

의미그룹을 이용한 단어 중의성 해소 (Word Sense Disambiguation using Meaning Groups)

김은진[†] 이수원^{**}
(EunJin Kim) (SooWon Lee)

요약 본 연구에서는 기계 가독형 사전 정보를 이용하여 중의어 의미 태깅을 위한 학습에 사용될 양질의 의미부착 데이터를 자동 생성하여 중의어 태깅 정확도를 향상시키는 방법을 제안한다. 또한 중의어의 의미별로 특징있는 주변 단어를 원소로 갖는 의미그룹을 생성하고, 의미그룹 내 중요 자질의 가중치 선정에 집중도 개념을 도입하여 태깅 정확도를 높인다. SENSEVAL-2 데이터를 사용하여 태깅 정확도를 평가한 결과 본 연구에서 제안한 방법이 기존의 방법보다 정확도를 향상시킨 것으로 나타났다.

키워드 : 단어 중의성 해소, 자동 의미 태깅, 의미그룹

Abstract This paper proposes the method that increases the accuracy for tagging word meaning by creating sense tagged data automatically using machine readable dictionaries. The concept of meaning group is applied here, where the meaning group for each meaning of a target word consists of neighbor words of the target word. To enhance the tagging accuracy, the notion of concentration is used for the weight of each word in a meaning group. The tagging result in SENSEVAL-2 data shows that accuracy of the proposed method is better than that of existing ones.

Key words : Word Sense Disambiguation, Automatic Sense Tagging, Meaning Group

· 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음
· 이 논문은 제36회 추계학술발표회에서 '의미그룹을 이용한 단어 중의성 해소'의 제목으로 발표된 논문을 확장한 것임
† 학생회원 : 숭실대학교 컴퓨터학부
rlwjr27@naver.com
****** 종신회원 : 숭실대학교 컴퓨터학부 교수
swlee@ssu.ac.kr
논문접수 : 2009년 12월 24일
심사완료 : 2010년 3월 25일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제6호(2010.6)

1. 서론

단어의 중의성 해소 문제는 중의성을 가지는 단어의 정확한 의미를 밝히는 문제로서 자연언어 처리 및 검색 분야에서 계속 연구되어온 난제이다. 특히 '지능형 웹'이라 불리는 시맨틱 웹이 각광을 받으면서 시맨틱 웹을 위한 온톨로지 매핑에서의 단어 중의성 문제가 화두로 떠오르고 있다.

본 연구에서는 기존의 단어 중의성 해소 연구가 가지고 있는 문제점을 해결하여 중의어 태깅 정확도를 향상시키는 방법을 제안하고자 한다. 대표적인 방법인 지도식 학습 방법은 태깅 성능이 뛰어나지만 양질의 학습 데이터 확보가 어렵다는 문제점을 가지고 있으며, 수작업으로 학습 데이터를 생성하는 작업은 많은 비용이 소요되는 작업이다[1]. 또한 사전에 생성된 학습 데이터를 이용하는 경우 역시 학습 데이터의 양과 질에 제한적이다[2]. 본 연구에서는 기계가독형 사전 정보를 이용하여 웹 문서 내 중의어 의미를 자동 태깅하여 학습 데이터를 자동으로 생성하는 방법을 제안한다.

단어 중의성 해소를 위한 기존의 많은 연구에서는 중의어의 문맥 내 주변단어를 중의어 의미 태깅을 위한 학습 자료로 선정한다. 그러나, 자질의 가중치를 선정하는 경우 해당 자질이 다른 의미들에 비하여 특정 의미를 위해 쓰인 상대적인 중요도를 반영하지 않는다[3]. 즉, 자질이 중의어의 다른 의미일 경우에 많이 사용됨에도 불구하고 특정 의미에서 높은 가중치를 가질 수 있다.

이러한 문제점을 해결하기 위해 본 연구에서는 중의어의 의미별로 특징있는 주변 단어로 '의미그룹'이라는 학습 모델을 생성하여 의미태깅에 활용한다. 의미그룹 내 단어의 가중치를 결정하는 경우 집중도라는 개념을 이용하여 타 의미그룹에 잘 등장하지 않으면서 특정 의미그룹에 자주 등장하는 정도를 반영한다. 본 연구에서는 이러한 집중도 개념을 적용하여 의미그룹 내 중요자질을 선정하고 단어의 가중치를 계산한다.

