DP 알고리즘에 의한 발음사전 전처리와 문맥종속 자소별 MLP를 이용한 영어 발음사전 생성기의 개선

Improvements of an English Pronunciation Dictionary Generator Using DP-based Lexicon Pre-processing and Context-dependent Grapheme-to-phoneme MLP

김 회 린*, 문 광 식**, 이 영 직*, 정 재 호**
(Hoi Rin Kim*, Kwang Sik Moon**, Young Jik Lee*, Jae Ho Chung**)

※ 이 연구는 정보통신부 출연 과제의 연구 결과물 입니다.

요 약

본 논문에서는 가변어휘 단어 인식기에 사용하기 위한 개선된 MLP 기반 영어 발음사전 생성기를 제안한다. 가변어휘 단어 인식기는 인식대상 도메인이 수시로 바뀌는 상황에서 현재의 인식 도메인에 의해 결정되는 임의의 한국어 어휘들에 대해 처리할 수 있다. 이 시스템을 영어 단어에 대해서도 처리할 수 있도록 하기 위해서는 미리 정의된 사전에 포함할 수 없는 영어 고유명사와 같은 단어의 발음열을 구할 수 있는 방법이 필요하다. 영어 발음사전 생성기를 구현하기 위하여 본 연구에서는 각 자소를 음소로 변환해 주는 문맥종속 다층 퍼셉트론 구조를 제안한다. 각 자소별 다층 퍼셉트론을 훈련하기 위해서는 표준 발음사전으로부터 각 자소에 대응하는 음소 학습용 데이터를 준비해야 한다. 이를 위해 본 연구에서는 적절한 거리척도를 사용하는 동적 프로그래밍 알고리즘을 사용한다. 훈련 및 평가를 위한 데이터로는 116,191개 영어 단어의 발음사전을 사용하였다. 평가 결과 각각 30~50개의 히든 노드를 가지는 26개 자소별 MLP와 예외 자소 발음사전을 가지고 표준 발음사전에 대하여 72.8%의 단어 정확도를 얻었으며, 이것은 기존의 규칙에 기반한 발음사전 생성의 정확도인 24.0% 보다 매우 우수한 결과임을 보여주었다.

ABSTRACT

In this paper, we propose an improved MLP-based English pronunciation dictionary generator to apply to the variable vocabulary word recognizer. The variable vocabulary word recognizer can process any words specified in Korean word lexicon dynamically determined according to the current recognition task. To extend the ability of the system to task for English words, it is necessary to build a pronunciation dictionary generator to be able to process words not included in a pre-defined lexicon, such as proper nouns. In order to build the English pronunciation dictionary generator, we use context-dependent grapheme-to-phoneme multi-layer perceptron(MLP) architecture for each grapheme. To train each MLP, it is necessary to obtain grapheme-to-phoneme training data from general pronunciation dictionary. To automate the process, we use dynamic programming(DP) algorithm with some distance metrics. For training and testing the grapheme-to-phoneme MLPs, we use general English pronunciation dictionary with about 110 thousand words. With 26 MLPs each having 30 to 50 hidden nodes and the exception grapheme lexicon, we obtained the word accuracy of 72.8% for the 110 thousand words superior to rule-based method showing the word accuracy of 24.0%.

*한국전자통신연구원 멀티모달I/F팀
**인하대학교 전자공학과
접수일자 : 1999년 1월 21일

I. 서 론

일반적으로 음성 인식기를 개발할 때는 우선 그 인식기가 인식할 대상 도메인 및 어휘를 미리 결정하고 이로부터 해당 음성 데이터를 수집한다. 다음으로 수집된 음성 데이터를 적절히 가공하고 학습하여 그 도메인에 포함되어 있는 단어나 음소 모델들을 얻게 된다. 이렇게 하여 얻은 모델들은 미리 정의된 어휘들에 대하여 최적화 되어 해당 어휘들에 대해서는 우수한 인식 성능을 보여주게 된다. 그러나, 만일 우리가 미리 정의하지 않았던 새로운 어휘를 인식대상 어휘로 등록하게 되면 이 시스템의 성능은 급격히 저하되게 마련이다. 새로운 어휘를 인식대상 어휘로 등록할 경우에 높은 성능을 제공하려면 그 새로운 어휘를 포함한 음성 데이터로 모델들을 재훈련하는 방법이 쉽게 접근할 수 있는 방법이나, 이 방법은 음성 데이터를 매번 새로 수집하고 이를 이용하여 재훈련 해야 하는 번거로움이 있다.

