

대화형 에이전트 인식오류 및 신조어 탐지를 위한 알고리즘 개발: 한글 음절 분리 기반의 단어 유사도 활용*

이정원
연세대학교 경영학과
(neckt@yonsei.ac.kr)

임 일
연세대학교 경영학과
(il.im@yonsei.ac.kr)

인공지능 스피커로 대표되는 대화형 에이전트는 사람-컴퓨터 간 대화형이기 때문에 대화 상황에서 오류가 발생하는 경우가 잦다. 에이전트 사용자의 발화 기록에서 인식오류는 사용자의 발화를 제대로 인식하지 못하는 미인식오류 유형과 발화를 인식하여 서비스를 제공하였으나 사용자가 의도한 바와 다르게 인식된 오인식오류 유형으로 나뉜다. 이 중 오인식오류의 경우, 서비스가 제공된 것으로 기록되기 때문에 이에 대한 오류 탐지가 별도로 필요하다. 본 연구에서는 텍스트 마이닝 기법 중에서도 단어와 문서를 벡터로 바꿔주는 단어 임베딩과 문서 임베딩을 이용하여 단순 사용된 단어 기반의 유사도 산출이 아닌 단어의 분리 방식을 다양하게 적용함으로써 연속 발화 쌍의 유사도를 기반으로 새로운 오인식오류 및 신조어 탐지 방법을 탐구하였다.

연구 방법으로는 실제 사용자 발화 기록을 활용하여 오인식오류의 패턴을 모델 학습 및 생성 시 적용하여 탐지 모델을 구현하였다. 그 결과, 오인식오류의 가장 큰 원인인 등록되지 않은 신조어 사용을 탐지할 수 있는 패턴 방식으로 다양한 단어 분리 방식 중 초성 추출 방식이 가장 좋은 결과를 보임을 확인하였다.

본 연구는 크게 두 개의 함의를 가진다. 첫째, 인식오류로 기록되지 않아 탐지가 어려운 오인식오류에 대하여 다양한 방식 별 비교를 통해 최적의 방식을 찾았다. 둘째, 이를 실제 신조어 탐지 적용이 필요한 대화형 에이전트나 음성 인식 서비스에 적용한다면 음성 인식 단계에서부터 발생하는 오류의 패턴도 구체화할 수 있으며, 오류로 분류되지 않더라도 사용자가 원하는 결과에 맞는 서비스가 제공될 수 있음을 보였다.

주제어 : 대화형 에이전트, 음성 인식오류, 발화 분석, 신조어 탐지, 문서 유사도

논문접수일 : 2023년 8월 30일 논문수정일 : 2023년 9월 12일 게재확정일 : 2023년 9월 12일
원고유형 : Fast Track 교신저자 : 이정원

1. 서론

인공지능(artificial intelligent: AI) 스피커로 대표되는 대화형 에이전트(conversational agent)는 인공지능 기술이 스피커, 모바일 등의 디바이스에 적용되어 목소리나 채팅으로 다양한 서비스를 이용할 수 있는 대화형 인터페이스이다(이지희 등, 2017).

인공지능 스피커는 VUI(voice user interface) 기반으로 사용자의 목소리를 통해 디바이스에 명령을 하고 인공지능 에이전트의 음성으로 피드백을 받는 대화의 형태로 사용한다. 즉, 음성 대화 기반의 사람-컴퓨터 인터페이스이다. 이는 대화 메시지, 음성 인식의 정확도 등의 언어적 품질과 음성 에이전트의 목소리의 음색, 속도, 높낮이,

* 이 연구는 BK21 2023년도 연구비와 연세대학교 연구경쟁력 강화사업 연구비(#2023-22-0207)의 지원을 받아 수행되었습니다.

크기, 정서표현 등의 준언어적(paralanguage) 요소들을 포함하고 있다. 이러한 요소들을 시스템 상에서 처리하는 것을 대화형 에이전트라 하며 이는 VUI에서 매우 중요한 역할을 담당하고 있다(김민정 등, 2020).

이렇듯 대화형이기 때문에 현재의 기술력으로는 대화 상황에서 오류가 발생하는 경우가 잦다. 에이전트가 사용자의 음성을 제대로 인식하지 못하는 경우도 있고, 사용자가 발화한 문장의 맥락적 의미를 이해하지 못하는 경우도 있다. 이러한 오류는 그저 매끄러운 대화 상황을 방해하는 정도의 사소한 문제로 여겨질 수도 있지만 실제로는 사용자에게 여러 부정적인 영향을 미친다. 우선, 시스템과의 상호작용 상황에서 일어나는 오류는 사용자 경험(user experience: UX) 측면에서 사용자에게 좌절감을 느끼게 한다. 뿐만 아니라, 인공지능 스피커 혹은 인공지능 비서와 같이 정보 전달, 검색이나 명령 수행을 주 기능으로 삼는 시스템의 경우, 사용자에게 검색이나 명령의 결과를 전달하고 때로는 행동이나 태도의 변화를 이끌어내는 것이 중요한 경우도 있다(Hancock et al., 2011).

에이전트 사용자의 발화 기록에서 인식오류를 탐지하는 것은 그 발생 단계와 원인에 따라서 크게 두 가지 유형으로 나뉜다. 사용자의 발화를 제대로 인식하지 못하는 미인식오류 유형과 발화를 인식하여 서비스를 제공하였으나 사용자가 의도한 바와 다르게 인식된 오인식오류 유형이다. 미인식오류의 경우, 발화 기록에서 오류로 기록되기 때문에 이에 대한 탐지와 개선에 관한 연구들이 많이 진행되어 왔다. 반면에 오인식오류의 경우, 서비스가 제공된 것으로 기록되기 때문에 이에 대한 오류 탐지가 별도로 필요함에도 이에 대한 연구는 매우 드물다.

이 연구에서는 텍스트 마이닝 기법 중에서도

단어와 문서를 벡터로 바꿔주는 단어 임베딩(word embedding)과 문서 임베딩(document embedding)을 이용하여 단순 사용된 단어 기반의 유사도 산출(Ceaparu et al., 2004)이 아닌 한글 음절 분리를 다양하게 적용하여 연속 발화 쌍의 유사도를 산출함으로써 새로운 오인식오류 탐지 방법을 제안한다. 본 연구는 먼저 인공지능 스피커 사용자의 발화 기록에서 연속된 발화를 추출하여 단순 연속 발화인지 오인식오류로 인한 재발화인지 여부를 판별하였다. 추출된 발화 쌍들을 이용하여 단어 및 문서 임베딩 모델 중에서 Word2Vec 모델 및 Doc2Vec 모델의 학습을 진행하였다. 이때, 발화 전체 내용 기반, 단어 기반, 단어의 형태소 원형 기반 외에도 한글 음절 분리로 초성, 중성, 종성을 분리하고 발음이 유사한 군집을 만들어 해당 음절 기반으로 학습하도록 했다. 그 후 학습된 모델로부터 얻은 연속 발화 쌍의 유사도를 계산하여 그 정확도를 살펴보았다.

2. 선행연구

여기서는 본 연구와 관련 있는 선행 연구로서 인식오류 탐지모델과 사용자의 발화 분석을 위한 자연어 처리기법에 대한 기존 연구를 살펴본다. 그리고 한국어의 특징과 그에 따른 자연어 처리의 문제점을 바탕으로 새로운 분석 방법의 가능성에 대해서 기술한다. 이와 관련된 선행연구를 정리하면 <표 2>와 같다.

