

다각도 정보융합 방법을 이용한 지능형 에이전트 시스템

이 현 숙 *

An Intelligent Agent System using Multi-View Information Fusion

Hyun-Sook Rhee *

요 약

본 논문에서는 데이터마이닝모듈과 정보융합모듈을 핵심구성요소로 가지는 지능형에이전트 시스템을 설계하고 다각도 정보를 융합하여 진단전문가시스템으로 활용할 수 있는 가능성을 제시한다. 데이터마이닝모듈에서는 퍼지신경망 OFUN-NET에 의하여 다각도의 데이터를 분석하고 퍼지 클러스터 정보를 지식베이스로 구축한다. 정보융합모듈과 응용모듈에서는 가능성정도로 제공되는 진단결과와 불확실 결정상태나 비대칭의 발견과 같은 전문가의 진단에 유용한 정보를 제공해 주고 있다. 또한 DDSM 벤치마크 데이터베이스로부터 획득한 디지털 유방 x선 영상의 BI-RADS 기반 특징데이터를 가지고 실험한 결과는 기존의 방법보다 높은 분류 정확도를 보여주면서 컴퓨터보조진단시스템으로서의 가능성을 보여주고 있다.

▶ Keywords : 지능형 에이전트시스템, 컴퓨터보조진단시스템, 데이터마이닝, 정보융합, 특징데이터

Abstract

In this paper, we design an intelligent agent system with the data mining module and information fusion module as the core components of the system and investigate the possibility for the medical expert system. In the data mining module, fuzzy neural network, OFUN-NET analyzes multi-view data and produces fuzzy cluster knowledge base. In the information fusion module and application module, they serve the diagnosis result with possibility degree and useful information for diagnosis, such as uncertainty decision status or detection of asymmetry. We also present the experiment results on the BI-RADS-based feature data set selected form DDSM benchmark database. They show higher classification accuracy than conventional methods and the feasibility of the system as a computer aided diagnosis system.

•제1저자 : 이현숙 •교신저자 : 이현숙

•투고일 : 2014. 11. 5, 심사일 : 2014. 11. 12, 게재확정일 : 2014. 11. 18.

*동양미래대학교 소프트웨어정보과(Dept. of Software Engineering, Dongyang Mirae University)

※ 본 논문은 동양미래대학교 2014학년도 교내 학술연구과제 지원사업에 의하여 연구되었음.

▶ Keywords : intelligent agent system, computer-aided diagnosis system, data mining, information fusion, feature data

I. 서 론

최근 컴퓨터가 처리하는 데이터의 양이 방대해지면서 이를 효과적으로 처리하는 빅데이터기술과 인간을 대신하여 특정한 일을 처리해 주는 에이전트시스템에 대한 연구가 활발히 진행되고 있다[1-2]. 이러한 연구는 최적의 맞춤형 정보를 찾기 위한 인터넷 사용자, 패킷의 지연이 가장 적은 라우터를 배치하고자하는 네트워크 설계자, 유체의 저항을 최소로 하는 항공기를 설계하고자하는 항공 공학자, 오진을 적게 하고자하는 의사, 컴퓨터시스템의 바이러스를 탐지하고자하는 시스템 운영자, 증권정보를 분석하여 투자하는 애널리스트 등의 전문가의 의사결정이나 특정 작업이 포함되는 다양한 분야에 활용되고 있다. 에이전트시스템의 핵심 기술은 이미 알려진 사전 정보를 학습하여 최적의 모델을 만들어 비슷한 상황에 대하여 적절한 정보를 제공해 주는 것이다. 이러한 학습기능은 특정 문제영역에 정보를 습득하고 새로운 환경에 반응하는 정보를 제공해 주어야하는 에이전트시스템의 필수 기능이다. 어떤 관점에서는 학습만이 데이터의 무제한적인 증가로 원하는 정보에 접근하기 어려운 현재 정보시스템의 문제를 해결하기 위한 유일한 방법일 것이다. 이와 같은 에이전트 시스템은 처리해야 할 정보의 양이 방대할 뿐 아니라 계속하여 새로운 정보가 발생하므로 적응 능력을 가진 학습방법을 포함하는 최적화과정이 필수적으로 요구되어 인공지능연구의 핵심 분야가 되어 왔다[3].

또한 에이전트시스템에 대한 연구는 의료영상의 해석과 분석에 적용되어 실효성 있는 컴퓨터보조진단시스템을 구성하는데 기여하고 있다[4]. 특히 유방암은 여성에게 가장 흔하게 발병하는 암으로 <유방암백서>에 따르면 국내여성의 유방암 발병률은 매년 가파르게 상승하여 최근 15년 사이 4배 이상 켜중 뛰었다. 그러나 다행스럽게도 <2013 보건복지 통계연보>에 따르면 의학적인 완치기준인 유방암의 5년 생존율은 91%에 이른다. 특히 유방암은 50%이상 정기검진을 통해 발견되는 것으로 알려져 있으며, 유방촬영술에 의해 획득한 맘

모그램은 진단의 중요한 단서가 되고 있다. 하지만 다각도의 의료영상의 처리과정에서 발생하는 전문가의 판독오류와 해석과정의 불일치는 오진율을 높이고 있다. 이에 의료장비의 발전과 지능형에이전트시스템 연구의 진보는 다각도에서 수집되는 디지털영상데이터를 처리하여 유방암을 조기 발견함으로써 궁극적으로 생존율을 높이는데 기여할 수 있게 된다.

