

학습을 통한 발음 변이 규칙 유도 및 적용을 이용한 영역 의존 발음 사전 자동 생성

전 재 훈¹, 정민화²

¹서강대학교 컴퓨터공학과
jhjeon@snu.ac.kr

²서울대학교 언어학과
mchung@snu.ac.kr

Automatic Generation of Domain-Dependent Pronunciation Lexicon with Data-Driven Rules and Rule Adaptation

Je Hun Jeon¹ and Minhwa Chung²

¹Department of Computer Science, Sogang
University
jhjeon@snu.ac.kr

²Department of Linguistics, Seoul National
University
mchung@snu.ac.kr

요 약

본 논문에서는 학습을 이용한 발음 변이 모델링을 통해 특정 영역에 최적화된 발음 사전 자동 생성의 방법을 제시하였다. 학습 방법을 이용한 발음 변이 모델링의 오류를 최소화 하기 위하여 본 논문에서는 발음 변이 규칙의 적용 기법을 도입하였다. 발음 변이 규칙의 적용은 대용량 음성 말뭉치에서 발음 변이 규칙을 유도한 후, 상대적으로 작은 용량의 음성 말뭉치에서 유도한 규칙과의 결합을 통해 이루어 진다. 본 논문에서 사용된 발음 사전은 해당 형태소의 앞 뒤 음소 문맥의 음운 현상을 반영한 발음 사전이며, 학습 방법으로 얻어진 발음 변이 규칙을 대용량 문자 말뭉치에 적용하여 해당 형태소의 발음을 자동 생성하였다. 발음 사전의 평균 발음의 수는 적용된 발음 변이 규칙의 확률 값들의 한계 값 조정에 의해 이루어졌다. 기존의 지식 기반의 발음 사전과 비교 할 때, 본 방법론으로 작성된 발음 사전을 이용한 대화체 음성 인식 실험에서 0.8%의 단어 오류율(WER)이 감소하였다. 또한 사전에 포함된 형태소의 평균 발음 변이 수에서도 기존의 방법론에서 보다 5.6% 적은 수에서 최상의 성능을 보였다.

1. 서론

어떤 단어가 발음될 때 사람에게 따라 또는 주위 상황에 다르게 발음 되는데, 이러한 발음 변이들은 자동 음성 인식의 인식을 저하의 주요한 원인 중 하나로 자리잡고 있다. 특히 대화체 등의 자연스러운 발화에 있어서 발음 변이가 더 빈번 하게 발생 하므로, 발음 변이를 모델링 하는 것은 인식률의 향상에 크게 기여 할 수 있을 것이다.

대부분의 자동 음성 인식기(ASR) 시스템은 이러한 발음 변이를 발음 사전에 반영하여 처리한다. 이러한 발음 변이를 단순히 사전에 추가할 경우 인식률의 향상을 기대할 수 없다. 추가 된 발음 변이들은 사용자들의 다양한 발음 현상을 인식 하게 할 수도 있지만,

그만큼의 탐색 영역 복잡도를 증가시키게 된다. 따라서, 복잡도의 증가를 방지하는 선에서 가능한 많은 발음 변이들을 포함시킬 필요가 있다.

현재까지 발음 변이의 모델링에는 주로 2 가지 방법론이 제시 되어 왔다. 이러한 방법론[1]은 정보의 종류에 따라 학습에 의한 방법과 지식 기반에 의한 방법으로 분류 될 수 있다. 학습에 의한 방법은 음성 신호로부터 발음 변이를 추출 하는 상향식 방법이고[2], 지식 기반 방법은 현재 활용 가능한 언어학적 지식을 이용하는 하향식 방법[3]이다. 이들 방법론들의 각각의 장단점으로 인해 상호간의 우위를 결정할 수 없지만, 지식 기반의 방법론은 대화체의 자연스러운 발화의 변의 정보 표현에 부적합 면이 있다.

본 논문에서는 대화체 음성 인식을 위한 발음 변이 모델링 방법으로 학습에 의한 규칙의 추출 방법을 도입하였다. 학습에 의한 발음 변이 모델링은 일반적으로 아래와 같이 네 단계로 구성 된다.

