

# 다중 자질 결정 목록을 이용한 단어 의미 중의성 해결

## (World Sense Disambiguation using Multiple Feature Decision Lists)

서 희 철<sup>†</sup>    임 해 창<sup>\*\*</sup>  
(Hee-Cheol Seo) (Hae-Chang Rim)

**요 약** 본 논문에서는 결정 목록을 이용해서 단어 의미 중의성을 해결하는 방법을 제안한다. 결정 목록은 하나 이상의 규칙으로 구성되며, 각 규칙에는 신뢰도가 부여되어 있고, 규칙은 불린 함수(=조건, precondition)와 부류(=의미, class)로 구성되어 있다. 분류 대상이 만족하는 불린 함수를 가진 규칙들 중에서 가장 신뢰도가 높은 규칙에 의해서 분류 대상의 부류가 정해진다. 기존 방법에서는 하나의 자질로 하나의 불린 함수를 구성하는 단일 자질 결정 목록을 이용해서 단어 의미 중의성을 해결했다. 이 경우, 자료 부족 문제와 전처리 과정의 오류에 민감하게 반응한다는 문제점이 있다. 본 논문에서는 기존의 단일 자질 결정 목록의 문제점을 해결하기 위해서, 하나 이상의 자질로 불린 함수를 구성하는 다중 자질 결정 목록을 제안하고, 다중 자질 결정 목록을 이용하여, 단어 의미 중의성을 해결하는 방법을 기술하고 있다. 단일 자질 결정 목록과 다중 자질 결정 목록을 비교하기 위해서, 1개의 한국어 의미 부착 말뭉치와 5개의 영어 의미 부착 말뭉치를 대상으로 단어 의미 중의성 해결 실험을 했다. 실험 결과 6개의 말뭉치 모두에서 다중 자질 결정 목록이 단일 자질 결정 목록에 비해서 더 좋은 결과를 나타냈다.

**키워드** : 다중자질 결정목록, 단어 의미 중의성 해결

**Abstract** This paper proposes a method of disambiguating the senses of words using decision lists, which consists of rules with confidence values. The rule of decision list is composed of a boolean function(=precondition) and a class(=sense). Decision lists classify the instance using the rule with the highest confidence value that is matched with it. Previous work disambiguated the senses using single feature decision lists, whose boolean function was composed of only one feature. However, this approach can be affected more severely by data sparseness problem and preprocessing errors. Hence, we propose multiple feature decision lists that have the boolean function consisting of more than one feature in order to identify the senses of words. Experiments are performed with 1 sense tagged corpus in Korean and 5 sense tagged corpus in English. The experimental results show that multiple feature decision lists are more effective than single feature decision lists in disambiguating senses.

**Key words** : multiple feature decision lists, word sense disambiguation

### 1. 서 론

자연어에는 하나의 표현이 두 개 이상의 해석을 가지는 경우가 있다. 이런 경우 중의성이 있다고 한다. 사람들은 자연어의 중의성에도 불구하고 의사 소통이 가능하

지만, 기계는 자연어의 중의성으로 인해서 자연어를 제대로 처리할 수 없다. 그러므로 기계가 자연어를 처리하기 위해서는 자연어의 중의성을 해결할 수 있어야 한다.

자연어에는 어휘 중의성과 구문 중의성이 있다. 어휘 중의성은 단어가 가지는 중의성이며, 하나 이상의 품사를 가지는 단어에 의해서 발생하는 단어 품사 중의성과 하나 이상의 의미를 가지는 단어에 의해서 발생하는 단어 의미 중의성이 있다. 그리고 구문 중의성은 문장이 가지는 중의성으로, 하나 이상의 구문 구조를 가지는 문

<sup>†</sup> 학생회원 : 고려대학교 컴퓨터학과  
hseo@nlp.korea.ac.kr

<sup>\*\*</sup> 종신회원 : 고려대학교 컴퓨터학과 교수  
rim@nlp.korea.ac.kr

논문접수 : 2002년 9월 25일

심사완료 : 2003년 3월 17일

장에 의해서 발생하는 중의성이다[1].

본 논문에서는 자연어의 어휘 중의성 중에서 단어 의미 중의성을 해결하는 방법을 제시한다. 의미 중의성을 가지는 단어를 다의어(Polysemous Word)<sup>1)</sup>라고 하며, 단어 의미 중의성을 해결하는 것을 단어 의미 중의성 해결(Word Sense Disambiguation)이라고 한다. 예를 들어, 단어 “배”는 “과일”, “신체일부” 그리고 “교통수단” 등의 의미를 가지는 다의어이고, “철수는 배를 타고 강을 건넜다.”와 같은 문장에서 다의어 “배”가 “교통수단”의 의미로 이용되었음을 판별하는 작업이 단어 의미 중의성 해결이다[2].

단어 의미 중의성을 해결하는 방법은 규칙에 의한 방법[4][5], 사전을 이용하는 방법[6][7][8] 그리고 말뭉치를 이용하는 방법이 있다. 규칙에 의한 방법은 사람이 수작업으로 생성한 규칙을 이용해서 의미를 판별하는 것으로 규칙 생성을 위한 시간이 많이 필요하다는 단점이 있다.

사전을 이용하는 방법은 사전에 나타난 다의어의 각 의미 해설 부분에 나타난 정보를 이용해서 의미를 결정한다. 이 방법은 자동으로 의미를 결정할 수 있지만, 사전의 해설 부분에 나타난 정보가 충분하지 않다는 점과, 사전 해설 부분의 비일관성으로 인해서 좋은 성능을 기대할 수 없다[9].

말뭉치를 이용하는 방법은 두 가지가 있다. 하나는 원시 말뭉치(raw corpus)를 이용하는 방법이고, 다른 하나는 의미 부착 말뭉치(sense tagged corpus)를 이용하는 방법이다. 원시 말뭉치를 이용하는 방법은 원시 말뭉치에 나타난 다의어의 용례들을 다의어의 의미에 따라 무리짓고(clustering), 새로운 다의어 문맥을 적당한 무리에 대응하는 것으로 다의어의 의미를 결정한다<sup>2)</sup>. 이 방법은 다의어의 의미를 사전에 나타난 의미로 부여하는 것이 아니라, 무리 번호로 부여한다. 그러므로, 다의어의 의미를 판별하기 위해서는 무리 번호와 다의어의 의미간의 대응과정이 필요하다.

의미 부착 말뭉치는 다의어의 의미가 부착된 말뭉치이다. 의미 부착 말뭉치를 이용하는 방법은 다의어의 각 의미와 함께 나타난 단어 정보를 자질로 이용하고, 기계 학습 방법을 이용해서 다의어의 의미를 결정한다. 의미 결정을 위해서 사용되는 기계 학습 방법으로는 기본 베

이지언 모델[11], 신경망[12], SVM[13], 예제 기반 모델 [14][15], 부스팅(boosting)[16], 결정 목록[17][18] 등이 있다.

단어 의미 중의성 해결 시스템은 단어 의미 중의성을 해결하는 시스템, 그 자체로 완전한 시스템이라기보다는 자연어 처리의 기반 요소로서[3], 자연어를 처리하는 여러 시스템에서 필요로 한다. 한국어 문서를 영어 문서로 변환하는 한영 기계 번역 시스템은 한국어 문서를 정확하게 번역하기 위해서 단어 의미 중의성 해결 시스템을 필요로 한다. 예를 들어, 다의어 “배”는 의미에 따라서 “ship(교통수단)”, “pear(과일)”, “stomach(신체일부)” 등의 영어 대역어가 존재하는데, “배”의 의미를 제대로 판별하지 못한 경우에는 적절한 대역어를 선정하지 못하게 된다.

그리고 정보 검색 시스템에서 사용자 질의와 관련된 문서를 제시하기 위해서도 단어 의미 중의성 해결 시스템이 필요하다. 예를 들어, “수상 교통 수단”으로서의 “배”를 찾고자 하는 이용자에게 관련된 문서를 제시하기 위해서는 검색 시스템에서 단어의 의미를 정확하게 판별할 수 있어야 된다. 기계 번역, 정보 검색 시스템 외에, 음성 인식 및 합성, 철자 교정, 품사 태깅 그리고 다양한 문서 처리 분야에서 단어 의미 중의성 해결 시스템을 필요로 한다[2].

단어 의미 중의성 해결 시스템에 대한 연구가 활발히 진행되면서, 단어 의미 중의성 해결 시스템들의 성능을 비교, 평가하는 대회인 SENSEVAL이 생겼다. SENSEVAL은 1998년과 2001년에 개최되었으며, 1998년에는 영어 다의어를 대상으로 시스템들을 평가했고, 2001년에는 영어뿐만 아니라, 한국어, 일본어, 중국어, 스페인어 등 10개의 언어에 대해서 시스템을 평가했다.

SENSEVAL에 참가한 시스템은 자율 학습 시스템(unsupervised learning system)과 교사 학습 시스템(supervised learning system)으로 나눌 수 있다. 자율 학습은 의미 부착 말뭉치 없이 학습하는 것을 의미하며, 교사 학습은 의미 부착 말뭉치에서 학습하는 것을 의미한다. SENSEVAL에서는 자율 학습 시스템보다는 교사 학습 시스템이 더 좋은 결과를 나타냈다. 그리고 교사 학습 시스템들 중에서 결정 목록을 사용하는 시스템이 매우 좋은 성능을 보였다[19][20].

