



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년11월20일
 (11) 등록번호 10-1463425
 (24) 등록일자 2014년11월13일

(51) 국제특허분류(Int. Cl.)
 G06F 19/00 (2011.01) G06F 17/15 (2006.01)
 G06F 17/16 (2006.01)

(21) 출원번호 10-2012-0135925
 (22) 출원일자 2012년11월28일
 심사청구일자 2012년11월28일
 (65) 공개번호 10-2014-0068436
 (43) 공개일자 2014년06월09일
 (56) 선행기술조사문헌
 Heesung Yoon 외 4인, 'A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer', A compar Journal of Hydrology 396, 2011.01.,*
 Xiaoqi Peng 외 2인, 'Outlier Detection Method Based on SVM and Its Application in Copper-matte Converting', Control and Decision Conference, 2010.05.26., pp.628-631*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 한국지질자원연구원
 대전광역시 유성구 과학로 124 (가정동)

(72) 발명자
 윤희성
 대전 서구 월평동로 83, 109동 506호 (월평동, 다모아아파트)
 하규철
 대전 유성구 은구비남로 34, 804동 1301호 (노은동, 열매마을8단지)
 김용철
 대전광역시 유성구 과학로 124 한국지질자원연구원 기숙사 104호

(74) 대리인
 김정수

전체 청구항 수 : 총 3 항

심사관 : 박재용

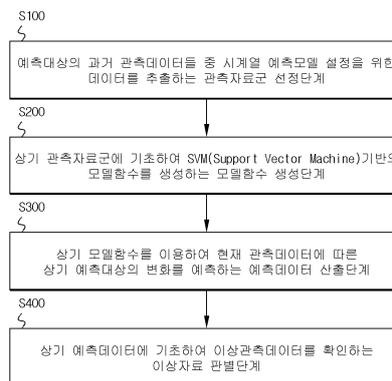
(54) 발명의 명칭 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법

(57) 요약

본 발명은 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법에 관한 것으로, 보다 상세하게는 과거의 관측자료들을 이용하여 예측된 결과에 기초하여 이상관측자료를 판별할 수 있도록 한 것이다.

(뒷면에 계속)

대표도 - 도1



특히, 본 발명은 지지벡터기계(Support Vector Machine, SVM)알고리즘을 이용하여, 과거에 측정된 관측자료에 의한 예측결과와 과거의 예측자료에 의한 예측결과를 비교함으로써, 실시간으로 이상관측자료를 판별할 수 있다.

구체적으로, 지하수위의 자연적인 변화에 영향을 주는 강우 등의 기상 자료와 지하수위 관측 자료를 이용하여, 지하수위 예측 시계열 모델을 만들고 이를 이용함으로써, 지하수위에 대한 이상관측자료의 탐지능력을 크게 향상시킬 수 있다.

따라서, 수자원 관리 분야, 특히 지하수 관리 분야는 물론, 다양한 기후 변화와 연관된 자연자원의 관리 분야 및 이와 연관 내지 유사한 분야에서 신뢰성 및 경쟁력을 향상시킬 수 있다.

이 발명을 지원한 국가연구개발사업

과제고유번호	GP2012-013
부처명	지식경제부
연구관리전문기관	산업기술연구회
연구사업명	주요사업-기관고유입무형-기본
연구과제명	대수층 인공함양 지하수 확보 융복합 핵심기술 개발
기 여 율	1/1
주관기관	한국지질자원연구원
연구기간	2012.01.01 ~ 2012.12.31

특허청구의 범위

청구항 1

삭제

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

예측대상인 지하수위 및 상기 지하수위에 영향을 미치는 강우량을 포함하는 영향인자의 과거 관측데이터들 중 이상관측이 발생하지 않은 자료군을 추출하는 관측자료군 선정단계;

상기 관측자료군에 기초하여 SVM(Support Vector Machine)기반으로, 과거의 관측데이터를 이용하여 예측을 수행하는 실측기반모델함수 및 과거의 예측데이터를 이용하여 예측을 수행하는 예측기반모델함수를 생성하는 모델함수 생성단계;

상기 실측기반모델함수 및 예측기반모델함수 각각에 대하여, 현재 관측데이터에 따른 상기 지하수위의 변화를 예측하는 예측데이터 산출단계; 및

상기 실측기반모델함수의 예측데이터 및 상기 예측기반모델함수의 예측데이터를 비교하고, 두 예측데이터의 차이값이 이상판단기준값보다 큰 경우, 해당 데이터를 이상관측데이터로 결정하는 이상자료 판별단계;를 포함하고,

상기 모델함수 생성단계는,

SVM기반의 시계열 예측을 위한 모델의 구조를 설정하는 구조설정단계; 및

SMO(Sequential Minimal Optimization)알고리즘으로 예측함수를 설정하고 최적화하는 최적화예측함수 생성단계;를 포함하여 상기 최적화예측함수에 의해 상기 모델함수를 생성하며,

상기 구조설정단계는,

상기 현재 관측데이터에 따른 입력벡터(Input vector);

상기 관측자료군에서 추출되는 지지벡터(SV, Support vector);

상기 입력벡터 및 지지벡터를 커널트릭(Kernel trick)으로 고차원 변환하는 커널함수(Kernel function); 및

상기 커널함수를 이용하여 예측데이터를 산출하는 출력함수(Output)를 포함하여, 상기 모델의 구조를 설정하고,

상기 최적화예측함수 생성단계는,

상기 SVM기반의 예측함수를 상기 SMO알고리즘으로 설정하는 단계;

상기 예측함수를 계산하기 위한 목적함수를 설정하는 단계; 및

상기 목적함수에 최적화 조건을 설정하여 상기 최적화예측함수로 변환하는 단계를 포함하는 것을 특징으로 하는 컴퓨터가 수행하는 지하수위의 이상관측자료 탐지 방법.

