情報保護學會論文誌
第18卷 第2號, 2008. 4

송·수신 이메일의 학습을 통해 긍정 오류를 줄이는 개선된 베이지안 필터링 기법*

김 두 환†, 유 종 덕, 정 수 환‡
숭실대학교

Improved Bayesian Filtering mechanism to reduce the false positives by training both Sending and Receiving e-mails*

Doohwan Kim†, Jongduck You, Souhwan Jung‡
Soongsil University

요 약

본 논문에서는 기존의 베이지안 필터링 방식에서 발생하는 긍정 오류를 줄이기 위한 개선된 베이지안 필터링 기법을 제안한다. 기존의 베이지안 필터링 방식에서는 이메일 서버에서 학습한 DB를 일괄적으로 개별 사용자들에게 적용한다. 또한 수신 이메일 위주의 학습 방식은 양질의 정상 DB를 학습하는데 어려움을 준다. 이러한 문제로 인해 기존의 베이지안 필터링 기법에서는 정상 이메일을 스팸 이메일로 판단하는 긍정 오류가 발생한다. 제안 기법에서는 사용자의 송신 이메일을 양질의 정상 DB 정보로 판단하여 베이지안 정상 DB에 자동으로 학습한다. 뿐만 아니라 개별 사용자에게 독립적인 베이지안 DB를 제공하여 사용자 개개인의 이메일 송·수신 특성을 고려한 필터링 서비스를 제공한다. 제안 기법은 기존의 베이지안 필터링 기법보다 필터링의 정확성에서 평균 3.13 % 향상된 결과를 보인다.

ABSTRACT

In this paper, we propose an improved Bayesian Filtering mechanism to reduce the False Positives that occurs in the existing Bayesian Filtering mechanism. In the existing Bayesian Filtering mechanism, the same Bayesian Filtering DB trained at the e-mail server is applied to each e-mail user. Also, the training method using receiving e-mails only could not provide the high quality of ham DB. Due to these problems, the existing Bayesian Filtering mechanism can produce the False Positives which misclassify the ham e-mails into the spam e-mails. In the proposed mechanism, the sending e-mails of the user are treated as the high quality of ham information, and are trained to the Bayesian ham DB automatically. In addition, by providing a different Bayesian DB to each e-mail user respectively, more efficient e-mail filtering service is possible. Our experiments show the improvement of filtering accuracy by 3.13 %, compared to the existing Bayesian Filtering mechanism.

Keywords : Bayesian Filtering Mechanism, False Positives, Ham DB, Sending E-mails, Train

접수일: 2007년 8월 17일; 채택일: 2008년 1월 30일
* 본 연구는 숭실대학교 교내연구비 지원에 의해 이루어진 연구 결과임.
† 주저자, shapja@cns.ssu.ac.kr
‡ 교신저자, souhwanj@ssu.ac.kr

Ⅰ. 서 론

이메일은 인터넷 환경에서 정보 공유의 한 방식으로 널리 사용되어지고 있다. 그러나 전 세계적으로 전송되어지는 이메일 중 약 50%는 이메일 수신자가 원하지 않은 광고성 스팸 이메일이다[1]. 개인이나 기업은 이러한 스팸 이메일을 차단하기 위해 매년 수 억 원의 비용과 많은 시간을 낭비하고 있다. 이러한 문제를 해결하기 위해 이메일 서비스를 제공하는 ISP (Internet Service Provider) 사업자들은 단순 패턴 필터링 기법과 지능형 패턴 필터링 기법 등과 같은 다양한 스팸 이메일 차단 기법을 사용한다.

단순 패턴 필터링 방식은 적용 및 구현이 용이하지만 사용자가 필터링하고자 하는 단어와 블랙리스트, 또는 IP 등의 목록을 직접 필터링 리스트에 업데이트해야 하기 때문에 매우 불편하다[2]. 이러한 단순 패턴 필터링 방식의 문제를 보완하기 위해 지능형 패턴 필터링 기법이 사용된다. 지능형 패턴 필터링 기법의 대표적인 방식으로 베이지안 필터가 있다. 베이지안 필터링 기법은 이메일 서버에서 스팸 이메일과 정상 이메일에 대한 구분을 명확히 하여 베이지안 DB에 학습해야만 지속적으로 높은 필터링 성능을 보일 수 있다[3]. 하지만 이러한 베이지안 DB의 학습을 위해서는 사용자의 피드백 과정이 필수적으로 사용되어야 하는 불편함이 있다[4][5]. 또한 이메일 서버에서 사용자가 수신한 이메일 위주로 베이지안 DB를 학습할 경우, 정상 DB와 스팸 DB의 학습 불균형을 가져다주며 사용자 개개인의 이메일 송·수신 패턴을 고려하지 않는 일괄적인 필터링을 하게 함으로써 정상 이메일을 스팸 이메일로 판단하는 긍정 오류(false positive)를 발생 시킨다. 베이지안 필터의 이러한 문제점을 극복하기 위해 베이지안 DB에 존재하는 토큰(단어)정보에 가중치를 부여하여 필터링 기준을 보다 명확하게 하는 가중치 기반의 베이지안 필터링 기법이 제안되었으며[5], 베이지안 필터를 독립적으로 사용하지 않고 단순패턴 필터링 방식과 계층적으로 조합하여 사용하는 방식이 제안되고 있다[6][7][8]. 그러나 이러한 기법들은 베이지안 DB를 구축할 때 여전히 수신 이메일에 대해서만 학습함으로 정상 DB과 스팸 DB 학습의 불균형을 발생시킨다. 또한 이메일 서버에서 학습한 베이지안 DB를 모든 사용자에게 일괄적으로 적용하기 때문에 사용자 개개인의 이메일 송·수신 특성에 맞는 이메일 필터링을 제공하기 어렵다. 따라서 위 방식들을 개선하여 긍정 오류를 해결할 수 있는 필터링 방식이 필요하다.

