

효율적인 소셜 검색을 위한 토픽기반 소셜 관계 랭크 알고리즘

김 영 안^{*}, 박 건 우[°]

Topic Sensitive_Social Relation Rank Algorithm for Efficient Social Search

Young-an Kim^{*}, Gun-woo Park[°]

요 약

지난 10여 년간, 정보기술 분야의 패러다임은 기계중심에서 인간중심으로, 기술기반에서 사용자가 쉽게 정보시스템에 참여하고 활용 할 수 있는 사용자 기반으로 변화되었다. 즉 소셜 네트워크를 이용하여 정보를 상호 공유하는 소셜 검색의 형태로 변화하고 있으며, 이와 같이 사람과 사람을 연결 해 주는 소셜 네트워크 서비스는 웹 서비스와 융합을 통해 친구 맺기, 친구 찾기, 유사한 관심사를 갖고 있는 사람들과의 정보공유, 선호도 검색, 정보 추천시스템 등 다양한 분야에 활용되고 있다.

본 논문에서는 토픽 기반 소셜 관계 랭크(TS_SRR : Topic Sensitive_Social Relation Rank) 알고리즘을 제안한다. 제안 알고리즘은 소셜 네트워크 서비스를 웹 검색 엔진과 통합하는 것을 목적으로 하며, 소셜 관계 지수, 즉 Social Relation Rank와 검색 결과에 대한 선호도 사이의 상관관계를 분석하였다. 실험 과정에서 소셜 네트워크 안에 존재하는 일반적인 사람들은 정보 공유시 특정 분야에 대해 관심사가 유사할 경우 잘 모르는 타인들에 비해 친밀도가 높은 친구를 더 신뢰한다는 것을 확인 할 수 있었다. 따라서 제안 알고리즘은 소셜 검색의 신뢰성을 향상 시킬 수 있을 것으로 판단된다.

Key Words : Social Search, Topic, Social Relation Rank, Similar, Credible

ABSTRACT

In the past decade, a paradigm shift from machine-centered to human-centered and from technology-driven to user-driven has been witnessed. Consequently, Social search is getting more social and Social Network Service (SNS) is a popular Web service to connect and/or find friends, and the tendency of users interests often affects his/her who have similar interests. If we can track users' preferences in certain boundaries in terms of Web search and/or knowledge sharing, we can find more relevant information for users.

In this paper, we propose a novel Topic Sensitive_Social Relationship Rank (TS_SRR) algorithm. We propose enhanced Web searching idea by finding similar and credible users in a Social Network incorporating social information in Web search. The Social Relation Rank between users are Social Relation Value, that is, for a different topics, a different subset of the above attributes is used to measure the Social Relation Rank. We observe that a user has a certain common interest with his/her credible friends in a Social Network, then focus on the problem of identifying users who have similar interests and high credibility, and sharing their search experiences. Thus, the proposed algorithm can make social search improve one step forward.

* 주저자 : 국방대학교 국방과학학과 교수, roundsun@kndu.ac.kr, 정회원

° 교신저자 : 육군본부, pgw4050@emerald.yonsei.ac.kr, 정회원

논문번호 : KICS2013-02-099, 접수일자 : 2013년 2월 21일, 최종논문접수일자 : 2013년 5월 10일

I. 서 론

소셜 네트워크는 웹 환경에서 개인을 중심으로 구성원 간의 연결망을 이루고 있다^c. 기존 오프라인에서 친구나 동문회, 동호회를 중심으로 형성된 연결망이 웹으로 이동한 하나의 커뮤니티(Community) 형태이다²⁾. 또한 웹 사용자간 정보교환 등 필요에 의해서 인위적으로 맺어진 온라인상의 인간 관계망이라 할 수 있다. 최근 들어 이러한 인간 관계망은 관계망이 어떻게 형성되고 해체되는가와 형성된 관계망을 어떻게 유지하고 있는가에 대한 관계망 분석의 측면에서 핵심적인 연구가 이루어지고 있다¹⁾²⁾. 즉, 이러한 인간 관계망에 근간을 둔 소셜 네트워크¹⁾는 웹 환경에서 개인을 중심으로 웹 사용자 간 직접 또는 간접연결이라는 논리를 통해 상호 연결강도(link strength)를 표현 할 수 있다. 이는 개인과 유사한 내재적 정보를 가진 소셜 네트워크를 웹 검색³⁾에 적용한다면 검색 결과의 만족도를 향상 시킬 수 있는 중요한 요소로 작용 할 것이다.

본 논문에서는 일반적인 소셜 관계 랭크(SRR : Social Relation Rank) 및 토픽 기반 소셜 관계 랭크(TS_SRR : Topic Sensitive_Social Relation Rank) 알고리즘을 제안한다. 제안 알고리즘은 소셜 네트워크 서비스를 웹 검색 엔진과 통합하는 것을 목적으로 하며, 소셜 네트워크 서비스에서 활동하고 있는 사람들 중 성향이 유사하고 신뢰할 수 있는 사람들의 식별을 통해 정보 검색 효율을 향상시키기 위한 소셜 검색 알고리즘이다. 이를 위해 소셜 관계 지수, 즉 Social Relation Rank와 검색 결과에 대한 선호도 사이의 상관관계를 분석하였다. 소셜 관계 지수는 주제에 따라 변화하기 때문에 주제별 서로 다른 소셜 관계지수를 산출 한다. 실험 과정에서 소셜 네트워크 안에 존재하는 일반적인 사람들은 정보 공유시 특정 분야에 대해 관심사가 유사할 경우 잘 모르는 타인들에 비해 친밀도가 높은 친구를 더 신뢰한다는 것을 확인 할 수 있었다. 따라서 소셜 네트워크 안에서 유사한 관심사와 높은 신뢰성을 가지고 있으면서 친밀도가 높은 사용자를 식별하는데 중점을 두었고, 정보 요구 시 이와 같이 관심사가 유사하면서 신뢰할 수 있는 친구와 정보 검색 결과를 공유하고 상호 자신의 정보를 서로 공유 할 수 있도록 하는데 중점을 두었다. 실험 및 평가 결과 제안 알고리즘이 기존의 알고리즘들에 비해 높은 성능을 나타내는 것을 확인 할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대하여 기술하고, 3장에서는 토픽별 웹 사용자의