청구항 8

삭제

청구항 9

삭제

청구항 10

제 7항에 있어서,

상기 커널함수는,

가우시안(Gaussian) 형태의 "radial basis kernel" 함수를 포함하는 것을 특징으로 하는 컴퓨터가 수행하는 지하수위의 이상관측자료 탐지 방법.

청구항 11

삭제

청구항 12

제 7항에 있어서,

상기 최적화예측함수 생성단계는,

상기 목적함수를 계산하기 위하여 라그랑제 승수(Lagrangian multiplier)를 이용한 안장점(Saddle point) 문제를 적용하는 것을 특징으로 하는 컴퓨터가 수행하는 지하수위의 이상관측자료 탐지 방법.

명세서

기술분야

[0001] 본 발명은 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법에 관한 것으로서, 보다 상세하게는 과거의 관측자료들을 이용하여 예측된 결과에 기초하여 이상관측자료를 판별할 수 있도록 한 것이다.

[0002] 특히, 본 발명은 지지벡터기계(Support Vector Machine, SVM)알고리즘을 이용하여, 과거에 측정된 관측자료에 의한 예측결과와 과거의 예측자료에 의한 예측결과를 비교함으로써, 이상관측자료를 판별할 수 있는 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법에 관한 것이다.

배경기술

[0003] 기후변화는 오늘날 가장 심각한 환경문제의 하나로, 그 규모가 전 지구적으로 광범위하게 일어나고 있으며, 우리나라에도 여러 분야에서 그 징후가 나타나고 있을 뿐만 아니라, 이에 따른 영향이 사회전반에 걸쳐 나타나고 있다.

[0004] 또한, 기후변화는 홍수, 가뭄, 태풍 등의 기상변화 및 생태계의 변화와 연관되어 있으며, 이러한 기후변화에 대응하기 위해서는, 기후변화의 영향에 따른 평가가 분야별로 요구되며, 이를 위해서는 자료의 신뢰성 확보가 가

장 중요하다고 할 수 있다.

- [0005] 특히, 우리의 생활과 가장 밀접한 수자원은, 기후변화뿐만 아니라 산업구조의 급격한 변화와 인구증가에 따른 용수수요의 증가 및 생태환경에 대한 인식변화 등으로 인하여, 지표수 이용에 많은 한계를 보이고 있으며, 이에 따른 대체 수자원으로서 지하수의 역할이 증대되고 있다.
- [0006] 그러나, 지하수의 개발과 이용에 대한 체계적인 관리가 이루어지지 않을 경우, 지하수를 포함하는 수자원에 대한 정확한 파악이 어렵고, 무분별한 지하수의 개발 등으로 인하여, 지하수의 고갈 등과 같은 관리적인 문제를 유발하고 있다.
- [0007] 한편, 유무선 통신망을 이용한 원격관리시스템 및 자동관측(계측)기술이 발전하면서, 국가지하수관측망, 보조지하수관측망, 해수침투관측망 등 지하수 실시간 관측 시스템들이 구축되고 있다.
- [0008] 그러나, 전자소자와 프로그램을 이용한 장치 및 시스템에서는, 하드웨어의 이상이나 소프트웨어의 연산오류 등이 발생할 수 있으며, 이로 인해 관측(계측)자료에 오류가 발생할 수 있다.
- [0009] 따라서, 관측자료의 신뢰성 확보 및 이를 통한 효율적인 관측망 운영과 지하수자원의 지속가능한 활용을 위해서는, 관측자료에 대한 실시간 이상관측자료 탐지방법이나 오류확인방법 등이 요구되고 있다.
- [0010] 이러한 방법 중 하나로는, 실제 관측된 자료와 과거 관측자료 통계값의 비교를 통해서 이상자료를 탐지할 수 있겠으나, 이러한 방법은 해당 지역 지하수위의 자연적인 변화 양상을 고려하기 힘들다는 점에서 제한적이라 할 수 있다.
- [0011] 하기의 종래 특허문헌에서도, 해당 지역 지하수위의 자연적인 변화 양상을 고려하여 이상자료를 실시간으로 탐지함으로써, 충분한 신뢰성을 보장할 수 있는 관측자료를 제공하는 방법에 대해서는 나타나 있지 않다.

선행기술문헌

특허문헌

- [0012] (특허문헌 0001) 대한민국 등록특허공보 제10-1013182호 "지하수의 수위 변화를 이용한 지하수 양수량 산출방법, 수위측정센서 및 양수량 산출방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체"

발명의 내용

해결하려는 과제

- [0013] 상기와 같은 문제점을 해결하기 위해서, 본 발명은 지하수위의 자연적인 변화에 영향을 주는 강우 등의 기상 자료와 지하수위 관측 자료를 이용하여 지하수위 예측 시계열 모델을 만들고 이를 활용하여 이상 관측 자료를 탐지할 수 있는 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법을 제공하는데 목적이 있다.
- [0014] 다시 말해, 예측대상에 대하여 과거의 관측자료들을 이용하여 예측된 결과에 의해 이상관측자료를 판별할 수 있도록 한 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법을 제공하는데 목적이 있다.
- [0015] 특히, 본 발명은 지지벡터기계(Support Vector Machine, SVM)알고리즘을 이용하여, 과거에 측정된 관측자료에 의한 예측결과와 과거의 예측자료에 의한 예측결과를 비교함으로써, 이상관측자료를 판별에 대한 신뢰성을 향상시킬 수 있는 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법을 제공하는데 목적이 있다.

