

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2020년 4월 23일 (23.04.2020)



(10) 국제공개번호
WO 2020/080826 A1

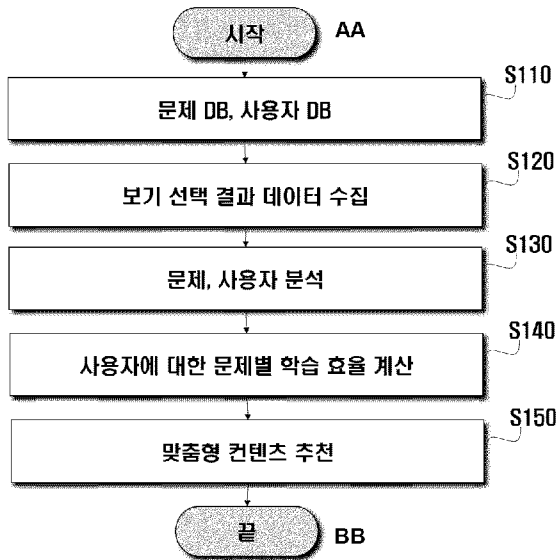
- (51) 국제특허분류: G06Q 50/10 (2012.01) G06N 20/00 (2019.01)
G06Q 50/20 (2012.01)
- (21) 국제출원번호: PCT/KR2019/013590
- (22) 국제출원일: 2019년 10월 16일 (16.10.2019)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2018-0123240 2018년 10월 16일 (16.10.2018)KR
- (71) 출원인: 주식회사 퀴이드 (RIIID INC.) [KR/KR]; 06167 서울시 강남구 테헤란로87길 29, 10층 (삼성동, 애플타워), Seoul (KR).
- (72) 발명자: 차영민 (CHA, Yeong Min); 13258 경기도 성남시 중원구 시민로 66번길 21 (중앙동, 중앙동힐스테이트), Gyeonggi-do (KR). 신동민 (SHIN, Dong Min); 08758 서울시 관악구 봉천로 466, 909호 (봉천동, 해주타워), Seoul (KR). 이재민 (LEE, Jae Min); 06735 서울시 서초구 강남대로39길 15-10, 1120호(서초동), Seoul (KR).

이용구 (LEE, Young Ku); 02196 서울시 중랑구 사가정로71길 11, 101동 1604호 (면목동, 용마2차한신아파트), Seoul (KR). 허재위 (HEO, Jae We); 08745 서울시 관악구 봉천로 49길 12-6, 305 (봉천동), Seoul (KR).

- (74) 대리인: 특허법인도담 (DODAM IP LAW FIRM); 13494 경기도 성남시 분당구 판교역로 231, 에스동 909호(삼평동, 에이치스퀘어), Gyeonggi-do (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE,

(54) Title: MACHINE LEARNING METHOD, DEVICE, AND COMPUTER PROGRAM FOR PROVIDING PERSONALIZED EDUCATIONAL CONTENTS ON BASIS OF LEARNING EFFICIENCY

(54) 발명의 명칭: 학습 효율을 기반으로 개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 방법, 장치 및 컴퓨터 프로그램



- S110 ... Question DB and user DB
- S120 ... Collect example selection result data
- S130 ... Analyze questions and users
- S140 ... Calculate learning efficiency of users for each question
- S150 ... Recommend customized contents
- AA ... Start
- BB ... End

(57) Abstract: The present invention relates to a method by which a service server provides user-customized learning content, the method comprising: a step (a) of constructing a question database including one or more multiple-choice questions containing one or more examples about a specific subject, providing the questions to user devices, and collecting, from the user devices, users' example selection data about the questions; a step (b) of estimating probabilities of correct answers to the questions for each of the users by using the example selection data of each of the users; and a step of calculating, for each question, probability change rates of correct answers to all of the questions, included in the question database, of the users on the assumption that certain users have selected respective examples for certain questions, and arranging the questions included in the question database in descending order of the change rates so as to recommend same to the users.

WO 2020/080826 A1

LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

- 국제조사보고서와 함께 (조약 제21조(3))
- 청구범위 보정 기한 만료 전의 공개이며, 보정서를 접수하는 경우 그에 관하여 별도 공개함 (규칙 48.2(h))

(57) 요약서: 본 발명은 서비스 서버에서, 사용자 맞춤형 학습 콘텐츠를 제공하는 방법에 대한 것으로, 특정 과목에 대해 적어도 하나 이상의 보기를 포함하는 객관식 문제를 적어도 하나 이상 포함하는 문제 데이터베이스를 구성하고, 상기 문제를 사용자 디바이스에 제공하고, 상기 사용자 디바이스로부터 상기 문제에 대한 사용자의 보기 선택 데이터를 수집하는 a 단계; 상기 사용자 각각의 상기 보기 선택 데이터를 이용하여, 상기 사용자 각각에 대해 상기 문제에 대한 정답 확률을 추정하는 b 단계; 임의의 사용자가 임의의 문제에 대해 각각의 보기를 선택한 것으로 가정하여, 상기 문제 데이터베이스에 포함된 문제 전체의 상기 사용자에 대한 정답 확률의 변경율을 각각의 문제에 대해 계산하고, 상기 문제 데이터베이스에 포함된 문제를, 상기 변경율이 높은 순서대로 정렬하여 상기 사용자에게 추천하는 단계를 포함하는 것을 특징으로 한다.

명세서

발명의 명칭: 학습 효율을 기반으로 개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 방법, 장치 및 컴퓨터 프로그램 기술분야

- [1] 본 발명은 데이터를 기반으로, 사용자 맞춤형 콘텐츠를 제공하는 방법에 대한 것이다. 보다 구체적으로 본 발명은 수집된 문제 풀이 결과 데이터를 기계 학습 프레임워크에 적용하여 사용자의 문제에 대한 학습 효율을 추정하고 이를 기반으로 교육 콘텐츠를 추천하는 방법에 대한 것이다.

배경기술

- [2] 지금까지 교육 콘텐츠는 일반적으로 패키지로 제공되어 왔다. 예를 들어 종이에 기록되는 문제집은 권당 최소 700문제가 수록되어 있으며, 온라인 또는 오프라인 강의 역시 1-2 시간 단위로 최소 한달간 공부할 양을 묶어서 한번에 판매된다.
- [3] 그러나 교육을 받는 학생들 입장에서는 개별적으로 취약한 단원과 취약한 문제 유형이 모두 상이하기 때문에 패키지 형태보다는 개인 맞춤형 콘텐츠에 대한 니즈가 존재한다. 제한된 시간에 가장 효율적인 학습을 하기 위해서는 학습효율이 가장 높은 문제를 골라서 학습하는 것이 문제집의 7백 문제 전체를 푸는 것보다 훨씬 효율적이기 때문이다.
- [4] 그러나 피교육자인 학생들 스스로 자신에게 적합한 문제를 골라내는 것은 매우 어렵다. 나아가 학원, 출판사 등 종래의 교육 업계에서도 주관적 경험과 직관에 의존하여 학생 및 문제들을 분석하기 때문에 개별 학생들에게 최적화된 문제를 제공하는 것은 쉽지 않다.
- [5] 이와 같이 종래의 교육 환경에서는 피교육자가 가장 효율적으로 학습 결과를 낼 수 있는 개인 맞춤형 콘텐츠를 제공하는 것이 쉽지 않으며, 학생들은 일률적으로 제공되는 교육 콘텐츠에 대해 성취감과 흥미를 금방 잃게 되는 문제가 발생한다.

발명의 상세한 설명

기술적 과제

- [6] 본 발명은 상기와 같은 문제를 해결하는 것을 목적으로 한다. 보다 구체적으로 본 발명은, 대량의 문제 풀이 결과 데이터를 기반으로, 사용자 및/또는 문제를 분석하고, 이를 기반으로 각각의 사용자에 대한 문제별 학습 효율을 계산하여, 학습 효율에 따라 학습 콘텐츠를 추천하는 방법에 대한 것이다.

과제 해결 수단

- [7] 본 발명의 실시예를 따르는 서비스 서버에서, 사용자 맞춤형 학습 콘텐츠를 제공하는 방법은, 특정 과목에 대해 적어도 하나 이상의 보기를 포함하는 객관식 문제를 적어도 하나 이상 포함하는 문제 데이터베이스를 구성하고, 상기 문제를

사용자 디바이스에 제공하고, 상기 사용자 디바이스로부터 상기 문제에 대한 사용자의 보기 선택 데이터를 수집하는 a 단계; 상기 사용자 각각의 상기 보기 선택 데이터를 이용하여, 상기 사용자 각각에 대해 상기 문제에 대한 정답 확률을 추정하는 b 단계; 임의의 사용자가 임의의 문제에 대해 각각의 보기를 선택한 것으로 가정하여, 상기 문제 데이터베이스에 포함된 문제 전체의 상기 사용자에게 대한 정답 확률의 변경율을 각각의 문제에 대해 계산하고, 상기 문제 데이터베이스에 포함된 문제를, 상기 변경률이 높은 순서대로 정렬하여 상기 사용자에게 추천하는 단계를 포함하는 것을 특징으로 한다.

발명의 효과

- [8] 본 발명의 실시예를 따르면, 사용자 및/또는 문제 분석 결과를 토대로, 사용자에게 최적화된 학습 콘텐츠를 제공할 수 있어, 사용자의 학습 효과를 극대화시킬 수 있는 효과가 있다.

도면의 간단한 설명

- [9] 도 1은 본 발명의 실시예를 따라 사용자 맞춤형 콘텐츠를 추천하는 과정을 설명하기 위한 순서도
- [10] 도 2는 본 발명의 실시예를 따라 사용자의 문제별 학습 효율을 계산하는 과정을 설명하기 위한 순서도
- [11] 도 3은 본 발명의 다른 실시예를 따라 사용자의 문제별 학습 효율을 계산하는 과정을 설명하기 위한 순서도

발명의 실시를 위한 형태

- [12] 본 발명은 이하에 기재되는 실시예들의 설명 내용에 한정되는 것은 아니며, 본 발명의 기술적 요지를 벗어나지 않는 범위 내에서 다양한 변형이 가해질 수 있음은 자명하다. 그리고 실시예를 설명함에 있어서 본 발명이 속하는 기술 분야에 널리 알려져 있고 본 발명의 기술적 요지와 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다.
- [13] 한편, 첨부된 도면에서 동일한 구성요소는 동일한 부호로 표현된다. 그리고 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시될 수도 있다. 이는 본 발명의 요지와 관련이 없는 불필요한 설명을 생략함으로써 본 발명의 요지를 명확히 설명하기 위함이다.

[14]

- [15] 최근 IT 디바이스의 보급이 확대되면서, 사용자 분석을 위한 데이터 수집이 용이해지고 있다. 사용자 데이터를 충분히 수집할 수 있으면, 사용자의 분석이 보다 정밀해지고 해당 사용자에게 가장 적합한 형태의 콘텐츠를 제공할 수 있다.

- [16] 이러한 흐름과 함께 특히 교육 업계에서 사용자 맞춤형 교육 콘텐츠 제공에 대한 니즈가 높다.

- [17] 간단한 예를 들어, 어떤 사용자가 영어 과목에서 “동사의 시제”에 대한 이해도가 떨어지는 경우, “동사의 시제”에 대한 개념을 포함하고 있는 문제를

추천할 수 있으면 학습 효율은 보다 높아질 것이다. 그런데 이와 같이 사용자 맞춤형 교육 콘텐츠를 제공하기 위해서는 각각의 콘텐츠 및 사용자 개인에 대한 정밀한 분석이 필요하다.

