

상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천

백지원¹, 김민정¹, 박찬홍², 정호일³, 정경용^{4*}

¹경기대학교 컴퓨터과학과, ²상지대학교 정보통신공학과,

³대림대학교 컴퓨터소프트웨어과, ⁴경기대학교 컴퓨터공학부

Non-hierarchical Clustering based Hybrid Recommendation using Context Knowledge

Ji-Won Baek¹, Min-Jeong Kim¹, Roy C. Park², Hoill-Jung³, Kyungyong Chung^{4*}

¹Department of Computer Science, Kyonggi University, South Korea

²Department of Information Communication Engineering, Sangji University, South Korea

³Department of Computer Software, Daelim University, South Korea

⁴Division of Computer Science and Engineering, Kyonggi University, South Korea

요약 현대 사회에서 사람들은 시간적인 여유, 경제적인 문제 등에 따라 여행지에 대해 심각한 고민을 한다. 따라서 본 논문에서는 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천을 제안한다. 제안하는 방법은 사용자의 위치, 장소, 날씨 등의 상황에 따라 선호하는 여행지에 대한 지식을 추천받을 수 있는 개인화된 방법이다. 설문조사를 통해 수집된 데이터로부터 14개의 속성을 기반으로 유사한 특성을 가진 사용자들을 비계층적 군집 기반 하이브리드 추천을 이용하여 군집한다. 이는 암묵적 데이터와 명시적 데이터에 가중치를 부여하여 보다 정확한 추천을 한다. 이를 통해 사용자는 불필요한 시간을 소모하지 않고 선호하는 여행지를 추천받을 수 있다. 성능평가는 정확도, 재현율, F-measure를 이용한다. 평가 결과 정확도는 0.636, 재현율은 0.723, F-measure는 0.676으로 평가되었다.

• **주제어** : 상황정보, 군집, 암묵적 선호도, 추천 시스템, 지식 추천

Abstract In a modern society, people are concerned seriously about their travel destinations depending on time, economic problem. In this paper, we propose a non-hierarchical clustering based hybrid recommendation using context knowledge. The proposed method is a personalized way of recommended knowledge about preferred travel places according to the user's location, place, and weather. Based on 14 attributes from the data collected through the survey, users with similar characteristics are grouped using a non-hierarchical clustering based hybrid recommendation. This makes more accurate recommendation by weighting implicit and explicit data. The users can be recommended a preferred travel destination without spending unnecessary time. The performance evaluation uses accuracy, recall, F-measure. The evaluation result was shown 0.636 accuracy, 0.723 recall, and 0.676 F-measure.

• **Key Words** : Context Information, Clustering, Recommendation System, Knowledge Recommendation

Received 11 September 2019, Revised 22 September 2019, Accepted 30 September 2019

* **Corresponding Author** Kyungyong Chung, Division of Computer Science and Engineering, Kyonggi University, 154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, South Korea. E-mail: roypark1984@gmail.com

