(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0021840
(43) 공개일자 2022년02월22일

(51) 국제특허분류(Int. Cl.)
G06Q 50/26 (2012.01) *G06F 16/951* (2019.01)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06Q 50/26 (2013.01)
G06F 16/951 (2019.01)
(21) 출원번호 10-2020-0163665
(22) 출원일자 2020년11월30일
심사청구일자 2020년11월30일
(30) 우선권주장
1020200102512 2020년08월14일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김서연
서울특별시 서초구 방배로 270, 바동 1204호
이민정
경기도 고양시 덕양구 세솔로 149, 1603호 402호
(뒷면에 계속)
(74) 대리인
김홍석

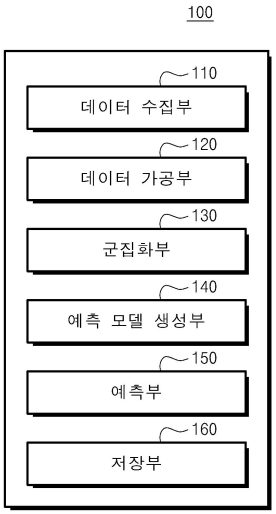
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 **치안 및 공공 데이터 기반 범죄 신고 수 예측 장치 및 방법**

(57) 요 약

치안 및 공공 데이터 기반 범죄 신고 수 예측모델 생성 장치 및 방법이 개시된다. 상기 예측모델 생성 장치는 크롤링 기법을 이용하여 데이터를 수집하는 데이터 수집부, 상기 데이터를 전처리하는 데이터 가공부, 및 전처리된 데이터를 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측모델을 생성하는 예측모델 생성부를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 20/00 (2021.08)
(72) 발명자
이지윤
경기도 의정부시 용민로 304, 507동 1503호
목충협
서울특별시 영등포구 당산로 214, 406동 501호
조용원
서울특별시 동대문구 왕산로 69-2, 902호
황하은
서울특별시 송파구 오금로32길 14, 104동 904호
김성범
서울특별시 송파구 올림픽로 99, 122동 1802호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103170
과제번호	2018-0-00440-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	ICT융합산업원천기술개발(R&D)
연구과제명	위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

크롤링 기법을 이용하여 데이터를 수집하는 데이터 수집부;

상기 데이터를 전처리하는 데이터 가공부; 및

전처리된 데이터를 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측모델을 생성하는 예측모델 생성부를 포함하는 예측모델 생성 장치.

청구항 2

제1항에 있어서,

상기 데이터는 미리 정해진 시간 구간 동안 발생한 범죄 신고 데이터를 포함하고,

상기 예측모델 생성부는 상기 범죄 신고 데이터를 이용하여 인공 신경망(Artificial Neural Network)을 학습시켜 상기 예측 모델을 생성하는,

예측모델 생성 장치.

청구항 3

제2항에 있어서,

범죄 신고 수의 패턴 유사도에 따라 관할관서들을 군집화하는 군집화부를 더 포함하고,

상기 예측모델 생성부는 군집들 각각에 대한 예측모델을 생성하는,

예측모델 생성 장치.

청구항 4

제3항에 있어서,

상기 군집화부는 DTW(Dynamic Time Warping)를 거리척도로 활용한 K-means 알고리즘을 이용하여 군집화하는,

예측모델 생성 장치.

청구항 5

적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 예측모델 생성 방법에 있어서,

크롤링 기법을 이용하여 데이터를 수집하는 단계;

상기 데이터를 전처리하는 단계; 및

전처리된 데이터를 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측모델을 생성하는 단계를 포함하는 예측모델 생성 방법.

청구항 6

제5항에 있어서,

상기 데이터는 미리 정해진 시간 구간 동안 발생한 범죄 신고 데이터를 포함하고,

상기 예측모델을 생성하는 단계는, 상기 범지 신고 데이터를 이용하여 RNN(Recurrent Neural Network)을 학습시켜 상기 예측모델을 생성하는,

예측모델 생성 방법.

청구항 7

제6항에 있어서,

상기 전처리하는 단계 이후에,

범죄 신고 수의 패턴 유사도에 따라 관할관서들을 군집화하는 단계를 더 포함하고,

상기 예측모델을 생성하는 단계는, 군집들 각각에 대한 예측모델을 생성하는,

예측모델 생성 방법.

청구항 8

제2 데이터를 수집하는 제2 데이터 수집부;

상기 제2 데이터를 전처리하는 제2 데이터 가공부; 및

제5항 내지 제7항 중 어느 하나의 항에 기재된 예측모델 생성 방법에 의해 생성되고 전처리된 제2 데이터를 입력으로 하는 예측모델을 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측부를 포함하는 예측 장치.

청구항 9

제8항에 있어서,

상기 제2 데이터는 미리 정해진 시간 구간 동안 발생한 범죄 신고 데이터를 포함하는,

예측 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 치안 및 공공 데이터 기반 범죄 신고 수 예측 방법에 관한 것으로, 특히 범죄 신고 수, 기상 데이터 등을 이용하여 예측모델을 생성함으로써 예측 시점에서의 범죄 신고 수를 미리 예측할 수 있는 범죄 신고 수 예측 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 인공지능 등의 정보통신의 발달로 치안 서비스의 지능화가 진행되고 있으며, 미국 등 주요 선진국들은 지난 10년 동안 범죄에 대한 경찰 빅데이터 분석을 통하여 한정된 경찰 인력을 효과적으로 활용하기 위해 많은 노력을 하고 있다.