과제의 해결 수단

- [0016] 상기와 같은 목적을 달성하기 위해서, 본 발명에 따른 시계열 예측 모델을 이용한 이상관측자료 탐지 방법은,

예측대상의 과거 관측데이터들 중 시계열 예측모델 설정을 위한 데이터를 추출하는 관측자료군 선정단계; 상기 관측자료군에 기초하여 SVM(Support Vector Machine)기반의 모델함수를 생성하는 모델함수 생성단계; 상기 모델함수를 이용하여 현재 관측데이터에 따른 상기 예측대상의 변화를 예측하는 예측데이터 산출단계; 및 상기 예측데이터에 기초하여 이상관측데이터를 확인하는 이상자료 판별단계를 포함한다.

- [0017] 또한, 상기 모델함수 생성단계는, SVM기반의 시계열 예측을 위한 모델의 구조를 설정하는 구조설정단계; 및 상기 SVM기반의 예측함수를 설정하고 최적화하는 최적화예측함수 생성단계를 포함하고, 상기 최적화예측함수에 의해 상기 모델함수를 생성할 수 있다.
- [0018] 또한, 상기 구조설정단계는, 상기 현재 관측데이터에 따른 입력벡터(Input vector); 상기 관측자료군에서 추출되는 지지벡터(SV, Support vector); 상기 입력벡터 및 지지벡터를 고차원 변환하는 차원변환함수; 및 상기 커널함수를 이용하여 예측데이터를 산출하는 출력함수(Output)를 포함하여, 상기 모델의 구조를 설정할 수 있다.
- [0019] 또한, 상기 최적화예측함수 생성단계는, 상기 SVM기반의 예측함수를 설정하는 단계; 상기 예측함수를 계산하기 위한 목적함수를 설정하는 단계; 및 상기 목적함수에 최적화 조건을 설정하여 상기 최적화예측함수로 변환하는 단계를 포함할 수 있다.
- [0020] 또한, 상기 예측데이터 산출단계는, 상기 모델함수에 과거의 관측데이터 및 현재의 관측데이터를 적용하여 예측된 실측기반 예측데이터; 및 상기 모델함수에 과거의 예측데이터 및 현재의 관측데이터를 적용하여 예측된 예측기반 예측데이터 중 적어도 하나를 산출할 수 있다.
- [0021] 또한, 상기 이상자료 판별단계는, 상기 실측기반예측데이터 및 예측기반예측데이터를 비교하고, 두 예측데이터의 차이값에 기초하여 이상관측데이터를 판별할 수 있다.
- [0022] 또한, 본 발명에 따른 지하수위의 이상관측자료 탐지 방법은, 예측대상인 지하수위 및 상기 지하수위에 영향을 미치는 강우량을 포함하는 영향인자의 과거 관측데이터들 중 이상관측이 발생하지 않은 자료군을 추출하는 관측자료군 선정단계; 상기 관측자료군에 기초하여 SVM(Support Vector Machine)기반으로, 과거의 관측데이터를 이용하여 예측을 수행하는 실측기반모델함수 및 과거의 예측데이터를 이용하여 예측을 수행하는 예측기반모델함수를 생성하는 모델함수 생성단계; 상기 실측기반모델함수 및 예측기반모델함수 각각에 대하여, 현재 관측데이터에 따른 상기 지하수위의 변화를 예측하는 예측데이터 산출단계; 및 상기 실측기반모델함수의 예측데이터 및 상기 예측기반모델함수의 예측데이터를 비교하고, 두 예측데이터의 차이값이 이상판단기준값보다 큰 경우, 해당 데이터를 이상관측데이터로 결정하는 이상자료 판별단계를 포함한다.
- [0023] 또한, 상기 모델함수 생성단계는, SVM기반의 시계열 예측을 위한 모델의 구조를 설정하는 구조설정단계; 및 SMO(Sequential Minimal Optimization)알고리즘으로 예측함수를 설정하고 최적화하는 최적화예측함수 생성단계를 포함하고, 상기 최적화예측함수에 의해 상기 모델함수를 생성할 수 있다.
- [0024] 또한, 상기 구조설정단계는, 상기 현재 관측데이터에 따른 입력벡터(Input vector); 상기 관측자료군에서 추출되는 지지벡터(SV, Support vector); 상기 입력벡터 및 지지벡터를 커널트릭(Kernel trick)으로 고차원 변환하는 커널함수(Kernel function); 및 상기 커널함수를 이용하여 예측데이터를 산출하는 출력함수(Output)를 포함하여, 상기 모델의 구조를 설정할 수 있다.
- [0025] 또한, 상기 커널함수는, 가우시안(Gaussian) 형태의 "radial basis kernel" 함수를 포함할 수 있다.
- [0026] 또한, 상기 최적화예측함수 생성단계는, 상기 SVM기반의 예측함수를 상기 SMO알고리즘으로 설정하는 단계; 상기 예측함수를 계산하기 위한 목적함수를 설정하는 단계; 및 상기 목적함수에 최적화 조건을 설정하여 상기 최적화예측함수로 변환하는 단계를 포함할 수 있다.
- [0027] 또한, 상기 최적화예측함수 생성단계는, 상기 목적함수를 계산하기 위하여 라그랑제 승수(Lagrangian multiplier)를 이용한 안장점(Saddle point) 문제를 적용할 수 있다.

