

운율구 추출 및 음소 지속 시간의 트리 기반 모델링

Tree-based Modeling of Prosodic Phrasing and Segmental Duration

이 상 호*, 오 영 환*

(Sang Ho Lee*, Yung Hwan Oh*)

* 본 연구는 한국과학기술원 인공지능연구센터의 지원에 의해서 이루어졌습니다.

요 약

본 논문에서는 한국어 TTS 시스템을 위한 운율구 추출, 운율구 사이의 휴지 기간, 음소의 지속 시간 모델링 방법을 설명한다. 실험을 위해 여러 장르로 구성된 400문장을 선정하고, 이를 전문 여성 아나운서가 발성하였다. 녹음된 음성 신호에 대해 음소 및 운율구 경계를 결정하고, 문장에 대해서는 형태소 분석, 발음표기 변환, 구문 분석을 수행하였다. 400문장(약 33분) 중 240문장(약 20분)을 이용하여 결정 트리 및 회귀 트리를 학습시킨 후, 160문장(약 13분)에 대해 실험하였다. 운율 모델링을 위한 특징들이 제안되었고, 학습된 트리들을 해석함으로써 특징들의 유효성이 평가되었다. 실험 문장에 대해 운율구 경계의 유무를 결정하는 결정 트리의 오류율은 14.46%이었고, 운율구 사이의 휴지 기간과 음소 지속 시간을 예측하기 위한 회귀 트리들의 평균 제곱 오류율(RMSE)이 각각 132 msec, 22 msec이었다. 수집된 모든 자료(400문장)로 학습한 결과, 운율구 경계 결정 오류율, 휴지 기간 및 지속 시간 RMSE의 10-fold cross-validation 추정치가 각각 13.77%, 127.91 msec, 21.54 msec이었다.

ABSTRACT

This study describes the tree-based modeling of prosodic phrasing, pause duration between phrases and segmental duration for Korean TTS systems. We collected 400 sentences from various genres and built a corresponding speech corpus uttered by a professional female announcer. The phonemic and prosodic boundaries were manually marked on the recorded speech, and morphological analysis, grapheme-to-phoneme conversion and syntactic analysis were also done on the text. A decision tree and regression trees were trained on 240 sentences (of approximately 20 minutes length), and tested on 160 sentences (of approximately 13 minutes length). Features for modeling prosody were proposed, and their effectiveness was measured by interpreting the resulting trees. The misclassification rate of the decision tree was 14.46%, RMSEs of the regression trees, which predict pause duration and segmental duration, were 132 msec, 22 msec respectively for the test set. As a result of training trees on 400 sentences, the 10-fold cross-validation estimates of the misclassification rate and the RMSEs were 13.77%, 127.91 msec, and 21.54 msec respectively for prosodic phrasing prediction, pause duration, and segmental duration.

I. 서 론

운율이란 분절음(segmental sound)에 실려 그 실체를 나타내는 초분절적 자질(suprasegmental feature)로서 크게 음의 강약, 고저, 장단, 경계로 나뉜다[6]. 그러므로, 문서 음

성 변환 시스템(text-to-speech system)은 이 네가지 운율을 제어하는 운율 생성 모듈을 포함하게 되며, 운율 생성 모듈의 성능이 합성음의 자연성을 좌우하게 된다[24].

일반적으로 운율 생성 문제는 운율구 추출, 억양 레이블(intonation label) 결정과 같은 패턴 분류(pattern classification) 문제와 억양, 지속 시간 예측과 같은 함수 근사(functional approximation) 문제를 포함하게 된다. 그러므로 여러 연구자들은 신경회로망[17], 은닉 마르코프

* 한국과학기술원 전산학과
접수일자: 1998년 5월 21일

모델[18], 트리 기반 모델[10] 등을 운율구 추출[14,26], 음소 지속 시간 예측[21], 억양 레이블 결정[22], 기본 주파수 예측[19,25] 등에 사용하여 왔다. 특히 기존의 패턴 인식 모델 중 트리 기반 모델은 특징 값으로 카테고리 변수(categorical variable)와 실변수(real-valued variable)를 동일한 시각에서 모두 처리할 수 있고, 결정 트리(decision tree) 혹은 회귀 트리(regression tree)를 사람들이 쉽게 해석할 수 있다는 장점이 있기 때문에, 다른 여러 분야에서도 적용되고 있다[23].

본 연구에서는 트리 기반 모델링 기법을 이용하여 위의 네가지 운율 중에서 운율구 추출(prosodic phrasing) 및 운율구 사이의 휴지 기간(pause duration), 음소 지속 시간(segmental duration)을 예측하고자 한다. 한국어 운율구 추출 및 음소 지속 시간에 관한 기존의 연구로는 문장 성분의 바이그램(bigram)을 이용한 운율구 추출 모델링[1], 회귀 트리에 의한 음소 지속 시간 모델링[5], 선형 회귀 모델에 의한 음소 지속 시간 모델링[2] 등이 있다.

본 논문에서 사용한 트리 기반 모델링 방법은 Gini 인덱스와 평균 제곱 오류를 이용하여 노드의 불순도(impurity)를 측정하고[10], Chou의 알고리즘에 의해 노드를 분할한 후[11], cost-complexity pruning 방법에 의해 트리를 선택한다[10]. 다양한 장르로 구성된 400문장에 대해 결정 트리 및 회귀 트리를 구하여 한국어 TTS 시스템의 운율 생성 모듈로 이용하고자 한다.

본 논문의 구성은 다음과 같다. 다음 장에서는 트리를 생성하기 위해 사용된 문장과 음성 코퍼스의 구성 및 특징에 대해 알아보고, 트리 생성 방법과 본 논문에서 사용하는 표기법에 대해 간단하게 설명한다. III, IV장에서는 각각 운율구 추출, 휴지 기간 예측, 음소 지속 시간 예측을 위한 특징 값들을 분석하고, 이를 이용하여 얻어진 트리들의 성능에 대해 알아본다. 400문장으로 학습된 결정 트리 및 회귀 트리의 최종 성능을 V장에서 설명하고, VI장에서 결론을 맺는다.

II. 코퍼스 및 트리 기반 모델링

2.1 코퍼스 구성 및 분석

TTS 시스템은 어떠한 제약도 가지 않은 입력 문장에 대해 합성음을 출력하여야 하기 때문에 운율 생성 모듈을 학습시키기 위한 코퍼스는 여러 장르에서 발췌된 문장들로 구성되어야 할 것이다. 이를 위해 본 연구에서 사용한 코퍼스는 초등학교 교과서, 논문 요약, 소설, 영화 해설 등에서 발췌되었으며 단문, 복문, 평서문, 의문문, 감탄문 등 다양한 형태의 문장들로 이루어진 400문장(3,724어절)이다. 이를 전문 여성 아나운서가 방음실에서 발성하였고, 이를 녹음하여 약 33분 분량의 음성 코퍼스를 구할 수 있었다.

문장 코퍼스는 기계발된 문서 분석기를 이용하여[7],

형태소 분석, 발음표기 변환, 구문 분석을 수행하고 분석 오류를 수정하였다. 형태소 분석에 사용된 품사 집합의 크기는 심벌을 포함해서 총 52개이나 발음표기로 변환한 후의 음소에 할당되는 품사는 표 1에 보이는 23개를 사용한다. 즉, 실질 형태소에 해당하는 체언과 용언 등을 세 분류하지 않는다. 본 연구에서 사용한 문서 분석기는 구문 분석시 어절 품사를 사용하게 되는데, 어절 품사는 어절내의 좌측과 우측에 위치하는 품사들의 쌍으로 표현된다. 또한 연속되는 품사들의 역할이 구문적으로 다른 의미를 가질 경우에는 하나의 품사로 축약되는 경우가 있다. 예를 들어, "공부할 때다"의 형태소 분석 결과는 "공부/nc+하/xv+ㄹ/ed 때/nc+이/pp+다/ef"이지만 "공부할"에서 "공부"와 "하"가 더해져 용언의 의미를 갖게 되어 어절 품사는 "vbed"이고, 좌품사와 우품사는 각각 "vb", "cd"가 된다[7]. 따라서, 어절의 좌품사로 사용되는 품사는 체언, 용언, 관형사, 부사, 감탄사 5개뿐이며, 우품사로 사용되는 품사는 용언, 서술격조사, 선어말어미, 동사 파생접미사, 형용사 파생접미사, 부사 파생접미사를 제외한 17개이다. 한편, "때다"의 경우, 발음 표기는 그대로 "때다"이므로 음소의 품사는 "ㄷ/nc+ㅏ/nc+ㄷ/ef+ㅏ/ef"가 된다. 이와 같이 음소에 품사를 할당할 때는 한국어의 품사 접속 규칙에 어긋나는 현상이 발생된다.

