



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0150174
(43) 공개일자 2022년11월10일

(51) 국제특허분류(Int. Cl.)
G06Q 50/20 (2012.01) G06N 20/00 (2019.01)
(52) CPC특허분류
G06Q 50/20 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2021-0096586
(22) 출원일자 2021년07월22일
심사청구일자 2021년07월22일
(30) 우선권주장
1020210057368 2021년05월03일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
진진시스템(주)
서울특별시 광진구 아차산로 334, 한강빌딩 5층 (자양동)
(72) 발명자
이상근
서울특별시 강남구 선릉로 221(도곡동, 도곡렉슬 아파트)
김강민
서울특별시 중랑구 상봉중앙로8길 44,A동302호(상봉동)
(뒷면에 계속)
(74) 대리인
윤귀상

전체 청구항 수 : 총 9 항

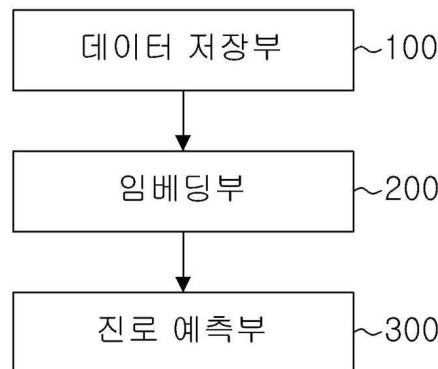
(54) 발명의 명칭 **체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨터 프로그램이 기록된 컴퓨터-판독가능매체**

(57) 요약

본 발명은 체육인 생애주기별 맞춤형 진로예측 장치에 있어서, 체육인 정보가 저장된 데이터 저장부; 상기 체육인 정보로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 임베딩부 및 상기 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로정보를 생성하는 진로예측부를 포함하고, 상기 진로정보는, 체육인이 진출할 수 있는 적어도 하나 이상의 진로에 대한 확률값인 것을 특징으로 하는 체육인 생애주기별 맞춤형 진로예측 장치를 제공할 수 있다.

대표도 - 도1

10



(72) 발명자

박준형

서울특별시 동대문구 이문로1길 21,5동 1102호(회
기동, 청량리신현대아파트)

정용호

서울특별시 관악구 청룡1길 10, 508호(봉천동)

명세서

청구범위

청구항 1

체육인 정보가 저장된 데이터 저장부;

상기 체육인 정보로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 임베딩 부 및

상기 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로 정보를 생성하는 진로예측부를 포함하고,

상기 진로정보는,

체육인이 진출할 수 있는 적어도 하나 이상의 진로에 대한 확률값인 것을 특징으로 하는 체육인 생애주기별 맞춤형 진로예측 장치.

청구항 2

제1항에 있어서,

상기 임베딩부는,

스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습하는 임베딩 학습부;

상기 추출된 수상실적 정보 및 이력정보를 실질형태소만 포함하도록 정제하여 정제 정보를 생성하는 정제 정보 생성부 및

상기 학습된 말뭉치 임베딩 벡터에 기초하여 상기 정제 정보로부터 상기 단어 임베딩 벡터를 생성하는 임베딩 추출부를 포함하는 체육인 생애주기별 맞춤형 진로예측 장치.

청구항 3

제2항에 있어서,

상기 임베딩부는,

패스트텍스트(fastText)를 사용하여 상기 스포츠 도메인 말뭉치 및 상기 정제 정보로부터 음절 단위의 단어를 추출한 다음 상기 말뭉치 임베딩 벡터 및 상기 단어 임베딩 벡터를 생성하는 것을 특징으로 하는 체육인 생애주기별 맞춤형 진로예측 장치.

청구항 4

제1항에 있어서,

상기 진로 예측부는,

상기 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 체육인의 이력 정보를 개월 단위로 구분하는 토큰을 추가하여 상기 단어 임베딩 벡터를 상기 토큰에 적용하는 전처리부를 포함하는 체육인 생애주기별 맞춤형 진로 예측 장치.

청구항 5

제4항에 있어서,

상기 전처리부는,

다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 상기 이력 정보를 년 주기로 분할하는 것을 특징으로 하는 체육인 생애주기별 맞춤형 진로예측 장치.

청구항 6

체육인에 대한 맞춤형 진로 예측 방법에 있어서,

체육인 정보를 데이터 저장부에 저장하는 데이터 저장단계;

상기 체육인 정보로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 임베딩 단계 및

상기 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로 정보를 생성하는 진로 예측단계를 포함하는 체육인 생애주기별 맞춤형 진로예측 방법.

