

Detection of Similar Answers to Avoid Duplicate Question in Retrieval-based Automatic Question Generation

Yong-Seok Choi[†] · Kong Joo Lee^{††}

ABSTRACT

In this paper, we propose a method to find the most similar answer to the user's response from the question-answer database in order to avoid generating a redundant question in retrieval-based automatic question generation system. As a question of the most similar answer to user's response may already be known to the user, the question should be removed from a set of question candidates. A similarity detector calculates a similarity between two answers by utilizing the same words, paraphrases, and sentential meanings. Paraphrases can be acquired by building a phrase table used in a statistical machine translation. A sentential meaning's similarity of two answers is calculated by an attention-based convolutional neural network. We evaluate the accuracy of the similarity detector on an evaluation set with 100 answers, and can get the 71% Mean Reciprocal Rank (MRR) score.

Keywords : Question-Answer Database, Automatic Question Generation, Similar Answer Detection

검색 기반의 질문생성에서 중복 방지를 위한 유사 응답 검출

최 용 석[†] · 이 공 주^{††}

요 약

본 연구는 검색 기반의 질문 자동 생성 시스템에서 사용자가 이미 답변한 내용을 재질문하지 않도록 사용자의 응답과 유사도가 높은 응답을 질문-데이터베이스에서 찾는 방법을 제안한다. 유사도가 높게 검출된 응답의 질문은 이미 사용자가 아는 내용일 확률이 높기 때문에 질문 후보군에서 제거한다. 유사 응답 검출에는 두 응답간의 동일 단어, 바꿔쓰기 표현, 문장 내용을 모두 사용하였다. 바꿔쓰기 표현은 통계기반의 기계번역에서 사용하는 구절 테이블을 사용하여 구축하였다. 문장 내용은 두 문장을 주의-기반 컨볼루션 신경망으로 압축하여 유사도를 계산하였다. 평가를 위해 구축한 100개의 평가 응답에 질문-응답 데이터베이스로부터 가장 유사한 응답을 추출해서 얻은 결과는 MRR값 71%의 성능을 보였다.

키워드 : 질문-응답 데이터베이스, 질문 자동 생성, 유사 응답 검출

1. 서 론

질문 자동 생성은 다양한 입력 형태 - 원문 텍스트, 데이터베이스, 의미표현 -로부터 기계가 질문을 자동으로 생성해 내는 것이다. 생성된 질문을 이용하여 학습자의 학습 정도를 평가해 볼 수 있기 때문에 교육 분야에서 학습자 평가에 사용될 수 있다[1].

일반적으로 질문 자동 생성은 두 단계를 거쳐 이루어진다.

첫째는 무엇을 질문할 것인지 질문의 내용/주제를 결정하는 것이다. 일단 질문의 내용/주제가 결정되면 두 번째는 질문의 내용을 기반으로 실제 질문 문장을 만드는 것이다.

본 연구에서는 질문 자동 생성의 두 과정 - 질문의 내용 결정과 질문 문장 생성 - 을 하나로 결합하여 처리하고자 한다. 일반적으로 단문 대화(short-text conversation) 시스템 분야에서는 질문 데이터베이스로부터 가장 적절한 응답을 검색하여 사용자의 질문에 대한 답변으로 제시한다[2]. 본 연구에서는 단문 대화 시스템 분야에서 사용되는 검색 기반 방식을 활용하여 질문 데이터베이스로부터 가장 적절한 질문을 추출한다. 이 방법은 항상 문법에 맞는 질문을 생성해 낼 수 있다는 장점이 있다. 또한 질문 데이터베이스로부터 다양한 형식의 질문을 생성해 낼 수 있다는 장점도 있다. 그러나 질문을 추출할 때에는 반드시 학습자의 그 이전 응답을 고려하

※ 이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No.NRF-2015R1C1A2A01051685).

† 준 회 원 : 충남대학교 전자전파정보통신공학과 박사과정

†† 중 심 회 원 : 충남대학교 전자전파정보통신공학과 교수

Manuscript Received : October 8, 2018

First Revision : November 27, 2018

Accepted : December 3, 2018

* Corresponding Author : Kong Joo Lee(kjoollee@cnu.ac.kr)

(System) Q1: What is daemon thread?	
U1: Daemon thread is a low priority thread, which runs in the back ground. Garbage collection is done in daemon.	U2: Daemon thread can be created by setDaemon method.
(System) Candidates of next Question:	
(1) What is a parent thread of daemon?	
(2) Which method is used to create the daemon thread?	
(3) What is the example of daemon thread?	
(4) What is a thread?	
(5) What about the daemon threads?	

Fig. 1. Example of System's Question and User's Answers

여 다음 질문을 선택해야 한다.

Fig. 1은 시스템의 질문 Q1에 대해 서로 다른 학습자 U1과 U2의 응답과 시스템의 다음 질문에 대한 후보 예제이다. 학습자 U2의 응답은 질문에 벗어난(off-topic) 답변이다. 이 때 시스템이 질문 데이터베이스로부터 추출한 다음 질문의 후보가 (1)~(5)로 주어졌을 때, 우선 후보 질문 (4), (5)는 학습자의 응답과 무관하게 다음 질문으로 부적절하다. 왜냐하면 질문 Q1과 동일 질문이거나 유사 질문이기 때문이다¹⁾. 이와 같은 유사 질문은 질문 데이터베이스 구축 시, 자동 군집화(clustering)를 통해 우선적으로 유사 질문임을 인식할 수 있다.

후보 질문 (1)은 두 사용자 모두에게 다음 질문으로 적합하다. 그러나 후보 질문 (3)은 학습자 U2에게는 다음 질문으로 적합한 반면 학습자 U1에게는 적합하지 않다. 왜냐하면 학습자 U1의 답변에 이미 질문 (3)의 답변이 (일부) 들어가 있기 때문이다. 또한 후보 질문 (2)은 학습자 U1의 다음 질문으로 적합한 반면 학습자 U2에게는 적절하지 않다. U2의 답변에 질문 (2)의 답변이 포함되어 있기 때문이다. 본 연구는 사용자의 응답이 질문에 적합한 답변인지에 대한 평가는 고려하지 않는다. 비록 사용자의 응답이 질문에 대한 정답이 아니더라도 사용자의 응답을 고려한 다음 질문 선택이 가능하다.

