Regular paper
Journal of the KIECS. pp. 781-786, vol. 14, no. 4, Aug. 31. 2019, t. 96, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2019.14.4.781

데이터마이닝을 활용한 유전자 질병 분석을 위한 MKSV시스템 구현

정유정* · 최광미**

For Gene Disease Analysis using Data Mining Implement MKSV System

Yu-Jeong Jeong* · Kwang-Mi Choi**

요 약

오늘날 다양한 생명현상을 다루고있는 질병연구와 같은 효율적인 목적을 달성하기 위해서는 이들 연구로부터 획득한 빅데이터를 처리하여 효과적인 현실적 가치를 부여할 수 있어야 한다. 본 논문에서 제안한 MKSV알고리즘은 최적의 확률분포를 추정하여 입력패턴을 결정 한 후 데이터마이닝 기법으로 분류한 결과 효율적인 계산량과 인식률을 획득할 수 있었다. MKSV 알고리즘은 유전자 데이터의 확률적 흐름을 시뮬레이션하여 빅데이터의 데이터마이닝 과정을 통해 데이터를 분류하여 빠르고 효과적인 성능 향상을 보임으로써 현 사회에 급증하는 질병과 유전자의 관련성을 연구하는 데 유용할 것이다.

ABSTRACT

We should give a realistic value on the large amounts of relevant data obtained from these studies to achieve effective objectives of the disease study which is dealing with various vital phenomenon today. In this paper, the proposed MKSV algorithm is estimated by optimal probability distribution, and the input pattern is determined. After classifying it into data mining, it is possible to obtain efficient computational quantity and recognition rate. MKSV algorithm is useful for studying the relationship between disease and gene in the present society by simulating the probabilistic flow of gene data and showing fast and effective performance improvement to classify data through the data mining process of big data.

키워드
Hidden Markov Model, SVM Model, Bigdata, DNA data
은닉 마코브 모델, SVM 모델, 빅데이터, 유전자 데이터

Ⅰ. 서론

다가오는 4차 산업혁명으로 인한 우리사회 전반에 급속한 영향을 미칠것으로 예측된다. 그 중 생명공학의 급속한 발전은 대규모의 바이오데이터의 생성과 처리가 화두가 되고 있다. 이러한 바이오데이터를 효율적으로 분석하기위한 빅데이터를 이용한 데이터마이닝 기법의 분류 방법들이 다양하게 연구되어지고 있다. 하지만 복잡한 생명 현상을 규명하기위해서는 유전체의 서열과 단백질이 가지는 고유 기능과 단백

* 조선대학교 sw융합교육원(narimono@hanmail.net)
**교신저자 :조선대학교 sw융합교육원
· 접 수 일 : 2019. 06. 30
· 수정완료일 : 2019. 07. 23
· 게재확정일 : 2019. 08. 15
· Received : Jun. 30, 2019, Revised : Jul. 23, 2019, Accepted : Aug. 15, 2019
· Corresponding Author : Kwang-Mi Choi
Dept. of Convergence Edu, Chosun University,
Email : iplab@nate.com

