사용자 행동을 이용한 쓰레기편지 여과의 성능 개선

김 재 훈[†] · 김 강 민[††]

요 약

인터넷의 급속한 성장으로 전자편지는 정보 전달의 중요한 수단으로 사용되고 있다. 그러나 수신자가 원하지 않는 전자편지(쓰레기 편지)들이 무분별하게 배달될 수 있으며, 이로 인해 사회적으로는 물론이고 경제적으로도 큰 문제가 되고 있다. 이와 같이 쓰레기 편지를 차단하거나 여과하기 위해서 많은 연구자와 회사에서 꾸준히 연구를 진행하고 있다. 일반적으로 쓰레기 편지를 결정하는 기준은 수신자에 따라서 조금씩 차이가 있다. 또한 쓰레기 편지와 정보성 편지에 따라서 수신자가 취하는 행동이 다르다. 이 논문은 이런 사용자 행동을 쓰레기 편지 여과 시스템에 반영하여 그 시스템의 성능을 개선한다. 제안된 시스템은 크게 두 단계로 구성된다. 첫 번째 단계는 사용자 행동을 추론하는 단계이고 두 번째 단계는 추론된 사용자 행동을 이용해서 쓰레기 편지를 여과하는 단계이다. 두 단계 모두에서 점진적인 기계학습 방법(TiMBL - IB2)을 이용한다. 제안된 시스템을 평가하기 위해 12명의 사용자로부터 12,000통으로 이루어진 전자편지 말뭉치를 구축하였다. 실험 결과는 사용자에 따라 81% ~ 93%의 분류 정확도를 보였다. 사용자의 행동 정보를 포함하는 편지 분류 결과는 그렇지 않은 결과에 비해 평균 14%의 분류 정확도가 향상되었다.

키워드 : 한국어처리, 쓰레기 편지, 정보 여과, 기계학습

Performance Improvement of Spam Filtering Using User Actions

Jae-Hoon Kim[†] · Kang-Min Kim[††]

ABSTRACT

With rapidly developing Internet applications, an e-mail has been considered as one of the most popular methods for exchanging information. The e-mail, however, has a serious problem that users can receive a lot of unwanted e-mails, what we called, spam mails, which cause big problems economically as well as socially. In order to block and filter out the spam mails, many researchers and companies have performed many sorts of research on spam filtering. In general, users of e-mails have different criteria on deciding if an e-mail is spam or not. Furthermore, in e-mail client systems, users do different actions according to a spam mail or not. In this paper, we propose a mail filtering system using such user actions. The proposed system consists of two steps: One is an action inference step to draw user actions from an e-mail and the other is a mail classification step to decide if the e-mail is spam or not. All the two steps use incremental learning, of which an algorithm is IB2 of TiMBL. To evaluate the proposed system, we collect 12,000 mails of 12 persons. The accuracy is 81 ~ 93% according to each person. The proposed system outperforms, at about 14% on the average, a system that does not use any information about user actions.

Key Words : Korean Language Processing, Spam Mail, Information Filtering, Machine Learning

1. 서 론

인터넷의 급속한 성장과 더불어 전자편지(e-mail)는 중요한 정보 전달의 수단으로 사용되고 있으나, 원하지 않는 많은 전자편지들이 배달되어 사생활을 침해하거나 바이러스를 감염시키거나 유포하는 등 사회적으로 큰 문제가 야기되고 있다[1]. 이와 같이, 수신자가 원하지 않는 전자편지를 일반적으로 쓰레기 편지(spam mail)이라고 하는데, 이들은 몇 가지의 공통적인 성질을 가지고 있다[2]. 첫째, 쓰레기 편지는 수신자가 원하지 않는다(불원성). 둘째, 쓰레기 편지는 일반적으로 영리적인 목적을 가진다(상업성). 셋째, 쓰레기 편지는 대량으로 보내진다(대량성). 마지막으로 대부분의 쓰레기 편지는 불법적이다(불법성)[1]. 또한 쓰레기 편지는 수신자나 전자편지 서비스 제공 업체들에게 경제적으로 많은 피해를

† 종신회원 : 한국해양대학교 컴퓨터공학과 부교수
†† 준 회 원 : (주)태광ENG 연구소 연구원
논문접수 : 2006년 1월 23일, 심사완료 : 2006년 3월 14일

1) 불법적인 예들은 다음과 같다. ① 기술적인 조작으로 발송자의 신원을 숨긴다. ② 반사회적이거나 악의적이고 불쾌한 내용이 포함되어 있다. ③ 편지 주소는 수신자의 동의 없이 수집되거나 판매된 것이다.

주고 있다. ITU[2]의 조사보고서에 따르면, 2003년 한 해 전 세계적으로 쓰레기 편지로 인한 경제 손실 비용이 약 25조 원[3]에 달하는 것으로 추정되었다[3]. 같은 해 국내에도 쓰레기 편지로 인한 피해액이 약 1조 3천억 원에 이르는 것으로 추정되었다[1].

이 문제를 해결하기 위해서 많은 연구자들은 쓰레기 편지를 여과하는 방법[4-6]에 대해서 연구하고 있으며, 공개되거나 상용화된 쓰레기 편지 여과 시스템[4]이 사용되고 있다[7]. 정부에서도 '정보통신망이용촉진및정보보호등에관한법률'을 제정하여 법적·제도적 장치를 마련하였고, 2003년부터 '불법스팸대응센터[5]'를 운영하고 있다. 또한 각 연구기관 및 정보통신 기업체들은 쓰레기 편지에 대한 기술적인 대처 방안을 연구하고 관련 시스템을 제안하였다[1].

