김일성종합대학학보
(자연과학)
주체106(2017)년 제63권 제9호

JOURNAL OF **KIM IL SUNG** UNIVERSITY
(NATURAL SCIENCE)
Vol. 63 No. 9 JUCHE106(2017).

# 최대엔트로피원리에 기초한 대규모재귀신경망 언어모형학습에 대한 연구

**리 현 순**

최근 음성인식체계들에서는 신경망을 언어모형학습에 리용하여 성능을 개선하고있다.[1] 특히 재귀신경망언어모형(RNNLM)은 단어리력을 제한하지 않고 임의의 길이를 가지는 단어문맥을 리력으로 하여 련속공간에서 파라메터추정을 진행하는것으로 신경망언어모형의 성능을 높이고있다.

그러나 RNNLM은 대규모학습자료를 리용하는 경우 숨은층의 크기를 증가시키는데 따라 계산량이 늘어나고 학습속도가 떨어져 대어휘련속음성인식체계들에는 아직까지 도입되지 못하고있다.[2, 3]

이로부터 론문에서는 최대엔트로피원리에 기초한 대규모재귀신경망언어모형학습방법을 제안한다.

## 1. 최대엔트로피원리에 기초한 대규모재귀신경망언어모형구축방법

### 1) 최대엔트로피원리에 기초한 재귀신경망

최대엔트로피방법은 자연언어에 존재하는 서로 다른 형태의 특징들을 언어모형에 통합하기 위한 방법이다.

단어리력 $h$가 주어진 조건에서 단어 $w$에 대한 최대엔트로피모형은 다음과 같다.

$$P_m(w|h)=\frac{\exp\sum_i \lambda_i f_i(h,\ w)}{\sum_{\tilde{w}}\exp\sum_i \lambda_i f_i(h,\ \tilde{w})} \tag{1}$$

여기서 $f_i(h,\ w)$는 $(h,\ w)$에 대한 $i$의 특징함수, $\lambda_i$는 $i$의 특징무게, $m$은 특징무게들의 모임이다.

매 특징 $i$에 대한 특징함수 $f_i(h,\ w)$는 다음의 성질을 가진다.

$$f_i(h,\ w)=\begin{cases}1, & \text{특징 } i\text{가 } (h,\ w)\text{에 존재하는 경우}\\ 0, & \text{기타 경우}\end{cases}$$

식 (1)에 따르면 최대엔트로피모형의 특징무게모임은 숨은층이 없는 신경망으로 학습할수 있다. 즉 최대엔트로피모형의 특징무게모임은 신경망모형의 측면에서 볼 때 입력층과 출력층들이 직접 련결되여있는 경우의 무게행렬로 볼수 있다.

론문에서는 재귀신경망의 입력층과 출력층들을 직접 련결하고 숨은층이 없는 재귀신경망을 리용하여 특징무게파라메터들을 학습한다.

2) 최대엔트로피원리에 기초한 재귀신경망언어모형의 학습

N-그람특징을 리용하는 최대엔트로피모형의 파라메터수는 $V^N$ 이다.

최대엔트로피모형구축에서 문제는 특징들의 차수가 높아지고 어휘규모가 커지는데 따라 추정해야 할 파라메터개수가 실현불가능하게 늘어나는것이다.

론문에서는 추정해야 할 거대한 무게행렬의 기억량을 줄이기 위하여 매 N-그람리력들을 하나의 값으로 넘기는 하쉬함수를 리용하였다.

3-그람리력들을 크기 $S$를 가지는 배렬로 넘기는 하쉬함수는 다음과 같다.

$$g(w(t-2),\ w(t-1)) = (w(t-2)\times P_1 \times P_2 + w(t-1)\times P_1)\%S \tag{2}$$

여기서 $P_1$, $P_2$는 임의의 큰 씨수들이며 하쉬배렬의 크기 $S$는 최대엔트로피모형의 파라메터수 $V^N$ 보다 훨씬 작게 설정한다.

이때 식 (1)은 다음과 같이 표현할수 있다.

$$P(w\,|\,h) = \frac{\exp\sum_{i=1}^{N}\lambda_i f_i(g(h),\ w)}{\sum_{w}\exp\sum_{i=1}^{N}\lambda_i f_i(g(h),\ w)} \tag{3}$$

하쉬함수를 리용하여 리력들을 배렬로 넘기는 경우 서로 다른 리력들이 같은 하쉬값을 가지게 되는 경우가 있다. 이 경우 하쉬크기를 증가시켜 충돌확률을 감소시킨다. 리력들이 같은 위치로 넘어가는 경우 작은 크기의 하쉬배렬을 리용하는 모형은 가지자르기한 모형과 류사하게 동작한다.

최대엔트로피모형의 학습은 통계적그라디엔트하강법을 리용하는 재귀신경망모형의 학습과 같은 방법으로 진행된다.

최대엔트로피모형추정을 위한 한 주기 RNNLM학습은 다음과 같이 진행한다.

① 시각 $t$를 0으로 설정하고 특징무게행렬을 포함한 무게행렬들을 초기화한다.

② $t$를 1만큼 증가시킨다.

③ 입력층에 현재 단어 $w_t$를 표현하는 단어벡토르 $\boldsymbol{w}(t)$를 입력한다.

④ 입력된 문맥의 N-그람특징들에 대한 하쉬배렬을 식 (2)에 따라 다음과 같이 만든다.

우선 N-그람특징($N$=3인 경우 3개의 하쉬배렬이 필요하다.)들에 대한 하쉬배렬을 창조하고 0으로 초기화한다. 다음 매 하쉬배렬값을 계산하는데 해당 N-그람특징에 대한 하쉬배렬값이 0일 때에는 N-그람이 존재하지 않는다는것을 나타내며 0이 아닐 때에는 특징무게행렬에서 해당한 N-그람특징무게의 번호를 나타낸다.

