

# K-평균 군집을 이용한 마이크로타겟팅을 위한 SNS 빅데이터 활용 모델링에 관한 연구

송재오<sup>o</sup>, 이상문<sup>\*</sup>

<sup>o</sup>한국교통대학교 컴퓨터정보공학과

e-mail: jeos@icomtool.com<sup>o</sup>, smlee@ut.ac.kr<sup>\*</sup>

## A Study on the Application Modeling of SNS Big-data for a Micro-Targeting using K-Means Clustering

Jeon Song<sup>o</sup>, Sang Moon Lee<sup>\*</sup>

<sup>o</sup>Dept. of Computer Science & Information Engineering, Korea Nat'l Univ. of Transportation

### ● 요약 ●

본 논문에서는 SNS에 존재하는 특정 제품과 브랜드 또는 기업에 대한 평가, 의견, 느낌, 사용 후기 등의 소비자 생각을 수집하여 기업에서 향후 신제품 개발이나 시장 진출 및 확대 등의 경영활동에 활용할 수 있도록 SNS 빅데이터를 분석하고, 이를 활용하여 보다 소집단화 되고 개인화 되어가는 Micro-Trend 중심의 마케팅 활동을 할 수 있는 Micro-Targeting 관련 분석 정보를 제공 모델링하는 것을 제안한다. 본 연구에서는 SNS 데이터의 수집, 저장, 분석에 대한 내용을 다루고 있으며, 특히 마이크로타겟팅을 위한 정보를 머하아웃(Mahout)의 유클리드 거리 기반의 유사도와 K-평균 군집 알고리즘을 활용하여 구현하고자 하였다.

키워드: 마이크로타겟팅(Micro-Targeting), 빅데이터(Big Data), 소셜네트워크(SNS)

## I. 서론

인터넷뿐만이 아니라 스마트폰, 태블릿PC 등이 널리 보급되고 사용되는 스마트 라이프 환경이 발전함에 따라 SNS를 사용하는 이용자가 꾸준히 증가하고 있으며 기업의 경영활동 및 정치활동 전반에서 중요한 요소가 되고 있다. 2012년 11월 오바마 선거캠프는 이런 데이터의 중요성에 집중하여 미국 대통령에 당선되는 결과를 가져오게 되었고, 그 과정에서 마이크로타겟팅이 사용되었으며 널리 알려지게 되었다.

이러한 마이크로타겟팅을 많은 기업에서도 소비자를 대상으로 활용하고자 노력 중이다. 본 논문에서는 화장품이라는 소비재에 대한 SNS 이용자의 소비패턴을 직관적 아닌 분석을 통해 기업이 마이크로타겟팅을 할 수 있는 모델링 연구를 제안한다.

## II. 관련 연구

### 1. 마이크로타겟팅(Micro-Targeting)

소셜 네트워크 서비스의 등장과 함께 생겨난 개념 중 하나로 하이퍼타겟팅(Hyper-Targeting)이라고도 한다. 하나의 모집단을 추출하고 그 모집단 속에서 관심사나 성격 등의 미세 단위별 행동과 반응을 분석해 유기적인 관계를 이끌어내 마케팅 활동의 효과를 극대화하는 방법이다. 또한, 사람들이 자신의 정보를 소셜 네트워크

프로필에 공개하기 때문에 기본적 접근이 가능하며 지역, 나이, 성별, 관심사 등에 따라 특정한 프로필을 가진 타겟에게만 광고를 노출할 수 있어 최근 많은 기업들이 관심을 보이고 있다.

### 2. 유클리드 거리 기반의 유사도

해당 구현방법은 사용자 사이의 거리에 기반한다. 선호값으로 좌표 체계를 구성한 많은 다차원 공간에 사용자를 하나의 위치로 보며, 두 사용자 위치 사이의 유클리드 거리  $d^3$ 를 계산한다. 즉, 유클리드 거리가 큰 값을 가지면 사용자 간의 거리가 멀다는 의미이기 때문에 사용자 간 유사성이 떨어져서 값 자체로는 유의미한 유사도 측정으로 보이지 않는다.

