

Bilinear Graph Neural Network-Based Reasoning for Multi-Hop Question Answering

Sangui Lee[†] · Incheol Kim^{**}

ABSTRACT

Knowledge graph-based question answering not only requires deep understanding of the given natural language questions, but it also needs effective reasoning to find the correct answers on a large knowledge graph. In this paper, we propose a deep neural network model for effective reasoning on a knowledge graph, which can find correct answers to complex questions requiring multi-hop inference. The proposed model makes use of highly expressive bilinear graph neural network (BGNN), which can utilize context information between a pair of neighboring nodes, as well as allows bidirectional feature propagation between each entity node and one of its neighboring nodes on a knowledge graph. Performing experiments with an open-domain knowledge base (Freebase) and two natural-language question answering benchmark datasets (WebQuestionsSP and MetaQA), we demonstrate the effectiveness and performance of the proposed model.

Keywords : Open Domain Question Answering, Knowledge Graph, Complex Question, Graph Neural Network

다중 홉 질문 응답을 위한 쌍 선형 그래프 신경망 기반 추론

이 상 의[†] · 김 인 철^{**}

요 약

지식 그래프 기반의 질문 응답 문제는 자연어 질문들에 대한 깊은 이해뿐만 아니라, 대규모 지식 그래프 상에서 올바른 답변을 찾기 위한 효과적인 추론 능력을 필요로 한다. 본 논문에서는 다중 홉 추론을 요구하는 복잡한 자연어 질문에 대해 연관 지식 그래프 위에서 답변 추론을 효과적으로 수행할 수 있는 심층 신경망 모델을 제안한다. 제안 모델에서는 지식 그래프 상의 각 개체 노드와 이웃 노드 간의 양방향 특징 전파를 허용할 뿐만 아니라, 두 이웃 노드 쌍 간의 맥락 정보까지 활용할 수 있는, 표현력이 뛰어난 쌍 선형 그래프 신경망(BGNN)을 이용한다. 본 논문에서는 오픈 도메인의 지식 베이스인 Freebase, 자연어 질문 응답을 위한 벤치마크 데이터 집합들인 WebQuestionsSP와 MetaQA를 이용한 실험들을 통해, 제안 모델의 효과와 우수성을 확인하였다.

키워드 : 오픈 도메인 질문 응답, 지식 그래프, 복잡 질문, 그래프 신경망

1. 서 론

대규모 지식 베이스를 토대로 인간을 대신해 자연어 질문에 스스로 답할 수 있는 지능형 에이전트는 실세계 다양한 분야에서 폭넓게 활용될 수 있다. 특히 최근 들어 심층 신경망(deep neural network)을 이용한 자연어 처리 기술들이 획기적으로 발전함에 따라, 지식 베이스 기반 질문 응답

(Knowledge Base Question Answering, KBQA) 문제와 이를 해결하기 위한 다양한 모델들에 관한 관심도 함께 증가하고 있다. 일반적으로 비정형 문서 집합(corpus)에 비해, 지식 베이스(knowledge base)는 대부분 트리플(triplet)과 같은 구조화된 형태로 지식을 저장하고 있으므로 질문 응답(QA)에 활용하기가 상대적으로 용이하다. 하나의 지식 트리플은 주어 개체(subject entity), 관계 서술자(relational predicate), 목적어 개체(object entity)로 구성되는데, 이것은 두 개체 간의 관계를 나타내는 하나의 사실(fact)로 간주할 수 있다. 따라서 다수의 트리플들로 이루어진 지식 베이스는 개체 노드(node)들과 그들 간의 관계를 표현하는 간선(edge)들로 구성되는 하나의 큰 지식 그래프(knowledge graph)로 볼 수 있다. 이러한 지식 그래프를 토대로 자연어 질문에 대한 답변을 찾기 위해서는 지식 그래프 위에서 전개되는 효율적인 추론(reasoning) 혹은 탐색(search) 과정이 요구된다.

* 본 연구는 정보통신기획평가원의 재원으로 정보통신방송 기술개발사업의 지원을 받아 수행한 연구과제(클라우드에 연결된 개별 로봇 및 로봇그룹의 작업 계획 기술 개발, 2020-0-00096)입니다.

** 이 논문은 2020년 춘계학술발표대회에서 "쌍 선형 그래프 신경망을 이용한 지식 그래프 기반 질문 응답"의 제목으로 발표된 논문을 확장한 것임.

† 준 회 원 : 경기대학교 컴퓨터과학과 석사과정

** 종 신 회 원 : 경기대학교 컴퓨터과학과 교수

Manuscript Received : June 11, 2020

First Revision : July 20, 2020

Accepted : July 22, 2020

* Corresponding Author : Incheol Kim(kic@kyonggi.ac.kr)

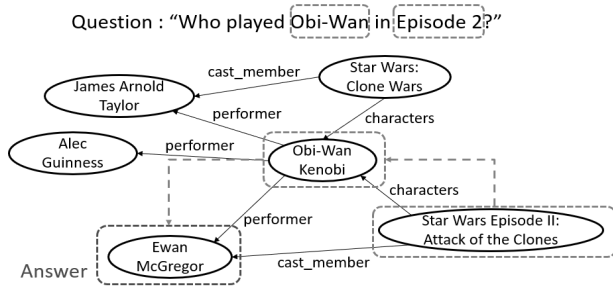


Fig. 1. Example of Multi-hop Inference on a Knowledge Graph

한편, 지식 그래프 기반 질문 응답 문제는 추론의 복잡도에 따라 단순 질문(simple question)과 복잡 질문(complex question)으로 나뉜다. 일반적으로 질문에 대한 답변을 위해 지식 그래프 위에서 단일 홉(single hop) 추론이 필요하나 아니면 다중 홉(multi-hop) 추론이 필요하나에 따라 단순 질문과 복잡 질문을 구분한다. 본 논문에서는 특히 다중 홉 추론을 요구하는 지식 그래프 기반 질문 응답 모델을 제안한다.

