

## 사용자 행동 패턴을 기반으로 가중치를 부여한 스팸 메일 필터링

한아성<sup>0</sup>, 김현준, 조근식  
인하대학교 컴퓨터.정보공학과  
{tinypre<sup>0</sup>, dannis}@eslab.inha.ac.kr, gsjo@inha.ac.kr

### Weighting based User Behavior Pattern for Filtering Spam Mail

Asung Han<sup>0</sup>, Hyun-jun Kim, Geun-sik Jo  
School of Computer Science & Engineering,  
Inha University

#### 요 약

스팸 메일의 비율은 지속적으로 증가하여 최근 전체 이메일의 92.6%가 스팸 메일인 것으로 드러났다. 본 논문에서는 시간의 경과에 따른 사용자의 액션 패턴을 기반으로 사용자의 관심에 따른 가중치를 적용하여 스팸 메일 여부를 가리는 방법을 다룬다. 액션간의 관계와 액션 사이의 시간에 따라 가중치를 차별화함으로써 얼마나 높은 필터링 성능을 보일 수 있는 지, 또한 학습 속도 향상에 얼마나 기여할 수 있는 지를 측정할 것이다. 실험에서는 실제 메일 데이터를 이용하여 베이지안 분류자, 가중치가 부여된 베이지안 분류자와 본 논문이 제안하는 시스템의 학습 성능의 향상 속도를 비교할 것이다. 또한 제안된 시스템이 Concept Drift와 적응 학습, 그리고 개인화를 어떻게 다룰 지를 보일 것이다.

#### 1. 서 론

최근 스팸 메일의 비율은 지속적으로 증가하여 매년 20%의 증가율을 보여왔다[1]. 주요 인터넷 보안업체의 조사에 따르면 2006 년 하반기에는 전체 이메일의 92.6%가 스팸 메일인 것으로 드러났다[2]. 이를 위해 다양한 메일 여과(filtering) 방법이 제안되었으며 지금도 활발히 연구되고 있다. 베이지안 분류(Bayesian classification)를 비롯, SVM(Support vector machine), 신경망을 이용한 분류(Neural network classification), 사례기반추론(Case-based reasoning) 등의 기계 학습(Machine Learning)과 통계적 기술 등은 스팸 메일 필터링을 위해 사용되는 대표적인 필터링 기술의 예이다.

본 논문에서는 시간의 경과에 따른 사용자의 액션 패턴을 기반으로 사용자의 관심에 따른 가중치를 적용하여 적응 학습(Adaptive learning)을 통해 분석해 스팸 메일 여부를 가리는 방법을 다룰 것이다. 사용자의 관심은 지속적으로 변화하므로, 이를 적절하게 학습하여 필터링에 반영해야 한다. 이것은 Concept Drift 라는 개념으로, 사용자는 스팸 메일로 간주했던 메일에 대해 얼마간의 시간이 흐른 뒤 관심을 보일 수도 있고, 반대로 어떤 주제와 관련한 메일에 대한 관심이 차츰 줄어드는 경우가 있을 수 있다. 따라서 본 논문에서는 액션들간의 관계를 분석하여 하나의 메일에 대해 사용자가 어떤 액션을 취한 뒤 다른 액션을 취하는 지, 나아가 그 액션 사이의 시간까지 고려하여, 취해진 액션들이 얼마만큼의 의미를 가지며 그것이 스팸 메일

필터링에 얼마나 영향을 미치게 될 지를 분석하여 가중치를 주고자 한다.

본 논문은 사용자의 액션에 기반한 스팸 메일 필터링 기술[3]을 기반으로 하고 있으며 기존 논문에서의 액션 카테고리 분류에 대한 기존의 부정확성을 줄이고 각 액션들에 대한 사용자의 의도를 더욱 명확히 하여 학습속도를 높이는 동시에, 학습된 메일에 추가되는 액션에 대한 의미를 지속적으로 반영함으로써 학습의 성능 향상을 목표로 한다.

#### 2. 관련연구

##### 2.1 다양한 스팸 메일 필터링 방법

메시지 규칙 기반의 필터링 기술이란 메일의 내용으로부터 특징들을 추출하여 그것들이 스팸 메일의 특성에 부합되는 지를 판단하여 스팸 여부를 결정하는 방법을 말한다[4]. 이 기술은 구현과 사용이 비교적 간편하여 많은 필터링 방법에 사용되었으나 스팸 메일이 변화함에 따라 지속적으로 메시지 규칙을 갱신시켜야 하는 한계가 있다.

