

코퍼스 방식 음성합성에서의 개선된 운율구 경계 예측

AP, IP Prediction For Corpus-based Korean Text-To-Speech

권 오 일* · 홍 문 기** · 강 선 미** · 신 지 영***
Ohil Kwon · Munki Hong · SunMee Kang · Jiyoung Shin

ABSTRACT

One of the most important factor in the performance of Korean text-to-speech system is the prediction of accentual and intonational phrase boundary. The previous method of prediction shows only the 75-85% which is not proper in the practical and commercial system. Therefore, more accurate prediction must be needed in the practical system. In this study, we propose the simple and more accurate method of the prediction of AP, IP.

Keywords: Corpus, CART, TTS, AP, IP, Gini, Classification Tree

1. 서 론

최근에 음성 인식 및 음성 합성 기술을 이용한 제품에 대한 수요가 높아지면서, 한국어 음성처리 성능 향상을 목적으로 다각도로 연구되어지고 있다. 특히 음성 합성 관련 분야는 ARS, 음성 메일, 기타 음성 관련 제품들을 통하여 급속히 확산되고 있는 실정이다.

본 논문에서 소개하는 TTS 시스템은 현재까지 개발된 시스템들의 다음과 같은 문제점을 개선하는 데 중점을 두고 연구되었다.

첫째, 기존의 CART를 사용한 연구들은 주로 음성DB로부터 높은 운율 예측률을 생성하는 데에 많은 시간과 노력을 기울였다. 하지만 언어적 분석의 한계로 보통 80% 전 후반대의 예측률로 합성기의 운율을 생성하게 되는데 20% 내외의 오류가 청취자로 하여금 듣기 거부한 소리를 내게 하고 있다.

둘째, 자연어 처리에 필요한 구문분석기를 사용하여 DB에 필요한 파라미터들을 사용하는 경우가 많은데 이는 DB에서 중요한 파라미터로서 작용하지만 구축하기도 어렵고 분석한 내용 또한 오류가 많아서 운율 예측을 하는데 파라미터로 사용하기가 어렵다.

이에 반해 본 연구에서는 운율 예측에 있어 높은 예측률이 아닌 듣기 자연스러운 물을 CART로부터 얻어내어 위 두 가지 문제를 모두 해결하고자 한다.

* (주) 현대오토넷

** 서경대학교 컴퓨터과학과

*** 나사렛대학교 언어치료학과

2. 본 론

음성 합성기를 기능별로 살펴보면 언어 처리부, 운율 생성부, 음성 합성부로 나눌 수 있다. 본 논문은 운율 생성부에 해당하는 운율 경계부를 추출하기 위한 방법[1]에 관한 연구이며, NONE, AP, IP의 3 단계로 나누어 운율 경계 강도를 측정하였다. 실험에 사용된 문장은 본 연구를 위해서 언어학 전문가로부터 선정된 400 문장으로서, 관측된 어절 수, 총 4,450 개를 데이터 베이스로 사용하였다. 음성 합성기에 사용되는 운율 요소들을 예측하는데 일반적으로 많이 사용하는 방법에는 HMM을 이용하거나 CART 통계 툴(tool)을 들 수가 있다. 이 중 우리는 확률상[2]으로 더 높게 나오는 CART 툴을 사용하기로 하였다. 다음과 같은 품사 태거[4]로부터 얻을 수 있는 파라미터(IP 파라미터 제외)를 사용하여 AP와 IP를 예측 모델을 구현하였다.

2.1 사용된 파라미터

- IP : 해당 어절의 IP 유무
- DPOS : 해당 어절의 대표품사
- DLPOS : 해당 어절의 좌품사
- DPLPOS : 앞 어절의 좌품사
- DPPLPOS : 앞 앞 어절의 좌품사
- DNLPOS : 다음 어절의 좌품사
- DNNLPOS : 다음 다음 어절의 좌품사
- DRPOS : 해당 어절의 우품사
- DPRPOS : 앞 어절의 우품사
- DPPRPOS : 앞 앞 어절의 우품사
- DNRPOS : 다음 어절의 우품사
- DNNRPOS : 다음 다음 어절의 우품사
- C_LOCFRBEG : 문장 앞에서부터의 어절위치
- C_LOCFREND : 문장 뒤에서부터의 어절위치

위 파라미터들을 사용하여 CART의 DB를 구축하였으며, 각 품사정보에는 72 개의 품사가 있는데 이를 특징적인 카테고리 39 개로 구분하여 품사 파라미터로 재구성하여 사용하였다. 그 종류는 다음과 같다.