### 3. K-평균 군집

K-평균 알고리즘(K-means algorithm)은 주어진 데이터를 k개의 클러스터로 묶는 알고리즘으로, 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다.

$$V = \sum_{i=1}^k \sum_{j \in S_i} [x_j - \mu_i]^2$$

이 값을 최소화하는  $S_i$ 를 찾는 것이 알고리즘의 목표가 된다. 알고리즘은 우선 초기의  $\mu_i$ 를 임의로 설정하는 것으로 시작한다. 그 다음에는 다음의 두 단계를 반복한다. 클러스터 설정: 각 점에

대해, 그 점에서 가장 가까운 클러스터를 찾아 배당한다.  $\mu_i$ 를 각 클러스터에 있는 점들의 평균값으로 재설정해준다.

만약 클러스터가 변하지 않는다면 반복을 중지한다.

### III. 본 론

#### 1. 연구 대상 선정 및 시스템 구현

본 연구에서는 일상에서 널리 사용되고 소비자의 이용 반응이 즉각적으로 표현되는 소비재로서 화장품을 선택하였으며 아래의 그림 1과 같이 관련 SNS를 수집하고 분석하여 마이크로타겟팅에 대한 모델링이 가능하도록 시스템을 구현하였다.

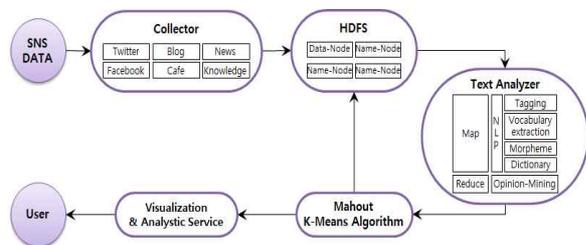


그림 319. 모델링 시스템 구성도

#### 2. 사용자간의 유클리드 거리와 유사도

제품에 대한 사용자들의 선호도가 유사할수록 유클리드 거리 d3 값이 작아야 한다. 따라서 본 연구에서 사용된 시스템은  $1/(1+d)$ 의 값을 반환하도록 구현하였다. 거리가 0일 때는 결과값이 1이고, 거리 d값이 증가함에 따라 값이 작아지는 것을 확인할 수 있다.

아래 표1은 마하트의 EuclideanDistanceSimilarity 클래스를 이용하여 화장품 소비자들 간의 유사도를 측정된 결과중 일부이다.

표 1. EuclideanDistanceSimilarity 사용 유사도 측정

	미백	여드름	동안	거리	소비자1과의 유사도
소비자1	3,0	2,5	5,0	0,000	1,000
소비자2	2,5	5,0	2,0	3,937	0,203
소비자3	-	-	2,5	2,500	0,286
소비자4	-	3,0	5,0	0,500	0,667
소비자5	3,0	2,0	4,0	1,118	0,472

#### 3. K-평균 군집과 마이크로타겟팅

마하트의 K-Means Algorithm은 KmeansCluster 또는 KmeansDriver 클래스로 실행할 수 있다. KmeansCluster 클래스는 인메모리 형식으로 군집 처리를 한다. 반면에 KmeansDriver 클래스는 K-평균 알고리즘을 MapReduce 로 실행하는 시작점이다.본 연구에서는 그림2의 예시와 같이 화장품의 미백, 여드름, 동안 기능에 대한 군집과 유사도를 측정할 수 있었으며 실제로는 화장품의 50여가지 기능에 대한 실험을 하였다.

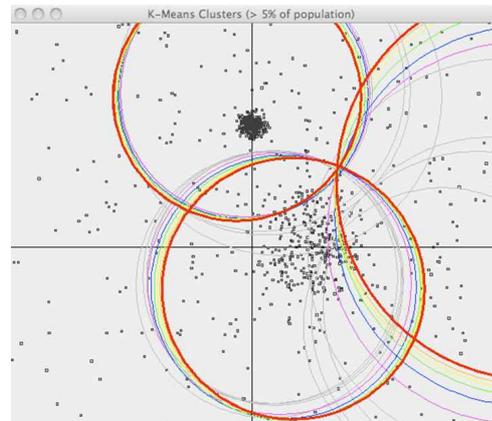


그림 320. K-평균 군집과 분포

마하트의 K-평균 군집과 유사도 측정을 통한 SNS 빅데이터 분석을 화장품 제품에 적용하여 그림3과 그림4와 같이 마이크로타겟팅을 위한 모델링을 구현할 수 있었다.

