

# 딥 러닝 기반의 멘션 페어 모델을 이용한 한국어 상호참조해결

박천음, 최경호, 이창기  
강원대학교 컴퓨터학과  
{parkce, gangsparkle, leeck}@kangwon.ac.kr

## Korean Coreference Resolution using the Deep Learning based Mention Pair Model

Cheon-Eum Park, Gyeong-Ho Choi, Chang-Ki Lee  
Dept. of Computer Science, Kangwon National University

### 요 약

최근 자연어처리에 딥 러닝이 적용되고 있다. 딥 러닝은 기존의 기계학습 방법들과 달리, 자질 추출 및 조합 등과 같이 사람이 직접 수행해야 했던 부분들을 자동으로 처리할 수 있는 장점이 있다. 본 논문에서는 기존 상호참조해결에 적용했던 SVM 대신 딥 러닝을 이용할 것을 제안한다. 실험결과, 딥 러닝을 이용한 시스템의 성능이 57.96%로 SVM을 이용한 것보다 약 9.6%만큼 높았다.

### 1. 서론

언어란 사람이 자신의 생각을 표현하기 위한 의사소통 중 하나이며 그 의미는 명확하게 전달되어야 한다. 그러나 자연어는 매우 다양하게 존재하며 그 표현 역시 무궁무진하다. 따라서 이런 자연어를 컴퓨터로 처리하기 위해서는 단순히 일반적인 규칙들로 처리하기에 한계가 있다. 이에 따라, 주로 기계학습 기반의 방법들이 연구되고 있으며, 최근에는 이 중에서 신경망(artificial neural network)을 확장시킨 딥 러닝(deep learning)이 다양하게 활용되고 있다.

기존의 기계학습 알고리즘은 주어진 자질(feature)에 대한 최적의 가중치를 찾아낸다. 그러나 기존의 기계학습 방법에서는 사람이 자질들을 직접 추출(feature extraction)하거나 조합(feature design)해야 하기 때문에 많은 시간과 노력이 필요하다. 따라서 이러한 문제를 해결하기 위하여 최근에는 자연어처리에 딥 러닝을 적용하려는 다양한 시도가 있었다[1].

딥 러닝은 여러 히든 레이어(hidden layer)를 포함하고 있는 신경망이다. 신경망의 히든 레이어는 비선형의 활성화 함수(activation function)를 통해 방출되는 유닛(unit)들로 구성되며, 이러한 유닛들을 기반으로 비선형 조합(non-linear activation)을 수행하면 표현력을 강화시키게 된다. 즉, 가중치에 따라 파라미터들이 조절되는 것으로 자질 추출 및 조합이 자동으로 진행되는 것이다. 이에 따라 각 레이어마다 보다 더 높은 수준의 추상화(abstraction)를 진행하게 된다[2]. 그러나 히든 레이어가 많아질 수록 학습에 따른 계산량이 많아지며 오류율의 역전파(back-propagation)에 대한 학습이 어려워진다. 따라서 딥 러닝은 사전학습(pre-training), 드랍아웃(drop-out)등의 방법들을 고안하여

위의 문제들을 해결하였다[3, 4].

상호참조해결(coreference resolution)은 어떤 문서 내에서 등장하는 단어들 중, 서로 같은 개체(entity)를 가리키는 단어들을 찾아 서로 간에 연관(linking)을 지어주는 것이다. 흔히 문장 내에서 이미 언급한 표현을 재사용할 때 대명사를 이용하여 표현하는데, 먼저 언급한 표현을 선행사(antecedent)라 하고 대명사와 같이 다시 표현하는 단어를 대용어(anaphora)라 한다. 또한 이런 선행사를 대명사 이외의 단어, 즉 두 단어가 가리키고 있는 개체는 명확히 같지만, 대용어 부분의 단어가 선행사와 다른 단어로 표현되는 경우, 이것을 상호참조(coreference)라 한다. 예를 들어, 대용어일 경우에는 다음과 같다.

“오바마 대통령의 고향은 호놀룰루이다.”,

“그는 현재 백악관에 산다.”

위의 두 문장에서 [오바마 대통령]는 선행사가 되며, [그]가 대용어가 된다. 또한 상호참조일 경우에는 다음과 같다.