- [18] 종래에는 콘텐츠와 사용자를 분석하기 위해 해당 과목의 개념들을 전문가에 의해 수작업으로 정의하고 해당 과목에 대한 각 문제가 어떤 개념을 포함하고 있는지 전문가가 개별적으로 판단하여 태깅하는 방식을 따랐다. 이후 각 사용자가 특정 개념에 대해 태깅된 문제들을 풀어본 결과 정보를 토대로 학습자의 실력을 분석하는 것이다.
- [19] 사용자가 문제를 풀이하는 과정을 살펴보면, 문제 출제자의 의도는 한 두개일 수 있지만 사용자가 해당 문제를 틀리는 이유는 수십 가지일 수 있다. 예를 들어 어떤 문제의 유형이 "가정법 과거"이며, 출제자의 의도는 "조동사의 과거형과 가정법을 구분할 줄 아는가"에 대한 경우를 가정할 수 있다.
- [20] 이 경우, 사용자가 해당 문제를 틀리는 이유는 출제의도와 동일하게 "조동사의 과거형과 가정법을 구분할 줄 몰라서"일 수 있지만, 지문의 특정 단어를 몰랐기 때문일 수도 있다. 이 경우, 해당 사용자에게 가장 효율적인 문제는 상기 단어를 학습하기 위한 문제일 것이다.
- [21] 그러나 전문가가 모든 경우의 수를 고려하여 해당 사용자를 분석하는 것은 불가능하기 때문에, 일반적으로 사용자는 문제 유형 태그 및 출제자 의도 태그에 따라 "가정법 과거", "조동사 과거", "가정법" 등의 개념에 대한 학습이 필요한 것으로 판단하고 그와 관련된 문제가 추천될 것이다.
- [22] 이와 같이, 전문가가 미리 정의한 태그 정보를 기반으로 사용자를 분석하는 방법은 태그 정보가 사람의 주관에 의존하며, 인간의 행동에 대한 다양한 경우의 수가 포괄되지 않는 문제점이 있다. 사람의 주관이 개입되지 않고 수학적으로 생성된 태그 정보들이 수학적으로 문제에 부여되는 것이 아니기 때문에 결과 데이터에 대한 신뢰도가 높을 수 없는 문제가 있었다.
- [23] 따라서 본 발명의 실시예를 따르는 데이터 분석 서버는 학습 데이터 분석에 머신 러닝 프레임워크를 적용하여 데이터 처리 과정의 사람의 개입을 배제하는 것을 목적으로 한다.
- [24] 이에 따르면, 해당 과목의 개념을 사전에 별도로 정의하지 않고, (즉, 문제에 포함된 개념에 대한 태그 정보를 생성하지 않고) 사용자의 문제 풀이 결과 로그를 기반으로 사용자 및 문제를 분석할 수 있다. 예를 들어 데이터 분석 서버는 사용자의 문제 풀이 결과 로그를 수집하고, 사용자와 문제로 구성된 다차원 공간을 구성하고, 사용자가 문제를 맞았는지 틀렸는지를 기준으로 상기 다차원 공간에 값을 부여하여, 각각의 사용자 및 문제에 대한 벡터를 계산하는 방식으로 사용자 및/또는 문제를 모델링할 수 있다.
- [25] 이 경우, 사용자 모델은 전체 문제에 대한 사용자 각각의 특성에 대한 정보를 포함할 수 있으며, 문제 모델은 전체 사용자에게 대한 문제 각각의 특성에 대한 정보를 포함할 수 있다. 나아가 문제 모델은 특정 문제의 해당 과목을 구성하는

개념에 대한 포함도, 사용자 모델은 특정 사용자의 해당 과목을 구성하는 개념에 대한 이해도로 표현될 수 있다.

- [26] 나아가 본 발명의 실시예를 따르면, 상기 문제 모델과 사용자 모델을 이용하여, 임의의 사용자가 임의의 문제에 대해 정답 보기를 선택할 확률, 즉, 사용자의 문제별 정답 확률을 계산할 수 있다. 이때 본 발명에서 상기 사용자 벡터, 상기 문제 벡터들이 어떤 속성, 또는 피처를 포함하고 있는지는 제한하여 해석될 수 없음을 유의해야 한다.
- [27] 예를 들어 본 발명의 실시예를 따르면, 상기 사용자 벡터는 상기 사용자가 임의의 개념에 대해 이해하고 있는 정도, 즉 개념의 이해도를 포함할 수 있다. 나아가 상기 문제 벡터는 상기 문제가 어떤 개념들로 구성되어 있는지, 즉 개념 구성도를 포함할 수 있다.
- [28] 나아가 본 발명의 실시예를 따르면, 특정 사용자에 대한 문제 데이터베이스에 포함된 문제들 각각에 대한 학습효율을 계산할 수 있어, 학습효율이 높은 순서대로 사용자에게 문제가 추천될 수 있다.
- [29] 예를 들어 본 발명의 제 1 실시예를 따르면, 정답 확률이 낮은 문제가 학습 효율이 높을 것으로 가정할 수 있다. 이에 따르면 특정 사용자가 이미 이해하고 있기 때문에 맞힐 문제는 배제하고, 틀릴 문제들만 골라서 추천하게 될 것이다. 상기 실시예를 따르면 사용자는 틀릴 것이 확실한 문제부터 학습할 수 있다.
- [30] 본 발명의 제 2 실시예를 따르면, 사용자가 어떤 문제를 맞추게 되면, 해당 문제를 포함해 새롭게 학습된 모델이 예측하는 문제 전체의 상기 사용자에게 대한 정답 확률의 증가율이 높은 문제가 학습 효율이 높을 것으로 가정할 수 있다. 이에 따르면, 사용자에게 문제 데이터베이스 전체를 보다 잘 풀 수 있게 되는 문제가 추천될 것이다. 이에 따르면, 사용자는 전체 문제의 정답 확률을 높일 수 있는 문제부터 학습하게 될 것이다.
- [31] 나아가 본 발명의 제 3 실시예를 따르면, 사용자의 실제 시험 점수를 높일 수 있는 문제가 학습 효율이 높을 것으로 가정할 수 있다. 이에 따르면, 문제 데이터베이스 외부에서 출제된, 동일한 과목에 대한 실제 시험 점수를 사용자 별로 추정하고, 상기 실제 시험 점수를 높일 수 있는 문제가 추천될 것이다. 이에 따르면 사용자는 실제 시험 점수를 높일 수 있는 문제부터 학습하게 될 것이다.
- [32] 도 1은 본 발명의 실시예를 따라 사용자 맞춤형 콘텐츠를 추천하는 과정을 설명하기 위한 순서도이다.
- [33] 단계 110 및 단계 130는 본 발명의 실시예를 따르는 데이터 분석 시스템에서 학습 데이터를 수집하고 문제와 사용자를 모델링하는 단계이다.
- [34] 단계 110에서 데이터 분석 서버는, 문제 데이터베이스 및 사용자 데이터베이스를 생성할 수 있다.
- [35]
- [36] 예를 들어 데이터 분석 서버는 시중에 나와 있는 각종 문제들을 수집하여 문제 데이터베이스를 생성할 수 있다. 상기 문제는 듣기 평가 문제를 포함하고,

- 텍스트, 이미지, 오디오, 및/또는 동영상 형태일 수 있다. 나아가 데이터 분석 서버는 사용자 식별 정보를 포함하는 사용자 데이터베이스를 생성할 수 있다.
- [37] 단계 120에서 데이터 분석 서버는 상기 문제 데이터베이스를 사용자 디바이스에 제공하고, 사용자가 해당 문제를 풀이한 결과 데이터를 수집할 수 있다.
- [38] 이때 데이터 분석 서버는 수집된 문제 풀이 결과 데이터를 사용자, 문제, 결과에 대한 리스트 형태로 구성할 수 있다. 예를 들어 $Y(u, i)$ 는 사용자 u 가 문제 i 를 푼 결과를 의미하며, 정답인 경우 1, 오답인 경우 0의 값이 부여될 수 있다.
- [39] 나아가 본 발명의 실시예를 따르는 데이터 분석 서버는 사용자와 문제로 구성된 다차원 공간을 구성하고, 사용자가 문제를 맞았는지 틀렸는지를 기준으로 상기 다차원 공간에 값을 부여하여, 각각의 사용자 및 문제에 대한 벡터를 계산할 수 있다. 이때 상기 사용자 벡터와 문제 벡터가 포함하는 피처는 특정되지 않는 것으로 해석되어야 한다. (단계 130)
- [40] 예를 들어, 본 발명의 실시예를 따르는 데이터 분석 서버는, Y 가 충분한 크기이면, 개념을 별도로 정의하지 않고, Y 로부터 사용자의 개념 이해도 L , 문제의 개념 구성도 R 를 추정할 수 있다. 이때 본 발명의 실시예를 따르면 상기 L 의 각 원소는 0에서 1 사이의 값을 가지며, R 의 각 행의 원소들의 합은 1이 될 수 있다.
- [41] 예를 들어 L 의 1번째 행의 값이 $[0, 0, 1, 0.5, 1]$ 인 경우, 이는 제 1 사용자가 제 1 개념 및 제 2 개념은 전혀 이해하지 못하고, 제 3 개념 및 제 5 개념은 완벽히 이해하고, 그리고 제 4 개념은 절반만큼 이해한 것으로 해석될 것이다.
- [42] 예를 들어 R 의 1번째 행의 값이 $[0, 0.2, 0.5, 0.3, 0]$ 이라 할 때, 제 1 문제는 제 2 개념이 20% 정도 포함, 제 3 개념이 50% 정도 포함, 제 4 개념이 30% 정도 포함된 것으로 해석될 것이다.
- [43] 한편 본 발명의 실시예를 따르는 데이터 분석 시스템은, L 과 R 이 충분한 신뢰도로 추정되면 사용자와 문제 사이의 상호 관계를 낮은 계수 행렬을 통해 수학적으로 연결할 수 있다.
- [44]
- [45] *예를 들어 서비스 서버는 분석 대상 전체 사용자 수가 n , 분석 대상 전체 문제 개수가 m 인 경우, 해당 과목을 구성하는 미지의 개념의 개수를 r 로 상정하면, 사용자의 개념별 이해도 행렬 L 을 n by r 행렬로 정의하고, 문제의 개념별 포함 정도에 대한 행렬 R 을 m by r 행렬로 정의할 수 있다. 이 경우, L 을 R 의 전치행렬 R^T 과 연결하면, 개념 또는 개념의 개수를 별도로 정의하지 않고 사용자와 문제의 상호 관계를 분석할 수 있다.
- [46] 이를 적용하면 L 의 1번째 행의 값이 $[0, 0, 1, 0.5, 1]$, R 의 1번째 행의 값이 $[0, 0.2, 0.5, 0.3, 0]$ 인 위의 예에서, 제 1 사용자가 제 1 문제에 대한 이해도 $X(1,1)$ 은 $X(1,1) = 0.5 - 0.5 * 0.3 = 0.65$ 와 같이 계산될 수 있다. 즉, 제 1 사용자는 제 1 문제를 65퍼센트 이해하는 것으로 추정될 수 있다.

