지능정보연구 제18권 제1호 2012년 3월(pp.91~107)

Naïve Bayes 방법론을 이용한 개인정보 분류

김남원
서울대학교 일반대학원 경영학과
(telem0909@gmail.com)

박진수
서울대학교 경영전문대학원
(jinsoo@snu.ac.kr)

인터넷의 성장과 개인의 참여는 사생활 정보 보호에 관련된 비효율적 관리 방안에 대한 문제의식을 불러일으키고 있으며 이를 해결하기 위한 여러 연구들이 이루어지고 있다. 본 연구에서는 기존에 존재하는 문서 분류 방법론을 이용하여 개인의 사적 공간을 나타내는 프라이버시의 항목 중 개인을 식별할 수 있거나 개인이 민감해 할 수 있는 사생활 정보를 담고 있는 문서를 탐지 혹은 분류하는 방법에 대해서 다룬다. 논문의 실험에서 기존의 학습데이터에 추가적으로 개인정보의 유형에 관련된 하위 학습 데이터를 추가함으로써 자동 문서 분류 알고리즘의 성능 측정치를 높이는 것을 시도하였다. 또한 개인정보의 유형에 따라 알고리즘에 효과적으로 적용하는 방향을 제시하기 위하여 기존 논문에서 나타난 개인정보의 유형들을 분석하였다. 개인정보 관련 문서로 분류된 학습 대상과 함께 개인정보에 영향력이 있는 개인정보 유형들을 추가 학습시켜 알고리즘이 학습하는 문서 자질(feature)의 질(quality)을 높였다. 높아진 학습 자질의 질로 인하여 기존의 Naïve Bayes 방법론을 이용한 평가 측정치가 높아질 수 있었다.

논문접수일 : 2012년 03월 10일 게재확정일 : 2012년 03월 18일
투고유형 : 국문일반 교신저자 : 박진수

1. 서론

현대에 정보 시스템 활용이 증가하면서 개인들의 온라인 상 활동이 활발해지고 생활의 경제성과 편의성이 증가하였다. 반면 이러한 활동들은 사생활 정보가 노출되어 여러 피해 사례와 논쟁을 야기시킴으로써 개인적, 사회적인 낭비를 낳아 중요한 이슈가 되고 있다(Mason, 1986; Gross and Acquisti, 2005). 인터넷의 성장과 개인의 참여는 사생활 정보 보호에 관련된 비효율적 관리 방안에 대한 문제의식과 이를 해결하기 위한 도전을 환기시킨다(Kobsa, 2002; Boyd and Ellison, 2008).

최근 10년 사이 인터넷이 대중화되면서 많은 사람들이 온라인 활동에 참여하기 시작하였다. 특히, 최근 들어 널리 퍼진 소셜 네트워크 서비스(SNS)는 개인의 근황 및 소식에 대한 의사소통을 원활하게 해줌으로써 인간관계를 유지하는데 큰 역할을 하고 있다(Lampe and Ellison, 2008; Tong et al., 2008). 기업들은 SNS의 고객 정보를 이용하여 고객에 맞는 서비스를 제시하고 홍보하는 활동을 하고 있다. 개인과 기업적 측면에서 인터넷 활동이 필요하다는 것을 부인할 수는 없을 것이다.

하지만 SNS의 활용의 결과로 떠도는 정보들은 사생활 피해라는 부정적인 측면을 초래하고 있다.

* 본 연구는 서울대학교 경영연구소의 연구비 지원에 의해 수행되었음.
본 연구는 서울대학교 경영대학 정보통신경영연구센터의 연구비 지원에 의해 수행되었음.

개인의 정보가 의도치 않게 타인에 의해 악용되어 개인의 이미지 손실, 금전적 피해, 피싱(phishing)의 사례들을 발생시키고 있다(Jagatic et al., 2007). 개인정보의 유출에 대한 책임은 서비스를 제공하는 업체도 종종 짊어지며 해당 업체에 관련된 이미지 손실과 다양한 재정적 손실을 주게 된다. 사생활 정보 유출로 인한 개인과 조직의 손실들은 나아가 모두 국가 차원의 손해를 일으킨다.

사생활 침해에 대한 많은 보고에 따르면 이는 외부로부터의 침범뿐만이 아니라 개인들에 의해 공개적으로 게시된 개인정보를 우연히 혹은 악의적으로 이용하기 때문이다(LoPucki, 2001; Solove, 2006). 최근 스마트폰의 인기와 더불어 트위터(Twitter)나 페이스북(Facebook) 같은 SNS가 활기를 띠면서 웹 포스팅과 블로깅을 통한 개인정보 유출이 더욱 심각하게 우려되고 있다(Hinduja and Patchin, 2008; Livingstone, 2008; Meeder et al., 2010).

본 논문은 사생활 정보 침해가 없는 바람직한 인터넷 생태계를 만들기 위해서는 SNS의 활동을 억제하기 보다는 정보 유출 가능성에 대해 감독하는 시스템이 필요하다는 점에서 착안되었다. 데이터의 내용 분석을 통하여 사생활 정보를 담고 있는 인터넷 문서를 탐지하여 그 위험성을 이용자들에게 알리고 추후에 있을 수 있는 개인정보 유출을 미연에 방지하는데 그 목적이 있다.

본 연구에서는 기존에 존재하는 분서 분류 방법론을 이용하여 개인의 사적 공간을 나타내는 프라이버시의 항목 중 개인을 식별할 수 있거나 개인이 민감해 할 수 있는 사생활 정보를 담고 있는 문서를 탐지 혹은 분류하는 방법에 대해서 다룬다. 여러 도메인에서 이미 충분히 입증된 Naïve Bayes 문서 분류 방법(Bayes, 1763; Mitchell, 2010)을 개인정보 영역에 적용하여 그 실효성을 살펴본다. 또한 분석된 결과에 포함된 개인정보를 유형별로 다차원적 분석을 함으로써 향후 연구를 위한 방향을 제시한다.

Naïve Bayes 방법론을 비롯한 기계학습 방법론에서는 훈련 데이터를 학습 시킨 후 이를 실험 데이터에 적용하여 문서 분류기의 성능을 평가한다. 방법론을 적용하기 위한 대상으로 개인정보 유출 위협이 가장 큰 인터넷 문서인 소셜 네트워크 서비스(SNS)를 선택하였으며 여러 소셜 네트워크 서비스 가운데 트위터(Twitter)를 통해 일반 대중들에게 모두 공개되고 있는 데이터를 무작위로 수집하였다. 수집한 데이터는 개인정보에 민감한 정보를 담고 있는지에 따라 학습을 위한 두 분류(privacy, non-privacy)로 나뉜다. 기계학습 문서 분류기는 분류된 훈련 데이터를 초기 학습하고 실험 대상이 되는 문서들을 알고리즘으로 분류한다. 그리고 실험으로 나타난 결과를 통해 개인정보 영역에서 더 효과가 높은 알고리즘을 선택하여 추가 실험을 진행한다.

