

# 한국어 상대시간관계 추출을 위한 LSTM 기반 모델 설계

임재균<sup>1</sup>, 정영섭<sup>2</sup>, 이영준<sup>1</sup>, 오교중<sup>1</sup>, 최호진<sup>1</sup>

<sup>1</sup>한국과학기술원 전산학부  
{rayote, yj2961, aomaru, hojinc}@kaist.ac.kr

<sup>2</sup>순천향대학교 빅데이터공학과  
bytecell@sch.ac.kr

## Design of LSTM-based Model for Extracting Relative Temporal Relations for Korean Texts

Chae-Gyun Lim<sup>1</sup>, Young-Seob Jeong<sup>2</sup>, Young Jun Lee<sup>1</sup>, Kyo-Joong Oh<sup>1</sup>, Ho-Jin Choi<sup>1</sup>

<sup>1</sup>School of Computing, KAIST

<sup>2</sup>Department of BigData Engineering, Soonchunhyang University

### 요약

시간정보추출 연구는 자연어 문장으로부터 대화의 문맥과 상황을 파악하고 사용자의 의도에 적합한 서비스를 제공하는데 중요한 역할을 하지만, 한국어의 고유한 언어적 특성으로 인해 한국어 텍스트에서는 개체간의 시간관계를 정확하게 인식하기 어려운 경향이 있다. 특히, 시간표현이나 사건에 대한 상대적인 시간관계는 시간 문맥을 체계적으로 파악하기 위해 중요한 개념이다. 본 논문에서는 한국어 자연어 문장에서 상대적인 시간표현과 사건 간의 관계를 추출하기 위한 LSTM(long short-term memory) 기반의 상대시간관계 추출 모델을 제안한다. 시간정보추출 연구에는 TIMEX3, EVENT, TLINK 추출의 세 가지 과제가 포함되지만, 본 논문에서는 특정 문장에 대해서 이미 추출된 TIMEX3 및 EVENT 개체를 제공하고 상대시간관계 TLINK를 추출하는 것만을 목표로 한다. 또한, 사람이 직접 태깅한 한국어 시간정보 주석 말뭉치를 대상으로 LSTM 기반 제안모델들의 상대적 시간관계 추출 성능을 비교한다.

주제어: 시간정보추출, 상대적 시간관계, 시간 표현, LSTM

### 1. 서론

시간정보추출에 관한 연구는 자연어 입력데이터로부터 시간이나 사건 표현을 추출할 뿐만 아니라 그 표현들 사이의 시간적 관계를 발견하는 것을 포함한다. 시간에 관한 문맥 정보는 자연어 표현에 암시된 의미적 특징을 포착하기 위해 이용될 수 있기 때문에, 시간정보를 추출하는 연구는 질의응답시스템 또는 구조화되지 않은 텍스트에 대한 처리를 하는 응용 등에서 중요한 비중을 차지한다. 이러한 연구가 전세계적으로 진행되면서 SemEval-2013의 shared task인 TempEval-3 [1]이 널리 알려져 있다. TempEval-3는 다량의 영어 문서에 대한 시간 및 사건 표현, 시간관계 추출에 대한 모델을 제시하고 성능을 평가한다. 최근에는 임상 도메인에서 시간정보를 추출하는 데 중점을 둔 SemEval의 다른 shared task인 Clinical TempEval [2]이 지속적으로 추진되고 있다.

그러나, TempEval과 같은 시간정보를 분석하는 연구는 대부분 영어를 대상으로 하며, 한국어 문서에 대한 시간정보추출 연구는 많이 이루어지지 않았다. 또한, 한국어에서 시간정보를 추출하기 위해 의존구문분석 트리(dependency parse tree)나 POS 태그와 같은 언어학적 정보가 자주 활용되지만, 한국어에 대한 언어분석 성능은 아직 충분하지 않다. 시간(TIMEX3)이나 사건(EVENT)

추출에 비해서 시간적 관계(TLINK)를 찾을 때 문장 내 의미론적 특성이 더욱 중요할 뿐만 아니라, 상대적인 시간관계를 이해하려면 시간 문맥도 파악할 필요가 있다. 언어분석 결과의 오류가 과급된다면 시간정보추출의 전반적인 성능 저하에 영향을 미친다.