표 1. 품사 정의표

기호	품사	예
nc	체언	학교, 나, 뽕
vb	용언	벌다, 높다
dn	관형어	새, 인, 무슨, 흰
ad	부사어	그리고, 마치
it	독립어	어머니, 앓
ps	주격조사	이, 기
po	목적격조사	-을, -를
pd	관형격조사	-의
pv	호격조사	아, 이
pa	부사격조사	-에게, -로
pp	서술격조사	-이-
pt	주제격조사	-은, -는
px	보조사	-도, -조차
pc	접속조사	-나, -와
ec	인접어미	며, -이
ex	보조격 인접어미	지, 고
ed	관형사형 전성어미	는, -ㄴ, -ㄹ
en	명사형 전성어미	은, -기
ep	선어말어미	-았-, -겠-
ef	어받어미	-나, -리
xv	동사 파생접미사	-되-, -시키
xj	형용사 파생접미사	-답-, -스님
xa	부사 파생접미사	-개-, -이

본 연구에서 사용한 구문 구조의 표현 방식은 어절간의 의존 관계를 표현하는 의존 트리이다. 의존 트리는 어절과 어절간의 관계를 지배소와 의존소 관계로 표현하고, 한국어의 경우는 지배소가 항상 의존소 뒤에 나타난다는 지배소 후위의 법칙을 가정하여 좀 더 제한된 구문 형태

로 표현한다[3,4]. 구문 구조의 예를 들면, “학교에서 우리가 쓰는 것을 알아봅시다”의 경우 “((학교/nc+에서/pa (우리/nc+가/ps 쓰/vb+는/ed)) 것/nc+을/po) 알/vb+아/ex+보 /vx+다시다/ef)”로 표현할 수 있는데, 여기서 괄호안의 우측 어절이 좌측 어절을 지배하고, 괄호안의 대표 어절은 우측 어절이 된다. 다시 말해, “쓰는”은 “학교에서”와 “우리가”의 지배소가 되며, “학교에서”와 “우리가”는 “쓰는”의 의존소가 된다.

음성 코퍼스는 자동 음소 분할 프로그램을 이용하여 음소별로 분할하고[8], 분할 오류를 수정한 후 운율구 경계를 표시하였다. 운율구 경계는 청각적으로 발화의 끊김을 지각하는 부분을 표시하였으며[1], 억양의 급변화, 어절내 마지막 음절의 장음화, 휴지(pause)의 삽입 등이 음성 신호에 나타나게 된다. 운율구 경계의 유형은 어절과 어절 사이에만 나타나는 것으로 가정하고, 경계의 유무 두 가지만을 사용하였다. 휴지가 존재하지 않을 경우는 휴지 기간을 0 msec로 가정하였으며, 문장과 문장 사이의 휴지를 제외한 나머지 휴지들에 대해 기간을 측정하였다. 표 2에 나타난 44개의 음소를 사용하여 음성 코퍼스로부터 24,531개의 음소를 얻었고, 문장의 마지막 어절을 제외한 운율구 경계는 1,091개가 발생되었다.

표 2. 음소 기호표

조음 방법	음소(기호)
이중모음	야(ya), 예(yv), 예(yc), 와(wa), 웨(wE)
	요(yo), 워(wv), 유(yu), 와(Wi)
단모음	아(a), 에(E), 어(v), 에(e), 오(o)
	외(we), 우(u), 위(wi), 오(U), 이(i)
파찰음	ㄱ(j), ㄷ(C), ㄸ(c)
비찰음	ㄱ(s), ㅅ(S), ㅇ(h)
파열음	ㅂ(b), ㅃ(P), ㅍ(p), 종성 ㅂ(b0)
	ㄷ(d), ㅌ(T), ㅊ(t), 종성 ㄷ(d0)
	ㄱ(g), ㅋ(K), ㆁ(k), 종성 ㄱ(g0)
비음	ㅁ(m), ㄴ(n), 종성 ㅇ(N)
	종성 ㅁ(m0), 종성 ㄴ(n0)
유음	ㄹ(r), 종성 ㄹ(l)

음성 코퍼스로부터 얻어진 음소들의 지속 시간 분포 및 휴지 기간의 통계적 특성은 그림 1, 2에 V표식 상자수염 그림(notched box-and-whisker plot)과 히스토그램을 통해 나타내었다. 그림 2의 휴지 기간에 대한 히스토그램을 보면 50 msec 이하의 휴지 기간도 상당히 관측됨을 알 수 있다. 이는 발화의 끊김을 지각함에 있어서 휴지의 삽입 보다는 운율구 마지막 음절의 장음화 현상에 더 의존한다는 것을 의미한다[27]. 음소의 평균 지속 시간, 표준 편차는 68 msec, 38 msec이었고, 휴지 기간의 경우는 각각 168 msec, 176 msec이었다.

본 논문에서 사용하는 트리 기반 모델링 기법의 성능을 검증하기 위해서, 400문장을 240문장의 학습 데이터와 160문장의 실험 데이터로 나누었다. 각 데이터들은 여러

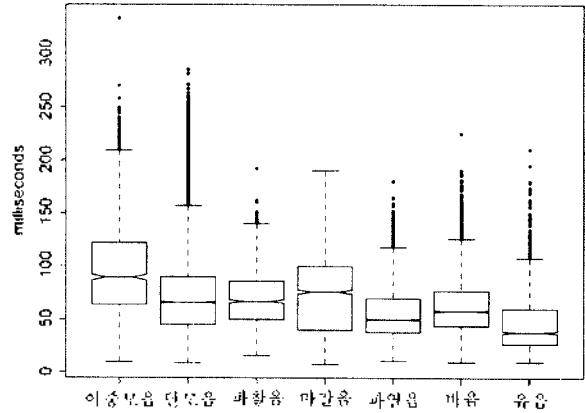


그림 1. 음소 클래스에 따른 음소 지속 시간의 V표식 상자수염 그림

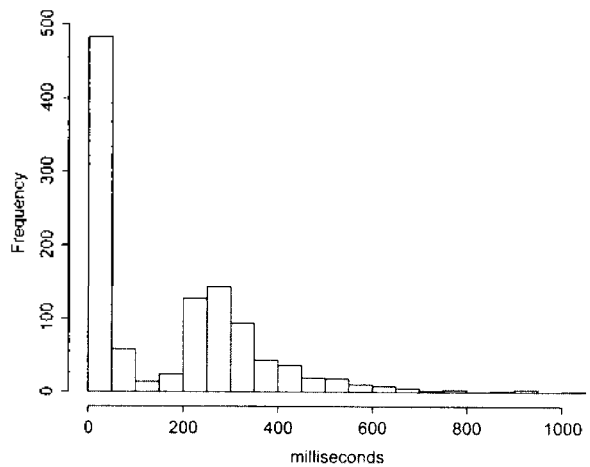


그림 2. 휴지 기간의 히스토그램

장르를 동일한 비율로 포함하도록 나뉘었고, 학습 데이터와 실험 데이터의 통계 특성이 전체 코퍼스의 통계 특성과 거의 일치하였다.

2.2 트리 기반 모델링

비모수적 통계 모델링 방법인 트리 기반 모델링 기법은 트리의 예측 오류를 최소화하는 방향으로 특징 벡터 x 의 공간을 연속적으로 나누는 기법으로[10], 예측될 변수 y 가 카테고리 변수인 경우와 실변수인 경우, 각각에 대해 결정 트리와 회귀 트리를 생성하게 된다. 트리 기반 모델링 방법의 장점들로는 특징 벡터 $x = (x_1, x_2, \dots, x_k)$ 를 이루는 각각의 원소 x 들의 변수 타입에 무관하게 동일한 틀 안에서 학습이 이루어진다는 점과 학습된 트리의 해석이 매우 용이하다는 점이 있다.

