코퍼스 기반 무제한 단어 중국어 TTS

*ZhengYu, *하주홍, **김병창, *이근배
*포항공과대학교 컴퓨터공학과
**위덕대학교 컴퓨터멀티미디어공학부

Corpus Based Unrestricted vocabulary Mandarin TTS

*ZhengYu, *Ju-Hong Ha, **Byeongchang Kim, *Gary Geunbae Lee
*Dept. Of CSE, POSTECH
**Division of Computer and Multimedia Engineering, UIDUK University
E-mail: {zhengyu, miracle, bckim, gblee}@postech.ac.kr

Abstract

In order to produce a high quality (intelligibility and naturalness) synthesized speech, it is very important to get an accurate grapheme-to-phoneme conversion and prosody model. In this paper, we analyzed Chinese texts using a segmentation, POS tagging and unknown word recognition. We present a grapheme-to-phoneme conversion using a dictionary-based and rule-based method. We constructed a prosody model using a probabilistic method and a decision tree-based error correction method. According to the result from the above analysis, we can successfully select and concatenate exact synthesis unit of syllables from the Chinese Synthesis DB.

I. 머리말

중국어 음성 합성에 대한 연구는 비교적 늦은 80년대로부터 시작됐다. 중국어 음성 합성 연구 중에 비교적 대표적인 시스템은 중국 과학연구원 음성학 연구소의 KX-PSOLA, 중국청화대학의 TH_SPEECH, 중국과학기술대학의 KDTALK, Microsoft 중국 연구소의 MSR Asia Mandarin TTS, 그리고 Bell 실험실의 Mandarin TTS이다. 상위의 시스템들은 대부분 PSOLA방법을 사용하여 합성을 하였으며 명료성은 상당한 수준에 도달하였지만 다른 언어들이 가지고 있는 문제점과 마찬가지로 자연성에 있어서 아직 높은 수준까지는 되지 못했다[1].

중국어는 톤이 있는 언어로 각각의 음절에 고유의 톤을 포함하고 있어 문장의 운율적인 변화가 다양하다. 이러한 다양한 운율을 자연스럽게 표현하기 위하여 본 논문에서는 말뭉치 기반의 음성 합성 방법을 사용하였다. 또한 중국어 텍스트 분석에서는 발음 생성에 많은 영향을 주는 미등록어 처리를 하였고 보다 자연스러운 합성음 생성하기 위하여 확률적 방법과 에러 수정 후처리 방법을 사용하여 띄여 읽기 처리를 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 중국어 단어 분할과 품사태깅, 미등록어 처리의 자연어 처리 모듈에 대해 기술하고, 3장에서는 문자 발음 변환에 대하여 살펴본다, 4장에서는 띄어읽기 단위인 언절경계 추정에 대해 기술한다. 5장에서는 음성DB로부터 합당한 음성단위를 선정하여 합성하는 방법에 대하여 기술한다.

II. 텍스트의 자연어 처리와 분석

(1)말뭉치와 사전 구성

중국어의 언어 처리를 위하여 본 논문에서는 북경인민일보 1998년 1월부터 6월까지 6개월 분의 기사를 말뭉치[2]로 사용을 했다. 이 말뭉치는 약 650만개의 단어로 구성되어 있고 수작업으로 단어 분할과 태깅이 미리 되어 있다.

본 논문에서 구현한 중국어 언어처리 부분은 시스템과 사전의 유연성과 적응성을 최대한 유지하기 위하여 모든 사전을 자동 학습을 통하여 구축했다. 단어 분할과 품사 태거에 사용할 사전들로는 어휘품사사전, 단어품사 확률사전, 이진 품사 확률사전들로 구성 되였다. 자동 학습된 어휘사전의 단어는 약 18만개다.

(2)단어의 분할

본 논문에서 구현한 단어 분할 시스템은 그림1에서 보는 것과 같다.

입력된 중국어 문장은 전처리 과정을 거치게 된다. 전처리 과정은 특수 부호 및 단어들을 이용하여 문장을 짧은 문장으로 분할함으로서 단어 분할의 정확도와 처리 시간을 줄이는 작용을 한다. 특수 부호 및 단어들은 표기, 숫자, 외국어, 기타 비 중국어 부호와 출현 확율이 높고 단어 조합 능력이 낮은 한자를 말한다.

