

문형 정보를 이용한 규칙 기반 한국어 의존구문분석*)

김성태^o, 김민호, 김현아, 권혁철
부산대학교

kst8798@gmail.com, ailabminho@gmail.com, gusdk39@gmail.com, hckwon@pusan.ac.kr

Rules-based Korean Dependency Parsing using Sentence Pattern Informations.

Sung-Tae Kim^o, Min-Ho Kim, Hyuna Kim, Hyuk-Chul Kwon
Busan University

요 약

본 논문에서 제안하는 구문분석기는 품사 태거를 사용하지 않고 문장에서 나오는 모든 형태소 분석 후보에 의존관계를 부여하는 광범위 의존구문분석기이다. 중의성이 발생할 수 있는 문장에 대해 나오는 모든 후보 구문분석 트리를 출력하며 규칙을 통해 순위화를 진행한다. 또한 문형 정보 말뭉치의 적절한 사용을 위해 이전 연구의 한계점을 극복한 규칙과 알고리즘을 구현하고 문형 정보를 통해 후보 구문분석 트리의 순위화를 강화하였다. 뿐만 아니라 순위화가 어려운 [명사-관형사구] 자질에 대해 문형 정보를 사용하여 순위화를 강화하였다. 그 결과, 1순위의 구문 분석 트리에 대한 UAS(Unlabeled Attachment Score)가 0.52% 향상되었고, 후보트리에 대한 평균 정답 순위는 12.2%의 성능향상을 보였다.

주제어: 문형 정보, 규칙 기반, 의존 구문 분석, 중의성 표현

1. 서론

구문분석은 문장 성분 간의 관계를 구조적으로 파악하는 과정으로, 크게 구구조 구문분석과 의존구문분석으로 나뉜다. 한국어는 영어와는 달리, 조사와 어미가 다양하게 결합하여 어순이 고정적이지 않고 문장 구성 성분의 생략이 빈번해 의존구문분석이 더 적합하다[1].

의존구문분석은 결정적 의존구문분석과 비결정적 의존구문분석으로 나뉜다. 결정적 의존구문분석은 전이 기반 구문분석이라고도 부르며[2], 근거리 의존 관계를 찾는 데 장점을 지니고 상대적으로 적은 문맥 자질로도 분석이 가능한 지역적 모델이다. 하지만 지역적으로 선택한 지배-의존관계 분석의 오류전파현상 때문에 긴 문장을 분석할 경우 정확도가 떨어지는 단점이 있다. 비결정적 의존구문분석은 만들 수 있는 모든 의존트리 중 가장 높은 점수의 의존트리를 찾는 그래프 기반 구문분석이다[3]. 전역적 탐색을 하므로 긴 문장에서도 오류 전파가 없으며 상대적으로 높은 정확도를 제공하지만 속도 측면에서 느린 단점이 있다[4].

전이 기반으로 구문분석하는 기계학습[13,14]과 딥러닝 구문분석[5,15,16]의 경우 말뭉치에 의존적이고 확률적으로 접근하기 때문에 자료 부족 문제(Data Sparsity Problem)에 의한 구문분석오류를 수정하기가 어렵다. 반면, 규칙기반의 구문분석은 언어학적인 지식

을 기반으로 분석하므로 분석에 많은 시간이 투자되지만 말뭉치에 의존적이지 않으며 자료 부족 문제(Data Sparsity Problem)에 대해 강건하고 오류분석이 가능하다. 본 논문에서 제안하는 구문분석 방법은 형태소 단위 규칙 기반의 비결정적 의존구문분석으로, 언어학적 지식기반에 근거해 점수를 주고 차트 파싱 알고리즘[6]을 통해 구문분석 트리를 생성한다. 중의성을 표현하기 위해 가장 높은 점수의 구문분석 트리만 출력하는 것이 아닌, 점수의 순위화를 통해 후보 구문분석 트리를 전부 출력한다.

본 논문의 구성은 다음과 같다. 2장에서 규칙, 통계, 딥러닝 순으로 한국어 의존구문분석에 관한 연구에 대해 설명한다. 3장에서는 중의성이 발생할 수 있는 종류와 문형 정보를 사용하여 구문분석 트리의 순위화를 강화하는 방식을 설명한다. 4장에서는 3장의 문형 정보에 대한 실험과 결과를 분석하여 기술하고 5장은 본 논문의 요약과 향후 연구에 대해 알아본다.

