AI Solution for AI needs customer

Mr.MIND

# RNN과 Sentence Embedding

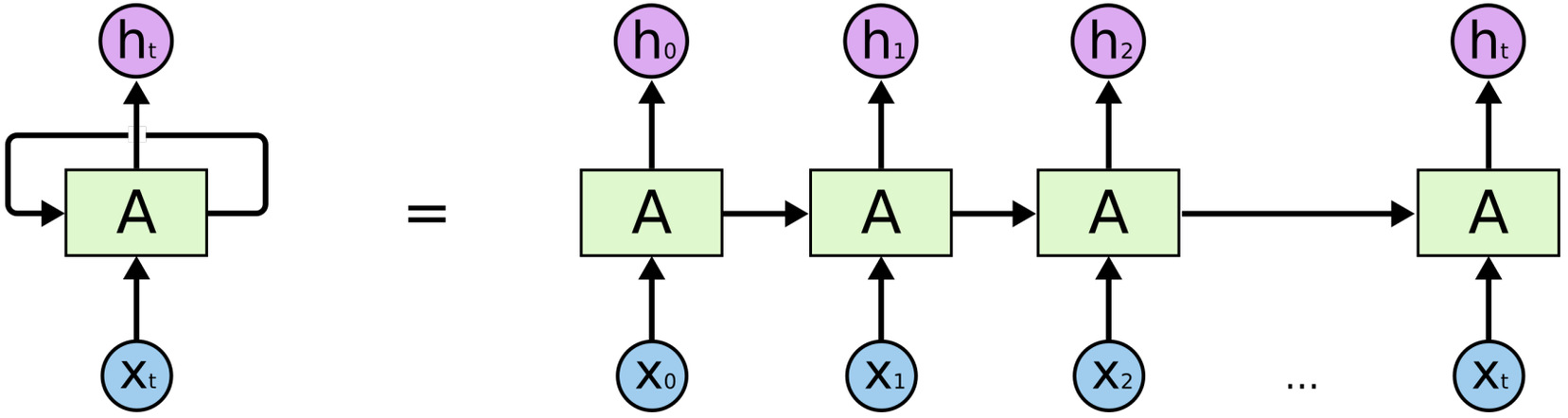
최태균 Mindmap.ai 선임연구원
| tgchoi03@gmail.com | 010-2004-1188

V1.0 | Released in 2018.01
Document by **Mr.MIND AI Consulting**

# 1. RNN이란?

# RNN모델

## Recurrent Neural Network Model




Unrolled된 Recurrent Neural Network 모델

순서가 있는 input 벡터들이 $x_1, x_2, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_T$ 있을 때
매 time step에 대한 RNN은 다음과 같은 연산을 한다.

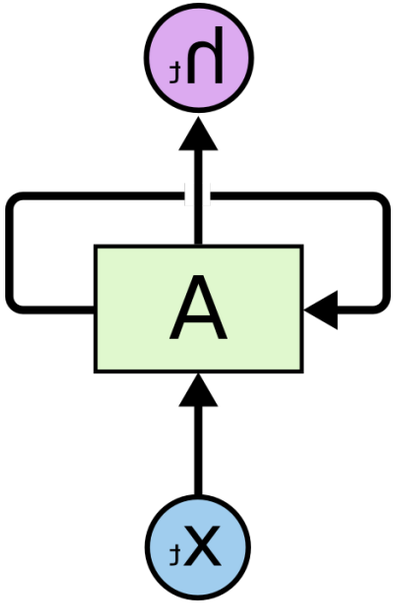
$$h_t = \sigma\left(W^{(hh)} h_{t-1} + W^{(hx)} x_{[t]}\right)$$

$$\hat{y}_t = \text{softmax}\left(W^{(S)} h_t\right)$$

# RNN모델

## Recurrent Neural Network Model

- 주된 아이디어는 모든 time step에서 파라메터 weight $W$에 대해서 공유한다는 점이다.




$$h_t = f_W(h_{t-1}, x_t)$$

new state

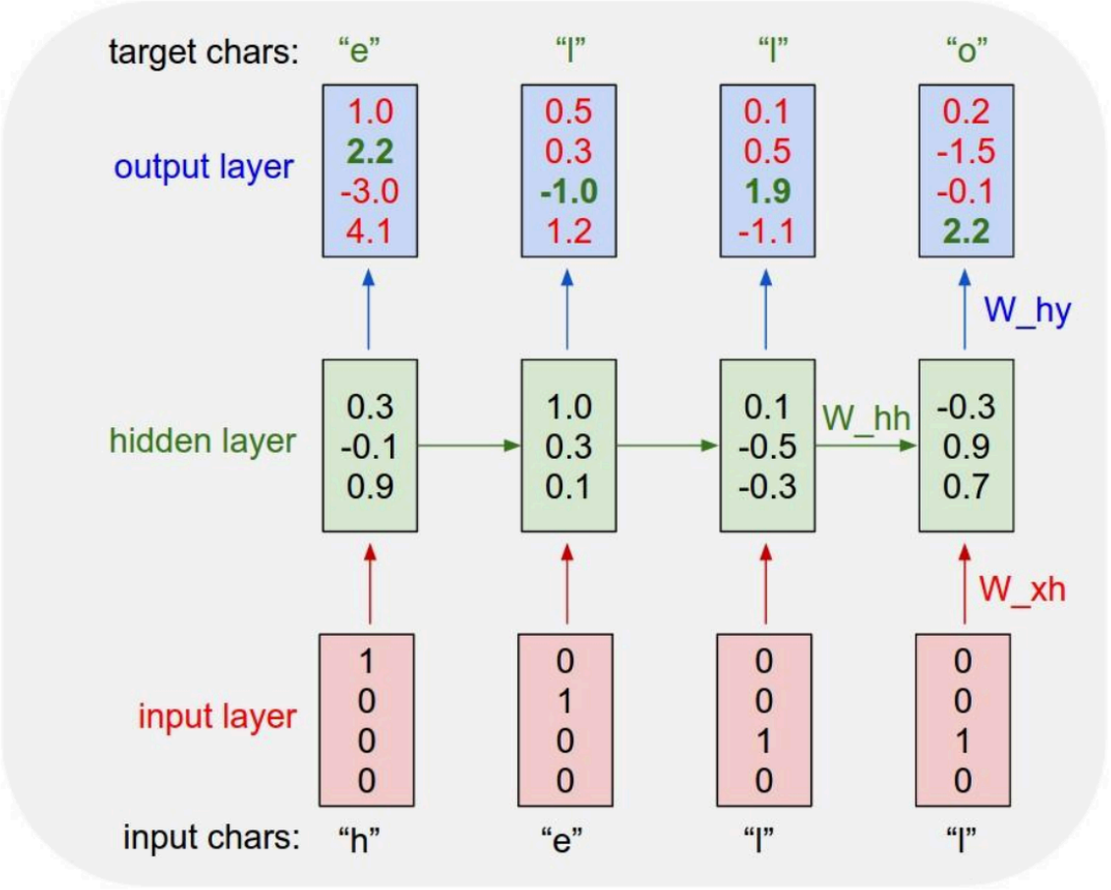
some function with parameters W

old state

input vector at some time step

# RNN모델

## Character-level language model example




입력 Vocabulary Set : ['h', 'e', 'l', 'o']

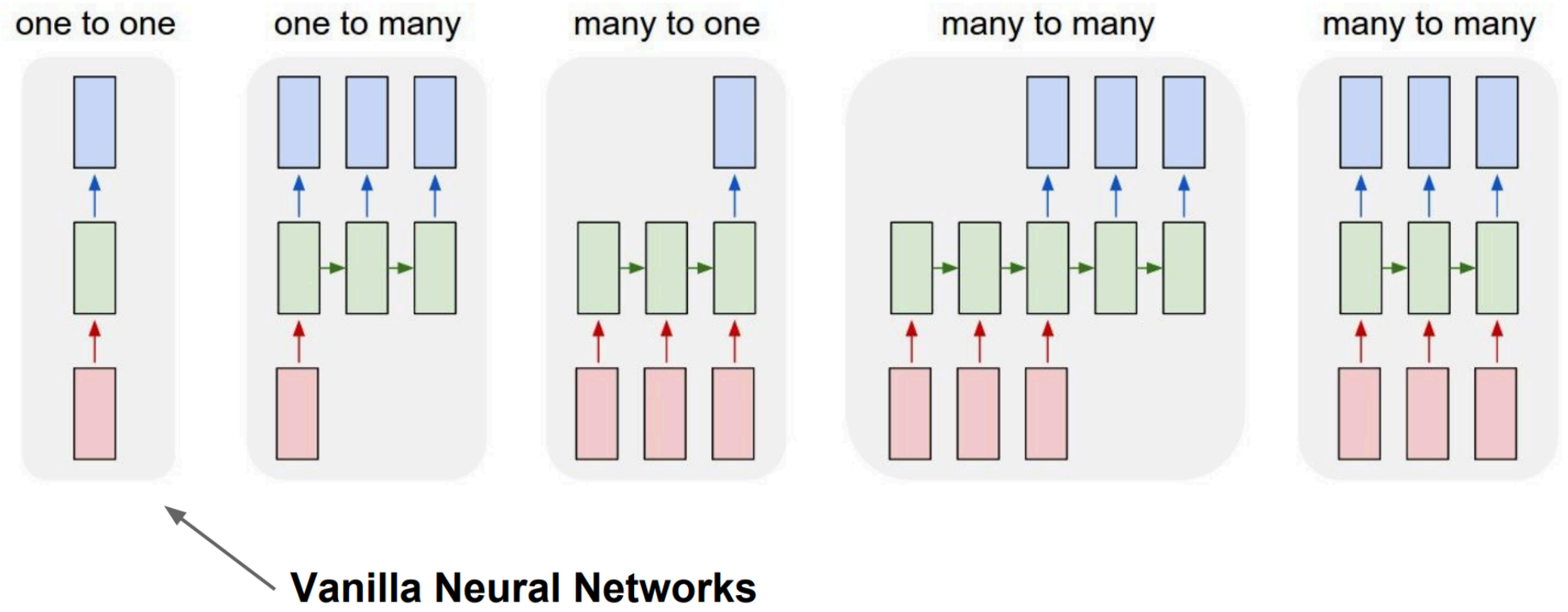
학습하고자 하는 Sequence : "hello"