본 논문에서는 다각도의 의료 영상데이터를 처리하기 위한 지능형에이전트시스템의 구성하고 공개적으로 널리 알려진 세계에서 가장 규모가 큰 Digital Database for Screening Mammography(DDSM)[5]의 데이터를 사용하여 제안된 시스템의 실효성을 검토하고자한다. 이를 위하여 본 논문에서는 데이터마이닝모듈과 정보융합모듈을 핵심구성요소로 가지는 지능형에이전트 시스템을 설계하고 다각도 정보를 융합하여 진단전문가시스템으로 활용할 수 있는 가능성을 제시한다. 데이터마이닝모듈에서는 퍼지신경망 OFUN-NET에 의하여 다각도의 데이터를 분석하고 퍼지 클러스터 정보를 지식베이스로 구축한다. 정보융합모듈과 응용모듈에서는 가능성정도로 제공되는 진단결과와 불확실 결정상태나 비대칭의 발견과 같은 전문가의 진단에 유용한 정보를 제공해 주고 있다. 또한 DDSM 벤치마크 데이터베이스로부터 획득한 2가지 뷰의 디지털 유방 x선 영상의 BI-RADS 기반 특징데이터를 가지고 실험한 결과는 기존의 방법보다 높은 분류 정확도를 보여주면서 제안된 시스템의 컴퓨터보조진단시스템으로서의 가능성을 확인한다. 물론 보다 견고한 시스템으로 발전하기 위해서는 유방암진단전문가의 전문지식을 필요로 하기 때문에 본 연구에서는 DDSM의 BI-RADS 기반 특징데이터를 그대로 추출하여 인코딩하여 사용한다.

본 논문의 구성은 다음과 같다. II장에서는 그 동안 연구되어 온 관련연구로서 에이전트시스템의 중요한 기능인 데이터 마이닝기법을 소개한다. 또한 본 논문의 연구대상인 유방암진단을 위한 컴퓨터 보조진단 시스템에 대하여 고찰함으로써 연구 배경을 서술하고자한다. III장에서는 본 논문에서 제안하는 지능형에이전트시스템의 구성도를 보여주며 각 모듈의 기능과 퍼지 성능 척도 활용방법을 기술한다. IV장에서는 DDSM 공개데이터베이스로부터 획득한 입력데이터를 준비하여 핵심

모듈에 적용하고 기존의 분류알고리즘과 비교함으로써 진단 시스템으로서의 가능성을 확인하고자 한다. 마지막으로 V장에서는 본 논문의 연구내용을 요약하고 앞으로의 연구방향을 제시한다.

II. 관련 연구

본 장에서는 그 동안 연구되어온 관련연구로서 에이전트시스템의 중요한 기능인 데이터마이닝기법을 소개한다. 데이터마이닝은 많은 데이터로부터 활용가능한 데이터를 추출하는 작업으로 광범위한 데이터 처리 및 분석방법이 관련되지만 본 연구의 대상인 퍼지클러스터링을 중심으로 요약한다. 또한 본 논문의 연구대상인 유방암진단을 위한 컴퓨터 보조진단 시스템에 대하여 고찰함으로써 연구배경을 서술하고자한다.

1. 퍼지 클러스터링 기반 데이터마이닝 기법

데이터 마이닝이란 대용량의 데이터베이스에서 이전에 알려지지 않고(unknown), 유효하고(valid), 활용가능한(actionable) 정보를 꺼내는 작업을 말한다. 일반적으로 데이터 마이닝은 실제적인 마이닝(Mining)과 특징, 클래스 및 패턴 등을 발견하는 과정을 말한다. 이러한 데이터 마이닝은 KDD(Knowledge Discovery in Database)와 동격의 과정으로 볼 수 있으며 그 처리주기는 다음과 같다[6]. 첫째로, 수많은 데이터베이스, 혹은 인터넷의 자료 중에서 자신이 처리하고자 하는 목적을 결정(Objective Determination)하는 과정이다. 두 번째 과정은 선택된 데이터를 준비하는 과정(Data Preparation)이다. 이 과정은 데이터 선택(Data Selection), 데이터 선처리(Data Preprocessing), 데이터 변환(Data Transformation)의 세 개의 세부과정으로 이루어져 있으며 적용하려는 영역데이터에 따른 여러 처리과정이 필요하다. 세 번째 과정은 이전의 과정에서 변환된 데이터를 가지고 실제로 마이닝을 처리하는 과정이다. 그리고 네 번째 과정은 마이닝된 결과를 분석하는 과정(Analysis of Results)이다. 이 과정에서 분석된 결과에 대하여데이터 가시화(Data Visualization) 작업을 수행한다. 마지막으로 마이닝된 결과를 실제 회사에서의 업무나 사용자가 원하는 작업으로 흡수하는 과정이다. 이러한 전체적인 처리 주기 중 첫 번째와 두 번째는 문제영역의 데이터에 종속적인 전처리과정이고 네 번째는 결과를 분석하여 사용하는 후처리과정에 해당된다. 그러므로 그 동안 연구의 초점은 세 번째 과정인 실제적인 마이닝기법과 알고리즘에 모아졌다.

