



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년02월05일
(11) 등록번호 10-2213358
(24) 등록일자 2021년02월02일

(51) 국제특허분류(Int. Cl.)
G06Q 30/02 (2012.01)

(52) CPC특허분류
G06Q 30/0278 (2013.01)
G06N 20/00 (2019.01)

(21) 출원번호 10-2018-0172144

(22) 출원일자 2018년12월28일

심사청구일자 2018년12월28일

(65) 공개번호 10-2020-0087312

(43) 공개일자 2020년07월21일

(56) 선행기술조사문헌

KR101896668 B1*

KR1020140131327 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

아주대학교산학협력단

경기도 수원시 영통구 월드컵로 206 (원천동)

(72) 발명자

한경식

경기도 수원시 영통구 웰빙타운로 70(이의동, 광교 호반베르디움)

(74) 대리인

박기갑, 유민규, 한선희, 안병규

전체 청구항 수 : 총 11 항

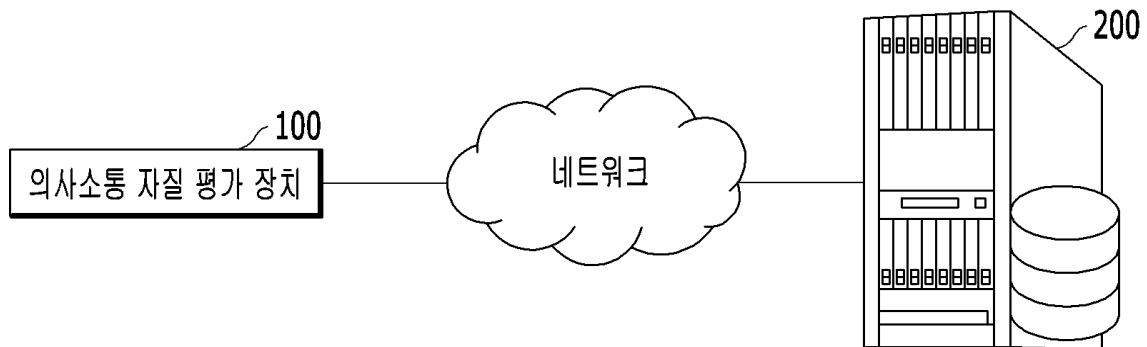
심사관 : 배혜정

(54) 발명의 명칭 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법

(57) 요약

본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치는, 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 수집부; 상기 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성하는 의사소통 자질 정보 생성부; 상기 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화 하고, 상기 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출하는 특징 추출부; 상기 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 모델 구축부; 및 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 상기 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가하는 평가부를 포함할 수 있다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	171107575
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	대학ICT연구센터지원사업
연구과제명	MR-IoT융합 재난대응 인공지능 연구센터
기 여 율	1/1
과제수행기관명	아주대학교 산학협력단
연구기간	2018.06.01 ~ 2021.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

머신러닝 모델 기반 저자의 의사소통 자질 평가 장치에 의한 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법에 있어서,

- (a) 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 단계;
- (b) 상기 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성하는 단계;
- (c) 상기 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화 하는 단계;
- (d) 상기 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출하는 단계; 및
- (e) 상기 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 단계; 및
- (f) 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 상기 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가하는 단계,

를 포함하고,

상기 (b) 단계는,

상호간 연관성을 갖는 자질 파라미터 질문에 대한 응답과, 설문 응답자의 특성 데이터를 고려하되, 상기 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하고, 다변량 선형 회귀 분석을 통해 상기 특성 데이터와 상기 의사소통 자질 정보의 상관관계를 고려하여 상기 의사소통 자질 정보를 생성하고,

상기 (d) 단계는,

상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고,

상기 특징과 상기 의사소통 자질 정보와의 상관관계에 따라, 상기 의사소통 자질 정보에 점수가 부여되고,

상기 (e) 단계는,

상기 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법.

청구항 2

제1항에 있어서,

상기 (a) 단계는,

상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고,

상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법.

청구항 3

제1항에 있어서,

상기 설문 응답자의 특성 데이터는,

상기 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격 중 적어도 어느 하나를 포함하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법.

청구항 4

제1항에 있어서,

상기 (c) 단계는,

전체 온라인 텍스트 저자 중 재샘플링 기법에 기초하여 후보 온라인 텍스트 저자를 선별하고,

상기 의사소통 자질 정보에 대한 미리 설정된 기준값에 기초하여 상기 후보 온라인 텍스트 저자 각각을 상위 그룹 및 하위 그룹으로 분류하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법.

청구항 5

삭제

청구항 6

제1항에 있어서,

상기 언어학 기법은,

언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고,

상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고,

상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법.

청구항 7

머신러닝 모델 기반 저자의 의사소통 자질 평가 장치에 의한 의사소통 자질을 평가하기 위한 모델을 구축하는 방법에 있어서,

(a) 온라인 텍스트를 수집하는 단계;

(b) 설문 응답자에 의한 상기 온라인 텍스트의 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 단계;

(c) 상기 설문 데이터에 기초하여 의사소통 자질 정보를 생성하는 단계;

(d) 상기 온라인 텍스트의 저자 정보 및 상기 온라인 텍스트로부터 특징을 추출하는 단계; 및

(e) 상기 특징을 입력으로 하는 분류 학습 모델에 기초하여 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 단계,

를 포함하고,

상기 (c) 단계는,

상호간 연관성을 갖는 자질 파라미터의 응답과, 설문 응답자의 특성 데이터를 고려하되, 상기 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하고, 다변량 선형 회귀 분석을 통해 상기 특성 데이터와 상기 의사소통 자질 정보의 상관관계를 고려하여 상기 의사소통 자질 정보를 생성하고,

상기 (d) 단계는,

상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고,

상기 특징과 상기 의사소통 자질 정보와의 상관관계에 따라, 상기 의사소통 자질 정보에 점수가 부여되고,

상기 (e) 단계는,

상기 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축하는 것인, 의사소통 자질을 평가하기 위한 모델을 구축하는 방법.

청구항 8

제7항에 있어서,

상기 (a) 단계는,

상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고,

상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함하는 것인, 의사소통 자질을 평가하기 위한 모델을 구축하는 방법.

청구항 9

제7항에 있어서,

상기 언어학 기법은,

언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고,

상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고,

상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출하는 것인, 의사소통 자질을 평가하기 위한 모델을 구축하는 방법.

청구항 10

머신러닝 모델 기반 저자의 의사소통 자질 평가 장치에 있어서,

온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 수집부;

상기 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성하는 의사소통 자질 정보 생성부;

상기 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화 하고, 상기 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출하는 특징 추출부;

상기 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 모델 구축부; 및

신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 상기 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가하는 평가부,

를 포함하고,

상기 의사소통 자질 정보 생성부는,

상호간 연관성을 갖는 자질 파라미터의 응답과, 설문 응답자의 특성 데이터를 고려하되, 상기 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하고, 다변량 선형 회귀 분석을 통해 상기 특성 데이터와 상기 의사소통 자질 정보의 상관관계를 고려하여 상기 의사소통 자질 정보를 생성하고,

상기 특징 추출부는,

상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고,

상기 특징과 상기 의사소통 자질 정보와의 상관관계에 따라, 상기 의사소통 자질 정보에 점수가 부여되고,

상기 모델 구축부는,

상기 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치.