이러한 문제를 해결하기 위해서는 임의의 등록 어휘에 대하여 비교적 우수한 성능을 얻을 수 있는 가변어휘 음성 인식기를 구현해야 한다. 가변어휘 음성 인식기를 구현할 때에는 크게 3가지 사항을 고려해야 한다. 첫번째로는 가변어휘 인식기의 음소 모델 훈련에 사용할 음성 데이터 베이스에 관한 것이다. 이 데이터 베이스는 해당 언어의 모든 음소를 포함하고 있어야 하며, 가능한 모든 음소 문맥 환경을 포함하고 있어야 한다[1,2]. 두번째로 고려할 사항은 새로 등록되는 어휘에 대한 발음사전을 실시간으로 자동 생성해야 한다는 것이다. 한국어에 대해서는 음운학적 규칙 및 예외 단어 발음사전으로 비교적 쉽게 구현할 수 있다. 그러나 영어의 경우에는 일반적으로 미리 정의된 대규모 발음사전 및 간단한 규칙으로 이를 구현할 수 있지만, 언어의 특성상 규칙의 정확도가 매우 떨어지는 단점이 있다. 따라서 규칙으로 처리하는 방법을 대체할 수 있는 새로운 방법이 필요하다. 세번째로 고려해야 할 사항은 연속음성을 처리하기 위하여 해당 도메인에 적합한 언어 모델링이 가능하도록 도메인에 무관한 언어 모델을 개발하거나 도메인에 신속히 적응하는 언어 모델링 기법을 개발하여야 한다. 본 연구에서는 두번째 문제에 대한 대안을 제시한다.

당 연구실에서 기 개발한 가변어휘 인식기[3,4,5,6]는 훈련시에 사용한 훈련 데이터와 무관한 임의의 단어를 인식 후보 단어로 등록하기만 하면 등록된 어휘의 음성을 인식해 낼 수 있는 단어 인식기이다. 발음사전을 실시간으로 자동 생성하기 위하여 grapheme-to-phoneme(GTP) 알고리즘을 개발하여 사용하고 있지만, 이는 한국어에 대해서만 가능하고 영어에 대해서는 별도의 알고리즘을 개발해야 한다. 영어에 대한 GTP 알고리즘은 한국어와 달리 몇가지 규칙으로 구현하기에는 너무 많은 예외 발음 방법이 존재하기 때문에, 당 연구실에서는 신경망의 일종인 multi-layer perceptron(MLP)을 이용하여 영어 단어를 구성하고 있는 각 자소의 음소를 추정해내는 알고리즘을 개발하였고, 출력된 각 영어 음소를 이와 가장 유사한 한국어 음소로 매핑해 주는 방식으로 임의의 영어 단어를 한국어식의 발음 기호열로 변환 시켜주는 프로그램을 개발하였다[7,8].

그러나, 이 MLP 기반 영어 발음사전 생성기는 단어 정확도가 116,191단어를 가진 표준 영어 발음사전에서 43.7%에 불과하므로 예외 단어 발음사전을 많이 가지게 되며, 또한 116,191단어를 가진 표준 영어발음사전 이외의 새로운 단어가 입력되면 해당 단어의 음소들을 부정확하게 생성하게 된다. 그 원인을 분석해 보면, 첫번째로 MLP 훈련에 사용하는 자소 대응 음소의 데이터가 부정확한 것들이 다수 포함되어 있었다는 점이고, 두번째로는 모든 자소를 하나의 MLP 네트워크로 훈련하여 정확도가 저하된 점이다. 따라서, 본 논문에서는 이미 개발된 MLP 기반 영어 발음사전 생성기의 성능을 개선하기 위하여 초기의 영어 자소와 음소를 대응시키는 전처리 과정을 동적 프로그래밍 알고리즘으로 개선시키는 방법과 자소별로 입력노드 수가 다른 자소별 MLP를 사용하는 방법을 제안하고 실험을 통하여 그 성능을 평가한다.

II. DP 알고리즘을 이용한 발음사전 전처리

2.1 전처리의 필요성

MLP 기반 영어 발음사전 생성기의 훈련 및 평가를 위해, CMU에서 만든 116,191 단어를 가진 표준 영어발음사전을 이용한다. 이 중 90%의 자소/음소 쌍은 훈련 데이터로 사용하고, 나머지 10%는 평가 데이터로 사용한다. MLP를 훈련시키기 위해서는 각 자소에 대응하는 음소를 찾아내어 훈련 데이터를 만들어야 한다. 그러나, 기존 방식[8]은 모음과 자음사이의 연결관계를 이용한 단순한 규칙을 사용하였기 때문에 성능이 좋지 않으며, 이 데이터로 MLP를 훈련시키면 MLP 기반 영어 발음사전 생성기의 성능은 떨어지게 된다.

따라서, MLP의 훈련 데이터를 만들 때, 단순한 대응 방식이 아닌 DP 알고리즘을 사용하여 각 자소에 대응하는 음소로 매핑시킨 훈련데이터를 그림 1과 같은 과정을 통해 만든다면 MLP 기반 영어 발음사전 생성기의 성능을 크게 향상시킬 수 있다.