본 논문에서는 LSTM(long short-term memory)을 적용한 신경망을 기반으로 한국어 문장에서 상대적인 시간관계를 나타내는 TLINK를 추출하는 딥러닝 모델을 제안한다. 시간정보추출 작업 중에서 TLINK 추출만을 대상으로 하는 모델을 설계하고, 시간관계 형성에 필요한 TIMEX3 및 EVENT 개체가 이미 추출되었다고 가정하여 모델의 입력으로 제공한다. 또한, LSTM 기반 모델의 입력 벡터를 생성하기 위해 전체 한국어 시간정보 주석 말뭉치를 분석하여 워드임베딩 모델을 구성하며, 이를 사용하여 TIMEX3 및 EVENT 개체를 임베딩 벡터로 변환한다. 이러한 TIMEX3/EVENT 개체의 임베딩 벡터들은 서로 다른 구조를 지닌 3가지 상대시간관계 추출 모델들을 학습할 때 활용된다. 3개의 모델은 서로 다른 개수의 LSTM 레이어로 이루어져 있으며, 실험을 통해 제안모델들의 상대시간관계 추출 성능의 차이를 비교한다.

본 논문은 다음과 같이 구성된다. 2장에서는 말뭉치로부터 시간관계를 추출하는 기존 연구를 소개하고, 3장에서는 제안하는 LSTM 기반 상대시간관계 추출 모델에 대

해 자세히 설명하며, 4장에서는 제안모델의 성능에 대한 실험결과를 나타내며, 5장에서 결론을 맺는다.

## 2. 관련 연구

시간정보추출 분야에서 딥러닝 모델을 활용하여 시간관계를 추출하려는 여러 연구들이 있다. Peng Zhou(2016)는 주어진 입력 문장 내에서 중요한 정보들이 아무 곳이나 위치하는 문제를 해결하기 위해 Attention-Based Bidirectional LSTM Networks 모델을 제안하였으며 F1-score 84%의 성능을 보였다[3]. Few Cheng(2017)은 관계 추출에 뛰어난 결과를 나타내는 DP(dependency path) 기반의 신경망 모델을 시간관계 분류에 적용하였고 F1-score 54%의 성능을 확인하였다[4]. Pengda Qin(2017)은 관계 분류에서 여러 개체들 간의 semantic knowledge가 완전히 활용되지 않는 문제를 해결하기 위해 Entity-pair-based Attention Mechanism 를 제안하였고 F1-score 84.7%의 결과를 산출하였다[5]. Julien Tourille(2017)은 narrative container identification을 위한 신경망 모델을 제안하였다. 이 모델은 최근에 입력되는 데이터에 편향되는 경향이 있는 LSTM의 특성에 대응하기 위해 문장을 역순으로 읽는 LSTM을 추가로 활용한다. 두 개의 LSTM을 통해서 주어진 입력에 대한 많은 정보를 학습할 수 있으며, 이 모델은 F1-score 61.3%의 성능을 보였다[6]. Yuanliang Men(2017)은 개체들 간의 시간관계를 복원하기 위해 shortest DP(dependency path)를 이용한 LSTM 기반 구조를 제안하였고 F1-score 47%의 결과를 나타냈다[7].

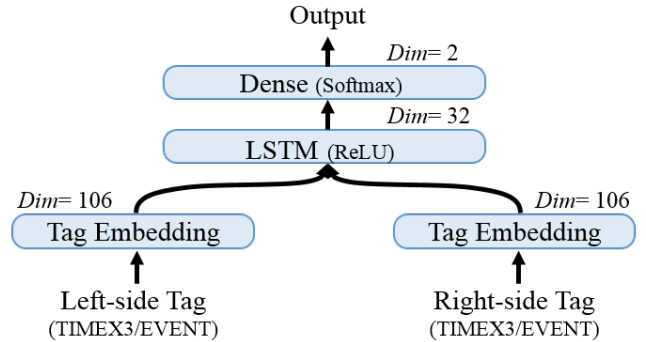
본 논문에서는 LSTM 기반 딥러닝 모델을 활용하여 TIMEX3과 EVENT 개체들 사이의 상대적인 시간관계를 학습한다. 기존의 연구는 주로 영어 코퍼스를 대상으로 시간관계를 추출하는 목적을 가지고 있지만, 본 논문에서는 한국어 코퍼스를 집중적으로 분석하고 한국어 특성을 반영하여 시간 및 사건 표현 간의 관계를 추출하는 것이 목표이다.