질간의 상호 작용이 질병과 밀접한 관계를 가지고 있다는 것은 일반적으로 알려진 사실이다[1]. 특히 유전체의 기능을 밝히는 연구과정은 유효한 데이터를 선택하기 위한 특징 선택과정이 매우 중요하다. 이 과정에서 선택된 유전자들은 분류자를 생성하기 위해 입력 값을 필요로 하는데, 이때 선택되는 특징의 종류와 갯수는 분류 결과에 큰 영향을 미치게 된다. 또한 데이터에 따라 효과적인 결과를 내는 분류자가 다르므로 이에 대한 연구도 이루어지고 있으나 데이터의 불완전성이나 알고리즘의 한계 등으로 인하여 완벽한 분류자를 밝히기는 어렵다. 현대를 살아가면서 미래에 대한 다양한 예측이 필요하며 인공지능에서도 다양한 예측 알고리즘을 제시하고 있다. 각 알고리즘의 특징이 달라 적용 분야에 따라 진단이나 예측의 정확도가 차이가 나는 문제점이 있다[2]. 본 논문에서 제안한 MKSV 알고리즘은 HMM likelihood 최대점을 사용하였고 매개변수 학습 효과를 이용하여 특징을 분류하고 데이터마이닝 분류기법을 통해 질병 유전자 분류를 사용하여 분석함으로써 성능을 높였다. 유전자 데이터에 포함된 특성들의 발생과정을 확률적으로 모델화한 것으로 적은 계산량으로 좋은 인식률을 얻을 수 있었으며 이는 유전자 데이터의 확률적 흐름을 시뮬레이션을 통해 유전자 발현 데이터에 적용하여 데이터마이닝 함으로써 더 효율적인 분류 결과를 얻을 수 있었다.

II. 본 론

2.1 데이터마이닝 기법

데이터마이닝은 일반적으로 대규모의 데이터에서 가치있는 정보를 추출하기위한 과정을 말하는데 보다 세부적으로 살펴보면 데이터 내의 관계나 패턴 규칙 등을 발견하기위해서 대량의 데이터로부터 자동화 혹은 반자동화 도구를 활용해 탐색 분석 모델링하는 과정을 말한다[3].

일반적으로 데이터마이닝은 기 수집된 대용량 데이터에서 가치있는 정보를 추출하는 것을 목표로 통계적 분석 방법론과 함께 기계학습 인공지능 기법을 결합하여 사용한다[4-5].

특히 의학분야의 경우 특정 질환을 앓고 있는 환자들의 진단이나 예후를 파악하기위한 연구들에서 많이 활용되고 있으며 대표적으로 특정 질환을 앓고있는 환자 데이터를 이용하여 각질환의 발생 패턴이나 확률을 예측하기위한 연구들이있다[6-8].

데이터마이닝의 대표적인 기능으로는 분류,추정,예측,연관분석,군집분석등이 있다.

특히 군집분석(clustering)은 주어진 데이터를 의미있거나 유용한 그룹 또는 군집으로 나누는 것으로, 데이터에 내재된 있는 그대로의 구조를 파악하기 위한 것이다. 군집분석에서는 사전에 정해진 범주를 가정하지 않기 때문에 비 표식 자료를 바탕으로 자율학습방법으로 수행된다. 대표적인 군집분석으로 클러스터링이 이에 해당한다[9].

2.2 자기조직화 함수(Self-Organizing Maps)

신경회로망의 SOM(: Self-Organizing Maps)알고리즘은 클러스터의 개수가 알려져 있을 때 주어진 다차원 데이터들을 가장 근접한 클러스터에 사상(mapping)시켜주는 방법이다[10].

Elastic network를 구성하는 map에 임의로 선택한 원소를 입력으로 주면서, 동시에 map의 가중치(weight)를 반복적으로 수정하여 입력 데이터들의 클러스터 이동이 없을 때까지 반복한다. 가중치 벡터의 갱신을 위해서 사용되는 가중치 벡터 갱신 함수는 식(1)과 같다.

$$w^j(t+1) = w^j(t) + \alpha(t)\left[x(t) - w^j(t)\right]$$
$$\alpha(t) = 0.1(1 - t/10^f) \text{ 단, } f\text{는 상수} \quad (1)$$

가중치 벡터 갱신 함수에 따라 각 출력 노드의 가중치 벡터는 그 출력 노드에 포함된 출력 노드를 승자로 택한 입력 데이터 방향으로 이동하게 된다. 이 움직임의 변화는 초기에는 매우 산만하나, 입력 벡터의 수가 어느 정도 이상이 되면 거의 변하지 않고 안정화된다.

이 방법은 복잡한 다차원 데이터의 클러스터링에 적합하며 결과의 가시화가 쉽고, 클러스터링 결과를 사용자가 제어할 수 있다는 장점을 가지고 있다[11].

