



### 1. 서론

최근 청소년의 교우관계에 있어서 집단 따돌림과 불량 그룹이 사회적으로 큰 문제가 되고 있다[1,2]. 첫째, 청소년 사이에서 특정 학생을 따돌리거나 괴롭히는 등의 집단 따돌림 현상이 급증하고 있다[1,3]. 이러한 집단 따돌림은 단순한 괴롭힘의 형태에서 벗어나 자살, 살인, 가정 붕괴 등과 같은 심각한 사회적 문제를 일으키고 있다. 둘째, 바람직하지 못한 교우관계를 바탕으로 일진회와 같은 불량 그룹 발생이 증가하고 있고, 이러한 불량 그룹은 폭력행사나 금품 갈취 등 집단적 일탈 행위의 문제를 일으키고 있다[2]. 집단 따돌림이나 불량 그룹은 기본적으로 부적절한 교우관계에서 나타나는 단편적인 모습이라 할 수 있다. 따라서 개별 문제에 대한 해결책보다는 전반적인 교우관계 분석을 바탕으로 한 해결책이 필요한 실정이다.

본 논문에서는 학생들의 교우관계를 보다 객관적으로 분석하는 방법을 제안한다. 교우관계를 분석하기 위해 사회학에서 연구된 방법은 크게 두 가지로 나눌 수 있다[4]. 첫째, 학생과 상담을 하거나 생활기록부를 관리하는 등의 관찰자 중심 방법이 있다. 둘째, 학생의 제보나 학생과의 면담을 이용하는 대상자 중심 방법이 있다. 그러나 두 방법 모두 관찰자 혹은 대상자의 주관적인 판단에 의존적이기 때문에 교우관계를 바르게 파악하지 못할 가능성이 매우 크다. 또한, 분석 방법이 비교적 단순하여 왜곡된 형태의 교우관계를 파악할 가능성이 높다. 이러한 문제를 해결하기 위해서는 첫째, 교우관계 파악을 위한 데이터가 학생 생활에서 얻을 수 있는 객관적인 것이어야 하고, 둘째, 그 데이터를 바탕으로 한 분석 방법이 보다 체계적이어야 한다. 본 논문에서는 객관적인 데이터로서 통신이력 데이터를 사용하며, 체계적 분석을 위하여 데이터 마이닝 기법을 활용한다.

본 논문에서 제안하는 교우관계 분석 방법은 통신이력 데이터에 기반한다. 인간은 타인과 대화하기 위해 전화, 이메일, 메신저, 편지 등과 같은 통신도구를 사용한다. 이들 통신도구 중에서, 비교적 최근에 개발된 전화, 메신저, 이메일 등은 통신 과정에서 **통신이력 데이터** (communication history record)가 저장된다. 그런데, 이들 통신이력 데이터는 인간관계 정보를 묵시적으로 내포하고 있다고 볼 수 있다. 즉, 통신을 많이 할수록 일반적으로 가까운 인간관계를 형성한다고 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 파악하는 방법을 제시한다.

본 논문에서는 통신이력 데이터를 기반으로 집단 따돌림 및 불량 그룹을 조기에 파악할 수 있는 체계적인 접근법을 제안한다. 제안하는 접근법의 구성은 다음과 같다. 첫째, 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. 결국, 이렇게 계산된 교우간 친밀도는 어떤 학생이 다른 학생을 얼마나 친밀하게 생각하는지를 나타내는 정량적 지표라 할 수 있다. 둘째, 계산한 교우간 친밀도를 사용하여 집단 따돌림의 가능성이 높은 학생을 파악한다. 이는 집단 따돌림 가능성이 높은 학생은 다른 학생으로부터의 통신 시도가 적을 것이라는 직관에 기반한다. 즉, 다른 학생들로부터 특정 학생으로의 친밀도가 낮으면 그 학생은 집단 따돌림의 가능성이 높다고 판단한다. 셋째, 교우간 친밀도를 데이터 마이닝 기법[5,6] 중 하나인 클러스터링[7-12]으로 분석하여 의미 있는 교우집단을 파악한다. 클러스터링 기법은 유사성을 가지는 동질적인 집단을 판별하는데 사용되므로, 이를 사용하면 친밀도가 높은 의미 있는 교우집단을 분석할 수 있다. 클러스터링 기법을 위해 본 논문에서는 교우간 친밀도를 사용하여 두 학생 사이의 교우간 **유사도** (*similarity*)를 정의한다. 직관적으로 설명하면, 두 학생의 공통된 친구가 많을수록 두 학생의 교우간 유사도는 높게 나타난다. 그런 다음, 클러스터링 기법의 하나인 ROCK 알고리즘[7,13]을 사용하여 의미 있는 클러스터, 즉 교우집단을 분석한다.

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 분석한다. 마지막으로, 클러스터링 기법을 적용하여 의미 있는 집단을 파악하고, 그 의미를 해석한다. 제안한 방법은 통신이력이라는 비교적 객관적인 데이터에 묵시적으로 포함되어 있는 교우간 친밀도 정보를 추출하고, 이를 집단 따돌림 및 교우집단 분석에 사용한다. 따라서, 본 논문에서 제안한 방법을 교우관계 분석에 활용할 경우, 교사 및 학부모가 문제 학생 및 문제 그룹을 조기에 파악하여, 학생들을 매우 객관적으로 지도할 수 있다고 사료된다.

본 논문의 구성은 다음과 같다. 제2장은 관련 연구로

서 기존의 교우관계 분석 연구를 소개한다. 제3장에서는 통신이력 데이터 기반의 교우간 친밀도를 정형적으로 정의하고, 이를 계산하는 방법을 제안한다. 제4장에서는 교우간 친밀도를 바탕으로 집단 따돌림 가능성이 높은 학생을 추출하는 방법과 의미 있는 교우집단을 파악하는 방법을 제안한다. 제5장에서는 제안한 방법의 구현과 실제 실험 결과를 제시하고, 그 의미를 분석한다. 마지막으로, 제6장에서는 결론을 내리고 향후 연구 과제를 제시한다.

## 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[4]. 사회는 다양한 개인들의 집합체이며, 이들 간의 관계는 수직적, 수평적으로 복잡하게 얽혀있다. 사회학에서는 이러한 복합적인 사회적 관계가 형성되는 원인에 주목하여, 원인과 원인들 간의 상호관계를 밝혀내려 노력해 왔다. 그리고 복합적인 사회적 관계 분석의 하나인 인간관계 분석을 위하여, 갈등론, 상징적 상호작용론, 구조기능주의 등 다양한 조사 방법론이 연구되었다[4,14,15].

