

통신 객체들 간의 친밀도 측정 (Proximity Measurement between Communication Objects)

권정은[†] 송지환^{**}
(Jeong Eun Kwon) (Jihwan Song)

김명호^{***}
(Myoung Ho Kim)

요약 미국과 유럽 연합 등을 포함한 여러 나라에서 통신 서비스 제공자들에게 통신 기록들을 일정 기간 보관하도록 의무화하고 있다. 저장된 통신 기록들은 범죄의 수사 및 감시, 기소 등의 목적으로 사용되지만, 그 크기가 매우 크다. 저장된 통신 기록으로부터 범죄자들과 용의자들의 관련성과 같은 정보를 효율적으로 추출하기 위해, 통신 기록의 대부분을 차지하는 불필요한 데이터를 제거해야 한다. 본 논문에서는 통신 기록들을 이용하여 통신 객체들 사이의 친밀도를 측정하는 방법을 제안한다. 주요 감시 대상자들이 주어졌을 때, 측정된 친밀도는 감시 대상자와의 친밀도 크기에 따른 데이터의 선별적인 분석 혹은 낮은 친밀도를 갖는 데이터의 제거 등에 사용될 수 있다. 실험 결과는 제안된 방법에 의해 계산된 친밀도 결과와 사람에 의해 계산된 결과가 서로 상관관계를 갖고 있음을 보여준다.

키워드 : 통신 객체, 친밀도, 통신 기록

Abstract Many countries including the US and EU oblige telecommunication service providers to retain com-

munication logs for a certain amount of time. The retained data are used for the purpose of the investigation, detection, or prosecution of serious crimes, but of huge size. In order to efficiently extract information such as a correlation between criminals and suspects, we must eliminate unnecessary data that occupy a large portion of communication logs. In this paper, we propose how to measure the proximity between communication objects using communication logs. The proximity let the collected data be analyzed efficiently: analyzing the data in the decreasing order of proximities or removing the data with low proximities before analyzing. The experimental results show that there is a correlation between proximities of our proposed measurement and estimation by people.

Key words : communication object, proximity, communication log

1. 서론

통신 네트워크 감청법에 의해, Google, Nortel, Verizon 등과 같은 서비스 제공자는 유, 무선 전화의 통신 기록뿐만 아니라 인터넷 기반의 통신 서비스인 이메일, FTP, HTTP 등의 통신 기록을 일정 기간 동안 저장한다. 이때, 개인 정보 보호를 위해 통신 내용은 일반적으로 저장하지 않는다[1]. 또한, 이들은 경찰, FBI 등과 같은 사법 집행 기관이 저장된 통신 기록을 요청할 경우 이를 제공할 의무가 있다. 저장된 통신 기록의 양은 매우 크기 때문에, 사법 집행 기관은 이를 빠르고 정확하게 분석하기 위해서는 감청을 수행하는 목적과 관련이 높다고 판단되는 데이터부터 분석하는 방법이 필요하다.

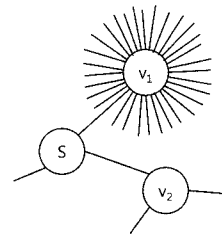


그림 1 스팸을 보내는 통신 객체를 포함한 통신 기록 그래프

본 논문에서는 타겟 통신 객체(예, 감시 대상자)와 친밀한 관계인 통신 객체일수록 감청을 수행하는 목적과 관련이 높은 내용의 통신을 한다고 가정하였다. 이 가정을 기반으로 타겟 통신 객체와 다른 통신 객체 간의 친밀도를 측정하는 방법을 제안한다. 그림 1은 저장된 통신 기록을 이용하여 만든 통신 기록 그래프의 일부이며,

· 본 연구는 한국과학재단 특정기초연구(R01-2006-000-10809-0)의 지원으로 수행되었음.

