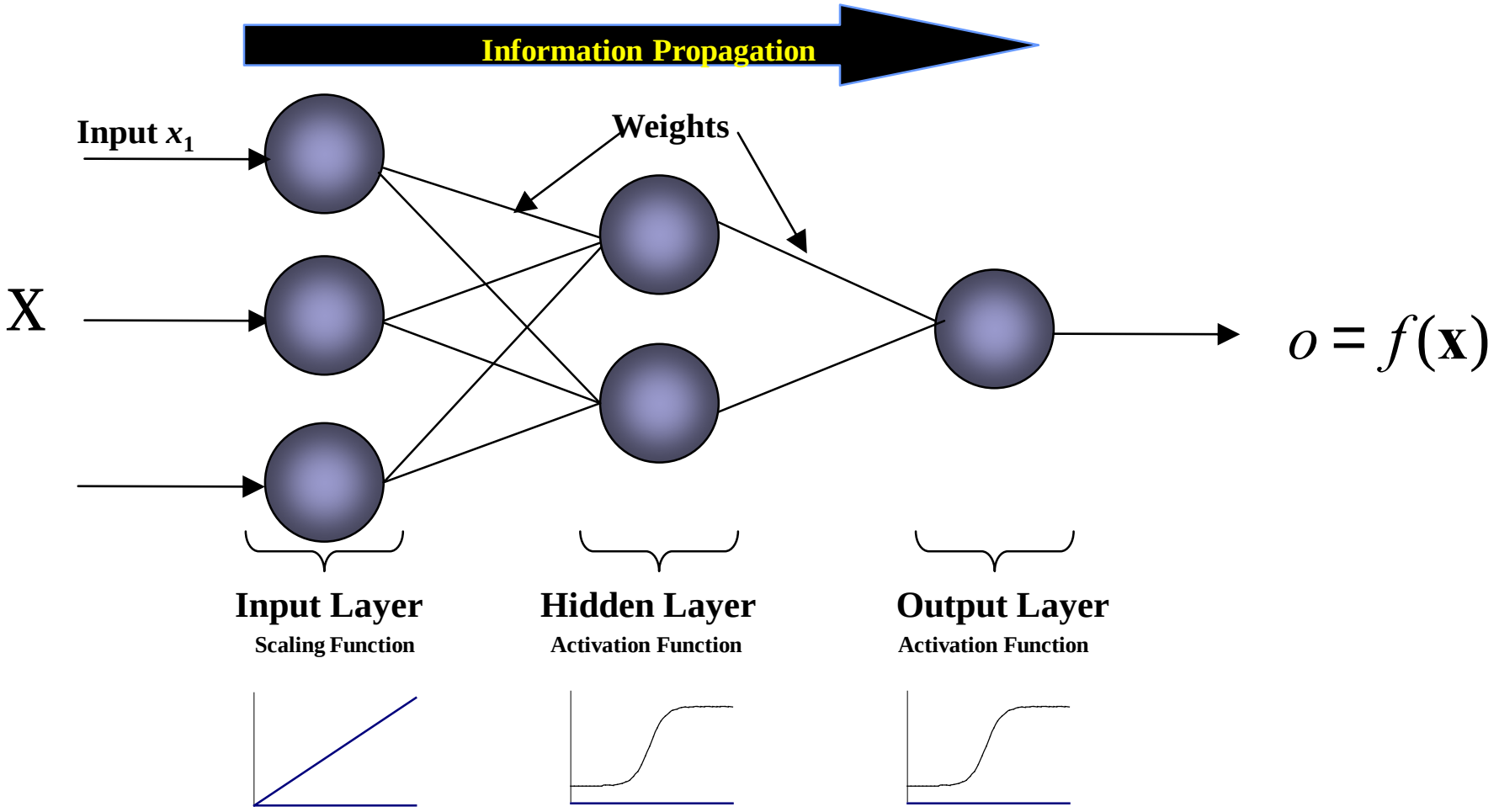
IoT 인공지능 빅데이터 개론 및 실습
2019.05.17

Recurrent Deep Neural Network의 개념과 텍스트 분석에의 응용

Kyomin Jung
Department of Electrical and Computer Engineering
Seoul National University

