

## 한국어 음성합성기의 운율 예측을 위한 의사결정트리 모델에 관한 연구

A Study of Decision Tree Modeling for Predicting the Prosody of Corpus-based Korean  
Text-To-Speech Synthesis

강 선 미\* · 권 오 일\*\*  
Sunmee Kang · Ohil Kwon

### ABSTRACT

The purpose of this paper is to develop a model enabling to predict the prosody of Korean text-to-speech synthesis using the CART and SKES algorithms. CART prefers a prediction variable in many instances. Therefore, a partition method by F-Test was applied to CART which had reduced the number of instances by grouping phonemes. Furthermore, the quality of the text-to-speech synthesis was evaluated after applying the SKES algorithm to the same data size. For the evaluation, MOS tests were performed on 30 men and women in their twenties. Results showed that the synthesized speech was improved in a more clear and natural manner by applying the SKES algorithm.

**Keywords:** Speech Synthesizer, TTS(Text-to-Speech), prosody prediction, CART, SKES, decision tree

### 1. 서 론

사람의 가장 자유로운 의사전달 수단인 음성을 이용한 컴퓨터 인터페이스 기술에 대한 요구는 매우 급격히 증대되고 있다. 자동 음성 입력기의 개발을 위한 음성인식 기술의 향상과 더불어 인간이 말을 하는 것과 같은 자연스러운 말을 생성해내는 음성 합성기의 요구도 확대되어 가고 있다. 임의의 문장에 대한 자연스러운 소리의 생성을 위해서 현재로는 문서-음성 변환 시스템(Text-to-Speech: TTS)이 가장 넓게 사용되고 있다. 이는 작은 합성 단위음성과 언어 처리를 이용하여 임의의 문장에 대한 음성을 생성한다. 언어 처리를 이용하여 입력된 문장을 적당한 합성 단위의 조합으로 대응시키고, 문장으로부터 적당한 억양과 지속시간을 추출하여 합성음의 운율을 결정한다.

초기 음성합성은 발성기관을 모델로 하여 신호처리 기술을 이용하여 단위음성을 원하는 모습의 음성으로 만들어 가는 방식을 시도하였으나 단위음성에 가해진 신호처리에 의해 음질의 저하가 발

\* 서경대학교 컴퓨터과학과

\*\* 현대오트넷(주)

생되고, 접합점에서 스펙트럼 및 에너지의 불연속성과 왜곡이 문제가 된다. 이러한 문제점을 해결하고자 코퍼스기반의 음성 합성(corpus-based speech synthesis)이 제안되었다. 코퍼스기반의 음성합성은 발성기구에 기초한 완전한 제어규칙에 따라 작업하는 것이 아니라 대규모의 음성 데이터베이스를 준비하여 그 중에서 가장 정확한 스펙트럼이나 파형을 효율적으로 찾아내는 것을 목표로 하고 있다[1].

코퍼스기반의 TTS 개발의 핵심 기술은 텍스트 전처리 알고리즘, 형태소 분석, G2P, 구문분석과 같은 텍스트 분석 알고리즘, 운율 예측 알고리즘 그리고 음성 합성 DB의 제작 기술을 들 수 있다. 상기 기술들은 서로 간에 매우 밀접한 관계가 있으며, 음질 향상을 위해서는 서로 유기적인 성능 평가에 의한 개선 작업이 요구된다[2].

본 논문에서는 한국어 TTS의 운율 예측 알고리즘을 결정트리 알고리즘인 CART(Classification And Regression Tree)[3]-[8]를 이용하여 훈련과 예측을 실시하고, SKES(Spearman or Kruskal - Wallis test and Exhaustive Search)[9][10]를 적용한 음성으로 음질을 비교 평가하여 운율예측에 대한 새로운 모델의 가능성을 제시하고자 한다.

## 2. CART를 이용한 운율 생성 모델

우리말의 운율 단위 중에서 가장 작은 것이 음절(syllable)이다. 음절은 모여서 음운 단위인 음운 단어(phonological word)를 만들고, 음운 단어는 모여서 강세구(accentual phrase: AP)를 만들며, 강세구들은 모여서 억양구(intonation phrase: IP)를 만들어 낸다. 발화(utterance)는 하나 이상의 억양구가 모여 만들어지는 가장 상위의 운율 단위로 본다[11]. 어절경계위치는 강세구의 경계와 억양구의 경계를 의미하며, 음성 합성기의 자연성을 향상시키기 위해서는 어절 경계가 잘 예측되어야 한다.

운율정보는 두 가지 경우로 나누어 볼 수 있다. 첫째는 운율경계위치를 찾는 것으로 ‘관측 어절이 억양구 또는 강세구의 경계인가?’ 하는 문제로 이는 ‘패턴 분류 문제’로 볼 수 있으며 결정트리 생성을 통하여 문제를 해결 할 수 있다. 둘째는 ‘휴지의 길이가 얼마인가?’ 또는 ‘음소의 길이, 세기는 어느 정도인가?’ 하는 문제로 이는 ‘함수 근사화의 문제’로 볼 수 있으며 이는 회귀트리 생성을 통하여 문제를 해결할 수 있다.