전자편지의 배달 경로는 <발신자, 발신 서버, 수신 서버, 수신자>이며, 쓰레기 편지를 차단하는 기술은 이 배달 경로상의 각 구성요소에 따라 조금씩 다를 수 있다. 예를 들면, 발신 서버 혹은 수신 서버의 경우에는 전자편지 서비스의 보편적인 사용자들이 원하지 않는 쓰레기 편지의 배달을 차단한다. 즉 쓰레기 편지의 근원지[6](blacklist) 정보를 이용하여 주로 쓰레기 편지의 배달을 차단한다. 쓰레기 편지가 수신자에게 배달되지 않도록 하는 방법에는 1) 편지주소 수집을 차단하는 방법, 2) 대량의 쓰레기 편지의 전송을 차단하는 방법, 3) 쓰레기 편지의 발신자 신원을 확인하는 방법, 4) 쓰레기 편지를 여과하거나 막는 방법이 있다[1]. 이 논문은 수신자 측에서 쓰레기 편지를 여과하는 방법에 관련된다. 수신자 측에서 쓰레기 편지의 여과 방법은 전자편자를 분석하여 쓰레기 편지의 여부를 판단하는 방법이며, opt-in 방법과 opt-out 방법이 있다. Opt-in 방법은 수신자가 동의하지 않은 편지를 사전에 차단하는 기술이고, opt-out 방법은 수신된 전자편지가 쓰레기 편지인지를 식별하는 기술이다. 이 논문은 opt-out 방법에 관련된다.

앞에서 말한 쓰레기 편지의 정의에서 알 수 있듯이 쓰레기 편지는 수신자에 따라서 그 기준이 많이 다를 수 있다. 예를 들면, 어떤 사용자는 책에 관련된 광고성 전자편지는 유용한 정보가 될 수 있으나 또 다른 사용자에게는 쓰레기 편지가 될 수 있다. 이처럼 쓰레기 편지의 정의는 수신자에 따라서 다를 수 있기 때문에 마이크로소프트사의 Outlook과 같은 많은 전자편지 수신 도구들은 수신자가 직접 쓰레기 편지를 차단하기 위한 규칙을 정의할 수 있는 기능을 가지고 있다. 또한 SpamBayes와 같은 시스템은 기계학습 방법을 이용해서 수신자가 직접 원하지 않는 메일을 여과할 수 있는 기능들을 제공한다.

이 논문에서는 기계학습을 이용한 쓰레기 여과 시스템에 사용자 행동 정보를 추가하여 쓰레기 편지 여과 시스템의 성능을 개선한다. 여기서 말하는 사용자 행동이란 전자편지를 읽을 때, 사용자 취하는 행동을 말한다. 예를 들면, 쓰레기 편지를 읽으면 일반적으로 제목만 보고는 바로 지우는 경향이 있으나, 정보성 편지를 읽으면 읽는데 다소 시간이 소요될 뿐 아니라 특정 폴더에 일정 기간 동안 보관한다. 이와 같이 쓰레기 편지와 정보성 편지에 따라 사용자들의 행동(action)이 서로 상이하다. 이 논문에서는 전자 편지에 대한 사용자 행동을 추론하여 이를 쓰레기 여과 시스템에 이용한다. 제안된 쓰레기 편지 여과 시스템은 크게 두 부분으로 구성된다. 하나는 사용자의 행동을 추론하는 부분이고 또 다른 하나는 추론된 사용자 행동을 이용하여 쓰레기 편지를 여과하는 부분이다. 이 모든 부분에서 기계학습 방법 중 하나인 사례기반 학습 방법을 이용한다. 일반적으로 기계학습 방법을 이용하기 위해서는 학습 데이터가 필요하다. 이 논문에서는 학습 데이터를 직접 구축하였다. 그 이유는 사용자 행동이 포함된 학습 데이터가 전혀 존재하지 않기 때문이다. 사용자 행동을 이용했을 경우 그렇지 않을 경우보다 평균 14% 정도의 성능 개선을 보였다.

이 논문의 구성은 다음과 같다. 2장은 관련연구로 쓰레기 편지 여과에 사용되고 있는 기계학습 방법과 전자편지 말뭉치에 대해서 간단히 살펴보고, 사용자 행동의 활용에 대해서 간단히 살펴보고자 한다. 3장에서는 이 논문에서 제안하는 사용자 행동 정보를 이용한 편지 여과 시스템의 구조를 자세히 설명하고, 4장에서는 제안된 시스템의 성능을 평가한다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 제시하고자 한다.

2. 관련 연구

이 장에서는 기계학습을 이용한 쓰레기 편지 여과 방법과 시스템을 평가하기 위한 편지 말뭉치에 대해서 살펴본다. 또한 인간과 컴퓨터 상호작용 분야에 사용되고 있는 사용자의 행동에 대해서 살펴보고자 한다.

