

예제 기반 챗봇을 위한 기계 학습 기반의 발화 간 유사도 측정 방법

양민철¹, 이연수¹, 임해창^{2*}

¹고려대학교 컴퓨터·전파통신공학과, ²고려대학교 컴퓨터·통신공학부

A Machine Learning based Method for Measuring Inter-utterance Similarity for Example-based Chatbot

Min-Chul Yang¹, Yeon-Su Lee¹ and Hae-Chang Rim^{2*}

¹Department of Computer and Radio Communications Engineering, Korea University

²Division of Computer and Communications Engineering, Korea University

요 약 예제 기반 챗봇은 사용자 발화와 가장 유사한 예제 발화를 대화 예제 데이터베이스로부터 검색하여 응답을 생성한다. 가장 유사한 발화를 찾는 것은 응답의 적절성과 직결되는 것임에도 불구하고, 유사 발화 검색을 위해 어떠한 자질을 사용할 것인지, 어떠한 방식이 좋은 지에 대한 기존 연구는 부족하였다. 본 연구에서는 검색의 정확도와 예제의 활용도를 높이기 위해 다양한 어휘적, 의미적 자질을 이용한 기계 학습 방법을 제안한다. 실험 결과 1) 대화 예제 데이터베이스의 활용도 2) 예제 발화의 매칭의 정확률 3) 답변의 질적인 측면에서 제안하는 방법은 기존의 방법에 비해 더 나은 성능을 보였다.

Abstract Example-based chatBot generates a response to user's utterance by searching the most similar utterance in a collection of dialogue examples. Though finding an appropriate example is very important as it is closely related to a response quality, few studies have reported regarding what features should be considered and how to use the features for similar utterance searching. In this paper, we propose a machine learning framework which uses various linguistic features. Experimental results show that simultaneously using both semantic features and lexical features significantly improves the performance, compared to conventional approaches, in terms of 1) the utilization of example database, 2) precision of example matching, and 3) the quality of responses.

Key Words : Dialogue system, Example-based chatbot

1. 서론

챗봇(chatBot)은 사람과 컴퓨터가 자연어 형태로 다양한 주제의 일상적인 대화를 주고받는 시스템이다. 개발 초기에는 정신병 환자의 치료에 쓰이거나 단순한 재미를 위해 사용되었지만 최근에는 정보 획득, FAQ 시스템 등 목적 지향 시스템과 통합되어 자연스러운 대화 인터페이스를 제공하는 요소 기술로 사용되고 있다. 하지만 채팅은 표-1과 같이 대화의 목적이나 도메인이 미리 정해져

있지 않기 때문에 사용자가 입력한 광범위한 주제의 발화들에 대해 적절히 응답하는 것이 매우 어렵다. 따라서 대부분의 채팅 시스템은 예제 대화 또는 말뭉치에서 광범위한 대화 지식을 습득한 후 입력된 사용자 발화에 대해 가장 유사한 예제 발화를 검색한 후 이에 대한 응답 또는 규칙을 이용하여 시스템 응답을 생성하는 기법을 사용한다. 예제의 형태와 응답 생성 방식이 약간의 차이가 있으나 유명한 ELIZA[1], ALICE[2], WOZ[3]도 일종의 예제기반 방식이라고 할 수 있다.

본 연구는 교육과학기술부 한국연구재단(KRF-2007-361-AL0013) 및 2단계 BK21사업의 지원을 받아 수행되었음.

*교신저자 : 임해창(rim@nlp.korea.ac.kr)

접수일 10년 06월 14일

수정일 10년 06월 24일

게재확정일 10년 08월 10일

【표 1】 목적 지향 시스템과 채팅 시스템의 비교

목적 지향적	채팅
특정 서비스 혹은 정보를 사용자에게 제공하는 것이 목적	미리 정해진 목적이 없음
도메인에 따른 한정된 내용의 발화만을 처리	자유롭고 일상적인 대화로 광범위한 내용의 발화를 처리
ex) TV 편성 안내 대화시스템 오늘 밤 10시 KBS에서 어떤 드라마 하지?	ex) 오늘 본 영화는 너무 슬펐어

예제 기반 챗봇은 더 유사한 발화를 찾을수록 더 적절한 응답이 될 것이라는 가정을 한다. 따라서 어떠한 기준으로 입력된 사용자 발화와 가장 유사한 예제 발화를 찾을 것인가는 매우 중요한 문제이다. 이를 위해서는 다음과 같은 것을 고려해야 한다.

