

# 영상처리기술을 이용한 SVM 기반 스팸메시지 필터링 기법

강효성<sup>o</sup> 홍충선\*  
 경희대학교 컴퓨터공학과  
 { kanghs, cshong } @khu.ac.kr

## Spam Message Filtering Based on SVM Using Image Processing Technology

HyoSung Kang<sup>o</sup> ChoongSeon Hong\*  
 Department of Computer Engineering, Kyung Hee University

### 요약

인터넷의 급속한 성장과 더불어 전자메일 사용자의 수가 증가하면서 인터넷 상에 개인 전자메일주소의 노출로 인한 스팸메시지의 양도 덩달아 증가하게 되었다. 이것은 메일 사용자에게 악용되어 바이러스 전파 및 개인정보 노출의 원인이 되고 있다. 이에 많은 스팸메시지 차단 기법들이 제시되었지만 이는 텍스트를 포함한 메시지에 특화되어 있어 문자가 포함되어 하나의 이미지로 전송되는 이미지스팸에는 차단이 어려운 단점이 존재한다. 본 논문에서는 이런 이미지스팸(Image Spam)에 대해서 LBP(Local Binary Pattern)를 추출하여 SVM(Support Vector Machine)의 기계학습을 통해 스팸메시지를 필터링하는 기법을 제안한다.

### 1. 서론

경제협력개발기구(OECD)의 브로드밴드 포털자료에 따르면 인구대비 유선 초고속 인터넷 보급률에서 한국은 2000년에서 2005년까지 1위를 차지했다. 2006년부터 2위로 조금씩 밀려 2011년 7위로 떨어지기는 했지만 여전히 OECD 평균 이상의 순위를 두고 있다[1].

이런 인터넷의 급속한 성장과 더불어 전자메일의 사용자 수도 점점 늘고 있다. 하지만 송수신에 편리하고 비용이 들지 않는 전자메일의 장점을 이용해 많은 개인, 기업들은 상업적 광고를 포함한 스팸메시지를 무분별하게 발송하고 있으며 그 양도 매년 꾸준히 증가하고 있다[2].

작년 하반기 국민 1인당 하루에 수신하는 이메일 스팸 양을 측정한 결과 평균 0.66통으로 나타났다. 이는 작년 상반기 0.51통과 비교해 늘어난 수치이다. 포털 다음이 0.66통으로 가장 많았고, 네이버와 네이트가 각각 0.42통으로 나타났다[3].

이런 증가 수치는 여러 스팸필터링시스템을 우회하기 위해 스팸발신자가 여러 특수문자, 영문+한글 혼합문자, 세로쓰기 등 문자를 변칙적으로 적용하여 메시지를 보내기 때문이다.

최근에는 내용을 텍스트가 아닌 이미지 형태로 바꾸어 텍스트는 거의 없이 이미지나 HTML만을 포함하는 이미지스팸(Image Spam)이 등장하였다. 그동안 대부분의 스팸메시지는 텍스트로 구성이 되어 스팸메시지 차단을 위해 텍스트 필터링 기술이 도입되어 왔다. 하지만 이미지스팸은 텍스트를 전혀 포함하지 않아 기존 스팸차단 술

루션기술 마저도 우회하기 때문에 스팸발신자들은 이미지스팸의 이런 특징을 이용해 대량으로 스팸메시지를 발송하고 있다[4].

이미지스팸의 양은 매년 꾸준히 증가하여 2010년에는 개인이 한달동안 수신한 메일중 스팸메시지의 비율이 전체 스팸메시지의 52%를 차지하기도 하였다[5].

본 논문에서는 이런 이미지스팸에 대한 효율적인 차단을 위해 LBP(Local Binary Pattern)를 통한 이미지 처리 후 SVM(Support Vector Machine) 학습으로 스팸메시지를 필터링하는 기법을 제시하고자 한다.

### 2. 관련연구

#### 2.1 텍스트 스팸 필터링



그림 1. 스팸 분류기 학습 컴포넌트 [6]

본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 육성지원 사업의 연구결과로 수행되었음 (NIPA-2014-(H0301-14-1003)).

\*Dr. CS Hong is the corresponding author.

