工學碩士 學位論文

사용자 행동과 점진적 기계학습을 이용한 쓰레기 편지 여과 시스템의 설계 및 구현

Design and Implementation of SPAM Filtering System Using User Action and Incremental Machine Learning

指導敎授 金 載 熏

2006年 2月

韓國海洋大學校 大學院

컴 퓨 터 工 學 科

金 江 珉

本 論文을 金江珉의 工學碩士 學位論文으로 認准함

委員長　工學博士　柳 吉 洙　印

委　員　工學博士　辛 沃 根　印

委　員　工學博士　金 載 熏　印

2005年　12月

韓國海洋大學校 大學院

컴퓨터工學科　金 江 珉

목 차

표 목차

그림 목차

Design and Implementation of SPAM Filtering System Using User Action and Incremental Machine Learning

Kang-min Kim

Department of Computer Engineering, Graduate School
Korea Maritime University, Busan, Korea

Abstract

With rapidly developing Internet applications, an e-mail has been considered as one of the most popular methods for exchanging information because of easy usage and low cost. The e-mail, however, has a serious problem that users can receive a lot of unwanted e-mails, what we called, SPAM mails, and then the user's mailbox can be grown exponentially. The users need for spending time to pick out the SPAM mails and give a great loss economically. To alleviate the problem, many researchers and companies proposed some filtering technologies.

On the other hand, in e-mail client systems, users do different actions according to usefulness of information on mails, and some classification and recommendation systems like GroupLens use the actions to improve the performance. This paper presents a mail filtering system using user actions and

incremental machine learning. E-mail data and user actions are collected through some user interface implemented in CGI/Perl. Our proposed system makes use of two models: One is anaction inference model to draw a user action from an e-mail and the other is a mail classification model to decide if an e-mail is SPAM or not. All the two models are derived using incremental learning, of which an algorithm is IB2 of TiMBL.

To evaluate our proposed system, we collect 10,000 mails of 12 persons from Hanmail (www.hanmail.net), which is one of the most popular e-mail service providers in Korea. The accuracy is 81 ~ 93% according to each person. Our proposed system outperforms a system that does not use any information about user actions. Consequently, we have shown that information about user actions is useful for e-mail filtering

제 1 장 서 론

월드 와이드 웹(WWW: World Wide Web)을 중심으로 하는 인터넷의 급속한 성장과 더불어 전자 편지 서비스(e-mail service)는 이제 의사교환의 필수적인 매체로 사용되고 있다. 그러나 전자 편지 서비스를 악용한 '쓰레기 편지(SPAM mail)' 는 사회적인 문제점으로 대두되고 있다. 여기서 쓰레기 편지는 '수신자가 원하지 않는 전자 편지' 이므로 수신자에 따라 그 정의가 달라질 수 있다. 그럼에도 불구하고 쓰레기 편지에는 다음과 같은 공통적인 성질이 있다. 일반적으로 쓰레기 편지는 수신자가 원하거나 요청하지 않았다는 성질, 영리 목적의 상업성, 그리고 대량성을 공통적으로 지닌다(Sorkin, 2001). 쓰레기 편지는 우선 수신자를 성가시고 짜증나게 하고, 전자 편지 서비스 제공 업체에게는 저장 장치의 용량을 부족하게 만드는 문제를 일으킨다. 경제적 측면에서 쓰레기 편지의 폐해는 다음과 같다. ITU[1]의 조사보고서에 따르면, 2003년 한 해 전 세계적으로 쓰레기 편지로 인해 발생한 경제 손실 비용이 약 25조 원[2]에 달하는 것으로 추정되었다(ITU, 2004). 같은 해 쓰레기 편지로 인한 국내 피해액 역시 약 1조 3천억 원에 이르는 것으로 추정되었다(한국정보보호진흥원, 2004).

쓰레기 편지가 심각한 문제로 부각되자 각국 정부에서는 법적, 제도적 대책을 수립하고, 쓰레기 편지 규제에 적극적으로 나서기 시작하였다. 국내의 경우 정보통신부의 관장 하에 '정보통신망이용촉진및정보보호등에관한법률' 을 제정하고, 정보통신부 산하기관인 한국정보보호진흥원에서는 2003년부터 '불법스팸대응센터[3]' 를 운영하고 있다. 또한 각 연구기관 및 정보통신 기업체들은

---

1) ITU(International Telecommunication Union)
2) 미화 250억 달러: 한화 약 25조원
3) 불법스팸대응센터 인터넷 주소: http://www.spamcop.or.kr/spamcop.html

쓰레기 편지에 대한 기술적인 대처 방안을 연구하고 관련 시스템을 제안하였다(한국정보보호진흥원, 2004).

이와 같은 문제를 보완하기 위해 이 논문에서는 쓰레기 편지와 정보성 편지에 대한 사용자들의 행동(action)이 각각 상이하다는 점에 착안하여 사용자의 행동을 쓰레기 편지 분류를 위한 자질로 사용하는 쓰레기 편지 여과 시스템을 제안한다. 사용자 행동의 예를 들면 쓰레기 편지로 간주되는 편지는 읽지 않고 편지 제목만 확인한 후 바로 삭제할 수 있다. 반대로 중요한 정보성 편지일 경우 따로 보관함을 만들어서 보관하거나, 다른 사람에게 전달 또는 답장을 보내는 행동을 할 수 있다. 제안하는 시스템은 이러한 사용자의 행동을 저장하여 쓰레기 편지/정보성 편지 분류를 위한 사례기반 학습(Daelemans & Zavrel, 2001)의 특징으로 활용하는 것이다. 이는 편지 내에서 추출할 수 있는 정보만을 이용하여 편지를 분류하는 방법에 비해 더 나은 성능을 기대할 수 있다.

이 논문의 구성은 다음과 같다. 2장은 관련연구로 현재 활용하고 있는 쓰레기 편지 여과 기법들과 쓰레기 편지 여과에 사용되고 있는 기계학습 방법, 그리고 사용자의 묵시적 피드백의 개념과 편지 말뭉치에 대해 정리하였다. 3장에서는 이 논문에서 제안하는 사용자 행동 정보를 이용한 편지 여과 시스템의 구성 및 주요 모듈의 개념을 설명하고, 4장에서는 제안된 시스템의 성능을 평가한다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 제시하고자 한다.

제 2 장 관련 연구

이 논문에서 제안하는 사용자의 행동을 통한 쓰레기 편지의 여과 시스템은 사용자의 행동을 모델링하고 학습하여 쓰레기 편지와 정보성 편지를 분류한다. 이 장에서는 일반적으로 쓰레기 편지 여과작업에 사용되는 기술적 대응 방법들을 살펴보고, 쓰레기 편지 여과작업에 활용하고 있는 대표적인 기계학습들에 대한 연구와 사용자의 묵시적 피드백(implicit feedback), 그리고 쓰레기 편지 여과 시스템의 성능 평가를 위한 편지 말뭉치에 대해 소개하고자 한다.

2.1 쓰레기 편지 차단을 위한 기술적 대응 방법

이 절에서는 쓰레기 편지 문제를 해결하기 위해 많은 연구 기관 및 업체들이 제안하고 활용하는 기술적 대응방안에 대해 소개하고자 한다.

2.1.1 편지 주소 수집 차단 기술

편지 주소 수집 차단 기술은 쓰레기 편지 발송자들이 자동 편지 주소 수집 프로그램을 이용하여 사용자들의 편지 주소를 대량으로 수집하지 못하도록 하는 기술이다. 편지 주소 수집 차단 기술의 대표적인 예는 e-mail masking 기법이다. 이 기법은 편지 주소를 여러 가지 형태로 변환하여 자동 편지 수집 프로그램이 사용자의 편지 주소를 수집하지 못하도록 하는 대표적인 방법이다(한국정보보호진흥원, 2004). e-mail masking에는 자바스크립트 변환 방식, ASCII 코드 변환 방식, 편지 주소의 이미지 파일 생성 방식 등이 있다. 이들 방식의 공통점은 편지를 브라우저에 출력하기 전에 여러 가지 형태로 변형한다는 것이

다. 자바스크립트 변환 방식은 자바스크립트를 통해 편지의 ID, @, 도메인을 각각 분리하여 자동 프로그램이 추출할 수 없도록 하는 기술이다. ASCII 코드 변환 방식은 편지 주소 부분을 모두 ASCII 코드로 변환하는 기술인데, 편지 주소 수집 프로그램이 이를 복호화하기 위해서는 ASCII 값을 모두 영문 또는 기타 문자로 재변환해 주어야 한다. 마지막으로 편지 주소의 이미지 파일 생성 방식은 편지 주소의 추출을 방지하는 가장 확실한 방식으로 모든 문자를 각각의 이미지 파일로 만들고, 그들을 조합하여 편지 주소를 출력하는 방식이다.

2.1.2 대량 쓰레기 편지 발송 대응 기술

대량 쓰레기 편지 차단 기술은 쓰레기 편지가 대량으로 발송되는 것을 제재하기 위한 일련의 기술적 대처방안이다(Graham, 2002). 이 기술의 대표적인 예로 HIP(Human Interactive Proof) 기술과 인터넷 편지 우표제가 있다. HIP 기술은 사람은 풀 수 있지만 기계적인 프로그램은 쉽게 풀지 못하는 인지적 단계의 퍼즐을 제공하는 방식이다. 이 기술은 쓰레기 편지 발송자들이 사이트 자동 가입 프로그램을 이용하여 대형 편지 서비스 업체의 계정을 대량으로 만든 후 이를 통해 쓰레기 편지를 발송하는 것을 차단하기 위한 것이다. 그림 2.1에서 확인할 수 있듯이 아주 단순한 단어를 무작위로 골라 글자를 약간 훼손시킨 뒤 복잡한 배경 화면 위에 표시한다[4](한국정보보호진흥원, 2004).