소셜 관계 지수 산정을 통한 소셜 관계 랭크 알고리즘을 제시하고, 4장에서는 제안하는 알고리즘을 통해 산출된 소셜 관계 랭크와 웹 사용자간 검색 결과 간 상관관계를 분석하고, 이를 통해 제안하는 알고리즘이 검색의 효율성과 신뢰성 향상에 중요한 영향을 미친다는 것을 검증한다. 마지막 5장에서는 결론 및 향후 연구를 기술한다.

II. 관련연구

2.1. 소셜 네트워크(Social Network)

소셜 네트워크는 하나 이상의 상호 의존적인 관계에 의해 구성된 개인 또는 집단으로 구성된 사회적 구조체(Social Structure)로 정의할 수 있다. 대표적인 SNS(Social Network Service)에는 Friendster¹⁾, Orkut²⁾, Facebook 한국의 Cyworld 등이 있다. 사회학, 통신공학, 경제학 등에서 폭넓게 연구 중인 SNA(Social Network Analysis)는 소셜 네트워크의 형태와 특성을 알고리즘 적으로 연구하는 것이다. 정보과학 분야에서는 기 구성된 소셜 네트워크의 현상을 웹 환경에 응용함으로써 검색의 효율성을 향상시키기 위한 연구가 활발히 수행되고 있다⁴⁾⁷⁾.

소셜 네트워크 분석에 대한 접근법은 전체 관계망에서의 위치와 그 효과를 측정하는 위치적 접근법(Positional Approach)과 연결망의 직접적인 관계에 초점을 둔 관계적 접근법(Relational Approach)이 있다. 기본적인 출발점은 행렬과 그래프를 이해하는 것이다. 즉 구성원 (i, j) 사이의 관계가 있고 없음을 '1'과 '0'으로 나타내는 행렬을 인접 행렬(Adjacency Matrix)이라 하며, 완전 연결망의 기본 형태이다[10]. 이러한 소셜 네트워크는 사용자의 명확한 참여 의사표시에 의하여 형성되고, 구성원 각각의 활동 정도가 서로 연관되어 연결성의 증강이 도모되는 것을 알 수 있다. 또한, 실제 사회에서와 유사한 형태(사람과 사람간의 관계 설정)로 구성되고 변화되는 것을 관찰할 수 있다. 이러한 특성 때문에 최근의 연구들은 소셜 네트워크의 구성원에 대한 조사를 토대로 그 성격을 규명하는 데 집중되어 있다.

2.2. 개인화된 검색 관련 랭킹 알고리즘

개인화된 검색 결과의 랭킹 관련 연구는 Personalized PageRank 방법⁸⁾, Topic_Sensitive PageRank 방법⁹⁾,

1) www.friendster.com
2) www.orkut.com

두 방법을 결합한 방법^[10] 등이 있다.

Topic_Sensitive PageRank^[8]는 수집된 문서를 주제어와 PageRank의 지수를 계산하여 주제별 문서의 순위를 구한다. 사용자의 질의어가 입력되면 질의어와 주제별 문서의 순위를 비교하여 수치화하고, ODP(Open Directory Project)³⁾의 16개의 카테고리를 이용하여 카테고리별 수치화된 문서를 순위화하여 랭킹을 부여하는 방법이다. 하지만 사용자 질의어의 하위 주제어를 선택할 때 확률을 기반으로 하고 있기 때문에 주제가 편향되어 선택될 수 있다.

Personalized PageRank^[9]는 대중적 가치를 기반으로 순위를 조정하여 PageRank에 개인화 개념을 추가한 알고리즘이다. 각 사용자는 관심 있는 문서의 집합을 선택한다. 선택된 문서의 집합들 중 의미 단어를 추출하여 Personalized PageRank Vector를 구성을 제안한다. 하지만 사용자에 의해 선택된 문서에서 관심어를 추출하여 벡터로 구성하는 것은 사용자가 원하지 않는 단어를 추출할 수 있기 때문에 많은 에러를 포함한다.

두 방법을 결합한 방법^[10]은 사용자에 의해 선택된 문서를 분석하여 초기의 사용자 프로파일로 설정하고, 프로파일을 학습하여 Topic_Sensitive PageRank에 적용하여 검색 결과를 랭킹하는 알고리즘을 제안했다. 하지만 사용자에 의해 선택된 문서에서 사용자 프로파일을 설정하고 학습하는 것은 사용자가 원치 않는 단어를 포함할 수 있기 때문에 많은 에러를 포함하여, 이러한 에러는 랭킹된 검색 순위에 영향을 미친다.

