

통계적 단어 대조를 이용한 음식점 추천 챗봇 애플리케이션 구현

민동희¹, 이우범^{1*}

¹상지대학교 정보통신공학과

Implementation of a Chatbot Application for Restaurant recommendation using Statistical Word Comparison Method

Dong-Hee Min¹, Woo-Beom Lee^{1*}

¹Department of Information and Communication Engineering, Sangji University

요약 사용자로부터 입력되는 비정형 데이터를 대화 형태로 이해하여 사용자가 원하는 정보에 대한 맞춤 서비스를 제공하는 챗봇은 모바일 서비스의 중요한 분야로서 주목받고 있다. 그러나 사용자의 자연 언어 형태의 질의 대화를 완전하게 이해하여 서비스할 수 있는 방법은 아직 미흡한 실정이다. 따라서 본 논문에서는 사용자가 음식점 추천을 위하여 입력하는 대화 문장으로부터 지역, 음식분류, 음식점명 등의 의미 단어를 추출하고, 추출된 단어를 SNS의 음식점 추천 관련 해시태그를 기반으로 구축된 지식 데이터베이스의 내용과 대조하여 통계적으로 단어 유사성이 가장 큰 사용자 목적 정보를 제공한다. 본 논문에서 구현한 음식점 추천 챗봇 시스템의 성능 평가를 위해서 웹 기반의 모바일 환경을 구축하여 다양한 사용자 질의 정보에 대한 접근 편의성을 측정하고, 기존 유사 서비스와 비교하여 터치 횟수와 화면 전환 횟수에서 각각 37.2%와 73.3%의 감소함을 보였다.

• 주제어 : 챗봇, 음식점 추천, 통계적 단어 대조, 접근성 평가

Abstract A chatbot is an important area of mobile service, which understands informal data of a user as a conversational form and provides a customized service information for user. However, there is still a lack of a service way to fully understand the user's natural language typed query dialogue. Therefore, in this paper, we extract meaningful words, such a region, a food category, and a restaurant name from user's dialogue sentences for recommending a restaurant. and by comparing the extracted words against the contents of the knowledge database that is built from the hashtag for recommending a restaurant in SNS, and provides user target information having statistically much the word-similarity. In order to evaluate the performance of the restaurant recommendation chatbot system implemented in this paper, we measured the accessibility of various user query information by constructing a web-based mobile environment. As a results by comparing a previous similar system, our chabot is reduced by 37.2% and 73.3% with respect to the touch-count and the cutaway-count respectively.

• Key Words : Chatbot, Restaurant Recommendation, Statistical Word Comparison, Accessibility measurement

Received 01 March 2019, Revised 25 March 2019, Accepted 30 March 2019

* Corresponding Author Woobeom Lee, Department of Information&Communication Engineering, Sangji University, 83, Sangjidae-gil, Woonju-city, Gangwon-do, Korea. E-mail: beomlee@sangji.ac.kr

I. 서론

챗봇(Chatbot; Chatting Robot)은 사용자에게 친숙한 메신저 형식으로 문자 대화를 통해 질의를 해석하여 사용자에게 서비스를 제공하는 인공지능 소프트웨어이다. 이러한 챗봇은 필요 앱의 검색과 설치로 발생하는 테크노스트레스(technostress)의 감소와 사용의 편의성으로 스마트 라이프 원칙을 보장하기 때문에 관심이 급격히 증가하고 있다.

또한 능동적 양방향 소통이 가능한 지능형 챗봇은 웹의 장점인 서비스 개선, 업데이트의 용이함과 애플리케이션의 장점인 접근성과 이동성을 동시에 보유하고 있어 챗봇이 메신저 플랫폼 시장을 주도할 것으로 예측하고 있다[1][2].

챗봇의 핵심기술은 대화형 시스템으로서 사용자의 비정형 형태의 자연어 질의에 대한 오류 정규화, 형태 분석, 구문분석, 의미분석 등을 요구하며, 적절한 응답을 위해서 많은 양의 질의 패턴을 미리 구축해 두어야 한다. 또한 논리적 추론까지 가능해야 하기 때문에 현실의 상황을 정보화하고 활용하는 지능화된 서비스를 제공할 수 있어야 한다[3][4].