[4] CAPTCHA(Completely Automated Public Test to tell Computers and Humans Apart) 기술

그림 2.1 HIP의 예

Fig. 2.1 An example of HIP

인터넷 편지 우표제는 편지 서비스 업체가 대량의 편지 발송자에게 '우표'를 구입하게 하여 일정의 발송 비용을 부과하는 방식이다(한국정보보호진흥원, 2004). 우표를 첨부하지 않은 편지는 편지 서비스업체 등의 서버에서 사용자에게 전달하지 않고 휴지통에 버려지게 된다. 따라서 발송자는 편지의 양이 늘어날수록 발송 비용이 증가하게 되고 때문에 대량의 쓰레기 편지 발송을 억제하는 효과가 있다.

2.1.3 쓰레기 편지 발송자 신원 확인 기술

발송자 신원 확인 기술은 쓰레기 편지를 차단하기 위해 발송자의 신원을 확인하여 검증되지 않은 발송자로부터 발송된 편지는 쓰레기 편지로 분류하는 기술이다. 이 기술의 대표적인 예는 caller-ID, challenge/response, PKI(Public Key Infrastructure)를 이용한 방식이 있다.

caller-ID 방식은 미국 Microsoft사에서 고안한 기술로 DNS(Domain Name Server) 등록정보를 통해 수신되는 편지가 신뢰할 만한 발송자가 보낸 것인지를 확인할 수 있는 기술이다. 편지 발송자는 특수한 caller-ID로 형식으로 DNS에 자신의 IP주소들을 공개하고, 수신자는 각 편지가 믿을 만한 도메인에서 온

것인지를 평가하기 위해 신뢰할 수 있는 발송 서버의 IP리스트를 요청하여 해당 IP가 리스트 내에 존재하면 정상적인 편지로 판별한다. challenge/response 방식은 수신자 측의 편지 서버가 일단 편지를 받은 후 발송자의 신원을 확인하는 과정을 거친 후에 확인된 편지만을 수신자에게 전달하는 방식이다. PKI 방식은 발송지가 편지에 대해 디지털 서명을 한 후 편지를 발송하고, 이에 대한 공개키를 제3의 검증기관에 등록하여 수신자가 수신된 편지의 발송자 정보와 그 내용이 위(변)조된 것이 아닌지 공개키 검증기관을 통해 확인할 수 있도록 하는 방식이다.

2.1.4 쓰레기 편지 여과 기술

쓰레기 편지 여과 기술은 분류 대상인 편지를 분석하여 쓰레기 편지의 여부를 판단하는 것이다. 이 기술은 수신자가 동의하지 않은 편지를 사전에 차단하는 것인지, 수신된 편지가 쓰레기 편지인지 아닌지의 여부를 식별하여 사후 차단하는 것인지에 따라 opt-in 방식과 opt-out 방식으로 나눌 수 있다(Mertz, 2002). opt-in 방식은 수신자가 수신을 원하는 편지 목록[5]을 편지 서비스 제공업체나 편지 관리 프로그램에 등록하여 자신이 등록하지 않은 곳에서 전송되는 모든 편지를 원천 차단하는 방식이다. opt-out 방식은 쓰레기 편지의 일반적인 특성을 추출하고 이를 이용하여 쓰레기 편지를 차단하는 작업이다. 이 논문이 제안하는 시스템은 opt-out 방식을 사용하므로 opt-out 방식에 대해 구체적으로 설명한다. opt-out 방식에서 사용하는 편지의 특성은 편지가 포함하는 단어 정보, 발송자 정보, 편지에 포함된 이미지 정보, 쓰레기 편지의 패턴 정보가 있다.

---

[5]수신자가 수신을 원하는 편지 목록(white list): 친구나 사업관계자 등 검증된 편지 주소 목록

편지가 포함하는 단어 정보를 특성으로 사용하는 여과작업은 사용자가 수신하는 쓰레기 편지에서 반복 사용되는 특정 단어를 찾아 이후 이 단어를 포함한 편지를 차단하는 작업이다. 발송자 정보를 이용한 차단 작업은 쓰레기 편지를 자주 보내는 전송자 편지 주소나 해당 편지 서버의 IP 주소, URL등의 정보를 저장하여 차단하는 방법이다. 하지만 전송자 정보를 위·변조하는 것이 용이한 현실에서 크게 실효성은 없다(이상호, 2005). 편지가 포함하는 이미지 정보를 이용한 여과작업은 특히 음란물 등을 걸러내는데 많이 활용될 수 있으나, 상대적으로 분류 정확도가 낮고 분류 작업을 위한 속도가 현저히 떨어진다는 단점이 있다. 쓰레기 편지의 패턴 학습 정보를 이용한 여과작업은 쓰레기 편지의 패턴을 지속적으로 인식하여 차단하는 작업을 말한다. 쓰레기 편지의 패턴으로 단어의 조합, 문장의 조합이나 출현 횟수의 누적치 및 가중치 등을 기반으로 쓰레기 편지의 여부를 판별하는 것인데, 많은 계산량을 요구한다는 단점이 있지만 비교적 분류 정확도가 높다(Graham, 2002).

2.2 기계학습을 이용한 쓰레기 편지 여과

쓰레기 편지 여과 시스템을 통과하기 위해 발송자들은 점차 지능적인 수법을 사용하여 쓰레기 편지를 발송하고 있다. 이에 따라서 쓰레기 편지 여과 시스템을 연구하는 사람들은 작업을 효과적으로 수행하기 위해 기계학습 기법을 사용하고 있다. 이 절에서는 최근 쓰레기 편지 여과 시스템에서 가장 널리 사용되고 있는 기계학습 기법 중에서 베이지안 분류를 이용한 기계학습, 지지벡터를 이용한 기계학습, 사례기반 기계학습 기법에 대해 소개한다.

2.2.1 베이지안 분류를 이용한 쓰레기 편지 여과

베이지안 분류를 이용한 나이브 베이지안 분류(naïve Bayesian classifier)는 단어의 발생 분포가 쓰레기 편지 집합과 정보성 편지 집합에서 서로 다르다는 점에 착안한 분류 방법이다(이상호, 2005). 쓰레기 편지 여과작업에서 활용하는 나이브 베이지안 분류는 다음과 같다. 입력 편지는 n개의 단어 $\{t_1, t_2, ..., t_n\}$로 이루어져 있고, 나이브 베이지안 분류에서 구하고자 하는 목적 클래스 $c^* \in \{SPAM, HAM\}$는 다음과 같다.

$$\begin{aligned} c^* &= \arg \max_{c \in \{SPAM, HAM\}} P(c \mid t_{1..n}) \\ &= \arg \max_{c \in \{SPAM, HAM\}} P(c) P(t_{1..n} \mid c) / P(t_{1..n}) \\ &= \arg \max_{c \in \{SPAM, HAM\}} P(c) P(t_{1..n} \mid c) \\ &= \arg \max_{c \in \{SPAM, HAM\}} P(c) \prod_{i=1}^{n} P(t_i \mid c) \end{aligned} \qquad (2.1)$$

수식 (2.1)의 마지막 식은 임의의 클래스로부터 단어열이 발생할 확률 $P(t_{1..n}|c)$를 $\prod_{i=1}^{n} P(t_i|C)$로 근사화하였는데 이것은 편지 내의 각 단어들이 위치에 관계없이 상호 독립적이라고 가정하여 계산의 복잡도를 줄이고, 구현을 편하게 하기 위해서이다.

나이브 베이지안 분류를 이용한 쓰레기 편지 여과 시스템의 특성을 파악한 쓰레기 편지 발송자들은 최근 'word salad' 라는 방법을 이용하여 쓰레기 편

지를 정보성 편지처럼 보이게 하였다(이상호, 2005). 이 방법은 쓰레기 편지 내에 정상적인 단어들을 추가하는 방법인데, 정보성 편지 내에서 자주 발생하는 단어들을 모은 뒤, 그 단어들을 원래 편지에서 사용하는 단어들의 개수보다 훨씬 많이 넣게 되면 여과 시스템은 그 편지를 정보성 편지로 간주하게 되는 것이다. 쓰레기 편지의 내용을 조작하여 쓰레기 편지 여과기를 통과하는 방법은 'word salad' 뿐만 아니라 'daily news', 'slice and dice', 'lost in space' 등 여러 가지가 있다. 결론적으로 분류 정확도의 향상을 꾀하기 위해서는 편지 내의 단어 이외에 다른 요소들[6]의 활용이 필요하다.

