PhD in Computer Science and Engineering
Bologna, April 2016

Machine Learning

Marco Lippi

marco.lippi3@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Recurrent Neural Networks

Material and references

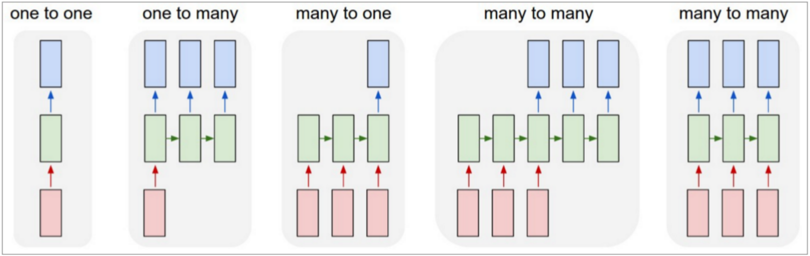
Many figures are taken from

Cristopher Olah's tutorial
(`colah.github.io`)

and

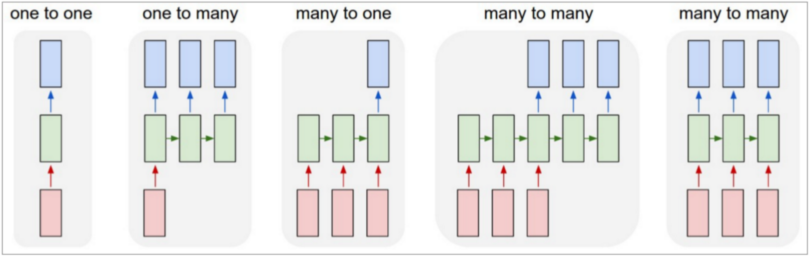
Alex Karpathy's post
"The unreasonable effectiveness of Recurrent Neural Networks"

In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



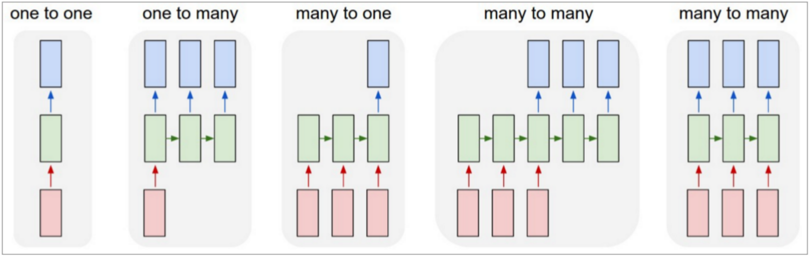
Processing sequences

In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



1. Standard tasks such as image classification

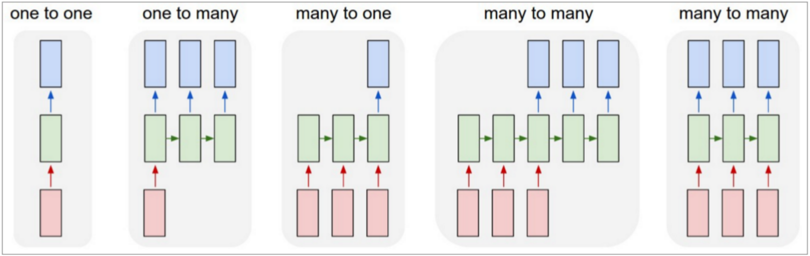
In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



2. Fixed input, sequence output: e.g., image captioning

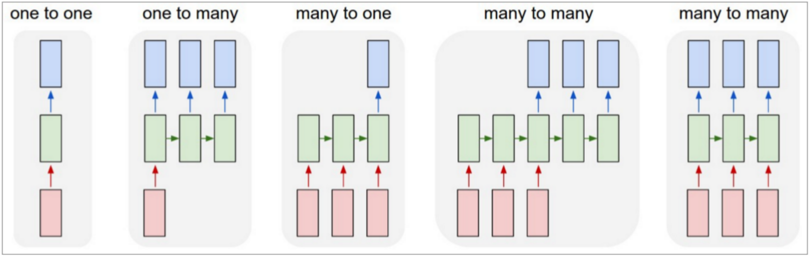
Processing sequences

In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



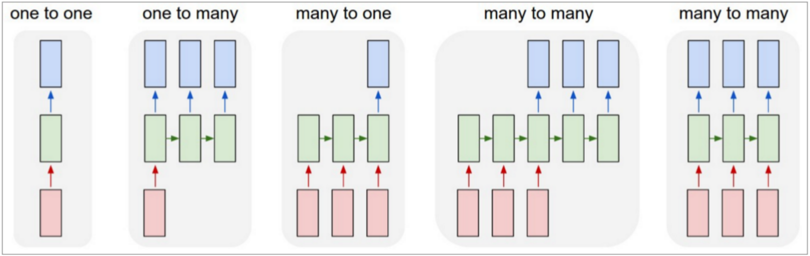
3. Sequence classification: e.g., sentiment classification

In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



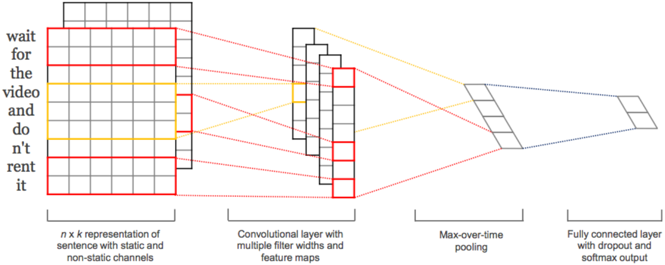
4. Sequence input/output: e.g., machine translation

In many tasks we should deal with **variable-size input** (text, audio, video) while most algorithms can deal with just **fixed-size** input



5. Synced sequence input/output: e.g., video frame classification

One possibility: use **convolutional neural networks** with **max pooling** over all feature maps

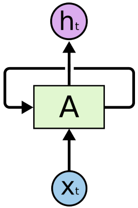


[Figure by Kim, 2014]

Problem: **transalation invariance**. . .

