

U-health 개인 맞춤형 질병예측 기법의 개선

Improvement of Personalized Diagnosis Method for U-Health

민병원*, 오용선**

목원대학교 산학협력단*, 목원대학교 정보통신공학과**

Byoung-Won Min(minfam@mokwon.ac.kr)*, Yong-Sun Oh(ysunoh@mokwon.ac.kr)**

요약

중래에 헬스케어 영역에서 주로 사용해왔던 기계학습 기법을 U-health 서비스 분석단계에 적용하기에는 여러 가지 문제점들이 있다. 첫째, 아직 U-health 분야의 연구가 초기단계에 불과하여 기존의 기법들을 U-health 환경에 적용한 사례가 매우 부족하다. 둘째, 기계학습 기법은 학습시간이 많이 소요되기 때문에 실시간으로 질환을 관리해야만 하는 U-health 서비스 환경에는 적용하기 어렵다. 셋째, 그동안 다양한 기계 학습 기법들이 제시되었으나 질환 연관변수에 가중치를 부여할 수 있는 방법이 없어, 개인 맞춤형 질병예측 시스템으로 구축할 수 없는 한계를 가진다.

본 논문에서는 이러한 문제점들을 개선하고, U-health 서비스 시스템의 바이오 데이터 분석 과정을 프로 세스로 해석하기 위하여, 개인 맞춤형 질병예측 기법인 PCADP를 제안하였다. 또한 이러한 PCADP를 바탕으로 U-health 데이터 및 서비스 명세의 의미 있는 표현을 위하여 U-health 온톨로지 프레임워크를 시멘 틱스형으로 모델링하였다. 또한 PCADP 예측 기법은 U-health 환경에서 판별 기법이 갖추어야 할 조건인 유연성과 실시간성이 기존의 방식에 비하여 향상되었고, 판별과정의 모니터링 및 시스템의 지속적인 개선 측면에서도 효율적으로 작용함을 확인하였다.

■ 중심어 : | 질병예측 | 개인맞춤형 | U-health |

Abstract

Applying the conventional machine-learning method which has been frequently used in health-care area has several fundamental problems for modern U-health service analysis. First of all, we are still lack of application examples of the traditional method for our modern U-health environment because of its short term history of U-health study. Second, it is difficult to apply the machine-learning method to our U-health service environment which requires real-time management of disease because the method spends a lot of time in the process of learning. Third, we cannot implement a personalized U-health diagnosis system using the conventional method because there is no way to assign weights on the disease-related variables although various kinds of machine-learning schemes have been proposed. In this paper, a novel diagnosis scheme PCADP is proposed to overcome the problems mentioned above. PCADP scheme is a personalized diagnosis method and it makes the bio-data analysis just a 'process' in the U-health service system. In addition, we offer a semantics modeling of the U-health ontology framework in order to describe U-health data and service specifications as meaningful representations based on this PCADP. The PCADP scheme is a kind of statistical diagnosis method which has characteristics of flexible structure, real-time processing, continuous improvement, and easy monitoring of decision process. Upto the best of authors' knowledge, the PCADP scheme and ontology framework proposed in this paper reveals one of the best characteristics of flexible structure, real-time processing, continuous improvement, and easy monitoring among recently developed U-health schemes.

■ keyword : | Computer Aided Diagnosis Probability | Personalized | U-health |

I. 서론

최근 건강이 사회의 중심가치가 됨에 따라 의료서비스 전문화 및 다양화 등 개인 맞춤형 의료서비스가 요구되고 있으며 대면 중심의 의료체제에서 예방 사후관리, 맞춤형 의료서비스 등 새로운 의료서비스 수요가 증대하고 있다. 또한 식생활 습관 변화, 스트레스 등으로 인한 질환의 증가에 따른 건강에 대한 관심 고조로 환자들이 소비 주체로서의 의료서비스를 요구하고 있으며, 병원 의료정보화로 고비용의 의료 기관 중심에서 일상 가정보건, 전문 의료서비스 중심으로 서비스가 변화하고 있다. 그러므로 의료기관을 방문하지 않고도 언제, 어디서나 건강상태 확인 및 의료서비스를 제공받기 원하는 니즈가 강하게 나타나, 향후 U-health 서비스에 대한 수요가 크게 확대될 것으로 예상된다[1-6].

U-health 서비스는 대개 측정센서로 사용자의 생체신호를 얻어오는 센싱 단계, 생체신호를 1차적으로 가공하는 모니터링 단계, 데이터로부터 패턴인식, 데이터 마이닝과 같은 기계학습(machine learning)으로부터 새로운 건강지표를 구하는 분석 단계, 건강 상태를 사용자에게 알려주는 피드백 단계와 같이 4단계로 구성된다. 특히, 축적된 생체신호 데이터로부터 새로운 건강지표를 구하는 분석단계가 U-health 서비스의 핵심이자 기반이 되는 기술이라고 볼 수 있다[7-9].

현재 많은 사람들이 종래에 헬스케어 영역에서 주로 사용해왔던 기계학습 기법을 분석단계에 적용하면 기대하는 효과를 거둘 수 있을 것이라고 판단하고 있다. 하지만, 아직 U-health 분야의 연구가 초기단계에 불과하여 기존의 기법들을 U-health 환경에 적용한 사례가 많이 부족하다. 또한 적시에 언제 어디서나 지속적인 모니터링으로 질환을 관리해야만 하는 U-health 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, 그 보다도 많은 변수들 중에서 일부를 선택하여 학습속도를 빠르게 하는 것이 더

큰 관심사인 것으로 알려져 있다[10-12].

본 논문에서는 U-health 환경에 부합하는 개인 맞춤형 질병예측 기법(PCADP; Personalized Computer Aided Diagnosis Probability)을 제안한다. 이는 U-health 환경에서 판별 기법이 갖추어야 할 조건인 유연한 구조, 실시간 처리, 지속적인 개선, 판별과정의 모니터링 등에 부합하는 통계학적 질병예측 방법이다. 기존의 시스템에서 문제점으로 지적되고 있는 데이터 및 응용서비스의 신뢰도를 확보하고, 스트레스, 비만, 당뇨와 같은 일반적인 질병 예측까지도 수용할 수 있는 확장된 예측시스템을 구축하기 위하여, 본 기법에서는 IA(Interrelation Appearance) 배열과 CI(Combination Interrelation) 배열을 새롭게 고안하여 활용하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 U-Health 환경에 부합한 질환 유무 판별 기법에 대한 국내외 관련 연구들을 고찰하였고, 제3장에서는 U-health 개인 맞춤형 질병예측 기법에 대해서 기술한다. 제4장에서는 제안된 내용들의 구체적인 가능성을 제시하기 위하여 검증실험을 수행하였고, 실험결과를 기반으로 종래의 질병예측 기법들과 비교 평가 및 고찰을 한다. 마지막으로 제5장에서는 본 논문의 결과와 향후 연구과제를 제시하였다.

II. 기존의 U-Health 질환 유무 판별 기법에 대한 고찰

2.1 지수형 회귀분석 기법

지수형 회귀분석 기법(logistic regression)은 식(2.1)로 표현되는 확률값에 의하여 질환 판별 변수인 V 값을 예측하는 기법이다.

$$P(V|X_1, \dots, X_n) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)} \quad (2.1)$$

여기서 X 는 연속 또는 불연속의 변수로서 질환 예측에 사용되는 변수들이고, β 값은 축적된 데이터들을 기반으로 변화해가면서 최적의 모델을 찾아가기 위한 과

라미터들이다[13]. 이 기법은 비교적 직관적이고 널리 이용되는 방법이지만, 학습 데이터가 바뀔 때 따라 지속적으로 β 값들을 바꾸어야 하는 점에서 유연하지 못하며, 질환 예측을 위한 분석 시간이 비교적 길다는 단점이 있다.

