청구범위

[청구항 1] 데이터 분석 프레임워크의 신규 사용자에 대한 진단용 문제 세트를 구성하는 방법에 있어서,
복수의 문제를 포함하는 문제 데이터베이스를 구성하고, 상기 문제에 대한 사용자의 풀이 결과 데이터를 수집하고, 상기 풀이 결과 데이터를 상기 데이터 분석 프레임워크에 적용하여 상기 문제 및/또는 사용자의 모델링 벡터를 계산하는 a 단계;
상기 문제 데이터베이스에서 상기 진단용 문제 세트 구성을 위한 후보 문제를 적어도 하나 이상 추출하는 b 단계;
상기 후보 문제에 대한 풀이 결과 데이터가 존재하는 사용자 및 상기 사용자의 풀이 결과 데이터가 존재하는 다른 문제를 확인하는 c 단계;
상기 후보 문제에 대한 상기 사용자의 풀이 결과 데이터만 상기 데이터 분석 프레임워크에 적용하여 가상 사용자 모델링 벡터를 계산하는 d 단계;
상기 가상 사용자 모델링 벡터를 적용하여 상기 다른 문제의 가상 정답률을 계산하는 e 단계;
상기 가상 정답률과 상기 사용자의 상기 다른 문제의 실제 풀이 결과 데이터를 비교하고, 비교한 결과를 상기 사용자의 수에 따라 평균화하여 상기 후보 문제의 예측률을 계산하는 f 단계를 포함하는 것을 특징으로 하는 문제 세트 구성 방법.

[청구항 2] 제 1항에 있어서,
상기 예측률이 임계값 이내인 후보 문제들을 상기 진단용 문제 세트로 구성하는 단계를 포함하는 것을 특징으로 하는 문제 세트 구성 방법.

[청구항 3] 자율학습 (unsupervised learning)기반의 데이터 분석 프레임워크를 통한 분석 결과를 해석하는 방법에 있어서,
복수의 문제를 포함하는 문제 데이터베이스를 구성하고, 상기 문제에 대한 사용자의 풀이 결과 데이터를 수집하고, 상기 풀이 결과 데이터를 상기 데이터 분석 프레임워크에 적용하여 상기 문제 및/또는 사용자에 대한 클러스터를 적어도 하나 이상 형성하는 a 단계;
상기 클러스터에서 적어도 하나 이상의 제 1 데이터를 랜덤하게 추출하고, 상기 제 1 데이터를 해석하기 위한 제 1 레이블을 선정하는 b 단계;
상기 클러스터에 포함된 데이터 중, 상기 제 1 데이터와 임계값 이내의 유사도를 가지는 데이터에 상기 제 1 레이블을 부여하는 c 단계;
상기 제 1 데이터와 임계값 이외의 유사도를 가지는 데이터 중, 적어도 하나 이상의 제 2 데이터를 랜덤하게 추출하고, 상기 제 2 데이터를 해석하기 위한 제 2 레이블을 선정하는 d 단계;

상기 클러스터에 포함된 데이터 중, 상기 제 2 데이터와 임계값 이내의 유사도를 가지는 데이터에 상기 제 2 레이블을 부여하는 e 단계;
상기 제 1 레이블 및 상기 제 2 레이블을 이용하여 상기 클러스터를 해석하는 f 단계를 포함하는 것을 특징으로 하는 해석 방법.

[청구항 4] 제 3항에 있어서,
특정 과목의 학습 요소를 트리 구조로 나열하여 상기 과목의 학습 요소에 대한 메타데이터 세트를 생성하는 단계;
상기 학습 요소를 분석 그룹 단위로 분류하여 상기 메타데이터의 인덱싱 정보를 생성하는 단계;
상기 메타데이터의 인덱싱 정보를 상기 제 1 레이블 및 상기 제 2 레이블로 활용하는 단계를 포함하는 것을 특징으로 하는 해석 방법.