Neural Networks



Recurrent Neural Networks

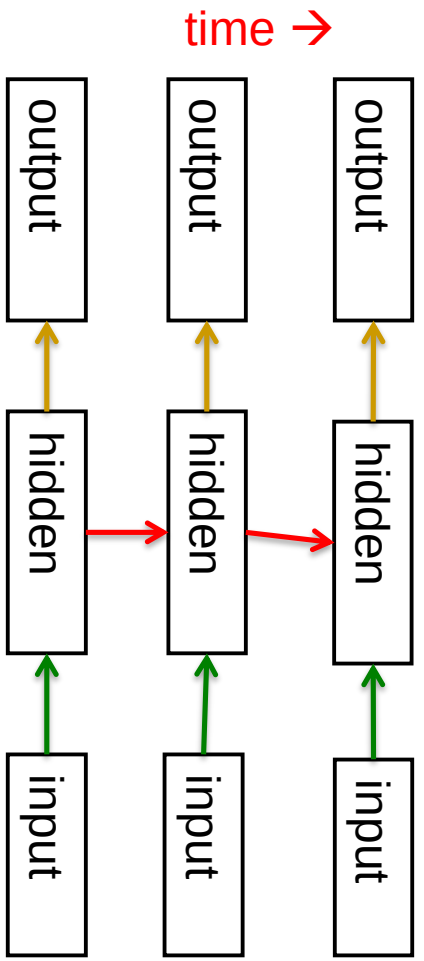
- Feed forward neural networks (including CNN)
  - Information only flows one way
  - One input pattern produces (same) one output
  - No sense of time (or memory of previous state)

- Recurrence
  - Nodes are allowed to connect back to a previous nodes and/or to themselves
  - Sense of time and memory

- Biological nervous systems show high levels of recurrence

Recurrent Neural Networks for Modeling Sequences

- Recurrent neural networks are a very natural way to model sequential data:
  - They are equivalent to very deep nets with one hidden layer per time step.
  - Except that they use the same weights at every time step and they get input at every time step.
- They have the ability to remember information in their hidden states for a long time.



Basic RNN formulation

- **Input sequence :** $x = (x_1, \dots, x_T)$
- **Hidden vector sequence :** $h = (h_1, \dots, h_T)$
- **Output vector sequence :** $y = (y_1, \dots, y_T)$
- **Hidden vector update function :**
  $h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$
- **Output vector update function :**
  $y_t = W_{hy}h_t + b_y$

Applications of RNN

- There are many applications that require learning a temporal sequence of events via RNN
  - Sequence Recognition: Produce a labelling for a given input sequence. Ex : speech recognition
  - Sequence Generation: Generate the rest of a sequence when the neural network sees only part of the sequence. Ex: Time series prediction, text completion, music generation

An Example

- Ilya Sutskever (2011) trained a special type of recurrent neural net (a multiplicative RNN) to predict the next character in a text sequence [1].

- After training for a long time on a string of half a billion characters from English Wikipedia, he got it to generate new text.
  - It generates by predicting the probability distribution for the next character and then sampling a character from this distribution.

[1] Ilya Sutskever, et al. **Generating** text with **recurrent** neural networks, ICML 2011.

