論 文
57-6-23

의료진단 및 중요 검사 항목 결정 지원 시스템을 위한 랜덤 포레스트 알고리즘 적용

Application of Random Forest Algorithm for the Decision Support System of Medical Diagnosis with the Selection of Significant Clinical Test

尹 泰 鈞* · 李 寬 洙†
(Taegyun Yun · Gwan-Su Yi)

Abstract – In clinical decision support system (CDSS), unlike rule-based expert method, appropriate data-driven machine learning method can easily provide the information of individual feature (clinical test) for disease classification. However, currently developed methods focus on the improvement of the classification accuracy for diagnosis. With the analysis of feature importance in classification, one may infer the novel clinical test sets which highly differentiate the specific diseases or disease states. In this background, we introduce a novel CDSS that integrate a classifier and feature selection module together. Random forest algorithm is applied for the classifier and the feature importance measure. The system selects the significant clinical tests discriminating the diseases by examining the classification error during backward elimination of the features. The superior performance of random forest algorithm in clinical classification was assessed against artificial neural network and decision tree algorithm by using breast cancer, diabetes and heart disease data in UCI Machine Learning Repository. The test with the same data sets shows that the proposed system can successfully select the significant clinical test set for each disease.

Key Words : Clinical Decision Support System, Random Forest, Medical Diagnosis, Feature Selection

1. 서 론

의학정보가 방대해지고 다양한 임상 사례가 증가함에 따라 질환 진단 및 예후 추정 등 의사의 전문적인 의사 결정 시 객관성을 확보하고 종합적인 정보 지원을 하는 임상 의사결정 지원 시스템 (Clinical Decision Support System: CDSS)의 역할이 중요해지고 있다. 임상 의사결정 지원 시스템은 규칙 기반 전문가 시스템 (Rule based Expert system)과 데이터 기반 기계 학습 추론 시스템 (Data driven Machine Learning Inference System) 으로 나누어진다. 규칙 기반 전문가 시스템은 의학 전문가에 의해 사전에 정해진 규칙을 사용하여 질환을 진단하는 시스템이고 데이터 기반 기계 학습 추론 시스템은 질환이 판정된 환자들의 임상 정보들을 수집한 후 기계 학습을 이용하여 주어진 임상 정보로부터 질환을 추론하도록 하는 시스템이다. 정확하고 종합적인 메커니즘이 밝혀지지 않은 대다수의 질환들에 대해서는 다양한 사례 데이터를 기반으로 하는 분석의 중요성이 매우 크다. 이러한 데이터 기반 시스템을 구성하는 추론엔진으로 인공신경망(Artificial Neural Network: ANN), 의사결정나무(Decision Tree: DT) 및 베이지안 네트워크 등 기존에 잘 알려진 기계학습 분류 알고리즘들을 적용하고 그 효율성을 검증하는 연구가 활발히 시도 되었다 [1, 2, 3, 4]. 데이터 기반 기계학습 추론 과정은 적절한 알고리즘을 적용한다면 진단의 정확성을 제고함과 동시에 다양한 사례에 포함된 판별 특징 정보를 추출할 수 있는 장점을 갖고 있다. 그러나 기존의 연구들은 질환 판별기로서의 정확성 향상에 편중하고 있고 각 질환 별 또는 사례군 별 판별 특징의 정보를 추출하는 방법을 제공하거나 그 효율을 분석하고 있지 않다.

임상데이터의 판별 특징 (feature)을 구성하는 각종 임상 검사항목들은 질환을 표지하는 생체 내 주요 물질로서, 임상 시험을 통해 이미 확정된 것들이나, 다양한 사례 분석을 통해 이들과 질환들 간의 상관관계 정보를 발굴하여 다양한 임상 프로세스에 활용할 수 있다. 각 판별 특징의 질환별 또는 사례군 별 판별 중요도를 분석하면 추정된 질환 후보군 각각의 판별이나 질환 간의 감별에 특별히 효율적인 검사항목들에 대한 정보를 얻을 수 있다. 또한, 특수한 사례들이 보여주는 다양한 임상 조건에 적합한 검사항목 정보를 확보할 수 있게 된다.