$$h_t = \sigma\left(W^{(hh)}h_{t-1} + W^{(hx)}x_{[t]}\right)$$

$$\hat{y}_t = \text{softmax}\left(W^{(S)}h_t\right)$$

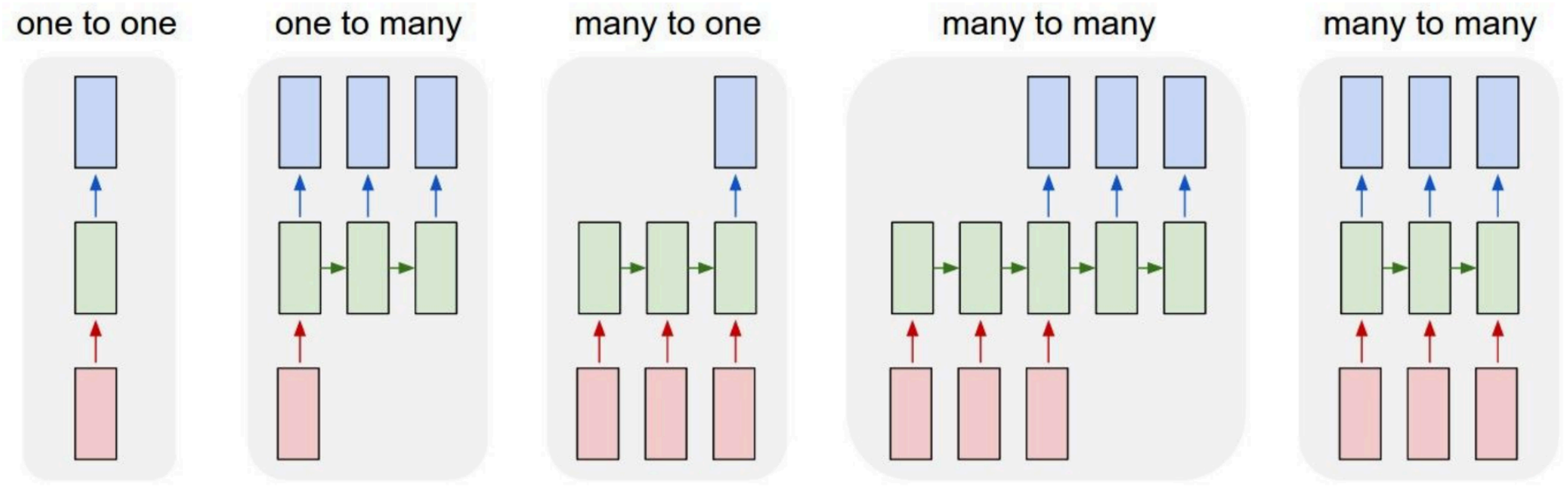
## RNN모델

# 다양한 RNN 모델 구조



## RNN모델

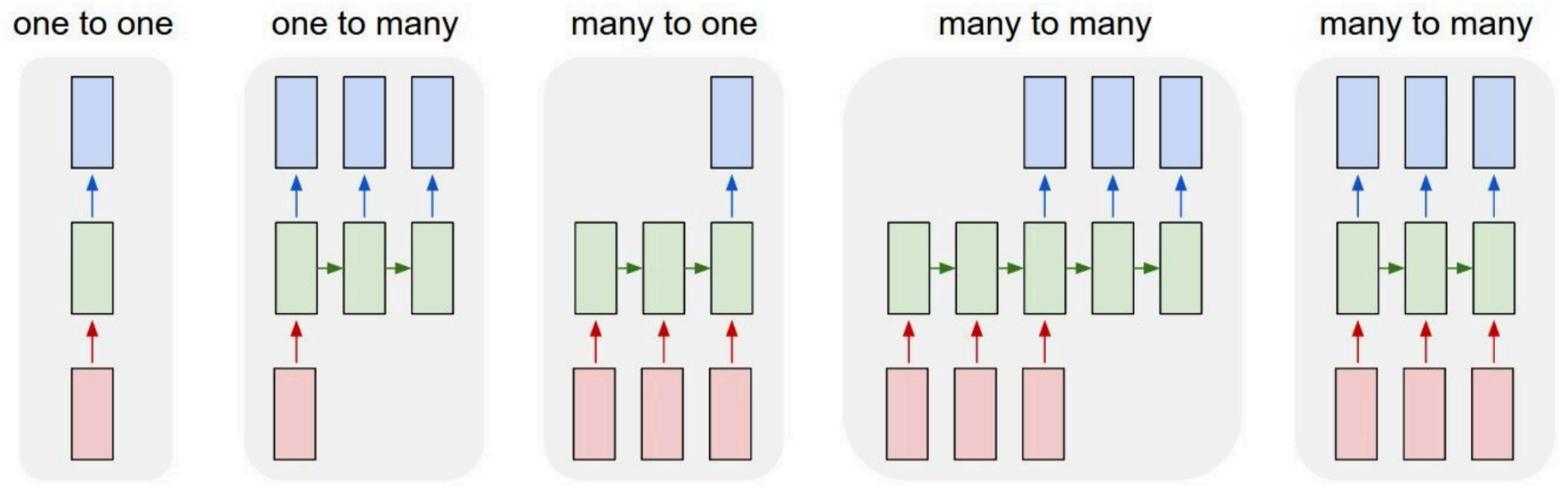
# 다양한 RNN 모델 구조




e.g. **Image Captioning**
image -> sequence of words

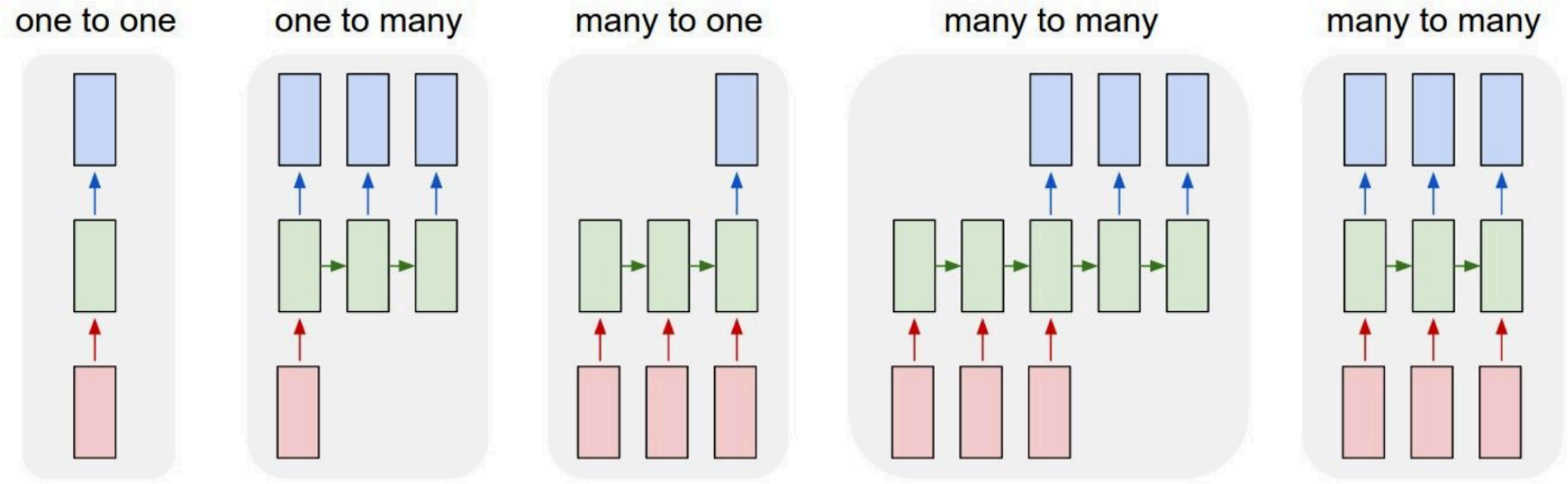
## RNN모델

# 다양한 RNN 모델 구조



## RNN모델

# 다양한 RNN 모델 구조

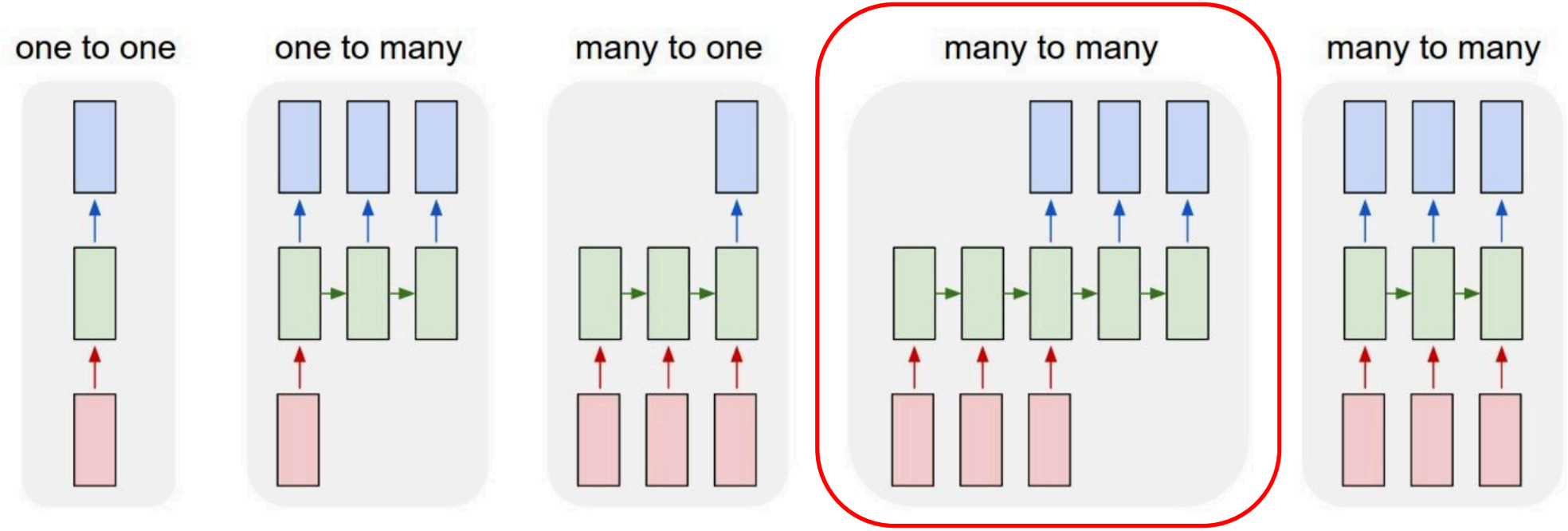



e.g. **Machine Translation**
seq of words -> seq of words

## RNN모델

# 다양한 RNN 모델 구조

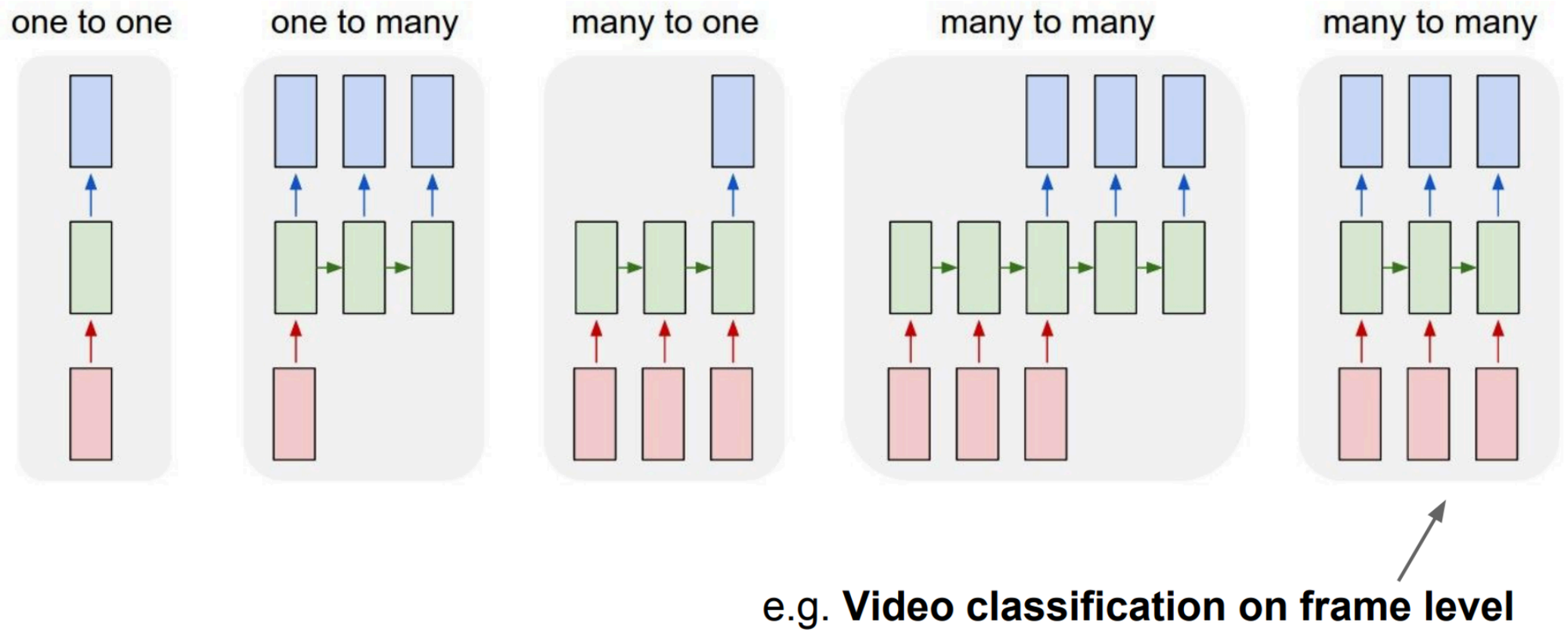
챗봇을 구현할 때 다루게 될 기술




e.g. **Machine Translation**
seq of words -> seq of words

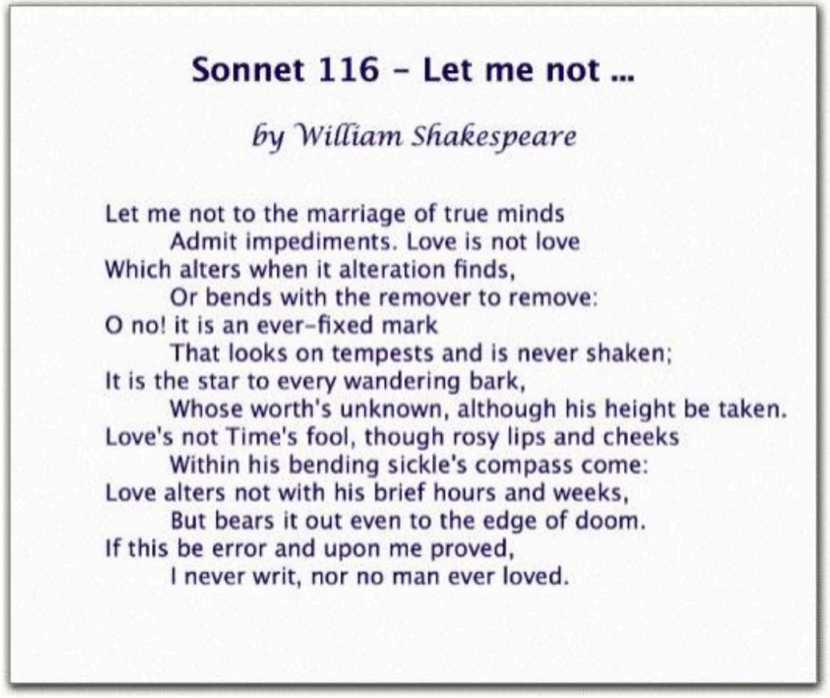
## RNN모델

# 다양한 RNN 모델 구조



## RNN모델

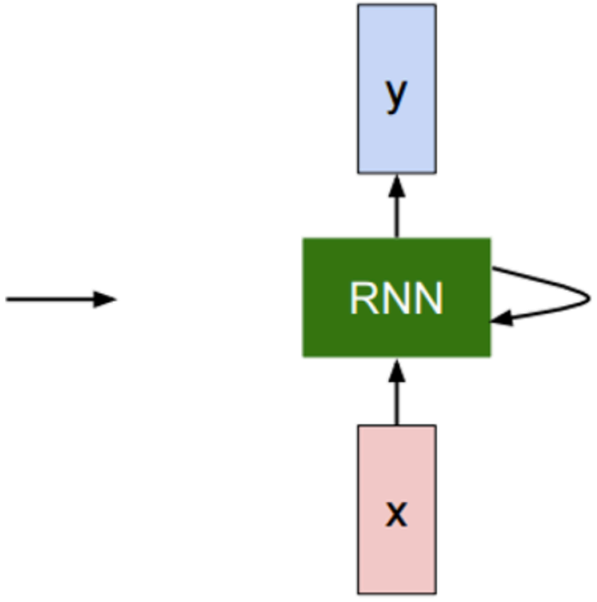
RNN으로 희곡을 학습시킨다면..?



**Sonnet 116 – Let me not ...**

*by William Shakespeare*

Let me not to the marriage of true minds
Admit impediments. Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken;
It is the star to every wandering bark,
Whose worth's unknown, although his height be taken.
Love's not Time's fool, though rosy lips and cheeks
Within his bending sickle's compass come:
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.
If this be error and upon me proved,
I never writ, nor no man ever loved.



## RNN모델 예시

RNN으로 희곡을 학습시킨다면..?

