(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0140789
(43) 공개일자 2023년10월10일

(51) 국제특허분류(Int. Cl.)
A61B 5/16 (2006.01) *A61B 5/00* (2021.01)
G10L 15/26 (2006.01) *G16H 10/20* (2018.01)
G16H 50/20 (2018.01)
(52) CPC특허분류
A61B 5/167 (2013.01)
A61B 5/4803 (2013.01)
(21) 출원번호 10-2022-0039465
(22) 출원일자 2022년03월30일
심사청구일자 2022년03월30일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
최기홍
서울특별시 중구 다산로29길 46-11
최준연
서울특별시 강남구 영동대로 22, 804동 3303호
윤서원
서울특별시 강남구 개포로 409, 5동 403호
(74) 대리인
김등용

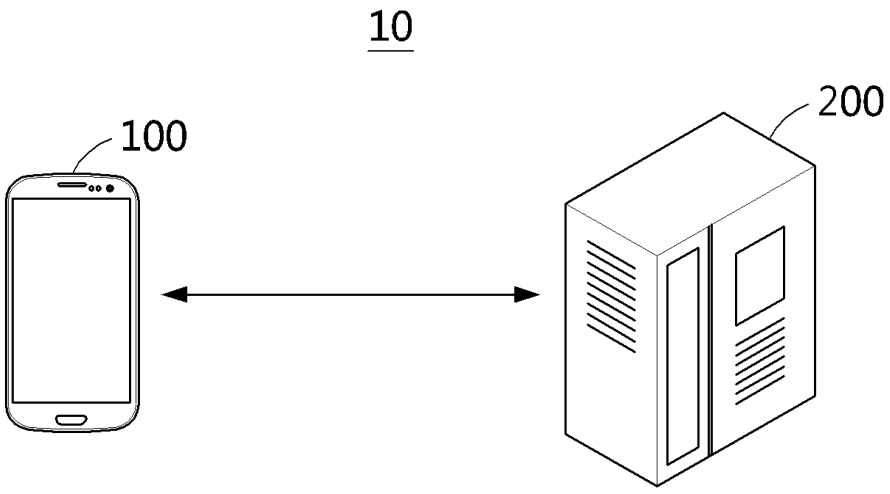
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **자연어 처리 기반 성격평가 장치 및 방법**

(57) 요 약

자연어 처리 기반의 성격평가 장치가 개시된다. 상기 성격평가 장치는 단말로 복수의 인터뷰 문항들과 적어도 하나의 지시문을 포함하는 메시지를 송신하고, 상기 단말로부터 상기 복수의 인터뷰 문항들 각각에 대한 제1 응답과 상기 적어도 하나의 지시문에 대한 제2 응답을 수신하는 설문 제공부, 상기 제1 응답과 상기 제2 응답 중 적어도 하나를 전처리하는 전처리부, 및 상기 전처리부에 의한 전처리 결과를 이용하여 상기 단말의 사용자의 성격을 예측하는 성격 예측부를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
A61B 5/7264 (2023.08)
G10L 15/26 (2013.01)
G16H 10/20 (2021.08)
G16H 50/20 (2018.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711120643
과제번호 2020R1A2C2099665
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 머신러닝 기반 '성격 평가' 알고리즘 개발
기 여 율 1/2
과제수행기관명 고려대학교
연구기간 2020.09.01 ~ 2021.08.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711139704
과제번호 2020R1A2C2099665
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 머신러닝 기반 '성격 평가' 알고리즘 개발
기 여 율 1/2
과제수행기관명 고려대학교
연구기간 2021.09.01 ~ 2022.02.28

명 세 서

청구범위

청구항 1

단말로 복수의 인터뷰 문항들과 적어도 하나의 지시문을 포함하는 메시지를 송신하고, 상기 단말로부터 상기 복수의 인터뷰 문항들 각각에 대한 제1 응답과 상기 적어도 하나의 지시문에 대한 제2 응답을 수신하는 설문 제공부;

상기 제1 응답과 상기 제2 응답 중 적어도 하나를 전처리하는 전처리부; 및

상기 전처리부에 의한 전처리 결과를 이용하여 상기 단말의 사용자의 성격을 예측하는 성격 예측부를 포함하는 성격평가 장치.

청구항 2

제1항에 있어서,

상기 복수의 인터뷰 문항들 각각은, 주관식 응답을 요하는, 상기 사용자의 성격을 예측하기 위한 문항이고,

상기 적어도 하나의 지시문은 상기 사용자를 잘 드러낸다고 생각하는 복수의 형용사 단어들에 대한 응답을 요청하는 지시문과 상기 사용자의 성격에 대한 장점과 단점에 대한 응답을 요청하는 지시문을 포함하는,

성격평가 장치.

청구항 3

제1항에 있어서,

음성 데이터 형태의 상기 제1 응답과 음성 데이터 형태의 상기 제2 응답 중 적어도 하나를 텍스트 데이터 형태로 변환하는 텍스트 변환부를 더 포함하는,

성격평가 장치.