Fig. 1은 지식 그래프를 이용한 다중 홉 추론의 한 예를 보여준다. Fig. 1에서 개체는 타원형으로, 개체 사이의 관계는 간선으로 표현되어 있다. Fig. 1의 상단에서처럼 자연어 질문이 주어지면 지식 베이스로부터 질문과 연관된 지식 그래프를 추출해야 한다. 지식 그래프를 추출할 때 질문에 언급된 개체를 중심으로 추출하며, 추출한 지식 그래프로부터 답을 추론할 수 있어야 한다. 다중 홉 추론을 요구하는 복잡 질문인 “Who played Obi-Wan in Episode 2?”에 대해 답변인 Ewan McGregor를 찾으려면 먼저 질문과 연관 있는 트리플 <Star Wars Episode II: Attack of the Clones, characters, Obi-Wan Kenobi>을 찾아야 한다. 이 트리플로부터 질문의 답변인, Obi-Wan Kenobi를 연기한 Ewan McGregor를 포함하는 트리플 <Obi-Wan Kenobi, performer, Ewan McGregor>을 찾을 수 있어야 한다.

지식 그래프를 이용한 다중 홉 추론에 관해 기존 연구 [1]에서는 (키(key), 값(value)) 쌍의 메모리 슬롯에 트리플들을 저장하고, 자연어 질문과 키(key)의 연관성에 따른 값(value)들의 가중치 합(weighted sum)을 이용해 메모리 슬롯을 반복적으로 참조하여 답변 개체를 예측하는 KV-MemNN 모델을 제안하였다. 이때 키(key)는 트리플의 주어 개체와 관계 서술자로 이루어진 튜플(tuple)에 해당하고 값(value)은 목적어 개체에 해당한다. 한편, 기존 연구[2]에서는 그래프 합성곱 신경망(Graph Convolutional Network, GCN)[3]과 페이지랭크(PageRank) 점수를 이용한 지식 그래프 추론을 수행하는 GraftNet 모델을 제안하였다. 그러나 이러한 기존 연구들은 KV-MemNN 모델과 같이 지식 그래프 추론 과정에 그래프 이웃 노드들과의 맥락 정보(structural context)를 반영하려는 별도의 노력이 없거나, 그래프 합성곱 신경망(GCN)에 기초한 GraftNet 모델처럼 이웃 노드들의 특징 정보를 단순 선형 응집(Linear Aggregation, LA)해서 얻은 맥락 정보만을 활용하였다. 하지만 대규모 지식 그래프 위에서 다중

홉 추론을 요구하는 다양한 복잡 질문(complex question)들에 대한 올바른 답변을 찾아내려면, 각 노드와 이웃 노드 간의 맥락 정보뿐만 아니라 이웃 노드와 이웃 노드 간의 맥락 정보도 효과적으로 활용할 수 있는 그래프 신경망(Graph Neural Network) 기반의 지식 그래프 추론 모델이 요구된다.

이러한 기존 연구들의 한계점을 고려하여, 본 논문에서는 새로운 그래프 신경망 기반의 질문 응답 모델을 제안한다. 제안 모델에서는 그래프 노드들 간의 양방향 특징 정보 전파(bidirectional propagation) 기능을 포함하고 있을 뿐만 아니라, 두 이웃 노드 쌍 간의 맥락 정보도 활용할 수 있는 쌍 선형 응집(Bilinear Aggregation, BA) 연산이 추가된 쌍 선형 그래프 신경망(Bilinear Graph Neural Network, BGNN)을 새롭게 채용하고 있다. 본 논문에서는 제안 모델의 성능 평가를 위해, 오픈 도메인 지식 베이스인 Freebase와 질문 응답 벤치마크 데이터 집합들인 WebQuestionsSP[4] 및 MetaQA[5]를 이용한 다양한 실험들을 수행하고, 그 결과를 소개한다.

2. 관련 연구

2.1 오픈 도메인 질문 응답

오픈 도메인 질문 응답에 대한 지식으로 Wikipedia나 Freebase가 자주 이용돼왔다. Wikipedia는 문서 집합으로 이뤄진 비정형 데이터로 바로 사용하기에는 힘들다는 단점이 있다. 그래서 DBpedia[6] 같이 Wikipedia로부터 지식 베이스를 구축하려는 시도도 있었다. 한편, Wikipedia와 달리, Freebase는 지식 베이스이기 때문에 데이터가 이미 정형화 되어 있어 바로 사용하기 쉽다는 장점이 있다. 그런 이유로 Freebase는 오픈 도메인 질문 응답에서 자주 사용됐다. 본 논문에서도 Freebase를 사용하여 오픈 도메인 질문 응답을 수행하였다.

Freebase로부터 답변을 찾을 수 있는 오픈 도메인 질문 응답 벤치마크 데이터 집합으로 WebQuestions[7]가 제시되었지만, WebQuestions[7] 데이터 집합의 약 18.5%는 Freebase로부터 답변을 찾을 수 없는 것으로 나타났다[4]. 이후에 WebQuestions[7] 데이터 집합 중 Freebase로부터 답변을 찾을 수 없는 질문 응답 데이터를 제외한 데이터 집합인 WebQuestionsSP[4]이 제시되었다.

오픈 도메인 질문 응답에서 필요한 작업 중 하나는 자연어 질문 내 언급된 개체들을 지식 베이스 또는 문서 내에 일치하는 개체들에 연결하는 작업이다[8, 9]. 이 작업을 개체 연결(entity linking)이라고 한다. 개체 연결 시스템(entity linking system)에 따라서도 질문 응답 모델의 성능이 달라질 수 있다[8]. 본 논문에서는 S-MART[9]의 개체 연결 결과를 이용하였다.

오픈 도메인 질문 응답 연구에서 초기 연구들은 주로 단순 질문 응답 문제를 다루었다. 단순 질문 응답은 자연어 질문에서 주어 개체와 관계 서술자들을 추출한 후, 이들과 대응되는 지식 베이스 내의 목적어 개체를 찾는 방식으로 답변을 구할

수 있었다. 단순 질문 응답은 대개 단일 홉 추론을 요구하는 쉬운 문제에 속하기 때문에, 당시에 제안된 모델들은 다중 홉 추론을 요구하는 복잡 질문 응답에 대해서 높은 성능을 보이지 못했다. 이후에, 그러한 복잡 질문 응답 문제를 해결하기 위한 다양한 시도들이 있었는데 메모리 네트워크를 이용한 연구, 강화학습을 이용한 연구, 그리고 그래프 신경망을 이용한 연구들이 이어져 왔다.

