

# 색인어 정규화 및 응답 필터링을 이용한 검색기반 채팅 모델

이현구<sup>○</sup>, 김민경, 김진태, 김학수, 이연수\*, 최맹식\*  
강원대학교 컴퓨터정보통신공학과, ㈜엔씨소프트\*

nlphglee@kangwon.ac.kr, kmink0817@kangwon.ac.kr, wlsxo1119@kangwon.ac.kr, nlpdrkim@kangwon.ac.kr,  
yeonsoo@ncsoft.com, mschoi@ncsoft.com

## Retrieval-based Chat Model using Index-Term Normalization and Answer Filtering

Hyeon-gu Lee<sup>○</sup>, Minkyong Kim, Jintae Kim, Harksoo Kim, Yeonsoo Lee\*, Maengsik Choi\*  
Kangwon National University Computer and Communication Engineering  
NCSOFT Corp.\*

### 요 약

채팅 모델은 인간과 컴퓨터가 신변잡기 대화를 나눌 수 있게 해주는 시스템으로 빠른 속도로 발전하는 인공지능 음성언어 비서 시스템에 필수적으로 사용되는 기술이다. 본 논문에서는 검색기반 채팅 모델에서 발생하는 검색 효율 문제와 정확하지 못한 답변을 출력하는 문제를 해결하기 위해 색인어 정규화와 응답 필터링이 적용된 검색기반 채팅 모델을 제안한다. 색인어 정규화를 통해 99.3%의 색인 커버리지를 확보하였으며 필터링 모델을 통해 기존 검색 모델에서보다 향상된 사용자 만족도를 얻었다.

주제어: 채팅 모델, 색인어 정규화, 문장 임베딩, 필터링 모델

### 1. 서론

최근 애플의 시리(Siri), 아마존의 알렉사(Alexa)는 물론 국내 SKT 누구(Nugu), KT 기가지니(GIGA Genie), 네이버 클로바(Clova)와 같은 인공지능 음성언어 비서 시스템이 활발히 연구되고 있으며 관련 시장이 빠른 속도로 성장하고 있다[1]. 이러한 인공지능 음성언어 비서 시스템에서 인간과 컴퓨터가 신변잡기 대화를 나눌 수 있도록 하는 채팅 모델은 가장 필수적인 기술로 입력된 문장을 통해 적절한 답변을 출력하는 시스템이다. 채팅 모델은 보유한 대화쌍에서 가장 유사한 내용을 선별하는 검색기반 모델과 입력된 문장을 통해 답변문장을 생성하는 생성기반 모델이 있다. 검색기반 모델은 제한된 데이터를 효과적으로 검색하고 응답의 정확도를 높여야하며 생성기반 모델은 문법적 오류 및 의미적 정확성을 향상시켜야 하는 이슈를 가지고 있다. 본 논문에서는 검색기반 채팅 모델의 이슈를 해결하기 위해 색인어 정규화 및 응답 필터링을 적용하여 검색 커버리지와 검색 결과 정확도를 향상시키는 검색기반 채팅 모델을 제안한다.

### 2. 관련 연구

검색기반 방식의 연구는 채팅 모델뿐만 아니라 질의응답에서도 많이 사용되고 있으며 검색 커버리지를 높이기 위해 질의를 정규화[2]하거나 질의확장[3]을 통해 어휘 일치율을 향상시키는 연구가 진행되었다. 또한 어휘 정보만 사용하는 기존의 검색 모델과 달리 의미정보를 함

게 반영하고 구문 패턴을 통해 말잇기를 하는 등 답변 문장의 품질을 향상시키기 위한 채팅 모델 연구도 진행되었다[4]. 검색 모델 외에도 질의를 분석하여 얻어진 자질을 통해 검색 결과를 재순위화 하여 문장의 품질을 향상시키는 후처리 방식의 연구도 진행되고 있다[5]. 본 논문에서는 검색의 커버리지를 향상시키기 위해 개체명과 시제, 보조 용언의 양상 정보를 통한 색인어 정규화를 사용하고 문장 임베딩을 활용한 응답 필터링을 통해 응답의 정확도를 향상시키는 검색기반 채팅 모델을 제안한다.

