CS224d: Deep NLP

Lecture 15: Applications of DL to NLP

Richard Socher

richard@metamind.io

Overview

- Model overview: How to properly compare to other models and choose your own model
  - Word representation
  - Phrase composition
  - Objective function
  - Optimization

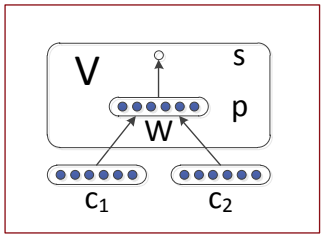
- Character RNNs on text and code
- Morphology
- Logic
- Question Answering
- Image – Sentence mapping

Model overview: Word Vectors

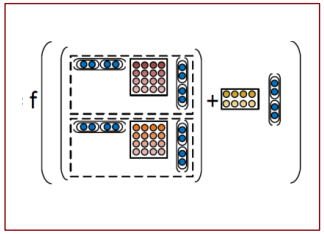
- Random
- Word2Vec
- **Glove**
- Dimension – often defines the number of model parameters
- Or work directly on characters or morphemes

Model overview: Phrase Vector Composition

- Composition Function governs how exactly word and phrase vectors interact to compose meaning

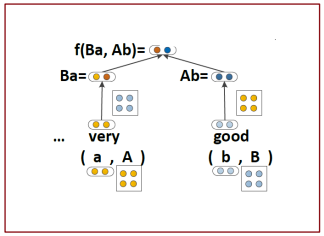


- Averaging: p = a + b
  - Lots of simple alternatives



- Recursive neural networks

- Convolutional neural networks



- Recurrent neural network

Composition: Bigram and Recursive functions

- Many related models are special cases of MV-RNN

$$p = f\left(W\begin{bmatrix} Ba \\ Ab \end{bmatrix}\right)$$

- Mitchell and Lapata, 2010; Zanzotto et al., 2010:

$$p = Ba + Ab = id\left([I_{n\times n} I_{n\times n}]\begin{bmatrix} Ba \\ Ab \end{bmatrix}\right)$$

- Baroni and Zamparelli (2010): A is an adjective matrix and b is a noun vector $p = Ab = id\left([0_{n\times n} I_{n\times n}]\begin{bmatrix} Ba \\ Ab \end{bmatrix}\right)$

- RNNs of Socher et al. 2011 (ICML, EMNLP, NIPS) are also special cases

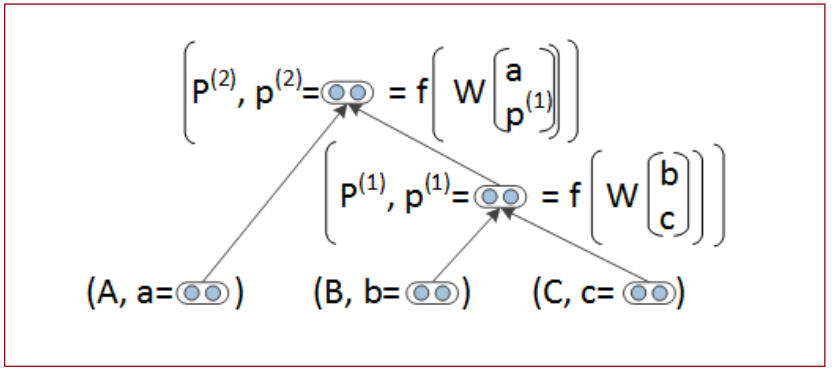
$$p = f\left(W\begin{bmatrix} I_{n\times n}a \\ I_{n\times n}b \end{bmatrix}\right)$$

- **Recursive neural tensor** networks bring quadratic and multiplicative interactions between vectors

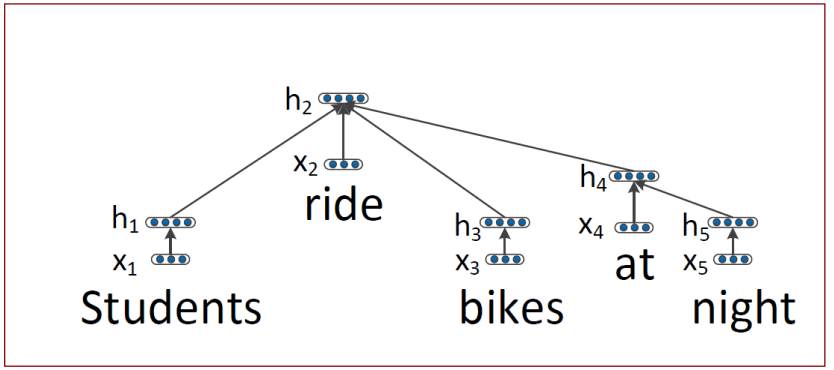
Additional choice for recursive neural nets

- Dependency trees focus more on semantic structure

1. Constituency Tree



2. Dependency Tree



3. Balanced Tree

Composition: CNNs

- Several variants also:
  - No pooling layers
  - Pooling layers: simple max-pooling or dynamic pooling
  - Pooling across different dimensions

- Somewhat less explored in NLP than RNNs[2]

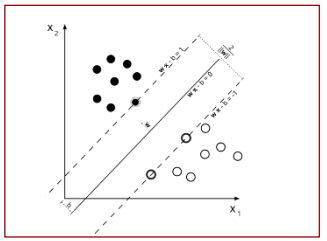
- Not linguistically nor cognitively plausible

Composition: Recurrent Neural Nets

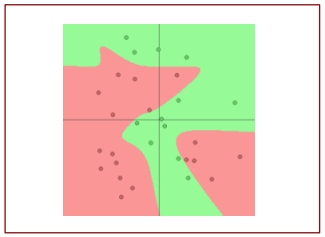
- Vanilla
- **GRU**
- **LSTM**
- Many variants of LSTMs
  "LSTM: A Search Space Odyssey" by Greff et al. 2015

Model overview: Objective function

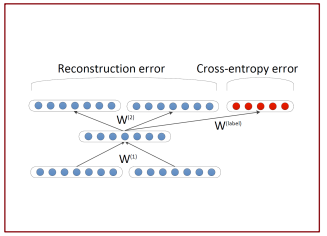
- **Max-margin**



- Cross-entropy
  - Supervised to predict a class
  - Unsupervised: predict surrounding words



- Auto-encoder
  - My opinion: Unclear benefits for NLP
  - Unless encoding another modality



