

위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측

최소영¹, 김주창¹, 정경용^{2*}

¹경기대학교 컴퓨터과학과 학생, ²경기대학교 컴퓨터공학부 교수

Prediction of Dietary Knowledge using Multiple Regression Analysis for Preventing Stomach Diseases

So-Young Choi¹, Joo-Chang Kim¹, Kyungyong Chung^{2*}

¹Student, Data Mining Lab., Department of Computer Science, Kyonggi University

²Professor, Division of Computer Science and Engineering, Kyonggi University

요 약 현대사회는 1인가구가 증가함에 따라 불규칙한 식습관으로 인해 영양이 불균형하게 포진되어있다. 이러한 식습관은 위장질환, 소화기 질환 등 만성질환의 발병률을 증가시켰다. 본 논문은 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측을 제안한다. 제안하는 방법은 식이지식 예측을 통해 사용자의 위장질환과 식이영양을 관리하는 방법이다. 헬스 플랫폼에서 스마트 기기를 통해 수집된 사용자의 PHR을 통합한다. 통합된 데이터로부터 다중회귀분석을 이용하여 사용자의 식이와 활동량 변화를 분석한다. 사용자의 식이 성분과 소모 칼로리, 기초대사와 같은 상황정보를 입력으로 적절한 식이성분, 위장질환 수치의 변화를 예측하고 필요할 것으로 나타나는 영양성분을 사용자에게 권장한다. 이를 통해 현대인들은 균형 잡힌 식사를 통해 위장질환을 관리할 수 있다.

주제어 : PHR, 식이, 다중회귀분석, 헬스케어, 헬스 플랫폼, 시계열

Abstract Modern society is undergoing nutritional imbalance according to the diet as the number of one person increases. This is increasing the incidence of chronic diseases such as gastrointestinal diseases and digestive diseases. This study suggests the prediction of dietary knowledge using multiple regression analysis for preventing chronic stomach diseases. The proposed method manages user's stomach diseases and dietary nutrition through the prediction of nutrition knowledge. It collects user's PHR through smart device and integrates in the health platform. The integrated data analyzes the dietary and activity of the user through multiple regression analysis. It predicts the required nutrients and provides services to users through applications. Therefore, it suggests recommended dietary components and consumed calories, appropriate dietary components based on the user's basal metabolism, and gastrointestinal levels. With the personalized health management, modern people can manage gastrointestinal diseases through a balanced diet.

Key Words : PHR, Nutrition, Multiple Regression Analysis, Healthcare, Health Platform, Time Series

*This work was supported by the GRR program of Gyeonggi province.

*Corresponding Author : Kyungyong Chung(dragonhci@gmail.com)

