Multimedia Message Service(MMS)상에서 전송되는 스팸이미지 필터링 시스템

박영만*
고려대학교 컴퓨터정보통신대학원*
e-mail:eoxjf@korea.ac.kr*

Multimedia Message Service(MMS) Spam Image Filtering System

Young-Man Park*
Graduate School of Computer Information & Communication, Korea University*

요 약

휴대전화 사용의 대중화로 인하여 개개인의 휴대전화로 수신되는 스팸메시지의 양도 덩달아 증가하게 되었다. 이것은 휴대전화 사용자가 불법광고 노출의 원인이 되고 있다. 이에 많은 스팸메시지 차단 기법이 제시되었지만 이는 텍스트기반의 문자메시지에 특화되어있어 문자가 포함되어있는 이미지스팸에는 차단이 어렵다는 문제점이 존재 한다. 이에 본 논문에서는 휴대전화로 오는 이미지메시지 중 스팸이미지를 검출해 내는 모바일 스팸이미지 필터링 시스템을 제시하고자 한다. 제시하고자 하는 시스템은 스팸이미지를 분석하여 이미지의 패턴을 검사하여 특정 패턴이 포함된 이미지에 대해서 스팸이미지로 분류하여 필터링하게 됨으로써, 실제 휴대전화로 수신되는 스팸이미지를 이용한 실험을 진행하였다. 그 결과 기존 텍스트기반 스팸필터링시스템에서 할 수 없었던 스팸이미지 필터링을 할 수 있음을 확인 하였다.

1. 서론

모바일 스팸이란 이윤을 목적으로 불특정 다수의 휴대전화 사용자에게 보내지는 광고 메시지를 말한다. 현재 우리나라 스팸유통량 조사결과를 살펴보면 국민1인당 1일 평균 0.23통을 수신하는 것으로 조사되었다[1].

문자기반의 스팸메시지의 경우 사용자들은 이통사 무료 부가서비스인 지능형 스팸차단 서비스를 사용하기 때문에 어느정도 스팸문자로부터 보호되고 있다[2]. 하지만 최근 각 이통사별로 스팸차단율이 급감하고 있다[1][2]. 이는 스팸메시지의 유형이 바뀌고 있기 때문이다.

과거에는 문자기반의 스팸메시지가 주를 이루었지만, 스패머들은 지능형 스팸차단 서비스를 회피하기 위한 방법을 찾기 시작하였다. 최근에는 회피하기 위한 방법으로 스패머들이 문자기반의 스팸메시지에서 이미지기반의 스팸이미지메시지를 사용하기 시작하였다 [3].

최근 이러한 모바일 이미지 스팸을 구별하기 위해 다양한 기법이 연구되고 있다. 하지만 모바일메시지의 특성상 전송이 실시간으로 이루어져야 하므로 분석시간이 오래 걸리는 기법들은 사용할 수 없고, 따라서 분석방법이 제한적일 수밖에 없다. 이러한 문제를 극복하기 위하여 이 논문에서는 이미지 자체를 스캔하지 않고 다양한 분야에 사용하고 있는 시그니처 기법에 주목하였다[4].

본 연구는 모바일 스팸이미지 필터링 시 이미지의 특정 시그니처를 추출 후 그 정보를 DB화 하여, 스팸이미지를 분석 및 차단하는 방법을 제시하고자 한다.

2. 관련 연구

2.1 텍스트 기반 스팸 필터링

2000년대 들어서면서 휴대전화 보급이 활성화됨에 따른 스팸문자메시지 문제가 사회적 이슈로 대두되기 시작하였다[5]. 때문에 국내에서 모바일 스팸필터링 방법에 대한 다양한 연구가 진행되었다. [6]에서는 문장의 스타일 정보[7]의 사용과 메시지 오류 교정 과정을 포함하는 시스템을 사용하였다. 이 연구는 문자메시지의 철자오류 및 띄어쓰기 오류수정, 어휘 자질 및 언어학적 분석을 통한 스타일 자질을 추출하는 방법을 함께 사용하였다. 이 연구에서 사용한 방법은 텍스트 기반의 스팸메시지 분류에는 효과적임을 증명하였다. 그러나 최근 증가하고 있는 이미지스팸의 경우 이 연구에서 제시하는 방법이 텍스트 기반의 분석방법이므로 이미지스팸의 경우 분석이 불가능하다는 한계점이 존재한다.

2.2 SVM기반 이미지 스팸 필터링

[8]에서 제시한 해당 알고리즘은 SVM(Support Vector Machine) 기반의 이미지 스팸필터링을 사용하고 있다. 이미지 스팸임을 판단하기 위하여 eLBP(edge-Local Binary Pattern)를 통해 추출된 스팸, 비스팸패턴에서 무작위로 추출하여 학습데이터로 선정 후 SVM기반의 학습을 수행하였다[9]. 학습이 완료된 SVM은 전체 이미지스팸의 패턴을 입력받아 학습된 경계를 기준으로 패턴을 분류한 후 실제 이미지들에 대해서 스팸여부를 판단하게 된다.