“오바마는 작년 봄에 한국에 방문하였다.”,

“그 이후에 미국 대통령은 말레이시아를 방문하기로 예정되어 있었다.”,

“그는 현재 백악관에서 집무 중에 있다.”

위와 같은 문장들에서는 [오바마]가 역시 선행사가 되며, [미국 대통령]과 [그]는 [오바마]와 같은 개체로서 상호참조 되는 단어가 된다. 따라서 상호참조해결은 대용어를 포함한 더 확장된 개념으로 서로 같은 개체들 간의 집합관계를 만드는 것이다.

### 2. 관련 연구

상호참조해결은 MUC(Message Understanding Conference

)에서 시작 되어 MUC5, MUC6, MUC7 그리고 conll-2011까지 다뤄진 중요한 정보추출(information extraction)방법 중 하나이다. 상호참조해결을 효과적으로 수행하기 위해서는 기본 적으로 규칙 기반과 통계 기반의 방법으로 나뉜다. 먼저 규칙 기판으로는 Stanford에서 제안한 다단계 시브(multi-pass sieve)를 이용한 방법[5]이 영어에 대하여 우수한 성능을 보였으며, 이를 한국어에 적용한 연구 역시 좋은 성능을 보였다[6]. 다음으로 통계 기판으로는 기계학습을 적용하여 상호참조해결을 진행하는데, 이에 사용되는 기본적인 모델로 멘션 페어(mention pair), 멘션 랭킹(mention ranking), 엔티티 멘션(entity mention) 등이 있다[7]. 그러나 아직 한국어 상호참조해결에 대하여 멘션 페어를 이용하였을 경우, 규칙 기반 보다 좋은 성능을 보이지 못하고 있다[8]. 이에 따라 [9]에서는 먼저 규칙 기판을 통해 상호참조해결을 진행하고 그 결과로부터 자질을 추출하여 Support Vector Machine(SVM)[10]을 적용하는 가이드 방법을 제안하였으며, 그 결과로 규칙 기반에 비해 다소 향상된 성능을 보였다.

본 논문에서는 기존 [7]에서 다룬 멘션 페어 모델에 대해 SVM을 이용하는 방법 대신 최근 우수한 성능을 보이고 있는 딥 러닝을 적용할 것을 제안한다.

### 3. 딥 러닝 기반 상호참조해결

멘션 페어 모델은 두 개의 멘션이 하나의 쌍을 이루 뒤 상호참조인지 아닌지에 대해 판별하는 방법이다. 즉, 활성 멘션( $m_k$ , 현재 멘션)과 후보 선행사( $m_j$ ) 간의 관계에서 서로 상호참조인지 분류하는 것으로, 이것에 대한 표현은  $instance(m_j, m_k)$ 와 같다. 이 중, 각 인스턴스(instance)는  $(m_j, m_k)$ 에 대한 자질 조합으로 본 논문에서는 총 31개의 자질을 baseline으로 설정하였다[9]. [8]에서는 한국어 상호참조해결에 SVM을 이용하여 두 멘션의 자질이 유사하면 상호참조 하는 것으로  $(m_j, m_k)$ 와 같이 쌍을 이루며, 그 외의 멘션들은 네거티브(negative)가 된다( $m_{j+1}, m_{j+2}, \dots, m_{k-1}$ ).

딥 러닝의 기본 구조는 입력 레이어(input layer)와 여러 층의 히든 레이어 그리고 출력 레이어(output layer)로 구성되어 있다. 자연어처리에서의 딥 러닝은 입력을 one-hot 표현(해당 유닛에 입력된 단어만 1, 나머지 0, 예를 들어 크기 5의 단어사전 중 첫 번째가 "프랑스"라면 프랑스를 입력 받은 유닛은 [10000]과 같다)으로 받게 되는데, 이 표현은 단어사전의 크기에 비례하여 차원이 증가하게 되므로 많은 부하가 따른다. 따라서 자연어처리에서의 이런 단점을 보완하기 위해 단어표현(word embedding)을 적용한다.

