

다중상황의 군집분석과 연관규칙을 이용한 지식추론 모델

신동훈¹, 김민정², 오상엽³, 정경용^{4*}

¹경기대학교 컴퓨터과학과 석사과정, ²경기대학교 컴퓨터과학과 학부생,

³가천대학교 컴퓨터공학과 교수, ⁴경기대학교 컴퓨터공학부 교수

Knowledge Reasoning Model using Association Rules and Clustering Analysis of Multi-Context

Dong-Hoon Shin¹, Min-Jeong Kim², SangYeob Oh³, Kyungyong Chung^{4*}

¹Master's Course, Department of Computer Science, Kyonggi University,

²Student, Department of Computer Science, Kyonggi University,

³Professor, Division of Computer Engineering, Gachon University,

⁴Professor, Division of Computer Science and Engineering, Kyonggi University

요 약 사람들은 바쁜 현대사회 속에서 시간적 제재를 받고 있다. 이에 따라 사람들은 건강에 나쁜 영향을 미치는 간편한 인스턴트 식품을 섭취하고 간단한 운동조차하기 어려운 상황에 놓여있다. 또한 불필요한 정보과부화 현상으로 인해 개인의 특성에 적합하고 정확한 추론을 하는 것에 대한 중요성이 커지고 있다. 따라서 본 논문에서는 다중상황의 군집분석과 연관규칙을 이용한 지식추론 모델을 제안한다. 제안하는 방법은 상황정보에 따른 군집을 기반으로 연관규칙을 생성함으로써 사용자들에게 개인화된 헬스케어 방법을 제공한다. 이를 통해 각 질병에 대한 위험도를 추론함으로써 해당 질병에 대한 발병률을 낮출 수 있다. 또한 성능 평가를 통해 제안하는 모델이 비교 모델보다 수치상으로 F-measure 값이 0.027 더 높게 나타나며, 비교 모델 보다 우수하게 평가된다.

주제어 : 다중상황, 군집분석, 연관규칙, K-평균 알고리즘, 추론

Abstract People are subject to time sanctions in a busy modern society. Therefore, people find it difficult to eat simple junk food and even exercise, which is bad for their health. As a result, the incidence of chronic diseases is increasing. Also, the importance of making accurate and appropriate inferences to individual characteristics is growing due to unnecessary information overload phenomenon. In this paper, we propose a knowledge reasoning model using association rules and cluster analysis of multi-contexts. The proposed method provides a personalized healthcare to users by generating association rules based on the clusters based on multi-context information. This can reduce the incidence of each disease by inferring the risk for each disease. In addition, the model proposed by the performance assessment shows that the F-measure value is 0.027 higher than the comparison model, and is highly regarded than the comparison model.

Key Words : Multi-Context, Cluster Analysis, Association Rules, Knowledge Reasoning