본 논문에서는 먼저 CART를 이용하여 운율 제어 규칙을 찾았으며, 이를 위해 여성화자가 발화한 6,245 문장과 남성화자가 발화한 5,760 문장을 사용하였다.

### 2.1 운율 경계 예측

운율 경계위치에 대한 제어 규칙을 찾기 위해서 발화문장에 대해 전문가가 억양구 또는 강세구 경계를 표시하였고, 해당 문장에 대해 형태소 분석을 하여 운율 경계위치와 그와 관련하여 서로 영향을 주는 특징 변수들의 상관 데이터를 수집하였다. 대상이 되는 문장은 다양한 유형을 포함하고 있다. 표본 데이터를 수집하는 과정에서 3 어절 미만으로 이루어진 문장과 16 어절 이상으로 이루어진 문장은 기본적으로 제외하고, 형태소 분석 결과에서 미등록어[12] 처리루틴을 통과하는 어절이 30% 이상인 것을 제외하여 여성화자의 경우 총 28,163 건의 데이터를 얻었으며, 남성화자의 경우

30,637 건의 데이터를 얻었다.

본 논문에서는 운율경계위치 예측을 위해 결정트리 모델링 기법을 이용하였고, 10-fold cross - validation 방법에 의해 최적 트리를 결정하고, 노드의 불순도를 나타내는 척도로는 Gini 인덱스를 사용하였고 SE(Standard Error) Rule은 최소비용트리를 선택하였다.

### 2.1.1 사용된 특징 변수들

<표 1>과 <표 2>는 이 논문에 사용된 독립변수와 종속변수를 나타낸다. 운율경계를 결정하는 결정트리를 생성하는데 필요한 특징 변수로서는 문장에서 관측어절의 앞과 뒤에 오는 최대 2 개의 어절에 대해서 품사를 고려하였다. 본 논문에서는 관측어절의 왼쪽에 위치한 품사를 좌품사, 오른쪽에 위치한 품사를 우품사라고 명명한다. 또한, 관측어절의 시작어절과 마지막 어절로부터 상대적인 위치, 음절수와 관측어절에 인접하는 문장부호 등을 고려하였다. 'C\_'로 시작하는 변수는 범주형 변수를 의미하고, 'N\_'으로 시작하는 변수는 연속형 숫자 변수를 의미한다.

표 1. 운율 경계 예측에 사용된 독립변수

변수	설명
C_PPLPOS	관측어절의 앞. 앞 어절의 좌품사
C_PPRPOS	관측어절의 앞. 앞 어절의 우품사
C_PLPOS	관측어절의 앞 어절의 좌품사
C_PRPOS	관측어절의 앞 어절의 우품사
C_LPOS	관측어절의 좌품사
C_RPOS	관측어절의 우품사
C_NLPOS	관측어절의 뒤 어절의 좌품사
C_NRPOS	관측어절의 뒤 어절의 우품사
C_NNLPOS	관측어절의 뒤. 뒤 어절의 좌품사
C_NNRPOS	관측어절의 뒤. 뒤 어절의 우품사
C_PAUSE	관측어절의 휴지나 종결기호의 유무
C_NPAUSE	관측어절의 뒤 어절의 휴지나 종결기호의 유무
C_QT_CL	관측어절의 인용의 종료로 나타내는지 여부
C_QT_NXO	관측어절의 다음어절의 인용의 시작을 나타내는지 여부
N_IDXBEG	관측어절의 첫 어절부터 시작한 어절 인덱스
N_IDXEND	관측어절의 마지막 어절부터 시작한 어절 인덱스
N_IDXFRI	관측어절의 억양구 발생 이후 어절 인덱스
N_SYLL_C	관측어절의 음절 수
N_SYLL_N	관측어절의 뒤 어절의 음절 수
N_SYLLIP	관측어절의 억양구 발생 이후의 누적 음절 수
N_SYLLAP	관측어절의 강세구 발생 이후의 누적 음절 수

표 2. 운율 경계 예측에 사용된 종속변수

변수	설명
C_IP	관측어절의 억양구 경계 여부
C_AP	관측어절의 강세구 경계 여부

2.1.2 좌품사와 우품사의 결정 규칙

형태소 분석기에서 사용된 품사집합을 좌품사와 우품사를 정하는 데에 그대로 적용하였으나, 어절과 어절 간의 관계를 잘 나타낼 수 있도록 몇 가지 부분에 대해서는 예외적으로 결정하였다. 현재의 품사집합에서는 주격 조사와 목적격 조사와의 구분이 없는 점과 접사종류와 연결어미의 세분화도 추후 고려해야 할 것이다.

형태소 분석에서와 달리, 좌품사와 우품사를 결정하는데 있어서 문장기호는 제외되었으며 그 밖의 특징 변수로 사용되었다. 이는 문장기호가 훈련과 예측과정에서 일관적으로 적용되기 어렵고, 순수하게 어절과 어절의 품사정보만을 규칙으로 반영하는 것이 오류를 적게 포함하는 것이라는 판단해서이다. 또한, 빈번하게 출현하는 명사 등의 나열에서 오는 모호함을 피하고 어절 상호간의 보다 분명한 관계를 나타내고자 하는 의도에서 예외 규칙을 적용하였다.

