

학생들의 관계성 파악을 위한 빅-데이터 분석에 관한 연구

황득영* · 김진묵**

요 약

최근 학교 사이버 폭력 문제가 심각하게 발생하여 문제의 심각성이 날로 급증하고 있다. 그 중에서도 스마트폰을 이용한 사이버 폭력의 심각성이 매우 높아 사회적으로 큰 문제로 인식되고 있다. 언어적인 사이버폭력은 물리적인 폭력보다 영향 범위가 넓고 지속 시간이 길어서 학생들과 같이 미숙한 존재에게는 그 피해정도가 매우 심각하다. 그러므로 본 논문에서는 학생들이 사용하는 언어와 학급에서의 관계성을 분석해서 학급 내의 친구들 사이에서 발생할 수 있는 사이버폭력의 징후를 사전에 파악한다. 그리고 파악된 사이버폭력의 위험성에 대해서 학부모나 학급 교사, 학교 안전지킴이 등에게 미리 알려주어 사이버폭력 사고를 미연에 방지할 수 있도록 하고자 한다. 이를 위해서 학생들이 사이버 학급에서 사용하는 메시지를 설계 및 구현하고자 한다. 구현한 메시지에서 학생들이 사용한 어휘들을 빅-데이터 분석 기법 중에 하나인 텍스트 마이닝을 사용해 금지어 사전을 작성하고, 학생별, 학급별 금지어 사용빈도와 친구관계를 분석할 수 있다.

A study of Big-data analysis for relationship between students

Deuk-Young Hwang* · Jin-Mook Kim**

ABSTRACT

Recent, cyber violence is increasing in a school and the severity of the problems encountered day by day. In particular, the severity of the cyber force using the smart phone is recognized as a very high and great problems socially. Cyberbullying have long damage degree and a wide range time duration against of existed physical cyber violence. Then student's affects is very seriously. Therefore, we analyzes the relationship and languages in the classroom for students to use to identify signs of cyber violence that may occur between friends in the class. And we support this information to identified parent, classroom teachers and school sheriff for prevent cyberbullying accidents in the school. For this research, we will design and implement a messenger in the cyber classroom. It have many components that are Big-data vocabulary, analyzer, and communication interface. Our proposed messenger can analyze lingual sign and friendship between students using Big-data analysis method such as text mining. It can analysis relationship by per-student, per-classroom.

Key words : Cyberbullying, School messenger, Big-data analysis, Data mining, Lingual security service

접수일(2015년 0월 0일), 수정일(1차: 2015년 0월 0일),
게재확정일(2015년 6월 0일)

* 강원대학교 삼척캠퍼스 / 컴퓨터공학과

** 선문대학교 / IT교육학부, 교신저자

★ 본 논문은 2014년 강원대학교 학술연구조성비 지원에
의하여 연구되었음(관리번호-220140079).

1. 서론

학습능률 향상, 학부모와 교사, 학생과 교사 사이의 원활한 의사소통을 위해서 학교 내에서 스마트폰이나 태블릿 PC를 사용한 사이버학습의 사용 사례가 증가하고 있다. 하지만 이러한 스마트 장치들은 순기능만을 가지는 것이 아니라, 사이버폭력, 집단 따돌림 등의 역기능으로 인해서 기존의 물리적인 폭력 행위보다 심각한 사회적 문제를 초래하기도 한다.

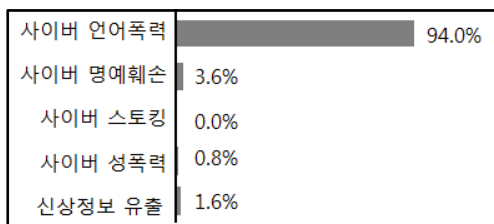
특히 지난 2012년 학교 폭력으로 인해서 대구 중학생 자살 사건이 발생함으로써 정부에서도 “학교 폭력 근절 종합대책”을 발표하고, 현실적인 해결책 마련에 최선을 다하고 있다. 하지만 사이버폭력은 학부모나 교사, 학교 안전 지킴이 등이 보지 않는 공간이나 시간에 발생하는 경우가 많아 근본적인 해결책을 제시하기 어렵다.

