
저자 (Authors)	오상현, 김성희 Sang-Heon Oh, Sung-Hee Kim
출처 (Source)	한국정보통신학회 종합학술대회 논문집 23(2) , 2019.10, 71-74(4 pages)
발행처 (Publisher)	한국정보통신학회 The Korea Institute of Information and Communication Engineering
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09262372
APA Style	오상현, 김성희 (2019). 챗봇 기반의 지능형 시각화 프레임워크. 한국정보통신학회 종합학술대회 논문집, 23(2), 71-74
이용정보 (Accessed)	한국방송통신대학교 203.232.176.*** 2020/02/03 17:50 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

챗봇 기반의 지능형 시각화 프레임워크

오상현* · 김성희

동의대학교

Intelligent Visualization Interaction Framework Based on Chatbot

Sang-Heon Oh* · Sung-Hee Kim

Dong-eui University

E-mail : shoh9179@gmail.com / sh.kim@deu.ac.kr

요 약

현재까지의 데이터분석은 관계형 데이터베이스 내에 수집된 데이터를 기반으로 SQL을 통해 주로 전문가들에 의해 이루어졌다. 빅데이터 시대에 접어들면서 데이터의 폭발적인 증가로 인해 이제는 전문가들뿐만 아니라 일반인들에게도 데이터분석은 필요한 영역이 되고 있다. 기업이나 기관에서는 보다 효과적인 데이터 분석을 위해 시각화를 활용하고 있으며, 데스크탑을 시작으로 모바일, 대형 디스플레이 등 다양한 환경으로 변화하고 있다. 하지만 시각화에 대한 인터랙션은 데스크탑 기반으로 발전해 왔으며, 모바일과 대형 디스플레이 환경에서의 최적화된 방법에 대한 연구가 필요하다. 따라서 본 논문에서는 접근성이 높고 쉽게 사용가능한 챗봇 기반의 지능형 시각화 프레임워크를 제안한다. 사용자는 챗봇을 통해 데이터 분석에 대한 요구사항을 텍스트로 입력하고, 이는 기계학습을 통해 쿼리로 변환되어 출력된 결과 데이터를 시각화하는 지능형 시각화 프레임워크를 제안한다.

ABSTRACT

Data analysis so far has been done primarily by experts through SQL based on data collected within relational databases. However, with the explosive growth of data, data analysis is not only important to domain experts, but also necessary, to the general public. Companies and institutions are using visualizations for effective data analysis. And data analysis is being conducted in various types of environments, from traditional desktop to mobile or to large displays. However, interaction with visualizations has evolved on a desktop-based basis and needs to be studied on how to optimize in mobile and large display environments. Therefore, this paper proposes an intelligent visualization framework based on chatbot that is highly accessible and readily available. Users can enter text through chatbot, which creates and executes queries through machine learning, then proposes an intelligent visualization framework that visualizes output result data.

키워드

자연어처리, 챗봇, 기계학습, 지능형 시각화, 프레임워크

1. 서 론

현재까지의 데이터 분석은 SQL(Structured Query Language)을 기반으로 데이터를 처리하는 데이터베이스

에 의해 주로 이루어져 왔고, 이러한 데이터를 시각화하여 의미 있는 정보를 찾아내는 과정으로 이루어졌다. 빅데이터 시대에 접어들면서 특정 분야의 전문가뿐만 아니라 이제는 일반인들도 데이터에 대

* speaker

해 이해하고자 한다. 도메인(Domain) 지식을 가진 전문가, 중요한 의사결정이 필요한 경영진, 그리고 모바일 기기를 통해 수집된 헬스케어 관련 데이터, IoT 기기를 통한 심박수, 혈압 등과 같은 생체 데이터들을 스스로 생산해내며, 이를 효과적으로 활용하고자 하는 일반인들도 이제는 시민 데이터 과학자의 역할로 데이터를 직접 분석해야하는 필요성을 느끼고 있다[1]. 또한 중요한 의사결정이 필요한 기업이나 기관에서는 보다 효과적으로 데이터를 분석하고 이해하기 위해 시각화를 함께 활용하여 의사결정에 활용하며[2], 이를 통해 분석된 결과를 설명하거나 레포트를 작성하기도 한다.

