

# 클러스터링 및 랭킹 기법을 활용한 트위터 인플루엔셜 추출 연구<sup>†</sup>

(A study on finding influential twitter users by  
clustering and ranking techniques)

최준일<sup>1)</sup>, 장중혁<sup>2)\*</sup>  
(Jun-Il Choi and Joong-Hyuk Chang)

**요약** 최근 SNS 서비스의 확산과 스마트 장치의 일반적인 보급에 따라 수많은 사용자들이 트위터를 비롯한 SNS를 이용하고 있다. 본 논문에서는 영향력 높은 트위터 사용자를 찾기 위한 방법으로 클러스터링 및 랭킹 기법을 적용하는 연구를 수행하였다. 트위터에서 일반적으로 중요하다고 여겨지는 follow, Retweet을 포함한 총 5가지의 랭킹 요소를 제안하고 해당 랭킹 요소들을 클러스터링을 위한 초기 중심점 선정 시 기준으로 설정함으로써 클러스터링 결과의 향상을 위한 비교 실험을 수행하였다. 이를 통해 트위터 인플루엔셜 추출의 새로운 방향을 제시할 수 있을 것으로 사료된다.

**핵심주제어** : SNS, 트위터, 인플루엔셜, 클러스터링, 랭킹

**Abstract** Recently, a lot of users are using social network services as the spread of SNS and generalization of smart-phone. In this study, we apply clustering and ranking method for finding twitter influential users. First, we propose five ranking elements. The five elements include the number of follow, the number of retweet, IRP, IFP and influ-score. These elements are used by centroid point of clustering methods. This study can help to find novel approaches for finding twitter influential users.

**Key Words** : SNS, Twitter, Influential, Clustering, Ranking

## 1. 서론

\* Corresponding Author : jhchang@daegu.ac.kr

† 이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2012R1A1B4000651)

Manuscript October 14, 2014 / Revised December 10, 2014 / Accepted December 16, 2014

1) 다이나맥스, 제1저자, 본 논문은 대구대학교 석사과정 재학 중 수행된 연구결과임

2) 대구대학교 컴퓨터IT공학부, 교신저자

최근 급속도로 발전하고 있는 스마트폰, 태블릿 PC 시장의 확대와 이로 인한 SNS의 확산이 가속화 되고 있다. SNS는 기존의 다양한 연구에서 밝혀진 바와 같이 강력한 파급효과를 가지고 있다.

예를 들어 사건 사고에 대한 증계, 투표 시 여론 조사, 고객 서비스를 위한 고객과의 소통 등 다양한 분야에서 SNS 서비스를 이용하고 있다[1,2,3]. 그러나 과연 SNS의 정보들이 과연 정확한가? 에 대한 의구심을 가

질 필요가 있다. 수도 없이 산재하는 다양한 데이터 속에서 과연 어떤 정보가 진실된 정보인지 믿을만한 정보인지 판단하는 것은 사용자 입장에서 매우 중요하다. 만약, 특정한 사용자가 보편적으로 신뢰성이 있다고 판단된다면 이러한 사용자가 제공하는 정보 또한 신뢰할 수 있다고 판단하기에 용이할 것이다.

이러한 신뢰성 있는 사용자, 혹은 영향력이 있는 사용자들을 인플루엔셜이라고 하는데 여기서 영향력의 의미는 SNS 뿐만 아니라 사회학적인 관점에서 보았을 때 특정한 군집 내에서 다른 사람들에게 과급효과를 미칠 수 있는 사람들을 의미한다. 트위터에서는 보통 과급 효과가 큰 사용자들은 유명한 인물(가수, 연예인, 정치인)인 경우가 대다수이다[4]. 그러나 이러한 유명도만으로는 특정 사용자가 인플루엔셜인지 판단하기에는 사용자의 주변 환경이나 접하는 매체에 따라 변수가 발생하기 마련이다.

따라서 본 논문에서는 SNS 서비스 중에서도 보편적으로 사용되는 트위터를 주제로 선정하고 트위터 사용자들 중에서도 인플루엔셜 추출을 위해 클러스터링 기법과 랭킹 함수들을 적용하여 사용자 중심의 인플루엔셜 추출을 위한 연구를 수행하였다.