청구항 7

제6항에 있어서,

상기 임베딩 단계는,

스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습하는 학습단계;

상기 추출된 상기 수상실적 정보 및 이력 정보를 실질형태소만 포함하도록 정제하여 정제 정보를 생성하는 정제 정보 생성단계 및

상기 학습된 말뭉치 임베딩 벡터에 기초하여 상기 정제 정보로부터 상기 단어 임베딩 벡터를 생성하는 임베딩 추출단계를 포함하는 체육인 생애주기별 맞춤형 진로예측 방법.

청구항 8

제6항에 있어서,

상기 진로 예측단계는,

상기 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 체육인의 이력 정보를 개월 단위로 구분하는 토큰을 추가하여 상기 단어 임베딩 벡터를 상기 토큰에 적용하고, 다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 상기 이력 정보를 년 주기로 분할하는 전처리단계를 포함하는 체육인 생애주기별 맞춤형 진로예측 방법.

청구항 9

제 6항 내지 8항 중 어느 한 항에 따른 체육인 생애주기별 맞춤형 진로예측 방법을 수행하는 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

발명의 설명

기술 분야

본 발명은 체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨

[0001]

터 프로그램이 기록된 컴퓨터-판독가능매체에 관한 것으로, 더욱 자세하게는 체육인의 진로 예측을 위해 체육 종목별로 상이한 체육인의 수상 실적, 경기 참여 실적 및 이력을 분석하고 과거 체육인들의 진로를 바탕으로 각 체육인 개인의 생애주기별로 적절한 진로를 예측해주는 체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨터 프로그램이 기록된 컴퓨터-판독가능매체에 관한 것이다.

배경 기술

- [0003] 전문 체육인이 급속도로 증가하고 있지만, 전문 체육인의 특성상 일부 성공한 소수의 체육인을 제외하고 다수의 체육인은 진로 및 커리어 개발에 상당한 어려움을 겪는다. 이에 따라 체육인의 종목과 개인적인 특성 및 이력을 고려한 적절한 개인별 진로 가이드가 필요하다. 이를 위해선 개별 체육인이 가지고 있는 수상 실적, 이력 등을 바탕으로 과거 체육인의 진로 정보를 통해 개별 체육인에게 진로 가이드를 해주어야 한다.
- [0004] 하지만 기존의 체육 관련 진로 가이드 시스템은 설문조사에 기반하여 적합한 종목을 추천해주는 것에 머물렀으며, 체육인들의 수상 실적과 이력은 자연어 형태로 기록되어 정량적인 분석이 어렵고, 체육 종목은 수십가지가 넘고, 이러한 종목 간의 이력 정보는 서로 상이한 문제가 있다.
- [0005] 또한 과거의 수천명의 방대한 체육인의 정보를 참조하여 적합한 진로를 추천해주는 것은 상당한 시간과 노력이 필요하다. 체육 분야로 진로를 삼은 체육인들은 소수의 국가대표, 프로선수 등으로 성공하는 경우를 제외하고는 다수의 체육인들은 진로에 어려움을 겪고 있기 때문에 체육인들의 커리어를 고려한 적절한 진로 가이드가 필요함에 따라 효율적인 진로 가이드를 위해 종목마다 상이한 체육인의 이력을 자동으로 분석하고, 이를 바탕으로 과거 체육인의 정보를 통해 진로를 예측해줄 수 있는 기술이 필요하다.
- [0006] 한편, 최근 인공지능 기술 중 하나인 딥러닝 기술의 여러 분야(e.g. 금융, 기상 등)에서 우수한 성능을 보였다.
- [0007] 이에 따라, 딥러닝 기술과 오랜 기간 동안 축적된 약 백여개의 종목에 대한 전문 체육인과 비전문 체육인의 데이터들로 체육인들의 진로를 예측해주어 가이드를 해줄 수 있는 연구가 필요한 실정이다.

선행기술문헌

특허문헌

- [0009] (특허문헌 0001) 대한민국 등록특허 제10-2143586호(2020.08.05)

발명의 내용

해결하려는 과제

- [0010] 상기와 같은 문제를 해결하고자, 본 발명은 체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨터 프로그램이 기록된 컴퓨터-판독가능매체에 관한 것으로서, 체육인들의 개별 이력 및 수상 실적을 분석하고 이를 과거 체육인들의 진로에 기반하여 적절한 진로를 추천해주기 위해, 개별 체육인의 이력 및 수상 실적을 분석하여 과거 체육인들의 진로를 통해 적절한 진로를 추천하여 가이드를 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0012] 상기와 같은 과제를 해결하기 위하여, 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치는, 체육인 정보가 저장된 데이터 저장부; 상기 체육인 정보로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 임베딩부 및 상기 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로정보를 생성하는 진로예측부를 포함하고, 상기 진로정보는, 체육인이 진출할 수 있는 적어도 하나 이상의 진로에 대한 확률값인 것을 특징으로 한다.
- [0013] 또한, 상기 임베딩부는 스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습하는 임베딩 학습

부; 상기 추출된 수상실적 정보 및 이력정보를 실질형태소만 포함하도록 정제하여 정제 정보를 생성하는 정제 정보 생성부 및 상기 학습된 말뭉치 임베딩 벡터에 기초하여 상기 정제 정보로부터 상기 단어 임베딩 벡터를 생성하는 임베딩 추출부를 포함할 수 있다.