[6]에서 제안한 해당 알고리즘은 텍스트기반의 스팸필터링 전처리 단계에 가진다. 텍스트에서 스팸으로 등록

된 단어들과 비교하여 필터링하기 위해 전처리 단계에서 텍스트메시지의 특수문자를 제거, 문장 사이 자동 띄어쓰기 추가, 수사어절 표준화, 실제 사진에 없는 은어와 같은 불용어 제거 단계를 우선적으로 거친다. 그리고 본 논문과 마찬가지로 필터링 단어의 패턴으로 학습된 SVM을 통해 메시지의 스팸여부가 판단된다.

하지만 해당패턴은 문자를 변칙적으로 적용된 기존 텍스트 스팸메시지의 단점을 어느 정도 보완하였지만 텍스트를 포함하지 않은 이미지스팸의 분류에는 적합하지 않다.

### 3. 제안사항

#### 3.1 제안된 알고리즘

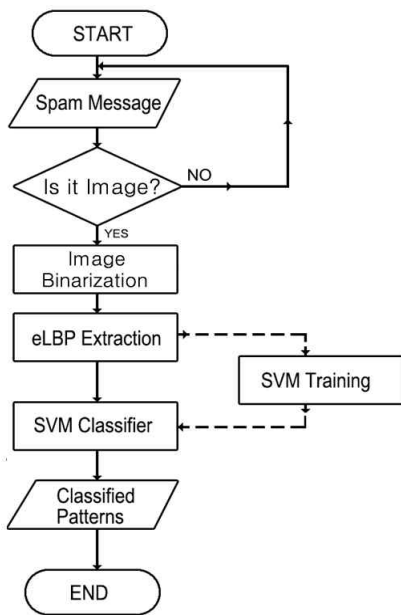


그림 2. 이미지처리 이용 SVM기반 스팸필터링 흐름도

그림 2는 본 논문에서 제안하는 이미지처리를 이용한 SVM기반 스팸필터링 흐름도이다. 스팸메시지가 들어오면 시스템은 해당 메시지가 이미지스팸인지 여부를 우선 판단한다. 이미지스팸으로 판단이 되면 이미지스팸에 포함된 이미지파일을 이진화하여 그레이 스케일로 바꾼 후 LBP(Local Binary Pattern) 정보를 추출한다. LBP는 이미지의 질감정보를 2진화 데이터로 표현하는 기술로 여기서는 이미지 내에서 적은 차원을 가지면서 많은 질감정보를 표현하는 eLBP(edge-LBP)를 사용하였다[7]. eLBP를 통해 추출된 스팸, 비스팸패턴에서 무작위로 약 20%를 각각 무작위로 추출하여 학습 데이터로 선정한다. 선정된 학습데이터는 SVM(Support Vector Machine) 학습에 활용된다. 학습이 완료된 SVM은 전체 이미지스팸의 패턴을 입력받아 학습된 경계를 기준으로 패턴을 분류한다. 본 논문에서는 최종 패턴의 분류율을 가지고 구현된 스팸필터링의 성능을 분석한다.

#### 3.2 이미지의 이진화

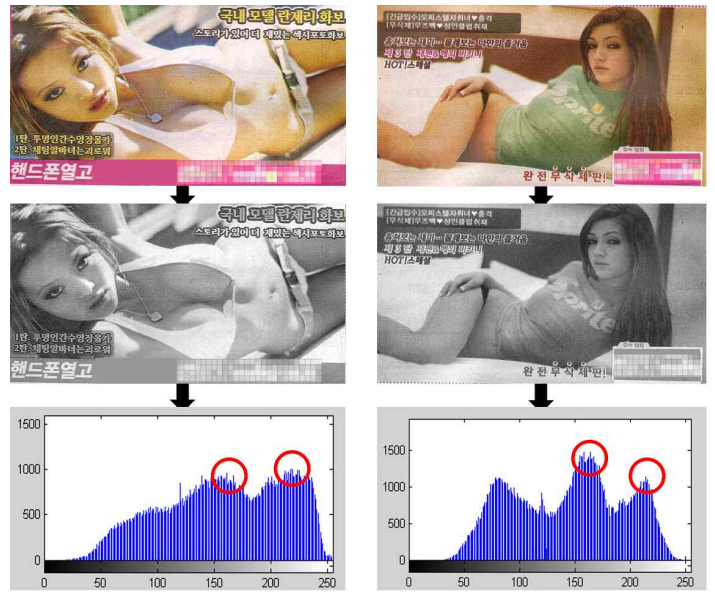
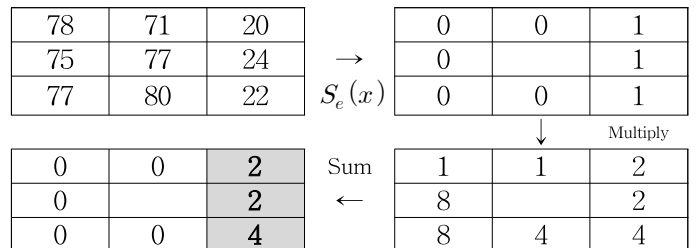