본 논문의 구성은 다음과 같다. 2장에서는 트위터와 관련된 컴퓨터 공학 중심의 연구들과 관련 정보에 대해서 살펴보고 3장에서는 전체적인 연구 체계의 설계 및 구현에 대해서 살펴본다. 4장에서는 실험을 수행하여 제안하는 기법에 대해서 평가하며 5장에서 결론을 맺는다.

## 2. 관련 연구

### 2.1 클러스터링 알고리즘

클러스터링 알고리즘은 유사한 객체들끼리 그룹화하기 위한 기법들을 의미한다. 본 논문에서는 이러한 기법 중에서도 K-means, K-medoids를 사용하는데 문서 클러스터링 방식으로 연구를 진행하였다.

#### 2.1.1 K-means Clustering Method

간단하다는 장점을 가진 K-means 알고리즘은 가장 잘 알려진 기법이기도 하다. 단점으로는 초기 중심점의

설정이 필요한 점, 이상치에 민감한 점 등이 있다.[5]

그러나 K-means 기법이 가지는 간편성과 빠른 속도 때문에 다양한 연구에서 많이 이용되고 있다.

K-means 기법의 수행절차는 Table 1과 같다.

Table 1 The overall process of the K-means

입력 : k, D(데이터 집합) 출력 : k 개의 클러스터 절차 : 1. D로부터 임의로 k개의 객체들을 초기 군집 중심점으로 선정 2. 클러스터 내 객체들의 평균값을 계산하여 가장 근접한 객체들을 클러스터에 (재)할당 3. 클러스터 평균값을 재계산 4. 클러스터 결과가 바뀌지 않거나 임계값 이하가 될 때까지 2~3 과정을 반복
--

#### 2.1.2 K-medoids Clustering Method

K-means와 유사한 기법이나 K-means의 차기 중심점이 평균값으로 계산되는 점과 달리 차기 중심점이 실제 객체로 선정된다는 점이 가장 큰 차이점이다. 알고리즘의 특성 때문에 K-means에 비해 계산 비용이 많이 소모된다는 단점이 있다.[5]

K-medoids 기법의 수행절차는 Table 2와 같다.

Table 2 The overall process of the K-medoids

입력 : k, D(데이터 집합) 출력 : k 개의 클러스터 1. 초기화 : n개의 객체 중 임의로 k개의 중심점 지정 2. 가장 가까운 중심점에 각각의 객체를 할당 3. 각각의 중심점 m에 대해서 3.1. 각각의 중심점이 아닌 객체 O를 임의 지정 3.2. m와 O에 대해 각각의 총 비용을 계산/비교 4. m, O 중에 저비용 객체를 차기 중심점 지정 5. 중심점이 변하지 않을 때까지 2~4 과정을 반복
---

#### 2.1.3 계층형 클러스터링

본 논문에서 직접적으로 계층형 클러스터링을 사용

하지는 않지만 K-means, K-medoids의 수행 결과로 도출되는 클러스터간의 거리 계산 및 비교를 위해 계층형 클러스터링의 거리 측정 개념이 사용되었다. 계층형 클러스터링에서 사용되는 클러스터간 거리 측정 방법은 다음의 Table 3과 같다.

클러스터 간 유사도 측정을 위해 3가지 방법 중에서도 Maximum Distance를 사용하여 최장 거리에 위치한 객체들 간의 유사도를 측정하여 클러스터의 유사도를 계산하였다. 결과물로 도출된 클러스터 간의 유사도가 낮을수록 클러스터링 결과가 우수하다고 판단할 수 있으며 클러스터링의 근간이 되는 랭킹 요소 또한 적합한 랭킹 요소라고 판단할 수 있을 것이다.

Table 3 Distance measures for hierarchical clustering

Distance Measure	내용
Minimum Distance	최단 거리 객체들 간의 유사도 측정
Maximum Distance	최장 거리 객체들 간의 유사도 측정
Average Distance	객체들간 평균을 도출하여 유사도 측정

### 2.2 트위터 관련 연구

Yizhou Sun[6]의 연구는 본 논문의 근간이 되는 연구로 데이터마이닝/데이터베이스와 하드웨어/컴퓨터 구조라는 2가지 영역으로 분류되는 20개의 학회를 기본 데이터 셋으로 선정하고 100명의 저자들을 각 분야에 맞게 클러스터링하고자 하였다. 향상된 결과를 위해서 데이터 셋에 대해 랭킹 기법을 그대로 적용하는 것보다 클러스터링을 수행한 후 랭킹 함수를 적용하게 되면 순위 결과가 보다 사실에 가까워짐을 확인하였다. 이와 같은 기본적인 아이디어에 입각하여 본 논문에서도 유사하게 트위터 데이터 셋에 대해 클러스터링을 수행한 뒤 결과 데이터 셋에 대해 랭킹 기법을 적용하고자 하였다.