마이닝기법의 대표적인 예는 군집화에 의한 비교사 학습(Unsupervised learning) 방법으로 속성이 비슷한 것들끼리 묶어 나누는 것으로, 대상 데이터가 너무 많아 전체를 파악하기 어려울 때, 부분을 살펴 전체의 윤곽을 잡도록 해준다. 이와 같은 군집화기법은 전문가시스템, 패턴인식, 영상처리, 음성인식 등 학습이 필요한 모든 지능형 에이전트시스템의 기본 단계이다. 특히 요즘 인터넷을 통한 검색엔진의 탐색과정에도 활용되어 비슷한 정보를 그룹핑하여 보여주고 있다. 그러나 기존의 클러스터링 방법론은 대부분 hard partitioning에 의한 방법으로 주어진 데이터 상호간의 경계가 명확하다는 가정에서 각 패턴을 하나의 클래스에 소속시키는 방법이다. 그러나 이 모델은 우리가 다루는 데이터의 경계가 대부분 불명확하므로 실제 데이터 상호간의 군집성을 묘사하기에 부적절하며 주어진 데이터 분포의 성질을 잃어버리는 결과를 가져온다. 이에 기존의 군집화기법에 퍼지이론을 적용한 Fuzzy C-Means(FCM) 알고리즘을 비롯하여 FCM의 통계적인 속성을 보완하여 신경망구조와 결합한 OFUN-NET 등 다양한 기법이 연구되어 여러 도메인에 적용되어왔다[7-8]. 특히 x-선 의료 영상의 진단 시스템의 경우 결정이 어려운 환자의 영상에 맞는 분류정보를 제공하는 것이 중요하다. 이에 퍼지이론을 적용하여 각 데이터의 클래스 소속값을 함께 학습함으로써 보다 정확하게 데이터를 표현하고 정확한 분류정보를 얻을 수 있도록 하는 퍼지 클러스터링 기법이 도입되었다. 대표적인 클러스터 분석 방법으로 널리 알려진 FCM(Fuzzy C-Means) 알고리즘은 Bezdek에 의하여 개발되어 최적 분할, 패턴분류 및 영상 처리 등의 여러 분야에 활용되었다[7]. FCM 알고리즘과 같은 퍼지 클러스터링 방법은 분류에 의하여 기본적인 지식베이스를 만들어 준다. 이때 의사결정을 위해 필요한 정보를 가능성으로 표시해 주기 때문에 진단과 같은 문제영역에서 더욱 유용하게 활용될 수 있다. 또한 퍼지정보를 처리하는 방법에 의하여 생성한 정보의 불확실성의 정도나 서로 다른 뷰로부터 생성된 정보의 비대칭여부를 측정하여 활용할 수도 있다. 이렇게 퍼지클러스터링 연구가 진행됨에 따라 신경망 분야의 학습 능력을 보강하기 위하여 FCM 알고리즘의 목적함수를 비교사 신경망에 결합시킨 퍼지 신경망에 대한 연구도 진행되었다. 제안된 퍼지 신경망의 학습 알고리즘은 클러스터링의 결과가 만들어 내는 오류 값을 요약하는 퍼지 함수를 설정한 후 그 값이 최소가 되도록 학습의 방향을 유도하는 메카니즘에 의해 진행된다. 또한 제안된 방법은 입력과 출력 사이의 관계를 기술하기 어려운 경우도 쉽게 처리하는 비교사 학습신경망의 장점도 함께 가지고 있다[9].

2. 컴퓨터보조진단시스템

에이전트시스템의 응용분야는 인간이 관여하는 정보통신의 전체를 포괄할 정도로 광범위하다. 본 논문에서 초점을 맞추고 있는 퍼지클러스터링에 의한 학습능력을 활용하는 발전 가능한 분야는 다양할 것이다. 특히 의료장비의 발전과 패턴인식분야의 연구결과를 접목하여 컴퓨터를 활용한 디지털 유방 X선 영상진단시스템과 같은 의료정보화분야에서 괄목할만한 의미있는 결과를 도출하고 있다.

컴퓨터보조진단시스템의 연구가 진행되면서 환부조직의 구조와 모양이 다양해짐에 따라 분류가 어려워지면서 Wu 등은 앙상블 네트워크 모델을 제안하여 분류의 정확도를 높이는 연구를 하였다[10]. 제안된 모델은 종괴 환부를 분류하기 위하여 가중치 평균과 퍼셉트론 평균의 두 가지 알고리즘을 사용하였다.