2. 관련 연구

한국어 구문분석 연구는 한국어의 지배 후위원칙을 고려하여 다양한 방식으로 연구가 진행되고 있다. 지배소-의존소 관계자질을 규칙으로 정의한 규칙기반 의존구문분석 연구[7-10], 대용량 구문분석 말뭉치에서 지배소-의존소 관계자질을 통계적으로 추출하여 분석하는 통계/기계학습 기반 의존구문분석 연구[11-14], 기본 자질(raw features)을 신경망 모델(Neural Network Model)을 활용해 여러 단계를 걸쳐 복잡한 자질로 합성

*) 본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [2013-0-00179, (엑소브레인-3세부) 컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식 자원 구축]

해 내는 딥 러닝 의존구문분석 방법[5,15,16]이 있다.

차트 파싱 알고리즘은 정점(vertex)과 에지(edge)에 의한 유방향 그래프로 표현한 파싱 알고리즘이다[6]. 차트를 이용하여 비결정적 구문분석을 할 경우, 중의성을 가지는 문장에 대하여 모든 가능한 파스 트리를 생성하기 때문에 지배소-의존소 관계가 많아지고 복잡해진다.

[7]은 한국어 통사 규칙에 바탕을 두고 차트 파싱 알고리즘을 개선하여 불필요한 지배소-의존소 관계를 제거함으로써 구문분석의 효율을 증가시켰다.

또한 지배소-의존소 관계의 복잡도를 최소화하고자 구 묶음(chunking)을 통한 연구가 주로 수행되었다. 형태소 과생성을 유발하는 복합 동사구와 의존 명사 등을 포함하는 어절에 대해 구문 형태소 단위로 처리하여 모호성을 줄이는 방법[8], 구문분석 단계 전에 수행될 명사구와 동사구 단위화 방법과 지배노드와 의존노드 사이의 범위를 제한하는 방법[9], 한국어 사전의 정보와 각 구문들 사이의 문법 관계를 이용해 최장 묶음을 만드는 방법[10] 등 다양한 방법으로 연구되었다.

그 후 대량의 구문분석 말뭉치가 구축되면서 말뭉치로부터 유의미한 통계적 특성을 반영하여 어말-어두 공기 정보에 기초한 어휘 중의성 해소[11], 한국어의 지배가능 경로 문맥을 이용한 수식 거리 확률 모델[12], ME(Maximum Entropy)와 SVM(Support Vector Machine)을 이용한 결정적 의존구문분석[13], 주어나 목적어와 같은 의존 관계명을 부착과 의존구조를 동시에 분석하는 CRFs 모델[14] 등 기계학습 기반으로 다양한 연구가 진행되어 왔다.

최근 딥 러닝을 통한 학습 방법이 좋은 성능을 보이면서 딥 러닝을 이용한 전이 기반 구문분석이 수행되고 있다[5,15,16].

대부분의 선행연구[5,13-16]는 품사 태거를 사용하여 각 어절에서 1순위 형태소분석 후보를 이용해 구문분석을 시도하며, 탐욕적 방법(Greedy Method)을 통해 하나의 구문분석트리만 출력하는 전이 기반 구문분석이다. 이와 같은 방식의 구문분석은 의미적으로 발생할 수 있는 중의적 표현에 한계점이 있다.

3. 광범위 규칙 기반 의존구문분석

본 논문의 구문분석은 선행연구[5,13-16]와 달리 모든 형태소분석 후보를 토대로 차트를 이용해 파싱 알고리즘을 적용한 구문분석이다. 분석 속도 측면에서는 불리한 점을 보이지만, “숲속에서 나는 새를 보았다”와 같은 문장에서 발생하는 중의적인 표현을 모두 표현할 수 있는 강점을 지닌다. 후보 구문분석 트리는 규칙과 자질을 통하여 가중치를 조정하여 순위화한다.

3.1 중의적 문장에 대한 순위화

문장을 분석 하고자 할 때 예문 1과 같이 대명사가 참조하는 개체가 모호하여 구문분석 시 중의성이 발생할 수 있다.

예문 1 ‘모두가 좋다고 평가한 사람도 자기에게 해롭게 구는 사람이라고 하면 나쁘다고 생각하는 일이 많다.’

예문 1에 대한 개체들을 나열하면 다음과 같다.