2.3 최적 상태 state sequence 결정: viterbi 알고리즘

주어진 관측열에 대응하는 최적의 상태열을 찾는 방법에는 동적 프로그래밍(Dynamic programming)기법 중의 하나인 Viterbi알고리즘을 적용한다[9].

관찰된 열 $O=(O_1,O_2,...O_t)$이 주어졌을 때, 하나의 가장 좋은 상태열 $q=(q_1,q_2,...q_T)$을 찾기 위해서는 식(2)과 같다. $\delta_t(i)$ 는 시간 t에서 하나의 경로를 따르는 가장 큰 확률을 뜻하는 식(3)와 같다.

$$\delta_t(i)=\max_{q_1,q_2,\cdots\cdots q_{t-1}} P[q_1,q_2,\ldots,q_{t-1},q_t=O_1,O_2,\ldots O_t|\lambda] \quad (2)$$

$$\delta_{t+1}(j)=\max_i[\delta_t(i)a_{ij}]b_j(O_{1+1}) \quad (3)$$

귀납에 의하여 식(2)는 식(3)과 같이 확장이 가능하며, 이를 이용하면 시간 t 뿐만 아니라 $t+1$에 대해서도 최대 확률을 가지는 상태들의 순서를 구할 수 있다[12].

III. 실험결과

본 논문에서서는 공개적으로 오픈된 마이크로어레이 실험데이터 유전자 발현 정보들 중에서 위암 데이터를 사용하였다.

본 논문에서 제안한 알고리즘은 데이터의 흐름 이론과 확률에 기반을 둔 클러스터링 알고리즘으로 빠르고 효과적이다. 표준화 데이터 확률을 계산하는 절차는, 다양한 유사성 척도 중에서 가장 많이 보편적으로 쓰이는 유클리디안 유사성 척도를 이용하여 기존 방식의 문제점을 보완한다. MKSV알고리즘을 설명하기 위해 데이터내의 data work의 확률 값으로 마코브 행렬 값을 사용하는데, 이 행렬은 각 열의 확률 합이 1을 넘지 않는다.

본 논문에서 제안한 확률 추정, 최적상태 순서 결정, 매개변수 추정 문제는 Forward Algorithm훈련이 끝난 다음에는 가장 큰 확률을 지닌 패스를 찾기 위해 Viterbi Algorithm에 의해 계산하여 모델 매개변수 추정과 은닉마코브모델의 구조를 통해 최적의 매개변수를 구한다.

두 번째로 은닉마코브모델에서 최적의 매개변수 값을 찾고, SVM벡터 입력 값으로 사용한다. 아주 많이 흩어진 벡터들을 두개의 클래스로 나누어 분류(Classfication)시키는 특성을 지닌다. SVM은 구분되지 않은 새로운 데이터를 가지고 구분되는 즉 Positive(+) 정상적인 유전자 판정, Negative(-) 비정상적인 암 유전자 판정으로 분류(Classification) 시키는 학습 알고리즘의 하나이다.

은닉마코브모델에서 매개변수 학습을 통해 일차적으로 분류한 다음, SVM 벡터 입력 값을 받아들여 가능한 높은 차원의 공간으로 확장시키고 이때 Positive와 Negative로 나누는 최적의 분할 공간을 구하기 위해 학습 시킨다. 따라서 최적의 초평면을 구하기 위해 훈련 데이터들이 되도록 적확히 분류되도록 해야 하며 margine을 최대로 해야 한다. 본 실험에서는 은닉마코브모델은 학습 과정과 인식 과정으로 구분되어진다. 우선 학습 과정에서는 입력패턴의 특징점을 상태천이 확률분포로 나타내고, 어떤 상태에서 특정하나 심벌이 나타날 수 있는 확률분포를 갖는 과정을 마코브프로세서로 가정하고, 학습데이터를 통하여 이들 확률분포를 추정한다. 이 추정된 확률분포를 바탕으로 입력된 입력 패턴이 그 모델에서 발생하였을 확률을 계산하여 인식을 하게 된다. 위의 과정에서 알아본 바와 같이, 은닉마코브모델은 어떤 관측할 수 있는 과정에서 상태가 있다는 가정을 통하여 확률과 각각의 전이가 일어난 관측된 심벌이 현재의 상태에 의존하는 관측확률을 구하게 된다.