### 3. 검색기반 채팅 모델

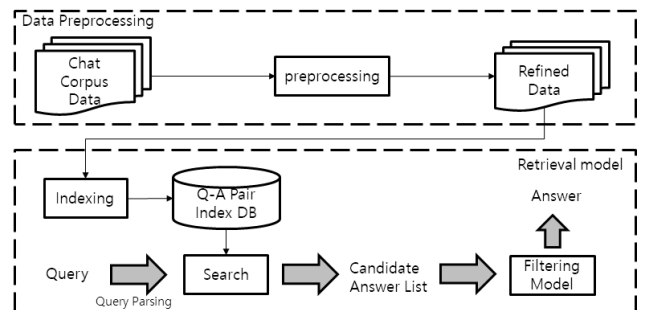


그림 1. 제안 모델의 구조도

[그림 1]은 제안 모델의 구조도를 보여준다. 본 논문에서 제안하는 검색기반 채팅 모델은 주어진 색인 질의/

응답 쌍 데이터를 전처리 및 정규화하여 색인하는 색인 모델, 사용자 질의가 입력됐을 때 사용자 질의와 유사한 색인 질의를 찾고 색인 질의와 쌍으로 존재하는 응답을 검색하는 검색 모델, 마지막으로 검색된 후보 응답이 적절한지를 판단하는 응답 필터링 모델로 구성된다. 색인 모델과 검색 모델에서 사용하는 색인어 정규화는 개체명, 보조 용언, 시제를 통해 생성하고 후보 응답이 적절한지 판단하는 응답 필터링 모델은 auto-encoder 방식의 sequence-to-sequence 모델을 통해 사용자 질의, 색인 질의, 색인 응답의 문장 임베딩을 생성하고 생성된 문장 임베딩을 통해 이진 분류하여 사용자 질의와 응답이 적합한지를 판단한다.

### 3.1. 색인어 정규화

본 논문에서는 검색의 커버리지(coverage)를 높이기 위해 색인 질의와 사용자 질의 색인어를 정규화한다. 정규화 작업은 형태적으로나 구문적으로 상이한 문장이라도 동일한 색인어로 변환하는 기술이다. 본 논문은 다음과 같은 순서로 정규화를 적용한다.

- 1) 연속된 기호(. ? !)를 하나의 기호로 변환
- 2) 자주 사용되지 않는 기호를 제거
- 3) 질문 패턴 사전을 통해 매칭된 “~까”, “~나”, “?” 등을 질문을 나타내는 심볼 “@Q” 로 변경
- 4) 형태소 분석결과 미등록어(NA)로 구분된 경우 의미를 유추할 수 있는 앞 세글자만 추출
- 5) 개체명 인식결과를 통해 개체명에 해당하는 형태소를 개체명으로 치환
- 6) 보조 용언과 시제를 통해 양상(modality)을 파악하고 양상별 심볼로 정규화
- 7) 체언류, 용언류의 내용어(content word)가 아닌 형태소 제거

정규화 과정 3)의 질문 패턴 사전은 질문을 표현할 때 사용되는 어휘들을 기록해둔 사전으로 질의에서 사전에 매핑되는 어휘가 나타날 시 치환하는 역할을 한다. 4)의 미등록어 처리 부분은 잘못된 어휘나 출현 빈도가 낮은 어휘의 커버리지를 높이기 위한 부분으로 “안녕하세요” 과 같이 잘못된 어휘로 인한 미등록어를 “안녕하” 로 치환하여 커버리지를 높인다. 6)의 양상별 심볼은 형태소가 조합되면서 나오는 양상을 그룹화[6]하여 “밥을 먹고 싶다.” 와 “밥을 먹기 바란다.” 같이 의미는 같지만 형태가 다른 질의를 정규화 해준다. [그림 2]는 색인어 정규화를 통해 생성된 색인어를 보여준다.

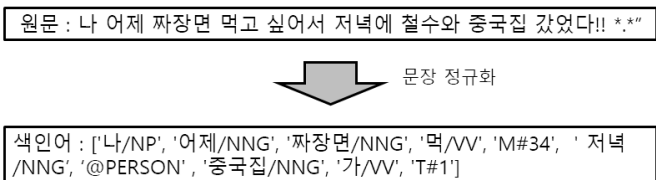


그림 2. 문장 정규화의 예

[그림 2]에서 색인어 정규화를 통해 생성된 색인어 중 “M#34” 는 희망을 나타내는 보조 용언이 치환된 것이고 “T#1” 은 과거를 나타내는 시제이다. 또한 “철수” 의 경우 개체명 인식결과 사람을 나타내는 “@PERSON” 으로 분류되어 형태소 “철수/NNP” 를 개체명 “@PERSON” 으로 변경한다.

