

## 통신이력을 사용한 교우 친밀도 분석\*

문양세 최훈영 김혜숙<sup>o</sup> 김진호  
강원대학교 컴퓨터과학과

ysmoon@kangwon.ac.kr, papcan@hanmail.net, {hskim, jhkim}@kangwon.ac.kr

### Analysis of Degrees of Familiarity among Students Using Communication History Records

Yang-Sae Moon, Hun-Young Choi, Hea-Suk Kim<sup>o</sup>, and Jinho Kim  
Dept. of Computer Science, Kangwon National University

#### 요 약

최근 청소년의 교우관계에 있어서 집단 따돌림이 사회적으로 큰 문제가 되고 있다. 본 논문에서는 학생들의 교우관계를 보다 객관적으로 분석하는 방법을 제안한다. 우선 분석을 위한 객관적인 데이터로 교우관계 정보를 명시적으로 내포하고 있는 통신이력 데이터를 사용한다. 다음으로 교우간 친밀도(*degree of familiarity*) 개념을 정형적으로 정의하고, 여러 통신도구에서 발생한 통신이력 데이터를 기반으로 교우간 친밀도를 수학적으로 계산하는 방법을 제안한다. 다음으로, 계산한 교우간 친밀도를 사용하여 집단 따돌림의 가능성이 높은 학생을 파악한다. 이는 집단 따돌림 가능성이 높은 학생은 다른 학생으로부터의 통신 시도가 적을 것이라는 직관에 기반한다. 마지막으로, 제안한 방법의 실용성을 입증하기 위하여, 실제 구현 및 분석 실험을 수행하고, 그 의미를 해석한다. 이 같은 결과를 볼 때, 본 연구는 학생들의 교우관계를 보다 객관적으로 파악할 수 있는 매우 우수한 결과라 사료된다.

#### 1. 서론

최근 청소년의 교우관계에 있어서 특정 학생을 따돌리거나 괴롭히는 등의 집단 따돌림 현상이 급증하고 있다[1]. 본 논문에서는 집단 따돌림 분석을 위해 통신이력 데이터를 사용한다. 통신도구 중에서, 비교적 최근에 개발된 전화, 메신저, 이메일 등은 통신 과정에서 통신이력 데이터가 저장된다. 그런데, 이들 통신이력 데이터는 인간관계 정보를 명시적으로 내포하고 있다고 볼 수 있다. 또한, 통신이력 데이터는 통신을 수행하는 과정에서 자동적으로 생성되므로, 교우관계를 파악할 수 있는 비교적 객관적인 데이터로 활용할 수 있다. 따라서, 본 논문에서는 이러한 통신이력 데이터를 사용하여 객관적 교우관계를 파악하고, 이를 기반으로 집단따돌림을 분석하는 방법을 제시한다.

본 논문에서는 통신이력 데이터를 기반으로 집단 따돌림을 조기에 파악할 수 있는 체계적인 접근법을 제안한다. 먼저, 교우간 친밀도(*degree of familiarity*) 개념을 제시하고, 여러 통신도구에서 생성된 통신이력 데이터를 기반으로 교우간 친밀도를 계산하는 수학적 기법을 제안한다. 제안하는 기법에서는 학생  $x$ 가 학생  $y$ 로 통신을 많이 할수록, 학생  $x$ 는  $y$ 를 친밀하게 생각한다는 가정을 사용한다. 즉, 학생  $x$ 에서  $y$ 로의 교우간 친밀도를 학생  $x$ 가  $y$ 로 얼마나 자주 통신하느냐로 정형적으로 정의하고, 이를 수학적으로 계산한다. 다음으로, 계산한 교우간 친밀도를 사용하여 집단 따돌림의 가능성이 높은 학생을 파악한다. 이는 집단 따돌림 가능성이 높은 학생은 다른 학생으로부터의 통신 시도가 적을 것이라는 직관에 기반한다.

통신이력 데이터 기반의 교우관계 분석이 유용함을 입증하기 위하여, 본 논문에서는 제안한 기법을 구현하고 실제 분석 실험을 수행한다. 분석에 사용한 데이터는 설문 조사를 통해 수집한 통신이력 데이터로서, 우선 이들 데이터를 분석하여 교우간 친밀도를 구한다. 그런 다음, 이 친밀도 데이터를 바탕으로 집단 따돌림 가능성이 높은 학생들을 추출하고, 그 의미를 분

석한다. 본 논문에서 제안한 방법을 교우관계 분석에 활용할 경우, 교사 및 학부모가 문제 학생을 조기에 파악하여, 학생들을 매우 객관적으로 지도할 수 있다고 사료된다[2].

