

비지도학습 머신러닝에 기반한 베타파 상관관계 분석모델

최성자

가천대학교 소프트웨어 중심대학 초빙교수

Beta-wave Correlation Analysis Model based on Unsupervised Machine Learning

Sung-Ja Choi

Department of software, College of IT, Gachon University

요 약 뇌파 파형중 베타파를 이용한 인간의 인지상태를 판별한다. 베타파는 인간의 인지상태중 스트레스 영역에 해당하는 특성이 있고, 이 영역에서 스트레스의 오버대역폭을 추출하기 위해서 저대역폭과 고대역폭 사이의 베타파간 상관관계를 분석해야 한다. 그러므로 본 논문에서는 효과적으로 베타파 상관관계를 분석하고 추출하기 위해 비지도학습 머신러닝을 이용한 Kmean 클러스터링 분석모델을 제시한다. 제시된 모델은 베타파 영역을 유사한 영역의 클러스터 군으로 분류하고 해당 클러스터링 범주에서 이상파형을 판별한다. 이상파형 판별을 위해 클러스터군의 밀집도와 정상범주 이탈영역을 기준으로 스트레스 위험군을 판별하고 판별된 스트레스 위험군에 대한 대처방안을 제공할 수 있다. 제시된 모델을 활용하면 뇌파파형을 통한 인지상태의 스트레스 지수분별이 가능하고, 개인의 인지상태에 대한 관리 및 응용이 가능하다. 또한 스트레스와 오피스증후군을 갖는 사람들에게 뇌파관리를 통해 개인의 삶에 대한 질적 향상에 도움을 준다.

주제어 : 뇌파 파형, 베타파, 머신러닝, 비지도학습, KMeans

Abstract The characteristic of the beta wave among the EEG waves corresponds to the stress area of human perception. The over-bandwidth of the stress is extracted by analyzing the beta-wave correlation between the low-bandwidth and high-bandwidth. We present a KMeans clustering analysis model for unsupervised machine learning to construct an analytical model for analyzing and extracting the beta-wave correlation. The proposed model classifies the beta wave region into clusters of similar regions and identifies anomalous waveforms in the corresponding clustering category. The abnormal group of waveform clusters and the normal category leaving region are discriminated from the stress risk group. Using this model, it is possible to discriminate the degree of stress of the cognitive state through the EEG waveform, and it is possible to manage and apply the cognitive state of the individual.

Key Words : Machine Learning, Unsupervised Learning, KMeans, SPARK, EEG

1. 서론

정보통신 기술(ICT)의 융합으로 뇌공학 발달은 비약적인 발전을 하고 있으며, 대표적인 뇌공학 분야로써 뇌파 활용을 들 수 있다. 뇌파를 활용하는 대표적인 분야로

는 BCI(Brain Computer Interface)시스템이 있다. 이 시스템은 장애인을 보조하기 위한 디바이스가 제공되며, 수면장애나 신경질환등 정신상태 감지가 가능하고, 동작과 뇌파 상호간의 연관관계를 실시간으로 추적이 가능하다. 또한 인간과 컴퓨터간의 인터페이스 활용을 위해 뇌

*Corresponding Author : Sung-Ja Choi(irecomm@naver.com)

Received January 7, 2019

Accepted March 20, 2019

Revised February 7, 2019

Published March 28, 2019

파가 활발히 활용되고 있다[1].

