

- (73) 특허권자
조선훈교산학협력단
광주광역시 동구 필문대로 309 (서석동)
- (72) 발명자
김영식
광주광역시 서구 풍암2로 66 금호2차아파트 201동
1005호
- (74) 대리인
김효성

심사관 : 권현수

(54) 발명의 명칭 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치 및 그 동작 방법

데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치 및 그 동작 방법이 개시된다. 본 발명은 다수의 데이터들을 유사 패턴에 따라 자동으로 분류하기 위한 정보를 결정하기 위해 수행하는 데이터 마이닝(data mining)에 따른 기계 학습 훈련 과정에서 데이터들에 대한 분류의 적합도를 측정할 수 있는 장치 및 그 동작 방법을 제공함으로써, 데이터 마이닝에서의 데이터 분류의 신뢰성을 향상시킬 수 있다.

[illegible]

(52) CPC특허분류
G06F 2216/03 (2013.01)

(56) 선행기술조사문헌
KR1020170030401 A*
KR1020180041478 A*
KR101484186 B1
JP5012078 B2
KR1020170143298
*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711050933
부처명	과학기술정보통신부
연구관리전문기관	한국연구재단
연구사업명	중견연구(총연구비3억초과~5억이하)
연구과제명	양자 연산 저항성을 갖는 사물인터넷 기반 커넥티드카 보안 기반 기술 연구
기 여 율	1/1
주관기관	조선대학교 산학협력단
연구기간	2017.03.01 ~ 2018.02.28

명세서

청구범위

청구항 1

데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치에 있어서,

데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들 - 상기 복수의 데이터들 각각에는 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 미리 지정된 인덱스 값이 삽입되어 있되, 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들에 삽입되어 있는 인덱스 값들은 서로 동일함 - 이 복수의 데이터 분류 그룹들로 분류되면, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산하는 연산부; 및

상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력하는 측정 값 출력부

를 포함하고,

상기 연산부는

상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 데이터가 순차적으로 하나씩 분류될 때마다, 현재 순번에서 분류되는 데이터에 삽입되어 있는 인덱스 값과 이전 순번에서 분류된 데이터에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인하여 상기 복수의 데이터 분류 그룹들 각각에 대해서 양 인덱스 값이 서로 동일한 것으로 확인된 총 충돌 개수를 카운트하는 충돌 개수 카운트부;

상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 분류된 데이터의 총 분류 개수를 확인하는 분류 개수 확인부; 및

상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 총 충돌 개수와 상기 총 분류 개수에 기초하여 상기 분류 적합도 측정 값을 연산하는 측정 값 연산부

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치.

청구항 2

제1항에 있어서,

상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값과 선정된(predetermined) 기준 값을 비교하여 상기 복수의 데이터들에 대한 데이터 분류의 적합도를 향상시키기 위해서 상기 기계 학습 훈련을 추가로 수행할 것을 지시하는 추가 학습 훈련 지시 메시지를 화면 상에 출력하는 메시지 출력부

를 더 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 측정 값 연산부는

하기의 수학적 식 1에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 상기 분류 적합도 측정 값을 연산하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치.

[수학적 식 1]

$$I_t = 1 - \frac{C_t}{N_t}$$

여기서, I_t 는 상기 복수의 데이터 분류 그룹들 중 t 번째 데이터 분류 그룹에 대해 연산되는 분류 적합도 측정 값, C_t 는 t 번째 데이터 분류 그룹에서 카운트된 총 충돌 개수, N_t 는 t 번째 데이터 분류 그룹에서 확인된 총 분류 개수임.

청구항 5

제2항에 있어서,

상기 메시지 출력부는

상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는지 여부를 확인하는 기준 값 초과 확인부;

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 데이터 분류 그룹의 개수가 선정된 기준 개수를 초과하는지 여부를 확인하는 기준 개수 초과 확인부; 및

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리하는 메시지 출력 처리부

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치.

청구항 6

제5항에 있어서,

상기 메시지 출력 처리부는

미리 지정된 서로 다른 복수의 분류 개수의 범위들 별로 미리 지정된 서로 다른 중요도 점수가 대응되어 기록되어 있는 중요도 점수 테이블을 저장하여 유지하는 중요도 점수 테이블 유지부;

미리 지정된 서로 다른 복수의 점수 범위들 별로 미리 지정된 서로 다른 반복 수행 횟수가 대응되어 기록되어 있는 수행 횟수 테이블을 저장하여 유지하는 수행 횟수 테이블 유지부;

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 적어도 하나의 제1 데이터 분류 그룹을 확인한 후 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해, 상기 중요도 점수 테이블을 참조하여 상기 적어도 하나의 제1 데이터 분류 그룹 각각에서 분류된 데이터의 총 분류 개수에 따른 분류 개수의 범위에 대응하는 중요도 점수를 확인하는 중요도 점수 확인부;

상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해 확인된 중요도 점수를 합산하여 합계 점수를 연산하는 합계 점수 연산부;

상기 합계 점수가 연산되면, 상기 수행 횟수 테이블로부터 상기 연산된 합계 점수에 따른 점수 범위에 대응하는 제1 반복 수행 횟수를 추출하는 수행 횟수 추출부; 및

상기 제1 반복 수행 횟수가 추출되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리함과 동시에 상기 제1 반복 수행 횟수를 화면 상에 출력하는 정보 출력 처리부

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치.

