



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2007년12월11일
(11) 등록번호 10-0783586
(24) 등록일자 2007년12월03일

(51) Int. Cl.

G06F 9/50 (2006.01) G06F 9/38 (2006.01)

G06F 9/46 (2006.01)

(21) 출원번호 10-2007-0038823

(22) 출원일자 2007년04월20일

심사청구일자 2007년04월20일

(56) 선행기술조사문헌

자원 신뢰성 측정을 통한 효율적인 그리드 자원 스케줄링 모델, 한국시물레이션학회 논문지 제15권 제4호, 2006. 12

US06675164 B2, 2004.01.06

전체 청구항 수 : 총 10 항

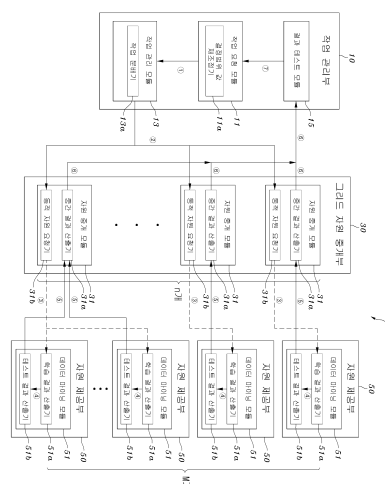
심사관 : 윤혜숙

(54) 그리드 기반 혼합형 데이터 마이닝 장치 및 방법

(57) 요약

본 발명은 그리드 기반 혼합형 데이터 마이닝 장치 및 방법에 관한 것으로서, 대용량의 데이터 마이닝을 위해 지리적으로 분산된 컴퓨팅 자원을 이용함으로써, 데이터 및 그리드 자원의 활용도를 높일 수 있으며, 대용량의 데이터 마이닝 작업에 대하여 서로 다른 그리드 자원의 연산 능력에 적합하도록 동적으로 작업을 분배하고, 대용량의 작업 처리 및 실행에서 발생가능한 병목 현상을 감소시킴으로써, 처리 시간을 단축시킬 수 있으며, 신뢰할 만한 데이터 패턴 인식률에 근접할 때까지 데이터 패턴의 인식 여부에 대한 결정범위값을 재조정하고, 이를 갱신하여 자원 제공부의 테스트 비용을 단축시킬 수 있는 그리드 기반 혼합형 데이터 마이닝 장치 및 방법을 제공하기 위한 것으로서, 그 기술적 구성은 대용량의 데이터 마이닝 작업을 위한 그리드 컴퓨팅의 자원을 요청하는 작업 요청 모듈과, 상기 요청에 따른 작업 수행에서 동일한 계산 능력을 갖도록 작업 분배기로 대용량의 데이터 마이닝 작업을 분배하는 작업 관리 모듈과, 작업이 완료된 데이터의 데이터 마이닝 인식률을 테스트하는 결과 테스트 모듈을 포함하는 작업 관리부; 상기 작업 관리 모듈의 분배를 할당받아 동적으로 자원을 요청하고, 작업이 완료된 데이터를 통합하여 산출된 결과를 상기 결과 테스트 모듈로 전송하는 다수의 자원 중개 모듈을 포함하는 그리드 자원 중개부; 및 상기 동적 자원 요청을 수신하여 데이터 마이닝을 위한 학습 결과 및 테스트 결과를 산출하여 상기 자원 중개 모듈로 전송하는 다수의 데이터 마이닝 모듈을 포함하는 자원 제공부; 를 포함한다.

대표도 - 도1



특허청구의 범위

청구항 1

대용량의 데이터 마이닝 작업을 위한 그리드 컴퓨팅의 자원을 요청하는 작업 요청 모듈과, 상기 요청에 따른 작업 수행에서 동일한 계산 능력을 갖도록 작업 분배기로 대용량의 데이터 마이닝 작업을 분배하는 작업 관리 모듈과, 작업이 완료된 데이터의 데이터 패턴 인식률을 테스트하는 결과 테스트 모듈을 포함하는 작업 관리부;

상기 작업 관리 모듈의 분배를 할당받아 동적으로 자원을 요청하고, 작업이 완료된 데이터를 통합하여 산출된 결과를 상기 결과 테스트 모듈로 전송하는 다수의 자원 중개 모듈을 포함하는 그리드 자원 중개부; 및

상기 동적 자원 요청을 수신하여 데이터 마이닝을 위한 학습 결과 및 테스트 결과를 산출하여 상기 자원 중개 모듈로 전송하는 다수의 데이터 마이닝 모듈을 포함하는 자원 제공부;

를 포함하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 2

제1항에 있어서,

상기 작업 요청 모듈은

상기 결과 테스트 모듈에서 신뢰할 수 있는 데이터 패턴 인식률에 근접할 때까지 데이터 패턴의 인식 여부에 대한 결정 범위를 조정하는 결정 범위값 제조정기;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 3

제1항에 있어서,

상기 작업 관리 모듈은

상기 자원 중개 모듈의 수 및 분배할 작업 수를 초기화하여 각 데이터 마이닝에서 요구하는 자원량을 기반으로 작업을 정렬하고, 자원 중개 모듈의 수가 전체 작업수 미만인 경우에는 작업 중개 모듈에 순차적으로 분배하며, 자원 중개 모듈의 수가 전체 작업수를 초과하는 경우에는 상기 순차적 분배 및 가중치를 이용하여 역순으로 분배하되, 모든 작업이 할당될 때까지 반복하는 작업 분배기;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 4

제1항에 있어서,

상기 자원 중개 모듈은

상기 작업 관리 모듈의 분배를 할당받아 동적으로 자원을 요청하는 동적 자원 요청기; 및

작업이 완료된 데이터의 검사 결과를 산출하도록 통합시켜 상기 결과 테스트 모듈로 전송하는 중간 결과 산출기;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 5

제4항에 있어서,

상기 동적 자원 요청기는 할당된 작업량이 상기 자원 제공부의 작업 가능량을 초과하면, 할당된 작업량을 수용 가능한 자원 제공부가 검색될 때까지 순차적으로 모두 검색하되, 수용가능한 자원 제공부가 검색되지 않으면 데이터 패턴 인식 여부에 대한 결정 범위값을 조정하여 재검색하고, 수용가능한 자원 제공부가 검색되면 해당 자원 제공부와 연결하여 해당 자원을 할당받는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 6

제1항에 있어서,

상기 자원 제공부는

상기 동적 자원 요청기의 요청을 수신한 각 데이터 마이닝 모듈에서 데이터 마이닝을 위한 학습 결과를 산출하는 학습 결과 산출기; 및

산출된 학습 결과를 검사하여 상기 중간 결과 산출기로 전송하는 테스트 결과 산출기;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 7

제1항에 있어서,

상기 자원 제공부는 학습 결과 및 테스트 결과를 역전과 방법의 가중치를 이용하여 산출하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 장치.

