

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2014-0034994 (43) 공개일자 2014년03월21일
(51) 국제특허분류(Int. Cl.) G06Q 50/22 (2012.01) G06Q 50/30 (2012.01)		(71) 출원인 고려대학교 산학협력단 서울시 성북구 안암로 145 (안암동)
(21) 출원번호 10-2012-0100643		(72) 발명자 김경현 서울특별시 서초구 잠원동 신반포한신아파트 25차 334동 304호
(22) 출원일자 2012년09월11일 심사청구일자 2012년09월11일		이형우 대전 유성구 엑스포로 448, 206동 304호 (전민동, 엑스포아파트)
		김의기 서울 성북구 개운사길 79, 101호 (안암동5가)
		(74) 대리인 특허법인엠에이피에스

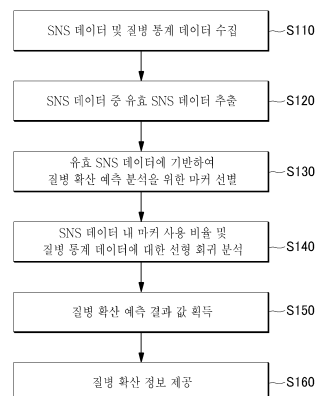
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **소셜 네트워크 서비스 데이터에 기반한 질병 확산 상황 예측 장치 및 방법**

(57) 요약

질병 확산 상황 예측 시, 기설정된 질병에 대한 의사 환자 분을 데이터를 수집하고, 기설정된 질병에 대한 의사 환자 분을 데이터를 수집하여 저장하고, 원본 SNS 데이터를 수집하여 저장하고, 저장된 SNS 데이터로부터 질병과 관련된 기설정된 기준 단어 중 적어도 하나가 포함된 유효 SNS 데이터를 추출하고, 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 의사 환자 분을 데이터와의 상관도에 기초하여 유효 SNS 데이터의 단어들로부터 마커를 결정하고, 저장된 SNS 데이터 내에서의 마커의 사용 비율을 산출하고, 마커의 사용 비율 및 의사 환자 분을 데이터에 기초하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출한다.

대표도 - 도10



특허청구의 범위

청구항 1

질병 확산 상황 예측 장치에 있어서,

질병 통계 서버가 제공하는 기설정된 질병에 대한 의사 환자 분율 데이터를 수집하여 저장하는 질병 통계 자료 수집부;

소셜 네트워크 서비스(SNS, Social Network Service) 서버가 제공하는 원본 SNS 데이터를 수집하여 저장하는 SNS 자료 수집부;

상기 저장된 SNS 데이터로부터 상기 질병과 관련된 기설정된 기준 단어 중 적어도 하나가 포함된 유효 SNS 데이터를 추출하는 유효 데이터 추출부;

상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 상기 의사 환자 분율 데이터와의 상관도에 기초하여 상기 유효 SNS 데이터의 단어들로부터 마커를 결정하는 마커 선별부;

기설정된 기간의 상기 저장된 SNS 데이터로부터 상기 마커의 사용 횟수 및 상기 기설정된 기간의 상기 저장된 SNS 데이터의 총 개수를 포함하는 마커 매트릭스를 생성하는 마커 매트릭스 생성부;

상기 마커 매트릭스를 이용하여 상기 저장된 SNS 데이터 내에서의 상기 마커의 사용 비율을 산출하고, 상기 마커의 사용 비율 및 상기 의사 환자 분율 데이터에 기초하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출하는 질병 확산 예측부; 및

상기 산출된 질병 확산 예측 값에 기초한 질병 확산 정보를 생성하여 제공하는 질병 확산 정보 제공부를 포함하는, 질병 확산 상황 예측 장치.

청구항 2

제 1 항에 있어서,

상기 마커 선별부는,

상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 상기 의사 환자 분율 데이터에 대해 LASSO(Least-Absolute Shrinkage And Selection Operator) 분석을 수행하고,

상기 LASSO 분석을 통해 산출된 상기 유효 SNS 데이터에 포함된 단어 별 상관 계수에 기초하여 상기 마커를 결정하는 질병 확산 상황 예측 장치.

청구항 3

제 1 항에 있어서,

상기 마커 선별부는,

상기 유효 SNS 데이터의 단어 별로 기본형을 판단하여, 상기 기본형 별 총 사용 횟수를 상기 단어 사용 빈도로 산출하는, 질병 확산 상황 예측 장치.

청구항 4

제 1 항에 있어서,

상기 질병 확산 예측부는,

상기 기설정된 기간의 상기 저장된 SNS 데이터에 대한 상기 마커의 총 사용 횟수 및 상기 기설정된 기간의 상기 저장된 SNS 데이터의 총 개수에 기초하여 마커 사용 비율 값을 산출하고,

상기 마커 사용 비율 값 및 상기 의사 환자 분율 데이터에 대해 선형 회귀(linear regression) 분석을 수행하여 상기 질병 확산 예측 값을 산출하는, 질병 확산 상황 예측 장치.

청구항 5

제 4 항에 있어서,

상기 질병 확산 예측부는,

기설정된 제 1 기간의 상기 저장된 SNS 데이터 및 상기 의사 환자 분율 데이터를 상기 제 1 기간 이전의 제 2 기간의 데이터로서 중복 사용하여 상기 선형 회귀 분석을 수행하는, 질병 확산 상황 예측 장치.

청구항 6

제 1 항에 있어서,

상기 SNS 자료 수집부는,

상기 원본 SNS 데이터 중 기설정된 추출 조건을 만족하는 SNS 데이터를 선별하여 수집하되,

상기 추출 조건은,

기설정된 특정 언어의 사용, 하이퍼링크(hyper link)의 불포함, 기설정된 스팸(SPAM) 단어의 불포함, 및 SNS 계정 별로 계정 사용자 이외의 사용자에게 의해 작성된 데이터 불포함 중 적어도 하나를 포함하는 것인, 질병 확산 상황 예측 장치.

청구항 7

제 1 항에 있어서,

상기 유효 데이터 추출부는,

상기 원본 SNS 데이터 중 상기 기준 단어와 관련하여 기설정된 안부성 문장 또는 구 또는 단어를 포함하는 SNS 데이터의 제거, 및 상기 기준 단어와의 동음이의어를 포함하는 SNS 데이터의 제거 중 적어도 하나의 제거를 처리하는, 질병 확산 상황 예측 장치.

청구항 8

제 1 항에 있어서,

상기 질병 확산 정보 제공부는,

상기 질병 확산 예측 값의 날짜 별 변화를 그래프 형태로 출력하는 질병 확산 상황 예측 장치.

청구항 9

질병 확산 상황 예측 장치를 통한 질병 확산 상황 예측 방법에 있어서,

(a) 소셜 네트워크 서비스(SNS, Social Network Service) 서버가 제공하는 원본 SNS 데이터 및 질병 통계 서버가 제공하는 기설정된 질병에 대한 의사 환자 분율 데이터를 각각 수집하여 저장하는 단계;

(b) 상기 저장된 SNS 데이터로부터 질병과 관련된 기설정된 기준 단어 중 적어도 하나를 포함하는 유효 SNS 데이터를 추출하는 단계;

(c) 상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 상기 의사 환자 분율 데이터와의 상관도에 기초하여 상기 유효 SNS 데이터의 단어들로부터 마커를 결정하는 단계;

(d) 기설정된 기간 단위 내의 상기 저장된 SNS 데이터 내에서의 상기 마커의 사용 비율 값을 산출하는 단계; 및

(e) 상기 마커의 사용 비율 값 및 상기 의사 환자 분율 데이터에 대해 선형 회귀(linear regression) 분석 처리하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출하는 단계; 및

(f) 상기 산출된 질병 확산 예측 값에 기초한 질병 확산 정보를 생성하여 사용자 단말에 제공하는 단계를 포함하는 질병 확산 상황 예측 방법.

청구항 10

제 9 항에 있어서,

상기 (c) 단계는,

(c-1) 상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도를 산출하는 단계;

(c-2) 상기 단어 별 사용 빈도와 상기 의사 환자 분을 데이터에 대해 LASSO(Least-Absolute Shrinkage And Selection Operator) 분석 처리하는 단계; 및

(c-3) 상기 LASSO 분석 처리의 결과 산출된 상기 유효 SNS 데이터에 포함된 단어 별 상관 계수에 기초하여 상기 마커를 결정하는 단계를 포함하는, 질병 확산 상황 예측 방법.

청구항 11

제 9 항에 있어서,

상기 (d) 단계는,

기설정된 기간 단위 내의 상기 저장된 SNS 데이터에 대한 상기 마커의 총 사용 횟수 및 상기 저장된 SNS 데이터의 총 개수에 기초하여 마커 사용 비율 값을 산출하는, 질병 확산 상황 예측 방법.