I. 서론

현대인들은 바쁜 일상을 떠나 휴식을 위해 여행을 떠나고 있다. 여행은 일상에서 받은 스트레스를 해소하고, 다양한 경험을 통해 삶의 의미를 찾으며 자신을 되돌아보게 되는 시간을 가질 수 있다. 이에 따라 여행의 중요성이 높아지고 있다. 하지만 많은 현대인들은 시간적인 여유나 경제적인 요인 등으로 여행을 다니기 힘든 상황에 놓여있다. 또한 여행을 가게 되더라도 평균 3.7일로 매우 짧은 여행 일수를 나타내고 있다[1-2]. 한국관광공사[1]와 문화체육관광부[2]가 국민 1,005명을 대상으로 설문조사를 진행한 결과, 여행 계획이 없는 응답자들 중에서 이유로는 마음의 여유 부족, 여행비용 부족, 여가시간 부족으로 응답하였다. 이는 시간과 비용이 결정에 중요한 요인으로 나타났다. 또한 여행 계획이 있는 응답자들 중 70%는 국내여행을 떠날 예정이라고 응답하였다. 이처럼 시간적이나 금전적으로 여유가 없는 사람들은 해외여행보다 국내 여행을 선호하고 있다. 여행은 계획 단계부터 많은 시간과 비용이 소모되며, 많은 사람들은 짧은 시간 안에 여행계획을 수립하는 것에 어려움을 느끼고 있다. 따라서 사용자의 상황에 따라 적합한 여행정보에 대한 추천이 필요하다. 현재 대부분의 여행지 정보를 제공하는 웹 사이트, 애플리케이션 등의 시스템들은 사용자의 리뷰를 제공한다. 이에 따라 다른 사용자들이 다녀간 여행지를 중 선호도가 가장 높은 일부의 여행지만을 추천하기 때문에 실제 사용자에게 적합한 추천을 하지 못하고, 여행지에 대한 정보가 부족하다[3]. 사람들은 여행지를 결정할 때 날씨나 상황에 따른 영향을 많이 받게 된다. 또한 여행을 가는 날짜와 시간이 제한적이므로 대부분의 사람들은 이를 고려하여 자신의 위치정보에 따라 행선지를 결정한다. 모든 사람들이 같은 장소에서 동일한 상황을 겪는 것은 아니다. 따라서 사용자의 상황정보를 고려하고 여행지에 대한 지식을 제공하여 개인화된 여행지를 추천하는 방법이 필요하다[4].

본 논문에서는 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천을 제안한다. 제안하는 방법은 성향이 유사한 사용자들을 군집한다. 또한 사용자의 위치, 날씨, 교통 등의 상황정보를 인식하여 지식을 생성하고, 개인적인 선호도에 의한 여행지를 추천하며 상황지식을 제공하여 추천에 대한 정확도를 높인다.

본 논문의 구성은 다음과 같다. 2장에서는 유사 성향 사용자 그룹을 위한 군집과 추천 기술의 최근 동향에 대해 기술하고 3장에서 제안하는 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천에 대해 기술한다. 4장에서는 성능 평가 및 결과에 대해 기술하고, 5장에서는 결론을 맺는다.

II. 관련 연구

2.1 유사 성향 사용자 그룹을 위한 군집

사람들은 각각의 다양한 성향을 가지고 있다. 유사한 성향을 가진 사람들 간에는 공감대가 형성되며 원활한 의사소통이 가능하다. 군집분석 방법을 적용하여 유사 성향의 그룹을 생성한다. 군집 분석은 정확한 답이 없는 비지도 학습으로 객체간의 유사한 특성을 추출하여 군집을 형성한다[5, 6]. 군집은 주관적인 특징보다 객관적인 특징을 결정짓는데 효과적이며, 심리학, 자연과학, 공학 등 다양한 분야에서 활용이 가능하다. 군집 분석은 일반적으로 계층적과 비계층적 군집분석으로 구분된다.

계층적 군집분석은 군집의 수를 정하지 않고 개체간의 거리에 의해 가까운 개체끼리 군집함으로써 단계적으로 확장하여 군집을 형성해 나가는 방법이다[7-9]. 이는 데이터의 차원수가 커지면 복잡해지기 때문에 분석하기 어렵고, 군집을 병합한 후에 다시 분할하기 어려우므로 신중한 결정이 요구된다. 계층적 군집 분석은 최단연결법과 최장연결법, 중심연결법, 중위수연결법, 평균연결법이 있다[10-12].