본 연구의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고, 3장에서는 본 연구에서 제안하는 중의성 해소 방법을 소개하며, 4장에서는 3장에서 제안하는 방법의 성능을 평가한다. 마지막으로 5장에서는 결론 및 향후 연구를 제시한다.

2. 관련 연구

중의어 태깅을 위한 학습 방법은 크게 세 가지 방법으로 분류될 수 있다. 첫 번째 방법은 지도식 학습 방법이고, 두 번째 방법은 비지도식 학습 방법이며, 세 번째 방법은 두 방법의 중간 형태인 준지도식 학습방법이다.

2.1 지도식 학습 방법

지도식 학습 방법[1,2]은 의미부착 데이터(학습 데이터)가 존재하여야 하며 이 데이터를 중의어 의미 태깅을 위한 학습에 사용한다. 이 방법은 학습 데이터로부터 중의어의 의미별로 문맥 상 주변 단어 또는 문법적인 정보(품사, 어미 등)를 추출하여 중의어 의미 태깅을 위한 학습 모델을 생성하며, 나이브 베이지안 등의 학습방법이 대표적이다[3]. 지도식 학습 방법은 상대적으로 높은 성능을 보이지만 의미부착 데이터를 생성하는 비용이 발생한다.

2.2 비지도식 학습 방법

비지도식 학습 방법은 사전에 의미부착한 데이터를 필요로 하지 않는다. 이 방법은 코퍼스 내 문장 또는 문단에서의 단어의 공기정보를 이용하여 중의어의 각 의미를 나타낼 만한 클러스터를 자동으로 생성하여 중의어 의미 태깅에 사용한다[4,5].

하지만 비지도식 학습 방법은 지도식 학습 방법에 비해 중의어 의미 태깅 성능이 좋지 않은 것으로 알려져 있다. 또한 비지도식 학습 방법으로 생성된 클러스터의 개수가 곧 중의어의 의미항목의 개수가 되는데 클러스터를 생성할 때 사용하는 임계치에 따라 생성되는 클러스터의 개수가 달라지므로 의미항목의 개수의 일관성이 부족하다.

2.3 준지도식 학습 방법

준지도식 학습 방법은 지도식 학습방법과 비지도식 학습방법의 중간 형태이다. 이 방법에서는 미리 정의된 소량의 의미 소스와 의미부착이 되지 않은 데이터를 중의어 의미 태깅을 위한 학습에 사용한다. 의미 소스로부터 추출된 의미별 정보와 중의어를 포함한 문맥 정보와의 유사도 비교를 통해 중의어의 의미를 태깅하거나 의미별 정보를 이용해 의미부착 되지않은 코퍼스에서 추가적인 정보를 추출하여 중의어의 의미 태깅을 위한 학습에 활용한다. 준지도식 학습 방법에서는 의미 소스로서 시소러스나 어휘사전을 가장 많이 사용한다[6,7].

[7]에서는 사전의 뜻풀이 단어와 문맥 단어 간 코퍼스 내 평균 상호 정보량을 이용한다. 하지만 [7] 또한 별도의 의미부착 데이터를 학습에 이용하지 않으므로 사전의 뜻풀이 양에 태깅 성능이 제한적이다.

3. 의미그룹을 이용한 단어 중의성 해소

본 연구에서는 기존 단어 중의성 해소 연구의 이슈 중 하나인 학습 데이터 부족 문제를 해결하기 위해 의미소스로 기계 가독형 사전을 이용하여 의미부착 되어 있지 않은 웹 문서를 의미부착 데이터로 자동 변환하는 방법을 제안한다. 본 연구에서는 웹이라는 언어자원이 풍부한 공간으로부터 의미부착 데이터를 자동 생성하여

학습 데이터 부족 문제를 해결한다는 점에서 기존의 연구와 차별화된다.