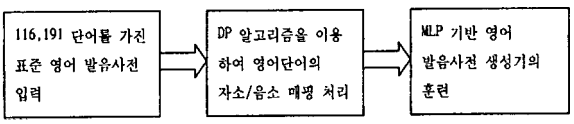

그림 1. DP 알고리즘을 이용한 MLP 기반 영어 발음사전 생성기의 훈련 과정

Fig. 1. Training procedure of MLP-based English pronunciation dictionary generator using DP algorithm.

2.2 DP 알고리즘을 이용한 전처리 및 성능 평가

MLP를 훈련시키기 위해서는 각 자소에 대응하는 음소를 찾아내어 훈련 데이터를 만들어야 한다. 그런데 영어 발음사전내의 각 단어의 자소열과 음소열의 수가 일반적으로 일치하지 않는다. 즉, 어떤 경우에는 자소열과 음소열의 시간적 크기가 선형적으로 쉽게 대응이 되지만 대부분의 경우에는 비선형적으로 대응이 된다. 이러한 대응관계를 얻기 위하여 기존에는 모음과 자음의 연결관계를 단순히 이용해서 비선형적으로 매핑을 시켰으나 오류가 많았다. 이를 개선하기 위하여 자소열과 음소열의 매핑을 DP(dynamic programming) 알고리즘으로 처리하면 보다 정확한 대응 관계를 얻을 수 있게 된다. 본 논문에서 제안한 DP 알고리즘 기반 대응 관계 처리 알고리즘을 상세히 기술하면 아래와 같다. 여기에서 각 자소와 음소 간의 거리 값은 발음 지식에 근거하여 적절히 설정한다.

① 영어 발음사전을 읽은 후, 가로축은 자소를 세로축은 음소를 넣어서 DP 공간을 만든다.
② 첫번째 자소와 첫번째 음소를 대응시킨다.
③ 그림 2와 같이 각 점에서 path의 진행방향은 3방향이며, 3방향 중 거리가 가장 가까운 점으로 path는 진행한다.
④ 현재 O - /O/로 대응이 되어 있고, 다음 자소가 O이고 그것과 다음 음소와의 거리가 2, 3, 4, 5이면 path는 [c]방향으로 진행한다.
⑤ 현재 R - /ER/로 대응이 되어 있고, 다음 자소와 다음 음소와의 거리가 2, 3, 4, 5이면, path는 [c]방향으로 진행한다.
⑥ 현재 W - /HH/로 대응이 되어 있고, 다음 자소와 다음 음소가 각각 H와 /W/라면, path는 [c]방향으로 진행한다. 즉, H와 /W/를 대응시킨다.
⑦ 현재 자소가 H, 그 다음 2개의 자소가 E, D이고, D는 단어의 마지막 자소이며, 다음 음소가 자음이라면, path는 [b]방향으로 진행한다.
⑧ 다음 자소가 J이며, 그 다음 자소가 모음이고, 현재 음소와 다음 음소가 모음이면, path는 [b]방향으로 진행한다.
⑨ [c]방향의 거리가 2나 3, 4, 5, 6이고, [b]와 [c]방향의 거리가 같다면 우선권은 [c]에 있다.
⑩ 다음 자소가 H면, [b]방향이 우선권을 갖는다.
⑪ [b], [c]방향의 거리가 모두 10나 20이고, [a]방향의 거리가 10보다 작으며, 현재 자소가 C가 아니라면, path는 [a]방향으로 진행한다.
⑫ 현재 음소가 /Y/이고 마지막 음소가 아니며, 바로 전 자소와 다음 자소가 자음이고, 다음 자소가 R이 아니면, path는 [a]방향으로 진행한다.
⑬ 동일 음소가 반복될 경우, 해당 자소들과 가장 거리가 짧은 음소만 살리고 나머지는 묵음 표기인 "#"으로 대체하며, 거리가 같을 경우에는 앞의 음소만 살리고, 나머지는 "#"으로 대체한다.
⑭ path가 [a]방향으로 진행하면, 한 개의 자소에 2개 이상의 음소가 대응되는 것이므로, 이럴 경우는 MLP를 훈련시켜도 원래 발음을 복원시킬 수가 없다. 따라서 이런 경우는 예외 단어 발음사전에 등록시키고, MLP 훈련 데이터에서 제외시킨다.

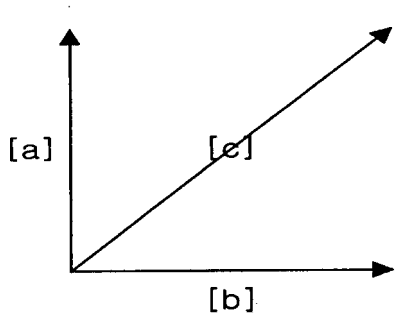

그림 2. DP 매칭 path의 3방향
Fig. 2. Three directions of DP matching.

