(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년10월25일
(11) 등록번호 10-1076909
(24) 등록일자 2011년10월19일

(51) Int. Cl.
G06F 15/00 (2006.01) *G06F 9/44* (2006.01)
(21) 출원번호 10-2005-0057153
(22) 출원일자 2005년06월29일
심사청구일자 2010년06월22일
(65) 공개번호 10-2006-0048713
(43) 공개일자 2006년05월18일
(30) 우선권주장
10/879,626 2004년06월29일 미국(US)
(56) 선행기술조사문헌
US20040064446 A1
US20030236883 A1

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
하저, 디렉 엠.
미국 98052 워싱턴주 레드몬드 원 마이크로소프트 웨이마이크로소프트 코포레이션 내
머피, 엘리사 이.
미국 98052 워싱턴주 레드몬드 원 마이크로소프트 웨이마이크로소프트 코포레이션 내
(뒷면에 계속)
(74) 대리인
제일특허법인

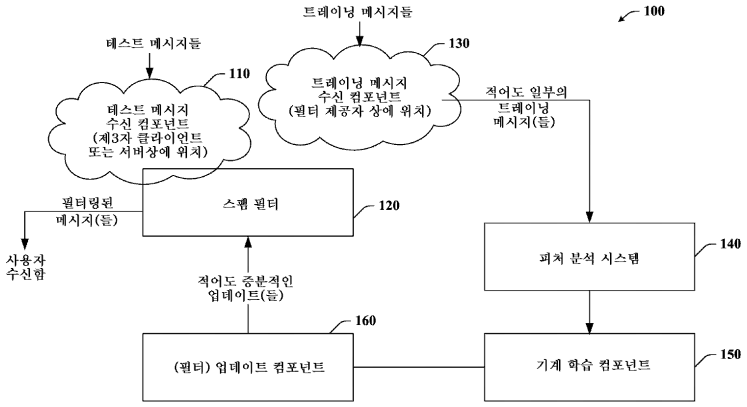
전체 청구항 수 : 총 15 항 심사관 : 복진요

(54) 증분적 안티 스팸 룩업 및 업데이트 시스템 및 서비스 방법

(57) 요 약

본 발명은 준 실시간 또는 실시간으로 스팸 필터들을 증분적으로 업데이트하는 것을 용이하게 하는 고유한 시스템 및 방법을 제공한다. 증분적 업데이트들은 부분적으로 차이 학습에 의해 생성된다. 차이 학습은 새로운 스팸 필터를 새로운 데이터에 기초하여 트레이닝하는 단계와, 새로운 스팸 필터와 기존의 스팸 필터 사이의 차이점들을 찾는 단계를 수반한다. 차이점들은 적어도 부분적으로는 파라미터 변경 사항들(상기 두 필터들 사이의 피처의 가중치 변화)의 절대값을 비교함으로써 판정된다. 파라미터들의 빈도와 같은 다른 요인들도 이용될 수 있다. 또한, 특정한 피처들 또는 메시지들에 대하여 이용 가능한 업데이트들이 하나 이상의 룩업(lookup) 테이블들 또는 데이터베이스들을 이용하여 룩업될 수 있다. 증분적 및/또는 피처 특정 업데이트들이 이용될 수 있는 경우, 이들은 예컨대 클라이언트 등에 의해 다운로드될 수 있다. 증분적 업데이트들은 자동으로 제공되거나 또는 클라이언트나 서버의 설정들에 따른 요청에 의해 제공될 수 있다.

대 표 도



(72) 발명자

헐텐, 지퍼레이 제이.

미국 98052 워싱턴주 레드몬드 원 마이크로소프트 웨이마이크로소프트 코포레이션 내

굿맨, 조수아 티.

미국 98052 워싱턴주 레드몬드 원 마이크로소프트 웨이마이크로소프트 코포레이션 내

로운스웨이트, 로버트 엘.

미국 98052 워싱턴주 레드몬드 원 마이크로소프트 웨이마이크로소프트 코포레이션 내

특허청구의 범위

청구항 1

안티 스팸 업데이트 시스템에 있어서,

스팸 및 양호한 메시지들을 구분하도록 트레이닝되는 기존 스팸 필터(existing spam filter); 및

업데이트 컴포넌트를 포함하고,

상기 업데이트 컴포넌트는,

상기 기존 스팸 필터의 하나 이상의 피처들(features)에 각각 연관된 하나 이상의 이전의 가중치들(old weights)을 업데이트된 정보를 갖는 새로운 필터의 하나 이상의 피처들에 각각 연관된 하나 이상의 가중치들과 비교하고,

하나 이상의 피처들에 각각 연관된 하나 이상의 새로운 가중치들을 추가하는 것과 상기 하나 이상의 피처들에 각각 연관된 하나 이상의 이전의 가중치들을 소정의 변화량(some amount of change)을 나타내는 하나 이상의 피처들에 각각 연관된 하나 이상의 가중치들로 대체하는 것 중 적어도 하나를 행하며,

상기 업데이트 컴포넌트는 적어도 부분적으로 기계 학습 컴포넌트(machine learning component)를 사용하여 트레이닝되는, 안티 스팸 업데이트 시스템.

청구항 2

제1항에 있어서,

상기 안티 스팸 업데이트 시스템은 클라이언트 또는 서버 중 적어도 하나에서 실행되는, 안티 스팸 업데이트 시스템.

청구항 3

제1항에 있어서,

상기 업데이트 컴포넌트는 매치(match) 기반 또는 해시(hash) 기반의 데이터를 사용하여 트레이닝되는, 안티 스팸 업데이트 시스템.

청구항 4

제1항에 있어서,

하나 이상의 업데이트들이 상기 기존 스팸 필터에 제공되기에 앞서 클라이언트 또는 서버가 업데이트들을 수신하기 위하여 지불하였는지 여부를 판정하는 지불 검증 컴포넌트(payment verification component)를 더 포함하는 안티 스팸 업데이트 시스템.

청구항 5

제1항에 있어서,

상기 업데이트 컴포넌트는 상기 업데이트된 정보를 복수의 IP 주소들을 통하여 상기 기존 스팸 필터에 제공하여 DOS(Denial-Of-Service) 공격들을 완화시키는, 안티 스팸 업데이트 시스템.

청구항 6

제1항에 있어서,

상기 변화량은 임계치 값을 충족하는 양인, 안티 스팸 업데이트 시스템.

청구항 7

제1항에 있어서,

상기 업데이트된 정보는 하나 이상의 피처들의 하나 이상의 피처 변화들에 각각 연관된 상기 하나 이상의 새로운 가중치들에 대응되는 하나 이상의 증분적 업데이트 파일들(incremental update files)을 포함하는, 안티 스팸 업데이트 시스템.

청구항 8

제1항에 있어서,

상기 메시지의 피처는 IP 주소 및 URL 중 적어도 하나를 포함하는, 안티 스팸 업데이트 시스템.

청구항 9

제1항에 있어서,

상기 기존 스팸 필터와 상기 새로운 필터 사이의 하나 이상의 피처들의 하나 이상의 피처 변화들에 각각 연관된 최소화된 개수의 상기 하나 이상의 새로운 가중치들을 가지는 기계 학습 스팸 필터들(machine learning spam filters)을 구축하여 증분적 업데이트 크기들을 최소화하는 것을 용이하게 하는 컴포넌트를 더 포함하는 안티 스팸 업데이트 시스템.

청구항 10

제1항에 있어서,

상기 업데이트 컴포넌트는 복수의 상이한 업데이트들을 상기 기존 스팸 필터에 순차적으로 적용하는, 안티 스팸 업데이트 시스템.

청구항 11

제1항에 있어서,

상기 업데이트 컴포넌트는 복수의 업데이트들의 적어도 부분 집합을 병합(merges)하여 상기 업데이트된 정보의 다운로드 효율을 향상시키는, 안티 스팸 업데이트 시스템.

청구항 12

제1항에 있어서,

상기 업데이트 컴포넌트는 하나 이상의 피처들에 각각 연관된 피처에 특정되는 업데이트들(feature-specific updates)을, 적어도 하나의 피처가 적어도 하나의 다른 피처에 독립적으로 업데이트되도록 독립적인 방식으로 선택적으로 제공하는, 안티 스팸 업데이트 시스템.

청구항 13

안티 스팸 질의 시스템(anti-spam query system)에 있어서,

스팸 및 양호한 메시지들을 구분하도록 트레이닝된 기존 기계 학습 스팸 필터; 및

룩업 컴포넌트(lookup component)를 포함하고,

상기 룩업 컴포넌트는,

상기 기존 기계 학습 스팸 필터로부터 질의를 수신하고 - 상기 질의는 상기 기존 기계 학습 스팸 필터가 스팸 또는 양호한 것으로서 분류할 수 없는 메시지의 적어도 하나의 피처를 포함함 -,

상기 수신된 적어도 하나의 피처에 기초하여, 상기 적어도 하나의 피처에 각각 연관된 적어도 하나의 새로운 가중치가 상기 기존 기계 학습 스팸 필터에 추가될지 또는 상기 기존 기계 학습 스팸 필터의 적어도 하나의 이전의 가중치를 대체할지를 판정하고 - 상기 적어도 하나의 이전의 가중치는 상기 적어도 하나의 피처에 각각 연관됨 -,

상기 기존 기계 학습 스팸 필터를 업데이트하기 위하여 상기 적어도 하나의 피처에 각각 연관된 상기 적어도 하나의 새로운 가중치를 제공하는, 안티 스팸 질의 시스템.

청구항 14

제13항에 있어서,

상기 룩업 컴포넌트는 증분적 룩업 질의들(incremental lookup queries)을 파일에 기록하거나 또는 상기 증분적 룩업 질의들을 디스크 상에 저장하고 상기 증분적 룩업 질의들을 메모리에서 결합시키는, 안티 스팸 질의 시스템.

청구항 15

제13항에 있어서,

상기 룩업 컴포넌트는,

피처들 및 관련 가중치들의 집합과 트레이닝 중에 생성되는 하나 이상의 모델들 또는 필터들을 포함하는 백엔드(back-end) 데이터베이스; 및

상기 백엔드 데이터베이스와 스팸 필터 사이의 통신을 전달하는 미들웨어 계층(middle-ware layer); 및

사전 정의된 또는 자동적인 빈도로 상기 미들웨어 계층을 호출하여 가장 최근의 업데이트된 모델을 획득하고, 온라인 모델을 로컬적으로(locally) 저장된 모델 파일과 병합하는 스팸 필터를 포함하는, 안티 스팸 질의 시스템.

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

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

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

[0016] 본 발명은 적법한 정보(예컨대 양호한 메일)와 원하지 않는 정보(예컨대 정크 메일)를 식별하기 위한 시스템들 및 방법들에 관한 것이고, 보다 구체적으로는 메시지 처리 중에 준 실시간 또는 실시간 업데이트를 기존의 트레이닝된 스팸 필터에게 제공하는 것에 관한 발명이다.

[0017] 인터넷과 같은 글로벌 통신 네트워크의 출현은 막대한 수의 잠재적 고객들에게 접근할 수 있는 상업적인 기회를 제공하게 되었다. 전자 메시징, 특히 전자 메일("이메일")은 원치 않는 광고들 및 판촉("스팸"이라고 일컬어짐)을 네트워크 사용자들에게 퍼뜨리기 위한 수단으로서 점차 만연되어 가고 있다.

[0018] 컨설팅 및 마켓 리서치 회사인 Radicati Group, Inc.는 2002년 8월 기준으로 20억 통의 정크 이메일 메시지들이 매일 전달되는 것으로 추정하였고, 이 수는 매 2년마다 3배가 되는 것으로 예상된다. 개인들 및 기관들(예컨대 기업체 및 정부 기관들)은 정크 메시지들에 의해 점차 불편을 겪고 있으며 종종 방해를 받는다. 이와 같이, 정크 이메일은 이제 신용 컴퓨팅에 대한 중대한 위협이 될 것이다.

발명이 이루고자 하는 기술적 과제

[0019] 정크 이메일 또는 스팸을 방해하기 위한 주요 기법은 필터링 시스템들 및/또는 방법들을 이용하는 것이다. 그러나, 스패머들은 필터들을 회피하기 위하여 자신들의 기법을 지속적으로 바꾼다. 따라서 스패머들의 기법이 변화함에 따라 필터들을 빠르게 자동으로 업데이트하고, 메시징 클라이언트들 및/또는 서버들에 의해 운영되는 최종 애플리케이션들에게 업데이트를 전파하는 것이 바람직하다.

[0020] 예컨대, 클라이언트들이 이용하는 메시징 프로그램들에는 대략 1억 개의 카피가 있을 수 있다. 또한, 새로운 스팸 필터들은 매일 생성될 수 있다. 스팸 필터들은 다소 크고 이들을 필터의 카피를 운영하는 각 클라이언트에게 분배하는 것이 매일 필요할 수 있기 때문에, 이는 클라이언트 및 필터 제공측 모두에서 금지되지 않는 경우 문제가 될 수 있다. 특히, 클라이언트들은 큰 파일들을 지속적으로 다운받을 필요가 있을 수 있으므로, 프로세서 메모리의 상당량을 소비하고 프로세싱 속도를 저하시키게 된다. 필터 제공자는 모든 사용자들 및/또는 클라이언트들의 모든 필터 카피들을 업데이트해야할 수 있기 때문에, 막대하고 비현실적인 양의 대역폭 및 서버들이 필요할 수 있다. 하루에 한 번 넘게 새로운 필터들을 제공하는 것은 완전히는 아니더라도 이러한 조건 하에서는 거의 불가능할 수 있다.