그러므로 본 논문에서는 학생들에게 스마트폰이나 태블릿 PC를 사용하지 못하도록 하거나, 카카오톡 등과 같은 메신저를 사용하지 않도록 하는 것만으론 현실적인 해결책으로 부족한 점을 감안하여 학급에서 사용할 수 있는 메신저를 제안하고, 제안한 시스템을 사용하는 학생들 사이에 언어폭력지수와 학생 사이의 친화 정도를 측정해서 사이버폭력을 사전에 방지할 수 있도록 빅-데이터 분석 기술을 적용하고자 한다.

2. 관련 연구

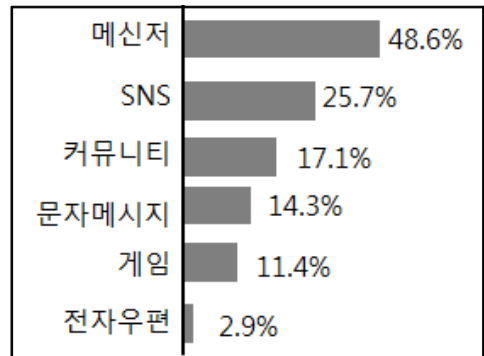
2.1 사이버폭력의 심각성 분석

2013년 한국인터넷진흥원의 초, 중, 고등학교 학생들, 학부모, 교사를 대상으로 실시한 조사결과에 따르면, 사이버폭력 중에서 언어폭력이 94%로 가장 높게 나타났다. (그림 1)에 요약된 내용을 나타내었다.



(그림 1) 주요 사이버폭력 발생 비율

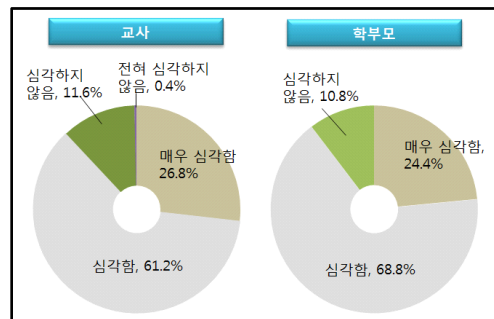
사이버폭력의 수단으로 사용되는 매체로는 메신저(48.6%), SNS(25.7%), 커뮤니티(17.1%), 문자메시지(14.3%), 게임(11.4%), 전자우편(2.9%) 순서로 나타났다. 학생들이 가장 사이버폭력을 많이 경험한 것은 메신저에서라고 응답했다. (그림 2)에 관련 내용을 나타내었다.



(그림 2) 주요 사이버폭력 매체 비율

위와 같이 초, 중, 고등학교 학생들 사이에서 발생하는 사이버 폭력 중에서 언어폭력이 가장 높게 발생하고 있고, 그 중에서도 스마트폰이나 태블릿 PC 등에서 메신저를 사용한 사이버 언어폭력이 가장 심각한 것으로 파악되었다.

이와 관련하여 교사와 학부모들은 사이버 언어폭력의 심각성에 대해서 매우 심각함(교사: 26.8%, 학부모: 24.4%) 또는 심각함(교사: 61.2%, 학부모: 68.8%)으로 인지하고 있었다. 이를 (그림 3)에서 나타내고 있다.



(그림 3) 언어폭력에 대한 심각성 파악정도

앞서 살펴본 것처럼 사이버 언어폭력에 대한 심각성에 대해서도 학부모, 교사, 학생들이 모두 인식하고 있고, 그 방법으로 메시지를 가장 많이 사용하고 있다는 점도 파악하고 있었다.

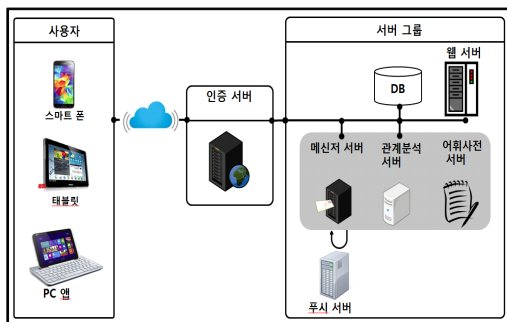
하지만 현실에서 사이버언어폭력에 대한 대응 정도를 조사한 결과에 따르면, 매우 미온적인 대응을 하는 것으로 조사되었다. 가장 두드러지게 조사된 결과로 예로 들어보면, 약 37.8%의 학생들이 사이버 언어폭력으로 인한 피해를 입어도 어떠한 대응도 하지 않는다고 답변했다. 이처럼 미온적인 대응을 하는 이유로는 57.1%의 학생들이 “신고해 봤자 별 소용이 없을 것 같다”라고 응답했다.