[0014] 또한, 상기 임베딩부는, 패스트텍스트(fastText)를 사용하여 상기 스포츠 도메인 말뭉치 및 상기 정제 정보로부터 음절 단위의 단어로 추출한 다음 상기 말뭉치 임베딩 벡터 및 단어 임베딩 벡터를 생성하는 것을 특징으로 한다.

[0015] 또한, 상기 진로 예측부는, 상기 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 체육인의 이력 정보를 개월 단위로 구분하는 토큰을 추가하여 상기 단어 임베딩 벡터를 상기 토큰에 적용하는 전처리부를 포함할 수 있다.

[0016] 또한, 상기 전처리부는, 다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 상기 이력 정보를 년 주기로 분할하는 것을 특징으로 한다.

[0018] 또한, 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 이용한 진로예측 방법은, 체육인 정보를 데이터 저장부에 저장하는 데이터 저장단계; 상기 데이터 저장부로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 임베딩 단계 및 상기 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로정보를 생성하는 진로 예측단계를 포함하는 체육인 생애주기별 맞춤형 진로예측 방법을 제공할 수 있다.

[0019] 또한, 상기 임베딩 단계는, 스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습하는 학습단계; 상기 추출된 상기 수상실적 정보 및 이력 정보를 실질형태소만 포함하도록 정제하여 정제 정보를 생성하는 정제 정보 생성단계 및 상기 학습된 말뭉치 임베딩 벡터에 기초하여 상기 정제 정보로부터 상기 단어 임베딩 벡터를 생성하는 임베딩 추출단계를 포함할 수 있다.

[0020] 또한, 상기 진로 예측단계는, 상기 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 체육인의 이력 정보를 개월 단위로 구분하는 토큰을 추가하여 상기 단어 임베딩 벡터를 상기 토큰에 적용하고, 다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 상기 이력 정보를 년 주기로 분할하는 전처리단계를 포함할 수 있다.

[0022] 또한, 본 발명의 실시예에 따라 체육인 생애주기별 맞춤형 진로예측 방법을 수행하는 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체를 제공할 수 있다.

발명의 효과

[0024] 상기와 같은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨터 프로그램이 기록된 컴퓨터-판독가능매체는 체육인의 종목을 고려하고, 개인의 이력 정보를 파악하여 맞춤형 교육 및 진로 가이드를 제공함으로써 체육인이 최적의 커리어를 개발할 수 있도록 한다.

[0025] 또한, 위에서 언급된 본 발명의 실시 예에 따른 효과는 기재된 내용에만 한정되지 않고, 명세서 및 도면으로부터 예측 가능한 모든 효과를 더 포함할 수 있다.

도면의 간단한 설명

[0027] 도 1은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 순서도.
 도 2는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 임베딩 추출부의 순서도.
 도 3은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 진행 단계를 도시한 예시도.
 도 4 및 도 5는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 통해 체육인의 진로를 예

측했을 때, 실제 진로에 대하여 예측된 진로를 평가한 값을 나타낸 표.

도 6은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 방법을 순차적으로 나타낸 흐름도.

도 7은 도 6의 S20 단계에서 이루어질 수 있는 단계들을 개략적으로 나타낸 흐름도.

