

CRF를 이용한 한국어 운율 경계 추정*

*김승원 **김병창 *정민우 *이근배
*포항공과대학교 지능소프트웨어 연구실 **대구가톨릭대학교 컴퓨터정보통신공학부
*{rockzja, stardust, gblee}@postech.ac.kr **bckim@cu.ac.kr

Using CRF (Conditional Random Fields) to Predict Phrase Breaks in Korean

*Seungwon Kim **Byeongchang Kim *Minwoo Jeong *Gary Geunbae Lee
*Dept. of CSE, POSTECH **School of Computer & Information Communication
Engineering, Catholic University of Daegu

요 약

본 논문은 한국어 TTS (Text-To-Speech)에서 운율 경계를 추정하는 문제를 클래스 분류 문제로 보고 CRF (Conditional Random Fields)를 적용하여 운율 경계를 추정 하였다. 우리는 품사와 운율 경계로 구성된 말뭉치를 사용하여 품사, 어휘, 단어의 길이, 문장에서의 단어 위치와 같은 다양한 속성의 언어적 자질을 추출하여 CRF를 훈련시켰으며, 자질들을 서로 조합하여 최고의 성능을 보이는 자질 집합을 골랐다. 또한 가우스 평활 (Gaussian Smoothing)을 적용하여 데이터의 희소성 문제를 줄였다. 실험 결과에서 본 방법이 기존의 방법보다 성능이 좋을 뿐만 아니라 운율 경계를 추정하기 위한 자질을 독립시켰기 때문에 다른 시스템과의 호환성도 높다는 것을 알 수 있었다.

1. 서 론

올바른 운율 경계는 문장 안에서 운율 구조를 이끌어 내어 자연스럽게 알아 듣기 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 잡음이 들어가 성능이 떨어진다는 것은 잘 알려진 사실이다.

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 우리는 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 룰 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.

우리는 운율 경계 추정 문제를 클래스 분류 문제로 보고 단어의 품사, 어휘, 단어의 길이, 문장에서의 단어 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

*본 연구는 한국과학재단 (KOSEF) 특장기초연구 (R01-2003-000-10181-0)지원으로 수행되었음.

Maximum)문제를 해결할 수 있다.

본 논문의 나머지 부분은 다음과 같이 구성된다. 2절에서 CRF에 대해 간단히 소개하고, 3절에서 한국어의 특징과 CRF에 사용한 언어적 자질에 대해 설명한다. 4절에서 추출된 자질을 사용하여 운율 경계 추정에 CRF를 적용한 실험 결과를 보이며, 5절에서 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 적당하다. 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 λ_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 $y^* = \arg \max_y P_{\Lambda}(y|x)$ 로 나타낼 수 있고, 이는 비터

비 (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.

$$L_{\Lambda} = \sum_i \log P_{\Lambda}(y_i|x_i) \quad (2)$$

$$= \sum_i \left(\sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t) - \log Z_{x_i} \right)$$

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.

$$\sum_i \sum_t f_k(y_{t-1}, y_t, x_i, t) \quad (3)$$

$$= \sum_i \sum_{y'} P_{\Lambda}(y'|x_i) \sum_t f_k(y'_{t-1}, y'_t, x_i, t)$$

CRF는 전체 최대 값 (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 우리는 BFGS 방법을 사용하였다.

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

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

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

$$\sum_i \sum_t f_k(y_{t-1}, y_t, x_i, t) - \frac{\lambda_k}{\sigma_k^2} \quad (5)$$

$$= \sum_i \sum_{y'} P_{\Lambda}(y'|x_i) \sum_t f_k(y'_{t-1}, y'_t, x_i, t)$$

3. 운율 경계 추정

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. 셋째, 한국어는 기본적으로 SOV 언어지만 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. 넷

째, 음운체계의 변화는 하나의 형태소 안, 하나의 단어 안의 형태소들 사이, 하나의 구 안에 단어들 사이에서만 일어날 수 있고 구들 사이에서는 일어나지 않는다.

3.2 언어적 자질 선택

품사는 세종 품사 45 분류대로 분류를 했으며 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 품사와 운율 경계가 표기된 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> 품사와 운율 경계가 표기된 예

기존의 언어 분석 시스템을 이용하여 다음과 같은 자질들을 추출했다.