[0003] 최근에는 범죄예측 모델을 통한 범죄테러 감지, 용의자의 차량 인식, 추적 등을 통하여 위험 상황을 사전에 예측하여 선제적인 대응 및 수사 지원을 위한 시스템 개발이 이루어지고 있다.

[0004] 사회 안전과 치안 유지를 위해 경찰 인력이 점점 증가하지만, 인력 충원만으로는 제약이 있기 때문에 새로운 기술의 도입이 시급하다. 종래의 기술은 전통 통계 방식 기반 범죄 예측을 수행하기 때문에 예측 성능 확보에 어려움이 있다. 이에 본 발명에서는 머신러닝 기술을 활용하여 범죄 예측 모델 기반으로 범죄 발생 건수 예측, 예측 모델 요인 진단을 통해 결과 해석 기능을 제공하고자 한다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 공개특허 제2014-0100173호 (2014.08.14. 공개)

(특허문헌 0002) 대한민국 공개특허 제2017-0098281호 (2017.08.29. 공개)

발명의 내용

해결하려는 과제

[0006] 본 발명이 이루고자 하는 기술적인 과제는 치안 및 공공 데이터 기반 범죄 신고 수를 예측하는 장치 및 방법을 제공하는 것이다.

과제의 해결 수단

[0007] 본 발명의 일 실시예에 따른 예측모델 생성 장치는 크롤링 기법을 이용하여 데이터를 수집하는 데이터 수집부, 상기 데이터를 전처리하는 데이터 가공부, 및 전처리된 데이터를 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측모델을 생성하는 예측모델 생성부를 포함한다.

[0008] 본 발명의 일 실시예에 따른 예측모델 생성 방법은 적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되고, 크롤링 기법을 이용하여 데이터를 수집하는 단계, 상기 데이터를 전처리하는 단계, 및 전처리된 데이터를 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측모델을 생성하는 단계를 포함한다.

[0009] 본 발명의 일 실시예에 따른 예측 장치는 제2 데이터를 수집하는 제2 데이터 수집부, 상기 제2 데이터를 전처리하는 제2 데이터 가공부, 및 제5항 내지 제7항 중 어느 하나의 항에 기재된 예측모델 생성 방법에 의해 생성되고 전처리된 제2 데이터를 입력으로 하는 예측모델을 이용하여 예측 시점에서의 범죄 신고 수를 예측하는 예측부를 포함한다.

발명의 효과

[0010] 본 발명의 실시예에 따른 범죄 신고 수 예측 장치 및 방법에 의할 경우, 예측 시점의 범죄 신고 수를 예측할 수 있고, 예측 결과에 기반하여 관할관서별 신고와 범죄를 효율적으로 처리할 수 있는 효과가 있다.

도면의 간단한 설명

[0011] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 본 발명이 일 실시예에 따른 예측모델 생성 장치의 기능 블럭도이다.

도 2는 도 1에 도시된 예측부에 의해 예측된 범죄 발생의 중요 변수를 도시한다.

도 3은 본 발명의 일 실시예에 따른 예측 장치의 기능 블럭도이다.

도 4는 도 1에 도시된 예측모델 생성 장치에 의해 수행되는 예측모델 생성 방법 또는 도 3에 도시된 예측 장치에 의해 수행되는 예측 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0013] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0014] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0015] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등

도 마찬가지로 해석되어야 한다.

[0016] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0017] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0018] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0020] 도 1은 본 발명의 일 실시예에 따른 예측모델 생성 장치의 기능 블럭도이다.

[0021] 도 1은 참조하면, 예측모델 생성 장치(100)는 데이터 수집부(110), 데이터 가공부(120), 군집화부(130), 예측모델 생성부(140), 예측부(150), 및 저장부(160) 중 적어도 하나를 포함할 수 있다. 즉, 상술한 예측모델 생성 장치(100)의 일부 구성은 실시예에 따라 생략될 수도 있으며, 예측모델 생성부(140)는 적어도 프로세서 및/또는 메모리를 포함하는 컴퓨팅 장치로 구현될 수 있다.

[0022] 예측모델 생성 장치(100)는 데이터 수집부(110)에 의해 수집된 데이터 또는 소정의 입력 장치를 통해 수신되어 저장부(160)에 저장되어 있는 데이터를 이용하여 범죄 신고 수(또는 범죄 발생 수) 및/또는 범죄 발생 요인을 예측하는 예측모델을 생성할 수 있다. 실시예에 따라, 예측모델 생성 장치(100)는 생성된 예측모델을 이용하여 범죄 신고 수(또는 범죄 발생 수) 및/또는 범죄 발생 요인을 예측할 수 있다. 이 경우, 예측모델 생성 장치(100)는 예측 장치로 명명될 수도 있다.

