

모발분석 및 처리를 위한 한국형 의료 정보 시스템 구축

조 영 임^{*}

요 약

현재 세계 48개국의 의료기관에서 임상을 위해 널리 사용 중인 모발분석(TMA)은 중요 미네랄 비율을 분석하여 체내에 과잉, 결핍 및 불균형 상태를 평가하고 인체에 미치는 영향을 예측하여 건강유지 방향을 제시하는 임상 영양학 및 독성학 검사방법을 말한다. 그러나 국내 모발분석방법에는 몇가지 문제점이 있다. 즉, 모발분석결과를 처리하고 해석할 수 있는 한국형 의료정보 데이터베이스가 없으므로 미국에 의뢰하고 있는데, 외화낭비는 물론 보내오는 모발분석 검사결과지가 영문이고 철저한 보안 위주파일이므로 국내의료기관에서의 활용도가 매우 낮다. 또한 모발분석 결과가 서구식 데이터베이스로부터 분석된 것이므로 검사결과와의 신뢰성 문제도 발생한다. 따라서 본 논문에서는 이러한 문제점을 해결하기 위해 국내최초로 TMA 기반 한국형 의료 정보 시스템을 구축하였다. 이 시스템은 복잡한 모발분석 자료의 분류를 다단계 통계분석 방법에 의한 결정트리 분류기를 통해 수행하고, 다중 퍼지 규칙방식의 데이터베이스를 구축하여, 지능형 퍼지추론 방법에 의해 모발분석 자료를 분석한다. 본 시스템의 성능을 실제 작업환경에서 측정한 결과 시스템을 사용하는 경우가 사용하지 않았을 경우보다 업무능률과 사용자 만족도가 각각 86%, 92% 증가하였다.

Implementation of Medical Information System for Korean by Tissue Mineral Analysis

Young-Im Cho^{*}

ABSTRACT

TMA(Tissue Mineral Analysis) is very popular method in hair mineral analysis for health care professionals in over 48 countries' medical center. Assessment of nutritional minerals and toxic elements in the hair is very important not only for determining adequacy, deficiencies and unbalance, but also for assessing their relative relationships in a body. In Korea, there are some problems in TMA method. Because of not having a medical information database which is suitable for Korean to do analyze, the requested TMA has to send to TEI-USA. However, as the TMA results from TEI-USA is composed of English documents and graphic files prohibited to open, its usability is very low and a lot of dollars has to be paid. Also, it can make some problems in the reliability of the TMA results, since the TMA results are based on the database of western health and mineral standards. To solve these problems, I developed the first Medical Information System of TMA in Korea here. The system can analyze the complex tissue mineral data with multiple stage decision tree classifier. It is also constructed with multiple fuzzy database and hence analyze the TMA data by fuzzy inference methods. The effectiveness test of this systems can be shown the increased business efficiency and satisfaction rate 86% and 92% respectively.

Key words: 모발 분석, 의료 정보 데이터베이스, 통계 분석, 퍼지 규칙베이스

접수일 : 2002년 8월 1일, 완료일 : 2002년 9월 18일
보건복지부의 2002년도 벤처 및 중소기업 기술개발지원
과제 중 보건의료정보관련 제품개발 과제에 선정되어 지
원사업으로 연구한 논문입니다.

^{*} 정회원, 평택대학교 컴퓨터학과 조교수

1. 서 론

인체의 상태를 평가하는 모발분석(TMA:Tissue Mineral Analysis)은 머리카락 속에 있는 30여 가지의 미네랄과 8가지의 중금속의 양과 중요 미네랄 비율을 분석하여 체내에 과잉, 결핍 및 불균형 상태를 평가하고 대사타입(metabolic type)을 제시하여 건강을 유지하는 방향을 제시하는 임상 영양학 및 독성학 모발조직 검사방법을 말한다[1-3]. 또한 호르몬의 세포수준 활성을 제시(갑상선 및 부신)하며 영양, 신경 및 호르몬의 불균형을 통해 질환발생 가능성을 설명한다. 모발분석자료는 매우 복잡할 뿐 아니라 38종의 미네랄과 미네랄 성분비가 광범위한 수치범위(-0.00001~+60000)를 갖고 있어서 다양한 통계적 분석 및 추론기법이 필요하다.

모발분석에 관해서는 미국 TEI(Trace Elements Inc.)에서 세계 최대의 데이터베이스를 보유하고 있는데 이는 약 20여년간 임상실험한 약 35만건 이상의 자료로부터 구축된 것으로 현재 미네랄 관련 가장 많은 임상자료를 보유하고 있으며, 세계 40여개국에 있는 의료기관에 의학 서비스를 제공하는 FDA 승인된 세계 최대 모발분석 전문기관이다[4]. 한편, 국내에서 모발검사를 하는 기관들이 있는데 자체 데이터베이스를 보유하고 있지 않아서 의료기관에서 사용하기에 어려운 점이 많다[5,6]. 현재 국내에서는 대학병원[7] 등을 중심으로 미네랄 관련 임상실험을 하고 있으며 점차로 모발분석에 대한 인식이 확산되고 있는 추세이다.

그러나 국내에서는 모발분석 방법이 없기 때문에 미국 TEI에 대한 의존도가 높는데, 검사하려는 환자 자료를 미국 TEI에 보내면, 자체 대규모 데이터베이스로부터 환자 개인의 미네랄 양과 성분비, 대사 및 영양상태 등을 종합적으로 분석된 환자 및 의사용 검사결과지(영문)를 국내 의뢰기관에 제공하게 된다.

그러나 미국에서는 데이터베이스 및 분석방법을 제공하지 않고 있으므로 현재 국내 TMA 분석 방법에서는 다음과 같은 문제점을 갖는다.

첫째, TMA 분석 기기는 있으나 분석결과를 전문적으로 해석하여 환자와 의사들에게 제공할 수 있는 프로그램 및 의료정보 데이터베이스가 없다. 현재는 TMA 분석을 위해 미국 TEI에 의뢰하여 결과를 받

아보는 방법을 이용하고 있는데, 이 방법은 현재 국내에서 TMA 분석을 의뢰하는 수요자가 증가하는 추세를 감안할 때 시간과 효율측면에서 적절하지 못하다.

둘째, 미국에서 보내오는 TMA 검사결과가 영문이라 한글화된 결과지가 필요하며 암호화된 파일형태이므로 시스템적 활용도가 낮다. 또한 한국적 특성이 아닌 서구식 데이터베이스에 의한 결과물이므로 검사결과와 신뢰성이 부족할 수밖에 없으며 한국적 임상 결과에 대한 검증자료가 없다

셋째, 검사비용으로 인한 연간 18억원 이상의 외화낭비가 초래된다.

따라서 본 논문에서는 이러한 문제점을 해결하기 위해 모발분석을 위한 알고리즘을 개발하고 모발분석결과에 대한 한국형 의료정보 시스템을 구축하고자 한다. 또한 통계분석적 오류와 데이터베이스 구축 및 추론상의 오류를 최소화하기 위해 다중 퍼지규칙 베이스와 다단계 통계분류 방법을 이용하고 한글화 데이터베이스 구축을 위해 한글화 작업도 병행하고자 한다. 이렇게 함으로써 미국에 검사비용 및 결과지 해석비용으로 지불하던 외화 낭비를 줄이고 환자의 추가 의료비를 절감하여 미국 TEI의 독점을 방지하고자 한다. 본 개발 시스템은 미국의 시스템의 단순한 한글화가 아니며, TMA 관련 데이터베이스 및 한국형 의료정보 시스템 구축에 관한 연구로는 국내에서 최초로 시도되는 연구이다.