본 논문에서는 SENSEVAL에서 우수한 성능을 보인 결정 목록을 이용해서 의미를 결정한다. 베이지언 모델, 신경망, SVM과 같은 기계 학습 방법은 다의어 주변에 나타난 모든 자질들을 고려하지만, 결정 목록은 다의어 주변의 일부 중요한 자질만을 고려해서 의미를 결정한다.

1) 개별 어휘의 여러 의미가 서로 연관된 단어를 다의어라고 하고, 개별 어휘의 의미가 서로 연관되지 않은 단어를 동형어라고 한다[1]. 그러나, 단어 의미 중의성 해결 측면에서 동일하게 취급될 수 있으므로[2], 본 논문에서는 다의어라는 용어로 통일해서 이용한다.  
2) 이 방법을 단어 의미 중의성 해결과 구별해서 단어 의미 분별(word sense discrimination)이라고도 한다[10].

다. 다의어의 의미는 다의어 주변에 나타난 모든 자질들에 의해서 결정되기보다는 다의어의 의미와 관련된 일부의 중요한 자질에 의해서 결정된다(one sense per collocation[21]). 그러므로, 단어 의미 중의성 해결에서는 결정 목록이 다른 기계 학습 방법에 비해서 좋은 결과를 보일 수 있다.

결정 목록을 이용하는 기존 연구에서는 하나의 자질만을 고려하는 단일 자질 결정 목록을 사용해서 의미를 결정했다. 단일 자질 결정 목록은 하나의 자질만 사용해서 분류하므로, 자료 부족 문제와 전처리 과정에서 발생한 오류에 민감하게 반응하는 문제점이 있다. 이 점을 해결하기 위해서, 본 논문에서는 하나 이상의 자질을 고려하는 다중 자질 결정 목록을 제안하고, 다중 자질 결정 목록으로 단어 의미 중의성을 해결한다. 본 논문에서 제안하는 다중 자질 결정 목록과 기존의 단일 자질 결정 목록을 비교, 평가하기 위해서 1개의 한국어 의미 부착 말뭉치와 5개의 영어 의미 부착 말뭉치에서 실험했다.

본 논문의 구성은 2장에서 결정 목록에 대해서 살펴보고, 하나 이상의 자질로 구성된 다중 자질 결정 목록을 생성하는 방법을 설명한다. 3장에서는 다중 자질 결정 목록을 이용해서 단어 의미 중의성을 해결하는 방법을 알아보고, 4장에서는 본 논문에서 제안한 방법이 기존 방법에 비해서 얼마나 유용한지를 실험으로 살펴본다. 마지막으로 5장에서 결론을 내리고, 향후 연구에 대해서 언급한다.

## 2. 결정 목록

### 2.1 정의

결정 목록[22]은 기계 학습 방법의 한 종류로써, 표 1과 같이 정의된다.

표 1에서 DL은  $r$ 개의 규칙(Rule)로 구성된 결정 목록이다. 규칙은 규칙의 신뢰도(confidence)에 의해서 정렬되어 있으며, 불린 함수(Boolean Function:  $bf$ )와 부류값(class value:  $v$ )으로 구성된다. 불린 함수는 규칙의 조건(precondition)으로써, 하나 이상의 자질들의 AND( $\wedge$ )연산으로 되어있다. 불린 함수를 구성하는 자질들을

표 1 결정 목록

$DL = Rule_1, Rule_2, \dots, Rule_r$ $Confidence(Rule_i) > Confidence(Rule_j), \text{ for } i < j$ $Rule_i = (bf_i, v_i), v_i \in C, C \text{는 부류들의 집합}$ $bf_i = f_{i1} \wedge f_{i2} \wedge \dots \wedge f_{in}, f_{ij} \in F, F \text{는 자질들의 집합}$
--

모두 가지고 있는 분류 대상이 불린 함수를 참으로 만든다. 부류값은 규칙의 결과(consequence)로써, 불린 함수를 만족하는 분류 대상에 할당되는 부류이다.

결정 목록은 분류 대상을 신뢰도가 높은 규칙부터 낮은 규칙순으로 비교하며, 분류 대상을 만족하는 첫 번째 규칙으로 분류한다. 결정 목록으로 분류하는 과정은 *if-then-elseif-then-... else* 형식으로 다음과 같이 표현될 수 있다.

```

if  $bf_1$  then  $v_1$ 
elseif  $bf_2$  then  $v_2$ 
...
elseif  $bf_r$  then  $v_r$ 
    
```

위에서  $bf_i$ 는 결정 목록의  $i$ 번째 규칙이 가지는 불린 함수이고,  $v_i$ 는 결정 목록의  $i$ 번째 규칙이 가지는 부류값이다.

### 2.2 단일 자질 결정 목록

분류 대상을 분류하기 위해서는 결정 목록이 생성되어 있어야 된다. 결정 목록은 규칙을 생성하는 단계와 규칙에 신뢰도를 부여하는 단계로 생성된다.

규칙 생성 단계는 불린 함수를 생성하는 단계와 불린 함수와 부류를 연결하는 단계로 이루어진다. 불린 함수는 자질들의 AND연산으로 구성되며, 자질들의 가능한 조합으로 만들어지므로, 자질의 수가 2개인 경우에는 3개의 불린 함수가 생성되고, 자질의 수가 3개인 경우에는 7개의 불린 함수가 생성되며, 자질의 수가  $n$ 개인 경우에는  $2^n - 1$ 개의 불린 함수가 생성된다. 예를 들어, 두 개의 자질  $f_1, f_2$ 로 생성되는 불린 함수는 다음과 같다.

$$\begin{aligned}
 bf_1 &= f_1 \\
 bf_2 &= f_2 \\
 bf_3 &= f_1 \wedge f_2
 \end{aligned}$$

자연어 처리 시스템들은 대체로 많은 수의 자질을 이용하므로, 자질들의 조합으로 생성되는 불린 함수는 매우 많다. 예를 들어, 단어 의미 중의성 해결 시스템에서는 다의어와 함께 나타난 단어들을 자질로 이용하므로, 수천개의 자질이 생성된다. 1000개의 자질만 이용하더라도  $2^{1000} - 1$ 개의 불린 함수가 생성되며, 이는 처리하기에는 너무 많은 불린 함수이다.

이 문제를 해결하기 위해서, [24]와 [18]은 단일 자질 결정 목록(single feature decision lists)을 사용했다.

3)  $n$ 개의 자질로 만들어 질 수 있는 불린 함수의 수는 다음과 같이 구한다.  
 ${}_nC_1 + {}_nC_2 + \dots + {}_nC_n = 2^n - {}_nC_0 = 2^n - 1$

단일 자질 결정 목록은 하나의 자질로 구성된 불린 함수를 가지는 결정 목록이다. 즉, 불린 함수를  $bf_i = f_{\bar{a}}$ 의 형태로 표현하며, 자질의 수만큼 불린 함수를 생성했다. 그리고 생성된 불린 함수와 부류를 연결함으로써 규칙을 완성했다. 생성된 규칙의 신뢰도를 계산하기 위해서 [24]에서는 수식 (1)을, [18]에서는 수식 (2)를 사용했다.

$$\begin{aligned}
 Confidence_{yarowsky}(Rule_i) &= Confidence_{yarowsky}(bf_i, v_i) \\
 &= Confidence_{yarowsky}(f_{\bar{a}}, v_i) \\
 &= LogLikelihood_{yarowsky}(f_{\bar{a}}, v_i) \\
 &= \log \frac{Pr(f_{\bar{a}}|v_i)}{Pr(f_{\bar{a}}|\neg v_i)} \quad (1)
 \end{aligned}$$

$$\begin{aligned}
 Confidence_{agirre}(Rule_i) &= Confidence_{agirre}(bf_i, v_i) \\
 &= Confidence_{agirre}(f_{\bar{a}}, v_i) \\
 &= LogLikelihood_{agirre}(f_i, v_i) \\
 &= \log \frac{Pr(v_i|f_i)}{Pr(\neg v_i|f_i)} \quad (2)
 \end{aligned}$$

수식 (1)과 수식 (2)에서  $Confidence(Rule_i)$ 는  $i$ 번째 규칙  $Rule_i$ 의 신뢰도를 나타낸다.  $Rule_i$ 는 불린 함수  $bf_i$ 와 부류  $v_i$ 로 구성되며, 불린 함수  $bf_i$ 는 하나의 자질  $f_{\bar{a}}$ 로만 구성되어 있다.  $LogLikelihood(f_{\bar{a}}, v_i)$ 는 불린 함수를 구성하는 자질  $f_{\bar{a}}$ 과 부류  $v_i$ 의 관련성을 나타내며,  $Pr(x|y)$ 는  $y$ 가 주어졌을 때  $x$ 가 발생할 조건부 확률이다.

