



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0015714
(43) 공개일자 2022년02월08일

(51) 국제특허분류(Int. Cl.)

G10L 25/63 (2013.01) G06T 19/00 (2011.01)
G10L 17/02 (2013.01) G10L 17/04 (2013.01)
G10L 17/08 (2013.01) G10L 25/78 (2013.01)

(52) CPC특허분류

G10L 25/63 (2013.01)
G06T 19/006 (2013.01)

(21) 출원번호 10-2020-0096085

(22) 출원일자 2020년07월31일

심사청구일자 2020년07월31일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자

이승룡

경기도 성남시 분당구 구미로144번길 25, 102동 102호(구미동, 삼환빌라)

방재훈

경기도 용인시 기흥구 덕영대로 1732, 351호(서천동, 전자정보대학)

(74) 대리인

유미특허법인

전체 청구항 수 : 총 15 항

(54) 발명의 명칭 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법 및 그 장치

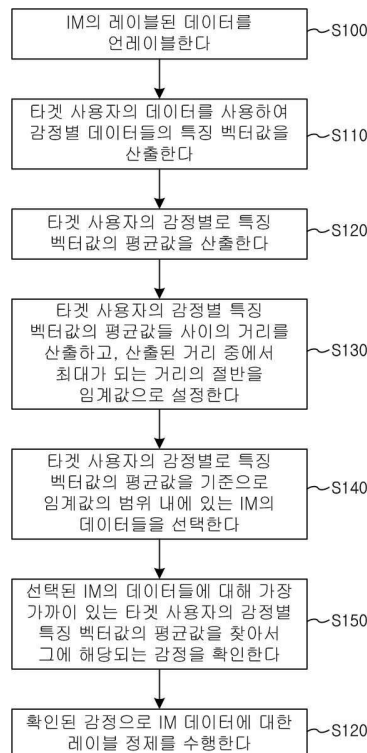
(57) 요약

개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법 및 그 장치가 개시된다.

이 방법에서, 타겟 사용자로부터 초기에 획득된 감정 음성 데이터(이하 “타겟 사용자의 데이터”로 지칭함)가 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터가 아닌 경우, 미

(뒷면에 계속)

대 표 도 - 도1



리 형성된 다수의 사용자에게 대한 초기 모델(Initial Model)(이하 “초기 모델”로 지칭함)의 레이블된 데이터를 언레이블한 데이터 중에서 상기 타겟 사용자의 데이터의 감정별 데이터와 유사한 데이터를 선택하여 강화한다. 또는 이 방법에서, 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터인 경우, 상기 초기 모델의 데이터 중에서 상기 타겟 사용자의 감정 데이터와 가장 유사한 사용자의 데이터 중 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화한다. 또는 이 방법에서, 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수 이상이면, 상기 타겟 사용자의 감정별 데이터의 개수가 불균형 상태인 경우, 오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화한다.

(52) CPC특허분류

G10L 17/02 (2013.01)

G10L 17/04 (2013.01)

G10L 17/08 (2013.01)

G10L 2025/783 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103238
과제번호	1711103238
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신기술인력양성/대학ICT연구센터육성지원
연구과제명	[정부] 2020 ITRC 지능형 의료 플랫폼 개발(4/4)
기 여 율	3/10
과제수행기관명	경희대학교 산학협력단
연구기간	2017.06.01~2020.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103071
과제번호	1711103071
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	디지털콘텐츠원천기술개발(R&D, 정보화)
연구과제명	다양한 디지털 제품의 UX 평가를 위한 Lean UX 핵심기술 및 플랫폼(4/8)
기 여 율	3/10
과제수행기관명	경희대학교 산학협력단
연구기간	2017.04.01~2024.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711099211
과제번호	1711099211
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	(유형1-2)중견연구(연평균연구비 1억원~2억원 이내)
연구과제명	Big Knowledge 관리를 위한 진화형 지식학습 기술(2/3)
기 여 율	3/10
과제수행기관명	경희대학교 산학협력단
연구기간	2019.09.01~2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711099622
과제번호	1711099622
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	한-EU 공동연구지원사업
연구과제명	스마트 환경에서 치매환자를 위한 다중 센서기반 상황인지 컴퓨팅 기술 개발
기 여 율	1/10
과제수행기관명	경희대학교 산학협력단
연구기간	2019.12.01~2020.11.30

명세서

청구범위

청구항 1

개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법으로서,

타겟 사용자로부터 초기에 획득된 감정 음성 데이터(이하 “타겟 사용자의 데이터”로 지칭함)가 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터가 아닌 경우, 미리 형성된 다수의 사용자에 대한 초기 모델(Initial Model)(이하 “초기 모델”로 지칭함)의 레이블된 데이터를 언레이블한 데이터 중에서 상기 타겟 사용자의 데이터의 감정별 데이터와 유사한 데이터를 선택하여 강화하는 단계, 또는

상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터인 경우, 상기 초기 모델의 데이터 중에서 상기 타겟 사용자의 감정 데이터와 가장 유사한 사용자의 데이터 중 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화하는 단계, 또는

상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수 이상이면, 상기 타겟 사용자의 감정별 데이터의 개수가 불균형 상태인 경우, 오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화하는 단계

를 포함하는 화자 적응 모델링 방법.

청구항 2

제1항에 있어서,

상기 초기 모델의 레이블된 데이터를 언레이블한 데이터 중에서 상기 타겟 사용자의 데이터의 감정별 데이터와 유사한 데이터를 선택하여 강화하는 단계는,

상기 타겟 사용자의 감정별 데이터의 특징 벡터값을 산출하는 단계,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값을 산출하는 단계,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리를 산출하는 단계,

산출된 거리에 기반하여 임계값을 설정하는 단계,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값을 기준으로 상기 임계값 이내의 범위 내에 속한 상기 초기 모델의 언레이블한 데이터를 선택하는 단계,

선택된 초기 모델의 언레이블한 데이터와 상기 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리에 기초하여 상기 선택된 초기 모델의 언레이블한 데이터에 해당하는 감정을 확정하는 단계, 그리고

상기 선택된 초기 모델의 언레이블한 데이터에 대해 확정된 감정으로 레이블 정제를 수행하는 단계

를 포함하는, 화자 적응 모델링 방법.

청구항 3

제2항에 있어서,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값은 MLE(Maximum Likelihood Estimation)에 의해 산출되는,

화자 적응 모델링 방법.

청구항 4

제2항에 있어서,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리는 유클리디언 거리로서 산출되는,
화자 적응 모델링 방법.

청구항 5

제2항에 있어서,
상기 임계값은 상기 산출된 거리 중 최대가 되는 거리의 절반으로 설정되는,
화자 적응 모델링 방법.

청구항 6

제1항에 있어서,
상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화하는 단계는,
상기 타겟 사용자의 감정별 데이터 각각에 대해 데이터 분포 요소를 산출하는 단계,
상기 초기 모델의 사용자별로 감정별 데이터 각각에 대해 데이터 분포 요소를 산출하는 단계,
상기 타겟 사용자의 감정별 데이터 각각에 대해 산출된 데이터 분포 요소와 상기 초기 모델의 사용자별로 감정별 데이터 각각에 대해 산출된 데이터 분포 요소 사이의 거리를 산출하는 단계,
산출된 거리에 기초하여 상기 타겟 사용자와 상기 초기 모델의 사용자별 유사성을 산출하는 단계, 그리고
산출된 유사성이 가장 높은 상기 초기 모델의 사용자의 데이터 중에서 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 상기 초기 모델의 사용자의 감정의 데이터를 선택하여 상기 타겟 사용자의 부재 데이터에 해당하는 감정의 데이터로서 강화하는 단계
를 포함하는, 화자 적응 모델링 방법.

청구항 7

제6항에 있어서,
상기 데이터 분포 요소는 중앙값, 분산, 왜도 및 첨도를 포함하는,
화자 적응 모델링 방법.

청구항 8

제1항에 있어서,
상기 오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화하는 단계는,
상기 타겟 사용자의 감정별 데이터에 대한 불균형비를 산출하는 단계 - 상기 불균형비는 상기 타겟 사용자의 감정별 데이터 중에서 가장 많은 개수와 가장 작은 개수의 비를 나타냄 -, 그리고
상기 불균형비가 미리 설정된 임계 비율 이상이어서 상기 타겟 사용자의 감정별 데이터가 불균형 상태임을 나타내는 경우, 상기 타겟 사용자의 감정별 데이터 중에서 가장 작은 개수의 감정 데이터에 대해 가상 데이터 증강 방식으로 가상 데이터를 생성하여 데이터를 강화하는 단계
를 포함하는, 화자 적응 모델링 방법.

청구항 9

제8항에 있어서,
상기 가상 데이터 증강 방식은 minority class의 데이터를 합성하여 증강시키는 SMOTE(Synthetic Minority Oversampling Technique) 방식인,
화자 적응 모델링 방법.

