



Korean Medicine Symptom Recommendation Based on Textual Description of the Patient's Condition. (Digestive disorder version)

Ho Jang¹, Sang-Hun Lee², Sangjun Yea¹

¹Intellectual Information Team, Korea Institute of Oriental Medicine

²Future Medicine Division, Korea Institute of Oriental Medicine

A B S T R A C T

We developed a suggestion system that recommends related Korean Medicine symptoms from natural language description of patient's health condition. First, we collected medical information from 200 patients who suffer digestive disorder. Then, three groups of Korean medicine doctors belonging to each of the major Korean medicine associations examined clinical report forms (CRFs) of the patients and specified primary symptoms of the patients. The Korean medicine doctors also attached text evidences of their decision based on the descriptions in the CRFs. Those text descriptions and the symptoms were used as the input text and the label of the system. The features of the multi label support vector machine are term frequencies of the input text. Those terms are the morphemes from Korean morphological analyzer. The model performance measure was the area under cover (AUC) values that is calculated based on the predicted ranks of the true symptom. All AUC values of the models from three independent datasets generated by the three different groups of Korean medicine doctors are more than 0.85 which suggests our system shows consistent and robust performances regardless of the given datasets. Our study is an artificial intelligence (AI) application to Korean medicine and our approach is useful to explore proper symptom names from the thousands of Korean medicine terms. It can be expanded to application to other kinds of disorders easily. Better performance can be achieved by adapting advanced text processing technique and AI models.

© 2020 KKITS All rights reserved

KEYWORDS: CDSS, Text mining, Text classification, Korean medicine, Symptom recommendation

ARTICLE INFO: Received 16 September 2020, Revised 29 September 2020, Accepted 13 October 2020.

*Corresponding author is with the Intellectual Information Team, Korea Institute of Oriental Medicine, Korea Institute of Oriental Medicine, 1672

Yuseong-daero, Yuseong-gu, Daejeon, 34054, Korea.
E-mail address: tomita@kiom.re.kr

1. 서론

최근 인공지능 기법이 다양한 분야에 적용되고 있다. 의료분야에서는 의료 종사자의 의사결정을 도와주는 임상 의사결정 지원 시스템(Clinical decision support system, CDSS)에서 인공지능 기술을 활용하는 비지식기반(Non-knowledge-based) 시스템이 주요하게 다루어지고 있다[1]. 이러한 기술의 발전은 의료 전문가가 효율적이면서도 정확한 의료서비스를 할 수 있도록 하고 결과적으로 환자들의 의료 복지에 기여하게 된다.

한의학은 수천 년 계승되어 발전되어온 진단 및 치료 기술이다. 한의학은 망문문절(望聞問切)을 통해 다양한 질환에 대한 진단을 수행하고 약물과 침구 등을 이용해 치료를 수행하고 있다. 한의학에서도 CDSS 및 인공지능 기법을 도입하여 진단 및 치료의 질을 향상 시키려는 시도가 수행되어오고 있다. 김상균 외 연구자들은 한의 서적으로 부터 생성된 온톨로지에 기반해서 증상 조합에 대한 한의 처방을 추천해주는 시스템을 개발 했다[2]. 정원모 외 연구자들은 Artificial neural network (ANN) 모델을 이용해서 환자의 증상 조합에 기반해서 침 자리를 추천해주는 연구를 수행했다[3]. Xiaoyu Mi 외 연구자들은 ANN 모델을 이용해서 증상 조합에 기반해서 한약 처방을 추천해주는 모델을 개발 했다[4].

이처럼 다양한 추천 시스템 및 기법들이 증상명을 입력으로 받아서 수행되고 있다. 그러나 한의학은 오랜 시간에 걸쳐 발전되어왔기에 한의학 용어들은 매우 다양하며 많은 동의어가 있다. 덧붙여 한의학에는 다양한 이론적 기반을 둔 학파들이 있고 학파별로 사용하는 용어들에게는 다소 차이가 있다. 이런 점이 증상에 기반한 추천 시스템의 연구 및 사용자 이용의 접근성을 저해하고 있다.

텍스트 분류(Text classification) 기법은 자연어

입력이 특정 군(Class)에 포함되는지 여부를 자동으로 판단하는 기법으로, 스팸 메일을 필터링 하거나 감성 분석을 하는 등 여러 도메인에 적용이 되어 왔다[5-6]. Naive Bayesian, SVM (Support vector machine) 방법 등이 현재의 딥러닝이 대두된 이후에도 널리 사용되고 있다. 그리고 SVM은 모델의 특징(Feature)의 개수가 많고 Sparse 한 특성을 가지는 텍스트 데이터의 분류 문제에 적합하여 여러 텍스트 분류 모델 비교 연구에서 우수하고 안정된 성능을 보였다[7].

증상 입력에 기반한 추천 시스템에서의 사용자 접근성을 개선하기 위해서, 한글 형태소 분석 및 SVM을 사용하는 텍스트 분류 기법을 적용하여 자연어 서술 기반 한의 증상명 추천 시스템을 개발 하였다. 환자의 상태에 대한 자연어 입력으로 부터 관련된 증상 용어들을 자동 추천하여, 의료종사자가 한의 증상명을 탐색하는데 도움을 주어 의료의 질을 향상시킬 수 있을 것으로 기대된다.

2. 연구 방법

<그림 1>는 연구의 전 과정을 보여준다. 임상에서 환자의 증상과 관련된 자연어로 기술된 원시 데이터가 수집 되었고, 각 한의학과 한의사들이 환자의 질병과 관련성이 높은 증상들을 선별해서 AI모델의 학습 데이터를 만들었다. 시스템의 입력으로 사용되는 자연어 데이터는 데이터 전처리 과정을 거쳐서 모델에서 사용할 수 있는 형태로 변환이 되었다.

학파의 한의사가 지정한 증상명이 모델에서 얼마나 높은 순위로 예측 되었는지를 Area under cover (AUC)를 이용해서 계산하고 이를 모델 성능 평가에 사용하였다. 모델은 5-fold cross validation 으로 평가되었다. 데이터 처리, 시스템 구축, 성능 평가를 위해 Python과 R이 사용되었다.

2.1 원시 데이터 생성

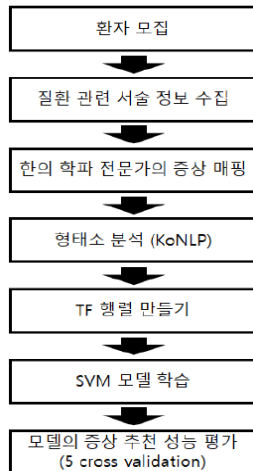


그림 1. 연구 절차
Figure 1. Development process

연구 참여자는 소화기계 증상을 호소하는 환자로, 부산대학교 한방병원에서 모집이 되었다.* 한의사가 환자들에게 질환에 대한 문진을 한 후, 문진 결과를 텍스트 데이터로 생성하는 작업을 수행하였다. 총 200명의 환자에 대해 원시 데이터(Raw data)가 수집되었다. 이후 주요한 한의학파에서 선별된 한의사들이 원시 자연어 데이터를 분석해서 2차 데이터를 생성했다.