본 논문에서는 사용자의 연속된 대화형 질의문으로부터 유의미한 단어를 추출하고, 추출된 단어의 통계적 단어 대조 방법을 이용하여 음식점 추천 정보를 제공하는 챗봇 애플리케이션을 제안한다[5]. II 장에서는 사용자 질의문으로부터 유의미한 단어를 추출하기 위한 통계적 방법을 소개하고, III 장에서는 챗봇 시스템의 구조와 구성요소에 대해 서술하며, IV 장에서는 기존 애플리케이션과 제안하는 챗봇의 접근 편의성을 비교하여 성능을 평가한다. 그리고 마지막 V 장에서 결론을 맺는다.

II. 통계적 단어 대조 방법

본 논문에서는 사용자의 질의문을 해석 및 처리하기 위하여 통계적 단어 대조 방법을 사용한다. 사용자가 메신저에서 입력하는 자연어 질의문은 문장의 다양성과 의미적 중의성, 화자의 심리가 더해져 의미해석에 있어서 모호함을 보인다. 따라서 단순히 질의에 대한 답안만 수집하여 사용자가 원하는 서비스를 제공하는 것은 한계가 있다. 따라서 본 논문에서는 사용자의 전체 질의문장에서 사용자가 원하는 정보를 추출할 수

있는 의미단어를 추출하기 위하여 통계적인 방법을 적용한다.

통계적 단어 대조를 위해서는 먼저 유의미 단어 추출을 수행한다. 유의미 단어 추출은 질의문장 내에 존재하는 조사 및 의미 없는 불용어를 미리 보유한 데이터베이스 정보를 이용하여 제외함으로써 문장에서 음식 정보와 관련된 의미를 갖는 단어만 남기도록 단어 단위의 분석을 수행한다. 이때 검색 서비스를 요구하는 사용자는 일반적으로 검색 정보를 획득하기 위해 SNS의 해시태그와 같이 단어만으로 제시하는 사용 형태를 보인다. 해시태그는 SNS에서 사용되는 기호로서 해시 기호 뒤에 특정 단어를 쓰면 그 단어에 대한 정보만 분류하여 사용자에게 제시할 수 있다.

따라서 본 논문에서 사용하는 통계적 단어 대조 방법은 SNS에서 음식점 추천에 관련된 해시태그의 사용 빈도를 분석한 통계적 데이터를 이용하여, 연관된 대표 단어를 의미 단어로 추출한다. 그 다음 통계적 단어 유사성에 의해 추출된 단어를 적용하여 사용자의 요구를 분석하고 그에 상응하는 서비스를 결정하여 제공한다.

Table 1. Example of meaning words extraction

| Sentenses | Preprocessing | Meaning Word |
|-----------|---------------|--------------|
| 근처에 | 근처/~에(조사) | 근처 |
| 한식점을 | 한식점/~을(조사) | 한식점 |
| 추천하는 | 추천/~하는(조사) | 추천 |
| 찾아줄래 | 찾아줄래(목적내포) | X(불용) |
| 삼계탕 | 삼계탕 | 삼계탕 |

표 1은 챗봇에 입력되는 질의문의 한 예로서 한 메시지 질의 혹은 대화 시에 발생하는 다중의 질의 집합이 될 수도 있다. 표 1에서 문장 ‘근처에’를 질의로 입력받았을 때 전처리 과정에서 조사 ‘~에’를 삭제한다. 그리고 ‘찾아줄래’의 경우 애플리케이션의 목적과 동일한 의미를 가짐으로 유의미 단어에서 제외하고 불용 처리한다. 또한 ‘삼계탕’의 경우 음식을 나타내는 명사이기 때문에 유의미 단어로 추출된다.

유의미 단어 집합을 이용하여 통계적 단어 대조 방법에 의해서 사용자가 원하는 정보를 제공하기 위해서는 제안하는 챗봇 시스템은 미리 음식 관련 위치, 음식분류, 식당명 등에 대한 데이터베이스를 보유하고

있어야 한다. 따라서 본 논문에서는 한국관광공사에서 제공하는 TourAPI를 기본 데이터베이스 정보로 활용한다. 통계적 대조를 위한 챗봇 시스템은 사용자와 소통하기 위한 앱 플랫폼, 사용자 질의 분석 및 대조를 위한 서버, 데이터베이스, 웹페이지, TourAPI 등으로 구성된다.