기존의 데이터 시각화에 대한 연구와 발전은 데스크톱을 기반으로 개발이 이루어졌다. 화면의 크기에 제약을 받지 않으며, 확대나 필터링 등의 인터랙션은 마우스를 통해 이루어졌다. 하지만 모바일 환경으로 변화하면서 화면의 크기는 작아졌고 보여줄 수 있는 정보의 양이 한정되어 인터랙션의 필요성이 늘어났지만, 터치 인터페이스로 인해 마우스로 이루어졌던 기존의 정교한 인터랙션은 더욱 어려워졌다. 또한 벽면의 대형 디스플레이를 통한 시각화도 대시보드 형태로 많이 구현되고 있다. 따라서 화면과 떨어져 있는 상황에서 데이터를 보게 되기 때문에 인터랙션은 누군가 컴퓨터 앞에서 마우스로 직접 제어해야 하는 상황이다[3].

이러한 기기 변화에 따른 어려움 이외에도 시각화에 대한 전문지식이 없는 일반인들이 데이터 시각화를 활용하기 위한 진입 장벽이 높은 상황이며, 이를 극복하기 위해 Tableau, Power BI에서는 자연어처리를 통해 사용자가 원하는 질의로 시각화를 자동으로 추천하여 제공하고 있다. 하지만 소프트웨어 역시 데스크톱 환경에서만 활용이 가능하며, 연속적인 질의를 통한 인터랙션보다는 단발성의 질의로 시각화를 추천하는 방식을 사용한다[4].

따라서 일반 사용자들도 접근성이 높고 쉽게 활용할 수 있는 챗봇을 활용하고자 한다. 사용자는 쉽게 챗봇과 대화하며 원하는 정보를 시각화된 자료로 볼 수 있기 때문에 더욱 효과적인 데이터 분석과 이해로 이어질 수 있다.

본 논문에서는 챗봇 기반의 지능형 시각화 인터랙션 프레임워크를 제안하고자 한다. 사용자는 챗봇에서 원하는 정보의 텍스트를 입력하고, 입력된 데이터는 딥러닝(Deep Learning) 기술을 통해 텍스트를 학습하여 쿼리에 필요한 의미 있는 키워드를 추출해낸다. 이를 통해 SQL을 생성하고 해당 쿼리를 시각화하는 지능형 프레임워크를 제안하고자 한다. 또한 이에 고려해야 하는 사항들을 분석하며, 차후의 연구에 밑바탕이 될 수 있도록 한다.

II. 관련 연구

최근 자연어처리를 통한 SQL 작성 관련 연구가

이루어지고 있다. 정윤기 외 2인은 관계형 데이터베이스 시스템에서 자연어처리를 통해 입력된 문장을 SQL문으로 변환하는 연구를 진행하였다. 해당 연구에서 쿼리의 From절에 입력되는 컬럼을 효과적으로 예측하는 것에 대해 개수가 증가하면 정확도가 떨어지는 문제를 발견하였으며, 이미지 분류에 주로 사용되는 합성곱 신경망을 기반으로 예측 모델을 만들어 정확도를 높이려 하였다[5].

심층 신경망의 자연어처리에 대한 모델 성능을 비교하는 연구도 진행되었으며, 해당 연구에서는 CNN(Convolution Neural Networks), Bi-RNN(Bidirectional Recurrent Neural Networks), 그리고 CNN과 Bi-RNN을 합친 모델에 대해서 가지의 말뭉치를 학습시켜 각 모델을 비교하였다. 결과적으로 CNN과 Bi-RNN을 합친 모델에서 소요 시간이 크다는 단점이 있었지만 전반적으로 가장 높은 정확도를 보였고, 각 모델이 말뭉치에 따라 다른 결과를 보이는 연구 결과가 있었다[6].