메모리 네트워크를 이용한 연구들[1, 10]에서는 지식 베이스 또는 문서들을 메모리 슬롯에 저장하였다. 특히 키-값 메모리 구조를 제안한 KV-MemNN[1]은 기존의 메모리 네트워크 MemN2N[10]보다 뛰어난 성능을 보였다.

강화학습을 이용한 연구들[11, 12]에서는 강화학습을 통해 지식 베이스로부터 답변을 구하였다. NSM[11]은 해당 연구에서 제안하는 모델을 강화학습으로 최적화하였고, MINERVA[12]는 강화학습으로 지식 그래프 상에서 답변을 구할 수 있는 추론 경로를 탐색하도록 하였다.

하지만 이같은 선형 연구들의 문제점은 지식 그래프 상에서 관계로 연결된 개체와 개체 사이의 맥락 정보를 답변 추론에 충분히 반영하지 못하였다는 것이다. 따라서 해당 문제점을 해결하고자, 개체들 간의 맥락 정보를 효과적으로 활용할 수 있는 그래프 신경망을 이용하려는 연구들이 최근에 등장하였다[2].

2.2 그래프 신경망

과거에 제안됐던 합성곱 신경망(Convolutional Neural Network, CNN)은 1차원의 문자열 또는 2차원의 영상을 처리하는데 성공적인 효과를 보였다. 그러나 기술이 발전됨과 동시에 최근에는 실제계의 개체들과 개체들 사이에 다양한 관계가 존재하는 그래프를 처리해야 하는 상황이 많아졌다. 대표적인 예로 장면 그래프(scene graph), 사회망(social network), 단백질 같은 분자 구조 등등이 있다.

그러한 그래프 형태의 구조적인 데이터를 효과적으로 처리하기 위해 GGS-NNs(Gated Graph Sequence Neural Networks) [13], GCN(Graph Convolutional Network)[3], GAT(Graph Attention Network)[14], GraphSAGE[15], BGNN(Bilinear Graph Neural Network)[16] 등 다양한 그래프 신경망 모델들이 연구되었다. 질문 응답 문제에서도 지식 베이스로부터 추출한, 질문과 연관된 지식 그래프를 처리하기 위해 그래프 신경망을 이용한 연구들이 진행되었다. [17]의 연구에서는 질문을 이용하여 지식 베이스로부터 개체들과 관계들을 추출하면서 의미 그래프(semantic graph)를 생성하였고, 이 그래프를 의미 분석(semantic parsing)하여 임베딩하기 위해 GGS-NNs[13]을 적용하였다. GraftNet[2]의 연구에서는 다중 홉 추론을 수행할 때, GCN[3]의 특성을 이용하여 단순 선형 응집에 의한 이웃 노드들의 맥락 정보를 활용하였다. 이웃 노드들의 맥락 정보를 구할 때 노드와 노드 사이의 관계마다 질문과의 연관성에 따른 가중치를 줌으로써 질문과 연관성이 큰 맥락 정보가 더 큰 비중을 가질 수 있도록 하였다. 그래프 신경망은 문서들을 이용한 질문 응답 연구[18, 19]에서도 사

용됐는데, 문장 내에 언급된 개체들, 문장들과 문단들을 계층화하여 그래프 형태로 만들어서 GAT[14]를 적용하거나[18], 답변 후보들, 문장들과 문장 내에 언급된 개체들을 연결한 그래프에 GCN[3]의 구조를 적용하였다[19]. 본 논문에서는 GraftNet[2]처럼 GCN[3]의 특성을 이용하여 단순 선형 응집에 의한 이웃 노드들의 맥락 정보를 구하는 한편, BGNN [16]의 특성을 이용하여 쌍 선형 응집에 의한 이웃 노드들 간의 맥락 정보도 구하여 효과적인 다중 홉 추론을 수행할 수 있도록 하였다.

3. 지식 그래프 기반의 질문 응답

3.1 문제 정의

본 논문에서는 지식 베이스로부터 추출된, 질문과 연관된 지식 그래프 상에서 자연어 질문에 대한 답을 추론하는 문제를 다루고 있다.

지식 베이스는 $K = (V, E, R)$ 로 정의할 수 있다. 여기서 K 는 지식 베이스, V 는 개체들의 집합, E 는 트리플 $t = (s, r, o)$ ($s, o \in V, r \in R$)들의 집합, R 는 개체들 사이의 관계 서술자들의 집합이다. Q 는 자연어 질문들의 집합, 질문은 단어들의 집합으로 정의할 수 있다. $q \in Q$ 에 대해, K 에서의 q 와 연관된 지식 그래프 g 를 $g = (q, K)$ 로 정의할 수 있다.

질문 응답(QA) 작업은 g 에서 q 에 대한 답 $v \in V$ 찾기로 정의할 수 있다. 다중 홉 추론은 v 를 찾기 위해 g 에서 q 에 언급된 개체 s 로부터 v 까지의 경로 $d = (t_1, t_2, \dots, t_{|d|})$ 찾기로 정의할 수 있다. d 를 구성하는 t_{n-1} 의 o 는 t_n 의 s 와 같은 개체이다.

3.2 제안 모델

Fig. 2는 본 논문에서 제안하는 모델의 전체 구조도를 나타낸다. 입력으로 질문과 지식 그래프를 갖고, 출력으로 질문에 대한 답변인 개체를 갖는다. 여기서 지식 그래프는 [2]의 전처리(preprocessing)를 이용하여 지식 베이스인 FreeBase로부터 추출한, 질문과 연관된 지식 그래프를 이용한다. 제안하는 모델은 질문 인코딩(Question Encoding), 지식 그래프 추론(Knowledge Graph Reasoning), 답변 예측(Answer Prediction)으로 이루어져 있다.

