工學碩士學位論文

개체명 부착 말뭉치에서 자동 오류 수정

2015年　**6**月

昌原大學校　大學院

컴 퓨 터　工 學 科

金　　中　　韓

工學碩士學位論文

개체명 부착 말뭉치에서 자동 오류 수정

Automatic Named-Entity Corpus Error Correction

指導敎授 車揖遠

이 論文을 工學碩士學位論文으로 提出함.

2015年 **6**月

昌原大學校 大學院

컴 퓨 터 工 學 科

金 中 韓

金中韓의 碩士學位 論文을 認准함.

審 査 委 員 長　　이 광 휘　　㊞

審 査 委 員　　정 성 욱　　㊞

審 査 委 員　　차 정 원　　㊞

2015年　**6**月

昌 原 大 學 校　大 學 院

목차

그림 목차

표 목차

제 Ⅰ 장

서론

개체명은 인명(PERSON), 지명(LOCATION), 조직명(ORGANIZATION) 등 특정한 의미를 가지는 개체의 이름을 말한다. 이러한 개체명은 문장에서 특정 범주에 속하는 의미 있는 정보를 가지고 있다. 개체명 정보는 정보 추출(Information Extraction), 정보 검색(Information Retrieval), 질의응답 시스템(Question/Answering System) 등 자연어처리 분야에서 다양하게 사용된다. 특히 시멘틱웹, 질의응답 시스템 등 높은 정밀도를 요구하는 자연어처리 응용에 대한 연구가 활발히 진행되고 있다.

개체명을 문장에서 인식하기 위해 개체명 인식 시스템은 말뭉치로부터 기계학습(Machine Learning)을 통해 구현한다. 기계학습을 하기 위해서는 개체명 태그가 부착되어 있는 학습 말뭉치가 필요하다. 학습 말뭉치는 기계학습에서 매우 중요한 부분으로 인식되고 있다. 이는 기계학습의 예측 성능이 학

습 말뭉치의 크기와 정확도에 따라 큰 차이를 보이기 때문이다. 따라서 학습 말뭉치의 정확도를 높이기 위한 방법에 대한 연구들이 이루어졌다[1,2,3]. 그리고 학습 말뭉치의 부족으로 인해 발생하는 자료희귀문제(Data Sparseness Problem)를 회피하기 위해 대량의 말뭉치가 필요하게 된다. 하지만 대량의 말뭉치를 제작하는 것은 많은 시간과 노력이 요구된다. 실질적으로 사람이 160 텍스트를 태깅하기 위해 8시간이 소요된다[4]. 또한 다수의 사람이 제작하는 말뭉치의 일관성을 유지하기는 매우 힘들다. 이는 사람이 말뭉치를 제작하기 위해 태그의 범주를 정의하고, 태그를 부여하는 과정에서 발생한다. 여러 사람이 함께 작업을 진행함으로 태그의 범주가 정의되어 있어도 각자 주관적인 생각으로 인해 서로 다른 태그를 부여함으로써 일관성의 문제가 발생한다. 이러한 이유로 학습에 사용되는 학습 말뭉치에 태그가 부착된 말뭉치만 사용하는 지도 학습(Supervised Learning)을 대체하기 위해 비지도 학습(Unsupervised Learning)이나 반지도 학습(Semi-Supervised Learning)에 대한 연구도 진행되었다. 그러나 이러한 연구에도 불구하고 학습 말뭉치의 중요성은 줄어들지 않고 있다[5].

개체명 말뭉치를 제작하기 위해 사람이 직접 수작업으로 진행하거나 개체명 부착 도구를 사용하는 방법으로 말뭉치 제작에 필요한 시간과 노력을 줄인다.

청와대에서 6월 1일 메르스에 대한 공식 입장을 발표했다.

문장에서 나타나는 청와대는 지명 또는 조직을 의미하는 단어로 애매성을 가지는 개체명이다. 해당 문장에는 조직을 의미하는 개체명으로 사용되고 있지만 개인의 편차로 인해 지명에 대한 태그를 부여해 오류를 발생시킬 수도 있

다. 또한 실수로 인해 개체명 태그를 잘못 부여하거나 부여하지 않는 경우에도 오류가 발생한다. 이러한 오류는 학습 단계에서 문제를 발생시켜 정확도를 떨어뜨리는 요인이 되고, 말뭉치를 검수하는 과정에서도 오류를 수정하기 위해 많은 시간과 노력이 필요하게 된다.

본 논문에서는 여러 연구자들이 손으로 직접 제작한 개체명 부착 말뭉치에 존재하는 개체명 오류를 자동 시스템을 통해 1차적으로 오류를 줄여줌으로써 말뭉치 제작에 소요되는 시간과 노력을 줄이고 정확한 말뭉치 구축을 목표로 한다. 논문의 구성은 다음과 같다. Ⅱ장에서는 한국어 개체명 부착 말뭉치를 제작하는 시스템의 특징과 방법을 설명하고 개체명 인식 시스템에 대한 특징과 성능을 알아본다. Ⅲ장에서는 제안 시스템의 구조도와 특징에 대해서 기술한다. Ⅳ장에서는 여러 실험을 통해 시스템을 평가 하고 분석한다. 끝으로 Ⅴ장에서는 결론과 향후 과제를 서술한다.

제 Ⅱ 장

관련 연구

본 장에서는 개체명 부착 말뭉치의 제작 방법에 따라 분류되는 말뭉치를 설명하고, 한국어 개체명 부착 말뭉치 생성 도구에 대한 관련 연구와 개체명 인식 시스템 관련 연구, RDR(Ripple-Down Rules)에 대한 관련 연구를 살펴본다.

1. 개체명 부착 말뭉치의 분류

개체명 부착 말뭉치는 제작 방법에 따라 2가지로 분류 할 수 있다.

1) 사람이 직접 작성한 말뭉치 : Gold-standard 말뭉치

2) 시스템이 자동으로 생성한 말뭉치 : 정답에 가까운 말뭉치

Gold-standard 말뭉치는 사람이 직접적으로 생성하거나 개체명 부착 도구를 사용해 생성한 말뭉치를 말한다. 개체명 부착 말뭉치를 제작하기 위해 특정 단어가 의미할 수 있는 범주를 고유의 식별자인 태그(Tag)로 정의한다. 예를 들어 인명, 지명, 조직명 등에 대한 범주를 정의하고, 고유의 식별자인 태그를 부착하여 구축하게 된다. 이를 위해 다양한 정의들 중에서 하나의 태그를 올바르게 정의하여 사용하게 된다[6]. 아래의 문장은 개체명 부착 말뭉치의 예제이다.

<청와대:OGG[1)]>에서는 <6월 1일:DT[2)]> 공식 입장을 발표했다.

