

# CRF를 이용한 운율경계추정 성능개선

김병창\*(대구가톨릭대), 김승원(포항공대), 이근배\*\*(포항공대)

## <차 례>

- |                                   |                 |
|-----------------------------------|-----------------|
| 1. 서론                             | 3.4 운율 경계 추정 과정 |
| 2. Conditional Random Fields(CRF) | 4. 운율 경계 추정 실험  |
| 3. 운율 경계 추정                       | 4.1 말뭉치         |
| 3.1 한국어의 특징                       | 4.2 성능 측정법      |
| 3.2 언어적 자질 선택                     | 4.3 실험 결과       |
| 3.3 CRF 훈련 과정                     | 5. 결론 및 향후 계획   |

## <Abstract>

### Improvements on Phrase Breaks Prediction Using CRF (Conditional Random Fields)

Byeongchang Kim, Seungwon Kim, Gary Geunbae Lee

In this paper, we present a phrase break prediction method using CRF(Conditional Random Fields), which has good performance at classification problems. The phrase break prediction problem was mapped into a classification problem in our research. We trained the CRF using the various linguistic features which was extracted from POS(Part Of Speech) tag, lexicon, length of word, and location of word in the sentences. Combined linguistic features were used in the experiments, and we could collect some linguistic features which generate good performance in the phrase break prediction. From the results of experiments, we can see that the proposed method shows improved performance on previous methods. Additionally, because the linguistic features are independent of each other in our research, the proposed method has higher flexibility than other methods.

\* Keywords: Text-to-Speech System, Prosody prediction, Phrase break prediction, Conditional random fields(CRF).

\* 본 논문은 2005학년도 대구가톨릭대학교 연구비지원에 의한 것임.

\*\* 본 논문은 한국과학재단(KOSEF) 특정기초연구(R01-2003-000-10181-0)지원으로 수행되었음.

## 1. 서 론

올바른 운율 경계는 문장 안에서 올바른 운율 구조를 이끌어내어, 자연스럽게 알아듣기 쉬운 음성을 만들어 주기 때문에 운율 경계를 추정하는 일은 TTS (Text-To-Speech) 시스템에서 중요한 작업 중의 하나이다. 또한 운율 경계는 악센트 추정[1]이나 음 길이를 추정[2]하는데 있어서도 유용한 정보로 사용된다. 잘못된 운율 경계는 단락 전체의 이해도를 떨어지게 만들 뿐만 아니라, 다른 운율 요소를 추정하는데 잘못된 정보를 제공해 나쁜 영향을 주게 된다. 유용한 정보를 갖고 있지 않는 자질을 운율 요소 추정에 사용하면 계산량이 늘어나고 성능이 떨어지게 된다.

음성 언어는 음운 구(Phonological Phrase), 억양 구(Intonation Phrase), 발화(Utterance)와 같은 운율 단위를 가지고 계층적 구조를 이루지만[3], 텍스트 언어에서는 운율 단위와는 다른 단어나 구 같은 문법 단위에 의해 구조가 결정된다. 그러나 일반적으로 문법적 정보는 운율 구조를 추정하는데 중요한 정보를 제공한다고 가정한다.

그 동안 운율 경계 추정을 위해 은닉 마코프 모델(Hidden Markov Models)[4], 결정 트리(Decision Tree)[5], 변환 률 기반의 학습[6], 순환 신경망(Recurrent Neural Network)[7], 유한 상태 변환기(Finite State Transducer)[8], 메모리 기반의 학습[9], 조건부 ME 모델[10] 등 다양한 방법들이 제안되었다. 그 밖에도 확률 방법만을 사용하면 발생하는 데이터 희소성 문제와 제한된 범위의 문맥 정보만을 다루기 때문에 발생하는 문제를 C4.5와 같은 결정 트리 방법을 이용하여 후처리로 에러를 정정하는 방법[11]을 사용하기도 하였다. 또한 운율 경계 추정에서 가장 기본으로 쓰이는 품사의 종류가 너무 많아 추정에 나쁜 영향을 준다고 생각하고 그러한 품사를 효율적으로 묶어 품사의 개수를 줄인 뒤 사용하는 방법[12]도 있었다.