청구항 11

제10항에 있어서,

상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고,

상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치.

청구항 12

제10항에 있어서,

상기 언어학 기법은,

언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고,

상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고,

상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출하는 것인, 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치.

발명의 설명

기술 분야

[0001] 본원은 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법에 관한 것이다.

배경 기술

[0002] 소셜 네트워크 서비스는 개인에게 다양한 정보의 접근을 가능하게 하고, 다양한 사용자들과의 상호작용을 위한 환경을 제공한다. 소셜 네트워크 서비스는 현대인의 일상과 밀접한 관계가 있으며, 다양한 정보의 생성 및 제공으로 인해 그 활용도 또한 증가하고 있다.

[0003] 한편, 소셜 네트워크 서비스는 누구나 접근 가능한 특성으로 인해 정보의 신뢰성 및 정확성에 대한 보장이 명확하지 않으며, 오해의 소지가 있는 정보 또는 잘못된 정보가 쉽고 빠르게 배포될 수 있는 문제점이 있다. 그러나, 소셜 네트워크 서비스를 통해 독자에게 제공되는 모든 정보의 신뢰성을 확인하는 것은 어려운 실정임으로, 독자 개인의 지적 능력, 경험, 지식 등의 기준으로 정보의 신뢰도를 판단해야 하는 한계점이 존재한다. 또한, 종래의 신뢰도 평가 기법들은 단순한 방법으로 온라인 게시물 자체의 신뢰도를 평가하는데 그쳤으며, 온라인 게시물의 저자 자체의 신뢰도를 바탕으로 해당 저자가 작성한 온라인 게시물의 신뢰도를 평가하려는 시도는 존재하지 않았다.

[0004] 본원의 배경이 되는 기술은 한국공개특허공보 제10-2014-0076667호에 개시되어 있다.

발명의 내용

해결하려는 과제

[0005] 본원은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 온라인 텍스트에 대한 독자의 평가에 기초한 머신러닝 모델을 통해 온라인 텍스트 저자의 의사소통 자질을 평가할 수 있는 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법을 제공하는 것을 목적으로 한다.

[0006] 본원은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 신규 온라인 텍스트 저자의 의사소통 자질을 평가할 수 있는 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법을 제공하는 것을 목적으로 한다.

[0007] 다만, 본원의 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들도 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0008] 상기한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법은 (a) 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 단계, (b) 상기 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성하는 단계, (c) 상기

온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화 하는 단계, (d) 상기 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출하는 단계 및 (e) 상기 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 단계 및 (f) 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 상기 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가하는 단계를 포함할 수 있다.

- [0009] 본원의 일 실시예에 따르면, 상기 (a) 단계는, 상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고, 상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함할 수 있다.
- [0010] 본원의 일 실시예에 따르면, 상기 (b) 단계는, 설문 응답자의 특성 데이터를 더 고려하여 의사소통 자질 정보를 생성하고, 상기 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하여 상기 의사소통 자질 정보를 생성하고, 다변량 선형 회귀 분석을 통해 상기 특성 데이터와 상기 의사소통 자질 정보의 상관관계를 산출하되, 상기 설문 응답자의 특성 데이터는, 상기 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격 중 적어도 어느 하나를 포함할 수 있다.
- [0011] 본원의 일 실시예에 따르면, 상기 (c) 단계는, 전체 온라인 텍스트 저자 중 재샘플링 기법에 기초하여 후보 온라인 텍스트 저자를 선별하고, 상기 의사소통 자질 정보에 대한 미리 설정된 기준값에 기초하여 상기 후보 온라인 텍스트 저자 각각을 상위 그룹 및 하위 그룹으로 분류할 수 있다.
- [0012] 본원의 일 실시예에 따르면, 상기 (d) 단계는, 상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고, 상기 (e) 단계는, 상기 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축할 수 있다.
- [0013] 본원의 일 실시예에 따르면, 상기 언어학 기법은, 언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고, 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고, 상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출할 수 있다.
- [0014] 본원의 일 실시예에 따른 의사소통 자질을 평가하기 위한 모델을 구축하는 방법은, (b) 설문 응답자에 의한 상기 온라인 텍스트의 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 단계, (c) 상기 설문 데이터에 기초하여 의사소통 자질 정보를 생성하는 단계, (d) 상기 온라인 텍스트의 저자 정보 및 상기 온라인 텍스트로부터 특징을 추출하는 단계 및 (e) 상기 특징을 입력으로 하는 분류 학습 모델에 기초하여 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 단계를 포함할 수 있다.
- [0015] 본원의 일 실시예에 따르면, 상기 (a) 단계는, 상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고, 상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함할 수 있다.
- [0016] 본원의 일 실시예에 따르면, 상기 (d) 단계는, 상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고, 상기 언어학 기법은, 언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고, 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고, 상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출할 수 있다.
- [0017] 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치는 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집하는 수집부, 상기 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성하는 의사소통 자질 정보 생성부, 상기 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화 하고, 상기 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출하는 특징 추출부, 상기 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축하는 모델 구축부 및 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 상기 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가하는 평가부를 포함할 수 있다.
- [0018] 본원의 일 실시예에 따르면, 상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함하고, 상기 자질 파라미터는, 출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성을 포함할 수 있다.

[0019] 본원의 일 실시예에 따르면, 상기 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 상기 특징을 추출하고, 상기 언어학 기법은, 언어 조사 및 단어 계산 분석 및 코사인유사성을 포함하고, 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출하고, 상기 코사인 유사성은, 상기 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크, 리트윗 중 적어도 어느 하나를 추출할 수 있다.

[0020] 상술한 과제 해결 수단은 단지 예시적인 것으로서, 본원을 제한하려는 의도로 해석되지 않아야 한다. 상술한 예시적인 실시예 외에도, 도면 및 발명의 상세한 설명에 추가적인 실시예가 존재할 수 있다.

발명의 효과

[0021] 전술한 본원의 과제 해결 수단에 의하면, 온라인 텍스트에 대한 독자의 평가에 기초한 머신러닝 모델을 통해 온라인 텍스트 저자의 의사소통 자질을 평가할 수 있는 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법을 제공할 수 있다.

[0022] 전술한 본원의 과제 해결 수단에 의하면, 신규 온라인 텍스트 저자의 의사소통 자질을 평가할 수 있는 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치 및 방법을 제공할 수 있다.

도면의 간단한 설명

[0023] 도 1은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 평가 시스템의 구성을 도시한 도면이다.

도 2는 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 평가 장치의 구성을 도시한 도면이다.

도 3은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 의사소통 자질 평가를 위한 자질 파라미터에 대한 질문의 예를 도시한 도면이다.

