

# 음성 데이터의 지능적 분류 및 컨텍스트 분석 시스템 구현

최현석, 주성환, 김대천, 박예찬, 염상길, 추현승  
성균관대학교 컴퓨터공학과

e-mail: cheehs, shjoo840, daecheon, dpcks001, sanggil12, choo@skku.edu

## Intelligent Classification and Context Analysis System of Voice Data

HyeonSeok Choi, SungHwan Joo, DaeCheon Kim, YeChan Park,  
Sanggil Yeom, HyeonSeung Choo  
Dept of Computer Engineering, Sungkyunkwan University

### 요 약

사람은 의사소통을 위해 음성, 글자, 몸짓 등 다양한 매개체를 활용한다. 오늘날 스마트폰의 발달로 문자의 비중이 높아지고 있지만 음성 대화는 여전히 사람들 사이에서 가장 많이 사용되어지는 의사소통 수단이다. 음성 대화는 녹음해서 음성 데이터로 남길 수 있다. 음성을 녹음하는 과정은 간편하지만 녹음과 일에서 원하는 데이터를 찾는 것은 많은 시간이 소모된다. 본 논문에서는 음성 데이터를 인식하여 텍스트화 시키고 문자화 된 데이터를 분석하여 사용자에게 효율적으로 분류하는 시스템을 제안한다. 이 시스템으로 사용자는 음성 데이터의 내용을 들어보지 않고 파악할 수 있으며 원하는 내용을 찾을 수도 있다.

### 1. 서론

인간은 문자나 음성을 통하여 하루에도 수백 번씩 대화하고 정보를 공유한다. 문자는 분실, 삭제 혹은 훼손하지만 않으면 계속 남아있기 때문에 시간과 공간의 제약 없이 언제든 다시 열람할 수 있다. 그에 비해 음성은 '휘발성'이라는 특징이 있다. 음성 대화는 화자와 청자 서로의 기억에 의해 저장되기 때문에 쉽게 왜곡되거나 잊히곤 한다.

그래서 사람들은 음성의 한계점을 보완하기 위해 크게 두 가지 방법을 활용한다. 하나는 음성을 녹음하여 음성 데이터화하는 방법이고 다른 하나는 음성을 필기하여 문자 데이터화하는 방법이다. 두 방법은 각각 장단점이 있다. 녹음은 녹음기기만 있으면 간편하게 대화 속의 모든 정보를 저장할 수 있다는 장점이 있다. 하지만 음성 데이터는 검색기능을 지원하지 않기 때문에 활용하기 어렵다. 문자는 검색기능을 지원하지 않기 때문에 효율적으로 활용할 수 있다. 하지만 대화를 기록하기 어렵고 억양과 발음 등 음성 데이터의 특정 정보가 사라진다.

본 논문은 음성 데이터를 편리하게 관리 할 수 있는 시스템을 제안한다. 본 논문의 시스템은 음성 데이터를 인식하여 텍스트화 하고 이를 형태소분석기로 분석해서 키워드를 추출한다. 그 후 Text-Mining 기법을 활용해서 키워드의 중요도를 측정하고 중요한 키워드를 선별한다. 또한 음성데이터의 어떤 시점에 키워드가 등장했는지 저장하여 데이터베이스화한다.

### 2. 관련 연구

#### 2.1. 음성 인식 API

음성 인식 API란 음성 데이터에서 화자를 인식한 후 텍스트로 변환해주는 기술이다. 현대에는 음성인식 기술의 중요성과 활용성이 점점 커지고 있는 추세이다. 그래서 IT 회사들이 각자의 음성 인식 API를 제작하였다. 공개적으로 사용 가능한 음성 인식 API로는 가장 범용 적으로 사용되고 있는 Google사의 Cloud Speech API, 한국어 인식이 높은 Daum사의 Newton, Microsoft사의 Bing Speech API등이 있다 [1]. 현재 음성인식 기술을 활용한 상품들이 많이 상용화 되고 있다. 음성 인식률은 2016년에 90%를 넘어섰고 인식률을 더 높이기 위해 IT회사들의 투자가 많이 이루어지고 있다.

#### 2.2. 음성 명령기

음성 인식과 음성 명령 기술은 현재 급성장하고 있는 인공지능에서 중요한 역할을 맡고 있다. AI가 사용자의 명령을 수행 하려면 음성 인식 기술을 통해 인식할 수 있어야 한다. 가장 가까이에서 접할 수 있는 스마트폰 음성 명령기를 예로 들어보자. 스마트폰 음성 명령기는 초기엔 단문의 명령어를 수행하는 정도의 제한적인 기능을 제공하였다. 현재는 장문의 문장도 읽어 들일 수 있도록 발전하였다. 그러나 음성 명령기의 활용성은 아직까지는 다소 떨어지는 편이다. Apple사의 Siri같은 경우 미국의 아이폰 사용

자의 13%정도만 매일 사용한다고 대답을 할 정도로 아직 일상생활에 익숙하게 사용되는 수준은 아니다 [2].