예제에서 본 바와 같이 질문 자동 생성의 경우 사용자의 이전 답변을 고려하여 다음 질문을 생성해야 한다. 그렇지 않을 경우 학습자가 이미 답변한 내용을 다시 질문하게 되는 경우가 발생하기 때문이다. 또한 이러한 문제는 검색 기반의 질문 자동 생성뿐만 아니라 다른 방식의 질문 자동 생성의 경우에서도 고려해야 하는 문제이다. 본 연구는 이와 같은 상황에서 학습자의 기 응답을 고려하여 중복된 질문을 다시 생성하지 않도록 하는 방법을 제안하고자 한다.

사용자의 응답을 고려한 질문을 생성하기 위해 질문 데이

1) 질문에 벗어난 답변을 제시한 U2에게는 같은 내용을 제질문할 수 있다.

터베이스는 질문 Q와 응답 A의 쌍으로 구축해 놓는다. 시스템의 i번째 질문 Q_i 에 대해 학습자의 응답 A_u 가 주어졌을 때 다음 질문 Q_{i+1} 의 후보군을 $Q_{cand} = \{(Q_{c1}, A_{c1}), (Q_{c2}, A_{c2}), \dots, (Q_{cn}, A_{cn})\}$ 이라고 하자. 본 연구에서는 응답 A_u 와 유사도가 일정 값(threshold) 이상인 문장을 유사 응답으로 간주한다. A_u 과 유사 응답으로 인식되는 A_{cj} 의 (Q_{cj}, A_{cj}) 를 Q_{cand} 으로부터 제거한다. 질문 Q_{cj} 를 다음 질문으로 제공했을 경우, 이 질문에 대한 응답을 사용자가 이미 했거나 알고 있을 확률이 높기 때문이다.

본 연구에서는 두 응답간의 유사도를 단어 일치도, 단어 임베딩 사이의 거리, 구절 단위의 바뀌쓰기 정보, 컨볼루션 네트워크를 사용한 문장 단위의 자질 등을 함께 사용하여 측정해 본다.

본 논문의 구성은 다음과 같다. 2장에서 관련연구를 살펴보고 유사 응답 인식을 3장에서 기술한다. 실험 환경 및 실험 결과를 4장에서 살펴보고 결론을 맺는다.

2. 관련 연구

2.1 질문 자동 생성

가장 널리 사용되는 질문 자동 생성 방법[3]은 질문 내용이 담긴 평서문을 의문문으로 변환하거나 또는 질문할 내용을 질문 템플릿에 채워 넣음으로써 질문 문장을 완성하는 것이다. [4]의 연구에서는 문서로부터 질문하고자 하는 주제를 선택한 후, 질문 템플릿을 이용하여 질문을 생성하였다. 예를 들어, 질문의 주제가 '정의(definition)'이면 'What is meant by X?'라는 질문 템플릿을 사용하여 질문을 생성하였다. 이 때 사용한 템플릿은 수동으로 구축되었다.

최근의 연구에는 딥러닝 기법을 이용하여 평서문을 의문문으로 자동 변환하고 이를 질문으로 생성하였다. 70,000개 이상의 문장-질문 쌍으로 구성된 SQuAD 학습데이터[5]를 이용하여 seq2seq 모델을 학습시켰고, 기존의 규칙 기반의 질문 생성 시스템에 비해 좋은 성능을 얻을 수 있었다[6].

[7]의 연구에서는 대량의 데이터를 쉽게 모을 수 있는 Community QA (CQA) 사이트에서 학습데이터를 수집하여 사용하였다. 우선 질문을 생성하기 위한 질문패턴(Question Pattern)을 CQA사이트에서 수집한 질문-응답 쌍으로부터 자동으로 추출하였다. 자동 수집한 질문패턴은 질문과 관련 있는 토픽 단어가 함께 제시된다. 예를 들어 'co-founded by' 가 문장에 있을 때, 사용 가능한 질문패턴은 "who founded # ?"이다. 이렇게 자동으로 추출한 질문패턴을 학습데이터로 이용하여 딥러닝 모델의 일종인 Recurrent Neural Network(RNN) 모델을 학습시켰다. RNN 모델의 입력은 평서문의 문장이고 가장 적절한 질문패턴이 출력되도록 학습한다. 질문을 생성하는 과정은 첫째, 문서로부터 질문주제를 선택한다. 둘째, RNN 모델을 이용하여 문서에 대한 가장 적절한 질문패턴을 추출한다. 셋째, 질문패턴의 #에 질문주제를 채워 넣어 질문을 완성한다. 여러 개의 질문이 생성된 경우에는 각 단계에서 사용된 점수 값과 다른 속성 정보를 활용하여 질문을 순위화시켰다.

2.2 유사 문장 검출

두 문장의 유사도를 측정하는 것은 자연언어처리 전반에서 중요한 기본 요소 중 하나이기 때문에 여러 연구자들에 의해 오랫동안 연구되어 온 분야 중 하나이다[8]. 일반적으로 문장 사이의 유사도는 단어, 구문, 의미의 세 가지 정보를 사용하여 측정한다. 단어 정보는 두 문장 사이에 동일한 단어가 얼마나 사용되었는지와 같은 직접적인 정보로부터 TF/IDF, 그리고 단어를 Word2Vec과 같은 벡터로 표현한 후 벡터간 거리 계산 등을 사용한다[9, 10]. 이와 같은 단어 정보는 개별 단어만을 평가하기 때문에 문장에 나타난 단어 간 순서나 단어 사이의 구문 정보는 배제된다[11]. 구문 정보는 문장을 구문 분석하여 사용된 문법 규칙이나 문법 패턴 등을 유사도 측정에 사용한다[12].