사회학의 분석 방법에서 개인과 개인 간의 관계, 즉 인간관계나 교우관계 분석에 많이 사용되는 방법은 다음과 같다[14,15].

- 상호작용분석: 집단 내 구성원들 간에 발생하는 상호작용 패턴을 분석하여 구성원 사이의 관계를 찾아내는 방법으로, Bales의 표준적 관찰분류목적체계, Comel의 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[14-17]. 첫째, 이들 방법은 관찰을 통해 결론을 이끌어내므로, 부정확한 관찰이나 잘못된 주관에 의한 결론 유추의 가능성이 항상 존재한다. 둘째, 몇 개의 관찰에 근거하여 일반적인 결론을 이끌어내는 오류를 범할 수 있다. 셋째, 전체 집단이 아닌 관찰자나 제보자의 선별된 경험을 바탕으로 일부에 대한 결론을 전체에 대한 결론으로 유추할 수 있다. 넷째, 원인과 결과가 비논리적으로 연결되는 비논리적 추론이나 결론에서 잘못된 원인을 이끌어내는 오류를 범할 수 있다.

지금까지 설명한 바와 같이, 사회학에서 제시된 기존 방법은 데이터 수집 과정의 문제로 인하여 분석 결과의

객관성을 확보하기 어려운 문제점이 있다. 즉, 분석 데이터의 수집이 관찰자의 주관적인 판단, 대상자의 직접 입력, 타인의 제보 등에 의존하는 관계로 데이터 자체의 객관성이 결여될 수 있다. 따라서, 일상적인 생활에서 발생하는 정보 속에서 분석 데이터를 찾아내고, 이를 교우관계에 활용하는 시도가 필요하다. 본 연구에서는 일상 생활에서 발생하는 통신이력 데이터에서 교우간 친밀도 정보를 찾아내고, 이를 사용하여 교우관계를 분석한다.

자연과학 및 공학 분야에서는 웹이나 데이터에 내재된 사람들간의 사회적 네트워크(social network)를 찾아내거나 표현하는 연구가 진행되었다[18-21]. 우선, 참고 문헌 [18]에서는 학생들 간의 사회적 네트워크를 그래프로 표현하는 방법을 제시하였다. 다음으로, 참고 문헌 [19]에서는 테러리스트가 고용되는 관계를 사회적 네트워크로 표현하는 방법을 연구하였다. 그리고, 참고 문헌 [20]에서는 비즈니스 과정에서 발생하는 패턴에 기반하여 사회적 네트워크를 마이닝하는 연구가 수행되었다. 또한, 참고 문헌 [21]에서는 웹 상에서 사회적 네트워크를 찾아내는 방법을 제안하였다. 이와 같이, 자연과학 및 공학에서 수행된 기존 연구는 주어진 환경에서 특정 응용에 적합한 사회적 네트워크를 찾거나 표현하는 방법으로 볼 수 있으며, 본 논문에서 초점을 맞춘 집단 따돌림 분석이나 불량 집단의 판별과 같은 “교우관계 분석 응용”에는 직접 이용하기 어려운 문제점이 있다. 따라서, 본 논문에서는 일반적인 사회적 네트워크가 아닌 “교우관계 분석 응용”에 중점을 둔 새로운 분석방법을 제안한다.

## 3. 통신이력 데이터 기반의 교우 친밀도 계산

교우관계는 학생들 사이의 친밀한 정도로 표현한다고 할 수 있다. 본 장에서는 학생들 사이의 친밀한 정도를 객관적 데이터인 통신이력 데이터로 계산하는 체계적인 방법을 제시한다.

### 3.1 통신이력 데이터

과학과 컴퓨터의 발전과 더불어, 인간은 타인과의 대화를 위하여 여러 가지 통신도구를 사용한다. 이러한 통신도구의 예로는 전화, 이메일, 문자 메시지, 메신저, 편지 등이 있을 수 있다. 그런데, 이들 많은 통신도구들은 통신을 시도하는 과정에서 통신이력을 남기게 된다. 즉, 통신이력 데이터는 통신을 시도하는 과정에서 저장되는 발신자, 착신자, 통신 시간 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, 통신 과정에서 발생하는 통신이력 데이터는 컴퓨터에 의해 자동적으로 저장된다. 따라서 저장된 통신이력 데이터는 객관성을 신뢰할 수 있고, 이 데이터를 교우관계 분석에 사용하면 보다 객관적인 분석이 가능하게 된다.

일반적으로 통신회사는 통신 과정에서 발생하는 통신이력 데이터를 보관한다. 실제로 통신이력 데이터는 통신사실확인자료 신청을 통해 얻을 수 있으며, 이력 데이터에는 통신일시, 발신자, 수신자 등의 정보가 수록되어 있다[22]. 표 1은 이러한 통신이력 데이터의 구성 예를 나타낸다. 표 1에서 보는 바와 같이 대부분의 통신이력 데이터는 발신자(혹은 송신자)와 착신자(혹은 수신자) 정보를 포함하고 있다. 본 논문에서는 이러한 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 2의 표기법을 사용한다.

**3.2 교우간 친밀도**

본 절에서는 교우간 친밀도를 정형적으로 정의하고, 이를 계산하는 방법을 제안한다. 학생  $x$ 에서 학생  $y$ 로의 교우간 친밀도 개념은 “학생  $x$ 가  $y$ 로 통신 시도를 많이 할수록, 학생  $x$ 는  $y$ 로 친밀하다”는 직관을 바탕으로 정의한다. 이러한 개념에 따라 교우간 친밀도를 다음과 같이 정의한다.

**정의 1.** 통신이력 데이터가 주어졌을 때, 학생  $x$ 에서 학생  $y$ 로의 교우간 친밀도  $dof(x,y)$ 는 학생  $x$ 가  $y$ 로 시도한 통신 횟수에 대한 학생  $x$ 가 시도한 전체 통신 횟수의 비율로 정의한다. 즉, 학생  $x$ 에서  $y$ 로의 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 학생  $y$ 로 자주 통신을 시도할수록 친밀도가 높아진다. 일반적으로 친한 친구일수록 많은 통신을 시도할 것이므로, 이는 일반적인 상황에서 매우 합리적이고 직관적인 척도라 할 수 있다.

