

# 한국어 질의 응답에서의 화제성을 고려한

## 딥러닝 기반 정답 유형 분류기

조승우<sup>0</sup>, 최동현, 김응균  
카카오

{jonathan.swcho, heuristic.085, jason.eg}@kakaocorp.com

### Deep learning-based Answer Type Classifier

### Considering Topicality in Korean Question Answering

Seung Woo Cho<sup>0</sup>, DongHyun Choi, EungGyun Kim  
Kakao Corp

#### 요약

한국어 질의 응답의 입력 질문에 대한 예상 정답 유형을 단답형 또는 서술형으로 이진 분류하는 방법에 대해 서술한다. 일반적인 개체명 인식으로 확인할 수 없는 질의 주제어의 화제성을 반영하기 위하여, 검색 엔진 쿼리를 빈도수로 분석한다. 분석된 질의 주제어 정보와 함께, 정답의 범위를 제약할 수 있는 속성 표현과 육하원칙 정보를 입력 자료로 사용한다. 기존 신경망 분류 모델과 비교한 실험에서, 추가 자질을 적용한 모델이 4% 정도 향상된 분류 성능을 보이는 것을 확인할 수 있었다.

**주제어:** 정답 유형 분류, 화제성 반영 질의 주제어, 오픈 도메인 질의 응답, 합성곱 신경망

#### 1. 서론

질의응답 기술은 사용자가 요구하는 정보를 자연어 질문으로 입력하였을 때, 시스템이 사용자의 질문에 부합하는 정답을 문서로부터 찾아서 제시하는 기술이다[1]. 이 때, 가장 연관성이 높은 N개의 문서들을 분석하여 얻어진 N개의 정답들을 각각 제시하거나[2,3], 또는 여러 후보 정답들을 하나의 단일 정답으로 통합하여 제시하기도 한다[1]. 후자의 경우, 최종 사용자에게 단 하나의 결정된 정답만을 제시하기 때문에, 스마트폰과 같은 소형 디스플레이를 가지고 있는 디바이스나, 보이스 인터페이스를 사용하는 디바이스에 적용하기 적절하다.

여러 개의 후보 정답들을 통합하여 하나의 정답으로

만드는 경우, 사용자 질의로 인해 요구되는 정답의 형태가 단답형(주제어에 대한 특정 속성을 질의하는 경우)인지, 또는 서술형(주제어에 대한 정의나, 사건에 대한 원인 등을 질의하는 경우)인지에 따라서 통합하는 방식이 크게 달라질 수 있다. 사용자 질의가 요구하는 정답의 형태가 단답형일 경우, 각 문서에서 얻어지는 정답의 형태는 대동소이하나, 서술형의 정답이 요구될 경우 각 문서에서 추출되는 정답의 형태는 크게 달라질 수 있다. 표 1은 다음 검색 엔진<sup>1</sup>에서 단답형 질문인 “마우스 발명자” 와 서술형 질문인 “컴퓨터란” 을 질의했을 때 얻어진 각 최상위 5개 문서에 대하여, 사람이 직접 질의에 대한 정답을 추출한 결과이다.

표 1. 단답형 질문과 서술형 질문의 문서별 후보 정답 예

단답형 질의	마우스 발명자	서술형 질의	컴퓨터란
후보 정답 1	더글러스 엔젤바트	후보 정답 1	미리 정해진 방법에 따라 입력된 자료를 처리함으로써 문제를 해결하는 다양한 형태의 전자공학적인 자동장치
후보 정답 2	더글러스 엔젤바트	후보 정답 2	반도체 집적 회로를 이용하여 주어지는 명령을 자동으로 맡아 하는 정보 처리기
후보 정답 3	더글러스 엔젤바트	후보 정답 3	프로그램에 의하여 정보를 처리하는 장치
후보 정답 4	더글러스 엔젤바트	후보 정답 4	전자 회로를 이용한 고속 자동 계산기
후보 정답 5	더글러스 엔젤바트	후보 정답 5	프로그램에 따라 작업이나 계산을 수행하는 기계