질문 인코딩은 순환 신경망(Recurrent Neural Network)의 한 종류인 LSTM(Long Short-Term Memory)으로 자연어 질문을 인코딩하여 특징을 추출한다. 추출된 특징은 지식 그래프 추론에서 초기에 사용된 후 매 레이어(layer)마다 질문에 언급된 개체들의 노드 특징들 간의 맥락 정보로 다시 인코딩된다.

지식 그래프 추론은 지식 그래프 상에서 질문에 대한 답변을 찾는 다중 홉 추론을 수행한다. 지식 그래프가 입력으로 들어오면, 각 노드에 대해 쌍 선형 그래프 신경망을 이용하여 이웃 노드들 간의 맥락 정보를 얻는다. 이렇게 얻은 맥락 정보와 질문 인코딩에서 얻은 질문의 특징을 활용하여 노드의 특징을 갱신하는 개체 노드 갱신(Entity node update)을 수행한다.

마지막으로 답변 예측은 지식 그래프 추론을 거쳐 나온 개

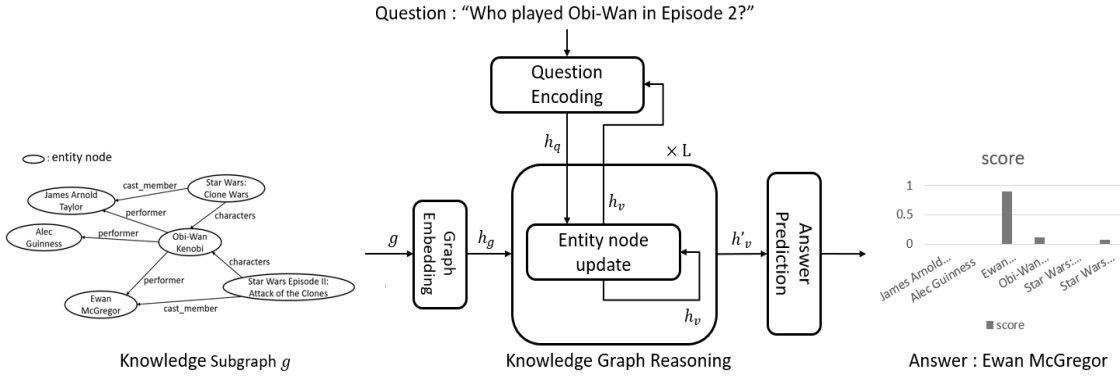


Fig. 2. Overall Architecture of the Proposed Model

제 노드들에 대해 시그모이드(sigmoid) 함수를 적용한다. 그 결과로 가장 높은 점수를 받은 개체 노드가 질문에 대한 답변으로 선택된다.

3.3 지식 그래프 추론

지식 그래프 추론에 앞서 신경망을 학습하기 위해 질문, 개체와 관계 서술자를 임베딩한다. 질문과 관계 서술자는 GloVe 임베딩 값을 이용한다. 개체 노드의 초기 특징값은 GloVe 임베딩 값과 지식 그래프를 미리 학습한 임베딩 값을 이용한다.

지식 그래프 추론 과정에서 다중 홉 추론을 수행하기 위해 쌍 선형 그래프 신경망의 구조를 채용한다. 즉, 기존에 선형 응집만 수행하던 연구들과 다르게 본 논문에서는 쌍 선형 응집(BA)을 추가하여 같이 사용한다. 갱신되는 노드를 목표 노드(target node)라고 할 때, 쌍 선형 응집을 사용하면 목표 노드와 이웃 노드 간의 맥락 정보뿐만 아니라 목표 노드의 이웃 노드들 간의 맥락 정보도 활용할 수 있다. 쌍 선형 응집으로 얻은 맥락 정보는 노드들 간의 요소별 곱(element-wise product)으로 인해 서로 유사한 특징 정보가 강조된 정보를 갖는다. 선형 응집과 쌍 선형 응집을 같이 사용함으로써 노드에 대한 다양한 맥락 정보를 토대로 효과적인 다중 홉 추론을 기대할 수 있다.

쌍 선형 응집은 계산 방식에 따라 BA-A와 BA-T로 나눌 수 있고, 그에 따라 쌍 선형 그래프 신경망의 형태가 BA-A를 사용한 BGNN-A와 BA-T를 사용한 BGNN-T로 나눌 수 있다. Fig. 3은 BA-A와 BA-T를 나타낸 그림이다.

BA-A는 목표 노드와 이웃 노드, 이웃 노드와 이웃 노드끼리 요소별 곱을 수행하여 맥락 정보를 구하는 방식으로 Equation (1)과 같다.

$$BA-A^l(v) = FFN\left(\frac{1}{2} \sum_{v'} \sum_{v''} (h_v^{(l-1)} \odot h_{v'}^{(l-1)}) + \sum_{v''} (h_v^{(l-1)} \odot h_{v''}^{(l-1)})\right) \quad (1)$$

$N(v)$ 은 목표 노드의 이웃 노드들의 집합을 뜻한다. Equation (1)에서 이웃 노드끼리 요소별 곱을 수행하면 같은 계산을 2번 하게 되므로 2로 나눈다.

BA-T는 목표 노드와 이웃 노드만 요소별 곱을 수행하여

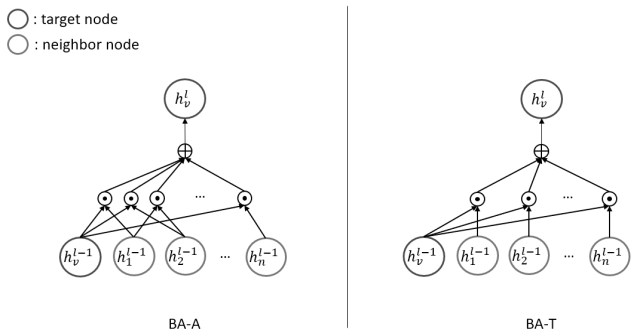


Fig. 3. Two Different Types of Bilinear Aggregation: BA-A and BA-T

맥락 정보를 구하는 방식으로 Equation (2)와 같다.

$$BA-T^l(v) = FFN\left(\sum_{v'} (h_v^{(l-1)} \odot h_{v'}^{(l-1)})\right) \quad (2)$$

본 논문에서는 목표 노드와 이웃 노드 간의 맥락 정보를 활용하는 BA-T를 이용한다. 3.2절 성능 평가 실험에서 BA-A를 사용한 모델과 BA-T를 사용한 모델의 다중 홉 추론 능력을 비교한다.

