

기계 학습을 이용한 한의학 처방 분석 방안

한국한의학연구원 책임연구원
오준호*

A Strategy for Disassembling the Traditional East Asian Medicine Herbal Formulas With Machine Learning

Oh Junho*

Senior Researcher. Korea Institute of Oriental Medicine.

Objectives : We propose a method to disassemble Traditional East Asian Medicine herbal formulas using machine learning.

Methods : After creating a model using Byte Pair Encoding(BPE) and G-Score, the model was trained with training data. Afterwards, the learned model was applied to the test data, of which the results were compared with expert opinion.

Results : The results acquired through the model were not significantly different from those of modern expert opinions. However, there were cases where the meaning was partially unclear, while there were cases where new knowledge could be obtained through the disassembling process.

Conclusions : It is expected that disassembling herbal formulas through the proposed method in this study will help save resources required to understand complex ones.

Key words : herbal formulas, machine learning, Byte Pair Encoding, G-score, Korean Medicine, Traditional East Asian Medicine(TEAM).

* Corresponding Author : Oh Junho.

Senior Researcher. Korea Institute of Oriental Medicine.

1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054

Tel +82-42-868-9317 E-mail: junho@kiom.re.kr

저자는 본 논문의 내용과 관련하여 그 어떠한 이해 상충도 없습니다.

Received(February 3, 2023), Revised(May 3, 2023), Accepted(May 3, 2023)

Copyright © The Society of Korean Medical Classics. All rights reserved.

© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서론

한의학의 치료 방법은 크게 침과 약으로 나눌 수 있다. 이 가운데 약을 이용한 치료 방법은 대부분 본초를 조합하여 만들어진 방제(이하 처방)이다. 한의학 고문헌에 적혀 있는 처방의 종류는 매우 많다. 그 전체 수량을 파악하기 어려우나 현대 중국에서 만든 《中醫方劑大辭典》에는 10만여 종의 처방이 실려 있다.¹⁾ 실제 임상에서는 기본 처방을 그대로 쓰기보다 환자의 증상에 따라 본초를 가감하거나 용량을 조절하여 사용하는 일이 많다. 이런 경우까지 고려한다면 한의학의 치료 처방은 그 수를 모두 헤아리기 어렵다.

처방의 약성과 효능을 이해하는 일은 한의사가 임상에서 환자를 치료하기 위해 매우 중요하다. 하지만 처방의 종류가 이처럼 많으므로 처방을 하나하나 모두 이해하기는 거의 불가능하다. 전통적으로 처방의 성격과 효능을 이해하기 위해 가장 많이 사용되는 방법은 대상 처방을 더 작은 단위로 분해하여 설명하는 방식이다.²⁾ 이는 마치 단어의 의미와 문장의 패턴을 알고 있으면 처음 보는 문장도 이해할 수 있는 것과 유사한 이치이다. 문장이 처방이라면 단어와 패턴은 본초와 본초 조합이라고 할 수 있다. 부분의 합을 통해 전체를 설명하기에는 분명 한계가 존재한다. 그러나 이러한 방식은 다양한 처방을 이해하기 위한 기초적인 접근법이다. 이러한 방법론은 오늘날에도 시도되고 있다.³⁾

4~5가지 본초로 이루어진 처방의 경우는 구성

본초 각각을 가지고 처방을 설명하더라도 이해가 어렵지 않다. 그러나 7~8가지가 넘는 약제로 구성된 처방의 경우에는 본초 하나하나로 처방을 설명하기 어렵다. 이때는 해당 처방이 어떻게 형성되었는지 거꾸로 추정하는 방법이 유용하다. 예를 들어 육군자탕(六君子湯)의 경우, 사군자탕(四君子湯)에 진피(陳皮)를 더한 처방이 이공산(異功散)이고, 이공산에 반하(半夏)를 더한 처방이 육군자탕이라고 설명하는 것이다. 이렇게 다른 처방으로 분해할 수 있다면 대상 처방을 보다 간단히 이해할 수 있다.⁴⁾

임의의 대상 처방을 이해하는 전통적이면서도 효과적인 방법은 위의 2가지 방법을 적절히 결합하여 사용하는 것이다. 먼저 알려진 기존 처방을 바탕으로 대상 처방을 더 작은 단위로 분해해 보고, 이렇게 설명되지 않는 나머지 부분을 약성과 효능을 바탕으로 설명하는 방법이다. 그러나 대상 처방의 본초 개수가 많다면 가능한 본초 조합이 많이 늘어나기 때문에 이와 같은 방식으로 처방을 분해하는데 상당한 시간과 노력이 소요된다. 만약 자동화된 방법으로 대상 처방을 더 작은 부분으로 나눌 수 있다면 처방 분석에 도움이 될 수 있을 것이다. 이에 본 연구에서는 기계 학습(machine learning)을 이용하여 처방을 분해하는 방법을 제안하고 몇 가지 처방에 적용한 결과를 소개하고자 한다.

기계 학습 방법의 우수성을 주장하기 위해서는 기계 학습을 적용할 대상 과제, 이를 검증할 테스트 데이터, 그리고 성능을 측정할 정량적인 평가 방법이 존재해야 한다. 하지만 본 연구에서 다루고 있는 처방 분해 문제의 경우 객관적으로 성능을 평가할만한 테스트 데이터와 정량적인 평가 방법이 존재하지 않는다. 따라서 본 연구의 목적은 기계 학습을 이용하는 방법을 제안하고 학계의 관심을 환기하는 것에 그칠 수밖에 없다는 점을 밝혀둔다.

1) 彭怀仁 等. 中醫方劑大辭典. 人民卫生出版社. 北京. 1993. 1판 기준 96,592종, 2판에는 2,400여 종이 추가되었다.
2) 처방 분석과 관련된 연구는 매우 많으며 컴퓨터를 이용한 처방 분석 연구도 다수 수행되었다. 한국에서 이루어진 초기 연구로는 다음과 같은 선행 연구가 있다. 이병욱, 정훈, 권영규. 인터넷상 동의보감 방제의 약물구성 정보 표현 방법에 관한 연구. 한국의사학회지. 2008. 21(2). pp.41-47.
김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. 대한한의학원전학회지. 2008. 21(1). pp.143-151.
김정훈, 이병욱. 본초 조합을 이용한 方劑의 類方分析 DB 구축 연구 -『東醫寶鑑』과『方劑學』을 중심으로 -. 대한한의학원전학회지. 2008. 21(1). pp.123-141.
3) 방민우, 김기욱, 이병욱. 본초 구성을 이용한 방제의 효능 추론 및 분류 방법에 관한 연구. 대한한의학방제학회. 2017. 25(1). pp.29-38.

4) 윤홍길 외 3인. 기존방제를 기반한 처방분석 방법에 관한 연구. 대한한의학원전학회지. 2021. 34(4). pp.65-78.

