

---

# 소셜 네트워크 키워드 분석을 통한 영화 추천 시스템 설계

양새동 · 이종원 · 저순 · 편도길 · 정희경

배재대학교 컴퓨터공학과

## Design for Recommended System of Movies using Social Network Keyword of Analysis

Xi-tong Yang · Jong-Won Lee · Xun Chu · Do-Kil Pyoun · Hoe-Kyung Jung

Department of Computer Engineering, PaiChai University

E-mail : withchyang1@gmail.com, starjwon@naver.com, 846045855@qq.com, dokil25@kbtus.ac.kr,

hkjung@pcu.ac.kr

### 요 약

IT 기술 발전과 스마트 기기의 보급으로 인해 다양한 웹 서비스가 개발되었다. 특히, 소셜 네트워크 서비스는 기존 웹 서비스와 달리 정보의 생산과 소비의 구별이 없으며 사용자들 간에 자유롭게 의사를 소통할 수 있게 한다. 그리고 정보 공유, 기존 인간관계와 새로운 인간관계들을 형성을 강화시켜준다.

본 논문에서는 소셜 네트워크 서비스가 제공하는 사회적 관계망을 사용하여 사용자들의 의사소통과 정보 공유에 사용되는 키워드를 수집하고 분석한다. 그리고 키워드에 적합한 영화들을 추천하는 시스템을 설계한다.

### ABSTRACT

Was developed of the web service in Due to the dissemination for IT skills development and smart appliances. In particular, Social network service for should be able to communicate feel free to a user across without distinguishing between production and consumption information in contrast to the existing web service. And strengthen to the information sharing relationships between existing human relation and new human relation.

In this paper, a social network service in providing a social networking from users using their communication and information sharing is used to collect and analyze in the keyword. And a design of recommended system of movies for appropriate keyword.

### 키워드

MongoDB, Recommend System, Social Network, Web Crawler

### 1. 서론

스마트 기기의 확산과 웹 서비스 기술이 발전하면서 장소 구별 없이 어디에서나 사람들과 소통할 수 있는 소셜 네트워크 서비스의 사용자들이 증가되고 있다[1]. 그로인해 소셜 네트워크 서비스의 정보의 중요성 또한 부각되고 있다.

소셜 네트워크 서비스를 대표하는 페이스북 북이나 트위터에는 다양한 사용자들의 개인 생활과 제품 평가, 사회정치 사건 등의 다양한 정보들이 공유되고 있다[2,3]. 또한, 감상한 영화들의 평점

과 평가들도 공유되고 있으며 다양한 웹 사이트와 영화 평가 사이트도 소셜 네트워크 서비스와 연동하고 있다.

본 논문에서는 소셜 네트워크 서비스에 사용되는 키워드를 수집하여 MongoDB에 저장한다. 그리고 Mahout를 사용하여 키워드를 영화를 분류하고 사용자의 키워드의 빈도를 측정하여 키워드에 적합한 영화를 추천하는 시스템을 설계한다.

## II. 관련 연구

본 장에서는 분산 처리 시스템인 Hadoop과 데이터 마이닝과 추천 시스템에 사용되는 Mahout, NoSQL인 MongoDB에 대해 기술한다.

### 2.1 Hadoop

분산 처리 시스템인 Hadoop은 Java로 개발되었으며 MapReduce를 사용하여 대용량의 데이터를 분산 처리하고 HDFS(Hadoop Distributed File System)를 사용하여 파일들을 저장하는 시스템이다. 기존 Hadoop 1.x의 문제점으로 지적되고 있던 확장성과 작업 실행 병목 현상을 해결하기 위해 Hadoop 2.x인 YARN을 개발했다[4].

### 2.2 Mahout

Mahout은 Hadoop 환경에서 사용할 수 있는 Machine Learning Library이다[5]. Java로 개발된 Mahout은 다양한 작업과 분류, 클러스터링, 패턴 마이닝 등을 제공하며 많은 프레임워크를 호환한다. Mahout은 일반적으로 Hadoop 환경에서 실행하며 다양한 알고리즘과 병렬 처리를 통해 대용량의 데이터를 처리할 수 있다.

### 2.3 MongoDB

MongoDB는 웹 어플리케이션과 인터넷 기반의 데이터베이스를 관리하는 시스템이다. MongoDB 엔진은 C++로 개발되었으며 데이터를 JSON 형태로 저장된다[6]. 또한, 다양한 인덱싱을 제공하며 데이터 복제를 통해 가용성을 향상 시킨다. 그리고 여러 서버에 데이터를 나누는 기능과 MapReduce를 지원한다.

집한다. 두 번째로는 오픈 소스로 구현되어 있는 Crawler를 사용하여 자료를 수집한다. 수집한 자료는 MongoDB에 저장한다.

### 3.2 키워드 분석

소셜 네트워크의 키워드 분석에는 Hadoop과 Mahout을 사용한다. 키워드 분석의 처리 순서는 그림 2와 같다.

