Sentiment Analysis

Announcements

- HW1 is due today at 11:59pm
- HW2 will come out Monday
- If you use late days for HW1, MUST STATE IT IN YOUR SUBMISSION.
- Reading: Today: C 4.2 NLP
- Monday: C 8.1-8.3 Speech and Language,
  8.1 NLP
- Recommend reading chapters in Yoav Goldberg on neural nets and Jurafsky and Martin as well.
  - Both are more intuitive

Today

- Sentiment analysis tasks: definition
- Sentiment resources
- Traditional supervised approach
- Neural net approach

What is sentiment?

- Expression of positive or negative opinions
- .. Towards a topic, person, event, entity
- .. Towards an aspect (e.g., service, food or ambience in a restaurant)

Why sentiment analysis?

- Sentiment is common in online platforms
  - People write about their personal viewpoints
- Useful to understand what people think about political issues, political candidates, important events of the day
- Useful for generating summaries of reviews: restaurants, products, movies

The sentiment analysis task(s)

- Subjective vs objective
- Polarity: Positive, negative or neutral
- Do we have sentiment towards a target? Or aspect based sentiment?
- What/who is the sentiment source?

Subjective vs Objective

- *At several different layers, it's a fascinating tale. ["Who's Spying on Our Computers", George Melloan Wall St Journal.* (Book review)

- *Bell Industries Inc increased its quarterly to 10 cents from 7 cents a share.*

Examples from Weibe et al 2004

Positive/Negative/Neutral

- From UseNet:
- Negative: *I had in mind your facts, Buddy, not hers.*
- Positive: *Nice touch. "Alleges" whatever facts posted are not in your persona of what is "real"*
- Neutral*: March appears to be an estimate while earlier admission cannot be entirely ruled out," according to Chen, also Taiwan's chief WTO negotiator*

Examples from Weibe et al 2004 and Rosenthal 2014

Subjective Phrases

- The foreign ministry said Thursday that it was "***surprised, to put it mildly***" by the U.S. State Department's ***criticism*** of Russia's human rights record and ***objected*** in particular to the "***odious***" section on Chechnya. [Moscow Times, 03/08/2002]

- *Subjectivity analysis identifies text that reveals an author's thoughts, beliefs or other private states.*

Examples from Weibe et al 2004

Subjective Phrases and Sources

- The foreign ministry said Thursday that it was "***surprised, to put it mildly***" by the U.S. State Department's ***criticism*** of Russia's human rights record and ***objected*** in particular to the "***odious***" section on Chechnya. [Moscow Times, 03/08/2002]

- Who was surprised?
- Who was critical?

Examples from Weibe et al 2004

Sentiment towards Target

- *I pretty much enjoyed the whole movie.* Target = whole movie, sentiment = positive.
- *Bulgaria is criticized by the EU because of slow reforms in the judiciary branch, the newspaper notes.* Target = Bulgaria, sentiment = negative
- *Stanishev was elected prime minister in 2005. Since then, he has been a prominent supporter of his country's accession to the EU.* Target = country's access to the EU, sentiment = positive

Examples from Breck & Cardie

Datasets (Sem-eval datasets also used)

Corpus	Average Word Count	Average Character Count	Subjective Phrases	Objective Phrases	Vocabulary Size	Character Length Restrictions
LiveJournal	14.67	66.47	3035 (39%)	4747 (61%)	4747	30-120
MPQA	31.64	176.68	3325 (41%)	4754 (59%)	7614	none
Twitter	25.22	118.55	2091 (36%)	3640 (64%)	8385	0-140
Wikipedia	15.57	77.20	2643 (37%)	4496 (63%)	4342	30-120

2000 sentences in each corpus

MPQA : extensively annotated dataset by Stoyanav, Cardie and Weibe 2004. 15 opinion oriented qustions, 15 fact oriented questions. Along with text spans from 252 articles.

Rosenthal and McKeown 2013)

Example Sentences

LiveJournal	i will have to stick to my canon film slr until in a few years i can afford to upgrade again :)
MPQA	The sale infuriated Beijing which regards Taiwan an integral part of its territory awaiting reunification , by force if necessary.
Twitter	RT @tash jade: That's really sad, Charlie RT "Until tonight I never realised how fucked up I was" - Charlie Sheen #sheenroast
Wikipedia	Perhaps if reported critically by a western source but certainly not by an Israeli source.

Subjective Objective

Sentiment Lexicons

- General Inquirer
- SentiWordNet
- Dictionary of Affect (DAL)

Dictionary of Affect in Language

- Dictionary of 8742 words built to measure the emotional meaning of texts
- Each word is given three scores (scale of 1 to 3)
  - pleasantness - also called evaluation (*ee*)
  - activeness (*aa*)
  - and imagery (*ii*)

$$\mathrm{sub}(c) = \begin{cases} \text{objective} & \text{if } |\sqrt{ee^2 + aa^2}| < \alpha \\ & \text{and } ii > 0 \\ \text{subjective} & \text{otherwise} \end{cases}$$

C. M. Whissel. 1989. **The dictionary of affect in language**. In R. Plutchik and H. Kellerman, editors, Emotion: theory research and experience, volume 4, London. Acad. Press.

Emoticons

- 1000 emoticons were gathered from several lists available on the internet
- We kept the 192 emoticons that appeared at least once and mapped each emoticon to a single word definition

emoticon	:)	:D	<3	:(	;)
definition	happy	laughter	love	sad	wink

Methods

- Pre-processing steps
  - Emoticon keys and contraction expansion
  - Chunker and tagger*
- Lexical Features*
- Syntactic Features*
- Social Media Features

*Apoorv Agarwal, Fadi Biadsy, and Kathleen R. McKeown. 2009. **Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams**. In Proceedings of EACL ’09

Preprocessing

LiveJournal	[i]/NP$_{sub}$ **[will have to stick]/VP$_{obj}$** [to]/PP$_{obj}$ [my canon film slr]/NP$_{obj}$ [until]/PP$_{obj}$ [in]/PP$_{obj}$ [a few years]/NP$_{sub}$ [i]/NP$_{sub}$ [can afford to upgrade]/VP$_{obj}$ [again :)]/NP$_{sub}$
MPQA	[The sale]/NP$_{sub}$ [infuriated]/VP$_{obj}$ [Beijing]/NP$_{obj}$ [which]/NP$_{sub}$ **[regards]/VP$_{sub}$ [Taiwan]/NP$_{obj}$ [an integral part]/NP$_{sub}$ [of]/PP$_{obj}$** [its territory awaiting reunification,]/NP$_{obj}$ [by]/PP$_{obj}$ [force]/NPsub [if]/$_{obj}$ [necessary.]/$_{sub}$
Twitter	**[RT@ tash jade:]/NP$_{obj}$** [That]/Np$_{obj}$ [is]/VP$_{sub}$ [really]/$_{sub}$ [sad,]/$_{sub}$ [Charlie RT]/NP$_{obj}$ [”]/NP$_{obj}$ [Until]/PP$_{obj}$ [tonight]/NP$_{sub}$ [I]/NP$_{sub}$ [never]/$_{sub}$ [realised]/VP$_{sub}$ [how]/$_{sub}$ [fucked]/VP$_{sub}$ [up]/PP$_{obj}$ [I]/NP$_{sub}$ [was]/VP$_{sub}$ [”]/$_{obj}$ [- Charlie Sheen # sheenroast]/NP$_{obj}$
Wikipedia	[Perhaps]/$_{sub}$ [if]/$_{obj}$ [reported]/VP$_{sub}$ **[critically]/$_{sub}$** [by]/PP$_{obj}$ [a western source but]/NP$_{sub}$ [certainly not]/$_{sub}$ [by]/PP$_{obj}$ [an Israeli source.]/NP$_{sub}$

Xuan-Hieu Phan, CRFChunker: **CRF English Phrase Chunker**
http://crfchunker.sourceforge.net/, 2006

Lexical Features

- POS Tags*
- N-grams*
- Performed chi-square feature selection on the n-grams

*Apoorv Agarwal, Fadi Biadsy, and Kathleen R. McKeown. 2009. **Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams**. In Proceedings of EACL ’09

Syntactic Features

- Use the marked up chunks to extract the following:*
  - n-grams: 1-3 words
  - POS: NP, VP, PP, JJ, other
  - Position: target, right, left
  - Subjectivity: subjective, objective
  - Min and max pleasantness

*Apoorv Agarwal, Fadi Biadsy, and Kathleen R. McKeown. 2009. **Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams**. In Proceedings of EACL '09

Social Media Features

Feature	Example
Capital Words	WHAT
Out of Vocabulary	dunno
Emoticons	:)
Acronyms	LOL
Punctuation	.
Repeated Punctuation	#$@.
Punctuation Count	5
Exclamation Points	!
Repeated Exclamations	!!!!
Question Marks	?
Repeated Questions	???
Ellipses	...

Single Corpus Classification

Unbalanced

Experiment	LiveJournal	MPQA	Twitter	Wikipedia
n-gram size	100	2000	none	none
majority	58%	59%	64%	63%
Just DAL	76.5%	75.7%	83.6%	80.4%
Dictionaries+SM	**77.1%**	**76.1%**	84%	**81.4%**
Wordnet	76.7%	75.6%	84%	80.7%
Wordnet+SM	**77.1%**	**76.1%**	**84.2%**	**81.4%**
Dictionaries	76.6%	75.7%	83.9%	80.7%
SM	**77%**	**76.1%**	83.7%	81.2%

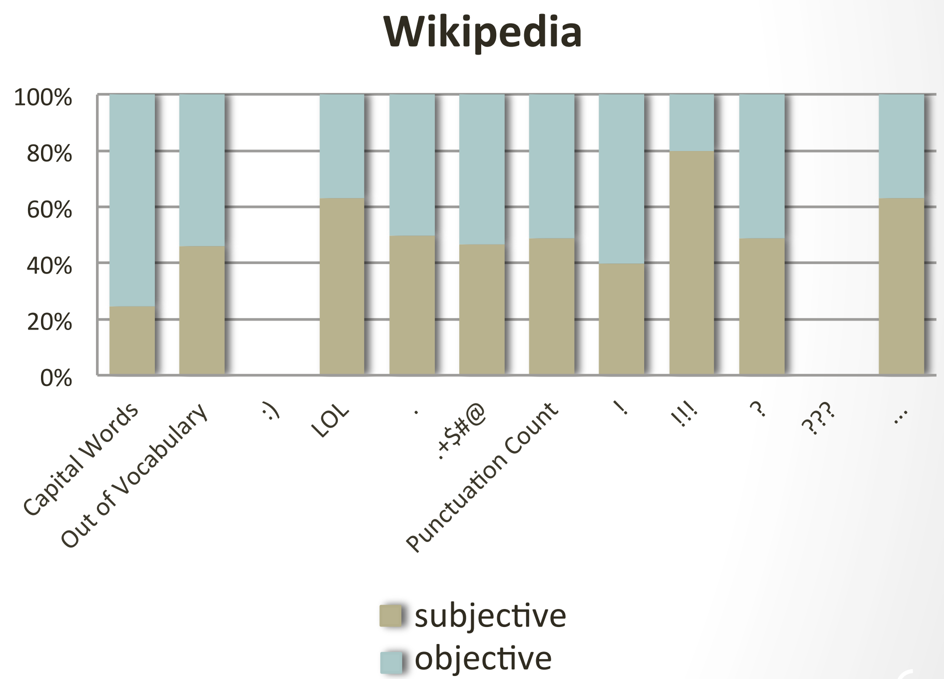
Balanced

Experiment	LiveJournal	MPQA	Twitter	Wikipedia
n-gram size	100	200	none	none
majority	50%	50%	50%	50%
Just DAL	74.7%	75.7%	81.9%	79.3%
Dictionaries+SM	**76.7%**	**76.2%**	**82.6%**	80.2%
Wordnet	75.1%	75.8%	82.4%	79.1%
Wordnet+SM	**76.6%**	75.3%	**82.6%**	80.3%
Dictionaries	75.3%	75.8%	82.4%	79.1%
SM	**76.2%**	**76.3%**	82.2%	**80.4%**

- Logistic Regression i
- 10 runs of 10-fold cross-validation
- Statistical significance using the t-test with p = .001

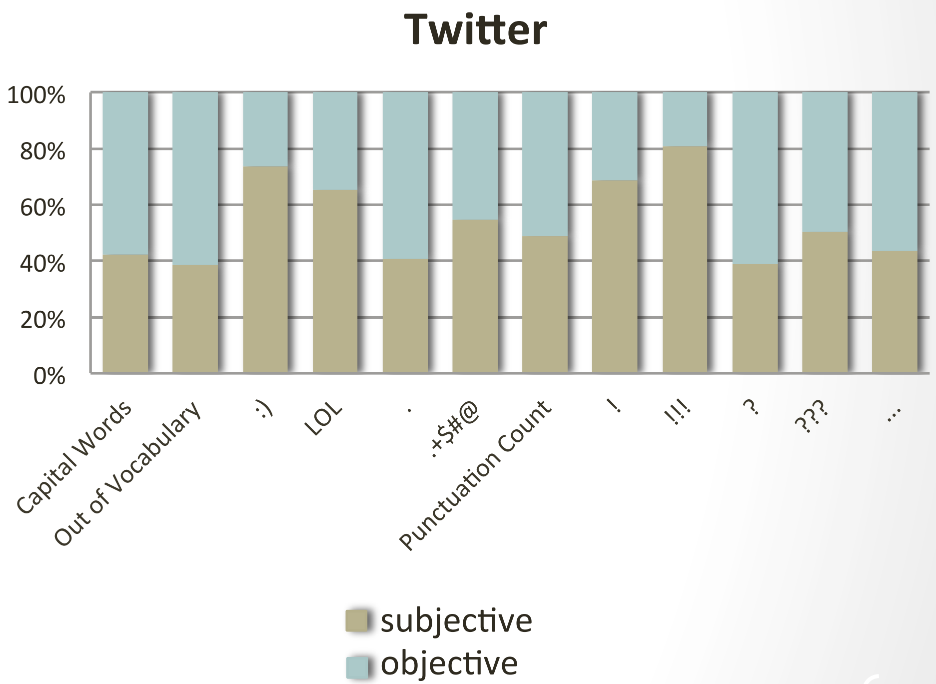
Social Media Error Analysis

- Wikipedia
  - Punctuation was useful as a feature for determining that a phrase is objective if it is a small phrase. However, several subjective phrases were incorrectly classified because of this