청구항 10

개인화된 감정 인식을 위한 강인한 화자 적응 모델링 장치로서,

미리 형성된 다수의 사용자에게 대한 초기 모델(Initial Model)(이하 “초기 모델”로 지칭함)의 레이블된 데이터를 언레이블한 데이터 중에서 타겟 사용자로부터 초기에 획득된 감정 음성 데이터(이하 “타겟 사용자의 데이터”로 지칭함)의 감정별 데이터와 유사한 데이터를 선택하여 강화하는 제1 모델링부,

상기 초기 모델의 데이터 중에서 상기 타겟 사용자의 감정 데이터와 가장 유사한 사용자의 데이터 중 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화하는 제2 모델링부,

오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화하는 제3 모델링부, 그리고

상기 타겟 사용자의 데이터가 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터가 아닌 경우 상기 제1 모델링부를 통한 데이터 강화를 수행시키거나, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터인 경우 상기 제2 모델링부를 통한 데이터 강화를 수행시키거나, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수 이상이면서, 상기 타겟 사용자의 감정별 데이터의 개수가 불균형 상태인 경우 상기 제3 모델링부를 통한 데이터 강화를 수행시키는 처리 제어부

를 포함하는 화자 적응 모델링 장치.

청구항 11

제10항에 있어서,

상기 제1 모델링부는,

상기 초기 모델의 레이블된 데이터를 언레이블시키거나, 또는 상기 초기 모델의 언레이블한 데이터 중 상기 타겟 사용자의 데이터로서 선택되는 경우 선택된 데이터에 대한 감정 레이블을 수행하는 레이블링기,

상기 타겟 사용자의 감정별 데이터의 특징 벡터값을 산출하는 특징 벡터값 산출기,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값을 산출하는 평균값 산출기,

상기 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리를 산출하거나, 또는 상기 초기 모델의 언레이블한 데이터와 상기 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리를 산출하는 거리 산출기,

상기 거리 산출기에 의해 산출된 거리에 기반하여 임계값을 산출하는 임계값 산출기,

상기 평균값 산출기에 의해 산출된 평균값을 기준으로 상기 임계값 이내의 범위 내에 속한 상기 초기 모델의 언레이블한 데이터를 선택하는 데이터 선택기, 그리고

상기 데이터 선택기에 의해 선택된 초기 모델의 언레이블한 데이터와 상기 평균값 산출기에 의해 산출된 평균값 사이의 거리에 기초하여 상기 선택된 초기 모델의 언레이블한 데이터에 해당하는 감정을 확정하는 감정 확정기

를 포함하는, 화자 적응 모델링 장치.

청구항 12

제11항에 있어서,

상기 임계값 산출기는 상기 거리 산출기에 의해 산출된 거리 중 최대가 되는 거리의 절반을 상기 임계값으로 산출하는,

화자 적응 모델링 장치.

청구항 13

제10항에 있어서,

상기 제2 모델링부는,

상기 타겟 사용자의 감정별 데이터 각각에 대해 데이터 분포 요소 및 상기 초기 모델의 사용자별로 감정별 데이터 각각에 대해 데이터 분포 요소를 산출하는 분포 요소 산출기,

상기 타겟 사용자의 감정별 데이터 각각에 대해 산출된 데이터 분포 요소와 상기 초기 모델의 사용자별로 감정별 데이터 각각에 대해 산출된 데이터 분포 요소 사이의 거리를 산출하는 거리 산출기,

상기 거리 산출기에 의해 산출된 거리에 기초하여 상기 타겟 사용자와 상기 초기 모델의 사용자별 유사성을 산출하는 유사성 산출기, 그리고,

산출된 유사성이 가장 높은 상기 초기 모델의 사용자의 데이터 중에서 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 상기 초기 모델의 사용자의 감정의 데이터를 선택하여 상기 타겟 사용자의 부재 데이터에 해당하는 감정의 데이터로서 강화하는 데이터 선택기

를 포함하는, 화자 적응 모델링 장치.

청구항 14

제13항에 있어서,

상기 데이터 분포 요소는 중앙값, 분산, 왜도 및 첨도를 포함하는,

화자 적응 모델링 장치.

청구항 15

제10항에 있어서,

상기 제3 모델링부는,

상기 타겟 사용자의 감정별 데이터에 대한 불균형비를 산출하는 불균형비 산출기 — 상기 불균형비는 상기 타겟 사용자의 감정별 데이터 중에서 가장 많은 개수와 가장 작은 개수의 비를 나타냄 —, 그리고

상기 불균형비가 미리 설정된 임계 비율 이상이어서 상기 타겟 사용자의 감정별 데이터가 불균형 상태임을 나타내는 경우, 상기 타겟 사용자의 감정별 데이터 중에서 가장 작은 개수의 감정 데이터에 대해 가상 데이터 증강 방식으로 가상 데이터를 생성하여 데이터를 강화하는 데이터 증강 수행기

를 포함하는, 화자 적응 모델링 장치.

발명의 설명

기술 분야

[0001] 본 발명은 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 전통적인 감정인식 기술은 다수의 사용자로부터 감정 음성 데이터를 수집하여 이를 기계 학습 알고리즘을 통해 학습하고, 학습에 의해 생성된 모델을 모든 사용자에게 동일하게 적용하는 것이다.

[0003] 그런데, 이러한 기술은 모든 사용자에게 고른 정확도를 나타내기 어렵고 정확도 편차가 크다는 단점이 존재한다. 따라서, 최근에는 사용자마다 개인화된 모델을 생성하고 이를 사용자에게 적용하여 제공하는 개인화된 감정 인식 기술이 대두되었다.

[0004] 기존의 개인화된 감정 인식 기술은 다수의 사용자로부터 구축되어진 감정 인식 모델을 중심으로 타겟 사용자로부터 수집되는 레이블(Label)된 감정 음성 데이터를 반영하여 선형 파라미터 변경 및 레이블 정제(Label Refinement) 작업을 통해 모델을 변경하는 방식으로 동작한다.

[0005] 그러나, 기존의 개인화된 감정 인식 기술에서는 정확도 높은 개인화된 모델로 변경하기 위해서 충분한 데이터를 요구하며, 사용자의 감정 음성 데이터가 균일하게 입력되어야지만 최상의 성능을 나타낼 수 있다. 그런데, 개인화된 감정 인식 초기 단계에서는 타겟 사용자의 레이블된 데이터를 충분히 확보하기 어렵고 또한 불균형 환경에 빠지기 쉬워 개인화된 모델을 생성하기 어려운 콜드 스타트 문제(Cold-Start Problem) 빠지기 쉽다.

[0006] 따라서, 기존의 개인화된 감정 인식 기술에서 콜드 스타트 문제를 해결하기 위한 강인한 개인화된 감정 인식 모델 생성 기술이 요구된다.

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 과제는 개인화된 음성 기반 감정 인식 과정에서 타겟 사용자의 데이터가 충분히 모이지 않은 작은 데이터, 부재 데이터, 불균형 데이터 환경에서도 개인화 훈련 모델 생성에 충분하고 균형잡힌 개인화된 훈련 데이터셋을 도출할 수 있는 강인한 화자 적응 모델링 방법 및 그 장치를 제공하는 것이다.

과제의 해결 수단

[0008] 상기한 바와 같은 본 발명의 과제를 달성하고, 후술하는 본 발명의 특징적인 효과를 실현하기 위한, 본 발명의 특징적인 구성은 하기와 같다.

[0009] 본 발명의 일 측면에 따르면, 화자 적응 모델링 방법이 제공되며, 이 방법은,

[0010] 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법으로서, 타겟 사용자로부터 초기에 획득된 감정 음성 데이터(이하 “타겟 사용자의 데이터”로 지칭함)가 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터가 아닌 경우, 미리 형성된 다수의 사용자에 대한 초기 모델(Initial Model)(이하 “초기 모델”로 지칭함)의 레이블된 데이터를 언레이블한 데이터 중에서 상기 타겟 사용자의 데이터의 감정별 데이터와 유사한 데이터를 선택하여 강화하는 단계, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터인 경우, 상기 초기 모델의 데이터 중에서 상기 타겟 사용자의 감정 데이터와 가장 유사한 사용자의 데이터 중 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화하는 단계, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수 이상이면, 상기 타겟 사용자의 감정별 데이터의 개수가 불균형 상태인 경우, 오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화하는 단계를 포함한다.

[0011] 본 발명의 다른 측면에 따르면, 화자 적응 모델링 장치가 제공되며, 이 장치는,

[0012] 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 장치로서, 미리 형성된 다수의 사용자에 대한 초기 모델(Initial Model)(이하 “초기 모델”로 지칭함)의 레이블된 데이터를 언레이블한 데이터 중에서 타겟 사용자로부터 초기에 획득된 감정 음성 데이터(이하 “타겟 사용자의 데이터”로 지칭함)의 감정별 데이터와 유사한 데이터를 선택하여 강화하는 제1 모델링부, 상기 초기 모델의 데이터 중에서 상기 타겟 사용자의 감정 데이터와 가장 유사한 사용자의 데이터 중 상기 타겟 사용자의 부재 데이터에 해당하는 감정에 대응하는 데이터를 선택하여 강화하는 제2 모델링부, 오버 샘플링 알고리즘을 사용한 가상 데이터 증강 방식으로 상기 타겟 사용자의 감정별 데이터를 강화하는 제3 모델링부, 그리고 상기 타겟 사용자의 데이터가 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터가 아닌 경우 상기 제1 모델링부를 통한 데이터 강화를 수행시키거나, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수보다 적으면서, 상기 타겟 사용자의 데이터의 감정 데이터가 부재 데이터인 경우 상기 제2 모델링부를 통한 데이터 강화를 수행시키거나, 또는 상기 타겟 사용자의 데이터가 상기 미리 설정된 임계 개수 이상이면, 상기 타겟 사용자의 감정별 데이터의 개수가 불균형 상태인 경우 상기 제3 모델링부를 통한 데이터 강화를 수행시키는 처리 제어부를 포함한다.