[0023] 데이터 수집부(110)는 예측모델의 생성과 예측 동작을 수행하는 과정에서 필요한 데이터를 수집할 수 있다. 데이터는 미리 정해진 시간 구간 동안의 데이터로 제1 데이터 및/또는 제2 데이터 중 적어도 하나를 포함할 수 있다. 데이터 수집부(110)는 크롤링 기법을 통해 데이터를 제공하는 서버(또는 서버에 의해 제공되는 포털 사이트)로부터 데이터를 수집할 수 있다. 데이터 수집부(110)에 의해 수집된 데이터는 저장부(160)에 저장될 수 있다.

[0024] 제1 데이터는 범죄 신고 수(또는 범죄 발생 수)를 예측하는 제1 예측모델의 학습 과정 및/또는 제1 예측모델을 이용한 범죄 신고 수(또는 범죄 발생 수) 예측 과정에 이용되는 데이터로써, 범죄 신고 데이터를 의미할 수 있다. 아래 표 1은 범죄 신고 데이터의 일 예를 나타낸다.

표 1

[0025]

월일	시간	신고 건수	…
4월 2일	22~23시	4	…
4월 2일	23~24시	2	…
4월 3일	00~01시	2	…
4월 3일	01~02시	2	…
4월 3일	02~03시	3	…
4월 3일	03~04시	2	…
4월 3일	04~05시	6	…
4월 3일	05~06시	5	…
4월 3일	06~07시	8	…
4월 3일	07~08시	3	…
4월 3일	08~09시	3	…
4월 3일	09~10시	5	…

…	…	…	…

[0026] 제1 데이터에는 범죄 신고가 발생한 일자, 시간, 장소, 신고 내용, 범죄의 유형, 중 적어도 하나를 포함할 수 있다. 범죄 신고가 발생한 시간은 신고가 접수된 시간을 의미하거나 신고가 접수된 시간 구간(시간대)을 의미할 수도 있다. 시간 구간은 미리 정해진 시간 단위로 구분될 수 있다. 예컨대, 시간 구간은 하루 24시간을 n(n은 정수) 시간 단위로 분할한 것으로, 0~1시, 1~2시, …, 23~24시 중 어느 하나를 의미할 수 있다. 범죄 신고가 발생한 시간이 시간 구간으로 정의된 경우, 제1 데이터는 해당 시간 구간 동안에 접수된 신고 건수를 포함할 수도 있다. 장소는 범죄가 발생한 장소 또는 신고된 장소를 의미할 수 있다. 실시예에 따라, 장소는 장소를 식별할 수 있는 좌표(예컨대, GPS 좌표), 주소(신고자가 위치한 위치의 주소 또는 범죄가 발생한 위치의 주소), 행정동 단위, 경찰서, 지구대, 파출소의 관할 구역 등을 의미할 수 있다. 신고 내용은 범죄 신고자의 신고 내용으로 신고 내용이 담긴 음성 데이터, 신고 내용이 기재된 텍스트 데이터, 또는 신고 내용을 요약한 요약 데이터 등을 의미할 수 있다. 범죄의 유형은 신고된 범죄의 유형을 의미하는 것으로, 가정, 강도·마약, 교통, 사기, 살인, 여성, 자살, 절도, 질서 유지, 청소년, 폭력, 풍속, 기타 등 중 어느 하나를 의미할 수 있다. 대한민국의 경우, 범죄 신고 데이터는 긴급 전화번호 112를 통해 접수된 범죄 신고 데이터에 해당한다. 상술한 제1 데이터에 포함된 내용은 데이터 수집부(110)의 크롤링 결과일 수도 있으나, 적어도 일부는 후술할 데이터 가공부(120)에 의해 가공된 결과일 수도 있다.

[0027] 제2 데이터는 범죄 발생 요인들 중 가장 중요도를 갖는 적어도 하나의 범죄 발생 요인을 예측하는 제2 예측모델의 학습 과정 및/또는 제2 예측모델을 이용한 범죄 발생 요인 예측 과정에 이용되는 데이터를 의미할 수 있다. 제2 데이터는 기상 데이터, 집회 데이터, 공휴일 데이터, 지역 공공 데이터, 및 유동인구 데이터 중 적어도 하나를 의미할 수 있다. 상술한 제2 데이터에 포함된 내용은 데이터 수집부(110)의 크롤링 결과일 수도 있으나, 적어도 일부는 후술할 데이터 가공부(120)에 의해 가공된 결과일 수도 있다.

[0028] 기상 데이터는 과거의 기상 데이터 및/또는 미래의 기상 예보 데이터를 포함할 수 있다. 또한, 기상 데이터는 일정 구역(도, 시, 구 등)에 관한 과거의 기상 데이터 및/또는 기상 예보 데이터를 포함할 수도 있다.

[0029] 집회 데이터는 집회를 하기 위해 신고된 집회 신고 데이터, 과거의 집회 발생 데이터 등을 의미할 수 있으며, 집회 일시, 집회 장소, 집회 인원 등에 관한 정보를 포함할 수 있다.

[0030] 공휴일 데이터는 법정 공휴일, 임시 공휴일 등에 관한 데이터를 의미할 수 있다.