1. 발화에 맞는 발음 변이 생성
2. 표준 발음과 발음 변이 정렬
3. 정렬된 자료에서 발음 변이 규칙 추출
4. 추출된 규칙 일반화

본 연구는 학습을 통한 발음 규칙의 오류를 최소화하기 위해 적응 기법을 도입하여 추출된 규칙을 일반화하였고, 이 규칙들의 적용 하여 특정 영역의 발음 현상 특징을 반영하는 발음 사전 생성 방법론을 제시하였다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 변의 발음 생성, 규칙 학습 및 적용 방법론을 기술하고, 3장에서는 영역 의존적 발음 사전 생성 방법을 설명하였다. 4, 5장에서는 제안된 방법론을 통한 인식 실험과 결론을 기술하였다.

2. 학습에 의한 발음 변이 규칙 유도

발음 변이 규칙은 표준 발음과 발음 변이 간의 정렬을 통해 학습 될 수 있다[4][5]. 본 논문에서는 표준 발음이 아닌 한글 자모와 발음 변이 간의 정렬을 통해 문자에서 발음 변이 간의 규칙을 추출 하도록 하였고, 아래의 형태와 같이 확률 기반의 발음 변이 규칙 형태를 이용하여 학습하였다.

$$r: L_2L_1 \underline{G} R_1R_2 \rightarrow T \text{ with } P_r$$

이 규칙은 하나의 한글 자모 G가 P_r의 확률을 가지고 음소 T로 변환된다는 것을 의미하고, L은 해당 자모 G의 왼쪽에 위치하는 음소 문맥, R은 오른쪽에 위치하는 음소 문맥을 의미한다. L과 R의 길이는 최대 2개까지 존재 할 수 있고, 순차적으로 생략 될 수도 있다.

규칙의 구조는 아래 [그림1] 과 같이 구성된다. 이 규칙에서 문맥의 길이가 짧은 규칙이 같은 음소 문맥을 포함하는 긴 문맥 규칙의 상위 규칙이 된다.

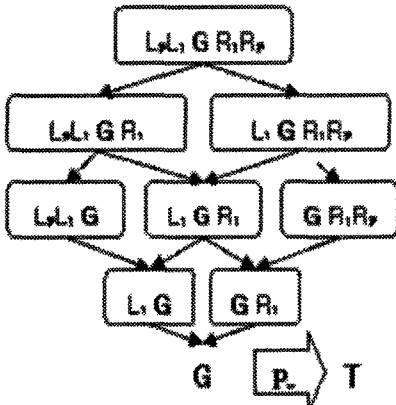


그림 1. 규칙의 계층구조적 표현법

2.1 변이 발음 생성

발음 변이 규칙을 유도 하기 위해 실제 발화에 해당하는 발음 열을 생성하는 것은 규칙 학습을 위한 필수 요소 이다. 발화에 대한 발음 변이 생성 방법은 크게 수작업에 의한 방법과 인식기를 이용해 자동 생성시키는 방법으로 나누어 볼 수 있다. 수작업 전사는 전문 지식이 필요할 뿐만 아니라 작업에 많은 시간과 노력이 필요하다. 또한 최적의 발음 전사를 이용해 규칙을 유도하더라도 해당 인식기에서 사용되는 음향 모델의 특성으로 인하여 WER의 감소를 보장할 수 없다. 이에 반해 인식기의 특성을 반영 할 수 있는 자동 전사 방법은 수작업 전사보다 빠른 시간에 이루어질 수 있고 저비용으로도 가능하다.

자동 전사에 의한 변의 발음 생성은 각 발화에 대해 주어진 발음 사전에서 최적의 변이 발음을 선택하는 강제 인식 방법[4] 또는 음소 인식[5]을 이용한 방법이 있다. 강제 인식 방법은 변이 발음 들이 주어진 발음 사전에 의해 제한 될 수 밖에 없는 단점이 있고, 음소 인식에 의한 방법은 강제 인식 보다는 자유로운 변이 발음을 획득 할 수 있지만 인식기 자체 오류가 빈번히 발생하는 단점이 있다.

본 연구에서는 이러한 한계와 오류를 줄이기 위하여 한글 자모와 음소의 혼합도 매트릭스[7]를 이용한 음소 단위 강제 인식을 수행 하였다. 초기 매트릭스는 각 자모가 모든 음소로 변환 될 수 있다고 정의하였다. 이 혼합도 매트릭스를 이용하여 각 발화의 정답 문장에 대한 [그림2]와 같이 음소 인식 HMM 네트워크를 구성하고, 이 네트워크를 이용해 음소 단위 강제 인식을 수행 하였다. 매트릭스의 각 자모에 대한 혼합도는 음소 인식 결과를 바탕으로 혼합도 확률 수정하고, 그 확률 값이 한계치 이하의 값을 가지는 음소 변이는 제거 하였다. 이러한 방법으로 혼합도 매트릭스의 확률 값이 안정 될 때까지 음소 강제 인식을 반복 수행하여 최종적 발음 변이 열을 생성 하였다.