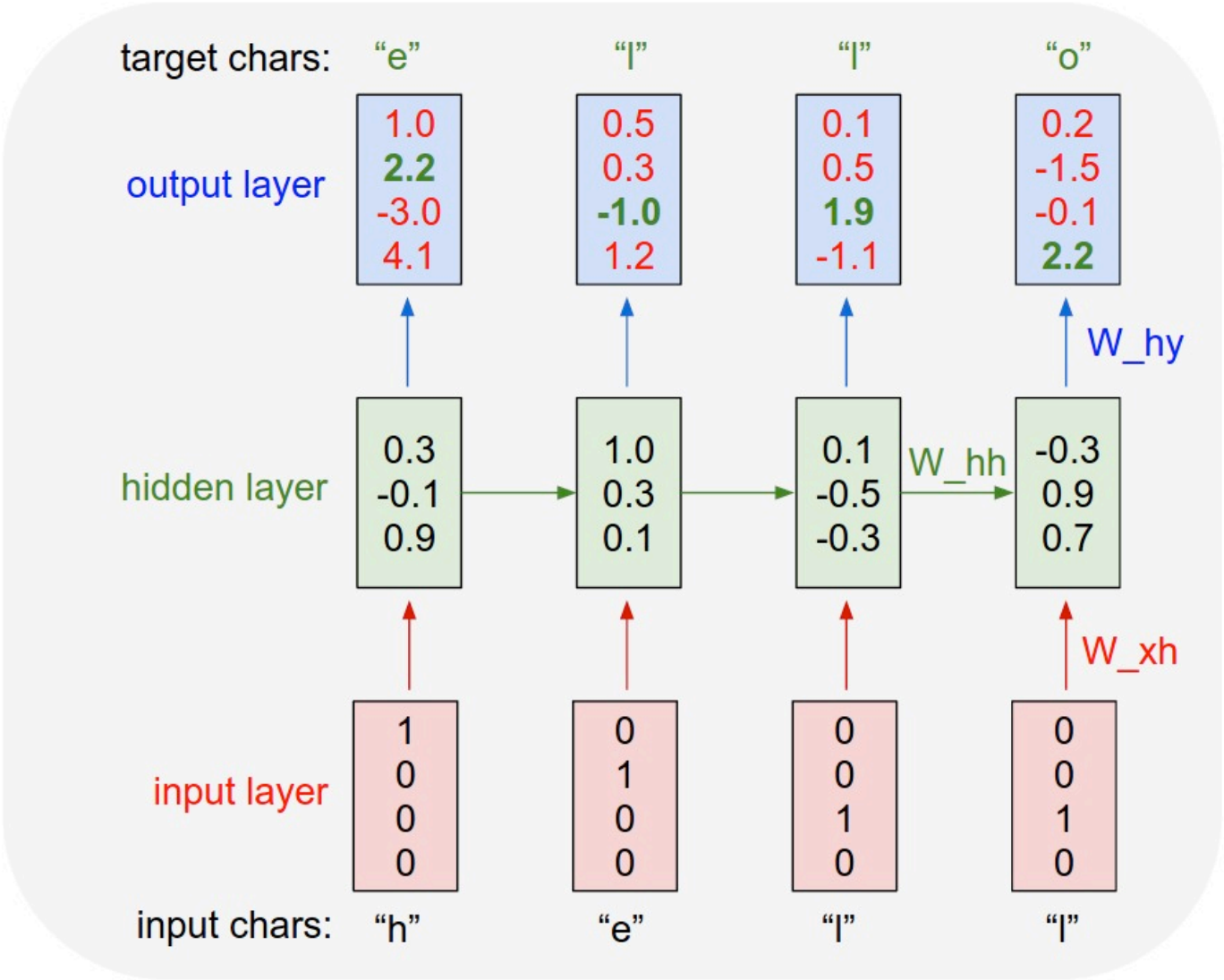
Optimization

- Initialization (word vector and composition parameters)!!

- Optimization algorithm
  - SGD
  - SGD + momentum
  - L-BFGS
  - AdaGrad
  - Adelta

- Optimization tricks
  - Regularization (some define as part of model)
  - Dropout

Character RNNs on text and code



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Character RNNs on text and code

- Haven't yet produced useful results on real datasets
- Shows that RNNs can memorize sequences and keep memory (mostly LSTMs)
- Most interesting results simply train on dataset and sample from it afterwards (first shown by Sutskever et al. 2011: Generating Text with Recurrent Neural Networks)
- Results from an LSTM (karpathy.github.io) →

Shakespeare

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.

Wikipedia

Naturalism and decision for the majority of Arab countries' capitalide was grounded by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal in the [[Protestant Immineners]], which could be said to be directly in Cantonese Communication, which followed a ceremony and set inspired prison, training. The emperor travelled back to [[Antioch, Perth, October 25|21]] to note, the Kingdom of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known in western [[Scotland]], near Italy to the conquest of India with the conflict.

Latex (had to be fixed manually)

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

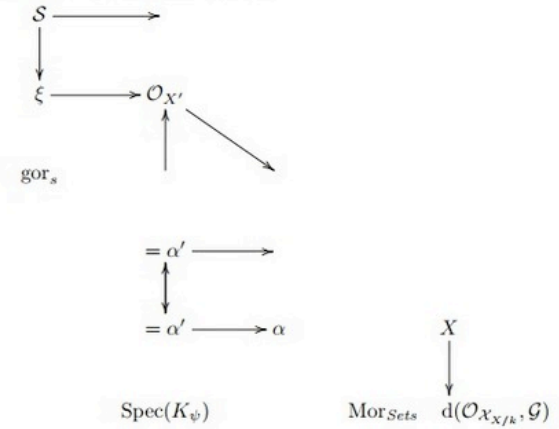
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-}1(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_t}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

Code! (Linux source code)

```
/*
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
  int error;
  if (fd == MARN_EPT) {
    /*
     * The kernel blank will coeld it to userspace.
     */
    if (ss->segment < mem_total)
      unblock_graph_and_set_blocked();
    else
      ret = 1;
    goto bail;
  }
  segaddr = in_SB(in.addr);
  selector = seg / 16;
  setup_works = true;
```

Morphology

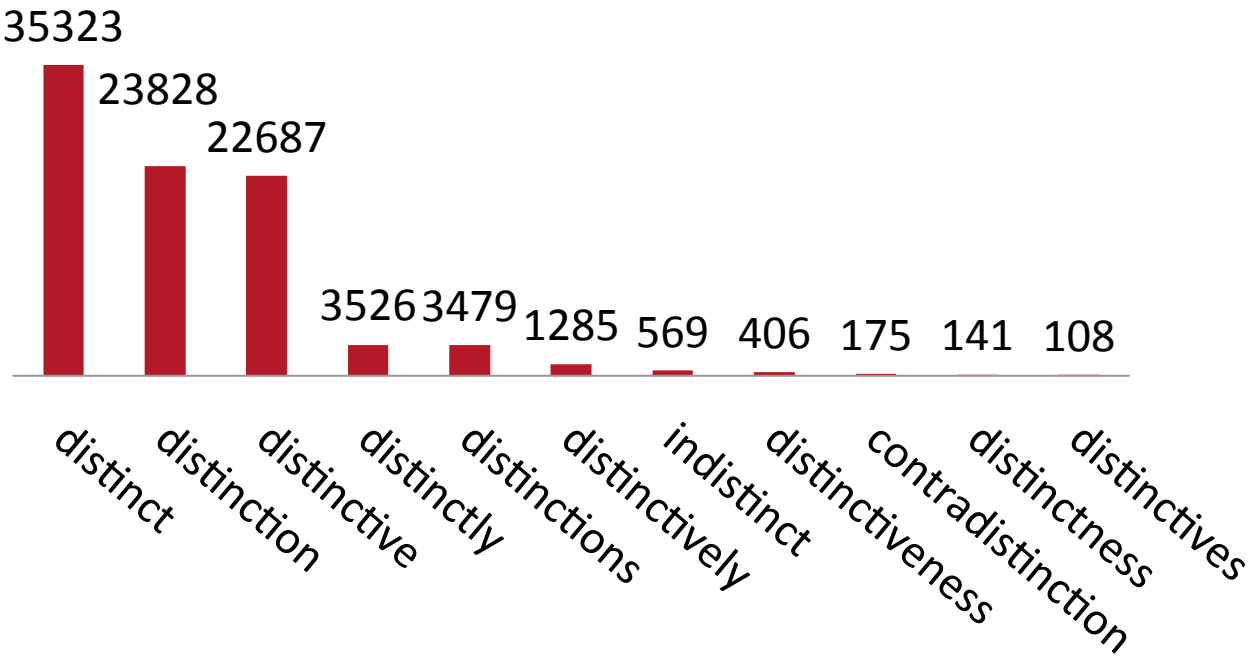
- Better Word Representations with Recursive Neural Networks for Morphology – Luong et al. (slides from Luong)

- *Problem with word vectors*:
  poorly estimate rare and complex words.

	(Collobert & Weston, 2010)	(Huang et. al., 2012)
distinct	different distinctive broader narrower	unique broad distinctive separate
distinctness	morphologies pesawat clefts pathologies	companion roskam hitoshi enjoyed
affect	exacerbate impacts characterize	allow prevent involve enable
unaffected	unnoticed dwarfed mitigated	monti sheaths krystal

Limitations of existing work

- Treat related words as independent entities.
- Represent unknown words with a few vectors.

