

고전 임상의학의 통계적 요인 추출방안에 관한 연구

박길하, 박찬석¹, 박래수², 이정화², 안상우²

공주대학교 의료정보관리학과, ¹충청대학 보건의료정보과, ²한국한의학연구원

A Study of the Method of Statistical Factors Analysis from Old Oriental Medical Records in Korea

Park GilHa, Park ChanSeok¹, Park LaeSu², Lee JeongHwa², Ahn SangWoo²

Kongju National University, ¹Chung Cheong University, ²Korea Institute of Oriental Medicine

This Study is to propose the method about statistical factor from the old oriental medical record in Korea. This Study reviews the statistical analysis recently published in the old oriental medical books which are in the Research Report by the Korea Institute of Oriental Medicine at 2007. The results reveal a disease factors, prescription factors, medicines factors, population factors and historical factors by the statistics. The results show that the old oriental medical record for a information system need the interpreting with a information analysts, a statistics analysts and an oriental medical doctors.

Keyword : Statistical Factor, Oriental Medical Record, Information System

I. 서론

의료정보학의 기원은 1950년대 프라하의 Antonin Svoboda 교수였는데, 주로 수학과 정보과학 응용분야와 약에 대한 정보를 접목시키면서 태동하였다. Jaroslav Hajek와 Zbynek Sidak는 1967년에 출판된 “heory of rank tests”에서 의학적인 문제를 해결하기 위해 통계적 순위에 기반을 두었으며 이것이 유럽에서 의료정보의 시초가 되었다(Haux, 2006). 현대 의료정보학은 1960년대 미국과 유럽의 몇몇 국가에서 병원에서 발생하는 정보를 효율적으로 활용하는 방안을 연구하던 중 등장하였으며 의료보험제도의 시행으로 정보화가 급격히 진행되었다(김창엽 등, 1999).

한방에서 의료정보는 한의학에서 환자의 진료, 의학교

육, 의학 연구 및 의료경영에 필요한 각종 정보를 효율적으로 체계화하여 관리하는 것으로 정보과학의 범주에서 한의학 연구를 진행하는 학문이다(신현규, 2004). 이에 따라 한의학 분야에서도 정보화가 꾸준히 진행되었고 모든 한방 병원에서 어떤 형태로든 정보시스템을 사용하고 있다(채영문, 1994).

최근 의료인들의 정보결핍에 대한 실증연구에 의하면 대부분의 의료인들에게서 정보결핍이 나타났으며 환자를 직접 진료하는 의사들에게서 정보 결핍이 매우 심하게 나타났다(박찬석, 고석하, 2003). 특히 직위가 높은 의사 집단에서 의사결정시 정보 결핍 정도가 매우 광범위하게 나타났다. 이것은 국내 의료계에서 도입되어 운영중인 의료정보시스템이 단편적 계산이나 반복적 업무 처리를 위해 구축되어 있어 장기적인 전략이나 임상연구를 위한 분석적 정보처리기능과 의학적 컨텐츠 활용에는 한계를 보인 객관적 증거이다(박찬석 등, 2004).

위와 같은 의료 정보화의 현실에서 청강의안의 현대적 해석은 고전의안과 현대의안을 체계적으로 연구하고 임상

✉ 접수 ▶ 2007년 11월 7일 수정 ▶ 2007년 12월 3일 채택 ▶ 2007년 12월 18일
✉ 교신처 ▶ 안상우, 대전광역시 유성구 엑스포로 483 한국한의학연구원
Tel 042-868-9442 Fax 042-863-9463 E-mail answer@kiom.re.kr

적으로 비교분석 할 수 있는 매우 중요한 연구(차용석, 2007)로 평가할 수 있다. 그러나 해지된 청강의안의 정보화는 온톨로지 기반으로 실시되어 통계학적으로 많은 한계를 보여주고 있고 잠재되어 있는 중요한 요인들의 추출이 필요하다.

이에 따라 본 연구에서는 해지된 청강의안을 바탕으로 잠재되어 있는 의학적 요인들의 추출과 통계학적 요소들의 활용에 대해 제안함으로써 귀중한 의학적 자산의 활용을 극대화하고자 한다.