본 연구에서는 운율 경계 추정 문제를 클래스 분류 문제로 변환하고, 형태소의 품사, 어휘, 어절의 길이, 문장에서의 어절 위치와 같은 다양한 언어적 정보를 CRF(Conditional Random Fields)[13]를 위한 자질로 사용했다. CRF를 사용하면 기존의 은닉 마코프 모델 기반의 방법보다 데이터의 희소성 문제에 대해 효과적이며, 국소최대(Local Maximum) 문제를 해결할 수 있다.

본 논문의 나머지 부분은 다음과 같이 구성된다. 2장에서 CRF에 대해 간단히 소개하고, 3장에서는 한국어의 특징과 CRF에 사용한 언어적 자질에 대해 설명한다. 4장에서 추출된 자질을 사용하여 운율 경계 추정에 CRF를 적용한 실험 결과를 보이며, 5장에서 본 방법론에 대한 결론과 향후 개선 방향에 대해 서술한다.

## 2. Conditional Random Fields(CRF)

CRF는 조건부 확률을 최대화하기 위해 훈련된, 방향성이 없는 그래프 모델이다 [13]. 선형 체인 CRF 모델은 유한 상태 기계에 대응되는 그래프 구조를 가지며, 연속 레이블링에 적합하다. 매개 변수  $\Lambda = \{\lambda, \dots\}$ 를 갖는 선형 체인 CRF는 입력 순열  $x = x_1 \dots x_T$ 가 주어진 상태 순열  $y = y_1 \dots y_T$ 에 대한 조건부 확률로 정의된다.

$$P_{\lambda}(y | x) = \frac{1}{Z_x} \exp\left(\sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t)\right) \quad (1)$$

$Z_x$ 는 모든 상태 순열의 확률의 합을 1로 만들어 주는 정규화 상수이고,  $f_k(y_{t-1}, y_t, x, t)$ 는 이진 값을 갖는 자질 함수이지만 실수 값도 가질 수 있다. 그리고  $\lambda_k$ 는 자질  $f_k$ 와 연관된 가중치를 학습한다. 자질 함수들은 현재 시간 단계  $t$ , 관측된 순열  $x$ , 상태 전이  $y_{t-1} \rightarrow y_t$ 의 어떤 양상을 측정할 수가 있다.

식 (1)에서 정의된 모델이 주어졌을 때, 입력 순열  $x$ 에 대해 가장 가능성이 높은 상태 순열은  $y^* = \underset{y}{\operatorname{argmax}} P_{\Lambda}(y | x)$ 로 나타낼 수 있고, 이는 비터비(Viterbi) 알고리즘을 사용한 동적 프로그래밍으로 쉽게 계산할 수 있다. 동적 프로그래밍 기반으로 상태의 주변 확률(Marginal Probability)이나 순열에서 각 위치의 전이를 계산 하는 것은 은닉 마코프 모델(Hidden Markov Models)과 매우 유사한 추론 과정이다.

매개 변수들은 상태 순열과 그것들에 각각 대응되는 주어진 입력 순열의 조건부 확률을 최대화하는 최대 유사도(Maximum Likelihood)에 의해 추정된다.

훈련 집합  $\{(x_i, y_i) : i = 1, \dots, M\}$ 의 로그 유사도(Log-Likelihood)는 다음과 같다.

$$\begin{aligned} L_{\Lambda} &= \sum_i \log P_{\Lambda}(y_i | x_i) \\ &= \sum_i \left( \sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t) - \log Z_{x_i} \right) \end{aligned} \quad (2)$$

각 자질의 실험상의 값과 모델  $P_{\Lambda}(y | x)$ 에 따라 예측된 값이 같다는 점에서 식 (2)를 최대화하는 것은 식 (3)을 만족하는 것과 일치한다.