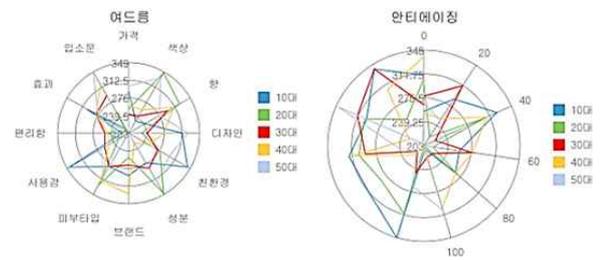


그림 321. 연령별 기능성 화장품 Micro-Trend 분석 결과

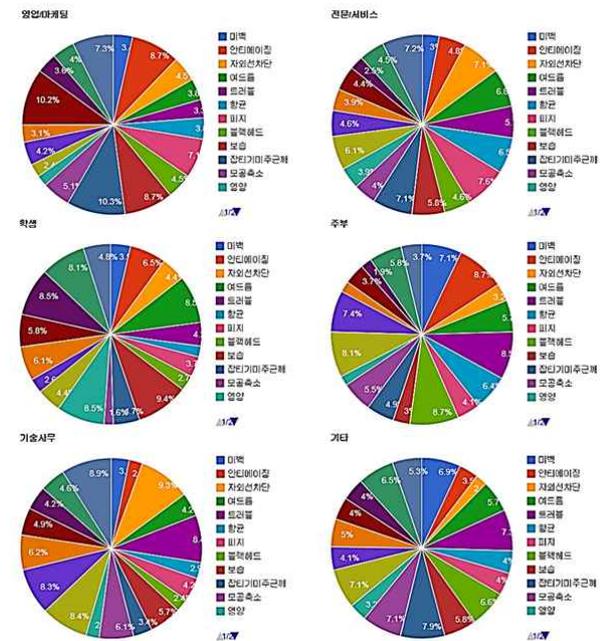


그림 322. 직업군별 기능성 화장품 Micro-Trend 분석 결과

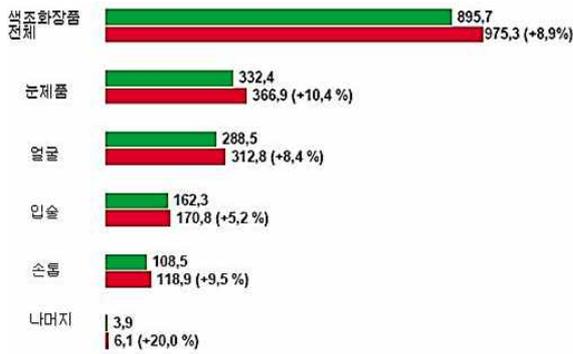


그림 323. 남녀별 색조 화장품 관심도 분석

위의 그림5는 SNS 빅데이터 분석을 통한 남녀 성별 색조 화장품에 대하여 유사도가 높은 신체 사용부위를 시각화한 것이다.

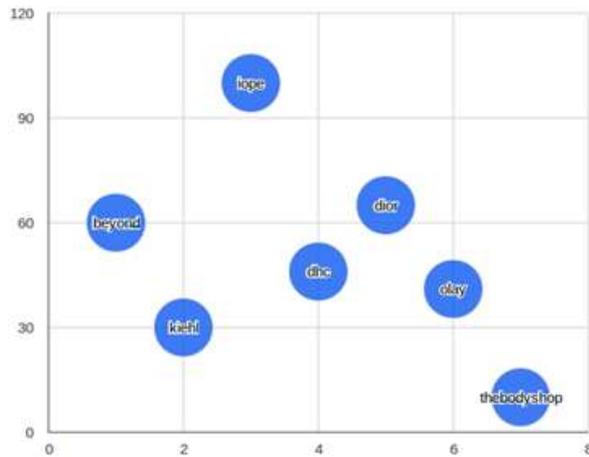


그림 324. R을 이용한 화장품 브랜드 K-평균 군집 시각화

그림6은 머하웃을 이용해 화장품 브랜드에 대한 선호도를 K-평균 군집으로 구하고 R을 이용하여 시각화한 것이다.

#### IV. 결론

본 논문에서의 마이크로타겟팅은 특정 제품에 대하여 SNS를 통해 대중적 의견과 생각 등이 담겨진 데이터를 수집하고 분석하여 활용하는 것을 목적으로 소비자 완제품을 분석 대상으로 선정해야 했다. 이에 화장품 제품군 하나를 선정하고 상세하게 분석하고자 하였다.

