

트위터 상의 악의적 이용 자동분류*

Automatic Classification of Malicious Usage on Twitter

김민철 (Meen Chul Kim)**
심규승 (Kyu Seung Shim)***
한남기 (Nam Gi Han)****
김예은 (Ye Eun Kim)*****
송민 (Min Song)*****

목차

- | | |
|-------------------|------------|
| 1. 서론 | 4. 실험결과 분석 |
| 2. 선행 연구 | 5. 결론 |
| 3. 악의적 이용 자동분류 실험 | |

초록

웹 2.0과 소셜미디어의 출현은 빅 데이터의 생성을 주도하고 있다. 하지만 이와 정비례하여 권리침해 및 사회질서위반 등의 역기능도 가파르게 증가하고 있는 실정이다. 이에 본 연구에서는 소셜미디어를 통해 생산되는 방대한 양의 데이터 중에서 악의적 이용을 필터링하여 건전한 사이버 생태계 구현에 기여할 수 있는 자동화 기법을 고안하고자 하였다. 실험 결과, 악의적 이용의 효과적 분류를 위해 사전에 이용한 규칙기반 자동분류에서 통계적으로 유의미한 성능 향상률을 보였다. 이러한 연구결과를 바탕으로 효과적인 악의적 이용 자동분류를 위한 방법을 제시하였다.

ABSTRACT

The advent of Web 2.0 and social media is taking a leading role of emerging big data. At the same time, however, informational dysfunction such as infringement of one's rights and violation of social order has been increasing sharply. This study, therefore, aims at defining malicious usage, identifying malicious feature, and devising an automated method for classifying them. In particular, the rule-based experiment reveals statistically significant performance enhancement.

키워드: 소셜미디어 마이닝, 트위터, 악의적 이용, 자동분류

Social Media Mining, Twitter, Malicious Usage, Automatic Classification

* 본 연구는 2012년 정부(교육과학기술부)의 재원으로 한국연구재단 한국사회과학연구지원사업(SSK)의 지원을 받아 수행된 것임(NRF-2012-2012S1A3A2033291).

** 연세대학교 문헌정보학과 대학원(andrewewans@yonsei.ac.kr)

*** 연세대학교 문헌정보학과 석사과정(moomyungz@naver.com)

**** 연세대학교 문헌정보학과 학사과정(hng88@naver.com)

***** 연세대학교 문헌정보학과 석사과정(arcoiris@yonsei.ac.kr)

***** 연세대학교 문헌정보학과 부교수(min.song@yonsei.ac.kr)

논문접수일자: 2013년 1월 15일 최초심사일자: 2013년 1월 21일 게재확정일자: 2013년 2월 2일
한국문헌정보학회지, 47(1): 269-286, 2013. (<http://dx.doi.org/10.4275/KSLIS.2013.47.1.269>)

1. 서론

웹 2.0의 출현과 함께, 최근 웹에서는 정보생산, 가공 및 전달 등 정보행위와 관련한 일련의 체계가 수요자, 즉 이용자 중심으로 변화하고 있다. 이러한 변화는 이용자들에게 손쉽게 정보 및 콘텐츠 등을 생산하고 공유할 수 있는 환경을 제공하였다. 이를 통해 이용자들은 기존의 수동적인 콘텐츠 소비자(Consumer)에서, 지식과 정보의 생산과 소비를 동시에 주도하는 프로슈머(Prosumer), 즉 생산적 소비자로 거듭나게 되었다. 이러한 정보주체의 진화는 '빅데이터(Big Data)'라 불리는 정보의 '초(超)범람'을 야기하였으며, 이러한 데이터는 생성량·주기·형식 등이 기존 데이터에 비해 너무 크기 때문에, 데이터의 수집·저장·검색·분석은 새로운 국면을 맞이하게 되었다.

한편 소셜미디어(Social Media)는 이용자들의 자발적이고 적극적인 참여를 기반으로 빅데이터의 출현을 가능케 한 웹 2.0의 대표적인 서비스이다. 일반적으로 '이용자가 입력한 프로파일'에 기반하여 서로 정보와 의견을 공유하면서 대인관계망을 넓힐 수 있는 플랫폼'으로 정의되는 소셜미디어는, 사용에 전문적인 지식이나 기술이 필요하지 않고, 실시간으로 콘텐츠를 생산하고 공유할 수 있으며, 유사한 관심사를 가진 이용자들 간에 사회적 네트워크 형성이 가능하다는 장점들 때문에 전 세계적으로 많은 이용자들에게 각광받고 있는 서비스이다(정유진, 배국진 2008; Boyd and Ellison 2007). 대표적인 소셜미디어인 페이스북(facebook)과 트위터(twitter)의 이용자는 전 세계적으로 각각 8억 5,000만 명과 5억 명으로 추산되고 있다

(윤미선, 이종혁 2012).

이러한 소셜미디어의 출현은 이용자를 기존의 데스크탑을 통한 정보접근에서 벗어나 다양한 스마트 기기를 통해 시공간의 제약을 받지 않고 실시간으로 온라인 커뮤니케이션 채널에 접근할 수 있게 함으로써 사람들의 커뮤니케이션 양식을 새롭게 변화시키고 있다.

하지만 개인과 개인 간의 교류가 급격히 증가함에 따라 오프라인에서 이루어지던 집단 따돌림 등의 공격적인 행위가 스마트 기기를 활용하여 이루어지는 형태로 발전한 '스마트 불링(Smart Bullying)' 등의 역기능도 출현하기 시작했다. 방송통신심의위원회에서는 이를 특별히 '불법·청소년유해정보'로 규정하고 불법·청소년유해정보 신고접수 및 처리, 인터넷상의 각종 피해에 대한 상담안내를 수행하고 있는 실정이다. IT통계포털에 따르면 권리침해 및 사회질서위반 부문 2009년부터 2011년까지의 심의 및 시정 사례는 16,687건, 38,439건, 48,690건으로 매년 가파른 증가세를 보이고 있다(〈표 1〉 참조). 이러한 현상은 웹 서비스의 비대면성, 익명성, 그리고 빠른 전파성 등에 기인하는 것으로 분석되고 있다(이규안 2012).

〈표 1〉 최근 3년간 악의적 이용 실태

구분	2009년	2010년	2011년
권리침해	5,114	2,364	3,501
질서위반	11,573	36,075	45,189
계	16,687	38,439	48,690

출처: IT통계포털. 정보화/IT활용 역기능. [online]. [cited 2012.11.7].

그렇기 때문에 소셜미디어를 통해 실시간으로 생산되는 방대한 양의 데이터 중에서 유해

한 정보를 필터링하고, 나아가 건전한 사이버 생태계 구현에 기여할 수 있는 자동화 기법의 고안이 필요한 시점이다.

그러나 지금까지 소셜미디어를 대상으로 한 국내외 대부분의 연구는 오피니언 마이닝, 이용 행태 분석, 그리고 비교적 패턴예측이 쉬운 스팸 필터링 등의 주제에 국한되어있는 실정이다.

