

Performance Improvement Methods of a Spoken Chatting System Using SVM

HyeokJu Ahn[†] · SungHee Lee^{††} · YeongKil Song^{†††} · HarkSoo Kim^{††††}

ABSTRACT

In spoken chatting systems, users' spoken queries are converted to text queries using automatic speech recognition (ASR) engines. If the top-1 results of the ASR engines are incorrect, these errors are propagated to the spoken chatting systems. To improve the top-1 accuracies of ASR engines, we propose a post-processing model to rearrange the top-n outputs of ASR engines using a ranking support vector machine (RankSVM). On the other hand, a number of chatting sentences are needed to train chatting systems. If new chatting sentences are not frequently added to training data, responses of the chatting systems will be old-fashioned soon. To resolve this problem, we propose a data collection model to automatically select chatting sentences from TV and movie scenarios using a support vector machine (SVM). In the experiments, the post-processing model showed a higher precision of 4.4% and a higher recall rate of 6.4% compared to the baseline model (without post-processing). Then, the data collection model showed the high precision of 98.95% and the recall rate of 57.14%.

Keywords : Spoken Chatting System, Re-Rank Model, Data Collection Model, RankSVM, SVM

SVM을 이용한 음성채팅시스템의 성능 향상 방법

안 혁 주[†] · 이 성 희^{††} · 송 영 길^{†††} · 김 학 수^{††††}

요 약

음성채팅시스템에서 사용자의 음성 질의는 자동음성인식기를 통하여 텍스트 질의로 변환된다. 만약 자동음성인식기의 1순위 결과가 틀린다면 이 오류는 그대로 음성채팅시스템에 전파된다. 자동음성인식기의 1순위 정밀도를 향상시키기 위하여 본 논문에서는 RankSVM을 이용하여 자동음성인식기의 n개 결과를 재순위화하는 후처리 모델을 제안한다. 채팅시스템을 학습하기 위해서는 대용량의 채팅 문장들이 필요하다. 만약 새로운 채팅 문장들이 학습데이터에 자주 추가되지 않는다면 채팅시스템의 응답은 급방 진부해질 것이다. 이러한 문제를 해결하기 위하여 본 논문에서는 SVM을 이용하여 TV와 영화 시나리오로부터 채팅 문장들을 자동으로 선택하는 데이터 수집 모델을 제안한다. 실험에서 제안된 후처리 모델은 후처리를 하지 않은 모델보다 정확률에서 4.4%, 재현율에서 6.4% 더 좋은 결과를 보였다. 그리고 제안된 데이터 수집 모델은 98.95%의 높은 정확률과 57.14%의 재현율을 보였다.

키워드 : 음성채팅시스템, 재순위화 모델, 데이터수집 모델, RankSVM, SVM

1. 서 론

음성채팅기술은 기존의 채팅 기술에 음성인식기술을 더한 것으로 기존의 키보드를 이용한 대화방법에서 한 단계 성장

한 대화방법을 제시하였다. 기업에서는 음성채팅기술을 통한 다양한 기기를 사용자들에게 선보여왔는데 대표적으로 네이버 케이션과 애플의 시리(Siri)[1], 삼성의 S보이스[2], LG의 Q보이스[3]와 같은 스마트폰 음성채팅기술이 주를 이루고 있다. 음성채팅기술은 화자의 말을 정확히 인식하는 데서 시작한다. 다시 말해서 음성인식 정확률에 따라 전체 시스템의 성능이 변하기 때문에 음성인식에 대한 연구는 매우 중요하다. 그러나 음성인식 성능에 관한 연구만으로는 채팅시스템의 정확률을 향상시킬 수 없다. 목적 지향의 채팅이 아닌 일상적인 대화를 이끌어가는 채팅시스템은 질문/답변의 채팅 말뭉치를 기반으로 하는 경우가 많다. 이때, 말뭉치가 많은 데이터를 보유하지 않으면, 시스템이 화자의 말을 정확히 인식했으나

※ 이 논문은 산업통상자원부/미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음[10041678, 다중영역 정보서비스를 위한 대화형 개인 비서 소프트웨어 원천 기술 개발]. 또한 본 연구는 우수기술연구센터 사업 중 “링크드데이터 기반 대화형 질의응답 검색 프레임워크 개발(과제번호 : 10048448)” 과제의 지원을 받아 수행되었음.

† 준 회 원 : 강원대학교 컴퓨터정보통신공학 전공 석사과정

†† 비 회 원 : 강원대학교 컴퓨터정보통신공학 전공 석사과정

††† 준 회 원 : 강원대학교 컴퓨터정보통신공학 전공 박사과정

†††† 정 회 원 : 강원대학교 컴퓨터정보통신공학 전공 교수

Manuscript Received : January 7, 2015

First Revision : April 27, 2015

Accepted : April 28, 2015

* Corresponding Author : HarkSoo Kim(nlprkim@kangwon.ac.kr)

그에 대한 만족스러운 답변을 할 수 없는 경우가 발생한다. 따라서 화자의 말을 인식하여 매칭할 수 있는 다양한 문장들이 채팅시스템 내에 존재해야 한다. 그래서 본 논문에서는 음성채팅기술의 성능 향상을 위해 음성인식 성능을 높이는 방법과 다양한 채팅 표현들을 수집할 수 있는 방법을 제안한다.

