



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0112008
 (43) 공개일자 2014년09월22일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
 (21) 출원번호 10-2014-7014084
 (22) 출원일자(국제) 2012년10월26일
 심사청구일자 없음
 (85) 번역문제출일자 2014년05월26일
 (86) 국제출원번호 PCT/US2012/062156
 (87) 국제공개번호 WO 2013/063416
 국제공개일자 2013년05월02일
 (30) 우선권주장
 13/660,533 2012년10월25일 미국(US)
 61/551,833 2011년10월26일 미국(US)

(71) 출원인
 탑시 랩스, 아이앤씨.
 미국, 캘리포니아 94105, 샌프란시스코, 6층, 세
 컨드 스트리트 140
 (72) 발명자
 고쉬, 리샙, 아이어
 미국 94107 캘리포니아주 샌프란시스코 에이퍼티.
 108 브래넨 스트리트 548
 맨리, 스캇, 파크
 미국 94619 캘리포니아주 오클랜드 로슨 스트리트
 3024
 (74) 대리인
 양영준, 백만기, 정은진

전체 청구항 수 : 총 33 항

(54) 발명의 명칭 **소셜 네트워크 상에서의 감정 검출, 측정 및 정규화를 위한 시스템 및 방법**

(57) 요약

소셜 네트워크 상에서 특정 이벤트 또는 토픽에 대해 사용자들의 그룹에 의해 표현되는 감정들을 검출, 측정, 집계 및 정규화하여, 정규화된 감정들이 그러한 특정 이벤트 또는 토픽에 대한 일반 대중의 감정들을 진정으로 반영하게 하는 능력을 제공하기 위한 시스템들 및 방법들을 고려하는 새로운 접근법이 제안된다. 게다가, 소셜 네트워크 상에서 표현되는 개별 사용자의 수집 및 측정된 감정들은 또한 사용자의 표현 시에 개별 사용자의 감정을 진정으로 반영하기 위해 반영하는 기준선 감정에 대해 정규화될 수 있다.

특허청구의 범위

청구항 1

동작 시에, 소셜 네트워크 상에서의 복수의 사용자의 활동(activity)들에 대한 데이터를 수집하는 데이터 수집 엔진(data collection engine);

동작 시에, 상기 소셜 네트워크 상에서의 상기 사용자들의 활동들의 상기 수집된 데이터에서 표현되는 바와 같은 특정 이벤트 또는 토픽에 대한 상기 복수의 사용자 각각의 감정(sentiment)을 검출 및 측정하고, 상기 이벤트 또는 토픽에 대한 상기 복수의 사용자의 상기 검출된 감정들을 집계하고, 상기 사용자들의 상기 집계된 감정들을 기준선 감정(baseline sentiment)에 대해 정규화하여, 상기 정규화된 감정들이 상기 이벤트 또는 토픽에 대한 일반 대중의 감정들을 진정으로(truly) 반영하게 하는

감정 분석 엔진(sentiment analysis engine)

을 포함하는, 시스템.

청구항 2

제1항에 있어서,

상기 소셜 네트워크는 자신의 사용자들/멤버들이 서로 게시(post), 공유, 통신 및 상호작용하는 것을 가능하게 하는 공개적으로 액세스 가능한 웹-기반 플랫폼 또는 커뮤니티인, 시스템.

청구항 3

제1항에 있어서,

상기 소셜 네트워크는 페이스북, 구글+, 트위터, 링크드인(LinkedIn), 블로그들, 포럼들, 또는 임의의 다른 웹-기반 커뮤니티들 중 하나인, 시스템.

청구항 4

제1항에 있어서,

상기 소셜 미디어 네트워크 상에서의 상기 사용자의 활동들은 트윗(tweet)들, 상기 트윗들에 대한 응답들 및/또는 리트윗(re-tweet)들, 게시들, 다른 사용자들의 게시들에 대한 코멘트들, 의견들, 피드(feed)들, 접속들, 참조들, 다른 웹사이트들 또는 애플리케이션들에 대한 링크들, 또는 상기 소셜 네트워크 상에서의 임의의 다른 활동들 중 하나 이상을 포함하는, 시스템.

청구항 5

제1항에 있어서,

상기 데이터 수집 엔진은 상기 사용자들로부터 최근 활동 데이터를 수집하기 위해 상기 소셜 네트워크를 주기적으로 크롤링(crawling)함으로써 상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들에 대한 상기 데이터를 수집하는, 시스템.

청구항 6

제1항에 있어서,

상기 데이터 수집 엔진은 상기 소셜 네트워크에 의해 제공되는 애플리케이션 프로그래밍 인터페이스(application programming interface, API)를 이용함으로써 상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들에 대한 상기 데이터를 수집하는, 시스템.

청구항 7

제1항에 있어서,

상기 감정 분석 엔진은 사람들이 상기 소셜 미디어 네트워크 상에서 자신을 표현하는 방식들을 고려하는 다수의 감정 텍스트 스코어링 스킴(sentiment text scoring scheme)을 통해 상기 소셜 네트워크 상에서의 상기 사용자

들의 활동들의 상기 수집된 데이터에서 상기 사용자들에 의해 표현되는 상기 감정들을 검출 및 식별하는, 시스템.