발명의 효과

[0013] 본 발명에 따르면, 개인화된 음성 기반 감정 인식 과정에서 타겟 사용자의 데이터가 충분히 모이지 않은 작은 데이터, 부재 데이터, 불균형 데이터 환경에서도 개인화 훈련 모델 생성에 충분하고 균형잡힌 개인화된 훈련 데이터셋을 도출할 수 있다.

[0014] 또한, 개인화된 모델링 기술을 통해 보다 정확한 감정 인지가 가능해짐으로써, 다양한 IoT 지능형 서비스 분야에 적용될 수 있다.

[0015] 또한, 로봇 및 콜 센터에 적용하여 입력되는 사용자들의 감정 음성을 축적해 업데이트함으로써 개인화된 감정 인식 서비스 제공이 가능하다.

[0016] 또한, 이를 활용한 조울증/우울증과 같은 심리 치료에 필요한 시장에 측정 기법으로 활용이 가능하다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 실시예에 따른 타겟 사용자의 소량의 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이다.

도 2는 도 1에 따른 흐름을 음성 데이터의 분포도를 사용하여 설명한 도면이다.

도 3은 도 1에 도시된 화자 적응 모델링 방법에서 IM의 레이블된 감정별 데이터에 대해 언레이블된 데이터의 예를 도시한 도면이다.

도 4는 도 1에 도시된 화자 적응 모델링 방법에 따라 타겟 사용자에게 대해 획득된 화남(Anger), 슬픔(Sadness), 행복(Happiness)의 감정 각각에 대한 특징 벡터값의 평균값이 MLE 계산에 의해 산출된 예를 도시한 도면이다.

도 5는 도 1에 도시된 화자 적응 모델링 방법에 따라 임계값을 설정하여 데이터를 선택하는 예를 도시한 도면이다.

도 6은 도 1에 도시된 화자 적응 모델링 방법에 따라 강화된 타겟 사용자 데이터(우측)와 기존 방식에 따라 확보된 타겟 사용자 데이터의 예를 도시한 도면이다.

도 7은 본 발명의 실시예에 따른 타겟 사용자의 부재 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이다.

도 8은 도 7에 따른 흐름을 음성 데이터의 분포도를 사용하여 설명한 도면이다.

도 9는 도 7에 도시된 화자 적응 모델링 방법에 따라 타겟 사용자의 데이터와 IM 데이터 각각에 대해 데이터 분포 요소를 산출한 예를 도시한 도면이다.

도 10은 도 7에 도시된 화자 적응 모델링 방법에 따라 타겟 사용자의 데이터와 IM 데이터 각각에 대해 산출된 데이터 분포 요소에 기초한 유사성 산출 예를 도시한 도면이다.

도 11은 본 발명의 실시예에 따른 타겟 사용자의 불균형한 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이다.

도 12는 본 발명의 실시예에 따른 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법의 개략적인 흐름도이다.

도 13은 본 발명의 실시예에 따른 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 장치의 개략적인 구성 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0019] 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 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0020] 본 발명에서 설명하는 장치들은 적어도 하나의 프로세서, 메모리 장치, 통신 장치 등을 포함하는 하드웨어로 구성되고, 지정된 장소에 하드웨어와 결합되어 실행되는 프로그램이 저장된다. 하드웨어는 본 발명의 방법을 실행할 수 있는 구성과 성능을 가진다. 프로그램은 도면들을 참고로 설명한 본 발명의 동작 방법을 구현한 명령어(instructions)를 포함하고, 프로세서와 메모리 장치 등의 하드웨어와 결합하여 본 발명을 실행한다.

[0021] 이하, 본 발명의 실시예에 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법에 대해 설명한다.

[0022] 본 발명의 실시예에서는 개인화된 감정 인식을 위한 강인한 화자 적응 모델링으로 3가지 방식을 제공한다. 구체적으로, 첫 번째 방식은 타겟 사용자의 감정 음성 데이터가 소량인 경우에 대한 방식이고, 두 번째 방식은 타겟 사용자의 일부 감정 음성 데이터가 부재인 경우에 대한 방식이며, 세 번째 방식은 타겟 사용자의 감정별 음성 데이터가 불균형 상태인 경우에 대한 방식이다.

[0023] 먼저, 첫 번째 방식에 대해 설명한다. 즉, 타겟 사용자에게 대해 수집된 감정 음성 데이터가 소량인 경우에 수행되는 강인한 화자 적응 모델링 방식에 대한 것이다. 이 방식은 소량의 데이터(Small Data) 환경에서 기존에 구축되어 있던 다수의 사용자 초기 모델에서 타겟 사용자와 유사한 음성을 선택한 후 레이블을 정제하여 혼련 데이터셋, 즉 타겟 사용자 중심의 실제(Real-case) 데이터로 선택하는 방식이다.

[0024] 도 1은 본 발명의 실시예에 따른 타겟 사용자의 소량의 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이고, 도 2는 도 1에 따른 흐름을 음성 데이터의 분포도를 사용하여 설명한 도면이다. 설명 전에, 본 발명의 실시예에서는 이미 다수의 사용자에게에 대한 초기 모델(Initial Model, IM)(11)이 구축되어 IM 데이터베이스(10)에 저장되어 있고, 타겟 사용자에게에 대해 초기에 획득된 감정 음성 데이터(21)가 미리 획득되어 타겟 사용자 DB(20)에 저장되어 있는 것으로 가정한다.

[0025] 도 1 및 도 2를 참조하면, 먼저, IM(11)의 레이블된 데이터를 언레이블(UN Label)한다(S100). 즉, IM(11)에 해당되는 다른 사용자들의 감정 정보를 제거한다. 그 이유로는 음성으로 표출하는 감정은 각 사용자마다 다르게 표현될 수 있기 때문이다. 예를 들어, 특정 사용자의 뉴트럴(Neutral) 감정의 음성이 다른 사용자의 감정 음성 정보에서 화남(Anger) 감정의 음성과 유사할 수 있기 때문이다. 따라서, 다른 사용자와 유사한 음성으로 타겟 사용자의 혼련 데이터 세트를 강화할 때 IM(11)에서 레이블을 제거하여 활용하는 것이 바람직하다. 도 3을 참조하면, 좌측의 IM(11)의 레이블된 감정별 데이터에 대해 언레이블된 데이터의 예가 우측에 도시되어 있다.

[0026] 다음, 타겟 사용자 데이터베이스(20)에 저장되어 있는 획득된 타겟 사용자의 데이터(21)를 사용하여 감정별 데이터들의 특징 벡터값(TfeatureVector)을 산출한다(S110). 여기서, 감정별 데이터들의 특징 벡터값을 산출하는 방법은 이미 잘 알려져 있으므로 여기에서는 상세하게 설명하지 않는다.

[0027] 그 후, 타겟 사용자의 감정별로 특징 벡터값의 평균값을 산출한다(S120). 이러한 특징 벡터값의 평균값은 MLE(Maximum Likelihood Estimation) 기법을 사용하여 산출될 수 있다. 예를 들어, 타겟 사용자의 감정별 특징 벡터값의 MLE(TMLE)는 [수학식 1]과 같이 산출될 수 있다.

[0028] [수학식 1]

$$TMLE_{ei} = \frac{1}{N} \sum_{j=1}^N TfeatureVector_{ji}$$

[0029]

[0030] 여기서, e는 대응하는 감정의 인덱스이고, I는 특징 벡터의 인덱스이며, j는 데이터의 인덱스이고, N은 데이터의 개수이며, TfeatureVector는 타겟 사용자의 감정별 데이터들의 특징 벡터값이다.

[0031] 도 4를 참조하면, 타겟 사용자에게에 대해 획득된 화남(Anger), 슬픔(Sadness), 행복(Happiness)의 감정 각각에 대한 특징 벡터값의 평균값이 MLE 계산에 의해 산출됨을 알 수 있다.

[0032] 다음, 타겟 사용자의 감정별 특징 벡터값의 평균값들 사이의 거리를 산출하고, 산출된 거리 중에서 최대가 되는 거리의 절반을 임계값으로 설정한다(S130). 여기서, 감정별 특징 벡터값의 평균값들 사이의 거리는 다음의 [수학식 2]와 같이 유클리디언 거리(Euclidean Distance)로서 산출될 수 있다.

[0033] [수학식 2]

$$d(TMLE_{ei}, TMLE_{ej}) = \sqrt{\sum_{k=1}^{FN} (TMLE_{eik} - TMLE_{ejk})^2}$$

[0034]

[0035] 여기서, TMLE는 [수학식 1]에서 산출된 타겟 사용자의 감정별 특징 벡터값의 평균값이고, e_i 및 e_j는 서로 다른 감정 인덱스이며, k는 특징 벡터의 인덱스이고, FN은 특징의 개수이다.

[0036] 이와 같이, 타겟 사용자의 감정별 특징 벡터값의 평균값들 사이의 거리가 산출되면 다음의 [수학식 3]을 사용하여 산출된 거리 중에서 최대값의 절반, 즉 1/2의 거리를 임계값(Maximum Threshold Distance, MTD)으로 산출한

다.

