

# 데이터마이닝 기법을 이용한 수술후 환자 진단

이경영 이주철 박순철  
전북대학교 정보통신공학과  
anthem@internet.chonbuk.ac.kr

## Application of Data Mining Techniques To postoperative patient Condition Diagnostic Predictions

Kyung Young Lee, Ju Cheel Lee, Soon Choel Park  
Dept. of Information and communication,  
Chonbuk National University

### 요약

정보화를 통한 업무의 효율성 제고에 대한 인식이 폭넓게 확산돼 있다. 의료분야에서도 비교적 단순한 원무관리 시스템이나 환자의 증상이나 각종 자료 등을 기록하고 의료진간의 공유를 가능하게 하는 전자의료기록 관리시스템의 구축이 필요하다. 또한 이들 시스템을 통하여 획득한 환자의 자료를 분석하여 의료진의 환자질병진단을 지원하고자 하는 연구가 활발히 진행되고 있다. 본 논문에서는 의료자료 분석에 요구되는 기법을 제시하며, 획득한 환자의 자료를 데이터마이닝 기법인 신경망 모델을 적용하여 결과를 분석한다.

### 1. 서론

최근 들어 지식자원의 효과적 관리 및 전략적 활용이 무엇보다 중요하다고 보는 새로운 경영 패러다임인 지식경영(Knowledge Management)의 중요성이 강조되고 있다. 다년간의 조직운명을 통해 축적된 전산자료를 분석하여 획득하고 이를 토대로 의사결정을 지원하는 정보 시스템의 필요성이 부각되고 있다.

타 분야와 마찬가지로 대부분의 의료기관에서도 수준 높은 의료서비스를 저렴하고 신속하게 제공하는 단계에 있다. 그러나 대부분의 병원은 단순한 원무관리를 지원하는 시스템 수준이다. 환자의 증상이나 각종 병력자료, 검사자료 등을 기록하고 의료진간의 공유를 가능하게 하는 전자의료기록 관리시스템의 보급은 미비한 상태이다.

선진 의료기관들의 대부분은 전자의료기록 관리시스템의 구축을 이미 완료한 상황이다. 한층 더 나아가 시스템을 통하여 획득한 환자의 자료를 분석하여 의료진의 환자 질병진단을 지원하고자 하는 연구가

꾸준히 진행되어지고 있다. 이러한 목적은 의료진의 활용을 유도함에 있다. 의료진의 의료지식과 진단의 폭을 넓힐 수 있다는 점과 질병의 진단이나 예후를 좀더 객관화함으로써 의료진단의 표준화를 기할 수 있다는 점이다.

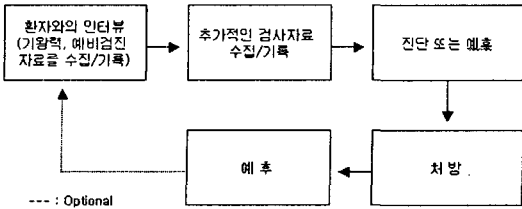
본 연구의 목적은 전자의료기록 관리시스템의 지원 환경 하에서 데이터마이닝 기법을 활용하여 환자의 인적, 병력, 증상, 의료검사자료와 질병간에 존재하는 패턴을 추출하여 의료진의 정확한 질병 진단이나 예후를 가능하게 한다. 또한 의료자료를 신경망 모델에 적용하여 결과를 분석하는데 있다.

### 2. 데이터마이닝 기법을 이용한 의료 진단의 개요

#### 2.1 의료진단 절차 및 진단시스템의 필요성

일반적으로 의료진단은 [그림1]과 같은 절차를 통해 이루어진다.

의료진의 최종 진단은 여러 형태를 가질 수 있다. 예를 들어 첫 번째 진단이 될 수도 있고 처방이 정해진 후 그 결과를 검토한 다음에 내려질 수도 있



[그림 1] 의료진단 절차

다. 또한 100% 신뢰할 만한 최종 진단을 내리기가 어려운 경우도 있다.

[표 1]은 진단이나 예후를 위한 정보시스템의 필요성 및 수반되는 문제점을 나열하고 있다.

[표 1] 자동 의료진단 시스템의 필요성 및 문제점

필요성	문제점
· 의사결정의 모호성 감소	· 의사 결정에 고려되는 일부자료
· 진단의 객관화 및 표준화	(예, 혈액, 기력등)의 전자화의
· 전문 인력비용 감소	어려움 및 신뢰도 감소
· 높은 검사비용 절감 및	· 의료진의 정보화 마인드 부재
시간 단축	· 예후의 경우 정보시스템의 지식
· 대체검사 항목의 부재로	베이스를 구성할 만한 자료
인한 예후 신뢰도 확보	수집의 어려움

## 2.2 데이터마이닝의 개요

컴퓨터의 사용이 보편화됨에 따라 과학 기술, 경제, 산업등 사회의 각 분야에서 많은 양의 데이터가 손쉽게 수집, 저장되고 있다.