2.2 의사결정나무 기법

의사결정나무 기법(decision trees method)은 [그림 1]의 가상 시나리오 모델에서와 같이 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석방법이다[14-16]. 이 예에서 첫 번째 노드(node)에서는 분류기준 없이 모든 병원 입원 대상자 중 암환자가 차지하는 비율이 20%이었다. 그러나 두 번째 노드에서는 연령이 50세 이상인 경우 암환자는 35%로 증가되었다. 이와 같이 의미 있는 변수를 기준으로 하는 지속적인 분류를 통해 암환자일 가능성 높은 군을 가려내는 모델을 형성하게 된다. 즉 이 모델에 의하면 50세 이상이고 양성종양이 있으며 가족력이 있는 경우에는 암환자로 분류될 확률이 20%에서 60%로 증가된 예를 보여준다. 이와 같이 의사결정나무 모델은 분류 또는 예측 과정이 나무구조에 의한 추론 규칙에 의해서 표현되기 때문에 의사결정과정을 쉽게 이해하고 설명할 수 있는 장점을 가지고 있다.

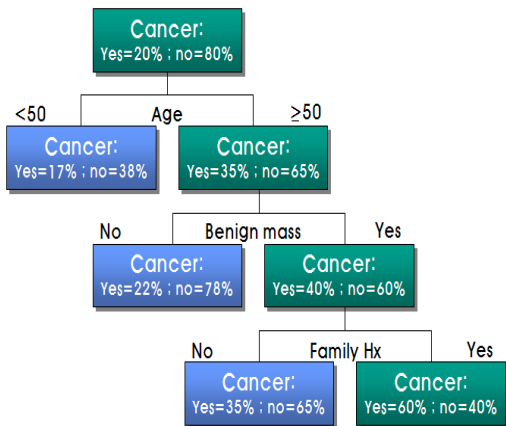


그림 1. 의사결정나무 모델의 예

III. U-health 개인 맞춤형 질병예측 기법

3.1 질병예측 알고리즘 구조

개인 맞춤형 질병예측 알고리즘에서는 질환집단과 정상집단의 생체신호와 증상간의 상호관계 패턴을 각각 저장하고 예측을 하고자 하는 대상의 질병을 예측하기 위하여 그 대상의 생체신호와 증상을 이용한 확률식을 도출해 낸다. [그림 2]는 전체적인 질병예측 알고리즘 구조를 나타낸다.

학습단계에서는 질환집단과 정상집단의 생체신호와 증상간의 상호관계 패턴을 각각 저장한다. 의사결정나무단계에서는 학습단계에서 만들어진 자료를 이용하여 노드(node)와 에지(edge)로 연결된 뿌리나무를 만들기 위해서 질병의 분류에 따라 <생체신호조합, 증상조합> 쌍의 데이터를 분류하여 조합한 후 allelic association 과 Hardy-Weinberg equilibrium 정보를 통해 데이터에서 중요한 DT값(Decision Tree value)을 찾는다. 예측 단계에서는 예측을 하고자 하는 대상의 질환을 판단하기 위하여 그 대상의 생체신호와 증상을 이용한 확률식을 고안하고 질환 정도를 예측한다.

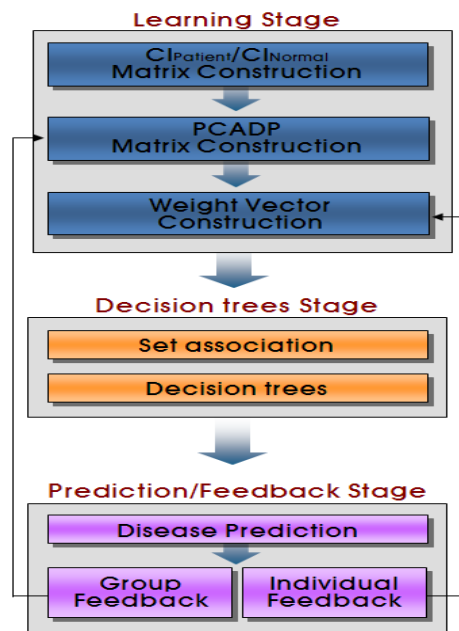


그림 2. 개인 맞춤형 질병예측 알고리즘 구조

학습단계는 다시 세 가지의 세부과정으로 나뉜다. 첫 번째 과정에서는 질환집단과 정상집단의 <생체신호 조합, 증상 조합> 쌍의 출현빈도를 기록할 수 있는 CI(Combination Interrelation) 배열을 작성한다. 따라서 CI 배열은 생체신호 및 증상뿐 아니라 생체신호 조합 및 증상 조합 정보를 취급한다. 두 번째 과정에서는 <생체신호 조합, 증상 조합> 쌍 들이 질환을 가진 집단에서 나타날 확률 정보를 계산하여 PCADP 배열을 생성한다. 세 번째 과정에서는 각 개인에게 맞춤형 서비스를 제공하기 위하여 각각의 사용자마다 WA(Weight Assignment) 행렬을 준비한다. 의사결정나무단계에서는 질병유무와 관계 정도를 측정된 χ^2 -통계치인 AA 통계치와 Hardy-weinberg equilibrium을 귀무가설로 하여 <생체신호 조합, 증상 조합> 쌍의 분산 정도를 측정된 χ^2 -통계치인 HWD 통계치를 적용하여 DT값을 찾게 된다. 예측단계에서는 학습단계에서 생성된 PCADP 배열과 Weight Vector를 기반으로 특정 <생체신호 조합, 증상 조합> 쌍을 가지는 대상의 질환 유무를 판정하는 단계이다. 마지막 피드백단계에서는 질환 유무 판정 예측이 종료된 뒤, 예측된 결과에 대한 사용자의 반응을 통해서 시스템을 보정하는 과정으로 개인사용자 피드백 메커니즘과 집단 사용자 메커니즘으로 구분하여 질환 유무 판별 프레임워크의 정확도를 진화적으로 개선하며, 사용자 맞춤 진단의 정밀화가 가능하다.

3.2 학습단계

학습단계(Learning Stage)는 예측단계에서 사용될 PCADP 배열과 Weight Vector를 준비하는 단계이다. 본 논문에서 Weight Vector는 해당 질병의 전문가의 도움을 얻어서 값을 결정하는 것으로 가정한다. 본 절에서는 학습을 통해서 <생체신호 조합, 증상 조합> 쌍이 특정 질병을 가진 집단에서 나타날 확률 정보를 담고 있는 PCADP 배열을 만드는 과정을 중심으로 기술하기로 한다. PCADP 배열을 얻기 위해서는 환자 군 및 정상 군으로부터 <생체신호 조합, 증상 조합> 쌍의 출현 빈도를 담고 있는 $CI_{Patient}$ 및 CI_{Normal} 배열을 우선적으로 구축하여야 한다.

3.2.1 IA 배열

이해의 편의상 특정 군에 있어서 <생체신호 조합, 증상 조합> 쌍의 출현빈도 기록에 사용되는 CI 배열을 설명하기 전에 <생체신호, 증상> 쌍의 출현빈도를 기록할 수 있는 IA(Interrelation Appearance) 배열을 소개하기로 한다. IA 배열의 각 행은 하나의 생체신호와 연관되며, IA 배열의 각 열은 하나의 증상과 연관된다. 따라서 IA 배열의 각 원소는 하나의 <생체신호, 증상> 쌍과 연관되어 <생체신호, 증상> 쌍의 출현빈도를 기록할 수 있다.