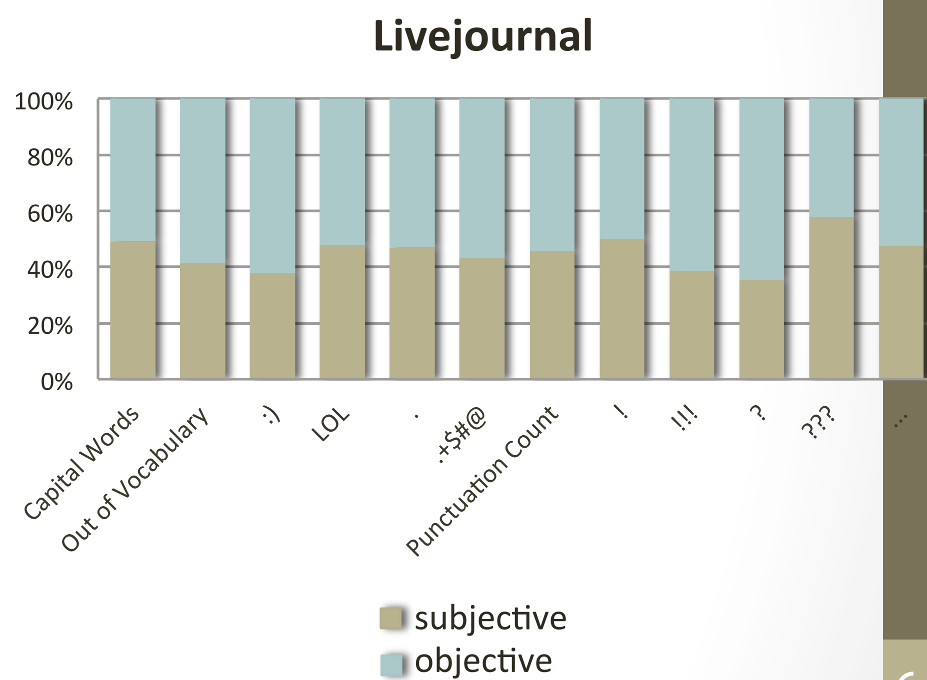
Social Media Error Analysis

- Twitter
  - Ellipses help indicate that a sentence is objective. The accuracy improved from 82% to 92% for sentences with this feature
  - All other social media features were incorrectly classified as objective/subjective depending on the social media preference.



Social Media Error Analysis

- LiveJournal
  - Out of Vocabulary words and punctuation were the most useful social media features.
  - In all datasets the punctuation feature caused close to 50/50 exchange but the feature was best in LiveJournal.



Neural Network Approaches to Sentiment

- Take a standard RNN
- Take a labeled dataset (e.g., IMDB sentiment data set)
- Initialize with pre-trained word embeddings (wordtovec or glove)
- Use sigmoid to predict binary sentiment labels: positive vs negative.

Language is made up of sequences

- So far we have seen embeddings for words
  - (and methods for combining through vector concatenation and arithmetic)
- But how can we account for sequences?
  - Words as sequences of letters
  - Sentences as sequences of words
  - Documents as sequences of sentences

Recurrent Neural Networks

- Represent arbitrarily sized sequences in fixed-size vector

- Good at capturing statistical regularities in sequences (order matters)

- Include simple RNNs, Long short-term memory (LSTMs), Gated Recurrent Unit (GRUs)

Learning word meaning from their morphs

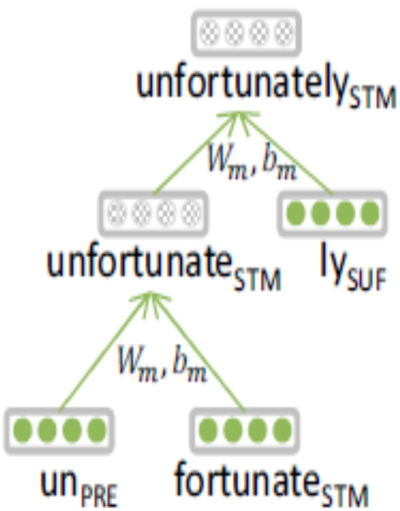


Figure 1: **Morphological Recursive Neural Network.** A vector representation for the word "unfortunately" is constructed from morphemic vectors: $un_{\mathtt{pre}}$, $fortunate_{\mathtt{stm}}$, $ly_{\mathtt{suf}}$. Dotted nodes are computed on-the-fly and not in the lexicon.

[Thang et al. 2013]

Logical entailment using compositional semantics via RNNs

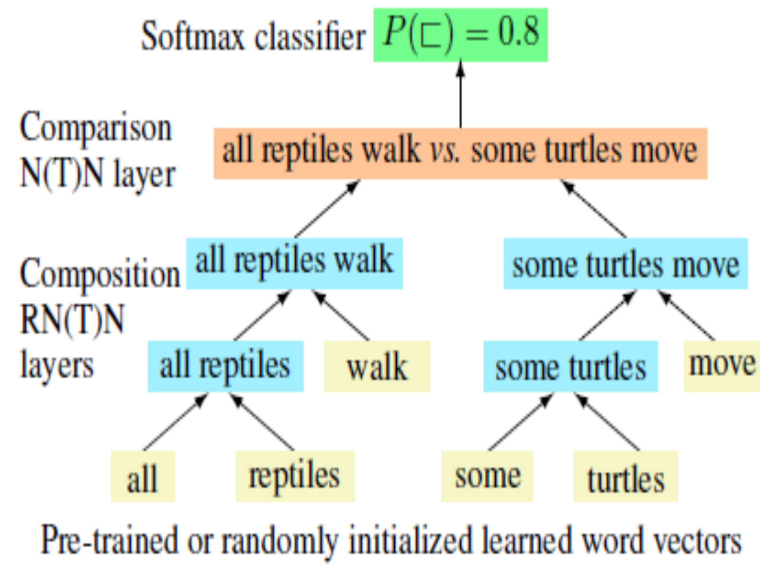


Figure 1: In our model, two separate tree-structured networks build up vector representations for each of two sentences using either NN or NTN layer functions. A comparison layer then uses the resulting vectors to produce features for a classifier.

[Bowman et al. 2014]

Machine Translation (Sequences)

- Sequence-to-sequence
  - Sutskever et al. 2014

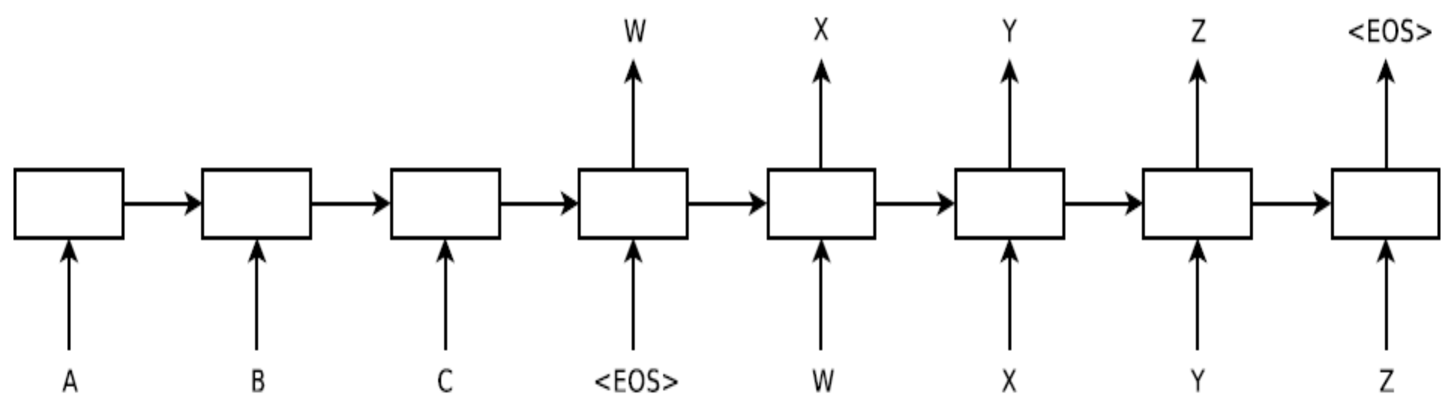


Figure 1: Our model reads an input sentence "ABC" and produces "WXYZ" as the output sentence. The model stops making predictions after outputting the end-of-sentence token. Note that the LSTM reads the input sentence in reverse, because doing so introduces many short term dependencies in the data that make the optimization problem much easier.

RNN Abstraction

- RNN is a function that takes an arbitrary length sequence as input and returns a single d_{out} dimensional vector as output
  - Input: $x_{1:n} = x_1\ x_{2\ \ldots}\ x_n$ (x_i ε $R^{d\text{-}in}$)
  - Output: y_n ε $R^{d\text{-}out}$

$$\mathbf{y_n} = \mathrm{RNN}(\mathbf{x_{1:n}})$$

$$\mathbf{x_i} \in \mathbb{R}^{d_{\mathrm{in}}} \quad \mathbf{y_n} \in \mathbb{R}^{d_{\mathrm{out}}}$$

$$\mathbf{y_{1:n}} = \mathrm{RNN}^{\star}(\mathbf{x_{1:n}})$$

$$\mathbf{y_i} = \mathrm{RNN}(\mathbf{x}_{1:i})$$

$$\mathbf{x_i} \in \mathbb{R}^{d_{\mathrm{in}}} \quad \mathbf{y_i} \in \mathbb{R}^{d_{\mathrm{out}}}$$

Output vector y used for further prediction

RNN Characteristics

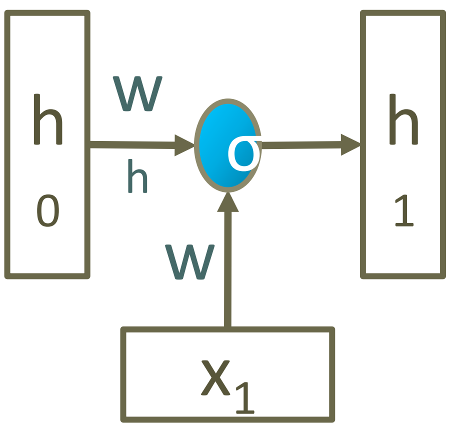
- Can condition on the entire sequence without resorting to the Markov assumption
- Can get very good language models as well as good performance on many other tasks

RNNs are defined recursively

- By means of a function R taking as input a state vector h_{i-1} and an input vector x_i
- Returns a new state vector h_i
- The state vector can be mapped to an output vector y_i using a simple deterministic function
- And fed through softmax for classification.