이렇게 전처리 과정을 거친 후에는 말뭉치에서 자동 학습된 어휘품사사전을 기반으로 전처리 과정에서 얻은 짧은 문자열에서 가능한 모든 단어를 추출한 후 Chih-Hao Tsai가 구현한 단어 분할 시스템[3]에서의 4가지의 휴리스틱 규칙들에 따라 단어 분할을 진행한다. 먼저 위에서 추출한 단어들에서 오른쪽으로 최장 일치되는 3개의 단어들을 하나의 묶음(chunk)단위로 묶고, 최장 길이의 묶음에서 처음 단어를 선택한다. 최장 길이의 묶음이 여러 개이면 다음 규칙으로 넘어간다. 두 번째 규칙은 평균 단어 길이가 최장인 묶음의 첫 단어를 선택하고, 평균 단어 길이가 최장인 묶음이 여러 개이면 다음 규칙을 적용한다. 세 번째 규칙은 단어 길이의 변화가 가장 작은 묶음에서 첫 단어를 추출한다. 이러한 묶음도 여러 개가 존재한다면 마지막 네 번째 규칙을 적용하게 된다. 여기에서는 자동 학습을 통해 생성된 중국어 문자들의 출현 빈도 사전을 이용하여 각 묶음 내의 각각의 문자들의 빈도 합이 가장 큰 묶음의 첫 단어를 추출하게 된다.

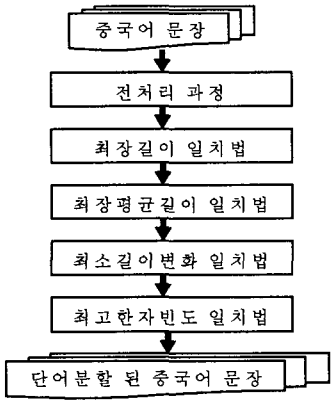

그림 1 중국어 단어 분할

(3)HMM기반 중국어 품사 태깅

단어 분할 시스템의 결과로 얻어진 중국어 문장들은 어휘품사사전을 이용하여 문장 내에 단어들이 부여될 수 있는 모든 품사들을 할당한다. 이렇게 초기의 태깅된 문장을 생성한 다음, 각각의 단어들에 문맥상 가장 적합한 품사를 부여하는 것이 우리가 얻고자 하는 품사 태깅 결과이다. 즉 주어진 문장 $W=w_1w_2.....w_n$에 대해 문맥상 가장 적합한 품사들 $T=t_1t_2.....t_n$을 구하는 것이다. 수식으로 표현하면,

$$T^* = \arg\max_T P(T/W)$$

----------(1)

로 표현할 수 있다. (1)에서 $P(T/W)$는

$$P(T/W) = \frac{P(T)P(W/T)}{P(W)}$$ ---------(2)

로 나타낼 수 있다. 식 (2)를 Markov가정에 따라 아래 식 (3)과 (4)에 의해 단순화 되고,

$$P(T) \approx P(t_1) \prod_{i=2}^{n} P(t_i/t_{i-1})$$

---------(3)

$$P(W/T) \approx \prod_{i=1}^{n} P(w_i/t_i)$$ --------(4)

$P(W)$는 상수이므로,

$$T^* = \arg\max_T \prod_{i=1}^{n} P(t_i/t_{i-1})P(w_i/t_i)$$

--------(5)

식 (5)에 따라 Viterbi 검색 알고리즘을 거치게 되면 입력된 문장 속의 단어들에 가장 적합한 품사들로 태깅된 최종 문장을 출력하게 된다[그림2].

