Journal of the Korea Convergence Society
Vol. 12. No. 1, pp. 99-104, 2021

ISSN 2233-4890 / eISSN 2713-6353
https://doi.org/10.15207/JKCS.2021.12.1.099

시간흐름을 고려한 특징 추출과 군집 분석을 이용한 헬스 리스크 관리

강지수[1], 정경용[2], 정호일[3*]
[1]경기대학교 컴퓨터과학과 학생, [2]경기대학교 컴퓨터공학부 교수, [3]대림대학교 컴퓨터정보학부 교수

Health Risk Management using Feature Extraction and Cluster Analysis considering Time Flow

Ji-Soo Kang[1], Kyungyong Chung[2], Hoill Jung[3*]
[1]Student, Department of Computer Science, Kyonggi University
[2]Professor, Division of Computer Science and Engineering, Kyonggi University
[3]Professor, Division of Computer Information, Daelim University

요 약 본 논문에서는 시간 흐름을 고려한 특징추출과 군집분석을 이용한 헬스 리스크 관리를 제안한다. 제안하는 방법은 세단계로 진행한다. 첫 번째는 전처리 및 특징추출 단계이다. 이는 웨어러블 디바이스를 이용하여 라이프로그를 수집하여 불완전데이터, 에러, 잡음, 모순된 데이터를 제거하며 결측 값을 처리한다. 그 다음 특징추출을 위해 주성분 분석을 통해 중요 변수를 선택하고, 상관계수와 공분산을 통해 데이터 간의 관계와 유사한 데이터들의 분류를 진행한다. 또한 라이프로그에서 추출한 특징을 분석하기 위해 시간의 흐름을 고려하여 K-means 알고리즘을 통해 동적 군집을 진행한다. 새로운 데이터는 오차 제곱합의 증가분을 기반으로 유사성 거리 측정 방법을 통해 군집을 진행하고, 시간의 흐름을 고려하여 군집에 대한 정보를 추출한다. 따라서 특징 군집을 통해 헬스 의사결정 시스템을 이용하여 신체적 특성, 생활습관, 질병여부, 헬스케어 이벤트 발생위험, 예상 정도 등의 요소를 통해 리스크를 관리할 수 있다. 성능평가는 Precision, Recall, F-measure을 사용하여 제안하는 방법과 퍼지방법, 커널기반 방법을 비교한다. 평가결과 제안하는 방법이 우수하게 평가된다. 따라서 제안하는 방법을 통해 유병자와의 유사도를 이용하여 정확한 사용자의 잠재적 건강 위험을 예측 및 적절한 관리가 가능하다.

주제어 : 라이프로그, 헬스케어, 특징 추출, 상관관계 군집 분석, 데이터 마이닝

Abstract In this paper, we propose health risk management using feature extraction and cluster analysis considering time flow. The proposed method proceeds in three steps. The first is the pre-processing and feature extraction step. It collects user's lifelog using a wearable device, removes incomplete data, errors, noise, and contradictory data, and processes missing values. Then, for feature extraction, important variables are selected through principal component analysis, and data similar to the relationship between the data are classified through correlation coefficient and covariance. In order to analyze the features extracted from the lifelog, dynamic clustering is performed through the K-means algorithm in consideration of the passage of time. The new data is clustered through the similarity distance measurement method based on the increment of the sum of squared errors. Next is to extract information about the cluster by considering the passage of time. Therefore, using the health decision-making system through feature clusters, risks able to managed through factors such as physical characteristics, lifestyle habits, disease status, health care event occurrence risk, and predictability. The performance evaluation compares the proposed method using Precision, Recall, and F-measure with the fuzzy and kernel-based clustering. As a result of the evaluation, the proposed method is excellently evaluated. Therefore, through the proposed method, it is possible to accurately predict and appropriately manage the user's potential health risk by using the similarity with the patient.