2.2.2 지지벡터를 이용한 쓰레기 편지 여과

지지벡터기계(SVM: Support Vector Machine)는 Vapnik에 의해서 1995년 이진 분류 문제를 해결하기 위해서 제안된 지도학습 방법의 일종이다(민도식 외, 2003). 지지벡터기계는 선형적으로 분리할 수 있는 학습 집단에 대해서 데이터를 +1 와 -1 클래스의 두 개의 집합으로 완전하게 분리시킬 수 있는 결정면을 이용하여 분류 작업을 수행한다(Vapnik & Cortes, 1995). 지지벡터기계의 알고리즘은 수식 (2.2)와 같다.

$$y_i = \begin{cases} +1 \text{ if } \overrightarrow{w} \cdot \overrightarrow{x_i} + b > 0, \\ -1 \text{ if } \overrightarrow{w} \cdot \overrightarrow{x_i} + b \leq 0, \end{cases} \qquad (2.2)$$

$\overline{w}$는 가중치벡터, $\overline{x_i}$는 입력벡터, b는 기준치이며, $\overline{w}$와 b는 학습 데이터로부

[6] 대상 편지가 포함하는 이미지 처리, 대상편지에 대한 사용자 행동 등

터 학습된다. 편지 데이터인 학습 문서 집합을 $D=\{(\overrightarrow{x_i},y_i)\}$라고 하면, 입력 데이터 $\overrightarrow{x_i}$가 범주에 속하는 경우 y_i가 +1[7]의 값을 가지고, 속하지 않으면 -1[8]의 값을 가진다.

지지벡터기반을 이용한 편지 분류 작업은 적은 학습 데이터를 이용하여 비교적 정확한 분류 결과를 제공한다. 하지만 학습 속도가 늦고, 분류 시스템의 부하가 크다는 단점이 있다(민도식 외, 2003).

2.2.3 사례기반 학습을 이용한 쓰레기 편지 여과

사례기반 학습(instance-based learning)은 기존의 사례를 근거로 하여 새로운 사례를 판단하여 범주화하거나 분류하는 학습방법이다. 지속적으로 변화하는 사용자의 요구를 반영해야 하는 쓰레기 편지 여과작업의 특성상 이 논문이 제안하는 쓰레기 편지 여과 시스템은 기존의 학습 모델의 점진적 갱신이 가능한 사례기반 학습을 사용한다. 따라서 사례기반 학습에 대해 자세히 설명한다. 그림 2.2는 학습부와 실행부로 구성되는 일반적인 사례기반 학습의 구성을 나타내고 있다. 학습부에서는 유사한 사례를 군집화하거나, 색인화하여 저장하고, 실행부에서는 주어진 입력 데이터와 저장된 사례들 간의 유사도를 계산하여 분류 작업을 수행한다(Daelemans *et al.*, 2004).

---

7) SPAM: 쓰레기 편지
8) HAM: 정보성 편지

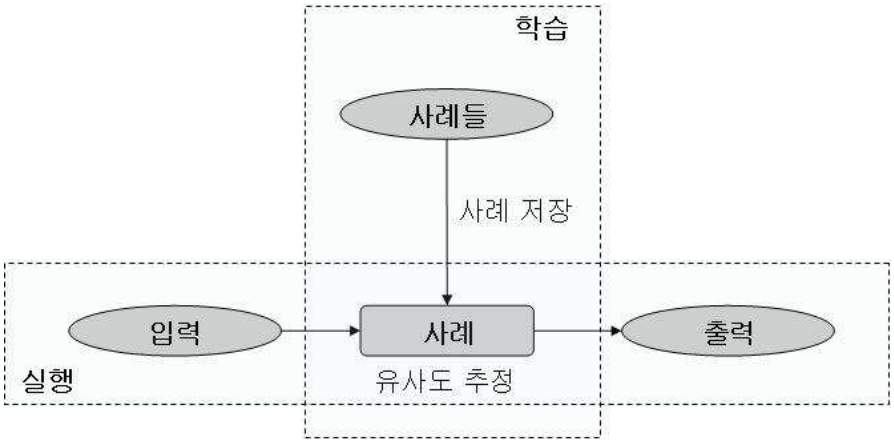

그림 2.2 사례기반 시스템의 흐름도

Fig. 2.2 A flow diagram of instance-based classification systems

사례기반 학습에서 입력 $X=(x_1, x_2, \ldots, x_n)$와 사례 $Y=(y_1, y_2, \ldots, y_n)$ 은 특별한 의미를 지닌 자질로 구성된 패턴이며, 두 패턴의 유사도 $\triangle(X, Y)$ 는 수식 (2.3), (2.4)와 같이 정의된다.

$$\Delta(X,Y)=\sum_{i=0}^{n} w_i \delta(x_i, y_i) \tag{2.3}$$

$$\delta(x_i, y_i)=\begin{cases} \dfrac{x_i - y_i}{\max_i - \min_i} & \text{if } i\text{번째 자질 = 숫자} \\ 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases} \tag{2.4}$$

여기서 $\max_i$와 $\min_i$는 각각 번째 자질이 가질 수 있는 최대값과 최소값을 의미하고, w_i는 i번째 자질의 가중치이다. 제안하는 시스템은 가중치를 계산하기 위한 방법으로 수식 (2.5)와 같은 이득률(gain ratio)를 사용하지만, 가

중치w_i를 구하기 위한 계산 방법에는 여러 가지가 있다.

$$w_i = \frac{H(C) - \sum_{v \in V_i} P(v) \times H(C|v)}{si(i)} \tag{2.5}$$

$$H(C) = - \sum_{c \in C} P(v) \times H(C|v) \tag{2.6}$$

$$si(i) = - \sum_{v \in V_i} P(v) log_2 P(v) \tag{2.7}$$

수식 (2.5)에서 C는 클래스 레이블, V_i는 i번째 자질의 값, $H(C)$는 클래스 레이블의 엔트로피이고 수식(2.6)은 $H(C)$를 정의한다. 수식 (2.7)은 자질 값의 엔트로피 $si(i)$를 정의한다.

일반적으로 사례기반 학습을 이용한 분류 작업은 학습 데이터 양이 적을 때 비교적 분류 정확도가 낮고, 처음 학습할 때 시간이 오래 걸린다는 단점이 있지만 일단 사례학습이 끝나면 다음부터 주어지는 새로운 사례에 대해서는 매우 우수한 학습 효율을 보여준다(김재훈 & 이공주, 2003). 편지 여과 작업에서 마찬가지로 처음 편지 데이터를 학습할 때는 학습 시간이 오래 걸리고 분류 정확도가 낮지만 사례기반 학습의 특성상 분류 모델이 만들어진 후에는 재학습 시간이 빠르고 학습 데이터 양에 따라 분류 정확도의 향상을 기대할 수 있다.

2.3 묵시적 피드백

묵시적 피드백(implicit feedback)은 사용자에 대한 지식 습득과정에서 사용자의 어떠한 직접적인 참여도 요구하지 않는 대신 입력되는 데이터에 대한 사용자의 행위(action)를 기록, 저장하여 사용자와 데이터 간의 연관성(쓰레기 편지의 여부)을 정의하는 작업을 말한다(Hanani *et al.*, 2001).

Morita와 Shinoda(1994)는 그들의 연구에서 사용자가 문서를 읽는데 소비하는 시간과 정보 요구 사이의 상관관계를 발견하였으며, Konstan(1997)이 제안한 협력 여과 시스템(collaborative filtering system)인 'GroupLens'는 문서를 읽는데 투자하는 시간을 관찰하여 이것을 사용자의 문서에 대한 관련 정도로 판단하였다. 그리고 Goecks과 Shavlik(2000)은 대상 웹 페이지에 대한 사용자의 하이퍼링크(hyperlink) 클릭 동작과 마우스 스크롤 행위를 활용하였다.

사용자의 문서 데이터에 대한 저장, 삭제, 출력 등의 행동 역시 관심사를 표현하는 증거로 활용할 수 있는데, 특히 Kim과 Oard(2004)는 최근 사용자의 행동을 검토(examination), 저장(retainment), 참조(reference), 주석 첨가(annotation)로 분류하고 이를 묵시적 피드백의 증거 데이터로 활용하는 방법을 제안하고 표 2.1과 같이 분류하였다.

표 2.1 Kim 의 사용자 행동의 분류법

Table 2.1 Classification of Kim's user action

최소 범위 / 행위의 종류	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

세로축은 사용자의 행동의 종류를 나타내고, 가로축은 분류 대상의 최소 범위를 나타낸다. 예를 들어 사용자의 복사/붙이기 행위는 문서의 일부분을 최소 범위로 하는 참조 행위로 해석할 수 있다. 그러나 현재까지 Kim과 Oard의 분류법을 활용한 시스템이 구현된 사례가 없다. 이 논문은 시스템은 위 분류법를 참고하여 사용자 행동을 모델링하고, 이를 이용한 쓰레기 편지 여과 시스템을 구현하였다.

2.4 편지 학습 말뭉치

객관적인 편지 여과 시스템의 성능을 평가하기 위해서는 공개된 편지 말뭉치가 필요하다. 영어로 구성된 편지 말뭉치 구축 작업은 표 2.2에서 확인할 수

있듯이 매우 활발하게 이루어지고 있다. 그리고 대표적인 편지 말뭉치로 Enron e-mail corpus(Klimt & Yang, 2004)가 있는데, 영문으로 구성된 뉴스그룹 데이터와 정보성 편지로 구성되어 있다. 이 편지 말뭉치에서 주목할 만한 점은 다른 편지 말뭉치와 달리 정보성 편지의 집합으로 구성되어 있다는 것이다.