2.1 기계학습을 이용한 쓰레기 편지 여과 방법

기계학습을 이용한 쓰레기 편지 여과는 두 종류의 클래스(spam과 ham)를 가지는 분류(classification) 문제로 간주할 수 있다. 따라서 대부분의 기계학습 방법[8]들이 그대로 쓰레기 편지 여과 시스템에 사용될 수 있다. 이들 중에도 쓰레기 편지 여과에 가장 널리 사용되는 기계학습 방법은 나이브 베이지안(naïve Bayesian) 분류자[9-10]이다. 이 방법은 단어(혹은 자질)와 클래스 사이의 조건확률만 이용하기 때문에 비교적 모델이 단순하며, 아주 쉽게 구현할 수 있다. 더구나 성능면에서도 다른 기계학습 방법들에 비해서 그다지 나쁘지 않았다[4, 10]. SpamAssassin과 같이 공개된 대부분의 쓰레기 편지 여과 시스템에는 나이브 베이지안 학습 기능을 가지고 있다[11]. 규칙기반 기계학습 방법[12]에서는 쓰레기 편지들이 공통적으로 가지고 있는 단어를 찾아내어 이를 쓰레기 편지를 판단하는 기준으로 사용하였다. 이는 수동으로 여과

2) ITU(International Telecommunication Union)

3) 미화 250억 달러: 한화 약 25조원

4) SpamBayes(http://spambayes.sourceforge.net/), SpamAssassin(http://spamassassin.apache.org/), SpamPal(http://www.spampal.org), 등.

5) 불법스팸대응센터(http://www.spamcop.or.kr/)는 정보통신부 산하기관인 한국정보보호진흥원에서 운영하고 있다.

6) 쓰레기 편지의 근원지는 유해 사이트(harmful site) 등이 여기에 속한다.

〈표 1〉 편지 말뭉치 현황

말뭉치 이름	URL	말뭉치 크기(전자편지 수)	상태	특징
SpamArchive	www.spamarchive.org	222,506	완료	쓰레기 편지
SpamAssassin public mail corpus	spamassassin.apache.org	약 100,000	완료	쓰레기 편지, 정보성 편지
Corpus of Junk Emails	clg.wlv.ac.uk	N.A	완료	쓰레기편지
Toasted Spam File	www.toastedspam.com	약 120,000	진행중	쓰레기 편지
Spam Hall of Shame	www.sput.nl	약 500,000	진행중	쓰레기 편지
Dolphinwave archive of spam	www.dolphinwave.org	약 90,000	진행중	쓰레기 편지
Spam Honeypot Archive	schnarff.com	N.A.	진행중	쓰레기 편지
The Enron corpus	www.cs.cmu.edu/~enrom	약 600,000	진행중	정보성 편지 뉴스 그룹 데이터

규칙을 입력하여 쓰레기 편지를 여과하는 방법과 매우 유사한 방법이다. 이 밖에도 지지벡터기계(support vector machine) 분류자[13-14], 사례기반(instance- or memory-based) 분류자[15] 등이 쓰레기 편지 여과에 두루 사용되고 있다.

2.2 편지 학습 말뭉치

쓰레기 편지 여과 시스템의 객관적인 성능을 평가하기 위해서 전자편지 말뭉치(e-mail corpus)가 필요하다. 영어의 경우에는 <표 1>에서 보는 바와 같이 공개된 여러 종류의 전자편지 말뭉치가 있으나, 한국어의 경우에는 공개된 전자편지 말뭉치가 존재하지 않는다. 따라서 한국어 전자편지를 대상으로 쓰레기 편지 여과 시스템을 객관적으로 평가하기는 불가능한 실정이다. 영어의 경우에도 대부분의 편지 말뭉치들은 쓰레기 편지인지 정보성 편지인지에 관련된 정보만 포함하고 있으며, 이 논문과 관련된 사용자 행동에 관련된 정보는 포함되어 있지 않다. 따라서 이 논문에서는 사용자 행동 정보의 유용성을 보이기 위해서 한국어 전자편지 말뭉치를 구축하였으며, 이 말뭉치는 쓰레기 편지에 관련된 정보뿐 아니라 사용자 행동에 관련된 정보도 함께 포함되어 있다. 구체적인 내용은 4장에서 다룰 것이다.

2.3 사용자 행동의 활용

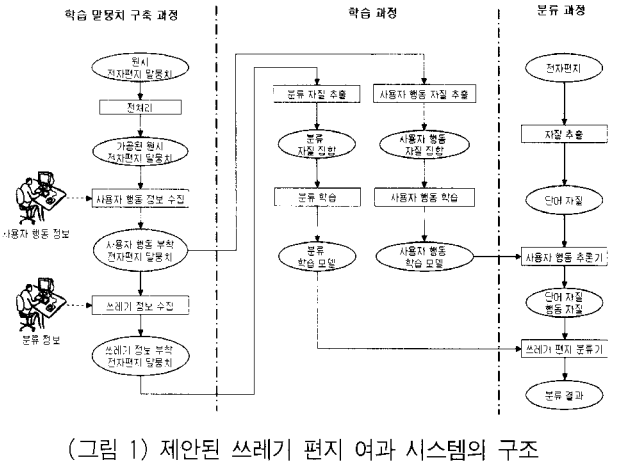
인간과 컴퓨터 상호작용(human-computer interaction)에서는 사용자의 행위 혹은 행동이 중요한 정보가 된다. 사용자의 행동에는 문서를 읽거나 머무는 시간[16], 마우스의 클릭 동작이나 스크롤 동작[17], 문서의 저장 혹은 삭제[18] 등이 있다. 최근 쓰레기 편지 여과 시스템에서 사용자의 행동을 <표 2>와 같이 검토(examination), 저장(retainment), 참조(reference), 주석 첨가(annotation)로 분류하여, 이를 묵시적 피드백(implicit feedback)의 증거 데이터로 활용하는 방법을 제안하였다[18]. <표 2>를 보면 문서의 일부분을 복사하거나 붙이는 행동을 참조라고 한다. 이 논문에서는 <표 2>의 분류법을 참조하여 사용자 행동을 새로 정의하고, 이를 이용한 쓰레기 편지 여과 시스템을 구현한다.

