감성대화 말뭉치로 보는 청소년의 문제 도출

김상민*, 이병천O, 우지영*
*순천향대학교 빅데이터공학과,
O순천향대학교 빅데이터공학과
e-mail: kimddol98@naver.com, bung0208@naver.com, jywoo@sch.ac.kr

Identifying issues facing youth through emotional dialogue corpus

Sangmin-Kim*, Byeongchun-LeeO, Jiyoung Woo*
*Dept. of Bigdata Engineering, Soonchunhyang University,
ODept. of Bigdata Engineering, Soonchunhyang University

● 요 약 ●

현대사회에서는 다양한 방법, 통로로 자신들의 의견을 표현하고 또한 감정들을 표출한다. 이렇게 표출된 다양한 문장 및 감정들을 통해 각 연령별로 어떤 문제를 가지고 있는지, 무슨 상황에 놓여있는지 등을 알 수 있다. 본 논문에서는 이렇게 모여진 감성대화 말뭉치를 이용해 청소년들이 문장에서 추출한 단어들과 감정, 상황과 어떠한 연관성을 보이는지 확인해보고자 연구를 진행하였다. 청소년들이 남성의 경우 학교폭력 및 따돌림과 관련한 문제, 여성의 경우 가족관계와 관련한 문제와 연관성이 크다는 것을 확인하였다.

키워드: 청소년(youth), 감성대화(emotional dialogue corpus), 연관성(association)

I. Introduction

현대사회에서 다양한 방법으로 많은 사람들이 각자의 감정을 표현한다. 사람들이 표현하는 내용 중 자신이 무슨 감정을 느끼고 있는지, 관심사는 무엇이고 어떤 상황에 처해있는지 등 다양한 내용이 있다. 이렇게 모여진 것이 감성대화 말뭉치 데이터 이다. 이번 연구의 목적은 여러 개의 문장들에서 추출된 단어들과 각 컬럼별로 구성되어있는 상황 및 감정 등을 이용한다. 추출한 단어들을 이용해 청소년들의 관심사에 대해서 파악하고 상황과 감정에 대해서도 함께 보는 등 각각의 요소들이 청소년과 어떤 연관성이 있는지에 대해서 파악해보고자 한다.

II. The Proposed Scheme

AI hub에서 제공하는 감성대화 말뭉치라는 텍스트 데이터를 사용한다. 이 데이터 셋은 크라우드 소싱 수행으로 일반인 1,500명을 대상으로 하여 음성 15,700문장 및 코퍼스 27만 문장 구축 및 세대별 감성 대화 텍스트 구축을 통해 40,879개의 발화정보로 이루어진 텍스트 데이터이다.

전체 데이터의 성별에 따른 주요 키워드를 살펴보면 그림1과 같다. 남성,여성 공통 관심사는 대인관계, 남성은진로,취업,직장, 여성은 연애,결혼,출산 나타났다.

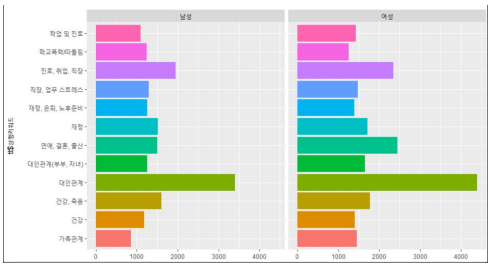


Fig. 1. 성별에 다른 주요 키워드 분포

감정분석 결과는 그림2와 같다. 남성은 불안, 분노 > 슬픔 > 기쁨 순의 감정이 보이고, 여성은 대체적으로 비슷한 분포를 보이지만 불안, 기쁨에 대한 감정이 많이 보이는 것으로 나타났다.

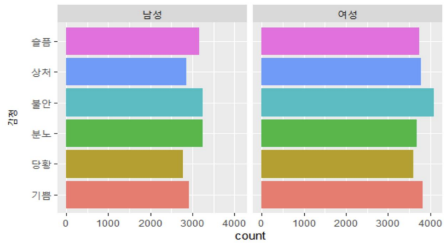


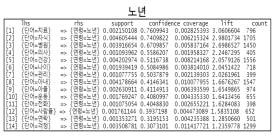
Fig. 2. 성별에 다른 감성 분포(전체 연령대)

Apriori 알고리즘을 이용해 연관분석을 전체 연령대에 수행한 결과는 다음과 같이 나타났다.

청소년과 연관성 높은 단어:
성적, 선생님, 학교, 아빠, 엄마, 공부, 시험 등

청년과 연관성 높은 단어:
연애, 면접, 취업, 결혼, 동료, 상사, 직장 등

중년과 연관성 높은 단어:
노후, 은퇴, 직원, 업무, 아이들, 스트레스, 운동 등



노년과 연관성 높은 단어:
치료, 자식, 병원, 의사, 건강, 관리, 운동, 전화 등

청소년과 연관성이 높은 단어로 성적, 선생님, 아빠, 엄마 등 학교와 가족과 관련된 단어들이 많이 나오는 것을 볼 수 있다.