2.1. 인식오류 탐지모델에 관한 선행연구

인식오류 탐지 및 교정하는 연구들은 크게 음성 인식 모델 자체의 성능 한계를 바탕으로 성능 개선은

위한 연구와 음성 데이터에 있는 발화자의 부정확한 발음을 식별하는 연구로 나누어 볼 수 있다. 음성 인식 모델 성능 개선을 위한 기존 연구들은 음성으로부터 인식된 텍스트를 사용하여 오류를 탐지 및 교정하는 연구, 음성으로부터 화자의 발음 오류 자체를 식별하는 연구, 음성 발화 그룹을 나누어 그룹의 특성을 고려한 음성 인식 모델을 제작하는 연구들이 시도되었다.

우선 텍스트를 사용하여 오류를 탐지 및 교정하는 Sarma and Palmer (2004)의 연구에서는 텍스트 데이터를 분석하여 동시에 발생하는 단어들의 통계치와 음성 음식과 동시에 얻어진 단어들의 통계값을 활용하여 후처리하는 방식을 사용하여 오류 탐지 및 교정을 하는 방식을 제안하였다. 음성으로부터 화자의 발음 오류 자체를 식별하는 Yoon et al. (2010)의 연구에서는 제2 외국어 발화자로부터 주로 발생하는 발음 오류의 음소를 자동으로 탐지하기 위하여 발생된 시점에서 음소가 오류인지 탐지하는 방법을 제안하였다. 마지막으로 음성 발화 그룹을 나누어 특성 고려한 유채권과 이경미(2011)의 연구에서는 성인과 유아의 음성 특성에 따라 주된 발음 오류 음소를 탐지하여 문장에서 어떠한 발음 오류가 발생했는지 탐지하였다. 이러한 연구들은 음성 오류가 발생하는 시점 또는 사전에 오류를 검출하기 위한 모델을 제안한다. 그러나 이러한 오류 발생 상황에서 사용자 관점의 대화형 에이전트 작동 과정에 따라 오류를 구분하고, 오류로 기록되지 않는 부분에 대해서 탐지하는 연구는 매우 드물다.

사용자 관점에서 오류는 사용자가 시스템을 사용할 때 좌절감을 경험하게 한다(Ceaparu et al., 2004). 이러한 좌절감은 사람들이 시스템을 지속적으로 사용하고자 하는 의도를 떨어뜨릴 수 있다. 특히나 관계를 쌓아 나가야 하는 경우에는, 장기적으로

호의적인 관계를 발전시키는 데 있어 장애물이 될 수 있다. 또한, 현재 사용되고 있는 사회적인 관계를 맺는 에이전트의 경우, 설득 에이전트(persuasive agent) 또는 도우미(helper)로서 사용자에게 정보나 의견을 전달하고, 때에 따라 태도 변화를 유도하는 경우도 있다. 오류가 지속적으로 발생하게 되면 정보원(information source)으로서의 신뢰도가 떨어지며 정보 전달이나 태도 변화의 영향력이 감소하게 된다(Salem et al., 2015). 다만, 대인 간 커뮤니케이션 상황에서 일어나는 부정적인 기대 위반과 인간-에이전트 상호작용에서 일어나는 오류 사이에는 몇 가지 차이점이 존재한다.

첫째, 인간-에이전트 관계는 대체로 위계적이거나, 인간이 지배적인 위치에 있는 경우가 많다. 예를 들어, 애플의 시리(Siri)나 KT의 기가지니와 같은 인공지능 스피커는 ‘가상 비서(virtual assistant)’로 홍보되곤 한다.

둘째, 대인 간 커뮤니케이션에서 기대 위반은 보통 서로가 가지고 있는 사회적인 기대치, 혹은 사회적 규범(social norm)을 위반한 경우에 발생한다. 그러나 에이전트 오류는 사회적 규범뿐 아니라 당연히 기술적인 문제에 의해서도 발생한다. 에이전트가 사람의 음성을 인식하지 못하거나 역으로 에이전트가 발화하는 기술이 완전하게 구현되지 못했거나 연결 문제가 생겼을 경우 얼마든지 오류가 발생할 수 있다.

마지막으로, 에이전트 오류가 발생했을 경우 사람들은 그 책임을 대부분 기계에게 돌린다. 즉, 보다 높은 자기봉사적 편향(self-serving bias)을 보인다. 자기봉사적 편향이란, 긍정적인 결과에 대해서는 자신에게로 공로를 돌리는 반면, 부정적인 결과에 대해서는 외부에 그 책임을 돌리는 경향을 가리킨다(Groom et al., 2010; You et al., 2011). 물론 이러한 자기봉사적 편향은 대인 간

커뮤니케이션 상황에서도 마찬가지로 흔하게 일어나는 현상이지만, 컴퓨터 시스템이나 에이전트의 경우에는 이러한 현상이 더 강하게 나타난다. 실제로 컴퓨터 시스템이 어떻게 개발되며 작동하는지 그 원리를 알고 있는 컴퓨터 공학 전공자들조차도 컴퓨터 오류가 발생할 경우, 일반적으로 컴퓨터의 의사 결정에 책임을 돌린다는 연구 결과도 있다(Friedman, 1995).

일반적인 대화형 에이전트의 작동 과정은 다음과 같다. 먼저 발화 단계는 사용자가 대화형 에이전트에게 음성으로 명령을 전달하는 단계로 시작하여 대화형 에이전트가 그 음성을 인식하는 인식 단계로 이어진다. 그 후 명령을 이해하고 명령에 대한 결과를 처리하여 서비스를 제공하는 명령 처리 단계로 구분할 수 있다. 또한, LuperFoy et al. (1998)는 음성 에이전트에서 발생할 수 있는 오류를 7가지 단계로 구분하였다. 이를 대화형 에이전트의 작동 과정에 비추어 보면, 0단계는 발화 단계, 1~2단계는 인식 단계, 4~5단계는 명령 처리 단계인 수행 단계에 해당한다. 이를 적용시키면 <표 1>과 같다.

이러한 연구들에 따르면, 에이전트 사용자의 발화 기록에서 인식오류를 탐지하는 것은 그 발생

단계와 원인에 따라서 크게 두 가지 유형으로 나뉜다. 사용자의 발화를 제대로 인식하지 못하는 미인식오류 유형과 발화를 인식하여 서비스를 제공하였으나 사용자가 의도한 바와 다르게 인식된 오인식오류 유형이다. 가장 큰 특징은 미인식 오류의 경우, 발화 기록에서 오류로 기록되지만 오인식오류의 경우, 서비스가 제공된 것으로 기록되기 때문에 이에 대한 오류 탐지가 별도로 필요하다.

오인식오류가 발생하는 원인은 첫째, 신조어나 줄임말을 사용하는 경우 그 의미를 파악하기 어려워져 가장 가깝게 발음되는 서비스로 제공된다. 예를 들어, ‘꼬리에 꼬리를 무는 그날 이야기’를 ‘꼬꼬무’로 줄여서 발화한 경우 가장 가깝게 발음되는 ‘코코몽’으로 제공될 수 있다. 두번째, 인식되는 시점보다 발화가 앞서서 이루어져서 발화의 일부가 인식되지 않는 경우이다. 예를 들어, ‘골목식당’을 발화한 사용자가 인식이 가능한 시점보다 빠르게 발화한다면 ‘목식당’으로 인식하여 다른 서비스가 제공될 수 있다. 이외에도 발음상의 이유와 같은 사람-사람 간 대화에서도 나타나는 문제들도 발생할 수 있다. 따라서 이에 대한 개선을 위해 오인식오류를 탐지하여 사용자가 의도한 서비스를 제공할 수 있도록 해야 한다.

