

# Coreference Resolution for Korean Using Random Forests

Seok-Won Jeong<sup>†</sup> · MaengSik Choi<sup>†</sup> · HarkSoo Kim<sup>\*\*</sup>

## ABSTRACT

Coreference resolution is to identify mentions in documents and is to group co-referred mentions in the documents. It is an essential step for natural language processing applications such as information extraction, event tracking, and question-answering. Recently, various coreference resolution models based on ML (machine learning) have been proposed. As well-known, these ML-based models need large training data that are manually annotated with coreferred mention tags. Unfortunately, we cannot find usable open data for learning ML-based models in Korean. Therefore, we propose an efficient coreference resolution model that needs less training data than other ML-based models. The proposed model identifies co-referred mentions using random forests based on sieve-guided features. In the experiments with baseball news articles, the proposed model showed a better CoNLL F1-score of 0.6678 than other ML-based models.

**Keywords :** Coreference Resolution, Random Forest, Sieve-Guided Features

## 랜덤 포레스트를 이용한 한국어 상호참조 해결

정석원<sup>†</sup> · 최맹식<sup>†</sup> · 김학수<sup>\*\*</sup>

### 요 약

상호참조 해결은 문서 내에 존재하는 멘션들을 식별하고, 참조하는 멘션끼리 군집화하는 것으로 정보 추출, 사건 추적, 질의응답과 같은 자연어 처리 응용에 필수적인 과정이다. 최근에는 기계학습에 기반한 다양한 상호참조 해결 모델들이 제안되었으며, 잘 알려진 것처럼 이런 기계학습 기반 모델들은 상호참조 멘션 태그들이 수동으로 부착된 대량의 학습 데이터를 필요로 한다. 그러나 한국어에서는 기계학습 모델들을 학습할 가능한 공개 데이터가 존재하지 않는다. 그러므로 본 논문에서는 다른 기계학습 모델보다 적은 학습 데이터를 필요로 하는 효율적인 상호참조 해결 모델을 제안한다. 제안 모델은 시브-가이드 자질 기반의 랜덤 포레스트를 사용하여 상호참조하는 멘션들을 구분한다. 야구 뉴스 기사를 이용한 실험에서 제안 모델은 다른 기계학습 모델보다 높은 0.6678의 CoNLL F1-점수를 보였다.

**키워드 :** 상호참조 해결, 랜덤 포레스트, 시브-가이드 자질

### 1. 서 론

상호참조 해결(Coreference Resolution)은 문서 내에 존재하는 멘션(mention)들을 식별하고, 참조하는 멘션끼리 군집화(clustering)하는 것으로 정보 추출, 사건 추적, 질의응답과 같은 자연어 처리 응용에 필수적인 과정이다. 상호참조 해결은 주로 규칙 기반 방식이나 기계학습(machine learning) 기반 방식으로 이루어진다. 대표적인 규칙 기반 방식인 Stanford[1]의 다단계 시브(multi-pass sieve) 시스템이 좋은

성능을 보이면서, 기존에는 규칙 기반 방식의 상호참조 해결이 각광을 받아왔지만, 규칙 기반 방식은 상호참조를 검사하는 규칙을 정의하고 규칙들 간의 최적화를 수행하는 것이 어렵다는 단점이 있다. 최근에는 기계학습에 기반한 다양한 상호참조 해결 모델들이 제안되었으며, 잘 알려진 것처럼 이런 기계학습 기반 모델들은 멘션 태그들이 수동으로 부착된 대량의 학습 데이터를 필요로 한다. 그러나 한국어에서는 기계학습 모델들을 학습할 가능한 공개 데이터가 존재하지 않는다. 그러므로 본 논문에서는 다른 기계학습 모델보다 적은 학습 데이터를 필요로 하는 효율적인 상호참조 해결 모델을 제안한다. 제안 모델은 시브-가이드 자질 기반의 랜덤 포레스트를 사용하여 상호참조하는 멘션들을 구분한다. 본 논문의 구성은 다음과 같다. 2장에서는 제안 시스템과 관련 있는 연구들에 대해 기술하고 3장에서는 제안 시스템에 대해 설명한다. 4장에서는 실험 및 결과에 대해 기술하며 5장에서 글을 맺는다.

\* 본 연구는 엔씨소프트 산학연구협력 과제의 지원을 받아 수행되었음. 또한 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.R-20160906-004163, 빅데이터 자동 태깅 및 태깅 기반 DaaS 시스템 개발).

