

한의 정보의 계산적 방법 조사

김상균 · 장현철 · 김진현 · 김 철 · 예상준 · 송미영*

한국한의학연구원 정보연구센터

Computational Methods for Traditional Korean Medicine : A survey

Sang Kyun Kim, Hyun Chul Jang, Jin Hyun Kim, Chul Kim, Sang Jun Yea, Mi Young Song*

Information Research Center, Korea Institute of Oriental Medicine

Traditional Korean Medicine (TKM) has been actively researched through various approaches, including computational methods. This paper aims at providing an overview of domestic studies using the computational techniques in TKM field. A literature search was conducted in Korean publications using OASIS system, and major studies of data mining in TKM were identified. A review was presented in six diagnosis fields, including sasang constitution diagnosis, eight constitution diagnosis, tongue diagnosis, pattern diagnosis for stroke, diagnosis based on ontology, diagnosis for cause of disease. They collect clinical data themselves for experiments and primarily applied a algorithm of decision tree, SVM, neural network, case-based reasoning, ontology reasoning, discriminant analysis. In the future, there needs to identify which algorithm is suitable to diagnosis or other fields of TKM.

Key words : computational method, data mining, traditional korean medicine, survey

서 론

한의학의 과학화 및 세계화를 위해서 한의학 연구에 대한 객관화된 결과 제시는 필수적이다. 이를 위해서 많은 한의학 연구자들은 수행한 연구 결과의 우수성을 보여주기 위해서 다양한 실험이나 논리적인 방법들을 이용한다. 최근에는 IT 기술이 발전함에 따라 기존의 효율적인 데이터 분석이나 계산 알고리즘을 한의학에 적용하려는 시도가 많아지고 있다. 특히 한의 진단 분야는 진단 결과가 얼마나 정확한지 또는 기존의 방법보다 얼마나 더 나은 진단이 가능한지를 객관적인 수치를 이용해서 보여주는 것이 중요하기 때문에 이러한 방법을 많이 이용하고 있다.

본 연구에서는 기존에 제안된 한의학 연구들 중에서 계산적 방법(computational method)을 이용한 연구들을 조사하고 리뷰하고자 한다. 일반적 의미의 계산적 방법은 데이터 마이닝(data mining, 통계나 인공 지능 기술을 이용해서 많은 양의 이질적인 데이터들에서 패턴을 추출하는 과정¹⁾을 통해 새로운 지식을 발견하는 방법을 의미한다. 하지만 한의학 분야의 연구 논문들에서는 서양의학의 연구 논문과 같이 대용량의 데이터를 이용하지는

않았다. 이는 아직까지 연구에 이용할 수 있는 한의학 정보 데이터베이스나 임상 데이터들이 부족하기 때문이다. 대신 연구 내용에 대한 객관적인 결과를 보여주기 위해 여러 데이터 마이닝 알고리즘들을 이용하고 있다.

본 연구에서는 데이터 마이닝 방법을 사용하는 한의학 논문들을 조사함으로써, 한의학 분야에서 사용되는 데이터마이닝 알고리즘들을 분석하고, 문제점 및 향후 연구 방향에 대해서 제안하고자 한다.

연구방법

본 연구에서는 데이터 마이닝을 이용하는 한의학 연구들을 검색하기 위해서 한의학 분야 논문에 대한 검색 및 원문 서비스를 제공하는 오아시스²⁾ 시스템을 이용하였다. 검색 키워드는 계산적 방법과 관련이 있는 데이터마이닝(data mining), 의사결정(decision support), 알고리즘(algorithm), 지식(knowledge), 전문가(expert), 진단(diagnosis), 온톨로지(ontology), 통계(statistic), 확률(probability) 등의 한글과 영문 키워드를 사용하였으며 검색 결과에서 데이터 마이닝을 이용하는 논문만을 골라 총 22건의 논문을 리뷰하였다. 논문의 선별 과정에서 일부 연구들은 수집된 자료들을 분석하기 위해서 ANOVA³⁾와 같은 통계 방법을 이용

* 교신저자 : 송미영, 대전시 유성구 전민동 461-24 한국한의학연구원

· E-mail : skkim@kiom.re.kr, · Tel : 042-868-9526

· 접수 : 2011/07/27 · 수정 : 2011/08/26 · 채택 : 2011/09/20

하는 연구들도 있었지만 통계 방법만을 이용하는 것은 데이터마이닝 연구가 아니기 때문에 이러한 연구들은 제외하였다.