표 2.2 편지 말뭉치 현황

Table 2.2 A list of mail corpus

말뭉치 이름	말뭉치 구축 사이트	말뭉치 크기	특징
SpamArchive	www.spamarchive.org/	222,506개(1.3 GB)	쓰레기 편지(수집 완료)
SpamAssassin public mail corpus	spamassassin.apache.org/publiccorpus	약 100,000개	쓰레기 편지 + 정보성 편지 (수집 완료)
Corpus of Junk Emails	clg.wlv.ac.uk/projects/junk-email/	약 800MB	쓰레기 편지(수집 완료)
Spam Archives (Corpora)	www.iit.demokritos.gr/skel/i-config/	N.A.	쓰레기 편지(수집 완료)
Toasted Spam File	www.toastedspam.com/stupid/	현재 123,522개	쓰레기 편지(수집 진행중)
Spam Hall of Shame	www.sput.nl/spam/spam-hall.html	현재 약 500,000개	쓰레기 편지(수집 진행중)
Dolphinwave archive of spam	www.dolphinwave.org/spam/	현재 93,924개	쓰레기 편지 분석 정보 (수집 진행중)
Spam Honeypot Archive	schnarff.com/honeypot.html	N.A.	쓰레기 편지(수집 진행중)
Spamming the Domain Owner	k.geocities.com/sjwest01/	현재 7533개	쓰레기 편지(수집 진행중)
African Emails	www.climate.unibe.ch/Ebeyerle/emails	현재 약 25,000개	헤더(header) 정보 N.A.
Stephen Newton's Spam Museum	spammuseum.blogspot.com/	N.A.	헤더(header) 정보 N.A. 쓰레기 편지
The Enron Corpus	www.cs.cmu.edu/~enron/	약 600,000개(1.8G)	정보성 편지 뉴스 그룹 데이터 포함

현재 영어 편지 말뭉치 구축작업이 활발하게 이루어지고 있는 반면, 한국어 편지를 대상으로 하는 공개된 편지 말뭉치는 없다. 따라서 쓰레기 편지 여과시스템을 개발하는 연구자들은 자신의 시스템 실험을 위해 각자 평가 데이터를 마련해야 하는 실정이다. 쓰레기 여과 시스템의 평가를 위한 공개 한국어 편지 말뭉치 구축 작업이 필요한 상황이다.

제 3 장 사용자 행동과 점진적 기계학습을 이용한 쓰레기 편지 여과 시스템의 설계 및 구현

이 논문에서 제안하는 점진적 기계학습을 이용한 쓰레기 편지 여과 시스템은 편지 말뭉치와 사용자 행동 정보를 이용한 학습 말뭉치 구축 과정, 사례기반 기계학습을 이용한 행동 추론 모델, 분류 모델을 만드는 학습 과정, 행동 추론 모델과 분류 모델을 이용한 편지 분류 과정으로 나눌 수 있다(그림 3.1 참고).

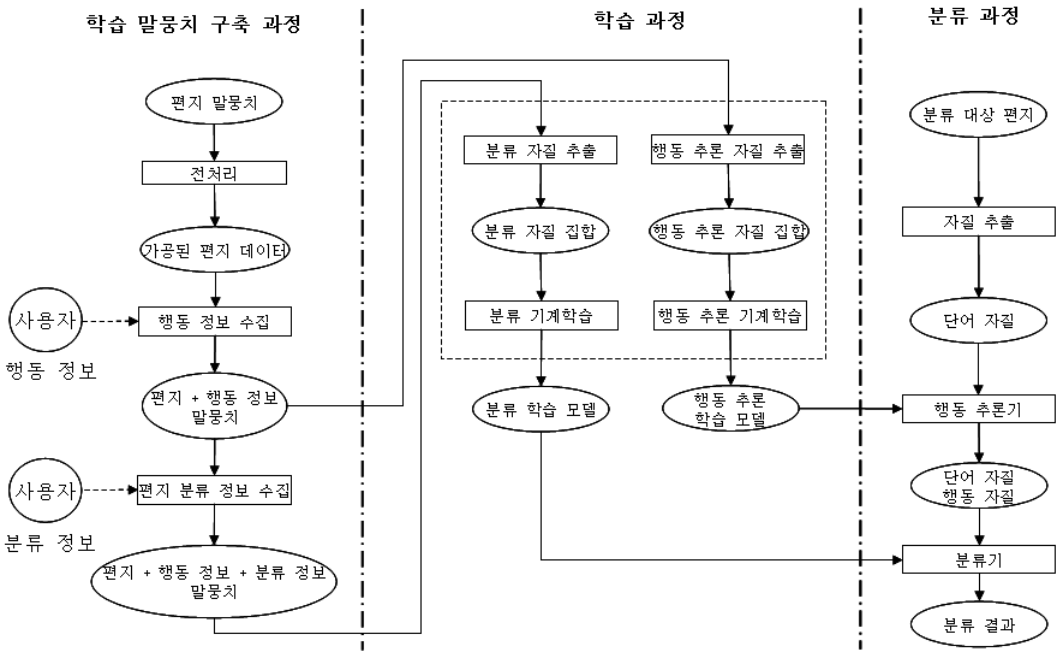

그림 3.1 사용자 행동과 점진적 기계학습을 이용한
쓰레기 편지 여과 시스템의 전체 구조

Fig. 3.1 Overview of the proposed SPAM mail filtering system

3.1 학습 말뭉치 구축 과정

제안하는 시스템은 쓰레기 편지 여과작업에 사용될 행동 추론 모델과 편지 분류 모델을 구축하기 위해 편지 말뭉치와 편지에 대한 사용자의 행동을 이용하여 학습 말뭉치를 구축한다. 학습 말뭉치 구축 과정은 편지 말뭉치를 전처리(preprocessing)하는 과정, 전처리 된 데이터를 이용한 사전 구축작업, 사용자 인터페이스를 통한 사용자 정보 수집 과정으로 구성된다.

3.1.1 전처리 과정

전처리 과정은 HTML문서로 구성되어 있는 전자 편지를 쓰레기 편지 여과작업에 활용할 수 있는 정보들의 집합으로 가공하는 과정이다. 전처리 과정은 다음과 같다. 먼저 편지를 분석하여 헤더(header) 정보를 추출하고 HTML 태그를 제거한 후 명사 데이터를 추출하는데, 이 논문에서는 명사 데이터를 추출하기 위해 명사 추출 시스템(김재훈 & 김준홍, 2001)의 을 활용한다. 전체적인 전처리 과정은 그림 3.2와 같으며, 전처리 과정의 입출력 데이터 예제는 그림 3.3에서 확인할 수 있다.

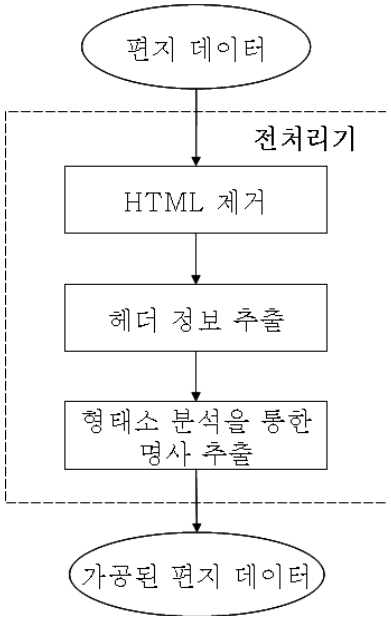

그림 3.2 전처리 과정

Fig. 3.2 Preprocessing steps

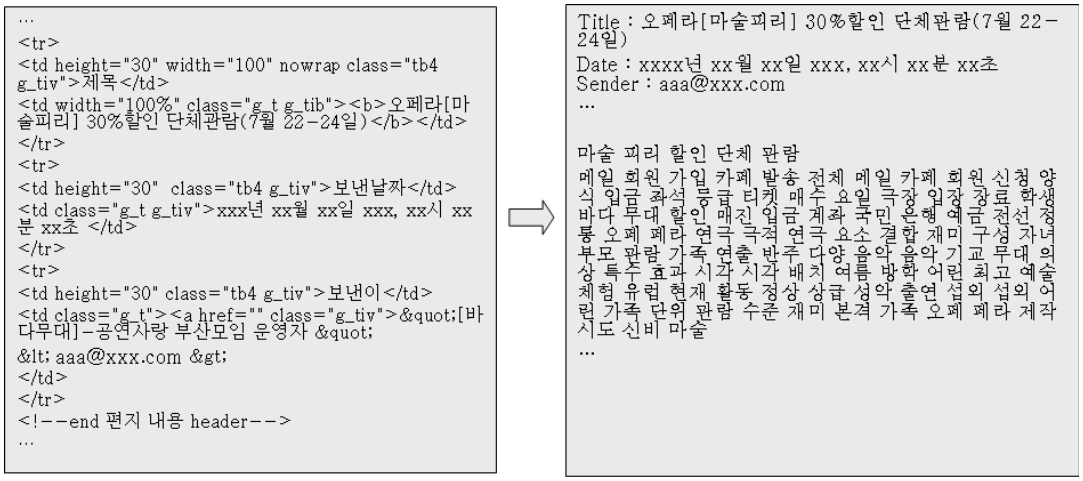

그림 3.3 전처리 전후 실제 데이터 변환 예

Fig. 3.3 An example of preprocessing data

3.1.2 사전 생성

사전은 전처리 과정을 통해 가공된 편지 데이터를 저장하여 이후 사례기반 기계학습의 자질 구축에 사용된다. 사전은 데이터 양을 줄이기 위해 편지로부터 추출한 명사정보를 바이그램(Bi-gram) 형태로 표현하는데, 이는 한자 문화권의 영향을 받은 한국어 명사가 보통 두 개의 음절로 구성되는 경우가 많다는 사실에 기인하였다(강승식, 2003). 그리고 사전에서 고빈도어(high frequency term)와 저빈도어(low frequency term)는 제거되는데, 고(저)빈도어는 여과작업에 의미 없는 데이터일 경우가 대부분이기 때문이다. 그림 3.4는 바이그램 표현의 예를 나타내고 있으며, 그림 3.5[9]는 실제 사전 데이터를 표현하고 있다.

