

특집논문 (Special Paper)

방송공학회논문지 제24권 제1호, 2019년 1월 (JBE Vol. 24, No. 1, January 2019)

<https://doi.org/10.5909/JBE.2019.24.1.48>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 목적 지향 대화를 위한 효율적 질의 의도 분석에 관한 연구

김 학 동<sup>a)</sup>, 고 명 현<sup>a)</sup>, 임 헌 영<sup>a)</sup>, 이 유 림<sup>b)</sup>, 지 민 규<sup>c)</sup>, 김 원 일<sup>d)†</sup>

## Effective Text Question Analysis for Goal-oriented Dialogue

Hakdong Kim<sup>a)</sup>, Myunghyun Go<sup>a)</sup>, Heonyeong Lim<sup>a)</sup>, Yurim Lee<sup>b)</sup>, Minkyu Jee<sup>c)</sup>, and Wonil Kim<sup>d)†</sup>

### 요약

본 연구는 목적 지향 대화 시스템 내에서 단일 한국어 텍스트 형식의 질문으로부터 질의자의 의도를 파악하는 것을 목적으로 한다. 목적 지향 대화 시스템은 텍스트 또는 음성을 통한 사용자의 특수한 요구를 만족시켜주는 대화 시스템을 의미한다. 의도 분석 과정은 답변 생성에 앞서 사용자의 질의 의도를 파악하는 단계로, 목적 지향 대화 시스템 전체의 성능에 큰 영향을 준다. 생활화학제품이라는 특정 분야에 제안 모델을 사용하였고, 해당 분야와 관련된 한국어 텍스트 데이터를 이용하였다. 특정 분야에 독립적이며 범용적인 의도를 의미하는 화행과, 특정 분야에 종속적인 의도를 의미하는 개념열로 나누어 분석한다. 화행과 개념열을 분석하기 위하여 단어 임베딩 모델, 합성곱 신경망을 이용한 분류 방법을 제안한다. 단어 임베딩 모델을 통하여 단어의 의미정보를 추상화하고, 추상화된 단어의 의미정보를 기반으로 합성곱 신경망을 통하여 개념열 및 화행 분류를 수행한다.

### Abstract

The purpose of this study is to understand the intention of the inquirer from the single text type question in Goal-oriented dialogue. Goal-Oriented Dialogue system means a dialogue system that satisfies the user's specific needs via text or voice. The intention analysis process is a step of analysing the user's intention of inquiry prior to the answer generation, and has a great influence on the performance of the entire Goal-Oriented Dialogue system. The proposed model was used for a daily chemical products domain and Korean text data related to the domain was used. The analysis is divided into a speech-act which means independent on a specific field concept-sequence and which means depend on a specific field. We propose a classification method using the word embedding model and the CNN as a method for analyzing speech-act and concept-sequence. The semantic information of the word is abstracted through the word embedding model, and concept-sequence and speech-act classification are performed through the CNN based on the semantic information of the abstract word.

Keyword : Intent-Analysing, Goal-oriented Dialogue, CNN(Convolutional Neural Network), concept-sequence, speech-act

a) 세종대학교 디지털콘텐츠학과(Department of Digital Contents, Sejong University)

b) 세종대학교 인공지능언어공학과(Department of Artificial Intelligence and Linguistic Engineering, Sejong University)

c) 세종대학교 소프트웨어융합학과(Department of Software Convergence, Sejong University)

d) 세종대학교 소프트웨어학과(Department of Software, Sejong University)

† Corresponding Author : 김원일(Wonil Kim)

E-mail: [wikim@sejong.ac.kr](mailto:wikim@sejong.ac.kr)

Tel: +82-3408-3795

ORCID: <https://orcid.org/0000-0002-1489-8427>

※ 이 논문은 2017년도 정부(교육부)의 지원으로 한국연구재단 기초연구 사업의 지원을 받아 수행된 연구임(No.2015R1D1A1A01060693).

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2015R1D1A1A01060693).

· Manuscript received November 14, 2018; Revised December 26, 2018; Accepted January 9, 2019.

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

## I. 서 론

대화 시스템은 항공권/숙박 등의 예약, 특정 분야에 대한 지식 요구, 상품 주문/취소/교환 등의 상담과 같은 특정한 목적 달성이 요구되는 목적 지향 대화 시스템과, 표현을 가능한 한 인간답게 하여 사용자로 하여금 흥미를 느끼는 데에 목적을 둔 재미를 위한 대화 시스템으로 나뉜다. 최근 들어, Google의 Google Assistant, Amazon의 Alexa, Apple의 Siri와 같이 두 대화 시스템을 결합하여 특정목적을 달성 하며 자연스러운 대화도 수행 가능한 서비스가 제공되기 시작하였다<sup>[1]</sup>. Gartner 등의 리서치 기관에서는 이러한 인공지능 비서 서비스를 향후 몇 년 간 가장 주목해야할 인공지능 기술 서비스 중 하나로 보고 있다<sup>[2]</sup>.

대화 시스템은 일반적으로 자연어 이해 모듈, 대화 운영 모듈, 자연어 생성 모듈로 나뉜다. 자연어 이해 모듈에서는 질의의 의도를 파악하고, 대화 관리 모듈에서는 질의의 의도, 대화 맥락을 고려한 답변 선택이 이루어지며, 자연어 생성 모듈에서는 이전 단계로부터 얻어진 정보들을 바탕으로 문장을 생성하는 과정이 수행된다<sup>[3][4]</sup>.