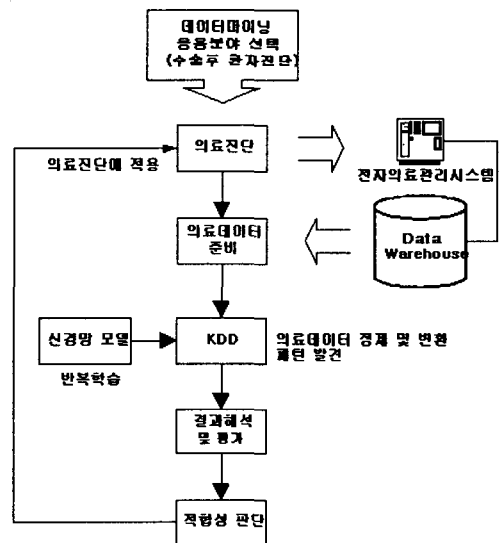
지식탐사(KDD, Knowledge Discovery in Database)는 이러한 데이터에서 유용하고 이해 가능한 지식을 추출하는 연구이다. 데이터마이닝은 KDD에서 가장 중요한 단계로 데이터로부터 지식을 추출하는 단계이다. 데이터마이닝에서 생성된 지식은 좋은 분류율을 가져야하고 이해하기 쉬워야 한다.

[그림 2]에서와 같이 신경망 기법을 이용한 의료진단의 수행 1단계는 전자의료관리시스템과 연계된 의료진단으로 데이터를 준비한다. 2단계는 지식탐사(KDD)절차를 통해 데이터 정제 및 변환한다. 이때 신경망 기법을 이용하여 반복학습을 통해 패턴을 찾아 지식을 추출하고 예측률 및 오류율을 측정한다. 마지막 해석 평가 단계에서는 데이터마이닝을 통하여 얻어진 지식을 해석하며 기존의 지식과 상충되는 지 비교하여 이를 해결한다.

### 2.2.1 자료분석 기법 (신경망 모델)

신경망은 인간 두뇌의 신경세포를 모델링하여 지능을 구현하는 기법이다. 사람의 뇌처럼 신경망은 Input set에서 학습을 하고, 패턴을 찾기 위해 학습 결과를 모델의 파라미터를 조정하는 데 사용한다.

신경망은 비선형 기법이지만 비선형 회귀분석처럼 특정한 설정이 필요 없다. 또한, 신경망에서는 나이와 성별이 합쳐진 효과 같은 상호작용 효과가 뚜렷이 파악된다. 신경망을 데이터마이닝 알고리즘에 적용할 경우 비선형 적인 패턴분할을 수행할 수 있는 신경망의 장점 때문에 매우 정확한 지식을 생성할 수 있다.



[그림 2] 데이터마이닝을 이용한 의료진단 과정

## 3. 연구방법

### 3.1 자료의 성격

University Of California At Irvine의 부속기관인 Machine Learning Center에서 관리하는 의료관련 자료를 사용했으며 구체적인 자료별 입력/목표변수와 특성을 [표 2]에 서술하고 있다.

### 3.2 신경망 모델 구현

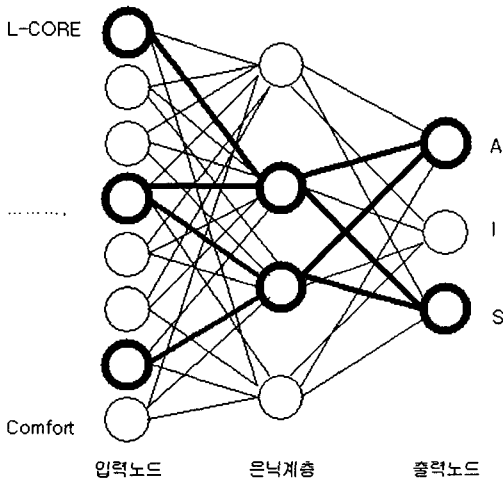
환자의 진단자료를 모형추정용(Training), 모형검증용(Validation), 모형시험용(Testing) 자료를 구성하였으며, 결과의 일반화를 도모하기 위해 진단자료를 13회 반복학습을 했으며, 진단자료의 연속형 변수의 빈 값은 평균값을 대입해 처리하였다.

[그림 3]의 신경망 모델의 8개의 입력과 3개의 출력 노드를 가지고 있을 뿐만 아니라 은닉노드를 가

[표 2] 수술 후 환자 상태 Data 입력 / 목표변수

자료명	입력변수	목표변수
Postoperative Patient Data	· L-Core	Intensive Care (I)
	· L-Surf	
	· L-O <sup>2</sup>	
	· L-BP	Prepared To Go Home (S)
	· Surf-STBL	
	· Core-STBL	
	· 8P-STBL	General Hospital (A)
	· COMFORT	
	· Decision ADM-DECS	
※ Postoperative Patient Data - 수술 후 회복실내에 있는 환자의 검사자료를 근거로 퇴원여부를 결정한다. - 저 체온증과 같은 긴급을 요하는 환자의 자료는 포함되지 않았다. (48% Accuracy) - Number Of Instance : 90 - Number Of Attributes : 9 Including the decision (Class Attribute) - Missing Attribute Values: 3		

지는 여러 개의 중간 계층으로 구성되어 있다. 입력노드는 은닉노드와 연결되어 있으며 은닉 노드는 출력 노드에 전부 연결되어 있다.