국내 연구에서는 인플루엔셜과 관련하여 [1]의 연구가 가장 많이 참조되었는데 해당 논문에서는 사회과학, 컴퓨터 과학 분야 등 연구의 방향에 따라 다양한 관점의 인플루엔셜을 탐색할 수 있다고 하였다.

[7]의 연구에서는 공통 관심사를 가진 사용자 군집을

탐색하기 위해서 트윗 분류를 위해 18가지 범주로 선별하여 할당한 뒤 K-means, MCL 클러스터링 기법을 적용하여 결과와 분류된 정보를 비교하여 평가하였다. 이러한 논문과의 비교를 통해 본 연구에서는 선별된 정보와 같이 감독 학습목적이 아닌 비감독 학습이 목적으로 결과적으로 정확도와 재현율을 이용한 비교 분석이 불가능하다는 차이점을 확인할 수 있었다.

[8]의 연구에서는 follow가 다른 트위터에서 중요하게 여겨지는 요소들에 비해서 얼마나 많은 영향력을 미치는지에 대한 연구를 수행하였다. 분석 요소로는 follow, retweets, mention과 같이 3가지를 비교분석하였다. 실험 결과로 follow는 사용자의 유명도를 표현하는 데에는 적절하지만 retweets, mentions과 같이 대중적으로 사용자들과 관련되는 개념과는 동떨어진 것을 확인하였다. 즉, 이러한 차이점 때문에 follow를 단독적으로 사용하는 경우 유사한 사용자들을 군집화 하는 것에는 한계점이 발생할 수 있음을 확인하였다.

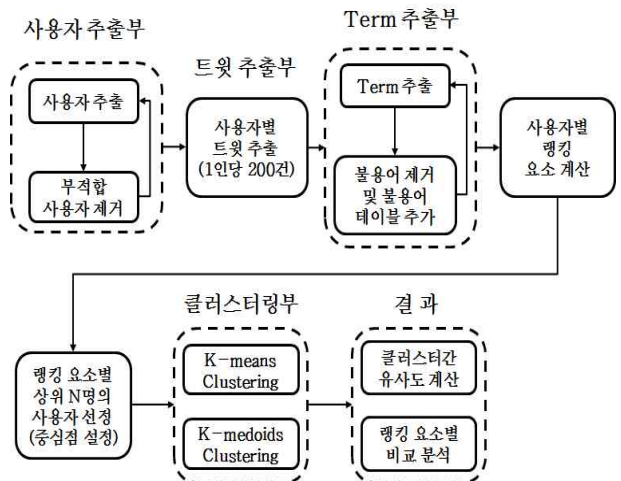


Fig. 1 The overall system architecture

### 3. 제안하는 클러스터링 및 랭킹 함수

본 연구에서 제안하는 시스템의 전체적인 구조는 Fig. 1과 같다. 시작은 트위터 사용자들의 추출로부터 시작하여 실험 데이터 셋에 부적합한 사용자(비영어권 사용자, 저빈도 트윗 사용자)들을 추출된 목록에서 제거한다. 다음으로 추출된 선별 인원에 대해서 1인당 200건씩의 트윗을 추출하였다. 추출된 트윗에 대해 사

용 용어들을 추출하고 불용어 및 금칙어를 제거하는 작업을 반복적으로 수행하였다. 이와 같이 구성된 데이터셋을 기본으로 하여 랭킹 및 클러스터링 기법을 적용한다. 우선 사용자별 랭킹 요소들을 계산하고 랭킹 요소별로 상위 n 명의 사용자들을 추출하여 클러스터링 기법의 중심점으로 설정한다. 클러스터링 기법의 방법으로는 상기 기술한 바와 같이 K-means, K-medoids 방법을 사용하고 클러스터링 결과 평가를 위해 클러스터간 유사도 계산과 랭킹 요소별 비교분석을 수행하였다.

### 3.1 데이터 셋 전처리 과정

기본 데이터 테이블은 크게 사용자 정보, 개인 트윗, 용어 테이블로 구성된다. 세부적인 데이터 셋 구축과정은 다음의 Fig. 2와 같다.