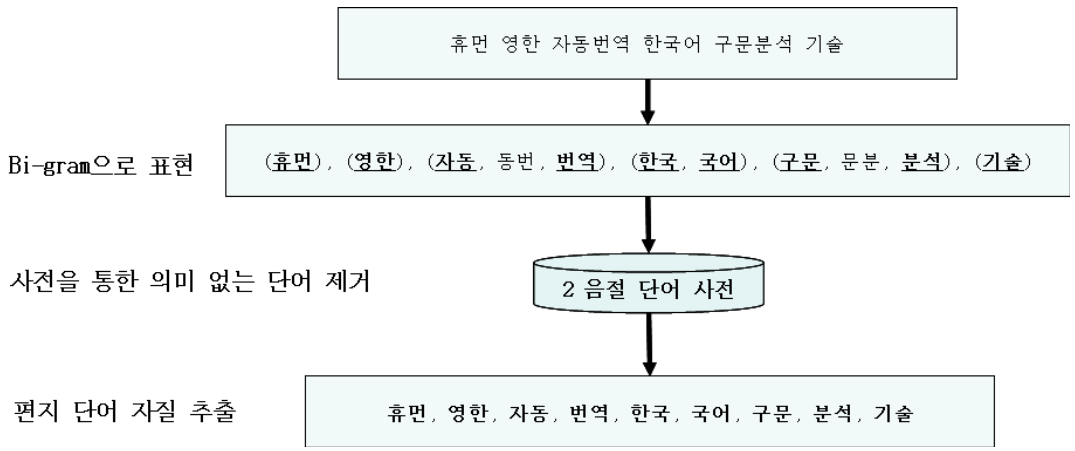

그림 3.4 명사 정보의 바이그램 표현의 개념

Fig. 3.4 Bi-gram representation of nouns

[9] 사전에서 단어 앞에 'T_' 가 붙은 것은 제목에서 추출한 단어임을 의미한다.

Index	term	freq.
...		
65	AGE	4
66	AIM	1
67	AJAX	1
68	AMD	9
69	AMD's	5
70	ANGEL	1
71	AOL's	2
72	AOPEN	1
73	AON	1
74	APEC	2
75	ASCAP	2
76	ATM	1
77	About	6
78	Acer	1
80	Acquire	1
81	Act	3
82	Adware	1
...		

Index	term	freq.
...		
1101	T_special	1
1102	T_square	1
1103	T_start	1
1104	T_suit	2
1105	T_summer	1
1106	T_technology	1
1107	T_vs	1
1108	T_wins	1
1109	T_workshop	1
...		

a) 한글 사전 데이터

Index	term	freq.
...		
3623	가격	3
3625	가결	1
3626	가공	1
3628	가능	13
3633	가방	4
3635	가사	2
3636	가상	6
3638	가수	16
3640	가슴	4
3647	가을	5
3649	가입	68
3651	가장	36
3653	가정	2
3654	가족	35
3656	가지	21
...		

Index	term	freq.
...		
1551	T_행운	3
1556	T_혁명	1
1557	T_현금	1
1558	T_혼자	1
1560	T_확실	1
1561	T_확정	1
1566	T_회복	3
1567	T_후보	6
...		
7659	할인	12
7660	함께	1
7662	합격	2
7663	합기	2
7664	합류	4
7666	합세	4
7669	합작	2
7670	합창	5
...		

b) 영어 사전 데이터

그림 3.5 사전 데이터

Fig. 3.5 Dictionary data

3.1.3 사용자 인터페이스를 통한 행동 정보 수집

제안하는 시스템은 사용자 인터페이스를 통해 사용자로부터 편지에 대한 행동 패턴 정보를 수집한다. 사용자 인터페이스는 그림 3.6에서 확인할 수 있듯이 각 편지에 대해 사용자의 행동 정보를 입력 받는 부분, 사용자가 자신이 입력한 행동을 확인할 수 있는 부분, 분류 결과를 표시하는 부분으로 구성되어 있다.

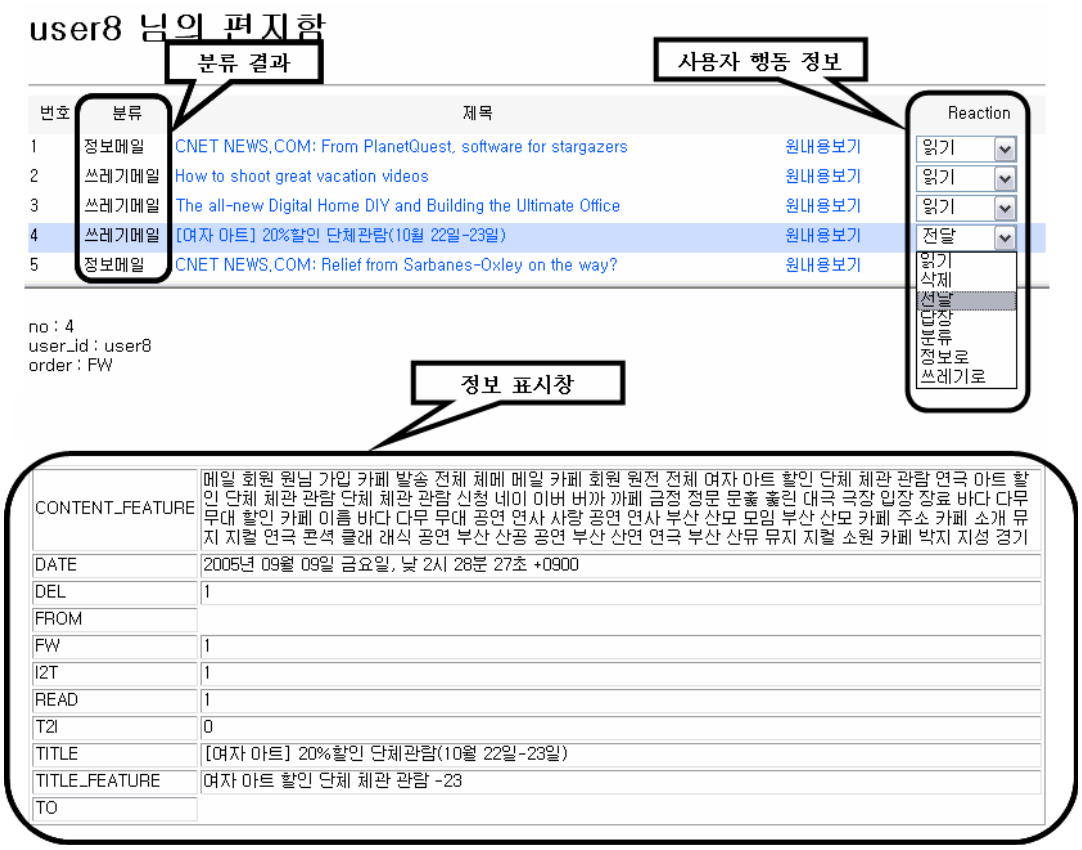

그림 3.6 사용자 인터페이스의 구성

Fig. 3.6 Organization of User Interface

제안하는 시스템이 사용자 인터페이스를 통해 수집하는 사용자 행동 패턴은 읽기(reading), 삭제하기(delete), 분류하기(classification), 전달하기(forward), 답장 보내기(reply)이다. 그리고 제안하는 시스템은 사용자가 제공할 수 있는 분류 정보를 정보 편지로 분류(I2T: Information to Trash), 쓰레기 편지로 분류(T2I: Trash to Information)로 정의하였다.

3.2 학습 과정

학습 과정은 학습 말뭉치를 이용하여 행동 추론 학습기와 분류 학습기의 입력 데이터를 추출하고, 이를 이용하여 행동 추론 모델과 편지 분류 모델을 구축하는 과정이다. 제안하는 시스템은 각 모델을 구축하기 위해 점진적 사례기반 기계학습 도구인 TiMBL[10]을 사용한다.

3.2.1 행동 추론 모델과 편지 분류 모델의 구축을 위한 학습 자질 추출

제안하는 쓰레기 편지 여과 시스템은 두 종류의 학습 모델이 사용된다. 하나는 기본 편지로부터 사용자의 행동을 추론하기 위한 행동 추론 모델이고, 하나는 추론된 사용자의 행동을 이용하여 편지를 분류하기 위한 편지 분류 모델이다. 행동 추론 모델의 학습 자질은 편지로부터 추출한 정보들로 구성된 편지부와 사용자 행동 정보에서 추출한 행동부로 이루어져 있다. 편지부는 해당 편지에서 추출한 단어의 사전 색인(index)번호와 출현 빈도수를 쌍으로 하여 구성하고, 각 편지에 포함된 이미지의 개수 정보[11]를 저장한다. 행동부는 사용자 인터페이스로부터 얻은 사용자 행동 정보를 이진수 형태로 저장하고 있다. 편지 분류 모델의 학습 자질은 행동 추론 모델의 학습 자질에 해당 편지에 대한 사용자의 분류 정보[12]를 덧붙여 구성한다. 제안하는 시스템은 데이터 크기를 줄이기 위해 두 개의 학습 자질을 희소 벡터(sparse vector) 형태로 표현한다.