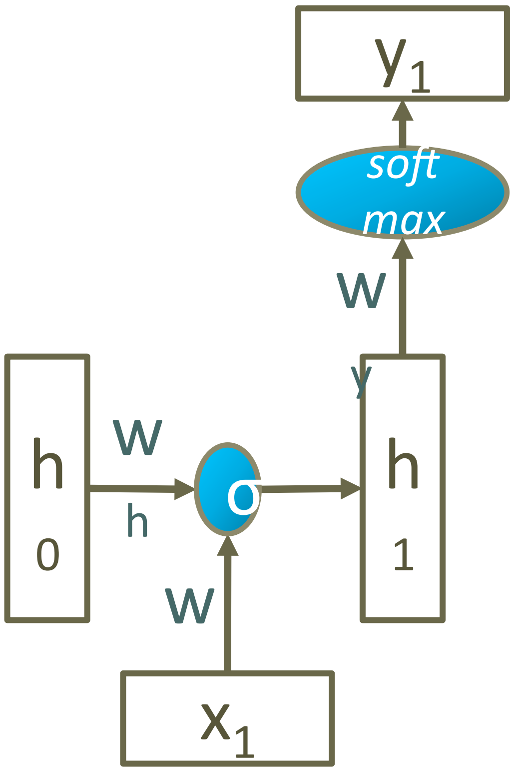
Recurrent Neural Networks

$$h_t = \sigma(W_h h_{t-1} + W_x x_t)$$



Slide from Radev

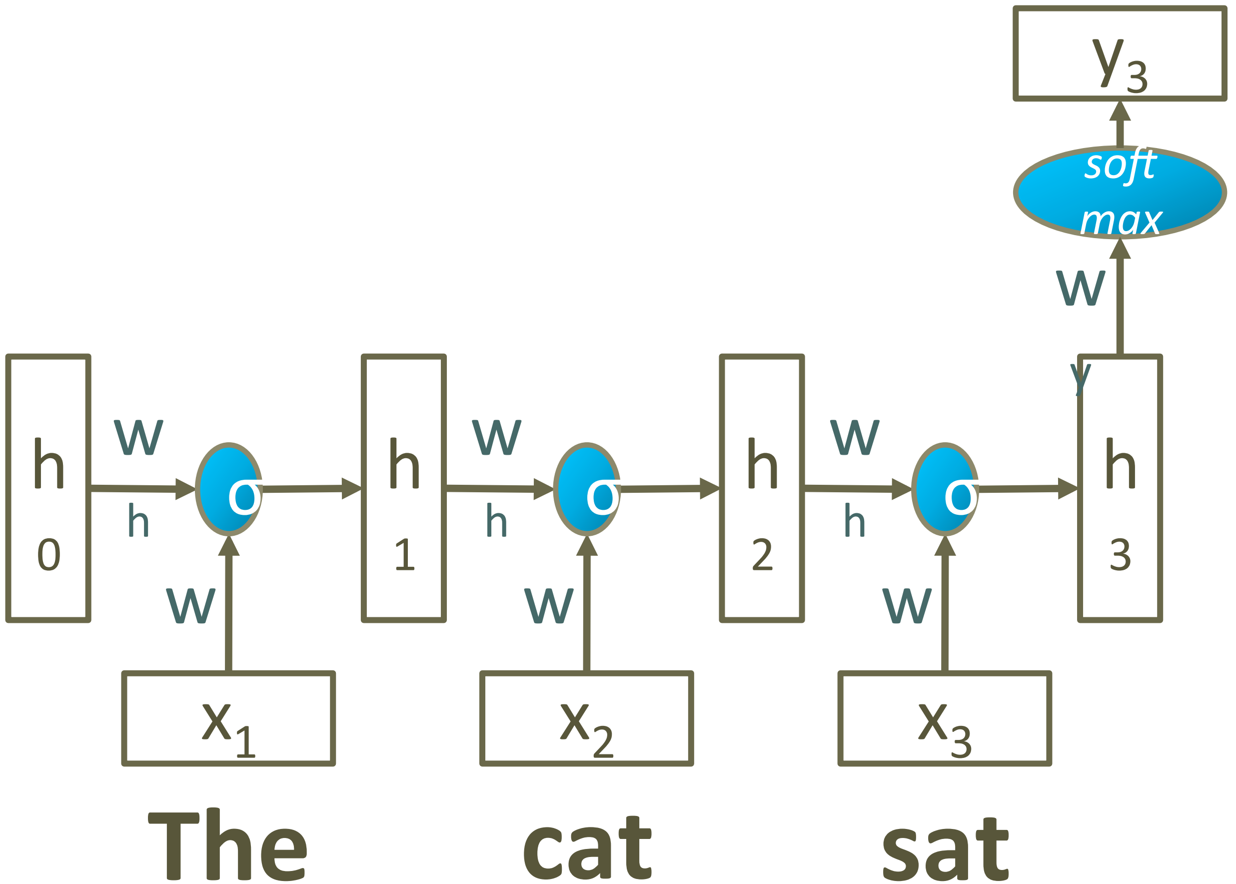
RNN



$$h_t = \sigma(W_h h_{t-1} + W_x x_t)$$
$$y_t = softmax(W_y h_t)$$

Slide from Radev

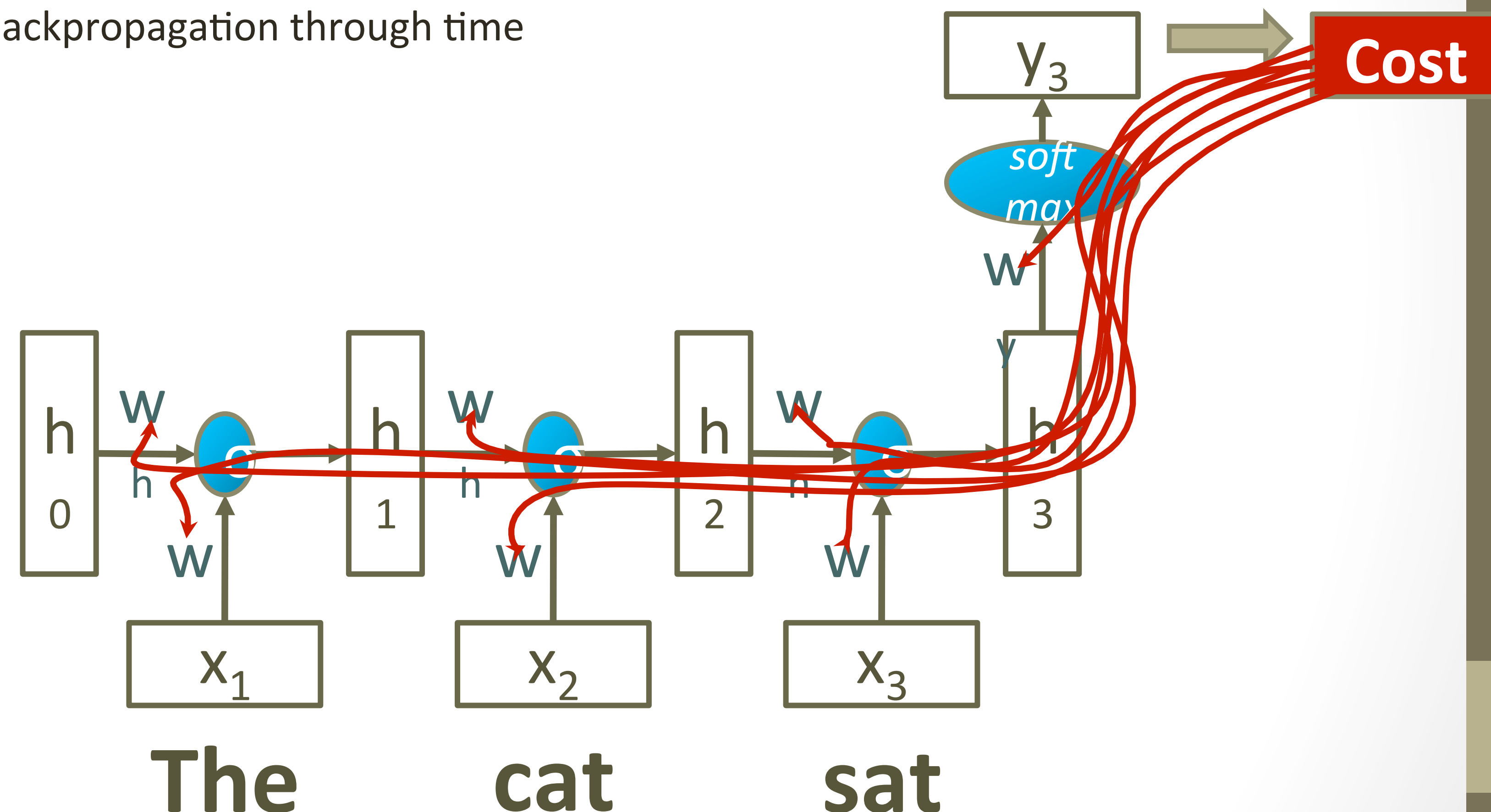
RNN



Slide from Radev

Updating Parameters of an RNN

Backpropagation through time



Slide from Radev

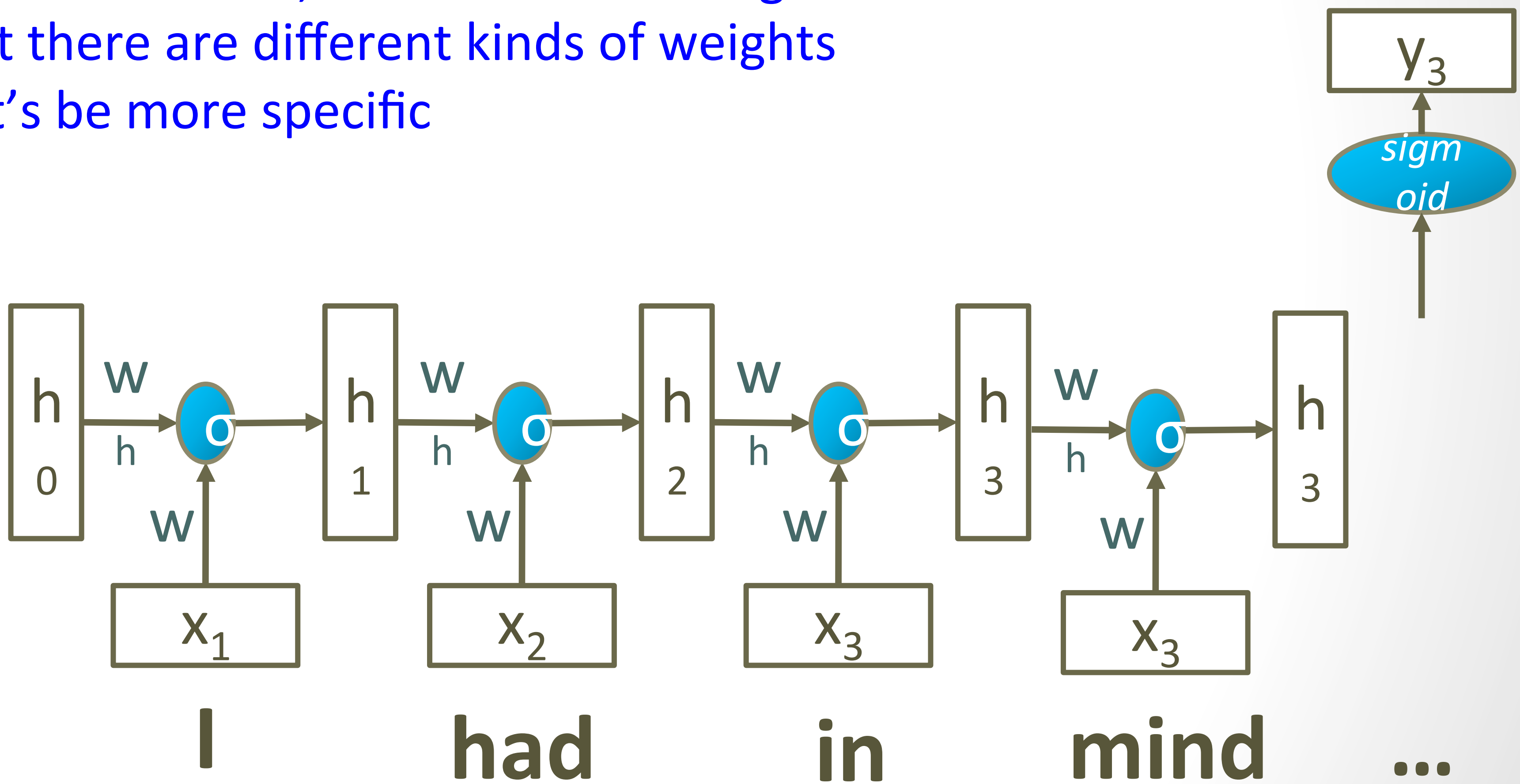
Example

- For each sentence in the training corpus, classify, compare to gold standard and compute loss, backpropagate.
  - Recall that we may use mini-batches so that we're not back-propagating for each example

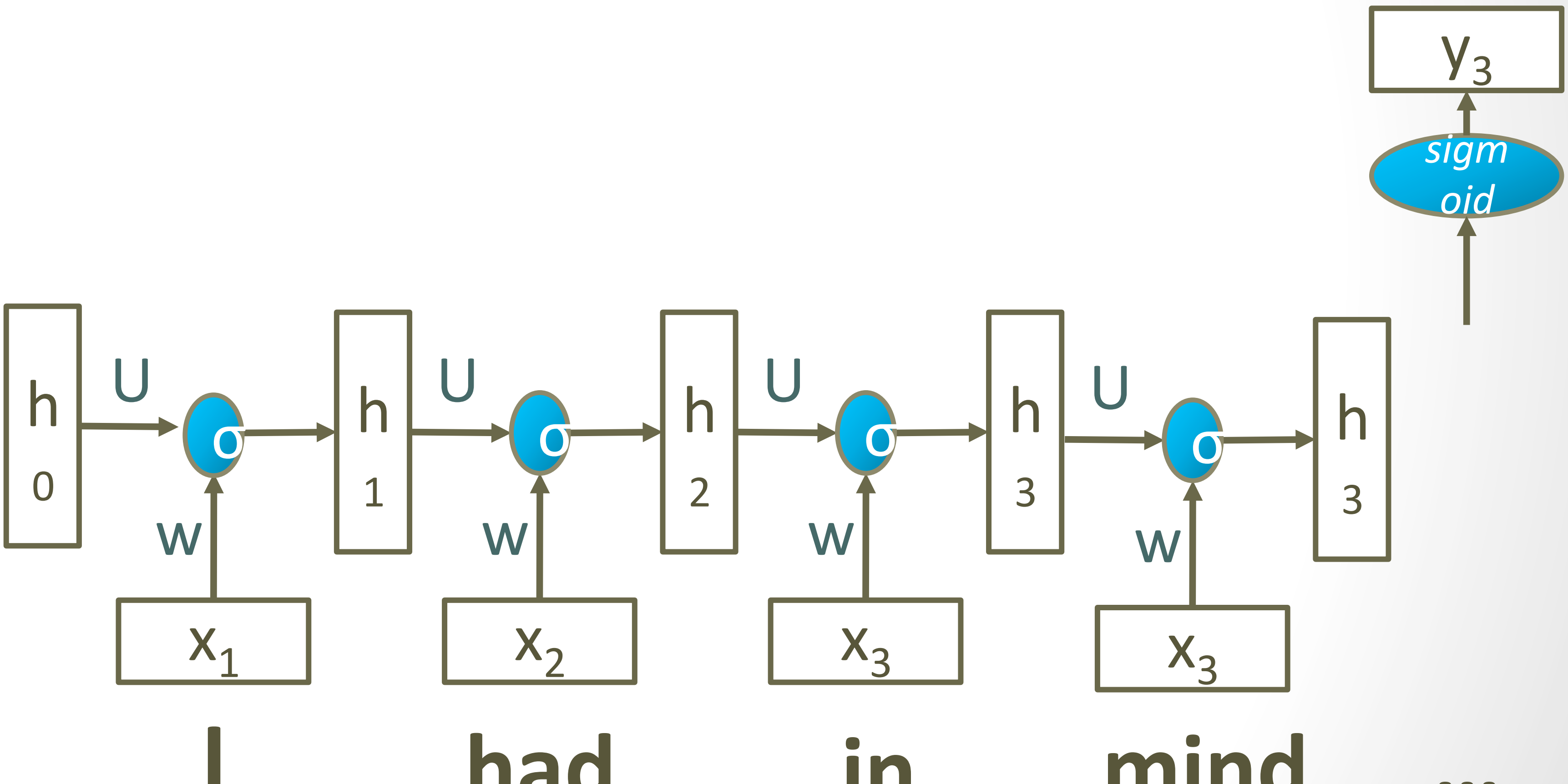
- *I had in mind your facts, Buddy, not hers.*

RNN – I had in mind your facts, buddy, not hers.

In this overview, w refers to the weights
But there are different kinds of weights
Let's be more specific



RNN – I had in mind your facts, buddy, not hers.



What is W?