본 논문에서는 생활화학제품이라는 특정 분야와 관련된 질의에 대응하는 목적 지향 대화 시스템의 전체과정 중 사용자 질의 의도 파악 과정에 대하여 다룬다. 2장에서는 본 연구와 관련된 선행 연구들에 대하여 논의한다. 3장에서는 본 연구에서 제안하고자 하는 분석 모델에 대하여 설명한다. 4장에서는 실험에 사용된 데이터들에 대하여 설명하고 실험의 과정과 결과에 대한 내용을 다루며, 5장에서 결론을 내린다.

## II. 관련 선행 연구

목적 지향 대화 시스템 내 사용자 질의 의도 파악과 관련된 연구의 최근 동향은 의도를 화행(Speech-act)과 개념열(Concept-sequence)로 나누어 파악하는 추세이다<sup>[5,6,7]</sup>. 화행은 특정 분야에 독립적인 범용적 의도를 의미하며, 개념열은 예약, 지식 요구, 메뉴 주문/취소/교환 등과 같이 특정 분야에 종속적인, 특수한 목적에 해당되는 의도를 의미한다.

목적 지향 대화 시스템의 초기연구에는 화행, 개념열과 같은 세부 계층의 의도를 고려하지 않은 연구가 주를 이루었으며, 대표적으로 Smith 외(1995)<sup>[8]</sup>가 있다. 이는 목적 지향 대화 시스템 내에서 사용자 질의에 대한 의도를 파악하는 연구로, 사용자의 다음 발화 문장과 의도를 미리 정의하고, 사용자의 입력이 이루어지면 정의된 문장들과 가장 관련 있는 음성 인식 후보를 선택함으로써 음성 인식의 오류를 줄이도록 한 연구이다. 또한 황재원 외(2008)<sup>[9]</sup>는 감정 자질을 이용하여 문장 또는 문서에 대한 감정 분류를 수행하는 연구이다. 지지벡터기계(SVM : Support Vector Machine)에 단어 벡터로 표현된 문서를 입력하여 긍정 또는 부정의 감정으로 이진분류 하였다.

한 편으로는, Reithinger 외(1995)<sup>[5]</sup>, 은종민 외(2005)<sup>[6]</sup>와 같이 대화 시스템 내 사용자의 발화에 대한 응답을 생성하는 데 중요한 단서 역할을 수행하는 화행에 대한 분석 연구가 있었다. [5]에서는 일정 생성 분야의 대화기록으로 구성된 Verbmobil 말뭉치에 대한 n-gram 정보를 활용하여 화행을 예측하는 연구를 수행하였고, 이를 통하여 탐색에 대한 복잡도를 나타내는 지표인 Perplexity를 감소시키는 성과를 보여주었다. [6]에서는 호텔 예약 분야의 말뭉치를 활용하여 품사 태그 정보, 품사 태그에 대한 Bigram 정보, 문맥 정보에 대한 카이 제곱 통계량 분석을 수행하여 자질을 추출하였다. 추출된 자질을 이용하여 지지 벡터 기계를 학습하고, 학습된 지지 벡터 기계를 이용하여 입력 발화에 대한 화행을 분류하였다. 그 결과 제안 시스템을 통하여 90%이상의 정확률을 얻는 성과를 보여주었다.

이현정 외(2008)<sup>[7]</sup>와 같이 화자의 의도를 파악하기 위하여 화행과 더불어 특정 분야에 종속적인 의도를 의미하는 개념열을 분석하는 연구 또한 있었다. [7]에서는 최대 엔트로피 모델(MEM : Maximum Entropy Model)을 이용하여 일정 관리 분야의 대화 데이터에 대한 화행과 개념열을 각각 동시에 분류하였다. 본 논문에서는 생활화학제품 분야와 관련된 문장 단위의 한국어 텍스트 데이터에 대한 발화 임베딩, 임베딩을 거친 문장 벡터에 대한 합성곱 신경망(CNN : Convolution Neural Network)의 학습을 통하여 화행과 개념열을 각각 분류하는 의도 분석 방안을 제안한다.

표 1. 학습 및 테스트용 데이터 셋 - 입력값

Table 1. Dataset for training and test - Input attribute

No	sentence	POS
1	제가 페녹시에탄올 제품에 알레르기가 있는데 아토세이프 밀크코튼을 사용해도 괜찮을까요? (I have a allergic to phenoxyethanol products. Is it OK with I use Atosafe Milk Cotton?)	[('제가', 'NNP'), ('페녹시에탄올', 'NA'), ('제품', 'NNG'), ... (이하생략)] ([(I', 'NNP'), ('phenoxyethanol', 'NA'), ('products', 'NNG'), ... (The rest is omitted)])
2	실수로 LG 생활건강 샤프란이 3ml 정도 입에 들어갔는데 병원에 가봐야 할까요? (Should I go to the hospital when I accidentally enter the mouth of LG Living Health Shafran about 3ml?)	[('실수', 'NNG'), ('로', 'JKB'), ('LG', 'SL'), ... (이하생략)] ([(accident', 'NNG'), ('-ly', 'JKB'), ('LG', 'SL'), ... (The rest is omitted)])
3	포름알데하이드 성분이 포함된 섬유유연제를 아기가 10분 정도 손으로 만졌는데 위험하가요? (Is it dangerous for baby to touch a softener containing formaldehyde ingredient during about 10 minutes?)	[('포름알데하이드', 'NA'), ('성분', 'NNG'), ('이', 'JKS'), ... (이하생략)] ([(formaldehyde', 'NA'), ('ingredient', 'NNG'), ('is', 'JKS'), ... (The rest is omitted)])
:		
1496	퓨어 키친 워셔 성분이 미산성차이염소산수라는데 안전한 성분인가요? (Pure kitchen washer ingredient is a weak acid hypochlorous acid. Is it a safe ingredient?)	[('퓨어', 'NNP'), ('키친', 'NNP'), ('워셔', 'NA'), ... (이하생략)] ([(Pure', 'NNP'), ('kitchen', 'NNP'), ('washer', 'NA'), ... (The rest is omitted)])