도 4 및 도 5는 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 온라인 텍스트의 예를 도시한 도면이다.

도 6은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 자질 파라미터의 다변량 회귀 분석 연산 결과를 도시한 도면이다.

도 7은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법의 흐름을 도시한 도면이다.

도 8은 본원의 일 실시예에 따른 의사소통 자질을 평가하기 위한 모델을 구축하는 방법의 흐름을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0024] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0025] 본원 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다.

[0026] 본원 명세서 전체에서, 어떤 부재가 다른 부재 "상에", "상부에", "상단에", "하에", "하부에", "하단에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.

[0027] 본원 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함" 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다.

[0028] 도 1은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 평가 시스템의 구성을 도시한 도면이고, 도 2는 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 평가 장치의 구성을 도시한 도면이다.

[0029] 도 1을 참조하면, 의사소통 자질 평가 시스템(10)은 의사소통 자질 평가 장치(100) 및 소셜 네트워크 서버(200)를 포함할 수 있다. 소셜 네트워크 서버(200)는 모든 온라인 텍스트 저자가 작성한 온라인 텍스트를 저장하는

데이터베이스를 포함할 수 있다. 또한 의사소통 자질 평가 장치(100) 및 소셜 네트워크 서버(200)는 네트워크로 연결될 수 있다. 네트워크는 단말들 및 서버들과 같은 각각의 노드 상호간에 정보 교환이 가능한 연결 구조를 의미하는 것으로, 이러한 네트워크의 일 예에는, 3GPP(3rd Generation Partnership Project) 네트워크, LTE(Long Term Evolution) 네트워크, 5G 네트워크, WIMAX(World Interoperability for Microwave Access) 네트워크, 인터넷(Internet), LAN(Local Area Network), Wireless LAN(Wireless Local Area Network), WAN(Wide Area Network), PAN(Personal Area Network), wifi 네트워크, 블루투스(Bluetooth) 네트워크, 위성 방송 네트워크, 아날로그 방송 네트워크, DMB(Digital Multimedia Broadcasting) 네트워크 등이 포함되나 반드시 이에 한정되지는 않는다.

[0030] 도 2를 참조하면, 의사소통 자질 평가 장치(100)는 수집부(110), 의사소통 자질 정보 생성부(120), 특징 추출부(130), 모델 구축부(140) 및 평가부(150)를 포함할 수 있다.

[0031] 수집부(100)는 소셜 네트워크 서버(200)로부터 상기 네트워크를 통해 온라인 텍스트를 수집할 수 있다. 또한, 수집부(100)는 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집할 수 있다. 설문 데이터를 수집하기에 앞서 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문에 대해 설명하면, 2016년 6월 내지 8월 올림픽 관련 트윗 즉, 온라인 텍스트와 2017년 1월 내지 4월 지카, 에볼라, 치쿤구니아 바이러스 관련 트윗에 대해 Twitter Search API (<https://dev.twitter.com/rest/public>)를 통해 약 12,000,000개의 온라인 텍스트를 수집하고, 상기 트윗 중 10개 미만의 트윗을 작성한 온라인 텍스트 저자는 필터링한다. 필터링 후, 15,260명의 저자가 산출되었으며, 상기 저자 중 무작위로 1,000의 저자와 각 저자의 트윗을 선정하여 설문에 활용하였다.

[0032] 설문은 SurveyMonkey를 통해 진행되며, 수집부(110)는 SurveyMonkey로부터 설문 데이터를 수집할 수 있다. 설문 에 앞서 설문 응답자에게 설문 조사의 목표와 절차를 공지하고 설문의 동의를 요청하였으며, 동의한 설문자에 한하여 설문을 수행한다. 설문은 크게 두가지 섹션으로 구분될 수 있으며, 첫번째 섹션은 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격에 대한 설문을 설문 응답자에게 제공하며, 각 설문당 1점 내지 5점의 리커트(Likert) 척도로 응답할 수 있다. 예를 들어, 나이의 경우, 10대 1점, 20대 2점, 30대 3점, 40대 4점, 50대 5점으로 구분될 수 있으며, 설문 응답자는 본인이 해당하는 나이 대를 선택하는 방식으로 설문을 진행할 수 있다. 상기 성격에 대한 설문은 5대 성격 특성을 활용하였으며, 성격 특성에는 외향적(사회적 수준), 공감성(조직적, 철저성), 양심적(조직적, 철저성), 신경과민적(감정적 안정성 수준), 개방성(지식과 새로운 경험에 대한 창의성과 욕망 수준)을 포함할 수 있다.

[0033] 설문의 두번째 섹션에서는 각 설문 응답자에게 동일한 온라인 텍스트 저자가 작성한 온라인 텍스트를 제공하고, 의사소통 자질 평가를 위한 자질 파라미터에 대한 질문을 설문 응답자에게 제공한다. 상술한 바와 같이 각 질문은 1점 내지 5점의 척도로 응답할 수 있다. 수집부(110)는 전술한 첫번째 섹션의 설문으로부터 설문 응답자의 특성 데이터를 수집할 수 있고, 두번째 섹션의 설문으로부터 설문 데이터를 수집할 수 있다. 즉, 상기 설문 응답자의 특성 데이터는 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격 중 적어도 어느 하나를 포함할 수 있다. 또한, 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함할 수 있다.

[0034] 도 3은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 의사소통 자질 평가를 위한 자질 파라미터에 대한 질문의 예를 도시한 도면이다.

[0035] 상기 자질 파라미터는 예시적으로, 출처 신뢰도(Source credibility (SC)), 대인관계 매력도(Interpersonal attraction (IA)), 대화 능력(Communication competence (CC)), 및 상호작용 의도성(Intent to interact (INT))을 포함할 수 있다. 도 3을 참조하면, 각 자질 파라미터당 2개 이상의 질문을 포함하며, 각 질문은 리커트 척도에 기반하여 1점(전적으로 동의하지 않음), 2점(동의하지 않음), 3점(보통), 4점(동의), 5점(전적으로 동의) 중 하나를 선택하는 것으로 응답할 수 있다. 또한, 도 3을 참조하면 크론바흐 알파(Cronbach's α)의 결과에 따르면, 각 자질 파라미터의 질문 상호간은 높은 연관성을 나타내는 것을 확인할 수 있다. 예시적으로, 대인관계 매력도에서 Ia1과 Ia2의 질문은 사회적 매력과 관련되고 Ia3과 Ia4는 업무적 매력과 관련된다고 할 수 있다.

[0036] 의사소통 자질 정보 생성부(120)는 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성할 수 있다. 설문 데이터를 통한 의사소통 자질 정보는 전술한 4가지 자질 파라미터에 의존적인 변수일 수 있다. 그러나, 도 3의 크론바흐 알파의 결과에 도시된 바와 같이, 각 자질 파라미터의 질문은 높은 신뢰도를 가지면서 상호 밀접한 상관관계가 있으므로, 자질 파라미터에 대한 응답으로부터 의사소통 자질 정보를 산출하는 것이 유효한 접근법이라고 할 수 있다.