### 3.2. 검색 모델

검색 모델을 사용하기에 앞서 색인을 위한 색인 데이터의 구조는 [그림 3]과 같다.

명칭	의미
ID	색인 질의/응답 쌍의 ID
Question	색인 질의의 원본 문장
Answer	색인 응답의 원본 문장
Index Term	색인 질의를 정규화하여 표현된 색인어

그림 3. 색인 데이터의 구조

[그림 3]의 구조에서 Index Term은 색인 질의를 3.1절에서 언급한 정규화 방법을 통해 생성된 색인어로 검색 모델에서 검색 할 때 사용되는 색인 정보이다. 나머지 정보는 검색이 됐을 때 필요에 따라 정보를 가져오기 위해 같이 색인된 정보로 검색할 때 사용되지 않는 정보이다.

질의/응답 데이터를 색인 한 후 사용자 질의와 유사한 색인 질의를 검색한다. 본 논문에서는 사용자 질의에 가장 유사한 색인 질의/응답 쌍을 검색하기 위해 식 (1)의 okapi BM25를 사용하여 사용자 질의와 색인 질의간의 유사도를 계산한다.

$$score(Q, S) = \sum_{i=1}^n IDF(q_i) \cdot \frac{f(q_i, S) \cdot (k_1 + 1)}{f(q_i, S) + k_1 \cdot (1 - b + b \cdot \frac{|S|}{avgdl})} \quad (1)$$

$$IDF(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

식 (1)에서  $Q$ 는 사용자 질의,  $S$ 는 색인 질의,  $q_i$ 는 사용자 질의에 포함된  $i$ 번째 단어,  $avgdl$ 은 평균 문장 길이,  $N$ 은 색인된 전체 질의의 수이다. 이때 초모수(hyper parameter)로 사용되는  $k_1$ 과  $b$ 는 문장 검색에 최적화된 1.0과 0.18로 설정한다[7].

### 3.3. 응답 필터링 모델

본 논문에서는 검색 모델을 통해 검색된 결과가 사용자 질의에 알맞은 응답이 아닌 경우를 감소시키기 위해 문장 임베딩을 사용하는 응답 필터링 모델을 제안한다. 응답 필터링 모델에 사용되는 문장 임베딩 모델은 입력 문장으로부터 생성되는 문장이 입력과 동일하게 나오도록 하는 auto-encoder 방식[8]의 sequence-to-sequence를

모델을 사용한다. [그림 4]는 auto-encoder방식의 문장 임베딩 생성 모델의 구조이다.

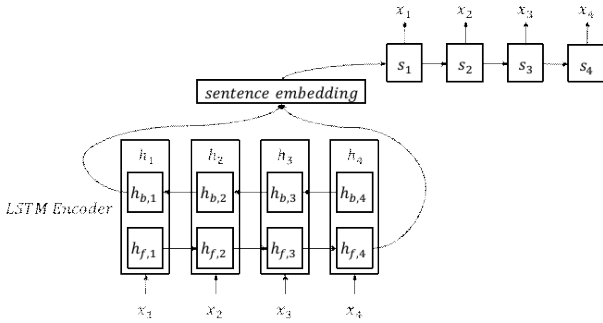


그림 4. Auto-encoder방식의 문장 임베딩 생성 모델

[그림 4]에서 보는 것과 같이 양방향 LSTM 순환 신경망(Bi-directional LSTM Recurrent Neural Network)[9]을 통해 입력 문장을 임베딩하고 생성된 임베딩을 디코딩하여 입력과 똑같은 문장이 생성되도록 학습한다. 다음으로 새로 입력된 문장의 문장 임베딩을 생성하고자 할 때 양방향 LSTM 순환 신경망을 통해 생성된 인코딩 정보만을 문장 임베딩으로 사용한다. 문장 임베딩 모델은 응답 필터링 모델에 적용하기 전 사전 학습하여 사용하며 600,000개의 구어체 자막 데이터를 사용하여 학습한다.

