

# 지자체 사이버 공간 안전을 위한 금융사기 탐지 텍스트 마이닝 방법\*

최석재

경희대학교 빅데이터 연구센터  
(sjchoi@khu.ac.kr)

이중원

경희대학교 경영학과  
(kanakan@khu.ac.kr)

권오병

경희대학교 경영학과  
(obkwon@khu.ac.kr)

최근 SNS는 개인의 의사소통뿐 아니라 마케팅의 중요한 채널로도 자리매김하고 있다. 그러나 사이버 범죄 역시 정보와 통신 기술의 발달에 따라 진화하여 불법 광고가 SNS에 다량으로 배포되고 있다. 그 결과 개인정보를 빼앗기거나 금전적인 손해가 빈번하게 일어난다. 본 연구에서는 SNS로 전달되는 홍보글인 비정형 데이터를 분석하여 어떤 글이 금융사기(예: 불법 대부업 및 불법 방문판매)와 관련된 글인지를 분석하는 방법론을 제안하였다. 불법 홍보글 학습 데이터를 만드는 과정과, 데이터의 특성을 고려하여 입력 데이터를 구성하는 방안, 그리고 판별 알고리즘의 선택과 추출할 정보 대상의 선정 등이 프레임워크의 주요 구성 요소이다. 본 연구의 방법은 실제로 모 지방자치단체의 금융사기 방지 프로그램의 파일럿 테스트에 활용되었으며, 실제 데이터를 가지고 분석한 결과 금융사기 글을 판정하는 정확도가 사람들에 의하여 판정하는 것이나 키워드 추출법(Term Frequency), MLE 등에 비하여 월등함을 검증하였다.

**주제어** : SVM, 금융사기, 사이버 범죄, 위기관리, 텍스트마이닝

논문접수일 : 2017년 7월 12일    논문수정일 : 2017년 9월 17일    게재확정일 : 2017년 9월 20일

원고유형 : 일반논문    교신저자 : 권오병

## 1. Introduction

SNS는 개인간 의사소통뿐 아니라 마케팅의 중요한 채널로도 자리매김하고 있다(Michaelidou et al., 2011). 그러나 사이버 범죄 역시 SNS의 확산에 따라 진화하고 있으며(Mazurczyk et al., 2016), 그 악영향 중 하나로 불법 광고(illegal ads)로 SNS에 다량으로 배포되어 개인정보를 불법 수집하거나 심지어 금전적인 해악을 끼치는 등, 금융사기는 오늘날 경제를 위협하는 매우 큰 위험 요소로 받아들여지고 있다(Jans et al.,

2010). 특히 불법 광고는 불법 대출 광고나 불법 방문판매 등 여러 가지 유형이 있으며, 그 내용이나 형식이 수시로 변경되고 있어 실시간으로 패턴을 분석하여 대응할 필요성을 가지게 한다.

불법 광고는 메시지의 패턴이 있으므로 자동화된 분석이 가능하다(Balamurugan et al., 2007). 따라서 합법적인 광고와 불법적인 광고를 판별하는 규칙을 지속적으로 찾고, 이를 시스템에 반영해야 함에도 불구하고, 아직까지는 기존의 불법 광고 사례를 알고 있는 경험에만 의존한다는 것, 피해자 설문 과정을 거쳐 뒤늦게 불법 광고

\* 이 논문은 2017년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066740).

판정 규칙을 수정하고 있다. 그러나 불법 광고는 그 내용이나 형식이 수시로 변경되고 있으므로 일반적이고 고정적인 규칙이 아니라, 실시간으로 패턴을 분석하여 대응할 필요성이 있다. 불법 광고는 업데이트된 작성 방법으로 규칙을 피하기 때문에 효과적으로 필터링이 되지 않고, 이로 인해 불법 광고의 패턴을 발견 및 대응하지 못한 많은 소비자들이 상당 기간 위협에 노출되는 한계가 있었다.

이러한 이유로 SNS에서도 위기관리(emergency management)가 필요하다. 위기 관리는 본래 홍수나 화재와 같은 자연 재해에 대응하기 위한 것(Waugh and Streib, 2006)이나, 이제는 자연재해와 같이 물리적 공간에서 발생하는 것은 물론, 네트워크 상에서 사람에 의해 일어나는 재해도 포함되어야 한다. 사이버 범죄는 컴퓨터를 이용하여 전 세계적으로 일어나기 때문에 피해 규모가 수십억 달러에 이르러(Hassan et al., 2012) 자연재해에 못지 않기 때문이다.

네트워크 상에서 일어나는 위기 관리는 위기의 패턴을 찾는 준비 단계, 네트워크 상에서 문제점을 찾는 대응 단계, 기본 서비스 제공 및 구호 활동의 복구를 포함하는 복원 단계의 세 단계로 나눌 수 있다(Waugh and Streib, 2006). 이 세 가지 활동을 가능하게 하는 본질적인 것 중 하나가 예측 분석이다. 즉 과거 또는 현재 실시간으로 나타나는 SNS상의 징후에 의거하여 사이버 범죄가 발생하고 있는지, 또는 발생 가능성이 높아지는지를 파악할 수 있어야 한다. 그러나 SNS의 매체는 다양하고, 데이터의 규모는 방대하기 때문에 이를 인력에 의하여 수동적으로 하는 것은 매우 많은 비용이 필요하거나 불가능하여 개별적인 신고에 의하여 대응하는 방식에 머물러 있다(Kumari et al., 2015).

따라서 본 연구의 목적은 SNS 범죄에 대해 자동적인 예측 분석이 가능한 방법을 제안하는 것이다. 이를 위해 SNS상의 비정형 데이터에 대한 텍스트 분석 및 분류 알고리즘 기법을 활용하여 사이버 범죄의 징후를 인식하고자 한다.

본 논문은 다음과 같이 구성되었다. 먼저 SNS 상에서의 위기관리와 관련된 기존 문헌을 분석한 것을 제2장에 보였다. 그리고 본 연구에서 제안하는 금융 사기글 판정 방법을 제3장에 기술하였으며, 실제 모 지자체의 사례를 바탕으로 성능 검증을 위한 실험 설계와 그 결과를 제4장에 보였다. 마지막으로 본 연구의 학술적, 실무적 시사점과 추후 연구방향, 그리고 결론을 각각 제5장과 제6장에 나타내었다.

## 2. SNS 위기관리

제2차 세계대전 때부터 본격적으로 연구되기 시작한 위기 관리는 원래는 재무 금융 보험에서의 이슈였으나(Dionne, 2013), 최근 위기 관리는 범죄 사고 혹은 재해 등의 다양한 이슈에 관련되어 있다. 그것이 자연재해이든, 기술, 에너지 또는 문명과 관련된 것이든 어떤 종류의 위협에 대하여도 인지하고 대응할 수 있어야 한다(McEntire, 2004). 특히 정부는 인간의 삶을 위협할 수 있는 자연재해는 물론 사람이 만든 재해에 대하여 위협도를 줄이고자 노력해 왔다(Petak, 1985). 정부가 빠르게 변화하는 위기에 대응할 수 있는 지식 관리 시스템을 갖추고 있으면 구성원들의 안전을 보장할 수 있으므로(Yates and Paquette, 2011), 정부는 사회 구성원이 정보를 기반으로 위협을 인지할 수 있는 방안을 구축할 필요가 있다(Comfort, 2007). 성공적인 위기 관리는

기업의 성과를 증진시키며(Grace et al., 2015; DeAngelo and Stulz., 2015), 재앙과 손실을 회피하거나 최소화해 주는 등 기업 경영을 보호하는 매우 강력한 수단이다(Sadgrove, 2016). 예를 들어 적극적으로 기후에 영향을 받는 비즈니스를 수행하는 기업은 실제로 기후 예측 및 대응 관리가 그 기업의 성과에 긍정적인 영향을 주는 것으로 실증분석 결과가 나타나고 있다(Pérez-González and Yun, 2013).