발명의 구성 및 작용

[0021] 이하에서는 본 발명의 간략한 요약을 제시하여 본 발명의 일부 태양의 기본적인 이해를 제공하고자 한다. 이러한 요약은 본 발명의 광범위한 개요가 아니다. 이는 본 발명의 주요/핵심 요소들을 식별하거나 본 발명의 범위를 기술하고자 하는 것이 아니다. 이것의 유일한 목적은 본 발명의 일부 개념들을 단순화된 형태로 제시함으로써 이후 제시되는 상세한 설명에 대한 도입 역할을 하게 하기 위함이다.

[0022] 본 발명은 실시간 또는 준 실시간 방식으로 스팸 필터들에게 부분적 또는 증분적 업데이트의 형태로 새로운 정보 또는 데이터를 제공하는 것을 용이하게 하는 시스템 및/또는 방법에 관한 것이다. 필터가 최신 정보로 업데

이트될 수 있는 준 실시간 메커니즘을 제공하는 것은 유입되는 스팸 공격들에 대한 효율적인 방어책을 제공하는 한 가지 전략일 수 있다.

[0023] 특히, 본 발명은 기존 필터에 대하여 정보에 증분적인 부분들을 전달함으로써 필터를 새로운 양호한 메시지들 및/도는 새로운 스팸에 대하여 최신의 상태로 유지하는 것을 용이하게 한다. 이는 부분적으로 차이 학습에 의해 이루어질 수 있으며, 여기서는 기존 필터의 하나 이상의 파라미터들이 새로운 필터 상의 이들 파라미터와 비교될 수 있다. 소정의 변화량을 나타내는 파라미터들은 그에 따라 업데이트되어 전체 필터의 모든 카피를 대체해야 할 필요를 완화시켜 준다. 따라서, 기존 필터와 새로운 필터 사이의 "차이점"이 기존 필터를 업데이트하도록 전달될 수 있다. 그 결과, 각 업데이트는 업데이트의 빈도에 따라 그 크기가 상대적으로 작아질 수 있다. 이는 부분적으로는 업데이트되는 정보가 주로 새로운 양호한 메시지들 또는 새로운 스팸에 기초한다는 사실에 기인하며, 시간 당 수신되는 스팸 또는 양호한 메시지들은 그리 많지 않다. 따라서, 소정의 시간 프레임에 있어서 그만큼 많은 업데이트를 수행하는 것이 스팸과의 전쟁에서 보다 효과적이고 효율적일 수 있다.

[0024] 본 발명의 일 태양에 따르면, 증분적 업데이트들은 부분적으로 서버들에 의해 결정될 수 있다. 서버들은 자신의 필터들의 어느 부분들을 업데이트할지 결정하고, 그러한 업데이트들을 획득하고, 이후 이들을 수신하기로 선택 또는 지불한 사용자들 또는 클라이언트에게 제공한다.

[0025] 본 발명의 또 다른 태양에 따르면, 증분적 업데이트들은 부분적으로 사용자 또는 클라이언트에 의해 웹 기반의 서비스를 통해 결정될 수 있다. 특히, 클라이언트는 현재의 필터가 스팸 또는 양호한 메시지인지 분류하는 데 어려움이 있다는 메시지를 수신할 수 있다. 웹 기반의 서비스는 메시지들에 관한 데이터 또는 정보, 또는 최근에 양호한 메시지들 또는 스팸인 것으로 결정된 메시지들의 피처(feature)들을 포함하는 룩업 테이블 또는 데이터베이스를 제공할 수 있다. 메시지로부터 소정의 정보를 추출함으로써 클라이언트는 웹 기반 서비스에게 질의(query)하여 자신의 필터를 위한 업데이트된 정보가 존재하는지 여부를 판정할 수 있다.

[0026] 예컨대, 클라이언트가 메시지를 수신하고 클라이언트의 필터가 그것이 스팸인지 또는 양호한 것인지를 분류하는 데 어려움을 겪는다. 클라이언트는 메시지의 일부, 예컨대 발신자의 IP 주소, 메시지 내의 URL, 또는 메시지의 해시(hash)를 추출하여 웹 기반의 룩업 서비스로부터 업데이트된 정보를 요청할 수 있다. 일례로, 질의가 웹 기반의 서비스에 제출될 수 있다. 그 대신, 또는 이에 부가하여, 클라이언트는 룩업 서비스에 의해 구축되고 최신의 정보로 유지되는 하나 이상의 룩업 테이블 또는 데이터베이스를 참조할 수 있다. 적어도 하나의 업데이트가 발견되면, 클라이언트의 필터는 그에 따라 업데이트될 수 있다. 클라이언트가 업데이트들의 시퀀스를 필요로 한다고 서비스가 결정하면, 서비스는 단순히 가장 최근의 업데이트를 제공함으로써 다운로드될 필요가 있는 업데이트들의 전체 수를 감소시킬 수 있다.

[0027] 소정의 스팸 필터에서 업데이트될 수 있는 상이한 파라미터들은 수천 개가 있을 수 있다. 이러한 필터들의 속성으로 인해, 하나의 파라미터 값에 대한 하나의 작은 변화가 거의 모든 파라미터들의 값에 소정의 변화를 야기할 수 있다. 따라서, 필터의 어느 부분을 업데이트하여 가장 효율적인 스팸 보호책을 제공할지 결정하는 방식에는 여러 가지가 있을 수 있다. 본 발명의 일 태양에 있어서, 파라미터들에 대한 절대 변화값이 조사될 수 있다. 값의 변화가 가장 큰 파라미터들이 업데이트되도록 선택될 수 있다. 그 대신, 임계 변화량(예컨대 절대값에 기초함)이 설정될 수 있다. 이러한 임계치를 초과하는 임의의 파라미터들이 업데이트될 것으로서 표지될 수 있다. 다른 요인들, 예컨대 들어오는 메시지들에 있어서 파라미터 또는 피처의 빈도 등도 함께 고려될 수 있다.

[0028] 본 발명의 다른 태양에 있어서, 증분적 업데이트들은 피처에 관하여 특유한 것일 수 있으며, 서버 또는 클라이언트의 설정(preferences)에 의해 원하는 빈도로 이루어질 수 있다. 또한, 필터들은 구 필터와 신 필터 사이의 파라미터 변화의 수를 최소화하도록 구축될 수 있다. 결과적으로, 어떠한 필터 업데이트(예컨대 데이터 파일)의 전체 크기 및 업데이트되는 파라미터들의 수는 그렇지 않은 경우보다 실질적으로 적을 수 있다.

[0029] 상술한 목적 및 관련된 목적을 달성하기 위해, 본 발명의 소정의 태양이 이하의 상세한 설명 및 첨부된 도면과 관련하여 본 명세서에 기술된다. 그러나, 이들 태양은 본 발명의 원리가 구현될 수 있는 다양한 방식들 중 일부만을 나타내는 것이며, 본 발명은 이러한 태양들 및 그 균등물을 포함하는 것이다. 본 발명의 다른 장점 및 신규한 특징들은 이하의 본 발명의 상세한 설명과 첨부된 도면을 참조함으로써 보다 분명해질 수 있다.

[0030] 이하 본 발명을 첨부된 도면을 참조하여 설명할 것이며, 여기서 동일한 참조 번호는 동일한 요소를 가리키도록 사용되었다. 이하의 상세한 설명에서, 설명을 위해 다수의 특정한 세부 사항이 본 발명의 완전한 이해를 돕기 위해 제시되었다. 그러나 본 발명이 이러한 특정 세부 사항 없이도 실시될 수 있음은 자명하다. 다른 실시예

에서, 잘 알려진 구조 및 장치들이 본 발명의 기술을 용이하게 하도록 블록도의 형태로 도시되었다.

[0031] 본원에서 사용되는 용어인 "컴포넌트" 및 "시스템"은 컴퓨터 관련 개체, 즉 하드웨어, 하드웨어와 소프트웨어의 조합, 소프트웨어, 또는 실행주인 소프트웨어를 가리키는 것이다. 예컨대, 컴포넌트는 프로세서에서 실행중인 프로세스, 프로세서, 객체, 실행 가능 명령어(executable), 실행의 쓰레드(thread of execution), 프로그램 및 컴퓨터일 수 있으나, 이에 한정되는 것은 아니다. 예컨대, 서버상에서 실행되는 애플리케이션과 서버 모두가 컴포넌트일 수 있다. 하나 이상의 컴포넌트가 프로세스 및/또는 실행의 쓰레드 내에 위치할 수 있으며, 컴포넌트는 하나의 컴퓨터에 국지화되고/되거나 둘 이상의 컴퓨터들 사이에 분산될 수 있다.

[0032] 본 발명은 기계 학습 또는 비 기계 학습 스팸 필터들에 대하여 적어도 부분적인 또는 증분적인 업데이트들을 제공하는 것과 관련하여 다양한 추론 방식 및/또는 기법을 도입할 수 있다. 본 명세서에서 사용되는 "추론"이라는 용어는 일반적으로 이벤트들 및/또는 데이터를 통해 포착된 관측값들의 집단으로부터 시스템, 환경, 및/또는 사용자의 상태들을 추단 또는 추측하는 프로세스를 가리킨다. 추론은 특정한 문맥(context) 또는 동작을 식별하는 데 사용되거나, 또는 예컨대 상태에 대한 확률 분포를 생성할 수 있다. 추론은 확률적일 수 있다. 즉 관심 상태에 대한 확률 분포를 데이터 및 이벤트들에 관한 고찰에 기초하여 계산하는 것일 수 있다. 추론은 또한 이벤트들 및/또는 데이터로부터 더 높은 레벨의 이벤트들을 구성하는 데에 사용되는 기법을 지칭할 수 있다. 이러한 추론은 이벤트들이 가까운 시간적 근접성으로 상관되는지 여부에 관계없이, 그리고 이벤트들 및 데이터가 하나 또는 여러 이벤트 및 데이터 소스들로부터 유래한 것인지에 관계없이, 관측된 이벤트들 및/또는 저장된 이벤트 데이터의 집합으로부터 새로운 이벤트들 또는 동작들을 구축하는 결과를 낳을 수 있다.

[0033] 본 발명의 다양한 태양은 기계 학습 및 비 기계 학습 필터들에 대해 적용될 수 있다. 통상적인 구현예에 있어서, 기계 학습 스팸 필터들은 기계 학습 알고리즘을 이용하여 메시지들로부터 추출되는 개개의 특성의 가중치를 계산함으로써 무엇이 양호한 메시지들과 스팸을 특징짓는지에 관한 정의를 학습한다. 메시지가 호스트 애플리케이션에 의해 수신되는 경우, 메시지 스트림은 스팸 또는 양호한 메시지 지시자들로서 가중치를 부여받는 피처들 또는 특성들에 대하여 해석(parse) 또는 조사될 수 있다. 가중치가 부여된 이들 피처는 이후 결합되어 그 메시지가 스팸인지 아닌지에 대한 전체 확률을 생성하게 된다. 메시지가 특정한 확률의 "임계치(threshold)"를 충족하는 경우, 그 메시지는 호스트 애플리케이션 세팅에 기초하여 지정된 동작을 취할 수 있다. 예컨대, 양호한 메시지들은 수신자의 수신함으로 향하고 스팸 메시지들은 특별한 폴더로 향하거나 삭제될 수 있다.

[0034] 스패머들은 지속적으로 자신들의 방법을 개변시켜 나간다. 이전의 피드백 루프들 및 기계 학습 기술을 통해, 새로운 필터들이 자동적으로 바로 생성될 수 있다. 그러나, 이러한 새로운 필터들을 실시간 또는 준 실시간으로 사용자들에게 효율적으로 전파하는 것도 마찬가지로 중요하다. 새로운 필터들의 전파는 두 부분으로 나뉠 수 있다. 처음은 크기의 문제이다. 전파되는 필터들은 크기가 커서 전체로서 용이하게 분배되기가 어려울 수 있다. 다행히, 이는 적어도 부분적으로는 구 필터와 신 필터 사이의 "차이점"을 룩업 시스템을 통해 전달함으로써 극복될 수 있다. 이하 설명할 바처럼, 차이점은 복수의 요인들 및/또는 설정들(클라이언트 또는 서버)에 기초하여 결정될 수 있다.

[0035] 둘째는 필터 업데이트들의 관리에 대한 문제이다. 한편, 많은 사람들은 새로운 코드 및 새로운 데이터를 포함하는 스팸 필터들에 대한 모든 변경 사항이 자동으로 전파되는 것을 원한다. 달리 말해, 많은 관리자들이 전체 조직에 이를 자동으로 전파하기에 앞서 새로운 파일들 및/또는 데이터를 테스트 머신에 설치하고자 한다. 이하 도면을 참조하여 논할 바처럼, 온라인 룩업 서비스를 통해 차이가 있는 피처들(가중치들)만을 메모리에 전달함으로써 큰 필터를 전파할 필요가 경감된다.

[0036] 도 1을 참조하면, 본 발명의 일 태양에 따라 이전의 또는 기존의 스팸 필터에 대하여 차이점 정보를 제공하는 것을 용이하게 하는 안티 스팸 업데이트 시스템(100)의 전반적인 블록도가 도시되어 있다. 시스템(100)은 들어오는 테스트 메시지들을 스팸 또는 양호한 메시지로 분류하기 위하여 적어도 하나의 스팸 필터(120)를 이용하는 테스팅 메시지 수신 컴포넌트(110)를 포함한다. 테스트 메시지 수신 컴포넌트(110)는 제3자 클라이언트 또는 서버(예컨대 홈 컴퓨터)에 위치할 수 있다. 스팸 필터(120)는 SVMs(Support Vector Machines), 최대 엔트로피 모델들(논리 회귀), 퍼셉트론들(perceptrons), 의사 결정 트리들(decision trees) 및/또는 신경망들 중 임의의 것을 이용하여 차별적으로(discriminatively) 트레이닝될 수 있다.