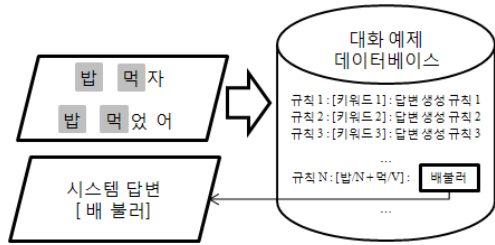
첫째, 대화 시스템에서 발화 간 유사성을 비교할 때는 화자의 의도를 고려해야 한다. 대화에서는 화자와 청자가 있다. 따라서 상대방에 대한 명령이나 요청, 질문, 칭찬, 비난 등 다양한 양태가 나타나고, 동일한 사건이라도 대화가 이루어 지는 시제에 따라 현재, 미래, 과거로 다르게 표현할 수 있다. 따라서 자연스러운 대화가 되기 위해서는 이러한 다양한 상황에 대해 다르게 답변을 해야 한다. 둘째, 예제 기반 대화 시스템에서 대화 예제의 양이 아무리 크더라도 정확히 일치하는 경우는 거의 없기 때문에 최대한 유사한 발화를 찾아 활용할 수 있는 전략이 필요하다. 특히 채팅 시스템의 경우 목적 지향적 시스템과 달리 정확한 응답이 아니더라도 대화를 중단하는 것 보다는 가능하면 이어가는 것이 중요하므로 유사 발화 검색에서도 이러한 점이 고려되어야 한다. 셋째, 발화 간 유사도를 측정하는 기준은 보유하고 있는 예제 데이터베이스의 특성이 반영되어야 한다. 만약 데이터 베이스의 크기가 매우 크고 유사한 발화들이 매우 많이 분포하고 있다면 단순화된 패턴 매칭만으로는 분별력이 없어지게 된다. 반대로 대화 집합의 크기는 매우 작은데 여러 가지 기준으로 완전(exact) 매칭만을 고집한다면 매칭된 발화가 거의 없어 대부분의 대화가 이어질 수 없을 것이다. 따라서 유사 발화를 검색하는 기준에는 사용자의 의도를 반영하는 다양한 의미적 자질과 어휘적 자질이 고려되어야 하고 또한 이 기준을 적용할 때는 유사한 예제가 많을 때는 분별력이 있도록 다양한 자질을 고려하되 예제가 적을 때에는 완전히 일치하지 않더라도 다양한 자질들의 중요도를 고려하여 유사도를 측정할 수 있어야 한다.

본 연구에서는 이를 위해 최근 정보 검색에서 많이 사용하는 기계학습 기법 중 하나인 랭킹 학습을 이용하였다. 본 연구에서 사용한 SVM 기계 학습은 자동으로 자질 간의 관계를 파악 하여 비슷한 경향성을 가진 자질끼리는 가중치를 배분하여 상대적으로 적은 가중치를 할당해 주고, 경향성이 확연히 다르고 유사함의 여부를 분류하는데 있어 명시적인 자질이라면 높은 가중치를 할당해 준다. 여기에 더하여 랭킹 학습은 단순히 유사 한지 여부보다 매우 유사한 발화 쌍의 자질과 덜 유사한 발화 쌍의 자질 간 차이까지 고려하여 가중치를 결정하기 때문에 가장 유사한 발화를 찾는 데 유용한 방법이라고 볼 수 있다. 이런 학습 방법을 예제 기반 챗봇에 적용하면 다음과 같은 이점이 있다. 첫째, 다양한 어휘적, 의미적 자질들을 동시에 고려하고 결합하는데 있어 매우 효과적인 방법이다. 둘째, 발화 간 유사한 정도를 측정함으로써 예제 데이터베이스에 유사한 발화들이 매우 많은 경우 혹은 매우 적은 경우 등 어떠한 상황에서든지 가장 유사한 발화를 선택할 수 있는 분별력을 가지게 된다. 셋째, 예제 데이터베이스의 특성을 자동으로 반영하므로 데이터베이스의 크기와 발화의 분포에 대해 매우 견고한 기준이 된다.