- [0037] 또한, 설문 응답자의 특성 데이터는 의사 소통 자질 정보에 영향을 미칠 수 있다. 예시적으로, 노년층의 설문 응답자는 온라인 텍스트 저자의 의사소통 자질에 부정적인 인식을 가지고 있고, 여성인 설문 응답자는 모든 질문에 대해 남성 응답자보다 긍정적인 인식을 보였다. 이에 의사소통 자질 정보 생성부(120)는 설문 응답자의 특성 데이터를 더 고려하여 의사소통 자질 정보를 생성할 수 있다.
- [0038] 의사소통 자질 정보 생성부(120)는 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하여 의사소통 자질 정보를 생성할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 다변량 선형 회귀 분석을 통해 특성 데이터와 의사소통 자질 정보의 상관관계를 산출할 수 있다. 상기 상관관계를 분석하기 위한 다양한 기법 중 다변량 선형 회귀 분석의 경우, 회귀 결과의 유효성을 측정하기 위해, DurbinWatson 통계 테스트, 분산팽창계수(variance inflation factor, VIF)검사 및 SPSS 통계 분석의 추가적인 측정 및 연산이 요구될 수 있다. 예시적으로, DurbinWatson 통계 테스트는 회귀 분석에서 잔차의 자기 상관 관계를 측정하는 것을 의미한다. DurbinWatson 통계의 결과값은 0 내지 4의 값으로 산출될 수 있다. 결과값이 2인 경우에는 표본에 자기 상관관계가 없음을 의미하고, 결과값이 0에 가까울수록 양(positive)의 자기 상관 관계를 나타낸다. 또한, 결과값이 4에 가까울수록 음(negative)의 자기 상관 관계를 나타낸다.
- [0039] 분산팽창계수(VIF)검사는 최소 자승 회귀 분석에서 다중 공선성의 심각도를 측정하는 것을 의미한다. 분산팽창계수는 공선성에 의해 추정된 회귀 계수의 분산이 얼마나 증가했는지 측정할 수 있는 지수를 제공할 수 있다. 통상적으로 분산팽창계수가 4를 초과하는 경우, 추가조사가 필요하며, 10을 초과하는 경우에는 상당한 다중성을 나타낸다고 해석할 수 있다.
- [0040] 단계적 선형 회귀 특성을 활용한 SPSS 통계 분석은 중요하지 않은 변수들을 동시에 제거하면서, 여러 개의 독립적인 변수를 다시 압축할 수 있다. 이는 가장 작은 상관 변수를 제거할 때마다 여러 회귀 분석으로 이어질 수 있다. 결과적으로, SPSS 통계 분석은 분포를 가장 잘 나타내는 독립 변수를 남기는 분석이다.
- [0041] 전술한 바와 같이, 의사소통 자질 정보 생성부(120)는 의사소통 자질 정보를 생성함에 있어서, 다변량 선형 회귀 분석을 통해 특성 데이터와 의사소통 자질 정보의 상관관계를 고려할 수 있다. 특성 데이터와 의사소통 자질 정보의 상관관계에 의하면, 설문 응답자의 특성 데이터는 의사소통 자질 정보에 영향을 미칠 수 있다는 것을 확인할 수 있다. 따라서, 의사소통 자질 정보 생성 시 설문 응답자의 특성 데이터를 고려하는 것은 중요한 사항이라고 할 수 있다. 의사소통 자질 정보 생성부(120)는 상기 자질 파라미터에 대한 응답의 평균을 통해 하나의 점수로 산출되는 의사소통 자질 정보를 생성할 수 있다. 이 때, 설문 응답자에 의한 설문 데이터는 설문 응답자 개인의 기준에 의한 점수에 기초하여 생성된 것이므로, 이러한 개인의 기준이 전체 의사소통 자질 정보에 영향을 미칠 수 있다. 따라서, 설문에 사용된 온라인 텍스트는 합의(consensus)가 고려될 필요가 있다. 본원의 일 실시예에 따르면, 의사소통 자질 정보 생성부(120)는 설문에 의한 의사소통 자질 정보의 평균 점수의 표준편차 σ 를 고려할 수 있다. 또한 의사소통 자질 정보 생성부(120)는 온라인 텍스트의 비 언어적 요소(예를 들어, 해시태그, 문장부호, 이모티콘, 링크(URL), 리트윗)를 독립 변수로 고려할 수 있다.
- [0042] 도 4 및 도 5는 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 온라인 텍스트의 예를 도시한 도면이다.
- [0043] 도 4는 의사소통 자질 정보가 서로 다른 세가지 온라인 텍스트를 도시하고, 도 5는 의사소통 자질 정보가 상대적으로 높은 온라인 텍스트를 도시한다. 도 4를 참조하면, 도 4의 (a)는 다수의 해시태그로 구성된 온라인 텍스트를 도시한다. 멘션없이 해시태그만으로 트윗을 작성한 경우 낮은 의사소통 자질 정보로 산출될 수 있다. 도 4의 (a)의 의사소통 자질 정보, 즉 점수는 2.59로 산출되었고 표준편차 σ 는 1.0으로 산출되었다. 이는 다수의 설문 응답자들이 해당 온라인 텍스트에 낮은 점수를 주었으나 소수의 설문 응답자들이 높은 점수를 주었다는 것을 의미한다. 도 4의 (b)는 반복되는 멘션과 해시태그를 모두 포함하는 트윗을 도시하며, 이러한 경우에도 도 4의 (a)와 같이 낮은 의사소통 자질 정보로 산출될 수 있다. (점수 2.98, 표준편차 σ 1.3) 도 4의 (c)는 동일한 주제에 대한 유사한 트윗을 도시한다. 도 4의 (c)의 경우, 설문 응답자에게 중복되고 무의미한 트윗으로 인식되어 낮은 점수가 산출될 수 있다. (점수 2.98, 표준편차 σ 1.2)
- [0044] 반면 도 5는 상대적으로 높은 의사소통 자질 정보로 산출된 트윗을 도시한다.(점수 3.75, 표준편차 σ 0.55) 도 5를 참조하면, 설문 응답자들은 해시태그나 반복되는 멘션을 포함하지 않은 트윗에 높은 점수를 주었다는 것을 확인할 수 있다. 살펴본 바와 같이, 상기 비 언어적 요소는 의사소통 자질 정보에 영향을 미치지므로, 의사소통 자질 정보를 산출함에 있어서, 비 언어적 요소를 고려하는 것은 중요한 사항이라 할 수 있다.
- [0045] 특징 추출부(130)는 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화

할 수 있다. 예시적으로, 특징 추출부(130)는 의사소통 자질 정보에 의한 점수가 높은 상위 그룹과 점수가 낮은 하위 그룹으로 구분할 수 있다. 의사소통 자질 정보에 의한 점수의 높고 낮음을 구분하는 기준을 고려하여 의사소통 자질을 평가하기 위한 모델을 구축함으로써 보다 정확한 모델이 구축될 수 있다. 상기 모델의 구축은 보다 뒤에서 살펴보기로 한다.

