

# 멀티모델 기반의 개인화된 메일 필터링 시스템

박유나<sup>0</sup> 장환 이복주  
단국대학교 컴퓨터 공학과  
{ypark, hwan, blee}@dankook.ac.kr

## Designed of personalized mail Filtering System using Support vector machines

Youna Park<sup>0</sup> Hwan Chang Bogju Lee  
Dept. of computer engineering, Dankook university

### 요 약

전자우편은 인터넷의 성장과 함께 필수적인 정보교환 수단으로 자리잡고 있다. 그 신속성과 용이성을 이용하여 많은 기업과 업체들이 손쉽게 광고 수단으로 이용하여 이로 인하여 개인과 기업에 큰 피해를 초래하고 있다. 필요한 스팸메일을 선정하여 분류하는 데 개인과 조직에 많은 정신적, 물리적인 스트레스를 요구한다. 본 논문에서는 통계적 학습 방법인 SVM을 이용하여 지속적으로 변화하는 다양한 스팸메일을 분류하고자 한다. 실험결과는 스팸메일 분류에 안정적인 성능을 보여줄 뿐 아니라 다양한 종류의 스팸메일을 카테고리별로 구분해 내는데 높은 성능을 보여준다.

### 1. 서론

인터넷을 통한 정보의 교환은 급속한 성장과 보급으로 인하여 사회 전체를 변화시키고 있다. 특히 전자우편은 신속성과 용이성으로 현재 중요한 정보 교환의 수단으로 이용되고 있다. 그러나 이러한 정보화의 역기능 중 하나로 스팸메일(Spam-mail)이 문제가 되고 있다. 스팸메일이란 수신자의 의사와 관계없이 불특정 다수에게 무작위로 발송되는 광고성 이메일을 일컫는 말이다. 전자우편은 주고 받는데 비용이 거의 들지 않기 때문에 많은 기업이나 업체들이 광고수단으로 이용하고 있다. 한국정보보호진흥원의 2001년 정보화 역기능 실태조사 보고서에 따르면 스팸메일에 대한 반응이 바로 지우는 경우가 43.6%였으며 흥미 있는 경우만 읽어보는 경우가 40.2%에 해당하였다. 무작위로 쏟아져 들어오는 스팸메일은 개인에게 정신적, 물리적, 스트레스를 줄 뿐 아니라, 조직의 업무 진행을 지연시키며, 나아가 공공 자원인 네트워크 독점에 의한 피해 등 사회 전 분야에 걸쳐 부정적인 영향을 미치고 있다. 또한 관심 있는 정보를 찾아내는 작업이 거의 불가능하다[2].

최근 스팸메일이 심각한 손실로 인하여 국내뿐 아니라 세계 여러 나라에서 법률로 지정하여 해결하려 하고 있으나 법안을 따르지 않는 스팸메일을 걸러내는 데에는 많은 어려움이 따르고 있다. 본 논문에서는 통계적 학습 방법인 Support vector machines을 사용하여 지속적으로 변화하는 다양한 내용의 스팸메일을 걸러내고 분류할 수 있도록 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 논문에서 구현에 사용된 문서 자동 분류와 사용된 기법에 관하여 언

급하였으며 3장에서는 구현한 스팸메일 분류 시스템, 4장에서는 구현된 시스템으로 실험한 결과에 대하여 언급 하고 마지막으로 5장에서는 결론과 추후 과제에 대하여 언급한다.

### 2. 관련 연구

본 절에서는 문서 자동 분류와 문서 자동 분류 기법 중 하나인 본 논문에서 사용된 서포트 벡터 머신에 관하여 논한다.