$$\begin{aligned} & \sum_i \sum_t f_k(y_{t-1}, y_t, x_i, t) \\ &= \sum_i \sum_y P_A(y' | x_i) \sum_t f_k(y'_{t-1}, y'_t, x_i, t) \end{aligned} \quad (3)$$

CRF는 전체 최대값(Global Maximum)으로 수렴하는 학습을 보장하는 볼록 가능도 함수(Convex Likelihood Function)를 갖는 일반적인 최대 엔트로피 모델의 많은 장점들을 갖고 있으며, GIS나 IIS[14] 같은 전통적인 최대 엔트로피 학습 알고리즘으로 CRF를 훈련한다. 하지만 BFGS[15] 방법이 더 빨리 수렴하기 때문에 본 연구에서는 BFGS 방법을 사용하였다.

훈련 데이터에 대해 발생할 수 있는 과적합(Overfitting) 문제를 해결하기 위해 가우스 사전 평활(Gaussian Prior Smoothing)을 적용하면 로그 유사도 식 (2)에 대해 다음과 같은 별점을 주게 된다.

$$L_A = \sum_i \log P_A(y_i | x_i) - \sum_k \frac{\lambda_k^2}{2\sigma_k^2} \quad (4)$$

여기서  $\sigma_k^2$ 는 분산이며, 식 (4)을 최대화 하는 것은 식 (5)를 만족하는 것과 같다.

$$\begin{aligned} & \sum_i \sum_t f_k(y_{t-1}, y_t, x_i, t) - \frac{\lambda_k}{\sigma_k^2} \\ &= \sum_i \sum_y P_A(y' | x_i) \sum_t f_k(y'_{t-1}, y'_t, x_i, t) \end{aligned} \quad (5)$$

### 3. 운율 경계 추정

운율 경계는 TTS시스템에서 자소열-음소열 변환과 운율 추출에 많은 영향을 줄 수 있다. 운율 경계 앞뒤의 자소들은 서로 영향을 받지 않고 고유의 발음으로 변환된다. 또한, 억양 추출과 띄어읽기 등의 성능은 운율 경계 추정에서부터 시작한다.

지금까지 운율 경계를 추정하기 위해 두 가지 접근방법이 있었다. 첫 번째는 구문 구조와 운율 구조가 서로 관계가 있다는 가정 하에 구문 구조를 이용한 운율 경계 추정이었다. 이 방법은 실시간 구문 분석기와 구문-운율 변환기가 필요하며 대체적으로 많은 계산량을 요구한다.

두 번째는 형태소 품사열을 이용한 통계적인 방법이다. 운율 경계 추정을 위한

통계적 방법은 일반적으로 많은 양의 훈련 데이터를 필요로 한다. 이러한 훈련 데이터 부족 현상을 해소하기 위해 smoothing 기법을 사용하지만 그 현상을 완벽하게 해소할 수는 없다. 본 연구에서는 이러한 데이터 부족 현상을 극복하고, 국소 최대(Local Maximum) 문제를 해결하기 위해 CRF를 운율 경계 추정에 사용하였다 [6].

### 3.1 한국어의 특징

한국어 음성의 언어학적 특징은 다음과 같다[11]. 첫째, 한국어는 교착어이기 때문에 한국어의 단어는 분명한 형태소 경계를 갖는 하나 이상의 형태소로 구성된다. 둘째, 한국어는 많은 종류의 조사, 어미, 선어말어미를 갖는 후치사적 언어이다. 그러한 기능 형태소들은 명사의 역할, 동사의 시제, 조동사, 단어의 수식 관계를 결정한다. 위의 두 특징에 의해 형태소의 품사가 운율 경계 추정에 중요한 자질이 될 수 있다.

셋째, 한국어는 기본적으로 SOV(Subject Object Verb) 언어이지만 고정된 단어 순서를 갖는 영어와 같은 언어에 비해 문장의 끝에 동사가 반드시 나타나야 하는 제약을 제외하고는 상대적으로 단어의 순서가 자유롭다. 하지만 한국어에서도 법(Modalities)을 나타내는 조동사는 반드시 주 동사가 따라와야 한다는 단어 순서의 제약이 있다. 따라서 자질의 순서보다는 어떤 자질이 나타났는가가 더 중요한 정보가 될 수 있다.

