ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

태그의 문맥 정보를 이용한 웹 자원 추천 시스템

Tag Based Web Resource Recommendation System

송 제 인[1] 정 옥 란[1*]
Je-In Song Ok-Ran Jeong

요 약

최근의 여러 웹서비스에서는 태깅 기능을 제공함으로써 사용자가 작성하는 게시물의 주제를 표현하도록 유도하고 있다. 태그를 이용하면 글이나 사진에 대한 글쓴이의 감정과 같은 문맥적인 정보의 효과적인 추출이 가능하기 때문에, 기계적인 방식보다 글의 내용에 대해서 더 나은 의미 파악이 가능하다. 따라서 이를 추천시스템에 적용한다면 사용자의 만족도를 높일 수 있는 추천이 가능할 것이다. 본 논문에서는 게시글에 속한 태그들 간의 관계를 계산하고, 효율적인 유사도 측정 알고리즘을 통해 게시글과 사용자등의 웹 자원을 추천하는 방법을 제안한다. 마지막으로, 실험을 통해 제안한 방법의 유효성을 검증하고, 사용자의 만족도를 측정하였다.

☞ 주제어 : 추천, 태깅, 문맥 정보, 웹 서비스, 웹 자원

ABSTRACT

Recent web services provide tagging function to users, and let them express the topic of the contents of their articles. Moreover, we can extract context information like emotion of the writer efficiently by using tags attached to the articles or images. And we are able to better understand article than traditional algorithm. (eg. TF-IDF) Therefore, if we use tags in recommendation system, we can recommend high quality resources to the users. This study proposes a recommendation method that provide web resources (articles, users) through simple algorithm based on related tag set extracted from the article. Through the experiments, we show that the result was satisfactory, and we measure the satisfaction of users.

☞ keyword : Recommendation, Tagging, Context Information, Web service, Web resource

1. 서 론

최근의 페이스북(Facebook), 인스타그램(Instagram), 트위터(Twitter)등의 소셜 네트워킹 서비스(Social Network Service), 블로그, 이외에도 여러 웹서비스에서 사용자가 키워드를 자신의 글이나 사진에 직접 달아 놓는 행위인 태깅(Tagging)이라는 개념은 많은 사람들에게 익숙하다. 사용자는 태깅을 통해서 글이나 사진에 대한 주제나 문맥, 감정에 관한 정보를 표현할 수 있고, 서비스 제공자 입장에서는 이러한 정보를 이용해서 더 좋은 서비스를 제공하는데 활용할 수 있다. 특히, 사진이나 비디오 같은 웹 자원 같은 경우 그 자체로부터 텍스트로 표현되는 문맥적인 정보를 충분히 제공받기가 힘들기 때문에 태그를 활용함으로써 정보의 부족 문제를 해결할 수 있다. 이외에도 태깅을 이용하여 웹 자원의 인덱싱(Indexing)에 활용할 수 있으며, 실제로 인스타그램의 경우엔 검색을 태그 기반으로 제공하고 있다. 본 연구에서는 이를 활용한 추천 시스템에 대한 연구를 진행하였다.

본 연구에서는 사용자의 태깅 활동을 기반으로, 태그 간의 관계를 계산하고, 이를 통해 웹 자원을 효과적으로 추천하는 방법을 제안한다. 게시글에 남겨진 태그는 해당 글을 가장 잘 표현할 수 있는 주제라고 볼 수 있기 때문에, 태그를 기반으로 추천을 한다면 글의 주제를 추출하기위한 별도의 복잡한 계산 없이도 효과적으로 글을 이해할 수 있는 장점이 있다.

본 연구에서 제안하는 추천 방법은 크게 3 단계로 구성되어있다. (1)사용자의 게시글 및 상호작용을 기록하는 단계 (2)연관 태그를 계산하는 단계, 마지막으로 (3)사용

[1] Dept. of Software, Gachon Univ., Seongnam, 461-701, Korea
* Corresponding author (orjeong@gachon.ac.kr)
[Received 12 September 2016, Reviewed 27 September 2016, Accepted 21 October 2016]
☆ 본 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업지원과 미래창조과학부 및 정보통신기술진흥센터의 ICT/SW창의연구과정지원사업(SW중심대학)의 지원을 받아 수행한 것임.
(NRF-2015R1C1A2A01051729, R2215-14-1006)

자 추천 알고리즘을 통한 사용자를 추천하는 단계와 게시글 추천 알고리즘을 통한 관련 게시글 추천의 단계로 구성되어 있다. 자세한 방법은 3장에서 기술한다.