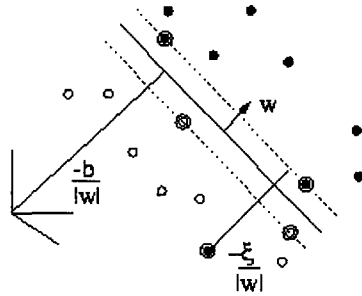
#### 2.1. 문서 자동 분류 (Automatic Text Classification)

문서 자동 분류란 하나의 문서가 주어졌을 때 여러 범주 중 가장 적절한 범주로 문서를 분류하는 것을 일컫는다. 이 연구는 1960년대부터 시작되어 1990년대에 들어서는 컴퓨터 처리 능력의 향상과 디지털 문서의 기하급수적 증가, 수작업 문서 분류의 한계로 인해 많은 수요가 있는 분야이다. 그리고 사용자의 요구에 적합한 정보의 추출을 위한 연구로 발전되고 있다. 특히 1990년도 전반부터 기계학습 (Machine Learning) 기법을 적용한 자동 문서 분류기에 관한 연구가 시작되면서 현재에 와서는 다양한 기계학습 기법의 적용이 시도되고 있다. 문서의 자동 분류 기법은 클러스터링 (Clustering)과 범주화 (Categorization)로 나눌 수 있으며 이 둘의 분류 기준은 문서가 사전에 분류 되어 있는지의 여부이다. 여기서는 사전 분류 체계에 기초하여 학습을 통하여 해당 범주의 문서들에 대한 모델을 생성하고 분류 대상이 되는 문서를 가장 적합한 범주에 할당하는 기법인 범주화(Categorization)를 이용한다. 문서 자동 분류의 구성 요소 중에 분류 속성 (Feature)을 선정하는 것을 속성 공간의 선정/축소 (Feature Selection/ Reduction)라고 하며 정보성이 낮거

나 없는 단어를 효과적으로 배제하는 작업이다. 문서 범주화 기법은 문서 전처리, 속성 선정/축소, 분류기 생성으로 구성되어 있다.

2.2. 서포트 벡터 머신 (Support vector machines)

SVM은 1995년 Vladimir Vapnik에 의해 이원 패턴 인식 문제를 해결하기 위해 제안된 방법으로 두 개의 범주를 구성하는 데이터들을 가장 잘 분리해낼 수 있는 결정면을 찾아내는 모형을 결정하는 알고리즘이 [그림 1]와 같다.



[그림 1] SVM의 구조

기본적인 선형 SVM은 데이터를 +1과 -1 클래스의 두 개의 집합으로 완전하게 분리시킬 수 있는 한계면(hyper-plane)이 존재하는 것이며, 수식으로 다음과 같이 나타낼 수 있다[4].

$$y_i \in \{-1, 1\} \quad (1)$$

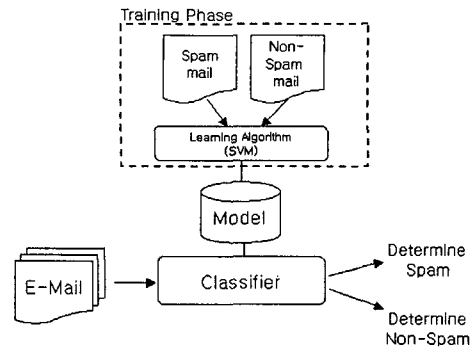
$$y_i(w^* \cdot x_i - b) \geq 1$$

w는 가중치 벡터, x는 입력벡터, b는 기준치이며, w와 b는 학습 데이터로부터 학습된다. 학습 문서 집합을  $D = \{(x_i, y_i)\}$ 라고 하면, 입력 데이터  $x_i$ 가 범주에 속하면  $y_i$ 는 +1의 값을 갖고, 속하지 않으면 -1값을 갖는다. 이를 통해 SVM은 최적의 w와 b를 찾는다.

SVM은 현재 뛰어난 성능으로 문자인식, 얼굴인식, 문서 분류 등 여러 분야에서 응용되고 있다.[4]에서 Joachims은 SVM이 높은 입력 데이터 차원, 관계 없는 속성이 적은 경우, 희소 문제를 해결하기에 적합하다고 결론지었다. 본 논문에서는 스팸메일의 분류를 위해 [4]에서 기존에 사용되어 왔던 나이브 베이지안 방법보다 문서 분류 성능이 뛰어나다고 하는 SVM을 학습 및 분류 알고리즘으로 사용하였다[9].