The embedding for word x_i

The matrix of embeddings for all words

None of the above

What is h?

the previous state

the hidden layer

weights for the previous state

None of the above

What is x_t*W

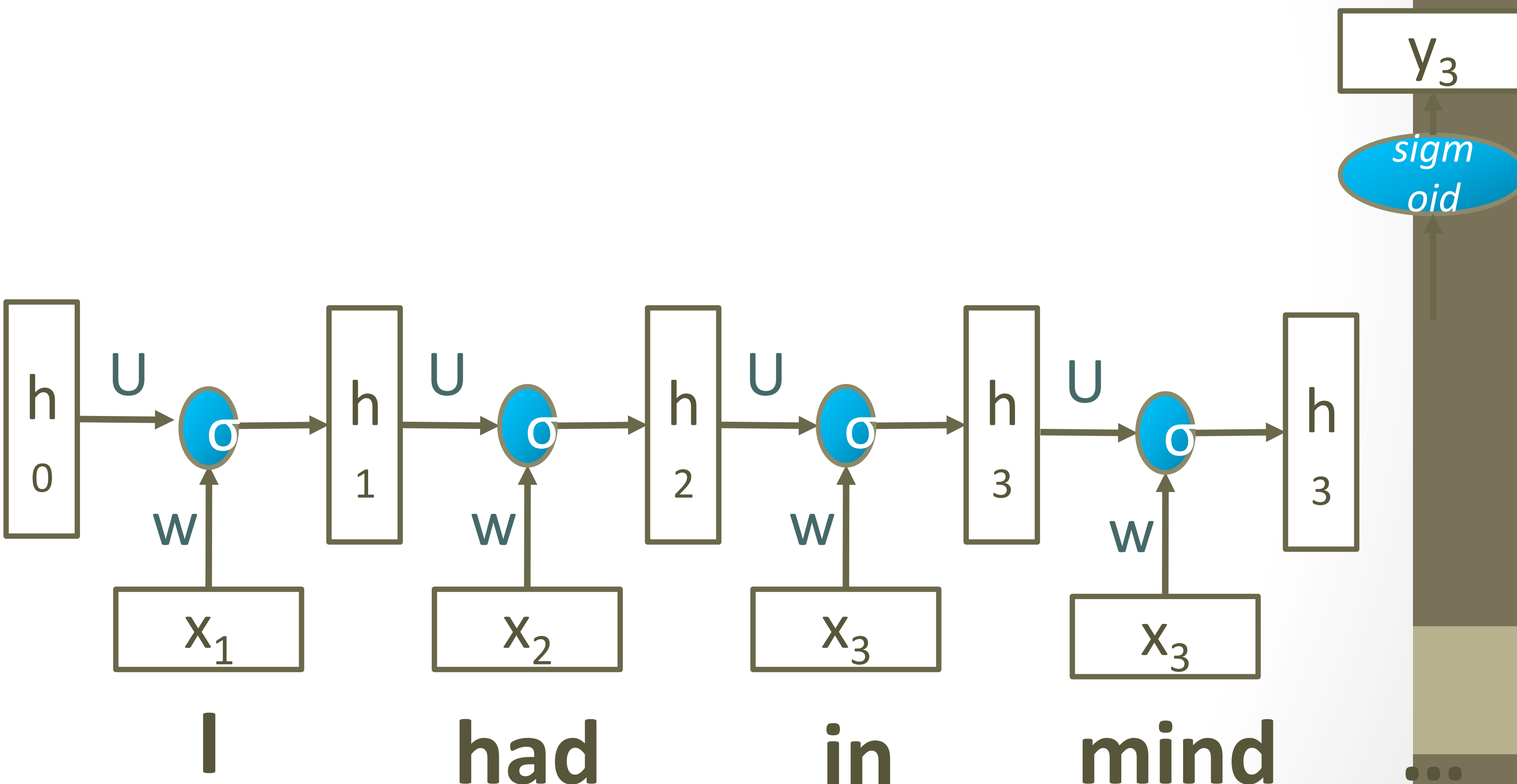
the hidden layer

the word embedding for x_t

the previous state

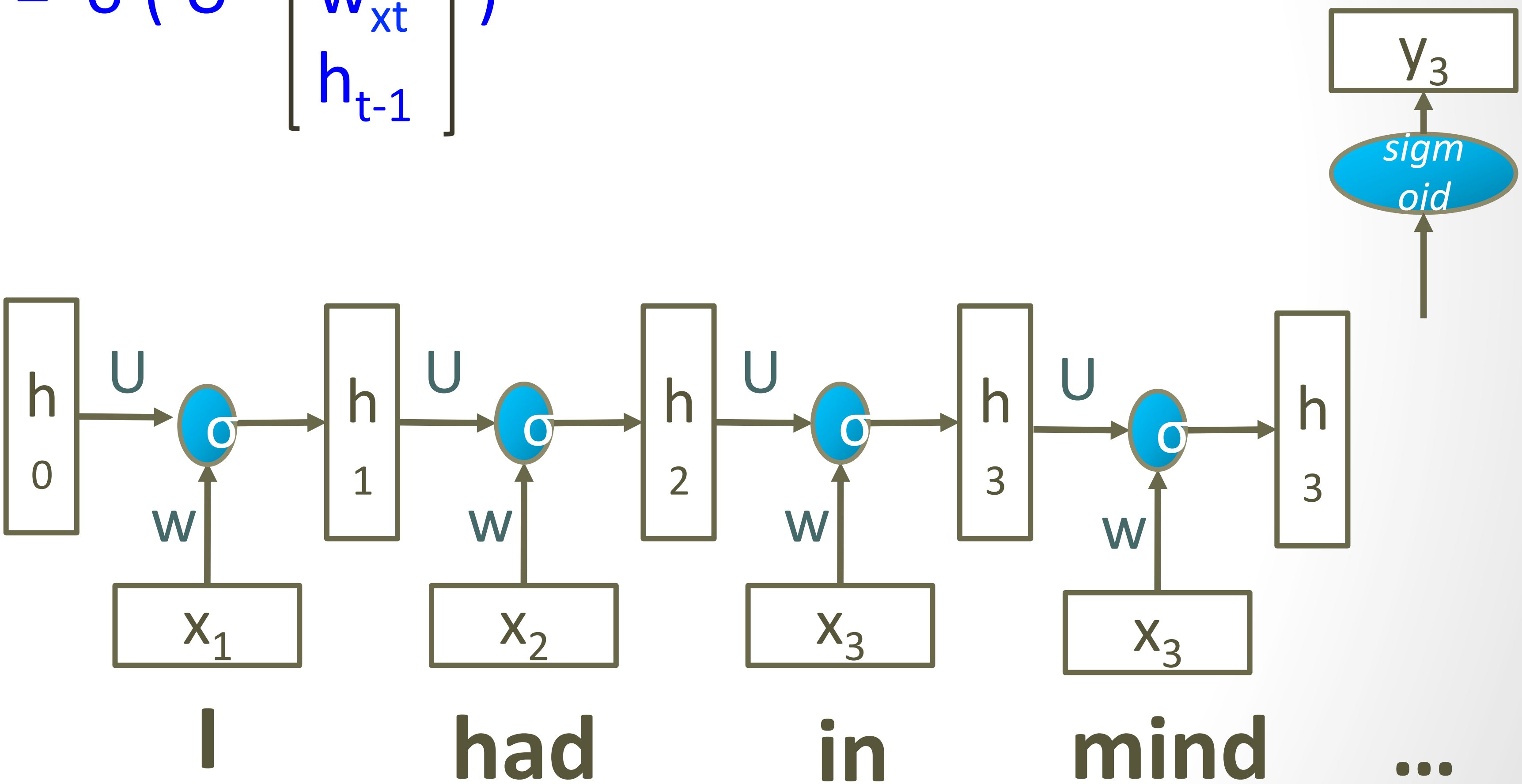
None of the above

RNN – I had in mind your facts, buddy, not hers.



RNN – I had in mind your facts, buddy, not hers.

$$h_t = \sigma \left(U \begin{bmatrix} w_{xt} \\ h_{t-1} \end{bmatrix} \right)$$



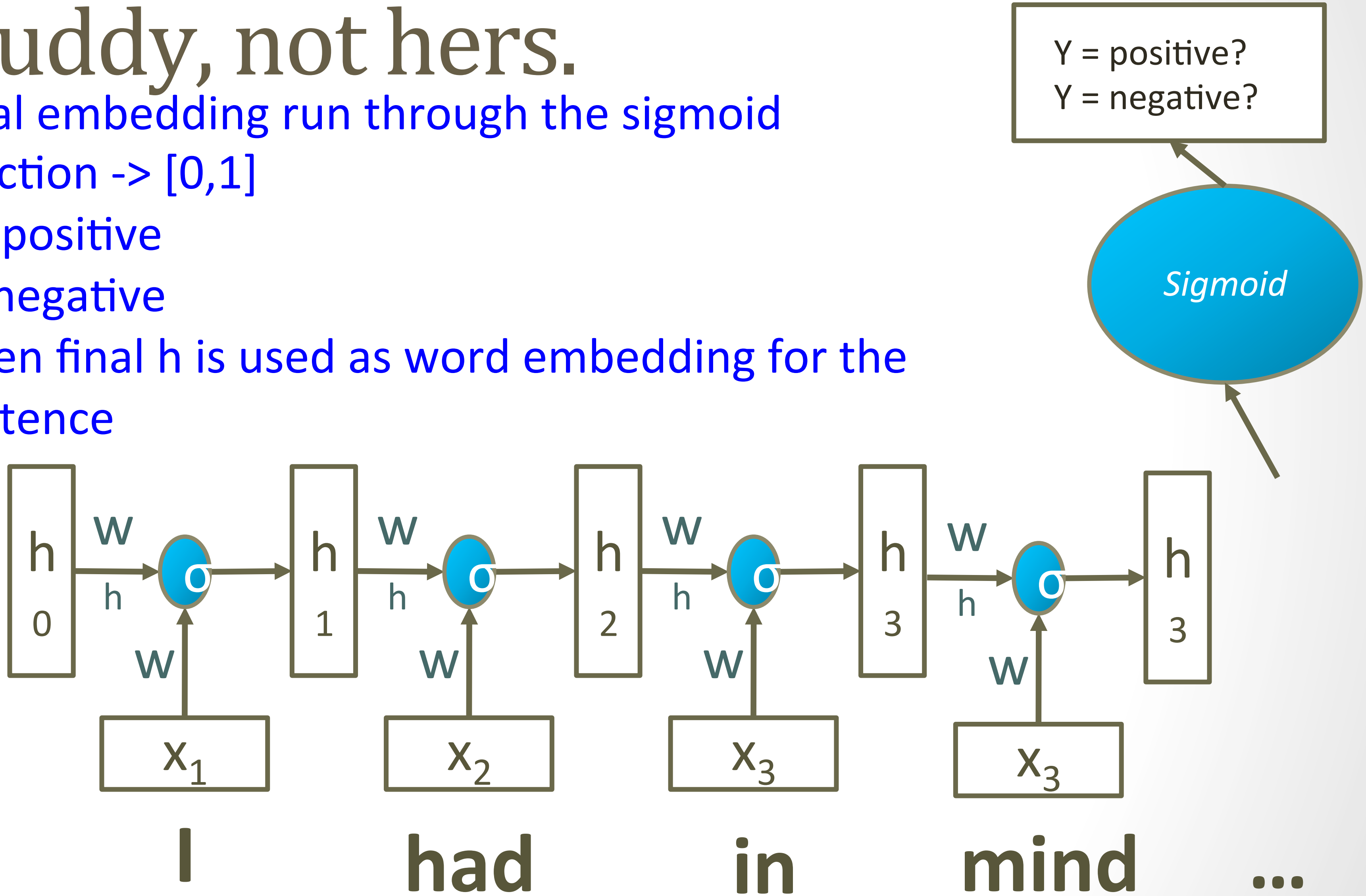
RNN – I had in mind your facts, buddy, not hers.

Final embedding run through the sigmoid function -> [0,1]

1 = positive

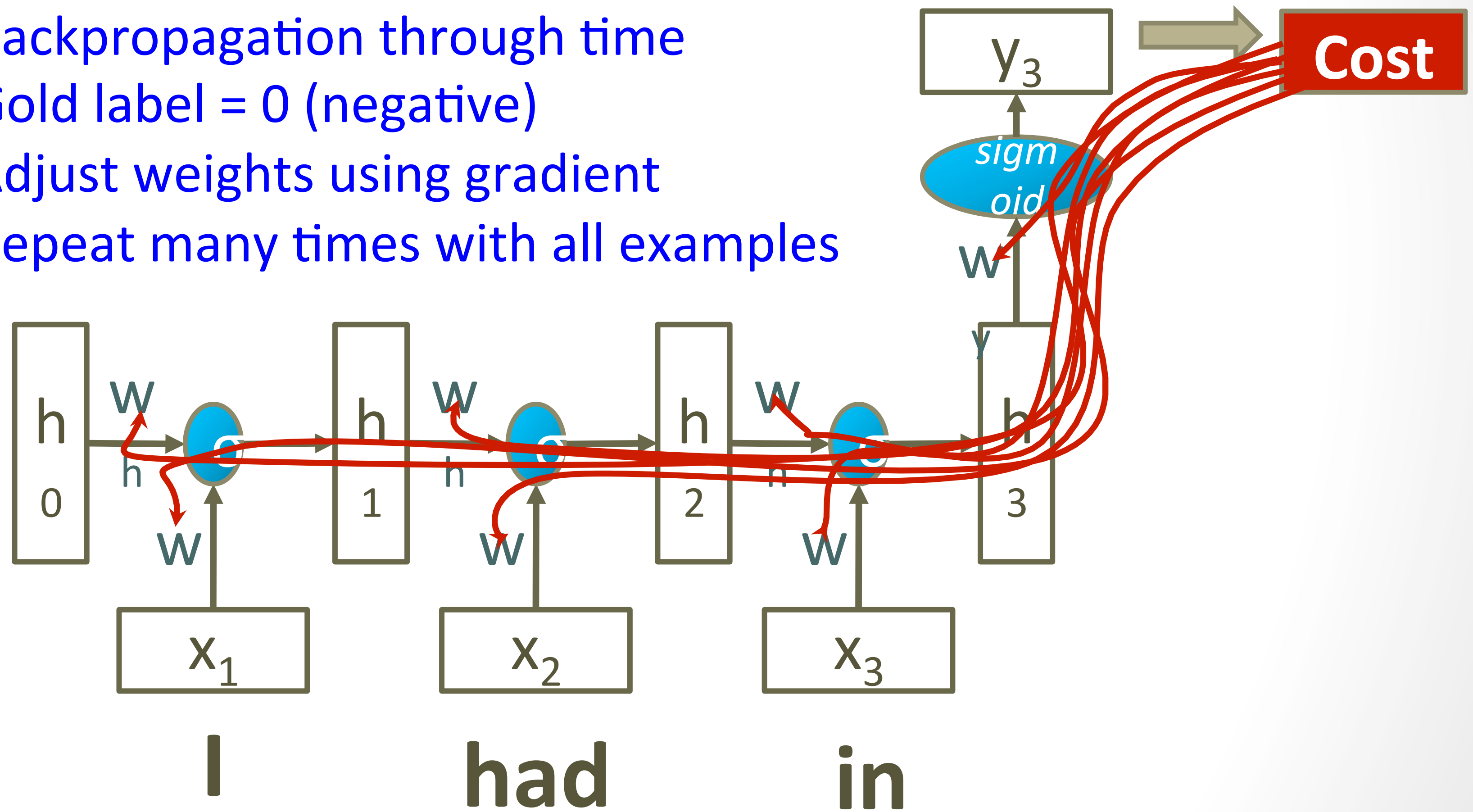
0= negative

Often final h is used as word embedding for the sentence



Updating Parameters of an RNN

Backpropagation through time
Gold label = 0 (negative)
Adjust weights using gradient
Repeat many times with all examples



Slide from Radev

Problem with RNN

- Vanishing gradients
- By the time we back-propogate all the way through the network, the weights approach zero -> vanishing gradient
- Error signals (gradients) in later steps diminish quickly and do not reach earlier input signals
    - -> Hard to capture long-distance dependencies

What is a long distance dependency?

