

# 데이터 마이닝을 이용한 대변과 약물간의 연관성 분석 -방약합편을 중심으로-

송영섭<sup>1)</sup> · 양동훈<sup>2)</sup> · 박영재<sup>1)</sup> · 박영배<sup>1)\*</sup>

1) 경희대학교 한의과대학 진단·생기능의학과학교실 / 2) 자생한방병원

---

## Abstract

---

A study of relationship between excrement and materia medica in Bangyakhappyeon based on the data mining analysis

Young Sup Song<sup>1)</sup> · Dong-hoon Yang<sup>2)</sup> · Young-Jae Park<sup>1)</sup> · Young-Bae Park<sup>1)\*</sup>

1) Dept. of Biofunctional Medicine & Diagnostics, College of Oriental Medicine, Kyung Hee University

2) Jaseng Hospital of Oriental Medicine

### Purpose

Nowadays excrement-related disease that repeats constipation and diarrhea is on the increase due to the change of dietary and lack of exercise, etc.

We analyzed Bangyakhappyeon in order to find out the materia medica which is used for the excrement patterns.

### Methods

The database used in present thesis consist of disease pattern, nature of medicinals and materia medica from Bangyakhappyeon was constructed. We analyzed the nature of medicinals of excrement patterns(or symptom) by frequency analysis and network analysis, and also searched main materia medica of excrement patterns(or symptom) by frequency analysis and rule mining.

### Results

We analyzed the nature of medicinals of excrement patterns(or symptom) in Bangyakhappyeon. And we researched the high frequency materia medica, high specificity materia medica and high frequent paired-drugs as main materia medica of excrement patterns(or symptom).

### Conclusion

This study found the information about frequency relationship between excrement patterns(or symptoms) and materia medica.

### Key Words

excrement, Bangyakhappyeon, herbs, database

---

\* 교신저자 : 박영배 / 소속 : 경희대학교 한의과대학 진단·생기능의학과학교실

TEL : 02-958-9195 / E-mail : bmppark@khu.ac.kr

투고일 : 2012년7월19일; 수정일 : 2012년8월14일; 게재확정일 : 2012년8월20일

## I. 서론

데이터 마이닝은 많은 양의 데이터 안에서 어떠한 방법(순차 패턴, 유사성 등)을 통해 알고 싶은 정보를 찾는 과정이다. 데이터 마이닝은 데이터에서 필요한 부분의 정보를 구하거나, 의사결정을 하게 해주는 정보를 제공하는데 사용되는 기술이라고 할 수 있다.

한의학 서적의 데이터베이스 구축은 현재 진행되고 있는 한의학 자료의 객관화, 정량화 작업에 도움을 줄 수 있다. 진단과 치료 평가 과정에 있어서 상당한 부분이 의사의 주관적인 판단에 근거하고 있는 한의학의 특성상 각각의 자료와 서적들의 정리와 체계적인 분류가 쉽지 않다. 그러나 점점 방대하게 늘어나는 자료와 서적, 임상 결과 등을 체계적으로 분류하고 정리해야 할 필요성이 높아지고 있다. 방대한 자료를 정리하는 과정에서 데이터 마이닝을 이용한 서적, 임상결과 등의 분석 작업은 한의학의 객관화와 각종 임상결과의 정량화에 도움이 된다.

한의학 자료의 데이터베이스화가 이루어지면서 DB구축에 관한 연구가 나타나고<sup>1)</sup> 전자도서관을 구축하고자 하는 노력도 보이고 있다.<sup>2)</sup> DB가 구축되고 양이 많아지면서 데이터 마이닝을 통한 자료 분석도 점점 늘어나고 있는데 이<sup>3)</sup>는 온톨로지를 기반으로 한 한의학 처방 지식 관리시스템을 연구하였고, 문<sup>4)</sup> 등은 한의학 온톨로지를 통한 진단 방법에 대해 연구하였다. 김<sup>5)</sup> 등은 상한론 온톨로지 구축 방법에 대해 연구하였고 홍<sup>6)</sup> 등은 상한론, 태양병 조문을 통해 약물과 증상간의 연관성을 분석하였다. 중국에서는 문헌 60만 건, 약물 11,000건, 처방구성 85,000건 등 데이터베이스 구축이 보고되고 있고<sup>7)</sup>, 이를 기반으로 데이터 마이닝을 이용한 한의학 연구들이 이뤄지고 있다.<sup>8)9)</sup>

최근 식생활 변화, 운동부족, 인스턴트 식생활 등으로 인해 반복된 변비나 설사를 호소하는 환자가

많아지고 있다. 이런 대변 장애에 있어 큰 비중을 차지하는 질병 중 과민성 대장 증후군은 복통이나 복부 불편감의 증상을 특징으로 하며 변비나 설사를 동반하는 만성적인 질환이다. 이러한 대변 증상에 사용되는 약물에 대해 데이터 마이닝으로 분석해보기 위해 임상적으로 많이 활용되고 있는 방약합편을 선택하였다.

방약합편은 조선시대 황도연(黃度淵)이 지은 의방서(醫方書)로 동의보감을 기본으로 한 것으로, 황도연의 저술인 의방활투(醫方活套)와 손익본초(損益本草)를 합하고, 이에 용약강령(用藥綱領), 구급(救急), 금기(禁忌) 등 10여 종을 더 첨가하여 1884년에 간행되었는데 그 후로 한의학계에 대단한 영향을 미친 책이다. 본 연구는 방약합편에 대한 전체 처방, 약물을 데이터베이스화 한 양<sup>10)</sup>의 연구를 기본으로 대변 증상에 대해 Apriori라는 연관규칙분석 방법을 통하여 병증과 처방, 약물간의 연관성에 대해 분석하였다.