본 연구에서는 기존의 중의성 해소 연구들과 마찬가지로 ‘중의어는 의미별로 문맥상의 특징있는 주변 단어를 가질 것이다’라는 가정으로부터 시작한다. 그리고 중의어의 의미별 문맥상의 특징있는 주변 단어를 ‘의미별 연관단어’로 정의한다. 또한 의미별 연관단어 집합을 ‘의미그룹’이라 정의한다. 의미그룹의 개수는 중의어의 의미 개수와 동일하며 의미별 연관단어는 의미그룹 내에서 가중치를 갖는다.

본 연구에서는 기계 가독형 사전 정보로부터 추출한 단어를 이용하여 웹 문서 문맥을 태깅하여 의미부착 데이터를 자동으로 생성한다. 이 때 특정 중의어에 대하여 의미 부착 데이터를 생성하기 위해 사전에서 의미별로 추출한 단어들을 ‘seed’로 정의한다. 의미별로 모아놓은 seed의 집합을 ‘기초 의미그룹’이라 정의하며 seed는 기초 의미그룹 내에서 가중치를 갖는다.

그림 1은 본 연구의 단어 중의성 해소 시스템을 나타낸 것이다. 본 연구에서는 먼저 사전으로부터 색인어를 추출한 후 불용어를 필터링하여 seed를 원소로 갖는 기초 의미그룹을 생성한다. 그 후 추출한 웹 문서 문맥에 대해 전 단계에서 생성된 기초 의미그룹 내 seed를 이용하여 문맥 내 중의어의 의미를 부착하여 의미부착 데이터를 생성한다. 생성된 의미부착 데이터로부터 의미그룹 내 의미별 연관 단어의 가중치를 결정하여 최종 의미그룹을 구성한다. 최종 의미그룹을 생성한 후 중요 자질을 선정하여 중의어의 최종 의미를 태깅한다.

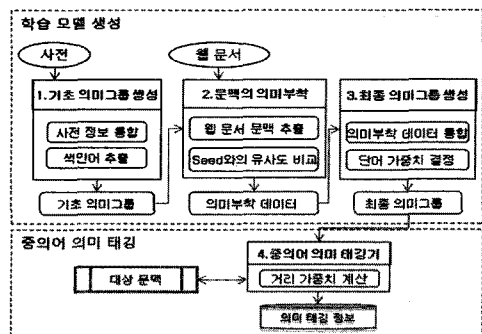


그림 1 의미그룹을 이용한 단어 중의성 해소 시스템

3.1 기초 의미그룹 생성

기초 의미그룹은 다음 단계에서 의미부착 데이터를 생성을 위한 seed를 원소로 갖는다. 기초 의미그룹을 생성 단계에서는 불용어 필터링을 통해 seed를 선별해내고 단일 사전이 아닌 여러 기계 가독형 사전을 통해 중요한 seed의 가중치를 높인다.

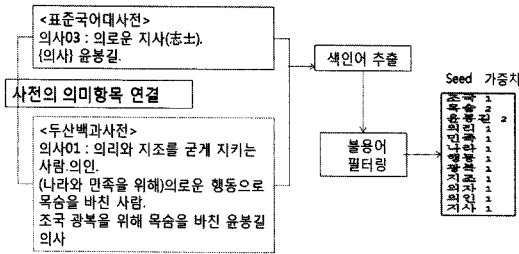


그림 2 기초 의미그룹 생성 과정의 예

그림 2는 '의사'라는 중의어에 대한 기초 의미그룹 생성 과정을 나타낸 것이다. 본 연구에서는 기계 가독형 사전으로 표준국어대사전과 두산백과사전을 사용하며 최종 의미 태깅에 적용하는 의미구분체계는 가장 의미항목이 많은 표준국어대사전을 따른다. 먼저 여러 기계 가독형 사전의 의미항목을 수작업으로 연결한 후 통합한 사전의 뜻풀이와 예문, 유의어 등의 사전 정보로부터 색인어를 자동으로 추출한다. 본 연구에서는 불용어 리스트를 여러 개의 대상 중의어들로 검색된 문서그룹에 등장하는 고빈도 단어들로 자동 선정한 불용어를 필터링한다.