2장에서는 태그를 이용한 추천에 관한 연구들을 다루고, 3장에서는 본 연구에서 제안하는 추천 방법(게시글 및 사용자추천)의 각 부분에 대해서 자세히 설명한다. 마지막 4장에서는 제안한 방법에 대한 실험과 분석을 하면서 본 논문을 끝맺고자 한다.

2. 관련 연구

태그는 사용자가 글에 부여하는 글이 가지는 메타데이터라고 할 수 있다. 사용자가 직접 작성하는 메타데이터인 만큼 글의 주제에 대한 정확한 정보를 담고 있다. 따라서 이러한 태그를 활용한다면 대량의 정보 속에서 원하는 주제의 글을 효율적으로 찾아 낼 수가 있다. 이미 트위터, 페이스북, 인스타그램 등의 소셜 네트워킹 서비스에서는 태깅 기능(해시태그)을 가지고 있어, 사용자들이 특정한 관심사를 쉽게 찾거나 공유할 수 있게 도와주고 있다.

태그를 활용하지 않을 경우엔 글의 주제 파악을 위해 별도의 형태소 분석과 TF-IDF등의 부가적인 계산이 필요하지만, 사용자가 직접 정의한 태그를 활용함으로써 복잡한 계산을 없애고 효율적으로 해당 글에 대한 이해 및 구조화 등이 가능하다는 장점을 가지고 있다.

태깅을 이용한 추천에 앞서, 본 연구에서 제시하는 추천 방법은 다음과 같은 가정을 따른다. Harris, Z.는 [2]에서 “같은 문맥에서 쓰이거나 등장하는 단어는 비슷한 의미를 지니는 경향이 있다”고 밝혔다. 이에 근거하면 태그를 통해서 글의 주제를 효과적으로 알아 낼 수 있다는 사실에서 나아가 사용자가 작성한 태그들 중 자주 같이 등장하는 태그라면 비슷한 의미를 지닐 가능성이 높다고 가정할 수 있다. 실제로 위의 가정에 근거하여 언어의 분포와 의미의 상관관계에 관한 많은 연구들이 이루어졌으며, 위의 가정에 근거한 언어 모델 word2vec을 이용하면 정확도 높은 모델을 구현할 수 있다고 알려져 있다. [3] 본 연구에서는 이러한 가정에 근거하여 태그들을 분석함으로써 연관 있는 태그들을 찾아내고, 이러한 태그들 간의 관계를 활용하여 실제 웹서비스 아키필드[4]에 게시글 및 사용자를 추천하는 기능을 구현하는데 활용하고자 하였다.

추천 시스템에는 크게 협업 필터링, 내용 기반 접근법 두 가지 접근이 있다. 협업 필터링은 사용자의 행동(평가내역)을 분석하고 비슷한 사용자간의 선호 아이템을 추천하거나, 아이템간의 유사도를 측정하여, 특정 사용자가 선호하는 아이템과 유사한 아이템을 추천해줄 수 있는 방법이다. 내용 기반 접근법은 TF-IDF등의 방법을 이용, 아이템의 컨텐츠 자체를 분석해 추천해주는 방법이다. [1] 이러한 특징을 가진 태그를 활용한 효과적인 글 추천 방법에 대한 다양한 연구들이 있다.

먼저 태그를 이용하여 사진이나 영화에 대한 분석을 시도하는 연구들이 진행되었다. [5]에서는 소셜 네트워크의 태그와 시간 정보를 반영하여 개인화된 추천 시스템을 연구하였다. [6]에서는 사진공유 웹서비스 플리커(Flickr)의 5200만개의 사진을 분석하여 사용자들이 어떠한 방식으로 사진에 태그를 추가하는지, 어떠한 태그를 제공하는지를 연구하였고 이를 바탕으로 사용자에게 특정 사진에 대한 태그를 추천해주는 방법을 제안하였다. [7]에서는 태깅을 사용자로 하여금 온라인상의 자원을 찾고 구조화하고 이해하는데 도움을 주는 강력한 매커니즘으로 이해하여 tagommenders라는 추천 시스템을 제안하였다. 영화에 달린 태그에 기반하여 사용자가 선호하는 태그로부터 영화에 대한 선호도를 예측하고 추천하는 시스템이다.

또한 SNS상에서의 사용자의 활동 정보를 이용한 연구들이 진행되었다. [8]에서는 SNS의 데이터를 전통적인 추천 알고리즘과 결합하여 추천의 효율을 높이고자 하였다. 이를 위해 사용자의 태깅 활동과 사회적 관계를 이용하여 사용자간의 유사도를 측정하는 새로운 방법을 제안했다. 나아가 사용자간의 유사도를 활용하여 아이템을 추천하는 방법을 제안했다. [9]에서는 사용자들의 태깅 활동을 잠재적인 정보(사용자 개인적인 흥미, 선호도, 목적 등)의 근원으로써 정의하고, 태그를 개인화된 추천에 활용하였다. 사용자들의 태깅활동에 근거해서 그들의 태그와 유사한 웹페이지를 추천해주는 태그 기반의 추천 시스템을 제안하였다.