<표 1> 음성 에이전트 작동 단계에 따른 오류 단계와 유형

	작동 단계			오류 단계	
	발화 단계	인식 단계	수행 단계		
1	X	X	X	Lv. 0	사용자의 음성 입력 실패
2	O	X	X	Lv. 1	에이전트가 음성 인식 실패
3	O	△	△	Lv. 2	에이전트가 음성을 잘못 인식
4	O	O	△	Lv. 3	사용자가 의미 상 틀린 문장을 발화
5	O	O	△	Lv. 4, 5	원하는 결과 정보 부족
6	O	O	△	Lv. 6	백엔드 시스템 문제

2.2. Word2Vec과 Doc2Vec

자연어 처리 기법 중에는 단어를 벡터(vector)로 매핑 시켜 벡터로 표현하는 과정인 단어 임베딩 방식과 여러 단어들이 포함된 문서를 벡터로 매핑 시켜 벡터로 표현해주는 문서 임베딩 방식이 있다. Word2Vec은 단어 임베딩 방식 중에서도 널리 쓰이는 벡터 공간 모델(vector space model) 중 하나이다(Mikolov et al., 2013; Church, 2017). 각 문서의 모든 단어는 전체 개수를 크기로 하는 벡터로 전환된다. 즉, Word2Vec은 다른 단어와의 관계를 반영하는 알고리즘을 바탕으로 산출된 단어 벡터라고 할 수 있다. 여기에서 비슷한 분포를 가진 단어나 문서는 유사한 뜻을 가진다는 분산 가설(distributional hypothesis)을 기반으로 한다(Harris, 1954). ‘주방에 있는 등 쪼뼉’라는 문장에서 ‘주방’은 멀리 위치한 ‘쪼뼉’보다 가까이 있는 ‘등’과 의미상 가깝다고 보는 것이다. 이러한 의미상 유사한 단어를 서로 인접시켜 임베딩 시키므로 벡터로 표현된 단어들은 문법적인 부분만 아니라 의미적인 부분까지 반영되어 거리가 가까울수록 서로 비슷한 벡터를 가지게 된다(정지수 등, 2019; Rong, 2014). 또한 단어 벡터에 포함된 단어들은 모두 수치화되어 있기 때문에 단어와 단어 간의 거리를 활용한 벡터 연산이 가능하고 이로 인해 추론을 할 수 있게 된다(최가람, 최성필, 2018). Word2Vec은 단어와 단어 간의 관계를 매칭하는 과정에서 산출되는데, 입력층과 출력층을 가지고, 그 사이에 은닉층 1개를 포함하는 얇은 신경망 구조로 되어 있다(유원준, 안상준, 2021). 즉, 각 단어들로 구성된 행렬(matrix)이 사전에 설정된 Word2Vec 행렬과 연산된 결과가 소프트맥스(softmax) 활성화 함수(activation function)를 거쳐 벡터로 변환되어 보고자 하는 단어의 벡터와 비교된다(Di et al.,

2021). 이 과정이 문서 단어들에 대해 반복 학습되고, 이어서 두 벡터 간의 차이(error)를 최소화하는 방식이 역전파(backpropagation)가 이뤄지며, 앞서 연산에 적용된 가중치를 반복적으로 바꾸는 과정을 거쳐 우수한 예측 성능을 보이는 Word2Vec 행렬이 도출되는 것이다(이종혁, 2021).

이러한 Word2Vec이 단어를 벡터로 표현하는 단어 임베딩 방식이라면, Doc2Vec은 문서를 벡터로 표현하는 문서 임베딩 방식이다(Lau & Baidwin, 2016). Word2Vec의 학습 알고리즘에는 2가지 모델이 있다. 첫번째는 문맥을 바탕으로 단어를 예측하는 CBOW(continuous bag of words) 모델과 단어를 바탕으로 문맥을 예측하는 skip-gram 모델이 있다(Mikolov et al., 2013). Doc2Vec은 CBOW와 유사한 PV-DM(distributed memory version of paragraph vector)과 skip-gram과 유사한 DBOW(distributed bag of words) 방식을 선택할 수 있다(이종혁, 2021). 기존 Word2Vec의 입력 부분에 문서 번호만 추가하여 각 문서에 고유 번호를 붙여 새로운 단어처럼 모델에 추가한다. 입력 부분의 주변 단어들과 문서 번호가 출력 부분의 중심 단어를 예측하는 방식으로 이뤄지며, 이 과정을 거쳐 문서도 번호별로 벡터를 갖게 된다. 이를 바탕으로 문서 간 유사도 측정이 가능해진다. 이때, 코사인 유사도(cosine similarity)를 사용하는데 두 문서의 벡터 간 코사인 각도를 측정해 유사도로 재구성한 것이다. 벡터 간 코사인 유사도는 0에서 1사이의 값을 갖고, 0은 둘의 관계가 없으며 1은 둘이 완전히 동일함을 나타낸다(이한동, 김종배, 2017). 본 연구에서는 각 발화 전체를 하나의 문서로, 발화에서의 단어들을 각각 벡터로 매핑하여 문서 간 유사도를 구하기 위해 PV-DM 모델이 사용되었다.

2.3. 한글 음절 분리

한글 음절은 한글 자모 첫소리인 초성과 가운뎃소리 글자인 중성, 끝소리 글자인 종성으로 이루어진 한글의 단위를 말한다. 현대 한글에서 표현할 수 있는 모든 글자는 모두 11,172 글자로, 초성 19개, 중성 21개, 종성 27개로 종성이 없는 경우를 포함하여 조합하여 만들 수 있는 글자 수이다(신지영, 2011).

영문 신조어의 경우, 주로 축약어로 각 단어의 맨 앞에 위치한 글자만 뽑아서 쓰는 경우가 많고, 해당 알파벳 뒤에 ‘.’(마침표)를 붙인다. 이에 반해 한글 줄임말의 경우, 본래의 어형에서 음절이나

형태소가 줄어든 말이거나 두 단어 이상의 표현을 줄이기 위해 한 음절씩 따와서 만든 말을 칭한다. 예를 들어, ‘안 물었고, 안 궁금해’를 ‘안물안궁’이라는 신조어로 사용하는 경우가 대표적이다. 이러한 축약어 외에도 파생어의 경우, 단어 및 용어에 접두사 또는 접미사 등의 보조를 붙여 새로운 단어 및 용어를 만들어내는 경우로 ‘~까’나 ‘~충’을 붙이거나 ‘주린이’의 예에서 그 사례를 볼 수 있다. 혹은 합성의 경우처럼 축약과 결과는 비슷하나 의미를 잃어버리는 경우로 ‘new’와 ‘baby’의 합성으로 ‘뉴비’라는 신조어가 사용되는 예와 ‘웃프다’와 같은 예가 있다(신조어, 2011).