청구항 7

데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법에 있어서,

데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들 - 상기 복수의 데이터들 각각에는 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 미리 지정된 인덱스 값이 삽입되어 있되, 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들에 삽입되어 있는 인덱스 값들은 서로 동일함 - 이 복수의 데이터 분류 그룹들로 분류되면, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산하는 단계; 및

상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력하는 단계를 포함하고,

상기 연산하는 단계는

상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 데이터가 순차적으로 하나씩 분류될 때마다, 현재 순번에서 분류되는 데이터에 삽입되어 있는 인덱스 값과 이전 순번에서 분류된 데이터에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인하여 상기 복수의 데이터 분류 그룹들 각각에 대해서 양 인덱스 값이 서로 동일한 것으로 확인된 총 충돌 개수를 카운트하는 단계;

상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 분류된 데이터의 총 분류 개수를 확인하는 단계; 및

상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 총 충돌 개수와 상기 총 분류 개수에 기초하여 상기 분류 적합도 측정 값의 연산을 처리하는 단계

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법.

청구항 8

제7항에 있어서,

상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값과 선정된(predetermined) 기준 값을 비교하여 상기 복수의 데이터들에 대한 데이터 분류의 적합도를 향상시키기 위해서 상기 기계 학습 훈련을 추가로 수행할 것을 지시하는 추가 학습 훈련 지시 메시지를 화면 상에 출력하는 단계

를 더 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법.

청구항 9

삭제

청구항 10

제7항에 있어서,

상기 연산을 처리하는 단계

하기의 수학적 식 2에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 상기 분류 적합도 측정 값을 연산하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법.

[수학적 식 2]

$$I_t = 1 - \frac{C_t}{N_t}$$

여기서, I_t 는 상기 복수의 데이터 분류 그룹들 중 t 번째 데이터 분류 그룹에 대해 연산되는 분류 적합도 측정 값, C_t 는 t 번째 데이터 분류 그룹에서 카운트된 총 충돌 개수, N_t 는 t 번째 데이터 분류 그룹에서 확인된 총 분류 개수임.

청구항 11

제8항에 있어서,

상기 추가 학습 훈련 지시 메시지를 화면 상에 출력하는 단계는

상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는지 여부를 확인하는 단계;

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 데이터 분류 그룹의 개수가 선정된 기준 개수를 초과하는지 여부를 확인하는 단계; 및

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리하는 단계

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법.

청구항 12

제11항에 있어서,

상기 추가 학습 훈련 지시 메시지의 출력을 처리하는 단계는

미리 지정된 서로 다른 복수의 분류 개수의 범위들 별로 미리 지정된 서로 다른 중요도 점수가 대응되어 기록되어 있는 중요도 점수 테이블을 저장하여 유지하는 단계;

미리 지정된 서로 다른 복수의 점수 범위들 별로 미리 지정된 서로 다른 반복 수행 횟수가 대응되어 기록되어 있는 수행 횟수 테이블을 저장하여 유지하는 단계;

상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 적어도 하나의 제1 데이터 분류 그룹을 확인한 후 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해, 상기 중요도 점수 테이블을 참조하여 상기 적어도 하나의 제1 데이터 분류 그룹 각각에서 분류된 데이터의 총 분류 개수에 따른 분류 개수의 범위에 대응하는 중요도 점수를 확인하는 단계;

상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해 확인된 중요도 점수를 합산하여 합계 점수를 연산하는 단계;

상기 합계 점수가 연산되면, 상기 수행 횟수 테이블로부터 상기 연산된 합계 점수에 따른 점수 범위에 대응하는 제1 반복 수행 횟수를 추출하는 단계; 및

상기 제1 반복 수행 횟수가 추출되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리함과 동시에 상기 제1 반복 수행 횟수를 화면 상에 출력하는 단계

를 포함하는 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법.

청구항 13

제7항, 제8항, 제10항, 제11항 또는 제12항 중 어느 한 항의 방법을 수행하는 프로그램을 기록한 컴퓨터 판독 가능 기록 매체.

청구항 14

제7항, 제8항, 제10항, 제11항 또는 제12항 중 어느 한 항의 방법을 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

본 발명은 다수의 데이터들을 유사 패턴에 따라 자동으로 분류하기 위한 정보를 결정하기 위해 수행하는 데이터 마이닝(data mining)에 따른 기계 학습 훈련 과정에서 데이터들에 대한 분류의 적합도를 측정할 수 있는 기법에 대한 것이다.

[0001]

