

U-health 환경에 부합하는 통계기반의 질환 유무 판별 기법

송지수⁰ 한동수
한국정보통신대학교
{bulldozer⁰, dshan}@icu.ac.kr

A Statistical Method for Disease Identification in u-Health

Jisoo Song⁰ Dongsoo Han
Information and Communications University

요 약

U-healthcare는 종래의 헬스케어 영역에 유비쿼터스 기술을 접목하여 개인의 건강상태 관리를 도와주는 서비스이다. 이의 기반이 되는 기술인 질환 유무를 판별하는 기법은 그 동안 헬스케어 영역에 적용시켜 왔다. 하지만, 적시에 언제 어디서나 지속적인 모니터링이 요구되는 U-health환경에서는 기존의 기계학습기법을 그대로 적용하는 데에는 어려움이 있다. 본 논문에서는 통계기반의 질환 유무 판별 기법을 제안한다. 본 판별 기법은 질환 판별에 이용되는 생체신호와 신체증상의 종류로 배열 구조를 설정하고 축적된 데이터로부터 생체신호와 신체증상간의 쌍에 누적 빈도 수를 기록하여 학습한 뒤 고안한 판별식을 적용시켜 사용자의 질환을 판별하는 기법이다. 제한적인 검증이지만 약 360명의 실제 환자 데이터를 이용하여 기법을 검증하였고, 빠른 속도와 지속적인 개선이 가능한 기법임을 알 수 있었다. 추후 정확한 데이터를 기반으로 다른 기법과의 비교 검증으로 엄밀한 검증이 요구된다.

1 서론

종래의 헬스케어 영역에 초고속 통신망 인프라 구축, 휴대 가능한 생체신호 측정 센서 등의 개발과 같은 다양한 유비쿼터스(Ubiquitous) 기술을 접목하여 시간과 장소의 구애 받지 않고 개인의 건강상태를 지속적으로 관리할 수 있는 새로운 서비스가 대두되고 있다. 이러한 서비스를 U-healthcare라 부르는데 이를 이용하여 만성질환의 효율적인 관리나 의료비 감소 등에 큰 도움을 줄 것으로 기대하고 있다[1][2].

U-healthcare는 대개 측정센서로 사용자의 생체신호를 얻어오는 센싱단계, 생체신호를 1차적으로 가공하는 모니터링 단계, 데이터로부터 패턴인식, 데이터마이닝 등과 같은 기계학습(machine learning)으로부터 새로운 건강지표를 구하는 분석단계, 건강상태 변화를 사용자에게 알려주는 피드백단계와 같이 4단계로 구성되어 있다[3]. 특히, 축적된 생체신호 데이터로부터 새로운 건강지표를 구하는 분석단계가 이러한 서비스에 핵심이자 기반이 되는 기술이라고 볼 수 있다.

현재 많은 사람들이 종래에 헬스케어 영역에서 주로 사용해왔던 기계학습 기법을 분석단계에 적용하면 기대하는 효과를 거둘 수 있을 것이라고 판단하고 있다. 하지만, 아직 U-health분야의 연구가 초기단계에

불과하여 기존의 기법들을 U-health환경에 적용한 사례가 많이 부족하다. 또한 적시에 언제 어디서나 지속적인 모니터링으로 질환을 관리해야만 하는 U-healthcare 서비스의 특성상 학습시간이 많이 소요되고 판정결과 과정을 역 추적하기가 힘든 대부분의 기계학습 기법들을 분석단계에 그대로 적용하기에는 한계가 있다. 또한, 다양한 기계학습 기법이 있지만 축적된 데이터를 이용하여 개인의 건강상태를 예측 또는 판정하는 성능에서는 서로 큰 차이를 보이지 않으며, 그보다도 많은 변수(feature)들 중에서 일부를 선택하여 학습속도를 빠르게 하는 것이 더 큰 관심사인 것으로 알려져 있다. [4][5][6]

