

목소리의 감성적인 필터링을 이용한 의사소통 도우미

이수호^{0,1}, 신채영¹, 신영찬¹, 지승용¹, 최지혜^{0,2}, 배대웅¹, 조윤진², 김지우¹, 신세영¹, 김종현^{1*}
 강남대학교 소프트웨어응용학부¹, 강남대학교 중등특수교육과²
 jonghyunkim@kangnam.ac.kr*

Communications Helper with Emotional Filtering of Voice

SooHo Lee^{0,1}, ChaeYoung Shin¹, YoungChan Shin¹, SeungYong Ji¹, JiHye Choi^{0,2}, DaeWoong Bae¹, YunJin Jo²,
 JiWoo Kim¹, SeYoung Shin¹, Jong-Hyun Kim^{1*}
 Dept. of Software Application¹, Dept. of Special Education in Middle School², Kangnam University

요약

본 논문은 일상적인 대화를 사운드와 텍스트로 분리하고, 각각에 대해서 감성을 추출 및 분석함으로써 상황에 맞는 대화의 내용과 분위기를 전달해줄 수 있는 의사소통 도우미를 제안한다. 텍스트는 대화의 내용을 직관적으로 전달하되 상황에 따라 다르게 표현 또는 해석된다. 예를 들어 말하는 화자의 목소리 톤에 따라서 일상적인 대화도 짜증이나 슬픔의 내용으로 변경될 수 있으며, 그 변화는 주파수 형태로도 나타난다 (그림 1 참조). 본 연구는 목소리의 감성 분석을 통해 정확한 의사소통을 도와줄 수 있는 기반 프레임워크를 제안한다. 결과적으로는 청각장애인 및 미취학 아동들이 화자의 대화를 정확하게 인지하여 의사소통 환경을 개선하고자 한다.

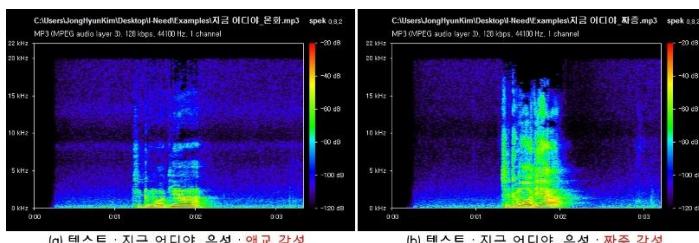


그림 1. 감정에 따라 달리지는 음성 주파수.

1. 서론

청각장애인은 청각적 신호를 전달받지 못하기 때문에 화자의 말 전체를 이해하기 보다 자신이 들은 단어에 초점을 맞춘 대화를 시도하여 대화 유지가 어렵다. 이러한 청력 손실을 보상해주는 지원에는 보청기, 수화나 구화, 필담 등이 있다. 그러나 청각적 보상을 받았음에도 불구하고 주위의 소음과 말소리에 해당되는 주파수 일부를 왜곡해서 듣거나 듣지 못하게 되어 의사소통

이나 학습 등에서 방해를 받는다[1]. 그러므로 일반아동에 비해 자기표현이 어려운 청각장애인은 언어적으로 제한되어 있기 때문에 일반생활에서의 기본욕구의 좌절 등에서 초래되는 부정적 심리를 감소시켜주는 것이 중요하다. 본 연구는 이러한 청각장애인의 일상생활에서 필요한 요구들을 바탕으로 이에 적합한 새로운 기능과 형태의 보조공학기기 기획을 통하여 장애인들의 의사소통에 도움을 줄 수 있는 새로운 프레임워크를 제안한다.

2. 제안하는 프레임워크

제안하는 시스템의 알고리즘 개요는 일반적인 대화 음성 파일을 입력으로 아래와 같은 순서로 수행된다. 기존 접근법과의 가장 큰 차이는 목소리를 사운드와 텍스트로 분할하여 각 감성을 추출한 뒤 최종적으로 사용자에게 상대방의 감성을 동시에 제공해주어 텍스트만 가지고 있는 정보의 한계를 개선시켰다는 점이다 (그림 2 참조).

1. 대화로부터 텍스트와 사운드를 추출하여 각각에 대해서 감성을 분석
2. 텍스트와 사운드의 감성을 비교/분석하여 말을 했을 때의 화자의 감성을 예측
3. 대화의 내용과 이를 분석했을 때의 화자의 감성을 동시에 전달

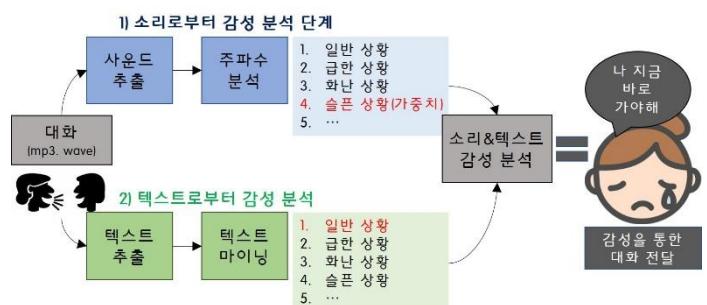


그림 2. 의사소통 도우미 프레임워크의 알고리즘 개요.

* 구두 발표논문

* 학부생 주저자 논문임

2.1. 사운드로부터 감성 분석

우리는 5개의 감성 카테고리(평범, 행복, 슬픔, 화남, 무서움)를 이용하여 사운드로부터 감성을 추출한다. 사운드로부터 감성을 추출하기 위해 5가지 감정의 평균과 표준편차를 이용하여 감성을 분류하였으며, 본 연구에서는 이를 수행하기 위해 *OpenVokaturi* 라이브러리[2]를 이용하였다 (그림 3 참조).