Gold-standard 말뭉치를 제작하기 위해서는 많은 시간과 노력이 필요하다. 하지만 영어, 독일어, 일본어 등의 몇몇 언어를 제외하고는 공개된 개체명 부착 말뭉치가 부족하다[6]. 이러한 개체명 부착 말뭉치 부족 문제를 해결하기 위해서 자동으로 개체명을 부착하여 생성하는 정답에 가까운 말뭉치를 구축하는 연구가 진행되었다. [7]에서는 인명, 지명, 조직명 3가지 개체명 타입의 개체명 사전을 사용해 웹에서 해당 개체명이 많이 나타나는 텍스트를 추출한다. 추출된 텍스트는 3가지의 개체명 사전에 포함되는 개체명이 많이 나타난다. 이 텍스트를 복합 명사 분리 과정을 통해 정제하고, 정제된 텍스트에서 개체명으로 인식된 개체명의 좌·우 문맥 패턴을 학습하여 개체명 말뭉치를 생성하는 방법을 제안하였다. 또한 비교적 개체명 인식이 힘든 한국어 제목 개체명

1) 조직명에 대해 정의된 개체명 태그
2) 일자에 대해 정의된 개체명 태그

을 인식하고 개체명 사전을 생성하여 개체명 말뭉치를 구축하는 방법도 연구되었다[8]. 그리고 디비피디아의 개체명 리스트와 위키피디아의 링크를 사용해 개체명 말뭉치를 구축하는 연구가 있었다[9].

2. 한국어 Gold-standard 말뭉치 생성 도구

한국어 개체명 말뭉치를 제작하는데 있어 소요되는 시간과 노력을 줄이기 위해 지원 도구에 관한 연구도 진행되었다. [10]에서는 문화유산정보에 대한 개체명 말뭉치 구축을 지원하는 도구가 연구되었다. 이 연구에서는 규칙 패턴을 개체명기반 패턴, 음절기반 패턴, 문맥기반 패턴으로 나눈다. 각각의 패턴의 후보에 대해서 식(1)을 사용해 가장 높은 Score를 가지는 후보를 제시한다.

$$\text{Score}_i = \begin{cases} \text{개체명후보}_i\text{를 추천한 개체명기반 패턴의 빈도수} \times \lambda_1 \\ + \text{ 개체명후보}_i\text{를 추천한 음절기반 패턴의 빈도수} \times \lambda_2 \\ + \text{ 개체명후보}_i\text{를 추천한 문맥기반 패턴의 빈도수} \times \lambda_3 \end{cases} \quad (1)$$

단, $\lambda_1 > \lambda_2 > \lambda_3$

여기서 i는 각 패턴에 의해 추천되는 후보를 뜻하고 λ는 각 패턴에 주어지는 가중치를 나타낸다. Score는 각 패턴에서 i번째 개체명 후보가 나타나는 빈도수와 각 패턴의 가중치를 곱한 합으로 계산한다. 그 중 가장 높은 Score로 계산된 개체명 후보가 선택된다.

[11]는 개체명 부착 도구를 관리자(Administrator) 도구, 어노테이터(Annotator) 도구, 컨쥬게이터(Conjugator) 도구로 분리하여 개체명 부착 말

뭉치 제작을 지원한다. 관리자 도구에서는 개체명 말뭉치의 분배 및 데이터 관리를 목적으로 하고, 어노테이터 도구에서 실질적인 개체명 부착을 지원한다. 또 컨쥬게이터 도구는 어노데이터 도구의 결과물을 통합하는 과정에서 같은 단어에 대해서 서로 다른 개체명 태그를 부착한 경우 올바르게 부착된 태그만을 선택하기 위해 사용한다.

3. 개체명 인식

개체명 인식에 대한 연구는 오래 전부터 많이 연구되었다. 초기 연구는 HMM(Hidden Markov Model)을 사용하여 8개의 범주에 대한 개체명을 부착하는 연구가 진행되었다[12]. 해당 연구에서는 영어의 문자 자질[3]인 대·소문자 자질을 사용하여 학습하였고, 93%의 높은 성능을 보였다.

최근에는 소셜 네트워크 서비스인 트위터의 텍스트를 사용해 개체명을 정규화시켜 개체명을 인식하였다[13]. 또한 위키피디아의 링크 정보와 병렬 말뭉치를 사용한 개체명 인식도 연구되었다[14].

한국어에 대한 개체명 인식에 대한 연구도 활발히 이루어지고 있다[15, 16, 17, 18]. [15]에서는 개체명과 개체명을 구성하는 단어들의 상관관계를 분석하고, 이를 활용해 HMM기반의 개체명 인식 모델을 제안하였다. [16]에서는 개체명 인식과 개체명 클래스 분류를 각각 분리하여 개체명을 인식하는 방법을 제안한다. 개체명 인식에서는 독립 가정(Independence Assumption)이 필요한 HMM보다 성능의 뛰어남이 증명된 CRFs(Conditional Random Fields)를

[3] 어떠한 대상을 구분할 수 있는 특징을 말한다. 예를 들어 딸기와 포도를 구별하기 위해 사용되는 자질은 열매의 색, 열매의 형태가 자질이 될 수 있다.

사용하였고, 인식 결과를 ME(Maximum Entropy)를 사용하여 개체명을 분류하였다. [17]은 Structural SVM을 사용한 개체명 인식을 제안하였다. 기존의 CRFs 방법[16]보다 학습시간을 줄이기 위해 Pegasos 알고리즘을 수정하여 성능을 유지하면서, 4%가량의 학습시간을 향상하였다.

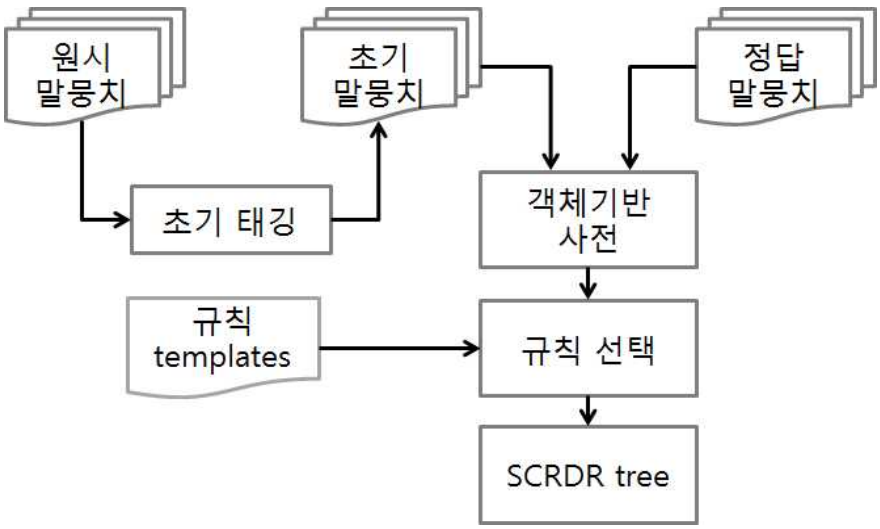
위의 모든 방법은 지도 학습을 이용한 방법이다. 이 방법들은 학습 말뭉치가 많이 필요하다. [18]에서는 학습을 위해 상대적으로 학습 말뭉치가 적어도 학습 가능한 공동 학습(Co-Training)을 이용한 개체명 인식 시스템을 제안되었다.

4. RDR(Ripple-Down Rules)

RDR(Ripple-Down Rules)은 1993년 Edwards와 Compton에 의해 처음 제안되었다[1]. RDR은 지식베이스시스템의 데이터를 분류하기 위해 사용되었다. 지식의 대부분이 정답으로 분류될 때에 오답으로 분류되는 일부에 대한 규칙을 생성함으로써 작은 수의 규칙만으로도 지식 데이터를 표현할 수 있다. 이후 SCRDR(Single Classification Ripple Down Rules)과 MCRDR(Multiple Classification Ripple Down Rules), NRDR(Nested Ripple Down Rules) 등의 다양한 형태의 RDR이 제안되었다[19]. SCRDR은 규칙에 따라 하나의 분류로만 선택되도록 하는 반면 MCRDR은 여러 개의 독립적인 분류가 선택될 수 있다.

