

A Method on Retrieving Personalized Information Based on Mutual Trust in Real and Online World

Myeonghun Kim[†] · Sangwook Kim^{**}

ABSTRACT

Two remarkable problems of recent online social network are information overflow and information overload. Since the mid-1990s, many researches to overcome these issues have been conducted with information recommender systems and context awareness based personalization techniques, the importance of trust or relationship between users to discover influential information has been increasing as recent online social networks become huge. But almost researches have not regarded trust or relationship in real world while reflecting them in online world. In this paper, we present a novel method how to discover influential and spreadable information that is highly personalized to a user. This valuable information is extracted from an information set that consists of lots of information user missed in the past, and we assumes important information is likely to exist in this set.

Keywords : Trust Measure, Social Network, Personalized Information, Recommender System

현실과 가상 세계에서 상호 신뢰도에 기반한 개인화 정보의 식별 방법

김명훈[†] · 김상욱^{**}

요약

최근 소셜 네트워크의 2가지 주목할 만한 특징은 바로 정보의 홍수와 정보 소비의 과부하 문제이다. 1990년대부터 상황 인지 기반의 개인화 기법들이나 추천 시스템을 통하여 이러한 문제들을 해결하고자 하였는데, 특히 커뮤니티 멤버간의 신뢰도나 친밀도를 기반으로 개인화 정보를 발견하기 위한 접근방법이 최근 소셜 네트워크가 거대화됨에 따라 그 중요성이 부각되고 있다. 그러나 이러한 대부분의 접근법들은 현실이 아닌 온라인 세계에서 상호 신뢰도(또는 친밀도)만을 고려하는 경향이 있어 보다 더 높은 개인화 성능(적절성)을 가지기 위해서는 현실 세계까지 고려할 필요가 있다. 본 연구에서는 높은 적절성을 가지는 개인화 정보를 식별하는 새로운 방법을 제시하며, 이러한 개인화 정보를 발견하기 위하여 현실 세계와 온라인 세계에서 정보 제공자와 수신자의 상호 신뢰도를 측정하고 이를 기반으로 정보의 중요도를 판단한다. 그리고 본 기법은 개인에게 신규로 제공되는 정보가 아닌 과거에 소비하지 못하고 놓친 정보들을 대상으로 적용한다.

키워드 : 신뢰도 측정, 소셜 네트워크, 개인화 정보, 정보 추천 시스템

1. 서론

소셜 네트워크 서비스(SNS)의 최근 발전 방향은 개인화(Personalization)와 지능화이다. 여기서 개인화와 지능화는 소셜 네트워크의 구성원에게 전달될 정보를 각 개인의 선호

컨텐츠, 주변 노드와의 친화도 등 다양한 개인 중심의 상황 정보를 고려(Context Awareness)하여 최적으로 적합하다고 판단되는 정보를 제공하는 것을 말하며[1], 상황 인지 기반의 추천 기법과 밀접한 관련성을 가지며 발전하고 있다.

이러한 특징은 소셜 네트워크 서비스의 거대화에서 비롯된다. 실례로 현재 Facebook에서 생성된 모든 정보의 개수는 약 2조를 상회하며, 활발히 활동하는 사용자의 숫자는 약 5억 명으로 추정되어 산술적으로 각 사용자에게 전달되는 총 정보의 개수는 최소 4천개가 넘는다. 또한 이런 개인에게 전달되는 총 정보의 양의 증가 속도는 점차 상승 추세를 보이고 있다[2]. 이는 극소수의 활발한 활동을 하는 사용자들을 제외한 절대 다수의 일반 사용자에게는 정보 소비 행위가 부담으로 작용할 수 있기 때문에 개인에게 중요하다고

* 본 논문(저서)은 교육부 및 한국연구재단의 BK21 플러스 사업(경북대학교 컴퓨터학부 Smart Life 실현을 위한 SW 인력양성사업단)으로 지원된 연구임(21A20131600005).

** 이 논문은 2016년도 한국정보처리학회 추계학술발표대회에서 '사회적 영향력과 여의 유사도 분석에 기반한 가치정보의 추천 기법'의 제목으로 발표된 논문을 확장한 것임.

[†] 준회원: 경북대학교 컴퓨터학부 박사과정

^{**} 정회원: 경북대학교 컴퓨터학부 교수

Manuscript Received: December 23, 2016

Accepted: January 10, 2017

* Corresponding Author: Sangwook Kim(kimsw@knu.ac.kr)

판단되는 개인화 정보를 식별하여 제공할 필요가 있다.

정보의 오버플로우(Overflow)[3]와 오버로드(Overload)[4]라는 대표적인 2가지 문제의 해결을 위하여 본 연구에서는 온라인 소셜 네트워크에서 각 개인에게 높은 중요도를 가지는 정보를 정보 제공자와 정보 수신자 상호간의 신뢰도(Trust)를 기반으로 식별하는 기법에 대해 제시한다. 각 개인의 온라인 세계(Online World)의 상황 정보를 기반으로 하는 대부분의 기존 기법들과는 달리 현실 세계(Real World)에서의 상호 신뢰도에 기반 한 상황 정보까지 반영하여 더욱더 개인에게 적절한 정보를 식별하는 기법을 제안하고, 또한 이 적절한 정보가 새롭게 개인에게 제공된 정보가 아닌 기존에 제공되었지만 사용자가 놓친 정보의 무리 속에 있음을 추가적으로 밝히고자 한다.

참고로 본 연구에서 정의하는 현실 세계에서의 신뢰도는 통화, 문자, 이메일 3가지의 커뮤니케이션을 통해 정의된다. 비록 온라인을 매개로 하고 있을지라도 그 내용은 대부분 개인의 현실 생활을 반영하기 때문이다. 구전 방식의 커뮤니케이션은 정보를 디지털화 하기 어렵기 때문에 본 연구에서는 제외하였다.

2. 관련 연구

온라인 세계 기반의 상호간 사회적 친화도 또는 신뢰도를 측정하는 대표적인 기법으로 TidalTrust[5], MoleTrust[6], SocialMF[7], LOCALBAL[8] 등이 있다. 이 중에서 J.A. Golbeck이 제안한 TidalTrust는 온라인 소셜 네트워크의 구조적인 관점에서 인접하지 않은 두 사용자 간의 신뢰도를 추론하는 방법으로, Fig. 1과 같이 한 사용자를 가리키는 Source Node와 신뢰도 측정 대상이 되는 Sink Node 간의 $trust_{ac}$ 를 중간 노드 b 를 통하여 점진적으로 추론하는 기법이다.