Key Words : Lifelog, Healthcare, Feature Extraction, Correlation Clustering Analysis, Data Mining

*This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. 2018R1C1B5047242).
*Corresponding Author : Hoill Jung(hijung1982@gmail.com)
Received December 8, 2020 Revised January 5, 2021
Accepted January 20, 2021 Published January 28, 2021

1. 서론

현대사회에서 기대 수명이 증가함에 따라 건강한 삶을 위한 헬스케어가 지속적으로 연구되고 있다[1]. IT기술의 발달과 보급에 힘입어 최근 헬스케어와 관련한 기술의 화두는 u-health와 IT-의료가 융합된 헬스케어이다. IT-의료 융합 헬스케어의 최종 목표는 환자 중심의 치료하는 개념에서 점차 개인화와 맞춤형 서비스가 가능해야 하고, 웰빙을 위한 건강증진 및 질병예방에 있다. 이러한 변화에 힘입어 국내는 물론 해외 주요 국가들은 의료정보화 산업육성 및 정보화, 표준화, 법제도, 인력양성 등의 기반조성에 투자를 확대하고 있으며, 세계적으로 인공지능을 활용한 헬스케어 연구가 활발하게 이뤄지고 있다[2]. 또한 스마트 디바이스가 대중화되면서 일상생활에서도 의료서비스가 가능해지고 있다. 이는 사람들이 일상생활에서 스마트 디바이스를 통해 헬스 라이프로그를 수집하고, 분석하여 스스로 건강을 관리함을 의미한다. 라이프 로그는 사용자의 일상에서 건강과 관련된 상황데이터의 기록이다. 이는 걸음 수, 수면시간, 식사량, 혈당량 등을 포함한다. 헬스 라이프로그는 시간의 흐름에 따라 수집되는 시계열 데이터이다. 라이프로그는 사용자의 기본정보 뿐만 아니라 생활 습관 전반에 대한 헬스 데이터가 축적되기 때문에 적절히 분석한다면 위험 패턴을 감지하는 것이 가능하다. 헬스 라이프로그에서 잠재적인 건강위험을 예측을 통한 질병의 조기 발견과 치료가 필요하다. 따라서 본 논문에서는 시간흐름을 고려한 특징 추출과 군집 분석을 이용한 헬스 리스크 관리를 제안한다. 제안하는 방법은 라이프로그의 특징을 추출하고 특징 군집을 형성하여 분석한다. 추출한 특징과 군집에서 사용자와 유병자간의 유사도를 시간흐름을 고려하여 계산한다. 이는 질병 생존자의 건강상태에 따른 군집은 유사한 정보를 가질수록 비슷한 생활 습관을 가지는 특징에 기인한다. 사용자와 유병자 간의 유사도를 통해 잠재적인 헬스 리스크를 예측하여 관리한다.

2. 라이프로그를 이용한 헬스케어 기술

로그 데이터는 ICT에서 시간에 따라 발생하는 모든 이벤트와 행동에 대한 기록을 나타낸다[3]. 헬스 로그 데이터는 ICT의 발전으로 임상 데이터, 유전 데이터, 라이프로그 등 인간의 건강과 관련된 다양한 기록을 나타낸다. ICT와 헬스케어의 융합으로 웨어러블 기기, 개인건강기기, 식이영양 관리앱 등을 통해 사용자의 일상에서 건강과 관련된 모든 상황데이터가 라이프로그로 기록된다. 또한, 헬스 플랫폼에서 임상에서 발생하는 모든 로그들과 유전자 데이터 등 다양한 로그들이 수집되고 저장되고 있다. 헬스 로그는 수동적으로 수집되는 기록과 능동적으로 수집되는 기록으로 구분된다. 스마트 밴드, 활동량, 걸음 수, 수면시간 등의 로그는 엠비언트 인텔리전스에 의해 사용자의 조작과 상관없이 수동적으로 기록된다. 식사량, 병원진료, 의약품 섭취, 혈당량 등은 사용자가 직접 능동적으로 기록해야 한다. 이러한 특성에 따라 헬스 로그 데이터는 기록되는 주기가 서로 상이하고, 통합이 어려운 문제가 있다. 또한, 사용자에 헬스 로그의 범위가 변화하여 인간의 다양성으로 인해 보편적인 서비스에 대한 정확도 낮은 문제가 있다[4]. 예를 들어, 능동적 로그 중 혈당량은 주로 당뇨 유병자의 개인건강기기에 의해 기록된다. 이러한 혈당량 로그 데이터를 이용한 서비스는 당뇨 환자에게 한정되어 보편적인 사용자에게 수요가 낮다. 헬스 로그 데이터는 다양한 경로를 통해 수집되고 있으며, 활용 목적에 적합한 로그 선택, 전처리, 분석, 적용 등이 중요하다. 헬스 로그 데이터는 개인의 민감한 정보를 포함하고 있다. 특히, 개인의 단말에 의에 수집되는 특성이 있어 정보 유출에 대한 위험도가 높다. 높아지는 헬스 로그 데이터의 요구에 따라 이를 위한 보안이나 활용에 대한 정책이나 기술이 필요하다. 현재 헬스 로그 데이터를 활용하는 정밀의료(precision medicine)가 주목받고 있다. 정밀의료란 유전체 정보를 기반으로 같은 질병이라도 유전체에 따라 더욱 세밀하고 정확한 처방을 하는 개인맞춤의료 서비스이다. 이는 기존에 사용되고 있던 맞춤의학(personalized medicine)에서 나아가 유전적, 환경적, 생활양식 등 더욱 많은 데이터에 기초하여 환자에게 효과적인 치료방법을 탐색한다. 이를 위해 개인의 정확한 헬스 로그 데이터가 필요하다.