〈표 2〉 선행연구

구분	연구자	내용	
인식오류 탐지모델	음성인식 오류 탐지 및 개선	Sarma and Palmer (2004)	텍스트를 사용하여 오류 탐지 및 교정
		Yoon et al. (2010)	음성으로부터 화자의 발음 오류 식별
		유재권과 이경미(2011)	성인과 유아의 음성 특성에 따라 주된 발음 오류 탐지
	사용자 관점의 오류 경험	Ceaparu et al. (2004)	오류 경험과 좌절감
		Salem et al. (2015)	오류 경험과 신뢰도
		Groom et al. (2010) You et al. (2011)	오류 경험과 자기봉사적 편향
		Friedman (1995)	오류 경험와 책임
음성 에이전트 오류 구분	LuperFoy et al. (1998)	음성 에이전트 오류 7단계	
Word2Vec과 Doc2Vec	Word2Vec	Mikolov et al. (2013) Church (2017) Harris (1954) 정지수 등(2019) Rong (2014) 최가람과 최성필(2018) 유원준과 안상준(2021) Di et al. (2021) 이종혁(2021)	Word2Vec 개념과 구조
	Doc2Vec	Lau and Baidwin (2016) 이종혁(2021)	Doc2Vec 개념과 구조
	코사인 유사도	이한동과 김종배 (2017)	코사인 유사도 개념
한글 음절 분리	한글 음절	신지영(2011)	한글 음절 특성
	신조어	신조어(2011)	신조어 특성

인식오류에서 가장 크게 나타나는 신조어의 특징은 고유 명사에서 주로 나타나고, 이때 축약어 또는 합성어의 형태가 가장 많이 나타난다. 단순 축약어의 경우, 같은 음절이 사용되는 경우가 많으므로 음절단위로 분리하여 유사도를 판단하는 경우, 그 특징을 가장 잘 반영할 수 있을 것으로 보인다. 그러나 합성어와 같이 축약된 상태에서 변형이 있는 경우에는 초성은 유지되지만 중성이나 종성은 발음으로 나타나는 소리로 변형되거나 생략되는 경우가 많기 때문에 음절단위에서 초성을 추출하는 방식이 성능을 높일 수 있다. 예를 들어, ‘하트시그널’이라는 프로그램명은 ‘할시’로 줄이고 변형하여 사용하는 것에서 그 방식을 이해해볼 수 있다. 따라서 인식오류에서도 빈번하게 요인으로 작용하는 신조어 탐지에서 이러한 특성을 반영하여 한글 음절 분리 과정을 접목할 필요성이 충분하다고 할 수 있다.

3. 연구 방법

3.1. 데이터

본 연구에서 대상으로 삼은 인공지능 스피커 기기는 사용자의 사용성과 편리성에 기여하고 인간-기계 상호작용을 주로 하는 기기를 대상으로 하였다. 다양한 시장 부문에 걸쳐 인공지능 스피커 서비스를 제공하는 대규모 글로벌 회사의 패널 데이터 세트를 사용하였다. 익명을 유지하고자 하는 회사의 바람에 따라 회사를 “주요 기업”이라고 명시한다. 인공지능 스피커 서비스의 데이터셋에는 각 사용 날짜와 시간, 사용자의 음성 명령에 대한 음성-텍스트 데이터, 스마트 스피커가 사용자의 질문에 구두로 응답하여 제공된 서비스의

종류 등이 포함된다. 2018년 10월부터 2021년 10월까지의 3년 간의 데이터에서 무작위로 10,000개 가구를 선택하여 발화가 기록되지 않은 255,853건을 제외한 24,111,355건의 발화가 분석되었다.

3.2. 연속 발화 분석

오인식오류는 명령이 실행되었지만 원하는 결과가 아닌 경우라는 특징을 가지고 있다. 사람들은 어린 아이나 외국인에게 말할 때와 마찬가지로, 컴퓨터나 로봇과 상호작용할 때 언어를 변조한다는 것이 밝혀져 왔다(Kriz et al., 2010). 즉, 명령이 실행되었지만 원하는 결과가 아닌 경우, 에이전트의 언어 능력이 낮다고 판단하여 입력을 더 정확하게 하기 위해 언어가 조절된다. 따라서 발화를 조절 후 사용자의 음성 행동 패턴을 나타내는 재발화가 수행된다. 따라서 오인식오류를 추출하기 위해 20초 이내의 발화 간격을 기반으로 시작 발화, 중간 발화, 최종 발화를 나타내는 플래그 레이블링이 먼저 수행되었다. 20초의 기준은 명령 실행에 필요한 평균 시간을 기반으로 책정되었다.

연속된 발화들을 각각 주어와 서술어로 분리하여 명령 수행의 종류를 결정하는 서술어만 다르고 주어는 같은 경우, 오인식으로 판별할 수 있다. 하지만 서술어가 같은 경우에는 대부분 축약어 사용 혹은 아직 등록되지 않은 단어의 사용으로 인해 탐지 모델을 개발할 필요성이 있다. 즉, 이러한 연속된 발화가 오인식오류로 인한 재발화인지 단순히 빠르게 다른 서비스를 이용하기 위해 발화를 이어서 한 것인지를 판단하기 위해 다음과 같은 탐지 모델을 제안한다.

3.3. 탐지모델

연속 발화에서 오인식오류로 인한 재발화는

입력을 더 정확하게 하기 위해 언어가 조절되기 때문에 앞선 발화와 다음 발화가 유사하다는 특징을 갖고 있다. 따라서 연속된 발화-발화 간 유사도를 측정하여 일정 수준 이상인 경우 언어가 조절되었다고 볼 수 있다.

탐지 모델 개발을 위해 연속 발화를 기반으로 Doc2Vec을 수행하였다. 이 방법은 Word2Vec의 확장으로, Word2Vec 모델에 문장 또는 단락 행렬을 추가한다. 기존 단어 벡터에 단락 벡터를 추가하는 것을 PV-DM(paragraph vector-distributed memory) 모델이라고 한다(Douzi et al., 2017). 이때 연속된 발화 간의 유사도는 Doc2Vec으로 얻은 벡터의 코사인 유사도 값을 사용하여 측정할 수 있다(이종혁, 2021).

이때, 모델 학습을 위해 연속된 발화는 쌍으로

그룹화되었으며, 오인식오류와 그렇지 않은 경우 각 500개의 쌍, 총 1,000개의 쌍에 대해 레이블링을 하는 코딩이 수행되었다. 각 발화를 하나의 단락으로 하여 단락 벡터 간 유사성을 분석하고, 각 단락 내 단어 벡터 생성을 위해 분리 방식을 <표 3>에 나와 있는 총 13개의 기준을 기반으로 학습 및 정확도 검증이 진행되었다. 즉, 단어 벡터와 단락 벡터를 구분하여 연결할 때, 각 단어 벡터를 나누는 기준을 다르게 하여 오인식오류로 인한 유사한 연속 발화인지 여부가 제대로 판단되었는지 여부를 정확도로 환산하였다. 발음 유사 군집은 훈민정음 제자 원리 및 상형의 원리에 따라 초성의 경우, 어긋넛소리, 헛소리, 입술소리, 잇소리, 목소리로 구분하고, 중성의 경우, 천지인 구분을 사용