본 논문에서는 통계를 기반으로 하여 U-health 환경에 부합하는 질환 유무 판별 기법을 제안한다. U-health 환경에서 판별 기법이 갖추어야 할 조건인 판별 속도나 유연성, 지속적인 개선 등에 부합하게끔 통계학적 질환 판별 기법을 고안하였다. 본 판별 기법은, 특정질환과 관련이 있는 생체신호나 신체증상의 종류와 타임 그리고 가중치만을 정하면 질환에 관계없이 배열 구조를 구축할 수 있고, 축적된 환자 집단과 정상 집단의 데이터로 배열을 학습한다. 학습된 배열을 기반으로 하여 간단한 계산식을 통하여 사용자에게 질환 유무 정도를 제공할 수 있다. 배열을 사용하는 본 기법은 다양한 형태의 구조를 지원하면서도 상대적으로 빠른 학습과 판별 속도를 가지고 있고, 쉽게 판정결과의 역 추적도 가능한 장점이 있다.

본 논문에서는 서울백병원으로부터 약360명의 환자데이터 (생체신호 19종류, 신체증상 19종류)를 이용하여 스트레스 질환을 대상으로 판별 기법을 검증하였다. 학습데이터로 사용한 데이터가 명확하게 스트레스 집단과 정상집단으로 구분되어 있지 않아 정확한 검증을 할 수는 없었지만, 배열 구조를 변형시켜가면서 최적의 배열 구조를 찾아가는 것이 가능하였고, 학습과 판별에 있어서 기존에 알려진 기계학습 기법보다 빠른 속도를 보였다. 하지만 더 정확한 검증을 위해서는 추후 기계학습 기법과 성능을 비교하는 등의 더 엄밀한 검증이 필요하다.

2장에서는 종래에 헬스케어 영역에서 사용되었던 주요 기계학습 기법들에 관하여 설명하고 3장에서는 U-health 환경에서 판별기법의 조건을 나열한다. 4장에서는 본 논문에서 고안한 판별기법을 설명하고 3장에서 나열한 조건에 부합하는 부분을 자세히 알아본다. 5장에서는 실제 측정 데이터를 이용하여 검증한 결과를 설명하고 6장에서 결론을 내린다.

2 관련연구

질환 유무를 판별하기 위하여 그 동안 다양한 분야를 이용하여 시도를 해왔지만 대표적인 방법으로 기계학습 기법이 있다. 기계학습은 인공지능의 한 분야로써 컴퓨터를 학습시키는 것을 뜻하는 것이다. 자세히 말하면, 기계학습 기법은 데이터 집합을 분석하여 새로운 규칙성을 찾고 이를 프로그래밍하는 기법이다. 기계학습은 통계학과 동일하지만 데이터 분석하여 규칙을 찾는다는 점에서 비슷한 성격을 나타낸다. 기계학습과 통계학은 기계학습이 구현 시 알고리즘이 더 복잡하고 치밀 하다는 것에서 차이점을 나타낸다. 기계학습은 검색엔진, DNA 분석, 주식시장 분석 등 광범위한 분야에서 적용되고 있다. 질환 판별에도 지금까지 많은 기법들이 적용되어 왔고 그 유효성을 검증해왔다[10].

지수형 회귀분석 기법(Logistic Regression)은 아래의 식에 의해서 질환 판별 변수인 V값을 예측하는 기법이다.

$$P(V|X_1, X_2, \dots, X_n) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

여기에서 X는 연속 또는 불연속의 변수로써 질환 예측에 사용되는 변수들이고 β 값을 축적된 데이터들을 기반으로 변화해가면서 최적의 모델을 찾아내는 기법이다[7]. 비교적 직관적이고 널리 이용되는 방법이지만, 학습 데이터가 바뀔때 따라 β 값을 바꾸어야 하는 점에서 유연하지 못하며 시간이 걸린다는 단점이 있다.