II. 본론

1. 연구 아이디어

한의학 처방이 최초 어떻게 탄생하였고 발전했는지 자세히 고증하기는 어렵다. 하지만 오랜 시간 이루어진 한의학 처방의 발전 과정에 대해서 우리는 2가지 사실을 미루어 짐작할 수 있다.

첫째, 한의학 발전 초기에는 본초와 처방의 구분이 명확하지 않았고 본초가 곧 치료 처방 그대로 사용되었을 것이다. 그러므로 초기의 처방은 1가지 본초로 이루어진 단방이었을 것이다. 둘째, 본초 구성이 복잡한 처방은 상대적으로 단순한 처방이 확장되어 만들어졌다. 초기의 처방이 단방이었다는 첫 번째 추측으로부터 오늘날의 복합방의 존재를 설명하려면 단순한 처방에서 복잡한 처방이 출현하였다고 보아야 하기 때문이다. 물론 팔미지황환(八味地黃丸)에서 육미지황환(六味地黃丸)이 탄생한 예와 같이 때때로 본초가 줄어드는 방향으로 이루어진 변화도 있었다. 또 더 복잡한 처방이 덜 복잡한 처방보다 우월하다고 할 수도 없다. 그렇지만 임의의 처방이 주어졌을 때 그 조상에 해당하는 처방은 그보다 구성 본초의 개수나 용량이 더 단순한 처방일 가능성이 크다.

이런 인식은 동아시아 전통사회에도 존재하였다. 《통감외기(通鑑外紀)》 등을 보면 삼황오제(三皇五帝) 시대에 신농씨(神農氏)가 약초를 맛보고 약물을 발견하였다고 하였다.⁵⁾ 또 《의평(醫評)》⁶⁾이나 《의학원류론(醫學源流論)》⁷⁾에는 고대에는 1가지 약재로 1가지 병을 치료하였는데 후에 몇 가지 약재를 합하여 처방이 만들어졌다고 하였다. 또 《동의보감(東醫寶鑑)》에서도 “상고시대에는 한 가지 약으로 한 가지 병을 치료했는데 한(漢)의 장중경에 이르러 여러 약으로 한 가지 병을 치료하게 되었다.”⁸⁾고 하

5) 윤용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p3. 재인용.

6) 의평(醫評). “醫道起於神農之著本草，以一藥治一病，但一病有數證。後之聖人，取藥之對證者合幾味而成方，故治病必先有藥而後有方，方成之後，再審其配合之法。”

7) 의학원류론(醫學源流論). “古方以一藥治一證，合數證而成病，即合數藥而成方.”

8) 동의보감(東醫寶鑑). 잡병편(雜病篇). 용약(用藥). “上古用一藥治一病，至漢張仲景，用群藥治一病.”

였다.

이러한 전제 아래 사고 실험을 진행해 보자. 어떤 시대에 몇 가지 본초로 이루어진 특효 처방이 있었고 당시에 많은 사람이 이 처방을 매우 애용하였다고 해보자. 이 처방을 ‘조상 처방’이라고 하자. 시간이 흘러 임상가들은 자신의 임상 경험에 따라 이 처방을 다양한 형태로 변주하였고 그중 몇몇은 또다시 새로운 처방으로 전승되었다. 이를 ‘후손 처방’이라고 하자. 다른 본초가 가감되기도 하고 별도의 처방이 합쳐지기도 하면서 조상 처방은 후손 처방으로 변성해 나갔다.

만약 시간이 흘러 조상 처방이 어떤 본초로 이루어졌는지 잊히고 말았다고 해보자. 다행히 우리는 여러 가지 후손 처방의 본초 조합을 알고 있다. 그렇다면 그로부터 이 조상 처방의 모습을 추측할 수 있다. 후손 처방에서 공통으로 빈도 높게 출현하는 본초 조합을 찾아내면 그것이 조상 처방의 본초 조합일 가능성이 크기 때문이다.

오늘날 존재하는 대부분의 처방은 여러 조상 처방의 후손 처방이라고 할 수 있다. 그렇다면 후손 처방의 본초 조합 속에서 공통으로 나타나는 본초 조합을 찾아낸다면 조상 처방의 본초 조합을 추측할 수 있다. 후손 처방을 조상 처방의 조합 혹은 조상 처방의 가감방으로 분해해 낼 수 있다면 처방을 이해하는 데 많은 도움이 될 것이다. 본 연구는 이 아이디어를 바탕으로 구체적인 처방 분해 방법을 구현해 보았다.

2. 처방 분해 방법

본 연구에서 제안하는 처방 분해 방법은 Byte Pair Encoding(BPE)과 G-Score를 이용한 방법이다. BPE는 정보 압축 알고리즘의 하나로, 자주 반복되는 문자열을 치환하여 데이터의 양을 줄여나가는 방법이다. 예를 들면 아래와 같다.

<표 1> 페어 인코딩(Byte Pair Encoding, BPE)
모식도

<p>Step 0 (초깃값) 대상 문자열 : “가가가나라가가나라가다” 용어집 : 없음</p>
<p>Step 1 대상 문자열 : “A가나라A가나라가다” 용어집 : A=가가</p>
<p>Step 2 대상 문자열 : “AB라AB가다” 용어집 : A=가가, B=가나</p>
<p>Step 3 대상 문자열 : “C라C가다” 용어집 : A=가가, B=가나, C=AB</p>

최초 데이터 “가가가나라가가나라가다”에서 자주 반복되는 문자열 “가가”를 ‘A’로 치환하여 “A가나라A가나라가다”로 업데이트하였다. 여기서 다시 “가나”를 ‘B’로 치환하여 “AB라AB가다”를 얻었다. 이런 방식으로 치환해 나가면 중복되는 문자열이 짧게 줄어들기 때문에 그만큼 데이터 압축 효과를 얻을 수 있다. 압축 결과와 용어집을 가지고 있으면 과정을 거꾸로 반복하여 초깃값을 얻을 수 있다.⁹⁾

한의학 처방에도 이를 적용할 수 있다. 먼저 여러 처방에 가장 반복적으로 등장하는 본초 조합 1쌍을 찾아낸다. 이를 새로운 문자로 치환하고 용어집을 업데이트한다. 이러한 과정을 계속 반복해 나가면 가장 빈도가 높은 본초 조합부터 낮은 본초 조합 순서로 본초를 묶어 나갈 수 있다. 이렇게 묶어 나가면 마치 양과 껍질을 벗기듯 처방을 단계적으로 분해하는 효과를 얻을 수 있다.