## 3. 제안 기법

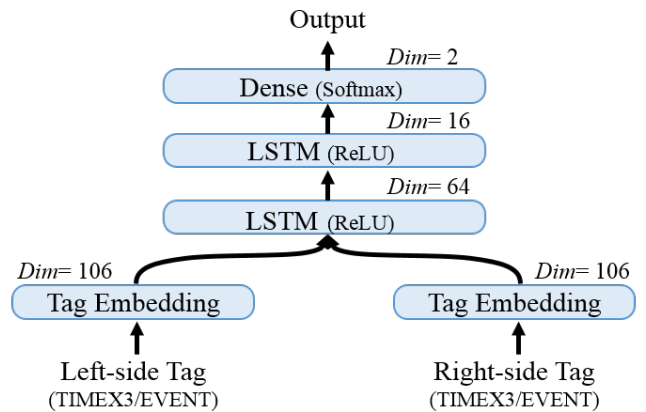
본 논문에서는 한국어 문장에서 시간 표현과 사건 간의 상대적인 시간관계를 추출하기 위해 TLINK에서 잠재성을 파악할 수 있는 LSTM 기반의 딥러닝 모델을 제안한다. 본 논문에서 제안하는 상대적 TLINK 추출 모델의 전체 디자인은 그림 1과 같다. 일반적으로, 시간정보추출은 주어진 말뭉치의 자연어 문장으로부터 TIMEX3 및 EVENT개체들을 추출하고, 그 개체들 사이의 시간적 관계를 발견하여 TLINK 개체로 추출하는 과정으로 이루어진다. 그러나, 본 논문에서는 딥러닝 모델에 기반한 TLINK 추출이라는 목적에 집중하기 위해서 TIMEX3 및 EVENT 개체들을 이미 추출한 상태라고 가정하고 제안모델의 입력으로써 제공한다. 먼저, 한국어 시간정보 주석 말뭉치에 포함된 자연어 문장으로부터 워드임베딩 모델을 학습한다. 각각의 문장은 단일 문자 단위로 토큰화되어 워드임

베딩 모델 학습에 사용되며, 임베딩 벡터는 100차원 공간으로 구성된다. 그 후, 워드임베딩 모델을 사용하여 TIMEX3 또는 EVENT 태그의 범위 내에 포함된 텍스트를 벡터화한다.

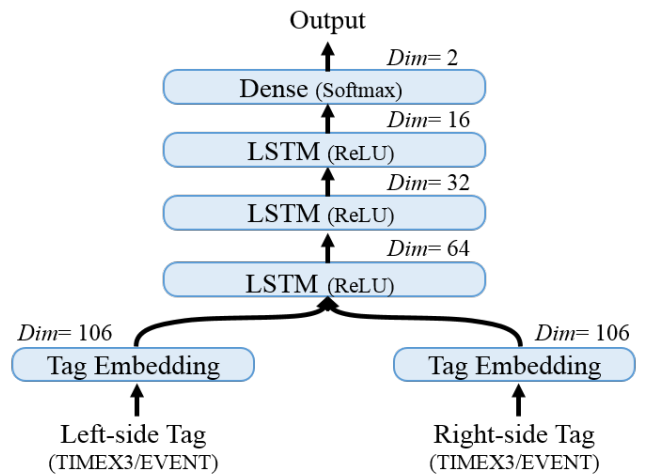
TLINK 태그는 관계유형 중 하나를 의미하는 relType 속성과 시간관계를 갖는 2개의 다른 개체들에 대한 참조로 구성된다. 따라서, 임베딩 벡터는 시간관계를 형성하는 2개의 개체에 대해 개별적으로 생성해야 하며, 이들