기존 연구를 통해 태깅을 추천에 활용함으로써 효과적인 추천이 가능함을 확인하였다. 본 연구에서는, 사용자들의 태깅 활동을 기반으로 같은 문맥에서 자주 쓰이는 단어일수록 서로 같은 의미를 지닐 가능성이 높다는 기존의 언어의 분포에 관한 성질[2]에서 아이디어를 얻어 주어진 태그집합 간의 유사도를 통한 관계 추출을 통해 게시글 혹은 사용자와 같은 웹 자원을 추천하는 방법을 제안한다. 이를 실제 웹서비스, 건축작품 관리를 위한 소셜 플랫폼 ‘아키필드’[4]에 태그를 기반으로 한 효율적인

게시글 및 사용자 추천 기능을 구현하여 사용자들의 서비스 이용 만족도를 극대화 하고자 하였다.

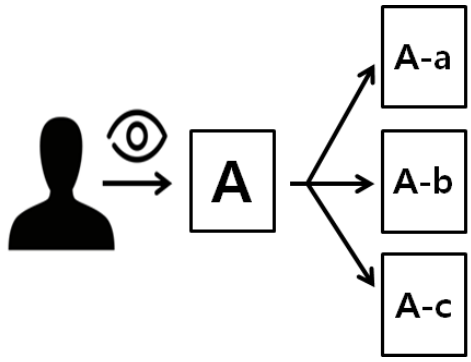
3. 태그 기반 웹자원 추천 시스템

본 장에서는 주어진 태그집합에서 태그들 간의 관계를 추출하고, 추출된 태그 사이의 관계를 이용하여 관련 게시글 및 사용자를 추천하는 방법에 대해 자세히 설명한다.

3.1 제안하는 추천 방법

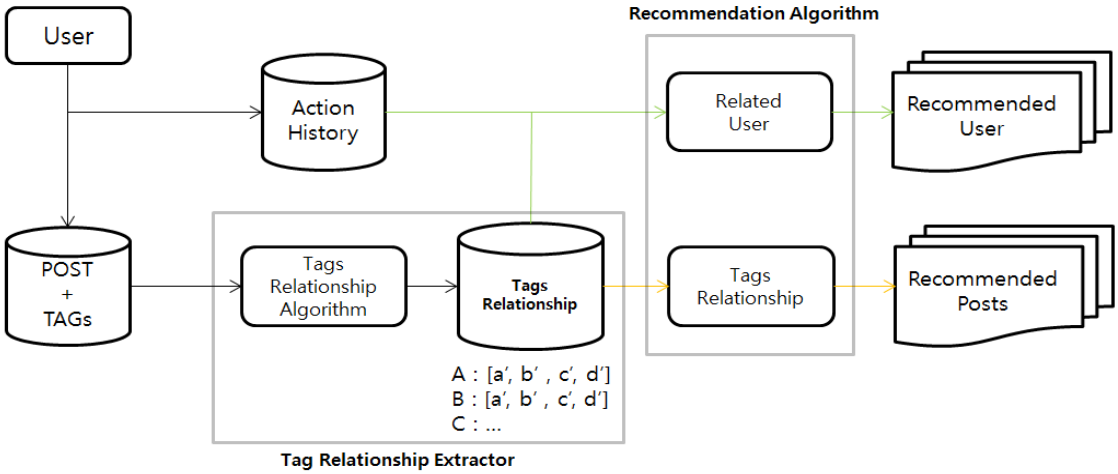
본 연구에서 활용한 웹서비스[4]에서는 그림 1과 같이 사용자가 현재 조회하고 있는 글('A')의 다음 글로 어떤 글('A-a', 'A-b', 'A-c')을 추천해줄 것 인지, 그리고 활발한 커뮤니티의 형성을 위해서 특정 사용자에게 비슷한 관심사를 가진 다른 사용자를 어떻게 추천해줄 것인가의 문제를 가지고 있었다. 이를 해결하기위해 본 논문에서 제안하는 추천 방법에서는 각 게시글에 달린 태그들을 하나의 태그집합으로 생각하고, 태그들 간의 관계를 분석하여 그 결과를 추천에 활용한다.