비계층적 군집 분석[13]은 군집의 수를 설정하고 군집의 중심으로부터 거리가 가까운 개체들을 하나씩 포함하여 군집을 생성하는 방식이다. 이는 방대한 양의 자료를 분류하는데 용이하지만 사용자가 군집의 수를 정해주어야 하고, 초기값에 따라 분석결과가 상이하다. 또한 계층적 군집 분석과 비교하여 연산 속도가 빠르고 반복적인 학습을 통해 순차적 군집이 가능하다. K-평균 알고리즘은 비계층적 군집 방법으로 분리형 군집화 알고리즘이다. 이는 주어진 데이터를 k개의 클러스터로 그룹을 생성하고, 각각의 군집에 대해 하나의 중심점을 계산한다. 거리가 가까운 클러스터의 중심에 객체가 할당되고, 중심에 모인 개체들이 하나의 군집을 형성한다[14]. 반복적인 실험을 통해 군집의 수를 지정해야 하고, 연속형 변수를 사용해 중심 재계산을

여러번 반복해서 진행해야 성능이 최적인 군집을 형성하게 된다. 그림 1은 K-평균 알고리즘을 이용한 군집을 나타낸다. 반복적인 실험을 통해 최적의 K를 발견하고, 유사한 특징을 추출하며 사용자에게 개인화 추천을 위해 사용한다.

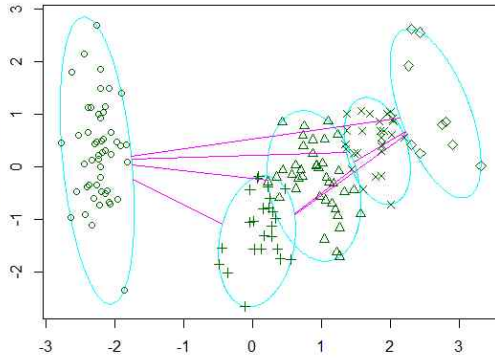


Fig. 1. Clustering using K-means Algorithm

그림 1에서는 최적의 K개수를 5개로 하여 군집을 형성한다. 각각의 군집에는 유사한 특성을 가진 개체가 속해있으며 각 군집간의 거리를 통해 다른 군집과의 관계도 발견할 수 있다.

2.2 추천 기술의 최근 동향

추천 시스템은 사용자의 선호도와 개인적인 정보 등을 기반으로 하여 적합한 상품이나 콘텐츠를 추천하는 방법이다[15]. 이는 주요 콘텐츠 서비스뿐만 아니라 음악, 영화, 도서, 등 다양한 분야에서 활용된다. 추천 시스템의 추천 방법은 콘텐츠 기반 필터링, 협력 필터링, 하이브리드 필터링이 있다. 콘텐츠 기반 필터링은 아이템에 대한 프로파일 내용 정보를 분석하여 유사도를 기반으로 추천한다. 이는 사용자의 수가 적을 때 용이하게 이용하지만, 다양한 사용자에 대한 추천의 정확도가 낮다는 단점이 있다[16].

협력적 필터링은 사용자들 간의 특정 아이템에 대해 선호도가 같을 경우 다른 아이템에서도 유사한 선호도를 보일 것으로 예측하는 방법이다[17]. 협력적 필터링은 아이템의 다양성으로 중요해지고 있지만 초기 사용자가 적을 경우에 유사도를 측정할 충분한 데이터가 존재하지 않는 문제점이 존재한다. 하이브리드 필터링[18]은 콘텐츠 기반 필터링과 협력적 필터링의 장점을 혼합함으로써 단점을 보완하고 각 필터링의 장점을 극대화한다. 최근에는 딥러닝, 강화학습을 적용한 추천 방법이 급증하고 있다. 그림 2는 하이브리드 추천 프

로세스를 나타낸다. 좌측은 사용자간의 유사도를 나타내며 우측은 아이템간의 유사도를 나타낸다.