은닉마코브모델은 세 가지 요소로 구성되는데 상태의 개수, 시간에 따른 상태의 변화를 결정하는 상태천이 확률분포, 그리고 각 상태에서의 출력 심볼의 확률분포이다. 각 상태들은 직접적으로 관찰 가능하지는 않으나 대신 각 상태들이 일정확률을 가지고 만들어내는 심벌을 보고 원래 상태를 추정하는 방법이다.

이러한 정의를 이용하여 다음 두 단계를 거치게 된다. 즉 모델형성 과정과 형성된 모델을 이용하여 관측 심벌의 확률 값을 구하는 두 과정을 거치게 된다. 1단계에서는 모델 형성과정 $P(O|\lambda)$를 최대로 하는 모델 파라미터 $\lambda=(A,B,\Pi)$를 구하는 문제이고 2단계에서는 모델 인식과정에서 관측된 심벌의 시퀀스 $O=O_1O_2...O_T$와 모델 $\lambda=(A,B,\Pi)$가 주어졌을 때 likelihood $P(O|\lambda)$를 구하는 문제이다. 1단계에서 얻

어진 관측 심벌의 시퀀스를 이용한 은닉마코브 모델링은 Viterbi알고리즘을 모델 파라미터 $\lambda=(A,B,\Pi)$를 설정하였다. 2단계의 인식단계에서는 각각 설정된 유전자별 은닉마코브모델과 부합하는 확률은 Forward알고리즘을 이용하여 산출하였다. 3단계에서는 은닉마코브모델에서 생성된 매개변수 생성과 학습을 통해 얻어진 변수를 통해 구해진 값을 입력 값으로 한다. 본 논문에서 SVM계산을 위해 필요한 커널함수는 RBF 커널함수를 사용했으며, SVM모델의 복잡성과 평활도에 대한 정도를 서로 보정해주는 C값과 커널 함수를 사용할 때 필요한 모수의 값은 한 번의 훈련 데이터를 가지고 결정한 후 그 값을 사용하였다. 본 논문에서는 이기종 플랫폼에서 동일한 생물학적 목적을 가지고 수행된 공개적으로 오픈된 마이크로어레이 실험데이터 유전자 발현 정보들 중에서 위암 데이터를 사용하였다. 실험 데이터는 9개의 샘플을 사용하였으며 MKSV알고리즘을 적용시키기에 앞서 각 데이터에 대해 유전자 선택 방법을 사용하였다. 유전자 수는 유전자 선택에서 첫 번째로 결정해야 하는 사항이다. 적절한 유전자의 수를 결정하는 것은 실험이나 경험에 의존하게 되므로 매우 어렵다. 본 논문에서는 이 실험은 직접 하지 않고 Tag and Hang의 실험을 통해 결정된 215개의 유전자를 선택했다[13]. 뽑힌 유전자만을 가진 데이터로 5개의 샘플 훈련 데이터와 4개의 훈련 테스트 데이터를 적용시켰다.

유전자 발현 데이터를 사용 할 수 있는 기존의 Clustering 도구를 사용하여 SOM알고리즘과 K-means알고리즘은 실험을 통하여 만들어진 노드들을 관찰해 보았다.

본 논문에서는 비교실험 대상인 K-means알고리즘 실험 조건은 K-means반복을 위한 매개변수 중 실행수는 50번, 정확도를 높이기 위해 여러 번의 실험 결과 Threshold값은 80%로, 최대 반복횟수 값이 높을수록 불필요한 유전자 분류들이 많아져서 가장 안전하게 결과 값이 나오는 50으로 설정하여 실험을 해보았다.