**예제 1.** 표 3과 같이 학생  $x$ 의 통신 시도 횟수가 있다고 하자. 표에서 보듯이, 학생  $x$ 는 전화, 이메일, 문자 메시지의 세 가지 통신도구를 사용하고 있다. 이 경우, 학생  $x$ 에서  $y$ 로의 교우간 친밀도는 정의 1에 의해 다음과 같이 0.24로 계산 된다.

$$dof(x,y) = \frac{\sum_{i=1}^3 f_i(x,y)}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^3 f_i(x,y) \right)} = \frac{\sum_{i=1}^3 f_i(x,y)}{\sum_{i=1}^3 \left( \sum_{\text{for all } y \in S} f_i(x,y) \right)} = \frac{20+15+50}{100+50+200} = \frac{85}{350} = 0.24 \quad \square$$

그런데, 각각의 학생은 자신이 선호하는 통신도구를 다를 수 있으므로, 이를 고려한 교우간 친밀도 계산이 필요하다. 즉, 친밀도를 계산할 때 각 학생이 선호하는 통신도구에 대해 가중치를 부여하는 것이 바람직하다. 이를 위해 각 학생 별로 자신이 사용하는 전체 통신 도구 중에서 특정 통신도구의 사용 비율을 가중치로 계산하고, 이를 교우간 친밀도 계산에 적용한다. 결국, 학생  $x$ 의 통신 도구  $k$ 에 대한 가중치란 학생  $x$ 가 전체 통신 도구 중에서  $k$ -번째 통신도구를 사용하는 비율로 볼 수 있다. 이러한 통신 도구 가중치를 다음과 같이 정형적으로 정의한다.

**정의 2.** 통신이력 데이터가 주어졌을 때, 학생  $x$ 의

표 1 통신이력 데이터의 구성 예제

통신도구	통신이력 데이터 구성
전화(핸드폰)	날짜, 발신번호, 수신번호, 사용항목, 사용내역, 통화 초/건 등
메신저	날짜, 시간, 보낸 사람, 받는 사람, 메시지
이메일	보낸 사람, 받는 사람, 날짜 및 시간, 제목, 내용

표 2 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 통신도구를 사용하여 학생 $x$ 가 학생 $y$ 로 통신을 시도한 횟수 ( $1 \leq i \leq m$ )

표 3 학생  $x$ 의 통신 시도 횟수 예제

구분	전화( $f_1(-,-)$ )	이메일( $f_2(-,-)$ )	문자 메시지( $f_3(-,-)$ )
전체 통신 횟수	100	50	200
학생 $y$ 로의 통신 횟수	20	15	50

$k$ -번째 통신도구에 대한 가중치  $weight(x, k)$ 는 학생  $x$ 가 시도한 전체 통신 횟수 중에서  $k$ -번째 통신도구를 사용한 비율에 전체 통신도구의 개수를 곱한 값으로 정의한다. 즉, 가중치  $weight(x, k)$ 는 다음 식 (2)와 같이 정의한다.

$$weight(x, k) = \frac{\text{학생 } x \text{가 } k\text{-번째 통신도구를 사용한 횟수}}{\text{학생 } x \text{가 시도한 전체 통신 횟수}}$$

$$= m \cdot \frac{\sum_{\text{for all } y \in S} f_k(x, y)}{\sum_{i=1}^m \left( \sum_{\text{for all } y \in S} f_i(x, y) \right)} \quad (2)$$

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 횟수**라 부른다. 그리고, 이러한 가중치 기반 통신 횟수를 사용하여 가중치 기반 친밀도를 정의하면 다음과 같다.

**정의 3.** 통신이력 데이터가 주어졌을 때, 학생  $x$ 에서 학생  $y$ 로의 **가중치 기반 친밀도**  $w dof(x, y)$ 는 학생  $x$ 가  $y$ 로 시도한 가중치 기반 통신 횟수와 학생  $x$ 가 시도한 전체 통신 횟수의 비율로 정의한다. 즉, 학생  $x$ 에서  $y$ 로의 가중치 기반 친밀도  $w dof(x, y)$ 는 다음 식 (3)과 같이 정의한다.

$$w dof(x, y) = \frac{\text{학생 } x \text{가 } y \text{로 시도한 가중치 기반 통신 횟수}}{\text{학생 } x \text{가 시도한 전체 통신 횟수}}$$

$$= \frac{\sum_{i=1}^m (weight(x, i) \cdot f_i(x, y))}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^m f_i(x, y) \right)} \quad (3)$$

**예제 2.** 예제 1에서 사용한 표 3과 같이 학생  $x$ 의 통신 시도 횟수가 주어졌다고 하자. 이 경우, 학생  $x$ 의 통신도구 별 가중치는 다음과 같이 구할 수 있다.

$$weight(x, 1) = 3 \cdot \frac{\sum_{\text{for all } y \in S} f_1(x, y)}{\sum_{i=1}^3 \left( \sum_{\text{for all } y \in S} f_i(x, y) \right)}$$

$$= 3 \cdot \frac{100}{100 + 50 + 200} = \frac{300}{350} \cong 0.86$$

$$weight(x, 2) = 3 \cdot \frac{50}{100 + 50 + 200} = \frac{150}{350} \cong 0.43$$

$$weight(x, 3) = 3 \cdot \frac{200}{100 + 50 + 200} = \frac{600}{350} \cong 1.71$$

상기 결과에 따라, 학생  $x$ 의 경우, 전체 통신도구 중에서 전화( $f_1$ )는 한 번 시도할 경우 0.86번 사용한 것으로 가중치를 부여하고, 이메일( $f_2$ )은 0.43의 가중치를 부여하며, 문자 메시지( $f_3$ )는 1.71의 가중치를 부여한다. 즉, 학생  $x$ 가 자주 사용하는 문자 메시지에 많은 가중치를 부여하는 것이다. 이를 바탕으로, 학생  $x$ 에서  $y$ 로의 가중치 기반 친밀도를 구하면 다음과 같다.

$$w dof(x, y) = \frac{\sum_{i=1}^3 (weight(x, i) \cdot f_i(x, y))}{\sum_{\text{for all } y \in S} \left( \sum_{i=1}^3 f_i(x, y) \right)}$$

$$= \frac{0.86 \cdot 20 + 0.43 \cdot 15 + 1.71 \cdot 50}{350} \cong \frac{109.2}{350} \cong 0.31$$

정의 1의 친밀도와 정의 3의 가중치 기반 친밀도의 차이는 개인의 통신도구 선호도를 고려하는지 그렇지 않는지의 차이이다. 특히, 예제 1과 2의 친밀도 값에 차이가 있음을 알 수 있다. 이는 학생  $x$ 가 다른 통신도구에 비해 문자 메시지를 많이 사용하기 때문에, 가중치를 고려할 경우 문자 메시지 사용 비율이 높은 학생  $y$ 로의 가중치 기반 친밀도가 단순한 친밀도보다 높게 나오는 것이다. 즉, 정의 1의 교우간 친밀도는 개인의 통신도구 선호도를 고려하지 않고 단순히 통신 횟수의 합을 사용하는 반면, 정의 3의 가중치 기반 친밀도는 개인의 통신도구 선호도인 가중치를 고려하여 통신 횟수의 합을 계산하여 사용한다. 본 논문에서는 개인의 선호도를 반영하기 위하여 정의 3의 가중치 기반 친밀도를 사용한다. 그리고, 혼란이 없는 한 앞으로 친밀도를 가중치 기반 친밀도와 동일한 개념으로 사용한다.