청구항 8

대용량의 데이터 마이닝 작업을 위해 분산된 그리드 컴퓨팅 자원을 작업 관리 모듈에 요청하여 상기 요청에 따른 작업이 동일한 계산량을 갖도록 작업 분배기에서 분배하는 단계;

상기 분배를 할당받은 그리드 자원 중개부 내 다수의 자원 중개 모듈은 동적 자원 요청기를 이용하여 각기 다른 자원을 보유한 자원 제공부에 자원 사용을 동적으로 요청하는 단계;

상기 자원 사용 요청을 받은 자원 제공부 내 다수의 데이터 마이닝 모듈은 역전과 방법의 가중치를 이용하여 학습 결과 및 이에 따른 테스트 결과를 산출하여 상기 각 결과를 자원 중개 모듈 내 중간 결과 산출기에서 통합하는 단계;

상기 통합된 결과를 작업 관리부 내 결과 테스트 모듈에서 분석 및 테스트하여 신뢰할 수 있는 데이터 패턴 인식률에 도달할 때까지 상기 단계를 반복적으로 구동하되, 결정 범위값 재조정기로 데이터 패턴의 인식 여부에 대한 결정 범위값을 조정하여 갱신하는 단계;

를 포함하는 그리드 기반 혼합형 데이터 마이닝 방법.

청구항 9

제8항에 있어서,

상기 작업 분배기의 작업 분배 방법은

상기 작업 관리 모듈에서 그리드 자원 중개부 내 자원 중개 모듈의 수 및 작업의 수를 초기화하는 단계;

각 데이터 마이닝 작업에서 요구하는 자원량을 기반으로 작업을 정렬하는 단계;

상기 자원 중개 모듈의 수가 전체 작업수 미만인 경우, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 모든 작업이 자원 중개 모듈에 할당될 때까지 분배하는 단계;

상기 자원 중개 모듈의 수가 전체 작업의 수를 초과하는 경우, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 분배 및 마지막 자원 중개 모듈로부터 가중치를 이용하여 역순으로 분배하는 과정을 모든 작업이 자원 중개 모듈에 할당될 때까지 반복하는 단계;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 방법.

청구항 10

제8항에 있어서,

상기 동적 자원 요청기의 동적 자원 요청 방법은

각 자원 중개 모듈의 동적 자원 요청기에서 모든 자원 제공부의 상태를 감시 및 갱신하는 단계;

할당된 작업량이 상기 자원 제공부의 작업 가능량 범위 내이면, 해당 자원 제공부와 연결을 설정하여 해당 자원

을 할당받는 단계;

할당된 작업량이 상기 자원 제공부의 작업 가능량을 초과하면, 할당된 작업량을 수용가능한 자원 제공부가 검색 될 때까지 순차적으로 모두 검색하는 단계;

수용가능한 자원 제공부가 검색되지 않으면 데이터 패턴의 인식 여부에 대한 결정범위값을 조정하여 재검색하되, 수용가능한 자원 제공부가 검색되면 해당 자원 제공부와 연결하여 해당 자원을 할당받는 단계;

를 포함하는 것을 특징으로 하는 그리드 기반 혼합형 데이터 마이닝 방법.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- <14> 본 발명은 그리드 기반 혼합형 데이터 마이닝 장치 및 방법에 관한 것으로, 더욱 상세하게는 대용량의 데이터 마이닝을 위해 지리적으로 분산된 컴퓨팅 자원을 이용함으로써, 데이터 및 그리드 자원의 활용도를 높일 수 있으며, 대용량의 데이터 마이닝 작업에 대하여 서로 다른 그리드 자원의 연산 능력에 적합하도록 동적으로 작업을 분배하고, 대용량의 작업 처리 및 실행에서 발생가능한 병목 현상을 감소시킴으로써, 처리 시간을 단축시킬 수 있으며, 신뢰할만한 데이터 패턴 인식률에 근접할 때까지 데이터 패턴의 인식에 대한 결정 범위 값을 재조정하고, 이를 갱신하여 자원 제공부의 테스트 비용을 단축시킬 수 있는 그리드 기반 혼합형 데이터 마이닝 장치 및 방법에 관한 것이다.
- <15> 일반적으로, 데이터 마이닝(Data Mining)은 대량의 데이터에 존재하는 데이터 간의 상관 관계를 추출하는 방법으로, 대량의 데이터 베이스 내의 데이터 속에서 유도된 새로운 데이터 모델을 발견하여 미래에 실행 가능한 정보를 추출하고, 이를 의사 결정에 이용하는 과정을 일컫는다.
- <16> 즉, 데이터에 존재하는 패턴 및 관계에 따른 정보를 발견하는 것으로, 통계 분석 및 모델링 기법을 적용하여 유용한 패턴 및 관계를 찾아내는데, 대량의 데이터 집합을 분석하여 특징 정보를 찾아내는 것을 데이터 마이닝이라 한다.
- <17> 여기서, 대용량의 데이터를 처리하기 위하여 분석 대상이 되는 데이터 집합의 범위를 분석자가 정의하고 범위에서 벗어나는 데이터를 소멸시켜, 현재 시점을 기준으로 가치있는 데이터의 집합에 대한 시간 처리를 실행한다.
- <18> 그러나, 대용량 데이터 마이닝은 가치있는 데이터의 집합이 대용량일 경우, 메모리의 공간은 일정한 반면, 가치있는 데이터의 패턴 및 관계를 찾아내는데 요구되는 작업 메모리 공간 및 처리 시간이 증가하고, 이에 따라 데이터 분석 및 처리에 요구되는 비용이 증가되는 등의 문제점이 있었다.

발명이 이루고자 하는 기술적 과제

- <19> 본 발명은 상기한 문제점을 해결하기 위하여 안출한 것으로, 대용량의 데이터 마이닝을 위해 지리적으로 분산된 컴퓨팅 자원을 이용함으로써, 데이터 및 그리드 자원의 활용도를 높일 수 있으며, 대용량의 데이터 마이닝 작업에 대하여 서로 다른 그리드 자원의 연산 능력에 적합하도록 동적으로 작업을 분배하고, 대용량의 작업 처리 및 실행에서 발생가능한 병목 현상을 감소시킴으로써, 처리 시간을 단축시킬 수 있으며, 신뢰할만한 데이터 패턴 인식률에 근접할 때까지 데이터 패턴의 인식 여부에 대한 결정 범위 값을 재조정하고, 이를 갱신하여 자원 제공부의 테스트 비용을 단축시킬 수 있는 그리드 기반 혼합형 데이터 마이닝 장치 및 방법을 제공하는 것을 목적으로 한다.