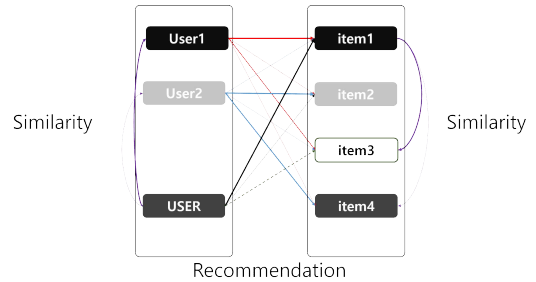


Fig. 2. Hybrid Recommendation Process

그림 2에서 사용자간의 유사도와 아이템간의 유사도를 계산한다. 예를 들어, User1은 item1을 선호하며 이와 유사한 item3을 추천한다. 이에 따라 사용자가 선호하는 아이템을 추천한다.

III. 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천

3.1 데이터 수집 및 구성

상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천은 사용자의 선호도에 따라 여행지에 대한 추천과 지식을 제공한다. 상황 지식은 경제요건, 실시간 날씨, 교통정보 등으로 생성하고, 사용자에게 적합한 여행지는 사용자 선호도와 여행 선호 조건을 이용한다.

사용자 선호도 수집을 위한 기초 데이터는 스마트관광 앱 개발 오픈 API[19]와 오픈 데이터 포털[20]에서 여행지 테마와 왕복시간에 따른 데이터를 수집하여 설문지를 구성한다. 이에 따라 인터넷 커뮤니티를 통해 각 테마와 왕복시간에 따른 선호도 설문조사를 진행하여 총 129명의 데이터를 수집한다. 분석의 정확도를 향상시키기 위해 데이터 중 NULL 값이 있거나 불성실하게 응답한 19명의 데이터를 제외한 전처리된 110명의 데이터를 활용한다. 여행지 테마별 선호도 설문지는 관광지 정보 데이터에서 테마를 기준으로 문화, 예술, 자연, 체험, 학습 등의 14개의 속성으로 구성한다. 여행지 왕복시간에 따른 선호도 설문지는 사용자의 현재 위치로부터 관광지까지 사용자가 평소 선호하는 왕복시간을 2시간 이내, 4시간 이내, 6시간 이내, 6시간 이상으로 구성한다. 선호도는 1-5까지의 가중치를 두

고, 1에 근접 할수록 선호하지 않음을 의미하고, 5에 근접할수록 선호함을 의미한다. Table 1은 여행지 테마와 소요시간에 따른 선호도를 나타낸다.

Table 1. Preference of Theme and Turnaround time

Theme	User					
	u1	u2	u3	u4	...	u110
Religion	1	1	1	4	...	4
History	1	2	1	3	...	5
Tradition	1	2	2	3	...	1
Experience	4	3	3	2	...	4
Learning	2	2	2	3	...	1
Industry	1	2	2	3	...	1
Culture	1	4	3	1	...	4
Art	5	3	5	2	...	2
Play	5	4	5	4	...	4
Nature	2	5	3	5	...	5
Sports	5	1	2	5	...	1
Healing	3	3	5	4	...	4
Camping	1	1	3	2	...	4
Shopping	3	4	4	4	...	4
2hours	4	5	3	3	...	4
4hours	4	3	4	4	...	5
6hours	5	1	5	5	...	5
6hours over	2	3	1	1	...	2

Table 2는 사용자가 여행에 선호하는 조건을 나타낸다. 이는 왕복 소요시간, 비용(만원), 동반객의 수로 구성된다.

Table 2. User Preference of Trip-condition

Attribute	User					
	u1	u2	u3	u4	...	u110
Turnaround time	7	5	2	8	...	11
Costs	120	450	375	240	...	400
Companion	2	8	7	4	...	7

Table 2에서는 선호하는 조건을 통해 선호하는 조건을 알 수 있다. 이는 암묵적인 추천을 위해 사용되는 데이터이다.