단어표현이란 Neural Network Language Model(NNLM) [2, 12]을 기반으로 고안된 것으로 Input layer와 선형연산 되는 가중치 매트릭스( $W$ )를 말한다. NNLM은 기존의 확률적 언어모델을 딥 러닝에 적용한 것이며, 임의의 현재 단어(목표 단어)를 예측하기 위해 주변의  $n$ 개의 단어들을 입력으로 하고 목표 단어가 등장할 확률을 출력으로 한다. [2, 12]에서는 입력 레이어, 투사 레이어(projection layer), 히든 레이어, 출력 레이어 등

으로 신경망을 구성하였으며, 입력 벡터와 가중치 매트릭스(lookup table,  $LT$ 로 정의)의 선형 연산을 통해 차원을 축소하는 방법을 제안했다. 즉, 입력 레이어에서 각 유닛 하나의 크기가  $1 \times V$ 이고  $LT$ 의 크기가  $V \times l$ 이라고 하면, 이 둘의 곱은  $1 \times l$ 이 된다( $V$ 는 단어사전의 크기이고 본 논문에서는 6만개 사용, 그리고  $l$ 은 임의의 크기로 정의되며, 보통 50차원으로 적용한다). 이에 따라 기존  $1 \times V$ 의 차원이  $1 \times l$ 로 축소되며, 입력 레이어에서 유닛 각각의 결과는 투사 레이어에서 하나의 벡터로 합쳐지게 되고 이것은 입력 벡터에 비해 많은 차원 축소가 발생하게 된다. 그 이후 일반적인 딥 러닝 연산이 수행되고 역전파를 통하여  $LT$ 는 최적화 값으로 갱신되며, 이것을 단어표현이라 한다.

[그림 1]은 본 논문에서 사용한 신경망을 표현한 것이며, 이에 적용할 입력은 멘션(즉, 선행사  $A$ , 활성 멘션  $M$ )과 그에 해당하는 자질이 된다. 본 논문에서 사용한 자질은 [9]의 자질들을 기반으로 하며, 이 중에서 both\_pronoun과 both\_proper\_name은 제외 하였다. 이 둘은 선행사와 활성멘션에 대하여 대명사 또는 고유명사 정보를 각각 조합한 자질로서, 딥 러닝에서는 자질조합이 자동으로 일어나기 때문이다. 그리고 가이드 멘션 페어 모델에서 규칙기반 시스템의 결과 자질도 본 논문에서 사용하지 않기 때문에 포함시키지 않았다.

본 논문에 적용한 단어표현은 [1]에서 NNLM을 통해 추출한 한국어 단어표현을 사용하였다. 그러나 단어사전은 단일 단어로 구성 되었기 때문에 명사구를 사용하기에는 어려움이 있어서 멘션의 중심어(head)만을 이용하여 단어표현에 적용하였다.

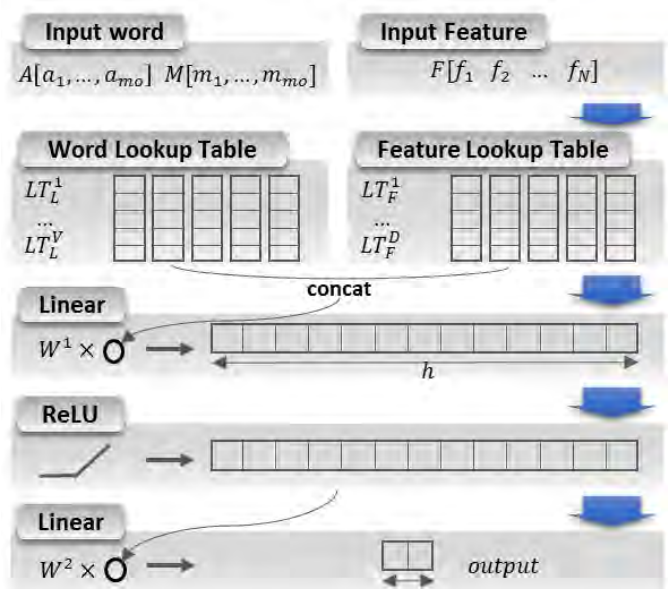


그림 1. 상호참조해결을 위한 신경망 구조

입력된 자질에는 위와 마찬가지로 자질 표현(feature embedding)을 적용하고, 단어표현을 적용한 결과와 함께 하나의 벡터(즉, concatenate)로 만들어 투사 레이어로 전달한다. 투사 레이어에서는 전달 받은 벡터와

