Journal of The Korea Society of Computer and Information
Vol. 23 No. 12, pp. 43-48, December 2018
www.ksci.re.kr
https://doi.org/10.9708/jksci.2018.23.12.043

A Implementation of Optimal Multiple Classification System using Data Mining for Genome Analysis

Yu-Jeong Jeong*, Gwang-Mi Choi*

Abstract

In this paper, more efficient classification result could be obtained by applying the combination of the Hidden Markov Model and SVM Model to HMSV algorithm gene expression data which simulated the stochastic flow of gene data and clustering it. In this paper, we verified the HMSV algorithm that combines independently learned algorithms. To prove that this paper is superior to other papers, we tested the sensitivity and specificity of the most commonly used classification criteria. As a result, the K-means is 71% and the SOM is 68%. The proposed HMSV algorithm is 85%. These results are stable and high. It can be seen that this is better classified than using a general classification algorithm. The algorithm proposed in this paper is a stochastic modeling of the generation process of the characteristics included in the signal, and a good recognition rate can be obtained with a small amount of calculation, so it will be useful to study the relationship with diseases by showing fast and effective performance improvement with an algorithm that clusters nodes by simulating the stochastic flow of Gene Data through data mining of BigData.

▸Keyword: BigData, Data Mining Gene Data, Classification System

I. Introduction

최근 생명공학의 급속한 발전으로 인해 대규모 바이오 데이터가 생성됨에 따라 이를 분석하는 여러 방법들이 연구되고 있다. 인간의 유전체 염기 서열이 해독됨으로써 인간 생명의 실체를 이해하는 기초가 마련되었으며, 더불어 암과 같은 질병의 원인 분석 및 진단 방법과 치료제 개발을 위한 새로운토대가 마련되었다고 할 수 있다.

그러나 인간의 복잡한 생명 현상을 규명하기 위해서는 유전체의 서열 정보만으로는 부족하며, 단백질간의 상호작용 및 유전자 발현 여부등 유전체의 기능을 밝히는 기능 유전체학(functional genomics)에 관한 연구가 필요하다.

지금까지 연구되고 있는 마이크로어레이 데이터 분석은 크게 네 가지로 나눌 수 있다. 특정 유전자 기능을 밝히기 위한 유전자 식별(gene identification) 방법[1], 그리고 동일하거나 유사한 기능을 하는 유전자들을 찾아내기 위한 클러스터링 방법(Clustering)[2], 세 번째로 질병의 유무 혹은 질병의 형태를 판단하기 위한 분류(Classification)방법, 마지막으로 유전자끼리 혹은 유전자 집단 간의 인과관계를 밝히기 위한 유전자 조절 네트워크(gene regulatory network)방법[3]이다. 특징 선택과정에서 선택된 유전자들은 분류자를 생성하기 위해 입력값이 되는데, 이때 선택되는 특징의 종류와 개수는 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. 이러한 상황에서 한 가지 특징 선택 방법이나 분류자만을 정해서 사용한다면 그것이 항상 좋은 결과를 도출한다고는 기대할 수 없다. 따라서 유전자의 효과적인 분석 결과를 얻기 위한 다양한 분류 방법이 연구되어지고 있다.

이러한 바이오 데이터 분석 중 무감독 학습 기반의 클러스터

• First Author: Yu-Jeong Jeong, Corresponding Author: Gwang-Mi Choi
*Yu-Jeong Jeong (narimono@hanmail.net), Visiting Professor, IT Convergence University, Chosun University
*Gwang-Mi Choi (iplab@nate.com), Visiting Professor, IT Convergence University, Chosun University
• Received: 2018. 09. 07, Revised: 2018. 11. 30, Accepted: 2018. 12. 12.

링 알고리즘은 같은 군집 내에 속한 표본들끼리는 유사성이 높고, 서로 다른 군집 간에 속하는 표본들끼리는 유사성을 작게 한다. 마이크로 어레이 실험 데이터에 대한 클러스터링 알고리즘 개발은 유전자의 기능 분석을 통해 유전자의 상호 관련성 분석 등의 중요 연구 분야에 크게 기여 할 수 있다[4]. 현재의 마이크로어레이 기술을 이용해서 효과적으로 암을 정확하게 분류하기 위해서는 특정 암의 분류와 밀접하게 관련이 있는 유용한 유전자를 선택하는 과정이 필수적이다. 이러한 유용한 유전자 선택 작업은 분류 성능을 높이는 데에도 기여하지만, 마이크로어레이 데이터를 이용한 암 분류에 있어서, 별도로 분리된 정보력이 있는 유전자들은 암의 분류와 예측을 통한 진단 분야뿐만 아니라 정확한 분류 이후의 치료분야에도 매우 중요한 의미를 지닌다.