본 연구에서는 양<sup>10)</sup> 등의 논문을 통해 밝혀진 방약합편의 병증, 약물, 약성에 관한 데이터베이스를 바탕으로 방약합편에 대변 증상에 해당되는 부분을 찾아 분석하였다. 분석 내용은 대변 증상과 藥性(五味, 四氣, 歸經, 效能, 毒性), 대변 증상에서 많이 쓰이는 약물인 고빈도 약물, 다른 증상에 비해 대변 증상에서 사용되는 빈도가 높은 약물인 고특이도 사용 약물에 관한 것이다.

## II. 연구방법<sup>10)</sup>

방약합편의 處方, 病症, 藥物 및 藥性を 엑셀로 데이터베이스화 한 자료를 바탕으로 방약합편의 활투침선에 있는 대변 증상에 대한 약성, 고빈도 약물, 고특이도 약물을 분석하였다(Figure1).

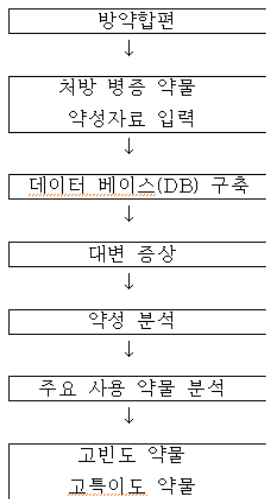


Figure 1. 연구의 흐름도

1) 병증 · 처방 · 약물 데이터 입력

a) 병증-처방-약물 데이터 입력

54개 병증에 제시되는 병증-처방 조합, 총 1413건을 입력하여 병증-처방 데이터베이스를 만들었다. 처방의 구성 약물자료로 처방-약물 데이터베이스를 만들었다. 각 처방에 해당하는 내용을 데이터베이스에 입력하였다. 병증-약물 데이터베이스는 병증-처방 데이터와 처방-약물 데이터를 통합하여 구축하였다. 이렇게 구축된 전체 데이터베이스 중

대변 증상에 해당되는 병증, 처방, 약물을 따로 찾아내어 입력하였다.

(1) 병증-처방 매트릭스

활동침선 조문을 분해하여 Microsoft Office Excel 2007을 이용하여 증상-처방의 Edge list의 형태로 입력하였다. 작성한 데이터는 Cyram netminer 3을 사용하여 매트릭스를 생성하였다(Figure 2).

(2) 처방-약물 매트릭스

방약합편 본문의 처방구성을 참조하여 처방-약물 Edge list의 형태로 모두 입력하였다. 위 방법과 같이 Cyram netminer 3을 사용하여 매트릭스를 생성하였다 (Figure 3).

(3) 병증-처방-약물 매트릭스

위의 병증-처방 매트릭스와 처방-약물 매트릭스를 통합하여 병증-처방-약물 데이터베이스를 완성하였다(Figure 4).

2) 약물의 약성 데이터 입력

방약합편에서 사용된 약물 305개의 속성을 데이

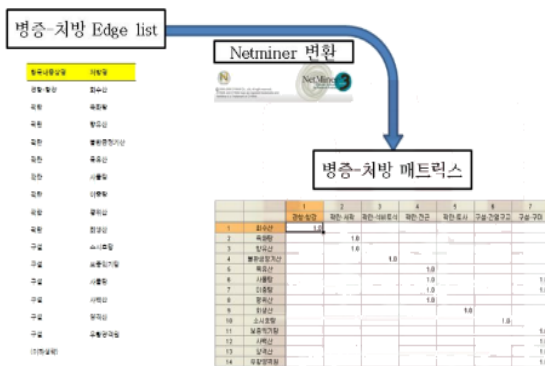


Figure 2. 병증-처방 매트릭스 변환과정

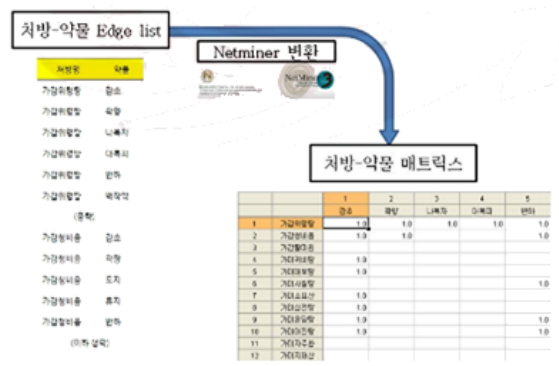


Figure 3. 처방-약물 매트릭스 변환과정



Figure 4. 병증-처방-약물 DB의 구축과정

터베이스로 구성하였다. 약물 속성은 四氣, 五味, 歸經, 效能, 毒性으로 한정하였다.

a) 약물명 표준화

수치(修治)한 약물은 수치하지 않은 약물과 같은 약물로 처리하였다(Table 1). 약물 중 같은 약물이

Table 1. 수치약물의 표준화

수치약물	수정후	수치약물	수정후	수치약물	수정후
주감초	감초	대부자포	부자	주홍화	홍화
자감초	감초	부자포	부자	주편금	황금
외강	건강	주지모	지모	주황금	황금
주당귀	당귀	주천마	천마	주황련	황련
당귀초	당귀	주작약	작약	주황백	황백
두충주초	두충	변향부	향부자		