- [47] 그러나 사용자가 특정 문제에 대한 이해도와 특정 문제를 맞출 확률은 동일하다고 할 수 없다. 위의 예에서 제 1 사용자가 제 1 문제를 65 퍼센트 이해한다면, 제 1 문제를 실제로 풀었을 때 정답일 확률은 어느 정도인 것인가
- [48] 이를 위해 본 발명은 심리학, 인지과학, 교육학 등에 사용되는 문항 반응 이론 (Item Response Theory)의 방법론을 도입하였다. 문항 반응 이론은 평가 문항들에 대한 응답을 이용하여, 피험자의 특성, 예를 들어 인지 능력, 물리적 능력, 기술, 지식, 태도, 인격 특징이나 평가 문항의 난이도를 측정하기 위한 검사 이론이다.
- [49] 상기 이론을 적용하여 충분히 큰 데이터를 사용하여 실험한 결과 문제 이해도 X 와 문제를 맞출 확률 P 은 리니어하지 않은 것으로 분석된다. 문제 이해도 X 와 문제를 맞출 확률 P 를 충분히 큰 데이터를 사용하여 실험한 결과, 2차원 그래프를 얻을 수 있다. 이를 설명할 수 있는 함수를 종래의 수학적 방법을 이용하여 계산하면, 사용자가 문제를 맞출 확률 P 를 추정하기 위한 함수 Φ 를 아래의 수학식 1와 같이 도출할 수 있다. 다시 말하면, 문제 정답 확률 P 는 함수 Φ 에 문제 이해도 X 를 적용하면 계산될 수 있다.

[50] [수식1]

$$\Phi(x)=0.25+\frac{0.75}{1+e^{-10(x-0.5)}}$$

- [51] 제 1 사용자의 제 1 문제에 대한 이해도가 65퍼센트인 위의 예에서, 제 1 사용자가 제 1 문제를 맞출 확률은 $P(1,1)=\Phi(X(1,1))=0.8632$ 와 같이 계산되어 86%에 해당한다. 즉, 제 1 사용자는 제 2, 4 개념은 전혀 이해하지 못하고, 제 3 개념을 완전히 이해하며, 제 1 문제는 제 2 개념이 20%, 제 3 개념을 50%, 제 4 개념이 30%로 구성된 문제인데, 위의 공식에 의하면 제 1 사용자가 제 1 문제 1을 풀게 되면 86퍼센트의 확률로 정답일 것으로 추정될 수 있다.
- [52] 한편, 사용자의 개념 이해도 L , 문제의 개념 구성도 R 를 추정하기 위해 아래에 예시된 방법을 따를 수 있다. 이는 확률적 그래디언트 알고리즘을 기반으로 사용자 i 의 개념 이해도 L_i 및 문제 j 의 개념 구성도 R_j 를 추정하기 위한 방법이며, 이는 예시에 불과하여 본 발명은 이에 한정되지 않음을 유의해야 한다.
- [53] 본 발명의 실시예를 따르는 데이터 분석 시스템은 사용자 i 의 개념 이해도 L_i , 문제 j 의 개념 구성도 R_j 의 초기 값을 임의의 값으로 부여할 수 있다. 예를 들어 데이터 분석 시스템은 사용자 i 의 임의의 개념 이해도 L_i 의 초기값 및 문제 j 의 임의의 개념 구성도 R_j 의 초기값을 0 내지 1 사이의 임의의 값으로 부여할 수 있다.
- [54] 나아가 문제 풀이 결과 데이터를 각각의 사용자 i , 각각의 문제 j 에 대해 정답인 경우 1, 오답인 경우 0으로 val를 설정한 리스트를 생성할 수 있다. 예를 들어 사용자별 문제 풀이 결과 데이터 세트를 유저 i 가 문제 j 를 풀었을 때 정답인 경우

(i, j, 1), 오답인 경우 (i, j, 0)인 형태의 리스트로 변환할 수 있다.

[55] 이후 데이터 분석 시스템은 문제 풀이 결과 리스트의 데이터를 랜덤하게 하나씩 로딩할 수 있다. 보다 구체적으로, 데이터 분석 시스템은 해당 시점의 사용자 i의 개념 이해도 개념 이해도 L_i , 문제 j의 개념 구성도 R_j 를 적용하여 특정 사용자 i가 특정 문제 j를 맞출 확률 $P_{(i,j)}$ 의 임시값 tmp 를 계산할 수 있다.

[56] 이는 해당 시점의 사용자 i의 개념 이해도 L_i , 문제 j의 개념 구성도 R_j 를 적용하여 계산한 확률 $P_{(i,j)}$, 즉 특정 사용자 i가 특정 문제 j를 맞출 확률과 해당 사용자 i의 해당 문제 j에 대한 실제 문제 풀이 결과값인 val (정답인 경우 1, 오답인 경우 0) 사이의 차이를 이용하여 L 및 R을 업데이트하기 위한 것이다.

[57] 선택한 데이터가 (i, j, val)인 경우, 데이터 분석 시스템은 해당 시점의 L_i (사용자 i의 개념 이해도), R_j (문제 j의 개념 구성도)를 이용하여 사용자 i가 문제 j를 맞출 확률의 임시값

$$tmp = \Phi(x(i,j)) = \Phi(L_i \cdot R_j^T)$$

를 계산할 수 있다.

[58] 예를 들어 선택한 데이터가 (1, 1, 0)인 경우, 제 1 사용자가 제 1 문제를 풀어서 틀렸다는 의미이다. 그런데 해당 시점의 L_1, R_1 를 이용하여 계산한 결과, 제 1 사용자가 제 1 문제를 푼다면 86퍼센트의 확률로 맞을 것이라는, 즉 $P=0.86$ 인 결과가 나왔다면, L_1, R_1 은 데이터 (1, 1, 0)를 반영하여 업데이트 될 필요가 있다.

[59] 이를 위해 본 발명의 실시예를 따르는 데이터 분석 시스템은 사용자 i가 문제 j를 맞출 확률의 임시값 tmp 및 사용자 i가 문제 j를 실제 풀이한 데이터인 val의 차이를 이용하여 선택한 데이터 (i, j, val)에 대한 근사 그래디언트 grad 를 계산할 수 있다. 이는 아래의 수학적 식 2를 이용하여 계산될 수 있다.

[60] [수식2]

$$grad = 10(val - tmp) / \{ tmp \times (1 + e^{-10(L_i R_j^T - 0.5)}) \}$$

[61] 나아가, grad와 함께 해당 사용자가 푼 문제의 개수 u_i , 해당 문제가 풀린 횟수 v_j 를 반영하여 아래와 같은 수학적 식 3에 따라 사용자 i의 개념 이해도 L_i , 문제 j의 개념 구성도 R_j 을 업데이트할 수 있다.

[62] [수식3]

$$L_i : (1 - mu \times step / u_i) \times L_i - step \times grad \times R_j$$

$$R_j : (1 - mu \times step / v_j) \times R_j - step \times grad \times L_i$$

[63] 이때, 상기 mu는 상기 수식의 성능을 좌우하는 파라미터로 종래의 기술에 따라 실험에 의해 선택되며, 상기 step은 상기 수식이 새로운 데이터를 받아들이는

- 속도에 대한 파라미터로 종래의 기술에 따라 실험에 의해 선택할 수 있다.
- [64] 본 발명의 실시예를 따르면, 전체 데이터에 대한 이와 같은 과정을 반복하고, 모든 데이터가 반영되면 다시 랜덤하게 리스트를 섞어서 상기 과정을 충분히 반복하면 L, R 을 신뢰도 있게 추정할 수 있다.
- [65] 그런데 본 발명의 실시예를 따르는 데이터 분석 프레임워크에서 사용자와 문제를 모델링 벡터로 표현하는 이유는 궁극적으로는 특정 사용자가 특정 문제를 맞출지 틀릴지를 정밀하게 예측하기 위한 것이다.
- [66] 사용자의 문제에 대한 정답률은 전술한 바와 같이 사용자의 정오답 결과를 수집하여 분석하는 방식으로 추정할 수 있지만, 보기의 선택 확률을 이용하여 추정할 수도 있다.
- [67] 예를 들어 제 1 사용자가 특정 문제에 대한 보기 선택 확률이 (0.1, 0.2, 0, 0.7)인 경우, 사용자는 높은 확률로 보기 4번을 선택할 것이고, 해당 문제의 정답이 4번인 경우, 제 1 사용자는 그 문제를 맞을 확률이 높을 것으로 예상할 수 있다.
- [68] 한편, 문제의 난이도는 전체 사용자들의 정오답 결과를 수집하여 분석하는 방식으로 추정할 수도 있지만, 보기의 선택 확률을 이용하여 추정할 수도 있다.
- [69] 예를 들어 4지 선다형으로 구성된 특정 문제에 대해 사용자들의 보기 선택 확률이 (0.5, 0.1, 0.3, 0.6)인 경우, 사용자들이 제 2 문제에서 보기 1번을 고를 확률과 보기 4번을 고를 확률이 비슷하므로, 제 2 문제는 난이도가 높은 것으로 분류할 수 있다.
- [70] 이를 위해 데이터 분석 서버는 전체 문제와 전체 사용자에 대한 풀이 결과 데이터를 사용자가 선택한 보기 요소를 포함하여 수집할 수 있다.
- [71] 보다 구체적으로 데이터 분석 서버는 문제 데이터베이스에 속하는 전체 문제들에 대해 사용자가 해당 문제에서 어떤 보기를 선택했는지, 보기 선택 결과를 수집하는 방식으로 풀이 결과 로그를 수집할 수 있으며, 추가적으로 해당 문제에 대한 정오답 결과를 수집할 수도 있다. 이때 데이터분석 서버는 사용자, 문제, 해당 사용자가 해당 문제에서 선택한 보기에 대한 리스트를 구성할 수 있다.
- [72] 이후 데이터분석 서버는 하나의 문제를 보기 단위로 확장하여 데이터 분석 처리를 수행할 수 있다.
- [73] 예를 들어 문제 i 가 사지선다형 문제인 경우, 문제 i 는 $(i, 1) (i, 2) (i, 3) (i, 4)$ 의 변수 4개로 확장되며, 각 문제-보기의 각 사용자의 선택 여부가 값으로 부여될 수 있다.
- [74] 예를 들어 $E(i, j)$ 는 문제 i 의 보기 j 를 의미하며, $Y'(u, E)$ 는 사용자 u 가 문제 i 의 보기 j 를 선택했는지를 의미하며, 본 발명의 실시예를 따르면 선택한 경우 1, 선택하지 않은 경우 0의 값이 부여될 수 있다.
- [75] 나아가 데이터 분석 서버는 사용자 및 문제-보기를 변수로 다차원 공간을 구성하고, 사용자가 해당 문제-보기를 선택했는지를 기준으로 상기 다차원 공간에 값을 부여하여, 각각의 사용자 및 문제-보기에 대한 벡터를 계산할 수

있다.

[76] 이후 데이터 분석 서버는 상기 사용자 벡터와 상기 문제-보기 벡터를 이용하여 임의의 사용자가 임의의 문제-보기를 선택할 확률, 즉 선택률을 추정할 수 있다.

[77] 이때 상기 사용자 벡터와 상기 문제-보기 벡터에 다양한 알고리즘을 적용하여 상기 선택률을 추정할 수 있으며, 본 발명을 해석함에 있어 선택률을 계산하기 위한 알고리즘은 제한되지 않는다.

[78] 예를 들어, 본 발명의 실시예를 따르면, 아래의 수학식 4와 같은 시그모이드 함수를 적용하면, 사용자의 문제-보기 선택률을 추정할 수 있다. (x는 문제-보기 벡터, 는 사용자 벡터)

[79] [수식4]

$$h_{\theta}(x) = 1 / (1 + e^{ -\theta * T * X})$$

[80] 나아가 본 발명의 실시예를 따르는 데이터 분석 서버는 사용자의 보기 선택률을 이용하여 문제의 정답률을 추정할 수 있다.

[81] 그런데 예를 들어 4지 선다형으로 구성된 특정 문제에 대해 특정 사용자의 보기 선택 확률이 (0.5, 0.1, 0.3, 0.6)이며, 정답 보기는 1번인 경우, 상기 사용자가 해당 문제를 맞출 확률은 얼마인지가 문제된다. 즉, 해당 문제에 대한 복수의 보기 선택률을 이용하여 해당 문제의 정답률을 추정하는 방법을 고려할 수 있다.