규칙의 신뢰도 계산에서 수식 (1)은 부류가 주어졌을 때 자질이 발생할 확률( $Pr(f_{\bar{a}}|v_i)$ )을 고려하는 반면에, 수식 (2)는 자질이 주어졌을 때 부류가 발생할 확률( $Pr(v_i|f_i)$ )을 고려한다. 그러므로, 수식 (1)은 학습 자료에 적게 나타난 부류와 자질간의 관계를 과대 평가할 수 있고, 수식 (2)는 학습 자료에 많이 나타난 부류와 자질간의 관계를 과대 평가할 수 있다. 다시 말하면, 수식 (1)을 이용하는 단일 자질 결정 목록은 학습 자료에서 빈도가 낮은 부류를 선호하게 되고, 수식 (2)를 이용하는 단일 자질 결정 목록은 학습 자료에서 빈도가 높은 부류를 선호하게 된다. 표 2는 예를 통해서, 수식 (1)과 수식 (2)에서 사용된  $LogLikelihood$ 값을 비교한다.

표 2는 학습 자료에서 부류가 나타난 빈도(괄호 안에 있는 수), 자질과 부류간의 공기 빈도, 자질과 부류간의  $LogLikelihood$ 값을 보여준다. 표에서 보듯이,

$LogLikelihood_{yarowsky}$ 는 학습 자료에서 빈도가 낮은 부류에 높은 값을 부여한다.  $LogLikelihood_{agirre}$ 는 학습 자료에서 부류의 빈도는 고려하지 않고, 자질과 부류간의 공기 빈도만을 고려한다. 학습 자료에 많이 나타난 부류와의 공기 빈도가 상대적으로 높기 때문에,  $LogLikelihood_{agirre}$ 는 학습 자료에서 빈도가 높은 부류에 높은 값을 부여한다. 예를 들어, feature2는 class1과 class2가 학습 자료에 동일한 빈도로 나타났다면 class1과의  $LogLikelihood$ 값이 더 높겠지만, 표 2에서 feature2는 학습 자료에 많이 나타난 부류 class2와의  $LogLikelihood$ 값이 더 높다.

하나의 자질만으로 불린 함수를 구성하는 단일 자질 결정 목록이 좋은 성능을 보이고 있지만, 다음과 같은 문제점이 있다.

① 자료 부족 문제로 인해서 과대 평가 혹은 과소 평가된 자질에 의해서 오류가 발생할 가능성이 높다. 예를 들어, 표 2에서 feature1은 부류 class1과 한번 공기했지만,  $LogLikelihood_{yarowsky}$ 값은 매우 높다. 그리고 feature2는 학습 자료에서 10번 나타난 부류 class1과 5번이나 공기했지만,  $LogLikelihood_{agirre}$ 값은 낮다. 이처럼 과대 혹은 과소 평가받는 자질-부류 관계가 있을 수 있으며, 하나의 자질만을 사용하는 경우에는 이런 자질들에 의해서 오류가 발생할 가능성이 높다.

② 전처리 과정에서 발생하는 오류에 민감하게 반응한다. 예를 들어, 다의어의 품사는 다의어의 의미를 결정하는데 매우 중요하므로, 하나의 자질로 의미를 결정할 때 중요한 자질이다. 그러므로, 품사 부착 오류로 인해서 다의어의 품사를 잘못 부착한 경우에는 오류가 생길 가능성이 매우 높게 된다.

③ 두 개 이상의 부류와 관계가 있는 자질의 특성을 반영하지 못한다. 예를 들어, 표 2에서 자질 feature2는 부류 class1, class2와 모두 공기한다. 그러나,

$$\begin{aligned}
 &LogLikelihood_{agirre}(feature_2, class_1) \\
 &< LogLikelihood_{agirre}(feature_2, class_2)
 \end{aligned}$$

이므로, 자질 feature2는 항상 class2로 분류한다.

표 2 예:  $LogLikelihood$ 값 비교

	class1(10)	class2(100)	$LogLikelihood_{yarowsky}$		$LogLikelihood_{agirre}$	
			class1	class2	class1	class2
feature1	1	1	3.32	-3.32	1	-1
feature2	5	6	3.05	-3.06	-0.27	0.26

2.3 다중 자질 결정 목록

본 논문에서는 단일 자질 결정 목록의 문제점들을 해결하기 위해서 하나 이상의 자질로 불린 함수를 구성하는 다중 자질 결정 목록(multiple feature decision lists)을 제안한다. 다중 자질 결정 목록에서는 매우 많은 불린 함수가 생성되는 문제가 있는데, 이는 학습 자료에 나타난 모든 자료를 대상으로 불린 함수를 생성하기 때문이다. 본 논문에서는 학습 과정이 아닌 분류 과정에서 각 분류 대상에 맞는 불린 함수를 생성한다. 각 분류 대상이 가지는 자질은 학습 자료에 나타나는 자질에 비해서 매우 적은 수이므로, 학습 과정에서 생성되는 불린 함수에 비해서 매우 적은 수의 불린 함수가 생성된다. 그리고 분류 대상에 나타난 자질들만으로 구성된 불린 함수이므로, 이 불린 함수를 가지는 규칙은 항상 분류 대상을 만족하는 규칙이 된다. 각 분류 대상에 나타난 자질들로 구성되는 불린 함수는 다음과 같다.

$$bf = f_1 \wedge f_2 \wedge \dots \wedge f_n, \quad f_i \in F_{instance}$$

$F_{instance}$ 는 분류 대상에 나타난 자질들의 집합

그러나, 각 분류 대상에 나타난 자질의 수가 학습 자료에 나타난 자질의 수에 비해서 매우 적은 수이긴 하지만, 분류 과정에서 처리할 수 없을 정도로 많은 불린 함수가 생성되는 경우가 있다. 이는 소수의 자질로도 많은 수의 불린 함수가 생성되기 때문이다. 예를 들어, 분류 대상이 가지고 있는 자질의 수가 10개이면 1023개 ( $= 2^{10} - 1$ )의 불린 함수가 생성되지만, 자질의 수가 20개이면 100만여개 ( $= 2^{20} - 1$ ), 30개이면 10억여개 ( $= 2^{30} - 1$ ) 그리고 40개이면 1조여개 ( $= 2^{40} - 1$ )의 불린 함수가 생성된다. 이처럼 많은 불린 함수를 생성해서 분류 대상을 분류하는 경우에는 너무 많은 시간이 필요하다. 그러므로 많은 수의 자질을 가지는 분류 대상에 대해서는, 불린 함수를 생성하는데 여전히 문제가 있다.

결정 목록은 분류 대상과 관계가 있는 규칙들 중에서 가장 신뢰도가 높은 규칙을 사용해서 분류한다. 즉, 분류 대상과 관계가 있으면서, 가장 신뢰도 높은 규칙만을 생성할 수 있으면, 결정 목록으로 분류 대상을 분류할 수 있다. 그러므로, 분류 대상과 관계가 있는 규칙들 중에서 신뢰도가 높은 규칙만을 생성하고 낮은 신뢰도의 규칙은 생성할 필요가 없다. 본 논문에서는 분류 대상과 관계가 있는 규칙들 중에서 신뢰도가 높은 규칙만을 생성함으로써, 불린 함수의 과생성을 막는다.

하나 이상의 자질로 구성된 불린 함수를 가지는 규칙의 신뢰도는 다음 수식 (3)을 사용해서 구한다.

$$\begin{aligned} Confidence(Rule_i) &= Confidence(bf_i, v_i) \\ &= Confidence(f_{i1} \wedge f_{i2} \wedge \dots \wedge f_{in}, v_i) \\ &\approx \sum_{j=1}^n Confidence(f_{ij}, v_i) \\ &= \sum_{j=1}^n LogLikelihood(f_{ij}, v_i) \end{aligned} \quad (3)$$

수식 (3)에서 불린 함수와 부류간의 신뢰도를 계산하기 위해서는, 학습 과정에서 불린 함수를 생성해야 되는데, 앞서 기술했듯이, 하나 이상의 자질로 구성된 불린 함수를 생성하는 일은 매우 어렵다. 그러므로, 본 논문에서는 자질들간의 독립가정을 사용해서 문제를 해결한다. 자질-부류쌍의 신뢰도는 [24]와 [18]에서 사용한 *LogLikelihood*값을 사용한다.

독립 가정에 의해서, 규칙의 신뢰도는 불린 함수를 구성하는 자질과 부류간의 신뢰도로 계산된다. 그러므로, 높은 신뢰도의 규칙은 높은 신뢰도의 자질-부류들로 구성되며, 각 부류와 관련된 규칙들 중에서 가장 신뢰도가 높은 규칙은 자질-부류간의 신뢰도가 양이 되는 모든 자질들로 구성된 불린 함수를 가지는 규칙이다. 각 부류와 관련된 가장 좋은 규칙은 다음과 같다.

$$\begin{aligned} BestRule(class|instance) &= (Bestbf(class|instance), class) \\ &= (f_1 \wedge f_2 \wedge \dots \wedge f_n, class) \end{aligned}$$

where for  $f_i \in F_{instance}$ ,  $F_{instance}$ 는 분류 대상에 나타난 자질들의 집합,  
for  $\forall f_i \in Bestbf(class|instance)$ ,  
 $LogLikelihood(f_i, class) > 0$   
and for  $\forall f_j \in Bestbf(class|instance)$ ,  
 $LogLikelihood(f_j, class) \leq 0$

*BestRule(class|instance)*은 분류 대상 *instance*를 부류 *class*로 분류하는 가장 신뢰도 높은 규칙으로, 분류 대상 *instance*가 만족하는 가장 좋은 불린 함수 *Bestbf(class|instance)*와 부류 *class*로 구성된다. 그리고 *Bestbf(class|instance)*는 분류 대상에 나타난 자질들 중에서 부류 *class*와의 신뢰도가 양( $LogLikelihood(f_i, class) > 0$ 인) 모든 자질들로 구성된다.