#### 2. 관련 연구

본 장에서는 사회학에서 주로 다루은 기존의 인간 및 교우관계 분석 방법을 설명한다. 사회학이란 기본적으로 인간과 그들이 구성하는 집단, 즉 사회와의 관계를 과학적으로 규명하려는 학문이다[3]. 사회학의 분석 방법에서 인간관계나 교우관계 분석에 많이 사용되는 방법은 다음과 같다[4,5].

- 상호작용분석: 집단 내 구성원들 간에 발생하는 상호작용 패턴을 분석하여 구성원 사이의 관계를 찾아내는 방법으로, Bales의 표준적 관찰분류목적체계, Comeli의 학급관찰방법 등이 대표적인 방법이다.
- 소시오메트리(sociometry): 집단 내의 개인 상호간의 매력, 배척, 무관심의 정도를 측정하여 집단에서 개인의 위치와 관계를 찾아내는 방법이다.

그러나, 사회학에서 사용하는 교우관계 분석 방법은 기본적으로 집단 구성원의 제보나 관찰자의 관찰을 통해 결론을 해석하기 때문에 여러 가지 오류가 발생할 가능성이 높다[4-7]. 따라서, 일상적인 생활에서 발생하는 정보 속에서 분석 데이터를 찾아내고, 이를 교우관계에 활용하는 시도가 필요하다.

#### 3. 통신이력 데이터 기반의 교우 친밀도 계산

##### 3.1 통신이력 데이터

통신이력 데이터는 통신을 시도하는 과정에서 저장되는 발신자, 착신자, 통신 시각 등의 이력 정보를 의미한다. 그리고 이들 통신이력 데이터는 기본적으로 인간관계 정보를 내포한다고 볼 수 있다. 즉, 잦은 통신을 시도하는 두 사람은 일반적으로 친밀한 인간관계를 가진다고 생각할 수 있고, 이를 학생들에게 적용하면 교우관계를 분석할 수 있다.

일반적으로 통신회사는 통신 과정에서 발생하는 통신이력 데이터를 보관한다. 실제로 통신이력 데이터는 통신사실확인자료

\* 본 연구는 첨단정보기술연구소(AITrc)를 통하여 한국과학재단(KOSEF)의 지원을 받았다.

신청을 통해 얻을 수 있으며, 이력 데이터에는 통신일시, 발신한 발신자 및 착신자 정보를 사용하여 교우관계를 분석한다. 즉, 통신 내용 자체를 이용하는 것이 아니라, 통신이력 데이터의 최소 정보만을 사용하여 분석을 수행한다. 통신이력 기반의 교우관계 분석을 위하여, 본 논문에서는 표 1의 표기법을 사용한다.

표 1. 주요 표기법.

표기법	정의/ 의미
$x, y$	분석하고자 하는 학생
$S$	분석하고자 하는 학생들의 집합 ( $x \in S, y \in S$ )
$m$	통신도구의 개수
$f_i(x, y)$	$i$ -번째 통신도구 $f_i$ 를 사용하여 학생 $x$ 가 $y$ 로 통신한 횟수

3.2 교우간 친밀도

학생  $x$ 에서 학생  $y$ 로의 교우간 친밀도는 "학생  $x$ 가  $y$ 로 통신 시도를 많이 할수록, 학생  $x$ 는  $y$ 로 친밀하다"는 직관을 바탕으로 다음과 같이 정의한다.

정의 1: 학생  $x$ 에서 학생  $y$ 로의 교우간 친밀도  $dof(x, y)$ 는 다음 식 (1)과 같이 정의한다.

$$dof(x, y) = \frac{\text{학생 } x \text{가 } y \text{로 시도한 통신 횟수}}{\text{학생 } x \text{가 시도한 전체 통신 횟수}} = \frac{\sum_{i=1}^m f_i(x, y)}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^m f_i(x, y) \right)} \quad (1)$$

정의 1에 따르면, 학생  $x$ 가  $y$ 로 자주 통신을 시도할수록 친밀도가 높아진다. 일반적으로 친할수록 많은 통신을 시도할 것이므로, 이는 매우 합리적이고 직관적인 척도라 할 수 있다.