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

train more

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

train more

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

train more

```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

## RNN모델 예시

# RNN으로 희곡을 학습시킨다면..?

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

# RNN모델 예시

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 compariciy 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

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

which has a nonzero morphism we may assume that $f_i$ is of finite presentation over $S$. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M^\bullet} = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of $X$. Thus $U$ is affine. This is a continuous map of $X$ is the inverse, the groupoid scheme $S$.

*Proof.* See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace $S$ by $X_{spaces,étale}$ which gives an open subspace of $X$ and $T$ equal to $S_{Zar}$, see Descent, Lemma ??. Namely, by Lemma ?? we see that $R$ is geometrically regular over $S$.

**Lemma 0.1.** *Assume (3) and (3) by the construction in the description.*

*Suppose* $X = \lim |X|$ *(by the formal open covering* $X$ *and a single map* $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ *over* $U$ *compatible with the complex*

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

*When in this case of to show that* $\mathcal{Q} \to \mathcal{C}_{Z/X}$ *is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition ?? (without element is when the closed subschemes are catenary. If* $T$ *is surjective we may assume that* $T$ *is connected with residue fields of* $S$*. Moreover there exists a closed subspace* $Z \subset X$ *of* $X$ *where* $U$ *in* $X'$ *is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) $f$ *is locally of finite type. Since* $S = \mathrm{Spec}(R)$ *and* $Y = \mathrm{Spec}(R)$.

*Proof.* This is form all sheaves of sheaves on $X$. But given a scheme $U$ and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme $X$ over $S$ at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

**Lemma 0.2.** *Let* $X$ *be a locally Noetherian scheme over* $S$, $E = \mathcal{F}_{X/S}$. *Set* $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. *Since* $\mathcal{I}^n \subset \mathcal{I}^n$ *are nonzero over* $i_0 \leq \mathfrak{p}$ *is a subset of* $\mathcal{J}_{n,0} \circ \overline{A}_2$ *works.*

**Lemma 0.3.** *In Situation ??. Hence we may assume* $\mathfrak{q}' = 0$.

*Proof.* We will use the property we see that $\mathfrak{p}$ is the mext functor (??). On the other hand, by Lemma ?? we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where $K$ is an $F$-algebra where $\delta_{n+1}$ is a scheme over $S$. □

# RNN모델 예시

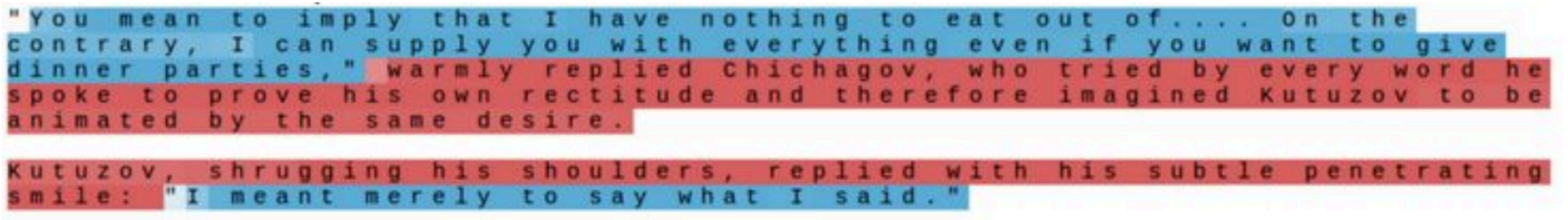
```
static void do_command(struct seq_file *m, void *v)
{
  int column = 32 << (cmd[2] & 0x80);
  if (state)
    cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
  else
    seq = 1;
  for (i = 0; i < 16; i++) {
    if (k & (1 << 1))
      pipe = (in_use & UMXTHREAD_UNCCA) +
        ((count & 0x00000000fffffff8) & 0x000000f) << 8;
    if (count == 0)
      sub(pid, ppc_md.kexec_handle, 0x20000000);
    pipe_set_bytes(i, 0);
  }
  /* Free our user pages pointer to place camera if all dash */
  subsystem_info = &of_changes[PAGE_SIZE];
  rek_controls(offset, idx, &soffset);
  /* Now we want to deliberately put it to device */
  control_check_polarity(&context, val, 0);
  for (i = 0; i < COUNTER; i++)
    seq_puts(s, "policy ");
}
```

```
/*
 *  Copyright (c) 2006-2010, Intel Mobile Communications.  All rights reserved.
 *
 *   This program is free software; you can redistribute it and/or modify it
 * under the terms of the GNU General Public License version 2 as published by
 * the Free Software Foundation.
 *
 *        This program is distributed in the hope that it will be useful,
 * but WITHOUT ANY WARRANTY; without even the implied warranty of
 *   MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
 *
 *  GNU General Public License for more details.
 *
 *   You should have received a copy of the GNU General Public License
 *    along with this program; if not, write to the Free Software Foundation,
 *  Inc., 675 Mass Ave, Cambridge, MA 02139, USA.
 */

#include <linux/kexec.h>
#include <linux/errno.h>
#include <linux/io.h>
#include <linux/platform_device.h>
#include <linux/multi.h>
#include <linux/ckevent.h>

#include <asm/io.h>
#include <asm/prom.h>
#include <asm/e820.h>
#include <asm/system_info.h>
#include <asm/setew.h>
#include <asm/pgproto.h>
```

## RNN모델 예시

RNN은 데이터의 무엇을 보는가?

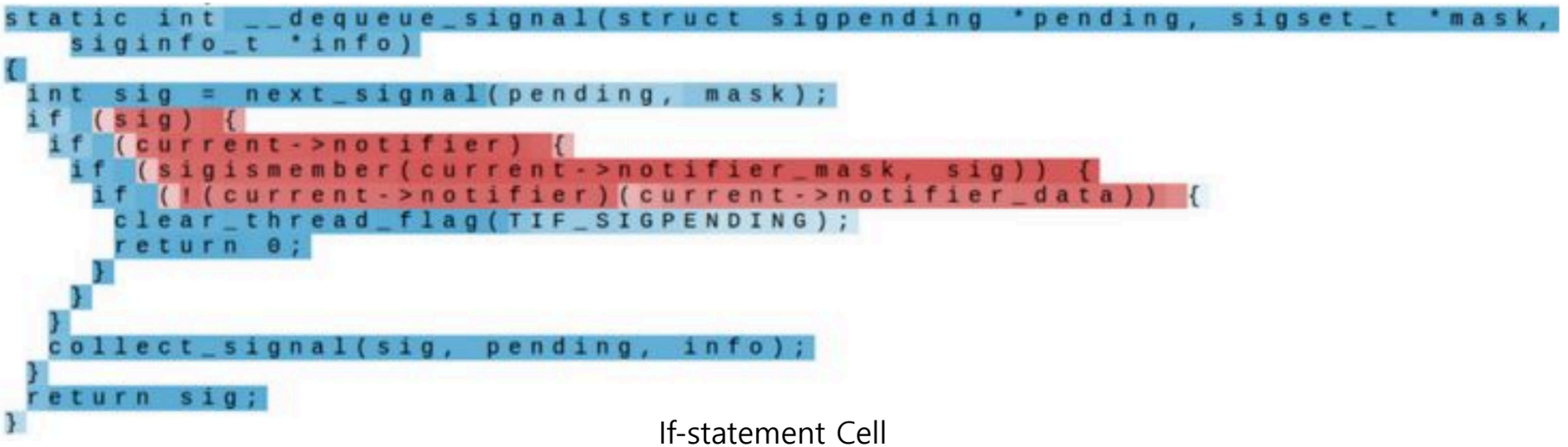
- RNN은 데이터에서 일정하게 반복되는 Sequence의 특징들을 포착한다.
  - New-line 지점
  - Quotation 영역
  - C 언어에서 If-statement 영역 등




Quote Detection Cell

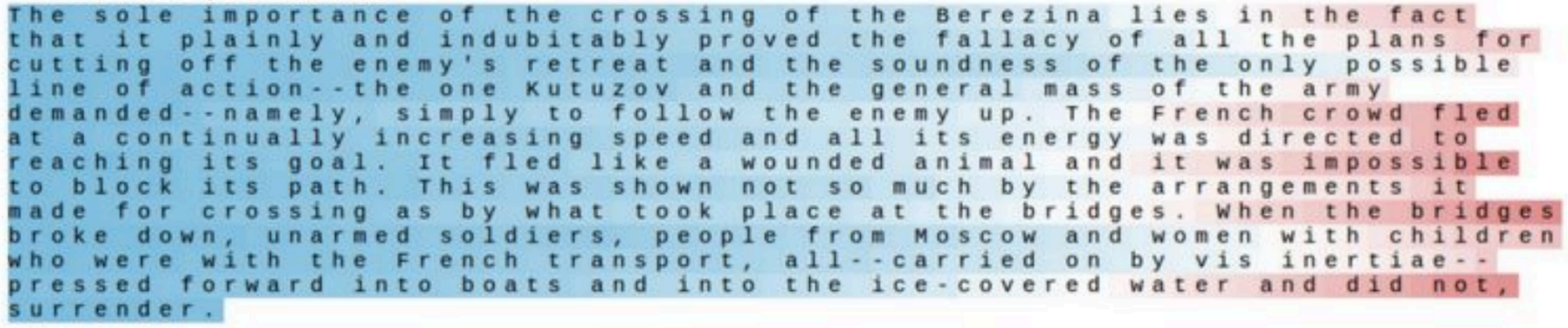
## RNN모델 예시

# RNN은 데이터의 무엇을 보는가?



```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
        siginfo_t *info)
{
  int sig = next_signal(pending, mask);
  if (sig) {
    if (current->notifier) {
      if (sigismember(current->notifier_mask, sig)) {
        if (!(current->notifier)(current->notifier_data)) {
          clear_thread_flag(TIF_SIGPENDING);
          return 0;
        }
      }
    }
    collect_signal(sig, pending, info);
  }
  return sig;
}
```

If-statement Cell