Table 2. 이명(異名)약물의 표준화

수정전	수정후	수정전	수정후	수정전	수정후	수정전	수정후
가자피	가자	염초	망초	소합유	소합향	자울	자완
건갈	갈근	백밀	밀	회향	소회향	적두	적소두
감초절	감초	은행	백과	대숙지	숙지	정력	정력자
대감초	감초	생모근	백모근	아교주	아교	조각	조각자
생감초	감초	파고지	보골지	봉출	아출	조각인	조각자
경미	갱미	백복신	복신	숙애	애엽	진사	주사
생견지황	견지황	소맥	부소맥	황단	연단	진구	진교
검인	검실	대부자	부자	연육	연자육	굴피	진피
흑축	전우자	부자포	부자	연자	연자육	굴홍	진피
초결명	결명자	생부자	부자	식염	염	천초	천초
양강	고량강	세비육	비자	오매육	오매	총백	총
고련근피	고련피	대복자	빈랑	편뇌	용뇌	산치	치자
고백반	고반	사과	사과락	초룡담	용담초	파극	파극천
황과루	과루인	사군자육	사군자	생우절	우방자	파두상	파두
감국	국화	축사	사인	석용황	용황	해표소	해표초
두시	담두시	산사육	산사	계심	육계	호도	호도육
당귀신	당귀	산사자	산사	관계	육계	마인	화마인
당귀초	당귀	황상엽	상엽	달계	육계	마자인	화마인
대당귀	당귀	서각설	서각	인동	인동등	생황금	황금
대조육	대조	주맥	소맥	인진	인진호	조금	조금
등심	등심초	자소엽	소엽			편금	황금

Table 3. 약성의 출전

출전		약물명						
본 초 학	가자	갈근	감수	감초	강활	강황	건강	건지황
	검실	견우자	결명자	경분	계내금	계지	고량강	고련피
	고반	고본	곤포	과투인	곽향	관동화	괴화	구기자
	구맥	구관	국화	금앵자	금은화	길경	니복자	노감석
	노회	녹각교	녹각상	녹용	단삼	단향	담두시	담반
	당귀	대극	대두황련	대복피	대자석	대조	대황	도인
	독활	동규자	두충	등심초	마황	만형자	망초	맥문동
	맥아	모려	목과	목단피	목적	목통	목향	몰약
	밀	박하	반하	방기	방풍	백강잠	백개자	백과
	백두구	백렴	백모근	백부자	백선평	백자인	백작약	백지
	백질려	백출	백편두	백합	백화사	별갑	보골지	복령
	복령피	복신	부소맥	부자	봉사	비자	비해	빈랑
	사간	사과략	사군자	사삼	사인	사향	산두근	산사
	산수유	산약	산조인	삼릉	상기생	상륙	상백피	상산
	상엽	상표초	생강	생지황	서각	석결명	석고	석곡
	석창포	선복화	선회	세신	소목	소엽	소자	소합향
	소회향	속단	속지황	승마	시체	시호	신국	야교
	이출	안석향	애엽	앵속각	양기석	여지핵	연교	연단
	연자육	영양각	오가피	오령지	오매	오미자	오수유	오약
	용골	용뇌	용담초	용안육	우방자	우슬	우절	우황
육리인	울금	웅담	웅황	원지	위령선	유향	육계	
육두구	육종용	의이인	익모초	익지인	인동등	인삼	인진호	
자석	자완	자하거	저령	적복령	적석지	적소두	적작약	
전갈	전호	정력자	정향	조각자	주사	죽여	죽엽	
지각	지골피	지모	지실	지유	진교	진주	진피	
차전자	창이자	창출	찬궁	찬남성	찬련자	찬마	찬문동	
천산갑	천오	천축황	천패모	천화분	청대	청몽석	청상자	
청피	초과	초두구	축초	총	치자	침향	택사	
토복령	토사자	통초	파극천	파두	편축	포황	필발	
필징가	하고초	하수오	현수석	해금사	해동피	해조	해표초	
행인	향부자	향유	현삼	현호색	형개	호골	호도육	
호박	호초	호황련	홍화	화마인	활석	황금	황기	
황련	황백	후박						
동 의 보 감	갱미	계장	계피	귤핵	금박	당귀미	도지	동변
	류지	리어	반하국	백초상	비마자인	사당	생강피	생구채
	생탈	생리	생토	생하엽	석연육	시상	양신	양육
	엽	우담남성	원청	은박	자석영	자초용	저제	전두등
	정향피	중유분	진창미	춘다	한식면	해대	해분	해삼
형개수	호동루	황미	황송절	흑당	흑두			

다른 이름으로 불리는 경우, 본초학의 약물명 기준으로 하나의 약물로 처리하였다(Table 2).

b) 약성의 입력

약성의 출전은 아래의 표와 같다.(Table 3) 약물에 해당하는 각 四氣, 五味, 歸經, 效能, 毒性을 출전에 의거하여 입력하였다.

3) 분석도구

빈도분석과 도표작성, 네트워크분석, Apriori 알고리즘 분석을 통한 연관규칙탐색은 각각 Microsoft Office Excel 2007, Cytam Netminer 3, SPSS Clementine 12.0을 이용하였다.

### Ⅲ. 분석방법

대변 증상이 포함되어 있는 방약합편의 활투침 선의 약성 및 약물 분석(고빈도 사용 약물 탐색)은 빈도 분석하였다. 五味, 四氣, 歸經, 效能, 毒性的 약성 분석과 대변 증상에 대한 고특이도 사용 약물은 각각 네트워크 분석과 연관규칙을 통하여 분석하였다.