그림 1은 통신이력 데이터를 바탕으로 교우간 친밀도를 계산하는 전체적인 과정을 나타낸다. 그림에서와 같이, 통신이력 데이터 집합이 입력으로 주어지고, 교우간 친밀도 집합이 결과로 출력된다. 그리고 이때의 교우간 친밀도는 정의 2의 가중치와 정의 3의 친밀도에 기반하여 계산된다.

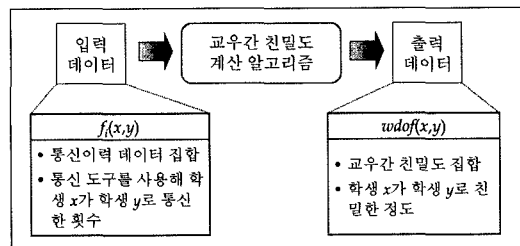


그림 1 교우간 친밀도 계산 과정

다음으로, 그림 2는 통신이력 데이터를 바탕으로 교우간 친밀도를 추출하는 알고리즘 *ComputeWdof()*를 나타낸다. 알고리즘의 입력은 학생들의 통신 과정에서 발생하는 통신이력 데이터  $f_i(x,y)$ 의 집합이다. 알고리즘은 통신이력 데이터를 바탕으로 각각의 학생  $x$ 에서 학생  $y$ 로의 친밀도를 계산하는 일련의 과정으로 구성되어 있다. 스텝 (1)~(3)을 보면, 정의 2의 식 (2)를 사용하여 각 학생  $x$ 의 통신도구 별 가중치를 구한다. 다음으로, 스텝 (4)~(6)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  $[1, m]$ 
(3)      $weight(x,k) = m \cdot \frac{\sum_{for\ all\ y \in S} f_k(x,y)}{\sum_{i=1}^m \left( \sum_{for\ all\ y \in S} f_i(x,y) \right)}$ ;
(4) for each  $x$  in  $S$ 
(5)   for each  $y$  in  $S$  ( $y \neq x$ )
(6)      $wdof(x,y) = \frac{\sum_{i=1}^m (weight(x,i) \cdot f_i(x,y))}{\sum_{for\ all\ y \in S} \left( \sum_{i=1}^m f_i(x,y) \right)}$ ;
end
    
```

그림 2 교우간 친밀도 계산 알고리즘

#### 4. 집단 따돌림과 교우집단 분석

본 장에서는 제3장에서 계산한 교우간 친밀도를 사용하여 집단 따돌림 가능성이 높은 학생과 의미 있는 교우집단을 판별하는 방법을 제안한다. 먼저, 4.1절에서는 특정 학생에 대한 친밀도의 합으로 집단 따돌림 가능성이 높은 학생을 판단하는 방법과 알고리즘을 기술한다. 다음으로, 제4.2절에서는 친밀도 데이터를 클러스터링 기법으로 분석하여 의미 있는 교우집단을 추출하는 방법과 알고리즘을 설명한다.

##### 4.1 집단 따돌림 학생 판단

일반적으로 집단 따돌림 학생이란 다른 학생들로부터 소외 받는 학생으로 볼 수 있다. 다시 말하면, 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 친밀도

를 계산하였을 때, 특정 학생으로의 교우간 친밀도의 합이 매우 낮은 학생은 집단 따돌림의 가능성이 높다. □

제3장에서 설명한 바와 같이, 학생  $x$ 에서  $y$ 로의 교우간 친밀도란 학생  $x$ 가  $y$ 로 얼마나 자주 통신을 시도하느냐로 정의된다. 따라서, 가정 1에 따르면 집단 따돌림 가능성이 높은 학생은 다른 학생으로부터 해당 학생으로의 통신 시도가 매우 적은 학생으로 해석할 수 있다. 그리고, 이러한 가정은 통신이력 데이터에서 집단 따돌림을 판단하는 비교적 객관적인 방법이라 할 수 있다.

가정 1에 기반하여 집단 따돌림 가능성이 높은 학생을 판별하기 위한 방법은 다음과 같다. 먼저, 다른 모든 학생으로부터 학생  $y$ 로의 교우간 친밀도의 합은

$\sum_{for\ all\ x \in S} wdof(x,y)$  를 계산한다. 그런 다음, 가정 1에 기반하여 학생  $y$ 로의 친밀도의 합, 즉  $\sum_{for\ all\ x \in S} wdof(x,y)$

의 값이 다음 식 (4)와 같이 매우 작은 경우에, 학생  $y$ 는 집단 따돌림 가능성이 높다고 판단한다.

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

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 주어진 임계치  $\theta$  이하이면, 학생  $y$ 는 **집단 따돌림**이라 정의한다. 즉, 다음 식 (5)를 만족하는 학생  $y$ 를 집단 따돌림이라 정의한다.

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

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

다음으로, 그림 3은 정의 4에 기반하여 집단 따돌림 가능성이 높은 학생을 판별하는 알고리즘 *Find-Bullied()*을 나타낸다. 알고리즘의 입력은 교우간 친밀도인  $wdof(x,y)$ 의 집합과 임계치  $\theta$ 이다. 알고리즘을 보면, 각 학생  $y$ 에 대한 교우간 친밀도 합을 사용하여 집단 따돌림 여부를 판단함을 알 수 있다. 스텝 (2)를 보

면, 학생  $y$ 로의 친밀도의 합  $\sum_{for\ all\ x \in S} wdof(x,y)$  이 주어진 임계치인  $\theta$  이하이면, 학생  $y$ 는 집단 따돌림의 가능성이 높다고 판단한다. 반면에, 스텝 (3)에서와 같이 친밀도의 합이  $\theta$ 보다 크면, 집단 따돌림의 가능성이 낮다고 판단한다.

```

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

그림 3 집단 따돌림 학생 판단 알고리즘

4.2 의미 있는 교우집단 분석

제3장에서 정의한 교우간 친밀도는 한 학생에서 다른 학생으로의 친밀한 정도를 표현한다. 그런데, 이와 같은 교우간 친밀도 자체만으로는 학생들 내부에 존재하는 교우집단을 파악하기는 어렵다. 따라서, 본 절에서는 유사도(similarity)를 바탕으로 동질적 집단을 판별하는 클러스터링 기법[7,8,10]으로 교우집단을 분석하는 방법을 제안한다. 클러스터링 기법으로는 두 객체간의 유사도가 정의된 경우에 효과적인 ROCK 알고리즘[7,13]을 사용한다.