[0031] 지역 공공 데이터는 해당 지역(예컨대, 미리 정의된 구역(구, 관할경찰서, 관할관서 등))의 보육시설(어린이집, 유치원 등)의 유무, 해당 월, 해당 지역의 범죄발생현황(5대범죄발생현황, 발생율, 발생빈도 등), 해당 지역의 공원 분포 여부, 해당 지역의 사망원인별 사망통계, 해당 지역의 공공체육시설 유무, 해당 지역의 노인여가복지시설 유무, 해당 지역의 장애인 현황(장애인 분포율 등), 해당 지역의 저소득한부모가족 현황(분포율 등), 해당 지역의 용도지역 구분, 해당 지역의 국공립중학교 유무, 해당 지역의 공동주택건립 여부, 해당 지역의 노인주거복지 시설의 유무, 해당 지역의 의료기관 유무, 해당 지역의 국공립고등학교 유무, 해당 지역의 빈집 현황, 해당 지역의 주민등록 인구 등 중 적어도 하나를 의미할 수 있다.

[0032] 유동인구 데이터는 해당 지역의 특정 시점(또는 특정 시간 구간)에서의 유동인구를 의미할 수 있으며, 통신사 등으로부터 제공받을 수 있다.

[0033] 실시예에 따라, 데이터 수집부(110)는 예측모델 생성 장치(100)에서 생략될 수 있다. 이 경우, 예측모델 생성 장치(100)는 미리 저장부(160)에 저장되어 있는 데이터를 이용하여 예측모델의 생성 및/또는 예측 동작을 수행할 수 있다.

[0034] 데이터 가공부(120)는 데이터 수집부(110)에 의해 수집된 데이터 및/또는 미리 수신되어 저장부(160)에 저장되어 있는 데이터를 전처리할 수 있다. 구체적으로, 제1 데이터에 포함된 범죄신고 데이터에 구체적인 시간이 포함되어 있는 경우, 미리 정해진 시간 구간 별로 그룹화하여 해당 시간 구간과 해당 시간 구간의 신고 건수를 포함하는 범죄 신고 데이터로 변환할 수 있다. 제1 데이터에 포함된 범죄신고 데이터에 구체적인 장소(GPS 좌표, 주소 등)가 포함되어 있는 경우, 데이터 가공부(120)는 해당 범죄 신고 데이터를 관할경찰서, 미리 정해진 해당 구역(구, 동 등), 및/또는 관할관서(관할 지구대 또는 관할 파출소)를 추가하거나 변환할 수 있다. 제1 데이터에 구체적인 신고 내용이 포함되어 있으나 범죄의 유형이 포함되어 있지 않는 경우, 범죄 신고에 기초하여 범죄의 유형을 결정하고 범죄 유형이 포함된 범죄 신고 데이터로 변환하거나 추가할 수 있다. 이를 위해, 데이터 가공부(120)는 범죄 신고 데이터에 포함된 음성 데이터를 텍스트 데이터로 변환할 수 있고, 텍스트 데이터에 기재

된 키워드를 추출하는 방식 등으로 범죄의 유형을 결정할 수 있다.

[0035] 또한, 데이터 가공부(120)는 제1 데이터 및/또는 제2 데이터에 결측치가 있는 데이터를 삭제할 수도 있다.

[0036] 데이터 가공부(120)에 의해 가공된 데이터는 저장부(160)에 저장될 수 있다. 또한, 제1 데이터 및/또는 제2 데이터에 전처리 동작이 불필요한 경우 데이터 가공부(120)는 예측모델 생성 장치(100)에서 생략될 수 있다. 이 경우, 예측모델 생성 장치(100)는 데이터 수신부(110)에 의해 수신된 데이터 또는 저장부(160)에 미리 저장되어 있는 데이터를 이용하여 예측모델의 생성 및/또는 예측 동작을 수행할 수 있다.

[0037] 군집화부(130)는 시계열적인 범죄 신고 데이터를 이용하여 관할관서들을 범죄 신고 패턴이 유사한 관할관서로 그룹화할 수 있다.

[0038] 시계열적인 범죄 신고 데이터는 제1 데이터를 의미할 수 있다. 이때, 제1 데이터는 미리 정해진 시간 단위(예컨대, 1일 또는 n 시간)로 관할관서별 범죄 신고 건수를 포함할 수 있다.

[0039] 또한, 군집화부(130)는 시계열 데이터의 패턴의 유사도(또는 거리)를 측정하고 측정된 유사도(또는 거리)에 따라 관할관서들을 그룹화할 수 있으며, 소정의 군집화 알고리즘이 이용될 수 있다. 예시적인 군집화 알고리즘은 DTW(Dynamic time warping)를 거리척도로 활용한 K-means 알고리즘이다.

[0040] 예측모델 생성부(140)는 데이터 수신부(110)에 의해 수신된 데이터, 데이터 가공부(120)에 의해 전처리된 데이터, 또는 미리 수신되어 저장부(160)에 저장되어 있는 데이터를 이용하여 범죄 신고 수 및/또는 범죄 발생 요인을 예측하는 예측모델을 생성할 수 있다. 예측모델은 제1 데이터 및/또는 제2 데이터를 이용하여 범죄 신고 수를 예측하는 제1 예측모델과 제2 데이터를 이용하여 범죄 발생 요인을 예측하는 제2 예측모델 중 적어도 하나를 포함할 수 있다.