단순 베이직인 분류 기법 (Naive Bayesian Classification)나 신경망 기법 (Neural Network) 역시 많은 분야에서 이용되고 있다. 두 기법들은 데이터들을

기반으로 각각의 상황에 따른 확률 값을 사람이 직접 정해야 하기 때문에 구축 시간을 필요로 하는 단점이 있지만, 정확한 확률 값인 경우에 상당한 정확도를 나타내는 것으로 알려져 있어서 심장 질환이나 폐렴 등 많은 의료분야에서 적용되고 있다[8].

규칙 기반 기법 (rule-based method)은 데이터와 결과의 일부를 보고 규칙성을 찾아내어서 전체적인 데이터에 적용시켜서 질환 판별을 하는 기법이다. 규칙을 자동으로 생성해주는 프로그램들이 최근 발전하면서 규칙을 유도하는 노력은 줄일 수가 있지만, 이 기법 역시 유연하지 못한 구조이고 환자상태를 모니터링 하기에는 적절하지 못한 기법이라고 할 수 있다.

KNN 기법(K-nearest neighbor method)은 학습을 통하여 전체적인 학습데이터를 분류시키고 예측단계에서 주위의 K개만큼의 데이터의 평균 값을 내어서 질환 유무 지표 값을 계산하는 기법이다. 이 기법은 모든 변수를 포함하여 판단할 수 있고 적절한 가중치를 부여할 수 있다는 점에서 정확한 추론이 가능하다. 하지만, 축적된 학습데이터를 분류하는 데에 시간이 많이 걸리는 단점이 있어서 많은 변수들 중에 일부를 선택하는 것이 U-health환경에서는 적절하다.

3 U-Health환경에서 질환 유무 판별 기법의 조건

3.1 실시간 처리

U-healthcare 서비스는 시간과 장소의 구애 없이 즉시 학습, 진단, 분석 그리고 관리 등이 이루어져야 하기 때문에, 서비스의 핵심 부분인 판별 기법은 빠르게 수행할 수 있게 최적화되어야 한다. 대부분의 기계학습 기법들은 학습하는 데 시간이 많이 소요가 되므로 수정이 필요하다.

3.2 유연한 구조

U-healthcare 서비스는 사용자의 건강을 종합적으로 체크하여 진단과 관리를 돕는 것을 목표로 하고 있다. 그러기 위해서는 다양한 질환에 대한 판별이 가능해야 하는 까닭에 판별 기법은 다양한 질환에 대해서 큰 변화 없이 대처가 가능해야 하고 각 질환 판별에 요구되는 서로 다른 변수들을 모두 포함시킬 수 있어야 한다. 몇몇 기법에서는 식에 포함시킬 수 있는 변수의 종류가 한정되어 있고 그에 따라서 몇몇 질환에는 적용할 수 없다는 것이 기존 기법들의 단점이다.

3.3 지속적인 개선

대부분의 질환 판별 기법들은 학습데이터의 성격에 의존하기 때문에 학습데이터의 변화에 민감하게 판별 성능이 바뀐다. 따라서, 끊임 없이 측정되는 데이터들을 학습데이터에 얼마나 효과적으로 반영을 하여 그 판별 기법을 개선해 나가는 것이 대단히 중요하다. 베이직인

기법이나 신경망 기법들은 학습데이터가 바뀌게 되면 사람이 직접 설정한 확률 값의 정확도가 떨어지게 되므로, 이 부분에 수정이 필요하다.