본 논문에서는 이메일 필터링 시 발생하는 긍정 오류의 발생비율을 줄이고, 사용자 개개인의 이메일 송·수신 패턴을 고려하는 베이지안 필터링 기법을 제안 한다. 제안하는 기법은 각 사용자에 대한 베이지안 DB를 독립적으로 구축하고, 사용자의 수신 메일뿐만 아니라 송신 메일에 대해서도 자동으로 베이지안 DB에 학습하도록 한다. 이렇게 학습된 DB는 이메일 필터링 시 개별 사용자에 대한 이메일 필터링 기준을 명확히 해 줌으로써 긍정 오류의 발생비율을 줄일 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 기존의 베이지안 필터링 방식과 한계점을 분석한다. 3장에서는 제안하는 베이지안 필터링 기법과 이메일 시스템을 설명하고 4장에서는 기존의 필터링 방식과 제안하는 필터링 방식을 비교 및 분석을 한다. 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 기존의 베이지안 필터링 방식과 문제점

이메일을 필터링 할 때 발생하는 긍정 오류 문제는 기존의 단순 패턴 필터링 방식에서 자주 발생하며 지능형 패턴 필터링 방식에서도 발생 가능하다. 이번 장에서는 지능형 패턴 필터링 방식의 기술을 분석하고 그 취약성을 설명한다.

2.1 지능형 패턴 필터링 방식 분석

2.1.1 기존의 베이지안 필터[3]

베이지안 필터링 기법은 지능형 패턴 필터링 방식 중에서 가장 널리 사용되는 필터링 기법이다. 먼저 수신된 이메일은 베이지안 필터 내부의 토크나이저 모듈을 통해서 여러 개의 토큰 (단어)들로 나누어진다. 각 토큰들은 기존에 학습된 정상 DB와 스팸 DB를 통해, 정상 이메일에서의 출현 빈도수와 스팸 이메일에서의 출현 빈도수를 가지게 된다. 이러한 출현 빈도수는 해당 토큰의 정상과 스팸확률 값을 계산하는데 사용된다. 각 토큰의 정상과 스팸 확률 값들 중 가장 확률 값이 높은 몇 개의 토큰만이 식(1)의 '베이지안 포뮬러' 공식의 입력 값이 되어 수신된 이메일의 전체 정상과 스팸 확률 값의 계산에 사용된다.

$$\Pr(spam|words)=\frac{\Pr(words|spam)\Pr(spam)}{\Pr(words)} \quad (1)$$

전체 확률 값이 일정 기준 이상이면 해당 이메일은 스팸으로 판단되고, 그 이하이면 정상으로 판단된다. 판단 결과에 사용된 토큰들은 정상 DB와 스팸 DB에 자동으로 학습된다.

2.1.2 규칙 기반과 베이지안 필터의 조합된 방식[5]

위 기법은 규칙 기반의 단순 패턴 필터링 기법과 베이지안 필터링 기법을 계층적으로 조합하여 사용하는 방식이다. 규칙기반의 필터링을 통해 우선적으로 수신자와 신뢰적 관계에 있는 송신자의 이메일만을 추출하여 정상 이메일로 판단하고 그렇지 않은 이메일에 대해서는 베이지안 필터를 통해 필터링하는 시스템이다. 텍스트로 이루어진 이메일은 규칙 기반 필터를 먼저 통과하고 HTML로 이루어진 이메일은 컨버터를 통해 텍스트로 변환되어 규칙 기반 필터를 통과한다. 베이지안 필터는 (1)에서 설명한 베이지안 필터와 동일한 동작과정을 거친다.

2.1.3 가중치가 부여된 베이지안 필터[6]

가중치가 부여된 베이지안 필터링 방식은 이메일로부터 분류된 각 토큰에 가중치를 부여하는 방식이다. 일반적으로 스팸 이메일에서 출현 빈도수가 높은 토큰들은 다른 스팸 이메일에서도 발견될 확률이 높기 때문에 해당 이메일에서 스팸 빈도수가 높은 토큰이 발견되었을 경우 그 이메일은 스팸일 확률이 높다. 출현 빈도수를 바탕으로 가중치를 부여할 경우, 기존의 베이지안 필터링 기법보다 향상된 성능을 보일 수 있다.