발명을 실시하기 위한 구체적인 내용

- [0028] 이하, 도면을 참조한 본 발명의 설명은 특정한 실시 형태에 대해 한정되지 않으며, 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있다. 또한, 이하에서 설명하는 내용은 본 발명의 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0029] 이하의 설명에서 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용되는 용어로서, 그 자체에 의미가 한정되지 아니하며, 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0030] 본 명세서 전체에 걸쳐 사용되는 동일한 참조번호는 동일한 구성요소를 나타낸다.
- [0031] 본 발명에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 또한, 이하에서 기재되는 "포함하다", "구비하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것으로 해석되어야 하며, 하나 또는 그 이상의 다른 특징들이나, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0032] 또한, 명세서에 기재된 "편부", "편기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0033] 이하, 첨부된 도면을 참조하여 본 발명의 실시 예에 따른 체육인 생애주기별 맞춤형 진로예측 장치, 이를 이용한 진로 예측 방법 및 이를 제공하기 위한 컴퓨터 프로그램이 기록된 컴퓨터-판독가능매체를 상세히 살펴보기로 한다.
- [0035] 참고로, 이하의 설명에서는 본 발명의 용이한 이해를 위해 체육인을 예시로 설명하지만, 체육인 이외의 다른 직업군에서도 본 발명의 실시예들이 동일한 방식으로 적용될 수 있다.
- [0037] 도 1은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 순서도이고, 도 2는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 임베딩 추출부의 순서도이며, 도 3은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치의 진행 단계를 도시한 예시도이다.
- [0039] 도 1 내지 3을 참조하면, 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치(10)는 체육인의 수상 실적 및 이력 정보를 이용하여 체육인의 진로를 예측하는 시스템으로, 데이터 저장부(100), 임베딩부(200) 및 진로 예측부(300)를 포함할 수 있다.
- [0040] 데이터 저장부(100)는 진로 예측 대상인 체육인과 관련된 모든 데이터를 저장하는 곳으로, 체육인 개인의 이력서, 수상정보, 체육인과 관련된 기사 등을 포함할 수 있다.
- [0041] 도 2를 참조하면, 임베딩부(200)는 데이터 저장부(100)로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성하는 곳으로, 임베딩 학습부(210), 정제 정보 생성부(220) 및 임베딩 추출부(230)를 포함할 수 있다.
- [0042] 먼저, 임베딩 학습부(210)는 스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습할 수 있다.
- [0043] 더욱 구체적으로, 임베딩 학습부(210)는 '국립국어원 모두의 말뭉치'의 신문 기사에서 각 스포츠별로 기사를 추출하여 스포츠 도메인 말뭉치를 수집할 수 있다. 이후, 임베딩 학습부(210)는 음절 단위로 단어 표현을 생성하는 서브 단어 표현 모델인 패스트텍스트(fastText) 방법을 이용하여 스포츠 도메인 말뭉치를 임베딩 학습에 맞게 가공하여 가공 데이터를 생성할 수 있다.
- [0044] 이때, 가공 데이터는 스포츠 도메인 말뭉치에 대해 학습할 단어 목록을 추출하고, 추출된 각각의 단어에 대한

하위 단어 정보일 수 있다.

- [0045] 이에, 임베딩 학습부(210)는 가공 데이터를 가지고 입출력층 및 프로젝션층으로 이루어진 학습 모델을 이용한 단어 임베딩 학습을 통해 학습할 단어를 학습하여 벡터화 한 말뭉치 임베딩 벡터를 생성할 수 있다.
