텍스트 기반 상담시스템의 효율성 제고를 위한 합성곱신경망을 이용한 자동답변추천 시스템

나훈엽○, 서상현, 윤지상, 정창훈, 전용진, 김준태
동국대학교, 컴퓨터 공학과
hoonyeob@dongguk.edu, shseo@dongguk.edu, js_yun@dongguk.edu,
gravity7508@dongguk.edu, yongjin117@dongguk.edu, jkim@dongguk.edu

Automated Answer Recommendation System Using Convolutional Neural Networks For Efficient Customer Service Based on Text

Hunyeob Na, Sanghyun Seo, Jisang Yun, Changhoon Jung, Yongjin Jeon, Juntae Kim
Dongguk University, Department of Computer Science

요 약

대면 서비스보다 비대면 서비스를 선호하는 소비자들의 증가로 인해 기업의 고객 응대의 형태도 변해가고 있다. 기존의 전화 상담보다는 인터넷에 글을 쓰는 형식으로 문의를 하는 고객이 증가하고 있으며, 관련 기업에서는 이와 같은 변화에 효율적으로 대처하기 위해, 텍스트 기반의 상담시스템에 대한 다양한 연구 및 투자를 하고 있다. 특히, 입력된 질의에 대해서 자동 답변하는 챗봇(ChatBot)이 주목받고 있으나, 낮은 답변 정확도로 인해 실제 응용에는 어려움을 겪고 있다. 이에 본 논문에서는 상담원이 중심이 되는 텍스트 기반의 상담시스템에서 상담원이 보다 쉽게 답변을 수행할 수 있도록 자동으로 답변을 추천해주는 자동답변추천 시스템을 제안한다. 실험에서는 기존 질의응답 시스템 구축에 주로 사용되는 문장유사도 알고리즘과 더불어 합성곱신경망을 이용한 자동답변추천 기법의 답변추천 성능을 비교한다. 실험 결과, 문장유사도 기반의 답변추천 기법보다 본 논문에서 제안한 합성곱신경망(Convolutional Neural Networks) 기반의 답변추천시스템이 더 뛰어난 답변추천 성능을 나타냄을 보였다.

주제어: 상담시스템, 자동답변추천, 합성곱신경망, 문장 분류

1. 서론

질의응답 시스템(Question Answering System)이란 사용자로부터 자연어로 구성된 질문을 입력받아 사용자가 원하는 답변을 자동으로 제공해주는 시스템이다. 최근에는 인공지능을 활용한 챗봇(Chat Bot)이 대두됨으로써 상담센터의 업무를 상당부분 대체할 수 있을 것으로 기대되고 있다. 실제로 은행이나 관공서를 방문하기보다는 전화, 혹은 인터넷을 통한 비대면 소통 채널을 선호하는 성향의 고객들이 늘어남에 따라, 국내외 다수 업체에서 챗봇 관련 서비스를 연구 개발하고 출시하여 서비스에 돌입하고 있다.

최근 많은 상담 센터에서는 상담원이 고객들의 다양한 상담을 전화 상담과 채팅을 통해 처리하고 있다. 채팅 상담의 경우 상담원 한명이 한 번에 처리할 수 있는 대화의 수가 제한적이기 때문에 시간, 비용적인 부분에서 효율적으로 처리하기 어렵다. 때문에 챗봇 서비스의 도입의 필요성이 대두되고 있다.

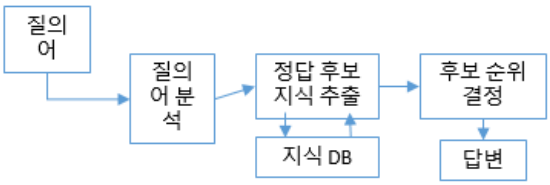
하지만 텍스트 생성모델의 제한과 규칙기반 챗봇의 한계점과 같은 문제로 인해 현실적으로 챗봇을 바로 실제 업무에 투입하기에는 어려움이 있다. 본 연구에서는 외식 주문 상담 업체에서 고객의 질의에 대한 자동 답변을 추천해주는 기능에 초점을 맞춰 주문 상담 센터의 업무 효율을 올릴 수 있는 질의응답 시스템을 설계하고 합성곱신경망을 활용하여 효율성을 높이는 방법을 제안한다.