분류 대상과 관련된 규칙들 중에서 가장 신뢰도가 높은 규칙의 부류로 분류 대상을 분류하며, 분류 대상의 부류는 다음과 같이 결정된다.

$$\begin{aligned} class(instance) &= \underset{class}{\operatorname{argmax}} Confidence \\ &\quad (BestRule(class|instance)) \\ &= \underset{class}{\operatorname{argmax}} \sum_{f: LogLikelihood(f, class) > 0} LogLikelihood(f_i, class) \end{aligned}$$

*class(instance)*는 분류 대상인 *instance*에 할당되는 부류를 나타내며, 분류 대상의 부류는 신뢰도 (*Confidence*)가 가장 높은 규칙(*BestRule*)에 의해서 결정된다. 각 부류별로 가장 신뢰도가 높은 규칙은 부류와

의  $LogLikelihood$  값이 양인 모든 자질들로 구성된다.

### 3. 결정 목록을 이용한 단어 의미 중의성 해결

문맥에 나타난 다의어의 의미는 다의어의 주변 단어들에 의해서 결정된다. 주변 단어들 중에는 다의어의 의미 결정에 기여하는 단어들도 있지만, 그렇지 않은 단어들도 있을 수 있다. 예를 들어, "Paulo wants to have much money in his **bank** account to buy a 35 foot sailboat."에서 다의어 "bank"의 의미를 결정하는데 기여하는 단어는 "account", "money"이다. 그리고 "Paulo", "wants", "to", "have", "much", "in", "his", "buy", "a", "35", "foot", "sailboat"는 의미 결정에 기여하지 않는다. 오히려, "sailboat"는 의미 결정에 부정적인 영향을 준다. 그러므로, 의미 중의성 해결 시스템은 의미 결정에 기여할 수 있는 단어들을 선별해서 사용하는 것이 바람직하다. 결정 목록은 분류 대상에 나타난 자질들 중에서 분류에 기여하는 자질만을 사용해서 분류한다는 점에서 단어 의미 중의성 해결에 적합한 모델이다. 그리고, 위의 예에서처럼 의미 결정에 기여하는 단어가 하나 이상인 경우에는, 의미 결정에 기여하는 모든 단어들을 고려하는 것이 의미결정의 신뢰도를 높일 수 있다. 그러므로, 하나 이상의 자질을 고려하는 다중 자질 결정 목록을 이용해서 의미를 결정하는 것이 바람직하다.

다중 자질 결정 목록을 사용해서 단어 의미 중의성을 해결하는 경우, 불린 함수는 다의어와 함께 나타난 단어들로 구성되며, 부류는 다의어의 의미가 된다. 분류 대상은 다의어를 포함하고 있는 문맥이 된다. 그리고, 의미 결정은 다의어를 포함하는 문맥과 관련된 규칙들 중에서 가장 신뢰도가 높은 규칙에 의해서 이루어진다. 이를 정리하면 다음과 같다.

자질 집합  $F = \{w_1, w_2, w_3, \dots, w_n\}$ ,  $w_i$ 는 의미 결정에 사용되는 자질

불린 함수  $bf = f_1 \wedge f_2 \wedge \dots \wedge f_n$ ,  $f_i \in F$

부류  $C = \{s_1, s_2, \dots, s_m\}$ ,  $s_i$ 는 다의어의  $i$ 번째 의미

규칙  $Rule_i = (bf_i, v_i)$ ,  $v_i \in C$

$$\begin{aligned} sense(pw|context) &= \underset{s}{\operatorname{argmax}} Confidence \\ &\quad (BestRule(s_i|pw, context)) \\ &= \underset{s}{\operatorname{argmax}} \sum_{f_j \in F, LogLikelihood(f_j, s_i) > 0} LogLikelihood(f_j, s_i) \end{aligned}$$

$$\begin{aligned} BestRule(s_i|pw, context) &= (Bestbf(s_i|pw, context), s_i) \\ &= (f_1 \wedge f_2 \wedge \dots \wedge f_n, s_i) \end{aligned}$$

where for  $f_j \in F_{context}$ ,  $F_{context}$ 는 다의어 문맥에 나타난 자질들의 집합.

$$\begin{aligned} &\text{for } \forall f_j \in Bestbf(s_i|pw, context), \\ &\quad LogLikelihood(f_j, s_i) > 0 \\ &\text{and for } \forall f_j \in Bestbf(s_i|pw, context), \\ &\quad LogLikelihood(f_j, s_i) \leq 0 \end{aligned}$$

위에서  $sense(pw|context)$ 는 다의어 문맥  $context$ 에 나타난 다의어  $pw$ 의 의미를 나타낸다.

예를 들어, 다의어 문맥 "Paulo wants to have much money in his **bank** account to buy a 35 foot sailboat."에서 다의어 "bank"의 의미는 다음과 같이 결정된다.

자질 집합  $F_{context} = \{Paulo, wants, to, have, much, money, in, his, bank, account, to, buy, a, 35, foot, sailboat\}$ <sup>4)</sup>

부류  $C = \{financial\_institution, slope\_of\_land\}$

$Bestbf(financial\_institution|bank, context) = money \wedge account$

$Bestbf(slope\_of\_land|bank, context) = sailboat$

$$\begin{aligned} sense(bank|context) &= \underset{s}{\operatorname{argmax}} Confidence \\ &\quad (BestRule(s_i|pw, context)) \\ &= financial\_institution \end{aligned}$$

위에서, 다의어  $bank$ 는 규칙  $(money \wedge account, financial\_institution)$ 에 의해서  $financial\_institution$ 의 미로 결정된다.

### 4. 실험 및 평가

단일 자질 결정 목록과 다중 자질 결정 목록을 비교, 평가하기 위해서 본 논문에서는 영어권에서 많이 사용되는 5개의 영어 의미 부착 말뭉치에서 실험했다. 의미 부착을 위한 전처리 단계에서 단어들을 토큰으로 분리하고, 각 단어에 품사를 부착했다. 다의어와 함께 나타난 단어 정보와 연어 정보를 자질로 사용했다.

#### 4.1 전처리

한국어 말뭉치의 전처리 과정에서는 한국어 품사 부착기[23]를 사용해서 형태소 분석과 단어의 품사를 결정했다. 영어 말뭉치의 전처리 과정에서는 영어 단어들을 토큰으로 분리하고, 영어 품사 부착기[25]를 사용해서 단어의 품사를 결정했다. 토큰화(tokenization)는 단어와 기호를 분리하는 것을 의미한다. 예를 들어, "I'm straight, white, no longer middle class, anti-IRA, have ..."는 "I 'm straight , white , no longer middle class , anti - IRA , have ..."로 토큰화된다. 예외로, 아포스트로피(apostrophe: ')는 뒤에 나타난 단어와 분리하지 않았다.

#### 4.2 자질 집합

본 논문에서 사용한 자질 집합은 다음 세 가지 자질

4) 단어만을 자질로 사용하는 경우

표 3 예: 다의어 **빚**이 있는 문맥에서 추출한 자질들

다의어 문맥	낮/일반명사+에는/부사격조사 굴/일반명사 속/일반명사+에서/부사격조사 잠/일반명사+을/목적격조사 자/동사+고 /어말어미+./기호, 밤/일반명사+에만/부사격조사 활동/동작성명사+하/용언화접사+1.다/종결어미+./기호.
지역 문맥	<ul style="list-style-type: none"> <li>•잠/-2 을/-2 자/-1 고/-1 /-1 밤/0 예만/0 활동/1 하/1 1.다/1 .1</li> <li>•잠/일반명사 을/목적격조사 자/동사 고/어말어미 ./기호, 밤/일반명사 예만/부사격조사 활동/동작성명사 하/용 언화접사 1.다/종결어미 ./기호.</li> <li>•잠을/-2 자고/-1 밤에만/0 활동한다./1</li> <li>•잠을 자고 고, 밤에만 활동하 하.다 1.다.</li> <li>•잠_목적격조사 일반명사_을 자_어말어미 동사_고 고_기호, 어말어미_, 밤_부사격조사 일반명사_에만 활동_용 언화접사 동작성명사_하 하_종결어미 용언화접사_1.다 1.다_기호. 종결어미_기호.</li> </ul>
광역 문맥	•낮 굴 속 잠 자 밤 활동
언어 문맥	•(잠을 자고,) (자고 밤에만) (밤에만 활동한다.)

유형으로 구성된다.