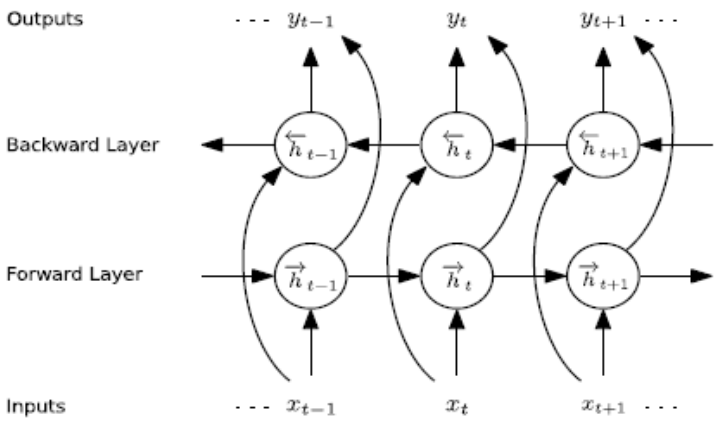
Some Text Generated One Character at a Time by Ilya Sutskever's RNN

In 1974 Northern Denver had been overshadowed by CNL, and several Irish intelligence agencies in the Mediterranean region. However, on the Victoria, Kings Hebrew stated that Charles decided to escape during an alliance. The mansion house was completed in 1882, the second in its bridge are omitted, while closing is the proton reticulum composed below it aims, such that it is the blurring of appearing on any well-paid type of box printer.

Demo available at http://www.cs.toronto.edu/~ilya/rnn.html

Bidirectional Recurrent Neural Networks

- Traditional RNNs only model the dependence of the current state on the previous state.
- BRNN extends to model dependence on both past states and future states.
- For example: to predict a missing word in a sequence, look at both the left and the right context.



An BRNN

$$\vec{h}_t = f(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}})$$
$$\overleftarrow{h}_t = f(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}})$$

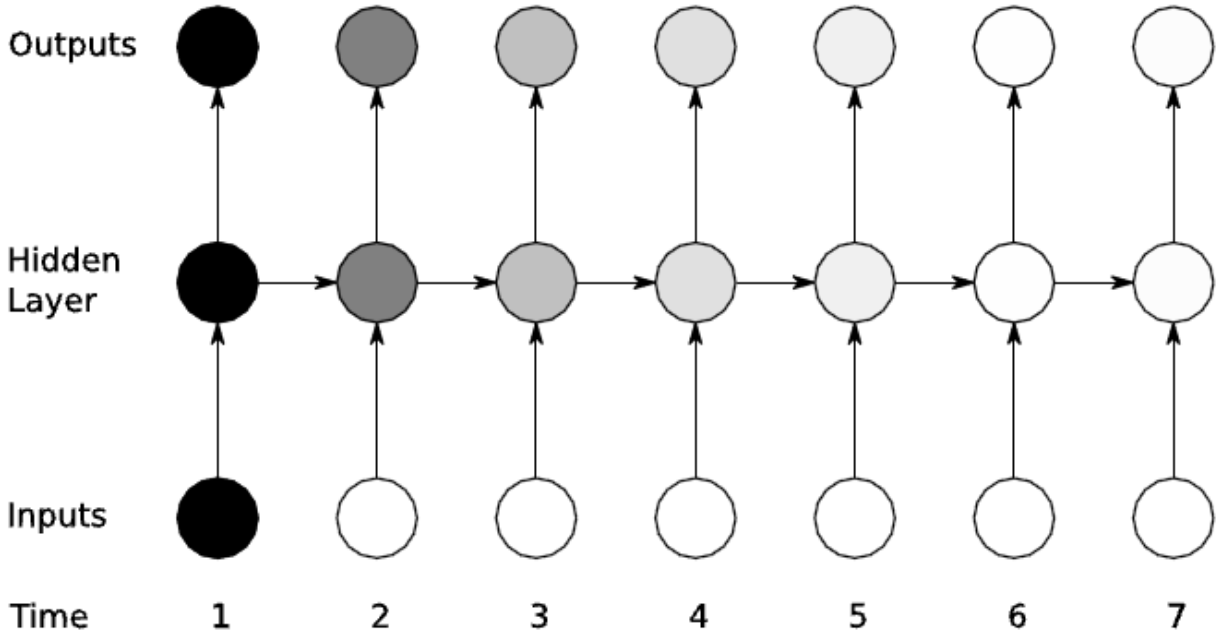
training sequence forwards and backwards to two separate recurrent hidden layers