2. 관련 연구

음성인식기 성능 향상에 대한 연구는 크게 음성이 텍스트로 변환될 때의 음성 주파수를 분석하는 연구와 음성인식 결과에 대한 후처리 연구로 나뉜다. 음성 주파수를 분석하여 이용하는 연구에서 Lee(2010)[4]는 발성속도의 변동에 따른 음성인식 성능 저하를 보상하기 위해 주어진 음성으로부터 발성속도를 정량적으로 나타낼 수 있는 변수를 도입하여 발성속도에 따라 선택적으로 시간축 변환을 적용하였다. Lee(2004)[5]는 남녀에 따른 음성 주파수를 비교하여 음성의 극성을 판별, 음성신호가 들어오면 주파수의 극성에 따라 음성을 인식하여 오류율을 낮추어 음성인식 성능을 향상시키는 방법을 제시하였다. Cho(2012)[6]와 Choi(2013)[7]는 왜곡된 음성 스펙트럼을 분석하여 스펙트럼 피크를 향상시키거나 주파수 기울기를 사람의 음성 스펙트럼에 맞추어 보상하는 방법을 사용하여 잡음에 대한 음성인식 손실을 보상하였다. Choi(2012)[8]는 기존의 MFCC(Mel Frequency Cepstral Coefficient)[9] 특징 추출 방법에서 HMM 음향 모델링을 거쳐 감마톤 특징을 추출하는 음향 모델링에 대한 연구를 진행하였다. 음성인식 결과에 대한 후처리 연구는 주로 형태소 분석이나 패턴 분석을 이용하여 이루어진다. Lim(2006)[10]은 음성인식 결과에서 띄어쓰기 오류로 인해 인식 결과에 대한 이후의 정보처리를 어렵게 하는 요인을 막기 위해 음성인식 결과를 형태소 분석하여 품사정보를 이용한 어절 재결합 기법을 기본 알고리즘으로 사용하고 추가로 Bi-gram 및 4-gram 정보를 이용하는 띄어쓰기 오류 교정방법을 제시하였다. Song(2005)[11]은 사용자가 직접 발화한 문장들로 구성된 순차패턴을 비교하여, 순차규칙상 얻을 수 있는 단어를 결정하여 음성인식기 단어 후보들의 확률값을 보정하는 방법을 제안하였다. 이상의 연구들 중 음성 주파수를 이용하는 연구는 음성인식 모델을 수정하는 방법이므로 자체 음성인식기를 보유하지 않았다면 연구를 진행할 수가 없다. 후처리 방법 역시 단순히 띄어쓰기에 의한 오류를 수정하는 방법이거나, 음성인식기의 어휘별 확률과 같은 정보를 사용하는 방법이기 때문에, 이와 같은 정보를 제공하는 음성인식기를 사용해야 한다. 그러나 대부분의 개발자들은 공개되어 쉽게 사용할 수 있는 구글음성인식기를 이용하기 때문에 위 연구들을 쉽게 적용할 수 없다. 그래서 본 논문에서는 RankSVM(Support Vector Machine for Ranking)[12]을 이용한 후보문장 재순위화에 대한 연구를

제안한다. 제안 방법은 음성인식 결과가 단일문장이 아닌 후보문장이 존재할 것이라는 생각에 착안하였고 기존의 SVM(Support Vector Machine)[13]을 응용한 RankSVM 기계학습기법을 이용하여 후보문장을 재순위화한다.

음성채팅시스템 성능 향상의 또 다른 이슈인 다양한 채팅 표현들을 위한 연구는 대표적으로 국내에서 가장 인지도 있는 인공지능 채팅시스템인 심심이(2002)[14]가 있다. 심심이는 유저들이 시스템에 직접 가르치는 방식으로 유저가 직접 발화/응답 쌍을 입력하면 시스템이 자동으로 인식한다. 그 외에 Kim(2009)[15]은 핵심어에 대한 리스트를 마련, 핵심어를 포함한 어절을 말뭉치에서 선택하고, 해당 어절 주위에 있는 어절의 출현 정보와 구문 정보를 이용하여 후보문장들을 생성했다. 그리고 복잡한 자질 정보를 이용하여 후보문장에 대한 응답문장을 선택하였다. 심심이(2002)에서 제공하는 방식은 다양한 표현을 구축할 수 있으나 유저가 직접 발화/응답 쌍을 생성한다는 점에서 많은 시간을 소요한다. Kim(2009)이 제안하는 연구는 발화/응답 쌍을 빠르게 생성할 수 있으나 응답에 대한 표현이 다양하지 않다는 단점이 있다. 따라서 본 논문에서는 다양한 채팅표현을 비교적 적은 시간들 들여 가져올 수 있는 대본 말뭉치를 이용한 채팅 말뭉치 구축에 대한 연구를 제안한다. 제안 방법은 질문과 그에 대한 답변이 명백히 존재하고 표현이 다양한 대본 말뭉치를 이용하여 SVM 기계학습기법을 이용하여 발화/응답 쌍이 될 수 있는 문장을 판별하여 음성채팅시스템 말뭉치에 추가한다.