또한 미세석회 환부가 암을 진단하는 중요한 단서가 됨이 알려지면서 이를 분류하기 위한 연구가 진행되었다. [11]에서는 정해진 특징을 사용하여 특징 추출 과정에서 분류에 영향을 주는 특징을 찾아가는 과정을 통하여 역전파 신경망기반의 컴퓨터 보조진단시스템을 구성하였다. 널리 사용되는 14가지의 특징을 추출한 후에 하나의 특징을 가지고 분류를 시작한 후 다른 특징을 첨가함으로써 분류율이 증가한다면 그 특징을 선택하고 그렇지 않으면 다음 특징을 고려하는 방식으로 전체 8192 개의 조합에 대하여 상당한 양의 실험을 계속 하였다. 이러한 실험의 결과로 14 가지의 특징을 고찰하여 신경망의 구조를 결정하고 미세석회패턴을 탐지하는 퍼지 논리 기반 특징결합방법을 제안하였다. 제안된 시스템은 특징을 선택하고 평가하는 과정에서부터 분류하는 과정까지 퍼지 신경망의 원리를 체계적으로 적용하였으나 20개의 양성환부와 20개의 악성환부 모두 40개의 미세석회환부가 학습에 사용되어 그 데이터가 작기 때문에 연구의 신뢰도가 높지 않은 것으로 판단된다. [12]에서 여러 가지 특징요소들을 결합하여 Linear discriminant analysis(LDA) 모델에 적용하여 292 DDSM 경우에 대하여 성능을 분석하였다. 다양한 특징결합의 시도로부터 형태학적인 정보를 가지는 영상처리 데이터와 x-선 영상 판독 전문가에 의한 데이터가 적절히 결합되었을 때 성능이 우수한 것으로 알려졌다. 또한 최근 DWT(Discrete Wavelet Transformation)를 사용하여 영상으로부터 잡음을 제거한 후 ROI(region of interest)를 선택하여 20개의 GLCM(Gray Level Co-occurrence Matrix) 특징데이터를 가지고 여러 분류알고리즘에 적용한 분류율을 비교한 연구도 진행되었다[13]. 진단오류를 줄이기

위하여 두가지 관점에서의 맘모영상의 분석을 통하여 얻은 정보를 융합하는 방법에 대하여 연구하면서 [14]에서는 DDSM 데이터셋으로부터 선택된 152 케이스의 두가지 관점의 영상으로부터 304개의 ROI를 가지고 실험한 결과 하나의 관점에서의 영상보다 높은 분류정확도를 나타낸 것을 확인하였다. 또한 [15]는 여러 분류기에 의한 다중 에이전트 방법을 이용한 다각도 정보 융합 알고리즘을 제안하여 DDSM으로부터의 64 케이스에 대한 128개의 ROI를 가지고 실험하였다. 그 결과 분류정확도는 높고 FPR(False Positive Rate)는 낮추는 실험 결과로서 그 방법의 타당성을 입증하였다.

III. 지능형 에이전트 시스템

에이전트시스템은 그동안 인공지능 연구결과를 실현한 것으로 감정의 개입 없는 객관적 데이터를 바탕으로 인간을 대신하여 특정영역의 작업을 대신 수행해 주는 시스템을 말한다. 이러한 시스템의 핵심 구성요소는 관련문제영역의 데이터를 수집하여 유용한 정보를 추출하는 데이터마이닝 기법을 중심으로 진행되어왔다. 데이터마이닝 기법의 핵심기능은 이미 알려진 데이터로부터 연관관계규칙과 패턴을 발견하여 해당 문제영역에서 필요한 의사결정에 활용하는 것이다. 특히 처리해야 할 데이터의 양이 많아지고 새로운 데이터가 계속하여 생성되는 환경에서는 적응능력을 가진 학습방법을 포함하는 최적화과정과 정보융합 방법들에 대한 연구가 시스템의 성능을 좌우하는 대표적인 이슈로 알려져있다.

본 논문에서는 유방암진단을 목표로 MLO 뷰와 CC 뷰 2 가지 관점에서의 영상을 분석하여 각 뷰에서의 normal, benign, malignant 3가지로 범주에 속할 가능성을 바탕으로 정보를 융합하여 전문가의 의사결정에 유용한 정보를 제공하는 (그림 1)과 같은 다각도 정보 융합 기반 지능형에이전트 시스템(MIAS)을 제안한다.

본 논문에서 제안된 MIAS 시스템은 퍼지클러스터링에 의하여 제공된 데이터를 바탕으로 유방 x-선 영상의 분류결과뿐만 아니라 퍼지정보이론을 적용하여 비대칭(asymmetry) 여부와 불확실성(uncertainty)의 정도와 같은 전문가의 의사결정에 실효성있는 정보를 제공해 주고있다.

유방촬영술(mammography)로부터 얻은 유방 x-선 영상(mammogram)의 종괴(mass)와 석회(calcification) 환부는 유방암 진단의 가장 중요한 단서로서 알려져 있다[4-5]. 진단 시스템은 디지털영상 으로부터 진단에 필요한 데이터를 준비하는 과정과 준비된 데이터를 활용하여 양성과 악성으로 분류하는 과정으로 구성되어 있다. 이러한 시스템은 악성을

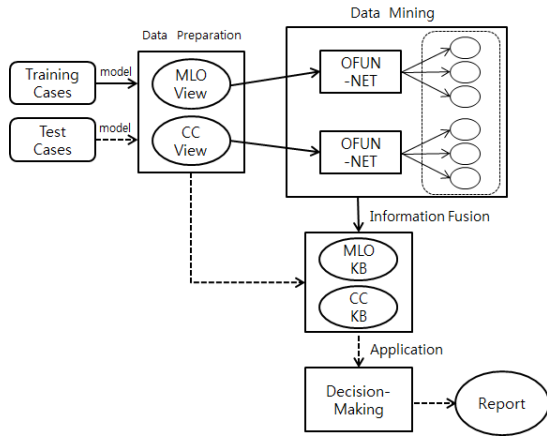


그림 1. 지능형에이전트시스템(MIAS) 구성도
Fig. 1. Intelligent Agent System Architecture

조기에 발견하여 적절한 조치를 취할 수 있도록 해 주고 양성인 경우 추적조사를 해야 할지, 조직검사를 해야 할지 등의 의사결정을 돕는 컴퓨터를 활용한 의료진단시스템의 가능성과 지능형에이전트시스템의 실제 사례를 보여주고 있다.