이후 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 기존 연구를 살펴보고 3장에서는 기계학습을 통한 발화 간 유사도 측정 방법에 대해 설명한다. 4장은 제안하는 방법의 검색 정확률과 응답 성능에 대한 실험 및 평가를 보여준다. 마지막으로 5장에서는 결론과 향후 연구에 대해 논의하도록 한다.

2. 관련 연구

예제 기반 챗봇에서의 예제 검색에 관한 기존 연구를 살펴 보면 다음과 같다. ELIZA[1], WOZ[3] 등의 채팅 시스템에서는 키워드 기반으로 답변 생성 규칙을 만들어 놓고 사용자 발화가 입력되면 예제 발화와 키워드 매칭으로 대응하는 답변 생성 규칙들을 찾아 이를 조합하여 응답을 생성한다. 그러나 이러한 방식은 동일한 키워드 열이라도 다른 의미의 발화가 될 수 있다는 점에서 응답의 적절성이 매우 떨어질 수 있다.



[그림 1] 키워드 완전 매칭의 오류

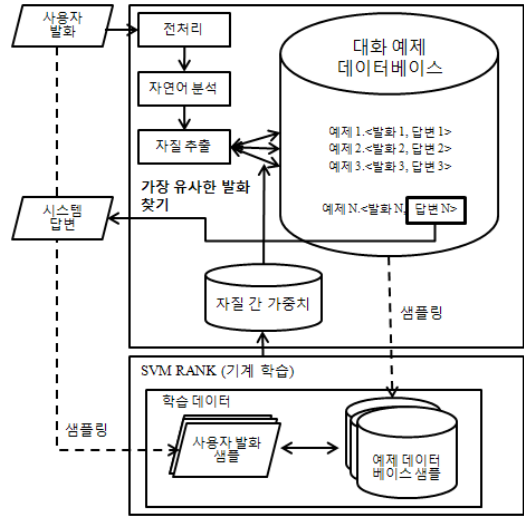
예를 들어 그림 1과 같이 “밥 먹자”와 “밥 먹었어”는 [밥/N 먹/V]이라는 동일한 키워드로 검색되지만 전자는 “청유”의 의미이고 후자는 단순한 “서술”의 의미로서 이에 대한 시스템 응답은 반드시 달라져야 한다. 그러나 단순한 키워드 매칭만으로는 이러한 차이를 구분할 수가 없다.

이러한 점을 보완하여 최근의 목적 지향적 대화 시스템에서는 어휘적 유사성 뿐 아니라 대화 히스토리 (discuss history), 화행(dialog act) 등 다양한 의미적 요소를 고려한다[4]. [4]에서는 자질 간 상관관계(correlation)를 통해 비교할 자질들을 줄여나가면서 두 발화의 자질에 대해 완전 매칭으로 유사한 발화를 찾는다. 하지만 매칭된 자질이 하나도 없을 때 어떤 대안으로 답변을 내줄지, 매칭된 자질 집합이 같은 발화가 여러 개 있을 때 어떤 방법으로 그 중 하나를 선택하여 답변을 내줄지에 대한 기준이 없다.

예제 기반 챗봇은 대량의 예제를 가정하지만 이를 확보하는 것은 쉽지 않다. 일부 주제에 대해서는 많을 수 있고 일부 주제는 예제가 부족할 수도 있다. 따라서 많은 예제가 있을 때는 정확하게 가장 유사한 것을 선택할 수 있어야 하고 예제가 부족할 때에도 대화를 자연스럽게 이어갈 수 있어야 한다. 그러나 기존 연구에서는 이러한 것을 해결하기에 매칭 기준이나 방법적인 측면에서 부족함이 있었다.