개인정보의 종류별로 개인정보 관련 문서에 어느 정도의 변별력 있게 나타나는지 알기 위해서는 개인정보에 해당하는 자질들을 분석할 필요가 있다. 개인정보에 대한 연구는 19세기 말부터 시작하여 꾸준히 연구되어 왔으며 많은 연구가들이 개인정보의 정의와 영역, 그리고 유형들에 대해서 소개하였다. 본 논문에서는 보다 정확한 개인정보의 자질 분석을 위하여 기존 문헌들을 종합하여 자질들의 목록을 도출하였다. 실질적으로 학습 알고리즘에서 쓰일 수 있는 개인정보 유형을 조사하여 개인정보 관련 문서가 가지는 자질의 질을 높인다. 그리고 개인정보 관련 문서의 자질로서 나타난 개인정보 유형을 학습데이터에 추가시킨 후 자동 문서 분류기의 성능을 측정한다.

본 논문은 개인정보 분야에서 문서를 자동 분류

하는 방법에 대해 고찰하고 문서 자질의 분석을 통해 개인정보 관련 문서 분류의 성능을 높일 수 있는 방법을 제안한다는 점에서 의의가 있다. 논문의 실험은 개인정보와 같은 특정 분야의 문서를 분류하기 위해서는 선입견을 제외한 상태에서 분류 자질 자체에 대해 깊이 숙고해야 한다는 점을 제시한다. 개인정보 관련 문서의 분류 방법에 대한 연구는 미래에 개인정보 누출로 인해 발생할 수 있는 개인적, 조직적, 그리고 국가적인 피해를 미연에 방지함으로써 우리가 입을 수 있는 손실을 줄일 수 있을 것이라고 기대한다.

2. 관련 연구

2.1 프라이버시와 개인정보

'프라이버시(Privacy)'의 본래 어원은 '사람의 눈을 피한다'라는 뜻의 라틴어 Privatue에서 유래한다. 프라이버시의 개념이 처음 언급된 것은 Thomas Cooley 판사가 1888년 출간된 그의 저서(Cooley, 1888)에서 프라이버시를 홀로 있을 수 있는 권리(right to be alone)라 표현된 데에서 비롯한다. 그 후 2년 뒤 Warren과 Brandeis(1890)의 '프라이버시 권리(The Right to Privacy)' 논문에서 프라이버시에 대한 개념을 '홀로 있을 수 있는 개인의 일반적인 권리(the more general right of individual to be let alone)'라고 규정하였다.

현대 사회가 고도로 발전되면서 조직 및 온라인 상에서 사람들과 관계를 맺으면서 프라이버시 문제는 점점 복잡해지고 다양해지고 이에 다양한 정의의 프라이버시가 언급되었다. 프라이버시 분야의 권위자 중 한 사람이었던 프라이버시 저널(Privacy Journal)의 편집장인 Smith(2000)는 프라이버시를 "방해, 침해, 당황, 책임이 없는 우리들 각자를 위한 물리적 공간에 대한 소망과 우리 자신에 관한 개인정보 공개의 시기와 방법을 통제하려는 시도"[1] 라고 현대에 맞게 새롭게 정의하였다. 프라이버시에 관련된 법적, 사회적, 윤리적 이슈를 다루는 국제단체인 프라이버시 인터내셔널[2](Privacy International)의 대표인 Davies(1996)는 프라이버시를 정보, 신체, 통신, 공간 프라이버시로 나누었다. 이 중 정보 프라이버시는 신용정보, 의료, 정부 기록과 같은 개인정보를 모으거나 이용하는 것을 관리하는 규율의 작성과 연관된 것이라고 정의하고 있다. Clarke(2000)는 프라이버시를 개인의 프라이버시, 개인 행위 프라이버시, 개인 통신 연락 프라이버시, 개인정보 프라이버시로 나누면서 개인정보에 대한 중요성을 역설하였다.

인터넷 상에서의 프라이버시는 앞에서 언급한 모든 프라이버시를 다 포함하진 않는다. Frederic(1998)은 인터넷 상의 프라이버시를 개인정보와 거의 동일시하고 있다. 개인정보가 프라이버시와 동일시되는 경향까지 나타나는 것으로 보아 이는 개인정보가 그만큼 프라이버시에서 상당히 중요한 비중을 차지하고 있음을 나타내는 것이다(한국정보보호진흥원, 2007).

개인이 여러 활동을 통해서 전달되는 많은 정보들은 정리되어 기록으로 저장되게 된다. 따라서 프라이버시 혹은 사적 공간은 다른 사람이 참고할 수 있는 정보의 형태를 취하게 될 때 개인정보로 바뀐다고 할 수 있다. 즉, 개인정보는 프라이버시의 발현 형태이며, 프라이버시가 정보라는 외형으로

1) the desire by each of us for physical space where we can be free of interruption, intrusion, embarrassment, or accountability and the attempt to control the time and manner of disclosure of personal information about ourselves.

2) 프라이버시와 개인정보 보호에 관련된 사항을 다루는 인권단체로 영국에 자리하고 있다.

나타남으로써 취급 및 관리가 가능하게 되는 것이다(황인호, 2002). 다양한 개인정보들은 디지털화되어 데이터베이스에 저장되고 이 데이터베이스가 각종 네트워크와 연결됨으로써 자료의 검색과 보존은 용이해졌지만 침해 위험성은 그만큼 더 높아졌다. 이러한 디지털 형태의 개인정보가 지닌 취약성과 정보화 시대에 개인정보가 지닌 사회적, 경제적 가치로 인해 오늘날 개인정보보호는 프라이버시 보호에서 가장 핵심적인 부분으로 이해된다(이창범, 조정현, 2003).

만약 개인이 공적으로 개인정보를 노출했을지라도 이것을 타인의 의도에 의해 다른 곳에 저장 혹은 전달한다면 이는 인간의 기본권인 개인정보 자기결정권에 대한 위반 행위이다(권건보, 2005). 소셜 네트워크 서비스에서 공적 장소에 개인정보가 노출되었더라도 이 정보에 대한 사용 권한은 엄연히 그 정보에 관련된 개인에게 있다. 많은 사람들은 타인의 정보를 함부로 이용하는 것이 개인정보 자기결정권이라는 기본 인권을 해친다는 사실을 잘 알지 못한다(Wolak et al., 2010). 이는 인터넷이 발전하는 속도에 비해 우리가 인터넷 상에서 갖추어야 할 윤리적 자세에 대한 교육과 제도가 제대로 준비되어 있지 않기 때문이라고 분석된다. 아직 윤리적으로 준비되지 않은 인터넷 이용자들을 대신하여 개인정보에 침해가 될 만한 요소들을 미연에 탐지하여 물질적, 정신적, 사회적, 법적인 피해를 최대한 줄일 수 있는 방안이 필요하다.