1. 서론

바쁜 현대사회에서는 시간적 부족으로 인한 잘못된 식습관과 체력관리가 많은 질병의 원인이 되고 있다. 따라서 평소 일상생활에서의 식습관 관리와 운동을 통한 체력 관리 및 건강관리가 필요하다[1]. 또한 최근 문화의 발달로 인해 정보의 접근성이 높아지고 있으며, 이에 따라 사람들은 무분별하고 과도하게 정보에 노출되고 있다. 이로 인해 많은 사람들은 자신이 원하는 정보를 얻는데 어려움을 느낀다. 우리가 흔히 접하는 매체에서의 식이관리나 운동요법은 많은 사람들의 평균적인 상황정보를 이용하여 추정한다. 이는 사용자들의 상황정보를 고려하지 못하기 때문에 사용자 개개인에게 적합하지 못한 결과가 나올 수 있다. 이에 따라 정보과다 현상과 비 개인화된 추천으로부터 사람들의 상황에 알맞은 정보 추천의 필요성과 수요가 증가하고 있다. 따라서 본 논문에서는 다중 상황의 군집분석과 연관규칙을 이용한 지식 추론모델을 제안한다. 다중 상황은 사용자에 따른 여러 가지의 상황 정보들의 집합이다[2]. 예를 들어 연령, 성별, 흡연여부, 음주여부 등의 복합적인 상황정보들을 다중 상황이라 한다. 사용자들의 건강과 관련된 다중 상황정보를 유사도에 따라 군집함으로써 새로운 사용자에게 적합한 군집을 선별하여 보편적인 정보 추천보다 사용자에게 맞는 적절한 추천이 가능하다. 이뿐만 아니라 사용자 개개인의 유병 여부, 가족력과 같은 세부적인 데이터들을 이용하여 연관규칙을 발견함으로써 한층 더 개인화된 지식추론이 가능하다. 이러한 만성질환관련 상황정보는 국민건강영양조사에서 조사한 통계자료를 기반으로 만성질환과 관련된 속성으로 구성한다[3]. 이는 국민건강영양조사 데이터를 필요한 부분과 불필요한 부분을 구분하여 전처리를 한다. 사람들의 상황정보를 군집하고, 상황정보간의 관계를 분석하여 연관규칙을 생성한다. 이를 바탕으로 사람 개인의 특성을 고려하여 개인화된 건강지식을 추론하고 사용자에게 제공한다.

2. 관련연구

2.1 K-평균 알고리즘 기반 군집 분석

군집분석은 개체들 간의 거리와 유사성을 이용하여 여러 개의 군집으로 분류하고, 다른 군집과의 상이성을 통해 개체관계를 분석하는 방법이다[4]. 이는 유사 특성을 가진 그룹 추출을 목표로 하며, 유사성에만 의존하여 군

집을 형성하는 비지도 학습 방법이다. 주관적인 군집결정보다 객관적인 군집을 결정짓는데 효과적이며, 자연과학, 심리학, 공학 등 다양한 분야에서 활용되고 있다. 군집분석은 계층적 군집분석과 비계층적 군집분석으로 분류한다. 계층적 군집은 사전에 k 를 정하지 않고 개체간의 거리에 의해 가까운 개체끼리 군집함으로써 단계적으로 군집을 형성해 가는 방법이다. 이는 각 개체들이 계층적으로 군집되는 과정을 보여주는 덴드로그램(Dendrogram)으로 인해 군집이 형성되는 과정을 파악하는데 용이하지만 자료의 크기가 커지면 분석하기 어렵고, 군집이 병합된 후에 다시 분할하기 어렵기 때문에 신중히 결정해야 한다[5]. 계층적 군집 분석의 예로는 최단 연결법과 최장 연결법이 존재하고, 뒤이어 초기의 방법들의 단점을 보완할 수 있는 중위수 연결법, 중심 연결법, 평균 연결법 등이 제안되었다. 비계층적 군집분석은 군집의 수 k 를 설정하고 군집의 중심으로부터 가까운 개체들을 하나씩 포함해 가는 방식이다. 방대한 양의 자료를 분류하는데 용이하지만 사용자가 군집의 수를 정해주어야 하고, 초기 값에 따라 분석결과가 달라질 수 있다. 비계층적 군집 분석으로 제일 많이 쓰이는 분석 방법에는 K-평균 알고리즘이 있다[6]. K-means 알고리즘은 데이터를 입력받아 k 개의 클러스터로 묶는 알고리즘이다. 각 군집은 하나의 중심을 가지고 각 개체는 군집의 중심으로부터 가장 근접한 군집으로 할당되며, 할당된 개체들이 모여서 하나의 군집을 형성한다. 사용자가 k 개의 군집수를 사전에 지정하여야 알고리즘이 실행된다. Fig 1은 미국의 폭력 범죄율에 관한 K-means 군집의 결과를 나타낸다.