In a **recurrent neural network** the hidden state at time t depends

- on the input
- on the hidden state at time $t-1$ (**memory**)

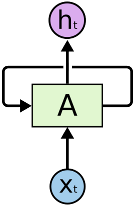


$$y_t = f_V(h_t)$$

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

We can parametrize:

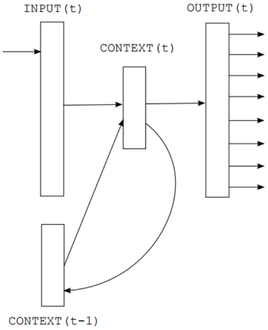
- input-hidden connections
- hidden-hidden connections
- hidden-output connections



$$y_t = W_{hy} h_t$$

$$h_t = tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

Recurrent Neural Network Language Model



[Figure by Mikolov et al., 2010]

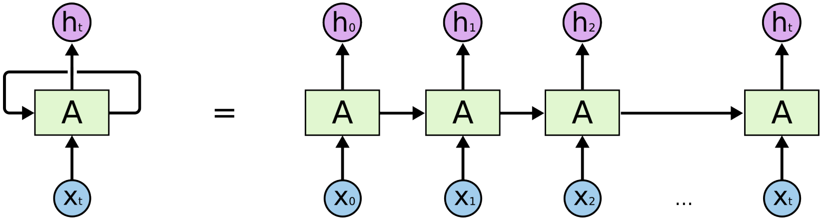
$$x(t) = w(t) + s(t-1)$$

$$s_j(t) = f(\sum_i x_i(t) u_{ji})$$

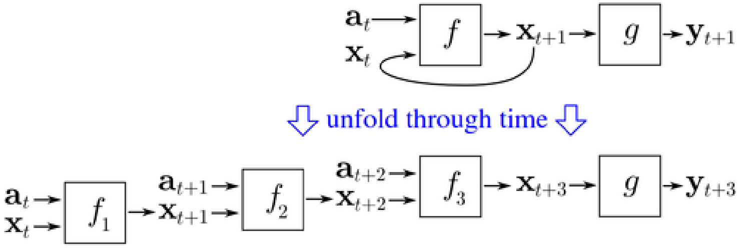
$$y_k(t) = g(\sum_j s_j(t) v_{kj})$$

Wrt NNLM, **no need** to specify context dimension in advance !

A classic RNN can be **unrolled** through time, so that the looping connections are made explicit



Backpropagation Through Time (BPTT)

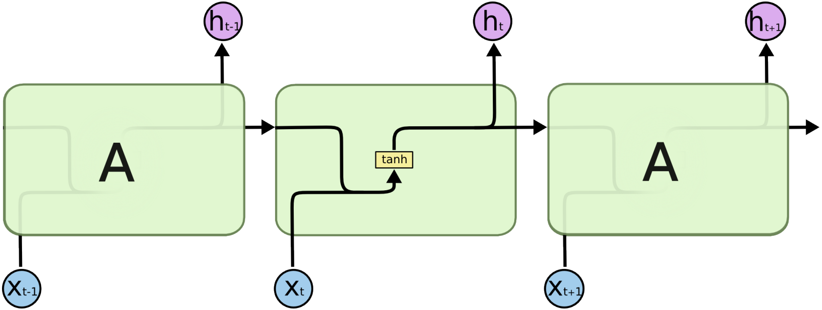


[Figure from Wikipedia]

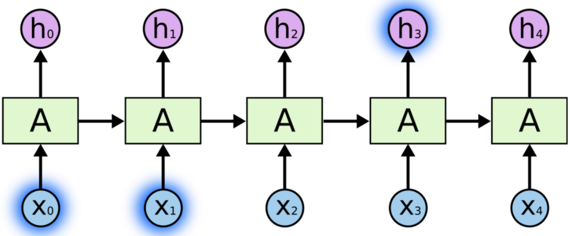
BPTT drawbacks:

- decide the value k for unfolding
- exploding or vanishing gradients
- exploding could be controlled with **gradient clipping**
- vanishing has to be faced with different models (LSTM)

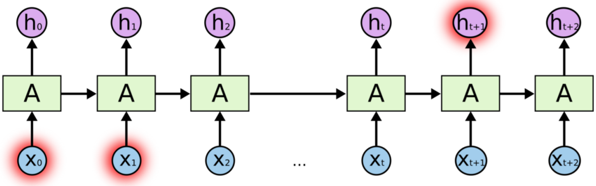
Recurrent Neural Networks (RNNs)



Some short- or mid-term dependencies can be afforded...



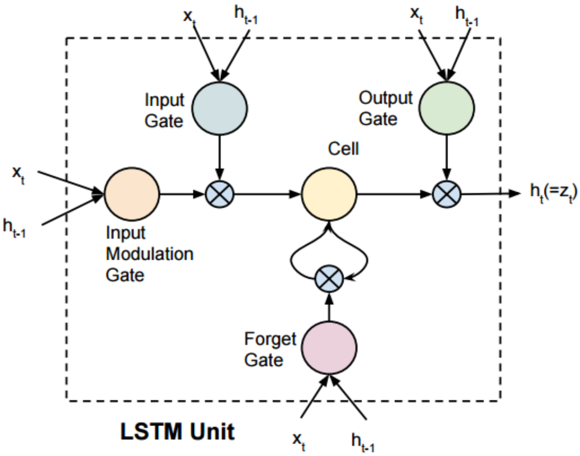
But the model fails in learning long-term dependencies !



The main problem is **vanishing gradients** in BPTT...

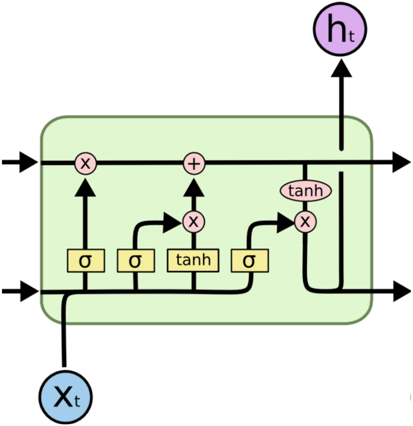
Long Short-Term Memory Networks (LSTMs)

An LSTM is basically an RNN with a different computational block
LSTMs were designed by Hochreiter & Schmidhuber in 1997 !



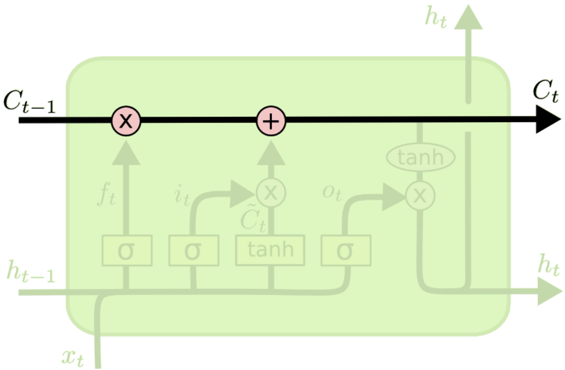
Long Short-Term Memory Networks (LSTMs)

An LSTM block features "a **memory cell** which can maintain its state over time, and **non-linear gating units** which regulate the information flow into and out of the cell" [Greff et al., 2015]