2.2 개인정보의 범위

세계 각국의 개인정보에 관한 법률상의 정의는 공통으로 “개인에 관한 정보”라는 점에서는 그 의의를 같이 한다(한국정보보호진흥원, 2007). 하지만 개인정보의 구체적인 범위는 실제 연구자나 사회적 규범에 따라 매우 다양한 요소들을 포함할 수 있다. 개인의 신분을 식별할 수 있는 정보까지를 개인정보로 보는 경우부터 민감한 정보만을 개인정보로 취급하는 경우(Parent, 1995)까지 개인정보의 범위를 분석하는 시각은 다양하다. Wacks (1993)의 경우에는 정리한 개인의 건강상태, 신체적 특징, 사상이나 신념과 같은 정신세계, 학력, 경력, 재산상태, 사회적, 경제적 지위 등 개인에 관한 사실, 판단, 평가를 나타내는 모든 정보를 개인정보로 파악하고 있다. 본 논문에서는 Wacks가 제시한 기준과 유사한 개인정보의 범위를 다룬 기존 문헌들을 조사하였다.

Weible(1993)는 자신의 박사 졸업 논문에서 개인정보의 종류에 대하여 정리를 하였고 윤상오(2009)의 논문에서는 Weible의 생각을 일반, 가족, 교육 및 훈련, 병역, 부동산, 동산, 소득, 기타 수익, 신용, 고용, 법적, 의료의 범주로 재정의하였다. 우리나라의 개인정보보호법 및 표준개인정보 보호지침에서는 개인정보를 내면의 비밀, 심신의 상태, 사회경력의 5개의 항목으로 정리하였다. 영국의 정보 위원회(Information Commissioner’s Office, 2001)에서는 정보보호법 시행과 함께 7가지 개인정보 항목을 정리하였다. 조동기와 김성우(2003)의 경우 위와 같은 일반적인 구분뿐 아니라 각 범주를 정태적 · 기술적 정보와 동태적 · 추론적 정보로 나누어 2차원적 구분이 되게 하였다. 이강신 외(2010)는 개인정보를 더 포괄적인 범주로 정리하여 개인정보의 특징을 분류하였다. 크게는 인적, 신체적, 정신적, 재산적, 사회적, 기타 요소 6가지로 나누었고 세부적으로는 재산적 정보를 개인금융정보와 신용정보로, 사회적 정보를 교육정보, 법적 정보, 근로정보, 그리고 병역정보로 나누어 구별하였다.

이처럼 개인정보의 유형을 분류하는 범주는 국

가와 기관, 그리고 연구자마다 각양각색이다. 하지만 구체적인 유형을 살펴보면 어느 정도 일치하는 분류가 있으며 또한 많은 예시들이 공통으로 개인정보 범주에 포함되고 있다.

3. 연구 방법론

3.1 Naïve Bayes 분류 방법

문서를 분류한다는 것은 문서가 관련성이 높다고 생각되는 하나 혹은 다수의 범주(category)에 문서를 대응시키는 것이다. 자동 문서 분류기는 훈련 문서 데이터의 자질(feature)들을 분석하여 학습한다. 그리고 학습한 정보를 이용하여 새로운 문서가 어떤 범주에 속할 지에 대하여 예측, 판단해 준다.

Naïve Bayes 분류법(Bayes, 1763)은 문서를 분류하는 방법에 있어서 가장 많이 쓰이는 알고리즘 중 하나이다. 이 방법론은 카테고리 안의 클래스(class)들을 정의하고 문서가 가지는 자질(feature)에 따라 특정 범주에 속할 확률을 계산한다. 그리고 가장 높은 확률을 가지는 범주로 문서를 분류한다. Naïve Bayes를 표현하는 기본 공식은 다음과 같다.

$$P(C|F) = \frac{P(F) \times P(F|C)}{P(C)}$$

C는 카테고리에 속한 클래스를 나타내며 F는 문서가 가지고 있는 자질을 나타낸다. P(C|F)는 문서가 F의 자질을 가지고 있을 때 C클래스에 포함될 가능성을 말한다. P(C|F)는 Naïve Bayes에서 결과적으로 구하고자 하는 목적 값이지만 직접 유추할 수 없으므로 $\frac{P(F) \times P(F|C)}{P(C)}$의 수식으로 변형시켜 값을 구한다. P(F)는 전체 문서를 대상으로 자질이 나타날 확률로서 모든 클래스에 대해 같은 값을 가지게 된다. P(C)는 전체 문서 중 C클래스에 속하는 비율이므로 C클래스에 속하는 문서들의 수를 전체 문서의 수로 나눈 값이 된다. P(F|C)는 C클래스에서 F자질이 나타날 확률이며 문서들 중 F자질이 나타나는 문서의 수를 전체 문서의 수로 나눈 값이 된다.

Naïve Bayes 알고리즘은 문서가 가지는 모든 자질 혹은 단어들은 서로 독립적이라고 가정한다. 이런 가정은 실제 조건과는 다르지만 Naïve Bayes는 기존 실험결과에서 매우 뛰어난 결과를 보여주고 있다.

Naïve Bayes 문서분류 알고리즘은 여타 선형 분류법, 의사결정트리 등의 다른 알고리즘보다 일반적으로 정확도가 높은 편이다(Domingos and Pazzani, 1996). 또한 Naïve Bayes 방법론은 소수의 데이터를 이용한 정교한 방법의 기법보다 학습 데이터 양이 많을 때 그 성능이 더 높아진다(Manning et al., 2008). 학습하는 시간은 다소 오래 걸릴 수 있지만, 시스템의 구축이 비교적 어렵지 않은 편이며 문서 분류의 속도가 매우 빠르다는 장점이 있다.

3.2 추가 자질 선택(Feature Selection)

Lewis(1992)는 그의 박사학위 논문 “Representation and Learning in Information Retrieval”에서 문서 분류기의 성능 향상을 할 수 있는 방법으로 기계학습 과정과 문서 자질의 개선을 제안한다. 이러한 자질들의 질을 향상 시킬 수 있는 방법에는 학습 과정에서 제대로 된 자질들을 선택하는 것(feature selection)과 구분된 문서에서 자질들을 잘 뽑아내는 것(feature extraction)의 두 가지 방법이 있다. 이 중 학습 과정에서 제대로 된 자질들을

선택하는 것은 기존에 존재하는 학습 대상에 어떤 하위 집합을 추가 설정함으로써 자질들의 전체 질을 향상시키는 것이다(Lewis, 1992).

특정 범주의 문서가 가지는 자질의 질을 향상시키는 방법에는 확실히 연관될 수 있는 작은 자질 집합을 학습 대상의 크기에 맞추어서 추가로 학습하는 방법이 있다. 그리고 이러한 자질들은 전통적으로 범주에서 특히 매우 빈번하게 나타나는 요소들로 구성된다. 많은 경우에 학자들이 이러한 자질들을 선택할 때 사람의 손을 거쳐서(manually) 생성되는 과정에 의존한다(Lewis, 1992).