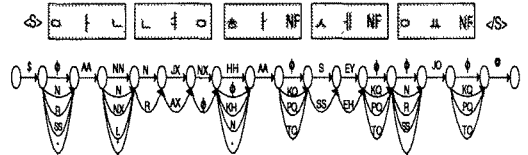


그림2. “안녕하세요”에 대한 자모 단위 강제 인식 네트워크

2.2. 발음 변이 규칙의 학습

음소 강제 인식에서 얻어진 음소 열과 자모 열을 바탕으로 정의된 구조적 규칙을 학습하였다. 학습에 사용된 음소 문맥의 종류는 한글 자음 57개 모음 21

개, 그리고 형태소가 다른 형태소와 결합 되는 형태소 경계 정보 18개로 정의 하였다. 한글 자음은 초성과 종성으로 분리하여 정의하고, 종성이 존재 하지 않은 경우 특수 자모(NF)를 사용 하였다.

기본적인 규칙의 학습 방법은 주어진 음소 문맥에서 해당 자모가 특정 음소로 변환 되는 수를 계산 하는 방법으로 진행 하였다. 여기에 인식기의 음향 모델의 특성을 반영하기 위해 음소 강제 인식의 계산 값(log-likelihood)의 요소를 더하여 아래와 같은 식으로 규칙의 확률 값을 추정 하였다.

$$\hat{P}(L_2L_1 \underline{G} R_1R_2 \rightarrow T) \approx \frac{\sum_{allT} \frac{E(\log P(T))}{\log P(L_2L_1 \underline{G} R_1R_2 \rightarrow T)}}{\text{Count}(L_2L_1 \underline{G} R_1R_2)}$$

위 식에서 $E(\log P(T))$ 는 각 음소 평균 값을 의미하고, $\log P(L_2L_1 \underline{G} R_1R_2 \rightarrow T)$ 는 주어진 음소 문맥에서 음소로 인식 되는 값을 나타낸다. 만약 평균 값보다 큰 값으로 인식 되었다면 그 규칙의 확률 값은 높아 진다는 것을 의미한다.

규칙의 학습 후에 음소 강제 인식에서 빈번하게 발생하는 삽입과 삭제에 대한 오류와 규칙 구조의 문제점인 모음으로만 구성된 음소 문맥에 대한 오류, 마지막으로 잘못된 연음으로 생성 되는 것을 방지 하는 연음 오류 페널티를 적용 하였다. 음소 강제 인식에서 앞 음절에 종성이 없고 뒤 음절의 초성이 경우 또는 격음이 오는 경우, 또는 초성에 음가가 없고 목음 구간 뒤에 위치 할 경우 빈번하게 삽입 된 것으로 인식 된다. 또 유사한 음가의 자음이 연속 되는 경우 하나의 자음으로 인식 되어 탈락 된 것으로 인식 되는 경우도 빈번히 발생한다. 위 두 경우의 값을 줄이기 위하여 삽입 과 삭제 페널티 모델을 적용 하였다. 자음의 음소 변동은 주로 인접한 자음의 영향을 받아 발생 한다. 하지만 정의된 구조적 음소 변동 규칙의 한계점으로 유사한 음소 문맥이 빈번히 발생할 경우 모음 음소 문맥으로 인한 음소 변동 규칙의 값도 높아 지게 된다. 이러한 규칙 구조의 단점을 해소 하기 위해 자음에 대한 음소 변동 규칙에서 모음만으로 이루어진 음소 문맥을 가지는 규칙에 대해 페널티 모델을 적용 하였다. 마지막으로 초성 음소 또는 종성 음소가 연속 해서 발생 하도록 만드는 규칙에 연음 오류에 대한 페널티를 적용 하였다.