본 논문에서는 기계학습 분류기(classifier)와 판별 특징 선택 기법을 결합한 새로운 형태의 임상 의사결정 지원 시스템을 제안한다. 이 시스템의 기계학습 분류기와 판별 특징 중요도 측정에는 랜덤 포레스트 (Random Forest) 알고리즘이 적용되었다. 본 시스템은 판별 특징 중요도 측정을 마친 후, 후진 제거기법을 통해 중요도가 떨어지는 판별 특징들을 제거해 가면서 분류기의 판별오류를 검사하여 질환 판별에 결정적인 검사항목 셋을 선정한다. 본 시스템의 핵심적인 랜덤 포레스트 알고리즘 [5] 은 처음 소개된 2001년 이후 다양한 분야에 적용되어 왔으나 임상 진단 데이터 분

* 正 會 員 : 韓國精報通信大學 工學部 碩博士統合課程
† 교신저자, 正會員 : 韓國精報通信大學 工學部 副教授
E-mail : gsyi@icu.ac.kr
接受日字 : 2008年 4月 17日
最終完了 : 2008年 4月 28日

류에 적용되지 않았었다. 본 논문에서는 UCI Machine Learning Repository [6]에서 제공하는 심장병, 유방암, 당뇨병 데이터를 사용하여 랜덤 포레스트 알고리즘의 판별 효율성이 기존에 임상 데이터에 적용되었던 대표적인 알고리즘(인공신경망 및 의사결정나무 학습 알고리즘)에 비해 우월함을 보였다. 또한, 각 질병에 대해 선정된 중요 검사항목의 판별 수행 능력과 임상적 중요성을 고찰하여 제안된 시스템의 유용성을 검증하였다.

2. 관련 연구 소개

기존에 연구되어온 임상 의사결정 지원 시스템은 크게 규칙 기반 전문가 시스템과 데이터 기반 기계 학습 추론 시스템으로 나뉠 수 있다. 규칙 기반 전문가 시스템은 전문가가 특정 영역의 문제를 해결하기 위하여 전문 지식을 바탕으로 사전에 여러 규칙을 도출하여 정리하고, 의사 결정 지원이 필요할 때 정리된 규칙을 바탕으로 자동적으로 질환 판별 등의 임상 의사 결정을 보조하는 시스템이다. 규칙 기반 전문가 시스템은 해당 영역의 전문가의 지식을 바탕으로 도출된 규칙을 사용하기 때문에 의사 결정 과정이 이해하기 쉽고 명확하다는 장점이 있으나 전문가가 도출한 규칙이 전문가에 따라 주관적으로 생성될 수 있으며 사전에 규칙이 도출되지 않은 영역의 문제에는 적용할 수 없다는 문제점을 지니고 있다. 이러한 문제점을 해결하기 위하여 다양한 기계 학습 알고리즘을 이용한 데이터 기반 기계 학습 추론 시스템이 연구되어져 왔다. 기계학습 추론에 사용되는 대표적인 알고리즘으로는 의사 결정 나무 (Decision Tree), 인공신경망 (Artificial Neural Network)과 베이지안 네트워크 (Bayesian Network) 등이 있다. 임상 의사결정 지원 시스템에는 인공신경망 [1], 데이터마이닝 방법 [2], 베이지안 네트워크 [3], 통계적 추론방법 [4] 등이 적용되어 왔다. 최근에는 이들을 결합한 방법이 소개되기도 하였다. 이 중에는 규칙 기반 추론방법과 데이터기반 기계학습 방법을 결합한 예도 나타났다[7].