[0041] 구체적으로, 예측모델 생성부(140)는 제1 데이터를 이용하여 시계열 모델, 예컨대, RNN(Recurrent Neural Network)과 같은 인공 신경망을 학습시킴으로써 제1 예측모델을 생성할 수 있다. 구체적으로, 예측모델 생성부(140)는 제1 데이터를 이용하여 예측 시점 이전의 범죄 신고 수를 입력받아 예측 시점으로부터 미리 정해진 시간 구간(예측 구간) 동안의 범죄 신고 수를 예측하도록 인공 신경망을 학습시킴으로써 제1 예측모델을 생성할 수 있다. 범죄 신고 수의 예측은 예측 동작을 수행하는 시점으로부터 k일(k는 정수) 후까지 수행될 수 있다. 예컨대, 제1 예측모델은 예측 동작을 수행하는 시점으로부터 7일 후까지의 범죄 신고 수를 일별로 예측할 수 있다. 아래 표 2는 예측 동작을 수행하는 시점으로부터 7일 후까지 일별로 범죄 신고 수를 예측하는 예를 나타낸다.

표 2

[0042]

예측 시점	예측 건수(범죄 신고 수)
1일 후(내일)	49
2일 후	34
3일 후	79
4일 후	23
5일 후	44
6일 후	61
7일 후	39

[0043] 실시예에 따라, 제1 예측 모델은 미리 정해진 시간 단위(예컨대, 6시간 단위) 별로 범죄 신고 수를 예측할 수도 있다. 다른 예로, 제1 예측 모델은 일단위 별로 범죄 신고 수를 예측하되, 적어도 일부의 예측 시점(예컨대, 1일 후)에 관하여는 미리 정해진 시간 단위 별로 범죄 신고 수를 예측할 수도 있다.

[0044] 실시예에 따라, 제1 예측 모델은 군집화부(130)에 의해 군집화된 각 군집들 각각에 대한 예측모델들을 포함할 수도 있다. 이 경우, 예측모델 생성부(140)는 각 군집들 각각에 대응하는 학습 데이터(제1 데이터 중 해당 군집에 관한 데이터)를 이용하여 군집들 각각에 대한 예측모델을 학습시킬 수 있다.

[0045] 또한, 예측모델 생성부(140)는 제2 데이터를 이용하여 범죄 발생 요인을 결정하는 제2 예측모델을 생성할 수 있다. 예컨대, 예측모델 생성부(140)는 제2 데이터를 이용하여, 범죄 발생에 가장 큰 영향을 미친 적어도 하나의 중요 변수를 추출하는 의사결정나무 모델(또는 부스팅 의사결정나무 모델)을 학습시킴으로써 제2 예측모델을 생성할 수 있다. 부스팅 의사결정나무 모델은 여러 개의 의사결정나무를 결합하여 하나의 의사결정나무보다 더 좋

은 결과를 출력하도록 하는 머신러닝 기법이다. 의사결정나무는 분류와 회귀가 모두 가능한 지도학습 모델 중 하나이다. 예시적인 부스팅 의사결정나무 모델은 CatBoost(Categorical Boost) 기법이다.

[0046] 예측부(150)는 예측모델 생성부(140)에 의해 생성된 예측모델을 이용하여 예측동작을 수행할 수 있다. 구체적으로, 예측부(150)는 제1 예측모델 및 제2 예측모델을 이용하여 범죄 신고 수를 예측할 수 있고, 제2 예측모델을 이용하여 범죄 발생의 중요 변수를 추가적으로 예측할 수도 있다.

[0047] 제1 예측모델 및 제2 예측모델을 이용하여 범죄 신고 수를 예측하는 경우, 예측부(150)는 제1 데이터(예컨대, 최근 30일 간의 범죄 신고 데이터)를 이용하여 미리 정해진 구역(예컨대, 관할 경찰서 또는 관할관서) 별로 범죄 신고 수를 예측할 수 있다. 이때, 하나의 제1 예측모델을 이용하여 각 구역의 범죄 신고 수를 예측하거나, 구역별로 특화된 예측모델(즉, 해당 구역이 속한 군집의 데이터를 이용하여 생성된 예측모델)을 이용하여 해당 구역의 범죄 신고 수를 예측할 수 있다. 또한, 예측부(150)는 제1 데이터를 이용하여 미리 정해진 예측 기간 동안이 범죄 신고 수를 예측할 수 있다.

[0048] 제2 예측모델을 이용하여 범죄 발생의 중요 변수를 예측하는 경우, 예측부(150)는 제1 데이터 및/또는 제2 데이터를 이용하여 예측 시점(예컨대, 3일 후)에서의 범죄 발생의 중요 변수를 추출할 수 있다. 제2 예측모델을 이용한 범죄 발생의 중요 변수 예측 결과는 도 2에 도시되어 있다.

[0049] 예측부(150)의 예측 결과를 통해 효율적인 인력 배치가 가능하다. 예컨대, 범죄 신고 수가 많을 것으로 예측되는 시점의 경찰 근무 인력을 탄력적으로 증원하거나, 범죄 신고 수가 많을 것으로 예측되는 구역의 경찰 근무 인력을 증원할 수 있다. 또한, 범죄 발생의 중요 변수에 따라, 효과적으로 사전 대비가 가능하다. 일 예로, 범죄 발생의 중요 변수가 '보육시설'인 경우, 보육시설의 순찰과 경계를 강화함으로써 범죄 발생을 억제하거나 조기 검거가 가능할 것이다.