#### 1) 빈도분석

약성과 약물의 빈도를 분석할 때 사용한 방법으로, 출현하는 변수의 빈도를 합하는 방식이다. 대변에 관한 고빈도 사용 약물은 빈도 분석 후 상위 약물을 도표로 작성했다.

#### 2) 네트워크분석

약성의 빈도를 링크속성으로 입력하고 유사성 다차원 척도법(similarity multidimensional scaling)을 이용하여 분석하였다. 유사성 다차원 척도법은 다차원 척도의 대표적인 모형으로 가장 많이 활용되고 있으며, 유사성 자료를 근거로 하여 변수가 한 차원 상의 공간에 위치로 전환하여 링크 속성이 클수록 가깝게 나타나도록 보여주는 방법이다.

#### 3) 연관규칙분석

연관규칙분석을 위해 Apriori 알고리즘을 사용하였다. 이 방법은 대규모 데이터에서 빈발 패턴을 찾는 적합한 방법으로 대표적인 연관규칙 분석 도구로 가장 잘 알려져 있다. Apriori 알고리즘은 연관 규칙을 찾아주는 알고리즘 중에서 가장 먼저 개발되었으며, 가장 많이 쓰이는 알고리즘이다. Apriori는 후보

항목집합들의 기하급수적인 성장을 체계적으로 제어하기 위해 지지도기반 가지치기의 사용법을 개척한 연관 규칙 탐사 알고리즘이다. 그 원리는 간단하게 만약 한 항목집단이 빈발한다면 그것의 모든 부분집합들 역시 빈발해야만 한다는 것이다. 고특이도 사용 약물탐색은 antecedent에 병증을 input으로, consequent에 약물을 output으로 입력하고 minimum confidence를 20%로 설정하고 lift가 높은 약물을 탐색하였다. 연관규칙분석의 분석지표는 아래와 같다.

##### (1) Support

Support는 전체 사건에서 X와 Y가 동시에 포함될 확률을 말한다.

$$\text{Support} = P(X \cap Y)$$

##### (2) Confidence

Confidence는 X라는 항목이 포함되었을 때 이 중에서 Y라는 항목이 함께 포함될 확률이다. 즉, Confidence는 선행 항목이 선택된 상황에서 선행 항목과 후행 항목이 동시에 선택될 확률을 말하는 것이다. 이는 연관성 규칙의 강도를 나타내며 다음과 같이 조건부 확률로 나타난다.

$$\text{Confidence} = P(Y | X) = \frac{P(X \cap Y)}{P(X)}$$

##### (3) Lift

항목 X가 포함되었을 때, 동시에 Y항목이 포함되는 경우가 임의로 선택되는 Y의 확률에 비해 얼마나 높아지는가(lift)를 나타내는 측정지표이다. 즉, Lift는 선행 항목이 선택되었을 때 임의로 선택되는 후행항목의 확률보다 선행 항목과 후행항목이 동시에

선택될 확률이 몇 배로 나타나는지를 말해주는 지표이다. Lift는 확률로 나타나지 않으며 다음과 같이 나타낼 수 있다. 이때 P(Y)는 전체항목 중에서 Y라는 항목이 포함될 확률이다.

$$Lift = \frac{P(Y \cap X)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$$

## IV. 결과

### 1. 대변 증상에 사용된 약물의 약성 분석

#### a) 四氣 빈도

대변은 溫(40%), 平(20%), 微溫(10%) 순으로 약성의 빈도가 높았다(Table 4)(Figure 5).

Table 4. 대변 四氣 빈도, 비율

사기	大寒	寒	微寒	凉	平	微温	温	熱	합계
빈도(회)	2	76	79	37	164	85	322	43	808
비율(%)	0.25	9.41	9.78	4.58	20.3	10.5	39.8	5.32	100

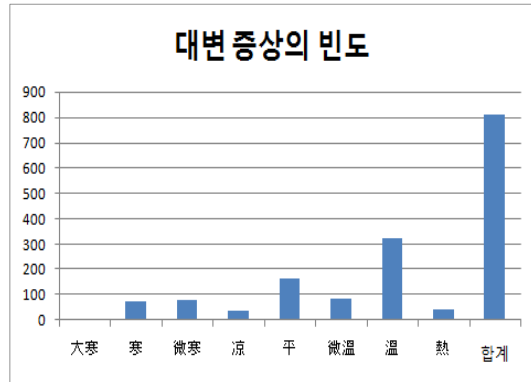
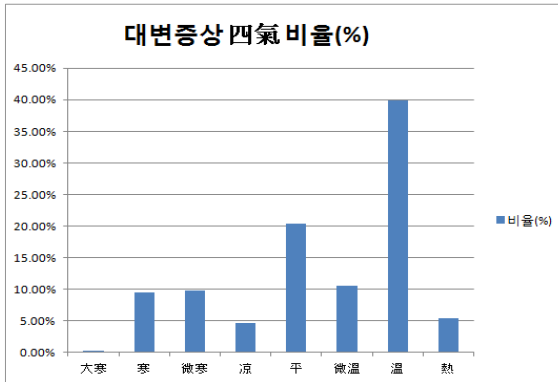