발명의 구성 및 작용

- <20> 상기한 바와 같은 목적을 달성하기 위하여 본 발명은 대용량의 데이터 마이닝 작업을 위한 그리드 컴퓨팅의 자원을 요청하는 작업 요청 모듈과, 상기 요청에 따른 작업 수행에서 동일한 계산 능력을 갖도록 작업 분배기로 대용량의 데이터 마이닝 작업을 분배하는 작업 관리 모듈과, 작업이 완료된 데이터의 데이터 패턴 인식률을 테스트하는 결과 테스트 모듈을 포함하는 작업 관리부; 상기 작업 관리 모듈의 분배를 할당받아 동적으로 자원을 요청하고, 작업이 완료된 데이터를 통합하여 산출된 결과를 상기 결과 테스트 모듈로 전송하는 다수의 자원 중

개 모듈을 포함하는 그리드 자원 중개부; 및 상기 동적 자원 요청을 수신하여 데이터 마이닝을 위한 학습 결과 및 테스트 결과를 산출하여 상기 자원 중개 모듈로 전송하는 다수의 데이터 마이닝 모듈을 포함하는 자원 제공부; 를 포함한다.

- <21> 그리고, 상기 작업 요청 모듈은 상기 결과 테스트 모듈에서 신뢰할 수 있는 데이터 패턴 인식물에 근접할 때까지 데이터 패턴의 인식 여부에 대한 결정 범위를 조정하는 결정 범위값 재조정기; 를 포함하는 것을 특징으로 한다.
- <22> 여기서, 상기 작업 관리 모듈은 상기 자원 중개 모듈의 수 및 분배할 작업 수를 초기화하여 각 데이터 마이닝에서 요구하는 자원량을 기반으로 작업을 정렬하고, 자원 중개 모듈의 수가 전체 작업수 미만인 경우에는 작업 중개 모듈에 순차적으로 분배하며, 자원 중개 모듈의 수가 전체 작업수를 초과하는 경우에는 상기 순차적 분배 및 가중치를 이용하여 역순으로 분배하되, 모든 작업이 할당될 때까지 반복하는 작업 분배기; 를 포함하는 것을 특징으로 한다.
- <23> 그리고, 상기 동적 자원 요청기는 할당된 작업량이 상기 자원 제공부의 작업 가능량을 초과하면, 할당된 작업량을 수용가능한 자원 제공부가 검색될 때까지 순차적으로 모두 검색하되, 수용가능한 자원 제공부가 검색되지 않으면 데이터 패턴의 인식 여부에 대한 결정 범위값을 조정하여 재검색하고, 수용가능한 자원 제공부가 검색되면 해당 자원 제공부와 연결하여 해당 자원을 할당받는 것을 특징으로 한다.
- <24> 또한, 상기 자원 중개 모듈은 상기 작업 관리 모듈의 분배를 할당받아 동적으로 자원을 요청하는 동적 자원 요청기; 및 작업이 완료된 데이터의 검사 결과를 산출하도록 통합시켜 상기 결과 테스트 모듈로 전송하는 중간 결과 산출기; 를 포함하는 것을 특징으로 한다.
- <25> 더불어, 상기 자원 제공부는 상기 동적 자원 요청기의 요청을 수신한 각 데이터 마이닝 모듈에서 데이터 마이닝을 위한 학습 결과를 산출하는 학습 결과 산출기; 및 산출된 학습 결과를 검사하여 상기 중간 결과 산출기로 전송하는 테스트 결과 산출기; 를 포함하는 것을 특징으로 한다.
- <26> 여기서, 상기 자원 제공부는 학습 결과 및 테스트 결과를 역전과 방법의 가중치를 이용하여 산출하는 것을 특징으로 한다.
- <27> 한편, 대용량의 데이터 마이닝 작업을 위해 분산된 그리드 컴퓨팅 자원을 작업 관리 모듈에 요청하여 상기 요청에 따른 작업이 동일한 계산량을 갖도록 작업 분배기에서 분배하는 단계; 상기 분배를 할당받은 그리드 자원 중개부 내 다수의 자원 중개 모듈은 동적 자원 요청기를 이용하여 각기 다른 자원을 보유한 자원 제공부에 자원 사용을 동적으로 요청하는 단계; 상기 자원 사용 요청을 받은 자원 제공부 내 다수의 데이터 마이닝 모듈은 역전과 방법의 가중치를 이용하여 학습 결과 및 이에 따른 테스트 결과를 산출하여 상기 각 결과를 자원 중개 모듈 내 중간 결과 산출기에서 통합하는 단계; 상기 통합된 결과를 작업 관리부 내 결과 테스트 모듈에서 분석 및 테스트하여 신뢰할 수 있는 데이터 패턴 인식물에 도달할 때까지 상기 단계를 반복적으로 구동하되, 데이터 패턴의 인식 여부에 대한 결정 범위값 재조정기로 결정 범위값을 조정하여 갱신하는 단계; 를 포함한다.
- <28> 여기서, 상기 작업 분배기의 작업 분배 방법은 상기 작업 관리 모듈에서 그리드 자원 중개부 내 자원 중개 모듈의 수 및 작업의 수를 초기화하는 단계; 각 데이터 마이닝 작업에서 요구하는 자원량을 기반으로 작업을 정렬하는 단계; 상기 자원 중개 모듈의 수가 전체 작업수 미만인 경우, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 모든 작업이 자원 중개 모듈에 할당될 때까지 분배하는 단계; 상기 자원 중개 모듈의 수가 전체 작업의 수를 초과하는 경우, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 분배 및 마지막 자원 중개 모듈로부터 가중치를 이용하여 역순으로 분배하는 과정을 모든 작업이 자원 중개 모듈에 할당될 때까지 반복하는 단계; 를 포함하는 것을 특징으로 한다.
- <29> 그리고, 상기 동적 자원 요청기의 동적 자원 요청 방법은 각 자원 중개 모듈의 동적 자원 요청기에서 모든 자원 제공부의 상태를 감시 및 갱신하는 단계; 할당된 작업량이 상기 자원 제공부의 작업 가능량 범위 내이면, 해당 자원 제공부와 연결을 설정하여 해당 자원을 할당받는 단계; 할당된 작업량이 상기 자원 제공부의 작업 가능량을 초과하면, 할당된 작업량을 수용가능한 자원 제공부가 검색될 때까지 순차적으로 모두 검색하는 단계; 수용가능한 자원 제공부가 검색되지 않으면 데이터 패턴의 인식 여부에 대한 결정범위값을 조정하여 재검색하되, 수용가능한 자원 제공부가 검색되면 해당 자원 제공부와 연결하여 해당 자원을 할당받는 단계; 를 포함하는 것을 특징으로 한다.