결 과

본 연구에서 조사한 22건의 논문을 분석한 결과 데이터마이닝 방법을 이용한 한의학 연구들은 사상체질 진단, 8체질 진단, 설진, 중풍 변증 진단, 온톨로지 기반 진단, 병인 진단등의 한의학 진단과 관련이 되어 있으며, 각 연구들에서는 진단 결과의 객관화를 위해서 데이터마이닝 방법을 이용하고 있었다. 특히 모두 인공 지능(artificial intelligence)⁴⁾의 한 연구 분야인 기계 학습(machine learning, 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야)⁵⁾ 방법을 이용하였다. 또한 모든 연구들은 각각 별도의 데이터베이스를 구축하고 이 데이터베이스에 대해서 실험한 결과를 제시하고 있었는데 이는 기존에 구축되어 있는 이용 가능한 한의학 정보 데이터베이스가 없을 뿐만 아니라 연구의 목적에 맞는 데이터베이스도 없기 때문이다. 이와 같이 연구를 위해서 데이터를 직접 구축하다보니 데이터의 양은 많지 않았다. 본 절에서는 각각의 진단 방법별로 연구들을 조사 및 분석한 결과를 기술한다.

1. 사상체질 진단

사상체질 분류를 위해서 기계 학습 알고리즘을 처음 적용한 시도는 박은경⁶⁾의 연구에서 제안되었다. 이 연구에서는 전문의에 의해서 체질이 결정된 200여명에게 사상체질분류검사지(QSCC II)^{7,8)}를 작성하게 하고 이를 의사결정나무(decision tree)⁹⁾를 이용해 분석하였다.

의사결정나무는 의사결정규칙을 나무 구조로 모델링하여 분류 또는 예측을 수행하는 방법이다. 이 방법은 사상체질 판별의 중요한 변수를 찾거나 각 체질들이 어떻게 분류되는지 발견하기 위해서 사용되고 있는데, 분석과정이 나무 형태의 그래프로 표현되기 때문에 다른 기계학습 방법들에 비해 연구자가 분석과정을 쉽게 이해하고 설명할 수 있는 장점을 가진다.

박은경의 연구에 의하면 기존의 사상체질 분류 방법들이 사상체질분류검사지에 의존하고 있으나 문항의 수가 많아 시간이 오래 걸릴뿐만 아니라 판별분석법(discriminant analysis)¹⁰⁾을 이용한 문항의 분석에 한계가 있기 때문에, 의사결정나무가 이러한 문제점을 해결하기 위한 새로운 분석 방법이 될 수 있다고 제안하고 있다. 즉, 의사결정나무를 이용하면 체질진단의 중요한 항목들을 찾을 수 있어 이 항목만 설문하면 보다 빠르게 체질진단의 결과를 얻을 수 있으며, 기존의 판별분석법과 같은 parametric 방법은 모집단의 자료가 특정분포(예, 정규분포)에 따른다는 기본 가정하에 분석하는 것이기 때문에 여러 제약이 많지만 nonparametric 방법 중의 하나인 의사결정나무는 이러한 제약이 없는 장점이 가진다. 이러한 이유로 박은경의 연구 이후에도 체질진단 연구에서 다른 방법보다 많이 이용되고 있다.

김규곤¹¹⁾의 연구에서는 박은영의 연구와 같이 사상체질분류 검사지를 이용해 422명의 표본을 얻은 후 판별분석과 의사결정

나무를 이용해서 체질을 분석하였으며 비교적 높은 수준의 진단 정확도를 보이고 있다고 제시하고 있다.

진회정의 연구에서는 사상체질 진단을 위해서 의사결정나무 방법을 다양하게 적용하였다. 예비연구¹²⁾에서는 의사결정나무 방법에서 가장 큰 빈도의 분류를 먼저 선택하고 나머지 분류에서 다시 큰 빈도를 선택하는 “one versus the rest” 방법으로 사상체질을 분류하였다. 이후에는 전형적인 사상체질을 가지는지 여부에 따라 표본 그룹을 둘로 나누고 두 단계의 의사결정나무 방법을 적용하는 방법¹³⁾을 제안하였다. 또한 통계 지식이 없는 일반 한의사가 의사결정나무 방법을 이용해 쉽게 사상체질 분류를 할 수 있도록 웹기반의 의사결정나무 자동 생성 프로그램을 개발¹⁴⁾하였다.

이와 같이 의사결정나무를 이용한 여러 연구가 있었지만 데이터마이닝에서 특정 문제에 적용하는 기법이 정해져 있는 것은 아니다. 또한 몇몇 연구자들이 사상체질 판별을 위해서 의사결정나무를 이용했다고 해서 사상체질 판별에 의사결정나무가 최적의 방법이라고는 할 수 없으며, 연구의 목적, 데이터, 연고자 하는 결과등에 따라 적용할 수 있는 기법들은 다를 수 있다. 홍진우¹⁵⁾는 WEKA(Waikato Environment Knowledge Analysis)¹⁶⁾를 이용해서 사상체질 연구에 있어 가장 높은 분류 타당성을 가지는 기계학습 알고리즘을 찾는 연구를 수행하였다. WEKA는 뉴질랜드 와이카토 대학에서 만든 자바 기반의 데이터마이닝 오픈소스 프로그램으로써 대부분의 기계학습 알고리즘들이 구현되어 있어 연구자들이 다양한 데이터 분석에 활용할 수 있도록 하고 있다. 이 연구에서는 79명에 대한 사상체질분류검사지 검사, MBTI(Myers-Briggs Type Indicator) 검사, BIA(Bio-Impedance Analysis) 검사 결과 데이터에서 30개의 속성을 선정하고 WEKA에서 제공하는 알고리즘을 가지고 평가를 수행하였다. 데이터의 종류와 양 그리고 알고리즘별 파라미터 설정 방법을 어떻게 설정하느냐에 따라 결과가 달라질 수 있고 또한 객관화된 평가 지표도 없는 상황이기 때문에 이 연구에서는 분류 결과의 타당성에 대해서는 결론을 언급하지 않았으며 대신 향후 연구 시 고려해야 할 내용들을 제안하고 있다.