본 논문에서 제안한 토픽 기반 소셜 관계 랭크 알고리즘은 소셜 네트워크 서비스를 웹 검색 엔진과 통합하는 것을 목적으로 하며, 소셜 관계 지수, 즉 Social Relation Rank와 검색 결과에 대한 선호도 사이의 상관관계를 분석하였다.

III. 토픽기반 소셜 관계 랭크 알고리즘 제안

연결 관계를 맺고 있는 웹 사용자들은 모든 토픽에 대해 같은 관심사를 갖고 있지 않으며, 동일 질의(Query)에 대한 검색 결과의 만족도 또한 서로 다르다. 즉 웹 사용자마다 같은 질의를 입력하더라도 전혀 다른 관심사(Interest)나 검색의도(Search Intent)를 가지고 검색하기 때문이다. 그림 1은 'IT 분야'에 중사하는 '웹 사용자 A'에게 보고서에 삽입 할 '사람 얼굴'에 대한 이미지 검색 후 만족스러운 검색 결과에

'네모박스'를 표시하도록 하고, 나머지 검색 결과에 대해서는 'X' 표시를 하도록 한 결과 화면이다.

그림 1에서와 같이 '웹 사용자 A'는 2개의 검색 결과에 대해서만 만족하는 것을 알 수 있다. 만약 '의학계'에 중사하는 다른 웹 사용자가 논문작성에 필요한 이미지를 검색했다면 점선의 원형으로 표기된 검색 결과를 선택 할 것이다. 이때 선택한 검색 결과가 만족스럽다면 유사한 직종의 다른 웹 사용자가 동일한 목적으로 동일 질의를 입력할 경우 자신이 선택했던 검색 결과를 추천 해 줄 것이다. 따라서 소셜 관계지수 산정을 위해 사용되는 여러 속성들 중 하나인 '직업적'인 특성만 유사하더라도 원하는 정보 획득이 보다 용이해 질 수 있다.

따라서 소셜 네트워크 구성원 간의 소셜 관계 지수에 랭킹을 부여 하는 알고리즘을 개발하고, 소셜 관계 랭크와 검색 패턴 사이의 상관관계를 분석하여 웹 검색에 적용한다면 검색의 효율성과 신뢰성을 향상시킬 수 있을 것이다. 즉, '나'와 소셜 관계 랭킹이 높은 웹 사용자의 검색 패턴 또한 같다면 소셜 네트워크 환경에서 검색에 유용하게 활용 될 수 있을 것이다.



그림 1. 질의'두뇌' 검색 결과에 대한 웹 사용자 선택 화면
Fig. 1. Search Result about Query 'human brain'

3.1. 요소(Factor)와 속성(Attribute) 선정

소셜 관계 지수를 산정하기 각 개인을 특징지을 수 있는 요소와 속성을 선정 하는 것은 매우 중요하다. 소셜 관계 지수 산정을 위해 'N' 개의 요소(Similarity, Access, Intimacy, Adjacent, Search Behavior, etc.)가 선정 된다. 이때 각 요소들은 'Li' 개의 속성을 갖는다. 이와 같은 속성들은 대표적인 SNS 사이트인 Friendster, Orkut, Facebook, Cyworld 등에서^[8-9] 사용되는 일반적인 입력 요소와 웹 사용자들의 검색 행위(Search Behavior)에서 발생하는 클릭 히스토리(Click-Through Data : 질의, url 로그 등) 등으로 선정된다.

3) <http://www.dmoz.org>

3.2. 토픽(Topic) 선정

소셜 관계 지수(랭킹)는 웹 사용자들의 관심사에 따라 변화 된다. 즉 평소 토픽 ‘Art’에 대해 강한 소셜 관계를 형성하고 있는 웹 사용자일지라도 관심 토픽이 ‘스포츠’로 변하면 해당 분야에 대한 소셜 관계는 약하게 변하거나 관계 자체가 없을 수도 있다. 따라서 토픽에 따라 변화되는 소셜 관계지수를 산정하기 위해 ‘M’개의 서로 다른 토픽을 선정해야 한다.

3.3. 속성과 토픽 간의 연관성 행렬 생성

소셜 관계 지수 산정을 위해 웹 사용자들이 갖는 내재적 속성과 토픽 간에 상호 연관성을 부여하여야 하며, 이때 속성이 특정 토픽에 영향을 받으면 ‘1’, 영향을 받지 않으면 ‘0’이 부여된다. <그림 2>는 {‘L’(Attributes) x ‘M’(Topics)} 크기의 연관성 행렬 (Association Matrix)을 나타낸다.

그림 2. ‘L’ x ‘M’ 연관성 행렬
Fig. 2. Association Matrix

3.4. SRR(Social Relation Rank)

SRR은 토픽의 영향을 받지 않는 일반적인 웹 사용자 간의 소셜 관계 랭크 알고리즘으로 수식 (1)과 같다.

$$SRR_{me_user_i} = \alpha \cdot \frac{\sum_{j=1}^{l_1} si_j}{SI_i} + \beta \cdot \frac{\sum_{j=1}^{l_2} acc_j}{ACC_i} + \dots + \gamma \cdot \frac{\sum_{j=1}^{l_3} inti_j}{INTI_i} \quad (1)$$