<표 3> 분리 방식 별 단어 벡터 예시

분리 방식	예시
전체 발화	['티비 꺼줘']
공백 기준 분리	['티비', '꺼줘']
형태소 기준 분리	['티비', '꺼', '줘']
형태소 원형 변형 후 분리	['티비', '끄', '어줘']
음절 단위 분리	['티', '비', '꺼', '줘']
음절 단위 분리 초성 추출	['트', '비', '기', '지']
음절 단위 분리 중성 추출	['', ' ', ' ', ' ']
음절 단위 분리 발음 유사 초성 군집	[헛소리, 입술소리, 어긋넛소리, 잇소리]
음절 단위 분리 발음 유사 중성 군집	[인, 인, 지, 지]
음절 단위 분리 초성 및 중성 결합	['트 ', '비 ', '기 ', '지 ']
음절 단위 분리 초성, 중성, 중성 결합	['트 ', '비 ', '기 ', '지 ']
음절 단위 분리 발음 유사 초성 및 중성 군집	[헛소리_인, 입술소리_인, 어긋넛소리_지, 잇소리_지]

하여 발음이 유사한 초성이나 중성끼리 군집화 과정을 거쳐 분리하는 것으로 진행하였다. 정확도 검증은 학습 세트와 검증 세트를 75 대 25 비율로 4번 랜덤하게 반복하여 연속 발화 쌍 간 코사인 유사도가 80% 이상인 경우 오인식오류로 분류하여 오인식오류 여부 결과와 비교하여 정확도(accuracy), 재현율(recall), 정밀도(precision), F1 점수로 확인하였다. 이때, 연속 발화 쌍 간 코사인 유사도가 80% 이상이라는 것은 각 발화에서 생성된 벡터 간 코사인 각도를 측정하였을 때, 관계가 없으면 0, 완전히 동일하면 1로 판단되고, 완전히 동일할 확률이 80% 이상이라는 의미로 문서 간 유사도가 매우 높음을 의미한다. 즉, 일부 언어가 조절되어 앞선 발화와 다음 발화가 유사한 경우를 판별하여 그 정확도를 살펴보았다.

4. 분석 결과

4.1. 기초 통계

분석에 사용된 단어 벡터의 각 단락 벡터 내 평균 구분된 수의 평균은 <표 4>와 같다.

<표 4> 분리 방식 별 단락 벡터 내 단어 수 평균

분리 방식	단락 벡터 내 단어 수 평균
전체 발화	1
공백 기준 분리	4.26
형태소 기준 분리	5.24
형태소 원형 변형 후 분리	5.24
음절 단위 분리 (초성, 중성, 종성, 발음유사군집, 일부 결합 등)	12.31

4.2. 전체 데이터의 정확도

각 분리 방식으로 학습 및 정확도 검증을 진행하였다. 레이블링된 연속 발화 쌍 중 25%를 4번 랜덤하게 반복하여 검증에 사용하였다. 이때, 연속 발화의 코사인 유사도가 80% 이상인 경우 오인식오류로 분류하여 오인식오류 여부 결과와 비교하여 정확도(accuracy), 재현율(recall), 정밀도(precision), F1 점수를 검증하였다. 1~4차의 각 결과는 <부록 1> ~ <부록 4> 와 같고, 이를 평균으로 종합하면 <표 5>와 같다. 또한, 검증에 사용되는 비율을 변경하여 30%를 검증에 사용하여 분석한 결과는 <표 6>과 같다.

그 결과, 기존에 유사도 검증에서 주로 사용되는 공백 또는 형태소 기준 분리보다 음절 단위 분리 후 초성을 추출한 것을 기준으로 학습 및 검증을 진행했을 때 보다 모든 검증 방식에서 높은 정확도를 보이는 것으로 나타났다. 이는 연속되는 발화에서 축약어를 사용한 줄임말이나 아직 정보가 등록되지 않은 단어를 찾아내는 것에 두 분석에서 모두 약 76%의 정확도로 분류할 수 있음을 나타낸다. 초성만을 추출했을 때, 유사한 발음이 잘못 인식된 경우도 탐지하기 위하여 발음이 유사한 군집을 만들어 진행하였을 때는 오히려 정확도가 떨어지는 결과를 보였다. 이는 초성에서의 발음이 잘못 인식되는 경우보다 제대로 인식되었으나 다른 명령이 수행된 경우가 더 많다는 것을 나타낸다고 볼 수 있다. 또한 검증 데이터의 비율을 변경하여도 큰 차이가 없이 분리 방식별 성능을 확인할 수 있다.

4.3. 신조어 정확도 분석

추가적으로 실질적으로 신조어 탐지가 가능한지를 살펴보기 위해 위키백과의 대한민국 신조어

〈표 5〉 분리 방식 별 1-4차 정확도 검증 종합 결과: 25% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	11.88	16.01	17.95	16.92
공백 기준 분리	21.30	22.63	23.75	23.18
형태소 기준 분리	63.10	62.91	63.80	63.35
형태소 원형 변형 후 분리	58.40	58.32	58.90	58.61
음절 단위 분리	71.83	71.01	73.75	72.36
음절 단위 분리 초성 추출	76.35	75.78	77.45	76.61
음절 단위 분리 중성 추출	53.78	53.65	55.50	54.56
음절 단위 분리 중성 추출	5.73	6.61	6.75	6.68
음절 단위 분리 발음 유사 초성 군집	55.48	55.35	56.70	56.02
음절 단위 분리 발음 유사 중성 군집	65.75	65.40	66.90	66.14
음절 단위 분리 초성 및 중성 결합	71.30	70.92	72.20	71.56
음절 단위 분리 초성, 중성, 중성 결합	38.35	38.53	39.15	38.84
음절 단위 분리 발음 유사 초성 및 중성 군집	70.43	69.92	71.70	70.80

〈표 6〉 분리 방식 별 정확도 검증 종합 결과: 30% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	9.33	9.87	10.00	9.93
공백 기준 분리	19.67	21.02	22.00	21.50
형태소 기준 분리	62.33	61.94	64.00	62.95
형태소 원형 변형 후 분리	55.33	55.26	56.00	55.63
음절 단위 분리	71.00	70.32	72.67	71.48
음절 단위 분리 초성 추출	76.00	75.66	76.67	76.16
음절 단위 분리 중성 추출	50.33	50.32	52.00	51.15
음절 단위 분리 중성 추출	4.33	5.23	5.33	5.28
음절 단위 분리 발음 유사 초성 군집	57.33	57.14	58.67	57.89
음절 단위 분리 발음 유사 중성 군집	66.00	64.81	70.00	67.31
음절 단위 분리 초성 및 중성 결합	71.00	69.81	74.00	71.84
음절 단위 분리 초성, 중성, 중성 결합	34.00	34.81	36.67	35.71
음절 단위 분리 발음 유사 초성 및 중성 군집	70.67	69.62	73.33	71.43