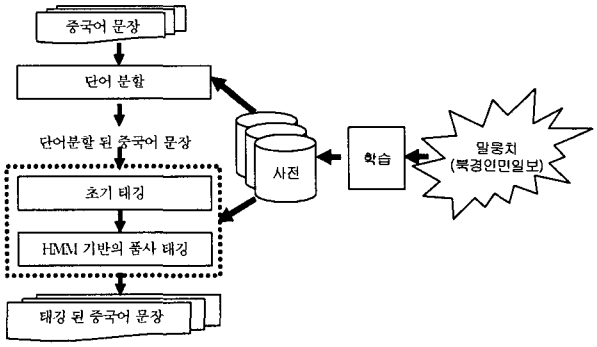

그림 2 중국어 HMM 품사 태깅

(4) 미등록어 추정

미등록어 추정은 SVM[4] 기반 2단계 처리 모듈을 사용한다. SVM은 기계 학습 방법들 중 가장 좋은 성능을 보이는 방법 중 하나이다. 하지만 다량의 데이터에 대해 수행 속도가 느리기 때문에 보다 효율적인 방법이 필요하다. 본 연구에서는 이러한 문제를 해결하기 위해 문장 전체를 검사하는 대신 문장 내 미등록어가 발생 가능한 구간을 미리 추정하고, 그 구간에 대해서만 미등록어 추정을 실시하여 고속의 TTS용 실시간 미등록어처리 알고리즘을 개발한다.

미등록어 후보 구간선정은 단일 한자가 연속적으로 발생하는 구간을 대상으로 하며 보다 정확한 성능을 위해 2단계 과정을 거치게 된다. 우선 말뭉치로부터 학습된 단일 한자 단어의 확률로 후보 구간을 줄이고, 개체명에 대한 간단한 바이그램 패턴으로 최종 미등록어 후보 구간을 선정한다.

미등록어 후보 구간선정 단계를 거쳐 나온 후보 구간들에 대해 보다 정밀한 처리를 위해 SVM 기반의 추정을 실시한다. 본 연구에서는 LIBSVM 2.0 버전[5]을 사용하였으며 커널 함수는 최적 파라미터를 구할 수 있는 RBF(radial basis function)를 사용하였다. SVM을 위한 자질은 현재 한자를 기반으로 좌우 두개의 한자의 어휘 정보와 태그 정보로 총 10개를 사용하였다.[6]

Ⅲ.문자 발음 변환

중국어 한자는 6765개가 있고 이 한자들은 1600여개의 발음으로 표현할 수 있다. 이중에서 1066여개의 다음자(polyphonic character)가 존재한다. 그러므로 정확하고 자연스러운 TTS를 위하여서는 고 성능의 문자 발음 변환을 요구한다.

그 과정을 그림3과 같다. 단어분할, 품사 태깅과 미등록어 처리가 된 문장이 입력이 되면 단어를 단위로 단어 중의 각 한자가 다음자(polyphone)에 속하는지를 판단한다. 만일 단어를 구성한 각 한자가 모두 다음자가 아니면 이 단어의 한자들은 애매성이 없기 때문에 한자 병음 (중국어 발음 표기) 사전을 기초로 발음 변환시키면 된다. 만일 단어 중 다음자 한자를 포함하면 다시 이 단어가 다음자 단어 병음 사전에 있는 단어인지를 확인하고 발음 변환시키면 된다. 나머지 문제는 단순히 사전 정보로 결정할 수 없는 단어들인데 이러한 단어의 병음은 규칙을 먼저 적용을 하고 만일 규칙으로 추정할 수 없으면 병음 말뭉치에 의하여 얻은 다음자 병음 확률 사전에 의하여 확률이 높은 병음을 선택한다. 좀 더 자세한 내용은 [7]을 참고하기 바란다.

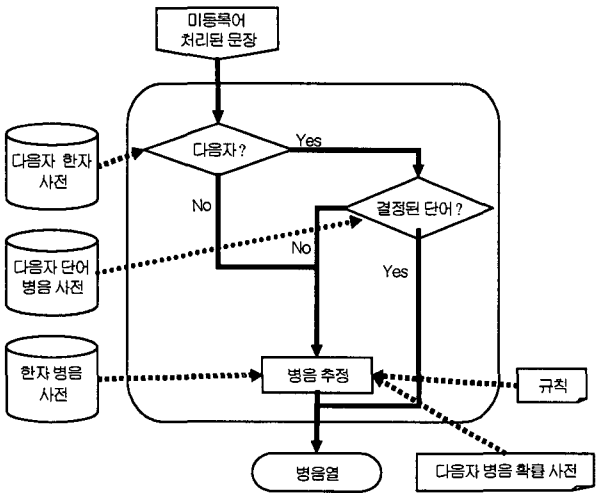

그림 3 중국어 자동 문자 발음 변환

Ⅳ. 언절 경계 추정

언절 경계 추정을 위한 통계적 방법은 일반적으로 많은 양의 훈련 데이터를 필요로 한다. 이러한 데이터의 부족 현상을 해소하기 위하여 본 연구에서는 확률적 추정과 동시에 결정트리로부터 규칙을 추출하여 언절 경계 추정에 사용한다.