3. 시간흐름을 고려한 특징 추출과 군집을 이용한 헬스 리스크 관리

3.1 헬스 라이프로그 전처리 및 특징 추출

시간흐름에 따른 라이프로그는 웨어러블 디바이스를 통해 개인의 일상생활에 대한 기록을 나타낸다. 라이프로그의 기록과 수집 방식에 따라 수동적 수집과 능동적 수집으로 구분한다. 수동적 수집 방식은 사용자의 조작과

상관없이 기록되는 경우이다. 반면에 능동적 수집 방식은 사용자의 자의적인 판단에 의해 직접 기록하는 경우이다. 라이프로그 데이터 수집은 스마트 IoT 디바이스를 사용한다. 모바일 디바이스는 사용자가 직접 소지하여 걸음수, 수면시간, 혈당, 의약품 섭취 횟수 등의 로그 데이터를 수집하여 서버로 전송한다. 수집한 데이터에서 시간흐름에 따른 사용자들의 생체신호 변화 및 헬스 리스크를 판단하기 위해 의미있는 특징을 추출한다. 특징 추출하기 위한 라이프로그 전처리는 불완전 데이터, 에러, 잡음, 모순된 데이터를 제거한다. 또한 결측값과 디폴트값의 경우 제거하거나 기본순위를 부여하여 처리하고 불완전하고 일관성이 없는 데이터는 삭제하여 기본적인 전처리를 수행한다. 라이프 로그는 걸음수, 수면시간, 혈당, 의약품 섭취 횟수 등의 로그 데이터로 구성되며 기록되는 주기가 서로 상이하고, 통합이 어려운 문제가 있다. 또한 선형, 비선형, 정형, 비정형 등 다양한 구조로 구성되어 각각의 구조에 따라 서로 다른 특징을 나타낸다. 따라서 특징 추출을 통한 데이터 전처리가 필요하다. 라이프 로그의 특징은 주성분 분석, 상관계수, 공분산, 푸리에 변환 분석 등을 이용하여 추출 가능하다. 주성분 분석은 데이터 구조에 대해 변화 없이 차원만을 축소시키는 장점이 있다[5]. 또한, 벡터 크기를 이용한 정렬을 통해 중요도가 높은 변수의 특정이 가능하다. 상관 계수와 공분산은 데이터간의 관계에 대한 분석이 가능하며, 유사한 데이터들의 분류를 통해 빅데이터를 효율적으로 처리한다. 스마트 디바이스에서 수집되어진 시간흐름에 따른 비정형 라이프로그 데이터는 주성분 분석을 통해 차원을 축소하고, 라이프로그간의 벡터 크기 정렬을 통해 가중치가 높은 변수를 추출한다. 다음은 가중치가 높은 변수들간의 상관계수를 계산하여 유사한 데이터를 분류하고 사용자 중심의 특징을 추출한다. 이는 시간이 지남에 따라 변할 수 있는 사용자의 동적 특성을 고려하여 특징 추출을 수행하기 위함이다.