청구항 4

제1항에 있어서,

상기 전처리부는, 상기 제1 응답과 상기 제2 응답 중 적어도 하나에 대한 분절화 동작을 수행하고, 상기 제1 응답과 상기 제2 응답 중 적어도 하나에 포함된 단어들 중 적어도 형용사, 동사, 및 명사에 해당하는 단어를 추출하고, 성격과의 연관성에 기초하여 추출된 단어들 중 일부만을 선별하는,

성격평가 장치.

청구항 5

제4항에 있어서,

상기 전처리부는, 추출된 단어들을 대상으로 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법을 적용하여 복수의 토픽들을 추출하고, 복수의 토픽들 각각과 성격의 연관성에 기초하여 추출된 단어들 중 일부만을 선별하는,

성격평가 장치.

청구항 6

제5항에 있어서,

상기 전처리부는, 각각이 가중치가 부여된 복수의 단어들을 포함하는 단어 리스트를 이용하여 상기 복수의 토픽들 각각과 성격의 연관성을 산출하고, 상기 복수의 토픽들 중에서 미리 정해진 임계치 이하의 연관성을 갖는 토

픽에만 포함된 단어를 제거하는,

성격평가 장치.

청구항 7

제5항에 있어서,

상기 전처리부는, 상기 복수의 토픽들 각각에 대한 선택 정보를 수신하고, 상기 복수의 토픽들 중에서 선택되지 않은 토픽에만 포함된 단어를 제거하는,

성격평가 장치.

청구항 8

제5항에 있어서,

상기 전처리부는, 토픽과의 연관성을 출력하는 연관성 판단 모델을 이용하여 상기 복수의 토픽들 각각과 성격의 연관성을 산출하고, 상기 복수의 토픽들 중에서 미리 정해진 임계치 이하의 연관성을 갖는 토픽에만 포함된 단어를 제거하는,

성격평가 장치.

청구항 9

제1항에 있어서,

상기 성격 예측부는 전처리 결과를 학습된 성격 예측 모델에 입력하여 상기 사용자의 성격을 예측하는,

성격평가 장치.

청구항 10

제9항에 있어서,

상기 성격 예측 모델은 트랜스포머 기반의 사전학습 언어 모델이고,

상기 성격 예측 모델은 상기 사용자의 적응적 상위요인 5가지와 부적응적 상위요인 5가지에 대한 예측 점수를 출력하는,

성격평가 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 성격평가 방법에 관한 것으로, 특히 자연어 처리에 기초한 자동화된 성격평가 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 개인의 성격 및/또는 정신건강 수준 등을 평가하는 심리평가 분야에서, 자기보고식(self-report) 검사가 널리 이용되었다. 자기보고식 검사는 몇몇 장점을 보이는 반면에, 태생적인 문제점들을 가지고 있다. 즉, 자기보고식 평가는 간편하고 사용성이 뛰어나지만, 안면타당도가 지나치게 높아 수검자가 반응을 왜곡하거나 편향하기 쉽다. 반응 왜곡 및 편향은 직무 면접이나 범죄 장면 등에서는 더욱 해결하기 어려운 문제이다.

[0003] 이에, 본 발명에서는 자기보고식 검사의 한계인 스스로 보고하는 성격, 왜곡, 및 편향이 가능한 응답이라는 한계를 극복할 수 있는 자연어 처리(Natural Language Processing, NLP) 기반의 성격검사 장치 및 방법을 제안하고자 한다. 구체적으로, 본 발명에서는 성격을 잘 드러낼 수 있는 질문에 대한 주관식 응답을 분석하는 모델(또는 알고리즘)을 개발하고, 모든 사람에게 통일된 질문을 하여 성격을 예측하고자 한다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 대한민국 공개특허 제2021-0040618호 (2021.04.14. 공개)

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적인 과제는 자연어 처리에 기반하는 성격검사 장치 및 방법을 제공하는 것이다.

과제의 해결 수단

[0006] 본 발명의 일 실시예에 따른, 자연어 처리 기반의 성격평가 장치는 단말로 복수의 인터뷰 문항들과 적어도 하나의 지시문을 포함하는 메시지를 송신하고, 상기 단말로부터 상기 복수의 인터뷰 문항들 각각에 대한 제1 응답과 상기 적어도 하나의 지시문에 대한 제2 응답을 수신하는 설문 제공부, 상기 제1 응답과 상기 제2 응답 중 적어도 하나를 전처리하는 전처리부, 및 상기 전처리부에 의한 전처리 결과를 이용하여 상기 단말의 사용자의 성격을 예측하는 성격 예측부를 포함한다.

발명의 효과

[0007] 본 발명의 실시예에 따른 성격검사 장치 및 방법에 의할 경우, 자연어 처리 기반 알고리즘을 활용하여, 종래의 자기보고식 성격평가의 한계점을 극복할 수 있는 효과가 있다. 본 발명을 통해 수검자는 자유로운 발화 및 제한되지 않은 다양한 반응을 보일 수 있으며, 보다 실생활에 가까운 성격이나 정신건강 상태를 드러낼 수 있고, 수검자의 반응의 질이나 극단 반응 성향 또한 분석에 포함될 수 있다.