```
The sole importance of the crossing of the Berezina lies in the fact
that it plainly and indubitably proved the fallacy of all the plans for
cutting off the enemy's retreat and the soundness of the only possible
line of action--the one Kutuzov and the general mass of the army
demanded--namely, simply to follow the enemy up. The French crowd fled
at a continually increasing speed and all its energy was directed to
reaching its goal. It fled like a wounded animal and it was impossible
to block its path. This was shown not so much by the arrangements it
made for crossing as by what took place at the bridges. When the bridges
broke down, unarmed soldiers, people from Moscow and women with children
who were with the French transport, all--carried on by vis inertiae--
pressed forward into boats and into the ice-covered water and did not,
surrender.
```

Line-length Tracking Cell

## 언어모델은 왜 RNN을 필요로 하는가?

### 왜 RNN을 언어모델에서 사용하는가?

- 고전적인 언어모델은 확률기반으로 모델을 만든다.

$$P(w_1, \ldots, w_m) = \prod_{i=1}^{m} P(w_i \mid w_1, \ldots, w_{i-1}) \approx \prod_{i=1}^{m} P(w_i \mid w_{i-(n-1)}, \ldots, w_{i-1})$$

- Markov assumption과 n-gram을 활용한 backoff 기법들을 활용한다.
- 이전 step에 있는 단어정보들을 크게 아우르기에는 자원소모가 크다.

  *Recent state of the art: Scalable Modified Kneser-Ney Language Model Estimation by Heafield et al.: "Using one machine with 140 GB RAM for 2.8 days, we built an unpruned model on 126 billion tokens"*

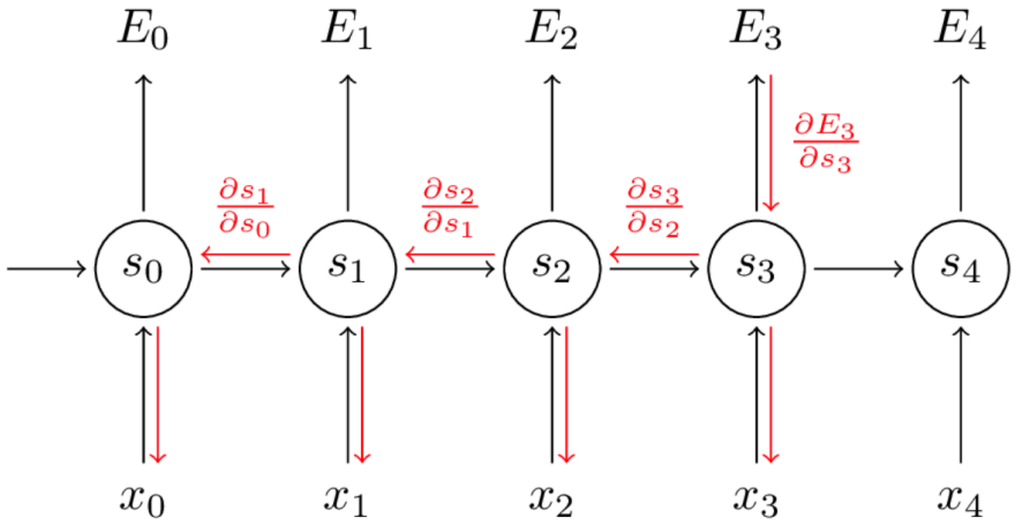
- **RNN의 경우 파라메터 Weight을 통해서 이전 time step에 있는 정보를 포함할 수 있다.**

# 2. RNN의 한계와 LSTM

# RNN의 한계

## Backpropagation Through Time (BPTT)

- RNN을 학습하게 될 때, 매 time step에 따른 Backpropagation을 하게되는데, 이를 BPTT라 한다.
- DNN (Deep Feedforward Neural Network)과 같은방식이지만, BPTT는 하나의 Weight에 대해서 Gradient를 더해준다.



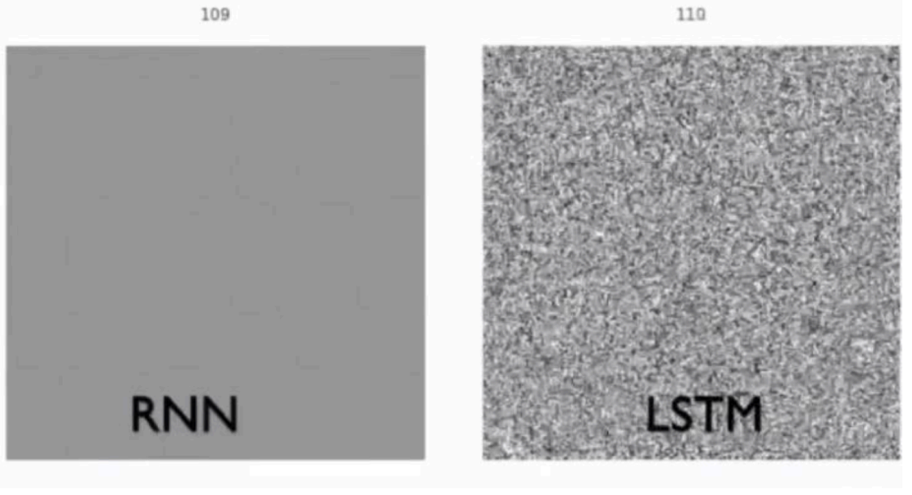

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial s_3} \frac{\partial s_3}{\partial s_k} \frac{\partial s_k}{\partial W}$$