CI 배열은 IA 배열을 확장하여 생체신호 및 증상뿐 아니라 생체신호 조합 및 증상 조합 정보를 취급한다. 따라서 CI 배열에는 <생체신호, 증상> 쌍의 출현빈도가 아닌 <생체신호 조합, 증상 조합> 쌍의 출현빈도를 기록한다.

3.2.2 CI 배열

질환 군과 정상 군에 대한 CI(Combination Interrelation) 배열, $CI_{Patient}$ 배열과 CI_{Normal} 배열이 얻어지면 <생체신호 조합, 증상 조합> 쌍 들이 질환을 가진 집단에서 나타날 확률 정보를 계산하여 CI 배열과 동일한 구조를 가진 PCADP 배열을 생성하고 기록하는 것이 가능하다. 본 논문에서는 원소 값을 결정하는 식에서 정상집단과 질환집단 각각의 학습 데이터를 고려하지 못한 점을 보완하기 위하여 정규화 기능을 계산식에 반영하였다. $CI_{Patient}$ 배열과 CI_{Normal} 배열에 누적 기록된 <생체신호 조합, 증상 조합> 쌍의 총 개수를 각각 $Total_{Patient}$, $Total_{Normal}$ 이라고 하면 PCADP 배열의 각 원소에는 식 (3.1)에 의하여 계산된 값이 저장된다.

$$PCADP_{ij} = \frac{Total_{Normal} \cdot (CI_{Patient})_{ij}}{Total_{Normal} \cdot (CI_{Patient})_{ij} + Total_{Normal} (CI_{Patient})_{ij}} \quad (3.1)$$

$PCADP_{ij}$ 확률값이 얻어지면 PCADP 배열을 이용하여 특정 <생체신호 조합, 증상 조합>쌍이 나타내는 상태를 가진 대상이 질환 군에 포함될 확률을 계산하는 것이 가능하다. 하지만 PCADP 배열 정보만을 사용하

여 대상의 질환 유무를 판단하는 것은 모든 생체신호 및 증상에 대해서 동일한 가중치를 적용하는 것과 동일한 효과를 가지고 있다. 즉 PCADP 배열에는 생체신호나 증상에 대한 가중치가 반영되어 있지 않아 질환에 따라서는 특정 생체신호나 증상이 여타의 생체신호나 증상에 비해 더 많은 영향을 미치는 것을 반영할 수 없다. 그러나 질환에 따라서는 특정 생체신호나 증상이 다른 생체신호나 증상에 비해 상대적으로 높은 영향을 주는 경우도 자주 발생한다고 볼 때 PCADP 배열의 정보만을 가지고 질환 유무를 판단하는 것은 적절하지 않다. 따라서 학습의 세 번째 단계에서는 가중치 정보를 반영하지 못하는 PCADP 배열의 한계를 보완하기 위해서 생체신호와 증상이 특정 질환에 관련되는 Weight Vector라는 개념을 도입한다.

3.2.3 Weight Vector

앞서 설명한 바와 같이 몇몇 생체신호나 증상은 예측 단계에서 다른 생체신호나 조건에 비해서 더 많은 가중치를 주는 것이 적절할 수 있다. 하지만 PCADP 배열은 이러한 가중치값 부여에 유연하지 않으므로 본 논문에서는 Weight Vector라는 개념을 도입하여 질환 예측 시 생체신호 및 증상별 가중치를 반영할 수 있도록 하였다. 예측에 사용되는 Weight Vector는 생체신호 Weight Vector와 증상 Weight Vector가 있고 여기서는 이것들을 각각 Weight생체신호 = {wb1, wb2, ..., wbn}, Weight증상 = {ws1, ws2, ..., wsm} 로 나타내기로 한다. Weight Vector의 초기값은 질환에 대한 전문가에 의뢰하여 결정하는 것으로 가정한다. 추후에는 별도의 실험을 통해서 최적화된 Weight Vector 값을 찾는 방법도 생각할 수 있다. 여기에서는 위의 Weight Vector 에는 생체신호 조합과 증상 조합에 대한 가중치를 별도로 기술하지 않았다. 생체신호 조합과 증상 조합에 대한 가중치를 해당 조합을 구성하는 생체신호 또는 증상의 Weight Vector 만 주어지면 평균값 산출과 정규화를 통해서 쉽게 구할 수 있다.

3.3 의사결정나무단계

본 논문에서 사용하고 있는 개인 맞춤형 질병예측 기

법은 복잡한 데이터에서 일정한 패턴을 찾아 이를 확률적으로 예측하는 방법이다. 그러나 다량의 데이터가 입력으로 들어오면 연산에 오랜 시간이 걸리고 패턴을 찾기가 어려워진다는 한계가 있다. 그래서 이러한 문제점을 해결하기 위해서 집합조합(set association)과 의사결정나무를 이용하여 질병예측시스템에 들어오는 데이터의 수를 줄여서 다량의 데이터에도 적용 가능하며 예측의 정확도를 높일 수 있도록 구성하였다.

일차적으로 set association과 decision tree를 이용하여 학습단계에서 만들어진 중요 PCADP 배열값을 선별한 후 이를 질환 예측 확률식의 입력값으로 주어, 다량의 데이터를 이용한 질환의 발병 여부 예측을 가능하게 하고 예측에 소요되는 시간을 줄이며 궁극적으로 질환 발병 예측의 효율성을 높이고자 하였다. set association 방법은 [그림 3]과 같다.

Set association은 allelic association과 Hardy-weinberg equilibrium 정보를 통해 데이터에서 중요한 DT값(decision tree value)을 찾는 방법이다. 이를 위해서 질병유무와 관계 정도를 측정할 χ^2 -통계치인 AA 통계치와 Hardy-weinberg equilibrium을 귀무가설로 하여 <생체신호 조합, 증상조합> 쌍의 분산 정도를 측정할 χ^2 -통계치인 HWD 통계치가 필요하다. 특별히 환자 그룹에서 HWD 편차가 높은 DT값들은 질병과 관련이 있다는 의미이다. 반면에 정상그룹에서 HWD 편차가 높은 DT값들은 오류를 의미하므로 해당 DT값을 제거하거나 '0'으로 맞추어주는 트리밍(trimming) 작업이 필요하다[15][16].

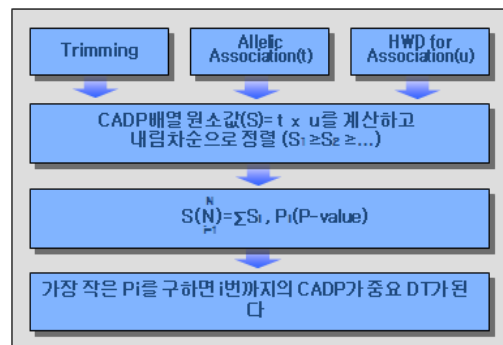


그림 3. Set association 방법의 요약

3.4 예측단계

학습단계를 통해서 PCADP 배열과 Weight Vector의 생성이 완료되면 질환 유무를 판정하는 것이 가능하다. 본 단계에서는 학습단계와 의사결정나무단계에서 생성된 PCADP 배열, Weight Vector 그리고 DT값을 기반으로 특정 <생체신호조합, 증상조합> 쌍 집합을 가지는 대상의 질환 유무를 판정하는 확률식을 고안하였다. 고안된 확률식을 통해서 계산되는 값은 대상이 질환집단에 속할 확률로서, 본 논문에서는 이 확률값을 해당 대상이 특정 질환을 가질 확률로 간주한다.