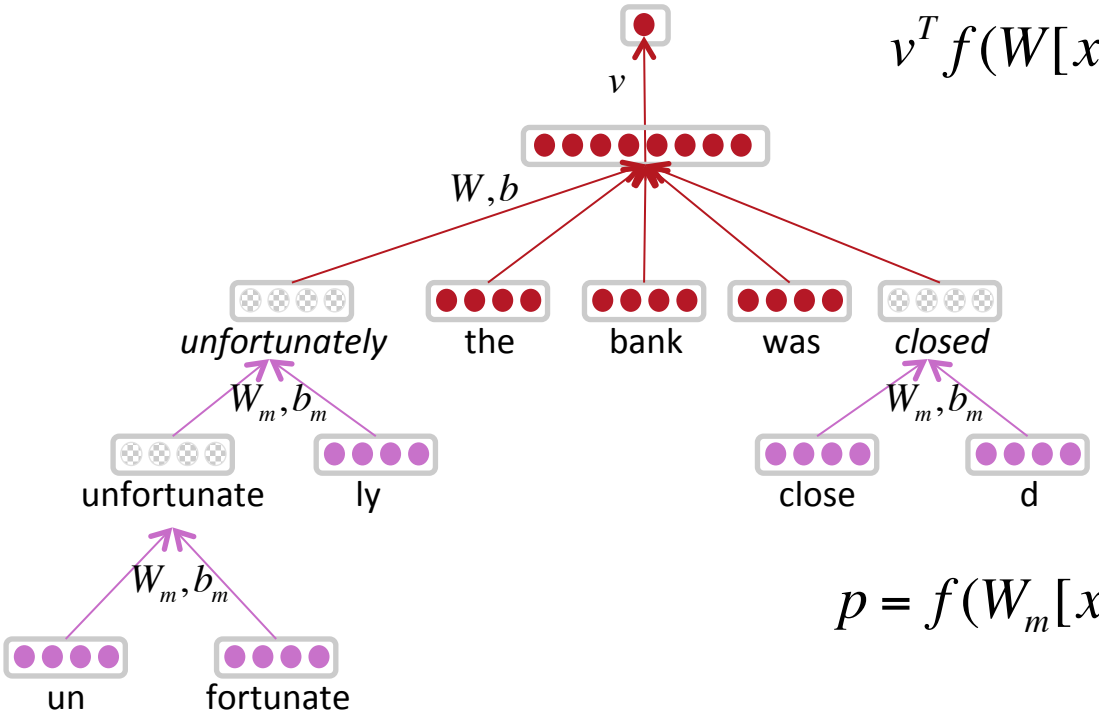


Word frequencies in Wikipedia docs (986m tokens)

Luong's approach – *network structure*

Neural Language Model

Morphology Model



$$v^T f(W[x_1; x_2; ...; x_n] + b)$$

$$p = f(W_m[x_{\text{stem}}; x_{\text{affix}}] + b_m)$$

- *Neural Language Model*: simple **feed-forward network** (Huang, et al., 2012) with **ranking-type cost** (Collobert et al., 2011).

- *Morphology Model*: **recursive neural network** (Socher et al., 2011).

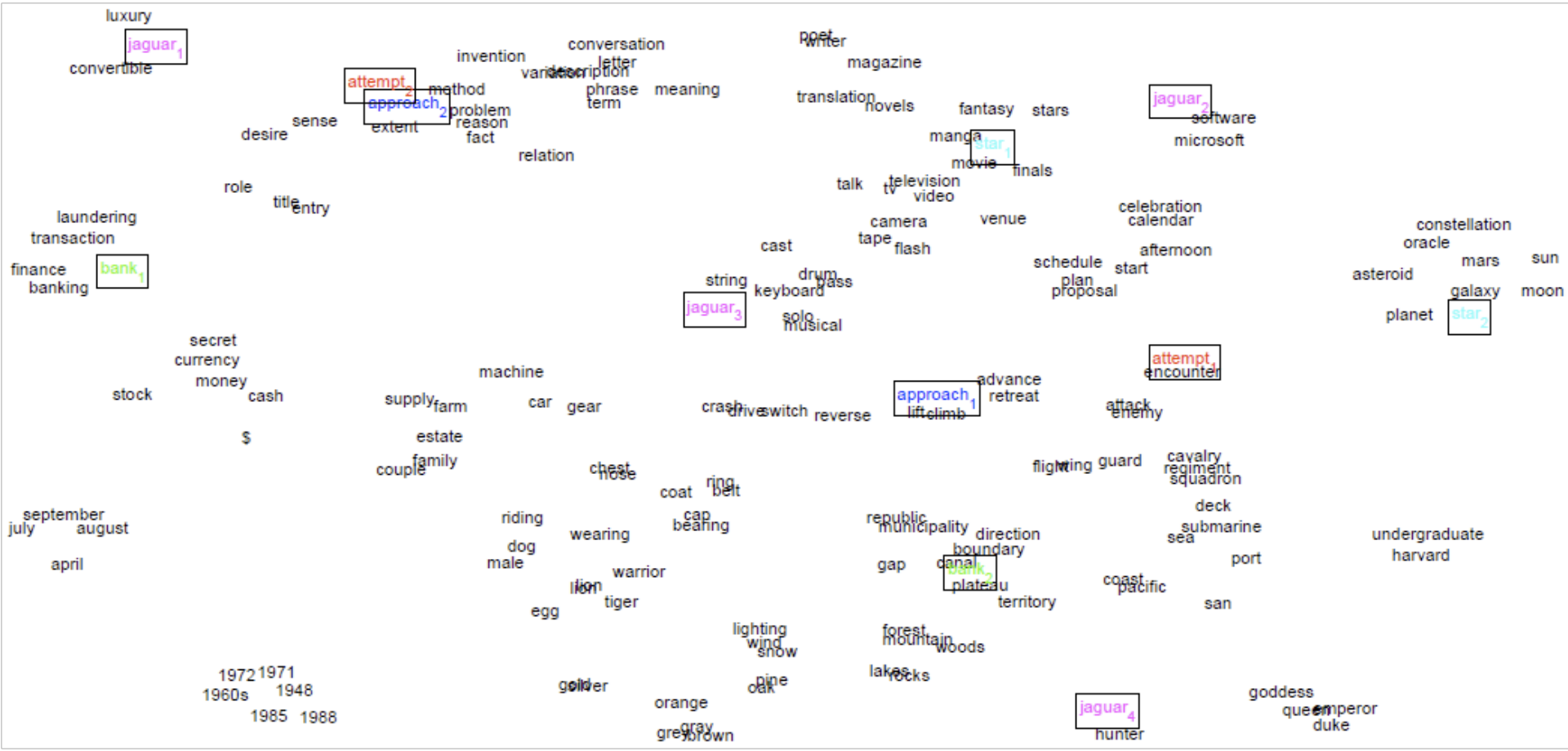
Analysis

- Blends word structure and syntactic-semantic information.

