

# SNS 게시물 분석에 기반한 사용자 관심분야 예측

김은상, 전영호, 박효주, 이기훈<sup>1)</sup>  
광운대학교 컴퓨터공학과  
e-mail: kihoonlee@kw.ac.kr

## Predicting User Interests Based on SNS Posting Analysis

Eun-Sang Kim, Young-Ho Jeon, Hyo-Ju Park, and Ki-Hoon Lee  
Dept. of Computer Engineering, Kwangwoon University

### 요 약

본 논문에서는 개인 맞춤형 콘텐츠 제공을 위해 SNS 게시글을 분석하여 개인의 관심분야를 예측하는 방법을 제안한다. 사용자의 SNS 게시글뿐만 아니라 사용자가 속한 그룹의 SNS 게시글도 분석하여 관심분야 예측의 정확도를 높인다. 사용자가 가입한 그룹은 사용자의 관심분야와 밀접한 관련이 있을 가능성이 높기 때문에 사용자가 속한 그룹의 게시글은 사용자의 관심분야를 예측하는데 있어 중요한 요소가 된다.

### 1. 서론

정보통신 기기의 발달은 소셜 네트워크 서비스(Social Network Service, 이하 SNS)의 확산을 촉진시키고 있다. 초기의 SNS는 일반화된 온라인 커뮤니티 형태로 특정 카테고리에 관심을 가진 집단이 폐쇄적인 서비스를 공유하는 것이었다. 이후 개인 중심이 되어 자신의 관심사나 개성을 공유하며 수천만 명의 고유한 개인 사용자를 모을 수 있는 SNS들이 제공되어 비즈니스와 각종 정보공유를 위한 용도로 활용되고 있다. 특히 동일한 관심분야를 가지는 사람들이 서로 인맥을 형성하고 정보를 교류하기 때문에, 이들이 제공하는 정보는 신뢰성이 높다는 특징을 가진다. 따라서 사용자의 관심분야를 예측하기 위해 SNS 데이터를 이용하는 것은 효과적인 방법이 될 것이다.

본 논문에서는 SNS 게시글을 분석하여 사용자의 관심분야를 자동으로 예측해내는 시스템을 제안한다. 이때 사용자의 게시글만을 분석하는 것이 아니라 사용자가 속한 그룹의 게시글도 함께 분석한다. 사용자가 속한 그룹은 사용자의 관심사를 반영하고 있다고 볼 수 있으므로 그룹의 게시글을 분석하면 관심분야 예측의 정확도를 크게 향상시킬 수 있을 것으로 기대된다. 이를 토대로 다양한 분야에서 개인 맞춤형 서비스를 제공할 수 있다.

### 2. 관련 연구

스마트폰에 관련된 SNS 게시글을 분석하여 스마트폰에

대한 사용자의 요구특성을 분석하는 연구가 진행되었다 [1]. SNS 게시글을 분석한다는 점은 유사하나 오직 스마트폰에 관련된 게시글만을 분석하므로 사용자의 관심분야를 예측하는데 적절하지 않다는 점에서 본 논문과의 차이가 있다.

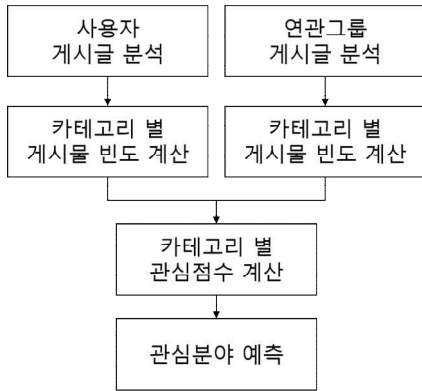
### 3. 제안하는 시스템

사용자에게 맞춤형 콘텐츠를 제공하기 위해서는 사용자의 관심분야를 알아낼 수 있어야 한다. 본 논문에서는 그림 1과 같이 사용자의 게시글과 사용자가 속한 그룹(이하 *연관그룹*)의 게시글을 분석하여 자동으로 관심분야를 예측하는 방법을 제안한다. 사용자의 게시글과 연관그룹의 게시글을 모두 분석하여 카테고리 별 게시글 빈도를 계산하고 이 빈도를 바탕으로 사용자의 최종 관심분야를 예측하게 된다.