이 신뢰도 추론 알고리즘은 Equation (1)로 표현되는데, Source Node가 Sink Node를 모를 때, Source Node는 자신의 모든 인접 노드들에게 얼마나 Sink Node를 신뢰하는지를 질의하고, 만약 인접 노드가 Sink Node를 모르는 경우 이 인접 노드를 기준으로 또 다른 인접 노드들에게 동일 과정을 반복하여 최종적인 Source Node와 Sink Node의 신뢰도를 측정한다. $adj(j)$ 는 Source Node a 에 연결된 모든 인접 노드(Adjacent Node)를, $trust_{ab}$, $trust_{bc}$, $trust_{ac}$ 는 각각 노드 a 와 b , 노드 b 와 c , 노드 c 와 a 사이의 신뢰도를 나타낸다.

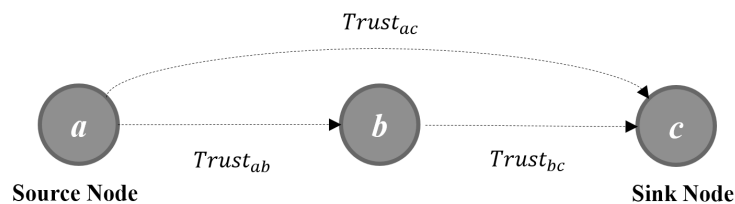


Fig. 1. Simplified Concept of TidalTrust

$$trust_{ac} = \frac{\sum_{j \in adj(j) | Trust_{ab} \geq max Trust_{ab} Trust_{bc}} Trust_{ab} Trust_{bc}}{\sum_{j \in adj(j) | Trust_{ab} \geq max Trust_{ab}} Trust_{ab}} \quad (1)$$

D.M. Romero가 제안한 IP(Influence-Passivity) 알고리즘[9]은 트위터(Twitter) 네트워크에서 정보 제공자의 영향력(I_i)과 정보 제공자에 대한 수신자의 수동성(P_j)을 측정하여 두 사용자간의 신뢰도를 정량적으로 판단하였다. 아래 Equation (2)와 Equation (3)은 영향력과 수동성을 측정하는 방법을 제공한다. u_{ij} 는 정보 수신자 i 가 정보 제공자 j 의 정보를 받아들이는 비율(Acceptance Rate)을, v_{ji} 는 거절하는 비율(Rejection Rate)을 의미한다.

$$I_i \leftarrow \sum_{j: (i,j) \in E} u_{ij} P_j \quad (2)$$

$$P_i \leftarrow \sum_{j: (j,i) \in E} v_{ji} I_j \quad (3)$$

상기 언급한 기법 이외에 다른 대표적인 기법으로 Facebook의 News Feed 알고리즘이 있다. $Feed = I \times P \times C \times T \times R$ 로 표현되는 정보의 노출 기법에 포함된 인자 중 C(Creator)는 정보 제공자가 수신자에게 미치는 영향력, 즉 상호 관계를 나타내는 변수이며, 비공개 되어있는 세부 알고리즘은 약 10만개의 사용자 상황 정보를 반영한다고 알려져 있다[10].

본 연구에서는 사용자의 온라인 환경에서의 상황 정보를 반영하는 위의 기법들과는 달리, 신뢰도를 추정하기 위하여 온라인 세계와 현실 세계의 상황 정보를 모두 반영하여 더욱더 높은 정확도를 가지는 신뢰도 측정 기법을 제안하고, 이 기법을 바탕으로 온라인 소셜 네트워크의 수많은 정보 중에서 사용자에게 중요한 정보를 식별하고자 한다.

3. 신뢰도 추정 기법

3.1 T-Value의 정의

T-Value(Trust value)는 온라인 소셜 네트워크에서 정보 수신자의 정보 제공자에 대한 신뢰도를 정량적으로 나타내며 Equation (4)와 같이 정의한다. 여기서 신뢰도는 정보 제공자의 정보에 관하여 수신자가 얼마나 민감한 반응을 보이는지를 나타내는 개념으로 구체화하며, 민감한 정보일수록

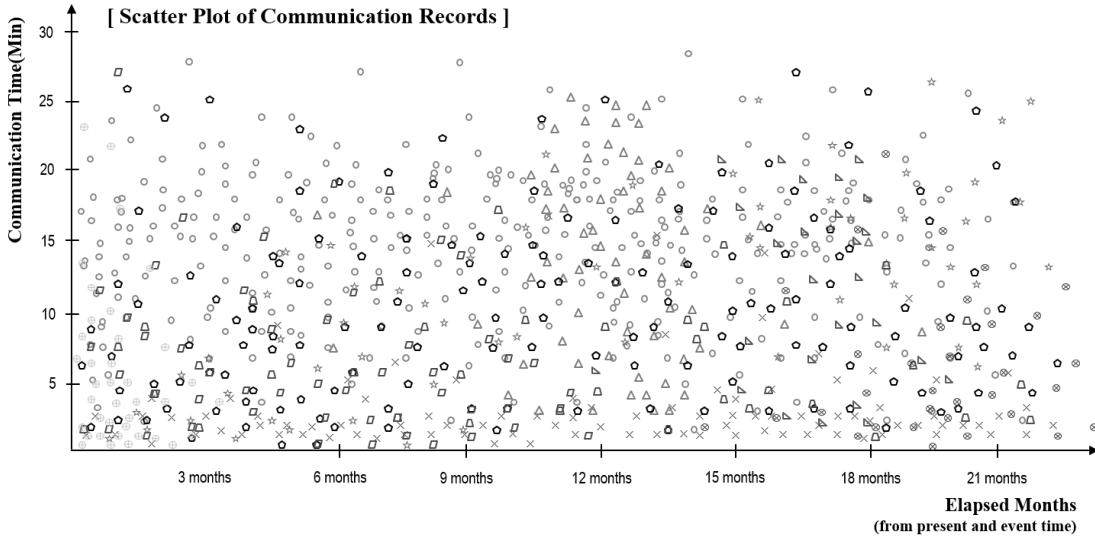


Fig. 2. Communication Records with Neighbors in Mobile Phone

수신자는 이 정보의 재 전달(2차 확산)에 더 적극적인 행위를 할 가능성이 높다.

$$T-Value_{ij} = TIR_{ij} + TIO_{ij} \quad (4)$$

정보 제공자 j 와 정보 수신자 i 간 현실 세계의 신뢰도는 TIR_{ij} (TIR: Trust In Real world)로 정의하고, 온라인 세계에서의 신뢰도는 TIO_{ij} (TIO: Trust In Online world)로 정의한다. TIR_{ij} 과 TIO_{ij} 는 각각 0에서 1사이의 값을 가지며, $T-Value_{ij}$ 역시 Normalization 과정을 거쳐 0과 1사이의 값을 가지도록 하였다.