그림 3. 이미지스팸[8]의 이진화와 히스토그램

그림 3은 이미지스팸을 이진화하여 히스토그램으로 나타낸 것이다. 성인광고 이미지스팸에 대한 히스토그램에서 평균적으로 그레이레벨이 155~170, 220~230 사이에서 높은 해당 그레이레벨의 빈도 수를 가진다. 이런 특징들은 SVM의 학습에서 정확한 분류 임계값을 정하는 기준이 된다. 또한 원본이미지를 그대로 SVM에 학습시킬 경우 차원의 수가 너무 많아 학습시간이 굉장히 오래 걸리고 분류결과도 정확하지 않은 단점들이 존재하기 때문에 이미지의 이진화 과정이 우선적으로 필요하다.

#### 3.3 edge LBP 추출



$$eLBP_{v-h} = 2$$

$$eLBP_{diag} = 2 + 4 = 6$$

$$\therefore eLBP_{total} = eLBP_{v-h} + eLBP_{diag} = 2 + 6 = 8$$

그림 4. 이미지의 eLBP 변환[7]

LBP는 이미지 정보를 2진화 패턴으로 추출하는 방법으로 이미지 질감에 대한 특징표현이 매우 우수하다. 그중 eLBP는 기존 LBP로는 표현하기 힘든 질감의 경계정보를 표현하기 위해 고안된 방식이다. 그림 4는 이미지에 대한 eLBP 변환 방법을 나타낸 그림이다[7]. 3×3 픽

셀블록 단위로 중심 화소의 화소값이 중심 화소 및 이웃 화소의 화소값의 차보다 크면 1 작으면 0의 값으로 표현한다. 그렇게 나온 값에 1, 1, 2, 2, 4, 4, 8, 8의 가중치를 각각 곱하여 수직, 수평 성분별로 따로 더한다. 그림 2에서  $eLBP_{v-h}$ 는 화소의 수직, 수평 정보를 표현한 값이고  $eLBP_{diag}$ 는 대각선 정보를 표현한 값이다. 최종적으로 이미지의  $eLBP_{total}$  값은  $eLBP_{v-h}$ 와  $eLBP_{diag}$ 를 합쳐 더하여 표현한다.

제안된 알고리즘의 성능 분석을 위하여 이미지스팸의 이미지에서 스팸패턴은 614개, 비스팸패턴은 1125개를 포함한 총 1739개의 패턴을 사용하였으며 이중 전체 패턴의 약 23%에 해당하는 스팸패턴 200개, 비스팸패턴 200개를 학습데이터로 선정해 SVM의 학습데이터로 선정해 분류 알고리즘의 학습을 진행한다.

### 3.4 SVM 학습

일반적으로 분류 알고리즘에는 결정트리, 신경망, SVM 등이 존재하며 그중에서도 SVM은 가장 뛰어난 성능과 분류율을 가진다[9]. 결정트리는 신경망 모델이 분류율을 최대로 하는 것을 목적으로 하는 것에 반해 SVM은 오류율을 최소로 하여 분류율을 높이고 나아가 지원벡터를 정해 분류를 구분하는 기준을 최대화 하는 것까지 고려한다. 이런 장점 때문에 제안하는 스팸필터링에서는 SVM 분류기를 선택하여 패턴학습을 진행한다.