구체적으로, 환자의 질환과 관련 있는 증상들이 학파 한의사들에 의해서 선정 되었다. 또한 한의사들은 그 증상과 관련된 자연어 서술 부분을 원시 데이터로부터 발췌하였다. 사상체질의학회, 대한형상의학회, 대한상한금궤의학회의 3개의 학파에서 각각 한의사 3인의 합의에 의해서 환자의 질환과 관련된 증상들이 추천되었다. 3개의 학파 별로 데이터셋이 만들어졌다.

* 환자는 부산대학교 한방병원 연구윤리심의위원회 승인 (승인번호 : PNUKH-IRB-2019005)하에 모집되었다.

2.2 학습 데이터 생성

증상의 추천은 텍스트 분류 모델(Text classification model)을 사용하여 이루어진다. 분류 모델은 Term frequency (TF) 행렬을 입력으로 사용한다.

행(Row)는 각 데이터이고 열(Column)은 형태소이다. 행렬의 값은 형태소가 한 번도 등장하지 않으면 0, n번 등장하면 n이다. 예를 들어 i번째 데이터가 “갈증을 자주 느낀다”이고 j번째 형태소가 ‘갈증’이면 TF 행렬의 i행의 j열의 값은 1이다. 행렬의 구성을 위해 Python의 KoNLPy 패키지의 Twitter 형태소 분석기가 사용되었다. 명사, 동사, 영단어, 숫자 등이 형태소로 추출 되었다. 5-fold cross validation으로 성능 측정을 위해서 최소 10개의 데이터를 가진 증상들만 학습 및 평가에 사용했다.

2.3 분류 모델 적용 및 성능측정

시스템은 다양한 증상을 판별해야하므로 Multi-label classification이 가능한 SVM모델이 사용되었다. Label은 증상이다. 증상 간의 비교를 위해 One-vs-one decision function이 사용되었다. 성능 평가를 위해 5-fold cross validation이 사용되었고, 각 증상에 해당하는 데이터들이 Fold별로 고르게 샘플링 되도록 계층적 샘플링을 사용했다. Fold별로 약 80%의 데이터가 모델 생성을 위한 학습 데이터로 사용되었고 나머지 20%가 테스트 데이터로 사용되었다. 학습된 모델은 입력에 대해 각 증상일 확률 값을 반환한다. 자연어 입력에 대한 전문가들에 의해 레이블링(Labeling)된 증상명이 모델에서 얼마나 높은 순위로 예측 되었는지를 아래의 수식을 이용해서 계산하고 이를 성능 평가에 사용하였다.

$$AUC = \frac{\sum_{i=1}^N \sum_{j=1}^M AUC(i, j)}{N \times M} \quad (1)$$

N은 테스트 데이터의 총 개수, M은 증상의 개수이다. $AUC(i, j)$ 는 다음과 같다.

$$AUC(i, j) = \sum_{k=1}^j I(model(i, k) = label(i)) \quad (2)$$

$model(i, k)$ 은 i 번째 테스트 데이터를 모델에 넣었을 때, 모델이 반환한 k 번째로 높은 확률 값에 대한 증상명이다. $label(i)$ 는 i 번째 테스트 데이터에 대해서 한의사들이 지정한 증상명이다. I 는 indicator 함수로 $model(i, k)$ 과 $label(i)$ 이 같으면 1을 다르면 0을 반환한다. 즉 i 번째 입력 데이터를 학습된 모델에 넣었을 때 모델이 반환한 j 번째 추천 증상까지 봤을 때 한의사들이 선정한 증상이 포함되면 1, 아니면 0의 값을 반환한다. 임의로 생성한 데이터에 의한 AUC값은 0.5이므로, 모델이 0.5 이상의 값을 반환하면 이것은 모델의 추천 성능이 무작위 결과 이상임을 함의한다. Python 패키지 Sklearn을 이용해서 데이터 샘플링, 모델 생성, 성능평가가 수행되었다.

3. 결과

3.1 데이터 기초 통계

모집된 환자들로부터 사상체질의학회 한의사들은 628개, 대한형상의학회는 802개, 대한상한금궤의학회는 468개의 자연어서술-증상명 쌍 데이터를 만들었다. 이 데이터들에 대해서, 사상체질의학회는 125개, 대한형상의학회는 164개, 대한상한금궤의학회는 167개의 구별되는 증상명을 이용해서 환

자의 증상을 지정하였다. 증상별 데이터 개수는 사상체질의학회는 평균 5개, 중앙값은 1개, 대한형상의학회는 평균 5개, 중앙값은 1개, 대한상한금궤의학회는 평균 3개, 중앙값은 2개였다.

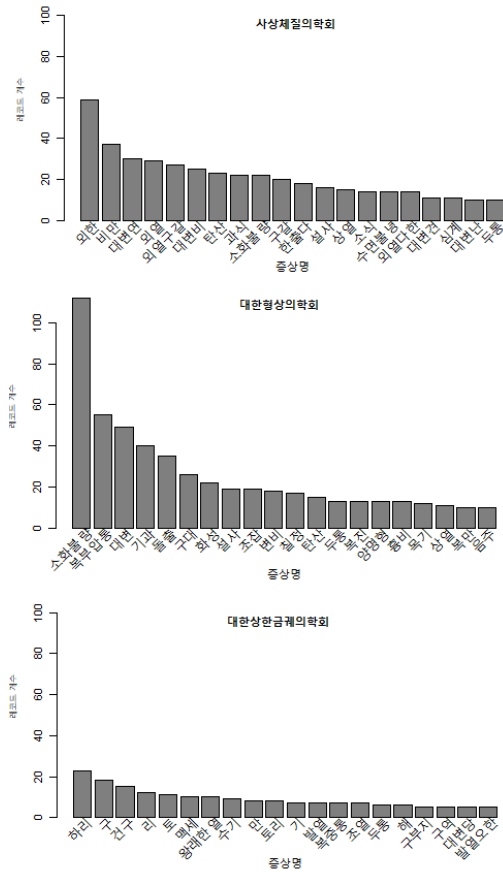


그림 2. 한의 학과별 증상별 데이터 개수
Figure 2. Number of data by symptom

<그림 2> 는 각 학과에서 사용한 증상 용어들을, 데이터 개수가 많은 순으로 상위 20개를 나열한 것이다. 사상체질의학회의 경우는 ‘외한’을 가지는 데이터의 수가 59개로 가장 많았고, 대한형상의학회는 ‘소화불량’이 112개로 가장 많았고, 대한상한금궤의학회는 ‘하리’가 23개로 가장 많았다.