- [0046] 여기서, 패스트텍스트(fastText) 방법은 워드 임베딩 및 텍스트 분류를 학습하기 위한 라이브러리로, 단어를 벡터로 만드는 방법 중 하나이다. 더욱 구체적으로 패스트텍스트(fastText) 방법은 하나의 단어 안에도 여러 단어들 존재하는 것으로 간주하며, 단어 안의 여러 단어들을 고려한 학습을 수행할 수 있다.
- [0047] 정제 정보 생성부(220)는 데이터 저장부(100)로부터 체육인의 수상실적 정보 및 이력 정보를 추출한 다음 정제하여 정제 정보를 생성하는 곳이다.
- [0048] 이때, 수상실적 정보는 "유소년 대회 입상", "프로리그 우승", "아마추어팀 리그 우승" 등 체육대회와 관련된 수상에 대한 모든 정보를 포함할 수 있으며, 이력 정보는 "프로 입단", "아마추어팀 감독 부임" 등 체육인의 이력에 관련된 모든 정보를 포함할 수 있다.
- [0049] 상기와 같이 추출된 수상실적 정보 및 이력 정보는 고유명사(인명, 팀명, 학교 등)를 보존할 수 있도록 한국어 형태소 분석기를 활용하여 데이터를 정제하는 것이 가장 바람직하다.
- [0050] 여기서, 한국어 형태소 분석기로는 Kiwi, HAM, HLX, Mecab 등 다양한 종류가 존재하는데 Kiwi 형태소 분석기를 이용하는 것이 가장 바람직하나, 이에 한정되지 않는다.
- [0051] 더욱 구체적으로, Kiwi 형태소 분석기를 이용하여 수상실적 정보 및 이력 정보에 대한 형태소 분석을 수행하고, 그 결과를 활용할 수 있다.
- [0052] 이에 따라, 정제 정보 생성부(220)는 수상실적 정보 및 이력 정보에 포함된 '을', '를' 같은 조사나 '하다', '였다'와 같은 어미들은 제거하여, 추후 단어 임베딩 벡터를 가지지 않도록 할 수 있다. 즉, 정제 정보 생성부(220)는 정제 정보가 일반명사, 동사 및 형용사 등의 실질형태소만 포함하도록 하여 불필요한 정보는 사전에 제거할 수 있는 장점이 있다.
- [0053] 임베딩 추출부(230)는 정제 정보로부터 단어 임베딩 벡터를 생성할 수 있다.
- [0054] 이때, 임베딩 추출부(230)는 상기 임베딩 학습부(210)에서와 같이 패스트텍스트(fastText) 방법을 이용하여 정제 정보에 포함된 단어들에 대한 하위 단어 정제정보를 추출한 다음 추출된 하위 단어 정제정보로부터 단어 임베딩 벡터를 생성할 수 있다.
- [0055] 이에 따라, 임베딩부(200)는 임베딩 학습부(210)를 통해 말뭉치 임베딩 벡터를 학습하고, 학습된 말뭉치 임베딩 벡터를 바탕으로 체육인의 수상실적 정보 및 이력 정보를 단어 임베딩 벡터를 생성하여 진로 예측부(300)로 하여금 체육인의 진로예측을 수행하도록 할 수 있다.
- [0056] 진로 예측부(300)는 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로정보를 생성할 수 있다.
- [0057] 여기서, 체육인의 진로정보는 체육인이 진출할 수 있는 적어도 하나 이상의 진로에 대한 확률값을 의미한다. 이때, 체육인과 관련된 진로는 은퇴, 행정가, 지도자, 국가대표, 청소년대표 및 심판 중 하나 이상일 수 있으며, 이에 한정되지 않고, 체육인과 관련된 진로는 모두 포함할 수 있다.
- [0058] 더욱 구체적으로, 진로 예측부(300)는 CNN(Convolutional Neural Network) 모델을 통해 임베딩 학습부(210)의 말뭉치 임베딩 벡터와 임베딩 추출부(230)의 단어 임베딩 벡터의 상관 관계에 대한 학습 모델을 획득할 수 있다.
- [0059] 이에, 진로 예측부(300)는 학습 모델에 기초하여 체육인의 진로정보를 생성하기 위한 목적에 따라 입력된 단어 임베딩 벡터에 대해 분할 및 조합 분석을 수행하여 진로정보를 생성할 수 있다.
- [0060] 이때, 진로 예측부(300)는 CNN(Convolutional Neural Network) 모델의 Input Layer, Convolutional Layer, Dropout Layer, Hidden Layer 및 Output Layer를 통해 수행될 수 있다.
- [0061] Input Layer는 단어 임베딩 벡터를 입력 받아 Convolutional Layer에 전달할 수 있다.
- [0062] Convolutional Layer는 단어 임베딩 벡터를 특정한 연속된 단어의 수로 분할 및 조합하면서 데이터를 분석할 수 있다. 이 경우, Convolutional Layer가 도출한 결과는 각기 다른 개수의 단어로 구성된 부분 단어 집합을 분석