가중치 매트릭스( $W_1$ )를 곱한 후의 결과( $h$ 만큼의 길이)를 활성화함수에 적용하고 히든 레이어로 전달한다. 여기서 활성화함수는 비선형 변환으로, 기존에 Sigmoid 또는 Tanh를 사용하였지만, 본 논문에서는 이 보다 더 성능이 뛰어난 ReLU를 적용한다[10]. 그 이후 히든 레이어는 다시 한번  $W_2$ 와 곱셈을 하게 되고 그 결과가 출력 레이어로 전달된다. 출력 레이어에서는 Softmax 함수를 이용하여 각 상태의 확률을 구한 뒤, Cross entropy를 목적함수로 적용하고 역전파 알고리즘을 이용하여 학습한다. 이 과정 중에서 과적합을 해결하기 위하여 드랍아웃 기술을 적용하였다(드랍아웃은 히든레이어에서 특정 확률로 임의의 유닛을 0으로 설정하는 것으로 보통 0.5의 확률을 적용한다. 이 방법은 Bagging과 유사한 효과를 내며, 과적합을 줄이는데 효과적이다).

4. 실험 및 결과

본 논문에서 한국어 상호참조해결을 수행하기 위해서 사용한 데이터 셋은 총 500문서로서 뉴스 도메인 150문서, 퀴즈 도메인 350쌍(퀴즈 도메인은 질문과 답 한 쌍을 한 문서로 다루며, 장학퀴즈 150쌍, WiseQA 200쌍)을 사용하였다. 이 중, 학습 데이터는 뉴스 도메인 130문서, 퀴즈 도메인 350쌍을 사용하였으며, 테스트 데이터로 뉴스 도메인 20문서를 사용하였다. 그리고 성능측정 척도(measure)는 [5]의 Conll 값(MUC, B-cube, Ceaf-e의 F1값 평균)을 적용하였으며, 멘션 경계(M.B., mention boundary), 중심어 경계(H.B., head boundary) 등, 두 가지 방법에 대한 성능을 측정하였다. 여기서 멘션 경계는 중심어를 수식하는 수식어구까지 포함한 멘션을 말하며(즉, 명사구 전체), 중심어 경계는 수식어구를 제외하고 중심어만 정의한 멘션을 말한다.

딥 러닝을 이용한 상호참조해결에서는 히든 유닛의 개수, 드랍아웃의 확률, 자질표현의 차원 수를 조절하여 최적의 파라미터를 구하였다. 그리고 최적의 파라미터에 기반한 딥 러닝을 멘션페어 방법에 적용하였을 경우, 기존의 SVM을 이용한 것보다 얼마나 더 뛰어난 성능을 보이는지 확인하기 위해 비교실험을 수행하였다.

먼저 최적의 파라미터를 찾는 실험에 대하여, [그림 2]는 히든레이어의 유닛 수를 100부터 1300까지 입력으로 하여 최적의 히든레이어 유닛 개수를 찾는다.

본 실험에서 사용된 초기 파라미터는 휴리스틱(heuristic)에 의해 드랍아웃은 0.5, 자질표현에 대한 차원의 수는 10으로 정의하였다.

[그림 2]의 결과로 히든레이어의 유닛이 700일 경우에 F1(conll)값[6]이 멘션 경계(M.B.)는 57.96%, 중심어 경계(H.B.)는 64.43%로 가장 우수한 성능을 보였다. 이에 따라 히든레이어의 유닛 수를 700으로 드랍아웃과 자질표현의 차원에 대해 실험을 진행하였다. 드랍아웃은 임의의 히든 유닛을 확률에 기반하여 제외하는 것으로, 입력되는 파라미터 값이 높아질수록 제외할 가능성이 커진다. 다음 [그림 3]은 드랍아웃에 대한 성능을 보인다.