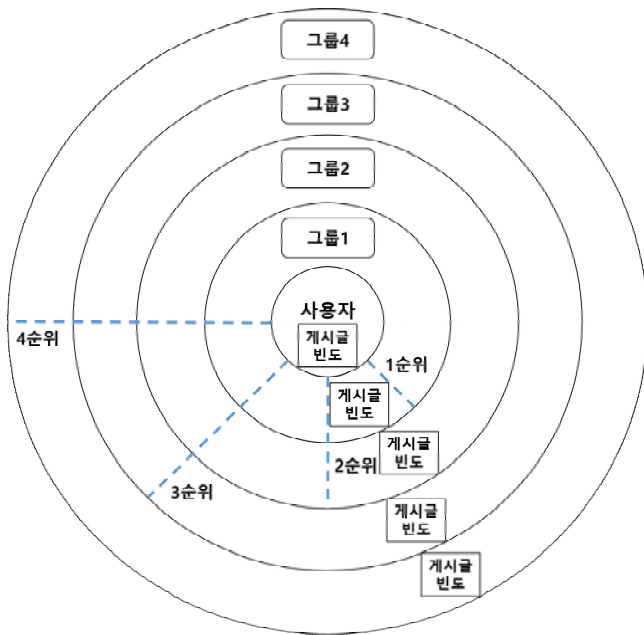
카테고리 별 게시글 빈도(%)는 (해당 카테고리에 속한 게시글 수 / 전체 게시글 수) \* 100을 통해 구한다. 이때 각 게시글이 어떤 카테고리에 속하는지를 판단하기 위해 지지 벡터 기계(SVM: Support Vector Machine)과 같은 기계학습 분류 알고리즘을 이용한다.

사용자가 다수의 그룹에 속해 있는 경우에는 사용자와의 관련성에 따라 그림 2와 같이 그룹들에 순위를 부여한다. 순위는 다양한 기준에 의해 정해질 수 있다. 예를 들어, 사용자가 관리자인 그룹을 1순위, 사용자가 자주 방문하는 그룹에 대해서 방문 횟수를 기초로 2순위부터 부여할 수 있다. 각 그룹에 순위에 따라 가중치를 부여한다.

1) 교신저자



(그림 1) SNS 게시물 분석에 기반한 관심분야 예측



(그림 2) 사용자가 속한 그룹의 순위화

각 카테고리에 대한 사용자의 관심점수는 사용자의 카테고리 별 게시물 빈도와 연관그룹의 카테고리 별 게시물 빈도를 그림 3의 계산식과 같이 가중 합산하여 구한다. 예를 들어 사용자 및 연관그룹의 카테고리 별 게시물 빈도가 그림 4와 같다고 할 때 그림 3의 계산식을 적용하면 그림 5와 같이 사용자의 카테고리 별 관심점수를 구할 수 있다. 이를 통해 해당 사용자는 IT, 경제, 사회, 연예, 생활, 스포츠 순으로 관심이 있다는 것을 예측할 수 있다.

$s^c = 100 \times \sum_{i=0}^n w_i f_i^c \quad (w_{i+1} \leq w_i, \sum_{i=0}^n w_i = 1)$	
$s^c$	카테고리(c) 별 관심점수
$w_i$	사용자 또는 연관그룹의 가중치 ( $w_0$ 는 사용자의 가중치)
$f_i^c$	사용자 또는 연관그룹의 카테고리 별 게시물 빈도
$n$	연관그룹의 개수

(그림 3) 카테고리 별 관심점수 계산식

구분	카테고리 별 게시물 빈도						가중치 (w)
	IT	경제	사회	연예	생활	스포츠	
사용자	34.5%	5.4%	19.2%	8.2%	22.5%	10.2%	0.4
1순위 연관그룹	11.9%	39.9%	19.7%	25.2%	3.3%	0.0%	0.3
2순위 연관그룹	33.8%	5.7%	0.0%	26.7%	11.5%	22.3%	0.15
3순위 연관그룹	40.9%	19.8%	39.3%	0.0%	0.0%	0.0%	0.08
4순위 연관그룹	5.0%	27.0%	10.1%	9.2%	18.9%	29.8%	0.07

(그림 4) 사용자 및 연관그룹의 카테고리 별 게시물 빈도

카테고리	IT	경제	사회	연예	생활	스포츠
관심점수	26.06	18.46	17.44	15.49	13.04	9.51

(그림 5) (그림 4)로부터 계산된 카테고리 별 관심점수

사용자의 실제 관심분야와 SNS 게시글을 이용하여 예측한 관심분야가 다를 경우에는 피드백을 받아 카테고리 별 관심점수를 수정한다. 이 피드백을 통하여 정확도를 높일 수 있다.

#### 4. 결론

본 논문에서는 SNS 사용자에게 맞춤형 콘텐츠를 제공하기 위해 자동으로 사용자의 관심분야를 예측하는 방법을 제안하였다. 사용자의 게시물뿐만 아니라 사용자가 속한 그룹의 게시물도 분석함으로써 관심분야 예측의 정확도를 향상시킬 수 있을 것으로 기대된다.

#### 감사의 글

본 연구는 중소기업청에서 지원하는 2014년도 산학연협력 기술개발사업(No. C0187264)의 연구수행으로 인한 결과물임을 밝힙니다.

#### 참고문헌

[1] 김태우, 백동현, “SNS의 스마트폰 게시물 내용 분석을 통한 사용자의 요구특성 분석,” *한국전자거래학회지*, Vol. 17, No. 4, pp. 197-208, 2012년 11월.