이에 본 연구에서는 '악의적 이용(Malicious Usage)'을 '사람을 비방할 목적으로 공연히 사실 또는 허위의 사실을 적시하거나 비속어 등을 사용하여 타인의 명예를 훼손하는 내용의 이용 행태'로 정의하고, 최근 각광받고 있는 소셜미디어 중의 하나인 '트위터'를 대상으로 웹에서 발생하는 '악의적 이용'의 특징을 분석하고자 하였다. 특히 실험문헌 구축과정에서 전문가의 적합성 판단(Relevance Judgement)을 통하여 악의적 이용의 효과적 분류를 위해 사용할 수 있는 자질과 학습데이터의 후보군을 질적으로 확인하였으며, 이에 기반하여 후속 연구자가 이용 가능한 학습문헌을 구축하였다. 또한 악의적 자질을 이용해 실질적으로 자동분류의 성능을 향상시켰으며, 이를 통해 악의적 이용을 자동으로 분류하여 건전한 사이버 생태계 구현에 기여할 수 있는 방법을 제시하였다.

트위터 상의 악의적 이용 자동분류를 위해 본 연구에서 다루고자 하는 연구질문은 다음과 같다.

- 연구질문 ①: 트위터 상의 악의적 이용과 비악의적 이용의 특징은 무엇인가?
- 연구질문 ②: 악의적 이용과 비악의적 이용에서 나타나는 악의적 자질과 비악의적 자질은 악의적 이용의 분류에 효과적인가?

본 연구의 나머지 부분은 다음과 같이 구성되어 있다. 2장에서는 소셜미디어 및 악의적 이용 분류에 대한 선행 연구들을 살펴본다. 3장에서는 악의적 이용 자동분류를 위한 '악의적 자질(Malicious Feature)'을 정의하고 실험집단 및 설계에 대해 설명한다. 4장에서는 실험결과와 평가에 대해 보고하며, 연구에 대한 총체적 결론과 후속연구 제안으로 5장을 맺는다.

2. 선행 연구

이 장에서는 소셜미디어 마이닝, 악의적 이용 및 스팸 필터링 등 소셜미디어와 관련된 선행 연구들을 살펴본다.

2.1 소셜미디어 마이닝

소셜미디어는 실시간으로 생성되는 방대한 양의 문자 데이터 외에도 의제설정 및 여론형성 등 여론선도(opinion leading)를 위한 영향력 있는 채널로 사용되고 있다는 점에서, 그 출현과 함께 정보검색과 텍스트마이닝의 주요 연구대상이 되어왔다.

Cha et al.(2010)은 트위터에서 누가 유력자가 될 수 있는지를 알아보기 위해 사용자의 팔로어(follower) 수, 트윗이 리트윗(retweet)된 횟수, 멘션(mention)된 횟수 등 다양한 기준을 총 5,200만 명의 트위터 이용자와 연관된 2,000만 개의 팔로우 링크와 1,700만의 트윗(게시글)에 적용하여 비교하였다. 그 결과 적극적인 트위터 사용자는 약 600만 명으로서 전체의 8.6%에 불과한 것으로 나타났으며, 트위터 사용자

들은 불균형한 영향력을 가지고 있기 때문에 트위터 내에서 영향력과 명성은 항상 일치하지는 않아 팔로어의 수가 반드시 유력함을 측정하는 결정적 지표일 수 없다는 결론을 제시했다. 따라서 본 연구에서는 팔로어의 수를 악의적 이용이 발생하는 맥락정보로 고려하지 않았다.

한편 이원태 등(2011)은 약 6백만 건의 한국어 트윗을 대상으로, 소셜미디어 유력자들(social media influentials)의 네트워크 특성과 역할을 규명하였다. 그 결과 한국의 트위터 유력자들은 글로벌 네트워크와는 달리 이용자 상호작용(user interaction)의 측면에서 매우 강력한 네트워크 호혜성을 특징적으로 보여주고 있는 것으로 나타났다. 또한 이들 트위터 유력자의 영향력은 그들이 보유한 팔로어 규모에 의존하기 보다는 적극적인 리트윗(RT)을 매개로 하는 수많은 팔로어들과의 개방적 소통방식에 의해 적극 추동된다는 점이 밝혀졌다. 이에 본 연구에서는 악의적 이용을 통한 집단 따돌림 등의 이용행태 분석을 위해 리트윗을 악의적 이용이 출현하는 주요한 근원으로 규정하였다.

이상의 연구들에서 착안하여 본 연구에서는 이용자 프로파일, 트윗, 리트윗, 그리고 멘션을 악의적 이용 분류를 위한 요소로 고려하고자 하였다.

한편 박건우 외(2011)는 감정 단어 사전 구축과 단어 사용 빈도수를 이용한 가중치 측정을 통해 소셜미디어에서의 오피니언 성향을 판단하는 시스템을 제안하였다. 본 연구에서는 웹 용어의 역동성을 고려하여 위의 연구에서 고려한 요소들을 악의적 이용 자동분류에 반영하였다.

2.2 악의적 이용 분류

현재까지 악의적 이용 분류에 관한 연구는 온라인 뉴스기사 또는 블로그 등의 악성 댓글과 스팸을 중심으로 이루어져 왔다.

김현정 등(2011)은 총 1,700개 댓글(학습문헌 1,500개, 실험문헌 200개)을 대상으로, 악성 댓글 자동판별 시 온라인에서 발생하는 언어과피 현상으로 인해 동일한 의미의 자질을 인식하지 못하는 문제를 해결하기 위해 유니코드를 이용하여 빈도수가 높은 글자는 초성으로 변경한 후 자질을 추출하는 FFP(Feature Frequency Profile) 방법을 제안하였다. 하지만 어의(semantic)별 가중치를 고려하지 않았다는 한계가 있다.

김묘실과 강승식(2006)은 총 2,200개 댓글(학습문헌 1,980개, 실험문헌 220개)을 대상으로, 형태소 분석기(KTL)를 이용하여 추출한 품사 정보를 조합하여 자질을 추출하고 자질의 tf-idf 가중치를 계산하여 SVM 알고리즘을 통해 악성 댓글 여부를 판단하는 시스템을 제안하였다. 하지만 웹용어 및 비속어의 경우 그 교착 정도가 심하기 때문에 기존의 형태소 분석 기법과 품사 태깅만으로는 그 의미를 온전히 식별할 수 없다는 한계를 갖는다. 본 연구에서는 그러한 한계를 극복하고자 새로운 악의적 자질로 구성된 어휘 사전을 문헌 학습과정에 사용하였다.

배민영과 차정원(2008)은 YAHOO Korea의 정치 뉴스 분야에서 무작위로 선택된 기사를 대상으로, 토픽 시그니처(Topic Signature)를 이용하여 댓글 분류 시스템을 구현하였다. 특히 악성 댓글의 특정 구간을 연구자가 직접 선택하여 분류결과에 평가를 용이하게 하였으나, 악성

자질로 선정된 용어들의 식별력이 방대한 양의 데이터가 실시간으로 생성되는 환경에서는 떨어질 수 있다는 한계를 갖는다.