II. 연구모형

본 연구에서는 해지된 청강의안의 처방내용들과 고전서의 온톨로지 방식의 정보화에 대한 내용들을 중심으로 문제점과 추가적 개발방안 및 통계학적 요소 추출에 대해 다음과 같은 주제로 연구를 실시하였다.

1. 해지된 청강의안을 이용하여 추출할 수 있는 의료 콘텐츠 추출방법은 무엇인가?
2. 청강의안의 정보화 추진방안은 무엇인가?
3. 데이터웨어하우스시스템을 이용한 콘텐츠 활용방안은 무엇인가?
4. 데이터마이닝을 이용한 콘텐츠 활용방안은 무엇인가?

III. 연구 방법

1. 소프트웨어 공학 관점에서의 접근

소프트웨어는 생성, 성장, 쇠퇴와 같은 단계를 거치면서 유기체적 특성을 지닌다. 특히 1960년대 NATO에서는 “소프트웨어 위기” 라는 용어를 만들어 내어 소프트웨어 개발의 낙후성을 지적하고 있다(고석하, 2003). 이것은 많은 비용과 시간을 투자한 소프트웨어가 최종 사용자들의 정보 욕구를 충족시키지 못하거나 운영비용의 급증으로 재개발을 해야 하는 상황을 지칭하는 표현이다. 이와 같은 위기의 주된 원인은 표준화되지 못한 코드, 최종 사용자들의 정보욕구 파악의 한계, 산업적 특성을 이해하지 못한 시스템 설계와 정보화에 대한 막연한 기대감으로 대표된다. 또한 소프트웨어 개발에 있어서 코드와 사용자 인터페이스 구현에 가장 많은 비용과 노력이 소요된다고 주장하

고 있다(Lozano et al, 2000). 이것은 사용자들의 요구사항에 대한 정확한 이해, 사용자와 의사소통 문제와 어플리케이션에 대한 다양한 테스트 미흡으로 귀결되고 있다(김정옥 등, 2003).

이렇듯 소프트웨어는 사용자들의 신뢰를 받지 못하면 경제적 시간적으로 매우 난감한 상황에 처한다는 것으로 국내 의료분야 정보화도 이와 유사한 형태를 보이고 있다(박찬석 등, 2004).

청강의안의 해지정보는 매우 다양한 측면의 자료들을 내포하고 있으나 정보시스템 설계에는 다음과 같은 한계점을 내포하고 있다.

첫째, 코드 분류의 체계화에 한계점이 있다. 특히 병명이나 처방에 대한 코드화가 미흡하여 약제와 질병에 대한 효율적이고 연계성에 한계가 있다.

둘째, 미 추출된 정보 요인으로 고전서 활용의 한계점이 대두된다. 해지된 청강의안 내용은 사회 환경적, 인구학적, 자연생물학적 요인과 의학적 내용들이 매우 다양하게 내제되어 있으나 고전한의학 정보화에서 전문가 집단의 참여가 미흡하여 충분한 콘텐츠 설계가 제외된 상태이다.

마지막으로 단순 소프트웨어 개발자에 의한 근시안적 정보시스템 설계로 정보 활용의 한계가 나타났다. 즉 한의학, 정보공학과 통계학 전문가들이 참여한 시스템 설계와 의료 콘텐츠 개발로 다양한 요인들의 연계에 의해 특이치방이나 역사적 사건과 연계된 폭 넓은 분석이 필요하다.

2. 통계학적 관점에서의 접근

데이터 마이닝은 KDD(Knowledge Discovery in Database)라고 불리우듯 대용량의 자료, 혹은 데이터 웨어 하우스로부터 쉽게 드러나지 않는 유용한 정보들을 찾아내는 과정을 말한다(최종후 등, 2005).