본 논문의 2장에서는 질의응답 시스템의 구조와 자연어 처리 분야에서 사용되어온 문장 유사도 알고리즘에 대한 관련 연구, 그리고 Word2Vec과 합성곱신경망에 대하여 살펴보고, 3장에서는 기존 연구들의 단점과 본 연구에서 제시하는 모델에 대해서 설명한다. 4장에서는 사용될 실험 데이터를 소개하고, 기존의 알고리즘들과 제시한 모델을 사용한 실험 결과를 비교 분석한다. 마지막으로 5장에서 결론과 향후 연구 방향에 대해서 논한다.

2. 관련 연구

2.1 질의응답 시스템

과거에는 문장 유사도 기반의 다양한 알고리즘을 이용하여 질의응답 시스템을 구축하였다. [그림1]은 기존 연구 중 코사인 유사도를 사용한 질의응답 시스템의 모형이다. 우선 필요한 지식데이터를 구축한 뒤에, 질의어가 들어오면 해당 질의어를 분석하고 질의에 대한 정답 후보가 주어져 있을 때, 구축한 지식데이터를 이용하여 정답후보에 대한 지식을 추출하고 이를 통해 후보 순위를 결정하여 정답을 결정하는 방식이다.[1][2]



[그림1] 자연어처리와 정보검색을 이용한 질의응답 시스템

2.2 문장 유사도 알고리즘

기존의 질의응답 시스템에서는 피어슨 상관계수 (Pearson Correlation Coefficient), 자카드 (Jaccard), 코사인 (Cosine), 타니모토 (Tanimoto), 레벤슈타인 (Levenshtein) 등 문장 유사도를 비교함으로써 질문에 대한 답변을 제시하는 방법들이 많이 사용되었다.[3][4][5] 하지만, 이러한 방법들은 단어가 본질적으로 다른 단어와 어떤 관련성을 가지는지 이해 할 수 없다는 단점이 존재한다.

우선, 피어슨 상관계수는 두 변수간의 관련성을 구하기 위해 사용되는 개념이다. X와 Y가 함께 변하는 정도를 X와 Y가 따로 변하는 정도로 나누어 계산하며 공식은 다음과 같다.

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2}}$$

r값의 범위는 $-1 \le r \le 1$이며, r=1일 경우, x와 y가 완전한 상관관계에 있는 경우를 나타낸다. r값이 0일 경우, 두 변수는 완전한 독립관계이다. 일반적으로 r값이 0.1과 0.3 사이라면, 약한 선형관계, 0.3과 0.7 사이라면 뚜렷한 선형관계, 0.7과 1.0 사이라면 강한 선형관계이다.

자카드 유사도는 두 집합 A와 B를 비교할 때, 교집합의 원소를 전체 원소로 나눈 것으로 계산한다.

$$r = \frac{|A \cap B|}{|A \cup B|}$$

두 종류의 개체가 가지고 있지 않은 것을 제외하고, 동일한 특성이 많을수록 r값이 증가한다.

코사인 유사도는 내적 공간의 두 벡터의 유사도를 측정한다. 두 벡터간 각도의 코사인으로 측정하며, 두 벡터가 동일한 방향을 향하고 있는지 여부를 측정한다. 코사인 유사도는 문서를 비교하는데 사용되기도 한다. 코사인 유사도의 계산식은 다음과 같다.

$$\mathrm{sim}(x,y) = \frac{x \cdot y}{\|x\| \, \|y\|}$$

타니모토 계수는 정보수집에서 자주 사용되는 방법으로 계산식은 다음과 같다.

$$T = \frac{A \cap B}{A \cup B - A \cap B}$$

레벤슈타인 거리는 두 개의 문자열이 얼마나 유사한지 알아내는 알고리즘으로 문자열 A와 B가 존재할 때, A가 B와 같아지기 위해서 몇 번의 연산이 필요한지 계산하는 개념이다. 본 논문에서는 레벤슈타인 거리의 개념에 단어 기반으로 한 one-hot encoding 방식을 적용시킨다면 문장 유사도를 측정이 가능할 것이라는 가설을 세우고 실험을 수행하였다.

2.3 유사도기반 답변추천 시스템의 문제점

앞에서 살펴본 기존의 문장 유사도를 기반으로 하는 예측은 다음과 같은 한계점을 갖고 있다.

첫째, 두 사용자 프로필 간의 상관관계는 두 사용자가 평가 한 항목을 기반으로만 계산할 수 있다. 즉, 상관공식의 합계 및 평균은 두 사용자가 평가한 항목에 대해서만 계산된다. 사용자가 수천 개의 항목 중에서 선택하여 평가할 수 있는 경우, 두 사용자가 평가한 항목의 중복이 많은 경우가 거의 없다. 따라서 계산 된 상관 계수 중 상당수는 적은 수의 관측치를 기반으로 하기 때문에, 계산 된 상관관계를 신뢰할 수 있는 유사도 척도로 간주할 수 없다.