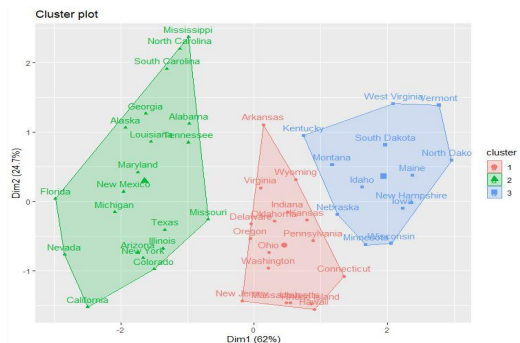


Fig. 1. Example of the K-means Clustering on U.S. Violence Crime Rates

2.2 연관규칙 탐색을 이용한 분석

다양한 분야에서 많은 양의 데이터들이 축적됨으로써 기존의 데이터로부터 잠재되어 있는 새로운 지식을 도출

하는 데이터 마이닝의 중요성이 커지고 있다[7,8]. 연관 규칙은 데이터로부터 변수간의 관계를 발견하여 일련의 규칙을 생성하는 분석방법으로 $X \rightarrow Y$ 로 나타낸다[9]. 이를 통해 기존의 대규모 데이터로부터 한층 더 확장된 지식패턴을 얻을 수 있다. 연관 규칙의 평가 척도로는 향상도(Lift), 신뢰도(Confidence), 지지도(Support)가 있다. 향상도는 X의 신뢰도와 Y의 신뢰도를 곱한 값에 대한 X, Y의 지지도를 뜻한다. 신뢰도는 X 데이터에 다른 X와 Y의 동시 해당 비율을 뜻한다. 지지도는 전체 데이터에 따른 X와 Y의 동시 해당 비율을 뜻한다. 연관규칙 알고리즘은 Apriori, DHP, FP-Growth 등이 존재한다 [10,11]. 대표적인 Apriori 알고리즘은 데이터들의 발생 빈도를 기반으로 연관규칙을 발견한다. 이는 모든 항목집합에서 사용자가 설정한 최소 지지도 이상을 갖는 빈발 항목집합을 추출한다. DHP 알고리즘은 트랜잭션의 규모와 개수를 축소하며 헤시 테이블을 만들어 항목을 비교하는 방법이다[12]. FP-Growth는 Apriori의 처리속도 문제를 FP-tree구조를 이용하여 보완한 방법이다[13]. 이는 후보항목집합을 만들지 않으며 링크와 트리를 이용한다.

3. 다중상황의 군집분석과 연관규칙을 이용한 지식추론 모델

3.1 국민건강영양 조사를 활용한 상황정보 구성

만성질환의 건강 데이터는 국민건강영양조사에서 제공하는 데이터를 이용한다. 이는 고혈압, 당뇨, 고지혈증 등에 따른 8,518건의 설문 데이터로 구성된다. 먼저 원본 데이터로부터 전처리 과정을 통해 사용하지 않을 변수를 제거한 뒤 순수 고혈압, 순수 당뇨, 순수 고지혈증 변수를 생성한다. 이 중 명목형 데이터인 성별, 연령, 흡연여부 변수는 정제과정을 거치지 않고, 연속형 데이터인 콜레스테롤, 혈당, 혈압, 체질량과 같은 변수는 정제과정을 거친다. 이를 마친 뒤 데이터를 선별하고 결측값과 무응답을 제거하여 5,326건의 데이터를 추출한다. 추출한 데이터로부터 55개의 변수를 가지는 데이터집합 $A = \{ID, age \dots DK9_pr, DK4_pr\}$ 를 구성한다. 또한 데이터집합 A로부터 비유병자를 제거한 만성질환 유병자만을 포함하는 데이터를 이용하여 17개의 변수를 가지는 데이터 집합 $B = \{go, dang \dots dbp, BMI\}$ 를 구성한다. 비유병자를 포함한 데이터집합 A는 K-means 알고리즘을 이용

하여 군집을 생성한다. 데이터집합 B는 Apriori 알고리즘을 이용하여 연관규칙을 생성한다.