본 연구에서 사용하는 트리 기반 모델링 기법은 CART (classification and regression trees)를 기반으로 한다. 트리 생성의 기본 절차인 노드의 분할과정과 트리 제거 방법은

각각 Chou의 분할 알고리즘과[11] CART의 cost-complexity pruning 방법을 적용하였다[10]. 노드의 불순도를 나타내는 함수로는 결정 트리일 경우는 Gini 인덱스를 사용하고, 회귀 트리일 경우는 평균 제곱 오류를 사용한다. 최적 트리를 찾기 위해 10-fold cross-validation 방법이 사용되며, 트리의 선택 방법은 OSE (0 standard error)법과 ISE 법을 모두 적용하였다. OSE 법은 오류의 cross-validation 추정치 $R^0(T)$ 가 최소인 트리 T 를 선택하는 방법이고, ISE 법은 추정치 표준 오류값 $SE(R^0(T))$ 를 $R^0(T)$ 에 더한 $R^0(T) + SE(R^0(T))$ 보다 작은 오류를 갖고 트리 노드의 개수가 최소인 트리를 선택하는 방법이다¹⁾. CART에서는 ISE 법을 제안하고 있지만[10], 최근의 연구 결과에 의하면 결정 트리를 구하는 실험에서 ISE 법은 OSE 법보다 성능이 떨어진다고 보고되었다[13]. 하지만, ISE 법에 의해 구해진 트리의 크기가 더 작고 트리의 해석이 용이하므로, 다음 상부터 트리의 해석 부분은 대부분 ISE 법에 의해 구해진 트리를 설명할 것이다.

본 논문에서 사용하는 표현법은 다음과 같다. 구해진 트리와 트리의 노드는 각각 T 와 t 로 표현하고, 트리의 단말 노드 집합은 \bar{T} , 단말 노드의 개수는 $|\bar{T}|$ 로 표현한다. 학습 혹은 실험에 사용된 N 개의 데이터에 대해 오류율 $R(T)$ 는 결정 트리일 경우는 $R(T) = \frac{1}{N} \sum_{t \in \bar{T}} M(t)$ ($M(t)$ 는 노드 t 에서 잘못 분류된 데이터 개수)이고, 회귀 트리일 경우

는 $R(T) = \frac{1}{N} \sum_{n=1}^N (y_n - d(x_n))^2$ ($d(x)$ 는 입력 특징 벡터

x 에 대한 트리 T 의 출력값)이다. 이 오류율들이 cross-validation에 의해 구해졌을 경우는 $R^0(T) \pm SE(R^0(T))$, 이를 줄여서 $\hat{R}(T)$ 로 표현하고, 실험 데이터에 적용한 후 구한 값이면 $R^*(T)$ 로 표현한다. 한편, 회귀 트리에 대한 실험에서는 자료의 분산 정도를 얼마나 감소시켰는지를 측정하기 위한 상대 평균 제곱 오류(relative mean squared error) $RE(T) = R(T)/R(\mu)$ ($R(\mu) = \frac{1}{N} \sum_{n=1}^N (y_n - \mu)^2$, $\mu = E\{y\}$)와, 예측치와 실제치간의 상관 계수(correlation coefficient) r 도 이용된다. 이 값들도 실험 데이터에 적용한 후 얻은 값이면 각각 $RE^0(T)$ 와 $r^0(T)$ 로 표현한다.

III. 운율구 추출 및 휴지 기간 예측

본 장에서는 운율구를 추출하고, 추출된 운율구 간의 휴지 기간을 예측하고자 한다. 3.1절에서는 문장으로부터 추출된 정보와 관측 어절 앞에서 발생한 운율구 경계를

이용하여, 관측 어절이 운율구의 마지막 어절인지를 예측해 본다. 이때 사용되는 이전 운율구 경계 정보는 수동으로 결정된 것을 사용하므로, TTS 시스템이 실행될 때에는 얻을 수 없는 정보이다. 물론, 결정 트리에 의한 결과를 이용할 수 있으나, 결정 트리의 결과에는 오류를 포함하므로 이를 그대로 이용하게 되면 오류의 전파에 의하여 더 높은 오류율을 유발시킬 가능성이 있다. 그러므로, 3.2절에서는 이전 운율구 경계 정보를 이용하지 않고, 문장으로부터 추출된 정보만을 이용하여 운율구를 추출하고, 휴지 기간을 예측한다.

본 장에서 사용되는 특징 벡터는 운율구를 추출하는 결정 트리와 운율구 사이의 휴지 기간을 예측하는 회귀 트리에 동일하게 이용되며, 사용되는 특징 값의 유효성을 실험에 앞서 알아본다. 특징 값과 예측 값이 모두 카테고리 변수일 경우는 상호 정보(mutual information)를[16], 모두 실변수일 경우는 상관 계수를 구하고, 서로 변수의 타입이 다를 경우는 분산 분석(ANOVA)에서 사용되는 F 값을 측정한다[12]. 특히, F 값은 패턴 인식 분야에서 특징 변수가 실변수이며 예측 변수가 카테고리 변수일 경우 특정 변수의 유효성을 측정하기 위해 사용되는 값이나, 회귀 트리에서 발생될 수 있는, 예측 변수가 실수이고 특정 변수가 카테고리인 경우에도 F 값이 특정 변수의 유효성을 보여준다는 가정하에 사용된다.

3.1 문장 정보와 이전 운율구 경계 정보를 이용한 예측

본 절에서는 문장으로부터 추출된 정보와 이전 운율구 경계 정보를 이용하여 운율구 추출과 운율구 사이의 휴지 기간을 예측한다. 본 실험에서는 총 18개의 특징들을 이용하여 결정 트리 및 회귀 트리를 생성한다. 400문장(3,724어절)에서 운율구 경계는 문장의 마지막 어절을 제외하고 1,091개가 발생되었다. 사용될 특징들의 상호 정보, F 값, 상관 계수 r 은 표 3과 같으며, 표에서 예측 변수가 경계 유부일 경우는 자료의 개수가 3,724개이고, 휴지 기간의 경우는 1,091개가 된다²⁾. 사용된 특징 변수와 그의 미들은 다음과 같다.

- Dlpos, Drpos : 관측 어절의 좌품사와 우품사. “-은”, “-는”과 같은 주제격 조사가 우품사로 사용될 경우, 운율구의 경계일 가능성이 높다.
- Dnlpos, Dnrpos : 관측 어절 다음에 위치하는 어절의 좌품사와 우품사. 관측 어절 다음에 위치하는 어절의 좌품사가 용언일 경우, 발화가 끊기지 않고 연결될 가능성이 높다. 만약 관측 어절이 문장의 마지막

1) N 개의 데이터에 대한 결정 트리 오류율 $R(T)$ 의 표준 오류(standard error) $SE(R(T))$ 는 이항 분포를 가정하여, $\sqrt{R(T)(1-R(T))/N}$ 로 구할 수 있다[10].

2) 변수의 첫 글자가 D일 경우는 그 변수가 카테고리 변수임을 뜻하고, C일 경우는 실변수임을 의미한다.

어절일 경우 이 변수들은 NA (Not-Applied) 값을 갖게 되고, 400문장에서 독립어는 문장의 첫 어절로만 사용되어 Dnlpos는 독립어를 제외한 NA, 체언, 용언, 관형어, 부사어 중 하나가 된다.