Cell state

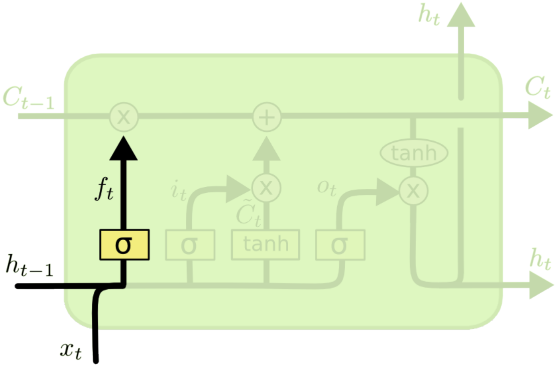
Just let the information go through the network
Other gates can **optionally** let information through



Forget gate

Sigmoid layer that produces weights for the state cell C_{t-1}
Decides what to **keep** (1) or **forget** (0) of past cell state

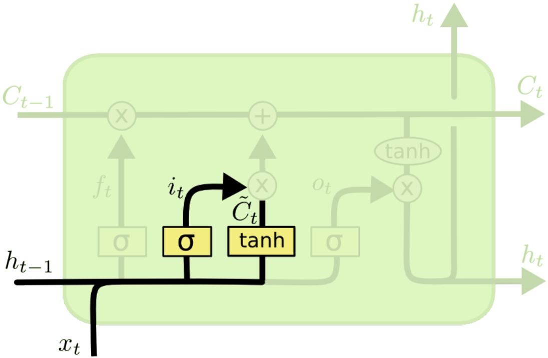
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$



Input gate

Allows novel information be used to update state cell C_{t-1}

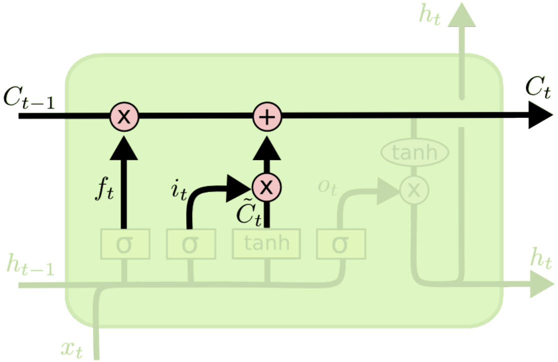
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$



Cell state update

Combine **old state** (after **forgetting**) with **novel input**

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

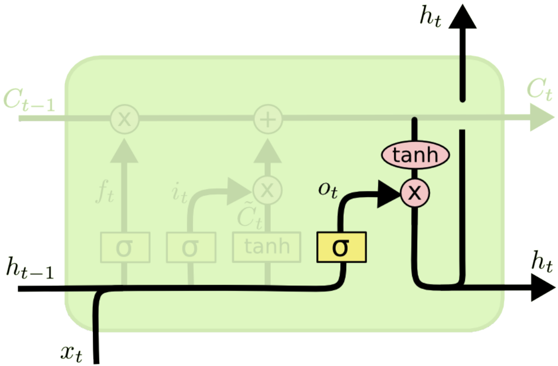


Output gate

Build the output to be sent to next layer and upper block

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

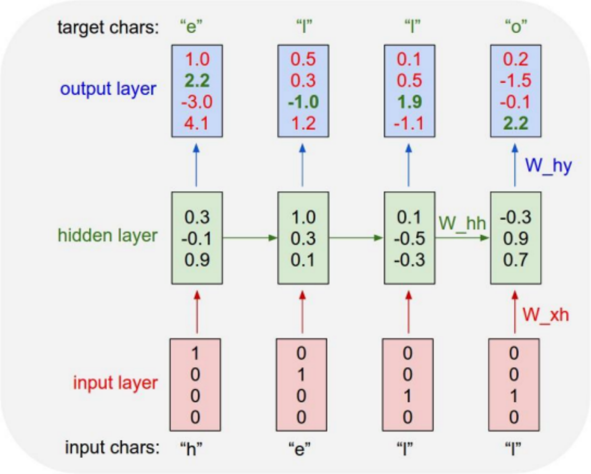
$$h_t = o_t \odot tanh(C_t)$$



Putting all together

- $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
- $o_t = \sigma(W_o[h_{t-1}, x_t] + b_O)$
- $h_t = o_t \odot tanh(C_t)$
- $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$

Example: character-level language model



[Figure by A. Karpathy]

Example: character-level language model

- Pick up a (large) plain text file
- Feed the LSTM with the text
- Predict **next character** given past history
- At prediction time, **sample** from output distribution
- Get LSTM-generated text

at first:

```
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

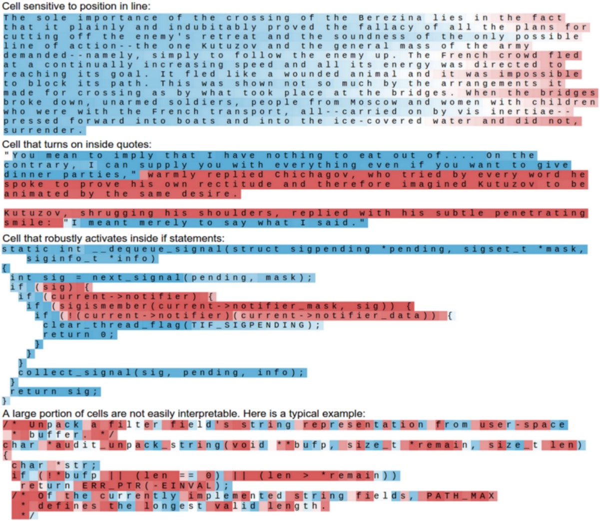
```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

↓ 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.
```

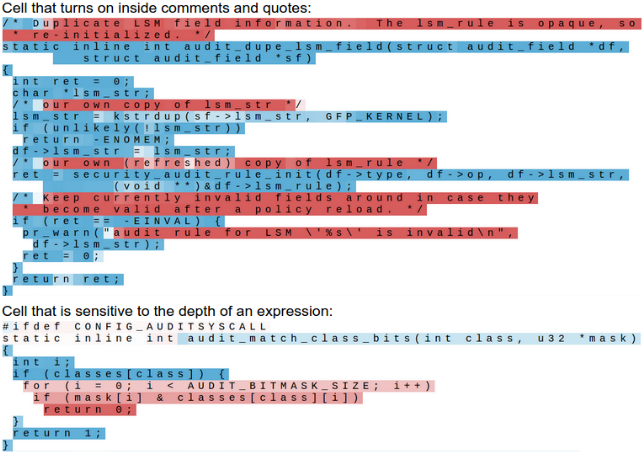
[Figure by A. Karpathy]

Interpretation of cells (blue=off, red=on)



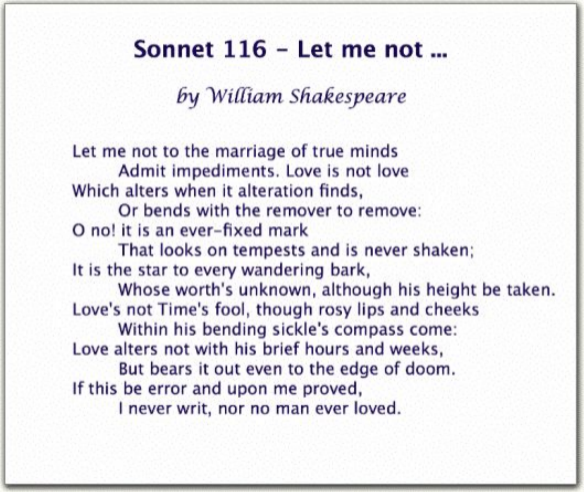
[Figure by A. Karpathy]

Interpretation of cells (blue=off, red=on)



[Figure by A. Karpathy]

Example: training with Shakespeare poems



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.

[Figure by A. Karpathy]

Application: text generation

Example: training with LaTeX math papers

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 $U = \lim_i X_i$. □

[Figure by A. Karpathy]

Example: training with LaTeX math papers