여기서 한 가지 고려해야 할 점이 있다. 단순히 ‘빈도가 높은 조합’이 유의미한 조합인가 하는 문제

이다. BPE 알고리즘은 데이터의 압축을 위해 고안된 방법이므로 가장 빈도가 높은 문자열, 즉 가장 반복적으로 등장하는 문자열을 우선 치환한다. 그래야 압축 성능이 좋아지므로 달리 고려할 사항이 없다. 하지만 한의학 처방의 경우는 문제가 다르다. 본초 각각의 사용 빈도에 편차가 심하기 때문이다. 감초(甘草)·대조(大棗) 등은 빈번히 사용되지만, 감수(甘遂)나 사간(射干) 등은 상대적으로 드물게 사용된다. 단순 빈도를 가지고 본초 조합을 묶어 나간다면 감초-생강, 혹은 감초-대조의 쌍처럼 사용 빈도가 높은 본초끼리 먼저 조합될 가능성이 크다. 따라서 단순히 빈도를 살펴서는 올바른 결과를 얻기 어렵다.

이러한 문제를 보완하기 위해 본 연구에서는 본초 조합을 찾을 때 단순 빈도를 대신하여 G-score를 이용하였다. G-score는 Ted Dunning이 제안한 방식으로 두 범주형 변수의 독립성을 비교하는 방법이다. 따라서 변수의 빈도와 상관없이 두 변수가 얼마나 독립적으로 나타나지 않는가를 설명해 준다. 두 변수가 함께 관찰된 실제 관찰값(observed value)을 O, 두 변수가 함께 관찰될 기댓값(Expected value)을 E라고 하였을 때, G-score의 근삿값은 아래와 같은 공식으로 계산할 수 있다.¹⁰⁾¹¹⁾

$$G \approx 2 \left(O \cdot \log \frac{O}{E} - (O - E) \right)$$

이처럼 BPE와 G-Score를 바탕으로 현존하는 많은 처방을 학습시켜 모델을 만들고 이 모델을 가지고 임의의 처방을 분해한다면 오늘날 존재하는 ‘후손 처방’을 ‘조상 처방’의 가감방으로 분해해 내는 효과를 가져올 것이라고 기대하였다.

9) Gage, Philip. A New Algorithm for Data Compression. The C User Journal. 1994. 12(2). pp.23-38.

10) Stefan Bordag. A Comparison of Co-occurrence and Similarity Measures as Simulations of Context. Computational Linguistics and Intelligent Text Processing. Alexander Gelbukh. Computational Linguistics and Intelligent Text Processing. Springer. 2008. pp.52-63.

11) 오준호. 한의학 고문헌 텍스트 분석을 위한 비지도학습 기반 단어 추출 방법 비교. 대한한의학원전학회지. 2019. 32(3). pp.47-57.

3. 처방 분해 실례

본 연구에서는 BPE와 G-Score를 바탕으로 처방 분해를 수행하기 위한 기계 학습 모델을 만들고, 학습 데이터를 이용해 이를 학습시킨 다음 테스트 데이터에 적용해 결과를 도출하였다.

3.1. 학습 데이터 준비 및 학습

학습 데이터로는 <한국전통지식포털>에 공개된 ‘전통의학처방’ 데이터¹²⁾를 선택하였다. 이 전통의학처방 데이터는 신뢰할 수 있는 기관에서 인터넷을 통해 공개한 처방 데이터로서 처방의 개수가 비교적 많고 각각의 본초 구성 정보가 함께 담겨 있다. 웹 스크래핑(web scraping) 기법을 이용하여 전체 데이터를 수집한 뒤 처방의 구성 본초 부분만을 정리하여 학습에 이용하였다. 이렇게 수집한 학습 데이터는 20,120종의 처방 데이터와 1,826종의 구성 본초로 이루어졌다.

수집한 데이터를 학습 데이터로 삼아 앞에서 설명한 BPE 및 G-score 방법을 통해 처방 분해 모델을 만들고 학습시켰다. 학습 과정에서 전체 출현 빈도가 5회 이하인 본초 혹은 본초 쌍은 제외했다. 또 관찰값과 기댓값 사이에 유의미한 차이가 없는 경우 (G-score가 0 이하인 경우)도 결과에서 제외했다. 이런 과정을 거쳐 더 이상 새로운 본초 쌍이 도출되지 않을 때까지 모델을 학습시켰다.

3.2. 테스트 데이터 처방 분해 결과

테스트 데이터로는 한방건강보험 기준처방(56종)¹³⁾을 사용하였다. 한방건강보험 기준처방은 인지도가 높으면서도 임상에서 사용 빈도가 높은 처방들이다. 학습된 처방 분해 모델을 적용하여 56종을 모두 분해하였으나 본고에서는 지면이 한정되어 있으므로 구성 본초의 개수가 많은 6가지 처방(오적산, 복령보심탕, 청서익기탕, 연교괘독산, 삼출진비탕, 반하백출천마탕)의 분해 결과를 중심으로 살펴보고

자 한다. 분해 결과를 정량적으로 평가하기 어렵기 때문에 현대의 전문가 방해(方解)와 비교하여 정성적인 비교를 수행하였다. 현대 방해는 운용잡의 저술¹⁴⁾을 참고하였다.¹⁵⁾ 처방에 대한 기본적인 설명은 <동의보감>에서 인용하여 실었으나¹⁶⁾ 처방 자체의 기원이나 효능, 다른 의서에서의 견해 등에 대해서는 따로 설명하지 않았다.

가. 오적산(五積散)

오적산에 대해 <동의보감>에서는 “풍한에 감축되고 손상되어 머리가 아프고 몸이 쭈시며, 사지가 싸늘하고 가슴과 배가 아프며, 구토와 설사를 하는 경우를 치료한다. 또, 날것, 찬 음식으로 속을 상하거나 걸음으로 풍랭을 받은 것에도 쓴다.”¹⁷⁾라고 하였다.

운용잡은 숙지황을 뺀 사물탕(四物湯), 지각길경탕(枳殼桔梗湯), 평위산(平胃散), 이진탕(二陳湯), 마황탕(麻黃湯), 계지탕(桂枝湯)을 함방하고 건강(乾薑)과 백지(白芷)를 더한 처방이라고 오적산을 방해하였다.¹⁸⁾

기계 학습을 통한 처방 분해 결과(<표 2> 참조)를 보면, 당귀-천궁-백지-작약(사물탕 유사), 반하-진피-감초(이진탕 일부), 후박-창출(평위산 일부), 지각-길경(지각길경탕)의 조합이 관찰되었다.

처방 분해 결과는 운용잡의 방해와 대동소이하였다. 분해 결과 [이진탕-평위산 조합]과 [백지가 추

14) 운용잡. 동의방제와 처방해설(개정판). 서울. 의성당. 2002.

15) 처방(방제) 해설에 대한 여러 가지 서적 가운데 운용잡의 서적을 선택한 이유는 다음과 같다. 첫째, 처방을 기본방과 부방의 관계로 설명하고 있다는 점이 본 연구의 관점과 일치한다. 둘째, 다루고 있는 처방의 수가 많고 해설이 풍부하며 관련 서적을 폭넓게 인용하고 있어 처방에 대한 균형 잡힌 설명을 기대할 수 있다. 셋째, 1998년 간행된 이래로 현재까지 꾸준히 증보되어 내용의 신빙성이 상대적으로 높다.