3.2 시간흐름에 따른 동적 군집 헬스 모델

라이프로그는 시간흐름에 따라 기록되는 주기가 서로 상이하고, 통합이 어려운 문제가 있다[3]. 따라서 헬스 라이프로그에서 추출한 특징을 분석하기 위해 시간흐름을 고려하여 동적 군집 헬스 모델을 제안한다. 군집에 적합한 모델인 K-means[6]를 사용하여 동적 군집을 수행한다. 적절한 군집의 수를 도출하기 위해 Ward 연결법을 이용한 계층적 군집을 진행한다[7]. 이는 오차 제곱합의 증가분을 기반으로 유사성 거리측정을 하고 군집을 단계적으로 결합하며, 군집들 간의 유클리드 제곱거리(euclidean squared distance)를 이용하여 군집을 구성하는 방법이다[8]. 각각의 데이터의 평균값으로 군집을 형성하고, 군집의 평균으로 다수의 군집을 구성하여 최적의 k개의 군집을 생성한다. 군집 구조와 수에 대한 전처리한 라이프로그간의 유사성을 기반으로 군집을 형성하고, 반복적으로 탐색하여 진행한다. 군집의 수를 도출하기 위해 통계분석 도구 R의 library인 rattle()함수를 사용한다. 최적의 K개수를 6개로 하여 군집을 형성, 각각의 군집에는 유사한 특성을 가진 개체가 속해있으며 각 군집간의 거리를 통해 다른 군집과의 관계를 발견한다.

시간흐름에 따라 사용자의 건강 상태를 고려한 동적 군집을 형성하여 정적 특성 정보들과 비교한다. 군집 분석 결과 클러스터의 개수를 6로 지정하였을 때 특징의 분포가 가장 균일하게 나타난다. 사용자의 헬스 라이프로그에서 추출한 특징을 군집화 함으로써 분석을 용이하게 하고 헬스 리스크의 판단이 가능하다.

3.3 특징 추출과 군집 기반 헬스 리스크 관리

헬스 리스크는 각각의 사용자에 따른 건강 위험 요인을 분석하고 효율적으로 관리하는 것이 필요하다. 시간흐름에 따라 변화하는 사용자의 건강관리는 상황 마이닝을 통해 진행한다. 상황 마이닝은 전처리 과정을 통해 진행한 라이프로그의 정적과 동적인 특징을 분석한다. 정적인 특징을 반영하기 위해 헬스 데이터로 질병관리청에서 제공받은 국민건강영양조사[9] 및 건강보험심사평가원의 보건의료빅데이터 개방시스템[10]을 사용한다. 또한 기존에 개발한 헬스 의사결정 지원시스템[11-13]을 활용하여 헬스 리스크의 요인 분석과 라이프로그를 이용한 생활 습관에 대한 연관성 분석을 수행한다. 헬스 리스크 관련 정보 요소를 기반으로 하여 구성된 특징 추출과 군집을 통해 헬스 의사결정의 시간흐름을 고려한 리스크를 판단한다. 시간흐름을 고려한 라이프로그의 특징 군집은 유사한 정보를 가질수록 비슷한 헬스 리스크를 가지는 경향이 있다. 따라서 특징 군집을 통해 헬스 의사결정 시스템을 이용하여 신체적 특성, 생활습관, 질병여부, 헬스케어 이벤트 발생위험, 예상 정도 등의 요소를 통해 리스크를 관리한다. 시간흐름에 대해 변화하는 특징 군집을 표현하기 위해 계층적 그래프 구조 모델을 사용한다.

본 연구에서는 동적과 정적인 데이터를 이용하여 시간흐름을 고려한 상황을 판단하여 사용자간의 특징 군집의 유사도를 통해 리스크를 예측한다. 유사도 가중치를 계산하기 위해서 상관계수[14]를 사용한다. 상관계수는 +1과

-1사이의 값을 가지며 값이 +1일 경우 양의 선형 관계, -1일 경우 음의 선형 관계를 나타낸다. 0일 경우 상관관계가 없다고 판단한다. 이는 사용자의 헬스 리스크가 +1에 근사할 때 질병에 걸릴 확률이 높으며, -1에 근사할 때 건강할 확률이 높다는 것을 의미한다. 기존의 헬스 리스크는 피어슨 상관계수를 이용하여 기본적으로 유사도를 판단한다. 하지만 데이터 결측의 문제로 인하여 리스크 판단이 어려운 문제가 발생한다. 이를 해결하기 위해 스피어만 순위 상관계수를 사용한다. 특정 값을 취하는 피어슨 상관계수와 달리 스피어만 순위 상관계수는 순위를 취하기 때문에 결측값 처리시 기본 순위를 부여함으로써 결측치에 대한 한계를 극복한다. 이는 특징 군집에 속해있는 데이터의 순위를 고려하여 데이터 결측에 반영하여 정확도, 재현율의 성능을 높이는 것이 가능하다. 또한 라이프로그에서 추출된 특징, 헬스 베이스에서 추출된 정적 특징을 반영한다.