3.2 웹 문서 문맥의 의미 부착

웹 문서 문맥의 의미 부착 단계에서는 전 단계에서 생성된 기초의미그룹 내 seed를 사용하여 웹 문서 내 중의어의 의미를 태깅하여 의미부착 데이터를 자동 생성한다.

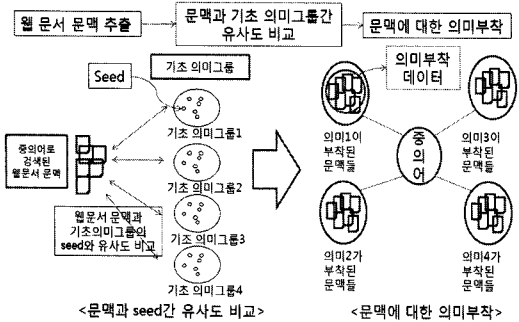


그림 3 문맥의 의미부착

그림 3은 웹 문서 문맥의 의미부착 단계를 도식화한 것이다. 먼저 중의어로 검색된 웹 문서로부터 중의어를 포함한 문맥을 추출한다. 본 연구에서는 추출한 웹 문서 문맥에 대하여 전 단계에서 생성된 기초 의미그룹의 seed와의 유사도 비교를 통해 의미를 부착시킨다. 여기서 문맥에 부착되는 의미는 웹 문서 문맥과 각 의미별 기초 의미그룹과의 코사인 유사도 비교를 통해 가장 높은 유사도를 보이는 의미로 할당한다.

3.3 최종 의미그룹 생성

최종 의미그룹 생성 단계에서는 전 단계에서 생성된 의미부착 데이터를 통합하여 의미그룹 정보를 추출한 후 의미그룹 내 단어 가중치를 계산한다. 본 연구에서는 의미그룹 내 단어 가중치를 결정할 경우 집중도 개념[8]을 도입한다. 집중도란 특정 의미그룹 내 의미별 연관단어가 다른 의미에 비해서 특정 의미를 반영하는 정도를 말한다.

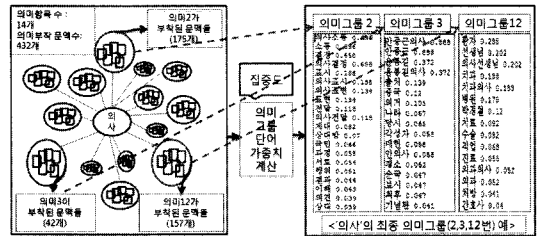


그림 4 최종 의미그룹 생성과정의 예

그림 4는 '의사'라는 중의어에 대한 의미그룹 내 단어 가중치를 계산하여 최종 의미그룹을 생성과정을 나타낸 것이다. 추출한 의미그룹의 부가 정보들을 이용하여 집중도 개념이 포함된 단어 가중치 결정 수식을 적용해 의미그룹 내 단어의 가중치를 결정한다. 본 연구의 단어 가중치 결정 수식은 [8]에서 사용된 기존의 집중도 수식을 수정하여 사용한다. 기존 집중도 수식에서는 다른 의미그룹 내에 해당 단어가 한 번이라도 등장하게 되면 마지막 항에 의해 전체 수식에 큰 영향을 미친다. 그러므로 본 연구의 수식에서는 다른 의미그룹 내의 문맥 비율을 반영하여 수식의 세 번째 항을 수정하여 사용한 다식 (1).