개체 노드 갱신은 Fig. 4와 같은 구조를 갖는다. 이는 Equation (3)과 같이 나타낼 수 있다.

$$h_v^l = (1 - \alpha) * LA^l(v) + \alpha * BA^l(v) \quad (3)$$

Equation (3)에서 α 는 하이퍼파라미터에 해당한다. α 의 값에 따라 선형 응집(LA)과 쌍 선형 응집(BA)의 적용 비율이 달라진다. LA에 대한 계산은 Equation (4)에서처럼 차례대로 목표 노드의 특징 $h_v^{(l-1)}$, 자연어 질문의 특징 $h_q^{(l-1)}$, 선형 응집에 의한 이웃 노드들의 맥락 정보 $h_N^{(l-1)}$ 을 이용한다. Equation (4)의 || 기호는 연결(concatenation)을 뜻한다.

$$LA^l(v) = FFN([h_v^{(l-1)} || h_q^{(l-1)} || h_N^{(l-1)}]) \quad (4)$$

선형 응집에 의한 이웃 노드들의 맥락 정보 $h_N^{(l-1)}$ 을 계산할 때는 목표 노드에 연결된 간선의 방향을 고려한다. 다중 홉 추론에서 목표 노드로 들어오는 간선과 연결된 이웃 노드들의 특

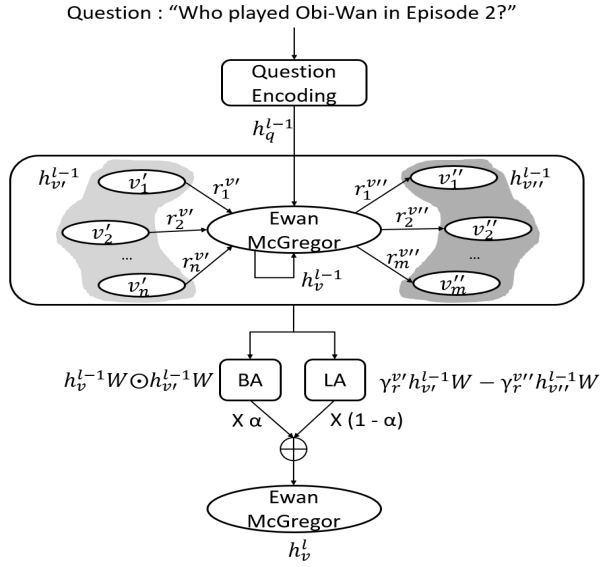


Fig. 4. Entity Node Update

질문은 목표 노드에 대한 추론 경로에 해당하는 정보로 볼 수 있다. 반대로, 목표 노드로부터 나가는 간선과 연결된 이웃 노드들의 특징들은 목표 노드에 대한 추론 경로에 해당하지 않는 정보로 볼 수 있다. 따라서 Equation (5)와 같이 선형 응집에 의한 이웃 노드들의 맥락 정보 $h_N^{(l-1)}$ 을 계산할 때 목표 노드로 들어오는 간선과 연결된 이웃 노드들의 특징들은 더하고 목표 노드로부터 나가는 간선과 연결된 이웃 노드들의 특징들은 빼도록 계산한다. 또한, 목표 노드가 이웃 노드들과 맺는 관계와 질문의 연관성에 따라 다른 주의집중(attention)을 주면서 가중치 합을 계산한다. 이렇게 함으로써 질문과 연관성이 높은 관계를 맺는 이웃 노드의 특징에 더 큰 가중치를 줄 수 있다.

$$h_N^{(l-1)} = FFN(\sum_{r \in R} \sum_{v' \in N_r^+(v)} \gamma_r^{v'} FFN(x_r, h_v^{(l-1)}) - FFN(\sum_{r \in R} \sum_{v'' \in N_r^-(v)} \gamma_r^{v''} FFN(x_r, h_v^{(l-1)}))) \quad (5)$$

Equation (5)에서 $r \in R$ 에 대해, N_r^+ 는 노드로 들어오는 간선과 연결된 이웃 노드들의 집합, N_r^- 는 노드로부터 나가는 간선과 연결된 이웃 노드들의 집합을 뜻한다.

질문 응답을 수행하기 위해서 신경망을 학습할 때 질문의 정보를 이용해야 한다. 따라서 Fig. 4처럼 개체 노드의 특징 $h_v^{(l-1)}$ 을 갱신할 때 자연어 질문의 특징 $h_q^{(l-1)}$ 을 이용한다. 또한, 선형 응집에 의한 이웃 노드들의 맥락 정보 $h_N^{(l-1)}$ 을 계산할 때도 질문을 이용한 주의집중 γ_r^v 을 준다. 자연어 질문의 특징 $h_q^{(l-1)}$ 은 초기값으로 Equation (6)처럼 LSTM의 직전 은닉 상태(hidden state)를 이용한다. 이후 그래프 신경망에서 각 레이어를 거칠 때마다 자연어 질문에 언급된 개체들의 노드들의 맥락 정보로 갱신한다. 자연어 질문과 지식 그래프 상의 개체 노드는 서로 다른 멀티모달(multimodal)이므로 처음에만 자연어 질문의 특징을 LSTM으로 인코딩하고 이후에는 질문에 언급된 개체들의 노드 특징들을 이용한 것이다.

$$h_q^l = \begin{cases} LSTM(q) & l = 0 \\ FFN(\sum_{v \in V_q} h_v^l) & l > 0 \end{cases} \quad (6)$$

Equation (6)의 $v \in V_q$ 는 질문에 언급된 개체의 노드를 뜻한다. Equation (5)에서 자연어 질문과 연관성이 높은 관계에 대해 더 큰 주의집중을 주기 위해 Equation (7)과 같이 주의집중 γ_r^v 을 계산한다.

$$\gamma_r^v = softmax(x_r^T h_q^{(l-1)}) \quad (7)$$

x_r 은 관계 서술자의 임베딩 벡터다. v 가 이웃 노드들과 맺는 관계에 대해 소프트맥스(softmax) 정규화가 행해진다.