- 품사 자질

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

- 어휘 자질

단어의 어휘를 그대로 사용한 자질이며 어절 사이를 기준으로 앞으로 최대 2개, 뒤로 최대 2개의 단어에 대해 어휘 자질을 추출했다.

- 단어 길이 자질

단어의 길이는 단어의 음절 개수이며 이는 기본 발음 단위가 된다. 어절 사이를 기준으로 앞으로 최대 2개, 뒤로 최대 2개의 단어에 대해 단어 길이 자질을 추출했다.

- 문장 위치 자질

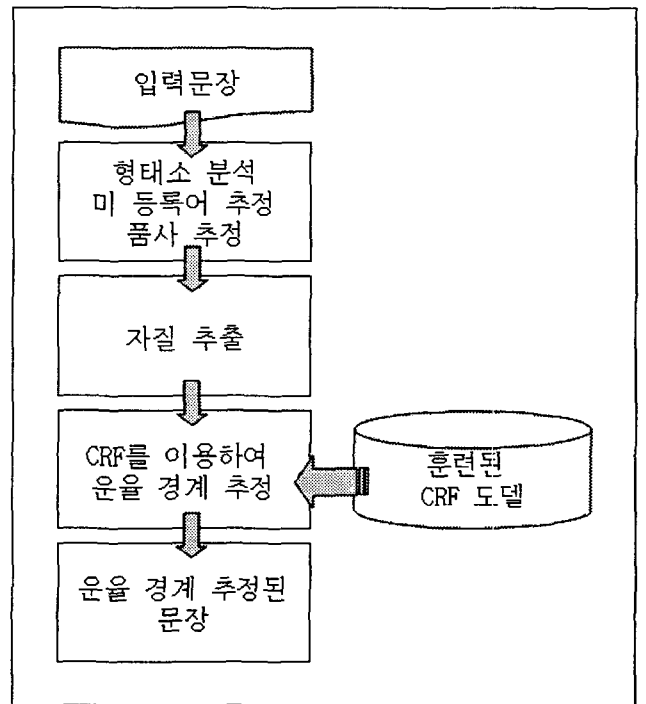
문장의 시작 위치에서 현재 위치까지의 음절 개수와 현재 위치에서 문장의 끝 위치까지의 음절 개수를 자질로 사용하였으며 현재 문장의 음절 개수로 해당 거리를 정규화를 한 자질도 사용했다.

<표 1>은 사용한 자질의 종류이다.

<표 1> 운율 경계 추정에 사용된 자질들

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

추출된 자질들과 CRF를 이용하여 <그림 2>처럼 운율 경계를 추정하였다.



<그림 2> 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 달려있다. 위의 말뭉치에서 품사 분류와 운율 경계 분류를 수작업으로 직접 하는 것은 시간이 많이 걸리므로 우리는 기존의 형태소 분석기 [16]로 품사를 달고 HTK의 자동 시간 정렬 기능을 이용하여 운율 경계를 달았

다.

4.2 성능 측정법

운율 경계 추정에 대한 성능은 하나의 측정 방법으로 평가할 수 없기 때문에 우리는 break-correct, juncture-correct [4] 방법으로 성능을 평가하였다.

여기서 N은 juncture의 총 개수로 문장 내에 어떤 종류의 운율 경계라도 담고 있는 모든 공백을 말한다. B는 운율 경계의 개수로 대 운율 경계와 소 운율 경계의 총 개수만을 말한다. D는 정답 문장에는 운율 경계가 달려 있지만 실험 문장에서는 운율 경계를 추정하지 못한 경우로 삭제 에러이다. S는 대 운율 경계를 소 운율 경계로 추정하거나 그 반대인 경우로 치환 에러이다. I는 정답 문장에는 운율 경계가 없지만 실험 문장에서는 운율 경계를 추정한 경우로 삽입 에러이다.

$$Break_Correct(B_C\%) = \frac{B - D - S}{B} \times 100\%$$

$$Juncture_Correct(J_C\%) = \frac{N - D - S - I}{N} \times 100\%$$

우리는 또 하나의 성능 측정 방법으로 운율 경계의 총 개수의 비율을 이용한 adjusted-score [17]를 사용하였다.

$$JC = \frac{Juncture_Correct}{100}$$

$$NB = \frac{N - B}{N}$$

$$Adjusted_Score(A_S) = \frac{JC - NB}{1 - NB}$$

4.3 실험 결과

우리는 CRF를 이용하여 운율 경계를 추정한 성능을 보기 위해 다음과 같이 실험을 했다. 앞서 선택한 자질들에 대하여 각각 단일 성능을 측정한 뒤 가장 좋은 성능을 내기 위해 자질들을 조합하여 최상의 자질의 집합을 선택하였다. <표 2>는 한가지 자질만 CRF에 적용한 결과이며 모든 실험에 10-fold cross validation을 사용하였다.