ROCK 알고리즘을 사용하기 위해서는 두 객체, 즉 두 학생간의 유사도를 정의하여야 한다. 이는 ROCK 알고리즘을 비롯한 대부분의 클러스터링 기법이 두 객체간의 유사도를 기반으로 클러스터링을 수행하기 때문이다. 따라서, ROCK 알고리즘을 사용하여 교우집단을 판별하기 위해서는, 통신이력 데이터를 바탕으로 두 학생간의 유사도를 먼저 계산하여야 한다. 그런데 이미 제3장에서 우리는 통신이력 데이터를 바탕으로 교우간 친밀도를 정의하고 계산하는 방법을 제안하였다. 따라서, 본 본문에서는 두 학생간의 유사도를 제3장에서 구한 교우간 친밀도를 사용하여 정의하고 계산한다.

직관적으로 생각했을 때, 두 학생의 친구들이 비슷하다면, 즉 두 학생이 공통된 친구들을 많이 가진다면, 두 학생은 유사한 교우관계를 가진다고 말할 수 있다. 이에 따라, 본 논문에서는 두 학생  $x$ 와  $y$ 의 교우간 유사도를 “학생  $x$ 의 친구들과 학생  $y$ 의 친구들이 비슷한 정도, 즉 두 학생  $x$ 와  $y$ 가 공통된 친구들을 가지는 정도”라 개념적으로 정의한다. 이 같은 개념으로 교우간 유사도를 구하기 위해서는, 우선 각 학생이 갖는 친구 목록을 구하여야 한다. 각 학생의 친구 목록은 제3장에서 구한 교우간 친밀도가 임계치  $\delta$ 보다 큰 학생들을 대상으로 구성한다. 친구 여부를 판단하는 임계치  $\delta$ 는 교사 등에 의해 주어지는 값으로서, 0 이상의 값을 가질 수 있다. 임계치  $\delta$ 는 어떤 학생이 다른 학생을 어느 정도 친밀하게 보아야 친구로 삼느냐는 기준치가 된다. 본 논문에서는 문제를 간략히 하기 위해, 임계치  $\delta$  값으로 0을 사용한다. 즉, 한 번이라도 통신을 시도한 학생은 친구 관계가

있다고 가정한다. 이에 따라서,  $w dof(x, y)$ 가  $0(=\delta)$ 보다 큰 경우, 학생  $y$ 를 학생  $x$ 의 친구 목록  $f list(x)$ 에 포함시킨다.

다음으로, 각 학생들의 친구 목록을 사용하여 다음과 같이 두 학생의 교우간 유사도  $sim(x, y)$ 를 정의한다.

정의 5. 학생  $x$ 와  $y$ 의 친구 목록을 각각  $f list(x)$ 와  $f list(y)$ 라 했을 때, 두 학생  $x$ 와  $y$ 의 교우간 유사도  $sim(x, y)$ 는 두 목록의 교집합과 두 목록의 합집합의 비율로 정의한다. 즉, 두 학생  $x$ 와  $y$ 의 교우간 유사도  $sim(x, y)$ 는 다음 식 (6)과 같이 정의한다.

$$sim(x, y) = \frac{f list(x) \cap f list(y)}{f list(x) \cup f list(y)} \tag{6}$$

정의 5의 교우간 유사도는 간략히 이야기해서 두 학생의 공통된 친구 비율을 의미한다. 즉, 식 (6)에서 분모는 합집합이므로 “두 학생이 가지는 모든 친구들”을 나타내고, 분자는 교집합이므로 “두 학생이 공통으로 가지는 친구들”을 나타낸다. 따라서, 정의 5의 식 (6)은 두 학생이 얼마나 많은 친구를 공통으로 갖는지를 나타내는 교우간 유사도로 사용할 수 있다.

그림 4는 교우간 친밀도를 바탕으로 교우간 유사도를 계산하는 알고리즘  $ComputeSim()$ 이다. 알고리즘의 입력은 교우간 친밀도인  $w dof(x, y)$ 의 집합과 임계치  $\delta$ 이다. 알고리즘을 보면, 우선 스텝 (1)~(3)에서 각각의 학생들을 대상으로 친구 목록인  $f list(x)$ 를 교우간 친밀도를 사용하여 구한다. 다음으로, 스텝 (4)~(6)에서는 앞서 구한 친구 목록을 사용하여 두 학생  $x$ 와  $y$ 의 교우간 유사도를 구한다. 이때 사용하는 스텝 (6)의 교우간 유사도는 식 (6)에 따른 것이다.

그림 4의  $ComputeSim()$ 을 사용하여 교우간 유사도를 구한 후에는 이를 사용하여 클러스터링을 수행한다. 이를 위해서 우선 다음과 같이 교우간 유사도와 클러스터와의 관계를 가정한다.

```

procedure ComputeSim (A set of  $w dof(x, y)$ 's, Threshold  $\delta$ )
begin
(1) for each  $x$  in  $S$ 
(2)   for each  $y$  in  $S$  ( $y \neq x$ )
(3)     if  $w dof(x, y) > \delta$  then Include  $y$  into  $f list(x)$ ;
(4) for each  $x$  in  $S$ 
(5)   for each  $y$  in  $S$  ( $y \neq x$ )
(6)      $sim(x, y) = \frac{f list(x) \cap f list(y)}{f list(x) \cup f list(y)}$ ;
end
    
```

그림 4 교우간 유사도 계산 알고리즘

표 4 설문에 참여한 학생들의 통신도구 사용 통계

구분	전화	문자 메시지	메신저	이메일
통신도구 별 사용 횟수	1,993	10,441	28,391	106
통신도구 별 사용 비율(%)	4.9	25.5	69.4	0.3

**가정 2.** 통신이력 데이터를 기반으로 교우간 유사도를 구하였을 때, 교우간 유사도가 높은 학생들은 동일한 교우 클러스터(교우집단)에 속할 가능성이 높다. □

가정 2에 따라서 교우간 유사도를 클러스터링 기법인 ROCK 알고리즘[7]에 적용하면 유사성이 높은 클러스터, 즉 의미 있는 교우집단을 판별해 낼 수 있다.<sup>1)</sup>

### 5. 구현 및 실험

본 장에서는 제안한 교우관계 분석 방법의 실제 구현 내용과 실험 결과를 설명한다. 제5.1절에서는 구현 내용과 실험 환경을 소개한다. 그리고, 제5.2절과 제5.3절에서는 집단 따돌림 분석 결과와 교우집단 분석 결과를 각각 설명한다.