이와 같이 학생들은 현실적으로 사이버 폭력, 특히 사이버 집단 따돌림이나 언어폭력에 대해서 이미 문제의 심각성을 파악하고 있고, 해결 방법도 알고 있지만 현실적인 해결책이 없다고 느끼고 있는 실정이다.

3. 제안시스템

3.1 제안시스템 구조

본 논문에서 제안한 시스템은 자체적으로 기존에 초등학교 안전지킴이 연구에서 수행된 시스템의 구조와 동작절차를 따른다[13].



(그림 4) 제안시스템 전체구조

제안시스템에서는 초. 중. 고등학교 학교 내에서 서버 그룹을 구성하고, 사용자가 스마트폰, 태블릿, PC 등을 사용해서 서버 그룹에 접속하여 메시지를 사용하도록 구성하였다.

메신저 서버를 통해서 사용자 간에 메시지를 주고 받는 과정에서 어휘사전 서버가 학생들이 사용하는 낱말들을 수집해 어휘 사전을 작성한다. 작성된 어휘사전을 기준으로 친구들 사이의 관계성을 분석할 수 있고, 언어폭력 지수를 계산할 수 있도록 구성하였다.

3.2 관계성 분석을 위한 자료구조

친구들 사이에서의 관계성 분석을 위해서 (그림 5)와 같이 자료 구조를 설계하였다.

User_ID	USER_ID : VARCHAR(20)
Friend_ID	FRIEND_ID: VARCHAR(20)
Request_UID	TALK_REQ_TIME: DATE
Request_Time	TALK_RES_TIME: DATE
Response_Time	TALK_REQ_CNT: INT(15)
Request_Count	TALK_UNTIL_TIME: CARCHAR(20)
Talk_Time	FRIENDLY_VALUE:INT(15)
Friendly_value	FRIENDLY_VALUE_n :INT(15)
Friendly_Value_n	FEEL_VALUE : INT(15)
Feel_value	FEEL_VALUE_n : INT(15)
Feel-value_n	

(그림 5) 친구 관계 분석을 위한 자료구조

위의 자료 구조를 사용해 학생 A와 학생 B 사이에서 친밀도를 분석하는 수식은 아래 수식1, 수식2와 같다.

- 기본 FRIENDLY_VALUE = 100.....(1)
 - 현재 FRIENDLY_VALUE = 기본 FRIENDLY_VALUE + 대화 시도 요청 수 + (대화 유지 시간 * 0.5)(2)
- (단, 1일 대화 시도 요청 최대값 <= 10,
대화 유지 시간 = 대화 종료 시간 - 대화 요청 시간(60분 단위)임)

친구들 사이의 친밀도 분석뿐만 아니라 메시지를 사용하는 학생의 건강이나 기분 상태를 파악하기 위해서 아래 수식 3, 수식 4를 사용한다.

- 기본 FEEL_VALUE = 0(3)
- 현재 FEEL_VALUE = 기본 FEEL_VALUE + 긍정적 건강 표현 이모티콘 사용 횟수 + 긍정적 기

분 표현 이모티콘 사용 횟수 - 부정적 건강 관련 이모티콘 사용 횟수 - 부정적 기분 표현 이모티콘 사용 횟수 (4)

위와 같이 (수식 1) ~ (수식 4)를 이용하여 학생별로 개인의 대화 시도 횟수, 기분이나 건강 상태를 나타내는 이모티콘 사용 횟수들을 누적하여 일 단위, 주 단위, 월 단위, 학기 단위, 년 단위로 학생 개인의 친밀 지수를 측정한다.

이때 제안시스템의 처리 능력을 감안하여 일 단위로 사이버 메시지를 사용하지 않는 야간에 배치 작업으로 처리할 수 있도록 설계하였다.