· 이 논문은 제35회 추계학술대회에서 '통신 객체들 사이의 친밀도 측정'의 제목으로 발표된 논문을 확장한 것임

† 정회원 : SK텔레콤 정보기술원
jekwon@sktelecom.com
** 학생회원 : 한국과학기술원 전산학과
jhsong@dbserver.kaist.ac.kr
*** 종신회원 : 한국과학기술원 전산학과 교수
mhkim@dbserver.kaist.ac.kr
논문접수 : 2008년 12월 19일
심사완료 : 2009년 2월 24일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제15권 제4호(2009.4)

이 그래프에서 노드 v_1 은 불특정 다수와 통신을 하는 스팜을 보내는 통신 객체이다. v_1 과 같은 통신 객체들은 다른 통신 객체와 우연히 통신을 하게 됨에 따라 통신 객체들과의 친밀도가 매우 낮으며, 다른 통신 객체보다 월등히 많은 개수의 통신 객체들과 통신을 한다는 특징을 가진다. 따라서, 타겟 통신 객체와 나머지 통신 객체들 간의 통신 기록 그래프 상에서 인접한 노드 수, 타겟 노드와의 에지 수, 타겟 노드와의 거리 등을 이용하여 통신 객체 간의 친밀한 정도를 판단할 수 있다. 이러한 친밀도는 분석의 우선순위 결정 등과 같은 응용에 이용될 수 있다.

2. 관련 연구

우리의 연구 조사 범위 안에서 통신 객체들 간의 친밀도를 직접 계산하는 과거 연구들은 찾을 수 없었다. 유사한 연구로서 웹 페이지들 간의 상대적인 중요도를 계산하여 웹 페이지의 랭킹에 사용되거나 혹은 어떤 한 분야나 어떤 개념으로 묶여있는 문서 간의 웹 페이지 랭킹에 사용되는 알고리즘들은 찾을 수 있었다.

PageRank 알고리즘[2,3]은 웹 상에 하이퍼링크로 연결된 문서들의 링크 구조를 분석하여 모든 문서들의 상대적인 중요도를 측정하는 알고리즘이다. 웹 상의 모든 문서와 하이퍼링크를 거대한 방향성 그래프로 나타내었고, 문서들을 나타내는 각 노드의 PageRank 값을 계산하여 중요도를 측정하였다.

HITS 알고리즘[4]은 어떤 한 분야나 어떤 개념으로 묶여있는 문서 간의 웹페이지 랭킹에 사용되는 알고리즘이다. HITS 알고리즘은 두 개의 값을 이용하여 웹페이지를 평가하는데, 어떤 웹페이지가 많은 페이지로부터 링크를 받고 있다면 그 페이지는 그 분야에서 중요한 비중을 차지한다는 것을 의미하는 Authority 값과 어떤 웹페이지가 그 그룹에 관련된 여러 가지 페이지에 관한 링크정보를 많이 보유하고 있다면 그 페이지 또한 비중 있는 페이지라는 것을 의미하는 Hub 값을 사용한다.

PageRank의 영향을 받은 질의 독립적인 링크 기반의 랭킹 알고리즘으로는 TrafficRank[5], BlockRank[6], TrustRank[7] 등이 있다. 한편, HITS의 영향을 받아 질의 종속인 링크기반 랭킹 알고리즘으로 SALSA[8], Randomized HITS[9], PHITS[10]들이 제안되었다.

그러나, 이러한 연구들은 웹 페이지들 간의 중요도를 계산한 반면, 통신 기록에서의 통신 객체들 간의 친밀도를 직접 계산하는 시도는 하지 않았다.

3. 통신 객체들 간의 친밀도 측정

3.1 용어 정의

3.1.1 통신 객체

통신 객체란 통신 네트워크 기반의 서비스를 이용하는 유, 무선 전화의 전화번호, 웹 서버의 IP 주소와 그러한 웹 서버에 접근하는 단말기의 IP 주소, 이메일 주소 등을 말한다.