개체 1 : ‘모두가 좋다고 평가한 사람’

개체 2 : ‘자기’

개체 3 : ‘해롭게 구는 사람’

‘사람’이라는 명사를 명확하게 나타내는 개체 1, 개체 3과는 달리, 개체 2는 이전에 언급한 개체를 대신해서 쓰는 대명사로 사용된다. 이전에 언급한 개체는 해당 문장에서 나온 개체 1일 수도 있고, 문장에서 언급되지 않는 개체 1과 개체 3과는 다른, 외부개체일 수도 있다. 따라서 개체2가 참조하는 개체에 따라 예문 1의 구문분석이 달라진다.

대명사로 사용된 개체가 해당 문장 내에서 언급한 개체를 참조한 경우, 즉, 예문 1에서 개체 2가 개체 1을 참조한 경우라면 해당 문장의 의미는 예문1-1처럼 해석가능하다.

예문 1-1 “모두가 좋다고 평가한 사람(A)도 자기(A)에게 해롭게 구는 사람(B)이라고 하면 나쁘다고 생각하는 일이 많다.”

개체 1과 개체 3은 서로 다른 개체이므로, 개체 1은 개체 3을 포함한 구 전체를 지배하는 형태의 분석이 자연스러운 구문분석이라고 할 수 있다. 서로 다른 두 개체가 [~은 ~이면 ~하다]같은 가정문법에 의해서 용언인 ‘나쁘다’에 함께 의존하는 것은 부자연스럽다. 따라서 [그림 1]과 같이 두 개체를 분리하여 지배-의존관계를 형성한 결과, (개체 1 - 생각하는)이 되고 (개체 3 - 나쁘다고)가 된다.

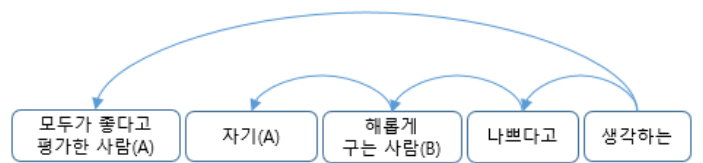


그림 1. 예문 1-1 해석에 대한 구문분석 트리

반면에 대명사로 사용된 개체가 해당 문장 외에서 언급된 외부개체 또는 문장에서 생략된 개체를 참조한 경우, 예문 1은 예문 1-2와 같이 해석가능하다. 외부개체 또는 생략된 개체를 C라고 가정하면, 해당 문장의 의미는 예문 1-2처럼 해석가능하다.

예문 1-2 “(C는) 모두가 좋다고 평가한 사람(A)도 자기(C)에게 해롭게 구는 사람(A)이라고 하면 나쁘다고 생각하는 일이 많다.”

보다 쉽게, ‘평가한 사람도’를 ‘평가한 사람일지도’로 바꿔 해석하면 더욱 명확해진다. 개체 1과 개체

3은 서로 같은 개체이기 때문에 [~은 ~이면 ~하다]와 같은 가정문법에 의해 구성되는 구조가 자연스러운 구문분석이다. 따라서 [그림 2]와 같이 두 개체를 하나의 지배-의존관계로 형성한 결과, (개체1 - 나쁘다고), (개체 3 - 나쁘다고)가 된다. 해당 중의성에 대한 순위화는 문장 내에서 언급한 개체를 참조한 예문 1-1에 가중치를 더 부여하였다.

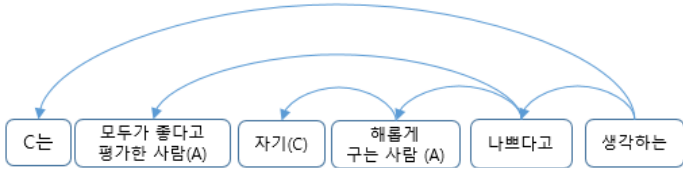


그림 2. 예문 1-2 해석에 대한 구문분석 트리

구문분석의 또 다른 어려움은 장소의 불확실성으로 인해 중의성이 발생한다는 점이다. 예문 2를 통해 이러한 불확실성을 확인할 수 있다.

예문 2 “미국에서 ‘미래주거단지 셀러브레이션 커뮤니티’를 기획해 논란을 일으키고 있는 예일대 로버트 등이 온다.”

예문 2에서 장소를 나타내는 단어인 {미국에서}가 어느 용언과 지배-의존관계를 형성하느냐에 따라 문장 전체의 의미와 구문분석이 달라진다.