- $a + \beta + \dots + \gamma = 1$, a, β, γ : Balance Factors
- $a = (\text{Total Attribute's number in Similarity}) / (\text{Total Attribute's number})$, $\beta = (\text{Total Attribute's number in Access}) / (\text{Total Attribute's number})$, ..., $\gamma = (\text{Total Attribute's number in Search Behavior}) / (\text{Total Attribute's number})$
- $A_j = \{si1, si2, \dots, silp, acc1, acc2, \dots, acc4q, inti1, inti2, \dots, intilr\}$, for each $i = 1, 2, 3, \dots, m$

- $si_j, acc_j, inti_j$: 웹 사용자 간 일치하는 속성들, 연관성 행렬에서 모든 행벡터(Row Vector)는 ‘1’
- $SI_i, ACC_i, INTI_i$: 토픽에 영향을 받는 속성들의 개수에 대한 총합

3.5. TS_SRR(Topic Sensitive_Social Relation Rank)

TS_SRR은 토픽의 영향을 받을 때 산정되는 웹 사용자 간의 소셜 관계 랭크 알고리즘으로 수식 (2)와 같다.

$$TS_SRR_{me_user_i} = \alpha \cdot \frac{t_i \cdot \sum_{j=1}^{l_1} si_j}{SI_i} + \beta \cdot \frac{t_i \cdot \sum_{j=1}^{l_2} acc_j}{ACC_i} + \dots + \gamma \cdot \frac{t_i \cdot \sum_{j=1}^{l_3} inti_j}{INTI_i} \quad (2)$$

- $a + \beta + \dots + \gamma = 1$, a, β, γ : Balance Factors
- $A_j = \{si1, si2, \dots, silp, acc1, acc2, \dots, acc4q, inti1, inti2, \dots, intilr\}$, for each $i = 1, 2, 3, \dots, m$
- t_i : 연관성 행렬의 행벡터
- $T_i = \{[ti, 0, 0, 0, 0, \dots; 0] = [1, 0, 0, 0, 0, \dots; 0], [0, ti, 0, 0, 0, \dots; 0] = [0, 1, 0, 0, 0, \dots; 0]\}$, where the value ‘1’ appears just only in the i th column
- $A_{j_topici} = [1, 1, 0, 0, 1, \dots, 1]$, where the value ‘1’ appears in the j th column
- $si_j, acc_j, inti_j$: 웹 사용자간 일치하는 속성에 대해 ‘0’과 ‘1’로 표현한 행벡터와 연관성 행렬의 행벡터 간 논리곱(AND) 연산을 통해 생성된 행벡터에서 ‘1’의 값을 갖는 속성
- $SI_i, ACC_i, INTI_i$: 토픽에 영향을 받는 속성들의 개수에 대한 총합

IV. 실험 및 평가

4.1. 데이터 셋(Data Set)

대표적인 SNS 사이트(Friendster, Orkut, Facebook, Cyworld, etc.)에서^[5-6] ‘나’와 관련된 있는 250명의 회원들을 이용하여 소셜 네트워크를 구성한다. 그중 50명은 ‘나’와 직접적인 관련을 맺고 있는 회원이며, 기타 인원은 간접적인 관계를 맺고 있는 회원이다. 또한 ‘나’ 중심의 소셜 네트워크가 아닌 다른 회원들을 중심으로 250개의 서로 다른 소셜 네트워크를 구성하여 실험 및 평가를 실시한다. 소셜 관계 랭크 알고리즘을 일반화하기 위해 소셜 네트워크는 한국인을 포함하여 다양한 나라(일본, 태국, 중국, 미국,

캐나다, 오스트레일리아, 스페인 등) 사람들로 구성한다.

4.2. 소셜 관계 지수 산정

4.2.1. 요소 및 속성 선정

실험을 위해 ‘N’개의 요소와 ‘L’ 개의 속성을 ‘오르컷(Orkut)’의 일반적인 필수 입력요소를 참조하여 각각 ‘2’, ‘11’개로 제한하였다. 그림 3은 ‘오르컷’^[5]의 입력 프로파일을 나타낸다.

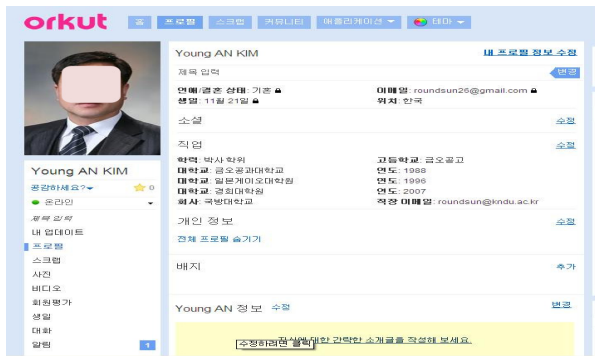


그림 3. ‘오르컷’의 입력 프로파일
Fig. 3. Input Profile of Orkut

4.2.2. 토픽선정

실험을 위해 ‘M’개의 토픽을 ‘오르컷’의 ‘ODP’를 참조하여 ‘15’개로 제한하였다. 그림 4는 ODP의 15개 디렉토리를 나타낸다.

SNS를 이용하기 위한 입력 항목인 개인 프로파일과, 다음으로 모든 주제를 포함할 수 있도록 정제되어져 있는 ODP에 해당하는 15개의 주제를 정의한다.

