



순차적 구문 분석 방법을 반영한 포인터 네트워크 기반의 한국어 의존 구문 분석기

Korean Dependency Parsing Using Sequential Parsing Method Based on Pointer Network



한장훈O, 박영준, 정영훈, 이인권, 한정욱, 박서준, 김주애, 서정연
서강대학교, 자연어 처리 연구실

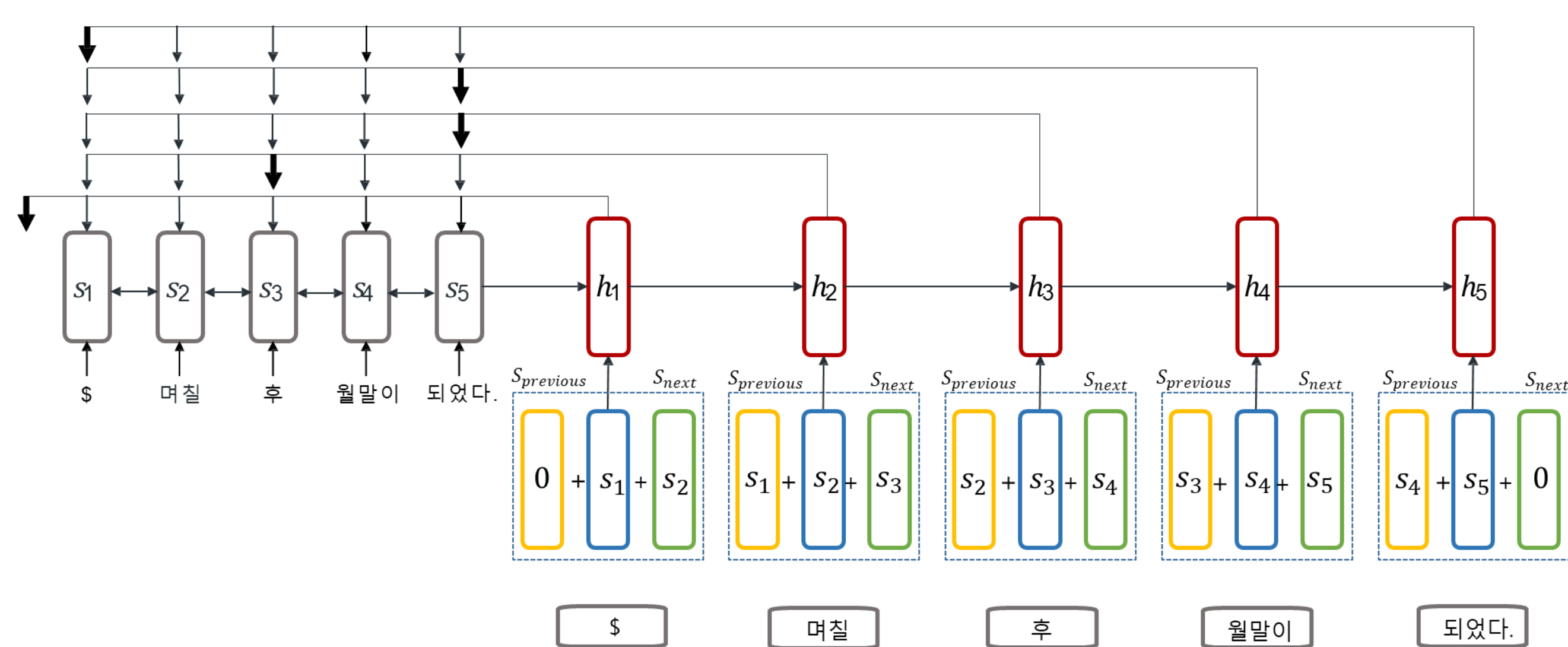
서론

- 의존 구문 분석은 문장 성분의 관계를 파악하여 문장의 구조를 이해하는 작업으로, 전통적인 의존 구문 분석 연구는 전이 기반 방식과 그래프 기반 방식으로 구분된다.
 - 전이 기반 방식 : Buffer와 Stack으로 전이 액션을 결정하는 지역적 탐색 방식
 - 그래프 기반 방식 : 가능한 모든 의존 구문 트리를 고려하는 방식
- 본 논문에서 제안하는 방법
 - 임베딩 : ELMo, BERT 임베딩
 - 모델
 - Left to Right Pointer Network
 - Right to Left Pointer Network
 - Deep Bi-Affine
 - 양상블 : 위의 세가지 모델을 Voting방법을 통해 양상블을 수행

