



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년06월30일
(11) 등록번호 10-1045452
(24) 등록일자 2011년06월23일

(51) Int. Cl.
G06Q 50/00 (2006.01)
(21) 출원번호 10-2004-0036088
(22) 출원일자 2004년05월20일
심사청구일자 2009년04월20일
(65) 공개번호 10-2005-0000309
(43) 공개일자 2005년01월03일
(30) 우선권주장
10/601,741 2003년06월23일 미국(US)
(56) 선행기술조사문헌
US20020188689 A1
WO2002071286 A2

(73) 특허권자
마이크로소프트 코퍼레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
스타벅크, 브라이언티.
미국98019워싱턴주두발노쓰이스트140번코트27517
룬트웨이트, 로버트엘.
미국98024워싱턴주폴시티287
번애비뉴사우쓰이스트4148
(뒷면에 계속)
(74) 대리인
주성민, 이중희, 백만기

전체 청구항 수 : 총 29 항

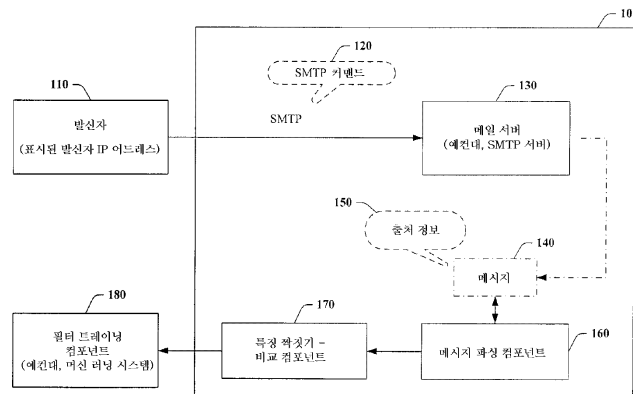
심사관 : 최석규

(54) 개선된 스팸 검출 시스템, 및 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법 및 컴퓨터-판독가능 기록 매체

(57) 요약

본 발명은 개선된 강건한 스팸 검출 시스템 및 방법을 개시하고 있다. 본 시스템 및 방법은 스팸머들이 회피하기 곤란한 특성을 찾아내고 스팸머들이 복제하기 곤란한 비스팸성 특성을 찾아내는 것을 개선 내지 촉진시키는 컴포넌트 및 동작을 포함한다. 상기한 특성의 예로는, 출처 특징 쌍을 조사하고, 문자열 및/또는 숫자열(스트링 또는 서브스트링)을 분석하고, 메시지 및/또는 특징의 사이즈에 대한 분석뿐만 아니라, 하나 또는 그 이상의 문자열(스트링 및/또는 서브스트링)의 여러 엔트로피 레벨을 검출하는 것을 포함한다.

대표도



(72) 발명자

헥커만, 데이비드이.

미국98008워싱턴주벨레뷰웨스트레이크삼마미쉬레인
노쓰이스트648

굿맨, 조슈아티.

미국98052워싱턴주레드몬드노쓰이스트38
번스트리트17424

길룸, 엘리엇씨.

미국95030캘리포니아주로스가토스오버룩로드18510

호웰, 나단디.

미국98133워싱턴주시에틀에이퍼티.에이노쓰105번스
트리트939

알던저, 케니쓰알.

미국98052워싱턴주레드몬드노쓰이스트78
번플레이스13527

특허청구의 범위

청구항 1

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍들 각각은,

MAIL FROM 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

HELO 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Received from 헤더 내의 IP 어드레스와 서브넷(subnet) 중 적어도 하나;

디스플레이 네임 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Message From 라인 내의 도메인 네임과 호스트 네임 중 적어도 하나; 및

최후의 Received from 헤더 내의 적어도 하나의 타임존(time zone)

중 적어도 하나를 포함하는, 스팸 검출 시스템.

청구항 2

삭제

청구항 3

제1항에 있어서,

도메인 네임은 호스트 네임으로부터 유도되는, 스팸 검출 시스템.

청구항 4

제1항에 있어서,

상기 서브넷은 처음의 다수의 비트들(a first number of bits)을 공통으로 공유하는 하나 이상의 IP 어드레스를 포함하는, 스팸 검출 시스템.

청구항 5

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍은 Message From 및 HELO 커맨드로부터의 도메인 네임과 호스트 네임 중 어느 하나인, 스팸 검출 시스템.

청구항 6

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍은 디스플레이 네임의 도메인 네임 및 호스트 네임과, Message From의 도메인 네임 및 호스트 네임인, 스팸 검출 시스템.

청구항 7

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍은 Message From 내의 도메인 네임과 호스트 네임 중 어느 하나와, Received From의 IP 어드레스와 서브넷 중 어느 하나인, 스팸 검출 시스템.

청구항 8

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍은 발신자측의 표시된(alleged) 타임존과, Message From의 도메인 네임인, 스팸 검출 시스템.

청구항 9

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 쌍은 발신자측의 메일링 소프트웨어의 타입과, SMTP 커맨드와 메시지 헤더 중 하나로부터 유도된 도메인 네임, 호스트 네임 및 사용자 네임 중 어느 하나인, 스팸 검출 시스템.

청구항 10

스팸 검출 시스템에 있어서,

메시지의 출처 정보(origination information)의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝(machine learning)을 이용하는 컴포넌트(740); 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하기 위해 추론 방식(inference scheme)을 이용하는 것에 관련하여 사용하기 위한 쌍(pair)들로 결합시키는 컴포넌트(770)를 포함하고,

상기 출처 정보는 SMTP 커맨드를 포함하며, 상기 SMTP 커맨드는 HELO 커맨드, MAIL FROM 커맨드 및 DATA 커맨드를 포함하는, 스팸 검출 시스템.

청구항 11

제10항에 있어서,

상기 DATA 커맨드는 Message From 라인, 발신자측의 표시된 타임존 및 발신자측의 메일링 소프트웨어를 포함하는, 스팸 검출 시스템.

청구항 12

제1항에 있어서,

메일 메시지에 대하여 일관되게 하나 이상의 경험적 학습법(heuristics)을 적용하여 일관된 특징 짝짓기(pairing)를 획득하는 컴포넌트를 더 포함하는, 스팸 검출 시스템.

청구항 13

제1항에 있어서,

상기 머신 러닝 필터를 트레이닝하는 것에 관련하여 메시지 헤더의 복수의 특징을 분석하는 컴포넌트를 더 포함하는, 스팸 검출 시스템.

청구항 14

제13항에 있어서,

상기 메시지 헤더의 특징은 적어도 하나의 메시지 헤더 타입의 존재와 부재 중 적어도 하나를 포함하고, 상기 메시지 헤더 타입은 X-Priority, 메일 소프트웨어 및 탈퇴(unsubscribing)를 위한 헤더 라인을 포함하는, 스팸 검출 시스템.

청구항 15

제14항에 있어서,

상기 메시지 헤더의 특징은 적어도 하나의 메시지 헤더 타입에 관련된 콘텐츠(content)를 더 포함하는, 스팸 검출 시스템.

청구항 16

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱(parsing)하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하게 하는 단계(830); 및

추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 각각의 쌍은,

MAIL FROM 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

HELO 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Received from 헤더 내의 IP 어드레스와 서브넷 중 적어도 하나;

디스플레이 네임 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Message From 라인 내의 도메인 네임과 호스트 네임 중 적어도 하나; 및

최후의 Received from 헤더 내의 적어도 하나의 타임존

중 적어도 하나를 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 17

제16항에 있어서,

파싱되는 상기 메시지의 적어도 일부분은 상기 메시지의 출처 정보에 대응하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 18

삭제

청구항 19

제16항에 있어서,

도메인 네임은 호스트 네임으로부터 유도되는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 20

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하는 단계(830); 및

추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 특징 쌍은 디스플레이 네임의 도메인 네임 및 호스트 네임과, Message From의 도메인 네임 및 호스트 네임인, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 21

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하는 단계(830); 및

추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 쌍은 Message From 및 HELO 커맨드로부터의 도메인 네임과 호스트 네임 중 어느 하나인, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 22

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들

을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하는 단계(830); 및
추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 특징 쌍은 Message From 내의 도메인 네임과 호스트 네임 중 어느 하나와, Received From의 IP 어드레스와 서브넷 중 어느 하나인, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 23

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하는 단계(830); 및
추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 특징 쌍은 발신자측의 표시된 타임존과 Message From의 도메인 네임인, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 24

스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법에 있어서,

적어도 하나의 메시지를 수신하는 단계(810);

메시지의 적어도 일부분을 파싱하여 헤더 특징을 자동으로 생성하는 단계(820) - 상기 파싱은 모든 헤더 특징들을 식별하기 위해 머신 러닝을 이용하는 것을 포함함 - ;

적어도 2개의 헤더 특징을 쌍으로 결합시켜, 각 특징 쌍이 적어도 하나의 추가 특징을 생성하는 단계(830); 및
추론 방식으로 상기 특징 쌍을 이용하여, 하나 이상의 특징에 기초한 스팸 이메일 검출에 후속하여 사용되는 머신 러닝 스팸 필터를 트레이닝하는 데 사용되는 트레이닝 데이터를 생성하는 단계(850)를 포함하고,

상기 특징 쌍은 발신자측의 메일링 소프트웨어의 타입과, SMTP 커맨드와 메시지 헤더 중 하나로부터 유도된 도메인 네임, 호스트 네임 및 디스플레이 네임 중 어느 하나를 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 25

제16항에 있어서,

하나 이상의 특징 쌍을 선택하여 상기 머신 러닝 필터를 트레이닝하는 단계를 더 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 26

제16항에 있어서,

새로운 메시지를 수신하는 단계;

상기 메시지 내의 출처 정보에 기초하여 특징 쌍들을 생성하는 단계;

상기 특징 쌍들을 머신 러닝 필터에 통과시키는 단계; 및

적어도 하나의 특징 쌍이 상기 메시지가 스팸일 가능성이 크다는 것을 나타내는지에 대한 판정을 획득하는 단계를 수행함으로써 상기 머신 러닝 필터가 스팸을 검출하도록 트레이닝된 후에, 상기 머신 러닝 필터를 채용하는

단계를 더 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 27

제16항에 있어서,

메시지 헤더의 복수의 특징을 분석하는 단계; 및

상기 분석된 특징을 이용하여 머신 러닝 필터를 트레이닝하는 단계

를 더 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 28

제27항에 있어서,

상기 메시지 내의 이미지 정보에 기초하여 복수의 특징을 분석하는 단계를 더 포함하는, 스팸 검출에 이용하기 위한 특징을 생성하는 것을 용이하게 하는 방법.

청구항 29

컴퓨터 상에서 실행될 때, 상기 컴퓨터가 제16항의 방법을 수행하게 하는 컴퓨터 프로그램 명령어들을 포함하는 컴퓨터 프로그램을 저장하고 있는, 컴퓨터-판독가능 기록 매체.

청구항 30

컴퓨터 실행가능 컴포넌트들이 저장되어 있는 컴퓨터 판독가능 기록 매체로서,

상기 컴퓨터 실행가능 컴포넌트들은

메시지의 출처 정보의 적어도 일부분에 관련된 모든 헤더 특징들을 자동으로 식별하기 위해 머신 러닝을 이용하는 컴포넌트; 및

상기 헤더 특징들을, 후속하여 스팸 검출에 사용되는 머신 러닝 필터를 트레이닝하는 것에 관련된 트레이닝 데이터를 생성하는 추론 방식(inference scheme)으로 사용하기 위한 쌍들로 결합시키는 컴포넌트를 포함하고,

상기 쌍들 각각은,

MAIL FROM 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

HELO 커맨드 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Received from 헤더 내의 IP 어드레스와 서브넷 중 적어도 하나;

디스플레이 네임 내의 도메인 네임과 호스트 네임 중 적어도 하나;

Message From 라인 내의 도메인 네임과 호스트 네임 중 적어도 하나; 및

최후의 Received from 헤더 내의 적어도 하나의 타임존

중 적어도 하나를 포함하는, 컴퓨터-판독가능 기록 매체.

청구항 31

제1항에 있어서,

적어도 하나의 메시지를 수신하기 위한 수단;

메시지의 적어도 일부분을 파싱하여 특징을 생성하기 위한 수단;

적어도 2개의 특징을 쌍으로 결합시켜, 특징 쌍 각각이 적어도 하나의 추가 특징을 생성하게 하기 위한 수단; 및

상기 특징 쌍을 이용하여 머신 러닝 스팸 필터를 트레이닝하기 위한 수단

을 더 포함하는, 스팸 검출 시스템.

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

[0025] 본 발명은 스팸 메시지의 식별을 위한 시스템 및 방법에 관한 것으로, 스팸머(spammer)들이 회피 및 복제하기 어려운 특성을 나타내는 스팸 메시지 식별 시스템 및 방법에 관한 것이다.