뇌파를 이용하여 인간의 인지 상태를 분별하기 위해서는 뇌파 대역폭을 구분하여 다양한 뇌파 파형(δ , θ , α , β , τ etc.)으로 나누어서 많은 연구에 활용되고 있다. 뇌파 대역폭 응용에 대한 연구 예를 살펴보면 다음과 같다. 인지 및 감정 인식을 위해 뇌파 신호로부터 얻을 수 있는 다양한 특징의 적합성을 조사하고 알파파(α -wave)와 베타파(β -wave)를 활용하여 EEG 파워 스펙트럼을 비모수적 방법을 사용하여 인간 감정의 지표로써 뇌파(Brainwave)를 사용한다[2]. 또한, 알파 및 베타 파에 집중시 측정된 체질량 지수 (BMI)와 뇌파 활동 간의 상관 관계에 초점을 맞추어 이상적인 체중 참가자는 다른 참가자보다 스트레스 지수가 적다는 결론을 도출하는데 뇌파 파형을 적용한다[3]. 한편, 인지 및 감정 중 스트레스 인지 상태에 대한 판별을 위해 다양한 연구가 진행되고 있다. Mediation의 뇌파 파형의 값을 기반으로 스트레스와 침착한 상태로 구분하고[4], 알파파와 베타파를 활용하여 에너지 스펙트럼(ESD: Energy Spectral Density)과 스펙트럼 중심(SC: Spectral Centroids)을 계산하고 k-Nearest Neighbor (k-NN) 알고리즘을 사용하여 정량화 및 스트레스 지수를 분류한다[5]. 뇌파 신호, 청력 변동 및 손 움직임을 분석하여 스트레스 수준을 나타내며 오피스 증후군을 지능형 시계로 감지한다[6].

뇌파를 활용한 인지상태 분별시 상당수의 연구에서 KMeans 알고리즘을 진행하고 있다. KMeans는 일반적으로 사용되고 있는 가장 좋은 클러스터링 기법 중 하나로써 지역 최소치에 수렴한다. KMeans 알고리즘을 활용한 데이터 클러스터링과 신경망을 이용함으로써 뇌파 기반의 불편감 추정기법을 제시하고[7], 인지 활동 테스트와 수학적 활동에 대한 테스트를 기반으로한 클러스터링에 KMeans를 활용한다[8]. 기존의 연구에서는 스트레스 지수판별을 위해 사용되는 방법이 복잡하고 뇌파 기기의 다른 외부기기가 요구되는 단점이 있다. 그래서, 본 논문에서는 스트레스에 민감한 저수준 및 고수준 대역폭의 베타파 상관관계를 활용하여 스트레스 지수를 추출하고, 베타파 상관관계의 분석을 위해 KMeans 비학습지도 머신러닝 모델을 제안한다. 제안 모델을 통하여, 개인의 인지상태에 대한 효과적인 관리 및 응용이 가능하고, 스트레스와 오피스증후군을 갖는 사람들에게는 뇌파 관리를 통해 개인의 삶에 대한 질을 향상시킬 수 있다.

본 논문의 구성은 다음과 같다. 1장에서는 연구의 배

경과 필요성에 대해 기술하고, 2장에서는 관련 연구에 대하여 살펴본다. 3장에서는 베타파 상관관계 분석을 위한 모델을 제시하며 4장에서는 실험 및 분석 결과를 통해 제안한 모델의 타당성을 입증하고 5장에서는 결론을 맺는다.

2. 관련연구

제안된 모델에 사용된 뇌파의 특징과 구분에 대하여 살펴보고, 스트레스 지수 추출을 위한 비지도학습 머신러닝인 KMeans 알고리즘의 이론적 배경에 대하여 기술한다.

2.1 뇌파의 특징과 구분

뇌파 (EEG) 신호는 뇌의 전기적 활동을 표현하고 기록하는 데 사용되며, Table 1에서와 같이 대역폭에 따라 다양한 피처를 제공한다.

