JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 2, pp. 321-327, Feb. 2019

머신러닝을 활용한 사상체질 분류 모델 선정과 서비스 플로우 디자인

이 재 화* · **이 현 학**
연세대학교 기술경영학협동과정

Selecting Sasang-Type classification model using machine learning and designing the service flow

Jae-Hwa Lee* · **Hyun-Hak Lee**
Department of Management of Techonology, Yonsei University, Seoul, 03722, Korea

[요 약]

현대의학과 달리 한의학은 데이터의 표준화가 이루어지지 않았고, 데이터를 누적할 수 있는 인프라가 매우 부족한 실정이다. 이러한 문제들은 한의학 도메인에서의 인공지능 연구는 물론 화두가 되고 있는 양방 협진도 어렵게 한다. 이러한 문제를 해결하기 위해 사상체질을 중심으로 머신러닝을 활용하여 사상체질 분류 알고리즘 모델을 선정하고, 선정한 모델을 바탕으로 의사와 환자가 이용할 수 있는 서비스를 디자인하였다. 또한, 데이터를 누적시킬 수 있는 데이터 베이스를 디자인하여 추후 누적된 데이터를 활용하여 분류 알고리즘 모델의 정확도를 개선할 수 있게 하였다.

[Abstract]

Unlike in modern medicine, medical data have not been standardized yet and infrastructure for accumulating data is poor in oriental medicine. These problems become obstacles in artificial intelligence research and bilateral cooperation in domain of oriental medicine. To solve this problem, we compared the classification accuracy using several machine learning algorithms to select the highest accuracy model. Also we designed an application using the selected classification algorithm model to offer convenience service to doctor and patient. In addition, we designed the database to accumulate medical data, so accumulated data can be used to enhance the accuracy of classification algorithm model.

색인어 : 머신러닝, 딥러닝, 한의학, 사상체질, 데이터베이스

Key word : MachineLearning, DeepLearning, Or7iental Medicine, Sasang-Type, Database

http://dx.doi.org/10.9728/dcs.2019.20.2.321

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 10 December 2018; **Revised** 02 January 2019
Accepted 20 February 2019

***Corresponding Author; Jae-Hwa Lee**

Tel: + ※ 개인정보 표시제한
E-mail: rhymeandflow@naver.com

Ⅰ. 서 론

최근 머신러닝, 딥러닝 알고리즘의 뛰어난 성능이 증명되면서, 많은 도메인에서 인공지능을 통한 연구 혹은 서비스가 파생되고 있다. 이러한 흐름에 맞춰 의료업계 또한 인공지능을 통해 큰 변혁의 시기를 맞이하고 있다. 특히, IBM이 개발하고 있는 왓슨(Watson)은 이러한 변혁을 선도하고 있다. 왓슨은 복잡하고 방대한 의료 데이터를 분석하여 의학적 통찰력을 도출하는 인공지능이다. 왓슨은 전자의무기록 (EMR), 환자 개인의 차트에 있는 진료 기록, 유전체 데이터, 임상 시험 데이터 등 수 많은 의료 빅데이터를 분석하여 의학적 통찰력을 도출하고 있다. 예를 들자면, 데이터에 기초하여 환자 개인의 질병을 진단하거나 예측하고, 그에 적절한 임상연구나 처방을 내려준다.[1]

또한, 왓슨은 특히 매일 쏟아져 나오는 수많은 의료 논문들, 임상 시험 결과들을 환자의 치료에 빠르게 반영 가능하다는 장점이 있다. 매년 종양학 부분에만 4만 건 이상의 논문이 쏟아지고, IBM은 2015년 기준 매일 20시간 이상 논문을 읽어도 모든 데이터를 얻을 수 없다는 통계를 내놨다.[2,3] 왓슨은 이렇듯 의사의 진단을 도와줄 뿐 아니라, 새로운 데이터를 끊임없이 학습하며 의사들의 짐을 덜어주는 역할을 할 수 있다.