청구항 12

제 9 항에 있어서,

상기 (d) 단계 이전에,

상기 기설정된 기간의 상기 저장된 SNS 데이터로부터 상기 마커의 사용 횟수 및 상기 기설정된 기간의 상기 저장된 SNS 데이터의 총 개수를 포함하는 마커 매트릭스를 생성하는 단계를 더 포함하되,

상기 (d) 단계는,

상기 마커 매트릭스를 이용하여 상기 마커의 사용 비율 값을 산출하는, 질병 확산 상황 예측 방법.

청구항 13

제 9 항에 있어서,

상기 (a) 단계는,

상기 원본 SNS 데이터 중 기설정된 특정 언어의 사용, 하이퍼링크(hyper link)의 불포함, 기설정된 스팸(SPAM) 단어의 불포함, 및 SNS 계정 별로 계정 사용자 이외의 사용자에게 의해 작성된 데이터 불포함 중 적어도 하나의 조건을 만족하는 SNS 데이터를 수집하는, 질병 확산 상황 예측 방법.

청구항 14

제 9 항에 있어서,

상기 (b) 단계는,

상기 원본 SNS 데이터 중 상기 기준 단어와 관련하여 기설정된 안부성 문장 또는 구 또는 단어를 포함하는 SNS 데이터의 제거, 및 상기 기준 단어와의 동음이의어를 포함하는 SNS 데이터의 제거 중 적어도 하나의 제거를 처리하여 상기 유효 SNS 데이터를 추출하는, 질병 확산 상황 예측 방법.

청구항 15

제 9 항에 있어서,

상기 (f) 단계는,

상기 질병 확산 예측 값의 날짜 별 변화를 그래프 형태로 출력하는 질병 확산 상황 예측 방법.

명세서

기술 분야

[0001] 본 발명은 소셜 네트워크 서비스(SNS, Social Network Service) 데이터에 기반하여 전염병의 확산 상황을 추정하는 질병 확산 상황 예측 장치 및 방법에 관한 것이다.

배경 기술

[0002] 전염병(즉, 전염성 질병)의 출현은 중대한 보건 문제로서, 기후 및 환경 문제 등과 더불어 세계적인 이슈(issue)이다. 최근 발생한 ‘신종 플루’, ‘SARS’, ‘구제역’ 및 ‘조류 인플루엔자’의 사례에서 보는 것처럼, 운송 수단의 발달과 글로벌화로 인해 질병의 전염은 국경을 무시하고 빠른 시간 내에 전 세계적인 재앙으로 확산될 가능성이 매우 높다.

[0003] 이전부터 전염병을 대비하기 위한 생물학적 백신 개발이 활발하게 진행되고 있으나, 사실상 전염병에 대한 대응책으로는 조기 예방이 가장 중요하다. 이는, 백신 제조에는 인플루엔자(즉, 독감)의 경우 4~6개월 가량의 긴 시간이 필요하므로 전염병이 창궐한 경우 백신 제조 기간 동안에 질병이 상당히 확산 되었을 가능성이 높기 때문이다.

[0004] 한편, 전염병의 일례로서 인플루엔자의 경우, 질병 관리 본부 등의 기관에서는 질병 확산 조기 예방 차원에서 인플루엔자 표본 감시 사업을 진행하며, 인플루엔자의 유행 수준을 주 단위로 발표하고 있다. 그런데, 전염병의 원인 병원체를 동정하고 확진하는데 소요되는 시간에 비해 병원체의 변이 진화 및 확산은 훨씬 빠른 속도로 진행되므로, 이러한 방식으로 질병 확산에 대응하는 예방 조치를 취하는 데에는 어려움이 있다. 실제로 인플루엔자 의사 환자(ILI, influenza-like illness) 정보의 취합과 분석에는 많은 시간이 소요되므로 인플루엔자 표본 감시 보고서는 보고일로부터 대략 2주 전에 발생한 환자들에 대한 정보를 포함하고 있는 실정이다.

[0005] 따라서, 전염병의 특정한 유행 시즌과 관계없이 지속적으로 바이러스 등의 병원체의 발생 상태를 감시하고 확산 동태를 파악 및 예측할 수 있는 방법이 필요하다.

[0006] 이와 관련하여, 대한민국공개특허 제2011-0056800호(역학 시뮬레이션 시스템 및 방법)에서는, 전국의 각 지역 내에서 사람 간의 접촉률과 각 지역의 인구 통계학적 특성 및 지역간 네트워크를 고려한 교통량을 적용하여 질병 전파의 방향 및 확산 정보와 환자 수를 계산하는 패치 모델을 통해 인플루엔자의 유행 시 전염 경로 및 지역 등을 예측하는 역학 시뮬레이션 시스템 및 방법을 개시하고 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명의 일 실시예는 SNS 데이터를 이용하여 질병의 확산 상황을 예측 분석하여 그 예측 정보를 제공하는 질병 확산 상황 예측 장치 및 방법을 제공하고자 한다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 측면에 따른 질병 확산 상황 예측 장치는, 질병 통계 서버가 제공하는 기설정된 질병에 대한 의사 환자 분율 데이터를 수집하여 저장하는 질병 통계 자료 수집부; 소셜 네트워크 서비스(SNS, Social Network Service) 서버가 제공하는 원본 SNS 데이터를 수집하여 저장하는 SNS 자료 수집부; 상기 저장된 SNS 데이터로부터 상기 질병과 관련된 기설정된 기준 단어 중 적어도 하나가 포함된 유효 SNS 데이터를 추출하는 유효 데이터 추출부; 상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 상기 의사 환자 분율 데이터와의 상관도에 기초하여 상기 유효 SNS 데이터의 단어들로부터 마커를 결정하는 마커 선별부; 기설정된 기간의 상기 저장된 SNS 데이터로부터 상기 마커의 사용 횟수 및 상기 기설정된 기간의 상기 저장된 SNS 데이터의 총 개수를 포함하는 마커 매트릭스를 생성하는 마커 매트릭스 생성부; 상기 마커 매트릭스를 이용하여 상기 저장된 SNS 데이터 내에서의 상기 마커의 사용 비율을 산출하고, 상기 마커의 사용 비율 및 상기 의사 환자 분율 데이터에 기초하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출하는 질병 확산 예측부; 및 상기 산출된 질병 확산 예측 값에 기초한 질병 확산 정보를 생성하여 제공하는 질병 확산 정보 제공부를 포함한다.

[0009] 또한, 본 발명의 다른 측면에 따른 질병 확산 상황 예측 장치의 질병 확산 상황 예측 방법은, (a) 소셜 네트워크 서비스(SNS, Social Network Service) 서버가 제공하는 원본 SNS 데이터 및 질병 통계 서버가 제공하는 기설

정된 질병에 대한 의사 환자 분율 데이터를 각각 수집하여 저장하는 단계; (b) 상기 저장된 SNS 데이터로부터 질병과 관련된 기설정된 기준 단어 중 적어도 하나를 포함하는 유효 SNS 데이터를 추출하는 단계; (c) 상기 유효 SNS 데이터에 포함된 단어 별 사용 빈도와 상기 의사 환자 분율 데이터와의 상관도에 기초하여 상기 유효 SNS 데이터의 단어들로부터 마커를 결정하는 단계; (d) 기설정된 기간 단위 내의 상기 저장된 SNS 데이터 내에서의 상기 마커의 사용 비율을 산출하는 단계; (e) 상기 마커의 사용 비율 값 및 상기 의사 환자 분율 데이터에 대해 선형 회귀 분석 처리하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출하는 단계; 및 (f) 상기 산출된 질병 확산 예측 값에 기초한 질병 확산 정보를 생성하여 사용자 단말에 제공하는 단계를 포함한다.