Table 1. Frequency bands of brainwave

Band	Type	Feature
0.1~3Hz	δ	Deep sleep
4~7Hz	θ	Sleep state
8~12Hz	α	Relaxation
13~15Hz	SMR	Caution
16~20Hz	Mid- β	Concentration
21~30Hz	β	Tension, Stress
30~50Hz	τ	Anxiety

델타파(δ -wave)는 0.1Hz에서 3Hz의 대역폭을 가지며 깊은 수면상태에서 검출되고, 세타파(θ -wave)는 4Hz에서 7Hz의 대역폭을 가지며 얇은 수면상태에서 검출된다. 알파파(α -wave, SMR)는 8Hz에서 15Hz사이에서 추출되며 릴렉스와 주의상태의 특성을 가지고 있다. 베타파(Mid- β , β -wave)는 6Hz에서 30Hz에서 추출되고 집중 및 스트레스 특성을 가진다. 감마파(τ -wave)는 30Hz에서 50Hz에서 추출되며, 감정중 걱정상태의 피처(Features) 특성을 가진다.

2.2 비지도학습 머신러닝

KMeans는 가장 대표적인 비지도학습 머신러닝 알고리즘으로써, k개의 클러스터로 묶어서 각 클러스터의 중심(centroid)와 클러스터 내 오브젝트와 거리의 차이에 대한 분산을 최소화한다[9].

$D = X_1, X_2, \dots, X_i, \dots, X_m$ a data set of m records

$X_i = (X_{i1}, X_{i2}, \dots, X_{in})$:

a each record is an n -dimensional vector

$$C_j = \text{Cluster}(X_i) = \arg\min \|X_i - \mu_j\|^2 \quad (\text{eq. 1})$$

$$\text{Distortion} = \sum_{i=1}^m (x_i - c_i)^2 = \sum_{j=1}^k \sum_{i \in \text{OwndBy}(u_j)} (X_i - \mu_j)^2 \quad (\text{eq. 2})$$

KMeans 클러스터링 처리과정을 Fig. 1에서 보여준다. 클러스터링 K 수를 입력하고 중심점을 선택한 후, 거리를 계산한다. 최소 거리인지 체크한 후 최소거리가 아니면 인접 클러스터군으로 이동한다. 최소거리를 찾기 위해 반복 수행한다.

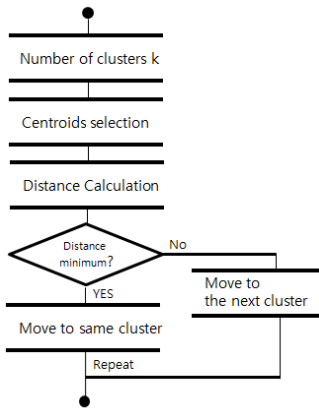


Fig. 1. Processing for KMeans clustering

2.3 상관관계 분석

인공지능 분야에서는 두가지 개체의 속성값들이 여러 개 일 경우 이들 속성 값들에 의한 두 개체 사이의 유사도를 구할 때 자주 사용한다. 또한, 선형 관계의 두변수의 강도와 방향을 측정 하기위해 상관 계수를 사용한다. 두 개의 수치 변수 X와 Y 사이의 선형 관계의 강도와 방향을 측정하며, 데이터 샘플에 대한 상관 계수는 r로 표시한다. 상관관계 분석을 위해 유클리드 유사도 상관관계를 적용한다.

$$Ed = \text{SQRT}(paw(x1-x2) + paw(y1-y2))$$

$$r = 1 + 1/Ed \quad (0 \sim 1) \quad (\text{eq. 3})$$

3. 베타파 상관관계 분석 모델

3.1 제안 모델

다양한 플랫폼의 뇌파분석 서버 시스템이 개발되었으며[10-12], 정보통신 융합으로 4차 산업혁명으로 패러다임이 변화되면서 뇌파분석 서버시스템 플랫폼의 확장이 요구된다. 따라서 본 논문에서는 4차산업혁명의 패러다임에 맞는 뇌파분석 모델을 제안한다. 이 모델은 빅데이터 기반의 분산처리가 가능하고 데이터 셋 간의 상관관계를 통해 데이터 분석을 한다. Fig. 2는 베타파 상관관계 분석을 위한 비학습지도 머신러닝의 KMeans 모델을 보여준다.