〈표 2〉 사용자 행동의 분류법

사용자 행동	문서 일부분	한 개의 문서	문서 묶음
검토	보기, 듣기	선택	
저장	출력	북마크, 저장, 획득, 삭제	동의
참조	복사, 붙이기, 인용	이동, 답변, 링크 걸기, 참조하기	
주석 달기	기호로 표시	문서 평가, 출판	재배치

3. 사용자 행동을 이용한 쓰레기 편지 여과 시스템

쓰레기 편지 여과 방식에서 opt-out 방식은 수신된 전자편지에 대해서 쓰레기 편지의 여부를 결정하며, 쓰레기 편지의 여부는 수신자에 따라서 다르다. 더구나 쓰레기 편지와 정보성의 편지에 따라서 수신자의 행동이 크게 다르다. 이 논문에서는 각 수신자에 적합한 쓰레기 편지의 기준과 각 수신자의 행동을 학습하기 위해서 점진적인 기계학습 방법을 사용한다. (그림 1)은 이 논문에서 제안하는 사용자 행동을 이용한 쓰레기 편지 여과 시스템의 구조이다. 제안된 시스템은 크게 편지 말뭉치와 사용자 행동 정보를 이용한 학습 말뭉치 구축 과정, 사례기반 기계학습을 이용한 행동 추론 모델과 분류 모델을 배우는 학습 과정, 행동 추론 모델과 분류



(그림 1) 제안된 쓰레기 편지 여과 시스템의 구조

모델을 이용한 전자편지 분류 과정으로 나눌 수 있으며, 이하에서 이들에 대해서 자세히 설명할 것이다.

3.1 사용자 행동

이 논문에서는 <표 3>과 같은 다섯 개의 기본적인 사용자 행동을 정의한다. 각 전자편지는 여러 개의 사용자 행동을 동시에 지닐 수 있다. 예를 들면 전자우편을 읽고 나서 삭제하거나 전달할 수 있다. 이와 같은 사용자 행동이 묵시적으로 제안된 시스템에 반영되어 시스템의 성능을 점차적으로 개선한다.

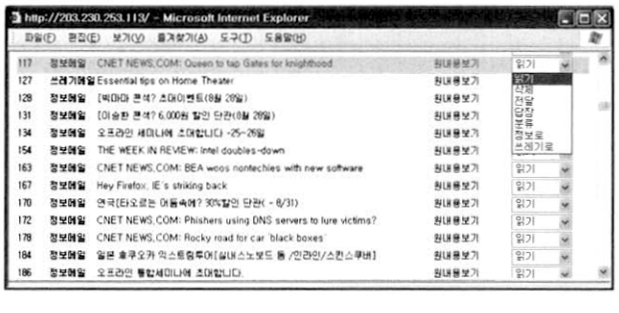
〈표 4〉 이 논문에서 정의한 사용자 행동

사용자 행동	정의
읽기(read)	수신된 전자편지를 열어서 읽는다.
삭제(delete)	수신된 전자편지를 제목만 확인하고 바로 지운다.
분류(classify)	수신된 전자편지를 하위폴더에 보관한다.
전달(forward)	수신된 전자편지를 다른 사람에게 전달한다.
답장(reply)	수신된 전자편지의 송신자에게 답장을 보낸다.

3.2 학습 말뭉치 구축 과정

이 논문에서 학습 말뭉치는 크게 원시 전자편지 말뭉치(raw e-mail corpus)와 사용자 행동 부착 전자편지 말뭉치(user-action tagged e-mail corpus) 그리고 쓰레기 정보 부착전자편지 말뭉치(spam tagged e-mail corpus)로 구성된다. 원시 전자편지 말뭉치는 전혀 가공이 되지 않은 전자편지 그 자체이다. 사용자 행동 부착 전자편지 말뭉치는 원시 전자편지 말뭉치에 속한 전자편지를 수신했을 때, 사용자가 취한 행동에 관련된 정보가 부착된 말뭉치이고, 쓰레기 정보 부착 전자편지 말뭉치는 원시 전자편지 말뭉치에 속한 전자편지가 쓰레기 편지인지에 관련된 정보가 부착된 말뭉치이다.

학습 말뭉치의 구축은 전처리 과정, 사용자 행동 정보 수집 과정, 쓰레기 정보 수집 과정으로 이루어진다. 전처리 과정은 전자편지의 헤더(header) 정보를 추출하고, HTML 태그를 제거한다. 사용자 행동 정보 수집 과정과 쓰레기 정보 수집 과정은 이 논문에서 특별히 제작된 전자편지 말뭉치 구축 도구 (그림 2)를 통해서 원시 전자편지 말뭉치에 사용자 행동 정보와 쓰레기 정보를 부착한다. (그림 2)의 제일 오른쪽 칸을 통해서 사용자 행동 정보와 쓰레기 편지 정보를 입력한다. 구축된 전자편자 말뭉치의 규모 등에 관련된 자세한 내용은 4장에서 기술할 것이다.