이에 반해서, 한의학 도메인에서는 이렇다 할 인공지능 연구가 진행되고 있지 않다. 대형 종합 병원들에 비하여 EMR이나 환자 차트가 정밀하게 관리되지 못하고 있고, 이에 따라 데이터 전처리는 더욱 어려워지며 이는 자연스럽게 인공 지능에 대한 연구를 어렵게 하고 있다. 이러한 상황은 이전부터 주욱 문제가 제기되어오던 한의학에 대한 과학적인 신뢰를 더욱 떨어뜨릴 수밖에 없다.

또한, 근래에 중풍, 안면마비, 신경계통 질환 등 다양한 중례에서 한, 양방 협진 치료의 효율성이 증명되며 정부에서도 여러 정책을 통해 협진을 유도하고 있으나, 협진이 제대로 이루어지는 병원은 극소수에 그치고 있고 연구 또한 중례 보고에만 그치고 있다.[4] 일본의 경우 2017년 일본한방생약제제협회가 실시한 조사에서, 일본의 양방 의사 중 한약을 처방하고 있는 의사는83.5%에 달하는 것으로 나타나는 등 비교적 협진 체계가 잘 이루어져 있음을 알 수 있다.[5] 이처럼 한의학을 과학화하려는 목적 이외에도 조금 더 협진을 활발하게 장려하기 위해서는 임상 데이터들의 표준화와 체계화가 필수이다.[6] 최도영 대한한의학회 회장 역시 한의학계가 당면한 가장 큰 과제는 표준화와 학술적 근거 마련이라고 밝힌 바 있다.[7] 이를 위해서는 지금과는 달리 데이터베이스의 구축을 통한 데이터 체계화가 필수이며, 체계화된 데이터를 통해 의료 근거를 마련해야 한다.[8]

본 논문에서는, 한의학의 과학화, 체계화를 위해 데이터를 수집하는 단계부터 저장, 그리고 이를 활용할 수 있는 적절한 알고리즘 도출과 서비스 플로우까지 일련의 단계를 모색하려 한다. 논문은 크게 두 파트로 나누어지며, 첫 번째 파트는 데이터 처리 및 적절한 알고리즘 선택에 대한 내용이고 두 번째 파트는 선택한 알고리즘을 어플리케이션에 적용시키고 데이터베이스와 연결하는 서비스 플로우에 관한 내용이다. 한의학 도메인이 굉장히 광대하기 때문에 하나의 주제에 집중하기 위해서 한의학 진단에서 제일 먼저 판별하는 사상체질 진단으로 시작한다. 기존에 병원에 저장되어 있는 사상체질 설문지 pdf 파일을 머신러닝 알고리즘에 적용하기 위하여 csv파일로 변환한다. 또한, 머신러닝 및 딥러닝 알고리즘을 여러 개 적용하여 가장 높은 정확도를 보이는 것을 어플리케이션에 적용하고 이를 데이터베이스에 연결하여 환자와 의사에게 편리함을 줄 수 있는 예측 결과 화면을 제공하며, 추후에는 데이터의 자동적인 누적과 이를 통한 알고리즘의 정확도 개선을 꾀하려 한다.

Ⅱ. 머신러닝을 사용한 사상체질 분류

2-1 사상체질

이제마는 사람마다 타고난 체질이 다르기 때문에 같은 병이라도 체질에 따라 그 치료가 달라야 한다고 주장하였으며, 이것을 바탕으로 동의수세보원에서 사상의학을 제창하였다. 사상의학에서는 사람을4가지 체질로 구분하며 각 특성에 따라 특별한 건강 정보와 치료 방법을 달리한다. 4가지 체질에는 태양인, 태음인, 소양인, 소음인이 있으며 한방병원에서는 먼저 체질을 판별한 후 이러한 체질을 기준으로 병을 진단하여 각 체질에 따른 처방을 한다.