본 연구에서 은닉마코브모델과 SVM모델을 결합하여 유전자 데이터의 확률적 흐름을 시뮬레이션한 HMSV 알고리즘 유전자 발현 데이터에 적용하여 클러스터링 함으로써 더 효율적인 분류 결과를 얻을수 있도록 제시해보고 HMSV 알고리즘의 성능을 향상하기 위해 은닉마코브알고리즘의 likelihood를 최대점을 사용하여 매개변수 학습 효과를 통해 암 분류의 특징을 분류하고 Support vector machine을 통해 암 유전자 분류를 사용하여 분석함으로써 성능을 높혀보고자 한다.

II. Preliminaries

1. Related works

1.1 SOM(Self-Organizing Maps)

신경회로망(Neuralnetwork) SOM(Self-Organizing Maps) 알고리즘은 클러스터의 개수가 알려져 있을 때 주어진 다차원 데이터들을 가장 근접한 클러스터에 사상(mapping)시켜주는 방법이다[5].

Elastic network를 구성하는 map에 임의로 선택한 원소를 입력하여, 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. α는 학습율로서 0과 1의 값을 가지게 되며 가중치 수정시에 승자 뉴런과 함께 재조정된다[6].

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)(x_i(t) - w_{ij}(t)) \qquad (1)$$

가중치 벡터 갱신 함수에 따라 각 출력 노드의 가중치 벡터는 그 출력 노드에 포함된(그 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 데이터 클러스터링에 알맞으면서 결과의 가시화가 쉽고, 클러스터링 결과를 사용자가 제어할 수 있다는 장점을 가지고 있다[7].

1.2 Support Vector Machine (SVM)

SVM에서 선형으로 나눌 수 없는 경우에는 커널함수(kernel function) Φ를 이용하여 고차원에 전사시키는 방법으로 비선형적인 초평면을 형성시킨다. 비선형의 경우 선형에서와 같은 계산을 하기 위해서 커널함수 내적 $\langle \Phi(x_i), \Phi(x_j) \rangle$의 계산이 필요하다. 커널함수의 내적계산을 $\langle \Phi(x), \Phi(x_i) \rangle = K(x, x_i)$라 한다면, 비선형에서의 결정함수는 식(2)와 같다.

$$\phi(x) - sign(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b) \qquad (2)$$

공간에서의 선형 분리 초평면은 원래의 공간에서 비선형 분리 함수로 얻을 수 있다. 하지만 이러한 확장은 차원이 증가함에 따라 급격하게 증가되고 많은 계산량이 요구되는 단점을 가지고 있다.

비선형 분리 함수로는 주로 식(3)과 같은 3가지 커널 함수를 이용한다.

$$\text{Polynomial} : K(x, x_i) = (1 + \langle x, x_i \rangle)^d$$
$$\text{Gaussian RBF} : K(x, x_i) = \exp(- \| x - x_i \|^2 / 2\sigma^2)$$
$$\text{Neural network} : K(x, x_i) = \tanh(\langle x, x_i \rangle + \Theta) \qquad (3)$$

커널함수의 선택에 따라 초평면의 형태가 다양하게 바뀌므로 적절한 커널함수의 선택이 이루어져야 한다. 커널함수의 선택은 아직까지 그 선택 기준이 없는 단점을 가지고 있어서 데이터에 따라 그 결과가 다르게 나타난다. 본 논문에서는 마이크로어레이 데이터일 경우 가장 많이 사용하는 Gaussian RBF 커널을 이용할 것이다[8-10].

1.3 The optimal state sequence: A viterbi algorithm

주어진 관측열에 대응하는 최적의 상태열을 찾는 방법에는 동적 프로그래밍(Dynamic programming)기법 중의 하나인 비터비(Viterbi)알고리즘[11]을 적용한다. 관찰된 열 $O = (O_1, O_2, \ldots O_t)$이 주어졌을 때, 하나의 가장 좋은 상태열 $q = (q_1, q_{2,\ldots} q_T)$을 찾기 위해서는 식(4)과 같이 정의한다.