3.2 만성 상황정보 군집 분석

Table 1은 비유병자와 유병자가 모두 존재하는 데이터집합 A를 기반으로 K-means 알고리즘을 사용하여 만든 군집을 나타낸다. Table 1의 상황정보들은 각각 만성질환 가족력여부(fh), 수축기혈압(sbp), 이완기 혈압(dbp), 고혈압 유병여부(HP), 신장(ht), 체중(wt), 허리둘레(wc), 체질량지수(BMI) 등을 의미한다. 이는 만성질환의 유병과 유의미한 관계를 갖는 상황정보들을 변수로 나타낸 것이다. 반복적으로 k의 값을 지정한 결과, k의 값을 5로 지정하였을 때 데이터의 분포가 가장 균일하게 나타나기 때문에 k의 값을 5로 지정한다. 군집분석을 한 결과 가족의 질병 여부, 혈압의 수치, 고혈압의 유병여부 등이 만성질환의 유병에 밀접한 관계가 있음을 알 수 있다.

Table 1. Context information cluster

Cluster	1	2	3	4	5
fh	0.735	0.782	0.940	1.079	0.721
sbp	127.535	103.220	126.353	152.394	111.125
dbp	84.833	67.817	76.236	85.674	73.009
HP	2.407	1.154	2.292	2.978	1.595
ht	170.757	160.412	156.596	157.586	169.646
wt	82.671	54.268	57.310	60.291	71.368
wc	94.997	72.398	80.579	83.592	86.442
BMI	28.420	21.126	23.427	24.291	24.868

Fig 2는 Table 1을 기반으로 유사한 성향이 나타나는 5개의 군집의 만성 상황정보에 따른 시각화이다. 군집결과 각 군집의 명확한 구별을 위해 5가지의 기호를 사용한다. 상황정보의 종류에 따른 8차원의 군집을 2차원으로 나타낸다.

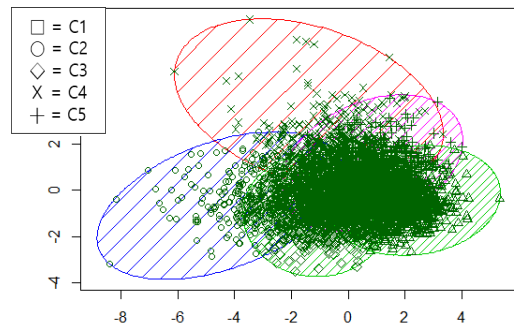


Fig. 2. Result of multi-context cluster

3.3 다중 상황정보에서 연관규칙 생성

만성질환 관련 상황 데이터를 기반으로 연관규칙 알고리즘을 사용하여 각 상황정보 사이의 관계를 분석한다. 연관규칙에서 상황정보는 고혈압(go), 당뇨(dang), 고지혈증(gozi)의 유병여부와 음주(drink), 흡연(smoke), 요단백(upro), 콜레스테롤(chol), 운동여부(exercise), 혈당(glu), 혈압(bp), 체질량(BMI)으로 구성한다. 이는 국민건강영양조사 데이터에서 선별한 고혈압, 당뇨, 고지혈증 중 적어도 하나가 True인 데이터 500건 중 결측값이 없는 210건을 사용한다. 연관규칙은 Apriori 알고리즘을 이용하여 각 변수들이 연관되어 나타나는 정도를 탐색하며, 최소지지도는 0.3 최소신뢰도는 0.2 최소항목집합은 2를 만족하는 규칙을 탐색한다[14]. Table 2는 만성 상황정보 연관규칙 탐색 결과를 나타낸다. Table 2에서 {exercise=0, bp=1} => {go=1}은 운동을 하지 않고 혈압이 높을 경우 고혈압 환자일 경우가 높다는 연관규칙을 나타낸다. 이와 같은 만성 상황정보 연관규칙을 이용하여 사용자의 생활습관을 통해 만성질환이 발생할 가능성을 추론할 수 있다.