표 1. 39 개의 카테고리로 분리된 품사셋

| 번호 | 품사종류 | 번호 | 품사종류 |
|----|-------------|----|----------------------|
| 1 | 보통명사(nc) | 21 | 대동적연결어미(ecq) |
| 2 | 의존명사(nb) | 22 | 종속적연결어미(ecs) |
| 3 | 인칭대명사(npp) | 23 | 보조적연결어미(ecx) |
| 4 | 지시대명사(npd) | 24 | 관형사형전성어미(exm) |
| 5 | 숫자(nnn) | 25 | 명사형전성어미(exn) |
| 6 | 수사(nn) | 26 | 선어말어미(efp) |
| 7 | 용언(p) | 27 | 종결어미(ef) |
| 8 | 보조용언(px) | 28 | 명사접미사(xn) |
| 9 | 관형어(m) | 29 | 동사파생접미사(xpv) |
| 10 | 부사(a) | 30 | 형용사파생접미사(xpa) |
| 11 | 단어접속부사(ajw) | 31 | 부사파생접미사(xa) |
| 12 | 문장접속부사(ajs) | 32 | 섬표(s _i) |
| 13 | 감탄사(i) | 33 | 종결(s _i) |
| 14 | 격조사(jc) | 34 | 여는따옴표(s') |
| 15 | 관형격조사(jcm) | 35 | 닫는따옴표(s') |
| 16 | 호격조사(jcv) | 36 | 이음표(s ₋) |
| 17 | 부사격조사(jca) | 37 | 단위(su) |
| 18 | 서술격조사(jcp) | 38 | 화폐단위(sw) |
| 19 | 보조사(jx) | 39 | 기타(sy) |
| 20 | 접속조사(jj) | | |

다음은 사용된 파라미터에서 대표품사(좌품사와 우품사의 관계)셋을 위 표 1. 품사셋을 근거로 CART에서 사용하기 위해 정의하였다.

표 2. 17 개의 카테고리로 분류된 대표 품사셋

| 번호 | 대표 품사 |
|----|--------------------|
| 1 | “명사류+부사격조사” |
| 2 | “명사류+목적격조사” |
| 3 | “동사류어간+주, 보조적연결어미” |
| 4 | “동사류어간+종속연결어미” |
| 5 | “부사어” |
| 6 | “명사류+주격조사” |
| 7 | “동사류어간+종결어미” |
| 8 | “동사류어간+명사형어미” |
| 9 | “명사류+속격조사” |
| 10 | “동사류어간+관형사형어미” |
| 11 | “동사류어간+인용형조사” |
| 12 | “관형어” |
| 13 | “의존명사” |
| 14 | “명사류+연결조사” |
| 15 | “접속사” |
| 16 | “명사류+보조사” |
| 17 | “명사” |

본 논문에서 운을 경계 강도 예측을 위해 사용한 트리기반 모델링 방법(Classification Tree)은 Splitting Method로 Gini 인덱스를 사용하였으며 예측한 결과를 테스트하기 위해 400 단어의 DB를 3 : 1로 나누어 Learning Data와 Test데이터로 사용하여 트리를 구성하였다. 운을 경계 강도는 AP, IP 이 두 가지를 구하며 AP는 IP파라미터를 포함하는 14 개의 파라미터를 사용하는 실험과 IP파라미터를 제외한 13 개의 파라미터를 사용하는 실험으로 나누어 총 3 개의 예측 실험을 하였다.

2.2 CART를 이용한 운을 경계 강도 예측: IP

본 실험에서는 IP(운을경계강도 2)를 예측하기 위해 다음과 같은 두 가지 방법으로 실험을 하였다.

[방법 1] no Linear Combination - CART의 디폴트 옵션.

[방법 2] Linear Combination을 사용하여 splitting의 성능 향상

보통 한 문장은 한 어절부터 10 어절 이상으로 이루어져 있으나 전처리 과정에서 문장을 80문자에 해당하는 어절의 앞뒤에서 끊어서 언어 처리부에 넘겨주게 된다.

언어 처리부에선 전처리에서 넘어온 문장을 가지고 IP예측을 하게 되는데 IP 예측에 있어 중요한 것은 CLASS 1(IP가 있는 부분)에서의 예측률이 자연스러운 발화에 큰 영향을 주게 된다. 그 이유는 IP가 있는데 없는 것으로 오류 예측된다면 쉬어야 할 곳에서 쉬지 못한다면 청취자로 하여금 숨가쁘게 들릴게 하며 정보전달에 있어서 정확히 인지하지 못하게 된다.