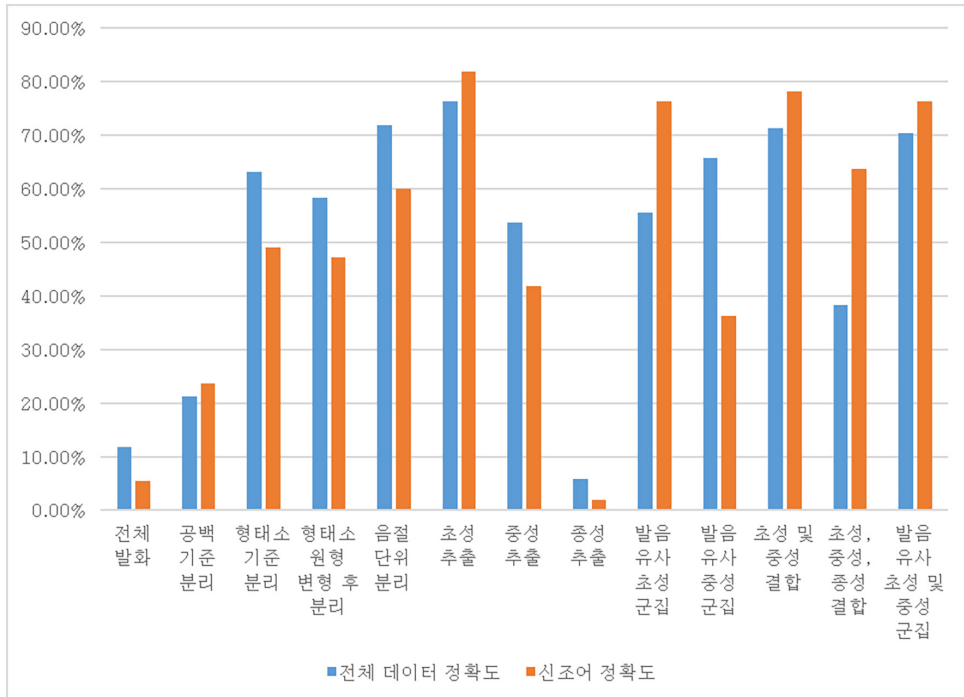
목록을 활용하여 총 55개의 표제어와 설명을 연결하여 각 분리 방식별로 정확도(accuracy), 재현율(recall), 정밀도(precision), F1 점수를 검증한 결과는 <표 7>과 같다.

오인식오류 발화 탐지와 마찬가지로 음절 단위 분리 후 초성을 추출한 것을 기준으로 한 모델을 활용하여 검증을 진행했을 때, 정확도 및 정밀도, F1 점수에서 가장 높은 결과를 보이는 것으로 나타났다. 잘못된 연결된 검증 쌍이 없으므로 재현율은 모두 동일하게 나왔다. 이는 신조어와 설명을 연결할 때 약 82%의 정확도로 분류할 수 있음을 나타낸다. 두 결과를 그래프로 비교해

보면 <그림 1>과 같다. 인식오류 분석보다 신조어 분석에서 가장 성능이 좋은 초성 분리 방식이 보다 높은 결과로 나타났는데, 이는 준말의 비율이 보다 더 많고, 단순 연속 발화와 같이 오인식오류가 아닌 경우가 본 분석에서 포함되어 있기 때문에 나타난 결과로 보인다. 반대로 초성이 포함되지 않은 방식에서는 오히려 정확도가 큰 폭으로 감소된 경우가 대부분으로 이는 신조어에서 축약과 합성의 특징이 보다 확연히 드러나기 때문으로 볼 수 있다.

<표 7> 분리 방식 별 신조어 목록 정확도 검증 결과

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	5.45	100.00	5.45	10.34
공백 기준 분리	23.64	100.00	23.64	38.24
형태소 기준 분리	49.09	100.00	49.09	65.85
형태소 원형 변형 후 분리	47.27	100.00	47.27	64.20
음절 단위 분리	60.00	100.00	60.00	75.00
음절 단위 분리 초성 추출	81.82	100.00	81.82	90.00
음절 단위 분리 중성 추출	41.82	100.00	41.82	58.97
음절 단위 분리 중성 추출	1.82	100.00	1.82	3.57
음절 단위 분리 발음 유사 초성 군집	76.36	100.00	76.36	86.60
음절 단위 분리 발음 유사 중성 군집	36.36	100.00	36.36	53.33
음절 단위 분리 초성 및 중성 결합	78.18	100.00	78.18	87.76
음절 단위 분리 초성, 중성, 중성 결합	63.64	100.00	63.64	77.78
음절 단위 분리 발음 유사 초성 및 중성 군집	76.36	100.00	76.36	86.60



〈그림 1〉 전체 데이터 및 신조어 정확도 비교

5. 결론 및 시사점

5.1. 결론

본 연구에서는 대화형 에이전트의 인식오류 중 오류로 기록되지 않아 탐지가 어려운 오인식 오류의 패턴을 파악하여 이를 탐지 모델에 분리 방식으로 반영하여 인식오류를 분석하였다. 이는 두 가지 관점에서 결론을 살펴볼 수 있는데 먼저 모델 관점에서는 Doc2Vec 모델에서 각 단락 벡터와 단어 벡터가 연결되는 점을 이용하고, 각 단어 벡터를 구분하는 분리 방식을 여러 유형으로 변형하여 모델 성능을 향상할 수 있었다. 또한, 패턴 관점에서는 인식오류의 패턴 및 신조어의 특징을 파악하여 탐지 분류 모델 개발에 접

목하였고 다른 분리 방식들과 비교함으로써 최적의 패턴이 반영된 모델을 찾았다. 즉, 먼저 실제 대화형 에이전트 사용자의 발화를 기반으로 본 연구에서 제시하는 다양한 분리 방법을 적용하였다. 각 분리 방법 별 모델을 학습 및 검증하여 오인식오류의 가장 큰 원인인 등록되지 않은 신조어 사용을 탐지할 수 있도록 그 패턴을 적용하였다. 그 결과 초성 추출 방식의 결과에서 가장 유의미한 분석 결과를 도출하였다. 특히, 단순 축약으로 신조어가 만들어진다면 초성 추출보다 음절 추출이 보다 효율적으로 패턴을 반영한 결과로 나와야 하지만 축약 시에 변형되는 경우가 있고, 이때에 초성보다 모음인 중성이 주로 변하기 때문에 음절보다 초성 추출 방식이 가장 높은 정확도를 보였다는 점이 신조어 생성 원리

와도 맞닿아 있어 패턴이 잘 반영된 결과라고 볼 수 있다. 또한, 초성의 경우 발음이 유사하여 잘못 인식되는 경우가 드물다는 점을 초성 추출 방식의 경우가 초성에 대한 발음이 유사한 군집을 만들어 진행한 경우보다 향상되었다는 점에서 찾을 수 있다. 그리고 초성과 중성 또는 초성, 중성, 종성을 모두 추출 후 결합한 경우보다 초성만 추출한 경우가 더 정확도가 높은 것으로 보아 주로 발음으로 인해 잘못 인식되는 경우는 초성보다는 중성이나 종성에서 나타난다는 것도 알 수 있다.