- The students were listening to Kathy McKeown speak.
- The students in 451 CS Building were listening to Kathy McKeown speak

Gated Architectures

- RNN: at each state of the architecture, the entire memory state (h) is read and written
- Gate = binary vector g ε {0,1}
    - Controls access to n-dimensional vector x•g
- Consider $$s' \leftarrow g \odot x + (1 - g) \odot (s)$$
    - Reads entries from x specified by g
    - Copies remaining entries from s (or h as we've been labeling the hidden state)

$$\underset{\mathbf{s}'}{\begin{bmatrix} 8 \\ 11 \\ 3 \\ 7 \\ 5 \\ 15 \end{bmatrix}} \leftarrow \underset{\mathbf{g}}{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}} \odot \underset{\mathbf{x}}{\begin{bmatrix} 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{bmatrix}} + \underset{(\mathbf{1}-\mathbf{g})}{\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}} \odot \underset{\mathbf{s}}{\begin{bmatrix} 8 \\ 9 \\ 3 \\ 7 \\ 5 \\ 8 \end{bmatrix}}$$

Example: gate copies from positions 2 and 5 in the input
Remaining elements copied from memory

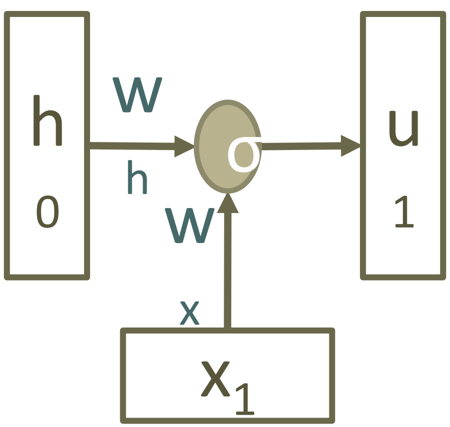
LSTM Solution

- Use memory cell to store information at each time step.
- Use “gates” to control the flow of information through the network.
  - Input gate: protect the current step from irrelevant inputs
  - Output gate: prevent the current step from passing irrelevant outputs to later steps
  - Forget gate: limit information passed from one cell to the next

[slides from Catherine Finegan-Dollak]

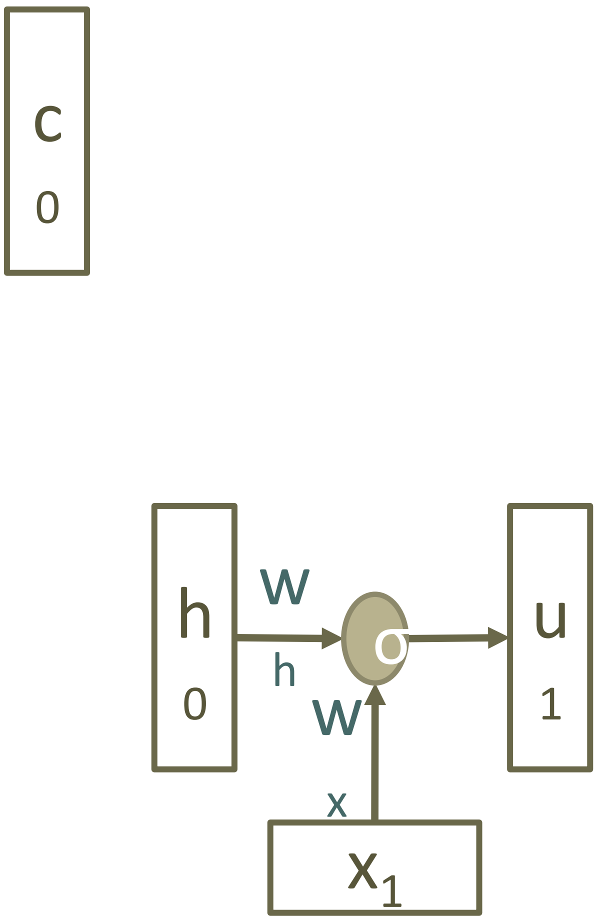
Transforming RNN to LSTM

$$u_t = \sigma(W_h h_{t-1} + W_x x_t)$$



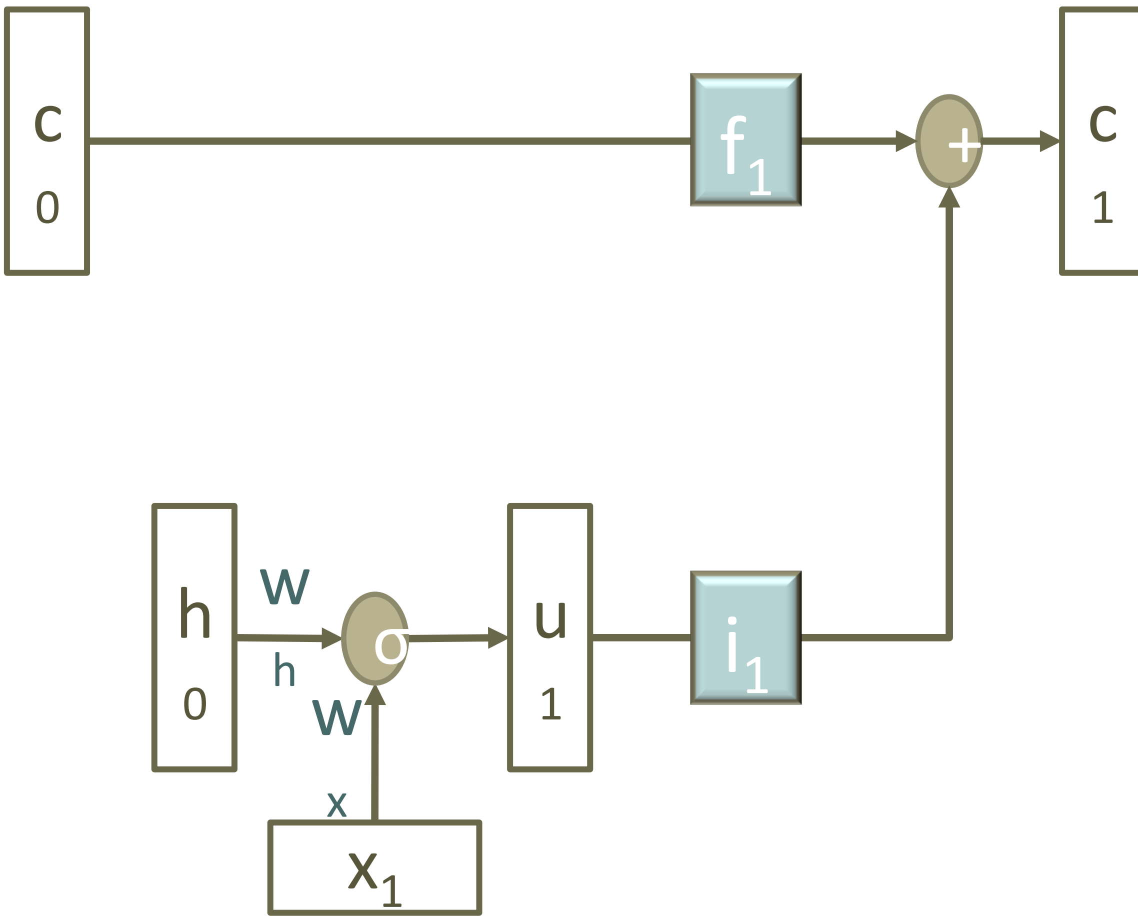
[slides from Catherine Finegan-Dollak]

Transforming RNN to LSTM



[slides from Catherine Finegan-Dollak]

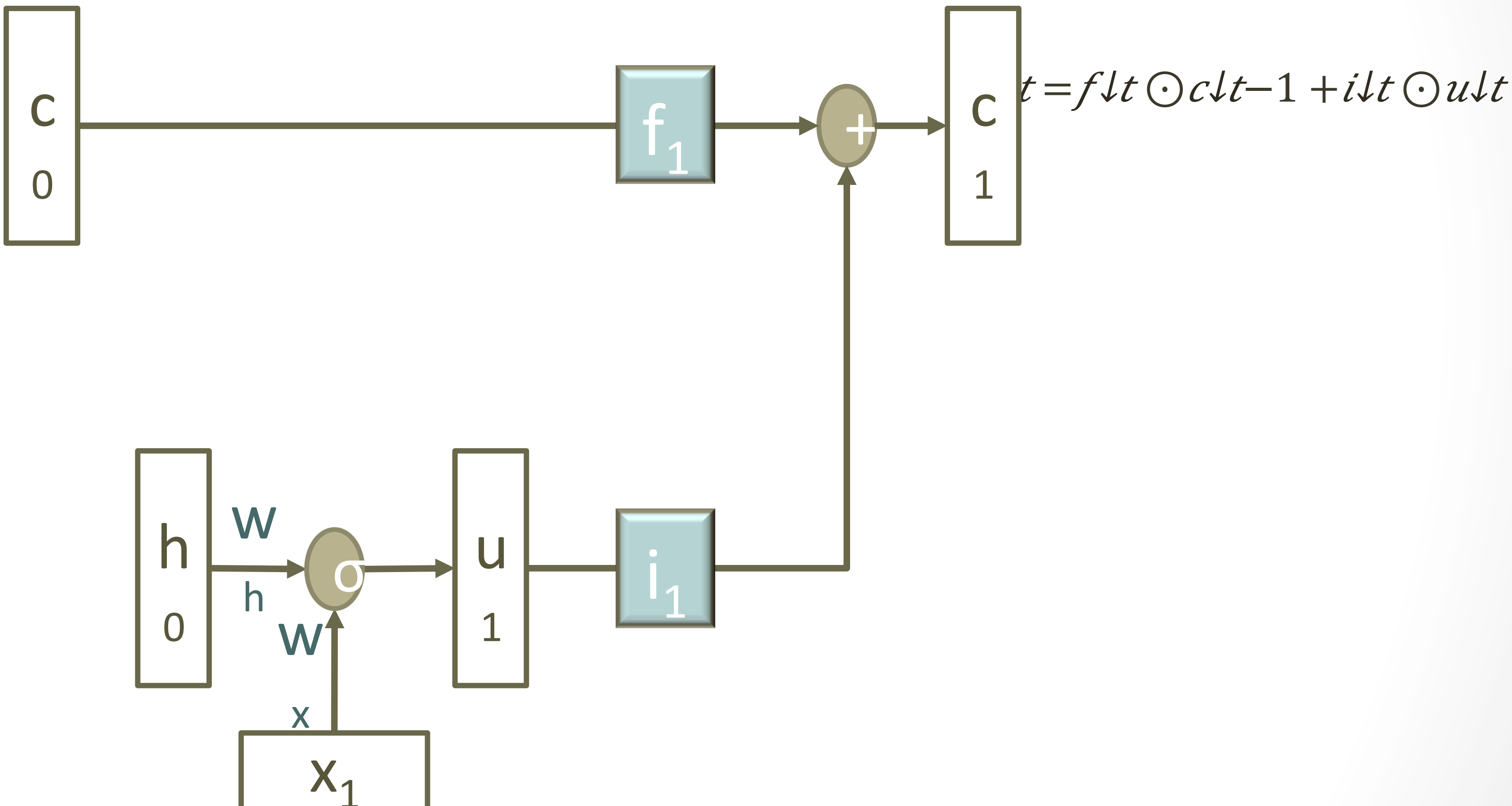
Transforming RNN to LSTM



$$t = f_t \odot c_{t-1} + i_t \odot u_t$$

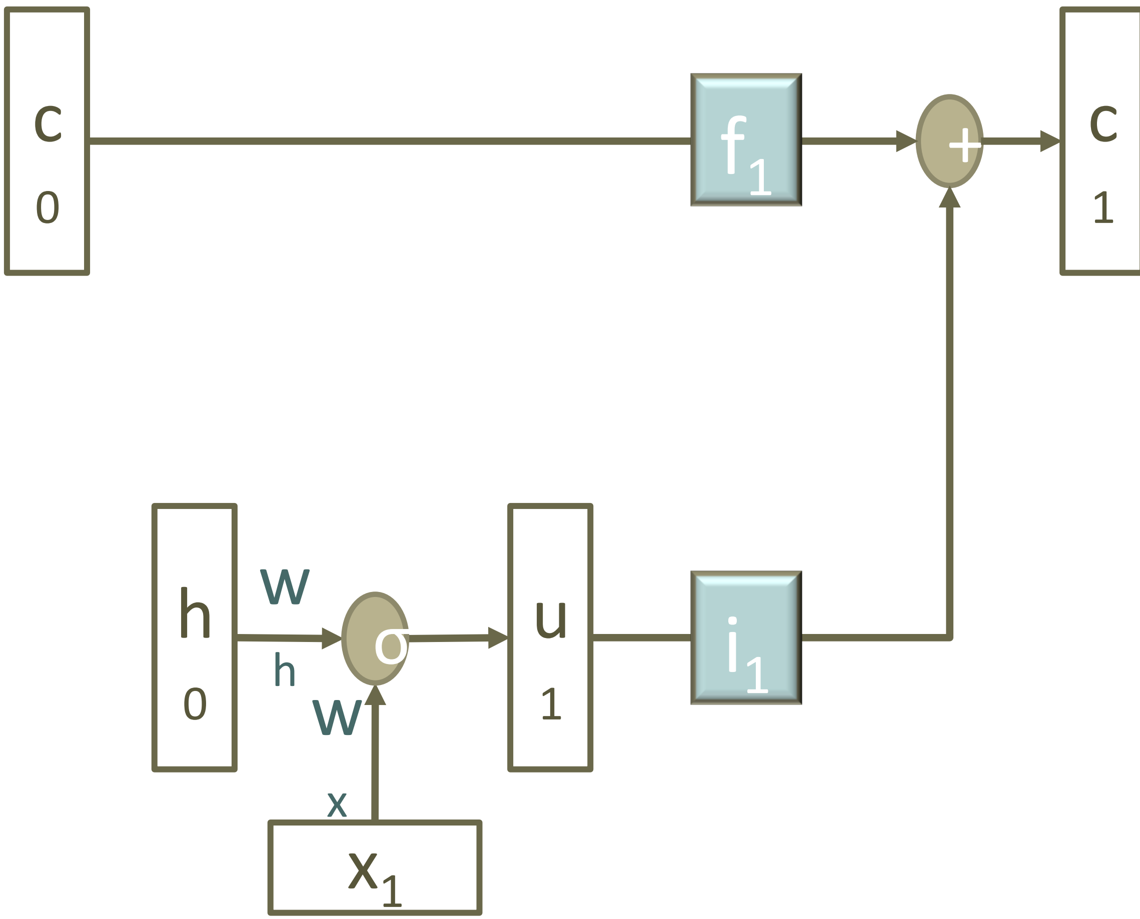
[slides from Catherine Finegan-Dollak]