[20]은 Brill이 제안한 Brill's Tagger[21]의 규칙 생성에 대한 알고리즘을 SCRDR로 변경하여 학습에 사용한다. [21]에서는 각 단어에 대해서 태그를 미리 부여하고 미리 정의된 16개의 templates를 적용해 규칙을 생성한다. 규

칙은 미리 부여한 태그와 정답 말뭉치의 태그가 서로 다를 때 생성된다. 따라서 이미 부여된 태그를 생성된 규칙에 적용되는 태그에 대해서 오류 수정을 진행한다. [20]은 규칙을 생성하는 templates를 27개로 늘리고 규칙을 생성하였다. 이렇게 생성된 규칙을 SCRDR에 적용하여 오류 수정에 사용하게 된다. SCRDR은 각 노드의 조건에 따라 'EXCEPT'과 'FALSE'이 발생한다. 'EXCEPT'이 발생했을 때 하위의 각 노드로 이동하며 각각의 조건에 따라 'FALSE'이 발생할 때까지 노드를 이동한다. 최종적으로 'EXCEPT'된 규칙에 의해 분류가 결정된다.



<그림 Ⅱ-1> RDRPOSTagger[20]의 학습 흐름도

<그림 Ⅱ-1>은 RDRPOSTagger[20]의 학습 흐름도이다. 원시 말뭉치를 형태소 품사 사전을 사용해 품사를 부착하는 초기 태깅을 진행하여 초기 말뭉치를 생성한다. 생성된 초기 말뭉치와 같은 문장의 정답 말뭉치를 입력으로 사용한다. 초기 말뭉치와 정답 말뭉치에서 같은 단어에서 다른 품사를 부착하

고 있는 단어에 대한 모든 규칙을 생성하게 된다. 이렇게 생성된 규칙 중 규칙 templates에 해당되는 규칙만을 선택한다. 최종적으로 선택된 규칙을 SCRDR에 적용하여 규칙을 정렬 한다.

제 Ⅲ 장

개체명 부착 코퍼스 자동 오류 수정

개체명 부착 말뭉치는 자연어처리 분야에서 중요한 부분으로 인식되고 있다. 개체명 부착 말뭉치는 형태소 분석 말뭉치와 구문 분석 말뭉치와 다르게 공개된 말뭉치의 리소스가 상당히 작은 편이다. 따라서 개체명 인식 시스템을 구축하기 위하여 많은 시간과 노력이 소모하여 개체명 부착 말뭉치를 제작해야 한다. 하지만 사람이 직접 제작하는 Gold-standard 말뭉치도 다수의 사람이 작업을 진행함으로써 각 개인의 편차로 인해 개체명의 일관성에 대한 문제가 발생한다. 또한 단순한 실수로 인해 발생할 수 있는 오류의 가능성도 가지고 있다. 따라서 이러한 일관성 문제 및 실수 등으로 인해 발생한 오류를 자동으로 수정 할 수 있으면 더 정확한 개체명 부착 말뭉치를 더 적은 시간과 노력으로 제작할 수 있다는 생각에서 출발했다.

본 장에서는 논문이 해결하려고 하는 문제를 정의하고 제약사항을 설명한다. 또한 논문에서 제안하는 시스템을 각 단계로 구분하여 진행 과정을 설명

한다.

1. 문제 정의

초기 말뭉치와 정답 말뭉치를 비교하여 초기 말뭉치에서 나타나는 오류에 대해서 분석하고 발생하는 오류를 4가지 유형으로 나누어 분류하였다. 일반적으로 오류는 삽입, 삭제, 치환으로 분류된다. 따라서 오류 유형 2는 삽입에 대한 오류, 오류 유형 3은 삭제에 대한 오류, 오류 유형 1, 4는 치환에 대한 오류로 분류할 수 있다.

<표 Ⅲ.1> 개체명 말뭉치에서 나타나는 오류 유형

유형		설명	예	
			초기 말뭉치	정답 말뭉치
분류	1	동일한 단어에 서로 다른 개체명	애플-CV[4]	애플-OGG
	2	초기 말뭉치에만 존재하는 개체명	애플-CV	-
인식	3	정답 말뭉치에만 존재하는 개체명	-	애플-OGG
	4	서로 다른 경계를 가지는 개체명	윈도7-TM	윈도-TM[5]

오류의 유형은 <표 Ⅲ.1>에 작성된 유형에 따라 다른 의미를 가진다. 유형 1은 개체명을 인식했으나 잘못된 개체명 태그를 부여한 경우이고, 유형 2는 개체명으로 인식한 단어가 개체명이 아닌 경우이다. 유형 1과 유형 2는 개체

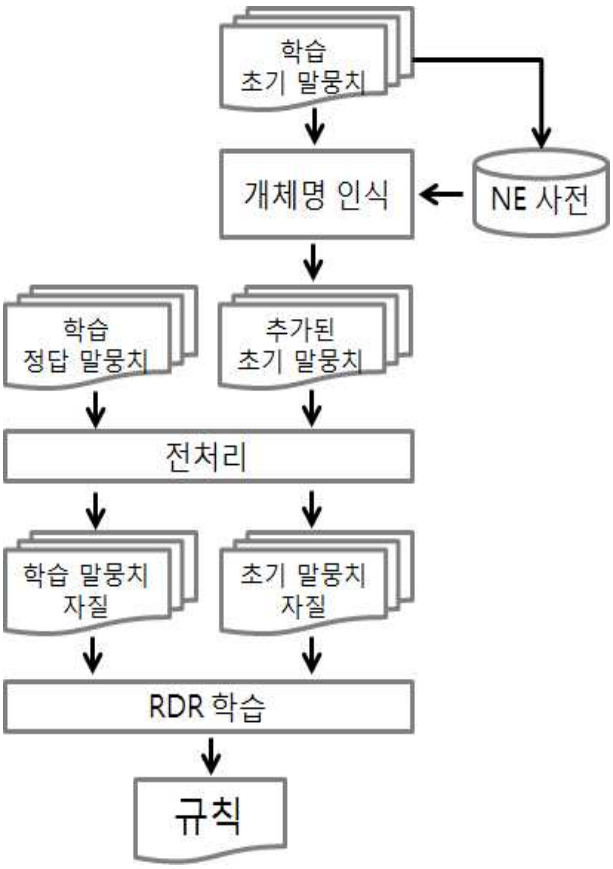
[4] 문화에 대해 정의된 개체명 태그
[5] 용어에 대해 정의된 개체명 태그

명 인식에 대한 오류보다는 분류에 대한 오류라고 볼 수 있다. 왜냐하면 'Null'도 하나의 개체명 태그라고 생각할 때, 'Null'이 아닌 다른 태그를 부착했다고 생각할 수 있기 때문이다. 이와 다르게 유형 3은 개체명인 단어를 인식하지 못한 경우이다. 또한 유형 4는 개체명의 일부분만을 인식해 잘못된 개체명이다. 그래서 유형 3과 유형 4는 개체명 인식의 문제로 볼 수 있다.

