

심리상담 내담자의 심리변화 시각화시스템

손우정, 손동현, 김주성, 강수환*

부산대학교, *라인플러스

handfriendship1998@gmail.com, europeanboy@naver.com, kjs_0322@naver.com,
lionwithnoname@gmail.com

Visualization system of the flow of emotion change of client in psychotherapy

WooJung-Son, DongHyun-Son, JuSung-Kim, SuHwan-Kang*

Pusan National Univ., *LINE Plus

요 약

심리상담이란 일상에서 마주할 수 있는 다양한 문제 상황으로부터 야기되는 정서적, 정신적 고통을 상담자와 내담자가 함께 공유하며 해결해나가는 과정을 말한다. 심리상담을 행하는 과정에서 상담자가 내담자의 고민과 심리를 정확하게 파악하는 것이 만족스러운 결과를 내는 것에 중요하다. 본 논문에서는 심리상담시 상담자가 AI 기술을 활용하여 심리상담을 받는 내담자의 심리변화를 상담자가 쉽게 파악할 수 있는 방법을 제안한다.

I. 서 론

심리상담이란 일상에서 마주할 수 있는 다양한 문제 상황으로부터 야기되는 정서적, 정신적 고통을 상담자와 내담자가 함께 공유하며 해결해나가는 과정을 말한다. 현재 상담업계에서는 심리치료, 상담 등의 분야에 전문 자격을 가진 상담사가 내담자의 말과 행동, 심리 등을 파악하고 직접 기록하는 형식으로 상담이 진행되는데, 자연어처리 기술의 발전으로 인공지능 기술을 심리상담 업계에 도입하기 위한 연구들이 활발히 진행되고 있다.

인공지능 기술을 상담에 활용하기 위해 행해진 대부분의 연구들은 인공지능 상담사를 도입하여 전문 상담사의 역할을 대신하는 것에 초점이 맞춰져 있다. 하지만 심리치료와 상담 분야의 인공지능 기술 적용이 아직 시작단계에 있어, 적용에 따른 효과를 철저히 검증하기 보다는 기술 자체를 소개하고 이를 활용하는 방안을 주안을 두는 것이 현재까지 진행된 대부분의 연구들이다. 심리적 개입을 통한 정신건강 상의 변화를 검증하는 것은 통상 오랜 관찰을 요하는데, 그것을 장기적으로 추적하고 그 효과성을 검증한 연구 사례가 부족하다.[1] 이러한 이유로 인공지능 상담 서비스의 도입과 상용화는 아직 시기상조이며, 현재 대부분의 상담센터에서는 전문상담사가 직접 상담을 수행하고 있다. 따라서 본 논문은 인공지능 기술을 상담사의 보조도구로써 활용함으로써 실제 심리상담 업계에 보다 직접적인 활용과 도움이 가능하도록 한 방법을 소개한다.

심리상담이 이루어지는 과정에서 상담사가 내담자의 말과 심리변화를 알아차리는 것은 올바른 결과를 이끌어내는 데 있어 중요한 요소이다. 하지만 상담사마다 다른 상담경력 및 인지능력, 해당 상담사의 컨디션 등은 같은 내담자의 동일 문제 상황을 마주하더라도 다른 결과를 도출할 수 있게 한다. 만약 내담자의 말을 듣고 그것의 감정 변화를 동일한 기준으로 관찰하며, 실제 상황에서 빠르게 적용할 수 있도록 시각화하는 시스템을 보조도구로써 사용한다면, 위에서 언급한 여러 외부적 변수로부터 강건한 상담을 행할 수 있다. 이는 내담자의 상담사 교체 요청, 상담사의 수퍼바이저(실제, 수련을 감독하는 경력 상담사)로의 도움 요청 및 인수인계 등 실제 빈번히 발생하는 다양한 상황에서 활용될 수 있다는 장점이 있다.

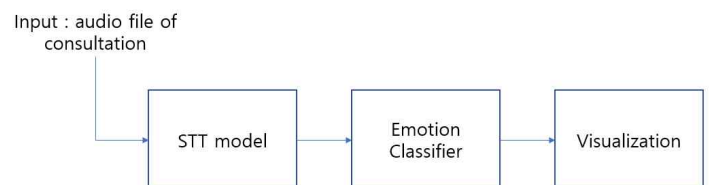
앞서 설명한 시스템을 도입하기 위해 본 연구에서는 인공지능 기술을 사용한 방법론을 제시한다. 우선 STT(speech to text) 기술을 이용해 심리상담 내용을 텍스트로 기록한다. 그 다음 감정분류모델을 이용하여 기록된 상담내용을 28개의 감정군으로 분류한다. 마지막으로 상담사가 쉽게 결과물의 의미를 파악할 수 있도록 시각화하여 타임라인 그래프로 표현한다. 이러한 프로세스를 통해 제한된 상담시간의 효율 가치를 극대화하고,

다양한 유형의 내담자에 대한 이해와 관찰의 '객관성'을 높이는 것에 기여하는 것이 본 시스템의 연구 목표이다.