이렇게 적용된 방법들 중 인공신경망과 같이 개별 판별 특징과 추론 결과의 선형적 상관관계를 추출할 수 없는 경우는 검사 항목의 중요도와 같은 개별 판별 특징의 정보를 이용하는데 한계가 있다. 판별 특징 즉 각 검사 항목의 개별 중요도 정보를 추출할 수 없고 각 판별 특징이 질환 판별과 같은 최종 클래스 판별에 기여하는 정도를 파악 할 수 없다는 단점을 지니고 있다. 그러나 통계적 추론이나 다른 데이터마이닝 기법과 같은 것들은 이러한 정보를 쉽게 추출할 수 있음에도 불구하고 아직까지 CDSS에 적용된 연구에서는 개별 판별 특징 정보를 이용하여 질환군 별 또는 사례군 별 중요 검사항목 군을 결정하는 것과 같은 방법은 시도되지 않았었다. 최근 Abdel-Aal 의 연구[8]에서와 같이 판별특징을 추론 분석에 이용한 예가 있으나, 이 방법도 질환 판별을 수행함에 있어 임상 데이터로부터 판별 특징을 추출하고 상이한 판별 특징 집합을 구성하여 신경망을 학습시키는 연구로서, 임상 검사 항목과 같은 판별 특징의 중요성에 주목하기 보다는 시스템의 분류 정확성을 향상시키기 위한 수단으로서 판별 특징을 이용하고 있다.

또한, 랜덤 포레스트 분류 알고리즘은 2001년 Breiman이 개발한 이후 여러 분야의 데이터에 대해 기존의 분류 알고리즘에 비해 좋은 성능을 보여 주목을 받고 있으나, 아직 임상 질환 판별을 위한 추론에 적용되지 않았다. 따라서 임상 데이터 판별에 대한 효율성을 정확히 측정하여 유용성 여부를 검증할 필요가 있다. 본 연구에서는 랜덤 포레스트를 이용한 분류기와 중요 판별 특징 결정기로 시스템을 구성하고 그 성능을 측정하여, 새로운 개념의 CDSS의 유용성을 보여주고자 한다.

3. 알고리즘 및 시스템

본 시스템은 질환 판별을 수행하는 랜덤 포레스트 분류기와 중요 임상 검사 항목을 선정하는 판별 특징 결정기로 이루어져있다. 판별 특징 값과 해당 클래스의 조합으로 이루어져 있는 사례의 집합이 훈련 데이터로 입력되면 개별 사례를 주어진 클래스로 분류하기 위하여 랜덤 포레스트를 구성 한다. 학습된 랜덤 포레스트는 새로운 임상 사례를 판별하는데 이용될 수 있으며 생성된 랜덤 포레스트와 훈련 데이터는 판별 특징 결정기에 입력되어 최종적으로 중요 판별 특징 셋을 갖는다.

분류기에 사용된 랜덤 포레스트는 무작위 (Random)로 추출한 사례의 집합들을 이용하여 많은 수의 의사결정나무를 생성하고 생성된 여러 의사 결정 나무의 판별 클래스 (Class)들을 가중 투표 (Voting)하여 최종 클래스를 결정하는 분류 (Classification)기법이다. 각각의 의사 결정 나무는 작은 편차를 가지기 위해 가지치기를 하지 않고 완전하게 생성된다. 랜덤 포레스트는 다른 분류 기법과 비교해 여러 가지 장점을 지닌다. 랜덤 포레스트는 부트스트랩 (Bootstrap) 기법을 이용하여 랜덤 포레스트 학습에 필요한 훈련 데이터를 생성하므로 적은 수의 임상 사례만으로도 일정 수준 이상의 정확성을 가지는 분류기를 생성할 수 있다. 또한 훈련 과정에서 무작위하게 추출된 훈련데이터로 많은 수의 의사 결정 나무를 생성하여 다양한 패턴을 포괄하기 때문에 훈련 데이터가 아닌 새로운 데이터가 판별을 위하여 분류기에 입력되었을 경우에도 높은 수준의 정확성을 가질 수 있다. 이외에도 랜덤 포레스트는 많은 수의 입력 데이터를 처리할 수 있고, 결측값 (missing value)을 예측할 수 있으며 과적합 (Overfitting) 문제가 발생하지 않고, 변수 중요도 (Variable Importance)를 제공한다.