- <30> 이하, 본 발명에 따른 실시예를 첨부된 예시도면을 참고로 하여 상세하게 설명한다.
- <31> 도 1은 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 장치를 개략적으로 도시한 블록도이다. 도면에서 도시하고 있는 바와 같이, 본 발명에 의한 그리드 기반 혼합형 데이터 마이닝 장치(1)는 작업 관리부(10)와 그리드 자원 중개부(30)와 자원 제공부(50)를 포함하여 이루어진다.
- <32> 상기 작업 관리부(10)는 대용량의 데이터 마이닝 작업을 처리함에 있어서, 다수의 그리드 컴퓨팅 자원이 동일한 계산 능력을 가지도록 상기 그리드 자원 중개부(30)에 작업을 분배하고, 데이터 마이닝 작업이 완료된 결과를 테스트하여 최적의 데이터 마이닝 인식률을 가질 수 있도록 결정범위 값을 조정하는 기능을 수행한다. 여기서 결정범위는 데이터 패턴의 인식 여부를 결정하는 범위에 대한 것을 의미하고 결정범위 값은 그 결정범위에 대한 값을 의미한다.
- <33> 여기서, 상기의 기능을 수행하기 위하여 상기 작업 관리부(10)는 작업 요청 모듈(11)과 작업 관리 모듈(13)과 결과 테스트 모듈(15)을 포함하여 이루어진다.
- <34> 그리고, 상기 작업 요청 모듈(11)은 대용량의 데이터 마이닝 작업을 위한 그리드 기반 자원들을 상기 그리드 자원 중개부(30)를 통하여 상기 자원 제공부(50)에 요청하도록, 작업 요청에 대한 요청 데이터를 상기 작업 관리 모듈(13)을 통해 상기 작업 관리 모듈(13)의 데이터와 함께 상기 그리드 자원 중개부(30)로 전달한다.
- <35> 또한, 상기 작업 요청 모듈(11)은 대용량의 데이터 마이닝 작업이 완료된 데이터의 결과를 테스트하는 결정범위 값 재조정기(11a)를 구비하는데, 작업이 완료된 데이터를 분석 및 테스트하여 신뢰할 수 있는 데이터 마이닝 인식률에 근접할 때까지 결정범위 값 재조정기(11a)에서 결정범위 값을 신뢰할 수 있는 데이터 마이닝 인식률에 근접하도록 조정하고, 결정범위 값 재조정에 대한 수식은 하기 수학식 1과 같다.

수학식 1

- <36>
$$DR = DR - \gamma \quad (IF \ CDMRR < RDMRR - \alpha)$$
- <37>
$$DR = DR + \gamma \quad (IF \ CDMRR > RDMRR + \alpha)$$
- <38> 여기서, DR(Decision Range)은 0과 1사이의 결정범위 값을 나타내며, γ 는 0과 1사이의 결정범위 변동 계수(Decision Range Floating Coefficient)를 나타내고, α 는 0과 1사이의 데이터 마이닝 인식률 한계 상수를 나타내며, CDMRR(Current Data Mining Recognize Rate)는 작업 관리부(10)로 전송된 데이터 마이닝의 결과를 결과 테스트 모듈(15)에서 검사한 결과로 현재의 데이터 마이닝 인식률을 나타내고, RDMRR(Reliable Data Mining Recognize Rate)은 데이터 마이닝 작업 시 사용자의 상황에 따라 변경 가능한 기 설정된 신뢰할만한 데이터 마이닝 인식률, 즉 기준 마이닝 인식률을 나타낸다. 여기서 데이터 마이닝 인식률은 데이터 마이닝에 의한 데이터의 패턴 또는 관계에 대한 인식률을 의미하는 것으로 데이터 패턴 인식률이라고도 한다.
- <39> 그리고, 상기 수학식 1의 첫번째 식을 설명하면, 현재의 데이터 마이닝 작업에 대한 데이터 패턴 인식률(CDMRR)이 기준 데이터 패턴 인식률 범위(RDMRR)의 하한값($RDMRR - \alpha$) 미만일 경우에는 결정범위 값 재조정기(11a)를 통하여 결정범위 값(DR)은 상기 작업 관리부(10)의 결과 테스트 모듈(15)을 통하여 산출된 결정범위 변동계수(γ)를 감산한다($DR = DR - \gamma$).
- <40> 더불어, 상기 수학식 1의 두번째 식을 설명하면, 현재의 데이터 마이닝 작업에 대한 데이터 패턴 인식률(CDMRR)이 기준 데이터 패턴 인식률 범위(RDMRR)의 상한값($RDMRR + \alpha$) 초과일 경우에는 결정범위 값 재조정기(11a)를 통하여 결정범위 값(DR)은 상기 작업 관리부(10)의 결과 테스트 모듈(15)을 통하여 산출된 결정범위 변동계수(γ)를 합산한다($DR = DR + \gamma$).
- <41> 이때, 상기 α 는 기준 데이터 패턴 인식률의 범위를 나타내기 위하여 이용된다.
- <42> 또한, 현재 데이터 패턴 인식률(CDMRR)이 상기 작업 요청 모듈(11)의 결정범위 값 재조정기(11a)에서 신뢰할만한 데이터 패턴 인식률인 기준 데이터 패턴 인식률의 범위($RDMRR - \alpha \leq CDMRR \leq RDMRR + \alpha$) 내에 존재하거나 또는 상기 결정범위 값 재조정기(11a)로 기준 데이터 패턴 인식률에 도달했을 경우에는 데이터 패턴 인식률이 신뢰할만하다고 판단하여, 현재 데이터에 대한 데이터 마이닝을 종료한다.
- <43> 한편, 작업 관리 모듈(13)은 상기 작업 요청 모듈(11)의 작업 요청을 메세지를 전달받아 상기 그리드 자원 중개부(30)에 각각의 하위 그리드 자원이 동일한 계산 능력을 가질 수 있도록 분배한다.
- <44> 이를 위하여, 상기 작업 관리 모듈(13)은 상기 작업 요청 모듈(11)로부터 작업 요청 메세지를 전달받아, 그리드

자원 중개부(30)로 작업 요청을 전송하는 기능 외에도, 각각의 하위 그리드 자원이 동일한 계산 능력을 가질 수 있도록 분배하는 작업 분배기(13a)를 포함한다.

<45> 여기서, 상기 작업 분배기(13a)는 상기 그리드 자원 중개부(30) 내에 다수의 하위 그리드 자원을 중개하도록 다수개 구비되는 다수의 자원 중개 모듈(31)에게 전달하여 동일한 계산 능력을 가질 수 있도록 작업을 분산시키는 기능을 수행하는데, 최초 데이터 마이닝 작업 요청 시 동일하나 연산 능력에 해당하는 자원을 이용할 수 있도록 데이터 마이닝의 작업량에 따라 요구되는 자원량을 기준으로 작업을 분배한다.

<46> 다시 말하면, 대용량의 데이터 마이닝 작업의 요청이 작업 요청 모듈(11)로부터 전달되면, 작업 관리 모듈(13)은 그리드 자원 중개부(30)에 포함된 자원 중개 모듈(31)의 수 및 분배할 작업의 수를 초기화한다.