- Dgrpos : 관측 어절의 지배소 어절 우품사. 지배소 어절은 문장 내에서 항상 우측에 위치하므로[4], 만약 지배소 어절의 우품사가 운율구 경계의 발생을 유발시킬 가능성이 높으면, 관측 어절에서는 운율구 경계가 발생되지 않을 가능성이 높다. 400문장에서 관형어, 독립어 혹은 호격 조사가 우품사인 어절들이 의존소를 가진 경우가 없었으므로, Dgrpos는 이들을 제외한 14개 품사 중 하나가 된다.
- Dcomma : 쉼표의 유무. 일반적으로 어절 다음에 쉼표가 있으면, 운율구 경계일 가능성이 높다.
- Deighth : 문장 내에서 관측 어절의 상대적 위치. 기존의 연구에 의하면, 운율구 경계는 비슷한 크기의 단위로 발화를 나누는 것으로 알려져 있다[20]. 관측 어절의 상대적 위치를 1에서 8사이의 값으로 양자화하여 사용한다.
- Cneojfrbeg, Cneojfrend : 문장의 첫 어절과 마지막 어절부터 몇 번째 어절에 해당하는지에 대한 정보. 문장의 앞부분에 가깝거나 혹은 끝 부분에 가까우면, 운율구 경계가 발생되지 않을 가능성이 높다[26].
- Cnsyltogovn, Cneojtogovn : 지배소까지의 음절과 어절 개수. 관측 어절 w_i 와 지배소 w_j ($i < j$)까지의 음절 혹은 어절 개수가 많다면, 어절들 w_1, \dots, w_i 의 군집 크기가 커진다는 의미이므로 관측 어절 w_i 다음에 운율구 경계일 가능성이 높다. 또한, 이미 운율구 경계라면 휴지 시간이 길어질 가능성이 높다.
- Cnphs, Cnsyls, Cneojis : 문장 내의 음소, 음절, 어절 개수. 문장을 이루는 음소, 음절 혹은 어절 개수가 작으면 운율구 경계의 발생 빈도가 낮다. 이 특징 값들은 문장 내의 모든 어절에서 동일한 값이 사용된다.
- Cbnsyll, Cbneojl : 이전 운율구 경계로부터 관측 어절까지의 음절 개수와 어절 개수. 이전 운율구 경계로부터 관측 어절까지의 어절 개수가 많을수록 운율구 경계가 발생할 가능성이 높다. 이는 일반적으로 사람이 발화를 할 때 이전 운율구 경계로부터 일정 시간이 지나면 다시 운율구 경계를 발생시킨다는 당연한 사실을 바탕으로 한 특징 값이다. 여기서 사용된 이전 운율구 경계는 음성 코퍼스로부터 추출된 정보를 이용한다
- Cbnsylr1, Cbneojr1 : Cbnsyll, Cbneojl을 문장 내의 음절과 어절 개수로 나눈 값. 일반적으로 영어권에서는 이전 운율구의 단어 개수로 나눈 값을 특징 값으로 이용한다[20]. 이 때 문장의 첫부분에서는 이전 운율구가 존재하지 않아 그 값을 구할 수 없게 되고, 이렇게 특징 변수들 중 관측되지 못한 결여 변수

(missing variable)가 있을 때 CART에서는 대리 변수 (surrogate variable)를 이용하여 이를 해결한다[10]. 본 논문에서 구현한 프로그램은 결여 변수를 해결하지 못하므로 문장 길이를 대신 이용한다. 특별히 문장 길이를 사용한 이유는 위의 Deighth 설명에서 보듯이 운율구의 길이가 전체 문장의 길이와 선형적인 관계에 있다는 가정을 바탕으로 한다.

표 3. 운율구 추출을 위한 특징 벡터의 상호 정보(I), F 값, 상관 계수(r)

특징 변수	경계 유무	휴지 시간
Dlpos	$I = 0.05$	$F = 19.59$
Drpos	$I = 0.30$	$F = 12.62$
Dnlpos	$I = 0.19$	$F = 19.97$
Dnrpos	$I = 0.22$	$F = 10.23$
Dqrpos	$I = 0.14$	$F = 2.68$
Dcomma	$I = 0.02$	$F = 135.22$
Deighth	$I = 0.04$	$F = 6.47$
Cneojfrbeg	$F = 0.46$	$r = 0.31$
Cneojfrend	$F = 87.59$	$r = 0.24$
Cnsyltogovn	$F = 359.36$	$r = 0.37$
Cneojtogovn	$F = 343.50$	$r = 0.36$
Cnphs	$F = 34.91$	$r = 0.41$
Cnsyls	$F = 34.80$	$r = 0.41$
Cneojis	$F = 42.71$	$r = 0.40$
Cbnsyll	$F = 508.39$	$r = 0.48$
Cbneojl	$F = 361.55$	$r = 0.40$
Cbnsylr1	$F = 517.26$	$r = -0.10$
Cbneojr1	$F = 401.17$	$r = -0.14$

240문장(2,286어절)으로 운율구 추출 결정 트리를 학습시키고 160문장(1,438어절)에 대해 실험한 결과, 표 4와 같았다. 표 4의 오류 추정치 $\hat{R}(T)$ 는 10-fold cross-validation에 의한 오류 추정치 $\hat{R}(T) \pm SE(\hat{R}(T))$ 이며, OSE 법에 의해 선택된 트리가 1SE 법에 의해 구해진 트리보다 단말 노드 개수가 큰 반면 성능면에서는 더 우수함을 보인다. 결정 트리의 confusion 행렬은 표 5와 같았으며, 경계 유무 결정의 평균 퍼센트 값은 OSE 법의 경우 83.96%, 1SE 법의 경우 83.35%이다. 한편, 회귀 트리의 경우는 표 6과 같은 성능을 나타내었고, 회귀 트리의 경우도 1SE 범보다는 OSE 법이 더 우수함을 알 수 있었다. 이는 1SE 법이 결정 트리뿐만 아니라 회귀 트리에서도 과다제거(overpruning)한다는 사실을 보여준다. 트리의 예측 값과 실제 값과의 상관 계수가 0.65 정도로 운율구 추출을 위해 사용되는 특

표 4. 운율구 추출 실험(성능)

	학습 ($N=2286$)	실험 ($N=1438$)
OSE	$\hat{R}(T) = 0.1351 \pm 0.0071$ $ \hat{T} = 11$	$R^*(T) = 0.1481$
1SE	$\hat{R}(T) = 0.1391 \pm 0.0072$ $ \hat{T} = 8$	$R^*(T) = 0.1515$

표 5. 운율구 추출 실험-I (confusion 행렬)

		경계	무경계	백분율
USE	경계	455	135	77.11%
	무경계	78	770	90.80%
ISE	경계	443	147	75.08%
	무경계	71	777	91.62%

표 6. 휴지 기간 예측 실험-I

	학습 (N=661)	실험 (N=130)
OSE	$\hat{R}(T) = 16850.06 \pm 1512.46$	$R^b(T) = 16716.94$
	$ \hat{T} = 10$	$\sqrt{R^b(\hat{T})} = 129.29$
		$RE^b(T) = 0.58$
		$r^b(T) = 0.65$
ISE	$\hat{R}(T) = 17886.29 \pm 1381.67$	$R^b(T) = 18257.26$
	$ \hat{T} = 6$	$\sqrt{R^b(\hat{T})} = 135.11$
		$RE^b(T) = 0.63$
		$r^b(T) = 0.61$

징 값들이 휴지 기간 예측에도 효과적으로 사용되었음을 보인다.

ISE 법에 의해 구해진 결정 트리와 회귀 트리는 그림 3, 4와 같았다³⁾. 결정 트리에서 사용된 특징 변수들을 관찰해보면, 카테고리 변수들 중 가장 상호 정보 값이 큰 Drpos(관측 어절의 우품사)와, F 값이 다른 변수들보다 큰 Cnsyltogovn(지배소까지의 음절 개수)와 Cbnsyll1(이전 운율구 경계로부터의 음절 개수)이 사용된 것을 알 수 있다. 루트 노드의 질문에서 독립어(it), 호격조사(pv) 등 발화의 끊김이 당연히 발생될 것이라고 예상되는 품사들과 체언(nc), 관형사형 전성어미(ed)와 같이 그 다음 어절과 연속될 가능성이 높은 품사들로 크게 나뉘어졌음을 알 수 있다. 그리고, Cnsyltogovn와 Cbnsyll1이 사용된 노드에서는 예상했던 바와 같이 모두 그 값들이 클 때 운율구 경계 발생율이 높았다. 한편, Cneojtogovn(지배소까지의 어절 개수)가 사용된 노드에서 Cneojtogovn가 0일 경우 운율구 경계가 발생되었는데, 이는 문장의 마지막 어절이 항상 자신을 지배소로 갖는다는 가정에 기인한 것이다³⁾.