그 다음으로 규칙 기반 필터링의 단점을 극복하기 위해 나온 것이 확률 및 통계를 이용한 방법이다. 가장 대표적인 방법인 베이지안 분류자는 메일 내용의 term 들을 분석해 스팸 메일 여부를 확률적으로 판단하는 방법으로, 환경의 변화에 대한 적응력이 뛰어나며 매우 좋은 성능을 보인다. 여기에서 조금 더

발전된 형태인 베이지안 네트워크, 나이브 베이지안 분류자(Naïve Bayesian Classifier), 가중치를 부여한 베이지안 분류자(Weighted Bayesian classifier(WBC)) 등과 같은 베이지안 분류 기반 시스템이 많이 이용되고 있다[5].

Bayesian 과 더불어 나온 SVM, LSI(Latent semantic indexing)과 같은 기계 학습을 이용한 방법들은 한 번 읽은 스팸 메일은 다시는 받아들이지 않는 방법으로, 계산이 복잡하고 구현, 관리가 어려워 실생활 적용이 비교적 어려운 실정이다[6].

### 2.2 Concept Drift 와 적응 학습

스팸 메일은 그 특성상 도메인이 급격하게 변하므로 그 변화를 잘 빠르게 인지하여 대처할 필요가 있다. 이를 Concept Drift(Interest Drift)라 하며, 사용자의 관심과 스팸 메일의 특징들의 변화를 추적하며 학습시키는 것을 말한다[7]. 이러한 필요성들에 의해 생겨난 개념이 적응 학습으로, 시간에 따른 사용자의 변화된 관심의 정도를 실시간으로 학습에 반영하는 것을 말한다[8]. 이 논문에서는 사용자가 직접 선호도를 체크하지 않아도 사용자의 액션만으로 스팸 메일 여부를 결정해 적응 학습하는 방법을 사용할 것이다.

### 3. 사용자 액션과 시간에 따른 메일 필터링

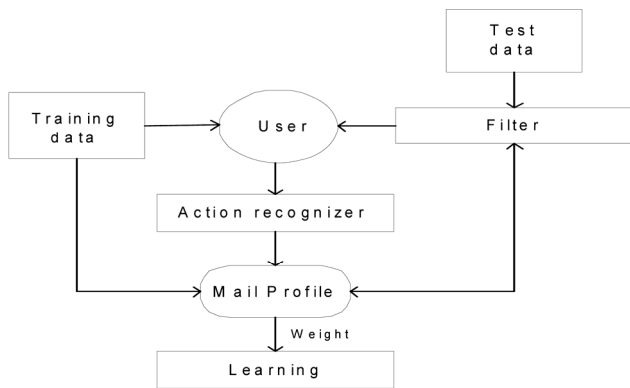


그림 1 시스템 구조도

그림 1 은 본 논문에서 제안하는 시스템의 구조를 보여준다. 메일에 대한 사용자의 액션 패턴, 그 액션이 이루어진 시간 등을 바탕으로 메일의 각 term 들에 대한 가중치를 구해 학습이 이루어진다. 또한 액션이 추가될 때마다 가중치가 갱신됨에 따라 재분류(re-classification)를 통해 학습할 수 있다[9]. 테스트 단계에서는 메일 내용의 각 term 들이 가중치가 부여된 베이지안 분류자를 통해 필터링되고, 메일에 대해

사용자의 액션이 취해지면 그 액션에 따라 학습시키는 과정을 반복하여 학습이 지속적으로 갱신될 수 있다.