하지만 제시한 SVM기반의 이미지 스팸필터링의 경우 LBP 변환 계산기간이 오래 걸리고 SVM의 학습시간까지 고려한다면 실제 상용 서비스에 적용하기에는 적합하지 않다.

3. 제시하는 방법

본 논문에서는 스팸이미지 분석을 위해 사용되는 패턴분석 기반 이미지 스팸필터링 시스템을 제시한다. 패턴 분석 방법은 침입탐지시스템에서 주로 사용하는 기법으로 많은 양의 데이터를 빠르게 분석할 수 있는 장점이 있다. 3.1장에서 본 논문에서 연구하는 시스템의 구조에 대해 기술하고, 3.2장에서 학습과정, 3.3장에서는 패턴분석방법에 대한 설명을 기술한다.

3.1 시스템 구조

본 논문에서 연구하는 이미지스팸 필터링 시스템의 전반적인 구조는 그림 1과 같다.

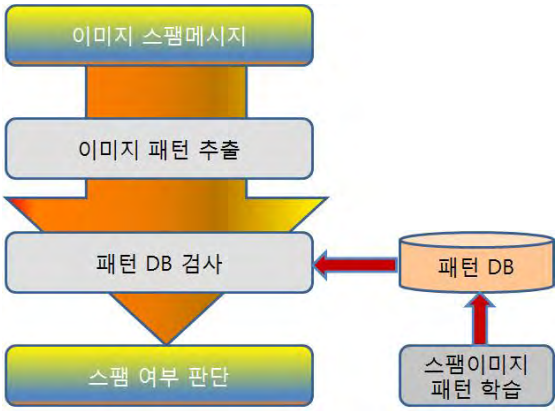

그림 1 시스템구조

최초 입력은 이미지 스팸메시지이고 출력은 이미지 스팸메시지를 분석한 결과이다. 결과는 스팸메시지, 비스팸메시지로 출력된다. 입력된 이미지 스팸메시지는 이미지태그를 모두 제거하고 실제 이미지부분만을 추출한다. 추출된 이미지에 대해서 패턴DB에 저장되어있는 패턴이 존재하는지 여부를 검사하는 과정을 거치게 되고, 스팸가능성을 판단하여 검사 결과가 출력되게 된다. 이때 스팸가능성을 판단하는 방법은 사전에 학습시켜놓은 패턴 DB를 기준으로 수행하게 된다.

그림 1은 본 논문에서 제시하는 이미지처리를 위해 사용되는 시그니처 기반 이미지 스팸필터링 흐름도이다. 스팸메시지가 들어오면 시스템은 해당 메시지가 이미지스팸인지 여부를 판단한다. 판단 방법은 이미지파일만 추출하여 패턴 검사를 수행하게 된다. 특정패턴 하나만 존재하는 경우에 대해서 스팸이미지로 판단할 경우 정상이미지가 스팸이미지로 잘못 판단되는 경우의 수가 많아 질 수 있다. 때문에 3개의 패턴이 일치하는 경우에만 스팸이미지로 판단하게 된다. 이때 사용되는 패턴은 사전에 스팸이미지로부터 미리 학습된 패턴을 이용하여 검사하게 된다.

패턴검사 시 검색속도 향상을 위해 휴리스틱 트리 탐색 알고리즘을 구현하여 사용 하였다. 각각의 이미지 파일에 대해서 전수검사를 진행해야 하므로 패턴검색 성능은 실제 상용 서비스에 적용하기 위해서 매우 중요한 부분이다. 이를 위해 패턴검색에서 사용되는 휴리스틱 트리 탐색 알고리즘을 구현하여 패턴검색 성능을 향상시켰다.

3.2 학습과정

그림 2 스팸이미지

학습의 경우 사전에 그림2와 같이 미리 수집된 스팸이미지로부터 패턴추출 과정을 거치게 된다. 이때 추출되는 패턴의 길이는 32바이트단위로 추출하게 되며, 각각의 이미지별로 총 3개의 패턴을 추출하는데, 추출되는 패턴의 위치는 랜덤하게 추출한다. 이러한 방법은 스팸메시지 판단 기준을 보다 엄격하게 하여 혹시 발생할 수 있는 비스팸메시지 차단을 방지하기 위함이다. 이와 같은 과정을 거쳐 만들어진 학습데이터는 패턴DB로 만들고 이미지 스팸을 판단하기 위해 사용된다.

4. 실험 및 결과

4.1 실험 데이터

본 연구에서는 실제 이미지스팸을 수집하여 실험에 사용하였다. 실제 이미지스팸 수신량이 텍스트기반의 스팸메시지보다 수신량이 미비하여 많은 양의 데이터를 사용하지 못하였다. 실험을 위해 총 1002개의 스팸이미지와 765개의 비스팸이미지를 이용하였다. 이중 스팸이미지 1002개의 데이터는 학습을 수행하였다. 스팸분류성능을 측정하기 위해 비스팸이미지와 스팸이미지 두 집합으로 구분하여 실험에 사용하였으며, 스팸이미지에대한 패턴검색 성능을 측정하였다.