[82] 본 발명의 실시예를 따라 보기 선택률을 문제 정답률로 환원하는 간단한 방식으로는 전체 보기의 선택률 대비 정답 보기의 선택률을 비교하는 방법을 고려할 수 있다.

[83] 이 경우 앞의 예에서 해당 사용자의 해당 문제에 대한 정답률은 0.5 / (0.5+0.1+0.3+0.6)로 계산할 것이다. 그러나 사용자는 문제를 풀이할 때는 보기 단위로 구분하여 해당 문제를 이해하는 것이 아니라, 전체 보기에 대한 구성 및 문제의 출제 의도를 포함하여 문제 단위로 이해하기 때문에 보기 선택률과 정답률은 단순 연결될 수 없다.

[84] 따라서 본 발명의 실시예를 따르면, 해당 문제의 전체 보기 선택률을 평균화하고 정답 보기의 평균화된 선택률을 전체 보기의 선택률에 적용하는 방식으로 보기 선택률로부터 해당 문제의 정답률을 추정할 수 있다.

[85] 앞의 예에서 보기의 선택 확률이 (0.5, 0.1, 0.3, 0.6)인 경우, 이를 전체 보기에 대해 평균화하면 각각의 보기 선택률은 (0.33, 0.07, 0.20, 0.41)로 스케일이 변경될 수 있다. 정답 보기가 1번인 경우, 보기 1의 평균화된 선택률은 0.33으로, 해당 사용자의 해당 문제에 대한 정답률은 33%로 추정될 수 있다.

[86] 나아가 본 발명의 실시예를 따르는 서비스 서버는 사용자의 문제-보기 선택 확률을 이용하여 문제의 정답률을 추정할 수 있으며, 이를 통해 특정 개념에 대한 사용자의 이해도를 추정할 수 있다.

[87] 나아가 데이터 분석 서버는 문제 모델 및 사용자 모델을 기반으로 해당

- 사용자의 특정 문제에 대한 학습 효율을 계산할 수 있다. (단계 140)
- [88] 본 발명의 실시예를 따르는 데이터 분석 서버는 전술한 바와 같이, 문제의 보기 단위의 모델과 사용자 모델을 이용하여 각각의 사용자에 대해 문제의 보기 단위의 선택 확률 및 해당 사용자의 문제 단위의 정답 확률을 계산할 수 있다.
- [89] 본 발명의 제 1 실시예를 따르면, 정답 확률이 낮은 문제가 학습 효율이 높을 것으로 가정할 수 있다. 이를 따르면, 서비스 서버는 정답 확률이 높은 문제는 사용자가 이미 이해하고 있는 것으로 보고, 사용자가 맞힐 문제는 배제하고, 틀릴 문제들만 골라서 추천할 것이다.
- [90] 보다 구체적으로, 서비스 서버는 문제 데이터베이스 세트 중 사용자가 이미 풀어본 문제를 제외하고, 나머지 문제들을 특정 사용자에 대한 정답률이 낮은 순서대로 정렬하여 해당 사용자에 대한 추천 문제 리스트를 작성하고 이를 제공할 수 있다. (단계 150)
- [91] 예를 들어 사용자-문제 정답률 P의 1번 행의 값이 [0.3, 0.4, 0.1, 0.9, 0.7]인 경우 사용자 1이 문제 1, 2, 3, 4, 5를 맞출 확률이 각각 30%, 40%, 10%, 90%, 70%인 것으로 해석될 것이다. 서비스 서버는 문제 3, 1, 2, 5, 4의 순서로 우선순위를 부여한 문제 추천 리스트를 사용자에게 제공할 수 있다. 이때 문제 5번을 사용자 1이 이미 풀었다면 서비스 서버는 문제 5번은 제외하고 문제 추천 리스트를 작성할 수 있다.
- [92] 상기 실시예를 따르면 서비스 서버는 문제 데이터베이스에 포함된 문제를 사용자 정답 확률이 낮은 순서로 정렬하여 사용자에게 추천할 수 있다. 사용자는 틀릴 것이 확실한 문제부터 학습하게 될 것이다.
- [93] 한편 데이터 분석 서버는 사용자 벡터, 문제-보기 벡터, 보기 선택률, 문제의 정답률, 사용자의 개념 이해도, 문제의 개념 구성도 중 적어도 하나 이상을 이용하여 사용자 맞춤형 콘텐츠를 추천할 수 있다. (단계 150)
- [94] 예를 들어 서비스 서버는 특정 개념에 대한 이해도가 낮은 사용자에게 해당 개념에 대한 심화 강좌를 추천할 수 있다. 또 다른 예로 서비스 서버는 복수의 개념을 모두 포함하는 문제에 대한 정답률이 낮은 사용자에게 상기 복수의 개념을 포함하도록 구성된 문제를 추천할 수 있다.
- [95] 한편 본 발명의 제 2 실시예를 따르면, 사용자가 어떤 문제를 맞히게 되면, 문제 전체의 상기 사용자에 대한 정답 확률의 변경률이 높은 문제가 학습 효율이 높을 것으로 가정할 수 있다. 이에 따르면, 사용자에게 문제 데이터베이스 전체를 보다 잘 풀 수 있게 되는 문제가 추천될 것이며, 사용자는 전체 문제의 정답 확률을 높일 수 있는 문제부터 학습하게 될 것이다.
- [96] 상기 제 2 실시예에 따라 사용자의 문제별 학습 효율을 계산하는 과정에 대한 설명은 첨부된 도 2에 대한 설명에서 후술된다. 도 2는 본 발명의 실시예를 따라 사용자의 문제별 학습 효율을 계산하는 과정을 설명하기 위한 순서도이다.
- [97] 임의의 시점에 사용자의 문제별 정답 확률 또는 임의의 문제의 보기별 선택 확률은 해당 사용자의 풀이 결과 데이터가 추가로 수집되면 이를 반영하여

업데이트될 것이다. 이를 고려하면, 특정 문제에 대해 해당 사용자가 문제 데이터베이스에 포함된 임의의 문제에 대해 정답 보기를 선택한 것으로 가정하면, 다른 문제들에 대한 해당 사용자의 정답 확률이 조금씩 변경될 것이다. (단계 210 내지 단계 230)

[98] 이후 서비스 서버는 전체 문제 데이터베이스에 포함된 문제들 각각에 대해 해당 사용자가 정답 보기를 선택한 것으로 가정하고, 다른 문제들 각각에 대해 정답 확률의 증가값을 계산하고 정답 확률의 변경값의 전체 합계 또는 정답 확률의 변경 값의 평균을 해당 문제의 학습 효율로 계산할 수 있다. (단계 240)

[99] 보다 구체적으로, 본 발명의 실시예를 따르는 데이터 분석 서버는, 특정 사용자의 임의의 문제의 보기별 선택 확률에 상기 사용자가 해당 보기를 선택한 경우, 다른 문제의 정답률 변경값의 평균값을 적용하면 해당 문제의 학습 효율을 계산할 수 있다.

[100] 예를 들어 사용자 A의 특정 문제 a의 제 1 보기에 대한 선택률이 a1, 제 2 보기에 대한 선택률이 a2, 제 3 보기에 대한 선택률이 a3, 제 4 보기에 대한 선택률이 a4인 경우를 고려할 수 있다.

[101] 그때, 사용자 A가 해당 문제에 대해 1 보기를 선택한 것으로 가정하면, 사용자 A의 모델 및 전체 문제들의 모델은 사용자 A의 제 1 보기 선택 이벤트가 적용하여 다소 변경될 것이며, 변경된 모델에 따라 각 문제들의 정답률이 다소 변경될 것이다.

[102] 이에 따라 데이터 분석 서버는 사용자 A의 제 1 보기 선택에 대한 가상 이벤트를 적용하여 전체 문제들의 정답률 변경 평균값 AVC_a1을 계산할 수 있다.

[103] 동일한 방법으로 데이터 분석 서버는 사용자 A가 제 2 보기를 선택한 것으로 가정하여 AVC_a2를 계산하고, 제 3 보기를 선택한 것으로 가정하여 AVC_a3를 계산하고, 제 4 보기를 선택한 것으로 가정하여 AVC_a4를 계산할 수 있다.

[104] 이후 데이터 분석 서버는 사용자 A의 문제 a를 통한 학습효율 E(A,a)를 아래의 수식을 통해 계산할 수 있다.

[105] [수식5]

$$E(A,a) = a1 * AVC_a1 + a2 * AVC_a2 + a3 * AVC_a3 + a4 * AVC_a4$$

[106] 나아가 데이터 분석 서버는 전체 문제에 대해 E(A)를 계산하고, E(A)를 해당 문제의 학습 효율로 활용할 수 있다.

[107] 이후 서비스 서버는 학습 효율이 높은 순서대로 전체 문제를 정렬할 수 있다. 즉, 해당 문제를 사용자가 학습하여 맞게 되면, 다른 문제들 전체의 정답 확률의 변경되는데, 상기 정답 확률의 변경값이 높으면, 사용자는 문제 데이터베이스 전체를 보다 잘 풀 수 있게 되기 때문에, 학습 효율이 높다고 보아 사용자에게 추천될 것이다. (단계 250)

[108] 한편 본 발명의 제 3 실시예를 따르면, 사용자의 실제 시험 점수를 높일 수 있는

문제가 학습 효율이 높을 것으로 가정할 수 있다. 이에 따르면, 문제 데이터베이스 외부에서 출제된, 동일한 과목에 대한 실제 시험 점수를 사용자 별로 추정하고, 상기 실제 시험 점수를 높일 수 있는 문제가 추천될 것이다. 이에 따르면 사용자는 실제 시험 점수를 높일 수 있는 문제부터 학습하게 될 것이다.

- [109] 상기 제 3 실시예에 따라 사용자의 문제별 학습 효율을 계산하는 과정에 대한 설명은 첨부된 도 3에 대한 설명에서 후술된다. 도 3는 본 발명의 제 3 실시예를 따라 실제 시험 점수를 올릴 수 있는 문제별 학습 효율을 계산하는 과정을 설명하기 위한 순서도이다.
- [110] 상기 실시예를 따르면, 문제 모델과 사용자 모델을 이용하여, 실제 시험 점수를 추정하는 과정이 필수적으로 요구된다. 문제 데이터베이스 외부에서 출제되는 실제 시험 점수를 어떻게 추정할 것인가?
- [111] 종래 특정 시험에 대한 피시험자의 예상 점수는 전문가들의 노하우에 따라 추정되는 것이 일반적이었다. 예를 들어 대학 수학 능력 시험의 경우, 전문가들의 노하우에 따라 실제 대학 수학 능력 시험과 유사하게 모의고사를 구성하고, 학생들이 모의고사를 풀어본 결과를 토대로 대학 수학 능력 시험의 예상 점수가 예측되게 된다.
- [112] 그러나 이와 같은 방법은 전문가들의 주관적 경험과 직관에 의존하기 때문에 실제 시험 결과와 크게 차이가 나는 경우가 적지 않다. 전문가들은 각자의 경험과 직관을 통해 모의고사를 설계하기 때문에, 이와 같은 모의고사를 통해서 학생들의 실력을 제대로 예측하거나 실제 시험 점수를 정확하게 예측할 수 없다.
- [113] 나아가 수험생들이 모의고사를 푸는 행위 자체는 능률적인 공부라고 보기 어렵다. 모의고사는 실제 시험과 문제 유형이 유사한지를 기준으로 구성되기 때문에 응시자의 실력과는 무관하게 출제된다. 즉, 모의고사는 시험점수를 추정하여 전체 학생들 중에서 자신의 위치를 확인하는데 목적이 있을 뿐, 응시자의 학습을 위해 구성된 문제가 아니며, 학습 효율을 높여주는 방식은 더더욱 아닐 것이다.
- [114] 이를 해결하기 위한 본 발명의 실시예를 따르면, 실제 시험 점수 데이터가 존재하는 사용자 그룹의 실제 시험 점수 데이터를 이용하여, 문제 데이터베이스 외부에서 출제되는 실제 시험 점수를 추정하고, 실제 시험 점수 추정값이 상승하기 위해 필요한 문제들이 추천될 수 있다.
- [115] 보다 구체적으로, 본 발명의 실시예를 따르면, 상기 사용자 그룹의 실제 시험 점수 데이터를 이용하여, (단계 320) 상기 사용자 그룹의 문제 데이터베이스에 포함된 모든 문제에 대한 정답 확률의 평균과 실제 시험 점수와의 관계를 나타내는 모델링 함수를 각 유저별로 계산할 수 있다. 즉, 실제 시험 점수 데이터를 이용하여, 전체문제의 정답 확률의 평균을 실제 점수로 변환하는 모델링 함수가 계산될 수 있다. (상기 모델링 함수를 계산하는 보다 구체적인 방법은 후술됨)