{미국에서 - 일으키다}에 초점을 두고 해석하면, {로버트}라는 개체가 {논란}을 일으키고 있는 장소는 {미국}이지만 그 개체가 현재 {미국}에서 오는 것인지는 명확하지 않다. 반면에 {미국에서 - 온다}에 초점을 두면, {로버트}라는 개체가 {논란}을 일으키고 있는 장소는 명확하지 않으나 그 개체가 {미국}에서 온다는 사실은 명확하다. 예문 2에 대해 발생하는 중의성에 대한 순위화는 지배소-의존소 관계 거리가 가까운 구문에 가중치를 부여하여 순위화를 조정하였다.

3.2 문형 정보를 이용한 후보트리 순위화 강화

이전 연구[18]에서 용언 상세전자사전의 한국어 용언 29,492항목에 대한 문형정보를 토대로 구문분석 후보의 순위화를 연구하였다. 본 연구에서는 [18]에서 구현한 문형사전의 자료구조를 [그림 4]의 자료구조로 수정하여 의미를 구분한 용언의 수가 6,469개 더 추가되었다. 이전과는 다르게, 한 문형에 대해 복수 개의 선택제약정보를 가지더라도, 해시 테이블에 모두 표현하도록 설계하였다.

또한 ‘문형 정보를 이용한 구문분석 후보의 순위화’에 대한 한계점들을 다음과 같이 언급하였다.

- 1) 용언의 의존소 개수가 문형과 다른 경우
- 2) 문형의 논항과 다른 선택제약을 취하는 경우
- 3) 용언만으로는 알맞은 문형을 선택할 수 없는 경우

첫 번째 경우, “동생은 이미 밥을 맛있게 먹었다.”

와 같은 문장에서 수식부사와 ‘[형용사/기(계/도록)]’을 의존소의 개수로 인정하지 않는 규칙을 추가하여 가중치의 누락을 방지하였다. 두 번째 경우는 “누군가가 밥을 먹었다”와 같이 한국어 어휘 의미망[17]에 존재하지 않는 단어에 대해 본 연구의 구문분석에서 사용하고 있는 카테고리-단어사전과 선택제약정보를 대응시켜 해결하였다. 세 번째의 경우, “우리는 선생님을 믿는 사람을 믿는다.”와 같이 문형 정보의 중복에 대한 선택 문제를 서브트리를 생성하는 과정에서 규칙과 함께 해결하였다.

서브트리 생성과정에서 문형 정보를 통해 가중치를 부여하는 방법은 [그림 3]의 의사코드(pseudocode)를 따른다. [표 1]은 [그림 3]에서 나오는 용어의 설명이다.

```
function Weighting_by_SubcatFrame
{
    if  $g_{pred} \in SF_{pred}$  and  $dep == arg_x$  and  $d_{prev} == sp_x$ 
        for all  $sg_{GT} \in GT$  do
            for all  $arg \in SF_{arg}$  do
                if  $sg_{GT} == arg$  and  $sp_{sg} \in SF_{sp}$ 
                     $w_{GT} += w_{SF}$ 
                end if
            end for
        end for
    end if
    apply  $w_{GT}$  to weight of  $GT$ 
}
```

그림 3. 문형 정보를 이용한 가중치 부여 의사코드

표 1. [그림 3] 의사코드에서 사용되는 용어 설명

용어	설명
GT	지배소의 부분 문장구조
sg_{GT}	지배소의 부분 문장구조에 직접적으로 의존한 하위지배소
g_{pred}	지배소의 용언
SF_{pred}	문형 정보의 용언 집합
dep, d_{prev}	의존소의 단어(격조사) 의존소의 선택제약단어(의존소 앞에 있는 단어)
arg, arg_x	문형 정보에 대한 격조사 (arg_x 는 문형 정보 중 x에 대한 격조사)
SF_{arg}	문형 정보의 격조사 집합
SF_{sp}, sp_x	문형 정보의 선택제약단어 집합 (sp_x 는 문형 정보 중 x에 대한 선택제약단어)
sp_{sg}	하위지배소의 선택제약단어
w_{GT}	지배소의 부분 문장구조 가중치
w_{SF}	문형 정보에 의한 가중치 (선택제약단어-격조사의 개수를 가중치로 부여)