모든 특징으로부터 유사도를 추출하고, 평균값을 도출하여 사용자의 헬스 유사도를 산출한다. 산출한 건강 유사도에 군집 기반 가중치를 부여하여 최종적인 헬스 리스크를 예측한다. 군집 기반 가중치는 특징군집에서 가까이 위치한 특징끼리는 유사한 특징을 가지는 특징을 이용한다. 사용자가 위치한 특징 군집과 유사한 특징을 갖는 군집간의 거리를 측정하여 군집 가중치를 산출한다. 군집 가중치는 0부터 1까지로 정규화하여 사용한다. 최종적으로 헬스 의사결정 지원 시스템에서 헬스 유사도와 군집 가중치의 곱을 통해 사용자의 헬스 리스크를 예측하는 것이 가능하다. 표 1은 사용자의 헬스 리스크 예측 결과를 나타낸다.

Table 1. Result of Health Risk Prediction

User	Health Similarity	Clustering Weight	Predicted Health Risk
1	+0.943	0.793	0.748
2	+0.312	0.183	0.057
3	-0.796	0.806	-0.642
...	...	...	...

시간흐름을 고려한 유사한 관계를 가지는 사용자와 환자의 특징을 기반으로 사용자의 헬스 리스크를 예측한다. 사용자의 헬스 리스크는 +1에서 -1까지의 헬스 유사도 및 군집 가중치와 비례한다. 즉, 사용자의 라이프로그가 유병자의 패턴과 유사하거나, 비슷한 특징을 갖는 군집이 유병자의 군집과 가까울 경우 +1에 가까워진다. 이는 사용자의 헬스 리스크가 높으며 각별한 관리가 필요함을 의미한다. 반대로 사용자의 라이프로그가 유병자의 패턴과 유사하지 않아 -1에 가까운 값을 갖거나, 비슷한 특징을 갖는 군집이 환자가 아닌 군집과 가까울 경우 -1에 근사한다. 이는 사용자의 생활 패턴이 환자의 것과 유사하지 않으며 잠재적인 헬스 리스크가 적다고 판단한다. 즉, 사용자의 헬스 리스크를 -1에 가깝게 유도함으로써 사용자가 건강한 헬스 리스크를 가지도록 관리한다. 사용자의 헬스 유사도를 환자의 것과 유사하지 않도록 -1에 가깝게 유도하고 특징 군집이 환자가 아닌 군집에 가깝도록 추천한다면 헬스 리스크를 낮게 관리하는 것이 가능하다. 따라서 제안하는 방법으로 사용자의 특징 군집과 환자의 유사도를 통해 헬스 리스크를 예측하고 사용자의 잠재적인 건강 위험을 낮춘다. 그림 1은 제안하는 시간흐름을 고려한 특징 추출과 군집 분석을 이용한 헬스 리스크 관리의 흐름도를 나타낸다.

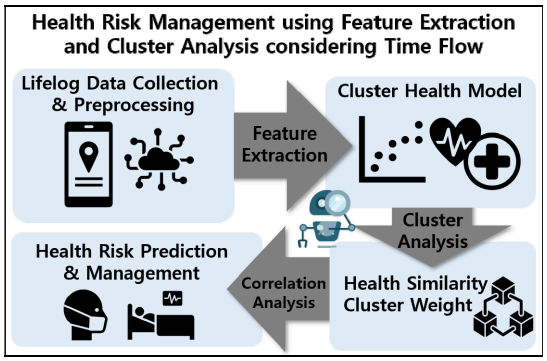

Fig. 1. Flowchart of health risk management

4. 성능평가

라이프로그의 특징 추출과 군집 분석을 이용한 헬스 리스크 관리의 성능 평가를 수행한다. 본 논문에서 제안하는 K-means 기반 헬스 리스크 예측을 통해 도출된 결과와 퍼지 군집 기반, 커널 군집 기반 헬스 리스크 예측의 결과를 비교 분석한다[15,16]. 실험 환경에서 운영체제와 하드웨어는 Windows 10, CPU는 AMD Ryzen 7 3700X 8-Core Processor 3.59GHz, 16GB RAM를 사용하였다. 또한 소프트웨어 환경은 Python 3.8.2와 R 3.6을 사용했다. 평가 지표는 Precision, Recall, F-measure을 사용한다[17].