데이터셋에서 자동으로 시각화를 생성하는 연구 사례가 있다. 이 연구에서는 주어진 데이터셋에서 시각화를 자동으로 생성하기 위한 신경 변환 모델인 Data2Vis를 소개하고 있다. 입력된 시퀀스로부터 다른 도메인의 시퀀스를 출력하는 모델인 시퀀스 투 시퀀스를 기반으로 하여 순환신경망(Recurrent Neural Networks)에 장단기기억망(Long-term Short-Term Memory)을 적용하였으며, JSON(JavaScript Object Notation) 형태의 데이터와 Vega-Lite의 코드 문법을 맵핑하고 이를 학습시켜 시각화를 만들어 내는 연구를 진행하였다[7].

유소엽 외 1인은 챗봇에서 지식 그래프를 활용한 연구를 진행했다. 해당 연구에서는 단어 임베딩에 대해 사전 학습이 되어있는 BERT를 기반으로 관계 추출 모델을 적용시켜 자동 확장할 수 있는 지식 그래프를 구축했다. 챗봇을 기반으로 사용자가 질문하면 학습된 모델을 통해 질문을 분석하고 이에 답이 될 수 있는 경우들을 지식 그래프로 출력하며, 가장 관련성이 깊은 답일수록 노드의 크기를 크게 표현하여 사용자들이 보다 직관적으로 알 수 있도록 시각화 한 사례이다[8].

III. 본론

i. 전체 프레임워크

본 논문에서 제안하는 프레임워크의 전체적인 프로세스는 그림 1과 같다. 먼저 사용자는 챗봇에게 원하는 정보를 요청한다. 예를 들어, ‘저번 달 내 몸무게를 보여줘.’ 라는 텍스트를 입력하면 기계학습 모듈에서 해당 텍스트를 전송받고 사용자가 원하는 정보의 쿼리를 생성해낸다. 쿼리를 실행하면 데이터베이스에서 저번 달의 몸무게를 조회하게 것이다. 이후 조회된 결과 데이터셋을 시각화 모듈에 전달하여 분석된 데이터 시각화를 사용자

에게 제공하게 된다.

따라서 사용자는 챗봇에게 원하는 정보를 텍스트로 입력하고 시각화 자료를 받아볼 수 있기 때문에 이는 사용자 입장에서 시퀀스 투 시퀀스로서의 개념으로 시각화 데이터 분석 서비스를 제공할 수 있다.

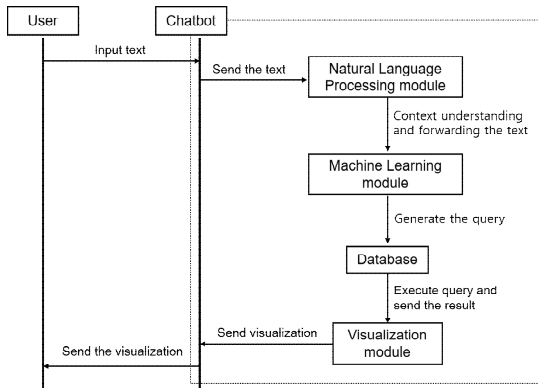


그림 1. 챗봇 기반의 지능형 시각화 프레임워크 전체 흐름도

이후 발생할 수 있는 사용자의 추가 질의에 대해 처리할 수 있어야 한다. 그림 2와 같이 사용자가 저번 달 자신의 몸무게를 물어보고 추가 질의로 ‘What about this week?’라는 질문으로 몸무게에 대한 질의임을 문맥을 이해를 통해 해석할 수 있어야 한다. 따라서 자연어처리 모듈에서 문맥을 이해하는 과정을 거친 후 기계학습 모듈로 문장을 전달해야 한다. 채팅은 단발성으로 연속적인 대화가 이루어지며, 이와 같은 대화의 문맥을 이해함으로써 사용자가 요구하는 정보에 대한 시각화를 보다 원활하게 표현할 수 있을 것이다.

ii. 문제점 및 고려사항

본 논문에서 제안하는 프레임워크를 구현하기 위해서는 여러 전제조건과 문제점들을 해결해야 할 필요가 있다.