기업의 온라인 활동이 활발해 짐에 따라 위기 관리의 영역은 오프라인뿐 아니라 온라인에도 포함되고 있다(Nykodym et al., 2005; Banerjee et al., 2008). 즉, SNS 같은 사이버 공간에서 발생하는 범죄도 위기관리 대상에 속한다. 최근에는 물리적 공간 외에도 사이버 공간이 많은 범죄가 발생한다는 점에서 관심을 받고 있다. 특히 SNS와 같은 최신 미디어 기술이 사이버 스토킹(cyber-stalking), 사이버 폭력(cyber bully), 아이디 도용(identity theft), 바이러스 유포(virus dissemination), 네트워크 침입(network intrusions), 피싱(phishing), 괴롭힘(harassment), 협박(threatening), 금융 사기(financial fraud)와 같은 사이버 범죄에 악용되고 있다는 점에 유념할 필요가 있다.

이러한 위기를 관리하기 위한 수단으로 최근 BI(business intelligence)나 빅데이터와 같은 지능형 방법들이 고려되고 있다. 먼저 SNS 같은 사이버 공간에서의 위기관리에 유용한 기술적 수단으로는 머신러닝 기법을 이용한 빅데이터 분석법이 많이 사용되고 있다. 예를 들어 Naïve Bayesian Model을 사용하여 정크 메일(junk mail) 여부를 판정한 연구(Sahami et al., 1998)에서는 규칙기반시스템(rule-based system)으로 규칙을 구성하는 것의 물리적인 한계를 지적하며 자동

화된 방법으로 필요한 문서를 분류할 필요성을 제기하였다. 예를 들어 이메일 상의 사이버 범죄 추적을 위해 코퍼스와 Gaussian Mixture Model을 활용한 연구도 있다(Sreenivasulu and Prasat, 2015).

또한 business intelligence 방법도 위기 관리에 사용되고 있는데 조기 경보 시스템 구축을 위해서 확률론을 활용하거나(Flores, 2009), 거시경제 모델을 활용하여 재무 문제를 해결하려는 연구도 있었다(Castell and Dacuycuy, 2009). 또한 리스크 분석에 게임 이론이 활용되기도 한다. 협동적 게임이나 비협동적 게임 모형 등 다양한 게임 이론에서의 모형을 활용하여 프로젝트 관리에서의 재무적 위기를 분석하거나(Zhao and Jiang, 2009) 위험이 적은 최적 온라인 광고 전략을 파악하는 것(Lin et al., 2012)이 가능하다. 또한 데이터 마이닝 기법은 비즈니스에서의 사기 이벤트를 파악하거나(Jans et al., 2010), 거짓 감사를 찾아내거나(Holton, 2009), 자연 위기를 관리하는 데에도 활용된 바 있다(Folino et al., 2010).

그러나 Jan 등(2010)의 연구에서 나타나는 바와 같이 사기 탐지를 위해서 사전에 준비된 도메인 지식을 사용하는 방법을 사용해왔기 때문에 기존 연구로는 사전에 지식을 확보하기 어려운 상황에서는 활용할 수 없다는 한계점이 있다. 또한 이들 연구는 기업체에서 관리하는 제한된 데이터로부터의 사기 방지 문제이고, 본 연구에서 대상으로 하는 SNS의 경우에는 매체가 다양하여 이를 인력으로 수동적으로 하는 것은 매우 많은 비용이 필요하다(Kumari et al., 2015). 따라서 본 연구에서는 오픈 네트워크 환경에서 수집되는 SNS 메시지를 머신러닝 기법을 이용하여 예측 분석하는 방안을 제안하고자 한다.

### 3. 방법론

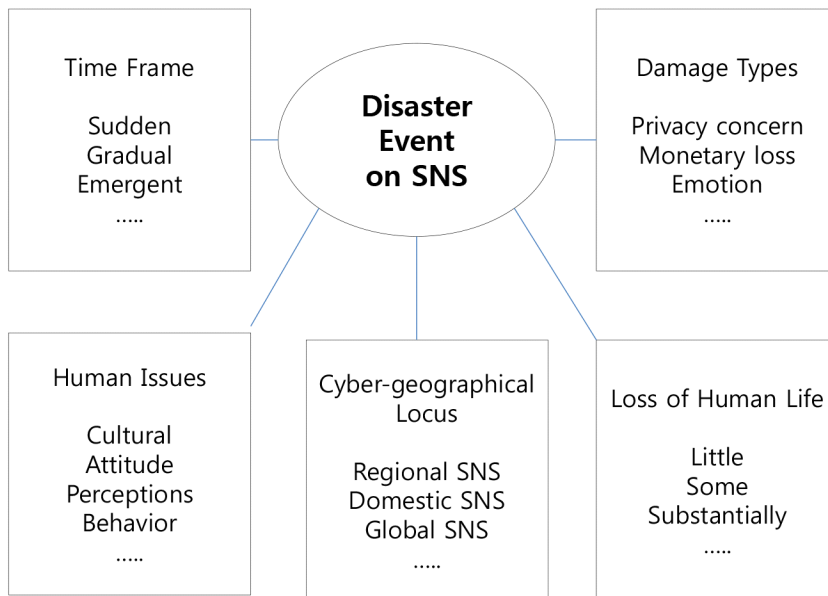
#### 3.1 전체적인 위기관리 프로세스

위기관리에 대한 프레임워크로는 Henderson (2004)의 연구가 대표적이다. 위기관리는 SNS 영역에서도 필요하므로 Henderson(2004)의 프레임워크를 기준으로 SNS 관리에 있어서의 위기 관리 이슈를 <Figure 1>과 같이 제안한다. 위기 관리는 시간, 손해의 종류, 인간 행동, 문화, 사이버 공간에서의 위치라는 다섯 개의 영역으로 나누어 관찰되는데, 먼저 시간의 측면에서 보면 급작스러운 위기 발생, 점진적인 발생, 그리고 이전에는 관찰된 바 없는 새로운 종류의 위기발생이라는 세 종류로 나뉘어진다. 또한 손해의 종류 측면에서 보면 불법대출, 불법(방문)판매, 사생

활 정보의 노출, 금전적 손해, 그리고 감정적 손해로 나뉘어진다. 인간의 행동주의적 관점으로는 인지, 태도, 행동 그리고 문화적인 이슈가 존재한다. 다음으로 사이버공간의 위치로 보면 지역적 SNS(기업 등 특정 사용자집단에만 허용)와 국내적 SNS, 전세계적 SNS(다국어 지원 등)로 나뉠 수 있다. 마지막으로 인간 생활적인 손실 관점에서 미약, 상당, 심각으로 나뉘어질 수 있을 것이다.

<Figure 1>에서 나타난 위기의 측면적 분류에 근거하여 <Table 1>과 같이 사이버 범죄를 위한 관리 프로세스를 제안한다.

먼저 사이버범죄 발생 전 관리는 위기를 인식하는 단계로부터 시작한다. 이는 불법 광고로 의심되는 글의 게재를 모니터링 하거나 그러한 글이 게재되는 사전 징후를 파악하는 것 등이 포함



<Figure 1> Matrix of conceptual characteristics of cybercriminality on SNS and emergency management

(Table 1) Emergency management process

Phase		Description
Pre-Cybercriminality	Risk identification	Hazard (e.g. suspected illegal ads posted) monitoring and forecasting
	Mitigation	Awareness about risks and prevention
	Risk transfer	Insurance, Reinsurance, Safety regulation
	Preparedness	Early warning systems
Post-Cybercriminality	Emergency Response	Damage assessment, Hazard alarm system, Accident announcement
	Recovery	Incorporation of cybercriminality mitigation components

된다. 다음으로 완화는 예상되는 사이버 범죄의 많은 영향을 최소화하거나 줄이기 위해 기관의 사이버 범죄 관리 자원을 준비하고 사전 배치하는 결정을 포함하게 된다. 위기 전가는 위기의 피해를 최소화하기 위한 사전 단계로서 보험 가입이나 재가입, 안전을 위한 각종 사이버 공간 상의 또는 정보통신 기술적 규제가 해당된다. 준비(preparedness)에는 사이버 범죄에 대응할 수 있는 정부 기관, 비정부 기관 및 국제기구의 다양한 유형, 수준 및 기술을 포함하는 조직 대응 시스템을 구축하려는 노력이 포함된다. 대비책에는 사이버 범죄를 예방하고 대응할 수 있는 가장 적절한 방법과 수단에 대한 공적 알람 및 경고가 포함된다.