[0037] 시스템(100)은 또는 다양한 트레이닝 메시지들을 수신할 수 있는 트레이닝 메시지 수신 컴포넌트(130)를 포함한다. 예로서는 피드백 루프 데이터(예컨대 들어오는 메시지들의 적어도 선택된 일부분을 양호한 메시지 또는 스팸으로 분류하는 데에 참여하는 사용자들로부터의 데이터), 사용자 불만, 꿀단지(honeypot) 데이터 등이 포함된다. 트레이닝 메시지 수신 컴포넌트(130)는 필터 제공자에게서 찾아볼 수 있다.

[0038] 들어오는 트레이닝 메시지들의 적어도 일부는 피처 분석 서브시스템(140)으로 향할 수 있고, 여기서는 이들의 스팸형(spam-like) 특성 및/또는 비 스팸형 특성에 관하여 이러한 메시지들을 분석 및 조사할 수 있다. 특히, IP 주소, URL 및/또는 특정한 텍스트와 같은 복수의 피처들은 각 메시지로부터 추출되어 분석될 수 있다. 기계 학습 컴포넌트(150)를 이용하여, 업데이트 컴포넌트(160)가 트레이닝될 수 있으며, 이는 차별적인 방법으로 수행될 수 있다. 그 대신, 업데이트 컴포넌트(160)는 매치 또는 해시 기반의 데이터를 이용하여 트레이닝될 수 있다. 피처 분석 서브시스템(140)으로 향하는 메시지들은 필터링되지 않거나 또는 필터링된 메시지 또는 이들의 조합일 수 있다. 필터링된 메시지들을 스팸 또는 양호한 메시지들로 분류하는 것이 업데이트 컴포넌트(160)의 트레이닝 또는 스팸 필터(120)에 대한 업데이트들을 생성하는 것에 영향을 미칠 필요는 없다.

[0039] 업데이트 컴포넌트(160)는 단어들, IP 주소의 리스트들, 호스트 명칭들(hostnames), URL들 등과 같이 들어오는 메시지들로부터 추출될 수 있는 복수의 메시지들 및/또는 피처들에 대하여 계산되는 가중치들과 같은 데이터를 포함한다. 이러한 데이터는 업데이트 컴포넌트(160)에 의해 제어되는 하나 이상의 데이터 파일 또는 데이터베이스로 조직될 수 있다.

[0040] 메시지 수신/분배 시스템(예컨대 하나 이상의 서버들)(110)에 의해 프롬프트(prompt)되는 경우, 업데이트 컴포넌트(160)는 스팸 필터(120)의 적어도 일부를 추가적인 정보로 적어도 증분적으로 증가시킬 수 있다. 예컨대, 업데이트 컴포넌트(160)는 새로운 피처-가중치 데이터를 추가하고/하거나 소정의 피처에 대한 새로운 가중치 데이터로 기존의 가중치 데이터를 대체함으로써 스팸 필터의 데이터 부분을 업데이트할 수 있다. 업데이트 컴포넌트(160)는 또한 가능한 경우 정해진 시간에 따라 증분적인 업데이트를 제공하여 임의의 1개의 업데이트의 상대적인 크기를 최소화하도록 커스터마이징될 수 있다. 업데이트들 자체는 시간을 기준으로 할 뿐만 아니라 들어오는 수신 메시지들의 수를 기준으로 하여 생성될 수 있다. 예컨대, 업데이트들은 매 시간 및/또는 수신되는 30번째 메시지마다 생성될 수 있다.

[0041] 이제 도 2를 참조하면, 스팸 방지를 용이하게 하는 안티 스팸 업데이트 시스템(200)의 블록도가 도시되어 있다. 일반적으로, 업데이트 시스템(200)은 새로운 파라미터 데이터를 기존의 스팸 필터(210)가 사용하는 종전의 파라미터 데이터와 비교한다. 시스템(200)은 들어오는 메시지들로부터 추출된 피처들을 조사하여 관련 피처들(예컨대 스팸 또는 양호한 메시지들인 것을 나타냄)을 식별하고 이들의 가중치, 점수 및 다른 관련 데이터를 확인할 수 있는 피처 추출-분석 컴포넌트(220)를 포함한다. 이러한 데이터는 파라미터 업데이트 데이터베이스(230)에 저장 및 유지될 수 있다. 데이터베이스 내의 새로운 파라미터 데이터는 파라미터 분석 컴포넌트(240)에 의해 기존 필터(210) 내의 종전 파라미터 데이터와 관련하여 분석됨으로써, 파라미터 데이터가 변경되었는지가 결정될 수 있다.

[0042] 예컨대, 파라미터의 가중치는 더 높은 또는 더 낮은 스팸 특성을 가리키도록 증가 또는 감소될 수 있다. 또한, 파라미터들은 기존 필터(210)에 대하여 추가 또는 삭제될 수 있다. 후자의 경우, 파라미터 또는 피처는 이것의 가중치가 0이 되는 경우 필터(210)로부터 제거될 수 있다.

[0043] 임의의 특정한 파라미터들의 집합 또는 부분 집합에 대하여 업데이트가 존재하는 것으로 판정되는 경우, 이러한 파라미터들은 업데이트 제어기(250)에 전달될 수 있다. 업데이트 제어기(250)는 데이터베이스(230)로부터 관련 파라미터 데이터를 액세스할 수 있고, 이러한 데이터를 기존의 필터(210)에게 전달할 수 있다. 본질적으로, 시스템(200)은 업데이트 서비스를 스팸 필터들에게 전달하여 이들을 최신 상태로 유지하고 새로운 형태의 스팸에 효율적으로 대처할 수 있게 한다.

[0044] 업데이트 시스템(200)은 클라이언트 또는 서버 상에서 자동으로 실행될 수 있다. 또한, 서비스는 구독(subscription)에 의해 동작할 수 있는데, 여기서는 지불 검증 컴포넌트(260)가 임의의 업데이트가 제공되기에 앞서 업데이트 또는 업데이트 서비스에 대하여 클라이언트 또는 서버가 지불하였는지 여부를 결정할 수 있다. 대신, 필터(210)는 룩업 또는 업데이트가 실행되기에 앞서 구독이 최신인지 여부를 검증할 수 있다.

[0045] 업데이트 룩업 시스템(예컨대 도 1 및 도 2에 도시)은 DOS(Denial-Of-Service) 또는 분산형 DOS 공격에 대한 본연적인 목표가 될 수 있다. 따라서 시스템은 예컨대 이를 복수의 IP 주소들 또는 상이한 IP 주소들에 대응하는 복수의 호스트 명칭들에게 분산함으로써 이러한 공격에 대하여 강화될 수 있다. 실질적으로, 예컨대 상이한 IP 주소들은 상이한 사용자들(또는 클라이언트들이나 서버들)에게 분산되어 공격자가 공격할 IP 주소들의 전체 리스트를 찾는 것을 보다 어렵게 할 수 있다.

[0046] 기계 학습 기법을 이용하면, 업데이트될 수 있는 상이한 수치 파라미터들이 수천 개 존재할 수 있는 바, 실질적으로 이들 모두가 적어도 소정의 적은 양만큼 변할 수 있기 때문이다. 결과적으로, 어떠한 업데이트가 이루어

질지 결정하는 것은 여러 상이한 방식으로 정해질 수 있다. 예컨대 한 가지 방식은 가장 많이 바뀐 파라미터들의 절대값들에 주목하는 것이다. 그러나, 최대 절대값 변화는 어느 파라미터를 업데이트해야할지를 가장 잘 나타내는 것이 아닐 수 있다. 이는 특히 파라미터들이 거의 관측되지 않는 피처에 관한 것일 경우에 그러하다. 따라서, 어느 파라미터를 업데이트해야할지 결정하는 경우에 고려해야 할 다른 요인들에는 발생, 빈도, 또는 가장 최근 데이터에 기초한 파라미터의 공통성(commonality)이 포함될 수 있다. 예컨대, 파라미터가 많이 바뀌었으나 대응되는 피처가 매우 적은 메시지들에게서 나타나는 경우(예컨대 매 100,000 메시지들마다 평균 3건), 이러한 피처에 대한 업데이트를 전달하는 것은 업데이트 서비스의 효율적인 사용이 아닐 수 있다.

[0047] 다른 방식으로는 파라미터들을 중요하게 만드는 일정량만큼(예컨대 소정의 최소값 또는 임계치만큼), 또는 보다 공통적인 피처들의 경우에는 보다 덜 공통적인 피처들과는 상이한 소정의 최소값만큼 변한 파라미터들의 절대값에 주목하는 방식이 있다. 특정 임계치가 충족되면, 파라미터가 업데이트될 수 있다. 그렇지 않으면, 이는 그대로 남게된다.

[0048] 또 다른 방식으로는 파라미터 변화의 수를 한정하고자 하는 필터들 또는 이들에 대한 업데이트들을 구축하는 방식이 있다. 평형 피처들(counterbalancing features)라고 지칭되는 일부 피처들은 서로 상호작용하여 최종적으로 필터의 동작에 영향을 미칠 수 있다. 평형 피처들이 트레이닝 동안 필터 내에서 제대로 고려되지 않은 경우, 필터의 성능이 바뀔 수 있다. 따라서, 파라미터 변화의 수를 한정하는 필터들을 구축하는 것은 평형 피처들이 적절히 고려되었는지를 추적할 필요를 또한 경감시켜 줄 수 있다.

[0049] 예컨대, "wet"이라는 단어에 대하여 0의 가중치를 갖고 "weather"라는 단어에 대하여 약간 음인 가중치를 가지고 현재 사용되는 필터 A를 가정하자. 그리고 "wet"을 포함하는(그러나 weather는 포함하지 않음) 다량의 스팸이 도착하였다고 가정하자. 또한 단어 "wet" 및 "weather"를 함께 포함하는 양호한 메일이 보통의 양만큼 있다고 가정하자. 새로운 필터 B는 "wet"에 대하여 스팸형 쪽으로 크게 가중치를 부여하고 "weather"에 대해서는 평형을 위한 음의(양호한) 가중치를 부여함을 학습하게 되어, 이들 단어가 동시에 발생하는 경우에는 이들 가중치가 상쇄되어 그 메일이 스팸으로서 분류되지 않게 된다. 여기서 필터 A에 비하여 볼 때 필터 B에서의 단어 "wet"은 가중치를 업데이트할 만큼 중요한 것으로(많은 양의 메일에서 발생하였음), "weather"는 그렇지 않은 것으로(적은 양의 메일에서 발생하였고 이미 작은 음의 가중치를 갖고 있었으므로 적은 양만큼 변함) 결정될 수 있다. 따라서, "wet"에 대한 업데이트는 전파될 수 있으나 "weather"에 대한 평형 업데이트는 그렇지 않을 수 있고, 이는 많은 수의 실수를 야기한다. 이러한 바람직하지 못한 업데이트의 생성을 방지하기 위해, 파라미터 변화의 수를 최소화하기 위한 필터가 도 3과 같이 구축될 수 있다.

[0050] 도 3에 따르면, 이전 데이터의 피처들 및 가중치들을 포함하는 구 필터 X(310)로 시작한다. 이제 기계 학습을 이용하여 신 필터 Y1(320)을 트레이닝한다. 소정의 휴리스틱(들)(330)에 비추어 중요한 X(310)와 Y1(320) 사이의 차이를 찾아낸다. 예컨대, 차이의 절대값, 차이로부터의 정보 획득, 차이의 절대값을 파라미터의 이용 빈도와 곱한 값 등이 측정될 수 있다. 선형 모델{예컨대 SVM 모델, 나이브-베이즈(Naive-Bayes) 모델, 퍼셉트론 모델, 맥센트(maxent) 또는 논리 회귀 모델}의 경우, 모델은 피처들(예컨대 메시지 내의 단어들)에 대한 가중치들로 이루어진다. 선형 모델에 있어서, 이는 이러한 측정치들 중 하나에 따라 가장 많이 변한 피처 가중치들을 찾는 것으로 이루어진다(340).

[0051] 이어서, 신 필터 Y2(350)는 필터들간의 모든 작은{또는 중요하지 않은(360)} 차이는 Y2(350)에서 X(310)에서와 같은 값을 가져야 한다는 제한 하에서 학습된다. 예컨대, 선형 모델에 있어서, 이는 많이 변하지 않은 피처들에 대한 가중치들은 Y2(350) 및 X(310)에서 동일하다는 것을 의미한다. 그러나, 많이 바뀐 피처들의 경우(예컨대 소정의 임계치 또는 휴리스틱을 충족하는 경우), 가중치들은 Y2(350)에서 상이하다. 이전의 "wet" 및 "weather"의 예를 참조하면, "wet"이 불량한 것으로 학습한 경우, 어떠한 용어가 매우 불량한 것으로서 학습될 수는 없는데, 이는 그에 대한 평형 가중치("weather")가 고정되기 때문이다. 따라서, 평형 피처들이 고려되었는지를 더 이상 추적해야 할 필요가 없다.