- 지역 문맥(local context)에 나타난 자질

지역 문맥은 다의어와 근접해서 나타난 단어들로 구성된다. 다의어와 근접해서 나타난 단어들은 다의어와 의미적으로 관련될 뿐만 아니라, 통사적으로도 관련되어 있다. 예를 들어, 다의어 “배”와 근접해서 나타난 단어 “먹\_동사”는 “배”와 의미적으로 관련될 뿐만 아니라, “배”를 목적어로 가지는 동사일 수도 있다. 한국어 단어 의미 중의성 해결에서 지역 문맥의 크기는 다의어의 앞에 나타난 두 단어와 뒤에 나타난 세 단어이며, 지역 문맥은 ①형태소와 형태소의 상대적인 위치, ②형태소와 형태소의 품사, ③어절과 어절의 상대적인 위치, ④형태소-형태소 바이그램(bigram), ⑤형태소-형태소의 품사 바이그램으로 구성된다<sup>5)</sup>. 영어 단어 의미 중의성 해결에서 지역 문맥의 크기는 다의어의 앞에 나타난 세 단어와 뒤에 나타난 세 단어이며, 지역 문맥은 ①단어와 단어의 상대적인 위치, ②단어와 단어의 품사, 그리고 ③ 단어의 품사와 상대적인 위치로 구성된다.

- 광역 문맥(topical context)에 나타난 자질

광역 문맥은 다의어를 포함하고 있는 문맥에 나타난 단어들로 구성된다. 문맥을 구성하는 단어들은 문맥의 의미를 전달하기 위해서 이용된 단어들이며 이들은 의미적으로 서로 관련되는 단어들이다. 그러므로, 다의어와 함께 나타난 단어들은 다의어의 의미와 관련이 있다. 예를 들어, “과수원”, “딸기”, “포도”, “막다” 등과 같은 단어들은 “배”의 의미를 결정하는데 기여한다. 광역 문맥의 크기는 다의어를 포함하는 문장과 앞뒤 문장이며, 광역 문맥은 단어의 의미간의 관련성으로 구성된 문맥

이므로 의미를 가지는 단어, 즉 명사 단어, 동사 단어, 형용사 단어, 부사 단어들로 구성된다. 한국어에서는 형태소가 광역 문맥을 구성하고, 영어에서는 단어들이 광역 문맥을 구성한다.

- 언어 문맥(collocation context)에 나타난 자질

언어 문맥은 다의어와 근접해서 나타난 연어<sup>6)</sup>들로 구성된다. 지역 문맥과 광역 문맥은 단어들로 구성되는 반면에, 언어 문맥에 나타난 자질은 단어쌍으로 구성된다. 이는 다의어의 의미가 단어쌍에 의해서 더 잘 판별되는 경우가 있기 때문이다. 예를 들어, “pay attention to”와 같은 숙어에서 다의어 ‘pay’는 ‘attention’과 ‘to’라는 각각의 단어보다는 ‘attention to’라는 단어쌍에 의해서 의미 결정이 더 잘 된다. 한국어 의미 중의성 해결에서 언어 문맥의 크기는 다의어의 앞 두 단어, 뒤 세 단어이며, 언어 문맥은 어절쌍으로 구성된다. 영어 의미 중의성 해결에서 언어 문맥의 크기는 다의어의 앞 두 단어, 뒤 두 단어이며, 언어 문맥은 ①단어쌍과 ②단어와 품사 쌍으로 구성된다. hedge one’s bets에서 다의어 bet의 의미는, 단어쌍(hedge his 혹은 hedge her 등)이 아니라, 단어-품사쌍(hedge PRP\$<sup>7)</sup>)에 의해서 결정되므로, 본 논문에서는 단어와 품사쌍을 언어 문맥에 포함시켰다.

표 3은 한국어 다의어 “빚”이 나타난 문맥에서 추출한 자질들을 보여준다.

표 4는 영어 다의어 bet이 나타난 문맥에서 추출한 자질들을 보여준다. 표 4에서 다의어 문맥에 있는 Motoring/VBG는 단어 Motoring과 단어의 품사 VBG를 의미한다. 지역 문맥에서 tend/-4는 다의어 bet의 앞 네 번째 단어가 tend임을 의미하고, tend/VBP는 지역에 나타난 단어 tend의 품사가 VBP임을 의미하고, VBP/-4는 다의어 bet의 앞 네 번째 단어의 품사

5) 지역 문맥과 언어 문맥의 크기를 정하기 위해서, 문맥의 크기를 1에서 6까지 바꾸면서 실험했다. 실험 결과 가장 좋은 결과를 나타낸 문맥 크기를 실험에 사용했다. 문맥 크기를 정할 때, 한국어는 한국어 의미 부록 말뭉치를 사용했으며, 영어는 SENSEVAL1 말뭉치를 사용했다.

6) 본 논문에서 연어는 두 단어로 구성되어 있다.

7) PRP\$는 소유격을 의미한다.

표 4 예: 다의어 *bet*이 있는 문맥에서 추출한 자질들

다의어 문맥	Motoring/VBG writers/NNS ./, like/IN politicians/NNS ./, tend/VBP to/TO hedge/VB their/PRP\$ bets/NNS ./
지역 문맥	<ul style="list-style-type: none"> <li>• ./-5 tend/-4 to/-3 hedge/-2 their/-1 bets/0 ./1</li> <li>• ./, tend/VBP to/TO hedge/VB their/PRP\$ bets/NNS ./</li> <li>• ./-5 VBP/-4 TO/-3 VB/-2 PRP\$/-1 NNS/0 ./1</li> </ul>
광역 문맥	• Motoring writers politicians tend hedge bets
언어 문맥	<ul style="list-style-type: none"> <li>• (hedge their) (hedge bets) (hedge .) (their bets) (their .) (bets .)</li> <li>• (hedge PRP\$) (hedge NNS) (hedge .) (their NNS) (their .) (bets .)</li> <li>• (VB their) (VB bets) (VB .) (PRP\$ bets) (PRP\$ .) (NNS .)</li> </ul>

표 5 실험 말뭉치

실험 말뭉치	다의어 수	말뭉치 크기 (다의어 문맥수)	의미 분류 사전
한국어 말뭉치	7개	13,002개	.
DSO 말뭉치[26]	191개	192,800개	WordNet 1.5
Bruce 말뭉치[28]	1개	2,369개	LDOCE
Leacock 말뭉치[11]	3개	12,983개	WordNet
SENSEVAL1 말뭉치	32개	16,582개	Hector 사전
SENSEVAL2 말뭉치	73개	12,939개	WordNet 1.7

가 *VBP*임을 의미한다. 언어 문맥에서 (*hedge their*)은 단어 *hedge*와 단어 *their*가 언어 문맥에 함께 나타났음을 의미하고, (*hedge PRP\$*)는 단어 *hedge*와 품사 *PRP\$*가 언어 문맥에 함께 나타났음을 의미한다.

#### 4.3 실험 말뭉치

본 논문에서는 한국어 의미 부착 말뭉치와 영어 의미 부착 말뭉치를 이용해서 실험을 했다. 한국어 말뭉치를 사용해서 한국어에서 기존의 단일 자질 결정 목록과 본 논문에서 제안하는 다중 자질 결정 목록의 유용성을 평가한다. 그리고 기존 연구에서 많이 사용되는 영어 의미 부착 말뭉치를 사용함으로써, 본 논문에서 제안하는 방법과 기존 연구를 비교 평가할 수 있다. 실험에 이용된 말뭉치는 표 5와 같다.

한국어 말뭉치는 한국어 1000만 어절 말뭉치에서 추출한 7개의 다의어에 대해서 의미 부착된 말뭉치이며, 동형의어 수준에서 의미를 분류했다. DSO 말뭉치[26]는 Brown 말뭉치와 WSJ 말뭉치에 있는 다의어의 용례에 대해서 수작업으로 구축한 의미 부착 말뭉치이다. Brown 말뭉치와 WSJ 말뭉치가 서로 다른 특성을 가지므로, DSO 말뭉치에서 실험을 하는 경우에는 Brown 말뭉치 부분과 WSJ 말뭉치 부분을 나누어서 실험을 하는 경우가 있다[27][16]. 본 논문에서도 DSO 말뭉치 전체에서도 실험을 하고, Brown 말뭉치 부분과 WSJ 말뭉치 부분으로 나누어서도 실험을 했다.

#### 4.4 실험 방법

SENSEVAL1과 SENSEVAL2 말뭉치는 단어 의미 중의성 해결 시스템들의 성능을 평가하기 위한 말뭉치로, 학습 말뭉치와 테스트 말뭉치가 구별되어 있다. SENSEVAL1 말뭉치는 10,598개의 학습 문맥과 5,984개의 테스트 문맥으로 구성되어 있다<sup>8)</sup>. SENSEVAL2 말뭉치는 8,611개의 학습 문맥과 4,328개의 테스트 문맥으로 되어 있다. SENSEVAL 말뭉치에 대해서는 학습 말뭉치에서 학습을 하고, 테스트 말뭉치에서 정확도를 평가했다.

한국어 의미 부착 말뭉치, DSO 말뭉치, Bruce 말뭉치, 그리고 Leacock 말뭉치는 학습 말뭉치와 테스트 말뭉치의 구별이 없다. 그래서 본 논문에서는 10-등분 교차 검증(10-fold cross validation)을 수행했다. 즉, 말뭉치를 동일한 크기로 10등분하고, 그 중에서 하나는 테스트 말뭉치로 사용하고 나머지 9개는 학습 말뭉치로 사용해서, 모두 10번의 실험을 수행했다.