이러한 일련의 특징 중에서 변수 중요도는 본 논문에서 제시하는 시스템 중 판별 특징 결정기 구성에 있어서 가장 중요한 특징이라고 할 수 있다. 판별 특징 결정기는 클래스 판별에 영향을 미치는 여러 판별 특징 중 중요도가 떨어지거나 클래스 판별에 부정적인 영향을 미칠 수도 있는 판별 특징을 제거하여 질환 판별에 중요한 영향력을 가지는 필수적인 소수의 판별 특징 셋을 선정해주는 역할을 한다. 랜덤 포레스트는 mean decrease accuracy와 mean decrease Gini

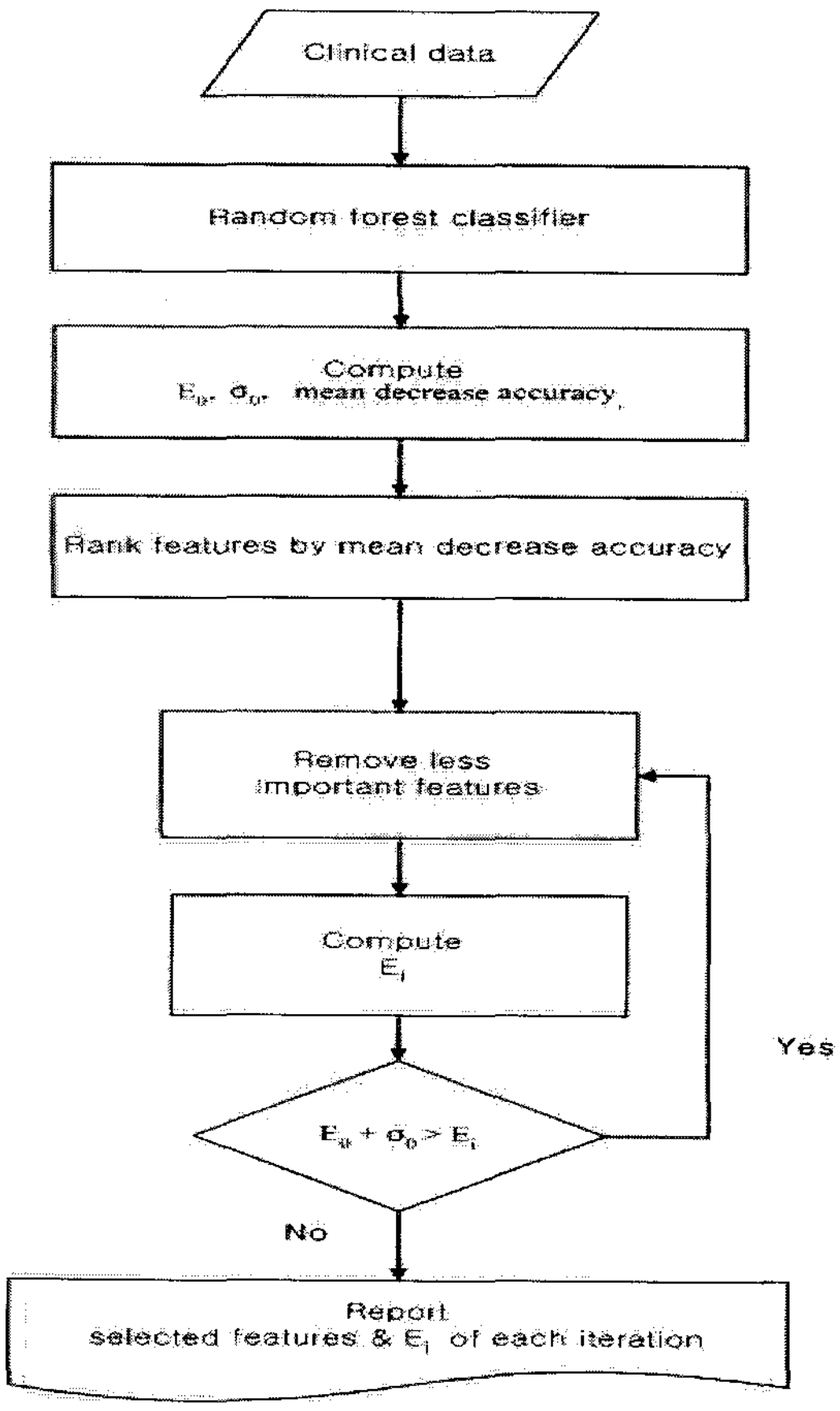


그림 1 제안된 시스템의 진단 판별 및 검사 항목 선정 흐름도

Fig. 1 Flow chart of disease classification and important clinical test selection in the proposed system

라는 두 가지 종류의 변수 중요도 측정치를 제공한다. 본 논문에서는 이 두 가지 측정치 중 mean decrease accuracy를 사용하였다. Mean decrease accuracy는 훈련의 결과 생성된 의사 결정나무들의 노드를 무작위 하게 배열하였을 경우와 훈련을 통해 배열된 경우를 비교하여 랜덤 포레스트의 오류 측정치인 Out-of-Bag (OOB) 오류값이 얼마나 차이를 보이는지에 따라 계산된다. 중요한 특징은 덜 중요한 특징에 비해 높은 mean decrease accuracy 값을 가진다.