Dnlpos(관측 어절 다음에 위치하는 어절의 좌품사)가 사용된 노드를 살펴보면, 용언(vb)이 사용되었을 때, 운율구 경계가 발생되지 않을 가능성이 높았다. 이러한 이유는 다음과 같이 해석될 수 있다. 우선, 트리에서 보듯이 지배소까지의 음절 개수가 많을수록 운율구 경계의 발생

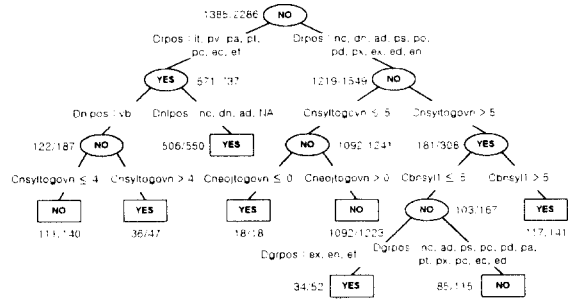


그림 3. ISE 법에 의해 구해진 운율구 경계 결정 트리-I

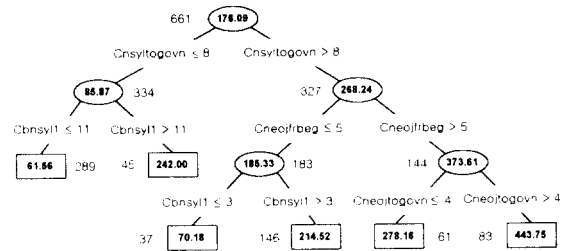


그림 4. ISE 법에 의해 구해진 휴지 기간 회귀 트리-I

빈도가 높았으며, 이는 지배소까지의 어절 개수의 경우도 동일한 특성을 보인다. 표 3에서 Cnsyltogovn와 Cneojtogovn의 F 값을 살펴보면, 각각 359.36, 343.50으로 높은 값을 갖고 있다. 실제로 코퍼스를 조사해본 결과, Cneojtogovn 값이 커짐에 따라 운율구 발생 확률이 높아졌다. Cneojtogovn 값과 발생 확률과의 관계를 더욱 자세히 살펴보면, 관측 어절이 다음 어절의 의존소가 되는 경우 (Cneojtogovn 값이 1인 경우)의 빈도가 총 400문장에서 2,195번 나타났으며, 그 중 운율구 경계가 발생된 경우가 279번으로 발생 확률은 0.127로 가장 낮은 확률을 나타내었다. 한편, Dnlpos와 Cneojtogovn와의 관계를 살펴보면, Dnlpos가 용언(vb)인 경우가 총 1,191번 있었으며, 그 중 관측 어절이 바로 다음 어절의 의존소가 되는 경우가 1,059번으로 확률 $P(Cneojtogovn = 1 | Dnlpos = vb)$ 는 $1059/1191 = 0.89$ 가 된다. Dnlpos가 체언(nc)의 경우, $1109/1859 = 0.59$, 관형어(dn)의 경우, $1/88 = 0.01$, 부사어(ad)의 경우, $26/186 = 0.13$ 이었으므로, 결국 용언의 경우가 가장 높은 확률로 관측 어절이 다음 어절의 의존소가 되며 아울러 운율구 경계 발생 가능성도 낮아지게 된다.

Dgrpos(지배소 우품사)가 사용된 노드를 살펴보면, 보조적 연결어미(ex), 명사형 전성어미(en), 어말어미(cf)가 지배소 우품사일 때 운율구 경계가 발생할 가능성이 높았다. 이러한 현상의 이유를 트리의 루트 노드로부터

3) 그림에서 결정 트리의 경우, 노드 좌우에 있는 x/y에서 y는 그 노드의 전체 데이터 개수이고 x는 노드의 레이블(YES 혹은 NO)에 해당하는 데이터 개수이다.

Dgprpos가 사용된 노드까지의 경로를 해석함으로써 알아 보자. 우선 관측 어절 우품사가 운율구 경계를 발생시킬 가능성이 낮으나 (루트 노드 절문), 지배소까지의 거리가 커서 (Cnsyltogo $v_n > 5$) 운율구 경계가 발생할 가능성이 높아진다. 그러나, 이전 운율구 경계로부터 일정 시간이 지나지 않았기 때문에 (Cbnsyl1 ≤ 5) 관측 어절에서 운율구 경계를 발생시킬 것인가를 지배소 우품사를 조사함으로써 결정한다. 즉, 지배소 우품사가 보조적 연결어미, 명사형 전성어미와 같이 지배소 다음에 운율구 경계가 발생되지 않을 가능성을 높이는 품사가 사용되었을 경우에는, 비록 이전 운율구 경계로부터 일정 시간이 지나지 않아도, 지배소 이후 어절 쪽에서 운율구 경계가 발생할 것으로 가정하고 (이는 화자로 하여금 Cnsyltogo v_n 가 나타내는 어절의 군집 크기보다 더 큰 군집이 될 가능성이 높다고 판단시키게 한다.) 관측 어절에서 운율구 경계를 발생시키게 된다. 실제로 400문장에서 보조적 연결어미와 명사형 전성어미는 각각 149번, 22번 나타났고, 이 중 각 어미 다음에 운율구 경계가 발생한 횟수는 각각 10번, 1번으로, 총 17개의 우품사 중 0.024의 확률을 가진 관형어를 제외하고 가장 운율구 경계 발생 확률이 낮았다. 특징 변수를 소개할 때 설명하였듯이 관형어는 Dgprpos로 사용되지 않았으므로, 실제로는 운율구 경계 발생 확률이 명사형 전성어미의 경우 0.045로 가장 확률이 낮았으며, 보조적 연결어미는 0.067로 두번째로 확률이 낮았다. 이상에서 지배소 우품사가 관측 어절의 운율구 발생과 관계가 있음을 알 수 있다. 끝으로 OSE 법에 의한 결정 트리에서는 이외에 변수 Cbneojr1(이전 운율구 경계로부터의 어절 개수를 문장의 어절 개수로 나눈 값)을 포함한 7개의 변수가 이용되었다.

그림 4의 회귀 트리에서 사용된 변수들을 살펴보면 상관 계수의 값이 큰 Cnsyltogo v_n 와 Cbnsyl1이 사용되었음을 알 수 있다. 이 트리에서 Cnsyltogo v_n 와 Cbnsyl1의 값이 클수록 휴지 시간이 길어졌고, 이는 어절 군집의 크기와 휴지 시간이 양의 상관 관계를 갖는다는 것을 나타낸다. 한편, 트리에서 Cneojr1beg가 사용된 노드를 살펴보면, 이 값이 클수록 휴지 시간이 길어지는 것을 관측할 수 있는데, 이는 다음과 같이 해석할 수 있다. 기존의 연구에 의하면, 발화에 사용된 모든 음소들의 지속 시간 합과 모든 휴지 시간의 합은 선형적으로 비례한다고 알려져 있다. 그러므로, 위의 특징들 중 문장을 이루는 음소 개수 (Cnphs)는 음소들의 지속 시간 합과 양의 상관 관계가 있을 것이며, 이는 비록 휴지 시간 합이 아니더라도 각 휴지 시간과 양의 상관 관계를 갖는다고 가정할 수 있어서, 표 3에 상관 계수가 0.41인 것을 볼 수 있다. 만약 문장 처음부터 관측 어절까지의 거리인 Cneojr1beg의 값이 크다면, 이는 문장이 길다는 것을 의미하며, 결과적으로 음소의 개수가 많아서 휴지 시간도 길어지게 된다. 실제로 Cneojr1beg와 Cnphs의 상관 계수를 구한 결과, 0.65의 양의 상관도를 보

였다. OSE 법에 의한 회귀 트리는 이외에 변수 Cbneoj1(이전 운율구 경계로부터의 어절 개수)과 Dcomma(쉼표의 유무)를 포함한 총 6개의 변수가 사용되었으며, 물론 쉼표가 있을 때 휴지 시간이 길어졌다.