본 논문의 2장에서는 제안된 시스템의 설계와 구조에 대해 소개한다. 3장에서는 시스템을 실험한 과정과 결과에 대해 설명한다. 마지막으로 결론과 의의, 한계와 필요한 후속 연구를 서술한다.

II. 본론

본 논문에서 제안하는 방법론의 시스템 설계는 [그림1]과 같다. 본 장에서는 STT model, Emotion Classifier, Visualization 순으로 시스템에 대해 기술하였다.



[그림1] 시스템 설계

2.1 음성텍스트변환(STT model)

음성 파일을 감정분류모델이 처리하기 쉽도록 텍스트 파일 형태로 바꾼다. 음성 파일을 텍스트 파일로 바꾸는 과정에는 mozilla 재단의 오픈소스 음성인식 엔진인 DeepSpeech의 사전훈련(pre-trained)된 모델을 이용하였다. DeepSpeech는 Google의 Tensorflow를 기반으로 하는데, MFCC(Mel Frequency Cepstral Coefficient)를 이용하여 음성의 특징벡터를 추출하며, RNN을 핵심 엔진으로 5개의 은닉노드 층을 두어 학습하는 음성인식 모델이다. 변환오류율은 6.5%로, KALDI 등의 타 음성인식 모델에 비해 상대적으로 간단하면서 높은 성능을 보여 선택하였다.[2]

본 연구에서는 DeepSpeech에서 제공하는 사전훈련 모델 중 'pbmm', 'scorer' 모델로 STT-model을 구성하였다. pbmm은 음성신호를 text로 전사(transcribe)하는 음향모델(acoustic model)이고, scorer은 pbmm과 별도로 전사된 text 후보들에 대해 점수(score)를 매기는 외부 언어모델(language model)이다.

사전훈련된 언어모델에 적용한 설정(configuration)은 다음과 같다.

beam_width는 1000으로 두었다. 이는 음성신호의 전사가 일어나는 언어 모델의 디코딩(decoding)과정에서 수행되는 beam search의 tree width이다. 언어모델의 가중치 상수인 language_model_weight은 0.93으로 두었다. 언어모델이 산출하는 점수는 말뭉치의 어휘 수가 많아질수록 작아지는데, 이를 보정해주는 상수 값인 WIP(word insertion penalty)를 1.18로 설정했다.

위와 같이 구성한 STT의 수행과정은 다음과 같다. 먼저 모델은 음성파일을 입력으로 받아들이는데, 입력의 sampling rate는 학습에 사용된 데이터의 sampling rate와 동일하게 16KHz로 샘플링하였다. 전사과정에서 중요한 요소 중 하나는 문장을 구분하는 것인데, [3]에서 제시한 0.8초를 기준으로 어휘 사이의 공백기간이 0.8초를 넘어가면 문장이 바뀌는 것으로 인식하게 하였다. 타임라인 그래프를 출력하여 시각화하는 것이 본 시스템의 목적이므로, 출력된 각각의 문장마다 타임스탬프를 함께 기록하였다. 마지막으로 전사된 문장과 타임스탬프를 csv형식으로 저장하였다.

2.2 감정분류모델(Emotion Classifier)

2.1절의 과정을 거쳐 텍스트로 전사된 음성내용을 입력으로 받아 해당하는 세그먼트에 적절한 감정을 라벨링해주는 분류모델을 설계한다. 학습에 사용된 데이터셋, 분류모델의 설계 순으로 기술하였다.

2.2.1. 데이터셋

해외 소셜 뉴스 웹사이트 'Reddit'의 영어 댓글을 수집하여 그와 적절한 감정을 라벨링하는 형식으로 구축한 구글의 공개 데이터셋 'GoEmotions'을 사용한다.[4] 총 58,009개의 세그먼트가 샘플로 사용되며, 각각의 샘플은 27가지의 감정으로 라벨링되어있고, 샘플의 sequence length의 최대 길이는 30이다. 전체 샘플 중 훈련 데이터셋으로는 43,410개, 테스트 데이터셋으로는 5,427개, 검증 데이터셋으로는 5,426개로 분배하여 사용하였다. 심리상담의 경우, 내담자의 발언에 항상 감정이 실려있지 않을 수 있으므로, 27개의 감정군에 감정적 중립상태인 'Neutral'을 추가하여 총 28가지의 감정군을 사용하였다.