질환 유무 판정 확률식 설명을 위해서 학습단계에서 PCADP 배열과 생체신호 Weight Vector X 와 증상 Weight Vector Y 가 준비되어 있다고 가정한다. 또한, $CP(p)$ 를 질환 유무가 검증되지 않은 대상(subject) p 의 <생체신호 조합, 증상 조합> 쌍 집합으로 정의한다. 그리고 $CP(p)$ 에 해당하는 값을 추출하여 PCADP 배열에서 찾아 배열 형태로 정리하여 얻어진 배열은 $PCADP(p)$ 배열이라고 한다. 그리고 $PCADP(p)$ 배열에 해당하는 생체신호 및 증상 Weight Vector는 $X(p)$ 와 $Y(p)$ 로 표기한다. 최종적으로 얻어진 배열과 증상 벡터가 각각 $m \times PCADP(p)$, m -Length 생체신호 Weight Vector $X(p)$, n -Length 증상 Weight Vector $Y(p)$ 라 할 때, $CP(p)$ 를 가진 대상의 질환 유무 판별은 식(3.2)을 사용하여 계산한다.

$$P(p \in Patient\ Group | CP(p)) = \sum_{i=1}^m \left[\frac{X(p)_i}{\sum_{j=1}^m X(p)_j} \left(\frac{\sum_{u=1}^n Y(p)_u DT(p)_i PCADP(p)_{i,u}^i}{\sum_{v=1}^n Y(p)_v} \right) \right] \quad (3.2)$$

3.5 피드백단계

앞에서 제안한 방식에 의해서 질환 유무를 판단한 것이 항상 올바른 결과라고 단정할 수 없다. 특히 유비쿼터스 환경에서는 제한된 생체신호 정보와 증상정보를 바탕으로 예측된 결과라고 볼 때 판정 결과의 신뢰성 확보가 쉽지 않다. 본 논문에서는 유비쿼터스 환경에서의 질환 유무 판별 프레임워크가 가지고 있는 취약성을

보완하고 질환 유무 판별 프레임워크의 정확도를 진화적으로 개선하기 위한 사용자 피드백 메커니즘에 관하여 기술한다. 사용자 피드백은 질환 유무 판정 예측이 종료된 뒤, 예측된 결과에 대한 사용자의 반응을 통해서 시스템을 보정하는 과정으로 질환 유무 판별 프레임워크의 마지막 단계이기도 하다. 여기에서는 사용자 피드백 메커니즘을 크게 개인 사용자 피드백 메커니즘과 집단 사용자 피드백 메커니즘으로 구분하여 기술한다.

3.5.1 개인 사용자 피드백 메커니즘

개인 사용자 피드백 메커니즘은 스트레스 질환과 같이 개인차가 많은 질환을 질환 유무 판별 프레임워크에 수용하기 위해서 고안한 것으로, 사용자가 자신의 생체신호와 증상을 질환 유무 판별 프레임워크에 전송한 뒤 얻어진 질환 유무 판정 결과를 받아보고 그 결과에 대한 사용자 피드백 정보를 제공한다. [그림 4]에 도시된 개인 사용자 피드백 구동과정에 있어서, 사용자가 판정된 결과에 대해서 긍정적인 경우에는 별다른 피드백 메커니즘이 구동되지 않고 해당 정보만 뒤에 기술할 집단 사용자 피드백 메커니즘을 위해서 기록으로 보관한다. 하지만 질환 유무 판정 결과에 대해서 사용자가 수긍하지 못하는 경우 사용자는 자신에게만 두드러지게 나타난다고 생각하는 생체신호 및 증상의 가중치를 조절할 수 있도록 한다.

질환 유무 판별 프레임워크는 사용자가 조정하는 가중치에 따라 개인별 생체신호 Weight Vector와 증상 Weight Vector를 생성하고 각각의 weight 값 변동에 따른 질환 유무 판정 결과의 변동을 즉각적으로 사용자에게 전달한다. 사용자는 자신이 변화시키는 가중치가 자신의 질환 유무 판정에 어느 정도 영향을 미치는지를 확인하고 자신에게 적절한 Weight Vector 값을 선택하게 된다. 이러한 Weight Vector 값 변경 과정은 개인에 따라 한 번에 끝날 수도 있고 몇 차례 반복될 수도 있다. 질환 유무 판정 시스템은 해당 사용자의 다음 번 질환 유무 판정에 있어서는 [그림 4]와 같이 변화된 개인의 Weight Vector를 기반으로 질환 유무를 판정함으로써 개인 사용자 피드백 메커니즘이 구성된다.

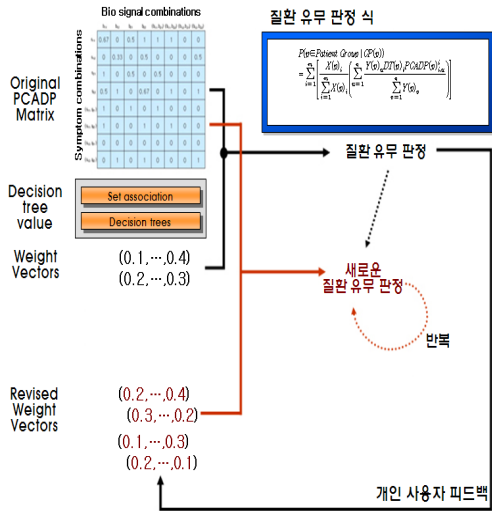


그림 4. 개인 사용자 피드백 메커니즘

3.5.2 집단 사용자 피드백 메커니즘

사용자 피드백 메커니즘을 통해서 얻어지는 정보에서 일부 데이터는 질환 유무 판정을 위한 새로운 학습 집단으로 활용하는 것도 가능하다. 집단 사용자 피드백 메커니즘은 이와 같이 새롭게 얻어지는 학습 집단을 활용하여 기존의 시스템을 진화시키는 일련의 과정이다. 집단 사용자 메커니즘은 피드백의 결과가 모든 사용자의 판정에 영향을 미친다는 점에서 개인의 질환 판정에 영향을 미치고 타인의 질환 판정에는 영향을 미치지 않는 개인 사용자 피드백 메커니즘과는 구분된다. 구체적으로 집단 사용자 피드백 메커니즘에서는 새롭게 얻어지는 새로운 학습 집단이 기존의 질환 유무 판정 시스템 구축에 사용된 학습 집단에 비해서 더욱 정제된 정보라는 것이 확인되면 이것을 기존 판별 프레임워크의 PCADP 배열에 반영하여 진화된 PCADP 배열을 얻게 된다. 기존의 질환 유무 판정 프레임워크는 이와 같이 진화된 PCADP 배열을 사용함으로써 새로운 시스템으로 진화하게 된다. [그림 5]는 이러한 집단 사용자 피드백 메커니즘을 개념적으로 나타내고 있다.

집단 사용자 피드백 메커니즘에서 두 가지 이슈가 제기된다. 첫 번째 이슈는, 기존의 PCADP 배열을 사용하여 예측한 결과의 피드백을 통해서 얻어진 데이터에서 어떻게 정상집단과 질환집단을 분리하여 학습 집단을

구성할 것인가 하는 문제이다. 또 다른 이슈는 새롭게 구성된 학습 집단을 통해서 얻어진 PCADP 배열과 기존의 PCADP 배열의 오류치를 어떻게 측정하고 오류치 측정에 사용할 질환집단을 어떻게 구성할 것인가 하는 문제이다.