5.2. 시사점 및 한계점

이 연구의 이론적 시사점으로는 첫째, 인식으로 기록되지 않아 탐지가 어려운 오인식오류를 줄일 수 있는 방법을 제시하고 있다는 점이다. 오인식은 어떤 오류가 발생했는지를 자동으로 데이터를 수집하기가 어려운데 이 논문에서는 연속 발화 데이터를 사용해서 오인식을 탐지하고 분석할 수 있는 방법을 제시하고 있다. 이 방법은 비슷한 문제를 가지고 있는 분야의 연구에 적용될 수 있을 것이다. 둘째, 신조어 중 가장 빈번하게 사용되는 축약어의 사용은 아직 의미가 시스템에 설정되지 않아 적절한 서비스 제공이 어렵고, 단순히 유사한 단어로 인식하여 서비스가 제공되는 경우가 많다. 본 연구에서 사용한 방법을 사용하면 새롭게 등장하는 신조어의 경우도 자동으로 탐지할 수 있는 가능성을 보였다. 즉, 본 논문에서 제시된 방법을 사용하면 오인식 단어가 기존의 어휘사전에 등록되지 않은 신조어일 확률이 높은 경우를 구분할 수 있다. 이를 바탕으로 탐지된 신조어를 자동으로 추가하는 알고리즘의 개발이 가능할 것으로 기대된다. 셋째,

이 논문에서는 다양한 음절 분리 방식을 제시하고 있으며 이를 다른 분야에도 적용할 수 있을 것이다. 즉, 오인식 탐지뿐 아니라 유사단어 분석과 같은 다른 텍스트 분석에서도 다양한 초성 분리 방식을 적용해 볼 수 있을 것이다.

본 연구는 실무적으로도 몇 가지 시사점을 제시한다. 첫째, 실제 사례인 대화형 에이전트 사용자의 실제 발화를 토대로 분석함으로써 실제 시스템에 적용성이 높은 결과를 도출하였다는 점이다. 만들어진 스크립트가 아닌 실제로 사용되는 발화에서 오인식오류가 발생한다는 점을 밝혔고, 이를 탐지하기 위해 오류의 유형을 살펴봄으로써 패턴을 적용하여 탐지 모델의 성능을 향상하였다. 둘째, 대화형 에이전트의 경우 발화된 내용이 앞이나 중간에 생략되는 경우가 많아서 일대일로 매칭이나 검색을 하면 오인식 탐지의 결과가 좋지 않은 경우가 많은데 본 논문에서 제시된 다양한 음절 분리 방식을 적용함으로써 대화형 에이전트의 인식률을 높일 수 있을 것이다. 셋째, 다양한 음절 분리 방식 중에서 초성으로 분리하는 것이 가장 정확도가 높다는 점을 실증하였다. 초성 분리 방식을 사용하면 오인식을 크게 줄일 수 있을 것으로 기대된다.

본 연구는 아래와 같은 한계점을 갖는다. 먼저, 본 연구에서는 오인식 탐지를 위한 방법은 제시하지만 오인식된 정확한 단어가 어떤 것인지를 알아내는 방법을 제시하지는 않는다. 후속 연구에서 이러한 방법이 연구된다면 대화형 에이전트의 성능을 더 높일 수 있을 것이다. 또한 오인식의 원인 중에서 신조어 이외의 다양한 오류 패턴에 대한 다각도 분석 및 개선점 도출 방법을 제시하지는 않는다. 즉, 다양한 오인식의 원인에 대한 시뮬레이션 모델 구축 및 분석을 통한 문제 해결 방법의 제시가 추후 연구에서 수행될 필요가

있다. 또한, 본 연구에서 검증을 위한 레이블링 코딩에 사용된 연속 발화 쌍은 무작위로 추출되었지만 다양한 오류 원인들을 모두 포함할 수는 없기 때문에 보다 다양한 오류 사례 연구 결과를 도출하지 못하였다. 오류 원인 패턴을 유형화하여 추가적인 사례 연구를 도출할 필요가 있다.

참고문헌(References)

[국내 문헌]

- 김민정, 한지은, 강효진, 권규현.(2020). 음성 사용자 인터페이스(VUI)의 사용성 범주 추출 : AI 비서를 중심으로. *한국HCI학회 논문지*, 15(1), 53-64.
- 신조어. (2011). 신조어 연어의 형성 원리. *인문논총*, 66, 269-297.
- 신지영. (2011). 한국어의 말소리. (*No Title*).
- 유원준, & 안상준. (2021). 딥 러닝을 이용한 자연어 처리 입문. <https://wikidocs.net/book/2155>.
- 유재권, 이경미. (2011). 한국어에서의 성인과 유아의 음성 인식 비교. *한국콘텐츠학회논문지*, 11(5), 138-147.
- 이종혁. (2021). 매체 간 뉴스 동질화 현상에 대한 탐색적 연구: Doc2Vec 을 통한 문서 유사도 측정의 활용. *언론정보연구*, 58(4), 5-48.
- 이지희, 전소원, 이종태. (2017). 융합기술의 사용자 수용에 시대적 압박이 미치는 영향에 대한 연구. *한국기술혁신학회 학술대회*, (), 1355-1368.
- 이한동, & 김종배. (2017). 문서 유사도 기법을 활용한 이슈 키워드 추출 방법-인터넷 뉴스 기사를 대상으로. *예술인문사회 융합 멀티 미디어 논문지*, 7(8), 383-391.
- 정지수, 지민규, 고명현, 김학동, 임현영, 이유림, & 김원일. (2019). 문서 유사도를 통한 관련

문서 분류 시스템 연구. *방송공학회논문지*, 24(1), 77-86.

최가람, & 최성필. (2018). 단어 임베딩 (Word Embedding) 기법을 적용한 키워드 중심의 사회적 이슈 도출 연구: 장애인 관련 뉴스 기사를 중심으로. *정보관리학회지*, 35(1), 231-250.

최우빈. (2020). 대화형 에이전트의 오류 메시지 전략과 의인화가 오류 회복에 미치는 영향. *국내석사학위논문 서울대학교 대학원*, 서울.

[국외 문헌]

- Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., and Shneiderman, B. (2004). Determining Causes and Severity of End-User Frustration. *International Journal of Human - Computer Interaction*, 17(3), 333-356. https://doi.org/10.1207/s15327590ijhc1703_3
- Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, 23(1), 155-162.
- Di Gennaro, G., Buonanno, A., & Palmieri, F. A. (2021). Considerations about learning Word2Vec. *The Journal of Supercomputing*, 1-16.
- Douzi, S., Amar, M., El Ouahidi, B., & Laanaya, H. (2017). Towards a new spam filter based on PV-DM (paragraph vector-distributed memory approach). *Procedia Computer Science*, 110, 486-491.
- Friedman, B. (1995, May). "It's the computer's fault" reasoning about computers as moral agents. In *Conference companion on Human factors in computing systems* (pp. 226-227).
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*,

- 53(5), 517-527.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- Kriz, S., Anderson, G., & Trafton, J. G. (2010, March). Robot-directed speech: Using language to assess first-time users' conceptualizations of a robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 267-274). IEEE.
- Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368*.
- LuperFoy, S., Loehr, D., Duff, D., Miller, K. J., Reeder, F., & Harper, L. (1998). An architecture for dialogue management, context tracking, and pragmatic adaptation in spoken dialogue systems. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Rong, X. (2014). word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.
- Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction* (pp. 141-148).
- Sarma, A., & Palmer, D. D. (2004). Context-based speech recognition error detection and correction. In *Proceedings of HLT-NAACL 2004: Short Papers* (pp. 85-88).
- Yoon, S. Y., Hasegawa-Johnson, M., & Sproat, R. (2010). Landmark-based automated pronunciation error detection. In *Eleventh annual conference of the international speech communication association*.
- You, S., Nie, J., Suh, K., & Sundar, S. S. (2011, March). When the robot criticizes you... Self-serving bias in human-robot interaction. In *Proceedings of the 6th international conference on human-robot interaction* (pp. 295-296).