Received March 22, 2019

Revised June 3, 2019

Accepted July 20, 2019

Published July 28, 2019

1. 서론

현대사회는 식생활의 서구화, 간편식품 산업의 확대, 1인가구의 증가 등으로 많은 사람들이 식이영양 관리에 대한 수요가 증가하고 있다. 식이영양 관리는 인간의 생존에 큰 영향을 미치며 건강을 유지하기 위해 지속적인 관리가 필요하다. 현대 사회는 다양한 식품첨가물이 포함된 수많은 가공식품이 생산되고, 자극적인 식품이 유행하고 있다. 영양성분은 사용자의 건강과 직접적으로 관련이 있어 식단관리를 통해 올바른 섭취가 필요하다[1,2]. 또한 1인가구가 증가함에 따라 영양성분을 고려한 식단보다 인스턴트 식품의 섭취태도가 크게 높아졌다. 이는 사회 전반적으로 청년층에게 영양불균형을 초래하여 위장 질환, 소화기 질환, 비만 등 다양한 질병의 원인이 된다. 실례로 건강보험심사평가원 데이터를 분석한 결과, 소화기 염증질환인 크론병 발병은 10대 기준 0.76명에서 1.3명으로 71%, 20대는 0.64명에서 0.88명으로 37% 증가했으며, 궤양성 대장염의 경우도 10대와 20대에서 상대적으로 발병률이 크게 높아졌다고 조사되었다[3]. 이에 따라 현대인에게 영양관리는 필수불가결한 것으로 나타나며 개인의 식이지식을 예측하여 개인화된 영양관리가 가능하도록 해야 한다[4]. 또한, 정보 기술의 발전으로 다양한 데이터의 수집이 용이해짐에 따라 축적되는 라이프 로그 빅데이터를 활용하기 위한 많은 연구가 진행되고 있다. 데이터가 방대해지고 운용이 많아짐에 따라 빅데이터를 이용한 산업이 확산되었고, 빅데이터 기술과 의료산업이 융합되어 헬스케어 산업이 크게 발전하고 있다. 특히, 엠비언트 인텔리전스 분야에서 정형, 비정형 데이터의 데이터 분석을 이용한 보다 정확한 예측이 가능해지고 있다[5]. 회귀모델은 데이터의 정확도가 예측의 오차에 큰 영향을 미친다. 특히, 모든 사용자가 의료 종사자와 비슷한 수준의 지식을 가질 수 없기 때문에 사용자가 직접 데이터를 입력하는 경우에 근삿값을 사용하거나 입력을 하지 않아 데이터에서 오차가 발생할 확률도 있다. 따라서 본 논문에서는 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측을 제안한다. 이는 사용자의 상황정보를 활용하여 식이영양의 변화와 위장질환의 발병 확률을 예측하는 건강관리 모델이다.

2. 관련연구

2.1 개인건강기록 기반 헬스케어

PHR(Personal Health Record)는 사용자의 라이프 로그, 질병이력, 유전요인 등의 정보로 구성된다[6]. 사용자의 PHR에 대한 분석을 통해 정밀의료가 가능하다. 정밀의료는 환자의 정보에 따라 최적화된 치료방안을 제공하는 개인화 의료행위이다[7,15]. 보다 정밀한 의료를 위해서는 PHR의 영향이 큰 비중을 차지한다. 하지만 라이프 로그, 질병기록과 같은 PHR 데이터는 환자의 사생활에 관련한 규제에 의해 상용화에 대한 논란이 있다[14,15]. 이에 대한 방안으로 제시된 개인 주도형 플랫폼은 사용자 개인에게 동의를 얻음으로써 데이터 규제문제를 해결한다[11]. 이러한 개인 주도 방식을 차용한 개인 건강기록 기반 정밀의료 플랫폼은 사용자 스스로 데이터를 입력, 저장, 수집하여 분석 및 활용이 가능한 구조이다[7]. 플랫폼과 연결된 스마트 디바이스를 통해 사용자들은 개인의 식이, 생활, 건강 상태를 점검할 수 있다. 또한 이는 정밀 의료에 필요한 데이터를 사용자 중심으로 연결하여 구체적인 정밀의료 데이터집합을 구성할 수 있다. 이를 통해 환자별 개인화된 진료서비스와 같이 더욱 선제적인 헬스케어가 제공 가능하도록 발전하고 있다.

2.2 ARIMA 기반 시계열 회귀분석

ARIMA(Auto Regressive Integrated Moving Average) 모델은 회귀 기반의 시계열 분석 기법 중 하나이다. 이는 관측된 과거의 데이터 값과 현재 값의 오차를 이용하여 현재 시계열 값을 설명하는 ARMA(Auto Regressive Moving Average)모델을 일반화한 방법이다[8,9]. ARMA 모델이 고정적인 시계열에 적용이 가능하다면, ARIMA는 분석 대상이 비고정적인 시계열에서 적용이 가능하다[8]. ARIMA는 자기회귀모델인 AR(Auto Regression)과 이동평균모델인 MA(Moving Average)가 결합된 모델이다. 데이터가 비고정적 시계열 데이터일 경우 변환이 필요하다. 주어진 시계열 데이터를 고정적으로 변환하는 방법은 R에서 '로그함수'와 '차분(diff)'을 이용한다. 로그함수와 차분을 차례로 적용한 다음 tseries 패키지의 `adf.test` 함수를 통해 변환된 시계열이 고정적인지 여부를 확인한다. 시계열이 고정적으로 나타난다면 `AUTO.ARIMA` 함수를 사용하여 최적화된 변수를 찾는다. 앞서 구한 변수를 통해 ARIMA 모델을 만든 후, 예측 함수를 사용하여 변화 예측이 가능하다.