배경 기술

- [0002] 최근, 인터넷이 널리 보급되고, 시간과 장소의 제약 없이 인터넷에 연결할 수 있는 환경이 갖추어지면서, 수많은 데이터들이 생산되고, 유통되고 있다.
- [0003] 이렇게, 수많은 데이터들이 생산되고, 유통됨에 따라, 이러한 데이터들의 패턴 등의 분석을 통해서 사람들의 생각, 성향 등을 파악하고, 이를 기초로 광고 서비스나 공공 서비스를 제공하는 기술들도 등장하고 있다.
- [0004] 대용량의 데이터들의 패턴이나 유사성 등을 분석해서 데이터 내에 숨겨져 있는 유용한 정보를 찾아내는 방법을 데이터 마이닝(data mining)이라고 한다. 이러한 데이터 마이닝을 통해 데이터들 간의 패턴이나 유사성을 분석해 낸 후 이러한 유사성 분석 결과에 근거하여 다양한 데이터들을 유사한 분류끼리 그룹화하기 위한 연구가 활발하게 진행되고 있다.
- [0005] 데이터 마이닝을 통해 데이터를 분류하기 위한 패턴을 분석하는 방법으로는 데이터 분류 작업을 다수회 반복 수행하는 기계 학습 훈련 방식이 많이 사용된다.
- [0006] 기계 학습 훈련 방식은 데이터 마이닝을 통해 데이터들을 분류해 보고, 각 분류 결과에 따른 오류 등을 확인하는 과정을 다수회 수행함으로써, 데이터 분류의 적합도를 높이는 방식이다.
- [0007] 이러한 기계 학습 훈련을 통해 데이터 분류를 위한 패턴 정보를 찾아내기 위해서는 기계 학습 훈련시마다 데이터 분류 결과에 대한 적합도를 측정할 필요가 있다. 만약, 적합도가 낮다면 기계 학습 훈련을 추가로 수행함으로써, 데이터 분류의 적합도를 높일 수 있기 때문이다.
- [0008] 따라서, 데이터 마이닝에 따른 기계 학습 훈련에 있어서 데이터 분류 결과에 대한 적합도를 측정하기 위한 기술의 연구가 필요하다.

발명의 내용

해결하려는 과제

- [0009] 본 발명은 다수의 데이터들을 유사 패턴에 따라 자동으로 분류하기 위한 정보를 결정하기 위해 수행하는 데이터 마이닝(data mining)에 따른 기계 학습 훈련 과정에서 데이터들에 대한 분류의 적합도를 측정할 수 있는 장치 및 그 동작 방법을 제공함으로써, 데이터 마이닝에서의 데이터 분류의 신뢰성 향상을 도모하고자 한다.

과제의 해결 수단

- [0010] 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치는 데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들 - 상기 복수의 데이터들 각각에는 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 미리 지정된 인덱스 값이 삽입되어 있되, 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들에 삽입되어 있는 인덱스 값들은 서로 동일함 - 이 복수의 데이터 분류 그룹들로 분류되면, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산하는 연산부 및 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력하는 측정 값 출력부를 포함한다.
- [0011] 또한, 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법은 데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들 - 상기 복수의 데이터들 각각에는 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 미리 지정된 인덱스 값이 삽입되어 있되, 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들에 삽입되어 있는 인덱스 값들은 서로 동일함 - 이 복수의 데이터 분류 그룹들로 분류되면, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산하는 단계 및 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력하는 단계를 포함한다.

발명의 효과

- [0012] 본 발명은 다수의 데이터들을 유사 패턴에 따라 자동으로 분류하기 위한 정보를 결정하기 위해 수행하는 데이터 마이닝(data mining)에 따른 기계 학습 훈련 과정에서 데이터들에 대한 분류의 적합도를 측정할 수 있는 장치

및 그 동작 방법을 제공함으로써, 데이터 마이닝에서의 데이터 분류의 신뢰성을 향상시킬 수 있다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 구조를 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법을 도시한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하에서는 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명하기로 한다. 이러한 설명은 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였으며, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 본 명세서 상에서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 사람에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다.

[0015] 도 1은 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 구조를 도시한 도면이다.

[0016] 도 1을 참조하면, 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치(110)는 연산부(111) 및 측정 값 출력부(112)를 포함한다.

[0017] 먼저, 데이터를 분류하기 위한 패턴 규칙을 결정하기 위한 목적으로 데이터 마이닝을 위한 기계 학습 훈련에 사용될 복수의 데이터들이 존재한다고 하였을 때, 관리자는 상기 복수의 데이터들 각각에 대해 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 인덱스 값을 미리 삽입해 둘 수 있다. 이때, 관리자는 상기 복수의 데이터들 중 동일한 그룹으로 분류되어야 할 데이터들에 대해 동일한 인덱스 값을 삽입해둘 수 있다.

[0018] 이러한 상황 하에서, 데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들이 복수의 데이터 분류 그룹들로 분류되면, 연산부(111)는 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산한다.

[0019] 이때, 본 발명의 일실시예에 따르면, 연산부(111)는 상기 분류 적합도 측정 값의 연산을 위한 구체적 구성으로, 총돌 개수 카운트부(114), 분류 개수 확인부(115) 및 측정 값 연산부(116)를 포함할 수 있다.

[0020] 총돌 개수 카운트부(114)는 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 데이터가 순차적으로 하나씩 분류될 때마다, 현재 순번에서 분류되는 데이터에 삽입되어 있는 인덱스 값과 이전 순번에서 분류된 데이터에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인하여 상기 복수의 데이터 분류 그룹들 각각에 대해서 양 인덱스 값이 서로 동일한 것으로 확인된 총 총돌 개수를 카운트한다.

[0021] 분류 개수 확인부(115)는 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 분류된 데이터의 총 분류 개수를 확인한다.

[0022] 측정 값 연산부(116)는 상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 총 총돌 개수와 상기 총 분류 개수에 기초하여 상기 분류 적합도 측정 값을 연산한다.

[0023] 이때, 본 발명의 일실시예에 따르면, 측정 값 연산부(116)는 하기의 수학적 식 1에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 상기 분류 적합도 측정 값을 연산할 수 있다.

수학식 1

$$I_t = 1 - \frac{C_t}{N_t}$$

[0025]

[0027]

여기서, I_t 는 상기 복수의 데이터 분류 그룹들 중 t 번째 데이터 분류 그룹에 대해 연산되는 분류 적합도 측정값, C_t 는 t 번째 데이터 분류 그룹에서 카운트된 총 충돌 개수, N_t 는 t 번째 데이터 분류 그룹에서 확인된 총 분류 개수이다.