즉, 대용량(massive)의 관측 가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 바탕으로 의사결정 등을 위한 정보로 활용하는 것이다. 실제 데이터 마이닝이 적용되는 과정은 탐색(Exploration)을 통해 평균, 이상치, 결측치 등을 발견하고 변형(Modification)으로 자료를 변환하며 모형화(Modeling)와 모델평가(Assessment)의 단계를 거치게 된다.(SAS Eminer 중심)

데이터 마이닝 기법으로는 군집분석(Cluster Analysis), 연결분석(Link Analysis), 판별분석(Discrimination Analysis) 등과 같은 기존의 통계 분석과 연관성규칙(Association

Rule), 의사결정나무(Decision Tree), 신경망모형(Neural Network), OLAP(On-Line Analytic Processing)등 변형된 형태의 분석 기법이 있다(강현철, 한상태, 최종후 외, 2001).

의학분야에서는 데이터마이닝을 활용한 질병, 처방, 상병 등 여러 변수들간의 상관분석을 통한 분류 및 세분화를 통해 질병별 패턴분석, 처방 패턴분석, 상병 패턴분석 등 여러 각도로 분석이 시도되고 있으며 다각도로 분석되고 있다.

그러나 OLTP(On-Line Transaction Process), 데이터웨어하우징(DataWarehousing) 등과 같은 데이터 저장 시스템 출현과 대용량, 초고속 컴퓨터의 보편화로 인하여 대용량 자료가 매순간 자동적으로 쌓이고 있고, 이는 비단 정보에 대한 자료를 관리하는 병원 및 기업뿐 아니라 연구 및 교육 기관들도 목적은 상이하지만 다양한 형태의 자료들이 수없이 시스템에 저장되고 있다. 이러한 자료들은 크기가 대용량일 뿐 아니라 자료의 형태가 다양하여 기존의 통계분석 도구로는(기법이나 통계 소프트웨어 모두 포함) 원하는 정보를 얻는데 한계가 있다.

현대 의료 산업에서는 정보 활용을 극대화하고자 다양한 형태의 정보시스템과 분석 기법들이 사용되고 있다. 그러나 과거 의료정보에 대한 현대적 해석과 활용은 매우 미흡한 수준이며 한의학 분야도 일부 고전의안에 대한 정보화는 온톨로지 중심의 정보화가 이루어져 있어 활용성과 파급효과가 미흡한 수준이다.

이에 따라 본 연구는 청강의안 해지본에 대해 소프트웨어 공학과 통계학 관점에서 접근하여 다양한 정보 활용방안과 한계점을 고찰하는데 목적이 있다.

IV. 연구결과

1. 청강의안 해지본의 컨텐츠 요인 추출 방법

고전의서의 해지는 한의학 분야에서 진료수준의 향상과 역사적 사실을 고찰하는 동시에 사회 문화적 측면과 경제 산업적 효율성을 증대시키는 매우 중요한 우리 고유의 자산이다(차웅석, 2007). 이와 같은 귀중한 자산에 대해 정보화는 온톨로지 중심으로 주요 처방전의 검색에 중점을 두고 있다. 이것은 고전의안의 처방 내역을 이미지로 검색하여 참고하는 시스템으로 통계학적으로 다음과 같은 한계를 내포하고 있다.

첫째, 통계적 요인의 추출이 불가능하여 다양한 의료 통계 추출에 한계점이 있다.

둘째, 해지된 내용에는 다양한 통계적 요인들이 잠재에 있지만 실질적으로 정보화에는 누락이 되어 활용에 한계가 있다.

셋째, 분석적 통계가 불가능하여 학술적으로 참조의 수준으로 활용되고 있다.

마지막으로 일반 국민들에게 서비스할 수 있는 컨텐츠가 거의 존재하지 않는다.

위와 같은 한계 상황에서 본 연구는 해지된 청강의안의 통계적 요인 추출 방법에 대해 <표 1>과 같은 방법의 제안을 통해 다양한 컨텐츠 추출을 제안하려 한다.