Words	(Collobert et al., 2011)	This work
commenting	insisting insisted focusing	commented comments criticizing
unaffected	unnoticed dwarfed mitigated	undesired unhindered unrestricted
distinct	different distinctive broader	divergent diverse distinctive
distinctness	morphologies pesawat clefts	distinctiveness smallness largeness
heartlessness	∅	corruptive inhumanity ineffectual
saudi-owned	avatar mohajir kripalani	saudi-based syrian-controlled

Solutions to the problem of polysemous words

- **Improving Word Representations Via Global Context And Multiple Word Prototypes by Huang et al. 2012**



Natural language inference

Claim: Simple task to define, but engages the full complexity of compositional semantics:

- Lexical entailment
- Quantification
- Coreference
- Lexical/scope ambiguity
- Commonsense knowledge
- Propositional attitudes
- Modality
- Factivity and implicativity

First training data

- Training data:
- *dance* ***entails*** *move*
- *waltz* ***neutral*** *tango*
- *tango* ***entails*** *dance*
- *sleep* ***contradicts*** *dance*
- *waltz* ***entails*** *dance*
-

Memorization (training set): Generalization (test set):

- *dance* ***???*** *move* *sleep* ***???*** *waltz*
- *waltz* ***???*** *tango* *tango* ***???*** *move*

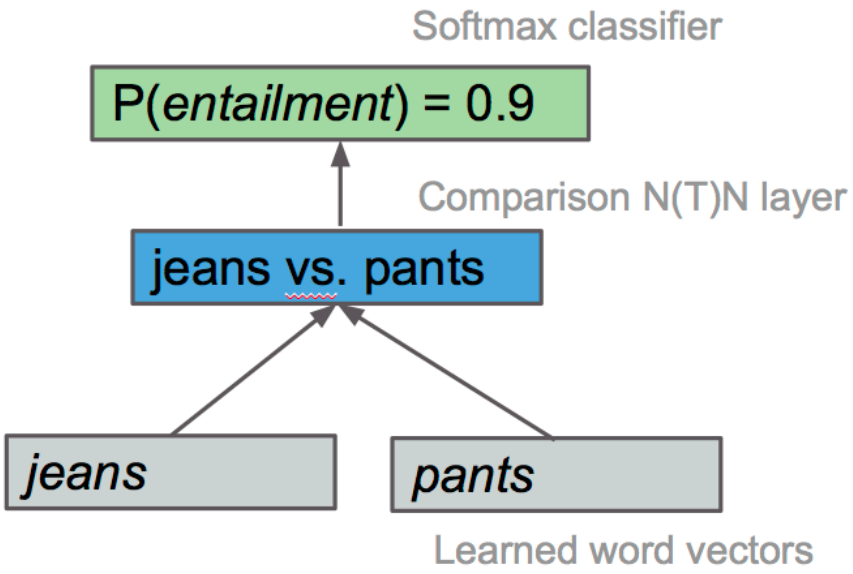
Natural language inference: definitions!

Slide from Bill MacCartne

	$x \equiv y$	equivalence	*couch* ≡ *sofa*
	$x \sqsubset y$	forward entailment (strict)	*crow* ⊏ *bird*
	$x \sqsupset y$	reverse entailment (strict)	*European* ⊐ *French*
	$x \wedge y$	negation (exhaustive exclusion)	*human* ^ *nonhuman*
	$x \mid y$	alternation (non-exhaustive exclusion)	*cat* | *dog*
	$x \smile y$	cover (exhaustive non-exclusion)	*animal* ⌣ *nonhuman*
	$x \# y$	independence	*hungry* # *hippo*

A minimal NN for lexical relations

- Words are learned embedding vectors.
- One plain RNN or RNTN layer
- Softmax emits relation labels
- Learn everything with SGD.



Recursion in propositional logic

Experimental approach: Train on relational statements generated from some formal system, test on other such relational statements.

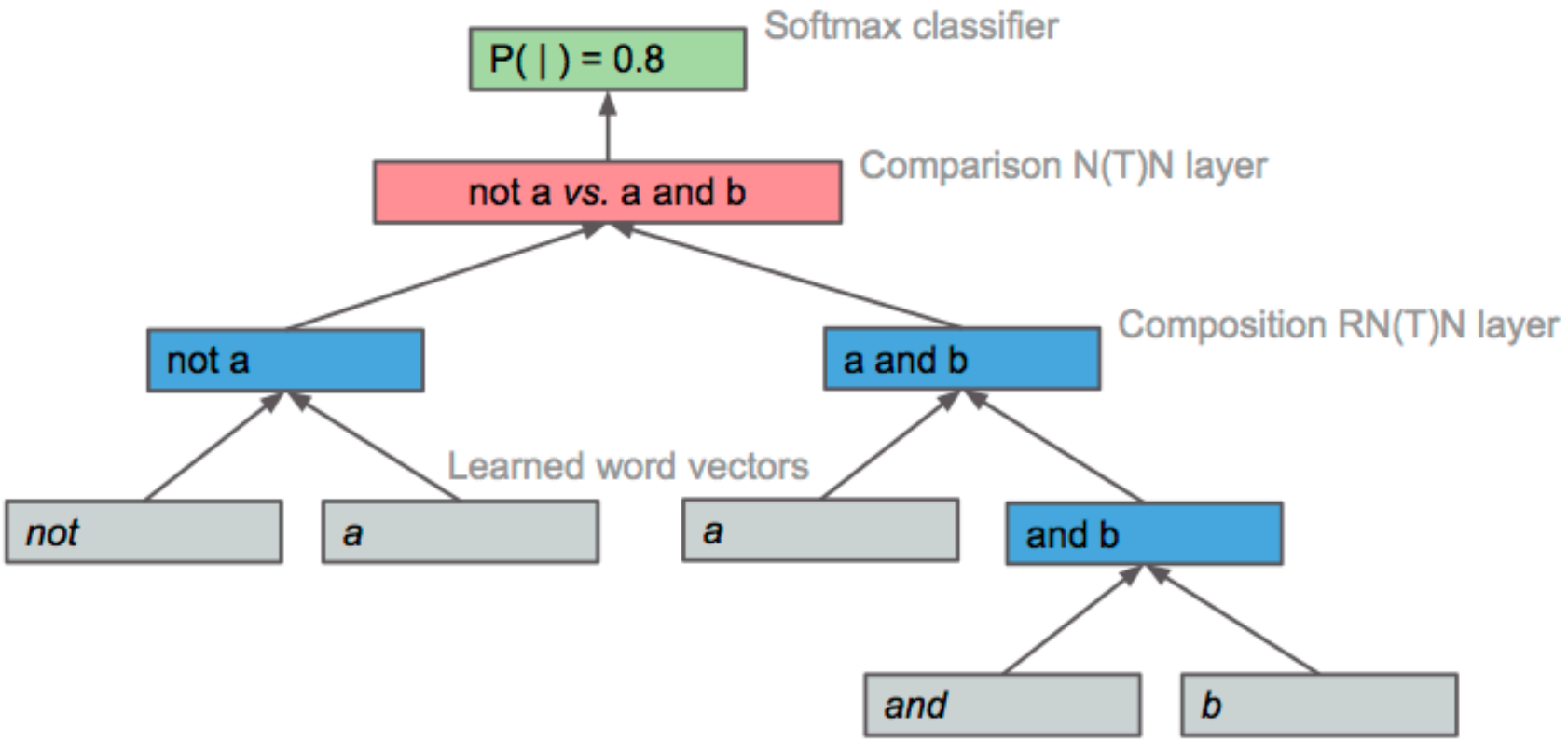
The model needs to:

- Learn the relations between individual words. (lexical relations)
- Learn how lexical relations impact phrasal relations.
  - This needs to be recursively applicable!

