



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0062885
(43) 공개일자 2020년06월04일

(51) 국제특허분류(Int. Cl.)

G06F 40/20 (2020.01)

(52) CPC특허분류

G06F 40/211 (2020.01)

G06F 40/268 (2020.01)

(21) 출원번호 10-2018-0148817

(22) 출원일자 2018년11월27일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

서정연

서울특별시 서초구 신반포로16길 15-20, 103동 2702호 (반포동, 반포힐스테이트)

안휘진

서울특별시 종로구 창의문로9길 8(부암동)

(뒷면에 계속)

(74) 대리인

유미특허법인

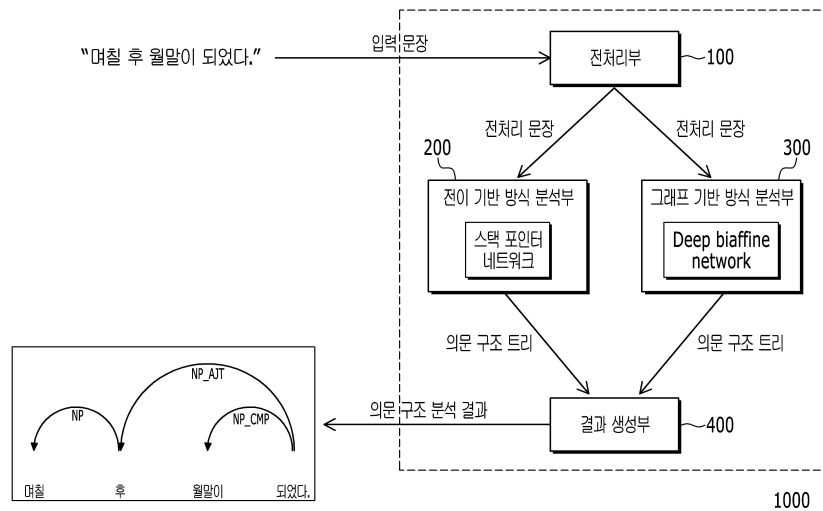
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 의존 구문 분석 장치 및 방법

(57) 요약

의존 구문 분석 장치로서, 입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성하는 전처리부, 스택 포인터 네트워크를 통해 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제1 의존 구문 트리를 생성하는 전이 기반 방식 분석부, Deep biaffine network를 통해, 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제2 의존 구문 트리를 생성하는 그래프 기반 방식 분석부, 그리고 상기 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 상기 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 상기 최종 의존 구문 트리를 이용하여 상기 입력 문장의 의존 구문을 분석하는 결과 생성부를 포함한다.

대표도



(72) 발명자

서민영

충청북도 충주시 동량면 지등로 232-2

박찬민

서울특별시 광진구 능동로4길 78, 102동 101호 (자양동, 우방아파트)

이재하

서울특별시 양천구 목동서로 38, 128동 401호 (목동, 목동아파트)

손정연

서울특별시 마포구 광성로6안길 23, 204호

김주애

서울특별시 은평구 갈현로47길 13, 402호 (갈현동)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711073577

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 SW전문인력역량강화

연구과제명 SW중심대학(서강대)

기 여 율 1/1

주관기관 서강대학교 산학협력단

연구기간 2018-01-01~2018-12-31

명세서

청구범위

청구항 1

의존 구문 분석 장치로서,

입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성하는 전처리부,

스택 포인터 네트워크를 통해 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제1 의존 구문 트리를 생성하는 전이 기반 방식 분석부,

Deep biaffine network를 통해, 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제2 의존 구문 트리를 생성하는 그래프 기반 방식 분석부, 그리고

상기 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 상기 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 상기 최종 의존 구문 트리를 이용하여 상기 입력 문장의 의존 구문을 분석하는 결과 생성부

를 포함하는 의존 구문 분석 장치.

청구항 2

제1항에서,

상기 전처리부는

상기 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수를 결정하고,

상기 형태소들의 수가 상기 미리 설정된 수와 동일한 제1 어절의 경우, 상기 제1 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하고,

상기 형태소들의 수가 상기 미리 설정된 수보다 작은 제2 어절의 경우, 상기 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 상기 제2 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하고,

상기 형태소들의 수가 상기 미리 설정된 수보다 큰 제3 어절의 경우, 상기 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 상기 제3 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 의존 구문 분석 장치.

청구항 3

제1항에서,

상기 전이 기반 방식 분석부는

상기 스택 포인터 네트워크의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 스택 포인터 네트워크의 디코더를 통해 상기 은닉 표상을 이용하여 상기 제1 의존 구문 트리를 생성하는 의존 구문 분석 장치.

청구항 4

제1항에서,

상기 그래프 기반 방식 분석부는

상기 Deep biaffine network의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 은닉 표상에 상기 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성하고, 상기 Deep biaffine network의 디코더를 통해 상기 전역 은닉 표상을 이용하여 상기 전

처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하고, 상기 의존소 표상 및 상기 지배소 표상 사이의 관계를 계산하여 상기 제2 의존 구문 트리를 생성하는 의존 구문 분석 장치.

청구항 5

제1항에서,

상기 결과 생성부는

상기 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하고, 상기 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하고, 상기 제1 가중치 의존소 확률 분포 및 상기 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 상기 최종 의존 구문 트리를 생성하는 의존 구문 분석 장치.

청구항 6

의존 구문 분석 장치가 입력 문장의 의존 구문을 분석하는 방법으로서,

입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성하는 단계,

스택 포인터 네트워크를 통해 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제1 의존 구문 트리를 생성하는 단계,

Deep biaffine network를 통해, 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제2 의존 구문 트리를 생성하는 단계, 그리고

상기 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 상기 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 상기 최종 의존 구문 트리를 이용하여 상기 입력 문장의 의존 구문을 분석하는 단계

를 포함하는 의존 구문 분석 방법.

청구항 7

제6항에서,

상기 전처리 문장을 생성하는 단계는

상기 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수를 결정하는 단계,

상기 형태소들의 수가 상기 미리 설정된 수와 동일한 제1 어절의 경우, 상기 제1 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계,

상기 형태소들의 수가 상기 미리 설정된 수보다 작은 제2 어절의 경우, 상기 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 상기 제2 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계, 그리고

상기 형태소들의 수가 상기 미리 설정된 수보다 큰 제3 어절의 경우, 상기 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 상기 제3 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계

를 포함하는 의존 구문 분석 방법.