본 연구에서 제안하는 규칙을 기반으로 하는 모델에서는 개체명 인식보다는 개체명 분류의 문제에서 보다 나은 성능을 보인다. 이는 기초 실험 통해 모든 유형을 학습할 때에는 오히려 많은 오류를 발생시켰다. 따라서 유형 1과 유형 2에 대해서만 오류 수정의 대상으로 제한한다. 하지만 오류 유형 3에 대한 오류를 단순히 배제하기는 힘들다. 그래서 개체명 사전을 사용한 개체명 인식을 통해 유형 1 또는 유형 2의 문제로 변경할 수 있다. 이를 위해 개체명 사전과 확장 개체명 사전을 사용해 초기 말뭉치에 추가적인 개체명 태그 부착한다. 개체명 사전과 확장 개체명 사전은 2,1절에서 설명한다. 하지만 유형 4는 다른 유형의 오류에 비해 개체명 말뭉치에서 나타나는 빈도가 미비하고 개체명 사전으로 처리하기에는 한계성이 있어 제외한다.

2. 학습 과정

<그림 Ⅲ-1>는 제안하는 시스템의 학습과정 흐름도이다.



<그림 Ⅲ-1> 제안 시스템의 학습과정 흐름도

학습 초기 말뭉치가 입력되면 해당 말뭉치에서 나타나는 개체명을 모두 포함하는 개체명 사전을 생성한다. 이후 개체명 사전에 등록된 개체명을 사용해서 학습 초기 말뭉치에서 미인식된 개체명을 추가적으로 인식하게 된다. 이는

초기 말뭉치에서 나타나는 개체명 중 다른 문장에서는 개체명으로 인식된 단어가 말뭉치를 제작하는 과정에서 누락된 경우인 오류 유형 3에 대한 오류를 감소시킬 수 있다. 이후 학습 정답 말뭉치와 추가로 개체명을 인식한 초기 말뭉치를 사용해 전처리 과정을 진행한다. 전처리 과정에서 각 말뭉치의 자질 정보를 추출하여 FeatureSet을 생성하고 오류 유형에 대한 매핑을 진행한다. 마지막으로 매핑된 두 종류의 말뭉치를 사용하여 RDR 학습을 진행한다. 학습 결과로 두 개체명 부착 말뭉치에서 서로 다르게 부착된 개체명 태그를 정답 개체명 태그로 수정할 수 있는 규칙들을 생성한다.

2.1 개체명 사전

오류 유형3은 RDR을 이용하여 생성한 규칙으로는 수정이 용이하지 않다. 이를 보완하기 위해 개체명 사전을 이용하여 추가로 말뭉치에 개체명을 부착하는 개체명 인식 과정을 거친다. 사전은 개체명 사전과 확장 개체명 사전으로 나뉜다. <표 Ⅲ.2>는 개체명 사전을 구성하는 자질 정보들이다.

<표 Ⅲ.2> 개체명 사전을 구성하는 자질 정보

개체명	개체명 태그	개체명을 구성하는 형태소	개체명의 이전 어절	개체명의 다음 어절

입력되는 개체명 말뭉치에서 개체명으로 인식된 개체명에 대한 단어, 개체명 태그 정보와 개체명의 형태소[6] 정보, 개체명의 앞·뒤 어절[7]에서 나타나는

[6] 뜻을 가지는 가장 작은 말의 단위

주변 정보를 모두 개체명 사전에 등록한다. 확장 개체명 사전은 정규식을 사용해 패턴화가 가능한 추가 자질 정보를 포함하는 사전이다. 아래의 문장 가)와 문장 나)는 정규식을 통해 패턴화가 가능한 개체명에 대한 예이다.

가) 개체명(단어) – 마이크로소프트(microsoft)

나) 개체명'단어' – 마이크로소프트(microsoft)

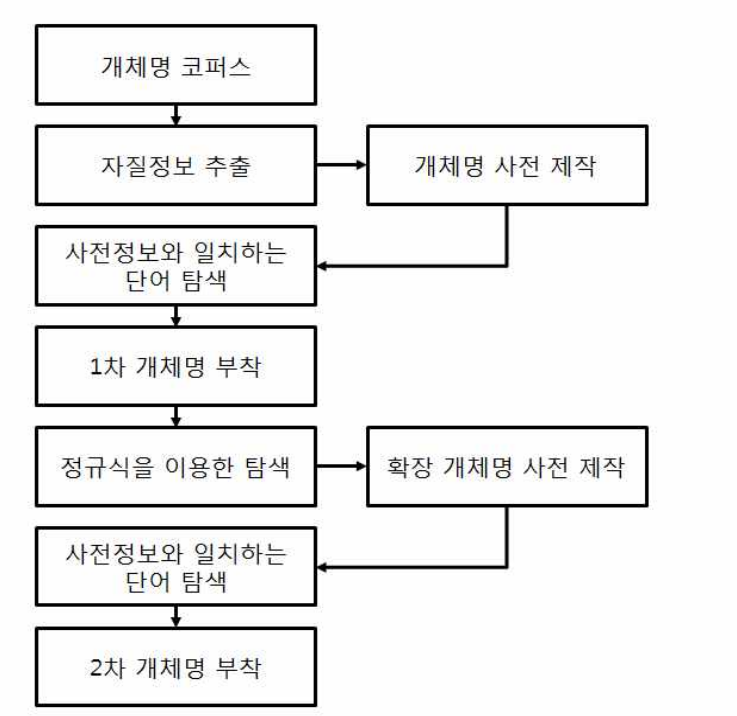
문장 가)에서 볼 수 있듯이 개체명 뒤에 오는 특수 기호로 싸인 단어에 대한 추가적인 개체명을 인식할 수 있다. <표 Ⅲ.3>은 학습 말뭉치와 평가 말뭉치에서 제작된 개체명 사전과 확장 개체명 사전에 등록된 개체명의 수이다.

<표 Ⅲ.3> 개체명 사전과 확장 개체명 사전에 등록되는 개체명 수

	학습 말뭉치 (3,500문장)	평가 말뭉치 (5,438문장)
개체명 사전	2,645개	3,761개
확장 개체명 사전	2,664개	3,795개

개체명 사전의 개체명 수와 확장 개체명 사전 개체명의 수가 약 20~30개의 개체명 수의 차이를 보인다. 이는 입력되는 말뭉치에서 정규식을 통해 패턴화 될 수 있는 개체명의 수가 상당히 제한적이기 때문이다.

[7] 문장을 구성하는 한 단위로 대체로 띄어쓰기 단위와 일치한다.



<그림 Ⅲ-2> 개체명 사전 생성 및 개체명 인식 흐름도

<그림 Ⅲ-2>는 개체명의 사전 생성 및 개체명 인식에 대한 흐름도이다. 입력되는 개체명 말뭉치을 사용해 자질 정보를 추출한다. 추출된 자질 정보를 모두 포함하여 개체명 사전을 생성한다. 생성된 개체명 사전을 사용해 입력된 개체명 말뭉치에 대해서 추가적인 개체명 인식을 진행하게 된다. 개체명의 인식은 1차와 2차로 구분되어 진행된다. 1차에서는 개체명 사전을 사용하여 개체명을 인식하고 2차에서는 확장 개체명 사전을 사용해 추가적으로 개체명을 인식한다. 개체명의 인식은 각 사전에 등록되어 있는 개체명과 동일한 형태소 및 주변 정보가 동일할 때에 개체명 태그를 부착한다.

2.2 자질 생성(FeatureSet 생성)

자질은 형태소 분석 단위로 구성되고 형태소 분석 정보, 구문 분석 정보, 개체명 분석 정보를 포두 포함한다. <표 Ⅲ.4>는 문장 다)를 FeatureSet으로 생성한 자질 정보의 예제이다.