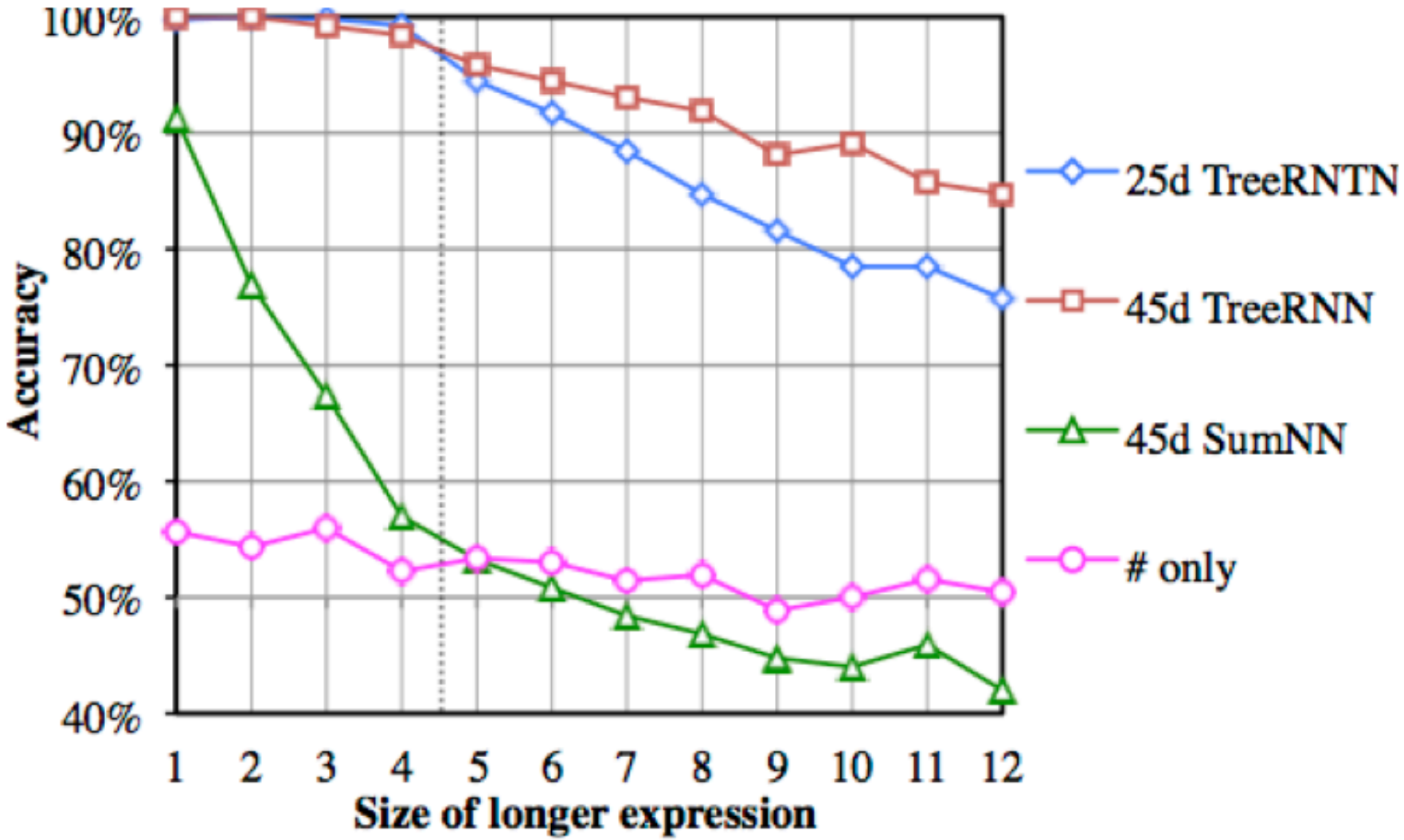
- a ≡ a, a ^ (not a), a ≡ (not (not a)), ...

Natural language inference with RNNs

- Two trees + learned comparison layer, then a classifier:



Natural language inference with RNNs



Question Answering: Quiz Bowl Competition

- **QUESTION:**
  He left unfinished a novel whose title character forges his father's signature to get out of school and avoids the draft by feigning desire to join. A more famous work by this author tells of the rise and fall of the composer Adrian Leverkühn. Another of his novels features the jesuit Naptha and his opponent Settembrini, while his most famous work depicts the aging writer Gustav von Aschenbach. Name this German author of The Magic Mountain and Death in Venice.

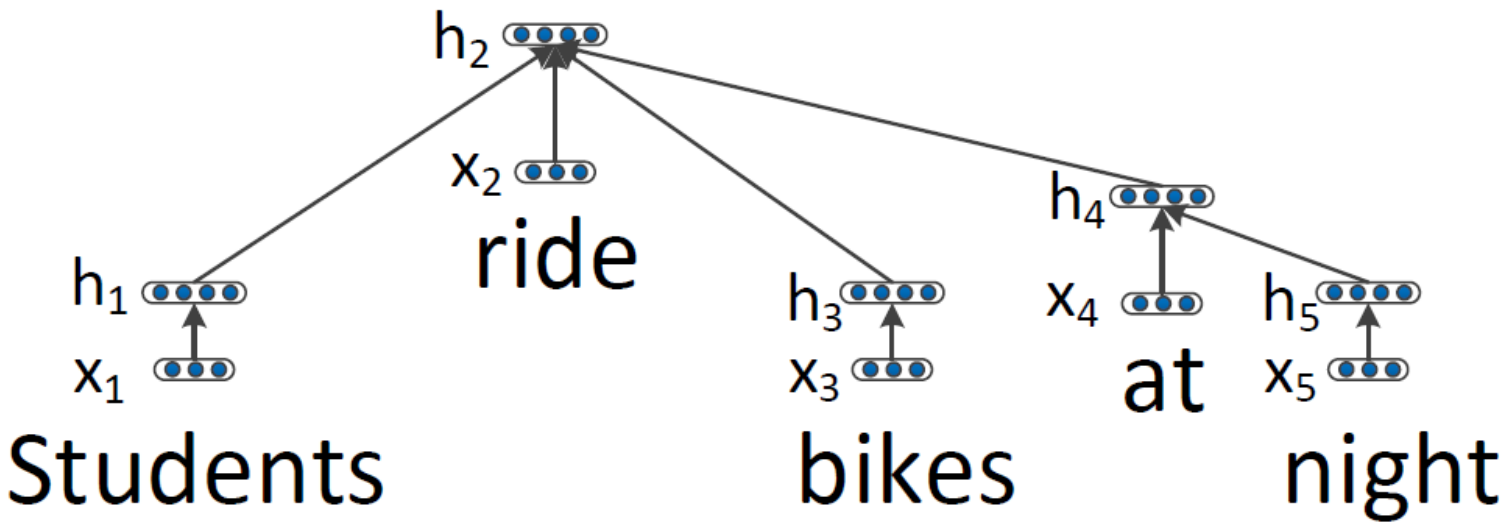
- Iyyer et al. 2014: A Neural Network for Factoid Question Answering over Paragraphs