---

10) TiMBL(Tilburg Memory Based Learner Version 5.1)
11) 일반적으로 쓰레기 메일은 오직 이미지로 구성되어 있는 경우가 많다.
12) 제안하는 시스템에서 대상 편지에 대해 사용자가 직접 입력한 쓰레기 편지와 정보성 편지에 대한 분류 결과는 행동부의 I2T 또는 T2I 항목이다.

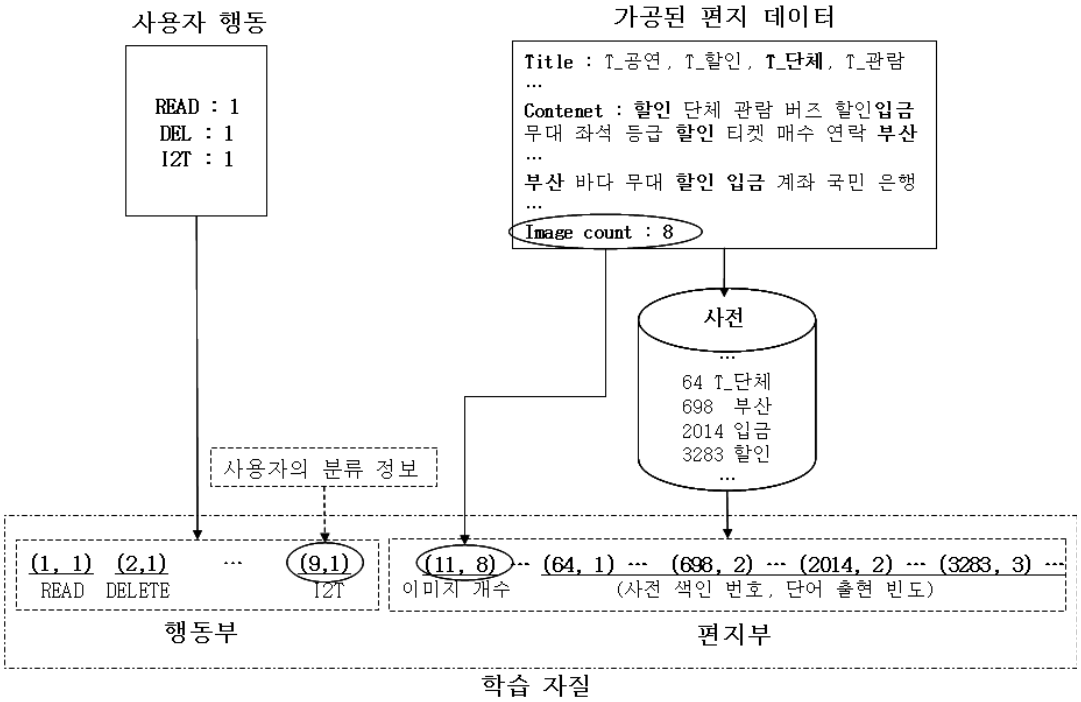

그림 3.7 자질 추출 예

Fig. 3.7 An example of the feature extraction

3.2.2 학습을 이용한 모델 생성

제안하는 쓰레기 편지 여과 시스템은 사용자의 요구 변화와 시대의 변화를 반영하기 위해 편지 분류에 사용하는 두 개의 모델을 점진적으로 갱신하는데, 이를 위해 점진적 사례기반 기계학습 도구인 TiMBL에서 제공하는 IB2 알고리즘[13]을 이용한다. 행동 추론 학습 모델과 편지 분류 학습 모델을 따로 생성하기 위해 두 번의 학습 과정이 필요하지만 학습 방법은 같다. 이 쓰레기 편지

[13] IB2: Incremental editing - 점진적 사례기반 학습 알고리즘

여과 시스템이 TiMBL을 이용하여 두 개의 모델을 생성하는 방법은 다음과 같다.

모델 생성을 위한 TiMBL 명령어

Timbl -f xxx.training -a 3 -F Sparse -N '자질 개수' -I xxx.instance -W xxx.weight

사용 옵션

-f : 학습 기능(xxx.training: 학습 자질 파일)
-a : 분류 알고리즘(3: IB2)
-F : 자질 데이터 형식(Sparse: 희소벡터)
-N : 자질 개수
-I : 사례 생성(사례 정보 파일: xxx.instance)
-W : 가중치 생성(가중치 정보 파일: xxx.weight)

3.3 분류 과정

제안하는 시스템은 대상 편지로부터 그림 3.7의 편지부를 추출하고 행동 추론 모델을 통해 사용자의 행동을 추론한 후에 편지 분류 모델을 이용하여 대상 편지가 쓰레기 편지인지 정보성 편지인지를 결정한다. 이 시스템이 쓰레기 편지를 분류 하는 과정은 그림 3.8과 같다.

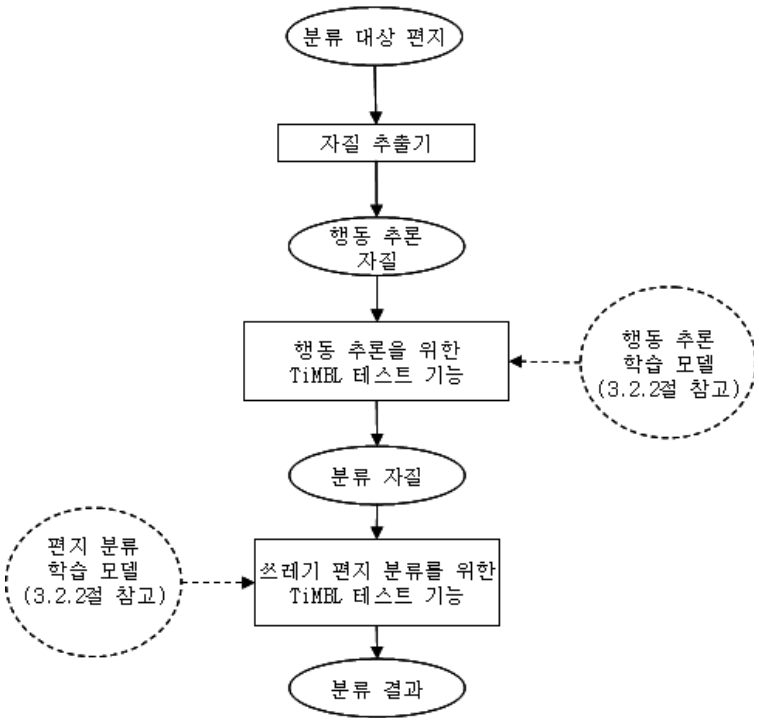

그림 3.8 제안하는 쓰레기 편지 여과 시스템의 편지 분류 과정

Fig. 3.8 A Process of mail classification in the proposed SPAM filtering system

편지 분류 작업은 TiMBL의 테스트 기능을 이용하여 수행하였는데, 편지 분류 작업을 위한 TiMBL의 사용법은 다음과 같다.

편지 분류를 위한 TiMBL 명령어

Timbl -t xxx.test -a 3 -F Sparse -N '자질 개수' -i xxx.instance -w xxx.weight -o '결과 저장 파일명'

사용 옵션[14)]

-t : 테스트 기능(xxx.training: 분류 대상 파일)
-a : 분류 알고리즘(3: IB2)
-F : 자질 데이터 형식(Sparse: 희소벡터)
-N : 자질 개수
-i : 사례 입력(사례 정보 파일: xxx.instance)
-w : 가중치 입력(가중치 정보 파일: xxx.weight)
-o : 분류 결과 출력

14) 학습 명령에서 -I, -W 옵션을 통해 출력된 xxx.instance, xxx.weight와 동일

제 4 장 실험 및 평가

이 실험은 학습 데이터 양에 따른 편지 분류 결과의 정확도를 측정하고, 최적의 분류 결과를 나타내는 학습 데이터의 수량과 쓰레기 편지 분류의 증거로 사용되는 사용자의 행동 패턴의 종류, 그리고 편지 여과작업에서 행동 정보의 유용성 정도에 초점을 두고 진행하였다.

4.1 실험 말뭉치

학습 및 실험에 말뭉치로 사용한 편지는 한메일[15]에서 제공하는 편지 백업 기능을 이용하여 12명의 사용자[16]로부터 추출한 10,000개의 편지 데이터와 해당 편지에 대한 각 사용자의 행동 정보이다. 편지 데이터의 수집 기간은 2005년 3월부터 6월까지 3개월이었으며, 각 사용자는 한메일에서 제공하는 쓰레기 편지 여과 기능을 사용하지 않고 편지 데이터를 수집하였다.

사례기반 기계학습 시스템의 학습 말뭉치 구축을 위해 각 사용자는 1,000개의 편지 말뭉치를 대상으로 행동 정보를 제공한다[17]. 수집된 각 사용자별 1,000개의 데이터는 900개의 학습 데이터와 100개의 실험 데이터로 구분한 후 테스트를 수행하였다.

행동 정보는 제안하는 시스템이 제공하는 사용자 인터페이스를 통하여 수집한 것이며, 수집된 행동 데이터의 통계는 그림 4.1과 같다.