마찬가지로 2가지의 신체증상 종류와 각각 2개의 타입을 정하면 VSPA배열에서 행의 좌표는 아래와 같이 표현이 된다.

$$\{s_{11}, s_{12}, s_{21}, s_{22}\}$$

결과적으로, 스트레스 유무 판별을 위하여 두 종류의 4 × 4 VSPA 배열에 구조가 완성이 된다. 나아가서 축적된 데이터가 {p₁, p₂, p₃}를 스트레스 환자들의 집단이고, {p₄, p₅, p₆}를 정상인들의 집단으로 분포가 되어 있을 때 각 사람들의 데이터 범위에 맞게 <생체신호, 증상> 쌍으로 나타내면 스트레스 환자들의 데이터는
 Pair (p₁)={<b₁₁, s₁₁>, <b₁₁, s₂₂>, <b₂₂, s₁₁>, <b₂₂, s₂₂>}
 Pair (p₂)={<b₁₁, s₁₁>, <b₁₁, s₂₁>, <b₂₁, s₁₁>, <b₂₁, s₂₁>}
 Pair (p₃)={<b₁₂, s₁₂>, <b₁₂, s₂₂>, <b₂₂, s₁₂>, <b₂₂, s₂₂>}
 와 같이 표현되고 정상인들의 데이터는
 Pair (p₄)={<b₁₂, s₁₂>, <b₁₂, s₂₁>, <b₂₁, s₁₂>, <b₂₁, s₂₁>}
 Pair (p₅)={<b₁₂, s₁₂>, <b₁₂, s₂₁>, <b₂₂, s₁₂>, <b₂₂, s₂₁>}
 Pair (p₆)={<b₁₁, s₁₁>, <b₁₁, s₂₂>, <b₂₁, s₁₁>, <b₂₁, s₂₂>}
 와 같이 표현 가능하다. 이를 실제 VSPA 배열의 값으로써 누적 출현 횟수를 기록하면 아래와 같은 값을 가지는 VSPA배열이 생성된다[9].

		생체 신호				생체 신호			
		b ₁₁	b ₁₂	b ₂₁	b ₂₂	b ₁₁	b ₁₂	b ₂₁	b ₂₂
데이터	s ₁₁	2	0	1	1	1	0	1	0
	s ₁₂	0	1	0	1	0	2	1	1
	s ₂₁	1	0	1	0	0	2	1	0
	s ₂₂	1	1	0	2	1	0	1	1
		스트레스 집단				정상 집단			

그림 1. VSPA(스트레스) 배열과 VSPA(정상) 배열

3.4 판별과정의 모니터링

사용자의 건강관리를 위해서는 질환 유무를 판별하는 데에서 판별기법의 역할이 끝나서는 안되고, 사용자가 상태 변화를 느끼고 조치를 취할 수 있게 하는 일련의 과정을 제시할 수 있어야 한다. 물론, 판별 기법에서 이러한 과정을 직접적으로 지원하지는 않았지만 최소한 판별에 영향을 끼쳤던 변수나 영향을 끼치지 않았지만 그 동안에 측정 데이터와는 크게 벗어나는 결과가 아닌지를 판단하여 그 다음 단계에서 수행할 수 있는 기초자료로써 결과를 제공할 수 있어야만 한다.

4 통계학적 질환 판별 기법

기계학습 기법과 통계학적 기법은 축적 데이터를 기반으로 규칙성을 찾는 점에서 동일한 목적을 가지고 있으나, 기계학습 기법의 경우 복잡한 알고리즘을 사용하여 속도 면에서는 떨어진다. 본 논문에서는 배열을 기반으로 하여 상대적으로 간단한 통계학적 판별 기법을 제안한다. 이는 속도뿐만 아니라 모든 질환에 대처 가능한 유연한 구조를 가지고 있으며 지속적인 개선이 가능하다는 장점을 지니고 있다. 통계학적 기법은 일반적으로 학습 데이터가 늘어남에 따라 높은 성능을 나타내므로 U-health 환경에서는 적합한 기법이라고 할 수 있다.