청구항 8

제7항에 있어서,

상기 다수의 감정 텍스트 스코어링 스킴은 상기 소셜 네트워크 상에서 상기 사용자들에 의해 표현되는 상기 감정들을 검출하기 위해 상기 수집된 데이터에서 식별되는 감정 표현들의 고유(specific) 특성들을 이용하는, 시스템.

청구항 9

제1항에 있어서,

상기 감정 분석 엔진은 상기 사용자들의 의견들이 표현되는 상기 소셜 네트워크의 자연적인 편향(natural bias)을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는, 시스템.

청구항 10

제1항에 있어서,

상기 감정 분석 엔진은 상기 이벤트 또는 토픽의 속성을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는, 시스템.

청구항 11

제1항에 있어서,

상기 감정 분석 엔진은 상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들의 타이밍을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는, 방법.

청구항 12

제1항에 있어서,

상기 감정 분석 엔진은 상기 사용자들의 상기 감정들의 정규화된 측정에 기초하여 상기 이벤트 또는 토픽에 대한 소셜 감정 스코어(social sentiment score)를 계산하는, 시스템.

청구항 13

동작 시에, 소셜 네트워크 상에서의 사용자의 활동들에 대한 데이터를 수집하는 데이터 수집 엔진;

동작 시에, 상기 소셜 네트워크 상에서의 상기 사용자의 상기 활동들의 상기 수집된 데이터에서 표현되는 바와 같은 특정 이벤트 또는 토픽에 대한 상기 사용자의 감정을 검출 및 측정하고, 상기 사용자의 상기 측정된 감정을 기준선 감정에 대해 정규화하여, 상기 정규화된 감정이 상기 이벤트 또는 토픽에 대한 상기 사용자의 실제 감정을 진정으로 반영하게 하는

감정 분석 엔진

을 포함하는, 시스템.

청구항 14

제13항에 있어서,

상기 데이터 수집 엔진은 상기 사용자에 대한 활동 수집 스케줄(activity collection schedule)에 기초하여 상기 사용자로부터 상기 데이터를 선택적으로 수집하는, 시스템.

청구항 15

제13항에 있어서,

상기 데이터 수집 엔진은 상기 소셜 네트워크 상에서의 상기 사용자의 상기 활동들과 관련된 타임스탬프

(timestamp)들에 기초하여 시간에 따른 상기 사용자에게 대한 활동 분포 패턴(activity distribution pattern)을 확립하는, 시스템.

청구항 16

제13항에 있어서,

상기 데이터 수집 엔진은 특정 이벤트들의 발생 시에 상기 사용자가 최고로 활동적일 가능성이 있는지의 여부 및/또는 상기 사용자가 최고로 활동적일 가능성이 있는 시기를 결정하는, 시스템.

청구항 17

제13항에 있어서,

상기 감정 분석 엔진은 개별 사용자에게 링크된 다른 콘텐츠에서 표현되는 감정들에 대해 상기 사용자의 상기 측정된 감정을 정규화하는, 시스템.

청구항 18

제13항에 있어서,

상기 감정 분석 엔진은 상기 사용자의 자연적인 성향(natural tendency)을 고려하는 상기 기준선 감정에 대해 상기 사용자의 상기 측정된 감정들을 정규화하는, 시스템.

청구항 19

소셜 네트워크 상에서의 복수의 사용자의 활동들에 대한 데이터를 수집하는 단계;

상기 소셜 네트워크 상에서의 상기 사용자들의 활동들의 상기 수집된 데이터에서 표현되는 바와 같은 특정 이벤트 또는 토픽에 대한 상기 복수의 사용자 각각의 감정을 검출 및 측정하는 단계;

상기 이벤트 또는 토픽에 대한 상기 복수의 사용자의 상기 검출된 감정들을 집계하는 단계;

상기 사용자들의 상기 집계된 감정들을 기준선 감정에 대해 정규화하여, 상기 정규화된 감정들이 상기 이벤트 또는 토픽에 대한 일반 대중의 감정들을 진정으로 반영하게 하는 단계

를 포함하는, 방법.

청구항 20

제19항에 있어서,

상기 사용자들로부터 최근 활동 데이터를 수집하기 위해 상기 소셜 네트워크를 주기적으로 크롤링함으로써 상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들에 대한 상기 데이터를 수집하는 단계를 추가로 포함하는, 방법.

청구항 21

제19항에 있어서,

상기 소셜 네트워크에 의해 제공되는 애플리케이션 프로그래밍 인터페이스(API)를 이용함으로써 상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들에 대한 상기 데이터를 수집하는 단계를 추가로 포함하는, 방법.

청구항 22

제19항에 있어서,

사람들이 상기 소셜 미디어 네트워크 상에서 자신을 표현하는 방식들을 고려하는 다수의 감정 텍스트 스코어링 스킴을 통해 상기 소셜 네트워크 상에서의 상기 사용자들의 활동들의 상기 수집된 데이터에서 상기 사용자들에 의해 표현되는 상기 감정들을 검출 및 식별하는 단계를 추가로 포함하는, 방법.

청구항 23

제19항에 있어서,

상기 소셜 네트워크 상에서 상기 사용자들에 의해 표현되는 상기 감정들을 검출하기 위해 상기 수집된 데이터에서 식별되는 감정 표현들의 고유 특성들을 이용하는 단계를 추가로 포함하는, 방법.