2.1.3 ZERO, AP, IP 예측

ZERO, AP, IP 경계 예측을 위해서 훈련데이터, 검증데이터, 시험데이터를 각각 50 : 30 : 20의 비율로 나누어 사용하였다. 검증데이터에 의해 오 분류율을, 학습데이터에 의해 결정트리를 모델링하고 시험데이터에 의해 검정을 실시하였다. 결정트리 알고리즘은 Chi-square test를 사용하였다[6].

카이제곱 통계량이 자유도에 비해서 매우 작다는 것은 설명변수의 각 범주에 따른 목표변수의 분포가 동질적이라는 것을 의미하며, 따라서 설명변수가 목표변수의 분류에 영향을 주지 않는다고 결론지을 수 있다. 자유도에 대한 카이제곱 통계량의 크고 작음은 p-값으로 표현될 수 있는데, 카이제곱 통계량이 자유도에 비해서 작으면 p-값은 커지게 된다. 결국, 분리기준을 카이제곱 통계량으로 한다는 것은 p-값이 가장 작은 설명변수와 그 때의 최적분리에 의해서 자식마디가 형성되게 한다는 것을 의미한다. p-값이 가장 작은 변수를 선택하여 SE(Standard Error)가 가장 작은 지점에서 분할하는 방법으로 적용하였다.

표 3. 학습데이터의 ZERO, AP, IP 예측 결과

실제 \ 예측	예측			
	ZERO	AP	IP	합계
ZERO	1,199 50.15%	1,095 45.08%	97 4.06%	2,391 17.32%
AP	597 7.65%	6,687 85.66%	522 6.69%	7,806 56.54%
IP	201 5.57%	1,559 43.19%	1,850 51.25%	3,610 26.15%
합계	1,997	9,341	2,469	13,807

<표 3>의 학습데이터에서의 예측결과를 살펴보면, ZERO 예측율은 50%로 낮은 예측결과를 나타내지만, IP를 ZERO로 판단하는 오 분류율은 5.57%이고 ZERO를 IP로 판단하는 오 분류율은 4.06%이다. IP를 ZERO로 잘못 예측하면 문장의 의미를 잘못 전달할 가능성이 있다. 본 실험 결과로는 음성 합성시 문장에 대한 의미가 잘못 전달될 가능성은 높지 않은 것으로 판단된다. 또한 운율 예측은 문장에 대한 의미론이 포함되어 있기 때문에 통계적인 수치에 의해 만족스러운 예측율을 얻

었다고 해도 실제 음성을 들었을 때와 약간의 차이가 있다. 그러므로 모델을 최적으로 만들려면 실제 음성을 청취하고 잘못된 부분을 수정해 나가는 과정이 필요하다.

## 2.2 음소의 지속시간, 평균피치 및 세기 예측

음소의 지속시간과 평균피치 및 세기에 대한 제어 규칙을 얻기 위해 이미 구축된 음성 데이터베이스에서 지속시간, 평균피치, 세기에 관계있는 독립변수들을 수집했다. 음소별로 서로 다른 음향학적인 특성을 쉽게 반영하기 위해 각각 해당 음소별로 데이터를 수집하였고, 다시 각각의 트라이폰에 대해서 수집하여 통계를 조사하고 정규화 하였다.

음소의 평균피치에 대해서 남성화자의 경우, 상한 220 Hz, 하한 60 Hz, 여성화자의 경우 상한 340 Hz, 하한 70 Hz의 문턱치를 두었고, 대상 트라이폰 중 상위와 하위 3~5%에 해당하는 평균피치를 가진 트라이폰은 제외하였다.

음소의 지속시간에 대해서도 각각의 트라이폰별 통계를 조사하여 평균 지속길이에 대해서 상위와 하위 5%에 해당하는 지속길이를 가진 트라이폰은 제외하였다.

본 논문에서는 음소의 지속시간과 평균피치 및 세기에 대해 10-fold cross-validation 방법에 의해 최적 회귀트리를 결정하고, 노드의 분할 방법으로 분산 감소량을 사용하였다.

### 2.2.1 사용된 특징 변수들

음소의 지속시간, 평균피치, 세기에 대해 회귀트리를 생성하는데 필요한 특징 변수는 <표 4>에 나타내었다. 어절경계에 대한 특징 변수들과 마찬가지로 'C\_'로 시작하는 변수는 범주형 변수임을 의미하고, 'N\_'로 시작하는 변수는 연속형 숫자 변수임을 의미한다.

표 4. 음소의 지속시간, 평균피치, 세기 예측에 사용된 독립변수

변 수	설 명
C_PHONE	해당 음소 (트라이폰에서의 중앙 음소인덱스)
C_LPHONE	인접 음소 (트라이폰에서의 왼쪽 음소인덱스)
C_RPHONE	인접 음소 (트라이폰에서의 오른쪽 음소인덱스)
C_ATTR	해당 음소의 속성 (예: 자음예사소리/단모음)
C_CLASS	해당 음소의 종류 (초성/중성/종성)
C_PENV_0	트라이폰 환경
C_PENV_1	트라이폰 환경
C_PENV_2	트라이폰 환경
C_PENV_3	트라이폰 환경
C_HDEL	ㅎ 음소 탈락 구분
C_ASPIRATION	해당 음절의 격음화 구분
C_GLOTTALIZATION	해당 음절의 경음화 구분
C_LENITION	해당 음절의 연음화 구분
C_ASPIRATION_N	다음 음절의 격음화 구분
C_GLOTTALIZATION_N	다음 음절의 경음화 구분
C_LENITION_N	다음 음절의 연음화 구분
C_SYKIND_AP	AP 구간내 음절 위치의 종류
N_SYLOC_AP	AP 구간내 음절의 위치