\*\* 본 논문은 한국정보과학회 “2016 한국컴퓨터종합학술회의”에서 발표된 “시브 자질 기반 랜덤 포레스트를 이용한 한국어 상호참조 해결”을 확장한 것임.

<sup>†</sup> 준 회 원 : 강원대학교 컴퓨터정보통신공학전공 박사과정

<sup>\*\*</sup> 정 회 원 : 강원대학교 컴퓨터정보통신공학전공 부교수

Manuscript Received : October 4, 2016

Accepted : October 13, 2016

\* Corresponding Author : HarkSoo Kim(nlprkim@kangwon.ac.kr)

## 2. 관련 연구

Stanford의 다단계 시브 시스템은 대표적인 규칙 기반 방식의 상호참조 해결 시스템이며, 간단한 방법으로 높은 성능을 보였다. 다단계 시브 시스템은 상호참조를 검사하는 규칙인 시브(sieve)를 다양하게 정의하여 적용한다. 각 시브를 정확률(precision)이 높은 것부터, 정확률은 떨어지지만 재현율(recall)이 높은 것으로 순차 적용하는 것이 특징이다. 박천음 외[2]는 Stanford의 다단계 시브를 한국어에 적용하였고, 영어권에서 구구조 분석을 기반으로 멘션 탐지를 수행하는 것과 달리, 의존구조 분석을 기반으로 멘션 탐지를 수행하였다. 중심화 이론(centering theory)[3]은 전통적인 방식의 대용어 처리 이론으로, 담화 내에서 주제 변화가 적은 쪽이 더 선호되는 경향, 즉, 응집성(coherence)이 있을 것이라는 가정을 바탕으로 대용어 처리를 수행하였다. 그러나, 중심화 이론에서 정의한 여러 중심들을 계속적으로 확인하는 것이 복잡한 과정이므로, S-list[4] 모델은 Prince 분류체계[5]를 통한 순위화, 간단한 구조 등 단순화된 방법의 대용어 처리를 제안하였다. 본 논문에서는 S-list를 한국어에 맞게 수정하여 적용하였고, 이를 통해 상호참조의 검사 순서를 결정한다. 랜덤 포레스트[6]는 기계 학습 알고리즘의 하나로 학습 데이터의 부분 집합들로부터 서로 다른 다수의 결정 트리(decision tree)를 학습함으로써, 단일 결정 트리에 비해 일반화 성능이 높은 것으로 알려진 모델이다. 본 논문에서는 랜덤 포레스트를 이용하여 상호참조 해결을 수행하였다.

## 3. 랜덤 포레스트를 이용한 상호참조 해결

본 논문에서 제안하는 시스템의 구조도는 아래와 같다.

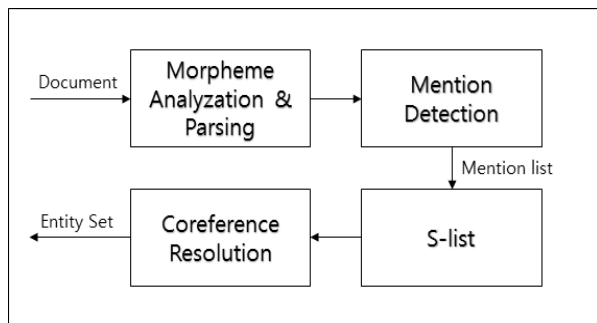


Fig. 1. System Architecture

Fig. 1과 같이 시스템은 문서의 형태소 분석 및 구문 분석 결과로부터 멘션을 탐지하며, 탐지된 멘션들은 S-list에 기반한 우선 순위에 따라 적절한 선행 멘션들과 상호참조 여부를 검사한다. 시스템의 결과로, 같은 대상을 참조하는 멘션들의 집합인 개체(entity)들이 반환된다.

### 3.1 멘션 탐지

멘션 탐지는 상호참조 해결의 대상이 되는 모든 멘션들을 찾아내는 것으로, 본 논문의 멘션 탐지는 구문 분석 결과를 바탕으로 이루어진다. 구문 분석으로 찾은 문장 성분 중 명사구들이 멘션이 된다. 아래의 Fig. 2는 제안 시스템이 수행한 멘션 탐지의 예이다.