<표 2> 단일 자질의 성능

자질	B_C%	J_C%	A_S
POS ±7	59.58	85.60	0.537
Word ±2	51.81	70.43	0.490
Wlen ±2	41.65	58.29	0.457
Dis	48.21	62.41	0.483

<표 3>는 자질들을 조합하여 CRF에 적용함으로써 최고의 성능을 보이는 자질의 집합을 찾은 결과이다. 자질 옆에 있는 숫자는 자질의 정보를 본 주변 범위를 나타내며 어절 간의 공백을 기준으로 좌, 우 단어에 대해 정보를 보았다.

<표 3> 최상의 자질 선택 결과

자질 조합	B_C%	J_C%	A_S
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

실험 결과를 살펴 보면 품사 자질은 좌, 우 5개의 품사를 보는 것 보다 7개의 품사를 보는 것이 성능이 더 좋으나 어휘 자질과 조합되었을 때 5개 만을 사용한 것이 더 좋게 나온다. 단어의 길이 자질은 좌 2개, 우 2개를 보았을 때 가장 좋은 성능을 보이며, 문장에서 운율 경계의 위치를 나타내는 거리 자질은 성능에 많은 도움이 되고, 문장의 길이로 정규화를 하면 최고의 성능을 보이게 된다.

<표 4>은 훈련 데이터에 과적합 되는 것을 막기 위해 운율 경계 추정에 가우스 평활 (Gaussian Smoothing)을 적용한 결과를 보여준다. break-correct는 분산이 5일 때 가장 높으며, juncture-correct와 adjusted-score는 분산이 0.5일 때 가장 높다.

<표 4> 가우스 평활을 적용한 결과

가우스 분산	B_C%	J_C%	A_S
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

5. 결론 및 향후 계획

우리는 CRF (Conditional Random Fields)를 이용하여 한국어에서 운율 경계 추정을 하였다. 여러 개의 자질들을 조합하여 사용함으로써 최상의 성능을 보이는 자질의 집합을 찾아 냈으며, 일반적으로 운율 경계 추정에 있어서 자질들은 서로 의존적인데 비해 우리는 자질

들을 서로 독립적으로 만들고 그 자질들을 조합하여 사 용함으로써 다른 기계 학습 모델보다 호환성이 좋은 운 울 경계 추정 시스템을 만들었다. 앞으로 성능향상을 위해서 구문 분석 결과와 같은 다양한 자질을 추가하고 규칙 기반의 에러 정정을 추가하여 성능을 높일 수 있 는 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, M. Collins. Improving intonational phrasing with syntactic information, In *ICASSP*, 2000.
- [6] C. Fordyci, M. Ostendorf. Prosody prediction for speech synthesis using transformational rule-based learning, *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1998.
- [7] Z. Ying, X. Shi. An RNN-based algorithm to detect prosodic phrase for Chinese TTS, *ICASSP*, pp.809-812, 2001.
- [8] A. Bonafonte, P. D. Agüero. 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, W. Daelemans, V. Hoste. 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, *Proceedings of the 8th international conference on spoken language processing (interspeech2004-ICSLP)*, 2004.
- [11] B. Kim, G. G. Lee. Decision-tree based error correction for statistical phrase break prediction in Korean, *Proceedings of the 18th international conference on computational linguistics (coling2000)*, pp.1051-1055, 2000.
- [12] I. Read, S. Cox. Using Part-Of-Speech For Predicting Phrase Breaks, *Proceedings of 8th International Conference on Spoken Language Processing*, Vol 1, 2004.
- [13] J. Lafferty, A. McCallum, F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proceedings of International Conference on Machine Learning, ICML-01*, 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, *Proceeding of CoNLL-2002*, pp.49-55, 2002.
- [16] 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.
- [17] E. Sanders. Using Probabilistic methods to predict phrase boundaries for a text-to-speech system. Master's thesis, University of Nijmegen, 1995.