먼저, 데이터베이스 내의 데이터들과 사용자 질의에 관련한 도메인, 데이터셋의 문제를 고려해야 한다. 사용자가 요구하는 정보에 따른 데이터셋을 데이터베이스에서 보유하고 있어야 하며, 해당 쿼리를 만들어내는 기계학습 모듈에서도 데이터베이스 내의 테이블명, 컬럼 등의 정보와 동기화가 이루어져야 한다. 또한 문맥을 이해한 후의 쿼리 생성이 이루어지기 때문에 자연어처리 모듈이 쿼리 생성의 정확도에 영향을 줄 수 있다. 따라서 각 모듈의 높은 정확성을 보여야 한다.

또한 Tableau의 Ask Data 기능, Power BI의 자동완성 기능은 데스크탑 기반의 소프트웨어에서 제공하고 있다. 큰 디스플레이를 통해 시각화를 표현할 수 있으며, 이를 제어할 수 있는 키보드와 마

우스 등의 인터랙션 방법이 다양하고 정교하다는 장점이 있다. 하지만 챗봇을 활용하기 위한 모바일 환경에서는 데스크탑 환경에 비해 정보를 표현할 수 있는 디스플레이는 매우 작으며, 발생할 수 있는 인터랙션은 터치, 채팅을 통한 대화의 방법으로 제한되어 있다. 따라서 사용자의 편의성을 고려하여 챗봇 기반에서 시각화를 표현하기 위한 인터페이스를 설계해야 한다.

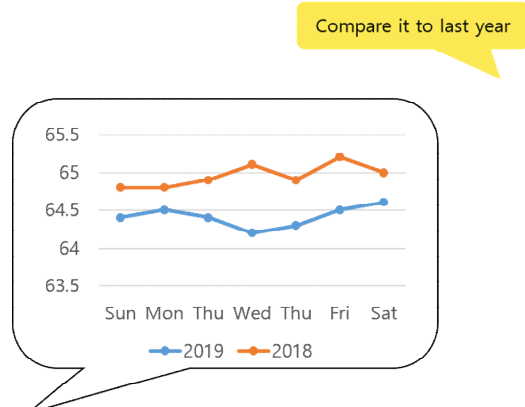
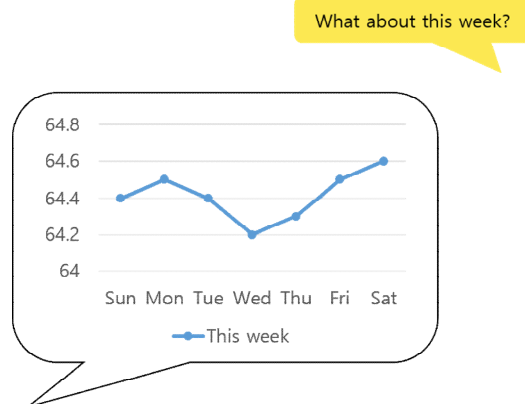
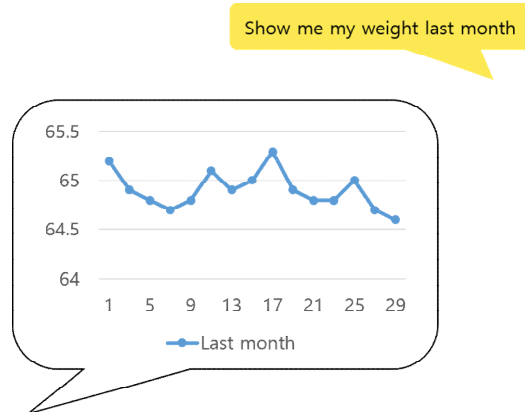


그림 2. 챗봇 기반의 지능형 시각화 프레임워크 예상 시나리오

시각화 모듈에서는 쿼리를 통한 결과 데이터셋에 어떠한 규칙에 따라 시각화해야 할 것인지 생각해야 한다. 지난 달 몸무게를 선 그래프로 표현해보면, 일반적으로 x축에는 지난달을 나타내는 시간 속성, y축에는 몸무게 수치와 같은 속성을 맵핑하여 그릴 수 있다. 하지만 각 축의 속성을 바꿔도 그래프를 충분히 표현할 수 있고, 다양한 그래프로도 시각화가 가능하다. 따라서 사용자에게 익숙한 차원의 그래프와 각 시각화의 특징을 잘 살려서 데이터셋을 표현해야 할 필요가 있다.