### 3.1 WBC 에 의한 사용자 액션 기반의 적응 학습을 이용한 스팸 메일 필터링

기존의 많은 스팸 메일 필터링 시스템들은 사용자로부터 직접 스팸 메일에 대한 피드백이나 rating 결과를 받아 처리해야 한다는 단점이 있어왔다. 이러한 불편을 해소하기 위해 메일에 대해 사용자가 행하는 액션을 자동으로 피드백 받아 필터링시 가중치 부여에 활용하는 방법이 소개되었다. 이러한 방법은 추가적으로 메일에 대한 선호도를 체크할 필요가 없으므로 사용상의 번거로움과 시간낭비를 막을 수 있다[10][11]. 특히 메일에 대한 기본 액션을 open, delete, save, reply, block, nothing 의 6 가지 액션으로 분류, 이 액션들의 조합과 액션 시간에 따라 가중치를 주어 학습한 후 WBC 방법을 통해 스팸 메일 여부를 필터링 하게 된다. 이렇게 개인화된 학습과 적응 학습을 포함함으로써 기존의 베이지안 분류자[12] 및 나이브 베이지안 분류자에 비해 향상된 성능을 볼 수 있다. 본 논문에서는 사용자의 액션을 세밀하게 분류함에 따라 가중치를 차별화하며 시간에 따라 그 가중치의 정도를 달리 함으로써 향상된 학습 성능을 얻을 수 있도록 한다.

### 3.2 사용자 액션에 따른 구분

각 액션은 positive 액션과 negative 액션으로 나뉘는데, open, save, reply 가 positive 에 속하고 delete 와 block 이 negative 에 해당한다. 각 액션들은 디폴트 가중치를 가지고 있으며 이를 기반으로 최종 가중치를 계산하게 된다. 디폴트 가중치는 사용자의 입장에서 어떤 액션을 취했을 때의 의도와 그 의미의 중요도를 고려하여 heuristic 방법에 의해 결정하였다.

표 1 액션에 대한 디폴트 가중치

Action	Default Weight	Category
Save	0.6	Positive
Reply	0.4	
Open	0.2	
Delete	0.4	Negative
Block	0.8	

### 3.3 가중치 부여를 위한 액션 조합의 분류

사용자 액션 인식기는 사용자의 액션에 따라 메일 ID, 액션, 시작 시간, 종료 시간 등을 메일 프로파일에

저장한다. 이 프로파일에 축적된 액션들을 바탕으로 가중치를 부여, 업데이트 해 메일을 학습시킨다. 기본적으로 가중치를 주는 방법은 덧셈 연산을 바탕으로 하며 프로파일에 쌓인 액션의 조합(현재 액션에 대한 가중치를 구하기 위해 바로 앞 액션과의 관계를 살핌)들을 세 가지 경우로 분류하여 각각에 대해 계산식을 달리하여 가중치를 계산하였다. 각 액션에 대한 가중치 계산은 액션 자신의 디폴트 가중치(표 1)를 기반으로 한다.

앞서 언급했던 것과 같이 사용자의 액션에 대한 의도를 파악해본 바, 모든 액션의 조합들은 그 패턴과 시간 그리고 액션이 일어난 횟수 등에 따라 제각기 다른 의미를 지니고 있다고 볼 수 있다[10][11][13][14]. 예를 들어 메일을 'open' 후 'delete'했을 때, 그것을 읽은 즉시 삭제하는 경우 혹은 읽은 후 한참 후에 삭제하는 경우 두 가지 모두 메일이 스팸 메일일 가능성과 스팸 메일이 아닐 가능성을 내포하므로 필터링의 정확성이 저하되는 원인이 될 수 있다. 따라서 메일을 읽고 삭제하는 과정 사이의 시간에 따라 가중치를 달리 주는 방법을 고안 했다. 먼저, 현재 액션에 대한 가중치를 구하기 위해 바로 전 단계에서 일어난 액션과의 관계를 다음과 같이 세 가지 경우로 분류한다.  $a_i$  는  $i$  번째 액션,  $a_{i-1}$  은  $a_i$  의 바로 전 단계에서 일어난 액션,  $P$  는 positive 액션의 집합이라 할 때 다음과 같이 나타낼 수 있다.

case 1.  $a_{i-1} \in P$  and  $a_i \in P$

case 2.  $a_{i-1} \in P$  and  $a_i = 'block'$  (1)

case 3.  $a_{i-1} \in P$  and  $a_i = 'delete'$

case1 의 경우 그 메일이 스팸 메일이 아닐 가능성은 positive 액션을 한 가지씩 취했을 경우보다 훨씬 높아진다고 볼 수 있다. 따라서 이 메일에 대한 논스팸 카테고리에 대한 가중치를  $w_N$ , 스팸 카테고리에 대한

가중치를  $w_S$  라 할 때 다음과 같은 결과를 보이며 결과적으로  $|w_N - w_S|$ 가 논스팸 카테고리로 학습된다.

$$w_N \gg w_S$$

또한  $a_i$  에 대한 가중치를  $w_i$ ,  $a_i$  의 디폴트 가중치를  $w_{d_i}$  라 할 때  $w_{d_i} < w_i$  가 되고, 시간이 지남에 따라