청구항 24

제19항에 있어서,

상기 사용자들의 의견들이 표현되는 상기 소셜 네트워크의 자연적인 편향을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는 단계를 추가로 포함하는, 방법.

청구항 25

제19항에 있어서,

상기 이벤트 또는 토픽의 속성을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는 단계를 추가로 포함하는, 방법.

청구항 26

제19항에 있어서,

상기 소셜 네트워크 상에서의 상기 사용자들의 상기 활동들의 타이밍을 고려하는 상기 기준선 감정에 대해 상기 사용자들의 상기 집계된 감정들을 정규화하는 단계를 추가로 포함하는, 방법.

청구항 27

제19항에 있어서,

상기 사용자들의 상기 감정들의 정규화된 측정에 기초하여 상기 이벤트 또는 토픽에 대한 소셜 감정 스코어를 계산하는 단계를 추가로 포함하는, 방법.

청구항 28

소셜 네트워크 상에서의 사용자의 활동들에 대한 데이터를 수집하는 단계;

상기 소셜 네트워크 상에서의 상기 사용자의 상기 활동들의 상기 수집된 데이터에서 표현되는 바와 같은 특정 이벤트 또는 토픽에 대한 상기 사용자의 감정을 검출 및 측정하는 단계;

상기 사용자의 상기 측정된 감정을 기준선 감정에 대해 정규화하여, 상기 정규화된 감정이 상기 이벤트 또는 토픽에 대한 상기 사용자의 실제 감정을 진정으로 반영하게 하는 단계

를 포함하는, 방법.

청구항 29

제28항에 있어서,

상기 사용자에 대한 활동 수집 스케줄에 기초하여 상기 사용자로부터 상기 데이터를 선택적으로 수집하는 단계를 추가로 포함하는, 방법.

청구항 30

제28항에 있어서,

상기 소셜 네트워크 상에서의 상기 사용자의 상기 활동들과 관련된 타임스탬프들에 기초하여 시간에 따른 상기 사용자에 대한 활동 분포 패턴을 확립하는 단계를 추가로 포함하는, 방법.

청구항 31

제28항에 있어서,

특정 이벤트들의 발생 시에 상기 사용자가 최고로 활동적일 가능성이 있는지의 여부 및/또는 상기 사용자가 최

고로 활동적일 가능성이 있는 시기를 결정하는 단계를 추가로 포함하는, 방법.

청구항 32

제28항에 있어서,

개별 사용자에게 링크된 다른 콘텐츠에서 표현되는 감정들에 대해 상기 사용자의 상기 측정된 감정을 정규화하는 단계를 추가로 포함하는, 방법.

청구항 33

제28항에 있어서,

상기 사용자의 자연적인 성향을 고려하는 상기 기준선 감정에 대해 상기 사용자의 상기 측정된 감정들을 정규화하는 단계를 추가로 포함하는, 방법.

명세서

배경 기술

- [0001] 관련 출원들에 대한 상호 참조
- [0002] 본 출원은 2011년 10월 26일자로 출원된 미국 특허 출원 제61/551,833호 및 2012년 10월 25일자로 출원된 미국 특허 출원 제13/660,533호에 대해 우선권을 주장하며, 이로써 상기 출원들은 본 명세서에 참고로 포함된다.
- [0003] 페이스북, 트위터 및 구글 플러스와 같은 소셜 미디어 네트워크들은 최근에 웹-기반 통신 플랫폼들로서 기하급수적으로 성장하였다. 날마다 수억 명의 사람들이 다양한 형태의 소셜 미디어 네트워크들을 사용하여 서로 통신하고 접속을 유지하고 있다. 따라서, 트위터 상에 게시(post)되는 트윗(tweet)들과 같은 소셜 미디어 네트워크들 상에서의 사용자들로부터의 결과적인 활동(activity)들은 경이적이게 되었고, 다양한 종류의 측정들 및 분석을 위해 수집될 수 있다. 구체적으로, 이러한 사용자 활동 데이터는 소셜 네트워크들의 소셜 데이터 자원들로부터 이들 각각의 공개적으로 이용 가능한 애플리케이션 프로그래밍 인터페이스(Application Programming Interface, API)들을 통해 검색되고, 인덱싱되고, 처리되고, 추가적인 분석을 위해 국지적으로 저장될 수 있다.
- [0004] 실시간으로 수집되는 소셜 네트워크들로부터의 이러한 스트림 데이터는 시간에 따라 수집되고 저장되는 데이터와 함께 다양한 측정들 및 분석을 위한 기초를 제공한다. 측정들 및 분석을 위한 측정 기준들의 일부는
- [0005]
 - 언급들의 수 - 키워드, 용어 또는 링크에 대한 언급들의 총 수;
- [0006]
 - 유력자들에 의한 언급들의 수 - 유력한 사용자들에 의한 키워드, 용어 또는 링크에 대한 언급들의 총 수;
- [0007]
 - 중요한 게시들에 의한 언급들의 수 - 리트윗(retweet)되었거나 링크를 포함한 트윗들에 의한 키워드, 용어 또는 링크에 대한 언급들의 총 수;
- [0008]
 - 속도 - 이전의 시간 윈도우들(예를 들어, 7일)에서 키워드, 용어 또는 링크가 "오프겨지는(taking off)" 정도를 포함하지만, 이에 제한되지 않는다.
- [0009] 데이터의 콘텐츠에 대해 수행되는 전술한 측정들 및 분석에 더하여, 또한 소셜 네트워크들 상에서의 사용자들의 활동들(예로서, 트윗들 및 게시들)을 통해 표현되는 사용자들의 집계된 감정(sentiment)들을 분석하는 것도 중요하다. 비제한적인 예로서, 그러한 집계된 감정들은 특정 기간에 걸쳐 특정 토픽에 대해 사용자들의 그룹에 의해 표현되는 긍정적인, 중립적인 그리고 부정적인 트윗들의 백분율에 의해 측정될 수 있다. 소셜 네트워크들 상에서 표현되는 사용자들의 감정들의 그러한 측정은 그들의 견해들/의견들의 실시간 판단 척도들을 제공하지만, 그러한 측정은 최고로 활동적인, 따라서 소셜 네트워크들 상에서 자신의 기분을 표현할 가능성이 가장 큰 사용자들의 유형, 각각의 개별 사용자에게 의한 표현의 타이밍 및 선호되는 방식 등을 포함하지만 이에 제한되지 않는 다양한 인자들로 인해 편향될 수 있다. 따라서, 측정될 때, 소셜 네트워크들 상에서 특정 문제들 또는 이벤트들에 대해 표현된 사용자들의 감정들은 일반 대중의 감정들의 진정하고 정확한 반영이 아닐 수 있다.
- [0010] 전술한 관련 기술의 예들 및 이들과 관련된 제한들은 배타적이 아니라 예시적인 것으로 의도된다. 관련 기술에