3.1.2 통신 기록

통신이란 통신 객체들간에 정보를 주고 받는 행위로, 통신 기록은 일정 기간 동안 행해진 통신에 대한 정보를 저장해둔 것이다. 통신 기록은 발신지, 수신지, 보낸 시각 정보로 구성된다.

만약 시각 T_n 에 통신 객체 V_i 가 V_j 가 존재하고, V_i 가 V_j 에게 통신 네트워크 기반의 서비스를 이용하여 정보를 보내는 행위를 하였을 경우, 저장된 통신 기록에는 $\langle V_i, V_j, T_n \rangle$ 의 형태로 표현된다. 그림 2(a)는 통신 기록의 예를 나타낸다.

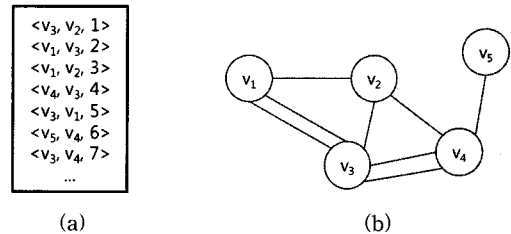


그림 2 통신 기록과 통신 기록 그래프

3.1.3 통신 기록 그래프

통신 기록 그래프란 통신 기록을 그래프 형태로 표현한 것이다. 통신 기록 그래프는 $G=(V,E)$ 인 방향성이 없는 멀티그래프로, 여기서 V 는 통신 객체를 나타내는 노드들의 집합, E 는 통신 객체 간의 통신을 나타내는 에지들의 집합을 의미한다. 그림 2(b)는 그림 2(a)의 통신 기록에서 시각 1부터 시각 7까지의 통신 기록을 나타낸 통신 기록 그래프이다.

통신 기록 그래프 상의 노드들과의 친밀도를 측정하는 기준이 되는 노드, 즉 타겟 통신 객체를 나타내는 노드를 타겟 노드라고 하며, 집합 S 에 저장된다. 집합 S 는 집합 V 의 부분집합이다.

3.1.4 Proximity(친밀도)

Proximity는 통신 기록 그래프 상의 노드와 타겟 노드가 얼마나 밀접한 관계인지를 나타낸다. Proximity가 큰 노드는 타겟 노드와 밀접한 노드일 확률이 큰 노드이고, 반대로 proximity가 작은 노드는 불특정 다수에게 보내는 스팜과 같이 타겟 노드와 관련이 작은 노드이다.

3.2 Proximity의 특성

각 노드의 proximity는 다음의 세 가지 특성을 반영한 값이다.

3.2.1 인접한 노드 수

각 노드의 proximity는 해당 노드와 통신한 내역이 있는 노드 수의 영향을 받는다. 그림 1의 통신 기록 그래프에서 s는 타겟 노드고 v₁과 v₂가 s에 인접해 있다. v₁는 수많은 노드와 통신을 하였고 v₂는 3개의 노드와 통신을 하였으므로, 노드 v₁와 v₂ 중, 다수와 통신을 한 v₁보다는 v₂가 s와 밀접한 관계라고 판단할 수 있다. 따라서, s에 대해 v₂가 v₁보다 큰 proximity를 가지도록 처리한다.

3.2.2 타겟 노드와의 에지 수

각 노드의 proximity는 타겟 노드와의 에지 수의 영향을 받는다. 그림 3에서 s가 감시 대상 노드이고, s는 v₂보다 v₁과의 에지가 더 많이 존재하므로 v₁은 v₂보다 s와 더 밀접한 관계라고 판단할 수 있다. 따라서, s에 대해 v₁은 v₂보다 더 큰 proximity를 가진다고 간주한다. v₃과 v₄의 proximity는 v₁의 proximity의 영향을 받게 되는데, v₃과 v₄의 경우 타겟 노드인 s와 직접 연결되지 못하였으므로 동일한 proximity를 갖도록 처리한다.