앱은 메시지 형식으로 사용자에게 이동성과 접근성이 보장된 서비스를 제공한다. 웹페이지는 안드로이드 애플리케이션이 보안 문제로 애플리케이션의 데이터베이스 직접 접근을 용인하지 않기 때문에 웹 페이지 및 URL을 이용한 데이터베이스 위치 은닉을 통해 애플리케이션과 서버의 통신 매개체 역할을 수행한다. 서버는 웹으로부터 전달 받은 문장의 통계적 대조를 위하여 전처리 과정과 데이터베이스 색인 기능을 수행한다. 이 때 데이터베이스 색인 기능은 서버가 처리한 유의미 단어 집합과 상응하는 서비스 형태를 내포하고 있는 단어의 대조 기능을 의미한다.

III. 음식 추천 챗봇 시스템

본 논문에서 음식 추천을 위한 챗봇 시스템은 그림 1과 같이 사용자 애플리케이션을 중심으로 웹서버, TourAPI 서버, AI 서버의 연동형 구조의 플랫폼으로 구성된다.

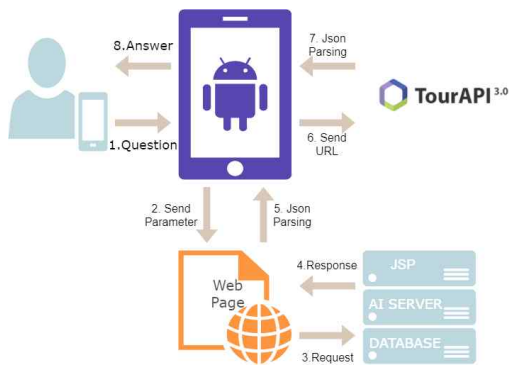


Fig. 1. Chatbot System Architecture

그림 2의 애플리케이션은 TourAPI 서버, 웹서버, AI 서버와의 통신에 의한 사용자 I/O 업무를 수행한다. 먼저 사용자가 문장 및 단어로 구성된 질의를 입력하면 문자열 체크를 이용해 오작동 유무를 확인 후 UI에 사용자 질의를 등록하고, 해당 내용을 웹페이지를 통해

AI 서버로 전달한다. 그리고 AI 서버가 반환한 JSON(JavaScript Object Notation) 포맷 코드를 해독하여 TourAPI 서버에 접속하기 위한 URL을 생성한다. 이때 만약 URL 생성을 위한 필요 코드가 부족할 경우에는 애플리케이션은 사용자에게 추가적인 데이터를 대화형식으로 요청한다.

그 다음 애플리케이션은 URL을 이용하여 TourAPI 서버에 사용자 질의에 해당하는 정보 제공을 요청하고, 반환된 TourAPI 서버 데이터를 메시지 체커에서 재가공하여 UI에 출력한다. 응답을 종료한 후에는 애플리케이션은 기존에 수집한 코드를 폐기하고 대기한다.

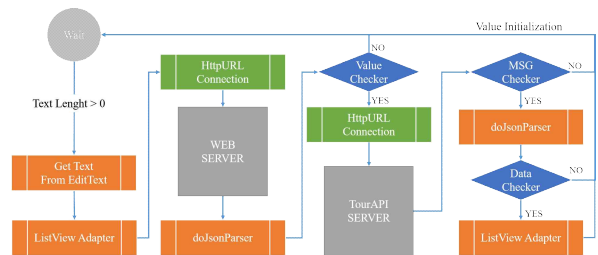


Fig. 2. Application Process Procedure

AI 서버는 그림 3과 같은 처리 과정으로 데이터베이스 접근과 사용자 질의 문장의 전처리를 수행한다. 서버는 사용자의 질의가 웹을 통해 전송되면 문장을 통계적 단어 대조 방법을 적용하기 위하여 전처리 과정으로 유의미 단어화 추출 작업을 진행한다. 단어화 작업은 데이터베이스에 미리 수집된 단어를 이용하여 서비스와 관련된 단어를 추출하고, TourAPI에 유효한 코드로 반환하여 코드를 JSON 형식으로 웹에 출력한다.

