

전자 제출 접수증

ePCT-Filing을 통해 제출한 PCT 국제출원을 수리관청(RO/KR)에서 접수하였음을 알려드립니다.
출원번호와 접수일이 자동으로 부여되었습니다(행정지침 제7부).

접수번호:	612024007707679	
국제출원번호:	PCT/KR2024/014280	
접수일:	23 September 2024	
수리관청:	Korean Intellectual Property Office	
참조기호:	SPC2024-1319	
출원인:	연세대학교 산학협력단	
출원인 수:	1	
발명의 명칭:	데이터베이스 파라미터 튜닝 방법 및 그 시스템	
제출된 문서:	SPC20241319-appb.xml (SPC2024-1319_출원명세서.xml)	58213
	SPC20241319-dpcf-000003.zip	605448
	SPC20241319-fees.xml	1795
	SPC20241319-poat-000001.xml	2424
	SPC20241319-requ.xml	11556
	SPC20241319-vlog.xml	716
제출자:	Ilshin KANG (고객 아이디: user_KR_KANG_ILSHIN_4000)	
접수 타임스탬프:	23 September 2024 15:43 UTC+9 (KST)	
제출에 관한 공식 요약:	8C:A4:35:C1:EB:21:00:1F:FC:AE:05:57:5A:5C:A3:3D:F2:12:E5:D1	

/RO/KR/

PCT 출원서

(전자적 형태가 원본)

0	수리관청 전용	
0-1	국제출원번호	PCT/KR2024/014280
0-2	국제출원일자	2024년 09월 23일 (23.09.2024)
0-3	수리관청 명칭 및 "PCT 국제출원"	대한민국 특허청 PCT 국제출원
0-4	서식 PCT/RO/101 - PCT 출원서	
0-4-1	우측에 기재된 바와 같이 작성되었다.	ePCT-Filing Version 4.14.004 MT/FOP 20240908/1.1
0-5	신청	
	아래 서명인은 본 국제 출원서가 특허협력조약에 의해 처리될 것을 청구합니다.	
0-6	출원인이 지정한 수리관청	대한민국 특허청 (RO/KR)
0-7	출원인 또는 대리인의 서류참조기호	SPC2024-1319
I	발명의 명칭	데이터베이스 파라미터 튜닝 방법 및 그 시스템
II	출원인	
II-1	이 사람은	오직 출원인 (applicant only)
II-2	우측 지정국에 관한 출원인	모든 지정국 (all designated States)
II-4ko	성명	연세대학교 산학협력단
II-4en	Name:	UIF (UNIVERSITY INDUSTRY FOUNDATION), YONSEI UNIVERSITY
II-5ko	주소	대한민국 03722
II-5en	Address:	서울특별시 서대문구 연세로 50 50, Yonsei-ro Seodaemun-gu Seoul 03722 Republic of Korea
II-6	국적	대한민국 KR
II-7	거주국	대한민국 KR
II-10	이메일 주소	sshyun@yonsei.ac.kr
II-10(a)	이메일 사용동의 수리관청, 국제조사기관, 국제사무국, 국제예비심사기관이 필요 시 이 이메일 주소를 사용하여 이 국제 출원과 관련하여 발 행된 통지서를 송부할 것에 동의한다.	오직 전자적 형태의 통지서만 송부 (서면 통지서는 미발송)
II-11	특허고객번호	2-2005-009509-9
III-1	출원인 및/또는 발명자	
III-1-1	이 사람은	오직 발명자 (inventor only)
III-1-3	우측 지정국에 관한 발명자	모든 지정국 (all designated States)
III-1-4ko	성명	박상현
III-1-4en	Name (LAST, First):	PARK, Sanghyun
III-1-5ko	주소	대한민국 03636
III-1-5en	Address:	서울특별시 서대문구 세무서8길 30, 101동 1502호 101-1502, 30 Semuseo 8-gil Seodaemun-gu Seoul 03636 Republic of Korea

PCT 출원서

(전자적 형태가 원본)

III-2	출원인 및/또는 발명자	
III-2-1	이 사람은	오직 발명자 (inventor only)
III-2-3	우측 지정국에 관한 발명자	모든 지정국 (all designated States)
III-2-4ko	성명	권세인
III-2-4en	Name (LAST, First):	KWON, Sein
III-2-5ko	주소	대한민국 03708 서울특별시 서대문구 연희로 93, 702호
III-2-5en	Address:	#702, 93 Yeonhui-ro Seodaemun-gu Seoul 03708 Republic of Korea
III-3	출원인 및/또는 발명자	
III-3-1	이 사람은	오직 발명자 (inventor only)
III-3-3	우측 지정국에 관한 발명자	모든 지정국 (all designated States)
III-3-4ko	성명	이지은
III-3-4en	Name (LAST, First):	LEE, Jieun
III-3-5ko	주소	대한민국 03716 서울특별시 서대문구 동교로 291, 101동 1004호
III-3-5en	Address:	101-1004, 291 Donggyo-ro Seodaemun-gu Seoul 03716 Republic of Korea
III-4	출원인 및/또는 발명자	
III-4-1	이 사람은	오직 발명자 (inventor only)
III-4-3	우측 지정국에 관한 발명자	모든 지정국 (all designated States)
III-4-4ko	성명	염찬호
III-4-4en	Name (LAST, First):	YEOM, Chanho
III-4-5ko	주소	대한민국 03726 서울특별시 서대문구 연희로10가길 51-12, 305호
III-4-5en	Address:	#305, 51-12 Yeonhui-ro 10ga-gil Seodaemun-gu Seoul 03726 Republic of Korea
III-5	출원인 및/또는 발명자	
III-5-1	이 사람은	오직 발명자 (inventor only)
III-5-3	우측 지정국에 관한 발명자	모든 지정국 (all designated States)
III-5-4ko	성명	김휘균
III-5-4en	Name (LAST, First):	JIN, HUIJUN
III-5-5ko	주소	대한민국 08762 서울특별시 관악구 남부순환로159길 32
III-5-5en	Address:	32 Nambusunhwan-ro 159-gil Gwanak-gu Seoul 08762 Republic of Korea

PCT 출원서

(전자적 형태가 원본)

IV-1	대리인 또는 대표자 아래에 기재된 자는 관할 국제기관에 대하여 우측에 표시된 자격으로 출원인을 대리하는 것으로 선임되었다.	대리인
IV-1-1ko	성명	특허법인시공
IV-1-1en	Name:	SIGONG IP&LAW FIRM
IV-1-2ko	주소	대한민국 06729 서울특별시 서초구 강남대로 275, 9층
IV-1-2en	Address:	9F, 275 Gangnam-daero Seocho-gu Seoul 06729 Republic of Korea
IV-1-3	전화번호	+8225778874
IV-1-5	이메일 주소	admin@sigong-ip.com
IV-1-5(a)	이메일 사용동의 수리관청, 국제조사기관, 국제사무국, 국제예비심사기관이 필요 시 이 이메일 주소를 사용하여 이 국제 출원과 관련하여 발행된 통지서를 송부할 것에 동의한다.	오직 전자적 형태의 통지서만 송부 (서면 통지서는 미발송)
IV-1-6	대리인번호	9-2023-100041-2
V	지정국	
V-1	본 출원서의 제출로, 규칙 4.9(a)에 따라, 부여될 수 있는 모든 종류의 권리 보호를 위하여, 그리고 해당하는 경우 지역특허 및 국내특허 모두를 위하여 당해 국제출원일에 PCT에 기록되는 모든 계약국이 지정된다.	
V-2	V-2란은 출원서 제출시 또는 규칙 26의 2.1에 의해 그 이후 출원서 제6기재란에 위 특정 관련 계약국의 국내 선출원에 대한 우선권주장이 포함되어 있을 경우 당해 계약국의 국내법에 의해 해당 국내 선출원의 효력이 상실되는 것을 방지하기 위한 목적으로 당해 계약국의 지정을 제외하는 데에만 사용될 수 있다 (지정 제외시 이의 취소 불가능).	KR
VI-1	선국내출원에 대한 우선권 주장	
VI-1-1	출원일	2024년 09월 09일 (09.09.2024)
VI-1-2	출원번호	10-2024-0122379
VI-1-3	파리협약 당사국명 또는 WTO 회원국명	KR
VI-2	우선권서류 신청 수리관청에 대하여 위에 명시된 선출원의 인증등본을 준비하여 국제사무국에 송부하여 줄 것을 신청한다.	VI-1
VI-3	인용에 의한 보완 조약 제11조(1)(iii)(d) 또는 (e)에서 규정하는 국제출원의 요소, 또는 규칙 20.5(a)에서 규정하는 발명의 설명, 청구범위 또는 도면의 일부, 또는 규칙 20.5(2)(a)에서 규정하는 발명의 설명, 청구범위 또는 도면의 요소 또는 일부가 이 국제출원에는 포함되어 있지 않지만 조약 제11조(1)(iii) 규정의 요소 중 하나 이상이 수리관청에 최초로 접수된 날에 우선권주장의 기초가 된 선출원에 완전히 포함되어 있는 경우, 그 요소 또는 부분은 규칙 20.6에 따른 확인을 조건으로, 규칙 20.6과 관련하여 이 국제출원에 있어서 인용에 의해 보완된다.	
VII-1	국제조사기관(ISA) 선택	대한민국 특허청 (ISA/KR)
VIII	선언서	선언서 개수
VIII-1	발명자의 신원에 관한 선언	-
VIII-2	국제출원일에 특허출원 및 특허를 받을 수 있는 출원인의 자격에 관한 선언	-
VIII-3	국제출원일에 선출원의 우선권을 주장할 수 있는 출원인의 자격에 관한 선언	-
VIII-4	발명자 선언(미국에 대한 지정의 경우에 한함)	-
VIII-5	신규성을 해치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언	-