<표 1> 청강의안 해지본의 정모공학적 컨텐츠 요인 추출 예시

선조 8권 7년 8월 16일 (정사) 003[환경적 요인]
 약방이 양위진식탕을 쓰도록 권하니 따르다.
 약방이 아뢰기를, “양위진식탕[처방 요인]을 지어다가 드신다면 수라가 잘 내려가게 되실 것입니다.” 하였는데, 입제하니 상이 그대로 따랐다. 약방이 아뢰기를, “육부가 모두 미열이 있고 심신 위 3장이 조금 허합니다[병증 요인]. 이 허열 때문에 날 것과 찬 것을 즐겨[식습관 요인] 드시므로 중기가 점차로 손상되어 식욕이 없는 것입니다. 원지 백복령 신국 맥문동 지각[혼용 약재 요인]을 가미한 양위진식탕은 한편으로는 허한 면을 보하고 한편으로는 전신을 소통시키는 것이니[효과 요인], 다려서 드시기도 하고 환복용 요인을 지어 드시기도 하되, 정도[용량 요인]에 맞추어 드시는 것이 마땅할 줄 압니다.
 대체로 병을 조리하는 요법은 오로지 약을 복용하는 데만 있는 것이 아닙니다. 되도록 마음을 맑게 하고 욕구를 억제하며 음식을 적절하게 들되 일정한 시간[지시 요소]을 놓치지 않으면 위장의 기운이 점차로 회복되고 심장의 열이 자연히 내려가게 되어 위장이 제대로 움직이고 비장이 제대로 소화시켜 수라가 잘 내려가게 될 것[효능/효과요소]입니다. 하니, 아뢴 대로 하라고 답하였다.

청강의안 해지본을 기준으로 추가적으로 분석된 의학적 요인들은 다음과 같다.

첫째, 인구 사회학적 요인 : 성별, 나이, 체질, 몸무게, 키, 직업 등

둘째, 발병증상 : 심신 위 3장이 허함, 육부에 미열, 소화가 잘 안됨 등

셋째, 식습관요인 : 날것과 찬 것을 즐겨함 등

넷째, 처방요인 : 양위진식탕 + 원지, 백복령, 신국, 맥문동, 지각을 혼용

다섯째, 복용법 : 탕, 환을 혼용

여섯째, 지시사항 : 마음을 맑게 하고 욕구를 억제, 음식을 적절하게 일정한 시간에 섭취

일곱째, 처방효능 : 위장의 기운이 점차로 회복되고 심장의 열이 자연히 내려가게 되어 위장이 제대로 움직이고 비장이 제대로 소화시켜 수라가 잘 내려가게 될 것

위와 같이 청강의안 해지본에 대해 한의학적 관점에서 콘텐츠 중심으로 재해석이 필요하며 해석된 자료에 대해 정보화를 위한 설계가 반드시 필요하다.

2. 고전의안 정보화 추진방안

<표 1>과 같이 재해석된 정보를 기준으로 데이터베이스를 구축하면 <그림 1>과 같은 전통 한의학 정보시스템 모형을 만들 수 있다. 환자를 진료하면 병명, 처방, 치료기술, 약재, 치료결과 등을 얻을 수 있다. 이와 같은 임상적으로 귀중한 정보들이 표준화 및 코드화의 미흡으로 활용성이

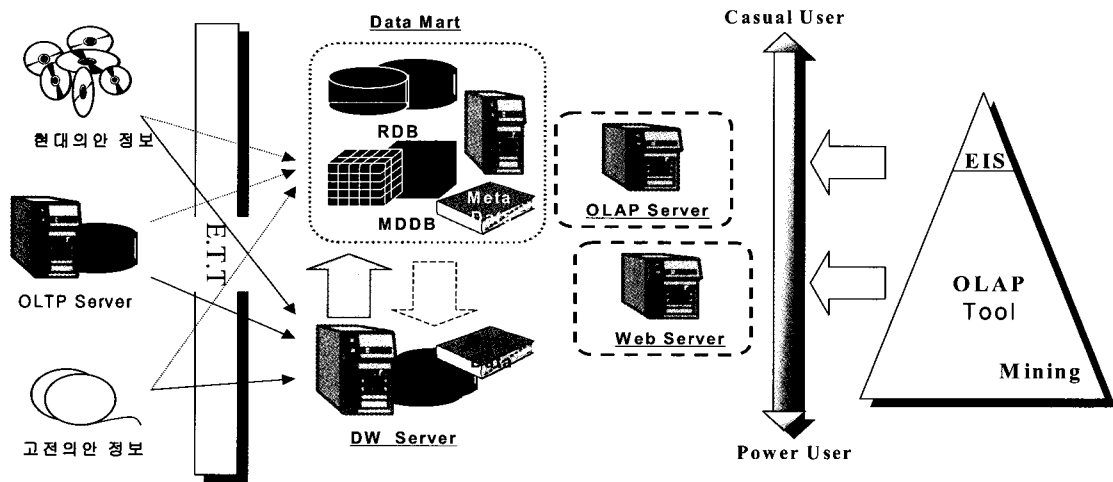
떨어진다면 매우 큰 자산을 잃는 경우가 된다. <그림 1>은 환자 진료시 임상적으로 활용할 수 있는 정보 요소들을 포함하였고, 처방 및 진료기록에 대한 고전 정보와 현대 진료정보를 비교 분석 할 수 있는 데이터베이스 모형을 예시한 것이다.