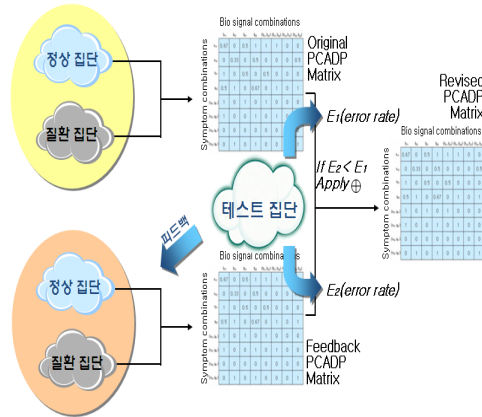


그림 5. 집단 사용자 피드백 메커니즘

IV. PCADP 알고리즘 검증

4.1 검증실험에 사용한 데이터

개인 맞춤형 질병예측기법인 PCADP의 검증실험에 사용한 데이터에 관하여 간략히 설명을 하고 검증실험 결과를 기술한다. 스트레스 판별에 관한 검증실험은 데이터가 충분치 않아 기본적인 검증실험만을 진행하였다.

본 검증실험에서 사용한 스트레스 실험 데이터는 모두 서울 백병원에 우종민 교수의 스트레스 클리닉에서 제공받은 자료를 바탕으로 구성하였다[17]. 스트레스 자료는 피실험자의 나이, 성별, 신장, 몸무게 등의 신체 기록과 혈압, 혈당, 체온 등의 생리학적 자료와 광맥과 측정기를 이용하여 RR(Respiration Rate) Interval을 측정된 자료 그리고 SRI(Stress Reservation Index) 설문지의 결과로 구성되어 있다. PCADP 배열을 검증하기 위하여 백병원으로부터 받은 자료를 선별하고 분석하는 과정을 수행하여 PCADP 배열을 구성하는 학습데이터로 사용하였다.

스트레스의 경우 성별이나 나이가 다를 때 스트레스

수치의 차이가 많이 나는 것으로 판단하여 성별과 나이 별로 따로 PCADP 배열을 구축하여 학습데이터를 학습하였고, RR Interval을 다양한 방법으로 분석하여 SDNN(Standard Deviation of the NN intervals), LF, HF등의 생체신호의 요소들을 얻었다. SRI는 각각의 질문과 연관성이 있는 증상들을 선별하여 각 질문의 점수를 바탕으로 증상의 요소들을 얻었다. 스트레스 검증 실험에 사용한 생체신호와 증상들은 각각 19개로 구성하였고 다음과 같다.

- 생체신호 : GLUCOSE, SYSTOLIC, DIASTOLIC, BODY_FAT, HF_NU, LF_NU, LF_HF, HF_MSEC.2, LF_MSEC.2, VLF_MSEC.2, TINN_SEC, HRV_INDEX, PNN50_PERCENTAGE, RMSSD_SEC, CV_PERCENTAGE, SDNN_SEC, MEAN_RR_SEC, HEART_RATE

- 증상 : UNEASINESS, TACHYCARDIA, STIMULUS_EVASION, SOLICITUDE, SCATTERBRAINED, PALPITATION, OVERSENSITIVE, OUTRAGE, IRRESOLUTION, HEADACHE, HAND_TREMOR, GLOOM, EVADING_CROWD, EMPTINESS, DYSPEPSIA, DIZZINESS, CHRONIC_FATIGUE_SYNDROME, ABUSE, ABDOMINAL_PAIN

4.2 검증실험 결과

스트레스 검증실험은 크게 두 가지 방법으로 진행하였다. PCADP 배열을 구축하기 위해서는 스트레스 집단과 정상집단으로 뚜렷이 구분된 학습 데이터를 필요로 한다. 하지만, 백병원으로부터 받은 자료는 스트레스 집단과 정상 집단으로 구분한 상태에서 생리학적 자료나 RR Interval 또는 SRI값을 얻은 것이 아니므로 검증 실험에서 임의로 데이터를 두 집단으로 구분할 필요가 있었다.

첫 번째 방법으로는 실제 스트레스 정도를 측정하는 데 사용되는 SRI의 값을 이용하여 구분하는 것이다. 하지만 실제로 알려져 있는 대로 SRI값으로 스트레스를 구분하는 방법은 매우 엄격하여 스트레스 집단의 수가 너무 부족하게 되므로 18명의 SRI값을 적당한 비율로 분리하여 각각 스트레스 집단과 정상집단의 학습데이터를 사용하였다. 본 실험에서는 그 비율을 15%, 30%, 50%로 정하여 각각의 정확도를 비교해보았다.

두 번째 방법으로는 스트레스 집단을 구분하는 데 사용되는 SDNN이나 LF/HF 값과 같은 HRV feature를 이용하여 집단을 구분하는 것이다. SDNN은 time domain method로부터 얻어지는 feature로써 가장 쉽게 구할 수 있는 값이면서도 기본적인 값으로 스트레스를 받는 사람들이 일반적으로 낮은 값을 가지는 것으로 알려져 있다. LF/HF값은 frequency domain method로부터 얻어지는 feature이고 교감신경계(sympathetic)의 변화에 민감하게 반응을 한다. 일반적으로 스트레스는 교감신경계 활동을 증대시키게 되고, 결과적으로 LF/HF ratio 값을 상승시키는 것으로 알려져 있다. 모든 검증 실험은 5-cross validation 방법을 사용하여 예측 정확도를 측정하는 방법으로 PCADP 배열을 검증하였다.

4.2.1 SRI로 구분

상하위 15%, 30%, 50%로 나누어 실험하였고, 다양한 형태의 PCADP 배열을 구성하여 각각의 정확도를 통하여 결과를 분석하고자 하였다. 그 형태는 총 7가지로 다음과 같다.

- Biosignal x Symptom (19 x 19)
- Biosignal x 1 (19 x 1)
- Symptom x 1 (19 x 1)
- Biosignal x Biosignal (19 x 19)
- HRV x 1 (14 x 1)
- HRV x HRV (14 x 14)
- HRV x Symptom (14 x 19)

위의 7가지 형태를 각각 스트레스 집단을 구분하는 세가지 방법인 상하위 15%, 30%, 50%에 의해서 검증 실험을 거쳤다. 즉, 각 형태마다 세 번씩 총 21가지의 검증 실험을 진행하였고 결과는 [그림 6]과 같다.

[그림 6]에서와 같이 Symptom을 반영한 배열에서의 정확도는 모두 90%이상을 나타내게 되면서 SRI와 그로부터 추출한 Symptom간의 correlation이 지나치게 높다는 것을 알게 되었다. 결국 SRI로 스트레스 집단과 정상 집단은 구분한 학습 데이터로는 정확한 검증 실험을 거칠 수 없었고, Symptom을 제외한 Biosignal x 1 배열에 정확도를 상승시키는 실험을 진행하였다. 정확도를 상승시킬 수 있는 방법으로는 사용한 feature들의

수를 조절하거나 feature들의 range의 수를 조절하여 정확도 상승을 꾀할 수가 있다.

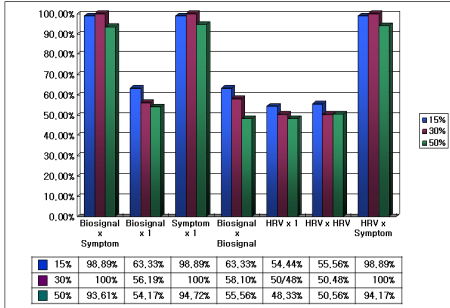


그림 6. PCADP 배열을 이용한 SRI 실험 결과

[그림 7]의 결과에서 Biosignal2는 19개에서 9개로 선택하여 줄인 새로운 Biosignal들을 의미하고 괄호에 값은 range의 수를 의미한다. 따라서 range의 수를 기존의 5개에서 10개로 늘리면서 또한 불필요한 feature를 제거함으로써 정확도를 상승시키는 효과를 가져왔다.