PCT 출원서

(전자적 형태가 원본)

IX	체크 리스트	용지 수	전자적 파일 첨부
IX-1	출원서(선언서 포함)	4	✓
IX-2	발명의 설명	13	✓
IX-3	청구범위	3	✓
IX-4	요약서	1	✓
IX-5	도면	5	✓
IX-6a	발명의 설명의 서열목록 부분	-	-
IX-7	용지매수 소계	26	
	첨부 항목	서면 첨부	전자적 파일 첨부
IX-8	수수료 계산 용지	-	✓
IX-9	개별위임장 원본	-	✓
IX-20	요약서에 수반되어야 할 도면 번호	1	
IX-21	국제출원의 출원 언어	한국어	
X-1	출원인, 대리인 또는 대표자의 서명 또는 날인	 특허법인시공 강일신 대표자	
X-1-1	성명		
X-1-2	서명인의 성명		
X-1-3	권한 (출원서를 통해 서명자의 자격이 명백하지 않은 경우에는 그 자격도 표시)		

수리관청 전용

10-1	국제출원으로 제출된 서류의 실제 접수일	2024년 09월 23일 (23.09.2024)
10-2	도면	
10-2-1	접수	
10-2-2	미접수	
10-3	국제출원으로 제출된 서류를 완성하는 서류 또는 도면의 추후 기간내 제출에 따른 정정된 실제 접수일	
10-4	PCT 제11조(2)에 따라 제출이 요구된 보완서로서 기간내 제출된 보완서의 접수일	
10-5	국제조사기관(ISA)	ISA/KR
10-6	조사료 납부시까지 지연된 조사용 사본의 송부	

국제 사무국 전용

11-1	국제 사무국의 기록원본 접수일	
------	------------------	--

PCT(부속문서 - 수수료 계산용지)

(전자적 형태가 원본)

이 페이지는 국제 출원서의 일부가 아니며 페이지수에 포함되지 않는다

0	수리관청 전용			
0-1	국제출원번호	PCT/KR2024/014280		
0-2	수리관청의 우편 소인 일자			
0-4	Form PCT/RO/101 (부속문서) PCT 수수료 계산 용지			
0-4-1	우측에 기재된 바와 같이 작성되었다.	ePCT-Filing Version 4.14.004 MT/FOP 20240908/1.1		
0-9	출원인 또는 대리인의 서류참조기호	SPC2024-1319		
2	출원인	연세대학교 산학협력단		
12	규정 수수료 계산	수수료 금액/계수	총 금액 (CHF)	총 금액 (KRW)
12-1	송달료 T	↔		45000
12-2-1	조사료 S	↔		450000
12-2-2	국제조사기관	KR		
12-3	국제 출원 수수료 최초 30장 i1	1330		
12-4	최초 30장 초과 장수	0		
12-5	최초 30장 초과 1장당 추가 수수료 (X) 0	0		
12-6	총 추가금액 i2	0		
12-7	i1 + i2 = i	1330		
12-12	XML 전자출원 감면 R	-300		
12-13	총 국제출원 수수료(i-R) I	↔	1030	
12-19	총 금액(T+S+I+P)	↔	1030	495000
12-21	결제 방법	기타 : 현재 지불금액이 없습니다.		

PCT 위임장

(전자적 형태가 원본)

0-1	PCT 위임장 (특허 협력 조약에 의거하여 제출된 국제 출원) (PCT 규칙 제90.4조)	
0-1-1	우측에 기재된 바와 같이 작성되었다.	ePCT-Filing Version 4.14.004 MT/FOP 20240908/1.1
1	아래에 서명한 출원인	연세대학교 산학협력단
1-1-1	우측에 기재된 사람을 아래의 자격으로 선임한다.	특허법인시공 SIGONG IP&LAW FIRM 대한민국 06729 서울특별시 서초구 강남대로 275, 9층 9F, 275 Gangnam-daero Seocho-gu Seoul 06729 Republic of Korea
1-2	자격	대리인
1-3	우측 기관에 대하여	모든 관할 국제 기관
1-4	아래의 국제 출원에 관한 서명의 출원인을 대리함	
1-4-1	발명의 명칭	데이터베이스 파라미터 튜닝 방법 및 그 시스템
1-4-2	출원인 또는 대리인의 서류참조기호	SPC2024-1319
1-4-3	국제출원번호(이용 가능한 경우)	
1-4-4	수리관청	대한민국 특허청 (RO/KR)
1-5	그리고 아래 서명인을 대신하여 지불하거나 지불받았다.	
2-1	출원인 서명	
2-1-1	성명	연세대학교 산학협력단
2-1-2	서명인의 성명	홍종일
2-1-3	권한 (출원서를 통해 서명자의 자격이 명백하지 않은 경우에는 그 자격도 표시)	대표자
3	일자	2024년 09월 23일 (23.09.2024)

발명의 설명

발명의 명칭: 데이터베이스 파라미터 튜닝 방법 및 그 시스템 기술분야

- [1] 본 개시는 데이터베이스 파라미터 튜닝 방법 및 그 시스템에 관한 것으로, 보다 구체적으로, 데이터베이스와 관련된 파라미터와 그 성능 지표를 증강시켜 잠재 공간에 인코딩한 후 잠재 공간의 튜닝을 통해 데이터베이스 파라미터를 튜닝하는 방법 및 그 시스템에 관한 것이다.

배경기술

- [2] 최근 데이터의 양이 방대해지고 있으며 대용량 데이터를 효율적으로 처리하기 위해 데이터를 관리하는 데이터베이스의 성능에 대한 관심도 높아지고 있다. 이러한 데이터베이스의 성능을 높이기 위한 방법에는 데이터베이스 파라미터 튜닝 방법이 존재한다. 데이터베이스 파라미터 튜닝은 사용자가 사용하는 데이터베이스에 존재하는 파라미터 설정 값을 적절하게 조절하여 데이터베이스의 성능을 향상시키는 작업이다. 종래 데이터베이스 파라미터 튜닝은 데이터베이스 관리자(Database Administrator; DBA)가 파라미터의 값을 바꿔가며 튜닝을 진행하였으나 이에 대해 몇 가지 한계점이 존재하였다.
- [3] 최근에는 머신러닝을 활용한 자동화 튜닝 연구가 활발히 진행되고 있으며, 베이시안 최적화(Bayesian Optimization; BO)와 강화학습(Reinforcement Learning; RL) 같은 알고리즘이 사용되어 왔다. 하지만 데이터베이스 파라미터가 고차원의 공간을 가지기 때문에 최적화가 어려운 문제점이 존재한다. 또한, 상위 파라미터를 선택하는 알고리즘을 사용하면 알고리즘을 적용할 때마다 그 결과가 달라지고, 성능 향상에 도움이 되는 파라미터가 선별되지 않아 튜닝에서 제외될 수 있다는 문제가 존재하며, 선택된 파라미터 이외의 다른 파라미터에 대한 정보를 놓칠 수 있다. 나아가, 워크로드 간 유사성을 바탕으로 튜닝을 진행하는 방식은 정확도가 떨어질 수 있으며, 잘못된 매핑이 이루어질 경우 비효율적인 파라미터가 선택될 가능성이 있다. 따라서, 이러한 한계를 극복할 수 있는 데이터베이스 파라미터의 최적화 방법이 요구된다.

발명의 내용

기술적 과제

- [4] 본 개시의 실시 예들을 통해 해결하고자 하는 기술적 과제는 우리는 sample augmentation 하는 과정을 추가하여 데이터를 생성할 때의 time consuming 문제를 해결할 수 있는 모델을 제공하는 것이다.
- [5] 또한 본 개시의 실시 예들을 통해 해결하고자 하는 기술적 과제는 기존 데이터베이스 파라미터 튜닝 모델의 기본적인 architecture에서 벗어나 전체 파라미터와 workload에 대한 정보를 압축한 latent space를 최적화하는 모델을 제공하는 것이다.

- [6] 본 개시가 해결하고자 하는 과제들은 상술한 과제에 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제 해결 수단

- [7] 본 개시의 실시 예에 따른 데이터베이스 파라미터 튜닝 방법은 적어도 하나의 컴퓨팅 장치에 의해 수행되고, 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 단계, 상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 단계, 상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 단계, 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 단계, 상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화되도록, 상기 잠재 공간을 튜닝하는 단계, 및 상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 단계를 포함할 수 있다.
- [8] 일 실시 예에서, 상기 제 1 데이터베이스의 성능 지표는 상기 제 1 데이터베이스의 쿼리 처리량, 상기 제 1 데이터베이스의 데이터 처리 지연 시간, 상기 제 1 데이터베이스에서 초당 처리되는 작업의 수, 상기 제 1 데이터베이스의 쓰기 증폭 비율, 및 상기 제 1 데이터베이스의 공간 증폭 비율 중 적어도 어느 하나를 포함할 수 있다.
- [9] 일 실시 예에서, 상기 제 2 파라미터 세트를 생성하는 단계는, 상기 제 1 파라미터 세트에 대해 라틴 하이퍼큐브 샘플링(Latin Hypercube Sampling; LHS)을 적용하여 상기 제 2 파라미터 세트를 생성하는 단계를 포함할 수 있다.
- [10] 일 실시 예에서, 상기 입력 데이터 세트는 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트를 연결(concatenate)하여 생성된 것이고, 상기 입력 데이터 세트의 차원은 상기 제 1 및 제 2 파라미터 세트가 포함하는 파라미터의 개수 및 상기 제 1 및 제 2 성능 세트가 포함하는 성능 지표의 개수의 합인 것을 특징으로 할 수 있다.
- [11] 일 실시 예에서, 상기 목적 함수는 상기 제 1 데이터베이스의 성능 지표와 관련된 함수이고, 상기 잠재 공간을 튜닝하는 단계는, 상기 잠재 표현에 대해 베이지안 최적화(Bayesian optimization)를 사용하여 상기 목적 함수를 최대화하는 잠재 표현을 결정하는 단계를 포함할 수 있다.
- [12] 일 실시 예에서, 상기 방법은 상기 복수의 튜닝된 파라미터를 상기 제 1 데이터베이스에 적용하는 단계를 더 포함할 수 있다.