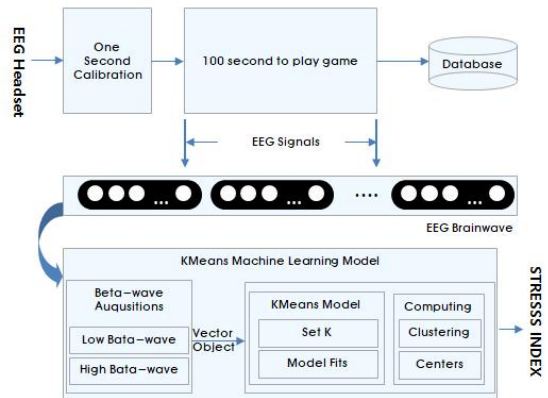


Fig. 2. KMean model for beta-wave correlation

제안 모델의 기능별 특성은 다음과 같다. (1)뇌파 신호 수신모듈(Brainwave Aquisition)은 1초 단위로 뉴로스카이 헤드셋의 블루투스 통신을 통하여 뇌파분석시스템으로 전송된다. 해당 데이터는 데이터베이스에 저장된다. 데이터베이스에 저장된 뇌파 데이터 셋을 (2)베타파 획득모듈(Beta-wave Aquisition)을 통해 sparkSQL 쿼리를 실행하여 저수준 대역의 베타파와 고수준 베타파를 획득한다. 해당 데이터 셋은 데이터프레임유형으로 되어 있기 때문에 벡터객체화 과정을 (3)KMeans 모델 구성 모듈에서 수행한다. Feature를 지정하고 클러스터링을 위한 데이터 셋으로 변환시킨다. 또한 클러스터링 군집 개수 변수 k 를 지정한 후 KMeans 모델을 최적화 한다. 구축된 모델을 (4)Cmputing 모듈에서 각 클러스터링 군집에 대해 클러스트화 하고 최단 거리의 평균값을 각 클러스터링 군집마다 추출한다. (5)스트레스 위험군 판별모듈에서 결과값은 스트레스 지수의 판별을 위해 상관관계와 함께 지표로 쓰인다. 해당 모델은 인지 및 임상 신경과학에서 뇌파 피쳐 특성을 연구하기위한 객관적·경험적인 중요한 도구가 될 것이다.

3.2 모델 구현 알고리즘

KMeans의 수행 과정에 베타파 상관관계 분석을 위한 수행과정 부분이 추가된다. Table 2에서와 같이 상세 모델 구현 알고리즘은 다음과 같다. 입력값으로 클러스터 개수를 지정하고, 데이터 프레임을 입력으로 처리하고 출력은 평균값을 획득한다. 수식(1)에서 제시된 연산을 통해 수행한다. 모델 구축을 위해 스파크 세션을 생성한다. 스파크 세션은 데이터 읽기, 메타데이터 활용, 세션 세팅, 그리고 클러스터 리소스에 접근하기 위한 시작점이다. 스파크 세션을 통해 sqlContext를 생성하여 데이터 베이스의 데이터 셋을 접근한다. 스파크 쿼리를 통해 데이터 프레임 형태의 데이터 객체를 추출한다.

Table 2. Algorithm for beta wave analyze model

Input k: Cluster Number
Dataframe: N's Data Object(low-beta, high-beta)
Output value: Centroid value
Algorithm
Sesstion, sqlContext create
Dataframe acquisitions
Vector Object create
Initial value: D= {d1, d2, ..., dn} , Centroid of each cluster
Repeat:
Center value=Distance(di, centroid) Recalculate centroid
Until no change
Calculate: (stress range ^ correlation coefficient)
Decide beta-wave outlier

4. 실험 및 성능평가

실험을 위해 분산처리 프레임워크 기반의 플랫폼을 먼저 구축하고 데이터가공을 위해 전처리과정에서 데이터셋을 획득한다. KMeans 모델을 적용하여 클러스터링 후 데이터분석을 통해 이상대역폭을 추출한다[13].