Lewis는 학습 알고리즘을 시행할 때 우리가 언어적 편견에 빠져서 학습할 대상의 자질을 잘못 분석하는 경우가 많다고 말한다. 자질들의 질(quality)을 떨어뜨리는 경우에는 자질들이 실제로 개체(instance)를 구분할 수 없거나 알고리즘의 가정들을 해치는 경우가 있다. 또한 중요한 개념이 자질에서 빠지게 되거나 너무 많은 종류의 개념들을 자질에 포함시켜 자질들의 질을 떨어뜨릴 수 있다.

문서가 가지는 자질의 질을 향상 시킬 수 있는 집단들을 결정할 때에는 실제 이 집단 내의 자질들이 문서 분류에 있어서 어느 정도 영향력이 있는지를 판단해야 한다. 가장 일반적으로 많이 쓰이는 자질 선택(feature selection) 방법론으로는 Mutual Information 방법이 있다. Mutual Information은 변수 간 독립성의 척도를 나타내는데 다른 자질 선택 방법론보다 노이즈로 인한 영향이 적어 신뢰도가 더 높다(Peng et al., 2005).

4. 실험 디자인 및 구현

4.1 실험대상 : 트위터(Twitter)

트위터는 사람 간에 관계를 맺고 개인적, 사회적인 소식을 주고 받는 소셜 미디어적 성격의 SNS 서비스로 2006년 8월에 시작하였다. 하나의 단문 메시지 트윗(tweet)은 140글자 내외로 작성할 수 있으며 타임라인(Timeline) 화면에 출력된다. 마이크로 블로그는 굉장히 가벼운 단위의 정보가 움직이므로 컴퓨터 뿐 아니라 스마트폰 등을 이용하여 쉽게 이용할 수 있다.

일반적으로 사람들이 트위터에서 작성하는 메시지는 인터넷 이용자가 모두 타임라인 화면에 게시되는 공개적 글이 된다. 공개적인 글에는 모든 사람들에게 알리기 위한 메시지에서부터 특정인을 지목해서 보내는 메시지가 있다. 상대방만 볼 수 있도록 메시지를 보내고 싶으면 쪽지 기능과 유사한 Direct Message(DM)을 이용할 수 있다.

다른 소셜네트워크 서비스와는 달리 트위터는 단문 메시지 중심이며 소셜 미디어적인 성격이 굉장히 강하기 때문에 이용자들은 즉흥적으로 글을 올리는 경향이 더 심화된다. 한국인터넷진흥원은 2011년 1월에 4일 동안 트위터 ID 200개를 대상으로 이름, 위치정보, 구체적인 인맥 정보 등 개인을 파악할 수 있는 34개 항목을 스마트폰 또는 인터넷을 통해 수집 가능한지에 대해 조사하여, 이름(88%), 인맥정보(86%), 사진 등 외모정보(84%), 위치정보(83%), 관심분야 등 취미정보(64%), 스케줄 정보(63%), 가족 정보(52%) 등을 조사대상 중 절반 이상에서 쉽게 파악할 수 있었으며 의료정보(29%), 정치성향 정보(19%) 등 민감 정보로 분류되는 정보도 상당히 높은 수치를 보인 것으로 파악되었다(방송통신위원회, 2011). 한국진흥원에서 발표한 위의 통계의 범주 역시 개인정보의 유형들에 속하는 부분들이며 유출 수준에 있어서 상당히 높은 수위를 보여준다.

4.2 문서 데이터 단위와 학습 대상 선정

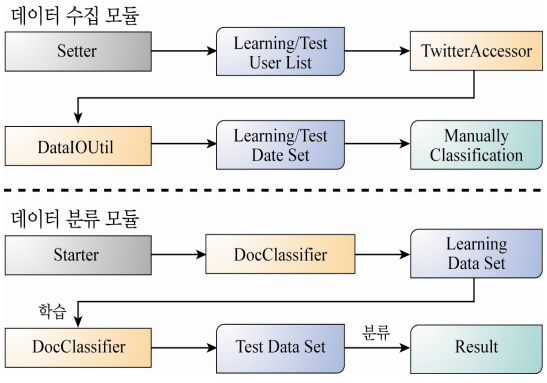
본 논문의 실험에서는 한 명의 트위터 유저 타

임라인에서 보이는 글 200건을 하나의 문서 단위로 취급한다. 학습 대상과 실험 대상이 될 유저는 무작위로 선택하였다. 학습 대상은 전체 500개의 문서이며 이 중 개인정보 범주의 문서가 250개, 비개인정보 범주의 문서가 250개이다.

4.3 프로토타입 구현

프로토타입은 Microsoft Window 7 64bit Version 기반에서 자바 프로그래밍 언어를 사용한 어플리케이션으로 작성되었다. 자료 구조 저장을 위해 데이터베이스를 사용하지 않고 텍스트 기반의 I/O로 구성하여 가벼운 프로토타입을 완성하였다.

전체 프로토타입의 아키텍쳐 구성은 <그림 1>과 같다. 전체는 크게 두 가지 모듈로 구분될 수 있다. 첫 번째 모듈인 Setter는 트위터 유저의 리스트를 받아서 데이터를 저장하는 역할을 한다. 데이터는 학습을 위한 데이터(Learning Data)와 실험을 위한 데이터(Test Data)로 구분되어 저장된다. 두 번째 모듈인 Starter는 Setter에서 생성된 학습 데이터를 사용하여 개인정보에 관련된 문서들을 학습한 후, 실험 데이터를 분류하여 결과를 도출하는 역할을 한다.



<그림 1> 실험 프로세스

5. 실험 및 결과

개인정보 문서 분류에 쓰일 알고리즘을 성능 측정을 위하여 기계 학습 문서 분류에서 가장 널리 쓰이는 Vector Space 선형 분류법(Lee et al., 1997; Steinbach et al., 2000)을 Naïve Bayes 분류법과 비교 분석하였다. 각 분류법에 대하여 각각 90개의 샘플로 이루어진 실험 데이터를 이용하여 측정치를 구하였다.

개인정보 관련 문서(P)와 일반 문서 분류(NP)에 대하여 precision(정밀도), recall(재현율), F-measure[3]를 측정하였다. 실험 1에서는 Vector Space 분류법을 이용하였고 실험 2에서는 Naïve Bayes 분류법을 이용하였다. 실험 결과 <표 1>에서 나타난 것과 같이 Naïve Bayes 분류 방법론이 Vector Space 분류 방법론보다 훨씬 우수함을 알 수 있었다. 분류의 정확도 역시 Naïve Bayes 분류법이 더 우수한 측정치를 보여주었다.