- [13] 본 개시의 다른 실시 예에 따른 데이터 파라미터 튜닝 시스템은 프로세서, 및 인스트럭션(instruction)들을 저장하는 메모리를 포함하고, 상기 인스트럭션들은 상기 프로세서에 의해 실행될 때, 상기 프로세서로 하여금, 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 동작, 상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 동작, 상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 동작, 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 동작, 상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화 되도록, 상기 잠재 공간을 튜닝하는 동작, 및 상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 동작을 수행하도록 할 수 있다.
- [14] 일 실시 예에서, 상기 입력 데이터 세트는 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트를 연결(concatenate)하여 생성된 것이고, 상기 입력 데이터 세트의 차원은 상기 제 1 및 제 2 파라미터 세트가 포함하는 파라미터의 개수 및 상기 제 1 및 제 2 성능 세트가 포함하는 성능 지표의 개수의 합인 것을 특징으로 할 수 있다.
- [15] 일 실시 예에서, 상기 목적 함수는 상기 제 1 데이터베이스의 성능 지표와 관련된 함수이고, 상기 잠재 공간을 튜닝하는 동작은, 상기 잠재 표현에 대해 베이지안 최적화(Bayesian optimization)를 사용하여 상기 목적 함수를 최대화하는 잠재 표현을 결정하는 동작을 포함할 수 있다.
- [16] 본 개시의 또 다른 실시 예에 따른 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터 프로그램 코드를 저장하고, 프로세서에 의해 상기 컴퓨터 프로그램 코드가 실행될 때, 상기 프로세서는, 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 동작, 상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 동작, 상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 동작, 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 동작, 상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화 되도록, 상기 잠재 공간을 튜닝하는 동작, 및 상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된

복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 동작을 수행할 수 있다.

발명의 효과

- [17] 본 개시의 실시 예에 따르면, 적은 양의 데이터 세트에 대해 데이터 증강을 수행하여 데이터 세트를 증가시키는 데 소요되는 시간이 절약될 수 있다.
- [18] 또한 본 개시의 실시 예에 따르면, 고차원의 데이터 파라미터 세트를 저차원의 잠재 공간으로 인코딩한 후 파라미터 튜닝을 수행함으로써, 일부 파라미터가 아닌 전체 파라미터에 대한 튜닝이 가능할 수 있다.
- [19] 나아가 본 개시의 실시 예에 따르면, 잠재 공간으로 인코딩하는 과정에서 데이터베이스 파라미터뿐만 아니라 해당 파라미터에 대응하는 데이터베이스의 성능 지표도 함께 사용됨으로써, 다양한 워크로드에 대한 데이터베이스 튜닝이 가능할 수 있다.
- [20] 본 개시에 따른 효과들은 상술한 효과에 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [21] 도 1은 본 개시의 실시 예에 따른 전체 시스템의 구성을 나타내는 예시적인 블록도이다.
- [22] 도 2는 도 1의 예측 모듈의 동작을 예시적으로 나타낸다.
- [23] 도 3은 도 1의 증강 모듈의 동작을 예시적으로 나타낸다.
- [24] 도 4는 도 1의 튜닝 모듈의 동작을 예시적으로 나타낸다.
- [25] 도 5는 본 개시의 실시 예에 따른 데이터 파라미터 튜닝 방법을 예시적으로 나타내는 흐름도이다.
- [26] 도 6은 본 개시의 실시 예에 따른 파라미터 튜닝 방법을 수행하기 위한 컴퓨팅 장치의 하드웨어 구성을 나타내는 블록도이다.

발명의 실시를 위한 형태

- [27] 이하, 첨부된 도면들에 기재된 내용들을 참조하여 본 발명에 따른 예시적 실시 예를 상세하게 설명한다. 다만, 본 발명이 예시적 실시 예들에 의해 제한되거나 한정되는 것은 아니다. 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 것이나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다.
- [28] 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다. 특정한 경우, 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 개시에서 사용되는 용어는 단순한 용

어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

- [29] 본 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 본 명세서에서 사용된 단수형은 특별히 언급하지 않는 한 복수형도 포함한다. 또한, 본 명세서 전체에서 기재된 "a, b 및/또는 c 중 적어도 하나"의 표현은, 'a 단독', 'b 단독', 'c 단독', 'a 및 b', 'a 및 c', 'b 및 c', 또는 'a, b, c 모두'를 포괄할 수 있다.
- [30] 한편, 본 명세서에서 사용되는 "제1 및/또는 제2" 등의 용어는 다양한 구성요소들을 설명하기 위하여 사용될 수 있으나, 이는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 목적으로만 사용될 뿐, 해당 용어로 지칭되는 구성요소로 한정하기 위한 것은 아니다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않는 한, 제1 구성요소는 제2 구성요소로 명명될 수 있으며, 제2 구성요소 또한 제1 구성요소로 명명될 수 있다.
- [31] 또한, 본 명세서에 기재된 "...부", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다. 또한, 본 명세서에서 본 개시의 실시 예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 본 개시의 실시 예는 하나 이상의 마이크로프로세서의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다.
- [32] 본 개시에 따른 실시 예에서, 인공지능과 관련된 기능은 프로세서 및 메모리를 통해 구현될 수 있다. 이 때, 프로세서는 CPU(Center Processing Unit), AP(Application Processor), DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU(Graphic Processing Unit), VPU(Vision Processing Unit)와 같은 그래픽 전용 프로세서 및 NPU(Neural network Processing Unit)와 같은 인공지능 전용 프로세서 중 어느 하나일 수 있다. 프로세서는 메모리에 저장된 기 정의된 동작 규칙 또는 인공지능 모델에 따라 입력 데이터를 처리할 수 있다. 또는, 프로세서가 인공지능 전용 프로세서인 경우, 인공지능 전용 프로세서는 특정 인공지능 모델의 처리에 특화된 하드웨어 구조로 설계될 수 있다. 본 개시에 따른 일부 실시 예에서, 인공지능과 관련된 기능은 복수의 프로세서들을 통해 구현될 수 있다.
- [33] 본 개시에 따른 실시 예에서, 기 정의된 동작 규칙 또는 인공지능 모델은 기계학습을 수행하도록 구성될 수 있다. 여기서, 기계학습을 수행하도록 구성된다는 것은, 기 정의된 동작 규칙 또는 인공지능 모델이 학습 알고리즘을 기반으로 다수의 학습 데이터들을 이용하여 학습되어 원하는 특성(또는 목적)을 수행하도록 구성됨을 의미한다. 이러한 학습은 본 개시에 따른 인공지능이 구현되는 장치 자체

에서 이루어질 수도 있고, 별도의 서버 및/또는 시스템을 통해 이루어질 수도 있다.

- [34] 인공지능 모델은 뉴럴 네트워크(또는 인공 신경망)로 구현될 수 있으며, 기계학습과 인지과학에서 생물학의 신경을 모방한 통계학적 학습 알고리즘에 기반하여 동작할 수 있다. 뉴럴 네트워크는 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜 문제 해결 능력을 가지는 모델 전반을 의미할 수 있다. 뉴럴 네트워크는 복수의 신경망 레이어(layer)들로 구성될 수 있으며, 예시적으로 뉴럴 네트워크는 입력 레이어(input layer), 은닉 레이어(hidden layer) 및 출력 레이어(output layer)를 포함할 수 있다. 복수의 신경망 레이어들 각각은 적어도 하나의 노드(node) 및 적어도 하나의 가중치(weight)를 포함할 수 있으며, 이전(previous) 레이어의 연산 결과와 가중치 간의 연산을 통해 신경망 연산을 수행할 수 있다. 복수의 신경망 레이어들이 가지고 있는 적어도 하나의 가중치는 인공지능 모델의 학습 결과에 의하여 최적화될 수 있다. 예를 들어, 학습 과정동안 인공지능 모델에서 획득한 손실(loss) 값 또는 비용(cost) 값이 감소 또는 최소화되도록 적어도 하나의 가중치가 갱신될 수 있다. 뉴럴 네트워크는 임의의 입력으로부터 예측하고자 하는 결과를 추론할 수 있다.
- [35] 인공지능 모델의 학습 방법은 학습 방식에 따라 입력 데이터 및 출력 데이터가 훈련 데이터로써 제공되어 문제(입력 데이터)에 대응하는 정답(출력 데이터)이 정해져 있는 지도학습(supervised learning), 출력 데이터 없이 입력 데이터만 제공되어 문제(입력 데이터)에 대응하는 정답(출력 데이터)이 정해지지 않은 비지도 학습(unsupervised learning) 및 현재 상태(state)에서 어떤 행동(action)을 취할 때마다 보상(reward)이 부여되고, 이러한 보상을 최대화하는 방향으로 학습을 진행하는 강화학습(reinforcement learning) 등으로 구분될 수 있다. 또는, 학습 모델의 구조인 아키텍처에 따라 구분될 수도 있다.
- [36] 본 개시의 실시 예에서, 인공지능 모델은 GoogleNet, AlexNet, VGG Network 등과 같은 CNN(Convolution Neural Network), R-CNN(Region with Convolution Neural Network), RPN(Region Proposal Network), RNN(Recurrent Neural Network), S-DNN(Stacking-based deep Neural Network), S-SDNN(State-Space Dynamic Neural Network), Deconvolution Network, DBN(Deep Belief Network), RBM(Restricted Boltzman Machine), Fully Convolutional Network, LSTM(Long Short-Term Memory) Network, Classification Network, Generative Modeling, eXplainable AI, Continual AI, Representation Learning, AI for Material Design, 자연어 처리를 위한 BERT, SP-BERT, MRC/QA, Text Analysis, Dialog System, GPT-3, GPT-4, 비전 처리를 위한 Visual Analytics, Visual Understanding, Video Synthesis, ResNet 데이터 지능을 위한 Anomaly Detection, Prediction, Time-Series Forecasting, Optimization, Recommendation, Data Creation 등 다양한 인공지능 구조 및 알고리즘 중 적어도 하나를 이용할 수 있으며, 상술한 예시들은 본 개시의 실시 예에 따라 이용되는