[0037] [수학식 3]

$$MTD(TMLE_{ei}) = \frac{1}{2} \operatorname{argmax}(d(TMLE_{e_i}, TMLE_{e_j}, \dots))$$

[0038]

[0039] 그 후, 상기 단계(S120)에서 산출된 타겟 사용자의 감정별로 특징 벡터값의 평균값을 기준으로 상기 단계(S130)에서 산출된 임계값의 범위 내에 있는 IM(11)의 데이터들을 선택한다(S140).

[0040] 여기서, 임계값의 범위 내에 있는 IM(11)의 데이터를 선택하기 위해 감정별 특징 벡터값의 평균값과 IM(11)의 데이터 사이의 거리가 다음의 [수학식 4]와 같이 유클리디언 거리로서 산출될 수 있다.

[0041] [수학식 4]

$$d(TMLE_{e_i}, IDS_m) = \sqrt{\sum_{k=1}^{FN} (TMLE_{e_{ik}} - IDS_k)^2}$$

[0042]

[0043] 여기서, TMLE는 [수학식 1]에서 산출된 타겟 사용자의 감정별 특징 벡터값의 평균값이고, e_i 는 감정 인덱스이며, k 는 특징 벡터의 인덱스이고, IDS_m 은 언레이블된 IM(11)의 데이터(Initial Data Set, IDS)이며, FN은 특징의 개수이다.

[0044] 따라서, [수학식 4]에 의해 산출된 거리($d(TMLE_{e_i}, IDS_m)$)가 상기 단계(S130)에서 산출된 임계값($MTD(TMLE_{ei})$) 내에 속하는 경우 해당 IM(11)의 데이터가 타겟 사용자의 음성 데이터로서 선택되게 된다. 구체적으로, [수학식 4]에 의해 비교한 후, [수학식 4]에 의해 산출된 거리($d(TMLE_{e_i}, IDS_m)$)가 임계값($MTD(TMLE_{ei})$)보다 큰 경우에는 임계값 범위를 벗어나므로 해당되는 IM(11)의 데이터는 폐기되고, [수학식 4]에 의해 산출된 거리($d(TMLE_{e_i}, IDS_m)$)가 임계값($MTD(TMLE_{ei})$) 이하인 경우에는 임계값 범위 내에 있는 것이므로 해당되는 IM(11)의 데이터가 선택된다.

[0045] 그 후, 선택된 IM(11)의 데이터들에 대해 가장 가까이 있는 타겟 사용자의 감정별 특징 벡터값의 평균값을 찾아서 그에 해당되는 감정을 확인한다(S150). 이러한 단계는 전술한 단계(S140)과 동시에 수행될 수 있다. 예를 들어, 도 5에 도시된 바와 같이, IM(11)의 데이터와 타겟 사용자의 감정별 특징 벡터값의 평균값 사이의 거리가 [수학식 4]에 의해 산출되고, 산출된 거리와 임계값의 비교에 의해 IM(11) 데이터를 선택할 때 동시에 산출된 거리와 임계값의 차이가 최소일 때의 감정을 확인함으로써 동시에 수행될 수 있다.

[0046] 마지막으로, 상기 단계(S150)서 확인된 감정으로 IM(11) 데이터에 대한 레이블 정제를 수행한다(S160).

[0047] 이와 같이, 본 발명의 실시예에 따르면, 타겟 사용자에 대해 획득된 초기 감정 음성 데이터가 소량인 경우 다수의 사용자 초기 모델인 IM(11)의 언레이블된 데이터 중에서 타겟 사용자의 감정별로 유사한 데이터를 선택하여 강화할 수 있다.

[0048] 도 6을 참조하면, 타겟 사용자 데이터(21)에 대해 기존 방식에 따라 확보된 타겟 사용자 데이터(좌측)와 본 방식에 따라 강화된 타겟 사용자 데이터(우측)의 예가 도시되어 있다. 도 6에서 알 수 있는 바와 같이, 기존의 방식에 비해 본 방식에 따라 확보된 데이터의 수량이 더 많아서, 결과적으로 개인화된 감정 인식을 위한 화자 적응 모델링이 강화되었음을 알 수 있다.

[0049] 다음, 두 번째 방식에 대해 설명한다. 즉, 타겟 사용자에 대해 수집된 감정 음성 데이터 중 특정 감정에 대한 데이터가 부재인 경우에 수행되는 강인한 화자 적응 모델링 방식에 대한 것이다. 일반적으로, 사람은 일상생활에서 같은 비율로 감정을 표현하지 않는다. 타겟 사용자가 특정 감정을 오랫동안 표현하지 않으면 해당 감정 표현에 대한 표본 없이 훈련 모델을 생성해야만 한다. 이 경우에는 특정 감정은 시스템에서 인식되지 않으며 정확도가 0%가 된다. 따라서, 이 방식은 IM(11)에 있는 다양한 사용자의 음성 중 타겟 사용자와 유사한 음성 패턴을 가지고 있는 사용자의 음성이 부재 감정 데이터와 유사하다고 가정하여 해당 사용자의 감정 음성 데이터

로 대체하여 강화하는 방식이다.

[0050] 도 7은 본 발명의 실시예에 따른 타겟 사용자의 부채 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이고, 도 8은 도 7에 따른 흐름을 음성 데이터의 분포도를 사용하여 설명한 도면이다. 마찬가지로, 설명 전에, 이미 다수의 사용자에게 대한 초기 모델(Initial Model, IM)(11)이 구축되어 IM 데이터베이스(10)에 저장되어 있고, 타겟 사용자에게 대해 초기에 획득된 감정 음성 데이터(21)가 미리 획득되어 타겟 사용자 DB(20)에 저장되어 있는 것으로 가정한다.

[0051] 도 7 및 도 8을 참조하면, 먼저, 타겟 사용자의 감정 음성 데이터(21)별로 데이터 분포 요소들을 산출한다(S200). 예를 들어, 데이터 분포 요소로는 중앙값(Median), 분산(Variance), 왜도(Skewness), 첨도(Kurtosis)가 사용되지만, 본 발명에서는 이것으로만 한정하는 것은 아니다.

[0052] 구체적으로는, [수학식 5]를 사용하여, 타겟 사용자의 감정 음성 데이터 각각에 대해 중앙값, 분산, 왜도, 첨도를 산출한다.

[0053] [수학식 5]

$$\begin{aligned} \text{Median}(\text{SortedFeatureVector}_{fi}) &= \text{SortedFeatureVector}_{fi_{N/2}} \\ \text{Variance}(\text{FeatureVector}_{fi}) &= \frac{1}{N} \sum_{k=1}^N (\text{FeatureVector}_{fi_k} - \text{means}_{fi})^2 \\ \text{Kurtosis}(\text{FeatureVector}_{fi}) &= \frac{\frac{1}{N} \sum_{k=1}^N (\text{FeatureVector}_{fi_k} - \text{means}_{fi})^4}{\left(\frac{1}{N} \sum_{k=1}^N (\text{FeatureVector}_{fi_k} - \text{means}_{fi})^2\right)^2} - 3 \\ \text{Skewness}(\text{FeatureVector}_{fi}) &= \frac{\frac{1}{N} \sum_{k=1}^N (\text{FeatureVector}_{fi_k} - \text{means}_{fi})^3}{\left(\frac{1}{N} \sum_{k=1}^N (\text{FeatureVector}_{fi_k} - \text{means}_{fi})^2\right)^{\frac{3}{2}}} \end{aligned}$$

[0054]

[0055] 여기서, fi는 특징 벡터의 인덱스이고, FeatureVector_{fi}는 특정 특징 값이며, mean은 FeatureVector_{fi}의 평균값, 즉, 전술한 TMLE와 같고, N은 데이터의 개수이다.

[0056] 다음, IM(11)의 데이터 중에서 타겟 사용자의 데이터(21)에 존재하는 감정에 대응되는 데이터 각각에 대해 데이터 분포 요소들을 산출한다(S210). 즉, IM(11)의 감정 음성 데이터 중에서 타겟 사용자의 데이터(21)에 존재하는 감정 데이터의 데이터 분포 요소들과 IM(11)에 있는 다른 사용자들의 감정 음성 데이터의 데이터 분포 요소들을 비교하여 유사성(Similarity)을 찾는 것이다.

[0057] 여기서의 데이터 분포 요소들은 상기 단계(S200)에서의 데이터 분포 요소와 동일해야 한다. 즉, 여기서의 데이터 분포 요소는 중앙값, 분산, 왜도, 첨도이다. 마찬가지로, 전술한 [수학식 5]가 사용되어 IM(11)의 데이터 각각에 대한 데이터 분포 요소들이 산출될 수 있다.

[0058] 도 9를 참조하면, 타겟 사용자의 감정 음성 데이터에 화남(Anger)과 슬픔(Sadness)의 감정들에 대한 데이터만이 존재하고, 다른 감정, 예를 들어 행복(Happiness) 감정에 대한 데이터가 부재인 경우, 타겟 사용자의 화남과 슬픔 감정의 데이터를 각각에 대한 데이터 분포 요소들이 각각 산출되고, 또한 IM(11)의 데이터들 중에서 역시 화남(Anger)과 슬픔(Sadness) 감정의 데이터들에 대해서만 데이터 분포 요소들이 각각 산출됨을 알 수 있다.

[0059] 그 후, 감정별로 타겟 사용자 데이터의 데이터 분포 요소들과 IM(11)의 데이터의 데이터 분포 요소들 사이의 유사성을 산출한다(S220). 여기서, 감정별 데이터 분포 요소들 사이의 유사성은 유클리디언 거리에 기초하여 산출될 수 있다.