결정 목록에서 사용되는 *LogLikelihood*를 계산하기 위해서는, 분모가 0이 되지 않도록 평탄화(smoothing)를 해야된다. 본 논문에서는 평탄화를 위해서 다음의 예

8) SENSEVAL1 말뭉치에는 명사, 동사, 형용사 다의어 그리고 품사가 결정되지 않은 다의어(indeterminate word)가 있다. 본 논문에서는 품사가 결정된 단어 즉, 명사, 동사, 형용사 다의어에 대해서만 실험을 했다. 명사, 동사, 형용사 단어에 대한 말뭉치의 크기이다.



디티브 평탄화를 사용했다[29].

$$\Pr(w_i | w_{i-1}) = \frac{\delta + c(w_i, w_{i-1})}{\delta |V| + \sum_w c(w_i, w_{i-1})}$$

위의 수식에서  $\Pr(w_i | w_{i-1})$ 은  $w_{i-1}$ 이 주어졌을 때,  $w_i$ 가 나타날 조건부 확률이고,  $c(w_i, w_{i-1})$ 은  $w_i$ 와  $w_{i-1}$ 이 함께 나타난 빈도이고,  $|V|$ 는  $w_i$ 의 종류 수이고,  $\delta$ 는 실험 상수이다. 본 논문에서는 실험에 의해서  $\delta$ 값을 0.000001로 정했다<sup>9)</sup>.

본 논문에서는 학습 말뭉치에서 가장 많이 나타난 의미로 의미를 결정하는 방법(most frequent class:MFC), 단일 자질 결정 목록(SFDL), 그리고 다중 자질 결정 목록(MFDL)을 비교, 평가했다. 그리고 [24]에서 제안한 규칙의 신뢰도 수식과 [18]에서 제안한 규칙의 신뢰도 수식에 대해서 비교 실험을 했다. 실험 결과에서 (Y)와 (A)는 각각 [24]과 [18]에서 제안한 규칙의 신뢰도 수식을 사용했음을 나타낸다.

4.5 실험 결과 및 평가

본 논문에서는 정확률을 이용해서 시스템의 성능을 평가하며, 정확률은 다음과 같이 구한다.

$$\text{정확률}(\%) = \frac{\text{정확하게 맞춘 다의어 문맥수}}{\text{다의어 문맥수}} \times 100$$

각 말뭉치별 실험 결과가 표 6에서부터 표 12까지 있다. 각 표에는 본 논문에서 수행한 실험 결과와 함께 기존 연구의 실험 결과를 함께 기록했으며, SENSEVAL1과 SENSEVAL2 말뭉치에 대해서는 SENSEVAL에 참가한 시스템들 중에서 상위 3개 시스템의 결과를 표시했다.

DSO 말뭉치를 이용한 기존 연구로는 [26][14][16][27]가 있다. [26]와 [14]는 DSO 말뭉치 전체를 실험에 사용했고, 예제 기반 모델을 사용해서 의미를 결정했다. [16]과 [27]는 DSO 말뭉치에서 21개의 단어<sup>10)</sup>를 선별해서 실험하고 하나의 자질로만 구성된 불린 함수를 사용하는 결정 목록으로 의미를 결정했다. 본 논문에서는 DSO 말뭉치 전체에서 실험한 결과와 21개의 단어에 대한 실험 결과를 제시했다. 표 7은 DSO 전체 말뭉치에 대한 실험 결과이고, 표 8은 21개의 단어에 대한 실험 결과이다.

표 6 한국어 말뭉치에 대한 실험 결과

	정확률(%)	불린 함수 수(개)
MFC	72.95 <sup>+0.66</sup>	
SFDL(Y)	85.73 <sup>+1.32</sup>	43.46
SFDL(A)	89.60 <sup>+0.80</sup>	38.36
MFDL(Y)	91.87 <sup>+0.95</sup>	53.16
MFDL(A)	90.60 <sup>+0.71</sup>	38.36

표 7 DSO 말뭉치에 대한 실험 결과(1)

	정확률(%)		
	Brown	WSJ	Brown + WSJ
MFC	47.47 <sup>+0.36</sup>	59.72 <sup>+0.43</sup>	53.26 <sup>+0.26</sup>
SFDL(Y)	43.30 <sup>+0.54</sup>	57.84 <sup>+0.28</sup>	52.27 <sup>+0.26</sup>
SFDL(A)	60.13 <sup>+0.39</sup>	72.95 <sup>+0.43</sup>	67.31 <sup>+0.31</sup>
MFDL(Y)	62.89 <sup>+0.31</sup>	75.56 <sup>+0.49</sup>	70.37 <sup>+0.22</sup>
MFDL(A)	62.71 <sup>+0.36</sup>	75.03 <sup>+0.48</sup>	69.68 <sup>+0.30</sup>
Ng[26]	54.0	68.6	
Ng[14]	58.7	75.2	

표 8 DSO 말뭉치에서의 실험 결과(2)

- 21개의 단어에 대한 실험 결과

	정확률(%)	
	Brown	WSJ
Escudero[16]	58.96 <sup>+1.85</sup>	71.26 <sup>+1.15</sup>
Martinez[27]	57.2	65.2
SFDL(Y)	42.77 <sup>+1.69</sup>	58.66 <sup>+1.08</sup>
SFDL(A)	58.53 <sup>+0.93</sup>	72.57 <sup>+1.19</sup>
MFDL(Y)	60.83 <sup>+1.49</sup>	75.77 <sup>+1.17</sup>
MFDL(A)	60.77 <sup>+1.63</sup>	74.84 <sup>+1.12</sup>

Bruce 말뭉치를 이용하는 기존 연구로는 [28][26][2]가 있다. [28]은 말뭉치에서 600개의 문맥을 무작위로 추출해서 테스트 말뭉치로 이용하고, 나머지를 학습 말뭉치로 이용했다. 그리고 분해 가능 모델(decomposable model)을 사용해서 의미를 결정했다. [26]와 [2]에서도 [28]과 동일한 방법으로 학습 말뭉치와 테스트 말뭉치를 만들었으며 100번 반복 실험했다. [26]는 예제 기반 방법을 사용했으며, [2]는 분류 정보 모형을 사용해서 의미를 결정했다. 실험 결과는 표 9와 같다.

Leacock 말뭉치는 명사 다의어 line, 형용사 다의어 hard, 그리고 동사 다의어 serve에 대해서 의미 부착된 말뭉치이다. Leacock 말뭉치를 이용한 기존 연구로는 [11]이 있으며, [11]에서는 단어의 의미마다 200개의 문맥으로 학습 말뭉치를 구성하고, 나머지 문맥을 테스트 말뭉치로 이용해서 실험했다. 그리고 기본 베이지언 모델을 사용해서 의미를 결정했다. 실험 결과는 표 10과 같으며, 각 단어에 대한 정확률과 함께 세 단어 모두에 대한 정확률을 함께 제시하고 있다.

9) SENSEVAL1 말뭉치를 대상으로,  $\delta$ 값을 바꾸면서 실험한 결과. 0.000001에서 좋은 결과를 얻었다.  $\delta$ 값은 0.1, 0.01, 0.001, ..., 0.00000001, 0.000000001까지 바꾸면서 실험했다.

10) 의미가 많으면서 말뭉치에 자주 나타나는 단어를 선택했다. 명사 단어는 age, art, body, car, child, cost, head, interest, line, point, state, thing, work 13개이고, 동사 단어는 become, fall, grow, lose, set, speak, strike, tell 8개이다.

표 9 Bruce 말뭉치에 대한 실험 결과

	정확률(%)
MFC	52.91 <sup>+2.85</sup>
SFDL(Y)	84.74 <sup>+1.83</sup>
SFDL(A)	84.40 <sup>-2.12</sup>
MFDL(Y)	91.65 <sup>+1.41</sup>
MFDL(A)	88.53 <sup>+1.76</sup>
Bruce[28]	79
Ng[26]	89
이호[2]	89

표 10 Leacock 말뭉치에 대한 실험 결과

	정확률(%)			
	line	hard	serve	전체
MFC	53.52 <sup>+2.15</sup>	80.49 <sup>+1.54</sup>	41.10 <sup>+2.22</sup>	58.77 <sup>+1.18</sup>
SFDL(Y)	78.09 <sup>+1.78</sup>	85.91 <sup>+1.82</sup>	85.55 <sup>+1.08</sup>	83.29 <sup>+1.18</sup>
SFDL(A)	82.05 <sup>+1.59</sup>	89.31 <sup>+1.31</sup>	84.32 <sup>+1.63</sup>	85.33 <sup>+1.21</sup>
MFDL(Y)	87.36 <sup>+1.10</sup>	92.68 <sup>+0.78</sup>	89.59 <sup>+1.54</sup>	89.96 <sup>+0.66</sup>
MFDL(A)	83.35 <sup>+1.09</sup>	89.40 <sup>+1.09</sup>	87.49 <sup>+1.47</sup>	86.64 <sup>+0.85</sup>
Leacock[11]	76	77	78	

