Введение в RNN

Тарасов Денис

20 ноября 2017

Содержание

1. Что такое RNN
2. Преимущества перед обычными НС
3. BPTT
4. Vanishing/Exploding Gradients
5. Примеры
6. Библиография

Что такое RNN

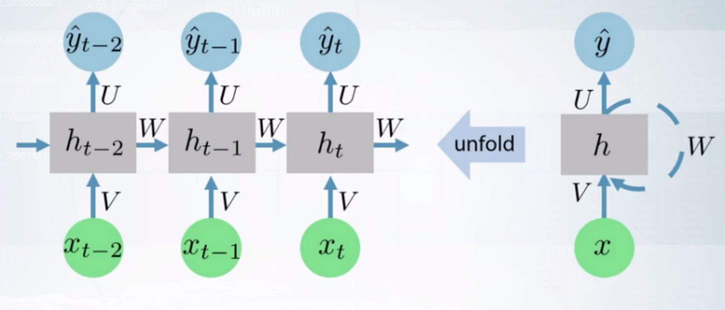


Figure: Coursera, Intro to Deep Learning by HSE

x – входные данные

$\hat{\mathrm{y}}$ – предсказание

h – hidden state

$\mathrm{h_t = f_h(Vx_t + Wh_{t-1} + b_h)}$

$\mathrm{\hat{y}_t = f_y(Uh_t + b_y)}$

Преимущества перед обычными НС

- Можно работать с последовательностями входных данных (например, текст или видео)

Обучение RNN

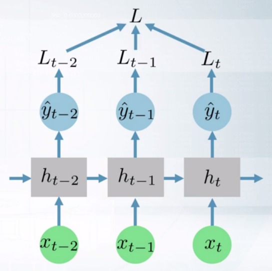


Figure: Coursera, Intro to Deep Learning by HSE

Пусть y_t – настоящее значение, $\hat{y}_t$ – предсказание, а $L_t(y_t, \hat{y}_t)$ – некторая функция потерь, то:

$$\text{loss} = L = \sum_i L_i(y_i, \hat{y}_i)$$

Backpropagation Through Time

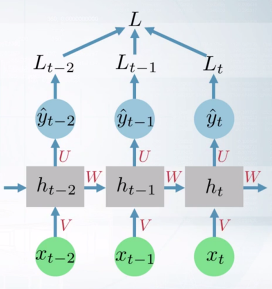


Figure: Coursera, Intro to Deep Learning by HSE

Forward pass: $\mathrm{h_t}, \hat{\mathrm{y}}_\mathrm{t}, \mathrm{L_t}, \mathrm{L}$

Backpropagation Through Time

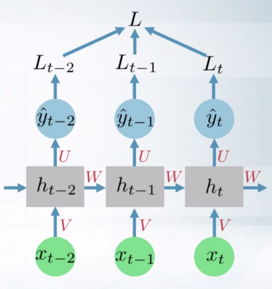
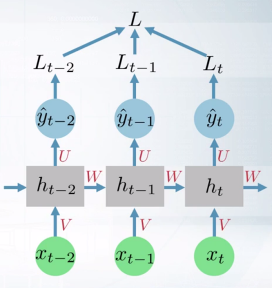


Figure: Coursera, Intro to Deep Learning by HSE

Backward pass: $\frac{\partial \mathrm{L}}{\partial \mathrm{U}}, \frac{\partial \mathrm{L}}{\partial \mathrm{V}}, \frac{\partial \mathrm{L}}{\partial \mathrm{W}}, \frac{\partial \mathrm{L}}{\partial \mathrm{b_x}}, \frac{\partial \mathrm{L}}{\partial \mathrm{b_h}}$

Backpropagation Through Time

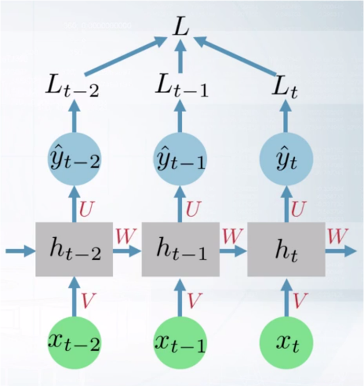


$$\frac{\partial L}{\partial U} = \sum_i \frac{\partial L_i}{\partial U}$$

$$\frac{\partial L_t}{\partial U} = \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial U}$$

$$\hat{y}_t = g(Uh_t + b_y)$$

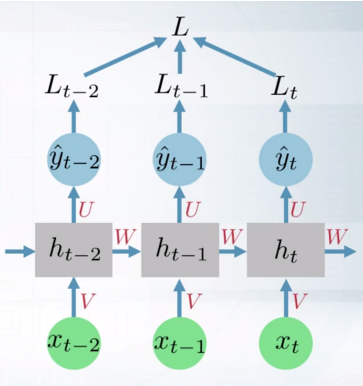
Backpropagation Through Time



$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}} = \sum_{\mathrm{i}} \frac{\partial \mathrm{L_i}}{\partial \mathrm{W}}$$

$$\mathrm{h_t} = \mathrm{f_h}(\mathrm{Vx_t} + \mathrm{Wh_{t-1}} + \mathrm{b_h})$$

Backpropagation Through Time



$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial h_t}\left(\frac{\partial h_t}{\partial W} + \frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} + \cdots\right) =$$

$$= \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial h_t}\sum_{k=0}^{t}\frac{\partial h_t}{\partial h_{t-1}}\cdots\frac{\partial h_{k+1}}{\partial h_k}\frac{\partial h_k}{W}$$

Backpropagation Through Time

Минусы:

- долго работает
- проблема с затуханием градиента

Vanishing/Exploding Gradients

Может возникнуть из-за многоразового повторения функций активации (например, tanh, сигмоида, чьи значения по модулю меньше 1)

Vanishing/Exploding Gradients

Способы решения:

- LSTM
- faster hardware
- ReLU (такая функция активации, что: $f(x) = \max\{0; x\}$)
- Truncated BPTT

Примеры

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

- Пример статьи, созданной с помощью RNN
  http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Примеры

```
/*
     * Don't miss if the descriptor is running this event
     */
    spin_lock_irqsave(&event_srcu->spu_list_lock, flags);
    while (npids && !sig_setup) {
        struct smp_instance *smp_processor = list[i];

        if (pending_identify_sig)
            complete(&(pids.event));
        smp_mb();
        if (pid == 0) {
            pr_warn("Error: enabling event %x\n", pid);
            pid_state += s->pid;

            while (pid) {
                if (smp_processor_id() == -1) {
                    event_for_each_pid(pid, event, upid)
                    break;
                }
            }
        }
    }
```

- Пример кода, созданного с помощью RNN

Примеры

- Синтетический Обама
  http://grail.cs.washington.edu/projects/AudioToObama/
  (Видео)

Библиография

- http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- https://www.coursera.org/learn/intro-to-deep-learning/lecture/zGHtr/simple-rnn-and-backpropagation
- http://colah.github.io/posts/2015-08-Backprop/