대한 다른 제한들은 본 명세서를 읽고 도면들을 검토할 때 명백해질 것이다.

도면의 간단한 설명

[0011] <도 1>

도 1은 소셜 네트워크들 상에서의 감정 검출, 측정 및 정규화를 지원하기 위한 시스템 도면의 일례.

<도 2>

도 2는 소셜 네트워크들 상에서의 감정 검출, 측정 및 정규화를 지원하기 위한 프로세스의 흐름도의 일례.

발명을 실시하기 위한 구체적인 내용

[0012] 유사한 도면 부호들이 유사한 요소들을 지시하는 첨부 도면들에 본 접근법이 제한이 아니라 예로서 예시된다. 본 명세서에서 "일" 또는 "하나" 또는 "일부" 실시예(들)에 대한 언급들은 반드시 동일한 실시예를 언급하지는 않으며, 그러한 언급들은 적어도 하나를 의미한다는 점에 유의해야 한다.

[0013] 소셜 네트워크 상에서 특정 이벤트 또는 토픽에 대해 사용자들의 그룹에 의해 표현되는 감정들을 검출, 측정, 집계 및 정규화하여, 정규화된 감정들이 그러한 특정 이벤트 또는 토픽에 대한 일반 대중의 감정들을 진정으로 (truly) 반영하게 하는 능력을 제공하기 위한 시스템들 및 방법들을 고려하는 새로운 접근법이 제안된다. 여기서, 사용자들에 의해 표현된 집계되는 감정들의 측정은 사용자들의 의견들이 표현되는 소셜 네트워크의 자연적인 편향(natural bias), 토론의 이벤트 또는 토픽의 속성 및 소셜 네트워크 상에서의 사용자들의 활동들의 타이밍 중 하나 이상에 기초하여 정규화될 수 있다. 게다가, 소셜 네트워크 상에서 표현되는 개별 사용자의 수집 및 측정된 감정들은 또한 사용자의 표현 시에 사용자의 감정을 진정으로 반영하기 위해 각각의 개별 사용자의 자연적인 성향(natural tendency) 및/또는 개별 사용자에게 링크된 다른 콘텐츠에서 표현되는 감정들을 반영하는 기준선 감정(baseline sentiment)에 대해 정규화될 수 있다.

[0014] 아래에서 언급될 때, 소셜 미디어 네트워크 또는 소셜 네트워크는 그의 사용자들/멤버들이 서로 게시, 공유, 통신 및 상호작용하는 것을 가능하게 하는 임의의 공개적으로 액세스 가능한 웹-기반 플랫폼 또는 커뮤니티일 수 있다. 비제한적인 예들로서, 그러한 소셜 미디어 네트워크는 페이스북, 구글+, 트위터, 링크드인(LinkedIn), 블로그들, 포럼들 또는 임의의 다른 웹-기반 커뮤니티들일 수 있지만, 이에 제한되지 않는다.

[0015] 아래에서 언급될 때, 소셜 미디어 네트워크 상의 사용자의 활동들은 트윗, 트윗에 대한 응답 및/또는 리트윗, 게시, 다른 사용자의 게시에 대한 코멘트, 의견(예로서, 공감(Like)), 피드(feed), 접속(예로서, 다른 사용자들 친구로 추가함), 참조, 다른 웹사이트 또는 애플리케이션에 대한 링크, 또는 소셜 네트워크 상의 임의의 다른 활동들을 포함하지만, 이에 제한되지 않는다. 생성 시간이 콘텐츠와 항상 명확하게 관련되지 않는 수 있는 통상적인 웹 콘텐츠와 달리, 소셜 네트워크 상의 사용자의 활동들의 하나의 고유한 특성은 활동들 각각과 관련된 명확한 타임스탬프(time stamp)가 존재하여, 소셜 네트워크 상에서의 시간에 따른 사용자의 활동들의 패턴을 확립하는 것이 가능하게 된다는 것이다.

