

대화형 개인 비서 시스템의 언어 인식 모듈(SLU)를 위한 미등록어(OOV) 처리 기술

이창수^o, 고영중
동아대학교

blue772001@gmail.com, youngjoong.ko@gmail.com

A Out-of-vocabulary Processing Technology for the Spoken Language Understanding module of a Dialogue Based Private Secretary Software

ChangSu Lee^o, YoungJoong Ko
Donga University, Computer Engineering

요 약

대화형 개인 비서 시스템은 사람의 음성을 통해 인식된 음성 인식 결과를 분석하여 사용자에게 제공할 정보가 무엇인지 파악한 후, 정보가 포함되어 있는 앱(app)을 실행시켜 사용자가 원하는 정보를 제공하는 시스템이다. 대화형 개인 비서 시스템의 가장 중요한 모듈 중 하나는 Spoken Language Understanding(SLU)이며, 발화의 “의미 분석”을 수행하는 모듈이다. 본 논문은 음성 인식 결과가 잘못 되어 SLU 분석이 실패하는 것을 방지하기 위하여 명사, 개체명의 음성 인식 결과를 보정 시켜주는 방법을 제안한다. 제안하는 미등록어(OOV) 처리 모듈은 미등록어 탐색 모듈과 미등록어 변환 모듈로 구성되며, 발화에서 미등록어 처리 모듈을 통해 미등록어 단어를 분류하고, 미등록어 변환 모듈을 통해 미등록어 단어를 사전에 존재하는 유사한 단어로 변환하는 방법을 제안한다. 제안한 방법을 적용하였을 때의 실험 결과, 전체 미등록어 단어 중 최대 52.5%가 올바르게 수정되었으며, “음성 인식 전의 원본 문장”과 비교하여 미등록어 처리 모듈을 적용한 결과가 음성 인식 결과를 그대로 사용했을 경우에 비해 최대 문장 단위 17.4%의 개선된 일치율을 보였다.

주제어: 대화시스템, Spoken Language Understanding, Conditional Random Fields, 미등록어 처리, 편집 거리

1. 서론

모바일기술의 발전으로 우리는 시간과 공간에 구애 받지 않고 실시간으로 원하는 정보를 얻을 수 있게 되었다. 예를 들어 버스 도착 정보를 실시간으로 확인 할 수 있으며, 스마트폰을 이용하여 실시간으로 영화 예약이 가능하게 되었다. 하지만, 최근에는 정보를 제공하는 앱(app)의 종류가 다양해지면서 정보가 기하급수적으로 늘어남에 따라 원하는 정보를 손쉽게 찾는 방법이 어려워졌다. 대화형 개인 비서 시스템은 인간에게 가장 편리한 인터페이스 중 하나인 음성 인식을 통해 사용자가 원하는 정보를 파악하고, 기존의 양질의 앱들을 통합, 활용하여 사용자에게 다양한 정보를 제공하는 시스템이다. 이러한 대화형 개인 비서 시스템의 중요한 모듈 중 하나는 Spoken Language Understanding(SLU) 모듈이다[1]. SLU 모듈은 사용자의 음성 결과를 분석하여 사용자가 원하는 정보가 무엇인지를 파악하는 “의미 분석” 모듈이

다. 대화형 개인 비서 시스템의 SLU 모듈은 모바일의 특성 상 기존의 SLU 모듈에 비해 더 자세한 분석이 요구된다[2]. 그러므로 대화형 개인 비서 시스템에서의 음성 인식 결과의 오류는 SLU의 성능에 큰 영향을 미친다. 게다가 다양한 연구에서 음성 인식 결과의 오류는 오류가 전과되는 특성으로 인해 여러 어플리케이션(개체명 인식, 구문 분석, 정보 검색 등)의 성능을 저하시키는 요인이 되고 있다.