기본적으로 추출된 변이 규칙은 그 수가 많고 일부 규칙은 사용 되는 빈도도 적을 뿐만 아니라, 올바르게 발음을 생성하는 것도 있을 수 있다. 따라서 추출된 규칙들의 수를 적절히 제한 할 필요가 있다. 일반적으로 사용 되는 방법은 빈도 수의 한계 값을 이용한 방법[8]이 일반적이다. 본 연구에서는 구조적 규칙의 특성을 반영 할 수 있는 엔트로피 변화도 값[4] 이 한계치 값을 넘지 못하는 규칙을 제거 하였다. 아래의 식은 한계치 값을 정의한 식이다.

$$\delta_{CP} = \frac{\Delta H_{CP}}{1 + e^{-4(C_{app}/C_{th})^{-1}}}$$

C_{app} 는 규칙이 적용 가능한 음소 문맥 조건이고, C_{fir} 는 그 규칙이 적용된 빈도 수 이다. C_{th} 는 각 조건의 절대 빈도수가 일정 수를 넘지 않을 경우 제한하도록 하는 상수 이다. ΦH_{CP} 엔트로피의 차이를 나타내는 식으로 아래와 같다.

$$H(C_{app}(C), C_{fir}(C)) \equiv -C_{fir}(C) \log \frac{C_{fir}(C)}{C_{app}(C)}$$

$$H_{before} = H(C_{app}(C), C_{fir}(C)) + H(C_{app}(P), C_{fir}(P))$$

$$H_{after} = H(C_{app}(C) + C_{app}(P), C_{fir}(C) + C_{fir}(P))$$

$$\Delta H_{cp} = \frac{|H_{before} - H_{after}|}{C_{app}(C) + C_{app}(P)}$$

푸루닝 결과로 초기 학습 규칙의 80% 정도의 규칙 이 성능의 저하 없이 제거 되었다.

2.3. 규칙의 적용 방법

본 연구에 사용된 음성 발음치는 한국어 대화체에 대한 것으로 자세한 소개는 4장에서 할 것이다. 이 대화체 발음치는 1 만 발화와 15 만 형태소로 구성 되어 있다. 이 발음치만으로 학습된 음소 변동 규칙은 안정적인 성능을 보이지 않았다. 따라서 기본 음소 변동 규칙 학습을 위해 대용량 발음치인 88만 발화 180만 형태소의 낭독체 와 방송 뉴스를 사용 하였다.

우선 대용량 발음치에서 학습된 규칙을 기본 규칙으로 사용하고, 대화체 발음치에서 학습된 규칙을 신규 규칙으로 사용하여 가중치를 가지고 결합 하도록 하였다. 가중치를 가진 결합은 주어진 음소 문맥에 대해 두 규칙을 검색하여 기본 규칙에 존재 하지 않는 규칙이 신규 규칙에 존재 할 경우 기본 규칙의 상위 규칙의 확률 값과 가중치 결합을 하는 것이다. 이 방법은 기본 규칙에 존재 하지 않은 음소 변동 규칙, 즉 하나의 자모가 다른 음소로 변화 되는 규칙 이나, 다른 음소 문맥을 가지는 규칙을 추가하는 방법이다. 신규 규칙과의 결합할 때, 기본 규칙의 상위 규칙을 사용하는 이유는 구조적 규칙 구조의 특징인 상위 규칙 이 하위 규칙을 포함 하는 특징 때문이다. 가중치 결합 방법은 아래와 같은 식으로 표현 된다.

$$\hat{P}(LGR \rightarrow T) \approx \begin{cases} \lambda P_{\text{parts of } LGR \rightarrow T} & \\ + (1-\lambda) P_{\text{new}}(LGR \rightarrow T), \text{ where } P_{\text{base}} = 0 & \\ P_{\text{base}}(LGR \rightarrow T) & , \text{ else} \end{cases}$$

이 식에서 $P_{Parents\ of\ base}(LGR \rightarrow T)$ 는 기본 규칙에서 주어진 음소 문맥을 가질 수 있는 상위 규칙 중 가장 큰 확률 값을 가지는 규칙을 의미 한다. 신규 규칙과 기본 규칙에서 같은 음소 문맥을 가지는 규칙이 존재하는 경우 신규 규칙의 확률 값은 무시하게 된다.

3. 발음 사전 자동 생성

본 연구에서는 목적 영역에 최적화된 발음 사전 생성을 위하여 기존의 발음 사전에서 신규 발음을 추가하는 방법을 사용하지 않고 학습된 규칙을 통하여 발음 사전을 다시 생성하였다. 발음 사전의 자동 생성을 위해 학습된 규칙을 언어 모델을 학습이나 적용을 위해 사용할 문자 말뭉치에 적용하여 각 형태소에 대한 확률 값을 가지는 발음 변이를 생성하도록 하고, 그 확률 값의 한계 값에 의해 사전에 포함 하도록 하여 다양한 발음 변이 수를 가지는 발음 사전을 생성하도록 하였다.