<sup>1</sup> <https://www.daum.net>

해당 표에서 보여지듯이, 질의에서 단답형 답변이 요구될 경우 대부분 비슷한 형태의 후보 정답들을 얻어낼 수 있지만, 서술형 답변이 요구될 경우 여러 종류의 완전히 다른 정답들이 도출됨을 알 수 있다. 따라서 각 경우에 대하여 적용되는 정답 통합 방법 또한 크게 달라져야 할 것이다. 본 논문에서는 주어진 사용자의 질의에 대하여, 질의에 대한 정답이 단답형인지, 또는 서술형인지를 예측하는 질의 답변 단답서술형 판별기에 대하여 서술한다.

주어진 질의가 단답형 답변을 요구하는지, 또는 서술형 답변을 요구하는지 여부는, 표 2에서 보여지듯이 해당 질의의 문맥 정보만 가지고는 판단이 불가하다. “경찰서 번호가 뭐야” 라는 질의의 경우 “경찰서” 라는 질의 주제어의 “번호” 라는 속성을 묻는 질문이기 때문에 단답형으로 판별되어야 하지만, “원자 번호가 뭐야” 라는 질문의 경우 질의 주제어인 “원자 번호”의 정의를 필요로 하기 때문에, 해당 질문은 서술형으로 판별되어야 한다. 따라서 질의의 문맥 정보만이 아닌, 질의에 등장한 단어들이 하나의 개념으로 고려될 수 있는 주제어인지, 또는 주제어와 속성의 조합인지 알아내는 것이 문장의 정확한 판별을 위해 매우 중요하다. 또한, 이러한 질의 주제어들은 화제성에 따라 새로이 추가되거나 제거될 수 있다. 표 2의 예시 “태양의 계절이 뭐야”에서 보여지듯이, 기존에는 단답형 정답이 필요했을 질의들도 새로운 화제성을 가진 개체의 등장(이 경우는 ‘태양의 계절’이라는 드라마)에 의하여 새로이 서술형으로 분류되어야 할 필요성이 있다.

**표 2. 답변의 단답형/서술형 판별을 질의의 문맥만으로 판단하기 어려운 사례**

정답 유형	질의 내용
서술형	경찰서 번호가 뭐야
단답형	원자 번호가 뭐야
서술형	태양의 계절이 뭐야

본 논문에서는 이러한 문제를 해결하기 위하여, 포털 사이트 쿼리 로그로부터 질의에 포함될 수 있는 후보 질의 주제어들을 추출하고, 이를 자질로 사용하여 주어진 질문이 서술형 답변을 요구하는지, 또는 단답형 답변을 요구하는지 판별하는 합성곱 신경망(Convolutional Neural Network) 기반 문장 분류 네트워크를 제안한다. 제안된 시스템의 성능을 추가적으로 향상시키기 위하여, 문장의 육하 원칙 정보 및, 수작업으로 구축된 속성 표현들이 같이 사용되었다. 실험 결과, 본 논문에서 제안된 질의 주제어 기반 단답서술형 분류기는 Baseline 시스템에 비하여 약 4%의 성능 향상을 보였다.

2장에서는 관련 연구에 대하여 간략히 서술한다. 3장에서는 화제성을 가진 질의 주제어 추출 방법 및, 추출된 질의 주제어를 이용한 질의 단답서술형 판별기에 대하여 서술한다. 4장에서는 실험 결과 및 결과 분석이 제시되고, 5장에서는 향후 연구 및 결론이 제시된다.