Table 2. Association rules

Rule	Sup.	Conf.	Lift
{exercise=0, bp=1}=>{go=1}	0.322	1	3.105
{chol=1, BMI=1, dang=1, drink=1} =>{gozi=1}	0.305	1	3.025
{glu=1, BMI=1, chol=1} =>{dang=1}	0.305	1	3.025
{exercise=0, glu=1, smoke=1} =>{dang=1}	0.330	1	3.025
{exercise=0, drink=1} =>{gozi=1}	0.330	1	3.025
{go=1, chol=1, bp=1} =>{gozi=1}	0.305	1	2.878
{gozi=1, bp=1} =>{go=1}	0.347	1	2.878
{BMI=1, smoke=1, exercise=0, dang=1}=>{go=1}	0.305	1	1.787
...

3.4 만성질환 예방을 위한 지식 추론모델

다중 상황정보를 기반으로 군집 분석과 연관 규칙을 이용하여 만성질환을 예방하기 위한 지식 추론 모델을 제안한다. 이는 K-means 알고리즘을 이용한 군집 분석과 데이터 마이닝의 연관규칙을 사용한다. Fig 3은 제안하는 지식 추론모델의 프로세스를 나타낸다. 이는 국민건강영양조사에서 제공하는 데이터집합을 사용한다. 훈련

데이터집합은 전처리한 데이터집합 A와 A에서 비유병자를 제외한 데이터집합 B로 구성한다. 데이터집합 A를 우선적으로 군집 분석함으로써 가족의 질병여부, 혈압의 수치, 고혈압의 유병여부 등이 만성질환 유병의 밀접한 관계가 있음을 알 수 있다. 또한 데이터집합 B를 이용하여 다중 상황정보간의 연관 규칙을 생성하고 이를 통해 질병의 위험도를 추론할 수 있다. 새로운 사용자의 상황 정보 데이터를 수집하여 만들어진 지식추론 모델을 사용하여 사용자의 유병여부를 확인할 수 있다.

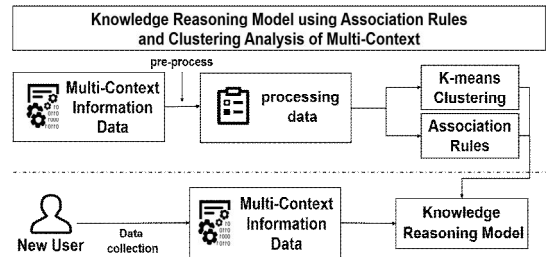


Fig. 3. Process of knowledge reasoning model

4. 성능평가

4.1 K-평균 알고리즘의 군집 수 결정

K-means 알고리즘은 k의 값을 직접 정해주어야 군집이 이루어진다. 일반적으로 군집 내의 분산은 작고, 군집 간의 분산은 클수록 좋은 군집이다. 통상적으로는 식 (1)과 같은 MinSSE로 정의한다. 이는 같은 자료에 여러 차례의 K-means를 실행하였을 때 가장 작은 SSE(Sum of Squared Error), 즉 군집의 중심점들이 각 군집의 특성을 잘 나타낸다. Table 3과 같이 k가 5일 경우 MinSSE 값이기 때문에 k 값을 5로 지정한다.

$$MinSSE = \sum_{i=1}^K \sum_{x \in C_i} d(x, \mu_i)^2 \quad \text{식 (1)}$$

Table 3. Result of SSE values

	k=1	k=2	k=3	k=4	k=5
SSE	34.1	25.2	41.3	54.2	12.5
	k=6	k=7	k=8	k=9	k=10
	23.4	18.6	66.1	24.5	38.9