한 결과일 수 있다.

- [0063] Dropout Layer는 분석 결과로부터 일부 정보를 탈락시킬 수 있다. 이러한 탈락 과정에 의해, 인공지능 분석 모델이 학습데이터에 과하게 조정되는 오버 피팅 현상을 방지하고, 현실에서 발생 가능한 비정형 텍스트의 임의성을 추가함으로써, 분석 결과의 품질이 향상될 수 있다.
- [0064] 이후, 데이터는 Hidden Layer와 Output Layer에 의해 데이터가 순차적으로 처리되고, Output Layer를 통하여 진로 정보 생성을 위한 하나의 수치 정보가 도출된다. 이 경우, 진로 정보에 주어진 확률값을 비교하여, 오차의 정도를 분석 모델에 반영하는 반복적인 학습 과정이 수행될 수 있다. 전체 학습 과정이 수행되는 동안, 단어 임베딩 벡터에 대하여 반복적인 학습을 수행할 수 있다.
- [0065] 이때, 진로 예측부(300)는 진로 정보 각각의 초기 확률값을 1로 설정하여 학습할 수 있다.
- [0066] 여기서, 초기 확률값은 원핫 벡터(one-hot vector)의 형태로 표현한 값으로, 하나의 요소는 1이고, 나머지는 0으로 표현할 수 있으며, 각각의 값은 100%, 0%의 의미를 가질 수 있다.
- [0067] 본 발명의 진로 예측부(300)는 진로 정보 각각에 대하여 정답에 대한 확률값을 원핫 벡터의 형태인 1로 설정하여 학습함으로써, 체육인이 다양한 진로로 동시 진출이 가능한 점을 고려할 수 있다.
- [0068] 예컨대, 진로 정보가 [국가대표, 지도자, 행정가]로 구성되었을 때, 해당 체육인이 국가대표를 지낸 후 지도자가 된 경우, 초기 확률값을 [1, 1, 0]으로 부여한 뒤 학습할 수 있다.
- [0069] 이에 따라, 진로 예측부(300)는 체육인에 대해 진로 정보를 한가지로 한정하지 않고 해당 체육인에 적합한 진로 정보를 모두 도출하여, 체육인으로 하여금 선택의 폭을 넓힐 수 있도록 할 수 있다.
- [0070] 또한, 진로 예측부(300)는 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 체육인의 이력 정보를 6개월 단위로 구분하는 토큰을 추가하여 단어 임베딩 벡터를 상기 토큰에 적용하는 전처리부를 포함할 수 있다.
- [0071] 이때, 전처리부는 다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 이력 정보를 년 주기로 분할할 수 있다.
- [0072] 이와 같이, 진로 예측부(300)는 이력 정보를 년 주기로 분할한 다음, 토큰을 통해 상기 이력 정보를 개월 단위로 분류함으로써 체육인의 이력 정보를 시간 흐름에 따라 잘 구분할 수 있다. 이때, 진로 예측부(300)는 이력 정보를 3년 주기로 분할하고, 토큰을 6개월 단위로 분류하는 것이 가장 바람직하나, 이에 한정되지 않는다. 이를 통해 진로 예측부(300)는 체육인의 진로 예측 시 보다 빠르고 정확하게 진로 정보를 생성하는 장점이 있다.
- [0073] 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치(10)는 상기에서 설명한 단계에 따라 체육인의 진로 정보를 생성할 수 있다.
- [0074] 예컨대, 도 3을 참조하면, 상기와 같은 단계를 통해 신문 기사나 다양한 기사에 '2011년 8월 서울 제3회 재능기 전국 중학교 배구대회 참여'와 같이 기재되어 있는 체육인 관련 수상실적 정보 및 이력 정보를 추출하고, 추출된 정보를 임베딩하여 벡터로 표현한 다음, 상기 벡터를 가지고 '국가대표 확률 42%, 지도자 확률 13% 등'과 같이 체육인의 진로를 예측할 수 있다.
- [0076] 도 4 내지 도 5는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 통해 체육인의 진로를 예측했을 때, 실제 진로에 대하여 예측된 진로를 평가한 값을 나타낸 표이다.
- [0078] 도 4 내지 도 5를 참조하면, 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 통해 체육인의 진로를 예측했을 때, 그 체육인이 실제 진로를 얼마나 맞추었는지를 확인할 수 있다.
- [0079] 이때, 체육인의 진로는 총 6개의 범주(은퇴, 행정가, 지도자, 국가대표, 청소년대표, 심판)로 하였으며, 성능 평가 척도는 총 3가지로 1-best accuracy, Recall@2, Full accuracy이다.
- [0080] 1-best accuracy는 총 6개의 범주 중 모델이 가장 강하게 예측하는 것을 하나 고르고, 그것이 실제 진로에 있었는지를 평가한다. Recall@2는 6개의 범주 중 모델이 두 개의 진로 예측을 하고, 그 두 개 중 하나라도 실제 진로에 있었는지를 평가한다. Full Accuracy는 6개의 범주 중 1개 이상의 진로를 예측하고 예측한 진로가 실제 진로와 완전히 정확한지를 평가한다. 1-best accuracy는 수식 1을 통해 계산할 수 있고, Recall@2는 수식 2를 통

해, Full accuracy는 수식 3을 통해 계산할 수 있다.

[수식 1]

$$1 - \text{best accuracy} = \frac{\text{Correct predictions}}{\text{All predictions}}, \text{Correct prediction} := \text{prediction} \in \{\text{label}_1, \dots, \text{label}_k\}$$

이때, label은 실제 체육인의 진로이다. 1-best accuracy는 Correct predictions를 All predictions으로 나눈 값이다. 모델의 모든 예측은 All predictions 이고, Correct predictions 는 모델이 예측한 하나의 진로가 실제 진로에 포함되어 있는 경우이다.

[수식 2]

$$\text{Recall@2 accuracy} = \frac{\text{Correct recall@2s}}{\text{All predictions}}, \text{Correct recall@2} := \text{prediction} \cap \text{labels} \neq \emptyset$$

이때 label은 실제 체육인의 진로이다. Recall@2 accuracy는 Correct recalls를 All predictions로 나눈 값이다. 모델의 모든 예측은 All prediction이고, Correct recall@2는 모델이 2개를 예측했을 때 그 두 개중 하나라도 실제 진로에 있는 경우이다.

[수식 3]

$$\text{Full-accuracy} = \frac{\text{Correct full predictions}}{\text{All predictions}}, \text{Correct full prediction} := \text{prediction} = \text{labels}$$

이때 label은 실제 체육인의 진로이다. Full-accuracy는 Correct full predictions를 All predictions로 나눈 값이다. 모델의 모든 예측을 All prediction이고, Correct full prediction은 모델이 예측한 N개가 정확하게 라벨과 동일할 경우이다.