두 문장의 의미 유사도는 문장을 고정된 벡터로 표현하고 두 벡터간의 거리를 계산하는 다양한 방법을 활용하여 수치화할 수 있다. 딥러닝 모델의 발전에 따라 다양한 신경망을 이용하여 두 문장 간의 유사도를 측정하는 연구들이 진행되고 있다. [8]의 연구에서는 LSTM 모델을 사용하여 두 문장 간의 유사도를 학습시켰다. 두 문장을 각각 LSTM 모델을 이용하여 고정된 크기의 벡터로 표현한 후, 두 벡터간의 Manhattan 거리를 두 문장의 유사도 값으로 사용하였다. SICK (Sentence Involving Compositional Knowledge) 데이터[13]로 학습하여 기존의 방법에 비해 좋은 결과를 얻을 수 있었다. [14]의 연구에서는 문장의 구문 구조까지 고려하기 위하여 Tree-LSTM을 제안하였다. 구문 분석이 된 문장을 Tree-LSTM을 이용하여 고정된 벡터로 표현한다. Tree-LSTM에서는 입력된 트리 구조를 따라 자식 노드의 hidden state를 합하여 현재 노드를 구성하였다. 이렇게 구성된 문장 표현을 이용하여 문장 사이의 유사도를 계산하였다. 본 연구에서는 단어 정보, 구절 정보, 문장의 의미 정보를 모두 사용하여 두 문장 간의 유사도 계산 함수에 적용해 본다.

3. 유사 응답 검출

3.1 질문 자동 생성 시스템 구성

질문 자동 생성 시스템의 전체 구성은 Fig. 2와 같다. 시스템의 i 번째 질문 Q_i 에 대한 학습자 응답 A_u 가 주어지면 Q_i 와 A_u 를 바탕으로 질문-응답 데이터베이스로부터 다음번 질문 Q_{i+1} 의 후보를 추출한다. 후보 질문을 추출하는 방법은 이전 질문인 Q_i 와 동일 분야에 대한 질문을 생성할 수도 있고 또는 새로운 분야에 대한 질문을 추출할 수도 있다.

추출된 후보 질문 중, 학습자의 응답 A_u 와 유사 응답을 검출하고 이에 해당하는 질문을 유사 질문으로 걸러낸다(Question Filter). 여과된 질문을 질문 로그(Question Log)에 남김으로써 다음번에도 역시 같은 질문이 반복되지 않도록 한다.

3.2 학습자 응답에 따른 유사 응답 인식

시스템의 i 번째 질문 Q_i 에 대해 학습자의 응답 A_u 가 주어졌을 때 질문-응답 데이터베이스로부터 추출한 다음 질문 Q_{i+1}

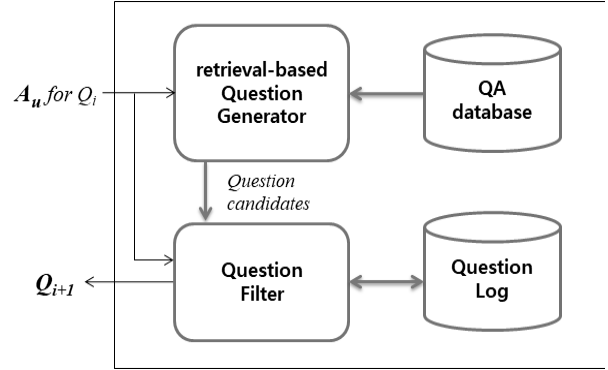


Fig. 2. The Overview of the Automatic Question Generation System

i 의 후보군을 $Q_{cand} = \{(Q_{c1}, A_{c1}), (Q_{c2}, A_{c2}), \dots, (Q_{cn}, A_{cn})\}$ 이라 하자. 본 연구에서 사용하는 질문-응답 데이터베이스에는 각각의 질문 Q 에 대해 응답 A 가 연결되어 있기 때문에 후보 집합 Q_{cand} 역시 (Q, A) 의 짝으로 표현한다.

본 연구에서는 질문 Q_i 에 대한 학습자의 응답 A_u 와 후보 질문 집합 Q_{cand} 각각의 응답과의 유사도를 계산하기 위해 [15]에서 제안한 방법 중 단어, 구절, 문장 단위의 자질을 사용하였다. 학습자가 제시한 응답 A_u 와 후보 질문 집합 Q_{cand} 의 A_c 간의 유사도를 계산하기 위해 Equation (1)을 사용한다.

$$\text{sim}(A_u, A_c) = \sum_k (\lambda_k \cdot f_k(A_u, A_c)) \quad (1)$$

f_k 는 단어, 구문, 문장 단위에서 학습자의 응답 A_u 와 응답 A_c 사이에 유사도를 계산하는 함수이다. λ_k 는 자질 함수들의 가중치를 반영하는 값이다.

본 연구에서는 λ_k 를 모두 1.0으로 사용하였다. 유사도 값 $\text{sim}(A_u, A_c) > \text{threshold}$ 일 때, 응답 후보 (Q_c, A_c) 를 Q_{cand} 에서 제거한다. 다음에서는 각 단위의 유사도 함수 f_k 에 대해 살펴본다.

1) 단어 단위 자질의 유사도 함수

단어 단위의 자질은 두 가지 정보를 사용한다. 첫째, 학습자 응답과 응답 후보 간의 기능을 제외한 자카드 유사도(Jaccard similarity)[16]이다(f_{wj}). 둘째는 학습자 응답과 응답 후보를 벡터로 표현한 후 코사인 유사도 값을 계산한다(f_{wemb})[15]. 이 벡터는 응답 문장에 포함된 단어(기능어 제외)들의 단어 임베딩의 평균을 취한 값이다.

2) 구절 단위 바꿔쓰기 자질의 유사도 함수

두 문장이 동일한 단어를 사용하지 않았더라도 유사한 의미를 나타낼 수 있다. 이와 같이 동일한 의미를 다른 표현으로 쓰는 것을 바꿔쓰기(paraphrase)[17]라고 한다. (예 1)의 (a), (b) 문장은 바꿔쓰기의 예제이다. 유사한 의미를 담고 있는 두 문장에서 (a)의 “a collection of”를 (b)의 “a set of”로 바꿔쓸 수 있다.