$E_3$의 W에 대한 Gradient 값을 구하는 과정

# RNN의 한계

## Gradient Vanishing / Explosion

- BPTT 중에 파라메터 weight $W$에 대해서 반복적으로 곱하면, 결국 매 time step의 gradient는 지수적으로 증가하거나 감소할 것이다.
- $tanh$ 함수에 대한 미분값의 범위는 0~1 사이이다.
- 역으로 Feed-forward시 time step이 길어지게 되면 앞에 있던 hidden state 정보에 대해서 잃게 된다. => 다음 step에 대한 예측이 어려워짐. (Long Term Dependency Problem)

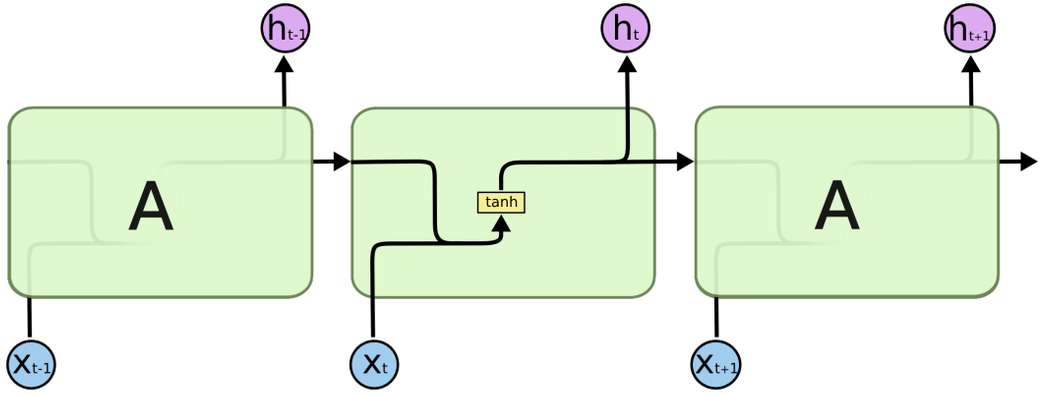



110 정도의 time step에서
RNN의 gradient값은 0에 수렴인 상태이다.

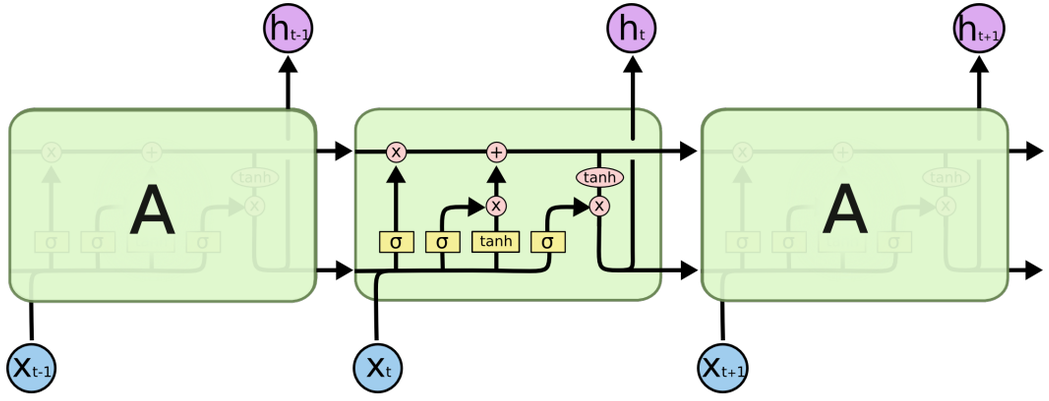
# LSTM모델

## Long Short Term Memory Network

- RNN에서 Long Term Dependency를 해결하기 위해 제안한 모델




RNN 모델

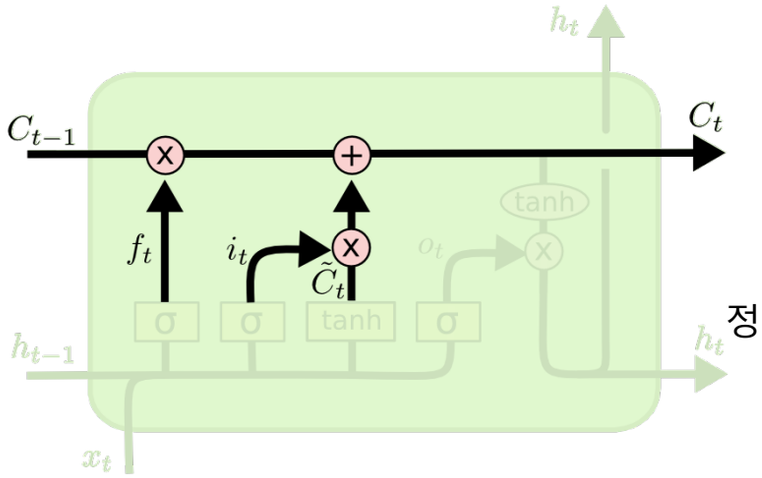



LSTM 모델

## LSTM모델

### Long Short Term Memory Network

- 주된 아이디어는 cell state를 통해서 이전 time step에 있는 hidden state 정보를 보존하거나 버리는 방법이다.




새로운 정보를 갱신하는 정도를 표현

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

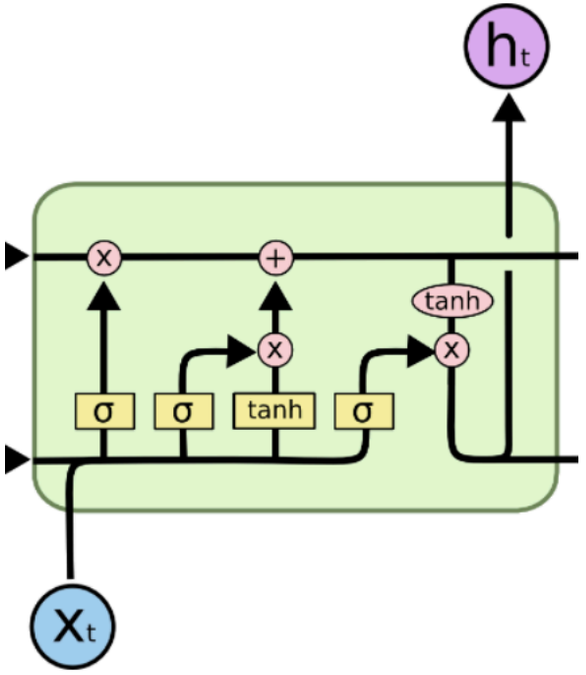
정보를 보존하는 정도를 표현

매 time step에서 cell state에 대한 정보를 갱신한다.

## LSTM모델

### LSTM모델 구성

- Forget(f), input(i), output(o) gate와 input 정보에 대한 weighted activate function(g) 으로 구성한다.




$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \tanh \end{pmatrix} W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

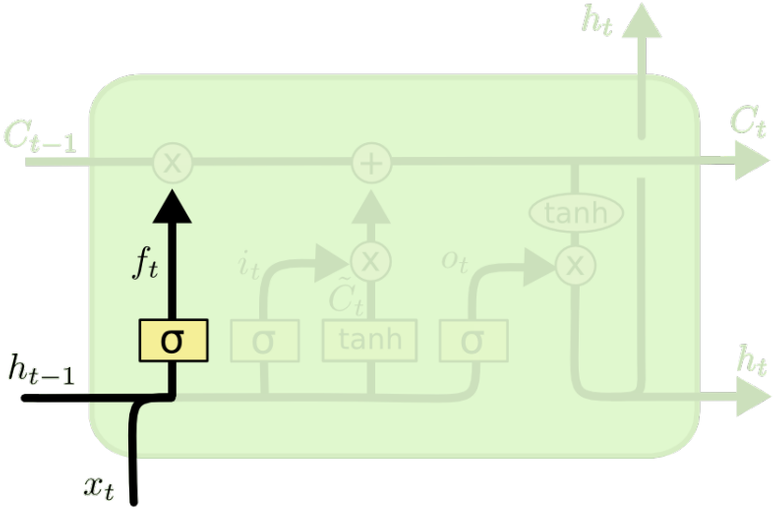
$$c_t^l = f \odot c_{t-1}^l + i \odot g$$

$$h_t^l = o \odot \tanh(c_t^l)$$

## LSTM모델

### Forget Gate

- 이전 time step 까지의 cell state에 대해서 현재 time step부터 어느정도로 정보를 덜 반영할 것인지를 결정한다.
- 직전 hidden state (short-term state) 정보와 현재 input vector의 조합된 정보를 통해서 0~1사이의 값의 vector로 나타낸다.

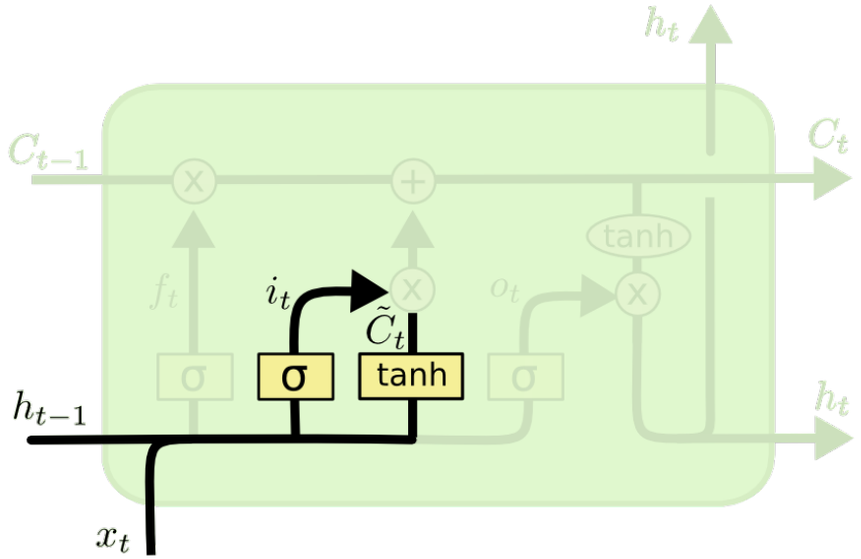



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

## LSTM모델

### Input Gate

- 이전 time step 까지의 cell state에 대해서 현재 time step에 대한 정보를 얼마나 줄 것인지를 결정한다.



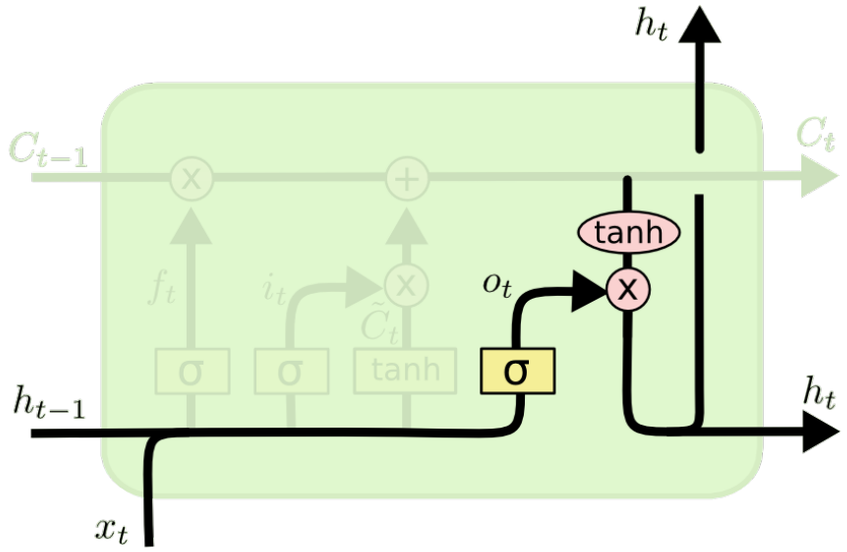

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

# LSTM모델

## Output Gate

- 현재 time step에서 갱신된 cell state를 얼마나 hidden state으로 반영할 것인지 결정한다.



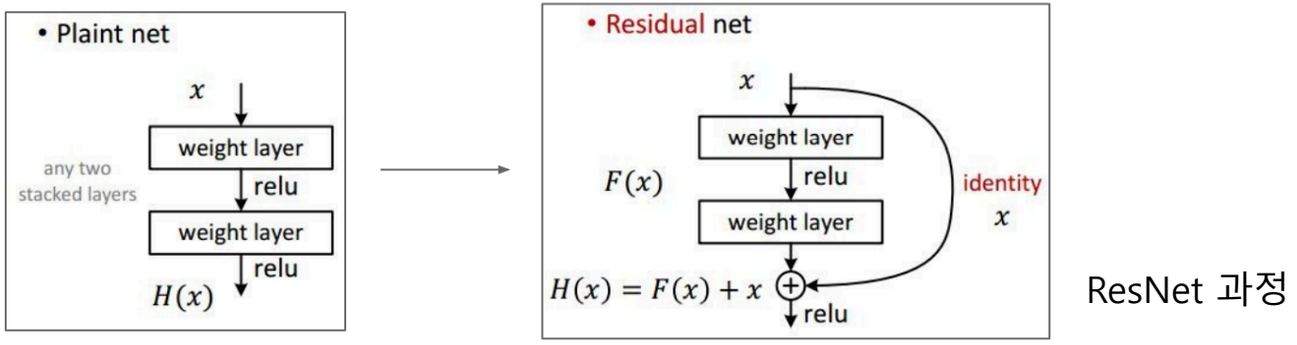