- [116] 상기 모델링 함수가 계산되면, 상기 사용자 그룹에 속하지 않는 사용자, 즉, 실제 시험 점수 데이터가 존재하지 않는 사용자의 상기 문제 데이터베이스에 포함된 모든 문제들에 대한 정답 확률의 평균을 실제 시험 점수로 변환할 수 있다. (단계 330)
- [117] 보다 구체적으로, 외부 시험 점수 데이터가 존재하지 않는 다른 사용자들에 대해, 단계 310에서 사용자 모델이 충분한 신뢰도로 계산된 경우, 상기 사용자 모델을 이용하면 문제 데이터베이스에 포함된 문제들에 대한 정답 확률을 예측할 수 있으며, 예측된 정답 확률을 상기 모델링 함수에 적용하면 정답 확률을 외부 시험 점수로 변환할 수 있다.
- [118] 이후 임의의 사용자가 임의의 문제에 대해 임의의 보기를 선택한 것으로 가정하면, (단계 340) 사용자 모델 및 문제 모델은 해당 사용자의 보기 선택 이벤트가 적용되어 다소 변경될 것이며, 변경된 모델에 따라 각 문제들의 정답률이 다소 변경되고, 상기 모델링 함수를 적용하여 변환한 실제 시험 점수 예측값 역시 변경될 것이다. (단계 350)
- [119] 즉, 서비스 서버는 전체 문제 데이터베이스에 포함된 문제들 각각에 대해 해당 사용자가 각각의 보기를 선택한 것으로 가정하고, 실제 시험 점수 예측의 변경값을 계산하고, 실제 시험 점수 예측의 변경값의 전체 합계 또는 변경 값의 평균을 해당 문제의 학습 효율로 계산할 수 있다.
- [120] 한편, 본 발명의 다른 발명의 실시예를 따르면, 상기 사용자 그룹, 즉 실제 시험 점수 데이터가 존재하는 사용자 그룹의 실제 시험 점수 데이터를 이용하여, (단계 320) 상기 사용자 그룹에 대한 추정 점수가 상기 실제 시험 점수 데이터와 유사하도록 모의고사 문제 세트를 문제 데이터베이스로부터 구성할 수 있다. (상기 모의고사 문제 세트를 구성하는 보다 구체적인 방법은 후술됨)
- [121] 상기 모의고사 문제 세트는, 전문가에 의해 실제 시험과 동일한 문제 유형으로 구성되는 종래의 모의고사와 달리, 추정 점수가 실제 시험 점수와 유사한지를 기준으로 구성되는 특징이 있다.
- [122] 상기 모의고사 문제 세트가 구성되면, 상기 사용자 그룹에 속하지 않는 사용자의 사용자 모델을 이용하여, 상기 모의고사 문제 세트의 추정 점수를 계산할 수 있다. (단계 330) 상기 모의고사 문제 세트는 추정 점수가 실제 시험 점수와 유사하도록 구성되었기 때문에, 실제 시험 점수 데이터가 존재하지 않는 사용자들의 상기 모의고사 문제 세트에 대한 추정 점수를 실제 시험 점수의 예측값으로 갈음할 수 있다.
- [123] 이후 임의의 사용자가 임의의 문제에 대해 각각의 보기를 선택한 것으로 가정하면, (단계 340) 사용자 모델 및 문제 모델은 해당 사용자의 보기 선택 이벤트가 적용되어 다소 변경될 것이며, 변경된 모델에 따라 각 문제들의 정답률이 다소 변경되고, 상기 모의고사 문제 세트에 대한 추정 점수 역시 변경될 것이다. (단계 350)
- [124] 즉, 서비스 서버는 전체 문제 데이터베이스에 포함된 문제들 각각에 대해 해당

사용자가 각각의 보기를 선택한 것으로 가정하고, 실제 시험 점수 예측 (즉, 모의고사 문제 세트에 대한 추정 점수)의 변경값을 계산하고, 실제 시험 점수 예측의 변경값의 전체 합계 또는 변경 값의 평균을 해당 문제의 학습 효율로 계산할 수 있다.

- [125] 이후 서비스 서버는 실제 시험 점수 예측값의 상승률이 높은 순서로 문제 데이터베이스를 정렬하여 사용자에게 추천할 수 있다. (단계 360)
- [126] 한편, 문제 데이터베이스에 포함된 문제 각각에 대한 정답 확률과 외부 시험 점수와의 관계를 나타내는 모델링 함수를 계산하는 과정은 다음과 같다.
- [127] 전술한 바와 같이, 데이터 분석 서버는 전체 사용자 중 외부 시험 점수 데이터가 존재하는 사용자 그룹을 확인할 수 있다. 이는 상기 사용자 그룹의 외부 시험 점수 데이터와 각 사용자 모델, 문제-보기 모델로부터 구할 수 있는 정답 확률과 관계를 이용하여, 외부 시험 점수 예측 효율이 높은 문제 세트를 구성하기 위한 것이다.
- [128] 나아가 본 발명의 실시예를 따르는 데이터 분석 서버는, 문제 세트의 성능, 즉 문제 세트가 외부 시험 점수의 예측 효율이 높은지 여부를 검사하거나 또는 상기 문제 세트의 점수를 외부 시험 점수로 변환하는 모델링 함수의 성능을 검사하기 위해 상기 사용자 그룹의 일부를 테스트 그룹으로 분할할 수 있다.
- [129] 예를 들어 데이터베이스에 속하는 전체 사용자 중, 외부 시험 데이터가 존재하는 사용자가 {A, B, C, D, E, F, G, H} 인 경우, 데이터 분석 서버는 {A, B, C, D, E}를 트레이닝 세트로 구분하고, {F, G, H}를 테스트 세트로 구분할 수 있다.
- [130] 보다 구체적으로 데이터 분석 서버는 외부 시험 점수 데이터가 존재하는 사용자 그룹을 임의로 두 그룹으로 나누어, 하나의 그룹은 데이터 분석 프레임워크에 상기 문제 세트 및 시험 점수 변환 모델링 함수를 구성하기 위한 데이터를 제공하는 트레이닝 세트로 그룹핑하고, 다른 하나의 그룹은 상기 문제 세트 및 모델링 함수의 성능을 검사하기 위한 데이터를 제공하는 테스트 세트로 그룹핑할 수 있다.
- [131] 다시 말하면, 외부 시험 점수 데이터가 존재하는 사용자 그룹 중, 상기 트레이닝 세트에 속하는 사용자들의 데이터는 문제 세트 및 시험 점수 변환 모델링 함수 구성에 사용되며, 상기 테스트 세트에 속하는 사용자들의 데이터는 상기 문제 세트 및 시험 점수 변환 모델링 함수의 성능을 검사하는데 사용될 수 있다.
- [132] 이후, 데이터 분석 서버는 상기 사용자 그룹의 외부 시험 점수 데이터 및 사용자 모델링 벡터를 이용하여 문제 데이터베이스에서 외부 시험 점수 예측 효율이 높은 문제 세트를 구성할 수 있다.
- [133] 특히 본 발명의 실시예를 따르는 문제 세트는 종래의 모의고사처럼 외부 시험과 유사한지를 기준으로 구성되는 것이 아니라, 외부 시험 점수에 대한 예측 효율이 높은지를 기준으로 구성되어 문제 개수 및/또는 문제 유형 등이 외부 시험에 종속될 필요가 없는 특징이 있다.
- [134] 보다 구체적으로 특히 데이터 분석 서버는, 문제 데이터베이스에 포함되는

각각의 문제에 대해, 외부 시험 점수 데이터가 존재하는 사용자 그룹의 모델링 벡터를 이용하여 계산한 정답 확률과 상기 사용자 그룹의 외부 점수 데이터의 관계를 다차원 평면에 표현할 수 있다.

- [135] 예를 들어 문제 데이터베이스에 포함된 문제가 100문제인 경우, 데이터 분석 서버는 x축을 문제 개수만큼 100개를 생성하고, y 축을 외부 시험 점수로 구성하여 다차원 평면을 생성할 수 있다.
- [136] 예를 들어 외부 시험 점수 데이터가 존재하는 사용자 중, 트레이닝 세트에 속하는 사용자가 A, B, C, D, E, 5명인 경우, 데이터 분석 서버는 사용자 A에 대해 미리 계산한 모델링 벡터를 이용하여 제 1 문제의 정답 확률을 계산할 수 있으며, 이를 제 1 x축 (x_1)의 x 값으로 설정하고, 사용자 A의 외부 시험 점수를 y 값으로 설정하여 사용자 A의 제 1 평면의 좌표를 결정할 수 있다.
- [137] 데이터 분석 서버는 동일한 방식으로 사용자 B 내지 사용자 E의 제 1 평면의 좌표를 결정하여 제 1 문제 (x_1)의 외부 시험 점수와의 관계를 다차원 공간 중 x_1 평면에 표현할 수 있다.
- [138] 나아가 데이터 분석 서버는 동일한 방식으로 제 2 문제 내지 제 100 문제에 대한 평면을 형성하고 각각의 문제 평면에 대한 사용자 A 내지 E의 (x, y) 좌표를 결정하여 제 2 문제 내지 제 100 문제의 외부 시험 점수와의 관계를 다차원 공간에 표현할 수 있다.
- [139] 이 경우, 상기 사용자 그룹의 개별 문제에 대한 추정 정답 확률과 외부 시험 점수는 실질적으로 2차원 평면에 표현될 것이며, 본 발명의 실시예를 따르는 데이터 분석 서버는 2차원 평면에 표현된 값들의 관계를 수학적으로 표현하는 모델링 함수로 변경할 수 있다.
- [140] 예를 들어, 본 발명의 실시예를 따르는 데이터 분석 서버는, 트레이닝 세트에 속하는 사용자 A, B, C, D, E의 외부 시험 점수 및 사용자 A, B, C, D, E의 사용자 모델링 벡터를 이용하여 계산한 제 1 문제에 대한 정답 확률의 관계를 수학적으로 표현하는 모델링 함수 $f_1(x_1)$ 를 계산할 수 있다.
- [141] 나아가 제 2 문제 내지 제 100 문제에 대한 데이터를 같은 방식으로 처리하여 모델링 함수 $f_1(x_2)$ 내지 $f_1(x_{100})$ 구할 수 있다. 이때, $f_1(x_1)$ 내지 $f_1(x_{100})$ 은 동일한 기준에 따라 결정되어야 한다.
- [142] 즉, 문제 세트를 구성하기 위해 계산되는 모델링 함수의 형태는 데이터베이스에 속하는 전체 문제에 대해 동일한 형태로 결정되는 것이 적절하다. 예를 들어 제 1 문제의 모델링 함수 $f_1(x_1)$ 가 $ax+b$ 형태의 1차원 함수로 표현되면, 제 2 문제 내지 제 100 문제에 대한 모델링 함수 $f_1(x_2)$ 내지 $f_1(x_{100})$ 의 모델링 함수 역시 1차원 함수 형태로 표현되는 것이 적절할 것이다.
- [143] 문제 세트는 문제 데이터베이스에 속하는 문제 중, 외부 시험 점수 예측 효율이 높은 문제들로 구성되어 하기 때문에 비교 기준이 동일해야 하기 때문이다.
- [144] 나아가 본 발명의 실시예를 따르는 모델링 함수는 전체 문제에 대해 동일한 기준이 적용되면, 그 형태는 제한되지 않는다. 즉, 모델링 함수는 트레이닝

세트의 데이터를 표현하기에 최적화된 형태로 작성되면 족하며, 종래 적용되는 다양한 알고리즘에 따라 계산될 수 있다.