채한¹⁷⁾은 신경망(neural network)¹⁸⁾을 이용해 사상체질을 분류하는 연구를 수행하였다. 신경망은 인공 뉴런들을 상호 연결한 네트워크 모델로써, 네트워크 상에 흐르는 내부 또는 외부 정보에 따라 모델 구조를 바꾸는 학습을 수행하게 된다. 본래 신경망은 생명체 신경의 특수한 정보처리 기능을 모방하고자 하는 의도에서 만들어진 인공지능 알고리즘의 하나이다. 신경망의 훈련 방법은 훈련 데이터가 뭔지를 아는 상태에서 훈련 데이터를 학습하는 supervised learning과 훈련 데이터가 무엇인지 모르고 숨겨진 구조를 찾는 unsupervised learning, 그리고 성공과 실패에 대한 보상 경험을 쌓으면서 작업을 개선 및 수행하도록 하는 reinforcement learning 방법으로 나뉘어진다. 채한의 연구에서는 102명의 사상체질검사지, MBTI, BIA 검사 데이터를 unsupervised learning을 이용한 SOM(Self-organizing Feature Map)이라는 신경망을 이용해서 사상체질 분석을 수행하였다. 특히 사상체질을 분류함에 있어 임상적 문맥(clinical context)를 제

의하고 unsupervised learning 방법을 이용함으로써, 기존의 다른 기계학습 알고리즘들과 달리 데이터 고유의 특성을 반영한 분류 결과를 도출하였으나, 반면에 이로 인해 예측율은 다른 방법보다 높지는 않았다.

위의 연구들에서는 사상체질을 분류할 때 사상체질분류지의 다양한 항목을 이용하였다. 하지만 일부 연구들에서는 안면이나 음성, 체형, 맥 정보만을 이용한 연구들도 존재한다.

도준형¹⁹⁾과 강제환²⁰⁾의 연구에서는 모두 사상체질 분류를 위해 SVM(Support Vector Machine)²¹⁾ 방법을 이용하였다. SVM은 분류 성능이 뛰어나 일반적으로 많이 이용되는 기계학습 알고리즘 중의 하나로써, 내부적으로 집단 간의 최대 마진 영역을 보장하는 최적의 방정식을 찾아 집단을 분류하는 방법이다. 특히 체질 분류를 위해서 성격, 피부등 다양한 정보를 이용하는 다른 연구들과 달리 도준형의 연구에서는 안면만을 가지고 체질을 분류하였으며 강제환의 연구에서는 음성만을 가지고 체질을 분류하였다. 이 연구들에서는 특정 정보만을 가지고 체질을 분류하기 때문에 여러 복합적인 정보를 이용해 체질 분류를 하는 것보다는 분류의 정확도가 떨어진다. 이를 개선하기 위해 판정불능 집단을 판정가능 집단에서 구분함으로써, 안면이나 음성으로 체질 판별이 가능한 집단에 대해서는 정확한 분류가 가능하도록 하였다.

이의주²²⁾의 연구에서는 강제환의 연구와 같이 음성분석을 이용해서 체질진단을 수행하였다. 이 연구에서는 사상체질검사지를 통해 체질이 판별된 90명에 대해서, 모음 /아/의 발음에 대해서 평상시 높이와 각각 다른 높이(고음, 중간음, 저음)의 음으로 발생했을 때의 파라미터 변화를 분석하고 이를 이용해 판별 분석을 수행하였다.

김종원²³⁾은 체질진단의사가 측정한 환자의 신체치수 즉, 체형기상 데이터를 이용하여 체질을 판별하는 연구를 수행하였다. 체질 판별을 위해서 체형 측정항목별로 변수를 선정 및 변환하고 이를 기반으로 판별분석을 수행하였다. 또한 이 연구에서는 분석을 위해서 체질진단을 받은 550명의 체형기상 데이터를 수집하였다.