[0050] 저장부(160)에는 데이터 수집부(110)에 의해 수집된(또는 수신된) 데이터, 데이터 가공부(120)에 의해 전처리된 데이터, 군집화부(130)에 의한 군집화 결과, 예측모델 생성부(140)에 의해 생성된 예측모델, 예측부(150)에 의한 예측 결과 등이 저장될 수 있다.

[0052] 도 3은 본 발명의 일 실시예에 따른 예측 장치의 기능 블럭도이다. 도 3에 도시된 예측 장치(300)를 설명함에 있어서, 앞서 기재된 내용과 중복되는 내용에 관하여는 그 기재를 생략하기로 한다.

[0053] 도 3을 참조하면, 예측 장치(300)는 데이터 수집부(310), 데이터 가공부(320), 예측부(330), 및 저장부(340) 중 적어도 하나를 포함할 수 있다. 즉, 예측 장치(300)의 일부 구성은 실시예에 따라 생략될 수도 있으며, 예측 장치(300)는 적어도 프로세서 및/또는 메모리를 포함하는 컴퓨팅 장치로 구현될 수 있다.

[0054] 예측 장치(300)는 데이터 수집부(310)에 의해 수집된 데이터 또는 소정의 입력 장치를 통해 수신되어 저장부(340)에 저장되어 있는 데이터를 이용하여 범죄 신고 수(또는 범죄 발생 수) 및/또는 범죄 발생 요인을 예측하는 예측모델을 이용하여 범죄 신고 수(또는 범죄 발생 수) 및/또는 범죄 발생 요인을 예측할 수 있다. 예측모델은 도 1에 도시된 예측모델 생성 장치(100)의 의해 생성되거나 도 1에 도시된 예측모델 생성 장치(100)에 의해 수행되는 예측모델 생성 방법을 통해 생성된 예측모델을 의미할 수 있다.

[0055] 데이터 수집부(310)는 예측 동작을 수행하는 과정에서 필요한 데이터를 수집할 수 있다. 데이터는 미리 정해진 시간 구간 동안의 데이터로 제1 데이터 및/또는 제2 데이터 중 적어도 하나를 포함할 수 있다. 데이터 수집부(310)에 의해 수집된 데이터는 저장부(340)에 저장될 수 있다. 데이터 수집부(310)의 동작은 도 1에 도시된 데이터 수집부(110)의 동작과 동일할 수 있으므로 구체적인 설명은 생략하기로 한다.

[0056] 실시예에 따라, 데이터 수집부(310)는 예측 장치(300)에서 생략될 수 있다. 이 경우, 예측 장치(300)는 미리 저장부(340)에 저장되어 있는 데이터를 이용하여 예측 동작을 수행할 수 있다.

[0057] 데이터 가공부(320)는 데이터 수집부(310)에 의해 수집된 데이터 및/또는 미리 수신되어 저장부(340)에 저장되어 있는 데이터를 전처리할 수 있다. 데이터 가공부(32)의 구체적인 동작은 도 1에 도시된 데이터 가공부(120)의 동작과 동일할 수 있으므로 구체적인 설명은 생략하기로 한다.

[0058] 데이터 가공부(320)에 의해 가공된 데이터는 저장부(340)에 저장될 수 있다. 또한, 제1 데이터 및/또는 제2 데이터에 전처리 동작이 불필요한 경우 데이터 가공부(320)는 예측 장치(300)에서 생략될 수 있다. 이 경우, 예측 장치(300)는 데이터 수신부(310)에 의해 수신된 데이터 또는 저장부(340)에 미리 저장되어 있는 데이터를 이용

하여 예측 동작을 수행할 수 있다.

[0059] 예측부(330)는 도 1에 도시된 예측모델 생성부(140)에 의해 생성된 예측모델을 이용하여 예측 동작을 수행할 수 있다. 구체적으로, 예측부(330)는 제1 예측모델을 이용하여 범죄 신고 수를 예측할 수 있고, 제2 예측모델을 이용하여 범죄 발생의 중요 변수를 예측할 수도 있다. 예측부(330)의 구체적인 동작은 도 1의 예측부(330)의 동작과 동일할 수 있으므로 구체적인 설명은 생략하기로 한다.

[0060] 저장부(340)에는 데이터 수집부(310)에 의해 수집된(또는 수신된) 데이터, 데이터 가공부(320)에 의해 전처리된 데이터, 예측모델, 예측부(330)에 의한 예측 결과 등이 저장될 수 있다.

[0062] 도 4는 도 1에 도시된 예측모델 생성 장치 또는 도 3에 도시된 예측 장치에 의해 수행되는 예측모델 생성 방법 또는 예측 방법을 설명하기 위한 흐름도이다.

[0063] 우선, 예측모델을 생성하기 위한 데이터 또는 예측 동작을 수행하기 위한 데이터가 수집된다(S100). 데이터는 제1 데이터 및/또는 제2 데이터를 포함할 수 있다. 실시예에 따라, 데이터는 소정의 저장 장치를 통하여 미리 수신되어 저장되어 있을 수 있으며, 이 경우 S100 단계는 생략될 수 있다.