발명의 효과

- [0028] 상기와 같은 해결수단에 의해, 본 발명은 예측대상에 대하여 과거의 관측자료들을 이용하여 예측된 결과에 의해 이상관측자료를 실시간으로 정확하게 판별할 수 있는 장점이 있다.
- [0029] 특히, 예측대상과 더불어 예측대상에 영향을 미치는 영향인자에 대한 예측을 병행하여 처리함으로써, 실시간 이

상관측자료의 판별에 대한 신뢰성을 크게 향상시킬 수 있는 장점이 있다.

- [0030] 구체적으로, 지하수위의 자연적인 변화에 영향을 주는 강우 등의 기상 자료와 지하수위 관측 자료를 이용하여, 지하수위 예측 시계열 모델을 만들고 이를 이용함으로써, 지하수위에 대한 이상관측자료의 탐지능력을 크게 향상시킬 수 있는 효과가 있다.
- [0031] 따라서, 수자원 관리 분야, 특히 지하수 관리 분야는 물론, 다양한 기후 변화와 연관된 자연자원의 관리 분야 및 이와 연관 내지 유사한 분야에서 신뢰성 및 경쟁력을 향상시킬 수 있다.

도면의 간단한 설명

- [0032] 도 1은 본 발명에 의한 시계열 예측 모델을 이용한 이상관측자료 탐지 방법을 설명하기 위한 흐름도이다.
- 도 2는 본 발명에 의한 지하수위의 이상관측자료 탐지 방법을 설명하기 위한 흐름도이다.
- 도 3은 본 발명에 의한 이상관측자료 탐지 방법의 개념을 설명하기 위한 도면이다.
- 도 4는 본 발명에 의한 예측모델의 구조를 설명하기 위한 개념도이다.
- 도 5는 도 2의 과정을 적용하기 위한 강우량 및 지하수위의 관측데이터를 나타낸 그래프이다.
- 도 6은 도 2에 나타난 실측기반모델함수에 의해 예측된 결과를 나타낸 그래프이다.
- 도 7은 도 2에 나타난 예측기반모델함수에 의해 예측된 결과를 나타낸 그래프이다.
- 도 8은 도 2에 나타난 실측기반모델함수 및 예측기반모델함수의 차이를 나타낸 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0033] 본 발명에 따른 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법에 대한 예는 다양하게 적용할 수 있으며, 이하에서는 첨부된 도면을 참조하여 가장 바람직한 실시 예에 대해 설명하기로 한다.
- [0034] 도 1은 본 발명에 의한 시계열 예측 모델을 이용한 이상관측자료 탐지 방법을 설명하기 위한 흐름도이다.
- [0035] 도 1을 참조하면, 시계열 예측 모델을 이용한 이상관측자료 탐지 방법은, 예측대상의 과거 관측데이터들 중 시계열 예측모델 설정을 위한 데이터를 추출하는 관측자료군 선정한다(단계 S100). 여기서, 추출되는 관측자료군은, 과거 관측데이터들 중 이상 관측값이 발생하지 않았다고 판단되는 자료군으로 선정함이 바람직하다.
- [0036] 관측자료군이 선정되면, 해당 관측자료군에 기초하여 SVM(Support Vector Machine)기반의 모델함수를 생성한다(단계 S200).
- [0037] 모델함수의 생성 과정을 살펴보면, SVM기반의 시계열 예측을 위한 모델의 기본적인 구조를 설정한 후, SVM기반의 예측함수를 설정하고 최적화하여 최적화예측함수를 생성할 수 있으며, 생성된 최적화예측함수에 의해 상기 모델함수를 생성할 수 있다.
- [0038] 또한, 시계열 예측을 위한 모델의 기본적인 구조는, 현재 관측데이터에 따른 입력벡터(Input vector) 및 관측자료군에서 추출되는 지지벡터(SV, Support vector)와, 입력벡터 및 지지벡터를 고차원 변환하는 차원변환함수, 그리고 차원변환함수를 이용하여 예측데이터를 산출하는 출력함수(Output)를 포함하여 구성될 수 있다.
- [0039] 또한, 최적화예측함수는, SVM기반의 예측함수를 설정하고, 예측함수를 계산하기 위한 목적함수를 설정한 후, 목적함수에 최적화 조건을 설정하여 변환하는 과정을 통해 생성될 수 있다.
- [0040] 모델함수가 생성되면, 생성된 모델함수를 이용하여 현재 관측데이터에 따른 예측대상의 변화를 예측하고 예측데이터를 산출한다(단계 S300). 예를 들어, 예측데이터는 과거의 관측데이터 및 현재의 관측데이터를 적용하여 예측된 실측기반 예측데이터 및 과거의 예측데이터 및 현재의 관측데이터를 적용하여 예측된 예측기반 예측데이터를 포함할 수 있다.
- [0041] 실측기반 예측데이터 및 예측기반 예측데이터를 포함하는 예측데이터가 산출되면, 산출된 예측데이터에 기초하여 이상관측데이터를 확인(판별)한다(단계 S400). 예를 들어, 이상관측데이터의 판별은 실측기반 예측데이터 및 예측기반 예측데이터를 비교하고, 두 예측데이터의 차이가 큰 경우를 이상관측데이터로 판단할 수 있다.

[0042] 이상에서 설명된 도 1의 이상관측자료 탐지 방법은 다양한 자연자원에 대하여 적용이 가능하며, 이하에서는 수자원, 특히 지하수에 적용한 경우에 대하여 살펴보기로 한다.

[0043] 도 2는 본 발명에 의한 지하수위의 이상관측자료 탐지 방법을 설명하기 위한 흐름도이다. 이하에서, 도 3 내지 도 8을 참조하여 도 2의 구체적인 일 실시예를 살펴보기로 한다.