김재욱²⁴⁾의 연구에서는 사상체질 감별을 위해 맥진을 이용하였다. 이 연구에서는 체질이 판별된 430명에 대해서 (주)대요메디의 3-D MAC기기를 이용해 맥파를 측정하고 SVM 방법을 이용해 체질을 분류하였다. 또한 도준형과 강제환의 연구와 같이 분류의 정확도를 높이기 위해서 판정 불가 집단을 구분하여 분석하였으나 그럼에도 불구하고 맥진을 이용한 사상체질 분류의 정확도가 다른 방법들보다 떨어진다고 보고하고 있다.

2. 8체질 진단

신용섭²⁵⁾은 8체질 진단 전문가 시스템을 개발하였다. 8체질 의학은 권도원 et al.²⁶⁾에 의해 소개된 體質論으로 맥진을 이용해 8체질을 감별하는 치료 방법이다. 8체질의학에서 각 체질은 각각 독특한 맥상을 가지고 있어서 맥진이 진단에 중요하게 사용되나, 8체질의학에서의 맥진은 기존 전통한의학의 맥진과 차이가 있어 임상에서 체질을 정확히 감별하기 위해서는 오랜 숙련과정이 필

요하다. 따라서 신용섭의 연구에서는 망진과 문진을 통해 체질별 특성을 계층 구조화²⁷⁾하고, 이를 기반으로 8체질 진단 전문가 시스템을 개발하였다. 이 연구에서는 진단 방법으로 사례 기반 추론(CBR:Case Based Reasoning)²⁸⁾을 이용하였다. 사례 기반 추론은 유사한 과거 사례의 해결 방법을 기반으로 새로운 문제를 해결하는 추론 방법을 말한다. 예를 들어 변호사가 기존의 판례를 이용해 현재 사건을 변호하는 것도 사례 기반 추론 방법이라고 할 수 있다. 이 시스템에서는 기존 연구에서 개발된 체크 리스트를 이용하여 우선 925명에 대한 사례베이스(case base)를 구축하고 CBR 기법 중의 하나인 최근접 이웃 추출법(nearest neighbor method)을 구현하였으며, 진단 결과를 높이기 위해서 계층적 분석 기법(analytic hierarchy process)이라는 속성 가중치 부여 방법을 이용하여 한의사의 의견을 반영할 수 있도록 하였다. 실제 사용자는 이 시스템에서 제공하는 설문 체크 리스트에 차례로 응답하면 최종적으로 8체질 결과를 볼 수 있도록 구현되어 있다.

3. 설진

설진은 혀의 상태와 여러 특징들을 진단에 사용하는 한의학 진단 방법이다. 하지만 정량화와 표준화 문제로 임상에서 많이 쓰이지 않고 있다. 따라서 몇몇 연구들에서는 판별함수를 이용하거나 새로운 통계함수를 개발^{29,30)}해서 설태 정보를 객관적으로 구분하는 연구를 수행하였다. 하지만 여기에서는 판별함수를 이용해서 설태를 분석하는 연구에 대해서만 조사하였다.

판별분석 방법을 이용해서 설태를 분석하는 연구는 이전³¹⁾의 연구와 최은지³²⁾의 연구가 있다. 두 연구 모두 연구방법은 비슷하며 단지 이전의 연구에서는 다양한 설태 중 백태와 황태를 분류하는 판별함수를 설계하였고, 최은지의 연구에서는 백태 중 후태와 백태를 분류하는 판별함수를 설계하였다. 논문에서는 기존에 제작된 디지털 설진 시스템^{33,34)}을 이용해서 720명의 설 영상을 수집하고 분석하였다. 이 중에서 황태와 백태라고 확진된 62명과 백태와 후태로 확진된 52명에 대해서 판별분석을 수행하였다.

4. 중풍 변증 진단

변증은 사진을 통해 환자의 증상을 분석하고 원인을 파악함으로써 증을 도출해내는 한의학의 진단 과정이다. 하지만 한의학 변증에 대한 표준화나 객관화에 대한 연구는 미비한 상황이다. 최근 한방의료기관의 입원환자의 질환 중 상위를 차지하는 중풍에 대해서 중풍변증진단 표준을 만드는 연구가 수행되었다. 이 연구에서는 전국 11개 한의과대학의 전문가들로 구성된 한의중풍진단 표준화위원회와 다기관 임상시험센터를 구축하고 변증 지표 및 변증 표준안을 마련하였다. 이를 위해 증례기록지를 만들고 이를 기반으로 증례를 수집하였으며 수집한 데이터를 가지고 중풍을 판별할 수 있는 판별함수를 설계하였다. 특히 박세욱³⁵⁾의 연구에서는 2006년 11월부터 2007년 3월까지 10곳의 한방병원에 입원한 502명의 증례 데이터를 수집하였으며, 이정섭³⁶⁾의 연구에서는 2008년 10월부터 2009년 6월까지 13개 한방병원에 입원한 중풍환자 데이터 541례에 대해서 분석하였다.