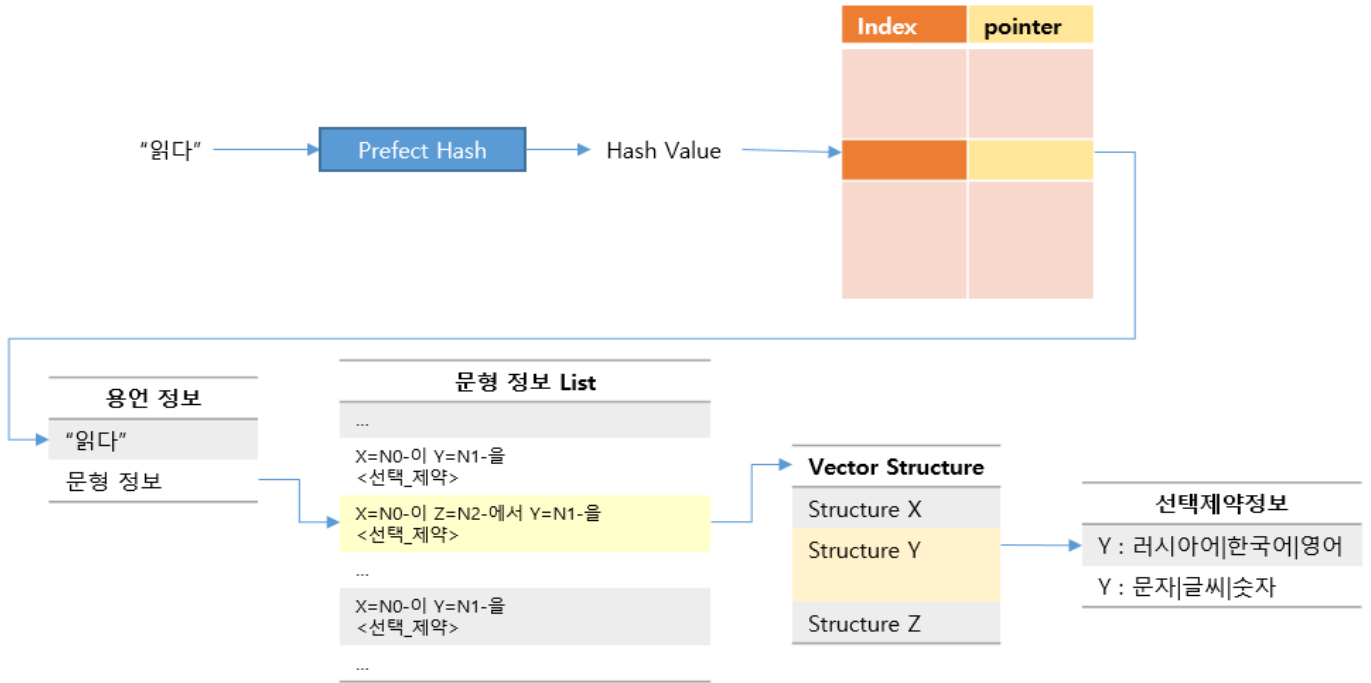


그림 4. 문형 사전의 자료 구조

예를 들어, ‘동생은 밥을 먹는다.’ 라는 문장과 이 문장에 따른 문형 정보 ‘(인간)-이 (음식)-을 먹다.’ 가 있다. 이때 의존소의 부분 문장구조는 [그림 5]이고 지배소의 부분 문장구조는 [그림 6]이라고 가정하자. g_{pred} 와 SF_{pred} 는 ‘먹다’로 일치하고 arg_x 는 ‘이’, dep 는 ‘은’이 된다. [18]의 논항일반화에 따라 ‘은’과 ‘이’는 동등한 격조사로 분류되므로 dep 와 arg_x 는 일치한 것으로 본다. 다음 sp_x 는 ‘인간’, d_{prev} 는 ‘동생’이고 ‘동생’은 ‘인간’의 하위어이므로 의존소에 대한 선택제약단어가 만족한다. 이후 GT 를 통해 나머지 문항 정보를 탐색한다. 문형 정보에서의 ‘(음식)-을’과 GT 에서의 ‘밥-을’은 ‘선택제약단어 - 격조사’ 정보가 일치하므로 해당 문장의 후보 구문분석 트리는 문형 정보를 반영한 가중치가 부여된다.

