진단 자동화를 위한 PSO 분류화 시스템의 설계 [1]

맹보연, 최옥주, 이민수
이화여자대학교 컴퓨터공학과
e-mail : {mngby, pensica}@ewhain.net,
mlee@ewha.ac.kr

Design of a Particle Swarm Optimization-based Classification System for automatic diagnosis

Boyeon Meang, Ok-ju Choi, Minsoo Lee
Dept. of Computer Science and Engineering, Ewha womans University

요 약

무선 센서들의 진보에 따라 환자의 상태를 모니터링 하거나 정보를 저장 후 원거리에 있는 의사들의 진단 제공이 가능하게 되었다. 하지만 환자의 데이터의 양에 비해 의사의 수가 적으므로 환자가 진단을 제공 받는데 시간적인 한계가 있다. 따라서 본 연구에서는 환자의 상태를 1 차적으로 자동 진단하는 시스템을 제안한다. 전체 데이터의 적용을 위해 Circadian rhythm 에 기반한 데이터 직접방법을 제안하고 데이터를 효율적으로 분류하기 위해 PSO(Particle Swarm Optimization)을 기반으로 하는 분류화 알고리즘을 적용하여 시스템의 수행속도 향상을 도모하였다.

1. 서론

최근 기술의 진보는 무선 센서들을 더욱 더 작고 가볍게 만들었으며 저전력 소모를 가능하게 했다. 이러한 기술 진보로 인해 무선 센서들은 넓은 범위의 사용 가능성과 함께 의료 분야에서 사용 되어 왔다. 사람이 착용한 옷이나 센서 등 여러 장비 등을 통해 몸을 중심으로 센서와 PDA 와 같은 휴대 가능한 기기 간에 데이터의 결합이나 교환을 하는 것을 WBAN(Wireless Body Area Network)라고 한다[1]. 현재 국내외에서 연구 및 개발되는 의료 분야 응용 서비스들은 WBAN 의 생체 및 행위정보 데이터에 대하여 주로 단순한 모니터링을 하고 있거나 의료 서버에 단순히 저장 후 의사의 진단을 제공하는 수준으로 기능이 한정되어 있다. 또한 환자의 수에 비해 진단을 제공할 수 있는 의사의 수가 적기 때문에 환자가 의사의 진단을 받기까지의 시간적 한계가 있다. 따라서 환자의 생체 데이터를 자동으로 분류하여 1 차 진단을 내릴 수 있는 시스템이 필요하다. 이에 착안하여 본 연구에서는 WBAN 환경에서 환자의 데이터를 실시간으로 측정 하여 의료 서버에 저장 후 자동으로 상태를 판단하는 시스템을 제안한다. 데이터를 측정하여 무선으로 전송 받아 데이터베이스에 저장한 후 수행속도를 향상시키기 위한 PSO 기반 분류화 기법으로 룰을 구축하여 질병에 대해 사전 검진하고 예방 또는 만성적인 노약자나 잠재적인 환자를 찾아내는 분류화 방법을 설계한다.

논문의 구성은 다음과 같다. 2 장에서 관련연구로 KNN 분류화 기법과 베이지안 분류화 기법을 소개하고, 3 장에서 환자의 데이터를 자동으로 분류하는 시스템을 설계하는 방법을 제안한다. 마지막으로 4 장에서 결론 및 향후 연구 과제를 기술 한다.

2. 관련 연구

2.1 인접 이웃 분류화 기법

인접 이웃 분류는 특정 훈련 사례를 사용하여 예측을 수행하는 사례기반 학습이다. 인접 이웃 분류는 지역 정보를 기반으로 예측을 수행한다. 분류하고자 하는 데이터가 입력되면 보유하고 있는 참조 데이터 셋 중에서 가장 가까운 k 개의 데이터의 클래스를 통하여 분류하는 비모수 분류 방법이다[2]. 인접 이웃 분류는 모델 구축은 하지 않지만, 시험 사례와 훈련 사례들 사이의 인접성 값을 일일이 계산하여야 하므로 분류 비용이 상당히 크다. 반면에 본 연구에서 제시한 분류 방법은 모델 구축에 상당한 컴퓨팅 자원을 필요로 하지만 모델 구축 후, 시험 사례의 분류는 상당히 빠르다.

2.2 베이지안 분류화 기법

베이지안 분류 방법은 베이지안 분석 확률 이론을 사용하여 예제 집합에 대한 최적의 분류를 구하도록 시도하는 학습 알고리즘의 일종으로 최소 오류율을 가지는 학습 시스템을 가지는 것이 목표이다. 이미 알고 있는 확률들을 이용하여 특정한 데이터가 어느

[1] 이 논문은 정부(교육과학기술부)의 재원으로 한국과학재단 지원을 받아 수행된 연구임(No. R01-2008-000-20029-0).

카테고리에 들어갈 지를 결정한다. 즉, 확률이 가장 높은 카테고리에 들어갈 가능성이 가장 높다. 베이지안 분류화는 모델을 생성하는데 상당한 노력과 시간이 요구되지만 새로운 변수를 추가하기가 쉽고 불완전한 데이터를 다루는 데 적합하다[3].