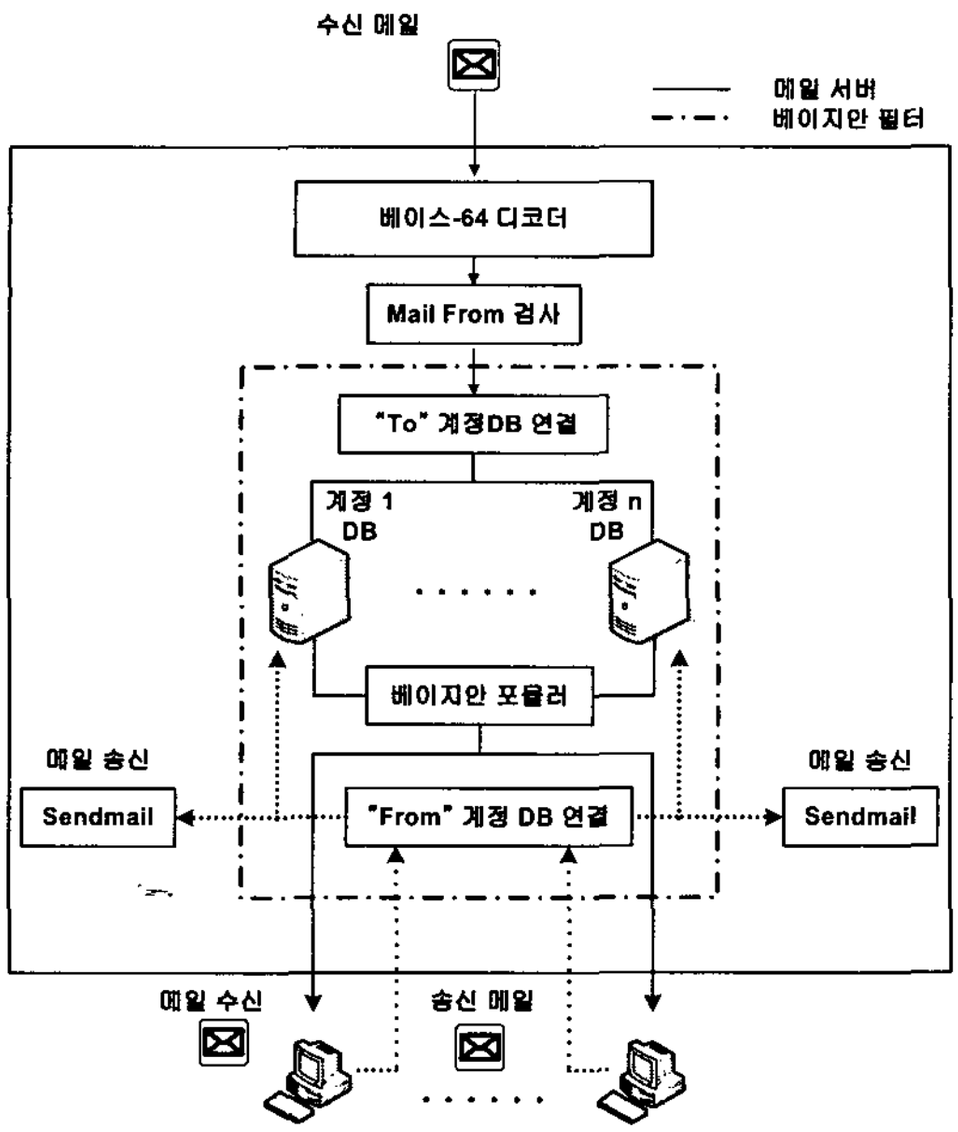
2.2 지능형 패턴 필터링 방식의 문제점

일반적으로 사용되는 베이지안 필터의 경우, 초기 학습과정에서 사용자 피드백 과정이 필수적으로 행해져야 지만 양질의 필터링 데이터베이스를 학습할 수 있으며, 이러한 피드백 과정 없이는 높은 필터링 성능을 유지하기가 어려운 단점이 있다. 또한 이메일 서버에서 학습한 필터링 DB를 다수 사용자에게 동일한 기준으로 적용하기 때문에 사용자의 이메일 특성을 고려하지 못한다. 기존의 이러한 베이지안 필터의 단점을 극복하고자 규칙기반 필터와 베이지안 필터의 조합된 방식을 사용하는 필터링 기법이 제안되었으며, 특정 단어에 가중치를 부여하여 필터링 판단기준을 보다 명확히 하는 가중치가 부여된 나이브 베이지안 필터링 기법이 제안되었다. 규칙기반 필터와 베이지안 필터의 조합된 방식은 정상 이메일의 필터링 성능에서는 우수한 성능을 보이지만 사용자가 일일이 필터링을 위한 단어나 키워드를 직접 입력해야하는 단점이 있으며, 이메일 서버에서 필터링 규칙을 일괄적으로 다수 사용자에게 적용시킨다. 가중치가 부여된 나이브 베이지안 필터링 기법도 마찬가지로 초기학습을 위한 피드백이 필수적이며, 잘못된 가중치는 더 높은 필터링 오류를 발생 시킬 수 있다. 또한 지능형 패턴 필터링 기법과 이를 보완한 여러 기법들은 베이지안 DB를 학습 할 때 수신 이메일에 한하여 DB를 학습하기 때문에 정상 DB와 스팸 DB의 학습에 취약성이 존재한다. 또한 이메일 서버에서 학습된 DB를 다수 사용자에게 일괄적으로 적용함으로써 사용자 개개인의 이메일 송·수신 패턴을 고려하지 못한 필터링을 하게 되어 긍정 오류나 부정 오류와 같은 문제점을 발생시킨다.

베이지안 필터의 성능 개선을 위한 가장 효과적인 방법은 스팸 이메일과 정상 이메일에 대한 기준을 명확히 구분하여 베이지안 DB에 학습하고, 스팸 DB와 정상 DB의 학습 균형을 이루는 것이다. 또한 이메일 사용자는 학생, 직장인, 일반 가정주부 등 다양한 그룹으로 구분 지을 수 있으며 각 그룹들은 각기 다른 스팸 이메일과 정상 이메일의 판단 기준을 가지기 때문에 사용자 개개인의 이메일 송·수신 특성에 맞게 독립적인 베이지안 DB를 구축해야 한다.

Ⅲ. 제안하는 베이지안 필터링 메커니즘

논문에서 제안하는 베이지안 필터링 기법은 송신 이메일을 베이지안 필터링 DB에 자동으로 학습함으로써 정상 DB의 양을 늘려주어 베이지안 DB의 학습 균형을 가져다준다. 또한 각 사용자 계정별로 독립적인 필터링 DB를 구축함으로써 각각의 이메일 사용자에 대해서 이메일 송·수신 패턴을 고려한 필터링이 가능하도록 한다. 따라서 제안하는 기법을 이메일 서버에 적용하여 사용할 경우 기존의 베이지안 필터에서 발생할 수 있는 긍정 오류 문제를 개선하고 각 이메일 사용자에 따른 독



〔그림 1〕 제안하는 메일 필터링 기법

립적인 DB 구축을 통해 사용자에게 종속적인 이메일 필터링이 가능하도록 한다.

제안 기법에서, 수신된 이메일은 디코더를 통해 텍스트로 변환되고 수신 이메일의 "To" 주소와 맵핑되는 사용자 계정 DB에 연계된다. 사용자 계정 DB를 통해 이메일에서 분류된 각 토큰들은 스팸과 정상 확률 값을 가지게 된다. 각 토큰의 확률 값은 '베이지안 포뮬러'의 입력 값이 되어 수신된 이메일의 스팸 확률 계산에 사용된다. 송신 이메일의 경우에는 해당 이메일의 "From" 주소와 맵핑되는 사용자 계정 DB로 맵핑된다. 송신 이메일은 토큰으로 자동 분류되어 사용자 계정의 정상 DB에 학습된다. 학습 이후에 해당 이메일은 Sendmail 데몬을 통해 외부 이메일 서버로 전송된다.