다음으로 사후 복구 단계에는 긴급 상황에 대한 대응 및 복구가 포함된다. 긴급 상황에 대한 대응을 위해서는 먼저 피해 규모에 대한 평가와 위기 발생 경보, 그리고 사건 사고 발생 사실을 알리는 알람이 필요하다. 그리고 복구 작업이 이어지는데 이는 사이버범죄를 완화하기 위한 여러 방안들을 적용하는 것이다. 예를 들어 인간, 금융 및 IT 인프라의 재구성, 주문 및 정보의 흐름 복원, SNS 사이트의 커뮤니티 재구축 단계가 포함된다(e.g. 사이버 범죄를 일으킨 개인이나 기

관에 대한 SNS 커뮤니티 접근 또는 계정 사용 금지).

본 연구에서는 <Figure 1>에서 나타난 종류 중에서 국내 사이버공간을 염두에 두고 지자체 주민들에게 상당한 금전적 손해를 가할 수 있는 급작스러운 불법대출 및 불법(방문)판매 글의 게재 행동을 실시간 탐지하고 대응하여 SNS 위기관리를 하는 분야에 집중하고자 한다. 그리고 위기관리 프로세스 관점에서는 사전적 사이버범죄 발생 전 관리의 첫 단계에 해당하는 위기의 인식 단계에 집중한다.

### 3.2 위험 인식

위기관리의 첫 단계는 불법 판매 등의 리스크를 인식하는 것이다. 불법 판매 리스크 인식은 어떠한 메시지가 출현하였을 경우, 그 메시지가 불법 판매 홍보글인지를 판정하는 데서부터 시작한다. 불법 판매 홍보글을 본 사람은 홍보 내용이 정상적인 것으로 오해하고 구매 및 결제 과정을 진행하기 때문이다. 다행히 한글로 작성된 비정형 문서로부터 의미 있는 정보를 추출하는 방법이 최근 개발되고 있으므로 이들을 위험 인식 문제에도 활용할 수 있다 (Cui et al., 2016;

Choi et al., 2015; Choi et al., 2016; Kim and Kwon, 2016). 따라서 여기에서는 한글로 작성된 메시지가 불법인지 여부를 판정하는 절차를 다루도록 한다.

SNS와 같은 대규모 데이터에서 리스크 인식을 자동으로 수행하기 위해서는 빅데이터 기법에서 사용되는 머신러닝 판별 알고리즘을 적용할 필요가 있다. 빅데이터 머신러닝 알고리즘은 크게 예측분석 기법과 판별분석 기법으로 나뉘는데 판별분석 알고리즘은 데이터를 둘 혹은 그 이상의 범주로 나눈다. 본 연구에서도 어떤 메시지가 불법 홍보글인지 여부를 판정하는 것으로 판별 알고리즘을 적용하면 성과를 기대할 수 있다. 특히, 여러 판별 알고리즘 중 SVM에 초점을 맞추고 진행하였는데, 그 이유는 SVM은 여러 일반적인 분류 상황은 물론, 텍스트 분석에서도 효과적이라고 알려져 왔기 때문이다(Joachims, 1997). 따라서 본 연구에서는 SVM을 불법 광고 한글 텍스트 분석에 적용하는 방안을 맞추어 진행하도록 하되, 다른 알고리즘도 같이 사용하여 그 결과를 비교해 보기로 하겠다.

SNS 데이터와 같은 비정형데이터로부터 특정 판별 알고리즘 모델을 생성하기 위한 학습용 데이터셋을 구축하는 과정은 다음과 같다. 먼저 SNS로부터 수집된 메시지 및 연관 데이터를 수집하여 메시지를 전처리 한다. 전처리 과정은 다시 형태소 분석 단계, 불용어 제거 단계, 유효 품사 선택 단계로 나뉘는데 형태소 분석 단계에서는 기존에 만들어진 한글 형태소 분석기를 사용하게 된다. 텍스트를 형태소 분석하게 되면, 문장이 단어 또는 단어보다 작은 단위로 잘게 나뉘어진다. 그 결과 복잡한 모양으로 된 문장이 정형 데이터와 유사한 형태소 단위로 바뀌어 기계적 분석이 가능하게 된다. 형태소 분석기의 선택

에 있어서 본 연구에서는 본 연구팀이 직접 개발한 RHINO 3.0을 이용하였다. RHINO 3.0 형태소 분석기는 초당 4만 어절의 분석 속도로 동음이의어를 구분해가며 한글 텍스트 데이터의 형태소 구분 및 품사 판정을 수행한다.

불용어 제거 단계에서는 숫자, 문장부호, 이중공백 등 비 어휘적 요소들을 텍스트에서 제거한다. 숫자, 문장부호, 이중공백 등은 어휘적인 의미를 거의 갖지 못할 뿐만 아니라, 일반적인 글에서도 많이 사용되므로 이후의 과정이 원활히 진행되도록 우선적으로 제거하는 것이 바람직하다. 숫자를 제거하는 이유는, 숫자는 통계 분석 등 정형 데이터 분석에서는 매우 중요한 정보가 되어주지만, 텍스트를 이용하여 합법과 불법을 판정하는 문제에서는 큰 도움이 되지 않는다. 오히려 숫자는 합법 메시지는 물론, 불법 메시지에서 모두 자주 사용된다는 점 때문에 불용어로 간주되어야 한다. 이와 같이 특성 선정(feature selection)을 통해 가급적 차원을 축소하는 것은 분석 모델을 만들기 위한 계산량을 줄여주는 역할도 한다.

전처리 세 번째 과정인 유효 품사 선택 단계에서는 ‘대출, 신청, 할부, 조건’ 등과 같은 명사류와, ‘♥, ☆, ♪, ♫’와 같은 기호류 두 가지만을 남긴다. 명사류는 사물을 언급하는 기능을 가지는 것으로서 메시지가 궁극적으로 말하고자 하는 바가 다른 품사들에 비해 가장 잘 표현된다. ‘대출, 신청, 할부, 조건’ 등 광고에 사용된 대출과 관련된 어휘는 이 글이 illegal한 것인지는 밝혀줄 수 없다고 하더라도, 적어도 해당 메시지가 우리가 찾으려고 하는 불법 대출과 관련된 영역에 속하는 글임을 알게 해준다. 감성 분석을 위한 비정형 텍스트 분석에서는 동사류도 많이 분석되지만(Bayer, 2010), 본 연구와 같이 도메인

구분 및 불법 여부를 판정하는 분류 상황에서는 품사의 기여도가 낮다. ‘싸다, 내려가다, 저렴하다’ 등 많은 동사들이 다양한 상황에서 사용되기 때문이다.

기호류는 정상적인 텍스트를 분석할 때는 어휘적인 의미가 크지 않아 불용어의 한 종류로 포함시켜 제거하지만, 같은 광고 메시지더라도 불법적인 텍스트일수록 기호류의 사용이 많다는 연구(Sahami et al., 1998)를 고려하면 불법 광고를 분류하는 본 연구와 같은 경우에는 오히려 판정의 좋은 keyword가 되므로 남겨둘 필요가 있다. 기호류가 불법 광고에서 더 많이 사용되는 이유는 무엇보다 메시지를 더 눈에 잘 띄게 해주기 때문이다. 예를 들어, ‘당일 조회, 당일 대출’ 보다는 ‘♥♥당일 조회, 당일 대출♥♥’이 시각적으로 더 두드러진다. 합법적인 광고 메시지에서 기호류를 사용하기는 하지만, 불법 광고는 보다 높은 수준으로 사용하는 경향이 있다. 위와 같이 판정에 유효한 품사만을 선택한 결과로 이후의 단계에서 분석 대상으로 삼아야 할 데이터의 범위가 크게 줄어든다.

둘째, 선별된 데이터에 대하여 ‘합법legal’ 또는 ‘불법illegal’을 부여하는 단계이다. 전처리 과정을 통해 선별된 데이터를 학습 데이터로 만들려면 각 데이터가 합법인지 불법인지를 미리 분류해 놓아야 한다. 합법글은 우선 대출 및 대부업과 전혀 관련 없는 글이 이에 속한다. 일부 ‘대출, 조회, 사채’와 같은 단어가 들어갔다고 하더라도 개인적인 글에서 사용된 경우일 수 있다. 이러한 메시지는 ‘합법legal’으로 분류해 놓는다.