사용 모델

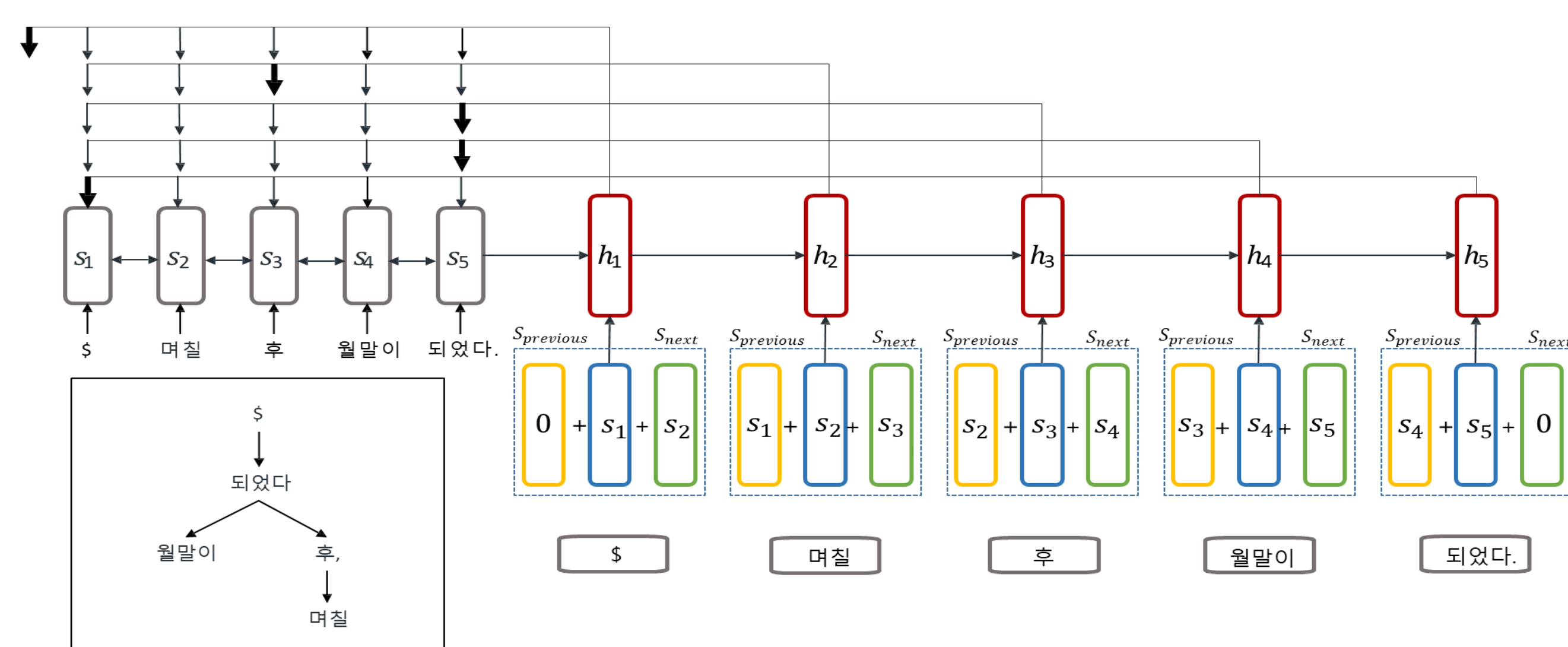
1. Left-to-Right 포인터 네트워크

- ✓ BiAffine과 Stack Pointer보다 좋은 성능
- ✓ 스택포인터 보다 빠른 속도 ($2N-1 \rightarrow N$)