확률적 추정에서 사용하는 수식은 아래와 같다.

$$P(b_j/t_1t_2t_3)=\lambda_1\frac{C(t_1t_2b_jt_3)}{\sum_{j=0,1,2}C(t_1t_2b_jt_3)}+\lambda_2\frac{C(t_2b_jt_3)}{\sum_{j=0,1,2}C(t_2b_jt_3)}+\lambda_3\frac{C(t_2b_j)}{\sum_{j=0,1,2}C(t_2b_j)}$$

---(6)

식(6)에서 $\lambda_1+\lambda_2+\lambda_3=1$ 이다.

확률적 추정방법으로 얻은 결과는 결정 트리로부터 얻은 규칙을 사용하여 에러 후처리하여 언절 경계를 수정한다. 언절 경계 후보를 기준으로 좌우 각각 5개 단어의 품사태그와 언절 경계 후보 바로 앞 단어의 위치정보, 단어 길이, 현재 문장 길이 등을 언절 경계 수정 시 자질(feature)로 사용한다.

확률적 추정방법으로 얻은 결과와 정답 언절 경계 태그열을 비교하고 이를 C4.5 기계학습 방법 알고리즘 [8]을 이용하여 언절에러 수정 규칙으로 추출한다.

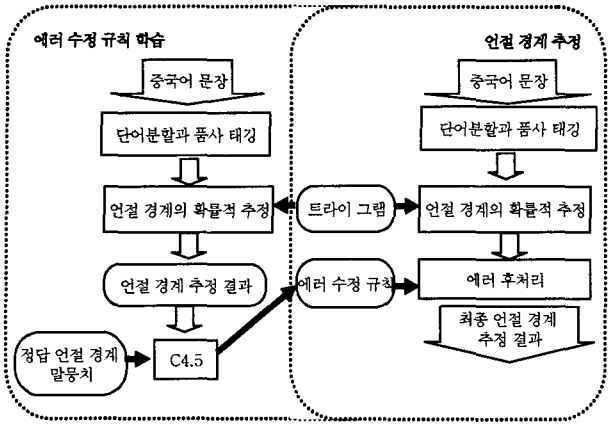

그림 4 언절 경계 추정

V. 합당한 합성 음절 선정

음성합성 DB는 합성단위가 음절이므로 음절을 단위로 구축이 되어 파일로 저장된다. 본 연구는 보이스웨어가 제공한 중국어 합성 DB를 사용한다. 본 음성 DB는 중국표준어로 녹음이 되었으며 16kHz 샘플링주파수와 16bit의 해상도로 이루어졌고 음성 DB에 관한 스크립트가 있다. 음성 DB는 2227개 문장를 포함하고 있으며 6시간 반의 녹음 분량이다.

음절 선정 과정은 환경 벡터가 포함된 병음 열을 받아 최적의 합성단위를 찾아 연쇄시킴으로써 합성음을 생성하는 과정이다. 본 논문에서는 2 단계로 구성된 합성단위 탐색알고리즘을 사용하였다. 첫째는 음절당 가장 가능성이 있는 출현(instance)들을 찾는 것이고(즉 음절 당 N개 후보를 선정함), 둘째는 N개의 후보들의 열로부터 최소의 오차를 갖는 합성단위 열을 찾는 것이다. 다음은 선정된 합성단위 열을 연쇄시켜 합성음을 생성한다. 그림5는 이 전체 과정을 보여 주며 본 연구는 [9][10]의 방법을 따른다.