현재 미등록어(Out-of-vocabulary, OOV) 연구는 다양한 방법으로 연구가 진행되고 있다[3][4][5]. 하지만 대화형 개인 비서 시스템은 SLU 외에 여러 모듈이 결합되기 때문에 속도 상 기존의 미등록어 처리 방법을 적용하기가 어려우며, 또한 이전 방법들은 많은 외부 데이터와 자원을 필요로 하는 문제가 있다. 그러므로 음성 인식 결과 중 오류가 많은 명사, 개체명의 오류를 수정하는 빠르고 효율적인 미등록어 처리 방법이 필요하다. 본 논문에서는 음성 인식 결과에서 명사, 개체명의 잘못된 음성 인식 결과를 수정하는 빠르고 효율적인 미등록어 처리 모듈을 제안한다. 본 논문에서 제안하는 미등록어 처리 모듈은 2개의 세부 모듈로 구성된다. 먼저 미등록어 탐색 모듈은 기계 학습 중 하나인 Conditional Random Fields(CRF)이용하여 사용자 발화에서 미등록어 후보 단

본 연구는 산업자원통상부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음.
[10041678, 다중영역 정보서비스를 위한 대화형 개인 비서 소프트웨어 원천 기술 개발]

어를 분류하고, 사전을 활용하여 미등록어 후보 단어들 중에서 미등록어 단어를 최종적으로 분류하는 모듈이다 [6]. 또한 미등록어 변환 모듈은 미등록어 단어를 기존의 사전에 존재하는 가장 유사한 단어로 변환하는 모듈이다. 우리는 미등록어 단어를 분류하고, 단어 간의 국제 음성 기호(IPA) 유사도와 단어 안의 각 음절의 초성, 중성, 종성 유사도를 편집거리 방법으로 각각 계산하여 미등록어 단어를 수정하는 방법을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 1장의 서론에 이어 2장에서는 관련연구에 대해 살펴보고, 3장에서는 미등록어 처리 모듈(미등록어 탐색 모듈, 미등록어 변환 모듈)에 대해 자세히 설명한다. 음성 인식 결과에 제안한 방법을 적용한 실험 성능을 4장에서 비교, 설명하고, 마지막으로 5장에서는 결론 및 향후 연구과제에 대하여 기술한다.

2. 관련 연구

음성 인식 결과를 이용하는 대부분 시스템들은 미등록어(OOV)에 대한 문제를 가지며, 미등록어는 대부분 명사, 개체명 같이 정보가 풍부한 단어로 구성되어 있다 [5]. 그러므로 Spoken Language Understanding(SLU) 모듈 같은 발화의 “의미 분석”을 수행하는 모듈에서는 미등록어 문제는 성능을 크게 저하시킨다. 게다가 음성 인식 결과를 이용하는 정보 검색, 개체명 인식과 같은 연구에서는 미등록어의 음성 인식 오류가 전파되어 어플리케이션의 성능을 저하 시키는 요인이 되고 있다.

현재 미등록어 문제를 해결하기 위한 다양한 연구가 진행되고 있다. 발화에서 미등록어를 탐색하기 위해 sub-word와 generic word model을 이용하여 filler models 방식을 적용한 연구가 있었으며[3], 발화에서 미등록어 구간을 찾기 위해 Confidence scores를 계산하는 Confidence estimation models을 이용한 방법이 연구되었다[4]. 또한 이 2가지 방법을 결합한 Hybrid models 또한 연구되었으며, Confusion Networks 와 CRF를 이용하여 미등록어 구간을 탐색했다[5]. 하지만 이들 연구는 많은 외부 데이터와 외부 자원이 필요하며, 대화형 개인 비서 시스템과 같이 여러 모듈이 결합되는 시스템에서는 속도 문제와 많은 요구 자원을 필요로 하는 측면에서 적용하기가 어렵다.

또한 한국어를 기준으로 미등록어를 해결하기 위한 연구도 진행되었다. 언어 지식 기반 미등록어 인식 방법이 있었으며 이 방법은 형태소 패턴, 어절 내 형태소 결합 정보와 같은 언어 지식을 바탕으로 미등록어를 인식하는 방법이다[7][8]. 이러한 방법은 형태소 분석이 앞서 수행되어야 하는데 우리의 시스템에서는 미등록어 분석 이후 작업으로 SLU 모듈을 위해 형태소 분석이 수행되므로 속도 상 형태소 분석을 2번 수행하는 것은 비효율적이다.