<그림 1>과 같이 설계된 데이터베이스에 정보들이 집적 되면 분석이 가능한 데이터웨어하우스 시스템을 구축해야 한다. 데이터웨어하우스의 특징은 주제 중심적 데이터 분석이 가능하며 요약된 정보 구조와 발견적 접근(Heuristic Approach) 방법을 통해 임상연구 및 환자진료에 필요한 의사결정 자료를 지원하게 된다.

<그림 2>는 데이터웨어하우스시스템 개념도로써 진료 기록에 대해 누적된 정보를 이용하여 한의학 주제 영역별 데이터 마트를 구축하고 이것을 이용하여 의사결정을 지



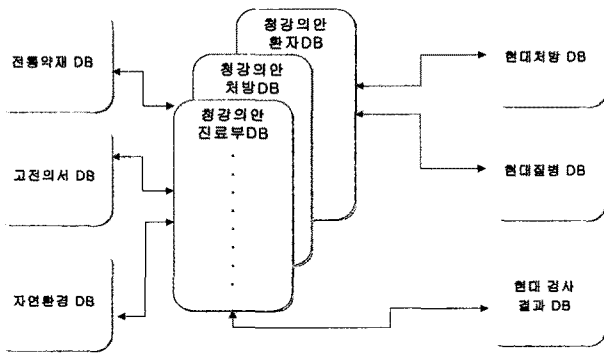
<그림 1> 데이터베이스 설계 모형 예시



<그림 2> 데이터웨어하우스시스템 개념도

원하는 체계로 구성되어 있다. 특히 고전의안 정보를 데이터 마트로 분류하는 과정에서 한의학자와 정보 분석가들의 역할이 중요하다.

<그림 2>와 같은 데이터웨어하우스시스템의 논리적 데이터베이스를 도식화하면 <그림 3>과 같이 요약할 수 있다. 재해석되어 구축된 청강의안 진료, 처방, 환자 데이터베이스는 전통약재, 고전의서와 현대 처방 및 질병 데이터베이스와 연동이 되는 시초가 될 것이며 궁극적으로 고전 한의학과 현대한의학을 비교 분석할 수 있는 초석이 될 것이다.