두 번째 실험에서는 SOM은 가장 많은 형태에 쓰이는 3×3의 SOM알고리즘을 이용하여 iteration 2000, alpha 값 0.05, radius 3.0 실험결과 9개의 노드가 만들어지고, 표본에 대한 클러스터도 구 할 수 있다. 은닉마코브모델에서 생성된 매개변수 생성과 학습을 통해 얻어진 변수를 통해 구해진 값을 입력 값으로 한다. 본 논문에서 SVM 계산을 위해 필요한 커널함수는 RBF커널함수를 사용했으며, SVM모델의 복잡성과 평활도에 대한 정도를 서로 보정해주는 C값과 커널 함수를 사용할 때 필요한 모수의 값은 한 번의 훈련 데이터를 가지고 Hsu and Lin에 따라 결정한 후 그 값을 사용하였다[14].

또한 기존의 방법과 본 논문에서 제안한 방법의 효율성을 검증하기 위해 민감도(Sensitivity)와 특이도(Specification)를 사용하여 생물학 데이터의 분류 성능을 평가하는 측도로 실험해보았다. 마지막으로 본 논문에서 제안한 MKSV알고리즘 프로그램 실험을 통해서 분류 정확도를 검증해 보았다. 각 샘플은 다른 종류의 암 조직으로부터 얻은 값이다.

그림 1은 데이터 셋을 실험에 적용하기 위하여 CSV 파일로 만들어 사용하였다.

그림 2는 K=4일때 centroid값으로 구분된 유전자를 표현한 그림이다.

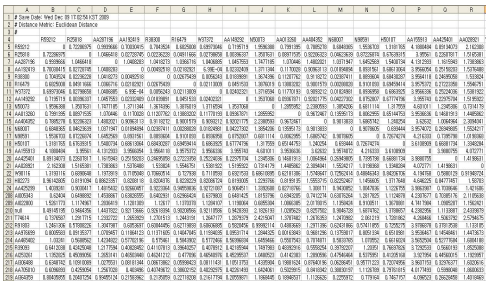

그림 1. 유클리디안 변환 테이블
Fig. 1 Euclidean transformation table

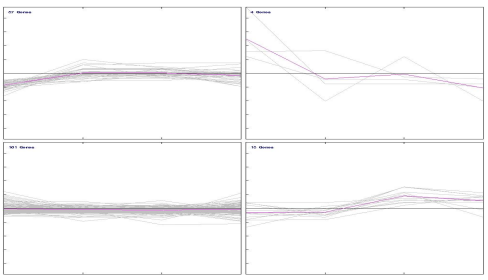

그림 2. K=4 일때 분류 샘플 유전자 표현 이미지
Fig. 2 K=4 classification sample gene expression image

두 번째 실험에서는 SOM은 가장 많은 형태에 쓰이는 3×3의 SOM알고리즘을 이용하여 iteration 2000, alpha 값 0.05, radius 3.0 실험결과 9개의 노드가 만들어지고, 표본에 대한 클러스터도 구 할 수 있다. 본 논문에서 제안한 MKSV알고리즘에 따르면 Step1은 원하는 데이터 파일을 불러들여서 Step2에서 HMM파라미터를 찾는다. 실험데이터의 초기 파라미터는 P(cancer|B)는 cancer데이터로 0.5, P(cancer|I)는 normal data로 가정하고 초기 값을 0.5로 설정을 한다. 상태전이 행렬은 일어날 수 있는 확률 상태 6가지로 설정을 한 다음 P(cancer|B)는 0.04, P(+cancer|B)는 0.94, P(+˜cancer|B)는 0.03, P(cancer|I)는 0.987, ,P(-cancer|I)는 0.007, P(-˜cancer|I)는 0.002의 파라미터 결과를 얻을 수 있었고, Step3 단계에서는 Viterbi알고리즘으로 최대 확률 값들의 누적치인 최적의 상태열을 추정하면 안정상태로 들어오는 값인 Step3의 결과 0.00116의 값을 얻을 수 있었다.