Transforming RNN to LSTM



[slides from Catherine Finegan-Dollak]

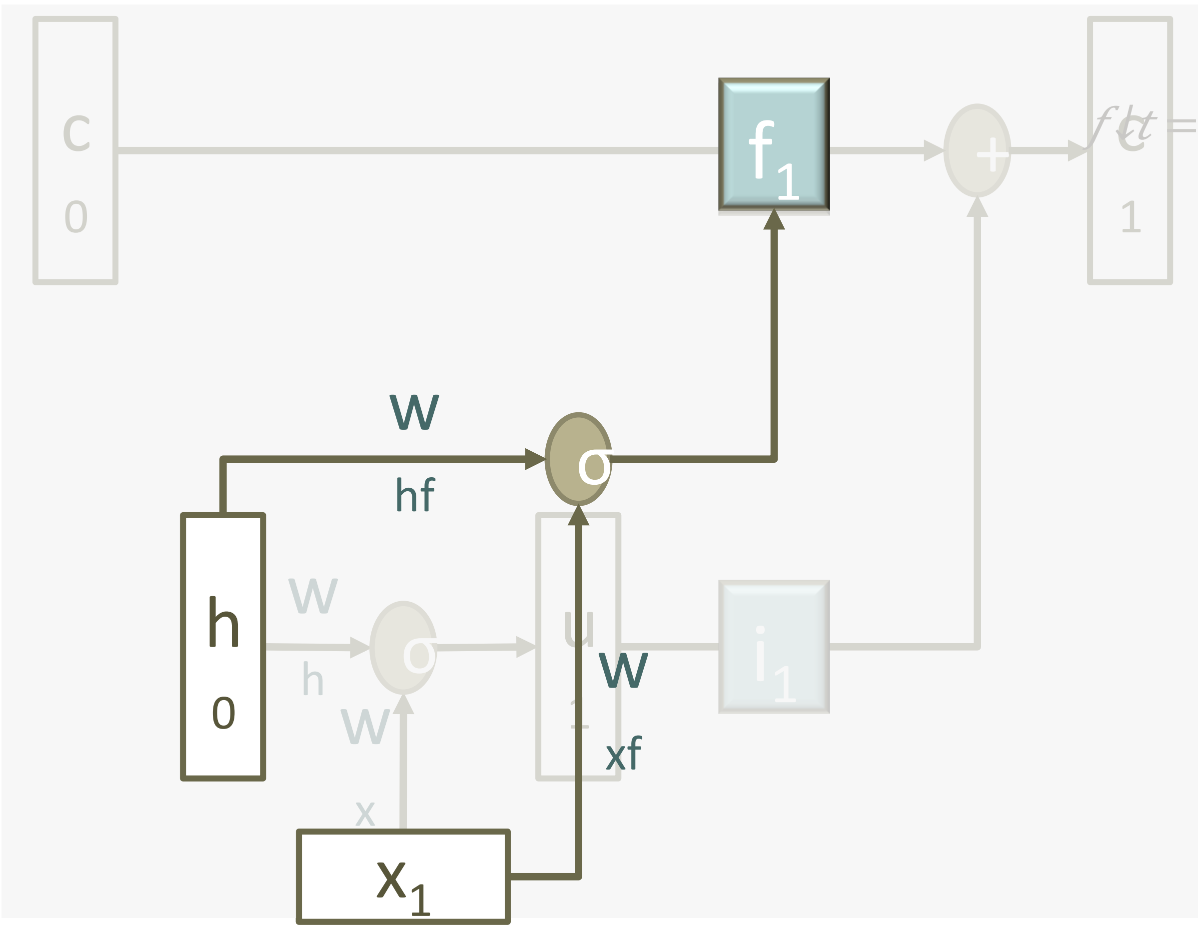
Transforming RNN to LSTM



$t = f_t \odot c_{t-1} + i_t \odot u_t$

[slides from Catherine Finegan-Dollak]

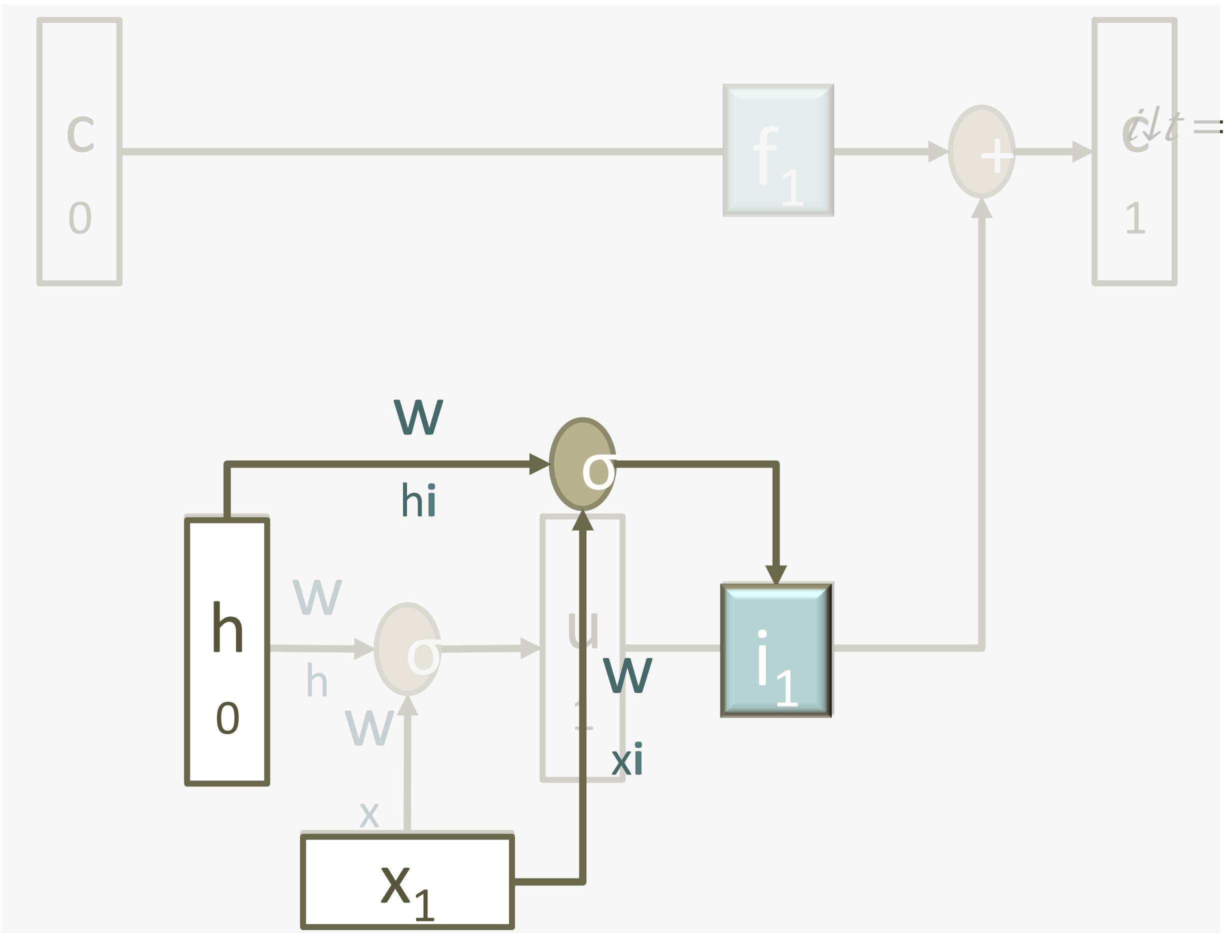
Transforming RNN to LSTM



$f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t)$

[slides from Catherine Finegan-Dollak]

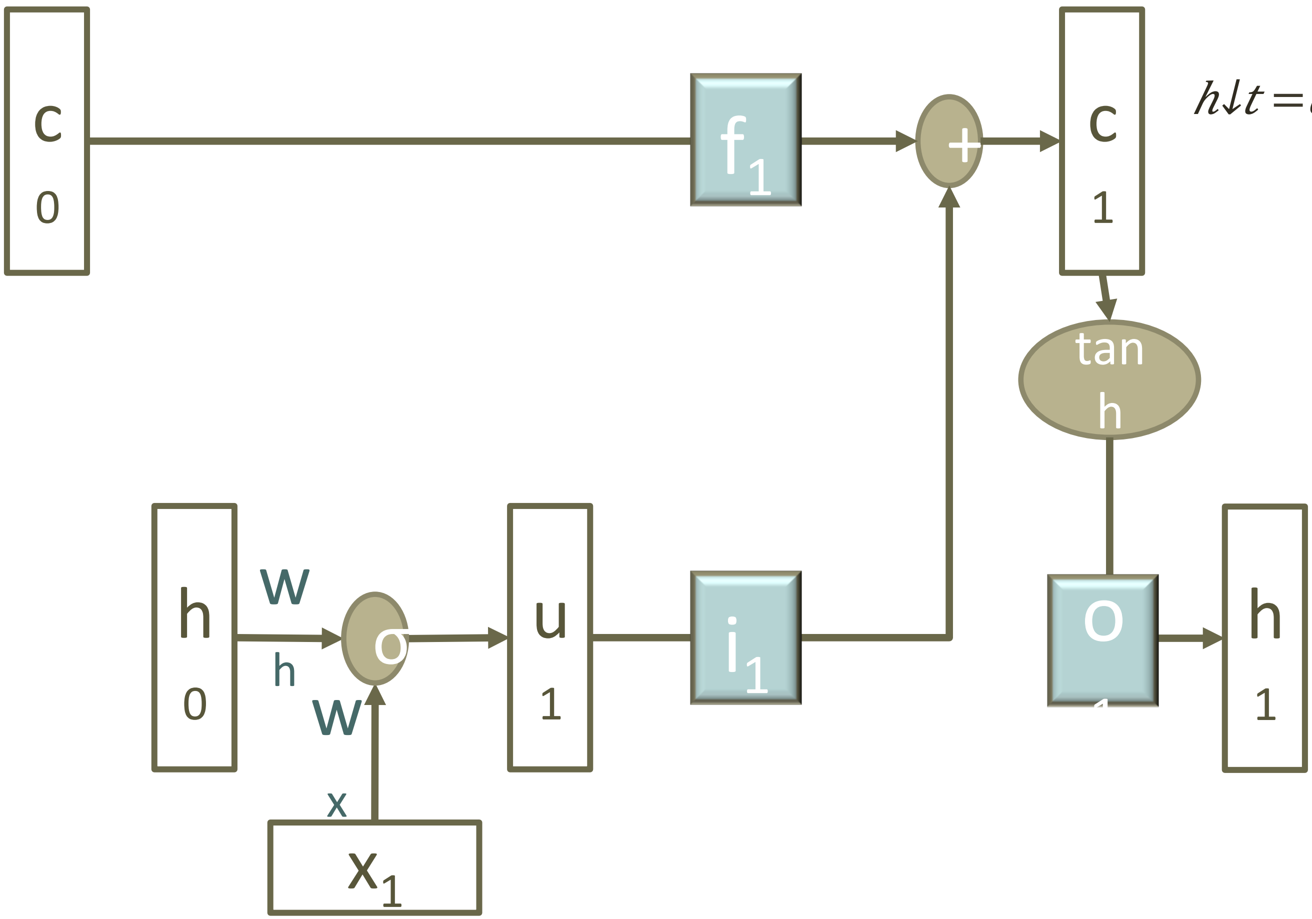
Transforming RNN to LSTM



$$i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t)$$

[slides from Catherine Finegan-Dollak]

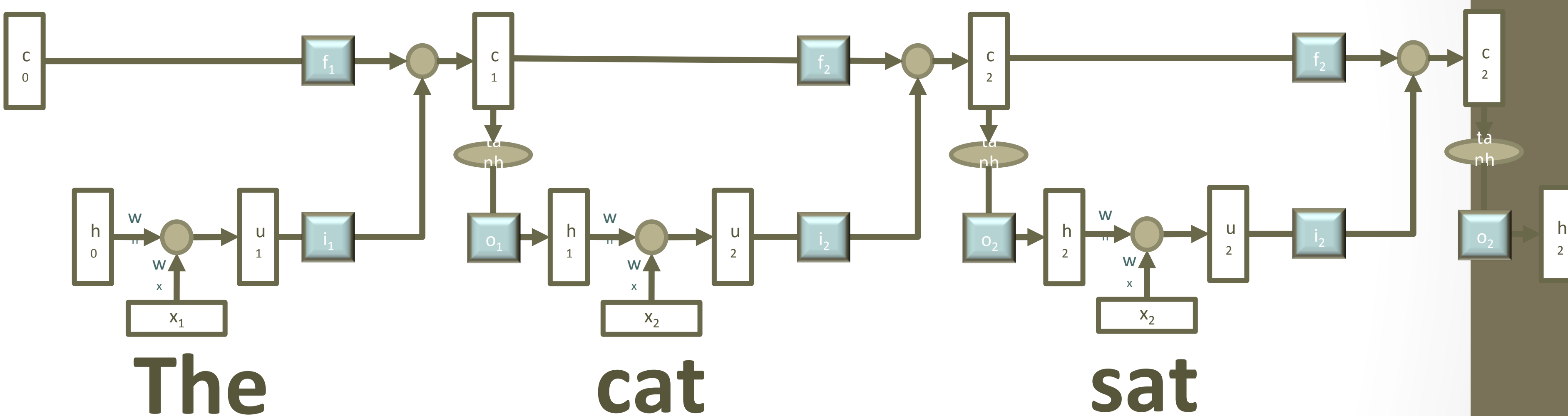
Transforming RNN to LSTM



$h_t = o_t \odot \tanh c_t$

[slides from Catherine Finegan-Dollak]