청구항 8

제6항에서,

상기 제1 의존 구문 트리를 생성하는 단계는

상기 스택 포인터 네트워크의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하는 단계, 그리고

상기 스택 포인터 네트워크의 디코더를 통해 상기 은닉 표상을 이용하여 상기 제1 의존 구문 트리를 생성하는 단계

를 포함하는 의존 구문 분석 방법.

청구항 9

제6항에서,

상기 제2 의존 구문 트리를 생성하는 단계는

상기 Deep biaffine network의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 은닉 표상에 상기 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성하는 단계, 그리고

상기 Deep biaffine network의 디코더를 통해 상기 전역 은닉 표상을 이용하여 상기 전처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하고, 상기 의존소 표상 및 상기 지배소 표상 사이의 관계를 계산하여 상기 제2 의존 구문 트리를 생성하는 단계

를 포함하는 의존 구문 분석 방법.

청구항 10

제6항에서,

상기 최종 의존 구문 트리를 생성하는 단계는

상기 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하는 단계,

상기 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하는 단계, 그리고

상기 제1 가중치 의존소 확률 분포 및 상기 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 상기 최종 의존 구문 트리를 생성하는 단계

를 포함하는 의존 구문 분석 방법.

발명의 설명

기술 분야

[0001] 본 발명은 의존 구문을 분석하는 기술에 관한 것이다.

배경 기술

[0002] 의존 구문 분석은 문장 성분간의 관계를 파악함으로써 문장의 구조를 이해하는 작업으로, 자연어 이해 영역의 대표적인 과제이다. 한편, 의존 구문을 분석하기 위한 기존의 방법은 머신 러닝 기반의 의존 구문 분석 방법과 딥 러닝을 활용한 의존 구문 분석 방법이 있다.

[0003] 머신 러닝 기반의 의존 구문 분석 방법은 부사의 통계 정보를 이용한 것으로, 부사에 담긴 통계 정보를 추출하여 전통적인 그래프 기반 방식으로 의존 구문 분석을 진행하며, 그래프 기반 의존 구문 분석 방식에서 추출된 부사 통계 자질을 활용하여 성능향상을 얻을 수 있다. 또한, 딥 러닝을 활용한 의존 구문 분석 방법은 포인터 네트워크를 이용한 것으로, 학습시 어절의 중심어에 대한 예측 정보와 레이블 예측 정보를 모두 활용하는 멀티 태스크 방식으로 학습하는 특징이 있다.

[0004] 그러나, 위와 같은 종래의 의존 구문 분석 방법들은 아래와 같은 한계를 가진다. 구체적으로, 머신 러닝 기반의 의존 구문 분석 방법은 핸드-크래프트(hand-crafted) 자질을 사용하는바, 다른 도메인의 문장에 대해 의존 구문 분석을 수행 하는 경우 또는 다른 방식의 의존 구문 분석 방법을 수행하는 경우 부사의 통계적인 정보가 효과적이지 않을 수 있다. 또한, 그래프 기반 의존 구문 분석 방식을 사용하므로 의존 구문 서브 트리에 대한 정보를

사용하지 못한다는 한계가 있다. 딥 러닝을 활용한 의존 구문 분석 방법의 경우에도 문장의 위계 정보(hierarchical information)을 활용하지 못하고, 문장의 의존 관계에 따라 생성되는 위계 정보를 충분히 반영하지 못하며, 이전 어절에 대한 예측 결과를 충분히 활용하지 못한다는 한계가 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 해결하고자 하는 과제는 스택 포인터 네트워크와 Deep biaffine network를 이용하여 전이 기반 방식 및 그래프 기반 방식의 의존 구문 분석 방법을 모두 수행하는 기술을 제공하는 것이다.

과제의 해결 수단

[0006] 본 발명의 일 실시예에 따른 의존 구문 분석 장치는 입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성하는 전처리부, 스택 포인터 네트워크를 통해 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제1 의존 구문 트리를 생성하는 전이 기반 방식 분석부, Deep biaffine network를 통해, 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제2 의존 구문 트리를 생성하는 그래프 기반 방식 분석부, 그리고 상기 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 상기 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 상기 최종 의존 구문 트리를 이용하여 상기 입력 문장의 의존 구문을 분석하는 결과 생성부를 포함한다.

[0007] 상기 전처리부는 상기 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수를 결정하고, 상기 형태소들의 수가 상기 미리 설정된 수와 동일한 제1 어절의 경우, 상기 제1 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하고, 상기 형태소들의 수가 상기 미리 설정된 수보다 작은 제2 어절의 경우, 상기 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 상기 제2 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하고, 상기 형태소들의 수가 상기 미리 설정된 수보다 큰 제3 어절의 경우, 상기 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 상기 제3 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.

[0008] 상기 전이 기반 방식 분석부는 상기 스택 포인터 네트워크의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 스택 포인터 네트워크의 디코더를 통해 상기 은닉 표상을 이용하여 상기 제1 의존 구문 트리를 생성한다.

[0009] 상기 그래프 기반 방식 분석부는 상기 Deep biaffine network의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 은닉 표상에 상기 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성하고, 상기 Deep biaffine network의 디코더를 통해 상기 전역 은닉 표상을 이용하여 상기 전처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하고, 상기 의존소 표상 및 상기 지배소 표상 사이의 관계를 계산하여 상기 제2 의존 구문 트리를 생성한다.

[0010] 상기 결과 생성부는 상기 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하고, 상기 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하고, 상기 제1 가중치 의존소 확률 분포 및 상기 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 상기 최종 의존 구문 트리를 생성한다.

[0011] 본 발명의 일 실시예에 따른 의존 구문 분석 장치가 입력 문장의 의존 구문을 분석하는 방법은 입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성하는 단계, 스택 포인터 네트워크를 통해 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제1 의존 구문 트리를 생성하는 단계, Deep biaffine network를 통해, 상기 각 어절들의 의존 관계를 예측하여, 상기 전처리 문장에 대한 제2 의존 구문 트리를 생성하는 단계, 그리고 상기 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 상기 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 상기 최종 의존 구문 트리를 이용하여 상기 입력 문장의 의존 구문을 분석하는 단계를 포함한다.

[0012] 상기 전처리 문장을 생성하는 단계는 상기 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수를 결정하는 단계, 상기 형태소들의 수가 상기 미리 설정된 수와 동일한 제1 어절의 경우, 상기 제1 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계, 상기 형태소들의 수가

상기 미리 설정된 수보다 작은 제2 어절의 경우, 상기 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 상기 제2 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계, 그리고 상기 형태소들의 수가 상기 미리 설정된 수보다 큰 제3 어절의 경우, 상기 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 상기 제3 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분하는 단계를 포함한다.