넷째, 음운체계의 변화는 하나의 형태소 안, 어절 안의 형태소들 사이, 구 안의 어절들 사이에서만 일어날 수 있고 구들 사이에서는 일어나지 않는다. 따라서 정확한 운율 경계의 추정은 정확한 발음의 추정에 필수적이라 할 수 있다.

### 3.2. 언어적 자질 선택

본 연구에서 언어처리를 위한 품사체계는 세종 품사 45 분류[16]를 따랐으며, 운율 경계는 h3으로 대 운율 경계(Major Phrase Break)를, h2로 소 운율 경계(Minor Phrase Break)를 표기했으며, h1으로 운율 경계가 없는(No Phrase Break)것을 표기했다. <그림 1>은 본 실험을 위해 <그림 2>의 과정에 따라 품사와 운율 경계가 표기된 말뭉치의 예이다.

NNG<회답>
JKO<을>
h1
VV<통하>
ETM<ㄴ>
h2
NNG<문제>
h1
NNG<해결>
JKO<을>
h3
NNG<모색>
XSV<하>
EC<고>
h1
VX<있>
EF<습니다>
SF<.>

<그림 1> “회답을 통한 문제 해결을 모색하고 있습니다.”에 대해  
품사와 운율 경계가 표기된 말뭉치

본 논문의 실험에서는 CRF모델의 훈련과 사용을 위하여 다음과 같은 자질들을 말뭉치로부터 추출하여 사용하였다.

● 품사 자질

운율 경계 추정에서 가장 중요하며 기본이 되는 자질이다. 현재 어절 사이를 기준으로 앞으로 최대 7개, 뒤로 최대 7개의 형태소에 대해 품사 자질을 추출했다.

● 어휘 자질

어절의 어휘를 그대로 사용한 자질이며, 현재 어절 사이를 기준으로 앞으로 최대 2개, 뒤로 최대 2개의 어절에 대해 어휘 자질을 추출했다.

● 어절 길이 자질

어절의 길이는 어절의 음절 개수이며 이는 기본 발음 단위가 된다. 현재 어절 사이를 기준으로 앞으로 최대 2개, 뒤로 최대 2개의 어절에 대해 어절 길이 자질을 추출했다.

● 문장 위치 자질

문장의 시작 위치에서 현재 위치까지의 음절 개수와 현재 위치에서 문장의 끝

위치까지의 음절 개수를 자질로 사용하였으며, 현재 문장의 음절 개수로 해당 거리를 정규화를 한 자질도 사용했다.

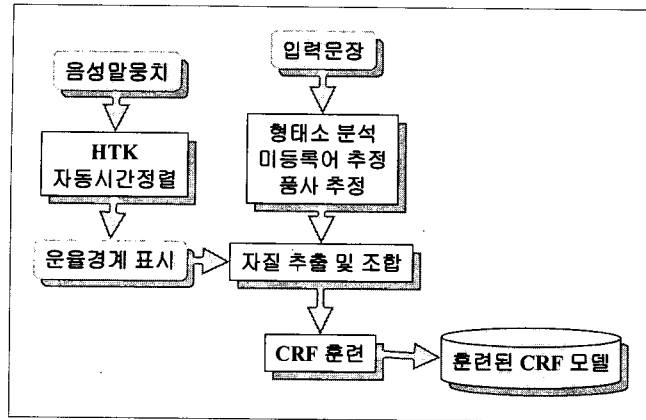
<표 1>은 본 연구에서 사용한 자질의 종류와 그 기호를 나타내고 있다. 자질기호 옆에 있는 숫자는 자질의 정보를 사용하는 주변 범위를 나타내며 어절 간의 공백을 기준으로 좌, 우 어절에 대해 정보를 사용했다.