[0028]

이때, 상기 수학식 1에서 나타난 분류 적합도 측정값이 0에 가까운 값을 가질수록 데이터 분류의 적합도가 높은 것으로 판정될 수 있고, 분류 적합도 측정값이 1에 가까운 값을 가질수록 데이터 분류의 적합도가 낮은 것으로 판정될 수 있다.

[0029]

관련해서, 데이터 마이닝을 위한 기계 학습 훈련에 사용될 상기 복수의 데이터들이 6개가 존재한다고 하고, 6개의 데이터들에 하기의 표 1과 같이 인덱스 값이 삽입되어 있다고 하며, 6개의 데이터들이 기계 학습 훈련에 따라 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3에 분류된다고 가정하여 충돌 개수 카운트부(114), 분류 개수 확인부(115) 및 측정값 연산부(116)의 동작을 예를 들어 설명하면 다음과 같다.

표 1

[0031]

데이터들	인덱스 값
데이터 1	인덱스 1
데이터 2	인덱스 1
데이터 3	인덱스 2
데이터 4	인덱스 3
데이터 5	인덱스 2
데이터 6	인덱스 3

[0033]

충돌 개수 카운트부(114)는 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3 각각에 대해 6개의 데이터들이 순차적으로 하나씩 분류될 때마다, 현재 순번에서 분류되는 데이터에 삽입되어 있는 인덱스 값과 이전 순번에서 분류된 데이터에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인하여 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3 각각에 대해서 양 인덱스 값이 서로 동일한 것으로 확인된 총 충돌 개수를 카운트할 수 있다.

[0034]

관련해서, 기계 학습 훈련이 수행됨에 따라 데이터 분류 그룹 1에 데이터 1, 데이터 2, 데이터 5가 순차적으로 분류된다고 하면, 충돌 개수 카운트부(114)는 데이터 1이 분류된 후 데이터 2가 분류될 때, 이전 순번에서 분류된 데이터인 데이터 1에 삽입되어 있는 인덱스 값과 현재 순번에서 분류된 데이터인 데이터 2에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인할 수 있고, 데이터 5가 분류될 때, 이전 순번에서 분류된 데이터인 데이터 2에 삽입되어 있는 인덱스 값과 현재 순번에서 분류된 데이터인 데이터 5에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인할 수 있다. 그리고, 충돌 개수 카운트부(114)는 각 확인 과정에서 양 인덱스 값이 서로 동일한 것으로 확인된 총 충돌 개수를 카운트할 수 있다. 관련해서, 데이터 1과 데이터 2 사이의 인덱스 값만 서로 동일하기 때문에 충돌 개수 카운트부(114)는 데이터 분류 그룹 1에 대해 총 충돌 개수로 1개를 카운트할 수 있다.

[0035]

이러한 방식으로, 충돌 개수 카운트부(114)는 데이터 분류 그룹 2와 데이터 분류 그룹 3에 대해서도 총 충돌 개수를 카운트할 수 있다.

[0036]

이때, 분류 개수 확인부(115)는 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3 각각에 대해, 각 데이터 분류 그룹들 별로 분류된 데이터의 총 분류 개수를 확인할 수 있다.

[0037]

관련해서, 분류 개수 확인부(115)는 데이터 분류 그룹 1에 대해 데이터 1, 데이터 2, 데이터 5가 분류되었기 때문에 데이터 분류 그룹 1에 대한 데이터의 총 분류 개수로 3개를 확인할 수 있고, 이러한 방식으로 나머지 데이터 분류 그룹인 데이터 분류 그룹 2, 데이터 분류 그룹 3에 대해서도 데이터의 총 분류 개수를 확인할 수 있다.

- [0038] 이렇게, 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3에 대해 총 충돌 개수와 총 분류 개수가 확인되면, 측정 값 연산부(116)는 상기 수학적 식 1의 연산에 따라 상기 분류 적합도 측정 값을 연산할 수 있다.
- [0039] 관련해서, 데이터 분류 그룹 1에 대해서는 상기 총 충돌 개수가 1개, 상기 총 분류 개수가 3개로 확인되었기 때문에 측정 값 연산부(116)는 상기 수학적 식 1의 연산에 따라 0.67이라고 하는 상기 분류 적합도 측정 값을 연산할 수 있다.
- [0040] 이러한 방식으로, 측정 값 연산부(116)는 데이터 분류 그룹 1, 데이터 분류 그룹 2에 대해서도 상기 수학적 식 1의 연산에 기초하여 상기 분류 적합도 측정 값을 연산할 수 있다.
- [0041] 보통, 데이터 분류의 적합도를 측정하는 척도로는 하기의 수학적 식 2에서 나타낸 바와 같은 지니(gini) 인덱스를 주로 사용한다.