16) 이하 <동의보감>의 원문과 번역문은 한국한의학연구원에서 제공하고 있는 <한의학교전DB>를 사용하였다. 한국한의학연구원. 한의학교전DB. [cited on Jan 8, 2023]. Available from: <https://mediclassics.kr>

17) 동의보감. 잡병편. 한(寒). 五積散 “治感傷風寒, 頭痛, 身疼, 四肢逆冷, 胸腹作痛, 嘔吐泄瀉. 或內傷生冷, 外感風冷, 并皆主之.”

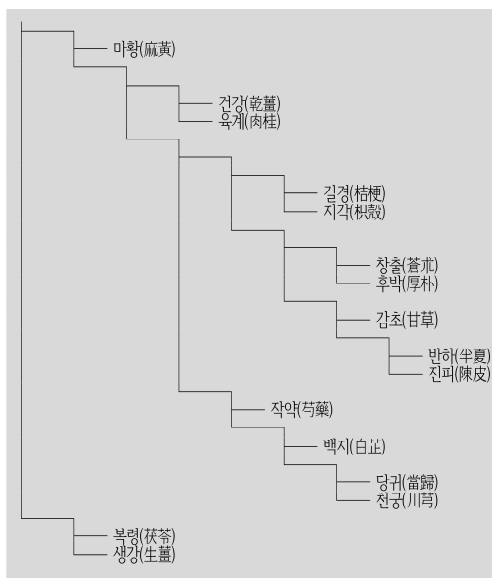
18) 운용잡. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.96.

12) 특허청. 한국전통지식포털. [cited on Jan 10, 2013]. Available from: <http://www.koreantk.com>

13) 행정규칙(보건복지부고시 제2018-157호). 한약제제 급여 목록 및 상한금액표.

가된 사물탕 조합]이 결합하여 오적산의 중심을 이루고 있다는 사실을 알 수 있었다. 다만 복령이 생강과 조합되면서 이진탕 조합[반하-진피-복령-감초]이 온전히 드러나지는 않았다. 또 백지는 당귀-천궁과 먼저 조합되었고, 건강도 육계와 우선 조합되어 윤용갑이 설명한 마황탕과 계지탕의 의미를 찾을 수 없었다.

<표 2> 오적산 분해도 (BPE 및 G-score 기반)



나. 복령보심탕(茯苓補心湯)

복령보심탕에 대해 《동의보감》에서는 “마음을 많이 써서 생긴 토혈을 치료한다.”¹⁹⁾라고 하였다.

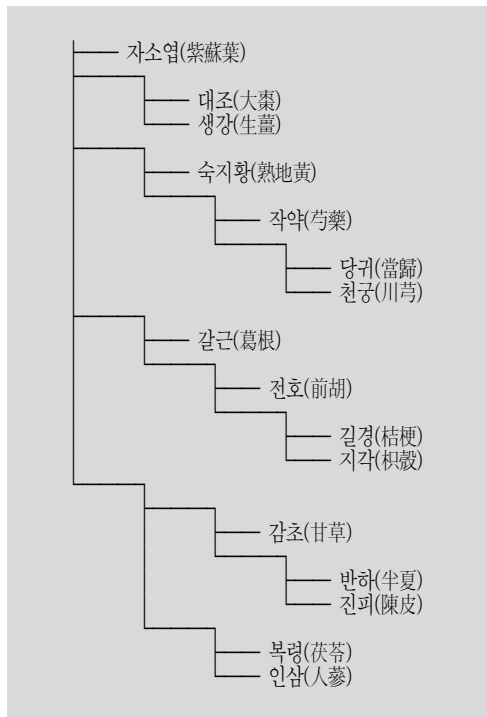
윤용갑은 사물탕, 백출을 뺀 사군자탕(四君子湯), 이진탕, 지각길경탕을 함방하고 갈근(葛根)·소엽(蘇葉)·전호(前胡)를 더한 처방이라고 하였다. 아울러 사물탕(四物湯)과 삼소음(參蘇飲)의 함방으로도 볼 수 있다고 하였다.²⁰⁾

기계 학습을 통한 처방 분해 결과(<표 3> 참조)를 보면, 당귀-천궁-작약-숙지황(사물탕), 인삼-복령(사군자탕 일부), 반하-진피-감초(이진탕 일부),

지각-길경(지각길경탕) 조합이 드러났다.

처방 분해 결과는 윤용갑의 방해와 같았다. 육군자탕(이진탕-사군자탕), 사물탕, 지각-길경(-갈근-전호) 조합이 복령보심탕의 중심을 이루고 있다는 사실을 파악할 수 있었다.

<표 3> 복령보심탕 분해도 (BPE 및 G-score 기반)



다. 청서익기탕(淸暑益氣湯)

청서익기탕에 대해 《동의보감》에서는 “늦여름에 습열이 사람을 혼증하여 사지가 노곤하고 정신이 없으며, 움직이기 싫어하고 몸에 열이 나면서 번갈이 있으며, 소변이 누렇고 잦으며, 대변이 무르고 잦거나 설사나 이질이 있으며, 음식 생각이 없고 숨이 차면서 자한이 있는 경우를 치료한다.”²¹⁾라고 하였다.

윤용갑은 이묘산(二妙散)[창출, 황백]에 택사를 더한 것이 이 처방의 기본방이며, 생맥산(生脈散)[황기, 감초, 이공산(異功散)[인삼, 백출, 감초, 진피, 청피],

19) 동의보감. 내경편(內景篇). 혈(血). 茯苓補心湯 “治勞心吐血.”

20) 윤용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.303.

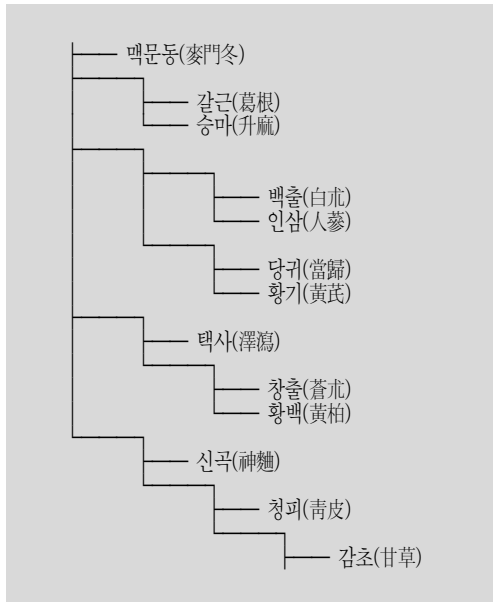
21) 동의보감. 잡병편. 서(暑) 淸暑益氣湯 “治長夏濕熱蒸人, 四肢困倦, 精神短少, 懶於動作, 身熱煩渴, 小便黃而數, 大便澀而頻, 或泄或痢, 不思飲食, 氣促自汗.”