LSTM for Sequences



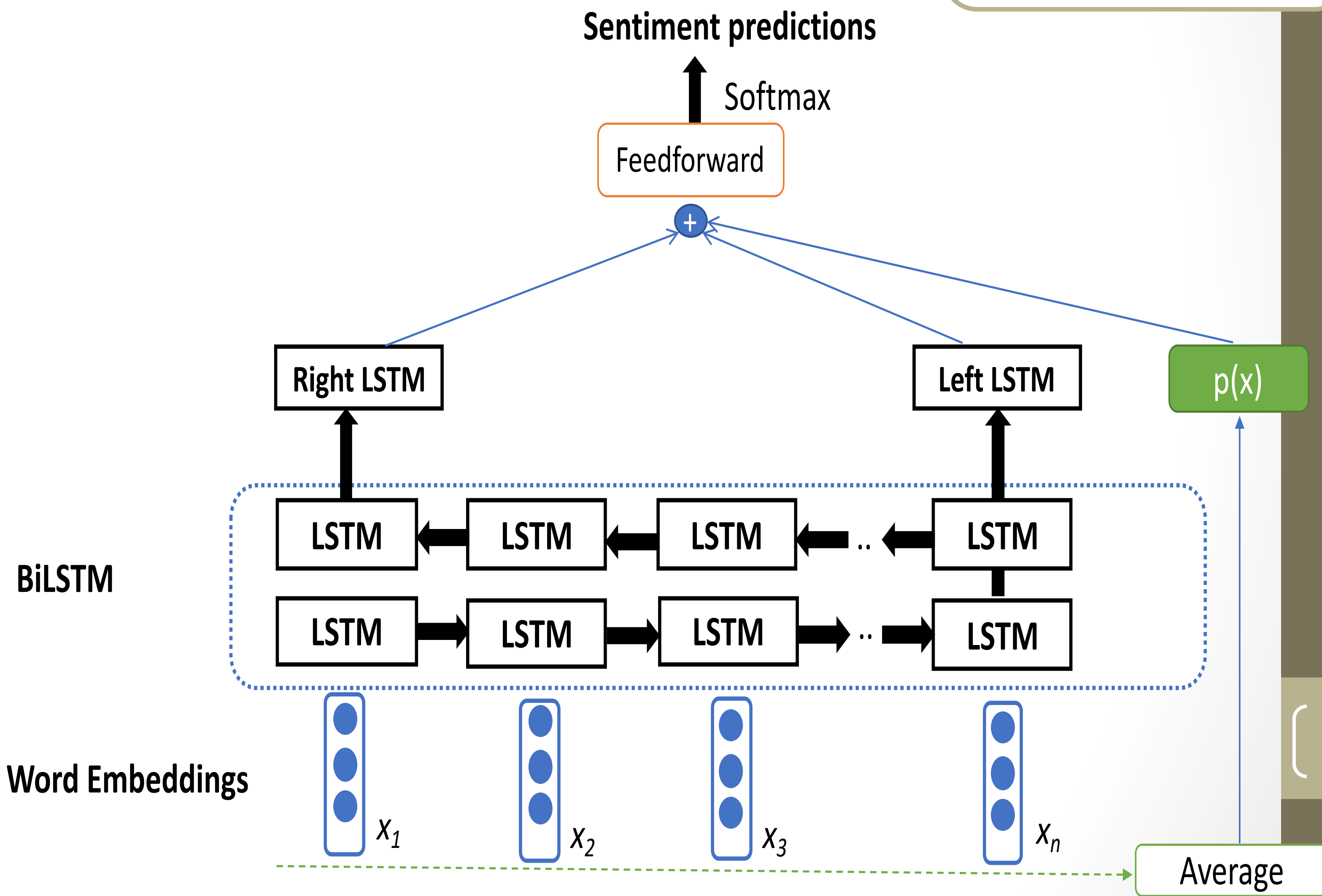
[slides from Catherine Finegan-Dollak]

Bi-LSTM for sentiment

- Pre-trained Word Embeddings

$v_{(x_i)} \in R^d$

d: embedding dimension



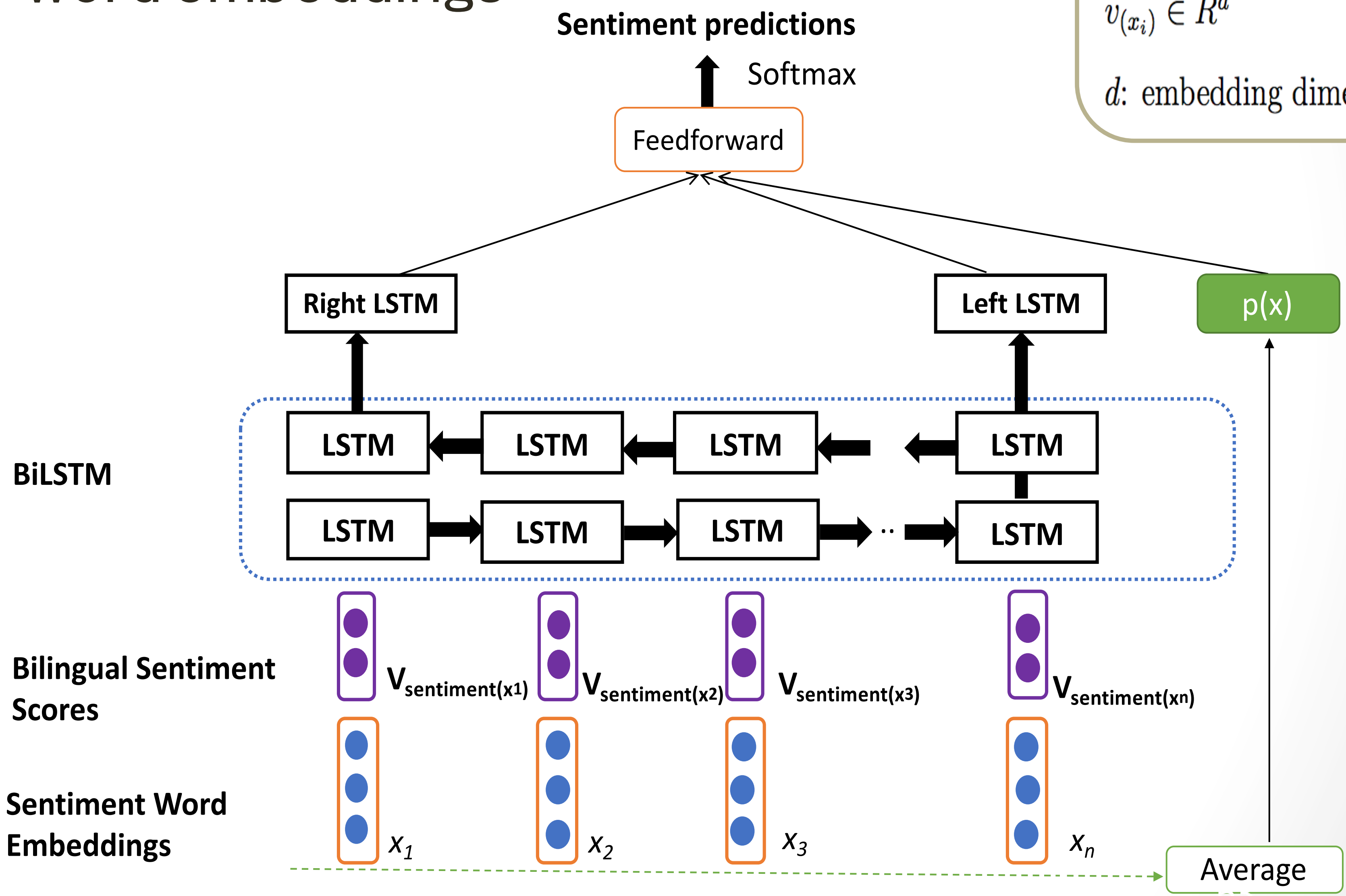
Bi-LSTM enriched with sentiment word embeddings

$L : \{pos, neg, neutral\}$

$v_{sentiment(x_i)} \in R^{|L|}$

$v_{(x_i)} \in R^d$

d: embedding dimension



Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank

- Socher et al, Stanford 2013 https://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf

- Problem with previous work: difficulty expressing the meaning of longer phrases

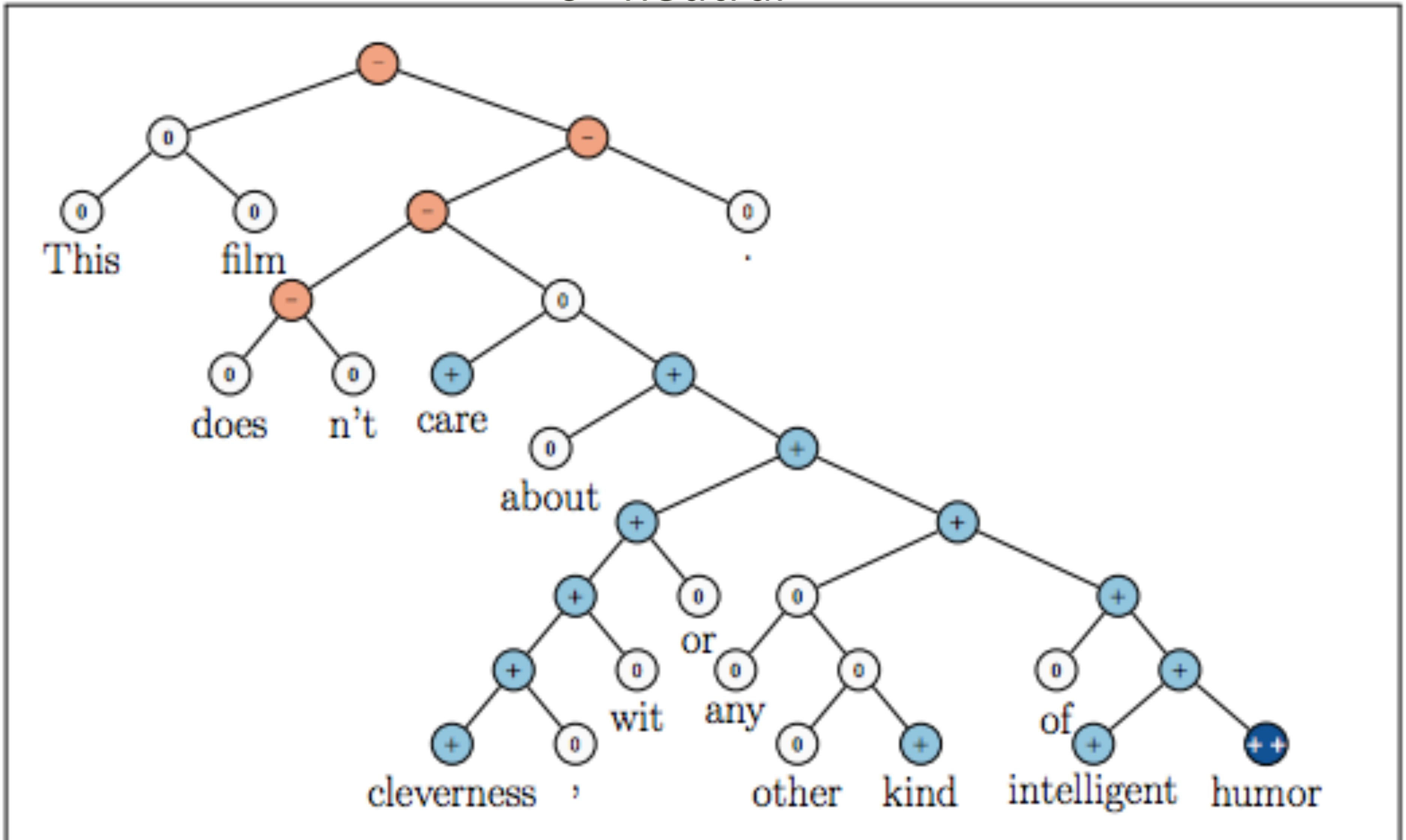
- Goal
  - To predict sentiment at the sentence or phrase level
  - Capture effect of negation and conjunctions
  - Sentiment Treebank
  - Recursive Neural Tensor Network