원활한 성능평가를 위하여 헬스 리스크가 0.5 이상이면 유병자로 판단한다. Precision은 예측한 헬스 리스크

기반으로 실제 사용자가 유병자임을 정확하게 예측한 성능을 의미한다. Recall은 실제 환자 중에서 헬스 리스크가 0.5 이상으로 유병자로 판단된 결과를 나타낸다. 마지막으로 F-measure는 전체 사용자 중에서 정확하게 예측한 결과이다. 표 2는 헬스 리스크 예측에 대한 군집 기법에 따른 Precision, Recall, F-measure의 평가 결과를 나타낸다.

Table 2. Result of Health Risk Prediction

	Precision	Recall	F-measure
Ours	0.879	0.861	0.870
Fuzzy	0.781	0.691	0.733
Kernel-based	0.762	0.805	0.783

성능평가 결과 K-means를 통해 수행된 군집의 결과는 다른 군집방법과 비교해서 우수한 성능을 보였다. K-means 기반 방법은 Precision이 0.879, Recall이 0.861, F-measure가 0.870으로 나타나며 퍼지 군집 또는 커널 군집 기반 모델보다 각각 약 14%p , 9%p 가량 높은 정확도를 보인다. 따라서 제안하는 시간흐름을 고려한 특징 추출과 군집 분석을 이용한 헬스 리스크는 헬스 의사결정 지원 시스템에서 라이프로그를 기반으로 유병자의 정적 데이터의 유사도를 통해 잠재적인 건강 위험을 높은 정확도로 예측한다. 이를 통해 사용자의 헬스 리스크를 탐지하고 적절하게 관리하는 것이 가능하다.

5. 결론

본 논문에서는 시간흐름을 고려한 특징 추출과 군집 분석을 이용한 헬스 리스크 관리를 제안하였다. 제안하는 방법은 헬스 의사결정 지원시스템에서 동적과 정적인 데이터를 이용하여 시간흐름을 고려한 상황을 판단하여 특징 군집의 유사도를 통해 리스크를 예측한다. 시간흐름에 따른 데이터 결측의 문제로 인하여 리스크 판단이 어려운 문제가 발생한다. 이를 해결하기 위해 스피어만 순위 상관계수를 사용한다. 특징 군집에 속해있는 데이터의 순위를 고려하여 데이터 결측에 반영하여 정확도, 재현율의 성능이 높음을 확인하였다. 또한 라이프로그에서 추출된 특징, 헬스 베이스에서 추출된 정적 특징을 반영하였고 상황정보와 헬스 리스크 등의 추론 규칙을 생성하였다. 추론 규칙을 헬스 의사결정 지원시스템에 반영하여 시간흐름을 반영한 헬스 리스크를 관리한다.

REFERENCES

[1] J. C. Kim & K. Chung. (2017). Depression Index Service Using Knowledge Based Crowdsourcing in Smart Health. *Wireless Personal Communications, 93(1),* 255-268.
DOI : 10.1007/s11277-016-3923-3

[2] K. Chung & Y. Hyun. (2020). Edge Computing Health Model Using P2P-Based Deep Neural Networks. *Peer-to-Peer Networking and Applications, 13(2),* 694-703.
DOI : 10.1007/s12083-019-00738-y

[3] H. I. Jung, H. Yoo & K. Chung. (2016). Associative Context Mining for Ontology-driven Hidden Knowledge Discovery. *Cluster Computing, 19(4),* 2261-2271.
DOI : 10.1007/s10586-016-0672-8

[4] J. W. Baek & K. Chung. (2020). Context Deep Neural Network Model for Predicting Depression Risk Using Multiple Regression. *IEEE Access, 8(1),* 18171-18181.
DOI : 10.1109/ACCESS.2020.2968393

[5] C. M. Kim, E. J. Hong, K. Chung & R. C. Park. (2020). Driver Facial Expression Analysis Using LFA-CRNN-Based Feature Extraction for Health-Risk Decisions. *Applied Sciences, 10(8),* 2956.
DOI : 10.3390/app10082956

[6] J. W. Baek, J. C. Kim, J. Chun & K. Chung (2019) Hybrid Clustering based Health Decision-making for improving Dietary Habits, *Technology and Health Care, 27(5),* 459-472. DOI : 10.3233/THC-191730