이후의 과정에서는 모델 학습을 위해서 데이터가 10개 이상인 증상들로 그리고 한의사들이 증상 관련 자연어 기술을 근거로 달아준 증상들로 한정했다. 사상체질의학회의 경우는 20개, 대한형상의학회의 경우는 14개, 대한상한금궤의학회의 6개의 증상이 남았다. <표 1>은 실제 모델의 개발에 사용된 증상별 데이터 개수 및 총 데이터 개수를 보여준다. 데이터의 총 개수는 사상체질의학회의 경우는 426개, 대한형상의학회의 경우는 364개, 대한상한금궤의학회의 경우는 87개였다. 대한상한금궤의학회의 경우는 총 167개의 증상들 중 데이터 개수가 10개 이상인 증상에 단 6개였기 때문에 468개 중 87개의 데이터만 연구모델 구축 및 평가에 사용되었다.

표 1. 학과별 증상별 데이터 개수
Table 1. Number of data by symptom

증상명 (사상체질의학회)	개수	증상명 (대한형상의학회)	개수	증상명 (상한금궤의학회)	개수
의한(畏寒)	59	소화불량 (消化不良)	112	하리(下利)	23
비만(痞滿)	37	복부압통 (腹部壓痛)	55	구(嘔)	18
대변연(大便軟)	30	대변(大便)	49	건구(乾嘔)	15
의열(畏熱)	29	설사(泄瀉)	19	토(吐)	11
의열구갈 (畏熱口渴)	27	변비(便秘)	18	맥세(脈細)	10
대변비(大便秘)	25	칠정(七情)	17	왕래한열 (往來寒熱)	10
탄산(吞酸)	23	탄산(吞酸)	15		
과식(過食)	22	두통(頭痛)	13		
소화불량 (消化不良)	22	복진(腹診)	13		
구갈(口渴)	20	흉비(胸痞)	13		
한출다(汗出多)	18	복만(腹滿)	10		
설사(泄瀉)	16	상열(上熱)	10		
상열(上熱)	15	음주(飲酒)	10		
소식(少食)	14	조잡(嘈雜)	10		
수면불녕 (睡眠不寧)	14				
의열다한 (畏熱多汗)	14				
대변건(大便堅)	11				
대변난(大便難)	10				
두통(頭痛)	10				
심계(心悸)	10				
총	426	총	364	총	87

3.2 전반적인 성능 평가

<그림 3>은 사상체질의학회의 데이터에 대한 모든 증상을 아우르는 전반적인 성능 및 Fold별 성능을 보여준다. X축은 모델이 반환한 추천 증상을 어디까지 고려할지를 의미하고, Y축은 그때 각 학과의 한의들이 지정한 증상이 모델의 추천 증상에 포함되는 경우의 총 개수를 의미한다. 각 Fold에서의 테스트 데이터의 개수는 85개 이상이고, AUC값은 모두 0.92이상 이었다. 모든 결과를 합한 AUC값은 0.93이었다.

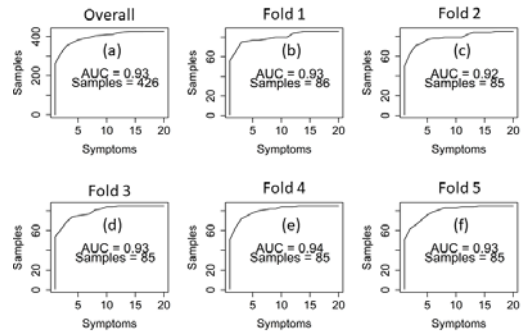


그림 3. 전반적인 성능 (사상체질의학회)
Figure 3. Overall performance

<그림 4>는 대한형상의학회의 데이터에 대한 성능을 보여준다. 각 Fold에서의 테스트 데이터의 개수는 72개 이상이고 AUC값은 모두 0.92이상 이었다. 모든 결과를 합한 전반적인 성능은 0.94였다.

<그림 5> 대한상한금궤의학회의 데이터에 대한 성능을 보여준다. 각 Fold에서의 테스트 데이터의 개수는 17개 이상이고 AUC값은 모두 0.78이상 이었다. 모든 결과를 합한 AUC값은 0.85였다. 대한상한금궤의학회의 경우는 데이터의 수가 적기 때문에 다른 학과들에 비해 고르지 않은 Fold별 AUC값을 보였다. 이후에는 학과별 증상별 추천 성능 및 성능과 관련하여 다빈도로 등장하는 형태소의 분석 결과를 설명한다.

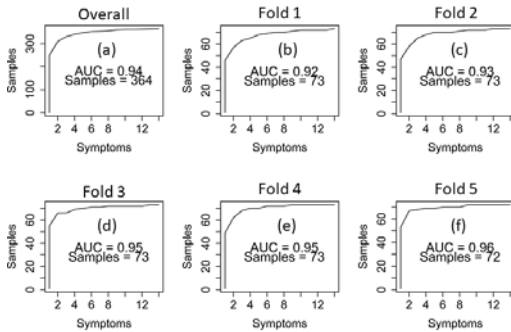


그림 4. 전반적인 성능 (대한형상의학회)
Figure 4. Overall performance

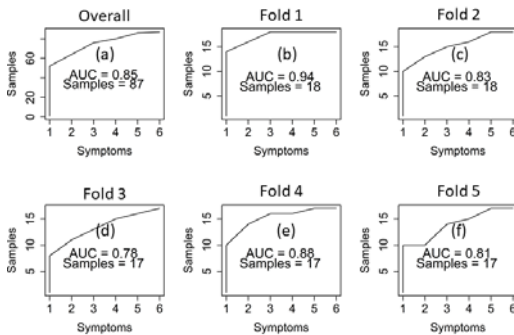


그림 5. 전반적인 성능 (대한상한금괴의학회)
Figure 5. Overall performance

3.3 증상별 추천 성능(사상체질의학회)

〈표 2〉은 사상체질의학회의 전문가가 생성한 데이터를 모델에 적용했을 때의 증상별 추천 성능을 AUC값 순으로 보여준다. 행은 각 증상에 등장했던 다빈도 형태소 10개를 설명한다. 괄호 안의 숫자는 데이터에서 형태소의 등장 빈도수이다. ‘외열다한’의 경우는 가장 높은 AUC값을 가지며, 모든 데이터에 ‘다한’ 및 ‘더위’라는 형태소가 포함되어 있었다. ‘외한’은 두 번째로 높은 AUC값을 보였다. 59개의 데이터 중 추위를 직접적으로 표현한 경우는 54개였다. ‘외열구갈’의 경우는 ‘더