본 논문의 구성은 다음과 같다. 3절에서 제안 방법에 대하여 설명하고, 4절에서 제안한 두 가지 방법의 실험 결과를 보인다. 그리고 5절에서 결론 및 향후 연구를 제시한다.

3. 제안 방법

3.1 전체 시스템 개요

본 논문에서는 채팅시스템의 성능 향상을 위해 입력에 대

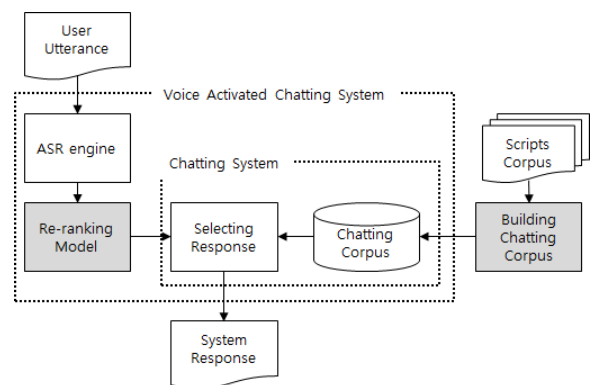


Fig. 1. The Position of Proposed Method in Voice Activated Chatting System

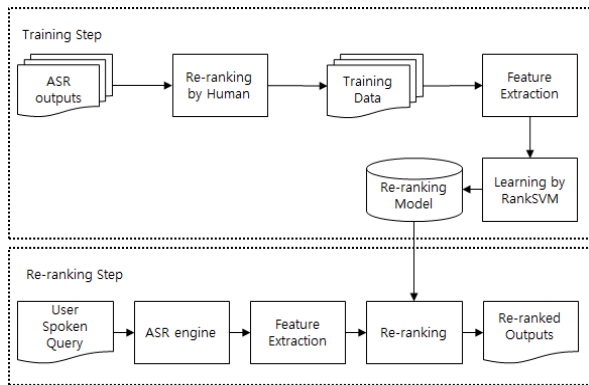


Fig. 2. Re-ranking Model for Google ASR Outputs

한 재현율을 높이기 위한 두 가지 방법을 제안한다. 첫 번째는 음성인식 성능 향상을 위한 방법이고, 두 번째는 채팅 말뭉치에 다양한 표현을 수집하기 위한 방법이다. 제안 방법은 각각의 기능을 하는 모듈로 음성채팅시스템에 사용될 수 있다. Fig. 1은 제안 방법이 전체 시스템의 어느 부분에 위치하는지를 보여준다.

음성채팅시스템은 입력으로 음성인식 결과를 채팅시스템에 넘겨준다. 이때 제안 방법의 첫 번째 모듈에서 음성인식 결과를 재순위화하여 보정된 결과를 넘긴다. 보정된 결과를 받은 채팅시스템은 색인된 질의/응답 말뭉치로부터 응답을 추출하게 되는데, 이때 질의/응답 말뭉치는 제안 방법의 두 번째 모듈에 의해 확장된 것이다.

3.2 RankSVM을 이용한 재순위화

1) 시스템 개요

본 논문에서는 음성인식 결과 후보문장 재순위화를 위해 RankSVM(Support Vector Machine for Ranking)[10]을 사용하였다. RankSVM은 일반적으로 초평면을 생성하여 데이터를 분류하는 SVM과 달리 초평면과의 거리를 계산하여 각각의 데이터에 점수를 부여하는 방식으로 데이터의 순위를 측정할 수 있는 기계학습기법이다. Fig. 2는 RankSVM을 이용한 재순위화 방법을 보여준다.

제안 방법은 학습 단계와 재순위화 단계로 구성된다. 학습 단계는 후보문장 재순위화 모델을 생성하기 위한 단계이다. 먼저 구글 ASR 결과로 나오는 1위문장 점수와 후보문장들에 정답을 직접 부착하여 정답 셋을 생성한 뒤, 필요한 자질을 추출하여 RankSVM 학습데이터를 생성한다. 학습데이터 생성이 완료되면 RankSVM을 이용하여 학습한 뒤, 재순위화 모델을 생성한다. 재순위화 단계는 사용자가 발화한 음성인식 결과를 제안 방법에 맞도록 재구성하는 단계로 음성인식 결과가 들어오면 해당 후보문장의 순위를 재순위화하여 그 결과를 출력한다.