2.3 소셜미디어 스팸 필터링

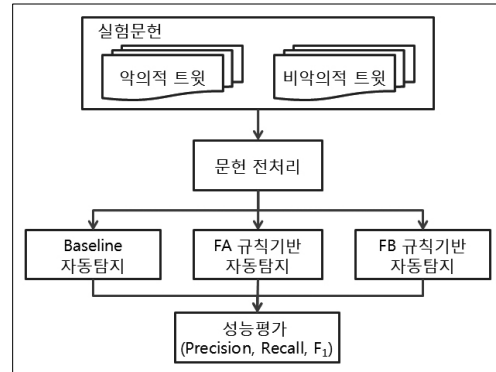
Lee, Caverlee, and Webb(2010)은 트위터와 마이스페이스(MySpace)를 대상으로 스팸머의 유형을 구분하고 그 유형에 따른 탐지에 유용한 자질(user image, profile, URL, hash tag 등)을 제안하였다.

전희원과 임해창(2007)은 블로그 본문과 댓글 간에는 유사도가 존재하는 반면, 스팸 댓글은 이러한 유사성을 보이지 않는다는 전제 하에 주제어에 대한 동시출현(co-occurrence) 확률 자질 정보를 역카이제곱 분류기에 적용하여 스팸 필터의 정확도를 높이고자 하였다.

하지만 스팸은 그 종류와 유형이 본 연구에서 다루고자 하는 악의적 이용에 비해 예측가능하기 때문에 본 연구에 적용하는 데에는 한계가 있다.

3. 악의적 이용 자동분류 실험

이 장에서는 악의적 이용 자동분류를 위한 실험문헌의 수집과정, 수집과정을 통해 밝혀진 '악의적 자질(Malicious Feature)'의 특성, 그리고 실험설계에 대해 설명한다. 전체적인 실험의 절차는 다음의 <그림 1>과 같다.



<그림 1> 실험 절차

3.1 실험집단

본 연구를 위한 실험문헌의 주제는 2012년 7월에서 8월까지 한 달간 트위터 한국어 홈페이지(<http://www.twitter.com/>)에 게재된 트윗(tweet) 중 악의적 이용의 사례가 가장 빈번히 발생하고 있을 것으로 예상되는 스포츠 및 연예 분야로 한정하였다. 분류기의 학습에 사용된 문헌 중 악의적 트윗은 해당기간 동안 수작업으로 수집된 700여 건의 악의적 트윗 중에서 전문가 10명 중 7명 이상이 악의적 이용이라고 동의한 500여 건의 트윗으로 선정하였다. 또한 학습을 위해 사용된 비악의적 트윗은 트위터에서 제공하는 오픈 Java Library인 Twitter4J의 API를 이용하여 무작위로 수집된 1,000건의 트윗 중 전문가의 판단을 거친 500건으로 선정하였다.

이후 수집된 문서들에서 실험과정에 필요한 사용자 ID와 트윗 본문을 추출하였고, 이 중 트윗 본문을 악의적 이용 자동분류를 위한 자질의 후보로 사용하였다. 실험문헌에 대한 요약정보는 다음의 <표 2>와 같다.

〈표 2〉 실험문헌 요약정보

구분	악의적 트윗	비악의적 트윗
이용자 수	441	469
총 건수	538	500
트윗 당 평균 단어	12.11	15.45
트윗 당 평균 길이	53.22	73.96
단어 당 평균 길이	4.79	4.40

※ 트윗 및 단어 당 평균 길이 단위: 글자 수
 단어 당 평균 길이 = $\frac{\text{트윗 당 평균 길이}}{\text{트윗 당 평균 단어}}$

3.2 악의적 자질(Malicious Feature)

본 연구의 조작적 정의에 기반 한 트위터 상의 악의적 이용의 행태는 크게 1) 트윗 내 비속어 사용 또는 허위 사실 유포 등을 통한 타인 비방과, 2) 이러한 내용을 포함한 멘션을 악의적인 의도로 다른 이용자와 주고받는 것으로 나눌 수 있다. 후자에게 악의적 이용(Malicious Usage)의 범위를 확대할 경우 악의적 상호작용이 발생하는 맥락 정보(Context)를 고려할 수 있다는 장점이 있다. 본 연구에서는 전문가의 적합성 판단(Relevance Judgement)을 통해 가능한 모든 악의적 이용의 사례 및 자질을 추출하려고 하였다. 하지만 수집된 실험문헌 중 악의적 트윗의 내용을 수작업으로 분석해본 결과, 악의적인 이용만을 위해 트위터를 사용하는 이용자는 존재하지 않았다. 따라서 이용자 프로파일은 악의적 이용의 분류를 위한 대상에서 제외하였다. 또한 이용자 간 악의적 이용의 상호작용 행태 및 그 맥락도 거의 발견되지 않았다. ‘허위 사실 유포’의 경우, 허위 사실이라는 판단에 연구자의 주관이 개입될 수 있기 때문에 본 연구의 악의적 이용의 대상에서는 제외하였다.

따라서 본 연구에서는 악의적 자질을 타인의 비방과 인신공격에 사용된 ‘비속어’에 한정하였다. 또한 ‘비속어’를 중심으로 발생하는 악의적 이용의 사례를 전문가의 적합성 판단을 통해 검증하였으며, 후속 연구자가 이용 가능한 학습문헌세트를 구축하였다(부록 참조). 한편, 악의적 이용 자동분류를 위한 분류실험에서는 위의 과정을 통해 질적으로 확인된 악의적 어휘를 포함한 사전을 구축하여, 지도학습의 학습자질로 이용하였다.

3.3 실험설계

본 연구에서는 실험문헌 수집 과정을 통해 판별된 각각 500개의 악의적 트윗과 비악의적 트윗을 이용해 지도 학습을 이용하여 악의적 이용 자동분류에 관한 실험을 수행하였다.

각각의 실험에는 최대 엔트로피(Maximum Entropy), 의사결정트리(Decision Tree), 나이브 베이즈(Naive Bayes) 분류 알고리즘이 사용되었으며, 10 묶음 교차 검증법(10 Cross-validation)으로 지도학습을 실시하였다.

한편 분류 알고리즘은 자바 기반의 오픈소스 툴킷인 MALLET을 이용하였으며, 자질 추출을 위해 Lucene Korean Analyzer 3.5 형태소 분석기를 사용하였다. 각 분류기의 성능은 분류 정확률(Precision)과 재현율(Recall), 그리고 F1 척도로 평가하였다.

악의적 이용 자동분류를 위한 실험과정은 기계학습으로 색인어를 추출하는 ‘베이스라인 실험’과 본 연구를 위해 구축한 사전을 이용하여 악의적 자질을 식별하는 ‘규칙기반 실험’으로 나뉜다.