스파크는 오픈소스 프레임워크 중 가장 빠른 속도로 분산처리 프로덕트 개발이 가능하다[14]. 베타파 상관관계의 분산처리 및 분석을 위해서 스파크와 하둡 플랫폼 기반으로 구축된다[15]. 또한 데이터 프레임을 획득하기 위한 스파크 쿼리는 자바 JVM과 Py4J 사이의 커뮤니케이션을 통해 수행한다. 머신러닝 수행을 위해 Spark ML

Library가 추가되어 KMeans ML 모델을 빌드하고 클러스터링을 완성한다. Table 3는 제안 모델을 구축하기 위한 플랫폼 및 주요 도구들에 대한 기능들이다.

Table 3. Using tools for the platform of construction

Tools	Function
Spark ML & sklearn	머신러닝 라이브러리
Spark-2.3.1-bin-hadoop2.6	윈도우용 스파크 엔진
Hadoop-2.6.5	분산처리 엔진
Python 3.7	스크립트 언어
Java EE 8.0	자바개발도구(엔터프라이즈)
Maria DB	오픈소스용 데이터베이스
Neurosky headset	뇌파 검출 기구
Pyneuro	뇌파 연동 라이브러리

뇌파의 데이터 셋을 추출하기 위해 20대 후반의 남자 대학생이 주기적으로 선호하는 게임(가위바위보등)을 수행한다. 게임 수행중 획득한 10개 채널의 데이터 셋을 데이터베이스에 실시간 저장한다. 뉴로스카이 뇌파 데이터는 1초 단위로 전송되며, 실험을 위해 100초 단위로 묶어서 처리한다. Fig. 3에서는 스파크 프레임 워크에서 실행된 뇌파 분석 화면을 보여준다. 대역폭에 따라서 attention, meditation, delta, theta, alpha, beta, gamma파의 파형을 보여준다. 수신된 결과값은 주파수 대역폭의 값이 상대적 분포 값으로 변경되어서 수신된다. Fig. 4에서는 베타파 분산처리를 하기위해 획득한 데이터 셋의 분포를 보여주고 있다. 저대역폭 베타파는 0 ~60,000, 고대역폭 베타파는 0~40,000사이에 집중된 것을 확인할 수 있다. 범주를 벗어난 베타파는 스트레스 이상현상이 발생 가능한 시간대의 파형으로 직관적으로 알 수 있다.

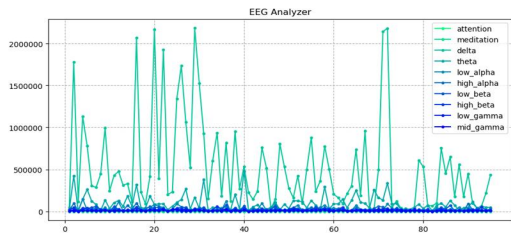


Fig. 3. Brainwave analyzer on spark

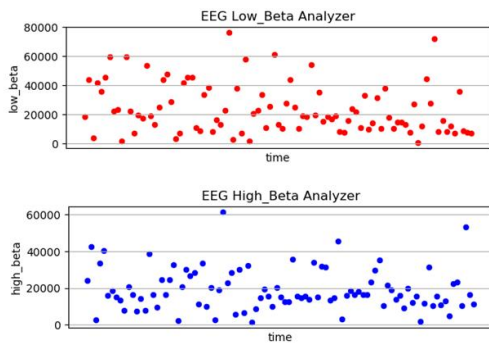


Fig. 4. Analyzer of beta-wave acquisitions dataset

직관에 의한 분석은 신뢰성이 떨어지므로 제안된 모델에 따라 이상현상 범위를 추출한다. KMeans 모델의 클러스터 k-값을 5로 지정하고 저수준 대역의 베타파와 고수준 대역의 베타파를 입력으로 데이터 프레임을 벡터화하여 피쳐값을 추출한다. 모델의 일관성을 유지하기 위해 SEED=1로 지정하여 수행한다. 모델의 클러스터링을 수행한 결과를 Fig. 5에서와 같이 보여준다. 베타파간의 유클리드 상관관계를 보여주며, 이상현상이 발생한 클러스터 군을 확인할 수 있다. 해당 결과를 스트레스 지수(삼분법: 상,중,하)로 변환함으로써 실시간 뇌파 관리가 가능하다. 위험 경고 발생 시(스트레스 지수=상) 릴렉스 가능한 행위를 하도록 한다.