2) 실험 데이터 정제 및 정답순위 부착

구글 음성인식 결과는 음성인식기(ASR : Automatic Speech Recognizer) 1위 문장의 점수, 그리고 사용자가 발화한 문장에 대한 결과로 출력되는 문장 후보들로 구성되어있다. 실제 정답과 ASR순위, RankSVM을 이용해 출력된 순위를 비교하기 위해 후보문장들에 정답을 직접 부착한다. 본 논문에서는 6명의 사용자가 구글 ASR을 통해 주어진 1,000개의 문장을 각자 발화하여 나온 6,000개의 결과를 기준에 맞게 부착하였다. Table 1은 후보문장들에 정답을 부착한 예를 보여준다.

Table 1. Example of Correct Rank Attachment

ASR Rank	Correct Rank	Candidate Sentence
Utterance		내 말을 이해를 못하겠어?
1	1	내 말을 이해를 못하겠어
2	6	내 말을 이해를 못하게써
3	6	내 말을 이해를 못하게 써
4	2	내 말을 이해 못하겠어
5	6	내 말을 이해를 못하겠어 아아

현재 공개된 구글 ASR은 최대 5개의 후보문장을 출력하며, 후보문장이 부족할 경우 1~4개를 출력한다. 후보문장들의 정답 부착은 해당 문장이 정답문장과 얼마나 일치하는지를 최우선으로 하여 실제 발화문장과 일치도, 비슷한 의미의 문장이 나오는지 고려하여 ASR 순위의 순서를 정한다. 후보문장이 실제 발화문장과 전혀 일치하지 않거나 문장 뒤에 추임새가 붙을 경우, 최하위 순위(후보문장 수보다 낮은 순위)를 매겨 다른 문장 후보와 차등을 준다.

3) 자질 생성

Table 2는 본 논문에서 사용한 자질의 종류이다.

Table 2. Features for Re-ranking Model

Feature Type	Feature No.	Feature Description
Considering the Google ASR Performance Features	1	Google ASR rank
	2	Score of a 1st-ranked candidate sentence
Morpheme Features	3	Morpheme bi-gram
	4	POS bi-gram
	5	The # of unknown words by domain dictionary
	6	Morpheme/POS pairs that not exist in domain dictionary
	7	The # of unknown words by general dictionary
	8	Morpheme/POS pairs that not exist in general dictionary

RankSVM에 쓰이는 자질 유형은 구글 ASR의 성능을 고려하는 자질과 형태소 자질로 나뉜다. 구글 ASR의 성능을 고려하는 자질은 총 2가지로 구글 ASR 순위 자질은 구글 음성인식기 자체에서 출력한 순위를 가지는 자질이다. 1위 후보문장 점수는 구글 음성인식기 결과 1위인 후보문장의 점수이며, 1위 외의 후보문장은 점수를 가지지 않기 때문에 점수에 대한 자질에서 제외하였다. 점수 분포에 대한 자질을 고려하기 위해 Table 3와 같이 1위 후보문장의 점수를 0.1단위로 분류하여 분포를 측정하였다.

Table 3. Distribution of 1st-ranked Candidate Sentence Scores

Candidates Score Range	The # of 1st-ranked Output	The # of 1st-ranked Output with Correct	Precision at the Range
0.0~0.1	8	2	0.2500
0.1~0.2	92	8	0.0869
0.2~0.3	314	53	0.1680
0.3~0.4	692	154	0.2225
0.4~0.5	1,094	347	0.3171
0.5~0.6	1,473	607	0.4121
0.6~0.7	1,341	776	0.5786
0.7~0.8	986	751	0.7540
Sum	6,000	2,698	

그 결과 구글 ASR 순위와 정답 순위가 모두 1위인 개수가 두드러지는 0.5 이상의 점수를 기준으로 하여 자질값을 분류하였다. 후보문장 점수에서 0.8 이상은 출력되지 않았기 때문에 0.0~0.8까지의 점수를 측정하였다. 문장의 특징을 추출하기 위해 형태소 분석 결과를 이용하였다. 문장을 형태소 분석한 결과로부터 추출할 수 있는 다양한 자질들(띄어쓰기, n-Gram, 후보문장 간의 단어 포함관계 등)을 실험한 결과 Table 2의 3~8번에 기술된 자질이 성능에 큰 영향을 미침을 확인하였다. 형태소 바이그램과 품사 바이그램은 후보문장을 형태소 분석하여 추출된 결과물에서 띄어쓰기를 포함하여 형태소 바이그램과 품사 바이그램 자질을 생성한다. 띄어쓰기까지 포함한 이유는 두개 이상의 음절이 한 어절로 표현되거나 두 어절 이상으로 표현될 수 있기 때문이다. 도메인사전 미등록어 개수에 대한 자질은 본 논문에서 실험데이터로 추출한 1,000개의 문장을 제외한 13,000문장의 채팅말뭉치를 기준으로 체언, 용언, 미등록어를 추출하여 구현한 사전을 이용한 자질이다. 해당 도메인사전을 이용하여 형태소 분석한 후보문장의 체언, 용언, 미등록어 중 사전에 없는 형태소의 형태소/품사 정보와 개수를 자질로 사용하였다. 일반 사전은 세종말뭉치[16]와 트위터 데이터를 이용하여