Table 1. 국내 한의 데이터마이닝 연구들의 알고리즘 및 연구 방법

| 논문 저자 | 알고리즘 | 연구 방법 |
|-----------------|-----------------|---|
| 박은경 et al. | 의사결정나무 | 사상체질 분류에 유효한 항목 분석 |
| 김규곤 et al. | 판별분석, 의사결정나무 | 판별분석과 의사결정나무 방법을 이용해 사상체질 분류 |
| 진희정 et al. (3편) | 의사결정나무 | 1. One versus the rest 방법을 적용한 사상체질 분류 2. 사상체질 판별을 위한 임상 지표 분석 3. 웹기반 의사결정나무 생성 프로그램 구현 |
| 채한 et al. | 신경망 | Unsupervised learning 방법을 사용하는 SOM을 이용해서 사상체질 분류 |
| 도준형 et al. | SVM | 안면 정보만을 가지고 사상체질 분류 |
| 강재환 et al. | SVM | 음성 정보만을 가지고 사상체질 분류 |
| 이의주 et al. | 판별분석 | 모음 /아/의 음성 정보를 가지고 사상체질 분류 |
| 김종원 et al. | 판별분석 | 신체체형 정보를 가지고 사상체질 분류 |
| 김재욱 et al. | SVM | 맥진 결과를 이용해 사상체질 분류 |
| 신용섭 et al. (2편) | 사례기반추론 | 1. 8체질의 체질별 특성을 계층 구조화 2. 8체질 진단 전문가 시스템 구현 (가중치 부여를 위해 계층 분석 방법 적용) |
| 이진 et al. | 판별분석 | 1. 설진에서 백태와 황태를 분류하는 판별분석 |
| 최은지 et al. | 판별분석 | 2. 설진에서 백태 중 후태와 박태를 분류하는 판별분석 |
| 박세욱 et al. | 판별분석 | 중풍변증진단을 위한 판별분석 |
| 이정섭 et al. | 판별분석 | 중풍변증진단을 위한 판별분석 |
| 최승훈 | 규칙기반추론 | 증상에 대한 변증을 추론하는 전문가 시스템 구현 |
| 박종현 et al. | 온톨로지추론 | 증상에 대한 변증을 추론하는 진단 시스템 구현 및 평가 |
| 김상균 et al. | 규칙기반추론 | 온톨로지를 이용한 규칙 기반 추론을 이용한 시맨틱 검색 시스템 구현 |
| 이혁재 et al. | 의사결정나무 | 의사결정나무를 이용해 병인 분류에 유의한 문항과 경로를 밝힘 |
| 홍진우 et al. | - | 사상체질 연구에 있어 가장 높은 타당성을 가지는 기계학습 알고리즘 탐색 |

5. 온톨로지 기반 진단

컴퓨터 과학이나 정보 과학 분야에서 얘기하는 온톨로지는 어떤 관심 분야를 개념화하기 위해 명시적으로 정형화한 명세서 (explicit and formal specification of a conceptualisation of a domain of interest)³⁷⁾라고 정의한다. 온톨로지는 본래 철학에서 기원했지만 최근 인공 지능 분야에서 도메인 지식에 대한 표준화된 표현 방법으로 많이 이용되고 있다. 또한 추론이라는 과정을 통해 기존 지식에서 새로운 지식을 발견할 수 있도록 하기 때문에 데이터마이닝 분야에서도 일부 이용되고 있다.

국내 한의 진단 시스템에서도 온톨로지를 이용한 연구들이 많이 존재하지만 여기에서는 온톨로지 추론 기법을 이용한 연구들을 조사하였다.

최승훈³⁸⁾은 한의진단명과 진단요건의 표준화 연구를 수행하면서 정리된 한의 변증 지식들을 기반으로 변증처방전문가 (KHU-PIPE : Kyung Hee University - Pattern Identification and Prescription Expert) 시스템을 개발하였다. 기존에 여러 한의 전문가 시스템들이 존재하지만 한의 지식에 대해서 표준화된 모델링이 수행되지 않아 여러 제약사항이 존재하고 있다. 반면에 이 연구에서는 진단에 대한 표준화 연구를 기반으로 하기 때문에 데이터에 대한 객관성을 확보하고 있다. 이 연구에서는 규칙 기반 추론(rule-based reasoning) 방법을 이용해서 변증 추론 기능을 구현하였으며, 환자의 증상을 입력하면 추론을 통해 변증과 처방 결과를 보여주는 시스템을 구현하였다.

박종현³⁹⁾의 연구에서는 한의 변증들을 OWL 온톨로지의 공리(axiom)를 이용해 정형화하고 추론할 수 있도록 하고 있다. 또한 온톨로지 추론을 수행하는 진단 프로그램을 구현하고 평가하는 방법을 제안하였다.