$w_i$  를 점차 작아지도록 함으로써 학습 성능을 향상시킬 수 있을 것이다. 여기에서 일부 액션의 조합에 대한 예외가 발생할 수 있는데, 이는 메일의 길이와 두 액션 사이의 시간차가 비례함에 따른 것이다. 'open'후 메일을 읽고 나서 'save'나 'reply'를 취했을 경우 본 논문에서는 메일의 길이와 상관없이 액션이 취해지는 시간만을 고려하므로 메일을 읽는 시간에 따른 변수를 전혀 고려하지 않는다. 이 부분은 사용자의 관심도에 영향을 미치기는 하나, 메일의 길이가 메일을 읽는 시간에 큰 영향을 미치지 않음에 따라 본 논문에서는 고려대상에 넣지 않았다[13].

case 2 의 경우 그 메일이 스팸 메일일 가능성은 스팸 메일이 아닐 가능성에 비해 상대적으로 높다고 볼 수 있다. 따라서 이 경우에 메일은 스팸 카테고리에 대해 학습되는 결과를 보인다.

$$w_N < w_S$$

가중치의 변화는 'block'의 디폴트 가중치인 0.8 에서 시작, 시간이 지날수록 'delete'의 디폴트 가중치인 0.4 까지 변화하도록 했는데 아무리 시간이 많이 지나도 'block'에 대한 가중치를 그 하위 단계인 'delete'에 대한 가중치보다 작아지지 않도록 하기 위함이다.

case 3 에서는 메일이 스팸 메일일 가능성과 스팸 메일이 아닐 가능성을 모두 배제할 수 없으므로 종전의 가중치 부여 방식과는 다른 방식을 취한다. 액션 간의 시간 차를  $t$  라 할 때, 부여되는 가중치는 다음과 같다.

$$w_S > w_N \rightarrow w_S < w_N$$

$$\lim_{t \rightarrow 0} (w_N - w_S) < 0, \lim_{t \rightarrow \infty} (w_N - w_S) > 0$$

따라서 case 3 에 대해서는 시간이 어느 정도 지났느냐에 따라 메일이 스팸 메일인 지 논스팸 메일인지가 결정된다.

### 3.4 가중치 부여 함수

식 (1)에 대한 가중치 함수를 구현하기 위해 시간을 적용시킬 수 있다.  $a_{i-1}$  이 시작되는 시점부터  $a_i$  가 시작되는 시점까지의 시간  $t$ 를 초단위로 측정해 이것을 가중치 함수  $f(t)$  에 적용한다.  $f(t)$  로는 logarithm

함수를 사용하였다. 즉  $f(t) = \log(t+1)$  의 형식을 취하는데, 이 로그 함수를 사용한 이유는 시간이 지남에 따른 액션의 중요성 혹은 의미의 퇴색이라 볼 수 있겠다. 이전 액션이 일어난 후 다음의 액션을 바로

취했을 경우 그것이 스팸 메일이 아닐 가능성은 한참 후 액션을 취했을 경우의 가능성보다 매우 높으므로 가중치는 시간에 반비례하는 양상을 띠게 되며, 시간에 따른 가중치의 감소폭 역시 점차 작아지게 된다. 따라서 이러한 로그의 성질을 반영하여 시간에 따라 가중치를 계산하도록 했다. 이 로그 함수를 이용한 식 (1)의 가중치 결정 함수는 다음과 같다.

$$\begin{aligned} \text{case 1. } w_i &= w_{d_i} + 0.2 - \frac{\log(t+1)}{\log(\max t)} * w_{d_i} \\ \text{case 2. } w_i &= w_{d_i} - \frac{\log(t+1)}{\log(\max t)} * w_{d_i} \\ \text{case 3. } w_i &= w_{d_i} - 0.1 - \frac{\log(t+1)}{\log(\max t)} * w_{d_i} \end{aligned} \quad (2)$$

여기에서  $\max t$  는  $t$  의 임계값(threshold)으로,  $t$  가 무한정 길어지지 않도록 하기 위함이다.

이렇게 계산된 각 액션에 대한 가중치들을 모두 반영하여 하나의 메일에 대한 두 개의 가중치  $w_N$  과  $w_S$  를 다음과 같이 계산하여 스팸과 논스팸의 두 카테고리 모두를 학습하게 된다. 즉, 두 가중치 중 큰 쪽으로  $|w_N - w_S|$  만큼 학습시키는 결과를 보인다.