이상과 같은 과정을 통해 발음사전 내의 각 자소와 음소를 매핑하면 그림 3과 같은 결과를 얻을 수 있다.

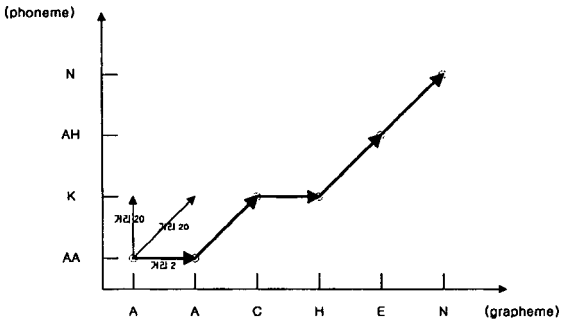

그림 3. DP 알고리즘을 이용한 자소와 음소 사이의 대응관계 설정의 예
Fig. 3. Example of grapheme-to-phoneme mapping by DP algorithm.

116,191 단어를 가진 표준 영어 발음사전 중 201개를 랜덤하게 추출해서, 기존 영어 발음사전 전처리 과정의 성능과 DP 알고리즘을 이용한 영어 발음사전 전처리 과정의 성능을 비교해 보면 자소별 음소 대응의 정확도가 88.56%에서 93.03%로 향상됨을 확인할 수 있었다. 따라서 개선된 훈련 데이터를 이용하면 MLP의 훈련이 보다 더 정확히 이루어질 수 있게 된다.

III. 문맥종속 자소별 MLP를 이용한 영어 발음사전 생성

3.1 제안된 MLP 기반 영어 발음사전 생성기

자소별 MLP는 3개의 층을 가진다. 입력층은 각 자소당 27개 노드로 구성되는 6개 자소 그룹과 추가 1개의 노드로 이루어진 총163개의 입력 노드들을 가지고 있다. 여기서 27개 노드들은 각각 26개의 영어 알파벳 자소와 1개의 묵음에 할당된다. 6개 그룹은 입력단어 스트링의 왼쪽과 오른쪽 자소 정보를 사용할 수 있도록 좌·우 3개 자소를 사용한다. 출력층은 묵음을 포함하여 최대 40개의 음소까지 매핑될 수 있다. 은닉층은 20개에서 60개까지의 노드들을 가질 수 있는데, 은닉층의 노드 수는 주어진 최소 및 최대 히든 노드 수 범위내에서 각 자소별 MLP의 허용 출력 노드 수에 비례하여 결정한다.

위와 같이 자소별 MLP를 구성하여 훈련 및 평가를 수행해 보면, 자소별 MLP에서 'A', 'E', 'I', 'J', 'O', 'U' 의 자소 정확율은 90% 미만을 보여준다. 이를 보완하기 위하여 이 6개 자소에 해당되는 MLP는 기존 MLP 보다 입력층의 노드 수를 더 확장한다. 즉, 입력층은 27개 노드들의 8개 그룹과 추가 1개의 노드로 구성된 총 217개의 노드들을 가지게 한다. 8개 그룹은 입력단어 스트링의 왼쪽과 오른쪽 자소 정보를 사용할 수 있도록 좌·우 4개 그룹씩 확장된다. 이를 다시 평가하여 확장을 해도 성능이 좋아지지 않는 자소는 제외하고 성능이 향상된 것 5개만 MLP 입력을 확장시킨다.

3.2 예외 자소 발음사전 구현

MLP 기반 영어 발음사전 생성기는 예외 자소 발음사전과 예외 단어 발음사전을 함께 사용한다. 예외 자소 발음사전은 자소별 MLP를 보정해 주는 역할을 하는 모듈이다. 자소별 MLP는 어떤 자소 문맥정보의 입력에 대해서 훈련된 특정 음소를 출력으로 내보내게 된다. 그러나 어떤 자소 문맥정보들은 우리가 원하지 않는 틀린 음소들을 출력한다. 따라서 이런 것들을 보정하는 모듈이 필요하다. 표 1를 가지고 예를 들어 설명하면 다음과 같다. 어떤 자소별 MLP가 'A' 라는 자소 문맥정보에 대해서는 'a' 라는 한가지 음소만을 출력한다고 하자. 그런데, 'A' 라는 자소 문맥정보가 각 단어들의 환경에 대해서 'b' 라는 한가지 음소만을 옳은 출력으로 사용한다면 이런 것들은 굳이 MLP를 이용할 필요 없이 단순이 'b' 라는 출력만 내주도록 예외 자소 발음사전에 등록하면 된다. 그러나, 실제 'A' 라는 자소 문맥정보가 각 단어들의 환경에 대해서 'a', 'b', 'c' .... 와 같이 여러가지 음소들을 옳은 출력으로 사용한다면 이것은 보정이 불가능하므로 MLP를 그대로 이용한다. 이와 같이 보정이 가능한 자소 문맥정보들과 이에 대응하는 음소들의 정보를 가지고 있는 모듈이 예외 자소 발음사전이다.