3.2 Trust In Real World (TIR)

Fig. 2는 특정 사람의 모바일 폰에서 추출한 커뮤니케이션의 이력을 나타낸다. 현실 세계에서의 정보 제공자와의 신뢰도를 나타내는 TIR은 아래의 무질서하게 산개된 커뮤니케이션 이력을 정보 제공자 별로 정량적으로 신뢰도 값을 산출하여 보다 쉽게 신뢰도를 파악할 수 있게 한다.

0에서 1사이의 값을 가지는 TIR_{ij} 는 두 사용자 i, j 간 현실 세계에서의 커뮤니케이션을 나타내는 모바일 폰의 통화, SMS, 그리고 이메일의 사용내역을 바탕으로 신뢰도를 정량적으로 측정할 값이다. Equation (5)와 같이 두 사용자 i, j 간 TIR_{ij} 는 커뮤니케이션의 유형(type)별 커뮤니케이션의 강도($COINT_{ij}$)의 합으로 산출하며, $COINT_{ij}$ 값은 Equation (6)과 같이 5개의 상황 변수를 기반으로 한다.

커뮤니케이션의 유형은 type=email, SMS, phone으로 3가지로 정의하여 현실의 사회적 활동을 반영하고자 하였다.

$$TIR_{ij} = \sum_{type} COINT_{ij(type)}, \quad 0 \leq TIR_{ij} \leq 1 \quad (5)$$

$$COINT_{ij} = ATI_{ij} \times F_{ij} \times D_{ij}^* \times \frac{\sigma_{ij}}{COEFF^{MRTI_{ij}}} \quad (6)$$

ATI_{ij} 는 두 사용자 i, j 가 활동적으로 커뮤니케이션을 한 시간적 간격(Active Time Interval)을, F_{ij} 는 두 사용자 i, j 의 전체 커뮤니케이션 횟수(Frequency)를, D_{ij} 는 커뮤니케이션 유형이 통화인 경우에 한정하여 두 사용자 i, j 간 통화시간(Duration)으로 정의하고, σ_{ij} 는 커뮤니케이션의 산포, 즉 두 사용자 i, j 가 얼마나 꾸준히 커뮤니케이션 해왔는가를 나타낸다. $MRTI_{ij}$ 는 가장 최근에 커뮤니케이션을 한 시점(Most Recent Time Interval)으로 현 시점과의 시간적 차이($|current\ time - event\ time|$)를 나타내는데, 커뮤니케이션의 횟수가 많고 산포가 크다고 할지라도 장시간이 지난 경우라면 그 사용자에 대한 신뢰도는 낮을 가능성이 높기 때문에 해당 상황 정보를 반영하고자 하였다. Fig. 3은 Equation (6)의 $COINT$ 변수를 구성하는 각 인자들의 정의를 그래프로 다시 표현한 것이다. $COEFF$ 는 $COEFF=e$ 를 나타내는 상수이므로 별도로 정의하지 않았다. Table 1에 TIR을 구성하는 모든 변수의 정의를 표시하였다.

3.3 Trust In Online World (TIO)

Fig. 4는 특정 사람(1st Receiver)의 모바일 폰에서 추출한 이메일의 흐름을 나타낸다. 온라인 세계에서의 정보 제공자와의 신뢰도를 나타내는 TIO는 아래의 그래프로 표현된 이메일 흐름을 이메일 송신자 별로 정량적으로 신뢰도 값을 산출하여 보다 쉽게 신뢰도를 파악할 수 있게 한다.

1st Spreader 4명이 1st Receiver로 보낸 메일 중 1st Receiver에 의해 2차적으로 확산된 메일의 분포를 나타낸다 (색깔 별로 구분). 위 Fig. 4에 의하면 1st Receiver는 Megan의

Table 1. Description and Mathematical Definition of each Variable of TIR

Variables	Description	Definition
TIR_{ij}	Degree of social affinity with person j in real world(call, sms, email).	See Equation (5)
$COINT_{ij}$	How closely communicate with person j in case that each communication type is a Call, SMS, E-mail. And it's an abbreviation of Communication Intensity.	$COINT_{ij} = ATI_{ij} \times F_{ij} \times D_{ij}^* \times \frac{\sigma_{ij}}{COEFF^{MRTI_{ij}}}$ $0 \leq COINT_{ij} \leq 1$
ATI_{ij}	A time difference between the recent communication time and the first communication time with a person j . It's an abbreviation of Active Time Interval.	$ATI_{ij} = \frac{ the\ recent\ time_{ij} - the\ first\ time_{ij} }{RTI_{ij}}$ $0 \leq ATI_{ij} \leq 1$ RTI_{ij} is an abbreviation of Reference Time Interval on node j . It means time difference between current time and the first time of communication with j $RTI_{ij} = the\ first\ time_{ij} - the\ current\ time_{ij} $
F_{ij}	How many a user communicate with a person j , it is a total sum of all communication behaviors.	$F_{ij} = \sum Comm_{ij}$ $Comm_{ij}$ means an event of Call or SMS or Email with j
D_{ij}	How long a user have spoken upon the telephone, it is a total sum of talking time with a person j . It can be only applied to a type of 'Call'.	$D_{ij} = \sum_k call_{ij(k)}$ k denotes k -th call among all calls with j
σ_{ij}	Standard deviation of communication with person j , it depends on communication time. Equation on the right side is same as the definition of standard deviation.	$\sigma_{ij} = \sqrt{\frac{F_{ij} \left\{ \sum_{k=1}^{F_{ij}} (CUT - COT_{ij(k)}) \right\}^2}{\sum_{k=1}^{F_{ij}} \left\{ \frac{\sum_{k=1}^{F_{ij}} (CUT - COT_{ij(k)})}{F_{ij}} - (CUT - COT_{ij(k)}) \right\}^2}}$ CUT : Current Time COT : Communication Time
$MRTI_{ij}$	It plays a role to know recency of communication with a person j , that is, how recent they interact with themselves.	$MRTI_{ij} = \frac{ CUT - RET_{ij} }{RTI_{ij}}, \quad 0 \leq MRTI_{ij} \leq 1$ CUT : Current Time RET_{ij} : Recent communication Time with j
$COEFF$	Coefficient related to MRTI, it shows COINT varies exponentially according to MRTI.	$COEFF = e$

이메일을 전달하는데 가장 소극적이었으며(Megan의 이메일을 가장 적게 확산시켰으며), Patricia의 이메일을 전달하는데 가장 적극적이었다.