<표 1> Vector Space, Naïve Bayes 분류법의 측정치

	Precision	Recall	F-measure
실험 1-P (Vector Space)	0.696	0.711	0.703
실험 2-P (Naïve Bayes)	0.791	0.756	0.773
실험 1-NP (Vector Space)	0.705	0.689	0.697
실험 2-NP (Naïve Bayes)	0.766	0.800	0.783

5.1 개인정보 유형별 영향력 측정

본 연구에서는 기존의 학습 데이터에 추가적으로 개인정보의 유형에 관련된 하위 학습 데이터를 추가함으로써 자동 문서 분류 알고리즘의 성능 측

3) F-measure = 2×Precision×Recall/(Precision+Recall).

정치를 높이는 것을 시도하였다. 개인정보의 유형에 따라 알고리즘에 효과적으로 적용하는 방향을 제시하기 위하여 본 논문에서는 기존 논문에서 나타난 개인정보의 유형들을 분석하여 Appendix 1의 <표 4>와 같이 종합하였다. 여러 가지 유형들은 언뜻 보기에는 모두 개인정보에 관련되어 있지만 실제로 문서상에서 자질로 나타나는지 아닌지 판별할 필요가 있다.

각각의 자질들이 개인정보 관련 문서에 어느 정도의 종속성을 지니는지 알아보기 위하여 Mutual Information 방법론을 이용한다. 개인정보의 유형마다 개인정보 문서를 분류함에 있어서 그 중요도가 다르게 나타남을 알 수 있다. 예를 들어 주소나 전자우편 같은 경우, 개인적인 정보가 아닐지라도 조직이나 기관에서 함께 쓰이는 정보의 유형들과 유사하여 개인정보 영역의 정보인지 아닌지에 대한 분간이 쉽지 않다. <표 2>와 같이 개인정보 관련 문서에 종속성을 띠는 개인정보 유형 집단을 F 집단이라 정하고, 추가된 실험에서는 집단 F를 개인정보 관련 문서의 중요 자질로 포함시켰다. 출현빈도가 지나치게 낮은 정보 유형은 모두 비종속적인 유형으로 분류하였다.

<표 2> 개인정보 유형 별 개인정보 관련 문서에서의 종속성 측정

	종속적인 개인정보 유형(F)	독립적인 개인정보 유형
식별정보	이름, 생일, 가족, 혼인관계	주소, 국적, 전자우편
민감정보	건강상태, 진단/수술기록, 통화/연락기록, 구속/전과, 휴가, 취미, 여가활동, 대여/관람 기록, 일정, 기호, 성생활	병력기록
생활정보	위치정보, 이동경로,	
교육, 훈련 정보	학교활동, 상벌기록	학력
고용정보	직업/고용이력, 고용주, 동료정보, 직업훈련, 교육	회사주소
재산소비	자동차, 보유현금, 저축현황, 연봉/이자, 구매형태, 대출상황, 계약상 합의	주식, 채권, 투자, 파산
사회활동	정당, 종교	기부

5.2 자질의 질 향상을 통한 알고리즘 성능 향상

본 논문에서는 개인정보 관련하여 높은 정확도를 보여준 개인정보 문자열을 실험 데이터에 추가적으로 학습시켜 개인정보 관련 문서 자질 집단의 질을 높임으로써 자동 분류 알고리즘의 성능을 향상시키는 방법을 제안한다. 실험 3에서는 앞서 실험 2에서 사용한 90개의 실험 데이터를 대상으로 실험 대상의 크기에 비례하여 개인정보 관련 유형 F에 대한 정보를 기존의 500개 학습 대상과 함께 학습[4]시킨 후 그 측정치를 비교 분석한 결과는 <표 3>과 같았다. 집단 F를 추가로 학습시킨 실험 3은 측정치에서 기존 집단으로만 학습시켰던 실험 2에 비하여 precision, recall, F-measure의 측정치에 있어서 개인정보 관련 문서와 일반문서 모두에서 높은 성능을 나타냈다.

<표 3> Naïve Bayes 방법론을 이용하여 개인정보 유형 F 집단을 추가 학습 시킨 실험 결과

	Precision	Recall	F-measure
실험 2-P (기초 학습)	0.791	0.756	0.773
실험 3-P (추가 학습)	0.872	0.911	0.891
실험 2-NP (기초 학습)	0.766	0.800	0.783
실험 3-NP (추가 학습)	0.907	0.867	0.886

위와 같은 Naïve Bayes 분류법 성능 향상 결과는

4) 기존 학습 데이터 10Kbyte 당 개인정보에 관련된 개인정보 유형 집단 F자질 전체를 1회 추가 학습.

개인정보 유형 중 실제 개인정보 문서에 자질로서 나타나는 것들만 추려서 추가 학습시킴으로써 가능하였다. 개인정보 관련 문서로 분류된 학습 대상과 함께 개인정보에 영향력이 있는 개인정보 유형들을 추가 학습시켜 알고리즘이 학습하는 문서 자질(feature)의 질(quality)을 높였다. 높아진 학습 자질의 질로 인하여 기존의 Naïve Bayes 방법론을 이용한 실험 결과 precision과 recall 등의 평가 측정치가 90%에 근접한 수치를 보이면서 알고리즘의 성능 향상을 확인할 수 있었다.

6. 결론 및 향후 연구 과제

6.1 연구 결과의 요약

본 논문은 증가하고 있는 개인정보 침해를 방지하기 위한 방법 중 하나로서 정보 유출 가능성에 대해 감독하는 시스템이 필요하다는 점에서 착안되었다. 따라서 자동 문서 분류기의 학습 알고리즘을 통하여 개인정보 관련 문서를 탐지하고, 이를 통해 추후에 있을 수 있는 개인정보 유출을 미연에 방지하는데 그 목적이 있다. 본 논문은 개인정보에 영향력이 있는 개인정보 유형들을 기존 학습 데이터에 추가 학습시켜 알고리즘이 학습하는 문서 자질(feature)의 질(quality)을 높이는 방법을 제시한다. 분석 결과를 통해 기존의 Naïve Bayes 방법론의 측정치를 높일 수 있음을 증명하였고 이를 실생활에 적용할 수 있다는 점을 시사한다.

본 논문의 실험에서는 한 명의 트위터 유저 화면에서 보이는 글들의 묶음을 하나의 문서 단위로 취급한다. 개인정보 문서 분류에 쓰일 알고리즘을 선정을 위하여 기계 학습 문서 분류에서 가장 널리 쓰이는 Vector Space 분류법과 Naïve Bayes 분류법을 비교 분석하였다. 각 분류법에 대하여 precision(정밀도), recall(재현율), F-measure, 그리고 accuracy(정확도) 등의 여러 측정치를 구한 결과 Naïve Bayes 분류법의 알고리즘 성능이 더 우수함을 알 수 있었다.