3.3 언어 지수 분석을 위한 자료구조

(그림 6)은 본 논문에서 제안하는 학생의 사이버 언어폭력 정도를 측정하기 위한 언어 지수 분석을 위한 자료구조이다.

User_ID	USER_ID : VARCHAR(20)
DIC_NUM	DIC_NUM: INT(20)
Glossary_type	GLOSSARY_TYPE : INT(1)
Count_of_Glossary	COUNT_OF_GLOSSARY : INT(10)
DIC_User_ID	DIC_REQ_ID: CHAR(20)
DIC_Other_ID	DIC_RES_ID: CHAR(20)
DIC_Date	DIC_DATE: DATE
DIC_Time	DIC_TIME: TIME
DIC_Type	DIC_VAR: CHAR(20)
DIC_State	DIC_STATE : INT(1)

(그림 6) 언어 지수 측정을 위한 자료구조

(그림 6)에 나타낸 것과 같이 기본적으로 언어 폭력성을 나타내는 DIC_NUM 값을 100으로 설정한 후, 금지어 사전에 등록된 욕설, 비속어, 은어 등을 사용할 때 마다 금지어의 등급에 따라서 차별적으로 점수를 차감한다. 이를 수식 5, 수식 6에 나타내었다.

• 기본 DIC_NUM = 100(5)

어휘 사전에 등록된 낱말들 중에서 금지어가 아닌 말 = 0, 욕설 = 2, 비속어나 은어 = 1 로 설정한다.

• 현재 DIC_NUM = 기본 DIC_NUM - (CoG * Type 1 + CoG * Type2)(6)

이처럼 욕설이나 비속어의 사용 횟수에 각각 곱셈을 수행하여 기본 언어지수에서 차감한다.

3.4 어휘 사전

2014년 1학기 동안 충남의 xx초등학교 1개 학급을 대상으로 학급 게시판과 메시지를 사용해서 스마트폰과 태블릿으로 입력된 어휘들을 사전에 구분해 어휘사전을 작성하였다. (그림 7)은 욕설을 수집한 결과를 나타내는 예제 화면이고, (그림 8)은 비속어, 은어 등을 수집한 예제 화면이다.

욕설	의미
개간년	개 같은 년
십색기	십새끼
개년	
개새끼	
개색기	
시발	
씨발	
병신	
빙신	
빙떡	
병떡	
십빠바올라	
씨발라머글년	

(그림 7) 어휘사전 예제(욕설)

A	B
개손해	완전 손해
개이득	완전 이득
개짱	아주 재밌다
개짱	진짜 최고
개취	정말 내 취향이다
개소	개인소장
개톡	개인카톡
갤	갤러리
강	그냥
강고	그냥 go
견어올녀	퇴근후에 후줄근하고 머리도 안감고, 맥주에 오징어를 즐기는 여성
겉조	겉어다니는 조각상

(그림 8) 어휘사전 예제(비속어, 은어)

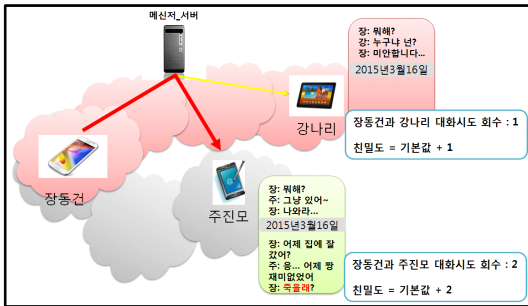
4. 실험 및 분석

본 논문에서 제안한 학급 친구사이의 친밀도 분석 및 언어 지수 분석을 위한 실험을 수행하였다. 이를 위해서 34명으로 구성된 초등학교 학급을 대상으로 1 학기 동안 분석된 결과를 나타내고, 나타난 의미 분석을 실시하였다. 본 연구를 위해서 제안한 시스템은 E-PKI 시스템을 통해서 학부모의 동의하에 개인 프라이버시 침해 문제를 해결하였다.

실험에서 친밀지수와 언어지수를 분석하기 위해서 공개 소프트웨어인 R Studio를 사용하였다.