```
\begin{proof}
We may assume that $\mathcal{I}$ is an abelian sheaf on $\mathcal{C}$.
\item Given a morphism $\Delta : \mathcal{F} \to \mathcal{I}$
is an injective and let $\mathfrak q$ be an abelian sheaf on $X$.
Let $\mathcal{F}$ be a fibered complex. Let $\mathcal{F}$ be a category.
\begin{enumerate}
\item \hyperref[setain-construction-phantom]{Lemma}
\label{lemma-characterize-quasi-finite}
Let $\mathcal{F}$ be an abelian quasi-coherent sheaf on $\mathcal{C}$.
Let $\mathcal{F}$ be a coherent $\mathcal{O}_X$-module. Then
$\mathcal{F}$ is an abelian catenary over $\mathcal{C}$.
\item The following are equivalent
\begin{enumerate}
\item $\mathcal{F}$ is an $\mathcal{O}_X$-module.
\end{lemma}
```

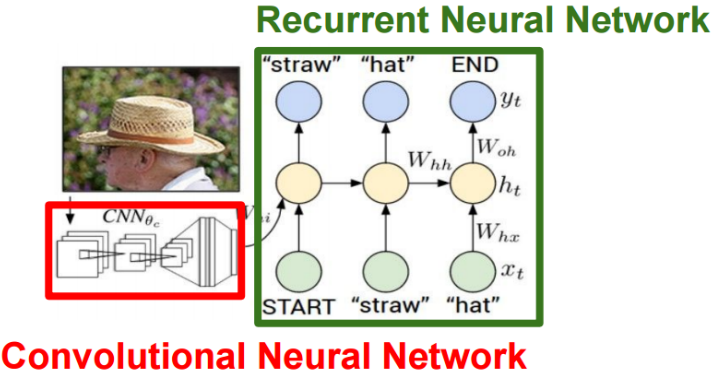
[Figure by A. Karpathy]

Example: training with Linux Kernel (C code)

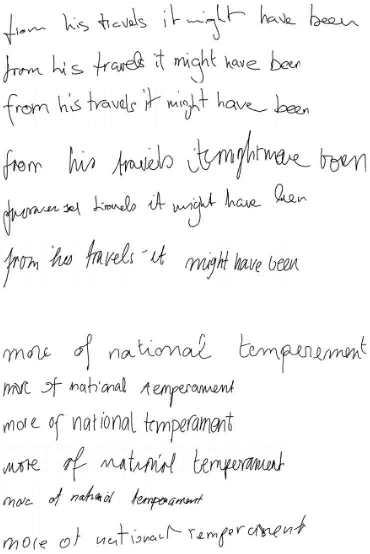
```c
/*
 * If this error is set, we will need anything right after that BSD.
 */
static void action_new_function(struct s_stat_info *wb)
{
  unsigned long flags;
  int lel_idx_bit = e->edd, *sys & ~((unsigned long) *FIRST_COMPAT);
  buf[0] = 0xFFFFFFFF & (bit << 4);
  min(inc, slist->bytes);
  printk(KERN_WARNING "Memory allocated %02x/%02x, "
    "original MLL instead\n"),
    min(min(multi_run - s->len, max) * num_data_in),
    frame_pos, sz + first_seg);
  div_u64_w(val, inb_p);
  spin_unlock(&disk->queue_lock);
  mutex_unlock(&s->sock->mutex);
  mutex_unlock(&func->mutex);
  return disassemble(info->pending_bh);
}