[0016] 도 1은 소셜 네트워크들 상에서의 감정 검출, 측정 및 정규화를 지원하기 위한 시스템 도면의 일례를 나타낸다. 도면들은 컴포넌트들을 기능적으로 분리된 것으로 도시하지만, 그러한 도시는 예시의 목적을 위한 것일 뿐이다. 이 도면에 도시된 컴포넌트들은 임의로 조합되거나, 개별 소프트웨어, 펌웨어 및/또는 하드웨어 컴포넌트들로 분할될 수 있다는 것이 명백할 것이다. 더구나, 그러한 컴포넌트들은 그들이 조합되거나 분할되는지에 관계없이 동일 호스트 또는 다수의 호스트 상에서 실행될 수 있다는 것도 명백할 것이며, 여기서 다수의 호스트는 하나 이상의 네트워크에 의해 접속될 수 있다.

[0017] 도 1의 예에서, 시스템(100)은 적어도 데이터 수집 엔진(data collection engine)(102) 및 감정 분석 엔진(sentiment analysis engine)(104)을 포함한다. 본 명세서에서 사용될 때, 엔진이라는 용어는 목적을 실현하는데 사용되는 소프트웨어, 펌웨어, 하드웨어 또는 다른 컴포넌트를 지칭한다. 엔진은 통상적으로 (보조 메모리 로도 지칭되는) 비휘발성 메모리 내에 저장되는 소프트웨어 명령어들을 포함할 것이다. 소프트웨어 명령어들이 실행될 때, 소프트웨어 명령어들의 적어도 서브셋이 프로세서에 의해 (주 메모리 로도 지칭되는) 메모리 내에 로딩된다. 이어서, 프로세서는 메모리 내의 소프트웨어 명령어들을 실행한다. 프로세서는 공유 프로세서, 전용 프로세서, 또는 공유 또는 전용 프로세서들의 조합일 수 있다. 통상적인 프로그램은 (I/O 디바이스들과 같은) 하드웨어 컴포넌트들에 대한 호출들을 포함할 것이며, 이는 통상적으로 드라이버들의 실행을 필요로 한다. 드라이버들은 엔진의 일부로서 간주될 수 있거나 간주되지 않을 수 있지만, 그 차이는 중요하지 않다.