둘째, 상관관계 접근법은 등급의 클래스에 대한 별도의 모델이 아닌, 사용자 간의 유사성에 대한 하나의 글로벌 모델을 유도한다. 이러한 접근 방식은 두 개의 사용자 프로필이 양의 상관관계인지, 상관관계가 없는지, 또는 음의 상관관계인지를 측정한다. 그러나 한 사용자가 제공한 평점은 두 사용자 프로필이 상관관계가 없더라도 다른 사용자의 평점에 대한 좋은 예측 요소가 될 수 있다. 예를 들어, 사용자 A의 긍정적 인 등급이 사용자 B의 부정적인 등급에 대한 완벽한 예측자인 경우가 있을 때, 사용자 A의 부정 등급은 사용자 B의 긍정적 인 등급을 의미하지는 않는다. 즉, 2개의 프로파일 사이의 상관관계가 0에 가까울 수도 있고, 잠재적으로 유용한 정보가 손실된다.

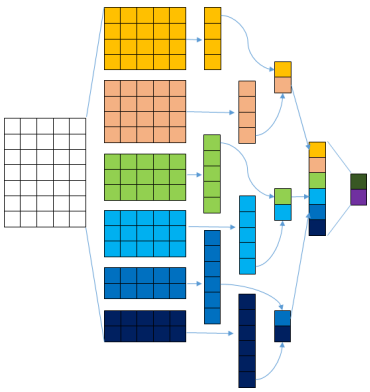
셋째, 두 명의 사용자가 등급이 서로 겹치면 유사 할 수 있다는 점이다. 즉 사용자가 공통 항목을 평가하지 않으면 해당 사용자 프로필은 상관관계를 가질 수 없다. 많은 도메인에서 평가할 수 있는 항목이 매우 많기 때문에 특히 시동 단계에서 많은 필터링 서비스에 방해가 될 수 있다. 그러나 사용자가 동일한 항목을 평가하지 않았다는 것을 발견하더라도 반드시 비슷한 생각을 하지는 않는다는 것은 아니다. 예를 들어, 사용자 A와 B는 높은 상관관계가 있고, B와 C도 서로 높은 상관관계가 있다. 이러한 관계는 사용자 A와 C 간의 유사성에 대한 정보도 제공한다. 하지만 사용자 A와 C가 공통 항목을 평가하지 않은 경우, 상관관계 기반 유사성 측정은 두 사용자 간의 관계를 발견하지 못한다. 이런 종류의 전이 유사성 관계를 발견 할 수 없다면 잠재적으로 유용한 정보가 손실된다.

2.4 Word2Vec

본 논문에서는 단어 임베딩 방법론 중 최근 많은 인기를 끌고 있는 Word2Vec을 활용한다. 단어 임베딩(Word Embedding)이란 고차원의 데이터를 그보다 낮은 차원으로 변환하면서 모든 데이터간의 관계가 성립되도록 처리하는 과정이다. 간단하게 정리하면 자연어 처리과정에서 단어를 벡터로 삽입하는 것이다. 다시 말해, 문자로 이루어진 단어를 숫자로 변환하는 것이다. 단어 자체를 아스키코드나 유니코드로 처리를 하여 사용해왔지만, 이것만으론 의미를 추론하기가 힘들었다. 단어 임베딩 기법에는 여러 모델이 있다. 초기모델인 NNLM모델과 RNNLM이 있고, 가장 최근에 발표되었고 현재 많은 인기와 동시에 많이 사용하는 Word2Vec모델이 있다. Word2Vec 모델을 사용하여 워드 임베딩을 진행하면 단어를 벡터화할 때 단어의 문맥적 의미를 보존하고 벡터로 바뀐 단어들은 코사인유사도와 같은 방식들로 그 거리를 잴 수 있고 단어사이의 거리가 가까울 경우 의미가 비슷한 단어끼리 벡터공간상에 맵핑되기 때문에 본 논문에서는 Word2Vec 모델을 활용하여 효과적으로 단어임베딩을 진행한다.[6]

2.5 합성곱신경망(Convolutional Neural Networks)

합성곱신경망은 이미지 분류와 컴퓨터 비전 시스템분야에서 활용이 되어왔지만, 최근 들어 자연어처리에 적용되기 시작하여 좋은 결과가 나오고 있다. 자연어 처리 분야에서는 이미지 픽셀대신, 행렬로 표현된 문장을 입력 값으로 받으며, 행렬의 각 열은 토큰, 일반적으로 단어에 대응된다. 각 행은 단어를 표현하는 벡터이다.