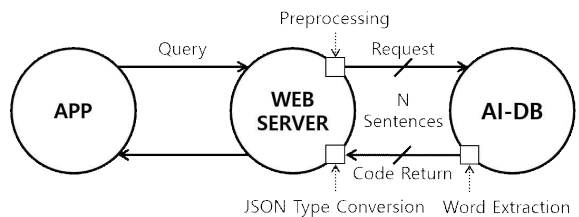


Fig. 3. Server Process Procedure

이 경우에 사용되는 데이터베이스는 통계적 단어 대조를 구현하기 위해 데이터베이스 테이블에 사용자가 원하는 서비스에 해당하는 단어들 이 TourAPI에 유효한 코드 형식으로 저장되어 있다. AI 서버로부터 단어 추출 작업 요청 시에는 통계적으로 검색을 위해 사용하

는 단어를 이용해 문장의 단어화 작업을 진행한다. 최종적으로 단어를 TourAPI에 유효한 코드로 치환하여 서버로 전달하여 데이터베이스의 통계적 정보로부터 유의미 단어 집합에 가장 최적의 결과를 추론한다. 이 경우에 사용자 예측 질의에 대한 데이터의 중복 저장으로 발생하는 저장 공간 낭비를 예방하기 위해서 데이터베이스는 단일 테이블로 구성되지 않는다.

본 논문에서 사용되는 TourAPI는 국내 유일의 실시간 다국어 관광 정보로 한국관광공사에서 제공하고 있는 서비스이다. 관광과 관련된 다양한 애플리케이션 개발 및 활용이 가능하며, 영문, 일문 등 9종의 언어를 이용해 정보를 제공하고 있다. 또한 정보 요청을 위해 URL을 작성하여 전달하는 방식으로 사용되며, URL 작성을 위해 TourAPI 메뉴얼에 기재된 특수한 코드를 사용한다.

IV. 실험

시스템 성능평가를 위한 애플리케이션은 스마트폰 LG V10 모델에서 Android Studio 2.3.3 환경으로 구현하였으며, 통계적 단어 대조 방법 및 전처리를 위한 서버는 Apache Tomcat 9.0과 Windows 10 환경에서 MySQL 5.7을 데이터베이스로 사용하였다.

본 논문의 성능 평가는 사용자 대화형 질의에 대한 해석 능력 및 편의성과 관련된 사용자의 애플리케이션 접근성을 측정하였다. 해석 능력 실험을 위해 질의 해석을 위해 조사를 포함한 2,251개의 점포 및 지역 데

이터를 제작하고, 코드화 과정에서 필요한 TourAPI의 특수한 점포 및 지역코드 1,113개를 수집하였다. 수집된 데이터는 그림 4와 같이 조사를 포함한 점포명, 지역명, 음식종류, 위치 및 평가 관련 단어는 데이터베이스의 QUERY 테이블에 (지역명, 점포명), (지역명, 음식종류), (현재위치)는 CODE 테이블에 저장하였다.

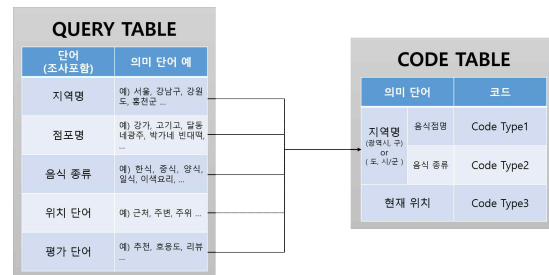


Fig. 4. Database Table Structure

그림 5는 (광역시, 구), (도, 시/군) 관련 지역 기반 음식점 추천 시뮬레이션의 예이다. (a)는 애플리케이션 실행 대기화면이고, (b)는 사용자와 애플리케이션 간 질의 응답을 보인다. 사용자가 질의 입력 시 통계적으로 (광역시, 구)와 관련된 단어 ‘서울시’를 입력하였으나, 챗봇은 URL 제작에 사용할 추가정보를 수집하기 위해 사용자에게 세부 지역(구)을 요청하는 질문을 한다. (c)는 시스템이 서비스를 제공하기 위해 필요한 정보를 모두 수집하였을 때 추천 점포명 ‘미친등심(구, 서울고깃집) 서울특별시 강남구 도산대로101길 7 (청담동)에 있어요’ 과 위치 정보를 서비스하는 결과이다. (d)는 같은 질의에 대해 또 다른 서비스 결과이다.



