

카페의 계산원과 고객의 대화 모델링

한철호^o, 장병탁

서울대학교 컴퓨터공학부

{chhan, btzhang}@bi.snu.ac.kr

Modeling Conversation between Cashier and Customer in Cafes

Cheolho Han^o, Byoung-Tak Zhang

Department of Computer Science and Engineering, Seoul National University

요 약

이 논문은 카페의 계산원과 고객의 대화에 대한 몇 가지 모델을 소개한다. 대화 모델링은 음성을 통한 로봇과의 대화에 필수적인 요소로 활발히 연구되고 있다. 그러나 현재 개발된 대화 시스템은 아직 상용화 수준에 이르기에는 미흡해 보인다. 이 논문 또한 마찬가지로 카페에서의 상황을 가정하여 좀 더 우수한 대화 시스템을 개발하기 위해 사용한 여러 방법과 주요 결과를 소개한다. 실험을 위해 먼저 계산원이 한 명의 고객과 번갈아 대화하는 내용의 음성 데이터를 수집하였다. 수집한 데이터를 직접 글자로 옮겨 적었으며, 이렇게 글자로 이루어진 데이터를 바탕으로 실험하였다. 글자 데이터는 수공예(hand-crafted) 특징(feature), 단어 가방(bag-of-words), 문서의 벡터화(doc2vec) 등을 이용해 전처리하였다. 전처리된 특징을 바탕으로 최근린(nearest neighbor, NN), 은닉 마코프 모델(hidden Markov model, HMM), 지지 벡터 기계(support vector machine, SVM), 하이퍼네트워크(hypernetwork, HN)를 통해 대화 의도 및 응답을 모델링하였다. 특히 은닉 마코프 모델을 통해 미리 정의된 26가지 대화 의도의 90%를 올바르게 분류할 수 있었다. 그리고 응답의 자연스러움을 9명의 사람이 평가했을 때, 7점 만점에 실제 응답 6.52, 생성 응답 4.47, 비교 모델 2.73을 얻어 꽤 자연스러운 응답을 할 수 있었다. 또한 응답을 분류 문제로 취급할 경우에 대해, 하이퍼네트워크를 통해 올바른 응답을 할 확률이 50%를 넘어섰다.

1. 서 론

대화 모델링은 음성을 통한 로봇과의 대화에 필수적인 요소로 활발히 연구되고 있다 [1-8]. 대화와 같은 시계열 데이터를 모델링하기 위해, 전통적으로 은닉 마코프 모델(hidden Markov model, HMM)이 자주 사용되었다 [2, 3]. 근래에는 하이퍼네트워크(hypernetwork, HN)[1]를 통한 시도나, 순환 신경망(recurrent neural network, RNN)[5-8]을 활용한 연구가 활발히 이루어졌다. 그러나 현재 개발된 대화 시스템은 아직 상용화 수준에 이르기에는 미흡해 보인다.

이 논문은 특별히 카페에서의 상황을 가정하여, 계산원과 고객의 대화에 대해 좀 더 우수한 대화 시스템을 개발하기 위해 시도한 여러 방법을 소개한다. 개략적인 대화 모델링 과정은 다음과 같다. 먼저 계산원이 한 명의 고객과 번갈아 대화하는 내용으로 음성 데이터를 수집하였다. 수집한 데이터를 직접 글자로 옮겨 적었으며, 이렇게 글자로 이루어진 데이터를 바탕으로 실험하였다. 글자 데이터는 몇 가지 방법을 통해 전처리하였고, 전처리된 특징을 바탕으로 대화 의도 및 응답을 모델링하였다. 전처리 기법으로는 수공예 특징(hand-crafted feature), 단어 가방(bag-of-words), 문서의 벡터화(doc2vec) 등이 사용되었고, 대화 의도 및 응답 모델링을 위해 최근린(nearest neighbor, NN), 은닉 마코프 모델, 지지 벡터 기계(support vector machine, SVM), 하이퍼네트워크 등의 모델이 대화 모델링에 사용되었다.

2장에서는 실험에 사용된 데이터 및 알고리즘, 3장에서는 수행한 실험 및 그 결과, 4장에서는 이번 연구를 통해 얻어진 결론 및 향후 연구를 소개한다.