### 2.3. Context analysis and Text-Mining

Text-Mining은 웹상의 데이터 중 대다수를 차지하는 비정형의 텍스트 데이터를 분석하여 유용한 정보를 추출하는 기법이다 [3]. 빅데이터 시대에 맞맞추어 Text-Mining에 대한 연구가 활발하게 이루어지고 있다.

현재 Text-Mining은 문서 분류 및 요약 기능이 상용화되어있다. 텍스트는 한 단어에 여러 가지 뜻을 내포할 수 있다. 단어가 내포하고 있는 뜻을 Context라고 한다. 텍스트를 단순히 표면적인 의미만 볼 것이 아니라 내포하고 있는 의미인 Context를 파악해야 정확한 정보를 얻을 수 있다 [4]. 따라서 현재 Text-Mining에서는 Context를 고려한 Contextual text mining이 중요시 된다.

## 3. 제안 방법

### 3.1. 시스템 개요

본 논문은 음성 데이터를 데이터베이스화하여 사용자가 편리하게 이용할 수 있게 하는 시스템을 제안한다. 이 시스템은 3가지의 과정을 거쳐서 데이터를 가공한다. 첫 번째는 음성인식 및 STT 과정, 두 번째는 키워드 추출 과정, 마지막으로 데이터베이스화 과정이다.



(그림 1) 음성 데이터의 데이터베이스화 과정

### 3.2. 음성인식 및 STT(Speech To Text)

음성 인식 및 STT 과정에서는 음성 인식 API로 음성 데이터를 인식하여 텍스트화 한다. 이 과정에서는 높은 정확도로 긴 길이의 음성 데이터를 STT하는 API가 필요하다. 현재 Microsoft Bing Speech API와 Google Cloud Speech는 최대 2분의 음성 데이터를 한번에 인식 할 수 있다. 또한 정확도가 다른 IT회사의 음성 인식 API보다 높다. 따라서 위의 두 음성인식 API를 활용하는 것이 적절하다.

### 3.3. 키워드 추출

STT의 과정을 통하여 텍스트화된 데이터는 형태소 분석 과정과 핵심 키워드 추출 과정을 거친다. 형태소 분석기는 문장에서 명사만 추려내고 조사나 접속사는 없앤다. 데이터의 Context를 분석하려면 추출된 키워드의 중요도를 측정할 알고리즘이 필요하다. 본 논문에서는 Text-Mining 기법인 TF-IDF(Term Frequency - Inverse Document Frequency)알고리즘을 이용하여 각 키워드의 중요도를 파악한다 [5]. TF-IDF의 계산 방법은 아래와 같다.

$$TF-IDF = TF \times \log\left(\frac{N}{IDF}\right) \quad (\text{단, } N \text{은 전체 문서 갯수})$$

(그림 2) TF-IDF 알고리즘

### 3.4. 데이터베이스화

음성 데이터에서 추출된 키워드는 키워드 등장 시간과 함께 데이터베이스에 저장한다. 또한 TD-IDF로 선별한 핵심 키워드는 보조키로 지정한다. 보조키를 통해 사용자는 데이터베이스만 봐도 음성 데이터의 주제를 파악할 수 있다. 이는 핵심 키워드가 음성 데이터의 내포된 의미인 context를 보여준다고 볼 수 있다.

## 4. 결론

본 논문에서는 보편적으로 사용하는 의사소통 수단인 음성 대화를 녹음하여 음성 데이터를 문자 데이터 수준으로 편리하게 관리 할 수 있는 시스템 제안한다. 본 논문의 시스템은 음성 데이터의 장점인 기록의 편리성을 취한다. 또한 키워드 추출 및 데이터베이스화를 통해 데이터를 직관적으로 볼 수 있게 하여 활용성을 극대화한다.

## ACKNOWLEDGEMENT

이 논문은 2016년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2010-0020210).

## 참고문헌

- [1] Google cloud speech API (<https://cloud.google.com/speech/>)
- [2] Tariq, K. "Siri, Google Now, Cortana: The Ultimate Digital Voice Assistant?" WCCFTech. com, 2016.
- [3] Mei, Qiaozhu, "Contextual text mining," University of Illinois at Urbana-Champaign, 2009.
- [4] 임일수, "Text-Mining on the Web," 연세대학교 경영대학, Sep. 2009.
- [5] 이성직, 김한준, "Keyword Extraction from News Corpus using Modified TF-IDF," 한국전자거래학회, 2009.