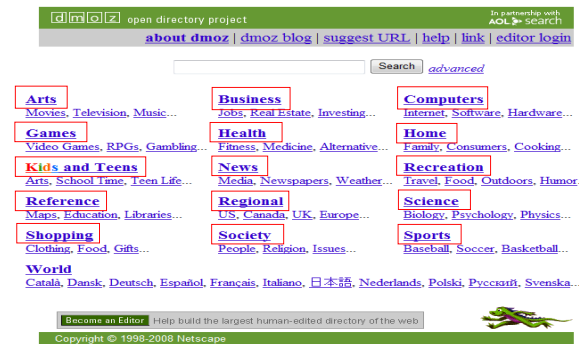


그림 4. ODP의 15개 디렉토리
Fig. 4. ODP Fifteen Directory

표 1. 11×15 연관성 행렬

Table 1. ‘11’ X ‘15’ Association Matrix

Factor	Similarity											Access		
	Gender	Relationship	Birth Date	Nation City	Language	High School College	Major	JOB Type	Interest	Temporal Interest	Sum (Sim)	Friend	Join Group	Sum (Acc)
Arts	1	0	0	1	1	1	1	1	1	1	8	1	1	2
Business	0	0	1	1	0	1	1	1	1	1	7	1	1	2
Computer	1	1	0	1	1	0	0	0	1	1	6	1	1	2
Game	1	0	1	1	0	0	0	0	1	1	5	1	1	2
Health	1	0	1	1	0	0	0	1	1	1	6	1	1	2
Home	1	1	1	1	1	0	0	1	1	1	8	1	1	2
Kids & Teens	1	1	1	1	1	0	0	0	1	1	7	1	1	2
News	0	0	0	1	1	1	1	1	1	1	7	1	1	2
Recreation	0	1	1	1	0	1	1	1	1	1	8	1	1	2
Reference	0	1	1	1	1	1	1	1	1	1	9	1	1	2
Regional	1	0	0	1	1	0	0	1	1	1	6	1	1	2
Science	1	0	1	1	1	1	1	1	1	1	9	1	1	2
Shopping	1	1	1	1	0	0	0	1	1	1	7	1	1	2
Society	1	0	1	1	1	0	0	1	1	1	7	1	1	2
Sports	1	1	1	1	0	0	0	0	1	1	6	1	1	2

4.2.3. 연관성 행렬 생성

본 연구에서는 사용자의 속성(Attributes)에 내재된 정보를 이용하여 주제 별 사용자 간 유사성 산정을 위해 위절에서 언급한 사용자의 속성인 개인 프로파일, 즉 성별, 관계현황, 생년 등 11개의 항목과 Arts, Business 등 15개의 주제에 대한 상호 연관성 정의가 우선시 되어야 한다.

실험을 위해 {L(속성 수)×M(토픽 수)} 연관성 행렬 중 2개 요소에 대한 (11×15) 크기의 연관성 행렬로 제한하였다. 표 1은 (11×15) 연관성 행렬을 나타낸다.

4.2.4. TS_SRR 알고리즘

성과 접근성의 두 가지 요소를 이용하여 일반적인 소셜 관계 지수와 토픽이 반영된 소셜 관계 지수를 산정하고 랭킹을 부여한다. 세부적인 변수들은 식(1), (2)에서와 같다.

- SRR_{(me_user)(none_topic)}

$$= SRR_{me_user_i} = \alpha \cdot \frac{\sum_{j=1}^{l_1} s_{ij}}{SI_i} + \beta \cdot \frac{\sum_{j=1}^{l_2} acc_j}{ACC_i} \quad (3)$$

- TS_SRR_{(me_userk)_topic}

$$= TS_SRR_{me_user_i} = \alpha \cdot \frac{t_i \cdot \sum_{j=1}^{l_1} s_{ij}}{SI_i} + \beta \cdot \frac{t_i \cdot \sum_{j=1}^{l_2} acc_j}{ACC_i} \quad (4)$$

• α (Attribute’s total number in Similarity / Total Attribute’s number = 9/11 = 0.82), β (Attribute’s total number in Access / Total Attribute’s number = 2/11 = 0.18)

4.2.5. 소셜 관계 지수 산정

표 2. 토픽별 소셜 관계 지수 산정
Table 2. Topic selected Social Relation Rank

Topic	user 1	user 2	user 00	user 12	user 13	user 00	user 33	user 34	user 00	user 49
Arts	0.50	0.68	...	0.65	0.30	...	0.44	0.41	...	0.47
Business	0.59	0.64	...	0.59	0.32	...	0.23	0.45	...	0.55
Computer	0.84	0.58	...	0.67	0.67	...	0.58	0.50	...	0.49
Game	0.80	0.50	...	0.80	0.39	...	0.30	0.59	...	0.41
Health	0.67	0.75	...	0.67	0.44	...	0.25	0.51	...	0.49
Home	0.77	0.68	...	0.65	0.53	...	0.44	0.53	...	0.59
Kids & teens	0.86	0.64	...	0.73	0.59	...	0.50	0.59	...	0.55
News	0.59	0.64	...	0.59	0.32	...	0.36	0.45	...	0.55
Recreation	0.65	0.56	...	0.53	0.41	...	0.32	0.41	...	0.59
Reference	0.69	0.60	...	0.59	0.39	...	0.40	0.49	...	0.62
Regional	0.67	0.75	...	0.67	0.34	...	0.42	0.51	...	0.49
Science	0.59	0.71	...	0.69	0.39	...	0.40	0.49	...	0.51
Shopping	0.73	0.64	...	0.59	0.43	...	0.36	0.45	...	0.55
Society	0.86	0.77	...	0.73	0.45	...	0.36	0.59	...	0.55
Sports	0.84	0.52	...	0.67	0.51	...	0.42	0.51	...	0.49

‘나’ 중심의 소셜 네트워크를 구성 한 후 제안한 알고리즘에 따라 소셜 관계지수를 산정했다. 표 2는 토픽별 사용자 간 소셜 관계 지수를 산정한 결과를 나타낸다.