[0008] 또한, 빅데이터에만 기반한 상향식(bottom-up) 방식의 기존 언어기반 성격모델의 한계를 극복하기 위해, 과학적으로 검증된 심리학 성격 이론을 기반으로 성격을 잘 드러낼 수 있는 문항들을 개발하였고, 이를 모든 수검자에게 동일하게 제공하여 small-sample일 때에도 성격 요인을 드러낼 수 있도록 입력(input)의 노이즈를 최소화하였다는 장점이 있으며, SNS(Social Network Service) 데이터 등 간편한 데이터가 아닌 심리학 분야 연구원들이 직접 인터뷰를 실시하여 수집한 양질의 데이터를 바탕으로 알고리즘을 개발했다는 장점이 있다.

[0009] 본 발명은 다양한 분야에서 활용될 수 있을 것으로 기대된다. 가장 우선적으로는 각 기업의 면접 등 인사장면이나 의료기관과 상담센터에서 활용할 수 있을 것이다. 또한, 범죄자 평가, 학교 등의 분야에서도 활용이 가능할 것으로 기대된다. 특히 인사평가와 범죄자 관련 성격평가야말로 기존의 자기보고식 검사의 한계로 인해 반응을 왜곡하고 숨기는 것이 가능했으므로, 본 평가기술을 활용할 여지가 매우 많다고 생각된다.

도면의 간단한 설명

[0010] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 본 발명의 일 실시예에 따른 성격 평가 시스템을 도시한다.

도 2는 도 1에 도시된 서버의 기능 블럭도이다.

발명을 실시하기 위한 구체적인 내용

[0011] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0012] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0013] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로

만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0014] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0015] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0016] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0017] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0018] 도 1은 본 발명의 일 실시예에 따른 성격 평가 시스템을 도시한다.

[0019] 도 1을 참조하면, 성격 측정 시스템 등으로 명명될 수도 있는 성격 평가 시스템(10)은 단말(100)과 서버(200)를 포함한다. 단말(100)과 서버(200)는 소정의 유무선 통신망을 통하여 서로 데이터를 주고 받을 수 있다. 또한, 단말(100) 및/또는 서버(200)는 적어도 프로세서(processor) 및/또는 메모리(memory)를 포함하는 컴퓨팅 장치로 구현될 수 있다. 컴퓨팅 장치는 모바일 폰, 스마트 폰, 태블릿 PC, PC(Personal Computer), HMD(Head Mounted Device), 서버 등을 포함할 수 있다.

[0020] 단말(100)은 서버(200)와의 통신을 통해 서버(200)로부터 메시지를 수신하고 이에 대한 응답을 서버(200)로 송신함으로써, 성격 평가 내지 성격 측정의 기초 자료가 되는 데이터를 서버(200)로 송신할 수 있다.

[0021] 구체적으로, 단말(100)은 서버(200)로부터 복수의 인터뷰 문항들을 수신하고, 이에 대한 (주관식) 응답을 서버(200)로 송신할 수 있다. 복수의 인터뷰 문항들은 사용자 또는 피검자의 성격을 평가하기 위한 문항들로써 하나의 인터뷰 문항씩 순차적으로 수신되거나 복수의 인터뷰 문항들이 함께 수신될 수도 있다.

[0022] 복수의 문항들 각각은 텍스트 데이터이거나 음성 데이터일 수 있다. 즉, 복수의 문항들 각각은 텍스트 데이터로 이루어져 단말(100)에 구비된 디스플레이 장치에 디스플레이될 수 있다. 복수의 문항들 각각은 음성 데이터일 수도 있는데, 이 경우 복수의 문항들 각각은 단말(100)에 구비된 음성 출력 장치를 통해 음성 신호로 출력될 수 있다. 실시예에 따라, 복수의 문항들 각각은 텍스트 데이터와 음성 데이터일 수 있다. 이 경우, 복수의 문항들 각각은 디스플레이 장치에 디스플레이되는 동시에 음성 출력 장치를 통해 음성 신호로 출력될 수도 있다.

[0023] 단말(100)은 소정의 입력 인터페이스를 통해 사용자 또는 피검자로부터 복수의 문항들 각각에 대한 응답을 수신하고, 수신된 응답을 서버(200)로 송신할 수 있다. 이때, 응답은 텍스트 데이터이거나 음성 데이터일 수 있다. 다만, 복수의 문항들 각각의 데이터 형태와 응답의 데이터 형태가 일치하는 것은 아니다. 즉, 복수의 문항들 각각이 텍스트 데이터임에도 불구하고 응답은 음성 데이터일 수 있고, 복수의 문항들 각각이 음성 데이터임에도 불구하고 응답은 텍스트 데이터일 수 있다.