표 3. [방법 1]의 예측 확률(Learn Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 2,410 | 86.058 | 2,074 | 336 | 83.42% |
| 1(IP) | 956 | 76.778 | 222 | 734 | |

표 4. [방법 1]의 예측 확률(Test Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 804 | 84.826 | 682 | 122 | 82.01% |
| 1(IP) | 280 | 73.929 | 73 | 207 | |

위에 표 3과 표 4는 일반적인 Gini 방법을 사용하여 학습 예측률과 오픈 테스트 예측률을 보인 것인데 CLASS 0은 예측이 비교적 잘 되었으나, CLASS 1은 80% 미만으로서 운을 적용하기에는 적당하지 않다. 특히 합성기가 ARS와 같은 시스템에 사용하고자 할 경우 정보전달 쪽에 우선점을 두고자 할 것이다. 그럼 어떻게 하면 CLASS 1이 높은 확률을 가지고 있는 룰을

만들 수 있겠는가에 대한 문제점을 생각할 수 있다.

그 방법의 제안으로 Gini와 함께 Linear Combination을 사용하는 것인데 이렇게 하면 Classification Tree를 구성할 때 각 노드의 조건이 one variable이 아닌 조합된 조건을 가짐으로써 좀더 효과적인 트리를 구성하게 된다. 다음 표 5를 보면 기여도면에서 상위 조건들이 Linear Combination을 사용하게 되어 높은 기여도를 보여주고 있다.

표 5. Cost에 따른 파라미터 기여도 비교

| [방법 1] no Linear Combination | | [방법 2] Linear Combination | |
|------------------------------|--------|---------------------------|--------|
| Variable | Score | Variable | Score |
| DRPOS | 100.00 | DRPOS | 100.00 |
| DNRPOS | 79.43 | DNRPOS | 88.61 |
| C_LOCFRE | 70.36 | C_LOCFRE | 80.59 |
| DNLPOS | 67.42 | DNLPOS | 75.78 |
| DNNRPOS | 34.65 | DNNLPOS | 34.02 |
| DNNLPOS | 33.35 | DNNRPOS | 13.69 |
| DMPOS | 9.10 | DLPOS | 3.74 |
| DPRPOS | 4.41 | DMPOS | 3.05 |
| DLPOS | 4.08 | C_LOCFRB | 1.00 |
| C_LOCFRB | 1.72 | DPPRPOS | 0.47 |
| DPPRPOS | 0.50 | DPRPOS | 0.20 |
| DPLPOS | 0.44 | DPPLPOS | 0.06 |
| DPPLPOS | 0.04 | DPLPOS | 0.00 |

Linear Combination을 사용했을 때의 예측 확률을 보면 다음과 같다.

표 6. [방법 2]의 예측 확률(Learn Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 2410 | 79.336 | 1912 | 498 | |
| 1(IP) | 956 | 83.264 | 160 | 796 | 80.45% |

표 7. [방법 2]의 예측 확률(Test Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 804 | 80.100 | 644 | 160 | |
| 1(IP) | 280 | 78.929 | 59 | 221 | 79.98% |

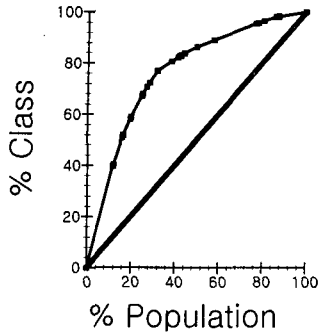


그림 1. [방법 1]의 이득 그래프

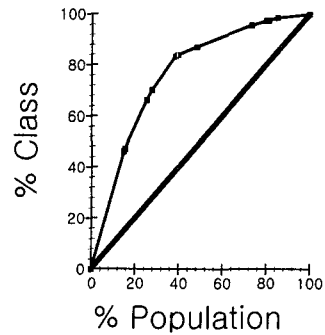


그림 15. [방법 2]의 이득 그래프

이득 그래프만을 봐서는 그림 1이 더 나아 보이지만 표 7을 보면 표 4에 비해서 CLASS 1의 확률이 더 좋아진 것을 알 수 있다. 전체 확률을 끌어올리는 것보다 정확한 정보 전달을 위해 필요한 부분확률을 높이는 것이 좀더 듣기에 거북하지 않는 운율을 만들어 낼 수 있다.

2.3 CART를 이용한 운율 경계 강도 예측: AP

본 실험은 AP(운율경계강도 1)를 예측하기 위한 실험으로써 다음 두 가지 방법으로 나누어 비교 분석하였다.