(예 1)

- (a) A graph is a collection of objects called vertices where some pairs of objects are connected by links called edges.
- (b) A graph is a set of objects where some pairs of objects are connected by links.

본 연구에서는 두 문장의 유사도 측정에 구문 단위의 바꿔쓰기 자질을 사용한다. 바꿔쓰기 표현들을 찾아내기 위해 문장단위로 정렬된 한-영 코퍼스를 이용하여 통계기반 기계번역(statistical machine translation:SMT)의 구문 테이블(phrase table)[18]을 구축하고 이를 이용하여 바꿔쓰기 표현들을 추출한다.

Table 1. Example of the Paraphrase Table

s_i	s_j	score($s_i; s_j$)
speech service option	voice service option	0.978723
interior core	inner core	0.949153
be in relation	associate with	0.940260
aggregator	collector	0.927419
diverse composition	various component	0.923077
higher likelihood	high probability	0.923077
arbitrary number of	any number of	0.918033
monitoring apparatus	monitoring device	0.909091
be prone to	tend to	0.872570
gratitude to you	thank you	0.853846

Table 1은 바꿔쓰기 테이블의 일부 결과를 나타낸 것이다. 바꿔쓰기 표현을 이용하여 두 문장 사이의 유사도를 계산한다. 구체적인 방법은 Fig. 3과 같다.

3) 문장 단위 자질의 유사도 함수

응답과 응답 후보 간의 주목 기반 컨볼루션 신경망(Attention-Based Convolution Neural Network; ABCNN)[19]을 활용하여 문장 사이의 유사도 점수를 계산한다. 주목(attention)을 사용하는 것은 응답 문장의 단어와 후보 응답 문장의 단어 사이의 가중치를 계산하기 위함이다.

문장을 구성하고 있는 단어 x 는 d 크기의 단어 임베딩 벡터로 표현하고 이를 컨볼루션 신경망의 입력으로 사용한다(Equation (2)). Equation (3)은 컨볼루션 신경망의 수식이며 h 는 컨볼루션의 필터 크기이다. 컨볼루션 신경망에서 계산된 결과를 입력으로 평균 풀링을 계산한다. Equation (4)는 다음 계층의 컨볼루션 신경망의 입력과 맞추기 위한 평균 풀링 연산이다. Equation (5)는 컨볼루션 신경망의 마지막 계층에서 문장을 하나의 벡터로 표현하기 위한 평균 풀링 연산이다.

$$x = [v_1, v_2, v_3, \dots, v_s] \quad (2)$$

$$v_i \in R^d, x \in R^{s \times d}$$

(0) 정렬 코퍼스로부터 구문테이블(phrase table)을 구축한다.

$$PT = \{ \langle s_i, t, p(t_i|s_i), p(s_i|t_i) \rangle \}$$

(1) 구문테이블 정보를 이용하여 바꿔쓰기 테이블(paraphrase table)을 구축한다.

score값을 다음과 같이 정의한다.

$$PP = \{ \langle s_i, s_j, score(s_j; s_i) \rangle \}$$

$$score(s_j; s_i) = \sum_t p(t|s_i) \times p(s_j|t)$$

(2) 두 응답 문장 A_u 와 A_c 사이의 바꿔쓰기 기반 유사도 h_{PP} 는 다음과 같이 계산한다.

$$h_{PP}(A_u, A_c) = \frac{\sum_{n=1}^N \frac{\sum_{j=0}^{|A_c|-n} Count_{PP}(A_c[j:j+n-1], A_u)}{|A_c|-n+1}}{N}$$

$A_c[j:j+n-1]$ 은 문장 A_c 에서 j 번째 단어부터 $j+n-1$ 번째 단어까지의 연속된 단어열을 의미한다. N 은 $N-gram$ 값이다. $Count_{PP}(A_c[j:j+n-1], A_u)$ 는 다음과 같이 계산한다.

(2.1) 만약 $A_c[j:j+n-1] \in A_u$ 이면

$$Count_{PP}(A_c[j:j+n-1], A_u) = 1$$

(2.2) (2.1)이 아니고

만약, $\langle A_c[j:j+n-1], s, score(s; A_c[j:j+n-1]) \rangle \in PP$

그리고 $A_c[j:j+n-1]$ 의 paraphrase $s \in A_u$ 이면

$$Count_{PP}(A_c[j:j+n-1], A_u) = score(s; A_c[j:j+n-1])$$

(2.3) (2.1)과 (2.2)를 모두 만족하지 않는 경우라면

$$Count_{PP}(A_c[j:j+n-1], A_u) = 0$$

Fig. 3. Similarity Function Between Sentences Using Paraphrases [15]

$$c_i = f(W \cdot x_{i:i+h} + b) \quad (3)$$

$$W \in R^{h \times d}; b \in R; f: \tanh$$

$$x^{(n)} = AVGPool_{wp}(c_1^{(n-1)}, c_2^{(n-1)}, \dots, c_{n+h-1}^{(n-1)}) \quad (4)$$

$$CNN_x = AVGPool_{ap}(c_1^{(n)}, c_2^{(n)}, \dots, c_{n+h-1}^{(n)}) \quad (5)$$

마지막 계층은 두 문장 A_u 와 A_c 사이의 유사도를 계산하기 위한 로지스틱 회귀(Logistic Regressing) 계층으로 선형 계층이다. 컨볼루션 신경망으로부터 출력된 문장을 표현한 벡터를 연결(concatenation)하고 추가로 자질을 포함하여 선형 계층(Linear Layer)의 입력으로 사용한다. 이 선형 계층의 출력이 두 문장 간의 유사도이다(Equation (6)). 추가로 포함 한 자질은 응답 문장의 길이(A_u-len)와 후보 응답 문장의 길이