4.2 군집분석과 연관규칙의 성능평가

제안하는 다중상황의 군집분석과 연관규칙을 이용한 지식 추론모델에 대한 평가는 추론에 대한 재현율 (Recall)과 정확도(Precision)를 이용한 F-measure로 평가한다[15]. 논문에서 분석한 다중 상황정보 군집과 만성 상황정보 연관규칙을 통해 추론한 결과를 이용하여 정확도와 재현율을 평가한다. 식 (2)는 만성질환에 대한 지식 추론의 F-measure 산출 방법을 나타낸다. 정확도와 재현율은 상관관계이기 때문에 한 개의 결과가 높다고 해서 성능이 좋아지는 것은 아니다. 따라서 정확도와 재현율을 이용한 F-measure로 나타낸다. 목표 변수는 고혈압 의사진단 여부, 당뇨 의사진단 여부, 고지혈증 의사진단 여부로 제안하는 방법을 통해 추론한다. 이는 목표변수의 값이 True인 데이터 500건 중 연관규칙 분석에 사용한 210건을 제외한 290건과 분석에서 사용하지 않은 데이터를 추가로 210건 선별하여 평가에 사용한다. 이중 고혈압은 140건, 당뇨는 90건, 고지혈증은 60건이고, 추가로 선별한 210건은 만성질환 값이 False인 데이터이다. Table 4는 오차행렬을 나타내는 표이다. True Positive는 데이터의 만성질환의 실제 값과 추론한 값이 True인 경우를 나타낸다. False Positive는 실제 값이 False이고, 예측 값이 True로 잘못 추론된 값을 나타낸다. 고혈압, 당뇨, 고지혈증 데이터는 True Positive와 False Negative로, 만성질환 값이 False인 데이터는 False Positive와 True Negative로 나타낸다.

Table 4. Confusion matrix

	Prediction Positive	Prediction Negative
Actual Value Positive	True Positive (TP)	False Negative (FN)
Actual Value Negative	False Positive (FP)	True Negative (TN)

$$F = 2 * \frac{Precision * Recall}{Precision + Recall} \quad \text{식 (2)}$$

Table 5는 추론 방법에 따른 성능평가 결과를 나타낸다. Table 5에서 KRM(Knowledge Reasoning Model)은 제안하는 다중상황의 군집분석과 연관규칙을 이용한 지식추론 모델을 나타낸다. BKRM(Bayesian Knowledge Reasoning Model)은 베이저안 지식추론 모델을 나타

낸다.

Table 5. Performance evaluation

	KRM	BKRM
Precision	0.656	0.632
Recall	0.754	0.723
F-measure	0.701	0.674

평가 결과 제안하는 KRM 방법의 정확도는 0.656, 재현율은 0.754, F-measure는 0.701로 다른 방법에 비해 우수하게 나타난다. 따라서 다중상황의 군집분석과 연관규칙을 이용한 건강추론 모델은 유병율이 높은 만성질환에 대해 베이저안 지식추론 모델 보다 정확한 추론이 가능하다.

5. 결론

현대인들은 바쁜 일상생활 속에서 제대로 된 식습관 조절과 체력관리가 이루어지지 못하고 있다. 이에 따라 당뇨, 고혈압, 고지혈증 등의 만성질환의 발병률은 높아지고 있다. 또한 정보의 과다화로 인해 개인에게 맞는 정보를 찾기란 쉽지 않은 일이다. 따라서 본 논문에서는 개인의 특성에 맞고 정확한 정보를 추론하기 위한 다중상황의 군집분석과 연관규칙을 이용한 지식 추론모델을 제안한다. 상황정보는 국민영양조사에서 제공하는 데이터를 두 분류로 전처리한 후 데이터집합 A는 군집분석에 이용하고, 데이터집합 B는 연관관계 규칙을 생성하여 사용자의 생활습관을 통한 만성질환의 발생 가능성을 추론한다. 이와 같이 사용자에 대한 다중상황의 군집분석을 이용하여 정확한 개인화된 정보를 얻을 수 있다. 또한 F-measure를 이용한 성능평가를 통해 베이저안 지식추론 모델보다 높게 평가되는 것으로 보아 만성질환에 대한 정확한 추론이 가능함을 알 수 있다.