[0052] 선택적으로, 이러한 프로시저는 반복되어 웨이트가 일관되게 상이한 피처들만을 찾아낼 수 있다. 예컨대, "weather" 파라미터 값이 바뀌지 않는 경우, "wet" 파라미터 값을 바꾸지 않을 것으로 결정할 수 있다.

[0053] 또한, 필터(들)은 Y1(320)을 이용하는 업데이트 대신 Y2(350)에 대한 업데이트를 이용하여 업데이트될 수 있다. Y2(350)와 X(310) 사이의 차이는 Y1(320)과 X(310) 사이의 차이보다 작은데, 이는 모델의 많은 부분이 동일한 제약을 받기 때문이다.

[0054] 이를 대신하는 기법으로는 보다 빠르게 바뀌는 부분 또는 모델에 보다 큰 영향을 주는 부분과 같은 데이터의 오

직 일부만을 업데이트하는 것이 있다. 예컨대, IP 주소 및 URL 데이터가 텍스트 데이터보다 빠르게(또는 느리게) 변할 수 있다. 또한, 이들 피처들을 다른 피처들과 무관하게 트레이닝하는 것이 쉬울 수 있다(예컨대 2004년 3월 25일자 미국 특허 출원 제10/809,163호, "Training Filters for IP Address and URL Learning" 참조). 따라서, 소정의 피처 집합을 일정하게 유지하면서 다른 피처들은 변경될 수 있도록 하는 모델이 구축될 수 있다.

[0055] 또한, 선택적으로 피처들의 부분집합(예컨대 다른 것들과 독립적인 적어도 하나의 파라미터)을 업데이트함으로써, 모델에 대한 장래의 업데이트들이 보다 쉽게 달성될 수 있다. 이러한 종류의 모델의 한 가지 예는 의사 결정 트리 모델이며, 여기서는 각 리프(leaf)가 다른 리프들에 있어서의 모델들과 별개로 업데이트될 수 있는 독립 모델을 포함한다. 연구 결과 이들 모델은 현재 구축되는 통상적인 모델과 동일한 수의 피처들을 가지나 보다 나은 전체 성능을 가질 수 있다는 점이 발견되었다.

[0056] 모델이 모델 구축{여기에는 피처들을 서로 관련된 그룹으로 클러스터링(clustering)하거나 또는 다른 메커니즘에 의해 임의로 피처 공간을 분할하는 것이 포함됨} 동안에 가중치들 서로간에 평형을 맞추지 않거나 또는 그것이 허용되지 않는 피처 부분 집합들을 갖도록 연역적으로 디자인될 수 있는 다른 방식이 존재한다. 그 대신, 의사 결정 트리에서와 같이, 메시지들은 예컨대 이들을 관련 그룹들로 클러스터링 함으로써 분할될 수 있다(이 경우 의사 결정 트리의 경우와 같이, 다른 클러스터에서 다른 가중치를 갖고 복제되는 피처들이 있을 수 있으나 이들은 독립적으로 업데이트될 수 있음).

[0057] 증분적 업데이트들은 또한 특정 고객(서버 또는 클라이언트)이 가장 많이 수신한 메시지들에 적용되는 피처들에 우선 초점을 두는 업데이트된 피처들과 함께 클라이언트, 서버, 또는 사용자가 수신하는 메시지들의 분포에 의해 적어도 부분적으로 결정될 수 있다. 따라서, 복수의 클라이언트는 예컨대 이들이 수신하는 메시지들의 유형에 따라 자신들의 필터들에 대한 상이한 업데이트들을 수신할 t 있다.

[0058] 업데이트들의 유형이 결정되면, 스팸 필터의 업데이트를 관리하는 것이 당면 과제일 수 있다. 메시지 시스템 관리자들은 종종 또는 가끔 자신들의 사용자들이 어떠한 소프트웨어(데이터 파일들을 포함함)를 사용하고 있는지 알고싶어 한다. 어떤 경우, 관리자들은 자신들의 모든 사용자들이 동일한 데이터를 실행할 것을 원하거나, 또는 그렇지 않으면 바람직한 환경에서 시험해 볼 기회를 갖기 전에 새로운 데이터 파일들을 분배하는 것을 원하지 않을 수 있다. 따라서, 이들은 사용자가 직접 업데이트 서비스와 통신하는 것을 원치 않을 수 있다.

[0059] 예컨대, 한 가지 시나리오에서, 관리자들은 특정한 파일들을 우선 다운로드하여 이들에 대한 동작성, 다른 시스템 파일들과의 충돌 등을 테스트하는 것이 이들을 사용자들에게 전달하는 것보다 먼저 이루어 질 것을 선호할 수 있다. 따라서 데이터 또는 코드에 대한 업데이트들이 우선 관리자에게 전달되고 다음으로 사용자들에게 전파되도록 하는 2단계의 전파를 용이하게 하는 것이 바람직하다. 어떤 경우에, 관리자들은 이미 필터 공급자를 신뢰하여 검증 없는 완전 자동 룩업 프로세스를 선호할 수도 있다.

[0060] 이러한 룩업 또는 업데이트 서비스는 이메일 클라이언트 또는 서버 상에서 동작하기 위한 코드를 필요로 할 수 있음을 이해하여야 한다. 또한, 룩업 또는 업데이트는 최종 사용자 또는 관리자에 의해 지정될 수 있는, 스케줄링된 간격으로 수행될 수 있다. 그 대신, 룩업 또는 업데이트는 어떠한 이벤트가 발생하는 경우, 예컨대 메시징 프로그램이 시작되거나 열렸을 때 수행될 수 있다. 업데이트를 이용할 수 있는 경우, 최종 사용자 또는 관리자는 통지를 받거나(예컨대 업데이트가 선택적임), 또는 업데이트가 자동으로 이루어질 수 있다. 최종 사용자 또는 관리자는 이들 옵션 가운데에서 선택할 수 있다. 마지막으로, 스팸 필터에 대한 업데이트는 메시징 프로그램을 시작하는 것조차 없이 즉시 발생하여 효력을 가질 수 있다.

[0061] 앞서 논한 바처럼, 스팸 필터들에 대한 업데이트는 적어도 증분적일 수 있으며, 이에 따라 스팸 필터의 가장 유용하거나 필요한 부분들이 업데이트되고 나머지 부분들이 그대로 유지되어 업데이트들 및 그와 관련된 데이터 파일들의 크기를 최소화할 수 있다. 대부분의 경우, 서버들은 어떤 업데이트를 수행할 것인지, 언제 이러한 업데이트를 수행할 것인지 및/또는 이러한 업데이트를 수행하는 방식에 대한 책임을 진다. 불행히도, 서버들은 이러한 결정을 내리는 데 느릴 수 있으며, 또는 이러한 업데이트의 타이밍이나 컨텐츠가 클라이언트 또는 사용자의 필터링 요구와는 다소 맞지 않을 수 있다. 어느 경우도 클라이언트에게는 문제일 수 있으며, 특히 기존의 스팸 필터가 특정 메시지(들)의 분류에 관해 단정하지 못하고 클라이언트가 서버가 프롬프트하는 업데이트를 기다림으로써 더 이상의 지연을 유지할 수 없는 경우에 특히 그러하다.

[0062] 도 4에는 클라이언트가 사용중인 동안 스팸 필터들을 업데이트하는 것이 가능한 룩업 서비스 시스템(400)의 블록도가 도시되어 있다. 룩업 서비스 시스템(400)은 앞서 도 1의 업데이트 시스템(100)과 유사할 수 있으며, 특

히 스팸 필터에 대하여 준 실시간 또는 실시간 전파를 위한 소정 유형의 업데이트 데이터를 생성한다는 점에서 그러하다. 그러나, 룩업 서비스 시스템(400)은 이에 더하여 서버 명령만이 아닌 클라이언트 또는 최종 사용자의 요청에 의해 스팸 필터에 대한 업데이트들을 제공할 수 있다.

[0063] 도 4에 따르면, 들어오는 테스트 메시지들은 메시지들이 스팸인지 아닌지를 분류하기 위한 적어도 하나의 스팸 필터(420)를 이용하는 테스트 메시지 수신 컴포넌트(410)에 전달될 수 있다. 테스트 메시지들은 파라미터들의 현재 세트가 주어졌을 때 스팸 필터(420)의 정확성을 결정하는 데 도움을 줄 수 있다. 테스트 메시지 수신 컴포넌트(410)는 제3자 서버 또는 클라이언트 상에 위치할 수 있다. 스팸 필터(420)는 기계 학습 또는 비 기계 학습 트레이닝을 받을 수 있다.

[0064] 업데이트 학습은 다음과 같이 수행될 수 있다. 즉, 들어오는 메시지의 적어도 일부가 트레이닝 메시지 수신 컴포넌트(필터 제공자 상에 위치함)(435)를 지나서 피처 분석 시스템(430)으로 향할 수 있다. 피처 분석 시스템(430)은 트레이닝 메시지들의 적어도 일부로부터 추출된 피처들 및 이들 각각의 가중치들에 기초하여 최근의 데이터를 생성하고 이들을 룩업 데이터베이스(440)에 저장할 수 있다.

[0065] 스패머들은 지속적으로 자신들의 스팸을 개변 및/또는 수정하므로, 기존의 스팸 필터(420)로는 스팸인지 또는 양호한 것인지 분류되지 못할 수 있는 메시지들이 일부 있을 수 있다. 클라이언트는 이러한 메시지들에 표지를 하고 그 메시지, 그 메시지의 해시 및/또는 그 메시지의 하나 이상의 피처에 기초하여 룩업 컴포넌트(450)에 대한 질의 또는 요청을 전달할 수 있다.

[0066] 룩업 데이터베이스로부터의 어떤 데이터가 이러한 요청을 충족하는 경우, 그러한 대응되는 정보가 스팸 필터(420)를 업데이트하도록 전달 또는 다운로드될 수 있다. 그 후, 업데이트된 스팸 필터는 불확실한 메시지들뿐만 아니라 새로운 메시지들에도 적용되어 분류 프로세스를 원활하게 할 수 있다.

[0067] 이제 도 5를 참조하면, 클라이언트(510)에 의해 이용되는 웹 기반의 업데이트 서비스를 용이하게 하는 온라인 룩업 시스템(500)의 개략도가 도시되어 있다. "종전의" 데이터로 트레이닝된 기존의 스팸 필터가 들어오는 메시지들(515)을 분류하도록 되어 있다고 가정하자. 불행히도, 클라이언트의 기존 필터는 일부 메시지들이 스팸인지 양호한 것인지를 결정하는 데 어려움을 겪고 있다. 메시지들(515)을 격리하거나 또는 서버가 프롬프트하는 업데이트가 도착하기를 기다리는 대신, 클라이언트(510)는 메시지들(515) 또는 그로부터 추출된 일부 피처들, 예컨대 IP 주소(520), URL(525), 호스트 명칭(530) 또는 다른 기타 특징(들)(535)을 취하여 온라인 룩업 시스템(500)에게 질의할 수 있다. 온라인 룩업 시스템(500)은 하나 이상의 룩업 테이블(540) 및/또는 하나 이상의 데이터베이스(545)를 포함할 수 있다. 룩업 테이블(540)은 예컨대 IP 주소(555)와 같은 피처(550)에 대하여 업데이트된 데이터를 포함할 수 있다. 클라이언트가 메시지의 IP 주소에 관한 질의를 수행하는 경우, 그 IP 어드레스는 적합한 룩업 또는 업데이트 테이블에서 룩업될 수 있다.

[0068] 마찬가지로, 데이터베이스(545)는 IP 주소(520)에 관한 임의의 업데이트에 대하여 참조 또는 검색될 수 있다. 데이터베이스(545)는 IP 주소(565)와 같은 업데이트된 피처(560)에 대하여 배열될 수 있다. 테이블 또는 데이터베이스 크기를 조절하기 위해, 업데이트된 정보를 갖는 피처들만이 룩업 테이블 또는 데이터베이스 내에 제공될 수 있다. 그러나, 가중치 또는 값이 변하였는지에 무관하게 실질적으로 모든 피처들을 갖는 룩업 테이블들 및/또는 데이터베이스들 역시 가능하다. 테이블 또는 데이터베이스 관계없이, 업데이트가 발견되면 이는 클라이언트에 의하여 직접 전달 또는 다운로드되어 스팸 필터를 업데이트할 수 있다. 따라서, 스팸 필터에 대한 업데이트들은 클라이언트 설정들에 기초할 수 있으며, 필요에 따라 발생할 수 있다.

[0069] 서버 또는 클라이언트 상의 메시징 시스템이 이전의 모든 업데이트를 수신하지 못한 경우, 하나 이상의 일련의 피처들 또는 업데이트들을 룩업할 필요가 있을 수 있다. 시스템은 마지막으로 기록된 룩업 이후의 룩업을 수행하여 이들을 순서대로 적용할 수 있다. 선택적으로, 업데이트 서버는 복수의 룩업 파일들을 병합하여 다운로드의 효율을 높일 수 있다. 마지막으로, 업데이트는 예컨대 HTTPS와 같은 보안 채널을 통해 이루어질 수 있다.

[0070] 새로운 데이터에 대한 증분적 룩업은 파일에 기록되거나 디스크 상에 저장되어 메모리에서 결합될 수 있다. 또한, 증분적 업데이트들은 모델의 소정 부분, 피처(들), 또는 파라미터(들)이 더 이상 필요하지 않음(예컨대 가중치가 0임)을 지정할 수 있고, 따라서 이들이 삭제되어 메모리 또는 디스크 공간을 절약하게끔 할 수 있다.