사상체질을 진단하기 위한 방법으로는 용모에서 나타나는 느낌을 보는 용모사기 방법과 체형·기상을 보는 체형기상 방법, 내면적인 성격을 살피는 성질재간 방법, 병의 특이한 증상과 약물 반응을 체크하는 병증약리 방법 등이 있다. 사상의학은 체질에 따라 처방법이 달라지므로 체질진단의 정확도가 가장 중요한 요소 중 하나라고 할 수 있다. 그러나 이를 활용하기 위해서는 사상이론에 대한 깊은 이해와 많은 임상 경험이 필요하기 때문에 실제로 사상의학을 임상에 바로 적용하기가 어려우며, 또한 한의사의 주관적 판단에 의거하여 진단하기 때문에 그 객관성에 대한 문제가 제기되어 왔다. 따라서 이를 해결하기 위해 사상체질 진단의 객관화와 신뢰도를 높일 수 있는 방법이 꾸준히 요구되고 있다.[9]

체질을 판별하는 여러 방법 중에서도 설문지는 표준화 과정

을 거쳐서 널리 사용되어 지고 있으며 가장 활발히 연구되고 있는 분야이다.[10],[11] 설문지는 단 시간에 환자의 정보를 신속하고 다양하게 취할 수 있으며, 데이터를 상수화 시키기 용이하다는 장점이 있다.

본 연구는 가산의료재단 광동한방병원의 도움을 받아 사상체질 분야에서 25년 간 연구와 진료를 해온 문병하 원장과 함께 이루어졌다. 병원에 내원한 환자들 중에 병원에 오랜 기간 내원하여 체질이 비교적 확실하게 판명된 환자들을 선별하여 설문지를 추려냈다.

표 1. 전처리 된 설문지 데이터
Table 1. pre-treated questionnaire data

Num	Sex	2a	2b	2c	6a	6b	6c	…	26d	Label
1	0	1	0	0	1	1	0	…	0	0
2	1	1	0	1	0	1	1	…	1	1
3	1	0	1	1	0	0	0	…	0	0
…	…	…	…	…	…	…	…	…	…	…
100	0	1	0	0	1	0	0	…	0	2

2-2 데이터 전처리

매우 희소한 태양인을 제외하고 다양한 성별과 연령대의 소양인, 소음인, 태음인을 각각 30명, 30명, 40명으로 총 100명을 추려내었고, 이를 다양한 알고리즘에 적용해보기 위하여 엑셀 시트에 정리하였다. 모든 설문을 넣지 않고, 전문가의 판단 하에 가장 유의성이 높은 13개의 문항만을 선별한 후, 액셀에 표와 같이 정리하였다. 각 설문 문항에 다양한 보기가 있고 중복 선택이 가능하기 때문에, 체크한 보기는 1, 체크하지 않은 보기는 0으로 정리하였다. 예를 들어 2a, 2b, 2c의 문항의 경우 각각 2번 문항의 첫 번째, 두 번째, 세 번째 체크박스를 의미한다. 또한, 라벨링의 경우에는 태음인을 0, 소음인을 1, 소양인을 2로 사용하였다. 수집하고 전처리한 데이터를 훈련 집합(Training Data Set)과 검증 집합(Test Data Set)으로 나누어 SVM, Gaussian, Ensemble, K-neighborsClassifier, DNN 총 5가지의 알고리즘을 각각 훈련, 검증하였다.

훈련 집합과 검증 집합을 분리하는 방법은 다양한 방법이 있으나, 훈련 집합과 검증 집합을 단순 비교해서 모델의 성능을 측정하는 방법은 타당도의 문제가 있다. 이러한 문제를 해결하기 위해서 자주 쓰이는 K-Fold Cross Validation을 사용하였다.

K-Fold Cross Validation은 데이터 집합을 K개의fold로 분할하여 K-1개의 fold를 훈련 집합으로 사용하고 나머지 하나의 fold를 검증 집합으로 사용하는 방법이다. Fold들을 조합하여 모델을 훈련하고 검증한다면 여러 개의 검증결과들을 얻을 수 있고 평균값을 이용해서 모델의 성능을 더욱 정확하게 평가할 수 있다는 장점이 있다. 데이터를 상호 배타적으로 K개로 분할해 훈련 데이터와 검증 데이터를 교차해 가면서 모형을 평가하기 때문에 실질적인 모형의 정확도를 추정하는 데에 있어 효율적인 방법이다.[12]

2-3 머신러닝/딥러닝 알고리즘

머신러닝 알고리즘은 python의 sklearn 라이브러리를 사용했으며, 딥러닝 알고리즘은 tensorflow를 backend로 keras를 사용하였다.