3. 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측

본 논문에서는 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측을 제안한다. 사용자가 식사, 걸음 수 등의 라이프로그를 스마트 디바이스에 기록하면, 사용자에 따른 PHR은 회귀식의 변수에 맞게 전처리되어 헬스 플랫폼에 로딩된다. 다수의 사용자의 데이터가 저장된 헬스 플랫폼은 다중회귀분석을 통해 각 사용자별 식이영양과 라이프로그를 분석한다. 분석결과를 기반으로 사용자에게 필요한 식이성분과 위장질환 위험 수치를 예측한다. PHR 실증 데이터는 비고정적 시계열 데이터이기 때문에 고정적 시계열로 변환한다. 제안하는 방법을 이용하여 사용자에게 필요한 영양소 공급을 위한 식단의 추천, 위장질환의 위험수치를 알림을 통해 개인화된 식이관리가 가능하다.

3.1 PHR 기반 식이 데이터 수집 및 처리

사용자는 자신의 스마트 디바이스에 아침, 점심, 저녁 식사와 걸음수, 운동 시간 등의 라이프로그를 지속적으로 입력한다[10]. 라이프로그는 지속적으로 발생하지만, 사용자마다 식사의 주기가 다르기 때문에 하루 단위로 데이터를 갱신한다. 수집된 개인의 데이터는 전처리 작업을 통해 공통 데이터 모델에 적합하게 변환한 후 헬스 플랫폼에 전송된다. 공통 데이터 모델은 의료기관별로 상이한 구조를 통합하기 위한 모델이다. 데이터를 통합하지 않을 경우 걸음수, 운동 시간, 활동량 등의 변수는 의미가 중복되어 데이터 분석 결과에 혼돈을 야기할 수 있다. 식이 데이터의 경우 사람마다 선호하는 음식과 식습관이 상이하기 때문에 대량의 변수를 고려해야 하는 문제가 있다. 식품에는 탄수화물, 지방, 단백질 등의 영양성분에 대한 데이터가 포함되어 있으며 각각 권장하는 섭취량이 다르기 때문에 올바른 식습관 분석을 위해 식품 데이터에 대한 세분화가 필요하다. 회귀식은 걸음수에 따른 소모 칼로리, 식품을 구성하는 탄수화물, 단백질, 지방, 소모 칼로리 등 헬스 플랫폼에 축적된 사용자의 빅데이터를 기반으로 도출된다. 디바이스를 통해 수집된 데이터는 애플리케이션 내에서 걸음 수, 운동 시간, 활동량 등의 속성은 소모 칼로리로 변환되어 저장하고, 식이데이터는 식품안전정보포털에서 제공하는 식품영양데이터를 기반으로 하여 식품을 구성하는 영양성분으로 분리되어 다시 저장된다. 이를 통해 사용자가 섭취하고 있는 음식으로부터 지방, 단백질, 탄수화물, 나트륨, 당류 등 영양성분에 대한 데이터의 수집이 가능하다. Table 1은 변수의 정의를 나타낸다. 변수는 Range와 Flag 2가지 타입으로 정의되고, Flag 타입의 STOM은 0이면 정상수치, 1이면 위장

질환 위험군으로 분류한다. Table 2는 사용자에 따른 전처리한 식이 데이터를 나타낸다. 이는 헬스케어 실증 데이터로 대상자 30명으로부터 2018년 12월 1일부터 2019년 2월 28일까지 90일간 수집된 데이터이다. 실증을 위한 변수는 ICT 기반 영양 관리 서비스[11]에서 식품을 구성하는 주요 영양성분과 위장질환과 관련된 변수로 구성한다.

