

## 신문기사 분석을 통한 이슈 라이프사이클에 관한 연구

### - 삼성 갤럭시노트7 사례 -

#### A Study on the Issue Lifecycle through the Analysis of News Texts

#### - A Case of Samsung Galaxy Note 7 -

허필희\*, 김양석\*\*, 이충권\*\*

(Pil Hee Heo, Yang Sok Kim, Choong Kwon Lee)

#### 요약

시장에 내놓은 제품이나 서비스가 문제를 일으켜서 해당 기업의 비즈니스와 이미지에 큰 타격을 입히는 사례가 자주 발생하고 있다. 발생한 문제에 적절하게 대응하고 피해를 최소화하는 것은 기업들에게 매우 중요한 일이다. 본 연구는 스마트폰 시장을 주도하고 있는 삼성이 개발하여 출시한 갤럭시노트7의 리콜 사태를 다루었던 뉴스를 수집하고 분석하였다. 이슈 라이프사이클에 기반하여 단계별로 뉴스의 특징을 표현하고 연관규칙을 이용하여 텍스트의 내용을 분석하고 시각화하였다. 본 연구의 결과는 이슈의 변화와 흐름을 이해하고 대응책을 모색해야 하는 기업들의 비즈니스 활동에 도움을 줄 것으로 기대된다.

■ 중심어 : 이슈 라이프사이클; 텍스트 분석; 연관규칙; 데이터 시각화

#### Abstract

It is often the case that products or services on the market are causing problems, which hurt the business and image of the company. Responding appropriately to the problem and minimizing the damage is very important to business organizations. This study collected and analyzed the news articles related to the recall of the Galaxy Note 7, which was developed and launched by Samsung Electronics, one of the smartphone market leaders. Based on the issue lifecycle, the characteristics of the news were expressed by stages and the contents of the news were analyzed and visualized using association rules. The results of this study are expected to help business organizations to understand the changes and trends of issues and search for counter measures.

■ keywords : Issue Lifecycle; Text Analysis; Association Rules; Data Visualization

### I. 서론

기업은 시장에 내놓은 제품이나 서비스로 소비자들의 평가를 받는다. 특히 식품, 자동차, 그리고 정보기술 등의 제품이 문제를 일으켜서 대규모 리콜사태를 초래하는 사례들이 지속적으로 증가하고 있다. 문제를 일으킨 제품의 리콜 과정에서 기업들은 다양한 채널을 통하여 소비자들의 인식을 조사하고 대응책을 마련해야 한다. 인터넷을 기반으로 하는 소셜네트워크와 언론미디어는 제품이나 서비스와 관련하여 발생하는 사건과 사고들에 관한 뉴스를 퍼뜨리는데 중요한 매체 역할을 하고 있다. 개인들이 인터넷에 남기는 댓글과 평가는 공신력 있는 언론기관들의 뉴스가 되고, 이렇게 발생한 뉴스는 다시 소셜 미디어를 통하여 개인들 간에 공유된다. 따라서 다수의 고객들을 상대로 제품과

서비스를 판매하는 기업들에게 관련 뉴스를 수집하고 분석하는 것은 중요한 업무가 되었다. 특히, 제품의 사건이나 사고와 관련된 뉴스는 기업의 운명을 결정할 수도 있기 때문에 이슈의 내용을 정확하게 분석하여 소비자 인식의 흐름을 종합적으로 이해하고 대응책을 마련하는 것은 매우 중요하다.