Step4에서 SVM은 은닉마코브 모델에서 매개변수 학습을 통해 일차적으로 분류한 다음, SVM 벡터 입력 값을 받아들여 가능한 높은 차원의 공간으로 확장시키고 이때 Positive 와 Negative로 나누는 최적의 분할 공간을 구하기 위해 학습 시킨다. 구분되지 않은 새로운 데이터를 가지고 구분되는 즉 Positive(+) 정상적인 유전자 판정, Negative(-) 비정상적인 암 유전자 판정으로 분류(Classification) 시킨다.

실험데이터 215개의 유전자 실험 결과 Positive Genes의 분류 중 실험 결과는 Positive Genes중 유전자 초기 Positive example로 선정되어 실험된 수는 141개이며, Positive에 따르는 유전자 분류 즉 Total Positive는 179개, Positive class에서 유전자를 포함하는 true Positive는 126개, Negative 유전자에서 긍정적인 수준을 포함하는 유전자 즉 False Negative의 수는 53개로 분류되었다.

Negative Gene 그룹 중에서는 유전자 초기에 Negative example로 선정된 샘플의 개수는 74개이며, Negative 유전자 분류에 따르는 Total Negative 개수는 26개, Negative class에서 유전자를 포함하는 True Negative는 12개, Positive 유전자에서 부정적인 negative를 포함하는 유전자 즉 False Positive는 14개로 분류되었다.

본 논문에서 제안한 MKSV알고리즘이 우수 하다는 것을 검증하기 위해 기존의 K-means와 SOM과 MKSV알고리즘의 실제 클래스와 분류기에 따른 결과를 비교하여 그 성능을 측정해 보는 실험이 민감도(Sensitivity)와 특이도(Specificity)인데 생물학 데이터의 분류성능을 평가하는 측도로써 사용한다.

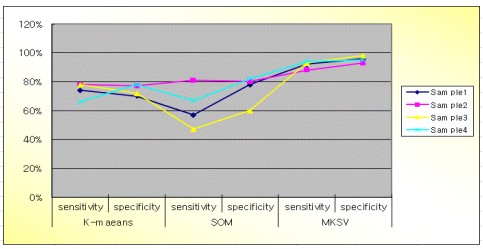

그림 3. K-means, SOM, MKSV의 민감도 특이도
Fig 3. Sensitivity specificity of K-means, SOM and MKSV

이는 K-means나 SOM보다 더 잘 분류 되는 것을 알 수 있었다. 하지만 sample2의 경우 다른 샘플의 경우와 차이가 나는 부분은 실험 유전자 중 나이등 여러 가지 환경적인 차이가 나는 유전자에 대한 영향이 있었던 것으로 보인다.

Ⅳ. 결론

본 논문에서는 독립적으로 학습된 알고리즘을 결합한 MKSV알고리즘에 대해 실험하였다. 모든 실험의 결과를 통해서 K-means, SOM, MKSV 알고리즘을 민감도, 특이도를 가지고 측정한 결과 본 논문에서 제안한 MKSV 알고리즘이 다른 알고리즘에 비해 향상된 분류 성능을 보이고 있음을 알았다. 본 논문에서는 방대한 빅 데이터를 데이터마이닝 작업을 통해 특별한 클러스터링 분석에서 나아가 유전자 질병을 조기 진단할수 있는 정보를 활용한 기능적인 방향을 제시할수 있다.. 향후 연구 과제는 다양하고 보다 체계적인 많은 데이터의 획득과 분석을 통해 빅데이터를 이용한 데이터의 획득과 분석을 통해 좀 더 효율적인 분류 유전자를 찾는 연구가 계속되어야하고, 이와 더불어 실험하는 사람이 데이터 마이닝을 이용한 데이터를 면밀히 조사하고 이에 아직 사용해보지 못한 또 다른 정규화 방법과 유의한 유전자 선택 방법

을 추가하여 더 많은 연구를 진행하고자 한다.