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y$$

past and future context determines the output

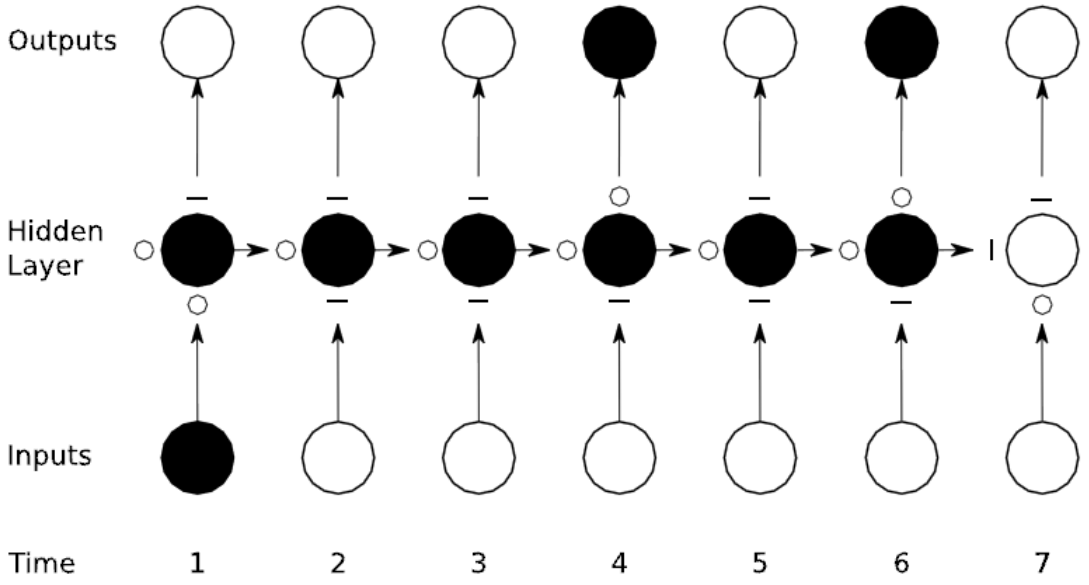
Vanishing Gradients

- Vanishing Gradients problem for RNNs
  - Influence of the inputs at time t decreases and vanishes over time



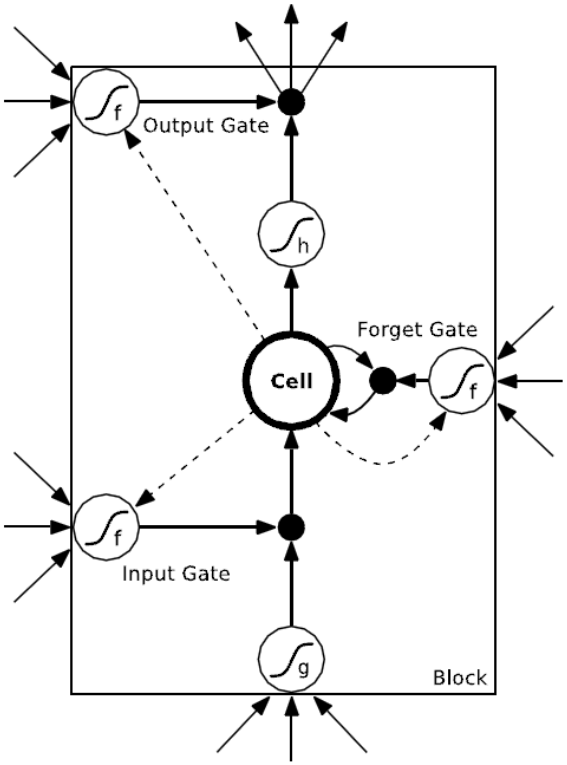
LSTM-RNNs

- LSTM can preserve gradient information
  - Hidden layer units formed with Long Short-Term Memory (LSTM) cells can store and access information over long periods of time