[0060] 예를 들어, 다음의 [수학식 6] 및 [수학식 7]을 사용하여 타겟 사용자의 데이터와 다른 사용자의 데이터 사이의 유사성이 산출될 수 있다.

[0061] [수학식 6]

$$d(TDDF_{e_i}, IDSDDF_{u_{e_i}}) = \sqrt{\sum_{k=1}^{DFN} (TDDF_{e_i k} - IDSDDF_{u_{e_i} k})^2}$$

[0062]

[0063] [수학식 7]

$$Similarity(TDDF_{e_i}, IDSDDF_{u_{ei}}) = \frac{1}{1+d(TDDF_{e_i}, IDSDDF_{u_{ei}})} * 100$$

[0064]

[0065] 여기서, $TDDF_{e_i}$ 는 타겟 사용자 데이터의 데이터 분포 요소이고, $IDSDDF_{u_{ei}}$ 는 IM(11) 데이터의 데이터 분포 요소이며, DFN은 데이터 분포 요소의 개수이다.

[0066] 즉, [수학식 6]을 사용하여 타겟 사용자 데이터(21)의 감정별 데이터 분포 요소와 IM(11) 데이터의 대응하는 감정별 데이터 분포 요소 사이의 유클리디언 거리를 산출한 후, [수학식 7]을 사용하여 타겟 사용자 데이터(21)의 감정별 데이터 분포 요소와 IM(11) 데이터의 대응하는 감정별 데이터 분포 요소 사이의 유사성을 산출한다.

[0067] 도 9의 예에서, 타겟 사용자의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들과 사용자1(User1)과 사용자2(User2) 각각의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들 사이에 산출된 유사성의 예가 도 10에 도시되어 있다. 도 10을 참조하면, (a)는 타겟 사용자의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들과 사용자1(User1)의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들 사이에 산출된 유사성으로 10.48이고, (b)는 타겟 사용자의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들과 사용자1(User2)의 화남과 슬픔 각각에 대해 산출된 데이터 분포 요소들 사이에 산출된 유사성으로 12.14임을 알 수 있다. 따라서, 타겟 사용자의 감정 음성 데이터와 유사한 사용자는 사용자2(User2)임을 알 수 있다.

[0068] 그 후, 산출된 유사성을 통해 타겟 사용자와 가장 유사한 감정 음성을 갖는 사용자의 감정 음성 데이터 중에서 타겟 사용자의 부재 감정에 해당되는 감정의 데이터를 타겟 사용자의 부재 감정의 데이터로서 선택한다(S230).

[0069] 도 10의 예를 참조하면, 타겟 사용자의 감정 음성과 가장 유사한 사용자가 사용자2(User2)이므로, 타겟 사용자의 부재 감정에 대응하는 사용자2(User2)의 감정 데이터가 타겟 사용자의 부재 감정의 데이터로서 선택된다. 도 9의 예를 참조하면, 타겟 사용자는 행복에 대한 감정 데이터가 부재이므로, IM(11) 데이터 중에서 타겟 사용자의 감정 음성과 가장 유사한 사용자2(User2)의 행복 감정에 해당하는 데이터가 타겟 사용자의 부재 감정인 행복 감정에 대한 데이터로서 확보되는 것이다. 이를 통해, 타겟 사용자의 행복 감정에 대한 데이터는 부재 데이터가 아닌 것으로 될 수 있다.

[0070] 이와 같이, 본 발명의 실시예에 따르면, 타겟 사용자에 대해 획득된 초기 감정 음성 데이터가 부재 감정 데이터인 경우 다수의 사용자 초기 모델인 IM(11)의 사용자들 중에서 데이터 분포 요소들 기반으로 타겟 사용자의 감정 음성과 가장 유사한 사용자의 감정 데이터 중에서 타겟 사용자의 부재 감정에 대응하는 감정의 데이터를 타겟 사용자의 부재 감정의 데이터로서 선택하여 강화할 수 있다.

[0071] 다음, 세 번째 방식에 대해 설명한다. 즉, 타겟 사용자에 대해 수집된 감정별 음성 데이터가 불균형 상태인 경우에 수행되는 강인한 화자 적응 모델링 방식에 대한 것이다. 이 방식은 타겟 사용자의 감정별 음성 데이터의 수량이 음성 인식을 위한 기계학습의 훈련셋으로 사용하기에 적합하지 않은 불균형한 상태인 경우에 균형 상태를 만들기 위해 수량이 적은 감정의 데이터를 가상 데이터로서 생성하는 방식이다.

[0072] 도 11은 본 발명의 실시예에 따른 타겟 사용자의 불균형한 감정 음성 데이터에 대한 화자 적응 모델링 방법의 개략적인 흐름도이다. 설명 전에, 본 발명의 실시예에서는 이미 다수의 사용자에 대한 초기 모델(Initial Model, IM)(11)이 구축되어 IM 데이터베이스(10)에 저장되어 있고, 타겟 사용자에 대해 초기에 획득된 감정 음성 데이터(21)가 미리 획득되어 타겟 사용자 DB(20)에 저장되어 있는 것으로 가정한다.

[0073] 도 11을 참조하면, 먼저, 타겟 사용자의 감정별 데이터의 수량에 기초한 불균형비(Imbalanced Ratio, IR)를 산출한다(S300). 여기서, 불균형비는 다음의 [수학식 8]과 같이 타겟 사용자의 감정 데이터 중에서 가장 많은 개수(Major Class)와 가장 작은 개수(Minor Class)의 비를 나타낸다.

[0074] [수학식 8]

[0075] $Imbalanced\ Ratio = Major\ Class / Minor\ Class$

[0076] 그 후, 산출된 불균형비가 불균형의 판단 기준이 되는 임계 비율, 예를 들어 2.0보다 작은지 여부를 판단한다(S310). 여기서, 임계 비율인 2.0은 하나의 예시에 불과할 뿐, 감정 인식 방식이나 기타 사용 등의 환경에 따라 다르게 설정될 수도 있다.

[0077] 만약 산출된 불균형비가 임계 비율인 2.0이상인 경우 타겟 사용자의 감정별 데이터가 불균형 상태로 판단되어,

데이터 개수가 작은 감정 데이터의 데이터 개수를 가상 데이터 증강(virtual data augmentation) 방식으로 생성한다(S320). 여기서, 가상 데이터 증강 방식은 minority class의 데이터를 합성하여 증강시키는 오버 샘플링 알고리즘으로 잘 알려져 있는 SMOTE(Synthetic Minority Oversampling Technique)가 사용될 수 있다.