지금까지의 실험에서 이전 운율구 경계로부터 관측 어절까지의 음절 개수와 어절 개수가 결정 트리 및 회귀 트리를 학습시키는데 매우 유용한 특징 변수임을 알 수 있었다. 그러나, 본 실험에서 사용한 이전 운율구 경계 위치는 음성 코퍼스로부터 추출된 정보이므로, TTS 시스템의 실행시에는 얻을 수 없는 정보이다. 그러므로, 다음 절에서는 이를 대신할 수 있는 특징 변수들에 대해 알아보고, 문장 정보만을 이용하여 결정 트리와 회귀 트리를 학습시킨다.

3.2 문장 정보만을 이용한 예측

앞 절에서 사용한 특징 변수들 Cbnsyl1, Cbneoj1, Cbnsylr1, Cbneojr1을 대신할 수 있는 변수로는 구문 정보로부터 크게 두 종류의 특징 변수를 다음과 같이 생각할 수 있고, 각 변수의 F 값과 상관 계수는 표 7과 같다.

표 7. 대용되는 특징 벡터의 F 값, 상관 계수(r)

특징 변수	경계 유무	휴지 시간
Cbnsyl2	$F = 544.38$	$r = 0.44$
Cbneoj2	$F = 501.83$	$r = 0.39$
Cbnsylr2	$F = 528.77$	$r = -0.11$
Cbneojr2	$F = 469.39$	$r = -0.15$
Cbnsyl3	$F = 245.26$	$r = 0.36$
Cbneoj3	$F = 222.49$	$r = 0.34$
Cbnsylr3	$F = 472.69$	$r = -0.07$
Cbneojr3	$F = 436.04$	$r = -0.11$

- Cbnsyl2, Cbneoj2 : 관측 어절 w_k 와 그 앞에 위치하는 임의의 어절 $w_l (k < l)$ 에 대하여, 그 사이에 위치하는 모든 어절들 $w_i (k < l \leq i)$ 이 항상 어절 w_k 의 지배소임을 만족할 때, 가장 앞부분에 위치하는 어절 w_k 와 어절 w_l 사이의 음절 개수와 어절 개수. 이 특징 값은 관측 어절까지의 극부적인 군집도를 나타낸다. 이 값이 크면 관측 어절까지 발화의 끊김이 없을 가능성이 높게 되어 관측 어절에서는 운율구 경계가 발생할 가능성이 높다.
- Cbnsylr2, Cbneojr2 : Cbnsyl2, Cbneoj2를 문장 내의 음절과 어절 개수로 나눈 값.
- Cbnsyl3, Cbneoj3 : 관측 어절 w_m 가 지배하는 어절들 $w_n (m < n)$ 중 가장 앞부분에 위치하는 어절 w_m 까지의 음절 및 어절 개수. 이 특징 값들은 Cnsyltogo v_n (지배소까지의 음절 개수)와 Cneojtogo v_n (지배소까지의 어절 개수)를 지배소 관점에서 본 것으로, 이 값들이 크면 관측 어절에서 운율구 경계가 발생할 가능성이 높다.
- Cbnsylr3, Cbneojr3 : Cbnsyl3, Cbneoj3을 문장 내의 음절과 어절 개수로 나눈 값.

앞 절과 마찬가지로 240문장으로 학습한 후, 160문장에 대해 실험하였고 그 결과는 표 8, 9와 같았다. 실험 결과, OSE 법의 경우 3.1절의 실험에서는 단말 노드의 개수가 11개임에 반해 본 실험에서는 15개로 증가하였고, 결정 트리의 성능은 이전 실험에 비해 오히려 약간 향상되었다. ISE 법의 경우는 단말 노드의 개수가 8개에서 7개로 감소되었고, 160문장에 대한 트리의 성능은 동일한 결과를 나타냈다. 휴지 시간 예측 실험에서는 표 10에 보듯이 3.1절의 실험에 비해 다소 떨어지는 결과를 나타내었다. 그러나, 트리의 단말 노드 개수가 OSE 법과 ISE 법 각각 6개, 5개로 이전 실험의 트리보다 더 적은 개수의 노드가 사용되었다.

표 8. 운율구 추출 실험-II(성능)

	학습 (N=2386)	실험 (N=1438)
OSE	$\hat{R}(T) = 0.1364 \pm 0.0071$ $ \hat{T} = 15$	$R^s(T) = 0.1446$
ISE	$\hat{R}(T) = 0.1408 \pm 0.0072$ $ \hat{T} = 7$	$R^s(T) = 0.1515$

표 9. 운율구 추출 실험-II (confusion 행렬)

		경계	무경계	백분율
OSE	경계	464	126	78.64%
	무경계	82	766	90.33%
ISE	경계	443	147	75.08%
	무경계	71	777	91.62%

표 10. 휴지 시간 예측 실험-II

	학습 (N=661)	실험 (N=430)
OSE	$\hat{R}(T) = 17951.78 \pm 1507.38$ $ \hat{T} = 6$	$R^s(T) = 17586.95$ $\sqrt{R^s(T)} = 132.61$ $RE^s(T) = 0.61$
ISE	$\hat{R}(T) = 19089.25 \pm 1537.80$ $ \hat{T} = 5$	$r^s(T) = 0.63$ $R^s(T) = 17664.46$
		$\sqrt{R^s(T)} = 132.90$ $RE^s(T) = 0.61$ $r^s(T) = 0.62$

그림 5는 ISE 법에 의해 생성된 결정트리이고, 그림 6은 OSE 법에 의해 생성된 회귀트리이다. 그림 5와 그림 3을 비교하면, 앞 절에서 사용된 변수 Cbsnyl1 대신에 Cbsnyl2가 사용되었음을 알 수 있다. 또한 그림 6과 그림 4를 비교하면, Cbsnyl1 대신에 Cbsnyl2가 사용되었음을 알 수 있다. 만약 Cbsnyl2와 Cbsnyl1간의 유효성 차이가 컸다면, 결정 트리와 회귀 트리의 형태가 앞 절의 트리들과 많은 차이를 보였을 것이다. 그러나, 그림에서 보듯이 앞 절

의 트리들에서 Cbsnyl1이 사용된 노드만이 교체된 것으로 미루어, 본 절에서 제안한 8개 특징들의 유효성이 앞 절의 음성 코퍼스에 기반을 둔 특징들의 유효성과 유사하다고 생각할 수 있다. 참고로, Cbsnyl1과 Cbsnyl2와의 상관 계수를 구한 결과, 0.61을 얻을 수 있었다.

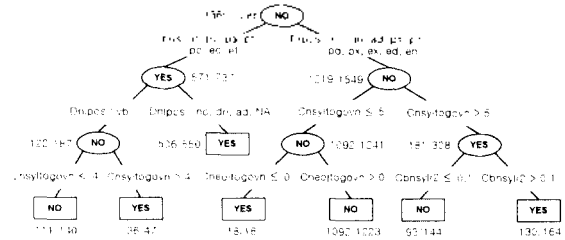


그림 5. ISE 법에 의해 구해진 운율구 경계 결정 트리-II

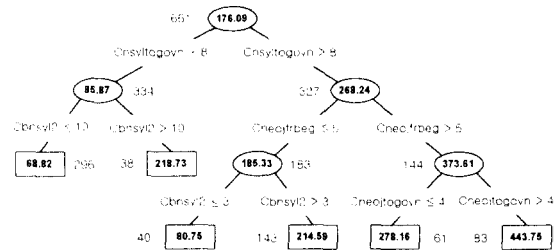


그림 6. OSE 법에 의해 구해진 휴지 시간 회귀 트리-II

IV. 음소 지속 시간 예측

본 장에서는 음소의 지속 시간을 예측하기 위해 다음 10개의 특징 변수들을 이용한다. 각 변수들의 유효성을 알아보기 위해 400문장(24,531개의 음소)에 대해 카테고리 변수일 경우 F 값, 실변수일 경우는 상관 계수를 구하였고, 표 11에 그 값을 보인다.