편집 거리 문제는 하나의 문자열을 다른 문자열로 변환 할 때 필요한 최소한의 연산의 개수를 구하는 문제이며, 주어진 문자열에 대해 비슷한 문자열을 찾는 알고리즘에 활용된다[9]. 문자 판독 장치나 문서 이미지 정보

분석, 맞춤법 검사기 등이 그 예이다[10][11]. 일반적인 편집 거리가 두 개의 문자열 사이에서 계산되는 것과 달리 최근에서 촘스키 위계상에 존재하는 형식 언어를 대상으로 하여 편집 거리를 계산하는 연구가 현재 진행되고 있다[12][13]. 우리 시스템은 단어 간의 유사도를 구하기 위해 이러한 편집 거리 방법을 활용한다.

3. 미등록어(OOV) 처리 모듈

본 논문에서 제안하는 미등록어 처리 모듈을 통해 음성 인식 결과를 수정하는 전체 과정의 실제 예는 [그림 1]에서 볼 수 있다.

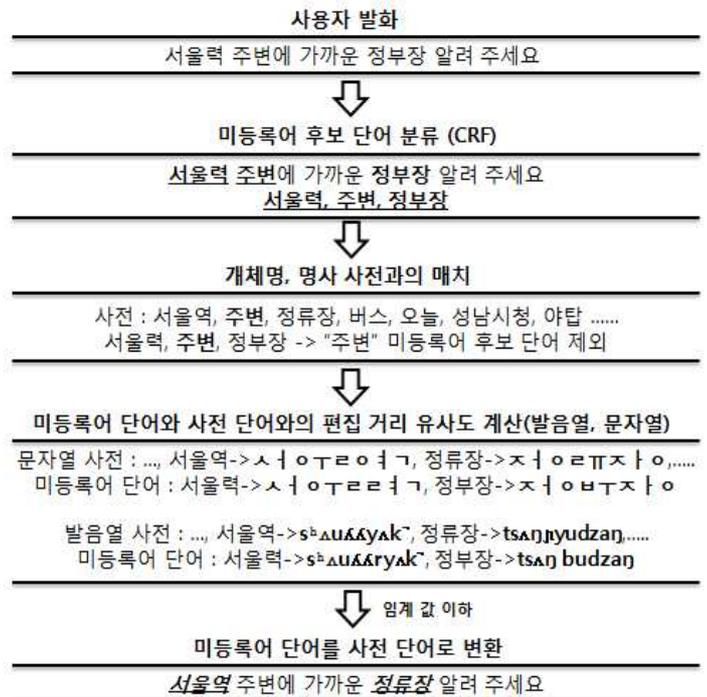


그림 1. 미등록어 처리 모듈 진행 과정의 실제 예

그림 1에서 보는 바와 같이 음성 인식 결과에서 잘못된 인식된 미등록어(명사, 개체명)를 수정하기 위한 미등록어 처리 모듈을 제안하며 이 모듈은 크게 2개의 세부 모듈로 구성된다.

1. 미등록어 탐색 모듈
2. 미등록어 변환 모듈

3.1 미등록어(OOV) 탐색 모듈

미등록어 탐색 모듈은 발화에서 미등록어 단어를 분류하는 모듈이며, 다음과 같이 2단계로 구성되어 있다.

1. CRF를 이용해 사용자의 발화에서 미등록어 후보 단어를 분류하는 단계
2. 분류된 미등록어 후보 단어에서 사전과의 매치를 통해 최종 미등록어 단어를 분류하는 단계

미등록어 탐색 모듈의 전체 구성도는 [그림 2]와 같다.