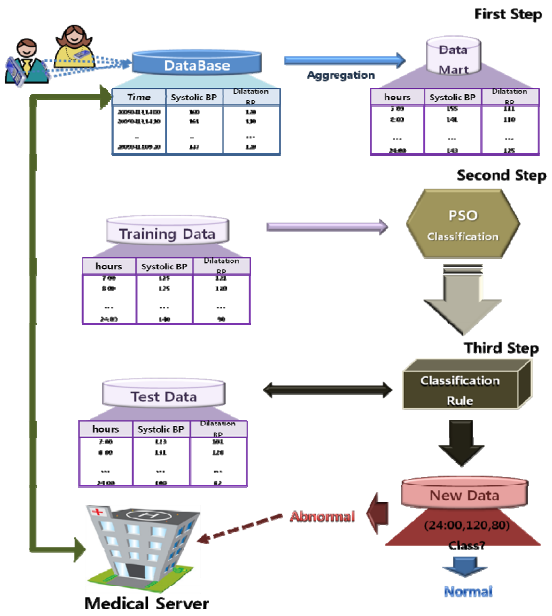
3. 진단 자동화를 위한 PSO 분류화 시스템

환자의 데이터가 실시간으로 데이터베이스에 저장되기 때문에 저장된 데이터의 양이 크므로 의사가 환자의 상태를 진단하고 정보를 제공하는데 시간적인 한계가 있다. 본 연구에서는 대용량의 데이터베이스에서 환자의 상태를 자동으로 진단하기 위한 PSO 기반의 분류화 시스템을 제안한다. 다음은 PSO 분류화 알고리즘이다[4].

```
Input : training data

iteration number

Generate initial particles
(= classification rule) randomly

for iteration number do
   for each particle do
      calculate the fitness value using training data
     find the pbest
   end for

   find the gbest

   for each particle do
      calculate velocity
      update particle
   endfor
endfor

output : rule corresponding to gBest
```

(그림 1) PSO 분류화 알고리즘

다음은 대용량의 PSO 분류화 알고리즘을 이용한 자동 진단 시스템이다.



(그림 2) PSO 분류화를 적용한 자동 진단 시스템

첫 번째 단계에서 효율적으로 데이터를 관리하기 위해 Circadian rhythm[5]에 따라 데이터를 시간대별, 낮과 밤, 24 시간으로 집적하여 데이터마트에 저장하고 룰을 랜덤으로 생성한다. 두 번째 단계에서는 훈련 데이터를 PSO Classification 에 적용하여 룰을 구축하고 룰 각각의 정확도를 계산한다. 다음은 정확도를 계산하는 방법이다[3].

$$\text{Accurate} = \frac{f_{11}+f_{10}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

(그림 2) 정확도 계산 방법

예측한 클래스와 실제 클래스가 일치하는 데이터의 개수는 $f_{11}+f_{00}$ 이고 일치하지 않는 데이터의 개수는 $f_{10}+f_{01}$ 이다.

세 번째 단계에서는 지정한 반복 횟수만큼 수행하며 가장 좋은 정확도를 가지는 룰을 선택한다. 그 룰을 비정상 여부를 분류해야 할 데이터에 적용한다. 만약 새로운 데이터의 클래스가 비정상일 경우, 그 데이터는 의사로부터 실시간 모니터링과 2 차 진단을 받기 위해 의료서버로 보내진다.

4. 결론 및 향후 연구

대용량의 데이터를 분석하는데 있어서 시공간적인 한계가 있기 때문에 본 연구에서는 환자의 데이터를 자동적으로 진단하는 시스템을 제안하였다. 데이터의 양을 효율적으로 줄이기 위하여 Circadian rhythm 을 기반으로 직접하는 방법을 제안하였고 수행 속도를 향상시키기 위하여 PSO 분류화 알고리즘을 적용하는 방법을 제안하였다. 향후 연구로, 제안한 분류화 시스템과 다른 분류화 시스템들과 수행 속도와 정확도 비교를 통한 검증이 현재 진행 중에 있다.

참고문헌

[1] 신준호, “WPAN/WBAN”, TTA Journal No.116, 2008
[2] H.P. Kriegel, et al, “Multi-represented Knn-Classification for Large Class Sets”, Lecture Notes in Computer Science, Vol.3453, pp. 511-522
[3] Tan, Addison Wesley, Introduction to Data Mining, 2005
[4] 이윤경, 윤혜정, 이민수, 이화여자대학교 컴퓨터학과, 마크로젠, Particle Swarm Optimization 알고리즘을 이용한 바이오칩 데이터의 군집화 및 분류화 기법. 한국정보과학회 Vol.34, No.2, pp. 151-154, 2007
[5] M. Smolensky and L. Lamberg, The Body Clock Guide to Better Health: How to Use your Body’s Natural Clock to Fight Illness and Achieve Maximum Health , Henry Holt and Co., 2000.