[0046] 특징 추출부(130)는 전체 온라인 텍스트 저자 중 재샘플링 기법(SMOT)에 기초하여 후보 온라인 텍스트 저자를 선별할 수 있다. 또한, 특징 추출부(130)는 의사소통 자질 정보에 대한 미리 설정된 기준값에 기초하여 후보 온라인 텍스트 저자 각각을 상위 그룹 및 하위 그룹으로 분류할 수 있다. 전술한 설문에 따르면, 각 설문은 리커트 척도로 응답할 수 있다. 이는 2.5점 이하의 점수는 설문 응답자의 부정적인 반응을 나타내고, 3.5점 이상의 점수는 설문 응답자의 긍정적인 반응을 나타낸다고 가정할 수 있다. 따라서, 특징 추출부(130)는 3.5점 이상의 의사소통 자질 정보를 갖는 온라인 텍스트 저자를 상위 그룹으로 분류하고, 2.5점 이하의 의사소통 자질 정보를 갖는 온라인 텍스트 저자를 하위 그룹으로 분류할 수 있다. 전체 온라인 텍스트 저자를 상위 그룹과 하위 그룹으로 분류한 결과, 하위 그룹이 20% 상위 그룹이 80%를 차지하였다. 특징 추출부(130)는 각 그룹에서 동일한 수의 후보 온라인 텍스트 저자를 선별하기 위해 재샘플링 기법 중 하나인 합성 마이너리티 오버 샘플링 기법(Synthetic Minority Over-sampling Technique, SMOTE)에 기초하여 후보 온라인 텍스트 저자를 선별할 수 있다. 특징 추출부(130)는 하위 그룹에서 오버 샘플링하여 상위 그룹과 하위 그룹이 동일한 비율로 후보 온라인 텍스트 저자를 선별할 수 있다. 합성 마이너리티 오버 샘플링 기법은 KNN 접근 방식을 적용한 기법으로 기 공지된 기법이므로 구체적인 설명은 생략한다.

[0047] 또한, 특징 추출부(130)는 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출할 수 있다. 추출된 특징은 의사소통 자질을 평가하기 위한 모델 구축에 활용될 수 있다. 특징 추출부(130)는 5가지 기준에 기초하여 온라인 텍스트로부터 특징을 추출할 수 있다. 구체적으로, 특징 추출부(130)는 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 특징을 추출할 수 있다. 저자 정보는 저자명, 저자 프로필 멘션, 나이, 팔로잉 수, 팔로워 수, 팔로워, 게시물 및 저자의 호감 게시물(예를 들어, 트위터의 '좋아요'를 입력한 게시물) 중 적어도 어느 하나를 포함할 수 있다. 구문 정보는 온라인 텍스트의 리트윗 수, 즐겨찾기, 해시태그, 온라인 텍스트 내의 문장 부호(물음표, 느낌표 등), 단어 수 및 단어 길이 중 적어도 어느 하나를 포함할 수 있다. 유사성 정보는 동일한 저자가 작성한 온라인 텍스트를 TF-IDF벡터로 변환하고 코사인 유사성을 연산하여 주체의 유사성을 산출한 것을 의미한다. 감정 정보는 TextBlob을 사용한 자연어 처리에 기반하여 감정 점수를 산출할 수 있다. 감정 점수는 -1 내지 1의 범위를 가지며, -1은 부정적인 감정 점수를 의미하고 1은 긍정적인 감정 점수를 의미한다.

[0048] 언어학 기법은 언어 조사 및 단어 계산 분석(Linguistic Inquiry and Word Count, LIWC) 및 코사인 유사성을 통해 특징을 추출할 수 있다. 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출할 수 있다. 언어 조사 및 단어 계산 분석에 의하면, 온라인 텍스트 내에 텍스트를 통해 추출할 수 있으나, 비 언어적 요소는 언어 조사 및 단어 계산 분석을 이용할 수 없으므로, 코사인 유사성을 통해 비 언어적 요소로부터 특징을 추출할 수 있다. 코사인 유사성은 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크(URL), 리트윗 중 적어도 어느 하나를 추출할 수 있다. 또한, 코사인 유사성은 동일한 저자가 작성한 온라인 텍스트를 TF-IDF벡터로 변환하여 코사인 유사성을 연산할 수 있다.

[0049] 도 6은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 장치의 자질 파라미터의 다변량 회귀 분석 연산 결과를 도시한 도면이다.

[0050] 도 6을 참조하면, 특징 추출부(130)는 다변량 회귀 분석 언어학 기법 및 코사인 유사성을 통해 추출된 특징과 의사소통 자질 정보와의 상관관계를 산출할 수 있다. 예시적으로, 6글자 이상의 단어(특징)가 많을수록 자질 파라미터(출처 신뢰도, 대인관계 매력도, 대화 능력, 상호작용 의도성)가 이 높은 점수를 부여 받을 수 있다. 6글자 이상의 단어를 사용하는 것은 트윗에서 교육과 사회 계층과 관련된 요소를 의미하며, 이러한 단어를 사용하는 것이 의사소통 자질 평가를 위한 설문에 긍정적인 영향을 미칠 수 있다. 또한, 정관사를 빈번하게 사용할수록 높은 점수의 의사소통 자질 정보가 산출될 수 있다. 구체적으로, 구상 명사, 물체 및 사물에 대한 관심을 나타내는 트윗일수록 상대적으로 높은 점수의 의사소통 자질 정보가 산출될 수 있다.

[0051] 모델 구축부(140)는 추출된 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축할 수 있다. 구체적으로, 모델 구축부(140)는 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 예시적으로, 모델 구축부(140)는 상기 분류 모델 중 랜덤 포레스트(Random Forest) 알고리즘에 기초한 기계학습을 통해 의사소통 자질 정보를 평가할 수 있는 모델을

구축할 수 있다. 모델 구축부(140)는 랜덤 포레스트 알고리즘 외에도 Logistic Regression (LR), Support Vector Machine (SVM), Adaptive boosted Decision Trees (ADT) 등의 알고리즘에 기초하여 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 그러나 상기 각각의 알고리즘에 기초한 모델의 정확성, 정밀성, 검출율을 산출한 결과, 랜덤 포레스트 알고리즘이 가장 우수한 성능을 나타낸 바, 모델 구축부(140)는 랜덤 포레스트 알고리즘에 기초한 기계학습을 통해 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 모델 구축부(140)에서 구축되는 의사소통 자질 정보를 평가할 수 있는 모델은 종래의 랜덤 포레스트 알고리즘에 기초한 모델과 종속 변수가 상이할 뿐만 아니라 라벨링 과정 또한 완전히 새롭게 수행될 수 있다.

[0052] 평가부(150)는 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있다. 평가부(150)는 신규 온라인 텍스트와 신규 온라인 텍스트의 저자 정보로부터 특징을 추출하고, 추출된 특징을 입력으로 하는 의사소통 자질 정보를 평가할 수 있는 모델을 통해 신규 온라인 텍스트의 저자의 의사소통 자질 정보의 점수를 산출할 수 있다.