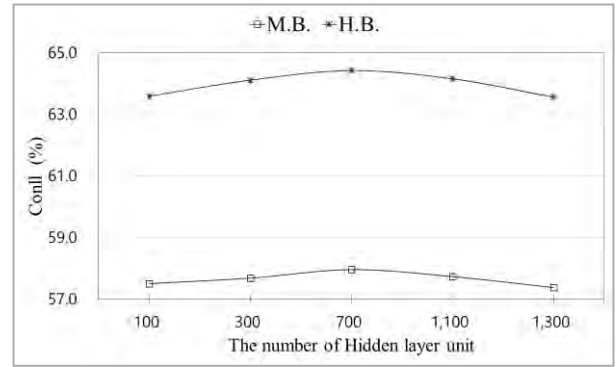


그림 2. 히든 레이어 유닛 최적화

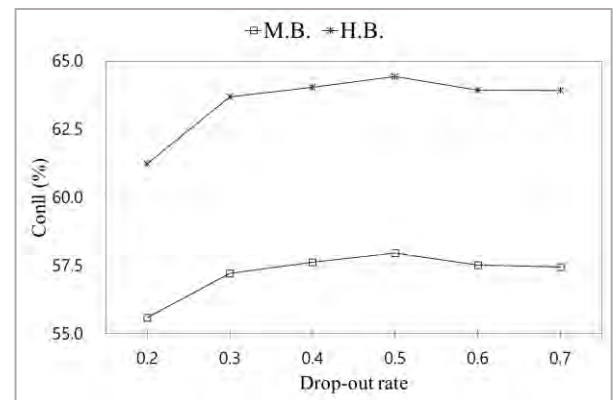


그림 3. 드랍아웃에 대한 확률 최적화

[그림 3]에 따라 파라미터 값이 낮거나 높을 때 낮은 성능을 보였으며(예를 들어, 멘션 경계에서 0.2는 55.60%, 0.7은 57.45%), 적절한 확률인 0.5일 때 멘션 경계에서 57.96%, 중심어 경계에서 64.43%로 최고의 성능을 보였다. 다음 [그림 4]는 자질표현의 차원에 따른 성능을 보인다.

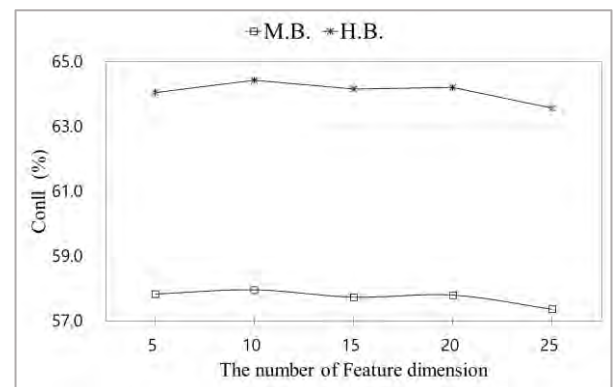


그림 4. 자질표현의 차원 수에 따른 최적화

[그림 4]에 대하여, 자질표현의 차원이 낮을 경우에는 자질과 곱해질 벡터가 적기 때문에 표현력이 약해져 낮은 성능을 보이며, 반대로 높은 경우에는 많은 자질 벡터를 갖게 되고 벡터 표현이 복잡해지기 때문에 좋은 성능을 보이지 못한다. 그 결과, 자질표현에 대한 차원이 10일 때 이전 실험들과 마찬가지로 최적의 성능을 보였으며, 멘션 경계에서 57.96%, 중심어 경계에서 64.43%이다. 이에 따라, 최적의 파라미터는

히든 레이어의 유닛이 700개, 드롭아웃 확률은 0.5 그리고 자질표현은 10차원으로 나타났다.

[그림 5, 6]은 멘션 페어 모델에 기반한 상호참조해결에 대하여 딥 러닝과 SVM의 차이를 보이며, [그림 5]는 멘션 경계, [그림 6]은 중심어 경계에 대한 성능이다.

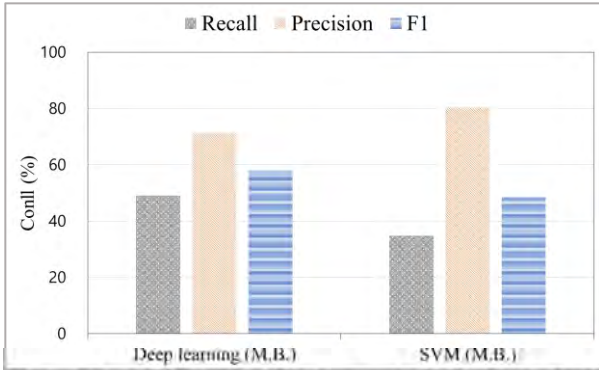


그림 5. 상호참조해결에 대한 딥 러닝과 SVM의 비교 (멘션 경계)