3.2 사용자 선호도 비계층적 군집

사용자 선호도 군집을 위해 비계층적 군집 방법을 사용한다. 비계층적 군집 방법은 연산 속도가 빠르고 순차적 군집 분석이 가능한 장점이 있다. 일반적으로 비계층적 군집 방법에는 k-means와 k-medoids[21]가

있다. k-means는 k의 값을 지정하여 객체들의 평균값으로 군집을 구성하는 방법이다. k-medoids는 여러 개의 객체 중에서 k개의 대표객체(medoid)를 지정하고, 나머지 객체들을 유사성에 따라 대표객체에 군집하도록 하는 방법이다. k-medoids는 k-means 보다 계산 비용이 높고 군집에 소요되는 시간이 길다. 따라서 본 논문에서는 k-means 알고리즘을 이용하여 사용자 선호도 군집을 형성한다. 군집분석의 결과 k값을 5로 지정하였을 때 데이터의 분포가 가장 균일하게 나타났다. 각 군집의 중심점은 여행지별 테마와 왕복시간에 따른 선호도 점수의 평균으로 구성한다. Table 3은 사용자의 선호도 군집 결과를 나타낸다. Table 3의 Cluster는 0~4까지의 5개의 군집으로 구성되며 values는 각 군집의 대표값을 나타낸다. 값들을 통해 각 군집들이 선호하는 테마와 시간을 쉽게 비교할 수 있다. Users는 각 군집에 속하는 사용자의 수를 나타낸다.

Table 3. Result of Clustering

Cluster	1	2	3	4	5
Users	15	18	31	28	18
Religion	4.65	2.33	2.06	1.79	3.20
History	4.21	2.65	2.87	2.39	2.75
Tradition	3.63	2.12	2.86	1.93	3.09
Experience	2.79	3.65	3.42	3.35	3.04
Learning	3.21	3.78	3.56	3.13	3.46
Industry	2.11	2.17	2.63	3.05	3.32
Culture	2.43	3.48	4.72	3.46	4.39
Art	2.79	2.77	4.29	3.79	4.18
Play	3.13	4.11	4.53	4.42	3.99
Nature	4.33	3.00	3.49	4.71	3.76
sports	2.46	4.53	3.08	2.87	3.58
Healing	3.49	3.98	3.68	4.14	3.80
Camping	2.14	4.32	3.03	2.76	3.00
Shopping	2.36	3.12	4.07	3.00	3.69
2hours	4.79	3.94	3.65	4.33	3.17
4hours	3.39	3.44	4.06	4.27	4.50
6hours	2.18	4.00	3.03	4.47	3.44
6hours over	2.07	2.28	2.19	1.67	1.61

3.3 상황 지식을 이용한 하이브리드 추천

본 논문에서는 사용자에게 여행지 추천과 해당 여행지에 대한 하이브리드 추천을 제공한다. 하이브리드 추천 방법[18]은 콘텐츠 기반 추천과 협력적 필터링을 혼합한 방법이다. 이는 희소성 문제와 초기평가문제를 해결하여 보다 정확한 추천이 가능하다. 또한 사용자의 상황 지식, 선호 조건, 선호도를 모두 고려할 수 있

는 방법이므로 추천 정확도를 향상시킨다. 이를 위해 사용자의 상황 지식과 설문조사에서 수집한 정보로 만든 군집을 통해서 사용자의 선호도에 따라 새로운 사용자와 가장 유사한 군집을 탐색한다. 선호도 데이터를 도출한 5개의 군집을 바탕으로 유클리디안 거리를 사용한다[22]. 이는 데이터 사이의 거리를 통해 유사도를 측정하는 기법이다. 군집된 사용자 선호도 테마 군집의 각 좌표에 속한 값들을 계산하여 유클리디안 거리를 계산하고, 이 거리가 작을수록 데이터의 유사성이 크다고 판단할 수 있다. 또한 유클리디안 거리를 통해 각 군집마다 어떠한 상관관계를 가지고 있는지 분석할 수 있다. 이를 통해 새로운 사용자가 각 속성에 대한 선호도를 입력하게 되면 해당 데이터가 어느 군집에 거리적으로 근접한지를 계산하고 유사도를 판단하여 사용자가 원하는 속성이 어떠한 것인지 암묵적으로 알 수 있다. 사용자와 가장 유사한 군집의 데이터를 발견하고 관광지 정보데이터에서 가장 적절한 테마가 포함되는 곳을 암묵적으로 선택한다. 사용자가 정한 시간 선호도에 따라 거리가 먼 곳은 2차적으로 필터링한다. 그림 3은 시스템 프로세스를 나타낸다.