구축한 사전으로 좀 더 포괄적인 단어들을 가지고 있다. 일반 사전 미등록어 개수에 대한 자질과 일반사전 미등록어 형태소/품사 자질을 생성하는 방법은 도메인사전과 같다.

3.3 대본 말뭉치로부터 채팅 말뭉치 구축

1) 시스템 개요

Fig. 3은 대본으로부터 채팅 말뭉치를 구축하는 연구를 제안하는 방법의 순서도이다. 시스템은 학습 단계와 적용 단계로 구성된다. 학습 단계는 드라마/대본 말뭉치에서 화자, 대사를 추출한 뒤, 시스템 입력 데이터 형식에 맞게 정제한다. 최종적으로 정제된 데이터는 발화/응답의 문장 쌍으로 구성된다. 실험에 쓰일 데이터를 구축하기 위해 발화/응답 문장 쌍에서 테스트 데이터를 따로 추출한다. 정확한 결과를 출력할 수 있는 말뭉치를 구축하기 위해 정제된 결과에서 SVM을 이용하여 데이터를 분리한 뒤, 발화 문장과 응답 문장의 내용이 대화가 될 수 있다면 'o', 그렇지 않으면 'x'를 직접 부착한다. ox레이블이 부착된 말뭉치를 SVM에서 해석할 수 있도록 자질을 추출한다. 적용단계에서는 추출한 자질을 바탕으로 만들어진 SVM데이터를 학습하여 그 결과를 추출 단계에서 가진 발화/응답 말뭉치와 비교하여 최종적으로 쓸 수 있는 결과물들을 말뭉치로 가진다.

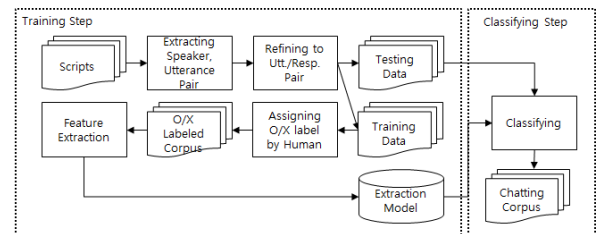


Fig. 3. A Chatting Corpus Building Model

2) 대본 말뭉치 정제

대본 말뭉치는 화자, 그리고 지문과 대사가 혼합되어있는 구조이다. 채팅시스템의 특성상 지문은 필요하지 않기 때문에 화자와 대사만을 추출한다. 이와 같이 분류하기 위해 직

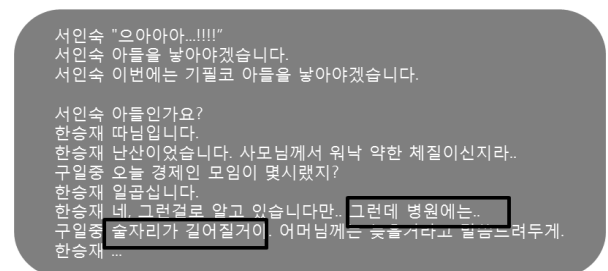


Fig. 4. Example of a Drama Script

접 대본을 읽어서 발화/응답 쌍이 될 수 있는 문장의 특징을 찾아 분류해야 한다. 우선 화자와 대사를 추출한 뒤 대사가 여러 문장인 경우 한 문장으로 통합하는 과정을 거친다. 추출할 때 한 화자의 대사만이 존재하거나 같은 화자가 연속으로 말할 경우 발화/응답 쌍을 만들지 않는다. Fig. 4와 Table 4는 대본 말뭉치를 추출한 뒤, 발화/응답 쌍으로 정제한 예를 보여준다.

Table 4. Example of a Utterance/Response Pair

Utterance	Response
아들인가요?	따님입니다.
사모님께서 워낙 약한 체질이신지라..	오늘 경제인 모임이 몇시됐지?
오늘 경제인 모임이 몇시됐지?	일곱입니다.
그런데 병원에는..	술자리가 길어질거야.

Fig. 4의 마지막 문장과 같이 여러 문장으로 이루어진 대사는 일반적으로 대사 전체가 발화/응답의 쌍으로 이어지기 어렵다는 관점에서 대사 전체를 사용하지 않고 발화 대사의 경우 맨 마지막 문장을, 응답 대사의 경우 맨 앞 문장만을 사용한다.