표 11. 음소 지속 시간 예측을 위한 특징 벡터의 F 값, 상관 계수(r)

특징 변수	음소 지속 시간
Dlph	F = 72.11
Dcph	F = 227.40
Drph	F = 351.18
Dlpos	F = 169.81
Dcpos	F = 140.26
Drpos	F = 570.20
Dwhineoj	F = 390.09
Dwhinphr	F = 6283.59
Cnsylineoj	r = -0.02
Cnsylinphr	r = -0.06

· Dlph, Dcph, Drph : 좌측 음소, 관측 음소, 우측 음소. 이 특징 값들은 음소의 문맥 정보를 표현하기 위해 사용되었고, 만약 관측 음소가 운율구의 처음 혹은 마지막에 위치하게 되면 해당되는 좌우측 음소들은

NA (Not-Applied) 값을 갖게 된다. 그러므로, Dcph는 표 2의 총 44개 카테고리 중 하나가 되나 Dlph와 Drph는 NA를 포함한 45개 카테고리 중 하나가 된다. 각 변수들의 F 값 중 Drph의 F 값이 가장 컸으며, 운율구 경계 앞에서의 장음화 현상에 기인한 것으로 생각된다.

- Dlpos, Dcpos, Drpos : 음소 문맥의 품사들. 이 특징 값들은 형태소 문맥에 따른 변이를 수용하기 위해서이다. 좌우 음소의 품사를 갖는 Dlpos, Drpos도 Dlph, Drph와 마찬가지로 NA 값을 가질 수 있다.
- Dwineoj, Dwihpnr : 어절과 운율구 내에서의 음소 위치 정보. 이 특징 값은 첫 음절, 중간 음절, 마지막 음절 세 개의 카테고리 중 하나가 되고, 만약에 한 음절로 된 어절 혹은 운율구일 경우는 마지막 음절에서 사용되었다고 가정한다.
- Cnsylneoj, Cnsylinphr : 어절내의 음절 개수와 운율구내의 음절 개수. 이 특징 값은 한국어가 영어에서 관측되는 "강세 중심의 리듬(stress-timed rhythm)"의 특성을 부분적으로 보인다는 기존의 연구를 바탕으로 사용되었다[9]. 즉 운율구내의 음절 수가 많을수록 각 음절이 차지하는 시간이 짧아지는 현상을 보이게 되며, 본 연구에서는 이 특성이 음소의 길이 변화에도 영향을 미칠 것으로 가정하고 어절과 운율구내의 음절 개수를 사용한다.

위의 특징 값 중 음소의 품사 정보는 기존의 음소 지속 시간 모델링에서 일반적으로 사용되지 않는 특징 값이다 [21]. 그러므로 본 연구에서는 음소의 품사 정보를 사용하는 것이 트리 기반 모델링 관점에서 과연 더 좋은 성능을 보일 것인가를 알아보기 위해 품사 정보 유무에 따른 두 가지 실험을 수행하였다. 앞 장의 실험과 마찬가지로 240 문장(15,037음소)으로 회귀 트리를 학습시킨 후, 160문장(9,494음소)에 대해 실험하였고 표 12와 같은 결과를 얻었다. 표에서 보는 바와 같이 품사 정보가 트리의 성능과 트리의 크기에 별로 영향을 끼치지 않음을 알 수 있다. 오히려, 품사 정보를 배제한 경우가 항상 약간 우수했다. 또한, ISE 법이 표에서 보는 바와 같이 트리의 크기가 클 때에는 실험 데이터에 대해 오히려 OSE 법보다 더 좋은 성능을 나타내며 트리의 크기도 작아지므로, 본 실험에서는 ISE 법이 선호된다. 본 연구에서 사용한 학습 데이터의 양이 충분히 크다고 생각할 수 없기 때문에⁴⁾, 품사 정보를 특징으로 이용하는 것이 음소 지속 시간을 예측하는 관점에서 의미가 있는지는 이 시점에서 결정하기 어렵다. 그러나 표에 나타난 네 가지 트리들을 본 연구실에서 개발된 TTS 시스템의 지속 시간 예측 모듈로 사용한 결과, 특징

정 트리의 경우가 합성음에 대해 청각적으로 우수하다고 볼 수 없었으므로, 최종적으로 품사 정보를 배제하고 ISE 법을 이용하기로 결정하였다.

표 12. 음소의 지속 시간 예측 실험

	품사 정보 V형	품사 정보 배제
	회귀 (N=15037)	회귀 (N=15037)
OSE	$R(T) = 473.04 \pm 9.18$ $R^2(T) = 487.36$ $\sqrt{R^2(T)} = 22.07$ $RE^2(T) = 0.328$ $r^{(T)} = 0.820$	$R(T) = 485.15 \pm 9.49$ $R^2(T) = 484.61$ $\sqrt{R^2(T)} = 22.01$ $RE^2(T) = 0.326$ $r^{(T)} = 0.821$
	$ T = 104$	$ T = 113$
ISE	$R(T) = 481.70 \pm 9.02$ $R^2(T) = 486.70$ $\sqrt{R^2(T)} = 22.06$ $RE^2(T) = 0.327$ $r^{(T)} = 0.820$	$R(T) = 494.16 \pm 9.29$ $R^2(T) = 479.08$ $\sqrt{R^2(T)} = 21.88$ $RE^2(T) = 0.322$ $r^{(T)} = 0.823$
	$ T = 74$	$ T = 73$

트리의 질문들을 조사해본 결과, 모든 변수들이 이용되었고, 루트 노드에서 사용된 질문은 Drph(우측 음소)가 NA인지를 비교하는 질문이었다. 즉 운율구의 마지막 음소에서 발생하는 장음화 현상을 처리하는 질문이 사용되었다. 실제 값과, 품사 정보를 배제한 ISE 법에 의한 트리의 예측 값과의 차이에 대한 V표식 상자-수염 그림은 그림 7과 같았고, 그림에서 보여지듯이 각 클래스 오류들의 중앙값이 0에 가까움을 알 수 있다.

V. 트리 기반 모델의 최종 성능

본 장에서는 구축된 문서 코퍼스(400문장)와 음성 코퍼스(약 33분)를 모두 이용하여 결정 트리 및 회귀 트리를 학습시키고 그 성능에 대해 알아본다. 운율구 추출 및 휴지 시간의 예측을 위해서는 3.2절에서 사용한 특징 값을 사

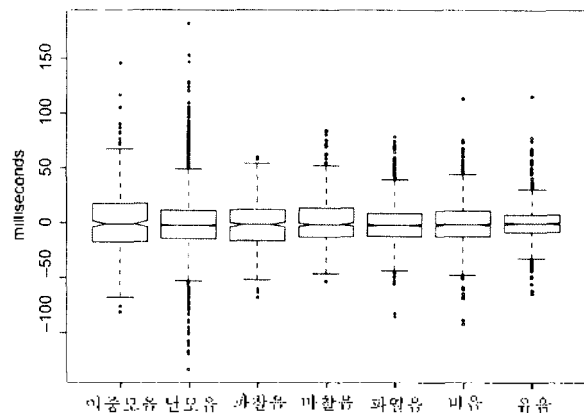


그림 7. 음소 클래스에 따른 예측 오차값의 V표식 상자-수염 그림

4) Michael D. Riley는 음소 지속 시간 예측을 위해 1500문장을 이용하여 트리를 생성하였다[21].

용하고, 음소 지속 시간 예측을 위해서는 음소의 품사 정보를 배제한 7개의 특징 값을 사용한다. 운율구 추출 및 휴지 기간 예측을 위해서는 OSE 법에 의해 트리를 선택하였고, 음소 지속 시간 예측을 위해서는 ISE 법을 이용하였다. 선택된 트리의 성능은 표 13과 같다. 한편, 음소의 지속 시간 예측 실험에서 음소의 품사 정보를 포함한 경우 $\|\hat{T}\|$, $\hat{R}(T)$ 는 각각 $98, 455.03 \pm 6.80$ 이었다.