(a) RelModel-1



(b) RelModel-2



(c) RelModel-3

그림 1. 상대적 시간관계 추출을 위한 LSTM 기반 딥러닝 모델

을 하나의 벡터로 연결(concatenate)하여 제안모델의 입력으로 사용한다. 주어진 개체의 임베딩 벡터는 TIMEX3 또는 EVENT의 특정 클래스를 나타내도록 one-hot encoding된 벡터와 태그 범위의 텍스트로부터 얻어진 워드임베딩 벡터로 구성된다. 각 태그의 임베딩 벡터는 해당 개체의 클래스에 대한 6차원 벡터와 워드임베딩의 100차원 벡터를 결합하여 총 106차원을 가진다. 이러한 방식으로 TLINK 내의 2개 개체에 대한 임베딩 벡터를 생성한 후에, 하나로 연결하여 212차원의 벡터를 제안모델의 입력으로써 사용한다.

LSTM 레이어에서는  $\tanh$  함수와  $ReLU$ 를 activation function으로써 사용한다. 그림 1(a)의 RelModel-1은 1개의 LSTM 레이어를 사용하고, 32차원의 벡터를 출력한다. 그림 1(b)의 RelModel-2는 2개의 LSTM 레이어들로 이루어져 있으며, 첫 레이어에서 64차원 벡터, 두 번째 레이어에서 16차원 벡터를 출력한다. 그림 1(c)의 RelModel-3은 3개의 LSTM 레이어들로 구성되어 있고, 첫 레이어에서 64차원 벡터, 두 번째 레이어에서 32차원 벡터, 마지막 세 번째 레이어에서 16차원 벡터를 출력한다.

그 다음 Dense 레이어는 activation function으로 softmax 함수를 사용하여 2차원 벡터의 형태로 최종 결과를 생성한다. 제안모델의 출력 벡터는 주어진 2개의 시간 또는 사건 개체들이 상대적인 TLINK를 형성하는지 여부에 대해서 one-hot encoding된 벡터이다. 이 출력 벡터에서 최대값을 가지는 쪽을 선택하여 TLINK 형성 여부를 결정한다.

#### 4. 실험

본 논문에서 사용한 데이터셋은 2393건의 문서와 6190개의 한국어 문장을 포함하고 있는 *Korean TimeBank* [8]이다. 이 데이터에서 TIMEX3, EVENT 및 TLINK 개체의 수는 각각 3290, 17547, 4545이고, 상대적인 시간관계를 가진 TLINK 개체는 333개이다. LSTM 기반 TLINK 추출 모델을 구현하기 위해 Windows OS 환경에서 Keras 2.0.6 라이브러리와 Python 3.5.3을 사용하였다. 문자 기반 워드임베딩 모델의 어휘사전 크기는 2076이다.

제안모델의 학습 과정에서 상대적 관계를 가진 TLINK 개체를 모두 사용하고, 동일한 개수의 TLINK가 아닌 개체의 쌍을 샘플링하였다. 따라서, TLINK 및 비 TLINK를 포함한 데이터의 총 개수는 666이 된다. 전체 데이터를 분할하여 90%는 학습 데이터로, 나머지 10%는 테스트 데이터로 사용하였다.

실험에서는 그림 1과 같이 LSTM 레이어의 개수를 달리 하며 설계된 3가지 모델들에 대한 상대적 시간관계의 추출성능을 측정하였다. 표 1은 실험결과를 나타낸 것이며, 제안모델들의 F1-score는 평균적으로 0.75 수준이었다. RelModel-1과 RelModel-2의 경우에는 0.9 이상의 높은 성능을 보였던 반면, RelModel-3는 상대적으로 낮은 0.33의 결과를 보였다. 이 결과로부터 주어진 2개의 TIMEX3/EVENT 개체들에 대해서 상대적인 시간관계의 유무를 알아내는 데 너무 많은 레이어를 구성하는 것은 오히려 좋지 않다는 것을 알 수 있다. 특히, RelModel-3는

표 1. 상대적 시간관계 추출 성능

제안모델	Precision	Recall	F1-score
RelModel-1	0.95	0.94	0.94
RelModel-2	0.97	0.97	0.97
RelModel-3	0.25	0.50	0.33
Avg.	0.72	0.80	0.75

테스트 데이터 중에서 상대시간관계가 존재하는 경우에는 성공적으로 분류했으나, 관계를 포함하지 않는 데이터에 대해서 모두 잘못 분류하는 결과를 보였다. 이러한 실험결과는 RelModel-3가 상대시간관계에 대해 과적합(overfitting)되었을 가능성이 높음을 시사한다.