이슈 라이프사이클(issue lifecycle) 모형에 따르면, 사회적 이슈는 다양한 사람들이 의제를 설정하고, 문제에 대한 인식을 형성하면서 발전적 과정을 갖는다고 하였다[1, 2]. 이슈 라이프사이클의 관점에서 중요한 것은 시간의 흐름에 따라 각 단계별로 논의되고 이슈가 되는 내용이 역동적으로 변화하면서 나아가는 것이다. 이러한 변화를 체계적으로 분석하고 종합적으로 이해하는 것은 이슈의 관리와 대응을 위한 방향과 전략의 수립에 있어서 중요하다. 따라서 기업들은 제품과 서비스를 출시하기 전부터 시장에서 성공적으로 정착하기까지 사회에서 발생하

\* 학생회원, 계명대학교 경영대학 경영정보학전공

\*\* 정회원, 계명대학교 경영대학 경영정보학전공

접수일자 : 2018년 09월 02일

수정일자 : 2018년 09월 20일

게재확정일 : 2018년 09월 27일

교신저자 : 이충권 e-mail : cklee@kmu.ac.kr

는 다양한 이슈들을 관리해야 한다.

본 연구에서는 뉴스를 라이프사이클의 단계별로 구별한 후, 각 단계별로 연관규칙 분석을 통해 사회적 이슈를 분석하는 방법을 제시하였다. 제안된 방법을 사용하여 2016년 발생한 삼성 전자 갤럭시노트7의 리콜사태에 적용하여 분석하였다.

## II. 연구 설계

### 1. 자료수집

본 연구는 삼성전자 갤럭시노트7의 기사를 수집, 분석하였다. 기사를 수집하기 위하여 네이버가 제공하는 뉴스 서비스 (<http://news.naver.com>)에서 '갤럭시노트7'을 입력하여 검색되는 일간지와 경제/IT신문에 게재된 기사들을 수집하였다. 검색기간은 갤럭시노트7이 출시된 2016년 8월 1일부터 사건이 종료된 2016년 11월 3일까지로 하였다. 수집한 자료의 제목을 검토하여 갤럭시노트7의 연관성을 확인하였고, 제목만으로 판단이 어려운 경우 본문을 확인하여 연관성이 적은 기사는 제외하였다. 또한 내용이 중복된 기사들도 제외하였다. 표 1에서 볼 수 있듯이, 일간신문 중에서 2,741건을 수집하고, 경제/전자/정보기술 관련 신문기사들 4,488건을 수집하였다.

표 1. 수집된 기사 데이터셋

종합일간지	기사 수	경제/IT	기사 수
조선일보	376	한국경제	812
동아일보	360	헤럴드경제	714
서울신문	346	파이낸셜뉴스	470
중앙일보	283	전자신문	414
국민일보	281	조선비즈	397
한겨레	214	포커스뉴스	341
경향신문	202	케이벤처	331
세계일보	174	ZDNet Korea	235
한국일보	174	한국금융신문	171
매일신문	163	프라이مج계	163
문화일보	70	테크홀릭	106
내일신문	58	초이스경제	100
아시아투데이	40	CCTV뉴스	83
		중소기업신문	73
		ITWorld	38
		CEO스코어데일리	34
		CIO Korea	4
		CIO코리아	2
합계	2,741	합계	4,488

### 2. 이슈 라이프 구간 설정

시간의 흐름에 따라 수집된 기사들을 시각화함으로써 이슈 라이프사이클의 전반적인 흐름을 파악할 수 있다. 본 연구는 Hainsworth와 Meng[3]의 발단, 조정/확장, 조직, 그리고 해결로 이어지는 4단계 모형을 적용하여 갤럭시노트7 사건의 라이프사이클을 다음과 같이 정의하였다. 발단 단계는 삼성전자 갤럭시노트7이 출시되기 전으로 신제품에 대한 기대와 관심이 늘어나고 있는 상황이었다. 조정/확장 단계에서는 첫 판매가 시작되면서 얼마 지나지 않아 갤럭시노트7의 충전 중 폭발이 발생하였고, 삼성전자 측에서 전량 신제품으로 교환해주겠다고 리콜 결정을 발표하였다. 조직 단계에서는 리콜이 이루어졌지만 연이어 계속적인 폭발사건이 발생하면서 기내 사용이 금지되었고, 임직원의 공식 사과와 강제 업그레이드까지 이루어졌다. 하지만 논란은 줄어들지 않고 마지막 해결 단계에서 문제를 해결하지 못해서 제품의 단종을 공식적으로 발표하게 되었다. 단계별로 정확한 기간과 발생한 뉴스기사를 표 2와 같이 정리하였다.