[0064] 수신된 데이터 또는 미리 수신되어 저장되어 있는 데이터는 예측모델을 생성하기 위한 형식 또는 예측 동작을 수행하기 위한 형식으로 가공(또는 전처리)된다(S200). 수신된 데이터 또는 미리 수신되어 저장되어 있는 데이터는 예측모델을 생성하기 위한 형식 또는 예측 동작을 수행하기 위한 형식일 수 있으며, 이 경우 S200 단계는 생략될 수 있다.

[0065] 이후, 복수의 관할관서들이 범죄 신고 수의 패턴에 따라 그룹화(또는 군집화)된다(S300). 이는 각 그룹 별로 상이한 예측모델을 생성하기 위한 것으로, 하나의 예측모델을 생성하는 경우, S300 단계는 생략될 수 있다. 또한, 예측모델 생성 방법이 아닌 예측 방법은 S300 단계를 포함하지 않을 수 있다.

[0066] 다음으로, 적어도 하나의 예측모델이 생성될 수 있다(S400). 예측모델 생성 방법에 관한 상세한 설명은 생략하기로 한다.

[0067] 추가적으로, 예측 단계가 수행될 수 있다(S500). 예측 단계는 생성된 예측모델을 이용하여 범죄 신고 수(또는 범죄 발생 수) 및/또는 범지 발생의 중요 변수들이 예측될 수 있다.

[0069] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPA(Field Programmable array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

[0070] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0071] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-optical Media), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0072] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0073] 100 : 예측모델 생성 장치