그런데, 각각의 학생은 자신이 선호하는 통신도구가 다를 수 있으므로, 이를 고려한 교우간 친밀도 계산이 필요하다. 이를 위해 각 학생 별로 자신이 사용하는 전체 통신도구 중에서 특정 통신도구의 사용 비율을 가중치로 계산하고, 이를 교우간 친밀도 계산에 적용한다. 결국, 학생  $x$ 의 통신도구  $k$ 에 대한 가중치란 학생  $x$ 가 전체 통신도구 중에서  $k$ -번째 통신도구를 사용하는 비율로 다음과 같이 정형적으로 정의한다.

정의 2: 학생  $x$ 의  $k$ -번째 통신도구에 대한 가중치  $weight(x, k)$ 는 다음 식 (2)와 같이 정의한다.

$$weight(x, k) = \text{통신도구 개수} \cdot \frac{\text{학생 } x \text{가 } k\text{-번째 통신도구 사용한 횟수}}{\text{학생 } x \text{가 시도한 전체 통신 횟수}} = m \cdot \frac{\sum_{\text{for all } y \in S} f_k(x, y)}{\sum_{i=1}^m \left( \sum_{\text{for all } y \in S} f_i(x, y) \right)} \quad (2)$$

앞서 정의 1에서 제시된 교우간 친밀도는 정의 2의 가중치를 적용하여 다시 정의할 수 있다. 즉, 학생  $x$ 가  $y$ 로 통신한 횟수인  $\sum_{i=1}^m f_i(x, y)$ 을 정의 1의 가중치를 적용하여 구하면  $\sum_{i=1}^m (weight(x, i) \cdot f_i(x, y))$ 가 된다. 이와 같이, 기본적인 통신 횟수에 개인의 통신도구 선호도인 가중치를 적용하여 통신 횟수를 구한 값을 가중치 기반 통신 횟수라 부르고 이를 사용하여 가중치 기반 친밀도를 정의하면 다음과 같다.

정의 3: 학생  $x$ 에서 학생  $y$ 로의 가중치 기반 친밀도  $wdof(x, y)$ 는 다음 식 (3)과 같이 정의한다.

$$wdof(x, y) = \frac{\text{학생 } x \text{가 } y \text{로 시도한 가중치 기반 통신 횟수}}{\text{학생 } x \text{가 시도한 전체 통신 횟수}} = \frac{\sum_{i=1}^m (weight(x, i) \cdot f_i(x, y))}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^m f_i(x, y) \right)} \quad (3)$$

정의 1의 친밀도와 정의 3의 가중치 기반 친밀도의 차이는 개인의 통신도구 선호도를 고려하는지 그렇지 않는지의 차이이다. 본 논문에서는 개인의 선호도를 반영하기 위하여 정의 3의 가중치 기반 친밀도를 사용한다.

그림 1은 통신이력 데이터를 바탕으로 교우간 친밀도를 추출하는 알고리즘 *ComputeWdof()*를 나타낸다. 알고리즘은 정의 2와 정의 3에 기반하여 통신이력 데이터를 바탕으로 각각의 학생  $x$ 에서 학생  $y$ 로의 친밀도를 계산하는 일련의 과정으로 구성되어 있다. 스텝 (1), (2)를 보면, 정의 2의 식 (2)를 사용하여 각 학생  $x$ 의 통신도구 별 가중치를 구한다. 다음으로, 스텝 (3), (4)에서는 정의 3의 식 (3)을 사용하여 학생  $x$ 에서 다른 학생  $y$ 로의 교우간 친밀도를 구한다.

procedure *ComputeWdof*(A set of  $f_i(x, y)$ 's)

begin

(1) for each  $x$  in  $S$

(2) for each  $k$  in  $[1, m]$   $weight(x, k) = m \cdot \frac{\sum_{\text{for all } y \in S} f_k(x, y)}{\sum_{i=1}^m \left( \sum_{\text{for all } y \in S} f_i(x, y) \right)}$ ;

(3) for each  $x$  in  $S$

(4) for each  $y$  in  $S (y \neq x)$   $wdof(x, y) = \frac{\sum_{i=1}^m (weight(x, i) \cdot f_i(x, y))}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^m f_i(x, y) \right)}$ ;

end

그림 1. 교우간 친밀도 계산 알고리즘.