승마갈근탕(升麻葛根湯)[승마, 갈근, 감초]의 의미를 담고 있다고 하였다.²²⁾

기계 학습을 통한 처방 분해 결과(<표 4> 참조)를 보면, 창출-황백(이묘산)에 택사를 더한 조합이 보이고, [인삼-백출 조합]과 [당귀-황기 조합]이 먼저 일군을 형성하고 있으며, 여기에 [승마-갈근 조합]이 나타났다. 또 [진피-감초-청피-신곡 조합]도 나타났다.

처방 분해 결과는 운용갑의 방해와 유사하였다. 분해 결과를 보면 [창출-황백-택사 조합]을 비롯하여 [당귀-황기-인삼-백출 조합], [감초-진피-청피-신곡 조합], [갈근-승마 조합]이 청서익기탕의 중심을 이루고 있었다. 다만 뒤의 3가지 조합은 보중익기탕(補中益氣湯)[인삼, 백출, 당귀, 황기, 승마, 시호, 진피, 감초]을 떠올리게 한다. 따라서 보중익기탕을 늦여름 습열 병증에 맞게 가감하고 이묘산과 합방한 처방이라고 운용갑의 방해를 보충할 수 있다. 인삼-백출-당귀-황기 조합이 처방 분해 결과 전면에 드러나면서 보중익기탕의 흔적을 보다 쉽게 파악할 수 있었다.

<표 4> 청서익기탕 분해도 (BPE 및 G-score 기반)



22) 운용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.177.

진피(陳皮)

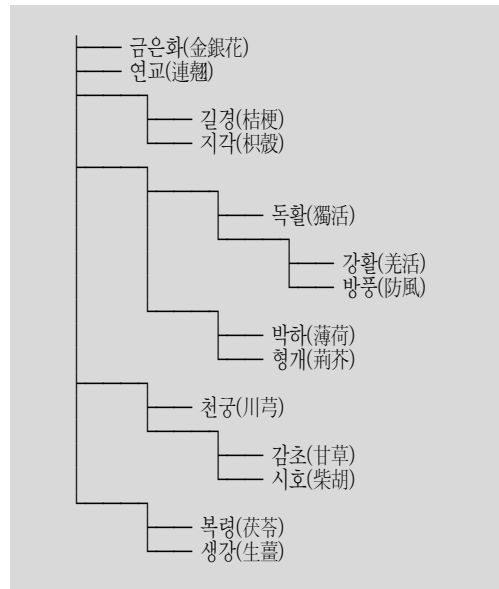
라. 연교패독산(連翹敗毒散)

연교패독산에 대해 《동의보감》에서는 “옹저 초기에 오한, 발열이 심하고, 심하면 머리가 아프고 손발이 당겨 상한과 비슷한 것을 치료한다.”²³⁾라고 하였다.

운용갑은 형방패독산(荊防敗毒散)에서 인삼 대신 금은화 연교를 더한 처방이라고 하였다. 형방패독산은 인삼패독산(人蔘敗毒散)에 형개와 방풍을 더한 처방이다.²⁴⁾ 인삼패독산은 강활·독활·시호·전호·길경·지각·천궁·복령·인삼·감초·생강·박하로 이루어진 처방이다.

기계 학습을 통한 처방 분해 결과(<표 5> 참조)를 보면, 특정 처방의 흔적이 잘 드러나지는 않지만 [지각-길경], [강활-방풍-독활], [형개-박하] 등 유사한 약성을 가진 본초 조합을 찾아볼 수 있었다. 금은화와 연교는 다른 본초와 조합을 이루고 있지 않았다.

<표 5> 연교패독산 분해도 (BPE 및 G-score 기반)



23) 동의보감. 잡병편. 옹저(癰疽) 連翹敗毒散 “治癰疽初發，憎寒壯熱，甚者頭痛拘急，狀似傷寒。”

24) 운용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.830.

윤용갑의 설명에 따르면 연교패독산은 인삼패독산의 변방이다. 그러나 이전에 살펴보았던 처방과 달리 다른 처방의 합방(合方) 관계로 설명하기 어려운 처방이다. 처방 분해 결과 역시 그러했다. 그렇지만 처방 분해 결과 약성이 비슷한 약재들끼리 조합을 이루고 있어 전체적인 처방의 성격을 파악하는데 도움을 받을 수 있었다.

마. 삼출건비탕(參朮健脾湯)

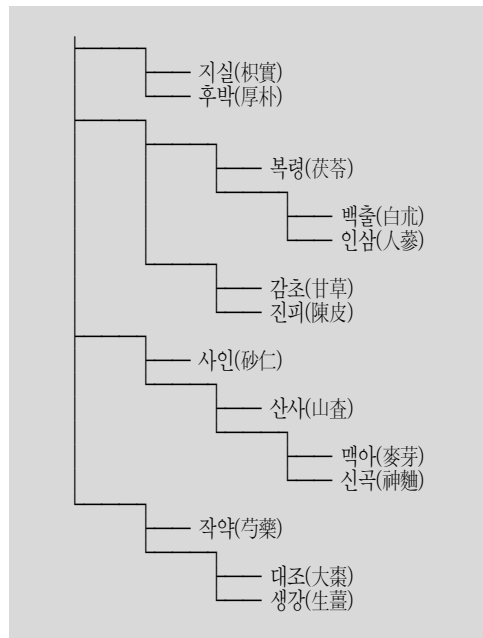
삼출건비탕에 대해 《동의보감》에서는 “비를 든든히 하고 위(胃)를 길러서 음식을 소화시킨다.”²⁵⁾라고 하였다.

윤용갑은 국벽지출환(麴蘗枳朮丸)에 사군자탕(四君子湯), 평위산(平胃散), 작약감초탕(芍藥甘草湯)을 합방하고 산사²⁶⁾를 배합한 처방이라고 하였다.²⁷⁾ 지출환은 지실과 백출로 이루어진 처방이고, 국벽지출환은 여기에 신곡과 맥아를 추가한 처방이다.

기계 학습을 통한 처방 분해 결과(<표 6> 참조)를 보면, 인삼-백출-복령-진피-감초(이공산, 사군자탕)에 지실-후박, 신곡-맥아-산사-사인 등의 소도제(消導劑)를 더한 처방이라고 할 수 있다.

윤용갑의 방해와 처방 분해 결과 사이에는 서로 차이가 있다. 윤용갑은 삼출건비탕의 기본방이 국벽지출환이라고 하였다. 즉 지실은 지출환의 의미, 후박의 평위산의 의미라고 설명한 것이다. 그러나 처방 분해 결과를 보면 삼출건비탕의 기본방은 이공산(異功散)이라고 할 수 있으며, 여기에 후박-지실, 신곡-맥아-산사-사인 등 유사한 효능을 가진 약재를 추가한 것이라고 해석된다. 양자 사이에 의견 차이가 드러나는 결과였다.