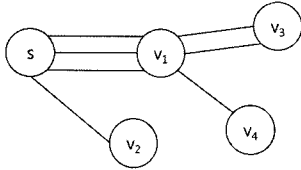


그림 3 타겟 노드와의 에지 수 예

3.2.3 타겟 노드와의 거리

각 노드의 proximity는 타겟 노드와의 거리의 영향을 받는다. 그림 4에서 s가 감시 대상 노드이고, s와 v₂의 거리는 2, s와 v₃의 거리는 3이다. 따라서, s에 대해 v₂는 v₃보다 더 큰 proximity를 갖도록 처리한다.

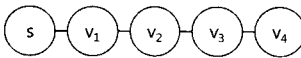


그림 4 타겟 노드와의 거리 예

3.3 Proximity 및 Global Proximity 계산

3.3.1 용어 정의

이절에서는 3.2절에서 언급한 proximity의 개념을 이용하여 global proximity를 정의 및 필요한 용어들을 정리하면 다음과 같다.

- globalProximity(p): 그래프 G상의 모든 타겟 노드들과 노드 p 사이의 proximity들의 합
- proximity(s,p): s는 타겟 노드이며, s에 대한 노드 p의 proximity

- level(s,p): s는 타겟 노드이며, s와 노드 p 사이의 홉 수 중 최소값
- numImmSucc(s,p): s는 타겟 노드이며, (p,q)∈E, q ∉ {타겟 노드들}를 만족하는 노드 q 중, level(s,p) < level(s,q)를 만족하는 노드 q의 수
- w(s): 타겟 노드 s의 가중치
- numEdge(p): 노드 p의 모든 에지 수이다.
- numEdge(p,q): 노드 p와 노드 q 사이의 에지 수이다.

3.3.2 Proximity 계산

Global proximity를 측정하기 위해서는 타겟 노드들의 집합 S의 임의의 노드 s와 V-(s)에 속하는 모든 노드들과의 proximity를 계산하는 것이 필요하다.

p∈S인 경우, 즉 p가 타겟 노드인 경우 proximity가 해당 노드가 타겟 노드와 얼마나 밀접한 관계인지를 나타내므로, 모든 타겟 노드의 proximity는 식 (1)과 같이 무한대 값을 가진다.

$$\forall p \in S, \text{proximity}(s,p) = \infty \quad (1)$$

p∉S이고 p가 s와 인접한 노드이면((s,p)∈E), proximity는 식 (2)와 같다.

$$\text{proximity}(s,p) = \frac{1}{\text{numImmSucc}(s,p)+1} \times \frac{\text{edgeNum}(s,p)}{\text{edgeNum}(s)} \times w(s) \quad (2)$$

edgeNum(s,p)/edgeNum(s)는 3.2.2의 타겟 노드와의 에지수를 반영하기 위한 것이고, sumImmSucc(s,p)+1은 3.2.1의 인접한 노드 수에 의한 것이다. 노드는 그 노드가 각 타겟 노드에 의한 proximity에 의해 global proximity가 결정되는데, 이 때 각 타겟 노드에 가중치를 주기 위한 것이 w(s)이다.

p∉S이고 p가 s와 인접한 노드가 아니면((s,p)∉E), proximity는 식 (3)과 같다.

$$\text{proximity}(s,p) = \sum_{q|(q,p) \in E, \text{level}(s,q) = \text{level}(s,p) - 1} \frac{\text{proximity}(s,q)}{\text{numImmSucc}(s,p)+1} \quad (3)$$