또한 대출 및 대부업과 관련된 글이더라도 합법적으로 허가를 받고 대출 및 대부업을 하는 경우의 글이 있다. 이러한 경우는 메시지만 보고서는 불법 여부를 판정하기 어렵다. 사용된 명사들

도 거의 비슷한 데다가, 사람의 주목을 끌기 위하여 합법 홍보글도 기호류를 사용하기 때문이다. 따라서 대출 및 대부업과 관련된 글이라고 생각되면, 그 메시지에 있는 전화번호 혹은 웹사이트 정보를 합법 업체 조회 사이트에 입력하여 합법 및 불법 여부를 직접 확인해 봐야 한다. 직접 정보를 입력하여 합법 및 불법 홍보글 여부를 확인하는 작업은 적지 않은 시간을 필요로 하지만, 이 과정을 통하여 사람은 빠른 시간 안에 구분하기 어려운 합법글과 불법글의 문자열 패턴을 기계가 인식하여 모델이 만들어지게 된다.

셋째, 가공된 데이터를 Corpus type으로 변환한다. 통계분석 및 기계학습 패키지인 R은 텍스트 데이터를 Corpus로 만들 수 있는 함수를 제공하고 있는데, 텍스트 데이터를 Corpus로 만들면 텍스트 마이닝에 필요한 다양한 도구를 활용하기가 쉬워진다.

넷째, 데이터를 Document-Term Matrix 형태로 만들고, 전체 문서에 대하여 각 단어의 TF-IDF 값을 구한다. TF-IDF 값은 각 데이터 사용된 단어의 발생 비율과 전체 데이터 집합에서의 발생 분포의 역수를 구하는 것으로서, 전체 문서의 관점에서 현재 문서의 특징을 단어 별로 알게 해주므로 데이터의 분류에 효과적으로 쓰일 수 있다. 데이터에 따른 단어 별 TF-IDF 값은 <Figure 2>와 같은 형태로 제시된다.

가로축은 각 문서(Document)에 해당하고, 세로축은 분석 대상 전체 문서 집합에서 발견된 어휘의 목록이다. 각각의 어휘가 각 문서에 어느 정도의 분포로 나타났는지가 TF-IDF 값으로 표현되었다. 본 연구에서는 열(column)에 들어가는 어휘를 분석 대상 문서에서 빈도 20위 이상의 것만으로 하였다. Document-Term Matrix에서 각각의 어휘는 차원을 형성하여 판정 기준이 되어주

id	write	consultant	name	time	apply	form	where	condition
(Original Words)	(기재)	(상담가)	(성명)	(시간)	(신청)	(양식)	(어디)	(조건)
1	0.182135	0.170809	0.16981	0.338039	0.166496	0.169612	0.169612	0.376247
2	0	0.19521	0.194069	0.38633	0.190281	0.193842	0.193842	0
3	0.194277	0.182196	0.181131	0.180287	0.180919	0.180919	0.180919	0
4	0.194277	0.182196	0.181131	0.360575	0.180919	0.180919	0.180919	0.40133
5	0.194277	0.182196	0.181131	0.180287	0.180919	0.180919	0.180919	0
6	0.194277	0.182196	0.181131	0.180287	0.180919	0.180919	0.180919	0
7	0.182135	0.170809	0.16981	0.338039	0.169612	0.169612	0.169612	0
8	0.194277	0.182196	0.181131	0.360575	0.180919	0.180919	0.180919	0
9	0.194277	0.182196	0.181131	0.360575	0.180919	0.180919	0.180919	0
10	0.208154	0.19521	0.194069	0.193165	0.193842	0.193842	0.193842	0

〈Figure 2〉 Document-Term Matrix with TF-IDF value

는데, 메모리가 충분하다면 더 높은 성능을 얻기 위해 더 많은 어휘를 넣을 수 있을 것이다.

다섯째, ‘합법’과 ‘불법’ 두 종류의 파일을 뒤쫓고, 일부를 학습데이터로 삼아 불법 메시지에서 나타나는 단어 별 TF-IDF의 특징을 구한다. 본 연구에서는 학습 데이터를 70%, 테스트 데이터를 30%로 삼았다.

여섯째, SVM 파라미터 최적값을 찾아 최종 예측 모델을 생성한다. SVM은  $\gamma$ 와  $cost$ 라는 두 개의 주요 인자를 받는데,  $\gamma$ 는 3차원 평면에서 추상적으로 구분한 초평면의 기울기이고,  $cost$ 는 구분선이 가질 수 있는 과적합 정도를 어느 정도까지 감수할지를 결정한다. 이 두 값은 테스트를 여러 번 반복 수행하면서 찾게 되는데 본 연구에서는 최적의 파라미터를 찾는 `tune.svm` 함수를 이용하여  $\gamma$ 는 0.5,  $cost$ 는 10으로 하였다. 이는 일반적인 경우보다는 특히  $cost$  값이 높게 책정된 편이다.

위의 과정을 거쳐 도출된 SVM 모델을 상시 가동하여 모니터링을 함으로써 위험 요소를 탐지한다. R로 만들어진 SVM 모델은 재사용이 가능하며, 실시간 분석에 활용될 수 있다. 모델을

배치 파일 안에 넣어 필요한 경우에는 언제든지 불러서 개별 데이터의 합법 및 불법 여부를 판정할 수 있다.

## 4. 실험

### 4.1 데이터셋

본 연구에서 제안한 방법론의 성능을 파악하기 위하여 성능 실험을 수행하였다. 실험에 사용된 데이터셋은 두 가지 부류를 가지고 있다. 하나는 목표 대상인 불법 대부업 광고 데이터로서 ‘불법illegal’ 태그가 부여된 것이고, 다른 하나는 합법 및 대부업과 관련이 없는 일반적인 데이터로서 ‘합법legal’ 태그가 부여된 것이다.

학습용 데이터를 수집하기 위해 트위터 상에서 특정 키워드를 가지고 있는 메시지를 수집하였다. 그러나 사전에 정의된 키워드 집합이 없기 때문에 최소한의 씨앗 단어들(seed words)로 키워드 집합을 확장해 나가는 방식을 취하였다. 씨앗 단어들 선정은 우선 ‘대출’과 ‘사채’ 두 개의



단어로 하고, 트위터 상에서 이 단어를 갖는 자료를 수집하였다. ‘대출’과 ‘사채’ 단어의 트위터 검색 시 체언을 중심으로 <Table 2> 및 <Table 3>과 같은 주요 어휘가 수집되었다.

다음으로 추출된 키워드 후보가 대부업과 얼마나 관련이 있는지 알아보기 위하여 키워드 후보를 다시 트위터 검색어로 하여 데이터를 재수집하고, 두 명의 연구원이 위의 키워드 후보로

<Table 2> Sample SNS messages and keywords searched by ‘대출(loan)’

Case	SNS Message	keywords
1	힘들고 어려울때(8등급까지) □사업자대출,전세 대출,고액 대출, 무직자,누구나 가능□ □당일조회 당일대출을 원칙으로 합니다□ □차별화되고 깔끔한 일처리로 확실한 금액차이를 보장합니다□ (When it is difficult and difficult (up to 8th grade) □Business loan, charter loan, high loan, unemployed, anyone□□On the same day, inquiries are made on the day of the loan□□Distinct and clean work processes ensure guaranteed value for money□)	8등급, 사업자, 고액, 무직자, 누구나, 가능, 당일, 조회, 원칙, 차별화, 일처리, 확실, 금액, 차이, 보장 (8th grade, operator, high amount, unemployed, anyone, possible, day, inquiry, principle, differentiation, work process, sure, amount, difference, guarantee)
2	◆무직 과다채무자대출가능 ◆◆ 만 20세이상이면 전국 누구나 ◆★7등급이상이면 금융권무조건 승인! (◆Unlimited borrowers can borrow ◆◆ Anyone over 20 years old ◆★If you are above grade 7, you are unconditionally approved for the financial sector!)	무직, 채무자, 가능, 전국, 7등급, 금융권, 승인 (unemployed, debtor, possible, nationwide, 7th grade, financial sector, approval)
3	아덜~ㅎㅎ,,,솔직히 말혀,,,분위기 무이자로,,대출받으거시지~?ㅎㅎㅎ대현맘들~울 아덜 분위기 좀 보고가셔요~^^** (Son ~haha,,,Be honest,,, with no interest in the atmosphere,,,Did you have a loan?HeheDae Hyun Mums ~I see the atmosphere of my son ~ ^^ **)	분위기, 무이자 (atmosphere, no interest)