[0071] 이제 도 6을 참조하면, 본 발명의 일 태양에 따른 안티 스팸 룩업 웹 기반 서비스의 예시 아키텍처(600)가 도시되어 있다. 아키텍처(600)는 예컨대 트레이닝되는 동안 생성되는 피처들의 부분 집합 및 관련 가중치들 및 모델들을 수용하는 데이데-티어(data-tier) 계층(또는 백엔드 데이터베이스), 데이터베이스와 스팸 필터 사이의 통신을 전달하는 미들웨어 계층 및 최근에 업데이트된 모델을 획득하기 위하여 미리 정해진 또는 자동적인 빈도

로 미들웨어 계층을 호출하고 온라인 모델을 국지적으로 저장된 모델 파일과 병합하는 스팸 필터와 같은 복수의 계층을 포함한다.

[0072] 보다 구체적으로, 데이터-티어 계층은 두 개의 저장소, 즉 트레이닝 저장소(610)(표준 트레이닝을 위하여 사용됨) 및 업데이트 저장소(620)를 수용한다. 이들 저장소는 플랫 파일(flat file) 또는 데이터베이스일 수 잇다. 전용 트레이닝 저장소(610)는 선택적으로 잦은 업데이트로부터 이득을 얻는 피처들의 부분 집합에 대한 피처들 및 가중치들만을 수용할 수 있다. 업데이트 저장소(620)는 전용 트레이닝 저장소(610)로부터의 정보의 부분 집합뿐만 아니라 배치된 제품과 관련된 몇몇 새로운 변수들로부터 비롯되는 이진 형태의 모델 출력을 포함하는 새로운 데이터베이스 또는 플랫 파일들의 집합이다. 이러한 정보의 부분 집합은 이하를 포함할 수 있다.

[0073] - 더욱 잦은 업데이트로부터 크게 이득을 얻는 피처들을 포함하는 새로운 모델들

[0074] - 그 예에는 URL 피처들, IP 피처들 및 새로운 특별 피처들이 포함됨

[0075] - 배치된 모델 파일들의 이전 버전들에 대한 각각의 새로운 확률 모델들의 관계 및/또는

[0076] - 새로운 모델 전송의 크기를 최소화하기 위한 새로운 모델들의 증분적 업데이트들

[0077] 미들웨어 계층(630)은 업데이트 저장소(620)와 스팸 필터 .dll 파일(640) 사이의 인터페이스로서 작용할 수 있다. 이는 웹 서비스 인터페이스들 및 스팸 필터와 온라인 룩업 서비스 사이에서 정보를 전달하는 기능을 제공한다. 이는 SOAP 서비스, HTTP 서비스, HTTPS 서비스, 또는 다른 인터넷 서비스일 수 있다.

[0078] 안티 스팸 룩업 서비스는 특히 소정의 다른 스팸 관련 시스템 및 방법과 결합될 때 강력하다. 특히, 이는 메시지 격리(message quarantining)와 결합될 때 특히 강력하다. 메시지 격리에 있어서, 일부 메시지들은 정크 폴더 또는 격리 폴더에 놓여지거나, 그렇지 않은 경우에는 임시적으로 보유된다. 이후 이들은 스팸 필터 업데이트 후에 재채점(rescore)된다. 사용자들이 정크 메시지들을 중앙 저장소에 보고하는 "정크 보고 버튼"과 같은 기법들이 또한 스팸 필터 업데이트에 대한 중요한 데이터를 제공할 수 있다. 또한, 메시지를 받지 말아야 하는 소정의 계정(예컨대 새로이 생성되어 사용되지 않은 계정)으로 데이터가 전달되는 꿀단지(honeypot)와 같은 기법들은 스팸 필터 업데이트에 대한 귀중한 소스이다. 또한, 피드백 루프에 있어서, 사용자들은 어떠한 메시지가 양호한지 또는 스팸인지에 대하여 투표(poll)를 한다. 이는 스팸 필터를 업데이트하기 위한 귀중한 데이터를 제공한다. 데이터가 상대적으로 편향되어 있지 않으므로, 이는 정크 보고 또는 꿀단지 데이터보다 유용할 수 있다.

[0079] 본 발명에 따른 다양한 기법이 이하 일련의 동작을 통해 기술될 것이나, 본 발명은 동작의 순서에 한정되지 않으며, 이는 본 발명에 따라 일부 동작들이 본 명세서에 도시 및 기술된 것과는 다른 순서로 및/또는 다른 동작들과 동시에 일어날 수 있기 때문이라는 점을 명심해야 할 것이다. 예컨대, 본 기술 분야의 당업자는 어떠한 방법은 일련의 상호 연관된 상태 또는 이벤트들(예컨대 상태도)로서 표시될 수 있음을 알 수 있다. 또한, 설명한 동작 모두가 본 발명에 따른 방법을 구현하는 데 필요한 것은 아니다.

[0080] 도 7을 참조하면, 사용중에 스팸 필터들을 적어도 준 실시간으로 업데이트하는 것을 용이하게 하는 예시 스팸 필터 업데이트 프로세스(700)의 흐름도가 도시되어 있다. 프로세스(700)는 기계 학습 기법 등에 의해 새로운 또는 보다 최근의 데이터(메시지들)로 신 필터를 트레이닝하는 단계(710)를 포함한다. 신 필터는 복수의 메시지 피처들 및 이들과 관련된 가중치들에 대하여 차별적으로 트레이닝될 수 있다. 피처들의 예에는 IP 주소들, URL들, 호스트 명칭들, 도는 메시지로부터 추출될 수 있는 임의의 단어들 또는 텍스트가 포함된다.

[0081] 단계(720)에서, 프로세스(700)는 신 필터와 구 필터(구 데이터로 트레이닝됨) 사이의 차이점을 찾아볼 수 있다. 발견 또는 탐지되는 모든 차이점들이 하나 이상의 별개의 데이터 파일들에 저장될 수 있다(730).

[0082] 선택적으로, 데이터 파일들은 데이터베이스에 저장되고/되거나 그에 포함된 컨텐츠들은 하나 이상의 룩업 테이블로 배열될 수 있다. 클라이언트들은 웹 기반의 룩업 서비스를 통해 이들 데이터 파일을 이용할 수 있다. 비록 도면에 도시되지는 않았으나, 클라이언트들은 룩업 서비스에게 임의의 메시지 또는 자신의 기존 스팸 필터를 이용하여 분류될 수 없는 메시지로부터의 피처들에 대한 특정한 업데이트들의 가용 여부를 질의할 수 있다. 업데이트를 이용할 수 있는 경우, 클라이언트는 자신이 원하는 것을 선택하여 다운로드함으로써 부분적으로 또는 증분적으로 기존의 스팸 필터를 업데이트할 수 있다.

[0083] 다시 도 7을 참조하면, 구 스팸 필터는 하나 이상의 데이터 파일들로 업데이트될 수 있다(단계 740). 따라서, 구 필터를 완전히 새로운 필터로 대체하는 것과는 반대로, 구 필터는 충분한 변화량을 보인 데이터로 증분적으

로 업데이트된다.

[0084] 실제적으로, 예컨대 파라미터들의 절대값들이 구 필터와 신 필터 사이에서 비교될 수 있다. 변화 임계치가 설정될 수 있다. 임의의 파라미터의 절대값 변화가 이러한 임계치를 충족하는 경우, 이러한 변화는 업데이트 컴포넌트 또는 데이터 파일에 저장될 수 있다. 메시지 내의 파라미터들의 빈도와 같은 다른 요인들은 특정 "변화"가 업데이트에 포함되는지 여부에 영향을 미칠 수 있다. 업데이트들은 데이터 파일들로서 저장될 수 있고, 룩업 테이블로서 배열될 수 있고/있거나, 검색 가능한 데이터베이스들에 저장될 수 있다.

[0085] 또한, 업데이트 요청은 서버들 및/또는 개인 클라이언트들에 의해 이루어질 수 있다. 예컨대, 서버 관리자들은 들어오는 메시지들 및 이들의 필터링을 조사할 수 있고, 다양한 요인들, 예컨대 특정 메시지들에 대한 사용자들의 불만이 증가하는 것 및/또는 격리소 내의 메시지들의 수 또는 유사도가 증가하는 것을 관측하는 것에 기초하여 특정한 업데이트들이 필요한지를 결정할 수 있다. 이러한 문제의 영역에 대처하기 위해, 서버들은 적어도 부분적으로 증분적인 필터 업데이트들을 요청할 수 있다. 결과적으로, 이들 업데이트는 서버측에 적용되고 이후 개인 클라이언트들에 적용된다.

[0086] 반대로, 클라이언트들이 직접 요청을 하고 증분적인 업데이트 데이터에 액세스할 수도 있다. 특정한 의심나는 메시지 또는 의심나는 메시지로부터의 피처들이 있는 경우, 클라이언트는 온라인 룩업 테이블 또는 데이터베이스를 통해 이러한 특정 메시지 또는 메시로부터의 피처들이 그에 대응하는 업데이트를 갖고 있는지 여부를 질의할 수 있다. 존재하는 경우, 관련 업데이트들이 클라이언트에게 다운로드되어 클라이언트의 필터에 적용될 수 있다. 서버 또는 이것의 스팸 필터들은 업데이트들에 의해 영향 받지 않는다. 따라서, 클라이언트들은 자신들이 수신하는 메시지들의 특정 유형에 적어도 부분적으로 기초하여 자신들의 스팸 필터들에 대한 업데이트들의 컨텐츠를 커스터마이징 또는 퍼스널라이징(personalizing)할 수 있다. 또한, 업데이트 또는 룩업 시스템에 대하여 보다 최근의 데이터에 관해 질의하는 것은 격리 프로세스를 통해 대기하는 것보다 빠를 수 있다. 또한, 구 필터는 부분적으로 증분적 또는 부분적으로 룩업 서비스/시스템에 의해 업데이트될 수 있다.

[0087] 스팸 필터들은 수천 개의 파라미터들을 트레이닝할 수 있으며, 각 파라미터는 자신과 관련된 값을 갖는다. 하나의 파라미터에 대한 작은 변화가 다른 파라미터들 모두에 있어서 적어도 작은 변화를 야기할 수 있다. 따라서, 어느 정도로는, 파라미터들 사이에 많은 수의 "차이점들" 또는 변화가 있을 수 있다. 변화의 수 및 필터 업데이트의 전체 크기를 최소화하기 위해, 도 8에 도시된 예시적인 프로세스(800)가 이용될 수 있다. 프로세스(800)의 결과로서, 필터들에 대한 업데이트들은 종전 및 새로운 데이터 사이의 보다 중요하고 의미 있는 변화에 대해 집중할 수 있다.

[0088] 도 8에 도시된 바처럼, 제1 신 필터(예컨대 필터 K)는 새로운 또는 최근에 수신된 메시지들로부터 추출된 데이터를 이용하여 트레이닝될 수 있다(810). 기계 학습 기법들이 이러한 트레이닝에 이용될 수 있다. 단계(820)에서, 새로운 필터 K와 종전의 또는 기존의 필터(현재 사용중인 필터) 사이의 차이가 예컨대 적어도 부분적으로 하나 이상의 휴리스틱에 기초하여 분리될 수 있다. 예컨대, 피처 가중치들이 비교되어 차이의 절대값이 결정될 수 있다(830). 메시지들에서 변화한 피처 또는 파라미터의 빈도 역시 고려될 수 있다. 많은 다른 휴리스틱이 또한 이용될 수 있다. 또한, 하나 이상의 임계치를 구성하여 차이의 절대값과 비교할 수도 있다. 임계치는 또한 피처(들)마다 결정되어 양호한 및/또는 스팸 메시지들의 다양한 피처들의 빈도 또는 발생 속도에 대응할 수 있다. 예컨대, 양호한 또는 불량한 메시지들에 있어서 거의 나타나지 않는 피처들에 대해서는 낮은 임계치가 설정될 수 있다.

[0089] 단계(840)에서, 제2 신 필터(예컨대 필터 Q)는 필터 J 및 K 사이의 모든 차이들(이는 작거나 또는 임계치 또는 휴리스틱을 충족시킬 정도로 크지 않았음)은 이들이 필터 J에서 가졌던 값들과 동일한 값을 가질 수 있다는 제한 하에서 트레이닝된다. 따라서, 이들 특징 피처들에 대한 가중치들은 제2 신 필터에서 일정하게 유지될 수 있다. 단계(850)에서, 구 필터 J와 제2 신 필터 사이의 차이들이 발견될 수 있다. 하나 이상의 임계치 또는 휴리스틱을 충족하는 이러한 차이들은 업데이트 데이터 파일에 저장될 수 있다. 제2 신 필터 Q의 피처들 중 많은 수가 구 필터 J에서와 같은 값들을 갖도록 제한되기 때문에, 두 필터 사이에는 더 작은 수의 변화가 있을 것임이 자명하다. 결과적으로, 필터 업데이트가 더 작다. 구 필터 J는 이후 단계(860)에서 업데이트될 수 있다.

[0090] 그 대신, 구 필터 데이터의 일부가 업데이트될 수 있다. 예컨대, IP 주소 또는 URL 데이터만이 임의의 텍스트 관련 피처들과 독립적으로 조사 및 업데이트될 수 있다. 일반적으로, 업데이트들은 순차적으로 적용될 수 있으며, 특히 서버나 클라이언트가 인터넷에 얼마간 접속하지 않아 현재 복수의 업데이트를 필요로 하는 상황에서 그러하다. 각 업데이트는 다운로드되어 순서대로 적용될 수 있다. 반대로, 필요한 업데이트들이 분석 및 병합되어 업데이트의 전체 사이즈를 감소시킬 수 있다. 예컨대, 서버의 최종 업데이트 이후로 가중치가 변화할 수

있다. 필터를 가중치가 변화할 때마다 업데이트하는 대신, 최종 및 최근 가중치를 적용하고 다른 "중간" 값들은 무시될 수 있다. 따라서, 업데이트가 줄어들거나 더 작아지게 된다.