필터링 모델은 사용자 질의, 색인 질의, 색인 응답의 문장 임베딩을 통해 사용자 질의와 검색된 후보 응답이 적절한지 여부를 판단하는 이진 분류 신경망을 사용한다. [그림 5]는 문장 임베딩을 사용하는 이진 분류 모델을 나타낸다.

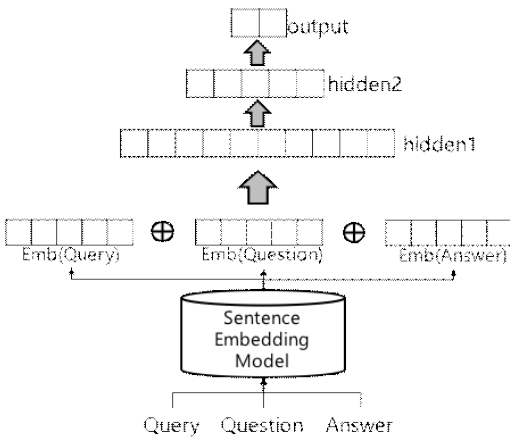


그림 5. 응답 필터링 모델

[그림 5]에서 Query는 사용자 질의, Question은 색인 질의, Answer는 색인 응답을 나타낸다. 사용자 질의, 색인 질의, 색인 응답을 각각 문장 임베딩 모델을 통해 임베딩을 생성하고 생성된 임베딩을 연결(concatenation)한 후 신경망 모델을 통해 이진 분류한다.

## 4. 실험 및 평가

### 4.1. 실험 준비

본 논문은 필터링 모델과 검색기반 채팅 모델 두 가지 성능을 평가한다. 필터링 모델을 학습 및 평가하기 위해 색인 데이터의 질의/응답 쌍을 positive 데이터, 색인 질의 Q를 검색하여 나온 검색 순위 4위 이하의 질의/응답 Q' /A' 로 만든 Q/A' 를 negative 데이터로 사용한다. 학습 데이터 82,814개, 평가 데이터 20,704개를 사용하며 총 103,518개의 데이터는 positive 데이터 49,417개, negative 데이터 54,101개로 구성된다. 다음으로 검색기반 채팅 모델의 색인 및 성능을 평가하기 위해 색인 커버리지를 확인하는 close테스트에 54,252개의 질의/응답 쌍 데이터, 사용자 만족도를 확인하기 위한 open테스트에 색인 데이터에 포함되지 않은 200개의 질문을 사용한다.

### 4.2. 응답 필터링 모델 실험 평가

본 논문에서 응답 필터링 모델의 성능을 평가하기 위해 정확도(accuracy), positive 레이블의 정확률(precision), 재현율(recall), F1-score를 사용한다. 표 1은 필터링 모델의 성능을 나타낸다.

표 1. 필터링 모델의 성능

	Performance
Accuracy	0.983
Positive Precision	0.967
Positive Recall	1.0
Positive F1-score	0.984

표 1에서 F1-score를 보아 사용자 질의와 검색된 색인 응답을 필터링 하는데 높은 성능을 보이고 있어 정확한 답변을 해야 하는 검색기반 채팅 모델에 효율적으로 사용할 수 있을 것으로 판단된다.

### 4.3. 검색기반 채팅 모델 실험 평가

검색기반 채팅 모델은 사용된 색인 커버리지를 확인하기 위한 close테스트와 채팅 모델의 사용자 만족도를 평가하기 위한 open테스트 두 가지를 평가한다. close테스트는 색인 질의를 검색했을 때 1순위로 검색된 질의가 검색 질의와 같은 여부를 측정한다. close테스트의 성능은 표 2와 같다.

표 2. 검색기반 채팅 모델 close 테스트 성능

	Accuracy
기본 검색 모델	0.989
필터링 모델 추가	0.993

표 2에서 기본 검색 모델은 3.2점까지 모델로 필터링 모델이 적용되지 않은 모델이며 필터링 모델 추가는 3.3점의 필터링 모델을 적용시킨 검색기반 채팅 모델이다. close테스트지만 약간의 오류가 있는데 이는 일부 문자가 다르지만 정규화된 질의어가 같아 발생하는 오류다.

다음으로 open테스트는 색인 질의에 포함되지 않은 200개의 질의를 검색해 나온 결과를 사람이 직접 정성평가한 결과이다. 정성평가는 1점부터 5점까지의 범위로 평가한다. 본 논문에서 정성평가의 기준은 다음과 같다.