static void num_serial_settings(struct tty_struct *tty)
{
  if (tty == tty)
    disable_single_st_p(dev);
  pci_disable_spool(port);
  return 0;
}
```

[Figure by A. Karpathy]

Application: image captioning



[Figure by A. Karpathy]



[Figure by Graves, 2014]

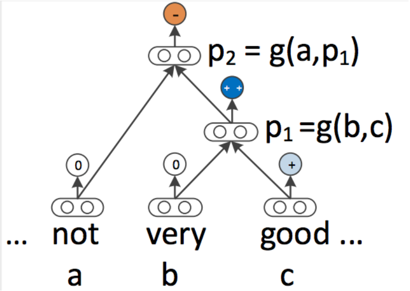
Variants of the LSTM model

Many different LSTM variants have been proposed:

- Peep-hole connections (almost standard now)
  all gates can also look at the **state**
- Coupling forget and input gates
  basically we input only when we forget
- Gated Recurrent Units (GRUs)
  where a single gate controls forgetting and update

See "LSTM: a search space odyssey" by Greff et al., 2015

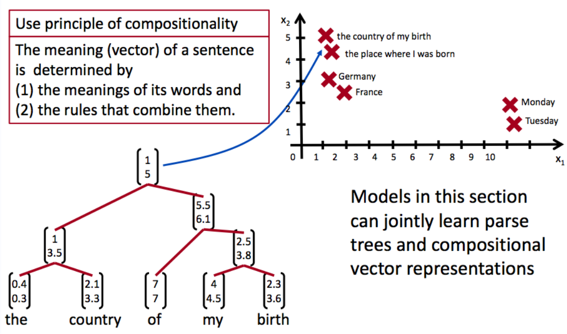
A generalization of RNNs to handle structured data in the form of a **dependency graph** such as a **tree**



[Figure by Socher et al., 2014]

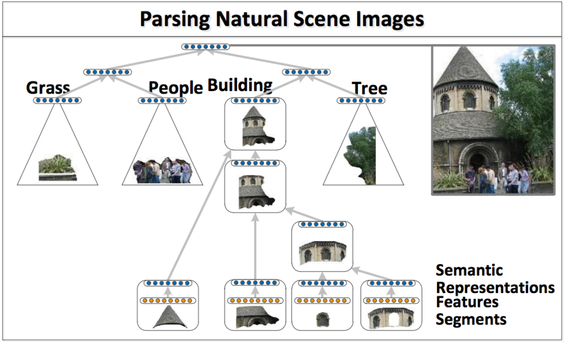
RNNs can be seen as RecNNs having a **linear chain** structure.

Exploit compositionality (e.g., parse trees in text)



[Figure by R. Socher]

Exploit compositionality (e.g., object parts in images)



[Figure by Socher et al., 2011]

Recursive Neural Tensor Networks (RNTNs)

Recently (2013) proposed by Stanford for sentiment analysis

- composition function over sentence **parse tree**
- exploit **parameters** (tensors) that are common to all nodes
- (tensor) backpropagation through structure

Later (2014-2015) also a **tree-structured** version of **LSTMs**

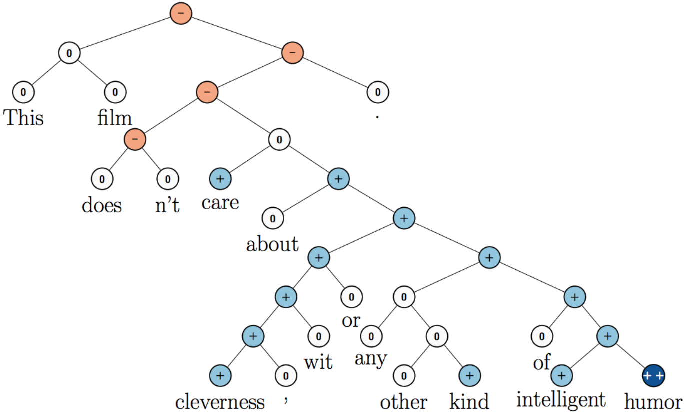
Recursive Neural Tensor Networks (RNTNs)

Sentiment Treebank built by **crowdsourcing**

- 11,855 sentences from movie reviews
- 215,154 labeled **phrases** (sub-sentences)

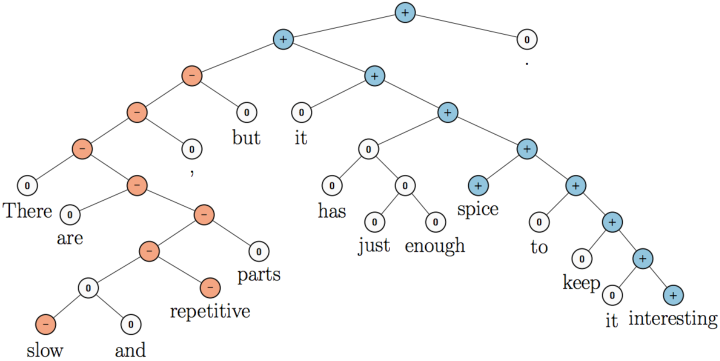
[Figure by Socher et al., 2014]

Recursive Neural Tensor Networks (RNTNs)



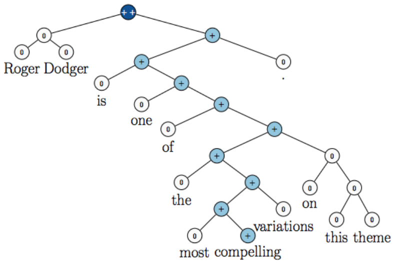
[Figure by Socher et al., 2014]

Recursive Neural Tensor Networks (RNTNs)



[Figure by Socher et al., 2014]

Recursive Neural Tensor Networks (RNTNs)



[Figure by Socher et al., 2014]

Recursive Neural Tensor Networks (RNTNs)

n	**Most positive n-grams**	**Most negative n-grams**
1	engaging ; best ; powerful ; love ; beautiful ; entertaining ; clever ; terrific ; excellent ; great ;	bad ; dull ; boring ; fails ; worst ; stupid ; painfully ; cheap ; forgettable ; disaster ;
2	excellent performances ; amazing performance ; terrific performances ; A masterpiece ; masterful film ; wonderful film ; terrific performance ; masterful piece ; wonderful movie ; marvelous performances ;	worst movie ; bad movie ; very bad ; shapeless mess ; worst thing ; tepid waste ; instantly forgettable ; bad film ; extremely bad ; complete failure ;
3	an amazing performance ; a terrific performance ; a wonderful film ; wonderful all-ages triumph ; A masterful film ; a wonderful movie ; a tremendous performance ; drawn excellent performances ; most visually stunning ; A stunning piece ;	for worst movie ; A lousy movie ; most joyless movie ; a complete failure ; another bad movie ; fairly terrible movie ; a bad movie ; extremely unfunny film ; most painfully marginal ; very bad sign ;
5	nicely acted and beautifully shot ; gorgeous imagery , effective performances ; the best of the year ; a terrific American sports movie ; very solid , very watchable ; a fine documentary does best ; refreshingly honest and ultimately touching ;	silliest and most incoherent movie ; completely crass and forgettable movie ; just another bad movie . ; drowns out the lousy dialogue ; a fairly terrible movie ... ; A cumbersome and cliche-ridden movie ; a humorless , disjointed mess ;
8	one of the best films of the year ; simply the best family film of the year ; the best film of the year so far ; A love for films shines through each frame ; created a masterful piece of artistry right here ; A masterful film from a master filmmaker , ; 's easily his finest American film ... comes ;	A trashy , exploitative , thoroughly unpleasant experience ; this sloppy drama is an empty vessel . ; a meandering , inarticulate and ultimately disappointing film ; an unimaginative , nasty , glibly cynical piece ; bad , he 's really bad , and ; quickly drags on becoming boring and predictable . ; be the worst special-effects creation of the year ;

[Figure by Socher et al., 2014]

Recursive Neural Tensor Networks (RNTNs)

Very nice model but:

- not easy to adapt it to other domains (Reddit, Twitter, etc.)
- need to compute parse tree in advance
- need classification at phrase-level (expensive)

Other applications

Modeling speech

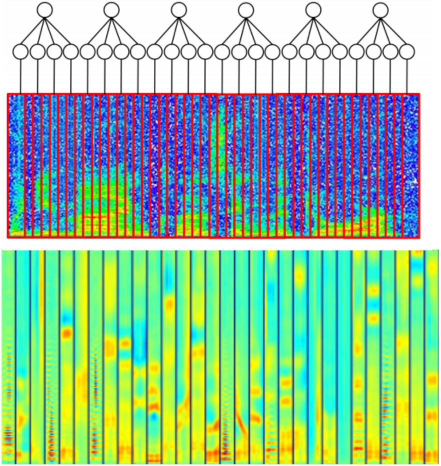
Several tasks. . .

- Speech recognition
- Speaker identification