- [0018] 도 1의 예에서, 엔진들 각각은 하나 이상의 호스팅 디바이스(호스트) 상에서 실행될 수 있다. 여기서, 호스트는 컴퓨팅 디바이스, 통신 디바이스, 저장 디바이스, 또는 소프트웨어 컴포넌트를 실행할 수 있는 임의의 전자 디바이스일 수 있다. 비제한적인 예들로서, 컴퓨팅 디바이스는 랩탑 PC, 데스크탑 PC, 태블릿 PC, 아이패드, 아이폰, 아이패드, 구글의 안드로이드 디바이스, PDA 또는 서버 기계일 수 있지만, 이에 제한되지 않는다. 저장 디바이스는 하드 디스크 드라이브, 플래시 메모리 디바이스 또는 임의의 휴대용 저장 디바이스일 수 있지만, 이에 제한되지 않는다. 통신 디바이스는 이동 전화일 수 있지만, 이에 제한되지 않는다.
- [0019] 도 1의 예에서, 데이터 수집 엔진(102) 및 감정 분석 엔진(104) 각각은 엔진들이 하나 이상의 통신 네트워크(도시되지 않음)를 통해 TCP/IP 프로토콜과 같은 특정 통신 프로토콜들에 따라 서로 통신하는 것을 가능하게 하는 소프트웨어 컴포넌트인 통신 인터페이스(도시되지 않음)를 구비한다. 여기서, 통신 네트워크들은 인터넷, 인트라넷, 광역 네트워크(WAN), 근거리 네트워크(LAN), 무선 네트워크, 블루투스, 와이파이 및 이동 통신 네트워크일 수 있지만, 이에 제한되지 않는다. 네트워크 및 통신 프로토콜들의 물리적 접속들은 당업자들에게 주지되어 있다.
- [0020] 도 1의 예에서, 데이터 수집 엔진(102)은 사용자들 각각으로부터 최근 활동 데이터를 수집하기 위해 소셜 네트워크를 주기적으로 크롤링(crawling)함으로써 소셜 네트워크 상에서의 사용자들의 활동들에 대한 데이터를 수집한다. 소셜 미디어 네트워크에서 액세스 가능한 막대한 양의 데이터가 주어지면, 데이터 수집 엔진(102)은 각각의 개별 사용자로부터 사용자에게 대한 활동 수집 스케줄(activity collection schedule)에 기초하여 선택적으로 데이터를 수집할 수 있다. 사용자의 활동들이 사용자의 활동 수집 스케줄에 따라 크롤링 시에 수집되지 않아야 할 경우, 데이터 수집 엔진(102)은 그 사용자와 관련된 콘텐츠를 생략(skip)하고, 자신의 활동이 그의 스케줄에 따라 활동이 수집되어야 하는 다음 사용자로 이동할 것이다. 데이터 수집 엔진(102)에 의한 그러한 선택적 데이터 수집은 수집되는 데이터의 신선도를 손상시키지 않고서 각각의 크롤링 시에 필요한 시간 및 자원들을 줄인다. 일부 실시예들에서, 데이터 수집 엔진(102)은 가능한 한 많은 데이터를 수집하기 위해 상이한 인터넷 주소들(IP들)로부터 오는 다수의 크롤러(crawler)를 실행하고 조정할 수 있다. 소셜 미디어 크롤링 엔진(106)은 또한 (HTTP) 요청마다 수집되는 새로운 데이터의 양을 최대화할 수 있다.
- [0021] 일부 실시예들에서, 데이터 수집 엔진(102)은 소셜 네트워크 상에서의 사용자의 활동들과 관련된 타임스탬프들에 기초하여 시간에 따른 각각의 사용자에게 대한 활동 분포 패턴/모델(activity distribution pattern/model)을 확립할 수 있다. 시간에 따른 그러한 활동 분포 패턴은 각각의 개별 사용자가 소셜 네트워크 상에서 최고로 또는 최소로 활동적인 시기 및 소셜 네트워크 상의 사용자의 활동들의 빈도를 반영할 수 있으며, 사용자에게 대한 활동 수집 스케줄을 설정하는 데 사용될 수 있다. 비제한적인 예로서, 사용자는 저녁 8-12시 사이에 소셜 네트워크 상에서 최고로 활동적일 수 있는 반면, 이른 아침 동안에 최소로 활동적일 수 있거나, 사용자는 주중에 아니라 주말에 최고로 활동적이다.
- [0022] 일부 실시예들에서, 데이터 수집 엔진(102)은 또한 사용자가 흥미를 갖는 특정 스포츠 이벤트 또는 제품 뉴스(예를 들어, 아이폰 출시)와 같은 특정 이벤트들의 발생 시에 각각의 개별 사용자가 최고로 활동적일 가능성이 있는지의 여부 및/또는 각각의 개별 사용자가 최고로 활동적일 가능성이 있는 시기를 결정할 수 있다. 대안으로서, 데이터 수집 엔진(102)은 사용자의 활동들이 소셜 네트워크 상에서 사용자와 접속된 하나 이상의 사용자의 친구의 활동들과 밀접하게 관련된 것으로 결정할 수 있다. 비제한적인 예로서, 사용자의 친구들 중 하나 이상이 활동하게 되는 경우, 예를 들어 흥미로운 토론을 시작하거나 온라인 게임에 참가하는 경우, 또한 사용자도 적극적으로 연루될 가능성이 있다.
- [0023] 일부 실시예들에서, 데이터 수집 엔진(102)은 소셜 네트워크에 의해 제공되는 애플리케이션 프로그래밍 인터페이스(API)를 이용함으로써 소셜 네트워크 상에서의 사용자들의 활동들에 대한 데이터를 수집할 수 있다. 비제한적인 예로서, 페이스북에 의해 제공되는 오픈그래프(OpenGraph) API는 소셜 네트워크 상에 다수의 자원(즉, 사용자의 활동들과 관련된 데이터)을 노출시키며, 모든 유형의 자원은 ID를 갖고, 그에 대해 이용 가능한 유형 및 방법들을 학습하기 위해 내성법(introspection method)이 이용될 수 있다. 여기서, ID들은 사용자 이름들 및/또는 번호들일 수 있다. 모든 자원들이 번호 ID들을 갖고, 일부만이 이름 ID들을 갖기 때문에, 번호 ID들만이 자원들을 참조하는 데 사용된다.
- [0024] 도 1의 예에서, 감정 분석 엔진(104)은 사람들이 일반적으로 소셜 미디어 네트워크 내에서 그리고 구체적으로 트위터 내에서 자신을 표현하는 방식들 및 그 뉘앙스들을 고려하는 다수의 감정 텍스트 스코어링 스킴(sentiment text scoring scheme)을 통해 특정 이벤트 또는 토픽에 관한/대한 소셜 네트워크 상에서의 사용자들의 활동들의 수집된 데이터에서 사용자들에 의해 표현된 감정들을 검출하고 식별한다. 트위터의 비제한적인 예

에서는, 전통적인 감정 측정 기술이 잘 처리하지 못하는 트윗의 140 문자 제약 내에서 사람들이 자신을 표현하는 방식에서의 큰 차이가 존재한다. 데이터 수집 엔진(102)에 의해 실시간으로 수집되고 시간에 따라 저장된 대량의 데이터의 분석 및 테스트에 기초하여, 감정 분석 엔진(104)은 트윗들 내의 다수의 "트위터리즘(twitterism)", 즉 사람들이 특정 이벤트 또는 사물들에 대해 느끼는 방식을 나타내는 것은 물론, 사람들이 트위터와 같은 소셜 네트워크 상에서 트윗들을 사용하여 자신을 표현하는 방식에 고유한 수집된 데이터 내의 감정 표현들의 고유(specific) 특성들을 식별할 수 있다. 이러한 식별된 감정 표현들의 특성들은 소셜 네트워크 상에서 사용자들에 의해 표현되는 감정들을 검출하기 위해 다수의 감정 텍스트 스코어링 스킴에 의해 이용된다. 여기서, 각각의 사용자의 감정은 매우 긍정적, 긍정적, 평면적, 부정적, 매우 부정적인 것으로서 특성화될 수 있다.