N_SYNUM_AP	AP 구간내 음절의 크기
N_AIP_AP_CURR	해당 AP 구간내 AP / IP 구분
N_AIP_AP_PREV	다음 AP 구간내 AP / IP 구분
N_SY_DUR	해당 음절의 지속시간
N_LDUR	인접 음소의 지속시간 (왼쪽)
N_RDUR	인접 음소의 지속시간 (오른쪽)
N_LPIT	인접 음소의 평균 피치 (왼쪽)
N_RPIT	인접 음소의 평균 피치 (오른쪽)
N_LINT	인접 음소의 세기 (왼쪽)
N_RINT	해당 음소의 억양구 경계 이후의 트라이폰 인덱스

표 5. 음소의 지속시간, 평균피치, 세기 예측에 사용된 종속변수

변 수	설 명
N_DUR	해당 음소의 지속시간
N_PIT	해당 음소의 평균피치
N_INT	해당 음소의 세기

음소의 지속시간, 평균피치 및 세기를 예측하는데 있어서 <표 5>의 특징 변수들 중에 각각의 운율예측에 맞는 변수를 선택하여 적용하였다. 트라이폰으로 운율을 예측하기 때문에 C\_PHONE, C\_LPHONE, C\_RPHONE는 사용할 수 없는 변수이고, 오른쪽과 왼쪽 음소의 지속시간, 평균피치, 세기는 운율 예측에 이용되지 않고 운율 예측 후 음소 합성시 변수로 사용된다.

2.2.2 음소의 지속시간에 대한 통계

지속시간은 억양과 함께 합성음의 자연성을 결정하는 중요한 요소이며, 발화 속도에도 영향을 받으므로 이를 제어하는 것이 자연성에 매우 큰 영향을 미친다고 볼 수 있다. 개별 음소로 나누어 해당 음소가 초성, 중성, 종성의 분류 및 조음방식, 어절 경계내의 상대적인 위치에 따른 음소 지속시간에 대한 통계치를 구하였다.

자음에 대하여 초성과 종성을 특징에 따라서 초성의 경우는 음소와 음소 사이에 해당 음소가 위치하는 경우와 어절 경계의 첫 번째, AP 경계의 첫 번째, IP 경계의 첫 번째에 위치하는 등의 4 가지 경우로 나누어 볼 수 있다. 종성의 경우는 음소와 음소 사이에 해당 음소가 위치하는 경우와 어절 경계의 마지막, AP 경계의 마지막, IP 경계의 마지막에 위치하는 등의 4 가지 경우를 고려하였다. 모음의 경우는 어절경계의 첫 음소로도 마지막 음소로도 올 수 있으므로 여러 경우의 수가 있겠지만, 음소 사이에 위치한 경우와 AP나 IP 경계의 첫 음소 및 마지막 음소에 대해서만 고려하였다.

2.2.3 음소의 지속시간에 대한 훈련 및 예측결과

음소의 지속시간에 대해 구해진 통계적인 수치를 가지고 회귀트리를 생성하여 얻은 제어규칙을 다시 원래의 발화 문장에 대해 적용해 봄으로써, 신뢰성에 대해 검증을 하였다.

실험을 통해서 얻은 음소별 표준편차를 가지고, 예측값과 실험값의 차가 표준편차범위의 50% 이내이면 '성공'으로 간주하였다. 다만, AP와 IP 경계에 대해서는 표준편차범위의 100% 이내이면 '성공'으로 간주하였다. 예측 결과는 <표 6>과 같으며, 평균 성공률은 70%대로 ZERO, AP, IP 예측시와

마찬가지로 만족스럽지는 않고 개선의 여지가 있다고 여겨진다.

표 6. 음소의 지속시간에 대한 예측 실험 결과

	여 성			남 성		
	성공	발생건수	성공률	성공	발생건수	성공률
초성	73,200	117,981	62.04	92,241	135,230	68.21
중성	110,582	143,271	77.18	124,574	163,958	75.98
종성	31,988	45,759	69.91	40,660	53,644	75.80
합계	215,770	307,011	<b>70.28</b>	257,475	352,832	<b>72.97</b>

2.3 CART에 의한 운율 예측시 고려되어야 할 문제

CART는 최적의 이진분리를 후진소거법으로 탐색한다. 즉, 각 노드에서 모든 가능한 분리에 대하여 분산을 계산하고, 분산감소량이 최대가 되는 분리 기준을 선택하는 알고리즘이다. 그러나 CART는 범주의 수가 많은 예측변수를 선호하는 심각한 편의를 가진다. 특히, 운율예측에서는 좌측과 우측의 음소를 고려하여 예측하므로 45개 이상의 큰 범주를 가지게 된다. 그러므로 운율 예측에서 CART 알고리즘의 사용을 위해서는 음소들을 그룹화 시켜 범주의 수를 줄이는 방법을 모색해야 한다.