질환 판별력의 중대한 훼손 없이 중요한 최소의 검사 항목을 선택하기 위해서 후진 제거법을 도입하였다. 랜덤 포레스트에서 도출되는 변수 중요도를 이용하여 모든 검사 항목을 중요도 순으로 정렬한 뒤에 중요성이 낮은 검사 항목부터 순차적으로 제거하고 남아있는 검사항목의 셋으로 랜덤 포레스트를 재구성한다. 재구성된 랜덤 포레스트의 OOB 오류값(Ei)을 초기에 모든 검사 항목을 대상으로 구성한 랜덤 포레스트의 OOB 오류값 (Eo)과 비교하여 일정한 문턱값 (σo)범위에 있을 때 까지 위의 과정을 반복한다. 문턱값 범위를 초과하여 반복 과정이 중단되면 이전 단계까지 제거한 검사 항목을 제외한 검사 항목 셋을 중요 검사 항목 셋으로 정의한다. 이러한 본 시스템의 수행과정은 그림 1에 요약하였다. 본 시스템은 R 언어 [9]로 작성하고 R package 형태로 구현하였다.

4. 성능 평가

제시된 알고리즘의 질환 판별력을 측정하기 위하여 UCI Machine Learning Repository에 공개된 세 종류의 임상 데이터를 획득 하였다. 세 종류의 임상 데이터는 클리브랜드 심장병 데이터 (13 가지 항목, 297개의 사례), 위스콘신 유방암 데이터 (9가지 항목, 683개의 사례), 당뇨병 데이터 (8가지 항목, 768개의 사례)이다. 각각의 항목에 결측값을 가지는 사례는 제외 되었다. 세가지 데이터 셋 모두 검사 항목의 측정치와 질환의 존재 유무가 명시되어있다. 판별 결과는 50번의 교차 검증 실험 측정치의 평균값을 사용하였으며, 70%의 데이터는 학습에 학습되었고 30%의 데이터는 학습된 분류기의 테스트에 사용되었다. 표 1은 실험에 사용된 데이터와 관련된 통계치를 보여준다.