1) KNeighborsClassifier

KNeighborsClassifier은 새로운 데이터를 입력받았을 때 가장 가까이 있는 데이터가 무엇인지를 중심으로 새로운 데이터의 종류를 판별하는 알고리즘이다. 즉, 서로 가까운 점들은 유사하다는 가정 하에 이루어지는 알고리즘이다. 이 때 K값, 즉 주변에 존재하는 데이터 개수를 몇 개로 볼 것인지가 가장 중요하며, 또한 거리를 어떻게 측정할지 결정하는 것도 중요하게 작용한다.[13]

2) Gaussian Naïve Bayes

Naïve bayes는 조건부 확률 기반 생성 모형에 속하며, 입력값과 출력값의 관계를 구하기 위해 우도, likelihood를 구하고 사전확률과 사후확률을 이용하여 입력값을 넣었을 때 출력값의 클래스 중 어떤 값이 나올지 확률을 계산하고 가장 높은 확률을 가진 클래스를 반환한다.[14]

3) Linear SVM / SVM Kernel

SVM 은 2차원 선, 초평면으로 분류하기 힘든 데이터셋을 분류하기 위한 머신러닝 모델이다. 즉, 차원데이터들을 한번 더 가공하여 새로운 데이터 차원을 추가하여 차원을 확장시켜 분류 평면을 만들고 다시 원래의 차원으로 투영시키는 원리이며 분류선을 기준으로 다른 클래스와 거리가 가장 큰 폭(margin)을 찾는 것이다. SVM에서 예측할 데이터와 학습된 데이터 포인트 사이의 거리를 가우시안 커널을 사용하여 계산한다. LinearSVC는 데이터셋에 이미 존재하는 차원 데이터들을 조합, 가공하여 새로운 차원을 추가했지만 어떤 차원을 추가하는

것이 효과적인지 아는 것은 어렵다. 따라서 새로운 차원을 만들지 않고 커널 기법을 사용하여 확장된 특성에 대한 데이터 포인트들의 거리를 계산하여 SVM을 적용한다. SVC() 생성자에서의 매개변수 kernel은 어떤 kernel 기법을 사용할지 선택하며 선택 가능한 커널 기법은 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed' 가 있고 매개변수 gamma와 C의 적절한 선택이 매우 중요하다.[15]

4) Ensemble Model – Random Forest / GradientBoosting

앙상블 모델은 여러 머신러닝 모델을 연결하여 더욱 성능이 뛰어난 모델을 만드는 기법이다. Random Forest와Gradient Boosting 두 모델 모두 결정트리에서 출발한다. Random Forest는 여러 결정 트리의 묶음으로, 서로 다른 방향으로 overfitting이 일어난 트리를 다수 생성하고, 각 트리의 결과값을 평균내서 과적합을 줄이는 방법이다. n_estimators, max_feature, max_depth와 같이 중요 매개변수를 변화시켜 accuracy를 높일 수 있다. Gradient Boosting은Random Forest와는 다르게 이전 트리의 오차를 줄이는 방식으로 트리를 만들어 나간다. 이 모델 또한n_estimators와 learning_rate의 적절한 balance가 중요하며 max_depth를 조정할 수 있다.[16], [17]