<Table 3> Sample SNS messages and keywords searched by ‘사채(private loan)’

Case	SNS Message	keywords
4	은행금리가 아무리 내려도, 돈이 궁한 사람들이 쓰는 사채금리가 내려가진 않는다. 돈 있는 사람은 싸게 쓰고, 없는 사람은 비싸게 쓰게 되는 것... (No matter how low the bank interest rate is, the bond interest rate of people who are in need is not lowered. People who have money spend it cheaply, and people who do not have enough money to spend it ...)	은행, 금리, 돈, 사람 (Bank, interest rate, money, person)
5	중고차담보대출/타던차그대로./저이율/사채아닙니다 중고차 똥차할부 2006년식이상차량 최저 8.9% ~20.9% 무방문,서류가필요없습니다 (Used car mortgage loan / used car as it is ../ low interest rate / not a used car shit car installment vehicle over 2006 year minimum 8.9% ~ 20.9% There is no need for documents, no documents)	중고차, 담보, 저이율, 똥차, 할부, 차량, 최저, 무방문, 서류, 필요 (used car, collateral, low interest rate, shaved car, installment, vehicle, minimum, no-show, documents, need)
6	최저금리8.9% ~ 19.9% 최대한도 4천만원, 취급수수료없음. 중도상환수수료 면제, 사채쓰지 마세요. 언제든지 상담하세요~ (Minimum interest rate 8.9% ~ 19.9% Maximum limit 40 million won, no handling fee. Please refrain from paying a modest repayment fee. Please feel free to contact us ~)	최저, 금리, 최대, 한도, 취급, 수수료, 중도, 상환, 수수료, 면제, 상담 (minimum, interest, maximum, limit, handling, commission, moderate, redemption, fee, exemption, consultation)

〈Table 4〉 Sample SNS messages and keywords searched by ‘연체(overdue)’

Case	SNS Message	Relatedness
7	신용 6/7/8등급에 연체만 없으면 100% 승인 통신연체도 8등급이시면 대납해드리고 진행해드립니다 (If you are credit 6/7/8th grade, and if you are not overdue, you will receive 100% approval.)	Yes
8	연체자 급전 현금마련 소액급전 현금지급 당일지급 전국출장가능합니다 (Payment of overdue cash Cash on delivery small payment on the day of payment is available nationwide)	Yes
9	가계 대출 연체율이 줄었지만 기업 대출 연체율은 5대 은행 대부분에서 금융위기 후 최대 폭으로 올라 전체 수익에 악영향을 미쳤다. (Although the delinquency rate for household loans has decreased, the delinquency rate for corporate loans has risen to the greatest extent after the financial crisis in most of the five major banks, which has had a negative impact on overall profits.)	No

검색된 데이터가 대부업과 조금이라도 관련이 있는지를 판정하였다. 이 작업에서는 데이터의 합법/불법(legal/illegal) 여부, 즉 불법 홍보글인지 여부는 확인하지 않고 단순히 대부업과 관련된 데이터인지만을 보았다. 예를 들어, ‘연체’라는 키워드 후보를 키워드로 데이터를 수집한 결과는 <Table 4>와 같다.

위와 같이 ‘연체’라는 어휘로 수집된 각각의 데이터에 대하여 두 명의 연구원이 개별적으로 대부업과 관련된 글인지를 판정하여 오른쪽에 기록하였다. 그 결과 위의 경우 ‘연체’라는 어휘는 앞의 두 데이터에서는 대부업 광고와 관련되어 있지만, 세 번째 데이터에서는 관련이 없게 나타났다. 각각의 데이터에 대하여 관련 여부 기록이 끝나면 총 데이터 건수에 대한 관련 비율을 조사하여, 두 명의 연구원에게서 각각 70% 이상의 비율을 보인 어휘만을 최종 키워드로 선정하였다.

다음으로 선정된 검색어를 이용하여 트위터, 뉴스, 카페, 블로그 등의 웹 자료에서 데이터를

본격적으로 검색 및 수집하였다. 앞의 수집 과정에서는 일반 PC에서 검색 API를 이용하여 수집하였으나, 본 수집 과정에서는 전문 업체의 크롤러(Crawler)를 활용하였다. 수집 대상은 2012년 1월 1일부터 2015년 12월 31일까지 3년의 기간 동안 작성된 데이터였다. 그 결과 총 820, 123 건의 데이터가 수집되었으며, 연도별로 구분하면 <Table 5>와 같다.

〈Table 5〉 The number of cases collected by years

Year	2013	2014	2-15	Total
Number of Cases	320,471	253,889	245,743	820,103

#### 4.2 전처리

학습 데이터의 품질을 개선하기 위하여 다음과 같이 전처리 작업을 하였다. 첫째, 고빈도 작성자를 우선적으로 추출하였다. 즉, 한 해 동안

<Table 6> Case frequency of one id and id type number

Ranges of the number of messages	10~19	20~29	30~39	40~49	50~59	60~69	70~79	80~89	90~99	100~
Number of IDs	3956	974	399	244	211	144	97	97	80	748

관련 키워드를 가진 글을 다량으로 작성한 아이디의 글만을 추려내었다. 일반적인 개인이 일정 기간에 대부업 관련 어휘를 가진 글을 수십 회 작성하는 일은 흔하지 않을 것이며, 특정 주제와 관련된 어휘로 반복적으로 글을 쓰는 개인은 그 주제와 관련된 일을 하고 있을 가능성이 높을 것이다. 이를 가정하고, 본 연구에서 수집된 대부업 관련 키워드를 가지고 있는 82만 건의 글에서 작성자 아이디가 동일한 글들을 분류하여 저장하였다. 그 결과 하나의 아이디로 대부업 관련 어휘가 들어간 글의 작성빈도와, 그에 해당하는 아이디의 수가 <Table 6>과 같이 정리되었다.

하나의 아이디로 대부업 관련 키워드가 들어간 글을 10건 이상 작성한 아이디는 3,956건이나 되며, 100건 이상 작성한 아이디도 748건에 이른다. 현재 단계는 학습 데이터, 특히 불법 판정을 받을 불법 홍보글 학습 데이터를 구축하는 것이므로, 여기서는 100회 이상 글을 작성한 748개의 아이디와 그 글을 선택하였다. 이 아이디로 작성된 글의 상당수는 불법 대부업 홍보글이다.

다음으로 중복글은 한 개만 남기고 모두 학습

데이터에서 삭제하였다. 학습 데이터 구축을 위해 특정 아이디로 작성된 글을 추적하는 과정에서는 실질적으로 동일한 데이터를 중복되게 보관할 이유가 없고, 검토할 대상을 줄이는 것이 효율적이기 때문이다. 중복 데이터 삭제는 두 가지 기준으로 이루어졌다. 첫 번째 기준은 <Figure 3>과 같이 작성자와 내용이 완전히 동일한 경우이다.

두 번째 기준은 <Figure 4>와 같이 작성자와 유도성 링크가 동일한 경우이다. 이 부류에 속하는 데이터들은 검열을 피하기 위하여 실질적인 내용은 같은데 띄어쓰기 등을 조금씩 다르게 한다. 그러나 글을 읽고 관심이 생긴 소비자가 최종적으로 연락할 곳은 있어야 하므로 홍보성 글은 대개 웹사이트 주소가 있는 유도성 링크나 연락할 전화번호를 남기게 된다. 이 정보를 기본키(Primary Key)로 삼으면 상당수의 데이터를 같은 것으로 보고 삭제할 수 있다. 다만 전화번호는 사람이 눈으로 보아야 이해할 수 있는 문자나 기호로 치환하는 경우가 많으므로, 본 연구에서는 추출이 보다 용이한 유도성 링크를 기본키로 삼

!!People who need living money but are afraid of high thresholds and high interest rates! We will loan 10% low interest loans to low-income / low-income people who suffer from loan-to-business loans.  
 !!People who need living money but are afraid of high thresholds and high interest rates! We will loan 10% low interest loans to low-income / low-income people who suffer from loan-to-business loans.  
 !!People who need living money but are afraid of high thresholds and high interest rates! We will loan 10% low interest loans to low-income / low-income people who suffer from loan-to-business loans.