2. 데이터 및 알고리즘

2.1 데이터

먼저 계산원이 한 명의 고객과 번갈아 대화하는 내용으로 130회의 대화, 총 706개의 문장으로 이루어진 20시간가량의 음성 데이터를 수집하였다.

2.2 전처리

2.2.1 음성-글자 변환

녹음된 음성 데이터를 직접 글자로 옮겨 적었으며, 이렇게 글자로 이루어진 데이터를 바탕으로 실험하였다. 한국어는 굴절(inflexion)과 같은 어휘의 변화가 많기 때문에, 이러한 변화에 다른 차이를 없애기 위해 각 단어를 형태소(morpheme)로 변환하였다. 이때 서로 다른 단어의 수는 461개였으며, 100개의 단어를 갖는 사전을 통해 충분히 높은 의도 분류 정확도를 얻을 수 있었다 [1].

2.2.2 글자-특징 변환

글자 데이터는 문장 별로 몇 가지 수공예(hand-crafted) 특징(feature), 단어 가방(bag-of-words), 문서의 벡터화(doc2vec) 등을 이용해 전처리하여 대화 모델링 학습을 위한 특징으로 표현하였다.

2.2.2.1 수공예 특징

수공예 특징으로서 발화 특징(utterance feature, UF)[2, 3]을 다양한 방법으로 정의하였다. 표 1은 발화 특징을 분류한 4 가지 기준을 나타낸다. 표 2는 기준3에 따라 13개의 발화 특징으로 분류된 예시 문장을 나타낸다.

표 1 여러 문장을 하나의 발화 특질로 묶는 기준

조건 기준	분류 조건
기준 1	메뉴, 가격, 기타
기준 2	메뉴(완성, 비완성), 가격, 기타
기준 3	기준 2 + 선택
기준 3	기준 3 + 행동

표 2 기준 3에 따른 각 발화 특질에 속하는 예시 문장

발화 특질	예시 문장
손님-메뉴 완성	두유 라떼 아이스로 하나 주세요.
손님-메뉴 비완성	생망고 생딸기 요구르트하구요.
손님-선택	아니요, 투샷으로 주세요.
손님-가격	5800원이요?
손님-행동	계산 또는 서명
손님-기타	노트북이랑 아이패드는 여기 있는 거 쓰면 돼요?
점원-메뉴 완성	두유 라떼 아이스
점원-메뉴 비완성	네 매실 토닉워터랑요.
점원-선택	카페모카 아이스요?
점원-가격	15000원입니다.
점원-행동	“손님-행동”에 대한 쌍
점원-기타	짜인해주시구요.
NULL	‘대화 시작/종료’를 나타내는 state의 evidence를 표현

2.2.2.2 단어 가방

단어 가방 모델 [1]은 음성 문장의 특질을 추출하기 위해 사용된다. 단어 가방 모델에서 사전은 언어 자료 혹은 말뭉치 (corpus)에 있는 서로 다른 단어들로부터 구성된다. 사전에 등록된 단어의 수는 전체 단어 461 개 중 100개 정도로 줄이더라도 충분한 성능을 얻을 수 있었다. 단어 가방 모델에서는 사전에 등록된 단어가 각 문장에 몇 번 나타나는지를 기록한 벡터를 만든다.

2.2.2.3 문서의 벡터화

문서의 벡터화 [4]는 문서를 벡터로 변환하는 방법이다. 파이썬(Python) 언어에서는 genism (genism.models.doc2vec)을 통해 수행할 수 있다. 문서의 벡터화에서도 단어를 공백으로 구분하고, 한국어의 굴절과 같은 변형에 따른 구분을 없애기 위해 형태소로 변환한 다음 문서의 벡터화를 적용하였다.

2.3 모델링

전처리된 특질을 바탕으로 최근린(nearest neighbor, NN), 은닉 마코프 모델(hidden Markov model, HMM), 지지 벡

터 기계(support vector machine, SVM), 하이퍼네트워크(hypernetwork, HN)를 활용하여 대화 의도 및 응답을 모델링하였다.