### III. 제안 방법

#### 1. 데이터 셋

화행 및 개념별 분류를 위하여 생활화학제품의 제품명이나 함유물질명을 의미하는 단어가 포함된 질의 문장, 문장의 형태소 분석 결과 및 품사 태깅 결과, 각 문장에 해당되는 정답 레이블을 학습 및 테스트용 데이터로 사용하였다. 본 논문에서 생활화학제품이란 “위해우려제품 지정 및 안전·표시기준, [시행 2018. 1. 22.] [환경부고시 제2018-12호, 2018. 1. 22., 일부개정], 별표 1”[10]에 명시되어 있는 위해 우려제품 23종을 의미한다. 전체 데이터 셋 1496개 인스턴스 중 1196개 인스턴스를 학습 데이터로 사용하고, 나머지 300개 인스턴스를 테스트용 데이터로 사용하였다. 표 1은

표 2. 학습 및 테스트용 데이터 셋 - 출력값

Table 2. Dataset for training and test - Output attribute

No	sa_label	cs_label
1	requesting (answer)	제품 위해도 판단 (judging product risk)
2	requesting (answer)	제품 위해도 판단 (judging product risk)
3	requesting (answer)	제품 및 성분 위해도 판단 (judging product and chemical component risk)
:		
1496	requesting (judgement)	제품 및 성분 위해도 판단 (judging product and chemical component risk)

전체 데이터 셋 중 질의 문장(sentence), 문장의 형태소 분석 결과 및 품사 태깅 결과(POS)를 표현한 것이며, 이 중 질의 문장은 제안 모델의 입력값으로, 형태소 분석 결과와 품사 태깅 결과는 제안 모델 내 단어 임베딩 모델의 입력값으로 사용된다. 표 2는 전체 데이터 셋 중 제안 모델의 출력값에 해당되는 화행 정답 레이블(sa\_label), 개념별 정답 레이블(cs\_label)을 표현한 것이다.

본 논문에서 사용할 화행 클래스는 Dore(1975)<sup>[11]</sup>의 화행 분류 체계를 기반으로, 설명이나 판단을 요구하는 목적 지향 시스템의 특성을 고려하여 정의하였다. 개념별 클래스는 자체적으로 생활화학제품 분야에 종속적인 의도를 고려하여 정의하였다. 최종적으로 분류되는 화행과 개념별의 클래스는 각각 9개, 7개이며 그 종류는 표 3, 표 4와 같다.

표 3. 화행 클래스

Table 3. Speech-act class

Class ID	Speech-act
1	labelling
2	repeating
3	answering
4	requesting (explaining)
5	requesting (judgement)
6	calling
7	greeting
8	protesting
9	practicing

표 4. 개념열 클래스  
Table 4. Concept-sequence class

Class ID	Concept-sequence
1	단어 정의 (word defining)
2	안전 물질 정보 (Safety material information)
3	유해 물질 정보 (Risk Material Information)
4	제품 위해도 판단 (judging product risk)
5	성분 위해도 판단 (judging chemical component risk)
6	제품 및 성분 위해도 판단 (judging product and chemical component risk)
7	none

## 2. 제안 모델

본 논문에서 제안하는 질의 의도 분석 모델은 그림 1과 같다. 입력된 질의 문장(Input)은 전처리 모델(Preprocessing Model)을 통하여 형태소 단위로 나뉘지며 각 형태소마다 품사 태그가 부착된다. 이후 형태소와 품사 태그는 임베딩 모델(Embedding Model)을 통하여 벡터 형태로 정형화되고, 정형화된 문장의 의미정보는 합성곱 신경망(Convolutional Neural Network) 기반 화행 분류기(Speech-act Classifier), 개념열 분류기(Concept-sequence Classifier) 각각의 입력값으로 사용된다. 최종적으로 두 분류기를 통하여 화행과 개념열이 9개의 화행 클래스, 7개의 개념열 클래스 중 하나로 각각 분류된다(Output).

여 화행과 개념열이 9개의 화행 클래스, 7개의 개념열 클래스 중 하나로 각각 분류된다(Output).