Table 1. Definition of variables

Name	Types	Values	Comment
calConsume	Range	[0, 388]	calorie consumed
calIntake	Range	[0, 6244]	calorie Intake
carbonhydra	Range	[0, 1359.44]	carbohydrate
fat	Range	[0, 427.84]	fat
protein	Range	[0, 409.24]	protein
sugar	Range	[0, 235.20]	sugar
Na	Range	[0, 11182.675]	natrium
cholest	Range	[0, 2705.01]	cholesterol
satFatAcid	Range	[0, 175.8]	saturated fatty acid
traFatAcid	Range	[0, 5.24]	trans fatty acid
STOM	Flag	[0 = Normal], [1 = Patient]	Stomach-disorder

Table 2. Preprocessed dietary data

Column	n00001	n00002	n00003	...	n00270
Date	180901	180901	180901	...	181130
User ID	u0001	u0003	u0004	...	u0030
calConsume	136	364	124	...	376
calIntake	2087.32	1464.19	1270.72	...	3565
carbohydra	395.44	205.03	230.28	...	270
protein	41.12	96.46	29.04	...	40
fat	37.92	28.7	25.92	...	260
sugar	0	1.54	2	...	15
Na	362.44	8055.43	3333.16	...	1015
cholest	0	167.34	161.96	...	0
satFatAcid	3.48	9.31	11	...	20
traFatAcid	0.08	0.63	0.24	...	0
STOM	0	1	0	...	0

3.2 다중회귀분석 기반 식이지식 예측 모델

주기적으로 입력되는 사용자의 PHR 실증데이터는 하루단위로 입력되어 날짜에 따른 시계열적 특성을 갖는다 [6,8]. 날짜의 흐름에 따라 음식섭취와 라이프로그가 일자별로 연속되어 저장된다. 시계열적 특성을 갖는 사용자의 PHR 데이터는 전처리 과정을 거쳐 식이데이터인 사용자의 식단은 탄수화물, 단백질, 지방으로 변환되고 운동량인 걸음 수, 이동거리는 소모 칼로리로 변환된다. 이와 같이 변환된 데이터는 통계적 계산을 위해 R을 활용

하여 다중회귀분석[12]을 수행한다. 이는 헬스케어 실증 데이터를 이용하며, 90일간의 데이터 중 65일의 데이터를 다중회귀분석에 활용한다. 12월 1일부터 65일 동안의 데이터를 분석에 활용하고, 그 이후의 데이터를 테스트에 활용한다. 각 사용자의 주간 영양소별 식이섭취, 운동량의 추이를 분석하여 사용자의 미래 식이, 위장질환 수치를 예측할 수 있다. 다중회귀분석은 하나의 종속변수와 다수의 독립변수 사이의 관계를 판단하여 분석하는 기법이다[12]. 종속변수의 변화를 통하여 n개의 독립변수 중 어떤 변수가 어떠한 영향을 주는지 분석하여 상관관계를 파악할 수 있다. 다중회귀분석은 회귀식의 변수를 결정하고, 이를 이용하여 다중회귀모델을 만든다. 모델은 상관계수 R, 결정계수 R², 추정값의 오차 등을 포함한다. 예측이 성공적인지 판단하기 위해서 R, R², 오차, p-value를 통해 통계적으로 유의미한지 회귀식을 검증한다. 식(1)은 사용자의 데이터를 다중회귀 분석하여 추정된 회귀식이다. 회귀식에 따르면 종속변수인 위장질환에 영향을 미치는 유의미한 독립변수는 carbonhydra, protein, Na, satFatAcid로 각각 탄수화물, 단백질, 나트륨, 포화지방산을 나타낸다.

$$\begin{aligned}
 STOM &= 0.0002 \times carbonhydra \\
 &+ (-0.0405) \times protein \\
 &+ 0.0002 \times Na \\
 &+ 0.002 \times satFatAcid \\
 &+ (-0.1952)
 \end{aligned}$$

식(1)

STOM은 위장질환을 나타내는 종속변수로 수치가 높을수록 위장질환에 걸릴 확률이 높음을 의미한다. 탄수화물과 나트륨, 포화지방산을 많이 섭취할수록 위장질환에 영향을 미치고, 단백질은 위장질환에 이롭다고 해석 가능하다. 따라서 추정된 회귀식에 의해, 탄수화물과 나트륨, 포화지방산을 줄이고, 단백질 섭취량을 늘리는 것이 위장질환에 걸릴 확률을 낮출 수 있다.