<표 1> 운율 경계 추정에 사용된 자질들

자질 번호	자질 기호	자질 내용
1.1	POS ±1	앞, 뒤 형태소 품사 1개씩
1.2	POS ±3	앞, 뒤 형태소 품사 3개씩
1.3	POS ±5	앞, 뒤 형태소 품사 5개씩
1.4	POS ±7	앞, 뒤 형태소 품사 7개씩
2.1	Word ±1	앞, 뒤 어절 어휘 1개씩
2.2	Word ±2	앞, 뒤 어절 어휘 2개씩
3.1	Wlen ±1	앞, 뒤 어절 길이 1개씩
3.2	Wlen ±2	앞, 뒤 어절 길이 2개씩
4.1	Dis	문장에서 현재어절의 위치
4.2	Dis_nor	문장에서의 현재어절의 위치를 정규화

### 3.3 CRF 훈련

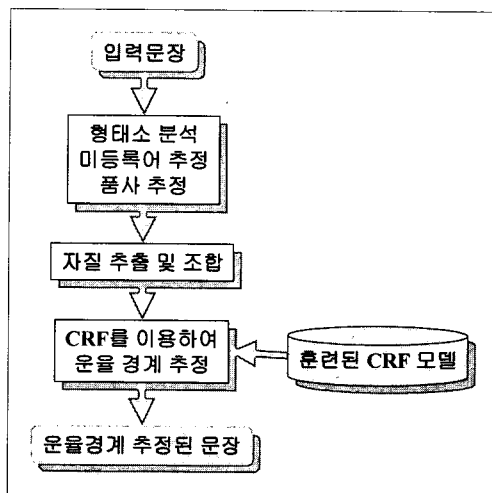
<그림 2>와 같이 운율 경계 추정에 사용되는 CRF 모델을 훈련시키기 위하여, 본 연구실에서 보유중인 언어분석 시스템을 사용하여 입력문장을 형태소분석, 미등록어 추정, 품사추정의 단계를 거치면서 기본적인 언어학적 정보를 추정하였다. 또한 입력문장을 발성한 음성말뭉치에 대해 HTK의 자동 시간 정렬 기능을 이용하여 운율경계를 표시하였다. 이 처리 결과로부터 <그림 1>과 같은 말뭉치를 얻을 수 있으며, 이 말뭉치에서 <표 1>의 자질들을 추출하고 조합하여 CRF 훈련에 사용하였다.



<그림 2> 운율 경계 추정을 위한 CRF 훈련

### 3.4 운율 경계 추정 과정

<그림 2>에서 훈련된 CRF 모델은, 입력문장에 대해 추출된 자질들과 함께 <그림 3>과 같은 과정을 거쳐 운율 경계를 추정하였다. 우선, <표 1>의 각 자질별로 단일 자질에 대한 CRF 모델을 훈련하고 성능을 측정하였다. 그 후, 각 자질들을 조합하고 각 조합별로 CRF 모델을 훈련하고 성능을 측정하였다. 이 실험 결과에서 어떤 자질의 조합이 본 방법에서 최적의 성능을 내는지 판단할 수 있었다.



<그림 3> 운율 경계 추정 과정



## 4. 운율 경계 추정 실험

### 4.1 말뭉치

본 연구에서는 1997년 10월 16일부터 2000년 3월31일까지의 MBC 뉴스 기사를 직접 녹음한 음성 파일과 텍스트를 사용하여 실험을 하였다. 말뭉치의 크기는 10,145 문장이며 122,021 어절로 구성되어 있고, 운율 경계 추정을 위해 말뭉치에는 품사와 운율 경계가 표시되어 있다. 위의 말뭉치에 표시된 품사 분류와 운율 경계 분류를 직접 수작업으로 하는 것은 시간이 많이 걸리는 작업이므로, 본 연구에서는 기존의 언어분석 시스템[17]으로 품사를 표시하고, HTK의 자동 시간 정렬 기능을 이용하여 운율 경계를 표시하였다. HTK의 자동 시간 정렬에서 문장내의 모든 공백에 대해 *pause*의 길이를 추정하였으며, *pause*의 길이가 70ms 이상이면 대운율 경계(h3), 70ms 미만이고 10ms 이상이면 소운율 경계(h2), 0ms이면 운율경계가 없는 것(h1)으로 구분하였다. 표시된 운율 경계에 대해 일부 문장을 수동으로 검증한 결과, 미세한 정렬에러는 있었지만 운율 경계에 대한 분류는 만족할 만한 수준이었다.