본 논문의 구성은 다음과 같다. 2장에서는 시스템 구축에 필요한 지능적 추론방법, 다중 퍼지규칙 베이스 구성 및 통계적 분류 방법을 제안하며, 3장에서는 한국형 의료정보 시스템 개요와 효율성 등을 설명하고, 4장에서 결론을 내리고자 한다.

2. TMA 의료정보 시스템

2.1 시스템 개요

본 논문에서 구축하려는 TMA 분석을 위한 한국형 의료정보 시스템은 그림 1과 같이 TMA 자료의 통계적 분석 단계, TMA 의료정보 데이터베이스 시스템 구축단계, 지능적 추론 시스템 구축 단계 등 크게 3단계로 구성되어 있다. 연구방법은 통계분석단계에서 통계적 분석이 완료된 후 시스템 구축단계에

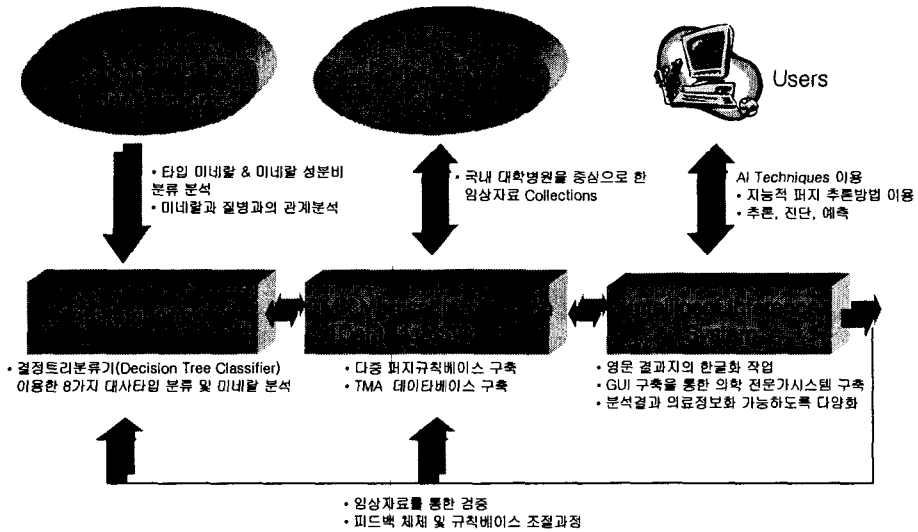


그림 1. 시스템 개요

서 TMA 데이터베이스를 구축하고, 임상자료의 피드백을 통해 시스템을 검증한다.

2.2 TMA 자료의 통계적 분석 단계

TMA 자료의 지능적 통계적 분석을 위해 대사타입의 분류 및 미네랄의 분석은 통계적 분류 및 분석 방법 중에서 결정트리분류기(Decision Tree Classifier)를 이용하여 TMA 자료의 대분류에 해당하는 인체의 8가지 대사타입(빠른 대사 4타입, 느린 대사 4타입)을 분류한다. 이 방법을 사용하는 이유는, 1단계 분류기는 클래스간의 변별력 중에서 평균이 최대가 되는 특징조합을 이용하나, 이 특징조합이 모든 클래스에 대하여 최적의 변별력을 항상 갖는 것이 아니므로 두 종류이상의 결정을 적절한 수준에 따라 몇 번이고 반복하여 입력패턴의 클래스를 결정해야 하므로 본 논문에서는 다단계분류기인 결정트리 분류방법[8-10]을 응용하여 분류한다.

그림 2는 결정트리분류기가 패턴집합을 분류한 것인데, 그 과정은 다음과 같다. 우선 루트노드에서 미네랄 비 $Ca/P(=2.63)$ 를 이용하여 느린 대사타입(Slow Metabolic Type)의 패턴집합을 $\{S_1, S_2, S_3, S_4\}$ 와 빠른 대사타입(Fast Metabolic Type) 패턴집합 $\{F_1, F_2, F_3, F_4\}$ 의 서브패턴집합으로 분류한 다음, 다시 내부노드에서 $Ca/K(=4.2)$ 를 이용하여 1차 분류된 서브패턴 집합 $\{S_1, S_2, S_3, S_4\}$ 를 패턴 $\{S_1, S_2\}$ 와

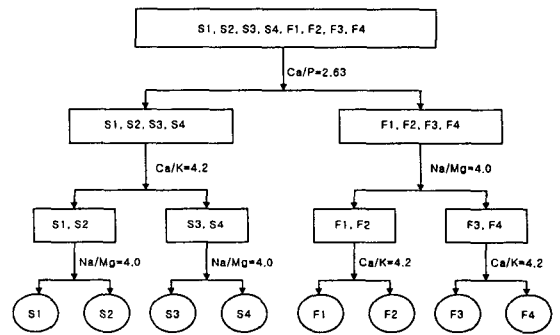


그림 2. 결정트리에 의한 TMA 타입 분류

$\{S_3, S_4\}$ 로 분류하고 $Na/Mg(=4.0)$ 를 이용하여 1차 분류된 서브패턴 집합 $\{F_1, F_2, F_3, F_4\}$ 를 $\{F_1, F_2\}$ 와 $\{F_3, F_4\}$ 로 분류하였다. 마지막 단계에서 $Na/Mg(=4.0)$ 를 이용하여 2차 분류된 패턴집합 $\{S_1, S_2\}$ 와 $\{S_3, S_4\}$ 를 S_1 과 S_2 그리고 S_3 과 S_4 로 최종 분류하고, 같은 방법으로 $Ca/K(=4.2)$ 를 이용하여 2차 분류된 패턴집합 $\{F_1, F_2\}$ 와 $\{F_3, F_4\}$ 를 F_1 과 F_2 그리고 F_3 와 F_4 로 최종 분류하여 결정트리를 구성하였다. TMA에서 가장 기본이 되는 대사타입의 분류 후에 각 대사타입 별 영양미네랄과 독성미네랄의 영양상태와의 관계는 상관분석을 통해 상관계수가 0.3보다 큰 미네랄들을 대상으로 분석하였다.

대사타입의 분류는 특성변수에 대한 클래스별 분포를 이용해도 같은 결과를 보여 준다. 다음은 특성

변수를 계급적으로 적용하여 클래스별 분포를 이용한 대사타입의 분류를 보여준다. 우선 Ca/P 클래스 분포에 의해 대사타입을 $\{S_1, S_2, S_3, S_4\}$ 와 $\{F_1, F_2, F_3, F_4\}$ 의 서브패턴집합으로 분류한 것이 그림 3이다.

그림 3에서 왼쪽의 4개 분포는 침도가 높고 표준편차가 작은 대사타입 F 의 Ca/P분포로

$$F_1 \sim (\hat{\mu}=1.97, \hat{\sigma}=0.38),$$

$$F_2 \sim (\hat{\mu}=1.85, \hat{\sigma}=0.28),$$

$$F_3 \sim (\hat{\mu}=2.01, \hat{\sigma}=0.32),$$

$$F_4 \sim (\hat{\mu}=2.34, \hat{\sigma}=0.26)$$

를 보여주고 있으며 아래쪽으로 누운 형태의 4개 분포는 침도가 낮고 표준편차가 큰 대사타입 S 의 Ca/P 분포로 $S_1 \sim (\hat{\mu}=9.31, \hat{\sigma}=6.65)$,

$$S_2 \sim (\hat{\mu}=5.35, \hat{\sigma}=3.45),$$

$$S_3 \sim (\hat{\mu}=5.75, \hat{\sigma}=3.60),$$

$$S_4 \sim (\hat{\mu}=3.79, \hat{\sigma}=1.64)$$

을 보여주고 있어 Ca/P의 분포에 의해 대사타입 F 와 S 가 분류됨을 알 수 있다.