3.1 제안 기법의 구성요소

제안하는 이메일 필터링 기법의 구성요소는 [그림 1]과 같이 이메일 서버와 베이지안 필터, 그리고 사용자 계정 DB 등으로 크게 나누어 볼 수 있다. 이메일 서버는 리눅스 기반으로 Sendmail 데몬과 POP3 또는 IMAP을 지원하며 DNS 서비스를 통해 외부 이메일 서버와 통신이 가능하다. 베이지안 필터는 Spamprobe에서 제공하는 오픈 소스로 일반적인 나이브 베이지안 필터와 동일한 기능을 가진다. 베이지안 필터는 HTML tags를 무시할 수 있으며 옵션에 따라 HTML tags를 단어로 구분지어 학습할 수 있다. 베이지안 필터는 이메일 헤더와 본문의 내용을 토큰으로 분류하여 DB에 학습한다. 제안 기법의 베이지안 DB는 각각의 사용자 계정마다 독립적인 DB가 운용되며 각 사용자 계정으로 송·수신된 이메일은 DB에 자동으로 학습되어 사용자의 이메일 송·수신 패턴이 고려된 필터링이 가능하도록 한다.

3.1.1 이메일 서버

- 리눅스 기반의 이메일 시스템
- Sendmail 데몬을 통한 SMTP 서버 기능
- Procmailrc 파일을 이용한 베이지안 필터 적용
- POP3 또는 IMAP을 통해 이메일 수신 가능

3.1.2 베이지안 필터

- Spamprobe v1.4 오픈 소스를 사용
- 이메일 서버로부터 송·수신된 이메일을 토큰으로 분류
- 토큰 정보를 개별 사용자의 DB에 검색하여 스팸 확률을 계산
- 계산된 확률 값에 따라 이메일을 스팸과 정상 폴더로 전달
- 판단된 메일의 토큰 정보를 DB에 학습

3.1.3 사용자 계정 DB

- PBL(Peter's Program based Library) DB를 사용
- DB의 양 : 20MB (4~5,000여개 이메일)
- 300GB의 이메일 서버 용량에서 최대 15,000명까지 수용 가능
- 각 사용자의 이메일 계정에 대해 독립적인 DB를 구축

3.2 제안 기법의 베이지안 DB 학습 방법

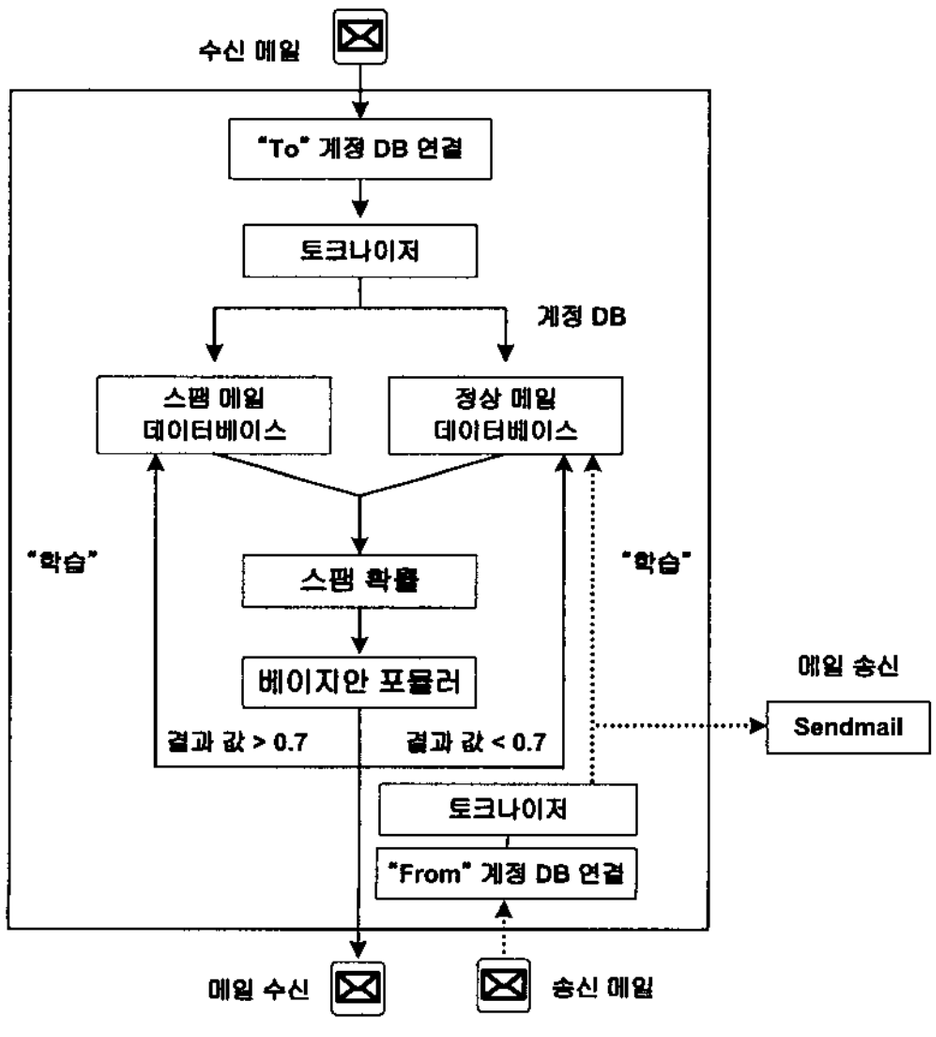
3.2.1 수신 이메일의 베이지안 DB 학습

[그림 2]와 같이 사용자 계정에 대한 각각의 DB는 수신되는 메일의 "To" 주소 계정마다 생성 된다. 이후에 수신되는 메일의 "To" 주소를 우선적으로 검사하여 해

당 이메일 계정의 DB에 이메일을 전달한다. 전달된 이메일은 베이지안 필터의 토크나이저를 통해 여러 개의 토큰으로 나뉜다. 각각의 토큰들은 해당 DB에 맵핑되어 각 토큰에 대한 스팸과 정상 확률이 결정된다. 계산된 확률 값들은 '베이지안 포뮬러'를 통해 이메일의 전체 스팸 확률 값 계산에 사용되고, 그 결과 값이 일정 기준 이상이면 스팸으로 판단하고 그 이하이면 정상으로 판단한다. 판단 결과에 따라 수신 이메일은 사용자 계정의 스팸 또는 정상 이메일 박스에 저장된다. 판단된 이메일의 토큰 정보는 다시 베이지안 DB에 자동으로 학습된다.

