

챗봇 프레임워크 성능 향상을 위한 점진적 학습 기법

박상현* (교신저자), 박진욱^o, 조수현*, 현재혁*, 황진성*
^o연세대학교 컴퓨터과학과

e-mail: {sanghyun, parkju536}@yonsei.ac.kr^o, {whtngjs415, jehyeok.hyun, progresivojs}@gmail.com*

Incremental Learning for Performance Enhancement of Chatbot Framework

Sanghyun Park*, Jinuk Park^o, Soohun Joe*, Jehyeok Hyun*, Jinseong Hwang*

^oDept. of Computer Science, Yonsei University

● 요약 ●

규칙 기반의 챗봇(Chatbot)은 개발자가 미리 지정한 키워드와 패턴을 통해 사용자의 의도(Intent)를 파악하기 때문에, 챗봇을 응용한 어플리케이션에서는 제한적인 활용도를 보인다. 본 논문에서는 위 문제를 해결하기 위해, 프레임워크 기반의 한글 자연어 처리 챗봇 성능 향상을 위한 점진 학습(Incremental Learning)을 제안한다. DialogFlow는 규칙 기반의 챗봇 프레임워크로서, 사용자 질의 패턴에 대한 사전 학습이 치명적이다. 제안하는 점진 학습 기법은 사용자 질의가 미리 학습되어 있지 않은 경우에도, 유사도 기반으로 질의의 의도를 결정할 수 있다. 이때 entity 조합과 기존에 학습된 질의들과의 유사도를 통해 의도를 결정하여, 프레임워크를 점진적으로 학습한다. 이를 적용하여 연세대학교 정보들을 제공하는 챗봇을 개발하고, 실험을 통해 제안된 점진 학습 기법은 기존 시스템보다 다양한 종류의 질의 처리가 가능하고, 더욱 빠른 응답 속도를 나타내는 것을 확인하였다. 또한 사용자가 증가함에 따라 점진 학습을 통해 성능이 더욱 증가하는 자가 학습 모형으로서의 우수함을 확인하였다.

키워드: 챗봇(Chatbot), 프레임워크(Framework), 질의의 의도·목적(Intent), 질의의 키워드·개체(Entity), 점진 학습(Incremental Learning)

I. Introduction

자연어 기반의 대화형 AI, 챗봇은 사용자가 직접 질의하여 정보를 찾는 주도적인 정보 습득이 가능한 구조를 지니고 있다. 이러한 주도적인 정보 습득 구조와 함께, 사회 관계망 서비스와 보편화된 스마트폰 등 다양한 커뮤니케이션 채널의 확대에 힘입어 챗봇은 차세대 소셜 플랫폼으로 떠오르고 있다[1].

현재 상용화할 수 있는 챗봇은 대부분 규칙 기반으로 이루어진 시스템으로서, 사전에 정의된 단어와 패턴에 의거하여 응답이 이루어진다. 따라서 미리 정의되고 시스템에 학습된 패턴이 아닌 경우, 사용자 질의에 응답할 수 없는 한계점이 발생한다[2].

우리는 위의 문제점을 해결하기 위해, 챗봇 시스템의 점진적인 학습 스타일을 제안한다. 사용자의 질의가 정의되지 않아 질문의 의도를 파악할 수 없는 경우, 시스템에서 그 의도를 유사도 기반으로 결정하여 시스템의 성능을 향상시킨다. 또한, 해당 질의와 의도를 새롭게 학습하여, 더욱 폭 넓은 질의 처리와 추후 질의들에 대한 빠른 응답 처리를 목표로 한다.

또한, 위의 기법을 적용한 연세대학교 챗봇인 ‘연챗’을 개발하여, 그 성능을 검증하고자 한다. 연세대학교 학교생활과 밀접한 정보들이 여러 웹사이트에 분산되어 있어, 학생들이 정보 획득의 불편함을

호소해왔다. 이에 주도적이고 쉬운 정보 습득을 위해, 구글 DialogFlow를 활용한 학교생활 관련 정보를 얻을 수 있는 챗봇의 개발을 목표로 한다.

본 논문의 주요 기여는 다음과 같다. (1) 규칙 기반의 챗봇 프레임워크가 지니는 근본적인 한계점인 사전 정의된 패턴을 극복하여 점진적 학습을 통해 시스템이 자가 학습하는 기법을 제안한다. (2) 연세대학교 챗봇인 연챗을 개발하고, 이를 통해 제안한 기법을 검증하였다.