발명의 효과

- [0010] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, SNS 데이터를 이용하여 전염병의 발생 및 확산에 대한 실시간 모니터링 및 예측 분석이 가능하여 질병에 대한 조기 예방 조치를 취할 수 있다는 효과가 있다.
- [0011] 그리고, 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 질병 확산 상황 예측을 위한 자료로서의 SNS 데이터를 이용함으로써, 간편하게 다양한 질병에 대한 각각의 질병 확산 상황 예측 분석을 수행하여 그 예측 정보를 제공할 수 있는 효과가 있다.
- [0012] 또한, 본 발명의 과제 해결 수단 중 어느 하나에 의하면, SNS 데이터 중 특정 언어(특히, 한글) 별로 적합한 데이터 가공 처리를 하여 실제 질병 통계 데이터와 상관도가 높은 SNS 데이터를 질병 확산 상황의 예측 분석의 조건으로 사용함으로써 예측의 정확도를 높일 수 있다는 효과가 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 일 실시예에 따른 질병 확산 상황 예측 장치의 구성을 나타내는 블록도이다.
- 도 2는 본 발명의 일 실시예에서 LASSO(Least-Absolute Shrinkage And Selection Operator) 분석의 결과에 따른 마커를 선별하는 방식을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일 실시예에서 SNS 데이터의 마커와 ILI 분율 데이터 간의 상관 관계를 나타내는 일례이다.
- 도 4는 본 발명의 일 실시예에서 SNS 데이터와 ILI 분율 데이터에 기초하여 질병 확산 예측을 수행한 결과를 나타내는 일례이다.
- 도 5는 본 발명의 일 실시예에서 질병 확산 예측 분석에 사용된 데이터의 기간 별 확산 예측 결과를 나타내는 일례이다.
- 도 6은 본 발명의 일 실시예에서 질병 확산 예측 분석에 사용된 데이터의 기간 별 예측 분석 결과 값들의 유효성을 나타내는 일례이다.
- 도 7은 본 발명의 일 실시예에서 제공하는 질병 확산 정보의 일례를 나타내는 그래프이다.
- 도 8은 본 발명의 일 실시예에 따른 질병 확산 예측 분석을 적용함에 따른 다른 종류의 질병에 대한 SNS 데이터를 이용한 질병 통계 데이터와의 상관 관계 및 질병 확산 예측 분석의 결과를 나타내는 일례이다.
- 도 9는 본 발명의 일 실시예에서 스마트 폰을 통해 제공되는 질병 확산 정보 화면의 일례를 나타내는 도면이다.
- 도 10은 본 발명의 일 실시예에 따른 질병 확산 상황 예측 방법을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0015] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0016] 도 1은 본 발명의 일 실시예에 따른 질병 확산 상황 예측 장치의 구성을 나타내는 블록도이다.
- [0017] 도 1에 도시한 바와 같이, 본 발명의 일 실시예에 따른 질병 확산 상황 예측 장치(100)는 SNS 자료 수집부

(110), 질병 통계 자료 수집부(120), 데이터베이스(130), 유효 데이터 추출부(140), 마커 선별부(150), 마커 매트릭스 생성부(160), 질병 확산 예측부(170) 및 질병 확산 정보 제공부(180)를 포함하여 구성된다.

- [0018] SNS 자료 수집부(110)는 소셜 네트워크 서비스 서버(10)가 제공하는 SNS 데이터(이하, ‘원본 SNS 데이터’라고 함)를 수집하여 데이터베이스(130)에 저장(이하, ‘저장 SNS 데이터’라고 함)하되, 사전에 저장되어 있던 일별 저장 SNS 데이터에 업데이트하여 저장한다. 참고로, SNS 서버(10)로부터 제공되는 SNS 데이터는 가공되지 않은 상태의 데이터 즉, 미가공 데이터(raw data)이다. 그리고, SNS 데이터는 복수의 계정 별 사용자가 SNS 서버(10) 상에 업로드하는 텍스트(text) 데이터를 의미한다.
- [0019] 본 발명의 실시예에서는 다수의 SNS 별로 제공하는 작성 툴(tool)을 이용하여 사용자가 일회에 작성(즉, 업로드)하는 텍스트 데이터의 기본 단위(예를 들어, ‘멘션’, ‘트윗’ 등의 용어로 지칭되는 단위)를 개별 SNS 데이터라 지칭하도록 하며, 이러한 개별 SNS 데이터는 음절, 단어, 구, 절, 및 문장 중 적어도 하나의 단위가 포함된다.
- [0020] 그리고, 본 발명의 일 실시예에서는 SNS 서버(10)가 SNS 데이터를 불특정 다수에게 랜덤 제공하는 오픈형 SNS 서버인 것을 예로서 설명하나, SNS 자료 수집부(110)가 인터넷에서 상용되고 있는 SNS 데이터 수집 장치 또는 애플리케이션을 통해 실시간으로 발행되는 SNS 데이터를 수집하는 것도 가능하다.
- [0021] 한편, SNS 자료 수집부(110)는 질병 확산 예측 분석을 위한 SNS 데이터를 효율적으로 검출하여 제공하기 위하여, 원본 SNS 데이터를 관계형 데이터베이스 형태로 저장할 수 있다. 예를 들어, 원본 SNS 데이터를 사용자 계정 별로 분류하여 테이블 형태로 저장할 수 있다.
- [0022] 또한, SNS 자료 수집부(110)는 원본 SNS 데이터를 기설정된 추출 조건에 따라 선별하여 데이터베이스(130)에 저장함으로써 저장 SNS 데이터를 구성할 수 있다.
- [0023] 구체적으로, SNS 자료 수집부(110)는 원본 SNS 데이터 중 기설정된 특정 언어의 SNS 데이터를 검출하거나, 하이퍼링크(hyper link)를 포함하는 SNS 데이터를 제거하거나, 기설정된 스팸(SPAM) 단어를 포함하는 SNS 데이터를 제거하거나, 각 SNS 계정 별로 해당 계정의 사용자 이외의 사용자에게 의해 작성된 SNS 데이터(예를 들어, 리트윗(RT, Retweet) 등)을 제거하는 것 중 적어도 하나의 처리를 통해 저장할 SNS 데이터를 선별할 수 있다. 참고로, 이상의 SNS 데이터 선별 방법은 원본 SNS 데이터에 대해 각 처리가 공통 또는 별개로 적용되거나, 동시 또는 단계 별로 적용될 수 있다.
- [0024] 본 발명의 일 실시예에서는, SNS 자료 수집부(110)가 한글로 작성된 SNS 데이터를 선별하여, 선별된 한글 SNS 데이터에 대해 나머지 추출 조건에 따른 선별 및 이를 이용한 질병 확산 예측 분석을 처리하는 것을 설명하도록 한다. 이는, 다른 종류의 언어 중에서도 한글이 가지고 있는 특성 중, 구성 방식, 다양한 어미 변화, 다양한 동음이의어, 다양한 단어 활용형, 신조어, 및 띄어 쓰기 등을 적용한 분석을 효율적으로 처리하는 방식을 설명하기 위한 것으로서, 다른 종류의 언어에도 각각 적합한 데이터 가공 전처리를 적용할 수 있다.
- [0025] 질병 통계 자료 수집부(120)는 질병 통계 서버(20)가 제공하는 기설정된 질병에 대한 질병 통계 데이터(본 발명의 일 실시예에서는, 의사 환자 분율 데이터)를 수집하여 데이터베이스(130)에 저장하되, 사전에 통계 보고일 별로 저장되어 있던 질병 통계 데이터에 업데이트하여 저장한다.
- [0026] 참고로, 본 발명의 실시예에 적용되는 질병 통계 서버(20)는 질병 관리 본부 등 질병에 대한 각종 통계 자료를 제공하는 기관 자체의 서버이거나 상기 기관과 연계된 서버로서, 질병 통계 데이터는 각 질병의 종류 별로 통계된 의사 환자 분율 데이터일 수 있다. 이때, 본 발명의 실시예에서 의사 환자 분율 데이터는 업데이트 당일로부터 적어도 하루 이전의 의사 환자 분율에 대한 정보를 포함하는 것이다.
- [0027] 이하에서는, 본 발명의 일 실시예에 따른 질병 확산 상황 예측 장치(100)가 전염병 중 ‘인플루엔자(독감)’를 확산 예측용 질병으로 설정하여, ‘인플루엔자’에 관한 확산 예측 분석을 수행하는 것을 설명하도록 한다. 이때, 의사 환자 분율 데이터로서 인플루엔자 의사 환자(ILI, influenza-like illness) 분율 데이터를 수집한다.
- [0028] 데이터베이스(130)는 SNS 자료 수집부(110)로부터 수집된 SNS 데이터 및 질병 통계 자료 수집부(120)로부터 수집된 ILI 분율 데이터를 포함하여, 질병 확산 예측 분석 처리에 필요한 기준 데이터들을 저장한다. 그리고, 데이터베이스(130)는 질병 확산 상황 예측 장치(100)의 각 부가 요청하는 데이터들을 검출하여 제공한다.
- [0029] 또한, 도 1에는 각 데이터를 저장하고 있는 데이터베이스(130)를 하나의 개념으로 나타내었으나, 데이터베이스(130)의 기능 및 저장되어 있는 데이터의 종류에 따라 다수의 분리된 데이터베이스(130) 형태로 포함되는 것도

가능하다.