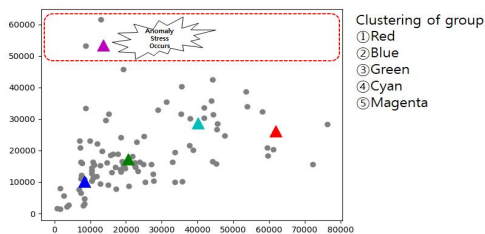


Fig. 5. KMeans model applied for beta-wave dataset

구축된 클러스터링 모델을 검증하기 위해 최소치 수렴한 K값을 찾아 검증한다. Fig. 6에서와 같이 X축의 값이 5에서 수렴하는 것을 확인할 수 있다. Elbow 방법을 통해 검증하며, 수식(2)에서 제시된 연산을 통해 수행한다. 클러스터의 수를 순차적으로 늘려가면서 결과를 모니터링 한다. 만약 하나의 클러스터를 추가했을 때 이전보다 훨씬 더 나은 결과를 나타내지 않는다면, 이전의 클러스터의 수를 구하고자 하는 클러스터의 수로 설정한다.

k-평균 알고리즘의 계산 복잡도에 크게 영향을 미치는 요소로 유클리드 공간 d와 클러스터의 수 k이다. n개의 d-차원 데이터 벡터들을 i번의 반복을 통해 k개의 클러스터로 묶는데 걸리는 복잡도는 $O(nkdi)$ 이다[16,17].

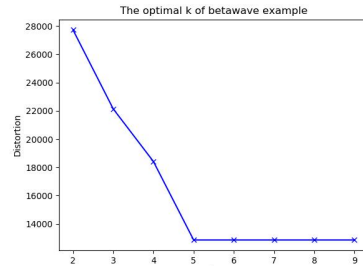


Fig. 6. Verification about KMeans model of optimized k-value in beta-wave dataset

5. 결론

뇌파를 활용한 인지상태 분별 시 기존의 연구에서는 스트레스 지수판별을 위해 사용되는 방법이 복잡하고 뇌파 기기와 다른 외부기기가 요구되는 문제점이 존재한다. 그래서 본 논문에서는 복잡한 과정의 스트레스 지수 판별을 개선하고, 외부 기기없이 인지·감정 상태를 분류할 수 있는 베타파 상관관계 분석모델을 제안했다. 제안한 모델은 피쳐 분류에서 스트레스 특성을 추출하고, 스트레스에 민감한 베타파를 기준으로 스트레스 지수를 추출하였다. 그리고 고수준 대역폭 베타파 값의 범주를 분할하여 이상현상 발생 대역폭에 대한 밀집도를 산출하여 값의 범주를 벗어나고 밀집도가 고수준인 경우 스트레스 이상현상이 발생할 확률이 높은 것으로 추정하였다. 스트레스 위험군에 대한 인지를 통해 개인의 뇌파 관리가 가능하며, 개인의 삶의 질을 향상 시킨다.

제시된 모델을 기반으로 비지도학습 머신러닝과 딥러닝 융합 기반의 뇌파분석 모델을 구현하는 연구와 이를 통해 다른 다양한 인지·감정 특성 추출을 통해 전체 뇌파 파형의 상관관계를 실시간으로 분석하는 연구가 진행되어야 한다.