[0026] 인터넷과 같은 글로벌 통신 네트워크의 등장에 의해 대규모의 잠재 고객들에 접근할 수 있는 상업적인 기회를 제공하게 되었다. 전자메일("email")과 같은 전자 메시징은 네트워크 사용자들에게 원하지 않는 광고나 홍보물("스팸"이라고도 함)을 유포하기 위한 보편적인 수단이 되고 있다.

[0027] 컨설팅 및 시장조사 회사인 'Radicati Group, Inc.'에 따르면, 2002년 8월 기준으로, 매일 20억의 정크 메일 메시지가 보내지고 있으며, 이 숫자는 2년마다 3배로 증대할 것으로 예측하고 있다. 개인 및 단체(예컨대, 기업, 정부기관)들은 이들 정크 메시지로 인한 불편이 증대되고 있으며 때로는 불쾌감을 느끼기도 한다. 따라서, 머지않아 스팸은 신뢰적 컴퓨팅에 대한 주요한 위협이 될 것이다.

[0028] 스팸 차단(thwart)에 사용되는 일반적인 방법은 필터링 시스템/방법을 채용하고 있다. 검증된 필터링 방법의 하나는 머신 러닝 기법(machine learning approach)에 기초하고 있다. 머신 러닝 필터는 인입 메시지에 대하여 그 메시지가 스팸일 확률을 지정(assign)한다. 이 방법에서는, 일반적으로 2분류의 샘플 메시지(예컨대, 스팸 메시지와 스팸이 아닌 메시지)로부터 특징을 추출한 다음, 러닝 필터를 적용하여 이들 2분류의 메시지를 확률적으로 구분한다. 메시지의 특징은 콘텐츠(예컨대, 제목 내의 전체 단어 및 구문 및/또는 메시지의 본문)와 관련되어 있으므로, 이러한 유형의 필터는 일반적으로 "콘텐츠-기반 필터"라고 한다. 이러한 유형의 머신 러닝 필터는 일반적으로 정상(good) 메시지에서 스팸 메시지를 검출하여 구별해내기 위하여 완전일치(exact match) 기법을 채용하고 있다.

발명이 이루고자 하는 기술적 과제

[0029] 그러나, 스팸머들은 스팸 메시지를 정상 메일처럼 보이도록 수정하여 통상의 머신 러닝 및/또는 콘텐츠-기반 필터를 무능화시키거나 또는 메시지에 여러가지 잘못된 문자를 포함시켜 문자 인식 시스템을 혼동키는 경우도 있다. 따라서 이러한 통상의 필터들은 스팸에 대한 제한된 보호를 제공하고 있다.

[0030] 다음은 본 발명의 일부 태양을 이해하는데 기초가 되는 본 발명의 개요를 설명한 것이다. 본 개요는 본 발명에 대한 넓은 의미의 개략을 제공하는 것은 아니다. 즉, 본 개요는 본 발명의 필수(key/critical) 구성요소를 확인하거나 본 발명의 범주를 설명하기 위한 것이 아니다. 본 개요는 후술하게 될 상세한 설명에 앞서 본 발명의 일부 개념을 서두로서 간략하게 설명하기 위함이다.

[0031] 스팸 필터는, 그것이 머신 러닝에 기초한 것이든 기타의 기법에 기초한 것이든 무관하게, 그 메시지가 스팸인지의 여부를 판정하기 위하여 메시지의 콘텐츠를 확인해야만 한다. 그러나, 스팸머들은 메시지의 많은 부분에 대해 속임수(disguise)를 부릴 수 있는 경우가 있다. 이들은 스팸성 단어를 오기입하거나 동의어를 사용하거나 단어를 이미지화하여 사용할 수 있다. 스팸 필터는 그 이미지 내의 단어를 광학문자인식(OCR) 소프트웨어를 사용하여 식별해낼 수 있지만, 스팸머들이 고의로 OCR 시스템으로는 곤란한 이미지를 사용하는 경우에는, 이에 대한 비용이 너무 많이 소요된다. 스팸머들이 메시지에 대하여 속임수를 부릴 수 있는 여지를 줄이기 위해서, 스팸머들이 날조(fake)하기 어려운 특징을 생성할 수 있다.

[0032] 상기한 특징은 이메일 또는 메시지 파싱 컴포넌트에 의해 검출된 사실(fact)과 관련된다. 상기한 메시지 파싱 컴포넌트는 메시지 내의 각 단어의 특징을 생성할 수 있다. 또한, 상기한 컴포넌트는 사용된 구두점의 유형에 무관하게 구두점이 사용된 때마다 특징을 생성할 수 있다. 상기한 특징은 머신 러닝 필터에서 이용되거나 또는 핸드빌트(hand-built) 규칙의 일부분으로서 다른 방식으로 이용될 수도 있다.

발명의 구성 및 작용

[0033] 본 발명은 스팸머들이 날조하기 어렵고 통상의 스팸 필터에서 사용된 것과는 다른 추가의 특징을 포함시킴으로써 스팸을 검출 및 방지하도록 한 시스템 및 방법을 제공한다. 이러한 특징의 일례로는 메시지의 특징 쌍들을 확인하는 것을 들 수 있다. 스팸의 일부 특징들은 떼어놓고 보면 위조(forge)가 용이하고 무가치한 것이지만, 이들을 모아놓고 보면 매우 귀중한 것이 될 수 있다. 쌍으로 취급할 수 있는 특징의 예로는 해당 메시지의 출처 정보에 관한 것 또는 그로부터 유도된 것을 들 수 있다. 특히, SMTP(Simple Mail Transfer Protocol)에서의 도메인 및 호스트 네임; HELO 커맨드에서의 도메인 및 호스트 네임; 'Received from' 헤더 내의 IP 어드레스 또는 서브넷(subnet); 디스플레이 네임에서의 도메인 또는 호스트 네임; 'Message From' 필드 내의 도메인 또는 호스트 네임; 및 최후의 received from 헤더 내의 타임존(time zone)은 모두 어떤 방식으로든(또는 그 조합한 경우) 일치해야 한다. 따라서, 상기한 것들의 임의의 쌍은 머신 러닝 필터 또는 기타의 규칙-기반 필터의 트레이닝에 유용할 수 있다.

[0034] 두번째 특징으로는 한 주행길이의(a run of) 문자열을 조사하는 것이다. 메시지들의 가장 일반적인 특징은 메시지 내의 단어들에 관련되어 있으며, 특히 스페이스에 의해 분리된 단어에 관련되어 있다. 그러나, 메시지의 일부에 일정한 문자열(스페이스의 유무 포함)이 존재한다는 사실은 스팸의 표시일 수 있다. 따라서, 본 발명은 각 문자열 또는 구두점과 스페이스를 포함한 가능한 모든 문자열에 대해 생성된 특징을 이용하는 시스템 및 방

법을 제공한다. 일부 스팸어들은 제목란 또는 메시지의 시작 또는 끝에 속임문자(chaff)를 포함시켜 대부분의 스팸 필터링 시스템에서 발견되는 완전일치 기법을 무력화시킬 수 있다. 상기한 속임문자에는 정상적인 메일에서는 좀처럼 나타나지 않는 "xz" 또는 "qp" 등의 문자 N-그램(character n-gram)이 포함된다. 따라서, 속임문자 또는 문자 N-그램이 존재하거나 나타난다는 것은 악성 메시지(즉, 스팸)일 확률이 높음을 나타낸다. 상기한 문자 N-그램은 위치 종속적(position-dependent)일 수 있다. 따라서, 본 발명에서는 이와 같은 위치 종속성을 포함하는 특징을 생성하여 이용할 수 있다.

[0035] 머신 러닝 시스템에 사용될 수 있는 세번째 유형의 특징은 상기한 속임문자의 검출에 회귀 문자열을 이용하는 것에 대한 대안에 관한 것이다. 상기한 세번째 특징은 예컨대 문자 N-그램 언어모델을 이용하는 높은 엔트로피의 문자를 검출하는 것과 관련된다. 이 모델에서는, 각 문자에 대해 출현 빈도를 지정하여, 소정의 문자열이 다른 문자열에 비해 보다 자주 나타남을 나타낸다. 예를 들어, 문자열 "he"(예컨대 "the", "hear", "she", "threat" 등에서 나타남)는 임의의 주행길이의 문자열 또는 임의의 문자 스트링에서 시퀀스 "xz"보다 자주 나타날 것이다. 따라서, 상기한 문자열 "xz"에 대한 엔트로피는 상기한 시퀀스 "he"에 대한 엔트로피보다 높을 것이다.

[0036] 높은 엔트로피와 더불어, 메시지 또는 제목란의 시작과 끝 등에서 문자의 평균 엔트로피도 검출할 수 있다. 또한, 문자들의 상대 엔트로피에 관한 특징도 유용할 수 있다. 예를 들어, 제목란의 시작에서의 평균 엔트로피가 제목란의 중간에서의 평균 엔트로피보다 0.5 높은 경우에 대해 특징을 부여(designate)할 수 있다. 또다른 특징의 예로는 메시지의 중간보다 1.0 높은 메시지 본문 끝에서의 평균 엔트로피에 대해 특징을 부여할 수도 있다. 또한, 이들 검출된 높은, 평균 및/또는 상대 엔트로피의 이벤트는 개별적인 특징으로서 이용될 수 있다.

[0037] 네번째 유형의 유용한 특징은 일반(generic) 헤더에 관련된 것이다. 전형적인 머신 러닝 알고리즘은 메시지의 제목란과 본문에 공통의 특징을 사용하거나 또는 메시지 헤더에서 발견된 다른 공통 필드에 기초한 특징들을 사용하고 있을 뿐이다. 본 발명에서는, 전형적인 필터에서와는 달리, 헤더 라인 유형의 유무를 포함한, 실질적으로 모든 헤더를 이용한다. 특히, 본 발명의 머신 러닝 시스템은 모든 유용한 헤더 특징을 자동 식별할 수 있으며, 경우에 따라서는 일부 헤더 라인을 제외시킬 수도 있다.

[0038] 본 발명의 또다른 태양에 따르면, 머신 러닝 기술에서 유용한 전자메일(이메일) 통신의 추가 특징으로서, 이미지 특징뿐만 아니라 특징의 확장된 사이즈도 포함한다. 대부분의 스팸은 사이즈가 크지 않기 때문에, 전술한 특징들 중 하나와 결합한 특징들의 서로 다른 사이즈들은 스팸의 식별에 도움이 된다. 예를 들어, 메시지의 사이즈에 대응되도록 특징을 생성할 수 있다. 즉, 100 바이트 및 200 바이트 이상, b 바이트(b는 1 이상) 미만의 메시지 사이즈에 대해, 각 사이즈에 대한 특징을 생성하거나 또는 그 사이즈 범위에 대한 특징을 생성할 수도 있다. 스팸어들은 메시지의 출처를 혼동시키거나 속이기 위하여 매우 긴 디스플레이 네임을 사용하기로 하므로, 이를 제목란 및 디스플레이 네임의 사이즈에도 적용할 수 있다. 마찬가지로, 일부 사용자들은 메시지를 열지 않고 오직 제목란에 의지하는 경우도 있으므로, 스팸의 제목란에 그 메시지의 주요부분이나 본문 전체를 포함시키기도 한다.

[0039] 전술한 특징들은 머신 러닝 시스템에서 정크메일 및/또는 스팸 필터의 트레이닝 및 개선에 사용됨으로써, 이러한 필터에 대해서는 스팸어들이 메시지를 수정하기 곤란하도록 한다. 또한, 스팸어들이 자신의 메시지를 메시지 시스템을 통과시킬 가능성은 거의 없게 된다.

[0040] 전술한 목적 및 이와 관련된 것들을 실현하기 위하여, 첨부 도면을 참조하여 설명하고 있는 상세한 설명을 통하여 본 발명의 태양에 대해 설명하기로 한다. 그러나, 이들 태양은 본 발명의 원리가 적용될 수 있는 여러가지 방식들 중 일부에 대한 예시일 뿐이며, 본 발명은 전술한 태양 및 그 균등물을 포함하는 것으로 이해되어야 한다. 본 발명의 다른 장점 및 신규한 특징들은 도면을 참조하여 설명하고 있는 하기의 상세한 설명으로부터 명백하게 될 것이다.

[0041] <실시예>

[0042] 이하, 유사한 구성요소에 대해 유사한 참조번호로써 지칭하고 있는 도면을 참조하여 본 발명에 대해 상세하게 설명하기로 한다. 이하에서는, 본 발명의 이해를 도모하기 위하여, 설명의 편의상, 특정한 예에 대해 기술하기로 한다. 그러나, 이들 특정한 예에 의하지 않더라도 본 발명을 실시할 수 있는 것으로 이해되어야 한다. 또한, 본 발명을 보다 잘 설명하기 위하여 공지의 구조 및 장치는 블록도로서 도시하고 있다.

