

Clustering method for similar user with Mixed Data in SNS

Hyoung-Min Song*, Sang-Joon Lee**, Ho-Young Kwak***

Abstract

The enormous increase of data with the development of the information technology make internet users to be hard to find suitable information tailored to their needs. In the face of changing environment, the information filtering method, which provide sorted-out information to users, is becoming important. The data on the internet exists as various type. However, similarity calculation algorithm frequently used in existing collaborative filtering method is tend to be suitable to the numeric data. In addition, in the case of the categorical data, it shows the extreme similarity like Boolean Algebra. In this paper, We get the similarity in SNS user's information which consist of the mixed data using the Gower's similarity coefficient. And we suggest a method that is softer than radical expression such as 0 or 1 in categorical data. The clustering method using this algorithm can be utilized in SNS or various recommendation system.

▶ Keyword : Similarity, Clustering, Mixed data, Gower's similarity coefficient

I. Introduction

2013년 한 해 전 세계에서 생산된 데이터 총량은 4.4ZB에 달하고 2020년에는 10배 증가하여 44ZB에 이를 것으로 예측되고 있다[1]. 이런 엄청난 양의 데이터로 인하여 정보선택의 문제가 발생하고 있으며, 사용자에게 불필요한 내용을 제거하고 필요한 정보만을 제공해주는 정보 필터링 기법과 여러 곳에 흩어진 콘텐츠를 하나의 콘셉트나 주제로 모아서 보여주는 콘텐츠 큐레이션 기법이 유행하고 있다[2][3].

일반적으로 추천시스템에 사용되는 정보 필터링기법으로는 협업적 필터링과 내용기반 필터링이 있다. 특히 협업적 필터링에서는 사용자와 성향이 유사한 사용자를 찾기 위해 유사도를 계산해야 하는데 기존 시스템에서 많이 사용되어왔던 피어슨 상관계수(Pearson correlation coefficient), 코사인 유사도(Cosine Similarity) 등의 유사도 계산법은 수치형 데이터에

만 활용이 가능하다. 그러나 SNS(Social Network Service) 상에 존재하는 사용자 정보는 수치형 데이터뿐 아니라 다양한 형태의 데이터로 이루어져있기 때문에 혼합형 데이터의 유사도를 구하는 방법이 필요하다. 본 논문에서는 Gower 유사도 계수(Gower's similarity coefficient)를 사용하여 수치형과 범주형의 혼합형 데이터로 이루어진 데이터의 유사도를 구하고 범주형 데이터의 경우 부울대수 형태의 극단적인 표현이 아닌 완화된 형식으로 계산하는 방법을 제안한다.

본 논문에서는 2장에서 관련 연구에 대하여 설명하고, 3장에서는 제안하는 알고리즘에 대하여 설명하였으며, 4장에서는 결과에 대한 분석을 실시하고, 5장에서 결론을 기술하였다.

II. Related Researches

1. Collaborative filtering

협업적 필터링은 아마존닷컴에서 추천시스템을 만들 때 사

*First Author: Hyoung-Min Song, Corresponding Author: Sang-Joon Lee
*Hyoung-Min Song(smart.lycaeus@gmail.com), Dept. of Computer Science, Graduate School, Jeju National University
**Sang-Joon Lee(sjlee@jejunu.ac.kr), Dept. of Computer Engineering, Jeju National University
***Ho-Young Kwak(kwak@jejunu.ac.kr), Dept. of Computer Engineering, Jeju National University
• Received: 2015. 11. 12, Revised: 2015. 11. 23, Accepted: 2015. 11. 30.

용한 방법으로 유명하며 대부분의 추천시스템에서 가장 많이 사용되는 기법이다. 이 기법은 여러 사용자의 평가값을 이용하여 특정 사용자의 평가값을 예측하는 방식으로 크게 메모리 기반과 모델기반 방식으로 나뉜다.