[0024] 또한, 단말(100)은 서버(200)로부터 적어도 하나의 지시문을 수신할 수도 있다. 적어도 하나의 지시문은 복수의 문항들과 함께 수신되거나 복수의 문항들과는 별도로 수신될 수 있다. 지시문 역시 텍스트 데이터이거나 음성 데이터일 수 있다. 즉, 적어도 하나의 지시문은 텍스트 데이터로 이루어져 단말(100)에 구비된 디스플레이 장치

에 디스플레이될 수 있다. 적어도 하나의 지시문은 음성 데이터일 수도 있는데, 이 경우 적어도 하나의 지시문은 단말(100)에 구비된 음성 출력 장치를 통해 음성 신호로 출력될 수 있다. 실시예에 따라, 적어도 하나의 지시문은 텍스트 데이터와 음성 데이터일 수 있다. 이 경우, 적어도 하나의 지시문은 디스플레이 장치에 디스플레이되는 동시에 음성 출력 장치를 통해 음성 신호로 출력될 수도 있다.

[0025] 단말(100)은 소정의 입력 인터페이스를 통해 사용자 또는 피검자로부터 지시문에 대응 응답을 수신하고, 수신된 응답을 서버(200)로 송신할 수 있다. 이때, 응답은 텍스트 데이터이거나 음성 데이터일 수 있다. 다만, 적어도 하나의 지시문의 데이터 형태와 응답의 데이터 형태가 일치하는 것은 아니다. 즉, 적어도 하나의 지시문이 텍스트 데이터임에도 불구하고 응답은 음성 데이터일 수 있고, 적어도 하나의 지시문이 음성 데이터임에도 불구하고 응답은 텍스트 데이터일 수 있다. 적어도 하나의 지시문에 대한 응답은 복수의 문항들 각각에 대한 응답들과 동시에 송신되거나 복수의 문항들 각각에 대한 응답들과는 별도로 송신될 수 있다.

[0026] 또한, 도 1에는 하나의 단말만이 도시되어 있으나, 본 발명이 권리범위가 이에 제한되는 것은 아니다. 즉, 실시예에 따라, 성격 평가 시스템(10)에는 복수의 단말들이 포함될 수도 있다.

[0027] 서버(200)는 단말(100)로부터 수신된 사용자 또는 피검자의 응답을 기초로 사용자 또는 피검자의 성격을 예측(또는 판단)하고 예측된(또는 판단된) 결과를 출력할 수 있다. 서버(200)이 보다 구체적인 동작은 도 2를 참조하여 상세히 설명하도록 한다.

[0028] 도 2는 도 1에 도시된 서버의 기능 블럭도이다.

[0029] 도 2를 참조하면, 성격 평가 장치, 성격 측정 장치, 성격 평가 서버, 성격 측정 서버 등으로 명명될 수도 있는 서버(200)는 설문 제공부(210), 텍스트 변환부(220), 전처리부(230), 및 성격 예측부(240) 중 적어도 하나 이상을 포함할 수 있다. 실시예에 따라, 서버(200)는 저장부(250)를 더 포함할 수도 있다.

[0030] 설문 제공부(210)는 복수의 인터뷰 문항들 및/또는 적어도 하나의 지시문을 포함하는 메시지를 단말(100)로 송신할 수 있다. 복수의 인터뷰 문항들 각각은 사용자 또는 피검자의 성격을 예측하기 위한 문항들로써 n(n은 2 이상의 자연수로써, 예시적인 값은 18임) 개의 인터뷰 문항들로 구성될 수 있다. 인터뷰 문항의 예시는 다음과 같다.

[0031] - 일상에서 여유가 있을 때, 주로 어떻게 시간을 보내길 원하세요?

[0032] - 스스로 세웠던 기준이나 목표를 어느 정도 달성하는 편인가요?

[0033] - 주변 사람과의 관계에서 어느 정도로 가깝게 지내기를 원하시나요?

[0034] - 정해진 기한 안에 마쳐야 할 일이 있을 때 평소에 이를 어떻게 처리하시나요?

[0035] - 가지고 있는 부정적인 감정이나 생각에 대해 이야기해 주세요.

[0036] 복수의 인터뷰 문항들 각각은 주관식 응답을 요하는 문항들이다. 설문 제공부(210)에 의해 제공되는 복수의 문항들은 저장부(250)에 저장되어 있을 수 있고, 설문 제공부(210)에 의해 수신되는 복수의 문항들 각각에 대한 (주관식) 응답은 저장부(250)에 저장될 수 있다.

[0037] 설문 제공부(210)에 의해 송신되는 적어도 하나의 지시문의 예시는 다음과 같다.

[0038] - 스스로를 잘 드러낸다고 생각하는 형용사 단어 k(k는 2 이상의 자연수로써, 예시적인 값은 6임)개를 기재해 주세요.

[0039] - 자신의 성격에 대해 장점과 단점을 포함하여 자유롭게 설명해 주세요(300자 내외).