4.1 친밀도 분석 실험

학급 단위의 친밀도 분석해서 초기 실험 데이터가 없는 상황이므로 데이터베이스에 친밀도 값들을 1주일 간격으로 저장하도록 설정한다. 친밀도값을 1~10 사이를 초기 친밀 구간, 11~50 사이를 상호 친밀 상태, 51 이상을 매우 친함으로 설정하고자 한다. 향후 친밀도 단계는 데이터베이스에 축적된 데이터를 기준으로 세부 실험을 수행해야 할 것으로 생각된다. 초기 친밀도 분석 사례로 3 명의 친구의 대화 내용을 예제를 제시하였다.



(그림 9) 친밀 지수 계산을 위한 예제

위의 실험에서 장동건과 주진모 학생은 친밀도값이 2인 것으로 나타났고, 장동건과 강나리 학생은 친밀도가 1인 것으로 나타나 상대적인 친밀도를 측정할 수 있음을 보였다.

학급 내 친구들 사이에 친한 정도를 분석하는 친밀도 분석을 위해서 R을 사용해 (수식 1) ~ (수식 4)에

나타낸 바를 코딩한 예제를 (그림 10)에 나타내었다.

```

관리자: C:\Program Files\WR Studio\W
> G_FRIENDLY_VALUE <- 100
> C_FRIENDLY_VALUE <- G_FRIENDLY_VALUE + Request_Count + (Talk_Time * 0.5)
> G_FEEL_VALUE <- 0
> C_FEEL_VALUE <- G_FEEL_VALUE + A_Imotion + A_Health - (P_Imotion + P_Health)
> print(C_FRIENDLY_VALUE("장동건 <=> 주진모"))
> 104 : 장동건과 주진모의 친밀지수는 긍정적인 편입니다.
> print(C_FEEL_VALUE("장동건"))
> 12 : 건강지수는 양호한 편입니다.
    
```

(그림 10) R studio를 사용한 실험 사례

위에서 나타낸 실험과 같이 장동건과 주진모 학생 사이에 장동건 학생이 주진모 학생에게 실험 당일을 기준으로 2회의 대화 요청을 하였고, 기존에 친밀지수에 가산이 되어 장동건과 주진모 학생사이의 친밀도는 누적된 점수 104점을 출력한 것으로 보아 두 학생사이의 친밀지수가 긍정적임을 출력하였다.

마찬가지로 장동건 학생의 건강지수를 출력해 보니 기준점수인 0점보다 높은 12점을 출력한 것으로 보아 장동건 학생의 건강이나 기분은 좋은 것으로 파악할 수 있다.

앞서 제시한 실험 결과는 초기 데이터베이스가 없는 상태에서 실험이 수행된 상태이다. 향후 데이터베이스를 보다 축적하고, 이를 바탕으로 학급 단위로 친밀도와 학생의 기분 상태 등을 파악한 것을 기반 데이터로 활용한다면 그 의미성을 찾을 수 있을 것이다.

5. 결 론

본 논문에서는 학생들이 사이버폭력으로부터 안전할 수 있는 학교 내부에서 사용하는 메신저를 설계하고 구현하였다. 제안한 시스템은 학교 내부에서 수업이나 방과 후 활동을 위해서 스마트폰이나 태블릿 PC 등을 사용할 때, 올바른 언어를 사용하는지 빅-데이터 분석 방법의 하나인 텍스트 마이닝 기술을 사용해서 언어폭력 정도를 파악할 수 있는 시스템을 제안하였다. 본 연구에서 제안한 사이버폭력 및 친구 관계성 파악을 위한 시스템은 아쉽지만 일 단위, 월 단위, 학기 단위, 년 단위로 금지어 사전을 수집하고 적용할

수 있다.

제안 시스템은 첫 번째 단계로 사이버언어 폭력을 구분할 수 있는 금지어 사전을 생성한다. 두 번째로 자체적으로 생성한 금지어 사전을 기준으로 학생들이 메신저에서 사용하는 어휘들을 분석하여 금지어(욕설, 비속어, 은어) 등의 사용 정도를 수치화함으로써 학생 개인의 사이버언어 폭력 지수를 구할 수 있다.

세 번째로 학급 내에서 친구들 사이에 사용하는 어휘를 분석해 학급의 분위기와 학생의 개인 건강상태나 기분 등을 파악할 수 있도록 하였다.