본 논문에서는 상기 문제를 보완하게 위해 F-Test 통계량에 의한 분할 방법을 적용하였다. F-Test는 목표변수가 연속형인 경우에는 두 개 이상의 그룹에 대해서 평균의 차이를 검정하는 ANOVA (Analysis of Variance) F 통계량을 분리기준으로 사용할 수 있다. F 통계량은 평균과 분산을 기초로 하여 <표 7>과 같은 분산 분석표로부터 계산된다.

$y_{ij}$ 를  $i$ 번째 설명변수의 범주에 속하는  $j$ 번째 관측개체의 목표변수의 값이라고 하고,  $\bar{y}_i$ 를  $i$ 번째 범주의 평균,  $\bar{y}$ 를 전체평균이라고 할 때, 분산 분석표는 <표 7>과 같다[13].

표 7. 분산 분석표

요인	자유도	평방향	평균평방	분산비
설명변수	$r - 1$	$SST = \sum_{i=1}^r n_i (\bar{y}_i - \bar{y})^2$	$MST = SST / (r - 1)$	
오차	$n - r$	$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$MSE = SSE / (n - r)$	$F = \frac{MST}{MSE}$
총 변동	$n - 1$	$TSS = \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$		

여기에서  $r$ 은 설명변수의 범주 개수,  $n_i$ 는  $i$ 번째 범주의 관측개체 수,  $n$ 은 마디에서의 관측개체 수이다. 이렇게 계산된 F 통계량은 자유도 ( $r-1$ ,  $n-r$ )인 F-분포를 따른다.

F 통계량이 자유도에 비해서 매우 작다는 것은 설명변수의 각 범주에 따른 목표변수의 평균차이가 유의하지 않다는 것을 의미하며, 따라서 설명변수가 목표변수의 예측에 영향을 주지 않는다고 결

론지를 수 있다. 카이제곱 통계량과 마찬가지로 자유도에 대한 F 통계량의 크고 작음은 p-값으로 표현될 수 있다. 즉, F 통계량이 자유도에 비해서 작으면 p-값은 커지게 된다.

### 3. SKES 알고리즘 적용 및 음절 비교

CART의 문제점으로 지적된 변수 선택의 편향이 비교적 적은 회귀트리를 만들기 위해, 분리 변수를 선택하는 단계와 그 변수의 분리 점을 찾는 단계로 나누어 노드의 분리 기준을 정하는 SKES 알고리즘을 적용하였다. 또한, 각 마디에서의 불순도(impurity)를 재는 측도인 엔트로피 지수를 분리 기준으로 사용하여 다지 분류(multi-split)도 가능한 C4.5 알고리즘에서 기준으로 사용하는 엔트로피 지수를 참조하여 알고리즘을 전개하였다[14][15].

#### 3.1 SKES(Spearman or Kruskal-Wallis test and Exhaustive Search) 알고리즘

본 알고리즘은 분리 변수를 선택하는 단계와 그 변수의 분리 점을 찾는 단계로 나누어 노드의 분리 기준을 정하는 알고리즘으로 제안되었다. 구성된 결정트리는 다양한 응용분야에 적용될 수 있다. 알고리즘의 전개는 참고문헌 [9]와 [10]을 참조하여 구현하였으며, 본 절에서는 대략적인 소개를 하겠다.

먼저 K개의 예측변수가 있는 자료가 있다고 하자. 자료의 예측변수 중에서  $X_1, \dots, X_m$ 은 연속형이고,  $X_{m+1}, \dots, X_K$ 는 범주형이라고 하자. 그리고 노드  $M_k(t)$ 는  $t$ 에서 범주형 예측변수  $X_k (k = m+1, \dots, K)$ 의 범주 수를 나타낸다고 하자. 목표변수와 예측변수의 쌍으로 이루어진 서로 독립인  $N$ 개의 자료는 <식 1>과 같다. 여기서  $x$ 는 예측변수이고  $y$ 는 목표변수이며,  $N$ 은 관측값의 수이다.

$$(x_i, y_i), \quad i = 1, \dots, N \quad (\text{식 1})$$

##### 3.1.1 분리 변수의 선택

연속형과 명목형 변수가 섞여 있는 혼합 자료에서 두 변수의 연관성을 측정하기 위해 두 변수간의 통계적 독립성 검정 방법을 사용한다. 독립성 검정을 수행하는데 사용되는 검정 통계량의 유의 확률인 p-값을 사용하여 혼합자료에서의 변수들과 클래스 사이의 연관성을 측정한다. 두 변수의 관계에 따라서 명목형 변수와 명목형 변수의 연관성 측정은 피어슨의 카이제곱통계량, 연속형 변수와 명목형 변수의 연관성 측정은 크루스칼-왈리스 검정 그리고 연속형 변수와 연속형 변수의 연관성 측정은 스피어만의 순위 상관 통계량이 사용되어 통계적으로 가장 유의하게 목표변수와 연관된 예측 변수를 노드의 분리 변수로 선택한다. 각 노드에서 분리 변수를 선택하는 것은 연속형인 경우와 범주형인 경우 각 범주에서 최소의 값을 나타내는 p-값 즉  $\hat{\alpha}(k_1)$ 과  $\hat{\alpha}(k_2)$ 을 구한 후, <식 2>와 같은 조건을 만족하는 변수를 분리변수로 선택한다. 노드  $t$ 에서  $X_k$ 를 분리 변수로 선택한다[16].