[0091] 증분적 업데이트의 저장은 특정 서버들 또는 클라이언트들에 따라 융통성이 있을 수 있다. 예컨대, 업데이트들은 별도의 파일에 저장되고 이후 원래의 (필터) 파일과 병합될 수 있다. 그러나, 업데이트 파일들은 이용된 후 곧 폐기될 수 있다. 따라서, 베이스 필터 파일이 유지되어 가장 최근의 차이들이 즉시(on the fly) 확인될 수 있다. 종종, 일부 피처들은 종국적으로 0의 가중치가 될 수 있다. 이들 피처는 필터로부터 제거되어 공간을 절약할 수 있다.

[0092] 본 발명의 다양한 태양들에 대한 추가적인 맥락을 제공하기 위해, 도 9 및 이하의 논의는 본 발명의 다양한 태양들이 구현될 수 있는 적합한 동작 환경(910)의 간략하고 일반적인 서술을 제공하고자 하는 것이다. 본 발명이 프로그램 모듈들과 같이 하나 이상의 컴퓨터들 또는 다른 장치들에 의해 실행되는 컴퓨터 실행 가능 명령어들의 일반적인 맥락에서 기술되었지만, 본 기술 분야의 당업자는 본 발명이 다른 프로그램 모듈들과의 조합 및/또는 하드웨어 및 소프트웨어의 조합으로 구현될 수 있음을 알 수 있다.

[0093] 그러나, 일반적으로 프로그램 모듈들은 특정한 작업들을 수행하거나 특정한 데이터 유형을 구현하는 루틴들, 프로그램들, 객체들, 컴포넌트들, 데이터 구조들 등을 포함한다. 동작 환경(910)은 적합한 동작 환경의 일례일 뿐 본 발명의 용도나 기능의 범위를 한정하고자 하는 것이 아니다. 본 발명과 함께 사용되기에 적합한 다른 잘 알려진 컴퓨터 시스템들, 환경들 및/또는 구성들에는 개인용 컴퓨터, 핸드헬드 또는 랩톱 장치들, 멀티프로세서 시스템들, 마이크로프로세서 기반의 시스템들, 프로그램 가능한 소비자 가전 제품들, 네트워크 PC들, 미니컴퓨터들, 메인프레임 컴퓨터들, 상기 시스템들 또는 장치들을 포함하는 분산형 컴퓨팅 환경들 등이 포함되나, 이에 한정되는 것은 아니다.

[0094] 도 9를 참조하면, 본 발명의 다양한 태양을 구현하기 위한 예시 환경(910)은 컴퓨터(912)를 포함한다. 컴퓨터(912)는 프로세싱 유닛(914), 시스템 메모리(916) 및 시스템 버스(918)를 포함한다. 시스템 버스(918)는 시스템 메모리(916)를 포함하는(이에 한정되지는 않음) 시스템 컴포넌트들을 프로세싱 유닛(914)에 연결시킨다. 프로세싱 유닛(914)은 임의의 이용 가능한 프로세서들일 수 있다. 이중 마이크로프로세서들 및 다른 멀티프로세서 아키텍처들이 또한 프로세싱 유닛(914)으로서 사용될 수 있다.

[0095] 시스템 버스(918)는 메모리 버스 또는 메모리 제어기를 포함하는 여러 유형의 버스 구조(들), 주변 버스 또는 외부 버스 및/또는 로컬 버스 등 임의의 이용 가능한 버스 아키텍처를 이용하는 것일 수 있으며, 여기에는 11비트 버스, ISA(Industrial Standard Architecture), MCA(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), and SCSI(Small Computer Systems Interface)가 포함되나, 이에 한정되는 것은 아니다.

[0096] 시스템 메모리(916)는 휘발성 메모리(920) 및 비휘발성 메모리(922)를 포함한다. 예컨대 시동중에 컴퓨터(912) 내의 요소들 사이에 정보를 전송하기 위한 기본 루틴들을 포함하는 BIOS(Basic Input/Output System)는 비휘발성 메모리(922)에 저장된다. 예컨대, 비휘발성 메모리(922)는 ROM(read only memory), PROM(programmable ROM), EPROM(electrically programmable ROM), EEPROM(electrically erasable ROM), 또는 플래시 메모리를 포함할 수 있으나, 이에 한정되는 것은 아니다. 휘발성 메모리(920)는 외부 캐시 메모리로서 동작하는 RAM(random access memory)를 포함한다. 예컨대, RAM은 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)과 같은 많은 형태로 이용 가능하나, 이에 한정되는 것은 아니다.

[0097] 컴퓨터(912)는 또한 이동형/고정형, 휘발성/비휘발성 컴퓨터 저장 매체를 포함한다. 도 9는 예컨대 디스크 저장소(924)를 도시한다. 디스크 저장소(924)는 자기 디스크 드라이브, 플로피 디스크 드라이브, 테이프 드라이브, Jaz 드라이브, Zip 드라이브, LS-100 드라이브, 플래시 메모리 카드 또는 메모리 스틱과 같은 장치들을 포함하나, 이에 한정되는 것은 아니다. 또한, 디스크 저장소(924)는 CD-ROM 장치, CD-R 드라이브, CD-RW 드라이브 또는 DVD-ROM 드라이브와 같은 광학 디스크 드라이브(이에 한정되지는 않음)를 포함하는 다른 저장 매체와 별도로 또는 조합하여 저장 매체를 포함할 수 있다. 디스크 저장 장치(924)를 시스템 버스(918)에 용이하게 접속할 수 있도록, 인터페이스(926)와 같은 이동형 또는 고정형 인터페이스가 통상적으로 사용된다.

[0098] 도 9는 적합한 동작 환경(91)에서 기술되는 기본적인 컴퓨터 자원들과 사용자들 사이의 매개체로서 동작하는 소

프트웨어를 설명하고 있음을 알 수 있다. 이러한 소프트웨어는 운영 체제(928)를 포함한다. 디스크 저장소(924)에 저장될 수 있는 운영 체제(928)는 컴퓨터 시스템(912)의 자원들을 제어 및 할당하는 작용을 한다. 시스템 애플리케이션들(930)은 시스템 메모리(916) 또는 디스크 저장소(924)에 저장된 프로그램 모듈들(932) 및 프로그램 데이터(934)를 통하여 운영 체제(928)에 의한 자원 관리의 혜택을 받는다. 본 발명은 다양한 운영 체제들 또는 운영 체제들의 조합으로 구현될 수 있음을 이해할 수 있다.

[0099] 사용자는 입력 장치(들)(936)을 통하여 컴퓨터(912)에 대하여 명령 또는 정보를 입력한다. 입력 장치들(936)에는 마우스, 트랙볼, 스타일러스(stylus), 터치패드, 키보드, 마이크, 조이스틱, 게임 패드, 위성 접시, 스캐너, TV 튜너 카드, 디지털 카메라, 디지털 비디오 카메라, 웹 카메라 등과 같은 포인팅 장치가 포함되나, 이에 한정되는 것은 아니다. 이러한 그리고 다른 입력 장치들은 인터페이스 포트(들)(938)을 통해 시스템 버스(918)를 거쳐 프로세싱 유닛(914)에 접속된다. 인터페이스 포트(들)(938)에는 예컨대 직렬 포트, 병렬 포트, 게임 포트 및 USB가 포함된다. 출력 장치(들)(94)은 입력 장치(들)(936)과 동일한 유형의 포트를 일부 사용한다. 따라서, 예컨대 USB 포트는 컴퓨터(912)에 대한 입력을 제공하고 컴퓨터(912)로부터 출력 장치(940)로 정보를 출력하는 데에 사용될 수 있다. 출력 어댑터(942)는 특별한 어댑터를 필요로 하는 다른 출력 장치들(940) 중에서 모니터, 스피커 및 프린터와 같은 일부 출력 장치(940)가 있음을 나타내기 위한 것이다. 출력 어댑터(942)는 예컨대 출력 장치(94)와 시스템 버스(918)사이의 접속 수단을 제공하는 비디오 및 사운드 카드들을 포함하나, 이에 한정되는 것은 아니다. 다른 장치들 및/또는 장치들의 시스템들이 원격 컴퓨터(들)(944)와 같은 입력 및 출력 기능을 제공할 수 있음에 주목하자.

[0100] 컴퓨터(912)는 원격 컴퓨터(들)(944)와 같은 하나 이상의 원격 컴퓨터들에 대한 논리적 접속을 이용하여 네트워크 환경에서 동작할 수 있다. 원격 컴퓨터(들)(944)는 개인용 컴퓨터, 서버, 라우터(router), 네트워크 PC, 워크스테이션, 마이크로프로세서 기반의 응용 기기, 피어 장치(peer device) 또는 다른 통상적인 네트워크 노드 등일 수 있으며, 통상적으로 컴퓨터(912)와 관련하여 기술된 요소들 중 다수 또는 전부를 포함한다. 편의상 원격 컴퓨터(들)(44)에 대해서는 메모리 저장 장치(946)만이 도시되었다. 원격 컴퓨터(들)(944)은 네트워크 인터페이스(948)를 통해 컴퓨터(912)에 논리적으로 접속되고, 통신 접속(950)을 통해 물리적으로 접속된다. 네트워크 인터페이스(948)는 LAN(local area networks) 및 WAN(wide area networks)와 같은 통신 네트워크들을 포함한다. LAN 기술에는 FDDI(Fiber Distributed Data Interface), CDDI(Copper Distributed Data Interface), Ethernet/IEEE 1102.3, 토큰 링(Token Ring)/IEEE 1102.5 등이 포함된다. WAN 기술에는 포인트 대 포인트 링크(point-to-point links), ISDN(Integrated Services Digital Networks)과 같은 회로 스위칭 네트워크들 및 이의 변형, 패킷 스위칭 네트워크들 및 DSL(Digital Subscriber Lines)이 포함되지만, 이에 한정되는 것은 아니다.

[0101] 통신 접속(들)(950)은 네트워크 인터페이스(948)를 버스(918)에 접속시키는 데 이용되는 하드웨어/소프트웨어를 지칭한다. 설명을 명확성을 위해 통신 접속(950)은 컴퓨터(912) 내에 위치하는 것으로 도시되었지만, 이는 컴퓨터(912)의 외부에 존재할 수도 있다. 네트워크 인터페이스(948)에 대한 접속을 위해 필요한 하드웨어/소프트웨어에는 예컨대 통상의 전화급 모뎀, 케이블 모뎀 및 DSL 모뎀을 포함하는 모뎀, ISDN 어댑터 및 이더넷 카드와 같은 내장 및 외장 기술이 포함될 수 있으나, 이는 예시일 뿐이다.

[0102] 이상 기술된 사항은 본 발명의 예시를 포함하는 것이다. 물론, 본 발명을 기술하기 위하여 생각할 수 있는 모든 컴포넌트들 또는 방법들의 조합을 기술하는 것은 불가능하지만, 본 기술 분야의 당업자는 본 발명의 여러 가지 조합 및 치환이 더 가능함을 알 수 있다. 따라서, 본 발명은 첨부된 청구 범위의 취지 및 범위 내에 있는 모든 변경, 개변 및 변형을 아우르는 것이다. 또한, 상세한 설명 및 청구 범위에서 "포함한다"는 용어가 사용되는데, 이러한 용어는 "내포한다"는 용어가 청구 범위에서 과도적인 단어로서 사용되는 경우에 마찬가지를 의미하는 것으로 파악되어야 한다.

발명의 효과

[0103] 본 발명은 실시간 또는 준 실시간 방식으로 스팸 필터들에게 부분적 또는 증분적 업데이트의 형태로 새로운 정보 또는 데이터를 제공하는 것을 용이하게 한다. 특히, 본 발명은 기존 필터에 대하여 정보에 증분적인 부분들을 전달함으로써 필터를 새로운 양호한 메시지들 및/또는 새로운 스팸에 대하여 최신의 상태로 유지하는 것을 용이하게 한다. 이는 부분적으로 차이 학습에 의해 이루어질 수 있으며, 여기서는 기존 필터의 하나 이상의 파라미터들이 새로운 필터 상의 이들 파라미터와 비교될 수 있다. 소정의 변화량을 나타내는 파라미터들은 그에 따라 업데이트되어 전체 필터의 모든 카피를 대체해야 할 필요를 완화시켜 준다. 그 결과, 각 업데이트는 업데이트의 빈도에 따라 그 크기가 상대적으로 작아질 수 있다.

도면의 간단한 설명

[0001] 도 1은 본 발명의 일 태양에 따른 스팸 필터들에 대한 기계 학습(machine learning) 업데이트들을 제공하는 것을 용이하게 하는 안티 스팸 업데이트 시스템의 블록도.

[0002] 도 2는 본 발명의 일 태양에 따른 증분적(incremental) 업데이트 시스템의 블록도.

[0003] 도 3은 본 발명의 일 태양에 따른 스팸 필터들 또는 그에 대한 업데이트(한정된 수의 파라미터 변화를 가짐)를 생성하기 위한 시스템 또는 메커니즘을 나타내는 개략도.

[0004] 도 4는 본 발명의 일 태양에 따른, 적어도 부분적으로 클라이언트 요청에 기초하는 안티 스팸 업데이트 시스템의 블록도.