네 번째로 학급에서 친구들 사이의 관계성을 분석 지수를 통해서 친한 친구와 그렇지 못한 친구 사이의 관계성을 파악할 수 있도록 하였다. 이를 통해서 학급에서 발생할 수도 있는 사이버언어 폭력의 징후를 사전에 파악하고 학부모, 학급 교사, 학교 안전 지킴이 등에게 해당 학생을 관찰하고, 지도할 수 있도록 하였다.

향후연구로 제안시스템에 대한 사이버폭력 차단 기능과 실시간 어휘 분석의 기능과 온라인 푸시 서비스를 통한 실시간 지도 및 관찰이 가능하도록 추가 연구를 수행하고자 한다.

ACKNOWLEDGMENTS

This study was supported by 2014 Research Grant from Kangwon National University(No. 220140079).

참고문헌

- [1] Jiwon Chung, Jeong-han Kang, "The Size of Intimate Peer Group and Juvenile Delinquency", Korean Journal of Sociology, Vol.46, No.5, pp.177-209, 2012.10.
- [2] Cho-Hee Yoon, Sang-Geun Park, In-Soo Shin, "A Meta-Analysis of the Effects of School Violence Prevention Programs in Korea", Asian Journal of Education, Vol.15, No.1, pp.189-215, 2014.3.
- [3] Yon-Ji Lim, Ha-Young Kim, "A Study of School Bullying Prevention with Professional Advisors", The Korea Journal of Sports Science, Vo.22, No.6, pp.1069-1086, 2013.12.
- [4] Cyber Violence, Report of Korea Internet & Security Agency, 2011.12.
- [5] Effects of Cyber Ethics, Report of Kroea Internet & Security Agency, 2013.12.
- [6] Trend of cyber violence, Kroea Internet & Security Agency, 2013.12.
- [7] Kyo-hyeon Park, Jee-hyong Lee, "Developing a Vulgarity Filtering System for Online Gaems using SVM", Proceeding of Computing Science and Engineering, Vol.33, No.2B, pp.260-263, 2006.10.
- [8] Hae-Kyung Seong, Moon-Goo Lee, "Implementaion of a Realtime Wireless Remote Control and Monitoring System", Journal of The Institute of Electronics Engineers of Korea, Vol.47, CI, No.6, pp.93-102, 2010.11.
- [9] Kyu-Seok Jeong, "Comparison between factors influencing on school violence among elementary school students and those among middle school students", Social science research, Vol.24, No.4, pp.323-338, 2008.12.
- [10] Yong-Bae Lee, "Analysis on Computer Education in Elementary Schools in North Korea and South Korea with Further Prospect", Journal of the Korea Convergence Society, Vol. 5, No. 4, pp. 49-60, 2014.
- [11] Sik-Wan Cho, Won-Jun Jang, Hyung-Woo Lee, "Development of User Oriented Vulnerability Analysis Application on Smart Phone", Journal of the Korea Convergence Society, Vol. 3, No. 2, pp. 7-12, 2012.
- [12] Byung-Seok Yu, Sung-Hyun Yun, "The Design and Implementation of Messenger Authentication Protocol to Prevent Smart Phone Phishing", Journal of the Korea Convergence Society, Vol. 2, No. 4, pp. 9-14, 2011.

[저 자 소 개]



황 득 영 (Deuk-Young Hwang)

1988년 2월 광운대학교
전자계산학과(이학사)
1990년 2월 광운대학교
전자계산학과(공학석사)
1999년 2월 광운대학교
전자계산학과 (공학박사)
1990년 3월 ~ 1994년 2월 : 전주 기
전대학교 전자계산학과 조교수
1994년 3월 ~ 현재 : 강원대학교 삼
척캠퍼스 컴퓨터공학과 교수
관심분야 : 프로그래밍 언어, 컴파일
러, 정보보안, 빅-데이터

email : dyhwang@kangwon.ac.kr



김 진 목 (Jin-Mook Kim)

1998년 2월 배재대학교
전자계산학과(이학사)
2000년 2월 배재대학교
컴퓨터공학과(공학석사)
2006년 2월 광운대학교
컴퓨터공학과(공학박사)
2006년 9월 ~ 2008년 2월 : 선문대
학교 컴퓨터공학과 연구교수
2006년 9월 ~ 현재 : 선문대학교 IT
교육학부 부교수

E-Mail : calf0425@sunmoon.ac.kr