식(4)에서 $\delta_t(i)$ 는 시간 t에서 하나의 경로를 따르는 가장 큰 확률을 뜻한다[11].

$$\delta_t(i) = \max_{q_1, q_2, \ldots\ldots q_{t-1}} P[q_1, q_2, \ldots, q_{t-1}, q_t = O_1, O_2, \ldots O_t | \lambda] \qquad (4)$$

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] b_j(O_{1+1}) \qquad (5)$$

귀납에 의하여 식(4)는 식(5)와 같이 확장이 가능하며, 이를 이용하면 시간 t 뿐만 아니라 $t+1$에 대해서도 최대 확률을 가지는 상태들의 순서를 구할 수 있다[12].

III. The Proposed Scheme

본 논문에서는 공개적으로 오픈된 마이크로 어레이 실험데이터 유전자 발현 정보들 중에서 췌장암 데이터를 사용하였다. 현대 의학의 발달에도 불구하고 췌장암은 치료를 위해 분자 유전학, 수술적 치료 및 환자 관리 등의 다양한 영역에서 상당한 발전이 있었음에도 불구하고, 췌장암은 여전히 조기진단을 위한 생물학적 표지자나 영상의학적 검사가 없어 조기진단이 어려운 상황이다[13-14]. 본 논문에서는 별도로 분리된 정보력이 있는 유전자들은 암의 분류와 예측을 통한 진단 분야뿐만 아니라 정확한 분류 이후의 치료분야에도 긍정적인 동기를 부여할 수 있다는 점에서 매우 중요한 의미를 지닐수 있도록 하였다. 실험 데이터는 9개의 샘플을 사용하였으며 HMSV 알고리즘을 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 Hang[15]의 실험을 통해 결정된 215개의 유전자를 선택했다. 뽑힌 유전자만을 가진 데이터로 5개의 샘플 훈련 데이터와 4개의 훈련 테스트 데이터를 적용시켰다. 유전자 발현 데이터를 사용 할 수 있는 기존의 데이터마이닝 Clustering 도구를 사용하여 SOM알고리즘과 K-means알고리즘은 실험을 통하여 만들어진 노드들을 관찰해 보았다.

본 논문에서는 비교 실험 대상인 K-means알고리즘 실험 조건은 K-means반복을 위한 매개변수 중 실행 수는 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 SVM계산을 위해 필요한 커널 함수는 RBF커널함수를 사용했으며, SVM모델의 복잡성과 평활도에 대한 정도를 서로 보정해주는 C값과 커널 함수를 사용할 때 필요한 모수의 값은 한 번의 훈련 데이터를 가지고 Hsu and Lin[16]에 따라 결정한 후 그 값을 사용하였다.

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

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. HMSV 알고리즘을 설명하기 위해 데이터내의 data work의 확률 값으로 마코브 행렬 값을 사용하는데, 이 행렬은 각 열의 확률 합이 1을 넘지 않는다.

1단계에서는 모델 형성과정 $P(O|\lambda)$를 최대로 하는 모델 파라미터 $\lambda=(A,B,\Pi)$를 구하는 문제이고 2단계에서는 모델 인식과정에서 관측된 심벌의 시퀀스 $O=O_1O_{2\ldots}O_T$와 모델 $\lambda=(A,B,\Pi)$가 주어졌을 때 likelihood $P(O|\lambda)$를 구하는 문제이다. 1단계에서 얻어진 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 커널 함수는 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.

다음 그림 1은 본 논문에서 제안한 HMSV 알고리즘 실험 결과이다.

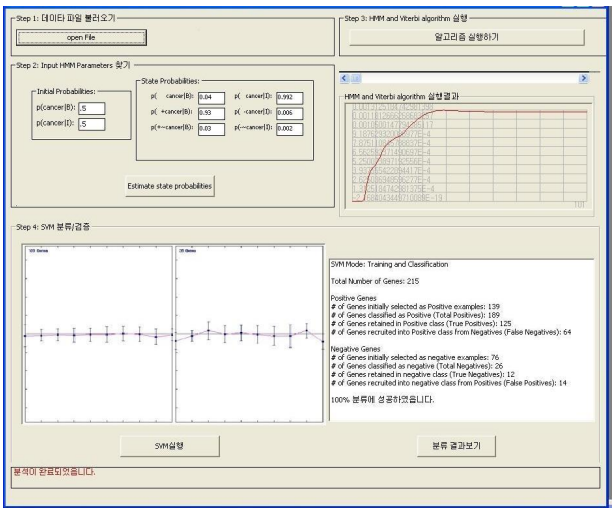

Fig. 1. The execution results of the HMSV algorithm proposed in this paper

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 0.93, P(+~cancer|B)는 0.03, P(cancer|I)는 0.992, ,P(-cancer|I)는 0.006, P(-~cancer|I)는 0.002의 파라미터 결과를 얻을 수 있었고, Step3 단계에서는 Viterbi 알고리즘으로 최대 확률 값들의 누적치인 최적의 상태열을 추정하면 안정상태로 들어오는 값인 그림 1의 Step3의 결과 0.00118의 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.