<47> 그리고, 작업 관리 모듈(13)의 작업 분배기(13a)가 각 작업에서 필요로 하는 자원량을 기초로 작업을 정렬하는데, 상기 초기화된 자원 중개 모듈(31)의 수 및 분배할 작업의 수를 비교하여 이에 따라 할당을 실시한다.

<48> 즉, 자원 중개 모듈(31)의 수가 분배할 작업의 수 미만이면, 하기 수학식 2에 의하여 분배된다.

수학식 2

<49> $A_i = i;$ for $i = 1, 2, \dots, TA$

<50> 여기서, i (Index)는 할당할 작업의 인덱스를 나타내며, A_i 는 i 번째 자원 중개 모듈(31)을 나타내고, TA (# of Total Agent)는 전체 자원 중개 모듈의 수를 나타낸다.

<51> 다시 말하면, 상기 작업 분배기(13a)에서는 상기 작업 요청 모듈(11)이 요청한 작업에 요구되는 자원량에 기초하여 정렬된 작업들은 그리드 자원 중개부(30) 내 다수개의 자원 중개 모듈(31)의 수보다 전체 데이터 마이닝 작업의 수 미만이면, 정렬된 데이터 마이닝 작업을 자원 중개 모듈(31)에 모든 작업이 분배될 때까지 순차적으로 분배하는 것이다.

<52> 반대로, 자원 중개 모듈(31)의 수가 분배할 작업의 수 미만이면, 하기 수학식 3에 의하여 분배된다.

수학식 3

<53> $A_i = i,$ (if $A_i \neq (TA + 1),$ for $i = 1, 2, \dots, TA$

<54> $w = (w * 2),$ (if $A_i = (TA + 1),$ for $i = 1, 2, \dots, TA$

<55> $A_i = A_i + w,$ (if $A_i = (TA + 1),$ for $i = 1, 2, \dots, TA$

<56> 여기서, i (Index)는 할당할 작업의 인덱스를 나타내며, A_i 는 i 번째 자원 중개 모듈(31)을 나타내고, TA (# of Total Agent)는 전체 자원 중개 모듈의 수를 나타내며, w (Weight Value)는 가중치를 나타낸다.

<57> 그리고, 상기 수학식 2에서와 마찬가지로 상기 작업 분배기(13a)에서는 상기 작업 요청 모듈(11)이 요청한 작업에 요구되는 자원량에 기초하여 정렬된 작업들은 자원 중개 모듈(31)에 순차적으로 분배하고, 마지막 자원 중개 모듈(31, A_{TA})로부터 가중치를 이용하여 역순으로 자원 중개 모듈(31)의 수(TA)만큼 실시하고, 모든 작업이 분배될 때까지 순차적 분배 및 가중치를 이용한 역순 분배의 두가지 분배 방식을 반복한다.

<58> 또한, 상기 그리드 자원 중개부(30)는 상기 작업 관리부(10)로부터 대용량의 데이터 마이닝 작업을 동일한 계산량을 가지도록 할당받아, 그룹별로 할당된 작업을 다시 하위 작업으로 나누어 그룹 내에 포함된 각 자원 제공부(50)에게 동적으로 대용량 데이터 마이닝 작업에 대한 자원을 요청하고, 작업이 완료된 데이터에 대하여 각 하위 그룹인 각 자원 제공부(50)의 완료 데이터 결과를 전달받아 중간 결과를 통합하여 산출한다.

<59> 이를 위하여, 그리드 자원 중개부(30)는 다수개의 자원 중개 모듈(31)로 이루어지는데, 상기 자원 중개 모듈(31)은 중간 결과 산출기(31a)와 동적 자원 요청기(31b)로 이루어진다.

<60> 여기서, 상기 다수개의 자원 중개 모듈(31)은 상기 작업 관리 모듈(13)의 작업 분배기(13a)의 할당에 따라 전달받은 작업을 동적으로 요청하는 동적 자원 요청기(31b)를 구비하는데, 이는 전달받은 동적인 자원 사용 메시지를 자원 제공부(50)로 전송한다.