메모리기반 방식은 두 사용자 간의 또는 항목 간의 유사도를 측정하여 가장 유사한 사용자들 또는 항목들의 평가치를 반영하여 추천한다[4]. 이런 추천 방법은 평가치 데이터가 희소할 경우, 유사도 계산이 불가능하거나 낮은 유사도의 이웃을 선택하게 되어 추천 품질의 수준이 나빠진다. 또한 유사도 측정 기법에 따라 추천의 정확성에 많은 영향을 미친다[5]. 사용되는 유사도 측정 방법으로는 피어슨 상관계수(Pearson's Correlation), 코사인 유사도(Cosine Similarity) 등이 있는데 이러한 방법은 대체로 수치형 데이터에 적합한 계산방법이어서 혼합형 데이터에는 적합하지 않다.

모델기반 방식은 선형대수, 신경망, 군집화 등을 기반으로 사전에 모델을 수립해 두고 추천하는 방식으로 신속한 추천이 가능하지만 모델을 구축하는데 많은 시간이 소요되며 사용자의 최신 동향을 반영하지 못하는 단점이 있다[6].

2. Gower Similarity Coefficient

일반적인 유사도 측정법은 데이터가 수치형이거나 범주형 둘 중 하나의 형식인 경우에만 계산 가능하지만 Gower 유사도 계수는 혼합형 데이터인 경우에도 사용할 수 있다. Gower의 유사도 계수는 식 1과 같이 정의된다[7].

$$S_{ij} = \frac{\sum_{m=1}^p w_{ijm} s_{ijm}}{\sum_{m=1}^p w_{ijm}} \quad (1)$$

여기서 $w_{ijm}=1$ 이고, 제약조건으로 다음의 규칙을 따른다. 첫째, 한 쌍의 레코드들 가운데 하나의 측정치가 알려져 있지 않을 때에는 $w_{ijm}=0$ 이다. 둘째, 이진이 아닌 범주형 변수에 대하여 만약 레코드들이 동일한 범주에 속해 있지 않을 경우에는 $s_{ijm}=0$ 이고, 동일한 범주에 속해 있을 경우에는 $s_{ijm}=1$ 이다. 셋째, 연속형 변수들에 대해서는 식 2와 같다.

$$s_{ijm} = 1 - \frac{|x_{im} - x_{jm}|}{\max(x_m) - \min(x_m)} \quad (2)$$

3. k-means Algorithm

군집화는 크게 계층형 군집화와 비계층형 군집화 두 가지 방법으로 연구되어 왔다. 일반적으로 계층형 군집화는 수행속도가 느리며 데이터의 극단값이나 오류데이터에 민감한 반응을 보인다. 또한 군집화 이전 단계의 오류를 다시 고칠 수 있는 방법이 없다. 따라서 계층적 군집화로는 처리할 수 있는 데

이터에 한계가 있고 시간복잡도 측면에서 비효율적이다. 그래서 비계층적 군집화의 일종인 k-means 기반의 군집화가 주로 사용되어 왔다[8].

군집화 알고리즘 중 분할법에 속하는 k-means 알고리즘은 구현이 쉽고, 패턴의 수가 n일 때 시간 복잡도가 $O(n)$ 인 장점이 있다. 하지만 초기 군집의 중심에 상당히 종속적이라는 단점이 있다[9]. k-means 알고리즘의 개념은 예를 들어 n개의 데이터가 있다면 이 데이터를 n보다 작거나 같은 k개의 군집으로 분할하고 분할된 군집의 중심과 소속 데이터와의 거리를 최소화하는 것이다.