단어 가중치 결정 수식은 식 (1)과 같다. 수식의 첫 번째 항은 특정 의미그룹 내 해당 단어의 비중이 높을 수록 커지며 두 번째 항은 의미그룹 전체에서 해당 단어가 출현한 문맥 수에 비해서 특정 의미그룹 내에서 해당단어가 출현한 문맥 수가 많을수록 커진다. 세 번째

$$W_{w,i,k} = \frac{cf_{w,i}(k)}{N_{w,i}} \times \frac{cf_{w,i}(k)}{cf_{w,all}(k)} \times \sum_{j \neq i} (cxt_rate_{w,i,j} \times (\frac{N_{w,j} - cf_{w,j}(k)}{N_{w,j}}))$$

$$cxt_rate_{w,i,j} = \frac{N_{w,j}}{\sum_{k \neq i} N_{w,k}}$$

$W_{w,i,k}$: 중의어 w의 의미그룹 i내 단어 k의 가중치
 $N_{w,i}$: 중의어 w의 의미그룹 i내 문맥 수
 $cf_{w,all}(k)$: 중의어 w의 의미그룹 전체에서 단어 k가 출현한 문맥 수
 $cf_{w,i}(k)$: 중의어 w의 의미그룹 i내 단어 k가 출현한 문맥 수
 $cxt_rate_{w,i,j}$: 중의어 w의 i가 아닌 의미그룹들에서 의미그룹 j의 문맥 할당률

식 (1) 단어 가중치 결정 수식

항은 다른 의미그룹 내에서 해당 단어의 비중이 낮을수록 커지며 cxt_rate를 통해 문맥을 많이 가지고 있는 의미그룹의 영향력을 높인다. 즉, 특정 의미그룹 내에서의 해당 단어의 비중이 높을수록, 다른 의미그룹 내에서의 비중이 낮을수록 해당 단어의 특정 의미그룹 내 가중치가 높아진다. 이와 같은 과정을 통해 최종적인 의미그룹을 생성한다.

3.4 중의어 의미 태깅

전 단계에서 생성된 최종 의미그룹을 이용해 중의어를 포함한 평가 문맥 내 중의어의 의미를 태깅한다. 최종 의미그룹 내에서 가중치 상위 N개의 단어들을 중요 자질로 사용한다. 본 연구에서는 N값을 5~7로 사용한다. 식 (2)는 문맥 내 중의어의 최종적인 의미를 결정하는 수식이다. 의미별로 평가 문맥 내 중요 자질의 의미그룹의 가중치를 합한다. 부가적으로 중요 자질이 중의어로부터 떨어져있는 거리에 따라 반비례하는 가중치를 적용한다. 이와 같은 과정을 거쳐 중의어의 최종 의미에 가장 높은 가중치를 갖는 의미로 할당한다.

$$Final_Meaning(w,c) = \underset{i}{\operatorname{argmax}} \sum_{k \in Feature(w,i)} \frac{W_{w,i,k}}{d(w,c,k)}$$

Final_Meaning(w,c): 평가 문맥c내 중의어w의 의미
 Feature(w,i): 중의어w의 의미그룹내 중요 자질
 d(w,c,k): 중의어w의 평가 문맥c내 주변단어가 w로부터 떨어져있는 거리
 W_{w,i,k}: 중의어w의 의미그룹i내 단어k의 가중치

식 (2) 최종 의미 결정 수식

4. 실험 및 평가

4.1 자료 수집

본 연구에서는 기계가독형 사전으로 표준국어대사전과 두산백과사전을 사용하여 중의어의 뜻풀이, 예문, 유의어 정보 등을 추출하였다.

대상 중의어로는 먼저 기존 연구와의 실험 결과 비교를 위하여 SENSEVAL-2대회에서 사용한 중의어 9개(눈, 말, 목, 밤, 점, 의사, 거리, 바람, 자리)로 설정하였다. 이 외에 '배, 병, 차, 신, 다리, 경기, 신장, 인도, 연기, 기간' 등의 10개의 일반적인 대상 중의어에 대해 평가 데이터를 수작업으로 구축하여 평가하였다. 본 연구에서는 학습 데이터 구축을 위하여 'Daum 블로그 검색'으로부터 중의어로 검색된 블로그 본문 내용을 수집하였다. 수집된 문서 수는 중의어 당 2000건이다. 수집된 문서 중 중의어를 기준으로 앞 뒤 2문장 정도의 웹 문서 문맥을 추출하였다.