2.3.1 최근린

최근린 (Nearest neighbor, NN) [4]은 기본적인 분류기 중 하나이다. 최근린은 화자의 의도를 분류하기 위해 사용될 수 있다. k-최근린은 이전에 관찰된 문장 중 입력 문장에 가장 가까운 k 개의 문장의 다수결을 통해 문장의 의도를 분류한다. 비교되기 위한 문장은 저장되어야 하고, 매 입력마다 비교되어야하므로, 대용량의 기억과 높은 계산 복잡도를 필요로 한다.

2.3.2 은닉 마코프 모델

은닉 마코프 모델(Hidden Markov model, HMM) [2, 3]은 순차적 데이터를 모델링하기 위해 사용되며, 데이터는 일부만 관찰이 가능하다. 은닉 마코프 모델을 의도 분류와 응답 생성을 위해 사용될 수 있으며 (그림 1), 학습 및 추론 알고리즘이 잘 알려져 있다.

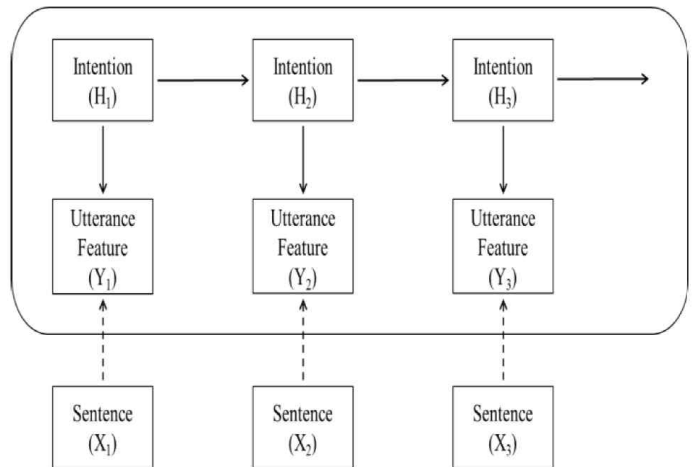


그림 1 은닉 마코프 모델을 통한 대화 모델링, 의도 H는 은닉 변수, 발화 특질 H는 관찰 변수이다. 각 문장 X는 수공예 특질을 통해 발화 특질로 변환된다.

2.3.3 지지 벡터 기계

지지 벡터 기계 (Support vector machine, SVM) [1]는 2000년대 초반에 활발히 쓰인 강력한 분류기이다. 가장 간단한 형태는 선형 지지 벡터 기계인데, 선형 분리 가능한 두 집단을 가장 큰 여유(margin)를 두고 분류할 수 있다. 커널 방법(kernel trick)을 통해 광범위한 데이터를 분류할 수도 있다.

2.3.4 하이퍼네트워크

하이퍼네트워크(Hypernetwork, HN) [1]는 일반화된 그래프이다. 그래프는 두 개의 점을 잇는 선을 갖는 반면, 하

이퍼네트워크는 여러 점을 잇는 하이퍼에지(hyperedge)를 갖는다. 하이퍼네트워크는 의도 분류 및 응답 생성에 활용될 수 있다.

3. 실험 및 결과

은닉 마코프 모델을 통해 미리 정의된 26가지 대화 의도의 90%를 올바르게 분류할 수 있었다 (그림 2). 또한 110회의 대화를 훈련 데이터(training data), 20회의 대화를 실험 데이터(test data)로 사용했을 때, 하이퍼네트워크를 통해 각 대화의 시작부터 끝까지 모두 올바르게 응답할 확률이 50%를 넘었다 (그림 3). 은닉 마코프 모델의 경우 응답의 자연스러움을 9명의 사람이 평가했을 때, 7점 만점에 실제 응답 6.52, 생성 응답 4.47, 비교 모델 2.73을 얻어 꽤 자연스러운 응답을 할 수 있었다 (그림 4).

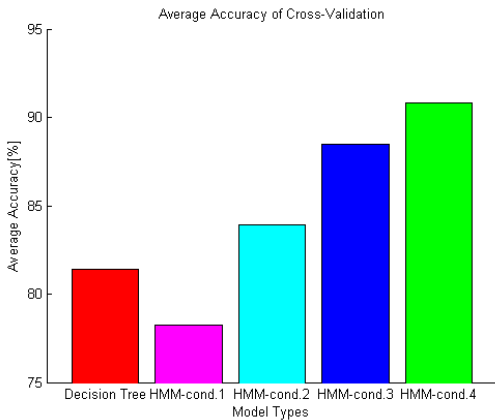


그림 2 의도 분류 정확도 비교: 의사 결정 나무 vs. 다양한 수공예 특질을 활용한 은닉 마코프 모델 [3]