3. 스팸메일 분류 시스템

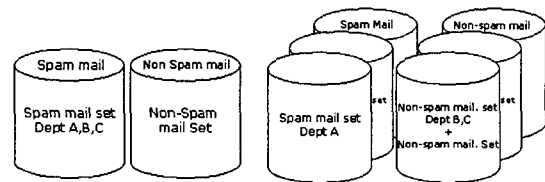
본 시스템의 목적은 학습을 통하여 새로운 메일이 들어왔을 때 스팸/비스팸 여부를 자동적으로 분류하는데 있다. 먼저 스팸메일의 분류를 위해서 SVM(Support vector machines)을 사용하여 학습하고 모델을 생성한다. 새 메일이 들어오면 SVM을 통해 생성된 분류기를 통해 스팸 여부를 판별한다. 분류를 위해 본 논문에서 구현한 시스템의 구성은 [그림 2]과 같다.



[그림 2] 시스템의 구성과 흐름도

3.1 학습 단계 (Training Phase)

스팸메일과 비스팸메일들을 먼저 데이터로 준비한다. 스팸메일은 여러 업종의 광고 메일을 사용하였으며 비스팸메일로는 광고와 관련이 없는 개인적인 내용의 메일을 선정하였다. 여기서 한글 형태소 분석기 (HAM)[6]을 이용하여 불필요한 정보를 제거하기 위하여 어근 추출과 접속사 제거를 수행하여 자질 공간을 축소한다. 추출된 키워드를 속성으로 간주하고 벡터 형태로 변환하여 데이터 베이스에 저장한다. 각 속성의 가중치는 Term Frequency (TF)를 문서 길이로 정규화한 L2 방식을 사용하였다. 벡터 형태로 저장된 문서 벡터들은 SVM알고리즘에 의해 학습되어 한계면 정보를 가지는 모델이 파일 형태로 저장된다. 모델 생성시 두 가지 방법으로 비교한다. [그림 3]과 같이 단순한 스팸/비스팸의 분류를 위한 싱글 모델과 카테고리별 분류를 위한 멀티 모델로 각각 모델을 형성한다.



[그림 3] Single model / Multi-model

3.2 분류 단계 (Classification Phase)

분류 단계에서는 메일전체를 각각 문서로 보아 스팸 여부를 판별한다. 위에서 생성된 학습 결과들과 각각의 문서를 비교하여 스팸메일에 해당하는 지를 판단하여 보고한다. 그리고 그 결과를 데이터베이스로 저장한다.

4. 실험 및 결과

4.1 실험 환경

실험 환경으로 인텔 펜티엄 1.6GHz CPU, 384 RAM을 가지는 서버를 사용하였고 한글 형태소 분석을 위한 HAM, 학습과 분류 엔진으로서 SVMlight 5.0[10]을 사용하여 VC++ 6.0과 MS-SQL을 사용하여 환경을 구축하였

다. 학습 및 테스트에 사용할 데이터들은 실제 메일로 세 가지 업종의 광고 메일을 사용하였으며 테스트에 사용하는 메일은 학습에 사용된 메일 이후에 온 메일을 사용하였다. 스팸 메일 또한 실제 메일을 사용하였으며 구성은 [표 1]과 같다.

	Training Data	Test Data
Spam	490	100
Adult	150	30
Product	260	52
Finance	40	8
Etc.	40	10
Non-Spam	200	40
Total	690	140

[표 1] 데이터 구성

#### 4.2 실험 결과 및 분석

문서 분류의 성능을 평가하기 위한 기준으로 정확도와 에러율을 사용한다. 시스템을 구현하여 싱글 모델로 분류 성능을 비교한 것이 [표 2]이다.

