

출력충분해를 리용한 대규모재귀신경망언어 모형의 학습속도개선에 대한 연구

리 현 순

최근 음성인식체계들에서는 지금까지 많이 사용되어오던 통계적 N-그램언어모형이 가지는 제한성으로부터 신경망을 언어모형학습에 리용하여 성능을 개선하고있다.

신경망언어모형에서는 단어리력들을 련속공간으로 넘기고 련속공간에서 언어모형과 라메터들의 확률을 추정하는것으로 하여 모형의 정확도가 높아지게 된다.[1]

재귀신경망언어모형(RNNLM)은 단어리력을 제한하지 않고 임의의 길이를 가지는 단어문맥을 리력으로 사용하여 련속공간에서 언어모형확률을 추정하는것으로 하여 신경망언어모형의 성능을 높이고있다. 현재 어휘규모가 크지 않은 조선어련속음성인식체계에서 RNNLM을 리용하여 인식률을 개선하고있다.[3] 그러나 RNNLM은 계산량이 방대한것으로 하여 대어휘련속음성인식체계들에는 아직까지 도입되지 못하고있다.[2]

이로부터 논문에서는 대규모언어모형학습에 재귀신경망을 리용하는데서 나서는 학습속도개선방법을 제안한다.

1. 출력충분해를 리용한 재귀신경망언어모형구축방법

1) 재귀신경망의 출력충분해

클래스나무에서 매 단어는 하나의 클래스와 그와 련관된 부분클래스들에 속한다.

w_i 가 문장에서 i 번째 단어라고 하면 련 $c_{1:D}(w_i) = c_1, \dots, c_D$ 는 클래스나무에서 단어 w_i 에 대한 경로를 반영한다. 여기서 D 는 나무의 깊이, $c_d(w_i)$ 는 w_i 에 대응하는 클래스 또는 부분클래스, $c_D(w_i)$ 는 w_i 와 련관된 잎(단어)을 표현한다.

단어리력 h 가 주어진 조건에서 w_i 의 N-그램확률은 다음과 같이 추정된다.

$$P(w_i | h) = P(c_1(w_i) | h) \prod_{d=2}^D P(c_d(w_i) | h, c_{1:d-1}) \quad (1)$$

여기에 기초하여 재귀신경망출력층의 어휘들을 무리짓기하여 클래스나무를 생성하는 방법으로 출력층을 분해할수 있는데 논문에서는 재귀신경망의 출력층을 3개의 부분층으로 분해한다.

첫째 출력층은 단축목록(shortlist)단어들과 OOS단어(단축목록밖의 단어)들을 위한 상위클래스들(가장 일반적인 클래스들)에 대한 분포를 추정한다.

단축목록에 있는 단어들은 매개가 부분클래스가 없이 자기자체의 클래스를 표현한다. 이 경우는 나무깊이 D 가 1과 같다.

상위클래스들은 OOS단어들을 처리하기 위한 나무의 뿌리로 되며 매 층이 softmax함수를 가지는 여러개의 부분클래스층들을 포함한다. 이 부분클래스층들이 두번째 층을 형

성한다.

세번째 층인 단어층들은 OOS단어들에 대한 단어확률을 추정한다.

결국 단어리력 h 가 주어진 조건에서 w_i 의 N-그램확률은 식 (1)에 따라 다음과 같이 추정된다.

웃층은 상위클래스들에 대한 분포 $P(c_1(w_i)|h)$ 를 계산하는데 리용되고 다음 부분클래스층들은 $1 < d < D$ 에 대하여 부분클래스확률 $P(c_d(w_i)|h, c_{1:d-1})$ 을 계산하는데 리용된다. 마지막으로 OOS단어확률 $p(C_D(w_i)|h, c_{1:D-1})$ 은 단어층에서 계산된다.

출력층분해를 리용한 재귀신경망은 입력층 x 와 숨은층 s , 출력층 y 로 이루어진다.

입력층과 숨은층사이에는 무게행렬 U 와 W , 숨은층과 출력층사이에는 무게행렬 V 를 가진다. 그리고 시각 t 에서 망의 입력은 $x(t)$ 이고 출력은 $y(t)$ 이며 $s(t)$ 는 숨은층의 상태이다. $x(t)$ 는 현재 단어를 표현하는 벡토르 $w(t)$ 와 시각 $t-1$ 에서 숨은층의 출력 $s(t-1)$ 로부터 얻어진다.

입력층, 출력층들은 각각 다음과 같이 계산된다.

$$s_j(t) = f\left(\sum_i w_i(t)u_{ji} + \sum_l s_l(t-1)w_{jl}\right) \quad (2)$$

$$y_k(t) = g\left(\sum_j s_j(t)v_{kj}\right) \quad (3)$$

여기서 $f(z)$ 는 시그모이드활성함수로서

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

로 표시되며 $g(z)$ 는 softmax함수로서

$$g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}} \quad (5)$$

로 표시된다.

2) 출력층분해를 리용한 재귀신경망언어모형의 학습

출력층분해를 리용한 재귀신경망언어모형의 학습은 출력층을 클래스나무로 분할하는 단어무리짓기단계와 전체 어휘를 가진 신경망파라미터학습단계로 나누어 진행한다.

알고리즘은 다음과 같다.

단계 1: 클래스나무생성을 위한 단어무리짓기

① 단어특징 학습

신경망의 출력으로서 단축목록만을 가지는 RNN을 학습하여 무게행렬 R 를 추정한다.

② 사영공간의 차원축소

행렬 R 에 대하여 표준주성분분석(PCA)방법을 적용한다.