표 11 SENSEVAL1 말뭉치에 대한 실험 결과

	정확률(%)			
	명사	형용사	동사	전체
MFC	41.70	62.15	53.10	50.85
SFDL(Y)	58.16	41.04	36.83	45.57
SFDL(A)	71.67	71.11	66.25	69.28
MFDL(Y)	77.67	76.17	70.69	74.43
MFDL(A)	73.85	73.60	69.97	72.18
hopkins[17](R)	86.5	77.7	70.6	77.1
ets-pu[27](R)	80.6	74.4	70.9	75.6
tilburg[31](R)	80.0	74.6	68.7	74.6

표 12 SENSEVAL2 말뭉치에 대한 실험 결과

	정확률(%)			
	명사	형용사	동사	전체
MFC	51.65	58.98	40.31	48.22
SFDL(Y)	38.60	36.33	26.69	33.23
SFDL(A)	63.23	65.23	50.66	58.34
MFDL(Y)	65.56	68.49	54.49	61.46
MFDL(A)	65.28	69.14	54.10	61.30
JHU[32](R)	68.2	73.2	56.6	64.2
SMUis[15]	69.5	68.8	56.3	63.8
KUNLP[33]	66.8	66.8	57.6	62.9

SENSEVAL1 말뭉치는 첫 번째 SENSEVAL 대회에 사용된 말뭉치이며, 참가한 시스템들 중에서 상위 3개 시스템의 결과를 함께 제시했다. SENSEVAL1 말뭉치에는 명사 다의어, 동사 다의어, 형용사 다의어, 그리고 품사가 결정되지 않은 다의어 등 4가지 유형의 다의어가 있다. 본 논문에서는 이 중에서 명사 다의어, 동사 다의어, 형용사 다의어들에 대해서만 실험을 했다. 반면에, SENSEVAL1에 참가한 시스템들은 품사가 결정되

지 않은 다의어에 대해서 의미 결정을 했다. 그러므로, MFC, 기존 방법, 제안하는 방법에 대한 “전체” 실험 결과는 품사가 결정되지 않은 다의어에 대한 정확률은 포함하고 있지 않고, SENSEVAL1에 참가한 시스템에 대한 “전체” 실험 결과는 품사가 결정되지 않은 다의어에 대한 정확률을 포함하고 있다. 실험 결과는 표 11과 같으며, (R)은 결과 제출 기한 이후에 제출된 결과를 의미한다.

SENSEVAL2 말뭉치는 두 번째 SENSEVAL 대회에 사용된 말뭉치이며, 영어, 한국어, 일본어, 중국어 등 10개의 언어에 대한 말뭉치가 있지만, 본 논문에서는 영어 말뭉치에 대해서만 실험을 했다. 영어 태스크에 참가한 시스템들 중에서 상위 3개 시스템의 결과를 함께 제시했다. 실험 결과는 표 12와 같으며, (R)은 결과 제출 기한 이후에 제출된 결과를 의미한다.

한국어 말뭉치에 대한 실험 결과 다중 자질 결정 목록이 단일 자질 결정 목록에 비해서 더 좋은 결과를 보였다. [24]의 신뢰도 수식을 이용하는 경우에는 약 6%의 성능 향상이 있으며, [18]의 신뢰도 수식을 이용하는 경우에는 약 1%의 성능 향상이 있었다.

단일 자질 결정 목록은 하나의 자질만 고려하므로, 문맥에서 하나의 자질을 잘못 선택하는 경우에는 잘못된 의미를 선택한다. 예를 들어, “돼지머리에, 통닭에, 사과, 배, 밥, 대추, 꽃감 그리고 술을 올렸다.”라는 문장에서, 단일 자질 결정 목록은 ‘술’ 하나의 자질만을 고려해서 ‘밥’을 ‘시간’ 의미로 결정하지만, 다중 자질 결정 목록은 ‘돼지머리’, ‘통닭’, ‘사과’, ‘배’, ‘대추’, ‘꽃감’ 등 하나 이상의 자질을 고려해서 ‘밥’을 ‘음식물’ 의미로 결정한다. 그리고 “긴 장대를 가지고 밥을 땀다.”라는 문장에서, 단일 자질 결정 목록은 ‘밥’과는 의미적으로 관련성이 없는 ‘긴’에 의해서 ‘밥’을 ‘시간’ 의미로 결정한다. 왜냐하면, 학습 말뭉치에서 “긴 밥”이 많이 나타나며, ‘긴’이 ‘따/1’보다 가중치가 높기 때문이다. 반면에 다중 자질 결정 목록은 ‘따/1’, ‘목적격조사\_따’, (밥을 땀다.) 등의 자질을 고려해서 ‘음식물’의 의미로 정확하게 의미 결정한다.

그러나 다중 자질 결정 목록도 잘못된 의미 결정을 하는 경우가 있다. 학습 말뭉치에서 의미별로 학습 자료가 편중된 경우, 학습 말뭉치에서 높은 빈도로 나타난 의미는 공기하는 자질의 수가 많은 반면, 적은 빈도로 나타난 의미는 공기하는 자질의 수가 적다. 이와 같은 자료 부족 문제로 인해서 하나의 의미와만 공기하는 자질이 있을 수 있다. 이런 자질들로 구성된 문맥은 다중 자질 결정 목록을 사용하더라도 정확하게 의미 결정되

지 않았다.

표 6에는 의미 결정의 정확률뿐만 아니라, 의미 결정에 사용된 불린 함수의 수가 있다. 불린 함수의 수는 하나의 다의어 문맥에서 다의어의 의미를 결정하는데 사용된 평균수이다. [18]의 신뢰도 수식을 사용하는 경우에는 단일 자질 결정 목록과 다중 자질 결정 목록에서 고려하는 불린 함수의 수가 동일하다. [18]의 신뢰도 수식에서는 하나의 자질은 하나의 부류와만 관련있기 때문이다. [24]의 신뢰도 수식에서는 단일 자질 결정 목록에 비해서 다중 자질 결정 목록에서 고려하는 불린 함수의 수가 10개 정도 많다.

5개의 영어 말뭉치에서 단일 자질 결정 목록(SFDL)에 비해서 다중 자질 결정 목록(MFDL)이 더 좋은 결과를 보였다. 그리고 단일 자질 결정 목록에서는 [18]의 신뢰도 수식이 [24]의 신뢰도 수식보다 더 좋은 결과를 보였지만, 다중 자질 결정 목록에서는 거의 차이가 없었다. 오히려 [24]의 신뢰도 수식이 근소하게 좋은 성능을 보였다. [24]의 수식은 학습 말뭉치에서 낮은 빈도로 나타나는 부류를 선호하지만, 본 논문에서 제안하는 다중 자질 결정 목록은 하나 이상의 자질을 고려하므로 학습 말뭉치에서 높은 빈도로 나타나는 부류를 선호하는 경향이 있다. 이런 두 가지 상반된 경향이 적절하게 조화를 이루어서 성능을 향상시키는 것으로 생각된다.

DSO 말뭉치, Bruce 말뭉치, Leacock 말뭉치에서는 다중 자질 결정 목록이 기존 연구들에 비해서 더 좋은 결과를 보였다. 그러나, SENSEVAL1 말뭉치와 SENSEVAL2 말뭉치에서는 기존 연구에 비해서 근소하게 낮은 성능을 보였다. 이는 DSO 말뭉치, Bruce 말뭉치, Leacock 말뭉치를 사용한 기존 연구는 하나의 모델을 사용해서 의미를 결정한 반면에, SENSEVAL1 말뭉치와 SENSEVAL2 말뭉치를 사용한 기존 연구는 하나 이상의 모델을 함께 사용해서 의미를 결정했기 때문으로 생각된다. 예를 들어, 첫 번째 SENSEVAL에 참가한 [17]은 수작업으로 구축한 규칙을 함께 사용했으며, [30]은 기본 베이지언 모델과 언어 처리 모델을 함께 사용해서 의미 결정을 했다. 그리고 두 번째 SENSEVAL에 참가한 [32]는 결정 목록과 기본 베이지언 모델을 함께 사용하고 있으며, [15]는 메모리 기반 모델과 단어 패턴 처리 모델을 함께 사용하며, [33]은 분류 정보 모형[2]과 언어 처리 모델을 함께 사용하고 있다. 본 논문에서는 결정 목록 하나를 사용해서 의미 결정했다. 그러므로, 하나의 모델을 사용하는 기존 연구들과 비교할 때, 본 논문에서 제안한 방법이 가장 좋은 성능을 보였다.

## 5. 결론 및 향후연구

본 논문에서는 결정 목록을 이용해서 단어 의미 중의성을 해결했다. 결정 목록을 이용하는 기존 방법들은 하나의 자질로만 구성된 불린 함수를 이용하였으며, 이 경우 자료 부족 문제와 전처리 과정에서 발생한 오류에 민감하게 반응하는 문제점이 있다. 본 논문에서는 하나 이상의 자질로 불린 함수를 구성함으로써 이런 문제를 해결했다.