3) 말뭉치에 레이블 부착

추출된 발화/응답 쌍을 가지는 말뭉치를 사람이 직접 읽고 대화가 성립되면 'o'를, 성립되지 않으면 'x'를 부착한다. 대화가 성립되는 문답은 채팅 코퍼스의 목적에 따라 달라질 수 있다. 그래서 본 논문에서는 어느 영역에 속하지 않는 일상 대화에 해당하는 것을 대상으로 하였다. 이에 따라 일정 또는 예약과 같이 특정 상황에 관련된 대화는 'x'를 부착하였다. 발화/응답 쌍에 공통적인 고유 명사가 존재하는 경우, 개체명 사건을 통해 다양한 활용이 가능하기 때문에 'o'를 부착하였다. Table 5는 ox레이블 부착이 완료된 말뭉치의 예이다.

Table 5. Example of O/X Tagged Corpus

Label	Utterance	Response
o	아들인가요?	따님입니다.
x	사모님께서 워낙 약한 체질이신지라..	오늘 경제인 모임이 몇시됐지?
x	오늘 경제인 모임이 몇시됐지?	일곱입니다.
x	그런데 병원에는..	술자리가 길어질거야.

4) 자질 추출

최종 말뭉치 추출을 위해 SVM에서 쓰이는 발화/응답 쌍에 대한 자질정보를 추출한다. 본 논문에서는 각 문장의 형

Table 6. Features for a Chatting Corpus Building Model

Feature No.	Feature Description
1	The boolean value of coreference words existence between utterance and response
2	The # of coreference words between Utterance and Response
3	The boolean value of a period existence
4	The word with an influential POS label

태소 분석 결과를 이용하여 자질정보를 추출하였다. 형태소 분석 결과를 이용한 자질들의 종류는 Table 6과 같다.

Table 6의 자질은 형태소 분석 결과를 이용한 여러 자질 중 가장 성능 향상이 뚜렷한 4개의 자질을 실험을 통해 선정하였다. 발화/응답 공통 어휘 존재 여부는 발화/응답 쌍에 공통된 주제가 있는지를 판단하는 자질로, 발화와 응답에 공통으로 존재하는 체언이 존재하는지를 판단하여 자질을 생성한다. 발화/응답 공통 어휘 수는 공통으로 존재하는 체언의 개수를 자질로 갖는다. 문장의 마침 기호는 발화가 물음, 응답이 대답이라는 가능성을 판단하는 자질로, 발화의 마침 기호가 '?'이고, 응답의 마침 기호가 '!', '.'인 경우를 판단하여 자질을 생성한다. 마지막으로 영향력 있는 품사에 대한 자질은 발화와 응답에서의 품사가 '동사', '종결어미', '고유명사', '고유명사가 붙은 종결어미' 그리고 응답문장에서의 '접속조사'가 발화/응답의 가능성이 높다고 판단하여 문장의 형태소에서 위와 같은 품사가 생성될 경우 그 형태소를 자질로 추가한다. 영향력 있는 품사에 대한 자질 추가에 대한 예는 Table 7과 같다.

Table 7. Example of Feature No. 4 in Table 6

Utterance	Response
싫어? 어/종결어미	그래 싫다! 그래/접속조사, 다/종결어미
김탁구 씨 맞습니까? 김탁구/NNP, 맞/VV, 습니까/EF	예, 제가 김탁군데요. 예/접속조사, 김탁군데요./고유명사종결어미
우리 기다렸어요? 기다리/동사, 어요/종결어미	기다리는 누가! 기다리/동사, 가/종결어미

4. 실험 및 분석

앞서 설명한 두 가지 제안 방법은 채팅시스템 내에서 사용하는 위치가 다르기 때문에 각각에 대한 실험을 진행하였다.

4.1 RankSVM을 이용한 재순위화 실험 방법 및 실험 결과

RankSVM을 이용한 재순위화에 대한 실험을 위해 총 6명의 인원이 구글 ASR을 이용하여 1,000개의 문장을 발화한 뒤 나오는 후보문장들에 정답 순위를 부착하여 사용하였다. 데이터에 있는 문장 외의 다른 문장들까지 측정하기 위해 5배교차검증법(5-fold cross validation)으로 실험하였다. 그리고 실험의 비교를 위해 구글 ASR순위와 제안연구방법을 정답과 비교하여 성능의 차이를 측정하였다.

본 논문은 음성인식 후보문장의 순위 변화와, 그에 대한 정확성을 측정하므로 구글 ASR과 제안연구의 성능 차이를 비교하였다. 성능 평가 방법으로는 Equation (1)과 Equation (2)의 정확률과 재현율을 사용하였다.

$$\text{정확률} = \frac{\text{시스템출력순위1위가정답인문장수}}{\text{시스템출력문장수}} \quad (1)$$

$$\text{재현율} = \frac{\text{시스템출력순위1위가정답인문장수}}{\text{1위가포함된정답문장수}} \quad (2)$$

정확률의 분모는 전체 문장에 대해 시스템 출력한 결과에서 1위인 후보문장이 포함된 문장이다. 시스템에서 출력한 후보문장은 무조건 1위가 포함되어있기 때문에 전체 문장수인 1,000개로 고정된다. 그리고 재현율은 후보문장에서 정답을 부착한 것 중에서 1위가 없는 경우도 있기 때문에 분모가 유동적이다. 6명의 결과에서 1위인 후보문장이 포함된 문장 수는 평균 705개이다. 분자는 시스템을 통해 출력된 1위 문장이 정답과 일치하는지를 기준으로 개수를 파악한다.