다) 지난해 3월 이후 기가 WiFi를 구축한 서울역·김포공항·엔제리너스·롯데리아·에버랜드 등이다.

<표 Ⅲ.4> 문장 다)에 대한 초기 말뭉치의 FeatureSet 정보의 예

	1	2	3	4	5	6	7	8	9	10	11
지난해	NNG	–	–	지역	NNG	3	SN	–	–	지난해	DT_B
3	SN	–	–	지난해	NNG	이후	NNG	이후	NNG	3월	DT_B
월	NNB	–	–	지난해	NNG	이후	NNG	이후	NNG	3월	DT_I
이후	NNG	–	–	3	SN	기가	NNG	구축	NNG	–	Null
기가	NNG	–	–	이후	NNG	WiFi	SL	WiFi	SL	기가 WiFi	TR_B
WiFi	SL	를	JKO	기가	NNG	구축	NNG	구축	NNG	기가 WiFi	TR_I
를	JKO	를	JKO	기가	NNG	구축	NNG	구축	NNG	–	Null
구축	NNG	–	–	WiFi	SL	서울역	NNP	서울역	NNP	–	Null
하	XSV	–	–	WiFi	SL	서울역	NNP	서울역	NNP	–	Null
ㄴ	ETM	–	–	WiFi	SL	서울역	NNP	서울역	NNP	–	Null
서울역	NNP	–	–	구축	NNG	등	NNB	–	–	서울역	LC_B
·	SP	–	–	구축	NNG	등	NNB	–	–	–	Null
김포공항	NNP	–	–	구축	NNG	등	NNB	–	–	김포공항	LC_B
·	SP	–	–	구축	NNG	등	NNB	–	–	–	Null
엔제리너스	NNP	–	–	구축	NNG	등	NNB	–	–	엔제리너스	OGG_B
·	SP	–	–	구축	NNG	등	NNB	–	–	–	Null
롯데리아	NNP	–	–	구축	NNG	등	NNB	–	–	롯데리아	OGG_B
·	SP	–	–	구축	NNG	등	NNB	–	–	–	Null
에버랜드	NNP	–	–	구축	NNG	등	NNB	–	–	에버랜드	OGG_B
등	NNB	–	–	서울역	NNP	–	–	–	–	–	Null
이	VCP	–	–	서울역	NNP	–	–	–	–	–	Null
다	EF	–	–	서울역	NNP	–	–	–	–	–	Null
.	SF	–	–	서울역	NNP	–	–	–	–	–	Null

자질 1은 형태소의 품사, 자질 2, 3는 해당 형태소가 개체명일 때 동일한 어절에 위치한 조사와 품사, 자질 4, 5는 이전 어절의 형태소와 품사, 자질 6, 7은 다음 어절의 형태소와 품사, 자질 8, 9는 해당 형태소의 지배소 형태소와 품사, 자질 10은 현재 형태소의 개체명 정보이다. 그리고 개체명 정보의 자질은 개체명 태그와 개체명 경계 정보인 B, I를 결합하여 사용한다.

<표 Ⅲ.4>에서 열한 번째 형태소인 '서울역'의 경우는 자신의 품사가 'NNP'[8]이고, 자기 자신을 포함하는 어절에는 조사 정보가 없다. 주변 정보로는 앞 어절의 형태소가 '구축', 품사가 'NNG'[9]이고, 뒷 어절의 형태소는 '등', 품사가 'NNB'[10] 이다. 그리고 개체명은 '서울역'이고, 개체명 태그는 'LC_B'[11]이다.

2.3 오류 유형 매핑

오류 유형 매핑은 전처리 과정 중에 진행된다. 이는 1절에서 설명한 바와 같이 오류 유형 1과 유형 2를 학습에 사용되는 규칙 생성 대상으로 하기 위해 진행된다. 개체명 사전을 사용해 추가적으로 개체명을 인식한 초기 말뭉치를 기준으로 형태소 단위의 개체명 자질 정보가 있는 형태소를 매핑하게 된다. <표 Ⅲ.5>은 문장 가)에 대한 FeatureSet을 매핑한 예제이다.

---

[8] 고유 명사 : 한국, 서울, 삼성 등
[9] 일반 명사 : 사전, 사과 등의 사물의 이름을 나타내는 단어
[10] 의존 명사 : 원, 달러 등의 자립해서 쓰일 수 없는 명사
[11] 지명에 대해 정의된 개체명 태그 LC와 개체명 경계 정보 (B: 개체명의 시작, I: 개체명의 연속)를 결합한 태그 정보를 뜻한다.

<표 Ⅲ.5> 문장 다)에 대한 초기 말뭉치의 FeatureSet을 매핑한 예

	1	2	3	4	5	6	7	8	9	10	11
지난해	NNG	–	–	지역	NNG	3	SN	–	–	지난해	DT_B
3	SN	–	–	지난해	NNG	이후	NNG	이후	NNG	3월	DT_B
월	NNB	–	–	지난해	NNG	이후	NNG	이후	NNG	3월	DT_I
기가	NNG	–	–	이후	NNG	WiFi	SL	WiFi	SL	기가 WiFi	TR_B
WiFi	SL	를	JKO	기가	NNG	구축	NNG	구축	NNG	기가 WiFi	TR_I
서울역	NNP	–	–	구축	NNG	등	NNB	–	–	서울역	LC_B
김포공항	NNP	–	–	구축	NNG	등	NNB	–	–	김포공항	LC_B
엔제 리너스	NNP	–	–	구축	NNG	등	NNB	–	–	엔제 리너스	OGG_B
롯데리아	NNP	–	–	구축	NNG	등	NNB	–	–	롯데리아	OGG_B
에버랜드	NNP	–	–	구축	NNG	등	NNB	–	–	에버랜드	OGG_B

<표 Ⅲ.5>는 <표 Ⅲ.4>와 다르게 열한 번째 자질 정보인 개체명 태그 정보가 'Null'인 형태소는 제외된다. 이는 개체명이 아닌 형태소를 제외함으로써 개체명으로 분류되는 형태소만을 규칙 생성에 사용할 수 있다.

2.4 규칙 생성

규칙을 생성하기 위해 사용된 모델은 RDRPOSTagger[20]를 수정하여 규칙 생성에 사용하였다. RDRPOSTagger는 초기 말뭉치와 정답 말뭉치를 비교하여 같은 형태소에 다른 품사가 부착된 경우, 이를 학습하여 잘못된 초기 말뭉치의 품사를 정답으로 수정할 수 있는 규칙을 생성한다. 이 모델에서는 오직 품사가 다른 형태소에 대해서만 학습하게 된다. 하지만 수정한 모델에서는 품사가 다른 경우와 같은 경우를 모두 학습하게 된다. 이는 기존의 모델에서는 형태소 태깅을 목적으로 하였지만 본 논문에서는 개체명 수정을 목적으로

하기 때문이다. 형태소를 부착하는 것보다 개체명을 부착하는 것이 더 많은 자질 정보를 사용하기 때문이다. 품사가 다른 경우만을 학습하였을 경우 기초 실험에서 더 많은 오류를 만들어내는 규칙을 생성하는 결과가 나타났다. 따라서 정확한 규칙을 생성하기 위해 초기 말뭉치와 정답 말뭉치에서 나타나는 모든 개체명을 형태소 단위로 학습하였다.