3.3 개인건강기록 기반 식이지식 개인화

다중회귀분석 기반 식이지식 예측은 많은 사용자들로부터 보편적인 변화를 예측하는 모델이다. 이를 통한 예측 결과는 사용자의 상황에 따라 다르게 해석될 수 있다. 예측한 향후 3일간의 섭취열량과 영양소의 변화에 대한 예측 값은 사용자의 건강상태에 따라 긍정적일 수도 부정적일 수도 있다. 실제 사용자의 건강상태는 식이성분

뿐만 아니라 다양한 변수에 의해 결정되며, 이에 따라 개인화된 식이 관리가 필요하다.

건강상태에 따라 일일 권장섭취량이 변화하며, 질병 유무에 따라 같은 식품이라도 긍정적인 경우와 부정적인 경우가 공존한다. 따라서 식이지식 관리의 개인화를 위해 두 가지 방법을 병행한다. 사용자의 PHR을 상황정보로 이용하여 다중회귀분석 기반의 식이지식 예측, 헬스케어 상황정보 모델[13], 식이지식 온톨로지[14]를 통해 적합한 개인화 서비스를 추천한다.

헬스케어 상황정보 모델은 사용자의 상황정보를 이용하여 사용자의 건강상태를 추론하는 모델이다[13]. 식이지식 서비스 온톨로지는 사용자의 건강상태에 적합한 식이지식을 추천하기 위한 모델이다[14]. 식품섭취 내역에 따라 예측된 섭취열량과 영양소를 헬스 상황정보 모델[13]을 이용하여 사용자의 건강상태와 비교한다. Fig. 1은 다중회귀분석을 이용한 식이지식 관리를 나타낸다. 이는 크게 두 단계로 구분할 수 있다. 첫 번째 단계에서 실제 사용자는 지속적으로 자신의 라이프로그를 입력하고, 식이지식 예측 모델을 통해 변화를 예측한다. 두 번째 단계에서 예측값과 사용자의 권장 섭취량 비교결과를 이용하여 사용자의 건강상태에 적합한 식품을 추천한다.

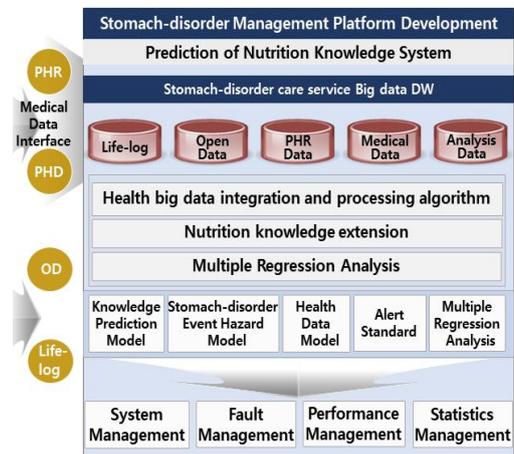


Fig. 1. Management of Nutrition Knowledge using Multiple Regression Analysis

4. 모델 및 성능평가

제안하는 모델인 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측의 성능평가를 수행하였다. 성능평가는 분석에 활용한 30명에 대한 65일간의 데이터를

제외한 25일 동안의 데이터를 사용한다. 이는 헬스케어 실증 대상자를 위장장애가 있는 그룹과 없는 그룹이 있으며 이를 구분하여 평가를 진행한다. Table 3은 성능평가 결과를 나타낸다. Test1은 30명 모두의 데이터를 활용하여 평가한 값을 나타낸다. Test2는 30명 중 위장질환을 가진 13명의 데이터를 이용하여 평가한 값을 나타낸다. Test3는 나머지 17명의 데이터를 이용하여 평가한 값을 나타낸다. 상관계수는 독립변수의 변동에 따른 종속변수의 변동을 나타낸다. 평균 절대오차는 측정값의 오차의 평균을 나타낸다[16]. 이는 측정값과 실제값의 차이이며, 정밀한 숫자로 낼 수 없어 측정값을 근사값으로 사용함으로써 나타나는 오차를 의미한다. 평균 제곱근 오차는 모델에서 예측값과 실제값의 차이를 다룰 때 사용하는 척도이다. 이는 정밀도를 표현하는 데 적합하다. 오차가 적게 나타날수록 회귀식이 적합함을 나타낸다. 위의 변수들은 모두 0에서 1사이의 값을 갖는다. 상대 절대 오차는 실제 값 전체의 평균을 나타낸다. 실제 값의 전체 절대 오차를 정규화하고, 이를 전체 예측값의 전체 절대 오차로 나눈 값이다. 상대 절대 오차 제곱근은 상대 절대 오차를 제곱근한 값을 나타낸다. 이는 오차를 예측되는 값과 동일한 차원으로 감소시킨 값이다.