김상균⁴⁰⁾은 기초 온톨로지를 이용한 시맨틱 검색 시스템을 구현하였다. 이 시스템에서는 한의학의 약재, 처방, 병증의 기본 요소들을 정형화하고 약재를 중심으로 인스턴스를 구축한 기초 온톨로지를 기반으로 규칙 기반 추론을 수행하였다. 특히 추론을

위해서 SWRL⁴¹⁾ 추론 언어와 pellet⁴²⁾ 추론기를 이용하였다.

6. 병인 진단

이혁재⁴³⁾는 병의 증상에 따라 치료하는 것이 아니라 병의 원인을 찾아 치료하는 病因論에 기반한 진단 방법을 연구하였다. 특히 병인론은 환자를 진단하는데 명확한 방법을 제공하는 장점이 있지만, 여러 병인이 중복되면 이를 구분하기 어려울 뿐만 아니라 이 경우 의사의 주관에 빠지기 쉽다는 문제가 있다. 따라서 이 연구에서는 이 문제를 해결하기 위해 의사결정나무 방법을 이용하여 병인 분류에 유의한 문항과 경로를 밝힘으로써 효율적이고 객관적인 병인 분류방법을 제안하고 있다. 실험을 위해서 병인론에 따라 치료를 받은 255명의 환자를 대상으로 기존에 구축된 병인별 설문 데이터를 수집하였으며 이를 의사결정나무로 분석해서 의미있는 항목을 도출하였다.

고찰 및 결론

데이터마이닝은 기존에 존재하는 정보들로부터 새로운 지식을 발견하도록 하기 때문에 정보 과학 분야에서 많이 이용되는 방법이다. 양방은 물론이고 중의학 분야에서도 많은 데이터마이닝 연구⁴⁴⁻⁴⁶⁾들이 제안되고 있다.

이에 본 논문에서는 한의학 분야의 데이터마이닝 연구를 조사하고 분석하였다. 조사 대상은 국내의 한의학 연구로 한정하였으며, 자료 수집을 위해서 국내 한의학 논문을 서비스하고 있는 오아시스에서 논문을 검색해서 22건의 데이터마이닝 관련 연구들을 얻을 수 있었다.

데이터마이닝 관련 22건의 연구는 대부분 사상체질 진단, 8체질 진단, 설진, 중풍 변증 진단 등 한의 진단과 관련된 연구들이었다. 또한 온톨로지 관련 연구를 제외한 진단 연구들에서는 연구를 위해서 연구팀에서 직접 데이터를 수집하였으며 수집된 개수도 1,000건을 넘지 않았다. 일반적인 데이터마이닝 연구가

대용량의 데이터를 가지고 실험되는 것에 비하면 데이터의 수가 적은 편이지만, 이는 기존의 축적된 데이터를 활용하는 것이 아니라 연구를 위해서 직접 수집했기 때문이다. 또한 연구의 목표가 새로운 지식을 발견하기 위해 데이터마이닝을 하기보다는 진단 결과의 객관성을 보장하기 위해 기계학습등의 알고리즘을 이용하고 있는 차이가 있다. 사실 데이터마이닝의 연구들이라면 현재 운영되고 있는 한의 정보 데이터베이스나 EMR과 같은 대용량의 임상 데이터베이스에 대한 데이터 분석 연구가 수행하는 것이 일반적이지만 한의학 분야에서 아직 이러한 연구는 찾아볼 수 없었다.

데이터마이닝을 위해 사용되는 알고리즘은 홍진우의 연구에서 보이는 것처럼 아주 다양하지만, 한의 연구에서 사용되고 있는 데이터마이닝 방법은 의사결정나무, SVM, 신경망, 사례기반 추론, 온톨로지 추론, 판별분석 방법등 일부에 한정되어 있었다. 특히 의사결정나무, SVM과 판별분석 방법이 가장 많이 사용되었다. 의사결정나무는 분석 결과가 나무형태의 그래프로 표현되기 때문에 다른 기계학습 방법들에 비해 연구자가 쉽게 이해하고 설명할 수 있는 장점이 있어 많이 이용되고 있다. 특히 분류보다는 진단 과정을 분석해서 진단에 중요한 지표를 발견하기 위해서 주로 사용되었다. SVM은 알고리즘 특성상 비교적 적은 데이터를 가지더라도 높은 분류율을 보이는 장점을 가진다. 이로 인해 안면, 음성, 맥진을 이용한 사상체질 분류 연구에서 이용되었지만 모두 연구기관이 동일하기 때문에 SVM이 이러한 연구에 보편적인 방법이라고 하기는 어렵다. 판별분석은 두 개 이상의 집단을 구분하기 위해 많이 사용되는 통계분석법으로, 통계, 패턴인식, 기계학습 분야에서 이용되는 방법이다. 하지만 판별분석을 사용하지 않은 연구들에서는 판별분석 방법이 가장 작은 집단의 사례수가 독립변수의 수보다 많아야 하고, 독립변수들이 정규분포를 이루어야 하며, 공분산 행렬이 동일해야 하는 등 가정해야 할 점들이 많아 문제가 있다고 지적하고 있다. 그래서 이러한 문제를 해결하기 위한 새로운 통계방법으로써 기계학습 알고리즘을 이용하는 게 낫다고 주장하고 있다.