Table 2. Examples of Quora Datasets

1	<p>Q: Which is the best, JavaScript or Bootstrap?</p> <p>A1: Javascript is a programming language. Bootstrap is (in the web dev context) a CSS framework. You are comparing cars with oranges.</p> <p>A2: JavaScript is the scripting language used for multiple purposes for web applications and sometimes in desktop applications as well. Whereas Bootstrap is a framework for making responsive web sites.</p> <p>A3: Well, first of all both are different things for different purpose. We use JavaScript as client side scripting language. And bootstrap is used to make website responsive etc.</p>
2	<p>Q: What are the similarities in a database and data warehouse?</p> <p>A1: Data warehouse or DWH is an enterprise scale database. Concept of DWH is mainly used for analytical processing while database can be considered as a part of transactional processing.</p> <p>A2: The key is that databases are primarily used to store data while data warehouses are used for data analysis.</p>

(A_{c-len}) 그리고 두 문장 사이의 공통된 단어(기능어 제외)에 대한 idf 값이다.

$$Output = W_l \cdot [CNN_{Au}; CNN_{Ac}; f_{Au-len}; f_{Ac-len}; f_{idf}] + b_l \quad (6)$$

$$Output \in R^2, W_l \in R^{2 \times (d1*2 + len(feature))}, b_l \in R^2$$

[19] 연구에서는 컨볼루션 신경망의 입력으로 응답 문장의 단어와 후보 응답 문장의 단어 사이의 가중치를 계산한 주목 (attention) 값을 함께 반영한다. Fig. 4는 [19]의 연구에서 제안한 주목 기반의 가중치를 반영하는 방법이다. 주목 행렬(attention matrix)은 두 문장 A_u 와 A_c 의 단어 사이의 가중치를 유클리드 거리(Euclidean distance)로 계산한다. Equation (7)은 주목 테이블을 계산하는 수식이다.

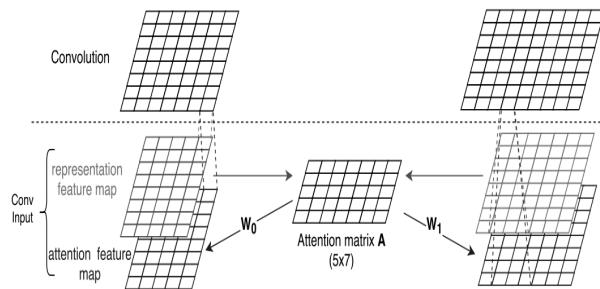


Fig. 4. Attention-Based Convolutional Neural Networks Structure [19]

$$Atten_{i,j} = euc(F_{Au,r}[i,:], F_{Ac,r}[j,:]) \quad (7)$$

$$F_{Au,r}, F_{Ac,r} \in R^{s \times d}$$

주목 테이블에 가중치(W_0, W_1)를 곱하여 주목 자질 맵(Attention Feature Map)을 만들고 입력 문장과 주목 자질 맵을 함께 컨볼루션 신경망의 입력으로 사용한다(Fig. 4).

쿼라(Quora)²⁾는 사용자 커뮤니티에 의해 운영되는 질의

응답 웹사이트이다. 주목 기반 컨볼루션 모델을 학습시키기 위해서 쿼라 사이트로부터 질문과 응답을 수집하였다. 하나의 질문에 여러 개의 답변이 있을 때, 이 답변들을 서로 유사 응답으로 간주하여 학습 데이터를 구축하였다. Table 2는 쿼라 코퍼스에서 하나의 질문에 대해 두 개 이상의 답변을 갖고 있는 예제들이다.

주목 기반 컨볼루션 신경망 모델의 학습을 위해서는 긍정 데이터뿐만 아니라 부정 데이터가 필요하다. 즉 유사 응답쌍 이외에 유사하지 않은 응답쌍이 필요하다. 이를 위해 쿼라 코퍼스에서 유사한 응답쌍을 만든 후, 한 문장을 임의의 다른 문장으로 교체하여 유사하지 않은 응답쌍을 만들었다.

4. 실험 환경 및 결과

4.1 IT 분야의 질문-응답 데이터베이스 구축

IT 분야의 질문-응답 데이터베이스 구축을 위해 다음 4개의 사이트로부터 질문과 응답 쌍을 추출하였다.

<https://www.javatpoint.com/interview-questions-and-answers>

<https://www.tutorialspoint.com/tutorialslibrary.htm>

<https://career.guru99.com/top-100-core-java-interview-questions/>

<https://www.javacodegeeks.com/2014/04/java-interview-questions-and-answers.html>

Table 3. Database of Questions and Answers

Domain	Number of Question-Answers
C/C++/C#/Java Programming language	885
Other Programming languages	1,155
Database	700
Web	1,217
JavaTech	880
Etc	1,482
TOTAL	6,319

2) <https://www.quora.com/>

Table 4. Examples of Questions and Answers

Domain	Question	Answer
Java	What is Function Overriding and Overloading in Java ?	Method overloading in Java occurs when two or more methods in the same class have the exact same name, but different parameters. On the other hand, method overriding is defined as the case when a child class redefines the same method as a parent class
C	What is the difference between method overloading and method overriding in C#?	Method parameters must be different in method overloading whereas it must be same in method overriding.
C	How C# supports static polymorphism?	Static polymorphism. They are: Function overloading Operator overloading
C	What is polymorphism?	In object-oriented programming paradigm, polymorphism is often expressed as 'one interface, multiple functions'.
Java	Can we override a method by using same method name and arguments but different return types?	The basic condition of method overriding is that method name, arguments as well as return type must be exactly same as is that of the method being overridden.
Java	Can we use different return types for methods when overridden?	The basic requirement of method overriding in Java is that the overridden method should have same name, and parameters
Database	What are the advantages of using a package?	Modularity Easy to design the applications Better performance Hiding information Added functionality Overloading

질문과 응답은 크게 6개 분야로 나눌 수 있다. 각 분야별 수집한 질문-응답은 Table 3과 같다. 질문-응답 데이터베이스의 일부를 Table 4에 제시하였다.