위의 평가 지표로 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치(10)의 성능을 측정하였다. 스포츠 도메인 임베딩 모델 기법의 성능을 검증하기 위해 스포츠 도메인 임베딩을 사용하지 않았을 경우, 단어 표현 모델인 word2vec 방법을 사용한 경우, 서브 단어 표현 모델인 패스트텍스트(fastText) 방법을 사용한 경우를 각각 검증하였다. 또한 각각의 임베딩 모델에 대해서 진로 예측을 위한 순환 신경망 모델인 LSTM과 합성곱 신경망 모델인 CNN을 사용하였다.

도 4는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 통해 배구 체육인 진로 예측 성능을 측정한 결과이다. 스포츠 도메인 임베딩 모델과 진로 예측 모델을 조합하여 각각 성능을 측정한 결과다. 표를 통해 서브 단어 표현 모델인 패스트텍스트(fastText) 방법으로 임베딩하고, 합성곱 신경망인 CNN을 사용했을 때, 진로 예측을 더 효과적으로 할 수 있음을 확인할 수 있다.

도 5는 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치를 통해 스피드 스케이팅 체육인 진로 예측 성능을 측정한 결과이다. 스포츠 도메인 임베딩 모델과 진로 예측 모델을 조합하여 각각 성능을 측정한 결과다. 표를 통해 서브 단어 표현 모델인 패스트텍스트(fastText) 방법으로 임베딩하고, 합성곱 신경망인 CNN을 사용했을 때, 진로 예측을 더 효과적으로 할 수 있음을 확인할 수 있다.

이러한 체육인 생애주기별 맞춤형 진로예측 방법에 대하여 하기에서 구체적으로 설명하기로 한다.

- [0102] 도 6은 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 방법을 순차적으로 나타낸 흐름도이고, 도 7은 도 6의 S20 단계에서 이루어질 수 있는 단계들을 개략적으로 나타낸 흐름도이다.
- [0104] 도 6을 참조하면, 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 방법은 데이터 저장단계(S10), 임베딩 단계(S20) 및 진로 예측단계(S30)를 포함할 수 있다.
- [0105] 데이터 저장단계(S10)는 체육인 정보를 데이터 저장부(100)에 저장하여 데이터셋이 구성되도록 할 수 있다.
- [0106] 임베딩 단계(S20)는 데이터 저장부(100)로부터 체육인의 수상실적 정보 및 이력 정보를 추출하여 단어 임베딩 벡터를 생성할 수 있다.
- [0107] 도 7을 참조하면, 임베딩 단계(S20)는 학습단계(S21), 정제 정보 생성단계(S22) 및 임베딩 추출단계(S23)를 포함할 수 있다. 임베딩 단계(S20)를 구성하는 각각의 단계에 대해서는 상기 장치에서 설명하였으므로, 자세한 설명은 생략하기로 한다.
- [0108] 학습단계(S21)는 스포츠 도메인 말뭉치를 기반으로 말뭉치 임베딩 벡터를 생성하여 학습할 수 있다.
- [0109] 정제 정보 생성단계(S22)는 데이터 저장부(100)로부터 체육인의 수상실적 정보 및 이력 정보를 추출한 다음 Kiwi 형태소 분석기를 이용하여 상기 수상실적 정보 및 이력 정보가 실질형태소만 포함하도록 정제한 정제 정보를 생성할 수 있다.
- [0110] 임베딩 추출단계(S23)는 정제 정보를 말뭉치 임베딩 벡터와 매칭하여 단어 임베딩 벡터를 생성할 수 있다.
- [0111] 마지막으로, 진로 예측단계(S30)는 단어 임베딩 벡터를 합성곱 신경망인 CNN(Convolutional Neural Network) 모델에 입력하여 체육인의 진로정보를 생성할 수 있다.
- [0112] 보다 구체적으로, 진로 예측단계(S30)는 CNN(Convolutional Neural Network) 모델을 통해 체육인의 진로정보를 생성하기 위한 목적에 따라 입력된 단어 임베딩의 분할 및 조합 분석을 수행하고 입력된 데이터의 패턴을 파악하여 진로에 대한 확률값을 도출할 수 있다.
- [0113] 이때, 진로 예측단계(S30)는 단어 임베딩 벡터를 상기 CNN 모델에 입력하기 전에 수행되는 전처리단계를 포함할 수 있다.
- [0114] 전처리단계는 패턴을 찾기 전에 체육인의 이력 정보를 개월 단위로 구분하는 토큰을 추가하여 상기 단어 임베딩 벡터를 상기 토큰에 적용하고, 다수의 훈련 데이터를 생성하기 위해 생애주기가 완성된 체육인의 상기 이력 정보를 년 주기로 분할할 수 있다.
- [0116] 이상에서 설명한 본 발명의 실시예에 따른 체육인 생애주기별 맞춤형 진로예측 장치는, 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 이러한 기록 매체는 컴퓨터 판독 가능 매체를 포함하며, 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 포함하며, 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.
- [0118] 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.