본 연구에서는 조사한 각각의 논문들에서 기술한 진단 결과의 정확도는 언급하지 않았다. 이는 각 연구마다 사용한 데이터의 양과 질이 다를뿐만 아니라, 사용한 데이터마이닝 방법과 각 방법을 적용할 때의 가정도 다르기 때문에 객관적으로 비교하기 힘들기 때문이다. 하지만 논문을 조사해본 결과 신경망과 맥진을 이용한 사상체질 연구와 진단 결과에 대한 분석이 없는 온톨로지 추론을 제외하고는 대부분의 연구에서 비교적 높은 정확도를 보인 것으로 나타났다. 그렇다고 해서 한의학 진단에서 데이터마이닝 알고리즘을 적용하는 것이 효율적이라고는 현재 쉽게 단정하기는 어렵다. 각각의 연구에서 사용한 데이터의 양이 적을 뿐만 아니라 아직 다양한 알고리즘의 시도가 부족하기 때문이다. 따라서 한의학 진단에 데이터마이닝 기술을 적용하기 위해서는 우선 방대한 한의학 지식 데이터베이스 구축이 필요하며 또한 이에 적합한 데이터마이닝 알고리즘에 대한 연구가 수행되어야 할 것이다. 또한 현재는 연구 동향이 진단 분야에 치우쳐 있지만, 한의학 데이터베이스가 구축이 되고 용어의 표준화가 이루어진

다면 약재, 처방등 다른 다양한 분야에서도 데이터마이닝을 이용한 한의 지식 분석도 가능할 것이다.

이와 같이 본 연구에서는 국내 한의학 논문에 대한 데이터마이닝 연구들을 조사하였다. 향후에는 국외의 한의학 논문들을 조사해서 본 연구를 보완할 예정이다.

참고문헌

1. http://en.wikipedia.org/wiki/Data_mining
2. <http://oasis.kiom.re.kr>
3. http://en.wikipedia.org/wiki/Analysis_of_variance
4. http://en.wikipedia.org/wiki/Artificial_intelligence
5. http://en.wikipedia.org/wiki/Machine_learning
6. 박은경, 이영섭, 박성식, 의사결정나무법을 이용한 체질진단에 관한 연구, 사상체질의학회지 13(2):144-155, 2001.
7. 金善豪, 高炳熙, 宋一炳, 四象體質分類檢査紙(QSCCⅡ)의 標準化 研究, 사상체질의학회지, 8(1):187-246, 1996.
8. 李廷燦, 高炳熙, 宋一炳, 四象體質分類檢査紙(QSCCⅡ)의 妥當化 研究, 사상체질의학회지, 8(1):247-294, 1996.
9. http://en.wikipedia.org/wiki/Decision_tree
10. http://en.wikipedia.org/wiki/Linear_discriminant_analysis
11. 김규곤, 김종원, 이의주, 김종열, 최선미, 데이터마이닝 기법을 이용한 사상체질 판별함수에 관한 연구, 동의생리병리학회지 18(6):1938-1944, 2004.
12. 진희정, 문진석, 고성호, 구임희, 이시우, 이도현, 송미영, 김종열, 사상체질 의사결정시스템 구축을 위한 체질 진단 자료를 이용한 예비연구, 한국한의학연구원은문집, 13(2):75-81, 2007.
13. 진희정, 김명근, 김종열, 사상체질 임상정보 분석을 위한 웹기반의 의사결정 나무 프로그램 개발, 한국한의학연구원은문집, 14(3):81-87, 2008.
14. 진희정, 이해정, 김명근, 김홍기, 김종열, 사상체질 판별을 위한 2단계 의사결정 나무 분석, 사상체질의학회지 22(3):87-97, 2010.
15. 홍진우, 김영인, 박소정, 김병철, 엄일규, 황민우, 신상우, 김병주, 권영규, 채한, 사상체질 진단검사를 위한 데이터마이닝 알고리즘 연구, 동의생리병리학회지 23(6):1234-1240, 2009.
16. <http://www.cs.waikato.ac.nz/ml/weka/>
17. 채한, 황상문, 엄일규, 김병철, 김영인, 김병주, 권영규, 신경망을 사용한 사상체질 진단검사 개발 연구, 23(4):765-771, 2009.
18. http://en.wikipedia.org/wiki/Neural_network
19. 도준형, 김성훈, 구임희, 김근호, 김종열, 판정불능을 포함한 안면 체질 분류 방법에 관한 연구, 사상체질의학회지 21(3):39-47, 2009.
20. 강제환, 도준형, 김종열, SVM을 이용한 음성 사상체질 분류 알고리즘, 사상체질의학회지 22(1):17-25, 2010.
21. http://en.wikipedia.org/wiki/Support_vector_machine