따라서 온라인 세계에서의 정보 제공자와 수신자 사이의 신뢰도를 나타내는 TIO는 수신자의 정보 전달 패턴으로부터 추론을 할 수 있는데, 정보제공자에 대한 수동성을 나타내는 Passivity를 먼저 측정하여 Equation (7)과 같이 간단히 정의할 수 있다.

$$TIO_{ij} = 1 - Passivity_{ij}, \quad 0 \leq TIO_{ij} \leq 1 \quad (7)$$

정보 제공자 j 와 정보 수신자 i 간 온라인 세계의 신뢰도는 1과 정보 제공자 j 에 대한 정보 수신자 i 의 정보 전달의 소극성 정도를 나타내는 $Passivity_{ij}$ 의 차이로 표현한다. 이 $Passivity_{ij}$ 는 실제 소셜 네트워크인 Facebook에서 수신자의 모든 정보를 Graph API로 취득하여 수신자의 2차 전달 여부를 모두 조사하여 산출한다.

0에서 1사이의 값을 가지는 TIO_{ij} 는 온라인 소셜 네트워크 환경에서 정보 수신자의 정보 전달 성향을 나타낸다. 즉 수신자의 정보 전달 유무는 정보 전달자와의 관계에 영향을 받는데, 이는 정보 수신자의 모든 개별 이웃 노드들에 대한 영향력을 정량적으로 판별할 수 있게 한다. Equation (8)과

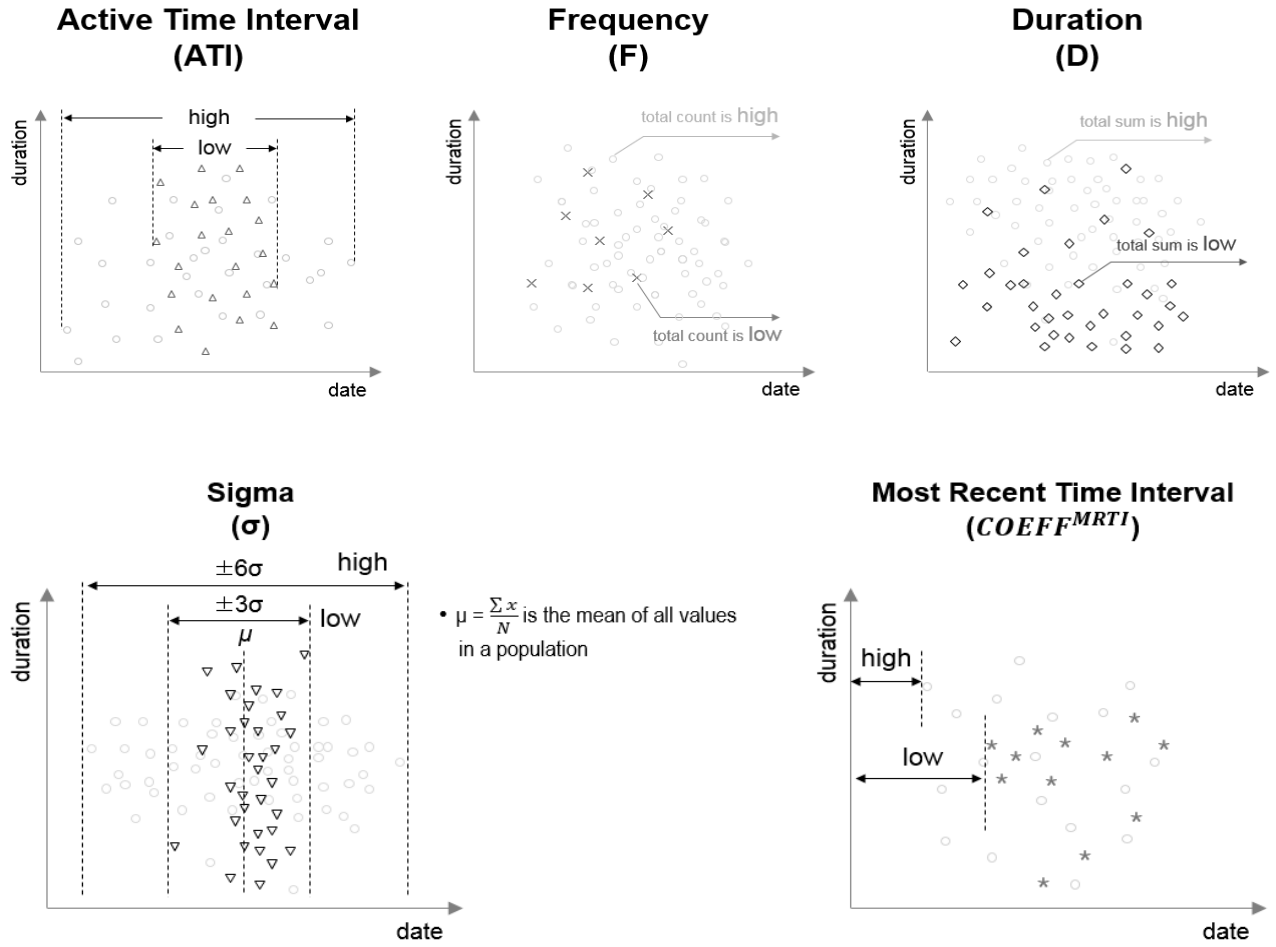


Fig. 3. Graphical Definitions of Factors in COINT Variable

같이 $Passivity_{ij}$ 는 정보 수신자 i 의 정보 제공자 j 에 대한 정보 전달 수동성을 나타내는 값으로서 기본적으로 모든 이웃 노드의 거절률($\sum \varphi_{ik}$)에 대한 노드 j 의 거절률에 대한 비로 표현한다. 거절률은 Equation (9)와 같이 온라인 소셜 네트워크에서 i 가 받은 k 의 정보의 총 개수($\sum_{(i,k) \in E} Info_{ik}$)와 i 가 전달한 k 의 정보의 총 개수($\sum_{(i,k) \in E} Info_{spread_{ik}}$)로 산출한다. TIO_{ij} 를 구성하는 변수는 다음과 같다. 참고로 Equation (8, 9)에서 노드 j 는 각각의 정보 제공자, 노드 k 는 노드 i 의 모든 이웃 노드를 나타내며, Table 2에 TIO_{ij} 를 구성하는 각 변수의 정의를 표시하였다.