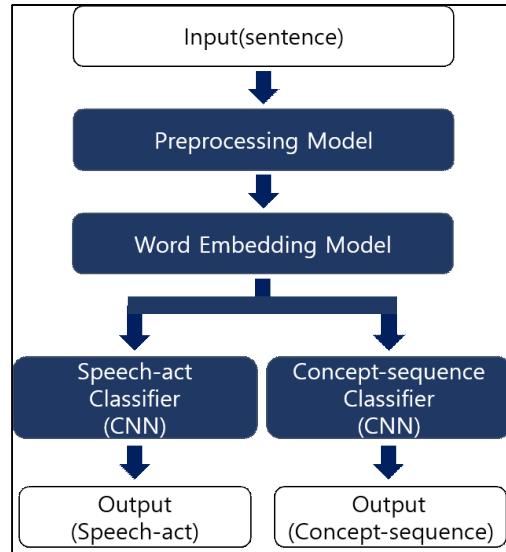


그림 1. 의도 분석 모델  
Fig. 1. Intent analysis model

## IV. 실험 결과 및 논의

### 1. 실험 진행

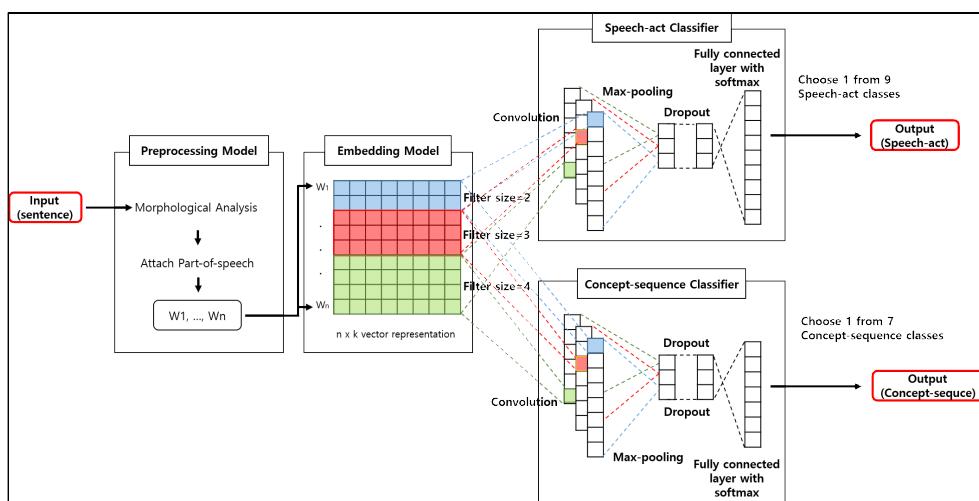


그림 2. 의도 분석 프로세스  
Fig. 2. Intent analysis process

제안 방법에 따른 실험을 위한 구체적 수행 절차는 그림 2와 같다.

### 1.1. 전처리

입력 문장(그림 2의 Input)을 단어 임베딩 모델에 입력하기 위하여 형태소 단위로 나누고 각 형태소에 품사 태그를 부착한다. 본 연구의 실험을 위한 전처리 과정에서는 Kkma 형태소 분석기를 사용하였다. 띄어쓰기 오류를 고려하여 입력 문장을 어절 단위로 쪼갠 뒤, 각 어절을 의미가 존재하는 동안 반복적으로 나누어 형태소를 얻는다. 이후 기분석 사전을 이용한 인접 조건 검사를 통하여 각 형태소의 품사를 파악하고, 해당 품사는 표 5와 같은 한글 형태소 품사-태그 표를 이용하여 태그로 맵핑되며, 최종적으로 품사 태그가 각 형태소에 부착된다(그림 2의 W1~Wn)<sup>[12]</sup>.

표 5. 한글 형태소 품사-태그 표 (일부분) [12]

Table 5. Korean morpheme part of speech-tag table (partial) [12]

대분류 (category)	KKMA 단일 태그 V1.0 (KKMA single tag V1.0)				...
	묶음 1 (gap 1)	묶음 2 (gap 2)	태그 (tag)	설명 (description)	
체언 (noun)	N	NN	NNG	보통명사 (common noun)	...
			NNP	고유 명사 (proper noun)	
			NNB	일반 의존 명사 (general dependent noun)	
			NNM	단위 의존 명사 (unit dependent noun)	
			NR	수사 (numeral noun)	
			NP	대명사 (pronoun)	
			:		

### 1.2. 단어 임베딩

입력 문장에 대한 전처리과정을 통하여 얻은 형태소, 품사 태그를 화행 분류기, 개념별 분류기에 입력하기 위하여 정형화한다. 본 연구의 실험을 위한 단어 임베딩 과정에서는 gensim 라이브러리의 Word2Vec 모델을 사용하였다. 해당 모델은 그림 3과 같이 각 형태소와 품사 태그를 one-hot-encoding 방식으로 입력받아 V차원 벡터로 표현된 중심 형태소 정보(그림 3의  $x$ )로부터 V차원 벡터로 표현된 인접 형태소의 정보(그림 3의  $y_1, y_2, \dots, y_c$ )를 예측하는 것을 목표로 하여(skip-gram 방식) 가중치 행렬  $W, W'$ 를 생성하는 방식으로 학습이 이루어진다. 이를 통하여 어휘의 의미(semant-

tic)정보와 구문(syntactic)정보를 정형화할 수 있기 때문에 자연어 처리 분야에서 타 임베드 방식에 비해 활용도가 높다<sup>[13]</sup>. 최종적으로 형태소와 품사 태그(그림 2의 W1~Wn)는 단어 임베딩 과정을 거쳐  $n \times k$ 차원의 벡터 형태(그림 2의  $n \times k$  vector representation)로 정형화된다.