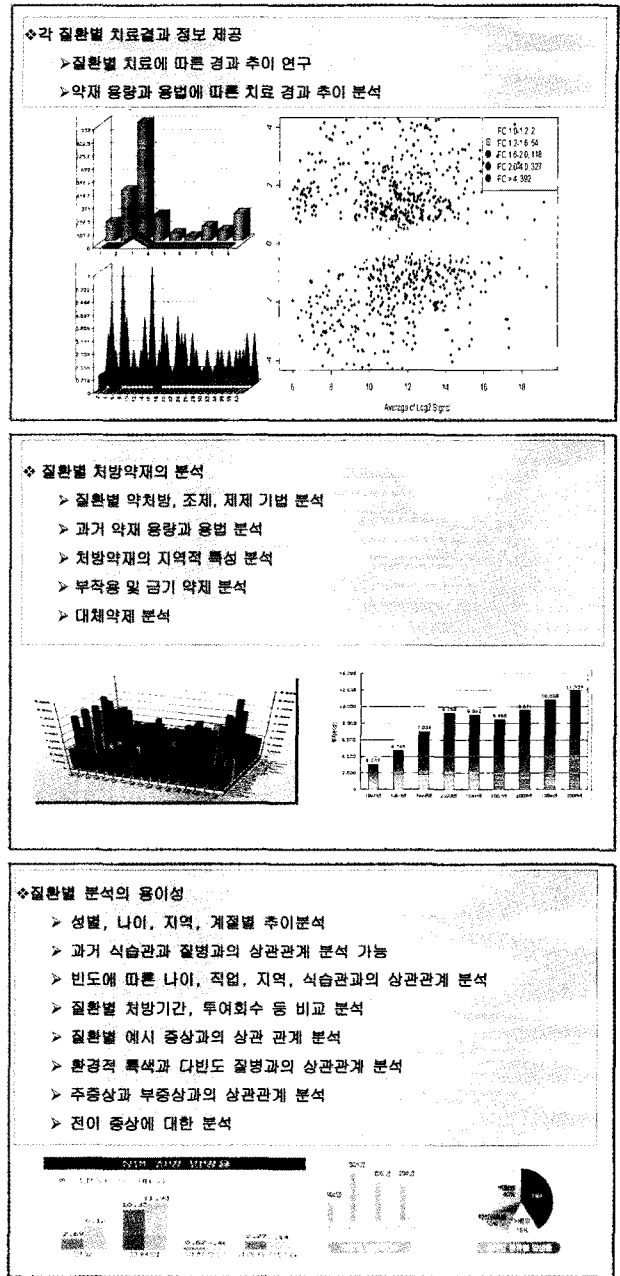
<그림 3> 발전된 한의학 데이터웨어하우스 설계 구성도 예시

3. 고전의안 통계적 활용 방안

이상과 같이 구축된 고전의서 정보시스템은 다양한 형태의 통계적 다각화 분석이 가능할 것이다(<그림 4> 참조). 특히 운동로지 분석에서는 불가능한 진료기록에 대한 컨테츠 분석과 주제 영역 분석이 가능해져 청강의안의 특이 처방이나 특이 치료기술을 연구하는데 매우 유용한 형태의 정보를 추출할 수 있다. 또한 사회 인구학적 요인, 자연환경적 요인, 역사학적 요인과 같이 진료에 직접적으로 연관이 적은 분야까지 분석이 가능해져 과거 우리사회의 사회상을 고찰하는데 귀중한 정보를 추출할 수 있을 것이다.