(Below is the Original Text)

!!생활자금이 필요하지만 높은 문턱과 높은 금리가 무서운 서민! 대부업 대출에 고통받는 저신용 / 저소득 서민에게 10%의 저금리로 대출해드립니다  
 !!생활자금이 필요하지만 높은 문턱과 높은 금리가 무서운 서민! 대부업 대출에 고통받는 저신용 / 저소득 서민에게 10%의 저금리로 대출해드립니다  
 !!생활자금이 필요하지만 높은 문턱과 높은 금리가 무서운 서민! 대부업 대출에 고통받는 저신용 / 저소득 서민에게 10%의 저금리로 대출해드립니다

<Figure 3> A case of identical contents

70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%], KT, LG [40%]	▶ Kakao Talk SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%],KT, LG [40%]	▶ Kakao Talk SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%],KT,LG [40%]	▶ Kakao Talk SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK[70%], KT, LG[40%]	▶ Kakao Talk SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
(Below is the Original Text)			
70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%], KT, LG [40%]	▶ 카카오톡 SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%],KT, LG [40%]	▶ 카카오톡 SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK [70%],KT,LG [40%]	▶ 카카오톡 SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>
70% 지급 (미납, 정책상 50%) 정보이용료 SK[70%], KT, LG[40%]	▶ 카카오톡 SKY3426	<a href="https://t.co/4LHWYV8dDQ">https://t.co/4LHWYV8dDQ</a>	<a href="https://t.co/ey92LF2QjQ">https://t.co/ey92LF2QjQ</a>

〈Figure 4〉 A case of identical URL address

고 동일한 데이터를 삭제하였다.

위의 과정을 거쳐 전체 82만건 데이터에서 4,398 건이 남게 되었다. 이 결과를 정부에서 운영하는 ‘전국 등록 대부업 및 대부중개업체 조회 서비스’ ([http://www.clfa.or.kr/page/popup\\_fcsc.asp](http://www.clfa.or.kr/page/popup_fcsc.asp))에서 불법성 여부를 조회하였다. 이 사이트에서 조회한 이유는 이 사이트가 불법성을 판정하는 가장 분명한 기준이 되어주기 때문이다. 이 사이트는 정부에 대부업 신고를 하고 정식으로 대부업 관련 사업을 하고 있는 업체의 목록을 가지고 있으므로, 이 사이트에 등록된 연락처를 가지고 있는 홍보글은 합법적인 홍보글로 볼 수 있다. 본 연구에서는 위의 사이트에 소개되어 있는 모든 전화번호를 조회하여 조회가 안 되는 경우만 불법 업체로 판정하였다.

조회 결과 4,398건 중 1,413건이 불법 데이터, 2,985건이 합법 또는 대부업과는 관련이 없는 데이터로 판명되었다. 두 종류의 데이터를 각각 illegal과 legal 판정을 위한 학습 데이터로 삼았다. 그리고 불법 홍보글로 판정된 글에 대해서는 작성자의 id, 메시지의 내용, 홍보 웹사이트 주소, 웹사이트 이미지, 전화번호, 카카오톡 id 등 추적에 도움이 될 수 있는 정보를 확보하여 지자체에 전달하였다.

### 4.3 결과

불법과 합법 두 종류의 학습 데이터가 구축되었으므로 이제 SVM 알고리즘을 이용하여 모델을 만들어 실험할 수가 있다. 본 연구에서 제안하는 방법에 대한 성능평가를 위해 다음과 같은 방법들과 비교하였다.

방법 1: MLE (Maximum Likelihood Estimation) - 최대우도 추정으로 금융사기 글인지의 여부를 판정하는 방법이다.

방법 2: Term Frequency - 불법대부나 불법방문 판매 사례에서 자주 등장하는 어휘(term)들을 정의하고, 그 어휘들이 등장하면 불법으로 판정하는 방법이다. 해당 어휘들은 불법방문판매의 경우 ‘당일 급전, 소액 급전, 한도 조회, 월변, 저신용, 후불유심, 무방문, 사금융, 대포폰, 서민 고통, 원라인, 무보증, 연체, 8등급, 연체금, 신용불량, 보이스피싱’ 등의 단어를, 불법방문판매의 경우 ‘다단계, 네트워크 마케팅, 후원수당, 추천수당, 기본 고수익, 무점포 사업’ 등의 단어를 사용하고 있어 이들을 판정 어휘로 활용하였다.

방법 3: Collective Intelligence - 불법대부광고와

불법방문판매에 대해 판정하는 복수의 경험자들의 합의에 의한 판정이다.

방법 4: Proposed - 본 연구에서 제안한 방법이다.

네 가지 방법에 대해 10-fold cross validation으로 성능을 측정한 결과 도출된 혼동행렬표를 기반으로 하여 구한 전반적인 정확도(overall accuracy)는 <Table 7>과 같다. 이 결과에 의하면 제안한 방법의 전반적인 정확도가 불법대부광고의 경우는 92.41%, 불법방문판매의 경우에는 77.75%로 비교 방법에 비하여 우수함을 보이고 있다. 또한 불법대부광고의 경우, 불법인데 불법이 아닌 것으로 잘못 판정한 경우는 423건 중 100건으로 24%이고, 불법이 아닌데 불법으로 잘못 판정한 경우는 895건 중 0건으로 0%이다. 불법방문판매의 경우에는 불법인데 불법이 아닌 것으로 잘못 판정한 경우는 172건 중 57건으로 33%이고, 불법이 아닌데 불법으로 잘못 판정한 경우는 300건 중 0건으로 0%였다. 따라서 SVM 기반의 제안한 예측 모델을 따르면 일부 불법 홍보글을 놓칠 수는 있지만, 불법으로 예측한 데이터는 거의 틀림없이 불법인 것으로 판정함을 파악하였다.

<Table 7> Accuracy Comparison of Algorithms

Data Set	Algorithm	Overall Accuracy
Illegal ads on loans	MLE	67.91%
	Term Frequency	58.54%
	Collective Intelligence	85.37%
	Proposed	92.41%
Illegal ads on visit sales	MLE	53.83%
	Term Frequency	48.78%
	Collective Intelligence	70.73%
	Proposed	77.75%

## 5. 토론

### 5.1 이론적 시사점

먼저 본 연구는 SNS와 같은 비정형 데이터 원천의 비정상성에 의하여 발생하는 위기 관리에 대한 틀을 제시하였다는데 의의가 있다. 첫째, Henderson(2004)의 연구를 이론적 근간으로 하여 SNS에서의 위기 관리의 다양한 특성을 정의하였으며, 위기관리 프로세스도 정의하여 추후 유사한 연구에서 참고하도록 하였다. 본 연구에서는 사이버범죄 발생 전(Pre-Cybercriminality), 관리의 첫 단계인 위기의 인식 단계에 집중하여 제안하였으나 추후 연구에서는 다른 특성과 다른 단계에서도 연구가 필요하다는 점을 제공하여 후속 연구 개발에 시사점을 제공할 수 있었다.

둘째, SVM과 같은 판별알고리즘을 한글 텍스트 판별 분석에 적용할 때 고려해야 하는 사항을 밝혀 연구자들에게 기여할 것으로 기대한다. 한글 형태소에 분석된 모든 품사류를 적용하는 것이 아니라, 판별에 가장 도움을 주는 명사류를 택하였고, 일반적인 분석에서는 불용어로 간주되는 기호류를 오히려 분석 대상 품사로 선정하였다. 이러한 방법은 계산의 현실성을 갖추면서도 분석의 효용성은 오히려 높이는 결과를 낳았다.

### 5.2 실무적 시사점

본 연구는 금융사기나 SNS 관리를 해야 하는 공공기관이나 지자체, 사기업에 다음과 같이 몇 가지 적용상 시사점을 준다. 첫째, 본 연구는 기계학습을 활용하여 금융 사기를 시도하는 SNS 메시지를 자동적으로 즉시 탐지할 수 있음을 보여주었다. 이것은 금융 사기 사건이 이미 발생한

후에 사후처리 식으로 대응하는 것보다 더 손실을 줄여주는 방법이다.