Figure 5. 대변 四氣 비율, 빈도

Table 5. 대변五味 빈도, 비율

오미	酸	苦	甘	辛	鹹	합계
빈도(회)	52	374	394	370	17	1,207
비율(%)	4.31	30.9	32.6	30.6	1.41	100

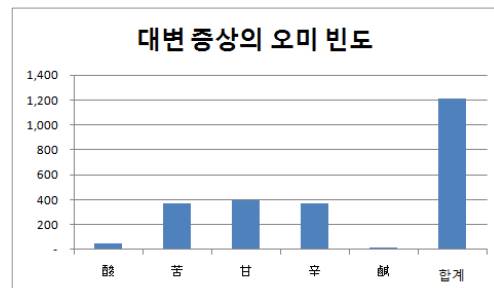
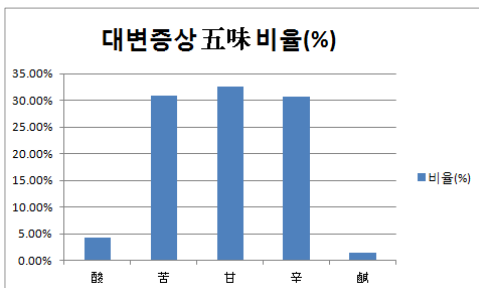


Figure 6. 대변五味 비율, 빈도

b) 五味 빈도

대변은 甘味(32%), 苦味(31%), 辛味(30%)가 90%를 차지하였다(Table 5)(Figure 6).

c) 歸經 빈도

대변은 脾(24%), 肺(16%), 胃(16%) 순으로 높은

빈도를 나타내었다(Table 6)(Figure 7).

d) 毒性 빈도

대변은 有毒(79%)에 해당하는 약물의 비율이 높았다(Table 7)(Figure 8).

Table 6. 대변 歸經 빈도, 비율

歸經	肝	心	脾	肺	腎	心包	膽	胃	小腸	大腸	膀胱	三焦	합계
빈도(회)	239	250	551	375	132	30	50	359	3	143	96	7	2,235
비율(%)	10.6	11.1	24.6	16.7	5.91	1.34	2.24	16.0	0.13	6.40	4.30	0.31	100

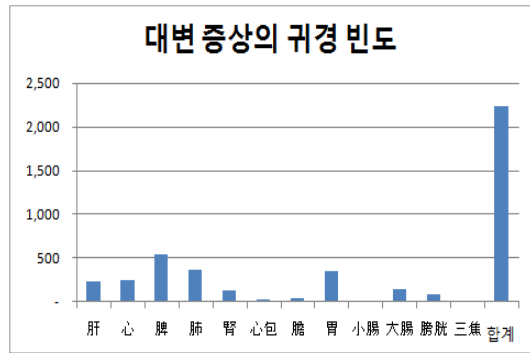
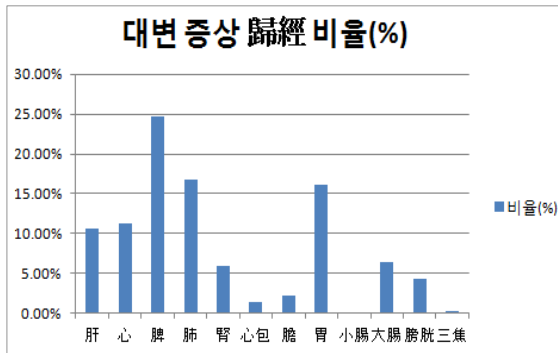


Figure 7. 대변 歸經 비율, 빈도

Table 7. 대변 毒性 빈도, 비율

毒性	有大毒	有毒	有小毒	합계
빈도(회)	4	26	3	33
비율(%)	12.1	78.7	9.1	100

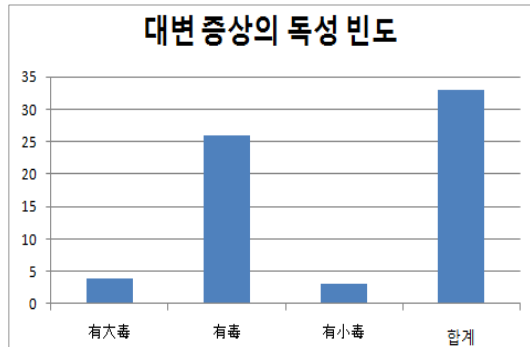
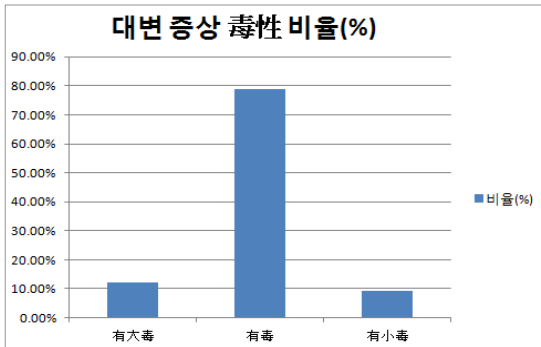


Figure 8. 대변 毒性 빈도, 비율



e) 效能 빈도

고 理氣(10%), 解表(9%)하는 약물이 뒤를 이었다 (Table 8)(Figure 9).

대변은 補虛(34%)하는 약물의 비중이 가장 높았

Table 8. 대변 효능 빈도, 비율

효능	解表	清熱	瀉下	祛風濕	芳香化濕	利水滲濕	溫裏	理氣	消息	驅蟲
빈도(회)	74	49	25	7	44	72	43	85	13	6
비율(%)	9.33	6.18	3.15	0.88	5.55	9.08	5.42	10.7	1.64	0.76

효능	止血	活血祛瘀	化痰止咳平喘	安神	平肝	開竅	補虛	收澀	涌吐	外用	합계
빈도(회)	3	27	33	3	1	4	273	30	1	0	793
비율(%)	0.38	3.40	4.16	0.38	0.13	0.50	34.4	3.78	0.13	0.00	100.