- [0030] 예를 들어, 본 발명의 일 실시예에 따른 데이터베이스(130)는 SNS 데이터, 질병 통계 데이터, 유효 SNS 데이터 추출을 위한 기준인 기준 단어 데이터, 스팸형 SNS 데이터를 제거하기 위한 기준인 스팸 단어 데이터, SNS 데이터 내 단어들의 각 기본형을 검출하기 위한 단어 활용형 데이터, 예측 분석에 유효한 SNS 데이터를 선별하기 위해 적합한 기준 단어의 사용을 검출하기 위한 동음이의어 데이터 및 안부성 문장/구/단어 데이터를 각각 저장하는 복수의 데이터베이스들(미도시)이 분리되어 구비될 수 있다. 참고로, 기준 단어 데이터, 스팸 단어 데이터, 단어 활용형 데이터, 동음이의어 데이터, 및 안부성 문장/구/단어 데이터의 각 데이터베이스들은 ‘언어 온톨로지’ 형태일 수 있으며, ‘언어 온톨로지’는 단어와 관계들로 구성된 일종의 사전일 수 있다.
- [0031] 유효 데이터 추출부(140)는 원본 SNS 데이터로부터 질병과 관련된 기설정된 기준 단어 중 적어도 하나가 포함된 유효 SNS 데이터를 추출한다. 이때, 유효 데이터 추출부(140)는 사용자가 요청한 예측 목표일의 원본 SNS 데이터로부터 유효 SNS 데이터를 추출하거나, 주/일/시 별로 원본 SNS 데이터 중 유효 SNS 데이터를 자동 추출할 수 있다.
- [0032] 구체적으로, 유효 데이터 추출부(140)는 원본 SNS 데이터 중 SNS 자료 수집부(110)가 선별한 SNS 데이터(즉, 상기 추출 조건을 만족하는 한글 SNS 데이터)로부터, 질병(즉, 인플루엔자)과 관련하여 사전에 설정된 기준 단어들을 포함하는 SNS 데이터를 유효 SNS 데이터로 추출한다. 일례로서, ‘인플루엔자’, ‘독감’, ‘감기’, ‘기침’, 및 ‘플루’ 등을 기준 단어로 사용할 수 있다.
- [0033] 또한, 유효 데이터 추출부(140)는 기준 단어를 포함하는 SNS 데이터 중 기준 단어와 관련하여 기설정된 안부성 문장 또는 구 또는 단어를 포함하는 SNS 데이터의 제거, 및 기준 단어와의 동음이의어를 포함하는 SNS 데이터의 제거 중 적어도 하나의 제거를 처리하여 유효 SNS 데이터를 추출할 수 있다.
- [0034] 구체적으로, 유효 데이터 추출부(140)는 사전에 저장되어 있는 안부성 문장/구/단어 데이터에 기초하여 개별 SNS 데이터를 분석하여, 예를 들어 ‘일교차가 심하네요. 감기 조심하세요.’와 같이 실제 질병의 확산 상황의 예측과는 상관이 없는 안부성 문장을 포함하는 SNS 데이터를 제거할 수 있다.
- [0035] 그리고, 유효 데이터 추출부(140)는 사전에 저장되어 있는 동음이의어 데이터에 기초하여 개별 SNS 데이터를 분석하여, 예를 들어 ‘눈은 감기고 시험은 일주일도 안 남았고……’와 같은 기준 단어(즉, ‘감기’)의 동음이의어를 포함하는 SNS 데이터를 제거할 수 있다.
- [0036] 예를 들어, 유효 데이터 추출부(140)는 기저장된 안부성 문장/구/단어와 일치하는 데이터가 개별 SNS 데이터 내에 포함되어 있거나, 동음이의어 별로 상용적으로 기준 단어의 동음이의어와 함께 사용되는 적어도 하나의 단어가 개별 SNS 데이터 내에 포함되어 있을 경우 각 SNS 데이터를 제거할 수 있다.
- [0037] 마커 선별부(150)는 유효 SNS 데이터에 포함된 단어 별 사용 빈도를 산출하고, 산출된 단어 별 사용 빈도와 의사 환자 분율 데이터 간의 상관도를 분석하여, 분석의 결과에 기초하여 단어 중 마커를 결정한다.
- [0038] 이때, 마커 선별부(150)는 유효 SNS 데이터들에 포함된 단어 별 사용 빈도와 의사 환자 분율 데이터에 대해 LASSO 분석을 수행한다. 그리고, 마커 선별부(150)는 LASSO 분석을 통해 산출된 단어 별 상관 계수가 기설정된 임계치 이상인 단어이거나 상관 계수의 값이 높은 순서에 따라 기설정된 순위 범위까지의 단어를 마커로 설정할 수 있다.
- [0039] 구체적으로, 마커 선별부(150)는 유효 SNS 데이터에 포함된 개별 SNS 데이터들을 각각 빈칸에 기준하여 개별 단어로 구분하고, 유효 SNS 데이터 내에서 개별 단어 별 사용 횟수를 산출한다. 그리고, 마커 선별부(150)는 사전에 저장된 단어활용형 데이터에 기초하여 상기 구분된 단어 별로 각각 매칭되는 기본형을 검출하여, 동일한 기본형을 갖는 단어들의 유효 SNS 데이터들 내 사용 횟수를 총합하여 단어 별 사용 빈도를 산출한다.
- [0040] 그리고, 마커 선별부(150)는 단어 사용 빈도가 높은 순서에 따라 기설정된 순위 범위까지의 단어를 판단하고, 상기 순위 범위까지의 단어를 마커 후보 단어로 선별한다.
- [0041] 예를 들어, 하기 표 1은 마커 선별부(150)를 통해 ‘인플루엔자’에 대한 유효 SNS 데이터로부터 선별된 마커 후보 단어들(즉, 유효 SNS 데이터 내 사용 횟수가 높은 단어들)을 나타내었다.
- [0042] [표 1]

가그린	독감	아이	있다	춡다
감기	두통	아프다	작년	치레
감기약	들다	악화	저도	치료
감기	따위	얕다	전염	치료제
감염	때문	악국	전염병	코감기
걱정	떠나다	악하다	접충	물룩
건강하다	떨다	어렵다	정말	콧물
걸리다	떨어지다	얼른	제대로	타미플루
결국	망할	엄마	조금	타이레놀
고생	대칠	없다	조류	테라플루
고열	목감기	예취	조심	폐렴
골골	목소리	예방	중합	플루
괜찮다	몸살	오늘	춡다	피곤
기운	무섭다	오다	주간	피로
기침	물려가다	오랫동안	주사	한달
까지	바이러스	오한	주의	한의원
깨끗하다	백신	웁다	주짜	해열제
꺼지다	벌써	완전	죽겠다	확정
끝나다	병원	요즘	증상	환자
나도	보내다	유행	증세	회복
나쁘다	불다	의료원	지겹다	효과적
냇다	빨리	이겨내다	지금	홀췌
녀석	세균	이미	지나가다	힘들다
높다	쉬다	이번	지독	
느낌	시작	이상하다	지랄	
달간	신종플루	인가	진짜	
달째	싫다	인듯	짜증	
담배	심하다	인플루엔자	차다	
당침	쌔쌔하다	일간	초기	
도움	아스피린	일째	최악	

[0043]

[0044]

한편, 마커 선별부(150)가 마커 후보 단어 별 사용 빈도와 의사 환자 분율 데이터를 LASSO 분석한 결과는 하기 도 2와 같이 나타낼 수 있다.

[0045]

도 2는 본 발명의 일 실시예에서 LASSO 분석의 결과에 따른 마커를 선별하는 방식을 설명하기 위한 도면이다.

[0046]

도 2에서는 마커 후보 단어 별로 의사 환자 분율 데이터와의 LASSO 분석을 통해 산출된 상관 계수(Weighting coefficient)를 나타내었다. 이처럼, 유효 SNS 데이터 내 단어들은 각각 의사 환자 분율 데이터와의 상관 관계도에 따라 '+' 또는 '-'의 상관 계수 값으로 표현되며, 마커 선별부(150)는 상관 계수가 기설정된 임계치 이상이거나 상관 계수 값이 높은 순서에 따라 기설정된 순위 범위까지의 단어를 마커로 결정한다.

[0047]

다시 도 1로 돌아가서, 마커 매트릭스 생성부(160)는 저장 SNS 데이터 내에서 기설정된 기간 단위(예를 들어, 일 단위 또는 주 단위 등)의 마커의 이용 횟수와 해당 기간 단위의 저장 SNS 데이터 총 개수를 추출하여 마커 매트릭스를 생성한다.

[0048]

구체적으로, 마커 매트릭스 생성부(160)는 질병 확산 예측부(170)의 입력 데이터로서 활용될 저장 SNS 데이터 내 마커의 기간 단위 총 사용 횟수 및 동일 기간의 저장 SNS 데이터의 총 개수를 포함하는 마커 매트릭스를 생성한다.