<표 6> 삼출건비탕 분해도 (BPE 및 G-score 기반)



마. 반하백출천마탕(半夏白朮天麻湯)

반하백출천마탕에 대해 《동의보감》에서는 “비위가 허약하여 생긴 담궐두통을 치료한다. 그 증상은 다음과 같다. 머리가 터질 것처럼 아프고 몸이 산처럼 무거우며, 사지가 싸늘하고 토하며 어지럽고, 바람이나 구름 속에 있는 것처럼 눈을 뜰 수 없다.”²⁸⁾라고 하였다.

윤용갑은 육군자탕에서 감초를 빼고 맥아-신곡, 창출, 택사-황백, 천마, 황기, 건강을 더한 처방이라고 하였다.²⁹⁾

기계 학습을 통한 처방 분해 결과(<표 7> 참조)를 보면, [반하-진피-생강], [창출-황백-택사-천마-건강], [인삼-백출-복령-황기], [맥아-신곡] 등의 하위 조합으로 구성되어 있음을 알 수 있다.

윤용갑의 해설과 처방 분해 결과가 유사했지만, 차이가 있는 부분도 있었다. 처방 분해 결과에서는

25) 동의보감. 잡병편. 내상(內傷). 參朮健脾湯 “健脾養胃, 運化飲食”

26) 윤용갑의 원저에는 “산조인”이라고 되어 있으나 산사의 오기(誤記)로 보여 수정하였다.

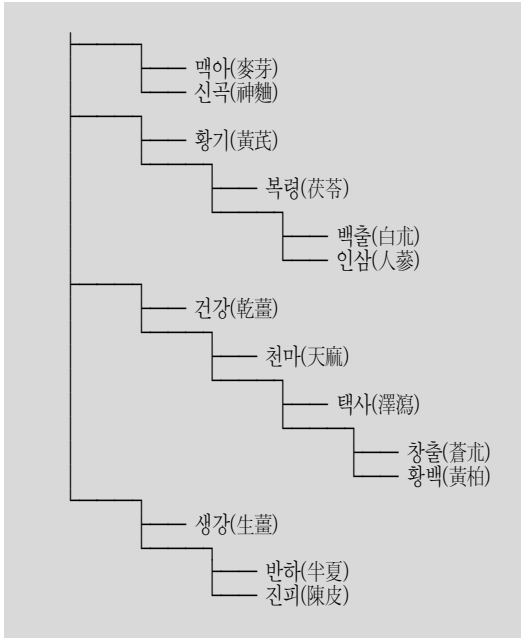
27) 윤용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.158.

28) 동의보감. 외형편(外形篇). 두(頭). 半夏白朮天麻湯 “治脾胃虛弱, 痰厥頭痛. 其證, 頭苦痛如裂, 身重如山, 四肢厥冷, 嘔吐, 眩暈, 目不敢開, 如在風雲中.”

29) 윤용갑. 동의방제와 처방해설(개정판). 서울. 의성당. 2002. p.339, 396.

[인삼-백출-복령 조합]이 황기와 먼저 결합하면서 육군자탕 조합이 잘 드러나지 않았다. 반면 율용갑이 지적하지 않았던 이묘산[창출-황백]에 택사가 결합한 조합을 쉽게 발견할 수 있었다. 양자의 결과를 서로 보완한다면, 반하백출천마탕은 육군자탕과 이묘산을 중심으로 형성된 처방이라고 할 수 있을 것이다. 반하백출천마탕 처방 이름에 있는 반하, 백출, 천마는 각각 처방 분해 결과의 3가지 주요 조합, 즉 [반하-진피-생강], [백출-인삼-복령-황기], [천마-창출-황백-택사]를 대표하는 것처럼 보이기도 한다.

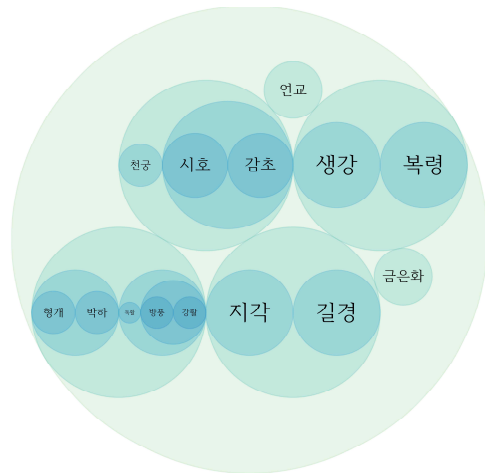
<표 7> 반하백출천마탕 분해도 (BPE 및 G-score 기반)



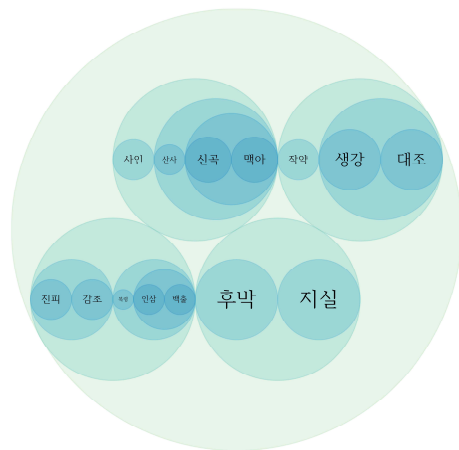
3.3. 처방 분해의 시각화

기계 학습을 이용한 처방 분해 방법은 본초 조합 사이의 관계를 정량적인 수치로 도출하기 때문에 그 결과를 다양한 차트나 다이어그램으로 시각화할 수 있다. 처방을 시각적으로 표현하면 본초의 친소 관계와 결합 관계를 더 직관적으로 살펴볼 수 있다. 연교패독산과 삼출견비탕의 시각화 사례를 신는다 (<그림 1~2>).³⁰⁾

시각화 사례에서 연교패독산은 연교·금은화 이외에 4가지 하위 본초 조합으로 구성되어 있음을 볼 수 있다. 삼출견비탕의 경우에도 4가지 하위 본초 조합을 확인할 수 있다. 각 하위 본초 조합 역시 본초 사이의 상대적인 관계를 파악할 수 있다. 예를 들어 삼출견비탕의 [진피-감초-복령-인삼-백출 조합]은 [진피-감초 조합]과 [복령-인삼-백출 조합]의 결합으로 이해할 수 있다.



<그림 1> 연교패독산 분해도



<그림 2> 삼출견비탕 분해도

30) 여기서 원의 포함 관계는 본초의 결합 관계를 나타내며

원의 크기는 의미를 갖지 않는다.