3.3.1 분류 및 학습 알고리즘

(1) 최대 엔트로피 분류기

최대 엔트로피는 주어진 정보를 바탕으로 가능한 최소로 편향된(biased) 추정을 제공하는 범용 기계 학습 기술이다. 즉, 최대 엔트로피는 누락된 정보들을 최대한 이용하지 않으며 (Jaynes 1957), 자질 간의 조건부 독립을 가정하지 않는다. 최대 엔트로피 알고리즘에 의해 입력된 문헌 D가 범주 C에 배정될 확률 추정 값 P(C|D)는 다음의 공식에 의해 계산된다.

$$P(C|D) = \frac{1}{Z(D)} \exp\left(\sum_i (\lambda_{i,c} F_{i,c}(D,C))\right)$$

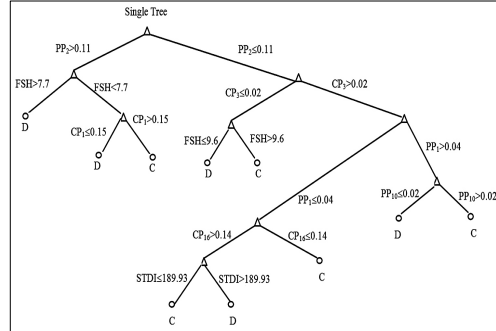
(2) 나이브베이즈 분류기

나이브베이즈는 베이지언 확률과 자질들의 확률은 서로 독립적이라는 가정에 의해 문서를 분류하는 알고리즘이다(Manning et al. 2008). 문헌집단 내에서 출현한 개별문헌 F_n이 범주 C에 속할 확률은 다음의 공식에 의해 계산되어 각각의 범주에 배정된다.

$$P(C|F_1, \dots, F_n) = \frac{1}{Z} P(C) \prod_{i=1}^n P(F_i|C)$$

(3) 의사결정트리 분류기

의사결정트리는 입력된 문헌을 하향식(top-down) 프로세스를 거치며 가장 유사한 문헌집합에 배정하는 분류 알고리즘이다(Yuan and Shaw 1995). 다음은 의사결정트리 알고리즘에 의해 입력된 문헌이 범주에 배정되는 과정의 예를 보여준다.



(4) 교차 검증법

10 묶음 교차 검증법은 전체 문헌의 크기를 10으로 나누어 9/10은 학습 문헌으로, 1/10은 실험 문헌으로 배정하고 각 문헌 집단에 배정되는 문헌들을 바꾸어가며 앞의 과정을 10번 반복하는 학습 방법이다. 교차 검증법을 위해 사용되는 공식은 아래와 같다.

$$\sum_i (Y_i - a - b_1 X_{i1} - \dots - b_p X_{ip})^2 / n$$

3.3.2 자동분류 실험

(1) 베이스라인 자동분류

베이스라인 실험에서는 악의적 자질 식별을 위한 별도의 과정 없이, 입력된 전체 문헌이 기계학습을 통해 사전 분류체계, 즉 악의적 트윗과 비악의적 트윗이라는 주제범주에 임의(random)로 배정된다.

(2) 규칙기반 자동분류

악의적 이용 분류를 위한 규칙기반 실험에서는 악의적 자질 식별을 위해 새롭게 구축한 사전을 이용하였다. 그 이유는 한국어의 교착어 특성으로 인한 웹용어의 다양한 변화를 탐지하

기 위해서는 새로운 어휘사전이 필요하기 때문이다. 본 연구에서는 실험을 위해 수집한 학습 데이터 중 악의적 트윗에 출현한 최신 비속어들과 이 비속어들의 접두·접미사 및 교착 패턴을 분석하여 자질 식별을 위한 사전에 추가하는 방법을 통해, '악의적 자질 사전(Malicious Feature Dictionary)'을 구축하였다. 비속어들의 접두·접미사 및 교착 패턴을 정성적으로 분석한 결과 최신 웹 비속어들은 다음의 접두·접미사와 교착하는 것으로 나타났다(〈표 3〉 참조).

〈표 3〉 웹 비속어 접두·접미사 예시

구분	종류	예시
접두사	꼴, 돌, 병 등	꼴데, 병립픽
접미사	퀴, 레기, 빠 등	소퀴, 효레기

또한 현재 웹서비스 및 온라인게임에서 사용되고 있는 기존의 XML 비속어 필터에서 자질을 추출하여 기존의 악의적 자질 식별을 위한 '웹서비스/온라인게임 비속어 사전'도 구축하였다.

한편 실험문헌의 분석 결과, 비악의적 이용의 특성이 뚜렷하게 나타나지 않았다. 따라서 존칭형 어미의 출현빈도가 높은 문헌은 악의적 트윗일 확률이 낮을 것이라는 가정 하에 존칭 어미의 비악의적 트윗 식별력 검증을 위해 '존칭 어미 사전'을 구축하였다.

다음의 〈표 4〉는 본 연구에서 규칙기반 자동 분류를 위해 사용된 용어사전의 종류와 자질의 예를 나타낸다.

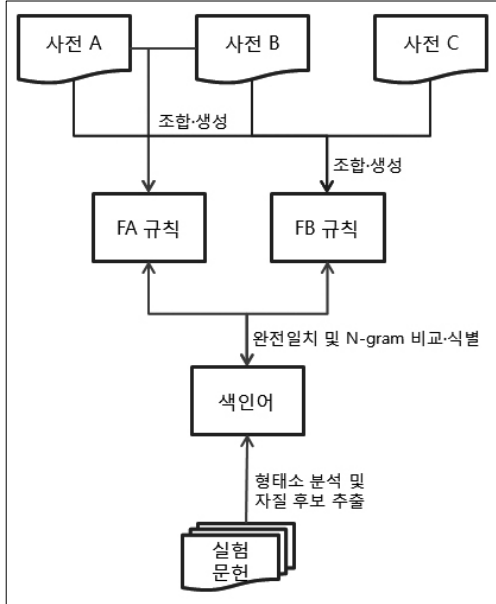
규칙기반의 악의적 이용 자동분류는, 사전의 조합 방식에 따라 (A)+(B) 사전을 사용한 'FA

규칙기반 실험'과 (A)+(B)+(C) 사전을 기반으로 한 'FB 규칙기반 실험'으로 나뉜다.

〈표 4〉 용어사전과 예시

사전 종류	예시
악의적 자질 사전(A)	개같은, 개깨끼, 개잡대, 개비오, 개빠따
웹서비스/온라인게임 비속어 사전(B)	미이친논, 미쳤냐, 미친, 밀구녕, 바이브레이터
존칭형 어미 사전(C)	시다, 드리다, 시오, 해요, 하오

규칙기반 분류실험에서 입력된 실험문헌은 형태소 분석을 통해 자질 후보를 생성하며, 추출된 후보 단어들은 FA 또는 FB 규칙과 비교되는 과정을 거친다. 각 규칙과 비교 시 사전에 등재된 어휘와 완벽히 일치(exact match)하는 경우에는 해당단어를 분류자질로 선정하며, 그 외의 형태소 판별이 불가능한 어절들은 1-gram 과 2-gram 기법에 의해 악의적 접두사 또는 접미사를 포함한 경우 악의적 자질로 식별된다. 선정된 자질들에게는 출현 빈도에 따라 단순 tf 용어 가중치를 부여 하였으며, 140자라는 트윗 길이의 특성을 고려하여 역문헌 빈도 등의 정규화 과정은 거치지 않았다. 이러한 과정을 통해 입력된 전체 문헌은 기계학습을 통해 사전 분류체계, 즉 악의적 트윗과 비악의적 트윗이라는 주제범주에 자동으로 배정된다. 다음의 〈그림 2〉는 규칙기반 실험의 학습자질 선정 과정을 나타낸다.