- [0078] 만약 산출된 불균형비가 임계 비율인 2.0보다 작은 경우에는, 타겟 사용자의 감정별 데이터가 균형 상태로 판단되어 가상 데이터 증강 방식의 수행이 중단된다(S330).
- [0079] 이와 같이, 본 발명의 실시예에 따르면, 타겟 사용자의 감정별 음성 데이터가 불균형 상태인 경우 가상 데이터 증강 방식을 통해 데이터 수가 작은 감정의 데이터 수를 증가시킬 수 있다.
- [0080] 이하, 위에서 설명한 3가지 방식을 사용하여 본 발명의 실시예에 따라 개인화된 감정 인식을 위한 강인한 화자 적응 모델링을 수행하는 방법에 대해 설명한다.
- [0081] 도 12는 본 발명의 실시예에 따른 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법의 개략적인 흐름도이다. 마찬가지로, 설명 전에, 본 발명의 실시예에서는 이미 다수의 사용자에게 대한 초기 모델(Initial Model, IM)(11)이 구축되어 IM 데이터베이스(10)에 저장되어 있고, 타겟 사용자에게 대해 초기에 획득된 감정 음성 데이터(21)가 미리 획득되어 타겟 사용자 DB(20)에 저장되어 있는 것으로 가정한다.
- [0082] 도 12를 참조하면, 먼저, 타겟 사용자의 데이터(21)에 대해 감정별 데이터 개수와 임계 개수를 비교한다(S400). 여기서, 임계 개수는 전술한 세 번째 방식에 따른 화자 적응 모델링 방식, 즉 타겟 사용자에게 대해 수집된 감정별 음성 데이터가 불균형 상태인 경우에 수행되는 강인한 화자 적응 모델링 방식을 수행하기에 충분한 개수를 판단하는 데 사용되는 기준값이다. 즉, 감정 데이터가 임계 개수보다 작으면, 세 번째 방식을 수행하기에 충분하지 않은 데이터 개수이므로, 전술한 첫 번째 방식이나 두 번째 방식의 수행에 의해 보다 많은 데이터 개수를 강화해야 함을 나타낸다. 그러나, 감정 데이터가 임계 개수 이상인 경우에는 전술한 첫 번째 방식이나 두 번째 방식의 수행없이 세 번째 방식을 수행하기에 충분한 개수이므로, 바로 세 번째 방식의 수행에 의해 데이터 강화가 수행된다. 일반적으로, 초기 타겟 사용자의 수집 데이터는 매우 적기 때문에 감정 데이터의 개수가 임계 개수보다 작지만, 그렇지 않은 경우도 있을 수 있으므로, 이러한 판단이 수행되어야 한다.
- [0083] 그 후, 감정별 데이터 개수가 임계 개수 이상인지가 판단되고(S410), 만약 감정별 데이터 개수가 임계 개수보다 작은 것으로 판단되면, 해당 감정의 개수가 0으로서 부재 데이터 여부가 판단된다(S420).
- [0084] 만약 부재 데이터가 아니고 소량의 데이터가 존재하는 것으로 판단되면, 도 1 내지 도 6을 참조하여 상기에서 설명한 첫 번째 방식에 따른 화자 적응 모델링 방식, 즉 타겟 사용자에게 대해 수집된 감정 음성 데이터가 소량인 경우에 수행되는 강인한 화자 적응 모델링 방식을 수행하여 해당 감정에 대응하는 데이터를 강화한다(S430).
- [0085] 그러나, 상기 단계(S420)에서 해당 감정이 부재 데이터인 것으로 판단되면, 도 7 내지 도 10을 참조하여 상기에서 설명한 두 번째 방식에 따른 화자 적응 모델링 방식, 즉 타겟 사용자에게 대해 수집된 감정 음성 데이터 중 특정 감정에 대한 데이터가 부재인 경우에 수행되는 강인한 화자 적응 모델링 방식을 수행하여 부재 감정에 대응하는 데이터를 강화한다(S440).
- [0086] 한편, 상기 단계(S410)에서 감정별 데이터 개수가 임계 개수 이상인 것으로 판단되면, 타겟 사용자의 감정 데이터가 불균형 상태인지가 판단된다(S450). 여기서의 불균형 상태의 판단은 전술한 바와 같이 불균형비를 사용하여 판단되는 것을 의미한다.
- [0087] 만약 타겟 사용자의 감정 데이터가 불균형 상태인 경우, 세 번째 방식에 따른 화자 적응 모델링 방식, 즉 타겟 사용자에게 대해 수집된 감정별 음성 데이터가 불균형 상태인 경우에 수행되는 강인한 화자 적응 모델링 방식을 수행하여 불균형한 감정 데이터를 강화한다(S460).
- [0088] 그러나, 타겟 사용자의 감정 데이터가 불균형 상태가 아닌 경우 화자 적응 모델링 방식의 수행이 중단된다(S470).
- [0089] 한편, 상기 단계(S430)와 상기 단계(S440)를 통해 강화된 타겟 사용자의 감정 음성 데이터에 대해 상기 단계(S410 내지 S470)이 반복 수행되어, 최종적으로 타겟 사용자의 감정 음성 데이터가 소량이 아니면서 부재한 감정 데이터도 없으면서 감정별 데이터 상태가 불균형한 상태가 아닌 훈련 데이터셋으로 강화된다.
- [0090] 이와 같이, 본 발명의 실시예에서는 타겟 사용자에게 대해 초기에 수집된 감정 음성 데이터의 수량이나 불균형 상태에 따라 적응적으로 화자 적응 모델링을 수행할 수 있다.
- [0091] 다음, 본 발명의 실시예에 따른 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 장치에 대해 설명한다.

본 장치는 전술한 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 방법을 수행하는 장치에 해당된다.

- [0092] 도 13은 본 발명의 실시예에 따른 개인화된 감정 인식을 위한 강인한 화자 적응 모델링 장치의 개략적인 구성 블록도이다.
- [0093] 도 13에 도시된 바와 같이, 본 발명의 실시예에 따른 화자 적응 모델링 장치(100)는 제1 모델링부(110), 제2 모델링부(120), 제3 모델링부(130) 및 처리 제어부(140)를 포함한다.
- [0094] 제1 모델링부(110)는 전술한 도 1 내지 도 6을 참조하여 설명한 첫 번째 방식에 따른 화자 적응 모델링을 수행한다. 즉, 제1 모델링부(110)는 타겟 사용자의 감정 음성 데이터(21)가 소량의 데이터인 경우, 다수의 사용자 초기 모델(IM)(11)에서 타겟 사용자와 유사한 음성을 선택한 후 레이블을 정제하여 타겟 사용자훈련 데이터셋, 즉 타겟 사용자 데이터를 강화한다.
- [0095] 이를 위해, 제1 모델링부(110)는 레이블링기(111), 특징 벡터값 산출기(112), 평균값 산출기(113), 거리 산출기(114), 임계값 산출기(115), 데이터 선택기(116) 및 감정 확정기(117)를 포함한다.
- [0096] 레이블링기(111)는 IM(11)의 감정 데이터에서 감정을 제거하는 언레이블링, 타겟 사용자의 데이터(21)에 대해 감정 레이블링을 수행한다.
- [0097] 특징 벡터값 산출기(112)는 타겟 사용자의 데이터(21)를 사용하여 감정별 데이터들의 특징 벡터값(TfeatureVector)을 산출한다.
- [0098] 평균값 산출기(113)는 타겟 사용자의 감정별로 특징 벡터값의 평균값을 산출한다. 여기서, 특징 벡터값의 평균값은 전술한 [수학식 1]과 같이 MLE 기법을 사용하여 산출될 수 있다.
- [0099] 거리 산출기(114)는 타겟 사용자의 감정별 특징 벡터값의 평균값들 사이의 거리를 전술한 [수학식 2]와 같이 유클리디언 거리로서 산출한다. 또한, 거리 산출기(114)는 전술한 [수학식 4]를 사용하여 감정별 특징 벡터값의 평균값과 IM(11)의 데이터 사이의 거리를 산출한다.
- [0100] 임계값 산출기(115)는 전술한 [수학식 3]을 사용하여 거리 산출기(114)에 의해 산출된 타겟 사용자의 감정별 특징 벡터값의 평균값들 사이의 거리 중에서 최대값의 절반, 즉 1/2의 거리를 임계값으로 산출한다.
- [0101] 데이터 선택기(116)는 타겟 사용자의 감정별로 특징 벡터값의 평균값을 기준으로 임계값의 범위 내에 있는 IM(11)의 데이터들을 선택한다.
- [0102] 감정 확정기(117)는 데이터 선택기(116)에 의해 선택된 IM(11)의 데이터들에 대해 가장 가까이 있는 타겟 사용자의 감정별 특징 벡터값의 평균값을 찾아서 그에 해당되는 감정을 확정한다.
- [0103] 제2 모델링부(120)는 전술한 도 7 내지 도 10을 참조하여 설명한 두 번째 방식에 따른 화자 적응 모델링을 수행한다. 즉, 제2 모델링부(120)는 타겟 사용자의 감정 중 데이터가 부재인 감정이 있는 경우, IM(11)에 있는 다양한 사용자의 음성 중 타겟 사용자와 유사한 음성 패턴을 가지고 있는 사용자의 감정 음성 데이터 중에서 타겟 사용자의 부재 감정에 대응하는 감정의 데이터로 대체하여 강화한다.
- [0104] 이를 위해 제2 모델링부(120)는 분포 요소 산출기(121), 거리 산출기(122), 유사성 산출기(123) 및 데이터 선택기(124)를 포함한다.
- [0105] 분포 요소 산출기(121)는 전술한 [수학식 5]를 사용하여 타겟 사용자의 감정 음성 데이터 각각에 대해 데이터 분포 요소들인 중앙값, 분산, 왜도, 첨도를 산출한다. 또한, 분포 요소 산출기(121)는 전술한 [수학식 5]를 사용하여 IM(11)의 사용자별 감정 음성 데이터 각각에 대해 중앙값, 분산, 왜도, 첨도를 산출한다.
- [0106] 거리 산출기(122)는 전술한 [수학식 6]을 사용하여 타겟 사용자의 감정별 분포 요소와 IM(11)의 사용자별 감정별 분포 요소 사이의 유클리디언 거리를 산출한다.
- [0107] 유사성 산출기(123)는 전술한 [수학식 7]을 사용하여 타겟 사용자의 데이터(21)와 IM(11)의 다른 사용자의 데이터 사이의 유사성을 산출한다.
- [0108] 데이터 선택기(124)는 산출된 유사성을 통해 타겟 사용자와 가장 유사한 감정 음성을 갖는 사용자의 감정 음성 데이터 중에서 타겟 사용자의 부재 감정에 해당되는 감정의 데이터를 타겟 사용자의 부재 감정의 데이터로서 선택한다.
- [0109] 제3 모델링부(130)는 전술한 도 11을 참조하여 설명한 세 번째 방식에 따른 화자 적응 모델링을 수행한다. 즉,

제3 모델링부(130)는 타겟 사용자에게 대해 수집된 감정별 음성 데이터가 불균형 상태인 경우에 균형 상태를 만들기 위해 수량이 적은 감정의 데이터를 가상 데이터로서 생성하여 데이터 강화를 수행한다.

[0110] 이를 위해, 제3 모델링부(130)는 불균형비 산출기(131) 및 SMOTE 수행기(132)를 포함한다.

[0111] 불균형비 산출기(131)는 전술한 [수학식 8]을 사용하여 타겟 사용자의 불균형비, 즉 타겟 사용자의 감정 데이터 중에서 가장 많은 개수(Major Class)와 가장 작은 개수(Minor Class)의 비를 산출한다.

[0112] SMOTE 수행기(132)는 불균형비 산출기(131)에 의해 산출된 타겟 사용자의 불균형비에 의해 타겟 사용자의 감정 데이터가 불균형 상태인 경우, 예를 들어 불균형비가 2.0 이상인 경우 데이터 개수가 작은 감정의 데이터 개수를 가상 데이터 증강(virtual data augmentation) 방식, 예를 들어 SMOTE 오버샘플링 알고리즘을 사용하여 생성한다.