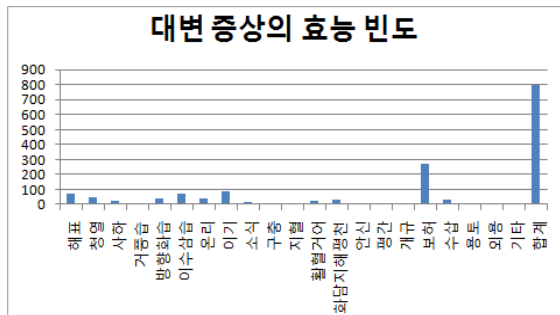
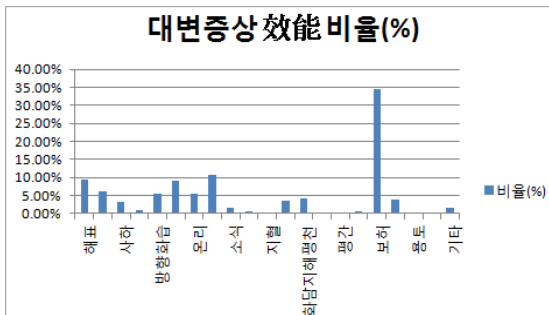


Figure 9. 대변 效能 비율, 빈도

\*기타 : 지혈,安神, 평간, 개규, 용토, 외용

Table 9. 대변 고빈도 약물 빈도, 비율

약물	감초	백출	인삼	당귀	진피	목향
빈도(건)	73	48	38	30	28	23
비율(%)	9.10	5.98	4.73	3.74	3.49	2.8

약물	백작약	적복령	복령	천궁	황련	후박
빈도(건)	23	22	18	18	18	18
비율(%)	2.86	2.74	2.24	2.24	2.24	2.24

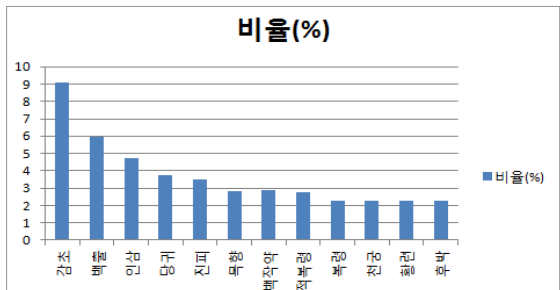
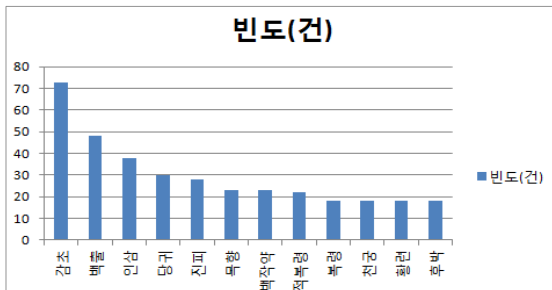


Figure 10. 대변 고빈도 약물 빈도, 비율

## 2. 대변 증상에 사용된 주요 약물

### a) 고빈도 약물

대변 증상에서의 고빈도 약물은 甘草(73건), 白朮(48건), 人蔘(38건), 當歸(30건), 陳皮(28건) 순으로 나타났다(Table 9)(Figure 10).

### b) 고특이도 약물

대변증상에서의 고특이도 약물은 肉荳蔻(5.36배), 大黃(2.6배), 木香(2.3배)등의 순으로 나타났다(Table 10)(Figure 11).

## V. 고찰

최근 들어 불규칙한 식사나 인스턴트, 패스트푸

드, 육식 위주의 식단, 운동 부족 등으로 인한 변비나 설사의 비중이 높아지고 있다. 또 병원에서 각종 검사를 받아도 특별한 기질적인 원인은 발견되지 않으나 반복적인 설사, 시원하지 않은 대변, 복부의 팽만, 배변 양상의 변화와 동반된 복통이나 복부 불편감을 특징으로 하는 과민성 대장증후군을 호소하는 환자도 증가하고 있다. 과민성 대장 증후군은 아직까지 명확한 원인이 밝혀지지 않았지만 정신적 스트레스가 주요 원인으로 알려져 있으며, 항진된 장운동기능, 내장의 과잉감각, 통증에 대한 예민도 증가, 감각기능의 이상, 중추신경계와 장관신경계의 상호작용 이상 등이 원인으로 추정된다.

임상적 중요성이 높아지고 있는 대변 증상에 대한 분석을 위해 연구 대상으로 방약합편을 채택하였다. 방약합편은 우리나라의 대표적 종합의전인 동의보감을 모태로 한 簡要醫書로서 구한말과 일제 통치 시에도 민중 속에 뿌리 깊게 이를 전승시켜 오늘

Table 10. 대변 고특이도 약물

Consequent	Antecedent	Support %	Confidence %	Lift	st-con	st-lift
육두구	대변	15.547	12.346	5.36	-0.60199	3.21229
대황	대변	15.547	11.111	2.631	-0.71455	0.985861
목향	대변	15.547	23.457	2.306	0.410728	0.720713
황련	대변	15.547	13.58	1.685	-0.48951	0.214077
택사	대변	15.547	14.815	1.513	-0.37695	0.073752
건강	대변	15.547	12.346	1.429	-0.60199	0.005221