표 2. 라이프사이클 단계별 기사 수

단계	기간	기사 수	비율
발단	8월 1일 - 8월 18일	869	12.02%
조정/확장	8월 19일 - 9월 8일	976	13.50%
조직	9월 9일 - 10월 10일	2,424	33.53%
해결	10월 11일 - 11월 03일	2,960	40.95%
합계		7,229	100.00%

시간의 흐름에 따라 총 7,229건의 뉴스 기사를 수집하였다. 단계가 진행되었고 폭발 사건이 계속 진행되면서 사회적으로 더욱 큰 이슈가 되어가는 것을 확인할 수 있었다. 시간의 흐름에 따라 발생하는 사건들을 시각적으로 표현 하는 타임라인 시각화(timeline visualization)는 의사결정과 행동의 영향을 평가하는데 있어서 유용하다. 그림 1은 타임라인을 2차원 평면 상에 표현하고, 8월 1일부터 11월 3일까지 갤럭시노트7 사건의 전반적인 흐름을 나타낸다.

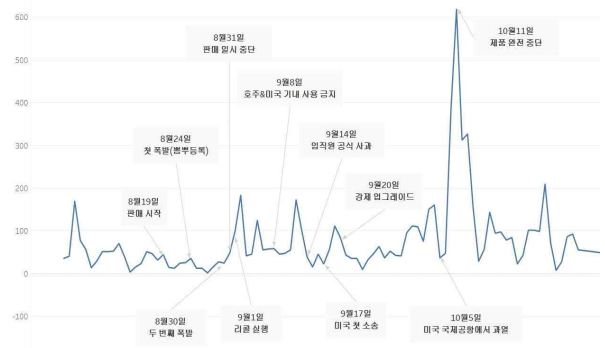


그림 1. 8월1일 ~ 11월15일 타임라인 그래프

### 3. 데이터 전처리

본 연구에서는 뉴스기사로부터 형태소를 추출하기 위하여 한나눔 형태소 분석기를 기반으로 개발한 텍스트 분석시스템을 사용하였다. 데이터 파일을 엑셀로 생성한 후 추출 단어의 유형을 폼사로 지정하여 형태소를 포함하는 산출물을 생성하였다. 산출물의 각 속성은 세미콜론(;)으로 나누었고, 형태소는 쉼표(,)로 구분되었다. 이 산출물을 입력하여 텍스트 분석을 실행하였다. 형태소 분석 단계에서 얻어진 결과물은 텍스트 분석에 필요치 않은 데이터를 포함하고 있다. 따라서, 이를 보정하기 위하여 네 가지 작업을 수행하였다. 첫째, 해당 분야의 특성을 반영하는 용어사전을 활용하여 불용어를 제거하였다. 둘째, 띄어쓰기와 관련된 문제를 해결하기 위하여 용어사전을 사용하였다. 셋째, 동일한 의미를 갖는 단어들이 다르게 표현된 경우에는 하나의 단어로 간주하였다. 넷째, 단어 표현에 명백한 오류가 있다고 판단되는 경우에는 올바른 표현으로 정정하였다.

### 4. 텍스트 분석 설계

본 연구에서는 오픈소스 기반의 데이터분석 프로그램인 레피드드라이너를 사용하였다. 레피드드라이너의 사용자 인터페이스는 분석 프로세스를 구성하여 결과를 얻는 과정이 매우 편리하고 직관적이다. 텍스트 분석을 위한 전처리 과정에서 얻어진 산출물을 입력하여 빈도 분석을 수행하였다. 빈도분석을 수행하기 위해서 레피드드라이너에 설치되는 패키지인 Text Processing을 사용하였는데, <<Read Excel>>, <<Remove Duplicate>>, <<Multiply>>, <<Filter Examples>>, <<Process Document from Data>> 등의 오퍼레이터들을 사용하여 그림 2와 같이 분석을 실행하였하였다.