Question Answering: Quiz Bowl Competition

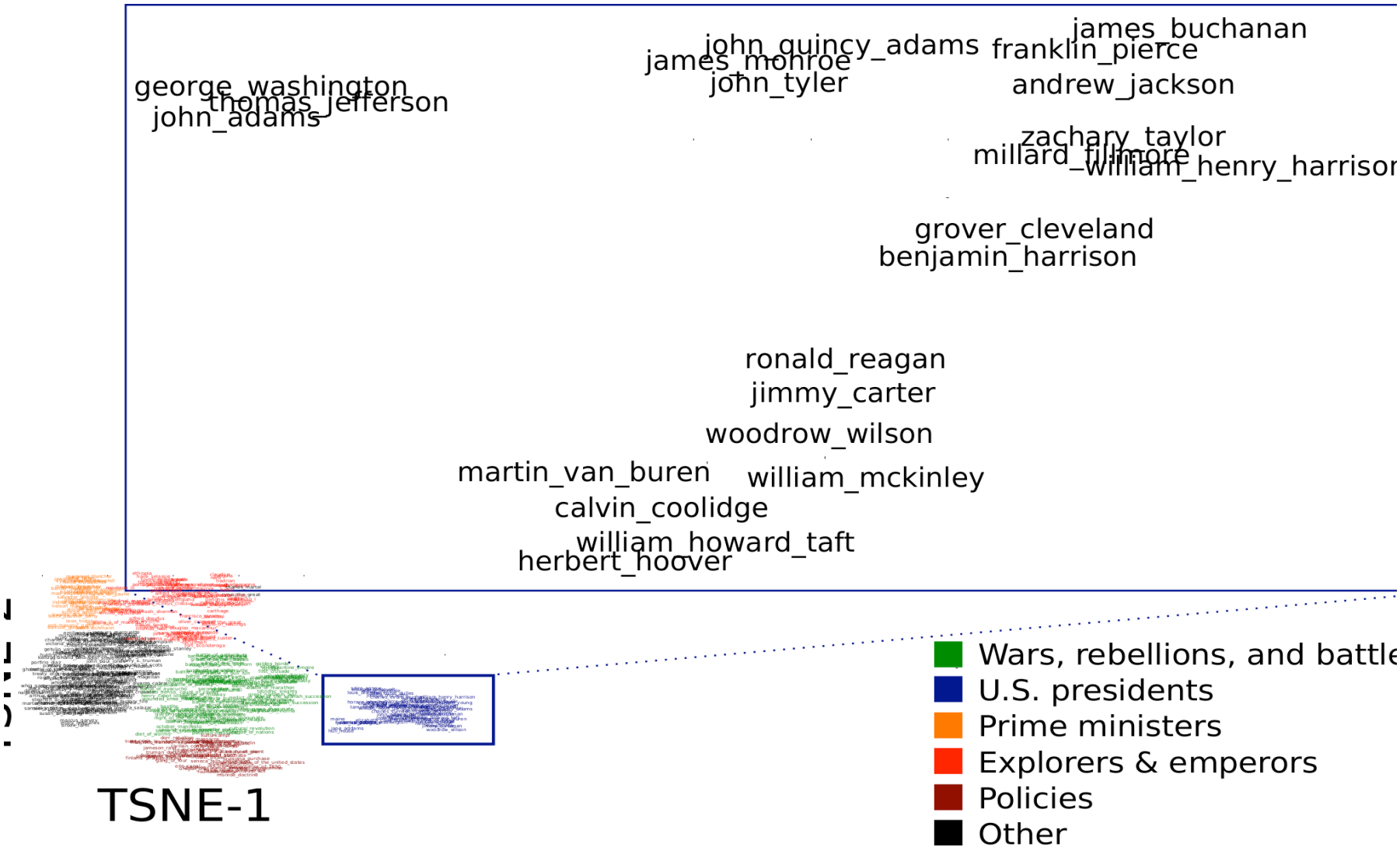
- **QUESTION:**
  He left unfinished a novel whose title character forges his father's signature to get out of school and avoids the draft by feigning desire to join. A more famous work by this author tells of the rise and fall of the composer Adrian Leverkühn. Another of his novels features the jesuit Naptha and his opponent Settembrini, while his most famous work depicts the aging writer Gustav von Aschenbach. Name this German author of The Magic Mountain and Death in Venice.
- **ANSWER:** Thomas Mann