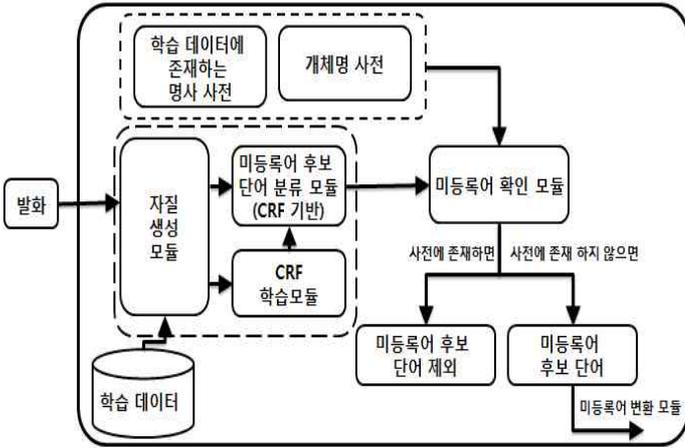


그림 2. 미등록어 탐색 모듈의 전체 구성도

[그림 2]에서 보는 바와 같이 다음과 같은 순서로 미등록어 탐색 모듈이 수행된다.

1. 사용자의 발화를 음성 인식을 통해 입력 받음
2. 미등록어 후보 단어가 태깅되어 있는 학습 데이터를 가지고 자질을 구축하여 생성한 CRF 모델을 이용해 발화의 미등록어 후보 단어(개체명, 명사)를 분류
3. 미등록어 후보 단어 중 개체명과 학습데이터에서 출현한 명사 사전과의 매치를 통해 사전에 존재하는 미등록어 후보 단어를 미등록어 단어에서 제외
4. 사전에 매치되지 않은 미등록어 후보 단어는 미등록어 단어로 간주하고 미등록어 변환 모듈에 전달

이와 같은 절차를 통해 발화에서 미등록어 단어를 분류했으며, 분류를 위해 사용한 자질은 [표 1]과 같다.

[표 1] 미등록어 후보 단어 분류를 위한 자질 집합

음절 단위 어휘 자질 (: 현재음절)	$w_{-1}/w_0, w_0/w_1, w_{-1}/w_0/w_1$ $w_{-2}/w_{-1}/w_0, w_0/w_1/w_2,$ $w_{-1}/w_0/w_1/w_2, w_{-2}/w_{-1}/w_0/w_1,$ $w_{-2}/w_{-1}/w_0/w_1/w_2$
음절 단위 문맥 자질	$t_{-1}, t_1, t_{-1}/t_1, t_{-2}/t_{-1}, t_1/t_2,$ $t_{-1}/t_1/t_2, t_{-2}/t_{-1}/t_1,$ $t_{-2}/t_{-1}/t_1/t_2$

w_0 은 현재 음절을 나타내며, w_{-1}, t_{-1} 은 현재 음절을 기준으로 바로 앞의 음절, w_1, t_1 은 현재 음절을 기준으로 바로 뒤에 음절을 의미한다.

3.2 미등록어(OOV) 변환 모듈

미등록어 변환 모듈은 단어 간의 유사도를 계산해 미등록어 단어를 사전에 존재하는 단어로 변환해 주는 모듈이다. 우리는 음성 인식 결과를 개선하기 위해 2가지의 서로 다른 유사도 계산 방법을 제안한다.

1. IPA를 중심으로 사전에 있는 단어와의 발음열 유사도를 편집거리를 이용해 계산 하는 방법
2. 단어를 각 음절별로 초성, 중성, 종성으로 분리하여 모음, 자음별 유사도를 편집거리를 이용해 계산하는 방법

미등록어 단어를 수정하기 위한 미등록어 변환 모듈의 전체 구성도는 [그림 3]와 같다.

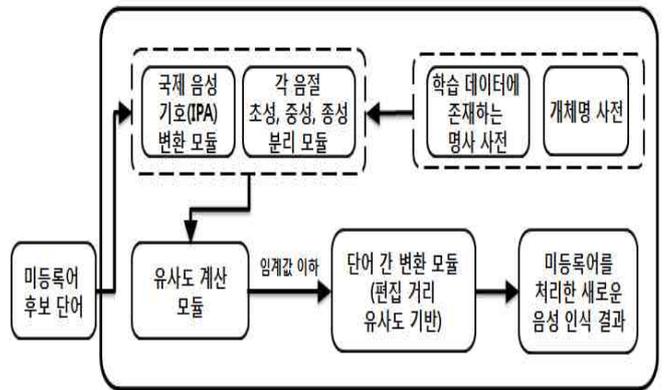


그림 3. 미등록어 변환 모듈의 전체 구성도

[그림 3]에서 보는 바와 같이 미등록어 변환 모듈은 다음과 같은 절차로 수행 된다.