수학적 식 2

$$I_G(p) = \sum_{i=1}^J p_i \sum_{k \neq i} p_k = \sum_{i=1}^J p_i (1 - p_i) = 1 - \sum_{i=1}^J p_i^2$$

- [0043]
- [0045] 상기 수학적 식 2에서 $I_G(p)$ 는 $i(i \in \{1, 2, \dots, J\})$ 라고 하는 인덱스 값이 할당되어 있는 J개의 원소들이 특정 그룹에 분류되었을 때 해당 그룹에서의 지니 인덱스를 의미하는 것으로, p_i 는 상기 그룹에 속해있는 전체 J개의 원소들 중 각 인덱스 값들이 할당된 원소들이 존재하는 비율을 의미한다.
- [0046] 이때, 상기 수학적 식 2에 따라 연산되는 지니 인덱스가 0에 가까울수록 상기 그룹 상에 원소들이 적절하게 분류가 잘 된 것으로 판정될 수 있고, 상기 지니 인덱스가 $1 - \frac{1}{J}$ 에 가까운 값을 가질수록 상기 그룹 상에 원소들이 랜덤하게 분류된 것으로 판정될 수 있다.
- [0047] 앞서, 수학적 식 1에서 나타낸 분류 적합도 측정 값은 하기의 표 2와 같은 수학적 특징에 의해 상기 수학적 식 2에 따라 연산되는 지니 인덱스에 수렴되는 특징을 가지고 있다.

표 2

[0049]

정리) 상기 수학적 식 1에서 나타낸 분류 적합도 측정 값은 상기 수학적 식 2에서 나타낸 지니 인덱스로 수렴함
증명)

먼저 $g(i)$ 를 다음과 같은 지시 함수(indicator function)라 하자.

$$g(i) = \begin{cases} 1, & \text{if } i = 1 \\ 0, & \text{otherwise} \end{cases}$$

이때, 특정 분류 그룹에 하나의 데이터가 분류되어 들어오는 것을 해당 분류 그룹에서 하나의 랜덤 소스를 가지면서 소스가 J 개의 클래스 중 하나에 해당하는 인덱스 값을 생성하는 것으로 모델링할 수 있다. 이때, 해당 소스를 에르고딕(ergodic) 랜덤 소스라 하면, $g(i)$ 의 기댓값은 다음과 같이 연산될 수 있다.

$$\frac{1}{N} \sum_{i=1}^N G_n(s) = E[g(D_n(U))] = \sum_{i=0}^{\infty} \Pr(D_n(U) = 1) g(i)$$

여기에서 D_n 은 다음과 같이 측정되는 거리 함수이다.

$$D_n(s) = \begin{cases} n & \text{if } \forall i < n, b_n(s^N) \neq b_{n-i}(s) \\ \min\{i | i \geq 1, b_n(s) = b_{n-i}(s)\} & \text{otherwise} \end{cases}$$

그리고, s 는 소스가 생성한 랜덤 값이고, 이에 대응되는 U 는 확률 변수이다. 그러면, $b_n(U)$ 이 통계적 독립 동일 분포(Independent and Identically Distribution: IID)를 가질 경우, 다음과 같은 수식이 성립한다.

$$C_J \text{가 클래스 인덱스들의 집합이라 할 때} \\ \Pr(D_n(U) = 1) = \sum_{b \in C_J} \Pr(b_n(s) = b_{n-1}(s) = b) = \sum_{b \in C_J} \Pr(b)^2$$

이에, $g(i)$ 의 기댓값은 다음과 같이 나타낼 수 있다.

$$C/N = E[g(D_n(U))] = \sum_{b \in C_J} \sum_{i=0}^{\infty} \Pr(b)^2 g(i) = \sum_{b \in C_J} \Pr(b)^2 g(1) = E[\Pr(b)]$$

따라서, 지니 인덱스는 다음과 같이 정리될 수 있다.

$$I_G = 1 - \sum_{i=1}^J \Pr(b_i)^2 = 1 - C/N$$

[0051]

이렇게, 상기 수학적 식 1에서 나타낸 분류 적합도 측정 값이 지니 인덱스로 수렴하는 특징을 가지고 있기 때문에, 연산부(111)는 복수의 데이터들이 기계 학습 훈련에 따라 복수의 데이터 분류 그룹들로 분류되었을 때, 해당 데이터의 분류에 대한 적합도를 표현하는 척도로서 상기 수학적 식 1에서 나타낸 분류 적합도 측정 값을 연산할 수 있다.

[0052]

이때, 상기 수학적 식 1에서 나타낸 분류 적합도 측정 값이 0에 가까운 값을 가질수록 데이터 분류의 적합도가 높은 것으로 판정될 수 있고, 분류 적합도 측정 값이 1에 가까운 값을 가질수록 데이터 분류의 적합도가 낮은 것으로 판정될 수 있다.

[0053]

이렇게, 연산부(111)를 통해, 상기 복수의 데이터 분류 그룹들 각각에 대한 상기 분류 적합도 측정 값의 연산이

완료되면, 측정 값 출력부(112)는 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력할 수 있다.

[0054] 이를 통해, 관리자는 각 데이터 분류 그룹들 별로 상기 복수의 데이터들이 얼마만큼의 적합도를 갖도록 분류가 되었는지 여부를 확인할 수 있고, 분류의 적합도가 낮은 것으로 판정된다면, 추가적으로 상기 복수의 데이터들에 대해 데이터 분류를 위한 기계 학습 훈련을 반복 수행함으로써, 데이터의 분류를 위한 패턴 특징을 찾아낼 수 있다.

[0055] 이때, 본 발명의 일실시예에 따르면, 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치(110)는 메시지 출력부(113)를 더 포함할 수 있다.

[0056] 메시지 출력부(113)는 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값과 선정된(predetermined) 기준 값을 비교하여 상기 복수의 데이터들에 대한 데이터 분류의 적합도를 향상시키기 위해서 상기 기계 학습 훈련을 추가로 수행할 것을 지시하는 추가 학습 훈련 지시 메시지를 화면 상에 출력한다.