부호의 설명

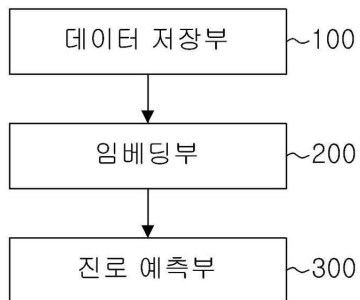
- [0120] 10: 체육인 생애주기별 맞춤형 진로예측 장치
100: 데이터 저장부

200: 임베딩부
210: 임베딩 학습부
220: 정제 정보 생성부
230: 임베딩 추출부
300: 진로 예측부

도면

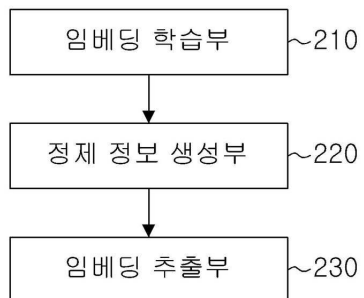
도면1

10

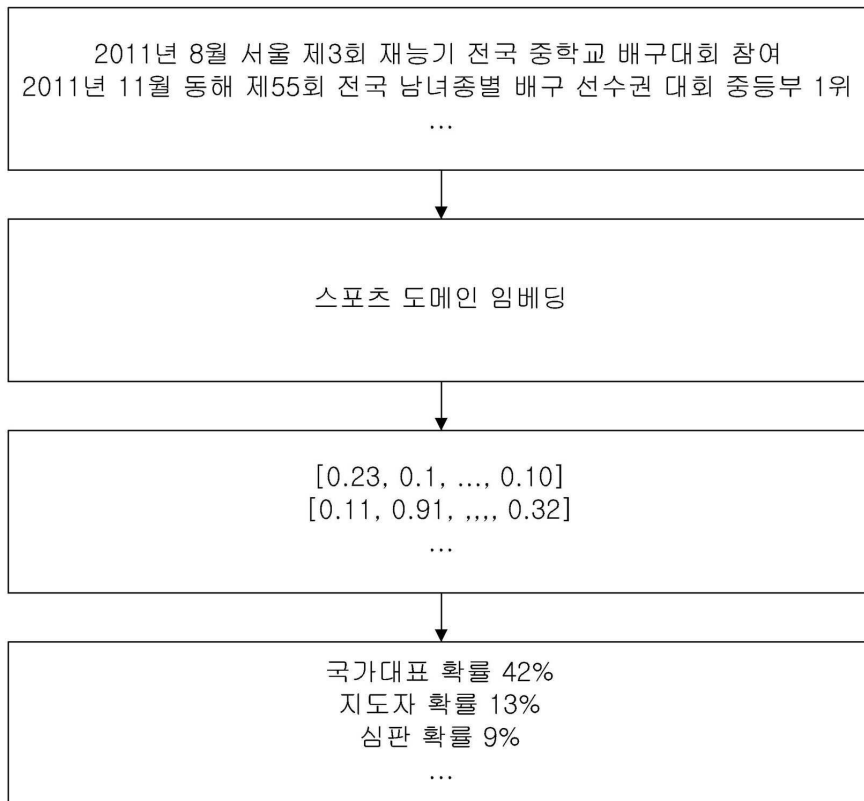


도면2

200



도면3



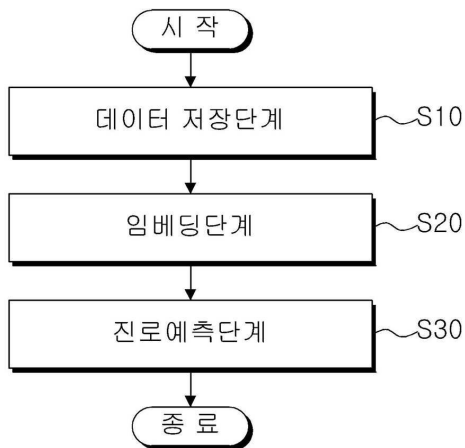
도면4

	1-best accuracy	Recall@2	Full accuracy
CNN	73.23	91.59	73.51
CNN+word2vec	73.6	89.29	73.42
CNN+fastText	74.06	89.67	73.67
LSTM	73.01	87.19	73.45
LSTM+word2vec	71.71	88.14	71.94
LSTM+fastText	69.4	84.4	69.6

도면5

	1-best accuracy	Recall@2	Full accuracy
CNN	78.21	96.65	77.37
CNN+word2vec	75.98	98.32	74.3
CNN+fastText	79.33	97.21	78.77
LSTM	70.42	95.77	69.72
LSTM+word2vec	62.44	92.96	63.38
LSTM+fastText	64.79	95.77	64.79

도면6



도면7

S20