위’와 ‘구갈’ 형태소가 높은 비중을 차지하였다. ‘상열’의 경우는 ‘오르다’, ‘올라오다’, ‘위로’, ‘(열이)나다’와 같은 더위와 관련된 형태소들이 상위에 있다. ‘한출다’의 경우는 땀이 많이 나는 것과 관련한 증상명인데 모두 ‘다한’이라는 형태소를 가지고 있었다. ‘비만’은 속이 더부룩한 상태와 관련이 있는데 ‘먹다’, ‘소화’, ‘가스’와 관련한 형태소들이 자주 등장했다. ‘대변연’은 변이 무른 양상을 나타낸다. ‘대변연’과 ‘설사’의 형태소 등장 패턴은 유사하나 ‘무르(다)’는 형태에 대한 형태소가 더욱 많이 등장하여 둘 사이의 구분에 기여한 것으로 보인다. ‘외열’증상에 대해서는 ‘더위’라는 형태소가 가장 빈번하게 등장했다. ‘수면불녕’은 양방의 불면증과 관련이 있는 증상명으로 ‘잠이)깨다’라는 형태소가 자주 등장하여 예측 성능에 기여를 한 것으로 보인다. 증상 ‘대변비’는 양방의 변비와 관련이 있다. 직접적으로 ‘변비’ 형태소가 가장 많이 등장했고 ‘(변을)보다’, ‘(변이)나오다’와 같은 배변을 묘사하는 형태소들이 많이 등장했다. 증상 ‘과식’의 경우는 ‘(과식을)하다’와 같은 형태소와 먹는 행위에 대한 ‘먹다’와 같은 형태소들이 다빈도로 등장했다. 또는 직접적으로 ‘과식’을 기술한 형태소가 나오기도 했다. 과식을 하는 상황 예를 들어 ‘스트레스’와 같은 것을 기술하는 경우도 5건이 있었다. ‘구갈’은 양방의 갈증과 관련성이 있는 증상으로 16개의 샘플에서 ‘구갈’을 직접적으로 사용했고 ‘갈증’을 기술한 샘플도 5개였다. ‘소화불량’과 관련된 형태소로는 17개의 데이터가 ‘소화’를 직접적으로 기술하였고 ‘먹다’, ‘불량’, ‘스트레스’와 같이 상황을 기술하는 형태소들이 등장했다. ‘탄산’은 신물이 올라오는 증상으로 ‘먹다’와 같은 형태소, ‘식도염’, ‘신물’, ‘역류성’, ‘올라오다’와 같은 형태소들이 다빈도로 등장한다. ‘두통’의 경우는

표 2. AUC값 및 자주 등장하는 형태소 상위 10개(사상체질의학회)
Table 2. Performance evaluation (AUC) by symptom and top 10 frequent terms

증상	AUC	데이터 개수	형태소 1	형태소 2	형태소 3	형태소 4	형태소 5	형태소 6	형태소 7	형태소 8	형태소 9	형태소 10
외열다한	1.000	14	다한(14)	다위(14)	추다(14)	민감(12)	수족(12)	하다(9)	냉음(7)	선호(6)	얼굴(5)	피로(5)
외한	0.996	59	추위(54)	수족(48)	민감(40)	선호(29)	찬물(18)	하다(14)	구갈(8)	다위(8)	먹다(8)	외한(6)
외열구갈	0.987	27	더위(26)	구갈(24)	수족(23)	민감(17)	다음(16)	선호(10)	냉음(9)	찬물(9)	먹다(7)	음수(5)
상열	0.983	15	오르다(7)	가끔(5)	위로(5)	나다(2)	되다(2)	오후(2)	올라오다(2)	월인(2)	이유(2)	내려가다(1)
한출다	0.983	18	다한(18)	추다(18)	하다(14)	피로(9)	나다(5)	상쾌(5)	커드랑이(3)	얼굴(3)	구갈(2)	다음(2)
비만	0.977	37	하다(22)	먹다(21)	소화(14)	속이다(12)	되다(9)	자다(9)	받다(8)	가끔(7)	가스(7)	느낌(7)
대변연	0.94	30	설사(16)	하다(13)	무르(10)	자주(10)	평소(10)	가다(9)	보다(9)	약간(8)	화장실(8)	가스(7)
설사	0.934	16	설사(16)	하다(15)	먹다(9)	자다(6)	자주(6)	나다(5)	가다(4)	문제(4)	우유(4)	기르다(3)
외열	0.931	29	더위(27)	수족(19)	민감(18)	찬물(11)	선호(8)	하다(8)	냉음(7)	먹다(6)	추위(6)	타다(6)
수면불녕	0.929	14	깨다(8)	하다(5)	편이(4)	가끔(3)	때문(3)	못자다(3)	않다(3)	야간(3)	오다(3)	항상(3)
대변비	0.928	25	변비(15)	먹다(12)	보다(8)	약간(8)	하다(8)	않다(7)	나오다(6)	생기다(6)	평소(6)	한번(6)
과식	0.911	22	하다(17)	먹다(15)	과식(8)	중기(5)	받다(5)	스트레스(5)	야식(5)	편이(5)	요새(4)	원래(4)
구갈	0.908	20	구갈(16)	다음(9)	갈증(5)	음수(5)	먹다(4)	자다(4)	mild(2)	냉음(2)	노력(2)	느끼다(2)
소화불량	0.905	22	소화(17)	먹다(15)	하다(12)	되다(9)	않다(9)	자다(9)	자주(9)	받다(6)	불량(6)	스트레스(5)
탄산	0.889	23	먹다(13)	하다(10)	받다(9)	식도염(8)	신물(8)	역류성(8)	올라오다(8)	자주(8)	증상(8)	가끔(6)
두통	0.87	10	헐다(6)	nrs(5)	두부(5)	머리(5)	통증(5)	두통(3)	나다(2)	느낌(2)	들다(2)	먹다(2)
심계	0.85	10	가슴(4)	하다(4)	가끔(3)	느낌(2)	되다(2)	받다(2)	스트레스(2)	헐다(2)	검사(1)	그냥(1)
대변난	0.745	10	변비(5)	보다(3)	않다(3)	음식(3)	한번(3)	거나(2)	걸리다(2)	느낌(2)	되다(2)	먹다(2)
대변건	0.727	11	약간(5)	가끔(4)	나오다(4)	보다(4)	자다(4)	하다(4)	평소(3)	다음(2)	대변(2)	되다(2)
소식	0.718	14	먹다(11)	하다(8)	자다(7)	간식(6)	않다(6)	과식(5)	소화(4)	속이다(4)	식탐(4)	공기(3)

‘간헐적’이라는 원 표현이 형태소 분석 후 변형되어 나온 ‘헐다’라는 형태소와 NRS (Numeric rating scale) 통증 척도와 관련된 형태소인 ‘nrs’, ‘통증’, ‘두통’과 같은 형태소들, 머리를 지칭하는 ‘두부’, ‘머리’와 같은 형태소들이 자주 등장했다. ‘심계’는 가슴이 두근거리면서 불안해하는 증상을 이르는데, ‘가슴’, ‘느낌’, ‘스트레스’ 등의 형태소들이 자주 등장했다. ‘대변난’은 변을 보기 어려운 상황과 관련이 있고 ‘보다’, ‘않다’, ‘느낌’과 같은 형태소들이 자주 등장을 한다. ‘변비’가 직접적으로 언급이 되기도 한다. ‘대변건’은 단단한 변의 상태를 나타내는 증상으로 다른 증상에 비해서 상대적으로 낮은 AUC를 보인다. 실제 자연어 기술에는 단단함을 기술한 표현이 등장하는데 형태소 분석과정에서 누락되었고 이것이 성능에 영향을 주었다. ‘소식’은 음식 섭취가 적은 것과 관련이 있는 증상으로 다른 증상들에 비해 가장 낮은 성능을 보인다. 주로 섭취의 양태와 관련된 용어들이 많이 등장한다. ‘과식’

하지 ‘않다’와 같은 부정 표현으로 기술이 되는데 TF행렬로 데이터를 변환하면서 부정 의미가 특징으로 표현이 되지 못했기 때문에 낮은 AUC를 보인 것으로 보인다.