태그는 사용자가 게시글을 남기면서 해당 글을 가장 잘 표현할 수 있는 주제를 남긴 것이라고 볼 수 있기 때문에, 사용자가 조회하고 있는 글의 다음 조회 후보로 태그를 기반으로 선정된 비슷한 게시글을 추천해 준다면, 글의 주제 추출을 위한 복잡한 알고리즘을 사용할 필요 없이, 간단한 알고리즘으로써 사용자에게 충분히 의미 있는 추천이 가능할 것이라고 생각하였다.



(그림 1) 게시글 추천
(Figure 1) Article Recommendation

이 외에도 사용자가 서비스를 이용하면서 남기는 게시글과의 여러 상호작용(댓글, 좋아요, 공유)들은 모두 데이터베이스에 기록이 된다. 우리는 사용자의 상호작용을 그래프로 표현하고, 관심이 비슷하면서 양질의 웹자원을 제공할 수 있는 사용자를 찾아내어 추천하고자 하였다. 나아가 관심사가 비슷한 사용자를 추천해줌으로써 관심사를 공유하는 커뮤니티 형성을 기대할 수 있다.



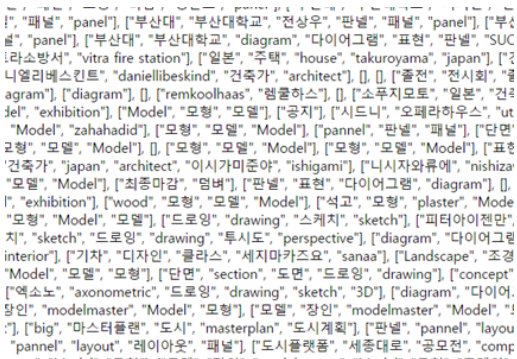
(그림 2) 제안하는 추천 방법의 전체적인 구조
(Figure 2) Overall Structure of Proposed Recommendation Method

본 연구에서 제안하는 추천 방법은 그림 2와 같이 크게 3 단계로 구성되어있다. (1)사용자의 게시글 및 상호작용을 기록하는 단계 (2)연관 태그를 계산하는 단계, 마지막으로 (3-a)사용자 추천 알고리즘을 통한 사용자 추천하는 단계와 (3-b) 게시글 추천 알고리즘을 통한 관련 게시글 추천의 단계로 구성되어 있다.

3.2 관련 태그 추출 방법

관련 태그를 추출하는 데에 활용할 수 있는 전통적인 방법으로는 여러 가지가 있다. 빈발 패턴 마이닝(Frequent Pattern Mining)기법의 Apriori[10], FP-Growth[11]등의 알고리즘이 있는데, 본 연구에서는 이를 기반으로 약간 변형된 방법으로 연관 있는 태그를 추출하였다. 그 방법은 언어에 관한 다음의 아이디어 “같은 문맥에서 쓰이거나 등장하는 단어는 비슷한 의미를 지니는 경향이 있다”라는 생각에 기초를 두고 다음의 방법을 제안한다.

관련 태그 추출 알고리즘의 입력으로는 그림 3과 같은 태그집합이 주어진다. 하나의 태그집합은 하나의 게시글에 달린 태그의 집합을 의미하며, 배열([태그1, 태그2, 태그3, 태그4 …])로 이루어져있다.



(그림 3) 입력 태그집합
(Figure 3) Input Tag Set

관련 태그 추출을 얻는 방법은 그림 4와 같다. 태그집합을 입력 받고, 각 태그 별로 나머지 태그에 대해서 같이 등장했던 빈도수를 세는 방법으로 해시 맵(key: tag, value : [tag1, tag2, tag3, ...])를 구해 놓은 이후에, 최종 결과를 JSON형식으로 저장해 놓는다. 관련 태그 추출의 결과는 표1과 같으며, 이는 이후에 관련 게시글 및 사용자 추천에 활용될 것이다.

```
transactions = read(Transaction File)
uniq_tags = set(transactions)          // Unique tags
tag_relations = {}                     // Hash
for current_tag in uniq_tags:
  init tmp_vector (length: uniq_tags, value: 0)
  for transaction in transactions:
    if current_tag in transaction:
      for tag in transcation:
        if current_tag != tag:
          tmp_vector[tag] ++;
    result[current_tag] = sort_by_value(tmp_vector)
dump as json
```

(그림 4) 관련 태그 추출 의사 코드
(Figure 4) Pseudo code for extracting related tags

(표 1) 예시 결과
(Table 1) Sample Result

태그	관련 태그	빈도수	태그	관련 태그	빈도수
건축가	architect	127	드로잉	Sketch	472
	일본	72		drawing	449
	Japan	71		스케치	443
	Sketch	23		3D	111
	Museum	21		section	88
	Pavilion	18		단면	79
렌더	3D	101	파빌리온	Pavilion	78
	render	96		Architect	21
	Perspective	89		Memorial	14
	투시도	89		Japan	13
	렌더링	27		일본	13
	스케치	15		Model	12