[0053] 도 7은 본원의 일 실시예에 따른 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법의 흐름을 도시한 도면이고, 도 8은 본원의 일 실시예에 따른 의사소통 자질을 평가하기 위한 모델을 구축하는 방법의 흐름을 도시한 도면이다.

[0054] 도 7에 도시된 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법 및 도 8에 도시된 의사소통 자질을 평가하기 위한 모델을 구축하는 방법은 앞선 도1 내지 도 6을 통해 설명된 의사소통 자질 평가 장치(100)에 의하여 수행될 수 있다. 따라서 이하 생략된 내용이라고 하더라도 도 1 내지 도 6를 통해 의사소통 자질 평가 장치(100)에 대하여 설명된 내용은 도 7 및 도 8에도 동일하게 적용될 수 있다.

[0055] 도 7을 참조하면, 단계 S710에서 수집부(100)는 소셜 네트워크 서버(200)로부터 상기 네트워크를 통해 온라인 텍스트를 수집할 수 있다. 또한, 수집부(100)는 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집할 수 있다. 예시적으로, 설문은 SurveyMonkey를 통해 진행되며, 수집부(110)는 SurveyMonkey로부터 설문 데이터를 수집할 수 있다. 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함할 수 있다. 상기 자질 파라미터는 예시적으로, 출처 신뢰도(Source credibility (SC)), 대인관계 매력도(Interpersonal attraction (IA)), 대화 능력(Communication competence (CC)), 및 상호작용 의도성(Intent to interact (INT))을 포함할 수 있다.

[0056] 단계 S720에서 의사소통 자질 정보 생성부(120)는 설문 데이터에 기초하여 온라인 텍스트 저자별 의사소통 자질 정보를 생성할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 설문 응답자의 특성 데이터를 더 고려하여 의사소통 자질 정보를 생성할 수 있다. 상기 설문 응답자의 특성 데이터는 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격 중 적어도 어느 하나를 포함할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하여 의사소통 자질 정보를 생성할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 다변량 선형 회귀 분석을 통해 특성 데이터와 의사소통 자질 정보의 상관관계를 산출할 수 있다.

[0057] 단계 S730에서 특징 추출부(130)는 온라인 텍스트 저자별 의사소통 자질 정보에 기초하여 복수의 온라인 텍스트 저자를 그룹화할 수 있다. 예시적으로, 특징 추출부(130)는 의사소통 자질 정보에 의한 점수가 높은 상위 그룹과 점수가 낮은 하위 그룹으로 구분할 수 있다. 특징 추출부(130)는 전체 온라인 텍스트 저자 중 재샘플링 기법(SMOT)에 기초하여 후보 온라인 텍스트 저자를 선별할 수 있다. 또한, 특징 추출부(130)는 의사소통 자질 정보에 대한 미리 설정된 기준값에 기초하여 후보 온라인 텍스트 저자 각각을 상위 그룹 및 하위 그룹으로 분류할 수 있다.

[0058] 단계 S740에서 특징 추출부(130)는 그룹화된 복수의 온라인 텍스트 저자의 온라인 텍스트 및 저자 정보로부터 특징을 추출할 수 있다. 특징 추출부(130)는 5가지 기준에 기초하여 온라인 텍스트로부터 특징을 추출할 수 있다. 구체적으로, 특징 추출부(130)는 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 특징을 추출할 수 있다. 언어학 기법은 언어 조사 및 단어 계산 분석(Linguistic Inquiry and Word Count, LIWC) 및 코사인 유사성을 통해 특징을 추출할 수 있다. 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출할 수 있다. 언어 조사 및 단어 계산 분석에 의하면, 온라인 텍스트 내에 텍스트를 통해 추출할 수 있으나, 비 언어적 요소는 언어 조사 및 단어 계산 분석을 이용할 수 없으므로, 코사인 유사성을 통해 비 언어적 요소로부터 특징을 추출할 수 있다. 코사인 유사성은 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크(URL), 리트윗 중 적어도 어느 하나를 추출할 수 있다. 또한, 코사인 유사성은 동일한 저자가 작성한 온라인 텍스트를 TF-IDF벡터로 변환하여 코사인 유사성을

연산할 수 있다.