단어 임베딩 실험 진행에 사용한 Word2Vec모델의 하이퍼 파라미터는 window size(한 번의 훈련/예측 동안 중심 형태소로부터 몇 개의 형태소를 훈련/예측할 것인지에 대한 파라미터로, 값이 5인 경우 중심 형태소로부터 양 옆 2개씩, 총 4개의 형태소를 훈련/예측하게 됨), dimension(하나의 형태소를 몇 차원의 벡터로 표현할 것인지에 대한 파라미터), epoch(훈련 반복 횟수 파라미터)이며, 최적의 결과를 나타내었던 하이퍼 파라미터 조합은 표 6과 같다.

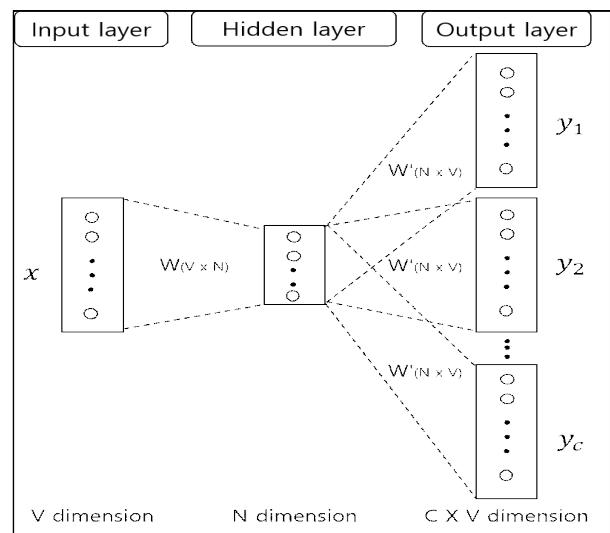


그림 3. Word2vec (skip-gram) 모델

Fig. 3. Word2vec (skip-gram) model

표 6. Word2Vec 모델의 하이퍼 파라미터

Table 6. Hyper parameters of Word2Vec model

Hyper parameters	Value
window size	5
dimension	200
epoch	50

### 1.3. 화행 및 개념별 분류

전처리 과정과 단어 임베딩 과정을 통하여 입력 문장을

벡터 형태로 정형화하였다. 이 정형화 결과(그림 2의  $n \times k$  vector representation)를 입력값으로 활용하여 화행과 개념열 분류를 수행한다. 분류를 위하여 합성곱 신경망(Convolution Neural Network)을 구축한다. 본 연구의 실험을 위한 합성곱 신경망 구축 과정에서는 Google의 텐서플로우(Tensorflow)를 이용하였다. 입력 벡터에 대하여 크기가 각각 2, 3, 4(그림 2의 Filter size)인 3종류의 필터를 각각 128 개씩 이용하여 크기 1의 간격(stride)으로 순회하는 방식의 합성곱 연산을 수행한다. 이를 통하여 필터 종류만큼의 자질 열(feature map)이 생성된다. overfitting 방지를 위하여 생성된 각 자질 열 요소 중 최댓값을 추출하는 최대 폴링 계층(Max-pooling layer)과 드롭아웃(Dropout) 계층을 사용하였고, 마지막으로 수식 1을 따르는 소프트맥스 계층(Softmax layer)을 통하여 각 클래스(화행의 경우 9개 클래스, 개념열의 경우 7개 클래스)들에 대한 정규분포 값을 얻는다. 최종적으로 모든 클래스 중 가장 큰 정규분포 값을 갖는 클래스를 선택하여 출력한다(그림 2의 Output). 본 세부 과정은 화행 분류와 개념열 분류 모두에 동일하게 적용된다.

화행 및 개념열 분류를 위하여 사용한 합성곱 신경망 모델의 하이퍼 파라미터는 filter size(합성곱 연산을 위한 필터의 크기를 의미하는 파라미터), stride(합성곱 연산 수행 시 필터가 순회하는 간격을 의미하는 파라미터), number of filters(필터의 개수를 의미하는 파라미터로, 값이 128이고 3종류의 필터를 사용한다면 각 필터의 개수가 128개, 총 개수는 384개), dropout(overfitting 방지를 위한 신경망 내 층 간 연결 시 일부 연결을 생략하는 것을 의미하는 dropout의 확률변수를 조정하는 파라미터로, 값이 0에 가까울수록 dropout이 많이 발생하여 생략되는 연결이 많고, 1에 가까울수록 dropout이 적게 발생하여 완전 연결에 가까워짐)이며, 최적의 결과를 나타내었던 하이퍼 파라미터 조합은

표 7. 합성곱 신경망 모델의 하이퍼 파라미터

Table 7. Hyper parameters of Convolutional Neural Network model

Hyper parameters	Value
filter size	2, 3, 4
stride	1
number of filters	128
dropout	0.8
epoch	50

표 7과 같다.

$$\text{Softmax}(x) = \frac{x_i}{\sum_{j=0}^k e^{x_j}}, (i = 0, 1, \dots, k) \quad (1)$$

## 2. 실험 결과

제안 모델의 성능을 평가하기 위한 평가지표는 재현율(Recall), 정밀도(Precision), F1-점수(F1-measure)이다. 정밀도(Precision)는 제안된 모델이 도출한 클래스 Cp 중 실제 클래스 Ca의 비율로, 수식 2를 따른다. 재현율(Recall)은 실제 클래스 Ca 중 제안된 모델이 올바르게 도출한 클래스 Cp의 비율로, 수식 3을 따른다. F1-점수는 수식 4와 같이 정밀도와 재현율의 조화평균의 값이다.

$$\text{Precision} = \frac{|Cp \cap Ca|}{|Cp|} \quad (2)$$

$$\text{Recall} = \frac{|Ca \cap Cp|}{|Ca|} \quad (3)$$

$$\text{F1-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

### 2.1. 제안 모델의 실험 결과

제안 모델의 최적화를 위하여 주요 하이퍼 파라미터를 변경해가며 실험을 반복 수행하였다. 파라미터 변화에 따른 화행 및 개념열 분류 결과는 표 8과 같다.