나머지 빠른대사 타입도 같은 방법으로 분류하였다.

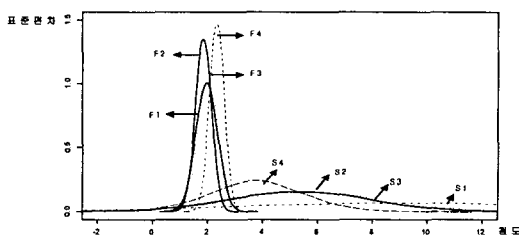


그림 3. Ca/P 클래스 분포를 이용한 대사타입의 분류

2.3 TMA 의료정보 데이터베이스 시스템 구축 단계

분석된 통계를 바탕으로 본 논문에서는 불확실한 다량의 정보로부터 지능적 추론이 가능한 데이터베이스 구축을 위해 다중 퍼지 규칙베이스 방식을 제안한다. 다중 퍼지 규칙베이스는 TMA 자료와 같이 다

양한 미네랄에 의해 의사결정을 해야 하는 시스템인 경우 규칙베이스가 매우 복잡해 질 수 있기 때문에 규칙베이스를 분산시켜서 연관성을 찾아 실시간으로 규칙들을 검색하고 추론하는 방법이 필요하다. 이것을 위해 본 논문에서는 규칙들의 결론부와 다른 규칙의 조건부와의 관계를 트리형태로 구성하여 다 단계로 구성함으로써 기존의 퍼지 규칙베이스보다 복잡한 규칙을 표현하기에 적합하도록 하는 방법을 제안한다. 또한 퍼지 이론을 본 논문에서 불확실성 처리를 위해 적용한 이유는 베이시안 이론 등 불확실성 처리방법들 중[11] 자료처리 및 제어시스템에 적당하기 때문이다.

본 논문에서는 다중 퍼지 규칙베이스 구축을 위해 다음과 같은 두 단계를 갖는데, 제 1 단계 구축 후 제 2 단계를 통해 다중 퍼지 규칙베이스 구축이 완성된다.

【제 1 단계】 대사타입 분류에 의한 퍼지 데이터베이스 구축

제 1단계는 빠른 대사와 느린 대사를 결정짓는 미네랄들의 상관관계 분석으로부터 대사타입에 따른 퍼지 데이터베이스를 구축하는 과정을 말한다. 이것을 위해 본 논문에서는 영양미네랄(Nutritional Mineral) 30종과 독성 미네랄(Toxic Mineral) 8종류 등에 관해 총 38종 미네랄과 영양상태와의 관계, 영양미네랄들의 비(ratio), 독성미네랄들의 비(ratio)와 영양상태와의 관계, 그리고 TMA를 통한 각종질환과 미네랄과의 경향에 관해 분석한 결과 다음 표 1을 얻었다.

표 1에서 문장 분석시 영향을 미치는 주요 미네랄 비율을 나타내는 언어적 변수인 “높다(high)” “낮다(low)” “적당(acceptable)” 등은 다음 그림 4와 같이 퍼지 집합[12]으로 표현하였고, x축은 미네랄 비율이며 y축은 퍼지 멤버십 값을 나타내는데 이러한 수치들은 2.2.2절의 TMA 통계적 분석에 의해 획득하였다.

이와 같은 7종의 미네랄의 비 뿐만 아니라 각 미네랄(Ca, Na, Mg 등 영양미네랄 15종, Hg, Pb 등 독성미네랄 8종, Bi, Ge 등 기타 미네랄 15종) 등에 대해서도 위 그림과 같이 퍼지 집합으로 표현하였다[13].

【제 2 단계】 각 미네랄 양과 비율에 따른 내분비선과의 관계 구축

제 2 단계는 제 1 단계의 대분류 과정을 통해 구축

표 1. 대사타입에 따른 퍼지 규칙베이스 구성(제 1 단계)

분류기준	Adrenal Function			Thyroid Function		대사타입
Ca/P>2.63	부교감신경발달	Na/Mg=4.0	값이 낮으면 기능저하	Ca/K=4.2	값이 높으면 기능저하	느린 대사타입 #1
	부교감신경발달	Na/Mg=4.0	값이 높으면 기능항진	Ca/K=4.2	값이 높으면 기능저하	느린 대사타입 #2
	부교감신경발달	Na/Mg=4.0	값이 낮으면 기능저하	Ca/K=4.2	값이 낮으면 기능항진	느린 대사타입 #3
	부교감신경발달	Na/Mg=4.0	값이 높으면 기능항진	Ca/K=4.2	값이 낮으면 기능항진	느린 대사타입 #4
Ca/P=2.63	적당 (기준값)					
Ca/P<2.63	교감신경 발달	Na/Mg=4.0	값이 높으면 기능항진	Ca/K=4.2	값이 낮으면 기능항진	빠른 대사타입 #1
	교감신경 발달	Na/Mg=4.0	값이 높으면 기능항진	Ca/K=4.2	값이 높으면 기능저하	빠른 대사타입 #2
	교감신경 발달	Na/Mg=4.0	값이 낮으면 기능저하	Ca/K=4.2	값이 낮으면 기능항진	빠른 대사타입 #3
	교감신경 발달	Na/Mg=4.0	값이 낮으면 기능저하	Ca/K=4.2	값이 높으면 기능저하	빠른 대사타입 #4

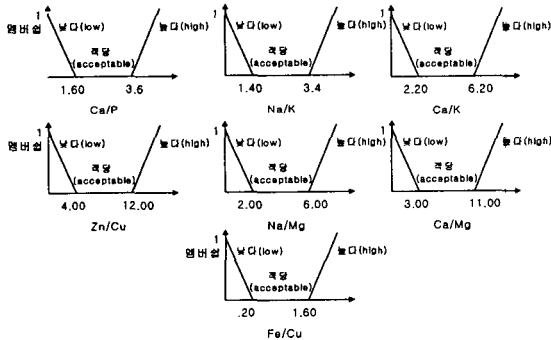


그림 4. 미네랄 비율의 퍼지 집합 표현

한 대사타입에 따른 퍼지 데이터베이스를 세분화하고 또한 내분비선과 미네랄과의 상관관계에 대한 데이터베이스를 구축하는 과정을 말한다. 제 2 단계 분류 및 퍼지 규칙베이스 설계를 위해 그림 5와 같은 분석 결과를 이용한다. 예를 들면, Ca/P, Ca/Mg, Ca/Na, Ca/K, Ca/Fe가 증가하면 Parathyroid가 증가하게 되고, 반대로 Fe/Cu가 증가하면 Parathyroid가 감소함을 의미한다. 내분비선 Thyroid와 Progesteron의 증가요인 미네랄은 각각 Na/Mg Fe/Cu와 Zn/Cu Fe/Cu 이고 감소요인 미네랄은 Ca/P Ca/K와 Na/K Ca/K이다. 또한 Pancreas의 증가요인 미네랄은 Ca/P Ca/Mg Ca/K Ca/Fe Ca/Na이고 감소요인 미네랄은 Zn/Mg Fe/Cu이다. 기타 본 논문에서