5) Deep Neural Network

Multi Layer Perceptron 은 하나의 퍼셉트론을 여러 층으로 쌓은 것으로 이 층이 깊을 때 Deep Neural Network(DNN)으로 명명한다. 왼쪽부터 시작하는 첫 번째 층은 입력층이고 맨 오른쪽에서 끝나는 마지막 층은 출력층이다. 그리고 그 사이에 있는 중간 단계에 있는 층들을 은닉층이라고 일컫는데 이 은닉층이 많을수록 깊은 신경망이 된다. 모델을 변형하는 방법은 여러 방법이 있는데 먼저 은닉 유닛, 은닉 층을 추가하는 방법이 있다. 또한, sigmoid / tanh / relu 등 다양한activation 함수를 변화시키고 그에 따라 softmax / adam 등 optimizer를 변화시키는 방법이 있다.[18]

2-4 분류 정확도 비교

각 분류 알고리즘들을 훈련 집합으로 훈련시키고 난 후, 예측 값을 검증 집합의 정답 레이블과 비교하여 일치하는 것을 기준으로 정확도를 출력하였다. 고전적인 머신러닝 알고리즘들 중에서는 SVM 모델과 Random Forest 모델이 0.7로 가장 높은 정확도를 보였으며, DNN 알고리즘 또한 0.8로 비교적 높은 정확도를 기록하였다. Deep Neural Network에는 Adam[19]과 Relu[20]를 사용하였는데, DNN의 정확도가 높게 나온 이유는 훈련셋에 과적합되는 것을 방지하기 위하여 층마다 DropOut을 시킨 것이 의미가 있는 것으로 사료된다.[21] 최종적으로 가장

표 2. 알고리즘 별 정확도
Table 2. Accuracy of Algorithms

Algorithm Name	Accuracy
KNeighbors Classifier	0.6
Gaussian Naïve Bayes	0.55
Support Vector Machine	0.7
Random Forest	0.7
Deep Neural Network	0.8

높은 정확도를 보인 DNN 분류 알고리즘을 채택하여 어플리케이션과 데이터베이스 디자인을 하였다.

Ⅲ. 서비스 플로우

3-1 어플리케이션 디자인

알고리즘의 정확도 비교에 더불어 환자들의 사상체질 문진검사 데이터 취득을 위한 스마트폰 어플리케이션 또한 개발하였다. 기존에는 병원을 내원한 환자들이 직접 펜을 이용하여 문진검사 서류를 작성하고 환자가 작성한 서류를 스캔하여 pdf 파일로 저장해 두었다가 나중에 필요할 때 해당 파일을 다시 찾아보는 방식을 채용하고 있었다. 환자들의 문진검사 결과 데이터를 체계적으로 데이터베이스화 하기 위하여 스마트폰, 태블릿에서 작동하는 문진검사 어플리케이션을 개발하였다. 안드로이드 기반으로 만들어진 어플리케이션은 그림과 같다. 처음 어플을 실행하면 문진검사에 대한 소개 후25개 문항에 대하여 환자는 자신이 해당하는 항목에 터치하여 체크하고 다음문항이나 이전문항으로 이동할 때에는 좌, 우 Swipe 혹은 하단의 버튼을 통해 이동할 수 있도록 하였다. 모든 항목에 대하여 답변을 완료하면 마지막에는 환자의 이름과 생년월일을 입력한 후 5초 후에 자동으로 종료되도록 설계하였다.

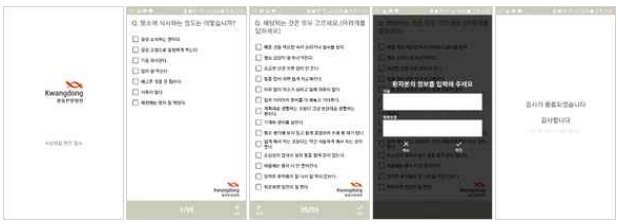

그림 1. 안드로이드 기반 사상체질 문진검사 어플리케이션
Fig 1. Android-Based Sasang Constitution Application

3-2 데이터베이스/웹 디자인

위에서 기술한 어플리케이션을 통해 환자가 작성한 25개 질의에 대한 응답들은 Boolean 타입의 데이터로 어플리케이션과 연결된 Google Firebase에 환자번호, 이름과 함께 저장된다. Firebase는 Google에서 공개한 클라우드 서버로 실시간 데이터베이스를 활용하여 새로운 데이터의 추가, 수정, 삭제 이벤트를 다른 서버나 클라이언트 단에서 감지하여 반응하는 것을 쉽게 제작할 수 있다.