4.1 VSPA 배열

VSPA 배열 (Vital sign-Symptom Pair Appearance Matrix)은 2차원 배열으로써 열(column)과 행(row)에 각각 질환의 생체신호 종류와 신체증상 종류로 구성된다. 생체신호나 신체증상의 종류와 타입(예를 들면 high, middle, low) 그리고 가중치만을 정하면 어떠한 형태도 단지 배열 구조를 구축할 수 있다. 한 질환의 배열 구조가 결정이 되면 질환 집단과 정상 집단의 두 가지 종류의 VSPA 배열을 구축한다. 배열의 원소에 값은 특정 쌍(Pair)이 축적된 데이터에서 출현한 횟수를 두 집단의 배열에 각각 기록한다. (단, 여기에서 축적된 데이터는 정확하게 두 집단으로 사전에 구분되어 있어야 한다) 완성된 두 종류의 VSPA 배열을 기반으로 실제 판별에 사용하는 DCAP배열을 생성할 수 있다.

실제 예로써 스트레스 유무 판별을 위하여 2가지의 서로 다른 생체신호 종류 {b₁, b₂}를 정하고 각각의 생체신호 b₁, b₂가 2개의 서로 다른 타입이 있다고 정할 때, VSPA배열에서 열의 좌표를 아래와 같이 표현할 수 있다.

$$\{b_{11}, b_{12}, b_{21}, b_{22}\}$$

4.2 DCAP 배열

4.1에서 생성된 두 종류의 VSPA배열을 기반으로 하여 실제 판별식에 이용되는 DCAP 배열 (Disease Combination Appearance Probability Matrix)을 생성할 수 있다. 배열의 원소에 값은 아래의 식과 같이 생체신호와 신체증상간의 쌍이 총 빈도수에 비하여 얼마나 자주 VSPA(스트레스) 배열에서 나타나는가에 대한 확률값으로 결정된다. 즉, 0에서 1사이의 값을 가진다.

$$DCAP_{ij} = \frac{VSPA(스트레스)_{ij}}{VSPA(스트레스)_{ij} + VSPA(정상)_{ij}}$$

4.1에서의 예를 바탕으로 DCAP배열을 생성하면 아래와 같다.

		생체 신호			
		b ₁₁	b ₁₂	b ₂₁	b ₂₂
강도	s ₁₁	0.67	0	0.5	1
	s ₁₂	0	0.33	0	0.5
키	s ₂₁	1	0	0.5	0
	s ₂₂	0.5	1	0	0.67

그림 2. DCAP 배열

4.3 질한 판별식

앞의 4.1과 4.2 배열은 실제 질한 판별을 위하여 학습 준비를 하는 것이고 축적된 데이터로 학습이 된 배열이 구축이 되면, 질한 판별이 가능해진다. 질한 판별을 하고자 하는 사용자를 p라고 하면 4.1에서의 표현방법과 같이 Pair(p)를 구성할 수 있고 아래의 식을 이용하여 계산하면 0부터 1까지의 질한 유무 정도를 판별할 수 있다. 여기에서 X(p)_i는 생체신호 종류 i의 가중치 값을 나타내고 Y(p)_j는 신체증상 종류 j의 가중치 값을 나타낸다. 가중치 값은 사전에 배열 구조를 설정할 때 값의 조절이 가능하고 이를 조절하여 최적의 모델을 만들어 낼 수 있다.

$$P(p \in \text{질한 그룹} | \text{Pair}(p)) = \frac{\sum_{j=1}^m \left[\frac{X(p)_j}{\sum_{j=1}^m X(p)_j} \cdot \left(\sum_{u=1}^n \frac{Y(p)_u \cdot \text{DCAP}_{i,u}}{\sum_{v=1}^n Y(p)_v} \right) \right]}{1}$$

위의 판별식은 질한 유무 판별뿐만 아니라 계산 과정을 역추적하여 판정 결과에 어느 정도의 영향을 끼쳤는지에 대한 정보를 제공하는 것이 가능하다. 즉, 판정 결과만을 내리는 1차적인 기능만을 하는 것이 아니라 사용자의 지속적인 건강 관리를 위해 필요한 처방에 도움을 주는 정보를 제공 가능한 장점이 있다.