인공지능 구조 및 알고리즘의 예를 나열한 것인 뿐, 본 개시의 실시 예에 따라 이용되는 인공지능 구조 및 알고리즘을 제한하는 것은 아니다.

- [37] 이하, 본 개시의 다양한 실시 예들은 첨부된 도면을 참조하여 상세하게 설명될 것이다. 실시 예를 설명함에 있어서 본 발명이 속하는 기술 분야에 익히 알려져 있고 본 발명과 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략할 것이다. 이는 불필요한 설명을 생략함으로써 본 발명의 요지를 흐리지 않고 더욱 명확히 전달하기 위함이다. 마찬가지로 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시되었다. 또한, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다. 본 명세서에서, 전문에 걸쳐 동일한 참조 부호는 동일한 또는 대응하는 구성 요소를 지칭할 수 있다.
- [38] 본 명세서에서, 데이터베이스 파라미터 튜닝(database parameter tuning)이란, 데이터베이스 파라미터를 최적화함으로써 데이터베이스의 성능을 최대화로 이끌어내는 방법을 나타낸다. 또한, 본 명세서에서 성능 지표는 데이터베이스 관리 시스템(database management system; DBMS)의 성능 지표를 나타내는 것이나, 본 명세서 전반에 걸쳐 데이터베이스 관리 시스템의 성능 지표와 데이터베이스의 성능 지표는 혼용될 수 있을 것이다.
- [39] 도 1은 본 개시의 실시 예에 따른 전체 시스템의 구성을 나타내는 예시적인 블록도이다. 도 1을 참조하면, 전체 시스템은 데이터베이스 파라미터 튜닝 시스템(10), 데이터베이스 관리 시스템(database management system; DBMS)(20), 및 데이터베이스(30)를 포함할 수 있다. 그리고, 데이터베이스 파라미터 튜닝 시스템(10)은 예측 모듈(11), 증강 모듈(13), 및 튜닝 모듈(14)을 포함할 수 있다.
- [40] 한편, 도 1에 도시된 데이터베이스 파라미터 튜닝 시스템(10)의 구성 요소들(모듈들)은 기능적으로 구분되는 기능 요소들을 나타낸 것으로서, 적어도 하나의 구성 요소(모듈)가 실제 물리적 환경에서는 서로 통합되는 형태로 구현될 수도 있음에 유의한다. 즉, 데이터베이스 파라미터 튜닝 시스템(10)의 구성 요소들(모듈들)은 하나의 물리적인 서버로 구현될 수도 있고, 복수의 서버에 나누어 구현될 수도 있다.
- [41] 데이터베이스 파라미터 튜닝 시스템(10)은 예측 모듈(11)을 통해 데이터베이스(30)와 관련된 복수의 파라미터에 기반하여, 데이터베이스(30)가 해당 복수의 파라미터를 가질 때의 성능 지표를 출력할 수 있다. 예측 모듈(11)은 복수의 파라미터를 입력 받아 성능 지표를 출력하기 위한 인공지능 모델인 제 1 모델(12)을 포함할 수 있다. 예를 들어, 제 1 모델(12)은 TabNet을 사용하여 구현된 모델일 수 있다. 이하 도 2를 참조하여, 예측 모듈(11)의 동작에 대해 구체적으로 설명한다.
- [42] 도 2는 도 1의 예측 모듈(11)의 동작을 예시적으로 나타낸다. 도 2를 참조하면, 복수의 파라미터 세트 $conf_0$, $conf_1$, $conf_2$, ..., $conf_k$ 가 나타나 있다. 각 파라미터 세트는 데이터베이스(30)와 관련된 복수의 파라미터를 포함할 수 있다. 예를 들어, 데이터베이스(30)와 관련된 복수의 파라미터는 데이터베이스의 블록 크기, 버퍼 풀 크기, 로그 파일 크기, 최소 토큰 크기, 캐시 크기, 압축률, 대기 지연 등을 포함

할 수 있다. 각 파라미터 세트($\text{conf}_0, \text{conf}_1, \text{conf}_2, \dots, \text{conf}_k$)는 동일한 종류의 복수의 파라미터에 대한 서로 다른 값을 포함할 수 있다.

- [43] 복수의 파라미터 세트는 제 1 모델(12)에 입력될 수 있고, 복수의 파라미터 세트에 대응하는 복수의 성능 세트가 출력될 수 있다. 예를 들어, 파라미터 세트(conf_0)에 대응하는 성능 세트는 데이터베이스(30)가 파라미터 세트(conf_0)에 대응하는 값들을 가지는 복수의 파라미터를 가질 때 데이터베이스(30)의 성능 지표에 해당할 수 있다. 예를 들어, 데이터베이스(30)의 성능 지표는 쿼리 처리량(throughput), 데이터 처리 지연 시간(latency), 초당 처리되는 작업의 수(work rate), 쓰기 증폭 비율(write amplification factor; WAF), 및 공간 증폭 비율(space amplification factor; SAF) 중 적어도 어느 하나를 포함할 수 있다. 한편, 데이터베이스(30)의 성능 지표는 데이터베이스 관리 시스템(20)의 종류에 따라(예를 들어, MySQL인지, RocksDB인지 등에 따라) 달라질 수 있을 것이다.
- [44] 다시 도 1로 돌아가서, 데이터베이스 파라미터 튜닝 시스템(10)은 증강 모듈(13)을 통해 복수의 파라미터 세트를 증강시킬 수 있다. 그리고, 증강된 파라미터 세트에 포함된 복수의 파라미터를 다시 예측 모듈(11)의 제 1 모델(12)에 입력하여, 증강된 복수의 파라미터에 대응하는 성능 지표를 출력할 수 있다. 이하 도 3을 참조하여, 증강 모듈(13)의 동작에 대해 구체적으로 설명한다.
- [45] 도 3은 도 1의 증강 모듈(13)의 동작을 예시적으로 나타낸다. 도 3을 참조하면, 증강 모듈(13)은 복수의 파라미터 세트($\text{conf}_0, \text{conf}_1, \text{conf}_2, \dots, \text{conf}_k$)에 대해 데이터 증강을 수행하여 복수의 증강된 파라미터 세트를 생성할 수 있다. 좀 더 상세하게는, 데이터 증강의 수행 결과 각 파라미터 세트($\text{conf}_0, \text{conf}_1, \text{conf}_2, \dots, \text{conf}_k$) 별로 n 개의 파라미터 세트가 생성될 수 있을 것이다. 예를 들어, 파라미터 세트(conf_k)에 대응하는 증강된 파라미터 세트는 $\text{AUG_conf}_{k+1}, \text{AUG_conf}_{k+2}, \text{AUG_conf}_{k+3}, \dots, \text{AUG_conf}_{k+n}$ 과 같을 것이다. 이들은 각각 제 1 모델(12)에 입력될 수 있고, 각 증강된 파라미터 세트에 대응하는 성능 세트가 출력될 것이다. 증강된 파라미터 세트에 대응하는 성능 세트는, 도 2를 참조하여 생성된 성능 세트가 증강된 데이터 세트에 해당할 것이다.
- [46] 예를 들어, 데이터 증강은 각 파라미터 세트($\text{conf}_0, \text{conf}_1, \text{conf}_2, \dots, \text{conf}_k$)에 대해 라틴 하이퍼큐브 샘플링(Latin Hypercube Sampling; LHS)을 적용함으로써 수행될 수 있다. LHS를 적용함으로써, 각 파라미터 세트($\text{conf}_0, \text{conf}_1, \text{conf}_2, \dots, \text{conf}_k$)에 대해 가능한 모든 값에 있어 균일한 샘플링이 수행될 수 있을 것이다. 그러나 본 개시는 이에 한정되지 않으며, LHS 이외의 다른 방법을 통해 데이터 증강이 수행될 수도 있을 것이다. 이와 같이 데이터 증강이 수행됨으로써, 데이터베이스 파라미터 튜닝을 위한 다수의 데이터 세트를 생성하는 데 소요되는 시간이 단축될 수 있다.
- [47] 다시 도 1로 돌아가서, 데이터 파라미터 튜닝 시스템(10)은 튜닝 모듈(14)을 통해 도 2의 파라미터 세트, 성능 세트, 도 3의 증강된 파라미터 세트, 증강된 성능

세트를 잠재 공간으로 인코딩할 수 있고, 잠재 공간 상에서 데이터베이스(30)의 성능 지표와 관련된 목적 함수를 최대화하도록 복수의 파라미터를 튜닝할 수 있다. 그리고, 데이터 파라미터 튜닝 시스템(10)은 튜닝 결과를 디코딩하여 출력된 튜닝된 파라미터 세트를 데이터베이스 관리 시스템(20)을 통해 데이터베이스(30)에 적용할 수 있다. 튜닝 모듈(14)은 잠재 공간으로의 인코딩 및 잠재 공간에서의 디코딩을 위한 오토인코더(15) 및 잠재 공간에 대응하는 잠재 표현을 입력받아 목적함수를 출력하기 위한 인공지능 모델인 제 2 모델(16)을 포함할 수 있다. 이하 도 4를 참조하여, 튜닝 모듈(14)의 동작에 대해 구체적으로 설명한다.