- <61> 그리고, 상기 동적인 자원 사용 메시지에 의하여, 서로 다른 연산 능력을 가지는 각 자원 제공부(50)에서 필요한 자원을 제공하여 데이터 마이닝 작업을 수행하는데, 상기 데이터 마이닝 작업이 완료된 각각의 하위 자원 제공부(50)의 결과는 자원 제공부(50)보다 상위 노드인 그리드 자원 중개부(30)의 중간 결과 산출기(31a)로 통합되어야 하며, 이와 같이 통합된 중간 결과를 상기 중간 결과 산출기(31a)는 작업 관리부(10)의 결과 테스트 모듈(15)로 전송한다.
- <62> 또한, 상기 자원 제공부(50)는 다수개 구비되어 상기 그리드 자원 중개부(30)의 하위 노드를 이루는데, 다수개의 자원 제공부(50)는 각각 서로 다른 연산 능력을 가지고 있으며, 그리드 자원 중개부(30)로부터 수신한 자원 사용에 대한 요청에 따라 데이터 마이닝 작업을 처리하고, 그 결과를 상기 다시 그리드 자원 중개부(30)로 전달한다.
- <63> 이를 위하여, 상기 자원 제공부(50)는 각각 데이터 마이닝을 수행할 수 있도록 데이터 마이닝 모듈(51)을 포함하며, 이는 학습 결과 산출기(51a)와 테스트 결과 산출기(51b)를 포함하여 이루어진다.
- <64> 여기서, 데이터 마이닝 모듈(51)에서는 자원 사용을 위한 데이터 마이닝 작업을 처리하고, 데이터 마이닝을 위한 학습(Learning) 결과는 학습 결과 산출기(51a)를 통하여 산출되며, 테스트 결과는 테스트 결과 산출기(51b)를 통해 산출되는데, 학습 결과 산출기(51a) 및 테스트 결과 산출기(51b)는 신경망 연구 모델 중 하나인 역전파 방법(Method of Back - Propagation)을 이용한다.
- <65> 또한, 상기 역전파 방법을 통하여 안정된 결과를 도출할 수 있는 가중치로 학습 및 테스트 결과를 학습 결과 산출기(51a) 및 테스트 결과 산출기(51b)로 산출하여, 이를 상기 그리드 자원 중개부(30)의 동적 자원 요청기(31b)로 출력한다.
- <66> 도 2는 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법을 개략적으로 도시한 흐름도이다. 도면에서 도시하고 있는 바와 같이, 본 발명에 의한 그리드 기반 혼합형 데이터 마이닝 방법은 대용량 데이터 마이닝에 대한 작업을 위하여 작업 요청 모듈에서 각기 다른 용량의 그리드 컴퓨터 자원을 요청한다(S10).
- <67> 그리고, 작업 요청 모듈의 요청 메시지를 수신한 작업 관리 모듈은 작업 분배기로 대용량의 데이터 마이닝 작업을 동일한 계산 능력을 갖도록 분배하며(S20), 작업 관리 모듈의 분배로 인하여 그리드 자원 중개부의 각 자원 중개 모듈로 작업이 할당되고, 각 자원 중개 모듈의 동적 자원 요청기를 통해 각 자원 제공부에 동적으로 그리드 컴퓨팅 자원을 사용하기를 요청한다(S30).
- <68> 또한, 수신한 자원 요청에 따라, 각 자원 제공부의 데이터 마이닝 모듈에서 데이터 마이닝을 위한 학습 결과를 학습 결과 산출기로 산출한다(S40).
- <69> 여기서, 산출된 결과를 이용하여 테스트 결과를 테스트 결과 산출기로 산출하여 그리드 자원 중개부의 각 자원 중개 모듈 내의 중간 결과 산출기로 전송한다(S50).
- <70> 더불어, 중간 결과 산출기로 각각의 테스트 결과를 통합하여 작업 관리부의 결과 테스트 모듈로 전송하여(S60), 이를 분석 및 테스트하고(S70), 현재 데이터의 패턴 인식률이 신뢰할 수 있는 데이터 패턴 인식률인 기준 데이터 패턴 인식률에 도달할 때까지 작업 요청 모듈의 결정범위 값 재조정기에서 결정범위 값을 조정하여(S90) 상기 단계(S10 - S70)를 반복구동하되, 현재 데이터 패턴 인식률과 신뢰할 수 있는 데이터 패턴 인식률인 기준 데이터 패턴 인식률과 동일하면 종료시킨다(S80).
- <71> 도 3은 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법 중 작업 분배 방법을 개략적으로 도시한 흐름도이다. 도면에서 도시하고 있는 바와 같이, 본 발명에 의한 그리드 기반 혼합형 데이터 마이닝 방법 중 작업 분배 방법은 작업 관리 모듈에서 그리드 자원 중개부 내 자원 중개 모듈의 수 및 자원 중개 모듈의 수에 따라 분배할 작업의 수를 초기화한다(S20-1).
- <72> 그리고, 작업 관리 모듈의 작업 분배기는 각 데이터 마이닝 작업에서 요구하는 자원량을 기반으로 작업을 정렬하고(S20-2), 자원 중개 모듈의 수와 전체 대용량 마이닝 데이터 작업 수를 비교한다(S20-3).
- <73> 여기서, 자원 중개 모듈의 수가 전체 대용량 마이닝 데이터 작업 수 미만인 경우에는, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 분배하되(S20-4), 모든 작업이 자원 중개 모듈에 할당될 때까지 계속적으로 구동한다(S20-7).
- <74> 또한, 자원 중개 모듈의 수가 전체 대용량 마이닝 데이터 작업 수를 초과하는 경우에는, 작업 분배기에서 산출된 요구 자원량으로 정렬된 데이터 마이닝 작업을 자원 중개 모듈에 순차적으로 분배하는 순차적 분배 과정을

진행하고(S20-5), 마지막 자원 중개 모듈로부터 가중치를 이용하여 역순으로 분배하는 역순 분배 과정을 진행하되(S20-6), 모든 작업이 자원 중개 모듈에 할당될 때까지 상기 순차적 분배 과정(S20-5)과 역순 분배 과정(S20-6)을 반복적으로 구동한다(S20-7).

- <75> 도 4는 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법 중 동적 자원 요청 방법을 개략적으로 도시한 흐름도이다. 도면에서 도시하고 있는 바와 같이, 각 자원 중개 모듈의 동적 자원 요청기에서 모든 자원 제공부의 상태를 감시 및 갱신한다(S30-1).
- <76> 그리고, 할당된 작업량이 자원 제공부의 작업 가능량의 범위 내에 있으면(S30-2), 해당 자원 제공부와 연결을 설정하여 해당 자원을 할당받으며 종료된다(S30-7).
- <77> 반면, 할당된 작업량이 자원 제공부의 작업 가능량의 범위 내에 있지 않으면(S30-2), 상기 자원 제공부를 모두 순차적으로 탐색하고(S30-3), 총 자원 제공부의 수가 탐색 수를 초과하면(S30-4), 즉 총 자원 제공부에 대한 탐색을 완료하지 못한 경우이면, 할당된 작업량이 자원 제공부의 작업 가능량 범위 내에 있는 해당 자원 제공부를 검색했는지의 여부를 묻는다(S30-5).
- <78> 여기서, 해당 자원 제공부의 수를 모두 검색하지 않은 상태에서(S30-4), 할당된 작업량이 자원 제공부의 작업 가능량 범위 내에 있는 해당 자원을 검색하지 못하였으면(S30-5), 총 자원 제공부를 순차적으로 모두 검색할 때까지 자원 제공부를 순차적으로 탐색하는 단계(S30-3)으로 이동한다.
- <79> 이때, 총 자원 제공부의 수가 탐색수와 같으면(S30-4), 즉 모든 자원 제공부에 대한 탐색이 완료된 상태임에도 불구하고 해당 자원 제공부를 검색하지 못한 경우에는 데이터 패턴의 인식에 대한 결정범위값을 조정하여(S30-6) 할당된 작업량이 자원 제공부 작업 가능량의 범위 내에 있는지의 여부를 확인하는 단계(S30-2)로 이동한다.
- <80> 또한, 총 자원 제공부의 수가 탐색수 미만이면(S30-4), 즉 모든 자원 제공부에 대한 탐색이 완료된 상태에서는 바로 데이터 패턴의 인식 여부에 대한 결정범위값을 조정하는 단계(S30-6)로 이동한다.
- <81> 더불어, 모든 자원 제공부를 모두 탐색하여(S30-4) 할당된 작업량이 자원 제공부의 작업 가능량 범위 내에 있는 해당 자원이 검색되었으면(S30-5), 해당 자원 제공부와 연결을 설정하여 해당 자원을 할당받는 단계(S30-7)로 이동한다.
- <82> 다시 말하면, 각 자원 중개 모듈의 동적 자원 요청기는 모든 자원 제공부의 상태를 감시 및 갱신하는데, 할당된 작업량이 상기 감시한 자원 제공부의 작업 가능량 범위 내에 있다면, 해당 자원 제공부를 연결하여 해당 자원을 할당받고, 반대로 할당된 작업량이 상기 감시한 자원 제공부의 작업 가능량 범위 내에 있지 않다면, 모든 자원 제공부를 순차적으로 탐색하되, 모든 자원 제공부를 검색해도 범위 내의 해당 자원 제공부를 검색하지 못한다면, 데이터 패턴의 인식 여부에 대한 결정범위값을 조정하여 다시 순차적으로 검색하고, 해당 자원을 할당받을 때까지 계속적으로 구동시키는 동적인 요청 방법을 수행한다.
- <83> 이상에서는 본 발명의 바람직한 실시예를 예시적으로 설명하였으나, 본 발명의 범위는 이같은 특정 실시예에만 한정되지 않으며 해당 분야에서 통상의 지식을 가진자라면 본 발명의 특허 청구 범위내에 기재된 범주 내에서 적절하게 변경이 가능 할 것이다.