V. 결 론

본 논문에서는 챗봇 기반의 지능형 시각화 프레임워크를 제안하였다. 이를 통해 챗봇을 기반으로 일반 사용자들이 높은 접근성을 가지며, 보다 쉽고 직관적인 데이터 분석을 위해 시각화 할 수 있을 것이다.

하지만 이를 구현하기 위해서는 앞서 해결해야 할 문제들이 있다. 자연어처리를 통해 기계학습 모듈에서 생성해야하는 쿼리의 정확성이 현재까지의 연구에는 미흡한 부분이 있으며, 데이터베이스와의 안정적인 동기화가 필요하다. 시각화 모듈에서 결과 데이터셋을 통해 어떠한 규칙에 따라 데이터를 시각화할 것인지 고려해야 하며, 사용자와의 대화 문맥을 이해하여 요구사항을 정확하게 파악하는 것 또한 중요한 이슈로 남아있다. 챗봇의 사용 목적에 따른 도메인과 데이터셋에 대한 문제, 그로 인해 다른 시각화의 특징을 고려하여 결과를 출력해야 하는 것 이외에도 본 논문에서 언급하지 못한 다양한 문제들도 발생할 것이다.

차후에는 본 논문에서 제안하는 프레임워크를 구현할 예정이며, 진행 중 발생하는 새로운 문제점들을 계속해서 제기하고 해결해 나갈 것이다. 빠른 속도로 발전하는 기술, 연구들과 함께 가까운 미래에는 어렵지 않게 구현될 것이고, 일반 사용자들도 자신만의 데이터를 손쉽게 분석하고 인사이트를 찾아갈 수 있기를 기대한다.

Acknowledgement

이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019R1C1C1005508).

References

[1] Computer world. "Citizen data scientists are essential, not optional." [Internet]. Available : <http://www.comworld.co.kr/news/articleView.html?idxno=49315>

[2] K. K. Ko, D. W. Jeong, M. A. Lee, "Is data visualization active in administrative studies?", *The Korean Association For Public Administration*, 1919-1944. Dec. 2018.

[3] B. S. Lee, M. Brehmer, P. Isenberg, E. K. Choe, R. Langner, R. Dachsel, "Data Visualization on Mobile Devices.", *Extended Abstracts of the 2018 CHI Conference*, 1-8, Apr. 2018.

[4] Tableau, "Automatically Build Views with Ask Data" [Internet]. Available : https://help.tableau.com/current/pro/desktop/en-gb/ask_data.htm

[5] Y. K. Jeong, D. M. Kim, J. W. Lee, "A CNN-based Column Prediction Model for Generating SQL Queries using Natural Language", *The Journal of Korean Institute of Information Scientists and Engineers*, Vol. 46, No. 2, pp. 202-207, Feb. 2019.

[6] T. G. Lee, K. S. Shin, "Performance Comparison of Natural Language Processing Model Based on Deep Neural Networks", *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 44, No. 7, pp. 1344-1350, Jul. 2019.

[7] V. Dibia and Ç. Demiralp, "Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks," in *IEEE Computer Graphics and Applications*, Vol. 39, No. 5, pp. 33-46, 1 Sept.-Oct. 2019.

[8] S. Y. Yoo, O. R. Jeong, "An Intelligent Chatbot Utilizing BERT Model and Knowledge Graph", *The Journal of Society for e-Business Studies*, Vol. 24, No. 3, pp. 87-98, Aug. 2019.