[48] 도 4는 도 1의 튜닝 모듈(14)의 동작을 예시적으로 나타낸다. 도 4를 참조하면, 입력 데이터 세트는 복수의 파라미터 세트($conf_0, conf_1, conf_2, \dots, conf_k$), 복수의 파라미터 세트를 제 1 모델(12)에 입력하여 출력된 성능 세트, 증강된 복수의 파라미터 세트(예를 들어, $AUG_conf_{k+1}, AUG_conf_{k+2}, AUG_conf_{k+3}, \dots, AUG_conf_{k+n}$), 증강된 복수의 파라미터 세트를 제 1 모델(12)에 입력하여 출력된 성능 세트를 연결(concatenate)하여 생성될 수 있다. 예를 들어, 입력 데이터 세트의 차원은 각 파라미터 세트(또는 각 증강된 파라미터 세트)가 포함하는 파라미터의 개수와, 성능 세트가 포함하는 성능 지표의 개수의 합에 대응할 수 있다.

[49] 입력 데이터 세트는 오토 인코더(15)의 인코더를 통해 잠재 공간(latent space)으로 인코딩될 수 있고, 대응하는 잠재 표현(latent representation)이 생성될 수 있다. 예를 들어, 잠재 표현의 차원은 입력 데이터 세트의 차원보다 작으며, 0 이상 1 이하의 값을 가질 수 있다. 인코더를 통해 생성된 잠재 표현은 베이지안 최적화(Bayesian optimization; BO)를 사용하여 튜닝될 수 있다. 잠재 표현의 튜닝은 목적 함수 $f(x)$ 의 최대화를 통해 수행될 수 있다.

[50] 여기서, 목적함수 $f(x)$ 는 데이터베이스(30)의 성능 지표와 관련된 함수(즉, 성능 지표를 변수로 포함하는 함수)이며, 제 2 모델(16)에 잠재 표현을 입력함으로써 출력될 수 있다. 예를 들어, 목적함수 $f(x)$ 는 쿼리 처리량, 데이터 처리 지연 시간, 초당 처리되는 작업의 수, 쓰기 증폭 비율, 및 공간 증폭 비율 중 적어도 어느 하나를 변수로 포함하는 함수일 수 있으며, 이 변수들의 값은 잠재 표현에 따라 달라질 수 있고, 잠재 표현의 튜닝 여하에 따라 성능 지표를 나타내는 목적함수도 튜닝될 수 있다. 즉, 목적함수 $f(x)$ 가 최대화된다는 것은 성능 지표를 최대화하는 파라미터의 조합이 결정되었음을 의미할 수 있다. 여기서 제 2 모델(16)은 제 1 모델(12)과 동일하게 TabNet을 사용하여 구현될 수 있으나, 다른 가중치를 가지도록 학습된 상이한 모델일 수 있다.

[51] 오토 인코더(15)는 잠재 표현에 대해 베이지안 최적화를 사용하여 목적 함수 $f(x)$ 를 최대화하는 잠재 표현을 결정할 수 있다. 다시 말해, 목적 함수를 최대화하도록 결정된 잠재 표현은 튜닝된 잠재 표현(튜닝된 잠재 공간)에 해당할 수 있다. 그리고 이와 같이 튜닝된 잠재 표현(튜닝된 잠재 공간)은 디코더를 통해 디코딩될 수 있고, 입력 데이터 세트와 동일한 차원을 가지는 출력 데이터 세트로서 생성될 수 있다. 즉, 출력 데이터 세트는 데이터베이스(30)와 관련된 복수의 튜닝된

파라미터, 그리고 데이터베이스(30)가 튜닝된 파라미터를 가질 때의 성능 지표를 포함할 수 있다.

- [52] 데이터베이스 파라미터 튜닝 시스템(10)은 적어도 하나의 컴퓨팅 장치로 구현될 수 있다. 예를 들어, 데이터베이스 파라미터 튜닝 시스템(10)의 모든 기능이 하나의 컴퓨팅 장치에서 구현될 수도 있고, 데이터베이스 파라미터 튜닝 시스템(10)의 제 1 기능은 제 1 컴퓨팅 장치에서 구현되고, 제 2 기능은 제 2 컴퓨팅 장치에서 구현될 수도 있다. 또는, 데이터베이스 파라미터 튜닝 시스템(10)의 특정 기능이 복수의 컴퓨팅 장치들에서 구현될 수도 있다. 예를 들어, 컴퓨팅 장치는 스마트폰, 데스크탑, 랩탑, 태블릿 등 연산 수단 및 통신 수단이 구비된 모든 종류의 장치를 포함할 수 있다. 나아가, 데이터베이스 파라미터 튜닝 시스템(10)은 가상 머신 등 클라우드 기술에 기반하여 서버 팜(server farm)에 포함된 하나 이상의 물리 서버(physical server)를 이용하여 구성될 수도 있을 것이다. 데이터베이스 파라미터 튜닝 시스템(10)이 컴퓨팅 장치로 구현된 실시 예는 도 6을 참조하여 설명된다.
- [53] 데이터베이스 관리 시스템(DBMS)(20)은 다수의 사용자가 데이터베이스(30) 내의 데이터에 액세스할 수 있도록 해주는 소프트웨어 도구의 집합으로, 사용자의 요청을 수신하여 데이터베이스(30)로부터 대응하는 응답을 출력할 수 있고, 데이터베이스(30)의 파라미터를 관리할 수 있다. 예를 들어, 데이터베이스 관리 시스템(20)은 MySQL, Oracle, MS-SQL, PostgreSQL, MongoDB, Redis, SQLite, RocksDB 등을 포함할 수 있으나, 본 개시는 이에 한정되지 않는다. 데이터베이스 관리 시스템(20)의 종류에 따라 성능 지표의 종류는 다소 달라질 수 있다. 상술한 바와 같이, 본 명세서에서 데이터베이스 관리 시스템(20)의 성능 지표와, 데이터베이스(30)의 성능 지표는 혼용될 수 있다.
- [54] 도 1에 도시된 구성 요소들은 네트워크를 통해 통신할 수 있다. 예를 들어, 네트워크는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN), 이동 통신망(mobile radio communication network), Wibro(Wireless Broadband Internet) 등과 같은 모든 종류의 유/무선 네트워크로 구현될 수 있다.
- [55] 도 5는 본 개시의 실시 예에 따른 데이터 파라미터 튜닝 방법을 예시적으로 나타내는 흐름도이다. 참고로, 도 5는 도 1의 데이터베이스 파라미터 튜닝 시스템(10)에서 수행되는 데이터베이스 파라미터 튜닝 방법의 단계/동작들을 나타내고 있다. 따라서, 이하의 설명들에서, 특정 단계/동작의 주체가 생략된 경우, 데이터베이스 파라미터 튜닝 시스템(10)에서 수행되는 것으로 이해될 수 있다.
- [56] 단계 S110에서, 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강이 수행될 수 있고, 이를 통해 제 2 파라미터 세트가 생성될 수 있다. 예를 들어, 데이터 증강 방법으로 라틴 하이퍼큐브 샘플링(LHS)이 사용될 수 있다.