[0044] 도 2 및 도 3를 참조하면, 예측대상인 지하수위 및 지하수위에 영향을 미치는 강우량을 포함하는 영향인자의 과거 관측데이터(t-1 시점 이전의 데이터)들 중 이상관측이 발생하지 않은 자료군(t-k 내지 t-k-τ)을 추출하여 관측자료군으로 선정한다(단계 S110). 예를 들어, 도 5에 나타난 바와 같이 2000년부터 2001년의 관측데이터를 자료군으로 추출할 수 있다.

[0045] 그리고, 선정된 관측자료군에 기초하여 SVM(Support Vector Machine)기반으로, 과거의 관측데이터를 이용하여 예측을 수행하는 실측기반모델함수(모델A) 및 과거의 예측데이터를 이용하여 예측을 수행하는 예측기반모델함수(모델B)를 생성한다(단계 S120).

[0046] 이때, 예측모델은 수학적 식 1 및 수학적 식 2와 같이 나타낼 수 있다.

[0047] (수학적 식 1)

$$[0048] \hat{y}_i = f_A(x_{i-1}, \dots, x_{i-a}, y_{i-1}, \dots, y_{i-b})$$

[0049] (수학적 식2)

$$[0050] \hat{y}_i^* = f_B(x_{i-1}, \dots, x_{i-a}, \hat{y}_{i-1}, \dots, \hat{y}_{i-b})$$

[0051] 여기서, x는 강우 등의 기상요소(설명변수), y는 지하수위(종속변수), fA는 모델A 함수, fB는 모델B 함수, a는 설명변수의 지연시간, b는 종속변수의 지연시간이다.

[0052] 이후, 수학적 식 1의 실측기반모델함수 및 수학적 식 2의 예측기반모델함수 각각에 대하여, 도 5에 나타난 바와 같이 2002년부터 2004년의 자료에 적용하여, 현재 관측데이터(도 3에서 t+1 시점의 데이터)에 따른 지하수위의 변화를 예측한다(단계 S130).

[0053] 예측결과에 따라 도 6에 나타난 바와 같이 실측기반모델함수의 예측데이터 및 도 7에 나타난 예측기반모델함수의 예측데이터를 비교하고, 도 8에 나타난 두 예측데이터의 차이값이 이상판단기준값보다 큰 경우(도 8에서 점선부분), 해당 데이터를 이상관측데이터로 결정한다(단계 S140).

[0054] 이하에서 시계열 모델 개발(생성) 방법에 대해 보다 구체적으로 살펴보기로 한다.

[0055] 본 발명은 지지벡터기계(Support Vector Machine; SVM)알고리즘에 의해 시계열 모델을 생성할 수 있다.

[0056] 먼저, SVM알고리즘을 살펴보면, SVM은 ANN(Artificial Neural Network)등의 일반적인 학습 모델들이 지닌 경험적 위험 최소화 원리(Empirical Risk Minimization; ERM)가 아닌, 구조적인 위험 최소화 원리(Structural Risk Minimization; SRM)에 의한 것으로, 일반적인 학습방법보다 높은 일반화 성능을 보여줄 수 있다.

[0057] 자료의 분류의 측면에서 살펴보면, ANN을 포함한 기존의 방법들은 학습 자료에 의한 분류 오류율을 최소화하는 목적으로 설계되지만, SVM은 분류될 자료 군들 사이에 존재하는 여백(Margin)을 최대화시키는 방향으로 설계됨으로써, 모델의 일반화 능력을 극대화시킬 수 있다.

[0058] SVM의 전체 구조는 도 4에 나타난 바와 같다.

[0059] 도 4를 참조하면, SVM은 ANN과 비슷한 모습을 보이지만 그 원리는 상이하다. ANN의 구조가 입·출력 벡터가 가지는 성분들을 노드로 설정하고 해당 노드에 주어진 자료들이 들어오게 되어있는 구조인 반면, SVM은 훈련에 이용될 입·출력 벡터들 중 모델을 ‘지지’ 하기에 적합한 벡터들이 선정되어 모델을 구성하게 된다. 이와 같이 선정된 벡터들을 지지벡터(Support Vector; SV)라 한다.

[0060] 선정된 SV로 구성된 SVM에 새로운 입력벡터(Input vector)가 들어오게 되면 지지벡터(SV)와의 계산을 통해 새로운 출력값을 예측하게 된다.

[0061] 이 과정에서 커널함수의 전이에 의해 SV와 입력 벡터는 보다 높은 차원으로 이동하게 되고, 이렇게 높여진 차원

내에서 비선형적인 부분이 보다 간편하게 계산될 수 있으며, 이러한 방법을 커널 트릭(Kernel trick)이라 한다.

[0062] 본 발명에서는, 커널함수로 수학식 3과 같이 가우시안(Gaussian) 형태의 "radial basis kernel" 함수를 이용할 수 있다.

[0063] (수학식 3)

$$K(\mathbf{x}_k, \mathbf{x}_l) = \exp\left(-\frac{\|\mathbf{x}_k - \mathbf{x}_l\|^2}{2\sigma^2}\right)$$

[0064]

[0065] 한편, SVM은 뛰어난 학습 및 일반화 능력을 지니고 있지만, SVM의 해는 2차규모의 Quadratic Programming을 풀이 함으로써 얻어지기 때문에, 학습 데이터 수가 늘어남에 따라 학습에 필요한 시간이 기하급수적으로 늘어나는 단점을 가지고 있다.