2.2.2. 모델 설계

감정분류모델은 구글의 Transformer기반 사전훈련 모델인 BERT(Bidirectional Encoder Representations from Transformers)의 소형화 버전인 BERT-base를 미세조정(fine tuning)하여 설계하였다.[5] [4]에서 소개한 방법에 따라 선행학습된 모델에 밀집 출력층을 추가하였고, 심리상담의 경우 내담자의 발언에 여러 가지 감정이 내포되어있을 수 있으므로 다중레이블이 가능하도록 시그모이드 크로스 엔트로피 손실함수를 사용하였다. 또한 발언의 문맥을 파악하는 것에 중점을 두어 토큰단위가 아닌 세그먼트단위로 분류를 수행하였다. 이외의 모델 설정(configuration)을 위한 파라미터는 hidden_size 768, hidden_act gelu, vocab_size 30522, hidden_dropout_prob 0.1, num_attention_heads 12, max_position_embeddings 512, num_hidden_layers 12, attention_probs_dropout_prob 0.1 와 같이 구성되었다. 미세조정을 위한 하이퍼파라미터는 batch_size 4, learning_rate 0.00005, epochs 4 가 사용되었다.

2.3 시각화(Visualization)

시각화에는 구글에서 제공하는 Google Charts 중 timeline API를 이용하였다. 2.2절의 수행결과를 tsv형식으로 저장하였다. tsv파일에는 텍스트로 전사된 문장, 발화 시작시간, 발화 종료시간, 감성 레이블, 감성 score(0~1

사이의 확률)이 기록되어 있다. 이 파일을 읽어들이는 자바스크립트 코드를 작성하여 x축은 시간, y축은 감성 레이블이 표기된 타임라인 그래프를 브라우저에 렌더링하였다. 결과 그래프는 III장에서 확인할 수 있다.

III. 실험 방식 및 시스템 실험 결과

본 시스템은 심리상담 중 내담자의 발언을 마이크 등으로 녹취하여 그것을 저장한 음성파일을 모델에 입력하면, 감성흐름의 시각화 결과를 컴퓨터의 브라우저에서 확인하는 상황에서 이용하는 것으로 한다. 심리상담에서는 주로 내담자의 사적인 고민 혹은 정신적, 정서적 문제가 드러날 수 있으므로 프라이버시 보호를 위해 내용을 발설하지 않는 것을 원칙으로 한다. 따라서 전문 상담사가 행하는 실제 상황 대신, 심리상담 중 과거경험을 이야기하는 상황을 가정하여 저자가 직접 해당 상황을 묘사하여 음성 파일에 녹취하였다.

제안한 시스템을 batch형식으로 수행하는 쉘스크립트 파일인 run.sh를 수행하면, 먼저 음성파일을 텍스트로 전사해주는 speech2text.py 파일이 실행되어 전사된 텍스트가 기록된 임시파일을 얻는다. speech2text.py의 설명은 II장의 2.1절에 기술되어 있다. 이후 2.2절의 감정분류모델이 구현된 mybert_classifier.py파일이 내부적으로 수행되어 전사된 텍스트파일에 감성레이블이 표기된 tsv파일을 얻는다. 마지막으로 타임라인 그래프를 그려주는 자바스크립트 코드를 이용해 결과를 시각화한다. 음성파일과 시스템을 시연한 결과 동영상은 아래 링크에서 확인할 수 있다.

<https://drive.google.com/drive/folders/1U3S5Q2zM26UPd-VvLpK6LQLfnfDurOTD?usp=sharing>

III. 결론

본 연구에서는 STT, 감정분류모델을 이용해 심리상담시 내담자의 감성흐름을 관찰하고 시각화하는 방법을 제시하고 실험하였다. 결과를 타임라인으로 보여줌으로써 누구나 쉽게 감정의 흐름을 이해할 수 있고, 상담사에 관계없이 같은 결과를 제공하므로 본 연구의 목표를 달성했다는 것에 의의가 있다.

한계점으로는 음성과 감정라벨이 주어진 테스트 데이터셋 확보의 어려움으로 시스템 성능평가의 어려움이 있었다. 화자식별 기능의 부재로 복수의 내담자가 존재하는 집단상담시 화자별 녹취가 필요하다는 단점이 있다. 본 연구에서는 내담자의 말을 녹취하는 방법을 택했지만 컴퓨터 혹은 기타 주변기기를 이용해 실시간 음성인식을 구현하여 시각화하는 방법도 생각해 볼 수 있다. 이에 관해서는 후속 연구를 기약한다.

참 고 문 헌

- [1] 김도연, 조민기, 신회천, 상담 및 심리치료에서 인공지능 기술의 활용: 국외사례를 중심으로, May, 2020.
- [2] Awni Hannun, Carl Case, Jared Casper et al., Deep Speech: Scaling up end-to-end speech recognition, Dec, 2014.
- [3] Mattias Heldner, Jens Edlund, Pauses, gaps and overlaps in conversations, Oct, 2020.
- [4] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko et al., GoEmotions: A Dataset of Fine-Grained Emotions, Jun, 2020.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Oct, 2018.