안드로이드 어플리케이션에서 Firebase로 전송된 데이터들이 추가되면 별도로 동작하는 서버에서 이를 감지하여 새로 추가된 데이터들을 2장에서 선택한 분류 알고리즘 모델을 이용하여 환자의 사상체질을 예측하고 예측한 결과를 다시 Firebase에 저장되어 있는 해당 환자의 정보를 업데이트 한다. 그리고 이 과정에서 Firebase는 NoSQL 형태로 데이터를 저장하기 때문에 추후에 데이터 쿼리의 편의를 위해 MySQL에도 환자의 정보와 문진검사 응답 데이터를 저장하도록 하였다.

서버 언어로는 Node.js를, 데이터베이스는 MySQL을 사용하였으며 처음에는 각 문항을 하나의 column으로 구성하였지만 중복응답이 있는 경우가 많아 문항의 각 응답 항목을 하나의 column으로 하여 해당된다고 응답한 경우에는 1, 그렇지 않은 경우에는 0으로 데이터를 저장하도록 하였다. Firebase의 실시간 데이터베이스에서는 트리 형태로 데이터를 저장하였으며 마찬가지로 문항의 각 응답 항목을 하나하나 key – value 형태로 저장하였다.

Firebase에 환자의 예상 사상체질 결과까지 업데이트가 완료되면 의사가 사용하고 있는 PC에서는 웹 브라우저로부터 push 알림을 받게 되고 웹페이지를 통해 현재 진료하고 있는 환자의 문진검사 응답과 예측된 사상체질을 열람할 수 있도록 제작하였다.

그림 2. 의사용 환자 문진표 확인 웹 페이지
Fig 2. web page that contains patient's information using by doctor

그림 3. 서비스 플로우 개요
Fig 3. Service Flow

Ⅳ. 결 론

본 논문에서는 사상체질의 판별을 위해 설문지 데이터를 전처리하여 다양한 알고리즘을 통해 검증해보고, 가장 성능이 좋은 알고리즘을 선별하였다. 또한, 선별한 알고리즘을 중심으로 환자용 어플리케이션, 의사용 웹페이지, 서버와 데이터베이스를 디자인하였다.

내원한 환자의 경우 편리하게 문진표를 작성할 수 있으며, 추후 알고리즘의 정확도가 더욱 개선될 경우 앱스토어에 정식 출시하여 헬스컨설턴팅 서비스를 받을 수 있도록 할 예정이다. 또한, 의사의 경우 디자인 한 웹페이지를 통해 알고리즘이 판별한 예상 체질을 보고 자신의 판단을 한 번 더 되돌아볼 수 있게 끔 하여 의사결정에 도움을 줄 수 있다. 뿐 만 아니라, 데이터베이스를 MySQL, NoSQL를 모두 사용하여 디자인했기 때문에, 시간이 지나 데이터가 누적되면 바로 정확도 개선을 위해 더 많은 샘플로 training을 할 수 있으며, 추후에는 침술이나 물리치료 등 더욱 많은 주제에 확대 적용할 수 있을 것이다.

이후 일정 기간이 지난 후에 본 논문에서 개발한 데이터베이스에 누적되어 있는 데이터를 샘플로 추가하여 훈련시키는 방법을 통해 모델의 성능을 더욱 향상시킬 것이다. 또한, 체질 판

별 이외에도 통증 부위 별 침술, 물리치료를 통한 회복 추이 등 다양한 주제들 중 하나를 선별하여 추가적인 기능을 갖도록 서비스를 확대할 예정이다.

참고문헌

[1] K. Y. Lee, “ArtificialIntelligence Technology Trends and IBM Watson References in the Medical Field”, *Korean Medical Education Review*, Vol. 18, No. 2, pp. 51-57, 2016.