(그림 2) 사용자 행동 정보와 쓰레기 정보를 부착하기 위한 전자편지 말뭉치 구축 도구

3.3 학습 과정

(그림 1)에서 보는 바와 같이 이 논문에서는 두 종류의 학습 모델, 사용자 행동 추론 모델과 쓰레기 편지 분류 모델을 사용한다. 사용자 행동 추론 모델은 원시 전자편지로부터 사용자의 행동을 추론하기 위한 모델이고, 쓰레기 편지 분류 모델은 추론된 사용자의 행동을 이용하여 전자편지를 분류하기 위한 모델이다. 각 모델은 자질 추출과 모델 생성 과정을 통해서 생성된다. 각 모델은 사례기반 기계학습 도구인 TiMBL-IB2[19]을 사용한다.

두 학습 모델은 서로 다른 자질을 사용한다. 사용자 행동 추론 모델의 자질은 전자편지의 본문에 포함된 그림의 수와 전자편지의 헤더와 본문에 출현한 명사들이며, 추출된 자질의 자질값은 빈도수이다. 헤더와 본문에 출현된 명사는 서로 다른 것으로 간주되며, 자질의 수를 줄이기 위해서 두 자로 구성된 명사만을 선정한다. 이와 같은 방법으로 추출된 자질은 사용자마다 서로 다르다. 쓰레기 편지 분류 모델의 자질은 사용자 행동 추론 모델의 자질과 사용자 행동(읽기, 삭제, 분류, 전달 답장)이며, 사용자 행동은 이진 자질이다.

학습 모델의 생성은 제안된 시스템의 두 학습 모델이 점진적 사례기반 기계학습 도구인 TiMBL에서 제공하는 IB2 알고리즘을 이용하여 점진적으로 개선된다.

3.4 분류 과정

먼저 입력된 편지로부터 사용자 행동 추론 모델을 생성하기 위한 자질을 추출한다. 추출된 자질과 TiMBL을 이용해서 가능한 사용자 행동을 추론한다. 추론된 사용자 행동은 0 ~ 31까지의 숫자로 표현된다. 이는 기본적인 사용자 행동을 이진화하여 표현한 것이다. 예를 들면 추론된 행동이 18이라면 "10010"으로 표현되며, 읽고 전달되었음을 의미한다. 이렇게 추론된 사용자 행동 자질과 TiMBL을 이용해서 입력된 전자편지가 쓰레기 편지인지를 결정한다.

4. 실험 및 평가

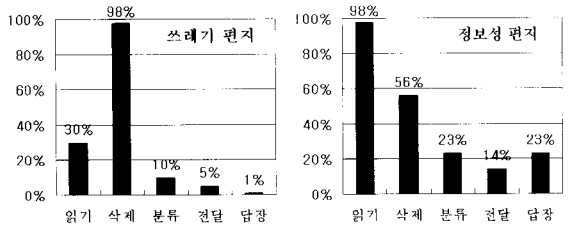
4.1 전자편지 말뭉치

학습 및 실험에 말뭉치로 사용한 편지는 한메일[7]에서 제공하는 편지 백업 기능을 이용하여, 10명의 수신자[8]로부터 추출한 10,000통의 원시 전자편지 말뭉치를 구축하였다. 전자편지의 수집 기간은 2005년 3월부터 6월까지 3개월이었으며, 각 수신자는 한메일에서 제공하는 쓰레기 편지 여과 기능을 사용하지 않고 편지 데이터를 수집하였다.

사용자 행동 정보는 3장에서 설명한 말뭉치 구축 도구를

7) 한메일 인터넷 주소: www.daum.net

8) 전자편지 수집자는 한국해양대학교 IT공학부 학부생 및 대학원생이며, 각 사용자들에게 개별적인 동의를 구했다.



(그림 3) 구축된 전자편지 말뭉치에서 사용자 행동의 분포

이용해서 부착되었다. 이 때 2명의 원시 전자편지 말뭉치에 대해서 서로 독립적으로 사용자 정보와 쓰레기 정보를 부착하도록 하였다. 즉, 같은 원시 전자편지에 대해서 서로 다른 사용자 행동이 쓰레기 편지 여과에 어떤 영향을 주는지를 관찰하기 위해서 이와 같은 작업을 수행하였다. 따라서 구축된 사용자 행동 부착 전자편지 말뭉치와 쓰레기 정보 부착 전자편지 말뭉치는 각 12,000통의 전자편지로 구성되었다.

각 사용자[9]는 1,000통의 전자편지를 대상으로 사용자 행동 정보가 부착되었으며, 그 중에서 900통은 학습 데이터로, 100통은 실험 데이터로 나누어 사용하였다. 부착된 사용자 행동 정보는 (그림 3)과 같은 분포를 가진다.

(그림 3)에서 쓰레기 편지의 경우 삭제 행동이 뚜렷했으며, 정보성 편지의 경우 읽기 행동이 뚜렷하게 나타났다. 또한 쓰레기 편지의 경우 삭제 행동과 이외의 행동의 차이가 큰 반면, 정보성 편지의 경우 가장 높은 행위인 읽기와 더불어 비교적 다양한 행동 패턴을 보였다.

4.2 평가 측도

이 논문에서는 식 (4.1)과 같은 정확도 A를 이용해서 쓰레기 편지 여과 시스템의 성능을 평가하였다.

$$A = \frac{E}{N} \tag{4.1}$$

여기서 N은 실험 데이터의 총 수이고, E는 정확하게 분류한 결과의 개수이다.