[0066] 본 발명에서는, 이러한 문제점을 해결하기 위하여, 순차적 최소규모 최적화(Sequential Minimal Optimization; SMO) 알고리즘을 이용함으로써, 계산해야 될 2차규모의 파라미터를 2개로 최소화시켜 이의 해석해를 계속적으로 구해가는 방법으로 SVM의 계산속도 향상시킬 수 있다. 이를 통해 실시간처리가 가능해질 수 있다.

[0067] SVM에 의한 예측함수는 수학식 4와 같이 나타낼 수 있으며, 예측함수를 구하기 위한 목적함수는 수학식 5와 같이 나타낼 수 있다.

[0068] (수학식 4)

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b$$

[0069]

[0070] (수학식 5)

$$\begin{aligned} & \underset{\mathbf{w}, b, \xi, \xi^*}{\text{minimize}} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{k=1}^N (\xi_k + \xi_k^*) \\ & \text{subject to} && \begin{cases} y_k - \mathbf{w}^T \phi(\mathbf{x}_k) - b \leq \varepsilon + \xi_k \\ \mathbf{w}^T \phi(\mathbf{x}_k) + b - y_k \leq \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \geq 0 \end{cases} \quad k = 1, 2, \dots, N \end{aligned}$$

[0071]

[0072] 여기서, \mathbf{w} 는 연결강도 벡터, b 는 편중 정도, ϕ 는 비선형 전이함수이며, 비선형 전이함수 ϕ 는 이하에서 커널함수 K 로 표현될 수 있다.

[0073] ξ 는 특정 손실함수를 통해 훈련 과정에서의 오차에 대해 벌칙을 주는 파라미터, ε 는 오차 허용률, C 는 최적화 문제에 경험 오차의 정도를 제공하는 길항 파라미터이다.

[0074] 수학식 5에서, 라그랑제 승수(Lagrangian multiplier)를 이용한 안장점(Saddle point) 문제를 통해 풀이한다.

[0075] 라그랑제 승수 (α, η)를 이용하여 수학식 5를 수학식 6과 같이 변환할 수 있다.

[0076] (수학식 6)

$$\begin{aligned} L(\mathbf{w}, b, \xi, \xi^*; \alpha, \alpha^*, \eta, \eta^*) = & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{k=1}^N (\xi_k + \xi_k^*) \\ & - \sum_{k=1}^N \alpha_k (\varepsilon + \xi_k - y_k + \mathbf{w}^T \phi(x_k) + b) \\ & - \sum_{k=1}^N \alpha_k^* (\varepsilon + \xi_k^* + y_k - \mathbf{w}^T \phi(x_k) - b) \\ & - \sum_{k=1}^N (\eta_k \xi_k + \eta_k^* \xi_k^*) \end{aligned}$$

[0077]

[0078] 또한, 수학식 6은 수학식 7의 최적화 조건을 만족시킴으로써, 수학식 8과 같이 풀이될 수 있다.

[0079] (수학식 7)

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \mathbf{w} = \sum_{k=1}^N (\alpha_k - \alpha_k^*) \phi(x_k)$$

$$\frac{\partial L}{\partial b} = 0 \quad \sum_{k=1}^N (\alpha_k - \alpha_k^*) = 0$$

$$\frac{\partial L}{\partial \xi_k} = 0 \quad C - \alpha_k - \eta_k = 0$$

$$\frac{\partial L}{\partial \xi_k^*} = 0 \quad C - \alpha_k^* - \eta_k^* = 0$$

[0080]

[0081] (수학식 8)

$$\text{maximize}_{\alpha, \alpha^*} \begin{cases} -\frac{1}{2} \sum_{k,l=1}^N (\alpha_k - \alpha_k^*) (\alpha_l - \alpha_l^*) K(\mathbf{x}_k, \mathbf{x}_l) \\ -\varepsilon \sum_{k=1}^N (\alpha_k + \alpha_k^*) + \sum_{k=1}^N y_k (\alpha_k - \alpha_k^*) \end{cases}$$

$$\text{subject to} \begin{cases} \sum_{k=1}^N (\alpha_k - \alpha_k^*) = 0 \\ 0 \leq \alpha_k, \alpha_k^* \leq C \end{cases}$$

$$\text{to obtain} \quad f(\mathbf{x}) = \sum_{k=1}^n (\alpha_k - \alpha_k^*) K(\mathbf{x}, \mathbf{x}_k) + b$$

[0082]

[0083] SMO에서는 $\beta_i = \alpha_i - \alpha_i^*$ 라 하고 수학식 8를 β 에 대한 2차 함수 문제로 만들어서, β 의 다양한 구간에 대해 최소 값의 해석해를 반복적으로 구함으로써, SV와 그 연결강도를 결정하여 SVM 예측함수를 제공할 수 있다.

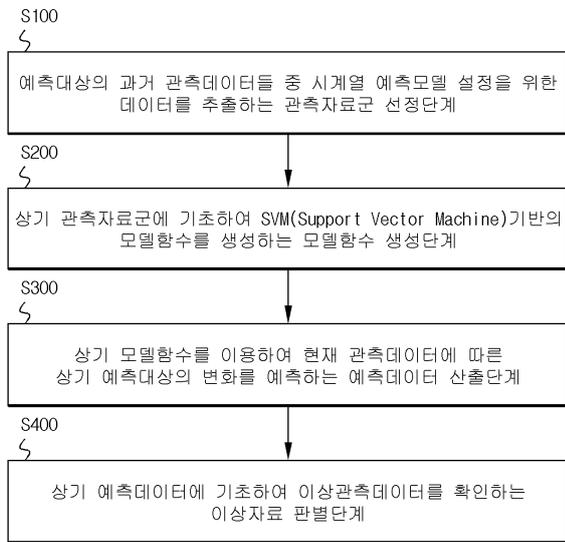
[0084] 따라서, 예측대상인 지하수위에 대하여, 과거의 관측자료들을 이용하여 예측된 결과에 의해 이상관측자료를 실시간으로 정확하게 판별할 수 있다.