Table 8은 RankSVM을 이용한 재순위화 연구에 사용한 실험데이터를 이용해 음성채팅시스템과 제안 연구의 정확률과 재현율을 측정한 결과이다.

Table 8. Performance of Re-ranking Model for Google ASR Outputs

Data	Google ASR		After Re-ranking Model	
	Precision	Recall	Precision	Recall
Human1	48.7	64.7	54.5	72.6
Human2	40.8	63.4	46.1	71.7
Human3	43.7	66.5	46.9	71.5
Human4	46.6	66.9	49.9	71.7
Human5	44.0	60.4	49.4	67.8
Human6	46.0	60.8	49.3	65.4
Average	45.0	63.8	49.4	70.2

<Table 8>에서 보는 것과 같이 제안 방법은 구글 ASR보다 평균적으로 4.4% 높은 정확률을 보였고, 6.4% 높은 재현율을 보였다. 또한 모든 실험참가자의 데이터에서 성능 향상이 있는 것을 확인할 수 있었다.

4.2 대본 말뭉치로부터 채팅 말뭉치 구축 실험 방법 및 실험 결과

대본 말뭉치로부터 채팅 말뭉치 구축에 대한 실험은 웹상에 공개되어있는 197개의 대본을 이용했고, 실험에 필요한 데이터를 위해 지문을 제거한 약 13만 개의 문장을 얻었다. 그중 임의로 9,164개의 문장을 선택하고 정제하여 얻은 5,560문장을 ox레이블링을 하여 실험을 진행하였다. ox레이블링 결과 5,560문장 중 280문장이 'o'로 레이블링 되었다. 실험은 SVM을 이용하여 5배교차검증법으로 진행하였다. 성능 평가 방법으로는 Equation (3)과 Equation (4)의 정확률과 재현율을 사용하였다.

$$\text{정확률} = \frac{\text{'o'로레이블링된시스템출력이맞은수}}{\text{'o'로레이블링된시스템출력문장수}} \quad (3)$$

$$\text{재현율} = \frac{\text{'o'로레이블링된시스템출력이맞은수}}{\text{'o'로레이블링된정답문장수}} \quad (4)$$

Table 9는 대본 말뭉치로부터 채팅 말뭉치 구축 연구의 실험 결과이다.

Table 9. Performance of a Chatting Corpus Building Model

Data	Precision	Recall
Data1	94.7	64.2
Data2	100.0	69.6
Data3	100.0	58.9
Data4	100.0	35.7
Data5	100.0	57.1
Average	98.9	57.1

실험 결과 'o'로 레이블링 된 문장과 시스템 출력 결과가 높은 정확률로 일치하는 것을 확인할 수 있다. 이는 드라마, 연구 대본을 정제하여 '발화/응답' 쌍으로 활용하는 것이 채팅시스템의 성능을 향상시키는 데 큰 도움이 됨을 증명한다. 특히 응답문장에서 '접속조사'가 포함된 문장은 대부분 질문에 대한 대답이 주를 이루므로 대본 말뭉치에서 중요하게 여길만한 자질로 생각된다.

5. 결 론

본 논문에서는 음성채팅시스템의 성능 향상을 위해 발화 결과로 출력된 후보문장에 RankSVM을 이용하여 재순위화한 뒤 구글 ASR의 성능을 향상시키는 방법과 채팅 말뭉치의 다양성을 위해 드라마/연극 대본을 정제하여 채팅 말뭉치를 추가하는 방법을 제안한다. 그 결과, 기존의 시스템보다 제안한 방법에서 성능이 향상됨을 보이고, 높은 정확률을 가진 채팅 말뭉치를 추가할 수 있었다. 본 논문에서 두 가지 방법에 대해 사용한 자질 외에도 많은 자질을 연구 중에 사용했지만, 그중 가장 성능이 높은 자질들을 간추려 결과를 반영하였다. 그리고 채팅 말뭉치를 추가함에 있어 정제된 대본 말뭉치에 대해 일일이 발화/응답 쌍이 되는지 확인할 수 없고, 대본의 특성상 사투리, 은어가 많이 포함되어있기 때문에 이를 선별하는 기계적 작업이 필요할 것이다. 따라서 추후에는 문장 후보들의 구조를 분석하여 성능을 향상시킬 수 있는 새로운 자질들을 찾는 방법과 수많은 대본을 기계적으로 정제할 수 있는 방법을 모색할 예정이다.