[0057] 이때, 본 발명의 일실시예에 따르면, 메시지 출력부(113)는 기준 값 초과 확인부(117), 기준 개수 초과 확인부(118) 및 메시지 출력 처리부(119)를 포함할 수 있다.

[0058] 기준 값 초과 확인부(117)는 상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는지 여부를 확인한다.

[0059] 기준 개수 초과 확인부(118)는 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 데이터 분류 그룹의 개수가 선정된 기준 개수를 초과하는지 여부를 확인한다.

[0060] 메시지 출력 처리부(119)는 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리한다.

[0061] 관련해서, 상기 복수의 데이터들에 대해 데이터 마이닝에 따른 기계 학습 훈련을 수행한 결과, 상기 복수의 데이터들이 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3, 데이터 분류 그룹 4로 분류되었다고 하고, 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3에 대해서 연산된 분류 적합도 측정 값이 상기 선정된 기준 값을 초과한다고 하며, 상기 선정된 기준 개수가 2개라고 한다면, 메시지 출력 처리부(119)는 4개의 데이터 분류 그룹들 중 상기 선정된 기준 개수인 2개를 초과하는 3개의 데이터 분류 그룹들에서 연산된 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하기 때문에 상기 기계 학습 훈련을 추가로 수행할 것을 지시하는 상기 추가 학습 훈련 지시 메시지 화면 상에 출력할 수 있다.

[0062] 즉, 특정 개수를 초과하는 데이터 분류 그룹에서의 분류 적합도가 낮은 경우, 전체적으로 데이터 마이닝에 따른 분류 적합도가 낮은 것으로 볼 수 있으므로, 메시지 출력 처리부(119)는 기계 학습 훈련의 추가 반복을 지시하는 상기 추가 학습 훈련 지시 메시지를 화면 상에 출력할 수 있다.

[0063] 이때, 본 발명의 일실시예에 따르면, 메시지 출력 처리부(119)는 중요도 점수 테이블 유지부(120), 수행 횟수 테이블 유지부(121), 중요도 점수 확인부(122), 합계 점수 연산부(123), 수행 횟수 추출부(124) 및 정보 출력 처리부(125)를 포함할 수 있다.

[0064] 중요도 점수 테이블 유지부(120)는 미리 지정된 서로 다른 복수의 분류 개수의 범위들 별로 미리 지정된 서로 다른 중요도 점수가 대응되어 기록되어 있는 중요도 점수 테이블을 저장하여 유지한다.

[0065] 예컨대, 상기 중요도 점수 테이블에는 하기의 표 3과 같이 정보가 기록되어 있을 수 있다.

표 3

분류 개수의 범위	중요도 점수
0~5개	10점
5~10개	20점
10~15개	30점
...	...

[0069] 수행 횟수 테이블 유지부(121)는 미리 지정된 서로 다른 복수의 점수 범위를 별로 미리 지정된 서로 다른 반복 수행 횟수가 대응되어 기록되어 있는 수행 횟수 테이블을 저장하여 유지한다.

[0070] 예컨대, 상기 수행 횟수 테이블에는 하기의 표 4와 같이 정보가 기록되어 있을 수 있다.

표 4

점수 범위	수행 횟수
30~50점	1회
50~90점	2회
90~120점	3회
...	...

[0074] 중요도 점수 확인부(122)는 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 적어도 하나의 제1 데이터 분류 그룹을 확인한 후 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해, 상기 중요도 점수 테이블을 참조하여 상기 적어도 하나의 제1 데이터 분류 그룹 각각에서 분류된 데이터의 총 분류 개수에 따른 분류 개수의 범위에 대응하는 중요도 점수를 확인한다.

[0075] 관련해서, 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 상기 적어도 하나의 제1 데이터 분류 그룹이 데이터 분류 그룹 1, 데이터 분류 그룹 2, 데이터 분류 그룹 3이라고 하고, 데이터 분류 그룹 1에 3개의 데이터가 분류되었고, 데이터 분류 그룹 2에 8개의 데이터가 분류되었으며, 데이터 분류 그룹 3에 13개의 데이터가 분류되었다고 하는 경우, 중요도 점수 확인부(122)는 데이터 분류 그룹 1에 대해, 상기 표 3과 같은 중요도 점수 테이블을 참조하여 데이터 분류 그룹 1에서 분류된 데이터의 총 분류 개수인 3개에 따른 분류 개수의 범위에 대응하는 중요도 점수인 10점을 확인할 수 있고, 데이터 분류 그룹 2에 대해, 상기 중요도 점수 테이블을 참조하여 데이터 분류 그룹 2에서 분류된 데이터의 총 분류 개수인 8개에 따른 분류 개수의 범위에 대응하는 중요도 점수인 20점을 확인할 수 있으며, 데이터 분류 그룹 3에 대해, 상기 중요도 점수 테이블을 참조하여 데이터 분류 그룹 3에서 분류된 데이터의 총 분류 개수인 13개에 따른 분류 개수의 범위에 대응하는 중요도 점수인 30점을 확인할 수 있다.

[0076] 합계 점수 연산부(123)는 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해 확인된 중요도 점수를 합산하여 합계 점수를 연산한다.

[0077] 수행 횟수 추출부(124)는 상기 합계 점수가 연산되면, 상기 수행 횟수 테이블로부터 상기 연산된 합계 점수에 따른 점수 범위에 대응하는 제1 반복 수행 횟수를 추출한다.