[0043] 본 명세서에서, 용어 "컴포넌트" 및 "시스템"은 컴퓨터 관련 실체로서, 하드웨어, 하드웨어와 소프트웨어의 결합, 소프트웨어를 의미할 수도 있으며, 소프트웨어에 의해 실행되는 것을 의미할 수도 있다. 예를 들어, 이것

에 국한되지는 않지만, 컴포넌트는 프로세서에서 실행되는 프로세스, 프로세서, 오브젝트, 실행가능물(executable), 일련의 실행(execution), 프로그램 및/또는 컴퓨터일 수 있다. 예를 들어, 서버에서 실행되는 애플리케이션이나 상기한 서버도 컴포넌트일 수 있다. 일련의 실행 및/또는 프로세스 내에 하나 또는 그 이상의 컴포넌트가 존재할 수 있으며, 하나의 컴포넌트가 하나의 컴퓨터 상에 로컬화되거나 및/또는 2개 또는 그 이상의 컴퓨터 사이에 분산될 수도 있다.

[0044] 본 발명은 머신 러닝 스팸 필터링을 위한 트레이닝 데이터의 생성과 관련한 각종의 추론(inference) 방식 및/또는 기술을 포함시킬 수 있다. 여기서, 용어 "추론"은 이벤트 및/또는 데이터를 통해 포획된 일조의 관측으로부터 시스템의 상태, 환경 및/또는 사용자에 관한 추리(reasoning) 또는 유추(inferring) 프로세스를 말한다. 상기한 추론은 특정한 상황(context)이나 액션을 식별하는데 이용될 수 있으며, 또는 예를 들어 상태에 대한 확률 분포를 생성할 수 있다. 상기한 추론은 확률적인 것으로, 데이터 및 이벤트를 고려한 주목 상태에 대한 확률 분포의 연산이다. 또한, 상기한 추론은 일조의 이벤트 및/또는 데이터로부터의 보다 높은 레벨의 이벤트를 구성하는데 이용된 기술을 참조할 수도 있다. 상기한 추론의 결과, 상기한 이벤트가 서로 시간적으로 근접하여 상관되어 있는지의 여부에 무관하게, 또한 상기한 이벤트 및 데이터가 하나 또는 수개의 이벤트 및 데이터 소스로부터 온 것인지에 무관하게, 일조의 관찰된 이벤트 및/또는 저장된 이벤트 데이터로부터 신규 이벤트 또는 액션을 구성하게 된다.

[0045] 여기서, 상기한 용어 '메시지'는 명세서에서 많이 사용되고 있지만, 상기한 용어는 전자메일에만 국한되지 않으며, 임의의 적절한 통신구조를 통해 배포될 수 있는 임의의 형태의 전자 메시지를 포함하는 것으로 이해되어야 한다. 예를 들어, 2명 또는 그 이상의 사람들 간의 회의를 가능케 하는 컨퍼런스 애플리케이션(인터랙티브 채팅 프로그램 및 인스턴트 메시징 프로그램)의 경우에도, 사용자들이 메시지를 교환하는 도중에 정상적인 채팅 메시지에 원하지 않는 텍스트가 전자적으로 끼어들어오거나, 및/또는 문두(lead-off) 메시지나 문말(closing) 메시지 또는 이들 양자 모두에 삽입될 수도 있으므로, 본 명세서에 개시된 필터링의 장점을 이용할 수 있다. 상기한 특정 애플리케이션의 경우에는, 상기한 필터는 상기한 바람직하지 않은 콘텐츠(예컨대, 광고, 홍보물 또는 선전 등)를 정크로서 포획 및 태그처리하기 위하여 특정 메시지 콘텐츠(텍스트 및 이미지)를 자동적으로 필터링하도록 트레이닝될 수 있다. 또다른 예로서, 셀룰러폰의 SMS 메시지도 필터링할 수 있다.

[0046] 도 1은 본 발명의 태양에 따른 스팸 검출을 도모하기 위해 특징 쌍들의 이용을 가능케 하는 시스템(100)을 개략적으로 나타낸 블록도를 예시하고 있다. 일부 특징들은 이미 개별적으로 이용되었다하더라도 쌍으로서 사용된 경우 특히 유용하다. 일반적으로, 메일이 어디에서 온 것인지에 대해 일치하는 정보가 존재한다. 예를 들어, IP 어드레스를 위조하는 것은 실용적이지 않다. 따라서, 다른 특징들이 IP 어드레스 특징과 합치하는지를 체크하도록 스팸 필터를 트레이닝할 수 있다. 일반적인 의미로 일치판정(match)을 이용할 수도 있다. 예를 들어, 동일한 서버를 사용하여 2개의 도메인(즉, Hotmail 및 MSN)으로부터 메일을 발신한 경우, HELO 커맨드 및 "from" 커맨드는 동일한 도메인을 제공할 필요가 없지만, 일반적으로 소정의 유효한 쌍으로만 나타날 것이다.

[0047] 도 1에 도시된 바와 같이, 하나 또는 그 이상의 발신자(110)로부터의 메시지는 이들 각각의 발신자(들)로부터 시스템(100) 내에 포함된 SMTP 서버 등의 메일 서버(130)에 전달된다. 상기한 메시지의 전달은 예를 들어 다수의 SMTP 커맨드(120)를 통해 실현될 수 있다. 다른 메일 전달 프로토콜도 가능하며 본 발명에 유사한 방식으로 적용될 수 있다.

[0048] 따라서, 상기한 메시지(140)가 상기한 배달의 수신자에 의해 받아들여지기 전에 일련의 SMTP 커맨드가 발행되어 분석될 수 있다. 특히, 상기한 특징 쌍들을 확립하는데 필요한 출처(origination) 정보(150)는 SMTP 커맨드(120)에서 찾을 수 있다. 상기한 출처 정보(150)로부터 특징을 유도하기 위하여, 상기한 정보(150)를 메시지 파싱 컴포넌트(160)에서 평가 및 파싱한다. 상기한 메시지(140)의 출처 정보(150) 중 적어도 일부가 파싱되면, 이 파싱된 부분 또는 특징은 특징 파싱 컴포넌트(170)에 통신된다. 상기한 특징 파싱 컴포넌트(170)는 상기한 특징의 임의의 가능한 조합을 분석하여, 그 결과의 특징 쌍이 추가의 특징으로서 사용되도록 한다. 이 특징 쌍들은 예를 들어 스팸 필터의 트레이닝 시, 상기한 시스템(100)에 연동하는 필터 트레이닝 컴포넌트(180)에 의해 이용된다.

[0049] 상기한 필터가 충분히 트레이닝되면, 머신 러닝 시스템과 함께 사용될 수 있으며, 또다른 그룹의 메일 메시지에 적용되어 그 그룹으로부터 스팸성 메시지를 필터링한다. 상기한 필터는 정기적으로 갱신될 수 있으며, 및/또는 필요한 경우 적절한 메일과 스팸 메일을 효과적으로 구분하기 위하여 새로운 필터를 형성할 수도 있다.

[0050] 도 2 내지 도 4는 본 발명의 태양에 따른 SMTP 커맨드에서 발견되는 출처 정보로부터 파싱되며, 다수의 유용한 쌍들(예컨대, 도면에서 화살표로 표시된 것들)에 결합될 수 있는 여러가지 특징들을 나타낸 것이다. 예를

들어, 도 2에서, 제1 SMTP 커맨드는 HELO 커맨드(200)로서, 발신측 머신이 자신이 네임을 예컨대 'HELO x.y.z' 임을 말해주고 있다. 여기서, 상기한 'x.y.z'가 'mail1.ddd.com'의 형태인 경우, "mail1.ddd.com"은 호스트 네임(210)이며, "ddd.com"은 도메인 네임(220)이라고 할 수 있다. 따라서, 호스트 네임(210)에서 도메인 네임(220)을 떼어낼 수 있다.

[0051] 또한, 발신자 IP 어드레스를 검출할 수 있다: 상기한 SMTP 프로토콜은 일반적으로 TCP/IP에서 사용되므로 이 통신에 사용된 IP 어드레스는 수신자에게 알려진다. 또한, 상기한 IP 어드레스(230)는 서브넷(240)이라고 하는 그룹으로서 판매 또는 사용되는 것이 보통이다. 서브넷(240)은 여러가지 방식으로 규정될 수 있지만, 실제로는 서브넷의 일례는 최초 24 비트를 공유하는 모든 IP 어드레스를 포함하는 것으로 규정된 것일 수 있다. 따라서, 상기한 HELO 커맨드(200)가 'HELO ddd.com'라고 말하고 있는 경우, 'ddd.com'으로부터 발신하고 있는 다수의 머신이 존재할 수 있지만, 이들 대부분의 발신측 머신은 동일한 서브넷(240)에 존재할 것이다.

[0052] 일반적으로, 일부 특징의 쌍들은 다른 것들에 비해 이해가 되지 않을 수도 있다. 예를 들어, HELO 호스트 네임(210)과 HELO 도메인 네임(220)의 짝짓기(pairing)는 이들 중 하나는 다른 하나로부터 유도된 것이므로 보다 덜 유용하다. 그러나, 서브넷(240)과 HELO 도메인 네임(220)의 짝짓기는, 이들 특징의 적어도 일부가 정상적인 환경에서 일치해야만 하므로 매우 유용한 것이다.

[0053] 상기한 HELO 커맨드(200)의 다음에, 그 메시지의 발신측 IP 어드레스 및 'x.y.z'를 포함하는 라인과 표시된(alleged) 타임존을 포함한 시각이 해당 메시지의 수신측의 'Received from' 라인에 첨부될 수 있다. 스팸 필터는 상기한 헤더를 스캔하여 HELO 커맨드(200)에서 보여주는 것들을 확인할 수 있다. 상기한 발신자측의 표시된 타임존(하나의 특징)은 상기한 메시지의 타임스탬프(또다른 특징)와 일치해야 한다. 또한, 표시된 타임존은 상기한 HELO 커맨드(200) 내에 표시된 발신자의 머신 네임 또는 IP 어드레스와 일치해야 한다. 이들이 불일치하는 경우 스팸임을 나타내게 된다.

[0054] 일반적으로, HELO 커맨드(200)에서의 IP 어드레스(예컨대, HELO x.y.z)는 'Received from' 라인 내의 IP 어드레스 또는 표시된 머신 네임과 동일 또는 일치해야 하지만, 스팸머들은 자신의 IP 어드레스에 대하여 올바른 호스트 또는 도메인 네임을 사용하지 않음으로써 메일을 위조할 수 있다. 따라서, 이것이 불일치하면 스팸인 것을 나타내게 된다. 여기서, IP 어드레스(230)의 서브넷(240)이 도메인 네임(220)과 일치하지 않을 확률은, IP 어드레스(230)가 도메인 네임(220)과 일치하지 않을 확률보다 낮다.

[0055] 도메인에 따라서는 자신의 머신에 대하여 HELO 커맨드(200)에 올바른 머신 네임을 정확하게 제공하도록 구성하고 있지 않은 경우도 있으며, 이러한 경우에 대해서, 필터는 특정 쌍을 후속 메시지에서 다시 발견한 경우 해당 쌍을 올바른 메시지로 받아들이도록 함으로써 무엇이 특정 쌍인지에 대해 학습(러닝)할 수 있다. 따라서, 상기한 필터는 필터의 트레이닝과 필터의 이용 시에 소정의 일관성이 유지되는 한, 출처 정보에서의 사소한 예러나 불일치뿐만 아니라 개인적 선호도도 수용하도록 트레이닝될 수 있다. 또한, 예컨대 유효한 호스트 네임과 유효한 IP 어드레스로 채워진 가능한 쌍들의 리스트를 생성하여, 이 리스트에 존재하지 않는 것으로 검출된 출처 정보는 스팸일 가능성이 높게 된다.

[0056] 본 발명의 태양에 따르면, 도 3에 나타낸 바와 같이 다음 커맨드로서 MAIL FROM 커맨드(300)가 예시되고 있다. 상기한 MAIL FROM 커맨드는 'Envelope From'으로서도 알려져 있으며 MAIL FROM 'a@b.c.d'의 형태를 갖는다. 여기서, 'b.c.d'는 상기한 'x.y.z'와 동일 또는 상이할 수도 있지만, 본 예시에서는 서로 동일해야 한다. 특히, 호스트 네임(310) 부분은 일치해야 한다. 예를 들어, 상기한 필터에서 유효한 일치를 나타내기 위해서는 'c.d' 부분은 'y.z' 부분과 동일 내지 대응되어야 한다. 상기한 호스트 네임(310)에서 도메인 네임(320)을 떼어내어, 상기한 HELO IP 어드레스(230)(도 2 참조)에서와 마찬가지로, 추가의 특징 쌍들을 제공할 수도 있다.