### 4.2 성능 측정법

운율 경계 추정에 대한 성능은 일반적으로 많이 사용하는 *break-correct*, *juncture-correct* [4]를 사용하여 성능을 평가하였다.

여기서 *N*은 *juncture*의 총 개수로 문장 내에서 운율 경계의 대상이 되는 모든 공백을 말한다. *B*는 운율 경계의 개수로 대운율 경계와 소운율 경계의 합이 된다. *D*는 실제적으로 문장에는 운율 경계가 표시되어 있지만 실험 결과에서는 운율 경계를 추정하지 못한 경우로 삭제 에러이다. *S*는 대운율 경계를 소운율 경계로 추정하거나 그 반대인 경우로 치환 에러이다. *I*는 실제적으로 문장에는 운율 경계가 없지만 실험 결과에서는 운율 경계를 추정한 경우로 삽입 에러이다.

$$BreakCorrect = \frac{B - D - S}{B} \times 100\% \quad (6)$$

$$JunctureCorrect = \frac{N - D - S - I}{N} \times 100\% \quad (7)$$

*Break-correct*는 추정해야 하는 운율 경계에 대해, 올바르게 추정한 개수의 비율이므로, 삽입 에러를 포함하지 않는다. *Juncture-correct*는 운율 경계의 대상이 되는 모든 공백에 대해, 운율 경계가 아닌 것을 포함해서 올바르게 추정한 개수의 비율

이다. 일반 문장에서 운율 경계의 대상이 되는 공백 중에 실제적으로 운율 경계가 아닌 것(h1)이 높은 비율을 차지하므로, 기준성능(baseline)이 높게 나올 수 있다.

따라서, 또 하나의 성능 측정 방법으로 운율 경계의 총 개수의 비율을 이용한 adjusted-score [18]를 사용하였다. 이는 운율 경계의 대상이 되는 공백에 대한 추정 결과가 모두 h1으로 되었을 때 기준성능이 0가 된다. 그리고 아무런 예러 없이 모든 경계에 대해 완벽하게 추정되었을 때는 1이라는 점수를 가지게 된다.

$$JC = \frac{Juncture\ Correct}{100} \quad (8)$$

$$NB = \frac{N - B}{N} \quad (9)$$

$$Adjusted\ Score = \frac{JC - NB}{1 - NB} \quad (10)$$

### 4.3 실험 결과

<표 2>는 <표 1>의 자질들을 단독으로 사용하여 CRF 모델을 훈련하고 적용한 결과이다. 본 논문에서는 모든 실험에 10-fold cross validation을 사용하였다.

<표 2> 단일 자질의 성능

사용한 자질번호	Break Correct(%)	Juncture Correct(%)	Adjusted Score
1.4	59.58	85.60	0.537
2.2	51.81	70.43	0.490
3.2	41.65	58.29	0.457
4.1	48.21	62.41	0.483

실험에 사용된 여러 가지 언어학적 자질중에 형태소의 품사가 가장 좋은 성능을 내는 자질로 판단되었다. 그리고 동일한 종류의 자질이라도 동시에 사용하는 형태소의 개수가 많을수록 더 높은 성능을 나타냈다. 단일 자질만을 사용해서 실험한 결과는 <표 3>과 비교해 보았을 때, 기존 연구의 확률적 방법과 결정트리를 사용한 후처리 방법의 성능보다 다소 떨어지는 현상을 보이고 있다[11].

&lt;표 3&gt; 확률적 방법과 결정트리를 사용한 추정 결과[11]

자질	Break Correct(%)	Juncture Correct(%)	Adjusted Score
확률적 방법을 사용한 추정	52.17	81.39	0.480
결정트리를 사용한 추정	51.77	81.71	0.488
확률적 방법과 결정 트리를 이용한 후처리	62.74	85.57	0.597

<표 4>는 단일 자질들을 조합하여 CRF 모델을 훈련하고 적용해서 얻은 실험 결과이다. 자질의 조합이 다수 존재하지만, 그 중 상위성능을 나타내는 조합을 구성해가는 과정만을 정리하였다.