Long Short-Term Memory

- LSTM block architecture
  - 3 gates
    - Input gate adjust the influence from input to cell
    - Forget gate adjust the influence from cell to cell over time
    - Output gate adjust the influence from cell to output

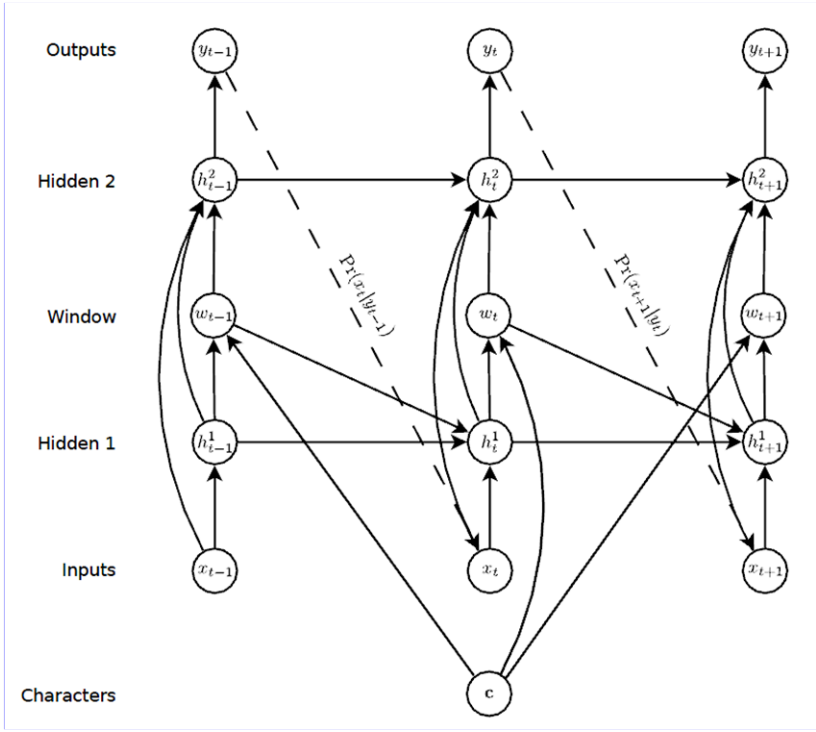


Applications – Text sequence generation

*experiment done by Alex Graves et al.

- Demo available at
  - http://www.cs.toronto.edu/~graves/handwriting.html

- On IAM-OnDB
  - Using character-level transcriptions
  - 57 distinct characters
  - 3 layers of 400 LSTM blocks each

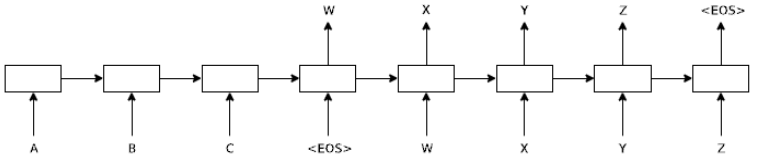


Applications – Machine translation

*experiment done by Ilya Sutskever et al.

- On WMT'14 English to French dataset
  - 12M sentences with 348M French words and 304M English words
  - 4 layers of 1000 LSTM blocks each