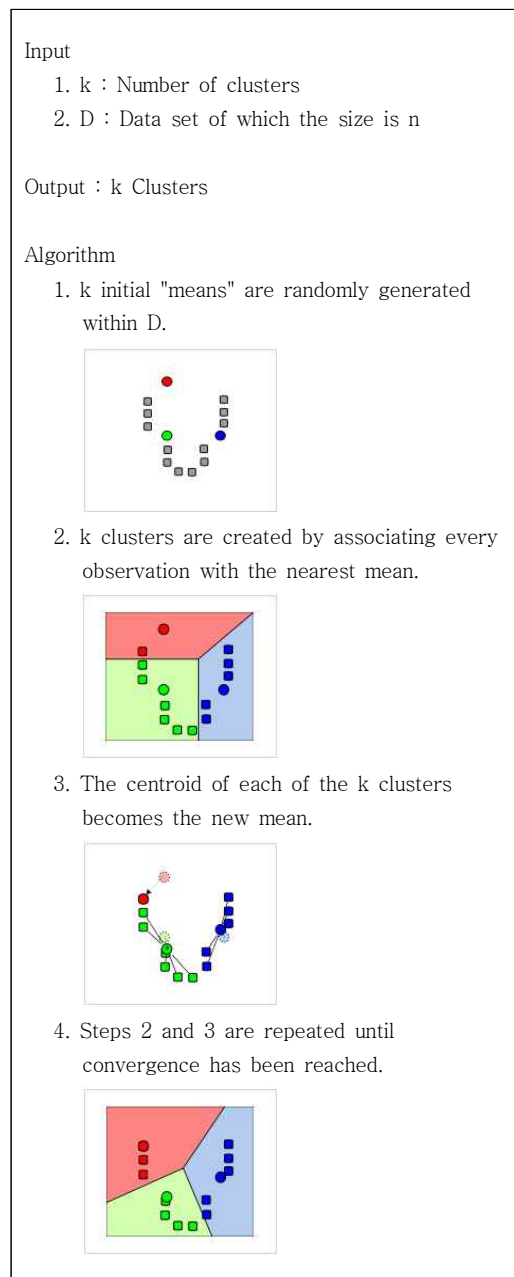


Fig. 1. k-means Algorithm[10]

III. The proposed algorithm

1. The Calculation of Similarity factor

본 논문에서는 페이스북 사용자의 프로필 데이터를 가지고 유사도를 구하고 유사 사용자를 군집화 한다. 페이스북 사용자의 프로필 정보는 수치형 또는 범주형 등 다양한 형태의 혼합형 데이터로 이루어져 있다. Gower 유사도 계수는 혼합형 데이터의 유사도 계산에 적합하지만 범주형 데이터의 유사도 계산 방법으로 완전매칭법을 사용하기 때문에 두 개의 데이터가 “같다” 혹은 “아니다”로 극단적으로 표현된다. 본 논문에서는 구조적 관계에 있는 범주형 데이터를 그림 2와 같이 다단계 범주형 데이터 형태로 변환하여 부분매칭을 할 수 있게 하였다. 그림 2에서 완전매칭법의 경우 서울과 인천 사이의 유사도나 서울과 부산 사이의 유사도는 모두 0으로 계산되지만 부분매칭법을 사용하게 되면 서울과 인천은 수도권이라는 같은 범주의 하위레벨이므로 부산보다 좀 더 유사한 관계가 된다.

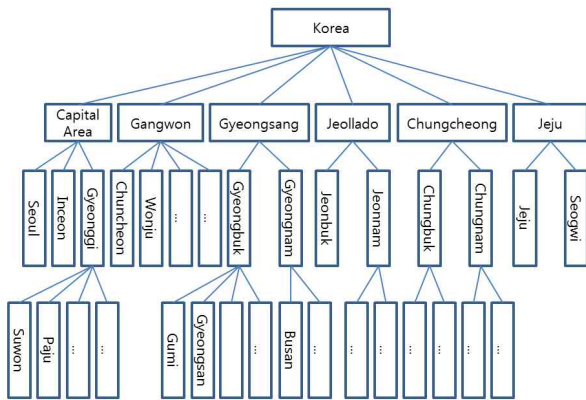


Fig. 2. Structure of Multi-stage Categorical Data

그림 2에서 노드와 노드 사이를 잇는 간선의 길이는 1로 정의하고 유사도는 식 3과 같이 정의한다.

$$S_{ijm} = 1 - distance \quad (3)$$

$$distance = \frac{length\ between\ nodes}{\max(length\ between\ nodes)} \quad (4)$$

예를 들어 그림 2에서 거주지가 서울인 사람끼리의 거리는 0이고 노드간 거리의 최대값은 6이므로 서울과 인천의 거리는 0.333, 서울과 수원위의 거리는 0.5이다. 따라서 서울과 서울의 유사도는 1, 서울과 인천의 유사도는 0.666, 서울과 수원의 유사도는 0.5이다.