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

## LSTM모델

어떻게 LSTM이 RNN의 한계를 해결할 수 있을까?

- LSTM에서 cell state에 대한 정보를 갱신하는 과정에서 덧셈연산의 역할을 크게 2가지로 볼 수 있다.
  - ➢ 이전 time step에 대한 hidden state 정보가 보존된다.
    (Long Term Dependency Problem 해결)
  - ➢ Backpropagation 중 덧셈에 대한 미분의 역할이 각 time step의 gradient 값에 대해 보존하는 역할을 한다. (Gradient Vanishing Problem 해결)
- 이와 같은 역할는 Residual Network와 Highway Network에서도 확인 할 수 있다.




ResNet 과정

## LSTM모델

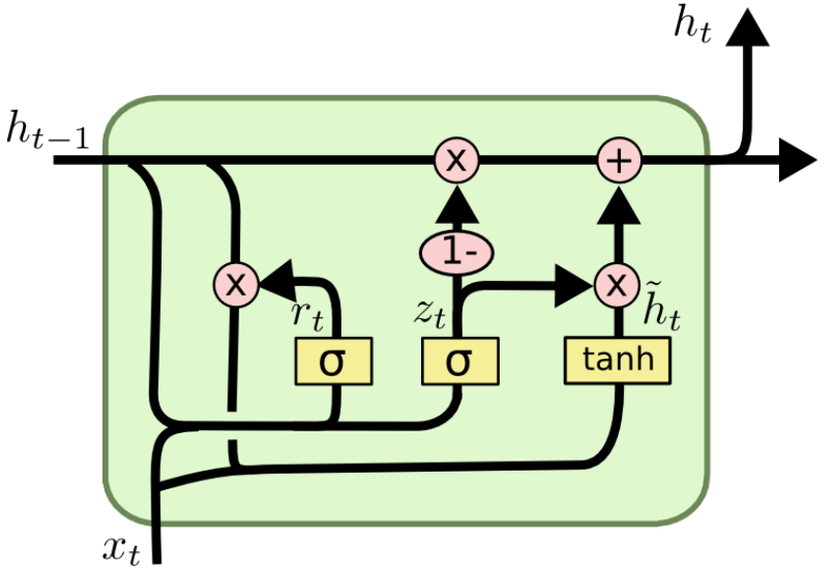
### LSTM의 효과

- Long-term Dependency과 Gradient Vanishing 문제를 해결
  - 긴 sequence에 대해서 앞선 time step에 대한 정보반영이 잘 된다.
  - 긴 sequence에 대한 Gradient Vanishing을 막을 수 있다.

# GRU모델

## Gated Recurrent Unit Model

- LSTM 모델을 변형한 모델
- LSTM의 cell state와 hidden state를 합친 것이 큰 차이
- LSTM 모델과 같이 많이 쓰이는 모델




$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

# 3. Sentence Embedding

## 문장을 하나의 벡터로 표현한다면?

하나의 문장을 어떻게 벡터로 표현할 수 있을까?

Word ⟶ Sentence ?

$$\text{linguistics} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix} \qquad \text{Natural language processing is fun.} = \begin{pmatrix} -0.132 \\ 1.129 \\ 0.827 \\ 0.110 \\ -0.527 \\ 0.156 \\ 0.349 \\ -0.286 \end{pmatrix}$$

## 문장을 하나의 벡터로 표현한다면?

하나의 문장을 어떻게 벡터로 표현할 수 있을까?

- Word Vector를 평균을 내본다.

$$v(\text{“natural language processing”}) =$$
$$1/3\,(v(\text{“natural”}) + v(\text{“language”}) + v(\text{“processing”}))$$

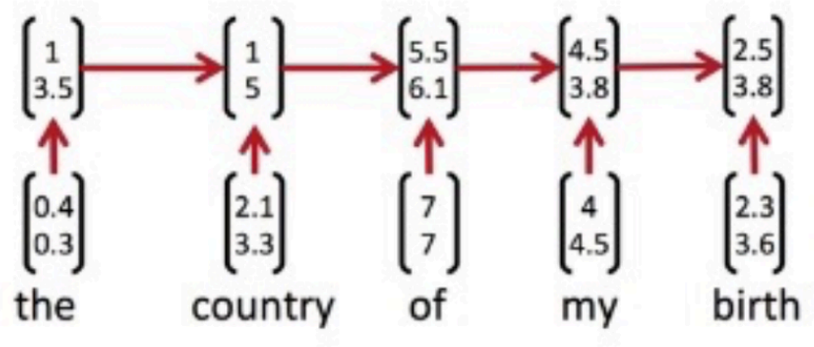
- 만약에 두 문장을 위 방식으로 Sentence Embedding Vector를 만든다면, 두 vector는 과연 의미적으로 서로 다른 문장이라 말할까?

1. You are going there to teach not play.
2. You are going there to play not teach

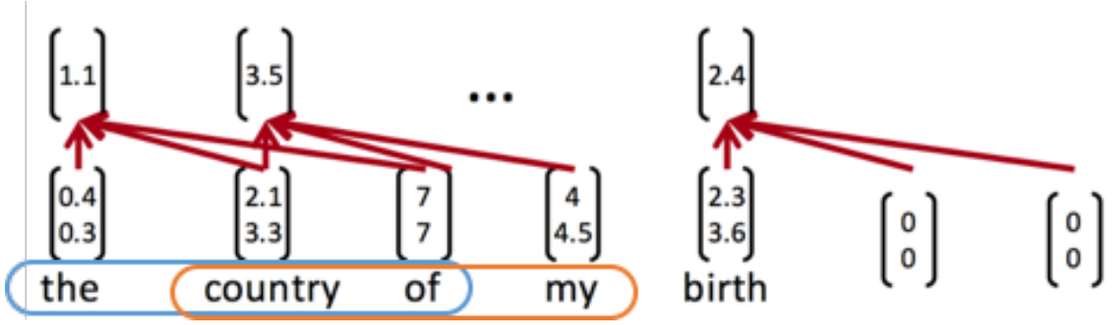
# 문장을 하나의 벡터로 표현한다면?

## Neural Network를 활용한 Sentence Embedding

- Recurrent Neural Network

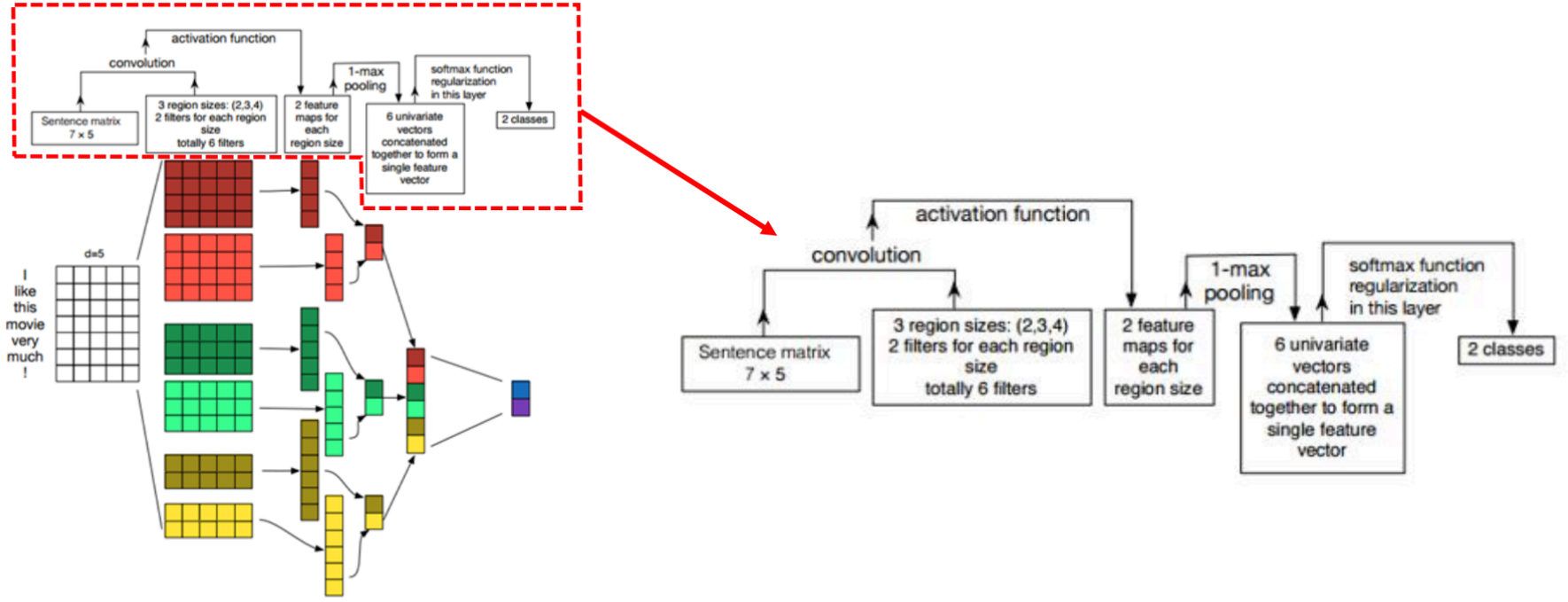



- Convolutional Neural Network



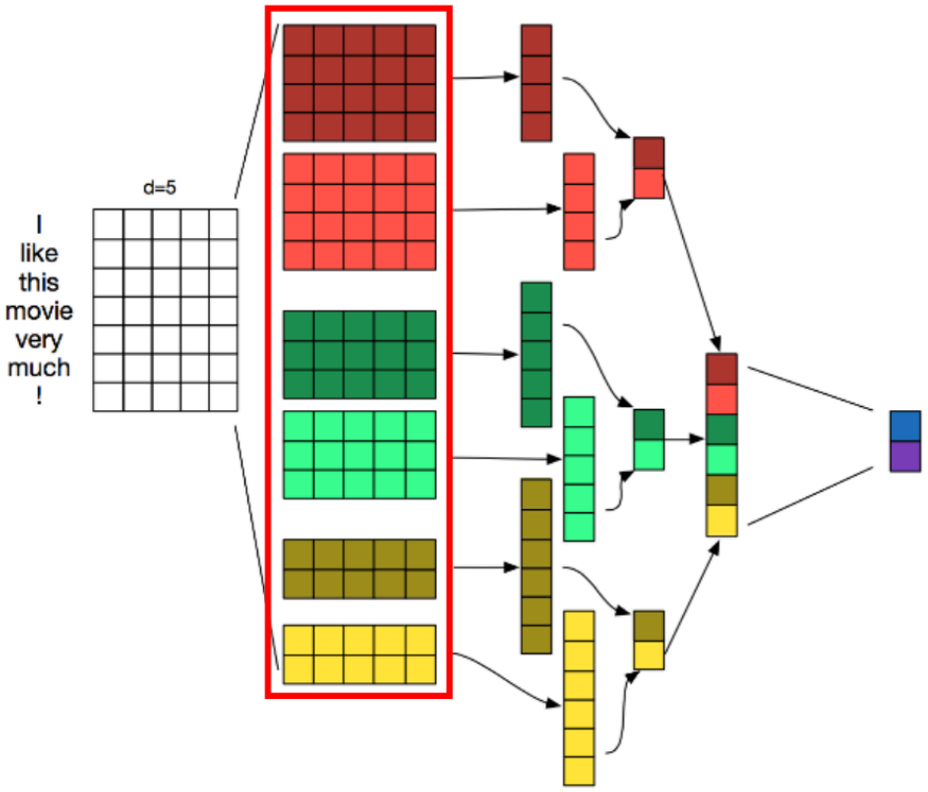
# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet



# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet




**3 region size: (2, 3, 4)**

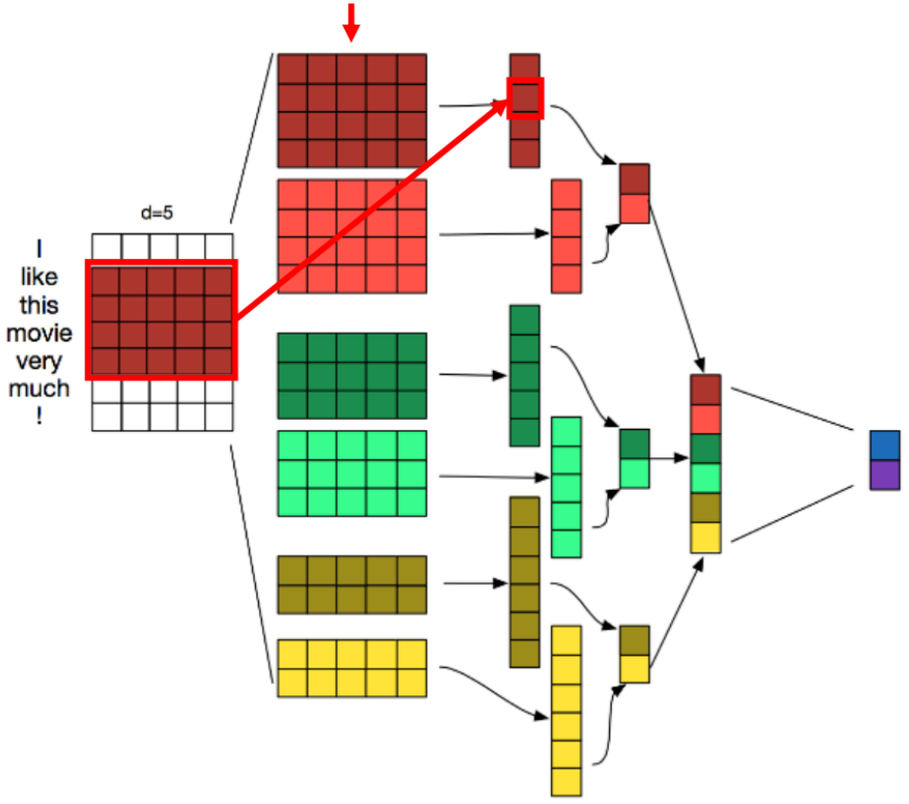
- 각각 bigram, trigram과 4-gram featur를 추출할 수 있다

**각 region size마다 2개의 filter**

- 같은 region size의 filte를 여러 개 사용할 수 있다. (e.g. bigram feature가 여러 개가 있을 수 있다)

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

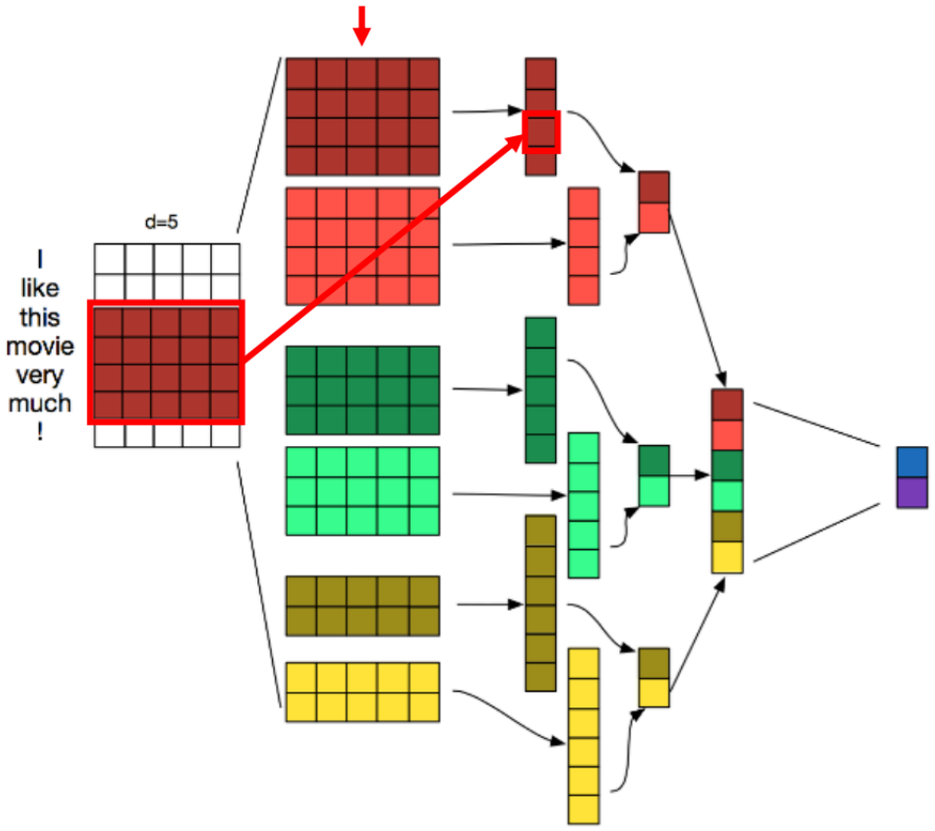



**Convolution**

**4-gram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

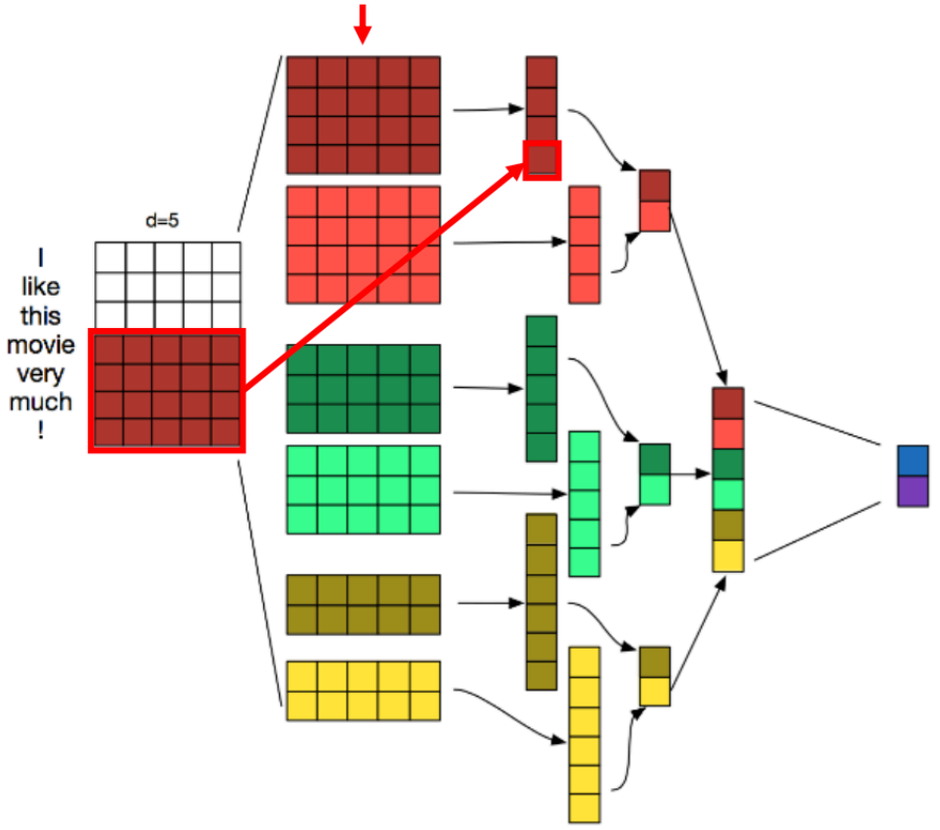



**Convolution**

**4-gram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

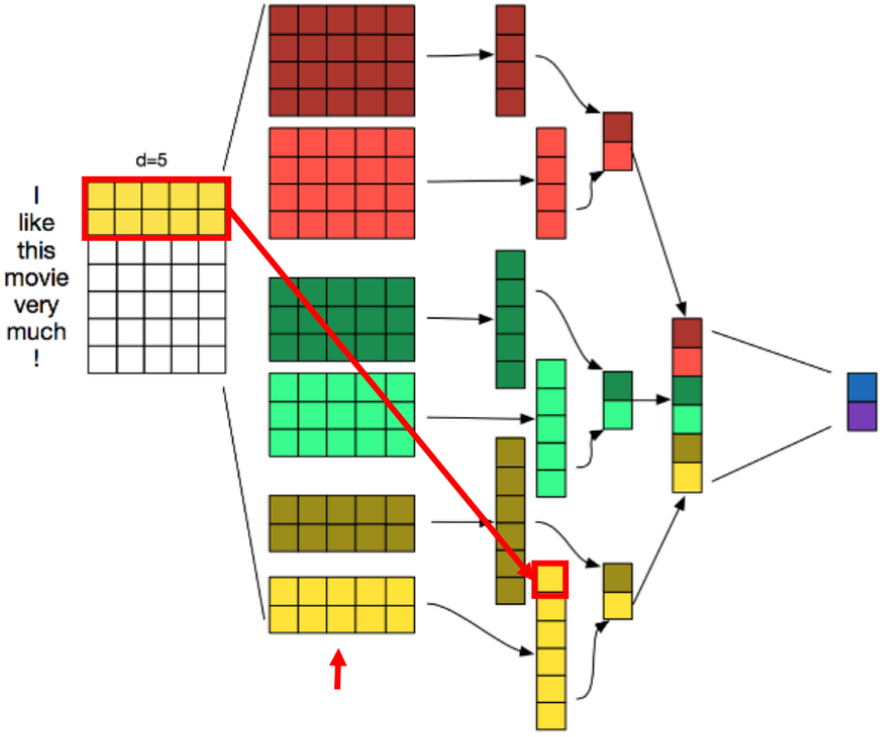



**Convolution**

**4-gram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

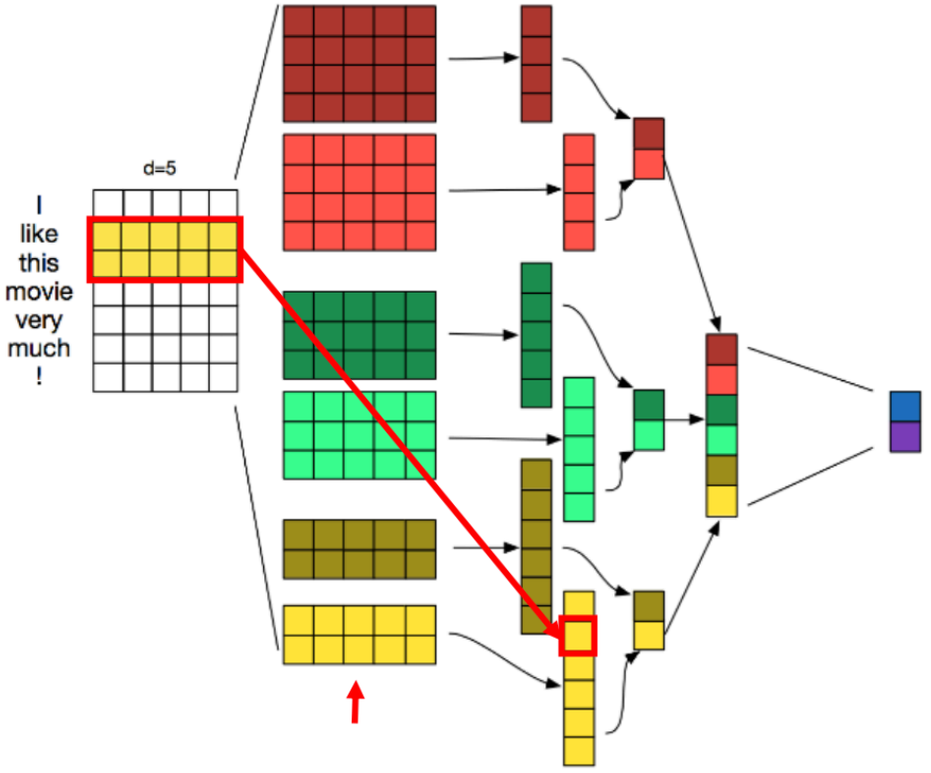



**Convolution**

**Bigram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

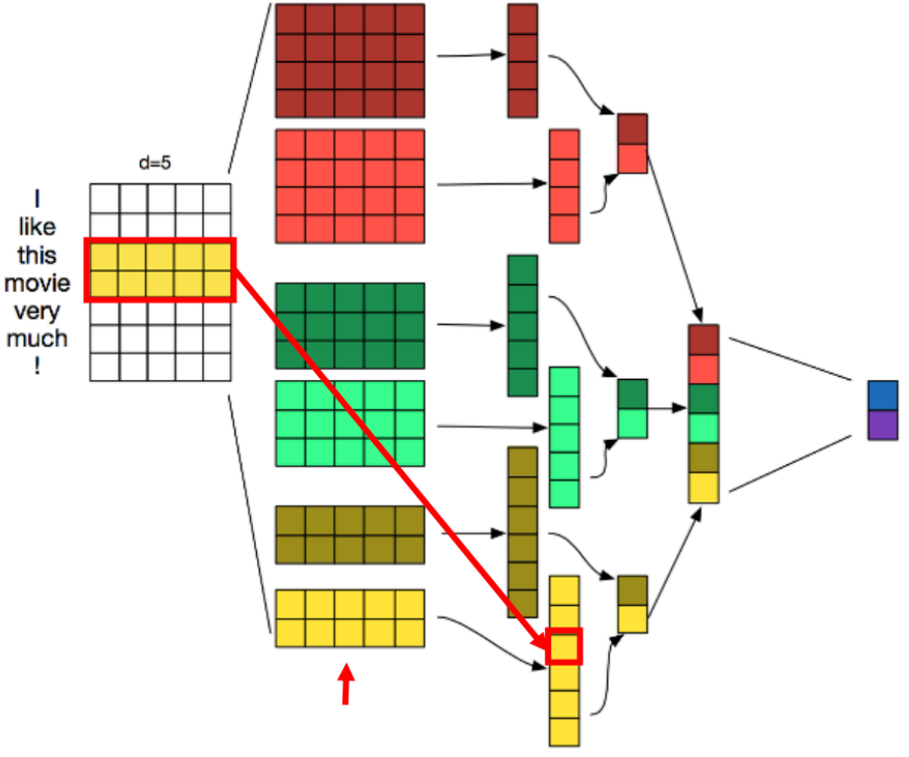



**Convolution**

**Bigram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

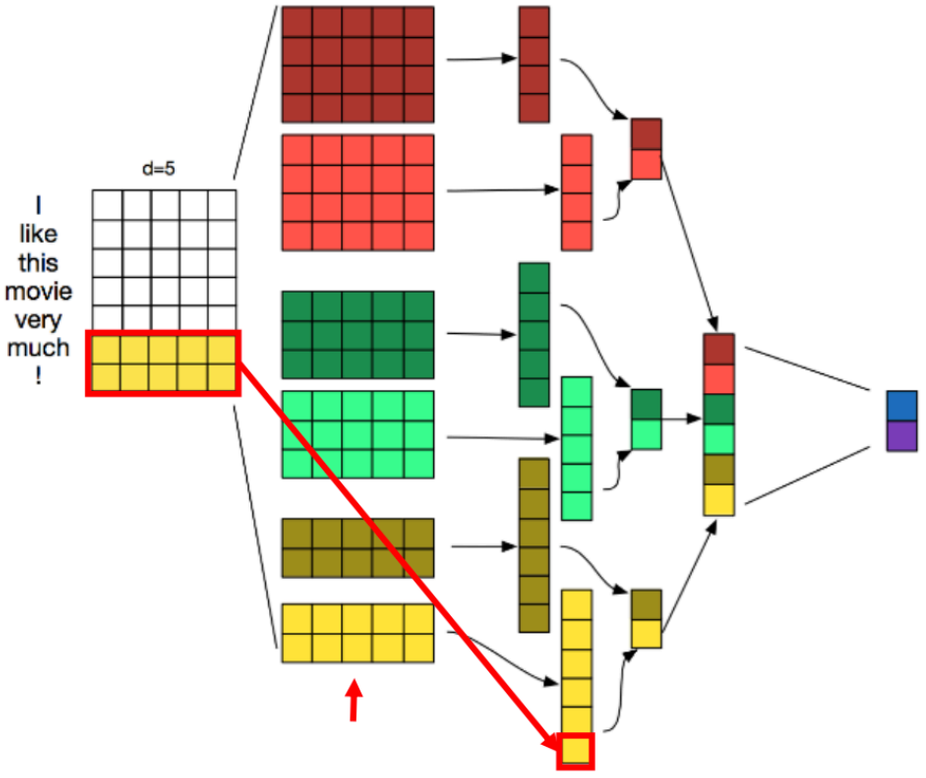



**Convolution**

**Bigram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

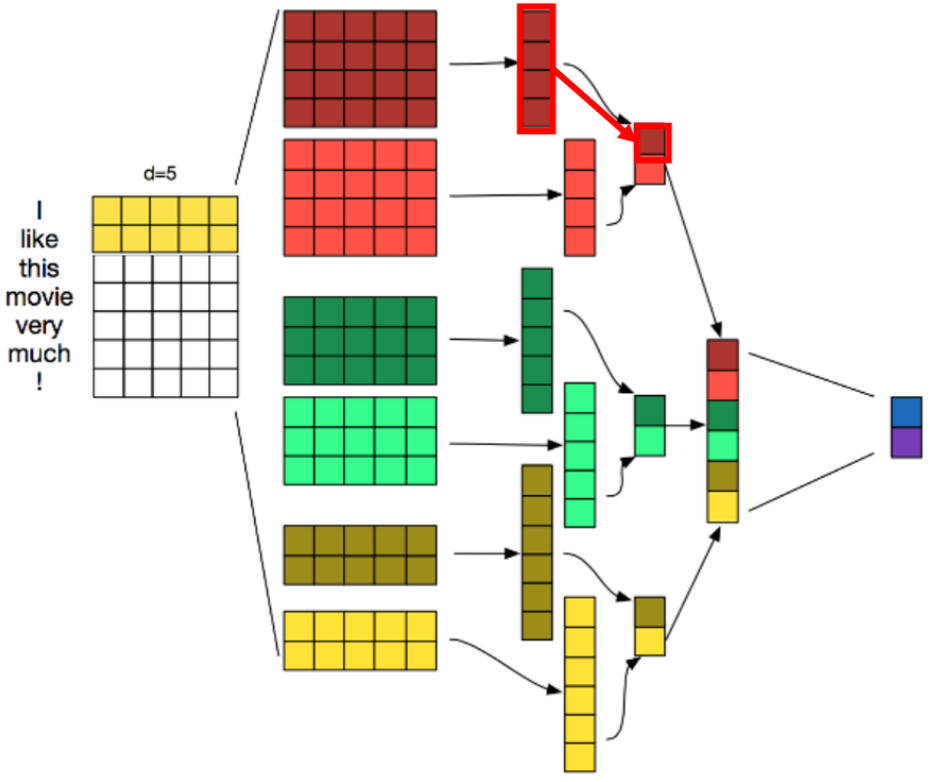



**Convolution**

**Bigram feature를 추출**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet

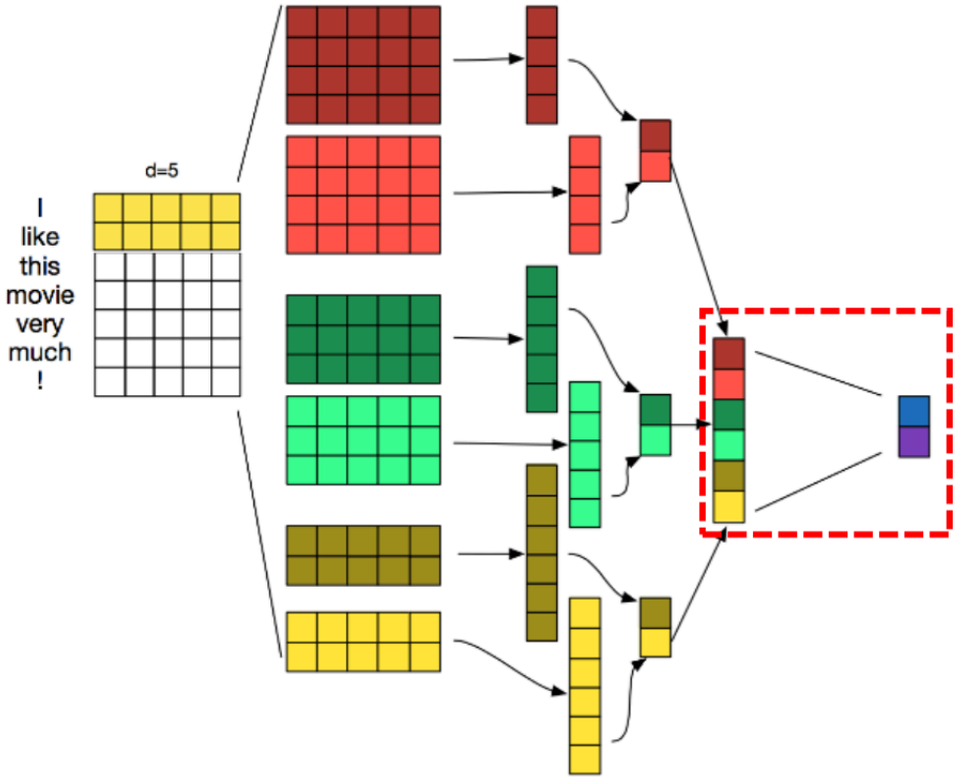



**Max Pooling**

**max-over-time**

# 문장을 하나의 벡터로 표현한다면?

## NLP ConvNet



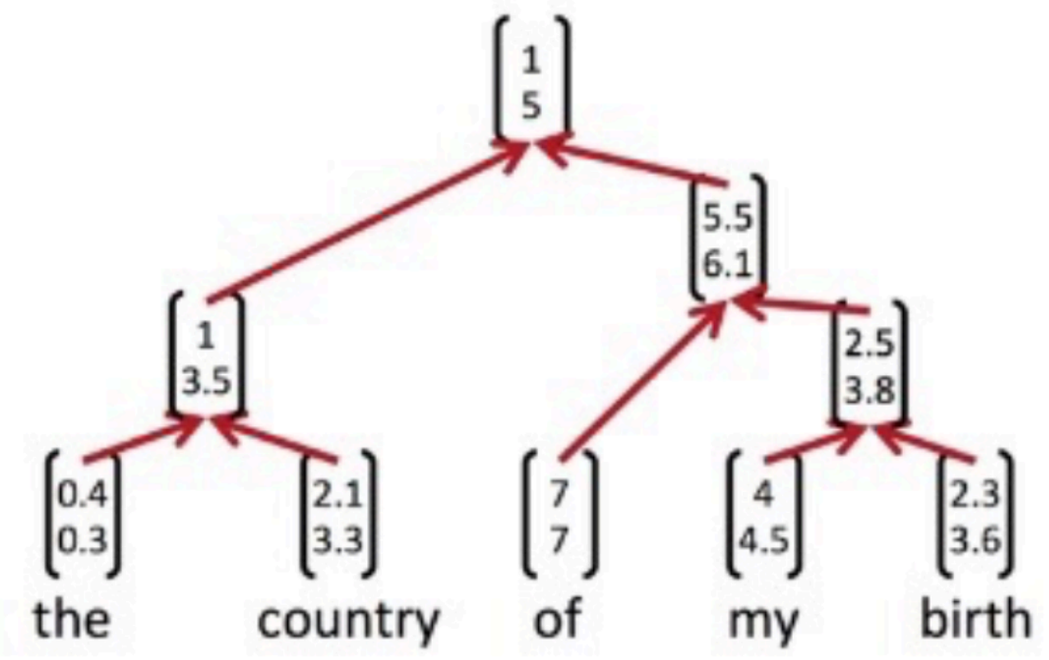

**Output layer**

**Classification (e.g. binary classification for pos/neg)**

## 문장을 하나의 벡터로 표현한다면?

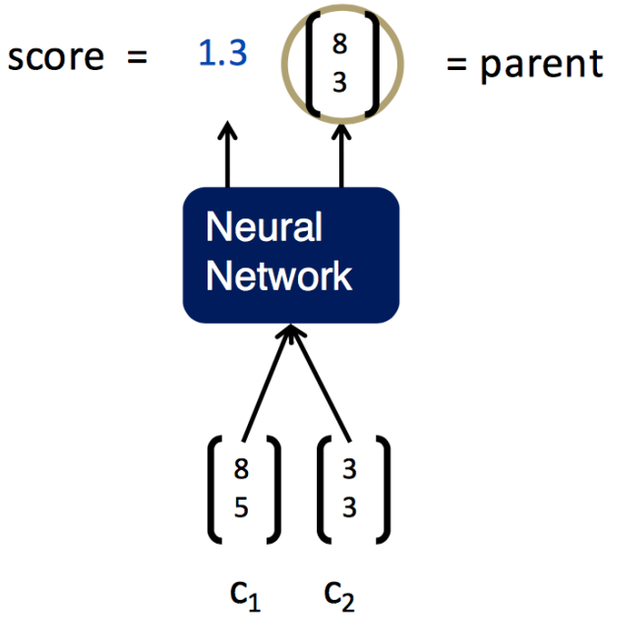
Neural Network를 활용한 Sentence Embedding

- Recursive Neural Network



# 문장을 하나의 벡터로 표현한다면?

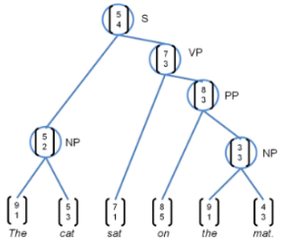
## Recursive Neural Network Definition




score = $U^T p$

$$p = \tanh\left(W\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + b\right),$$

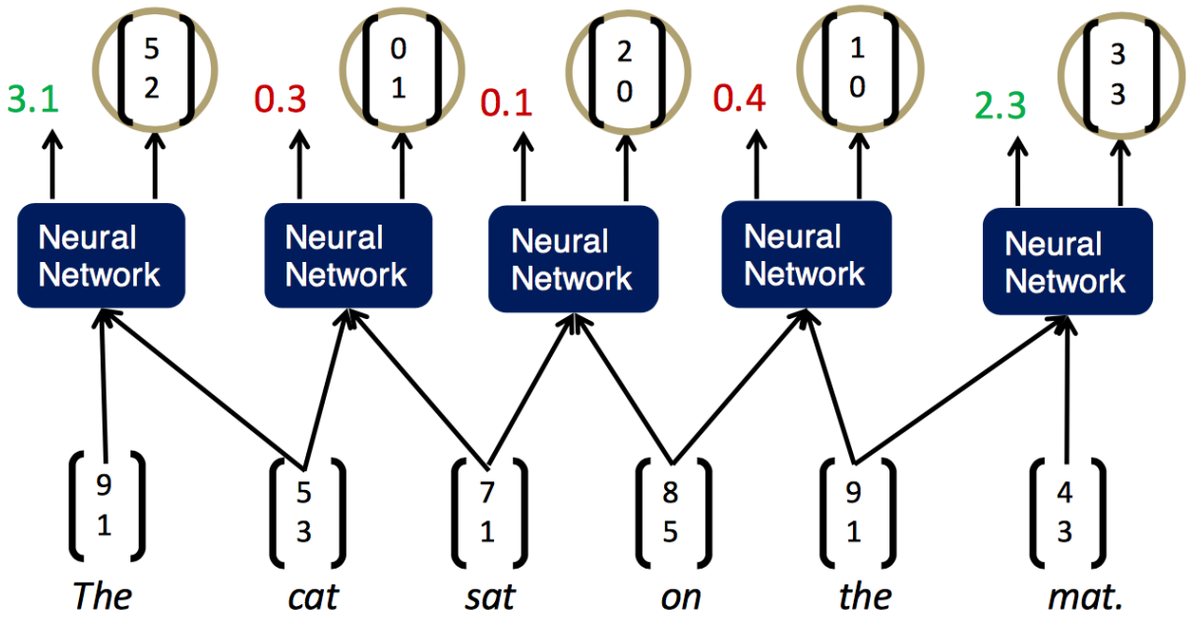
**Same** $W$ parameters at all nodes of the tree




21

# 문장을 하나의 벡터로 표현한다면?

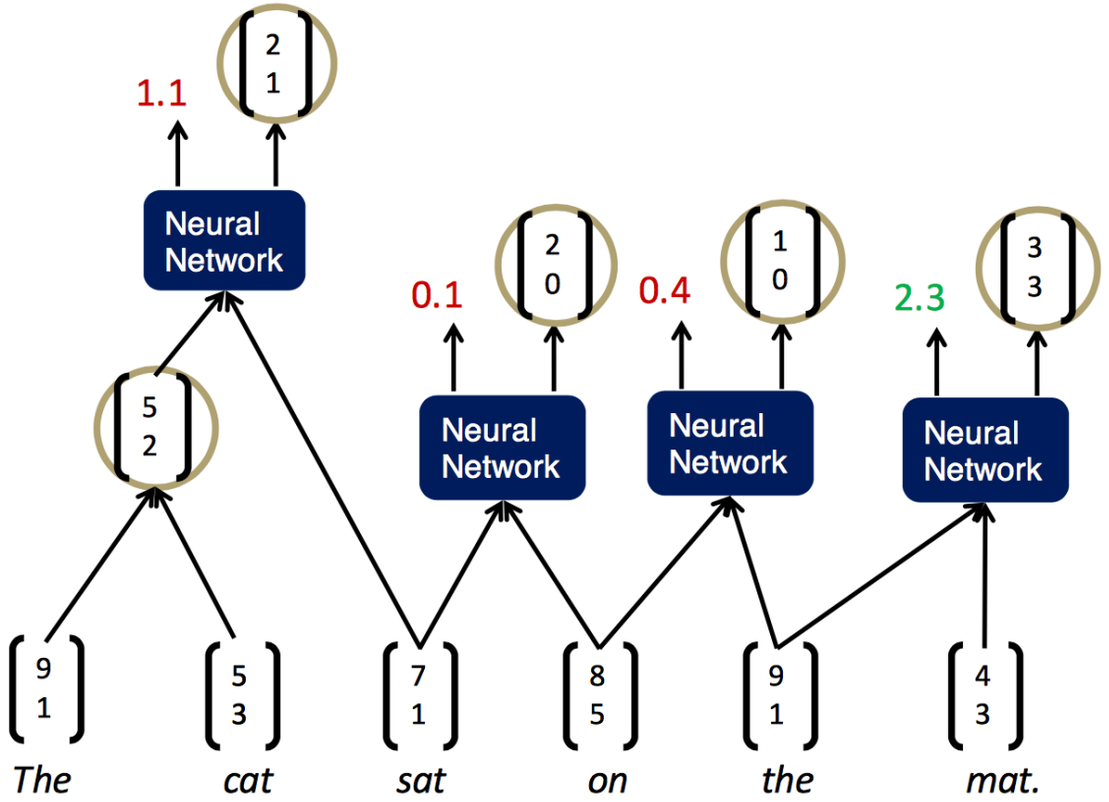