4.2 응답 유사도 함수 구현을 위한 학습 데이터 및 실험 환경

유사도 함수를 구현하기 위한 학습 데이터는 다음과 같다. 단어 단위의 유사도 함수에서 f_{wemb} 자질은 단어 임베딩을 사용한다. 단어 임베딩은 Google News³⁾를 대상으로 학습한 것을 사용한다.

구절 단위의 유사도 함수 구축을 위한 구문테이블에는 한-영 문장 단위로 정렬된 약 1,753,000 문장쌍⁴⁾을 사용하였다. 정렬된 문장의 평균 문장 길이는 영어가 25.83 단어, 한국어가 18.98 어절이다. 구문 테이블을 구축하기 위해 GIZA⁺⁵⁾를 이용하였다.

문장 단위의 유사도 함수 구축을 위해서는 퀴라(Quora) 사이트로부터 15,304개의 질문을 수집하였고 126,690 문장 쌍의 데이터를 대상으로 문장 단위 자질 모델을 학습하였다. 단어 임베딩의 크기는 300차원이며 Google News 대상으로 학습한 임베딩을 그대로 사용한다. 파라미터들은 Xavier Initialization 으로 초기화하였고, 파라미터의 값 설정은 [19]의 연구에서 제안한 것을 그대로 채용하였다. 컨볼루션의 계층의 크기는 2, 필터의 크기는 4, 출력은 50차원이다. 최적화 함수로는 Adagrad를 사용했다. learning rate와 l2 norm은 0.08과 0.0004로 설정하였다.

3) <https://code.google.com/archive/p/word2vec/>
 4) 세종 영한 병렬 코퍼스를 포함하여 신문기사, 특허분야 등 여러 도메인의 한-영 정렬 문장쌍임.
 5) <http://www.statmt.org/moses/index.php>

4.3 유사 응답 인식을 위한 평가 데이터 구축

평가 데이터를 구축하기 위해 Table 3의 질문-응답 데이터베이스에서 100개의 질문을 임의로 추출한 후, 4.1장에서 제시한 4개의 웹사이트와는 다른 참조 사이트로부터 100개의 질문에 대한 새로운 답안을 수집하였다. 부록에 평가집합의 일부를 보였다. 기술의 편의를 위하여 동일한 질문 Q에 대하여 4.1절 Table 3의 답안을 **A**로 4.3절에서 평가를 위해 새로 수집한 답안을 **A'**로 지칭한다. 이렇게 추출된 답안 **A'**와 답안 **A**는 질문 Q에 대한 유사 응답으로 간주할 수 있다. 즉, **A'**에 대한 정답을 **A**로 간주할 수 있다.

유사 응답이라고는 하나 평가집합 **A'**의 문장당 평균 단어 수는 26.8이고 정답집합 **A**의 평균 단어수는 43.8로 우선 문장 길이에서 많은 차이가 남을 알 수 있다. 평가집합의 응답 **A'**와 정답에 해당하는 응답 **A**의 단어 수준 유사도를 살펴보기 위하여 각각의 응답 **A'**와 정답 **A**, 그리고 정답 이외의 응답들에 대해 자카드 유사도 값[16]과 블루 스코어 값[20]을 계산해 보았다. Table 5에서 보는 바와 같이 (2) 정답 이외의 응답과의 유사도가 (1)보다 더 높기 때문에 단순 단어 수준의 비교만으로는 응답 **A'**에 대해 정답 **A**를 찾아내기 어렵다.

평가는 100개 평가집합의 응답 **A'**에 대해 질문-응답 데

Table 5. Word-level Comparisons Between Two Different Sets for Evaluation Data

		Answer A in QA database	Best scored Answer (not A) in QA database
Answer A' in Evaluation	Jaccard	0.2520	0.2572
	BLEU score	0.1223	0.1472

이터베이스(6,319 질문-응답 쌍)로부터 유사 응답 A 를 얼마나 정확히 찾아내는가를 살펴본다. 평가집합 100개의 응답 A' 과 전체 질문-응답 데이터베이스와의 유사도를 계산하고 유사도에 따른 순위를 매긴 후 정답에 해당하는 A 의 순위를 계산하여 MAP(Mean Average Precision)와 MRR(Mean Reciprocal Rank)[16]로 평가한다.

4.4 유사 응답 인식 실험 결과

평가집합 100문장에 대한 평가 결과는 Table 6과 같다. 세 가지 자질의 유사도 함수를 모두 다 사용한 경우 (7)이 가장 좋은 성능을 보였다. 영어권의 가장 유사한 연구[15]가 MAP 68.25%, MRR 70.73%의 정확도를 보고하고 있다. 물론 동일한 문제, 동일한 데이터 집합이 아니기 때문에 직접적인 비교는 어렵지만 본 연구의 실험 결과가 기존의 연구에 비해 뒤지지 않는 성능을 보이고 있음을 확인할 수 있었다.

Table 6. Evaluation Results

Similarity Functions	MAP	MRR
(1) word	53.17	61.24
(2) phrase	45.86	54.55
(3) sentence	59.08	67.97
(4) word + phrase	59.14	68.00
(5) word + sentence	58.58	65.70
(6) phrase + sentence	61.00	68.48
(7) word + phrase + sentence	63.21	71.00

다음 Table 7은 유사도 함수 (7)을 사용하여 평가집합에 대해 가장 유사한 답안으로 찾은 것이 동일 질문에 대한 답인 경우의 예제들이다. 즉, 응답 A' 에 대해 정답 A 를 제대로 찾은 경우들이다. 각 질문과 응답은 Stanford CoreNLP⁶⁾에서 원형으로 분석하여 사용하였다.

Table 8은 유사도 함수 (7)을 사용하여 각 응답 A' 에 대해 가장 유사한 응답 10개를 찾아내고 그에 해당하는 질문을 나열한 결과이다. 결과에서 보듯이 유사 또는 관련 질문들이 추출되는 것을 확인할 수 있었다.