#### 5.1 구현 및 실험 환경

실험을 위하여 우선 제3장에서 제안한 교우간 친밀도 알고리즘 *ComputeWdof()*를 구현하였다. 그리고 집단 따돌림 분석을 위하여 제4장에서 제안한 알고리즘 *FindBullying()*을 구현하였다. 또한, 교우집단 분석을 위하여, 제4장의 교우간 유사도 계산 알고리즘 *ComputeSim()*과 ROCK 알고리즘을 구현하였다. 모든 실험은 Intel Pentium 4 3.0GHz CPU, 392MB 메모리, 30GB 하드 디스크를 가진 Windows 2000 서버에서 수행하였으며, 알고리즘의 구현은 Borland Delphi 언어 [22]를 사용하였다.

통신이력 데이터는 설문 조사를 통하여 수집하였다. 실제로 통신이력 데이터는 통신회사에서 얻을 수 있으나, 통신비밀보호법[23]과 개인정보보호법[24]으로 인해 개인의 동의가 있어야만 이를 얻을 수 있다. 그런데, 많은 학생들을 대상으로 개개인의 동의를 얻기에는 기간 및 여건상 어려운 상황이어서 본 논문에서는 설문조사를 통해 실험에 사용할 통신이력 데이터를 수집하였다.<sup>2)</sup> 설문조사는 경기도 양평에 위치한 모 고등학교 2학년의 6개 학급 모든 학생을 대상으로 2005년 6월 16일부터 25일 사이에 실시하였다. 설문조사를 실시한 관계로, 수집된 데이터는 통신이력을 기반으로 얻을 수 있는 데이

타에 비해 그 수가 적고 정확도가 떨어지는 단점이 있다. 이로 인해, 실험 결과의 신뢰성이 떨어지는 문제가 있으나, 본 논문에서 제시한 방법론이 빠르게 동작함을 실험하고 그 결과를 해석하기에는 충분하다고 사료된다. 또한, 이는 통신이력 데이터를 사용할 수 있는 실제 환경에서는 자연스럽게 해결될 수 있는 문제이다. 설문조사 내용은 각 학생이 다른 학생에게 핸드폰, 문자 메시지, 메신저, 이메일을 사용하여 통신한 발신자, 착신자, 통신 횟수이다.

설문조사 대상 학생은 총 143명이고, 학생들이 통신도구를 사용한 비율은 표 4와 같다. 통계 결과에 따르면 학생들은 주로 메신저와 문자 메시지를 통신도구로 사용한다는 것을 알 수 있다. 반면에 전화의 비율은 비교적 낮으며, 이메일은 거의 사용하지 않는 것으로 나타났다.

#### 5.2 집단 따돌림 분석 결과

제4장에서 제안한 바와 같이, 집단 따돌림 가능성이 높은 학생이란 다른 학생으로부터 받은 친밀도 합

$\sum_{\text{for all } x \in S} wdof(x, y)$  이 낮은 학생  $y$ 를 의미한다. 따라서, 집단 따돌림 분석 실험에서는 임계치  $\theta$ 를 변화시키면서

친밀도 합  $\sum_{\text{for all } x \in S} wdof(x, y)$  이 임계치  $\theta$  이하인 학생들을 조사하였다.

분석 결과 각 학생의 교우간 친밀도 합의 평균은 0.90으로 나타났다. 친밀도 합의 평균이 1보다 작게 나온 이유는 다른 학생에게 단 한번도 연락을 하지 않은 학생이 있기 때문이다. 즉 다른 학생에게 보내는 친밀도의 합이 1이 아니라 0인 학생이 있기 때문에 나타난 결과이다. 평균을 중심으로 친밀도 합을 간단히 해석하면 다음과 같다. 먼저, 교우간 친밀도의 합의 평균보다 낮은 학생은 자신을 좋아하는 친구들이 상대적으로 적음을 의미한다. 다음으로, 친밀도의 합의 평균보다 높은 학생은 자신을 좋아하는 친구들이 상대적으로 많음을 의미한다.

그림 5는 임계치  $\theta$ 의 변화에 따른 학생 수의 변화를 나타낸다. 그림을 보면, 임계치가 0.00에서 1.70 사이에 거의 모든 학생(전체의 93.0%)이 고르게 분포하는 것을 알 수 있다. 직관적으로는 대다수의 학생들이 평균인 0.90 주위에 모여 있을 것으로 예상을 했으나, 실제로는 평균을 중심으로 고르게 분포하는 것으로 나타났다. 이는 학생들의 교우관계가 개인적으로 상당한 차이가 있

1) ROCK 알고리즘은 링크를 계산하는 *ComputeLinks()*와 클러스터를 구하는 *Cluster()* 알고리즘으로 구성된다. 그러나 클러스터링을 수행하는 이들 알고리즘은 본 논문의 초점이 아니므로 자세한 설명은 생략한다.  
 2) 교사나 학부모가 교우관계 파악을 목적으로, 학생들의 동의 하에 발신자와 착신자의 제한된 정보를 얻는 것은 현실적으로 충분히 가능하다. 그러나, 본 연구는 교우관계 분석 방법에 초점을 맞추었으므로, 단순한 방법인 설문 조사로 통신이력 데이터를 수집하였다.



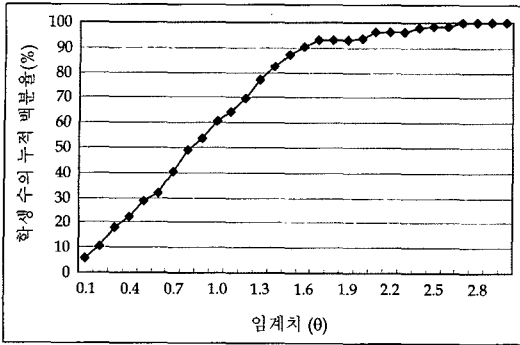


그림 5 임계치 변화에 따른 학생 비율

음을 나타낸다. 다음으로, 그림에서 교우간 친밀도 합이 임계치 0.10, 0.20와 같이 작은 값을 가지는 학생이 집단 따돌림의 가능성이 높은 학생이다. 표 5는 교우간 친밀도의 합이 0.20 이하인 경우의 학생 수와 학생 번호<sup>3)</sup>를 보여준다. 즉, 표 5에 나타난 학생들은 집단 따돌림의 가능성이 높다고 볼 수 있다. 표에서 알 수 있듯이, 친밀도의 합이 0.20 이하로 집단 따돌림의 가능성이 비

교적 높은 학생은 전체의 10.5%에 해당하고, 친밀도의 합이 0.10 이하로 집단 따돌림의 가능성이 매우 높은 학생은 전체의 5.6%에 달함을 알 수 있다. 이와 같이 파악된 집단 따돌림 학생에 대해서는 면담 등을 통해 특별한 지도와 관심이 필요하다고 사료된다.