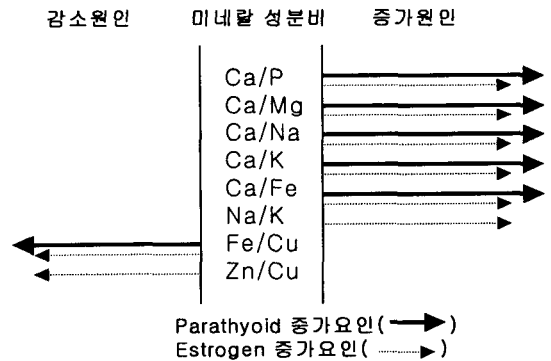


그림 5. 미네랄 비율과 내분비선과의 증감관계(제 2 단계)

다양하게 분석한 내분비선과 미네랄과의 관계 분석 자료를 이용하여 데이터베이스를 구축하였다.

다음 표 2는 미국에서 보내주는 환자의 대사타입에 따른 결과지에 포함되어 상태를 설명하는 각 영어 문장(paragraph)과 특정 미네랄 분포와 상관관계가 있다는 사실로부터 문장을 분석하는 과정의 일부를 나타낸 것이다. 여기서 왼쪽 수치값은 분석을 위해 본 논문에서 임의로 부여한 각 영어문장 ID이다. ID는 총 8자리로 구성되며 앞의 두자리는 문장 번호들의 대분류와 관련된 문장들의 번호이며, 다음 3자리는 미네랄의 순서 예를 들어, 'Na', 'Mg', 'Ca' 순으로 미네랄에 대한 설명이 나오면 'Na'는 '001'이 되고, 'Ca'는 '003'이 된다. 또한 다른 3자리는 미네랄을 설

표 2. 미네랄의 분포에 의한 문장 분석 과정의 예

문장 ID	미네랄 분포
13009014	Fe <= 0.4 And Mg < 3 And CaP < 2.63 And NaMg >= 4 And CaK < 4.2
13010001	Q2 < Mn < 4.2
13010002	Q16 < Mn <= 0.19
13010003	Q2 < Mn < 4.2
13010004	Q008 <= Mn <= 0.01
13011001	Q1 <= Cr <= 0.13
13011002	Q14 <= Cr
13011003	Cr <= 0.02 And Mn < 0.07 And Q03 < Se < 0.11 And CaP > 2.63 And CaK >= 4.2 And NaMg < 4
13012001	Q4 <= Se
13012002	Se < 0.03 And Hg < 0.09
13012003	Q22 <= Se <= 0.31
13012004	Se < 0.03 And Hg < 0.09
13013001	1.36 <= B <= 2.0
13013002	2.0 < B
13013003	2.0 < B
13013004	2.0 < B
13014001	Q01 <= Co <= 0.017 And CaP >= 2.63 And CaK >= 4.2 And NaMg < 4
13014002	Q004 <= Co <= 0.007
13014003	Q05 < Co And CaP >= 2.63 And CaK >= 4.2 And NaMg < 4
13014004	Q01 <= Co And CaP >= 2.63 And NaMg < 4
13014005	Q05 < Co And CaP >= 2.63 And CaK >= 4.2 And NaMg < 4
13014006	Q05 < Co And CaP >= 2.63 And CaK >= 4.2 And NaMg < 4
13014007	Q013 <= Co And CaP >= 2.63 And NaMg < 4
13015001	Q015 <= Mo < 0.025
13015002	Mo = 0.001 And Age > 12
13015003	Q011 <= Mo < 0.015
13015004	Q016 <= Mo <= 0.054 And Age < 12
13015005	Q011 <= Mo < 0.015
13015006	Q002 <= Mo <= 0.003 And 12 < Age And 3000 <= CuMo < 4000 And Hg < 0.09
13015007	Mo <= 0.001 And Age < 12
13016001	S < 3000
13017001	Ge <= 0.005
13017002	Q03 <= Ge And 47 <= Age
13017003	Q03 <= Ge And 47 <= Age
13017004	Q003 <= Ge < 0.006 And Age < 12
13017005	no data
13017006	no data
13017007	Ge > 0.014 And Age > 12 And CaP > 2.63 And CaK < 4.2 And NaMg >= 4
13018001	Q43 <= Ba
13018002	no data
13019001	Q059 < Bi
13020001	Q032 < Rb

명하는 문장들의 종류번호이다. 예를 들어, 공통문장에서 'Ca'을 설명하는 15번째 문장의 ID는 '13003015'가 된다. 이 표에서 본 바와 같이 영어 각 문장은 특정 미네랄의 양, 다른 미네랄과의 관계, 주요 미네랄의 비 등과 매우 밀접한 관계를 갖고 있고, 동일 미네랄 그룹내 ID 문장간 연관성은 전문가의 의견을 바탕으로 연관성을 고려하여 설계하였다. 이러한 2단계 과

정을 모두 거친 후 다중 퍼지 규칙베이스 구축이 완성된다.

논문에서 제안하는 다중 퍼지 규칙베이스에 의한 규칙 형태와 기존의 단순 퍼지 규칙베이스에 의한 규칙형태를 비교하면 다음 예와 같다. 여기서 보는 바와 같이 본 논문에서 제안하는 다중 방법은 퍼지 규칙의 조건부에 조건명제들의 수를 3개 이하로 제

한시켜 데이터베이스에서의 추론 방법을 효율화 했으며 계층적 구조를 이용하여 연관된 다음 규칙과의 관계로부터 규칙의 조건부가 추론되도록 하였다.

【다중 퍼지 규칙베이스에서의 규칙형태의 예】

if Ca/P가 높고(high) Na/Mg가 낮고(low) Ca/K가 높으면(high) then Slow type #1이다.

if Na/K가 낮고(low) Zn/Cu가 높고(high) Fe/Cu가 높으면(high) then Progesteron을 증가시킨다.

if Progesteron이 높고(high) Thyroid가 높으면(high) then Ca/K는 낮다(low).

【단순 퍼지 규칙베이스에서의 규칙형태의 예】

if Ca/P가 높고(high) Na/Mg가 낮고(low) Ca/K가 높고(high) Na/K가 낮고(low) Zn/Cu가 높고(high) Fe/Cu가 높고 Progesteron이 높고(high) Thyroid가 높으면 (high) then Ca/K는 낮다(low).

여기서 데이터베이스에는 Ca/Na, Na/Mg 값 등이 그림 5와 같이 수치적 자료로써 저장되어 있고, 높고 낮음 등을 나타내는 언어적 변수 등은 함수로써 저장되어 있어서 자연어 처리 표현에 효과적으로 이용할 수 있다.

2.4 지능적 추론 시스템 구축 단계

본 논문에서 구축한 지능적 추론 방법을 위한 퍼지 규칙베이스 형태는 다음과 같다. 본 논문에서는 규칙의 조건부에는 미네랄의 비를 이용하여 구축하였고, 결론부에는 대사타입 및 규칙의 조건부에 입력되는 미네랄의 비를 이용하였다. 본 논문에서 현재까지 구축한 규칙은 약 1000개 정도이다.