위의 예시에 판별식을 적용시켜 보면 Pair(p)가 {<b₁₁, s₁₁>, <b₁₁, s₂₁>, <b₂₂, s₁₁>, <b₂₂, s₂₁>}와 같은 쌍을 가질 때 판별식의 값은 아래와 같다.

$$P(p \in \text{스트레스 그룹}) = \frac{0.5}{0.5+0.5} * \left(\frac{0.67*0.5}{0.5+0.5} + \frac{0.5*0.5}{0.5+0.5} \right) + \frac{0.5}{0.5+0.5} * \left(\frac{1*0.5}{0.5+0.5} + \frac{0.5*0.5}{0.5+0.5} \right) = 66.75\%$$

이는 사용자 p가 스트레스가 있을 확률이 66.75%라는 것을 의미한다. 단, 여기에서는 각 변수 별 가중치 값을 같다고 가정하였다. (X(p) = Y(p) = {0.5, 0.5})

4.4 배열구조 및 가중치 값의 재구성

본 판별기법의 구현은 학습의 속도를 빠르게 하고 유연한 구조를 위하여 축적된 데이터에서 배열 구조와 변

수들의 가중치 값을 XML문서로써 입력 받아 처리하게끔 하였다. 사용자는 자유롭게 변수들의 종류와 그 타입 그리고 가중치 값을 설정하면 그에 맞는 XML문서가 자동생성 되고 이를 본 판별기법에서 배열구조를 생성하여 학습과 판별단계를 거치게 된다.

본 판별기법을 포함한 대부분의 판별기법들은 어떤 구조나 가중치 값으로 설정하는가에 따라서 그 정확도와 성능이 크게 좌우된다. 본 판별기법은 XML문서의 자동생성부터 학습, 판별까지의 단계가 단 몇 초 내로 수행이 가능하고 XML문서에서 자유롭게 배열구조와 가중치 값을 조절 가능하기 때문에 이를 달리해가면서 성능을 서로 비교해보는 것이 수월하다. 비교를 통하여 특정 질한에 맞는 최적의 배열구조와 가중치 값을 설정할 수 있고, 이 과정은 프로그램을 통하여 거의 자동적으로 수행 가능하다.

5 검증

본 논문에서는 서울백병원으로부터 약360명의 환자데이터 (생체신호 19종류, 신체증상 19종류)를 이용하여 스트레스 질환을 대상으로 판별 기법을 검증하였다. 환자데이터가 스트레스 환자와 정상 환자의 구분이 명확히 되어있지 않아 문진을 통하여 스트레스 정도를 알아보는 SRI(Stress Response Index)값을 이용하여 환자를 구분하였다.

하지만, SRI값은 신체증상의 값과의 연관성이 커서 이를 변수로 하여 스트레스를 판별하는 것은 의미가 없었다. 결국, 2차원 배열에서 신체증상 종류를 제외한 1차원 배열만으로 스트레스 판별을 수행하였고, 유연하면서도 최적의 구조를 찾을 수 있다는 점을 검증하기 위하여 배열 구조를 조정하면서 각각 5번의 크로스 확인(cross validation)을 통하여 정확도를 측정하였고 이를 평균을 내어서 아래의 그림과 같이 정리하여 나타내었다.