[0049]

질병 확산 예측부(170)는 마커 매트릭스 생성부(160)에서 생성된 마커 매트릭스를 이용하여 저장 SNS 데이터 내에서의 마커의 사용 비율을 산출하고, 마커의 사용 비율 및 의사 환자 분율 데이터에 기초하여 예측 목표일에 대응하는 질병 확산 예측 값을 산출한다.

[0050]

구체적으로, 질병 확산 예측부(170)는 마커 매트릭스에 포함된 저장 SNS 데이터 내에서의 마커의 총 사용 횟수 및 동일 기간의 저장 SNS 데이터의 총 개수에 기초하여, 마커의 총 사용 횟수를 저장 SNS 데이터의 총 개수로 나눈 값을 마커 사용 비율 값으로 산출한다. 그리고, 질병 확산 예측부(170)는 마커 사용 비율 값 및 의사 환자 분율 데이터에 대해 선형 회귀(linear regression) 분석을 수행하여 질병 확산 예측 값을 산출한다.

[0051]

한편, 도 3은 본 발명의 일 실시예에서 SNS 데이터의 마커와 ILI 분율 데이터 간의 상관 관계를 나타내는 일례이다.

[0052]

도 3에서는, 복수 개의 마커를 이용하여 선형 회귀 예측 분석을 수행한 결과와 동일 기간에서의 의사 환자 분율 데이터 간의 상관 관계를 나타내었다. 이처럼, SNS 데이터 내 질병 관련 단어의 사용 횟수에 기초하여 의사 환자 분율 데이터와의 선형 회귀 분석을 수행함으로써, 실제 질병 확산의 패턴과 거의 동일한 SNS 데이터에 기반

한 질병 확산 예측 패턴을 산출할 수 있다. 참고로, 사용되는 마커의 개수가 많을수록 산출된 질병 확산 예측 값과 의사 환자 분율 데이터 간의 상관 관계는 더욱 긴밀해진다.

- [0053] 한편, 인플루엔자의 발생 패턴은 유행 시즌 별로 유사하며, SNS 데이터 사용자 들의 글쓰기 습관 또한 일정하게 유지된다고 가정할 수 있다.
- [0054] 이에 따라, 질병 확산 예측부(170)는 질병의 이전 유행 시즌(예를 들어, 년 또는 월 단위)의 데이터가 확보되지 않은 경우, 현재 유행 시즌의 의사 환자 분율 데이터 및 SNS 데이터와 현재 유행 시즌의 의사 환자 분율 데이터 및 SNS 데이터로부터 추출된 마커 데이터 중 적어도 하나의 데이터를 이전 유행 시즌 데이터로서 중복(즉, 대체) 사용할 수 있다. 이는, 하기 도 4 내지 도 6에서 나타내는 바와 같이, 질병 확산 상황 예측의 정밀한 산출을 위해서는 이전 유행 시즌의 데이터들이 중요하기 때문이다.
- [0055] 예를 들어, 도 4는 본 발명의 일 실시예에서 SNS 데이터와 ILI 분율 데이터에 기초하여 질병 확산 예측 분석을 수행한 결과를 나타내는 일례이다.
- [0056] 도 4의 (a)는 현재 유행 시즌의 데이터만을 이용하여 선형 회귀 분석을 수행한 결과 값을 나타내었으며, (b)는 이전 유행 시즌의 데이터를 포함한 데이터를 이용하여 선형 회귀 분석을 수행한 결과를 나타내었다. 이때, 도 4는 각 기간의 날짜 별 질병 확산 예측 값 및 인플루엔자 의사 환자(ILI) 분율 데이터의 변화를 비교한 그래프를 나타내었다. 도 4에 나타난 바와 같이, 이전 유행 시즌의 데이터를 포함한 데이터를 이용한 선형 회귀 분석의 결과 값이 현재 유행 시즌의 데이터만을 이용한 선형 회귀 분석의 결과 값 보다 인플루엔자 의사 환자 분율 데이터와 상관 관계가 긴밀한 것을 알 수 있다.
- [0057] 또한, 도 5는 본 발명의 일 실시예에서 질병 확산 예측 분석에 사용된 데이터의 기간 별 확산 예측 결과를 나타내는 일례이다.
- [0058] 도 5에서는 목표 예측일을 n 이라고 가정할 경우, $(n-1)$ 전부터 $(n-21)$ 전까지의 데이터를 이용하여 질병 확산 예측 값을 산출한 결과를 나타내었다. 이때, 도 5의 (a)는 $(n-1)$ 일, (b)는 $(n-7)$ 일, (c)는 $(n-14)$ 일, 및 (d)는 $(n-21)$ 일의 질병 예측 결과 값 및 인플루엔자 의사 환자(ILI) 분율 데이터의 비교 결과를 나타낸 그래프이다.
- [0059] 도 6은 본 발명의 일 실시예에서 질병 확산 예측 분석에 사용된 데이터의 기간 별 예측 분석 결과 값들의 유효성을 나타내는 일례이다.
- [0060] 도 6에서는 도 5에서와 같이 $(n-1)$ 일부터 $(n-21)$ 일까지의 데이터를 이용하여 선형 회귀 분석을 하여 산출한 질병 확산 예측 값에 대한 회귀식 설명력 R^2 값을 나타내었다. 이때, 도 6에서와 같이, $(n-21)$ 까지의 데이터를 이용하여 산출된 질병 확산 예측 값의 R^2 값들이 각각 0.8 이상의 값으로서 유효한(즉, 유의미한) 예측 값인 것을 알 수 있다.
- [0061] 다시 도 1로 돌아가서, 질병 확산 정보 제공부(180)는 질병 확산 예측부(170)가 산출한 질병 확산 예측 값에 기초하여, SNS 데이터에 기반한 질병 확산 정보를 생성하여 제공한다. 예를 들어, 질병 확산 정보 제공부(180)는 산출된 질병 확산 예측 값의 날짜 별 변화를 그래프 형태로 출력할 수 있다.
- [0062] 이때, 질병 확산 정보 제공부(180)는 사용자 단말(미도시) 또는 자체적으로 포함한 출력부를 통해 질병 확산 정보가 출력될 수 있도록 포맷화된 질병 확산 정보를 제공할 수 있다. 참고로, 사용자 단말(미도시)은 개인 단말 이거나, 국가 질병 관리 기관 또는 질병 백신 개발 사업체 등 다양한 사용자의 단말을 의미한다.
- [0063] 도 7은 본 발명의 일 실시예에서 제공하는 질병 확산 정보의 일례를 나타내는 그래프이다.
- [0064] 이때, 도 7에서는 질병 확산 상황 예측의 예측 목표일이 인플루엔자 의사 환자(ILI) 분율 데이터가 발표되지 않은 날인 경우, 질병 확산 상황의 예측 값을 나타내었다. 즉, 도 7에서와 같이, 질병 관리 본부에서 인플루엔자 의사 환자(ILI) 분율 데이터 보고가 발표되지 않은 일자에도 질병 확산 상황 예측을 수행할 수 있다. 참고로, 동일 날짜에 의사 환자 분율 데이터를 수집하였다 하더라도, 사실상 의사 환자 분율 데이터는 업데이트 당일로 부터 대략 2주 정도 이전의 질병 확산 상황을 보고하는 것이므로, 본 발명의 실시예에서와 같이 SNS 데이터에 기초하여 질병 확산 예측 값을 산출함으로써 실시간으로 질병 확산 상황을 예측할 수 있다.
- [0065] 한편, 이상에서는 본 발명의 일 실시예에 따른 질병 확산 상황 예측 장치(100)가 ‘인플루엔자’에 대한 질병 통계 데이터 및 SNS 데이터를 이용하여 질병 확산 상황을 예측하는 것을 설명하였으나, 다른 질병에 대해서도

동일한 방식으로 질병 확산 상황을 예측할 수 있다.