[0013] 상기 제1 의존 구문 트리를 생성하는 단계는 상기 스택 포인터 네트워크의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하는 단계, 그리고 상기 스택 포인터 네트워크의 디코더를 통해 상기 은닉 표상을 이용하여 상기 제1 의존 구문 트리를 생성하는 단계를 포함한다.

[0014] 상기 제2 의존 구문 트리를 생성하는 단계는 상기 Deep biaffine network의 인코더를 통해 상기 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 상기 각 어절들에 대한 은닉 표상을 생성하고, 상기 은닉 표상에 상기 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성하는 단계, 그리고 상기 Deep biaffine network의 디코더를 통해 상기 전역 은닉 표상을 이용하여 상기 전처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하고, 상기 의존소 표상 및 상기 지배소 표상 사이의 관계를 계산하여 상기 제2 의존 구문 트리를 생성하는 단계를 포함한다.

[0015] 상기 최종 의존 구문 트리를 생성하는 단계는 상기 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하는 단계, 상기 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하는 단계, 그리고 상기 제1 가중치 의존소 확률 분포 및 상기 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 상기 최종 의존 구문 트리를 생성하는 단계를 포함한다.

발명의 효과

[0016] 본 발명에 따르면, 전이 기반 방식 및 그래프 기반 방식을 모두 사용하여 의존 구문 분석을 수행하는바 양 방식의 장점을 모두 사용할 수 있으며, 구체적으로, 의존 구문 서브 트리 정보를 자질로 사용할 수 있을 뿐만 아니라, 문장의 전역적인 정보 또한 사용할 수 있다.

도면의 간단한 설명

[0017] 도 1은 한 실시예에 따른 의존 구문 분석 장치를 설명하는 도면이다.
 도 2 내지 도 4는 전처리부가 전처리 문장을 생성하는 방법을 설명하는 도면이다.
 도 5는 전이 기반 방식 분석부가 스택 포인터 네트워크를 이용하여 전처리 문장의 의존 구문 분석을 수행하는 방법을 도시한 도면이다.
 도 6은 그래프 기반 방식 분석부가 Deep biaffine network를 이용하여 전처리 문장의 의존 구문 분석을 수행하는 방법을 도시한 도면이다.
 도 7은 의존 구문 분석 장치가 입력 문장의 의존 구문 분석을 수행하는 방법을 도시한 도면이다.

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.

[0020] 도 1은 한 실시예에 따른 의존 구문 분석 장치를 설명하는 도면이다.

[0021] 도 1을 참고하면, 의존 구문 분석 장치(1000)는 전처리부(100), 전이 기반 방식 분석부(200), 그래프 기반 방식 분석부(300) 및 결과 생성부(400)를 포함한다.

- [0022] 전처리부(100)는 입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성한다.
- [0023] 도 2 내지 도 4는 전처리부가 전처리 문장을 생성하는 방법을 설명하는 도면이다.
- [0024] 전처리부(100)는 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수 및 품사 정보를 결정하고, 형태소들의 수와 미리 설정된 수를 비교한다. 한편, 전처리부(100)는 형태소 분석이 완료된 문장이 입력된 경우 형태소 분석을 생략할 수 있다.
- [0025] 전처리부(100)는 구성하는 형태소들의 수가 미리 설정된 수와 동일한 제1 어절의 경우, 미리 설정된 수만큼의 형태소들 및 형태소들의 품사 정보들로 제1 어절을 구분한다.
- [0026] 예를 들면, 미리 설정된 수가 4이고, 제1 어절을 구성하는 형태소들이 4개인 경우, 전처리부(100)는 미리 설정된 수와 제1 어절을 구성하는 형태소들의 수가 동일하므로, 제1 어절에 대해 제1 어절을 구성하는 형태소들 및 각 형태소들의 품사 정보들로 구분하여 전처리 문장을 생성할 수 있다.
- [0027] 그러나, 전처리부(100)는 구성하는 형태소들의 수가 미리 설정된 수보다 작은 제2 어절의 경우, 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 제2 어절을 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.
- [0028] 예를 들면, 도 2를 참고하면, 제2 어절이 "것이든"이고 미리 설정된 수가 4인 경우, 제2 어절을 구성하는 형태소들이 "것", "이", 및 "든" 3개이므로, 전처리부(100)는 제2 어절을 구성하는 형태소들 중 가운데 위치한 형태소 "이"를 중복 사용하여 4개의 형태소가 되도록 형태소들을 재설정하고, 제2 어절에 대해 재설정된 형태소들 및 각 형태소들의 품사 정보들로 구분하여 전처리 문장을 생성할 수 있다.
- [0029] 한편, 전처리부(100)는 제2 어절을 구성하는 형태소들을 사용하지 않고 적어도 하나 이상의 더미(dummy) 형태소를 사용하여 제2 어절이 미리 설정된 형태소를 가지도록 형태소들을 재설정하고, 제2 어절에 대해 재설정된 형태소들 및 각 형태소들의 품사 정보들로 구분하여 전처리 문장을 생성할 수 있다.
- [0030] 예를 들면, 도 3을 참고하면, 제2 어절이 "열린"이고 미리 설정된 수가 4인 경우, 제2 어절을 구성하는 형태소들이 "열리" 및 "ㄴ" 2개이므로, 전처리부(100)는 두 개의 더미 형태소를 사용하여 제2 어절이 4개의 형태소들 가지도록 형태소들을 재설정하고, 제2 어절에 대해 재설정된 형태소들 및 각 형태소들의 품사 정보들로 구분하여 전처리 문장을 생성할 수 있다. 이 경우, 더미 형태소는 품사 정보가 없으므로 품사 정보는 생략될 수 있다.
- [0031] 또한, 전처리부(100)는 형태소들의 수가 미리 설정된 수보다 큰 제3 어절의 경우, 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 제3 어절을 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.
- [0032] 예를 들면, 도 4를 참고하면, 제3 어절이 "비생산적인"이고 미리 설정된 수가 4인 경우, 제3 어절을 구성하는 형태소들이 "비", "생산", "적", "이" 및 "ㄴ" 5개이므로, 전처리부(100)는 제3 어절을 구성하는 형태소들 중 가운데에 위치한 형태소 "적"을 제외한 나머지 형태소 4개를 추출하여 4개의 형태소가 되도록 형태소들을 재설정하고, 제3 어절에 대해 재설정된 형태소들 및 각 형태소들의 품사 정보들로 구분하여 전처리 문장을 생성할 수 있다.
- [0033] 한편, 도 2 내지 도 4에 도시된 품사 정보를 나타내는 정보는 자연어 처리 분야에서 품사 정보를 지칭하는 일반적인 의미로 쓰이는바 자세한 설명은 생략하며, 도 2 내지 도 4에서 도시된 형태소들 및 각 형태소들의 품사 정보들로 구분된 전처리 문장은 단지 예시이고, 형태소들 및 각 형태소들의 품사 정보들로 구분될 수 있으면 특정 실시예에 한정되지 않는다.
- [0034] 위에서 설명한 방법으로 인해, 만일 입력 문장이 "며칠 후 월말이 되었다."인 경우, 전처리부(100)는 입력 문장이 4개의 어절("며칠", "후", "월말이", "되었다.")로 구성되었음을 결정하고, 각 어절들을 4개의 형태소들 및 각 형태소들의 품사 정보들로 구분하여 4개의 어절들 각각에 대해 4개의 형태소들 및 품사 정보들로 구분된 전처리 문장을 생성할 수 있다.
- [0035] 이 경우, 미리 설정된 수가 4인 경우, 한국어 어절은 통상적으로 4개 이하의 형태소들로 구성되므로, 전처리부(100)에 의해 생성된 전처리 문장은 어절을 구성하는 모든 형태소들의 정보를 포함할 수 있어 어절이 의미하는 의미적 정보 및 복합적 정보를 모두 포함할 수 있다.
- [0036] 한편, 전처리부(100)가 문장을 어절 별로 분리하고, 각 어절에 포함된 형태소들 및 형태소들 각각에 대한 품사