Recursive Neural Networks

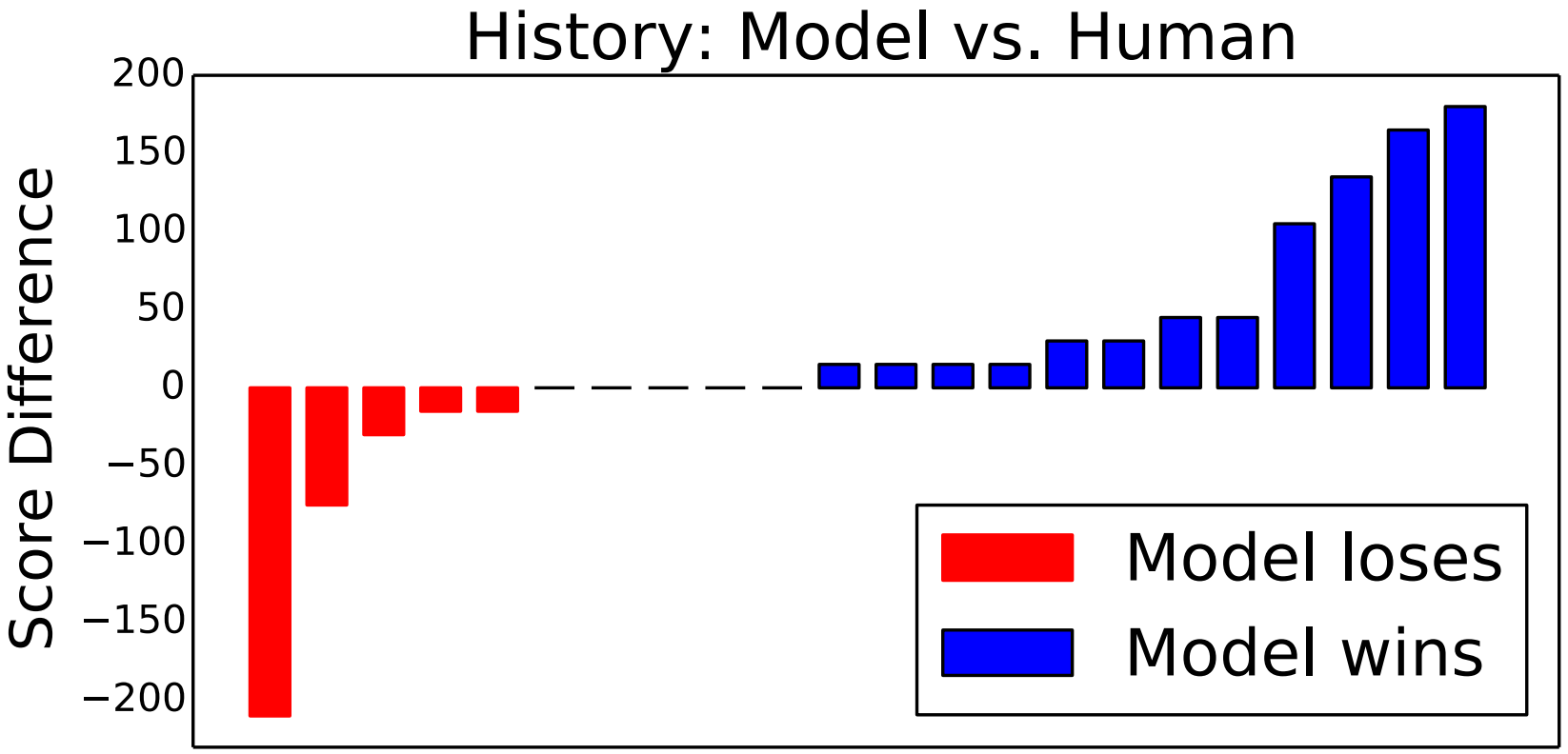
- Follow dependency structure



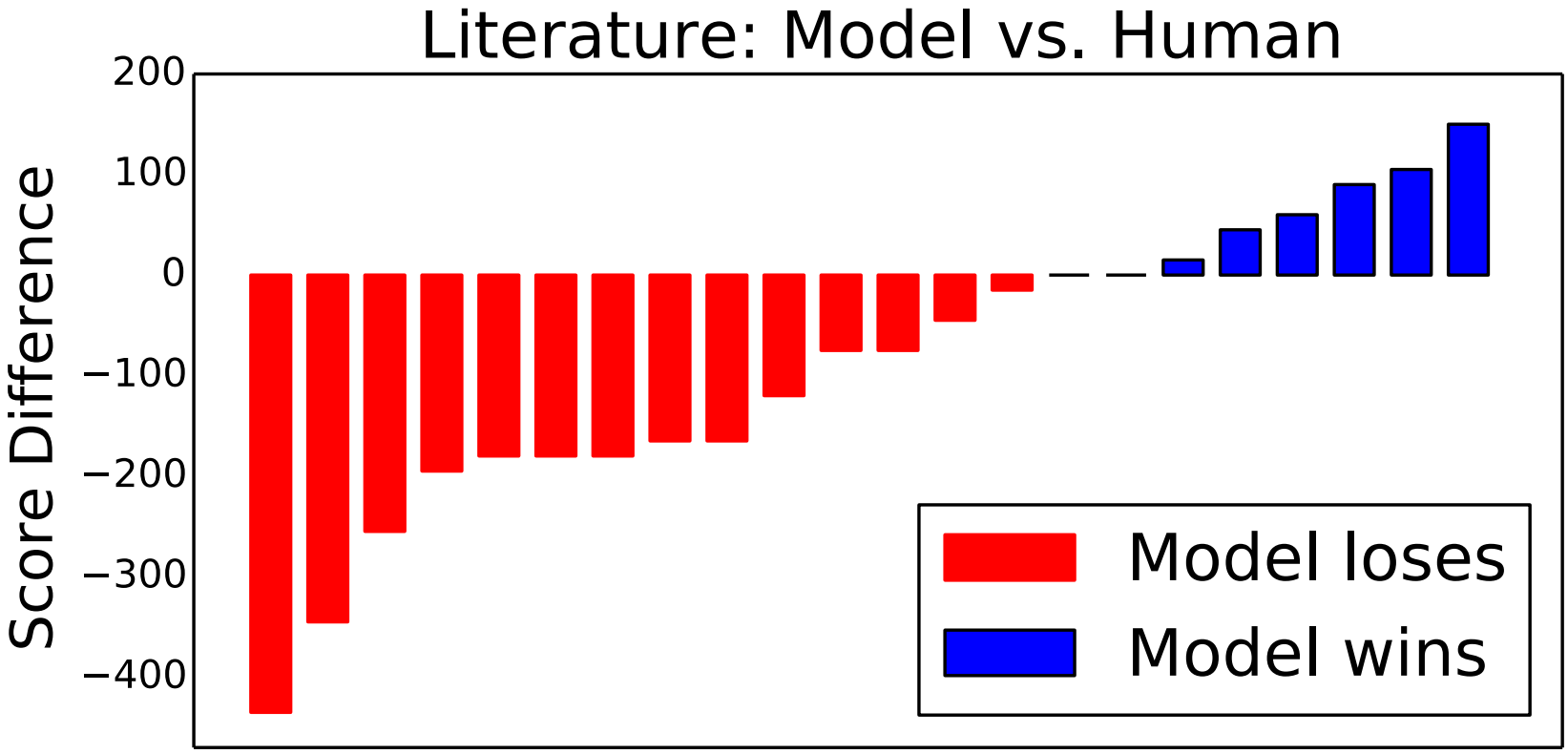
Pushing Facts into Entity Vectors



Qanta Model Can Defeat Human Players

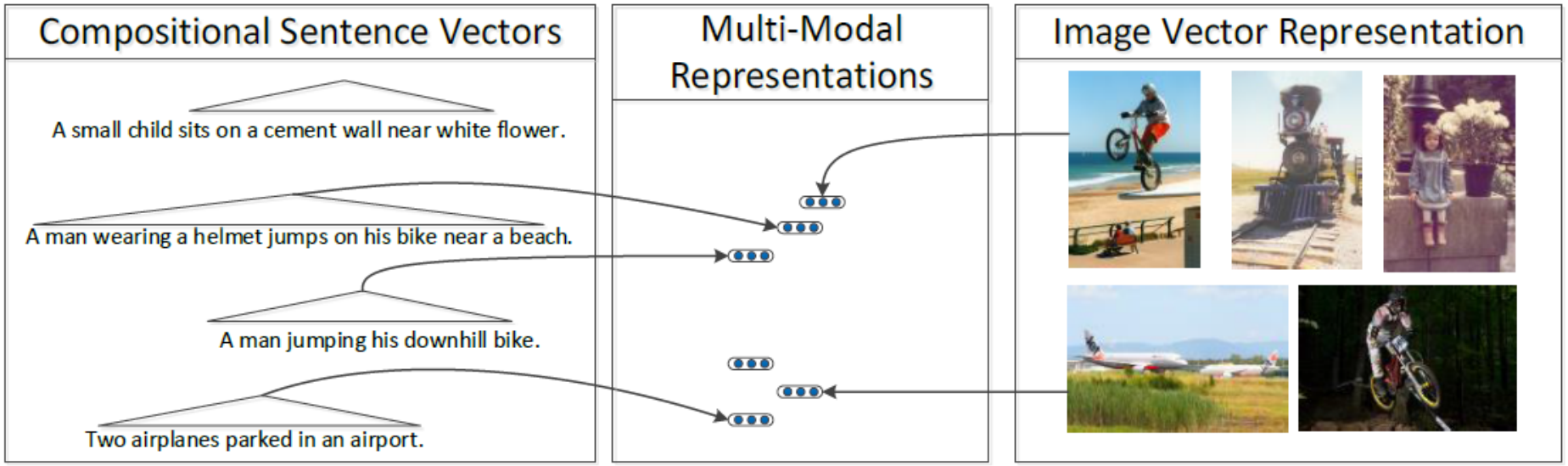


Literature Questions are Hard!



Visual Grounding

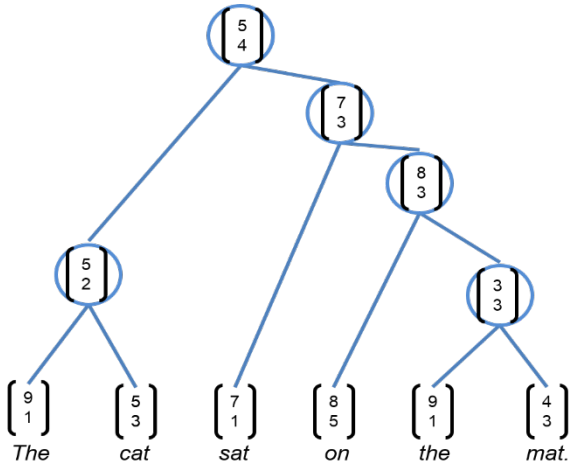
- Idea: Map sentences and images into a joint space



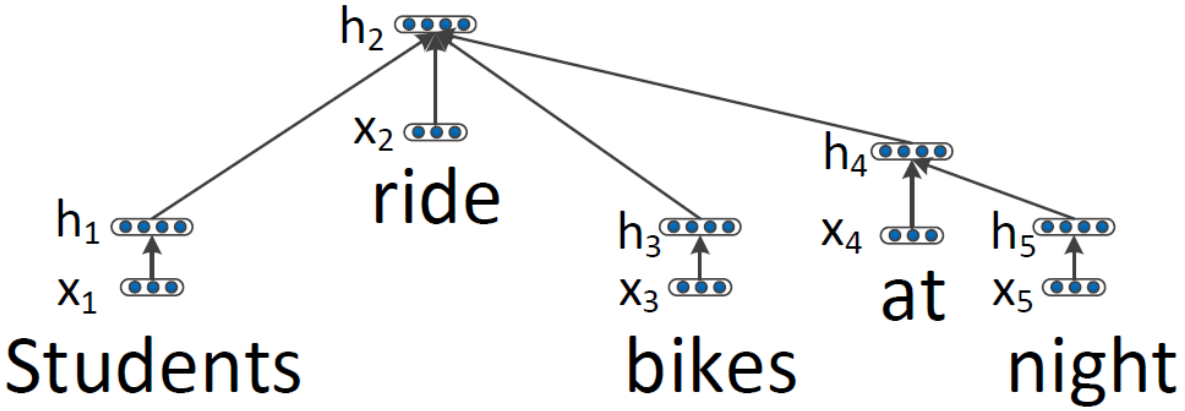
- Socher et al. 2013:
  Grounded Compositional Semantics for Finding and Describing Images with Sentences