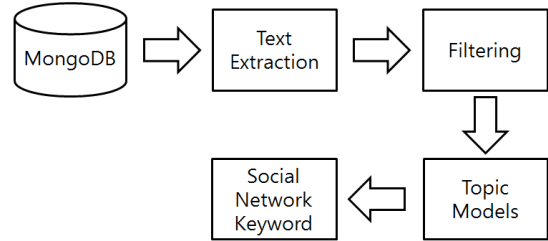


그림 2. 키워드 분석의 처리 순서도

먼저 Hadoop을 사용하여 MongoDB에 저장되어 있는 데이터베이스에서 텍스트를 추출한다. 그리고 추출한 텍스트에서 영화와 관련 없는 텍스트를 제거한다.

토픽 모델링 단계에서는 LDA(Latent Dirichlet Allocation) 알고리즘을 사용하여 그 달에 상영하고 있는 영화 리스트를 분류하고 자주 출현하는 키워드를 추출한다. 또한, 중영된 영화들은 월별로 분류한 뒤 해당 키워드들도 추출한다.

### 3.3 영화 추천

영화 추천에서는 그림 3과 같이 Mahout에서 제공하는 협업 필터링인 사용자 기반 협업 필터링과 아이템 기반 협업 필터링을 사용한다.

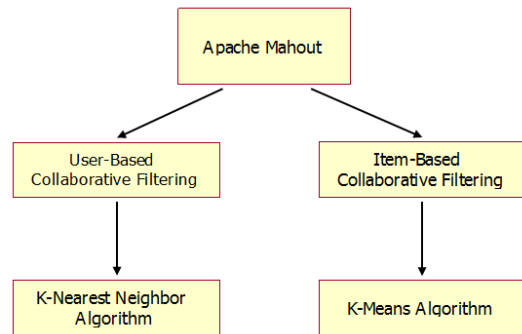


그림 3. 제안하는 시스템 구조도

사용자 기반 협업 필터링에서는 K-최대근접 이웃 알고리즘을 사용하여 사용자 간의 유사한 선호도를 가진 사용자의 영화 평가에 근거하여 영화를 추천한다. 아이템 기반 협업 필터링에서는 K-평균 알고리즘을 사용하여 유사도를 계산하고 영화를 추천한다.

## III. 시스템 설계

제안하는 시스템의 구조는 그림 1과 같으며, 페이스북 북과 트위터의 자료를 수집하는 Crawler와 Hadoop, Mahout, MongoDB로 구성되어 있다.

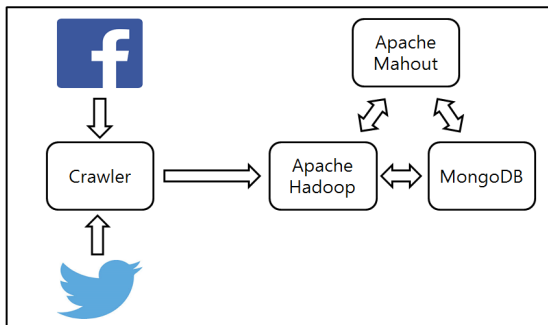


그림 1. 제안하는 시스템 구조도

### 3.1 정보 수집과 저장

페이스 북과 트위터에서 정보를 수집하는 방법은 두 가지로 분류된다. 첫 번째는 페이스 북과 트위터에서 제공하는 API를 사용하여 정보를 수

#### IV. 결 론

소셜 네트워크 서비스의 사용이 증가하면서 소셜 네트워크 서비스의 정보들을 사용하여 B2B나 B2C 마케팅 등에 활용되고 있다. 또한, 기업의 이익 극대화와 문제 개선으로도 활용되고 있다.

본 논문에서는 소셜 네트워크에서 사용되는 키워드를 활용하여 사용자에게 적합한 영화를 추천하는 시스템을 설계했다. 설계한 시스템은 오픈 소스로 개발되어 있는 프레임워크를 사용하여 간편하게 구축할 수 있다. 또한, 대용량 데이터의 수집과 분석, 저장할 수 있는 확장성을 제공한다. 그리고 Mahout이 제공하는 기능들을 사용하여 기계 학습 지식과 데이터 마이닝 지식 없이 추천 시스템의 기능들을 구현할 수 있다.

향후 연구로는 협업 필터링 기술의 문제점인 Cold-Start와 추천 시스템의 정확성을 개선하는 방향에 대한 연구가 필요하다.

#### 참고 문헌

- [1] Kazienko, Przemyslaw, Katarzyna Musial, and Tomasz Kajdanowicz, "Multidimensional social network in the social recommender system," IEEE Transactions on 41.4, Vol.41, pp.746-759, 2011
- [2] Philip. Bonhard, "Improving recommender systems with social networking," Proceedings Addendum of the 2004 ACM Conference on Computer-Supported Cooperative Work, 2004
- [3] 최홍구, 황인준, "트위터 문서 분석을 통한 감정 기반의 음악 추천 시스템," 정보과학회논문지, 제18권, 제11호, 2012.11
- [4] Vavilapalli, Vinod Kumar, et al., "Apache hadoop yarn: Yet another resource negotiator," Proceedings of the 4th annual Symposium on Cloud Computing, 2013
- [5] Anil, Robin, Ted Dunning, and Ellen Friedman, "Mahout in action," manning, 2011
- [6] Doychev, Doychin, A. Lawor, and Rachael Rafter, "An Analysis of Recommender Algorithms for Online News," CLEF, 2014