표 1 UCI Machine Learning Repository에서 획득한 질환별 실험 데이터 정보

Table 1 Summary statistics for disease data sets selected from UCI Machine Learning Repository

데이터	항목	양의 사례	음의 사례	총 사례	학습	테스트
심장병	13	137	160	297	208	89
유방암	9	239	444	683	478	205
당뇨병	8	268	500	768	538	230

제시된 방법의 질환 판별력은 기존에 활발히 연구되어진 인공 신경망과 의사결정나무의 판별력과 비교되었다. 표 2는 제시된 방법의 질환 판별력이 인공 신경망과 의사결정나무의 질환 판별력과 비견될 수 있음을 보여준다. 모든 검사 항목을 대상으로 학습된 랜덤 포레스트와 중요 검사항목을 선정하여 학습한 랜덤 포레스트의 분류 정확성은 의사 결정나무보다 항상 뛰어나며 일반적으로 판별력이 매우 우수하다고 알려져 있는 인공신경망과 유사한 성능을 나타낸다.

표 2 제안된 시스템의 질환 데이터 판별 정확성 비교

Table 2 Classification accuracy of the proposed method for three different diseases

데이터	제안된 방법		기존 방법	
	모든 검사 항목 사용	션정된 검사항목 사용	의사결정 나무	인공 신경망
심장병	83.1%	84.2%	77.1%	81.8%
유방암	97.5%	96.9%	94.7%	96.9%
당뇨병	75.7%	76.3%	74.6%	75.6%

제안된 방법을 사용하여 우리는 각각의 데이터 셋에서 판별력의 중대한 훼손 없이 중요한 검사 항목을 선정할 수 있었다. 클리브랜드 심장병 데이터의 13가지 항목은 3가지 항목으로 줄어들었으나 오히려 정확성의 실험 평균값은 83.1%

에서 84.2%로 증가함을 볼 수 있다. 위스콘신 유방암 데이터에서는 9가지 항목에서 4가지 항목을 선정함에 따라 분류 정확성이 약간 감소하였지만 이는 인공신경망의 분류 정확성과 같은 수치이다. 당뇨병 데이터에서도 8가지 항목에서 5가지 검사 항목으로 중요 검사 항목을 선정할 수 있었지만 분류 정확성은 75.7%에서 76.3%로 증가하였다.

표 3 선정된 검사 항목의 성능 비교

Table 3 Performance of different subsets of feature in the proposed system

데이터	모든 검사 항목 사용			선정된 검사항목 사용		
	정확성	민감도	특이도	정확성	민감도	특이도
심장병	83.1%	79.3%	86%	84.2%	78.2%	88.5%
유방암	97.5%	96.9%	97.3%	96.9%	96.1%	97.2%
당뇨병	75.7%	60.3%	85.3%	76.3%	60.9%	85.2%

표 3은 각각의 데이터를 대상으로 전체 항목으로 랜덤 포레스트를 구성하였을 때와 제시된 시스템으로 중요 항목을 선택하여 더 적은 수의 항목으로 랜덤 포레스트를 구성 하였을 때의 정확성 (Accuracy), 민감도 (Sensitivity), 특이도 (Specificity)를 나타낸다. 선택된 검사 항목으로 훈련된 랜덤 포레스트의 정확성, 민감도, 특이도는 전체 항목의 결과값들과 비교했을 때 큰 차이를 보이지 않는다. 오히려 특이도의 경우 클리브랜드 심장병 데이터에서 2.5%가량 증가하였다.

표 4는 각 데이터 셋에서 선택된 임상 관련 항목과 선택되지 않은 임상 관련 항목의 리스트를 보여준다. 클리브랜드 심장병 데이터에서는 chest pain type (cp), number of major vessels colored by floursopy (ca), Thal의 3가지의 중요 임상 관련 항목이 선택되었다. Duch et al.[10]은 동일한 클리브랜드 심장병 데이터 셋을 대상으로 심장병 질환을 판별하기 위한 최적의 규칙을 도출하였다. 수식 1과 수식2에서 볼 수 있듯이 Duch et al은 본 시스템이 실험 결과 선정한 동일한 3개의 심장병 관련 항목을 이용하여 판별 규칙을 만들었다. 이것은 제시된 방법이 심장병 판별에 임상적으로 중요한 의미를 지닌 검사 항목을 선정하였음을 입증한다. 또 하나의 검증 예로 Zhu와 Guan [11]은 당뇨병 데이터를 대상으로 상대적 중요 척도(Relative Importance Factor)라는 측정치를 제시하였는데 본 연구에서 제시한 방법을 통해 선택된 5개의 임상 관련 항목 중에서 3개의 항목이 높은 RIF 값을 가졌다.