1. 미등록어 탐색 모듈을 통해 미등록어 단어들을 입력 받음
2. 각 미등록어 단어에 대해 IPA 변환 모듈을 이용해 발음열을 구축하고, 각 음절 초성, 중성, 종성 분리 모듈을 통해 문자열을 구축
4. 유사도 계산을 위한 2가지 방법을 이용해 미등록어 단어와 사전 단어 간의 유사도를 계산
5. 단어의 유사도가 임계값 이하이면 사전에 있는 단어로 미등록어 단어를 변환

유사도를 구하기 위한 2가지 방법 모두 편집거리[9] 방법을 이용하여 최종적으로 발음열, 문자열의 유사도를 계산한다. 널리 알려진 바와 같이 편집 거리 문제는 하나의 문자열을 다른 문자열로 변환할 때 필요한 최소한의 연산 개수(삽입, 삭제, 치환 연산)를 구하는 문제이다.

단어 간 유사도 계산을 위한 첫 번째 방법은 IPA를 중심으로 편집 거리 유사도를 계산하는 방법이다. 각 단어의 발음열은 국립국어원의 표준 발음 변환기를 이용하여 구축했다. [그림 4]는 단어가 발음열로 변환되는 예를

보여준다.



그림 4. 국제 음성 기호로 변환한 단어의 발음열의 예

또한 [그림 4]의 방법을 이용해 미등록어 단어의 발음열을 구축한 뒤, 사전에 존재하는 단어들의 발음열 편집거리 유사도의 계산해 미등록어 단어를 변환하며, [그림 5]에서 예를 보인다.



그림 5. 발음열의 유사도를 구해 단어 간 변환을 수행하는 실제 예

두 번째 유사도 계산 방법은 미등록어 단어를 각 음절에 대해 초성, 중성, 종성으로 분리해 각 모음, 자음으로 구성된 문자열을 만들어 사전 단어와의 유사도를 구하는 방법이다. [그림 6]은 단어가 문자열로 변환되는 예를 보인다.

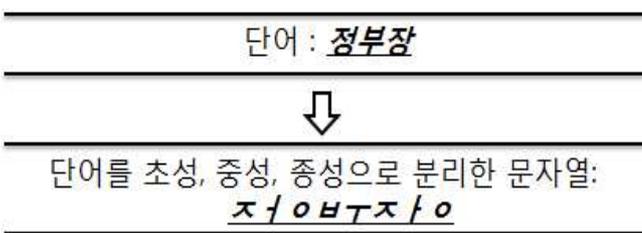


그림 6. 단어를 초성, 중성, 종성으로 분리한 예

또한 [그림 6]의 방법을 이용해 미등록어 단어를 변환한 후, 사전 단어들 간의 문자열 유사도를 계산해 미

등록어 단어를 수정하는 예는 [그림 7]과 같다.

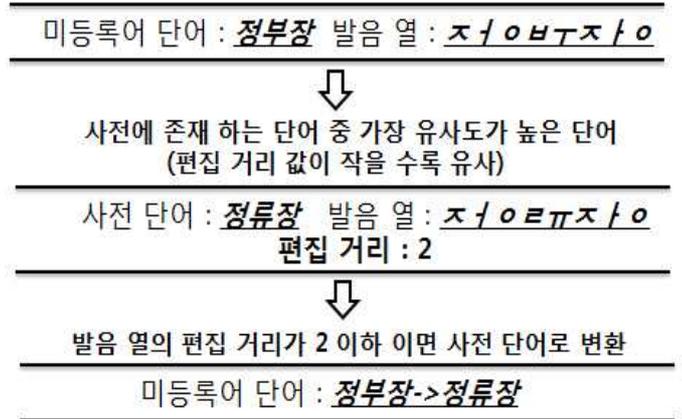


그림 7. 문자열의 유사도를 구해 단어 간 변환을 수행하는 실제 예

4. 실험결과

실험결과는 크게 4가지로 분류하여 평가했다.