5. 결론

본 연구에서는 검색 기반의 질문 자동 생성 시스템을 다루었다. 좀 더 인텔리전트한 질문 생성기가 되기 위해서는 응답자의 기응답과 관련이 있으면서도 응답자가 이미 응답하지 않은 내용만을 다음번 질문으로 선택해야 한다. 시스템은 사용자의 이전 대화를 고려하여 질문을 생성하지 않으면 사용자에게 했던 질문을 다시 하거나 사용자가 이미 알고 있는 내용의 질문을 다시 하게 되는 경우가 발생할 수 있다.

본 연구에서는 사용자의 이전 응답을 고려하여 다음번 질문을 생성하기 위하여 사용자의 응답과 가장 유사한 응답을 질문-응답 데이터베이스로부터 추출하고, 이 유사 응답에 해당하는 질문을 다음번 질문 후보군에서 제거하였다. 그렇게 함으로써 사용자가 이미 알고 있는 내용에 대한 질문을 방지할 수 있다. 본 연구에서 제안한 유사 응답 추출기는 단어 정

Table 7. Examples of the Most Similar Answer from QA Database

Answer in Evaluation Set	The most similar answer in QA Database	Question
a query be cover if all the column it use come from one or more index. cover query be generally consider favorable because datum access through the index can be more efficient .	cover query make the execution of the query faster because index be store in RAM or sequentially located on disk. it make the execution of the query faster . cover query make the field be cover in the index itself, MongoDB can match the query condition as well as return the result field use the same index without look inside the document.	what be the importance of covered query ?
use a stack when you want to get thing out in the reverse order than you put they in.	stack follow LIFO method and addition and retrieval of a datum item take only o -lrb- n -rrb- time. stack be use where we need to access datum in the reverse order or they arrival. stack be use commonly in recursive function call, expression parsing, depth first traversal of graph etc.	why do we use stack ?
MapReduce operation typically split the input datum set into fully independent chunk by map operation. the framework sort the output of the map and process they by reduce operation.	map : in Hadoop , a map be a phase in hdfs query solve. a map read datum from a input location, and output a key value pair accord to the input type . Reducer : in Hadoop, a reducer collect the output generate by the mapper, process it, and create a final output of its own .	what be `` map `` and what be `` reducer `` in Hadoop ?

6) <https://stanfordnlp.github.io/CoreNLP/>

Table 8. The Most Top-10 Similar Questions Found by the Similarity Functions

Answer A' in Evaluation Set	The most top-10 similar questions
<p>a foreign key be a column in a relational database table that provide a link between datum in two table. because it refer to the primary key of another table, it act as a cross-reference between the table.</p>	<p>Q0: what be a foreign key ? Q1: how can you build a relationship between the different database in access ? Q2: define one-to-one relationship while design table . Q3: which element of hbm.xml define map the unique id attribute in class to the primary key of the database table ? Q4: what type of relationship can be form between different datum ? Q5: which element of hbm.xml be use to map a Java class property to a column in the database table ? Q6: how many primary key can be declare on a table ? Q7: what be pl/sql table ? why it be use ? Q8: what do you mean by Free form import in Sqoop ? Q9: what be Inf ?</p>
<p>a friend function of a class be define outside the scope of the class, but have permission to access all private and protected member of the class.</p>	<p>Q0: define friend function . Q1: there be two class name classa and classb. both class be in the same package. can a private member of classa can be access by a object of classb ? Q2: what be the role of protected access specifier ? Q3: what do Scope Resolution operator ? Q4: define the private, protected and public in C++ ? Q5: what be scope of a private member variable of a c# class ? Q6: what be the purpose of a access specifier in C# ? Q7: what be a friend function ? Q8: what be scope of a public member variable of a c# class ? Q9: what be the C++ access specifier ?</p>
<p>user can extend service to they need, customize application, and access cloud service from anywhere through a internet connection.</p>	<p>Q0: what be the security benefit of cloud computing ? Q1: what be cloud computing ? Q2: what do you mean by software as a service ? Q3: what be a cloud ? Q4: what be the usage of utility computing ? Q5: what be Controller in Spring MVC framework ? Q6: what be the purpose of Service Requestor in web Service architecture ? Q7: what be the purpose of Service Provider in web Service architecture ? Q8: what be IIS ? Q9: what be the different layer in cloud computing ? explain working of they.</p>

보, 바뀌쓰기 표현, 문장 수준의 정보를 모두 결합하여 가장 유사한 응답을 찾아낸다. 평가집합에 대해 실험한 결과 71%의 MRR 값을 얻을 수 있었다.

본 연구에서 제안한 유사도 측정 함수는 질문-응답 데이터베이스의 응답 유사도 계산뿐만 아니라 일반 텍스트의 유사도 검사에도 응용이 가능할 것이다. 추후의 연구로는 유사도 함수를 좀 더 고도화하여 다양한 응용 분야에 적용 가능하도록 하는 것이다.

References

[1] N. T. Le, T. Kojiri, and N. Pinkwart, "Automatic Question Generation for Educational Applications - The State of Art," In *Advanced Computational Methods for Knowledge Engineering*. Springer, Cham, pp.325-338, 2014.
 [2] Z. Ji, Z. Lu, and H. Li, "An information retrieval approach to short text conversation," *arXiv preprint arXiv:1408.6988*. 2014.
 [3] J. F. Aquino, D. D. Chua, R. K. Kabling, J. N. Pingco and

R. Sagum, "Text2Test: Question Generator Utilizing Information Abstraction Techniques and Question Generation Methods for Narrative and Declarative Text," In *Proceedings of the 8th National Natural Language Processing Research Symposium*, pp.29-34, 2011.
 [4] P. Pabitha, M. Mohana, S. Suganthi, and B. Sivanandhini, "Automatic Question Generation System," In *International Conference on Recent Trends in Information Technology*, 2014.
 [5] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Austin, Texas, pp.2383-2392, 2016.
 [6] X. Du, J. Shao, and C. Cardie, "Learning to Ask: Neural Question Generation for Reading Comprehension," *arXiv preprint arXiv:1705.00106*, 2017.
 [7] N. Duan, D. Tang, P. Chen, and M. Zhou, "Question generation for question answering," In *Proceedings of the 2017*