둘째, 불법 대출, 불법 대부업과 같은 지방자치단체 또는 국가에서 민감하게 대응하는 이슈에 대하여 텍스트 마이닝 분석 기법이 유용하게 활용될 수 있음을 보였다. 그동안 연구적 차원에서 SVM을 비롯한 다양한 알고리즘을 적용한 연구는 적지 않았으나, 정부 기관에 적용한 사례는 찾기 어렵다. 본 연구에서는 82만 건의 SNS 데이터를 수집하고, 여기에서 실제 사이버 범죄 데이터를 추출 및 판별하는 모델을 생성하였다. 본 연구는 가능성을 탐지해 보는 파일럿 프로젝트였으나, 해당 지자체는 이의 실제 적용을 검토해 보고 있다.

다음으로, 본 연구와 같은 실제 적용 사례에서는 정답 분류가 있는 학습 데이터가 사전에 주어지지 않으나, 본 연구에서는 분석을 위한 학습 데이터 구축 과정을 밝혀 다른 사례에서의 문제 해결에 본 연구의 내용이 실무적인 도움이 될 것이다. 본 연구는 네트워크에 존재하는 방대한 데이터에서 씨앗 단어들(seed words)을 이용하여 전체 분석 대상 데이터를 82만 건으로 압축하고, 거기에서 다시 데이터의 외적, 내적 특징을 이용해 데이터를 압축하여 최종 분석 데이터를 사람이 직접 확인 가능한 수준인 4,400 건으로 줄였다. 이 4,400 건 데이터에 대하여 직접 불법 홍보글 여부를 확인함으로써 학습 데이터를 구축할 수 있었다.

### 5.3 추후 연구 방향

앞에서 밝힌 바와 같이 국내 사이버공간을 염두에 두고 지자체 주민들에게 상당한 금전적 손해를 가할 수 있는 급작스러운 불법대출 및 불법

(방문)판매 글의 게재 행동을 실시간 탐지하고 대응하여 SNS위기 관리를 하는 분야에 집중하고자 한다. 그리고 위기 관리 프로세스 관점에서는 본 연구에서는 사전적 사이버범죄 발생 전 관리의 첫 단계에 해당하는 위기의 인식 단계에 집중하였다. 따라서 추후에는 위기관리 프로세스의 이후 단계에 대한 연구를 진행할 것이다. 예를 들어 위험이 감지된 경우 이에 대응 또는 경감하기 위한 방지 행위를 정하는 지능적 방법이 필요하다. 또한 피해 복구에 대한 연구에도 지능적 기술의 도움이 가능하다. 예를 들어 금융 사기를 일으킨 기관은 계정을 바꾸는 등의 방법을 통해 다시 SNS에 재접근하여 유사 범죄를 일으키려 한다. 따라서 id는 다르지만 유사한 범죄 패턴을 보이는 데이터에 대해서 대응할 수 있어야 한다. 이를 위해 사례기반추론 등과 같은 유연성을 가진 탐색적 기법을 통해서 과거 범죄와의 유사도를 검토할 수 있다.

## 6. 결론

온라인 경제가 활성화되면서 금융사기와 같은 역기능을 방지하는 것은 매우 중요하면서도 그동안 상당한 인력과 비용이 수반되는 것이었다. 본 논문은 SNS를 매체로 하는 온라인 활동에서의 위기관리의 프레임워크 및 위기관리 프로세스를 제시하고 불법대출과 불법방문판매에 국한하여 수도권 모 지자체의 금융사기를 혁신적으로 방지할 수 있는 방법을 개발하여 파일럿 적용한 것이다. 그 결과 SNS에 실시간으로 등록되는 글에 대해서 자동적으로 불법대출과 불법방문판매에 해당하는지를 판정하는 알고리즘을 개발하였으며 정확도 면에서 그 성능은 사람에 의한 판

정 및 일반적으로 적용되던 방법보다 더 우수했다. 향후 불법대출과 불법방문판매 외의 영역에 대한 자동 모니터링으로까지 확산된다면 학술적으로, 실무적으로 더욱 기여할 수 있을 것이다.

## 참고문헌(References)

- Balamurugan, S., R. Rajaram, G. Athiappan and M. Muthupandian, "Data Mining Techniques for Suspicious Email Detection: A Comparative Study," *Proceeding of the IADIS European Conference Data Mining 2007*, (2007), 213~217.
- Banerjee, A., Barman, D., Faloutsos, M., & Bhuyan, L. N. Cyber-fraud is one typo away. *INFOCOM 2008. The 27th Conference on Computer Communications*. IEEE (2008) (pp. 1939-1947). IEEE.
- Bayer, M., W. Sommer and A. Schacht, "Reading emotional words within sentences: The impact of arousal and valence on event-related potentials," *International Journal of Psychophysiology*, Vol.78, No.3(2010), 299~307.
- Castell, M. R. F. and L. B. Dacuycuy, "Exploring the use of exchange market pressure and RMU deviation indicator for early warning system (EWS) in the ASEAN+3 region," *DLSU Business & Economics Review*, Vol.18, No.2(2009), 1~30.
- Comfort, L. K., "Crisis Management in Hindsight: Cognition, Communication, Coordination, and Control," *Public Administration Review*, Vol. 67, No.1(2007), 189~197.
- Choi, S., Jeon, J., Subrata, B., Kwon, O., "An efficient estimation of place brand image power based on text mining technology," *Journal of Korea Intelligent Information Systems*, Vol. 21, No.2(2015), 113~129.
- (최석재, 전종식, 권오병, "텍스트마이닝 기반의 효율적인 장소 브랜드 이미지 강도 측정 방법," *지능정보연구*, Vol.21, No.2(2015), 113~129.)
- Choi, S. Song, Y., Kwon, O., "Analyzing contextual polarity of unstructured data for measuring subjective well-being," *Journal of Intelligent Information Systems*, Vol.22, No.1(2016), 83~105.
- (최석재, 송영은, 권오병, "주관적 웰빙 상태 측정을 위한 비정형 데이터의 상황기반 긍정성 분석 방법," *지능정보연구*, Vol. 22, No.1(2016), 83~105.)
- Cui, M., Jin, Y. and Kwon, O., "A method of analyzing sentiment polarity of multilingual social media : A case of korean-chinese languages," *Journal of Intelligent Information Systems*, Vol.22, No.3(2016), 91~111.
- (최미나, 진윤선, 권오병, "다국어 소셜미디어에 대한 감성분석 방법 개발," *지능정보연구*, Vol. 22, No.3(2016), 91~111.)
- DeAngelo, H. and R. M. Stulz, "Liquid-claim production, risk management, and bank capital structure: Why high leverage is optimal for banks," *Journal of Financial Economics*, Vol.116, No.2(2015), 219~236.
- Dionne, G., "Risk management: History, definition, and critique," *Risk Management and Insurance Review*, Vol.16, No.2(2013), 147~166.
- Flores, C., "Management of catastrophic risks considering the existence of early warning systems," *Scandinavian Actuarial Journal*,