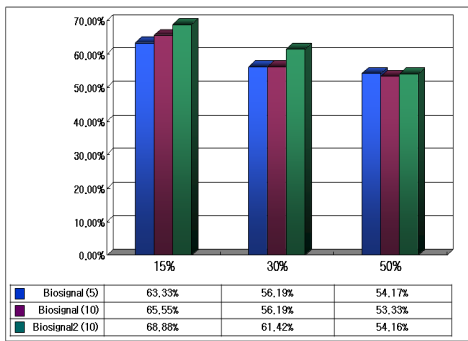


그림 7. Biosignal의 range수 변경 후 실험 결과

4.2.2 SDNN과 LF/HF로 구분

SRI의 검증방법과 동일하게 SDNN과 LF/HF값의 각각 상위 30% 기준으로 두 집단을 구분하여 PCADP 배열의 학습데이터로써 사용하였다. (단, SDNN의 경우는 상위 30%가 정상집단, 하위 30%가 스트레스집단이고, LF/HF의 경우는 반대로 상위 30%가 스트레스집단, 하위 30%가 정상집단이다).

생체신호 feature들의 수는 전체 19개에서 집단에 구

분하는데 사용한 feature를 제외한 18개를 사용하였으며, 증상 feature들의 수는 동일하게 19개를 사용하였다. Range의 수는 기존의 방법과 마찬가지로 5등분하여 사용하였다.

SDNN으로 구분한 3번의 실험과 LF/HF로 구분한 3번의 실험으로 구성하여 검증실험을 하였고, HRV feature들간의 상호적인 correlation이 생길 것으로 판단되어서 HRV feature가 아닌 glucose(혈당농도)로도 집단을 구분하여 3번의 실험을 더한 총 9번의 실험을 하였으며, 이를 SRI로 나눈 검증결과와 비교해 보았다. 그 형태는 총 9가지로 다음과 같다.

- Biosignal x Symptom (18 x 19) - SDNN
- Biosignal x 1 (18 x 1) - SDNN
- Symptom x 1 (19 x 1) - SDNN
- Biosignal x Symptom (18 x 19) - LF/HF
- Biosignal x 1 (18 x 1) - LF/HF
- Symptom x 1 (19 x 1) - LF/HF
- Biosignal x Symptom (18 x 19) - GLUCOSE
- Biosignal x 1 (18 x 1) - GLUCOSE
- Symptom x 1 (19 x 1) - GLUCOSE

총 9번의 실험들의 결과는 [그림 8]과 같다. SRI와 비슷하게 SDNN과 LF/HF가 각각 biosignal과의 correlation을 가지고 있는 것을 확인하였다. 따라서 기존의 데이터로 하나의 방법만으로 집단을 구분하는 것은 힘들고, 각각의 방법을 이용하여 가상의 스트레스 집단을 구성하여 검증실험을 진행하는 것이 정확한 검증결과를 얻을 수 있을 것으로 판단된다.

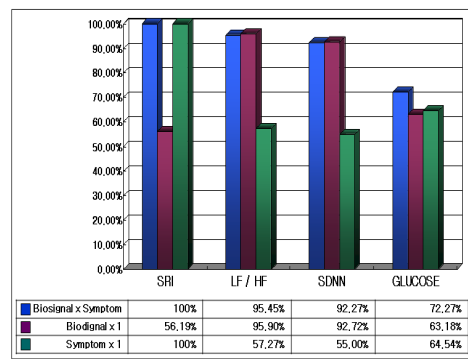


그림 8. SDNN과 LF/HF로 구분하여 실험 결과

4.2.3 스트레스 확률 분포도

PCADP 배열 matrix에서는 [그림 9]와 같이 스트레스군과 비스트레스군의 분포를 가지고 있다. 이 데이터는 주기적으로 진화하므로써 일반적인 생체신호와 문진에 따른 스트레스 확률을 측정하는 기본 자료로 사용된다.

예측 단계는 학습 과정에서 구성되는 바이오 시그널과 증상 쌍의 출현 패턴 정보에 기반하여 이루어진다. 학습단계에서는 정상집단과 환자집단의 생체신호 데이터와 신체증상의 상호관계 패턴의 분리를 시도한다. 본 연구에서는 상호관계 인식을 위한 충분한 학습 데이터가 이 단계에서 이용 가능하다고 가정한다.

학습단계에서는 정상집단과 환자집단으로부터 생체신호와 신체증상의 상호작용 빈도를 계산하여 조합 상호작용(Combination Interaction, CI) 배열에 저장한다. 각각의 CI 배열은 정해진 정상집단과 환자집단의 생체신호와 신체증상의 조합 패턴을 그것이 나타난 빈도수와 함께 저장한다.

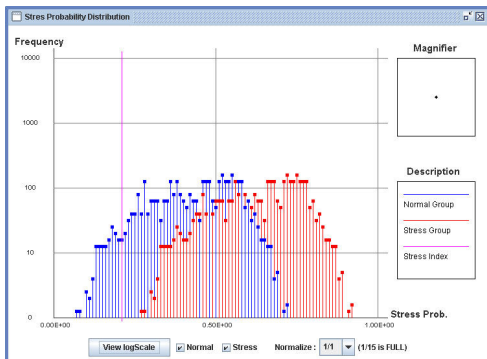


그림 9. 스트레스 확률 분포도

환자 그룹과 정상그룹의 CI 배열은 PCADP 배열로 통합된다. PCADP 배열에서 각 요소는 그 요소의 조합 패턴이 환자집단에서 나타날 확률값을 가진다. 예측 단계에서는 질병 유무를 판단하고자 하는 사람의 생체신호와 신체증상을 예측시스템에 입력하고 그 예측시스템은 그 사람이 환자집단으로 판단될 확률을 계산하여 보여준다. PCADP 배열이 CI 배열을 기반으로 만들어지기 때문에, 본 연구는 어떤 생체신호와 신체증상의

조합이 특정 진단에 기여하는지를 쉽게 판단할 수 있다. [그림 10]에서 색으로 표시된 부분은 대응되는 생체신호와 증상 조합의 요소들이고 0부터 1사이의 확률이 PCADP 계산식에 의해 계산된다.

Key	Value
BF in profile	1.0
BS in profile	1.0
BT in profile	0.0
Blood Sugar	1.0
Body Fat	84.0
Body Temperature	36.2999992371
Constipation	0
Drinking	0
Dyspepsia	4
ECC	3.8100113869
ECC in profile	1.0
HR in profile	1.0
Headache	0
Heart Rate	63
Muscular Pain	0
PHASE	1
Smoking	0
UserID	s1
Weight of BF	1
Weight of BS	1
Weight of BT	0
Weight of ECC	0
Weight of HR	0
근육통 in profile	1.0
두통 in profile	1.0
변비 in profile	0.0
소화불량 in profile	4.0
흡주 in profile	3.0
흡연 in profile	1.0

Key	Value
0:normal, 1:error	0
Bio signal	Body Fat
PHASE	1
ST_SYM0BS0	9.795306315845552
ST_SYM0BS1	12.648308155412224
ST_SYM0BS2	15.50130999497889
ST_SYM1BS0	11.846750495724448
ST_SYM1BS1	16.672399321505782
ST_SYM1BS2	21.49804814728711
ST_SYM2BS0	14.305766366970003
ST_SYM2BS1	17.973911589270003
ST_SYM2BS2	21.64205681157
ST_SYM3BS0	14.998638242293337
ST_SYM3BS1	21.112213612793337
ST_SYM3BS2	27.225788983293338
ST_SYM4BS0	16.887053745525558
ST_SYM4BS1	23.67176383457601
ST_SYM4BS2	30.456473923526442
ST_SYM5BS0	18.422240449773334
ST_SYM5BS1	25.823742364992007
ST_SYM5BS2	33.22524428021087
Stress Probability	21.112213%
Symptom	Drinking
UserKey	s11150894585234
Weight of bio signal	1.0
Weight of symptom	3.0

그림 10. PCADP 배열에서 얻어진 사용자의 건강지수 결과

4.3 비교평가 및 고찰

본 논문에서는 통계를 기반으로 U-health 환경에 부합하는 개인 맞춤형 질병예측 기법인 PCADP 알고리즘

을 제안하였고, U-health 환경에서 판별 기법이 갖추어야 할 조건인 유연한 구조, 실시간 처리, 지속적인 개선, 판별과정의 모니터링 등에 부합하도록 필수적인 요구사항으로 설정하였다. 또한, 보안도 고려되어야 하는 중요한 요구사항 이지만 본 논문에서는 다루지 않았다. 향후 U-health 서비스 시스템이 상용화되어 보급되기 위해서는 꼭 해결되어야 하는 이슈이다.