표 1. 예외 자소 발음사전에 등록하는 경우
Table 1. Case of registering to exception grapheme pronunciation dictionary.

경우	입력된 자소 문맥정보	MLP 출력 결과	해당 단어내 실제 음소 종류	비고
1	A	[illegible]	[illegible]	자소별 MLP 이용
2	A	[illegible]	[illegible]	예외 자소 발음 사전 이용
3	A	[illegible]	[illegible] ...	자소별 MLP 이용
4	A	[illegible]	[illegible] ...	자소별 MLP 이용

3.3 예외 단어 발음사전 구현

116,191개 단어를 가진 표준 영어 발음사전에서 예외 자소 발음사전과 자소별 MLP에 의해서 음소열이 정확하게 생성되지 않는 단어들을 등록한 것이 예외 단어 발음사전이다. 그림 4는 한국어 및 영어의 입력 단어가 해당 한국어 음소로 대응되는 과정을 보여준다. 이 그림에서 전처리 모듈은 입력 단어에서 숫자, 기호, 구두점 등 발음과 관계없는 것을 지우는 역할을 한다. 한국어 발음사전 생성기는 한국어의 음운학적 규칙과 예외 발음사전으로 만들어진다. 영어 예외 단어 발음사전은 훈련과정에서 116,191개 단어를 가진 표준 영어 발음사전에 대하여 MLP를 테스트함으로써 얻을 수 있다. 영어 음소를 한국어 음소로 매핑하는 것은 일대일 대응에 의해서 간단히 처리하였다.

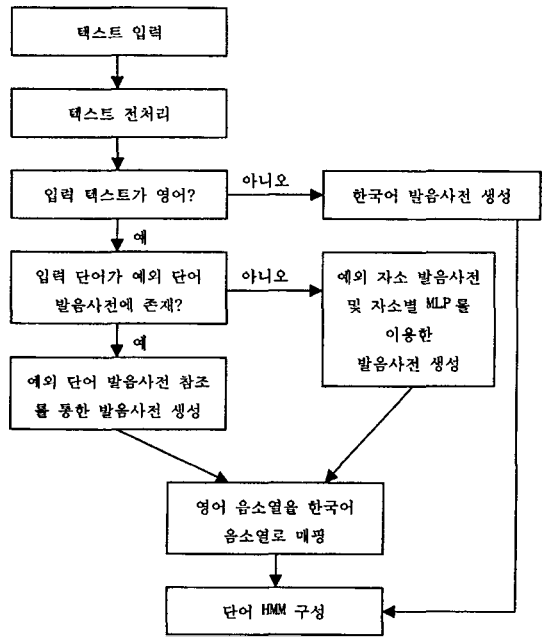

그림 4. 한국어 및 영어 입력 단어에 대해서 발음사전을 출력하는 과정
Fig. 4. Procedure of pronunciation dictionary output for Korean and English input words.

IV. 성능 평가

전처리 과정에서 DP 알고리즘을 적용하면 한 개의 자소에 2개 이상의 음소가 대응되는 경우가 생긴다. 이럴 경우는 MLP를 훈련시켜도 원래 발음을 복원시킬 수가 없다. 따라서 이런 경우는 예외 단어 발음사전에 등록시키고, MLP 훈련 데이터에서 제외시킨다. 이렇게 하여 예외 단어 발음사전에 들어간 단어는 116,191개 단어를 가진 표준 영어 발음사전 중 8,617개(7.4%)의 단어들이다. 따라서 MLP 훈련 데이터 및 평가 데이터를 만드는 데는 107,574개의 단어가 사용된다. MLP 훈련 및 평가 데이터는 107,574개 단어에 DP를 적용해서 생성되는 787,267개의 자소/음소 대응 쌍이며, 훈련 데이터와 평가 데이터의 비율은 이 787,268개의 데이터를 9:1로 나눈 비율이다. MLP 훈련 및 평가는 은닉층의 node 개수를 여러가지로 구분해서 실험하였다. 여기서는 20~30, 40~50, 20~50, 30~50으로 나누어서 실험하였다. 그리고 MLP의 iteration은 100회까지 수행하였다.

