

04

세분화된 한국어 형태소 규칙에 기반한 의존구문분석 모델 개발

소속 정보컴퓨터공학부

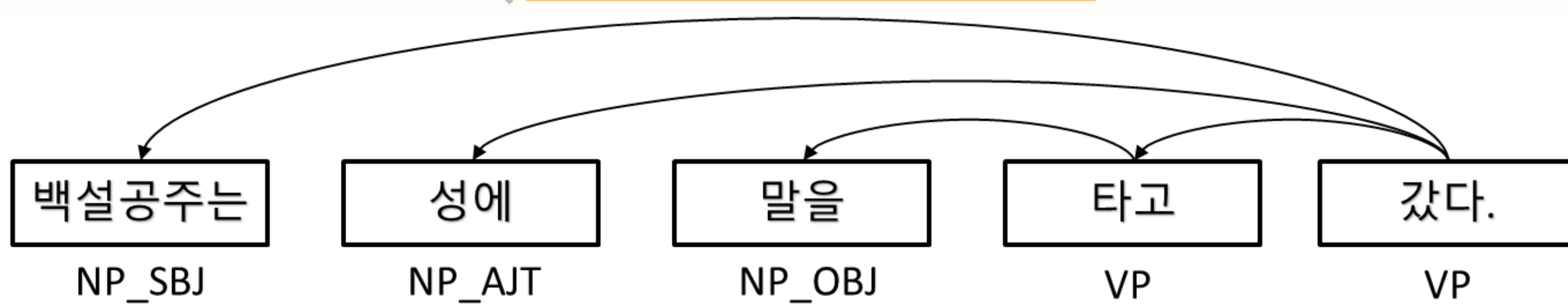
분과 A

팀명 노답삼형제

참여학생 김준기, 박기훈, 정대성

지도교수 권혁철

과제 개요 및 목표



과제 개요

- ✓ 의존구문분석이란 자연어 문장을 지배소, 피지배소 의존 관계로 분석하는 구문 분석 방법론으로 한국어에 적합한 자연어 처리 방식
- ✓ 시중에 koBERT, Klue/Roberta-Base(or Large), KoElectra=base 등의 의존구문분석 모델이 이미 있으나 개선할 여지가 남아 있음

과제 목표

- ✓ 부산대학교 AILAB의 KLTagger를 활용하여 성능이 뛰어난 의존구문분석 모델 개발
- ✓ 의존구문분석의 결과를 그래프로 시각화하여 출력해 사용자의 이해를 돕는 웹사이트 개발

과제 내용

구분	학습데이터	평가데이터	Metric	UAS	LAS
(1) Baseline	KLUE-DP	KLUE-DP	Macro	93.48	87.82
(2) 추가 형태소 임베딩+트랜스포머	KLUE-DP	KLUE-DP	Macro	93.73	88.03
(3) KLTagger	KLUE-DP	KLUE-DP	Macro	94.30	87.94
(4) KLUE 형태소 규칙	KLUE-DP	KLUE-DP	Macro	94.22	88.39
(5) KLTagger 규칙	KLUE-DP	KLUE-DP	Macro	94.88	88.48

동일한 학습데이터, 평가데이터와 평가기준을 사용해 비교

- ✓ 한국어 형태소 기반 KLTagger 사용시 UAS(Unlabeled Attachment Score) **0.57%** 향상(모델 2와 3 비교)
 - > 한국어 형태소 세부 분석은 **Parser의 성능**에 따라 성능이 크게 향상
- ✓ 각 형태소가 문장 내의 어절들과 의존 관계를 알려주는 Attention Score 방식을 활용한 규칙 제어 알고리즘을
 - 사용하자 KLUE 형태소 규칙은 UAS가 **0.74%** 향상(모델 1과 4 비교), KLTagger는 UAS가 **1.4%** 향상(모델 1과 5 비교)
 - > 형태소 분석에 **규칙 제어 알고리즘**이 효과적

결과

한국어 의존 구문 분석기

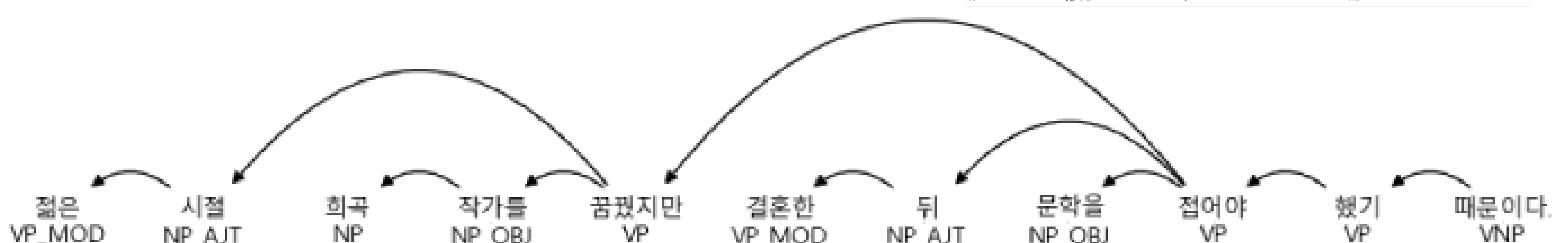
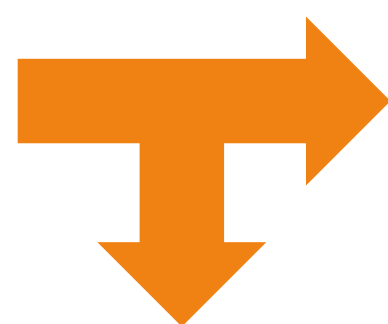
텍스트입력

젊은 시절 희곡 작가를 꿈꿨지만 결혼한 뒤 문학을 접어야 했기 때문이다.

분석을 원하는 문장 입력

입력문장 : 젊은 시절 희곡 작가를 꿈꿨지만 결혼한 뒤 문학을 접어야 했기 때문이다.

인덱스	단어형식	지배소인덱스	의존관계레이블
1	젊은	2	VP_MOD
2	시절	5	NP_AJT
3	희곡	4	NP
4	작가를	5	NP_OBJ
5	꿈꿨지만	9	VP
6	결혼한	7	VP_MOD
7	뒤	9	NP_AJT
8	문학을	9	NP_OBJ
9	접어야	10	VP
10	했기	11	VP
11	때문이다.	0	VNP



각 형태소의 의존 관계를 표와 그래프로 출력해 사용자의 이해를 돕는다