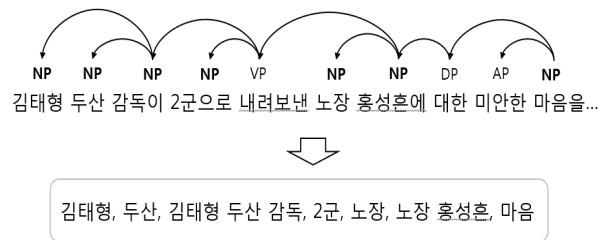


Fig. 2. An Example of Mention Detection

탐지된 멘션들 중에서 중심어가 겹치는 멘션의 경우 작은 멘션은 제거한다. 예를 들어, Fig. 2에서 “김태형 두산 감독”, “두산 감독”의 멘션을 탐지할 수 있다. 그러나 중심어가 “감독”으로 겹치므로 더 작은 멘션인 “두산 감독”은 제거된다. 또한, 일반적인 멘션 탐지에서는 구문 분석 결과에 따라 지나치게 긴 멘션까지 고려하는 경우도 있다. “김태형 두산 감독이 2군으로 내려보낸 노장 홍성흔에 대한 미안한 마음을...”이라는 문장에서 문장 성분들의 수식에 따라 “2군으로 내려보낸 노장 홍성흔”이나 “노장 홍성흔에 대한 미안한 마음”과 같은 멘션을 고려할 수 있다. 그러나 본 논문에서는 (그림 2)의 “노장 홍성흔”, “마음”처럼 중심이 되는 단어로 이루어진 작은 멘션들만을 고려한다.

### 3.2 S-list를 이용한 비교 순서 결정

상호참조 해결에서 멘션들 간의 비교 순서는 상호참조 해결 성능에 영향을 미치는 중요한 요소이다. 만약 상호참조하지 않는 멘션들을 먼저 비교하여 상호참조하는 것으로 판단하는 경우 오류가 누적되어 크게 잘못된 개체를 생성할 수 있다. 현재 멘션의 이전에 나타난 선행 멘션들 중 가장 가까운 것부터 비교하는 것이 일반적이지만, 본 논문에서는 Strube의 S-list[4]를 한국어에 맞게 수정하여 S-list의 우선 순위를 기반으로 비교하였다. S-list의 우선순위는 아래 Fig. 3과 같다.



Fig. 3. Priority in S-List

Fig. 3과 같이 각 멘션들은 두 종류의 우선순위 정보를 가진다. 첫 번째는 멘션의 형태 및 상호참조 결과에 따른

우선순위이다. 멘션이 고유명사, 영어로 이루어졌거나 한 번 이상 상호참조된 경우 OLD, 수식어를 가지거나 다어절인 경우 MED, 그 외의 경우는 NEW가 된다. 두 번째는 멘션이 문장 내에서 어떤 성분으로 사용되었는지에 따른 우선순위이다. 멘션에 보조사 ‘은/는’이 붙는 경우 topic, 주어로 사용된 경우 subject, 목적어로 사용된 경우 object, 그 외의 경우엔 etc가 된다. 우선순위 결정 시 첫 번째 우선순위를 먼저 비교한 뒤에 두 번째 우선순위를 비교한다. 또한, 우선순위가 동일한 경우엔 더 최근에 언급된 멘션을 먼저 비교한다. 아래 Fig. 4는 S-list 우선순위의 실제 적용 사례이다.

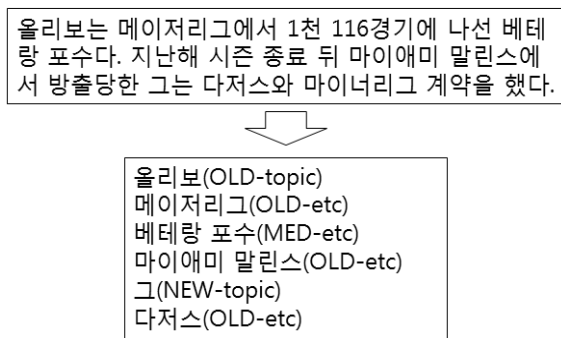


Fig. 4. An Example of S-List

탐지된 멘션들은 Fig. 4와 같은 우선순위 정보를 가지며, 상호참조 검사 시 해당 우선순위에 따른 순서대로 검사한다. 예를 들어, “베테랑 포수”의 선행 멘션을 검사할 때, 일반적인 경우 최근 언급된 순서인 “메이저리그”, “올리보”의 순서로 검사한다. 그러나 제안 시스템은 “올리보”의 우선순위가 OLD-topic으로 “메이저리그”의 OLD-etc보다 높으므로 “베테랑 포수”와 “올리보”간의 상호참조 여부를 먼저 검사한다. “그”의 선행 멘션을 검사할 경우엔 “올리보”, “마이애미 말린스”, “베테랑 포수”(“올리보”와 상호참조하므로 이 시점에서 는 OLD-etc가 됨), “메이저리그”의 순서로 검사한다.