그런데 그림 2에서 서울과 수원의 거리는 0.5인데 범주가 전혀 다르다고 할 수 있는 수도권과 강원 지역의 거리는 0.333으로 같은 범주내의 항목보다 거리가 더 가깝게 계산된다. 이런 경

우가 발생하는 것을 막기 위해 루트 노드를 지나는 거리는 1이라고 정의한다.

이와 같이 범주형 데이터를 완전매칭이 아닌 부분매칭방법으로 변환하여 계산한 후 Gower 유사도 계수 식 1의 S_{ijm} 에 대입하면 부울대수 형태의 극단적인 유사도를 좀 더 세분화된 유사도로 표현할 수 있다.

Table 1. Example Data for Calculation of Gower's Similarity Coefficient

User	Age	Gender	School	Residence	Hometown
...
I	35	1	Seoul	Jeju	Jeju
...
J	26	1	Deagu	Seogwi	Gyungsan
...

표 1의 데이터 예시에서 제안한 방식으로 Gower 유사도 계수를 구하면 식 5와 같이 계산되어 진다.

$$S_{ij} = \frac{\sum_{m=1}^{10} w_{ijm} S_{ijm}}{\sum_{m=1}^{10} w_{ijm}} \quad (5)$$

$$= \frac{0.888 + 1 + 0 + 0.666 + 0}{1 + 1 + 1 + 1 + 1}$$

$$= \frac{2.554}{5} = 0.5108$$

만일 학교, 거주지, 고향에 대한 데이터를 다단계 범주형 데이터 구조로 변환하지 않고 완전매칭법으로 Gower 유사도 계수를 구한다면 사용자 I와 사용자 J의 유사도는 0.3776으로 부분매칭법과 비교하여 상대적으로 낮게 구하여진다. 따라서 범주형 데이터를 구조화시켜 부분매칭법을 사용하는 것이 더 정확한 사용자간 유사도를 보여준다고 할 수 있다.

2. Clustering

k-means 알고리즘은 일반적으로 유클리안(euclidean) 거리를 이용해 중심과의 유사도(거리)를 구하지만 본 논문에서는 개선된 Gower 유사도 계수를 사용하여 유사도를 계산한다. 처음 k개의 중심이 선택되면 선택된 k개의 데이터와 모든 데이터와의 유사도를 계산하고 새로운 군집을 형성한다. 새로 형성된 군집의 중심은 수치형 데이터는 산술평균을 사용하고 범주형 데이터는 최빈값을 사용하여 만들어낸다. 새로 형성된 군집의 중심이 변하지 않을 때까지 반복 계산하는 것은 기존의 k-means 알고리즘과 같다.

IV. The result of the experiment

1. Data Collection and Preprocessing

실험을 위해 페이스북 사용자 100명의 프로필 정보를 수집

하였다. 수집된 데이터 중 결측치가 많은 32명의 사용자를 제거하고 68명의 데이터를 사용하였으며 데이터 타입에 따라 전처리 과정을 거쳤다. 연속형 데이터는 최소값, 최대값을 정하였고 범주형 데이터는 구조화시킬 수 있는 데이터에 한해 다단계 범주형으로 변환시켜 유사도를 세분화시켰다. 스포츠, 음악, 영화, TV, 책 등에 대한 선호도는 관련페이지를 좋아한 개수를 정규화(Normalization)하여 0~1까지의 연속형 변수로 변환시켜 유사도를 구하였다.

Table 2. user profile field and data type collected from Facebook

Field	Data type
Age	Numerical Data(10~80)
Gender	Categorical Data
School	Multi-stage Categorical Data
Residence	Multi-stage Categorical Data
Hometown	Multi-stage Categorical Data
Sports	Numerical Data(0~1)
Music	Numerical Data(0~1)
Movie	Numerical Data(0~1)
TV	Numerical Data(0~1)
Book	Numerical Data(0~1)

1.1 Data Preprocessing

출신학교, 거주지, 고향은 구조화된 형태로 표현할 수 있으므로 그림 2에서와 같이 다단계 범주형 데이터로 변환하여 사용하였다. 스포츠, 음악, 영화, TV프로그램, 책에 대한 선호여부는 관련된 페이지를 좋아한 개수를 사용하여 선호도(0~1)로 나타내었다. 선호도는 사용자가 관련 페이지를 좋아한 개수를 평균값의 두 배로 나누어 구하였다.