Ⅲ. 고찰

글을 마무리하기에 앞서 한의학 처방을 어떻게 이해할 것인가에 대해 조금 더 고찰해 보고자 한다. 먼저 처방을 모델링하는 문제이다. 세상의 현상을 분석하기 위해서는 모델이 필요하다. 모델은 인간이 만들어낸 구조이므로 필연적으로 실재하는 현상을 단순화시킬 수밖에 없다. 지나치게 단순화시킨 모델은 실재하는 의미를 훼손시킬 여지가 있고, 그렇지 않은 모델은 인간이 그 결과를 파악하기 어려울 수 있다. 따라서 단순하면서도 인간의 암묵지와 들어맞는 모델을 찾는 것이 관건이다.

본 연구에서 네트워크 모델 대신 BPE로 모델을 생성한 것은 이러한 문제 때문이다. 한의학의 처방은 본초로 이루어진 하나의 네트워크이다. 따라서 네트워크 모델을 이용하면 처방에 담겨 있는 정보를 충실히 담아낼 수 있다. 그러나 네트워크 모델은 지나치게 복잡해지기 쉽다. 처방의 경우 구성 본초의 개수가 늘어나면 네트워크의 복잡성은 크게 높아진다. 예를 들어 4가지 본초로 이루어진 사군자탕[인삼-백출-복령-감초]의 경우, 인삼-백출, 인삼-복령, 인삼-감초, 백출-복령, 백출-감초, 복령-감초처럼 6가지 관계를 형성하게 된다. 구성 본초가 10가지가 되면 본초 사이의 관계는 45가지로 불어난다. 또한 [인삼-백출]-복령, [인삼-백출]-감초처럼 조합의 크기를 늘리면 그 복잡도는 더욱 높아진다. 따라서 처방을 본초로 이루어진 네트워크로 정량화하거나 시각화할 수는 있지만, 인간이 그 결과를 직관적으로 이해하기 어려울 수 있다. 이에 본 연구에서는 본초와 본초 사이의 모든 관계를 고려하기보다는 가장 중요한 관계부터 묶어 나가는 방법을 통해 대상 처방을 가장 단순하게 구조화시키려고 하였다. 이 때문에 모델을 만들 때 하나의 본초가 여러 관계에서 중심이 될 수 있다는 문제를 제외할 수밖에 없었다.

다음으로 본초와 본초 사이의 관계를 어떻게 정량화할 것인가의 문제이다. 본 연구에서는 단순 빈도가 적합하지 못하다고 생각하여 G-score를 이용하였다. 임상에는 상견 질환이 있는가 하면 회소 질환도 있다. 전자에 사용되는 약제는 빈번히 사용되고 후자에 사용되는 약제를 드물게 사용될 것이다.

따라서 단순히 빈도만을 가지고 본초 사이의 관계를 정량화시킨다면 빈번히 사용되는 약제는 과장되고 드물게 사용되는 약제는 의미가 축소될 수 있다. 따라서 관찰값과 기댓값을 비교하는 방법이 필요하다. 이런 문제를 회피하기 위해서는 G-Score 외에도 Chi-Squared-Score, Mutual Information, Association strength, Jaccard distance 등 기존에 알려진 평가 방법을 고려할 수 있다.³¹⁾ 또한 이런 취지에 맞게 연구자가 새로운 지표를 제시할 수도 있을 것이다. 이런 연구가 지속된다면 어떤 정량화 방법이 가장 효과적인지에 대한 비교도 가능할 것이다.

다음은 본초 용량에 대한 문제이다. 똑같은 본초 조합이라고 하더라도 각 본초의 용량이 달라지면 처방의 방해(方解)가 달라질 수 있다. 본 연구에서는 확보할 수 있는 데이터의 한계 때문에 본초의 용량을 고려하지 않았다.³²⁾ 만약 본초 용량을 고려한다면 본초의 용량을 가중치로써 활용할 수 있을 것이고, 그에 따른 모델과 분석 방법도 고민되어야 할 것이다. 문제는 학습 데이터이다. 기계 학습 방법을 활용하기 위해서는 데이터가 동일한 기준 아래 정리되어야 하며, 충분한 양이 존재해야 한다. 하지만 전통의학에서 도량형은 하나의 연구 분야를 형성할 만큼 복잡한 문제이다. 도량형 기준이 시대마다 서적마다 조금씩 다르다는 점, 다양한 단위가 사용되었다는 점, 때때로 용량이 기재되지 않았거나 범위로 기재되었다는 점, 수량으로 기재되었을 때 약재마다 무게나 부피가 달라진다는 점, 처방마다 하루 복용 횟수가 다르므로 1회 복용량과 1일 복용량을 구분해야 한다는 점 등 해결해야 할 문제가 많다.

31) 오준호. 고의서에 나타난 경혈과 병증의 연관성 측정 및 시각화. Korean Journal of Acupuncture. 2016. 33(1). pp.18-32.

32) 아래와 같이 본초 용량을 고려한 선행 연구도 존재한다. 다만 한정된 서적에 수록된 비교적 적은 수의 처방을 대상으로 하였기 때문에 본 연구에서는 적용할 수 없었다. 박대식, 이부균, 이병욱. 방제의 본초 중량비를 활용한 방제 비교 방안에 관한 연구. 대한한의학방제학회지. 2013. 21(2). pp.121-132. 김기욱, 김태열, 이병욱. 본초 비율의 순위를 이용한 문헌의 특징 분석 방법. 대한한의학원전학회지. 2014. 27(4). pp.73-84.

앞으로 이런 문제를 극복한 동아시아 전통의학 처방 데이터가 구축된다면 처방에 관한 정량적 연구가 더 활성화될 수 있을 것이다.

다른 측면에서 처방의 성격을 이해하는데 구성 본초의 용량이 필수 불가결한가의 문제도 생각해 보아야 한다. 본초 용량이 추가적인 정보를 알려주지만, 처방에서 가장 중요한 것은 구성 본초의 종류이다. 용량이 배제된 구성 본초 데이터만으로도 처방의 상당 부분을 설명할 수 있기 때문이다. 따라서 구성 본초의 용량이 추가된 진보된 동아시아 전통의학 처방 데이터의 구축을 기다리는 동안에도 연구자들이 할 수 있는 일은 적지 않을 것이다.

마지막으로 전통적인 처방 이론에 대한 문제이다. 한의학에는 군(君)·신(臣)·좌(佐)·사(使)와 같이 처방 안에서 본초의 역할을 설명하는 이론이 존재한다. 본 연구에서는 이러한 전통 이론을 고려하지 않았다. 이에 대해서도 본초 용량 문제처럼 의미의 문제와 데이터의 문제로 나누어 볼 수 있다. 전통 이론 자체의 의미는 전문가마다 견해가 다를 수 있다. 별도의 주제이기 때문에 이에 대해 자세히 다룰 수는 없지만 군신좌사 이론은 본초의 역할을 비유적으로 표현한 설명 방법일 뿐 분명한 기준이나 필연적인 개연성이 내포되어 있지 않다. 옛사람들도 모든 경우에 이 방법을 사용한 것이 아니라 군신좌사로 처방을 설명하는 것이 효과적일 때만 사용하였다. 따라서 군신좌사 이론은 처방을 이해하는 데 참고할 수 있는 '이론'이지 반드시 고려해야 할 '원리'는 아니다.