Table 3. Performance evaluation

Performance	Evaluation		
	Test1	Test2	Test3
Correlation Coefficient	0.8277	0.6715	0.9711
Mean Absolute Error	0.2111	0.5171	0.1784
Root Mean Squared Error	0.2699	0.6814	0.1918
Relative Absolute Error	45.62%	48.16%	44.38%
Root Relative Squared Error	56.12%	57.18%	55.95%
Total Number of Instances	300	130	270

평가결과, Test2에서 전반적으로 평가가 낮게 나타난다. 이는 만성 위장질환자의 불규칙적인 식습관으로 인해 예측의 정확도가 낮아져 발생하는 현상으로 비교적 규칙적인 식습관을 가진 Test3에서는 상관관계나 오차가 우수하게 나타난다. 전체 대상자의 데이터를 사용한 Test1에 대한 상관계수는 0.8277로 높게 나타나 종속변수가 독립변수에 따라 변동될 확률 즉, 종속변수인 위장질환과 독립변수인 탄수화물, 단백질, 나트륨, 포화지방산의 관계가 밀접함을 알 수 있다. 평균절대오차는 0.2111로 측정값과 실제값의 차이가 크지 않아 회귀식을 추정하는데 미친 오차의 영향이 상대적으로 낮아 데이터의 정확도가 높음을 알 수 있다. 평균 제곱근 오차는 0.2699로 모델

을 통한 예측값과 실제 관찰값의 차이가 낮기 때문에 회귀식의 예측률이 높고, 추정이 잘 되었다고 판단할 수 있다. 따라서 성능평가에 의해 추정된 회귀식은 높은 추정치를 갖는다는 것을 알 수 있다. 이에 따라 탄수화물, 나트륨, 포화지방산의 과도한 섭취는 위장질환을 촉구하고, 단백질은 완화요인이 될 수 있다. 성능평가결과와 제안하는 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측 모델은 높은 추정치로 위장질환 수치를 추측할 수 있음을 알 수 있다. 또한 추정된 회귀식의 독립변수를 통해 사용자의 위장질환 수치에 대비한 식이성분을 권장할 수 있다.

5. 결론

현대는 가공된 식품이 생산 및 소비되고 자극적인 식품이 유행하고 있다. 또한 1인 가구 증가로 인해 현대인의 식품첨가물이 포함된 식품의 소비가 증가하며 그에 따른 섭취행태는 현대인에게 영양불균형을 초래하였다. 영양불균형은 소화기관에 무리를 주어 크론병, 위궤양, 대장염 등 다양한 위장질환을 일으킨다. 건강보험심사평가원의 데이터 분석결과 이러한 위장질환을 앓는 청년층의 비율이 점차 증가하고 있음이 밝혀졌다. 따라서 이에 대비한 해결 방안으로 위장질환 예방을 위한 다중회귀분석을 이용한 식이지식 예측 방법을 개발하였다. 이는 사용자가 애플리케이션에 섭취한 식이 데이터와 운동량을 입력하여 수집하였고, 수집된 데이터는 플랫폼에 맞게 전처리를 수행하였다. 이를 주기적으로 헬스 플랫폼으로 전송하여 다중회귀기법을 통해 사용자의 섭취 영양 상태와 운동량의 경향에 따라 분석하였다. 이를 통해 부족하거나 과다할 것으로 예상되는 영양소와 위장질환의 발생 가능성을 예측할 수 있었다. 예측 결과는 한국영양학회에서 제공하는 한국인 영양섭취기준에 대하여 결과와 포화를 판단하였다. 이는 사용자에게 필요한 영양성분, 운동량 등의 정보를 제공함으로써 사용자의 적합한 개인화된 식이지식 관리를 지원한다.