4.3 성능 평가와 분석

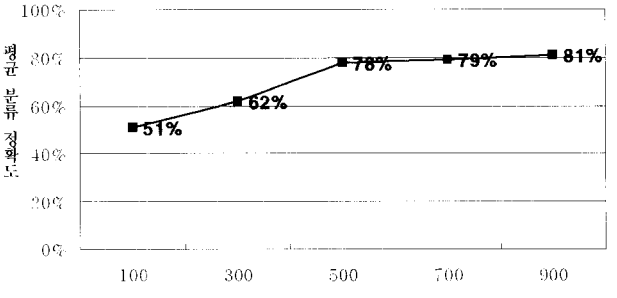
4.3.1 학습 데이터 양에 따른 분류 정확도

<표 4>는 각 사용자에 대해서 학습 데이터 양의 변화에 따른 분류 정확도를 관찰하였다. 각 사용자의 학습 데이터를 100통에서 900통까지 200통씩 점진적으로 증가시켜서 분류 정확도를 측정했는데, 대부분의 경우에는 학습 데이터 양의 증가에 따라 분류 정확도가 증가하였으나, 일부 사용자에 대해서는 약간의 진동이 있었다. (그림 4)는 학습 데이터 양의 변화에 따른 평균 분류 정확도의 변화를 보여주고 있다. (그림 4)에서 **사용자 1**과 **사용자 11**은 같은 원시 전자편지이지만 여과 성능은 각각 70%와 73%로 차이를 보였다. **사용자 2**와 **사용자 12**도 같은 원시 전자편지이지만 서로 다른 성능을

9) 이 논문에서 사용자는 전자편지를 수신하여 사용자 행동 정보 및 쓰레기 정보를 부착한 사람을 말한다.

〈표 4〉 학습 데이터 양에 따른 분류 정확도(단위 %)

사용자	학습 말뭉치의 전자편지 수					평균
	100	300	500	700	900	
1	40	72	80	81	79	70
2	50	65	73	71	71	66
3	*53*	*83*	*82*	*93*	*93*	*81*
4	48	75	76	72	67	67
5	44	56	88	91	87	73
6	64	78	75	80	86	76
7	25	53	75	69	64	57
8	53	64	78	88	83	73
9	*35*	*43*	*65*	*68*	*67*	*56*
10	41	65	64	64	75	62
11	34	76	88	87	82	73
12	56	66	76	83	85	73
평균	51	62	78	79	81	67



(그림 4) 학습 데이터 양에 따른 분류 정확도 변화

보였다. 이처럼 사용자가 취하는 행동이 쓰레기 편지에 크게 영향을 주었다.

4.3.2 사용자 행동이 여과 성능에 미친 영향

<표 4>에서 살펴보면 **사용자 3**이 가장 높은 분류 정확도를, **사용자 9**가 가장 낮은 분류 정확도를 보였다. <표 5>는 두 사용자의 사용자 행동 패턴을 볼 수 있다. 높은 정확도를 보인 **사용자 3**은 쓰레기 편지와 정보성 편지에 대해 명확한 행동 차이를 보였다. 예를 들어 쓰레기 편지는 경우 대부분 읽지 않고 삭제하였으며, 정보성 편지에 대해 대체로 회신하거나 분류(다른 편지함으로 이동)하는 행동을 보였다. 반면

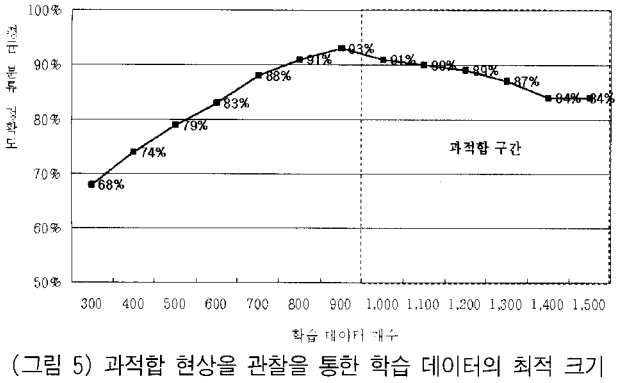
〈표 5〉 학습 데이터의 사용자 행동 패턴(단위 %)

사용자 행동	사용자 3		사용자 9	
	쓰레기	정보성	쓰레기	정보성
읽기	13	96	84	99
삭제	92	53	99	89
분류	0	26	3	4
전달	0	11	0	6
답장	0	34	0	11

사용자 9의 경우 비교적 쓰레기 편지와 정보성 편지에 대한 구분이 모호하였는데, 예를 들어 정보성 편지에 대해 분류나 회신 또는 전달 작업 없이 읽은 후 삭제하는 행동 패턴이 다수 있었다. 이와 같은 사용자 행동 패턴이 쓰레기 편지 분류에 나쁜 영향을 주었다고 판단된다.