또한, RelModel-1과 RelModel-2의 실험결과에서도 상대시간관계가 없는 테스트 데이터에 대해서 잘못 분류하는 비율이 높은 경향을 보였다. 우리는 이 문제가 발생한 이유는 상대적 시간관계가 없는 TIMEX3-EVENT 조합의 유형이 너무 다양하기 때문이라고 추정하였다. 추후 상대적 시간관계의 조합에 대한 패턴을 분석하고 규칙 기반 접근법을 함께 활용하는 시도가 필요하다고 생각된다.

#### 5. 결론

본 논문에서는 시간 문맥을 파악하고 한국어 문장에서 상대적인 시간관계를 추출하기 위한 LSTM 기반의 딥러닝 모델들을 제안하였다. 본 논문의 주요 목적은 주어진 한국어 코퍼스에서 TIMEX3과 EVENT 개체들 사이의 상대시간관계를 설명하기 위한 TLINK를 발견하는 것이므로, 이미 추출된 TIMEX3과 EVENT 개체를 입력으로 제공하여 TLINK 추출 모델에만 집중하였다. 실험에서는 서로 다른 구조를 가진 LSTM 기반 모델들의 상대적 시간관계 추출의 성능을 비교하였다.

향후 연구에서는 상대시간관계 TLINK 추출의 전반적 성능을 향상시키기 위해 한국어 시간정보 주석 말뭉치의 규모를 충분히 확장시킨 후, 상대적 시간관계의 패턴을 분석하고 모든 관계유형에 대한 추출 성능을 고려하는 실험을 수행할 계획이다.

#### 감사의 글

본 연구는 미래창조과학부 산업융합원천기술개발사업 "휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발" (No.2013-0-00131)과 ICT유망기술개발지원사업 "지능형 대화 서비스를 위한 화용 및 문맥 분석 기반 대화솔루션 개발(No. 2017-0-00868)"의 지원으로 수행되었음.

#### 참고문헌

- [1] N. UzZaman, H. Llorens, J. Allen, L. Derczynski, M. Verhagen, and J. Pustejovsky, "Tempeval-3: Evaluating events, time expressions, and temporal

- relations,” *arXiv preprint arXiv:1206.5333*, 2012.
- [2] S. Bethard, G. Savova, W.-T. Chen, L. Derczynski, J. Pustejovsky, and M. Verhagen, “Semeval-2016 task 12: Clinical tempeval,” In *Proceedings of SemEval*, pp. 1052-1062, 2016.
- [3] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, “Attention-based bidirectional long short-term memory networks for relation classification,” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2, pp. 207-212, 2016.
- [4] F. Cheng and Y. Miyao, Classifying Temporal Relations by Bidirectional LSTM over Dependency Paths, In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2, pp. 1-6, 2017.
- [5] P. Qin, W. Xu, and J. Guo, “Designing an adaptive attention mechanism for relation classification,” *Neural Networks (IJCNN), 2017 International Joint Conference on*, pp. 4356-4362, IEEE, 2017.
- [6] J. Tourille, O. Ferret, A. Neveol, and X. Tannier, “Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers,” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2, pp. 224-230, 2017.
- [7] Y. Meng, A. Rumshisky, and A. Romanov, “Temporal Information Extraction for Question Answering Using Syntactic Dependencies in an LSTM-based Architecture,” *arXiv preprint arXiv:1703.05851*, 2017.
- [8] Y.-S. Jeong, W.-T. Joo, H.-W. Do, C.-G. Lim, K.-S. Choi, and H.-J. Choi, “Korean TimeML and Korean TimeBank,” In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC)*, pp. 356-359, 2016.