References

- [1] iOS 8-Siri [Internet], <http://www.apple.com/kr/ios/siri/> (2014. 11. 26).
- [2] S-Voice [Internet], http://ko.wikipedia.org/wiki/S_보이스 (2014. 11. 26).
- [3] LG OptimusUI [Internet], http://ko.wikipedia.org/wiki/LG_옵티머스_UI (2014. 11. 26).
- [4] Ki-Seung Lee, "Study on the Improvement of Speech Recognizer by Using Time Scale Modification," *The Journal of the Acoustical Society of Korea*, Vol.23 No.6, pp.462-472, 2004.
- [5] Chang-young Lee, "Comparison of Male/Female Speech Features and Improvement of Recognition Performance by Gender-Specific Speech Recognition," *Journal of The Korea Institute of Information and Communication Engineering*, Vol.5, No.6, pp.568-574, 2010.
- [6] Jungho Cho, "A Spectral Compensation Method for Noise Robust Speech Recognition," *Journal of the Institute of Electronics Engineers of Korea*, Vol.49-IE, No.2, pp.9-17, 2012.
- [7] Sook-Nam Choi, Hyun-Yeol Chung, "Noise Robust Speech Recognition Based on Parallel Model Combination Adaptation Using Frequency-Variant," *The Journal of the Acoustical Society of Korea*, Vol.32, No.3, pp.252-261, 2013.
- [8] Tae-woong Choi, Soon-hyob Kim, "Gamma-tone Feature Extraction Acoustic Modeling for Improving Speech Recognition

Performance," *The Korean Institute of Information Technology*, Vol.10, No.11, pp.155-160, 2012.

- [9] Md. Afzal Hossan, Sheeraz Memon, and Mark A Gragory, "A Novel Approach for MFCC Feature Extraction," *ICSPCS*, pp.1-5, 2010.
- [10] DongHee Lim, SeungShik Kang, and DuSeong Chang, "Word Spacing Error Correction for the Postprocessing of Speech Recognition," *Korea Computer Congress*, Vol.33, No.1, pp.25-27, 2006.
- [11] WonMoon Song, EunJu Kim, and MyungWon Kim, "Post-Processing of Speech Recognition Using User Utterance Sequential Pattern," *Korea Computer Congress*, pp.709-711, 2005.
- [12] Thorsten Joachims, Support Vector Machine for Ranking, Cornell University, 2009, [Internet] http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html (2014.11.26).
- [13] Thorsten Joachims, Support Vector Machine(light), Cornell University, 2008, [Internet] <http://svmlight.joachims.org/> (2014. 11. 26).
- [14] Simesi [Internet], <http://developer.simesi.com/> 2002 (2014. 11. 26).
- [15] Jonghwan Kim, Duseong Chang, and Harksoo Kim, "Statistical Generation of Korean Chatting Sentences Using Multiple Feature Information," *Korean Journal of Cognitive Science*, Vol.20, No.4, pp.421-437, 2009.
- [16] Sejong Corpus [Internet], <http://www.sejong.or.kr/> (2014. 11. 26).



안 혁 주

e-mail : zingiskan12@kangwon.ac.kr

2014년 강원대학교 컴퓨터정보통신공학 전공 (학사)

2014년~현 재 강원대학교 컴퓨터정보통신공학 전공 석사과정

관심분야: FAQ 검색 시스템, 의견분류, 채팅 시스템



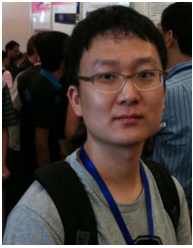
이 성 희

e-mail : tjdgml0820@naver.com

2014년 강원대학교 컴퓨터정보통신공학 전공 (학사)

2014년~현 재 강원대학교 컴퓨터정보통신공학 전공 석사과정

관심분야: 자연어 기계학습, 구문분석



송 영 길

e-mail : nlpyksong@kangwon.ac.kr
2008년 강원대학교 컴퓨터정보통신공학 전공 (학사)
2010년 강원대학교 컴퓨터정보통신공학 전공 (석사)
2010년~현 재 강원대학교 컴퓨터정보통신공학 전공 박사과정

관심분야: 개체명 사전생성, 개체명 인식, 채팅시스템



김 학 수

e-mail : nlpdrkim@kangwon.ac.kr
1996년 건국대학교 전자계산학과(학사)
1998년 서강대학교 컴퓨터학과(석사)
2003년 서강대학교 컴퓨터학과(박사)
2004년~2005년 CIIR in UMass Amherst (박사후연구원)
2005년~2006년 한국전자통신연구원(선임연구원)

2006년~현 재 강원대학교 컴퓨터정보통신공학 전공 교수
관심분야: 자연어처리, 대화시스템, 정보검색, 질의응답시스템, 기계 학습 응용