- [57] 단계 S120에서, 제 1 파라미터 세트가 제 1 모델에 입력될 수 있고, 제 1 파라미터 세트에 대응하는 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트가 출력될 수 있다. 그리고, 단계 S130에서, 제 2 파라미터 세트가 제 1 모델에 입력될 수 있고, 제 2 파라미터 세트에 대응하는 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트가 출력될 수 있다. 즉, 제 2 성능 세트는 제 1 성능 세트가 증강된 데이터 세트에 해당할 수 있다.
- [58] 단계 S140에서, 제 1 및 제 2 파라미터 세트, 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트가 잠재 공간으로 인코딩되어 잠재 표현이 생성될 수 있다. 여기서 입력 데이터 세트는 제 1 및 제 2 파라미터 세트, 제 1 및 제 2 성능 세트가 연결(concatenate)하여 생성된 데이터 세트일 수 있다. 잠재 표현은 제 1 모델과 상이한 제 2 모델에 입력될 수 있고, 목적 함수가 출력될 수 있다. 여기서, 목적 함수는 제 1 데이터베이스의 성능 지표와 관련된 함수(즉, 성능 지표를 변수로 하는 함수)일 수 있다.
- [59] 단계 S150에서, 출력된 목적 함수가 최대화되도록, 잠재 공간이 튜닝될 수 있다. 예를 들어, 잠재 공간의 튜닝은 잠재 표현에 대해 베이지안 최적화(BO)를 사용하여 목적 함수를 최대화하는 잠재 표현을 결정함으로써 수행될 수 있다. 단계 S160에서, 튜닝된 잠재 공간은 디코딩될 수 있고, 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트가 생성될 수 있다. 마지막으로 단계 S170에서, 출력 데이터 세트에 포함된 복수의 튜닝된 파라미터가 데이터베이스 관리 시스템을 통해 제 1 데이터베이스에 적용될 수 있다. 이로써, 제 1 데이터베이스의 성능 지표 또한 최적화될 수 있다.
- [60] 도 6은 본 개시의 실시 예에 따른 파라미터 튜닝 방법을 수행하기 위한 컴퓨팅 장치(500)의 하드웨어 구성을 나타내는 블록도이다.
- [61] 컴퓨팅 장치(500)는 하나 이상의 프로세서(510), 버스(530), 통신 인터페이스(540), 프로세서(510)에 의하여 수행되는 컴퓨터 프로그램을 로드(load)하는 메모리(520)와, 컴퓨터 프로그램(560)을 저장하는 스토리지(550)를 포함할 수 있다. 다만, 도 6에는 본 개시의 실시 예와 관련 있는 구성 요소들만이 도시되어 있다. 따라서, 본 개시가 속한 기술 분야의 통상의 기술자라면 도 6에 도시된 구성요소들 외에 다른 범용적인 구성 요소들이 더 포함될 수 있음을 알 수 있다. 즉, 컴퓨팅 장치(500)에는, 도 6에 도시된 구성 요소 이외에도 다양한 구성 요소가 더 포함될 수 있다. 또한, 경우에 따라, 도 6에 도시된 구성요소들 중 일부가 생략된 형태로 컴퓨팅 장치(500)가 구성될 수도 있다. 이하, 컴퓨팅 장치(500)의 각 구성 요소에 대하여 설명한다.
- [62] 프로세서(510)는 컴퓨팅 장치(500)의 각 구성의 전반적인 동작을 제어할 수 있다. 프로세서(510)는 CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 또는 본 개시의 기술 분야에 잘 알려진 임의의 형태의 프로세서 중 적어도 하나를 포함하여 구성될 수 있다. 또한, 프로세서(510)는 본 개시의 실시 예들에 따른 동작/방법을 실행하기

위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있다. 컴퓨팅 장치(500)는 하나 이상의 프로세서를 구비할 수 있다.

- [63] 다음으로, 메모리(520)는 각종 데이터, 명령 및/또는 정보를 저장할 수 있다. 메모리(520)는 본 개시의 실시 예들에 따른 동작/방법을 실행하기 위하여 스토리지(550)로부터 컴퓨터 프로그램(560)을 로드할 수 있다. 메모리(520)는 RAM과 같은 휘발성 메모리로 구현될 수 있을 것이나, 본 개시는 이에 한정되지 않는다.
- [64] 다음으로, 버스(530)는 컴퓨팅 장치(500)의 구성 요소 간 통신 기능을 제공할 수 있다. 버스(530)는 주소 버스(Address Bus), 데이터 버스(Data Bus) 및 제어 버스(Control Bus) 등 다양한 형태의 버스로 구현될 수 있다.
- [65] 다음으로, 통신 인터페이스(540)는 컴퓨팅 장치(500)의 유무선 인터넷 통신을 지원할 수 있다. 또한, 통신 인터페이스(540)는 인터넷 통신 외의 다양한 통신 방식을 지원할 수도 있다. 이를 위해, 통신 인터페이스(540)는 본 개시의 기술 분야에 잘 알려진 통신 모듈을 포함하여 구성될 수 있다.
- [66] 다음으로, 스토리지(550)는 하나 이상의 컴퓨터 프로그램(560)을 비임시적으로 저장할 수 있다. 스토리지(550)는 ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크, 착탈형 디스크, 또는 본 개시가 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체를 포함하여 구성될 수 있다.
- [67] 다음으로, 컴퓨터 프로그램(560)은 메모리(520)에 로드될 때 프로세서(510)로 하여금 본 개시의 다양한 실시 예들에 따른 동작/방법을 수행하도록 하는 하나 이상의 인스트럭션을 포함할 수 있다. 즉, 프로세서(510)는 로드된 하나 이상의 인스트럭션을 실행함으로써, 본 개시의 다양한 실시 예들에 따른 동작/방법을 수행할 수 있다.
- [68] 예를 들어, 컴퓨터 프로그램(560)은 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 동작, 제 1 파라미터 세트를 제 1 모델에 입력하여, 제 1 파라미터 세트에 대응하는 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 동작, 제 2 파라미터 세트를 제 1 모델에 입력하여, 제 2 파라미터 세트에 대응하는 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 동작, 제 1 및 제 2 파라미터 세트, 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 동작, 잠재 표현을 제 1 모델과 제 2 모델에 입력하여 출력된 목적 함수가 최대화되도록, 잠재 공간을 튜닝하는 동작, 및 튜닝된 잠재 공간을 디코딩하여, 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 동작을 수행하도록 할 수 있다.
- [69] 한편, 본 명세서에 개시된 실시 예들은 컴퓨터에 의해 실행 가능한 명령어를 저장하는 기록매체의 형태로 구현될 수 있다. 명령어는 프로그램 코드의 형태로 저

장될 수 있으며, 프로세서에 의해 실행되었을 때, 프로그램 모듈을 생성하여 개시된 실시 예들의 동작을 수행할 수 있다. 기록매체는 컴퓨터로 읽을 수 있는 기록매체로 구현될 수 있다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터에 의하여 해독될 수 있는 명령어가 저장된 모든 종류의 기록 매체를 포함할 수 있다. 예를 들어, ROM, RAM, 자기테이프, 자기디스크, 플래시 메모리, 광 데이터 저장 장치 등이 있을 수 있다.

[70] 본 개시의 실시 예에 따르면, 데이터베이스 파라미터에 대한 데이터 증강을 통해, 파라미터 튜닝 및 인공지능 모델의 학습을 위한 데이터 세트를 생성할 때 소요되는 시간이 절약될 수 있다. 또한 본 개시의 실시 예에 따르면, 전체 데이터베이스 파라미터의 정보를 저장원의 잠재 공간으로 압축함으로써, 데이터베이스 파라미터 중 일부 파라미터에 대해서만 튜닝이 이루어지는 대신, 전체 파라미터에 대해 튜닝이 수행될 수 있다.

[71] 상술된 내용은 본 개시를 실시하기 위한 구체적인 실시 예들이다. 본 개시는 상술된 실시 예들뿐만 아니라, 단순히 설계 변경되거나 용이하게 변경할 수 있는 실시 예들 또한 포함할 것이다. 또한, 본 개시는 상술된 실시 예들을 이용하여 용이하게 변형하여 실시할 수 있는 기술들도 포함할 것이다. 따라서, 본 개시의 범위는 상술된 실시 예들에 국한되어 정해져서는 안 되며 후술하는 특허청구범위뿐만 아니라 본 개시의 특허청구범위와 균등한 것들에 의해 정해져야 할 것이다.

[72] 한편, 본 개시는 한국 과학기술정보통신부의 (SW 스타랩) IoT 환경을 위한 고성능 플래시 메모리 스토리지 기반 인메모리 분산 DBMS 개발(과제고유번호: 2710008456, 과제번호: 2017-0-00477-007, 과제관리(전문)기관명: 정보통신기획평가원, 연구사업명: SW컴퓨팅산업원천기술개발, 과제수행기관명: 연세대학교 산학협력단, 연구기간: 2024.01.01~2024.12.31)의 일환으로 수행한 연구로부터 도출된 것이다. 본 개시의 모든 측면에서 과제 제공 주체인 한국 정보의 재산상 이익은 없다.

청구범위

- [청구항 1] 적어도 하나의 컴퓨팅 장치에 의해 수행되는 데이터베이스 파라미터 튜닝 방법에 있어서,
제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 단계;
상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 단계;
상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 단계;
상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 단계;
상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화되도록, 상기 잠재 공간을 튜닝하는 단계; 및
상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 단계를 포함하는,
데이터베이스 파라미터 튜닝 방법.
- [청구항 2] 제 1 항에 있어서,
상기 제 1 데이터베이스의 성능 지표는 상기 제 1 데이터베이스의 쿼리 처리량, 상기 제 1 데이터베이스의 데이터 처리 지연 시간, 상기 제 1 데이터베이스에서 초당 처리되는 작업의 수, 상기 제 1 데이터베이스의 쓰기 증폭 비율, 및 상기 제 1 데이터베이스의 공간 증폭 비율 중 적어도 어느 하나를 포함하는,
데이터베이스 파라미터 튜닝 방법.
- [청구항 3] 제 1 항에 있어서,
상기 제 2 파라미터 세트를 생성하는 단계는,
상기 제 1 파라미터 세트에 대해 라틴 하이퍼큐브 샘플링(Latin Hypercube Sampling; LHS)을 적용하여 상기 제 2 파라미터 세트를 생성하는 단계를 포함하는,
데이터베이스 파라미터 튜닝 방법.
- [청구항 4] 제 1 항에 있어서,
상기 입력 데이터 세트는 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트를 연결(concatenate)하여 생성된 것이고,

상기 입력 데이터 세트의 차원은 상기 제 1 및 제 2 파라미터 세트가 포함하는 파라미터의 개수 및 상기 제 1 및 제 2 성능 세트가 포함하는 성능 지표의 개수의 합인 것을 특징으로 하는,
데이터베이스 파라미터 튜닝 방법.