[방법 1] IP파라미터를 DB로 사용한 경우

[방법 2] IP파라미터를 DB로 사용하지 않은 경우

14 개의 파라미터 중 IP 파라미터는 1 절에서 소개한 실험에서 예측한 IP 예측치를 가져와서 파라미터로 사용한다. IP가 AP의 파라미터가 될 수 있는 이유는 IP구간이 있는 부분은 AP 구간도 있다고 규정지을 수 있기 때문이기도 하지만, 이때에 IP를 적용하기 때문에 IP가 기여도가 가장 높은 파라미터로 사용됨을 표 6의 [방법 1]에서 볼 수 있다. 따라서 다른 파라미터들은 비교적 낮은 수치를 갖게 되고 [방법 2]와 비교했을 때 Cost(Classification Tree를 구성하는데 드는 노드수에 비례한 어려움)를 절감할 수 있다.

표 8. IP파라미터에 따른 기여도 비교

| [방법 1] IP파라미터 사용 Minimum Cost = 0.334 | | [방법 2] IP파라미터 미사용 Minimum Cost = 0.616 | |
|--|--------|---|--------|
| Variable | Score | Variable | Score |
| IP | 100.00 | DRPOS | 100.00 |
| DRPOS | 42.59 | DNLPOS | 84.02 |
| DNLPOS | 37.28 | DNRPOS | 40.17 |
| DNRPOS | 31.70 | C_LOCFRE | 68.93 |
| C_LOCFRE | 26.53 | DNNLPOS | 31.77 |
| DLPOS | 2.38 | DMPOS | 14.85 |
| DMPOS | 2.32 | DNNRPOS | 10.98 |
| DNNRPOS | 2.10 | DLPOS | 10.68 |
| DNNLPOS | 1.71 | C_LOCFRB | 9.70 |
| DPPLPOS | 1.36 | DPRPOS | 8.97 |
| DPPRPOS | 1.16 | DPLPOS | 5.82 |
| DPRPOS | 0.67 | DPPLPOS | 5.71 |
| DPLPOS | 0.61 | DPPRPOS | 5.49 |
| C_LOCFRB | 0.57 | | |

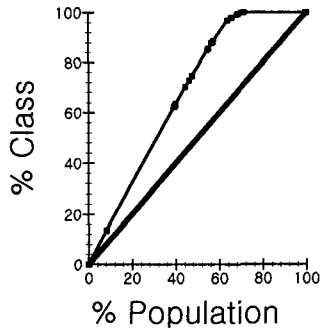


그림 3. [방법 1]의 이득 그래프

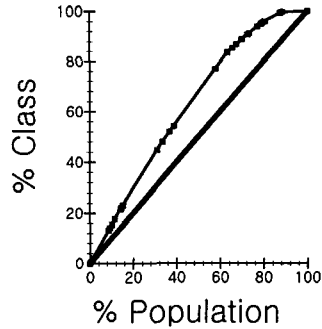


그림 4. [방법 2]의 이득 그래프

이득 또한 그림 3이 그림 4보다 조금 더 높은 것을 알 수 있다.

표 9. [방법 1]의 예측 확률(Learn Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체 예측확률 |
|---------|------------|-----------------|---------|---------|---------|
| 0(NONE) | 1577 | 73.811 | 1164 | 413 | |
| 1(AP) | 1789 | 96.758 | 58 | 1731 | 86.0% |

표 10. [방법 1]의 예측 확률(Test Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 491 | 70.468 | 346 | 145 | 84.50% |
| 1(AP) | 593 | 96.121 | 23 | 570 | |

표 11. [방법 2]의 예측 확률(Learn Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 1577 | 59.417 | 937 | 640 | 72.22% |
| 1(AP) | 1789 | 83.510 | 295 | 1494 | |

표 12. [방법 2]의 예측 확률(Test Data)

| Class | Total Case | Percent Correct | Class 0 | Class 1 | 전체예측확률 |
|---------|------------|-----------------|---------|---------|--------|
| 0(NONE) | 491 | 55.804 | 274 | 217 | 69.84% |
| 1(AP) | 593 | 81.450 | 110 | 483 | |

표 10과 표 12를 비교해서 보면 알 수 있듯이 앞서 예측한 IP파라미터를 AP에서 활용함에 따라서 예측 확률의 향상과 트리 구성의 비용절감 오류 노드수도 줄어드는 것을 볼 수 있다. 여기서 오류 노드수라 함은 Class 0(NONE)을 Class 1(AP)로 보거나 Class 1을 Class 0으로 잘못 예측한 경우이다. 여기서 Class 1의 확률이 중요한데 Class 1의 예측이 잘못되면 앞 어절의 종성이 뒤 어절의 초성으로 넘어가는 경우가 생기게 된다.