정보를 결정하는 방법은 본 발명의 기술 분야에서 이미 공지된 기술로서 자세한 설명을 생략한다. 예를 들면, 전처리부(100)는 입력 문장에 대해 오픈 소스 기반의 형태소 분석기인 코모란(Komoran)을 사용하여 입력 문장의 형태소 분석 및 품사 정보 태깅을 진행할 수 있다.

- [0037] 전이 기반 방식 분석부(200)는 스택 포인터 네트워크(Stack Pointer Network)를 통해 각 어절들의 의존 관계를 예측하여, 전처리 문장에 포함된 각 어절들의 의존 관계를 나타내는 전처리 문장에 대한 제1 의존 구문 트리를 생성한다.
- [0038] 도 5는 전이 기반 방식 분석부가 스택 포인터 네트워크를 이용하여 전처리 문장의 의존 구문 분석을 수행하는 방법을 도시한 도면이다.
- [0039] 도 5를 참고하면, 스택 포인터 네트워크는 인코더(10), 디코더(20) 및 스택 포인터(30)로 구성되며, BiLSTM-CNN으로 구성되는 인코더(10)는 전처리 문장에 포함된 각 어절들의 형태소들 및 품사 정보들을 고려해서 각 어절들에 대한 은닉 표상을 생성한다.
- [0040] 이 경우, 전처리 문장에 포함된 각 어절들은 미리 설정된 수의 형태소들 및 품사 정보들로 구분되었으므로, 전이 기반 방식 분석부(200)는 인코더(10)를 통해 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 전처리 문장에 포함된 각 어절들에 대한 은닉 표상을 생성할 수 있다.
- [0041] 전이 기반 방식 분석부(200)는 디코더(20)를 통해 인코더(10)로부터 전달받은 은닉표상을 이용하여 제1 의존 구문 트리(50)를 생성한다.
- [0042] 구체적으로, LSTM로 구성되는 디코더(20)는 인코더(10)로부터 전처리 문장의 마지막 어절의 은닉 표상을 전달받는다. 마지막 어절의 은닉 표상은 전처리 문장에 대한 전반적인 정보를 포함하므로, 디코더(20)는 마지막 어절의 은닉 표상을 이용하여 전처리 문장에 포함된 각 어절들의 의존 관계를 예측하여 전처리 문장에 대한 제1 의존 구문 트리(50)를 생성한다.
- [0043] 예를 들면, 도 5에서, 입력 문장 "며칠 후 월말이 되었다."의 전처리 문장이 전이 기반 방식 분석부(200)에 입력된 경우, 전이 기반 방식 분석부(200)는 루트 심볼 "\$"를 시작으로 전처리 문장의 각 어절들을 순차적으로 인코더(10)에 입력하여 은닉 표상을 생성한다.
- [0044] 구체적으로, 전이 기반 방식 분석부(200)는 인코더(10)를 구성하는 BiLSTM-CNN인 s_1 내지 s_4 를 이용하여 전처리 문장의 각 어절들에 대한 은닉 표상을 생성한다.
- [0045] 즉, 전이 기반 방식 분석부(200)는 인코더의 s_1 에 루트 심볼을 입력하여 루트 심볼의 은닉 표상을 생성한다. 이후, 전이 기반 방식 분석부(200)는 루트 심볼의 은닉 표상을 s_2 에 입력하고 동시에 어절 "며칠"을 s_2 에 입력하여 어절 "며칠"의 정보들과 이전 정보인 루트 심볼의 정보까지 고려하여 어절 "며칠"의 은닉 표상을 생성한다. 이 경우, 어절 "며칠"의 정보들은 미리 설정된 수의 형태소들 및 품사 정보들을 의미한다. 전이 기반 방식 분석부(200)는 s_2 내지 s_5 를 통해 전처리 문장의 각 어절들에 대해 미리 설정된 수만큼의 형태소들 및 품사 정보들이 반영된 마지막 어절의 은닉 표상을 생성하고, 마지막 어절의 은닉 표상을 디코더(20)에 입력한다.
- [0046] 전이 기반 방식 분석부(200)는 디코더(20)를 구성하는 LSTM인 h_1 내지 h_5 , 및 h_1 내지 h_5 로부터 결과값을 수신하여 biaffine attention을 이용하여 전처리 문장의 의존 구문 트리를 생성하는 포인터 네트워크(30)를 포함한다.
- [0047] 구체적으로, 디코더(20)에 입력되는 어절의 순서는 스택(40)의 자료 구조에 의해 결정된다. 스택(40)의 하단에는 루트 심볼이 위치하고, 매 단계에서 스택(40)의 최상위에 위치한 어절의 은닉 표상이 디코더(20)에 입력된다. 이 경우, 어절에 sibling 및 grand parent가 있을 경우 이에 대한 정보를 은닉 표상에 반영한다. 이후 은닉 표상은 디코더(20)의 LSTM을 거쳐 포인터 네트워크(30)의 입력으로 사용되며, 포인터 네트워크(30)는 biaffine attention을 이용하여 각 어절에 대한 의존소를 예측한다. 예측된 어절은 스택(40)의 최상위에 위치하게 된다. 스택(40)이 모두 비어지는 시점에서 의존 구문 분석이 완료되게 된다. 위와 같은 과정을 통해 전이 기반 방식 분석부(200)는 전처리 문장 "며칠 후 월말이 되었다."에 대한 제1 의존 구문 트리(50)를 생성할 수 있다.
- [0048] 그래프 기반 방식 분석부(300)는 Deep biaffine network를 통해, 각 어절들의 의존 관계를 예측하여, 전처리 문장에 대한 제2 의존 구문 트리를 생성한다.
- [0049] 도 6은 그래프 기반 방식 분석부가 Deep biaffine network를 이용하여 전처리 문장의 의존 구문 분석을 수행하

는 방법을 도시한 도면이다.