4.3.3 학습 데이터의 최적 크기

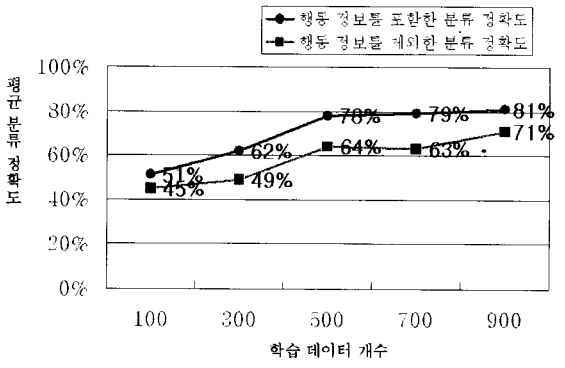
4.3.1절에서 언급했듯이 일부 사용자의 경우, 학습 데이터가 증가해도 분류 정확도가 떨어지는 현상이 발생하였다. 이와 같은 현상을 일반적으로 기계학습에서는 과적합(over-fitting)이라고 한다[8]. 이 절에서는 제안된 시스템의 과적합 현상을 관찰하기 위해서 10,000통의 학습 데이터에서 사용자의 정보를 무시하고 임의로 1,500통의 학습 데이터를 추출하여 분류 정확도를 관찰하였다(그림 5). 그림에서 보는 바와 같이 학습 데이터 양이 900통에 1,000통 사이에 과적합 현상이 나타났으며, 이 결과를 통해서 제안된 시스템의 적절한 학습 데이터의 양은 약 900통 정도임을 알 수 있었다.



(그림 5) 과적합 현상을 관찰을 통한 학습 데이터의 최적 크기

4.5 행동 정보의 유용성 평가

이 절에서는 사용자 행동이 쓰레기 편지 여과에 어느 정도 유용한지를 관찰하려고 한다. (그림 6)은 **사용자 3**에 대해서 사용자 행동의 사용 여부에 따른 여과 성능을 보이고 있다. (그림 6)에서 볼 수 있듯이 학습 데이터 양에 따라 최소 6%에서 최대 16%의 분류 정확도가 향상되었음을 확인할 수 있다. 결론적으로 사용자 행동이 쓰레기 편지 여과에 매우 효과적인 자질임을 알 수 있었다.



(그림 6) 사용자 행동의 사용 유무에 따른 분류 정확도

4.6 기존 쓰레기 편지 여과 시스템과의 비교

이 절에서는 제안된 쓰레기 편지 여과 시스템과 기존의 쓰레기 편지 여과 시스템들과의 비교·분석하고자 한다. <표 6>은 기존 쓰레기 편지 여과 시스템들과 제안한 시스템의 특징을 나타낸 것이다.

<표 6>에서 시스템 I[20]과 II[21]는 나이브 베이지안 분류 방법, 시스템 III[13]은 SVM을 이용하여 편지를 분류하였다. 시스템 IV는 이 논문에서 제안된 시스템이다. <표 6>에서 볼 수 있듯이 각 시스템은 자질, 분류 방법, 대상 언어 학습 및 실험 데이터의 양이 모두 다르기 때문에 각 시스템의 성능을 객관적으로 비교할 수 없다. 이는 2장에서 설명하였듯이 공개 한국어 전자편지 말뭉치가 없기 때문이다. 시스템 II는 잘못된 여과 결과에 대해 사용자가 피드백 정보[11]를 입력할 수 있는 인터페이스를 제공하지만, 학습 자질로 사용하지 않았다.

5. 결론 및 앞으로의 연구 과제

이 논문에서는 사용자의 행동 정보를 이용해서 쓰레기 편지 여과 시스템의 성능을 개선하였다. 학습 단계에서는 사례기반 학습기를 이용하여 사용자 행동 추론 모델과 분류 모델을 생성하고, 분류 단계에서는 대상 편지를 입력으로 사용자 행동을 추론하고, 이를 쓰레기 편지 여과 시스템의 입력 자

10) (그림 5)에서 보이고 있는 여과 성능이 다른 시스템과 최대한 비슷한 환경에 실험되었으므로 그 결과를 실었다.
11) 쓰레기 편지함에서 정보성 편지함으로 보내기와 그 역에 관한 일을 할 수 있는 인터페이스를 제공한다.

〈표 6〉 쓰레기 편지 여과 시스템의 비교[10]

시스템	특징		편지 말뭉치			분류 정확도(%)	사용자 행동 정보 사용
	분류 자질	분류 방법	언어	학습데이터	실험데이터		
I	제목, 본문	나이브 베이지안	한국어 영어	3538	1,517	95	X
II	제목, 본문, HTML 링크, 특정 HTML Tag와 같이 쓰인 단어	나이브 베이지안	영어	576 (329+247)	201 (148+53)	94	X
III	헤더, 본문	SVM (SVM Light)	한국어 영어	666 (441+225)	153 (100+53)	91	X
IV	제목, 본문, 본문에 포함된 이미지 개수, 편지에 대한 사용자의 행동 정보	사례기반 학습 (TiMBL)	한국어	9,000 (7,534+2,466)	1,000 (322+678)	93	O

질로 사용하여 쓰레기 편지 여부를 결정한다. 제안된 시스템의 성능을 평가하기 위해 12명의 사용자에 대해서 12,000통의 전자편지를 수집하였다. 사용자에 따라 조금씩 차이를 보이지만 학습 데이터가 900통일 때, 평균 약 81%의 정확도를 보였다. 또한 사용자 행동을 이용한 쓰레기 편지 여과 시스템이 그렇지 않은 시스템에 비해서 6% ~ 14%의 성능 향상을 보였다. 향후에는 제안된 시스템에서 정의한 사용자의 행동 외의 다양한 사용자의 행동 패턴을 연구하여 쓰레기 편지 분류 작업에 활용하고, 사례기반 학습 이외의 다양한 쓰레기 편지 분류 기법을 이용하여 사용자의 행동을 이용한 여과 성능 향상을 꾀할 수 있을 것이다. 무엇보다 쓰레기 편지 여과 시스템의 객관적이고 정확한 성능 평가를 위해 공개 한국어 편지 말뭉치 구축 작업이 필요하다. 또한 쓰레기 편지 여과뿐만 아니라 웹 문서 분류와 도서 추천 시스템 등의 다양한 분야에 사용자의 행동 정보를 활용하여 더욱 향상된 시스템을 구현할 수 있을 것으로 기대된다.