[0078] 관련해서, 앞서 설명한 예시와 같이 데이터 분류 그룹 1에 대해 중요도 점수가 10점, 데이터 분류 그룹 2에 대해 중요도 점수가 20점, 데이터 분류 그룹 3에 대해 중요도 점수가 30점으로 확인되었다고 한다면, 합계 점수 연산부(123)는 각 중요도 점수를 합산하여 60점이라고 하는 합계 점수를 연산할 수 있고, 수행 횟수 추출부(124)는 상기 표 4와 같은 상기 수행 횟수 테이블로부터 60점이라는 합계 점수에 따른 점수 범위에 대응하는 2회라고 하는 상기 제1 반복 수행 횟수를 추출할 수 있다.

[0079] 이렇게, 상기 제1 반복 수행 횟수가 추출되면, 정보 출력 처리부(125)는 상기 추가 학습 훈련 지시 메시지의 출력을 처리함과 동시에 상기 제1 반복 수행 횟수를 화면 상에 출력한다.

[0080] 즉, 앞서 설명한 예시와 같이 상기 제1 반복 수행 횟수가 2회라고 하는 경우, 정보 출력 처리부(125)는 화면 상에 상기 추가 학습 훈련 지시 메시지를 출력함과 동시에 2회라고 하는 상기 제1 반복 수행 횟수를 출력함으로써, 관리자로 하여금 기계 학습 훈련을 최소한 2회 이상 반복 수행하도록 유도하여 데이터 마이닝에 따른 분류 패턴이 보다 정확하게 결정될 수 있도록 지원할 수 있다.

[0081] 도 2는 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법을 도시한 순서도이다.

[0082] 단계(S210)에서는 데이터 마이닝 분석에 기초한 기계 학습 훈련이 수행됨에 따라 복수의 데이터들(상기 복수의 데이터들 각각에는 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들을 구분하기 위한 미리 지정된 인덱스 값이 삽입되어 있되, 동일한 데이터 분류 그룹으로 분류되어야 할 데이터들에 삽입되어 있는 인덱스 값들은 서

로 동일함)이 복수의 데이터 분류 그룹들로 분류되면, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들로 분류된 데이터에 삽입되어 있는 인덱스 값에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 분류 적합도 측정 값을 연산한다.

[0083] 단계(S220)에서는 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값을 화면 상에 출력한다.

[0084] 이때, 본 발명의 일실시예에 따르면, 상기 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법은 상기 복수의 데이터 분류 그룹들 각각에 대해서 연산된 상기 분류 적합도 측정 값과 선정된 기준 값을 비교하여 상기 복수의 데이터들에 대한 데이터 분류의 적합도를 향상시키기 위해서 상기 기계 학습 훈련을 추가로 수행할 것을 지시하는 추가 학습 훈련 지시 메시지를 화면 상에 출력하는 단계를 더 포함할 수 있다.

[0085] 또한, 본 발명의 일실시예에 따르면, 단계(S210)에서는 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 데이터가 순차적으로 하나씩 분류될 때마다, 현재 순번에서 분류되는 데이터에 삽입되어 있는 인덱스 값과 이전 순번에서 분류된 데이터에 삽입되어 있는 인덱스 값이 서로 동일한지 여부를 확인하여 상기 복수의 데이터 분류 그룹들 각각에 대해서 양 인덱스 값이 서로 동일한 것으로 확인된 총 충돌 개수를 카운트하는 단계, 상기 복수의 데이터 분류 그룹들 각각에 대해, 각 데이터 분류 그룹들 별로 분류된 데이터의 총 분류 개수를 확인하는 단계 및 상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 총 충돌 개수와 상기 총 분류 개수에 기초하여 상기 분류 적합도 측정 값의 연산을 처리하는 단계를 포함할 수 있다.

[0086] 이때, 상기 연산을 처리하는 단계 상기 수식 1에 기초하여 상기 복수의 데이터 분류 그룹들 각각에 대한 상기 분류 적합도 측정 값을 연산할 수 있다.

[0087] 또한, 본 발명의 일실시예에 따르면, 상기 추가 학습 훈련 지시 메시지를 화면 상에 출력하는 단계는 상기 복수의 데이터 분류 그룹들 각각에 대해, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는지 여부를 확인하는 단계, 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 데이터 분류 그룹의 개수가 선정된 기준 개수를 초과하는지 여부를 확인하는 단계 및 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리하는 단계를 포함할 수 있다.

[0088] 이때, 본 발명의 일실시예에 따르면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리하는 단계는 미리 지정된 서로 다른 복수의 분류 개수의 범위들 별로 미리 지정된 서로 다른 중요도 점수가 대응되어 기록되어 있는 중요도 점수 테이블을 저장하여 유지하는 단계, 미리 지정된 서로 다른 복수의 점수 범위들 별로 미리 지정된 서로 다른 반복 수행 횟수가 대응되어 기록되어 있는 수행 횟수 테이블을 저장하여 유지하는 단계, 상기 복수의 데이터 분류 그룹들 중 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인된 데이터 분류 그룹의 개수가 상기 선정된 기준 개수를 초과하는 것으로 확인되면, 상기 분류 적합도 측정 값이 상기 선정된 기준 값을 초과하는 것으로 확인되는 적어도 하나의 제1 데이터 분류 그룹을 확인한 후 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해, 상기 중요도 점수 테이블을 참조하여 상기 적어도 하나의 제1 데이터 분류 그룹 각각에서 분류된 데이터의 총 분류 개수에 따른 분류 개수의 범위에 대응하는 중요도 점수를 확인하는 단계, 상기 적어도 하나의 제1 데이터 분류 그룹 각각에 대해 확인된 중요도 점수를 합산하여 합계 점수를 연산하는 단계, 상기 합계 점수가 연산되면, 상기 수행 횟수 테이블로부터 상기 연산된 합계 점수에 따른 점수 범위에 대응하는 제1 반복 수행 횟수를 추출하는 단계 및 상기 제1 반복 수행 횟수가 추출되면, 상기 추가 학습 훈련 지시 메시지의 출력을 처리함과 동시에 상기 제1 반복 수행 횟수를 화면 상에 출력하는 단계를 포함할 수 있다.