- Conference on Empirical Methods in Natural Language Processing*, pp.877-885, 2017.
- [8] J. Mueller and A. Thyagarajan, "Siamese Recurrent Architectures for Learning Sentence Similarity," In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, pp.2786-2792, 2016.
- [9] J. Allan, C. Wade, and A. Bolivar, "Retrieval and novelty detection at the sentence level," In: *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. SIGIR '03*, pp.314-321, 2003.
- [10] T. C. Hoad and J. Zobel, "Methods for identifying versioned and plagiarized documents," In: *Journal of the American Society for Information Science and Technology Archive*, Vol.54, Issue 3, pp.203-215, 2003.
- [11] W. N. Zhang, T. Liu, Y. Yang, L. Cao, Y. Zhang, and R. Ji, "A Topic Clustering Approach to Finding Similar Questions from Large Question and Answer Archives," *PLoS one*, Vol.9, No.3, e71511, 2014.
- [12] K. Wang, Z. Ming, and T. S. Chua, "A syntactic tree matching approach to finding similar questions in community-based QA services," In: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '09*, pp.187-194, 2009.
- [13] M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini, and R. Zamparelli, "SemEval-2014 Task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment," In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pp.1-8, 2014.
- [14] K. S. Tai, R. Socher, and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks," *ACL*, pp.1556-1566, 2015.
- [15] Z. Yan, N. Duan, J. Bao, P. Chen, M. Zhou, Z. Li, and J. Zhou, "Docchat: An information retrieval approach for chatbot engines using unstructured documents," In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp.516-525, 2016.
- [16] C. D. Manning, H. Schütze, and P. Raghavan, "Introduction to information retrieval," Cambridge University Press, 2008.
- [17] C. Callison-Burch, P. Callison-Burch, and M. Osborne, "Improved statistical machine translation using paraphrases," In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp.17-24, 2006.
- [18] R. Zens and H. Ney, "Efficient Phrase-table Representation for Machine Translation with Applications to Online MT and Speech Translation," *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp.492-499, 2007.
- [19] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "AbCNN: Attention-based convolutional neural network for modeling sentence pairs," *arXiv preprint arXiv:1512.05193*, 2015.
- [20] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "BLEU, a Method for Automatic Evaluation of Machine Translation," In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp.311-318, 2002.

최 용 석



<https://orcid.org/0000-0002-7889-8004>

e-mail : yongseok.choi.92@gmail.com

2016년 충남대학교 정보통신공학과(학사)

2018년 충남대학교 전자전파정보

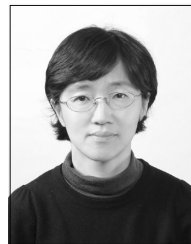
통신공학과(석사)

2018년~현재 충남대학교 전자전파정보

통신공학과 박사과정

관심분야 : 자연언어처리, 정보검색, 기계학습, 인공지능

이 공 주



<https://orcid.org/0000-0003-0025-4230>

e-mail : kjoolee@cnu.ac.kr

1992년 서강대학교 전자계산학과(학사)

1994년 한국과학기술원 전산학과(공학석사)

1998년 한국과학기술원 전산학과(공학박사)

1998년~2003년 한국마이크로소프트(유)

연구원

2003년 이화여자대학교 컴퓨터학과 대우전임강사

2004년 경인여자대학 전산정보과 전임강사

2005년~현재 충남대학교 전파정보통신공학과 교수

관심분야 : 자연언어처리, 기계번역, 정보검색, 정보추출

<Appendix>

Evaluation Set (in part)

Questions Q	Answers A in QA database	Answers A' in Evaluation Set
Q1: What is an overflow error?	It is a type of arithmetical error. It is happen when the result of an arithmetical operation been greater than the actual space provided by the system.	An error that occurs when the computer attempts to handle a number that is too large for it.
Q2: What is structure?	Structure is a user-defined data type that allows to store multiple types of data in a single unit. It occupies the sum of memory of all members.	A structure is another user-defined data type that can combine different kinds of data items. The structure is used to represent records.
Q3: When to use a DOM4J Parser?	You should use a DOM4J parser when – You need to know a lot about the structure of a document You need to move parts of the document around (you might want to sort certain elements, for example) You need to use the information in the document more than once You are a java developer and want to leverage java optimized parsing of XML.	DOM4J provides Java developers the flexibility and easy maintainability of XML parsing code.
Q4: What is deadlock?	Deadlock is a situation when two threads are waiting on each other to release a resource. Each thread waiting for a resource which is held by the other waiting thread.	A deadlock occurs when there is a circular chain of threads which each hold a locked resource and are trying to lock a resource held by the next element in the chain.
Q5: How to delete a session in servlet?	When you are done with a user's session data, you have several options: Remove a particular attribute: You can call public void removeAttribute(String name) method to delete the value associated with a particular key. Delete the whole session: You can call public void invalidate() method to discard an entire session. Setting Session timeout: You can call public void setMaxInactiveInterval(int interval) method to set the timeout for a session individually. Log the user out: The servers that support servlets 2.4, you can call logout to log the client out of the Web server and invalidate all sessions belonging to all the users.	By calling invalidate() method on session object, we can destroy the session.
Q6: What is the purpose of using COMMIT in DB2?	COMMIT is used to change the data permanently. It also allows to access data by other applications that can reference the committed data.	The COMMIT statement terminates the current transaction and ensures that changes made in the transaction are persistent.
Q7: How is MongoDB better than other SQL databases?	MongoDB allows a highly flexible and scalable document structure. For e.g. one data document in MongoDB can have five columns and the other one in the same collection can have ten columns. Also, MongoDB database are faster as compared to SQL databases due to efficient indexing and storage techniques.	MongoDB uses dynamic schemas, meaning that you can create records without first defining the structure, such as the fields or the types of their values.
Q8: What happens to the data stored in Memcached when server accidently gets shut down?	In Memcached, data is not permanently stored. It is not a durable data so, if the server is shut down or restarted then all the data stored in Memcached will be deleted.	Memcached is a memory caching system that store data and objects in RAM, so if you restart the server you will lost the information.