통신 기록 그래프는 방향성이 없는 멀티그래프로, 임의의 한 노드와 타겟 노드 사이의 여러 가지 길 (path)이 존재할 수 있다. 따라서 각 노드와 타겟 노드간의 친밀도는 이러한 여러 길에 의해 영향을 받으며, 위 식에서 시그마를 이용해 여러 proximity를 합하는 것은 이러한 특성을 반영하려는 것이다. sumImmSucc(s,p)+1은 3.2.1의 인접한 노드 수에 의한 것이고, 3.2.2의 타겟 노드와의 에지 수는 타겟 노드와 직접적으로 통신한 것이 아니므로 무시한다. 3.2.3의 타겟 노드와의 거리는 위 식이 자신보다 타겟 노드와의 홉 수가 작은, 즉 더 작은 레벨값을 가지는 상위 노드의 값 중 일부를 취하는 형태이기 때문에 만족된다.

식 (3)에서 레벨에 대한 조건을 주어, 레벨이 더 작은 노드에 의한 값을 취하고 감시 대상 노드에 대한 레벨이 동일한 노드 간에는 길이 존재한다 하더라도 값을 취하지 않는데, 이는 두 노드가 서로 값을 주고받으면서 무한루프에 빠지는 문제를 해결하기 위해서이다.

또한, 식 (4)와 같이 타겟 노드가 아닌 노드들의 임의의 타겟 노드 s 와의 proximity를 모두 더하면 s 의 가중치 값을 가진다.

$$\sum_{p \in V-S} \text{proximity}(s, p) = w(s) \quad (4)$$

이에 따라 그래프 상에 예지가 추가되어 proximity에 변화가 생길 때, 각 노드의 proximity가 증가하는 것이 아니라, 각 노드가 갖게 되는 proximity의 비가 변하게 된다. 이는 통신 기록 그래프 상의 예지의 형태에 따라 특정 노드가 global proximity에 미치는 영향력이 커지는 것을 막기 위한 것이다. 모든 타겟 노드는 global proximity에서 동일한 영향력을 가지게 되며, 타겟 노드에 가중치를 줌으로써 영향력을 조절할 수 있다.

3.3.3 Global Proximity 계산

주어진 통신 기록 그래프에서 타겟 노드들의 집합을 S 라 할 때, 임의의 노드 p 의 global proximity는 S 의 원소인 각 타겟 노드에 대한 proximity의 합으로, 이를 식을 나타내면 식 (5)와 같다.

$$\text{global_proximity}(p) = \sum_{s \in S} \text{proximity}(s, p) \quad (5)$$

타겟 노드가 아닌 노드들의 임의의 타겟 노드 s 와의 proximity를 모두 더하면 s 의 가중치 값을 가진다는 proximity의 특성에 의해, 통신 기록 그래프 상에서 타겟 노드를 제외한 나머지 모든 노드의 global proximity 값의 합은 식 (6)과 같이 집합 S 에 속하는 타겟 노드들의 가중치의 합과 같다.

$$\sum_{p \in V-S} \text{global_proximity}(p) = \sum_{s \in S} w(s) \quad (6)$$

4. 실험 및 평가

실험은 3개의 데이터 집합에 대해 이루어졌으며, 성능 평가는 상관계수를 이용해 제안된 방법의 결과 순위와 사람이 평가한 결과 순위를 비교하였다.

4.1 실험 환경

개인 정보 보호 상의 문제로 실제 통신 기록을 가지고 실험을 하는 것은 사실상 불가능하다. 따라서, 임의로 생성한 데이터를 이용해 제안된 방법의 성능을 평가하였다. 50개의 통신 객체와 120개 이상의 통신 내역을 가지는 3개의 저장된 통신 기록들을 생성하였다. 첫 번째 통신 기록은 통신 객체들이 크게 4그룹을 이루는 경

향을 가지도록 임의로 생성하였고, 두 번째 통신 기록은 크게 2그룹을 이루는 경향을 가지도록, 마지막 세 번째 통신 기록은 어떠한 경향성도 가지지 않도록 임의로 생성하였다. 타겟 노드는 통신 기록 그래프에서 연결된 노드의 수가 5개 이상인 노드들 중 임의로 3개를 선택하였다.