파라미터 변화에 따른 제안 모델의 실험 결과를 나타내고 있는 표 8에 따르면, 모델의 성능에 가장 큰 영향을 미치는 정형화 과정(Word2Vec)의 파라미터는 window size 파라미터이다. 중심형태소로부터 양 옆의 인접형태소를 1개씩만 학습하는 경우(window size 3)보다 2개씩 학습하는 경우(window size 5) 눈에 띄는 성능 차이를 보였으며, 값을 더 증가시키는 경우 미미하게 더 낮은 성능을 확인할 수 있었다. 보다 넓은 범위의 인접형태소를 학습함으로써 형태소의 위치 정보가 보다 풍부하게 고려될 수 있기 때문에 성능이 높아지는 것으로 보여 진다. 또한, 지나치게 넓은 범위의 인접형태소를 학습시키게 되면 하나의 window 범

표 8. 파라미터 변화에 따른 제안 모델의 실험 결과 (화행 / 개념열)

Table 8. Experiment result of proposed model according to parameter change (speech-act / concept-sequence)

Hyper parameters				Speech-act classification result			Concept-sequence classification result		
Word2Vec		CNN		Recall	Precision	F1	Recall	Precision	F1
window size	dimension	filter size	number of filters						
3	150	2,3	64	0.59	0.58	0.58	0.60	0.59	0.59
			128	0.62	0.61	0.61	0.62	0.62	0.62
			256	0.61	0.59	0.60	0.62	0.60	0.61
		2,3,4	64	0.62	0.64	0.63	0.63	0.66	0.64
			128	0.63	0.64	0.64	0.64	0.65	0.64
			256	0.64	0.66	0.65	0.66	0.65	0.66
	200	2,3	64	0.61	0.61	0.61	0.63	0.60	0.61
			128	0.64	0.63	0.63	0.64	0.62	0.63
			256	0.64	0.63	0.63	0.65	0.64	0.64
		2,3,4	64	0.63	0.64	0.63	0.63	0.65	0.64
			128	0.69	0.70	0.69	0.69	0.71	0.70
			256	0.69	0.69	0.69	0.70	0.69	0.69
5	150	2,3	64	0.71	0.69	0.70	0.72	0.69	0.70
			128	0.74	0.71	0.72	0.74	0.72	0.73
			256	0.73	0.72	0.72	0.73	0.73	0.73
		2,3,4	64	0.73	0.74	0.73	0.74	0.75	0.74
			128	0.78	0.78	0.78	0.79	0.79	0.79
			256	0.79	0.80	0.79	0.79	0.79	0.79
	200	2,3	64	0.74	0.73	0.73	0.75	0.74	0.74
			128	0.77	0.76	0.76	0.78	0.75	0.76
			256	0.77	0.75	0.76	0.79	0.77	0.78
		2,3,4	64	0.75	0.78	0.76	0.76	0.77	0.76
			128	0.80	0.82	0.81	0.82	0.84	0.82
			256	0.80	0.81	0.80	0.81	0.82	0.81

위 내 인접형태소의 수가 많아지게 되고, 이는 범위 내 같이 출현하는 인접형태소가 많아진다는 것을 의미하며, 그 결과 타 window 범위 내 인접형태소와의 구분이 어려워지기 때문에 지나치게 큰 값으로 설정하는 경우 성능이 낮아지는 것으로 보여 진다.

모델의 성능에 눈에 띄는 영향을 미치는 합성곱 연산(CNN)의 파라미터는 filter size 파라미터이다. 두 종류의 필터(filter size 2, 3)를 사용하는 경우보다 3종류의 필터(filter size 2, 3, 4)를 사용하는 경우 더 좋은 성능을 확인할 수 있었으며, 이보다 더 많은 종류의 필터를 사용하는 경우 성능이 더 낮아짐을 확인할 수 있었다. 이는 다양한 필터를 사용하게 되면 추상화 과정에서 더 많은 가중치가 사용되

고 그 결과 더욱 정교한 분류가 수행되기 때문에 성능이 높아지는 것으로 보여 진다. 또한 지나치게 많은 종류의 필터를 사용하는 경우, 사용되는 가중치가 지나치게 많아지게 되며 그 결과 모델의 복잡도가 높아져 성능이 낮아지는 것으로 보여 진다.