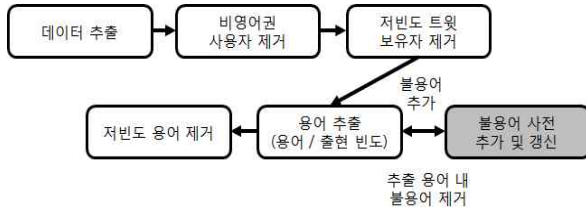


Fig. 2 Constructing data set

### 3.2 랭킹 요소의 선정 및 활용

랭킹 요소를 선정하기 위해서 트위터에서 중요하다고 판단되는 요소들을 활용한 5가지의 랭킹 요소를 선정하였다. 요소들의 종류와 특성은 다음의 Table 4와 같다.

Table 4 Features for ranking

요소	특성
follow	사용자별 follow 수를 의미
Retweet	추출된 트윗 중에서 Retweet이 된 트윗 수(한 트윗에 대해 중복된 Retweet은 미적용)
IFP	follow로부터 도출되는 함수
IRP	Retweet으로부터 도출되는 함수
Influ Score	IFP, IRP를 조합한 요소 (가중치 부여)

$$\text{InfluentialFollower}(p) = \frac{\text{Follower}_p}{\sum_1^{N_{\text{public}}} \text{Follower}_i}$$

Fig. 3 Evaluation of the IFP

follow와 Retweet은 단순한 사용자들이 가지는 수치를 의미하며 IFP는 Fig. 3과 같다. 또한, IRP와 Influ Score 의 계산은 다음과 같다.

$$\text{InfluentialRetweet}(p) = \frac{\text{tweet}_{RT-p}}{\sum_1^{N_p} \text{tweet}_p} / \frac{\text{tweet}_{RT-N_{\text{public}}}}{\sum_1^{N_{\text{public}}} \text{tweet}_{N_{\text{public}}}}$$

Fig. 4 Evaluation of the IRP

$$\text{Influ\_Score}(p) = \text{InfluentialFollower}(p) \times \sigma_1 + \text{InfluentialRetweet}(p) \times \sigma_2$$

Fig. 5 Evaluation of the Influ-Score

여기서 IFP는 [6]의 논문에서 사용되는 follow 인플루엔셜 수치를 참조하였다. 본 논문에서 해당 요소를 IFP로 표현하였고 특정 사용자의 follow 수 / 전체 사용자의 follow 합으로 계산한다.

IRP에서 좌항은 tweetRT-p(특정 사용자의 Retweet 받은 횟수) / 특정 사용자 트윗 전체 개수를 의미하며 우항은 tweetRT-Npublic(전체 사용자의 Retweet 받은 횟수) / 전체 사용자의 트윗의 합을 의미한다. IRP의 결과값은 좌항과 우항을 나눔으로써 평균화된다.

Influ Score는 IFP와 IRP에 가중치를 부여하여 이러한 가중치 변화를 통해서 최적의 가중치 비율을 찾아내 고자 하였다. 초기 가중치는 각각 0.5로 설정하였다.

이와 같이 선정된 랭킹 요소들은 클러스터링 시에 필요한 초기 중심점들을 선정할 때 사용된다. 랭킹 요소를 이용한 점수가 높은 사용자들을 n명 추출하여 중심점으로 선정하여 클러스터링을 수행한다.

### 3.3 클러스터링 기법의 적용

#### 3.3.1 유사도 계산

본 논문에서는 문서 클러스터링 기법과 동일하게 적용하여 사용자 1인을 단일 문서로 간주한다. 문서들 간의 유사도 계산을 위해 Cosine Similarity를 사용하였다.

Cosine 유사도를 사용하여 Table 5와 같이 표현 가능한 트위터 사용자들의 단어 벡터들을 계산한다. 결과 값으로는 1에 가까울수록 유사하고 0에 가까울수록 유사하지 않음을 의미한다.

Table 5 Vectors for twitter user's words

User \ Term	Gabriela Frias	Colleen Johnson	bbynygirl
person	0	0	4
picture	5	0	0
Diamond	0	0	0
Olympic	0	0	0
business	0	0	0
Storm	0	0	0

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Fig. 6 Cosine Similarity

### 3.3.2 K-means, K-medoids의 적용

K-means 클러스터링은 차기 중심점 선정 시 객체들 간의 평균을 계산하는 방식이다. 문서 클러스터링 방식에서는 주로 단어 벡터를 사용하므로 차기 중심점 또한 벡터들의 평균을 계산하는 방식을 사용한다.