기존의 방법론을 그대로 적용하였을 때, 실제 그것이 사용되기 위해서는 알고리즘 성능을 더욱 향상시킬 필요가 있었다. Lewis(1992)는 학습 알고리즘을 시행할 때 우리가 언어적 편견에 빠져서 학습할 대상의 자질을 잘못 분석하는 경우가 많다고 말한다. 본 논문에서는 이러한 자질들의 질을 향상 시킬 수 있는 방법 중 학습 과정에서 제대로 된 자질들을 선택하는 것(feature selection)을 이용하여 자동 문서 분류 알고리즘의 성능을 향상시키고자 하였다.

본 연구에서는 기존의 학습 데이터에 추가적으로 개인정보의 유형에 관련된 하위 학습 데이터를 추가함으로써 자동 문서 분류 알고리즘의 성능 측정치를 높이는 것을 시도하였다. 각각의 개인정보 유형들이 개인정보 관련 문서에서 출현한 빈도 횟수를 전체 출현 횟수로 나누어 각 자질 혹은 개인정보 유형의 개인정보 관련 영향력을 측정하였다. 그리고 개인정보 관련 문서로 분류된 학습 대상과 함께 개인정보에 영향력이 있는 개인정보 유형들을 추가 학습시켜 알고리즘이 학습하는 문서 자질(feature)의 질(quality)을 높였다. 높아진 학습 자질의 질로 인하여 기존의 Naïve Bayes 방법론을 이용한 평가 측정치가 높아질 수 있었다. 결과 중 precision과 recall 등의 측정치가 90%에 근접한 수치를 보이면서 실생활에서 알고리즘의 사용 가능성이 매우 높음을 보였다.

6.2 연구의 시사점

우선 학술적 내용에서 본 논문은 자동 문서 방법론의 적용과 개인정보 분야 부분에 여러 영향을 끼칠 수 있다. 첫째로 본 연구는 개인정보 분야에

서 문서를 자동 분류하는 방법을 제시했다는데 그 의의가 있다. 개인정보 보안 분야에 문서 분류 방법론을 적용하였으며, 개인정보 관련 문서 자질의 질을 높이는 방안을 이용하여 알고리즘의 성능을 더욱 높일 수 있음을 보여준다.

둘째, 본 논문은 문서를 분류함에 있어서 범주에 해당하는 문서의 자질의 질에 대한 검토가 중요하다는 점을 환기해준다. 우리가 문서 분류 시 자질을 판단할 때 선입견에 사로 잡혀 잘못된 문서 자질을 선택할 수 있다. 연구 실험 결과, 문서를 자동으로 분류하는 과정에서 분류 범주에 해당하는 문서 자질의 질을 향상시킴으로써 문서 분류 알고리즘의 성능을 크게 향상시킬 수 있었다.

셋째, 본 논문은 개인정보의 유형을 기존 문헌을 통해 종합하고 유형별로 텍스트 분석에 활용 가치가 있는지 없는지에 대한 자료를 제시하였다. 우리가 개인정보에 해당한다고 생각하지만 실제 문서의 자질에서는 그것이 발현되지 않는 경우가 존재한다. 논문에서 밝혀낸 개인정보 유형의 특징은 향후 개인정보를 연구하는 다른 학자들에게 활용도가 높은 정보를 제공한다.

학술적인 부분뿐만 아니라 실용적인 부분에서도 본 논문은 그 의미가 아주 크다. 논문이 제시하는 방법론은 개인정보를 담고 있는 문서를 분류할 수 있다는 점에서 우리에게 여러 가지 이익을 줄 수 있다.

첫째, 개인적인 측면에서는 작성한 문서의 개인정보 누출 가능성을 알림으로써 개인 차원에서 개인정보 누출을 스스로 자제하려는 노력을 할 수 있다. 또한, 기존에 존재하는 개인정보 관련 문서들을 검토하고 수정 혹은 삭제 할 수 있다. 이러한 영향은 결국 개인정보가 누출되는 것을 사전에 미리 방지함으로써 개인정보 누출로 인한 개인적 피해 사례를 미연에 막을 수 있다.

둘째, 조직적인 측면에서는 개인정보를 감지할 수 있는 지능형 시스템 구축을 하는데 큰 도움이 될 수 있다. 개인정보 누출로 인하여 이용자와 인터넷 서비스를 제공하는 회사 간의 법적인 문제를 줄일 수 있다. 또한 개인정보를 관리하는 시스템을 통하여 제공하고 있는 서비스에 대한 고객 신뢰도를 높일 수 있는 효과를 제공할 수 있다.

셋째, 사회와 국가적인 측면에서 개인정보 누출로 인한 피해 사례를 줄임으로써 경제적, 정신적인 여러 손실을 줄일 수 있다. 개인정보 누출이 원인이 되는 사회 문제 발생과 법정 소송 등의 문제점을 줄일 수 있다.

6.3 연구의 한계점 및 향후 발전 방향

본 연구는 Naïve Bayes 분류 알고리즘과 개인정보 유형 분석을 이용하여 개인정보 관련 문서를 분류하는 방법을 제안한다. 실험결과 개인정보 관련 문서 분류에 대한 측정치는 높은 수치를 보였다. 이 측정치들은 향후 연구를 통해 더욱 향상될 수 있으리라 생각한다.

첫째, 문서 내에서 불용어(stopword)의 추출과 어간화(stemming) 등의 중간 과정을 보완할 수 있다면 분류 측정치를 더욱 높일 수 있으리라 기대된다. 본 연구에서는 영어의 불용어나 어간화 분석에 대하여 다루지 않았다. 스탠포드 대학에서 개발한 POS-Tagger나 아파치(Apache)의 오픈소스인 루씬(Lucene) 등을 이용한다면 불필요한 용어들이 알고리즘에서 생략되거나 종합되어 더 정확한 분류 결과를 도출할 것이라 기대한다.

둘째, 트위터와 같은 소셜 미디어적 성격의 인터넷 문서 외에 다른 실험 데이터를 이용하여 알고리즘 성능을 판단해 볼 필요가 있다. 트위터와 같은 경우 개인적인 문서에는 지나치게 약술된 용어들이 많이 쓰이고 있으며 조직이나 기관에 관련

된 문서에서는 빈번하게 하이퍼링크 관련 정보들이 쓰이고 있다. 이러한 데이터 성향이 알고리즘 성능 측정치를 다소 높게 나타내게끔 유발했을 가능성을 염두에 둘 필요가 있다.

셋째, 기존에 작성된 인터넷 문서의 개인정보 누출 위험을 분류하는 것이 아닌, 글이 작성되는 순간 실시간으로 글을 분류하는 방법이 필요하다. 한 메시지에 대한 분류는 힘들겠지만 서로 관련된 여러 메시지들을 묶어서 분류할 수 있을 것이라 기대한다.

마지막으로, Naïve Bayes를 비롯한 다양한 기존의 문서 분류 방법론을 적용해볼 필요가 있다. 현재, 기초적인 방법론을 변형시킨 수많은 알고리즘들이 개발되어 있다. 본 논문에서 이용한 알고리즘은 가장 기초적인 Naïve Bayes 방법론을 사용한 것으로 변형되어 개발된 여타 알고리즘들은 더 높은 성능을 발현할 가능성이 있다.