[2] IBM Watson Hard At Work: New Breakthroughs Transform Quality Care for Patients [Internet]. Available from: http://www-03.ibm.com/press/us/en/pressrelease/40335.wss?i=1360645029661.

[3] Gil Hospital adopts IBM WFO for the first time in South Korea[Internet]. Available from: http://www-03.ibm.com/press/kr/ko/pressrelease/50591.wss#release.

[4] K. I. Han, “Reviewing Research of Eastern-Western Integrative Medicine Studies in Korea”, *Journal of Korean Medicine Rehabilitation*, Vol. 28, No. 1, pp. 53-60, January, 2018.

[5] 84 percent of modern medicine doctor prescribe oriental medicine in Japan[Internet]. Available from: https://news.joins.com/article/22702871.

[6] S. Y. Lee, “Policy Directions to Promote Integrated Services of the Western and the Oriental Medicines”, Health and Welfare Forum, Vol. 97, No. 0, 2004.

[7] Choi Do-young of the Korean Medical Association said, "We are focusing our efforts to succeed in the standardization project of oriental medicine."[Internet]. Available from: http://www.sisaon.co.kr/news/articleView.html?idxno=53884

[8] K. S. Moon, “Oriental Medical Ontology for Personalized Diagnostic Services”, *Journal of the Korea Society of Computer and Information*, Vol. 15, No. 1, pp. 23-30, 2010

[9] S. H. Shin, “Study on the Discrimination of Constitution Using Pulse Wave”, *Korean J. Oriental Physiology & Pathology*, Vol. 22, No. 6, pp. 1403-1409, December, 2008.

[10] S. H. Kim, “Study on the Standardization of Questionnaire for the Sasang Constitution Classification(QSCC II)”, *Journal of Korean Medicine*, Vol. 17, No. 2, pp. 337-393, 1996.

[11] Y. T. Kim, “Optimization of Oriental Medicine Survey Using Data Mining Techiniques”, *Journal of Knowledge Information Technology and Systems*, Vol. 6, No. 6, pp.15-22, 2011.

[12] Kohavi, Ron, "A study of cross-validation and bootstrap for accuracy estimation and model selection.", Ijcai, Vol. 14, No. 2, pp. 1137-1145, 1995.

[13] Peterson, Leif E, “K-nearest neighbor”, *Scholarpedia*, Vol. 4, No. 2, 2009.

[14] Kohabi, Ron, “Scaling up the accuracy of Naive-Bayes classifiers: a decision tree hybrid”, KDD, Vol. 96, pp. 202-207, 1996.

[15] Amari, Shun-ichi, and Si Wu, "Improving support vector machine classifiers by modifying kernel functions.", *Neural Networks*, Vol. 12, No. 6, pp. 783-789, 1999.

[16] Liaw, Andy, Matthew Wiener, “Classification and regression by randomForest”, R news, Vol. 2, No. 3, pp. 18-22, 2002.

[17] Friedman, Jerome H, "Greedy function approximation: a gradient boosting machine.", *Annals of statistics*, pp. 1189-1232, 2001.

[18] Schmidhuber, Jürgen, "Deep learning in neural networks: An overview.", Neural networks, Vol. 61, pp. 85-117, 2015.

[19] Kingma, Diederik P, Jimmy Ba, “Adam: A method for stochastic optimization.“, *International Conference on Learning Representations (ICLR)*, Vol. 5. 2015.

[20] Nair, Vinod, Geoffrey E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10),* pp. 807-814, 2010.

[21] Srivastava, Nitish, “Dropout: a simple way to prevent neural networks from overfitting.”, *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929-1958, 2014.

이재화(Jae-Hwa Lee)

2017년~현 재: 연세대학교 기술경영학협동과정 석사과정
※관심분야 : 머신러닝, 인공지능, 데이터베이스

이현학(Hyun-Hak Lee)

2017년~현 재: 연세대학교 기술경영학협동과정 석사과정
※관심분야 : 머신러닝, HCI, 서비스 디자인