[0085] 이상에서 본 발명에 의한 시계열 예측 모델을 이용한 이상관측자료 탐지 방법 및 지하수위의 이상관측자료 탐지 방법에 대하여 설명하였다. 이러한 본 발명의 기술적 구성은 본 발명이 속하는 기술분야의 당업자가 본 발명의 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다.

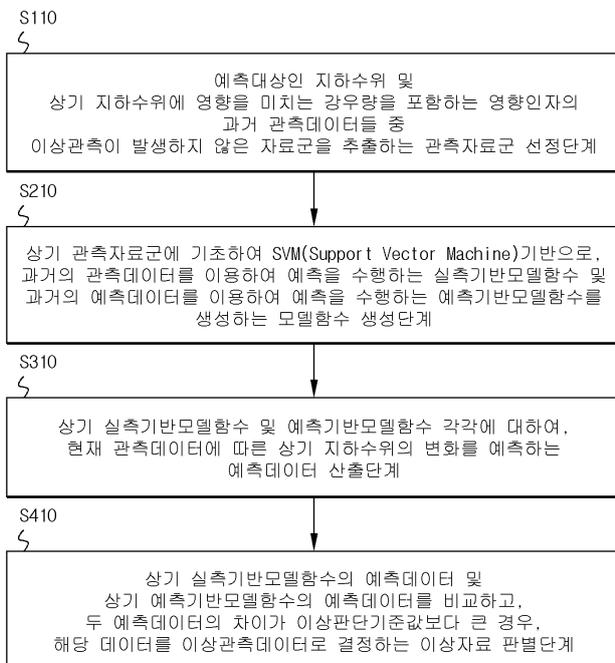
[0086] 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며, 한정적인 것이 아닌 것으로서 이해되어야 하고, 본 발명의 범위는 전술한 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지는 것이므로, 특허청구범위의 의미 및 범위 그리고 그 등가개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