### 3.3 랜덤 포레스트 모델

제안 시스템은 멘션들 간의 상호참조 여부를 판단하기 위해 랜덤 포레스트를 사용한다. 일반적으로 Maximum Entropy Model(MEM)[7], Support Vector Machine(SVM)[8]과 같은 통계 기반의 기계학습 알고리즘이 더 좋은 성능을 보이는 것으로 알려져 있다. 그러나, 충분한 학습 데이터가 주어지지 않은 상황에서는 결정 트리, 랜덤 포레스트와 같이 자동으로 규칙을 생성하는 모델들이 강세를 나타낼 것이라 예상하였다. 제안 시스템은 모델을 학습하기 위해 다단계 시브 시스템에서 주로 사용되는 규칙들을 자질로 사용한다. Table 1은 각각 제안 시스템에서 사용한 자질의 종류, 각 자질에서 나타날 수 있는 값, 자질의 예시를 보여준다. 예시는 “류현진”, “투수 류현진”의 두 멘션으로부터 추출된 자질이다.

Table 1. Input Features of the Proposed Model

Features	value	example
Exact String match	0, 1	0
Acronym match	0, 1	1
Head match	0, 1	1
Last morpheme match	0, 1	1
Pronoun type match	0, 1	0
Head match (Definite noun phrase)	0, 1	0
Domain type match	0, 1	0
Morpheme match	Exact match, inclusion, partial match, none match	inclusion
The number of identical eojools	Zero or more	1
Marched morpheme	Morpheme	류현진
Position of matched morpheme	Left, center, right	right
POS of matched morpheme	Part-of-speech	NNP
Proportion of matching	Exact match: 0, partial math: 1-3, none march: 4	2
Priority of the previous mention	OLD, MED NEW	OLD

Table 1에서 전체 문자열 비교, 약어 비교, 중심어 비교, 마지막 형태소 비교, 대명사 타입 비교, 한정 명사구일 경우 중심어 비교, 도메인 타입 비교 등은 규칙 기반 방식에서 주로 사용되는 자질이며, 제안 시스템에서도 동일하게 사용한다. 전체 문자열 비교는 두 멘션의 문자열이 완벽히 일치하는지를 검사하는 자질로 “류현진”, “류현진”과 같이 문자열이 완벽히 일치하는 경우를 검사한다. 개체들은 종종 약어의 형태로 축약되는데, 약어 비교는 축약된 개체들 간의 상호참조를 해결하기 위한 자질이다. 약어는 동적으로 휴리스틱하게 생성한다. 예를 들어, 비교할 두 멘션이 “김태형 두산 감독”, “김 감독”인 경우 “김태형 두산 감독”으로부터 “김 감독”, “두산 감독”, “김태형 감독” 등의 약어를 생성한 뒤 이 약어들과 “김 감독”을 비교한다. 중심어 비교의 경우 “투수 류현진”, “류현진”과 같이 중심어가 일치하는 경우를 해결하기 위한 자질이다. 멘션이 대명사로 나타나는 경우에는 단순한 문자열 비교로는 상호참조를 해결하기가 어렵다. 대명사 타입 비교 자질은 미리 대명사의 타입들을 정의해 놓고 정의된 대명사의 타입과 다른 멘션의 개체명 타입을 비교한다. 예를 들어, 비교할 두 멘션이 “투수 류현진”, “그”인 경우 대명사 “그”의 타입인 ‘PER’과 “투수 류현진”의 개체명 타입인 ‘PER’을 비교하여 상호참조를 해결한다. 도메인 타입 비교는 본 논문에서 사용한 데이터에 특화된 자질이다. 본 논문에서 사용한 야구 도메인에 대해 주요 단어들의 타입을 미리 정의하였다. 예를 들어, “투수”의 경우 ‘PIT

(pitcher), “심판”의 경우 ‘UMP(umpire)’로 정의된다. 주요 단어들의 타입을 개체명의 세부 타입과 비교하여 상호참조를 해결한다. 비교할 두 멘션이 “심판”, “박근영”인 경우 “심판”의 타입인 ‘UMP’와 “박근영”이라는 개체명의 세부 타입인 ‘UMP’를 비교하여 상호참조를 해결한다.