[0040] 상술한 바와 같이, 적어도 하나의 지시문은 스스로를 잘 드로낸다고 생각하는 형용사 단어 k(예컨대, 6)개에 대한 기재 요청 및/또는 성격에 대한 자유로운 기재 요청을 포함할 수 있다. 설문 제공부(210)에 의해 제공되는 적어도 하나의 지시문에 대하여, 사용자 또는 피검자는 단말(100)을 통하여 자신을 잘 드러낸다고 생각하는 형용사 단어 k(예컨대, 6)개를 텍스트 데이터 또는 음성 데이터의 형태로 송신할 수 있다. 설문 제공부(210)에 의해 제공되는 적어도 하나의 지시문은 저장부(250)에 저장되어 있을 수 있고, 설문 제공부(210)에 의해 수신되는 적어도 하나의 지시문에 대한 응답은 저장부(250)에 저장될 수 있다.

[0041] 텍스트 변환부(220)는 단말(100)로부터 수신된 응답을 텍스트 데이터로 변환할 수 있다. 즉, 단말(100)로부터 수신되는 복수의 문항들 각각에 대한 주관식 응답 및/또는 적어도 하나의 지시문에 대한 응답에 음성 데이터 형태의 응답이 있는 경우, 텍스트 변환부(220)는 음성 데이터 형태의 응답을 텍스트 데이터 형태의 응답으로 변환

할 수 있다. 실시예에 따라, 사용자 또는 피검자에 의한 응답은 텍스트 데이터 형태일 수 있고, 이 경우 텍스트 변환부(220)는 서버(200) 내에서 생략되는 구성으로 이해될 수도 있다.

[0042] 전처리부(230)는 설문 제공부(210)에 의해 수신된 응답, 설문 제공부(210)에 의해 수신되어 저장부(250)에 저장되어 있는 응답, 또는 텍스트 변환부(220)에 의해 텍스트 데이터의 형태로 변환된 응답에 대하여 전처리 동작을 수행할 수 있다. 전처리 동작은 분절화(토큰화) 동작, 단어 추출 동작, 단어 선별 동작 중 적어도 하나를 포함할 수 있다. 이를 위해, 전처리부(230)는 분절화 동작을 수행하는 분절화부, 단어 추출 동작을 수행하는 단어 추출부, 및 단어 선별 동작을 수행하는 단어 선별부 중 적어도 하나를 포함할 수 있다.

[0043] 전처리부(230)의 분절화부는 응답(복수의 인터뷰 문항들 각각에 대한 주관식 응답 및 적어도 하나의 지시문에 대한 응답 중 적어도 일부를 포함함)에 대하여 분절화 동작을 수행할 수 있다. 분절화 동작을 수행하기 위해서, 기존에 널리 알려진 기법들 중 적어도 하나의 기법이 이용될 수 있다.

[0044] 전처리부(230)의 단어 추출부는 응답(복수의 인터뷰 문항들 각각에 대한 주관식 응답 및 적어도 하나의 지시문에 대한 응답 중 적어도 일부를 포함함)에 포함된 단어들 중 적어도 일부의 단어만을 추출할 수 있다. 추출되는 단어는 적어도 형용사, 동사, 및 명사를 포함한다. 이를 위해, 단어 추출부는 분절화된 단어들 각각에 대한 품사(Part Of Speech, POS)를 결정하고 단어들 중에서 적어도 형용사, 동사, 및 명사에 해당하는 단어를 추출하고 나머지 단어들은 제거할 수 있다. 단어 추출부는 단어를 추출하기 위해 품사 태깅(Part Of Speech tagging, POS 태깅)을 수행할 수 있다. POS 태깅은 널리 알려진 기법 중 적어도 하나를 이용하여 수행될 수 있다. 예시적인 POS 태깅 기법에는, SVM(Support Vector Machine) 등과 같은 분류기(Classifier)를 이용하여 각 단어의 품사를 개별적으로 예측하는 개별 예측(Pointwise prediction) 기법, HMM이나 CRF(Conditional random field) 등과 같은 확률 모델(Probability model)을 이용하는 기법, 및 RNN, LSTM, BERT 등과 같은 신경망 기반 모델(Neural network-based model)을 이용하는 기법 등이 있다. 실시예에 따라, POS 태깅은 분절화 동작과 동시에 수행될 수도 있다.

[0045] 전처리부(230)의 단어 선별부는 잠재 디리클레 할당(LDA) 기법을 이용하여, 단어 추출부에 의해 추출된 단어들 중 적어도 일부의 단어만을 선별하고 선별되지 않은 단어는 제거할 수 있다. 이를 위해, 단어 선별부는 선별된 단어들을 대상으로 LDA 기법을 적용하여 복수의 토픽들을 추출할 수 있다. 추출된 복수의 토픽들 각각과 성격은 서로 다른 연관성을 가질 것이다. 따라서, 성격과의 연관성이 낮은 토픽에(또는 성격과의 연관성이 낮은 토픽에만) 포함된 단어들은 제거될 수 있다.