- Speaker gender classification
- Phone classification
- Music genre classification
- Artist classification
- Music retrieval

Modeling speech

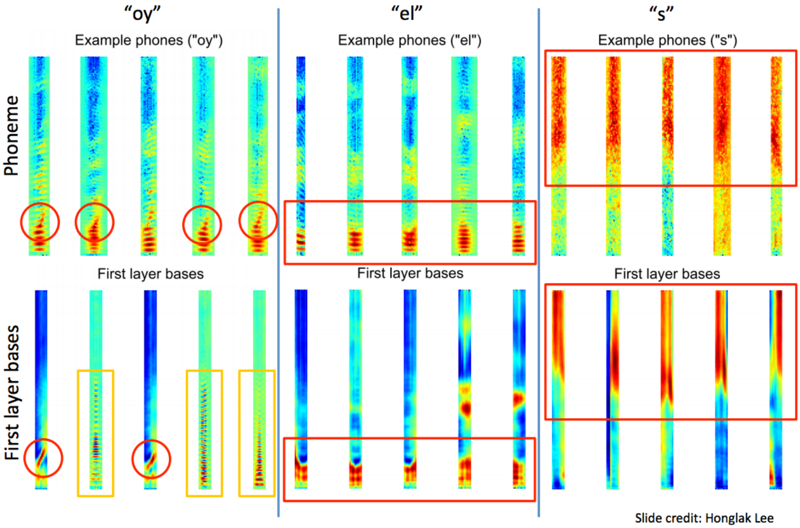
Feature extraction from acoustic signals. . .

. . . Tons of unsupervised data !

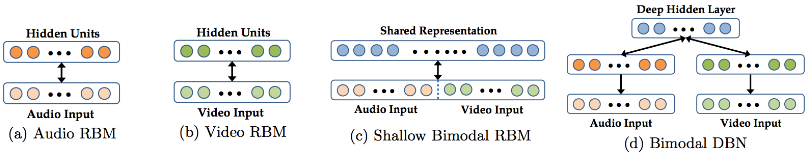


[Figure by H. Lee]

Modeling speech



[Figure by H. Lee]



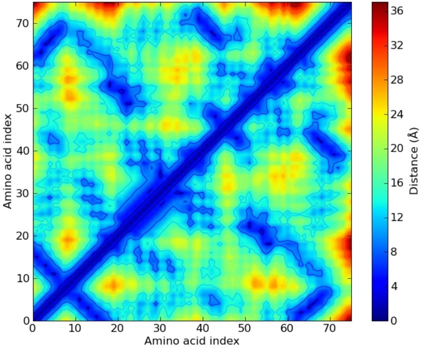
[Figure by Ngiam et al.]

Can we gain something from **shared representations**... ?
... Most probably yes !

Protein structure prediction (contact maps)

- Deep Spatio-Temporal Architectures [Di Lena et al., 2012]

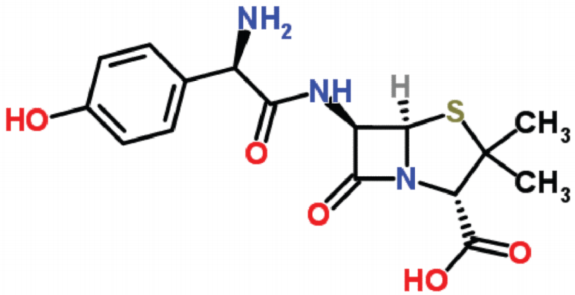
Sequence: ASCDEVVGSACH...CPPGAERMMAYGV



[Figure by Rafferty et al.]

Predicting the aqueous solubility of drug-like molecules

- Recursive neural networks [Lusci et al., 2013]



[Figure by Lusci et al.]

Ongoing Research and Concluding Remarks

Summary

Deep learning has obtained breakthrough results in many tasks

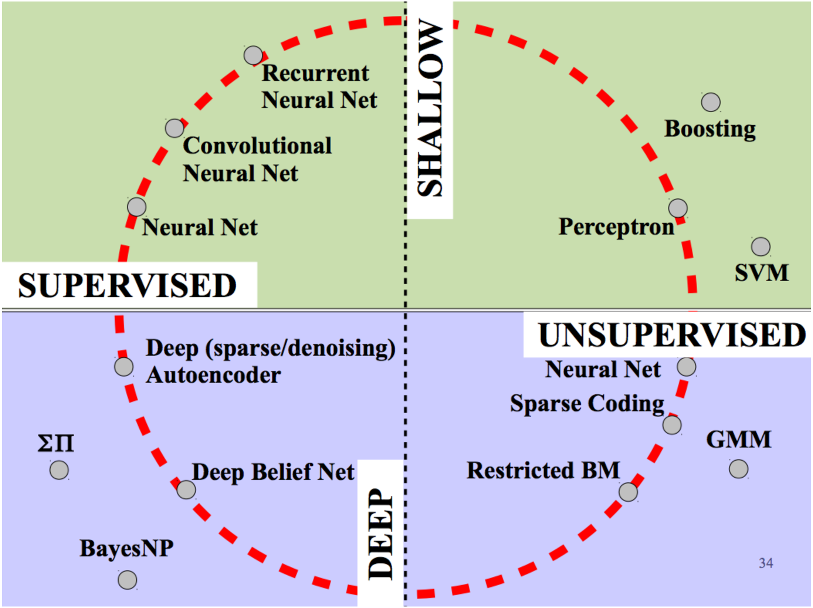
- exploit unsupervised data
- learning feature hierarchies
- some optimization tricks (dropout, rectification, ...)

Is this the **solution** to all AI problems ? Probably not but...

- for certain types of task **it is hard to compete**
- **huge datasets** and **many computational resources**
- **big companies** will likely play the major role
- huge space for applications **upon deep learning systems**

...But what is missing ?

A nice overview of methods



We are not yet there. . .

Still far from human-level in unrestricted domains...

The Caffe neural network library makes implementing state-of-the-art computer vision systems easy.

Classification

Click for a Quick Example

Maximally accurate | Maximally specific

structure	0.59865
roof	0.56189
building	0.44220
protective covering	0.42164
residence	0.34297

CNN took 0.123 seconds.

Again: symbolic vs. sum-symbolic

Still basically a **sub-symbolic approach**

- Representation learning: a step towards symbols...
- What are the connections with symbolic approaches ?
- What about logic and reasoning ?

Some steps forward:

- bAbI tasks @ Facebook
- Neural Conversational Models @ Google
- Neural Turing Machines (NTMs) @ GoogleDeepMind

Text understanding and reasoning with deep networks

- The (20) bAbI tasks
- The Children's Book Test
- The Movie Dialog dataset
- The SimpleQuestions dataset

bAbI tasks (Facebook)

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions
John moved to the playground.
Daniel went to the bathroom.
John went back to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets
Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Task 9: Simple Negation
Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge
John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

[Table by Weston et al.]

bAbI tasks (Facebook)

Task 11: Basic Coreference
Daniel was in the kitchen.
Then he went to the studio.
Sandra was in the office.
Where is Daniel? A:studio

Task 12: Conjunction
Mary and Jeff went to the kitchen.
Then Jeff went to the park.
Where is Mary? A: kitchen
Where is Jeff? A: park

Task 13: Compound Coreference
Daniel and Sandra journeyed to the office.
Then they went to the garden.
Sandra and John travelled to the kitchen.
After that they moved to the hallway.
Where is Daniel? A: garden

Task 14: Time Reasoning
In the afternoon Julie went to the park.
Yesterday Julie was at school.
Julie went to the cinema this evening.
Where did Julie go after the park? A:cinema
Where was Julie before the park? A:school

Task 15: Basic Deduction
Sheep are afraid of wolves.
Cats are afraid of dogs.
Mice are afraid of cats.
Gertrude is a sheep.
What is Gertrude afraid of? A:wolves

Task 16: Basic Induction
Lily is a swan.
Lily is white.
Bernhard is green.
Greg is a swan.
What color is Greg? A:white

Task 17: Positional Reasoning
The triangle is to the right of the blue square.
The red square is on top of the blue square.
The red sphere is to the right of the blue square.
Is the red sphere to the right of the blue square? A:yes
Is the red square to the left of the triangle? A:yes

Task 18: Size Reasoning
The football fits in the suitcase.
The suitcase fits in the cupboard.
The box is smaller than the football.
Will the box fit in the suitcase? A:yes
Will the cupboard fit in the box? A:no

Task 19: Path Finding
The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen? A: west, north
How do you go from office to bathroom? A: north, west

Task 20: Agent's Motivations
John is hungry.
John goes to the kitchen.
John grabbed the apple there.
Daniel is hungry.
Where does Daniel go? A:kitchen
Why did John go to the kitchen? A:hungry

[Table by Weston et al.]