예문) 바닥에/넘어지면/아프다.

오류 → 바닥에/넘어지거나프다

예문과 같은 결과로 정보전달에 문제가 생길 수 있으므로 [방법 1]과 같이 Class 1의 확률이 높은 것이 좋다.

3. 결 론

본론에 소개된 IP 예측 실험을 통해서 알 수 있는 것은 Linear Combination Split방법을 써서 전체 확률의 상승이 아닌 정보 전달을 위한 요소의 향상을 꾀하는데 있다는 점이다. 이렇게 얻은 IP 정보를 2 절에 소개된 실험 [방법 1]에서 입력 파라미터로 사용하여 cost를 줄이고 예측률을 향상되는 것을 표와 그림으로 확인할 수 있다.

그렇다면 이 파라미터를 사용하지 않는다면 얼마나 많은 비용이 들고 확률이 떨어질 것인가에 대해서 2 장의 실험 [방법 2]에서 IP 파라미터를 제외하고 실험에 주어진 나머지 13 개의 파라미터를 가지고 실험해 보았다. 이 실험결과를 통해서 알 수 있는 것은 cost가 비싸고 예측 확률의 저하로 합성기에 적용하기에는 힘들다는 것을 알 수 있다.

따라서 합성기에 적용하기 위한 운율 요소를 예측하기 위해서는 Cost를 줄일 수 있는 파라미터 선정을 하고, 그 파라미터들을 근본으로 하여 정보전달을 명확하게 하기 위한 요소를 찾아내어 그 확률을 높이는 것이 본 논문에서 제안된 내용이라고 볼 수 있다.

참 고 문 헌

- [1] International Telematics Exhibition & Conference. 2002. 한국커머스넷, 165-178.
- [2] 이상호, 오영환. 1998. "CART를 이용한 운율구 추출 및 휴지시간 모델링." 제15회 음성통신 및 신호처리 워크샵, 81-86.
- [3] 박상언. 2001. 코퍼스 기반 한국어 음성합성 시스템의 합성음 자연성 향상. 전남대 석사학위논문.
- [4] 이상호. 2000. 한국어 TTS시스템을 위한 운율의 트리기반 모델링. 한국과학기술원 박사학위논문.
- [5] 신지영. 2001. 말소리의 이해. 서울: 한국문화사
- [6] Huang, Xuedong, Alex Acero & Hsio Wuen Hon. 2000. *Spoken Language Processing*. Carnegie Mellon University, 679-89.
- [7] 심준혁, 김준석, 차정원, 이근배. 1999. "통계와 규칙을 이용한 강인한 품사태거." 제1회 MATEC99 평가대회 논문발표집, 전자통신연구소.
- [8] Kim, Byeongchang & Geunbae Lee. 1999. "Statistical/Rule-based Hybrid Phrase Break Detection." *Proceedings of the ICSP '99*, 595-599.
- [9] Sanders, Eric. 1995. *Using Probabilistic Methods to Predict Phrase Boundaries for a Text-to-Speech System*. MS Thesis of University of Nijmegen.
- [10] Charniak, E. 1994. *Statistical Language Learning*. MIT Press.

접수일자: 2002. 7. 14.

게재결정: 2002. 9. 8.

▲ 권오일

서울특별시 강남구 역삼동 823-21 대공빌딩 7층 (우: 135-080)
 (주) 현대오토넷 차장
 Tel: +82-2-3016-9457 Fax: +82-2-3016-9460
 E-mail: koi@haco.co.kr

▲ 홍문기

서울특별시 성북구 정릉동 16-1 북악관 502호 (우: 136-704)
 서경대학교 컴퓨터학과
 Tel: +82-2-940-7291 Fax: +82-2-919-5057

E-mail: mkhong@ihci.skuniv.ac.kr

▲ 강선미

서울특별시 성북구 정릉동 16-1 (우: 136-704)

서경대학교 컴퓨터과학과

Tel: +82-2-940-7291 Fax: +82-2-919-5057

E-mail: smkang@skuniv.ac.kr

▲ 신지영

충남 천안시 쌍용동 456번지 (우: 330-718)

나사렛대학교 언어치료학과

Tel: +82-41-579-7941

E-mail: jyshin@kornu.ac.kr