[0057] DATA 커맨드(400)에서와 같이, 상기한 메시지의 다음에 'From: e@f.g.h' 형태의 라인이 추가될 수 있다. 또한, 호스트 네임(410) 'f.g.h'는 상기한 'x.y.z' 및 'b.c.d'와 동일할 수 있다. 이와 달리, 적어도 도메인 네임(420) 'g.h'는 'y.z' 및 'c.d'와 일치해야 하지만, 항상 그럴 필요는 없다. 상기한 From 라인은 'Message from'이라고도 한다. 또한, 상기한 라인이 'From: "i"<e@f.g.h>'의 형태로 된 경우도 있다. 여기서, "i"는 디스플레이 네임(430)이라고 한다. 실제로, 많은 이메일 클라이언트에서는 'e@f.g.h' 대신에 상기한 디스플레이 네임 "i"만을 표시하고 있다. 그러나, 상기한 "i"를 "j@k.l.m"과 같은 형태로 표시하여, 사용자로 하여금 메시지 발신자의 실체를 오인하도록 하는 경우도 있다. 이러한 디스플레이 네임은 일반적인 경우가 아니므로, 이것만으로도 스팸이라고 할 수 있다. 그러나, "i"가 "j@k.l.m"의 형태로 존재한다면, 'k.l.m'이 다른 호스트 네임과 일치해야 하며, 또는 적어도 도메인 네임들이 일치(예컨대 'l.m'이 'g.h'에 대응함)해야 한다.

- [0058] 경우에 따라서는, 특정한 3자릿글자(triple)(예컨대, x.y.z)가 호스트 네임 또는 도메인 네임인지를 밝히기 어렵거나 마땅하지 않은 경우가 있을 수도 있다. 때로는, 추측(guess)이 필요한 경우도 있다. 예를 들어, HELO 커맨드가 'x.y.z' 형태의 어드레스를 제공하고, MAIL FROM 커맨드가 'y.z' 형태의 어드레스를 갖는다면, 'x.y.z'는 호스트 네임이며 'y.z'는 도메인 네임이라고 어느 정도의 확신을 가질 수 있다. HELO 커맨드가 'x.y.z' 형태의 어드레스를 제공하고, MAIL FROM 커맨드가 'b.c.d' 형태의 어드레스를 갖는다면, 'x.y.z' 및 'b.c.d'는 모두 호스트 네임 및 도메인 네임이며 'y.z' 및 'c.d'는 도메인 네임이라고 추측할 수 있을 것이다. 최선의 추측만을 특징으로서 채용하는 것이 아니라, 모든 추측을 특징으로서 채용할 수 있다. 다른 방법으로, 최선의 추측을 이용할 수도 있다. 이러한 추측을 수행하기 위한 간단하고 다양한 발견적 학습법을 추론해낼 수 있다. 일반적으로, 머신 러닝 시스템의 취급에 있어서, 상기한 추측이 주어진 유형의 메일에 대해 일관적인 한 상기한 추측이 항상 옳다고 하는 것은 중요하지 않으며, 동일한 특징 쌍들은 일관되게 일어난다.
- [0059] 또한, SMTP MAIL FROM 커맨드 내의 도메인 및 호스트 네임에 관련된 특징; HELO 커맨드 내의 도메인 및 호스트 네임에 관련된 특징; Received from 헤더 내의 IP 어드레스 또는 서브넷에 관련된 특징; 디스플레이 네임 내의 도메인 또는 호스트 네임에 관련된 특징; 'Message From'에서의 도메인 또는 호스트 네임에 관련된 특징; 최후의 Received from 헤더 내의 타임존에 관련된 특징; 및 발신자가 사용한 메일링 소프트웨어의 종류에 관한 특징은 모두 어떻게든 일치해야 한다. 상기한 IP 어드레스와 서브넷을 제외하고는, 스팸머들에 의해 리스트되어 있는 속성들의 거의 대부분이 위조될 수 있으므로, 전술한 특징들 중 임의의 쌍들은 유용하다. 따라서, IP 어드레스 또는 서브넷을 포함하고 있는 쌍들은 다른 임의의 특징들과 결합된 경우 매우 효과적이고 유용하다.
- [0060] 도 5는 메일 메시지의 Message From 라인 및 Received from 헤더(전체적으로 500으로 지칭함)로부터 유도될 수 있는 특징 쌍들을 예시하고 있다. 도시된 바와 같이, 도메인 네임 "domain.com"(510)은 IP 어드레스(520), 발신자측의 표시된 타임존(530), 및/또는 IP 어드레스(520)의 서브넷(540)과 일치하도록 짝짓기될 수 있다. 또한 다른 방법으로, 상기한 발신자측의 표시된 타임존(530)은 발신자측의 표시된 IP 어드레스(520)와 일치하도록 짝짓기될 수 있다. 도면에 예시되지 않은 다른 특징 쌍들뿐만 아니라 본 명세서에 도시되지 않은 다른 특징들에 대해서도 전술한 내용이 적용될 수 있다.
- [0061] 도 6 및 도 7은 개선된 스팸 검출을 도모하는 추가의 특징 생성 시스템을 도시하고 있다. 도 6은 한 주행길이의 문자열에 관련된 특징과 적어도 이들 문자열의 엔트로피에 부분적으로 기초하고 있는 특징들을 생성하기 위한 시스템(600)의 일례에 대한 블록도를 나타낸 것이다.
- [0062] 메시지 내의 대부분의 특징은 그 메시지에서 발견된 단어에 관련되어 있다. 그러나, 메시지 또는 메시지의 일 부분에 소정의 문자열이 존재하고 있다는 사실도 유용할 수 있다. 예를 들어, 스팸머들은 "rich" 또는 "RICH"에 대신하여 "R.I.C.H." 등과 같은 문자열을 사용하기도 한다. 패턴-매칭 기술을 이용하면, 이와 같이 "R.I.C.H."로 기록된 단어로부터 기본 단어 "RICH"를 용이하게 추출하여 찾아낼 수 있다.
- [0063] 또한, 스팸머들은 제목란 또는 메시지의 시작 또는 끝에 무작위 문자들(letters)을 속임문자로서 추가시키기도 한다. 이는 통상의 필터에서 사용하고 있는 완전일치 기술을 무능화시킨다. 이들 무작위 문자열은 정상적인 메일에서는 거의 찾아볼 수 없는 "xz" 또는 "qp" 등의 문자 N-그램을 포함하고 있을 가능성이 있기 때문에, 메시지에 이러한 것들이 나타난다는 것은 그 메시지가 악성(예컨대 스팸)이라는 강한 표시자일 수 있다. 또한, 스팸머들은 기호나 마침표 및 하이픈 등의 구두점을 임의로 추가하여, 스팸의 특성으로서 알려져 있는 단어 및/또는 구문을 변형시킴으로써, 전형적인 스팸 필터를 빠져 나갈 수 있다.
- [0064] 이러한 의도적인 조작을 둔화시키기 위하여, 상기한 시스템(600)에서는 의도적인 문자의 치환, 삽입 및 오기입을 식별 및 검출하기 위하여 가능한 문자열 각각에 대한 특징을 생성한다. 이를 위해, 상기한 시스템(600)은 텍스트를 하나의 문자씩 답보(walk through)해 가면서 주행길이 n(예컨대, n은 1 이상의 정수) 각각에 대한 특징을 생성하며, 그 결과 단어, 구두점, 스페이스 및 기타 콘텐츠를 효과적으로 포착(pick up)하게 된다.
- [0065] 예를 들어, 발신자(610)는 도시된 바와 같이 메시지(620)를 발신한다. 상기한 메시지(620)는 메시지 서버(630)에 전달되며, 상기 서버(630)에서 상기 메시지(620)는 문자 시퀀싱 컴포넌트(640)에 의해 처리되어 하나 또는 그 이상의 특징을 얻게 된다. 상기한 문자 시퀀싱 컴포넌트(640)는 스팸을 나타내는 특정 문자열, 스트링 및/또는 서브스트링을 검색함으로써 상기한 메시지의 적어도 일부분을 분석한다. 상기한 문자열, 스트링 및/또는 서브스트링은 완전한 단어 또는 스페이스로 분리된 단어일 필요는 없다.
- [0066] 예를 들어, 상기한 메시지(620)가 다음을 포함하고 있다고 가정해 보자.
- [0067] "Get Rich ~~quick~~ by Calling now!!!!!"

- [0068] 주행길이가 6인 경우, 다음의 문자열(650)의 예를 생성할 수 있다.
- [0069] "Get Ric"
- [0070] "et Rich"
- [0071] "t Rich "
- [0072] " Rich ~"
- [0073] 주행길이가 7인 경우, 다음의 문자열(650)의 예를 생성할 수 있다.
- [0074] "Rich ~-q"
- [0075] "ich ~-qu"
- [0076] "now!!!!"
- [0077] 상기한 문자열(650)이 식별 및 생성되면, 특징 생성 컴포넌트(660)는 각 문자열에 대한 특징(670)을 생성하게 된다. 그러면, 이 특징(670)은 필터 트레이닝 컴포넌트(680)에서의 예를 들어 스팸 필터의 트레이닝에 이용되게 된다.
- [0078] 동일 메시지에 대해, 하위길이(sublength)(서브스트링)와 개별길이(스트링) 양방을 추종(keep track of)하기 위하여, 예를 들어 1 내지 n까지의 복수의 주행길이를 사용할 수도 있다.
- [0079] 문자 N-그램에 관해서는, 상기한 N-그램이 어디에 나타나는지에 따라서 동일 또는 다른 특징을 사용할 수 있다. N-그램은 From 어드레스, 제목란, 텍스트 본문, html 본문 및/또는 첨부물에 위치할 수도 있다. 또한, N-그램 특징은 메시지 내에서의 위치에 따라서 생성되도록 할 수도 있다. 예를 들어, 속임문자(예를 들어 N-그램을 포함함)는 제목의 시작 또는 끝에 나타나는 경향이 있으므로, 제목란의 시작 또는 끝에 있는 희귀한 문자열은 중간에 위치하는 경우에 비해 보다 스팸이 가능성이 높다. 따라서, 제목란의 시작과 끝과 같이 원하는 위치에서만 속임문자 및/또는 N-그램을 검출하도록 상기한 시스템(600)을 프로그램할 수 있다. 마찬가지로, N-그램 특징도 메시지의 시작 또는 끝에 대해 위치 중속적일 수 있다.
- [0080] 상기한 시스템(600)은 외국어에 대해서도 유용하며, 특히 한국어 및 일본어(예컨대 히라가나 및 가타카나)와 같이 스페이스로 단어를 분리하고 있지 않은 경우에는 특히 유용하다. 전술한 바와 같이, 서로 다른 문자 길이의 거의 대부분의 문자열을 용이하게 검출할 수 있다. 또한, 상기한 시스템(600)은 스페이스가 거의 없는 경우, 영어에서 좀처럼 사용되지 않는 문자들이 다수 검출된 경우(즉, 큰 바이트의 문자), 또는 소정의 유니코드(unicode) 문자 타입이 검출된 경우와 같이, 해당 텍스트가 외국어로 된 것으로 의심되는 경우에만 호출될 수도 있다. 따라서, 상기한 문자 N-gram은 대문자(uppercase), 소문자(lowercase), 구두점 또는 스페이스 특성을 갖지 않는 것으로 판정된 문자에 대해서만 사용된다. 예를 들어, 메시지를 스캔해 본 결과, 스페이스가 거의 없고 및/또는 큰 바이트의 긴 스트링의 문자열이 검출된 경우에, 상기한 N-그램 시퀀싱을 호출(involve)한다. 이와 같이 문자 N-그램을 제한적으로 사용하게 되면, 전술한 바와 같이, 모든 이메일에 대해서 무제한(full) N-그램을 사용하기 위해서는 비용이 많이 소요되며 많은 시간이 걸리므로, [예컨대 모든 메시지에 대한] 무제한 N-그램보다 유익하다.
- [0081] 희귀 문자열을 사용하는 것은 속임문자를 검출하는 한 방법이지만, 많이 있을 수 있는 희귀 문자열 각각에 대한 리스트를 작성할 필요가 있다. 본 발명의 또다른 태양에 따르면, 속임문자를 검출하는 또다른 방법은 높은 엔트로피의 문자열을 검출하는 것과 관련된다. 높은 엔트로피의 문자열을 검출하는 것은 보다 비용면에서 효과적이며 보다 유효한 스팸 식별 방법일 수 있다.
- [0082] 도 6을 참조하면, 상기한 시스템(600)은 속임문자의 검출을 위한 또다른 및/또는 추가의 기술을 제공하는 엔트로피 검출 컴포넌트(690)를 구비하고 있다. 상기한 엔트로피 검출 컴포넌트는 해당 메시지가 스팸임을 나타내는 무작위 문자열의 실례(instances)를 검색함으로써 메시지의 적어도 일부분을 분석한다.
- [0083] 문자열의 엔트로피는 본질적으로 해당 문자열의 비유사성 또는 무작위성과 같다. 일반적으로, 문자열 "abc...z"의 확률 P를 $P(abc...z)$ 라고 하면, 이 문자열의 엔트로피는 다음과 같다.
- [0084] $-\log_2 P(abc...z)$