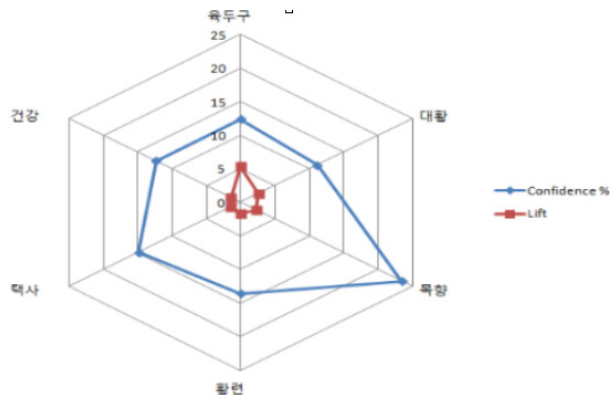


Figure 11. 대변 고특이도 약물

의 中興期를 맞이할 수 있게 한 알기 쉬운 한방치료 의학의 길잡이가 되어왔다. 방약합편은 그 내용의 簡要性和 實用性 때문에 지금도 한방진료에 관여하는 많은 인사들에게 빈도 높은 참고서로 애용되고 있다. 방약합편은 편찬 이래 지금까지 의사의 필수서가 되고 있으며 한방을 연구하는 사람이나 실무에 종사하는 사람이라면 누구나 반드시 갖추고 있어야 할 필독서로 인정되고 있다.<sup>11)</sup> 본 연구는 대변 증상을 전체 방약합편의 처방과 약물의 데이터베이스를 기반으로 하여 분석하였다.

한의학 지식의 정보화·현대화를 해결하기 위한 방법으로 Knowledge Discovery in Database(KDD)를 활용할 수 있다. KDD의 목적은 "대규모 데이터에서 유효하고(valid), 새롭고(noble), 잠재적으로 유용(potentially useful)하며, 궁극적으로 이해될 수 있는 패턴을 알아내는 것"으로 데이터 입력·사전처리, 그리고 변환과정을 거쳐 데이터베이스를 구축하고, 데이터 마이닝을 수행하여 결과를 해석·시각화하여 사용하고 관리하는 모든 과정을 말한다.<sup>10)</sup>

데이터베이스 구축과 자료의 변환은 엑셀과 데이터 마이닝 기법을 사용하였다. 데이터 마이닝은 대량의 데이터로부터 이전에 알려지지 않았지만 묵시적이고 잠재적으로 유용한 정보를 추출하는 작업이라고 정의할 수 있다. 현재와 같은 방대한 데이터 중에서 필요한 정보를 얻기란 쉬운 일이 아니므로 컴퓨터로 하여금 방대한 데이터 속에 숨겨진 정보를 발견하는 방법이 요구되었다.<sup>12)</sup> 특히 데이터베이스에 숨겨져 있는 패턴을 발견하는 연구 중에서 데이터 내에 존재하는 항목간의 종속 관계를 찾아내는 연관성 분석이 가장 많이 사용되고 있다. 연관성 분석은 하나의 거래나 사건에 포함되어 있는 항목들의 경향을 파악해서 연관성을 발견하는 것인데<sup>13)-15)</sup> 본 연구에서는 Apriori 알고리즘을 통하여 대변 증상과 약물과의 연관성에 대해 분석하였다.

본 논문에서는 대변증상의 치료 약물의 약성, 고

빈도 약물, 고특이도 약물에 대해서 알아 보았다. 약성이란 한약의 성질과 약리작용을 통틀어서 이르는 말로 약물의 성질과 맛, 작용 방향, 보사작용 등을 포괄한다. 四氣는 寒熱溫冷의 4가지로 나뉘는데 방약합편에서는 이를 좀 더 세분화하여 大寒, 寒, 微寒, 凉, 平, 微溫, 溫, 熱 8가지로 구분하였다. 五味는 약의 맛으로 酸, 苦, 甘, 辛, 鹹의 다섯 가지 맛으로 구분되어 있다. 약물의 效能은 解表, 清熱, 瀉下, 祛風濕, 芳香化濕, 利水滲濕, 溫裏, 理氣, 消息, 驅蟲, 止血, 活血祛瘀, 化痰止咳平喘, 安神, 平肝, 開竅, 補虛, 收澀, 涌吐, 外用(외용약재)의 20가지로 구분하였다. 有毒, 無毒은 약물을 복용하였을 때 인체에 대한 毒性을 지니고 있는지 아닌지를 나타내는 의미이다.

대변증상에 쓰이는 약물을 방약합편을 통해 분석해 본 결과, 四氣에서는 溫, 平, 微溫 순으로 빈도가 높고 五味에서는 甘味, 苦味, 辛味 순으로 빈도가 높았다. 歸經에서는 脾, 肺, 胃 순으로 빈도가 높았고 효능은 補虛, 理氣, 解表 순이었다. 毒性은 有毒에 해당하는 약물의 비율이 높았다.

고빈도 약물은 대변증상을 치료하는데 있어 방약합편에 나온 약물들 중 사용 빈도가 높은 약물을 말한다. 대변 증상에서 많이 쓰이는 약물로는 감초(73건), 백출(48건), 인삼(38건), 당귀(30건), 진피(28건) 순으로 나타났다.

고특이도 약물은 해당 증상에서 다른 증상에 비해 사용되는 빈도가 높은 약물을 말한다. 고특이도 약물의 분석 지표인 confidence는 선행 항목이 선택된 상황에서 선행 항목과 후행 항목이 동시에 선택될 확률을 말하는 것이고, lift는 선행 항목이 선택되었을 때 임의로 선택되는 후행항목의 확률보다 선행 항목과 후행항목이 동시에 선택될 확률이 몇 배로 나타나는지를 말해주는 지표이다. 대변 증상에서 lift값은 육두구가 5.36배, 대황이 2.6배, 목향이 2.3배 순으로 나타났다.