3. 유사 발화 검색 방법

본 장에서는 예제 기반 채팅 시스템에서 사용자 발화에 대해 가장 유사한 발화를 검색하기 위한 방법에 대해 논한다. 유사도를 학습하고 이를 이용하는 예제 기반 챗봇의 시스템 구조는 그림 2와 같다.



[그림 2] 시스템 구조

먼저 대화 예제 데이터베이스의 각 예제는 전처리와 형태소 분석, 키워드 분석, 시제, 문형, 긍부정, 양태, 술어, 논항 등 의미 분석이 미리 수행되어 저장된다. 사용자가 발화를 입력하면 전처리 과정을 거쳐 교정되고, 자연어 분석 모듈은 형태소 분석 및 구문 분석을 수행하고 시제, 긍부정, 양태 등 발화의 의미를 분석한다. 그 다음 <사용자 입력 발화, 예제 발화> 쌍에 대해 자질 추출기가 기계 학습에서 사용될 자질 값을 계산한다. 학습된 자질 간 가중치를 기반으로 전체 데이터베이스의 예제에 대해 유사도를 계산하고 그 중 가장 높은 발화에 대응하는 응답을 시스템 응답으로 출력한다.

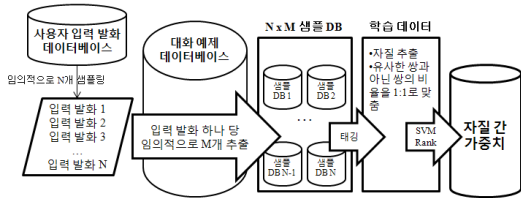
3.1 발화의 의미 분석

[표 2] 자연어 분석

발화 정보	설명
형태소/품사열	발화의 형태소/품사열
긍/부정	ex) 긍정문, 부정문
시제	ex) 과거, 현재, 미래
문형	ex) 평서문, 의문문, 명령문, 청유문
양태	화자의 주관적인 태도 ex) 지각, 추측, 희망, 금지 등
술어	발화에서 주어의 움직임, 상태, 성질 따위를 서술하는 말
논항	술어가 나타내는 의미를 완성시키기 위해 필요한 말 ex) 행동의 주체, 대상, 목적, 도구 등
키워드	품사가 체언, 용언, 수식어에 해당하는 단어

본 연구에서는 정확한 분석을 위하여 사용자 발화에 대해 철자 오류, 띄어쓰기 오류 등의 전처리 단계를 가진다. 그 다음 발화의 어휘적, 의미적 분석을 위해 표 2와 같은 자연어 분석을 수행한다.

3.2 발화 간 유사도 학습



[그림 3] 가중치 학습 과정

본 연구에서 자질 간 가중치 학습을 위한 절차는 그림 3과 같다. 먼저 사용자가 실제 시스템에 입력한 발화 집합에서 임의적으로 N개를 샘플링(sampling)한다. 그 다음 각 발화별로 전체 예제 데이터베이스에서 임의적으로 M개의 예제를 추출하여 작은 샘플 데이터베이스를 만든다. 이러한 N개의 <입력 발화-데이터베이스> 쌍에 대해 사람이 직접 학습 데이터를 구축한다. 평가자는 하나의 발화에 대해 샘플 데이터베이스상의 모든 발화를 대상으로 유사도를 평가한다. 그러면 총 N*M개의 입력발화-예제 발화 유사도 평가 집합이 생성될 수 있지만 생성된 학습 데이터에는 부정적인 예(negative example)가 너무 많으므로 기계학습에 이용하기에는 적당하지 않다. 따라서 비율이 1:1 이 되는 만큼만 부정적인 예를 사용하도록 하였다. 이렇게 구축된 학습 데이터에 자질을 분석하여 추가한 후 기계학습을 통해 자질의 가중치를 결정한다.