[0005] 도 5는 본 발명의 일 태양에 따른, 적어도 부분적으로 클라이언트 요청에 기초하는 안티 스팸 업데이트 시스템의 블록도.

[0006] 도 6은 본 발명의 일 태양에 따른 예시 안티 스팸 룩업 웹 서비스의 개략도.

[0007] 도 7은 본 발명의 일 태양에 따른 적어도 증분적으로 스팸 필터들을 업데이트하는 것을 용이하게 하는 예시 방법을 나타내는 흐름도.

[0008] 도 8은 본 발명의 일 태양에 따른, 이전 필터로부터의 최소 업데이트 양 또는 변화량을 나타내는 필터들을 생성하는 것을 용이하게 하는 예시적인 방법을 나타내는 흐름도.

[0009] 도 9는 본 발명의 다양한 태양들을 구현하기 위한 예시 환경을 나타내는 도면.

[0010] <도면의 주요 부분에 대한 부호의 설명>

[0011] 100 : 안티 스팸 업데이트 시스템

[0012] 120 : 스팸 필터

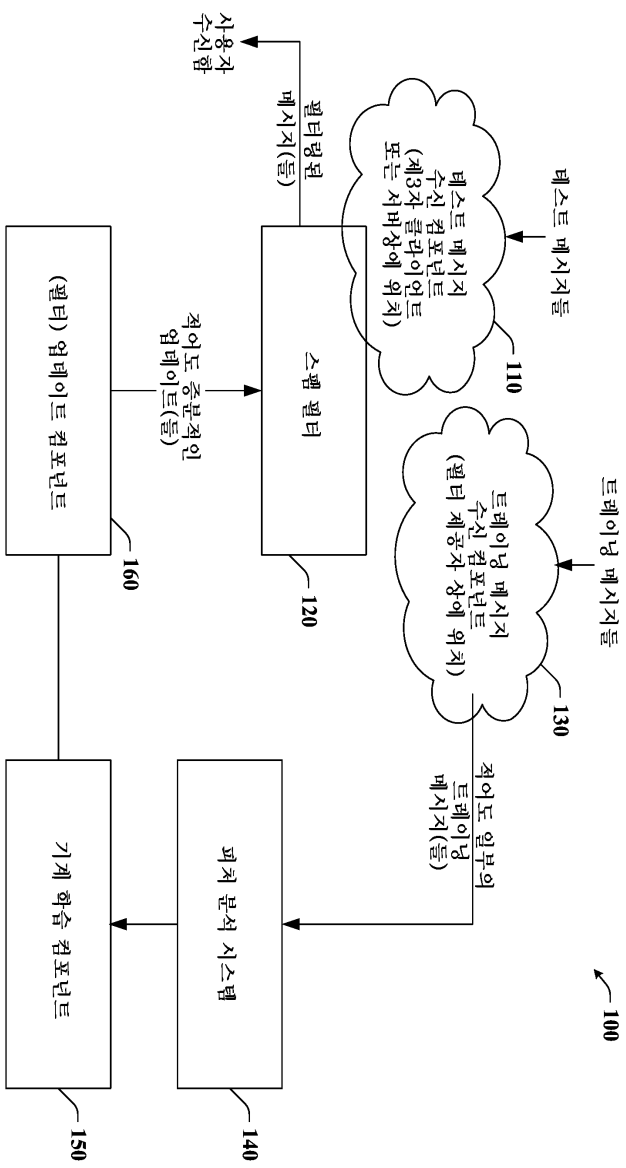
[0013] 140 : 피처 분석 시스템

[0014] 150 : 기계 학습 컴포넌트

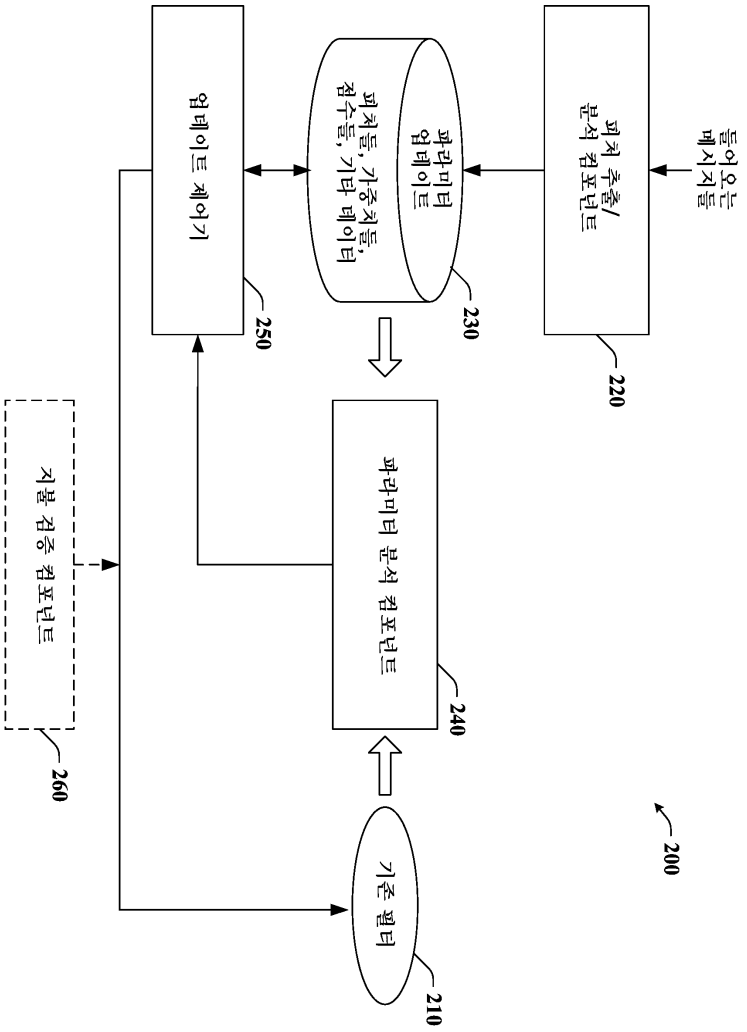
[0015] 160 : (필터) 업데이트 컴포넌트

도면

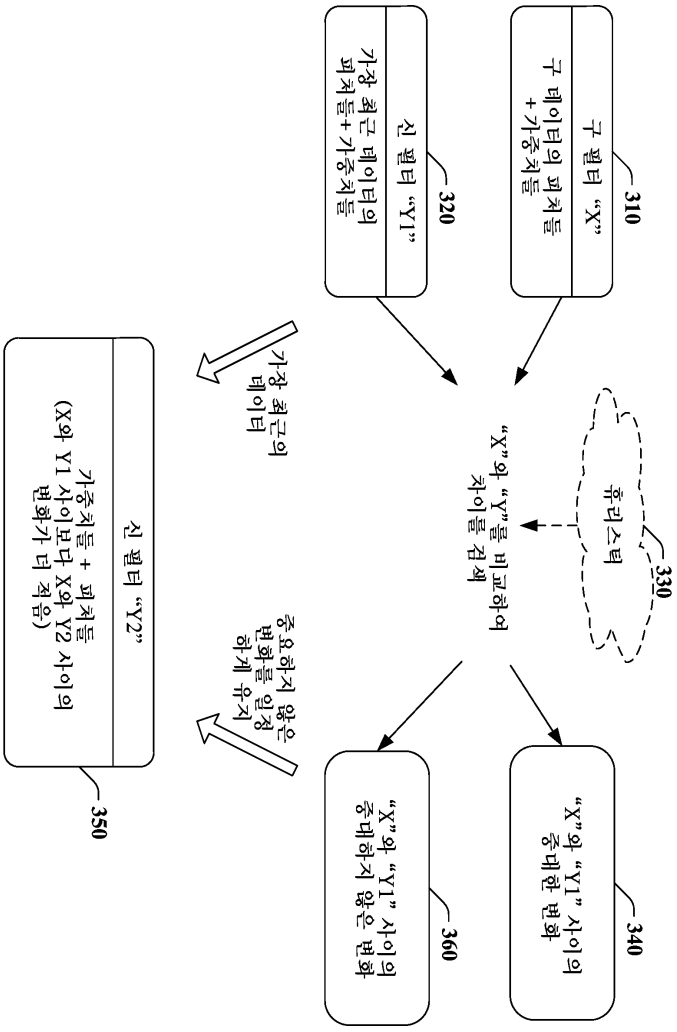
도면1


100
테스트 메시지들
테스트 메시지
수신 컴포넌트
(제3자 클라이언트
또는 서버상에 위치)
110
필터링된
메시지(들)
사용자
수신함
스팸 필터
120
트레이닝 메시지들
트레이닝 메시지
수신 컴포넌트
(필터 제공자 상에 위치)
130
적어도 일부의
트레이닝
메시지(들)
적어도 중요한인
업데이트(들)
(필터) 업데이트 컴포넌트
160
피처 분석 시스템
140
기계 학습 컴포넌트
150

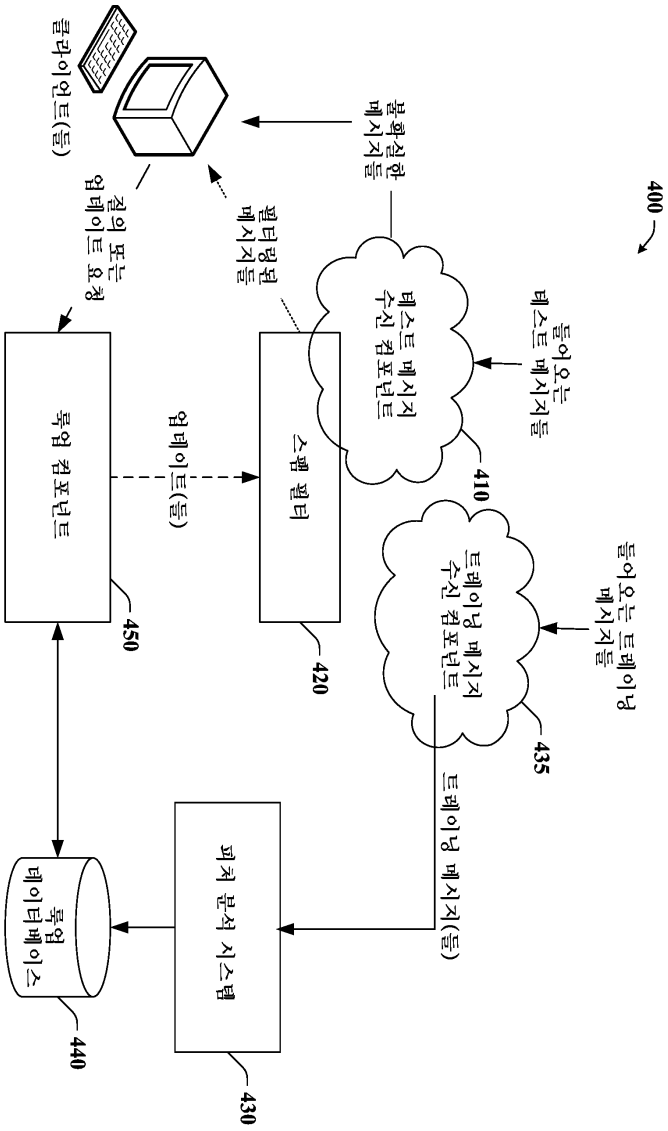
도면2


들어오는 메시지들
피처 추출/ 분석 컴포넌트
220
파라미터 업데이트
피처들, 가중치들, 점수들, 기타 데이터
230
업데이트 제어기
250
파라미터 분석 컴포넌트
240
기존 필터
210
지불 검증 컴포넌트
260
200

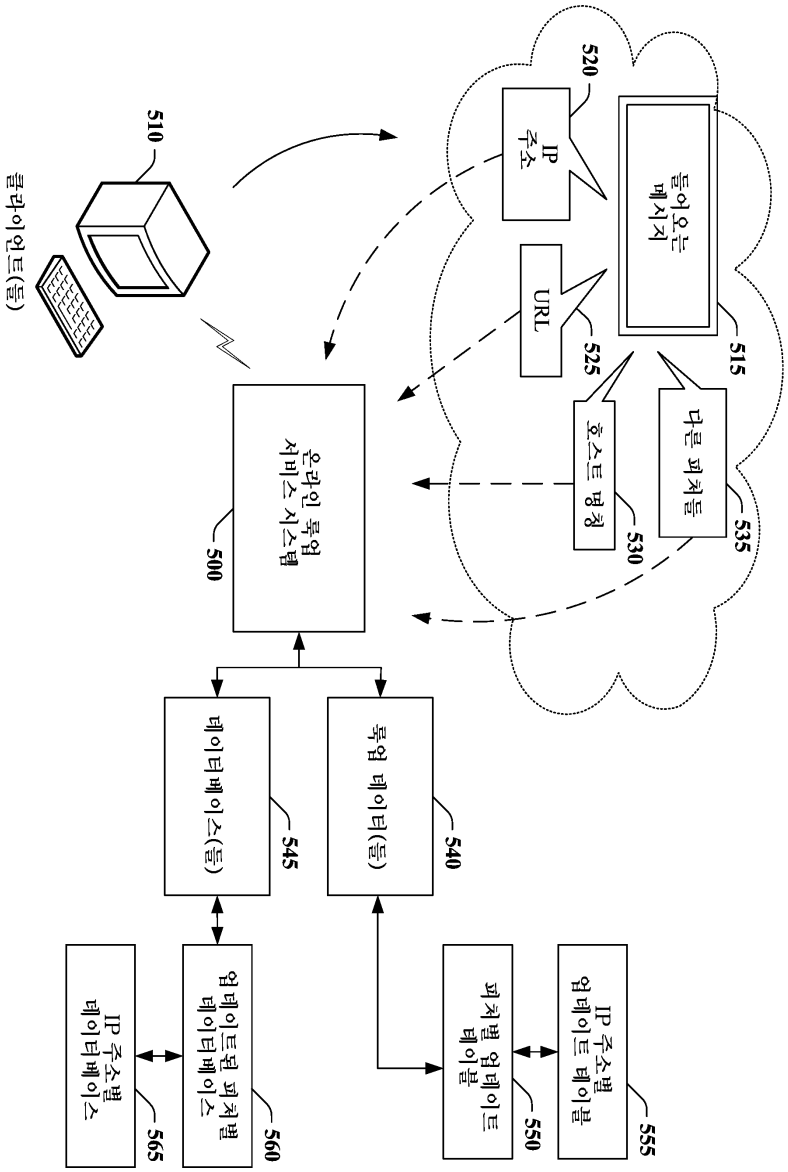
도면3


구 필터 "X"
구 데이터의 피처들 + 가중치들
310
신 필터 "Y1"
가장 최근 데이터의 피처들 + 가중치들
320
휴리스틱
330
"X"와 "Y"를 비교하여 차이를 검색
340
"X"와 "Y1" 사이의 중대한 변화
360
"X"와 "Y1" 사이의 중대하지 않은 변화
가장 최근의 데이터
중요하지 않은 변화를 일정하게 유지
신 필터 "Y2"
가중치들 + 피처들 (X와 Y1 사이보다 X와 Y2 사이의 변화가 더 적음)
350