K-medoids는 차기 중심점 선정 방식이 실제 객체 선정 방식이므로 평균 벡터를 도출할 필요 없이 랭킹 요소로 선별된 사용자의 단어 벡터를 그대로 사용한다.

Table 6 An example of the average vector

사용자 \ 단어	bit	state	office	viking
A	0	1	2	0
B	3	4	2	1
C	1	2	4	0
평균	1.33	2.33	2.67	0.33

## 4. 실험 및 평가

본 논문에서 제시한 인플루엔셜 추출 기법의 유용성을 검증하기 위해서 실제 트위터 사용자의 트윗을 대상으로 실험을 진행하였다. 실험에 사용된 데이터 집합은 최대 400명의 트위터 사용자의 트윗으로 수집된 것으로서, 각 트윗 내용을 대상으로 기존의 일반적인 불용어 처리 과정 등을 통해 얻어진 핵심어 집합을 구하고 이를 실험 데이터 집합으로 활용하였다.

### 4.1 실험 순서

실험 순서는 다음과 같다. 인플루엔셜 추출을 위해 우선적으로 클러스터링 알고리즘들에 대해서 동일한 데이터 셋에 대해 반복 수행을 진행한다.

클러스터링에 사용될 중심점 선정은 5가지 랭킹 요소 기준으로 선별된 상위 n명의 사용자들이다.

랭킹 요소들의 변화에 따른 상태 추이를 확인하기 위해 클러스터링의 대상이 될 인원수를 100, 200, 400명으로 조절하였고 각 항목에 대해 클러스터 개수를 3개에서 10개로 변화시키면서 추이를 파악하였다.

클러스터링 반복 횟수는 K-means는 10회, K-medoids는 100회 반복수행을 하였다.

K-medoids는 최소 20회 이상의 클러스터링 연산을 수행하고 변화임계값으로 클러스터 개수 - 1번의 변화율을 가질 때 정지하도록 설정하였다.

추가 실험으로 Influ Score의 구성이 되는 IFP와 IRP의 가중치 변화 실험 또한 진행하였다.

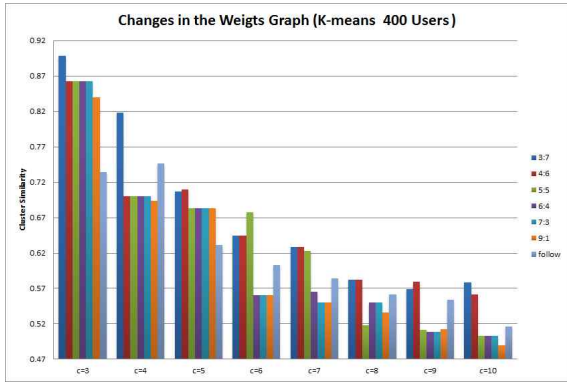


Fig. 7 Similarity between clusters in the Influ score

## 4.2 실험 결과 및 분석

### 4.2.1 Influ Score 가중치 변화 실험

본 논문에서 제안하는 요소 중 하나인 Influ Score의 유효성을 증명하기 위해 IFP 와 IRP의 가중치를 달리 하여 클러스터링 기법에 적용하여 비교적 우수하다고 기존 연구해서 판단된 follow와 IFP, Influ Score를 상호 비교하였다. 가중치는 3:7, 4:6, 5:5, 6:4, 7:3, 9:1과 같이 다양한 비율로 가중치를 부여하고 클러스터링 알고리즘에 적용하였다.

Fig. 7에서 확인 가능하듯이 IFP의 비중이 50% 이상이고 클러스터의 개수가 많아질수록 Influ Score의 점수가 나은 성능을 보여줄을 확인할 수 있다. follow는 일반적으로 사용자들 간의 인맥 관계를 위해 존재하는

개념으로 특정 1인이 다른 사람에게 신청을 하여 예의 상으로 맺어주는 관계가 존재한다. 이러한 경우 실제적인 교류가 없다면 이는 무의미한 인맥 관계가 될 수 있으며 [9]의 연구에서는 최소한 이러한 관계에서 2개 이상의 Reply 혹은 Mention과 같은 실제적인 교류가 존재하는 경우에만 해당 인맥 관계가 의미 있는 관계라는 의견을 제시하였다.