II. The Proposed Scheme

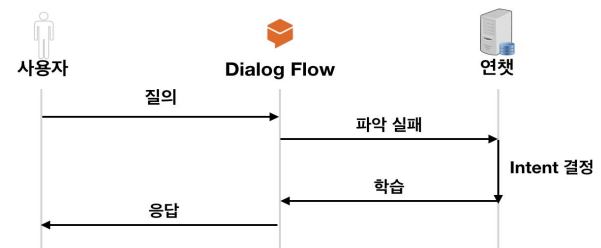


Fig. 1. Overview of the proposed learning scheme
DialogFlow에 미리 학습되어 있지 않은 사용자 질의를 처리 과정을

Fig. 1에 제시한다. DialogFlow에서 파악하지 못한 intent는, 먼저 entity를 추출하여 해당 조합과 가장 유사한 intent로 결정된다.

예를 들어, “도서관 2층에서 일요일에 책 볼 수 있어?” 라는 질의의 intent를 DialogFlow가 파악하지 못하면 ‘도서관’, ‘2층’, ‘일요일’ 이라는 entity를 추출한 뒤, 위의 조합을 포함하고 있는 ‘도서관 운영 시간’ intent로 매핑하여 DB로부터 도서관 2층의 일요일 운영시간 정보를 사용자에게 응답한다.

추가적으로, DialogFlow가 intent를 파악하지 못한 질의를 학습시킴으로써 이후 유사한 질의에 대해서는 자체적으로 응답할 수 있도록 한다. 이러한 프로세스를 통하여 사용자의 질의를 예상하여 미리 학습시키는 과정이 불필요하며 질의에 포함된 entity만을 중심으로 응답을 생성할 수 있다. 또한, 학습을 통해 사용자들이 증가할수록 성능 개선을 기대할 수 있다.

III. Conclusions

제안된 연챗은 2018년 11월 16일부터 11월 28일까지 연세대학교 학생들의 사용을 통해 총 2,433개의 질의를 입력받았다. Table 1은 그 중 처리한 질의 개수를 연챗과 기본 DialogFlow를 비교하여 나타낸다.

DialogFlow만을 이용할 때보다 391개의 질의를 더 응답하여 32.3%의 성능 개선을 보였다.

Table 1. Response rate for queries

Yonchat	DialogFlow	Improvements
65.9% (1604)	49.9% (1213)	32.3% (391)

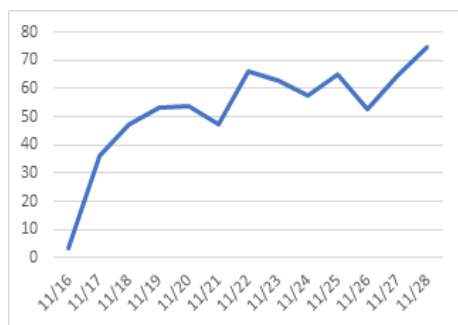


Fig. 2. Response rate per day (%)

또한, Fig. 2는 시간에 따른 연챗의 응답률 상승을 나타낸다. DialogFlow에 학습된 질의가 많아짐에 따라, 자체의 응답률이 16일 3.2%에서 28일 74.5%로 향상되었다.



Fig. 3. Response time per day(ms)

Fig. 3은 제안된 기법을 통해 응답 속도가 향상됨을 시사한다. 16일 평균 1,544ms의 응답시간은 점차 감소하여 28일 평균 1,006ms로 향상되었다. 특히 26일에는 926ms의 응답시간을 보였다.

본 연구에서는 규칙 기반 챗봇 프레임워크의 한계점을 극복하는 점진 학습 기법을 도입한 챗봇을 제안하였고, 연챗을 통해 효율성과 성능 향상을 입증하였다. 향후 연구에서는 유사도 측정의 효율성을 향상을 위해 단어 벡터의 도입한다면, 더욱 일반화된 의도 파악을 구축할 수 있을 것으로 사료된다.

ACKNOWLEDGMENT

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (IITP-2017-0-00477, (SW스타랩) IoT 환경을 위한 고성능 플래시 메모리 스토리지 기반 인메모리 분산 DBMS 연구개발)

REFERENCES

- [1] D. Braun, A. Hernandez-Mendez, F. Matthes, and M. Langen, “Evaluating natural language understanding services for conversational question answering systems,” SIGDIAL, pp. 174-185, 2017.
- [2] S. Kim, and S. Hwang, “Designing Conversational AI with Mixed Initiatives: A Case Study,” KCC, pp. 628-630, 2018.