- [0085] 평균 엔트로피 또는 문자(a, b, c, ..., z)별 엔트로피는 다음과 같이 표현되며,
- $$\frac{-\log_2 P(abc...z)}{\text{length}(abc...z)}$$
- [0086]
- [0087] 이것은 속임문자의 인식 및 식별과 유사한 방식으로 사용될 수 있다. 엔트로피 측정 단위는 "비트"이다.
- [0088] 문자열의 확률을 얻는 방법은 여러가지가 있을 수 있다. 예를 들어, 정상 및 악성 이메일 전체를 이용하거나, 및/또는 이메일 이외의 데이터베이스를 이용함으로써, 기존의 정상 이메일에 대해 문자 N-그램 언어 모델을 트레이닝할 수 있다. 다른 발견적 학습법을 이용하여 높은 엔트로피 또는 평균 엔트로피를 검출할 수도 있다. 예를 들어, 공통(common) 문자 쌍 또는 3자릿글자(예컨대 유효 문자열 또는 2 및 3개 문자, 각각)의 리스트를 작성할 수도 있다. 또한, 상기한 리스트에 나타나지 않는 임의의 주어진 문자열 내에서의 상기한 문자 쌍 또는 3자릿글자의 비율을 해당 문자열의 엔트로피 판정에 포함시킬 수도 있다.
- [0089] 실제로, 개선된 강건한 스팸 검출 시스템을 제공하기 위해서는 상대 엔트로피가 매우 유용할 수 있다. 보다 구체적으로, 제목란의 시작과 끝에서는 중간에 비해 평균 엔트로피가 높게 또는 비교적 높게 검출될 수 있다. 실제로, 예를 들어, 제목란의 시작에서는 중간에 비해 평균 엔트로피가 0.5 비트 높게 나타난다.
- [0090] 또한 다른 방법으로, 메시지의 시작 또는 끝에서의 평균 엔트로피는 메시지 전체의 평균 엔트로피, 또는 메시지의 중간에서의 평균 엔트로피에 비해 높을 수 있다. 예를 들어, 메시지의 끝에서의 평균 엔트로피는 메시지의 중간에서 보다 적어도 1 비트 높다(예를 들어, 단위를 비율 또는 인수(factor)로 변환할 수도 있음). 따라서, 많은 특징들이 사용될 수 있다.
- [0091] 무작위 문자열뿐만 아니라, 대부분의 스팸은 텍스트 대신에 이미지를 포함하고 있다. 이들 이미지는 단지 1이나 0 또는 다른 숫자의 시퀀스이다. 이 때문에, 스팸머들은 이미지 숫자열에 최소량의 스테틱(static)을 입력함으로써 통상의 스팸 필터링 시스템을 통과하는데 성공할 수 있다. 따라서, 이미지의 엔트로피도 문자열의 엔트로피와 유사한 방식으로 판정될 수 있다. 또한, 하나 이상의 메시지에서 검출되게 될 이미지들을 서로 비교할 수도 있다. 이들이 실질적으로 유사한 것으로 판정되면, 동일한 또는 실질적으로 동일한 이미지를 포함하는 모든 메일을 차단할 수 있다.
- [0092] 마지막으로, 문자열 및 이미지열의 엔트로피 이벤트에 관련된 특징(670)은 필터 트레이닝 컴포넌트(680)에서의 머신 러닝 필터의 트레이닝에 이용될 수 있다.
- [0093] 도 7을 참조하면, 본 발명의 일면에 따른 스팸 검출을 가능케 하는 특징 생성 시스템(700)의 일례를 블록도로 도시하고 있다. 시스템(700)은 메시지 서버(710)를 구비하고 있으며, 발신자(720)가 메시지(730)를 발신하면 그 수신측에 도달하기 전에 상기한 메시지 서버(710)에 전달된다. 메시지 서버(710)에서, 상기한 메시지는 메시지 헤더 분석 컴포넌트(740), 이미지 프로세싱 컴포넌트(750) 및/또는 메시지 및 특징 사이징 컴포넌트(760)에 의해 파싱되어 다수의 특징을 얻게 된다.
- [0094] 상기한 메시지 헤더 분석 컴포넌트(740)는 머신 러닝 필터의 트레이닝과 연계하여 메시지 헤더의 실질적으로 모든 특징을 분석한다. 특히 상기한 머신 러닝은 모든 유용한 헤더 특징을 자동적으로 식별하도록 채용될 수 있다. 이러한 방법의 일례로는, "X-Priority: 3" 등의 특정한 헤더뿐만 아니라, "X-Priority" 등의 헤더 라인 타입의 존부에 적어도 일부 기초하여 특징을 생성하는 것을 포함한다. 또한, 탈퇴(unsubscribing)에 관한 헤더 라인은 스팸의 식별에 보다 유용하다.
- [0095] 일부 헤더 라인은 사용자의 선호도에 따라서 제외시킬 수도 있다. 또한, 발신자가 사용한 메일 소프트웨어의 종류 등과 같은 헤더 라인 콘텐츠는 스팸의 검출에 유용하다. 모든 헤더 라인에 대해 그 존부, 및/또는 콘텐츠를 조사 및 분석하는 것은, 종래의 제목란과 메시지 본문에서의 특징만을 이용하고 있는 전형적인 머신 러닝 알고리즘에 비해 개선된 점들이다. 일부 머신 러닝 알고리즘에서도 이메일 헤더에 기초하여 특정한 특징을 채용하고 있지만, 이전의 방식에서는 헤더 라인 내의 모든 또는 실질적으로 거의 대부분의 특징을 사용하고 있지는 않다.
- [0096] 스팸머들은 이미지의 경우 텍스트에 비해 필터에서 분석하기가 보다 곤란하고 많은 시간이 소요된다는 이유로 텍스트 보다는 이미지를 사용하고자 하기 때문에, 이미지 프로세싱 컴포넌트(750)는 메시지 내에 포함된 이미지에 기초한 각종의 특징을 파싱해 낼 수 있도록 적용된다. 예를 들어, 이미지의 갯수, 위치(예컨대 메시지에 임베드되거나 외부로 링크됨) 및/또는 타입(예컨대, JPGs 및/또는 GIFs)을 메시지에서부터 확인하여 특징으로서 사용할 수 있다. 또한, 이미지의 X-Y 치수뿐만 아니라 그 사이즈(예컨대 바이트)도, 메시지 내에 임베드된 이미

지에 대해 최소의 프로세싱을 통해 판별해 낼 수 있다.

- [0097] 개인 디지털 사진을 포함하는 정당한 메일이 차단되는 것을 방지하기 위하여, 이미지의 사이즈가 일반적인 디지털 카메라에 의해 제작된 사이즈 및/또는 치수와 일치하도록 특수한(일반적으로 포지티브) 특징을 생성할 수도 있다. 이미지의 총 넓이뿐만 아니라 바이트 단위의 이미지 사이즈에도 특징을 관련시킬 수 있다. 마지막으로, 메시지 내의 이미지가 다른 곳(예컨대 메시지 외부)으로 링크하고 있는지의 여부에 관련된 특징은, 대부분의 스팸머들은 자신의 스팸에 외부로의 링크(external link)를 포함시키고 있으므로, 스팸임을 나타낼 수 있다.
- [0098] 또다른 방법으로, 메시지에 클릭가능한(clickable) 이미지를 포함하고 있어, 이미지 자체가 외부로 링크된 이미지가 아니라 클릭가능한 하이퍼링크로서 사용될 수도 있다. 이 경우, 메시지 내의 HTML 텍스트는 <AHREF="first URL">와 같은 태그 패턴을 포함하고 있다. 여기서, 'first URL'과 'second URL'은 서로 다르다. 따라서, 상기한 태그 패턴의 적어도 일부에 관련된 특징을 스팸 필터의 트레이닝에 이용할 수 있다. 일반적으로, HTML 속성 및 태그 패턴 내에서의 개별 위치는, 대부분의 스팸머들이 텍스트가 아닌 이미지를 사용하여 스팸 필터를 피하고자 하므로, 스팸임을 나타낼 수 있다. 따라서, 이러한 정보를 추출하여 필터의 트레이닝에 이용할 수 있다.
- [0099] 대부분의 스팸은 사이즈가 크지 않기 때문에, 많은 서로 다른 사이즈 특징들을, 메시지 사이징 컴포넌트(760)에서 그 사이즈를 하나 또는 그 이상의 버킷(bucket)으로 분할(granularize)하여 이용할 수 있다. 한 방법으로는 특징들을, 메시지 사이즈가 100, 200, 400 및 b 바이트(여기서 b는 1 이상의 정수)를 넘는 경우로 분할하는 것이다. 이 분할은 사이즈가 500 바이트인 메시지는 사이즈가 100, 200, 400을 넘는 특징들과 연관되도록 중첩하는 버킷들에 기초할 수 있다. 다른 방법으로, 중첩하지 않는 버킷에 기초할 수도 있다. 즉, 각 버킷들이 특정한 사이즈의 메시지에 속하게 하여, 하나의 사이즈 특징이 각 메시지에 연관되도록 한다. 예를 들어, 메시지 사이즈가 100 바이트 미만인 버킷과, 100 이상 200 바이트 미만의 버킷과, 200 이상 400 바이트 미만의 버킷과, b 바이트 미만의 버킷이 존재할 수 있다.
- [0100] 두번째 방법으로, 스팸 및/또는 스팸성 메시지는 예컨대 속임문자 등이 존재하므로 제목란과 디스플레이 네임이 큰 경향이 있으므로, 상기한 메시지 사이징 컴포넌트를 보다 작은 사이즈 규모를 갖는 제목란 및 디스플레이 네임에 적용할 수도 있다.
- [0101] 또한, 메시지를 식별하여 정보 비트를 식별하는 과정에서, 특징 생성 컴포넌트(770)가 상기한 정보로부터 특징들을 생성한 다음, 이들 또는 그 중 선택된 부분을 필터 트레이닝 컴포넌트(780)에서 연계하여 사용하도록 통신할 수도 있다.
- [0102] 본 발명에 따른 각종의 방법들을 일련의 동작을 통해 설명하기로 한다. 그러나, 본 발명에 따르면 일부 동작들이 도시된 순서와는 다른 순서로 일어날 수 있으며, 및/또한 다른 동작들과 동시에 일어날 수도 있으므로, 상기한 동작의 순서에 국한되지는 않는 것으로 이해되어야 한다. 예를 들어, 본 기술분야의 전문가라면, 방법들을 상태도(state diagram)와 같이 서로 연관된 일련의 상태 또는 이벤트로서 표현할 수도 있음을 이해할 것이다. 또한, 본 발명에 따른 방법을 구현함에 있어서 반드시 예시된 모든 동작을 수행할 필요는 없다.
- [0103] 도 8은 개선된 스팸 검출과 관련하여 특징 생성을 도모하는 방법(800)의 일례를 나타낸 흐름도이다. 상기한 방법(800)은 단계 810에서 메시지를 수신함으로써 개시된다. 단계 820에서, 적절한 이메일 파싱 컴포넌트에 의해 적어도 메시지의 일부분이 파싱되어, 그 메시지의 출처 정보에 대응하는 특징이 생성된다. 상기한 특징은 단계 830에서 쌍으로 결합된다. 단계 840에서, 가장 유용한 특징 쌍들이 선택되며, 단계 850에서 이러한 특징 쌍들에 의해 필터가 머신 러닝 알고리즘을 이용하여 트레이닝되게 된다. 상기한 방법(800)은 상기한 필터를 트레이닝하기에 충분한 정도로 원하는 만큼 반복 수행된다.
- [0104] 메시지의 출처 정보로부터 유도되는 정보는 많지만, 이들 특징들 중에는 정당한 메일과 스팸을 구분함에 있어서 다른 것들에 비해 보다 유용한 것들이 있다. 특히, IP 어드레스와 그 관련 서브넷 등과 같은 특징들은 스팸머들이 수정 또는 속임수를 부리기가 매우 어렵다. 따라서, 정당한 사용자의 경우, 이들 특징은 발신자의 표시된 머신 네임 및/또는 발신자의 표시된 타임존 등과 같은 다른 특징들과 일치해야 한다. 따라서, 이들 특징 쌍들을 조사하는 경우, 각 특징 쌍들이 일치한다는 것은 해당 메시지가 (예컨대 스팸이 아닌) 정당한 가능성이 보다 높음을 나타낸다. 반대로, 이들 쌍들이 일치하지 않는다면 해당 메시지가 스팸일 가능성이 높음을 나타낸다.
- [0105] 도 9는 본 발명에 따른 도 8에 도시된 트레이닝 필터를 채용하는 방법(900)의 일례에 대한 흐름도이다. 특히, 상기한 방법(900)은 메시지를 수신하는 단계 910; 하나 이상의 출처 특징을 생성하는 단계 920; 및 이들 쌍들을 짝짓기하여 보다 유용한 특징 쌍을 획득하는 단계 930을 포함한다. 단계 940에서, 상기한 특징 쌍들은 머신 러

닝 필터에 패스되어 특정 메시지의 스팸성 여부를 판정한다. 단계 950에서, 해당 메시지의 스팸성에 대해 상기한 머신 러닝 시스템으로부터의 판정을 얻게 된다. 예를 들어, 상기한 판정은 해당 메시지가 스팸일 가능성에 대한 확률 형태로 주어질 수 있다.