1. 미등록어 후보 단어 분류 모듈(CRF)의 성능
2. 전체 미등록어 단어 개선율
3. 미등록어 변환 모듈을 기준으로 한 단어 개선율
4. 문장 단위 개선율

4.1 실험데이터

우리는 개체명 사전을 구축하기 위해 ETRI 개체명 사전을 사용했으며, 미등록어 처리 모듈을 위한 데이터는 구글 음성 인식기를 이용하여 교통 도메인에 대해서 207개의 문장을 구축했으며, 문장 안에서의 미등록어 후보 단어는 대화형 개인 비서 시스템의 데이터를 구축하는 전문가가 태깅을 수행했다. 또한 5-fold cross validation[14]을 이용해 성능을 평가했다.

4.2 미등록어 후보 단어 분류 모듈 성능 결과

[표 2]는 미등록어 탐색 모듈 중 가장 중요한 미등록어 후보 단어 분류 모듈의 성능을 보인다.

[표 2] 미등록어 후보 분류의 성능

미등록어 후보 분류 성능
90.24%

실험 결과, 미등록어 후보 단어(개체명, 명사)를 높은 확률로 분류한 것을 확인 할 수 있었다.

4.3 전체 미등록어 단어 개선율

전체 미등록어 단어 개선율은 전체 미등록어 단어들 중에서 얼마나 많은 미등록어 단어가 미등록어 처리 모듈을 통해서 올바르게 수정되었는지에 대한 성능이다. 우리는 미등록어 변환 모듈을 2가지 방법(IPA, 단어의 각 음절별 초성, 중성, 종성 분리)을 이용했으며, 각 성능은 [표 3]과 같다.

[표 3] 전체 미등록어 단어 개선율

	IPA 방법	음절 별 초성, 중성, 종성 분리 방법
미등록어 단어 개선율 (음성 인식 결과를 0%로 가정했을 경우 기준)	52.2%	41.7%

음성 인식 결과를 그대로 사용했을 경우와 비교해 IPA 방법은 전체 미등록어 단어 중 52.2%가 올바르게 수정되었으며, 음절 별 초성, 중성, 종성 분리 방법은 41.7%가 올바르게 수정되어, 우리가 제안하는 방법이 효율적인 방법임을 알 수 있었다.

4.4 미등록어 변환 모듈 성능

미등록어 탐색 모듈을 통해 전체 미등록어 단어를 정확히 분류할 수는 없다. 즉, 미등록어 탐색 모듈은 오류가 발생하기 때문에 탐색 모듈을 통해 올바르게 분류된 미등록어 단어에 대해서만 계산한 단어 개선율을 분석한다. 즉, 미등록어 분류 모듈만의 성능을 [표 4]에서 보인다.

[표 4] 미등록어 변환 모듈을 기준으로 한 단어 개선율

	IPA 방법	음절 별 초성, 중성, 종성 분리 방법
미등록어 단어 개선율 (음성 인식 결과를 0%로 가정했을 경우 기준)	63.7%	56.4%

4.5 문장 단위 개선율

“음성 인식 전의 원본 문장”과 비교하여, 미등록어 처리 모듈을 적용하였을 경우가 음성 인식 결과를 그대로 사용하였을 때 보다 얼마나 원본 문장과 같아졌는지를 문장 단위 성능으로 나타냈으며, [표 5]에서 성능을 확인 할 수 있다.

[표 5] 미등록어 모듈 처리 결과 후 문장 단위 성능

음성 인식 결과	음성 인식 결과에 IPA 방법을 적용	음성 인식 결과에 음절 별 초성, 중성, 종성 분리 방법을 적용
67.6%	85%	84%

원본 문장과 비교하여 음성 인식 그대로의 결과는 문장 단위 67.6%가 같았으며, 제안하는 모듈을 적용한 결과 최대 85%까지 원본 문장과 같아짐을 확인 할 수 있었다.