4.1 자소 정확도 평가

자소 정확도란 각 단어에 있는 "A"부터 "Z"까지의 자소가 MLP를 거치면서 해당단어의 환경에 맞게 음소로 바뀌게 되는데, 이때 각 자소가 음소로 바뀌는 정확도를 말한다. 자소 정확도를 보면 자소가 모음("A", "E", "I", "O", "U")이거나 반모음의 성격을 가지는 것("J")일 때 자소 정확도가 90% 이하로 떨어지는 것을 볼 수 있다. 그 이유는 자음 자소는 비교적 그 자소에 맞는 한가지 음소만을 가지지만, 모음 자소는 한 자소가 문맥에 따라 여러가지 음소를 가질 수 있으므로 MLP 훈련 시 각 모음 자소가 여러가지 음소로 훈련이 될 때 훈련 정확도가 떨어지기 때문이다. 이를 보완하기 위해서 이 6개의 자소에 대해서는 입력층의 노드 수를 더 확장한다. 즉, 입력층은 27개 노드들의 8개 그룹과 추가 1개의 노드로 구성된 총217개의 노드들을 가지게 한다. 여기에서 8개 그룹은 입력단어 스트링의 왼쪽과 오른쪽 자소 정보를 더 많이 사용할 수 있도록 좌·우 4개 그룹씩 확장된다.

표 2는 각 자소별로 정확도를 나타낸 표이다. 이 표는 MLP를 100회 훈련한 뒤 실행한 결과이며, 6개 자소에 대한 자소 정확도는 볼드체로 진하게 표시하였다. 이 표의 내부에서 화살표 앞은 입력 노드의 그룹 수를 6개 사용한 경우이고, 화살표 뒤는 그룹 수를 8개로 확장한 경우를 표시하였다. 입력 노드를 확장 시켜도 성능 개선이 이루어지지 않는 자소는 입력 노드 수를 확장 시키지 않고 기존대로 27개 노드들의 6개 그룹과 추가의 1개의 노드로 구성된 총 163개의 노드를 사용한다.

표 3은 훈련 데이터, 평가 데이터, 전체 데이터에 대해서 각각 자소 정확도를 보여준다. 이 표에서 알 수 있듯이 훈련 데이터가 평가 데이터보다는 자소 정확도가 다소 높은 것을 알 수 있다.

표 2. 각 자소의 정확도(%)
Table 2. Accuracy(%) of each grapheme.

자소별 MLP의 히든 노드 수	자소		20~30	20~50	30~50	40~50
자소별 MLP로 복원시킬 수 있는 107,574개 단어에 대한 자소 정확도 (%)	a	all data	73.25	71.19→74.27	73.76→75.25	72.92→73.93
	b		99.80	99.77	99.75	99.75
	c		92.02	93.69	93.89	93.47
	d		99.25	99.29	99.33	99.24
	e		81.38→81.96	81.41→82.90	82.26→82.85	81.23→83.41
	f		99.48	99.36	99.29	99.18
	g		93.39	94.19	93.61	93.42
	h		96.62	97.05	96.48	96.73
	i		84.84→85.66	85.50→88.28	85.71→85.23	85.56→86.21
	j		83.47→84.16	82.63→84.16	82.59→83.66	82.30→83.99
	k		99.40	98.84	99.39	99.25
	l		99.57	99.63	99.65	99.67
	m		99.68	99.71	99.70	99.76
	n		98.48	98.73	98.73	98.60
	o		79.19	80.49	79.77→80.58	79.09→79.57
	p		99.44	99.52	99.48	99.36
	q		99.56	99.56	99.56	99.56
	r		93.41	93.39	93.05	93.47
	s		94.47	95.21	95.29	95.47
	t		97.84	97.96	98.17	97.66
	u		88.66	89.39→90.48	89.21	89.87→90.35
	v		99.72	99.73	99.72	99.72
	w		92.62	93.26	93.57	93.40
	x		96.76	95.95	96.36	97.17
	y		93.40	94.13	94.56	93.26
	z		95.56	95.50	95.62	95.52

표 3. 자소별 MLP 만으로 복원시킬 수 있는 107,574개 단어의 자소 정확도(%)
Table 3. Grapheme accuracy(%) of 107,574 words by the MLP method only.

구분	히든 노드 수	20~30	20~50	30~50	40~50
자소 정확도 (%)	Train data	90.80	91.39	91.43	91.31
	Test data	89.82	90.16	90.16	90.03
	All data	90.70	91.27	91.31	91.18

4.2 단어 정확도 평가

표 4는 MLP로 복원시킬 수 있는 107,574개의 단어 중에서 각 단어별로 평가한 자소 정확도의 분포에 따른 단어 정확도를 보여 준다. 예를 들어 자소 정확도 100%인 경우

표 4. 각 단어별 자소 정확도 분포에 따른 단어 정확도(%)
Table 4. Word accuracy(%) according to the grapheme accuracy of each word.