3.2.1 송신 이메일의 베이지안 DB 학습

송신하는 이메일에 대한 DB는 송신 이메일의 "From" 주소 계정마다 각각 생성된다. 해당 송신 이메일에 대한 DB의 학습 방법은 수신 이메일에 대한 DB의 학습 방식과 동일하다. 수신되는 이메일의 "To" 주소 계정과 송신하는 이메일의 "From" 주소 계정이 일치하면 동일한 DB에 해당 이메일의 토큰 정보가 학습되는 것이다. 사용자는 상당히 많은 양의 스팸 이메일과 상대적으로 적은 양의 정상 이메일을 동시에 수신하게 되는데, 수신 이메일에 대해서만 DB를 학습 할 경우, 정상 이메일에 대한 DB의 용량이 스팸 이메일에 대한 DB의 용량보다 상대적으로 작아져서 정상 이메일의 스팸 확률 값이 높아지는 문제가 발생할 수 있다. 일반적으로 사용자는 자신이 송신한 이메일의 패턴에 따라 답신을 받게 되는 경향이 있으므로, 송신하는 이메일의 데이터를 베이지안 정상 DB에 학습 할 경우 사용자의 이메일 송·수신 패턴에 따른 DB의 구축이 가능해진다.



〔그림 2〕 송·수신 이메일의 DB 학습 방법

Ⅳ. 제안 기법 성능 평가

본 장에서는 기존의 베이지안 필터링 기법과 제안 기법의 성능을 비교 및 분석한다. 제안 기법의 성능 비교를 위해 이메일 테스트 환경을 구축하고 빠른 성능평가를 위한 이메일 분류기, 이메일 대량 전송 모듈 등을 개발하였다. 제안 기법의 성능 평가를 위해 이메일의 사용자 군을 학생과 회사, 그리고 일반 가정주부로 나누었으며 제안 기법이 각 사용자 군에 따라 얼마만큼의 성능 개선 효과가 있는지를 평가한다. 성능 평가의 기준이 되는 파라미터는 이메일 필터링의 정확성과 오판 정도를 나타내는 정확도(Precision)와 재현율(Recall)을 사용한다[9].

4.1 테스트 환경

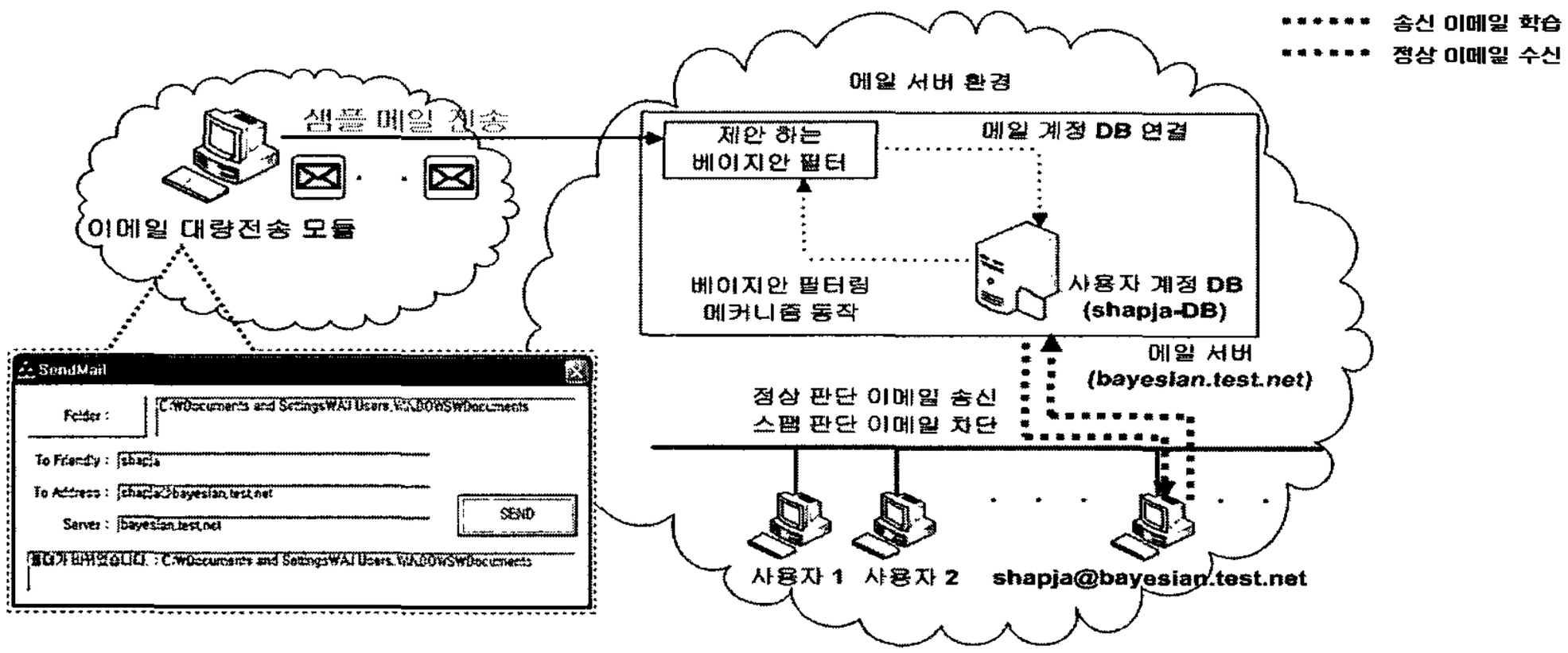
기존의 베이지안 필터링 기법과 제안 기법의 성능 비교를 위해 [그림 3]과 같은 이메일 테스트환경을 구축하였다. 데스크탑 PC에는 이메일 분류기와 이메일 대량 전송 모듈을 탑재하여 테스트에 사용되어지는 샘플 이메일을 대량으로 전송한다. 테스트에 사용된 데스크탑 PC는 3.41GHz, 2GB RAM의 시스템이며, 이메일 서버는 2.5GHz, 1.4GB RAM의 시스템이다. 사용자 별로 PBL DB를 각각 구성하며 사용자의 송신 이메일과 수신 이메일은 자동으로 DB에 학습된다.