[0089] 이상, 도 2를 참조하여 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법에 대해 설명하였다. 여기서, 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법은 도 1을 이용하여 설명한 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치(110)의 동작에 대한 구성과 대응될 수 있으므로, 이에 대한 보다 상세한 설명은 생략하기로 한다.

[0090] 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의 동작 방법은 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다.

[0091] 또한, 본 발명의 일실시예에 따른 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치의

동작 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 본 발명을 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), 및 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.

[0092] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

[0093] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

부호의 설명

[0094] 110: 데이터 마이닝을 위한 기계 학습 훈련에 있어서의 분류 적합도 측정 장치

111: 연산부 112: 측정 값 출력부

113: 메시지 출력부 114: 충돌 개수 카운트부

115: 분류 개수 확인부 116: 측정 값 연산부

117: 기준 값 초과 확인부 118: 기준 개수 초과 확인부

119: 메시지 출력 처리부 120: 중요도 점수 테이블 유지부

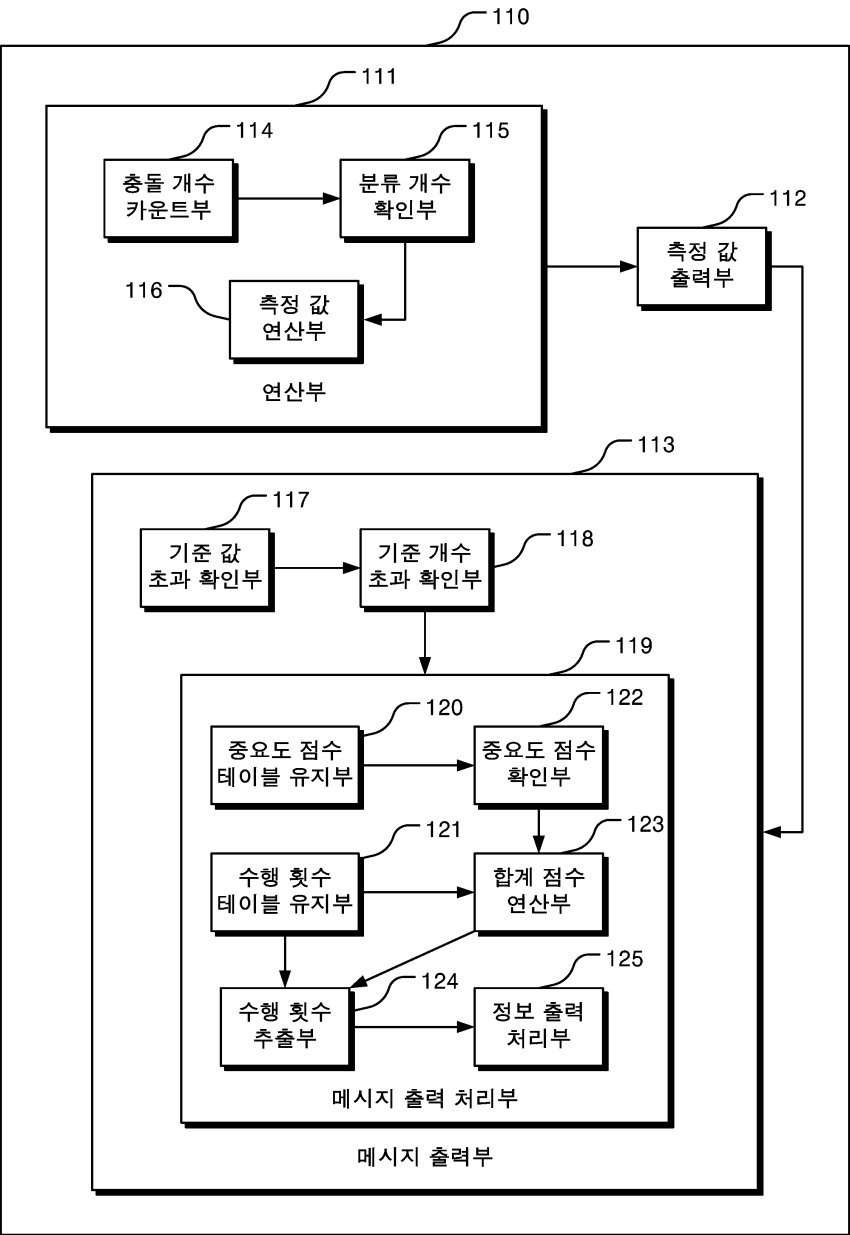
121: 수행 횟수 테이블 유지부 122: 중요도 점수 확인부

123: 합계 점수 연산부 124: 수행 횟수 추출부

125: 정보 출력 처리부

도면

도면1



도면2