Type	Sentence
Our model	Ulrich UNK , membre du conseil d' administration du constructeur automobile Audi , affirme qu' il s' agit d' une pratique courante depuis des années pour que les téléphones portables puissent être collectés avant les réunions du conseil d' administration afin qu' ils ne soient pas utilisés comme appareils d' écoute à distance .
Truth	Ulrich Hackenberg , membre du conseil d' administration du constructeur automobile Audi , déclare que la collecte des téléphones portables avant les réunions du conseil , afin qu' ils ne puissent pas être utilisés comme appareils d' écoute à distance , est une pratique courante depuis des années .
Our model	" Les téléphones cellulaires , qui sont vraiment une question , non seulement parce qu' ils pourraient potentiellement causer des interférences avec les appareils de navigation , mais nous savons , selon la FCC , qu' ils pourraient interférer avec les tours de téléphone cellulaire lorsqu' ils sont dans l' air " , dit UNK .
Truth	" Les téléphones portables sont véritablement un problème , non seulement parce qu' ils pourraient éventuellement créer des interférences avec les instruments de navigation , mais parce que nous savons , d' après la FCC , qu' ils pourraient perturber les antennes-relais de téléphonie mobile s' ils sont utilisés à bord " , a déclaré Rosenker .
Our model	Avec la crémation , il y a un " sentiment de violence contre le corps d' un être cher " , qui sera " réduit à une pile de cendres " en très peu de temps au lieu d' un processus de décomposition " qui accompagnera les étapes du deuil " .
Truth	Il y a , avec la crémation , " une violence faite au corps aimé " , qui va être " réduit à un tas de cendres " en très peu de temps , et non après un processus de décomposition , qui " accompagnerait les phases du deuil " .



Applications – Speech Recognition

- On TIMIT database
  - Audio data – phoneme classification

```
3050 5723 she
5723 10337 had
9190 11517 your
11517 16334 dark
16334 21199 suit
21199 22560 in
22560 28064 greasy
28064 33360 wash
33754 37556 water
37556 40313 all
40313 44586 year
```

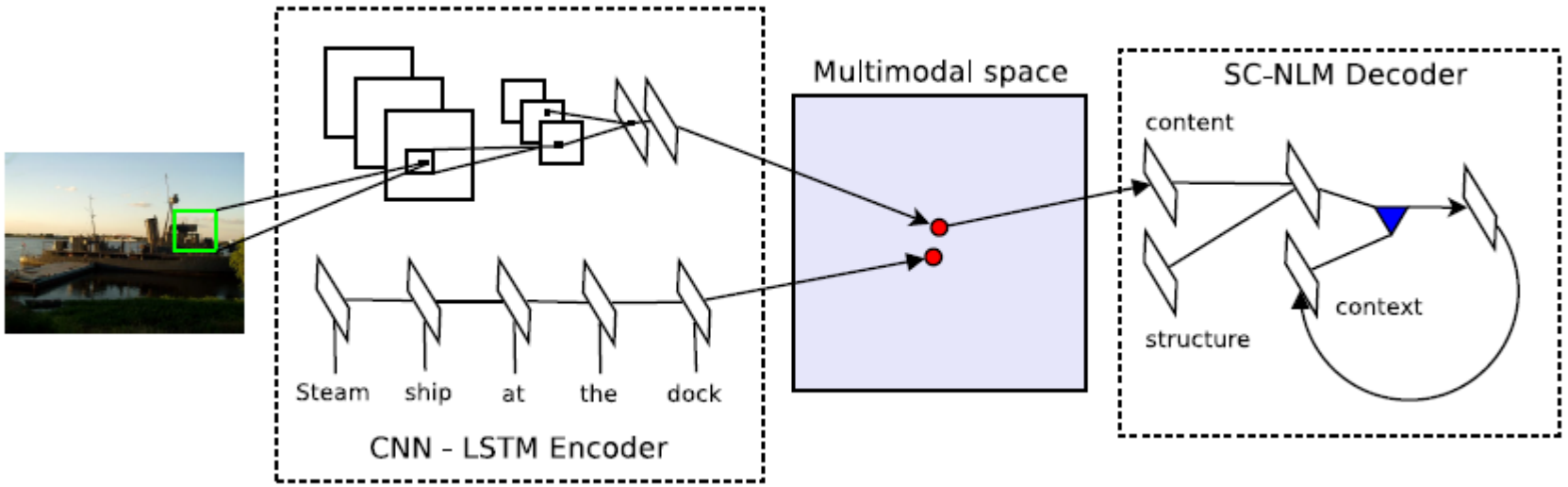
- 3 layers with 250 hidden LSTM block each
- Beats HMM (Hidden Markov Model) based models