BA에 대한 계산은 Equation (8)에서처럼 쌍 선형 응집 BA-T로 구한, 목표 노드와 이웃 노드 간의 맥락 정보를 이용한다. 실험상 모든 이웃 노드에 대해 쌍 선형 응집을 하는 것보다 목표 노드로 들어오는 간선과 연결된 이웃 노드들에 대해서만 하는 것이 더 좋은 성능을 보였다. 따라서 본 논문에서는 목표 노드로 들어오는 간선과 연결된 이웃 노드들에 대해서만 쌍 선형 응집을 수행한다.

$$BA^l(v) = \frac{FFN(\|h_v^{(l-1)}\| \|h_q^{(l-1)}\| FFN(\sum_{v' \in N_r^-(v)} (h_v^{(l-1)} \odot h_{v'}^{(l-1)})))}{FFN(\sum_{v' \in N_r^-(v)} (h_v^{(l-1)} \odot h_{v'}^{(l-1)}))} \quad (8)$$

답변을 추론할 때 모든 노드에 대해 답인지 아닌지 분류하는 이진 분류를 수행한다. 손실 함수로 바이너리 크로스 엔트로피 손실(binary cross entropy loss)을 사용한다.

4. 구현 및 실험

4.1 데이터 집합과 모델 학습

본 논문의 제안 모델은 Ubuntu 16.04 LTS에서 Python 딥러닝 라이브러리인 PyTorch를 이용하여 구현하였다. 모델의 학습 및 평가를 위한 질문 응답 데이터 집합으로는 WebQuestionsSP[4]와 MetaQA[5]를 사용하였다.

WebQuestionsSP는 지식 베이스인 FreeBase를 토대로 생성한 오픈 도메인 질문 응답 데이터 집합이다. 이 데이터 집합은 총 4737개의 자연어 질문과 답변들 중 3098개는 훈련용(training set), 1639개는 검증용(validation set)으로 구성되어 있다. 특히 전체 질문의 약 70%는 단일 홉 추론을 요구하는 단순 질문들(single-hop simple questions), 약 30%는 2-홉 질문들(double-hop complex questions)이다.

MetaQA는 영화와 관련된 질문들로 구성된 데이터 집합인 WikiMovies[1]를 토대로 생성된 데이터 집합이다. MetaQA는 단일 홉의 단순 질문들과 최대 3-홉까지의 복잡 질문들을 포함한 약 40만 개 이상의 질문과 답변들로 구성되어 있다. 본 논문에서는 MetaQA의 단일 홉 질문들과 2-홉 질문들을 실험에 이용하였다.

모델 학습을 위해 쌍 선형 그래프 신경망(BGNN)의 레이어 수(number of layers)는 3, 반복 학습주기(epoch)는 100, 학

습률(learning rate)은 0.0007, 배치 크기(batch size)는 WebQuestionsSP의 경우 10, MetaQA의 경우 20으로 설정하였다. 실험은 64GB의 메인 메모리와 Geforce RTX 2080 Ti 1개를 포함한 컴퓨터 환경에서 수행되었다.

4.2 성능 평가 실험

본 논문에서는 제안하는 모델의 성능 평가 척도로 Hits@K와 F1 score을 채택하였다. Hits@K는 답변이라고 예측한 것들 중 가장 점수가 높은 K개를 선택했을 때 실제 답변이 포함되어 맞춘 비율을 나타낸 것이다.

첫 번째 실험은 제안하는 모델의 지식 그래프 추론에서 쌍 선형 응집의 효과를 보이는 실험이다. 선형 응집만 썼을 때(LA), 쌍 선형 응집만 썼을 때(BA-A와 BA-T) 그리고 둘 다 사용했을 때(BGNN-A와 BGNN-T) 나타난 성능을 비교하였다. Table 1은 첫 번째 실험을 수행하여 측정한 표이다. 이 실험에서는 BGNN의 α 값을 BGNN-A에서는 0.1, BGNN-T에서는 0.3으로 설정하였다.

Table 1에서 보듯이 두 데이터 집합 WebQuestionsSP와 MetaQA에 대해 선형 응집(LA)과 쌍 선형 응집(BA)을 모두 사용하는 BGNN-T가 대부분의 척도에서 가장 우수한 성능을 보였다. 선형 응집(LA)과 쌍 선형 응집(BA)만 비교했을 때는, 선형 응집(LA)의 성능이 쌍 선형 응집(BA)보다 우수한 성능을 나타내었다. WebQuestionsSP에서 Hits@1을 기준으로 보면, 선형 응집과 쌍 선형 응집을 모두 사용하는 BGNN-T와 BGNN-A가 선형 응집(LA)만 사용했을 때보다 각각 약 1.5%와 약 0.3% 더 높은 성능을 보여주었다. 또, BGNN-T를 사용했을 때가 BGNN-A를 사용했을 때보다 약 1.2% 더 높은 성능을 보였다. 특히 MetaQA 데이터 집합의 단일 홉 단순 질문들보다 2-홉 복잡 질문들의 경우, 본 논문에서 제안한 BGNN-T와 BGNN-A가 다른 방식들에 비해 더 큰 격차로 높은 성능을 보여 주었다. 이러한 실험 결과는 복잡 질문 응답에서 제안 방법인 BGNN-T와 BGNN-A의 긍정적 효과를 입증해주는 것으로 볼 수 있다.

두 번째 실험은 복잡 질문들을 포함한 질문-응답 데이터 집합들에 대해 제안 모델과 기존 선형 모델들의 성능을 서로 비교하는 실험이다. Table 2는 WebQuestionsSP 데이터 집합에 대한, Table 3은 MetaQA 데이터 집합에 대한 실험 결과를 각각 나타낸다.