표 13. 트리 기반 모델의 성능

운율구 추출 ($N=3724$)
$\hat{R}(T) = 0.1377 \pm 0.0056$
$\ \hat{T}\ = 17$
휴지 기간 예측 ($N=1091$)
$\hat{R}(T) = 16361.04 \pm 1003.37$
$\sqrt{\hat{R}(T)} = 127.91$
$\ \hat{T}\ = 7$
음소 지속 시간 예측 ($N=24531$)
$\hat{R}(T) = 464.11 \pm 6.78$
$\sqrt{\hat{R}(T)} = 21.54$
$\ \hat{T}\ = 95$

이상의 실험에서 우리는 한국어의 운율을 모델링함에 있어서 트리 기반 모델링이 운율구 경계를 찾는 패턴 분류 문제와 휴지 기간, 음소 지속 시간을 예측하는 함수 근사 문제에 대해 모두 효과적인 것을 알았다.

VI. 결 론

본 논문에서는 트리 기반 모델링 기법을 이용하여 한국어 운율 중 운율구 추출, 휴지 기간, 음소 지속 시간을 모델링하였다. 총 400문장을 여러 장르로부터 발취하고, 여성 아나운서가 발성한 음성을 녹음한 후, 문장과 음성 신호에 언어 정보와 운율구 경계, 음소 경계 등을 표시하였다. 운율을 모델링함에 앞서, 특징 변수들의 상호 정보, F 값, 상관 계수를 이용하여 변수들의 유효성을 예측하였다.

운율구 추출의 경우는 관측 어절의 우품사에 해당되는 형식 형태소가 가장 유효한 특징이었고, 휴지 기간 예측에서는 관측 어절의 지배소까지의 음절 수가 효과적으로 사용되었다. 음소의 지속 시간 예측에서는 관측 음소가 운율구의 마지막 음소인지를 알아보는 질문이 루트 노드에서 사용되었다. 실험을 통하여 문장 정보만으로도 운율구 추출이 효과적으로 예측될 수 있음을 알 수 있었고, 음소의 지속 시간 예측에서는 음소의 품사 정보를 사용하는 것이 트리의 오류율 감소에 크게 영향을 미치지 못하였다.

트리 기반 모델링은 사람이 해석하기가 매우 쉽다는

장점이 있는 반면, 특징 공간의 분할 평면이 특징 값 축에 평행해야 하는 단점이 있다. 이를 보완하기 위해 변수들의 선형적 결합에 의한 분할 평면이 사용되거나[10], 퍼셉트론(perceptron)을 통한 비선형적 분할이 시도되기도 하였다[15]. 그러나 이러한 경우에는 트리 해석의 용이성이 떨어지게 되므로, 앞으로 트리 생성 방법의 기본적인 틀을 유지하며 트리의 성능을 향상시키는 연구를 계획 중이다.

참 고 문 헌

1. 김 상훈, 성 철재, 이 정철, "운율구 경계현상 분석 및 텍스트에서의 운율구 추출", 한국음향학회지, Vol. 16, No. 1, pp. 24-32, 1997.
2. 김 정수, 이 혜정, "언어정보 및 통계 데이터를 이용한 한국어 운율 생성", 제 13회 음성통신 및 신호처리 워크샵, pp. 227-231, 1996.
3. 김 형근, 확률적 의존 문법과 한국어 구문 분석, 석사 학위논문, 한국과학기술원 전산학과, 1995.
4. 서 광준, 어절사이의 의존관계를 이용한 한국어 구문분석기, 석사학위논문, 한국과학기술원 전산학과, 1993.
5. 성 유나, 이 양희, "한국어 음운 지속시간 모델화를 위한 특징 분석", 제13회 음성통신 및 신호처리 워크샵, pp. 301-306, 1996.
6. 이 기문, 김 진우, 이 상익, 국어 음운론, 학연사, 1984.
7. 이 상호, 오 영환, 서 정연, "한국어 문서 음성 변환 시스템을 위한 문서 분석기", 한국음향학회지, Vol. 15, No. 3, pp. 50-59, 1996.
8. 이 상호, 오 영환, "음성 코퍼스로부터 TTS 시스템을 위한 단위 음성 자동 생성", 한국음향학회 학술발표대회 논문집 제 16권, pp. 293-296, 1997.
9. 이 현복, 한국어의 표준발음, 교육과학사, 1989.
10. L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth Statistics/Probability Series, Belmont, CA, 1984.
11. P.A. Chou, "Optimal Partitioning for Classification and Regression Trees," *IEEE Trans. on PAMI*, Vol. 13, No. 4, pp. 304-354, 1994.
12. J.L. Devore, *Probability and Statistics for Engineering and the Science*, Brooks/Cole Publishing Company, 1991.
13. F. Esposito, D. Malerba, and G. Semeraro, "A Comparative Analysis of Methods for Pruning Decision Trees," *IEEE Trans. on PAMI*, Vol. 19, No. 5, pp. 476-491, 1997.
14. S. Fujio, Y. Sagisaka, and N. Higuchi, "Stochastic Modeling of Pause Insertion using Context-free

Grammar," *Proceedings ICASSP '95*, pp. 604-607, 1995.

15. H. Guo and S.B. Gelfand, "Classification Trees with Neural Network Feature Extraction," *IEEE Trans. on Neural Networks*, Vol. 3, No. 6, pp. 923-933, 1992.

16. R.W. Hamming, *Coding and Information Theory*, Prentice-Hall, Inc., 1980.

17. S. Haykin, *Neural Networks*, Macmillan College Publishing Company, 1994.

18. X.D. Huang, Y. Ariki, and M.A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, 1990.

19. A. Ljolje and F. Fallside, "Synthesis of Natural Sounding Pitch Contours in Isolated Utterances using Hidden Markov Models," *IEEE Trans. on ASSP*, Vol. 34, No. 5, pp. 1074-1079, 1986.

20. M. Ostendorf and N. Veilleux, "A Hierarchical Stochastic Model for Automatic Prediction of Prosodic Boundary Location," *Computational Linguistics*, Vol. 20, No. 1, pp. 27-54, 1994.

21. M.D. Riley, "Tree-based modelling of segmental duration," *Talking Machines: Theories, Models, Designs*, G. Bailly, C. Benoit, and T.R. Sawallis, editors, pp. 265-273, Elsevier Science, 1992.

22. K. Ross and M. Ostendorf, "Prediction of Abstract Prosodic Labels for Speech Synthesis," *Computer Speech and Language*, Vol. 10, pp. 155-185, 1996.

23. S.R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 21, No. 3, pp. 660-674, 1991.

24. R. Sproat and J. Olive, "An Approach to Text-to-Speech Synthesis," *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal, editors, pp. 611-633, Elsevier Science, 1995.

25. C. Traber, "F₀ generation with a database of natural F₀ patterns and with a neural network," *Talking Machines: Theories, Models, Designs*, G. Bailly, C. Benoit, and T.R. Sawallis, editors, pp. 287-304, Elsevier Science, 1992.

26. M.Q. Wang and J. Hirschberg, "Automatic Classification of Intonational Phrase Boundaries," *Computer Speech and Language*, Vol. 6, pp. 175-196, 1992.

27. C.W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P.J. Price, "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries," *J. Acoust. Soc. Am.*, Vol. 91, No. 3, pp. 1707-1717, 1992.

이 상 호(Sangho Lee)



1993년 2월: 동국대학교 전자계산학과 (학사)
 1995년 2월: 한국과학기술원 전산학과 (석사)
 1995년 3월~현재: 한국과학기술원 전산학과 박사과정 재학중
 *주관심분야: 음성 합성, 자연언어처리, 패턴인식

오 영 환(Yung-Hwan Oh)



1972년 :서울대학교 공과대학(학사)
 1974년 :서울대학교 교육대학원(석사)
 1980년 :Tokyo Institute of Technology 정보공학전공(박사)
 1981년 ~ 1985년 :충북대학교 컴퓨터공학과 조교수
 1983년 ~ 1984년 :University of California, Davis 연구교수
 1995년 ~ 1996년 :Carnegie-Mellon University 연구교수
 1985년 ~ 현재 :한국과학기술원 전산학과 교수
 *주관심분야: 음성인식, 음성합성, 음성코딩, 화자인식, 대화관리, 신경회로망, 전문가 시스템