[145] 각 문제에 대한 모델링 함수가 전체 문제에 대해 동일한 형태로 결정되면, 본 발명의 실시예를 따르는 데이터 분석 서버는 해당 문제의 모델링 함수 및 테스트 세트에 속하는 사용자의 외부 시험 점수의 편차를 이용하여 외부 시험 점수 예측 효율이 높은 문제세트를 구성할 수 있다.

[146] 예를 들어 데이터 분석 서버는 테스트 세트에 속하는 사용자 F, G, H의 사용자 모델링 벡터를 이용하여 계산한 제 1 문제 (x_1)의 정답 확률을 제 1 문제의 모델링 함수 $f(x_1)$ 에 적용한 값과 사용자 F, G, H의 외부 시험 점수의 차이값의 평균을 계산하고, 상기 평균값 (e)이 작으면 외부 시험 점수 예측 효율이 높은 것으로 판단할 수 있다.

[147] 상기 평균값 e 을 구하는 수학적식은 아래와 같으며, e 는 문제 세트의 외부 시험 점수 추정 효율을 의미할 수 있다. 보다 구체적으로, e 값이 작을수록 해당 문제 세트의 외부 시험 점수 추정 효율이 높은 것으로 해석될 수 있다.

[148] [수식6]

$$e_1 = (| y_F - f(x_{1F}) | + | y_G - f(x_{1G}) | + | y_H - f(x_{1HH}) |) / 3$$

[149] 이후 데이터 분석 서버는 동일한 기준에 따라 e_2 내지 e_{100} 을 계산할 수 있으며, 가장 작은 e 를 가지는 문제가 외부 시험 점수 예측 효율이 가장 높은 것으로 판단할 수 있다.

[150] 예를 들어 제 4 문제에 대한 e_4 가 e_1, e_2, e_3, e_5 내지 e_{100} 보다 가장 작은 경우, 제 4 문제가 외부 시험 점수 예측 효율이 가장 높은 것으로 판단할 수 있기 때문에 데이터 분석 서버는 제 4 문제를 문제 세트에 최초로 포함시킬 수 있다.

[151] 이후 데이터 분석 서버는 제 4 문제와 함께 분석되는 것을 전제로 제 1, 제 2, 제 3 및 제 5 내지 제 100 문제 중에서 외부 시험 점수 예측 효율이 높은 문제를 추출할 수 있다.

[152] 이를 위해 본 발명의 실시예를 따르는 데이터 분석 서버는, 제 4 문제에 대한 평면을 고정하고, 나머지 제 1, 제 2, 제 3 및 제 5 내지 제 100 문제에 대한 평면을 형성하여, 트레이닝 세트에 속하는 사용자 A, B, C, D, E의 외부 시험 점수 및 사용자 A, B, C, D, E의 사용자 모델링 벡터를 이용하여 계산한 나머지 제 1, 제 2, 제 3 및 제 5 내지 제 100 문제에 대한 정답 확률의 관계를 수학적으로 표현하는 모델링 함수 $f(x_4, x_1), f(x_4, x_2), f(x_4, x_3), f(x_4, x_5)$ 내지 $f(x_4, x_{100})$ 를 다시 한번 계산할 수 있다. (제 2 모델링 함수)

[153] 상기 제 2 모델링 함수값은 제 4 문제와 해당 문제에 대한 외부 시험 점수 추정값으로 해석될 수 있다.

[154] 이때, 위의 2차원 모델링 함수 $f(x_4, x_1), f(x_4, x_2), f(x_4, x_3), f(x_4, x_5)$ 내지 $f(x_4, x_{100})$ 들은 제 4 문제만을 가지고 모델링 함수를 구성할 때와 동일한 기준에 따라 결정되어야 할 것이다.

[155] 즉, 예를 들어 제 4 문제와 함께 제공되는 제 1 문제의 모델링 함수 $f(x_4, x_1)$ 는 cx_4+dx_1+e 형태로, 기존의 $f(x_4)$ 의 ax_4+b 와 같이 선형 모델의 기준을 그대로 따르면서 x_1 의 변수를 추가해 차원만 늘린 2차원 함수로 표현되어야 한다. 제 4 문제와 함께 제공되는 제 2 문제, 제 3 문제, 제 5 문제 내지 제 100 문제에 대한 모델링 함수 $f_2(x_4, x_2)$, $f_2(x_4, x_3)$, $f_2(x_4, x_5)$ 내지 $f_2(x_4, x_{100})$ 의 모델링 함수 역시 기존의 $f(x_4)$ 와 동일한 선형 모델이며, 각 문제 x 가 변수로 추가된 2차원 함수 형태로 표현되는 것이 적절할 것이다.

[156] 제 4 문제가 문제 세트에 포함되는 것을 전제로, 나머지 문제에 대한 2차원 모델링 함수가 결정되면, 본 발명의 실시예를 따르는 데이터 분석 서버는 해당 문제들의 2차원 함수 및 테스트 세트에 속하는 사용자의 외부 시험 점수의 편차를 이용하여 외부 시험 점수 예측 효율이 높은 문제세트를 구성할 수 있다.

[157] 예를 들어 데이터 분석 서버는 테스트 세트에 속하는 사용자 F, G, H의 사용자 모델링 벡터를 이용하여 계산한 제 1 문제 (x_1)의 정답 확률, 제 4 문제 (x_4)의 정답 확률을 2차원 모델링 함수 $f(x_4, x_1)$ 에 적용한 값 (이는 제 4 문제와 제 1 문제의 외부 시험 점수 추정값으로 해석될 것이다)과 사용자 F, G, H의 외부 시험 점수의 차이값의 평균을 계산하고, 상기 평균값이 작으면 외부 시험 점수 예측 효율이 높은 것으로 판단할 수 있다. 이를 수학식으로 표현하면 아래와 같다.

[158] [수식7]

$$e_{4,1} = (| y_F - f(x_{4F}, x_{1F}) | + | y_G - f(x_{4G}, x_{1G}) | + | y_H - f(x_{4H}, x_{1H}) |) / 3$$

[159] 이후 데이터 분석 서버는 동일한 기준에 따라 $e_{4,2}$ 내지 $e_{4,100}$ 을 계산할 수 있으며, 가장 작은 e 를 가지는 문제가 외부 시험 점수 예측 효율이 가장 높은 것으로 판단할 수 있다.

[160] 예를 들어 제 1 문제에 대한 $e_{4,1}$ 가 $e_{4,2}$, $e_{4,3}$, $e_{4,5}$ 내지 $e_{4,100}$ 보다 가장 작은 경우, 제 4 문제와 함께 제공되기에 제 1 문제가 외부 시험 예측 효율이 가장 높은 것으로 판단할 수 있기 때문에 데이터 분석 서버는 제 1 문제를 문제 세트에 포함시킬 수 있다.

[161] 이후 데이터 분석 서버는 제 1, 4 문제와 함께 분석되는 것을 전제로 제 2, 제 3 및 제 5 내지 제 100 문제 중에서 외부 시험 예측 효율이 높은 문제를 전술한 방법과 같은 논리를 따라 추출할 수 있다.

[162] 한편, 이와 같은 방법에 따라 최소한의 e 를 가지는 문제들로 문제 세트를 구성하는 경우, 문제 세트 구성의 완료 시점이 문제될 수 있다. 문제 세트에 포함되는 문제의 개수가 많을수록 추정의 정밀도는 높아질 수 있으나, 문제 세트를 구성하기 위해 개별 문제마다 모델링 함수를 계산하고, 우선 순위를 계산하기 위해 소요되는 리소스 사용량이 증가될 수 있다. 따라서 외부 시험 점수 추정에 최적화된 개수로 문제 세트를 구성하는 것이 효율적이다.

[163] 본 발명의 실시예를 따르면, 테스트 세트에 속하는 사용자들의 모델링 벡터를 이용하여 계산한 문제 세트에 대한 정답 확률을 해당 문제 세트의 모델링 함수에

적용한 값과 테스트 세트에 속하는 사용자들의 외부 시험 점수의 차이값의 평균, 즉, e 를 계산하고, 문제 세트의 업데이트에 따르는 e 의 변경 추이에 따라 문제 세트 구성 종료 여부를 결정할 수 있다.

- [164] 본 발명의 실시예를 따르는 데이터 분석 서버는 문제 세트의 e 의 업데이트 추이를 확인하여 문제 세트 업데이트 종료 여부를 결정할 수 있다. 보다 구체적으로, 문제 세트 업데이트를 종료하는 경우는 크게 두가지를 고려할 수 있다.
- [165] 첫번째는 해당 시점의 문제 세트에 속하는 사용자들의 외부 시험 점수를 충분히 추정할 수 있을 때이다. 즉, 특정 문제 세트만으로 외부 시험 점수 추정을 충분히 수행할 수 있는 경우이다. 이는 해당 시점의 문제 세트의 e 가 임계값 이하인지 여부를 확인하여 판단할 수 있다.
- [166] 두번째는 문제 세트에 다른 문제를 추가한다 하더라도 외부 시험 점수 추정의 효율이 더 이상 높아지지 않는 경우이다. 즉, 문제 추가의 효과가 없는 경우인데, 문제를 추가한다 하더라도 외부 시험의 점수 추정 효율의 변화가 없을 것으로 예상되는 경우이다. 이는 문제 세트의 업데이트에도 불구하고 업데이트 문제 세트의 추정 효율, 즉 e 가 변경되지 않고 임의의 범위 내에서 유지되는지 여부를 확인하여 판단할 수 있다.
- [167] 종합하면, 본 발명의 실시예를 따르는 데이터 분석 서버는 문제 세트의 업데이트를 수행하면서, 문제 세트의 외부 시험 점수 추정 효율, 즉, e 의 변경 추이를 확인하고, 특정 시점의 문제 세트의 추정 효율이 미리 설정된 임계값을 내려가거나, 상기 추정 효율이 임의의 범위 내에서 유지되는 경우 문제 세트의 업데이트를 종료하고 문제 세트를 구성할 수 있다.
- [168] 한편, 본 발명의 실시예를 따르는 모델링 함수는 문제 세트에 포함되는 문제들에 대해 동일한 형태인 것을 전제로 다양한 알고리즘에 따라 결정될 수 있다. 그런데 다양한 알고리즘을 적용하여 생성한 다수의 모델링 함수 중, 외부 시험 점수 추정에 가장 효율적인 함수를 선택하는 문제가 해결 과제로 남을 수 있다.
- [169] 본 발명의 실시예를 따르면, 모델링 함수의 효율성은 e , 즉, 문제 세트의 외부 시험 점수 예측 효율을 이용하여 계산할 수 있다.
- [170] 예를 들어, 트레이닝 세트 $\{A, B, C, D, E\}$ 에 속하는 사용자의 데이터를 이용하여 제 1 기준에 따라 제 1 모델링 함수 f_1 가 추출되고, 상기 제 1 모델링 함수를 적용하여 구성한 문제 세트가 {제 3 문제 (x_3), 제 4 문제 (x_4), 제 5 문제 (x_5)} 경우를 고려할 수 있다.
- [171] 나아가 동일한 트레이닝 세트 $\{A, B, C, D, E\}$ 에 속하는 데이터를 이용하여 제 2 기준에 따라 제 2 모델링 함수 f_2 가 추출되고, 상기 제 2 모델링 함수를 적용하여 구성한 문제 세트가 {제 1 문제 (x_1), 제 3 문제 (x_3), 제 5 문제 (x_5)} 인 경우를 고려할 수 있다.
- [172] 위의 예에서 본 발명의 실시예를 따르는 데이터 분석 서버는 테스트 세트 $\{F,$

G, H}의 데이터를 이용하여 각 문제 세트 및 각 모델링 함수의 e , 즉 문제 세트의 외부 시험 점수 예측 효율을 계산하고 더 작은 e 값을 가지는 모델링 함수를 선택할 수 있다.