[그림2] 텍스트 처리를 위한 합성곱 신경망.

본 논문에서는 이러한 벡터를 Word2Vec을 활용하여 임베딩을 진행하여 입력하는데, 이것이 합성곱신경망에서의 이미지가 된다. 비전에서 필터는 이미지의 지역 조각을 슬라이딩하지만 자연어처리에서는 일반적으로 행렬의 전체 행을 슬라이딩 한다. 그러므로 필터의 너비는 보통 입력 행렬의 너비와 같고 높이 또는 지역의 크기는 변한다. 하지만 슬라이드 윈도우는 2~5단어가 일반적이다. 그림으로 설명하면 [그림2]와 같다.

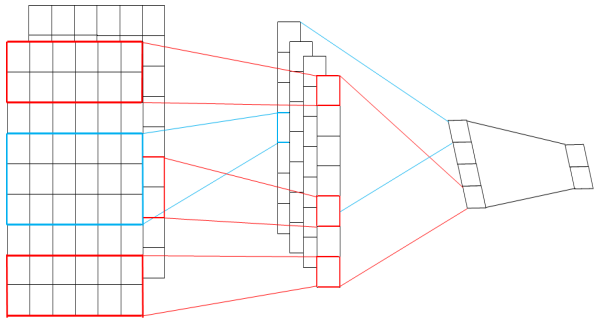
세 개의 필터 지역 크기는 2,3,4이다. 그리고 각각 두 개의 필터를 가지고 있다. 모든 필터는 문장 행렬에 대해서 합성곱을 수행하고 피쳐맵을 생성한다. 그 후 1-max-pooling이 전체 맵에 대해서 수행되어진다. 각 피쳐맵에서 가장 큰 수가 기록되게 된다. 그러므로 피쳐 벡터들은 모든 6개의 맵에서부터 만들어지고, 이 6개 피쳐 벡터들은 특성 벡터를 형성하기 위해 이어붙여진다. 마지막 소프트맥스 층은 이 특성 벡터를 입력으로 받아서 문장을 분류하는데 사용하게 된다.[7]

3. 합성곱신경망 기반 답변 추천 시스템

3장에서는 기존에 진행되었던 유사도기반 질의응답 시스템의 한계점, 그리고 이를 보완할 본 연구에서 사용할 Word2Vec과 합성곱신경망을 활용한 모델에 대해서 살펴본다.

3.1 Word2Vec과 합성곱신경망을 활용한 답변 추천

본 논문에서는 Yoon Kim이 제안한 Word2Vec과 합성곱신경망을 활용한 문장 구분 방법[8]을 활용하여 답변 추천 시스템을 만들어 사용한다. Yoon Kim이 제안한 모델의 구조는 [그림3]과 같다.



[그림3] Word2Vec + CNN 모델

위 모델은 우선 문장을 n*k 크기의 이미지로 변환을 한 뒤에 합성곱을 통해 피쳐맵을 만든다. 이렇게 만들어진 피쳐맵을 하나로 합친 뒤에, 완전 연결 층을 통해 결과 값을 만들어 낸다.

이 모델이 사용된 연구에서는 긍정과 부정으로 분류하는데 사용이 되었지만, 본 연구에서는 이렇게 나온 결과 값을 특성벡터로 활용하여 문장을 10개의 클래스로 분류하여 입력된 문장이 어떤 종류의 문의인지 구분하는데 사용한다. 이렇게 문의의 종류가 분류가 되면 해당 문의에 적합한 답변을 추천해줌으로써 상담원이 고객의 문의를 처리하는데 도움을 준다.

4. 실험

이 장에서는 앞서 소개한 5개의 유사도 기반 알고리즘을 활용한 답변 추천 시스템과 Word2Vec과 합성곱신경망을 사용한 답변 추천 시스템을 실험하여 그 결과를 비교한다.

4.1 실험 방법

데이터 셋은 [표1]처럼 400개의 데이터가 10개의 클래스로 각각 40개씩 나누어져있다. 데이터는 외식주문업체의 웹사이트에 등록되어 있는 FAQ 항목을 참고하고, 집단지성을 활용하여 유사 문장을 자체적으로 구축하였다. 이 데이터 셋을 3:1 비율로 학습에 사용될 트레이닝 데이터와 실험에 사용될 테스트 데이터로 나누었다.