음소 변동 규칙의 적용은 주어진 아래 [그림 3]과 같이 주어진 형태소에서 앞 뒤 형태소 정보와 음소 정보를 포함한 음소 문맥 정보로 변환 하고, 이 음소 문맥에 해당 하는 규칙을 적용하여 음소 변동 가능 리스트를 작성한다.

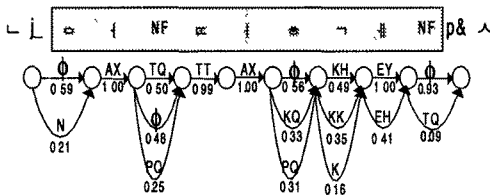


그림3. “어떻게” 에 대한 음소 변동 리스트

이렇게 이루어진 음소 변동 리스트를 탐색하여 여러 개의 발음 변이를 생성 한다. 각 발음 변이는 적용된 규칙의 확률 값들의 곱을 통해 자신의 확률 값을 가지게 되며 [표 1]과 같은 형태로 발음 사전에 저장 된다. 여기에서 새로운 발음 변이를 사전에 추가할 것인가를 결정하는 한계 값은 생성된 발음 변이와 최고의 값을 가지는 발음 변이 값의 비율로 정의 된다.

표 1: 한계 값 0.6 의 “어떻게” 에 발음 변이

Likelihood	L2 ,L1	Variants	R 1 , R2
0.0100	l ,	AX TQ TT AX KH	m - ,
0.0097	j -	EY	σ
0.0073	l ,	AX TT AX KH EY	m - ,
0.0072	j -	AX TQ TT AX KK	σ
	l ,	EY	m - ,
	j -	AX TQ TT AX K EY	σ
	l ,		m - ,
	j -		σ

4. 실험 결과

4.1 실험 말뭉치와 실험 조건

본 실험에서 사용한 음성 말뭉치는 대화체 말뭉치로서 여행 계획에 관한 여행사 직원과 고객의 대화를 가정한 대화를 녹음한 것이다. 이 말뭉치는 2개의 부분으로 이루어져 있는데, 첫 번째는(TP#1)는 50명의 화자의 100개의 대화로 구성 되었고, 두 번째는(TP#2)는 50명의 화자의 150 개의 대화로 구성되었다. TP#2를 녹을 할 때 각 화자에 대해 단순하고 짧은 문장을 사용하도록 의도 하여 대화체에서 빈번한 간투어의 사용을 줄이도록 하였다. 따라서 TP#1이 TP#2에 비해 발화의 길이가 길고, 간투어와 발음 변이가 보다 다양하게 나타난다. 실험을 위해 각 말뭉치를 [표 2]와 학습용과 인식용으로 분리 하였다.

표2: 학습 및 인식 실험용 말뭉치 크기

	학습	인식 실험
TP#1 (10 hours)	90 대화 5,383 발화	10 대화 622 발화
TP#2 (7.5 hours)	115 대화 4,621 발화	10 대화 440 발화

음향 모델의 학습 및 인식은 HTK를 이용하여 이루어 졌다. 견고한 음향 모델의 학습을 위해 정의된 학습용 말뭉치와 25시간 분량의 낭독체 음성 말뭉치를 사용하였다. 학습에 사용된 음소는 45개의 기본 음소와 잡음 모델, 묵음 등 총 48개이다. 인식에 사용된 언어 모델은 백오프 바이그램으로 학습용 말뭉치와 신문 뉴스를 이용해 학습 되었다. 사용된 발음 사전의 표제어의 개수는 2308개 이며, 표제어의 평균 발음 변이의 수는 다양하게 조정 되도록 하였다.

4.2 실험 결과 및 분석

본 실험에서는 3가지 발음 사전을 이용하였고, 각 발음 사전은 다양한 수의 발음 변이를 포함 하도록 하였다. 첫 번째 발음 사전(KBR)은 지식 기반의 규칙[3]으로 만들어진 것이고, 두 번째는 학습에 의한 발음 사전(DBR)으로 대용량 음성 말뭉치에서 학습된 발음 변이 규칙으로 생성된 사전이다. 마지막 세 번째 발음 사전(ADR)은 제안된 규칙 적용 기법을 통해 생성된 발음 사전이다.