4.2 성능 평가를 위한 개발 모듈

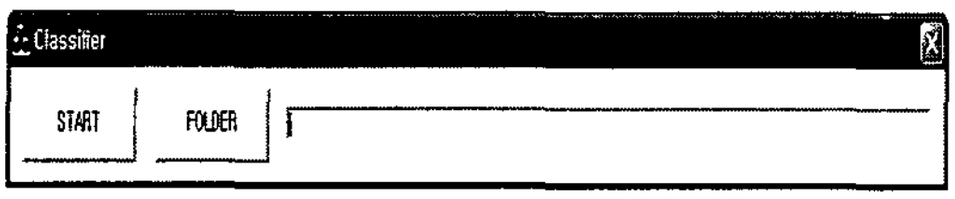
본 논문에서는 대량의 이메일을 정상과 스팸으로 자동으로 분류하고 각각 분류된 이메일을 이메일 서버로 대량으로 자동 전송할 수 있는 모듈을 개발하여 사용자에게 편의성을 제공하고 베이지안 DB의 학습과 필터링 성능 테스트를 보다 효율적으로 하였다.

4.2.1 이메일 분류기

[그림 4]는 대량의 정상 이메일과 스팸 이메일을 자동으로 분류하기 위한 모듈이다. 대량의 이메일이 저장



〔그림 3〕 이메일 대량 전송 모듈을 이용한 제안 기법 테스트 환경

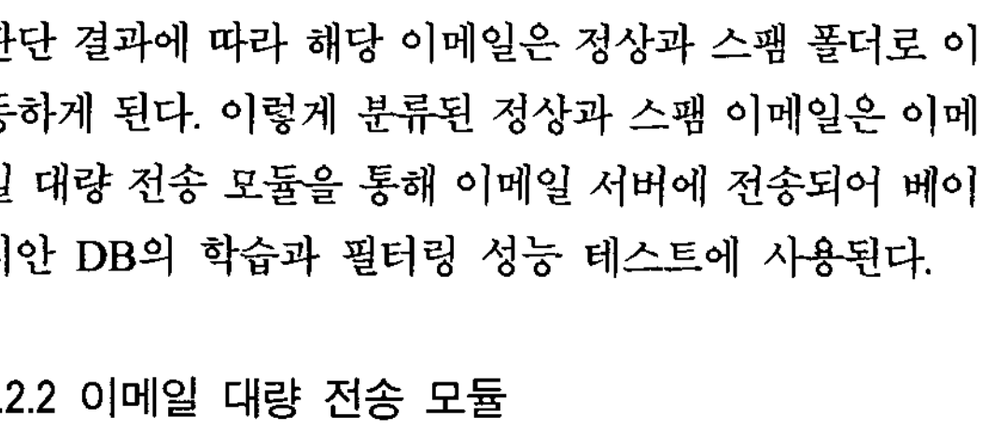


〔그림 4〕 이메일 분류기

된 폴더를 지정하여 'Start' 버튼을 누르면 해당 폴더에 저장된 이메일의 정보가 차례로 화면에 나타나고 선택된 이메일을 사용자가 정상 또는 스팸으로 판단하면 그 판단 결과에 따라 해당 이메일은 정상과 스팸 폴더로 이동하게 된다. 이렇게 분류된 정상과 스팸 이메일은 이메일 대량 전송 모듈을 통해 이메일 서버에 전송되어 베이지안 DB의 학습과 필터링 성능 테스트에 사용된다.

4.2.2 이메일 대량 전송 모듈

[그림 5]는 이메일 대량 전송 모듈로, 이메일 대량 전송 모듈은 이메일 분류기를 통해 수집된 이메일을 SMTP (Simple Mail Transfer Protocol)을 통해서 대량으로 이메일 서버에 전송하는 기능을 한다. 이메일 분류



〔그림 5〕 이메일 대량 전송 모듈

기를 통해 정상과 스팸 폴더에 저장된 이메일은 .eml 형식의 이메일 원문형태로 저장된다. 먼저 To address field에 이메일을 송신하고자 하는 사용자의 이메일 계정 주소를 입력한다. 다음으로 Server field에 이메일 계정을 관리하는 이메일 서버의 주소를 입력하여 SEND 버튼을 누른다.

그렇게 되면 자동으로 폴더 안에 저장된 모든 .eml 파일의 To 주소가 입력한 To address field의 주소로 바뀌어 입력된 이메일 서버의 주소로 자동으로 전송된다. 이메일 대량 전송 모듈을 통해서 1,000여개의 이메일을 이메일 서버로 전송할 때 이메일 용량에 따라 약 20분에서 30분가량이 소요된다. 이메일 대량 전송 모듈을 사용하게 되면 사용자가 일일이 하나의 이메일을 이메일 서버로 피드백 하는 번거로움을 줄일 수 있으며 자동으로 대량의 이메일을 전송할 수 있어 베이지안 DB 학습과 베이지안 필터의 성능 테스트에 매우 효과적으로 사용될 수 있다.