4. 집단 따돌림 분석

집단 따돌림 학생은 다른 학생이 해당 학생을 친밀하게 생각하지 않는다고 생각할 수 있다. 이러한 개념을 바탕으로 본 논문에서는 집단 따돌림 가능성이 높은 학생을 통신이력 데이터 기반의 교우간 친밀도를 사용하여 다음과 같이 가정한다.

가정 1: 통신이력 데이터를 기반으로 교우간 친밀도를 계산하였을 때, 특정 학생으로의 교우간 친밀도의 합이 매우 낮은 학생은 집단 따돌림의 가능성이 높다. □

가정 1에 따르면 집단 따돌림 가능성이 높은 학생은 다른 학생으로부터 해당 학생으로의 통신 시도가 매우 적은 학생으로 해석할 수 있다. 즉, 학생  $y$ 로의 친밀도의 합이 다음 식 (4)와 같이 매우 작은 경우에, 학생  $y$ 는 집단 따돌림 가능성이 높다고 판단한다.

$$\sum_{\text{for all } x \in S} wdof(x, y) \approx 0 \quad (4)$$

그런데, 식 (4)를 사용하여 집단 따돌림의 가능성이 높은 학생을 판별하기 위해서는 "0에 가깝다"는 이야기를 정량적으로 표현할 필요가 있다. 이를 위하여, 본 논문에서는 다음과 같이 임계치  $\theta$ 를 도입하여 집단 따돌림 학생을 정의한다.

정의 4: 학생  $y$ 로의 교우간 친밀도의 합이 추어진 임계치  $\theta$ 이하이면, 학생  $y$ 는 집단 따돌림이라 정의한다. 즉, 다음 식 (5)를 만족하는 학생  $y$ 를 집단 따돌림이라 정의한다.

$$\sum_{\text{for all } x \in S} wdof(x, y) \leq \theta \quad (5)$$

정의 4에 따라, 임계치가 교우관계 분석을 원하는 교사 등에 의해 주어졌을 때, 교우간 친밀도의 합이 주어진 이하인 학생을 집단 따돌림 가능성이 높다고 판단한다.

그림 2는 정의 4에 기반하여 집단 따돌림 가능성이 높은 학생을 판별하는 알고리즘 *FindBullying()*을 나타낸다. 알고리즘을 보면, 각 학생  $y$ 에 대한 교우간 친밀도 합을 사용하여 집단 따

둘림 여부를 판단함을 알 수 있다. 스텝 (2)를 보면, 학생 y로의 친밀도의 합이 주어진 임계치 이하이면, 학생 y는 집단 따돌림의 가능성이 높다고 판단한다. 반면에, 스텝 (3)에서와 같이 친밀도의 합이 보다 크면, 집단 따돌림의 가능성이 낮다고 판단한다.

```

procedure FindBullying (A set of wdo(x, y)'s, Threshold  $\theta$ )
begin
  (1) for each y in S
  (2)   if  $\sum_{\text{for all } x \in S} wdo(x, y) \leq \theta$  then bullying(y) = False;
  (3)   else bullying(y) = True;
end
    
```

그림 2. 집단 따돌림 학생 판단 알고리즘.

5. 구현 및 실험

5.1 구현 및 실험 환경

실험을 위하여 교우간 친밀도 알고리즘 ComputeWdof()과 집단 따돌림 분석 알고리즘 FindBullying()을 구현하였다. 모든 실험은 Intel Pentium 4 3.0GHz CPU, 392MB 메모리, 30GB 하드디스크를 가진 Windows 2000 서버에서 수행하였으며, 알고리즘의 구현은 Borland Delphi 언어를 사용하였다.

통신이력 데이터는 설문 조사를 통하여 수집하였다. 실제로 통신이력 데이터는 통신회사에서 얻을 수 있으나[9], 많은 학생들을 대상으로 개개인의 동의를 얻기에는 기간 및 여건상 어려운 상황이었어서 본 논문에서는 설문조사를 통해 통신이력 데이터를 수집하였다. 설문조사는 경기도 양평에 위치한 모 고등학교 2학년의 6개 학급 모든 학생을 대상으로 2005년 6월 16일부터 25일 사이에 실시하였다. 설문조사 내용은 각 학생이 다른 학생에게 핸드폰, 문자 메시지, 메신저, 이메일을 사용하여 통신한 발신자, 착신자, 통신 횟수이다.