<그림 2> 규칙기반 실험 학습자질 선정 과정

3.4 실험평가

베이스라인과 규칙기반 실험의 성능은 각각 분류 정확률, 분류 재현율, 그리고 F₁ 척도에 의해 평가되었다. 각각의 평가 방법은 다음의 <표 5>와 같다.

<표 5> 성능평가 방법

평가	구분	
	옳은 범주	틀린 범주
범주에 분류	a	b
범주에 비분류	c	d

※ 정확률(P) = a/(a+b)
 재현율(R) = a/(a+c)
 F₁ = 2PR/(P+R)

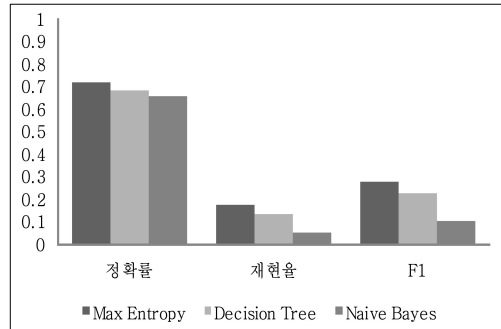
본 연구에서는 실험결과 해석의 일관성을 도모하기 위해 악의적 트윗 범주만을 분류성능 평가의 기준으로 제시하였다.

4. 실험결과 분석

4.1 베이스라인 실험결과

악의적 이용 분류를 위한 베이스라인 실험의 결과는 다음의 <그림 3> 및 <표 6>과 같다. 분류 정확률과 F₁ 척도 모두 최대 엔트로피가 가장 좋은 성능을 보였으며, 그 다음으로는 의사결정트리, 나이브 베이즈의 순이었다. 정확률의 경우 세 분류기 모두 70% 수준의 양호한 분류 성능을 나타냈다.

한편 분류 재현율과 F₁ 척도는 세 분류기 모두 정확률에 비해 크게 떨어지는 결과를 보였는데, 이는 자동으로 추출된 자질들의 식별력이 떨어져 자동분류 시 비악의적 트윗 범주의 문헌이 악의적 트윗 범주로 배정된 것으로 분석된다.



<그림 3> Baseline 실험결과

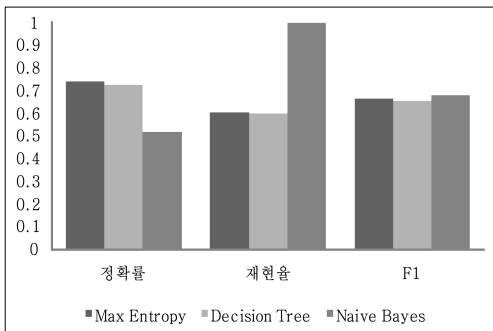
<표 6> Baseline 실험결과

구분	Max Entropy	Decision Tree	Naive Bayes
정확률	0.721	0.682	0.658
재현율	0.175	0.138	0.056
F ₁	0.281	0.230	0.103

4.2 규칙기반 실험결과

4.2.1 FA 규칙기반 자동분류

규칙기반 실험 중 FA 규칙기반 자동분류의 결과는 아래의 <그림 4> 및 <표 7>과 같다.



<그림 4> FA 규칙기반 실험결과

<표 7> FA 규칙기반 실험결과

구분 평가	Max Entropy	Decision Tree	Naive Bayes
정확률	0.742	0.726	0.518
재현율	0.604	0.597	0.998
F ₁	0.666	0.655	0.682

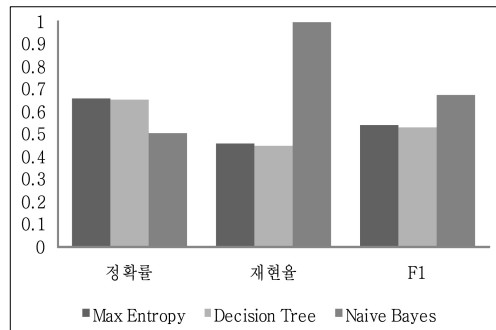
베이스라인 실험과 마찬가지로 분류 정확률에서는 최대 엔트로피가 가장 좋은 성능을 보였다. 그 다음으로는 의사결정트리, 나이브 베이즈의 순이었으며, 나이브 베이즈를 제외한 나머지 분류기들에서 70%를 상회하는 양호한 분류 성능을 나타냈다.

한편 분류 재현율과 F₁ 척도는 세 분류기 모두 70%에 가까운 성능을 보이며 베이스라인 실험에 비해 크게 향상되었으며, 나이브 베이즈는 100%에 가까운 분류 재현율을 보였다. 이는 본

연구를 위해 구축된 용어사전의 악의적 자질 규칙변화 식별력이 높아 분류 재현율이 향상된 것으로 분석된다.

4.2.2 FB 규칙기반 자동분류

규칙기반 실험 중 FB 규칙기반 자동분류의 결과는 다음의 <그림 5> 및 <표 8>과 같다.



<그림 5> FB 규칙기반 실험결과

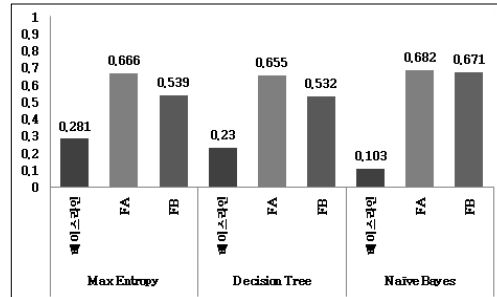
<표 8> FB 규칙기반 실험결과

구분 평가	Max Entropy	Decision Tree	Naive Bayes
정확률	0.656	0.654	0.505
재현율	0.457	0.448	0.996
F ₁	0.539	0.532	0.671

FA 규칙기반 실험과 마찬가지로 분류 정확률에서는 최대 엔트로피가 가장 좋은 성능을 보였다. 그 다음으로는 의사결정트리, 나이브 베이즈의 순이었으며, 나이브 베이즈를 제외한 나머지 분류기들에서 65%를 상회하는 양호한 분류 성능을 나타냈다.

한편 분류 재현율과 F₁ 척도는 세 분류기 모두 베이스라인 실험에 비해 크게 향상되었으나,

분류 정확률과 재현율, F_1 척도 모두 FA 규칙 기반 실험보다는 약간 하락하였다. 이는 FA 규칙기반에서 악의적 트윗으로 분류되었던 문헌들 중 준칭형 어미가 출현한 트윗이 비악의적 트윗의 범주로 배정된 것으로, 본 연구의 비악의적 자질의 정의를 위한 가정을 뒷받침하는 결과로 분석된다.