- [0050] 도 6을 참고하면, Deep biaffine network는 인코더(60)와 디코더(70)로 구성되며, 임베딩 레이어(Embedding layer)(61) 및 BiLSTM(62)으로 구성되는 인코더(60)는 전처리 문장을 구성하는 각 어절들에 대한 은닉 표상을 생성하고, 은닉 표상에 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성한다. 또한, MLP(Multi Layer Perceptron)(71) 및 Biaffine Attention(72)으로 구성되는 디코더(70)는 전역 은닉 표상을 이용하여 전처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하고, 의존소 표상 및 지배소 표상 사이의 관계를 계산하여 제2 의존 구문 트리(80)를 생성한다.
- [0051] 구체적으로, 임베딩 레이어(61)는 전처리 문장에 포함된 각 어절들의 형태소들 및 품사 정보들을 반영하는 은닉 표상을 각 어절마다 생성한다. 이 경우, 전처리 문장에 포함된 각 어절들은 미리 설정된 수의 형태소들 및 품사 정보들로 구분되었으므로, 그래프 기반 방식 분석부(300)는 임베딩 레이어(61)를 통해 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 해당 어절들에 대한 은닉 표상을 생성할 수 있다.
- [0052] BiLSTM(62)는 임베딩 레이어(61)로부터 전처리 문장에 포함된 각 어절들의 은닉 표상들을 수신하고, 수신한 은닉 표상들에 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성한다.
- [0053] 이 경우, 전역 은닉 표상은 전처리 문장의 전역적인 정보를 반영하는바, 특정 어절의 전역 은닉 표상은 전처리 문장 내 다른 어절들의 정보까지 반영하는 것을 의미한다. 즉, BiLSTM(62)를 통해 특정 어절의 은닉 표상을 생성하는 경우 문장 내 모든 어절들의 은닉 표상이 입력으로 들어가는바, 특정 어절의 전역 은닉 표상은 전처리 문장 내 다른 어절들의 정보까지 반영한다.
- [0054] 이후, MLP(71)는 전역 은닉 표상을 이용하여 전처리 문장에 포함된 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산한다. 또한, Biaffine Attention(70)는 전처리 문장에 포함된 각 어절들의 의존소 표상과 지배소 표상 사이의 관계를 Biaffine Attention을 이용하여 계산하여 전처리 문장의 의존 관계를 예측하고, 이를 통해 제2 의존 구문 트리(80)를 생성한다.
- [0055] 구체적으로, Biaffine Attention(70)는 그래프 기반 방법으로 의존소 및 의존 관계명을 예측한다. 생성된 은닉 표상 r_t 에 서로 다른 가중치를 가진 MLP 계층을 적용하여 $h_t^{\text{arc-dep}}$ 및 $h_t^{\text{arc-head}}$ 를 생성하며, 이와 같은 연산을 입력 문장 전체에 적용하여 $H^{\text{arc-dep}}$ 및 $H^{\text{arc-head}}$ 를 생성한다. 이후, Biaffine Attention(70)는 하기 수학식 1 및 수학식 2를 이용하여 입력 단어 x_t 에 대한 의존소 확률 s_t^{arc} 및 의존 관계명 확률 s_t^{rel} 를 계산한다.

수학식 1

$$s_t^{\text{arc}} = H^{\text{arc-head}} U^{\text{arc}} h_t^{\text{arc-dep}} + H^{\text{arc-head}} u^{\text{arc}}$$

수학식 2

$$s_t^{\text{rel}} = H^{\text{rel-head}} U^{\text{rel}} h_t^{\text{rel-dep}} + H^{\text{rel-head}} u^{\text{rel}}$$

- [0058] 수학식 1 및 수학식 2에서, s_t^{arc} 는 의존소에 대한 unnormalized log probability distribution이고, s_t^{rel} 은 의존 관계명에 대한 unnormalized log probability distribution이고, $h_t^{\text{arc-dep}}$ 는 의존 관계 예측을 위한 t번째 어절의 지배소 표상 벡터이고, $H^{\text{arc-head}}$ 는 의존 관계 예측을 위한 어절의 의존소 표상 벡터를 쌓아 만든 매트릭스이고, U^{arc} 는 의존 관계 예측을 위한 파라미터 매트릭스이고, u^{arc} 는 의존 관계 예측을 위한 파라미터 벡터이고, $H^{\text{rel-head}}$ 는 의존 관계명 예측을 위한 어절의 의존소 표상 벡터를 쌓아 만든 매트릭스이고, $h_t^{\text{rel-dep}}$ 는 의존 관계명

예측을 위한 t 번 째 어절의 지배소 표상 벡터이고, u^{rel} 는 의존 관계명 예측을 위한 파라미터 벡터이다.