[0046] 토픽과 성격의 연관성은 다양한 방식으로 판단될 수 있다. 우선, 미리 정해진 단어 리스트와 각 토픽에 포함된 키워드들을 비교함으로써 연관성(또는 연관도)이 측정될 수 있다. 단어 리스트에는 성격과의 연관성이 높은(또는 높다고 판단되는) 단어들이 포함되어 있다. 또한, 단어 리스트에 포함된 단어들 각각은 연관도에 따른 가중치가 산정되어 있을 수도 있다. 구체적인 예로, 토픽에 포함된 키워드들 각각에 대한 가중치의 합으로 연관도가 산출될 수 있다. 미리 정해진 임계치 이하(또는 미만) 또는 0의 연관도를 갖는 토픽은 성격과 무관한 토픽(또는 성격과 관련성이 떨어지는 토픽)으로 판단되고, 성격과 무관한 토픽(또는 성격과 관련성이 떨어지는 토픽)으로 판단된 토픽에 포함된 단어들(또는 토픽에만 포함된 단어들)은 제거될 수 있다. 여기서, 연관도가 0인 경우는 토픽에 포함되는 키워드들 중 어느 키워드도 단어 리스트에 포함되지 않는 경우이다.

[0047] 다음으로, 추출된 토픽들에 대한 관리자(또는 전문가)의 선택 정보를 수신하고, 수신된 선택 정보에 대응하지 않는 토픽이 성격과 무관한 토픽으로 판단될 수 있다. 이를 위해, 전처리부(230)의 단어 선별부는 추출된 토픽들을 소정의 출력 장치(예컨대, 디스플레이 장치)를 통해 출력하고, 성격과 연관된 토픽인지 여부에 대한 정보를 관리자(또는 전문가)로부터 수신할 수 있다. 결국, 성격과 무관한 토픽으로 판단된 토픽에 포함된 단어들(또는 토픽에만 포함된 단어들)은 제거될 수 있다.

[0048] 또다른 예로, 추출된 토픽들 각각에 대하여 성격과의 연관도를 출력하는 모델을 이용할 수도 있다. 이를 위해서는, 토픽에 포함된 키워드들을 수신하고 성격과의 연관성(또는 연관도)를 출력하는 연관성 판단 모델을 이용할 수 있다. 연관성 판단 모델은 토픽을 이루는 키워드들과 이에 대한 연관성 정보(예컨대, 연관도)를 포함하는 학습 데이터를 이용하여 소정의 신경망 모델을 학습시킴으로써 생성할 수 있다. 연관성 판단 모델의 출력이 미리 정해진 임계치 이하(또는 미만) 또는 0인 토픽에 포함된 단어들(또는 토픽에만 포함된 단어들)은 제거될 수 있다. 실시예에 따라, 연관성 판단 모델은 성격과의 연관성에 대하여 여부만을 출력할 수도 있다. 이 경우, 연관성이 없다고 판단되는 토픽에 포함된 단어들(또는 토픽에만 포함된 단어들)은 제거될 수 있다.

[0049] 성격 예측부(240)는 전처리의 결과를 기초로 사용자(또는 피검자)의 성격을 판단할 수 있다. 전처리 결과는 최

종적으로 선별된 복수의 단어들을 의미할 수 있다. 성격 예측부(240)은 복수의 단어들을 임베딩한 후 성격 예측 모델(성격 판단 모델)에 입력함으로써 사용자의 성격을 예측(또는 판단) 수 있다.

[0050] 전처리를 거친 이후, 단어의 맥락만 파악하는 것이 아닌 문장 내의 뜻을 파악하기 위해서 알고리즘 내에서 트랜스포머 기반 사전학습(pretrained model) 언어모델(KoBERT 등) 등 기존의 한국어 사전학습 모델 중 가장 우수한 성능을 보이는 임베딩 기법을 적용할 수 있다. 이를 통해 사용자로부터 받은 텍스트 처리의 데이터를 숫자로 변환한다. 이후 성격 예측 모델을 위해 KoBERT 등을 기반으로 미세 조정(Fine-tuning)된 모델을 적용하여 산출된 파라미터를 적용하여 성격 예측 점수가 산출된다. 파라미터 추출을 위해서는 딥러닝 기법과 함께 전통적인 기계학습 테크닉(예컨대, Boruta, Recursive Feature Elimination with Cross Validation(RFECV), Logistic Regression using Statsmodels, Random Forest, H2O, and L1 Regularization 등) 중 우수한 성능을 보이는 테크닉을 사용하여 학습된 알고리즘을 적용하여 산출된 파라미터를 적용하게 된다.