1. 데이터준비모듈

지능형에이전트 시스템이 특정목적을 가지고 의사결정을 하기 위하여 관련 데이터를 선택하여 준비하는 과정이 필요하다. 데이터준비 과정은 데이터선택, 데이터 전처리, 데이터 변환의 세 개의 세부과정으로 이루어져 있다. 이 과정은 주어진 데이터를 마이닝에 사용할 수 있도록 문제영역에서 관찰된 다차원데이터로부터 데이터를 묘사하는 구조를 잘 반영하는 특징을 추출하는 것이 핵심기능이다. 일반적으로 특징 추출은 같은 범주에 속하는 데이터들의 공통적인 속성(intraset features)과 서로 다른 범주에 속하는 데이터를 구별하는 속성(interset features)으로 나누어 고려되어 왔다. 이러한 특징선택은 어떤 시스템이든 근본적인 성능과 유용성을 결정짓는 중요한 영역으로 상관관계기법, 차원축소 및 상호정보처리 등의 정보이론이나 통계학적인 접근방법을 중심으로 연구되어왔다. 하지만 점차 다루는 데이터의 양이 방대해지고 복잡해지면서 인공지능분야의 휴리스틱 방법론이 적용되어 연구영역을 확대하고 있다.

데이터준비모듈에서는 유방암진단을 위한 의료전문에이전트시스템으로 구축하기 위하여 DDSM으로부터 두 가지 관점에서의 촬영방법-Mediolateral Oblique(MLO), Cranio-Caudal(CC)-으로부터 얻은 디지털영상을 사용하였다. 이 때 추출된 특징은 X선 영상판독 전문가의 의견을

BI-RADS 기준에 의한 수치값을 변경하여 사용한 것으로 본 논문의 중점연구대상은 아니고 결과를 가져다 시스템을 구성하는 하나의 모듈로 활용하였다.

2. 데이터마이닝모듈

데이터마이닝은 방대한 양의 데이터를 탐색하여 숨겨진 정보와 규칙 또는 요약정보를 꺼내는 작업을 말한다. 이를 위하여 여러 방법론이 적용되어왔지만 클러스터분석방법은 가장 일반적으로 적용되는 방법이다. 이는 비교사학습방법으로 비슷한 그룹으로 묶어가며 그 군집에 속하는 대푯값을 찾아가는 방법으로 새로운 데이터에 대하여 적용하기 위한 신경망모델이 보편적으로 사용되고 있다.

본 논문에서는 데이터마이닝을 위하여 클러스터링결과가 만들어내는 거리의 합을 표현하는 퍼지함수의 값 J_m 을 최소로 하도록 학습을 유도하는 퍼지신경망 OFUN-NET을 사용한다. OFUN-NET은 [9]에서 제안된 FNN-B를 바탕으로 클러스터의 대푯값 (v_1, v_2, \dots, v_c) 과 각 데이터의 클러스터 대푯값과의 일치도를 측정하는 소속함수값 u_{ij} 를 학습해 간다.

$$J_m = \sum_{j=1}^n \sum_{i=1}^c (u_{ij})^m (d_{ij})^2 \tag{1}$$

$$u_{ij} = \frac{1}{\sum_{s=1}^c \left\{ \frac{\|x_j - v_i\|}{\|x_j - v_s\|} \right\}^{2/(m-1)}} \tag{2}$$

이러한 학습 알고리즘은 클러스터링의 결과가 만들어 내는 오류 값을 요약하는 퍼지 함수를 설정한 후 그 값이 최소가 되도록 학습의 방향을 유도하는 메카니즘에 의해 진행된다. 또한 제안된 방법은 입력과 출력 사이의 관계를 기술하기 어려운 경우도 쉽게 처리하는 비교사 학습신경망의 장점도 함께 가지고 있다.

데이터마이닝모듈의 중심기능을 수행하는 OFUN-NET은 데이터준비모듈로부터 전처리과정과 변환과정등의 데이터준비과정을 통하여 두 가지 데이터집합 $\pi_{MLO} = [x_1, x_2, \dots, x_n]$, $\pi_{cc} = [y_1, y_2, \dots, y_n]$ 을 준비하여 각각 OFUN-NET의 입력으로 사용하여 학습한 결과 3가지 클러스터에 대한 대푯값을 만들어 MLO 지식베이스와 CC 지식베이스를 만들게 된다.

3. 정보융합모듈

정보융합모듈에서는 진단을 원하는 2가지 뷰의 한쌍의 테스트 데이터 t_1, t_2 에 대하여 데이터마이닝 모듈에서 OFUN-NET의 학습결과로 만들어진 MLO 뷰와 CC 뷰 각각 3개의 클래스에 대한 지식베이스안의 정보를 활용하여 다음을 계산한다.