- [173] 보다 구체적으로, 먼저 데이터 분석 서버는 테스트 세트에 속하는 사용자 F, G, H의 사용자 모델링 벡터를 이용하여 제 1 문제 세트인 {제 3 문제 (x3), 제 4 문제 (x4), 제 5 문제 (x5)}의 정답 확률을 제 1 모델링 함수 f_1 에 적용한 값과 사용자 F, G, H의 외부 시험 점수의 차이값의 평균, 즉 제 1 모델링 함수에 대한 e_1 을 계산할 수 있다.
- [174] 이후 데이터 분석 서버는 테스트 세트에 속하는 F, G, H의 사용자 모델링 벡터를 이용하여 제 2 문제 {제 1 문제 (x1), 제 3 문제 (x3), 제 5 문제 (x5)}의 정답 확률을 제 2 모델링 함수 f_2 에 적용한 값과 사용자 F, G, H의 외부 시험 점수의 차이값의 평균, 즉 제 2 모델링 함수에 대한 e_2 을 계산할 수 있다.
- [175] 이후 데이터 분석 서버는 더 작은 e 값을 가지는 모델링 함수 및 문제 세트를 선택하여, 사용자의 외부 시험 점수를 추정하는 이후의 절차를 진행할 수 있다.
- [176] 한편, 사용자에 대한 추정 점수가 문제 데이터베이스를 이용하지 않고 출제된 실제 시험 점수 데이터와 유사하도록 모의고사 문제 세트를 구성하는 과정은 다음과 같다.
- [177] 본 발명의 실시예를 따르면 모의고사는 아래와 같은 방법으로 구성할 수 있다.
- [178] 첫번째는 데이터베이스 문제 각각의 전체 사용자에 대한 평균 정답률을 이용하여 전체 사용자에 대한 모의고사의 평균 점수가 임의의 범위에 속하도록 문제 세트를 구성하는 것이다.
- [179] 예를 들어, 어학 능력 시험의 통계를 참고할 때 해당 시험의 전체 응시자의 평균 점수가 67점에서 69점인 경우, 데이터 분석 서버는 모의고사도 평균 점수가 67점에서 69점 범위에 속하도록 문제 세트를 구성할 수 있다.
- [180] 이때 해당 시험의 문제 유형 분포를 고려하여 모의고사의 문제 세트를 구성할 수 있다. 예를 들어 어학 능력 시험의 통계를 참고할 때, 실제 시험에 제 1 유형이 20 퍼센트, 제 2 유형이 30 퍼센트, 제 3 유형이 40 퍼센트, 제 4 유형이 10 퍼센트 내외로 출제되는 경우, 모의고사도 문제 유형 분포가 실제 시험과 유사하도록 구성할 수 있다.
- [181] 이를 위해 본 발명의 실시예를 따르면 문제 유형에 대한 레이블을 미리 생성하여 문제 데이터베이스에 인덱스 정보를 추가할 수 있다.
- [182] 예를 들어 데이터 분석 서버는 임의의 유형으로 분류할 수 있는 문제들의 레이블을 미리 정의하고, 해당 문제 유형을 따르는 문제 모델의 특성을 학습하여 문제들을 클러스터링하고, 클러스터링된 문제 그룹에 문제 유형에 대한 레이블을 부여하는 방식으로 인덱스 정보를 생성할 수 있다.
- [183] 또 다른 예로 데이터 분석 서버는 문제 유형에 대한 레이블을 미리 정의하지 않고 문제들의 모델링 벡터를 이용하여 문제들을 클러스터링하고, 클러스터링된 문제 그룹의 의미를 해석하여 문제 유형에 대한 레이블을

부여하는 방식으로 인덱스 정보를 생성할 수도 있다.

- [184] 본 발명의 실시예를 따라 모의고사를 구성하는 두번째 방법은 해당 시험에 대한 임의의 사용자들의 실제 점수 정보를 이용하는 것이다.
- [185] 예를 들어, 어학 능력 시험에 대한 앞의 예에서, 해당 시험을 응시한 사용자 A, B, C의 실제 점수가 각각 60, 70, 80점인 경우, 미리 계산된 사용자 A, B, C의 정답률을 적용하여 계산한 모의고사 추정 점수가 각각 60, 70, 80점이 되도록 모의고사 문제 세트를 구성할 수 있다.
- [186] 모의고사의 추정된 점수가 실제 점수와 근접하도록 문제 세트를 구성하는 상기 실시예를 따르면 모의고사와 실제 시험의 유사도가 실제 시험을 응시한 사용자의 점수 정보를 이용하여 수학적으로 계산될 수 있다. 따라서 모의고사의 신뢰도, 즉, 모의고사 점수가 실제 시험 점수에 근접한다는 신뢰도를 높일 수 있다.
- [187] 이때 본 발명의 실시예를 따르면 해당 시험의 문제 유형 분포 정보를 적용하여 모의고사 문제 세트를 구성할 수 있으며, 통계적으로 분석되는 다른 정보를 적용할 수도 있다.
- [188] 한편, 데이터 분석 서버는 모의고사 문제 세트를 구성하는 과정에서 문제들의 배점을 조정할 수 있다. 이는 문제 데이터베이스에 속하는 문제들은 별도의 배점 정보가 부여되어 있지 않으나, 실제 시험은 문제들 각각에 상이한 배점이 부여되기 때문이다.
- [189] 일반적으로 실제 시험은 어려운 문제에 높은 배점이 부여되고, 쉬운 문제에 낮은 배점이 부여된다. 이를 해석하면, 문제의 평균 정답률, 문제를 구성하는 개념의 개수, 문제 지문의 길이 등을 고려하여 실제 문제의 배점이 부여되며, 문제 유형에 따라 미리 특정된 배점이 부여될 수도 있다.
- [190] 따라서 본 발명의 실시예를 따르는 데이터 분석 서버는 해당 문제의 평균 정답률, 문제를 구성하는 개념의 개수, 문제 지문의 길이, 및 문제 유형 정보 중 적어도 하나 이상을 반영하여 모의고사 문제 세트를 이루는 각각의 문제들의 배점을 부여할 수 있다.
- [191] 이를 위해 데이터 분석 서버는 문제의 개념에 대한 레이블 생성을 위해 해당 과목의 학습 요소 및/또는 주제를 트리 구조로 나열하여 최소 학습 요소에 대한 메타데이터 세트를 생성하고, 상기 최소 학습 요소를 분석에 적합한 그룹 단위로 분류하여 문제를 구성하는 개념에 대한 인덱스 정보를 생성할 수도 있다.
- [192] 특히 본 발명의 실시예를 따르면, 해당 시험을 실제로 응시한 사용자들의 실제 점수와 해당 사용자들의 모의고사 문제 세트에 대한 추정 점수가 근접하도록 문제 세트를 이루는 각각의 문제들의 배점을 부여할 수도 있다.
- [193] 실제 시험과 유사도가 높은 모의고사 문제 세트가 구성되면, 본 발명의 실시예를 따르는 데이터 분석 서버는 사용자 각각의 모의고사에 대한 예측 점수를 추정할 수 있다. 실제 시험과 모의고사가 유사하다는 것을 전제로 모의고사의 점수를 실제 시험의 점수로 추정하는 것이다.

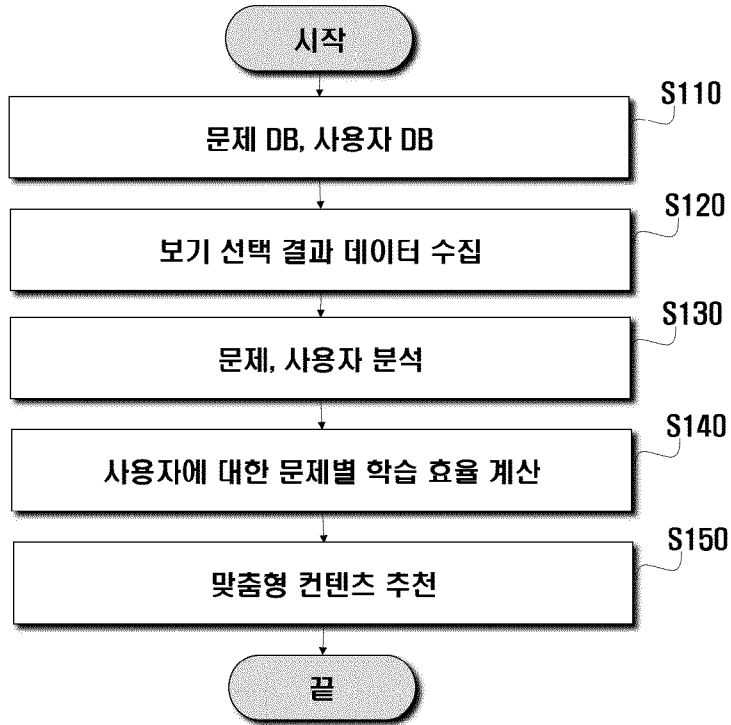
- [194] 특히 본 발명의 실시예를 따르면, 사용자가 모의고사를 직접 풀어볼 필요없이 모의고사의 점수를 높은 신뢰도로 추정할 수 있는 특징이 있다.
- [195] 본 발명의 실시예를 따르는 모의고사는 문제 데이터베이스에 포함된 문제로 구성되며, 데이터베이스에 속하는 각각의 문제에 대한 사용자의 정답률은 전술한 바와 같이 미리 계산되어 있다. 따라서 모의고사를 이루는 모든 문제들에 대한 개별 사용자의 정답률을 이용하여 해당 사용자의 모의고사 예상 점수를 추정할 수 있다.
- [196] 이때 본 발명의 실시예를 따르면 임의의 시험 점수를 추정하기 위한 모의고사 문제 세트를 복수개 구성하고, 특정 사용자의 복수의 모의고사에 대한 추정 점수를 평균화하여 해당 사용자의 실제 시험에 대한 예측 점수를 추정할 수 있다.
- [197]
- [198] 본 명세서와 도면에 게시된 본 발명의 실시 예들은 본 발명의 기술 내용을 쉽게 설명하고 본 발명의 이해를 돕기 위해 특정 예를 제시한 것뿐이며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 게시된 실시 예들 이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다.