## 2. 관련 연구

핵심 주제어는 인명, 지명, 상호명 같은 개체명부터 화제성을 가지고 있는 사건, 인물, 신조어까지 다양한 형태로 등장할 수 있는 단어로, 주로 개체명 인식을 통하여 탐지해 왔다[1,4]. 그러나 사전 정의된 범주의 개체명 인식으로는 화제성을 가진 주제어에 대응하기 어려우며, 개체명이 아닌 일반적인 개념을 대상으로 하는 주제어 또한 다수 존재한다는 점에서(Ex. 원자 번호가 뭐야) 한계를 가진다. 본 연구에서는 포털 사이트 쿼리를 이용함으로써 기존 방식의 이러한 한계점을 극복하고자 하였다.

속성 표현은 “길이” 나, “넓이” 와 같이 주어진 질의 문장의 정답 범위를 제약하는 표현으로, 기존의 많은 연구들에서 주제어와 속성 표현, 의문사의 등장 패턴을 목록화하고[1,4,5], 동의어, 유의어 정보로 이를 확장[5,6,7]시켜서 정답 유형을 분류하는 데 사용하였다. 그러나 이러한 규칙 기반 방식은 패턴을 벗어나는 일반적인 발화에 대해서는 커버리지가 낮고, 화제성에 따라 추가된 주제어가 속성 표현을 포함하였을 경우 처리가 불가능하다. 본 연구에서는 이러한 문제를 해결하기 위해, [8]에서 제안된 문장 분류 합성곱 신경망에 질의 주제어 정보가 추가된 질의 단답서술형 판별 네트워크 구조를 제안하였다.

## 3. 방법론

본 장에서는 먼저 포털 서비스 쿼리에서 화제성을 가진 질의 주제어를 추출하는 방법에 대하여 서술한다. 또한, 이를 이용하여 주어진 사용자 질의가 요구하는 답변이 단답형인지, 또는 서술형인지 판별하는 합성곱 네트워크 구조에 대하여 서술한다.

### 3.1 화제성 반영 질의 주제어 추출 및 사전 구축

본 절에서는 포털 사이트 검색 로그를 이용하여 화제성이 반영된 질의 주제어를 추출해내는 방법에 대하여 서술한다. 검색 서비스 사용자들은 일반적으로 검색 엔진에 대한 질의 과정을 통하여, 자신이 현재 알지 못하는 대상에 대한 정보를 파악하고자 한다. 이 때 사용자가 파악하고자 하는 대상은 새로이 만들어진 개념 또는 단어일 수도 있고, 현실에서 화제가 되고 있는 사건일 수도 있다. 따라서 검색 로그를 이용한다면, 화제성이 있는 질의 주제어 후보군을 얻어낼 수 있을 것이다.

이를 위하여 먼저 검색 로그 중, 질의 주제어를 명확하게 한정지을 수 있는 로그들을 기 정의된 패턴을 이용하여 필터링한다. 표 3은 질의 주제어를 포함한 검색 로그 추출을 위해 사용한 패턴 목록을 보여 준다.

**표 3. 질의 주제어 추출을 위한 검색 로그 필터링 패턴**

정규 표현식	$\wedge[가-힣A-Za-z0-9]+(이란 란 논란)\$$
--------	------------------------------------

표 3에 제시된 정규 표현식을 2019년 1월부터 6월까지의 검색 로그에 적용한 결과, 총 34,056,766개의 질의 주제어를 포함한 검색 로그(중복 포함)를 얻을 수 있었다. 얻어진 검색 로그에서 질의 주제어를 추출하는 것은 정규 표현식에서 해당 부분에 매칭되는 부분을 추출함으로써 이루어질 수 있다. 추출된 질의 주제어 후보들 중, 등장 빈도수가 6,000 회 이상인 3,375개의 질의 주제어들이 질의 분류기의 사전으로서 사용되었다. 표 4는 두 개 이상의 어절로 구성된 질의 주제어 중 상위 10개를 각 주제어들의 등장 빈도수와 함께 나타낸다. 드라마(“하나뿐인 내편”이나 “아스달 연대기”), 영화(“어벤져스 엔드게임”), 사건사고(“헝가리 유람선”)와 같이 해당 시기에 주목을 받던 주제어들이 성공적으로 추출된 것을 알 수 있다.

