

2022학년도 1학기 컴퓨터언어학

제14강 시퀀스 처리 (1)

박수지

서울대학교 인문대학 언어학과

2022년 4월 25일 월요일

오늘의 목표

- 1 PyTorch 라이브러리를 활용하여 순방향 신경망을 훈련시킬 수 있다.
- 2 순방향 신경망을 순환 신경망으로 확장하는 과정을 설명할 수 있다.

	입력	출력
순방향 신경망 (FFNN)	하나의 벡터	하나의 벡터
합성곱 신경망 (CNN)	여러 개의 벡터 (이미지)	하나의 벡터
순환 신경망 (RNN)	여러 개의 벡터 (시퀀스)	여러 개의 벡터

언어 모형이 하는 일

- 단어 시퀀스에 확률을 부여한다.
- 주어진 문맥에서 다음에 나타날 단어를 예측한다.

문제

$P(w_1 w_2 \cdots w_n)$ 를 어떻게 알아내는가?

조건부확률과 연쇄법칙

조건부확률 $P(w|h) = \frac{P(h, w)}{P(h)}$

연쇄법칙 $P(h, w) = P(h)P(w|h)$

연쇄법칙

$$P(h, w) = P(h)P(w|h) \Rightarrow P(w_{1:n}) = \prod_{i=1}^n P(w_i | w_{<i})$$

예시: 길이 $n = 3$

$$\begin{aligned}
 P(w_1 w_2 w_3) &= P(h = w_1 w_2, w = w_3) \\
 &= P(w_1 w_2) P(w_3 | w_1 w_2) \\
 &= P(h = w_1, w = w_2) P(w_3 | w_1 w_2) \\
 &= P(w_1) P(w_2 | w_1) P(w_3 | w_1 w_2) \\
 &= P(w_1 | <s>) P(w_2 | w_1) P(w_3 | w_1 w_2)
 \end{aligned}$$

언어 모형의 평가

훈련집합에서 본 적 없지만 시험집합에 실제로 존재하는 문장에 높은 확률을 부여해야 한다.

복잡도

시험집합 $W = w_1 \dots w_K$ 의 확률의 역수의 K제곱근 \Rightarrow 확률이 커지면 복잡도가 낮아진다.

$$PP(W) = P(w_1 w_2 \dots w_K)^{-\frac{1}{K}} = \sqrt[K]{\frac{1}{P(w_1 w_2 \dots w_K)}}$$

의문

왜 K제곱근을 취하는가?

- 길이 K인 연쇄의 확률: 조건부확률 K개의 곱
- 효과: 문장이 길어질수록 확률이 낮아지는 현상을 보완해 준다.

Feed-Forward Neural Network

$$h = f(\mathbf{W}\vec{x} + \vec{b})$$

\vec{x} 입력층

\vec{h} 은닉층

f 활성화 함수 (주로 ReLU)

Recurrent Neural Network

$$h_t = f(\mathbf{U}\vec{h}_{t-1} + \mathbf{W}\vec{x}_t + \vec{b})$$

t 시간

\vec{x}_t 현재 시점 t 에서의 입력층

\vec{h}_{t-1} 직전 시점 $(t - 1)$ 에서의 은닉층

\vec{h}_t 현재 시점 t 에서의 은닉층

f 활성화 함수 (주로 tanh)

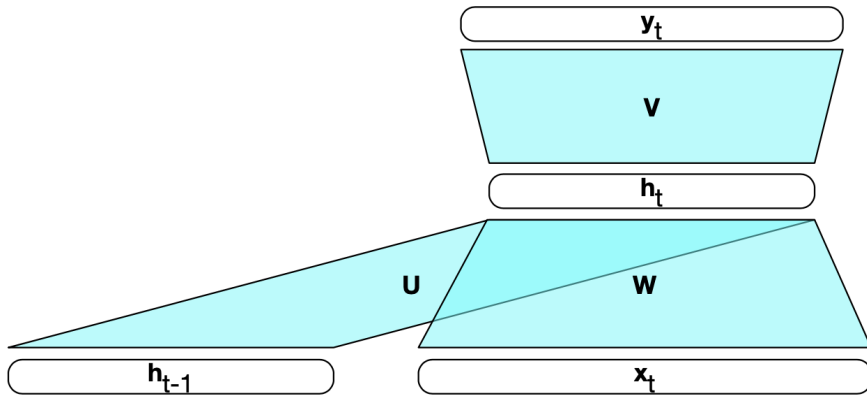


Figure 9.3 Simple recurrent neural network illustrated as a feedforward network.

순환 신경망의 장점

순방향 신경망 언어 모형의 문제점 해결

- 문장 전체를 처음부터 볼 수 있다.
- 단어 연쇄가 길어져도 가중치 패러미터의 수가 늘어나지 않는다.

순환 신경망의 한계점

- 계산이 느리다.
- 실제로 예전 단계의 정보에 접근하는 것은 어렵다. — 기울기 소실 문제