[7] K. Chung & H. Jung. (2020). Knowledge-based dynamic cluster model for healthcare management using a convolutional neural network. *Information Technology and Management, 21(1),* 41-50.
DOI : 10.1007/s10799-019-00304-1

[8] H. I. Jung & K. Chung. (2015). Ontology-driven Slope Modeling for Disaster Management Service. *Cluster Computing, 18(2),* 677-692.
DOI : 10.1007/s10586-015-0424-1

[9] Korea Health Statistics 2018: Korea National Health and Nutrition Examination Survey (KNHANES Ⅶ-3), 2017, Korea Centers for Disease Control and Prevention

[10] Health Insurance Review & Assessment Service, http://opendata.hira.or.kr.

[11] J. S. Kang, D. H. Shin, J. W. Baek & K. Chung. (2019). Activity Recommendation Model Using Rank Correlation for Chronic Stress Management. *Applied Sciences, 9(20),* 4284.
DOI : 10.3390/app9204284

[12] H. Jung & K. Chung. (2016). Life Style Improvement Mobile Service for High Risk Chronic Disease based on PHR Platform. *Cluster Computing, 19(2),* 967-977
DOI : 10.1007/s10586-016-0549-x

[13] J. S. Kang & J. W. Baek, K. Chung. (2020). PrefixSpan Based Pattern Mining Using Time Sliding Weight From Streaming Data. *IEEE Access, 8,* 124833-124844.
DOI : 10.1109/access.2020.3007485

[14] D. Chicco & G. Jurman. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics, 21(1),* 6.
DOI : 10.1186/s12864-019-6413-7

[15] Z. Zivkovic. (2004). Improved adaptive Gaussian mixture model for background subtraction. *In Proc of the 17th International Conference on Pattern Recognition, ICPR, 2,* 28-31.
DOI : 10.1109/icpr.2004.1333992

[16] M. Tang, D. Marin, I. B. Ayed & Y. Boykov. (2019). Kernel cuts: Kernel and spectral clustering meet regularization. *International Journal of Computer Vision, 127(5),* 477-511.
DOI : 10.1007/s11263-018-1115-1

[17] D. H. Shin, R. C. Park & K. Chung. (2020). Decision Boundary-Based Anomaly Detection Model Using Improved AnoGAN From ECG Data, *IEEE Access, 8,* 108664-108674.
DOI : 10.1109/access.2020.3000638

강 지 수(Ji-Soo Kang) [학생회원]

· 2016년 2월 ~ 2020년 2월 : 경기대학교 컴퓨터공학부 (공학사)
· 2020년 2월 ~ 현재 : 경기대학교 컴퓨터과학과 (석사과정)
· 2017년 2월 ~ 현재 : 경기대학교 컴퓨터공학부 데이터마이닝 연구실 연구원
· 관심분야 : 데이터 마이닝, 헬스케어, 빅데이터 분석, 추천 시스템, 인공지능
· E-Mail : kangjs920@gmail.com

정 경 용(Kyungyong Chung) [종신회원]

· 2000년 2월 : 인하대학교 전자계산공학과 (공학사)
· 2002년 2월 : 인하대학교 전자계산공학과 (공학석사)
· 2005년 8월 : 인하대학교 컴퓨터정보공학부 (공학박사)
· 2006년 3월 ~ 2017년 2월 : 상지대학교 컴퓨터정보공학부 교수
· 2017년 3월 ~ 현재 : 경기대학교 컴퓨터공학부 교수
· 관심분야 : 데이터 마이닝, 헬스케어, 빅데이터, 지능시스템, 인공지능, HCI, 인공지능, 추천 시스템
· E-Mail : draginchi@gmail.com

정 호 일(Hoill Jung) [정회원]

· 2010년 2월 : 상지대학교 컴퓨터정보공학부 (공학사)
· 2013년 2월 : 상지대학교 컴퓨터정보공학과 정보통신전공 (공학석사)
· 2017년 2월 : 상지대학교 컴퓨터정보공학과 정보통신전공 (공학박사)
· 2018년 3월 ~ 2019년 2월 : 원광대학교 컴퓨터·소프트웨어 공학과 교수
· 2019년 3월 ~ 현재 : 대림대학교 컴퓨터소프트웨어과 교수
· 관심분야 : 지능형 시스템, 데이터 마이닝, 추천, 감성공학, 지식 시스템
· E-Mail : hijung1982@gmail.com