3.3 관련 게시글 추천 방법

앞서 언급했듯이 본 연구에서 활용한 웹서비스에서는 그림 1과 같이 글 조회 시 해당 글과 관련된 글들을 추천해 주는 기능을 필요로 했다. 따라서 현재 조회하는 글이 가지는 태그집합과 유사도가 높은 태그집합을 가진 글들을 관련 글로 추천함으로써 사용자들에게 편의를 제공할 수 있다. 이는 앞서 구해 놓은 관련 태그를 이용한다. 관련 게시글의 추천 방법은 그림 5와 같다. 현재 조회하는 게시글의 태그집합을 얻어오고, 각 태그마다 상위 5개씩

의 관련 태그를 추출하여 이를 포함하는 관련 게시글 집합을 얻는다. 이렇게 얻은 관련 게시글 각각에 태그집합의 유사한 정도를 판단하는 식 1을 사용하여 임계점(thres_len)을 넘어서는 게시글을 추천 후보로 지정한다. 식 1은 현재 게시글에 속하는 태그의 집합 O_t와 비교 대상이 되는 게시글의 태그 태그집합 T_t의 교집합의 개수를 측정한다.

$$intersection(O_t, T_t) = Count(O_t \cap T_t) \quad (1)$$

```
thres_len = 1 <= n
originan_tagset = current_article.tagset
tag_relations = (calculated before)
articles = Articles which have at least 1 tag in tag_relations
result = [ ]
for article in articles:
   if intersection(original_tagset, article.tagset) >= thres_len
      result.append(article)
return result
```

(그림 5) 관련 게시글 추천 의사 코드
(Figure 5) Pseudo code for recommending related articles

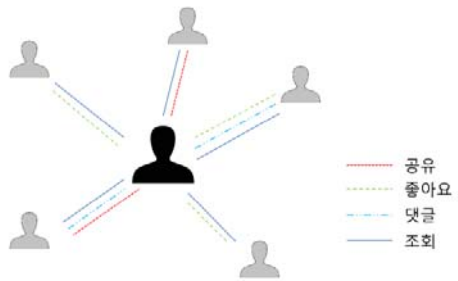
3.4 관련 사용자 추천 방법

사용자에게 비슷한 관심사를 가진 다른 사용자를 추천해 줌으로써 관심사를 공유하는 사용자 간의 활발한 커뮤니티 형성을 기대할 수 있다. 현재 사용자의 상호작용 기록을 조회하면 각 글들에 달려있는 태그들을 수집할 수가 있다. 이러한 태그의 집합을 마치 하나의 게시글에 달린 태그집합과 같이 생각한다면, 그림 6과 같은 방법으로 추천 후보들을 얻을 수 있다. 우리는 단순히 비슷하거나 관련된 사용자를 찾는 것에서 나아가 특히 영향력을 가진, 즉 컨텐츠를 많이 보유한 사용자를 추천해 줄수록 서비스를 이용하는 사용자 입장에서는 만족스러운 경험을 할 가능성이 높아질 것이라 생각했다. 따라서 그림 7과 같은 논리적인 그래프를 가정하였다.

특정 사용자는 자신이 남긴 게시글에 대한 댓글, 공유, 좋아요 등을 통해서 다른 사용자와의 상호작용을 갖게 된다. 이러한 상호작용이 활발한 사용자일수록 양질의 웹자원을 제공함과 동시에 다른 사용자와 긍정적인 상호작용을 주고받을 수 있을 것이라고 기대될 수 있다.

```
thres_len = 1 <= n
thres_score = 1 <= n
originan_tagset = user.action_history.tagsets
tag_relations = (calculated before)
articles = Articles which have at least 1 tag in tag_relations
result = [ ]
for article in articles:
   if intersection(original_tagset, user.action_history.tagsets) >= thres_len
      if score(user) > thres_score
         result.append(article)
return result
```

(그림 6) 관련 사용자 추천 의사 코드
(Figure 6) Pseudo code for recommending related users



(그림 7) 사용자 상호작용 그래프
(Figure 7) User Interaction Graph

따라서 위에서 그래프로 표현한 논리적인 관계를 수치화하여 사용자에 대한 점수를 매기고자 하였다. 점수는 식 2와 같이 사용자가 남긴 게시글의 공유 횟수, 좋아요 개수, 댓글 개수, 조회수를 단순 합산한 것으로 한다.