기존의 기계학습을 통한 질병예측방법은 인공지능의 한 분야로서 컴퓨터를 학습시켜서 데이터를 분석하여 새로운 규칙성을 찾고 이를 프로그래밍 하는 기법이다. 기계학습은 통계학과 동일하진 않지만 데이터를 분석하여 규칙을 찾는다는 점에서 유사한 특징을 지닌다. 한편, 기계학습은 구현 시 알고리즘이 더 복잡하고 치밀하다는 점에서 통계학습과 차이점이 있다.

[표 1]은 U-health 환경에서 판별기법이 갖추어야할 조건을 기존의 기계학습과 본 논문에서 제안한 PCADP 질병예측 기법을 비교평가 하여 정리한 내용이다.

지수형 회귀분석기법과 신경망기법은 질병예측에 사용되는 변수들을 축적해가면서 최적의 모델을 찾아가는 모델이다. 비교적 직관적이고 널리 이용되는 방법이지만, 데이터 차원의 문제와 블랙박스(Black Box'syndrome) 모델에 따른 각 변수간의 상호 연관성을 알기 어렵기 때문에 유연하지 못하며, 판별 시간이 많이 소요된다는 문제점이 있다. 또한 블랙박스 모델을 사용하므로 판별과정의 모니터링이 어렵다.

표 1. 기계학습기법과 PCADP 기법의 비교평가

예측기법	조건	유연한 구조	실시간 처리	지속적인 개선	판별과정의 모니터링
PCADP기법		○	○	○	○
지수형 회귀분석기법		△	△	△	N/A
신경망기법		△	△	○	N/A
의사결정규칙기법		○	○	○	N/A
규칙기반기법		△	○	○	N/A

○ : 지원됨, × : 지원안됨, △ : 불충분함, N/A : 적용할 수 없음

의사결정규칙기법은 분류 또는 예측과정이 나무구조에 의한 추론 규칙에 의해서 표현되기 때문에 데이터의 차원 문제와 변수 간 연관성 해석 그리고 유전형의 분포에 상관없으며 처리속도가 빠르고 결과를 쉽게 해석

할 수 있는 방법이다. 하지만 의사결정과정의 모니터링 방법이 제시되지 않아 U-health 환경에 부합되는 예측 기법으로서는 한계가 있다.

규칙기반 기법은 If-Then과 연산자 and로 이루어진 규칙(Rule)을 모은 규칙 세트(Rule Set)를 만들고 이 규칙 세트의 규칙에 따라 데이터를 맞춰보고 맞으면 데이터의 결과를 해당 규칙의 결과로 예측하는 방법이다. 규칙을 자동으로 생성해주는 프로그램들이 발전 하면서 규칙을 유도하는 노력을 줄일 수가 있고 명목형(Categorical) 변수 처리에 맞고 블랙박스(Blackbox) 모델이 아니므로 규칙 세트만으로도 사람이 그 내용을 이해하기 쉽다는 장점 때문에 다양한 분야에서 널리 사용되고 있다. 하지만 이 기법 역시 유연하지 못한 구조이고 환자상태를 모니터링 하기에는 적절하지 못한 기법이다. 다음은 본 논문에서 제시한 설계 목표가 어떻게 작용하여 만족시키는 지 정리한 내용이다.

4.3.1 유연한 구조

U-health 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. 본 논문에서는 유연한 구조를 지원하기 위해 온톨로지 기반 데이터 관리와 PCADP 배열 역시 이 시스템의 유연성을 향상시키는 특징이다. 온톨로지의 구조와 PCADP 배열 값은 사용자 피드백 정보에 따라 유동적으로 변화된다.

4.3.2 실시간 처리

U-health 서비스는 시간과 장소에 구애 없이 즉시 학습, 진단, 분석 그리고 관리 등이 이루어져야 하기 때문에, 서비스 핵심 부분인 판별 기법은 빠르게 수행할 수 있게 최적화되어야만 한다. 본 논문에서는 실시간 처리를 지원하기 위해 통계기반의 질병 예측 기법을 고안하였으며, 다량의 데이터가 입력으로 들어오면 연산에 오랜 시간이 걸리고 패턴을 찾기가 어려워진다는 한계를

해결하기 위해서 집합조합(set association)과 의사결정 나무를 이용하여 질병예측시스템에 들어오는 데이터의 수를 줄여서 다량의 데이터에도 적용 가능하며 예측의 정확도를 높일 수 있도록 구성하였다.

4.3.3 지속적인 개선

대부분의 질환 판별 기법들은 학습데이터의 성격에 의존하기 때문에 학습데이터의 변화에 민감하게 판별 성능이 바뀐다. 따라서, 끊임없이 측정되는 데이터들을 학습데이터에 얼마나 효과적으로 반영하여 그 판별 기법을 개선해나가는지 하는 것이 대단히 중요하다.

본 논문에서 제안한 U-health 온톨로지는 새로운 데이터와 사용자로부터의 피드백 정보가 축적됨에 따라 지속적으로 개선된다. 또한 PCADP 배열 역시 진화한다. 분석 결과에 대한 피드백에 따라 PCADP 배열은 보다 복잡하고 개인화된 분석을 위해 데이터를 변경함으로써 질병예측 방법의 진화를 촉진한다.

4.3.4 판별과정의 모니터링

사용자의 건강관리를 위해서는 단순히 질환 유무를 판별하는 것으로 판별기법의 역할이 끝나서는 안 되고, 사용자가 상태 변화를 인지하고 그에 따른 조치를 취할 수 있게 하는 일련의 과정을 제시하여야 한다. 물론, 판별기법에서 이러한 과정을 직접적으로 지원하지는 않았지만 최소한 판별에 영향을 끼쳤던 변수나 영향을 끼치지 않았지만 평상의 측정 데이터와는 크게 벗어나는 결과 등을 가려내어 그 다음 단계에서 수행할 수 있는 기초자료로 이를 제공할 수 있어야만 한다.

온톨로지 관리기 모니터는 서비스 내에서 주어진 정보를 활용하여 온톨로지 매니저가 온톨로지의 검색, 추론, 수집 등을 통하여 결과를 도출하는 과정을 시각적으로 보여주기 위한 도구이다. 온톨로지 관리기의 활용 방법으로 스트레스 그룹 판정을 위한 온톨로지 활용 서비스를 구현하였다. 스트레스 그룹 판정을 위해 각 바이오 신호별 가중치를 온톨로지를 통해 체계적으로 관리하기 위한 서비스로 증상 온톨로지와 온톨로지 API를 활용하였다.

V. 결론

U-health 서비스는 대개 측정센서로 사용자의 생체 신호를 얻어오는 센싱단계, 생체신호를 1차적으로 가공하는 모니터링 단계, 데이터로부터 패턴인식, 데이터마닝과 같은 기계학습으로부터 새로운 건강지표를 구하는 분석단계, 건강 상태를 사용자에게 알려주는 피드백단계 등 4단계로 구성되어 있다. 특히, 축적된 생체신호 데이터로부터 새로운 건강지표를 구하는 분석단계가 이러한 서비스에 핵심이자 기반이 되는 기술이라고 볼 수 있다.