[그림 4, 5]에 나타난 것과 같은 발음 사전의 종류와 그 발음 변이 수에 따라 WER이 다르게 나타났다. 최상의 성능은 제안된 방법으로 평균 발음 변이 1.6 개에서 얻을 수 있었다.

기준 실험은 기존에 사용되는 방법인 KBR을 이용한 인식 성능으로 TP#1, TP#2 모두에서 1.7 개의 발음 변이에서 최상의 성능을 보였다. 또 TP#1 은 TP#2 발음치는 보다 다양한 대화체 특징을 포함하여 그 복잡도가 높아 상대적으로 높은 WERs를 보였다. 첫 번째 실험으로 DBR을 생성하여 실험 하였다. 최고 성능에서 TP#1에서는 DBR과 유사한 수준의 WERs를 보인 반면, TP#2에서는 0.38%의 WER의 저하를 가져왔다. 하지만 두 발음치에서 모두 1.6개의 발음 변이수에서 최상의 성능을 보였다. 두 번째 실험에서는 ADR을 생성하여 인식을 수행 하였다. TP#1에서는 DBR을 사용 했을 때와 유사한 결과를 보였지만, TP#2에서는 WER이 줄어 든 것을 볼 수 있고, KBR과 유사한 결과를 얻을 수 있었다. 이 결과에서 대화체 발음치 TP#2의 발음 변이 현상이 낭독체에서 나타나는 현상과 다르게 나타남을 알 수 있다. 하지만 TP#1은 대화체에서 나타나는 발음 변이 현상을 단순히 발음 사전에 발음 변이를 발음 사전에 추가하는 것만으로는 인식률의 향상을 가져 오지 못했다. 마지막 실험으로 ADR을 바탕으로 음향 모델을 재학습하여 인식을 수행 하였다. 이 실험에서 TP#1에서 1.05%에 WER 이 줄어 들었다. TP#2와 달리 TP#1 같은 복잡성을 가진 대화체의 경우에는 음향 모델의 재학습이 인식률 향상의 필수적인 요소였다.

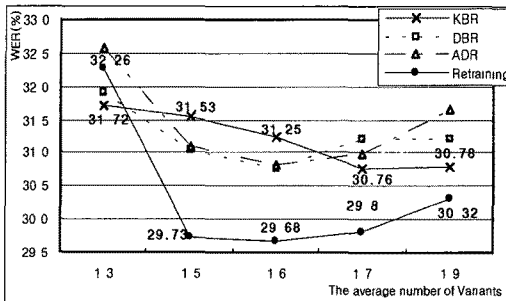


그림 4: TP#1에서의 사전에 따른 WERs

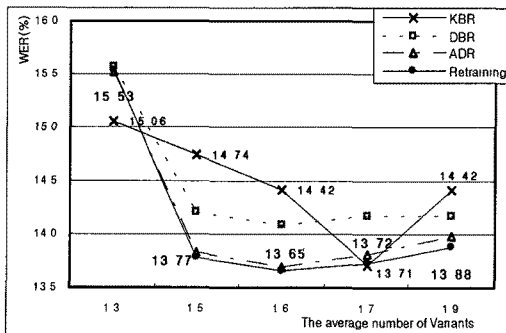


그림5: TP#2에서의 사전에 따른 WERs

실험에서 사용된 규칙 수를 보면 DBR의 경우

42,527개 이고, ADR의 경우는 38,970 개였다. 규칙 적응에서 약 10개의 신규 규칙이 추가 되었지만, 푸루닝 과정에서 많은 수의 규칙이 상위 규칙에 흡수 되었다. [그림 6]에서 나타나듯이 특히 길이가 긴 음소 문맥 규칙 수의 변화가 많았다. 긴 음소 문맥을 가지는 규칙이 푸루닝 되지 않기 위해서는 해당 음소 문맥이 학습 발음치에서 빈번히 발생하여야 하고, 그 변화 양상도 유사해야 함으로 보다 영역 의존 적이라 할 수 있다. 기본 규칙에서 이러한 긴 음소 문맥을 가진 규칙들이 신규 규칙과 적응 된 후 많은 부분이 푸루닝 되었음을 볼 수 있다.