$a_{m-1}$  까지 학습되었고  $a_n$  이 학습시킬 마지막 액션,  $U$  는 모든 액션의 집합이라 할 때 다음의 식을 통해 두 카테고리의 가중치를 계산할 수 있다.

$$\begin{aligned} w_N &= 1 + \sum_{i=m}^n w_i, \quad a_m \in P, \quad a_n \in P \\ w_S &= 1 + \sum_{i=m}^n w_i, \quad a_m \in U, \quad a_n \in U \end{aligned} \quad (3)$$

### 3.4 가중치가 부여된 베이지안 분류

이렇게 구해진 가중치들은 학습을 통해 각 term 들에 부여되고, 이 term 들에 대한 가중치를 이용해 새로운 메일에 대해 가중치가 부여된 베이지안 분류자를 적용하여 메일이 스팸 메일인 지, 논스팸 메일인 지를 필터링하도록 했다. 기존의 나이브 베이지안 분류자가 각 term 들에 대해 모두 같은 가중치를 갖는 반면, 가중치가 부여된 베이지안 분류자는 각 term 들에 대해

모두 가중치를 달리 줌으로써 상대적으로 높은 속도의 학습 성능을 기대할 수 있다. 가중치가 부여된 베이지안 분류자의 식은 다음과 같다.

$$c = \arg \max_{c_i \in C} P(c_i) \prod_{k=1}^n \left( \frac{w_{t_k}}{\max_j w_{t_j}} \right) P(t_k | c_i) \quad (4)$$

새로운 메일에 대한 각 term 들에 대한 가중치를 스팸과 논스팸 카테고리에 대해 각각 구하고 term 의 가중치들의 합이 더 큰 카테고리로 메일을 분류하도록 한다. 가중치들의 정규화를 위해 각 term 들의 가중치를 최대의 가중치를 갖는 term 의 가중치로 나누어 식에 반영하도록 했다.

### 4. 실험 및 결과

본 논문의 실험은 펜티엄 4, 2.79GHz, 1.49GB RAM 의 Microsoft Windows XP 환경에서 진행되었으며 Java 와 My-SQL 을 기반으로 구현되었다.

#### 4.1 데이터 집합

실험을 위해 사용한 데이터는 LingSpam[15] 메일 집합으로, 메일을 스팸 메일과 논스팸 메일의 두 가지로 분류해 사용하였다. 스팸 메일은 중복될 수 있으며, 한 번 테스트된 메일은 다시 학습을 위해 사용될 수 있다는 가정 하에 데이터를 분류하였다.

표 2 데이터 집합

	Training Data	Test Data
Spam mail	1328	1301
Nonspam mail	147	152
Total	1475	1453

#### 4.2 성능 측정

표 2 와 같이 두 개의 카테고리로 나뉘어진 실제 영문 메일을 임의로 섞어 스팸/논스팸 메일의 비율을 일정하게 하여 하루에 50 여 개의 메일을 받는다는 가정 하에 20 일 동안의 사용자의 관심도를 반영한 필터링의 정확도를 테스트하기로 했다.

학습은 하루에 한 번씩 이루어지도록 했으며 학습된 결과를 바탕으로 다음날 받은 메일에 대해 필터링하고,

이렇게 필터링된 메일에 대해 다시 학습하는 과정을 반복하여 학습의 정확도와 정확도의 증가 속도, 그리고 마지막 날까지의 필터링 결과를 관찰하기로 한다.

의 측정을 위한 정확도는 일반적으로 널리 사용되는 precision, recall, F1-measure[16]를 기반으로 한다.

$$Precision = \frac{\text{스팸으로 분류된 스팸 메일}}{\text{스팸으로 분류된 전체 메일}}$$

$$Recall = \frac{\text{스팸으로 분류된 스팸 메일}}{\text{전체 스팸 메일}} \quad (5)$$

$$F1\text{-measure} = \frac{2 * Precision * Recall}{Precision + Recall}$$

### 4.3. 실험 결과

날짜가 지남에 따른 테스트 데이터 필터링의 정확도는 그림 2와 같다.