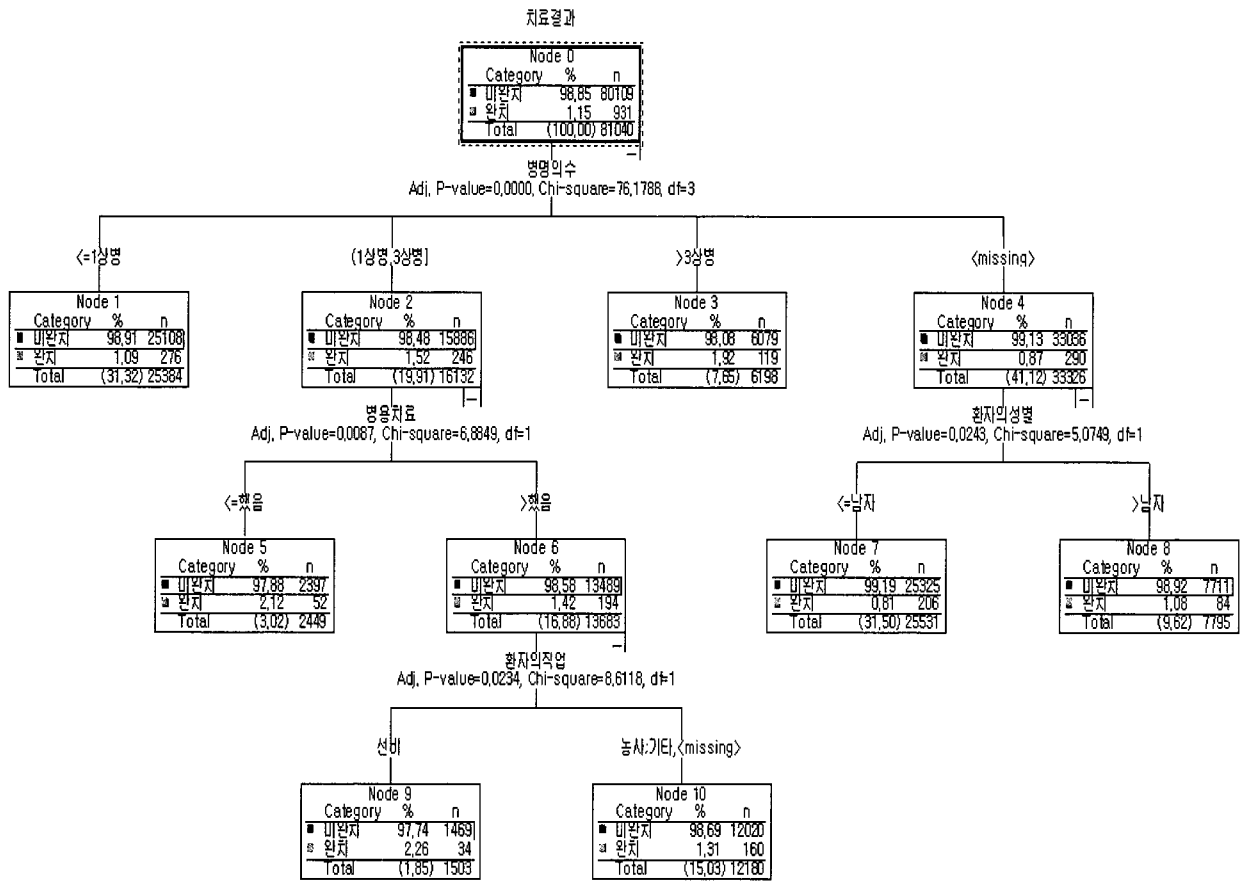
다음으로 데이터마이닝 기법을 통한 고전의서 활용방안에 대해 알아보자.

데이터마이닝은 III장에서 언급한 것과 같이 분석 툴(Tool)을 이용하여 정상적인 정보 유형과 비정상적인 정보 유형을 추출하여 새로운 가치로 해석하는 것이라 정의할 수 있다. 의학 분야에서는 병명, 처방, 치료기술에 따라 치료결과에 대한 다양한 임상적 분류를 가능토록 한다.



<그림 4> 고전의안 데이터베이스를 이용한 통계적 예시

<그림 5>는 SPSS 예제파일로 시장세분화 사례분석 자료에 변수를 수정하여 데이터마이닝 분석 툴인 AnswerTree 3.0을 이용하여 분석해 본 예시 결과이다. 사용변수는 나이, 성별, 자녀수, 치료일수, 치료방법, 병명의수, 환자직업, 용법, 치료결과, 빈도로 설정하였고 목표변수(Target)는 치료결과로 하였다.



<그림 5> AnswerTree를 이용한 진료정보 데이터마이닝 분석 예

분석결과 치료결과에 대한 첫 번째 분리변수는 상병(병명)임을 알 수 있었다. 즉 상병(병명)이 많을수록 치료효과가 좋은 것으로 나타났고(치료율 = 1.92%), 병명에 대한 결측치(병명을 알지 못함)가 있는 경우 치료효과가 떨어지는 것을 볼 수 있었다. 또한 상병수가 2개와 3개인 경우(1개, 3)가 함께 그룹화 되어 있다. 이것은 범주 간 통계 적인 차이가 없는 경우 자동적으로 범주를 병합시키기 때문이다. 그리고 병명이 2개 또는 3개에 해당하는 경우에는 병용치료 즉 탕과 침을 혼용해서 치료하였는가로 병용을 한 범주가 안한 범주보다 완치율(1.67%)이 좋은 것으로 볼 수 있다. 마찬가지로 병명의 수가 결측치인 경우에 있어서는 성별에 의하여 분리가 이루어졌는데, 환자의 성별이 여성인 경우 남성보다 완치율(1.08%)이 더 높게 나타난 것을 알 수 있다. 병명의 수가 2 또는 3개이면서 병용치료를 하지 안한 경우 환자의 직업에 의해 추가적으로 분리가 이루어졌다. 이때의 분리결과는 환자의 직업이 선비인 경우 완치율(2.26%)이 높게 나타났다.