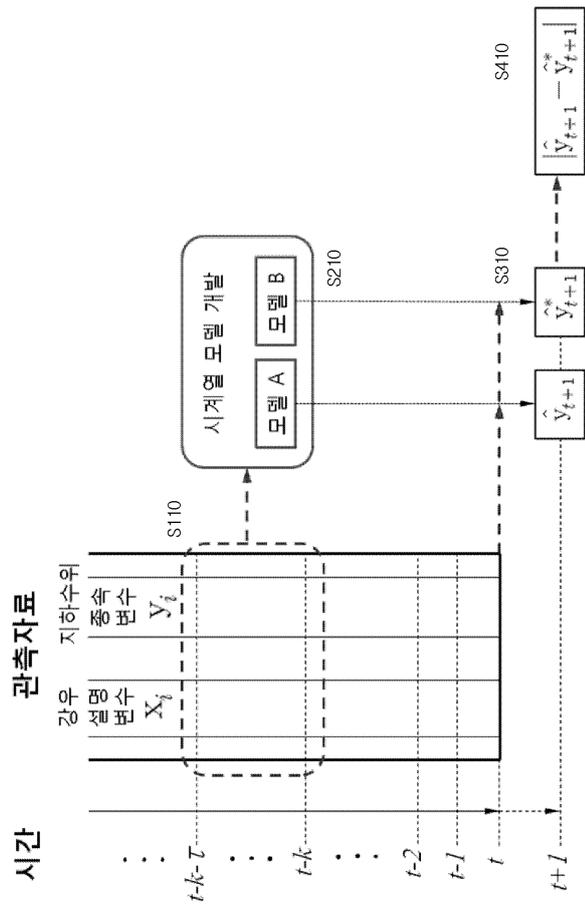
도면1



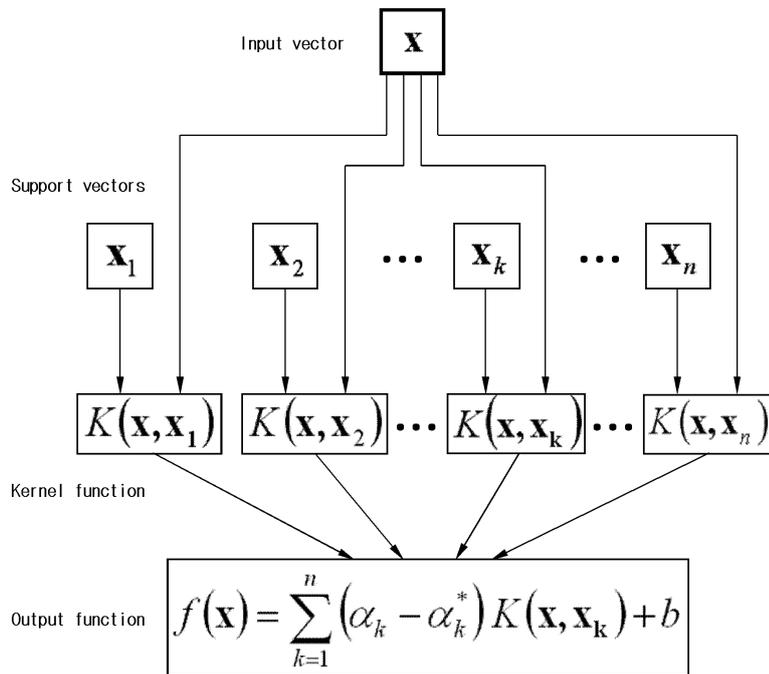
도면2



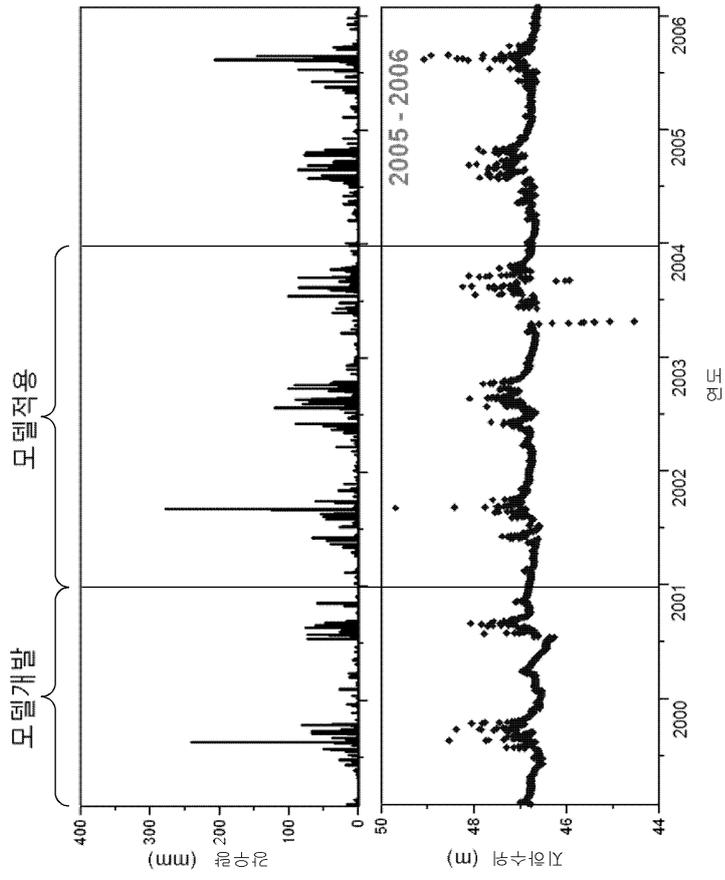
도면3



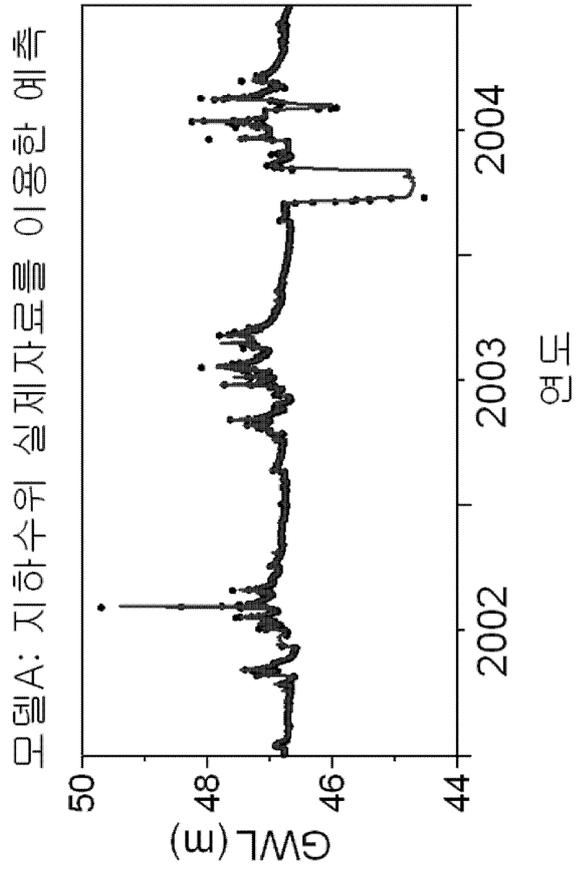
도면4



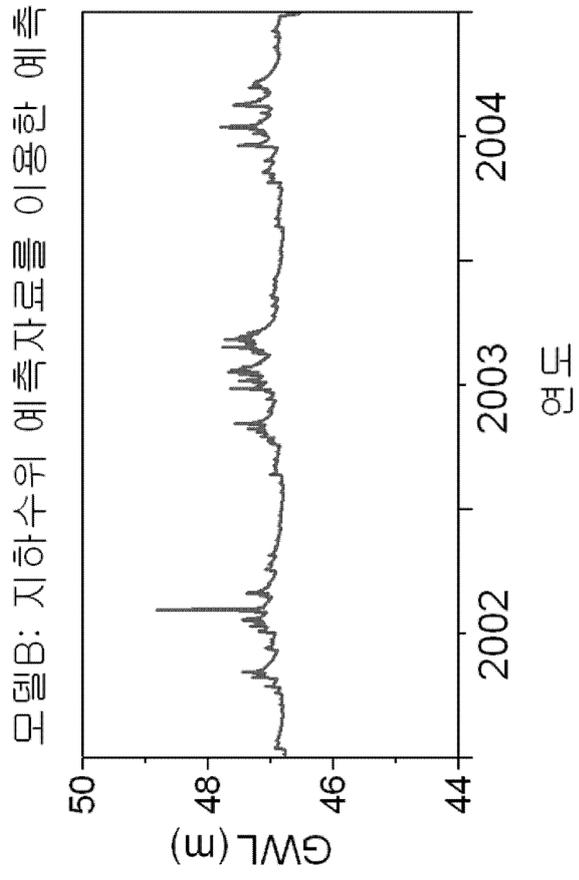
도면5



도면6



도면7



도면8