표 4. 두 개 어절로 구성된 질의 주제어 추출 예시

순위	질의 주제어	빈도수
1	아프리카 tv	2,212,778
2	농협 기업인터넷뱅킹	1,387,953
3	기업은행 인터넷뱅킹	831,165
4	하나뿐인 내편	678,059
5	공무원 연금관리공단	533,844
6	실시간 시청률	450,506
7	어벤져스 엔드게임	439,642
8	아스달 연대기	426,966
9	연애의 맛	399,295
10	헝가리 유람선	385,223

질의 주제어 사전과 별개로, 속성 표현 사전 또한 수작업으로 구축되었다. 속성 표현은 질의에서 주어진 주제어의 속성에 대해 묻기 위하여 사용되는 표현으로, “길이”나 “넓이”, “키”, “높이” 등이 있다. 화제성을 가지는 개체가 등장할 때마다 변동이 발생하는 질의 주제어 사전과 달리, 자주 사용되는 속성 표현은 대부분의 경우 큰 변동이 없이 동일하다. 본 연구에서는, 169개의 자주 사용되는 속성 표현들을 수동으로 선별하여, 실시간으로 얻어진 질의 주제어와 같이 질의 분류기의 사전으로서 이용될 수 있도록 하였다.

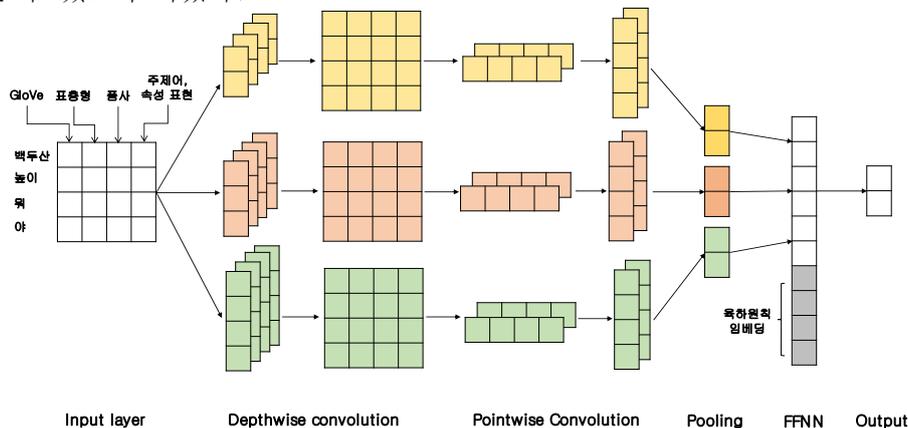


그림 1. 깊이별 분리 합성곱 신경망 기반 분류 네트워크 모델

### 3.2 네트워크 구조

본 절에서는 네트워크 구조의 입력에 대해서 먼저 설명을 한 후, 분류 모델 구조를 서술한다. 질의응답 시스템에 입력된 질의 문장은 기본적으로 형태소 분석되어 각 형태소의 표층형, 품사 정보를 가지게 된다. 표층형, 품사 정보, GloVe 의미 벡터[9]를 사용하는 기존 분류 모델[8]에 질의 문장의 화제성 반영 질의 주제어, 속성 표현, 육하원칙 정보를 추가하여 임베딩으로 입력한다.

화제성 반영 질의 주제어 목록과 속성 표현 목록에 있는 표현들이 질의 문장에 있는지 최장 길이 탐색한 후에, 표 5처럼 탐색된 후보 주제어와 속성 표현에 특수 태그를 부여한다. 단, 후보 주제어와 속성 표현의 범위가 서로 겹치면 속성 표현을 우선 적용하며, 질의 문장 구조를 고려하여 속성 표현은 첫 어절 이후의 탐색 결과만 적용한다.