<Fuzzy Rule Base 구성 형태>

Rule 1: IF x1 is A1 and y1 is B1 and ... z1 is C1 THEN k1 is R1
 Rule 2: IF x2 is A2 and y2 is B2 and ... z2 is C2 THEN k2 is R2
 ...
 Rule n: IF xn is An and yn is Bn and ... zn is Cn THEN kn is Rn

여기서 x1,y1,z1,k1 : 미네랄, 미네랄 비, 내분비선 등
 A1,B1,C1,R1 : 퍼지 집합 (high,low,medium...)

본 논문에서는 다단계 통계분석방법을 사용하여 분류한 자료로부터 구축한 다중 퍼지 규칙베이스에서 퍼지 추론을 효율적으로 하기 위해 분산 목표트리 방식인 TAEMS(Task Analysis Environment Modeling and Simulation)[14]를 적용하였다. TAEMS는 태스크를 분류한 후 분산 목표트리를 구성하여, 서로

다른 목표를 갖는 시스템의 분산목표 수행을 위해 시스템의 목표나 자료들 간의 상호 의존성을 나타냄으로써 효율성을 높이기 위한 분산 인공지능적인 방법이다. TAEMS에는 분산 목표들간의 스케줄링, 목표트리 분해 및 분석 방법 등을 포함하는 새로운 개념으로 미국의 U.of Massachusetts 멀티에이전트 연구실에서 연구되었다.

그림 6은 다중 퍼지 규칙베이스에서 규칙들간의 상호의존성을 TAEMS 방식을 적용해서 나타낸 그림이며, 본 논문에서 구현한 퍼지 규칙베이스의 구성 형태와 FUZZY_INFER()의 추론알고리즘(FUZZY_INFER_I, FUZZY_INFER_II)은 다음과 같다. 여기서 Dataset(Input Facts)는 규칙의 조건부에 입력되는 미네랄 또는 미네랄 성분비 값들을 말한다.

<FUZZY_INFER() 알고리즘>

Enter of input facts to Rule i (초기입력);
 TAEMS에 의한 퍼지 규칙 연관관계 트리구조 생성;
 Rule i 와 연관된 규칙들의 집합 생성; // 규칙들간 Complexity, Scheduling은 TAEMS 방식 적용
 퍼지추론 할 수 있는 입력형태로 변환;

If input fact == conditional part of Rule m(연관된 규칙)
 then FUZZY_INFER_I(RuleBase, input facts);
 // 입력이 규칙의 조건부와 모두 일치하고, 결론부가 단일일 경우
 즉, 규칙들간의 다중 상호 의존관계가 없는 경우 퍼지추론하는 과정

If input fact == some conditional part of Rule n(연관된 규칙)
 then FUZZY_INFER_II(RuleBase, input facts);
 { for(i=0; i<n; i++) // m, n은 TAEMS에 구축된 연관 규칙들의 갯수
 for(j=0; j<=m; j++)
 { x(i) of Rule[i] = K(i) of Rule[j];
 y(i) of Rule[i] = K(k) of Rule[k];
 x(i) of Rule[i] = K(m) of Rule[m];
 }
 }
 // 입력이 규칙의 조건부의 일부와 유사하고, 다른 규칙의 결론부로부터 값을 유도해야 할 경우
 즉, 규칙들간의 다중 상호 의존관계가 갖는 경우 TAEMS 트리구조에서 생성된 연관 규칙들로
 부터 값을 유도하는 과정

퍼지 규칙베이스에 있는 각 규칙들은 다음 그림 6과 같이 상호의존관계를 갖는다. 즉, 하나의 규칙을 구성하는 조건부, 결론부 퍼지명제들은 다른 규칙의 조건부에 값을 전달하는 상호의존관계로 구축되어 있다. 이렇게 하는 이유는 TMA 미네랄 자료가 단순히 몇 개의 미네랄에 의해 분류되지 않고 매우 복잡한 연관관계를 통해 분석되고 측정되기 때문이다.

이 구조에 따라 다중 퍼지 규칙베이스에서 10개의 상위레벨로부터 파생된 인체의 대사타입에 따라 분류·구축된 약 1000개의 규칙들의 각 조건부들을 찾아가는 과정을 예를 들어 나타내면 그림 7과 같다. 만약 ID가 '13002312'인 영어 문장은 출력해야 할 대 분류 'Nutrient Mineral Level' 중 두 번째 미네랄인 'Mg'에 관한 312번째 문장을 말한다. 이 ID인 문장이 추론되기 위해서는 위의 그림에서와 같은 3개의 퍼지 규칙이 연관되어 만족되어야 한다. 즉, 'THEN'이

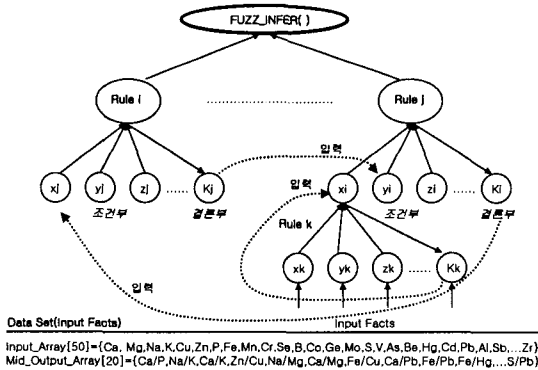


그림 6. 다중 퍼지 규칙베이스에서 규칙들 간의 상호의존관계

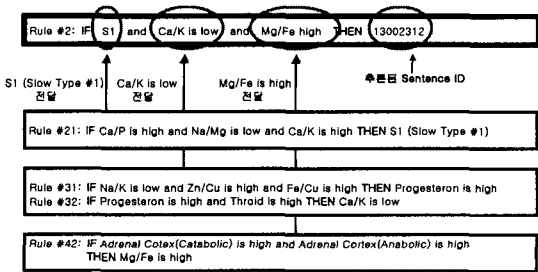


그림 7. 다중 퍼지 규칙베이스에서 정보 추론 과정

하의 결론부인 'S1', 'Ca/K is low', 'Na/Mg is high'가 각각 Rule #2의 조건부에 입력되어 TAEMS방식의 다단계 스케줄링 과정을 거쳐서 해당 ID 문장을 추론하게 된다. 각 규칙의 결론부가 상위레벨의 규칙의

각 조건부 문장들과 일치되는 것들을 찾아가면서 최종적으로 결론부의 문장 ID를 찾게 되며 추론된 해당 ID 문장은 다음과 같다.

【ID: 13002312인 영어문장】

“Magnesium deficiency has been shown to be associated with decreased antibody production. Published studies have revealed that the lymphocytes, which are the body’s defence against foreign invaders, are inhibited when there is a deficiency of Magnesium”

【ID: 13002312 번역된 한국어 문장】

“마그네슘 부족은 항체생성 감소와 관련되어 있다. 발표된 연구자료에 의하면 외부 침입자에 대항하여 인체의 방어작용을 하는 림프구가 마그네슘 부족할 때 억제된다.”

따라서 환자의 다양한 미네랄 수치와 미네랄들간의 비에 따라 지능적으로 추론된 각 ID에 따른 영어 문장들과 해당 문장들이 번역된 한국어 문장들이 데이터베이스로부터 추론되어 레포트형식을 갖추어서 결과지로서 환자와 의사에게 제공된다.