```
S: 1 So they had to fall a long way .
   2 So they got their tails fast in their mouths .
   3 So they could n't get them out again .
   4 That 's all . '
   5 ` Thank you , ' said Alice , ` it 's very interesting .
   6 I never knew so much about a whiting before . ' '
   7 I can tell you more than that , if you like , ' said the Gryphon .
   8 ` Do you know why it 's called a whiting ? ''
   9 I never thought about it , ' said Alice .
   10 ` Why ? '
   11 ` IT DOES THE BOOTS AND SHOES . '
   12 the Gryphon replied very solemnly .
   13 Alice was thoroughly puzzled .
   14 ` Does the boots and shoes ! '
   15 she repeated in a wondering tone .
   16 ` Why , what are YOUR shoes done with ? '
   17 said the Gryphon . '
   18 I mean , what makes them so shiny ? '
   19 Alice looked down at them , and considered a little before she gave
   her answer .
   20 ` They 're done with blacking , I believe .

q: `Boots and shoes under the sea , ' the _________ went on in a deep
voice , are done with a whiting .
C: Alice, BOOTS, Gryphon, SHOES, answer, fall, mouths, tone, way, whiting.

MemNNs (window + self-sup.): Gryphon
```

```
S: 1 He thought that Old Mr. Toad was trying to fool him .
   2 Presently Peter Rabbit came along .
   3 He found Jimmy Skunk sitting in a brown study .
   4 He had quite forgotten to look for fat beetles , and when he forgets to do
   that you may make up your mind that Jimmy is doing some hard thinking .
   5 `` Hello , old Striped-coat , what have you got on your mind this fine
   morning ? ''
   6 cried Peter Rabbit .
   7 `` Him , '' said Jimmy simply , pointing down the Lone Little Path .
   8 Peter looked .
   9 `` Do you mean Old Mr. Toad ! ''
   10 he asked .
   11 Jimmy nodded .
   12 `` Do you see anything queer about him ? ''
   13 he asked in his turn .
   14 `` Do you see anything queer about him ? ''
   15 he asked .
   16 Peter stared down the Lone Little Path .
   17 `` No , '' he replied , `` except that he seems in a great hurry . ''
   18 `` That 's just it , '' Jimmy returned promptly .
   19 `` Did you ever see him hurry unless he was frightened ? ''
   20 Peter confessed that he never had

q: `` Well , he is n't ________ now , yet just look at him go '' retorted Jimmy .

C: Do, came, confessed, frightened, mean, replied, returned, said, see, thought.