- [0059] 그래프 기반 방식 분석부(300)는 계산한 의존소 확률 s_t^{arc} 및 의존 관계명 확률 s_t^{rel} 중에서 가장 확률이 높은 의존소 및 의존 관계명을 선택하여 제2 의존 구문 트리(80)를 생성할 수 있다.
- [0060] 그래프 기반 방식 분석부(300)는 Deep biaffine network를 통해 전처리 문장을 구성하는 어절들의 모든 조합을 고려하여 어절들 간 최적의 조합을 생성하는바, 전처리 문장에 대해 그래프 기반 방식의 의존 구문 분석을 수행한다.
- [0061] 결과 생성부(400)는 제1 의존 구문 트리 및 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 전처리 문장에 대한 최종 의존 구문 트리를 생성하고, 최종 의존 구문 트리를 이용하여 입력 문장의 의존 구문을 분석한다.
- [0062] 구체적으로, 스택 포인터 네트워크를 통해 생성된 제1 의존 구문 트리와 Deep biaffine network를 통해 생성된 제2 의존 구문 트리는 중간 과정인 의존소 확률 분포 및 의존 관계명 확률 분포에서 차이가 있다. 따라서, 결과 생성부(400)는 양 네트워크의 중간 과정의 차이를 조합하여 최종 의존 구문 트리를 생성한다.
- [0063] 결과 생성부(400)는 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하고, 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하고, 제1 가중치 의존소 확률 분포 및 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 최종 의존 구문 트리를 생성한다.
- [0064] 예를 들면, 입력 문장 내 어절 단위로 설명하면, 스택 포인터 네트워크를 통해 입력 문장 "며칠 후 월말이 되었다." 내 어절 "월말이"의 의존소 확률 분포가 [0.3, 0.2, 0.1, 0.5]이고 가중치가 0.2이고, Deep biaffine network를 통한 동일한 어절의 의존소 확률 분포가 [0.2, 0.7, 0.1, 0.5]이고 가중치가 0.8인 경우, 결과 생성부(400)는 스택 포인터 네트워크의 가중치 의존소 확률 분포를 [0.06, 0.04, 0.02, 0.1]로 결정하고, Deep biaffine network의 가중치 의존소 확률 분포를 [0.16, 0.56, 0.08, 0.4]로 결정하고, 이들을 더한 의존소 확률 분포 [0.22, 0.6, 0.1, 0.5]를 최종 의존소 확률 분포로 결정할 수 있다. 또한, 결과 생성부(400)는 입력 문장 내 각 어절들에 대해 위와 같은 과정을 수행하며, 각 어절들 마다 결정된 최종 의존소 확률 분포를 갖는 최종 의존 구문 트리를 생성할 수 있다. 이 경우, 의존 구문 트리에 나타난 각 확률들은 해당 어절이 각 확률들에 대응하는 의존소에 해당할 확률을 나타낸다.
- [0065] 한편, 결과 생성부(400)는 복수의 스택 포인터 네트워크들 및 복수의 Deep biaffine network에 대해 의존 구문 트리를 각각 생성하여 최종 의존 구문 트리를 생성할 수도 있다. 예를 들면, 결과 생성부(400)는 5개의 스택 포인터 네트워크들 및 5개의 Deep biaffine network들로부터 10개의 의존 구문 트리를 각각 생성하고, 위와 같은 과정을 거쳐 하나의 의존소 확률 분포를 결정할 수도 있다.
- [0066] 결과 생성부(400)는 최종 의존소 확률 분포에서 가장 높은 확률 값을 결정하고, 결정된 확률 값에 대응하는 의존소를 해당 어절의 의존소로 결정할 수 있다. 결과 생성부(400)는 각 어절들에 대해 의존소를 결정하여 최종적으로 입력 문장의 의존 구문을 분석할 수 있다.
- [0067] 도 7은 의존 구문 분석 장치가 입력 문장의 의존 구문 분석을 수행하는 방법을 도시한 도면이다.
- [0068] 도 7에서, 도 1 내지 도 6에서 설명한 부분은 자세한 설명을 생략한다.
- [0069] 도 7을 참고하면, 의존 구문 분석 장치(1000)는 입력 문장을 구성하는 각 어절들을 미리 설정된 수의 형태소들 및 품사 정보들로 구분한 전처리 문장을 생성한다(S100).
- [0070] 구체적으로, 의존 구문 분석 장치(1000)는 입력 문장을 구성하는 복수의 어절들을 결정하고, 각 어절들을 구성하는 형태소들의 수를 결정한다.
- [0071] 만일 형태소들의 수가 미리 설정된 수와 동일한 제1 어절이 존재하는 경우, 의존 구문 분석 장치(1000)는 제1 어절을 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.
- [0072] 만일 형태소들의 수가 미리 설정된 수보다 작은 제2 어절이 존재하는 경우, 의존 구문 분석 장치(1000)는 제2 어절을 구성하는 형태소들 중 임의의 형태소를 중복 사용하여 제2 어절을 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.
- [0073] 한편, 만일 형태소들의 수가 상기 미리 설정된 수보다 큰 제3 어절이 존재하는 경우, 의존 구문 분석 장치

(1000)는 제3 어절을 구성하는 형태소들 중 임의의 복수의 형태소들을 하나의 형태소로 처리하여 제3 어절을 상기 미리 설정된 수만큼의 형태소들 및 품사 정보들로 구분한다.

[0074] 의존 구문 분석 장치(1000)는 스택 포인터 네트워크를 통해 각 어절들의 의존 관계를 예측하여, 전처리 문장에 대한 제1 의존 구문 트리를 생성한다(S200).

[0075] 구체적으로, 의존 구문 분석 장치(1000)는 스택 포인터 네트워크의 인코더를 통해 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 각 어절들에 대한 은닉 표상을 생성하고, 스택 포인터 네트워크의 디코더를 통해 은닉 표상을 이용하여 제1 의존 구문 트리를 생성한다.

[0076] 의존 구문 분석 장치(1000)는 Deep biaffine network를 통해, 각 어절들의 의존 관계를 예측하여, 전처리 문장에 대한 제2 의존 구문 트리를 생성한다(S300).

[0077] 구체적으로, 의존 구문 분석 장치(1000)는 Deep biaffine network의 인코더를 통해 미리 설정된 수의 형태소들 및 품사 정보들을 이용하여 각 어절들에 대한 은닉 표상을 생성한다.

[0078] 이후, 의존 구문 분석 장치(1000)는 은닉 표상에 전처리 문장의 전역적인 정보를 반영하여 전역 은닉 표상을 생성하고, Deep biaffine network의 디코더를 통해 전역 은닉 표상을 이용하여 전처리 문장을 구성하는 각 어절들이 의존소로 쓰였을 때의 표상인 의존소 표상 및 지배소로 쓰였을 때의 표상인 지배소 표상을 각각 계산하한다. 의존 구문 분석 장치(1000)는 의존소 표상 및 지배소 표상 사이의 관계를 계산하여 제2 의존 구문 트리를 생성한다.

[0079] 한편, 비록 도 5에서는 스택 포인터 네트워크를 통해 제1 의존 구문 트리를 생성하고, Deep biaffine network를 통해 제2 의존 구문 트리를 생성하는 것으로 도시되었으나, Deep biaffine network를 통해 제2 의존 구문 트리를 생성한 이후 스택 포인터 네트워크를 통해 제1 의존 구문 트리를 생성하거나, 제1 의존 구문 트리와 제2 의존 구문 트리를 동시에 생성하여도 무방하다.