표 5. 후보 주제어와 속성 표현에 대한 태그 부여 예시

항목	내용
입력 문장	블랙홀의 크기는
태그 부여	블랙홀/<Topic>_B 의 크기/<Attr>_B 는

육하원칙 정보는 속성 표현과 비슷하게 정답 유형의 범위를 제약시켜주는 효과가 있다. 한국어 문장 분류[8]의 응용 일환으로 개발된 합성곱 신경망 기반 분류기를 이용하여 입력된 질의 문장에 대한 육하원칙 태그를 획득한다.

추가 정보를 적용한 네트워크 모델은 그림 1과 같다. 질의 문장에 대한 입력 분석이 끝나고, 형태소 표층형은  $x^{sf} \in \mathbb{R}^i$ 으로, 품사 정보가  $x^{pos} \in \mathbb{R}^k$ , 화제성 반영 질의 주제어와 속성 표현 정보를  $x^{etag} \in \mathbb{R}^l$ , GloVe 정보는  $x^{glove} \in \mathbb{R}^g$ 로 임베딩될 때,  $i$ 번째 형태소의 입력 벡터  $x_i$ 는 아래 수식처럼 연결하여 사용한다.

$$x_i = [x^{sf}; x^{pos}; x^{etag}; x^{glove}] \quad (1)$$

질의 문장에 존재하는  $N$ 개 형태소에 대한 입력 벡터가 생성된 후, 깊이별 분리 합성곱 신경망(Depthwise separable convolution) [10]을 이용하여 질의 문장의 전체적인 문맥을 파악한다. 깊이별 분리 합성곱 신경망의 각 Kernel size는 2,3,4로 설정하고 Filter 개수는 동일하게 지정하였다. Batch normalization[11]을 도입하여 각 Filter별로 계산 값을 정규화하고 max-pooling layer를 거치도록 구현한다.

각 Filter의 최종 계산 수치  $C_s^{out} \in \mathbb{R}^c, s \in \{2,3,4\}$ 는 질의 문장의 육하원칙 정보를 임베딩한 벡터  $x^{5w1h} \in \mathbb{R}^d$ 와 함께 모두 하나의 벡터로 연결된다. 연결된 벡터는 Feed-forward neural network(FFNN)의 입력으로 사용되며, 입력 질문에 대한 클래스  $c \in \mathbb{R}^2$ 에 대한 분류 확률은 아래 수식처럼 FFNN층 결과에서 softmax를 취하여 계산한다.

$$h_s = [C_2^{out}; C_3^{out}; C_4^{out}; x^{5w1h}] \quad (2)$$

$$p(c|w_{1:N}) = \text{softmax}(W_{hy}h_s + b_y) \quad (3)$$

#### 4. 실험 및 분석

한국어 질의응답 로그 데이터로 구성된 말뭉치 정보와 정답 유형 분류 실험에 대한 설계 및 결과를 분석과 함께 나타낸다.

##### 4.1 실험 말뭉치 구성

한국어 질의응답 로그를 수동 분류하고 표 6에서와 같이 학습, 개발, 실험 말뭉치로 나누어 사용하였다. 각 말뭉치에 대한 모든 문장은 형태소 분석되었으며, 시작 위치와 끝 위치를 명시적으로 입력하기 위하여 별도 토큰이 앞뒤로 추가되었다.

표 6. 말뭉치 내 서술형 질문 및 단답형 질문 개수

항목	서술형 질문 수	단답형 질문 수
학습 말뭉치	2,618	2,988
개발 말뭉치	127	164
실험 말뭉치	200	200

##### 4.2 실험 결과

평가 기준의 베이스라인 모델은 기존 연구[8]를 재구현한 모델로 지정한다. 육하원칙, 화제성 반영 질의 주제어, 속성 표현 정보를 모두 사용한 제안 모델을 베이스라인 모델과 비교한다. 평가 수치는 문장 분류 정확도가 사용되었으며, 추가 분석을 위해서 단답형 분류에 실패한 문장, 서술형 분류에 실패한 문장, 둘 다 성공한 문장의 개수를 계산하였다. 또한, 추가 정보의 세부적인 효용성 평가를 위하여, 추가 정보를 하나씩 제거하는 Ablation 실험을 진행한다.