1) 3개의 클래스에 대한 일치도를 다음 식에 의하여 계산한다. 이 계산으로부터 MLO 뷰로부터의 데이터 t_1 의 각 클래스에 속할 가능성, u_{11}, u_{21}, u_{31} 와 CC 뷰로부터의 데이터 t_2 의 각 클래스에 속할 가능성, u_{12}, u_{22}, u_{32} 로부터 각 뷰 데이터의 각 클래스에 대한 소속함수값을 계산한다.

$$u_{ij} = \frac{1}{\sum_{c=1}^3 \left\{ \frac{\|t_j - v_i\|}{\|t_j - v_c\|} \right\}^{2/(m-1)}} \quad (3)$$

2) 비대칭여부를 측정한다. 서로 다른 뷰 데이터의 결과가 다른 경우 주어진 테스트 데이터의 결과를 비대칭(asymmetry)인 것으로 판정한다. 즉 다음 식으로부터 c_1 과 c_2 의 값이 다르면 비대칭인 경우로 판단하게 된다.

if ($c_1 \neq c_2$) then Report "This case is asymmetry"
 ,where $c_1 = \max(u_{i1})$ 인 값을 가지는 i
 $c_2 = \max(u_{i2})$ 인 값을 가지는 i

3) 하나의 뷰에서 얻은 데이터의 불확실성(uncertainty)를 측정한다. 다음 식의 조건을 만족하면 주어진 테스트 데이터는 시스템에서 학습하여 가지고 있는 지식베이스에 대하여 퍼지상태(fuzzy status)인 것으로, 아니면 분명한 상태(crisp status)인 것으로 정의한다.

$$\frac{u_{ij} - u_{sj}}{u_{ij}} \leq \frac{1}{c}, \quad (4)$$

where $u_{ij} = \max(u_{ij}), u_{sj} = \max(u_{ij}), s \neq i$

식 (4)는 주어진 테스트 데이터의 가장 큰 소속함수값과 두 번째 소속함수값의 차이가 작으면 주어진 데이터는 현재 학습한 정보로는 판단하기 어려운 상태임을 말해주며 그렇지 않고 그 차이가 충분히 크면 판정하기에 충분한 상태임을 말해주고 있다

4. 응용모듈

응용모듈에서는 정보융합모듈에서 만들어낸 정보를 바탕으로 주어진 에이전트시스템이 적절한 기능을 수행하도록 분류 및 진단 등의 실제 응용 업무에 도움이 되는 주문을 하고 전문가를 도와주는 역할을 수행한다. 본 시스템은 유방암진단과 관련 후속조치에 필요한 정보를 제공해 주는 것이기 때문에 다음과 같은 기능을 수행하고 있다.

1) 비대칭(asymmetry) 결과를 도출했거나 불확실(uncertainty)한 결정상태로 판정된 경우 본 시스템으로는 판정을 보류하고 전문가의 의견을 모아 진단하게 된다.

2) 1)의 두 가지 경우가 아니라면 시스템이 판정할수 있는 경우로서 3가지 진단결과(normal, benign, malignant) 중 가장 적합도가 높은 클래스로 판정하게 된다.

2-1) $\max(u_{i1})$ 와 $\max(u_{i2})$ 인 값을 가지는 클래스 i 가 정상 클래스에 속하는 것으로 판정되면 정상인 것으로 레포팅한다

2-2) 양성종양이라 판정되면 ML촬영, 확대촬영, 유방초음파 등의 향후 추적조사를 위한 진단을 한다.

2-3) 악성종양이라고 판정되면 조직검사 등의 향후 조치를 하게된다.

IV. 응용데이터 적용 및 분석

본 장에서는 III장에서 제안한 지능형 에이전트 시스템, MIAS의 타당성을 검증하기 위하여 [표 1]에 요약된 공개적으로 알려진 DDSM[5,16]으로부터 획득한 입력데이터를 준비하여 핵심모듈에 적용하여 진단시스템으로서의 가능성을 확인하고자 한다.

1. BI-RADS 기반 특징데이터 적용

본 실험에서 사용한 DDSM 벤치마크 데이터베이스[5,16]는 MGH(Massachusetts General Hospital), WFU(Wake Forest University School of Medicine), WU(Washington University of St. Louis Medical Center)으로부터 수집한 43개의 볼륨안에 2620개의 케이스를 가지고 있다. 본 논문에서는 [표 1]에서 제시한 450개를 선택하여 영상에 연결된 ics파일과 overlay file로부터 BI-RADS 기반의 특징 6개를 인코딩하여 가져와서 입력데이

터를 준비하였다. 이때 사용한 인코딩방법은 American College of Radiology에서 제안한 BI-RADS 표준에 따라 J. Y. Lo가 제시한 rank ordering 방법(17)을 활용하였으며 [표2]와 같이 정리할 수 있다. 각 케이스에 연결된 ics 파일로부터 환자의 나이(patient_age)와 조직의 밀도(density)를 얻을 수 있으며 overlay 파일로부터 mass shape, margins, (또는 calcification type, calcification distribution), assessment, subtlety를 얻어 각 환부 유형 (lesion type)에 따라 4 가지의 특징데이터를 얻을 수 있다.

이렇게 준비한 450 케이스에 대한 두가지 뷰의 450*6의 입력데이터 집합을 구성한다. 이를 제안된 시스템의 학습데이터로 준비하여 데이터마이닝모델의 두가지 퍼지신경망, OFUN-NET(9)을 통하여 각 분류의 중심값과 각 데이터의 클러스터에 대한 적합도를 소속함수값으로 가지는 지식베이스를 구성한다.