뿐만 아니라 [그림 3]의 의사코드를 응용하여 [명사-관형사구] 자질에 대해 가중치를 조정하였다. 예를 들어 ‘밥을 먹은 동생의 표정이 행복해 보인다.’ 라는 문장에서 [명사-관형사구] 자질에 대해 분석하면, (먹은-동생)과 (먹은-표정)인 두 개의 구문분석후보가 생성된다. 이전 연구에서는 관형사구가 가장 멀리 있는 명사에 의존하여 (먹은-표정)에 가중치를 부여하였다. 하지만,



그림 5. 의존소의 부분 문장구조 예



그림 6. 지배소의 부분 문장구조 예

만, 문형 정보 ‘(인간)-이 (음식)-을 먹다.’를 ‘(음식)-을 먹다+[관형형정성어미] (인간)-이’로 변형하여 [그림 7]과 같이 (먹은-동생)이 생성된 구문분석후보에 가중치를 부여하는 형식으로 순위화를 강화하였다.

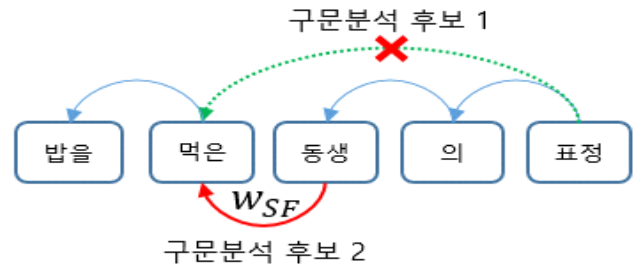


그림 7. [명사-관형사구] 자질 강화 가중치부여

4. 실험 및 평가

4.1 실험 환경

기존연구에서는 성능 평가를 위해 문어 말뭉치 433,707어절, 41,005문장의 규모로 구축된 ‘21세기 세종계획 구문분석 말뭉치’를 사용한다[19]. 하지만 세종구문분석 말뭉치는 구문분석의 기본단위가 어절로 정의되어 있어 형태소를 기본단위로 분석하는 본 연구의 구문분석에서 사용하기엔 부적합하다. 따라서 세종구문분석 말뭉치로부터 임의 추출한 600문장을 대상으로 본 논문의 구문분석에 맞춘 답안을 직접 작성하여 성능 평가에 사용하였다. 성능 평가에 사용된 문장들은 15어절 이상의 길이를 가진 복문형태의 문장들이다.

4.2 실험 결과

본 논문에서 소개한 광범위 의존구문분석은 하나의 구문분석 트리만을 선택하지 않고 복수 개의 후보트리

를 출력하므로, 선행연구의 결과들과 단순하게 비교하는 것은 무리가 있다.

적용한 문형 정보가 구문분석에 얼마나 영향을 미치는지 평가하기 위해, 모델을 3개로 나누었다. 모델 1은 이전 연구[18]의 모델이고 모델 2는 재구축한 문형 정보를 적용하였지만 [명사-관형사구] 자질에 대한 알고리즘은 사용하지 않은 모델, 모델 3은 [명사-관형사구] 자질에 대한 알고리즘까지 적용한 모델이다. 평가 기준은 각 모델의 구문분석 후보 트리 중 1순위의 UAS(Unlabeled Attachment Score)와 후보트리 내에서 정답이 나오는 평균 순위를 평가로 정했다.

제안한 모델 3은 기존 연구의 모델 1에 비해 UAS 0.52%의 성능향상을 이루었다. 또한, 후보트리에서 나오는 평균 정답 순위는 12.2%의 성능향상을 이루어 후보트리에 대한 순위화가 강화되었음을 보였다. 이 실험 결과에서 문형 정보를 사용할 경우, 구문분석의 정확도가 향상되고 특히, [명사-관형사구] 자질에 대한 구문분석의 순위화가 강화됨을 알 수 있다[표 2].

표 2. 성능 비교 분석

	UAS	평균 정답 순위
모델 1 [18]	91.22	18.8
모델 2	91.56	18.2
모델 3	91.74	16.5

5. 결론

본 논문의 구문분석은 문장에서 나오는 모든 형태소 분석 후보에 지배-의존관계를 부여하는 광범위 의존구문분석기이다. 중의성이 발생할 수 있는 문장에서 나올 수 있는 모든 후보 구문분석 트리를 출력하기 때문에 순위를 잘 조정하는 것이 중요하다. 이전연구[18]에서는 문형 정보를 이용하여 순위화를 강화하였으나 한계점으로 인해 많은 문제점이 발생하였다. 따라서 본 논문에서는 규칙과 [그림 3]의 알고리즘으로 문형 정보 적용의 문제점을 해결하였다. 또한, 순위화가 어려운 [명사-관형사구] 자질에 대한 분석을 문형 정보를 토대로 가중치를 부여해 순위화를 강화시켜 후보 트리 중 1순위의 UAS(Unlabeled Attachment Score)와 평균 정답 순위가 향상됨을 보였다.