베이스라인 모델과 제안 모델의 성능은 표 7에 기입되

었다. 표 7에서 베이스라인 모델과 비교하여 제안 모델은 4% 정도 더 높은 분류 정확도를 가진 것으로 나타났다. 유형별 분류 개선 정도는 단답형 분류에 실패한 문장(C1), 서술형 분류에 실패한 문장(C2), 둘 다 올바르게 분류한 문장(C3)을 보면, 단답형 질문에 대한 오류는 제안 기법을 통해서 월등히 개선되었으나, 서술형 질문에 대한 오류는 비슷한 수준을 유지하는 것으로 집계되었다.

표 7. 베이스라인 모델과 제안 모델 성능 비교

모델	C1	C2	C3	정확도
Baseline[9]	50	1	349	87.25 %
제안된 기법	33	2	365	<b>91.25 %</b>

표 8은 추가 정보에 대한 Ablation 실험 결과이다. 개별적인 추가 정보를 제거했을 때, 육하원칙(5W1H)이 0.5%, 화제성 반영 질의 주제어(Topic)는 0.75%, 속성 표현(Attr)은 3.25% 정도 성능이 하락하는 것을 발견하였다. 따라서, 속성 표현이 가장 주요한 추가 정보였으며, 화제성 반영 질의 주제어, 육하원칙이 순차적으로 높은 중요도를 가지는 것을 알 수 있었다.

표 8. 추가 정보에 대한 Ablation 실험 결과

실험 항목	C1	C2	C3	정확도
제안된 기법	33	2	365	<b>91.25 %</b>
- Topic	36	2	362	90.5 %
- Attr	46	2	352	88 %
- 5W1H	35	2	363	90.75 %
- Topic, Attr	44	2	354	88.5 %
- Topic, 5W1H	38	3	359	89.75 %
- Attr, 5W1H	47	2	351	87.75 %

##### 4.3 결과 분석

베이스라인 모델은 단답형 질문에 대한 분류보다 서술형 질문에 대한 분류 성능이 상대적으로 높은 것으로 측정되었다. 이는 로그에서 추출한 서술형 질문이 대부분 단순한 유형이기 때문이다. 대부분이 “명사+격조사+의 문사”의 조합으로 구성된 질의 문장이었고, 이러한 질의 문장들은 품사 정보를 적용하면서 분류에 어려움이 없었던 것으로 분석할 수 있다. 제안 모델은 이러한 베이스라인 모델보다 단답형 질문 분류에 대한 성능이 주로 개선되었다. 이는 속성표현이나 육하원칙을 추가하면서 분류 모델이 정답의 범위를 축약하도록 학습시킬 수 있었기 때문이다.

각 추가 정보에 대한 Ablation 실험 결과에서 속성 표현 정보는 단답형 질문 분류 성능을 크게 좌우하는 정보로 나타났다. 이는 단답형 질문 분류에 유용한 표층형 정보를 명시적으로 지정함으로써 단답형 질문 분류에 관한 표층형 임베딩 학습이 크게 개선되었기 때문이다. 별

도 태그에 대한 임베딩과 함께, 개선된 표층형 정보는 정답 범위를 효과적으로 줄여 분류 성능을 크게 향상시킬 수 있었다.

화제성 반영 질의 주제어는 서술형 질문 분류보다는 단답형 질문 분류 성능을 개선하는데 사용되었다. 이는 추출된 질의 주제어 정보가 “질의 주제어+속성 표현” 패턴 학습에 유용하게 사용되어 성능을 향상시킬 수 있었기 때문이다. 서술형 표현 정보만을 추가한 실험에서 상대적으로 성능 향상 폭이 낮아진 것도 특정 형태소 표층형에 대한 강조 효과가 약해짐으로써 발생한 것으로 판단할 수 있다.