이러한 특성에 근거하여 단순히 follow 수가 높다는 사실보다는 Retweet과 같은 실제적 교류 관계가 반영된 개념을 적절하게 반영한 경우에 클러스터링 시 더 나은 클러스터링 결과를 이끌어 낼 수 있음을 확인할 수 있었다.

결과로 IFP, IRP 각각에 대해 9:1비율의 Influ Score가 가장 우수한 점수를 도출하여 클러스터링 실험 시도에서는 초기값으로 해당 비율을 사용하였다.

### 4.2.2 클러스터링 결과

K-means 클러스터링 실험 결과로 무작위 선별 인원들을 중심으로 선정한 경우에 비교하여 다른 랭킹 요소들이 비교적 안정된 하강 곡선 및 평균 유사도 또한 낮은 양상을 보여 무작위 선별 방법에 비교하여 유효성이 있음을 증명하였다. 랭킹 요소별 비교 측면에서는 follow와 IFP, Influ Score가 비교적 높은 비유사도를 가졌고 100, 200, 400명과 같이 인원수를 달리한 데이터 셋에서 Influ Score가 가장 나은 성능을 보여주었다.

K-medoids 실험에서는 실험 인원에 따라 상이한 양

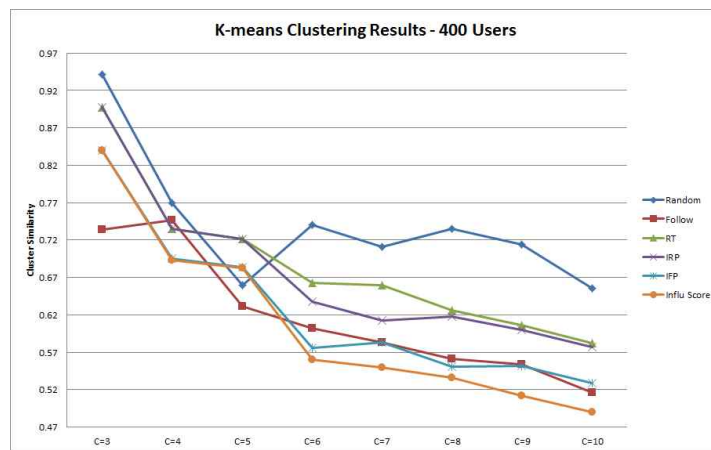


Fig. 8 Clustering results of the K-means (for 400 users)

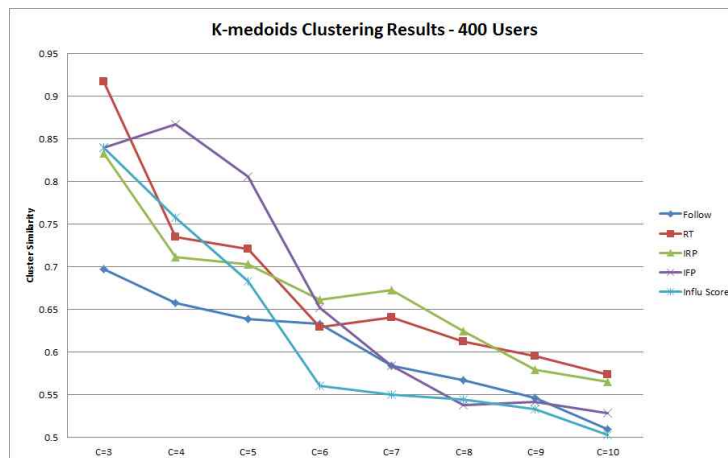


Fig. 9 Clustering results of the K-medoids (for 400 users)

상이 나타났는데 100명인 경우 Retweet, 200명인 경우 follow, 400명의 경우 Influ Score가 높은 성능을 보여 주었다. 이와 같이 안정적이지 못한 결과 데이터는 K-medoids 클러스터링 기법에서 차기 중심점 선정을 위해 기분류 된 군집 내에서 무작위 중심점 선정 방식을 사용하기 때문에 가장 최적의 중심점 선정을 위해 많은 계산 비용이 요구되는 한계점 때문인 것으로 판단 된다.