[표 3] 발화 간 유사도 기준과 예

유사도	설명
2	누가 봐도 동일한 의미인 경우 ex) 우리 이제 친구지 - 나랑 친구 하자
1	넓게 생각하면 같은 뜻으로 해석 가능하거나 같은 반응(답변)이 가능한 경우 ex) 우리 이제 친구지 - 나 너 좋아
0	누가 봐도 동일하지 않거나 반대의 의미인 경우 ex) 우리 이제 친구지 - 나 궁금해

표 3은 학습 데이터를 만들기 위한 발화 간 유사도 기준과 그 기준에 따른 학습 데이터의 예이다. 본 연구에서는 유사한 정도를 0-2까지 3단계로 태깅(tagging) 하였다.

본 연구에서 사용한 기계학습 방법은 랭킹 학습 기법이다. 예제 기반 챗봇의 특성 상 예제 데이터베이스의 모든 발화 중 사용자 입력 발화에 대해 가장 유사한 1개를 찾아 이에 대한 답변을 내주므로 유사한가 아닌가를 분류하기 보다는 유사한 발화들끼리도 어떤 발화가 더 유사한지 판별하는 것이 중요하다. 그래서 표-3과 같이 유사한 발화 쌍들끼리도 2점, 1점이라는 유사한 정도를 높음으로써 발화 간 랭크를 정하고 이를 기준으로 서로 짝지어 랭킹 학습하는 Pairwise Rank 방식을 이용한다.

3.3 자질 집합

본 연구에서는 3.1장에서 분석한 발화의 어휘적, 의미적 정보를 이용하여 기계 학습에 이용되는 자질을 만든다. 자질은 크게 어휘적 자질(Lexical feature)과 의미적 자질(Semantic feature)로 나눌 수 있다. 표 4는 본 연구에서 고려한 26개의 자질 집합이다.

[표 4] 자질 집합

분류	No.	설명	속성
어휘적 자질	F_1	형태소 기반 편집거리	0-1
	F_2	품사 기반 편집거리	0-1
	F_{3-14}	품사 별 키워드 매칭 비율 (용언, 체언 수식언에 속하는 12 개 품사)	0-1
의미적 자질	F_{15}	의미적 거리를 고려한 편집거리	0-1
	F_{16}	술어 매칭	Y/N
	F_{17}	논항 형태소 매칭	매칭 수
	F_{18-22}	논항 카테고리 별 매칭 수 (행위주체, 대상, 영향주체 등)	매칭 수
	F_{23}	공부정 매칭	Y/N
	F_{24}	시제 매칭	Y/N
	F_{25}	문형 매칭	Y/N
	F_{26}	양태 매칭	Y/N

예를 들어 공부정 매칭은 입력 발화와 예제 발화의 공부정이 일치하는지 여부를 말한다. 품사 별 키워드 매칭 비율 방법은 품사 간 매칭된 형태소의 개수에서 전체 형태소 개수만큼 나누어 정규화 함으로써 얼마나 서로 매칭되는지 비교할 수 있다. 편집거리는 정규화된 편집거리(normalized edit distance)[5] 계산 방법을 사용하여 발화 간 어휘적으로 얼마나 유사한지 비교할 수 있다. 마지막으로 의미적 거리를 고려한 편집거리에서 의미적 거리는 품사가 일반 명사(NNG)와 고유 명사(NNP)인 형태소에 대해서만 한글 워드넷[7]을 이용하여 [6]의 방법에 따라

계산한다.

4. 실험 및 평가

3장에서 제시한 각 자질의 유용성과 SVM Rank[8]를 통한 유사도 학습 방법의 효과를 평가하기 위해 테스트 집합에서 실제 9만개의 예제 데이터베이스를 대상으로 유사 발화를 검색하고 시스템 응답을 생성하여 유사 발화의 정확도와 사용자 만족도(답변 적정도)를 평가하였다. 본 장에서는 먼저 학습 데이터 구축 등 실험 환경에 대해 살펴보고 실험 결과를 소개한 후 이에 대한 분석을 하도록 한다.

4.1 실험 환경

실험에 사용된 대화 예제 데이터베이스는 실제 모바일 채팅 시스템을 통해 사용자가 각 입력 발화에 대한 응답 부분을 입력한 것으로 사람이 직접 구축한 것이다. 이런 이유로 각 예제 발화와 이에 대한 답변은 서로 적절한 질문과 응답쌍이라고 가정한다. 전체 데이터 크기는 동일한 예제를 제외한 91,128개로 구성되어 있다.