REFERENCES

[1] E. A. Larsen. (2011). *Classification of EEG signals in a brain-computer interface system*, Master's thesis,

- Institutt for datateknikk og informasjonsvitenskap.
- [2] Y. P. Ma et al. (2014, August). Emotional identification during mobile RF Radiation. *In Control and System Graduate Research Colloquium (ICSGRC), 2014 IEEE 5th* (pp. 285-289). IEEE.
- [3] R. S. S. A. Kadir et al. (2009, March). Analysis of correlation between body mass index (BMI) and brain wave using EEG for Alpha and Beta frequency band. *In Signal Processing & Its Applications, 278-283*.
- [4] K. Crowley, A. Sliney, I. Pitt & D. Murphy. (2010, July). Evaluating a brain-computer interface to categorise human emotional response. *In Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on* (pp. 276-278).
- [5] N. Sulaiman et al. (2012, February). Development of EEG-based stress index. *In Biomedical Engineering (ICoBE), 2012 International Conference on* (pp. 461-466). IEEE.
- [6] P. Reanaree, P. Tananchana, W. Narongwongwathana & C. Pintavirooj. (2016, December). Stress and office-syndrome detection using EEG, HRV and hand movement. *In Biomedical Engineering International Conference (BMEiCON),* (pp. 1-4). IEEE.
- [7] T. Teramae, D. Kushida, F. Takemori & A. Kitamura. (2010, August). Estimation of Feeling Based on EEG by Using NN and k-means Algorithm for Massage System. *In SICE Annual Conference 2010, Proceedings of* (pp. 1542-1547). IEEE.
- [8] A. Azhari & L. Hernandez. (2016). Brainwaves feature classification by applying K-Means clustering using single-sensor EEG. *International Journal of Advances in Intelligent Informatics, 2(3), 167-173*.
- [9] J. Han, J. Pei & M. Kamber. (2011). *Data mining: concepts and techniques*. Elsevier.
- [10] S. J. Choi & B. G. Kang. (2014). Prototype design and implementation of an automatic control system based on a BCI. *Wireless personal communications, 79(4), 2551-2563*.
- [11] S. J. Choi, B. G. Kang & G. J. Kim. (2018). The Brainwave Analyzer of Server System Applied Security Functions. *Journal of Digital Convergence, 16(12), 343-349*.
- [12] L. Li & D. Yao. (2007). A new method of spatio-temporal topographic mapping by correlation coefficient of k-means cluster. *Brain topography, 19(4), 161-176*.
- [13] J. H. Yang, Y. S. Park & S. H. Lee. (2017). Text extraction in images using simplify color and edges pattern analysis. *Journal of the Korea Convergence Society, 8(8), 33-40*.
- [14] I. K. Lim, D. J. Park & H. J. Cho. (2018). Development of Procurement Announcement Analysis Support System. *Journal of the Korea Convergence Society, 9(8), 53-60*.
- [15] M. A. Alsheikh, D. Niyato, S. Lin, H. P. Tan & Z. Han. (2016). Mobile big data analytics using deep learning and apache spark. *IEEE network, 30(3), 22-29*.
- [16] D. Aloise, A. Deshpande, P. Hansen & P. Popat. (2009). NP-hardness of Euclidean sum-of-squares clustering. *Machine learning, 75(2), 245-248*.
- [17] S. Dasgupta & Y. Freund. (2009). Random projection trees for vector quantization. *IEEE Transactions on Information Theory, 55(7), 3229-3242*.

최 성 자(Choi, Sung Ja)

[정회원]



- 1991년 2월 : 한남대학교 컴퓨터 공학과(공학사)
- 1997년 2월 : 한남대학교 컴퓨터 공학과(공학석사)
- 2005년 8월 : 한남대학교 컴퓨터 공학과(공학박사)
- 2019년 2월 ~ 현재 : 가천대학교 소프트웨어교육센터
- 관심분야 : Bio 센서, 뇌공학, 머신러닝, 아키텍처
- E-Mail : irecomm@naver.com