참 고 문 헌

[1] 한국정보보호진흥원, 알기 쉬운 스팸 대응 현황 자료집, http://www.kisa.or.kr/index.jsp, 2004.

[2] Sorkin, D. E.,"Technical and legal approaches to unsolicited electronic mail", San Francisco University Raw Review, vol. 35, pp.334, 2001.

[3] ITU, SPAM in the Information Society: Building Frameworks for International Cooperation, http://www.itu.int/osg/spu/publication/#2004, 2004.

[4] Zhang, L., Zhu, J. and Yao, T. "An evaluation of statistical spam filtering techniques", ACM Transactions on Asian Language Information Processing, vol. 3, No.4, pp.243-269, 2004.

[5] Tretyakov, K. "Machine Learning Techniques in Spam Filtering," Institute of Computer Science, University of Tartu Data Mining Problem-oriented Seminar, MTAT, Vol.3, pp.60-79, 2004.

[6] Cranor, L. F., and LaMacchia, B. A. "Spam!", Communications of ACM, Vol.41, No.8, pp.74-83, 1998.

[7] Wolfe, P., Scott C., and Erwin M. W. (2004), Anti-SPAM Toolkit, McGraw-Hill/Osborne.

[8] Mitchell, T. M., Machine Learning, McGraw-Hill, 1997.

[9] 이상호, "자동 생성 메일계정 인식을 통한 스팸 필터링", 정보과학회 논문지: 소프트웨어 및 응용, Vol.32, No.5, pp.378-384, 2005.

[10] Androutsopoulos, I., Koutsias, J., Chandrinos, K. V. and Spyropoulos. C. D. "Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach," Proceedings of the Workshop on Machine Learning and Textual Information Access, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp.1-13, 2000b.

[11] Schwartz, A., SpamAssassin, O'Reilly, 2004.

[12] Cohen, W. W., "Learning rules that classify e-mail," Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access, pp.18-25, 1996.

[13] 민도식, 송무희, 손기준, 이상조, "SVM 분류 알고리즘을 이용한 스팸 메일 필터링", 한국정보과학회 2003년 춘계학술대회 발표논문집, Vol.30, No.1, pp.552-554, 2003.

[14] Drucker, H. D., Wu, D. and Vapnik, V., "Support vector machines for spam categorization", IEEE Transactions on Neural Networks, Vol.10, No.5, pp.1048-1054, 1999.

[15] Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. and Stamatopoulos, P., "A memory-based approach to anti-spam filtering for mailing lists", Information Retrieval, Vol.6, pp.49-73, 2003.

[16] Morita, M. and Shinoda, Y., "Information filtering based on user behavior: Analysis and best match text retrieval", Proceedings of SIGIR, pp.272-281, 1994.

[17] Goecks, J. and Shavlik, J.,"Learning user's interests by unobtrusive observing their normal behavior", Proceedings of The 5th International Conference on Intelligent User Interfaces, pp.129-132, 2000,

[18] Kim, J. and Oard, D. W., "Observable behavior for implicit user modeling: A framework and user studies", Journal of the Korean Society for Library and Information Science, Vol.35, No.3, pp.173-189, 2001

[19] Daelemans, W., Zavrel, J. and Ko, van der S., TiMBL: Tilburg Memory-Based Learner Version 5.1 reference guide, Tilburg University, ILK Technical Report, ILK-0104, 2004.

[20] 임정택, 김형준, 강승식, "나이브 베이지안 분류자와 메일 주소 유효성 검사를 이용한 스팸 메일 필터링 시스템", 한국정보과학회 2005년 춘계학술대회발표논문집, Vol.32, No.2, pp.523-525, 2005.

[21] 김현준, 정재은, 조근식, "가중치가 부여된 베이지안 분류를 이용한 스팸 메일 필터링시스템", 정보과학회 논문지: 소프트웨어 및 응용, Vol.31, No.8, pp.1092-1100, 2004.

김 재 훈

e-mail : jhoon@mail.hhu.ac.kr
1986년 계명대학교 전자계산학과(학사)
1988년 한국과학기술원 전산학과 (공학석사)
1996년 한국과학기술원 전산학과 (공학박사)
1988년~1997년 한국전자통신연구원, 선임연구원
1997년~1999년 한국해양대학교, 컴퓨터공학과, 전임강사
2001년~2002년 USC, Information Sciences Institute, 방문연구원
1999년~현재 한국해양대학교, 컴퓨터공학과, 부교수
관심분야 : 자연언어처리, 한국어정보처리, 정보검색, 정보추출

김 강 민

e-mail : kkangmin@gmail.com
2003년 한국해양대학교 자동화정보공학부 (학사)
2006년 한국해양대학교 컴퓨터공학과 (공학석사)
2006년 3월~현재 (주)태광ENG 연구소 연구원
관심분야 : 스팸메일 필터링, 정보 검색, 기계학습