실험 데이터 215개의 유전자 실험 결과 Positive Genes의 분류 중 실험 결과는 Positive Genes중 유전자 초기 Positive example로 선정되어 실험된 수는 139개이며, Positive에 따르는 유전자 분류 즉 Total Positive는 189개, Positive class에서 유전자를 포함하는 true Positive는 125개, Negative 유전자에서 긍정적인 수준을 포함하는 유전자 즉 False Negative의 수는 64개로 분류되었다.

Negative Gene 그룹 중에서는 유전자 초기에 Negative example로 선정된 샘플의 개수는 76개이며, 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.

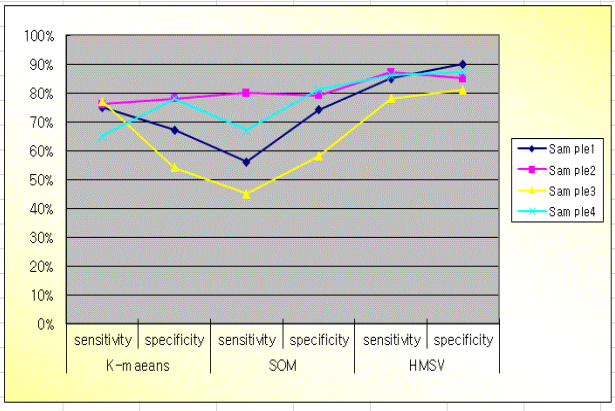

Fig. 2. Sensitivity and specification of K-means, SOM, HMSV

그림 2는 본 논문이 다른 논문보다 우수하다는 것을 증명하기위해 분류 기준으로 많이 사용하는 민감도와 특이도를 테스트한 결과 본 논문에서 제안한 HMSV 알고리즘을 그래프로 나타낸 것이다. 이는 K-means나 SOM보다 더 잘 분류 되는 것을 실험을 통하여 확인하였다.

IV. Conclusions

본 논문에서는 방대한 빅 데이터를 데이터마이닝 작업을 통해 유니크한 클러스터링 분석에서 나아가 게놈 질병을 조기진단할수 있는 게놈 정보를 활용한 기능적인 방향을 제시 할 수 있다. 본 논문에서는 독립적으로 학습된 알고리즘을 결합한 HMSV 알고리즘에 대해 실험하였다. sample1의 경우 sensitivity 85%, specification 90%, sample2의 경우 sensitivity 87%, specification 85%, sample3의 경우 sensitivity 78%, specification 81%, sample4의 경우 sensitivity 86%, specification 87%로 안정성 있게 높게 나왔다는 것을 알수 있었다. 이는 K-means나 SOM보다 더 잘 분류 되는 것을 알 수 있었다. 하지만 sample3의 경우 다른 샘플의 경우와 차이가 나는 부분은 실험 유전자 중 나이등 여러 가지 환경적인 차이가 나는 유전자에 대한 영향이 있었던 것으로 보인다.

본 논문의 연구 결과를 바탕으로 지속적인 연구를 해간다면 생명과 관련있는 부분이기 때문에 암 연구에서 정확한 진단을 해야하는 중요성은 커지고 있는데, 이에 대응하는 암 세포주의 특성을 해석 분류하여, 실제 임상 치료 전에 좋은 결과를 얻을 수 있는 테스트 모델로 적용하는데 HMSV 알고리즘이 큰 역할을 할 것이다.

이러한 알고리즘을 바탕으로 실제로 암세포의 분류와 관련된 새로운 유전자 탐색 및 메커니즘을 규명하는데 유용하게 사용할 수 있을 것으로 사료된다. 이러한 결과는 암세포 내에 존재하는 유전자 발현의 특징을 분류하고 조절함으로써 암 특징점을 찾아내 더욱더 중요한 암지표를 만들어 다양한 암의 분류 진단 및 메커니즘을 만들어 활용될 수 있을 것이라고 사료된다. 또한 방대한 양의 데이터를 분류하고 자유롭게 그 내용을 탐색할 수 있게 데이터의 내용을 탐색할 수 있는 빅데이터의 알고리즘을 추가적으로 설계하고 분석하여 활용할 수 있는 결과를 제공할 수 있는 부분으로 넓혀져 간다면, 암 환자의 데이터를 좀더 정확하게 예측하여 Electronic Health Records 의료 데이터를 학습하여 환자 상태를 예측하는 바이오 산업이나 헬스케어 분야의 많은 부분에서 큰 도움이 되는 역할을 할 것이다.