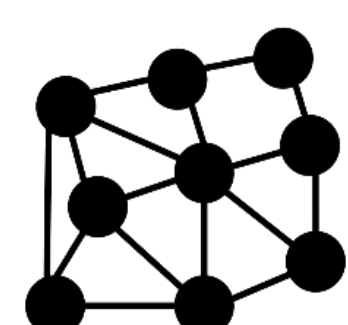
2. Right-to-Left 포인터 네트워크

- ✓ 지배소가 뒤에 위치하는 한국어의 특성을 반영
- ✓ 다른 모델들과 양상블 하였을 때 성능 향상에 기여



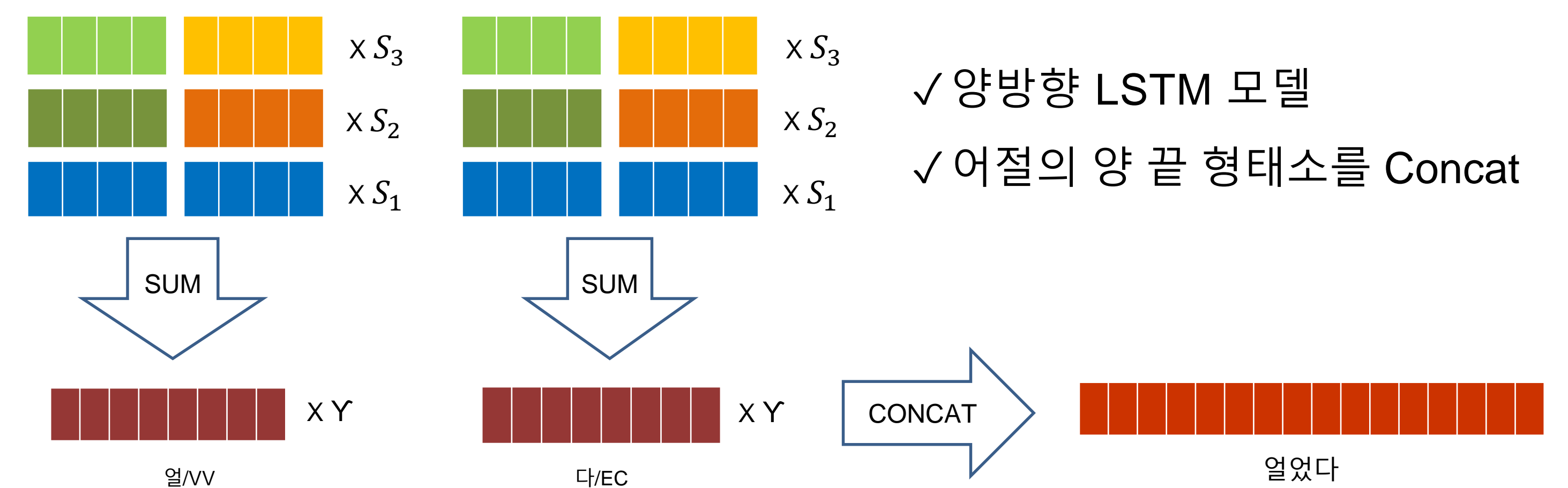
3. Deep Biaffine

- ✓ Embedding, BiLSTM, MLP, Biaffine Attention으로 구성
- ✓ 각 어절에 대한 의존소, 지배소 표상 생성
- ✓ Biaffine attention을 통해 의존 관계 및 관계명 예측