$$\tilde{k} = \begin{cases} k_1, & \text{if } \hat{\alpha}(k_1) \leq \hat{\alpha}(k_2) \\ k_2, & \text{if } \hat{\alpha}(k_1) > \hat{\alpha}(k_2) \end{cases} \quad (\text{식 2})$$

### 3.1.2 분리 점의 선택

노드의 최적 분리 기준은 3.1.1절의 소개된 방법에 따라 선택된 분리 변수만을 이용한 Breiman[5] 등의 후진소거법을 적용하였다. 또한 3.1.1 절의 단계를 거쳐 선택된 분리 변수  $X_{\tilde{k}}$ 의 유형에 따라 분리 점을 찾게 된다. 분리 점은 분리 변수가 연속형인 경우와 범주형인 경우에 대하여 관측 값  $x_{1\tilde{k}}, \dots, x_{N(t)\tilde{k}}$ 과 목표변수의 평균값  $\bar{y}_1(t), \dots, \bar{y}_{M_f(t)}(t)$ 을 사용하여 다음 과정에 의해 결정한다.

단계 1 :  $X_{\tilde{k}}$ 의 관측 값(혹은 목표값)  $x_{1\tilde{k}}, \dots, x_{N(t)\tilde{k}}$ 를  $x_{1\tilde{k}} \leq \dots \leq x_{N(t)\tilde{k}}$ 와 같이 순서화한다.

단계 2 :  $l(=1, \dots, (N(t)-1))$ 에 대해  $x_{l\tilde{k}} \neq x_{l+1, \tilde{k}}$ 일 때  $x_{1\tilde{k}}, \dots, x_{l\tilde{k}}$  값을 갖는 사례는 왼쪽 노드( $t_L$ )로,  $x_{l+1, \tilde{k}}, \dots, x_{N(t), \tilde{k}}$  값을 갖는 사례는 오른쪽 노드( $t_R$ )로 보내어 <식 3>과 같이 평균제곱오차의 가중평균을 구한다. 단,  $p_L$ (또는  $p_R$ )은  $N(t)$ 개의 사례 중 노드  $t_L$ (또는  $t_R$ )에 포함된 사례수의 비율을 나타낸다.

$$s_l^2(t) = p_L s^2(t_L) + p_R s^2(t_R) \quad (\text{식 3})$$

단계 3 :  $x_{l\tilde{k}} \neq x_{l+1, \tilde{k}}$ 에서 최소  $s^2(t)$ 을  $\tilde{s}^2(t)$ 라 정의 할 때, 예측변수  $X_{\tilde{k}}$ 의 분리 점은  $\tilde{s}^2(t)$ 에 대응하는  $l$ 의 값 즉,  $x_{l\tilde{k}}$ 로 결정한다. 따라서 노드  $t$ 의 분리 기준은  $X_{\tilde{k}} \leq x_{l\tilde{k}}$ 으로 한다.

3.1.1절과 3.1.2절에 소개된 방법에 의해 노드를 분리하는 과정을 노드 내 개체 수가 미리 지정해 놓은 값 보다 큰 값을 갖는 한 계속 진행되다가 모든 노드에서 정지규칙(stopping rule)이 만족될 때 비로소 끝나게 된다.

## 3.2 실험 결과

CART와 SKES 알고리즘에서 구해진 운율 규칙을 6,000 문장 DB에 적용하여 합성음을 제작하였다. 남자 16 명과 여자 14 명 총 30 명을 대상으로 방음이 되지 않는 강당에서 MOS\* 시험을 실시하였다. 피시험자는 본 시험과 관련되어서 특별한 훈련을 받지 않았고, 즉 시험 내용에 대한 사전 지식

1) MOS(Mean Opinion Score) 테스트 : 다수 피험자를 대상으로 여러 가지 합성음을 들려준 뒤 각 합성음에 대해 1~5 사이의 점수를 매겨 이를 평균하여 최종 점수를 구한다. 각 점수별 의미는 다음과 같다.

(1 : Bad, 2 : Poor, 3 : Fair, 4 : Good, 5 : Excellent)

이 없이 무작위로 선택되어서 시험에 응하게 되었다. 임의의 10문장을 두 가지 방법에 의해 제시된 운율 정보를 바탕으로 합성된 음을 각각 2회씩 읽어주고 비교 평가하도록 하였다. 실험에 사용한 평가문장은 특정한 단어에 국한되어 제작된 것이 아니며, 일반적으로 자주 활용되는 문장으로 다음과 같은 문장이 시험 문장으로 사용되었다.

- 음성안내를 시작하겠습니다. 목적지가 설정되었습니다. 안전 운전하시기 바랍니다.
- 한 개의 메시지가 도착했습니다. 자동으로 연결합니다.