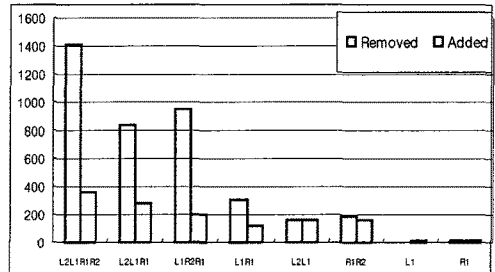


그림 6: 규칙 적응 후의 규칙 수의 변화

최상의 성능을 나타내는 부분에서, 발음 사전의 평균 변이 개수가 KBR에 비해 ADR에서는 95%로 줄어 들었다. ADR의 발음 변이 중 KBR 과 공통된 발음 변이를 가지는 것은 68% 였다. 이러한 발음 변이의 변화가 인식률의 향상을 가져 오는 부분도 있지만 저하를 가져 오는 부분도 있었다. [표 3]과 같이 발음 사전에 따른 WER의 변화가 나타났다. ADR과 KBR을 사용한 인식 실험 결과를 비교 해 볼 때, KBR에서 잘못 인식된 형태소 중 2.98%가 올바르게 인식 되었고, 반면 2.21%의 새로운 단어 인식 오류가 발생하였다. 전체적으로는 0.79%의 인식 오류가 줄어들었다.

표 3: WER(%)의 변화 양상

KBR	ADR	New errors	Reduced errors	WER change
25.53%	24.76%	2.21%	2.98%	- 0.79%

5. 결론

본 논문에서는 영역 의존적인 발음 사전의 자동 생성의 방법으로 학습에 의한 발음 변이 규칙을 제안하였다. 또한 영역 의존적 발음 변이 규칙의 학습을 위해 규칙 적응 방법을 제안하였다. 기존의 지식 기반의 방법론과 비교 했을 때 제안된 학습 기반의 방법이 0.8%의 인식률 향상을 가져 왔다. 또한 사용된 발음 사전의 발음 변이의 개수도 5.6개 줄이는 효과를 가져 왔다.

향후 과제로 본 실험에서 사용된 학습 기반의 규칙 중에서 인식률의 향상에 기여 하는 규칙과 저하를 가져오는 규칙에 대한 분별적 재학습이 이루어 진다면 추가적인 인식률 향상이 기대 된다.

6. 감사의 글

이 연구(논문)는 과학기술부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

7. 참고 논문

- H. Strik, "Pronunciation adaptation at the lexical level", Proc. of the ISCA Tutorial & Research Workshop (ITRW) on Adaptation Methods For Speech Recognition, Sophia-Antipolis, France, pp. 123-131, 2001.
- J. M. Kessens, et al, "A data-driven method for modeling pronunciation variation," The journal of Speech Communication, volume 40, pp. 517-534, 2003.
- K. N. Lee, J. H. Jeon, and M. Chung, "Automatic Generation of Pronunciation Variants for Korean Continuous Speech Recognition," The Journal of the Acoustical Society of Korea, volume 20, pp.35-43, 2001.
- Q. Yang, J. Martens, P. Ghesquiere, and D. Compernelle, "Pronunciation variation modeling for ASR : Large improvements are possible but small ones are likely to achieve," Proc. of Pronunciation Modeling and Lexicon Adaptation for Spoken Language, Colorado USA, 2002.
- I. Amdal, F. Korkmazskiy, and A.C. Surendran, "Joint pronunciation modeling of non-native speakers using data-driven methods," Proc. of ICSLP-2000, Beijing, China, pp. 622-625, 2000.
- J. M. Kessens, H. Strike, "On automatic phonetic transcription quality : lower word error rates do not guarantee better transcriptions," The journal of Computer Speech and Language, volume 18, pp. 123-141, 2004.
- D. Torre, L. Villarrubia, L. Hernandez, and J.M. Elvira, "Automatic Alternative Transcription Generation and Vocabulary Selection for Flexible Word Recognizers," Proc. of ICASSP-97, Munich, Germany, 1997.
- I. Amdal, F. Korkmazskiy and A.C. Surendran, "Data-Driven Pronunciation Modeling for Non-Native Speakers Using Associative Strength between Phones," ASR2000.
- Y.H. Park, and M. Chung, "Style-specific Language Model Adaptation for Korean

Conversational Speech Recognition," Proceedings of the International Conference of Natural Language Processing and Knowledge Engineering, pp. 591-596, 2003.