4.3 성능평가 파라미터

제안 기법의 성능 평가를 위해서 이메일 필터링 시 발생하는 긍정 오류와 부정 오류의 비율을 검사하고, 그 비율을 이용하여 이메일 필터링의 성능을 평가하기 위한 기준이 되는 정확도 (Precision)와 재현율 (Recall)을 식 (2), (3)과 같이 정의하여 사용한다.

$$\text{정확도 } P = \frac{\text{'스팸'으로 분류된 실제 '스팸' 메일 수}}{\text{'스팸'으로 분류된 메일 수}} \qquad (2)$$

$$\text{재현율 } R = \frac{\text{'스팸'으로 분류된 실제 '스팸' 메일 수}}{\text{전체 '스팸' 메일 수}} \quad (3)$$

스팸 이메일 판단하는 기준 값은 0.7이며, 기준 값을 통해 판단된 스팸과 정상 이메일의 개수를 조사한다. 또한 스팸으로 오판된 정상 이메일의 개수와 정상으로 오판된 스팸 이메일의 개수를 파악하여 정확도와 재현율을 구하기 위해 사용한다.

4.4 성능평가 결과 및 분석

다음 [표 1]은 기존의 베이지안 필터와 제안 기법의 DB 학습에 사용된 이메일 수와 테스트에 사용된 이메일의 수를 나타낸다. 제안 기법의 성능을 분석하는데 있어 객관성을 유지하기 위해 사용자 군을 학생과 기업, 그리고 일반 가정주부로 분류하여 테스트 하였으며, 각 사용자 군은 개별적으로 학습된 DB를 가진다. [표 1]에서 사용된 이메일은 본 논문의 4.2.1절 에서 소개한 이메일 분류기를 통해 분류된 이메일이다.

[표 2]는 기존의 베이지안 필터링 기법과 가중치 기반의 베이지안 필터링 기법, 그리고 제안 기법의 필터링 성능을 비교 결과이다. [표 2]에서 살펴볼 수 있듯이 제안 기법은 기존의 베이지안 필터링 기법과 가중치 기반의 베이지안 필터링 기법보다 정확도 측면에서 높은 필터링 성능을 가진다. 일반 학생의 경우, 정확도는 기존의 베이지안 기법보다 2.76 % 향상된 성능을 보이며 가중치 기반의 베이지안 기법보다 3.82 % 향상된 성능을 보인다. 회사원의 경우, 정확도는 기존의 베이지안 기법보다 3.57 % 향상된 성능을 보이며 가중치 기반의 베이지안 기법보다 2.38 % 향상된 성능을 보인다. 가정주부의 경우, 기존의 베이지안 기법과 가중치 기반의 베이지안 기법보다 각각 4.3% 와 1.96%의 향상된 결과를 보인다. 따라서 제안 기법은 기존의 베이지안 필터링 기법보다 평균 3.13% 향상된 필터링 정확성을 제공한다.

[표 2]의 성능 비교에서 살펴볼 수 있듯이 제안 기법은 이메일 필터링 시 발생하는 긍정 오류의 발생비율을 줄여주며, 이메일 스팸의 필터링 성능에서 가장 중요시되는 정확도 (Precision)에서 향상된 성능을 보여준다. 사용자 1의 일반 학생 군의 경우, 긍정 오류가 발생한 정상 이메일은 주로 국내외 학회에서 송신한 워크숍과

〔표 1〕 베이지안 DB 구성

	베이지안 DB 구성 및 테스트 메일 개수					
	사용자 1 (일반 학생)		사용자 2 (회사원)		사용자 3 (가정주부)	
	베이지안	제안 기법	베이지안	제안 기법	베이지안	제안기법
스팸 학습	379	379	355	355	369	369
정상 학습	수신 : 237	수신 : 237 송신 : 169	수신 : 229	수신 : 229 송신 : 165	수신 : 200	수신 : 200 송신 : 99
총 학습	616	785	584	749	569	668
총 테스트	325	325	325	325	325	325

〔표 2〕 제안 기법과 베이지안 필터링 성능 비교

	성능 비교						
	가중치 기반의 베이지안	사용자 1 (일반 학생)		사용자 2 (회사원)		사용자 3 (가정주부)	
		베이지안	제안 기법	베이지안	제안 기법	베이지안	제안 기법
긍정 오류	·	8	3	11	5	14	6
부정 오류	·	19	19	16	16	18	18
정확도 P	94.44 %	95.50 %	98.26 %	93.25 %	96.82 %	92.10 %	96.4 %
재현율 R	85.71 %	89.94 %	89.94 %	90.47 %	90.47 %	90.05 %	90.05 %
메일 학습 데이터	수신	수신	송·수신	수신	송·수신	수신	송·수신
DB 용량	·	9.063 MB	10.759 MB	8.759 MB	10.860 MB	8.168 MB	9.552 MB

관련된 안내 이메일이나 국제 표준 단체에서 송신한 메일링 리스트와 같이 학문적 성격을 가지는 이메일이다. 일반적으로 학생 군의 경우 이러한 학문적 연구 단체나 기관에 등록을 위한 이메일이나 연구와 관련된 이메일을 송·수신 하는 경향이 있다. 따라서 송신 이메일을 정상 DB에 학습할 경우 기존의 베이지안 필터에서 스팸으로 처리되는 이러한 정상 이메일을 올바르게 판단할 수 있다. 사용자 2의 회사원 군이나 사용자 3의 가정주부 군에서도 기존의 베이지안 필터에서는 회사 업무와 집안 생계활동에 관련된 회계, 금융, 광고, 홈쇼핑, 등의 정상 이메일을 스팸으로 처리하지만 제안 기법에서는 사용자의 이메일 송·수신 패턴을 고려하여 필터링 DB를 학습하기 때문에 긍정 오류의 발생비율을 줄여 줄 수 있다.

제안 기법은 기존의 베이지안 필터링 기법보다 DB의 증가 측면에서 사용자 군에 따라 평균적으로 1.2 MB 정도의 메모리 용량이 증가하는 단점이 있다. 하지만 본 논문의 3.1절의 (3)에서 제시한 이메일 서버의 용량을 통해 개인 사용자에게 약 20 MB의 메모리 용량을 할당 할 경우, 제안 기법에서 추가로 부담해야하는 메모리 부하는 전체 사용자 DB 용량의 약 6 % 정도이므로 충분히 고려할만 하다.