자소별 MLP의 히든 노드 수			20~30	20~50	30~50	40~50
자소별 MLP로 복원시킬 수 있는 107,574개 단어중에서 자소 정확도의 분포에 따른 단어 정확도 (%)	자소 정확도 100%인 경우	all data	53.52	55.62	55.79	55.33
	자소 정확도 90%인 경우		59.15	61.09	61.28	60.91
	자소 정확도 80%인 경우		85.32	86.55	86.52	86.26
	자소 정확도 70%인 경우		92.96	93.53	93.69	93.61
	자소 정확도 60%인 경우		97.63	97.87	97.99	97.93
	자소 정확도 50%인 경우		99.37	99.48	99.47	99.44

의 단어 정확도란, 한 영어 단어가 MLP를 통과할 때 음소열을 생성하게 되는데 이때 그 입력된 영어 단어에 100% 정확히 맞는 음소열을 생성하는 단어의 비율이다. 여기에서는 자소 정확도를 100%에서 50%까지 변화해 가면서 단어 정확도를 조사하였다.

116,191개 단어를 가진 표준 영어 발음사전 중 8,617개의 단어들은 전처리 과정에서 처리 불가능한 단어들이므로 MLP로도 복원이 불가능한 단어들이다. 따라서 전체 116,191개 단어 중에서 자소 정확도 100%인 경우의 단어 정확도는 107,574개 단어의 정확도를 가지고 쉽게 계산해 낼 수 있다. 표 5는 전체 데이터에 대한 단어 정확도를 보여준다. 이 표에서 알 수 있듯이 히든 노드의 수가 30~50일 때가 가장 좋으며, 다음으로 20~50, 40~50, 20~30순이다. 40~50은 실행시간이 다소 오래 걸리며 성능도 별로 좋지 않고, 30~50이나 20~50을 사용하는 것이 성능이나 속도면에서 유리하다.

표 5. 전체 데이터에 대한 자소 정확도 100%인 경우의 단어 정확도(%)

Table 5. Word accuracy(%)in the case that the grapheme accuracy is 100% for all data.

자소별 MLP의 히든 노드 수	116,191개 단어 중에서 자소 정확도 100%인 경우의 단어 정확도(%)
20~30	49.55
20~50	51.50
30~50	51.65
40~50	51.23

107,574개 단어 중에서 히든 노드 수가 30~50일 때 예외 자소 발음사전과 자소별 MLP로 자소 정확도 100%로 복원시킬 수 있는 단어 정확도는 78.60%이다. 따라서 전체 데이터(116,191개 단어)에 대한 단어 정확도는 72.77%가 된다. 여기에서 틀린 단어들은 다중 발음사전을 포함하여 모두 예외 단어 발음사전에 등록하게 된다. 그러므로 예외 자소 및 단어 발음사전이 포함된 MLP기반 영어 발음사전 생성기의 단어 정확도는 표준 영어 발음사전에 대해서는 100%가 되고, 새로운 단어에 대해서는 72.77%의 단어 정확도로 발음사전을 출력하게 된다.

4.3 규칙에 의한 발음사전 생성과의 성능 비교

규칙에 의한 발음사전 생성은 Naval Research Laboratory(NRL)에서 개발한 방법을 적용하였다. NRL에서는 총 42개의 음소를 사용하는 반면, MLP 기반 영어 발음사전 생성에서 사용한 CMU의 발음 표기는 총 39개의 음소를 사용한다. 또한, "YU", "AX", "WH", "j"는 NRL에서만 사용하는 음소이고, "JH"는 CMU에서만 사용하는 음소이다. 따라서 정확한 성능 비교는 어렵기 때문에 CMU에서 정한 116,191개 단어들을 NRL에서 만든 규칙에 넣어서 생성된 각 단어의 음소들을 CMU의 해당 단어의 음소들과 동일한지 단순 비교 하였다. 이때 해당 단어의 음소에 NRL에서만 쓰이는 4개의 음소("YU", "AX", "WH", "j")가 들어 있거나, CMU에서만 쓰이는 1개의 음소("JH")가 들어 있을 경우, 그 단어는 비교 대상에서 제외시켰다.

NRL에서 만든 규칙에 116,191개의 단어들을 입력한 결과, 이중 22,656개 단어가 제외된 93,535개의 단어가 비교 대상이었고, 비교된 93,535개 단어 중 22,452개의 단어만이 CMU의 영어 발음사전과 음소들이 정확히 일치하였다. 즉, 93,535개 단어 중에서 자소 정확도 100%인 경우의 단어 정확도는 24%로 매우 낮다. 한편, 최소 30개, 최대 50개의 히든 노드를 가지는 자소별 MLP의 성능은 51.65%이며, 예외 자소 발음사전을 함께 사용한 자소별 MLP의 성능은 72.77%이다. 따라서 DP 알고리즘을 사용한 MLP 기반 영어 발음사전 생성기의 성능이 월등히 좋음을 알 수 있다.