또한 규칙을 생성하는데 있어서 자질 정보를 효과적으로 사용하기 위해 자질 템플릿을 입력으로 사용하게 수정하였다. 이는 사용자가 템플릿을 수정함으로써 자질 정보를 보다 효과적으로 사용하게 할 수 있다. 따라서 규칙 생성에 있어 보다 정확한 규칙을 생성할 수 있게 한다.

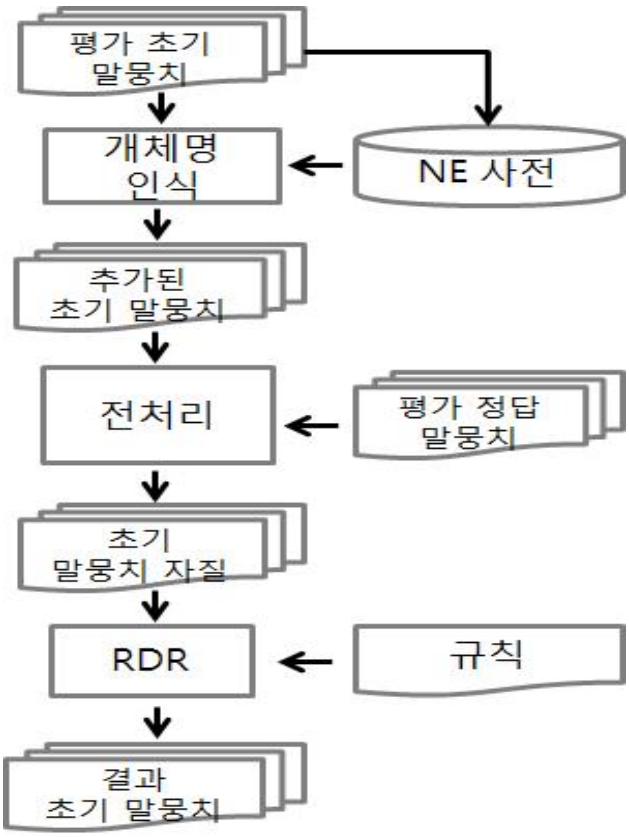
문장 라)는 자질 템플릿을 사용했을 때 생성될 수 있는 규칙에 대한 예제이다.

라) U02 : object[0][11] == "에버랜드" : object.conclusion = "LC_B"

문장 라)는 현재 형태소의 개체명이 '에버랜드'일 때 개체명 태그를 'LC_B'로 변경하라는 규칙을 말한다.

3. 개체명 오류 수정 과정

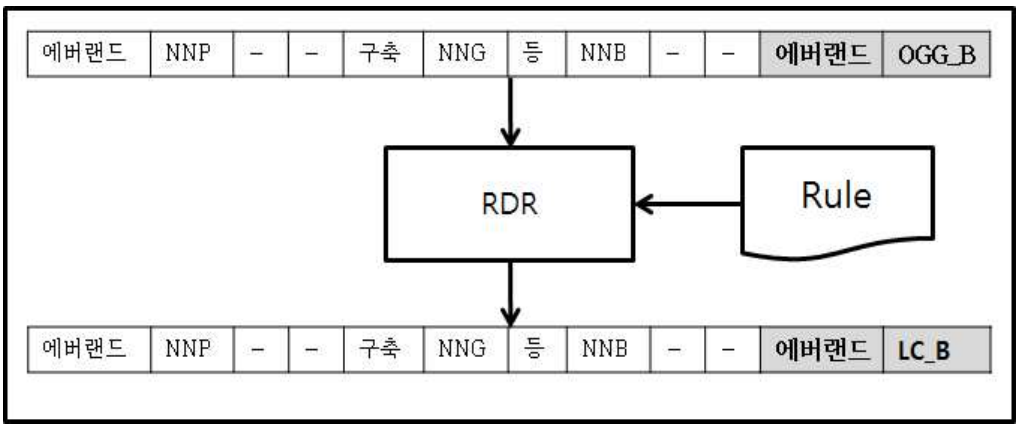
<그림 Ⅲ-3>은 개체명 오류 수정 과정의 흐름도이다.



<그림 Ⅲ-3> 개체명 오류 수정 과정의 흐름도

오류 수정 과정은 학습 과정과 유사하다. 입력되는 평가 초기 말뭉치로부터 개체명 사전을 생성하고 개체명 사전을 사용해 초기 말뭉치의 개체명을 추가로 인식한다. 추가로 개체명을 인식한 초기 말뭉치와 정답 말뭉치를 전처리

과정 중에 FeatureSet을 생성하고 개체명이 포함된 형태소만을 매핑한다. 이는 앞에서도 설명한 바와 같이 오류 유형을 제한하기 위해서 진행된다. 매핑된 결과 말뭉치와 학습 과정에서 생성된 규칙을 사용해 RDR 시스템을 적용하여 초기 말뭉치의 개체명 태그가 수정된 시스템 결과를 출력하게 된다.



<그림 Ⅲ-4> 개체명 오류 수정 예

<그림 Ⅲ-4>는 입력되는 초기 말뭉치를 FeatureSet으로 생성한 결과를 2.4절의 문장 나)의 규칙에 적용되었을 때의 개체명 태그가 수정된 결과 이다. 형태소 '에버랜드'의 개체명 태그 'OGG_B'가 규칙에 따라 'LC_B'로 수정되었다.

제 Ⅳ 장

실험 및 토의

1. 실험 환경

본 연구에서 사용하는 개체명 부착 말뭉치는 15개의 카테고리를 정의한 개체명 태그를 사용한다. 실험을 위해 사용된 말뭉치는 블로그에서 텍스트를 추출에 직접 제작한 8,938문장의 개체명 말뭉치를 학습 3,500문장과 평가 5,438문장으로 나누어 사용하였다. 또한 규칙 학습을 위해 규칙 선택 임계값(Threshold)는 '1'을 사용한다. 규칙 선택 임계값는 규칙을 생성하기 위한 변수값으로 동일한 규칙이 몇 번 반복되어 나타나는가에 따라 규칙을 생성하는 값이다. 이러한 규칙 선택 임계값를 '1'로 선택된 이유는 개체명이 나타나는 빈도가 높지 않으므로 보다 많은 규칙을 생성하기 위해서다. 기본 실험을 진행하여 최적의 템플릿을 구하고 학습을 진행하였다. 실험은 소량의 문서로 학

습하여 대량의 평가 말뭉치를 분석한다. 사용한 개체명 말뭉치에 대한 개체명 통계와 유형별 오류 개체명 수 및 오류율은 <표 Ⅳ.1>과 같다.

<표 Ⅳ.1> 블로그 문서의 개체명 수와 오류율

		학습문서		평가문서	
		초기	정답	초기	정답
문장 수		3,500	3,500	5,438	5,438
총 개체명 수		6,538	6,531	10,447	10,396
정답 개체명 수		6,222	6,531	9,951	10,396
오류 개체명 수 (비율)	유형1	100 (21%)		128 (18%)	
	유형2	168 (35%)		276 (38%)	
	유형3	161 (34%)		225 (31%)	
	유형4	48 (10%)		92 (13%)	
오류율 (단위 : %)		7.12		6.76	

제안된 시스템의 성능을 평가하기 위해 사용된 오류율을 계산하기 위한 수식은 식(2)을 통해 계산하였다.

$$\text{오류율} = (1 - \frac{\text{정답 개체명 수}}{\text{총 개체명 수} + \text{오류 유형 3의 개체명 수}}) \times 100 \quad (2)$$

오류율을 계산하는 예제로 초기 말뭉치의 정답 개체명의 수가 6,222개이고

정답 말뭉치의 개체명 수는 6,531개이다. 오류 유형 3의 개체명 수는 161개일 때 식(2)를 사용하여 계산된 오류율은 7.12%이다.