	Precision	Error-rate
Spam	92.2%	7.77%
Non-Spam	92.5%	7.5%
Avg.	92.35%	7.63%

[표 2] 스팸/ 비스팸 메일의 분류 성능

스팸과 비스팸 메일을 사용하여 분류한 경우 정확도와 에러율로 보아 비슷한 수준으로 안정적인 성능을 보이고 있다.

Category	Precision	Error-rate
Adult	93.3%	6.66%
Product	96.1%	3.84%
Finance	100%	0%
Avg.	96.46%	3.5%

[표 3] 카테고리별 분류 성능

[표 2]의 싱글모델과 비교하여 멀티모델을 통한 성능인 [표 3]에서 스팸 메일을 카테고리별로 분류하여 평가해 보았다. 각 카테고리별로 나누어 분류해본 경우 기존의 스팸과 비스팸으로 나누어 분류한 경우가 평균 92% 정도인데 비하여 평균 96% 이상의 더 높은 정확도를 보였다. 이를 토대로 멀티모델을 이용한 개인화된 메일 필터링을 가능하게 할 수 있을 것이라고 할 수 있다.

#### 5. 결론

본 논문에서는 스팸 메일 필터링을 위하여 Support vector machine을 시스템을 제안하고 구현하였으며 두

가지 모델을 형성하여 테스트하였다. 제안한 시스템은 본 실험을 통하여 스팸 메일과 비스팸 메일의 경우 비슷한 수준의 안정된 성능을 보였다. 또한 멀티모델을 사용하여 카테고리 별로 나누어 분류한 경우에는 싱글 모델에 비해서 높은 성능을 보였다. 이를 통하여 기존에 사용되었던 다른 알고리즘에 비하여 단순한 스팸 메일 분류 뿐 아니라 개인화된 메일 필터링을 가능하게 할 수 있다. 또한 기존의 스팸 메일 탐지 시스템에서 주로 사용되는 키워드 검색 기반 또는 룰베이스(Rule-Base)를 이용한 다중 필터를 사용할 경우 더욱 좋은 결과를 얻을 수 있을 것이다. 최근 많은 스팸 메일이 텍스트에서 이미지의 형태로 바뀌고 있다. 하지만, 기존의 스팸 메일 탐지 시스템은 텍스트에만 의존하기 때문에 이를 효율적으로 분류하기 힘들다. 그러므로 본 논문의 향후 과제는 텍스트 이외의 이미지와 같은 다른 포맷의 콘텐츠들까지 확대 적용한 스팸 메일 탐지 시스템에 관한 연구이다.

#### 참고 문헌

- [1] 이복주, 박유나, 조남섭, “ 기밀정보 유출 방지를 위한 전자우편 모니터링 시스템”, 데이터 마이닝 학회 논문지, 2002.
- [2] 안수산, 신경식, “ 데이터마이닝 기법을 활용한 스팸 메일 분류 및 예측모형 구축에 관한 연구”, 한국 지능 정보 시스템 학회 논문지, 2000.
- [3] Bogju Lee et al., “ TWIMC: An Anonymous Recipient E-mail System”, Proceedings of the 15th International Conference on Industrial and Engineering Application of Artificial Intelligence and Expert Systems (IEA/AIE-2002), Cairns, Australia, 2002.
- [4] T. Joachims, TextCategorization with Support Vector Machines - Learning with Many Relevant Features, Proc. of the European Conference on Machine Learning, 1998.
- [5] T. Joachims, Learning to Classify Text Using Support Vector Machines II Methods, Theory and Algorithms, Kluwer Academic Publishers, 2002.
- [6] <http://npl.kookmin.ac.kr>
- [7] <http://svmlight.joachims.org/>
- [8] 정영미. (1993). 정보 검색론. 서울: 구미무역 출판부
- [9] 조한철, 조근식, 나이브 베이지안 분류자와 메시지 규칙을 이용한 스팸 메일 필터링 시스템. 한국 정보 과학회 논문지. 2002.
- [10] 민도식, 송무희, 손기준, 이상조, “ SVM분류 알고리즘을 이용한 스팸 메일 필터링”, 한국 정보 과학회 논문지, 2003.