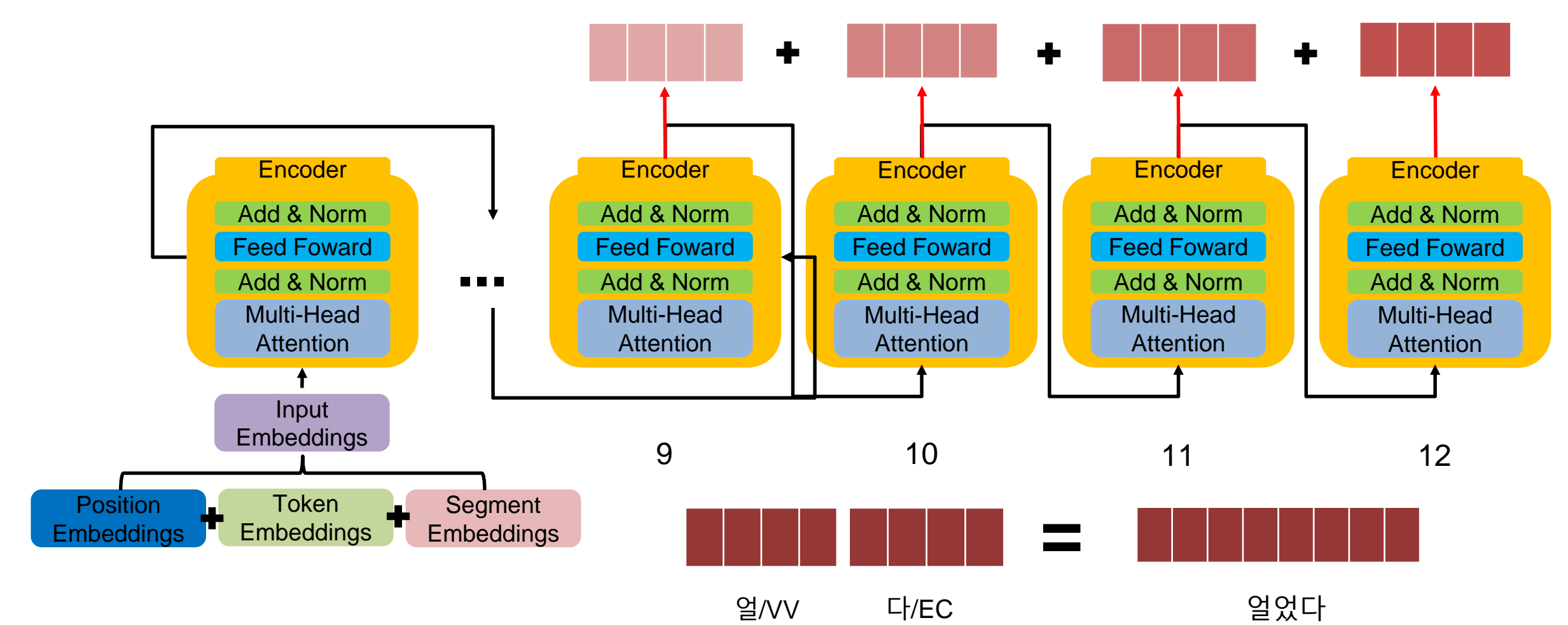
임베딩

1. ELMo 임베딩



2. BERT 임베딩

✓ 12개 중 마지막 4개의 인코더 블록을 합한 다음, 어절의 양 끝 토큰을 Concat



실험 및 결론

1. 실험 결과

✓ 같은 모델에 대해서는 다른 임베딩 사용

모델	UAS	LAS
BiAffine + BERT	94.07	91.92
BiAffine + ELMo + BERT	94.10	92.03
LR_Parser + BERT	94.14	92.07
LR_Parser + ELMo + BERT	94.23	92.12
RL_Parser + BERT	94.10	91.97
Ensemble	94.50	92.46

2. 결론

- ✓ 고성능 한국어 의존 구문 분석기 개발
- ✓ Left-to-Right 포인터 네트워크
- ✓ Right-to-Left 포인터 네트워크
- ✓ Deep Biaffine 네트워크
- ✓ 양상블 모델
- ✓ Dev 성능 UAS 94.50 / LAS 92.46