**5.3 교우집단 분석 결과**

제4장에서 제안한 교우간 유사도 알고리즘과 ROCK 알고리즘을 구현하여, 교우집단 분석을 실시하였다. 실험에서는 전체 학생을 대상으로 교우집단을 분석하기 위해 집단(클러스터)의 개수를 여섯 개로 설정하였다. 다시 말해서, 추출하고자 하는 집단의 개수를 실험 대상인 학교의 2학년 학급 수인 여섯 개로 설정하고 실험을 수행하였다.

직관적으로 생각할 때, 원하는 집단의 수와 실제 학급의 수가 동일하기 때문에 추출된 집단은 각각의 학급과 연관될 것이라 예상했다. 그러나 실제로 추출된 각 집단의 학생들은 모든 학급에 고르게 분포하는 것으로 나타났다. 그림 6은 클러스터링 결과 얻은 집단의 학생 분포를 그림으로 나타낸 것이며, 그림 7은 결과 집단의 반별 분포도를 나타낸다. 그림을 보면, 학생들의 교우집단이 자신이 속한 학급을 중심으로 형성되기 보다는 여러 학급에 걸쳐서 형성되고 있음을 알 수 있다. 특히, 2번 집

3) 세 자리로 구성되는 학생 번호의 첫 자리는 학급을 다음 두 자리는 고유 번호를 의미한다.

표 5 집단 따돌림 가능성이 높은 학생

임계치(θ)	학생 수(명)	백분율(%)	학생 번호
0.00	1	0.7	606
0.05	3	1.4	606, 505, 515
0.10	8	5.6	606, 505, 515, 127, 512, 504, 104, 511
0.15	11	7.7	606, 505, 515, 127, 512, 504, 104, 511, 324, 603, 107
0.20	15	10.5	606, 505, 515, 127, 512, 504, 104, 511, 324, 603, 107, 419, 209, 506, 617

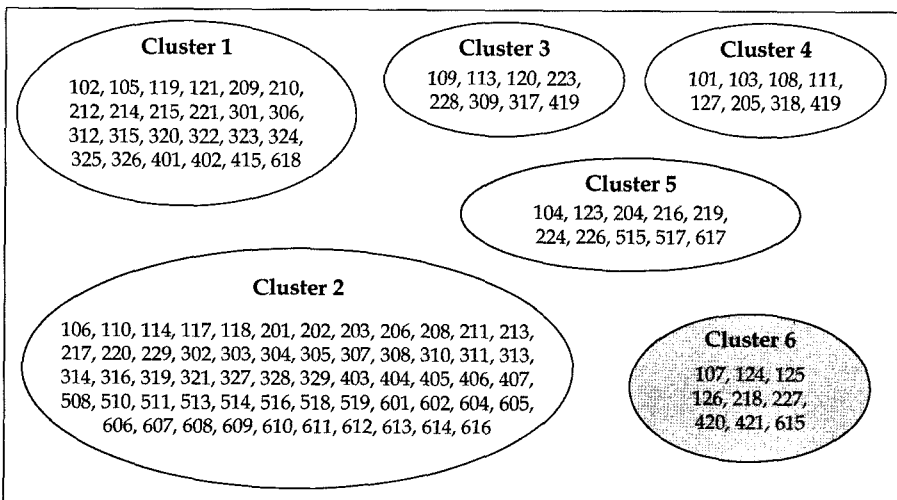


그림 6 클러스터링에 의해 파악된 의미 있는 집단의 학생 분포

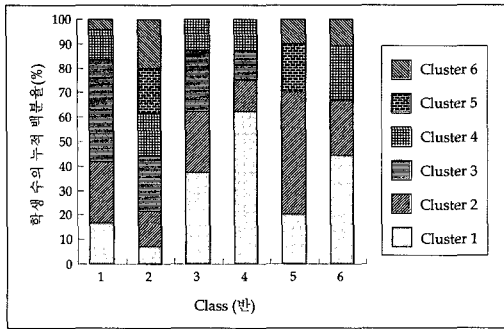


그림 7 클러스터링에 의해 파악된 집단의 반별 분포

단의 학생이 전체 학생의 41%로 나타났는데, 이는 각 학급을 중심으로 교우집단이 형성되기 보다는 전체 학급을 대상으로 교우집단이 형성된다는 점을 더욱 잘 보여준다. 이와 같은 결론은 각 학생을 학급 중심으로만 지도할 것이 아니라 전체 학급을 대상으로 지도하는 새로운 방안이 필요함을 의미한다.

그림 6과 7에서 3번, 4번, 5번, 6번의 각 집단은 구성원이 10명 이하인 소수 집단이다. 이와 같은 소규모 인원의 집단은 대다수 학생들이 형성한 교우집단과 차이가 있다고 볼 수 있다. 결국, 이들 소규모 집단의 학생들은 전체 학생들과는 다른 생활 패턴을 지닌 특수한 집단, 예를 들어 우수 집단, 동아리 집단, 소외 집단, 불량 집단 등을 형성하고 있을 가능성이 높음을 의미한다. 따라서, 이들 의미 있는 소규모 집단에 대해서도 면담 등의 특별한 생활 지도와 관심이 필요하다고 사료된다.

### 6. 결론

본 연구에서는 학생들의 교우관계를 파악하는 객관적이고 체계적인 방법을 제시하였다. 현재까지는 관찰자나 제보자의 주관적인 데이터를 기반으로 교우관계를 파악하였다. 반면에, 본 연구에서는 교우관계를 파악하기 위한 객관적인 데이터로 통신이력 데이터를 사용하며, 이를 바탕으로 집단 따돌림과 교우집단 등의 교우관계를 파악하는 체계적인 방법을 제안하였다.

본 연구의 공헌을 요약하면 다음과 같다. 첫째, 통신이력 데이터 기반의 교우간 친밀도를 정형적으로 정의하고, 이를 계산하는 체계적인 방법을 제안하였다. 둘째, 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. 셋째, 교우간 친밀도를 클러스터링 기법에 적용하여 의미 있는

교우집단을 분석하는 방법을 제안하였다. 이를 위해 우선 교우간 친밀도에서 두 학생의 교우간 유사도를 정형적으로 정의하고, 이를 계산하는 방법과 알고리즘을 제안하였다. 그리고 이러한 교우간 유사도를 사용하여 클러스터링을 수행하는 방법을 제시하였다. 마지막으로, 구현 및 실험을 통해 통신이력 데이터 기반의 교우관계 분석 방법의 실용성을 입증하였다. 이 같은 결과를 볼 때, 본 논문은 실제 학교 현장에서 학생들의 교우관계를 보다 객관적이고 체계적으로 파악할 수 있는 효과적인 방법론이라 사료된다.