이외의 자질들은 제안 시스템의 성능을 향상시키기 위해 추가한 자질들이다. 주로 어휘 정보를 이용한 자질들을 추가하였다. 일치 어절 수는 두 멘션이 일치한 경우의 어절 수로 예제에서는 “류현진”과 “투수 류현진” 중 한 어절이 일치하므로 1이 된다. 불일치하는 부분의 비율은 두 멘션에서 같은 형태소 외의 비율을 말한다. “류현진”과 “투수 류현진”은 절반이 불일치하므로 불일치 비율은 2가 된다.

#### 4. 실험 및 결과

##### 4.1 실험 데이터

본 논문의 실험을 위해 272개 야구 도메인 뉴스 기사에 대해 수동으로 정답을 부착하여 데이터를 구축하였다. 272개 뉴스 기사에서 총 개체 수는 1,473개로 나타났으며, 멘션 쌍의 수는 6,126개로 나타났다. 성능은 전체 문서에 대해 10배 교차 검증(10-fold cross validation)을 통해 측정하였고, 평가 지표는 MUC[9], B-cube[10], CEAF[11]와 3개 지표의 평균인 CoNLL[12]을 사용하였다.

##### 4.2 S-list 우선순위 성능 평가

제안 시스템 중 S-list의 성능을 평가하기 위해 랜덤 포

레스트 모델에서 S-list를 적용하였을 때와 적용하지 않았을 때의 성능을 비교하였다. 실험 결과는 Table 2와 같다. Table 2에서 ‘None’은 S-list의 우선순위를 고려하지 않고, 최근에 언급된 순서대로 비교하였을 경우의 성능이며, ‘S-list’는 S-list를 적용하여 비교하였을 경우의 성능이다. S-list를 적용한 경우가 그렇지 않은 경우에 비해 모든 지표에서 성능이 향상됨을 확인할 수 있다.

##### 4.3 랜덤 포레스트 모델 성능 평가

본 논문에서 제안한 랜덤 포레스트 모델의 성능을 평가하기 위해 다른 기계학습 모델과의 비교 실험을 진행하였다. MEM과 랜덤 포레스트는 Weka[13]를 사용하여 실험하였고, SVM은 LibSVM[14]을 사용하여 실험하였다. 실험 결과는 위의 Table 3과 같으며, 제안한 랜덤 포레스트 모델이 다른 기계학습 모델보다 높은 성능을 보이는 것을 확인할 수 있다. 이는 통계 기반 모델은 각각의 시브-가이드 자질들이 가중치를 가지는 반면, 제안한 랜덤 포레스트 모델에서는 각 자질들이 효과적으로 작동할 수 있는 순서가 결정되므로 다단계 시브 시스템의 장점을 더 반영할 수 있기 때문인 것으로 보인다. 표에는 나타나지 않았지만, 학습 속도 또한 랜덤 포레스트를 이용한 모델이 월등히 빠른 속도를 보였다. 실험 데이터에서 MEM이 약 3시간, SVM이 약 10시간의 학습 시간이 소요된 반면, 랜덤 포레스트는 약 30분이 소요되었다. 일반적으로 통계 기반의 기계학습 모델인 MEM, SVM이 높은 성능을 나타내는 것과 달리, 상황에 따라 다른 기계학습 모델도 좋은 성능을 낼 수 있음을 나타낸다.

Table 2. Effectiveness of S-List

	MUC			B-cube			CEAF			CoNLL		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
None	0.7525	0.6207	0.6565	0.7108	0.5929	0.6197	0.7407	0.6627	0.6824	0.7347	0.6254	0.6529
S-list	0.7565	0.6339	0.6725	0.7175	0.6031	0.6356	0.7607	0.6659	0.6953	0.7449	0.6343	0.6678

P : Precision, R : Recall, F1 : F1-score

Table 3. Performance Comparison with Other Machine Learning Models

	MUC			B-cube			CEAF			CoNLL		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
MEM	0.6866	0.5513	0.5853	0.6630	0.5132	0.5533	0.7156	0.5721	0.6192	0.6884	0.5455	0.5859
SVM	0.6864	0.5942	0.6107	0.6404	0.5694	0.5748	0.6862	0.6383	0.6393	0.6710	0.6007	0.6082
Random Forest	0.7565	0.6339	0.6725	0.7175	0.6031	0.6356	0.7607	0.6659	0.6953	0.7449	0.6343	0.6678

## 5. 결 론

본 논문에서는 S-list를 통해 멘션 간의 비교 순서를 결정하는 방법을 제안하였고, 다른 기계학습 모델보다 적은 학습 데이터를 필요로 하는 효율적인 상호참조 해결 모델을 제안하였다. 야구 뉴스 기사(272개 문서, 10배 교차 검증)를 이용한 실험에서 제안 모델은 다른 기계학습 모델들(SVM 모델: 0.6082, MEM 모델: 0.5859)보다 더 나은 CoNLL F1-점수(0.6678)를 보였다.