REFERENCES

[1] K. Chung & R. C. Park (2019). Cloud based U-healthcare Network with QoS Guarantee for Mobile Health Service, *Cluster Computing*, 22(1), 2001-2015.
 [2] R. Zhao, W. Ouyang, H. Li & X. Wang. (2015). Saliency

Detection by Multi-Context Deep Learning. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 1265-1274.

- [3] Korea Centers for Disease Control and Prevention (KCDC). <http://www.cdc.go.kr/>.
- [4] G. O. Kim, G. S. Lee & S. H. Lee. (2014). An Edge Extraction Method using K-means Clustering in Image. *Journal of Digital Convergence*, 12(11), 281-288.
- [5] P. Langfelder, B. Zhang & S. Horvath. (2007). Defining Clusters from a Hierarchical Cluster Tree: The Dynamic Tree Cut Package for R. *Bioinformatics*, 24(5), 719-720.
- [6] J. A. Hartigan & M. A. Wong. (1979). Algorithm AS 136: A K-means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C*, 28(1), 100-108.
- [7] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin. (2005). The Elements of Statistical Learning: Data Mining, Inference and Prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
- [8] J. C. Kim & K. Chung. (2019). Associative Feature Information Extraction using Text Mining from Health Big Data. *Wireless Personal Communications*, 105(2), 691-707.
- [9] K. T. Oh & S. Y. Lee. (2019). A Movie Recommendation System processing High-Dimensional Data with Fuzzy-AHP and Fuzzy Association Rules. *Journal of Digital Convergence*, 17(2), 347-353.
- [10] S. Zhang. (2011). An Apriori-based Algorithm of Association Rules based on Cloud Computing. *Communications Technology*, 44(6), 141-143.
- [11] C. Borgelt. (2005). An Implementation of the FP-growth Algorithm, *In Proc. of the International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations*, 1-5.
- [12] J. C. Kim & K. Chung. (2017). Emerging Risk Forecast System using Associative Index Mining Analysis. *Cluster Computing*, 20(1), 547-558.
- [13] D. M. Powers. (2011). Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.
- [14] J. Davis & M. Goadrich. (2006). The Relationship between Precision-Recall and ROC Curves. *In Proc. of the International Conference on Machine Learning*, 233-240.
- [15] P. S. Bradley & U. M. Fayyad. (1998). Refining Initial Points for K-Means Clustering. *In Proc. of the International Conference on Machine Learning*, 91-99.

신 동 훈(Dong-Hoon Shin)

[학사회원]



- 2019년 2월 : 동서대학교 컴퓨터공학부 (공학사)
- 2019년 3월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마케팅 연구실 연구원
- 2019년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 석사과정
- 관심분야 : 데이터 마이닝, 헬스케어, 딥러닝, AR/VR, 추천 시스템, 데이터 분석, 인공지능
- E-Mail : ehdgns8227@naver.com

김 민 정(Min-Jeong Kim)

[학사회원]



- 2017년 3월 ~ 현재 : 경기대학교 컴퓨터과학과 학부생
- 2018년 6월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마케팅 연구실 연구원
- 관심분야 : 데이터 마이닝, 빅데이터, 추천 시스템, 데이터 분석, 헬스케어
- E-Mail : minjeog0513@naver.com

오 상 엽(SangYeob Oh)

[정회원]



- 1991년 2월 : 광운대학교 대학원 전자계산학과 (이학석사)
- 1999년 2월 : 광운대학교 대학원 전자계산학과 (이학박사)
- 2007년 2월 ~ 현재 : 가천대학교 IT대학 인터랙티브미디어학과 교수
- 관심분야 : 인공지능, HCI, 차량 통신, 영상관리, 음성 및 음향 신호처리, 정보검색, 추천 시스템, 기계학습
- E-Mail : syoh1234@gmail.com

정 경 용(Kyungyong Chung)

[정회원]



- 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, 데이터 분석, 인공지능, 추천 시스템
- E-Mail : dragonhci@gmail.com