2. 실험 결과 분석

실험은 최소한의 학습 데이터로 더 큰 데이터를 평가하기 위해 작은 양의 문장을 학습하고 많은 양의 문장을 평가하였다. 하지만 작은 양의 문장에 나타나는 개체명은 빈도가 낮기 때문에 규칙 선택 임계값(Threshold)를 1로 설정하여 학습하였다. 규칙 선택 임계값를 높게 설정할 경우 생성되는 규칙이 적어 수정되는 개체명이 거의 나타나지 않았다. 또한 낮은 규칙 선택 임계값는 오류만 학습하는 기존의 RDR 시스템[20]에서는 오히려 오류를 발생시키는 규칙도 생성하는 문제점이 나타났다. 이를 해결하기 위해서 오류와 정답을 모두 학습해 규칙을 생성되게 하였다.

<표 Ⅳ.2> 개체명 사전을 이용한 개체명 부착 성능

		초기	개체명 사전 적용 후	확장 개체명 사전 적용 후
총 개체명 수 (증감)		10,447	10,614 (+167)	10,652 (+205)
정답 개체명 수 (증감)		9,951	10,002 (+51)	10,005 (+54)
오류 개체명 수 (증감)	유형1	128	128 (+0)	136 (+8)
	유형2	276	390 (+114)	416 (+140)
	유형3	225	172 (−53)	159 (−66)
	유형4	92	94 (+2)	95 (+3)
오류율 (단위 : %)		6.76	7.27	7.46

<표 Ⅳ.2>는 사전을 이용한 개체명 부착 방법에 대한 성능을 분석한 결과이다. 이 결과는 <표 Ⅳ.1>의 평가 문서에 대한 통계 데이터를 기반으로 개체명 사전을 사용한 개채명 인식과 확장 개체명 사전을 사용한 개체명 인식을 비교하였다. 사전을 사용한 개체명 인식에서는 오류 유형 3의 수가 초기 오류의 수에 비해 각각 53개, 66개가 감소되었다. 이는 개체명 인식의 문제로 분류한 오류 유형 3을 개체명 사전을 사용했을 때에 보완이 가능하다고 분석할 수 있다. 또한 정답 개체명의 수도 증가됨을 확인할 수 있다.

하지만 오류 유형 3을 감소시킨 개체명과 정답으로 인식된 개체명을 합한 수보다 오류 유형 2에 대한 개체명 수가 더 많이 증가함으로써 오류율은 0.5% 이상 증가되었다. 이는 개체명이 아닌 형태소를 개체명으로 인식함으로써 발생한다. 이러한 오류는 개체명 사전에 등록된 개체명이 초기 말뭉치로부

터 생성되기 때문에 초기 말뭉치가 가지고 있는 오류 개체명을 개체명으로 인식하고 사전에 등록함으로써 발생했다. 그리고 확장 개체명 사전에 등록된 개체명 수가 개체명 사전에 등록된 개체명 수보다 많으므로 오류 유형 3의 수를 더 많이 증가시켰다.

<표 Ⅳ.3> RDR 시스템 성능

		초기	RDR 적용		
			사전 미사용	개체명 사전	확장 개체명 사전
총 개체명 수 (증감)		10,447	10,230 (−217)	10,350 (−97)	10,422 (−25)
정답 개체명 수 (증감)		9,951	9,987 (+36)	10,037 (+86)	10,041 (+90)
오류 개체명 수 (증감)	유형1	128	92 (−36)	93 (−35)	100 (−28)
	유형2	276	59 (−217)	160 (−116)	186 (−90)
	유형3	225	225 (0)	172 (−53)	159 (−66)
	유형4	92	92 (0)	94 (+2)	95 (+3)
오류율 (단위 : %)		6.76	4.47	4.61	5.10

<표 Ⅳ.3>은 RDR 시스템 성능을 나타낸 표이다. 성능 비교는 사전을 사용하지 않고 RDR 시스템을 사용해 오류를 수정한 성능과 개체명 사전 및 확장 개체명 사전을 사용하였을 때의 성능을 각각 비교하여 작성하였다.

오류 수정 과정을 통해 오류 유형 1과 유형 2에 대한 오류의 수는 모두 감소됨으로써 오류율은 각각 4.49%, 4.61%, 5.10%로 감소되었다. 또한 유형 2에 대한 개체명 수가 상대적으로 많이 감소되어 총 개체명의 수는 감소되었지만 유형 2가 잘못 인식된 개체명에 대한 오류임으로 긍정적인 결과이다. 하지

만 오류 유형 1이 유형 2에 비해 작은 비율로 수정되었다. 이는 사전을 사용했을 때의 정답 개체명과 사용하지 않았을 때의 정답 개체명수를 비교했을 때 정답 개체명의 증가된 수가 2배 이상 차이가 난다. 또한 오류 2의 경우에도 사전을 사용하지 않은 경우 217개의 오류 개체명이 삭제되었지만 사전을 사용한 경우에는 감소된 개체명 수가 100여개가 작다. 이는 개체명 사전을 사용했을 때 추가된 오류 유형 2의 개체명이 100여개임을 보았을 때 실질적으로 추가된 개체명의 수정이 제대로 이루어지지 않았음을 알 수 있다. 이 결과로 학습 과정에서 생성된 규칙이 평가 말뭉치의 개체명과 다른 개체명이거나 다른 주변 정보를 가지고 있기 때문이라고 분석된다. 따라서 사전을 사용하지 않았을 때의 오류율이 6.76%에서 4.49%로 2.27%가 감소되었으므로 가장 뛰어난 성능이다.

제 Ⅵ 장

결론 및 향후연구

본 연구에서는 개체명 태그 부착 말뭉치의 오류를 자동으로 수정하는 방법을 제시하였다. 이를 위해 개체명 오류의 유형을 분석하고 개체명 분류(오류 유형 1, 유형 2)에 대한 문제와 개체명 인식(오류 유형 3, 유형 4)에 대한 문제로 분류하여 실험을 진행하였다. 개체명 오류 수정은 개체명 분류의 문제로 제한한 오류 유형 1과 유형 2에 대해서만 진행하였다. 이는 수정한 RDR시스템이 개체명 분류에 대해서만 뛰어난 성능을 보였기 때문이다. 하지만 개체명 사전을 사용해 초기 말뭉치에 추가로 개체명 인식을 진행하는 방법으로 개체명 인식에 대한 부분도 보완하였다.

성능은 3,500문장 학습하고 5,438문장으로 평가하여 2.27%의 오류율이 감소하였다. 학습하는 문장에 따라 규칙을 생성하는 시스템의 특성 때문에 학습 말뭉치의 영향이 매우 크게 작용하였다. 이는 학습을 위해 사용되는

학습 말뭉치의 크기가 커야 오류를 수정할 수 있는 다양한 규칙이 생성되어 개체명 사전을 사용하였을 때에도 오류율은 감소될 수 있다.