〈부록 1〉 분리 방식 별 1차 정확도 검증 결과: 25% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	11.40	16.14	18.40	17.20
공백 기준 분리	20.30	21.39	22.20	21.79
형태소 기준 분리	62.10	61.98	62.60	62.29
형태소 원형 변형 후 분리	56.00	55.91	56.80	56.35
음절 단위 분리	73.30	72.45	75.20	73.80
음절 단위 분리 초성 추출	75.50	75.05	76.40	75.72
음절 단위 분리 중성 추출	54.80	54.65	56.40	55.51
음절 단위 분리 중성 추출	5.10	5.72	5.80	5.76
음절 단위 분리 발음 유사 초성 군집	52.00	51.94	53.60	52.76
음절 단위 분리 발음 유사 중성 군집	66.50	66.21	67.40	66.80
음절 단위 분리 초성 및 중성 결합	71.50	71.04	72.60	71.81
음절 단위 분리 초성, 중성, 중성 결합	38.10	38.22	38.60	38.41
음절 단위 분리 발음 유사 초성 및 중성 군집	70.10	69.51	71.60	70.54

〈부록 2〉 분리 방식 별 2차 정확도 검증 결과: 25% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	12.60	16.73	18.80	17.70
공백 기준 분리	21.50	23.46	25.20	24.30
형태소 기준 분리	61.70	61.58	62.20	61.89
형태소 원형 변형 후 분리	58.10	58.05	58.40	58.23
음절 단위 분리	71.80	70.72	74.40	72.51
음절 단위 분리 초성 추출	77.30	76.82	78.20	77.50
음절 단위 분리 중성 추출	52.90	52.78	55.00	53.87
음절 단위 분리 중성 추출	6.10	6.53	6.60	6.57
음절 단위 분리 발음 유사 초성 군집	55.70	55.60	56.60	56.10
음절 단위 분리 발음 유사 중성 군집	65.60	65.23	66.80	66.01
음절 단위 분리 초성 및 중성 결합	71.10	70.65	72.20	71.41
음절 단위 분리 초성, 중성, 중성 결합	38.30	38.51	39.20	38.85
음절 단위 분리 발음 유사 초성 및 중성 군집	70.20	69.73	71.40	70.55

<부록 3> 분리 방식 별 3차 정확도 검증 결과: 25% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	13.60	16.79	18.40	17.56
공백 기준 분리	22.10	23.43	24.60	24.00
형태소 기준 분리	64.10	63.85	65.00	64.42
형태소 원형 변형 후 분리	59.70	59.57	60.40	59.98
음절 단위 분리	72.10	71.37	73.80	72.57
음절 단위 분리 초성 추출	76.50	75.83	77.80	76.80
음절 단위 분리 중성 추출	54.10	53.97	55.80	54.87
음절 단위 분리 중성 추출	5.20	6.25	6.40	6.32
음절 단위 분리 발음 유사 초성 군집	56.70	56.50	58.20	57.34
음절 단위 분리 발음 유사 중성 군집	66.20	65.76	67.60	66.67
음절 단위 분리 초성 및 중성 결합	70.90	70.61	71.60	71.10
음절 단위 분리 초성, 중성, 중성 결합	38.80	39.02	39.80	39.41
음절 단위 분리 발음 유사 초성 및 중성 군집	70.00	69.53	71.20	70.36

<부록 4> 분리 방식 별 4차 정확도 검증 결과: 25% 검증 데이터 사용

분리 방식	정확도 (%)	재현율 (%)	정밀도 (%)	F1 점수 (%)
전체 발화	9.90	14.39	16.20	15.24
공백 기준 분리	21.30	22.24	23.00	22.62
형태소 기준 분리	64.50	64.24	65.40	64.82
형태소 원형 변형 후 분리	59.80	59.76	60.00	59.88
음절 단위 분리	70.10	69.51	71.60	70.54
음절 단위 분리 초성 추출	76.10	75.44	77.40	76.41
음절 단위 분리 중성 추출	53.30	53.20	54.80	53.99
음절 단위 분리 중성 추출	6.50	7.93	8.20	8.06
음절 단위 분리 발음 유사 초성 군집	57.50	57.37	58.40	57.88
음절 단위 분리 발음 유사 중성 군집	64.70	64.38	65.80	65.08
음절 단위 분리 초성 및 중성 결합	71.70	71.40	72.40	71.90
음절 단위 분리 초성, 중성, 중성 결합	38.20	38.39	39.00	38.69
음절 단위 분리 발음 유사 초성 및 중성 군집	71.40	70.90	72.60	71.74

Abstract

Developing a New Algorithm for Conversational Agent to Detect Recognition Error and Neologism Meaning: Utilizing Korean Syllable-based Word Similarity

Jung-Won Lee* · Il Im**

The conversational agents such as AI speakers utilize voice conversation for human-computer interaction. Voice recognition errors often occur in conversational situations. Recognition errors in user utterance records can be categorized into two types. The first type is misrecognition errors, where the agent fails to recognize the user's speech entirely. The second type is misinterpretation errors, where the user's speech is recognized and services are provided, but the interpretation differs from the user's intention. Among these, misinterpretation errors require separate error detection as they are recorded as successful service interactions. In this study, various text separation methods were applied to detect misinterpretation. For each of these text separation methods, the similarity of consecutive speech pairs using word embedding and document embedding techniques, which convert words and documents into vectors. This approach goes beyond simple word-based similarity calculation to explore a new method for detecting misinterpretation errors.

The research method involved utilizing real user utterance records to train and develop a detection model by applying patterns of misinterpretation error causes. The results revealed that the most significant analysis result was obtained through initial consonant extraction for detecting misinterpretation errors caused by the use of unregistered neologisms. Through comparison with other separation methods, different error types could be observed.

This study has two main implications. First, for misinterpretation errors that are difficult to detect due to lack of recognition, the study proposed diverse text separation methods and found a novel method that improved performance remarkably. Second, if this is applied to conversational agents or voice

* Corresponding author: Jung-Won Lee
School of Business, Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-02-2123-5480, E-mail: neckt@yonsei.ac.kr

** School of Business, Yonsei University

recognition services requiring neologism detection, patterns of errors occurring from the voice recognition stage can be specified. The study proposed and verified that even if not categorized as errors, services can be provided according to user-desired results.

Key Words : Conversational Agent, Speech Recognition Error, Speech Analysis, Neologism Detection, Document Similarity

Received : August 30, 2023 Revised : September 12, 2023 Accepted : September 12, 2023

Corresponding Author : Jung-Won Lee

저 자 소개



이정원

연세대학교 일반대학원 경영학과 ODI-IS 전공 박사학위를 취득하였으며, 주요 관심분야는 인공지능, 텍스트 마이닝, 의인화 연구 등이다. Journal of Business Research에 논문을 게재하였다.



임 일

저자는 현재 연세대학교 경영대학 교수이다. University of Southern California에서 정보시스템 전공으로 박사학위를 받았으며 New Jersey Institute of Technology 대학에서 교수로 재직하였다. 추천시스템과 AI, 디지털 트랜스포메이션 분야에 관심이 있으며 MIS Quarterly, DSS, Information & Management, Computers in Human Behavior, Journal of Business Research, ACM Transactions on Information Systems 등에 논문을 게재하였다.