&lt;표 4&gt; 최상의 자질 선택 결과

자질 조합	Break Correct(%)	Juncture Correct(%)	Adjusted Score
1.1	53.61	84.43	0.500
1.2	56.77	85.31	0.528
1.3	58.26	85.46	0.532
1.4	59.58	85.60	0.537
1.3 + 2.1	66.06	87.37	0.566
1.3 + 2.2	66.82	86.74	0.550
1.4 + 2.1	60.01	87.49	0.570
1.3 + 2.1 + 3.1	66.15	87.95	0.574
1.3 + 2.1 + 3.2	66.65	88.13	0.580
1.3 + 2.1 + 3.2 + 4.1	71.20	88.64	0.598
1.3 + 2.1 + 3.2 + 4.2	75.62	89.80	0.616

<표 4>의 실험결과에서 품사 자질은 좌, 우 5개의 품사를 사용하는 것 보다 7개의 품사를 사용하는 것이 성능이 더 좋으나, 어휘 자질과 조합되었을 때는 5개의 품사를 사용한 것이 더 좋은 성능을 나타냈다. 단어의 길이 자질은 좌 2개, 우 2개를 사용했을 때 가장 좋은 성능을 보이며, 문장에서 운율 경계의 위치를 나타내는 거리 자질은 성능에 많은 도움이 되고, 문장의 길이로 정규화를 하면 최고의 성능을 보이게 된다. 이러한 자질의 조합을 사용한 성능은 기존의 확률적 방법과 결정트리를 사용한 후처리 방법보다 우수한 성능을 나타내고 있다[11].

<표 5>는 훈련 데이터에 과적합되는 것을 막기 위해 <표 4>의 가장 마지막 실

험에서 운을 경계 추정에 가우스 평활(Gaussian Smoothing)을 적용한 결과를 보여 준다. Break-correct는 분산이 5일 때 가장 높으며, juncture-correct와 adjusted-score는 분산이 0.5일 때 가장 높다. 이 경우의 추정결과에 대한 confusion matrix는 <표 6>과 같다.

<표 5> 가우스 평활을 적용한 결과

가우스 분산	Break Correct(%)	Juncture Correct(%)	Adjusted Score
0.1	71.42	90.22	0.632
0.5	74.53	90.26	0.633
5	75.71	89.76	0.614
10	75.62	89.80	0.616
20	75.00	89.94	0.621

<표 6> <표 5>의 가우스 분산 0.5일 때의 confusion matrix

추정결과 운을경계	h1	h2	h3	총계
h1	93,151	182	3,757	97,090
h2	3,428	88	954	4,470
h3	4,405	146	26,055	30,606

## 5. 결론 및 향후 계획

본 논문에서는 한국어 TTS시스템에 필요한 운을 경계 추정을 위하여 CRF (Conditional Random Fields)를 사용하였다. 또한, 최상의 성능을 보이는 자질의 집합을 찾아내기 위하여, 여러 언어학적 자질들을 조합하여 사용하였다. 그 결과, 기존의 확률적 방법과 결정트리를 사용한 후처리 방법의 성능보다 개선된 성능을 얻을 수 있었다. 일반적으로 운을 경계 추정에 있어서 자질들은 서로 의존적인데 비해, 본 연구에서는 자질들을 서로 독립적으로 만들고 그 자질들을 조합하여 사용함으로써 다른 기계 학습 모델보다 호환성이 좋은 운을 경계 추정 시스템을 만들었다. 앞으로 성능향상을 위해서 구문 분석 결과와 같은 다양한 자질을 추가하고 규칙 기반의 에러 정정을 추가하여 성능을 높일 수 있는 방법을 연구할 계획이다.