향후 연구 과제로는 다양하고 보다 체계적으로 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 방법을 추가하여 더 많은 연구를 진행하고자 한다. 하지만 본 연구의 한계점으로는 공개적으로 오픈된 아미크로어레이 데

이터이기 때문에 더 많은 정확한 데이터 자료를 취득을 하여야 하며 ,데이터가 연구자마다 다른 부분들이 많아서 이를 표준화 하고 통합하여 암세포 지표에 따른 다른 정규화방법을 찾아 질병치료에 유의한 유전자 선택 방법을 추가하여 더 많은 연구를 하여야 하는 어려움이 있으며, 또한 암 질병 빅 데이터를 만들면서 데이터 마이닝을 이용 하려면, 현대 사회에 필수적인 개인정보 보호와 관리가 필요한 부분이라 빅 데이터를 얻어 연구를 진행하고자 하는 부분에 많은 어려움이 있다.

REFERENCES

[1] G.K., Yang, Y. H., T. p. Speed, " Statistical issues in microarray data analysis," Functional Genomics, Methods and Protocols, 24 ,111-136, 2003

[2] Y.Chen, E. R. Dougherty and M. L., " Bittner, Ratio-Based Decision and the Quantitative Analysis of cCNA Microarray Images," Journal of Biomedical Optics 2 no.4,364-374, 1997

[3] Y. H. Yang, S. Dudiot, P. Luu, D. M. Lin, V. Peng, J. Nagi and T.P. Speed, "Normalization for cDNA Microarray data : a robust composite method addressing single and multiple slide systematic variation," Nucleic Acids Research no,2002.

[4] Pierre Baldlnd G. Wesley Hatfield, "*DNA Microarrays and gene expression* "(n.p.: Cambridge University Press, 2002)

[5] T. Kohonen, elf-Organizing Map (n.p.: Springer, 1997)

[6] Kim sul Lam, "Analysis of Influencing Factors of Medical Expenditure on Elderly Hypertension Outpatients - Focused on Region and Medical Use," (Master of Engineering Thesis, Chungbuk National University Graduate School, 8-9,2018.

[7] E. Berglund ; J. Sitte, "The parameterless self-organizing map algorithm," IEEE Transactions in Neural Networks 17 no.2 ,305-316,2006

[8] Smyth, G.K., Yang, Y.H., Speed, T.P, "Staticstics issues in microarray data analysis. Function Genomics," Methods and protocols 24,111-136, 2003

[9] Yang, Y.H., Dudoit, s., Luu,P., Lin, D.M., Peng, V., Nagi, J., Speed, T.P.(2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiplr slide systematic varation. Nucleic Acids Res 30.

[10] SukBuk Kang, oungMin Kim, JinKap Choi, BongSeon Kim, WonSub Yang. "Application Statistics." n.p.: Kyeongmunsa, 1993.

[11] Han hakyoung, "*Introduction to pattern recognition*"(n.p.: hanbit media, 2011)

[12] Cho sunho, "Segmented viterbi algorithm for speech recognition,",Master's Thesis, Korea University, n.d, 8-9.

[13] National Cancer Information Center. *http://www.cancer.go.kr*, 2017

[14] Lee Ji Sun, "Explanatory model on quality of life in patients with pancreatic cancer", doctor, Yonsei University Graduate School, 1-2, 2018

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

[16] Hsu, Chih_Wei and Chih-Hen Lin, "comparison of methods for multi-class support vector machines," IEEE Transactions in Neural Networks 13,415-425,2002

Authors

Yu-Jeong Jeong received the B.S., M.S. and Ph.D. 1993,1997 and 2010 Graduated from Department of Computational Statistics, Chosun University. She is currently a Professor in the Visiting Professor, IT Convergence University. She was a software engineer in Ssangyoung software Corportation. She is interested data minining, information retrieval, and database system and image processing.

Kwang-Mi Choi graduated from Chosun University in 1990 with a B.S in computer science, 1995 with a master's degree in computational statistics, and a Ph.D.2003. She is currently a Professor in the Visiting Professor, IT Convergence University. She is interested data minining,mobile internet service, and database system and data-warehouse design and application