Discussion: Compositional Structure

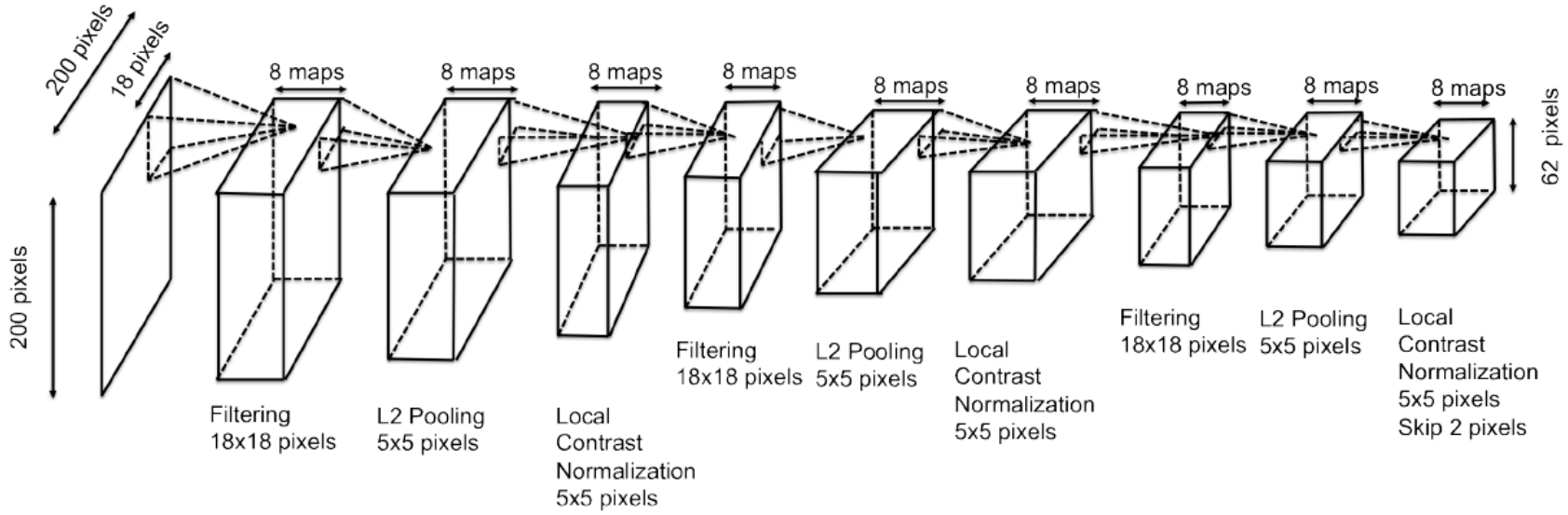
- Recursive Neural Networks so far used constituency trees which results in more syntactically influenced representations



- Instead: Use dependency trees which capture more semantic structure



Convolutional Neural Network for Images



- CNN trained on ImageNet (Le et al. 2013)
- RNN trained to give large inner products between sentence and image vectors:

$$J(W_I, \theta) = \sum_{(i,j)\in\mathcal{P}} \sum_{c\in\mathcal{S}\setminus\mathcal{S}(i)} \max(0, \Delta - v_i^T y_j + v_i^T y_c)$$

Results

A gray convertible sports car is parked in front of the trees. ✓
A close-up view of the headlights of a blue old-fashioned car. ✗
Black shiny sports car parked on concrete driveway. ✓
Five cows grazing on a patch of grass between two roadways. ✗

A jockey rides a brown and white horse in a dirt corral. ✓
A young woman is riding a Bay hose in a dirt riding-ring. ✗
A white bird pushes a miniature teal shopping cart. ✗
A person rides a brown horse. ✓

A motocross bike with rider flying through the air. ✓
White propeller plane parked in middle of grassy field. ✗
The white jet with its landing gear down flies in the blue sky. ✗
An elderly woman catches a ride on the back of the bicycle. ✗

Results

People in an outrigger canoe sail on emerald green water ✗
Two people sailing a small white sail boat. ✓
behind a cliff, a boat sails away ✗
Tourist move in on Big Ben on a typical overcast London day. ✗

A group of people sitting around a table on a porch. ✗
A group of four people walking past a giant mushroom. ✗
A man and women smiling for the camera in a kitchen. ✗
A group of men sitting around a table drinking while a man behind stands pointing. ✓

Describing Images	Mean Rank
Random	92.1
Bag of Words	21.1
CT-RNN	23.9
Recurrent Neural Network	27.1
Kernelized Canonical Correlation Analysis	18.0
DT-RNN	**16.9**

Image Search	Mean Rank
Random	52.1
Bag of Words	14.6
CT-RNN	16.1
Recurrent Neural Network	19.2
Kernelized Canonical Correlation Analysis	15.9
DT-RNN	**12.5**

Image – Sentence Generation (!)

- Several models came out simultaneously in 2015 that follow up
- Replace recursive neural network with LSTM and instead of only finding vectors they generate the description
- Mostly memorized training sequences (becomes similar again)

- Donahue et al. 2015: Long-term → Recurrent Convolutional Networks for Visual Recognition and Description

- Karpathy and Fei-Fei 2015: Deep Visual-Semantic Alignments for Generating Image Descriptions

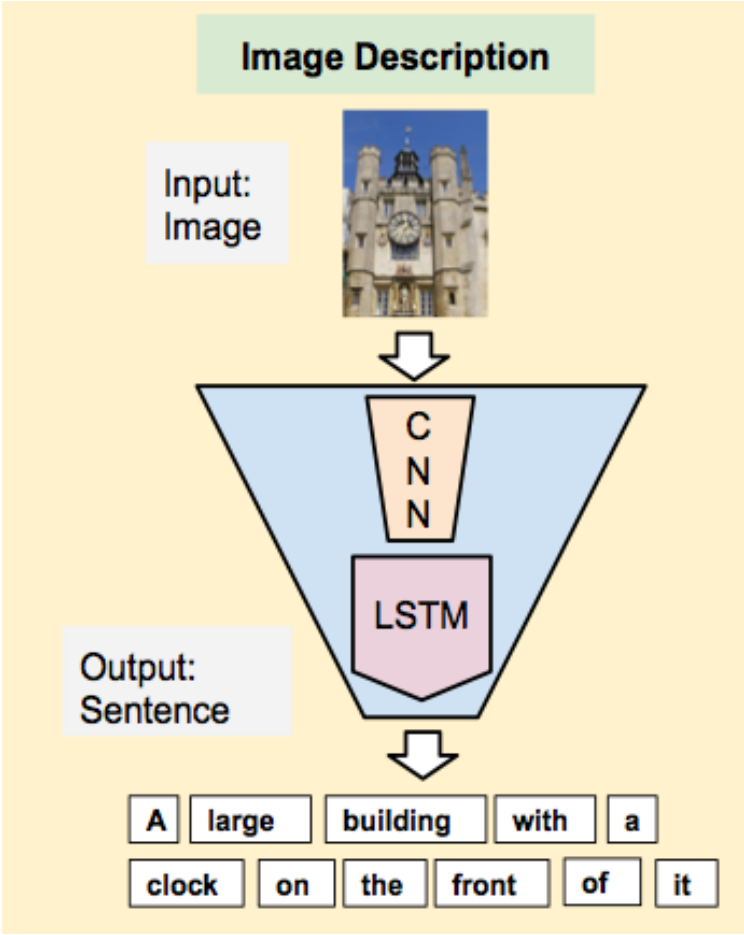


Image – Sentence Generation (!)

"little girl is eating piece of cake."

"baseball player is throwing ball in game."

"woman is holding bunch of bananas."

"a young boy is holding a baseball bat."

"a cat is sitting on a couch with a remote control."

"a woman holding a teddy bear in front of a mirror."

Next Lecture

- The future (?) of deep learning for NLP

- No video