4.2 평가 데이터

평가를 위해 SENSEVAL-2대회의 9개의 대상 중의어에 대한 Korean 평가 데이터를 사용한다. 자동 태깅한

웹 문서를 이용해 학습하고, SENSEVAL의 평가 문맥을 이용하여 평가한다. 평가 척도는 정확도를 사용한다(식 (3)).

$$정확도 = \frac{\text{중의어 의미 태깅에 성공한 문맥 수}}{\text{전체 테스트 문맥 수}} \times 100$$

식 (3) 최종 의미 결정 수식

4.3 실험 결과

표 1은 최종 의미그룹 내에서 가중치 상위 N개의 단어들을 중요 자질로 사용하여 태깅한 결과를 나타낸 것으로 MFS(Most Frequent Sense) 태깅 실험 결과와 SENSEVAL대회의 두 참가자(KUNLP, KSENS)의 실험결과[9], 나이브 베이저안 수식을 적용한 결과와 비교한 것이다. MFS 태깅 실험은 학습 데이터에 가장 많이 나오는 의미인 MFS로 중의어의 의미를 태깅하는 방법을 사용한다. SENSEVAL대회의 두 참가자의 시스템은 지도식 학습방법을 사용한 방법으로 대부분 MFS 태깅 방법을 취하고 있다.

9개의 대상 중의어에 대하여 본 연구의 실험결과와 정확도는 평균 82%로 MFS 태깅 실험의 정확도인 76%보다 높았으며 SENSEVAL대회의 두 참가자보다도 높은 정확도 결과를 보였다. '의사', '밤', '말'과 같은 중의어에서 MFS 태깅 실험이나 SENSEVAL대회의 두 참가자보다 높은 정확도를 보인 이유는 자동 생성한 의미 부착 데이터로부터 MFS뿐 아니라 다른 의미들을 구분할 수 있는 중요한 자질을 추출하였기 때문이다.

'바람'이라는 중의어에 대한 태깅 정확도가 낮은 이유는 '바람'의 3번째 의미인 '어떤 일에 더불어 일어나는 기세'라는 의미가 MFS인데 이 의미는 주변 자질로 특징있는 단어를 가지고 있지 않기 때문에 주변 단어를 중요자질로 사용하여 의미 태깅을 하는 본 시스템에서는 정확도 성능이 떨어지기 때문이다.

표 1 실험 결과 1 (단위 : %)

중의어 \ 방법	본 시스템	MFS	KUNLP	KSENS	베이저안
의사	89	71	71	71	73
밤	86	80	80	24	64
바람	87	91	91	55	66
목	97	97	97	97	90
말	48	12	36	27	9
거리	60	62	63	62	47
자리	88	88	88	88	80
점	81	81	81	62	81
눈	92	97	97	97	90
평균	82	76	77	61	68

또한 본 연구의 중의성 해소 시스템 태깅 결과 나이브 베이지안[3]을 사용하여 태깅한 결과와 비교하여 높은 정확도를 보였다. 나이브 베이지안 수식(식 (4))으로 의미를 태깅할 경우 의미별 사용 확률과 해당 주변 단어가 중의어의 특정 의미에서 사용된 비율들을 곱하여 가장 가능성이 높은 의미로 태깅하게 된다. 하지만 다른 의미에서 해당 주변 단어가 사용되는 비율을 반영하지 않는다. 그러므로 의미그룹 전반에 골고루 나온 단어들의 가중치를 낮춘 단어 가중치 수식을 적용한 본 시스템이 나이브 베이지안 수식을 적용한 방법보다 높은 태깅 정확률을 나타냈다.

$$\begin{aligned}
 S &= \arg \max_{S_j \in S} p(S_j | V_1, V_2, V_3, \dots, V_k) \\
 &= \arg \max_{S_j \in S} p(V_1, V_2, V_3, \dots, V_k | S_j) p(S_j) \\
 &= \arg \max_{S_j \in S} \prod_k p(V_k | S_j) p(S_j)
 \end{aligned}$$

V_k : 중의어를 포함한 문맥내 주변단어

S_j : 중의어의 각 의미

식 (4) 나이브 베이지안 수식

표 2는 10개의 일반적인 중의어에 대한 정확도 실험 결과를 나타낸 것이다. 10개의 대상 중의어에 대하여 본 연구의 실험결과의 정확도는 평균 84%로 MFS방법의 정확도 평균인 80%보다 높게 나타났다. 표 1의 실험결과와 마찬가지로 한 음절의 중의어보다 두 음절의 중의어가 정확도가 대부분 높게 나타났다. 이는 색인어 추출 과정에서 한 음절 단어가 제외되며, ‘차’와 같은 중의어의 ‘국산차’, ‘외계차’와 같은 복합명사 형태에서 단어를 분리할 때 두 음절 이상만을 고려하였기 때문이다.