REFERENCES

- [1] W. Raghupathi & V. Raghupathi. (2014). Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*, 2(1), 3.

- [2] H. S. Schwerin, J. L. Stanton, J. L. Smith, A. M. Riley Jr & B. E. (1982). Food, Eating Habits, and Health: a Further Examination of the Relationship Between Food Eating Patterns and Nutritional Health. *The American Journal of Clinical Nutrition*, 35(5), 1319-1325.
- [3] Health Insurance Review & Assessment Service(HIRA). (2018), www.hira.or.kr/.
- [4] M. E. Shils & M. Shike. (2006). *Modern nutrition in health and disease*. Lippincott Williams & Wilkins.
- [5] D. Krackhardt. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks*, 10(4), 359-381.
- [6] D. Kaelber & E. C. Pan. (2008). *The Value of Personal Health Record (PHR) Systems*. In AMIA Annual Symposium Proceedings. (2008. p. 343). American Medical Informatics Association.
- [7] F. C. Collins & H. Varmus. (2015). A New Initiative on Precision Medicine. *New England Journal of Medicine*, 372(9), 793-795.
- [8] G. P. Zhang. (2003). Time Series Forecasting using a Hybrid ARIMA and Neural Network model. *Neurocomputing*, 50, 159-175.
- [9] J. Contreras, R. Espinola, F. J. Nogales & A. J. Conejo. (2003). ARIMA Models to Predict Next-day Electricity Prices. *IEEE Transactions on Power Systems*, 18(3), 1014-1020.
- [10] E. Y. Jung, J. H. Kim, K. Chung & D. K. Park. (2013). Home Health Gateway based Healthcare Services through U-health Platform. *Wireless Personal Communications*, 73(2), 207-218.
- [11] National Information Society Agency. ICT-based Nutrition Management Service Empirical Data. (2018). https://www.nia.or.kr/.
- [12] C. H. Mason & W. D. Perreault Jr. (1991). Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research*, 28(3), 268-280.
- [13] J. Kim & K. Chung. (2014). Ontology-based Healthcare Context Information Model to Implement Ubiquitous Environment. *Multimedia Tools and Applications*, 71(2), 873-888.
- [14] H. Jung & K. Chung. (2016). Knowledge-based Dietary Nutrition Recommendation for Obese Management. *Information Technology and Management*, 17(1), 29-42.
- [15] J. C. Kim & K. Chung. (2018). Mining Health-risk Factors using PHR Similarity in a Hybrid P2P Network. *Peer-to-Peer Networking and Applications*, 11(6), 1278-1287.
- [16] J. C. Kim & K. Chung. (2019). Prediction Model of User Physical Activity using Data Characteristics-based Long Short-term Memory Recurrent Neural Networks. *KSII Transactions on Internet and Information Systems*, 13(4), 2060-2077.

최 소 영 (So-Young Choi)

[학사학위]



- 2016년 3월 ~ 현재 : 경기대학교 컴퓨터과학과 학부생
- 2017년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마이닝 연구실 연구원
- 관심분야 : 데이터 마이닝, 헬스케어, 빅데이터, 지능시스템, 인공지능
- E-Mail : qpalqp3396@gmail.com

김 주 창 (Joo-Chang Kim)

[학사학위]



- 2014년 2월 : 상지대학교 컴퓨터정보공학부 (공학사)
- 2016년 8월 : 상지대학교 컴퓨터정보공학과 (공학석사)
- 2017년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 박사과정
- 관심분야 : 데이터 마이닝, 딥러닝, 자연어 처리, 블록체인, 헬스케어
- E-Mail : kjc2232@naver.com

정 경 용 (Chung, Kyungyong)

[중신학위]



- 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, 정보검색, 추천 시스템
- E-Mail : dragonhci@hanmail.net