- [0059] 단계 S750에서 모델 구축부(140)는 추출된 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축할 수 있다. 구체적으로, 모델 구축부(140)는 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 예시적으로, 모델 구축부(140)는 상기 분류 모델 중 랜덤 포레스트(Random Forest) 알고리즘에 기초한 기계학습을 통해 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 모델 구축부(140)에서 구축되는 의사소통 자질 정보를 평가할 수 있는 모델은 종래의 랜덤 포레스트 알고리즘에 기초한 모델과 종속 변수가 상이할 뿐만 아니라 라벨링 과정 또한 완전히 새롭게 수행될 수 있다.
- [0060] 단계 S760에서 평가부(150)는 신규 온라인 텍스트를 수신하고, 상기 구축된 모델을 통해 신규 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있다. 평가부(150)는 신규 온라인 텍스트와 신규 온라인 텍스트의 저자 정보로부터 특징을 추출하고, 추출된 특징을 입력으로 하는 의사소통 자질 정보를 평가할 수 있는 모델을 통해 신규 온라인 텍스트의 저자의 의사소통 자질 정보의 점수를 산출할 수 있다.
- [0061] 도 8을 참조하면, 단계 S810에서 수집부(100)는 소셜 네트워크 서버(200)로부터 상기 네트워크를 통해 온라인 텍스트를 수집할 수 있다.
- [0062] 단계 S820에서, 수집부(100)는 온라인 텍스트 저자의 의사소통 자질 평가를 위한 설문 데이터를 수집할 수 있다. 예시적으로, 설문은 SurveyMonkey를 통해 진행되며, 수집부(110)는 SurveyMonkey로부터 설문 데이터를 수집할 수 있다. 상기 설문 데이터는 의사소통 자질 평가를 위한 자질 파라미터에 대한 응답을 포함할 수 있다. 상기 자질 파라미터는 예시적으로, 출처 신뢰도(Source credibility (SC)), 대인관계 매력도(Interpersonal attraction (IA)), 대화 능력(Communication competence (CC)), 및 상호작용 의도성(Intent to interact (INT))을 포함할 수 있다.
- [0063] 단계 S830에서 의사소통 자질 정보 생성부(120)는 설문 데이터에 기초하여 온라인 텍스트 저자의 의사소통 자질 정보를 생성할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 설문 응답자의 특성 데이터를 더 고려하여 의사소통 자질 정보를 생성할 수 있다. 상기 설문 응답자의 특성 데이터는 설문 응답자의 성별, 나이, SNS 사용기간, SNS 사용빈도 및 성격 중 적어도 어느 하나를 포함할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 설문 데이터와 설문 응답자의 특성 데이터를 이용한 통계적 분석에 기초하여 의사소통 자질 정보를 생성할 수 있다. 또한, 의사소통 자질 정보 생성부(120)는 다변량 선형 회귀 분석을 통해 특성 데이터와 의사소통 자질 정보의 상관관계를 산출할 수 있다.
- [0064] 단계 S840에서 특징 추출부(130)는 온라인 텍스트의 저자 정보 및 온라인 텍스트로부터 특징을 추출할 수 있다. 구체적으로, 특징 추출부(130)는 저자 정보, 구문 정보, 유사성 정보, 감정 정보 및 언어학 기법에 기초하여 특징을 추출할 수 있다. 언어학 기법은 언어 조사 및 단어 계산 분석(Linguistic Inquiry and Word Count, LIWC) 및 코사인 유사성을 통해 특징을 추출할 수 있다. 상기 언어 조사 및 단어 계산 분석은 상기 온라인 텍스트로부터 감정, 스타일, 조사 및 단어 중 적어도 어느 하나를 추출할 수 있다. 언어 조사 및 단어 계산 분석에 의하면, 온라인 텍스트 내에 텍스트를 통해 추출할 수 있으나, 비 언어적 요소는 언어 조사 및 단어 계산 분석을 이용할 수 없으므로, 코사인 유사성을 통해 비 언어적 요소로부터 특징을 추출할 수 있다. 코사인 유사성은 온라인 텍스트로부터 해시태그, 문장부호, 이모티콘, 링크(URL), 리트윗 중 적어도 어느 하나를 추출할 수 있다. 또한, 코사인 유사성은 동일한 저자가 작성한 온라인 텍스트를 TF-IDF벡터로 변환하여 코사인 유사성을 연산할 수 있다.
- [0065] 단계 S850에서 모델 구축부(140)는 추출된 특징을 입력으로 하는 분류 모델에 기초하여 상기 모델을 구축할 수 있다. 구체적으로, 모델 구축부(140)는 추출된 특징에 기초하여 기 설정된 기계학습 방식으로 온라인 텍스트의 저자의 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 예시적으로, 모델 구축부(140)는 상기 분류 모델 중 랜덤 포레스트(Random Forest) 알고리즘에 기초한 기계학습을 통해 의사소통 자질 정보를 평가할 수 있는 모델을 구축할 수 있다. 모델 구축부(140)에서 구축되는 의사소통 자질 정보를 평가할 수 있는 모델은 종래의 랜덤 포레스트 알고리즘에 기초한 모델과 종속 변수가 상이할 뿐만 아니라 라벨링 과정 또한 완전히 새롭게 수행될 수 있다.
- [0066] 본원의 일 실시 예에 따른, 머신러닝 모델 기반 저자의 의사소통 자질 평가 방법 및 의사소통 자질을 평가하기 위한 모델을 구축하는 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 구

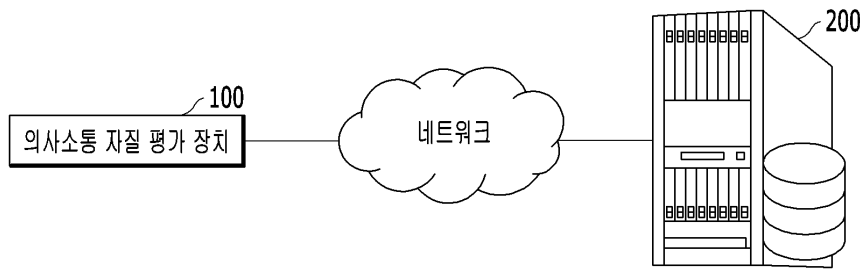
조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 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 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0067] 전술한 본원의 설명은 예시를 위한 것이며, 본원이 속하는 기술분야의 통상의 지식을 가진 자는 본원의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

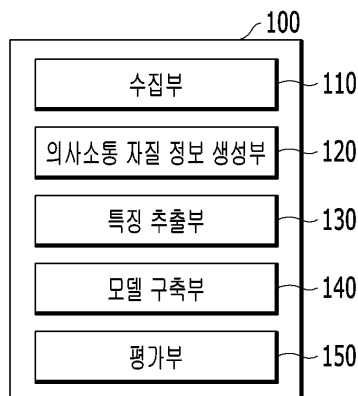
[0068] 본원의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본원의 범위에 포함되는 것으로 해석되어야 한다.

도면

도면1



도면2



도면3

Type	Question (1: Strongly Disagree, 5: Strongly Agree)	α
Source credibility (SC)	SC1 The author seems honest	0.89
	SC2 The author seems trustworthy	
	SC3 The author seems genuine	
Interpersonal attraction (IA)	IA1 The author could be a friend of mine	0.92
	IA2 I could establish a personal friendship with the author	
	IA3 I have confidence in the author's ability to get the job done	
	IA4 If I wanted to get things done, I could probably depend on the author	
Communication competence (CC)	CC1 The author seems effective in accomplishing what was set out to do	0.92
	CC2 The author's tweets seem easy to understand	
	CC3 The author's tweets seem written in a confident style	
Intent to interact (INT)	INT1 If the topic matches my interest, I may want to follow the author	0.93
	INT2 If the topic matches my interest, I may want to receive tweet updates from this author	

도면4

- #pythia #coroebus #goldenaeon #highachievers #olympian #olympics #octaeteris <https://t.co/mcfdMT5CG>
- #SHEbear #octaeteris #olympics #olympian #olympians #pythia #coroebus #goldenaeon #highachievers #judo #tokyo2020 <https://t.co/GCM6YFwdoz>
- #finlit #fintech #pythia #coroebus #goldenaeon #highachievers #olympics #olympian #roadtorio #teamUSA @ <https://t.co/GEupJvWXDo>
- #judo #olympics #olympian #olympians #rio2016 #coroebus #goldenaeon #highachievers #octaeteris <https://t.co/K7Y5Cf6YRs>
- #olympian #olympics #teamUSA <https://t.co/pJJuCDEiq>
- #olympics #olympian #judo #teamUSA <https://t.co/vXHiKuCC3i>

(a)

- RT @ <https://t.co/RynaBE5fih> #rio2016 #refugees @ RIO Olympics <https://t.co/RynaBE5fih>
- 🇧🇷 bongaboi: #rio2016 (at RIO Olympics) <https://t.co/Lh1b8gQt5Y>
- RT @ <https://t.co/Lh1b8gQt5Y>: Congrats to @ <https://t.co/Lh1b8gQt5Y> and @ <https://t.co/Lh1b8gQt5Y> for bringing home the gold in #swimming! GO #TeamUSA! #Rio2016 #Olympics
- RT @ <https://t.co/Lh1b8gQt5Y>: So moving to watch @ <https://t.co/Lh1b8gQt5Y>; win his 20th gold medal in Rio. Congratulations Michael! #Olympics
- RT @ <https://t.co/Lh1b8gQt5Y>: 🍷🍷 @ <https://t.co/Lh1b8gQt5Y> with the EQUALIZER for @ <https://t.co/Lh1b8gQt5Y> in their #Olympics quarterfinals match. 1-1 #USAvsSWE <https://t.co/Lh1b8gQt5Y>
- RT @ <https://t.co/Lh1b8gQt5Y>: #ad #RoadtoRio Watch #Olympics - biggest sports event! Opening Ceremony on Friday, August 5th @ 7:30pm EDT on @ <https://t.co/Lh1b8gQt5Y>
- RT @ <https://t.co/Lh1b8gQt5Y>: Canoe Single (C1) Men's Semifinal #CanoeSlalom #Olympics #Rio2016 <https://t.co/Lh1b8gQt5Y>