표 3. ‘소식’의 예측 추천 결과

Table 3. Prediction results of symptom whose AUC is the lowest

Cutoff	추천1	추천2	추천3	추천4	추천5
1	비만(10)	과식(2)	상열(1)	소화불량(1)	
2	비만(12)	소화불량(6)	과식(4)	대변비(2)	상열(1)
3	비만(12)	소화불량(8)	탄산(6)	과식(4)	대변비(3)
4	비만(12)	탄산(9)	소화불량(8)	과식(7)	대변비(5)
5	비만(12)	과식(11)	소화불량(10)	탄산(10)	소식(7)
6	과식(13)	비만(12)	소식(11)	소화불량(11)	탄산(11)
7	과식(14)	비만(13)	소식(11)	소화불량(11)	탄산(11)
8	과식(14)	비만(14)	소식(11)	소화불량(11)	탄산(11)
9	과식(14)	비만(14)	대변비(12)	소화불량(12)	소식(11)

<표 3>는 AUC 기준 가장 하위의 증상인 ‘소식’에 대해 cutoff를 달리 했을 경우의 결과를 보여준다. 각 행은 모델의 증상 예측결과를 상위 부터 cutoff개까지 참고하는 것을 고려하는 것이고, 각 열은 그때 모델이 추천한 증상의 개수를 가장

많이 추천한 순으로 보여준다. ‘소식’의 경우는 상위 5개 까지 보았을 때 14개 중 7개(50%)가 포함되었고 상위 6개 까지 보았을 때 11개(78.5%)를 포함하였다. 그러나 ‘소식’과 관련한 자연어 기술에 대해서 모델이 예측한 증상들은 ‘비만’, ‘과식’, ‘소화불량’, ‘탄산’ 같은 소화와 관련된 증상들이 주로 예측되어 유사한 증상의 군을 형성했다. 이는 대변의 양상을 나타내거나 열감을 나타내는 다른 증상 군들과는 구별되었다.

3.4 증상별 추천 성능(대한형상의학회)

〈표 4〉는 대한형상의학회의 전문가가 생성한 데이터를 모델에 적용했을 때의 증상별 추천 성능을 보여준다. 형상의학은 얼굴을 포함한 신체의 외형을 진단 및 치료에 고려한다. 이 학파의 전문가들이 선정한 일부 증상들은 환자의 사진을 보고 판단되었다. 우리는 사진에 근거하여 선정된 데이터들을 제외하고 자연어 기술에 근거한 데이터들로 한정하여 모델에 적용을 시켜보았다. 총 164개의 증상들 중 14개의 증상들이 자연어 기술에 근거한

10개 이상의 데이터를 가지고 있었다. ‘복부압통’은 복부에 느껴지는 통증과 관련이 있는 증상명으로 ‘중완부’, ‘하부’, ‘하복부’와 같은 통증의 위치 및 압통을 직접적으로 표현하는 형태소들이 자주 등장했다. ‘흉비’는 가슴의 답답함과 관련된 증상으로 ‘가슴’, ‘스트레스’, ‘긴장’과 같은 형태소가 주를 이루었다. 증상 ‘소화불량’은 가장 많은 데이터를 가진다. ‘먹다’, ‘소화’와 같은 직접적인 형태소들이 자주 등장을 하고 ‘(소화가 안되는) 느낌’과 같은 형태소들의 비중도 컸다. 증상 ‘대변’은 대변과 관련된 설사, 변비와 같은 다양한 양상을 포괄한다. 증상명 ‘설사’는 설사가 나오는 상황에 대한 묘사가 많았다. 형태소에서도 그러한 상황과 관련된 것들이 다수였다. ‘상열’은 열이 나는 상황에 대한 묘사나 열이 나는 부위에 대한 묘사가 많았다. 형태소에도 ‘머리’, ‘얼굴’과 같은 것들이 등장했다. ‘탄산’은 식도염이나 신물, 쓴물 올라오는 상태에 대한 묘사가 많았다. ‘복만’은 ‘가스(가차다)’, ‘또는 ‘더부(룩하다)’와 증상을 나타내는 묘사가 많았다. 증상 ‘변비’는 자연어로 ‘변비’를 직접 기술하거나 변을 보는 주기를 기술한다.

표 4. AUC값 및 자주 등장하는 형태소 상위 10개(대한형상의학회)
Table 4 Performance evaluation (AUC) by symptom and top 10 frequent terms

증상	AUC	데이터 개수	형태소 1	형태소 2	형태소 3	형태소 4	형태소 5	형태소 6	형태소 7	형태소 8	형태소 9	형태소 10
복부압통	1.0000	55	중완부(45)	압통(39)	하부(34)	긴장감(26)	하복부(14)	긴장(8)	복부(2)	통증(1)		
흉비	0.9890	13	가슴(13)	가끔(3)	하다(3)	나타나다(2)	스트레스(2)	자다(2)	긴장(1)	느낌(1)	되다(1)	들다(1)
소화불량	0.9795	112	먹다(51)	하다(47)	소화(41)	자다(35)	자주(26)	속이다(22)	느낌(20)	가끔(14)	되다(14)	받다(14)
대변	0.9562	49	설사(35)	자다(33)	생기다(30)	변비(13)	대변(11)	보다(10)	않다(9)	먹다(8)	무른변(8)	하다(8)
설사	0.936	19	설사(17)	하다(13)	자주(10)	먹다(7)	가다(5)	화장실(5)	음식(4)	가스(3)	기르다(3)	긴장(3)
상열	0.9285	10	느낌(3)	머리(3)	얼굴(3)	하다(3)	나서다(2)	들다(2)	나다(1)	등반(1)	되다(1)	두통(1)
탄산	0.9190	15	신물(6)	올라오다(6)	식도염(5)	역류성(5)	가끔(3)	느낌(3)	쓸다(3)	과식(2)	먹다(2)	속이다(2)
복만	0.8857	10	가스(5)	더부(4)	차고(4)	가끔(2)	먹다(2)	가다(1)	나다(1)	느낌(1)	밀가루(1)	변비(1)
변비	0.8809	18	변비(13)	먹다(5)	3일(4)	대변(4)	생기다(4)	자다(4)	보지(3)	되다(2)	배변(2)	하다(2)
음주	0.8785	10	소주(8)	1회(4)	2-3회(3)	맥주(3)	먹다(2)	설사(2)	하다(2)	2-3(1)	과식(1)	다음(1)
질정	0.8781	17	받다(11)	스트레스(11)	하다(8)	가다(4)	신경(4)	되다(3)	자다(3)	가끔(2)	긴장(2)	막히다(2)
복진	0.8681	13	긴장감(9)	중완부(9)	하부(5)	압통(2)	긴장(1)	복부(1)				
두통	0.8626	13	통증(8)	nrs(7)	혈다(6)	두부(5)	두통(4)	발생(4)	소화(4)	신경(4)	우리(4)	하다(4)
조잡	0.7285	10	먹다(4)	올라오다(3)	가끔(2)	공복(2)	느낌(2)	새벽(2)	속이다(2)	식도염(2)	역류성(2)	예전(2)