표 8. 문장별 알고리즘에 대한 항목별 평균 비교

실험대상 문장번호	남 자				여 자			
	CART		SKES		CART		SKES	
	명료도	자연성	명료도	자연성	명료도	자연성	명료도	자연성
1	4.00	3.00	3.81	3.40	4.00	2.88	3.77	2.88
2	4.04	3.18	3.72	3.31	4.00	3.00	4.00	3.00
3	3.72	3.22	3.36	3.45	3.66	3.00	3.44	2.88
4	3.00	2.77	3.31	3.09	3.33	2.33	3.55	2.55
5	3.59	3.13	3.81	3.40	3.44	2.88	3.88	3.22
6	5.45	2.90	3.59	3.18	3.33	2.66	3.33	2.77
7	3.13	2.59	3.72	3.31	2.88	2.22	3.55	3.11
8	3.13	2.45	3.50	3.04	3.00	2.44	3.55	2.55
9	3.50	3.00	3.22	2.72	3.11	2.22	3.22	2.55
10	2.95	2.55	3.18	2.68	2.33	2.33	3.00	2.33
전체	3.45	2.87	3.52	3.16	3.31	2.60	3.53	2.78

<표 8>는 각 문장별로 남녀로 실험한 결과를 나타낸 것이다. <표 9>는 문장별로 평가한 결과를 나타낸다. 전반적으로 SKES로 구현된 모든 합성음이 자연성에서 좋은 결과를 나타내었다.

표 9. 문장별 알고리즘에 대한 항목별 평균 비교

실험대상 문장번호	CART		SKES	
	명료도	자연성	명료도	자연성
1	4.00	2.96	3.80	3.25
2	4.03	3.12	3.80	3.22
3	3.70	3.16	3.38	3.29
4	3.09	2.64	3.38	2.93
5	3.54	3.06	3.83	3.35
6	3.41	2.83	3.51	3.06
7	3.06	2.48	3.67	3.25
8	3.09	2.45	3.51	2.90
9	3.38	2.77	3.22	2.67
10	2.77	2.45	3.12	2.58
전체	3.41	2.79	3.52	3.05

<표 10>은 각 방법에 따른 남녀 실험 결과를 평균 낸 것으로 전체적으로 명료성과 자연성에 대해서는 남자가 여자보다 더 나은 점수를 주었고, CART에 비하여 SKES 알고리즘이 두 가지 측면에서 더 좋은 결과를 얻은 것을 알 수 있다.

표 10. 성별에 따른 실험 결과 평균 값

실험대상	사용방법	CART		SKES	
		명료도	자연성	명료도	자연성
남자		3.45	2.87	3.52	3.16
여자		3.31	2.60	3.53	2.78
전체		3.41	2.79	3.52	3.05

### 3.3 결과 분석 및 문제점 제기

<표 8-10>에 나타난 실험 결과를 살펴보면, 남자는 여자에 비하여 비교적 좋은 점수를 주었으며, 실험에 사용된 문장별 시험 결과에서 명료성과 자연성이 서로 상반된 결과를 나타내는 경우가 발견되었다. 실험의 결과에 대한 신뢰성 평가를 필요로 하나, 현재 실험 데이터로는 충분한 검증이 어렵다고 판단된다. 남녀 성별의 차이에 따른 실험 결과에 대한 성향에 대한 원인 분석은 하지 못하였고, 실험에 임한 몇 명의 남녀의 결과로 전체적인 성별 구분에 대한 결론을 내리기에는 미흡하다고 본다. <표 10>을 제시한 이유는 <표 8>과 <표 9>에 나타난 실험결과로 남녀에 대한 동향을 알아보기 위함이다.

전체적으로 SKES의 적용이 자연성에서 좋은 결과를 나타내지만, 명료성은 시험 문장에 따라 오히려 낮은 경우를 발견할 수 있었다. CART와 SKES 알고리즘이 제시한 운율 예측 성능 비교 검증을 위해서는 문장 구성시 어떤 운율이 선택되어졌는지에 대한 보다 면밀한 비교 분석이 필요하며, 자연성과는 어떤 관계가 있는지도 규명할 필요가 있다고 본다. 단, 본 논문에서는 SKES 알고리즘의 적용에 의한 시험 결과를 CART의 경우와 비교함으로써 운율 예측에 대한 새로운 모델의 가능성을 제시하였다는데 의미를 둔다. 현재까지 얻은 실험결과로 확정적인 결론을 내리기에 신뢰성 문제를 충분히 고려했다고 할 수는 없다고 본다. 단지, 본 논문에서 제시한 실험 결과는 전체적인 성향과 가능성을 제시한 것으로 평가하고 있다.

## 4. 결 론

본 논문에서는 결정트리 알고리즘인 CART를 이용하여 학습 음성 DB 6,000 문장에서 얻어진 통계적인 결과를 이용하여 한국어 TTS의 운율을 예측하는 과정과 실험 결과를 제시하였고, 동일한 방법으로 SKES 알고리즘을 이용하여 얻어진 합성음과의 MOS 실험을 통한 성능 비교를 실시하였다.