본 연구는 방약합편에서 대변 증상에 쓰인 약재들을 정리하여 약물의 약성, 고빈도 약물, 고특이도 약물에 대해 알아보았다. 이 과정에서 나타난 한계점으로, 첫째는 대변 증상을 설사와 변비로 나누지 않고 하나의 증상으로 인식했다는 점이다. 설사와 변비에 사용되는 약물은 분명 상반되는 경우도 있으므로 이를 나누어 다시 분석해 볼 필요성이 있다. 방약합편에서 설사에 해당되는 병증과 변비에 해당되는 병증을 구분한 뒤, 그에 해당되는 약재들을 다시 분류하여 본 논문의 연구 방법과 동일한 방법으로 분석한다면 그 결과를 얻을 수 있을 것이다. 둘째는 약물의 용량을 고려하지 않고 약물의 포함 여부에 의해서만 분석했다는 점이다. 병을 치료하는 처방은 사용 약물과 그 약물의 용량으로 구성되는데, 방약합편은 각 약물의 용량의 표현이 달라 그 기준을 정확히 정하기 어려웠기 때문에 사용 약물의 용량까지 분석하지 못하였다. 추후 연구를 통해 방약합편의 약물 용량에 대한 표준화가 이루어져야 할 것으로 사료된다.

## VI. 결론

본 연구에서는 데이터베이스로 구축된 방약합편의 처방, 병증, 약물을 빈도분석, 네트워크분석, 연관규칙분석을 통하여 대변 증상에 사용되는 약물의 약성, 고빈도 약물, 고특이도 약물을 분석하였다.

1. 방약합편의 대변 증상에 쓰인 약물의 약성을 四氣, 五味, 歸經, 效能, 毒性 5가지 기준으로 알아보았다. 대변 증상에 쓰인 약물의 四氣 빈도는 溫(40%), 平(20%), 微溫(10%) 순으로 나타났고, 五味 빈도는 甘味(32%), 苦味(31%), 辛味(30%) 순으로 나타났으며 歸經 빈도는 脾(24%), 肺(16%), 胃(16%) 순으로 나타났다. 또한 대변 증

상에 쓰인 약물의 毒性은 주로 有毒(79%)에 해당하는 약물이 많았고, 效能 빈도는 補虛(34%), 理氣(10%), 解表(9%) 순으로 나타났다.

2. 방약합편의 대변 증상에서 고빈도 약물은 甘草(73건), 白朮(48건), 人蔘(38건), 當歸(30건), 陳皮(28건) 순이었고, 고특이도 약물은 肉苁蓉(5.36배), 大黃(2.6배), 木香(2.3배) 순으로 나타났다.

## 參 考 文 獻

1. 전병권, 韓醫學分野 專門情報 DB 構築에 관한 研究, 한국과학기술정보연구원, 2003, 1-188.
2. 한중현, 윤용갑, 한성국 외 1명, 한의학 문헌정보의 전자도서관 구축, 대한한의학회지. 1999; 7: 153-165.
3. 이현실. 합성 온톨로지 기반의 한의학 처방 지식 관리 시스템. 한국학술정보(주). 2006, 1-186.
4. 문경실, 박수현. 맞춤형 진단 서비스를 위한 한의학 온톨로지. 한국컴퓨터정보학회 논문집, 2010; 15: 23-30.
5. 김희열, 傷寒論 온톨로지 구축 방법론 연구, 대구 한의대학교 대학원 한의학과 박사논문. 2011, 1-62.
6. 홍대기, 육순형, 김민용, 박영재, 오환섭, 남동현, 박영배. 네트워크 모델을 통한 傷寒論 구조분석 연구- 太陽病 症狀-處方을 중심으로- 대한한의학회지, 2010; 32: 56-66.
7. Yi Feng, Zhaohui Wu, Xuezhong Zhou, Zhongmei Zhou, Weiyu Fan. Knowledge discovery in Chinese traditional medicine: State of the art and perspectives. Artificial Intelligence in Medicine, 2006; 38: 219-236.

8. Wu ZH, Zhou XZ, Liu BY, Chen JL. Text mining for finding functional community of related genes using TCM knowledge. In: Boulicaut JF, Esposito F, Giannotti F, Pedreschi D, editors. Proceedings of the 8th European conference on principles and practice of knowledge discovery in databases. Berlin: Springer-Verlag. 2004, 459-470.
9. Jiang Youguang, Li Renshu. Experiment on Data Mining in Compatibility Law of Spleen-stomach Prescriptions in TCM. World Science and Technology Modernization of Traditional Chinese Medicine and Materia Medica. 2005; 5: 33-35.
10. 양동훈, 데이터마이닝을 이용한 方藥合編의 병증 및 약물간 연관성 분석. 경희대학교 대학원 한의학과 박사논문. 2011, 1-77.
11. 申裁鏞, 『方藥合編解說』, 成輔社. 1988, 3-5.
12. Kirt Thearling, Increasing customer value by integrating data mining with campaign management software, Exchange Applications Inc. Direct Marketing Magazine. 1999.
13. Srikant R. and Agrawal R. Mining Sequential Patterns: Generations and Performance Improvements, Proc. of EDBT, Avignon, France, Mar. 1996.
14. Park J.S., Chen M.S., and Yu P.S., An Effective Hash-based Algorithm for Mining Association Rules, In Proc. Of ACM SIGMOD, 1995, 175-186.
15. Pei J., Han J., Mao R., CLOSET: An efficient algorithm for mining frequent closed itemsets, In Proc, ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery, 2000, 11-20.