〈그림 6〉 F_1 척도 성능 종합평가

4.3 종합평가

이상의 결과를 종합하여 보면, 베이스라인과 규칙기반 실험 모두의 경우에서 분류 정확률과 F_1 척도로 평가한 분류기의 분류 성능은 대체로 최대 엔트로피, 의사결정트리, 나이브 베이즈의 순으로 나타났다. 분류 알고리즘 별로 살펴보면, F_1 척도로 평가한 최대 엔트로피 분류기의 FA/FB 규칙기반 실험의 결과는 베이스라인에 비해 각각 137.01%, 91.81%의 성능 향상률을 보였다. 또한 의사결정트리 분류기에서도 FA/FB 규칙기반 실험이 각각 184.78%, 131.30%의 성능 향상률을 보였으며, 나이브 베이즈 분류기의 경우 562.14%와 551.46%로 가장 급격한 성능 향상을 관찰할 수 있었다. 동일 분류 알고리즘 내 모든 FA/FB 규칙기반 실험의 성능 향상률은 유의수준 $\alpha=0.001$ 에서 통계적으로 유의미한 것으로 나타났다. 이를 통해 소셜미디어에서의 효과적인 악의적 이용 자동분류를 위해서는 역동적으로 변화하는 악의적 자질의 실시간 색인과 그 변화 패턴의 예측이 필수적임을 유추해볼 수 있다.

악의적 이용 분류를 위한 실험의 종합결과는 다음의 〈그림 6〉 및 〈표 9〉와 같다.

〈표 9〉 성능 종합평가

분류기	구분	F_1	성능 향상률
Max Entropy	베이스라인	0.281	
	FA	0.666	137.01%*
	FB	0.539	91.81%*
Decision Tree	베이스라인	0.230	
	FA	0.655	184.78%*
	FB	0.532	131.30%*
Naive Bayes	베이스라인	0.103	
	FA	0.682	562.14%*
	FB	0.671	551.46%*

* $p<.001$

5. 결론

웹 2.0의 출현으로 인해 웹서비스 이용자들은 기존의 시혜적인 입장에서 벗어나 정보의 생산과 공유에 능동적으로 참여할 수 있게 되었다. 이러한 현상은 소셜미디어라는 새로운 개념의 커뮤니케이션 채널의 등장을 통해 더욱 가속화되었으며, 실시간으로 생산되는 빅 데이터의 수집·저장·검색·분석은 새로운 국면을 맞이하게 되었다.

한편 소셜미디어의 빠른 속도로 증가하는 이용자와 데이터양에 정비례하여, 권리침해 및 사회질서위반 등의 정보 역기능의 발생도 가속화되고 있는 실정이다. 그리하여 실시간으로 생산되는 방대한 양의 데이터 중에서 유해한 정보를 차단하고, 나아가 건전한 사이버 생태계 구현에 기여할 수 있는 자동화 기법의 고안이 시급한 실정이다.

이에 본 연구는 이러한 소셜미디어의 역기능 중의 일부를 '악의적 이용'으로 정의하고, '악의적 자질'의 분석과 알고리즘의 최소한의 개선을 통해 '악의적 이용'의 자동분류 성능을 향상시키고자 하였다.

그 결과, 실험에서 사용된 모든 분류 알고리즘에서 베이스라인 실험에 비해 '악의적 자질'을 색인하여 구축한 용어 사전을 기반으로 한 규칙기반 자동분류 실험에서 통계적으로 유의미한 성능의 향상을 볼 수 있었다.

이러한 결과로 미루어 볼 때, 교착어의 특성과 실시간으로 쏟아져 나오는 방대한 양의 데이터를 고려하여 악의적 자질 사전의 잦은 갱신과 이를 기반으로 한 패턴 예측이 선행되어야만 효과적인 악의적 이용 자동분류를 달성할 수 있을 것으로 분석된다.

그러나 본 연구는 실험문헌이 다소 적어 악의적 자질 사전에 색인된 비속어가 실시간 데이터 생성의 환경에서 출현하는 방대한 양의 악의적

자질을 탐지하지 못할 수도 있다는 것과 악의적 자질의 언어적 강도를 고려하지 못했다는 한계를 갖는다. 또한 수집된 데이터의 한계로 인해 악의적 이용이 발생하게 된 문맥 등의 비문자적 요인을 고려하지 못하였다.

하지만 본 연구는 악의적 이용의 효과적 분류를 위해 사용할 수 있는 학습데이터의 후보군을 질적으로 확인하였다는 것과 후속 연구자가 이용 가능한 학습문헌을 전문가의 적합성 판단(relevance judgement)을 통해 구축했다는 것에 그 의의가 있다. 또한 이를 통해 실질적으로 자동분류의 성능을 향상시켰으며, 악의적 이용의 분류를 위해 사용할 수 있는 실용적인 학습데이터의 대상을 확인했다는 의미를 갖는다. 또한 소셜미디어 상의 악의적 이용 자동분류를 위한 첫 연구라는 점에서 그 시사하는 바가 크다.

향후 보다 큰 규모의 데이터 세트나 실시간으로 트윗을 수집가능한 환경에서 언어적 강도와 감성 분석이 함께 수반된다면, 더욱 효과적인 악의적 이용 자동분류 기법을 고안할 수 있을 것으로 기대된다. 또한 악의적 이용의 발생과 관련된 정보이용행태 등의 맥락정보 분석과 허위사실 유포의 추적을 위한 인용구문(meme) 마이닝 알고리즘 구현이 함께 수반된다면, 악의적 이용 자동분류 성능의 보다 큰 향상을 기대할 수 있을 것이다.

참 고 문 헌

- [1] 김현정, 윤영미, 이병문. 2011. 향상된 FFP(Feature Frequency Profile)을 활용한 악성 댓글의 판별시스템. 『한국정보기술학회논문지』, 9(1): 207-216.
- [2] 김묘실, 강승식. 2006. SVM을 이용한 악성 댓글 판별 시스템의 설계 및 구현. 『제18회 한국정보과학회 언어공학연구회 학술발표 논문집』, 285-289.
- [3] 박건우, 임선열, 이태양, 김정훈, 최웅철. 2011. 감성 단어 분석 모듈을 이용한 모바일 소셜 네트워크 서비스에서의 성향 판단 시스템 구현. 『한국정보과학회 2011 가을 학술발표논문집』, 38(2): 97-100.
- [4] 박수영, 하용호, 김용혁. 2010. 트위터 정보 검색 분야의 최근 연구들. 『한국정보과학회 학술발표논문집』, 27(2): 25-29.
- [5] 방송통신심의위원회. 『불법·청소년유해정보 개념』. [online]. [cited 2012.11.7].
 <http://www.kocsc.or.kr/01_online/community_Useinfo.php/>.
- [6] 배민영, 차정원. 2008. Topic Signature를 이용한 댓글 분류 시스템. 『한국정보과학회』, 35(12): 774-779.
- [7] 윤미선, 이종혁. 2012. 소셜미디어 이용에 영향을 미치는 요인과 사회자본에 대한 연구: 계획된 행동 이론을 적용한 매개모델 분석. 『방송과 커뮤니케이션』, 13(2): 5-44.
- [8] 이규안. 2012. 사이버범죄의 새로운 양상과 대처방안 연구. 『한국전자통신학회 춘계학술지』, 6(1): 224-226.
- [9] 이원태, 차미영, 양해륜. 2011. 소셜미디어 유력자의 네트워크 특성. 『언론정보연구』, 48(2): 44-79.
- [10] 전희원, 임해창. 2007. 본문과 댓글의 동시출현 자질을 이용한 역 카이 제곱 기반 블로그 댓글 스캠 필터 시스템. 『제19회 한국정보과학회 언어공학연구회 학술발표 논문집』, 122-127.
- [11] 정영미. 2005. 『정보검색연구』. 서울: 구미무역.
- [12] 정유진, 배국진. 2008. 『소셜네트워킹서비스의 동향과 전망』. 서울: 정보통신정책연구원.
- [13] Boyd, D. M., & Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1). [online]. [cited 2012.11.7].
 <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html/>>.
- [14] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.
- [15] Manning, C. D., Raghavan, P., & Schuetze, H. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- [16] IT통계포털. 『정보화/IT활용 역기능』. [online]. [cited 2012.11.7].
 <http://www.itstat.go.kr/stat/graphView.htm?mclass_cd=JD1/>.