육하원칙 정보도 단답형 분류 성능에 영향을 미치는 것으로 드러났다. 육하원칙 정보는 where, when은 단답형으로, how, why는 서술형으로 쉽게 분류해 낼 수 있으나, what이나 who의 경우에는 서술형과 단답형이 혼재되어 있는 특징이 있다. what이나 who 정보와 함께 특정 형태소 표층형이 등장했을 때, 질의 문장을 단답형 질문으로 분류하게 되면서 유용한 표층형 정보를 속성 표현처럼 학습한 것으로 분석된다. 또한, 여기에 화제성 반영 질의 주제어 정보의 추가로 “질의 주제어+속성 표현” 패턴까지 처리할 수 있게 되면서 성능 개선에 영향을 주었던 것으로 보인다.

표 9는 추가 정보를 적용한 모델과 베이스라인 모델을 비교하여, 개선된 분류 사례를 나타낸다. 앞서 Ablation 실험 분석에 언급한 것과 같이, 육하원칙과 속성 표현은 주로 단답형 질문 분류에서 특정 형태소들이 입력된 문장에서 나타난 오분류를 수정하는 결과를 보여주었다. 화제성 중심 주제어 정보는 “질의 주제어+속성 표현” 패턴에 대한 개선 사례와 함께, 드라마 개체명에 대한 오분류를 개선한 것으로 나타났다.

표 9. 추가 정보 적용에 따른 분류 개선 사례

적용 항목	문장 내용
육하원칙	강원도지사 이름이 뭐야 (who)
	제주도 넓이 알려줘 (what)
화제성 반영 질의 주제어	빅스비 별명이 뭐야
	지붕 뚫고 하이킥이 뭐야
속성 표현	롯데월드 입장료 알려줘
	금굴하는 번호 검색해줘

표 10은 추가 정보를 모두 적용한 제안 모델에서 분류에 실패한 사례들을 나타낸다. 첫 번째 사례는 속성 표현이 포함된 질의 주제어에 대한 분류 학습이 잘 이루어지지 않은 사례이다. 오분류를 개선하기 위해서는 이런 유형의 화제성 반영 질의 주제어를 사전에 추가하고 추가된 질의 주제어에 대한 학습 데이터도 생성해야 한다.

두 번째 사례는 질의 문장에 대한 정답 정보가 나열형으로 나타나는 사례이다. 질의 주제어(“맨인블랙”)에 대한 속성 표현(“시리즈”)이 나타내는 대상이 하나가 아니라 여러 개의 대상이 될 수도 있기 때문이다. 현재 기준은 속성 표현이 나타내는 정보를 단일 정보로만 국

한시켜서 수동 분류하였기 때문에, 이런 사례에 대한 정답은 단답형으로 지정되어 있다. 그러나 나열형은 단답형/서술형에 속하지 않는 유형이므로, 나열형 질문에 대한 정의를 세우고, 나열형 질문에 맞는 속성 표현을 독립적으로 구축해야 한다.

마지막 사례는 특정 질의 주제어(“최저 임금”)에 한하여, 단답형 정보(“최저 임금의 금액”)와 서술형 정보(“최저 임금에 대한 정의”)를 둘 다 획득할 수 있는 사례이다. 나열형과 마찬가지로, 중의적 질문은 새로운 유형의 질문이다. 질의 주제어가 가진 각각의 의미를 검색 기술 및 정답 추출 기술에서 어떻게 활용할지 설계하는 것이 중요하며, 이러한 문장은 단답서술형이 아니라 별도의 유형으로 분류되어야 할 것이다.