$$TIO_{ij} = 1 - Passivity_{ij} = 1 - \left(\frac{\varphi_{ij} \omega_{ij}}{\sum_{k: (i,k) \in E} 1 - \tau_{ik}} \right) \quad (8)$$

$$1 - \tau_{ik} = \varphi_{ik} = 1 - \frac{\sum_{(i,k) \in E} Info_{spread_{ik}}}{\sum_{(i,k) \in E} Info_{ik}} \quad (9)$$

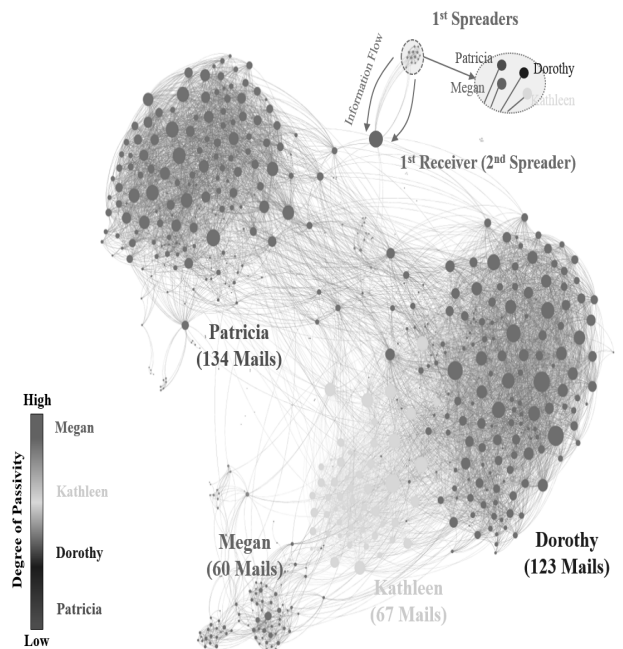


Fig. 4. Definition of TIO with Four Email Spreaders

Table 2. Description and Mathematical Definition of Each Variable of TIR

Variables	Description	Definition
TIO_{ij}	Degree of node i 's Activeness in spreading node j 's information.	see equation (8)
$Passivity_{ij}$	Degree of node i 's Passivity in spreading node j 's information. It depends on relationship with a spreader and it is an abbreviation of Passivity.	see equation (8)
φ_{ij}	Rejection rate. It indicates how many pieces of information were rejected by node i among all pieces of information node j delivered to node i .	$\varphi_{ij} = 1 - \tau_{ij} = 1 - \frac{\sum_{(i,j) \in E} Info_{spread_{ij}}}{\sum_{(i,j) \in E} Info_{ij}}, \quad 0 \leq \varphi_{ij} \leq 1$
ω_{ij}	Weight value of edge $e = (i, j)$, which depends on variables of COINT, ATI, F, MRTI, σ . It reflects how long, how frequently, how recently communicate with node j .	$\omega_{ij} = COINT_{ij} = ATI_{ij} \times F_{ij} \times \frac{\sigma_{ij}}{COEFF^{MRTI_{ij}}} \quad 0 \leq \omega_{ij}$
τ_{ij}	Acceptance Rate. It indicates how many pieces of information were accepted by node i among all pieces of information node j delivered to node i .	$\tau_{ij} = 1 - \varphi_{ij}, \quad 0 \leq \tau_{ij} \leq 1$
$Info_{ij}$ $Info_{spread_{ij}}$	$Info_{ij}$ means information received from node j . $Info_{spread_{ij}}$ means information spread by node i .	No mathematical definition. Just regard them as information itself.
$MRTI_{ij}$	All edges directly connected to node i .	$\forall (i, j, k) \in E$

4. 신뢰도의 평가

4.1 평가의 설계

본 평가의 목적은 앞의 장에서 언급한 TIR, TIO 기법을 바탕으로 온라인 소셜 네트워크 환경에서 정보 제공자와 수신자의 신뢰도를 측정했을 때, 기존의 기법들과 대비해서 더 정확한 신뢰도를 얻는지 확인하는 것이다. 그리고 이 정확함은 각 기법들로 측정된 신뢰도로 중요 정보를 식별하고 수신자에게 제공할 때 수신자의 더 많은 2차 전달 행위를 유도하는 것으로 판정한다. 즉,

- ① 타 기법과 TIR, TIO를 각각 적용한 정보 군(set)을 수신자에게 제공하고,
- ② 각 기법을 적용한 정보 군마다 수신자의 정보 전달률을 측정한다.

여기서 정보 전달률은 수신자가 받은 정보 중 수신자가 관심을 보이고 인접 노드로 전달한 정보의 비율을 의미하며 Equation (10)과 같이 정의한다. 이 정보 전달률(Spread_Rate)이 아래 진행할 평가의 성능 평가 지표가 된다.

$$SR_{user} = \frac{\sum Info_{spread}}{\sum Info_{received}}, \quad 0 \leq SR_{user} \leq 1 \quad (10)$$

정보 수신자(user)의 정보 전달률 SR_{user} 는 이 수신자가 받

은 전체 정보의 개수($\sum Info_{received}$)와 이 중 전달한 정보의 개수($\sum Info_{spread}$)의 비율이며 0과 1사이의 값을 가지도록 한다.

평가의 대상이 될 기법은 2장에서 언급한 TidalTrust와 IP 알고리즘이며, Fig. 5와 같이 각 기법들을 Facebook의 8명의 사용자에게 제공되는 Facebook 포스트(글)에 대하여 6가지 경우로 매달 변경 적용하여 8명의 사용자 각각이 반응한 정보의 개수를 카운트하여 최종 SR_{user} 값을 측정하였다.