Applications – Image to text

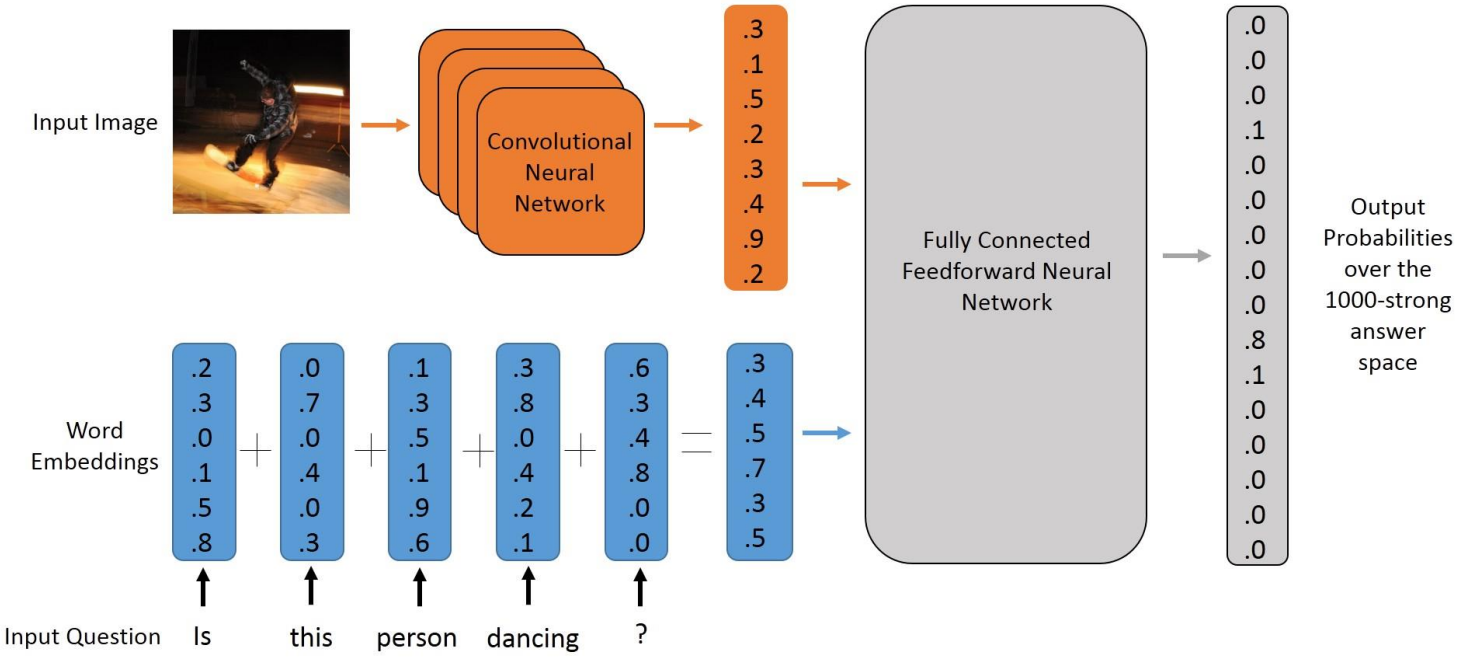
- Generate or predict a text sequence from a given image
  - image embedding is done by a CNN
  - word embedding is done by an LSTM-RNN
  - Decoding to a text sequence is done by another LSTM-RNN
    - Content vector is multimodal vector in embedded space

Applications – Image to text

- Demo available at
  - http://www.cs.toronto.edu/~rkiros/lstm_scnlm.html



Applications – Visual Question Answering



The output is conditioned on both image and text inputs. A CNN is used to encode the image and a RNN is used to encode the sentence.