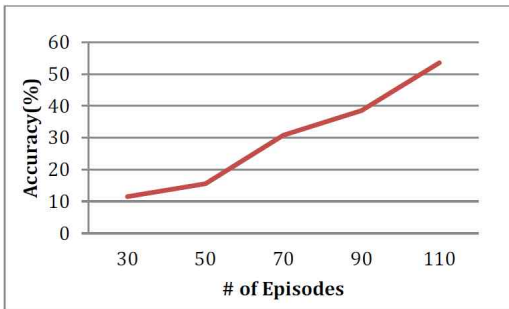


그림 3 응답 생성 정확도 vs. 학습 대화 수 [1].

4. 결론 및 향후 연구

대화 의도 및 응답 모델링을 통해 대화 시스템을 모델링하였다. 대화 모델링을 통해 카페에서 고객을 응대하는 계산원을 보조/대체하는 로봇을 개발한다면, 카페의 점주의 입장에서는 비용을 줄이고, 고객의 불평, 불만을 줄일 수 있을 것이다. 또한 고객 입장에서는 방대한 지식을 가진 로봇의 추천 등을 통해 편리함을 얻을 수 있을 것으로 기대된다. 추후 소음이 심하거나 화자가 여러 명인 상황에서의 대화를 모델링하는 기술이 필요할 것으로 보인다.

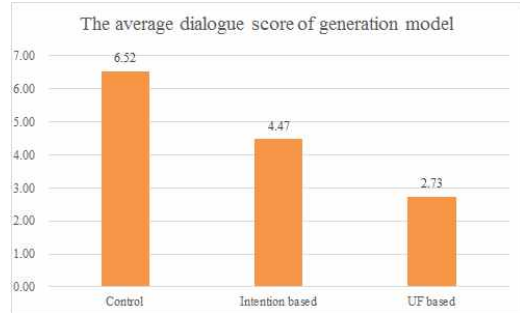


그림 4 9명의 사람들이 평가한 대화의 평균적인 자연스러움. Control은 실제 응답이며, Intention-based HMM 생성 시 추론된 의도와 관찰 문장을 이용한다.. UF-based HMM은 의도 대신 발화 특질과 관찰 문장을 이용한다 [3].

감사의글

이 논문은 2016 년도 정부(미래창조과학부, 국방부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW 스타랩), 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF), 국방과학연구소(UD130070ID-BMRR)의 지원을 받았음.

참고문헌

- [1] J.-H. Oh, H.-S. Chun, and B.-T. Zhang, "Generating cafeteria conversations with a hypernetwork dialogue model," in *Proc. the 14th International Symposium on Advanced Intelligent Systems (ISIS 2013)*, pp. 1424-1435, 2013.
- [2] 이승원, 김은솔, 장병탁, "Hidden Markov Model을 이용한 대화 의도 모델링" *한국정보과학회 동계학술발표회 논문집*, pp. 1417-1419, Nov. 2014.
- [3] S. Lee, J. Hwang, E.-S. Kim, and B.-T. Zhang, "An adaptive computational discourse system based on data-driven learning algorithm," *The 16th International Symposium on Advanced Intelligent Systems (ISIS 2015)*, 2015.
- [4] 최성호, 김은솔, 장병탁, "Paragraph Vector를 이용한 문장 의도 예측 기법," *2016 한국컴퓨터종합학술대회(KCC2016) 논문집*, pp. 977-979, Jun. 2016.
- [5] Shang, Lifeng, Zhengdong Lu, and Hang Li. "Neural responding machine for short-text conversation." *arXiv preprint arXiv:1503.02364* (2015).
- [6] Vinyals, Oriol, and Quoc Le. "A neural conversational model." *arXiv preprint arXiv:1506.05869* (2015).
- [7] Serban, Iulian V., et al. "Building end-to-end dialogue systems using generative hierarchical neural network models." *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*. 2016.
- [8] Ren, Yafeng, Yue Zhang, Meishan Zhang, and Donghong Ji. "Context-Sensitive Twitter Sentiment Classification Using Neural Network." *AAAI*, 2016.