발명의 효과

- <84> 이상에서 설명한 바와 같이 상기와 같은 구성을 갖는 본 발명은 대용량의 데이터 마이닝을 위해 지리적으로 분산된 컴퓨팅 자원을 이용함으로써, 데이터 및 그리드 자원의 활용도를 높일 수 있으며, 대용량의 데이터 마이닝 작업에 대하여 서로 다른 그리드 자원의 연산 능력에 적합하도록 동적으로 작업을 분배하고, 대용량의 작업 처리 및 실행에서 발생가능한 병목 현상을 감소시킴으로써, 처리 시간을 단축시킬 수 있으며, 신뢰할만한 데이터 패턴 인식률에 근접할때까지 결정범위값을 재조정하고, 이를 갱신하여 자원 제공부의 테스트 비용을 단축시킬 수 있는 등의 효과를 거둘 수 있다.

도면의 간단한 설명

- <1> 도 1은 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 장치를 개략적으로 도시한 블록도.
- <2> 도 2는 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법을 개략적으로 도시한 흐름도.
- <3> 도 3은 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법 중 작업 분배 방법을 개략적으로 도시한

흐름도.

<4> 도 4는 본 발명에 따른 그리드 기반 혼합형 데이터 마이닝 방법 중 동적 자원 요청 방법을 개략적으로 도시한 흐름도.