참고문헌

권건보, *개인정보보호와 자기정보통제권*, 서울 : 경인문화사, 2005.

방송통신위원회, *트위터에 노출된 나의 정보는 얼마나 될까?*, 방송통신위원회, 2011.

윤상오, "전자정부 구현을 위한 개인정보보호 정책에 관한 연구 : 정부신뢰 구축의 관점에서", *한국지역정보화학회지*, 12권 2호(2009), 1~29.

이강신, 이기혁, 박진식, 최일훈, *개인정보보호 기초와 활용*, 서울 : 미디어그룹 인포더, 2010.

이창범, 조정현, "APT(Asia-Pacific Telecommunity) 개인정보 및 프라이버시 보호 가이드라인 제정 방안 연구", *개인정보분쟁조정위원*, 2003.

조동기, 김성우, "인터넷의 일상화와 개인정보 보호", KISDI *이슈리포트*, 11권(2003), 10~11.

한국정보보호진흥원, *개인정보의 경제적 가치 분석에 관한 고찰*, 서울 : 한국정보보호진흥원, 2007.

황인호, "개인정보보호 제도에서의 규제에 관한 연구", *공법연구*, 30권 4호(2002), 232~232.

Bayes, T., "An essay towards solving a problem in the doctrine of chances. Philos", *Philosophical Transactions*, Vol.53(1763), 370~418.

Boyd, D. M. and N. B. Ellison, "Social Network Sites : Definition, History, and Scholarship", *Journal of Computer-Mediated Communication*, Vol.13, No.1(2008), 210~230.

Clarke, R., Beyond the OECD Guidelines : Privacy Protection for the 21st Century, (2000), Roger Clarke's Web-Site : http://www.rogerclarke.com/DV/PP21C.html.

Cooley, T. C., *Laws of Torts*, New York : Praeger, 1888.

Davies, S., *Big Brother : Britain's web of surveillance and the new technological order*, London : Pan, 1996.

Domingos, P. and M. Pazzani, "Beyond Independence : Conditions for the Optimality of the Simple Bayesian Classifier", *Proceedings of the 13th International Conference on Machine Learning*, (1996), 105~112.

Gross, R. and A. Acquisti, "Information revelation and privacy in online social networks", *WPES '05 Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, 2005.

Information Commissioner's Office, *Notification Handbook-A Complete Guide to Notification*. Information Commissioner, 2001.

Jagatic, T., N. Johnson, M. Jakobsson, and F. Menczer, "Social phishing", *Communications of the ACM*, Vol.5, No.10(2007), 94~100.

Kobsa, A., "Personalized Hypermedia and International Privacy", *Communications of the ACM*, Vol.45, No.5(2002), 64~67.

Lampe, C. and N. B. Ellison, "Changes in use and perception of facebook", *Proceedings of the 2008 ACM conference on Computer supported cooperative work CSCW '08*, (2008), 721~730.

Lee, D. L., H. Chuang, and K. Seamons, "Document Ranking and the Vector-Space Model", *IEEE Software*, Vol.14, No.2(1997), 67~75.

Lewis, D. D., *Representation and Learning in Information Retrieval*, Doctorial Dissertation : The Graduate School of the University of Massachusetts, 1992.

Livingstone, S., "Taking risky opportunities in youthful content creation : teenager's use of social networking sites for intimacy, privacy and self-expression", *New Media Society*, Vol.10, No.3(2008), 393~411.

LoPucki, L. M., "Human Identification Theory and the Identity Theft Problem", *Tex. L. Rev.*, Vol.80(2001), 89~135.

Manning, C. D., P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

Mason, R. O., "Four Ethical Issues of the Information Age", *MIS Quarterly*, Vol.10, No.1 (1986), 5~12.

Meeder, B., J. Tam, P. G. Kelley, and L. F. Cranor, "RT@IWantPrivacy : Widespread Violation of Privacy Settings in the Twitter Social Network", *Web 2.0 Privacy and Security Workshop, IEEE Symposium on Security and Privacy*, 2010.

Mitchell, T. M., *Machine Learning*, McGraw Hill, 2010.

Parent, W. A., "Privacy : A Brief Survey of the Conceptual Landscape", *Santa Clara Computer and High Tech. L. J.*, Vol.11(1995).

Peng, H., F. Long, and C. Ding, "Feature selection based on mutual information : criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.8(2005), 1226~1238.

Schauer, F., "Internet Privacy and The Public-Private Distinction", *Jurimetrics*, Vol.38, No.4 (1998), 555~564.

Smith, R. E., *Ben Franklin's Web Site : Privacy and Curiosity from Plymouth Rock to the Internet*, Sheridan Books, 2000.

Solove, D. J., "A taxonomy of privacy", *University of Pennsylvania Law Review*, Vol.154, No.3(2006), 477~560.

Steinbach, M., G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques", *KDD Workshop on Text Mining*, 2000.

Tong, S. T., B. Van Der Heide, L. Langwell, and J. B. Walther, "Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook", *Journal of Computer-Mediated Communication*, Vol.13, No.3(2008), 531~549.

Wacks, R., *Privacy*, New York : Oxford University Express, 1993.

Warren. S. D., L. D. Brandeis, "The Right to Privacy", *Harvard Law Review*, Vol.4, No.5 (1890), 193~220.

Weible, R. J., *Privacy and data : An empirical study of the influence of types of data and situational context upon privacy perceptions*, D.B.A. : Mississippi State University, 1993.

Wolak J., D. Finkelhor, K. J. Mitchell, and M. L. Ybarra, "Online "Predators" and Their Victims", *Psychology of Violence*, Vol.1, No.1(2010), 13~35.

〈Appendix 1〉

<표 4> 개인정보 유형에 대한 문헌 조사 종합

사항	Weible	조동기, 김성우	한국 행정자치부	영국 정보위원회	이강신 외
이름	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
주민등록번호	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
운전면허번호	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
주소	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
생년월일	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
출생지	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
본적지	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
국적	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
주거기록	일반정보	인적 사항	생활, 가정, 신분	개인속성	인적 사항
가족관계	가족정보	인적 사항	생활, 가정, 신분	가족, 사회 환경	인적 사항
혼인관계	가족정보	인적 사항	생활, 가정, 신분	가족, 사회 환경	인적 사항
동거관계	가족정보	인적 사항	생활, 가정, 신분	가족, 사회 환경	인적 사항
이혼기록	법적정보	인적 사항	생활, 가정, 신분	가족, 사회 환경	인적 사항
혈액형	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
성별	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
DNA	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
지문	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
심신장애	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
병력기록	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
수술기록	의료정보	신체, 의료	신체적특징	개인속성	신체적 정보
전화번호	일반정보	통신, 위치	생활, 가정, 신분	개인속성	인적 사항
전자우편	N/A	통신, 위치	생활, 가정, 신분	개인속성	인적 사항
위치정보	N/A	통신, 위치	생활, 가정, 신분	가족, 사회 환경	기타
회원 ID	N/A	통신, 위치	생활, 가정, 신분	개인속성	인적 사항
IP	N/A	통신, 위치	생활, 가정, 신분	개인속성	기타
통화기록	N/A	통신, 위치	N/A	개인속성	기타
접속로그	N/A	통신, 위치	생활, 가정, 신분	개인속성	기타
이동기록	N/A	통신, 위치	N/A	가족, 사회 환경	기타
학력	교육 및 훈련정보	교육, 훈련	사회경력	교육, 훈련	교육정보
학교성적	교육 및 훈련정보	교육, 훈련	사회경력	교육, 훈련	교육정보
자격증	교육 및 훈련정보	교육, 훈련	사회경력	교육, 훈련	교육정보