- [17] Jaynes, E. T. 1957. Information Theory and Statistical Mechanics. *The Physical Review*, 106(4): 620-630.
- [18] Kooti, F., Yang, H., Cha, M., Gummadi, K.P., & Mason, W. A. 2012. The Emergence of Conventions in Online Social Networks. *6th International AAAI Conference on Weblogs and Social Media*.
- [19] Lee, K., Caverlee, J., & Webb, S. 2010. Uncovering Social Spammers: Social Honey Pots + Machine Learning. *SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 435-442.
- [20] Song, J., Lee S., & Kim, J. 2011. Spam Filtering in Twitter Using Sender-Receiver Relationship. *RAID '11 Proceedings of the 14th international conference on Recent Advances in Intrusion Detection*, 301-317.
- [21] Yuan, Y., & Shaw, M., J. 1995. Induction of Fuzzy Decision Trees. *Fuzzy Sets and Systems*, 69: 125-139.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kim, H., Yoon, Y., & Lee, B. 2011. Prediction System for Abusive Postings using Enhanced FFP. *Journal of Advanced Information Technology and Convergence*, 9(1): 207-216.
- [2] Kim, M. S., & Kang, S. S. 2006. A Design and Implementation of Malicious Web Log Identification System by Using SVM. *The 18th Annual Conference on Human & Cognitive Language Technology*, 285-289.
- [3] Park, K., Lim, S., Lee, T., et al. 2011. Implementation of a Tendency-judging System for Mobile Social Network Services by using the module that Analyzes Emotional Words. *2011 Proceedings of Conference of Korean Institute of Information Scientists and Engineers*, 38(2): 97-100.
- [4] Park, S. Y., Ha, Y. H., & Kim, Y. H. 2010. Recent Studies on Twitter in the Field of Information Retrieval. *Journal of Korean Institute of Information Scientists and Engineers*, 27(2): 25-29.
- [5] Korea Communications Standards Commission. *The Concept of the Illegal Information*. [online]. [cited 2012.11.7]. <http://www.kocsc.or.kr/01_online/community_Useinfo.php/>.
- [6] Bae, M. Y., & Cha, J. W. 2008. Comments Classification System using Topic Signature. *Journal of Korean Institute of Information Scientists and Engineers*, 35(12): 774-779.
- [7] Yoon, M., & Lee, J. 2012. Antecedents of Social Media Use, Actual Use, and Social Capital:

- An Analysis of a Mediation Model. *Broadcasting & Communication*, 13(2): 5-44.
- [8] Lee, K. 2012. A Study on a New Aspect of Cyber Crime and Management Plan. *2012 Proceedings of Conference of Korea Institute of Electronic Communication Science*, 6(1): 224-226.
- [9] Lee, W., Cha, M., & Yang, H. 2011. Network Properties of Social Media Influentials: Focusing on the Korean Twitter Community. *Journal of Communication Research*, 48(2): 44-79.
- [10] Jeon, H. W., & Rim, H. C. 2007. A Comment Spam Filter System based on Inverse Chi-Square Using of Co-occurrence Feature Between Comment and Blog Post. *19th Proceedings of Conference of Korean Institute of Information Scientists and Engineers*, 122-127.
- [11] Chung, Y. M. 2005. Introduction to Information Retrieval. Seoul: Gumi Trading.
- [12] Jung, Y., & Bae, K. 2008. *Trend and Prospect of SNS*. Seoul: KISTI.
- [16] IT Statistics of Korea. *Adverse Effect of Informatization*. [online]. [cited 2012.11.7].
<http://www.itstat.go.kr/stat/graphView.htm?mclass_cd=JD1/>.

[부록] 악의적 이용 자동탐지를 위한 실험 데이터1)

□ 비악의적 트윗

No.	내용
1	RT@#####:어제새벽에와있던문자..자신도드라마와영화촬영하느라더정신없고바쁘텐데..늦은새벽에도이렇게모니터해주고연락주는착한재중이형..그리고자신의드라마인터뷰도는내내 내얘기만 나오면 더 신나게 [URL]
2	@##### ... (·▽·) 고양님도 피튀기는거 잘 보시나봐여.. 전 드라마에서도 총쏘는거 비슷한거 나옴 눈돌려여 못봐..
3	@##### 친구놈들은 축구하고 야구밖에 몰라요 오늘은 다들 더워서 집에 박혀있대요
4	남우현 우는 건 이제 지겹. 눈에 인공눈물 주머니 있나 근데 나도 드라마보면서 매일 쳐우니까 이해함. 근데 좀 작작올어라
5	@##### ππππ... 공연6시30분부터인데... 야구는더늦게끝날려나...
6	진짜 완전 이기적인말인데 난 얼마안된축덕이라서 아직축구에대해 알아가는 과정이지만 괜히 주위애들이 잘생긴 축구선수만 따져대면 짜증난다 근데 나도 얼빠졌다는게 함정
7	RT @##### : 올림픽 대표팀 에제자 향한 소속팀 사령탑들의 애절한 응원메시지! 감독뿐만 아니라 선수들의 개알같은 응원 메시지도 확인하세요~ 우리 선수들, 멋지게 승리하여 K리그에서도 그 활약 쪽 이어갑시다~ 화이팅!!:) [URL]
8	@##### 그림~드라마~영화 그리고 정신 차려보니 두시....
9	RT @##### : 악플로 자살한 연예인들의 악플러들보면 참 한심하고 답없고 왜 그러는지 모르겠고 눈살 찌푸려지지? 지금 니가 하고있는것이 바로 그 짓이야
10	RT @##### : 나는 가수다는 가수의 문제보다는 포맷의 문제가 아닐까. 그렇다면 포맷은 늘 똑같은데 할 때마다 드라마틱한 스포츠경기의 힘은 뭘까.
11	@##### 페이스북공유하기랑좋아요가리트윗이고리트윗하면내빨로워들한테보이는거오ㅋㅋ나도그냥좋아하는연예인들빨로잉해두고올라오는글만눈팅하는거라자세한거는잘몰랑ㅋㅋ
12	올해 가온차트 남자아이돌 솔로 순위~^ [URL]
13	계잡가면사람들이 뺨터드릴필요있는아이돌해서엠블랙얘기하는데보완하지못할망정새그룹데뷔ㅋㅋㅋㅋㅋ멘봉
14	@##### 혹시 세놈이 친하게 지내다 한놈 자살, 한놈은 전학가고.. 자살한 아이 아버지가 친구들 찾아다니면서 이들이 어땠는지 물어보는 영화 본적 있어요? 제목이랑 배우 얼굴은 기억이 안나고 내용만 기억나요 철길에서 야구하고 율미도가고그랬는데
15	전에 모 연예인 깎다가 털린 본계 트친이 생각난다.. 그사람은 아마 슈주를 깎았더랬지.. 2~3일만 지나면 잠잠해져요
16	진짜드라마가그냥만화책같은어이없고말도안되고유치찬란...우리나라드라마가한마디로막장이라면애넌그냥드라마자체가오타쿠야
17	세상에.. 이번 주말! 전민희작가님 사직구장에 시구하러 오신다고.. 소사맘 가야하나ππππππππππππ야구는 잘 안보지만....
18	RT @##### : 스타킹 몸짱킹!!중간점검때 제대로 공개합니다!!기대해주세요!!??*??*몸짱킹의 대장님 이영만트레이너의 선수들 연예인 그리고 일반 출연자들까지!놀라운 일이 펼쳐집니다!!*??*??*coming soon!!!!