<5> <도면의 주요 부분에 대한 도면 부호의 간단한 설명>

<6> 1: 그리드 기반 혼합형 데이터 마이닝 장치

<7> 10: 작업 관리부

11: 작업 요청 모듈

<8> 11a: 결정범위 값 재조정기

13: 작업 관리 모듈

<9> 13a: 작업 분배기

15: 결과 테스트 모듈

<10> 30: 그리드 자원 중개부

31: 자원 중개 모듈

<11> 31a: 중간 결과 산출기

31b: 동적 자원 요청기

<12> 50: 자원 제공부

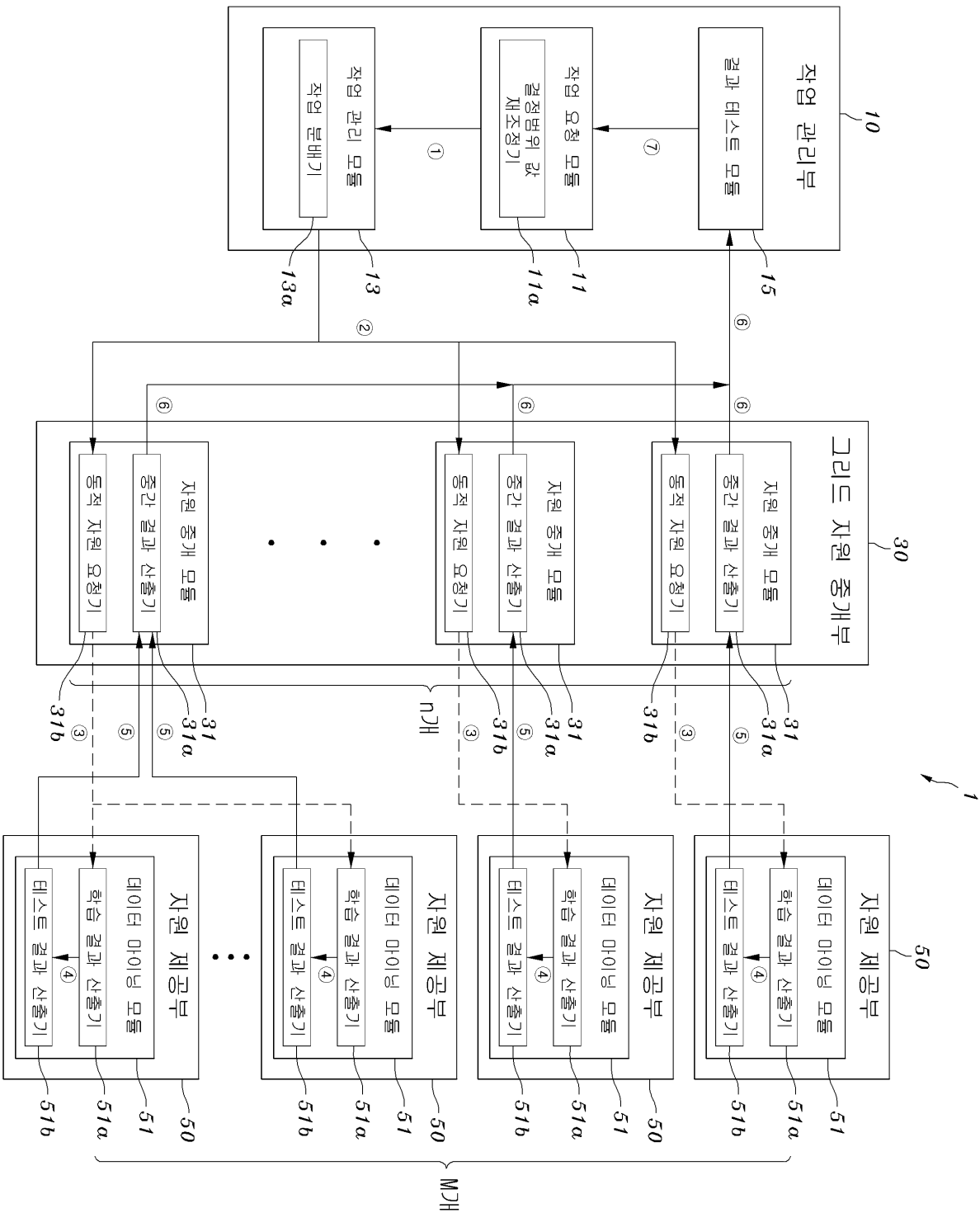
51: 데이터 마이닝 모듈

<13> 51a: 학습 결과 산출기

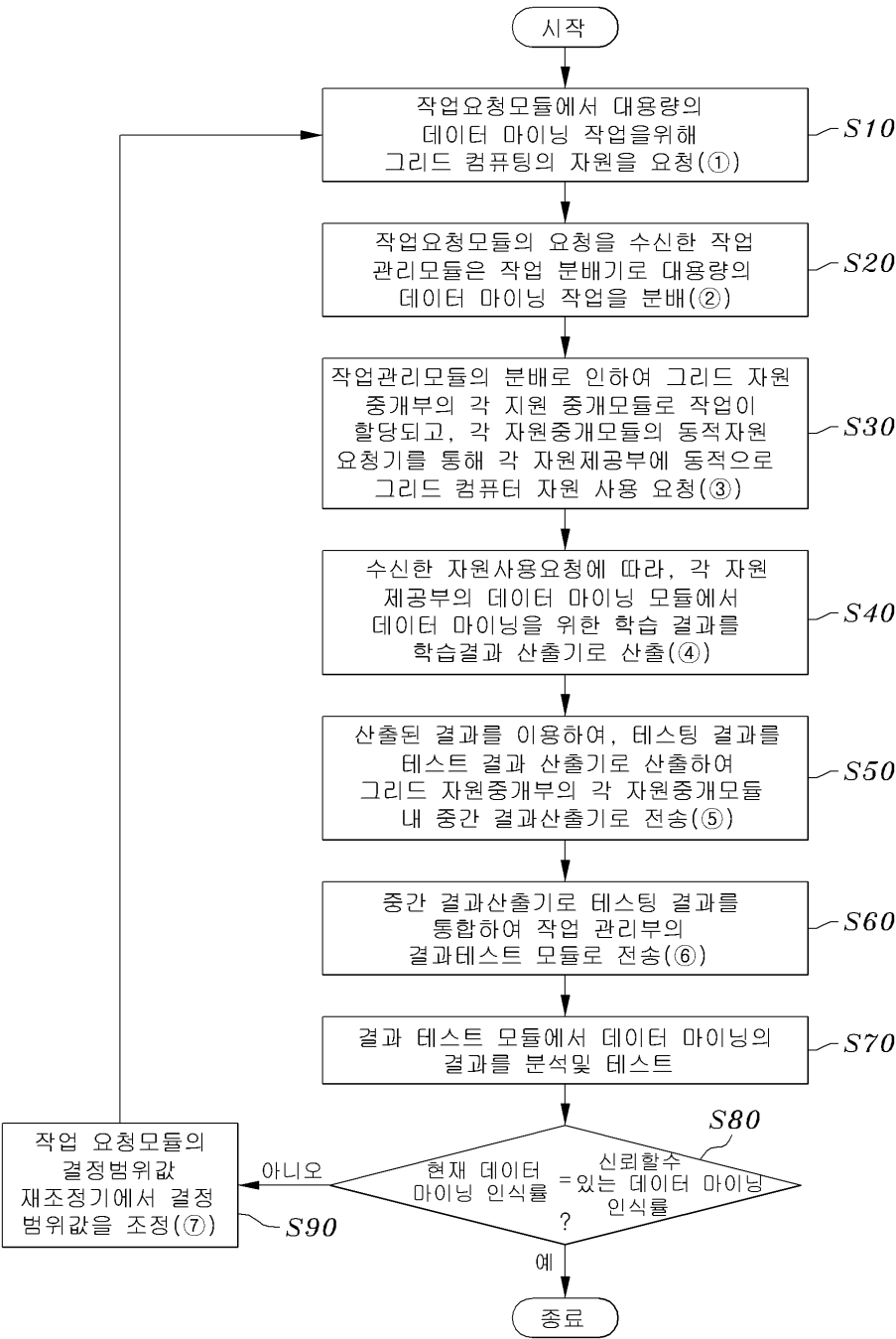
51b: 테스트 결과 산출기

도면

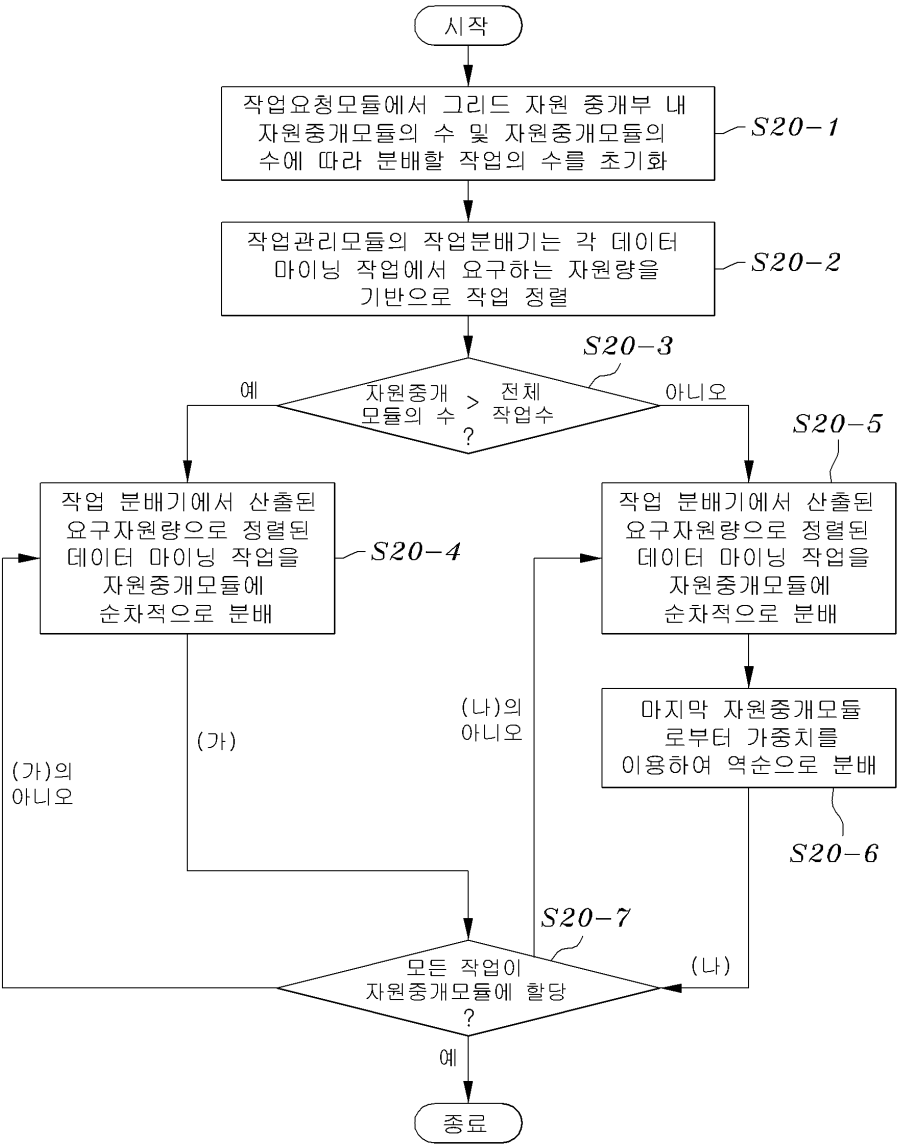
도면1



도면2



도면3



도면4