CART를 이용한 운율경계 예측시 좌품사와 우품사의 결정규칙은 형태소분석기의 결과를 그대로 적용하거나, 부분적으로 예외 규칙을 사용하였다. ZERO, AP, IP 경계 예측 실험 결과로는 전체적으로 만족할 만하지는 않지만, IP를 ZERO로, ZERO를 IP로 판단하는 오 분류율은 비교적 낮으므로,

음성 합성시 문장에 대한 의미가 잘못 전달될 가능성은 높지 않은 것으로 판단되어 실험 결과를 그대로 적용하였다. 음소의 지속시간은 평균 성공률은 70%대로 ZERO, AP, IP 경계 예측시 개선의 여지가 있다고 여겨진다. CART의 문제점으로 지적되는 범주의 수가 많은 예측변수를 선호하는 편인 음소들을 그룹화 시켜 범주의 수를 줄이는 방법으로 F-Test 통계량에 의한 분할 방법을 적용하였다.

또한, 변수선택의 편향이 적은 회귀 나무를 만들기 위해 제안된 SKES 알고리즘으로 같은 음성 데이터 크기에 적용하여 음질을 비교하였다. 평가를 위해 20 대 남녀 30 명을 대상으로 MOS시험을 실시하였고, 그 결과 자연성과 명료성 부분에서 SKES 알고리즘을 적용한 것이 각각 0.26 점과 0.11 점 향상한 것을 알 수 있었다. CART와 SKES 알고리즘 적용에 의해 얻어진 통계적인 수치가 어떠한 의미를 나타내며, 합성음의 성능 향상에 어떠한 영향을 미쳤는지에 대한 정확한 평가는 아직 규명하지 못했지만, 본 논문에서는 SKES 알고리즘의 적용에 의한 시험 결과를 비교함으로써 운율 예측에 대한 새로운 모델의 가능성을 제시하였다는데 의미를 둔다.

향후 두 알고리즘에 대한 좀 더 구체적인 선택 규칙에서의 성능 비교를 실시하여 그 원인을 규명하여 좀 더 안정적이고 적합한 운율예측 알고리즘에 대한 연구를 계획하고 있다. 또한, 운율 경계 예측의 정확도를 높이기 위해서 문장의 간단한 구문관계 정보가 필요하다. 즉, 구 또는 절간의 의존관계 정보만 있어도 충분히 운율 경계 예측의 정확도를 높이는데 효과적일 수 있다. 그러므로 이에 대한 연구도 본 논문에서 제시한 운율예측 알고리즘 연구와 더불어 연계해서 진행되어야 할 것으로 판단된다.

## 참 고 문 헌

- [1] 박상언. 2001. *코퍼스 기반 한국어 음성합성 시스템의 합성음 자연성 향상*. 전남대학교 대학원 석사논문.
- [2] 미디어젠. 2006. *음성 HMI 전용 단말기용VUI 모듈 개발 보고서*.
- [3] 이상호. 2000. *한국어 TTS시스템을 위한 운율의 트리기반 모델링*. KAIST 박사논문.
- [4] *An Overview of the CART Methodology*. <http://www.salford-systems.com>
- [5] Breiman, L., Friedman, J., Olshen, H. R., A. & Stone, C. J. 1984. *Classification and regression trees*. Belmont: Wadsworth.
- [6] Breiman, L. 1996. "Bagging predictors." *Machine Learning* 24, pp. 123-140.
- [7] 권오일, 홍문기, 강선미, 신지영. 2002. "코퍼스 방식 음성합성기에서의 개선된 운율구 경계 예측", *음성과학* 9(3), 25-34.
- [8] 홍문기, 신지영, 강선미. 2003. "한국어 음성합성기의 성능 향상을 위한 합성 단위의 유무성음 분리", *음성과학* 10(2), 7-14.
- [9] 김민호. 2004. *회귀나무 구성에서 변수선택 편향에 대한 연구*. 수원대학교 대학원 석사논문.
- [10] 권영식. 2005. *회귀나무에서 편향이 적은 변수선택에 관한 연구*. 성균관대학교 대학원 석사논문.
- [11] 신지영, 차재은. 2003. *우리말 소리의 체계 -국어 음운론 연구의 기초를 위하여-*, 서울: 한국문화사.
- [12] 이상호. 1995. *미 등록어를 고려한 한국어 품사 태깅 시스템 구현*. KAIST 석사논문.
- [13] 장석환. 2001. *분산분석*. 서울: 영지문화사.
- [14] Quinlan, J. R. 1993. *C4.5 Programs for machine learning*. San Mateo: Morgan Kaufmann.

- [15] Quinlan, J. R. 1996. Improved Use of Continuous Attributes in C4.5, *Journal of Artificial Intelligence Research* 4, 77-90.
- [16] 손진식, 송수정, 정창훈, 강역식 공역. 2004. *최신 통계학*. 서울: 東和技術.

접수일자: 2007. 5. 3

게재결정: 2007. 5. 29

▲ 강선미

서울특별시 성북구 정릉 4동 (우: 136-704)

서경대학교 컴퓨터과학과

Tel: +82-2-940-7291 Fax: +82-2-919-5075

E-mail: smkang@skuniv.ac.kr

▲ 권오일

경기도 이천시 부발읍 아미리 산 136-1 번지 (우: 467-866)

현대오토넷

Tel: +82-31-639-7817 Fax: +82-31-639-7817

E-mail: koi@hyundaiatonet.com