3. TMA 의료정보 시스템(IMES) 설계 및 구현

3.1 IMES 개요

본 논문에서 구현한 TMA 의료정보 시스템은 미

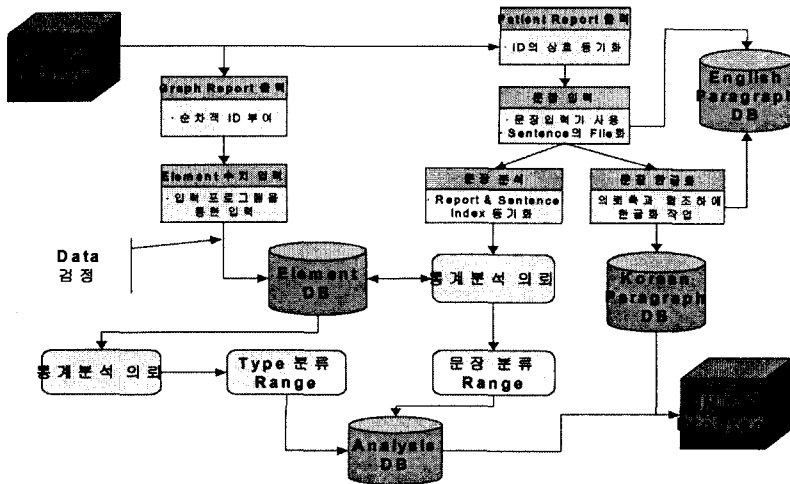


그림 8. IMES 시스템 구축과정

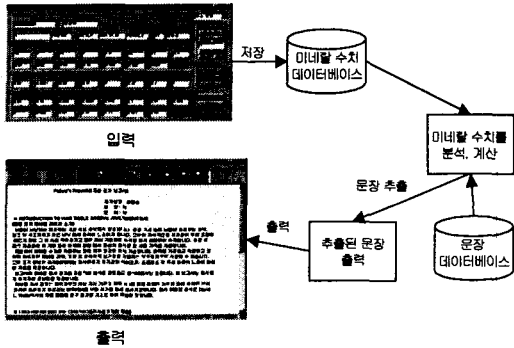


그림 9. IMES 시스템 I/O 및 구성요소간 연관성

국에서 보내오는 암호화된 데이터 파일을 텍스트파일화 하고 이들 미네랄로부터 유도되는 영어문장들과의 상관관계를 분석하여 데이터베이스화한 것으로, 분석된 미네랄 수치만 입력하면 국내에서도 미국에 의존하지 않고 자체적으로 TMA 분석 및 다양한 의학적 예측이 가능 한 국내 최초로 개발된 TMA 관련 한국형 의료 정보 시스템(IMES: Intelligent Medical Expert System)이다. IMES에는 미네랄로부터 분석된 다양한 형태의 한글화된 한국형 의료 정보 데이터베이스가 구축되어 있다. 따라서 IMES는 입력된 환자의 미네랄 수치에 의해 데이터베이스와 연동하여 각각의 조건을 검색하여 현재 환자의 상태 및 앞으로 발생할 상황들에 대한 예측을 다양하게 제공해준다.

본 논문에서 구현한 IMES의 구성절차는 그림 8과 같다. 본 시스템은 MS Visual C++과 MS ACCESS를 이용하여 윈도우 환경에서 구현하였다. IMES 시스템 I/O 및 구성 요소들간의 연관성은 그림 9와 같다. IMES 시스템은 미네랄 수치입력 인터페이스와

포트를 출력해주는 부분으로 크게 나누어진다. 미네랄 수치를 입력하면 입력한 수치는 수치 데이터베이스에 저장되고 그 입력된 수치값으로 문장 데이터베이스에서 문장을 추출하게 된다.

IMES 프로그램의 화면 및 기능은 다른 프로그램들과 유사하나 이 시스템의 장점은 한국적 TMA 데이터베이스를 구축하고 실행한 점이 특징이다.

데이터베이스는 다음과 같은 4개의 테이블과 필드들로 구성되어 있다. TMA 테이블은 총 78개의 필드로 구성되어 있으며, 환자에 대한 미네랄 수치와 그 환자의 인적사항과 기타 사항을 관리하는 테이블이다. Administrator Table은 5개의 필드로 구성되어 있으며, 병원과 진료과, 의사 그리고 담당자를 관리하는 테이블이다. IMES 시스템에서 관리자 설정과 입력에 사용되는 테이블이다. 관리자 설정에서 입력되어지는 사항에 의해 자동으로 입력창에 관리자를 탐색해서 출력해 준다. ID_Management Table은 3개의 필드로 구성되어 있으며, 환자의 재검사여부를 관리하는 테이블이며, Sent_Name 은 2개의 테이블로 구성되어 있으며 문장을 관리하는 테이블로, 문장 수정에 사용되며, 문장 탐색을 쉽게하기 위한 테이블이다.

3.2 IMES 시스템의 주요기능

초기 입력화면은 데이터베이스와 연결되어 있으며, 화면에 보이는 환자의 신상정보와 미네랄수치는 데이터베이스에 저장되어 있는 내용과 일치하도록 설계하였다. 새로운 데이터의 입력은 현재의 화면에서 입력을 하거나 초기화 후 입력을 하고 데이터베이스에 저장하거나 단지 레포트만을 출력할 수 있도록

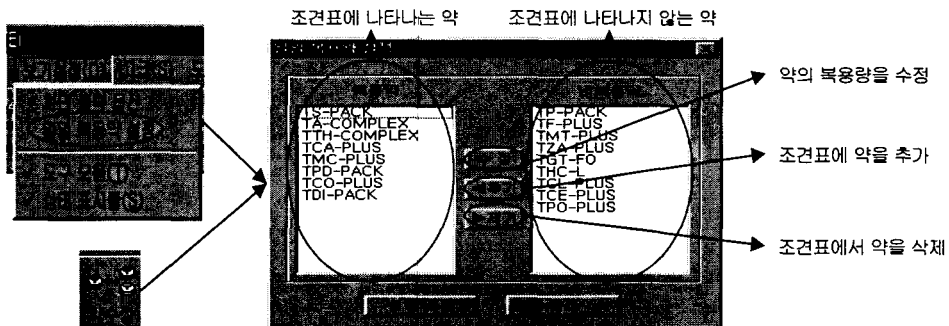


그림 10. 조건표 수정화면

록 하였다.

검색은 여러 가지 조건 중 검색하고자 하는 조건을 입력하여 그 조건에 맞는 내용을 데이터베이스에서 찾아 사용자가 쉽게 검색할 수 있도록 하는 기능을 추가하여 설계하였다. 검색의 예를 들면, 검색 조건이 SLOW#1에 남자만 검색하고 싶을 때에는 밑에 그림과 같이 '성별'과 'TYPE'을 조건에 맞게 설정하고 검색을 실행하면 그 조건에 맞는 데이터가 화면에 보여지게 된다.

레포트 출력기능은 입력화면을 실행한 후 사용이 가능하게 되는데, 이는 어떤 환자를 선택하여 그 환자의 미네랄 수치값을 입력해야 레포트를 보여 주기 때문이다. 레포트는 한글과 영어의 옵션에 따라 출력하도록 하였다.

출력결과로부터 조건표의 수정은 약의 개수나 약의 내용을 변경해야 할 경우 쓰여지게 되는데, 프로그램에서 그림과 같이 보기설정에서 일일 복용약 설정 실행에 의해 일일 복용약 설정 창에서 수정이 가능하도록 하였다. 이는 환자에게 제공하는 알약의 품목 변경시 관리자가 반영할 수 있도록 하기 위함이다.