4.2 평가 기준

실험에 사용된 통신 기록이 실제 통신 기록이 아니라 임의로 생성된 통신 기록이기 때문에 어떠한 노드가 타겟 노드와 밀접한 관계인지를 절대적으로 평가할 수 있는 기준이 없다. 그렇기 때문에 사람의 판단을 기준으로 삼아 통계학적 방법을 사용하였다. 실험자들에게 주어진 통신 기록 그래프 상의 노드들의 타겟 노드와의 친밀도 순위를 조사하고, 해당 순위와 논문에서 제안한 방법을 사용하여 구한 순위를 상관계수(coefficient of correlation)를 사용하여 평가하였다.

상관계수는 두 변량 X, Y 사이의 상관관계의 정도를 나타내는 수치로서, 본 실험에서는 실험자가 평가한 순위와 논문에서 제안한 방법이 평가한 순위 사이의 상관관계를 나타내는 상관계수를 구한다. 측정값 (x, y) 에 대해서 n 개의 측정값인 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 이 주어졌을 때 식 (7)과 같이 r 을 x, y 사이의 상관계수라고 한다.

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad (7)$$

\bar{x}, \bar{y} 는 각각 x, y 의 평균값, σ_x, σ_y 는 x, y 의 표준편차이다. 상관계수 r 은 항상 부등식 $-1 \leq r \leq 1$ 을 만족시키며, 양의 상관관계가 있을 때는 $r > 0$, 음의 상관관계가 있을 때는 $r < 0$ 이다. 무상관일 때는 0의 값을 갖는다.

4.3 실험 결과

실험은 15명의 사람들에게 감시 대상 노드들이 표시된 통신 기록 그래프를 제시하고, 각 감시 대상 노드와 통신한 노드들 중 그래프 상의 모든 감시 대상을 고려하였을 때 감시 대상 노드와 가장 의심스러운 대화를 나눈만한 노드의 순위를 매겨달라고 하였다. 이는 친밀도의 개념을 쉽게 이해시키기 위한 것으로, 친밀도가 큰 노드는 사건에 연루된 사람들과 관련된 노드들과 밀접한 사이이고, 따라서 해당 사건에 관련된 대화를 나눌 확률이 높다고 볼 수 있다. 즉 가장 의심스러운 대화를 나누는 노드라고 할 수 있다.

표 1은 통신 기록 #1, 통신 기록 #2, 그리고 통신 기록 #3에 대해, global proximity에 의한 노드들의 순위(GP)와 실험자에 의한 노드들의 순위(H) 중 상위 10개 순위를 나타낸다.

표 1 실험 결과

순위	통신 기록 #1		통신 기록 #2		통신 기록 #3	
	GP	H	GP	H	GP	H
1	20	20	24	4	19	19
2	43	43	4	24	35	35
3	14	30	29	29	15	43
4	30	14	42	20	24	4
5	28	23	41	42	14	15
6	23	9	20	12	5	5
7	9	28	12	26	43	24
8	21	21	34	41	34	14
9	24	44	44	34	23	23
10	27	17	23	23	11	18

그림 5는 통신 기록 #1, #2, 그리고 #3에 대한 실험 결과의 상관관계와 표준편차를 보여준다. 통신 기록 #1과 #2의 실험 결과에서 상관관계는 0.8 이상으로, global proximity에 의한 순위 결과와 실험자의 순위 결과가 서로 상관관계를 가진다고 할 수 있다. 하지만 통신 기록 #3의 실험 결과에서는 상대적으로 낮은 상관관계를 보였다.