③ 클래스나무생성

K -평균알고리즘을 리용하여 단축목록에 없는 단어들을 가지고 앞선단계에서 생성된 단어특징벡토르에 기초하여 하강형단어무리짓기를 진행한다. 클래스에 속하는 단어들의 개수가 실험적으로 설정된 턱값 W 보다 큰 경우에는 부분클래스들로 가른다. W 개이상의

단어들을 포함하는 매 클래스는 $|\sqrt{W}+1|$ 개의 부분클래스들로 나누어진다.

단계 2: 전체 어휘를 가진 재귀신경망학습

출력으로서 클래스나무구조를 가지는 전체 어휘 RNNLM을 BPTT알고리즘을 리용하는 통계적그라디언트하강법에 의해 학습한다.

한 주기 RNNLM학습은 다음과 같이 진행한다.

- ① 시각 t 를 0으로 설정하고 숨은층에서 신경세포들의 상태 $s(t)$ 를 초기화한다.
- ② t 를 1만큼 증가시킨다.
- ③ 입력층에 현재 단어 w_t 를 표현하는 단어벡터 $w(t)$ 를 입력한다.
- ④ 숨은층의 상태 $s(t-1)$ 을 입력층으로 복사한다.
- ⑤ 식 (2), (3)에 따라 $s(t)$, $y(t)$ 를 계산한다.
- ⑥ 출력층에서 오차 $e(t)$ 의 그라디언트를 교차엔트로피척도를 리용하여 계산한다.
- ⑦ 오차를 역전파하고 무게들을 대응시켜 다음과 같이 변화시킨다.

$$v_{jk}(t+1) = v_{jk}(t) + s_j(t)e_{ok}(t)\alpha \quad (6)$$

여기서 α 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 오차그라디언트로서 이것은 숨은층 $s(t)$ 로부터 이전 시각단계 $s(t-1)$ 의 숨은층으로 재귀적으로 전파된다. 즉

$$e_h(t-\tau-1) = d_h(e_h(t-\tau)^T W, t-\tau-1). \quad (7)$$

여기서 τ 는 시간지연을 나타내며 오차벡터는 성분별로 적용되는 함수 $d_h()$ 를 리용하여 얻어진다.

$$d_{hj}(x, t) = xs_j(t)(1-s_j(t)) \quad (8)$$

입력층 $w(t)$ 와 숨은층 $s(t)$ 사이의 무게행렬 U 는 다음과 같이 갱신된다.

$$u_{ij}(t+1) = u_{ij}(t) + \sum_{z=0}^T w_i(t-z)e_{hj}(t-z)\alpha - u_{ij}(t)\beta \quad (9)$$

여기서 T 는 시각의 개수이다.

한편 재귀무게행렬 W 는 다음과 같이 갱신된다.

$$w_{ij}(t+1) = w_{ij}(t) + \sum s_l(t-z-1)e_{hj}(t-z)\alpha - w_{ij}(t)\beta \quad (10)$$

- ⑧ 모든 학습표본들이 처리되지 못했으면 ②로 간다.

2. 성능평가

조선어음성인식프로그램 《룡남산》을 리용하여 대규모RNNLM의 성능평가실험을 진행하였다.

실험에서는 출력층분해를 리용한 대규모재귀신경망언어모형의 학습속도개선 및 정확도평가를 진행하였다. 대비를 위한 기준모형들로서는 출력층분해가 없는 재귀신경망언어모형과 저차원되돌이3-그램언어모형을 리용하였다. 언어모형학습자료로서 40M단어들로 구성되는 《로동신문》본문코퍼스를 리용하였으며 생성된 어휘크기는 60K이다.

먼저 출력층분해를 리용한 대규모재귀신경망언어모형의 학습속도개선평가실험을

Core i3(2GB, 3GHz)급 컴퓨터를 리용하여 진행하였다.

RNNLM1은 출력충분해가 없는 재귀신경망언어모형, RNNLM2는 출력충분해를 리용한 재귀신경망언어모형이라고 하자.

이때 숨은층의 크기를 150으로 고정하고 실험하였다.

RNNLM1에서는 한주기 학습하는데 보통 30h, 10주기 학습하는데 15일 이상의 시간이 걸린다. 한편 RNNLM2에서는 단축목록의 크기에 따라 한주기 학습하는데 보통 2~8h정도 걸린다.

다음 단축목록크기 2 000에 대하여 학습된 RNNLM2를 조선어런속음성인식체계 《룡담산》에 적용하여 단어오유률과 분기수평가를 진행한 결과는 표와 같다.

실험결과들은 출력충분해를 리용하는 재귀신경망이 학습속도를 훨씬 높여 대규모언어모형을 학습할수 있으며 단어오유률과 분기수측면에서 저차원되돌이모형에 비하여 우월하다는것을 보여 주었다.

표. RNNLM2의 단어오유률(WER)과 분기수(PPL)평가

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

맺 는 말

출력충분해를 리용하는 재귀신경망언어모형구축방법을 제안하고 음성인식체계에 적용할 때의 효과성을 검증하였다. 출력충분해를 리용하는 재귀신경망은 언어모형학습속도가 훨씬 빠를뿐아니라 지금까지 많이 리용되어오던 저차원되돌이3-그람언어모형보다 성능이 우월하다.

참 고 문 헌

- [1] 리현순; 정보과학, 2, 52, 주체104(2015).
- [2] H. Schwenk; in Computer, Speech & Language, 3, 21, 492, 2007.
- [3] Y. Bengio et al.; Journal of Machine Learning Research, 3, 1137, 2003.

주체106(2017)년 4월 5일 원고접수

A Method of Improving Study Speed of Large Scale of Recurrent Neural Network Language Model by Factorization of Output Layer

Ri Hyon Sun

We propose a new study method of large scale language model using recurrent neural network. We implemented study speed of large scale recurrent neural network language model by dividing output layer into 3 sublayers.

Key words: reccurent neural network, RNNLM, RNN, language model