MemNNs (window +self-sup.): frightened
```

Figure 2: **Correct predictions of MemNNs (window memory + self-supervision) on CBT** on Named Entity (left) and Verb (right). Circled phrases indicate all considered windows; red ones are the ones corresponding to the returned (correct) answer; the blue windows represent the queries.

[Table by Hill et al., 2016]

Movie Dialog dataset

Task 1: Factoid Question Answering (QA)

What movies are about open source? Revolution OS
Ruggero Raimondi appears in which movies? Carmen
What movies did Darren McGavin star in? Billy Madison, The Night Stalker, Mrs. Pollifax-Spy
Can you name a film directed by Stuart Ortiz? Grave Encounters
Who directed the film White Elephant? Pablo Trapero
What is the genre of the film Dial M for Murder? Thriller, Crime
What language is Whity in? German

Task 2: Recommendation

Schindler's List, The Fugitive, Apocalypse Now, Pulp Fiction, and The Godfather are films I really liked.
Can you suggest a film? The Hunt for Red October

Some movies I like are Heat, Kids, Fight Club, Shaun of the Dead, The Avengers, Skyfall, and Jurassic Park.
Can you suggest something else I might like? Ocean's Eleven

Task 3: QA + Recommendation Dialog

I loved Billy Madison, My Neighbor Totoro, Blades of Glory, Bio-Dome, Clue, and Happy Gilmore.
I'm looking for a Music movie. School of Rock
What else is that about? Music, Musical, Jack Black, school, teacher, Richard Linklater, rock, guitar
I like rock and roll movies more. Do you know anything else? Little Richard

Tombstone, Legends of the Fall, Braveheart, The Net, Outbreak, and French Kiss are films I really liked.
I'm looking for a Fantasy movie. Jumanji
Who directed that? Joe Johnston
I like Tim Burton movies more. Do you know anything else? Big Fish

[Table by Dodge et al., 2016]

SimpleQuestions dataset

What American cartoonist is the creator of Andy Lippincott?	(andy_lippincott, character_created_by, garry_trudeau)
Which forest is Fires Creek in?	(fires_creek, containedby, nantahala_national_forest)
What is an active ingredient in childrens earache relief ?	(childrens_earache_relief, active_ingredients, capsicum)
What does Jimmy Neutron do?	(jimmy_neutron, fictional_character_occupation, inventor)
What dietary restriction is incompatible with kimchi?	(kimchi, incompatible_with_dietary_restrictions, veganism)

Table 1: **Examples of simple QA**. Questions and corresponding facts have been extracted from the new dataset SimpleQuestions introduced in this paper. Actual answers are underlined.

[Table by Bordes et al., 2015]

Neural Conversational Model (Google)

Human: *why are we here ?*
Machine: *i 'm not sure .*
Human: *okay, bye !*
Machine: *bye .*
Human: *see you later .*
Machine: *bye .*

Conversation 2: Simple Q&A (contexts and multiple choice)

Human: *my name is david . what is my name ?*
Machine: *david .*
Human: *my name is john . what is my name ?*
Machine: *john .*
Human: *are you a leader or a follower ?*
Machine: *i 'm a leader .*
Human: *are you a follower or a leader ?*
Machine: *i 'm a leader .*

Conversation 4: Philosophical Q&A

Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

[Table by Vinyalis & Le, 2015]

Neural Conversational Model (Google)

Conversation 3: General knowledge Q&A

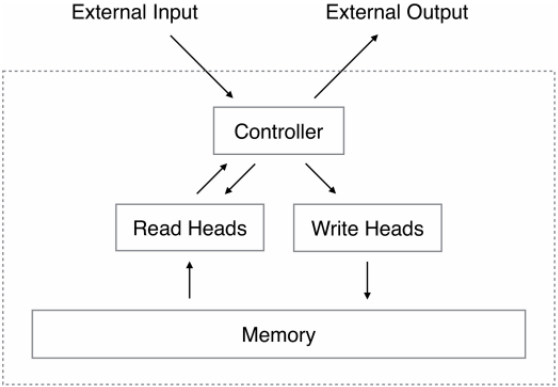
Human: *who is skywalker ?*
Machine: *he is a hero .*
Human: *who is bill clinton ?*
Machine: *he 's a billionaire .*
Human: *is sky blue or black ?*
Machine: *blue .*
Human: *does a cat have a tail ?*
Machine: *yes .*
Human: *does a cat have a wing ?*
Machine: *no*
Human: *can a cat fly ?*
Machine: *no .*
Human: *how many legs does a cat have ?*
Machine: *four , i think .*
Human: *how many legs does a spider have ?*
Machine: *three , i think .*
Human: *how many legs does a centipede have ?*
Machine: *eight .*
Human: *what is the color of the sky ?*
Machine: *blue .*
Human: *what is the color of water ?*
Machine: *water .*
Human: *what is the color of blood ?*
Machine: *it is the same as a black eye .*
Human: *what is the usual color of a leaf ?*
Machine: *it is a green one .*
Human: *what is the color of a yellow car ?*
Machine: *yellow .*
Human: *how much is two plus two ?*
Machine: *four .*

Conversation 5: Morality

Human: *what is moral ?*
Machine: *what empowered humanity , what intellectual the essence is .*
Human: *what is immoral ?*
Machine: *the fact that you have a child .*
Human: *what is morality ?*
Machine: *what is altruism ?*
Human: *what is the definition of altruism ?*
Machine: *if you don 't believe in god , then you don 't know .*
Human: *ok ... so what is the definition of morality ?*
Machine: *well , the truth is , you 're not a believer in god almighty .*
Human: *tell me the definition of morality , i am quite upset now !*
Machine: *i 'm not ashamed of being a philosopher !*
Human: *tell me the definition of morality .*
Machine: *i don 't have ethics .*
Human: *look , i need help , i need to know more about morality ...*
Machine: *i don 't know what ethics is .*
Human: *seriously , what is morality ?*
Machine: *what is the definition of living ?*
Human: *why living has anything to do with morality ?*
Machine: *you 're not a cop .*
Human: *okay , i need to know how should i behave morally ...*

[Table by Vinyalis & Le, 2015]

Neural Turing Machines (Google Deepmind)



[Figure by Graves et al., 2014]

- Inspired by Turing Machines
- Memory (tape) to read/write through deep networks
- Capable of learning "simple" algorithms (e.g., sorting)

Learning concepts

Examples from the ImageNet category “Restaurant”

Is it enough to learn **the concept of a restaurant** ?

Software packages and a few references

Tensorflow

Developed by Google

- Python
- Computational graph abstraction
- Parallelize over **both** data and model
- The multi-machine part is **not open source**
- Nice tool to visualize stuff (Tensorboard)
- A little slower than other systems
- Provides only a few (one?) pre-trained models

Theano

University of Montréal (Yoshua Bengio's group)

- Python
- Computational graph abstraction
- The code is somehow difficult (low-level)
- Long compile times
- Easily GPU and multi-GPU

Keras and Lasagne

Keras is a wrapper for Theano or Tensorflow

- Python
- Plug-and-play layers, losses, optimizers, etc.
- Sometimes error messages can be criptic. . .

Lasagne is a wrapper for Theano

- Python
- Plug-and-play layers, losses, optimizers, etc.
- **Model zoo** with plenty pre-trained architectures
- Still employs some **symbolic computation** as plain Theano

Torch

Facebook, Google, Twitter, IDIAP, ...

- mostly written in Lua and C
- sharing similarities to python (e.g. tensors vs. numpy arrays)
- module **nn** to train neural networks
- same code running for both CPU and GPU

One notable product by Torch: **Overfeat**
A CNN for image classification, object detection, etc.

Caffe

Berkeley Vision and Learning Center (BVLC)

- written in C++
- Python and MATLAB bindings (although not much documented)
- very popular for CNNs, not much for RNNs...
- quite a large model zoo (AlexNet, GoogLeNet, ResNet, ...)
- scripts for training without writing code

Just to give an idea of Caffe's performance...

- During training $\sim$60M images per day **with a single GPU**
- At test time $\sim$1 ms/image

Matlab

Plenty of resources

- Code by G. Hinton on RBMs and DBNs (easy to try !)
- Autoencoders (many different implementations)
- Convolutional neural networks
- . . .

Websites and references

- `http://deeplearning.net`
- `http://www.cs.toronto.edu/~hinton/`
- `http://www.iro.umontreal.ca/~bengioy/`
- `http://yann.lecun.com/`
- Introductory paper by Yoshua Bengio:
  "Learning Deep Architectures for AI"

Videos

- Geoffrey Hinton on DBNs:
  https://www.youtube.com/watch?v=AyzOUbkUf3M
- Yoshua Bengio (Deep Learning lectures):
  https://www.youtube.com/watch?v=JuimBuvEWBg
- Yann LeCun (CNNs, energy-based models):
  https://www.youtube.com/watch?v=oOB4evKlEmQ