[0080] 의존 구문 분석 장치(1000)는 제1 의존 구문 트리 및 상기 제2 의존 구문 트리에 서로 다른 가중치를 적용하여 전처리 문장에 대한 최종 의존 구문 트리를 생성한다(S400).

[0081] 구체적으로, 의존 구문 분석 장치(1000)는 제1 의존 구문 트리의 의존소 확률 분포에 제1 가중치를 적용하여 제1 가중치 의존소 확률 분포를 결정하고, 제2 의존 구문 트리의 의존소 확률 분포에 제2 가중치를 적용하여 제2 가중치 의존소 확률 분포를 결정하고, 제1 가중치 의존소 확률 분포 및 제2 가중치 의존소 확률 분포를 더한 최종 의존소 확률 분포를 의존소 확률 분포로 가지는 최종 의존 구문 트리를 생성한다.

[0082] 의존 구문 분석 장치(1000)는 최종 의존 구문 트리를 이용하여 입력 문장의 의존 구문을 분석한다(S500).

[0083] 구체적으로, 최종 의존 구문 트리의 의존소 확률 분포는 각 어절들이 특정 의존소일 확률을 의미하는바, 의존 구문 분석 장치(1000)는 각 어절들에 대하여 의존소 확률 분포 중 가장 높은 확률을 갖는 의존소를 특정 어절의 의존소로 결정하여 입력 문장의 의존 구문을 분석할 수 있다.

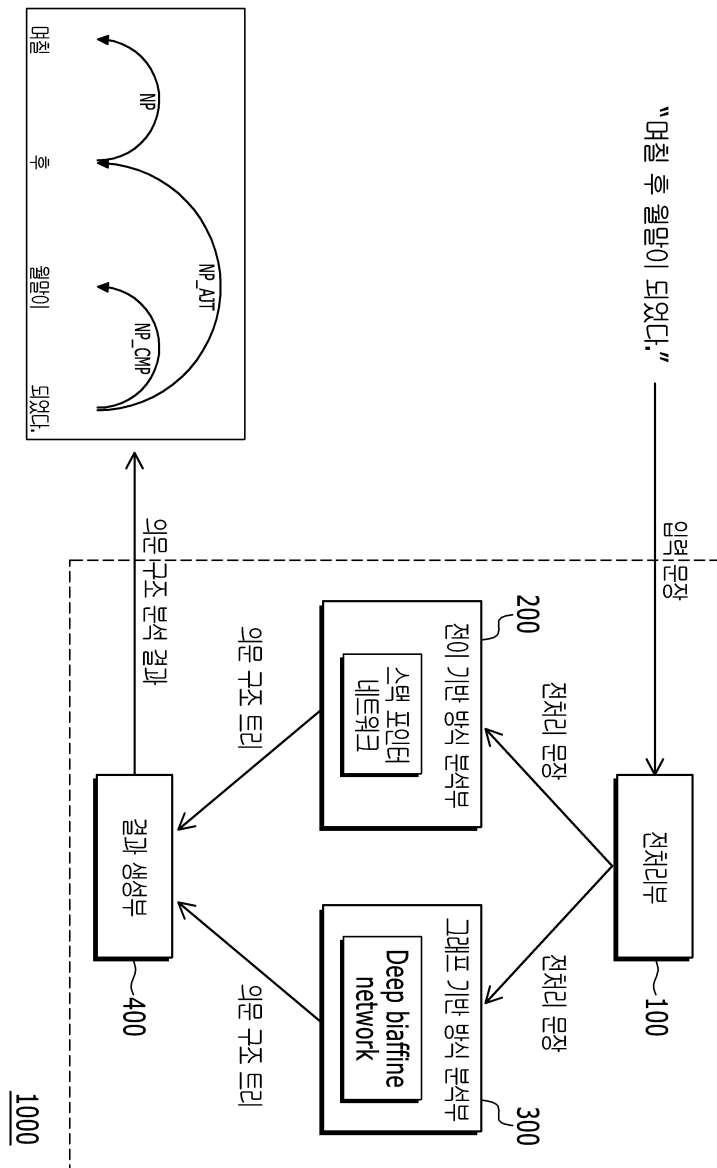
[0084] 본 발명에 따르면, 전이 기반 방식 및 그래프 기반 방식을 모두 사용하여 의존 구문 분석을 수행하는바 양 방식의 장점을 모두 사용할 수 있으며, 구체적으로, 의존 구문 서브 트리 정보를 자질로 사용할 수 있을 뿐만 아니라, 문장의 전역적인 정보 또한 사용할 수 있다.

[0085] 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.

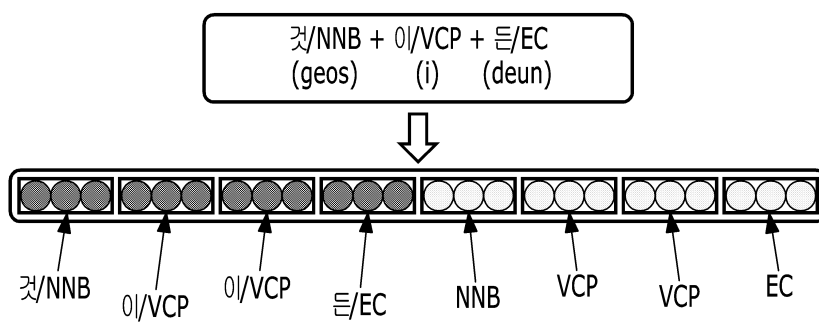
[0086] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