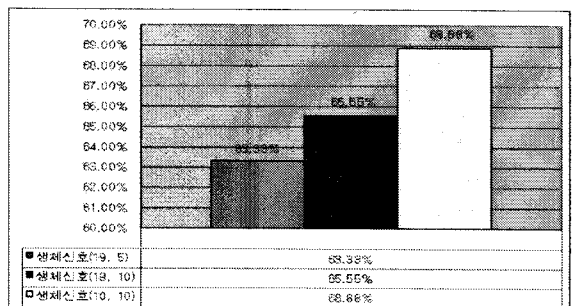


그림 3. DCAP 배열구조 변화에 따른 정확도

생체신호의 수와 그 타입의 수를 각각 n, m이라고 하였을 때 생체신호(n, m)으로 나타내었고, 위의 그림은 생체신호의 수 조절과 타입의 수 조절과 같은 구조 변

화로 성능에 최적의 배열 구조를 찾아낼 수 있음을 보인다. 또한, 각각의 판별 속도가 단 몇 초 내로 수행 가능하였고 이는 기존의 기계학습 기법의 속도를 상회하는 결과이다.

하지만 본 검증에서는 환자데이터의 특성상 2차원 배열을 대상으로 하지 못하였고, 실제 기계학습 기법과의 비교 검증을 수행하지 못한 점에서 제한적인 검증 결과일 뿐이다. 질환이 명확히 구분된 데이터를 준비하여 기계학습 기법과의 성능 비교를 통하여 더 정확한 검증이 추후에 필요하다.

6 결론

앞으로 다가올 U-healthcare 서비스는 빠르면서도 유연한 구조를 지녀야 하는 특성상 종래의 헬스케어 영역에 적용되어 왔던 기계학습 기법의 변화는 불가피하게 이루어져야 한다.

본 논문에서는 실시간 처리가 가능하면서 유연한 구조를 지니고 지속적인 개선으로 최적의 구조를 구축하는데에 용이한 통계기법을 기반으로 하여 질환 유무 판별 기법을 제안하였고 제한적이긴 하나 검증과정을 통하여 U-health환경에 부합될 수 있는 가능성을 확인하였다. 뿐만 아니라 데이터의 수가 늘어남에 따라 그 정확도와 성능이 높아지는 기법의 특성상 앞으로 개선가능성 역시 높을 것으로 기대한다. 하지만 추후에 정확하면서도 충분히 많은 환자데이터를 이용하여 기존의 다른 기법들과 비교하는 등의 더 엄밀한 검증이 요구된다.

7 참고문헌

- [1] 손미숙, u-health 서비스 지원을 위한 웨어러블 시스템, 전자통신동향분석, Vol. 21, No. 3, 2006
- [2] 한동수, 정명애, u-Healthcare 국내외 연구 동향 및 서비스 플랫폼, IITA, 2006
- [3] 정병주, u-Healthcare 서비스의 현황과 과제, 유비쿼터스사회연구시리즈 제10호, 2005
- [4] M.A. Klopotek, A New Bayesian Tree Learning Method with Reduced Time and Space Complexity, Fundamenta Informaticae, Vol 49, Issue 4, 2002
- [5] I. Inza, M. Merino, P. Larranaga, J. Quiroga, B. Sierra, M. Giral, Feature Subset by genetic algorithms and estimation of distribution algorithms A case study in the survival of cirrhotic patients treated with TIPS, Artificial Intelligence in Medicine 23, 2001
- [6] C.F. Cooper et al., An evaluation of machine-learning methods for predicting pneumonia mortality, Artificial Intelligence in Medicine 9, 1997
- [7] M.J. Fine, B.H. Hanusa, J.R. Lave, D.E. Singer, R.A.

- Stone, L.A. Weissfeld, C.M. Coley, T.J. Marrie, W.N. Kapoor, Comparison of severity of illness measures in patients with community-acquired pneumonia, J. Gen. Int. Med., 1995
- [8] M. Kukar, C. Grosz, I. Kononenko, J.J. Fettich, An Application of Machine Learning in the Diagnosis of Ischaemic Heart Disease, IEEE Symposium on Computer-Based Medical Systems, 1997
- [9] D.S. Han, J.S. Song, J. Matai, M.K. Lee, A Probability-Based Prediction Framework for Stress Identification, Healthcom2007, 2007
- [10] T.M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc., 1997.