표 2 실험 결과 2 (단위 : %)

중의어 방법	배	병	차	신	다 리	경 기	신 장	인 도	연 기	기 간	평 균
본 시스템	76	75	79	88	93	88	92	90	81	75	84
MFS	70	68	80	85	83	86	84	88	77	82	80

5. 결론 및 향후 연구

본 연구에서는 중의성 해소를 위해 기계가독형 사전으로부터 추출한 정보를 이용하여 웹 문서 문맥 내 중의어의 의미를 자동으로 부착하였다. 새롭게 만들어진 의미부착 데이터를 중의성 해소를 위한 학습에 이용하였다.

또한 본 연구에서는 중의어 의미 태깅을 위한 의미그룹을 생성하여 의미그룹 내 단어 가중치 설정을 위해 집중도 개념을 도입하였다. 이를 통해 각 의미그룹에 골고루 나오는 불용어를 자동적으로 순위를 낮추고 특정

의미그룹을 대표하는 단어의 가중치를 높임으로써 태깅 정확도를 높였다.

실험 결과, 본 연구에서 제안한 방법이 MFS로 태깅하는 방법과 다른 SENSEVAL참가자들의 태깅 방법보다 향상된 태깅 정확률을 보였으며 베이지안 수식을 적용하여 태깅한 결과보다 높은 태깅 정확률을 보였다.

본 연구에서는 사전에서 추출한 seed의 양이 적을 경우 의미부착 데이터의 양 또한 적어지고 생성된 의미부착 데이터의 신뢰성이 떨어지는 문제점을 가지고 있다. 이를 해결하기 위해 웹 문서의 문맥에 의미부착할 때 의미를 할당받지 않은 문맥을 활용하는 방법에 대한 추가 연구가 필요하다.

참고 문헌

- [1] H. Lee, D. H. Baek, H. C. Rim, "Word Sense Disambiguation using Classification Information," *Journal of KIISE*, vol.24 no.7, pp.779-789, 1997. (in korean)
- [2] H. C. Seo, H. C. Rim, "Word Sense Disambiguation using Multiple Feature Decision Lists," *Journal of KIISE*, vol.30, no.7-8, pp.659-671, 2003.(in korean)
- [3] D. S. Kim, J. W. Choe, "A Study on the Computational Model of Word Sense Disambiguation, based on Corpora and Experiments on Native Speaker's Intuition," *Journal of KSCS*, vol.17, no.4, pp.303-321, 2006. (in korean)
- [4] H. Schutze, "Automatic Word Sense Discrimination," *Proc. Computational Linguistics*, 24(1): pp.97-123, 1998.
- [5] S. I. Shin, K. S. Choi, "Automatic Clustering of Collocation for Detecting Practical Sense Boundary," *Proc. ACL*, 2004.
- [6] M. Lesk, "Automatic Sense Disambiguation Using Machine Readable Dictionaries," *Proc. SIGDOC Conference*, pp.24-26, 1986.
- [7] J. Hur, M. G. Jang, "Homonym Disambiguation based on Average Mutual Information," *Proc. of the KIISE Korea Computer Congress*, pp.159-166, 2005. (in korean)
- [8] S. J. Lee, "A Related Keyword Group Extraction Method for Recommendation in Keyword Shop," M.S. Thesis, Soongsil University, Seoul Korea, 2003. (in korean)
- [9] H. C. Seo, S. Z. Lee, H. C. Rim, and H. Lee, "KUNLP System using Classification Information Model at SENSEVAL-2," *Proc. SENSEVAL-2*, pp.147-150, 2001.