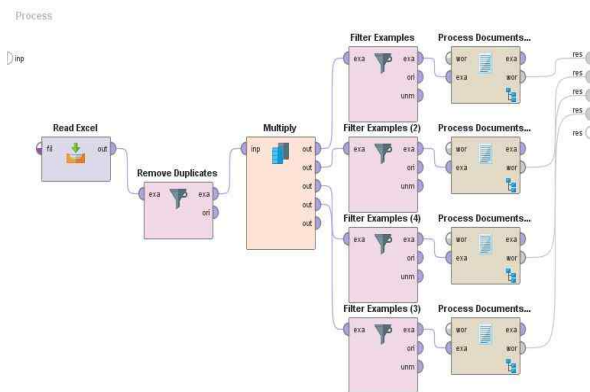


그림 2. 빈도분석 프로세스

빈도분석 과정을 통하여 얻어진 산출물을 활용하여 단어와 단

어들의 관계를 알아보기 위하여 연관분석을 수행하였다. 연관분석을 위하여 <<Read Excel>>, <<Remove Duplicate>>, <<Multiply>>, <<Filter Examples>>, <<Select Attributes>>, <<Subprocess>> 등의 오퍼레이터들을 그림 3과 같이 연결하여 분석을 수행하였다.

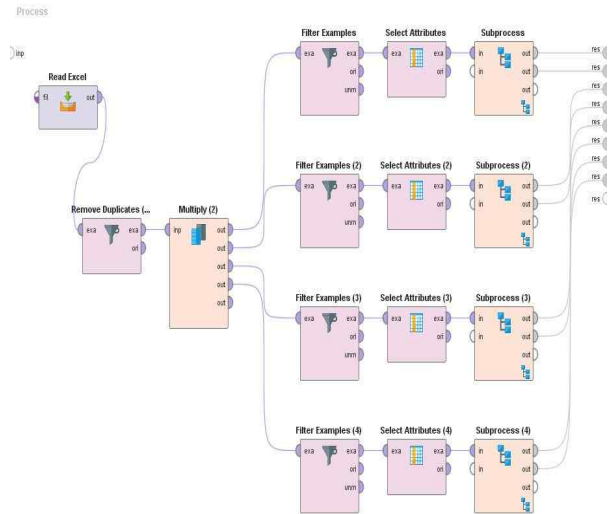


그림 3. 연관분석 프로세스

## III. 연구 결과 및 해석

### 1. 이슈 단계 빈도 분석

표 3은 각 단계별로 가중치 상위를 차지한 단어들을 살펴본 것이다. 1단계 발단은 갤럭시노트7이 판매가 되기 이전에 나타난 상황이고, 제품에 대한 관심과 기대가 높아서 갤럭시노트7의 특성이나 기능에 관한 단어들이 자주 나타나는 것을 볼 수 있다. 즉, 출시(5.7%)와 기능(5.5%), 그리고 휴대인식(3.5%) 등과 같은 단어들이 등장하였다. 2단계 조정/확장은 첫 판매가 시작되고 나서 갤럭시노트7의 폭발사건이 발생하고, 전량 신제품 교환과 리콜을 발표한 상황이었다. 배터리 폭발 및 리콜과 관련된 단어가 등장하였는데, 배터리(5.6%), 제품(4.2%), 리콜(3.7%), 그리고 폭발(2.8%) 등이 나타났다. 3단계 조직은 갤럭시노트7에 대한 리콜이 이루어졌지만 폭발사건이 계속 발생하여 비행기 내에서의 사용을 금지하게 되었다. 삼성전자 임직원의 공식적인 사과와 강제적인 업데이트까지 이루어졌지만 논란은 증폭되었다. 이러한 과정에서 제품의 교환(3.9%)이나 리콜(3.7%)과 관련된 단어들이 자주 등장하였다. 4단계 해결은 폭발 사고의 해결을 시도했지만 연이은 실패로 인하여 제품을 단종하겠다는 발표를 공식적으로 함으로써 실적(2.1%)과 단종(1.7%)과 같은 단어들 출현하였다. 전반적으로 총 단어출현 횟수와 문서