- [0066] 예를 들어, 도 8은 본 발명의 일 실시예에 따른 질병 확산 예측 분석을 적용함에 따른 다른 종류의 질병에 대한 SNS 데이터를 이용한 질병 통계 데이터와의 상관 관계 및 질병 확산 예측 분석의 결과를 나타내는 일례이다.
- [0067] 도 8에서는, ‘결막염’ 및 ‘눈병’에 대한 질병 확산 상황 예측을 수행한 결과를 나타내었다.
- [0068] 도 8의 (a)에서는 결막염 및 눈병 별로 하나의 마커를 사용한 마커 사용 비율 값과 유행성 결막염 평균 환자 데이터(즉, 질병 통계 데이터)를 선형 회귀 분석한 결과를 나타내었으며, (b)에서는 복수개의 마커를 사용한 마커 사용 비율 값과 유행성 결막염 평균 환자 데이터를 선형 회귀 분석한 결과를 나타내었다.
- [0069] 또한, 질병 확산 정보 제공부(180)가 제공하는 질병 확산 정보는 질병 확산 예측 값 및 의사 환자 분율 데이터의 변화를 비교한 그래프 형태의 정보를 포함하되, 각 사용자 단말에서는 목적에 따라 상기 질병 확산 정보를 이용할 수 있다.
- [0070] 일례로, 질병 통제 기관의 경우, 국가적으로 감염성 질병 위기 경보 단계를 설정하여 두고, 질병 확산 정보 제공부(180)로부터 수신된 질병 확산 예측 값에 기초하여 해당하는 위기 경보 단계를 발령할 수 있으며, 질병에 대한 국가 정책 결정의 근거로 사용하여 각 위기 경보 단계에 해당하는 대응 조치 정보를 출력할 수 있다.
- [0071] 또 다른 예로서, 개인 사용자의 스마트 폰의 경우, 사전에 설치된 애플리케이션을 통해 질병 상황 예측을 요청할 수 있으며, 이에 따라 질병 확산 상황 예측 장치(100)로부터 예측 목표일에 대한 질병 확산 예측 값을 포함하는 질병 확산 정보를 수신하여 출력할 수 있다.
- [0072] 도 9는 본 발명의 일 실시예에서 스마트 폰을 통해 제공되는 질병 확산 정보 화면의 일례를 나타내는 도면이다.
- [0073] 도 9에서는, 질병 확산 정보 제공부(180)가 사용자 단말(스마트 폰)로 예측 목표일의 질병 확산 예측 값을 포함하는 질병 확산 정보를 제공한 것을 나타내었다. 이때, 도 9에서와 같이, 사용자 단말(스마트 폰)의 화면에는, 기간 별 인플루엔자 의사 환자(ILI) 분율 데이터 및 인플루엔자 질병 확산 예측 데이터가 서로 비교될 수 있도록 그래프 형태로 출력될 수 있다. 또한, SNS 데이터에 포함된 GPS 정보를 이용한 지역별 질병 확산 상황이 지도 상에 매핑되어 출력될 수 있으며, 인플루엔자에 대한 자가 진단 및 예방 조치에 대한 정보가 출력될 수 있다.
- [0074] 이하, 도 10을 참고하여 본 발명의 일 실시예에 따른 질병 확산 상황 예측 방법에 대해서 상세히 설명하도록 한다.
- [0075] 도 10은 본 발명의 일 실시예에 따른 질병 확산 상황 예측 방법을 설명하기 위한 순서도이다.
- [0076] 먼저, 소셜 네트워크 서비스 서버가 제공하는 원본 SNS 데이터 및 질병 통계 서버가 제공하는 기설정된 질병에 대한 질병 통계 데이터를 수집하여 저장한다(S110).
- [0077] 이때, 원본 SNS 데이터(즉, 수집되어 저장되는 ‘저장 SNS 데이터’)는 매일 업데이트(예를 들어, 시간 별 또는 일 별)되며, 질병 통계 데이터는 질병 통계 서버로부터 월 단위, 주 단위 또는 일 단위로 사전에 설정된 질병 통계 데이터 보고일에 업데이트 될 수 있다.
- [0078] 또한, 원본 SNS 데이터 중 기설정된 특정 언어를 사용하거나, 하이퍼링크(hyper link)의 불포함, 기설정된 스팸(SPAM) 단어의 불포함, 및 SNS 계정 별로 계정 사용자 이외의 사용자에게 의해 작성된 데이터 불포함 중 적어도 하나의 조건을 만족하는 SNS 데이터를 수집하여 저장할 수 있다.
- [0079] 그런 다음, 수집한 원본 SNS 데이터로부터 질병과 관련된 기설정된 기준 단어를 적어도 하나 포함하는 유효 SNS 데이터를 추출한다(S120).
- [0080] 이때, 기준 단어를 포함하는 SNS 데이터 중 기준 단어와 관련하여 기설정된 안부성 문장 또는 구 또는 단어를 포함하는 SNS 데이터를 제거하거나, 기준 단어와의 동음이의어를 포함하는 SNS 데이터를 제거하여 유효 SNS 데이터를 추출할 수 있다.
- [0081] 다음으로, 유효 SNS 데이터 내의 단어 별 사용 횟수와 질병 통계 데이터에 기초하여, 유효 SNS 데이터 내의 단어 중에서 질병 확산 예측 분석을 위한 마커를 선별한다(S130).
- [0082] 구체적으로, 유효 SNS 데이터에 포함된 단어 별 사용 빈도를 산출하고, 단어 별 사용 빈도와 질병 통계 데이터에 대해 LASSO 분석을 하고, LASSO 분석 처리의 결과 산출된 유효 SNS 데이터에 포함된 단어 별 상관 계수에 기

초하여 마커를 결정한다.

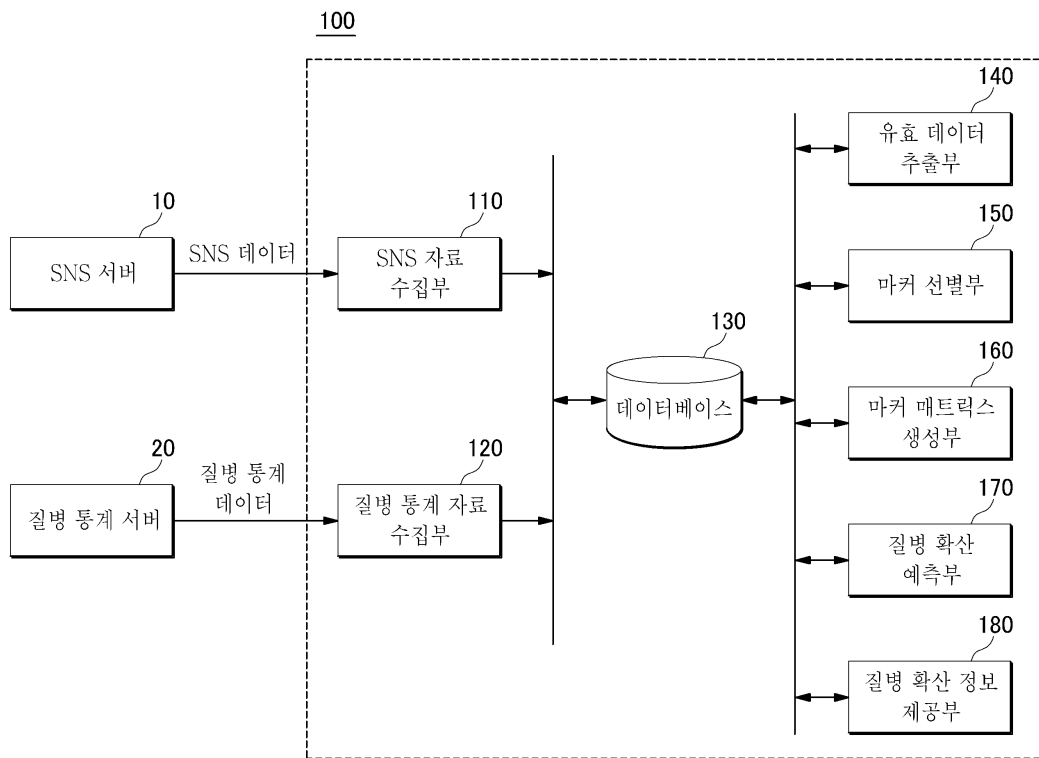
- [0083] 그런 후, 저장 SNS 데이터 내 마커 사용 비율 및 질병 통계 데이터에 대한 선형 회귀 분석 처리를 하여(S140), 예측 목표일의 질병 확산 예측 값을 산출한다(S150).
- [0084] 구체적으로, 저장 SNS 데이터에 대한 마커의 총 사용 횟수 및 저장 SNS 데이터의 총 개수에 기초하여 마커 사용 비율 값을 산출한 후, 마커 사용 비율 값 및 상기 의사 환자 분율 데이터에 대해 선형 회귀(linear regression) 분석 처리하여 상기 질병 확산 예측 값을 산출한다.
- [0085] 다음으로, 산출된 질병 확산 예측 값을 포함하는 질병 확산 정보를 제공한다(S160).
- [0086] 예를 들어, 날짜 별 질병 확산 예측 값 및 의사 환자 분율 데이터의 변화를 비교한 그래프를 질병 확산 정보로서 제공할 수 있다.
- [0087] 진술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0088] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0089] 또한, 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수도 있다.
- [0090] 본 발명의 범위는 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