References

[1] E. Kim, J. Jeong, and B. Lee, "A Big Data Based Random Motif Frequency Method for Analyzing Human Proteins," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 6, 2018, pp. 1397-1404.

[2] G. Park and Y. Bae, "Performance Comparison of Machine Learning in the Various Kind of Prediction,"*J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 1, 2019, pp. 169-178.

[3] J. Michael and S. Gordon, *Data Mining Techniques:For Marketing,Sales, and Customer Relationship Management.* Seoul: Hankyung Corporation, 2010, pp. 10-15.

[4] S. Kim, Y. Kim, and R. Kim, *Convergence & consilience: communication research methods in a multiple media environment.* 발행도시: Korean J. of Communication Science, 2012, pp .53-81.

[5] G. Shmueli, R. Patel, and C. Bruce, *Data Mining for Business Intelligence:Concepts, and Applications in Microsoft Office Excel with XLMine.* 발행도시: Wiley, 2010.

[6] J. Ryu, S. Kim, J. Park, and J. Lee, "Risk Factors of Impaired Fasting Glucose and Type 2 Diabetes Mellitus - Using Datamining," *Korea epidemiological society*, vol. 28, no. 2, 2006, pp. 138-151.

[7] Y. Kim, "Screening test data analysis for liver disease prediction model using growth curve," Master's Thesis, *Yonsei University*, 2002, pp. 1-68.

[8] K. Lee, S. Park, S. Kang, and H. Kang, "Development of Prediction Model for Diabetes Mellitus Using Data Mining Technique," *Korean Journal of Health Policy and Administration,* vol. 16, no. 2, 2006, pp. 21-48.

[9] Y. Kim, "Development of advertising effect prediction model for celebrity models using Big Data," Master's Thesis, *Hanyang University*, 2019.

[10] T. Kohonen, "Exploration of very large databases by self-organizing maps," In *Proc. Conf. on Neural Networks,* Houston, TX, USA, 1997.

[11] H. Han, "*Introduction to pattern recognitio,*" hanbit media, 2011.

[12] C. Yoo and C. Park, "Single channel subband blind source separation using temporal dependency of speech via viterbi algorithm," Master's Thesis, *Korea Advanced Institute of Science and Technology,* 2005.

[13] L. Tao, C. Zhang, and O. Mitsunori, "A comparative study of feature selection and multiclass classfication methods for tissue classification based on gene expression," *Bioinformatics*, vol. 20, issue 15, 2004, pp. 2429-2437.

[14] C. Hsu and C. Lin, "A Comparison of Methods for Multiclass Support Vector Machines," *Trans. of the IEEE, on Neural Networks*, vol. 13, no. 2, Mar. 2002, pp. 415-425.

저자 소개

정유정(Yu-Jeong Jeong)

1992년 조선대학교 전산학과 졸업(이학사)
1997년 조선대학교 대학원 전산통계학과 졸업(이학석사)
2010년 조선대학교 대학원 전산통계학과 졸업(이학박사)
2018년 ~현재 조선대학교 SW융합교육원 객원초빙교수
※ 관심분야:빅데이터, 데이터마이닝, 인공지능, 영상 처리

최광미(Kwang-Mi Choi)

1990년 조선대학교 전자계산학과 졸업(이학사)
1995년 조선대학교 대학원 전산통계학과 졸업(이학석사)
2003년 조선대학교 대학원 전산통계학과 졸업(이학박사)
2018년 ~현재 조선대학교 SW융합교육원 객원초빙교수
※ 관심분야 : 데이터마이닝, 데이터모바일 서비스, 데이타베이스시스템 응용연구