5.2 분석 결과

집단 따돌림 분석 실험에서는 임계치  $\theta$ 를 변화시키면서 친밀도 합이 임계치 이하인 학생들을 조사하였다. 분석 결과 각 학생의 교우간 친밀도 합이 평균은 0.90으로 나타났다. 친밀도 합의 평균이 1보다 작게 나온 이유는 다른 학생에게 단 한번도 연락을 하지 않은 학생이 있기 때문이다. 평균을 중심으로 친밀도 합을 간단히 해석하면 다음과 같다. 먼저, 교우간 친밀도의 합이 평균보다 낮은 학생은 자신을 좋아하는 친구들이 상대적으로 적음을 의미한다. 다음으로, 친밀도의 합이 평균보다 높은 학생은 자신을 좋아하는 친구들이 상대적으로 많음을 의미한다.

그림 3은 임계치  $\theta$ 의 변화에 따른 학생 수의 변화를 나타낸다. 그림을 보면, 임계치가 0.00에서 1.70사이에서 거의 모든 학생(전체의 93.0%)이 고르게 분포하는 것을 알 수 있다. 직관적으로는 대다수의 학생들이 평균인 0.90 주위에 모여 있을 것으로 예상했으나, 실제로는 평균을 중심으로 고르게 분포하는 것으로 나타났다. 이는 학생들의 교우관계가 개인적으로 상당한 차이가 있음을 나타낸다. 다음으로, 그림에서 교우간 친밀도 합이 임계치 0.10, 0.20와 같이 작은 값을 가지는 학생이 집단 따돌림의 가능성이 높은 학생이다. 이와 같이 파악된 집단 따돌림 학생에 대해서는 면담 등을 통해 특별한 지도와 관심이 필요하다고 사료된다.

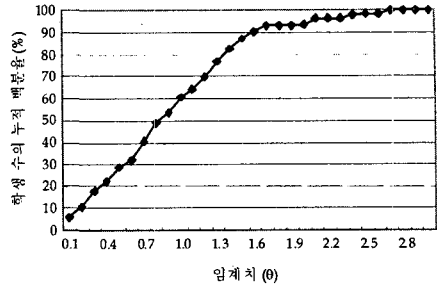


그림 3. 임계치 변화에 따른 학생 비율.

6. 결론

본 연구에서는 교우관계를 파악하기 위한 객관적인 데이터로 통신이력 데이터를 사용하며, 이를 바탕으로 집단 따돌림을 파악하는 체계적인 방법을 제안하였다. 본 연구의 공헌을 요약하면 다음과 같다. 먼저, 통신이력 데이터 기반의 교우간 친밀도를 정형적으로 정의하고, 이를 계산하는 체계적인 방법을 제안하였다. 다음으로, 교우간 친밀도를 기반으로 집단 따돌림 학생을 판별하는 방법을 제안하였다. 이를 위해 집단 따돌림 학생은 다른 학생들로부터의 교우간 친밀도가 낮은 학생으로 가정하였다. 즉, 다른 학생으로부터의 통신 시도가 낮을 경우 그 학생을 집단 따돌림 가능성이 높다고 가정하고, 이를 판별하는 방법과 알고리즘을 제안하였다. 마지막으로, 구현 및 실험을 통해 통신이력 데이터 기반의 교우관계 분석 방법의 실용성을 입증하였다. 이 같은 결과를 볼 때, 본 논문의 연구 결과는 실제 학교 현장에서 학생들의 교우관계를 보다 객관적이고 체계적으로 파악할 수 있는 매우 우수한 연구 결과라 사료된다.

참고 문헌

- [1] Lee, C., Preventing Bullying in Schools, Sage, 2004.
- [2] 김경동, 사회학의 이론과 방법론, 한국학술정보, 2003.
- [3] Chen, M. S., Han, J., and Yu, P. S., "Data Mining: An Overview from a Database Perspective," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883, Dec. 1996.
- [4] 강상철, 최건수, "상징적 상호작용론," *교육발전총론*, 제9권, 제1호, pp. 23-39, 1988.
- [5] 최순남, 인간행동과 사회환경(제3판), 법문사, 2005.
- [6] Snell, H. E., et al., Social Relationships and Peer Support, Paul H. Brooks Pub. Co., 1999.
- [7] Wajcman, J., Feminism Confronts Technology, Pennsylvania State Univ. Press, 1991.
- [8] 통신비밀보호법, 1993년 제정(법률 제4650호), 2004년 제8차 개정(법률 제7138호)