제안하는 다중 자질 결정 목록과 기존의 단일 자질 결정 목록을 비교 평가하기 위해서, 1개의 한국어 의미 부작 말뭉치와 5개의 영어 의미 부작 말뭉치에서 실험했다. 실험 결과, 모든 실험 말뭉치에서 본 논문에서 제안하는 다중 자질 결정 목록이 기존 단일 자질 결정 목록보다 좋은 결과를 보였다. 그리고 동일한 말뭉치를 사용한 기존 연구와 다중 자질 결정 목록을 비교할 때, 하나의 모델을 사용한 기존 연구들 중에서는 가장 좋은 결과를 보였다. 그러므로, 하나 이상의 자질을 고려하는 다중 자질 결정 목록이 단어 의미 중의성 해결에서는 매우 효과적인 방법임을 알 수 있다.

그리고 단일 자질 결정 목록에서는 [18]에서 제안한 신뢰도 수식이 [24]보다 더 좋은 성능을 보였지만, 다중 자질 결정 목록에서는 [24]의 신뢰도 수식이 [18]보다 더 좋은 결과를 보였다. [18]의 수식은 학습 자료에서 많이 나타난 부류를 선호하는 반면에, [24]는 학습 자료에서 적게 나타난 부류를 선호한다. 그리고 다중 자질 결정 목록은 학습 자료에 많이 나타난 부류를 선호하는 경향이 있으므로, [24]와의 상호 보완에 의해서 좋은 결과를 보인 것으로 생각된다.

다중 자질 결정 목록은 하나의 자질로만 의미 결정되는 문맥에서는 오류를 범하는 경우가 있다. 즉, 단일 자질 결정 목록에서 의미 결정할 수 있는 문맥을 의미 결정하지 못하는 경우가 있다. 이 문제를 해결할 수 있는 방법에 대해서 향후에 살펴보고자 한다.

그리고 결정 목록은 일부의 중요한 자질만을 이용해서 의미를 결정하므로, 의미 결정에 영향을 주는 중요한 자질이 없는 문맥에 대해서는 의미 결정할 수 없거나, 혹은 잘못 결정할 가능성이 높다. [32]는 이런 결정 목록의 단점을 보완하기 위해서, 중요한 자질을 포함하는 문맥은 결정 목록으로 의미 결정하고, 그렇지 않은 문맥은 다른 기계 학습 방법을 사용해서 의미를 결정했다. 다중 자질 결정 목록에서도 이런 문제가 발생할 수 있다. 그러므로, 향후에는 다중 자질 결정 목록의 문제를 보완할 수 있는 기계 학습 방법에 대해서 살펴보고, 다

중 자질 결정 목록과 함께 이용하는 방법에 대해서 연구할 계획이다.

마지막으로, 본 논문에서는 평탄화를 위해서 애디티브 평탄화(additive smoothing)을 사용했다. 이 평탄화 방법은 매우 간단한 방법으로 알려져 있으며, 이보다 개선된 많은 방법들이 있다. 그러므로, 평탄화 방법을 개선함으로써 성능의 향상을 기대할 수 있으며, 향후에는 다른 평탄화 방법을 도입하고자 한다.

### 참 고 문 헌

- [1] 이상주, 자동 품사 부착을 위한 새로운 통계적 모형, 고려대학교 컴퓨터학과 박사학위 논문, 1999.
- [2] 이호, 단어 의미 중의성 해결을 위한 분류 정보 모형, 고려대학교 컴퓨터학과 박사학위 논문, 1999.
- [3] Yorick Wilks and mark Stevenson, "The Grammar of sense: Is word sense tagging much more than part-of-speech tagging?," In *Technical Report CS-96-05, University of Sheffield*, 1996.
- [4] Kelly, E. and Philip S. *Computer Recognition of English Word Senses*, Amsterdam, North-Holland, 1975.
- [5] Weiss, S., *Learning to Disambiguate*, Information Storage and Retrieval, Vol. 9, pp. 33-41, 1973.
- [6] Lesk, M., "Automatic Sense Disambiguation: How to tell a Pine Cone from an Ice Cream Cone," In *Proceeding of the 1986 SIGDOC Conference*, New York: Association for Computing Machinery, 1986.
- [7] Luk, K. A., "Statistical sense disambiguation with relatively small corpora using dictionary definitions," In *Proceedings of the 33rd Annual Meetings of the Association for Computational Linguistics*, 1995.
- [8] Veronis, J. and Ide, Nancy, "Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries," In *Proceedings COLING-90*, pp. 389-394, 1990.
- [9] Nancy Ide and Jean Veronis, *Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art*, Computational Linguistics, Vol.24, No.1, pp.1-40, 1998.
- [10] Hinrich Schuetze, *Automatic Word Sense Discrimination*, Computational Linguistics, Vol.24, No.1, pp.97-123, 1998.
- [11] Claudia Leacock, Martin Chodorow, and George A. Miller, *Using Corpus Statistics and WordNet Relations for Sense Identification*, Computational Linguistics, Vol.24, No.1, pp.147-165, 1998.
- [12] Geoffrey Towell and Ellen M. Voorhees, *Disambiguating Highly Ambiguous Words*, Computational Linguistics, Vol.24, No.1, pp.125-1145, 1998.
- [13] Clara Cabezas, Philip Resnik and Jessica Stevens, Supervised Sense Tagging using Support Vector Machines," In *Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation System(SENSEVAL 2)*, pp. 59-62, 2001.
- [14] Hwee Tou Ng, "Exemplar Based Word Sense Disambiguation: Some Recent Improvement," In *Proceedings of the 2nd conference on Empirical Methods in Natural Language Processing*, 1997.
- [15] Rada F. Mihalcea and Dan I. Moldovan, "Pattern Learning and Active Feature Selection for Word Sense Disambiguation," In *Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation System(SENSEVAL 2)*, pp. 127-130, 2001.
- [16] Gerard Escudero, Lluís Marquez and German Rigau, "A Comparison between Supervised Learning Algorithms for Word Sense Disambiguation," In *Proceedings of CoNLL-2000 and LLL 2000*, 2000.
- [17] David Yarowsky, *Hierarchical Decision Lists for Word Sense Disambiguation*, Computers and the Humanities, Vol.34, No.1-2: pp. 179-186, 2000.
- [18] Eneko Agirre and David Martinez, "Exploring automatic word sense disambiguation with decision lists and the Web," In *Proceedings of the Semantic Annotation And Intelligent Annotation workshop organized by COLING*, Luxembourg 2000.
- [19] A. Kilgarriff and J. Rosenzweig, *Framework and Results for English SENSEVAL*, Computers and the Humanities, Vol.34, No.1-2: pp. 15-48, 2000.
- [20] A. Kilgarriff, "English Lexical Sample Task Description," In *Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation System(SENSEVAL-2)*, pp.17-20, 2001.
- [21] David Yarowsky, "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," In *Proceedings on 33rd Annual Meeting of the Association for Computational Linguistics*, pp.189-196, 1995.
- [22] Ronald L. Rivest, *Learning decision lists*, Machine Learning, Vol.2, No.3, pp.229-246, 1987.
- [23] 김진동, 임희석, 임해창, "Twoply HMM : 한국어의 특성을 고려한 형태소 단위의 품사 태깅 모델", 한국정보과학회 논문지(B), 제24권, 제12호, pp.1502-1512, 1997.
- [24] David Yarowsky, "Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French," In *Proceedings on 32nd Annual Meeting of the Association for Computational Linguistics*, pp.88-95, 1994.
- [25] Eric Brill, "Some Advances in rule-based part of

speech tagging," In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI 94)*, 1994.

[26] Hwee Tou Ng and Hian Beng Lee, "Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar Based Approach," In *Proceedings on 34th Annual Meeting of the Association for Computational Linguistics*, pp.88-95, 1996.

[27] David Martinez and Eneko Agirre, "One Sense per Collocation and Genre/Topic Variations," In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 2000.

[28] Rebecca Bruce and Janyce Wiebe, "Word Sense Disambiguation using Decomposable models," In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pp. 139-146, 1994.

[29] Stanley F. Chen and Joshua Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling," In *Technical Report TR 10-98*, Computer Science Group, Harvard University.

[30] Martin Chodorow, Claudia Leacock, and George A. Miller, *A Topical/Local Classifier for Word Sense Identification*, Computers and the Humanities, Vol. 34, No. 1-2, pp. 115-120, 2000.

[31] Jorn Veenstra, Antal van den Bosch, Sabine Buchholz, Walter Daelemans and Jakub Zavrel, *Memory based Word Sense Disambiguation*, Computers and the Humanities, Vol. 34 No.1-2, pp. 171-177, 2000.

[32] David Yarowsky, Silviu Cucerzan, Radu Florian, Charles Schafer, and Richard Wicentowski, "The Johns Hopkins SENSEVAL2 system descriptions," In *Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation System (SENSEVAL 2)*, pp. 163-166, 2001.

[33] Hee Cheol Seo, Sang Zoo Lee, Hae Chang Rim, and Ho Lee, "KUNLP system using Classification Information Model at SENSEVAL 2," In *Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation System (SENSEVAL 2)*, pp. 147-150, 2001.

임 해 창

정보과학회논문지 : 소프트웨어 및 응용  
제 30 권 제 1 호 참조



서 회 철

1998년 2월 고려대학교 컴퓨터학과 학사  
2000년 2월 고려대학교 컴퓨터학과 석사  
2000년 3월~현재 고려대학교 정보통신  
공동연구소 연구원. 관심분야는 자연어처  
리, 정보검색