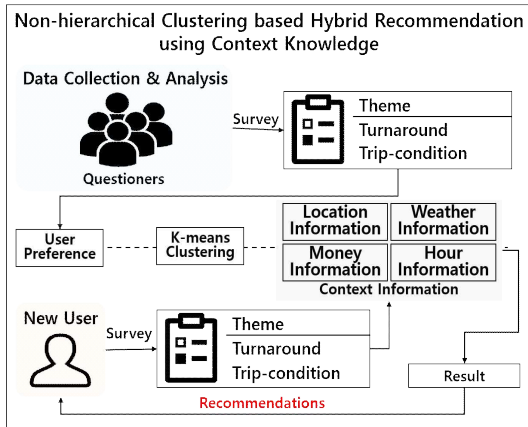


Fig. 3. System Process

예를 들어, 새로운 사용자가 설문조사를 하여 얻어진 데이터로 군집한 데이터와 유클리디안 거리 측정을 통해 A가 속한 군집을 찾고 속한 군집 내에서 선호하는 여행지를 차례대로 나열한다. 암묵적 추천받은 여행지에 따른 날씨 상황정보를 적용하여 사용자에게 가장 적절한 장소를 추천한다.

IV. 성능 평가 및 결과

제안하는 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천에 대한 평가는 추천에 대한 정확도와 재현율을 이용한 F-measure[23]로 평가한다. 성능평가를 위해 본 논문에서는 220개의 사용자 선호도 데이터에서 70%는 훈련 데이터로 사용하고, 10%는 모델 검증 데이터 20%는 테스트 데이터로 사용한다. 식 1은 정확도와 재현율을 이용한 F-measure를 나타낸다. 정확도는 제안하는 모델이 추천하는 장소 중에서 사용자가 실제로 선호하는지를 평가한다. 재현율은 사용자의 선호도에 따라 적합한 여행지가 얼마나 추천되었는지를 평가한다. 예를 들어 표 1에서 사용자 1의 예술, 놀이, 스포츠, 체험에 4점 이상으로 사용자가 선호하는 것으로 나타난다. 정확도는 제안하는 추천 모델이 사용자 1이 선호하는 여행 테마를 얼마나 정확하게 추천한 것인지를 나타내고, 재현율은 사용자 1이 선호하는 테마 중에 추천된 테마가 몇 개가 있는지를 나타낸다.

$$F = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad \text{식 1}$$

Table 4는 각 모델에 따른 성능평가 결과를 나타낸다. Model 1은 제안하는 상황지식을 이용한 하이브리드 추천 모델이고, Model 2는 Model 1에서 상황인식에서 위치 정보를 제외한 모델로 각각의 여행지에 대한 테마를 이용하여 추천하는 방법이다.

Table 4. Evaluation Results

	Model 1	Model 2
Precision	0.636	0.582
Recall	0.723	0.666
F-measure	0.676	0.605

평가 결과 제안하는 모델인 Model 1과 Model 2의 정확도는 각각 0.636, 0.582이 평가되었고 재현율은 0.723, 0.666로 나타나며, F-measure 0.676 과 0.605로 나타난다. 정확도와 재현율은 트레이드오프 관계이기 때문에 한 개의 결과가 높다고 해서 성능이 좋아지는 것은 아니다. F-measure 값은 정확도와 재현율의 성능을 하나의 숫자로 표현하는 방법인데 정확도와 재현율의 조화 평균값으로 나타낸다. Model 1이 Model 2보다

F-measure 값이 높게 평가되었다는 것은 Model 1이 더 높은 정확도와 재현율을 가지고 있다고 본다. 제안하는 비계층적 군집 기반 하이브리드 추천 모델은 사용자의 선호도에 따른 적절한 여행지를 예측할 수 있음을 알 수 있다.