V. 결 론

본 논문에서는 MLP 기반 영어 발음사전 생성기의 성능을 개선시키기 위하여 기존의 영어 자소와 음소를 대응시키는 전처리 과정을 DP 알고리즘으로 개선시키는 방법과 자소별로 입력노드 수가 다른 자소별 MLP로 훈련하는 방법을 제안하였다. 또한 예외 단어 발음사전의 크기를 줄일 수 있도록 예외 자소 발음사전을 사용하는 방법도 제안하였다.

제안된 영어 발음사전 전처리 과정의 성능은 기존 영어 발음사전 전처리 과정의 성능보다 자소/음소 대응 정확도가 88.56%에서 93.03%로 개선되었으며, 하나의 MLP network을 자소별로 입력노드 수가 다른 자소별 MLP로 나누어 훈련시킴으로써 단어 정확도는 히든 노드 수 40~50개에서 7.56%(43.67%에서 51.23%) 개선됨을 볼 수 있었다. 또한 히든 노드의 수를 30~50개로 사용하면 단어 정확도는 약 8%(43.67%에서 51.65%)까지 향상되는 것을 확인할 수 있었다. 또한, 훈련이 잘 안된 자소들은 예외 자소 발음사전을 사용해서 보정해 주면 29.10%(43.67%에서 72.77%)까지 향상되었다.

제안된 MLP 기반 영어 발음사전 생성기를 사용하면 예외 단어 발음사전의 단어 수를 대폭 줄일 수 있을 뿐만 아니라, 새로운 단어가 입력 되더라도 해당 단어의 발음열을 보다 정확히 온라인으로 생성할 수 있다. 이 발음사전 생성기를 사용한 가변어휘 단어 인식기는 음성명령 웹 브라우저[9]에서 영어를 포함한 각종 링크 명령어를 음성으로 제어하는데 사용할 수 있다.

참 고 문 헌

1. Yeonja Lim and Youngjik Lee, "Implementation of the POW (Phonetically Optimized Words) algorithm for speech database," *Proc. of ICASSP*, pp.89-91, 1995.
2. 서영주, 성철재, 이정철, 한민수, 이영직, "음성학적 지식에 기반한 한국어 변이음 집단화 수형도의 구현," *제13회 음성통신 및 신호처리 워크샵(KSCSP' 96)* 논문집, 13권, 1호, pp.344-347, 1996.
3. 김회린, 이항섭, "POW 3848 단어 인식기 구현 및 어휘 독립 실험," *제13회 음성통신 및 신호처리 워크샵 (KSCSP'96)* 논문집, 13권, 1호, pp.127-130, 1996.
4. 김회린, 이항섭, "음성학적 지식 기반 변이음 모델을 이용한 가변 어휘 단어 인식기," *한국음향학회지*, 제16권, 제2호, pp.31-35, 1997.
5. Oh-Wook Kwon, Chong-Kwan Un, Hoi-Rin Kim, "Performance of vocabulary-independent speech recognizers with speaker adaptation," *Jour. of ASK*, Vol.16, No.1E, pp.57-63, 1997.
6. 김회린, 이항섭, 권오욱, "음성인식에서 훈련 및 인식 과정에 사용되는 대상 어휘의 차이에 대한 음향모델의 성능 평가," *한국음향학회지*, 제17권, 제7호, pp.22-27, 1998.
7. T. J. Sejnowski and C. R. Rosenberg, "Parallel networks that learn to pronounce English text," *Complex Systems*, 1:145-168, 1987.
8. Hoi-Rin Kim, Youngjik Lee, and Jung-Chul Lee, "MLP-based English pronunciation dictionary generator for applying to variable vocabulary word recognizer," *Proc. of ICSP' 97*, pp.465-468, 1997.
9. Hang-Seop Lee and Hoi-Rin Kim, "Speech Web browser using variable vocabulary word recognition," *Proc. of AVIOS' 97*, San Jose, 1997.

▲김 회 린(Hoi-Rim Kim)
한국음향학회지 제16권 제2호 참조
한국전자통신연구원 멀티모달I/F팀 선임연구원

▲문 광 식(Kwang-Sik Moon)

1998년2월 : 인하대학교 전자공학과 (학사)
1998년3월-현재 : 인하대학교 전자공학과 석사과정 재학중
※주관심분야 : 음성인식, 화자인식, 디지털 신호처리

▲이 영 직(Youngjik Lee)

1979년2월 : 서울대학교 전자공학과 (학사)
1981년2월 : 한국과학기술원 산업전자공학과(석사)
1989년1월 : Polytechnic University 전기 및 전산과(박사)
1981년-1984년 : 삼성전자주식회사 컴퓨터개발실
1989년1월-현재 : 한국전자통신연구원 멀티모달I/F팀장, 책임연구원
※주관심분야 : 음성인식, 음성합성, 자동통역, 신경회로망, 패턴인식, 멀티모달I/F

▲정 재 호(jae-Ho Chung)
한국음향학회지 제16권 제3호 참조
인하대학교 전자공학과 부교수