[0051] 이를 통해, 각 사용자의 성격 점수로 환산이 이루어진다. 알고리즘의 경우, 예를 들어 사용자가 '친구'라는 단어와 '친구들'이라는 단어를 사용할 경우(혹은, 친구라는 단어와 유사한 단어군을 사용할 경우) 각각 외향성에 가중치(weight)값을 가지고 점수가 측정되게 되며, '친구들'이라는 단어는 '친구'라는 단어보다 더 높은 가중치 값을 지녀 더 높은 외향성을 예측하는 방식으로 작동한다. 반면 '혼자'라는 단어는 외향성에 음수의 가중치값을 지녀 외향성 점수를 낮추는 방식으로 작동하며, '우울' 혹은 '슬픔'과 같은 단어는 부정정서성 점수를 증가시키고 정서안정성 점수를 낮추는 방식으로 작동한다. 각 사용자의 언어적 응답은 각각 적응적 상위요인 5가지(외향성(또는 외향-내향성), 수용성, 성실성, 개방성, 정서안정성), 부적응적 상위요인 5가지(대인회피성, 자기중심성, 조절취약성(또는 주의곤란성), 사고지각특이성(또는 정신증성), 부정정서성)에 대한 점수와, 각각의 상위요인에 대한 3~4가지의 하위요인 점수도 함께 제공되어 성격 프로파일을 산출할 수 있다.

[0052] 다시 말해, 성격 예측 모델은 소정의 학습 데이터를 이용하여 신경망 모델 또는 언어 모델(예컨대, 사전 학습된 언어 모델로써 BERT 또는 KoBERT)을 학습시킴으로써 생성될 수 있다. 학습 데이터는 텍스트 데이터 형식을 갖는 사용자의 응답 데이터와 이에 대응하는 성격 점수를 의미할 수 있다. 성격 점수는 적응적 상위요인 5가지 각각에 대한 점수(하위요인 각각에 대한 점수를 포함할 수 있음) 및/또는 부적응적 상위요인 5가지 각각에 대한 점수(하위요인 각각에 대한 점수를 포함할 수 있음)를 포함할 수 있다.

[0053] 적응적 상위요인 중 외향성은 사람들과의 관계에서 활력을 얻으며 대인관계를 확장하고 주변에 영향력을 행사하고자 하는 성향을, 수용성은 원만한 관계를 중요시하며 타인에게 적극적으로 공감하고, 너그럽고 신뢰롭게 행동하는 성향을, 성실성은 질서와 체계를 중시하며 사회 관습과 기준을 달성하기 위해 노력하는 성향을, 개방성은 자기 내면과 세상에 대한 호기심이 풍부하며 새롭고 다양한 경험을 추구하는 성향을, 정서안정성은 감정을 억압하거나 두려워하지 않고 건강하고 성숙한 방식으로 수용하고 표현하는 성향을 의미한다.

[0054] 부적응적 상위요인 중 대인회피성은 긍정 정서를 경험하기 어렵고 대인관계에 대한 욕구가 낮아 사람들과의 정서적 교류를 피하는 성향을, 자기중심성은 자신의 이익과 만족을 중요시하여 타인의 욕구나 감정에 둔감한 성향을 주의곤란성은 맥락과 상황에 맞게 충동을 조절하기 어려워하고 중요한 자극에 주의를 집중하기 힘들어하는 성향을, 정신증성은 사고의 유연성이 부족하고 자기 세계에 몰입하여 현실 세계를 객관적으로 인식하기 어려워하는 성향을, 부정정서성은 평소 부정적 정서를 자주 경험하여 사소한 자극에도 쉽게 각성되고 압도되는 성향을 의미한다.

[0055] 외향성의 하위 요인 중 활력은 에너지가 많으며 활동적이고 열정적으로 생활하기를 즐기는 성향을, 사교는 새로운 만남을 좋아하고 사람들에게 쉽게 다가가 친밀한 관계를 맺는 성향을, 주도는 자기 주장이 강하여 앞장 서서 다른 사람을 이끌고자 하는 성향을, 내향은 사람들과 어울려서 지내기보다 독립적으로 생활하기를 선호하는 성향을 의미한다.

[0056] 수용성의 하위 요인 중 신뢰는 대인관계에서 믿은과 신의를 중시하여 투명하고 솔직하게 행동하는 성향을, 관대는 대인관계에서 너그럽고 무던하게 행동하여 원만한 관계를 유지하는 성향을, 이타는 타인의 생각과 감정에 공감하며 타인을 적극적으로 배려하고 돕는 성향을 의미한다.

[0057] 성실성의 하위 요인 중 체계는 규칙과 체계를 좋아하며 질서정연하게 사고하고 행동하는 성향을, 끈기는 목표를 완수하기 위해 근면하고 꾸준하게 노력하는 성향을, 성취는 높은 수준의 목표를 지향하고 이를 달성하려는 성향을 의미한다.

[0058] 개방성의 하위 요인 중 체험은 새롭고 다양한 활동을 직접 체험하기를 좋아하는 탐험적인 성향을, 탐구는 지적 호기심이 강하며 철학적이고 추상적인 사고와 추론을 즐기는 성향을, 심미는 예술과 아름다움에 대한 감수성이

풍부하며 미적 활동을 즐기는 성향을 의미한다.