⑤ 재귀신경망의 앞방향학습을 진행한다.

⑥ 최대엔트로피모형학습을 위해 출력층의 매 요소를 순환하면서 N-그람특징을 가지는 요소들에 대하여 특징들에 대응하는 특징무게값을 출력층의 값에 더해준다.

⑦ 출력층에서 오차 $\boldsymbol{e}(t)$의 그라디엔트를 교차엔트로피척도를 리용하여 계산한다.

⑧ 오차를 역전파하고 특징무게행렬을 포함하여 무게행렬들을 다음과 같이 변화시킨다.

$$v_{jk}(t+1) = v_{jk}(t) + s_j(t)e_{ok}(t)\alpha \tag{4}$$

여기서 $\alpha$는 학습률, $j$는 숨은층의 크기, $k$는 출력층의 크기, $s_j(t)$는 숨은층에서 $j$번째 신경세포의 출력이다. 그리고 $e_{ok}(t)$는 출력층에서 $k$번째 신경세포의 오차그라디엔트로서 $s(t)$로부터 이전 시각단계 $s(t-1)$의 숨은층에로 재귀적으로 전파된다. 즉

$$\boldsymbol{e}_h(t-\tau-1)=d_h(\boldsymbol{e}_h(t-\tau)^T\boldsymbol{W},\ \ t-\tau-1). \tag{5}$$

여기로부터 오차벡토르는 성분별로 적용되는 다음과 같은 함수를 리용하여 얻는다.

$$d_{hj}(x,\ t)=xs_j(t)(1-s_j(t)) \tag{6}$$

식 (6)에서 재귀무게행렬 $\boldsymbol{W}$는 다음과 같이 갱신된다.

$$w_{lj}(t+1)=w_{lj}(t)+\sum s_l(t-z-1)e_{hj}(t-z)\alpha-w_{lj}(t)\beta \tag{7}$$

특징무게행렬은 출력층의 매 요소에 대하여 1～N－1그람특징이 있는 경우에 그 특징에 대응하는 특징무게값을 갱신한다.

⑨ 모든 학습표본들이 처리되지 못했으면 단계 ②로 간다.

## 2. 성 능 평 가

조선어음성인식프로그람《룡남산》을 리용하여 최대엔트로피재귀신경망언어모형의 성능평가실험을 진행하였다.

실험에서는 Core i3(2GB, 3GHz)급콤퓨터를 리용하였다.

먼저 숨은층의 크기에 따르는 성능평가실험을 진행하였다.(표 1)

언어모형학습자료로서 10M단어들로 구성되는《로동신문》본문을 리용하였으며 생성된 어휘크기는 20K이다. 표 1에서 RNNLM1은 재귀신경망언어모형을, RNNLM2는 최대엔트로피재귀신경망언어모형을 나타낸다.

표 1에서 보는것처럼 RNNLM1의 성능은 숨은층의 크기를 증가시키는데 따라 개선되지만 RNNLM2의 성능은 숨은층의 크기에 의존하지 않는다. 따라서 RNNLM2에서는 작은 크기의 숨은층을 가지고도 성능이 높은 신경망언어모형을 학습할수 있다.

다음 최대엔트로피대규모재귀신경망언어모형을 구축하고 조선어련속음성인식체계《룡남산》에 적용하여 단어오유률과 분기수평가를 진행하였다.(표 2)

표 1. 숨은층의 크기에 따르는 분기수

| 숨은층 크기/개 | 50 | 100 | 150 | 200 | 300 | 400 |
|---|---|---|---|---|---|---|
| RNNLM1/개 | 146 | 141 | 137 | 132 | 126 | 123 |
| RNNLM2/개 | 124 | 119 | 118 | 118 | 117 | 117 |

표 2. RNNLM2의 단어오유률(WER)과 분기수(PPL)평가

| 언어모형 | WER/% | PPL/개 |
|---|---|---|
| 저차원되돌이3－그람 | 3.9 | 172 |
| RNNLM2 | 2.0 | 156 |

언어모형학습자료로서 40M단어들로 구성되는《로동신문》본문을 리용하였으며 생성된 어휘크기는 60K이다. 숨은층의 크기를 100으로 고정하였다.

실험결과들은 단어오유률과 분기수측면에서 저차원되돌이모형에 비하여 우월하다는것을 보여준다.

## 맺 는 말

최대엔트로피재귀신경망언어모형구축방법을 제안하고 음성인식체계에 적용할 때의 효과성을 검증하였다. 최대엔트로피원리에 기초한 재귀신경망에서는 숨은층의 크기를 증가시키지 않고도 대규모학습자료에 대하여 지금까지 많이 리용되여오던 저차원되돌이 3－그람 언어모형보다 성능이 높은 언어모형을 학습할수 있다.

## 참 고 문 헌

[1] Y. Bengio et al.; Journal of Machine Learning Research, 3, 1137, 2003.

[2] H. Schwenk; Speech & Language, 21, 3, 492, 2007.

[3] I. Allauzen et al.; In Proc. of ICASSP, 11, 5524, 2011.

주체106(2017)년 5월 5일 원고접수

## The Study of Constructing Large Scale Recurrent Neural Network Language Model based on Maximum Entropy Principle

*Ri Hyon Sun*

We proposed the method of constructing maximum entropy recurrent neural network language model and estimated the effectiveness of application to ASR.

We verified that the maximum entropy recurrent neural network can study language model with high accuracy on large scale corpus than traditional back-off 3-gram language model without increasing the size of the hidden layer.

Key words: maximum entropy, language model, recurrent neural network