---

15) 한메일 인터넷 주소: www.daum.net

16) 한국해양대학교 IT공학부 학부생 및 대학원생(1학년 3명, 2학년 3명, 3학년 2명, 4학년 2명, 대학원생 1명), 기타 1명 - 남 9명, 여 3명

17) 시스템의 성능 비교 작업을 위해 사용자 1번-11번, 2번-12번은 같은 편지를 대상으로 실험 수행

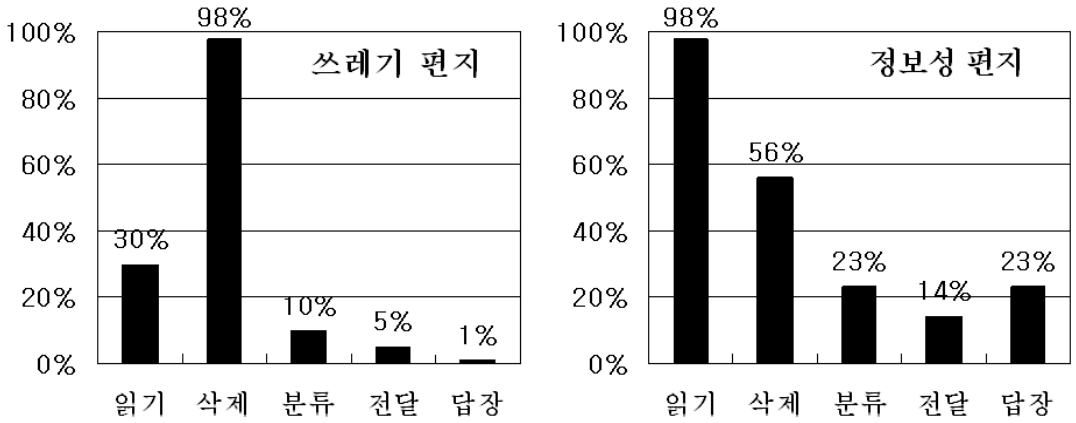

그림 4.1 학습 말뭉치의 쓰레기/정보성 편지에 대한 사용자 행동

Fig. 4.1 The ratio of action of SPAM/HAM in learning corpus

4.2 성능 평가 방법

이 논문에서는 제안한 여과 시스템의 정확도를 평가하기 위하여 다음과 같이 정확도 A를 계산하였다.

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

수식 4.1에서 N은 실험 데이터의 개수, E는 사용자의 분류결과와 일치하는 시스템의 분류결과의 개수이다.

4.3 분류 정확도 평가와 분석

4.3.1 학습 데이터 양에 따른 분류 정확도

제안하는 시스템은 사례기반 기계학습을 이용하여 추출한 분류 모델을 활용하여 여과작업을 수행하는데, 우선 학습 데이터 양에 따라 분류 정확도의 변화 추이를 조사하였다. 실험은 각 사용자의 학습 데이터를 100개에서 900개까지 200개씩 점진적으로 증가시켜서 분류 정확도를 측정하는 방식으로 진행하였고, 기본적으로 학습 데이터의 양이 늘어남에 따라 분류 정확도가 향상되었음을 표 4.1과 그림 4.2에서 확인할 수 있다.

표 4.1 학습 데이터 양에 따른 분류 정확도 변화

Table 4.1 A mail classification precision according to learning data

(단위 %)

사용자	학습 데이터 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

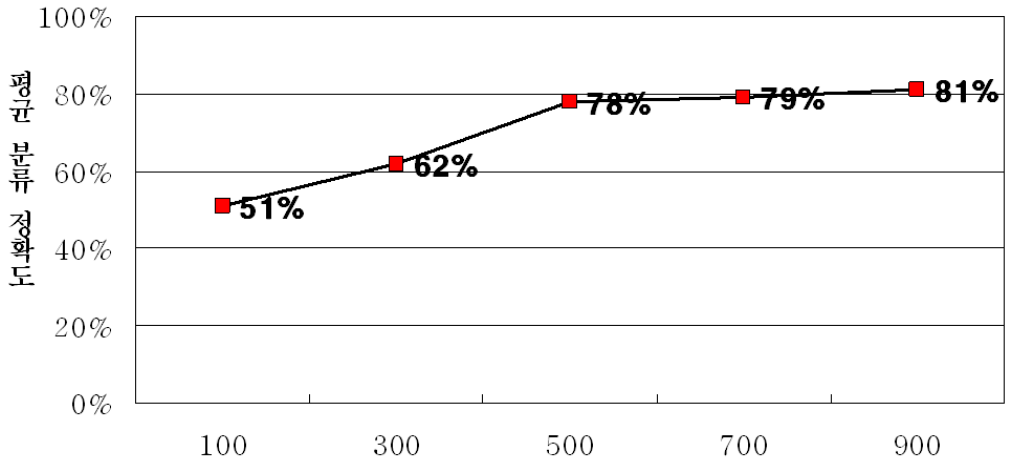

그림 4.2 학습 데이터 양에 따른 분류 정확도 변화

Fig. 4.2 A mail classification precision according to learning data

4.3.2 사용자별 분류 결과 차이 분석

표 1에서 살펴보면 **사용자3**이 가장 높은 분류 정확도를, **사용자9**가 가장 낮은 분류 정확도를 나타내었다. 두 사용자의 분류 정확도 차이의 원인은 행동 정보 입력 패턴에서 그 원인을 찾을 수 있다. 높은 정확도를 보인 **사용자3**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 **사용자9**의 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 있었다는 점은 시스템이 편지 분류작업을 수행하기 위한 조건이 비교적 나빴다는 것을 의미한다(표 4.2 참고).

표 4.2 분류 정확도에 따른 사용자 행동비율

Table 4.2 User action ratio according to classification precision

(단위 %)

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

4.3.3 최적의 분류 결과를 나타내는 학습 데이터 양

실험 결과 중 주목할 만한 점은 일부 사용자의 경우 학습 데이터 양이 늘었음에도 불구하고 분류 정확도가 떨어지는 현상이 발생하였는데, 이것은 학습 시스템의 과적합(over-fitting)현상으로 해석할 수 있다(Mitchell, 1997). 따라서 10,000개의 학습 데이터 중 학습데이터 1,500건을 추출하여 최적의 편지 분류 성능을 위한 학습 데이터 양을 조사하였다. 그림 4.3은 조사 결과를 나타내고 있는데 학습 데이터 양이 900개에 1,000개 사이에 최적의 분류 성능을 나타내었고, 이후 학습 시스템의 과적합 현상에 따라 분류 정확도가 지속적으로 낮아지는 현상을 발견하였다.

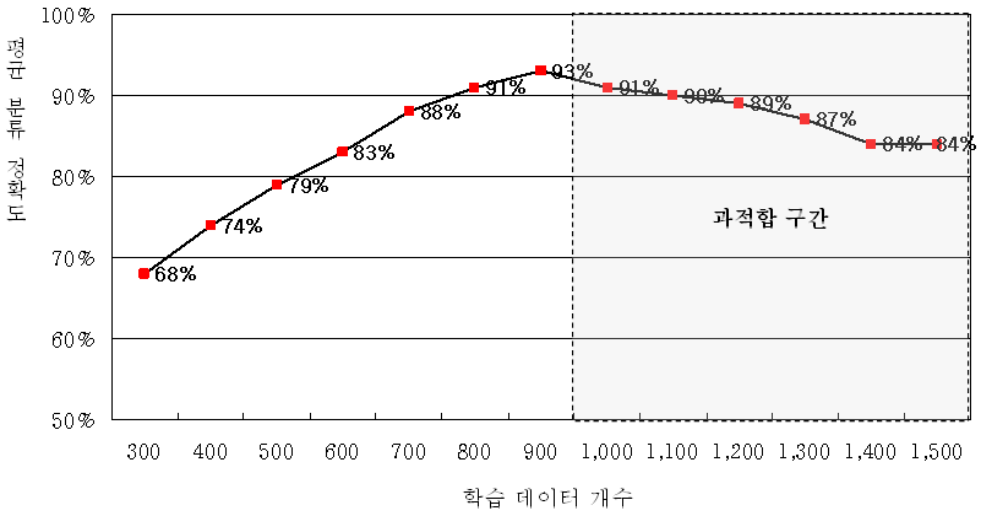

그림 4.3 학습 데이터 양에 따른 과적합 현상

Fig. 4.3 A problem of over-fitting according to amount of learning data

4.4 분류의 증거로 사용되는 행동 패턴 분석

시스템에서 정의한 사용자 행동 정보 모델은 읽기, 삭제, 분류, 전달, 답장으로 구성되어 있다. 쓰레기 편지와 정보성 편지 각각에 대해 사용자의 행동 정보를 조사하였다(그림 4.1 참고). 쓰레기 편지와 정보성 편지에서 각각 추출한 사용자 행동 비율이다[18]. 사용자는 쓰레기 편지/정보성 편지에 대해 삭제 행동과 읽기 행동을 가장 높은 비율로 하였고, 쓰레기 편지에 대해서 삭제의 행동과 이외의 행동과 그 비율의 차가 큰 반면, 정보성 편지에 대해서는 가장 높은 행위인 읽기와 더불어 비교적 다양한 행동 패턴을 보이는 것으로 조사되었다.

[18] 편지에 대한 사용자의 행동 정보는 중복 가능

4.5 편지 여과작업에서 행동 정보의 유용성 여부 평가

이 논문에서는 편지를 분류하기 위해 일반적으로 사용하는 편지에서 추출한 단어뿐만이 아니라 대상 편지에 대한 사용자의 행동 정보를 이용하는 기법을 제안하였다. 따라서 사용자 행동 정보를 제외한 여과 성능과 사용자 행동 정보를 포함한 여과 성능을 비교할 필요가 있는데, 결과는 그림 4.4에서 확인할 수 있듯이 학습 데이터 양에 따라 최소 6%에서 최대 16%의 분류 정확도가 향상되었음을 확인할 수 있다. 결론적으로 이 연구는 편지에 대한 사용자의 행동 정보가 쓰레기 편지 분류 작업에 효과적인 요소가 되었다는 사실을 보였다.