표 2에서 나타나듯이 토픽 ‘Arts’와 ‘Business’ 분야에서 ‘나’와 ‘user2’ 간의 소셜 관계 지수가 ‘0.68’, ‘0.64’로 가장 크지만, 토픽 ‘Computer’ 및 ‘Game’ 분야에 대해서는 ‘user1’과 ‘user12’의 소셜 관계 지수가 ‘0.84’, ‘0.80’으로 ‘user2’ 보다 더 큰 것을 확인할 수 있다.

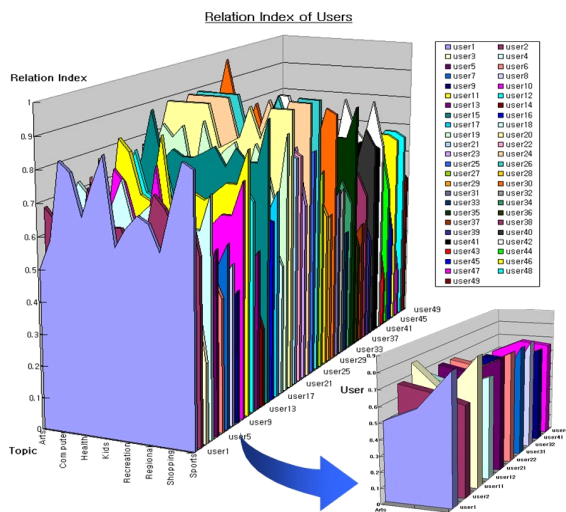


그림 5. 토픽에 따른 소셜 관계 지수 변화
Fig. 5. Social Relation Rank according to Topics

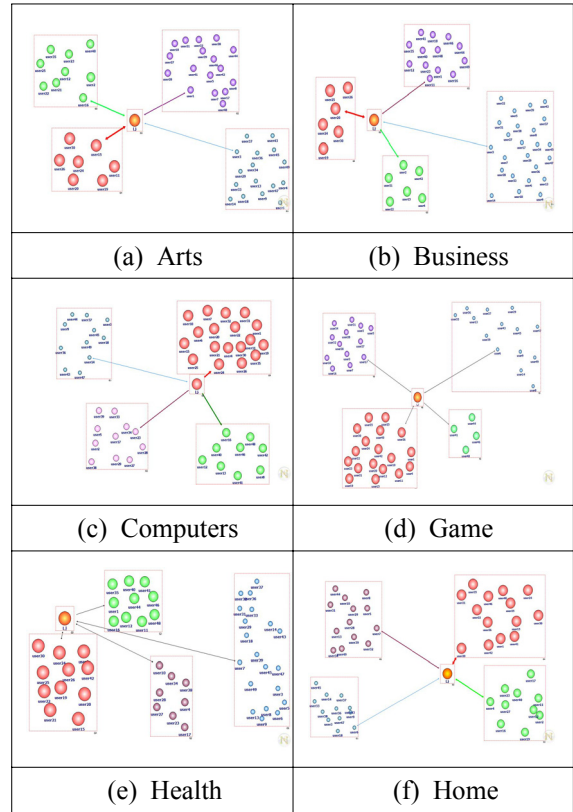
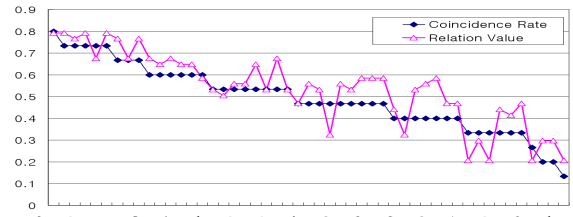
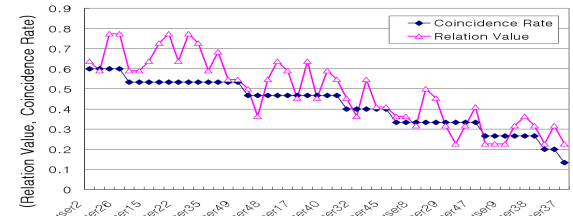


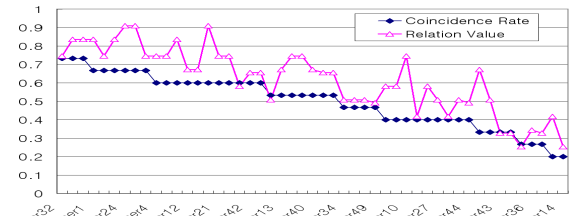
그림 6. 소셜 관계 지수에 따른 웹 사용자 클러스터링
Fig. 6. Web users Clustering according to Social Rank



