호에서 추출해내는 알고리즘이 감정분류에 핵심. |
| 38 | SVM Server구현 | SVM Train, Model생성, Android 음성 전처리, Google STT Http JSON요청, 감정결과, STT결과관리 |
| 안중환 | 60 | Connection server 구현 | 메시지 순서 관리, Connection 관리 |
| 53 | Android App 개발 | 통화 연동, 묵음 감지 녹음기, service Controller 구현 |
| 정다비치 | 50 | Android UI 구현 | Service View 레이아웃 제작 |
| 19 | 음성데이터 확보 | 한국어 음성 데이터 확보 |

**※ 각 팀원 별 기여도는 반드시 5%이상 차등 적용하여 기재**

6. 목표관리 평가서 (구, 평가 받고 싶은 카테고리) (완성도 포함 최소 4개 이상)

(평가 등급은 A+,A,A-,B+,B,B-,C+,C,C-,D 총 10개, 구현목표를 만족할 경우 B- 이상)

|  |  |  |
| --- | --- | --- |
| **구현목표** | **음성이 실시간으로 처리되어 전송되는가?** | **가중치** |
| **상세내용** | 음성이 청각장애인에게 TEXT + 감정으로 실시간 전송이 되는가?  - 전화와 실시간으로 청각장애인 기기에서 TEXT로 보여줘야 한다.  (TEXT는 실시간, 감정은 시간차를 허용).  - 대화가 가능할 정도의 처리 속도  - 유동IP에 관여 받지 않는 안정적인 데이터 통신 | 25 |
| **평가** | **근거** | |
| **A-** | - 긴 음성은 STT처리시간이 오래 걸려 묵음+최대길이방식으로 쪼개어 처리하였고, 감정결과를 3초이내에 처리할 수 있도록 리 샘플링, 묵음과 잡음을 제거하여 FFT시간을 줄이는 방법으로 속도를 높였다.  - 처음에는 Google STT보다 우리서버의 감정처리속도가 느릴 것으로 예상했는데 알고리즘을 점점 개선하여 Google STT보다 감정이 빠르게 처리되어 감정이 STT결과를 기다리게 되었다.  - Android특성상 유동IP에의해 통신상태가 변하여도 처리결과를 누락 없이 전송한다. | |
| **구현목표** | **음성을 이용한 감정분석이 가능한가?** | **가중치** |
| **상세내용** | 음성을 이용한 감정 파악이 가능한가?  - 감정이 실린 음성을 기쁨, 슬픔, 화남, 보통으로 표시하는가?  - 다양한 목소리(음색, 크기)에도 정확한 감정분석이 가능한가?  - 성별에 관계없이 올바른 감정결과를 얻을 수 있는가?  - 잡음환경에서의 감정분석이 가능한가?  - 화남/기쁨의 감정구분이 잘되는가? | 40 |
| **평가** | **근거** | |
| **A-** | - 학습된 SVM모델을 이용하여 감정을 파악한다.  - 음색, 높낮이, 크기에 상관없는 에너지의 분포 패턴을 사용하였기 때문에 대화하듯이 말한다면 대부분 정확한 감정결과를 얻을 수 있다. (확실한 감정으로 말하지 않는 연기, 글읽기는 부정확한 결과를 얻음)  - 남자와 여자의 패턴이 달라 남, 여 두 개의 모델을 만들어 선택하도록 하였다.  - Android 잡음제거 녹음방식 설정으로 1차 잡음 필터, 서버에서 음성신호처리 시 묵음, 잡음 제거로 2차적필터를 거쳐 잡음이나 묵음을 처리하였다.  - 화남/기쁨을 구별하는 것은 학회에서도 이슈이다. 차이점을 극대화시켜 모델을 만드는 방법으로 정확도를 높였다. | |
| **구현목표** | **시각화** | **가중치** |
| **상세내용** | 음성에 따른 감정을 충분히 시각적으로 표현하였는가?  - TEXT를 위젯 형식으로 통화App과 함께 볼 수 있는가?  - 감정결과값을 사용자가 쉽게 확인할 수 있는가? | 10 |
| **평가** | **근거** | |
| **B+** | - Service View형태로 통화중에 접고 펼 수 있게 하였다.  - 감정별로 그래프를 보여주며 가장 높은 감정에 따라 색상이 변한다.  - 감정별로 보여주는 얼굴의 표정들이 어떤 감정인지 확실히 알 수 있다. | |
| **구현목표** | **완성도** | **가중치** |
| **상세내용** | 1. 모바일 UI 및 사용편의성  2. 감정분류의 정확도  3. 서버의 학습 능력 및 피드백  4. 끊김이나 지연 없는 안정성 | 25 |
| **평가** | **근거** | |
| **A-** | - 일반 통화처럼 사용 가능  - 양방향 가능  - 다중 사용자 가능  - 통화 로그 확인 가능 | |