22. 이의주, 송광빈, 최한수, 유정희, 곽창규, 손은혜, 고병희, 음성분석에 의한 체질진단에 관한 연구, 대한한의학회지 26(1):93-102, 2005.
23. 김종원, 김규근, 이의주, 이용태, 신체계측법에 의한 사상체질별 체형기상 연구 1, 동의생리병리학회지 20(1):268-272, 2006.
24. 김제욱, 김성훈, 이유정, 전영주, 김근호, 김종열, 맥진을 이용한 사상체질 판별 방법의 개선 및 의의, 한국한의학연구원논문집, 15(2):93-100, 2009.
25. 신용섭, 박영배, 박영재, 김민용, 오환섭, 8체질의학을 위한 진단 전문가 시스템 개발 및 고찰, 대한한의학진단학회지 12(1):142-184, 2008.
26. Dowon Kuon, A Study Of Constitution-Acupuncture. Journal of the International Congress of Acupuncture & Moxibustion: Tokyo, Japan Acupuncture & Moxibustion Society, 10: 149-167, 1965.
27. 신용섭, 박영배, 박영재, 김민용, 이상철, 오환섭, 8體質 診斷 專門家 시스템 개발을 위한 基礎研究, 대한한의학진단학회지 11(1):25-47, 2007.
28. http://en.wikipedia.org/wiki/Case-based_reasoning
29. 김근호, 도준형, 유현희, 김종열, 설진 기기의 시스템 구성 및 진단 방법 개발, 한국한의학연구원논문집, 14(3):89-95, 2008.
30. 김근호, 유현희, 김종열, 디지털 자동 설진 시스템 구축을 위한 설태 이식 알고리즘 기초 연구, 동의생리병리학회지 23(1):97-103, 2009.
31. 이 진, 최은지, 유현희, 이해정, 이유정, 박경모, 김종열, 다차원 컬러벡터 기반 백태 및 황태 분류 판별함수 설계, 한국한의학연구원논문집, 13(2):47-52, 2007.
32. 최은지, 김근호, 유현희, 이해정, 김종열, 백태 중 후태 및 박태 분류 판별함수 설계, 한국한의학연구원논문집, 13(3):119-124, 2007.
33. 김제균, 디지털 설진 시스템의 개발, 경희대학교 학사논문, 2005.
34. 어윤혜, 디지털 설진 시스템을 위한 설 검출과 분류, 경희대학교 석사논문, 2006.
35. 박세욱, 강병갑, 장인수, 홍석, 한창호, 권정남, 선승호, 전찬용, 조기호, 박세진, 이인, 설인찬, 최선미, 다기관 임상연구를 통해 도출된 증풍변증표준안의 진단프로그램개발에 관한 연구 - I, 대한한학회지 28(3):126-137, 2007.
36. 이정섭, 김소연, 강병갑, 고미미, 김정철, 오달석, 김노수, 최선미, 방옥선, 전문가 변증과정을 반영한 증풍 변증 판별모형, 동의생리병리학회지 23(6):1460-1464, 2009.
37. T. R. Gruber. A translation approach to portable ontologies, Knowledge Acquisition, 5(2):199-220, 1993.
38. 최승훈, 인터넷 기반 한의진단전문가 시스템(KHU-PIPE) 개발, 동의생리병리학회지 16(3):528-531, 2002.
39. 박종현, 신상우, 정길산, 박경모, 김선호, 한의진단 Ontology 구축과 평가, 동의생리병리학회지 20(1):202-208, 2006.
40. 김상균, 장현철, 김진현, 김철, 예상준, 송미영, 한의 기초 온톨로지 기반 시맨틱 검색 시스템, 한국한의학연구원논문집, 17(2), 2011.
41. <http://www.w3.org/Submission/SWRL/>
42. <http://clarkparsia.com/pellet/>
43. 이혁재, 김민용, 오환섭, 박영배, 의사결정나무법을 이용한 病因분류에 관한 연구, 대한한의학진단학회지 12(2):27-40, 2009.
44. Yi Feng, Zhaohui Wu, Xuezhong Zhou, Zhongmei Zhou, Weiyu Fan, Knowledge discovery in traditional Chinese medicine: State of the art and perspectives, Artificial Intelligence in Medicine, 38: 219-236, 2006.
45. Suryani Lukman, Yulan He, Siu-Cheung Hui, Computational methods for Traditional Chinese Medicine: A survey, Computer Methods and Programs in Biomedicine, 88: 283-294, 2007.
46. Xuezhong Zhou, Yonghong Peng, Baoyan Liu, Text mining for traditional Chinese medical knowledge discovery: A survey, Journal of Biomedical Informatics, 43: 650-660, 2010.