Table 3. Preference about Sports

User	The number of pages the person likes about sports	Preference
A	7	$\frac{7}{10}$
B	5	$\frac{5}{10}$
C	8	$\frac{8}{10}$
D	5	$\frac{5}{10}$
E	16	1
F	0	0
G	2	$\frac{2}{10}$
H	4	$\frac{4}{10}$
I	0	0
J	3	$\frac{3}{10}$
합계	50	
평균	5	

일반적으로 데이터에서 선호도를 구하는 방법으로는 변수 값을 해당 데이터의 최대값으로 나누는 방법이 있다. 이 경우 최대값은 극단치를 제거하고 사용하는 것이 보통이다. 하지만

극단치의 기준이 애매하기 때문에 본 논문에서는 평균의 2배 이상인 값을 선호도 1이라 가정하였다. 예를 들어 표 2와 같이 특정사용자 A가 스포츠 관련 페이지를 좋아한 개수가 7개이고 모든 사용자들이 스포츠 관련 페이지를 좋아한 평균 개수가 5라면 평균의 2배인 10으로 나누어 A의 스포츠에 대한 선호도를 7/10로 계산하였다. 그리고 E의 경우와 같이 평균의 2배를 넘는 경우는 선호도를 1로 정의하였다.

Table 4. Preprocessing data before Clustering

User	Age	Gen ger	Scho ol	Resid ence	Hom etow n	Spo rts	Mus ic	Mov ie	TV	Boo k
...
12	36	M	Gyeo nggi	Seoul	Gyeo nggi	0.5	0.5	0.4	0.3	0.1
13	32	M	Jeju	Jeju	Jeju	0	0.2	0	0	0.2
...
15	30	F	Jeju	Jeju	Jeju	0.1	0	0.2	0.1	0.4
...
19	33	M	Jeju	Jeju	Jeju	0.1	0	0.1	0.1	0.1
20	31	F	Jeju	Jeju	Jeju	0	0.3	0	0	0.5
...
24	34	F	Jeju	Jeju	Jeju	0.1	0	0.1	0.5	0.1
...
35	37	F	Seoul	Jeju	Jinju	0	0	0.1	0.1	0.1
...
37	35	F	Seoul	Jeju	Guns an	0.1	0.4	0.2	0.1	0.5
...
39	37	F	Gyeo ngbu k	Jeju	Jeju	0.3	0.5	0.2	0.5	0.7
40	33	F	Jeju	Gwa ngju	Deag u	0.2	0.2	0.5	0.4	0.6
...
52	23	F	Buch eon	Inche on	Seoul	0	0.3	0	0	0.5
...
63	27	F	Deag u	Deag u	Deag u	0	0.1	0.1	0.2	0
...

2. The Clustering results

표 5와 표 6는 각각 Gower 유사도 계수와 개선된 방식을 사용하여 15번 사용자를 중심으로 1차 군집화한 결과이다. 표 5는 학교, 거주지, 고향 항목의 변수가 표 6에 비하여 다양하다.

표 5는 학교, 거주지, 고향 항목을 다단계 범주형으로 변환하지 않고 비교대상이 같으면 1, 다르면 0을 대입하여 유사도를 계산하였기 때문에 표 6의 방식에 비해 학교, 거주지, 고향이 유사하지 않다고 판단되는 데이터가 훨씬 많았다. 그래서 나머지 항목이 유사하다면 군집화되는 경우가 많아 다른 항목에 비해 범주형 데이터가 군집에 미치는 영향이 작았다.