[0025] 사용자들의 감정들이 사용자들의 수집된 활동들에 기초하여 검출되면, 감정 분석 엔진(104)은 특정 이벤트 또는 토픽에 대한 사용자들의 감정들(긍정적 또는 부정적인 감정들)을 평가하고 집계한다. 비제한적인 예로서, 새로운 아이폰의 발매 시기 쯤에 트위터 상의 아이폰 관련 트윗들을 분석하는 것은 사용자들의 21%가 긍정적인 반면에 사용자들의 18%가 부정적임을 보여줄 수 있다. 기간이 발매 후의 일주일 또는 한 달로 확장되는 경우, 사용자들이 새로운 아이폰에 대해 더 많은 시간의 경험을 가짐에 따라, 소셜 감정 스코어(social sentiment score)는 상이한 감정 스코어(긍정적 또는 부정적인 사용자들의 더 높은 백분율)를 지시할 수 있다.

[0026] 일부 실시예들에서, 감정 분석 엔진(104)은 사용자들의 의견들이 표현되는 소셜 네트워크의 자연적인 편향, 토론의 이벤트 또는 토픽의 속성, 및 소셜 네트워크 상에서의 사용자들의 활동들의 타이밍을 포함하지만 이에 제한되지 않는 하나 이상의 인자/편향을 고려하는 기준선 감정에 대해 사용자들의 집계된 감정들 및/또는 각각의 개별 사용자의 감정을 정규화한다. 여기서, 감정 분석 엔진(104)은 중간, 평균, 표준 편차, 일관성 또는 이러한 척도들의 임의의 조합과 같은 다양한 통계 척도들을 사용하여, 시간에 따른 사용자들의 측정된 감정들을 정규화할 수 있다. 그러한 감정 정규화는 특정 이벤트에 대한 각각의 개별 사용자 및/또는 일반 대중의 감정의 정확한 척도를 얻는 데 필요하다. 게다가, 감정 분석 엔진(104)은 각각의 개별 사용자의 자연적인 성향 및/또는 개별 사용자에게 링크된 다른 콘텐츠에서 표현되는 감정들에 대해 각각의 개별 사용자의 측정된 감정을 정규화할 수 있다.

[0027] 일부 실시예들에서, 감정 분석 엔진(104)은 각각의 개별 사용자 또는 그룹으로서의 사용자들의 감정들의 정규화된 측정에 기초하여 이벤트 또는 토픽에 대한 소셜 감정 스코어를 계산한다. 여기서, 이벤트에 대한 소셜 감정 스코어는 현재 이벤트에 대해 그리고/또는 (분석되는 사용자들의 활동들의 타임스탬프들에 의존하는) 특정 기간에 걸쳐 소셜 네트워크 상에서 표현된 개별 사용자 또는 사용자들의 정규화된 감정들을 나타내며, 그러한 소셜 감정 스코어는 개별 사용자의 진정한 감정 또는 일반 대중의 감정들을 반영한다.

[0028] 새로운 아이폰의 발매 쯤에 사용자들의 감정들을 분석하는 이전 예에서, 트위터 상의 사용자들의 트윗들에 기초하는 측정된 감정은 아이폰 발매에 대해 약간 긍정적일 뿐이다(21% 긍정 대 18% 부정). 그러나, 트위터 상에서 표현되는 감정들은 일반 대중의 감정들보다 부정적인 경향이 있으므로, 약간 긍정적인 감정 판독은 감정 분석 엔진(104)에 의해 트위터의 부정적 편향에 대해 정규화될 때 사실상 훨씬 더 긍정적이다.

[0029] 다른 비제한적인 예에 대해, 트위터 상에서 사용자들에 의해 표현되는 가장 강한 부정적 감정은 정치와 관련된 것들에 대한 것인 경향이 있는 반면, 가장 강한 긍정적 감정은 부정적 감정만큼 강하지 않으며, 여행, 사진 등과 같은 비논쟁적인 토픽들에 집중된다. 따라서, 감정 분석 엔진(104)에 의해 측정되는 감정 스코어들은 이러한 지식을 염두에 두고 정규화되어야 하며, 정치 이벤트에 대한 약간 긍정적인 판독은 정치적 용어들에 대한 대부분의 감정들이 압도적으로 부정적이라는 사실과 함께 정규화될 때 이벤트가 매우 양호하게 수용된다는 것을 실제로 지시할 수 있다.

[0030] 다른 비제한적인 예에 대해, 사용자 #1이 단어들/구절들에 대한 그의 선택에서 더 긍정적인 경향이 있는 반면(예를 들어, 그가 항상 "훌륭하다"고 말하는 반면), 사용자 #2가 단어들/구절들에 대한 그의 선택에서 더 유보적인 경향이 있는 경우(예를 들어, 그가 항상 "오케이(ok)"라고 말하는 경우), 사용자 #2에 의한 긍정적인 표현(예로서, "훌륭하다")은 그의 부정적인 편향에 대해 감정 분석 엔진(104)에 의해 정규화될 때 사실상 매우 긍정적인 반면, 사용자 #1에 의한 동일한 표현은 그의 긍정적인 편향에 대해 정규화될 때 단지 중립적일 수 있다.