SVM은 커널 함수의 활용에 따라서 그 성능의 확연한 차이가 존재한다. 커널함수를 통해 단순 이차원 평면상의 데이터분류 보다 고차원 형태의 분류도 가능하게 해주어 주로 많은 차원의 수를 가지는 실시간 이미지 정보의 패턴분류에 많이 이용된다[10]. 본 논문에서는 고차원을 가진 LBP를 학습하기 위해 비선형 SVM을 사용하였다. 이중 가우시안 커널(Gaussian Kernel)과 다형식 커널함수(Polynomial Kernel)를 통해 SVM 학습을 진행하여 분류율을 비교하였다.

### 4. 성능분석

성능분석은 SVM의 커널 함수마다 정확한 SVM의 분류 결과를 얻기 위해 각각 1000, 1500, 2000, 2500 에폭(Epoche)단위로 진행했으며 각 에폭 당 분류율의 평균을 구하여 스팸메일 필터링의 전체 분류율을 계산하였다.

Epoche	시간 (s)	총 패턴수	실패 패턴수	성공 패턴수	성공률 (%)
1000	156	1739	129	1610	92.54
1500	231		126	1613	92.75
2000	311		110	1629	93.67
2500	402		119	1620	93.1

표 1. Gaussian Kernel을 이용한 eLBP 분류결과

Epoche	시간 (s)	총 패턴수	실패 패턴수	성공 패턴수	성공률 (%)
1000	163	1739	64	1675	96.31
1500	250		61	1678	96.49
2000	334		62	1677	96.43
2500	432		65	1674	96.26

표 2. Polinomial Kernel을 이용한 eLBP 분류결과

표 1은 가우시안 커널을 이용한 SVM 이미지스팸의 분류 결과이며 표 2는 다형식 커널을 이용한 SVM 이미지스팸의 분류결과이다. 전자의 경우에는 평균 93%의 분류율을 보였고, 후자의 경우에는 평균 96%의 분류율을 보였다. 다형식 커널을 통한 이미지스팸의 필터링에서 약 3% 우수한 결과를 보인다.

### 5. 결론

본 논문에서 제안된 필터링 알고리즘은 앞에서 언급한 텍스트 기반 스팸메시지필터링의 단점을 보완한다. 하지만 스팸발신자가 보낸 이미지스팸의 패턴이 복잡하거나 해상도가 낮은 이미지에서는 분류가 어렵다. 또한 이미지의 LBP 변환 계산시간도 오래 걸려 SVM의 학습시간까지 고려한다면 알고리즘의 실적용에는 아직 미흡하다. 성능분석 결과에서도 테스트 결과 평균 90%이상 분류율을 가지지만 완전한 필터링에는 어느 정도 한계가 존재한다. 따라서 향후 연구로 LBP 성능을 개선하고 SVM과는 다른 새로운 신경망 알고리즘을 추가하여 성능을 향상시킨다면 실상용화가 가능한 스팸필터링이 가능할 것이라고 전망한다.

### 참고문헌

[1] OECD, “2011년 4분기 OECD 브로드밴드 통계보고서”, 2011. 9.  
 [2] 한국전산원, “국가정보화백서(National Informatization White Paper)”, pp. 23, 2002.  
 [3] 한국인터넷진흥원, “2013년 하반기 스팸 유통현황”, 2014. 3. 8.  
 [4] 한국정보통신기술협회, 『TTA용어사전』, <http://word.tta.or.kr/terms/terms.jsp?search=%C0%CC%B9%CC%1%F6+%BD%BA%C6%D4&how=like>.  
 [5] 시만텍, “시만텍 월간 스팸보고서”, 2010. 5.  
 [6] 조인희, “휴대폰 SMS를 위한 SVM 기반의 스팸필터링 시스템”, 한국통신학회논문지, Vol. 34, No. 9, pp.908-913, 2009.  
 [7] Marios Anthimopoulos, “A two-stage scheme for text detection in video images”, Image and Vision Computation, vol. 28, pp.1413-1426, 2010.  
 [8] 구글 이미지 검색, <http://joonystory.tistory.com/m/post/62>.  
 [9] 이진선, “대용량 분류에서 SVM과 신경망의 성능 비교”, 한국정보처리학회, 12권, 1호, pp.25-30, 2005.  
 [10] 박준철, “SVM을 이용한 가스 터빈 엔진의 성능저하 예측”, 한국항공우주학회 추계학술발표회논문집(1), pp. 375-378, 2005. 11.