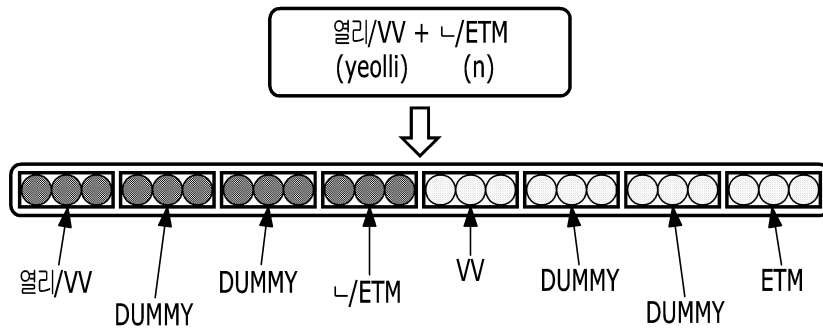
도면1



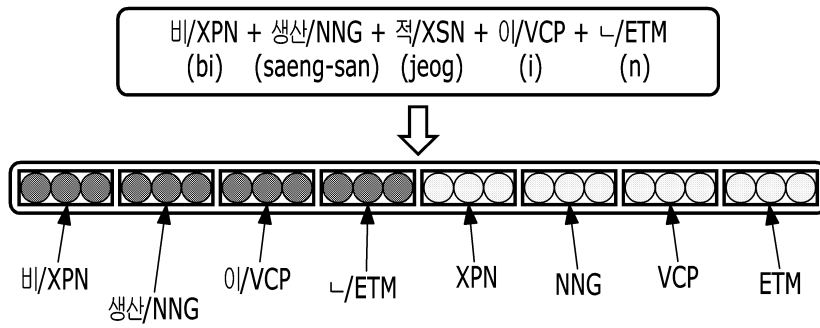
도면2



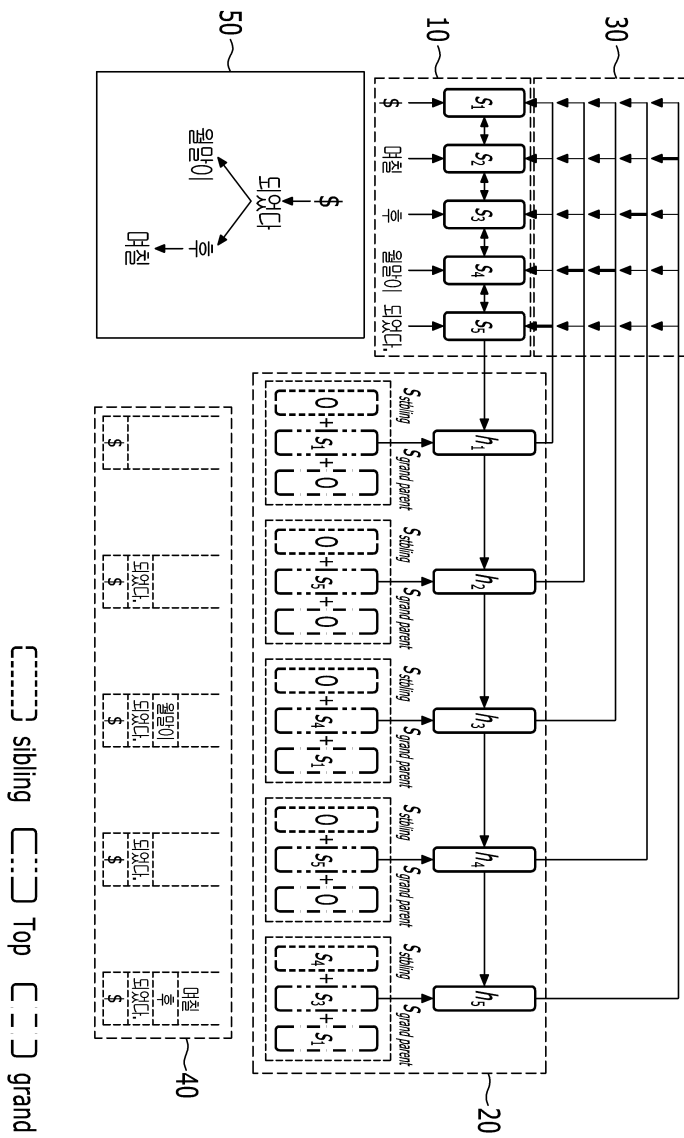
도면3



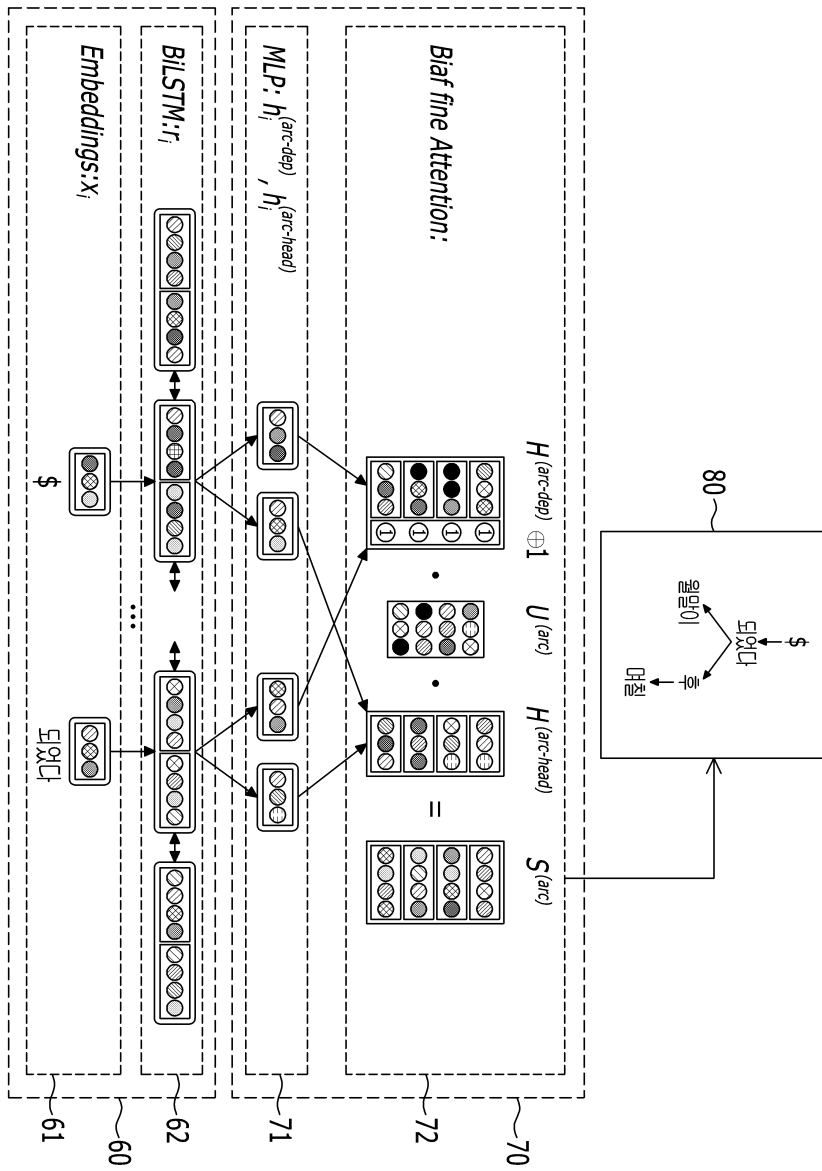
도면4



도면5



도면6



도면7