사항	Weible	조동기, 김성우	한국 행정자치부	영국 정보위원회	이강신 외
서클활동	교육 및 훈련정보	교육, 훈련	사회경력	교육, 훈련	교육정보
상벌기록	교육 및 훈련정보	교육, 훈련	사회경력	교육, 훈련	교육정보
직업	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
고용형태	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
병역상태	병역정보	고용, 경력	사회경력	고용 및 근로	근로정보
고용이력	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
고용주	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
회사주소	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
동료 정보	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
직업 훈련, 교육	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
직무 평가	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
근태기록	고용정보	고용, 경력	사회경력	고용 및 근로	근로정보
소유주택	부동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
소유토지	부동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
소유상점	부동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
소유건물	부동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
자동차	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
보유현금	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
저축현황	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
현금카드	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
주식	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
채권	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
유가증권	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
수집품	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
고가의 예술품	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
보석	동산정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
연봉	소득정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
이자소득	소득정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
임대소득	소득정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
기타소득	소득정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
보험가입현황	기타수익정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
보험수익자	기타수익정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
회사의 판공비	기타수익정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
투자프로그램	기타수익정보	재산, 소비	경제관계	금융 및 신용	개인금융정보

사항	Weible	조동기, 김성우	한국 행정자치부	영국 정보위원회	이강신 외
퇴직프로그램	기타수익정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
신용카드 번호	신용정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
대출상황	신용정보	재산, 소비	경제관계	금융 및 신용	신용정보
저당권설정여부	신용정보	재산, 소비	경제관계	금융 및 신용	신용정보
카드 연체, 미납	신용정보	재산, 소비	경제관계	금융 및 신용	신용정보
구매형태	신용정보	재산, 소비	경제관계	금융 및 신용	개인금융정보
교통위반기록	법적정보	정치, 사회	사회경력	민감정보	법적정보
파산, 담보기록	법적정보	정치, 사회	사회경력	민감정보	법적정보
구속기록	법적정보	정치, 사회	사회경력	민감정보	법적정보
전과기록	법적정보	정치, 사회	사회경력	민감정보	법적정보
납세기록	법적정보	정치, 사회	경제관계	민감정보	법적정보
제공받는 재화	소득정보	재산, 소비	경제관계	계약 활동	개인금융정보
법적 이용권	소득정보	재산, 소비	경제관계	계약 활동	개인금융정보
계약상의 합의	소득정보	재산, 소비	경제관계	계약 활동	개인금융정보
정당	조직정보	정치, 사회	사회경력	민감정보	정신적 정보
종교	조직정보	정치, 사회	사회경력	민감정보	정신적 정보
노조	조직정보	정치, 사회	사회경력	민감정보	정신적 정보
휴가	기타수익정보	여가, 생활	심신의 상태	가족, 사회 환경	기타
병가	기타수익정보	신체, 의료	심신의 상태	가족, 사회 환경	기타
취미	습관 및 취미	여가, 생활	N/A	가족, 사회 환경	기타
여가활동	습관 및 취미	여가, 생활	N/A	가족, 사회 환경	기타
대여기록	습관 및 취미	여가, 생활	N/A	민감정보	기타
관람기록	습관 및 취미	여가, 생활	N/A	가족, 사회 환경	기타
기부활동	법적정보	정치, 사회	사회경력	가족, 사회 환경	기타

Abstract

Personal Information Detection by Using Naïve Bayes Methodology

Namwon Kim* · Jinsoo Park**

As the Internet becomes more popular, many people use it to communicate. With the increasing number of personal homepages, blogs, and social network services, people often expose their personal information online. Although the necessity of those services cannot be denied, we should be concerned about the negative aspects such as personal information leakage.

Because it is impossible to review all of the past records posted by all of the people, an automatic personal information detection method is strongly required. This study proposes a method to detect or classify online documents that contain personal information by analyzing features that are common to personal information related documents and learning that information based on the Naïve Bayes algorithm.

To select the document classification algorithm, the Naïve Bayes classification algorithm was compared with the Vector Space classification algorithm. The result showed that Naïve Bayes reveals more excellent precision, recall, F-measure, and accuracy than Vector Space does. However, the measurement level of the Naïve Bayes classification algorithm is still insufficient to apply to the real world.

Lewis, a learning algorithm researcher, states that it is important to improve the quality of category features while applying learning algorithms to some specific domain. He proposes a way to incrementally add features that are dependent on related documents and in a step-wise manner. In another experiment, the algorithm learns the additional dependent features thereby reducing the noise of the features. As a result, the latter experiment shows better performance in terms of measurement than the former experiment does.

Key Words : Naïve Bayes, Document Classification, Personal Information, Privacy, Security, Social Network Service

* College of Business Administration, Seoul National University
** Graduate School of Business, Seoul National University

저 자 소 개

김남원

한국 과학기술원(KAIST) 산업공학과를 졸업하였으며, 서울대 경영대학에서 경영정보 전공으로 석사학위를 취득하였다. 주요 관심분야는 정보시스템 보안, 시맨틱 웹, Social Network에서의 개인정보보호 등이 있다.

박진수

The University of Arizona에서 경영정보시스템을 전공하여 경영학 박사를 취득했으며, University of Minnesota의 Carlson School of Management에서 조교수, 고려대학교 경영대학에서 조교수를 역임했다. 현재 서울대학교 경영전문 대학원/경영대학에 부교수로 재직 중이다. 현재 국제저널인 Journal of Database Management와 International Journal of Principles and Applications in Information Science and Technology의 편집위원으로 활동하고 있으며 MIS Quarterly, IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Computer, ACM Transactions on Information Systems(TOIS), Information Systems Frontiers, Communications of the AIS, Journal of Global Information Technology Management(JGITM), International Journal of Electronic Business, 경영정보학연구 등 국내외 유수 전문학술지에 다수의 논문을 게재하였다. 주요 관심분야는 온톨로지, 정보 시스템 통합, 지식 공유, 에이전트, 시맨틱 모델링, 웹 정보시스템 등이 있다.