표 1. 데이터정보
Table 1. Data information

Institution	Normal	Benign	Malignant
MGH	50	38	50
WFU	70	52	58
WUJ	30	60	42
Total	150	150	150

표 2. BI-RADS 특징데이터와 인코딩값
Table 2. BI-RADS mammographic features and encoding value

Feature Type	Description (Encoding Value)
Mass Shape	no mass (0), round (1), oval (2) lobulated (3), irregular (4)
Margins	no mass (0), well circumscribed (1) microlobulated (2), obscured(3) ill-defined (4), spiculated (5)
Calcification Type	no calcification (0) milk of calcium-like (1) eggshell or rim (2), skin (3) vascular (4),spherical or lucent (5) suture (6), coarse (7) large rod-like (8), round (9) dystrophic (10), punctate (11) indistinct (12), pleomorphic (13) fine branching (14)
Calcification Distribution	no calcification (0), diffuse (1) regional (2),segmental (3), linear (4) clustered (5)
Density	1, 2, 3, 4
Assessment	1, 2, 3, 4, 5
Subtlety	1, 2, 3, 4, 5

2. 분류정확도 비교 및 분석

이렇게 학습된 지식베이스를 가지고 본 에이전트시스템의 진단능력을 확인하기 위하여 테스트 데이터셋을 학습데이터와 같은 방법으로 준비하여 분류정확도를 측정하였다. 테스트 데이터셋은 학습에 참여한 데이터 중 각 클래스당 50개와 학습에 참여하지 않은 데이터를 각 클래스 당 50개씩 준비하여 모두 300가지 케이스에 대하여 300*6의 테스트 데이터쌍을 준비하였다.

제안된 MIAS 시스템에 준비한 두 가지 뷰 데이터셋을 가지고 분류정확도를 기준으로 분석한 결과는 [표 3]와 같이 요약할 수 있다. 준비된 두 쌍의 300*6의 테스트 데이터 셋을 기존의 잘 알려진 분류 알고리즘(7-9)[18], KNN(K-Nearest Neighbor), 지지벡터기계 (SVM:Support Vector Machine)을 이용하여 두 뷰의 결과가 다르게 나온 경우(Asymmetry)를 제외한 분류정확도이다. KNN과 SVM과 달리 FCM(Fuzzy C-means) 알고리즘은 제안된 방법과 같은 퍼지클러스터분석에 기반을 두고 있으므로 불확실결정상태를 측정하여 같은 조건에서 비교가 가능하도록 실험하였다. [표 3]의 결과는 MIAS시스템이 기존의 방법과 다른 접근방법이므로 순수한 정량적인 비교는 의미가 없을 수 있지만 진단시스템으로서 MIAS는 기존의 방법과 비교하여 타당한 결과를 얻었다는 것을 보여주기에는 충분한 실험결과이다. 또한 대표적인 퍼지클러스터링 알고리즘인 FCM과 같은 기준에서 보다 높은 분류정확도를 얻었다는 것으로 본 시스템의 타당성을 알 수 있다. 이와 같이 제안된 시스템은 판정보류한 경우에 대하여 비대칭여부나 불확실한 결정상태와 같은 정보를 전문가에게 제공하여 줄 수 있는 잇점도 가지고 있다. 물론 판정보류한 사례에 대하여 1차 단계에서 반영하지 않은 기하학적 특징요소를 가지고 판정하는 2단계 접근방법을 고려해 볼 수 있다. 또한 본 실험을 통하여 제안된 시스템은 응용독립적인 모듈과 의존적인 모듈을 분리하여 제시함으로 다양한 응용데이터에 적용을 쉽게 할 수 있다는 것도 확인하였다.

표 3. 분류정확도 비교
Table 3. Comparison of Classification Accuracy

	KNN	SVM	FCM	MIAS
분류 정확도	84.8%	87.7%	86.9%	88.1%

V. 결론

컴퓨터와 데이터처리기술의 발전은 의료영상의 해석과 분석에 활용되어 의료정보화에 기여하고 있다. 특히 다각도 의료영상은 진단의 중요한 단서로서 처리과정에서 발생하는 전문가의 판독오류와 해석과정의 불일치를 줄이기 위하여 영상처리기술과 정보이론이 적용되어 컴퓨터 진단시스템은 전문가의 의사결정을 도울 수 있는 체계적인 도구로서 연구되고 있다.

본 논문에서는 퍼지 클러스터 분석에 기반을 둔 데이터마이닝모듈과 정보융합모듈을 핵심구성요소로 가지는 지능형에이전트 시스템을 제안하고 응용모듈을 통하여 의료전문가 시스템에 적용하였다. 또한 DDSM으로부터의 실세계 데이터를 활용하여 유방암을 진단하는 실험을 통해 기존의 시스템과 비교하여 보다 높은 분류 정확도와 비대칭여부 등의 정보를 제공하므로 진단 전문가에게 보다 유용한 정보를 제공할 수 있음을 보여주면서 제안된 시스템의 타당성을 입증하였다. 이러한 시스템의 성능은 디지털라이저의 처리과정이나 특징추출 과정 등에 크게 좌우되므로 서로 다른 상황에서의 연구 결과와 정량적인 비교는 큰 의미가 없을 수도 있다. 그러나 본 연구는 지능형에이전트 시스템의 기능을 모듈별로 구성하고 두 가지 뷰의 mam모영상에 적용하여 진단시스템으로서 응용할 수 있는 프레임워크를 마련하여 향후 연구의 기초를 만들 수 있었다.