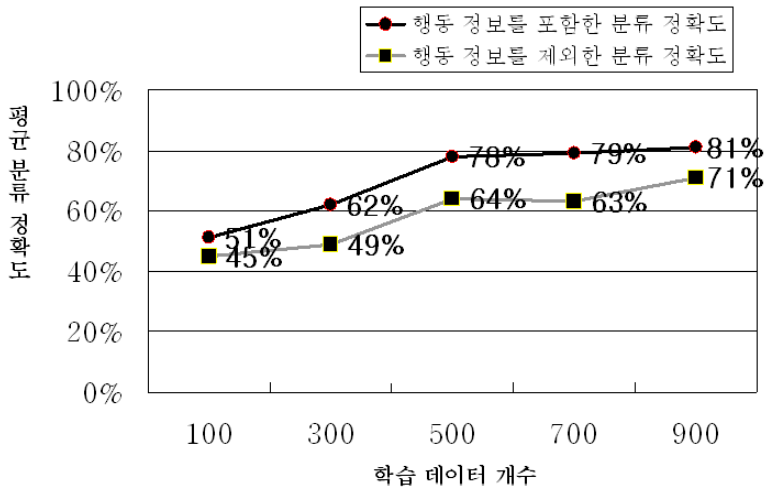

그림 4.4 사용자 행동 정보의 유무에 따른 분류 정확도

Fig. 4.4 Usefulness of user action in mail classification

4.6 기존 쓰레기 편지 여과 시스템과의 비교

이 절에서는 제안한 쓰레기 편지 여과 시스템과 기존의 쓰레기 편지 여과 시스템들과의 비교를 하고자 한다. 표 4.3은 기존 쓰레기 편지 여과 시스템들과 제안한 시스템의 특징을 나타낸 것이다.

표 4.3 쓰레기 편지 여과 시스템의 비교[19)]

Table 4.3 Characteristic comparison of existing SPAM filtering systems

시스템	특징		편지 말뭉치			분류 정확도(%)	사용자 행동 정보 사용
	분류 자질	분류 방법	언어	학습 데이터	실험 데이터		
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

표 4.3에서 시스템 I(임정택 외, 2005)과 II(김현준 & 정재은, 2003)는 나이브 베이지안 분류 방법, 시스템 III(민도식 외, 2003)은 SVM을 이용하여 편지를 분류하였다. 시스템 IV는 제안한 시스템이다. 시스템 II는 잘못된 여과 결과에 대해 사용자가 피드백 정보[20)]를 입력할 수 있는 인터페이스를 제공하지만 학습 자질로 사용하지는 않는다.

19) 편지 말뭉치 항목의 (xxx + yyy)에서 xxx는 쓰레기 편지의 개수를 의미하고, yyy는 정보성 편지의 개수를 의미한다.

20) 이 논문에서 사용하는 T2I, I2T 정보와 동일하다.

이 논문의 2.4절에서 설명하였듯이 공통적으로 사용할 수 있는 공개 한국어 편지 말뭉치가 없기 때문에 각 시스템들은 성능 평가를 위해 각각 편지들을 수집하였다. 따라서 객관적인 성능 비교는 어려울지 모르지만, 표 4.3에서 제시한 시스템은 대체로 비슷한 분류 정확도를 나타내었다.

제 5 장 결론 및 앞으로의 연구 과제

이 논문에서는 사용자의 행동 정보와 점진적 기계학습을 이용한 쓰레기 편지 여과 시스템을 제안하였다. 학습 단계에서는 사례기반 학습기를 이용하여 행동 추론 모델과 분류 모델을 만들고, 분류 단계에서는 대상 편지를 입력으로 행동 추론 모델과 편지 분류 모델을 거쳐 쓰레기 편지 여부를 판단하였다.

이 시스템의 성능을 평가하기 위한 학습 말뭉치로 국내 편지 서비스 제공 업체인 한메일의 개인 편지 10,000개를 사용자의 동의 하에 추출하였다. 실험은 학습 데이터 양에 따른 편지 분류 결과의 정확도, 분류의 증거로 사용되는 사용자의 행동 패턴의 종류, 편지 여과작업에서 행동 정보의 유용성 정도에 초점을 두고 진행하였다. 이 시스템은 학습 데이터가 900개일 때, 평균 약 81%의 정확도를 보였으며, 학습 데이터 양이 늘어감에 따라 과적합 문제로 인한 정확도 저하 구간이 발견되었지만 일반적으로 분류 성능이 좋아지는 사실을 발견할 수 있었다. 실험을 통해 얻은 중요한 사실은 사용자의 행동 정보를 포함한 편지 분류 작업이 사용자의 행동 정보를 제외한 편지 분류 작업에 비해 6%~14% 정도의 분류 정확도가 향상되었다는 것이다. 결과적으로 이 논문은 사용자의 행동이 편지 분류 작업에 있어서 유용한 정보가 된다는 것을 보였다.

향후에는 제안한 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 있다. 무엇보다 개발한 쓰레기 편지 여과 시스템의 객관적이고 정확한 성능 평가를 위해 공개 한국어 편지 말뭉치 구축 작업이 필요하다. 또

한 쓰레기 편지 여과작업뿐만 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

강승식 (2003), “음절 바이그램 단순화 기법에 의한 한국어 자동 띄어쓰기 시스템의 성능 개선”, *제15회 한글 및 한국어 정보처리 학술발표 논문지*, pp. 227-231.

김재훈, 김준홍 (2001), “도합유사도를 이용한 한국어 문서요약 시스템”, *한국인지과학회 논문지*, vol. 12, no. 2, pp. 35-42.

김재훈, 이공주 (2003), “사례기반 학습을 이용한 음절기반 한국어 단어 분리 및 범주 결정”, *정보처리학회 논문지*, vol.10, no. 1, pp. 47-56.

김영택 외 (2001), *자연언어처리*, 생능출판사.

김현준, 정재은 (2004), “가중치가 부여된 베이지안 분류를 이용한 스팸 메일 필터링시스템”, *한국정보과학회 논문지*, vol. 31, no. 8, pp. 1092-1100.

민도식, 송무희, 손기준, 이상조 (2003), “SVM 분류 알고리즘을 이용한 스팸 메일 필터링”, *한국정보과학회 논문지*, vol. 30, no. 1, pp. 552-554.

이상호 (2005), “자동 생성 메일계정 인식을 통한 스팸 필터링”, *정보과학회 논문지*, vol. 32, no. 8, pp. 378-384.

임정택, 김형준, 강승식 (2005), “나이브 베이지안 분류자와 메일 주소 유효성 검사를 이용한 스팸 메일 필터링 시스템”, *정보과학회 논문지*, vol. 32, no. 2, pp. 523-525.

한국정보보호진흥원 (2004), *알기쉬운 스팸 대응 현황 자료집*, http://www.kisa.or.kr/index.jsp.

한국정보보호진흥원 (2002), *이메일 추출 방지 프로그램의 원리 및 기능분석*, http://www.kisa.or.kr/index.jsp.

Daelemans, W., Zavrel, J. and Ko, van der S. (2004), *TiMBL: Tilburg Memory-Based Learner Version 5.1 reference guide*, Tilburg University, ILK Technical Report, ILK-0104, http://ilk.kub.nl/download/pub/papers/ilk0402.pdf.

Dasarathy, B. V. (1991), *Nearest Neighbor (NN) norms: NN pattern classification techniques*, McGraw-Hill Companies Inc.

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

Graham, P. (2002), "A plan for SPAM", http://paulgraham.com/spam.html.

Hanani, U., Shapira, B. and Shoval, P. (2001), "Information filtering: Overview of issues, research and systems", *User Modeling and User-Adapted Interaction,* vol. 11, no. 3, pp. 203-259.

Haykin, S. (1998), *Neural Networks, second edition,* Prentice-Hall Inc.

ITU, (2004), "SPAM in the information society: Building frameworks for international cooperation", http://www.itu.int/osg/spu/publication/.

Kim, J. and Oard, D. W. (2001), "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.

Mertz, D. (2002), *SPAM Filtering Techniques: Six approaches to Eliminating Unwanted E-mail*, http://www-128.ibm.com/developer/linux/library/spamf.html.

Mitchell, T. M. (1997), *Machine Learning*, McGraw-Hill Companies Inc.

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

Oard, D. W. and Marchionini, G. (1996), *A Conceptual Framework for Text Filtering*, Maryland University, Technical Report, EE-TR-96-25 CAR-TR-830 CLIS-TR-96-02 CS-TR-3643, www.ee.umd.edu/filter.ps.

Quinlan, J. R. (1993), *C4.5: Programs for machine learning*, Morgan Kaufmann Publishers Inc.

Solomonoff, R. J. (2002), *Progress in Incremental Machine Learning*, Preliminary Report, http://world.std.com/rjs/nips02.pdf.

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

Vapnik, V. and Cortes, C. (1995), "Support vector networks, machine learning", *Proceedings of The 7th Annual Workshop of Computational Learning Theory*, pp. 144-152.

Wolfe, P., Scott C., and Erwin M. (2004), *Anti-SPAM Toolkit*, McGraw-Hill Companies Inc.

Zdziarski, J. (2005), *Ending SPAM: Bayesian Content Filtering and The Art of Statistical Language Classification*, No Starch Press.