표 10. 모든 추가 정보가 적용된 모델에서의 실패 사례

문장 내용	예측	정답
유통 기한 알려줘	단답형	서술형
맨인블랙 시리즈 알려줘	서술형	단답형
최저 임금 알려줘	서술형	단답형

## 5. 결론

본 논문에서는 한국어 질의응답 시스템의 입력 질의 문장을 단답형/서술형으로 분류하는 모델을 제안하였다. 정답 유형 분류에 필요한 질의 주제어 정보를 포털 서비스 쿼리를 통해서 확보하였고 쿼리 빈도수를 활용하여 화제성을 반영하였다. 화제성 반영 질의 주제어와 속성 표현은 육하원칙 정보와 함께 명시적으로 분류에 대한 정보를 모델에게 전달함으로써, 정답의 범위를 유추할 수 있도록 학습을 유도하였다. 제안한 추가 정보를 모두 적용한 모델은 기존 모델에 비해 4% 정도의 분류 정확도가 개선되었다.

그러나 화제성 반영 질의 주제어와 속성 표현은 추가적인 연구로 범위를 확장할 필요가 있다. 화제성을 띤 주요 질의 주제어들만 선출하기 위하여, 쿼리 접미사와 단순한 빈도수만 확인하는 것이 아니라, 토픽 모델링이나 TF-IDF 같은 기법을 적용해야 한다. 또한, 속성 표현을 포함한 질의 주제어에 대한 사전 확장도 반드시 필요하다. 속성 표현은 현재 명사 형태로만 사전으로 적용되고 있으나, 정답을 제약할 수 있는 술부에도 이를 적용할 수 있어야 한다.

화제성 반영 질의 주제어와 속성 표현의 적용에 있어서도 새로운 설계가 필요하다. 나열형 질문이나 중의적인 질문에 대한 유형 정의와 함께, 각 새로운 유형별로 질의 주제어와 속성 표현을 찾는 연구가 진행된다면 더 높은 분류 성능을 획득할 수 있을 것이다.

## 참고문헌

- [1] 허정, 류범모, 장명길, 김현기, “오픈 도메인 질의 응답을 위한 검색문서 제약 및 정답유형 분류기술”, 정보과학회논문지 소프트웨어 및 응용, pp.118-132, 2012.
- [2] Yan, Ming., Xia, Jiangnan., Wu, Chen., Bi, Bin., Zhao, Zhongzhou., and Zhang, Ji., “A deep cascade model for multi-document reading comprehension”, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33. 2019.
- [3] Chen, Danqi., Fisch, Adam., Weston, Jason., and Bordes, Antoine, “Reading wikipedia to answer open-domain questions”, arXiv preprint arXiv:1704.00051, 2017
- [4] 허정, 류범모, 김현기, 옥철영, “인공지능: WiseQA 를 위한 정답유형”, 정보처리학회논문지 소프트웨어 및 데이터 공학, pp.283-290, 2015.
- [5] 윤성희, “자연어 질의유형 판별과 응답 추출을 위한 어휘 의미 체계에 관한 연구”, 한국산학기술학회논문지, pp. 539-545, 2004
- [6] Jeong, Yoonjae, and Sung-Hyon Myaeng, “Using WordNet hypernyms and dependency features for phrasal-level event recognition and type classification”, European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 2013
- [7] Hao, Tianyong., Xie, Wenxiu., Wu, Qingyao., Weng, Heng., and Qu, Yingying., “Leveraging question target word features through semantic relation expansion for answer type classification”, Knowledge-Based Systems, 2017
- [8] 최동현, 박일남, 임재수, 백슬예, 이미옥, 신명철, 김응균, 신동렬, “한국어 대화 엔진에서의 문장 분류”, 한글 및 한국어 정보처리 학술대회 논문집, pp.210-214, 2018.
- [9] Pennington, Jeffrey., Richard Socher., and Christopher Manning., “GloVe: Global Vectors for Word Representation”, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014.
- [10] Howard, Andrew. G., Zhu, Menglong., Chen, Bo., Kalenichenko, Dmitry., Wang, Weijun., Weyand, Tobias., Andreetto, Marco., and Adam, Hartwig., “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, arXiv preprint arXiv:1704.04861, 2017
- [11] Ioffe, Sergey, and Christian Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, International Conference on Machine Learning, 2015