Table 2에서 보듯이 제안하는 모델의 성능이 Hits@1에서 GraftNet[2]보다 약 1.7% 더 높고, F1 score에서는 GraftNet[2]보다 약 1.4% 더 높다. 비록 F1 score에서 GraftNet[2]보다 더 낮은 성능을 보였지만 Hits@1에서 더 높은 성능을 보여, 질문 응답에서 중요한 답변 추론 능력은 더 뛰어난 것을 확인할 수 있었다. 지식 그래프를 (키(key), 값(value)) 쌍의 메모리 슬롯으로 다루는 KV-MemNN[1]은 Hits@1에서 약 46.7%, F1 score에서 약 38.6%를 기록하여 가장 낮은 성능을 보였다.

한편, Table 3은 MetaQA 데이터 집합에 대해 제안 모델과 기존 선형 모델들의 성능을 비교한 실험 결과를 나타낸다.

Table 3에서 보듯이 제안하는 모델의 성능이 단순 질문에 대

Table 1. Performance Comparison with Different Reasoning Methods

method	Hits @1	Hits @5	Hits @10	Hits @20	F1 score
WebquestionsSP					
LA	66.9	79.7	83.3	86.0	58.2
BA-A	19.6	19.9	29.4	39.6	0.6
BA-T	44.7	68.9	76.5	82.2	35.4
BGNN-A	67.2	80.5	83.3	86.1	59.1
BGNN-T	68.4	80.9	83.9	86.0	61.0
MetaQA (1-hop)					
LA	95.7	99.5	99.6	99.7	90.7
BA-A	83.3	97.6	98.7	99.2	65.6
BA-T	94.1	99.3	99.6	99.7	90.8
BGNN-A	95.5	99.5	99.6	99.6	93.8
BGNN-T	95.9	99.6	99.7	99.7	94.7
MetaQA (2-hop)					
LA	97.3	98.6	98.6	98.6	91.5
BA-A	83.4	97.3	98.3	98.6	62.2
BA-T	95.2	98.4	98.6	98.6	86.6
BGNN-A	97.2	98.6	98.6	98.6	92.3
BGNN-T	97.7	98.6	98.6	98.6	92.6

Table 2. Performance Comparison with other Models on WebQuestionsSP

Model	Hits@1	F1 score
KV-MemNN[1]	46.7	38.6
GraftNet[2]	66.7	62.4
Ours	68.4	61.0

Table 3. Performance Comparison with other Models on the MetaQA dataset

Model	1-hop	2-hop
KV-MemNN[1]	96.2	82.7
GraftNet[2]	97.0	94.8
Ours	95.9	97.7

해서는 GraftNet[2]보다 약 1.1% 더 낮지만, 복잡 질문에 대해서는 GraftNet[2]보다 약 2.9% 더 높다. 이런 결과를 보이는 원인은 단순 질문의 경우 대개 질문과 연관된 하나의 트리플로부터 바로 답변을 구할 수 있어 이웃 노드들의 맥락 정보가 필요 없기 때문으로 보인다. 특히 GraftNet[2]의 경우 개체 노드마다 PageRank 점수를 부여해서 질문과 연관성을 계산할 때 이용하기 때문에 단순 질문에서 높은 성능을 보인 것으로 추측한다.

Table 4는 제안 모델의 정성적 평가를 위해 WebQuestionsSP 데이터 집합에 대해 제안 모델이 수행한 결과들 중 선정한 몇 가지 복잡 질문 응답의 예를 나타낸다. Table 4에서 보듯이, 질문 1, 2, 3과 같이 다중 홉 추론을 요구하는 많은 복잡 질문들에 대해, 제안 모델은 정답(ground truth answer)과 일치하는 예측 답변(predicted answer)들을 생성한 것을 확인할

Table 4. Question Answering Examples Resulted by the Proposed Model

Question	Ground Truth Answer	Predicted Answer
1. Who played Obi-Wan in Episode 2?	Ewan McGregor	Ewan McGregor
2. What club does Cristiano Ronaldo play for in 2010?	Portugal national football team, Real Madrid C.F.	Real Madrid C.F.
3. What voices does Seth MacFarlane play on Family Guy?	Kevin Swanson , Tom Tucker, Brian Griffin, etc.	Kevin Swanson
4. What was the last movie Sean Connery was in?	Sir Billi	James Bond
5. When was the last time the Dodgers won a championship?	1988 World Series	1981 World Series

수 있다. 예를 들면, 질문 1의 경우 영화 “Star Wars Episode II: Attack of the Clones”에 나오는 “Obi-Wan Kenobi”를 연기한 사람이 누구인지 묻는 예이다. 이 질문에 대한 정답인 “Ewan McGregor”를 지식 그래프 상에서 찾아내기 위해서는, 질문에서 언급된 개체 노드들이면서 정답 개체 노드와 관계를 맺고 있는 이웃 노드들인 “Obi-Wan Kenobi” 및 “Star Wars Episode II: Attack of the Clones”의 맥락 정보를 추론에 활용할 필요가 있다. 특히 “Obi-Wan Kenobi”가 “Star Wars Episode II: Attack of the Clones”에 등장한다는 사실 또한 질문과 연관이 있으므로, 지식 그래프 상에서 “Ewan McGregor”를 추론할 때 이웃 개체 노드들의 맥락 정보뿐만 아니라 이웃 노드들끼리의 맥락 정보도 활용되면 더 효과적일 수 있다. 따라서 제안 모델에서 채용하고 있는 쌍 선형 그래프 신경망(BGNN)이 해당 질문들의 정답을 추론해내는데 큰 도움을 준 것으로 판단한다.

하지만, 질문 4, 5와 같이 시간적 비교(temporal comparison)를 포함하는 일부 복잡 질문들에 대해서는 제안 모델이 정답을 잘 추론하지 못하는 경우가 발생하기도 하였다. 이를테면, 질문 4는 “Sean Connery”가 출연한 마지막 영화가 무엇인지 묻는 예이다. 이 질문 역시 다중 홉 추론을 요구하는 복잡 질문으로서, “Sean Connery”가 출연한 영화들을 찾고, 해당 영화들 중에 가장 최근에 출시된 영화를 찾아야 하는 질문이다. 제안 모델은 이 질문에 대해 잘못된 답변을 예측하였는데, 그 이유는 지식 그래프 상에서 “Sean Connery”가 출연한 영화들을 찾을 수는 있었지만, 그런 영화들 중에서 가장 최근에 출시된 영화가 어떤 것인지 제대로 판단하지 못하였기 때문이다. 이러한 제안 모델의 한계점은 질문 5의 경우에도 마찬가지로 오답을 예측하는 결과를 만들었다. 따라서 이와 같이 수치 비교 능력을 요구하는 다양한 복잡 질문들에 대해서도 답변 추론 능력을 더욱 향상시킬 수 있도록, 향후 제안 모델의 개선이 이루어져야 할 것으로 판단한다.