청구범위

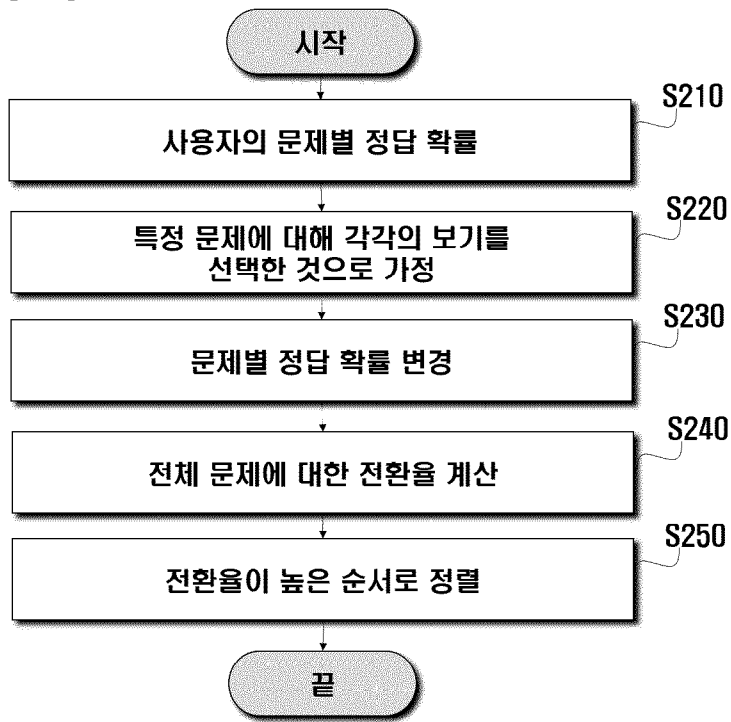
- [청구항 1] 서비스 서버에서, 사용자를 분석하는 방법에 있어서,
 특정 과목에 대해 적어도 하나 이상의 보기를 포함하는 객관식 문제를 적어도 하나 이상 포함하는 문제 데이터베이스를 구성하고, 상기 문제를 사용자 디바이스에 제공하고, 상기 사용자 디바이스로부터 상기 문제에 대한 사용자의 보기 선택 데이터를 수집하는 a 단계;
 상기 사용자 각각의 상기 보기 선택 데이터를 이용하여, 상기 사용자 각각에 대해 상기 문제에 대한 정답 확률을 추정하는 b 단계;
 상기 과목에 대해, 상기 문제 데이터베이스를 이용하지 않고 출제된 외부 시험 점수 데이터가 존재하는 제 1 사용자 그룹을 확인하고, 상기 외부 시험 점수 데이터가 존재하지 않는 제 2 사용자 그룹의 상기 외부 시험 점수를 예측하기 위한 모의고사 문제 세트를 상기 문제 데이터베이스로부터 구성하며, 상기 모의 고사 문제 세트는, 상기 정답 확률을 이용하여 계산한 상기 제 1 사용자 그룹에 대한 추정 점수가 상기 제 1 사용자 그룹의 외부 시험 점수 데이터와 유사하도록 구성되는 c 단계; 및
 상기 모의고사 문제 세트를 상기 제 2 사용자 그룹에게 제공하지 않고, 상기 정답 확률을 이용하여 상기 제 2 사용자의 상기 모의고사 문제 세트의 예측 점수를 추정하고, 상기 모의고사 문제 세트의 예측 점수를 상기 외부 시험에 대한 예측 점수로 추정하는 d 단계를 포함하는 사용자 분석 방법.
- [청구항 2] 제 1 항에 있어서, 상기 d 단계 이후에,
 임의의 사용자가 상기 문제 데이터베이스에 포함된 임의의 문제에 대해 임의의 보기를 선택한 것으로 가정하여, 상기 사용자에 대해 추정된 상기 외부 시험에 대한 예측 점수의 상승률을 각각의 문제에 대해 계산하고, 상기 문제 데이터베이스에 포함된 문제를, 상기 상승률이 높은 순서대로 정렬하여 상기 사용자에게 추천하는 단계를 포함하는 것을 특징으로 하는 사용자 분석 방법.
- [청구항 3] 제 2 항에 있어서, 상기 d 단계는,
 상기 제 1 사용자 그룹의 외부 시험 점수 데이터를 이용하여, 상기 제 1 사용자 그룹의 상기 데이터베이스에 포함된 각각의 문제에 대한 정답 확률과 상기 외부 시험 점수와의 관계를 나타내는 모델링 함수를 계산하는 단계; 및
 상기 모델링 함수를 이용하여 상기 제 1 사용자 그룹에 속하지 않는 상기 제 2 사용자 그룹의 상기 문제 데이터베이스에 포함된 각각의 문제에 대한 정답 확률을 외부 시험 점수로 변환하는 단계를 포함하는 사용자 분석 방법.

- [청구항 4] 제 1항에 있어서, 상기 c 단계는
상기 모의고사 문제 세트를 구성하는 각각의 문제들의 전체 사용자에게
대한 평균 정답률을 이용하여, 상기 전체 사용자에게 대한 상기 모의고사
문제 세트의 평균 점수가 미리 설정된 범위에 해당하도록 상기 모의고사
문제 세트를 구성하는 단계를 포함하는 것을 특징으로 하는 사용자 분석
방법.
- [청구항 5] 제 2항에 있어서, 상기 c 단계는,
미리 계산된 상기 외부 시험 문제 세트의 문제 유형 분포 정보를 따르도록
상기 모의고사 문제 세트를 구성하는 단계를 포함하는 것을 특징으로
하는 사용자 분석 방법.

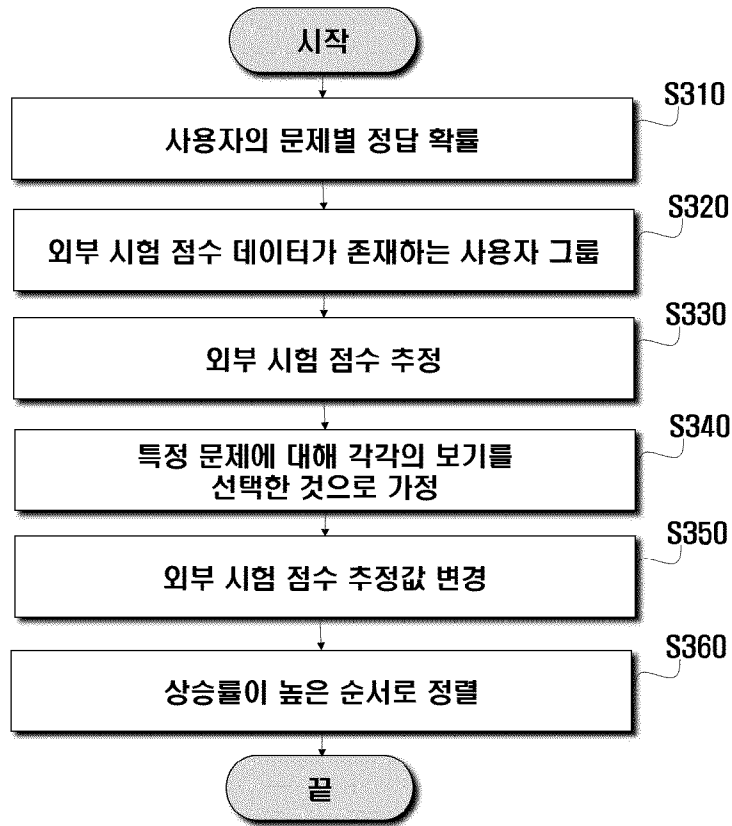
[도1]



[도2]



[도3]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/013590

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 50/10(2012.01)i, G06Q 50/20(2012.01)i, G06N 20/00(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q 50/10; G06N 99/00; G06Q 50/20; G09B 7/07; G06N 20/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: question, database, selection, answer, probability, external test, group, estimate, score

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-1853091 B1 (RHHD. CO.) 27 April 2018 See paragraph [0061], claims 1-2 and figure 1.	1-5
A	KR 10-1895963 B1 (RHHD. CO.) 04 October 2018 See claim 1 and figure 1.	1-5
A	KR 10-1285217 B1 (QUANTLAB INC.) 11 July 2013 See claims 2, 5.	1-5
A	KR 10-2010-0022427 A (KANG, Namhi et al.) 02 March 2010 See paragraphs [0038]-[0039] and claim 1.	1-5
A	KR 10-2017-0034106 A (AJOU UNIVERSITY INDUSTRY-ACADEMIC COOPERATION FOUNDATION) 28 March 2017 See claims 1-3, 13 and figure 1.	1-5
PX	KR 10-2015075 B1 (RHHD. CO.) 27 August 2019 See claims 1-5. * The above document is the published document for the earlier application that serves as the basis for claiming priority of the present international application.	1-5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

24 FEBRUARY 2020 (24.02.2020)

Date of mailing of the international search report

25 FEBRUARY 2020 (25.02.2020)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2019/013590

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-1853091 B1	27/04/2018	CN 108959331 A	07/12/2018
		JP 06625585 B2	25/12/2019
		JP 2018-194804 A	06/12/2018
		KR 10-1996249 B1	04/07/2019
		KR 10-2019-0025870 A	12/03/2019
		KR 10-2019-0025871 A	12/03/2019
		KR 10-2019-0025872 A	12/03/2019
		KR 10-2019-0025873 A	12/03/2019
		US 2018-0336792 A1	22/11/2018
		WO 2018-212394 A1	22/11/2018
KR 10-1895963 B1	04/10/2018	KR 10-2019-0049435 A	09/05/2019
KR 10-1285217 B1	11/07/2013	KR 10-2013-0037541 A	16/04/2013
KR 10-2010-0022427 A	02/03/2010	None	
KR 10-2017-0034106 A	28/03/2017	None	
KR 10-2015075 B1	27/08/2019	None	

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G06Q 50/10(2012.01)i, G06Q 50/20(2012.01)i, G06N 20/00(2019.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G06Q 50/10; G06N 99/00; G06Q 50/20; G09B 7/07; G06N 20/00 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 문제(question), 데이터베이스(database), 선택(selection), 정답(answer), 확률(probability), 외부시험(external test), 그룹(group), 예측(estimate), 점수(score)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-1853091 B1 ((주)뤼이드) 2018.04.27 단락 [0061], 청구항 1-2 및 도면 1 참조.	1-5
A	KR 10-1895963 B1 ((주)뤼이드) 2018.10.04 청구항 1 및 도면 1 참조.	1-5
A	KR 10-1285217 B1 (주식회사 퀴트랩) 2013.07.11 청구항 2,5 참조.	1-5
A	KR 10-2010-0022427 A (강남희 등) 2010.03.02 단락 [0038]-[0039] 및 청구항 1 참조.	1-5
A	KR 10-2017-0034106 A (아주대학교산학협력단) 2017.03.28 청구항 1-3,13 및 도면 1 참조.	1-5
PX	KR 10-2015075 B1 ((주)뤼이드) 2019.08.27 청구항 1-5 참조. * 위 문헌은 본 국제출원의 우선권주장의 기초가 되는 선출원의 공개된 공보임	1-5
<input type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 “X”에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2020년 02월 24일 (24.02.2020)	국제조사보고서 발송일 2020년 02월 25일 (25.02.2020)	
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 장기정 전화번호 +82-42-481-8364	

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-1853091 B1	2018/04/27	CN 108959331 A	2018/12/07
		JP 06625585 B2	2019/12/25
		JP 2018-194804 A	2018/12/06
		KR 10-1996249 B1	2019/07/04
		KR 10-2019-0025870 A	2019/03/12
		KR 10-2019-0025871 A	2019/03/12
		KR 10-2019-0025872 A	2019/03/12
		KR 10-2019-0025873 A	2019/03/12
		US 2018-0336792 A1	2018/11/22
		WO 2018-212394 A1	2018/11/22
		KR 10-1895963 B1	2018/10/04
KR 10-1285217 B1	2013/07/11	KR 10-2013-0037541 A	2013/04/16
KR 10-2010-0022427 A	2010/03/02	없음	
KR 10-2017-0034106 A	2017/03/28	없음	
KR 10-2015075 B1	2019/08/27	없음	