$$R1: (thal = 0 \vee thal = 1) \wedge ca = 0.0 \quad (1)$$

$$R2: (thal = 0 \vee ca = 0.0) \wedge cp \neq 2 \quad (2)$$

표 4 제안된 시스템의 질환 데이터별 선택 및 미 선택 검사 항목 기술

Table 4 Description of selected features and unselected features for each dataset using proposed system

	심장병	유방암	당뇨병
선택된 검사 항목	CP: chest pain type	Clump Thickness	Number of times pregnant
	CA: number of major vessels (0-3) colored by flourosopy	Uniformity of Cell Size	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
	Thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	Uniformity of Cell Shape	Body mass index (weight in kg/(height in m)^2)
	Bare Nuclei	Diabetes pedigree function	
		Age (years)	
선택되지 않은 검사 항목	Age	Marginal Adhesion	Diastolic blood pressure (mm Hg)
	Sex	Single Epithelial Cell Size	Triceps skin fold thickness (mm)
	Resting blood pressure	Bland Chromatin	2-Hour serum insulin (mu U/ml)
	serum cholestoral in mg/dl	Normal Nucleoli	
	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)	Mitoses	
	Resting electrocardiographic results(0,1,2)		
	maximum heart rate achieved		
	exercise induced angina (1 = yes; 0 = no)		
	oldpeak= ST depression induced by exercise relative to rest		
	the slope of the peak exercise ST segment(1,2,3)		

본 실험결과가 나타내듯이 선택된 적은 수의 임상 관련 항목으로도 만족할만한 수준의 질환 판별이 가능하다는 사실은 주목할 만하다. 게다가 위스콘신 심장병 데이터의 경우는 13가지 임상 관련 항목 중 3가지 항목만을 사용하여 질환을 판별하였음에도 불구하고 판별의 정확성이 13가지 항목을 모두 사용한 경우보다 높았다. 이것은 일부 임상 검사 결과는 올바른 질환 진단을 이끌어내는데 부정적인 영향을 끼칠 수 있음을 보여준다. 실제로 모든 임상 검사 항목이 궁극적으로는 질환을 진단하는데 중요할 수 있지만 특정 의사 결정 단계에서 모든 검사 결과를 참고하여 판단을 내리는 것은 아니다. 전문가는 최종 판단을 내릴 때까지 연속적으로 이어진 일련의 의사 결정 단계에서 소수의 중요한 항목을 참조하여 판단을 내리고 그 판단에 따라 다음 의사 결정 단계로 진행하는 과정을 수행한다. 현재 연구되어온 임상 의사결정 지원 시스템은 이러한 의사 결정 모델을 반영하지 못하고 있으나 현재 본 논문에서 새롭게 제시된 방법은 이러한 모델을 반영하는 임상 의사결정 지원 시스템을 개발하는데 소수의 중요 검사 항목을 선정한다는 면에서 중요한 의미를 지니고 있다.

5. 결 론

본 논문에서는 랜덤 포레스트에 후진 제거법을 적용하여 질환 판별에 중요한 최적의 검사 항목을 선택하는 시스템을 제안하였다. 제안 시스템은 기존에 개발된 데이터 기반 기계 학습 방법인 의사결정 나무와 인공 신경망의 질환 판별 정확성과 비교해 보았을 때 높은 정확성을 보였으며, 특히 더 적은 수의 검사 항목을 사용하면서도 함에도 불구하고 유사하거나 더 높은 성능을 보이는 중요검사항목을 선정하였다. 이것은 제안된 방법이 많은 수의 임상 검사 항목 중 중요한 소수의 검사 항을 선별 가능하게 함을 의미하며 따라서 제안된 방법은 초기 질환 진단의 효율성을 증가시키고, 질환별 또는 사례별로 특화된 검사항목 특징을 추출하는데 사용할 수 있을 것이다. 또한 제안된 시스템은 주어진 모든 검사 항목의 결과값을 이용하여 한 번에 질환 진단을 하기보다는 중요한 소수의 검사 항목 결과치의 조합을 이용하여 단계별로 의사결정을 내리는 실제 임상 의사결정 과정을 반영한다고 할 수 있다.