향후 연구로는 문형 정보를 이용하여 도치된 구문에 대해서도 순위화를 강화하여 어순이 자유롭지 않은 한국어의 특성에 의한 중의성 문제를 완화할 계획이다.

참고문헌

[1] Michael A.Covington, "A dependency parser for variable-word-order languages," Research Report AI-1990-01, University of Georgia, 1990
 [2] J. Nivre, "An efficient algorithm for projective dependency parsing," In Proc. Of IWPT, pp.149-160, 2003.

[3] R. McDonald, K. Crammar, F. Pereira, "Online Large-margin Training of Dependency Parsers," Proc. Of ACL, pp.91-98, 2005.
 [4] Ryan McDonald, Joakim Nivre, "Analyzing and Integrating Dependency Parsers," Computational Linguistics, Volume 37, Number 1, 2011.
 [5] 이창기, 김준석, 김정희, "딥 러닝을 이용한 한국어 의존구문분석," 제26회 한글 및 한국어 정보처리 학술대회 논문집, HCLT 2014
 [6] Margaret King, "Natural Language Parsing," pp.58-87, Academic Press, 1983
 [7] 김호영, 최재혁, 이상조., "한국어 통사 규칙에 바탕을 둔 개선된 차트 파싱 알고리즘", 한국정보과학회 학술발표논문집, Vol.17 No.1, 1990
 [8] 황이규, 이현영, 이용석., "형태소 및 구문 모호성 축소를 위한 구문단위 형태소의 이용", 정보과학회논문지 소프트웨어 및 응용, 27(7), 784-793, 2000
 [9] 김미영, 강신재, 이종혁, "단위(Chunks)분석과 의존문법에 기반한 한국어 구문분석," 한국정보과학회 학술발표논문집, 27(1B), 327-329, 2000
 [10] 박상규, 정창민, 조준모, 이상조, "최장 묶음을 이용한 효과적인 한국어 구문분석기," 한국정보과학회 학술발표논문집, 22(1), 961-964, 1995
 [11] 이하규, "어말-어두 공기 정보를 이용한 한국어 어휘 중의성 해소," 정보과학회논문지(B), 24(1), 82-89, 1997
 [12] 우연문, 송영인, 박소영, 임해창, 정후중, "지배 가능 경로 문맥을 이용한 의존구문분석의 수식 거리 확률 모델," 한국정보과학회 언어공학연구회 학술발표 논문집, 40-47, 2006
 [13] 이용훈, 이종혁, "기계학습 기법을 이용한 한국어 구문분석," 한국정보과학회 학술발표논문집, 35(1C), 285-288, 2008
 [14] 최맹식, 정석원, 김학수, "CRFs를 이용한 의존구조 분석 및 의존 관계명 부착," 정보과학회논문지 소프트웨어 및 응용, 41(4), 302-308, 2014
 [15] 나승훈, 김강일, 김영길, "Stack LSTM을 이용한 전이 기반 한국어 의존 파싱," 한국정보과학회 학술발표논문집, 732-734, 2016
 [16] 홍승연, 나승훈, 신종훈, 김영길, "BERT와 ELMo 문맥화 단어 임베딩을 이용한 한국어 의존 파싱," 한국정보과학회 학술발표논문집, 491-493, 2019
 [17] 윤애선, 황순희, 이은령, 권혁철, "한국어 어휘의 미망 「KorLex 1.5」의 구축," 정보과학회논문지 소프트웨어 및 응용, Vol.31, No.1, pp.92-108, 2009.
 [18] 류재민, "광범위 한국어 구문분석을 위한 규칙 기반의 통사 중의성 해소 방안," 부산대학교 석사학위논문, 2018
 [19] 홍정하, 김주영, 강범모, "세종 구문분석 말뭉치의 구축과 통사 범주 및 기능의 통계적 분포," 민족문화연구 49집, pp.285-331, 2008.