3.3 IMES 시스템 분석

3.3.1 분류방법에서의 오분류율 처리방법

분류방법의 적절성을 확인하는 절차 중의 하나는 오분류율(misclassification probability)을 계산하는 것이다. 본 연구에서 오분류율의 계산은 표본을 이용하여 다음과 같은 식 (1)에 의해 계산하였다.

$$ARE(actual\ error\ rate) = p_1 \int_{R_1} f_1(x) dx + p_2 \int_{R_2} f_2(x) dx \quad (1)$$

여기에서 R_1 과 R_2 는 표본의 크기에 의해 결정되는 분류지역(classification regions)을 의미한다. 이것을 사용하는 이유는 위 식(1)의 방법은 미지의 밀도함수(density function), $f_1(x)$ 와 $f_2(x)$ 의 함수로 구성되어 있으므로 표본정보만을 갖고 있는 본 논문에서는 모집단의 분포에 종속되지 않고 오분류율을 계산할 수 있어야 하기 때문에 APER(AParent Error Rate)을 이용하였다.

그림 11을 이용하여 $APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$ 로 계산되며 이는 오분류된 비율을 나타낸다. 본 논문에서 오분류율 측정 결과 8가지 대사타입(빠른 대사, 느린 대사)에 대해서는 오분류율이 거의 0에 가까웠으나 표본의 크기가 작은 미네랄간의 관계 등에서는 그림 12와 같은 오분류율이 발생하였는데, 표본크기가 200 이상일 경우 오분류율은 거의 0값을 나타내었다.

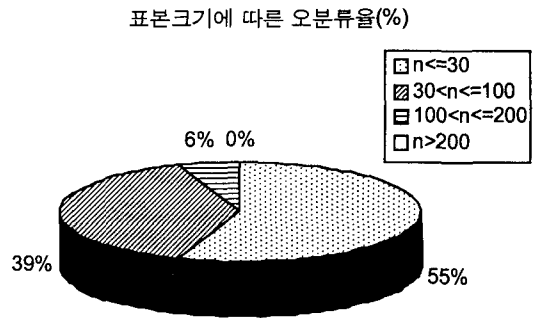


그림 12. 오분류율

		Predicted membership		
		π_1	π_2	
Actual membership	π_1	n_{1C}	n_{1M} $= n_1 - n_{1C}$	n_1
	π_2	n_{2M} $= n_2 - n_{2C}$	n_{2C}	n_2

여기에서 n_{1C} = 실제 π_1 에 속하면서 π_1 에 분류된 수
 n_{1M} = 실제 π_1 에 속하면서 π_2 로 잘못 분류된 수
 n_{2C} = 실제 π_2 에 속하면서 π_2 에 분류된 수
 n_{2M} = 실제 π_2 에 속하면서 π_1 으로 잘못 분류된 수

그림 11. 오분류율 측정방법

3.3.2 시스템의 효율성 및 활용성

미국 TMA는 철저한 보안전략으로 인해 대략적인 구조만 알 수 있어서 본 논문에서 구현한 한국의 IMES 시스템과의 차이점을 그림 13과 같이 나타낼 수 있다.

그림 13과 같이 미국 TMA는 35만건의 사례별로 구축된 단순 RDB로 구성되어 있으며 검사한 미네랄 수치입력시 단순한 수치에 의한 패턴매칭에 의해 자료추론을 하고 있으나, IMES는 미국 시스템에 비해 약 2천건의 자료로부터 통계적으로 분류하여 지능형 다중 퍼지 규칙베이스를 구축한 것이 주요 특징이다.

다중 퍼지규칙 베이스의 효율성을 분석한 결과 표 3에서와 같이 다중 퍼지 규칙베이스를 구축하였을 경우 단일 퍼지 규칙베이스를 구축하는 것보다 시스템의 효율성이 향상되었다. 이것은 표본크기가 2000이고 오분류율이 0이라는 조건하에서 측정된 결과이다.

평균 처리속도는 입력부분에서 미네랄 수치를 입력하여 영어 및 번역된 문장을 찾아내는 속도를 측정 한 것으로, 기존의 방법은 규칙의 입력부분의 비교횟수가 많아 처리속도가 느린 반면 제안한 방법은 규칙의 조건부 비교횟수를 줄여주어 결론부를 재입력함

으로써 평균 처리속도가 약 5배 정도 향상되었다.

결과의 정확성은 2001. 7~2002. 2까지 7개월간 국내에 보관된 환자의 임상자료 약 2000건에 대해 측정 한 것으로, 시스템에서 전체적으로 한국어로 번역되어 출력되는 최종 결과지가 미국에서 보내오는 영문으로 된 결과지와 일치하는 정도를 측정 한 것이다. 환자 한건당 결과지에 포함해서 출력해야 할 총 50개의 문장들 중 단일 퍼지규칙 베이스로 구축한 경우는 평균 34개 정도의 문장만이 일치하였으나 다중 퍼지 규칙 베이스로 시스템을 구축한 경우 평균적으로 46개 정도 일치하였고 4개 정도 비슷하거나 잘못된 문장을 출력하였는데, 이것은 제한한 방법이 기존에 비해 약 24% 정도 정확성이 향상된 수치이다. 현재 IMES 시스템의 데이터베이스에 저장된 총 영문 종류는 각 대사 및 미네랄 타입별로 약 700개 정도이다. 이것으로부터 약 35만건의 자료로부터 추론된 것보다 다중 퍼지규칙 베이스에 의한 2000건에 의한 추론 결과가 더 우수함을 알 수 있었다.

IMES 시스템의 활용성 평가를 위해 다음 그림 14와 같이 IMES 시스템을 사용한 경우와 사용안한 경우로 나누어서 분석을 하였다. 분석을 위해 실제 시스템 사용자를 대상으로 2002. 3월부터 6월까지 임상 실험 결과를 중심으로 분석을 하였다. 환자 1건당 평균 처리 시간이 2002. 3월 이전에는 약 50분 정도 소요되었는데, 측정결과 IMES를 사용할 경우 약 7분 정도가 소요되어 IMES를 사용하지 않은 경우보다 86% 일의능률이 향상되었음을 알 수 있었다.

이상으로부터 사용자들의 IMES 시스템에 대한 활용도 대비 일의능률에 대한 만족도(활용도를 100으로 했을 때의 일의 능률 값의 비)를 측정하면 다음과 같다. 평균 환자 1건당 출력 속도 포함하여 3분 정도를 가장 적당한 값이라 하면, 여기서 만족도는 기본값을 3분으로 하고 추가된 시간에 대해서 분석

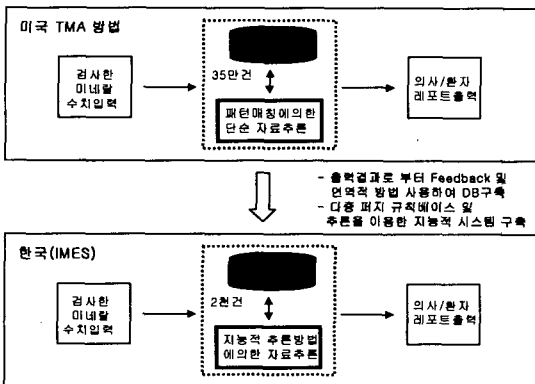


그림 13. 미국 TMA방법과 한국 IMES 시스템 비교

표 3. IMES 시스템의 효율성

비교항목 \ 구분	다중 퍼지규칙 베이스 (제안한 방법)	단일 퍼지규칙 베이스 (기존의 방법)
평균 처리속도(sec)	0.1sec	0.5sec
결과의 정확성(%)	92%	68%