4.2 Dataset

챕터 4.1에서 언급한 바와 같이 신뢰도는 정보 제공자에 대한 수신자의 SR 지수로 판정을 한다. SR판정을 위해서는 Equation (10)과 같이 수신자가 받은 정보의 개수($\sum Info_{received}$)와 전달한 정보의 개수($\sum Info_{spread}$)를 측정해야 하는데, 여기서 정보는 Facebook의 정보를 의미하며 이 정보의 구조는 Fig. 6과 같다.

7명의 피 실험자는 Facebook에서 자신에게 전달되는 정보를 Fig. 5와 같이 각 알고리즘을 6가지 경우로 적용하여 각 경우 마다 사용자가 자신의 인접 노드로 전달한 정보의 개수를 측정하고 SR 지수를 측정한다. 피 실험자 별로 적용되는 6가지 알고리즘은 동일한 정보 군으로 평가를 하는 것이 가장 이상적이지만, 실제 사람은 정보에 학습되는 효과가 있기 때문에 각 알고리즘을 다른 정보 군에 대하여 월별로 적용하였다.

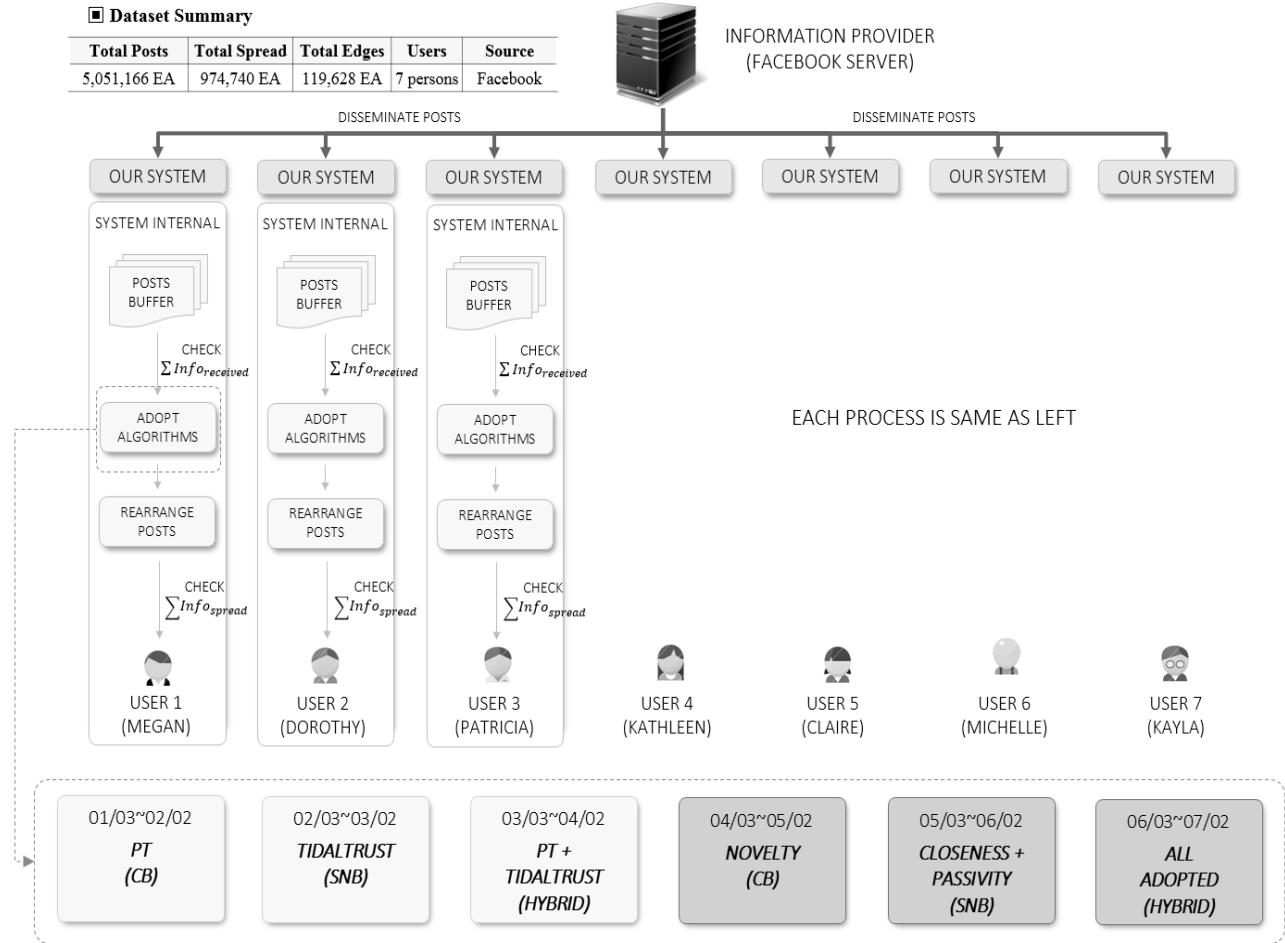


Fig. 5. Entire Progress of Evaluation on 7 Users in Facebook

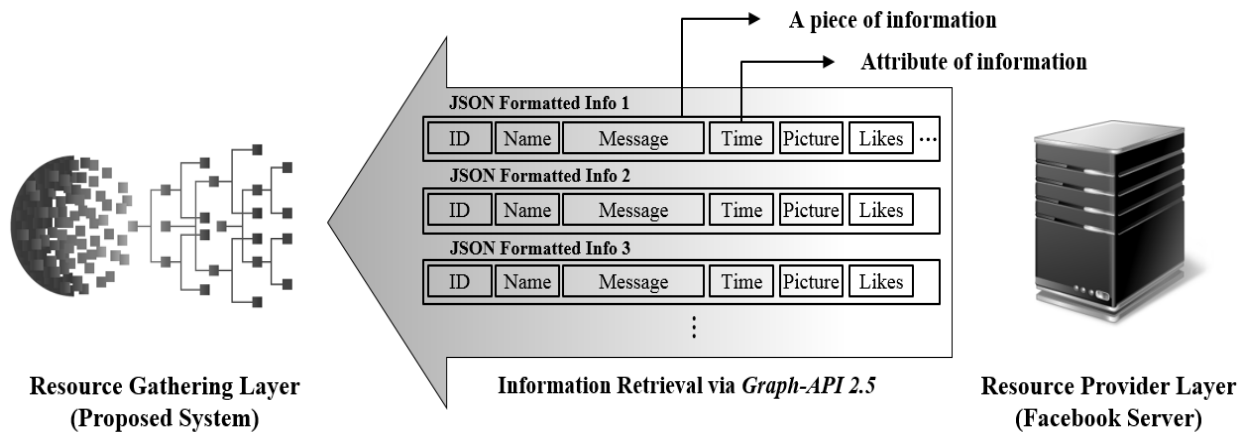


Fig. 6. Simple Expression of Information Structure

참고로 Facebook에서 제공하는 GraphAPI를 사용하여 피실험자 7명에 대해 취득한 총 정보의수는 5,051,166개이고 이들이 전달한 정보의 개수는 974,740개, 그리고 7명에게 직접 연결된 간선의 수(정보 제공자의 수)는 119,628개였다.