본 논문에서 개체명 부착 말뭉치를 수정하기 위해 사용된 주변 자질 정보 이외의 추가적인 자질 정보에 대한 연구가 진행되어야 한다. 또한 개체명 사전을 사용한 개체명 인식의 정확도를 향상시킬 수 있는 방법에 대한 연구도 진행되어야 한다.

참 고 문 헌

[1] Edwards G, and Compton P, “Peirs : A pathologist maintained expert system for the interpretation if chemical pathology reports”, Pathology 25:27-34, 1993.

[2] Wu X, “Knowledge acquisition from database”, Ablex Publishing Corp USA, 1995.

[3] Zhu X, Wu X, and Chen Q, “Eliminating Class Noise in Large Datasets”, Proceedings of the 20th ICML International Conference on Machine Learning(ICML 2003), Washington DC, pp.920-927, 2003.

[4] Ellen Riloff, “Automatically generating extraction patterns from untagged text”, Proceedings of the thirteenth national conference on Artificial intelligence(AAAI-96), Vol.2, pp.1044-1049, 1996.

[5] Edwards G, and Compton P, “Experience With Ripple-Down Rules”, AI 2005 SI, Vol.19, pp.356-362, 2006.

[6] Janyce M Wiebe, Rebecca F Bruce, and Thomas P O'Hara, “Development and use of a gold-standard data set for subjectivity classifications”, Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, pp.246-253, 1999.

[7] 안주희, 이승우, 이근배, “웹을 이용한 개체명 부착 말뭉치의 자동생성과 정제”, 한국정보과학회, 제14회 한글 및 한국어 정보처리 학술대회, pp.85-91, 2002.

[8] 박용민, 이재성, "한국어 제목 개체명 인식 및 사전 구축: 도서, 영화, 음악, TV프로그램", 한국정보처리학회, 정보처리학회논문지, 소프트웨어 및 데이터

공학, Vol.3, No.7, pp.285-292, 2014.

[9] 함영균, 위키피디아 스케일의 디비피디아 온톨로지 기반 개체명 코퍼스 구축 방법 연구, 한국과학기술원, 석사학위 논문, 2014.

[10] 최지예, 김명근, 박소영, "문화유산정보 말뭉치 구축을 위한 개체명 및 이벤트 부착 도구", 한국컴퓨터정보학회, Journal of The Korea Society of Computer and Information. Vol.17, No.9, 2012.

[11] 최동현, 김은경, 고은비, 최기선, "COAT: 시멘틱 어노테이션 말뭉치 구축 지원 도구", 한국정보과학회, 한글 및 한국어 정보처리 학술대회 Poster, 2011.

[12] Daniel M Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel, "Nymble: a High-Performance Learning Name-finder", Proceedings of the 5th Conference on Applied Natural Language Processing, pp.194-201, 1997.

[13] Xiaohua Liu, Ming Zhou, Furu Wei, Zhongyang Fu, and Xiangyang Zhou, "Joint Inference of Named Entity Recognition and Normalization for Tweets", Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp.526-535, 2012.

[14] Sungchul Kim, Kristina Toutanova, and Hwanjo Yu, "Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia", Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp.694-702, 2012.

[15] 황이규, 유보현, "HMM에 기반한 한국어 개체명 인식", 한국정보처리학회, 정보처리학회논문지B, Vol.10, No.2, pp.229-236, 2003.

[16] 이창기, 황이규, 오효정, 임수종, 허정, 이충희, 김현지, 왕지현, 장명길, "Conditional Ranfom Fields를 이용한 세부 분류 개체명 인식", 한국정보과

학회, 제18회 한글 및 한국어 정보처리 학술대회, pp.268-272, 2006.

[17] 이창기, 장명길, "Structural SVMs 및 Pegasos 알고리즘을 이용한 한국어 개체명 인식", 한국인지과학회, 인지과학 Vol.21, No.4, pp.655-667, 2010.

[18] 김주근, 반지도 학습법을 이용한 한국어 개체명 인식, 창원대학교 석사학위 논문, 2013.

[19] Cao T.M, and Compton P.A, "Simulation Framework for Knowledge Acquisition Evaluation", 28th Australasian Computer Science Conference ACSC2005, pp.353-360, 2005.

[20] Nguyen D.Q, Nguyen D.Q, Pham D.D, and Pham S.B, "RDRPOSTagger: A Ripple Down Rules-based Part-Of-Speech Tagger", EACL'14, pp.17-20, 2014.

[21] Eric Brill, "Transformation-based error-driven learning and natural language processing : a case study in part-of-speech tagging", Comput Linguist, pp.543-565, 1995.

ABSTRACT

Automatic Named-Entity Error Correction

by JungHan Kim

Dept. of Computer Engineering
Graduate School, Changwon National University

Learning corpus of natural language processing is extremely important.

In this paper, the RDR (Ripple-Down Rules) proposes a method to correct the errors of the Named-Entity tag. Generating a rule in order to use the feature for extand the RDR was extended to enable the template.

It collects the Blog document in learning and using the prepared test corpus, and A small amount of the document and study tested the effectiveness and efficiency of error correction in general situations by evaluating a large number of documents.

부록 A. 형태소 품사 집합

TAG	POS	TAG	POS
NNG	일반명사	IC	감탄사
NNB	의존명사	VCP	긍정지정사
NNP	고유명사	VCN	부정지정사
NP	대명사	VV	동사
NR	수사	VA	형용사
JKS	주격조사	VX	보조용언
JKC	보격조사	EF	종결어미
JKO	목적격조사	EC	연결어미
JKG	관형격조사	ETN	명사형 전성어미
JKB	부사격조사	ETM	관형형전성어미
JKV	호격조사	EP	선어말어미
JKQ	인용격조사	SF	마침표, 물음표, 느낌표
JC	접속조사	SP	쉼표, 가운뎃점, 콜론, 빗금
JX	보조사	SS	따옴표, 괄호표, 줄표
XPN	명사접두사	SE	줄임표
XSN	명사파생접미사	SO	붙임표(물결, 숨김, 빠짐)
XSB	부사파생접미사	SL	외국어
XSV	동사파생접미사	SH	한자
XSA	형용사파생접미사	SN	숫자
XR	어근	NF	명사추정범주
MM	관형사	NV	용언추정범주
MAG	일반부사	SW	기타기호
MAJ	접속부사	NA	분석불능범주

부록 B. 개체명 태그 집합

TAG	용어	설명
PS	PERSON	인명 / 인명의 별칭
FD	STUDY_FIELD	학문 분야
TR	THEORY	특정 이론, 법칙, 원리 등의 명칭
AF	ARTIFACTS	인공물
OGG	ORGANIZATION	기관 / 단체 명칭
LC	LOCATION	지역명
CV	CIVILIZATION	문명 / 문화에 관련된 용어
DT	DATE	날짜
TI	TIME	시간
QT	QUANTITY	수량 표현
EV	EVENT	특정 사건 / 사고 명칭
AM	ANIMAL	동물 명칭
PT	PLANT	식물 명칭
MT	MATERIAL	물질 명칭
TM	TERM	용어

이 력 서

성　　명: 김 중 한
생년월일: 1986년 08월 23일
출 생 지: 경상남도 마산시
주　　소: 경상남도 창원시 마산회원구 석전2동 294-22번지

학 력

2005-2013: 창원대학교 공과대학 컴퓨터공학과(B.S.)
2013-2015: 창원대학교 대학원 컴퓨터공학과(M.S.)

발표논문