V. 결론

최근 사람들은 일상 속에서 받은 스트레스를 해소하고 다양한 체험을 통해 삶의 의미를 찾고 싶어 한다. 하지만 시간적 여유와 경제적인 여유 부족으로 인해 마음 놓고 여행을 하지 못하는 상황에 놓여있다. 이에 따라 본 논문에서는 상황 지식을 이용한 비계층적 군집 기반 하이브리드 추천을 제안한다. 이는 사람들에게 여행지의 테마와 소요시간에 따른 선호도, 여행의 선호조건을 데이터를 설문조사를 통해 수집하였고, 수집한 데이터를 군집하였다. 도출된 군집을 바탕으로 새로운 사용자에게 선호도 값을 받아 유클리디안 거리를 사용하여 새로운 사용자의 선호도 값과 도출된 군집 사이의 거리를 통해 유사도를 측정한다. 새로운 사용자가 속한 군집을 확인하고 상황인식 정보에 따라 필터링을 거친 데이터를 테마별, 거리별로 여행지를 하이브리드 추천 방법으로 추천하여 주는 방안을 수행하였다. 그 결과 성능평가에서 제안하는 모델의 F-measure 값이 비교 모델보다 높게 나타난 것으로 보아 사용자의 기호에 맞춰 높은 정확도로 여행지를 예측할 수 있음을 알 수 있다.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2016R1D1A1A09917313)

REFERENCES

[1] Korea Tourism Organization, <http://www.visitkorea.or.kr>
 [2] Ministry of Culture, Sports and Tourism, <https://www.mcst.go.kr>
 [3] C. C. Yu, and H. P. Chang, "Personalized location-based recommendation services for tour planning in mobile

tourism applications," In International Conference on Electronic Commerce and Web Technologies, pp. 38-49, 2009.
 [4] J. Bao, Y. Zheng, Mokbel, and F. Mohamed, "Location-based and Preference-Aware Recommendation using Sparse Geo-Social Networking Data," In: Proc. of the 20th International Conference on Advances in Geographic Information Systems, ACM, pp. 199-208, 2012.
 [5] A. W. Edwards, and L. L. Cavalli-Sforza, "A Method for Cluster Analysis. Biometrics," pp. 362-375, 1965.
 [6] J. C. Gower, "A Comparison of Some Methods of Cluster Analysis. Biometrics," pp. 623-637, 1967.
 [7] P. Langfelder, B. Zhang, and S. Horvath, "Defining Clusters from a Hierarchical Cluster Tree: the Dynamic Tree Cut Package for R," *Bioinformatics*, Vol. 24, No. 5, pp. 719-720, 2017.
 [8] F. B. Baker, and L. J. Hubert, "Measuring the Power of Hierarchical Cluster Analysis," *Journal of the American Statistical Association*, Vol. 70, No. 349, pp. 31-38, 1975.
 [9] C. C. Bridges Jr, "Hierarchical Cluster Analysis. Psychological Reports," Vol. 18, No. 3, pp. 851-854, 1966.
 [10] J. C. Gower, and G. J. Ross, "Minimum Spanning Trees and Single linkage Cluster Analysis," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 18, No. 1, pp. 54-64, 1969.
 [11] P. Dawyndt, H. De Meyer, and B. De Baets, "The Complete linkage Clustering Algorithm Revisited," *Soft Computing*, Vol. 9, No. 5, pp. 385-392, 2005.
 [12] H. K. Seifoddini, "Single linkage versus Average linkage Clustering in Machine Cells Formation Applications," *Computers & Industrial Engineering*, Vol. 16, No. 3, pp. 419-426, 1989.
 [13] E. Diday, "Optimization in Non-hierarchical Clustering. Pattern Recognition," Vol. 6, No. 1, pp. 17-33, 1974.
 [14] S. Na, L. Xumin, and G. Yong, "Research on K-means Clustering Algorithm: An Improved K-means Clustering Algorithm," In Proc. of the International Symposium on Intelligent Information Technology and Security Informatics. IEEE. pp. 63-67. 2010.
 [15] H. Jung, K. Chung, "Knowledge-based Dietary Nutrition Recommendation for Obese Management," *Information Technology and Management*, Vol. 17, No. 1, pp. 29-42, 2016.
 [16] M. J. Pazzani, and D. Billsus, "Content-based recommendation systems," In *The adaptive web*. Springer, Berlin, Heidelberg, pp. 325-341, 2007.
 [17] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," In *The adaptive web*, pp. 291-324, 2007.