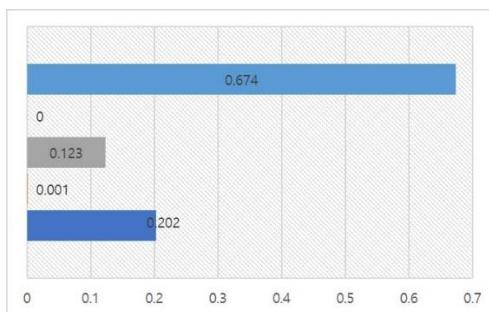


그림 3. 입력된 사운드로부터 분석된 감성 그래프 (**X**축: 감성 카테고리, **Y**축: 각 감성의 수치 값).

계산된 감정은 긍정(평범, 행복)인 s^p 과 부정(슬픔, 화남, 무서움)인 s^n 으로 카테고리를 분류였고, 각 카테고리에 속해 있는 감정은 평균으로 최종 사운드의 감정 수치 값을 계산한다 (그림 4 참조).

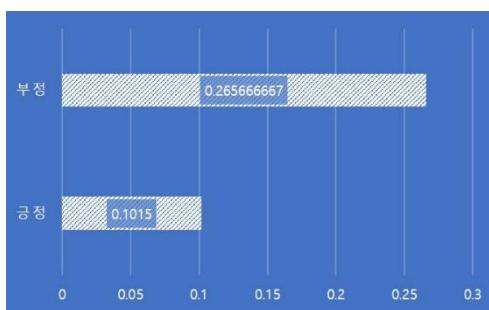


그림 4. 분류된 감성 카테고리에 따라서 계산된 사운드의 감성 수치 값.

2.2. 텍스트로부터 감성 분석

본 연구에서는 대화를 텍스트로 변환하기 위해 IBM의 왓슨 음성 인식[3]을 활용하였으며, 조건부 확률 모델인 나이브 베이즈(Naïve Bayes) 분류 기법을 이용하여 텍스트 내 감정을 분석하였다. 그 중에서 감성 단어 사전과 단어를 비교 대조하여 감정 점수를 산출하는 기법을 활용한다 (수식 1 참조).

$$p(c|x) = \frac{P(x|c)P(c)}{P(x)} \cong \operatorname{argmax} \prod P(x|c)P(c) \quad (1)$$

수식 1은 베이즈 정리를 수학적으로 표현한 것이며 보통 $p(A|B)$ 라고 할 수 있는데 B 가 나왔을 때 A 일 확률을 계산하는 것이다. 여기서 x 는 분류될 문장이고, c 는 분류할 클래스이다. 우리는 이 기법을 기반으로 긍정과 부정 감성 단어 사전과 대조하여 긍정적인 단어와

부정적인 단어의 빈도를 추출한 후 긍정 단어가 나타나면 증가, 부정 단어가 나타나면 감소의 점수를 부여해 텍스트의 감성을 수치화 한다. $p(c|x)$ 가 0으로 나오는 문제를 피하기 위해 라플라시안 스무딩을 적용하여 최종 텍스트의 감정 수치 값을 계산한다 (수식 2 참조).

$$p(c_i|x_k) = \frac{\operatorname{count}(c_i, x_k)+1}{|x_k|+|u|} \quad (2)$$

여기서 $\operatorname{count}(c_i, x_k)$ 는 k 번째 클래스에서 c_i 가 나타난 횟수이다. $|x_k|$ 와 $|u|$ 는 각각 k 번째 클래스의 모든 요소의 수와 모든 클래스에서 유일한 요소의 수를 나타낸다.

2.3. 텍스트와 사운드의 감성 합성

우리는 사운드와 텍스트의 긍정/부정에 대한 수치 값을 이용하여 최종 감성을 사용자에게 제안한다. 텍스트는 감성 단어 횟수에 의존하는 결과는 나타내지만, 사운드의 감성은 소리에 따라 변화가 크게 나타난다. 그렇기 때문에 본 연구에서는 각 감성을 합성할 때 사운드의 감성에 좀 더 높은 가중치로 설정하여 실험하였으며, τ_p 와 τ_n 를 비교하여 최종 감성을 결정한다.

$$\tau_p = \log\left(\frac{s^p\alpha+t^p\beta}{2}\right), \tau_n = \log\left(\frac{s^n\alpha+t^n\beta}{2}\right) \quad (3)$$

여기서 s^p 와 s^n 는 사운드의 긍정과 부정에 대한 감성 수치이며, 텍스트도 t^p 와 t^n 도 마찬가지이다. α 와 β 는 사운드와 텍스트에 대한 가중치이며, 본 연구에서는 각각 0.6, 0.4으로 설정하였으며, 사용자에게는 최종 감성과 개별적인 사운드와 텍스트의 감성도 같이 제공한다.

3. 결론

본 논문에서는 텍스트와 사운드의 감성을 이용하여 의사소통에 도움을 줄 수 있는 프레임워크를 제안하였으며, 일상적인 대화를 긍정/부정에 따라 실험했으며(표 1 참조), 대화의 감성 인식 측면에서 유의미한 결과를 얻었다.

지금 몇 시야, 지금 뭐해, 빨리 와, 빨리 왔네, 이게 뭐야, 잘했네, 등

표 1. 본 연구에서 실험한 일상적인 대화 템플릿.

참고문헌

- [1] KyoungRhan Park, A study on the characteristics of connective ending use by connective ending forms for the students with hearing impairment, *The Korean Society of Education for Hearing-Language Impairments*, 5(1):1-22, 2014.
- [2] OpenVokaturi, <https://developers.vokaturi.com/getting-started/overview>, 2018.
- [3] IBM, Watson Speech to Text, <https://www.ibm.com/watson/kr-ko/developercloud/speech-to-text.html>, 1996.