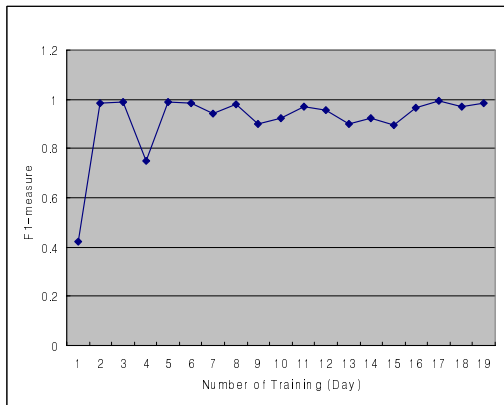


그림 2 본 논문이 제안하는 시스템의 시간에 따른 필터링 정확도

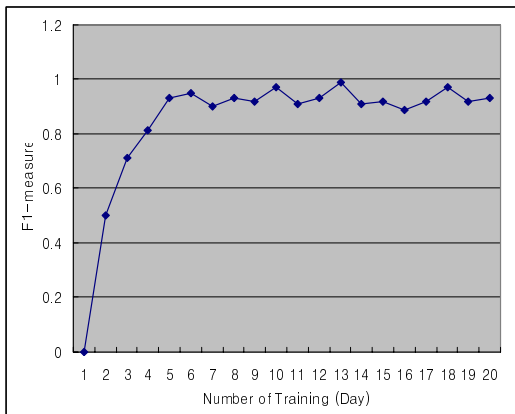


그림 3 가중치가 부여된 베이지안 분류자의

### 필터링 정확도

필터링 결과, 그림 2에서 나타난 것과 같이 본 논문이 제안한 시스템은 초기에는 학습시 필터링의 정확도가 매우 높고, 학습이 반복됨에 따라 그 변화폭이 상대적으로 크지만 학습이 진행됨에 따른 정확도의 증가 속도는 평균 95% 이상으로 비교적 높다고 보여진다. 이를 그림 3에서 나타난 기존의 가중치가 부여된 베이지안 분류자[2]와 비교한 결과 학습 성능의 향상 속도가 상대적으로 높은 것을 볼 수 있다. 수치적으로 우리의 시스템이 기존의 방법들에 비해 40%이상 학습 성능이 빠른 것으로 나타났다.

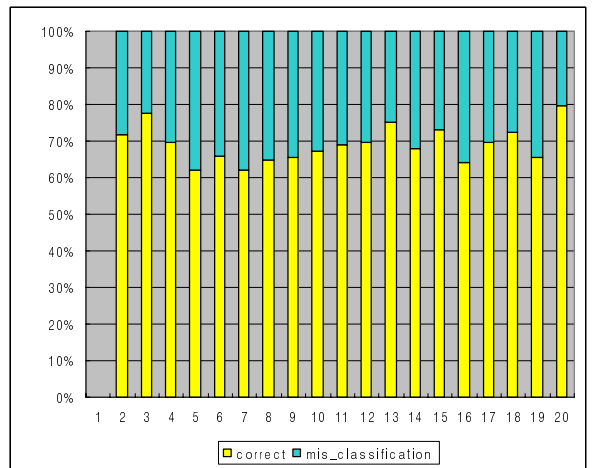


그림 4 스팸/논스팸 메일에 대한 카테고리별 확률

그림 4에서는 필터링 정확도를 조금 다르게 실험해 보았다. 각 카테고리에 대해 메일이 할당될 확률을 측정하는 것으로, 스팸/논스팸 메일이 얼마나 확실하게 스팸/논스팸 카테고리로 분류되었는지 그 비율을 계산하였다. 그림에서 보듯 초기 필터링 단계에서 60% 이상의 정확도로 메일이 분류되었다.

실험에서는 논스팸 메일에 대한 스팸 카테고리로의 할당 비율이 비교적 높은 것으로 나타났는데, 이는 논스팸 메일에 대한 term 이 스팸 메일의 term 에 비해 상대적으로 적었던 것에 원인이 있는 것으로 판단된다. 즉, 실험에 쓰여진 논스팸 메일의 양이 극히 적음에도 불구하고 그 내용의 분야가 매우 넓어 논스팸 카테고리 내에서 겹치는 term 들이 많지 않은 까닭에 논스팸 메일에 대한 분류의 정확도가 스팸 메일에 대한 정확도에 비해 상대적으로 낮게 나온 것으로 판명되었다.