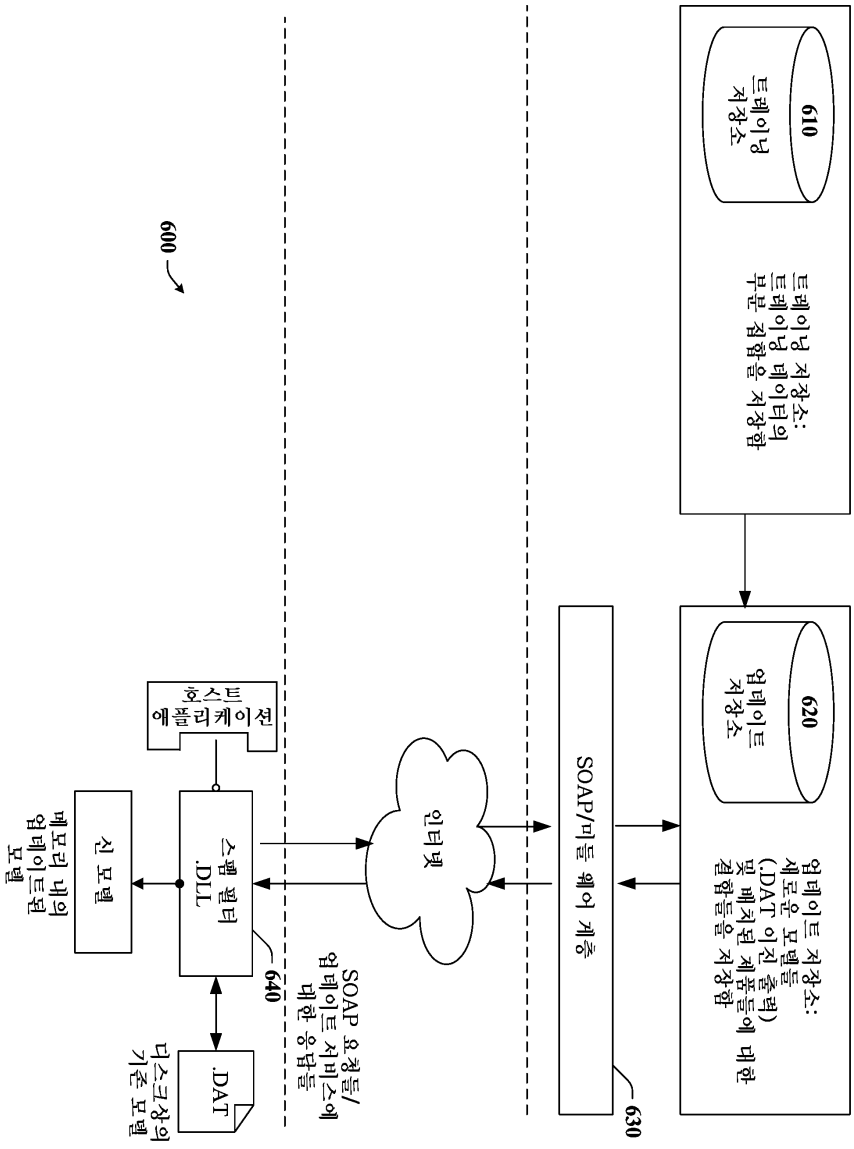
도면4


400
클라이언트(들)
불확실한 메시지들
필터링된 메시지들
쉘의 또는 업데이트 요청
들어오는 테스트 메시지들
테스트 메시지 수신 컴포넌트
410
들어오는 트레이닝 메시지들
트레이닝 메시지 수신 컴포넌트
435
트레이닝 메시지(들)
스팸 필터
420
업데이트(들)
룩업 컴포넌트
450
피처 분석 시스템
430
룩업 데이터베이스
440

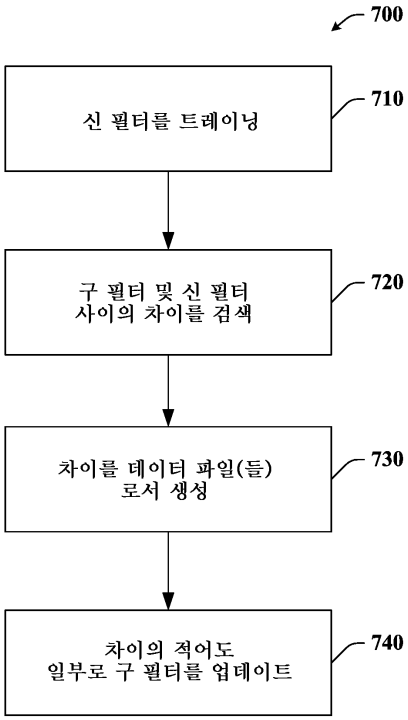
도면5


클라이언트(들)
510
IP 주소
520
들어오는 메시지
515
URL
525
호스트 명칭
530
다른 피처들
535
온라인 룩업 서비스 시스템
500
데이터베이스(들)
545
룩업 데이터(들)
540
IP 주소별 데이터베이스
565
업데이트된 피처별 데이터베이스
560
피처별 업데이트 테이블
550
IP 주소별 업데이트 테이블
555

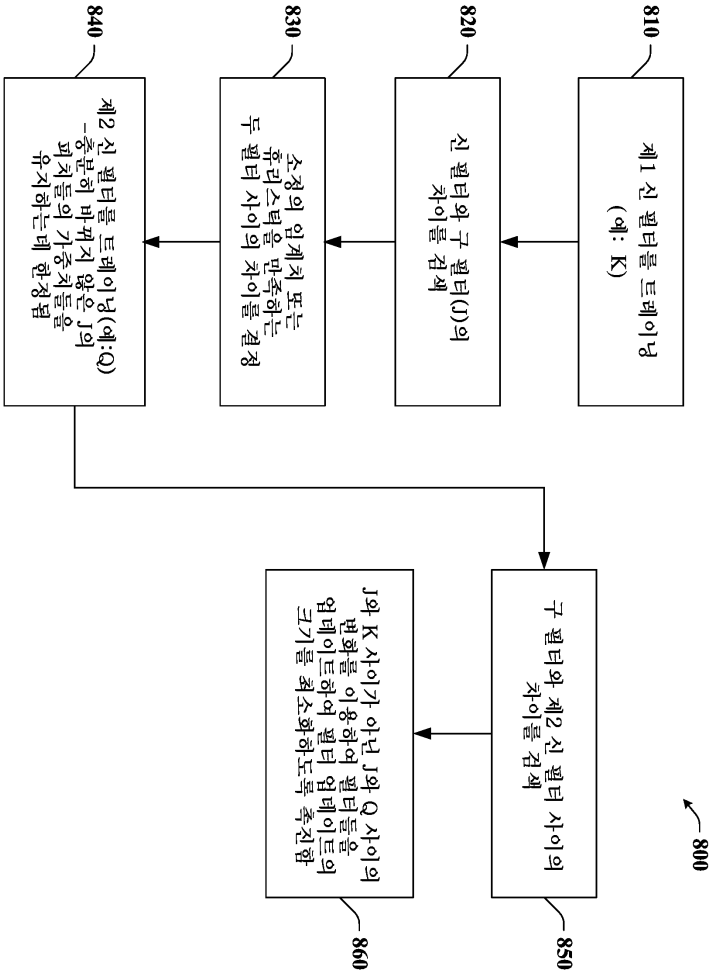
도면6


610
트레이닝
저장소
트레이닝 저장소:
트레이닝 데이터의
부분 집합을 저장함
620
업데이트
저장소
업데이트 저장소:
새로운 모델들
(.DAT 이진 출력)
및 패치된 제품들에 대한
결합들을 저장함
630
SOAP/미들 웨어 계층
인터넷
600
640
호스트
애플리케이션
스팸 필터
.DLL
신 모델
메모리 내의
업데이트된
모델
.DAT
디스크상의
기존 모델
SOAP 요청들/
업데이트 서비스에
대한 응답들

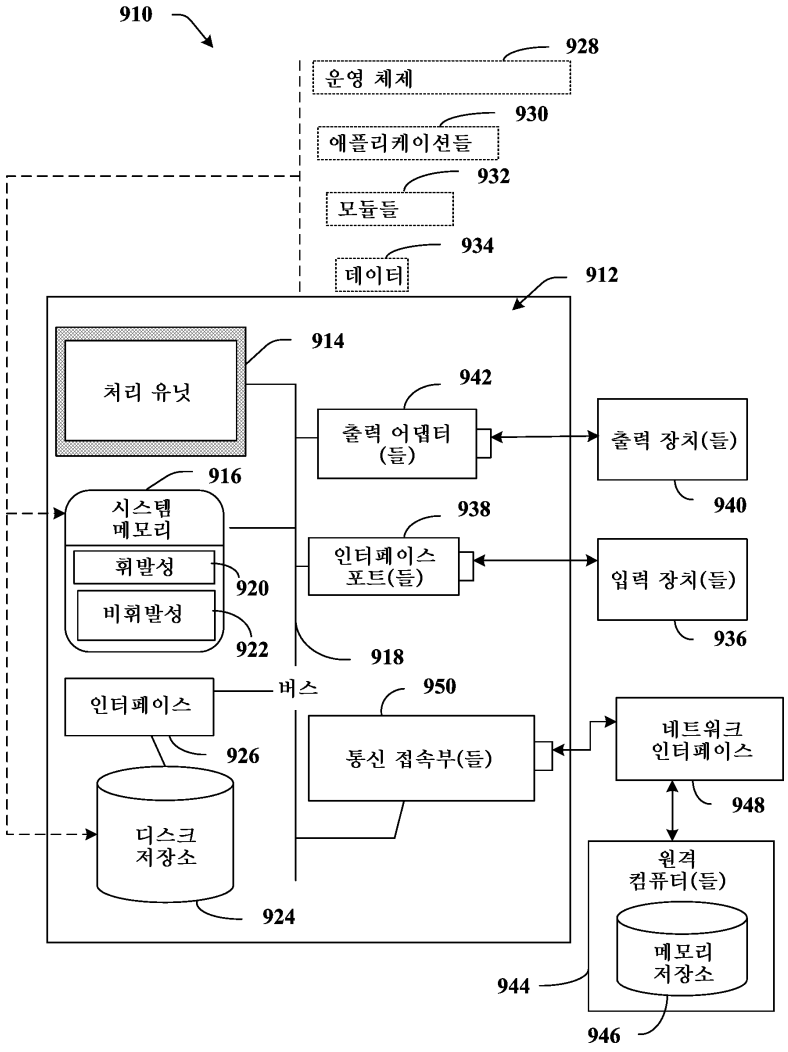
도면7


700
신 필터를 트레이닝
710
구 필터 및 신 필터
사이의 차이를 검색
720
차이를 데이터 파일(들)
로서 생성
730
차이의 적어도
일부로 구 필터를 업데이트
740

도면8


810
제1 신 필터를 트레이닝
(예: K)
820
신 필터와 구 필터(J)의
차이를 검색
830
소정의 임계치 또는
휴리스틱을 만족하는
두 필터 사이의 차이를 결정
840
제2 신 필터를 트레이닝(예:Q)
-충분히 바뀌지 않은 J의
피처들의 가중치들을
유지하는데 한정됨
850
구 필터와 제2 신 필터 사이의
차이를 검색
860
J와 K 사이가 아닌 J와 Q 사이의
변화를 이용하여 필터들을
업데이트하여 필터 업데이트의
크기를 최소화하도록 촉진함
800

도면9


910
928
운영 체제
930
에플리케이션들
932
모듈들
934
데이터
912
914
처리 유닛
942
출력 어댑터(들)
출력 장치(들)
940
916
시스템 메모리
휘발성
920
비휘발성
922
938
인터페이스 포트(들)
입력 장치(들)
936
918
버스
인터페이스
926
950
통신 접속부(들)
네트워크 인터페이스
948
디스크 저장소
924
원격 컴퓨터(들)
메모리 저장소
944
946