## 참고 문헌

- [1] K. Ross, M. Ostendorf, "Prediction of abstract prosodic labels for speech synthesis", *Computer Speech and Language*, 1996.
- [2] J. P. van Santen, "Assignment of segmental duration in text-to-speech synthesis", *Computer Speech and Language*, 1994.
- [3] A. Steven, "Chunks and dependencies: bringing processing evidence to bear on syntax", *Computational Linguistics and Foundations for Linguistic Theory CSLI*, 1995.
- [4] P. Taylor, A. W. Black, "Assigning phrase breaks from part-of-speech sequences", *Computer Speech and Language*, Vol. 12, No. 2, pp. 99-117, 1998.
- [5] P. Koehn, S. Abney, J. Hirschberg *et al.* "Improving intonational phrasing with syntactic information", *Proc. ICASSP*, 2000.
- [6] C. Fordyci, M. Ostendorf, "Prosody prediction for speech synthesis using transformational rule-based learning", *Proc. ICSLP*, 1998.
- [7] Z. Ying, X. Shi, "An RNN-based algorithm to detect prosodic phrase for Chinese TTS", *Proc. ICASSP*, pp. 809-812, 2001.
- [8] A. Bonafonte, P. D. Aguero. "Phrase break prediction using a finite state transducer", *11th International Workshop, Advances in Speech Technology 2004*, Maribor, Slovenia, 2004.
- [9] E. Marsi, M. Reynaert, A. van den Bosch *et al.* "Learning to predict pitch accents and prosodic boundaries in Dutch", *Proceedings of the 41th Annual Meeting of the ACL*, 2003.
- [10] Y. Zheng, B. Kim, G. G. Lee, "Using multiple linguistic features for Mandarin phrase break prediction in maximum-entropy classification framework", *Proc. Interspeech 2004-ICSLP*, 2004.
- [11] J. Lee, B. Kim, G. G. Lee, "Automatic corpus-based tone and break-index prediction using K-ToBI representation", *ACM Transactions on Asian Language Information Processing*, pp. 207-224, 2002.
- [12] I. Read, S. Cox, "Using part-of-speech for predicting phrase breaks", *Proc. Interspeech 2004-ICSLP*, Vol 1, 2004.
- [13] J. Lafferty, A. McCallum, F. Pereira, "Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data", *Proc. ICML*, pp. 282-289, 2001.
- [14] S. Pietra, V. Pietra, J. Lafferty, "Inducing features of random fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 4, 1995.
- [15] R. Malouf, "A comparison of algorithms for maximum entropy parameter estimation", *Proc. of CoNLL-2002*, pp.49-55, 2002.
- [16] 세종21계획 국어기초자료구축분과, 형태 분석 말뭉치 구축 지침 ver 1.0, 국립국어연구원, 2003.
- [17] G. G. Lee, J. Cha, J. Lee, "Syllable pattern-based unknown morpheme segmentation and estimation for hybrid part-of-speech tagging of Korean", *Computational Linguistics*, Vol 28, No 1, pp.53-70, 2002.
- [18] E. Sanders, *Using Probabilistic Methods to Predict Phrase Boundaries for a Text-to-Speech*

*System*, Master's thesis, University of Nijmegen, 1995

접수일자: 2006년 2월 15일

게재결정: 2006년 3월 10일

▶ 김병창(Byeongchang Kim) : 교신저자

주소: 717-702 경상북도 경산시 하양읍 금락1리 330번지

소속: 대구가톨릭대학교 컴퓨터정보통신공학부

전화: 053-850-2718

E-mail: bckim@cu.ac.kr

▶ 김승원(Seungwon Kim)

주소: 790-784 경상북도 포항시 남구 효자동 산 31번지

소속: 포항공과대학교 컴퓨터공학과

전화: 054-279-5581

E-mail: rockzja@postech.ac.kr

▶ 이근배(Gary Geunbae Lee)

주소: 790-784 경상북도 포항시 남구 효자동 산 31번지

소속: 포항공과대학교 컴퓨터공학과

전화: 054-279-2254

E-mail: gblee@postech.ac.kr