앞으로 다양한 실험을 통하여 프로토타입 시스템으로 제안된 지능형에이전트시스템의 타당성과 가능성을 확인할 필요가 있다. 또한 보다 견고한 시스템으로 발전하기 위해서는 특징선택방법이 중요하나 이는 또 다른 연구 분야로서 해당 영역의 전문적인 지식도 필요로 하기 때문에 연구의 시작단계부터 응용독립적인 모듈과 의존적인 모듈을 분리하여 시스템을 구성하는 방향으로 연구를 진행하는 것이 바람직하다는 결론을 얻게 되었다. 이러한 연구를 통하여 시스템의 각 모듈을 보다 정교하게 구성하고 실세계에 적용하면서 전문화된 컴퓨터보조진단시스템으로서 발전할 수 있도록 지속적인 연구가 진행되어야 할 것이다.

참고문헌

[1] X. Wu, X. Zhu, G-Q. Wu, and W. Ding, "Data Mining with Big Data", IEEE Trans on

Knowledge and Data Engineering, Vol.26, No.1, Jan. 2014

- [2] D.S. Kim, C.S. Kim, and K.W. Rim, "Modeling and Design of Intelligent Agent System", International Journal of control, Automation, and Systems, Vol. 1, No. 2, pp. 257-260, June 2003.
- [3] D. Vidhate, Dr. P. Kulkarni, "Cooperative Machine Learning with Information Fusion for Dynamic Decision Making in Diagnostic Applications", Int. Conf. on Advances in Mobile Network, Communication and its Applications, 2012.
- [4] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa and Y. Yang, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography : Recent Advances", IEEE Trans on Information Technology in Biomedicine, Vol.13, No.2, March, 2009.
- [5] M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer Jr., "The Digital Database for Screening Mammography", 5th IWDM, Medical Physics Publishers, 2001.
- [6] Mehmed Kantardzic, "Data Mining : Concepts, Models, and Algorithms", John Wiley & Sons, 2011.
- [7] Z. Vlad, M. D. Ofelia, and T-A. Maria, "Fuzzy Clustering in an Intelligent Agent for Diagnosis Establishment", Scientific Bulletin of the Petru Maior University of Tirgu Mures Vol. 6, 2009.
- [8] P. Vats, "A Noval Study of Fuzzy Clustering Algorithms for their Applications in Various Domains", JICTEE, 2014.
- [9] H. S. Rhee, "A Feature Selection Method Based on Fuzzy Cluster Analysis", Journal of Korea Information Processing Society, Vol.14-B, No.2, pp.135-140, 2007.
- [10] Wu, Y., He, J., Man, Y., & Arribas, J.I., "Neural Network Fusion Strategies for Identifying Breast Masses", proc. of the IEEE International Joint Conference on Neural Networks(IEEE-IJCNN'2004), 2004.

[11] R. Panchal and B. Verma, "Characterization of breast abnormality patterns in digital mammograms using autoassociator neural network," in International Conference on Image Processing 2006, Part III, LNCS, vol. 4234, pp. 127-136, Springer-Verlag, 2006.

[12] Brijesh Verma and John Zakos, "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques," IEEE Trans. on Information Technology in Biomedicine, vol. 5, no. 1, march 2001.

[13] M. Radovic, M. Djokovic, A. Peulic, and N. Filipovic, "Application of Data Mining Algorithms for Mammogram Classification", 13th conf. on Bioinformatics and Bioengineering(BIBE), 2013.

[14] L.Sun, L. Li, W. Xu, W. Liu, J. Zhang, and G. Shao, "A Novel Classification Scheme for Breast Masses Based on Multi-view Information Fusion", 4th Int. Conf. on Bioinformatics & Biomedical Engineering (iCBBE), 2010.

[15] H. Zhao, W. Xu, L. Li, and J. Zhang, "Classification of Breast Masses Based on Multi-view Information Fusion Using Multi-Agent Method", 5th Int. Conf. on Bioinformatics & Biomedical Engineering (iCBBE), 2011.

[16] Reference URL :
<http://marathon.csee.usf.edu/Mammography/Databse.html>

[17] J.Y. Lo, et al., "Computer-aided classification of breast microcalcification clusters: Merging of features from image processing and radiologists", Proc. SPIE 5032, Medical Imaging, 2003.

[18] Dheeda J. and T. Selvi.S, "Classification of Malignant and Benign Microcalcification Using SVM Classifier", Proc. of ICETEECT, 2011.

저 자 소 개



이 현 숙

1989년 : 서강대학교
전자계산학과(학사)

1991년 : 포항공과대학교
컴퓨터공학과(석사)

1997년 : 서강대학교
컴퓨터학과(박사)

1991년~1997년 :
한국전자통신연구소(ETRI)
연구원

2006-2007년 : UCF
(University of Central
Florida) 방문교수

1997년~현재 : 동양공업전문대학교
전산정보학부 교수

관심분야 : 지능형정보처리,
데이터마이닝
소프트웨어개발방법론,

Email : hsrhee@dongyang.ac.kr