가중치 학습을 위한 학습데이터는 입력 발화와 샘플 데이터베이스 쌍으로 구성되는데 입력 발화는 예제 데이터베이스에 사용되지 않은 발화 10,000개 중 무작위로 160개를 추출하였다. 하나의 입력 발화 당 3,000개로 구성된 샘플 데이터베이스는 실제 9만개의 전체 예제 데이터베이스에서 임의적으로 샘플링 하여 구축하였다. 3.2장에서 설명한 방식을 통해 최종적으로 총 4,851개의 발화 쌍이 학습데이터로 추출되었다.

평가 데이터는 예제 데이터베이스에 사용되지 않은 발화 10,000개 중 무작위로 추출한 150개 테스트 발화에 대해 학습한 가중치를 기반으로 검색한 유사 발화와 답변을 앞서 3.2장의 유사도 기준과 [9]의 사용자 만족도 기준에 따라 3명의 실제 사용자가 평가하여 평균값을 측정하였다.

[표 5] 검색 방법 비교

방법	검색 방법	사용한 자질
Key by Exact	완전 매칭	F ₃₋₁₄
Key + SEM by Exact	완전 매칭	F ₃₋₁₄ + F ₂₃₋₂₆
Word class by Exact	완전 매칭	F ₃₋₁₅
NED by SVM	랭킹 학습	F ₁₋₂
SEM by SVM	랭킹 학습	F ₃₋₂₆
All by SVM (제안하는 방법)	랭킹 학습	F ₁₋₂₆

실험을 통해 제안하는 방법과 비교하고자 하는 유사 발화 검색 방법은 표 5와 같다.

4.2 실험 결과 및 분석

각 검색 방법이 얼마나 유사한 발화를 찾을 수 있는가를 비교하기 위해 다음 식(1), (2)와 같이 정확률과 유사도 점수를 평가한다.

$$\text{정확률} = \frac{\text{sim}(u_i, e(u_i)) \geq 1 \text{인 경우의 수}}{\text{전체 입력 발화 개수}} \tag{1}$$

$$\text{유사도 점수} = \sum \text{sim}(u_i, e(u_i)) \tag{2}$$

- $\text{sim}(x, y)$ = x 발화와 y 발화 간 유사도
- u_i = i번째 입력 발화
- $e(x)$ = x 발화에 대해 시스템이 내준 가장 유사한 발화

정확률은 테스트 발화에 대해 9만개의 예제 데이터베이스에 유사한 발화가 있다고 가정하며 절대적인 수치보다는 각 방법을 비교하는데 의미가 있다. 유사도 점수는 전체 테스트 쿼리에 대해서 서로 다른 방법들이 얼마나 더 유사한 발화를 찾는가를 평가하기 위한 것으로서 이것 역시 절대적인 수치보다는 각 방법을 비교하는데 의미가 있다.

[표 6] 유사 발화도 성능

방법	정확률(%)	유사도 점수
Key by Exact	29.53	72
Key + SEM by Exact	28.19	62
Word class by Exact	44.97	92
NED by SVM	59.73	126
SEM by SVM	64.43	138
All by SVM	76.51	168

유사 발화 검색의 기존 연구와 다양한 자질을 사용하는 방법 및 SVM을 통해 학습을 한 경우에 대한 비교 실험 결과는 표 6과 같다. NED by SVM, SEM by SVM, All by SVM을 비교해 볼 때 어휘적 자질 또는 의미적 자질만을 각각 사용하는 것 보다 의미적, 어휘적 자질을 동시에 고려하는 것이 성능이 좋았다. 어휘적 자질의 가중치가 높게 나타났지만 어휘적 자질로 구별할 수 없는 것을 의미적 자질을 이용함으로써 더 잘 구별할 수 있었으며, 의미적 자질이 동일할 때에는 형태 적 자질이 가장 유사한 것을 찾는데 결정적인 역할을 했다. 또한 Key by Exact, Key + SEM by Exact, Word class by Exact, All