No.	Input Data	Class
1	온라인 주문은 몇 시부터 가능하나요?	1
2	온라인 주문은 아침부터 가능하나요?	1
...	...	...
40	기프티콘 주문방법을 알려주세요.	2
41	제가 기프티콘을 가지고 있는데 어떻게 주문해야 되나요?	2
...	...	...
400	다른 매장으로 주문을 하고 싶어요.	10

[표1] : 실험에 사용된 데이터 셋

이러한 데이터 셋 앞에서 소개한 5개의 문장 유사도 기반 알고리즘과 이 논문에서 제시하는 Word2Vec과 CNN을 사용한 답변 추천 시스템을 사용해 실험을 진행하여 정확도를 계산한다. 정확도(Accuracy)는 다음과 같은 방법으로 산출하였다.

$$Accuracy = \frac{\text{클래스를 맞게 분류한 문장의 갯수}}{\text{전체 테스트 문장의 갯수}}$$

4.2 실험 결과 및 분석

Model	Accuracy
Pearson Coefficient	61%
Cosine Similarity Coefficient	63%
Tanimoto Coefficient	64%
Jaccard Coefficient	64%
Levenshtein Distance	57%
Word2Vec & CNN	84%

[표2]: 문장 유사도 알고리즘들과 Word2Vec & CNN 모델의 정확도 비교

위 표는 기존 문장 유사도 알고리즘을 사용한 모델과 본 논문에서 제시하는 모델의 정확도를 비교하여 보여준다. 실험결과, 유사도 기반의 알고리즘의 정확도는 약 60%를 살짝 넘어서는 수준이었으며, 레벤슈타인 거리는 57%로 가장 정확도가 낮았고, 연산에 걸린 시간도 약 6시간가량 소요되었다. 본 논문에서 제시하는 Word2Vec & CNN을 활용한 모델은 84%의 정확도를 보이면서, 다른 알고리즘들에 비해 약 20% 정도 향상된 정확도를 보여주고 있다.

5. 결론

본 논문에서는 Word2Vec과 합성곱신경망을 사용하여 답변 추천 시스템을 구현하여 기존의 문장유사도 기반 모델들과 비교하는 실험을 진행하였다.

실험을 통해, Word2Vec과 합성곱신경망을 사용한 답변 추천 시스템이 기존의 문장유사도 기반 모델에 비하여 20%정도의 정확도 향상이라는 결과물을 얻을 수 있었다.

본 연구에서는 Word2Vec을 사용함으로써 각 단어별 벡터를 학습시키는 것이 가능해짐으로써 수치화를 통한 단어의 개념적 구분이 가능해졌다. 여기에 합성곱신경망을 통해 좀 더 높은 정확도를 얻을 수 있었다.

향후 연구에서는 순환신경망을 사용한 질의응답 시스템을 연구하여 합성곱신경망을 사용한 방법과 비교 연구하는 것이 필요하다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음 (2016-0-00017)

참고문헌

[1] 박세영, 윤희근, 김다영, 김동건, 김민정, 신우석 “자연어처리와 정보검색을 이용한 질의응답 시스템,” 2015년 한국컴퓨터종합학술대회 논문집, 2015

[2] 이승우, 이근배 “유한패턴매칭을 이용한 자연어 질의응답 시스템”, 정보과학회지 22(4), 2004.4

[3] W.H.Gomaa and A.A.Fahmy “A Survey of Text Similarity Approaches,” International Journal of Computer Applications Volume 68-No.13, 2013

[4] V.U.Thompson, C.Panchev, M. Oakes “Performance Evaluation of Similarity Measures on Similar and Dissimilar Text Retrieval,” Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 2015 7th International Joint Conference on, 2015

[5] S.Minmin, Q. Dongmei “The Application of Levenshtein Algorithm in the examination of the Question Bank Similarity,” 2016 International Conference on Robots & Intelligent System, 2016

[6] T. Mikolov, K. Chen, G. Corrado, J. Dean “Efficient Estimation of Word Representations in Vector Space,” arXiv:1301.3781, 2013

[7] Y.Zhang and B.C.Wallace “A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification” arXiv:1510.03820, 2016.

[8] Yoon Kim “Convolutional Neural Networks for Sentence Classification,” 2014 Conference of Emirical Methods in Natural Language Processing(EMNLP), 2014.