설 추출에 있어 부적절하다는 점을 확인할 수 있었다. 이와 같은 연구를 통해 사용자의 요구와 연구 목적에 맞는 다양한 인플루엔셜 추출 연구에 대해 하나의 방향을 제시할 것으로 사료된다. 향후 연구 과제로는 선정된 트위터 랭킹 요소들에 대해서 추가적으로 사용자들 간의 관계와 실제적인 사용자간의 교류를 증명할 수 있는 Reply, Mention을 활용한 링크 정보를 고려한 트위터 분석을 수행할 예정이다.

## 5. 결 론

본 연구에서는 하루에도 수천 건 이상의 정보가 생산되는 SNS의 분석을 통해 영향력 있는 사용자인 인플루엔셜의 탐색을 위해 진행되었다. 인플루엔셜의 탐색을 통해 개인 사용자의 사회적 욕구 충족, 다양한 요구에 맞는 정보 습득뿐만 아니라 기업 입장에서 마케팅 전략 수립을 위한 수단으로 사용가능한 점 등의 이점을 가진다. 클러스터링의 중심점 선정을 위해 트위터에서 중요하게 여겨지는 요소들을 5가지 선정하였다. 이러한 5가지 요소들을 중점으로 K-means, K-medoids 클러스터링 기법에서 사용될 중심점들을 선정하여 클러스터링을 수행하였다. 실험의 결과로 follow가 Retweet보다 일반적으로 클러스터링에 많은 영향을 미친다는 것을 확인하였으나 follow 단독으로는 사용자들 간의 적절한 군집화와 인플루엔셜 추출을 하는데 있어 한계점이 있음을 확인하였다. 추가적으로 클러스터링 기법상의 한계 때문에 K-medoids 기법의 경우 인플루엔

## References

- [1] S. Y. Park, Y. H. Ha and Y. H. Kim, "Recent Studies on Twitter in the Field of Information Retrieval", Korea Institute of Information Scientists and Engineers, Vol 27, pp.25-29, 2010.
- [2] C. G. Yang, W. C. Hwang, and H. Baek, "A Study on Social Media Market Competition based on User Gratification", Journal of the Korea Society Industrial Information System, Vol. 19. No. 2, pp.105-117.
- [3] K. A. CHA, "A Development of SNS Application for Location based Information Sharing using Smartphone", Journal of the Korea Society Industrial Information System, Vol. 18. No. 6, pp.1-8.
- [4] H. S. Park, H. W. Guak, M. Y. Cha, S. B.



Moon, "Influentials Ranking in Social Networks", Korea Institute of Information Scientists and Engineers, Vol 28, No.3. pp.23-30, 2010.

[5] S. S. Singh and N. C. Chauhan, "K-means v/s K-medoids : A Comparative Study", National Conference on Recent Trends in Engineering & Technology, 2011.

[6] Yizhou Sun, Jiawei Han, Peixiang Zhao, Zhijun Yin, Hong Cheng, and Tianyi Wu, "RankClus: Integrating Clustering with Ranking for Heterogeneous Information Network Analysis", EDBT, 2009.

[7] S. H. Kim, D. J. Choi, J. G. Kim, H. W. Jung and J. H. Lee, "Discovering Twitter User Group with Common Interests by Tweets", Korea Intelligent Information System Society, pp.129-131, 2011.

[8] Cha. M, Haddadi. H, Benevenuto. F, & Gummadi, K. P., "Measuring User Influence in Twitter: The Million Follower Fallacy", In Proceedings of 4th International AAAI Conference on Weblogs and Social Media, pp.10-17, 2010.

[9] B. A. Huberman, D. M. Romero, & F. Wu., "Social networks that matter : Twitter under the microscope", arXiv:0812.1045v1, 2008.



최 준 일 (Jun-Il Choi)

- 정회원
- 대구대학교 컴퓨터IT공학과 공학사
- 대구대학교 컴퓨터정보공학과 공학석사
- 다이내믹스 연구원
- 관심분야 : 데이터베이스, 빅데이터, 트위터, 인플루엔셜, SNS



장 중 혁 (Joong-Hyuk Chang)

- 정회원
- 연세대학교 컴퓨터과학과 이학사
- 연세대학교 컴퓨터과학과 공학석사
- 연세대학교 컴퓨터과학과 공학박사
- 대구대학교 정보통신대학 컴퓨터IT공학부 교수
- 관심분야 : 데이터공학, 데이터베이스, 데이터 스트림, 빅데이터, 인플루엔셜, SNS