110 : 데이터 수집부

120 : 데이터 가공부

130 : 군집화부

140 : 예측모델 생성부

150 : 예측부

160 : 저장부

300 : 예측 장치

310 : 데이터 수집부

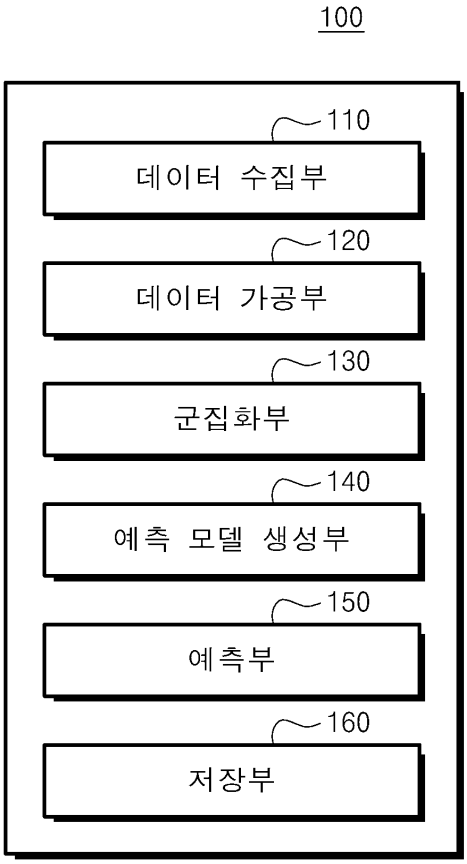
320 : 데이터 가공부

330 : 예측부

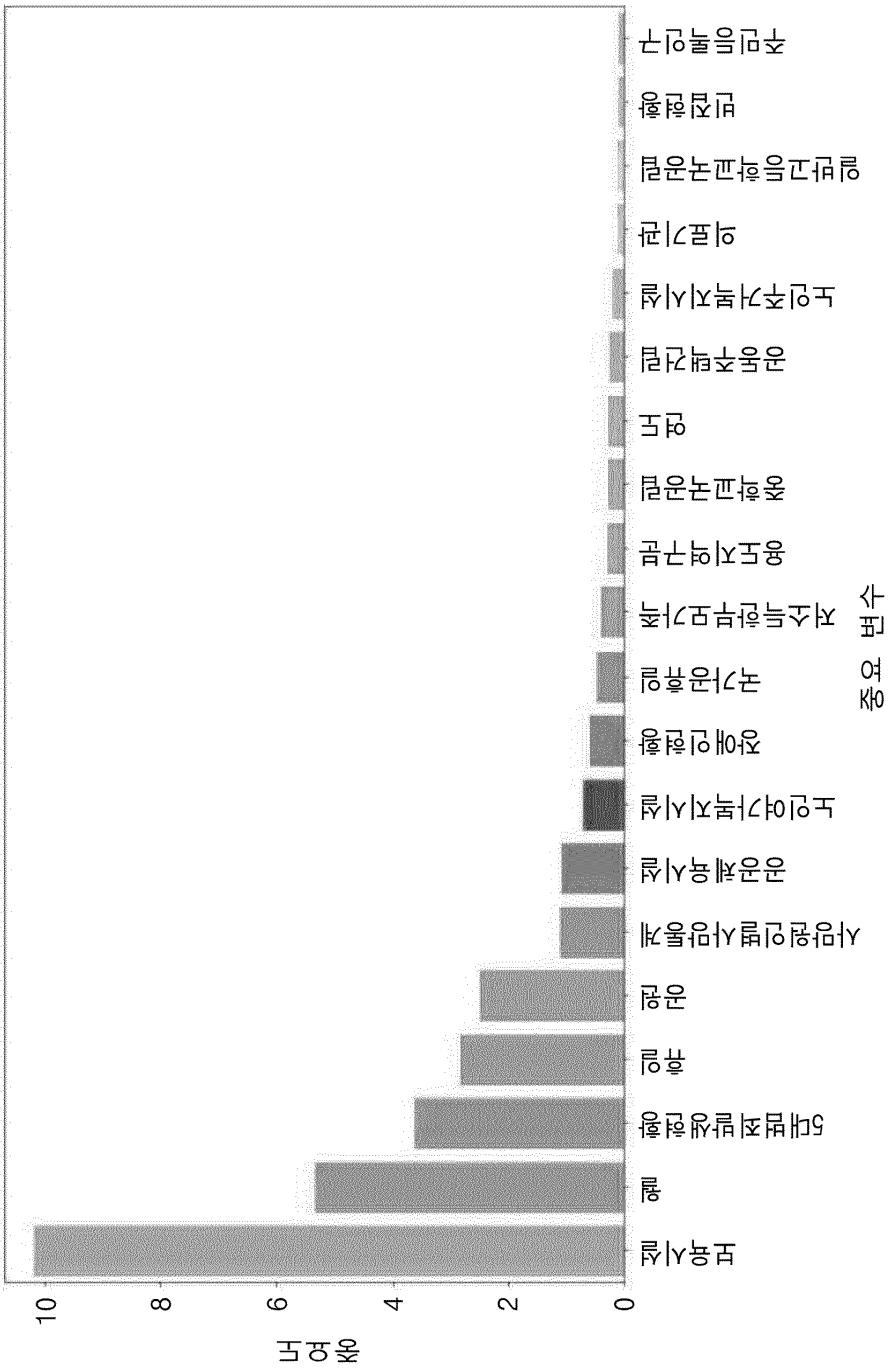
340 : 저장부

도면

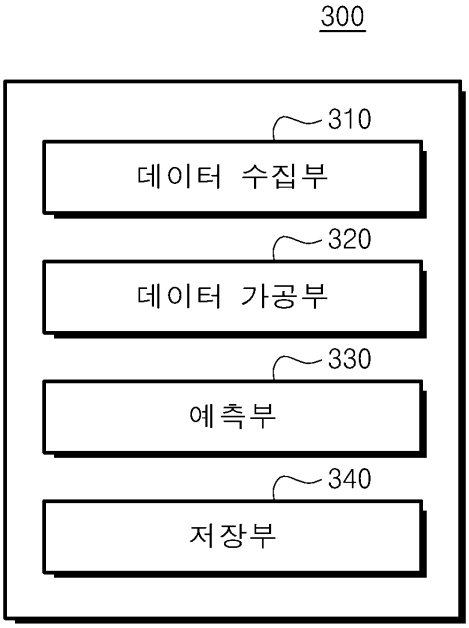
도면1


100
110
데이터 수집부
120
데이터 가공부
130
군집화부
140
예측 모델 생성부
150
예측부
160
저장부

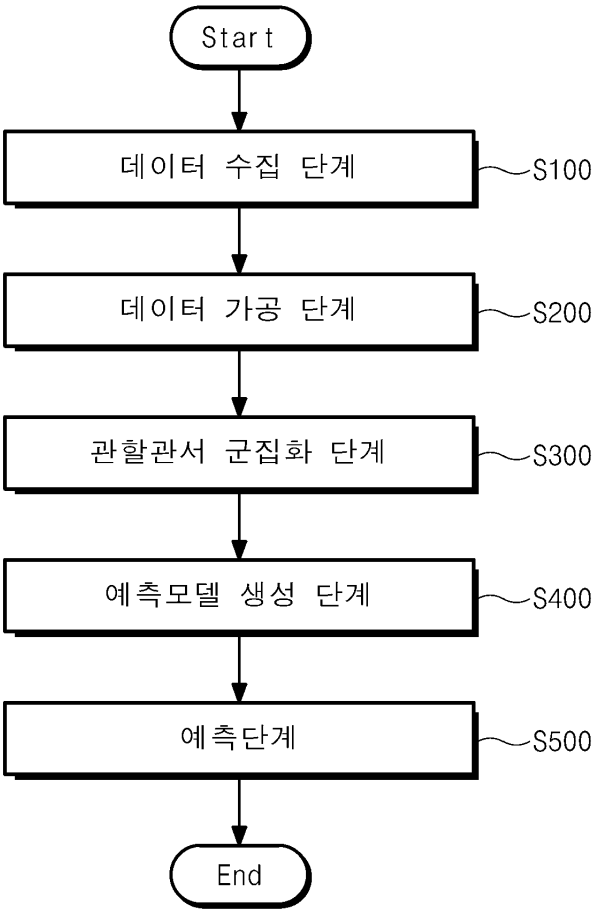
도면2


중요도
0
2
4
6
8
10
보육시설
월
5대범죄발생현황
휴일
공원
사망원인별사망통계
공공체육시설
노인여가복지시설
장애인현황
국가공휴일
저소득한부모가족
용도지역구분
중학교국공립
연도
공동주택건립
노인주거복지시설
의료기관
일반고등학교국공립
빈집현황
주민등록인구
여성
댐사

도면3


300
310
데이터 수집부
320
데이터 가공부
330
예측부
340
저장부

도면4


Start
데이터 수집 단계
S100
데이터 가공 단계
S200
관할관서 군집화 단계
S300
예측모델 생성 단계
S400
예측단계
S500
End