본 연구에서 제안된 시스템을 다양한 사례에 적용하면 기존에 존재하는 질환과 수많은 검사항목의 관계 중 중요 검사 항목을 도출하는 과정에서 질환 조합 판별, 질환 간의 감별 등에 특화된 검사 항목을 도출하거나 특정 사례군의 조건에 적합한 검사항목을 도출하는 용도로도 쓰일 수 있을 것으로 기대된다.

감사의 글

이 논문은 산업자원부 차세대 신기술 개발 사업의 일환인 "바이오기술을 응용한 진단검사용 지능형로봇 기술 개발" 과제(10024715-2006-11)의 지원으로 수행되었습니다. 관계 부처에 감사드립니다.

참 고 문 헌

[1] DL Hudson, ME Cohen, A neural network learning algorithm, for development of diagnostic decision strategies, IEEE Engineering in Medicine and Biology, 1990; 12:1451-1452.

[2] SJ Fakih, TL Das, LEAD: A methodology for learning efficient approaches to medical diagnosis, IEEE Trans. Information Technology in Biomedicine, 2006; 10 (2):220-228.

[3] RO Duda, R.O., PE Hart., Pattern Classification and Scene Analysis, Wiley-Interscience, New York, 1973.

[4] DL Hudson, ME Cohen, Neural Networks and Artificial Intelligence in Biomedical Engineering, IEEE Press/Wiley, 1999.

[5] Breiman L: Random forests, Machine Learning 2001, 45 pp. 5-32.

[6] http://www.ics.uci.edu/~mlearn/MLRepository

[7] ME Cohen, DL Hudson, Combining Evidence in Hybrid Medical Decision Support Models, Proceeding of IEEE EMBS, 2007

[8] R.E. Abdel-Aal, Improved classification of medical data using abductive network committees trained on different feature subsets, Computer Methods and Programs in Biomedicine, 2005, 80 pp. 141-153

[9] http://www.r-project.org/

[10] W. Duch, R. Adamczak, K. Grabczewski, A new methodology of extraction, optimization and application of crisp and fuzzy logical rules, IEEE Trans. Neural Networks 12, 2001, pp. 277-306.

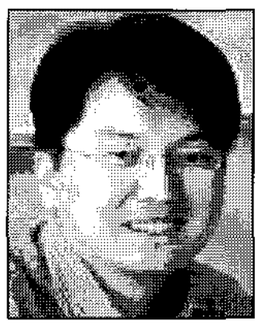
[11] F. Zhu, S. Guan, Feature selection for modular GA-based classification, Appl. Soft Comput, 2004, 4 pp. 381-393.

저 자 소 개

윤 태 균 (尹 泰 鈞)

1983년 5월 9일생, 2005년 한국정보통신대학교 IT경영학부 졸업, 현재 한국정보통신대학교 대학원 공학부 석박사 통합과정

Tel : 042-866-6815
Fax : 042-866-6814
E-mail : bacillusk@icu.ac.kr



이 관 수 (李 寬 洙)

1964년 11월 3일생, 1988년 서울대학교 동물학과 졸업, 1990년 한국과학기술원 생물공학과 졸업 (석사), 1993년 한국과학기술원 생물공학과 졸업 (박사), 2002년 ~ 현재 한국정보통신대학교 부교수

Tel : 042-866-6160
Fax : 042-866-6815
E-mail : gsyi@icu.ac.kr