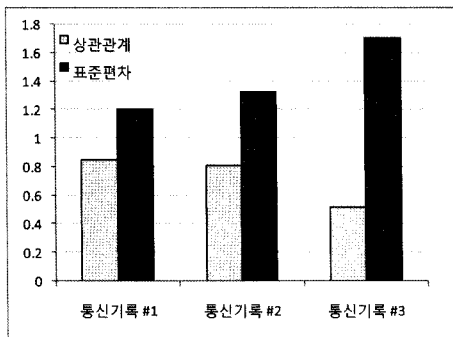


그림 5 실험 결과들의 상관관계 및 표준편차

표준편차는 실험자들이 매진 순위가 얼마나 유사한가를 측정하기 위한 것으로 통신 기록 #1의 실험결과에서 통신 기록 #3의 실험결과로 갈수록 표준편차가 커졌음을 알 수 있다. 이는 통신 기록 그래프가 통신 기록 #1에서 #3으로 갈수록 복잡하기 때문에, 표준편차가 커지고 상관관계가 작아졌다고 볼 수 있다. 즉 복잡한 데이터를 가지고 친밀도를 판단할수록 사람들의 판단이 힘들어지고 사람들간의 결과의 차이도 크다고 볼 수 있다. 그러나, 비교적 간단한 그래프에서는 본 논문에서 제안한 global proximity에 의한 결과와 사람들이 평가한 결과가 유사하게 나왔다.

5. 결론

본 논문에서는 통신 기록을 이용하여 통신 기록 그래프 상에서 타겟 통신 객체와 나머지 통신 객체들 간의

친밀도를 측정하는 방법을 제안하였다. 이 방법은 통신 내용 데이터를 고려하지 않고, 과거의 통신 형태로부터 통신 객체 간의 밀접한 정도를 측정한다. 친밀도 측정을 위해 타겟 통신 객체와 나머지 통신 객체들 간의 통신 기록 그래프 상에서 가지는 여러 특징을 찾아 이를 기준으로 친밀도를 측정하기 위한 식을 세웠다. 이를 통해 전자 감시 시스템에서 사법 집행 기관이 서비스 제공자로부터 실시간으로 받은 감청 데이터를 분석하여 수사에 이용할 때, 저장된 통신 기록으로부터 구한 친밀도 순위에 따라 분석을 하면 짧은 시간 내에 한정된 인력으로 효율적인 분석을 할 수 있을 것이다.

참고 문헌

- [1] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006.
- [2] Brin, S. and Page, L., "The anatomy of a large-scale hypertextual Web search engine," Computer Networks and ISDN Systems, Vol.30, No.1-7, pp. 107-117, 1998.
- [3] Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank citation ranking: Bringing order to the web," Technical report, Stanford Digital Library Technologies Project, 1998.
- [4] Kleinberg, J.M., "Authoritative sources in a hyper-linked environment," Journal of the ACM, Vol.46, No.5, pp. 604-632, 1999.
- [5] Tomlin, J.A., "A new paradigm for ranking pages on the World Wide Web," In Proc. of the 12th International World Wide Web Conference, pp 350-355, 2003.
- [6] Kamvar, S.D., Haveliwala, T.H., Manning, C.D. and Golub, G.H., "Extrapolation methods for accelerating PageRank computations," In Proc. of the 12th International World Wide Web Conference, pp. 261-270, 2003.
- [7] Gyöngyi, Z., Garcia-Molina, H. and Pedersen, J., "Combating web spam with TrustRank," In Proc. of the 30th International Conference on Very Large Databases, pp. 576-587, 2004.
- [8] Lempel, R. and Moran, S., "The stochastic approach for link-structure analysis (SALSA) and the TKC effect," Computer Networks and ISDN Systems, Vol.33, No.1-6, pp. 387-401, 2000.
- [9] Ng, A.Y., Zheng, A.X. and Jordan, M.I., "Stable algorithms for link analysis," In Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 258-266, 2001.
- [10] Cohn, D. and Chang, H., "Learning to probabilistically identify authoritative documents," In Proc. of the 17th International Conference on Machine Learning, pp. 167-174, 2000.