[청구항 5]

제 1 항에 있어서,
상기 목적 함수는 상기 제 1 데이터베이스의 성능 지표와 관련된 함수이고,
상기 잠재 공간을 튜닝하는 단계는,
상기 잠재 표현에 대해 베이지안 최적화(Bayesian optimization)를 사용하여 상기 목적 함수를 최대화하는 잠재 표현을 결정하는 단계를 포함하는,
데이터베이스 파라미터 튜닝 방법.

[청구항 6]

제 1 항에 있어서,
상기 복수의 튜닝된 파라미터를 상기 제 1 데이터베이스에 적용하는 단계를 더 포함하는,
데이터베이스 파라미터 튜닝 방법.

[청구항 7]

프로세서; 및
인스트럭션(instruction)들을 저장하는 메모리를 포함하고,
상기 인스트럭션들은 상기 프로세서에 의해 실행될 때, 상기 프로세서로 하여금,
제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 동작;
상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 동작;
상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 동작;
상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 동작;
상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화되도록, 상기 잠재 공간을 튜닝하는 동작; 및
상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 동작을 수행하도록 하는,
데이터베이스 파라미터 튜닝 시스템.

[청구항 8]

제 7 항에 있어서,

상기 입력 데이터 세트는 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트를 연결(concatenate)하여 생성된 것이고,
 상기 입력 데이터 세트의 차원은 상기 제 1 및 제 2 파라미터 세트가 포함하는 파라미터의 개수 및 상기 제 1 및 제 2 성능 세트가 포함하는 성능 지표의 개수의 합인 것을 특징으로 하는,
 데이터베이스 파라미터 튜닝 시스템.

[청구항 9]

제 7 항에 있어서,
 상기 목적 함수는 상기 제 1 데이터베이스의 성능 지표와 관련된 함수이고,
 상기 잠재 공간을 튜닝하는 동작은,
 상기 잠재 표현에 대해 베이지안 최적화(Bayesian optimization)를 사용하여 상기 목적 함수를 최대화하는 잠재 표현을 결정하는 동작을 포함하는,
 데이터베이스 파라미터 튜닝 시스템.

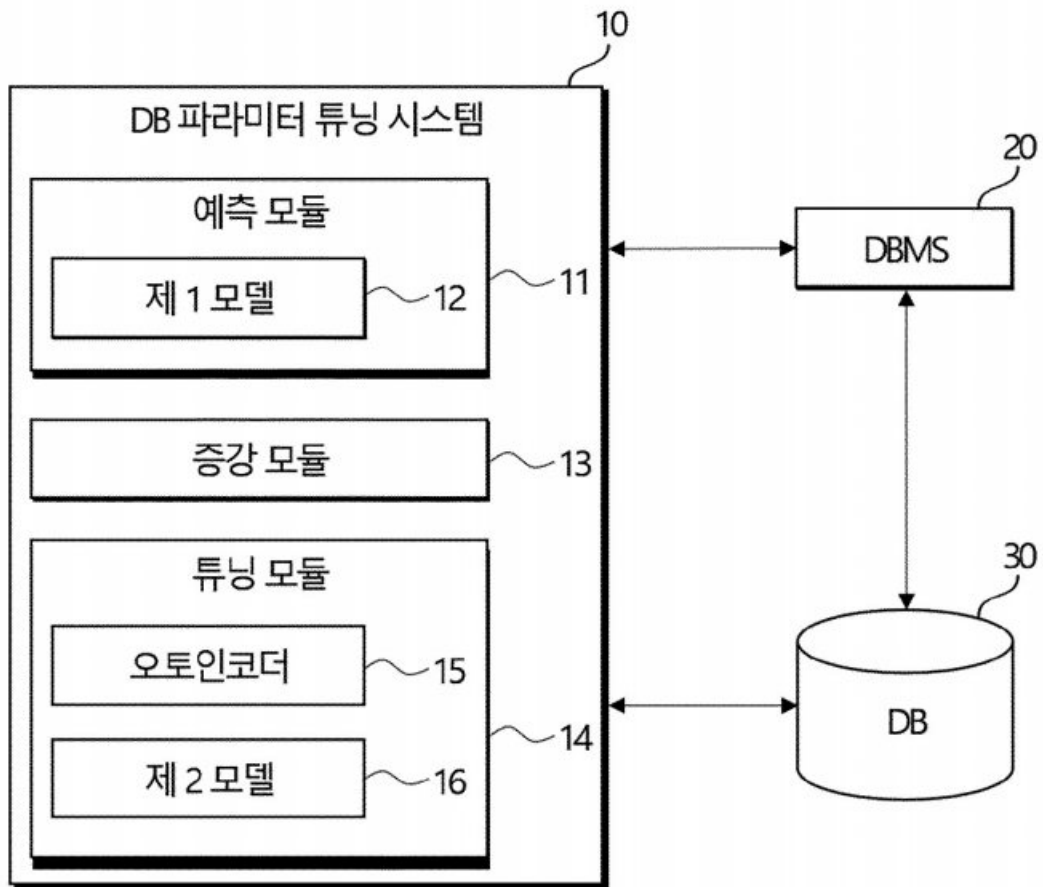
[청구항 10]

컴퓨터 프로그램 코드를 저장하는 컴퓨터로 읽을 수 있는 기록매체에 있어서, 프로세서에 의해 상기 컴퓨터 프로그램 코드가 실행될 때, 상기 프로세서는:
 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 동작;
 상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 동작;
 상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 2 파라미터 세트에 대응하는 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 동작;
 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 동작;
 상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 출력된 목적 함수가 최대화되도록, 상기 잠재 공간을 튜닝하는 동작; 및
 상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 동작을 수행하는,
 컴퓨터로 읽을 수 있는 기록매체.

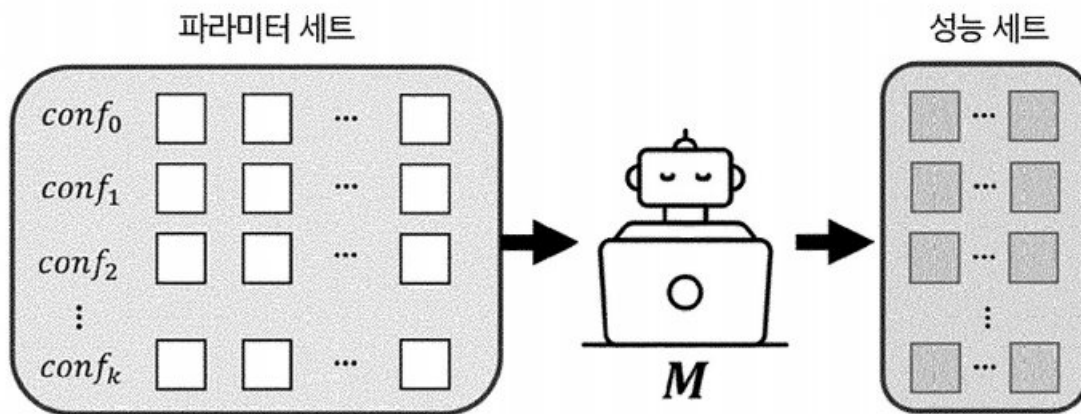
요약서

본 개시의 실시 예에 따른 데이터베이스 파라미터 튜닝 방법은 제 1 데이터베이스와 관련된 복수의 파라미터를 포함하는 제 1 파라미터 세트에 대해 데이터 증강을 수행하여 제 2 파라미터 세트를 생성하는 단계, 상기 제 1 파라미터 세트를 제 1 모델에 입력하여, 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 1 성능 세트를 출력하는 단계, 상기 제 2 파라미터 세트를 상기 제 1 모델에 입력하여, 상기 제 1 데이터베이스의 성능 지표를 포함하는 제 2 성능 세트를 출력하는 단계, 상기 제 1 및 제 2 파라미터 세트, 상기 제 1 및 제 2 성능 세트에 기반하여 생성된 입력 데이터 세트를 잠재 공간으로 인코딩하여 잠재 표현을 생성하는 단계, 상기 잠재 표현을 상기 제 1 모델과 상이한 제 2 모델에 입력하여 상기 잠재 공간을 튜닝하는 단계, 및 상기 튜닝된 잠재 공간을 디코딩하여, 상기 제 1 데이터베이스와 관련된 복수의 튜닝된 파라미터를 포함하는 출력 데이터 세트를 생성하는 단계를 포함할 수 있다.