## Parsing a sentence with an RNN




22

# 문장을 하나의 벡터로 표현한다면?

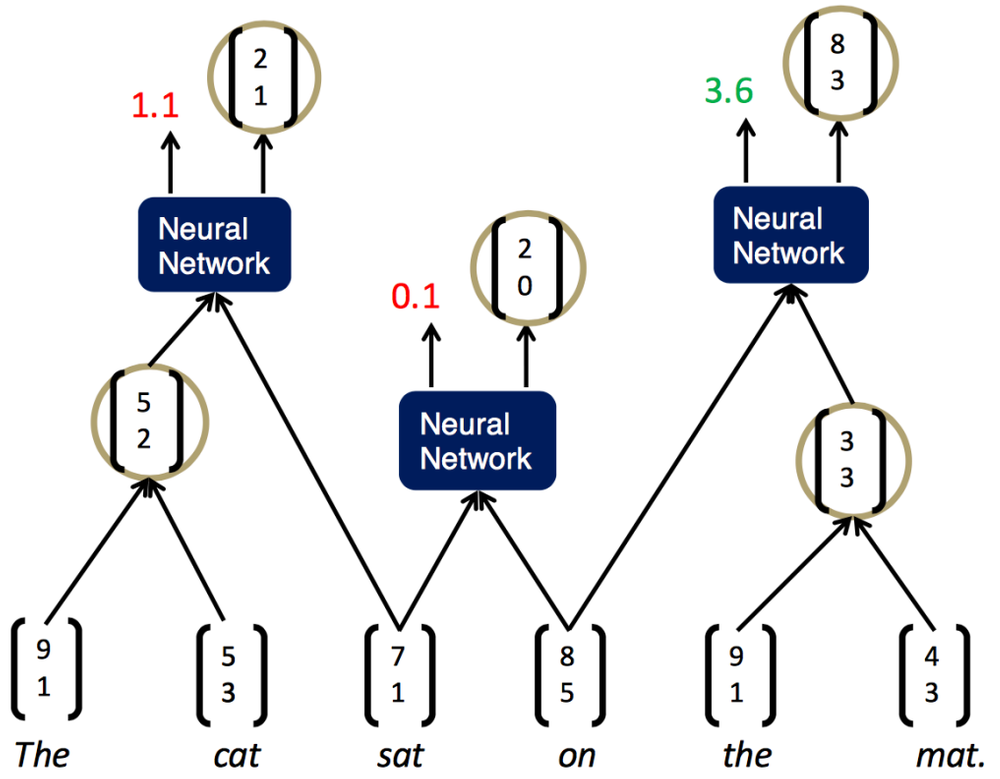
## Parsing a sentence




23

# 문장을 하나의 벡터로 표현한다면?

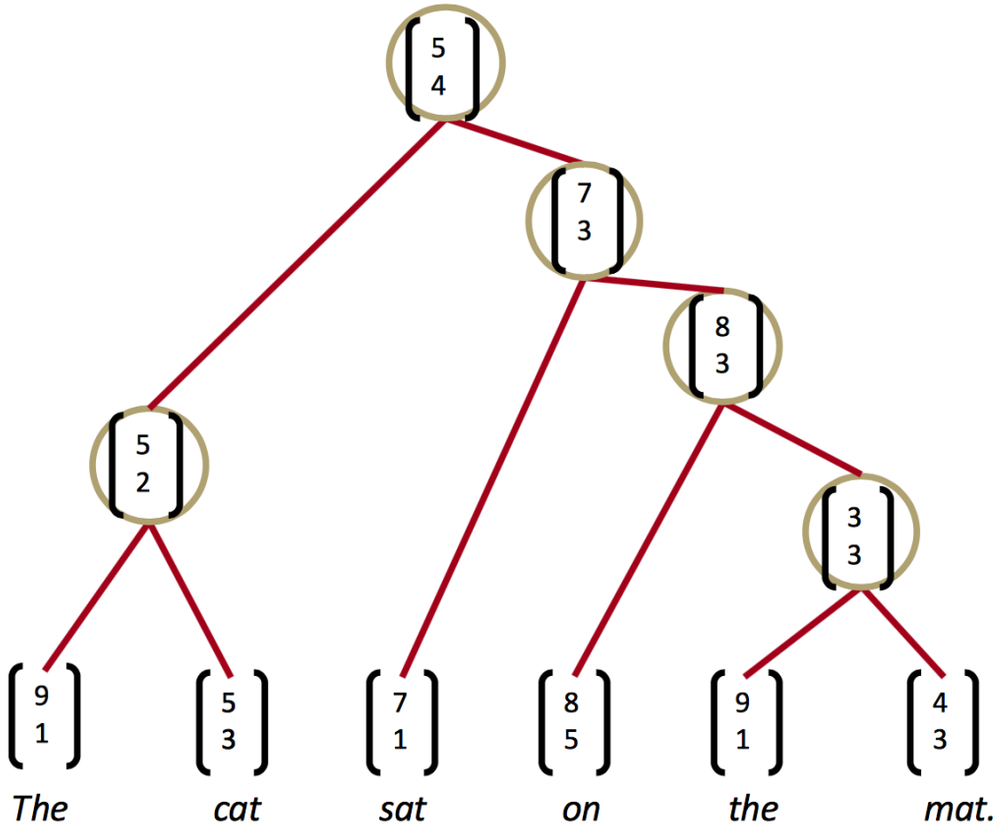
## Parsing a sentence




24

# 문장을 하나의 벡터로 표현한다면?

## Parsing a sentence

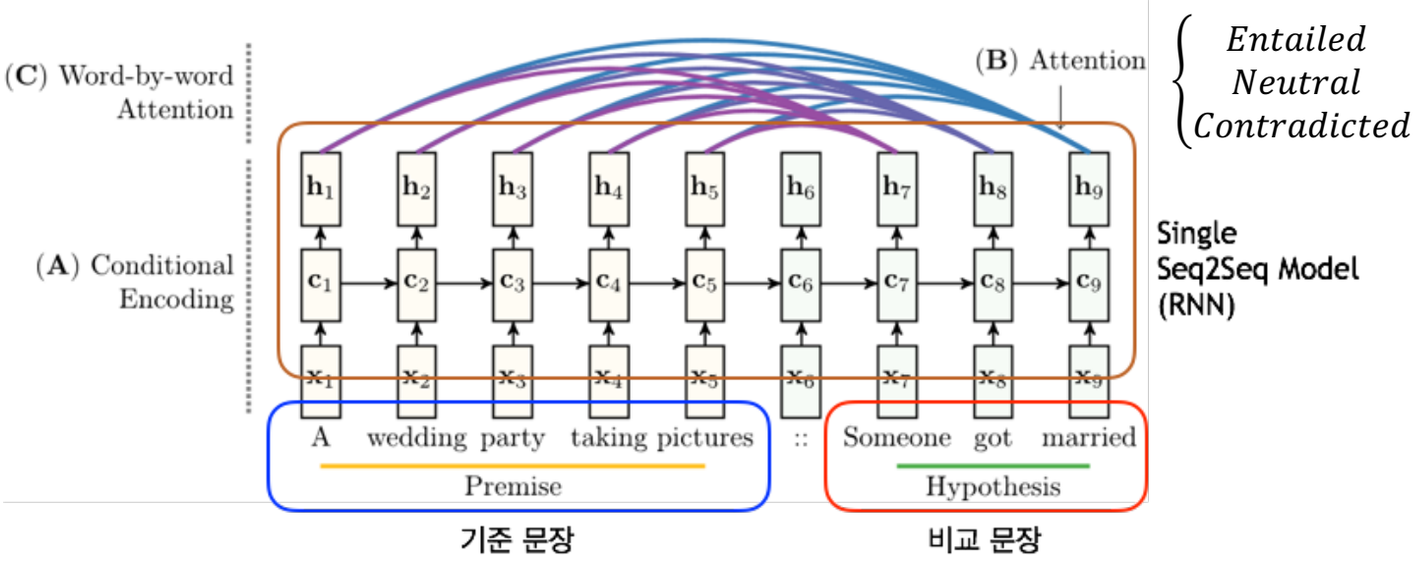



25

# Sentence Embedding의 활용

Sentence Embedding의 활용 ≈ Classification

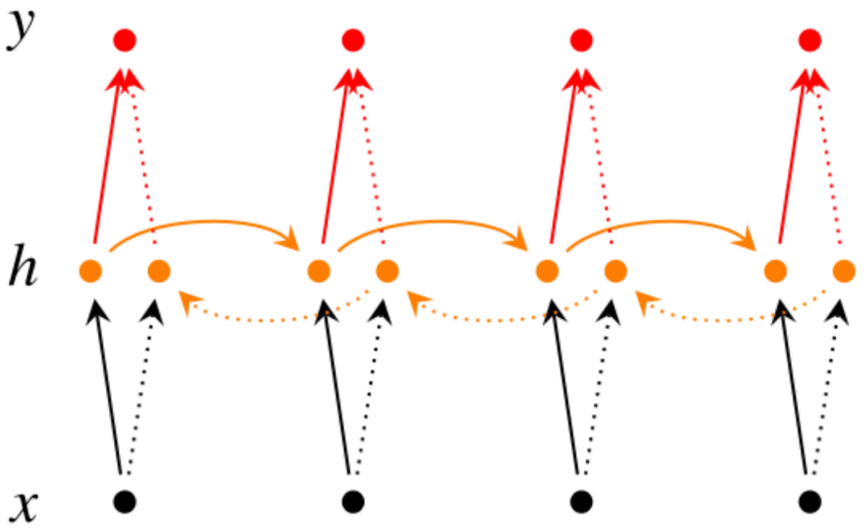
- Sentiment Analysis
- Document Similarity
- Sentence Entailment




Sentence Entailment 모델

# Bidirectional RNN

- Problem: For classification you want to incorporate information from words both preceding and following.




$$\vec{h}_t = f(\vec{W}x_t + \vec{V}\vec{h}_{t-1} + \vec{b})$$

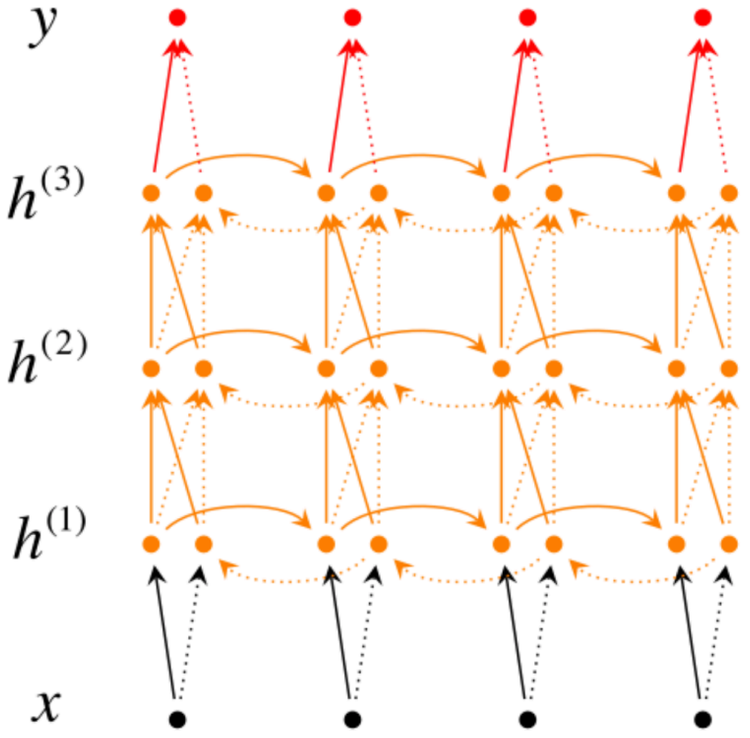
$$\overleftarrow{h}_t = f(\overleftarrow{W}x_t + \overleftarrow{V}\overleftarrow{h}_{t+1} + \overleftarrow{b})$$

$$y_t = g(U[\vec{h}_t; \overleftarrow{h}_t] + c)$$

$h = [\vec{h}; \overleftarrow{h}]$ now represents (summarizes) the past and future around a single token.

Bidirectional RNN

# Deep Bidirectional RNN




$$\overrightarrow{h}_t^{(i)} = f(\overrightarrow{W}^{(i)} h_t^{(i-1)} + \overrightarrow{V}^{(i)} \overrightarrow{h}_{t-1}^{(i)} + \overrightarrow{b}^{(i)})$$

$$\overleftarrow{h}_t^{(i)} = f(\overleftarrow{W}^{(i)} h_t^{(i-1)} + \overleftarrow{V}^{(i)} \overleftarrow{h}_{t+1}^{(i)} + \overleftarrow{b}^{(i)})$$

$$y_t = g(U[\overrightarrow{h}_t^{(L)}; \overleftarrow{h}_t^{(L)}] + c)$$

Each memory layer passes an intermediate sequential representation to the next.

## References


- Stanford CS224n Lectures
  http://web.stanford.edu/class/cs224n/syllabus.html
- Stanford CS231n Lectures
  http://cs231n.stanford.edu/2016/syllabus.html
- "Understanding LSTM Networks" by Colah's Blog
  http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- "Sentence Embedding" by Nishant Nikhil
  https://towardsdatascience.com/sentence-embedding-3053db22ea77

# 감사합니다