향후 연구로는 분석한 결과와 현실에서 나타나는 실제적인 모습이 일치하는지를 확인하는 작업이 필요하다. 더 나아가서, 분석된 결과를 실제 활용할 수 있는 교육적인 지침에 대한 연구가 필요하다.

### 참고 문헌

- [1] 김선애, "집단따돌림 연구경향 분석", 한국청소년연구, 제39호, pp. 73-107, 2004.
- [2] 백운학, "학교폭력과 집단 따돌림 현상의 실태와 대응 방안", 재활심리연구, 제6권, 제6호, pp. 223-245, 1999.
- [3] Lee, C., Preventing Bullying in Schools, Sage, 2004.
- [4] 김경동, 사회학의 이론과 방법론, 한국학술정보, 2003.
- [5] 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.
- [6] Moon, Y.-S., Whang, K.-Y., and Loh, W.-K., "Efficient Time-Series Subsequence Matching using Duality in Constructing Windows," *Information Systems*, Vol. 26, No. 4, pp. 279-293, June 2001.
- [7] Guha, S., Rastogi, R., and Shim, K., "ROCK: A Robust Clustering Algorithm for Categorical Attributes," In *Proc. Int'l Conf. on Data Engineering (ICDE)*, IEEE, Sydney, Australia, pp. 512-521, March 1999.
- [8] Zhang, T., Ramakrishnan, R., and Livny, M., "BIRCH: An Efficient Data Clustering Method for very Large Databases," In *Proc. Management of Data, ACM SIGMOD*, Montreal, Canada, pp. 103-114, June 1996.
- [9] Hwang, J.-J., Whang, K.-Y., Moon, Y.-S., and Lee, B.-S., "A Top-down Approach for Density-Based Clustering Using Multidimensional Indexes," *Journal of Systems and Software*, Vol. 73, Issue 1, pp. 196-180, Sept. 2004.
- [10] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," In *Proc. the 2nd Int'l Conf. on Knowledge Disco-*

very and Data Mining, pp. 226-231, Portland, Oregon, Aug. 1996.

[11] Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P., "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications," In *Proc. Int'l Conf. on Management of Data*, ACM SIGMOD, pp. 94-105, June 1998.

[12] Gibson, D., Kleinberg, J., and Raghavan, P., "Clustering Categorical Data: An Approach Based on Dynamical Systems," *The VLDB Journal*, Vol. 8, Issue 3-4, pp. 222-236, Feb. 2000.

[13] Dutta, M., Mahanta, A. K., and Pujari, A. K., "QROCK: A Quick Version of the ROCK Algorithm for Clustering of Categorical Data," *Journal of Pattern Recognition Letters*, Vol. 26, No. 15, pp. 2364-2373, Nov. 2005.

[14] 강상철, 최건수, "상징적 상호작용론", *교육발전총론*, 제9권, 제1호, pp. 23-39, 1988.

[15] 최순남, *인간행동과 사회환경(제3판)*, 법문사, 2005.

[16] Snell, H. E., et al., *Social Relationships and Peer Support*, Paul H. Brooks Pub. Co., 1999.

[17] Wajzman, J., *Feminism Confronts Technology*, Pennsylvania State Univ. Press, 1991.

[18] Saltz, J. S., Hiltz, S. R., and Turoff, M., "Student Social Graphs: Visualizing a Student's Online Social Network," In *Proc. of Conference on Computer Supported Cooperative Work*, ACM CSCW, Chicago, Illinois, pp. 596-599, Nov. 2004.

[19] Ko, T. H. and Berry, N. M., "Agent-Based Modeling with Social Networks for Terrorist Recruitment," In *Proc. of the 9th National Conference on Artificial Intelligence*, San Jose, California, pp. 1016-1017, July 2004.

[20] Aalst, W. M. P., and Song, M., "Mining Social Networks: Uncovering Interaction Patterns in Business Processes," *Lecture Notes in Computer Science*, Vol. 3080, pp. 244-260, 2004.

[21] Adamic, L. A., Buyukkokten, O., and Adar, E., "A Social Network Caught in the Web," *First Monday*, Vol. 8, No. 6, 2003.

[22] Cantu, M., *Mastering Borland Delphi 2005*, John Wiley & Sons, 2005.

[23] *통신비밀보호법*, 1993년 제정(법률 제4650호), 2004년 제8차 개정(법률 제7138호).

[24] *개인정보보호법*, 1994년 제정(법률 제4743호), 1999년 제1차 개정(법률 제5715호).

9월~2002년 2월 ㈜현대시스코 호처리개발실 선임연구원  
 2002년 2월~2005년 2월 ㈜인프라밸리 기술연구소 기술위원(이사). 2005년 3월~현재 한국과학기술원 첨단정보기술 연구센터 연구원. 2005년 3월~현재 강원대학교 컴퓨터과학과 조교수. 관심분야는 Data Mining, Knowledge Discovery, Stream Data, Storage System, Database Applications, Mobile/ Wireless Communication Services & Systems



**최 훈 영**  
 2003년 2월 강원대학교 컴퓨터공학과 학사. 2006년 2월 강원대학교 교육대학원 컴퓨터교육전공 석사. 2005년 3월~현재 한국과학기술원 첨단정보기술연구센터 연구보조원. 관심분야는 Computer Education, Data Mining, Knowledge Discovery



**김 진 호**  
 1982년 2월 경북대학교 전자공학과 학사. 1985년 2월 한국과학기술원 전산학과 석사. 1990년 2월 한국과학기술원 전산학과 박사. 1995년 8월~1996년 7월 미국 미시간 대학교 객원 교수. 2003년 2월~2004년 2월 미국 Drexel University 객원 교수. 1999년 3월~현재 한국과학기술원 첨단정보기술연구센터 연구원. 1990년 8월~현재 강원대학교 컴퓨터과학과 교수. 관심분야는 Data warehouse, OLAP, Data Mining, Real-time/Embedded Database, Main-memory database, Data Modeling, Web Database Technology



**문 양 세**  
 1991년 2월 한국과학기술원 과학기술대학 전산학과 학사. 1993년 2월 한국과학기술원 전산학과 석사. 2001년 8월 한국과학기술원 전자전산학과 전산학전공 박사. 1993년 2월~1997년 2월 현대전자산업(주) 통신사업본부 주임연구원. 2001년