- [0091] 10: SNS 서버
20: 질병 통계 서버
100: 질병 확산 상황 예측 장치
110: SNS 자료 수집부
120: 질병 통계 자료 수집부
130: 데이터베이스
140: 유효 데이터 추출부
150: 마커 선별부
160: 마커 매트릭스 생성부
170: 질병 확산 예측부
180: 질병 확산 정보 제공부

도면

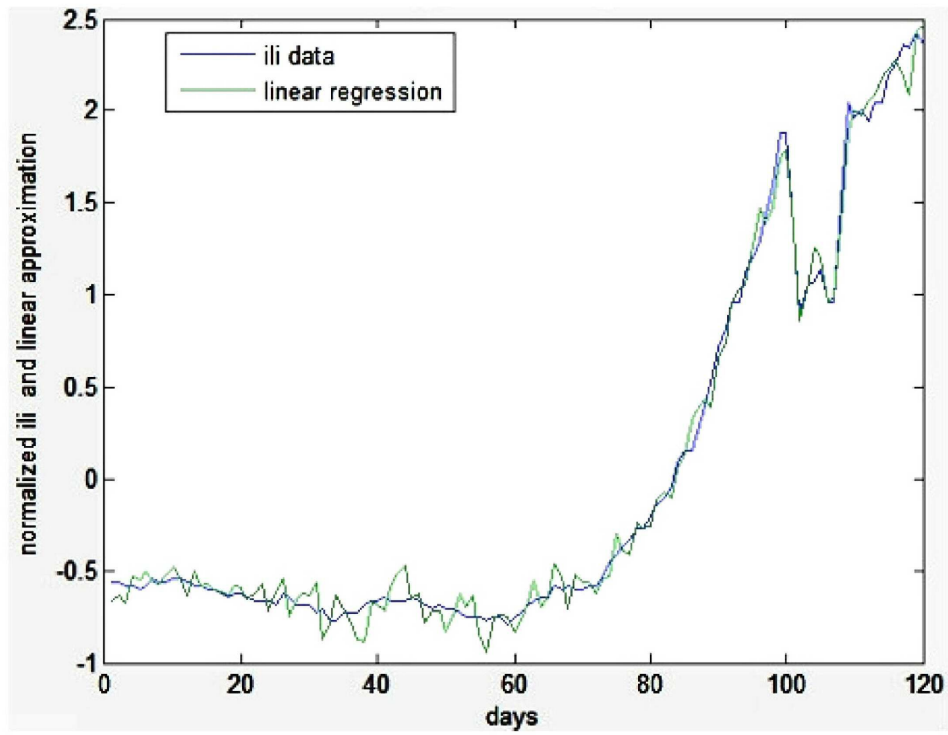
도면1



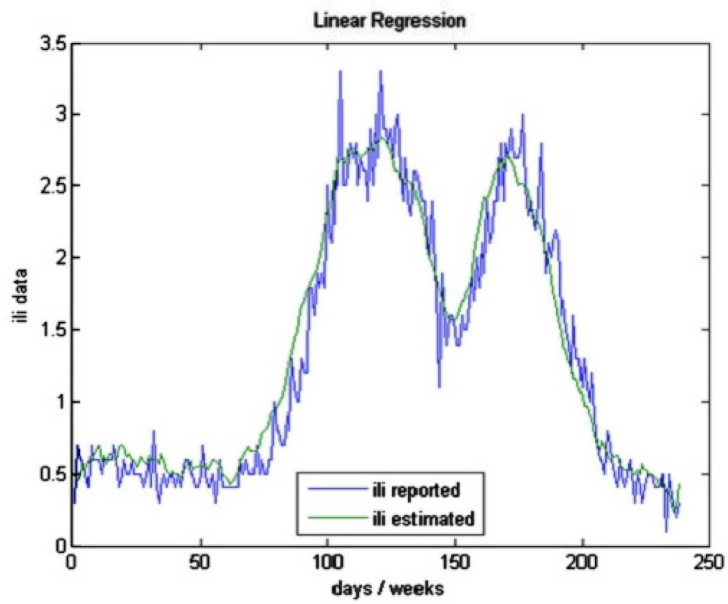
도면2

마커	Weighting Coeff.	마커	Weighting Coeff.	마커	Weighting Coeff.
가그린	1.527	벌써	-4.990	제대로	0.524
감기	4.165	병원	-4.070	조금	8.410
감기약	-0.689	보내다	4.423	조류	-0.469
감긴	-0.929	불다	-1.495	조심	-1.705
걱정	-1.045	빨리	8.169	종합	1.924
건강하다	-1.372	세균	-0.262	좋다	6.691
걸리다	-2.010	쉬다	1.508	주사	1.960
결국	-2.988	신종플루	1.362	주의	-2.651
고열	-0.784	싫다	-1.094	주짜	-1.488
골골	-2.891	심하다	0.920	죽겠다	10.513
기운	1.554	쌀쌀하다	-3.026	증상	1.795
기침	5.445	아스피린	0.352	증세	1.204
까지	4.598	아이	-2.322	지겹다	-0.925
깨끗하다	0.745	아프다	-2.111	지금	8.384
꺼지다	6.222	악화	0.228	지독	2.178
끝나다	-3.954	알다	2.766	지랄	0.699
나쁘다	3.344	약국	-0.259	진짜	-6.955
낮다	1.659	약하다	0.345	짜증	-2.842
녀석	-1.930	어렵다	2.655	차다	-2.315
높다	0.181	얼른	1.140	최악	-3.611
느낌	-1.301	엄마	-0.788	춡다	-1.578
달간	-0.462	예취	0.886	치레	0.400
달째	-0.815	예방	-0.926	치료	-1.221
담배	-1.365	오늘	-1.256	치료제	0.803
당첨	0.695	오다	2.830	코감기	-2.667
도움	-1.329	오랫동안	3.717	콜록	0.150
독감	3.010	오한	-0.528	콧물	-1.256
두통	1.713	웁다	0.563	타미플루	0.103
들다	-20.885	완전	-0.480	타이레놀	1.347
따위	-1.919	요즘	1.479	테라플루	-0.291
때문	-4.108	유행	0.527	폐렴	0.267
떠나다	0.790	의료원	1.604	플루	1.790
떨어지다	2.032	이겨내다	3.959	피곤	-3.779
망할	-0.975	이미	2.968	피로	-1.010
며칠	-1.393	이상하다	-1.716	한달	0.935
목감기	-0.410	인듯	-1.534	한의원	2.246
목소리	-2.931	일간	0.102	해열제	-0.077
몸살	2.286	일째	2.154	확정	-0.590
무섭다	-4.308	작년	0.804	환자	1.030
물러가다	-0.919	저도	-7.334	효과적	2.769
바이러스	-1.896	전염	-0.740	홀쩍	-0.845
백신	-1.931	접종	1.905	힘들다	5.250

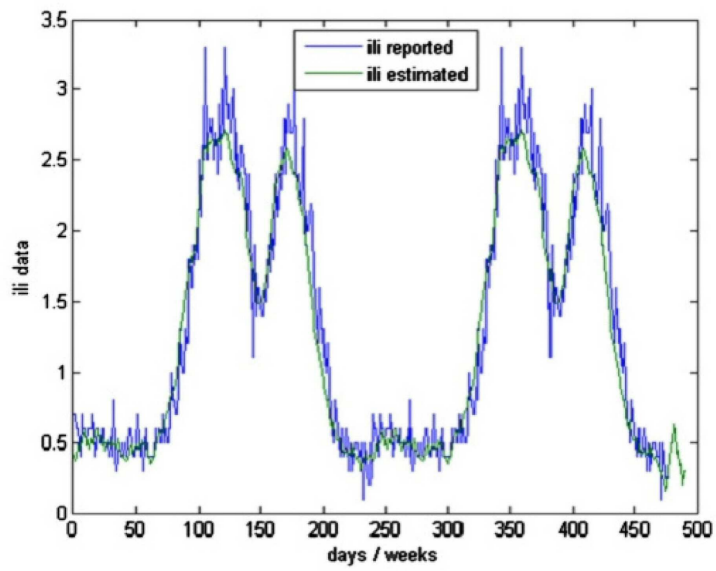
도면3



도면4

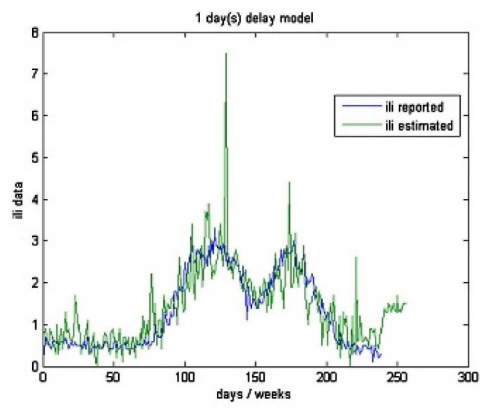


(a)