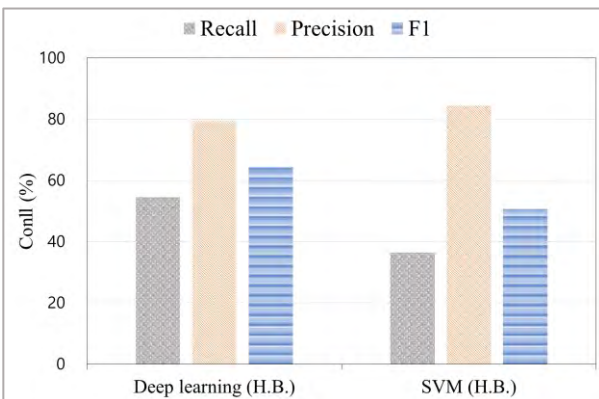


그림 6. 상호참조해결에 대한 딥 러닝과 SVM의 비교 (중심어 경계)

[그림 5, 6]에서의 딥 러닝과 SVM은 최적의 파라미터를 사용한 것이며, SVM의 경우에는 Cost값으로 3000과 이진 분류 문제에 사용하는 owps(one word per sente-nce) 포맷[10]을 적용하였다.

[그림 5]의 결과, SVM의 성능은 48.36%이며, 딥 러닝을 적용한 멘션 페어 모델의 성능이 약 9.6%로 많이 향상된 것을 볼 수 있으며, [그림 6]은 SVM이 50.55%로서 약 13.88%의 차이로 딥 러닝이 더 높다. 그리고 상호참조해결에서의 재현율(recall)은 유사한 멘션들 간의 연관성을 의미하는데 이 또한 멘션 경계는 약 14.27%, 중심어 경계는 약 18.12%의 차이로 딥 러닝이 더욱 좋은 성능을 보였다.

### 5. 결론 및 향후 연구

본 논문에서는 기존 SVM을 이용하여 멘션 페어 기반의 한국어 상호참조해결을 진행했던 것에 대하여 딥 러닝을 적용할 것을 제안하였다. 그 결과, baseline 자질을 사용할 경우, 딥 러닝의 성능은 F1 값이 멘션 경계에서 57.96%, 중심어 경계에서 64.43%를 보였다.

이에 따라, 기존의 SVM 보다 멘션 경계는 약 9.6%, 중심어 경계는 약 13.9%의 차이로 훨씬 뛰어난 성능을 보여, 딥 러닝이 SVM보다 더욱 적합하다는 것을 알 수 있었다.

향후 연구로는 상호참조에 보다 더 효과적인 자질을 찾고 더 많은 데이터 셋을 구축하여 성능을 더욱 향상 시킬 것이다.

### 감사의 글

본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음[10044577, 휴먼 지식증강서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발]

### 참고문헌

- [1] 이창기, 김준석, 김정희. 딥 러닝을 이용한 한국어 의존 구문 분석. 제26회 한글 및 한국어 정보처리 학술대회, p. 87-91, 2014.
- [2] R. Collobert, et al. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 12, 2011.
- [3] G. E. Hinton, et al. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 2006.
- [4] G. E. Hinton, et al. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.
- [5] H. Lee, et al. Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics* 39.4: 885-916, 2013.
- [6] 박천음, 최경호, 이창기. Multi-pass Sieve를 이용한 한국어 상호참조해결. *정보과학회논문지* 41.11, p. 992-1005, 2014.
- [7] A. Rahman, and V. Ng. Supervised models for coreference resolution. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. Association for Computational Linguistics, p. 968-977. 2009.
- [8] 최경호, 박천음, 이창기. SVM 기반의 Mention Pair Model을 이용한 한국어 상호참조해결. *한국정보과학회 2014 한국컴퓨터종합학술대회 논문집*, p. 598-600, 2014.
- [9] 최경호, 박천음, 이홍규, 이창기. SVM 기반의 가이드 멘션페어 모델을 이용한 한국어 상호참조해결. *한국정보과학회 2014 동계학술발표회 논문집*, p. 482-484, 2014.
- [10] C. Lee, and M.G. Jang. Fast training of structured SVM using fixed-threshold sequential minimal optimization. *ETRI journal* 31.2, p. 121-128, 2009.
- [11] X. Glorot, et al. Deep Sparse Rectifier Neural Networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. JMLR W&CP Volume*. Vol. 15, p. 315-323, 2011.
- [12] Y. Bengio, et al. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3, p. 1137-1155, 2003.