개선된 알고리즘으로 군집화를 진행한 결과 두 번의 중심이동 후 만들어진 군집에서 더 이상 중심의 변화가 없었다. 표 7은 군집화 완료 후 개선된 알고리즘을 사용한 군집과 Gower

유사도 계수를 이용한 군집간의 비교이다.

Table 5. preliminary Clustering using Gower's similarity Coefficient

User	Age	Gen	School	Residence	Home town	Sports	Music	Movie	TV	Book
15	30	F	Jeju	Jeju	Jeju	0.1	0	0.2	0.1	0.4
20	31	F	Jeju	Jeju	Jeju	0	0.3	0	0	0.5
23	33	F	Gyeonggi	Seoul	Bucheon	0.2	0.7	0.3	0.2	0.3
35	37	F	Seoul	Jeju	Jinju	0	0	0.1	0.1	0.1
37	35	F	Seoul	Jeju	Gunsan	0.1	0.4	0.2	0.1	0.5
39	37	F	Gyeongbuk	Jeju	Jeju	0.3	0.5	0.2	0.5	0.7
40	33	F	Jeju	Gwangju	Daegu	0.2	0.2	0.5	0.4	0.6
52	23	F	Bucheon	Incheon	Seoul	0	0.3	0	0	0.5
63	27	F	Daegu	Daegu	Daegu	0	0.1	0.1	0.2	0
67	27	M	Gyeonggi	Gyeonggi	Gyeonggi	0.2	0.4	0	0.1	0.4
new centroid	31	F	Jeju	Jeju	Jeju	0.1	0.3	0.2	0.2	0.4

Table 6. preliminary Clustering using advanced method

User	Age	Gen	School	Residence	Home town	Sports	Music	Movie	TV	Book
15	30	F	Jeju	Jeju	Jeju	0.1	0	0.2	0.1	0.4
20	31	F	Jeju	Jeju	Jeju	0	0.3	0	0	0.5
24	34	F	Jeju	Jeju	Jeju	0.1	0	0.1	0.5	0.1
35	37	F	Seoul	Jeju	Jinju	0	0	0.1	0.1	0.1
37	35	F	Seoul	Jeju	Gunsan	0.1	0.4	0.2	0.1	0.5
39	37	F	Gyeongbuk	Jeju	Jeju	0.3	0.5	0.2	0.5	0.7
new centroid	34	F	Jeju	Jeju	Jeju	0.1	0.2	0.1	0.2	0.4

Table 7. Result of Clustering

	Gower's similarity	Advanced method
Cluster 1	13,15,19,20,24,35,37,39,40,52,63	13,15,19,20,24,35,37,39,40
Cluster 2	1,3,5,6,7,11,16,18,28,30,50	1,3,5,6,7,16,18,28,30
Cluster 3	2,4,8,9,10,12,14,17,21,22,23,25,27,31,32,33,34,36,38,41,42,43,44,45,46,47,48,49,53,64	2,4,8,9,10,12,14,17,21,22,23,25,27,29,31,32,33,34,36,38,41,42,43,44,45,46,47,48,49,50,51,52,53,64,65,67,68
Cluster 4	26,29,51,54,55,56,57,58,59,60,61,62,65,66,67,68	11,26,54,55,56,57,58,59,60,61,62,63,66

표 7에서 군집 1을 보면 경우 상당히 비슷한 결과를 보여준다. 그 중 다른 결과를 만드는 52번 사용자와 63번 사용자의 데이터를 표 4에서 확인해보면 52번은 수도권지역 사용자이고 63번은 대구지역 사용자임을 알 수 있다. 군집1에 소속된 나머지 사용자의 특징을 살펴보면 전부 학교, 거주지, 고향 항목이 제주와 관련성이 높은 사람들의 집합이다. 따라서 개선된 방식이 더 정확한 군집을 만들어 준다는 것을 알 수 있다. 나머지 군집에서도 다른 결과를 보여주는 사용자를 살펴보면

Gower 유사도 계수를 사용한 방식은 학교, 거주지, 고향 항목에서 전혀 관련성이 없는 사용자를 유사하다고 평가하고 있는 것을 알 수 있다. 즉 극단값으로 구성된 범주형 데이터는 유사도 측정에 좋지 않은 영향을 끼친다고 할 수 있다.