일반적으로 이메일 필터링 시스템의 성능 평가를 위한 파라미터로 가장 중요시되는 부분은 긍정 오류의 발생비율이다. 이메일의 사용자 측면에서, 중요한 정상 이메일이 필터링 시스템으로 인해 스팸으로 판단되어 전달되지 못하는 상황이 스팸 이메일을 정상 이메일로 판단하여 사용자에게 수신되는 상황 보다 더 큰 문제점을 발생 시킨다. 실제로 이메일의 필터링 서비스를 제공하는 소프트웨어 개발업체나 UTM (Unified Threat Management)장비 개발업체들은 부정 오류의 발생비율을 줄이기 위한 노력보다 긍정 오류의 발생비율을 개선하기 위한 방법에 관심을 가지고 있다. 본 논문에서 제안하는 기법은 소량의 베이지안 DB 확장으로 긍정 오류의 문제를 효과적으로 개선하기 때문에 실제 이메일 필터링 시스템에 적용될 경우, 개선된 필터링 성능을 보일 수 있다.

V. 결 론

본 논문에서는 사용자의 이메일 송·수신 패턴을 고려한 베이지안 필터링 기법을 제안 하였다. 기존의 베이지안 필터링 기법에서는 서버에서 수신한 이메일을 기반으로 베이지안 DB를 구축하므로 사용자 개개인이 관심을 갖는 특정 분야나 사용자 개개인의 이메일 송·수신 특성 등을 고려하지 못한다. 제안 기법에서는 사용자의 송신 이메일을 양질의 정상 DB로 학습하여 이메일 시스템의 가장 중요한 성능 지표인 긍정 오류의 발생비율을 효과적으로 개선한다. 또한 개별 사용자마다 독립적인 필터링 DB를 구축하여 각 사용자의 스팸 이메일과 정상 이메일에 대한 판단 기준을 고려한 필터링을 가능하게 한다. 본 논문에서는 제안 기법의 실험적 검증을 통해서 일반 학생, 회사원, 가정주부 등의 사용자 군에게 평균 3.13 %의 향상된 필터링 정확성을 제공한다. 따라서 제안기법은 이메일 필터링 시스템의 가장 중요한 성능평가 파라미터인 긍정 오류의 발생 비율에서 기존의 베이지안 필터링 기법과 이를 개선하기 위한 다른 기법 보다 향상된 성능을 보인다.

참고문헌

[1] Roger Wattenhofer, Gordon V. Cormack, and Christof Fetzer, "Mastering Spam A Multifaceted Approach with the Spamato Sapm Filter System," *Swiss Ferderal Institute of Technology Zurich*, 2006.

[2] Barracuda Networks. "An Overview of Spam Blocking Techniques," *http://www.innovativeidea.com*, 2006.

[3] Ricardo Silva, Richard Scheines, "Bayesian Learning of Measurement and Structural Models," *Proceedings fo the 23th International Conference on Machine Learning, Pittsburgh*, PA, 2006.

[4] Y. Li, B. Fang, L. Guo, and S. Wang, "Research of a Novel Anti-Spam Technique Based on User's Feedback and Improved Naive Bayesian Approach," *Proc. ICNS'06,* 2006.

[5] 조한철, 조근식, "나이브 베이지안 분류자와 메세지 규칙을 이용한 스팸메일 필터링 시스템," *한국정보과학회 봄 학술발표논문집*, 2002년.

[6] 김현준, 정재은, 조근식, "가중치가 부여된 베이지안 분류자를 이용한 스팸 메일 필터링 시스템 설계," *정보과학회논문지, 소프트웨어 및 응용 제*

31권 제 8호, 8월, 2004년.

[7] E. Michelakis, I. Androutsopoulos, G. Paliouras, G. Sakkis, and P. Stamatopoulos, "Filtron: A Learning-Based Anti-Spam Filter," *Proc. CEAS 2004,* 2004.

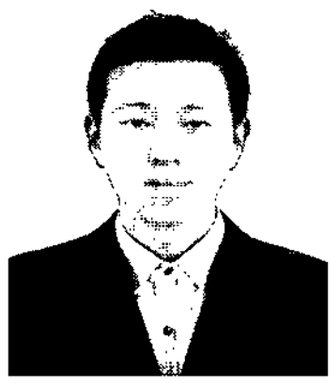
[8] V. Cheng and C.H. Li, "Personalized Spam Filtering with Semi-supervised Classifier Ensemble," *Proc. WIC/ACM,* 2006.

[9] Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Paliouras, G. and Spyropoulos, C. D., "An Evaluation of Naive Bayesian Anti-Spam Filtering," *Proc of the 11th European Conference on Machine Learning, pp.9-17,* 2000.

〈著者紹介〉

김 두 환 (Doo-Hwan Kim) 학생회원

2006년 2월 : 숭실대학교 정보통신전자공학부 졸업
2006년 8월~현재 : 숭실대학교 정보통신전자 석사과정
<관심분야> 무선네트워크 보안, VoIP 보안, 이동통신네트워크 보안,



유 종 덕 (Jong-duck You) 학생회원

2007년 2월 : 숭실대학교 정보통신전자공학부 졸업
2007년 3월~현재 : 숭실대학교 정보통신전자 석사과정
<관심분야> 차량 이동통신 보안, VoIP 보안, 네트워크 보안,

정 수 환 (Sou-hwan Jung) 종신회원

1985년 2월 : 서울대학교 전자공학과 학사
1987년 2월 : 서울대학교 전자공학과 석사
1988년~1991년 : 한국통신 전임 연구원
1996년 6월 : University of Washington 박사
1996년~1997년 : Stellar One SW Engineer
1997년~현재 : 숭실대학교 정보통신전자공학부 부교수
<관심분야> 이동 네트워크 보안, VoIP 보안, 네트워크 보안, RFID/USN 보안