출현 횟수는 일부 단어를 제외하고 유사한 패턴을 보여주었다. 또한 4단계의 경우에는 특별히 많이 출현한 단어가 없는 것으로 나타났다.

표 3. 단어 출현 횟수

단어	총 단어출현 횟수				문서출현 횟수			
	1 단계	2 단계	3 단계	4 단계	1 단계	2 단계	3 단계	4 단계
출시	5.7	3.2	1.6	1.6	4.1	2.6	1.4	1.5
스마트폰	5.2	2.9	2.0	2.0	3.7	2.5	1.8	1.8
제품	2.8	4.2	4.0	7.6	2.3	2.7	2.7	1.9
시장	2.3	1.9	1.9	1.9	1.9	1.6	1.8	1.8
사용	2.2	1.2	1.8	0.7	2.2	1.3	1.5	0.9
고객	2.1	2.1	1.6	1.5	1.6	1.5	1.2	1.2
소비자	2.0	2.1	2.2	1.3	1.9	1.9	2.1	1.2
미국	1.8	1.3	3.1	1.3	2.0	1.5	2.3	1.4
국내	1.5	1.3	1.5	1.1	1.9	1.7	1.6	1.3
판매	1.3	2.0	3.2	1.7	1.5	1.9	2.4	1.8
분기	0.8	0.8	2.1	4.1	0.6	0.7	0.9	1.7
실적	0.8	0.8	1.5	2.1	0.7	0.8	1.1	1.5
배터리	0.7	5.6	2.8	1.1	1.0	2.9	2.3	0.9
아이폰칠	0.6	1.2	1.2	1.6	0.6	0.7	0.8	0.7
문제	0.3	2.4	1.5	1.2	0.4	2.0	1.6	1.2
교환	0.0	2.1	3.9	2.1	0.0	1.4	2.1	1.3
리콜	0.0	3.7	3.7	1.1	0.0	2.4	2.8	1.2
중단	0.0	1.0	2.2	1.0	0.0	1.3	1.6	1.3