종래에 헬스케어 영역에서 주로 사용해왔던 기계학습 기법을 U-health 서비스 분석단계에 적용하기에는 여러 가지 문제점들이 있다. 첫째, 아직 U-health 분야의 연구가 초기단계에 불과하여 기존의 기법들을 U-health 환경에 적용한 사례가 많이 부족하다. 둘째, 기계학습 기법은 학습시간이 많이 소요되기 때문에 실시간으로 질환을 관리해야만 하는 U-health 서비스 환경에는 적용하기 어렵다. 셋째, 그동안 다양한 기계학습 기법들이 제시되었으나 질환 연관변수에 가중치를 부여할 수 있는 방법이 없어, 개인 맞춤형 질병예측 시스템으로 구축할 수 없는 한계를 가진다.

본 논문에서는 이러한 문제점들을 개선하고, U-health 서비스 시스템의 바이오 데이터 분석 과정을 프로세스로 해석하기 위하여, 온톨로지 기반 개인 맞춤형 질병예측 기법인 PCADP를 제안하였다. 또한 이러한 PCADP를 바탕으로 U-health 데이터 및 서비스 명세의 의미 있는 표현을 위하여 U-health 온톨로지 프레임워크를 시멘틱스형로 모델링하였다.

이를 실현한 알고리즘은 데이터 및 응용서비스 신뢰도를 확보하고 스트레스, 비만, 당뇨와 같은 일반적인 질병 예측까지 수용할 수 있도록 설계되었으며, 특정 질환과 관련 있는 생체신호나 신체증상의 종류와 형태 그리고 가중치만을 정하면 질환에 관계없이 배열 구조를 구축하여 다양한 형태의 구조를 지원하면서도 상대적으로 빠른 학습과 판별 속도를 나타내며, 판정 결과의 역 추적도 매우 용이하다는 장점이 있다. 예측 정확도의 향상을 위하여 정상집단과 환자집단의 크기 차이

를 고려하여 PCADP 계산식에 정규화 기능을 부여하였고, 개인 맞춤형 진단의 정밀화를 위한 피드백 메커니즘을 고도화하였다.

본 논문에서 제안한 PCADP 기법의 성능 및 효율성을 검증하기 위하여, 5-cross validation 방법을 사용하여 예측 정확도를 측정하는 방법으로 PCADP 기법을 검증하였다. 이는 가상의 질환집단을 구성하여 PCADP 검증실험을 진행하는 것으로, 본 논문의 제안 기법이 기존의 방식에 비하여 월등히 정확한 결과를 나타냄을 확인하였다. 또한 PCADP 예측 기법은 U-health 환경에서 판별 기법이 갖추어야 할 조건인 유연성과 실시성이 기존의 방식에 비하여 향상되었고, 판별과정의 모니터링 및 시스템의 지속적인 개선 측면에서도 효율적으로 작용함을 확인하였다.

본 논문에서 제안된 U-health 서비스의 상용화 단계까지는 많은 연구가 남아 있고, 시스템을 보다 안전하고 신뢰성 있게 만들기 위해 이루어져야 할 미해결의 작업들이 남아있다. 어떻게 건강 데이터를 안전하게 관리할 것인가?, 어떻게 모바일 기기들과 U-health 서버 간에 교환되는 메시지를 압축하여 성능을 향상시킬 것인가? 하는 등의 문제들에 대한 연구가 필요하다.

참 고 문 헌

- [1] 한국홈네트워킹산업협회, IPTV, u-Home, u-Health 서비스 중심으로, 도서출판 진한엠앤비, 2009.
- [2] u-Health Trials 2008, 한국정보사회진흥원, 2008.
- [3] u-Health 서비스 배경 및 추진현황, 한국정보사회진흥원, 2008.
- [4] 유비쿼터스 시대의 보건의료서비스 전망, 서울대학교병원, 2006.
- [5] 2008 u-Health 기술 워크샵, u-Health산업활성화포럼, 2008.
- [6] 2008 u-Health 로드쇼, u-Health산업활성화포럼, 2008.
- [7] 손미숙, u-health 서비스 지원을 위한 웨어러블 시스템, 전자통신동향분석, Vol.21, No.3, 2006.
- [8] 한동수, 정명애, u-Healthcare 국내외 연구동향 및 서비스 플랫폼, IITA, 2006.
- [9] 정병주, u-Healthcare 서비스의 현황과 과제, 유비쿼터스사회연구시리즈 제10호, 2006.
- [10] M. A. Klopotek, A New Bayesian Tree Learning Method with Reduced Time and Space Complexity, Fundamenta Informaticae, Vol 49, Issue 4, 2002.
- [11] I. inza, M. merino, P. Larranage, J. Quiroga, B. Sierra, and 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.
- [12] C. F. Cooper, An evaluation of machine learning methods for predicting pneumonia mortality, Artificial Intelligence in Medicine 9, 1997.
- [13] M. J. Fine, B. H. Hanusa, J. R. Lave, D. E. Singer, R. A. Stone, L. A. Weissfeld, C. M. Coley, T. J. Marrie, and W. N. Kapoor, Comparison of severity of illness measured in patients with community-acquired pneumonia, J.Gen. Int. Med., 1995.
- [14] D. H. Kim, S. Uhmm, S. W. Cho, K. B. Hahm, J. Kim, Predictive Model for Chronic Hepatitis Susceptibility from Single Nucleotide Polymorphism, BIOINFO, pp.35-38, 2006.
- [15] 최현주, 김승현, 위규범, 신경망을 이용한 천식 발병 예측 모델, 제27회 한국정보처리학회 춘계 학술발표논문집, 제14권, 제1호, pp.597-600, 2007.
- [16] 김명기, 김승현, 위규범, 의사결정규칙을 이용한 복합 질환의 예측, 제27회 한국정보처리학회 춘계 학술발표논문집, 제14권, 제1호, pp.593-596, 2007.
- [17] 우종민, 심박동수 변이(Heart Rate Variability) 측정법의 개념과 임상적 활용, 정신신체의학논문

집 제12권, 제1호, pp.3-14, 2004.

저 자 소 개

민 병 원(Byoung-Won Min)

중신회원



- 2005년 2월 : 중앙대학교 컴퓨터소프트웨어학과(공학석사)
- 2010년 2월 : 목원대학교 IT공학과(공학박사)
- 2005년 4월 ~ 2008년 2월 : 영동대학교 컴퓨터공학과 전임강사
- 2008년 3월 ~ 2010년 2월 : 목원대학교 산학협력단 전임강사
- 2010년 3월 ~ 현재 : 목원대학교 산학협력단 특임교수
<관심분야> : 온톨로지, U-Health, 모바일콘텐츠, SaaS, 모바일 클라우드, 스마트 도서관

오 용 선(Yong-Sun Oh)

중신회원



- 1983년 2월 : 연세대학교 공과대학 전자공학과(공학사)
- 1985년 2월 : 연세대학교 대학원 전자공학과(공학석사)
- 1992년 2월 : 연세대학교 대학원 전자공학과(공학박사)
- 2007년 9월 ~ 2008년 8월 : 한국전자통신연구원(ETRI) 초빙연구원
- 1988년 3월 ~ 현재 : 목원대학교 공과대학 정보통신공학과 교수
- 2006년 7월 ~ 현재 : 한국콘텐츠학회 회장
<관심분야> : 디지털통신시스템, 정보공학, 멀티미디어 콘텐츠, 맞춤형 이러닝