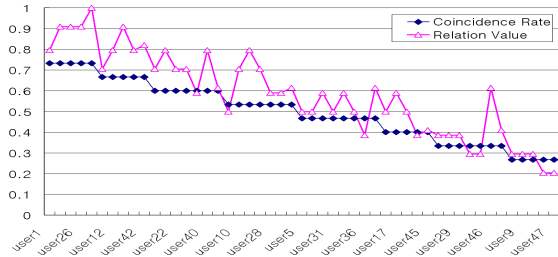
(a) Arts



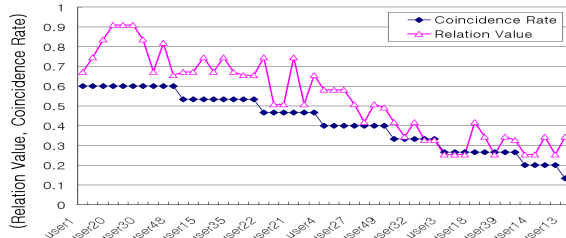
(b) Business



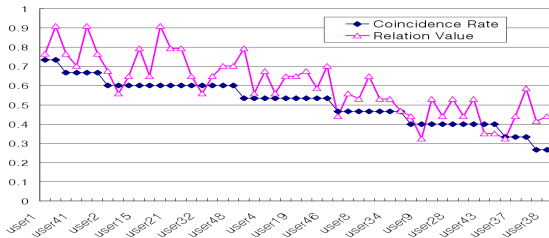
(c) Computers



(d) Game



(e) Health



(f) Home

그림 7. 소셜 관계 랭킹과 검색 패턴간의 상관관계
Fig. 7. Correlation between Social Relation Rank and Search Pattern

그림 5, 그림 6은 각각 토픽별 변화되는 소셜 관계 지수와 소셜 관계 지수별 상위에 랭크 된 웹 사용자로부터 하위에 랭크 된 웹 사용자까지 클러스터링 한 결과를 나타낸다. 그림에서 볼 수 있듯이 웹 사용자들의 관계 지수는 토픽에 따라 변화되며 이와 같은 사실을 통해 웹 사용자들은 특정 분야에 대해 검색하고자 할 경우에는 해당 분야마다 소셜 관계 지수가 높은 웹 사용자들을 재탐색 할 필요가 있다.

4.3. 검색 패턴 분석

ODP의 15개 토픽에 대해 각 토픽별 선호도가 높은 5개의 질의를 선정, 총 75개의 질의를 선정한다.

토픽별 소셜 관계 랭킹이 높은 웹 사용자들은 검색 패턴 또한 유사함을 검증하고자 소셜 관계 랭킹을 기반으로 검색 결과에 대한 선택 일치율을 비교 실험하였다. 즉, 토픽별 검색 결과 중 만족하는 검색 페이지를 15개씩 선택하게 하여 웹 사용자 상호간 선택 결과에 대한 일치 여부를 확인 하였다.

4.4. 소셜 관계 랭킹과 검색 패턴간의 상관관계 분석

그림 7은 6개의 토픽에 대해 소셜 관계 랭킹과 검색 패턴간의 상관관계를 분석한 결과를 나타낸다. 그림에 나타나듯이 소셜 관계 랭킹이 높은 웹 사용자들간의 검색 패턴이 소셜 관계 랭킹이 낮은 웹 사용자들에 비해 더 유사한 것을 확인 할 수 있다. 즉, ‘나’와 소셜 관계 랭킹과 검색 패턴 간에 상관관계가 매우 높은 user26, user20, user11, user2, user35, user27은 검색결과에 대한 일치율 또한 높게 나타남을 알 수 있다. 특히 그림을 통해 주제별 사용자의 유사성이 높고 낮음에 따라 검색 결과의 만족도 또한 비슷한 곡선의 형태를 나타내고 있음을 확인할 수 있다. 결과에서 나타나듯이 나머지 5개의 토픽에 대해서도 유사한 결과를 나타낸다.

4.5. 스마트 웹 검색

그림 8은 본 논문에서 제안하는 알고리즘을 적용하여 토픽별, 요소별 랭킹에 따른 추천 목록이 제시되는 것을 보여주고 있다. 제안 알고리즘의 최종 목표는 관심 분야별 소셜 관계 지수가 높은 사람들을 검색하여 추천해주며, 이와 같이 추천된 사람들로부터 검색 패턴을 상속 또는 공유하는 것이다.