[0106] 도 10에 도시된 바와 같이 스팸 검출을 개선하기 위한 추가의 특징이 생성될 수 있다. 도 10은, 방법(1000)의 일례에 대한 흐름도를 도시한 것으로서, 상기한 방법(1000)은 하나 또는 그 이상의 메시지를 수신하는 단계 1010; 메시지 및/또는 제목란의 텍스트를 담보하여 최대 주행길이 n 까지의 문자열 각각에 대한 특징을 생성하는 단계 1020; 하위길이의 문자열 각각에 대한 특징을 생성하는 단계 1030을 포함한다.

[0107] 또한, 단계 1040에서, N -그램의 위치[예를 들어, 제목란 및/또는 메시지 본문의 시작, 끝, 중간]에 기초하여 문자 N -그램의 특징을 생성할 수 있다. 단계 1050에서, 상기한 제목란 및/또는 메시지 본문의 중간과 비교하여 그 시작 및/또는 끝에서 문자열이 나타날 수 있으므로, 상기한 문자열의 상대적 엔트로피에 관련된 특징도 생성할 수 있다. 단계 1050에서, 문자열에 대한 높은 엔트로피 및 문자별 엔트로피(예컨대 평균 엔트로피)를 판정하여 특징으로서 채용할 수도 있다. 마지막으로, 단계 1060에서 상기한 특징들을 이용하여 머신 러닝 필터의 트레이닝을 수행한다. 상기한 방법(1000)은 원하는 수의 메시지, 엔트로피 이벤트 및/또는 문자열에 대해 실질적으로 트레이닝될 때까지 반복 수행될 수 있다.

[0108] 도 11은 스팸 검출을 실현하기 위하여 도 10에 따라 트레이닝된 필터를 채용하는 방법(1100)의 일례에 대한 흐름도이다. 상기한 방법(1100)은 메시지를 수신하는 단계 1110; (예컨대 복수의 길이로 된) 유효 문자열 리스트에 일치 및/또는 불일치하는 문자열을 찾기 위해 상기한 메시지의 적어도 일부를 스캔하는 단계 1120; 및 상기한 유효 문자열 리스트에서 발견되지 않은 문자열을 포함하여, 메시지 및/또는 제목란의 문자열의 적어도 일부에 대한 엔트로피를 검출하는 단계 1130을 포함한다. 단계 1140에서, 검출된 이벤트들은 특징으로서 사용되며 머신 러닝 필터에 패스된다. 단계 1150에서, 상기한 머신 러닝 시스템으로부터 상기한 메시지의 스팸성 여부에 대한 판정을 획득한다.

[0109] 도 12 및 도 13을 참조하면, 본 발명의 또다른 태양에 따른 머신 러닝 알고리즘에서 사용하기 위한 개선된 특징 생성을 가능케 하는 프로세스(1200 및 1300)의 일례에 대한 흐름도가 각각 도시되어 있다. 먼저, 상기한 방법(1200)은 하나 또는 그 이상의 메시지를 수신하는 단계 1210; 및 헤더 라인을 파싱 및 분석하여 특징을 생성하는 단계 1220을 포함한다. 또한, 단계 1230에서는, 선택적으로, 메시지 및/또는 특징의 사이즈(예컨대 메시지 사이즈, 디스플레이 내임 길이, 제목란의 길이)에 관련된 특징이 생성될 수도 있다. 단계 1240에서, 메시지 내의 임의의 이미지에 대해서는, 우선, 그 사이즈, 위치(메시지 내부 또는 외부 링크) 및/또는 수량을 분석하여 추가 특징을 생성할 수 있다. 단계 1250에서는, 생성된 거의 대부분의 특징들을 머신 러닝 시스템을 이용하여 필터의 트레이닝에 사용한다.

[0110] 도 12에 도시된 트레이닝 필터는 도 13에 예시된 프로세스(1300)에 기재된 바와 같이 새로운 메시지에도 적용될 수 있다. 단계 1310에서, 하나 또는 그 이상의 메시지가 수신된다. 단계 1320에서, 메시지에서부터 헤더 특징을 파싱해낸다. 단계 1320 및 단계 1330에서는, 선택적으로, 메시지 및/또는 특징 사이즈 및/또는 이미지 특성에 대응하는 특징들을 상기한 메시지에서부터 각각 파싱할 수도 있다. 단계 1340에서, 이들 특징들은 머신 러닝 필터에 패스되거나 또는 조사되게 된다. 단계 1350에서는, 적어도 부분적으로 파싱된 특징들에 기초하여 해당 메시지의 스팸성에 대한 정도(amount) 또는 확률을 나타내는 판정을 획득한다.

[0111] 본 발명의 다양한 태양에 대한 이해를 도모하기 위하여, 도 14 및 하기의 설명을 통해 본 발명의 각종 태양이 구현될 수 있는 적합한 오퍼레이팅 환경(1410)에 대해 간략하게 설명하기로 한다. 본 발명에서는 하나 또는 그 이상의 컴퓨터 또는 기타 장치에 의해 실행되는 프로그램 모듈과 같은 컴퓨터 실행가능 명령문의 일반 형태로 기술하고 있지만, 본 기술분야의 전문가라면 다른 프로그램 모듈과 결합한 형태 및/또는 하드웨어와 소프트웨어의 결합물로서 구현될 수도 있음을 이해할 것이다.

[0112] 그러나, 일반적으로 프로그램 모듈은 특정 태스크를 수행하거나 특정 데이터 타입을 구현하는 루틴, 프로그램, 오브젝트 컴포넌트, 데이터 구조 등을 포함한다. 상기한 오퍼레이팅 환경(1410)은 적합한 오퍼레이팅 환경의 일례일 뿐이며, 본 발명의 기능이나 용도 범위에 대한 어떠한 제한을 두기 위한 것은 아니다. 본 발명에 사용하기에 적합한 다른 알려져 있는 컴퓨터 시스템, 환경 및/또는 구성에는 개인용 컴퓨터, 핸드헬드 또는 랩탑 디바이스, 멀티프로세서 시스템, 마이크로프로세서-기반 시스템, 프로그래머블 가전제품, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 및 상기한 시스템 또는 디바이스를 구비하는 분산형 컴퓨팅 환경 등을 포함하며, 이것에 국한되지는 않는다.

- [0113] 도 14를 참조하면, 본 발명의 각종 태양을 구현하기 위한 환경(1410)의 일례를 나타낸 것으로서, 상기한 환경(1410)은 컴퓨터(1412)를 포함한다. 상기한 컴퓨터(1412)는 프로세싱 유닛(1414), 시스템 메모리(1416) 및 시스템 버스(1418)를 포함한다. 상기한 시스템 버스(1418)는 이것에 국한되지는 않지만 시스템 메모리(1416)를 포함하는 시스템 구성요소들을 상기한 프로세싱 유닛(1414)에 연결시킨다. 상기한 프로세싱 유닛(1414)은 임의의 가용한 프로세서일 수 있다. 듀얼 마이크로프로세서 및 기타의 멀티프로세서 구조도 상기한 프로세싱 유닛(1414)으로서 이용될 수 있다.
- [0114] 시스템 버스(1418)는 임의의 이용가능한 버스 구조를 이용하는 메모리 버스 또는 메모리 컨트롤러, 주변 버스 또는 외부 버스 및/또는 로컬 버스를 포함하는 임의의 종류의 버스 구조로 구성될 수 있으며, 상기한 이용가능한 버스 구조로는, 이것에 국한되지는 않지만, 11비트 버스, ISA(Industrial Standard Architecture), MSA(Micro-Channel Architecture), EISA(Extended ISA), IDE(Intelligent Drive Electronics), VLB(VESA Local Bus), PCI(Peripheral Component Interconnect), USB(Universal Serial Bus), AGP(Advanced Graphics Port), PCMCIA(Personal Computer Memory Card International Association bus) 및 SCSI(Small Computer Systems Interface)가 포함된다.
- [0115] 상기한 시스템 메모리(1416)는 활성 메모리(1420) 및 불활성 메모리(1422)를 포함한다. 기동시와 같이, 컴퓨터(1412) 내의 엘리먼트들 사이의 정보의 교환을 위한 기본 루틴을 포함하고 있는 BIOS(basic input/output system)는 불활성 메모리(1422)에 저장되어 있다. 상기한 불활성 메모리(1422)의 예로는, 이것에 국한되지는 않지만, ROM(read only memory), PROM(programmable ROM), EPROM(electrically programmable ROM) 또는 플래시 메모리를 포함한다. 활성 메모리(1420)는 외부 캐시 메모리로서 기능하는 RAM(random access memory)를 포함한다. 상기한 RAM의 예로는, 이것에 국한되지는 않지만, SRAM(synchronous RAM), DRAM(Dynamic RAM), SDRAM(synchronous DRAM), DDR SDRAM(double data rate SDRAM), ESDRAM(enhanced SDRAM), SLDRAM(Synchlink DRAM) 및 DRRAM(direct Rambus RAM) 등과 같은 형태로 이용된다.
- [0116] 또한, 상기한 컴퓨터(1412)는 착탈식/비착탈식, 활성/불활성 컴퓨터 스토리지 매체를 포함한다. 도 14는 디스크 스토리지(1424)를 예시하고 있다. 상기한 디스크 스토리지(1424)는, 이것에 국한되지는 않지만, 자기 디스크 드라이브, 플로피 디스크 드라이브, 테이프 드라이브, Jaz 드라이브, Zip 드라이브, LS-100 드라이브, 플래시 메모리 카드 또는 메모리 스틱을 포함한다. 또한, 디스크 스토리지(1424)는 개별 스토리지 매체를 포함할 수도 있으며, 또는 이것에 국한되지는 않지만, CD-ROM, CD-R 드라이브, CD-RW 드라이브, DVD-ROM 등과 같은 광학 디스크 드라이브를 포함하는 다른 스토리지 매체와 결합될 수도 있다. 상기한 디스크 스토리지 디바이스(1424)와 시스템 버스(1418) 사이의 접속을 제공하기 위하여, 인터페이스(1426)로서 일반적으로 착탈식 또는 비착탈식 인터페이스가 사용된다.
- [0117] 도 14는 적합한 오퍼레이팅 환경(1410)에서의 사용자와 기본 컴퓨터 리소스 사이의 중재자로서 기능하는 소프트웨어를 기술하고 있다. 이러한 소프트웨어는 오퍼레이팅 시스템(1428)을 포함한다. 상기한 오퍼레이팅 시스템(1428)은 디스크 스토리지(1424)에 저장될 수 있으며, 컴퓨터 시스템(1412)의 리소스를 제어 및 할당하도록 동작한다. 그러면, 시스템 애플리케이션(1430)은 시스템 메모리(1416) 또는 디스크 스토리지(1424) 중 어느 하나에 저장된 프로그램 모듈(1432) 및 프로그램 데이터(1434)를 통하여 상기한 오퍼레이팅 시스템(1428)에 의한 리소스 관리를 이용한다. 본 발명은 각종의 오퍼레이팅 환경에서 구현될 수 있으며, 결합된 형태의 오퍼레이팅 환경에서도 구현될 수 있는 것으로 이해되어야 한다.
- [0118] 사용자는 입력 장치(1436)를 통해 상기한 컴퓨터(1412)에 정보 또는 커맨드를 입력한다. 상기한 입력 장치(1436)의 예로는, 이것에 국한되지는 않지만, 마우스, 트랙볼, 스타일러스, 터치패드, 키보드, 마이크로폰, 조이스틱, 게임패드, 위성 수신기, 스캐너, TV 튜너 카드, 디지털 카메라, 디지털 비디오 카메라, 웹 카메라 등을 들 수 있다. 이들 및 기타의 입력 장치는 인터페이스 포트(1438)를 매개로 시스템 버스(1418)를 통해 프로세싱 유닛(1414)에 접속되어 있다. 인터페이스 포트(1438)로는 예를 들어 시리얼 포트, 패러럴 포트, 게임 포트 및 USB(universal serial bus)를 들 수 있다. 출력 장치(1440)는 일부 입력 장치(1436)와 동일한 타입의 포트를 사용할 수도 있다. 따라서, 예를 들어 USB 포트는 컴퓨터(1412)에 입력을 제공할 뿐만 아니라, 상기한 컴퓨터(1412)로부터 출력 장치(1440)에 정보를 출력하는데 이용할 수도 있다. 출력 어댑터(1442)는 출력 장치(1440)들 중에는 특수한 어댑터를 필요로 하는 모니터, 스피커 및 프린터 등의 출력 장치(1440)가 있기 때문에 제공되는 것이다. 상기한 출력 어댑터(1442)의 예로는, 이것에 국한되지는 않지만, 출력 장치(1440)와 시스템 버스(1418) 사이의 접속 수단을 제공하는 비디오 및 사운드 카드를 들 수 있다. 또한, 다른 장치 및/또는 장치 시스템들은 리모트 컴퓨터(1444)와 같이 입력 및 출력 능력을 모두 구비하고 있다.