기업에서의 마이크로타겟팅은 모집단을 추출하고 그 모집단 속에서 관심사나 성격 등의 미세 단위별 행동과 반응을 분석해 유기적인 관계를 분석하여 마케팅 활동을 하는 것을 의미한다. 즉, 소규모의 집단이나 개인별 취향을 분석하여 틈새시장을 통해 기업과 제품에 대한 시장을 확대하는 것이다. 이와 같은 의미에서 본 연구는 K-평균 군집을 통해 모집단을 추출하고 유클리드 거리 기반의 유사도를 통해 일반 소비자 집단에 대한 분석을 할 수 있으며, 해당 연구 결과를 반대적으로 활용하여 SNS를 이용한 소집단 및 개인을 대상으로 하는 마이크로타겟팅 모델을 수립하고 연구할 수 있었다.

수집되는 SNS 데이터가 축적될수록 분석결과는 정확해질 것으로 예측된다. 누적 수집 및 누적 분석되는 전체 데이터에 비교하여 현재 시장에서의 또는 특정 시점에서의 기업, 제품, 브랜드에 대한 수준 값을 연계할 수 있다면 향후 비즈니스 모델링 구현을 위한 분석과 시스템 구현도 가능할 것으로 예상된다.

#### 참고문헌

- [1] Jeong Min Seo, Jeo Song, Chae Ri Lee, Sang Moon Lee, "Functional Cosmetics Trend Analysis System using Big Data From The Girls High School Of SNS", KSCI Proc. Vol.21, No.1, pp.99-101, 2013
- [2] Jeo Song, Da Young Yun, Jae Moon Kim, Sang Moon Lee, "Cloud-based Enterprise Information Portal System Using SNS Big Data for Business Competitiveness", MITA international Conference, HongKong, Vol.10, 2014.
- [3] Da Young Yun, Hyun Hwa Lee, Sang Moon Lee, Jeo Song, "A Study the Automatic Summary System of Information using Big Data", KSCI Proc. Vol. 22, No. 1, pp.415-416, 2014
- [4] Nancy Lynch, Seth Gilbert, "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services", ACM SIGACT News, Vol.33, No.2, pp.51-59, 2002.
- [5] Jeffrey Dean, Sanjay Ghemawat, "MapReduce: a flexible data processing tool", InProc of Magazine Communications of the ACM, Vol.53, Issue 1, January, 2010
- [6] "Big Data: The next frontier for innovation, competition", productivity, McKinsey Global Institute, 2011
- [7] Soon Hwa Ju, "An Influence of SNS usage and product involvement on SNS marketing of cosmetics", Graduate School of Food and Drug Administration, Chung-Ang University, 2012.
- [8] "Analysis Report of Cosmetics", Korea Health Industry Development Institute, 2013.
- [9] "Small Change and Big Market-Make Capture Micro-Trends", Global Business Report, Vol.08, No.31, KOTRA, 2008.
- [10] "A Study on Fostering Global Cosmetics Industry", Ministry of Health and Welfare, Korea Cosmetics Association, 2009.
- [11] "Survey the Distributors of Medical Machine and Cosmetics", Ministry of Health and Welfare, Korea Health Industry Development Institute, 2012.
- [12] Sung Il Kim, Duik Hong Min, Jong Heuk Park, Hwan Ku Lee, Yoon Cheul Yoon, Hye Jeung Koh, Eun Hye Lee, "Study on Target Marketing Strategy Based on Each Consumer Group in Cosmetics Business", Korean Society for Aesthetics and Cosmetology Proc, Vol.4, No.2, pp.21-29, 2009.

- [13] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, Vol.1, pp.281-297, 1967
- [14] Sean Owen, Robin Anil, Ted Dunning, Ellen Friedman, MAHOUT IN ACTION, Manning Publications, USA, 2011
- [15] D. Arthur, S. Vassilvitskii, "How Slow is the k-means Method?," Proceedings of the Symposium on Computational Geometry, 2006
- [16] Jeo Song, Jin Kwan Kwon, Yoon Su Park, Sang Moon Lee, "An Implementation of Micro-Targeting Services for Cosmetics using SNS Big-Data Analysis", Korea Multi Media Society Proc. Vol.17, No.2, pp.128-131, 2014