## 2.2. 제안 모델과 타 모델 간 비교 실험 결과

제안 모델의 성능을 확인하기 위하여 동일 실험 데이터셋에 대한 제안 모델의 실험 결과와 타 모델(SVM : Support Vector Machine<sup>[9]</sup>, MEM : Maximum Entropy Model<sup>[7]</sup>)의 실험 결과를 비교하였다. SVM 모델의 경우 본 연구와는 다르게 화행, 개념열의 세부 계층을 사용하지

표 9. 제안 방법과 타 모델들 간 성능 비교

Table 9. Performance comparison between proposed model and other models

Model (general intention)	General intention classification result							
	Recall		Precision		F1			
<b>SVM</b>	0.72			0.73				
Model (speech-act / concept-sequence)	Speech-act classification result			Concept-sequence classification result				
	Recall	Precision	F1	Recall	Precision	F1		
<b>proposed model(CNN)</b>	0.80	0.82	0.81	0.82	0.84	0.82		
<b>SVM</b>	0.62	0.61	0.61	0.74	0.74	0.74		
<b>MEM</b>	0.65	0.63	0.64	0.77	0.77	0.77		

않고 긍정 또는 부정에 대한 이진분류를 수행하는 모델임을 고려하여, 이진분류를 수행하는 경우와 세부 계층을 사용한 경우 모두 실험하였다. 표 9는 제안 모델과 타 모델 간 비교 실험 결과를 나타내고 있다.

제안 모델(proposed model)을 사용하여 화행과 개념열을 분류하는 경우의 성능은 화행, 개념열 모두에서 SVM 모델과 MEM 모델을 사용하였을 때의 성능보다 높음을 확인할 수 있었다. 이는 제안 모델이 SVM 모델, MEM 모델에 비하여 비선형 문제에 적합한 모델이기 때문인 것으로 보여 진다.

SVM 모델은 세부 계층을 분류하는 경우(speech-act / concept-sequence)보다 그렇지 않은 경우(general intention) 더 높은 성능을 확인할 수 있었다. 이는 SVM 모델이 이진 분류와 같은 선형 문제를 수행하는 데 적합하기 때문인 것으로 보여 진다. SVM 모델이 9개의 화행 클래스, 7개의 개념열 클래스와 같이 세부 계층을 분류하는 경우 성능이 낮아지는 것을 확인할 수 있었다. 긍정 또는 부정에 대한 이진 분류와 같은 선형 문제 수행 시 요구되는 모델의 복잡도보다 세부 계층을 분류하는 비선형 문제 수행 시 요구되는 모델의 복잡도가 높고, SVM 모델은 제안 모델에 비하여 모델의 복잡도가 낮다. 따라서 제안 모델이 SVM 모델보다 화행 및 개념열 분류에 있어 더 높은 성능을 도출하는 것으로 보여 진다.

MEM 모델에 의한 세부 계층 분류의 경우 SVM 모델에 비하여 높은 성능을, 제안 모델에 비하여 낮은 성능을 보임을 확인할 수 있었다. 이진분류와 같은 선형 문제에 적합한 로지스틱 회귀 분석 모델을 다중 클래스 분류와 같은 비선형 문제도 해결할 수 있도록 softmax 회귀 분석 모델을 결

합하여 구축한 MEM 모델이 SVM 모델보다 다중 클래스 분류에 적합하기 때문에 더 높은 성능을 도출하는 것으로 보여 진다. 그러나 MEM 모델은 기존 로지스틱 회귀 분석 모델에 softmax 회귀 분석 모델을 추가하는 방법으로 구축되었기 때문에, 이진분류와 같은 선형 문제에 적합한 기존 로지스틱 회귀 분석 모델 자체의 특성은 변화가 없다. 따라서 비선형 분류에 효율적인 제안 모델과 MEM 모델이 화행 및 개념열 분류에 있어 성능 차이가 있는 것으로 보여 진다.

## V. 결 론

본 논문에서는 사용자의 의도를 예측하는 인공신경망 모델을 제안하였고 생활화학제품이라는 특정 분야에 제안 모델을 사용하였다. 사용자의 의도를 분야에 독립적이고 범용적인 화행과 분야에 종속적인 개념열로 정의하였다. 관련 선행 연구에서는 화행, 개념열과 같은 세부적 의도를 다루지 않은 연구, 문서에 대한 감정 이진분류를 수행한 연구, 화행만을 고려한 연구, 타 기계학습 모델을 사용한 화행, 개념열 분류 연구에 대하여 논의하였다.

연구 수행을 위하여 1496개의 생활화학제품 분야의 질의 문장 데이터를 활용하였으며, 입력 질의에 대한 형태소 분석과 품사 부착, 단어 임베딩 모델을 통한 질의 문장에 대한 정형화, 합성곱 신경망을 사용한 화행 및 개념열 분류 실험을 수행하고 그 결과를 살펴보았다. 생활화학제품 분야와 관련된 사용자 질의 의도 분석 실험에서 제안된 모델은 타 기계학습 모델에 비하여 재현율, 정밀도, F-1 점수가 높음을 확인할 수 있었다. 더 많은 학습 데이터 셋을 확보하면

성능을 보다 높일 수 있을 것으로 예상된다. 향후에는 사용자의 이전 발화를 고려한 의도 분류를 수행할 예정이다. 대화 시스템은 일반적으로 사용자와 시스템 간의 발화 교환이 여러 번 이루어지기 때문에, 이전 발화를 고려함으로써 화행 및 개념열의 예측이 더욱 정확해질 것으로 예상된다.