4.3 SR 지수의 측정 결과
타 기법들과의 SR지수를 평가하기 전에 TIR, TIO, TIR+TIO 3가지 알고리즘의 SR 성능을 비교한 결과는 Fig. 7과 같다. TIO는 TIR 대비 비교적 우수한 성능의 신뢰도

Table 3. Summary of Results on SR Evaluation

Result 1						Result 2	
<i>IP</i>	<i>TidalTrust</i>	<i>IP+TidalTrust</i>	<i>TIR</i>	<i>TIO</i>	<i>TIR+TIO</i>	<i>Non-Refed (new)</i>	<i>Refed (re-provided)</i>
13.13%	13.74%	15.43%	14.06%	15.98%	16.25%	11.83%	35.79%
3.12%P ↑(vs <i>IP</i>), 2.51%P ↑(vs <i>TidalTrust</i>), 0.82%P ↑(vs <i>IP+TidalTrust</i>)						23.96%	

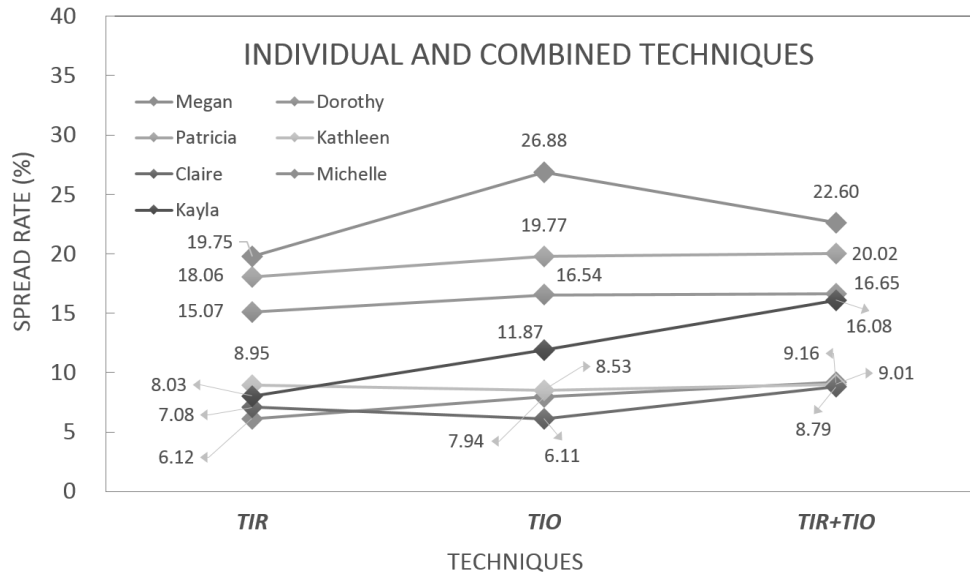


Fig. 7. Performance Comparison between Individual and Combined Techniques

측정 결과를 보이며(7명 중 5명), TIO와 TIR을 복합적으로 적용하였을 때 가장 우수한 성능을 보였다(6명 중 7명). 따라서 3가지 신뢰도 측정 접근법(TIR, TIO, TIR+TIO) 중 Equation 4의 T-Value로 온라인 소셜 네트워크에서 정보 수신자의 정보 제공자의 신뢰도를 측정하는 것이 최적의 접근 방법임을 알 수 있다. 그리고 Table 3은 2장에서 언급한 TidalTrust, IP 알고리즘과 본 연구에서 제시하는 기법과의 SR 지수의 측정 결과를 각각 나타낸 결과이다(Result 1). 4.1, 4.2장에서 언급한 바와 같이, 7명의 피 실험자의 SR 지수를 측정한 결과를 알고리즘 별로 종합한 값으로 제안 기법의 개별 성능 평가와 종합 성능 평가(기법의 복합 적용)를 비교할 수 있다. TIR은 IP+TidalTrust 대비 저조한 SR 값을 얻었으나, TIR과 TIO를 복합하여 적용한 경우 IP, TidalTrust, IP+TidalTrust 대비 각각 3.12%P, 2.51%P, 0.82%P 우수한 SR 지수 값을 산출할 수 있었다. 이는 본 연구에서 제시하는 TIR+TIO 복합 기법으로 온라인 소셜 네트워크에서 정보 수신자와 제공자와의 신뢰도를 측정할 때 IP, TidalTrust, IP+TidalTrust 대비 각각 3.12%P, 2.51%P, 0.82%P 더 정확한 신뢰도 값을 산출할 수 있음을 의미하며, 이 알고리즘을 적용하여 상호 신뢰도를 측정하고 이를 기반으로 정보를 제공하는 경우 더 높은 사용자의 만족도(더 높

은 선별도, 개인화의 적절성)를 얻을 수 있다는 결론을 얻을 수 있다. Fig. 8(a)에 이 결과를 다시 요약하였다.

Table 3의 Result 2는 개인화 정보가 사용자에게 신규로 공급되는 정보 군(Non-Refed)가 아닌 이미 한번 이상 전달되었다가 사용자가 놓친 정보 군(Refed)에 더 많이 포함되어 있음을 SR 지수를 통해 증명하고 있다. 더 정확히 말하면 사용자가 놓친 정보 들을 본 연구에서 제시하는 TIR+TIO 기법을 적용하여 대상 정보들의 중요도를 정량적으로 산출하고 높은 중요도를 가지는 개인화 정보를 우선 적으로 공급할 때 사용자의 정보 반응 행위(전달 행위)를 극대화 할 수 있음을 말하며, 이는 결국 전체 소셜 네트워크에 흐르는 정보의 질적 향상을 기대할 수 있다. Table 3의 Result 2와 Fig. 8(b)에 이를 요약 하였다. 사용자가 놓친 정보들을 수집하여 TIR+TIO 기법으로 상호간 신뢰도를 측정하고 이를 기반으로 정보의 중요도를 산출한 뒤 이를 재공급하는 것이 본 연구의 핵심이다.