‘음주’는 ‘맥주’나 ‘소주’와 같은 주종이나 주량에 대한 표현이 많았다. ‘질정’은 심적인 어려움이나 스트레스와 관련한 형태소가 자주 등장했다. ‘복진’은 ‘복부압통’과 유사한 자연어 기술을 보였고 주로 특정 부위의 긴장감에 대한 기술이 주를 이루었다. 증상 ‘두통’은 두통의 양상, 위치, 조건에 대한 내용 및 두통의 정도를 표현한 ‘NRS’를 기술하는 경우가 많았다. 증상 ‘조잡’은 기술된 내용에서 증상 ‘탄산’과 유사했으나 속이 쓰리거나 미식거리는 증상표현이 두드러졌다.

표 5. ‘조잡’의 예측 추천 결과

Table 5. Prediction results of symptom whose AUC is the lowest

Cutoff	추천1	추천2	추천3	추천4	추천5
1	소화불량(8)	질정(1)	탄산(1)		
2	소화불량(10)	탄산(6)	대변(1)	변비(1)	설사(1)
3	소화불량(10)	탄산(8)	조잡(3)	대변(2)	상열(2)
4	소화불량(10)	탄산(8)	상열(5)	조잡(4)	질정(4)
5	소화불량(10)	탄산(8)	상열(7)	조잡(7)	질정(5)
6	소화불량(10)	조잡(9)	탄산(9)	상열(8)	질정(6)
7	소화불량(10)	상열(9)	조잡(9)	질정(9)	탄산(9)
8	변비(10)	소화불량(10)	조잡(10)	상열(9)	질정(9)
9	변비(10)	상열(10)	소화불량(10)	조잡(10)	복만(9)
10	변비(10)	복만(10)	상열(10)	소화불량(10)	조잡(10)

〈표 5〉는 가장 낮은 예측 성능을 보인 ‘조잡’에 대한 모델의 예측 결과를 상위부터 Cutoff까지 보았을 때 빈번하게 등장한 증상 순으로 보여준다. ‘조잡’의 경우는 상위의 6개까지 고려할 때 10개 중 9개(90%)의 조잡 관련 데이터가 포함되었다. 모

델은 ‘조잡’에 대한 입력에 대해 ‘소화불량’과 ‘탄산’을 상위로 추천했다. 이는 ‘소화불량’이나 ‘탄산’의 특징들과 다른 증상들보다 ‘조잡’과 가까웠기 때문으로 보인다. 즉 모델은 유사한 증상 군을 상위로 추천한다.

3.5 증상별 추천 성능(대한상한금궤의학회)

〈표 6〉는 대한상한금궤의학회의 전문가가 생성한 데이터를 모델에 적용했을 때의 증상별 추천 성능을 보여준다. 이 학과는 상한론(傷寒論)에 바탕을 둔 독자적인 치료 체계를 가지고 환자의 진단 및 치료에 적용한다는 특징이 있다. 데이터가 개별 증상들에 고르게 분포하고 있어 AI모델에 적용하기에 충분한 데이터가 없어 6개의 증상들에 대해서 모델의 성능이 측정되었다. 가장 높은 성능을 보인 ‘구’는 소화가 안 되는 상황과 관련한 증상명이다. ‘과식’, ‘(소화)문제’ 등을 직접 언급하는 형태소의 빈도수가 높았다. ‘하리’는 설사를 이르는 증상 용어로 실제 자연어 입력에서도 ‘설사’를 직접 언급하는 경우가 많았다. ‘건구’는 속쓰림과 관련이 큰 증상으로 속쓰림이 발생하는 상황인 ‘공복’, ‘빈속’ 등이 직접 언급되거나 원 표현인 ‘속쓰림’이 자연어의 형태소 분석 과정에서 변형된 형태소인 ‘속이다’가 빈번

표 6. AUC값 및 자주 등장하는 형태소 상위 10개(대한상한금궤의학회)

Table 6. Performance evaluation (AUC) by symptom and top 10 frequent terms

증상	AUC	데이터 개수	형태소 1	형태소 2	형태소 3	형태소 4	형태소 5	형태소 6	형태소 7	형태소 8	형태소 9	형태소 10
구	0.9444	18	과식(4)	문제(4)	소화(4)	먹다(3)	불량(3)	하다(3)	되다(2)	발생(2)	속이다(2)	음식(2)
하리	0.9347	23	설사(22)	문제(2)	되다(1)	발생(1)	자다(1)	자주(1)	하다(1)			
건구	0.8555	15	공복(4)	속이다(4)	먹다(3)	빈속(3)	당뇨(2)	과식(1)	문제(1)	발생(1)	소화(1)	음식(1)
맥세	0.8500	10	어리다(7)	발생(2)	증상(2)	되다(1)	먹다(1)	부터(1)	소화기(1)	자다(1)	자주(1)	하다(1)
왕래한열	0.7333	10	환절기(4)	날긋(3)	문제(3)	하다(3)	소화기(2)	증상(2)	걸리다(1)	당뇨(1)	되다(1)	비열(1)
토	0.6363	11	나다(3)	먹다(2)	음식(2)	증상(2)	되다(1)	불량(1)	소화(1)	속이다(1)	자다(1)	하다(1)

하게 등장하였다. ‘맥세’는 과민한 장의 상태와 관련해서 빈번하게 사용되는 증상명으로 어려서부터 문제 상황이 발생 했다는 자연어 기술이 형태소 분석을 통해서 추출된 ‘어리다’라는 형태소가 빈번하게 등장하였다. ‘왕래한열’의 경우는 추위와 열감을 번갈아 느끼는 증상으로 해당 증상이 생기는 조건인 ‘환절기’, ‘날긋(이)’와 같은 형태소들이 많이 등장했다. ‘토’는 속쓰림, 식도염, 구취와 관련한 자연어 기술들이 주를 이루었으나 형태소 분석 과정에서 형태소로 추출되지 않았기에 가장 낮은 AUC값을 보였다.