현재 멘션 탐지 단계에서 상호참조하지 않는 멘션들을 탐지하여 잘못된 결과가 나타나는 경우가 존재한다. 향후 연구로 상호참조 가능성이 낮은 멘션들을 필터링하는 방법을 연구하고, 현재 모델로 해결이 불가능한 오류들을 처리하기 위해 다양한 자질들로 실험할 예정이다. 또한, 최근 활발히 연구되고 있는 딥러닝 기반의 모델을 연구할 예정이다.

## References

[1] Lee, Heeyoung, et al., "Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task," *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, 2011.

[2] Park, Cheon-Eum, Kyoung-Ho Choi, and Changki Lee, "Korean Coreference Resolution using the Multi-pass Sieve," *Journal of KIISE*, Vol.41, No.11, pp.992-1005, 2014.

[3] Brennan, Susan E., Marilyn W. Friedman, and Carl J. Pollard. "A centering approach to pronouns," *Proceedings of the 25th Annual Meeting on Association for Computational Linguistics*. 1987.

[4] Strube Michael, "Never look back: An alternative to centering," *Proceedings of the 17th International Conference on Computational Linguistics-Volume 2*, Association for Computational Linguistics, 1998.

[5] Ellen F. Prince, "Toward a taxonomy of given-new information," *Radical Pragmatics*, 1981.

[6] Breiman Leo, "Random Forests," *Machine Learning*, Vol.45, No.1, pp.5-32, 2001.

[7] Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, Vol.22, No.1, pp.39-71, 1996.

[8] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, Vol.9, No.3, pp.293-300, 1999.

[9] M. Vilain, J. Burger, J. Aberdeen, D. Connolly, and L. Hirschman, "A model-theoretic coreference scoring scheme," *Proceedings of the 6th Conference on Message Understanding*, Association for Computational Linguistics, pp.45-52, 1995.

[10] A. Bagga and B. Baldwin, "Algorithms for scoring coreference chains," *The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, Vol.1, pp.563-566, 1998.

[11] X. Luo, "On coreference resolution performance metrics," *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 25-32, 2005.

[12] S. Pradhan, L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel, and N. Xue, "Conll-2011 shared task: Modeling unrestricted coreference in ontonotes," *Proc. of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, Association for Computational Linguistics, pp. 1-27, 2011.

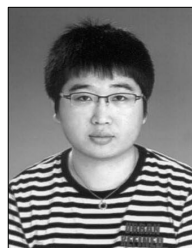
[13] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, Vol.11, Iss.1, 2009.

[14] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *Proc. of ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol.2, Iss.3, Apr., 2011, Article No.27, 2011.



### 정 석 원

e-mail : nlpsw@kangwon.ac.kr  
 2014년 강원대학교 컴퓨터정보통신공학과 (학사)  
 2016년 강원대학교 컴퓨터정보통신공학과 (석사)  
 2016년~현 재 강원대학교  
 컴퓨터정보통신공학전공 박사과정  
 관심분야: 상호참조 해결, 개체명 인식



### 최 맹 식

e-mail : nlpmschoi@kangwon.ac.kr  
 2009년 강원대학교 컴퓨터학부(학사)  
 2011년 강원대학교 컴퓨터정보통신공학과 (석사)  
 2011년~현 재 강원대학교 컴퓨터정보  
 통신공학전공 박사과정  
 관심분야: 형태소 분석, 구문 분석, 관계 추출





## 김 학 수

e-mail : nlpdrkim@kangwon.ac.kr

1996년 건국대학교 전자계산학과(학사)

1998년 서강대학교 컴퓨터학과(석사)

2003년 서강대학교 컴퓨터학과(박사)

2004년 University of Massachusetts,  
Amherst 박사후연구과정

2005년 한국전자통신연구원 선임연구원

2006년~현재 강원대학교 컴퓨터정보통신공학전공 부교수

관심분야: 자연어처리, 대화모델링, 정보검색, 정보추출,

질의응답시스템