- Vol.1(2009), 38~62.
- Folino, G., A. Forestiero, G. Papuzzo and G. Spezzano, "A grid portal for solving geoscience problems using distributed knowledge discovery services," *Future Generation Computer Systems*, Vol.26, No.1 (2010), 87~96.
- Grace, M. F., J. T. Leverty, R. D. Phillips and P. Shimpi, "The value of investing in enterprise risk management," *Journal of Risk and Insurance*, Vol.82, No.2(2015), 289~316.
- Hassan, A. B., F. D. Lass and J. Makinde, "Cybercrime in Nigeria: Causes, Effects and the Way Out," *ARPJ Journal of Science and Technology*, Vol.2, No.7(2012), 626~631.
- Henderson, L. J., "Emergency and disaster: Pervasive risk and public bureaucracy in developing nations," *Public Organization Review*, Vol.4, No.2(2004), 103~119.
- Holton, C., "Identifying disgruntled employee systems fraud risk through text mining: a simple solution for a multi-billion dollar problem," *Decision Support Systems*, Vol.46, No.4(2009), 853~864.
- Jans, M., N. Lybaert and K. Vanhoof, "Internal fraud risk reduction: results of a data mining case study," *International Journal of Accounting Information Systems*, Vol.11, No.1(2010), 17~41.
- Joachims, T., "Text categorization with support vector machines: Learning with many relevant features," *Technical Report LS8-Report*, Universitaet Dortmund, 1997.
- Kim, J. and Kwon, O., "A method of predicting service time based on voice of customer data," *Journal of the Korea society of IT services*, Vol. 15(2016), 197~210.
- (김정훈, 권오병, "고객의 소리 (VOC) 데이터를 활용한 서비스 처리 시간 예측방법," *한국 IT 서비스학회지*, Vol.15(2016), 197~210.)
- Kumari, A., K. Sharma, and M. Sharma, "Predictive Analysis of Cyber Crime Against Women in India and Laws Prohibiting Them," *International Journal of Innovations & Advancement in Computer Science*, Vol.4, No.3(2015), 1-6.
- Lin, M., X. Ke and A.B. Whinston, "Vertical differentiation and a comparison of online advertising models," *Journal of Management Information Systems*, Vol.29, No.1(2012), 195~236.
- Mazurczyk, W., T. Holt, and K. Szczypiorski, "Guest Editors' Introduction: Special Issue on Cyber Crime," *IEEE Transactions on Dependable and Secure Computing*, Vol.13, No.2(2016), 146~147.
- McEntire, David A. *The status of emergency management theory: Issues, barriers, and recommendations for improved scholarship*. University of North Texas. Department of Public Administration. Emergency Administration and Planning, (2004).
- Michaelidou, N., N. T. Siamagka and G. Christodoulides, "Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands," *Industrial Marketing Management*, Vol.40(2011), 1153~1159.
- Nykodym N., R. Taylor and J. Vilela, "Criminal profiling and insider cyber crime," *Digital Investigation*, Vol.2(2005), 261~267.
- Pérez-González, F., and H. Yun, "Risk management and firm value: Evidence from weather derivatives," *The Journal of Finance*,



- Vol.68, No.5(2013), 2143~2176.
- Petak, W. J., "A Challenge for Public Administration," *Public Administration Review*, Vol.45(1985), 3~7.
- Sadgrove, K. *The complete guide to business risk management*. Routledge, 2016.
- Sahami, M., S. Dumais and D. Heckerman and E. Horvitz, "A Bayesian Approach to Filtering Junk E-Mail," In *Learning for Text Categorization: Papers from the 1998 workshop*, Vol.62(1998), 98~105.
- Sreenivasulu, V., and R.S. Prasad, "A Methodology for Cyber Crime Identification using Email Corpus based on Gaussian Mixture Model," *International Journal of Computer Applications*, Vol.117, No.13 (2015), 29~32.
- Waugh, W. L., and G. Streib, "Collaboration and leadership for effective emergency management," *Public administration review*, Vol.66, No.1(2006), 131~140.
- Yates, D, and S. Paquette, "Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake," *International Journal of Information Management*, Vol.31(2011), 6~13.
- Zhao. L. and Y. Jiang, "A game theoretic optimization model between project risk set and measurement," *International Journal of Information Technology & Decision Making*, Vol.8, No.4(2009), 769~786.

## Abstract

# Financial Fraud Detection using Text Mining Analysis against Municipal Cybercriminality

Sukjae Choi\* · Jungwon Lee\*\* · Ohbyung Kwon\*\*\*

Recently, SNS has become an important channel for marketing as well as personal communication. However, cybercrime has also evolved with the development of information and communication technology, and illegal advertising is distributed to SNS in large quantity. As a result, personal information is lost and even monetary damages occur more frequently. In this study, we propose a method to analyze which sentences and documents, which have been sent to the SNS, are related to financial fraud.

First of all, as a conceptual framework, we developed a matrix of conceptual characteristics of cybercriminality on SNS and emergency management. We also suggested emergency management process which consists of Pre-Cybercriminality (e.g. risk identification) and Post-Cybercriminality steps. Among those we focused on risk identification in this paper.

The main process consists of data collection, preprocessing and analysis. First, we selected two words 'daechul(loan)' and 'sachae(private loan)' as seed words and collected data with this word from SNS such as twitter. The collected data are given to the two researchers to decide whether they are related to the cybercriminality, particularly financial fraud, or not. Then we selected some of them as keywords if the vocabularies are related to the nominals and symbols. With the selected keywords, we searched and collected data from web materials such as twitter, news, blog, and more than 820,000 articles collected.

The collected articles were refined through preprocessing and made into learning data. The preprocessing process is divided into performing morphological analysis step, removing stop words step, and selecting valid part-of-speech step. In the morphological analysis step, a complex sentence is transformed into some morpheme units to enable mechanical analysis. In the removing stop words step, non-lexical elements such as numbers, punctuation marks, and double spaces are removed from the text.

---

\* Humanitas BigData Research Center, Kyung Hee University

\*\* School of Management, Kyung Hee University

\*\*\* Corresponding Author: Ohbyung Kwon

School of Management, Kyung Hee University

26 Kyungheedae-ro, Dongdaemun-gu, Seoul 130-701, Korea

E-mail: obkwon@khu.ac.kr

In the step of selecting valid part-of-speech, only two kinds of nouns and symbols are considered. Since nouns could refer to things, the intent of message is expressed better than the other part-of-speech. Moreover, the more illegal the text is, the more frequently symbols are used.

The selected data is given 'legal' or 'illegal'. To make the selected data as learning data through the preprocessing process, it is necessary to classify whether each data is legitimate or not. The processed data is then converted into Corpus type and Document-Term Matrix. Finally, the two types of 'legal' and 'illegal' files were mixed and randomly divided into learning data set and test data set. In this study, we set the learning data as 70% and the test data as 30%.

SVM was used as the discrimination algorithm. Since SVM requires gamma and cost values as the main parameters, we set gamma as 0.5 and cost as 10, based on the optimal value function. The cost is set higher than general cases. To show the feasibility of the idea proposed in this paper, we compared the proposed method with MLE (Maximum Likelihood Estimation), Term Frequency, and Collective Intelligence method. Overall accuracy and was used as the metric. As a result, the overall accuracy of the proposed method was 92.41% of illegal loan advertisement and 77.75% of illegal visit sales, which is apparently superior to that of the Term Frequency, MLE, etc. Hence, the result suggests that the proposed method is valid and usable practically.

In this paper, we propose a framework for crisis management caused by abnormalities of unstructured data sources such as SNS. We hope this study will contribute to the academia by identifying what to consider when applying the SVM-like discrimination algorithm to text analysis. Moreover, the study will also contribute to the practitioners in the field of brand management and opinion mining.

**Key Words** : SVM; Financial Fraud Detection; Cybercrime; Crisis Management; Text Mining

Received : July 12, 2017 Revised : September 17, 2017 Accepted : September 20, 2017

Publication Type : Regular Paper Corresponding Author : Ohbyung Kwon

## 저자 소개



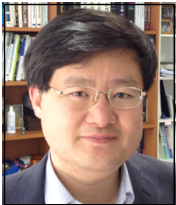
### 최석재

현재 경희대학교 빅데이터연구센터 전임연구원으로 재직 중이다. 고려대학교 국어국문학과에서 학사, 석사, 박사 학위를 취득하였고, 미국 카네기 멜론 Language Technologies Institute 객원연구원, 중국 연변과학기술대학 언어공학연구소 실장, 성신여대 국어국문학과 초빙교수 등을 거쳤다. 한국어 비정형 데이터에서 정보를 추출하는 방안이 주요 관심 분야이다.



### 이증원

한국산업기술대학교 e비즈니스과에서 학사, 경희대학교 일반대학원 빅데이터경영과에서 석사학위를 취득했다. 주요 관심분야는 텍스트 마이닝, 기계학습 등이다.



### 권오병

현재 경희대학교 경영대학 교수로 재직 중이다. 서울대학교에서 경영학사, 한국과학기술원경영과학과에서 공학석사 및 박사학위를 취득했다. 주 관심 분야는 빅데이터, IoT, 인간-로봇 인터페이스, 의사결정지원시스템 등이다.