Table 1. 청소년과 연관성이 높은 단어

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{단어=성적}	=> {연령=청소년}	0.002541773	0.85467757	0.002973956	5.1043508	941
[2]	{단어=선생님}	=> {연령=청소년}	0.003705965	0.85270354	0.004346135	5.0925614	1372
[3]	{단어=학교}	=> {연령=청소년}	0.003546597	0.79431337	0.004464985	4.7438405	1313
[4]	{단어=아빠}	=> {연령=청소년}	0.002895623	0.79348631	0.003649241	4.7389011	1072
[5]	{단어=엄마}	=> {연령=청소년}	0.004567628	0.72762478	0.006277450	4.3455594	1691
[6]	{단어=공부}	=> {연령=청소년}	0.004030102	0.72497570	0.005558947	4.3297384	1492
[7]	{단어=시험}	=> {연령=청소년}	0.001907005	0.69831850	0.002730853	4.1705348	706
[8]	{단어=동생}	=> {연령=청소년}	0.001477524	0.65508982	0.002255452	3.9123622	547
[9]	{단어=언니}	=> {연령=청소년}	0.001139881	0.64427481	0.001769247	3.8477723	422
[10]	{단어=부모님}	=> {연령=청소년}	0.003911251	0.62765496	0.006231531	3.7485143	1448
[11]	{단어=대학}	=> {연령=청소년}	0.001042640	0.52949246	0.001969131	3.1622630	386
[12]	{단어=친구}	=> {연령=청소년}	0.012265879	0.28234782	0.043442441	1.6862527	4541
[13]	{연령=청소년}	=> {단어=친구}	0.012265879	0.07325493	0.167440994	1.6862527	4541
[14]	{단어=가족}	=> {연령=청소년}	0.001812465	0.26324049	0.006885207	1.5721388	671

청소년 하면 친구와 관련한 단어들이 많이 나올 것 으로 예상하여 '친구'라는 단어와 다른 단어들과의 연관성을 본 결과 학교와 관련된 단어들 및 부정적인 감정에 관한 단어들과 연관성을 보였다.

Table 2. '친구' 단어와 다른 단어와의 연관성 분석 결과

{이야기}	=> {친구}	0.015583304	0.3166240	0.049217059	1.371531 619
{연락}	=> {친구}	0.014601480	0.4853556	0.030084084	2.102432 580
{학교}	=> {친구}	0.014198681	0.4437451	0.031997382	1.922186 564
{속상}	=> {친구}	0.011932934	0.2984887	0.039977846	1.292973 474
{고민}	=> {친구}	0.011152510	0.2529983	0.044081365	1.095921 443
{대화}	=> {친구}	0.010900760	0.2674490	0.040758270	1.158518 433
{당황}	=> {친구}	0.010422436	0.2795409	0.037284125	1.210897 414
{선생님}	=> {친구}	0.009339912	0.3354430	0.027843512	1.453050 371

'친구'라는 단어와 감정간의 패턴을 보았을 때 긍정적인 감정과는 연관성이 낮았고, 오히려 부정적인 감정과 연관성이 높고 빈도수 또한 높게 나온 것을 확인했다.

Table 3. '친구' 단어와 감정간의 연관성 분석결과

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{단어=걱정}	=> {감정_대분류=불안}	0.005205098	0.45587887	0.01141772	2.5735398	1927
[2]	{단어=친구}	=> {감정_대분류=당황}	0.008303306	0.19113349	0.04344244	1.1973585	3074
[3]	{감정_대분류=당황}	=> {단어=친구}	0.008303306	0.05201618	0.15962930	1.1973585	3074
[4]	{단어=친구}	=> {감정_대분류=상처}	0.008416754	0.19374495	0.04344244	1.1907678	3116
[5]	{감정_대분류=상처}	=> {단어=친구}	0.008416754	0.05172986	0.16270589	1.1907678	3116
[6]	{단어=친구}	=> {감정_대분류=분노}	0.006877104	0.15830380	0.04344244	0.9671802	2546
[7]	{단어=친구}	=> {감정_대분류=슬픔}	0.007420033	0.17080147	0.04344244	1.0267613	2747
[8]	{단어=친구}	=> {감정_대분류=기쁨}	0.005964118	0.13728782	0.04344244	0.8052134	2208
[9]	{단어=친구}	=> {감정_대분류=불안}	0.006461128	0.14872847	0.04344244	0.8396060	2392

다음으로 청소년과 감정에 대한 패턴을 확인해 보기 위해 연관성 분석을 실행한 결과, 전부 부정적인 감정들과 패턴을 보이는 것을 확인했다.