표 7. ‘조잡’의 예측 추천 결과
Table 7. Prediction results of symptom whose AUC is the lowest

Cutoff	추천1	추천2	추천3	추천4	추천5
1	구(7)	건구(2)	왕래한열(2)		
2	구(11)	건구(9)	왕래한열(2)		
3	구(11)	토(10)	건구(9)	왕래한열(2)	맥세(1)
4	건구(11)	구(11)	토(10)	왕래한열(7)	맥세(5)
5	건구(11)	구(11)	맥세(11)	토(11)	왕래한열(9)
6	건구(11)	구(11)	맥세(11)	왕래한열(11)	토(11)

<표 7>은 가장 낮은 AUC를 보인 증상인 ‘토’의 자연어 입력에 대해 모델이 예측한 증상들을 Cutoff별로 나타낸 결과이다. 상위 3개 까지 보았을 때 11개 중 10개를 맞출 수 있었다. ‘토’의 자연어 입력에 대해 ‘구’, ‘건구’ 같은 증상들이 상위로 추천되었다. ‘토’의 자연어 입력에 대해 ‘토’가 상위에 오지는 못했지만, ‘토’와 유사한 증상군인 ‘구’, ‘건구’를 추천함을 확인 했다.

4. 논의 및 결론

우리는 한의 증상 관련 자연어 기술을 입력으로 받아 증상명을 추천하는 시스템을 개발했다. 본 시스템은 자연어 입력으로부터 적절한 한의 증상명을 탐색하기 위한 용도로도 사용될 수 있을 뿐 아

니라, 한의 병증 및 치료 정보를 제공해주는 시스템([8-13])과 연계하여 그러한 시스템의 활용에 있어서도 유용성을 높여 줄 것으로 기대된다.

본 연구에서 수집된 소화기계 관련 증상을 가진 환자들 데이터를 분석한 결과, 다양한 한의학과의 한의사들이 사용하는 증상 용어에 차이가 있음을 확인 했다. 동일한 환자에 대해서 학과별로 유사하지만 동일하지 않은 용어로 기술한 경우가 있었다. 대한형상의학회에서 ‘칠정’으로 기술한 환자에 대해서 사상의학회의 경우는 ‘불안’으로 대한상한금궤학회의 경우는 ‘양명병’, ‘심번’으로 기술된 사례가 있었다. ‘설사’에 대해서는 사상의학회의 경우는 ‘설사’ 또는 ‘대변연’, 상한금궤학회의 경우는 ‘하리’로 표현하는 경우가 있었다.

다른 학과에서 생성된 데이터와 비교해서, 대한상한금궤학회 전문가가 생성한 데이터들은 가장 적은 수의 데이터만 모델 생성 및 성능평가를 위해서 이용되었다. 증상들 별로 할당된 데이터 수가 10개 이하인 경우가 다수였고, 이는 학습 및 평가를 하기 에는 부족한 양이기 때문이다. 이 학과의 전문가들이 생성한 데이터의 자연어 기술 부분은 큰 차이가 없으나 서로 다른 증상명으로 기술한 경우가 있었다. 상한금궤학과의 전문가들이 사용한 증상명에는 자연어 기술에 직접 드러나지 않은 고려사항들이 함의 되어 있기 때문으로 보인다.

모든 학과의 데이터셋에 대해, 학습된 모델의 전반적인 AUC값은 0.85이상이었다. 증상별로 성능을 평가 했을 때는, 가장 낮은 성능을 보인 증상이라 할지라도 모든 학과에서 0.6 이상의 AUC값을 보였다. 그러나 낮은 예측 성능을 보인 증상들이라도, 모델은 최소한 해당 증상과 유사한 증상들을 추천할 수 있음을 확인했다.

일부 증상들에 대해 모델의 예측 성능이 낮은 것에는 다양한 원인이 있을 수 있다. 같은 학과 내의 두 증상 사이에도 명확하게 구별되지 않는 경

우가 있었다. 예를 들어 ‘대변난’ 과 ‘대변비’ 와 같은 경우는 입력으로 사용된 자연어 서술로는 구별이 어려운 경우가 있었다. 다른 원인으로, 형태소 분석 기법의 한계 때문일 수 있다. 예를 들어 증상의 구별에 주요한 어떤 형태소가 형태소 분석 과정에서 추출이 되지 않았기 때문일 수 있다. 또는 우리의 분류 모델이 TF를 특징으로 사용했기에 생긴 한계 때문일 수 있다. 예를 들어 환자의 상태에 대한 자연어 기술이 TF 특징에서 잘 드러나지 않는 부정 표현으로 기술되어 있는 경우는 낮은 분류 성능을 보일 수 있다.

우리는 한글 형태소 분석을 위해 KoNLPy의 twitter를 사용했다. 명사, 동사, 영단어, 숫자를 형태소로 사용했다. <표2>, <표4>, <표6>에서 볼 수 있는 것처럼 ‘하다’, ‘되다’ 와 같은 표현이 다수 등장했다. 형태소 분석기에서 위의 단어들을 Stop word로 제외한 후 성능을 측정한 결과, 두 학파의 데이터에 대한 AUC값은 큰 변화가 없었지만 대한상한금궤의학회 데이터에 대한 AUC는 0.88로 상승하였다. 이는 적절한 Stop word를 선정하여 추가적인 성능의 향상을 기대할 수 있음을 시사한다. 또한 형태소 분석의 결과로 ‘더부룩하다’ 가 ‘더부’ 로, ‘간헐적’ 이 ‘헐다’ 로, ‘어릴때’ 가 ‘어리다’ 로 변환되었다. 비록 형태소 분석 후 원형을 알기 어렵게 되었지만, 이러한 형태소를 제외 시켰을 경우 예측 성능이 감소됨을 확인했다. 형태소분석기 별로 특징과 장단점이 다르므로 임상 텍스트 분석에 적합한 형태소 분석기를 비교 분석 후 선택하여 이런 한계를 극복하고 성능의 향상을 기대할 수 있다.

우리는 증상 추천을 위해서 TF를 특징으로 사용한 SVM을 사용했다. 최근에는 텍스트 분류 문제를 위해 워드 임베딩에 기반한 CNN (Convolutional neural network)등도 적용되고 있다. CNN은 이미지 분류에 좋은 성능을 보이는 것으로 잘 알려져 있

으나([14]) 워드 임베딩과 결합하여 텍스트 분류 문제에도 좋은 성능을 보인다[15]. 그러나 임베딩 벡터의 구성 및 데이터 개수의 제한 등 이 최신의 기법을 도입하기 위해 앞서 고려할 사항들이 있다. 향후에 추가적인 데이터를 생성하여 데이터의 양을 늘리거나 또는 적은 데이터 개수 문제를 해결할 수 있는 기술적인 방법을 적용하여 더욱 향상된 추천 성능을 기대할 수 있다.

우리의 시스템은 방대한 한의 증상명에 대한 탐색을 용이하게 함으로써 한의 임상의 및 일반 사용자들의 한의 의료에 대한 접근성 및 활용성을 향상시켜 한의학의 대중화에 기여할 수 있다.