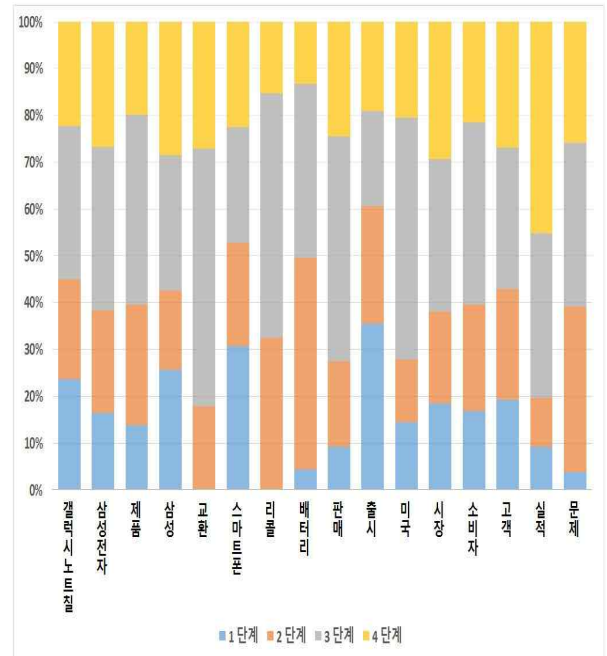


그림 4. 단계 단어별 누적그래프

흥미로운 것은 전체적으로는 수집된 기사는 4단계에서 가장 많았지만, 상위 20개 단어에 있어서는 3단계가 가장 큰 비중을 차지하였다. 4단계에서는 특별히 집중된 이슈가 없었지만, 3단계에서는 교환(리콜) 같은 단어가 부각되어 특정한 이슈로 많이 기사화되었기 때문인 것으로 판단된다. 4단계에서 가장 많은 비중을 차지한 단어는 ‘실적’이었고, ‘삼성’과 ‘시장’이 뒤를 이었다. 즉, 4단계에는 이슈가 전체적으로 안정기에 들어서면서 매출과 같은 단어를 중심으로 기업의 실적에 관한 논의가 이루어진 것으로 볼 수 있다.

## 2. 이슈 단계별 상대빈도

그림 4는 각 단계별 단어 빈도를 비율로 나타낸 것이다. 단계별로 단어들의 출현 비율을 나타내고 있다. 그러나 이슈의 단계별로 출현한 단어들의 비율을 한눈에 파악하기는 어렵다. 이를 한 번에 분석하기 위하여 단어의 상대빈도를 살펴보았다. 전체 기사에 출현한 단어들 중에서 상위에 해당하는 20개 단어에 대하여 각 단어의 단계별로 상대적인 출현빈도를 보여준다.

시각화 분석은 비슷한 데이터의 전반적인 흐름을 빠르고 쉽게 파악할 수 있다는 장점이 있다. 1단계에서는 삼성, 스마트폰, 출시 등이 상대적으로 많은 비중을 차지하였고, 2단계에서는 배터리, 문제, 리콜 등의 출현빈도가 상대적으로 높았다. 3단계에 많이 출현한 단어는 교환, 리콜, 판매, 미국 등이었다.

## 3. 연관규칙 분석

신문기사에서 관련성이 높은 단어들을 파악하기 위하여 동시 발생하는 경우가 많은 단어들을 찾아내는 것이 연관규칙이다. 한무명초 등[4]은 신문기사의 내용을 분석하기 위하여 연관규칙을 활용하였다. 관련성이 높은 단어들을 보여주는 연관규칙 프로세스를 실행하면, 전제(premises)가 되는 단어와 결론(conclusion)이 되는 단어 간의 연관규칙 결과를 보여준다. 연관규칙의 결과가 어느 정도의 설명력을 갖는지를 측정하기 위하여 지지도(support), 신뢰도(confidence), 향상도(lift)를 살펴보았다. 지지도를 기준으로 보았을 때, 결과단어는 삼성전자, 갤럭시노트7이 나타났다. 1단계의 전제단어는 삼성전자, 출시, 스마트폰, 제품 등의 단어가 연관규칙을 보였다. 2단계의 전제단어는 삼성전자, 출시, 배터리, 제품 등의 단어가 연관규칙으로

나타났다. 3단계의 전제단어는 삼성전자, 제품, 리콜 등과 같은 단어들이 연관규칙을 보였고, 4단계의 전제단어는 삼성전자, 스마트폰, 출시, 단종, 사태, 시장 등의 단어가 연관규칙으로 나타났다.

신뢰도를 기준으로 보았을 때 결과단어는 갤럭시노트7이 많이 나타났다. 1단계의 전제단어는 홍채인식, 탑재, 사용, 출시, 기능 등의 단어가 연관규칙을 보였고, 2단계의 전제단어는 배터리, 제품, 출시, 리콜, 폭발, 문제, 소비자 등의 단어들이 연관규칙으로 나타났다. 3단계의 전제단어는 발표, 중단, 결정, 국내, 등의 단어들이 연관규칙을 보였고, 4단계의 전제단어는 사태, 영향, 단종, 중단, 관계자, 분석 등과 같은 단어들이 연관규칙으로 나타났다.