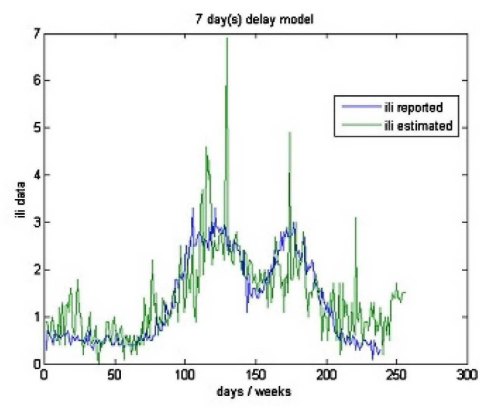


(b)

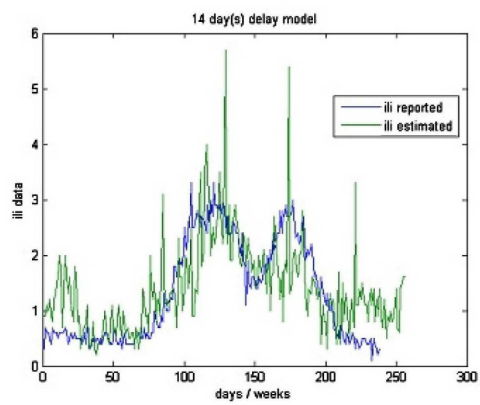
도면5



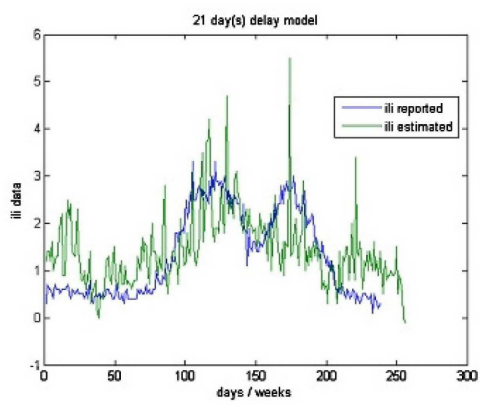
(a)



(b)



(c)

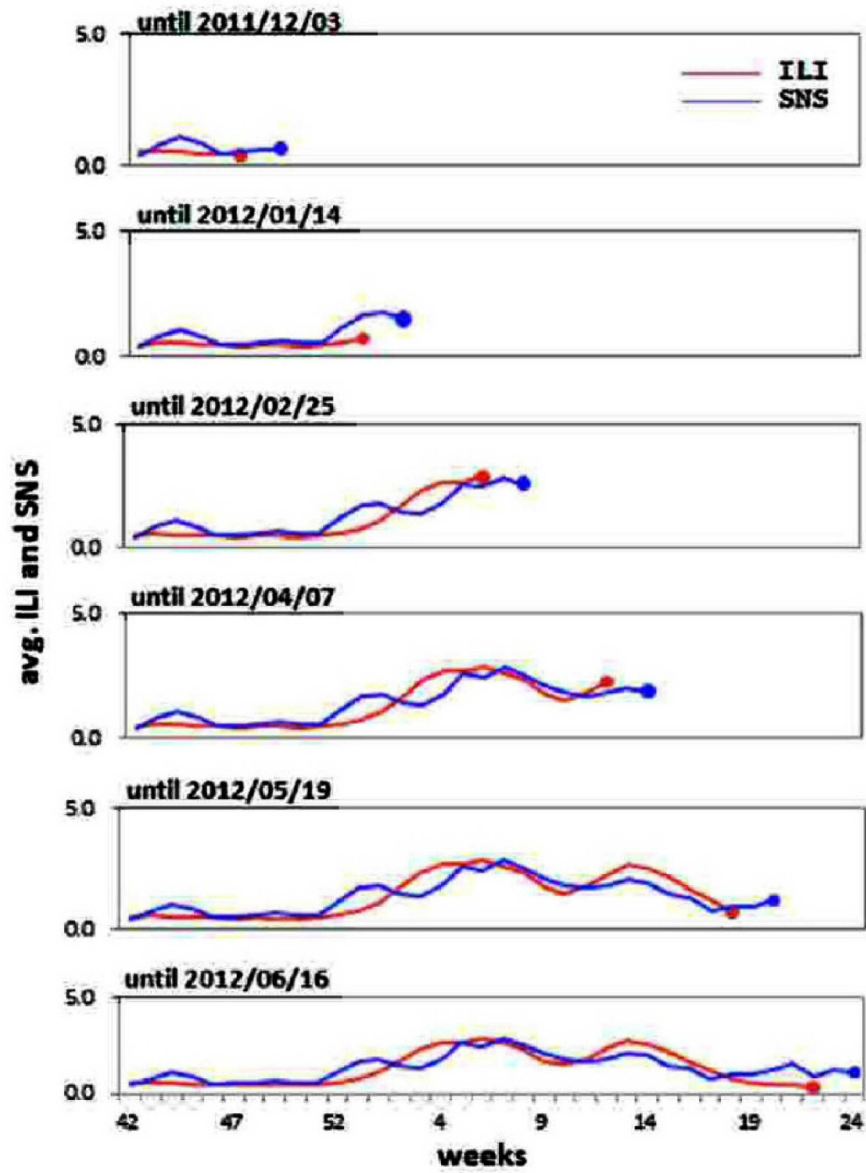


(d)

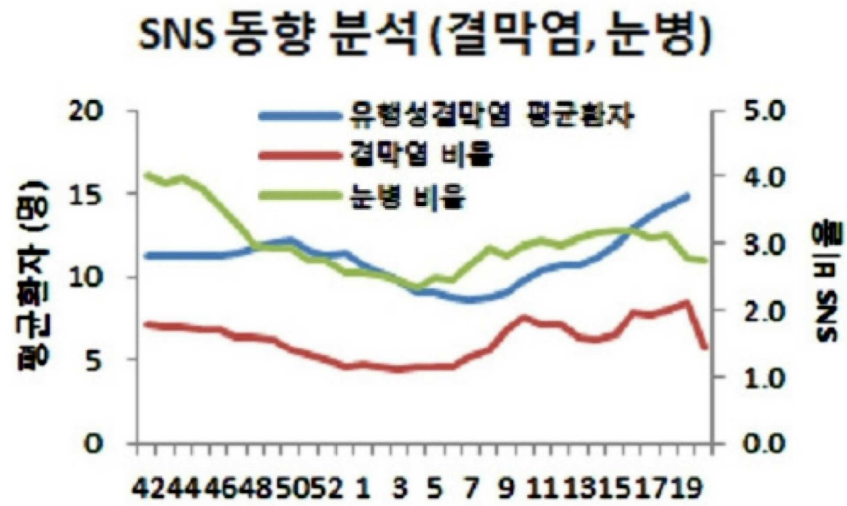
도면6

Delayed day(s)	$R^2_{ordinary}$
1	0.871
2	0.873
3	0.874
4	0.873
5	0.871
6	0.871
7	0.868
8	0.867
9	0.868
10	0.866
11	0.865
12	0.872
13	0.871
14	0.866
15	0.867
16	0.868
17	0.869
18	0.874
19	0.882
20	0.873
21	0.878

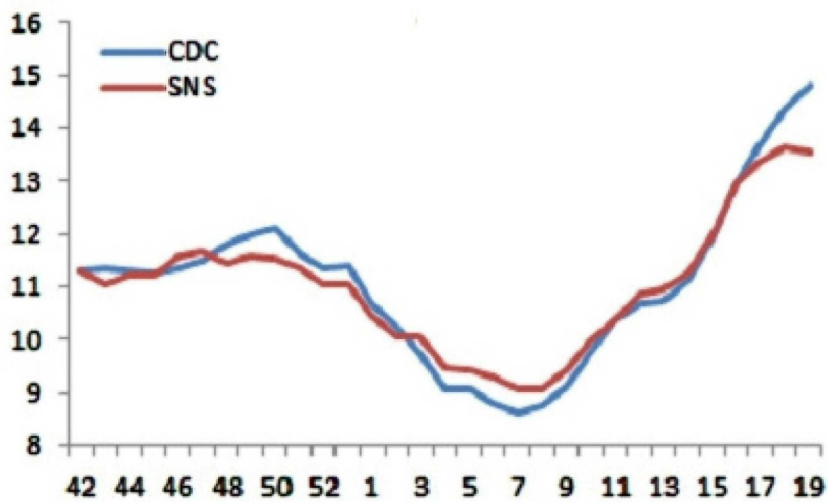
도면7



도면8



(a)



(b)

도면9



도면10