- 1점 - 응답이 전혀 없거나 말이 되지 않는 것
- 2점 - 질의에 대한 응답으로 조금 적절치 못한 것
- 3점 - 질의에 대한 응답으로 될 수도 있는 것
- 4점 - 질의에 대한 응답으로 괜찮은 것
- 5점 - 질의에 대한 완벽한 응답

정성평가는 4명의 평가자가 진행하며 표 3은 정성평가의 연구자별 평점을 보여준다.

표 3. 정성평가의 성능

	기본 검색 모델	필터링 모델 추가
Human 1	3.069	3.55
Human 2	2.708	3.35
Human 3	3.335	3.6
Human 4	2.968	3.55
평균	3.02	3.51

표 3에서 기본 검색 모델은 평가자들의 평균이 3.02로 대부분의 문장이 응답이 될 수 있는 것이지만 필터링 모델을 적용할 경우 평가자들의 평균이 3.51로 향상되어 모델 응답의 품질이 상승되는 것을 알 수 있었다.

#### 4.4. 응답 필터링 성능 향상 예시

본 논문에서 제안한 색인화 정규화 및 응답 필터링이 적용된 결과는 [그림 6]과 같다.

	질의	응답
기본 검색 모델	정말 어려운 문제야.	누가 뭐랬나요?
필터링 모델 추가		잘 할 수 있을 거예요.

그림 6. 필터링 모델을 통해 정답 품질이 상승한 예

[그림 6]은 기본 검색 모델의 응답 결과와 응답 필터링이 적용될 경우 품질이 상승된 결과 예시를 보여준다. 질의 “정말 어려운 문제야.” 라고 입력했을 때 기본 검색 모델은 가장 유사한 문장으로 “정말이라고, 정말!” 이 매칭되어 그에 해당하는 “누가 뭐랬나요?” 라는 엉뚱한 답변을 출력한다. 하지만 필터링 모델이 적용된 경우 “정말이라고, 정말!” 은 필터링되고 “정말 어려

워.” 와 매칭 되어 “잘 할 수 있을 거예요.” 와 같이 질의와 어울리는 문장을 출력하게 해준다.

#### 5. 결론 및 향후 연구

본 논문에서는 개체명, 보조 용언, 시제를 통해 색인어를 정규화하고 검색 결과를 필터링하여 품질을 향상시키는 검색기반 채팅 모델을 제안하였다. 실험 결과 색인어 정규화로 높은 검색 커버리지를 확보했고 문장 임베딩을 이용한 응답 필터링을 통해 응답의 품질을 향상시킬 수 있었다. 향후 연구로 검색 모델을 통해 결과가 검색되지 않는 경우를 해결하기 위해 생성기반 채팅 모델을 결합한 하이브리드 채팅 모델을 연구할 예정이다.

#### 감사의 글

본 연구는 엔씨소프트 산학연구용역 과제의 지원을 받아 수행되었음.

#### 참고문헌

- [1] 김학수, “인공지능 음성언어 비서 시스템의 자연언어처리 기술들”, *정보과학회지*, 35.8, pp. 9-18, 2017.
- [2] A. B. Abacha and P. Zweigenbaum, Medical question answering: translating medical questions into sparql queries, *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. ACM, pp. 41-50, 2012.
- [3] A. R. Aronson and T. C. Rindfleisch, Query expansion using the UMLS Metathesaurus, *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, pp. 485, 1997.
- [4] 전원표, 송영길, 김학수, “채팅 모델 구현을 위한 3단계 문장 검색 방법”, *한국마린엔지니어링학회지 제37권 제2호*, pp. 205-212, 2013.
- [5] 이현구, 김민경, 김학수, “의학문서 질의응답을 위한 정답 스니핏 검색”, *정보과학회논문제 제43권 제8호*, pp. 927-932, 2016.
- [6] 안동연, “Corpus를 기반으로 하는 한국어 술어의 양상 생성”, *KAIST 박사학위논문*, 1995.
- [7] R. Blanco, H. Zaragoza, Finding support sentences for entities, *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 339-346, 2010.
- [8] Y. Bengio, Learning Deep Architectures for AI, *Foundations and trends® in Machine Learning*, 2(1), pp. 1-127, 2009.
- [9] M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, 45(11), pp. 2673-2681, 1997.