Fig. 5. Simulation Example of Region-based Restaurant Recommendation Service

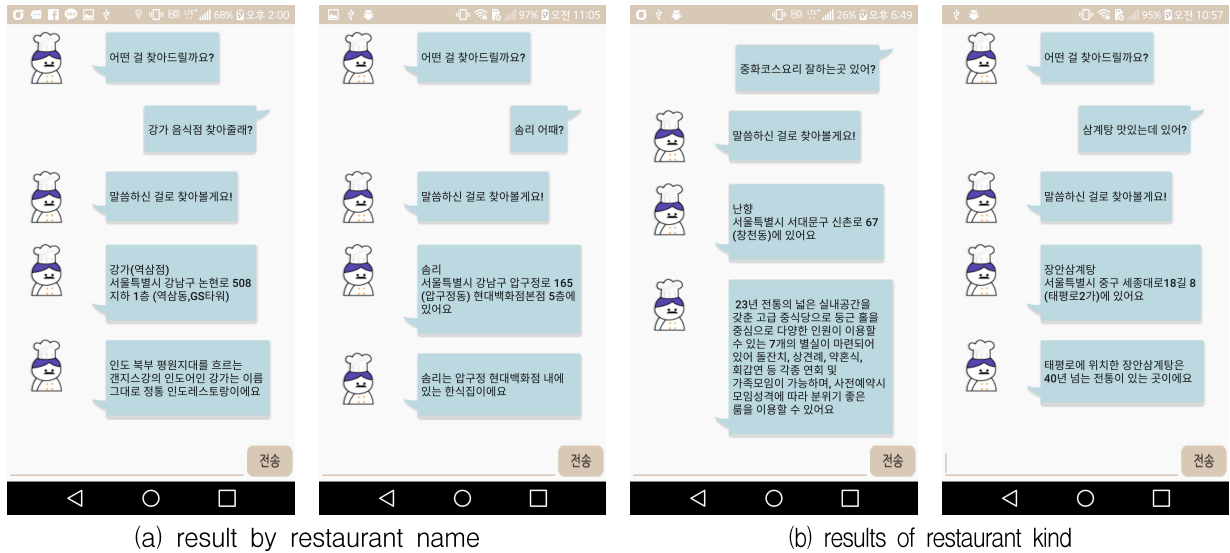


Fig. 6. Simulation Example of Restaurant Detailed Information Service

그림 6은 음식점 세부정보 제공 서비스의 시물레이션 예이다. (a)는 점포명 ‘강가’와 ‘솨리’로 서비스를 요구하였고, (b)는 음식종류 ‘중화코스요리’와 ‘삼계탕’으로 요구하였다. 그 결과로 점포의 위치 및 간단한 설명을 제공한다.

표 2는 본 논문의 챗봇 시스템에 대한 성능평가를 위해서 오픈 마켓에서 50만 건 이상 다운로드 받은 기존 음식점 추천 애플리케이션과의 최종 정보에 대한 접근성 비교 결과를 나타낸다.

실험 내용으로 ‘서울시 강남구 신사동 고기고’ 음식점을 가장 짧은 문구로 검색 및 조회까지 요구되는 사용자 터치 횟수와 화면 전환 횟수이다. 실험 기준으로 애플리케이션 첫 실행화면을 이용하였다. 단, 실험의 형평성을 위해 로딩, 로그인 및 회원 가입에 필요한 소비자원은 배제하였다. 실험 결과 기존 애플리케이션들이 서비스를 위해 소비하는 터치 횟수와 화면 전환 횟수는 각각 평균 12.75회, 3.75회로써 제안하는 시스템의 8회, 1회의 측정 결과는 각각 37%와 73%의 접근성 감소율을 나타낸다.



Fig. 7. Simulation Example of User location-based recommendation Service

그림 7은 사용자 위치기반 추천 서비스의 시물레이션 예이다. (a)는 사용자의 위치정보 권한 획득을 위한 챗봇의 요청 예이고, (b)는 사용자가 주변 음식점과 관련된 질의에 대한 서비스 결과로서 추천 음식점명과 위치 그리고 현 위치에서 목적지까지 거리로 응답하였다.