by SVM 방법들을 비교해 보면 기존의 완전 매칭이 위험할 수 있음을 알 수 있다. 기존 방법은 한정된 자질에 대해서 모두 매칭이 되거나 하나도 매칭이 안 되는 경우에는 대안을 찾을 수 없었다. 하지만 랭킹 학습을 통해 학습한 경우는 유사도를 계산함으로써 발화 간 랭킹을 매길 수 있고 이를 통해 어떤 상황에서든 가장 유사하다고 판단할 수 있는 발화를 찾을 수 있었다.

[표 7] 사용자 만족도 기준

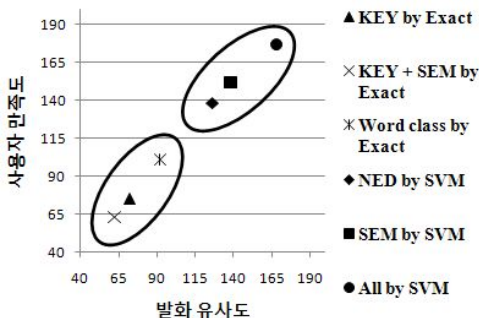
만족도	설명
2	상황에 맞는 답변 (reasonable reply)
1	상황에 어울리지 않지만 이해 가능한 답변 (weird but understandable reply)
0	상황에 어울리지도 않고 무의미한 답변 (non-sensical reply)

다음은 응답의 질에 관한 평가로서 평가 기준은 [9]에서 사용한 다음의 표 7과 같은 기준을 적용하였다.

[표 8] 사용자 만족도 성능

방법	0	1	2	총 점수
Key by Exact	0.67	0.14	0.18	75
Key + SEM by Exact	0.71	0.15	0.13	63
Word class by Exact	0.50	0.33	0.17	101
NED by SVM	0.36	0.36	0.28	138
SEM by SVM	0.27	0.44	0.29	152
All by SVM	0.21	0.41	0.39	177

답변에 대한 사용자 만족도 분포 결과는 표 8과 같다. 모든 자질을 동시에 고려하고 학습을 하는 방법이 응답의 전체 점수 뿐 아니라 다른 방법에 비해 2점의 분포가 높은 것을 알 수 있다.



[그림 3] 발화 유사도와 사용자 만족도 간 관계

발화 유사도와 사용자 만족도의 종합적인 분포와 이들

간 직접적인 연관 관계는 그림 3을 통해 볼 수 있다. 완전 매칭은 유사 발화를 찾지 못하는 경우가 많아 전체적으로 랭킹 학습 방법에 비해 응답 성능도 낮다. 또한 학습 방법을 사용하면서도 의미적, 어휘적 자질을 동시에 고려한 제안하는 방법이 발화 유사도, 답변 적정도 모두 가장 좋은 성능을 보였다. 마지막으로 발화 유사도가 높을수록 사용자 만족도가 높아짐을 볼 수 있는데 앞서 말한 예제 기반 챗봇의 가정이 참임을 알 수 있다.

[표 9] 무의미한 응답 반복 비율

방법	repetition(%)
Key by Exact	52.35
Key + SEM by Exact	59.73
Word class by Exact	30.20
All by SVM	0.00

응답에 대한 또 다른 평가로서 표 9와 같이 기본적인 무의미한 응답 반복(repetition)[9]의 비율을 평가하였다. Word class by Exact는 단어 클래스를 이용한 일반화를 통해 매칭 비율이 향상되어 무의미한 반복이 줄어들긴 하였으나, 대부분 기존 연구에서 사용하는 완전 매칭은 채팅 시스템에서 유사 발화가 반 이상 검색이 안되어서 무의미한 응답을 반복하였다. 제안하는 방법은 이완 매칭(relaxed matching)을 이용해 한정된 양의 데이터베이스 내에서 예제를 최대한 활용하여 모든 사용자 발화에 대해 가능한 답변을 내줌으로써 기존 매칭 방법의 문제점을 해결함을 볼 수 있었다.