[0031] 도 2는 소셜 네트워크들 상에서의 감정 검출, 측정 및 정규화를 지원하기 위한 프로세스의 흐름도의 일례를 나타낸다. 이 도면은 예시의 목적을 위해 기능 단계들을 특정 순서로 도시하지만, 프로세스는 단계들의 임의의 특정 순서 또는 배열로 제한되지 않는다. 관련 분야의 당업자는 이 도면에 도시된 다양한 단계들이 다양한 방식으로 생략, 재배열, 조합 및/또는 적응될 수 있다는 것을 알 것이다.

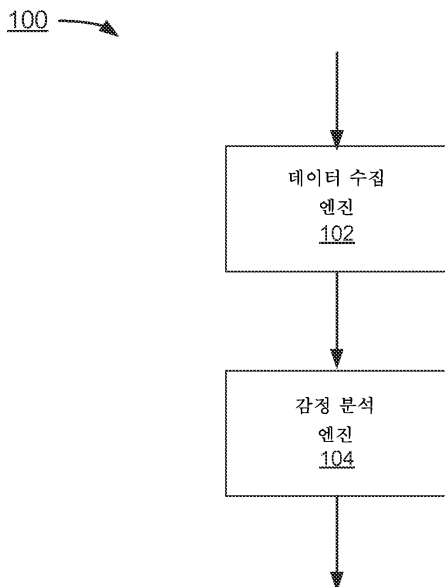
[0032] 도 2의 예에서, 흐름도(200)는 블록 202에서 시작되며, 여기서 소셜 네트워크 상에서의 복수의 사용자의 활동들에 대한 데이터가 수집된다. 흐름도(200)는 블록 204로 계속되며, 여기서 소셜 네트워크 상에서의 사용자들의 활동들의 수집된 데이터에서 표현되는 바와 같은 특정 이벤트 또는 토픽에 대한 사용자들 각각의 감정이 검출되고 측정된다. 흐름도(200)는 블록 206으로 계속되며, 여기서 이벤트 또는 토픽에 대한 복수의 사용자의 검출된 감정들이 선택적으로 집계된다. 흐름도(200)는 블록 208에서 종료되며, 여기서 사용자의 집계된 감정들 및/또는 각각의 개별 사용자의 감정이 기준선 감정에 대해 정규화되어, 정규화된 감정들이 이벤트 또는 토픽에 대한 일반 대중 및/또는 개별 사용자의 감정들을 진정으로 반영하게 된다.

[0033] 컴퓨터 분야의 당업자들에게 명백한 바와 같이, 일 실시예는 본 발명의 교시 내용에 따라 프로그래밍된 전통적인 범용 또는 특수 디지털 컴퓨터 또는 마이크로프로세서(들)를 사용하여 구현될 수 있다. 소프트웨어 분야의 당업자들에게 명백한 바와 같이, 본 발명의 교시 내용에 기초하여, 적절한 소프트웨어 코딩이 숙련된 프로그래머들에 의해 쉽게 준비될 수 있다. 당업자들에게 쉽게 명백해지는 바와 같이, 본 발명은 또한 집적 회로들의 준비에 의해 또는 전통적인 컴포넌트 회로들의 적절한 네트워크를 상호접속함으로써 구현될 수 있다.

[0034] 일 실시예는 본 명세서에서 제공되는 임의의 특징을 수행하도록 하나 이상의 호스트를 프로그래밍하는 데 사용될 수 있는 명령어들을 그 상에/내에 저장한 기계 판독가능 매체(매체들)인 컴퓨터 프로그램 제품을 포함한다. 기계 판독가능 매체는 플로피 디스크, 광 디스크, DVD, CD-ROM, 마이크로 드라이브 및 광자기 디스크를 포함하는 하나 이상의 유형의 디스크들, ROM, RAM, EPROM, EEPROM, DRAM, VRAM, 플래시 메모리 디바이스, 자기 또는 광 카드, (분자 메모리 IC들을 포함하는) 나노시스템, 또는 명령어들 및/또는 데이터를 저장하는 데 적합한 임의의 유형의 매체 또는 디바이스를 포함할 수 있지만, 이에 제한되지 않는다. 임의의 하나의 컴퓨터 판독가능 매체(매체들) 상에 저장될 때, 본 발명은 범용/특수 컴퓨터 또는 마이크로프로세서의 하드웨어를 제어하기 위한 그리고 컴퓨터 또는 마이크로프로세서로 하여금 본 발명의 결과들을 이용하여 사람 관찰자 또는 다른 메커니즘과 상호작용하는 것을 가능하게 하기 위한 소프트웨어를 포함한다. 그러한 소프트웨어는 디바이스 드라이버, 운영 체제, 실행 환경/컨테이너 및 애플리케이션을 포함할 수 있지만, 이에 제한되지 않는다.

도면

도면1



도면2

200 →