Table 4. 청소년과 감정간의 연관성 분석 결과

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{감정_대분류=구역질 나는}	=> {연령=청소년}	0.003946366	0.2244929	0.01757902	1.3407286	1461
[2]	{감정_대분류=걱정스러운}	=> {연령=중년}	0.009408072	0.3585916	0.02623618	1.3260843	3483
[3]	{감정_대분류=당황한}	=> {연령=청소년}	0.003214357	0.2192336	0.01466179	1.3093186	1190
[4]	{감정_대분류=부러운}	=> {연령=청소년}	0.003041484	0.2147625	0.01416208	1.2826163	1126
[5]	{감정_대분류=억울한}	=> {연령=청년}	0.006115382	0.3929873	0.01556127	1.2535167	2264
[6]	{감정_대분류=혐오스러운}	=> {연령=청소년}	0.003176541	0.2078473	0.01528305	1.2413167	1176
[7]	{감정_대분류=놀람이 나는}	=> {연령=청소년}	0.003657344	0.2069703	0.01767086	1.2360793	1354
[8]	{감정_대분류=환멸을 느끼는}	=> {연령=노년}	0.004872857	0.3034993	0.01605558	1.2206510	1804
[9]	{감정_대분류=낙의 시선을 의식하는}	=> {연령=청년}	0.006101876	0.3811372	0.01600966	1.2157181	2259
[10]	{감정_대분류=짜증나는}	=> {연령=청소년}	0.003892343	0.2028720	0.01918620	1.2116031	1441
[11]	{감정_대분류=낙담한}	=> {연령=청년}	0.005872279	0.3798043	0.01546133	1.2114667	2174
[12]	{감정_대분류=한심한}	=> {연령=중년}	0.005988428	0.3263173	0.01835155	1.2067330	2217
[13]	{감정_대분류=외로운}	=> {연령=청소년}	0.002693037	0.1998396	0.01347599	1.1934930	997
[14]	{감정_대분류=자신하는}	=> {연령=노년}	0.004632456	0.2963539	0.01563150	1.1919126	1715

마지막으로, 청소년들이 이런 부정적인 감정과 많은 연관성을 보이는데, 어떤 상황과 관련 있었는지에 대해 파악해보고자 한다. 청소년과 상황키워드를 연관성분석을 해 본 결과 남자 청소년의 경우에는 학교폭력/따돌림과 관련한 문제, 여자 청소년의 경우는 가족관계와 관련한 문제와 연관성을 보였다.

Table 5. 청소년과 상황에 대한 연관성 분석 결과

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{연령=청소년,성별=남성}	=> {상황키워드=학교폭력/따돌림}	0.02742198	0.37182727	0.07374924	6.972731	10152
[2]	{연령=청소년,성별=남성,상처질환=해당없음}	=> {상황키워드=학교폭력/따돌림}	0.02742198	0.37182727	0.07374924	6.972731	10152
[3]	{연령=청소년,성별=여성}	=> {상황키워드=가족관계}	0.03339420	0.35642622	0.09369176	6.733033	12363
[4]	{연령=청소년,성별=여성,상처질환=해당없음}	=> {상황키워드=가족관계}	0.03339420	0.35642622	0.09369176	6.733033	12363
[5]	{연령=청소년,감정_대분류=분노}	=> {상황키워드=학업 및 진로}	0.01200657	0.41032032	0.02926145	6.706977	4445
[6]	{연령=청소년,상처질환=해당없음,감정_대분류=분노}	=> {상황키워드=학업 및 진로}	0.01200657	0.41032032	0.02926145	6.706977	4445

III. Conclusions

3.1 분석 결과

본 연구를 진행하기 전, '친구'라는 단어는 당연하게도 '기쁨'과 같은 긍정적인 감정과 연관성이 클 것이라 생각했다. 하지만 연관분석 결과 '친구'라는 단어와 '당황','상처' 같은 감정과 큰 연관성을 보이는 의외의 결과를 찾을 수 있었다. 또한 남성 청소년은 '학교폭력/따돌림', 여성 청소년은 '가족관계'에 연관성이 있는 것으로 보아 감성대화 서비스를 이용하는 청소년층은 부정적인 감정을 가진 상태로 서비스를 이용할 것이라 예상할 수 있었다.

3.2 활용방안

사람이 부정적인 생각을 하게 되면 행동도 부정적일 가능성이 높을 것 이라 생각한다. 그러므로 챗봇이 부정적인 감정 등을 표출하는 청소년 이용자에게 상담기관의 연락처를 알려주어 상담을 권유하거나, 챗봇이 상담역할을 수행하는 시스템을 구축하여 감성대화 시스템 청소년 이용자에게 도움될 수 있도록 분석결과 활용방안을 생각해보았다.

REFERENCES

[1] A Study on Design and Implementation of Personalized Information Recommendation System based on Apriori Algorithm(Yong Kim , Dec. 2012)