### 5. 결 론

본 논문에서 우리는 메일에 대한 사용자의 액션들과 그 액션 사이의 시간에 따라 가중치를 차별화하여 학습한 후, 가중치가 부여된 베이지안 분류자를 이용해 스팸 메일을 필터링하는 시스템을 제안하였다. 이 시스템은 학습 성능 측면에서 기존의 베이지안 분류자와 가중치가 부여된 베이지안 분류자들에 비해 더 나은 결과를 보였고, 특히 짧은 시간 안에 학습 속도를 크게 향상시켰다. 이는 액션이 추가됨에 따라 가중치를 갱신하여 학습시키는 적응 학습을 적용한 결과 이루어진 것이라 볼 수 있다. 또한 본 논문은 개인화 측면 역시 다루고 있는데, 사용자가 본인의 관심 정도에 따라 메일에 대한 가중치 부여를 정교화하고, 각 개인의 관심 흐름을 지속적으로 반영함에 따른 것이다. 향후 우리가 연구해야 할 것으로, 액션 사이의 관계를 보다 세밀하게 분석함으로써 학습 속도를 더욱 높이는 동시에, 메일의 길이와 시간과의 관계 등 사용자의 액션의 의미를 보다 정확히 파악하여 필터링에 반영하는 방안 등을 모색할 것이다. 또한 사용자의 액션과 시간에 따른 가중치 부여를 Collaborative Filtering 등에 반영하여 학습 성능을 향상시킬 수 있는 방법도 연구 과제로 고려할 수 있을 것이다.

## 6. 참고 문헌

- [1] ZDNet Korea, <http://zdnet.co.kr>
- [2] 디지털 타임스, [http://www.dt.co.kr/contents.html?article\\_no=2006122802011457730008](http://www.dt.co.kr/contents.html?article_no=2006122802011457730008)
- [3] H. J. Kim, J. Shrestha, H. N. Kim, G. S. Jo, "User Action Based Adaptive Learning with Weighted Bayesian Classification for Filtering Spam Mail", AI 2006, LNAI 4304, pp. 790-798, 2006
- [4] Cohen, W. W., "Learning Rules that Classify E-Mail", AAI Spring Symposium, 1996
- [5] Androutsopoulos, I., Koutsias, J., Chandrinou, K. V., Paliouras, G., Spyropoulos, C. D., "An Evaluation of Naive Bayesian Anti-Spam Filtering", ECML 2000, pp. 9-17, 2000
- [6] Thomas, G., Peter, A. F., "Weighted Bayesian Classification based on Support Vector Machine", International Conference on Machine Learning, pp. 207-209, 2001
- [7] Koychev, I., Schwab, I., "Adaption to Drifting User's Interests", ECML200/MLnet Workshop, 2000
- [8] Dwi H. Widiantoro, Thomas R. Ioerger, John Yen, "An Adaptive Algorithm for Learning Changes in User Interests", International conference on Information and knowledge management, pp. 405-412, 1999
- [9] M. Albanese, A. Picariello, C. Sansone, L. Sansone, "Web Personalization Based on Static Information and Dynamic User Behavior", ACM international workshop, pp. 80-87, 2004
- [10] M. Morita, Y. Shinoda, "Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval", ACM SIGIR conference, pp. 272-281, 1994
- [11] Y. W. Seo, B. T. Zhang, "Learning User's Preferences by Analyzing Web-Browsing Behaviors", International conference on Autonomous agents, pp. 381-387, 2000
- [12] Sahami, M., Dumais, S., Heckerman, D., Horvitz, E., "A Bayesian Approach to Filtering Junk E-Mail", AAI, pp. 55-62, 1998
- [13] Burklen, S., Marron, P.J., Fritsch, S., Rothermel, K., "User Centric Walk: An Integrated Approach for Modeling the Browsing Behavior of Users on the Web", Annual Simulation Symposium, 2005
- [14] E. Agichtein, Z. Zheng, "Identifying 'Best Bet' Web Search Results by Mining Past User Behavior", ACM SIGKDD international conference, pp. 902-908, 2006
- [15] Androutsopoulos, I., Koutsias, J., Chandrinou, K. V., Paliouras, G., Spyropoulos, C. D., "An Evaluation of Naive Bayesian Anti-Spam Filtering", ECML2000, pp. 9-17, 2000
- [16] Y. Yang, X. Liu, "A Re-examination of Text Categorization Methods", ACM SIGIR'99, 1999