향상도 기준으로 보았을 때 결과단어는 갤럭시노트7, 기능, 스마트폰, 삼성전자 단어가 많이 나타났다. 1단계의 전제단어는 홍채인식, 탑재, 공개 등의 단어가 연관규칙을 보였고, 2단계의 전제단어는 폭발, 문제 등의 단어가 연관규칙으로 나타났다. 3단계의 전제단어는 삼성전자, 교환, 소비자 등의 단어가 연관규칙을 보였고, 4단계의 전제단어는 갤럭시노트7, 스마트폰, 단종, 발표 등의 단어들이 연관규칙으로 나타났다. 세 가지 기준인 지지도, 신뢰도, 그리고 향상도가 모두 비슷한 연관관계 규칙의 단어들로 나타났고, 갤럭시노트7 배터리 폭발 사건의 각 단계와 비슷한 연관규칙이 나타났다.

#### 4. 연관규칙 시각화

연관규칙 분석의 결과를 시각적으로 표현하는 데는 ISOM, KKLLayout, FRLayout, Circle 등의 방법이 있으나, 본 연구에서는 한눈에 파악이 빠르고 밀집이 잘 되어 있는 ISOM을 선택하였다. 연관분석 결과를 어떤 단어들이 연관 규칙이 있는지 알아보기 위하여 ‘갤럭시노트7’ 키워드를 중심으로 시각화하였다.

1단계 발단에서는 갤럭시노트7의 결과단어가 나타났을 때 스마트폰, 홍채인식, 삼성전자, 출시, 탑재, 사용, 기능과 같은 단어들이 연관규칙으로 나타났다. 2단계 조정/확장에서는 삼성전자, 배터리, 폭발, 리콜, 전량, 문제와 같은 단어들이 연관규칙으로 나타났고, 3단계 조직에서는 교환, 리콜, 삼성전자, 제품, 중단, 소비자, 판매의 단어가 연관규칙으로 나타났다. 마지막으로, 4단계 해결에서는 삼성전자, 사태, 제품, 상황, 미국, 아니, 실적, 때문, 단종, 중단 등의 단어들이 연관규칙으로 나타났다. 그림 5, 그림 6, 그림 7, 그리고 그림 8은 갤럭시노트7의 결과단어가 나타났을 때 연관되는 전제단어를 각 단계별로 Association Rule ISOM 그래프로 표현한 것이다.