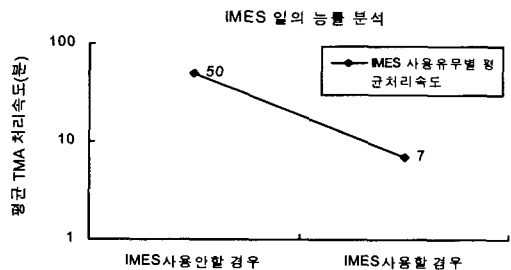


그림 14. IMES 시스템의 일의능률 분석

하였다. 따라서 50분, 7분에 대한 만족도를 측정하기 위해 본 논문에서 식 (2)와 같은 방법을 제안하였는데, 이 식에서 50은 기존에 IMES가 없는 상태에서의 처리속도(만족도 0)이며 3은 최소 기본 처리속도(만족도 100)를 의미한다.

$$IMES \text{ 만족도} = 100 * (Max - \text{소요시간}) / (Max - Min) \quad (2)$$

여기서 Max=50, Min=3 (3 ≤ 소요시간 ≤ 50) 을 나타낸다. 측정 결과 IMES를 사용안할 경우와 사용할 경우 각각 6%, 92% 사용자 만족도를 나타내었다.

4. 결 론

본 논문은 TMA(모발분석)를 위한 한국형 의료정보 시스템 구축에 관한 것으로 기존의 국내 TMA 방법에서의 다음과 같은 문제점을 해결하기 위해 연구하였다.

첫째, TMA 분석 기기는 있으나 분석결과를 전문적으로 해석하여 환자와 의사들에게 제공할 수 있는 프로그램 및 의료정보 데이터베이스가 없다. 둘째, 미국에서 보내오는 TMA 검사결과가 영문이라 한글화된 결과지가 필요하며 암호화된 파일형태이므로 시스템적 활용도가 낮다. 또한 한국적 특성이 아닌 서구식 데이터베이스에 의한 결과물이므로 검사결과와의 신뢰성이 부족할 수밖에 없으며 한국적 임상결과에 대한 검증자료가 없고 검사비용으로 인한 외화낭비가 초래된다.

본 논문에서는 이러한 문제점 개선을 위해 몇가지 방법을 제안하였다. 첫째, 통계적 TMA 자료분석을 위해 다단계 분류방법을 활용하여 다중 영향 요인들에 대하여 분류하였다. 둘째, 다중 퍼지 규칙베이스를 구축하여 지능적인 퍼지추론 방법을 이용하여 미네랄과 관련문장들을 추론하였다. 셋째, 한국형 TMA 의료정보 데이터베이스를 구축하였고 이를 위해 한글화 작업을 병행하였다. 이러한 방법에 의해 TMA 관련 의료 정보 시스템(IMES: Intelligent Medical Expert System)을 국내 최초로 구축하였다.

현재 개발된 IMES 시스템은 국내 임상실험한 자료가 미국 TEI에 비해 작았지만(약 2000 여건 정도) 업무효율이 IMES 시스템을 사용하지 않았을 경우보다 86%이상 향상되었고, 사용자 만족도가 92%정

도 되었고, 시간당 처리하는 평균 처리건수도 기존에 비해 약 7.5배이상 증가하게 되었으며 결과의 정확도도 약 92%정도로 측정되었고, 다중 규칙베이스 시스템의 평균 처리속도도 약 5배이상 증가하였다.

앞으로 계속 연구해야 할 과제는 많은 임상결과를 바탕으로 시스템을 개선하는 일이고 지능적 추론 방법과 다단계 통계적 분류방법과의 연관성에 관해 좀 더 연구하여 시스템의 오류율을 최소화해야 하는 일이며 웹상에서 서비스할 수 있도록 시스템을 확장하는 일이다.

참 고 문 헌

- [1] D.L.Watts, Trace Elements and Other Essential Nutrients—Clinical Application of Tissue Mineral Analysis, 2nd Writer’s BLOCK edition, 1997.
- [2] W.A.Baumgartner and V.A.Hill and W.H.Blahd, “Hair Analysis for Drugs of Abuse”. Journal of Forensic Science Vol 34. No 6. 1989.
- [3] Vir etal, “Serum and hair concentrations of Copper During Pregnancy”, American Journal of Clinical Nutrition Vol. 34, 1981.
- [4] <http://www.traceelements.com>.
- [5] <http://www.teikorea.com>.
- [6] <http://www.medinex.com>.
- [7] 장수익, 김경곤, 이복기, 김형준, 유수현, 강희철, 윤방부, “당뇨환자에서의 모발내 미네랄의 함량연구”, 최신의학 45권 5호 별책. 2002.
- [8] S.R.Safavian and D.Landgrebe, “A survey of decision tree classifier methodology”, IEEE Transactions on System, Man, and Cybernetics., Vol.21, No.3, pp.660-674, 1991.
- [9] D.E.Brown and V.Corruble and C.L.Pittard, “A Comparison of decision tree classifiers with backpropagation neural networks for multinomial classification problems”, Pattern Recognition, Vol. 26, No. 6, 953-961, 1993.
- [10] R.A.Johnson and D.W.Wichern, Applied Multivariate Statistical Analysis 3rd edition, Prentice Hall, 1992.
- [11] I.B.Turksen and Z.Zhong, “An Approximate

Analogical Reasoning Schema based on Similarity Measures and Interval Valued Fussy Sets,” *Fuzzy Sets and Systems*, North-Holland, Vol. 34, pp.323-346, 1990.

- [12] C.C.Lee, “Fuzzy Logic in Control Systems: Fuzzy Logic Controller, Part I”, *IEEE Transactions on System, Man, and Cybernetics.*, Vol.20, No.2, pp.404-418, 1990.
- [13] Y.I.Cho, “An Effective Hybrid Methodology of Neuro-Fuzzy Systems”, *International ICSC Congress Computational Intelligent Methods and Applications*, USA, NY, pp.69-74, June 22-25, 1999.
- [14] J.R.Graham and K.S.Decker, “Towards a Distributed Environment - Centered Agent Framework”, *Umass Technical Report*, 1998.



조 영 임

1987년 고려대학교 생물학과 졸업
 1988년 고려대학교 컴퓨터학과 졸업
 1990년 고려대학교 컴퓨터학과 석사(인공지능전공)
 1994년 고려대학교 컴퓨터학과 박사(인공지능전공)
 1996년 현재 평택대학교 컴퓨터 과학과 조교수
 1995~1996년 삼성전자 멀티미디어 연구소 선임연구원
 1999~2000년 University of Massachusetts, at Amherst, Dept. of Computer Science, Post-doc
 2000~현재 International Conference on Parallel and Distributed Systems(ICPADS), PC member
 2001~현재 한국정보과학회 학회지 편집위원
 2001~현재 한국공학교육학회 논문지 편집위원
 2002~현재 한국퍼지 및 지능시스템학회 편집위원
 2001~현재 (주)한국티아아이 기술자문 교수
 2002~현재 (주)크레비스 기술자문 교수

교신저자

조 영 임 (450-701) 경기도 평택시 용이동 111번지 평택대학교 컴퓨터학과