[0113] 다음, 처리 제어부(140)는 타겟 사용자의 데이터가 임계 개수보다 작은 경우, 타겟 사용자의 감정 음성 데이터가 소량이면 제1 모델링부(110)에 의한 데이터 강화가 수행되도록 하고, 타겟 사용자의 감정 음성 데이터가 부재이면 제2 모델링부(120)에 의한 데이터 강화가 수행되도록 한다.

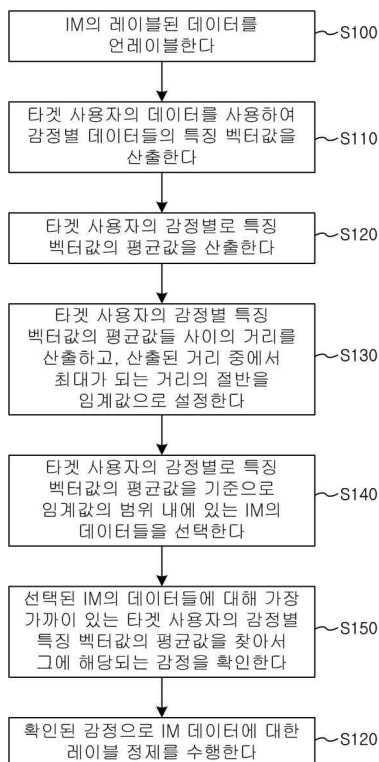
[0114] 또한, 처리 제어부(140)는 타겟 사용자의 데이터가 임계 개수 이상이고 불균형비가 임계 비율 이상이면, 제3 모델링부(130)에 의한 데이터 강화가 수행되도록 한다.

[0115] 이상에서 설명한 본 발명의 실시예는 장치 및 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하는 프로그램 또는 그 프로그램이 기록된 기록 매체를 통해 구현될 수도 있다.

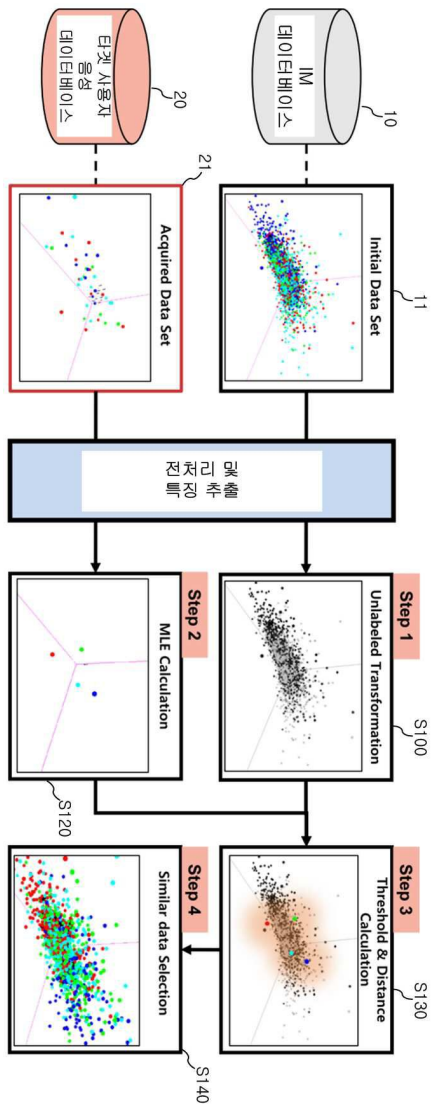
[0116] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