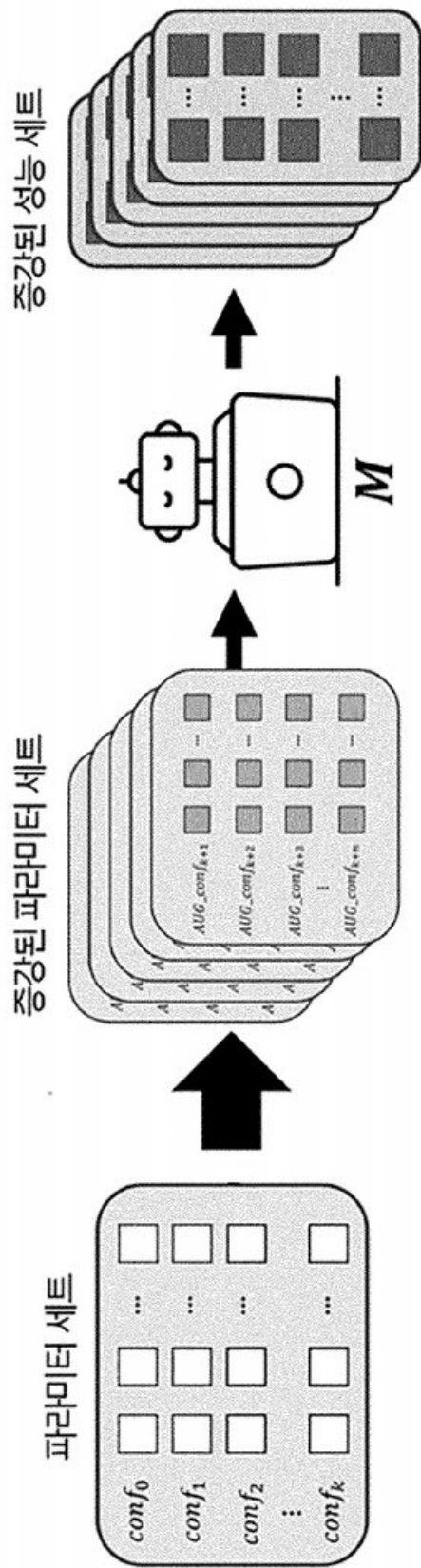
[도1]



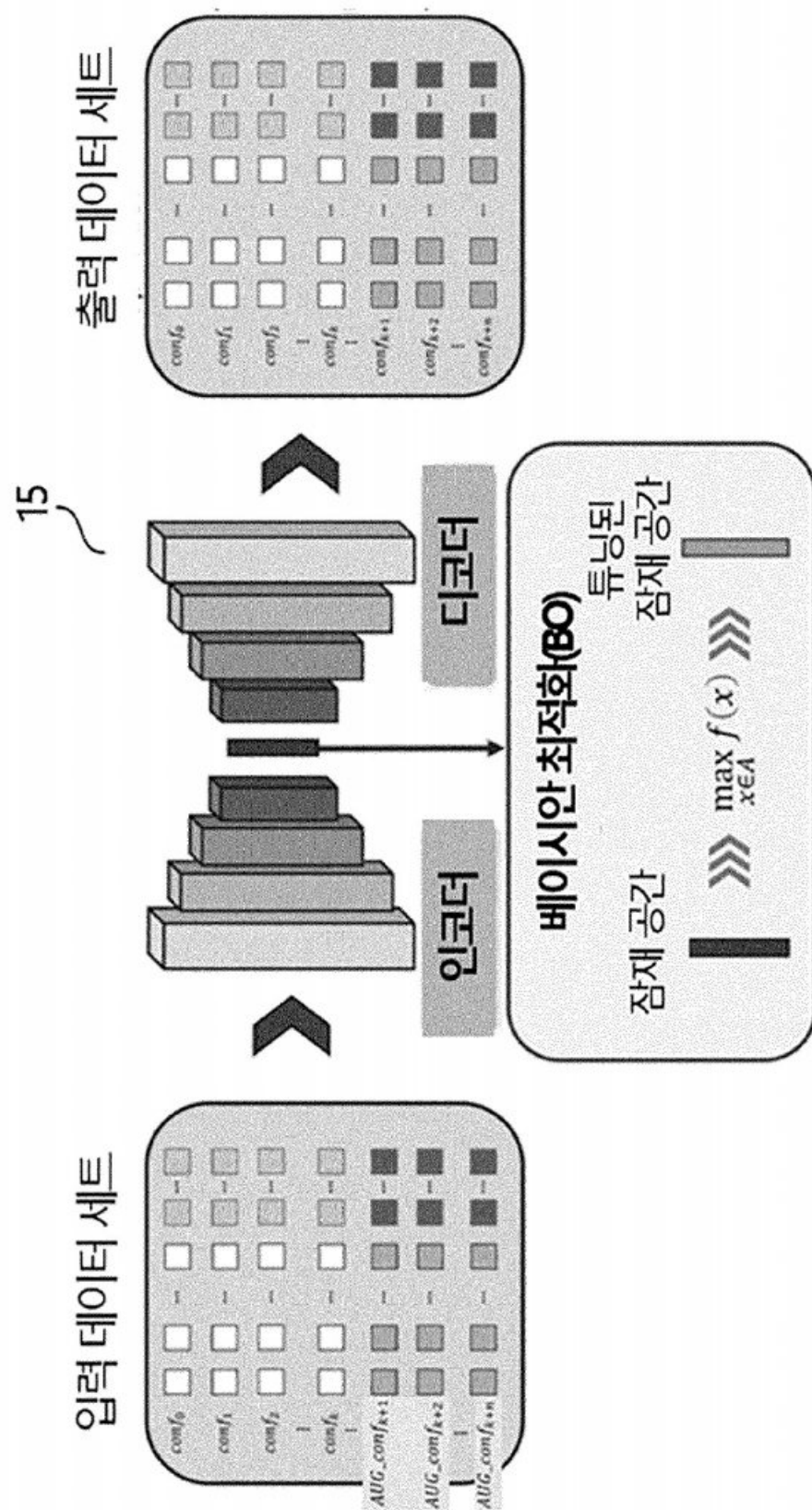
[도2]



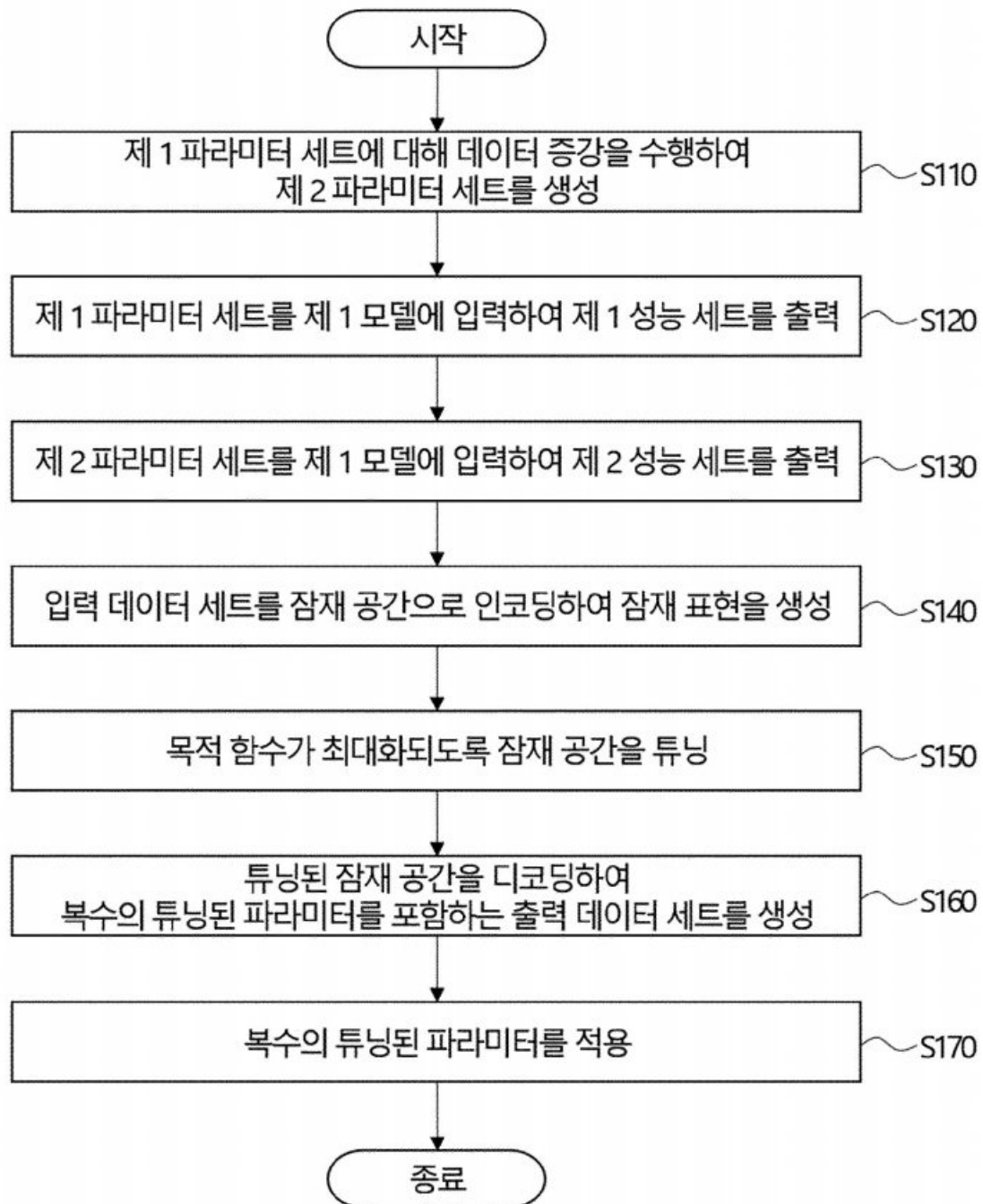
[도3]



[도4]



[도5]



[도6]