$$score(user) = shares + likes + comments + views \quad (2)$$

4. 실험 및 결과

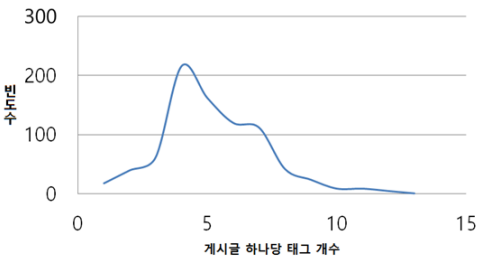
앞서 제안한 방법들을 실험하기 위해서, 웹서비스 아키필드의 5000개의 태그집합과 2000명의 사용자들에 대한 서비스와의 상호작용에 대한 기록을 얻어왔다. 결과적으로 간단한 알고리즘으로써 시스템 정확도면이나 사용자의 만족도 측면에서 좋은 결과를 보여주었다. 웹서비스에서 활용되는 추천 알고리즘이므로 추천 결과가 얼마나 정확한지를 유사도로써 측정하였고, 사용자 만족도를 측정하기 위해 구글 애널리틱스(Google Analytics)**[12]**를 활

용하여 사용자가 서비스에 머무르는 시간과, 연속적으로 방문하는 페이지가 얼마나 증가하였는지 등을 측정하였다. 구글 애널리틱스란, 웹로그 분석 툴로서 웹에서 발생하는 기록들을 분석해주는 기능을 제공한다.

4.1 관련 게시글 추천에 관한 실험

현재 조회 글이 가지고 있는 태그집합과 추천 결과로서 제안된 n개의 글들이 가지고 있는 태그집합을 비교해서 유사도를 측정 후 현재 글과 얼마나 유사한 글들이 추천되었는지를 식 3의 유사도를 이용해 판단한다.

그림 8에 보이는 것처럼, 게시글 하나당 달린 태그의 개수의 분포를 살펴보면 다음과 같다. 4~5개 보다 많아질수록 빈도수가 급격하게 낮아지는 것을 확인 할 수 있다. 이후에 이러한 태그 분포가 게시글 추천 여부를 결정하는 임계점(thres_len)에 간접적으로 어떠한 영향을 미치는지 확인할 수 있었다.

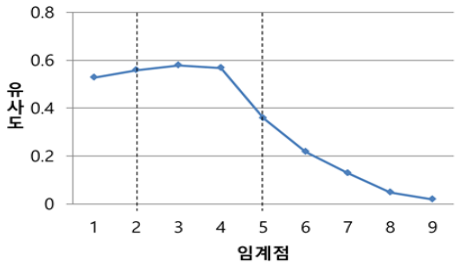


(그림 8) 태그 개수 분포
(Figure 8) Distribution of the number of tags

앞서 제시한 게시글 추천 방법을 적용해보면 다음과 같은 결과를 얻을 수 있다. 현재 조회하는 게시글에 대한 태그집합과, 각 추천 결과에 속하는 태그집합을 이용해 다음과 같은 Similarity식을 적용하였다. (O_t:Original Tag Set, R_{tk}:k-th Recommended Tag Sets)

$$Similarity(O_t,R_t) = \frac{1}{n}\sum_{k=1}^{n}\frac{Count(O_t \cap R_{tk})}{Count(O_t)} \quad (3)$$

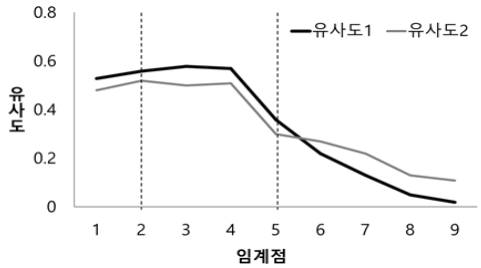
임계점을 높일수록 유사도가 증가할 것이라고 예상할 수 있지만, 오히려 그림 9에서 볼 수 있듯이 어느 선까지는 증가하다 이후에 감소하는 모양을 띄었다. 앞서 살펴본 태그집합의 길이에 대한 분포에서 알 수 있듯 3~5개 정도의 태그가 하나의 게시글에 달리는 빈도가 높았다. 따라서 임계점을 높여 게시글 간의 유사한 기준을 길게 할수록 오히려 유사도가 낮아지는 결과를 보인 것이다. 태그집합의 길이가 3~5개인 것이 대다수인데, 임계점을 7~9개로 잡아버리면 이 기준을 만족하는 후보들은 소수이기 때문이다. 따라서 적정한 임계점은 4이하라고 생각할 수 있다.