[0119] 상기한 컴퓨터(1412)는 리모트 컴퓨터(1444)와 같은 하나 또는 그 이상의 컴퓨터에 논리 접속된 네트워크 환경에서 동작할 수 있다. 상기한 리모트 컴퓨터(1444)는 개인용 컴퓨터, 서버, 라우터, 네트워크 PC, 워크스테이션, 마이크로프로세서 기반 어플라이언스, 피어 디바이스 또는 기타의 공통 네트워크 노드 등일 수 있으며, 일반적으로는 컴퓨터(1412)와 관련하여 전송한 엘리먼트의 다수 또는 모두를 포함하고 있다. 간단히 설명하면, 도면에서는 메모리 스토리지 디바이스(1446)만을 리모트 컴퓨터(1444)와 함께 도시하고 있다. 리모트 컴퓨터(1444)는 네트워크 인터페이스(1448)를 통해 컴퓨터(1412)에 논리접속되며 통신 접속(1450)을 매개로 물리적으로 접속된다. 네트워크 인터페이스(1448)는 LAN(local area network) 및 WAN(wide area network)과 같은 통신 네트워크를 포함한다. LAN 기술은 FDDI(Fiber Distributed Data Interface), CDDI(Copper Distributed Data Interface), Ethernet/IEEE 1102.3, 토큰링/IEEE 1102.5 등을 포함한다. WAN 기술의 예로는, 이것에 국한되지는 않지만, 포인트간 링크, ISDN 및 그 변형(variation) 등의 회로 스위칭 네트워크, 패킷 스위칭 네트워크 및 DSL(Digital Subscriber Lines)을 포함한다.

[0120] 상기한 통신 접속(1450)은 상기한 네트워크 인터페이스(1448)를 버스(1418)에 접속시키는데 이용된 하드웨어/소프트웨어이다. 도면에서는 상기한 통신 접속(1450)을 컴퓨터(1412)의 내측에 도시하고 있지만, 컴퓨터(1412)의 외측에 배치할 수도 있다. 네트워크 인터페이스(1448)의 접속에 필요한 하드웨어/소프트웨어의 예로는, 일반 전화급 모뎀, 케이블 모뎀 및 DSL 모뎀, ISDN 어댑터 및 이더넷 카드 등의 모뎀과 같이 내부 및 외부 기술을 포함한다.

[0121] 이상 본 발명의 실시예들에 대해 설명하였다. 본 발명을 설명함에 있어서, 모든 가능한 컴포넌트 또는 방법의 결합예를 설명하는 것은 불가능하지만, 본 기술분야의 전문가라면 본 발명에 대한 추가적인 결합예 및 치환예도 가능함을 이해할 것이다. 따라서, 본 발명은 청구된 특허청구범위의 사상과 범주에 포함되는 가능한 모든 변형, 수정 및 치환을 포함하는 것으로 이해되어야 한다. 또한, 본 명세서 및 특허청구범위에서 사용되고 있는 "포함한다"는 용어는, 특허청구범위에서 "구비한다"를 사용하는 경우 그 대체 단어인 것으로 해석되는 것과 마찬가지로, 포괄적인 것을 이해되어야 한다.

발명의 효과

[0122] 전송한 바와 같이, 본 발명에 따르면 스팸머들이 회피 및 복제하기 어려운 특성을 나타내는 스팸 메시지 식별 시스템 및 방법이 제공된다.

도면의 간단한 설명

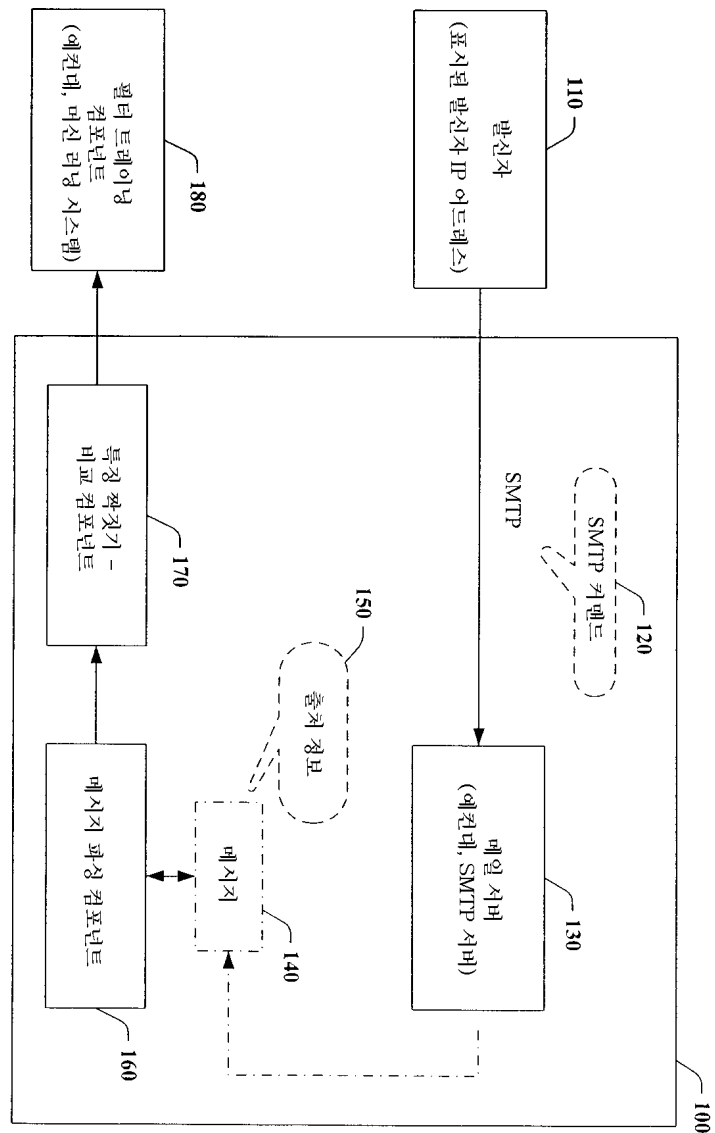
- [0001] 도 1은 본 발명의 태양에 따른 스팸 방지를 도모하는 시스템을 개략적으로 나타낸 블록도.
- [0002] 도 2는 본 발명의 태양에 따른 HELO 커맨드로부터의 특징 분기(break down)를 개략적으로 나타낸 블록도.
- [0003] 도 3은 본 발명의 태양에 따른 MAIL FROM 커맨드로부터의 특징 분기를 개략적으로 나타낸 블록도.
- [0004] 도 4는 본 발명의 태양에 따른 DATA 커맨드로부터의 특징 분기를 개략적으로 나타낸 블록도.
- [0005] 도 5는 본 발명의 태양에 따른 Received 라인 및 Message From 라인의 쌍으로부터의 특징 분기를 개략적으로 나타낸 블록도.
- [0006] 도 6은 본 발명의 태양에 따른 한 주행길이의(a run of) 문자열 및/또는 문자열 및/또는 이들 문자열의 엔트로피에 관련된 특징을 생성하기 위한 시스템을 개략적으로 나타낸 블록도.
- [0007] 도 7은 본 발명의 태양에 따른 메시지 헤더 콘텐츠 및/또는 사이즈 관련 특징 및/또는 메시지 내에 존재하는 이미지에 관련된 특징을 생성하기 위한 시스템을 개략적으로 나타낸 도면.
- [0008] 도 8은 본 발명의 태양에 따른 필터의 트레이닝을 위한 특징 쌍들을 포함하는 특징을 생성하기 위한 방법의 예를 나타낸 순서도.
- [0009] 도 9는 스팸 및/또는 스팸성 메시지의 식별에 도 8의 트레이닝 필터를 이용하기 위한 방법의 예를 나타낸 순서도.
- [0010] 도 10은 본 발명의 태양에 따른 필터의 트레이닝에 이용하기 위한 한 주행길이의 문자열 및/또는 상기한 문자열의 엔트로피에 기초하여 특징을 생성하기 위한 방법의 예를 나타낸 순서도.
- [0011] 도 11은 스팸 및/또는 스팸성 메시지의 식별에 도 10의 트레이닝 필터를 이용하기 위한 방법의 예를 나타낸 순

서도.

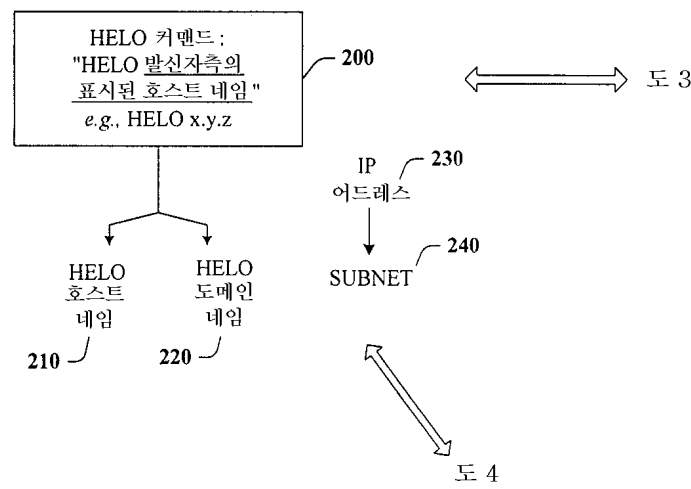
- [0012] 도 12는 본 발명의 태양에 따른 필터의 트레이닝에 이용하기 위한 특징을 생성하기 위한 방법의 예를 나타낸 순서도.
- [0013] 도 13은 스팸 및/또는 스팸성 메시지의 식별에 도 12의 트레이닝 필터를 이용하기 위한 방법의 예를 나타낸 순서도.
- [0014] 도 14는 본 발명에 따른 통신 환경의 예를 개략적으로 나타낸 블록도.
- [0015] <도면의 주요 부분에 대한 부호의 설명>
- [0016] 100: 전체 시스템
- [0017] 110: 발신자
- [0018] 120: SMTP 커맨드
- [0019] 130: 메일 서버
- [0020] 140: 메시지
- [0021] 150: 출처 정보
- [0022] 160: 메시지 파싱 컴포넌트
- [0023] 170: 특징 짚기-비교 컴포넌트
- [0024] 180: 필터 트레이닝 컴포넌트