5. 결론 및 향후 연구

본 논문에서는 예제 기반 챗봇에서 가장 중요한 기술 중 하나인 유사 발화 검색을 위한 새로운 방법을 제안하였고 제안하는 방법은 발화 간 유사성을 파악하는데 유용한 자질들을 추출하고 기계학습을 이용하여 선별한 자질을 활용한 방법이다.

실험을 통해 단순히 어휘적인 자질 만이 아니라 다양한 의미적 자질을 동시에 고려하는 것이 유사 발화 검색의 정확도를 향상시킬 수 있었고, 완전 매칭보다는 기계 학습 방법을 이용하여 각 자질에 최적의 가중치를 적용하고 자연스럽게 이완 매칭이 되도록 하는 방식이 정확도 뿐 아니라 예제의 활용도 측면에서 성능 향상이 크다는 것을 알 수 있었다. 또한 이러한 유사 발화 검색의 성능 향상이 응답 성능 향상에 매우 직접적으로 연관된다는 것을 응답 평가를 통해 알 수 있었다.

향후에는 사용자 발화와 예제 발화간의 관계 뿐 아니라 답변의 정보까지 고려할 것이다. 또한 기계 학습을 통한 유사도의 임계값을 설정하는 문제 및 검색 속도의 향상을 위한 자동화 방법 등을 연구하여 좀 더 완벽한 챗봇을 개발할 것이다.

참고문헌

- [1] Weizenbaum, J. "Eliza - a computer program for the study of natural language communication between man and machine", Communications of the ACM, Vol. 9, pp. 36-45, 1965.
- [2] Levin et al., "The ALICE System: A Workbench for Learning and Using Language", CALICO Journal Vol. 9, No. 1, pp. 27-56, 1991.
- [3] H. Murao, N. Kawaguchi, S. Matsubara, Y. Yamaguchi, and Y. Inagaki. "Example-based spoken dialogue system using WOZ system log", SIGdial Workshop on Discourse and Dialogue (SIGDIAL 2003), pp. 140-148, 2003.
- [4] C Lee et al. "Correlation-based Query Relaxation for Example-based Dialog Modeling", ASRU, pp. 474-478, 2009.
- [5] Andrés Marzal and Enrique Vidal. "Computation of normalized edit distance and applications", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15. pp. 926-932. 1993.
- [6] Jay J. Jiang and David W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy", In Proceedings of International Conference on Research in Computational Linguistics, Taiwan. 1997.
- [7] 윤애선 외, "한국어 어휘의미망 KorLex 1.5의 구축", 한국정보과학회 논문지, 제36권, 제1호, pp. 92-108. 2009.
- [8] T. Joachims. "Training Linear SVMs in Linear Time", Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD). 2006.
- [9] B.A. Shawar, and E. Atwell. "Different Measurement metrics to evaluate a chatbot system" Academic and Industrial Research in Dialog Technologies Workshop Proceedings. Association for Computational Linguistics. pp. 89-96. 2007.

양 민 철(Min-Chul Yang)

[학생회원]



- 2010년 2월 : 고려대학교 컴퓨터·통신공학부 (공학사)
- 2010년 3월 ~ 현재 : 고려대학교 대학원 컴퓨터·전파통신공학과 석사과정

<관심분야>
대화 시스템, 정보검색

이 연 수(Yeon-Su Lee)

[정회원]



- 2008년 8월 : 고려대학교 대학원 컴퓨터·전파통신공학과 (컴퓨터학석사)
- 2008년 9월 ~ 현재 : 고려대학교 대학원 컴퓨터·전파통신공학과 박사과정

<관심분야>
대화 시스템, 기계번역

임 해 창(Hae-Chang Rim)

[정회원]



- 1983년 12월 : University of Missouri-Columbia 대학원 전산학과 (전산학석사)
- 1990년 12월 : University of Texas-Austin 대학원 전산학과 (전산학박사)
- 1991년 9월 ~ 현재 : 고려대학교 컴퓨터학과 (현 컴퓨터·통신공학부) 교수

<관심분야>
자연어처리, 정보검색