1) 본 부록에 수록된 실험 데이터는 실제 실험을 위해 사용된 비악의적 트윗과 악의적 트윗 중 일부를 발췌하였으며, 개인정보 유출의 위험을 고려하여 게재 시간을 삭제하고 이용자의 아이디와 출현한 URL은 각각 '#####'과 '[URL]'로 대체하였음.

No.	내용
19	RT @#####: 축구는 눈물 안 나는데 핸드볼은 계속 눈물나. 오빠들은 그래도 좋은옷 입고 따뜻한밥 먹고 과외 받아가며 공부했는데 여동생은 아무도 안 들여다 봐주니까 혼자 부엌에 쭈그리고 앉아 공부했잖아. 상이라도 받아와야 "그래 잘했네" 하곤 또 부엌으로 보냈잖아
20	RT @#####: 리트윗한 사람들에게 올림픽 축구 한일전, 대한민국이 이기면 오늘 치킨 썸.
21	The Best Curation of K-pop] 기성 가수의 곡을 자신만의 스타일로 재해석한 커버곡! "메이건리"의 2NE1 #curation [URL]
22	내일부터 다시 돌아가야할 일상에 잠이 오지 않는다. 자고 싶을 때 실컷 자고 축구경기 체크해가며 응원다니던 그때로 돌아가고 싶다!!
23	김정민, 시구후눈물흘린사연'충격'-김정민이프로야구시구후눈물을쏟은이유를밝혔다. 김정민은지난9일광주구장에...기사보기More-스포츠한국 김정민, 시구 후 눈물 흘린 사연 ... [URL]
24	RT @#####: 리트윗한 사람들에게 올림픽 축구 한일전, 대한민국이 이기면 오늘 치킨 썸.
25	그로스크로이츠 믿음의 축구 돌네.
26	@##### 근데 난 이제 뿔티도하...ㅋ...ㅋㅋㅋㅋㅋㅋㅋㅋ 올림픽때무네 달린듯? 김박감이 콘섯해야.재입덕할듯요^o^
27	@##### 안온다능 올림픽선수단만온다능
28	@##### 신광훈은 항상 국대급이고 이번에 올림픽도 와일드 카드로 뽑혔어야 하는데리는 생각입니다. 올림픽 못한 환을 A매치에서 풀어줬음 하네요
29	RT @#####: 항상 이런식이죠! 정치적이벤트인 올림픽을 정치적으로 활용하고있는 IOC 에 강하게 항의해야 할 스포츠 관계자들이 박종우선수 개인에게 책임을 돌리는 듯한 이모습.. 잘싸우고 돌아온선수 기를 왜 죽이나요.
30	RT @#####: 올림픽 하면 생각나는 연구 하나. 은메달리스트보다 동메달리스트들이 더 큰 행복을 맞본다고 함. 전자는 비교의 기준이 상향(지금보다 더 나올 수..)인 반면 후자는 비교 기준이 하향(아예 메달 못 딸 뻔..)이기 때문.

□ 악의적 트윗

No.	내용
1	존나 귀찮게지랄이야 봉신퇴물걸스빠순이들 ㅋㅋ 퇴물로검색해서 지내육하나찾아보는주제에
2	솔직히 지금 잠수단 김구라보다 박유천식 이새끼가 더 드러운 새끼야 물론 김구라가 입에 담으면 안되는 말올한건 사실이지 하지만 이미 10년전 일이고 지금 자기 밥줄 다 끊어가며 반성하고 있잖아 ..
3	숨쉬기도 힘들텐데 담배피라 노래하라 춤추라 연기하라 살아있는게 용한 박유천에게 노벨천식상 줘라
4	SM은 이럴때 언플 좀 해라. 니넨 2천 벌금 내야되는데 죄와죄 개객끼들이 허위사실 유포하는 것도 벌금좀 때려야지 못쓰겠다. 봉도시는 옳은 소리해도 잡혀가는데 저색기들도 쳐넣어라. SM 빠수니 입장에선 빠치는 회사지만 배신의 아이콘 죄와죄만 하겠냐?
5	고상하게뚝뚝한척지랄염병을떨고있는죄와죄팬의전형적인글어디소시의오늘무대에자신감이없는소원분들은똥쳐헨썹해보세요[URL]
6	@##### #쩍긴뿔쩍어이개같은쩍년이[URL]이걸쳐봐이쌍년아샤이니에대해서조금만아는인간이면감히최민호쌍수드립은안친다이정신간죄와죄빠순년아니들이그러니까육을쳐먹는거야개년들
7	아씨발동방신기랑죄와죄랑위지말라고씨발새끼들아.눈빼꾸냐?아씨발존나기분좆갈네.워지좀마제발[URL]
8	지랄하네저좆같은년이ㅋ죽이고싶다^^"@#####:비스트겔모차르트반응이궁금해서염탐했는덱ㅋㅋㅋㅋ망상타절ㅋㅋㅋㅋㅋㅋㅋㅋ죄와죄좆바른덱ㅋㅋㅋㅋ[URL]"
9	군대로겨져 강나니는 또 왜나와 ㅋㅋㅋㅋㅋㅋㅋㅋ 좆스심장나와서 눈물토크같은거하면 레알 양심 o벗는거임
10	사사구때문에 풀리건들 분노하는거봐라ㅋㅋㅋㅋㅋㅋ 맞는말 했구만ㅋㅋㅋㅋㅋㅋ 풀스타전에 자부심느끼는거보면 중국인들이랑 다를게 뭘냐. 풀테가 좆크보며여살리는데 왜좆주만도 관중수가 못하나요ㅋㅋㅋ