다음으로 데이터의 문제이다. 현실적으로 군신좌사 이론을 반영한 처방 데이터를 만들 수 있을지 의문이다. 우선 신뢰하고 인용할만한 자료가 없다. 고의서에서도 일부 처방에 대해서만 군신좌사 분석 결과를 찾을 수 있을 뿐이다. 고의서 전체를 조사한다고 하더라도 많은 처방에 대해서는 군신좌사 분석 결과를 찾을 수 없다. 오늘날 전문가들이 처방마다 어떤 본초가 군약이고 어떤 본초가 신약인지 나누는 방식도 한계가 뚜렷하다. 2~3명의 전문가에게 100여 가지 처방을 주고 처방마다 군신좌사에 해당하는 본초가 무엇인지 물은 다음 교차 검증하였을 때 그

결과는 매우 다를 것이기 때문이다. 이 밖에 용량을 기준으로 한 방법 등 다른 대안도 있을 수 있다.³³⁾ 하지만 군신좌사 이론을 처방 데이터에 어떻게 도입할 것인가, 그것이 처방을 이해하는데 어떤 효용을 줄 수 있을 것인가에 대해서는 연구가 충분하지 않다.

IV. 결론

지금까지 한의학 처방을 분해하기 위해 BPE와 G-score를 이용한 방법을 제안하고, 학습 데이터를 이용해 이 방법을 학습시킨 다음 오적산·복령보심탕·청서익기탕·연교괘독산·삼출건비탕·반하백출천마탕 등 6가지 처방에 적용하여 결과를 살펴보았다.

기계 학습을 이용한 처방 분석 결과는 큰 틀에서 현대 전문가의 방해와 많이 다르지 않았다. 다만 오적산과 같이 부분적으로 의미가 잘 드러나지 않는 경우도 있었고, 반대로 청서익기탕과 같이 처방 분해를 통해 새로운 지견을 얻을 수 있는 경우도 있었다. 또한 삼출건비탕과 같이 양자의 견해 사이에 차이가 있는 경우도 있었고, 반하백출천마탕과 같이 양자의 견해를 서로 보완하여 더 나은 결론에 이를 수 있는 경우도 있었다.

본 연구는 연구 방법의 우수성을 증명하기보다는 한의학 처방을 분해하고 시각화시키고자 하는 하나의 시도이다. 테스트 결과를 평가할만한 지표나 방법이 고안되어 있지 않고 검토 결과도 적기 때문에 현재로서 본 연구에서 제안한 방법의 효용을 판단하기는 어렵다. 다만 테스트한 처방에서 전문가 의견과 유사한 결과를 보이고 있으므로 한의학의 처방을 이해하기 위한 하나의 접근 방법으로서 교육·연구·임상 현장에서 참고가 되기를 희망한다.

33) 박한수, 이병욱, 이부균. 군신좌사 개념을 도입한 방제 검색 및 비교 시스템에 관한 연구. 대한한의학방제학회지. 2014. 22(2). pp.45-54.

감사의 말씀

본 연구는 한국한의학연구원 주요사업 “AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축(KSN1922110)”의 지원을 받아 수행되었습니다.

References

1. Gage, Philip. A New Algorithm for Data Compression. *The C User Journal*. 1994. 12(2). pp.23-38.
2. Stefan Bordag. A Comparison of Co-occurrence and Similarity Measures as Simulations of Context. *Computational Linguistics and Intelligent Text Processing*. Alexander Gelbukh. *Computational Linguistics and Intelligent Text Processing*. Springer. 2008. pp 52-63.
https://doi.org/10.1007/978-3-540-78135-6_5
3. 김기욱, 김태열, 이병욱. 본초 비율의 순위를 이용한 문헌의 특징 분석 방법. *대한한의학원전학회지*. 2014. 27(4). pp.73-84.
<https://doi.org/10.14369/skmc.2014.27.4.073>
4. 김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. *대한한의학원전학회지*. 2008. 21(1). pp.143-151.
5. 김정훈, 이병욱. 본초 조합을 이용한 方劑의 類方分析 DB 구축 연구 -『東醫寶鑑』과『方劑學』을 중심으로 -. *대한한의학원전학회지*. 2008. 21(1). pp.123-141.
6. 박대식, 이부균, 이병욱. 방제의 본초 중량비를 활용한 방제 비교 방안에 관한 연구. *대한한의학방제학회지*. 2013. 21(2). pp.121-132.
<https://doi.org/10.14374/HFS.2013.21.2.121>
7. 박한수, 이병욱, 이부균. 군신좌사 개념을 도입한 방제 검색 및 비교 시스템에 관한 연구. *대한한의학방제학회지*. 2014. 22(2). pp.45-54.
<https://doi.org/10.14374/HFS.2014.22.2.045>
8. 방민우, 김기욱, 이병욱. 본초 구성을 이용한 방제의 효능 추론 및 분류 방법에 관한 연구. *대한한의학방제학회지*. 2017. 25(1). pp.29-38.
<https://doi.org/10.14374/HFS.2017.25.1.29>
9. 오준호. 고의서에 나타난 경혈과 병증의 연관성 측정 및 시각화. *Korean Journal of Acupuncture*. 2016. 33(1). pp.18-32.
<https://doi.org/10.14406/acu.2016.003>
10. 오준호. 한의학 고문헌 텍스트 분석을 위한 비지도학습 기반 단어 추출 방법 비교. *대한한의학원전학회지*. 2019. 32(3). pp.47-57.
<https://doi.org/10.14369/jkmc.2019.32.3.047>
11. 윤용갑. *동의방제와 처방해설(개정판)*. 서울. 의성당. 2002.
12. 윤홍걸 외 3인. 기존방제를 기반한 처방분석 방법에 관한 연구. *대한한의학원전학회지*. 2021. 34(4). pp.65-78.
<https://doi.org/10.14369/jkmc.2021.34.4.065>
13. 이병욱, 정훈, 권영규. 인터넷상 동의보감 방제의 약물구성 계보 표현 방법에 관한 연구. *한국 의사학회지*. 2008. 21(2). pp.41-47.
14. 彭怀仁 等. *中医方剂大辞典*. 北京. 人民卫生出版社. 1993.
15. 특허청. 한국전통지식포털. [cited on Jan 10, 2013]. Available from: <http://www.koreantk.com>
16. 한국한의학연구원. 한의학고전DB. [cited on Jan 8, 2023]. Available from: <https://mediclassics.kr>