5. 결 론

본 논문에서는 지식 그래프 위에서 다중 홉 추론을 요구하는 복잡한 자연어 질문에 효과적으로 답변을 추론할 수 있는 심층 신경망 모델을 제안하였다. 제안 모델에서는 지식 그래프 노드들 간의 양방향 특징 정보 전파 기능을 포함하고 있을 뿐만 아니라, 두 이웃 노드 쌍 간의 맥락 정보도 활용할 수 있는 쌍 선형

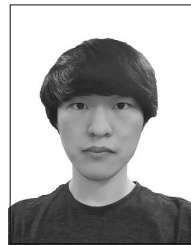
그래프 신경망(BGNN)을 새롭게 채용하고 있다. Freebase 오픈 도메인 지식 베이스와 WebQuestionsSP 및 MetaQA 질문 응답 벤치마크 데이터 집합을 이용한 다양한 실험들을 통해, 제안 모델의 효과와 우수성을 확인할 수 있었다. 향후 연구에서는 시간적 비교, 수치적 비교들이 포함된 복잡 질문들에 대해서도 정답률을 높일 수 있도록, 현재의 제안 모델을 추가 개선해볼 계획이다.

References

- [1] A. Miller, A. Fisch, J. Dodge, A. Karimi, A. Bordes and J. Weston, “Key-value memory networks for directly reading documents,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, pp.1400-1409, 2016.
- [2] H. Sun, B. Dhingra, M. Zaheer, K. Mazaitis, R. Salakhutdinov and W. Cohen, “Open domain question answering using early fusion of knowledge bases and text,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, pp.4231-4242, 2018.
- [3] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proceedings of the 5th International Conference on Learning Representations*, Toulon, 2017.
- [4] W. Yih, M. Richardson, C. Meek, M. Chang and J. Suh, “The value of semantic parse labeling for knowledge base question answering,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, pp.201-206, 2016.
- [5] Y. Zhang, H. Dai, Z. Kozareva, A. J. Smola and L. Song, “Variational reasoning for question answering with knowledge graph,” in *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, pp.6069-6076, 2018.
- [6] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. V. Kleef, S. Auer and C. Bizer, “DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia,” *Semantic Web*, Vol.6, No.2, pp.167-195, 2015.

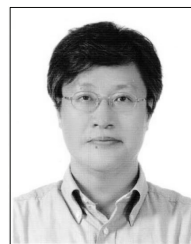
- [7] J. Berant, A. Chou, R. Frostig and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, pp.1533-1544, 2013.
- [8] W. Yih, M. Chang, X. He and J. Gao, "Semantic parsing via staged query graph generation: Question answering with knowledge base," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, pp.1321-1331, 2015.
- [9] Y. Yang and M. Chang, "S-MART: novel tree-based structured learning algorithms applied to tweet entity linking," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, pp.504-513, 2015.
- [10] S. Sukhbaatar, A. Szlam, J. Weston and R. Fergus, "End-to-end memory networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, pp.2440-2448, 2015.
- [11] C. Liang, J. Berant, Q. Le, K. D. Forbus and N. Lao, "Neural symbolic machines: Learning semantic parsers on freebase with weak supervision," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, pp.23-33, 2017.
- [12] R. Das, S. Dhuliawala, M. Zaheer, L. Vilnis, I. Durugkar, A. Krishnamurthy, A. Smola and A. McCallum, "Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning," in *Proceedings of the 6th International Conference on Learning Representations*, Vancouver, 2018.
- [13] Y. Li, D. Tarlow, M. Brockschmidt and R. Zemel, "Gated graph sequence neural networks," in *Proceedings of the 4th International Conference on Learning Representations*, San Juan, 2016.
- [14] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, "Graph attention networks," in *Proceedings of the 6th International Conference on Learning Representations*, Vancouver, 2018.
- [15] W. L. Hamilton, R. Ying and J. Leskovec, "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, pp.1025-1035, 2017.
- [16] H. Zhu, F. Feng, X. He, X. Wang, Y. Li, K. Zheng and Y. Zhang, "Bilinear graph neural network with neighbor interactions," arXiv preprint arXiv:2002.03575, 2020.
- [17] D. Sorokin and I. Gurevych, "Modeling semantics with gated graph neural networks for knowledge base question answering," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, pp.3306-3317, 2018.
- [18] Y. Fang, S. Sun, Z. Gan, R. Pillai, S. Wang and J. Liu, "Hierarchical graph network for multi-hop question answering," arXiv preprint arXiv:1911.03631, 2019.
- [19] M. Tu, G. Wang, J. Huang, Y. Tang, X. He and B. Zhou, "Multi-hop reading comprehension across multiple documents by reasoning over heterogeneous graphs," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, pp.2704-2713, 2019.

이 상 의



<https://orcid.org/0000-0001-9072-8094>
 e-mail : rmlrml125@kyonggi.ac.kr
 2019년 경기대학교 컴퓨터과학과(학사)
 2020년 ~ 현 재 경기대학교 컴퓨터과학과 석사과정
 관심분야 : 인공지능, 자연어처리, 로봇지능

김 인 철



<https://orcid.org/0000-0002-5754-133X>
 e-mail : kic@kyonggi.ac.kr
 1985년 서울대학교 수학과(이학사)
 1987년 서울대학교 전산과학과(이학석사)
 1995년 서울대학교 전산과학과(이학박사)
 1996년 ~ 현 재 경기대학교 컴퓨터과학과 교수
 관심분야 : 인공지능, 기계학습, 로봇지능