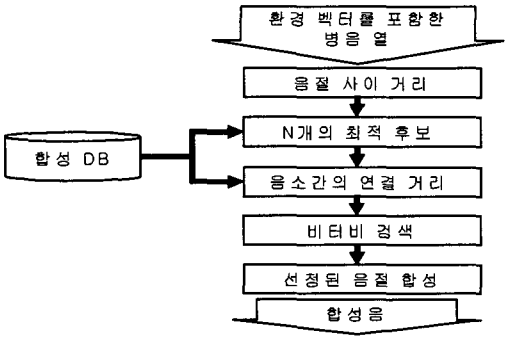

그림5 중국어 음절 선정 및 합성

VI.실험 및 분석

본 논문의 언어처리 부분에서 중국어 단어 분할의 정확도는 98.36%, 품사 태그의 정확도는 95.06%, 미등록어 추정에서 실험을 F-measure로 인명에 대해 87.94%, 지명에 대해 78.43%의 성능을 보였다 .문자 발음 변환에서는 95%의 정확도를 나타낸다. 그러나 현재 변환 룰은 계속 구축하고 있는 과정이기 때문에 그 정확도는 아직 향상의 여지가 많다. 전체적으로 본 연구는 현재 진행 중이며 Baseline성능에서 앞으로 많은 향상의 여지가 있다. 본 연구의 대표적 합성소리는 http://ressell.postech.ac.kr/~project/Research/POSTTSC에서 사전 청취가 가능하다. 보 데모에서 아직 뛰어읽기부분은 적용되지 않은 상태이다.

VII.결론

본 논문에서는 보다 자연스러운 TTS 합성음을 위한 자연어 처리 기법을 중심으로 코퍼스 기반 무제한 중국어 합성 시스템에 대해 살펴보았다. 이를 통해서 품사 정보와 발음, 언절 경계의 관계를 파악하고, 음성합성 단위 선정을 위한 언어 정보를 얻을 수 있었다. 그러나 아직도 여러 가지 문제점이 존재한다. 언어처리 부분에서 말뭉치의 부족 현상이 있었고 음성합성 부분에서 음성DB의 부족함으로 하여 적당한 합성단위를 선정하는데 어려움이 있다. 그리고 음절 레이블링이 HTK를 사용하여 자동으로 이루어졌으므로, 레이블링의 경계가 부정확한 부분들이 발생한다. 이 점은 연쇄합성음의 음질을 상당히 저하시킬 가능성이 있으므로, 향후 합성음 개선을 위하여 레이블링은 전문가에 의한 수정이 필요하다.

감사의 글

본 연구는 2003년 한국과학재단의 특정기초연구 사업인 "무제한 단어 중국어 TTS 시스템을 위한 자연어 처리"과제(년도:2003.9-2006.8)의 일환으로 수행되었습니다. 본 연구에 합성 DB를 제공한 (주)보이스웨어에 감사를 표합니다.

참고문헌

[1] 陶建華, 蔡蓮紅, "語音合成系統的關鍵技術", CTI論壇, 2001

[2]俞士汶, "現代漢語語料庫加工-詞語切分与詞性標注規范与手冊, 北京大學計算語言學研究所", 1999

[3]C. H. Tsai, "MMSEG:A Word Identification System for Mandarin Chinese Text Based on Two

Variants of the Maximum Matching Algorithm" , http://input.cpatch.org/cutphase/mmseg.htm, 1998
[4]V. Vapnik, "The Nature of Statistical Learning Theory", New York: Springer-Verlag, 1995
[5]C. C. Chang and C. J. Lin "LIBSVM: a Library for Support Vector Machines", a guide of beginners ,2003
[6]하주홍, 정옥, 이근배 "중국어 음성합성을 위한지지 지지 벡터 기반 실시간 미등록어 처리", 제 15회 한글 및 한국어 정보처리 학술대회, 2003
[7]Zhang Hong, Yu Jiangsheng, Zhan Weidong, "Disambiguation of Chinese Polyphonic Characters" The First International Workshop, 2001
[8]J. R Quinlan, "C4.5: Programs for Machine Learning Morgan Kaufmann, 1983
[9]Chu Min, Peng Hu, Chang Eric, "A concatenative Mandarin TTS system without prosody model and prosody modification", In SSW4-2000, 2000
[10]吳志勇, 蔡蓮紅, "漢語TTS中基于語音詞的基元選" 全國人机語音通訊技術會議2000, NCMMSC5, 2000