(그림 9) 임계점에 따른 유사도 변화
(Figure 9) similarity transition depending on the threshold

4.2 비교 실험

본 장에선 앞서 제안한 방법과 [8]에서 제안한 태그 기반 협업 필터링(Tag based collaborative filtering) 방법을 비교하였다. [8]에선 태그를 기반으로 사용자간의 유사도를 판단하고, 비슷한 사용자들이 관심을 가졌던 게시글을 추천하였다.



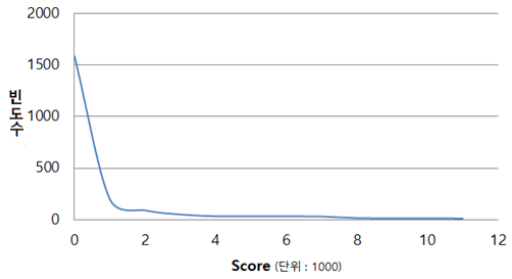
(그림 10) 임계점에 따른 유사도 변화 비교
(Figure 10) Comparison of similarity transition

[8]에선 본 연구에서 제안한 아이템 기반에 더하여 사용자간의 관계를 추가로 고려하여 추천을 하였다. 본 실험에서는 앞서 제안한 추천 방법(유사도1)에 사용자 관계 정보를 추가고려(유사도2) 하였을 때 추천 결과가 얼마나 개선되는지 보여주고자 하였다. 그림 10에서 볼 수 있듯이, 태그 길이에 관한 임계점을 높여 갈수록 유사도가 낮

아지는 경향을 보였다. 유사도가 낮은 구간에 대해서 친구 관계를 통해 얻어온 정보가 부분적으로 도움이 되는 구간이 있었다. 그러나 친구 관계를 이용하지 않더라도, 1~4정도의 임계점을 활용함으로써 충분히 높은 유사도를 얻을 수 있었기 때문에 본 연구에서 활용한 웹서비스**[4]**에서는 추가적으로 친구관계를 추천에 이용하지 않았다. 나아가 사용자 관계를 계산할 경우 추가적인 테이블 조인이 필요하므로 더 많은 계산 시간이 소비될 것이기 때문이다.

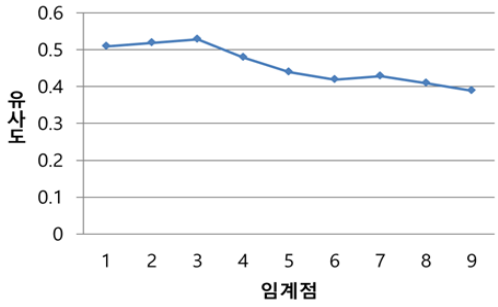
4.3 관련 사용자 추천에 관한 실험

사용자의 상호작용을 했던 각 게시 글들에 달려있는 태그들을 모은 이후에, 각 사용자를 하나의 게시 글로 생각한다면 게시글을 추천할 때 이용했던 방법으로 사용자에게 유사한 사용자를 추천해 줄 수가 있다. 단, 모은 태그에 대해서 중복이 발생할 경우마다 발생 빈도수를 기록한 후 횟수가 1인 태그는 노이즈로 생각하여 리스트에서 제외하였다. 앞서 제시한 사용자의 영향력을 측정하기 위한 식 2의 점수(score)의 분포를 얻어 보았다. 그림 11에 나타나는 것과 같이, 여러 사람의 상호작용을 받는 글을 적극적으로 게시하는 사용자를 1000점 이상이라고 가정했을 때, 그 비율은 전체 사용자 수의 10%정도에 불과했다. 따라서, 1000점을 점수에 관한 임계점(thres_score)으로 설정하였다.



(그림 11) 사용자 score 분포
(Figure 11) distribution of user's score

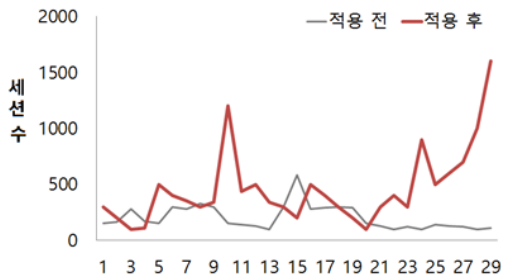
앞서 제시한 추천 방법을 통해 사용자에 대한 추천을 진행한 결과는 그림 12와 같다. 같은 임계점을 적용하였을 때, 게시글을 추천할 때와 다르게 상대적으로 급격히 기울기가 줄어들지 않는 이유는 사용자의 상호작용 리스트에 있는 태그들의 개수가, 게시글에 평균적으로 포함된 태그의 개수보다는 더 많을 수밖에 없기 때문이라고 해석할 수 있다. 따라서 사용자 추천에 관한 임계점(thres_len)은 3이하 정도가 적당하다고 판단 할 수 있다.