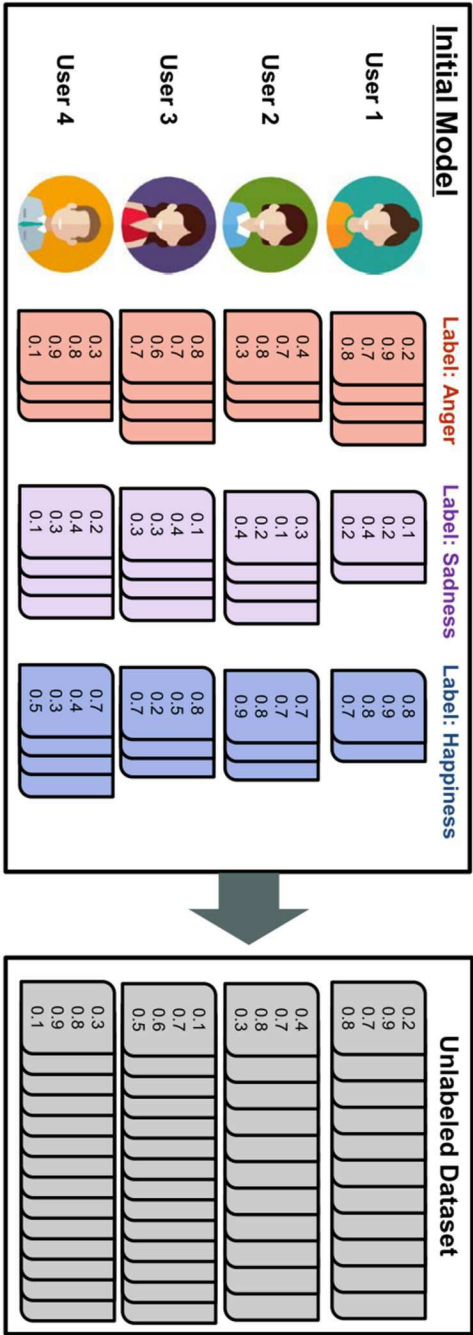
도면1



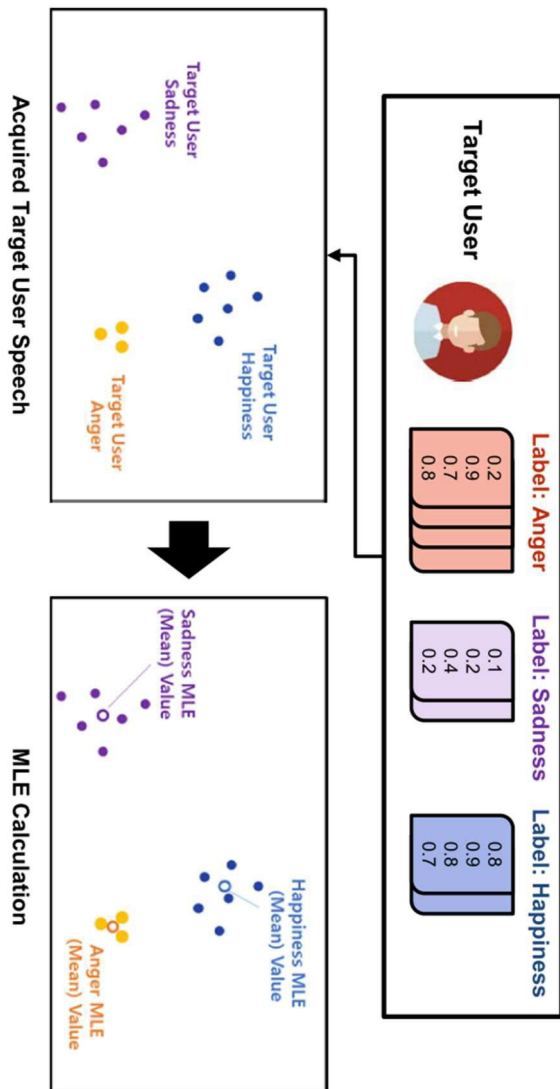
도면2



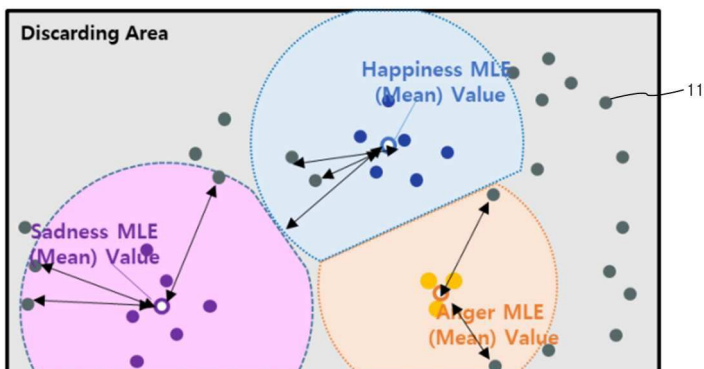
도면3



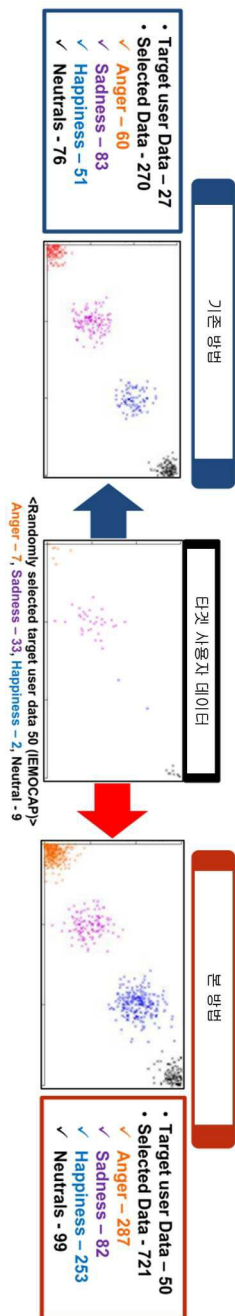
도면4



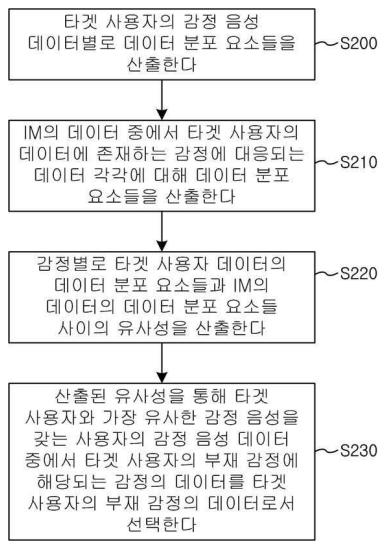
도면5



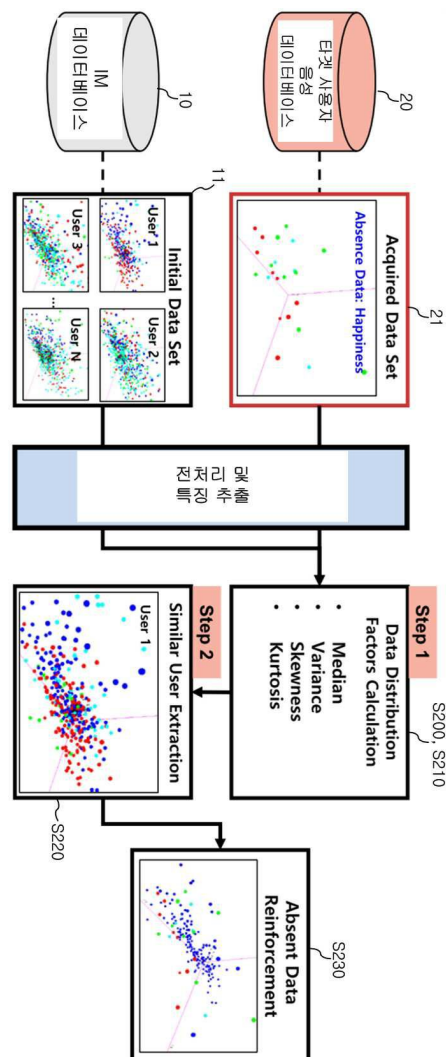
도면6

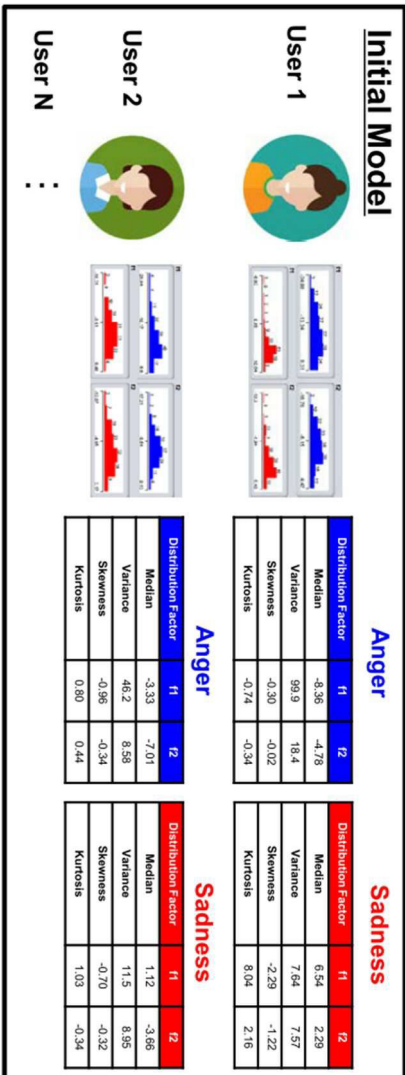
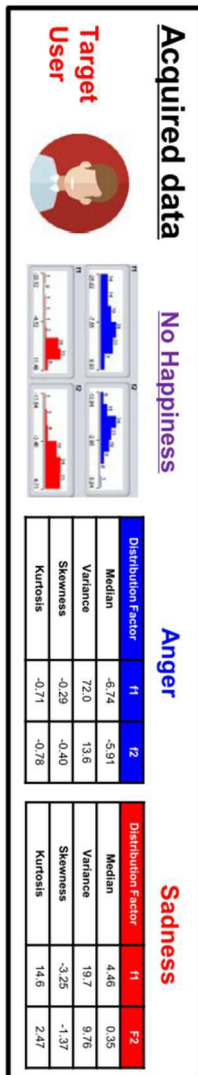


도면7

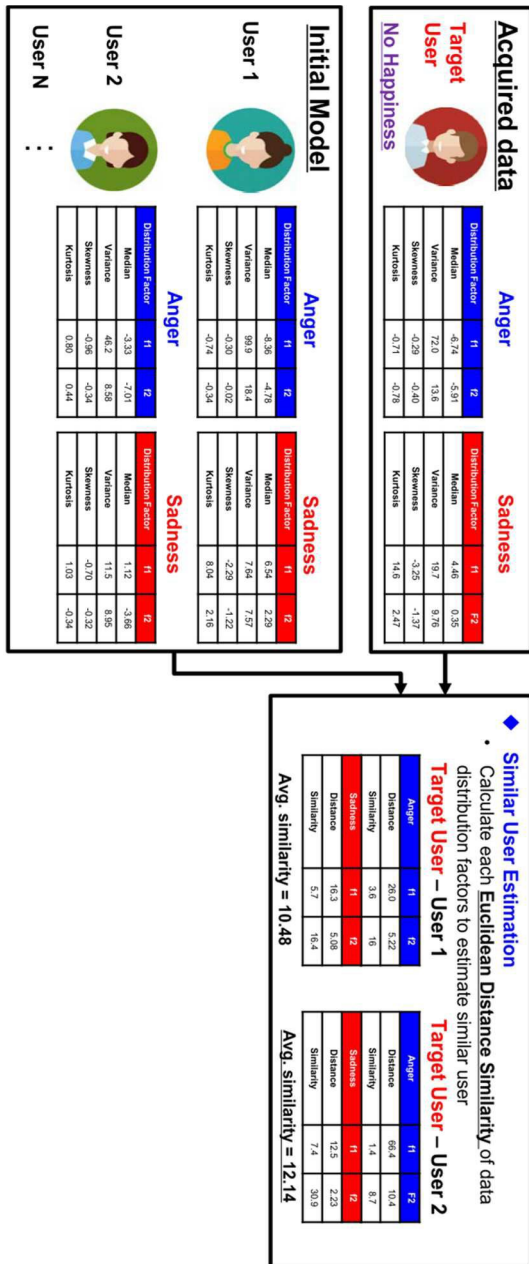


도면8

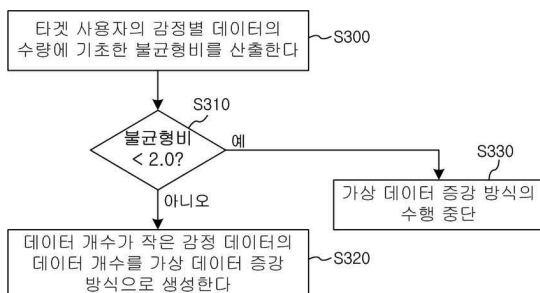




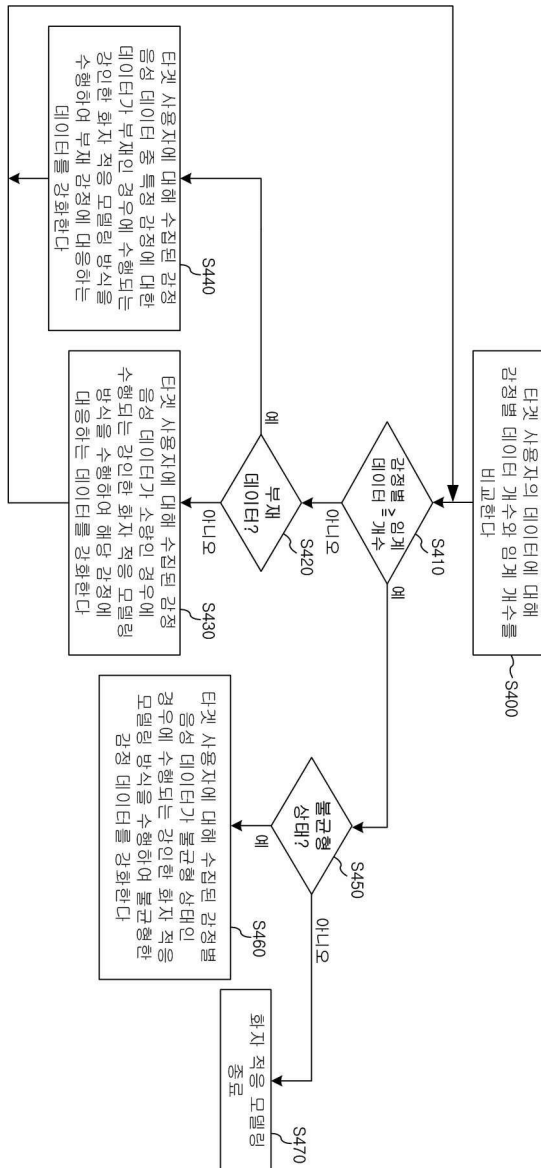
도면10



도면11



도면12



도면13