5. 결론

본 논문에서는 대화형 개인 비서 시스템에서 가장 중요한 모듈 중 하나인 SLU 모듈의 성능을 개선시킴과 동시에 전처리 작업으로 음성 인식 결과를 사용하는 다양한 연구에서 음성 인식 결과의 오류가 전파되는 문제를 해결하기 위해 음성 인식 결과를 보정하는 빠르고 효율적인 미등록어 처리 모듈을 제안했다. 우리가 제안한 미등록어 처리 모듈은 2개의 세부 모듈(미등록어 탐색 모듈, 미등록어 변환 모듈)로 구성되며 미등록어 탐색 모듈을 통해 미등록어 단어를 분류하고, 미등록어 변환 모듈을 통해 미등록어 단어가 바르게 수정되는 방법 보였다. 교통 도메인 데이터를 이용하여 실험한 결과, 음성 인식 결과를 그대로 사용한 방법에 비해 우리가 제안한 방법이 전체 미등록어 단어 중 최대 52.2%의 단어 개선율을 보였다. 또한 “음성 인식 전의 원본 문장”과 비교하여, 미등록어 처리 모듈을 적용했을 때의 결과가 음성 인식 결과를 그대로 사용했을 경우 보다 최대 문장 단위 17.4%의 개선된 일치율을 보였다.

향후 연구로는 국제 음성 기호(IPA)를 이용한 편집거리 유사도 방법이 단어의 각 음절 별 초성, 중성, 종성을 분리하여 편집거리 유사도 방법보다 더 효율적임을 알게 되었기 때문에 국제 음성 기호(IPA)를 사용하여 편집거리 유사도를 계산할 때 각 발음 기호 간의 밀접성을 분석하여, 가중치를 더 효율적으로 주는 방법을 연구 할 것이다. 또한 길이가 일정치 않는 발음열에 대해서 길이를 가장 효율적으로 일치시켜 분석하는 방법에 대해 연구 할 것이다.

참고문헌

- [1] Wang, Y.Y. and Acero, A. "Spoken language understanding : an introduction to the statistical framework. *IEEE signal Processing Magazine* 22(5) : 16-31 2005
- [2] 이창수, 고영중 "대화형 개인 비서 시스템의 언어 인식 모듈을 위한 개체명과 문장목적 동시 인식 방법" *정보과학회논문지 :소프트웨어 및 응용*, 2014
- [3] Bisani, M and Ney, H "Open vocabulary speech recognition with flag hybrid models." In *INTERSPEECH*, 2005
- [4] Burget, L., Schwarz, P., Matejka, P., Hannemann, M., Rastrow, A., White C., Khudanpur S., Hermansky, H. and Cernocky. J. "Combination of strongly and weakly constrained recognizers for reliable detection of OOVs." In *ICASSP*. 2008
- [5] Carolina P., Mark D., Denis F., and Fred J.,

- “Contextual information improves oov detection” in speech, 2010
- [6] Lafferty, J. McCallum, A.Pereira, F., "Conditional random fields : Probabilistic models for segmenting and labeling sequence data", ICML, pp.282-2289, 2001.
- [7] 양장모, 김민정, 권혁철 “언어정보를 이용한 한국어 미등록어 추정”, 한국정보과학회 봄 학술발표논문집, 제 23권 제1호, 957-960, 1996
- [8] 차정원, 이원일, 이근배, 이종혁 “형태소 패턴 사전을 이용한 일반화된 미등록어 처리”, 정보과학회 인공지능연구회 춘계학술대회 논문집, 37-42, 1997
- [9] 노강호, 김진욱, 김은상, 박근수, 조환규, “한글에 대한 편집 거리 문제”, *정보과학회논문지:시스템 및 이론*, 제37권, 제2호, pp.103-109, 2010.
- [10] Srihari, S.N. “Computer Text Recognition and Error” Correction. IEEE Computer Society Press, LosAlamitos, CA, USA, 1985
- [11] Kukich, K. “Techniques for automatically correcting words in text” *ACM Computing Surveys*, 24, 377-439, 1992
- [12] Bunke, H. “Edit distance of regular languages.” *In Proc. 5th Annual Symposium on Document Analysis and Information Retrieval*, 113-124 , 1996
- [13] Konstantinidis, S. “Computing the edit distance of a regular language” *Information and Computation*, 205, 1307-1316, 2007
- [14] Janez, D. "Statistical Comparisons of Classifiers over Multiple Data sets" *Journal of Machine Learning Research* 7(2006) 1-30.