References

- [1] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, *An overview of clinical decision support systems: benefits, risks, and strategies for success*, NPJ Digital Medicine. Vol. 3, No. 1, pp. 1-10, 2020.
- [2] S-K. Kim, H-C. Jang, J-H. Kim, Y-T. Oh, C. Kim, S-J. Yea, and M-Y. Song, *Traditional Korean medicine diagnosis system based on basic ontology*. Journal of Physiology & Pathology in Korean Medicine. Vol. 24, No. 6, pp. 1111-1116. 2010.
- [3] W-M. Jung, I-S. Park, Y-S. Lee, C-E. Kim, H. Lee, D-H. Hahm, H-J. Park, B-H. Jang, and Y. Chae. *Characterization of hidden rules linking symptoms and selection of acupoint using an artificial neural network model*, Frontiers of Medicine. Vol. 13, No. 1, pp 112-120. 2019.
- [4] X. Mi, H. Ikeda, F. Nakazawa, H. Matsuoka, E. Kataoka, S. Hamaya, and T.

- Kawanabe, *Prescription prediction towards computer-assisted diagnosis for kampo medicine*. International Conference on Computer Application Technologies IEEE. pp. 126-131, 2015.
- [5] T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*. European conference on machine learning. Springer, pp. 137-142. 1998.
- [6] J. Woo, *The spam detection model for web forums using text mining techniques*. Journal of Knowledge Information Technology and Systems, Vol. 7, pp 159-166, 2012.
- [7] G. Mujtaba, L. Shuib, N. Idris, W. L. Hoo, R. G. Raj, K. Khowaja, K. Shaikh, and H. F. Nweke, *Clinical text classification research trends: Systematic literature review and open issues*. Expert systems with applications, Vol. 116, pp. 494-520. 2019.
- [8] Z. Liu, Z. Zheng, X. Guo, L. Qi, J. Gui, D. Fu, Q. Yao, and L. Jin, *AttentiveHerb: A novel method for traditional medicine prescription generation*. IEEE Access, Vol. 7, pp. 139069-139085. 2019.
- [9] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, *A topic modeling approach for traditional chinese medicine prescriptions*. IEEE Transactions on Knowledge and Data Engineering, Vol. 30, No. 6, pp. 1007-1021. 2018.
- [10] K. Yang, R. Zhang, L. He, Y. Li, W. Liu, C. Yu, and X. Zhou, *Multistage analysis method for detection of effective herb prescription from clinical data*. Frontiers of medicine, Vol. 12. No. 2, pp 206-217, 2018.
- [11] S. Wang, E. W. Huang, R. Zhang, X. Zhang, B. Liu, X. Zhou, and C. Zhai, *A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional Chinese medicine patient records*, IEEE International Conference on Bioinformatics and Biomedicine (BIBM) pp. 411-418, 2016.
- [12] Y. B. Li, X. Z. Zhou, R. S. Zhang, Y. H. Wang, Y. Peng, J. Q. Hu, and B. Y. Liu, *Detection of herb-symptom associations from traditional chinese medicine clinical data*. Evidence-Based Complementary and Alternative Medicine, 2015.
- [13] F. Hu, Y. Qiao, G. Xie, Y. Zhu, Y. Jia, and P. Huang, *Symptom distribution regulation of core symptoms in insomnia based on informap-sa algorithm*. International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES). IEEE, pp 229-232, 2017.
- [14] O-S. Kwon, *Design of convolution neural network based on solar energy and enhancement of number recognition*. Journal of Knowledge Information Technology and Systems, Vol. 14, No. 1, pp 93-101, 2019.
- [15] Y. Kim, *Convolutional neural networks for sentence classification*, arXiv preprint arXiv:1408.5882, 2014.

자연어 기술 기반 한의 증상명 추천 시스템 (소화기계 질환 버전)

장호¹, 이상훈², 예상준³

¹한국한의학연구원 지능화추진팀 선임연구원

²한국한의학연구원 미래의학부 책임연구원

³한국한의학연구원 지능화추진팀 책임연구원

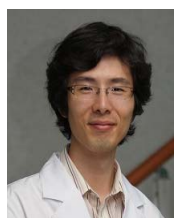
요 약

최근 다양한 정보통신 및 인공지능 기술들이 의료 서비스의 수준을 향상시키기 위해 연구되고 있다. 한의학 분야에서도 관련 연구자들이 다양한 연구를 수행하고 있다. 임상 한의사는 망문문절을 통해서 얻어진 환자의 정보와 다년간의 진료 경험을 통합해서 환자 변증을 특정하고 환자 상태에 맞는 처방을 결정한다. 이를 모사하여 알고리즘에 기반한 병증의 탐색 및 치료 추천을 위한 다양한 모델이 제안되어왔다. 증상특정-처방추천으로 이어지는 모델은 가장 널리 연구되어오던 접근 방법 중 하나이다. 그러나 한의 용어는 매우 광범위하고 한의 학과마다 사용하는 용어에 차이가 있기에 개발된 시스템의 원활한 이용에 한계가 있다. 따라서 우리는 환자의 상태와 관련한 자연어 기술로 부터 관련성이 높은 한의 증상명을 추천해주는 시스템을 개발했다. 이를 위해서 소화기계 질환을 가진 환자를 모집하고 환자의 상태를 자연어로 기술한 데이터를 만들었다. 한의 학과별 한의사들은 환자의 질병과 연관성이 큰 증상들을 선별하고 그 증상의 근거가 되는 자연어 기술 부분을 태깅해주었다. 이 데이터는 한글 형태소 분석 및 Multi-level support vector machine (SVM) 모델에 기반한 시스템의 입력 데이터로 활용되었다. 모델 성능 평가를 위해 5-fold cross validation 사용되었고 한의사들이 설정한 증상명이 학습 모델에서 어느정도 우선순위로 예측되었는지를 바탕으로 Area under curve (AUC) 값으로 계산했다. 전반적인 성능은, 사상체질의학회의 경우 93%의 AUC값을 보였고 대한형상의학회의 경우 94%의 AUC값을 보였다. 대한상한금궤의학회의 경우 85%의 AUC값을 보였다. 우리의 자연어 기반 증상명 추천 시스템은 한의 증상명 탐색의 편의성을 증진시켜 한의사들의 시스템 접근성을 도울 수 있다. SVM 모델을 바탕으로 만들어진 우리의 시스템은 이후 딥러닝에 기반한 고수준의 모델로 확장하여 성능의 향상을 도모하거나 다양한 질환들로 확장될 수 있다.



Ho Jang received his Ph.D. degree from School of Electrical Engineering and Computer Science at Gwangju Institute of Science and Technology in 2017. He has been working for Korea Institute of Oriental Medicine since 2018. His research interests include computational biology and machine learning.

E-mail address: jh@kiom.re.kr



Sanghun Lee He received his M.S. degree and his Ph.D. degree from Department of Korean Medicine at Wonkwang University in 2007 and 2011, respectively. After graduation, he has been researched on Modernization and Standardization of Korean medicine devices and informatics in Korea Institute of Oriental Medicine.

E-mail address: ezhani@kiom.re.kr



Sangjun Yea received the Ph. D in knowledge service engineering from KAIST, Korea in 2018. Since 2008, he is a principal researcher at Korea Institute of Oriental Medicine. The current research interests is biomedical data science.

E-mail address: tomita@kiom.re.kr

감사의 글

본 연구는 한국한의학연구원 'AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축 (KSN2012110)' 과제의 지원을 받아 수행되었습니다.