[그림 3] 환자 진단결과를 식별하는 신경망

신경망 모델 구현의 절차는 다음과 같다.

- ① 학습용 입력데이터에 대해 신경망의 실제 출력과 정확한 값에 해당하는 기대 출력 값을 가중치를 적용하여 비교한다.
- ② 응답결과가 어느 정도 정확할 때까지 가중치를

적용하여 반복 학습한다.

- ③ 신경망 구조가 안정화되면 목표변수에 대한 예측률 / 오류율을 판단한다.

#### 4. 결과 및 해석

[표 3]은 학습되어진 목표변수의 모형추정용 자료의 평균치를 나타내고 있다.

[표 3] 목표변수의 평균

목표변수	L-CO	L-SU	L-O <sub>2</sub>	L-BP	S-ST	C-ST	BP-S	CFT
I	L	M	E	H	M-S	S	M-S	M
S	M	M	G	M	S	S	S	M
A	M	M	G	M	S	S	S	M

· I (patient sent to Intensive Care Unit)  
 · S (patient prepared to go home),  
 · A (patient sent to general hospital floor)

목표변수 S(5회 반복학습)와 A(12회 반복학습)의 모형추정용 자료는 대부분의 Attribute들은 정상범위 내에 포함되어 있으나, 목표변수 S가 A에 비해 정상범위에 근접함을 보이고 있다. 목표변수 I(2회 반복학습)의 자료가 미비하기는 하나 자료를 분석하여 보면 대부분의 Attribute들이 정상범위를 벗어나 환자들 저 체온증과 고혈압 증상을 보이고 있다는 결과가 나타났다. 연구결과 수술후 환자의 몸 상태가 저체온증과 고(저)혈압이 있는 환자일수록 관심도가 높다는 규칙을 발견하였다. 또한 사용되어진 입력변수 L-O<sup>2</sup>와 COMFORT에 대해서는 아무런 영향을 미치지 않았다.

Machine Learning Center에서 제공하는 의료관련 자료와 실험한 자료의 오류율이 32.7%로 높게 나왔다. 본 논문에 사용된 의료자료는 위에서 제시한 자료의 성격에 나타난 것처럼 긴급을 요하는 환자의 자료가 누락되었으며, 입력변수에 대한 목표변수 A와 S가 유사하기 때문에 목표변수 A와 S에 대한 보다 정확한 예측률이 떨어진 결과이다. 신경망은 목표변수만을 제시하는 단순한 블랙박스로서 목표변수에 어떻게 도달하는 것에 대한 과정을 제시하지 못하고 있다.

#### 5. 결론 및 향후과제

데이터마이닝 기법을 활용하여 환자의 인적, 병력, 증상, 의료검사자료와 질병간에 존재하는 패턴을 추출하여 의료진의 정확한 질병 진단을 가능하게 하고 의료업계에 품질 향상을 꾀하는 방법을 제시하였다.

그러나 자료수집의 어려움으로 인해 국외 자료를 사용하였고, 신경망 기법을 통한 질병진단 및 예후에 관한 결과가 실제 의료진들의 진단결과와 비교하지 못함으로 인해 보다 객관적인 연구 결과를 제시하지 못하였다.

환자의 질병연구의 활성화를 위해서는 국내에서도 의료정보시스템 관련 연구의 활성화와 환자의 의무 기록의 전산화가 선행되어야 할 것이다.

앞으로의 과제는 자료분석을 위해서 다양한 기법들(로지스틱 회귀분석, 의사결정나무)을 적용하여 비교/분석함으로써 각각의 기법들의 강점과 한계 등을 살펴보는 것이 연구의 효율성과 신뢰도의 제고를 위해 필수적이다.

6. 참고문헌

- [1] <http://www.ics.uci.edu/~mlearn/MLSummary.html>
- [2] 장남식 · 홍성완 · 장재완. “데이터마이닝”, 대청출판사
- [3] Pieter Adriaans · Dolf Zantinge, “DataMining”
- [4] 강현철 · 한상태 · 최종후 · 김은성 · 김미경, “SAS Enterprise Miner4.0을 이용한 데이터마이닝 - 방법론 및 활용”, 자유아카데미
- [5] JACEK M.ZURADA, “Introduction to Artificial Neural Systems”
- [6] <http://www.Spss.co.kr>
- [7] Robert Groth, “DataMining Building Competitive Advantage”



박 순 철

1972-1979 인하대 응용물리학과  
 1986-1991 미국 루이지애나 주립대 전자계산 박사  
 1991-1993 ETRI DB 연구실 과제 책임자  
 1993-현재 전자정보공학부 부교수

관심분야 : DB, SAN, 정보검색



이 경 영

1994-1998 원광대 반도체학과  
 1999- 현재 전북대 전자정보공학부 인터넷연구실  
 관심분야 : DB, DataMining, 소프트웨어공학



이 주 철

1981-1988 전북대 상과대학 회계학과  
 2000- 현재 전북대 전자정보공학부 인터넷연구실  
 관심분야 : Network Security, DB