V. Conclusion

인터넷이나 SNS에 존재하는 데이터들은 여러 형태의 혼합형 데이터로 이루어져 있는 경우가 많다. 본 논문에서는 혼합형 데이터의 유사도 계산이 가능한 Gower 유사도 계수를 사용하였으나 구조화가 가능한 범주형 데이터는 트리 형태의 다단계 범주형 데이터로 변환하여 부분매칭법에 의하여 유사도를 구하였고 k-means 알고리즘을 이용하여 군집화 하였다. 제한된 방식은 극단적인 유사도를 세분화하여 정확성을 높이는 효과가 있다. 이는 사용자 군집 및 추천이 필요한 다양한 시스템에서 활용가능할 것이라 예상된다. 본 논문에서는 혼합형 데이터 군집의 중심값을 구하기 위해 수치형 변수는 산술 평균을 중심으로 하였고, 범주형 변수는 최빈값을 중심으로 설정하였다. 그러나 최빈값이 전체 군집에서 차지하는 비중이 낮다면 중심값으로써 의미가 약해진다. 향후에는 범주형 데이터의 중심값을 구하는 방법에 대한 추가적인 연구가 필요하다.

REFERENCE

- [1] "The Digital Universe of Opportunities", EMC & IDC, 2014
- [2] Soojin Lee, Taeryong Jeon, Gyeongdong Baek, Sungshin Kim, "A Movie Rating Prediction System of User Propensity Analysis based on Collaborative Filtering and Fussy System", Journal of Intelligence and Information System, Vol. 19, No. 2, pp. 242-247, 2009.
- [3] Yeo-Kwang Yoon, "A Study on Contents Curation of Portal Sites", Journal of the Korea Entertainment Industry Association(JKEIA), Vol. 8, No. 4, pp. 31-43, Dec. 2014.
- [4] Ahn Hyung Jun, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem", Information Sciences, Vol. 178, No. 1, pp.37-51, 2007.
- [5] Hyeongdo Kim, "Collaborative Tag-Based Recommendation Methods Using the Principle of Latent Factor Models", Journal of Society for e-Business Studies, Vol. 14, No. 4, pp.47-57, Nov. 2009.

- [6] Hyeong-Joon Kwon, Kwang-Seok Hong, "Personalization of LBS using Recommender Systems Based on Collaborative Filtering", Journal of Korean Society for Internet Information Vol. 11 No. 6, pp.1-11, Dec. 2010.
- [7] Galit Shmueli, Nitin R. Patel, Peter C. Bruce, "Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", Wiley, p.299, 2006.
- [8] Dongjin Park, Ingeuk Hwang, Teahun Ann, "A Clustering Algorithm for Categorical Data Oriented Database", Korea Institute of Industrial Engineers Autumn Conference, pp.355-362, Oct. 1998.
- [9] Shinwon Lee, "Comparison of Initial Seeds Methods for K-Means Clustering", Journal of Korean Society for Internet Information, Vol.13, No.6, pp.1-8, Dec. 2012.
- [10] https://en.wikipedia.org/wiki/K-means_clustering

Authors



Hyoung-Min Song is enrolled the M.S. degrees in Computer Engineering from Jeju Nat'l University, Korea, in 2015. He is interested in computing algorithm.



Sang-Joon Lee received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Chung-Ang University, Korea, in 1984, 1989 and 1992, respectively. Dr. Lee joined the faculty of the Department of Computer Engineering at Jeju National University, Jeju, Korea, in 1992. He is currently a Professor in the Department of Computer Engineering, Jeju National University. He is interested in intelligent system, computer algorithm.



Ho-Young Kwak received the B.S., M.S. and Ph.D. degrees in Computer Science from Hong-Ik University, Korea, in 1983, 1985 and 1990, respectively. Dr. Kwak joined the faculty of the Department of Computer Engineering at Jeju National University, Jeju, Korea, in 1990. He is currently a Professor in the Department of Computer Engineering, Jeju National University. He is interested in IT-Medical convergence, USN, software system.