4.2 실험 및 평가

표 1은 성능분석 시 사용된 장비의 사양정보이다. 장비의 성능이 패턴검색성능에 매우 큰 영향을 미칠 수 있다.

CPU	2.8GHz Dual Core Intel i5 Processor
MEM	8GB 1600MHz DDR3L
HDD	PCIe Flash Drive

표 1 테스트 시 사용된 장비사양

실험은 2가지로 나누어 진행하였다. 첫 번째로 사전에 수집된 스팸이미지 1002개를 이용하여 10회 반복 수행 후 스팸이미지의 평균패턴검색 시간을 측정하였다.

표2는 스팸이미지 1002개에 대해서 총 10회에 걸쳐 패턴검색성능을 측정한 결과 평균 탐색시간은 2,931㎲의 시간이 측정되었다.

스팸 이미지 개수	1002개
스팸 이미지 최소 크기	7,919 Byte
스팸 이미지 최대 크기	527,215 Byte
최소 이미지 탐색 시간	433 ㎲
최대 이미지 탐색 시간	42,016 ㎲
평균 총 탐색 시간	2,931 ㎲

표 2 패턴 검색 성능 테스트 요약

두 번째로 수집된 스팸이미지에 대해서 학습한 데이터를 기반으로 스팸이미지와 비스팸이미지 분류 성능을 측정하였다.

표3은 스팸 이미지와 비스팸 이미지의 스팸 분류 결과이다. 스팸이미지의 경우 스팸이미지 학습을 수행하였기 때문에 1002개 모두 스팸으로 분류되었으며, 비스팸 이미지의 경우 스팸 이미지 학습에 포함되지 않았기 때문에 비스팸으로 분류됨을 확인 할 수 있었다.

전체 이미지 개수	1713개
스팸 이미지 개수	1002개
비스팸 이미지 개수	711개
차단 개수	1002개
허용 개수	711개

표 3 스팸 이미지 검출 성능 테스트 요약

5. 결 론

본 논문에서 제시된 스팸 이미지 필터링 알고리즘은 앞에서 언급한 SVM 기반 이미지 스팸필터링의 성능상의 단점을 보완하고, 스팸 분류 방법을 패턴기반으로 구현하여 제시하였다. 제시하는 방법의 연구 실험 결과, 스팸분류성능의 경우 SVM 기반 이미지 스팸필터링 시스템은 약 7%의 분류성능 향상이 있었으며, 이미지 탐색시간의 경우 약 10만배의 성능 향상이 되었다.

하지만 학습되는 패턴의 수가 많아지면 패턴검색성능이 느려질 수 있다는 점과 다양한 데이터의 수집이 어렵다는 점이 존재한다. 때문에 학습된 패턴에 대해서 시일이 지나 현재 이미지 스팸 트렌드와 맞지 않는 데이터에 대해서는 삭제하여 성능상의 문제가 없도록 해야하는 번거로움이 존재 하며, 비스팸 이미지의 경우 스팸 이미지처럼 특정한 형태의 이미지가 아니기 때문에 보다 다양한 데이터의 테스트가 필요한 부분이므로, 본 논문에서 테스트한 결과를 신뢰할 수 없다는 단점이 존재 한다.

향후 학습 패턴이 많아지는 부분에 대한 방안 및 실망에서 사용되는 다양한 이미지에 대한 테스트가 이루어진다면 실상용화가 가능한 이미지 스팸필터링 시스템이 가능할 것이라고 전망한다.

참고문헌

[1] KISA, 2013년 상반기 스팸 유통현황 분석결과 보도자료, 11월, 2013

[2] KISA, 2012년 하반기 스팸 유통현황 분석결과 보도자료, 11월, 2013

[3] 아시아경제 뉴스, “문자 단속했더니 ‘그림스팸 극성’ ... 정부 차단책 마련, 5월, 2014

[4] Image Identifier using MPEG-7 Image Signature, 2013.

[5] J. M. Gómez et al., "Content Based SMS Spam Filtering", Proc. of the 2006 ACM Symposium on Document Engineering, pp. 107-114, 2006.

[6] Korean Mobile Spam Filtering System Considering Characteristics of Text Messages, 2010.

[7] M. Koppel et al., "Automatically categorizing written texts by author gender", Literary and Linguistic Computing, Vol. 17, No. 4, pp. 401-412, 2002.

[8] Spam MessageFiltering Based on SVM Using Image Processing Technology, 2014.

[9] Marios Anthimopoulos,“A two-stage scheme for text detection in video images”, Image and Vison Compution, vol. 28, p.1413-1426, 2010