5. 결 론

본 연구에서는 온라인 소셜 네트워크에서 정보 제공자와 수신자와의 상호 신뢰도를 측정하는 기법을 두 사람간의 현실 세계에서와 신뢰도(TIR)와 온라인 세계에서와 신뢰도

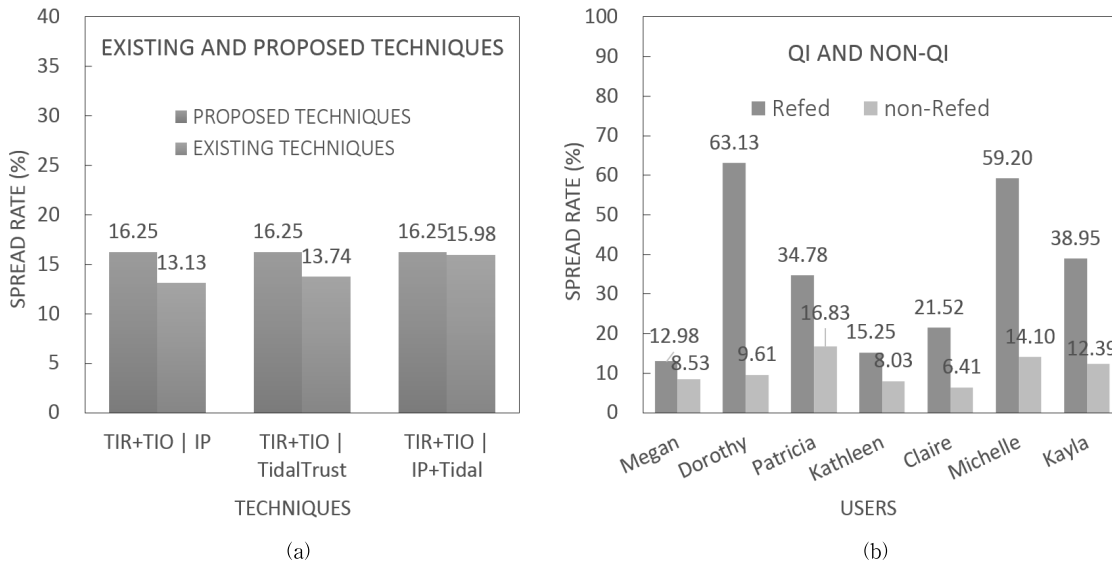


Fig. 8. Final Results of Evaluation

(TIO)를 복합적으로 적용한 기법을 제시하였으며, 이 기법과 SR지수를 기반으로 정보의 가치를 정량적으로 판단하였다. 그 결과 보편적으로 사용되는 IP, TidalTrust기법 대비 각각 3.12%, 2.51% 더 우수한 개인화 성능을 확보할 수 있었는데, 특히 본 기법을 신규 정보가 아닌 사용자에게 이미 제공되었으나 놓친 정보를 대상으로 적용하여 식별된 개인화 정보를 제공했을 때, 사용자의 정보에 대한 반응률이 23.96% 더 높음을 도출하였다.

본 연구는 온라인 소셜 네트워크 환경을 기반으로 진행되었으나, 개인이 소비하는 정보 중 소셜 네트워크 속의 정보는 극히 일부일 것이므로 향후 정보의 범위를 웹 전체로 확대하여 좀 더 일반화된 개인화 기법으로 개선하고자 한다. 표면 웹(Surface Web)에 한정하지 않고 최근 대두되고 있는 심층 웹(Deep Web)까지 포함할 때, 정보 제공자와 수신자의 ‘관계(Trust)’는 보다 더 정확한 값을 가질 것이다.

References

- [1] G. D. Abowd et al., "Towards a better understanding of context and context-awareness," in *Proceeding of International Symposium on Handheld and Ubiquitous Computing*, Springer Berlin Heidelberg, pp.304-307, 1999.
- [2] By the Numbers:200+ Amazing Facebook Statistics [internet], <http://expandeddrabbling-s.com/index.php/by-the-numbers-17-amazing-facebook-stats/8/>.
- [3] D. H. Lee and P. Brusilovsky, "Fighting information overflow with personalized comprehensive information access: A proactive job recommender," in *Proceeding of Autonomic and Autonomous Systems 2007*, pp.21-21, 2007.
- [4] F. E. Walter, S. Battiston, and F. Schweitzer, "A model of a trust-based recommendation system on a social network," *Journal of Autonomous Agents and Multi-Agent Systems*, Vol.16, No.1, pp.57-74, 2008.
- [5] J. Golbeck, "Generating predictive movie recommendations from trust in social networks," in *Proceeding of International Conference on Trust Management*, Springer-Verlag, Berlin Heidelberg, pp.93-104, 2006.
- [6] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proceeding of the 2007 ACM Conference on Recommender Systems*, pp.17-24, 2007.
- [7] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceeding of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 397-406, 2009.
- [8] J. Tang, X. HU, H. Gao, and H. Liu, "Exploiting Local and Global Social Context for Recommendation," in *Proceeding of the International Joint Conference on Artificial Intelligence*, pp.264-269, 2013.
- [9] D. M. Romero, W Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Proceeding of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer Berlin Heidelberg, pp.18-33, 2011.
- [10] M. McGee, "EdgeRank is dead: Facebook's News Feed algorithm now has close to 100K weight factors," *Marketing Land*, 2013.



김 명 훈

e-mail : mhkim@media.knu.ac.kr
2010년 경북대학교 전자전기컴퓨터학부
(학사)
2014년 LG Display 연구원
2017년 경북대학교 컴퓨터학부(석사)

2017년~현 재 경북대학교 컴퓨터학부 박사과정
관심분야: 소셜 네트워크, 정보 확산, 추천 시스템



김 상 욱

e-mail : kimsu@knu.ac.kr
1979년 경북대학교 전자계산기공학(학사)
1981년 서울대학교 컴퓨터과학(석사)
1989년 서울대학교 컴퓨터과학(박사)
1982년~현 재 경북대학교 컴퓨터학부
교수

관심분야: 모바일 미디어, 소셜 미디어, 인간과 컴퓨터의
상호작용, 사물인터넷