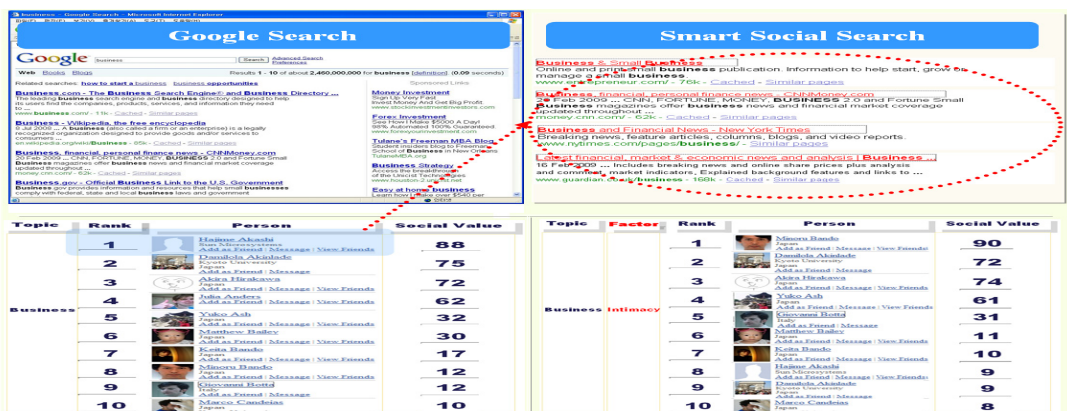


그림 8. 스마트 웹 검색 결과
Fig. 8. Result of Smart Web Search

V. 결 론

본 논문에서는 소셜 네트워크를 이용하여 웹 사용자 간의 소셜 관계 지수를 산정하여 랭킹을 부여하고 이들 사이의 검색 결과를 분석 하였다. 즉 소셜 관계 지수 기반으로 소셜 네트워크를 형성하고 선정한 질의를 통해 검색된 결과에 대한 웹 사용자들의 웹 페이지 선택 일치율을 비교 분석 함으로써 토픽별 개인과 소셜 관계 랭킹이 높은 웹 사용자 간에는 검색 패턴 또한 유사하다는 사실을 검증하였다. 실험 과정에서 소셜 네트워크 안에 존재하는 일반적인 사람들은 정보 공유시 특정 분야에 대해 관심이 유사할 경우 잘 모르는 타인들에 비해 친밀도가 높은 친구를 더 신뢰한다는 것을 확인 할 수 있었으며, 소셜 네트워크 안에서 유사한 관심사와 높은 신뢰성을 가지고 있으면서 친밀도가 높은 사용자를 식별하는데 중점을 두었다. 따라서 제안 알고리즘을 개인화 및 소셜 검색에 적용한다면 효율성 및 신뢰성을 향상 시킬 수 있을 것이다.

향후에는 사용자 속성과 검색에 있어서의 행동 패턴을 분석하여 실제 Google 검색엔진, SNS 서비스에 적용함으로써 검색의 효율성과 신뢰성을 향상 시킬 수 있는 방안을 위해 제안 알고리즘에 대한 복잡도 분석과 동일 클러스터에서 선택된 페이지와 다른 클러스터에서 선택된 페이지들에 대한 만족도 평가에 대한 지속적인 연구가 수행 되어져야 할 것이다.

Reference

- [1] J. S. Kim, "Human Network," *Policy of Inform. Commun.*, vol. 16, no. 16, pp. 1-19, Sep. 2004.
- [2] E. Lessor, D. Ransom, R. Sbab, and B. Pulver, "IBM collective intelligence white paper," *IBM Global Business Service*, Jan. 2012.
- [3] Y. A. Kim, S. G. Park, and S. H. Lee, "Design of efficient web search system based on reliable online community," *J. KICS*, vol. 37B, no. 1, pp. 67-74, Jan. 2012.
- [4] M. V. Vieira, B. M. Fonseca, R. Damazio, P. B. Golgher, D. C. Reis, and B. R.-Neto, "Efficient search ranking in social networks," in *Proc. 16th ACM Conf. Inform. Knowledge Manage. (CIKM 2007)*, pp. 563-572, Lisbon, Portugal, Nov. 2007.
- [5] A. Mohan, Z. Chen, and K. Weinberger, "Web-search ranking with initialized gradient boosted regression trees", in *Proc. JMLR Workshop and Conf. Proc.*, vol. 14, pp. 77-89, Haifa, Israel, June 2011.
- [6] E. Spertus, M. Sahami, and O. Buyukkokten., "Evaluating similarity measures: A large-scale study in the Orkut social newtork," in *Proc. 11th ACM Int. Conf. Knowledge Discovery in Data mining (ACM SIGKDD)*, pp. 678-684, Chicago, U.S.A., Aug. 2005.
- [7] C. Dwyer, S. R. Hiltz, and K. Passerini., "Trust and privacy: A comparison of Facebook and MySpace," in *Proc. 13th Amer. Conf. Inform. Syst. (AMCIS)*, paper 339, Keystone, U.S.A., 2007.
- [8] F. Yasuhiro, N. Makoto, and Y. Takeshi, "Efficient personalized pagerank with accuracy assurance," in *Proc. 18th ACM Int. Conf. Knowledge Discover and Data Mining (SIGKDD 2012)*, pp. 15-23, Beijing, China, Aug. 2012.
- [9] S. Hatakenaka and T. Miura, "Query and topic sensitive pagerank for general documents," in *Proc. 14th IEEE Int. Symp. Web Syst. Evolution(WSE)*, pp. 97-101, Trento, Italy, Sep. 2012.
- [10] Y. Zeng, H. Hao, N. Zhong, X. Ren, and Y. Wang, "Ranking and Combining Social network data for web personalization," in *Proc. 21th ACM Conf. Inform. Knowledge Manage. (CIKM 2012)*, pp. 15-18, Maui, U.S.A., Oct. 2012.
- [11] Y. H. Kim, *Social Network Analysis*, ParkYoung Company, 2003.

김 영 안 (Young-an Kim)



1988년 2월 금오공과대학 전
산학과 학사

1996년 3월 Keio University
Dept. of Information and
Computer Science 석사

2008년 2월 경희대학교 컴퓨
터공학과 박사

2009년 2월~현재 국방대학교

국방과학학과 교수

<관심분야> Ad-hoc Network, Routing Protocol,
DTN, VANET, WMN, 정보검색, SNA

박 건 우 (Gun-woo Park)



1997년 2월 충남대학교 컴퓨터
과학과 학사

2007년 2월 연세대학교 컴퓨터
과학과 석사

2011년 8월 국방대학교 국방과
학학과 박사

2012년 10월~현재 육군본부

정체단 정보통신망 사업담당

<관심분야> 정보검색, 네트워크 보안, SNA