Table 2. Accessibility Comparison of Our Chatbot

| Application | Touch | Cutaway | Reduce Rate | |
|-------------|-------|---------|-------------|------|
| Diningcode | 13 | 3 | -38% | -66% |
| Kakaoplace | 13 | 3 | -38% | -66% |
| Mangoplate | 13 | 6 | -38% | -83% |
| 전국맛집 | 12 | 3 | -33% | -66% |
| Ave | 12.75 | 3.75 | -37% | -73% |
| Our Chatbot | 8 | 1 | - | - |

V. 결론

본 논문에서 모바일 기반의 애플리케이션과 웹 접근

방법에 기반 한 통계적 단어 대조 방법을 적용한 음식점 추천 챗봇 애플리케이션을 구현하였다. 사용한 통계적 대조 방법은 SNS에서 음식점 추천에 관련된 해시태그의 사용 빈도를 분석한 통계적 데이터를 이용하였다. 구현한 챗봇 시스템의 성능 평가를 위해서 지역 기반 추천 서비스, 음식점 세부 정보 서비스, 사용자 위치 기반 추천 서비스 등의 사용자 요청 정보에 대한 실험을 진행한 결과, 사용자 질의에 대해 높은 이해력을 보였다. 또한 친숙한 메신저 UI에 의해서 사용법이 간편하며, 기존 애플리케이션보다 감소한 터치횟수와 화면전환으로 사용의 접근성과 편의성을 보장한다.

향후 자연어 해석 알고리즘 수정과 빅데이터 지원을 통한 불명확한 질의에 대한 처리 능력을 향상시킨다면, 다양한 분야에서 특화된 챗봇 시스템 구현에 큰 효과를 기대 할 수 있다.

REFERENCES

[1] H. G. Park, "Read myself Chatbot!" Excellence Marketing for Customer, Korea Marketing Association, vol 51(5), pp.40-50, May, 2017.

[2] J. T. Kim, "Trends and Application of Messenger-based Chatbot Technology", Defence Technology, vol 459, pp.118-127, May, 2017.

[3] G. W. Hong, "A Korean Mobile Conversational Agent System", Korea Society Of Computing Information, 13th ed., vol. 6, pp.263-271, 2008.

[4] B. K. Kim, "Anonymous Q&A System based on Social ChatBot", Proceedings of KIIT Summer Conference, pp.493-495, June, 2017.

[5] D. H. Min, C. H. An, K. H. Kang, and W. B. Lee, "Implementation of Restaurant Recommendation Application using the Mobile Chatbot Technology Based on Messenger", KIPS Conference 2017, vol 18(1), pp.110-112, Dec, 2017.

[6] M. J. Kang, "A Study of Chatbot Personality based on the Purposes of Chatbot", Korea Contents Association, 18th ed., vol. 5, pp.319-329, May, 2018.

[7] M. J. Kang, "A Study on Chatbot Profile Images Depending on the Purpose of Use", Korea Contents Association, 18th ed., vol. 12, pp.118-129, Dec, 2018.

[8] J. J. Park, "A Development of Chatbot for Emotional

Stress Recognition and Management using NLP", Korean Institute of Electrical Engineers, 67th ed., vol. 7, pp.954-961, July, 2018

[9] S. W. Oh, "(ro)Bot as media - An experimental discussion on news chatbot", Korean Association For Communication And Information Studies, 79th, pp.70-103, Oct, 2016.

저자 소개

민 동 희 (Dong-Hee Min)



2018년 2월 : 상지대학교
정보통신공학과(공학사)
2018년 3월~현재 : 상지대학원
컴퓨터정보공학부 정보통신공학전공

이 우 범 (Woo-Beom Lee)



1995 영남대학교 컴퓨터공학과 학사
1997 영남대학교 컴퓨터공학과 석사
2000 영남대학교 컴퓨터공학과 박사
2000.03-2004.02 대구과학대학교 컴퓨터공학과 교수
2004.03-2007.02 영남대학교 전자정보공학부 컴퓨터전공 특임교수
2007.03-현재 상지대학교 정보통신공학과 교수
관심분야 : 뉴로비전컴퓨팅, IT융합의료기(한방의료영상), 딥러닝(신경회로망)