(b)

- RT @ <https://t.co/FhnJuZ4Mbt>: Thousands Vaccinated in Brazil Yellow Fever Outbreak <https://t.co/FhnJuZ4Mbt>
- RT @ <https://t.co/FhnJuZ4Mbt>: Brazil Sees Largest Yellow Fever Outbreak in Years #RT <https://t.co/FhnJuZ4Mbt>
- RT @ <https://t.co/FhnJuZ4Mbt>: Thousands Vaccinated in Brazil Yellow Fever Outbreak #TODAY <https://t.co/FhnJuZ4Mbt>
- RT @ <https://t.co/FhnJuZ4Mbt>: Brazil Sees Largest Yellow Fever Outbreak in Years #TODAY <https://t.co/FhnJuZ4Mbt>
- RT @ <https://t.co/FhnJuZ4Mbt>: Thousands Vaccinated in Brazil Yellow Fever Outbreak #D19 <https://t.co/FhnJuZ4Mbt>
- RT @ <https://t.co/FhnJuZ4Mbt>: Brazil Sees Largest Yellow Fever Outbreak in Years... <https://t.co/FhnJuZ4Mbt>

(c)

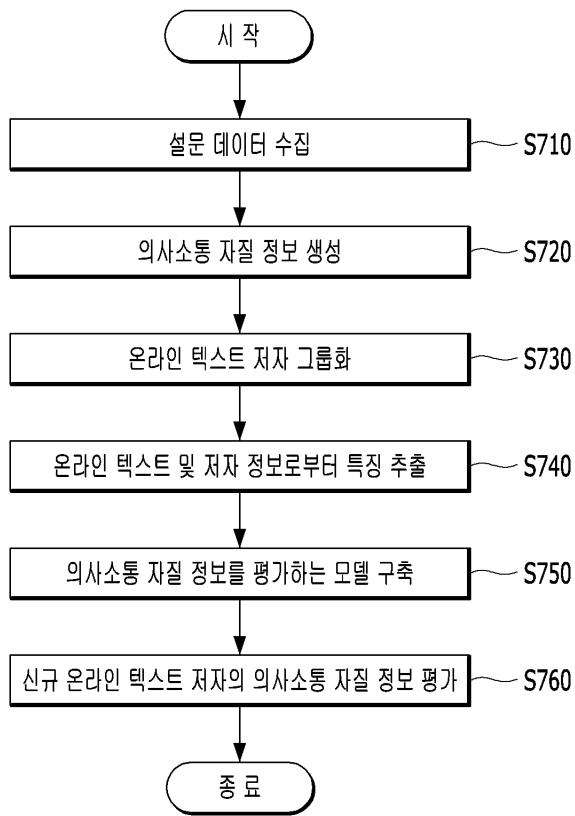
도면5

- Cool and rainy on the long run today. Seven weeks to Olympic qualification #RoadToRio @ <https://t.co/L5bVpvSyf3>
- We are proud to announce that @ <https://t.co/L5bVpvSyf3> has been selected to lead the South African men's marathon team at the Rio Olympic Games.
- RT @ <https://t.co/L5bVpvSyf3>: 2008 @ <https://t.co/L5bVpvSyf3>'s Marathon champ Sammy Wanjiru died on this day in 2011 at 24. Video tribute: <https://t.co/aG4AZ0WKpe>
- A win in a 2:11:27 and Rio Olympic marathon qualification. BOOM! <https://t.co/L5bVpvSyf3>
- Proudly South African. The ultimate achievement is to represent your country at the Olympic Games! @ <https://t.co/L5bVpvSyf3>
- Well done to the marathoners selected today to represent South Africa at the Rio Olympic Games.

도면6

Category	Credibility	Attraction	Competence	Intent	Comm Quality
Word count	-0.058	0.065	-0.001	0.008	0.001
Word per sentence	-0.020	-0.028	-0.040	-0.117*	-0.054
Sixtr words	0.153*	0.162*	0.152*	0.078†	0.155*
Dictionary	-0.068	-0.041	0.049	0.028	-0.040
1 st person singular	-0.038	-0.032	-0.021	-0.066	-0.046
1 st person plural	-0.022	-0.033	-0.074*	-0.021	-0.043
2 nd person	-0.023	-0.036	-0.025	-0.041	-0.036
3 rd person plural	-0.011	-0.025	-0.046	-0.005	-0.026
3 rd person overall	0.043	0.042	0.014	-0.007	0.031
Impersonal pronouns	0.017	0.044	0.037	0.015	0.034
Articles	0.126*	0.156*	0.107*	0.063†	0.117*
Auxiliary verbs	-0.045	-0.020	-0.054	0.079†	-0.020
Adverbs	-0.001	-0.032	0.003	0.056	0.000
Adjectives	0.013	0.016	0.037	0.013	0.022
Past tense	0.008	-0.007	-0.036	-0.012	-0.012
Present tense	0.036	0.075	0.030	-0.062	0.035
Future tense	0.024	0.044	0.021	0.014	0.032
Prepositions	0.124*	0.092*	0.079†	0.065†	0.106*
Conjunctions	0.061	0.077†	0.065†	-0.045	0.056
Negations	-0.056	-0.077†	-0.029	-0.089*	-0.071†
Numbers	0.073†	0.079*	0.056	-0.032	0.060
Quantifiers	0.087*	0.061†	0.048	-0.072*	0.047
Swear words	-0.067*	-0.014	-0.080	-0.005	-0.048
Average # hashtags	-0.091*	-0.043	-0.044	0.064	-0.042
Average # URLs	0.094†	0.053	0.073†	0.169*	0.079†
Average # mentions	-0.098*	-0.048	-0.086*	-0.115*	-0.093*
Average # retweets	-0.013	-0.022	-0.053	-0.010	-0.029
Average # questions	-0.071*	-0.112*	-0.074	-0.059†	-0.095*
Average # exclamations	-0.011	-0.016	-0.015	-0.008	-0.015
Average # stocks	-0.031	-0.001	-0.004	0.009	-0.009
Average # emojis	-0.043	-0.031	-0.030	0.026	-0.027
Average tweet length	0.001	-0.041	0.011	0.040	-0.002
Average sentiment	0.073*	0.069*	0.043	0.065*	0.068*
Tweets similarity	-0.170*	-0.146*	-0.152*	-0.166*	-0.179*
Durbin-Watson	1.434	1.353	1.434	1.208	1.401

도면7



도면8