### 참 고 문 헌 (References)

- [1] Kyung-soon Lee, Jin-xia Huang, Oh-woog Kwon, Young-kil Kim. "A Chatter Bot for a Task-Oriented Dialogue System," KIPS Transactions on Software and Data Engineering6, No.11, 499-506, 2017.
- [2] Oh-Woog Kwon, Teakgyu Hong, Jin-Xia Huang and Young-Kil Kim, "An Analysis for Dialogue Processing Technologies and Service Trends of Virtual Personal Assistants," Communications of the Korean Institute of Information Scientists and Engineers, Vol.35, No.8, pp.19~27, 2017.
- [3] Hyun-Jung Lee, Analysis and Prediction of Speakers' Intentions in a Dialogue-based NLIDB, PhD's Thesis of Sokang University, Seoul, South Korea, 2014.
- [4] Natural language processing technology for dialog system development, [https://m.blog.naver.com/PostView.nhn?blogId=naver\\_search&logNo=221027662050&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2F](https://m.blog.naver.com/PostView.nhn?blogId=naver_search&logNo=221027662050&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2F) (accessed Nov. 6, 2018).
- [5] N. Reithinger and E. Maier, "Utilizing Statistical Dialogue Act Processing in VERBMOBIL," Proceedings of the 33rd annual meeting on Association for Computational Linguistics, Cambridge, Massachusetts, pp.116-121, 1995
- [6] Jong Min En, Song Wook Lee, Jung Yun Seo, "An analysis of Speech Acts for Korean Using Support Vector Machines," The KIPS Transactions : Part B, Vol.12, No.3, pp.365~368, 2005
- [7] Lee, Hyunjung, Kim, Harksoo, Seo, Jungyun, "Domain action classification using a maximum entropy model in a schedule management domain," AI Communications, Vol.21, No.4, pp.221~229, 2008
- [8] R. W. Smith and D. R. Hipp, Spoken Natural Language Dialog Systems: a Practical Approach, Oxford University Press Inc, Oxford, United Kingdom, 1995.
- [9] Hwang, Jaw-Won; Ko, Young-Joong, "A Korean Sentence and Document Sentiment Classification System Using Sentiment Features," Korean Institute of Information Scientists and Engineers, Vol.14, No.3, pp.336~340, 2008
- [10] National Law Information Center, <http://www.law.go.kr/LSW/admRulLsInfoP.do?admRulSeq=210000110550#AJAX> (accessed Nov. 6, 2018).
- [11] Dore, J, "Holophrases, speech acts and language universals," Journal of Child Language, Vol.2, No.1, pp.21-40, 1975.
- [12] KKMA morpheme analyzer, <http://kkma.snu.ac.kr/> (accessed Nov. 6, 2018).
- [13] Y. Goldberg, O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method", arXiv preprint arXiv: 1402.3722, 2014.

---

### 저 자 소 개

#### 김 학 동



- 2016년 : 경성대학교 컴퓨터공학과 학사
- 2017년 ~ 현재 : 세종대학교 디지털콘텐츠학과 석,박사통합과정
- ORCID : <https://orcid.org/0000-0003-3816-1224>
- 주관심분야 : 머신러닝, 딥러닝, 자연어처리

#### 고 명 현



- 2016년 : 세종대학교 디지털콘텐츠학과 학사
- 2016년 ~ 현재 : 세종대학교 디지털콘텐츠학과 석사과정
- ORCID : <https://orcid.org/0000-0002-6036-4717>
- 주관심분야 : 텍스트 마이닝, 기계학습, 딥러닝

---

## 저자 소개

---

### 임현영



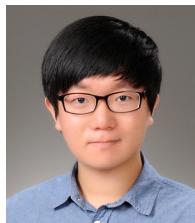
- 2017년 : 세종대학교 디지털콘텐츠학과 학사
- 2017년 ~ 현재 : 세종대학교 디지털콘텐츠학과 석사과정
- ORCID : <https://orcid.org/0000-0002-8547-6248>
- 주관심분야 : 컴퓨터 비전, 기계학습, 딥러닝

### 이유림



- 2018년 : 세종대학교 디지털콘텐츠학과 학사
- 2018년 ~ 현재 : 세종대학교 인공지능언어공학과 석사과정
- ORCID : <https://orcid.org/0000-0001-8309-090X>
- 주관심분야 : 텍스트 마이닝, 자연어 처리, 딥러닝

### 지민규



- 2018년 : 세종대학교 천문우주학과 학사
- 2018년 ~ 현재 : 세종대학교 소프트웨어융합학과 석사과정
- ORCID : <https://orcid.org/0000-0002-3089-1452>
- 주관심분야 : 텍스트 마이닝, 기계학습, 딥러닝

### 김원일



- 1981년 12월 ~ 1985년 7월 : (주)대한항공 전산실 재무 시스템 개발원
- 1982년 : 한양대학교 공과대학 금속공학 학사
- 1987년 : 미국 일리노이주 서던일리노이대학교 컴퓨터 과학 학사
- 1990년 : 미국 일리노이주 인디애나주 대학교 대학원 컴퓨터 과학 석사
- 1994년 : 미국 인디애나주 인디애나 대학교 대학원 컴퓨터 과학 전공
- 2000년 : 미국 뉴욕주 시러큐스 대학교 대학원 컴퓨터 & 정보과학 공학 박사
- 2000년 1월 ~ 2001년 3월 : 미국 펜실베니아주 외인시 소재 Bhasha, INC Technical Staff (연구원)
- 2002년 3월 ~ 2003년 8월 : 아주대학교 정보통신전문대학원 BK 교수
- 2003년 9월 ~ 2017년 2월 : 세종대학교 전자정보공학대학 교수
- 2017년 3월 ~ 현재 : 세종대학교 소프트웨어융합대학 교수
- ORCID : <https://orcid.org/0000-0002-1489-8427>
- 주관심분야 : 인공지능, 지능형 시스템, 딥러닝 등