(그림 12) 임계점에 따른 유사도 변화
(Figure 12) similarity transition depending on the threshold

4.4 사용자 만족도 측정

본 연구에서는 사용자의 만족도에 관한 부분을 측정하기 위해 구글 애널리틱스(Google Analytics)**[12]**를 활용하여 사용자가 연속적으로 방문하는 페이지가 얼마나 증가하였는지, 방문 페이지에서 바로 사이트를 떠난 비율은 얼마나 낮아졌는지 등을 사용자의 만족도로서 측정하였다. 앞서 제안한 기능의 적용 전/후를 각각 1달을 기준으로 측정하여 비교하였다. 그 결과, 사용자가 가지는 하나의 세션당 연속으로 조회하는 페이지뷰 수가 향상되었고, 사용자의 서비스 이탈률을 73% 가량 줄일 수 있었다. 이탈률은 사용자가 세션에서 페이지와 상호작용하지 않고 사이트를 떠난 단일 페이지 세션의 비율을 의미한다.



(그림 13) 한 달간의 세션 수 변화
(Figure 13) transition of the number of session for a month

(표 2) 사용자 만족도 변화
(Table 2) transition of user's satisfaction

	적용 전	적용 후
이탈률	30.82%	8.25%
페이지 뷰	26601	59968
세션당 페이지 뷰	3.93	5.51

5. 결 론

우리는 태그를 기반으로 비슷한 게시글 혹은 사용자를 얻어 추천에 활용하였다. [2]에서의 언어 분포의 성질에 따라서 같은 게시글에 등장하는 태그는 비슷한 문맥적 의미를 가질 것이라는 가정 하에 태그간의 관계를 계산하여 추천에 적용하였다. 실험을 통해서 태그 관계에 기반을 둔 단순한 알고리즘으로 게시글 및 사용자를 추천이 가능함을 보였다. 비교연구에서는 추천에 있어서 태그 자체 외에 별도의 요소를 고려하지 않아도 좋은 성능을 보임을 확인하였다. 최종적으로 본 연구에서 제시한 웹 자원 추천 방법을 아키필드[4]에 적용하여 사용자의 서비스 만족도를 높이는데 기여할 수 있었다.

참 고 문 헌 (Reference)

[1] Recommendation System, http://rosaec.snu.ac.kr/meet/file/20120728b.pdf

[2] Zellig S. Harris, "Distributional Structure", WORD, Vol. 10:2-3, pp.146-162, 2015.

[3] Tomas Mikolov, "Distributed Representations of Words and Phrases and their Compositionality" Advanced in Neural Information Processing Systems 26, 2013.

[4] Archifeeld, http://feeeld.com

[5] Jo Hyeon, "A recommendation algorithm which reflects tag and time information of social network", Journal of Korean Society for Internet Information, v.14, no.2, pp.15-24, 2013

[6] Borkur Sigurbjornsson, "Flickr Tag Recommendation based on Collective Knowledge", pp.327-336, WWW, 2008

[7] Shilad Sen, "Tagommenders: Connecting Users to Items through Tags", pp. 671-680, WWW, 2009

[8] Sogol Naseri, "Enhancing tag-based collaborative filtering via integrated social networking information", pp. 760-764, ASONAM '13, 2013

[9] Frederico Durao, "A Personalized Tag-Based Recommendation in Social Web Systems", pp. 40-49, Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, 2009

[10] Rakesh Agrawal, "Fast Algorithm for Mining Association Rules"

[11] JIAWEI HAN, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", Data Mining and Knowledge Discovery, 8, pp. 53-87, 2004

[12] Google Analytics, https://www.google.com/analytics

◐ 저 자 소 개 ◑

송 제 인 (Je-In Song)
2016년 가천대 소프트웨어 설계 경영학과 졸업(학사)
2016~ 현재 가천대학교 일반대학원 소프트웨어학과 (석사과정)
관심분야 : 데이터 마이닝, 소셜 네트워크
E-mail : wpdls601@gc.gachon.ac.kr

정 옥 란 (Ok-Ran Jeong)
2005년 이화여자대학교 컴퓨터공학과 (공학박사)
2005년~2006년 서울대학교 컴퓨터공학부 (박사후 연구원)
2007년 Univ. of Illinois of Urbana Champaign (visiting scholar)
2008년~2009년 성균관대학교 정보통신공학부 (연구교수)
2009년~2015년 가천대학교 소프트웨어설계 경영학과 (조교수)
2015년~현재 가천대학교 소프트웨어학과 (부교수)
관심분야 : 웹 마이닝, 정보검색, 추천 시스템, 소셜 컴퓨팅
E-mail : orjeong@gachon.ac.kr