도면

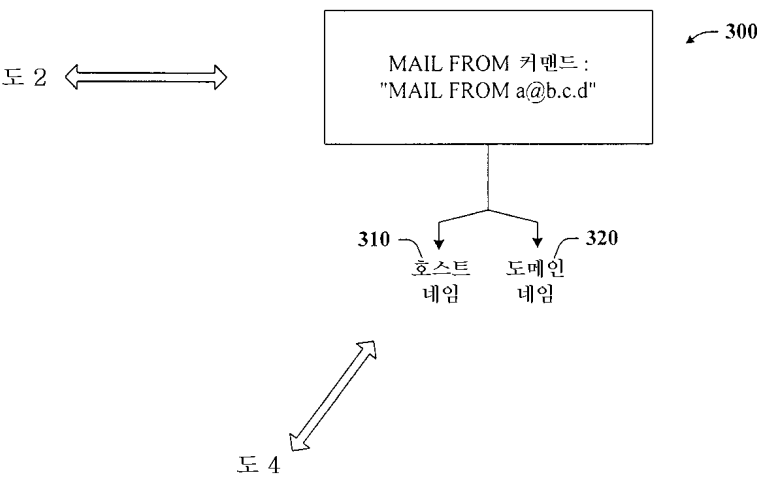
도면1



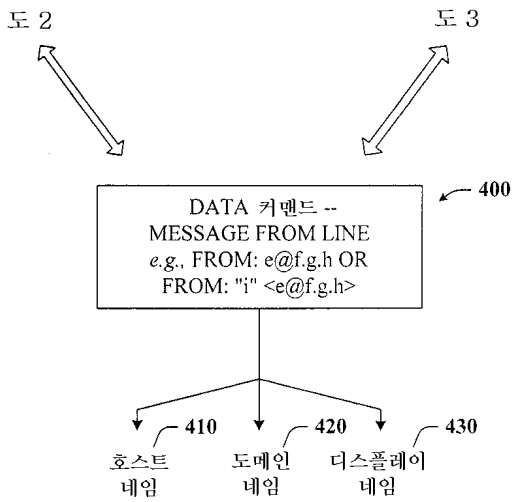
도면2



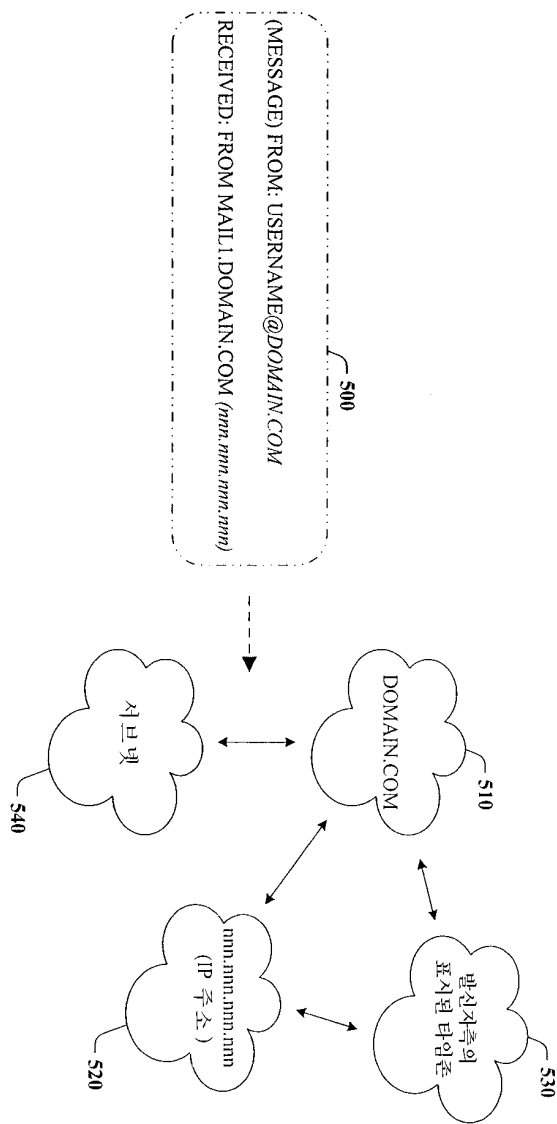
도면3



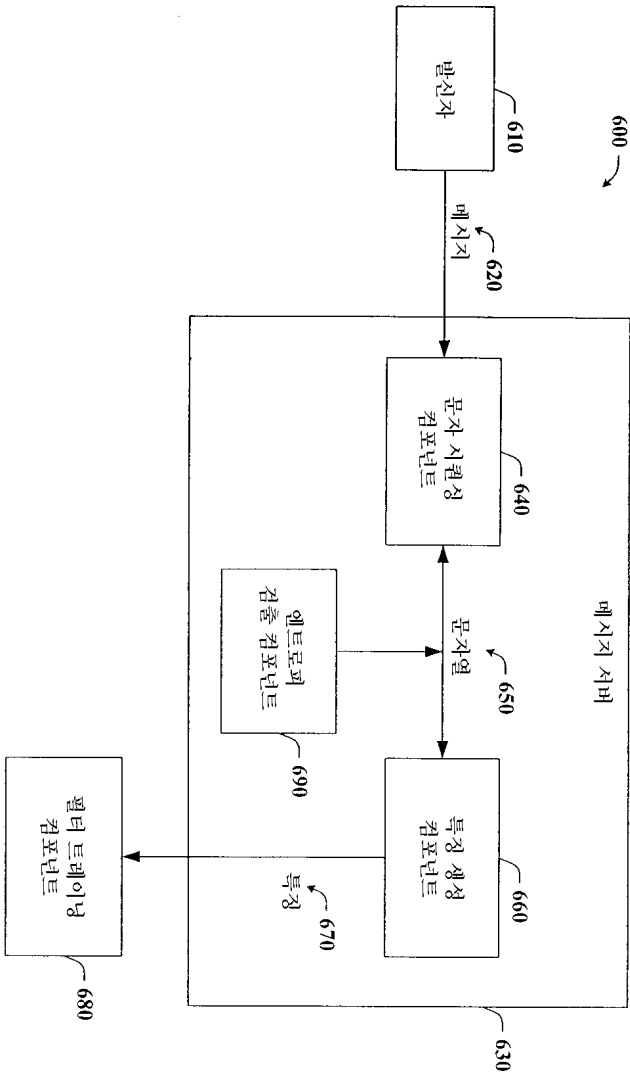
도면4



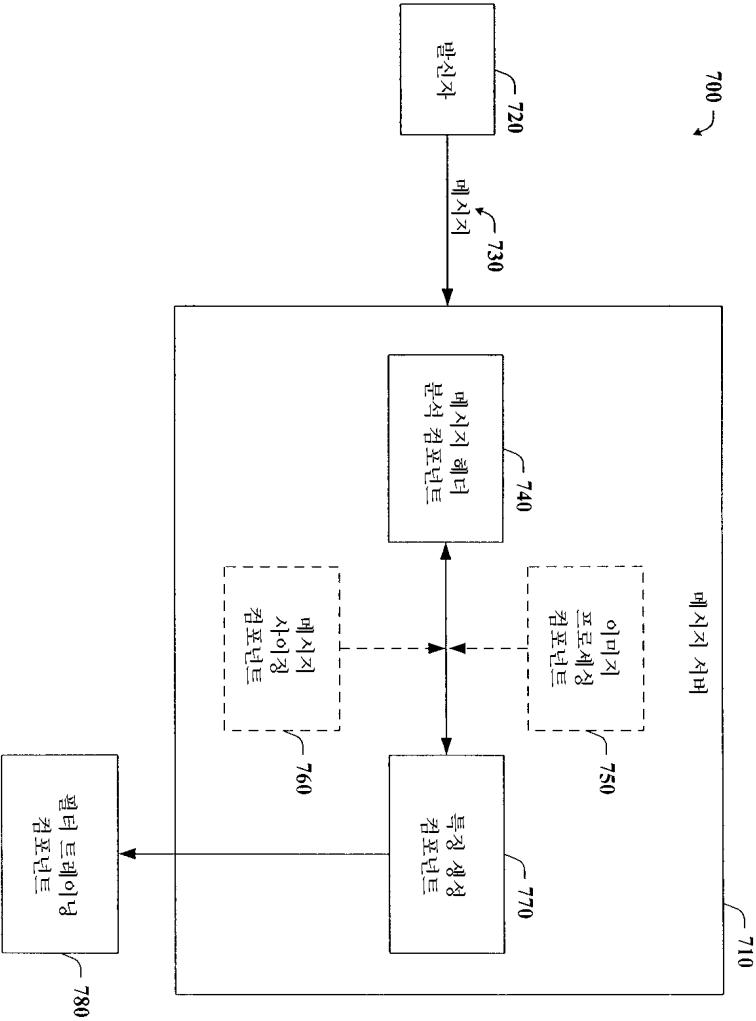
도면5



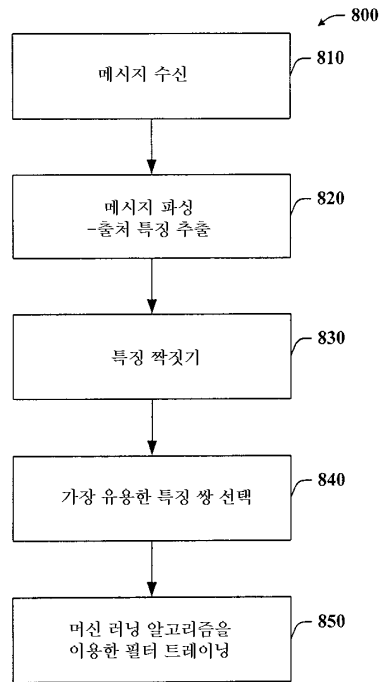
도면6



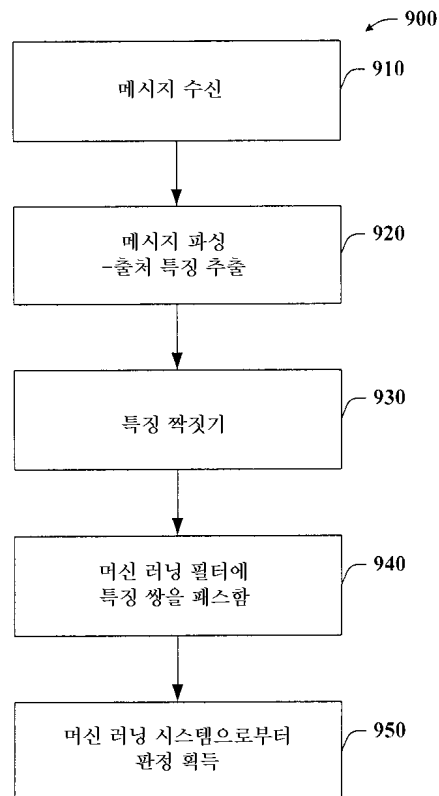
도면7



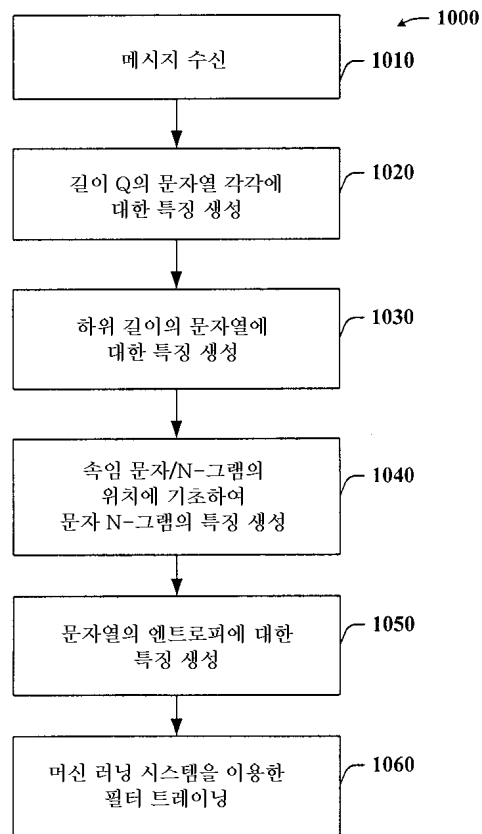
도면8



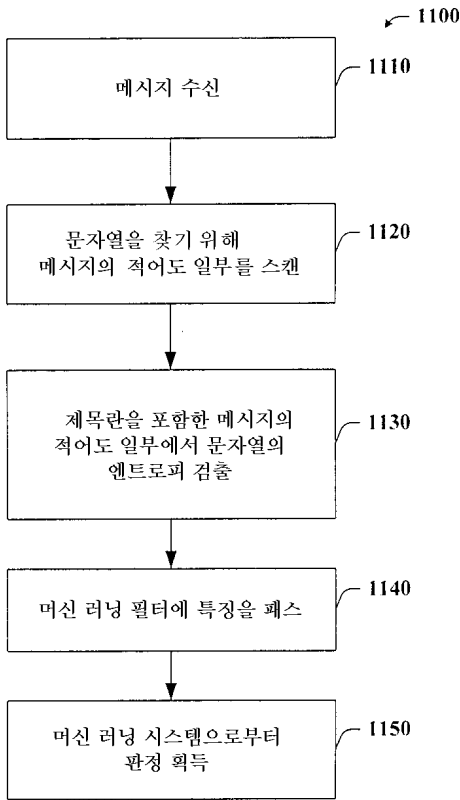
도면9



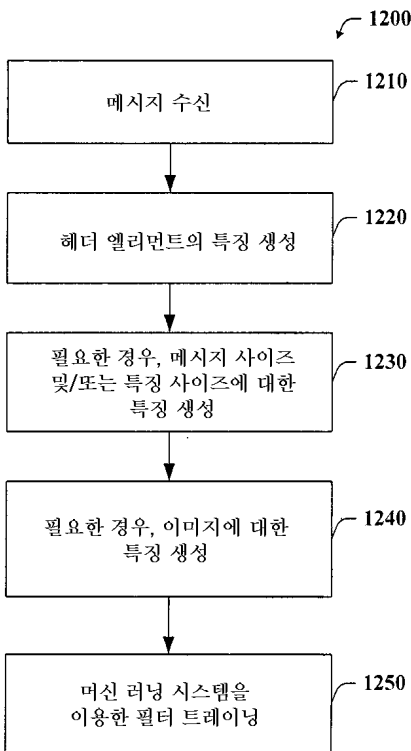
도면10



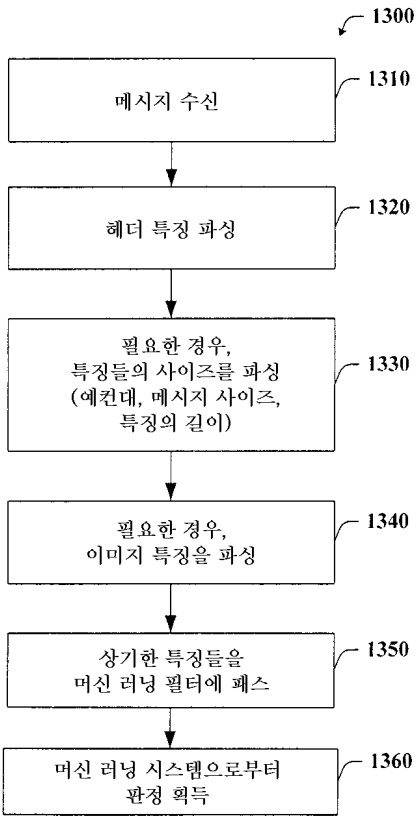
도면11



도면12



도면13



도면14

