

# Hierarchical Recurrent Neural Network for Document Modeling

Rui Lin et al.

Proceedings of the 2015 Conference on Empirical Methods  
in Natural Language Processing (EMNLP)

Presented by Boyoung Kim

April 26, 2018

# Introduction

- 문서는 여러 개의 문장, 문장은 여러 개의 단어로 구성되어 있음
- 목적
  - 다음 단어를 예측 (문장의 historical 정보 사용)

# Notation

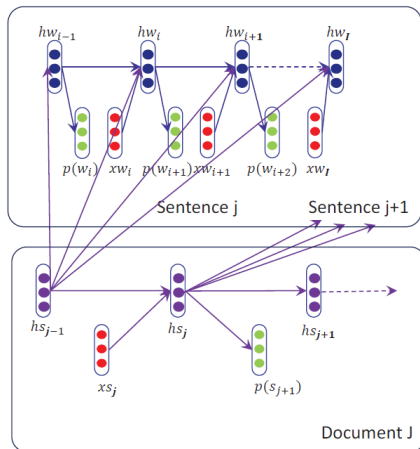
- Word-level

- $xw_i$  : 문장 내의  $i$ 번째 단어에 대한 one-hot vector
- $hw_i$  :  $i$ 번째 단어에 대한 history vector
- $p(w_{i+1})$  :  $i + 1$ 번째 단어를 예측하기 위한 output.

- Sentence-level

- $xs_j$  :  $j$  번째 문장에 대한 word-to-vec 표현.
- $hs_j$  :  $j$  번째 문장에 대한 history vector
- $p(s_{j+1})$  :  $j + 1$ 번째 단어들을 예측하기 위한 output.

# Hierarchical RNN Language Model (HRNNLM)



# Hierarchical RNN Language Model (HRNNLM)

- Sentence-level Model

$$hs_j = f(U_s hs_{j-1} + W_s xs_j)$$

$$p(s_{j+1}) = g(V_s hs_j)$$

여기서  $W_s, U_s, V_s$  는 weight matrix,  $f(\cdot) : HTanh$  함수,  $g(\cdot) : softmax$  함수

$\Rightarrow p(s_{j+1})$ 에  $j$ 번째 문장의 영향 강조하기 위해 다음과 같은 feature hashing 사용

$$p(s_{j+1})(t) = g(V_s(t)hs_j + \sum_{w \in s_j} D_{hash(w,t)})$$

여기서  $(t)$  는 vector의  $t$ 번째 원소 또는 matrix의  $t$ 번째 행,  $D_{hash(\cdot)}$  은  $hash(\cdot)$ 에 해당되는 weight,  $hash(\cdot)$ 는 feature를 고정된 길이의 인덱스로 매핑하는 함수,  $s_j$  는  $j$ 번째 문장의 단어집합

# Hierarchical RNN Language Model (HRNNLM)

- Word-level Model

$$hw_i = f(U_w hw_{i-1} + W_w xw_i)$$

$$p(w_{i+1})(t) = g(V_w(t)hw_i + D_{hash^D(w_i,t)}^w)$$

$f(\cdot)$  :  $HTanh$  함수,  $g(\cdot)$  : softmax 함수

$w_i$ 는  $i$ 번째 단어