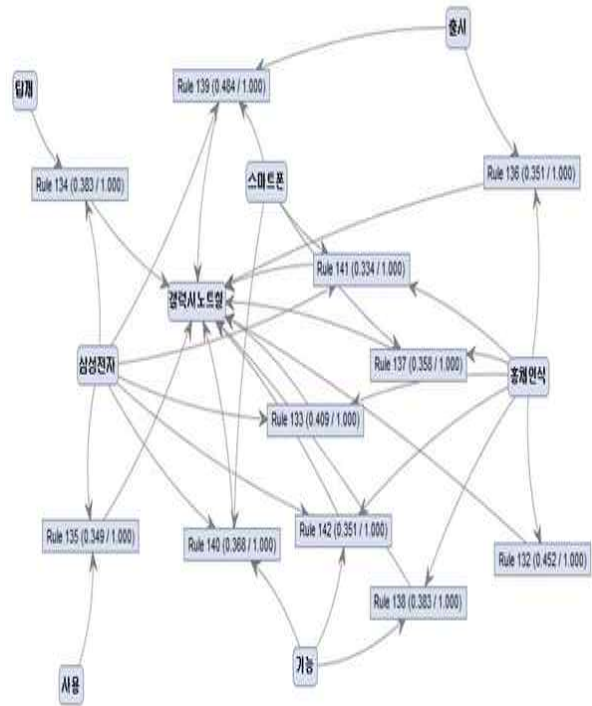


그림 5. 1단계 연관규칙 결과

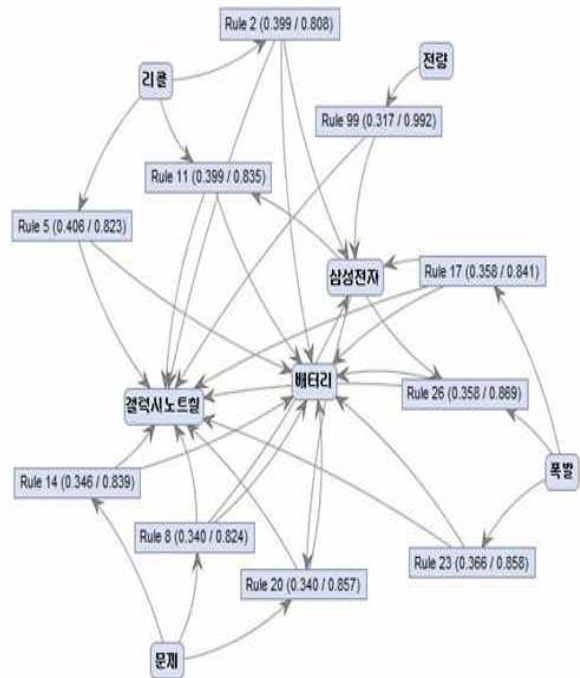


그림 6. 2단계 연관규칙 결과



을 파악하고 인위적으로 정의하였다. 단계들을 분리하고 판별할 수 있는 자동화 알고리즘의 개발을 고려할 필요가 있다. 셋째, 본 연구에서는 형태소 추출을 위해 한나눔 형태소 분석기를 기반으로 사용하였지만, 다양한 제약이 있어서 레피드마이너를 통해 수정을 하였다. 보다 체계적인 분석을 위해서는 형태소 사용의 개선이 필요하고 홍택은 등[5]의 연구처럼 감정을 분류하는 것도 필요하다. 넷째, 본 연구에서는 기본적인 연관분석 기법인 빈발패턴성장 알고리즘을 사용하였다. 다른 기법들을 사용한 결과와 비교하여 분석하거나 연관규칙 생성을 개선할 수 있는 방법에 대하여 보다 깊이있는 연구가 필요하다.

## REFERENCES

- [1] F. W. Geels, C. Penna, "Societal problems and industry reorientation: Elaborating the Dialectic Issue LifeCycle (DILC) model and a case study of car safety in the USA (1900 - 1995)," *Research Policy*, vol. 44, no. 1, pp. 67-82, 2015.
- [2] J. W. Kingdon, J. A. Thurber, *Agendas, Alternatives, and Public Policies*, Boston: Little, Brown. 1984.
- [3] B. Hainsworth and M. Meng, "How corporations define issue management," *Public Relations Review*, vol. 14, no. 4, pp. 18-30, 1988.
- [4] 한무명초, 김양석, 이충권, "텍스트 마이닝 기법을 활용한 동남권 신공항 신문기사 분석," *스마트미디어저널*, 제6권, 제1호, 47-53쪽, 2017년 3월
- [5] 홍택은, 김정인, 신주현, "인스타그램 이미지와 텍스트 분석을 통한 사용자 감정 분류," *스마트미디어저널*, 제5권, 제1호, 1-8쪽, 2016년 3월

## 저자 소개



### 허필희

계명대학교 경영정보학과 학사 졸업.  
계명대학교 경영정보학과 석사 졸업.

<주관심분야 : Big Data, 시각화>



### 김양석

1995년 서울시립대학교 경제학과 학사 졸업.

2004년 University of Tasmania 컴퓨터 공학 석사 졸업

2009년 University of Tasmania 컴퓨터 공학 박사 졸업.

<주관심분야 : Machine Learning and Data Analytics, Recommender Systems, Knowledge Engineering>



### 이충권

1995년 계명대학교 경영정보학과 학사 졸업.

1999년 Southeast Missouri State University MBA 졸업

2003년 University of Nebraska-Lincoln 경영학박사 졸업.

2003-2006년 Georgia Southern University 조교수

<주관심분야 : Big Data, Text Mining, IT Jobs>