V. 결론 및 기대효과

본 연구는 해지된 청강의안에 대해 의학 콘텐츠 중심으로 정보시스템의 재설계 방안과 활용 방안에 대해 실질적 예시를 통해 제안하고 있다. 즉 해지된 진료기록에 대한 한의학자, 정보 분석가, 통계학자 등 다양한 전문가 집단이 참여하여 주제 중심적이고 통계적인 요소들을 추출하는 작업이 필요하다고 요약할 수 있다. 이러한 다양한 전문가 집단은 의료산업의 특성과 의학 용어의 이해 폭이 넓고 경험이 있는 전문가들로 구성해야 이해의 오류 및 장벽을 최소화 할 수 있다.

이상에서 분석한 것과 같이 청강의안의 통계적 활용을 위한 정보시스템 설계 방안을 다음과 같이 제안할 수 있다.

첫째 다양한 전문가 집단이 포함된 콘텐츠 중심의 정보 시스템을 설계가 필요하다.

둘째 의료정보화에 경험이 있는 시스템 구축 전문가의 참여가 필요하다.

셋째 한의학자에 의한 해지본의 추가적 요인 추출이 필요하다.

넷째 현대한의학과 연계 된 표준화 및 코드화가 필요하다.

마지막으로 폭넓은 고전의서의 현대적 해석 및 정보 통합화 절실히 필요하다.

위와 같이 제안된 고전의서 정보화가 구축될 경우 청강의안의 체계적 정보화로 고유처방, 치료기술, 학문적 전통성을 확보할 수 있는 기초자료 제공이 가능하며, 온톨로지 방식의 정보처리의 한계를 극복하고 한의학 분야 관계자들의 마인드 재고와 교육 및 연구의 랜드마크 수립에 중요한 자원이 될 것이다. 또한 의학 분야에서 정보화의 선두화와 현대의학과 고전 의학을 연결하는 매개체로 급부상할 것이다.

VI. 참고문헌

1. 강현철, 한상태, 최종후 외, 『SAS Enterprise Miner 4.0을 이용한 데이터마이닝- 방법론 및 활용』, 자유아카데미, 2001.
2. 고석하, 『비즈니스시스템 분석 및 설계』, 대영사, 2003.
3. 김정옥, 유철중, 장옥배, 「사용성 중심설계를 지원하기 위한 사용자 인터페이스 프로토타입의 생성기법」, 『정보과학회논문지』, 2003;31(1):30-46.
4. 김창엽, 강길원, 이진석, 김병익, 김용익, 신영수, 「국내 병원정보시스템의 도입 현황」, 『대한의료정보학회지』, 1999;5(1):27-34.
5. 박찬석, 경원현, 고석하, 「종합병원정보시스템 사용자들의 정보결핍에 관한 연구」, 『대한의료정보학회지』, 2004;10(3):223-233.
6. 박찬석, 고석하, 「종합병원 구성원의 정보욕구에 의한 분류」, 『Journal of Information Technology Applications & Management』, 2003;10(2):44-57.
7. 신현규, 박경모, 『한의처방 Informatics 구축사업 연구』, 한국한의학연구원, 2004.
8. 차웅석, 『전통의학지식정보 기반구축연구』, 경희대학교, 2007.
9. 최종후, 소선하, 『사례로 배우는 데이터마이닝』, 자유아카데미, 2005.
10. 채영문, 『보건정보관리』, 수문사, 1994.
11. Haux, H, 「Health Information Systems - Past, Present, Future」, 『International Journal of Medical Informatics』, 2006;75(2):268-281.
12. Lozano, M. D., P. Gonzalez and I. Ramos, 「User Interface Specification and Modeling in an Object Oriented Environment for Automatic Software Development」, 『Technology of Object-Oriented Languages and Systems Proceedings 34th International Conference』, 2000: 373-381.