[18] K. Y. Chung, J. H. Lee, "User Preference Mining through Hybrid Collaborative Filtering and Content-based Filtering in Recommendation System," IEICE Transaction on Information and Systems, Vol.E87-D, No.12, pp. 2781-2790, 2004.

[19] TourAPI 3.0, <http://api.visitkorea.or.kr>

[20] Open Data Portal, <http://www.data.go.kr>

[21] P. Arora, and S. Varshney, "Analysis of k-means and k-medoids Algorithm for Big Data," Procedia Computer Science, 78, pp. 507-512, 2016.

[22] P. E. Danielsson, "Euclidean Distance Mapping," Computer Graphics and Image Processing, Vol. 14, No. 3, pp. 227-248, 1980.

[23] K. J. Dembczynski, W. Waegeman, W. Cheng, and E. Hüllermeier, "An Exact Algorithm for F-measure Maximization," In Advances in neural information processing systems, pp. 1404-1412, 2011.

저자 소개

백 지 원 (Ji-Won Baek)



2017년 2월 : 상지대학교 컴퓨터정보공학부 (공학사)
 2017년 2월 ~ 2018년 2월 : 인퍼니 연구원
 2018년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 (석사과정)
 2018년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마이닝 연구실 연구원

관심분야 : 인공지능, 데이터 마이닝, 헬스케어, 딥러닝, 빅데이터 분석

김 민 정 (Min-Jeong Kim)



2017년 3월 ~ 현재 : 경기대학교 컴퓨터과학과 학부생
 2018년 6월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마이닝 연구실 연구원
 관심분야 : 데이터 마이닝, 빅데이터, 추천 시스템, 클라우드

박 찬 흥 (Roy C. Park)



2008년 8월 : 상지대학교 산업공학과 (공학사)
 2010년 8월 : 상지대학교 컴퓨터정보공학과 (공학석사)
 2015년 2월 : 상지대학교 컴퓨터정보공학과 (공학박사)
 2015년 3월 ~ 2019년 2월 : 동서대학교 컴퓨터공학부 교수

2019년 3월 ~ 현재 : 상지대학교 정보통신공학과 교수
관심분야 : 클라우드, 빅데이터, 헬스케어, 인공지능, HCI, 정보검색, 추천 시스템

정 호 일 (Hoill Jung)



2010년 2월 : 상지대학교 컴퓨터정보공학부 (공학사)
 2013년 2월 : 상지대학교 컴퓨터정보공학과 (공학석사)
 2017년 2월 : 상지대학교 컴퓨터정보공학과 (공학박사)
 2018년 3월 ~ 2019년 2월 : 원광대학교 컴퓨터·소프트웨어 공학과 교수

2019년 3월 ~ 현재 : 대림대학교 컴퓨터소프트웨어과 교수
관심분야 : 지능형 시스템, 데이터 마이닝, 추천, 감성공학, 지식 시스템, 클라우드, 빅데이터

정 경 용 (Kyungyong Chung)



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, 정보검색, 추천 시스템