[0059] 정서안정성의 하위 요인 중 정서인식은 자신이 경험하는 정서에 관심을 기울이고 이를 정확하고 구체적으로 이해하는 성향을, 정서수용은 자신이 경험하는 감정을 왜곡하거나 억누르지 않고 자연스럽게 받아들이는 성향을, 정서표현은 감정을 드러내는 것을 두려워하거나 어려워하지 않고 자연스럽게 표현하는 성향을 의미한다.

[0060] 대인회피성의 하위 요인 중 무쾌는 기쁨, 즐거움 등 긍정적 감정을 잘 경험하지 못하는 성향을, 의심은 타인에게 부당하게 해를 입을까봐 사람을 신뢰하지 않고 경계하는 성향을, 고립은 대인관계에 대한 흥미나 욕구가 없어 사람들과 교류하지 않고 거리를 두는 성향을 의미한다.

[0061] 자기중심성의 하위 요인 중 자기애는 타인보다 자신이 우월하고 결점이 없으며 특별하게 대우받아야 한다고 믿는 성향을, 연기는 타인이 관심과 인정을 얻기 위해 자신을 과장하여 표현하는 성향을, 조종은 자신의 이익을 충족하기 위해 타인을 속이거나 자기 뜻대로 움직이려는 성향을, 냉혹은 다른 사람의 상황과 감정에 공감하지 못하여 가혹하고 냉담하게 행동하는 성향을 의미한다.

[0062] 주의곤란성의 하위 요인 중 충동은 상황에 따라 자기 욕구를 제어하지 않고 즉흥적으로 행동하는 성향을, 집착은 중요하지 않은 일에 주의가 고착되어 주의를 전환하기 어려워하는 성향을, 산만은 주의가 쉽게 분산되어 하나의 목표에 집중하기 어려워하는 성향을 의미한다.

[0063] 정신증성의 하위 요인 중 기이는 보통 사람과 달리 특이하고 비관습적인 방식으로 사고하거나 행동하는 성향을, 조현은 내면과 현실을 잘 구분하지 못하여 세계를 왜곡된 방식으로 경험하는 성향을, 경직은 유연성이 없어 모호한 상황을 참기 어려워하며 완고하게 사고하는 성향을 의미한다.

[0064] 부정정서성의 하위 요인 중 불안은 만성적으로 지나치게 긴장하고 경계하며 걱정하는 성향을, 분노는 작은 일에도 쉽게 분노하여 짜증을 내거나 성을 내는 성향을, 열등은 스스로에 대한 만성적으로 부정적인 평가를 하여 자신을 비하하고 비난하는 성향을, 의존은 불안정한 감정을 감당하기 위해 타인에게 전적으로 의지하는 성향을 의미한다.

[0065] 저장부(250)에는 서버(200)에 의한 성격 예측 동작을 수행하기 위한 프로그램, 운영체제 등이 저장될 수 있다. 또한, 저장부(250)에는 설문 제공부(210)에 의해 제공되는 복수의 인터뷰 문항들과 적어도 하나의 지시문, 설문 제공부(210)에 의해 수신되는 복수의 인터뷰 문항들 각각에 대한 주관식 응답과 적어도 하나의 지시문에 대한 응답, 텍스트 변환부(220)에 의한 텍스트 변환의 결과, 텍스트 변환 동작을 수행하기 위한 소프트웨어 및/또는 음성 인식 모델, 전처리부(230)에 의한 전처리 결과, 전처리 동작 중에 일시적으로 또는 비일시적으로 생성되는 데이터, 성격 예측부(240)에 의한 성격 예측의 결과, 성격 예측부(240)에 의한 성격 예측 동작 중에 일시적으로 또는 비 일시적으로 생성되는 데이터, 성격 예측을 위한 성격 예측 모델(즉, 학습된 모델) 등이 저장될 수 있다.

[0066] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPA(Field Programmable array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

[0067] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서,

분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0068] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-optical Media), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0069] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0070] 10 : 성격평가 시스템

100 : 단말

200 : 서버

210 : 설문 제공부

220 : 텍스트 변환부

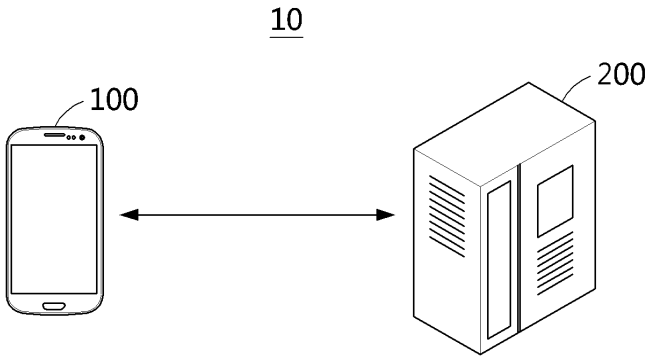
230 : 전처리부

240 : 성격 예측부

250 : 저장부

도면

도면1



도면2