Sentiment Treebank

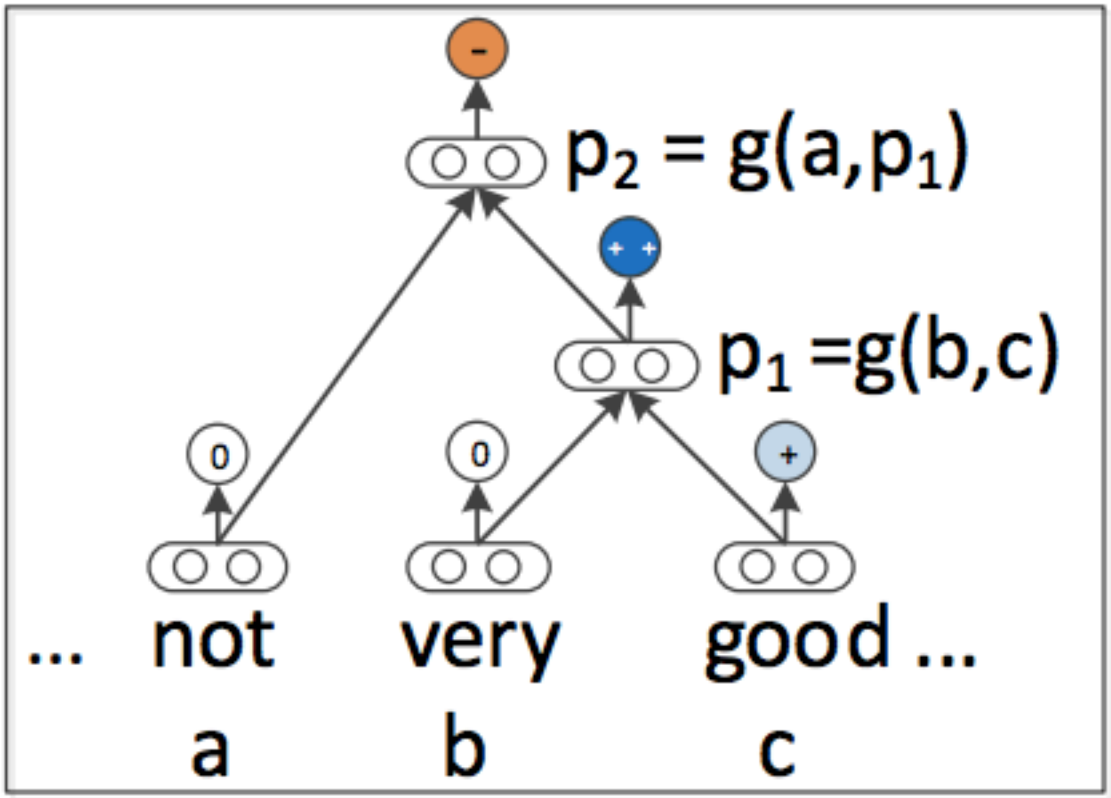
- Movie review excerpts from rottentomatoes.com (Pang & Lee 2005)
  - 10,662 sentences
- Parsed by Stanford Parser (Klein & Manning 2003)
  - 215,154 phrases
- Each phrase labeled for sentiment using Amazon Mechanical Turk (AMT)
  - 5 classes emerge: negative, somewhat negative, neutral, somewhat positive, positive

Example

-- very negative ++ very positive
- Negative + positive
0 neutral



Recursive Neural Models



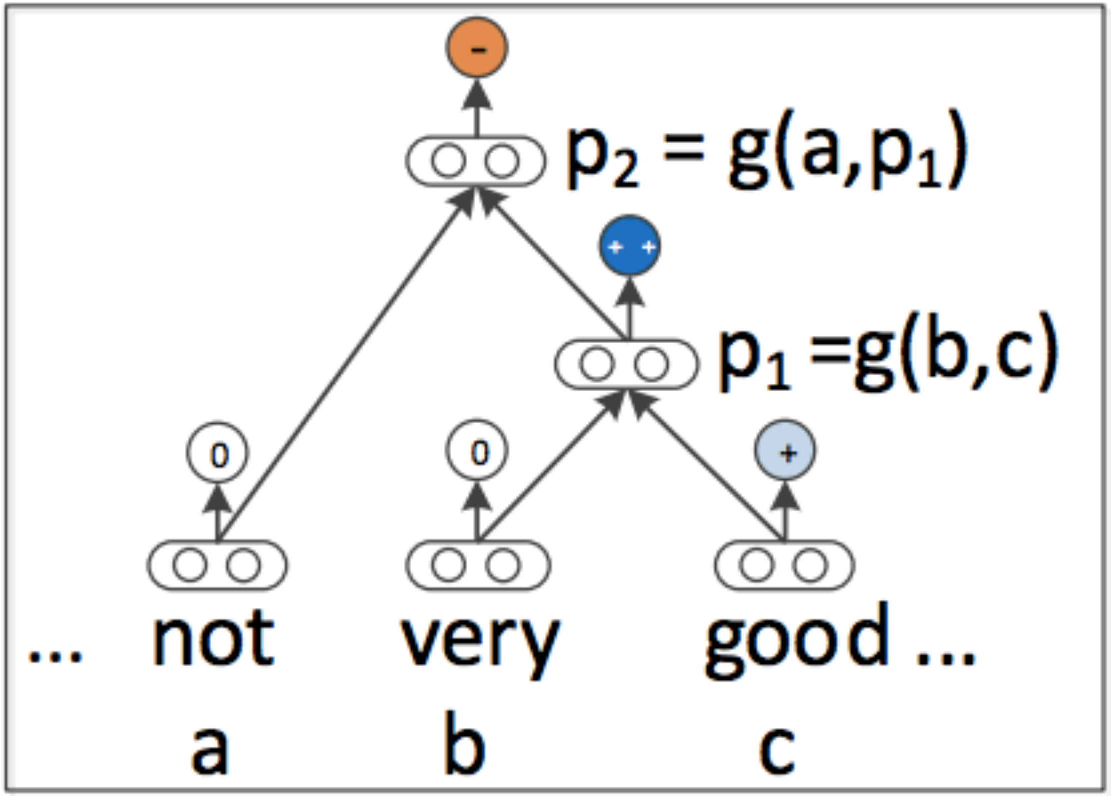
RNN: Recursive Neural Network

$$p_1 = f\left(W \begin{bmatrix} b \\ c \end{bmatrix}\right), p_2 = f\left(W \begin{bmatrix} a \\ p_1 \end{bmatrix}\right),$$

W are the weights to learn

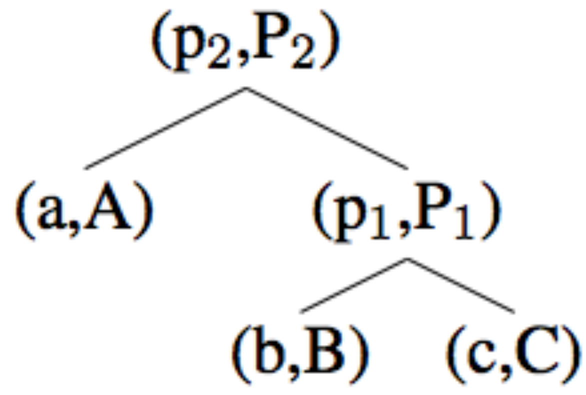
W ε $\mathbb{R}^{d\times 2d}$

f = tanh



MV-RNN Matrix vector RNN

- Introduce weight matrix associated with each non-terminal (P_2 for adjP) and terminal (A for a)
- a = not, b = very, c = good



$$p_1 = f\left(W\begin{bmatrix} Cb \\ Bc \end{bmatrix}\right), P_1 = f\left(W_M\begin{bmatrix} B \\ C \end{bmatrix}\right)$$

RNTN: Recursive Neural Tensor Network

- The MV-RNN has too many parameters to learn (size of vocabulary)
- Can we get compositionality with reduced parameters?
- $P_1 = f\left([a\ b] \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}\right)$

$$= f\left([a\ b] \begin{bmatrix} u_1 a + u_2 b \\ u_3 a + u_4 b \end{bmatrix}\right)$$

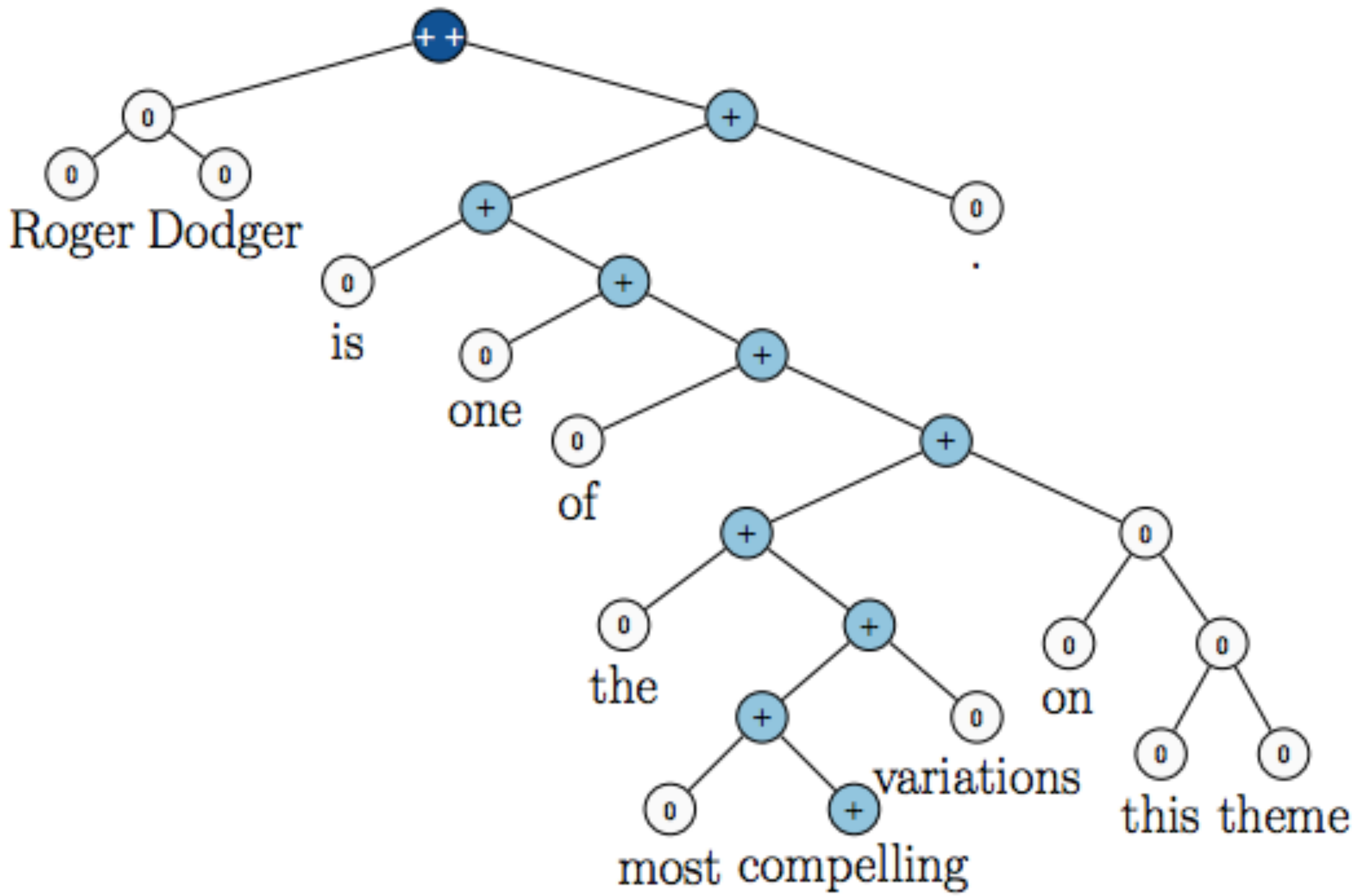
$$= f\left(u_1 aa + u_2 ab + u_3 ab + u_4 bb\right)$$

Results

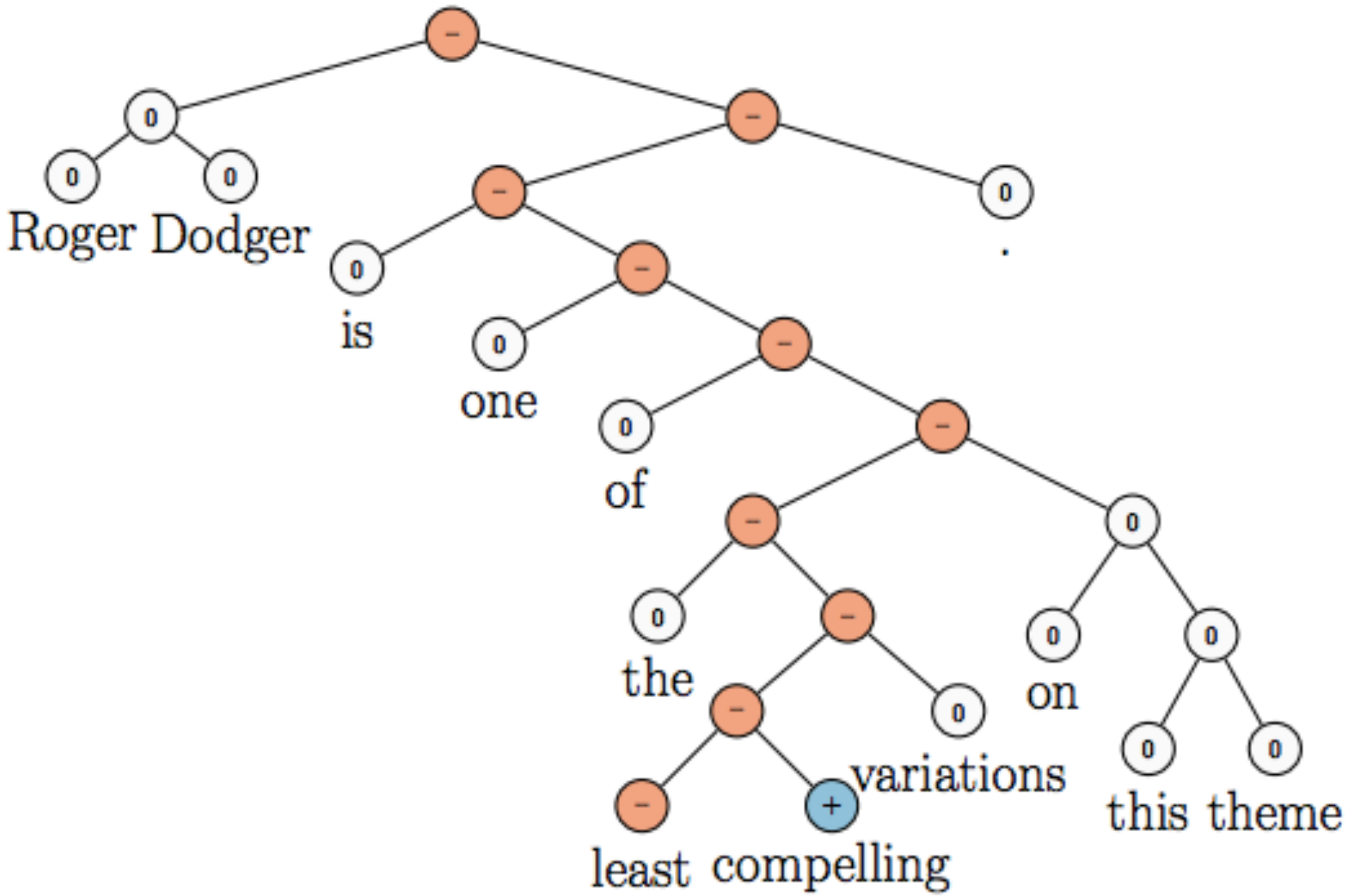
Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

Table 1: Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes.

Positive – "most compelling"



Negative – “least compelling”



Handling Conjunctions

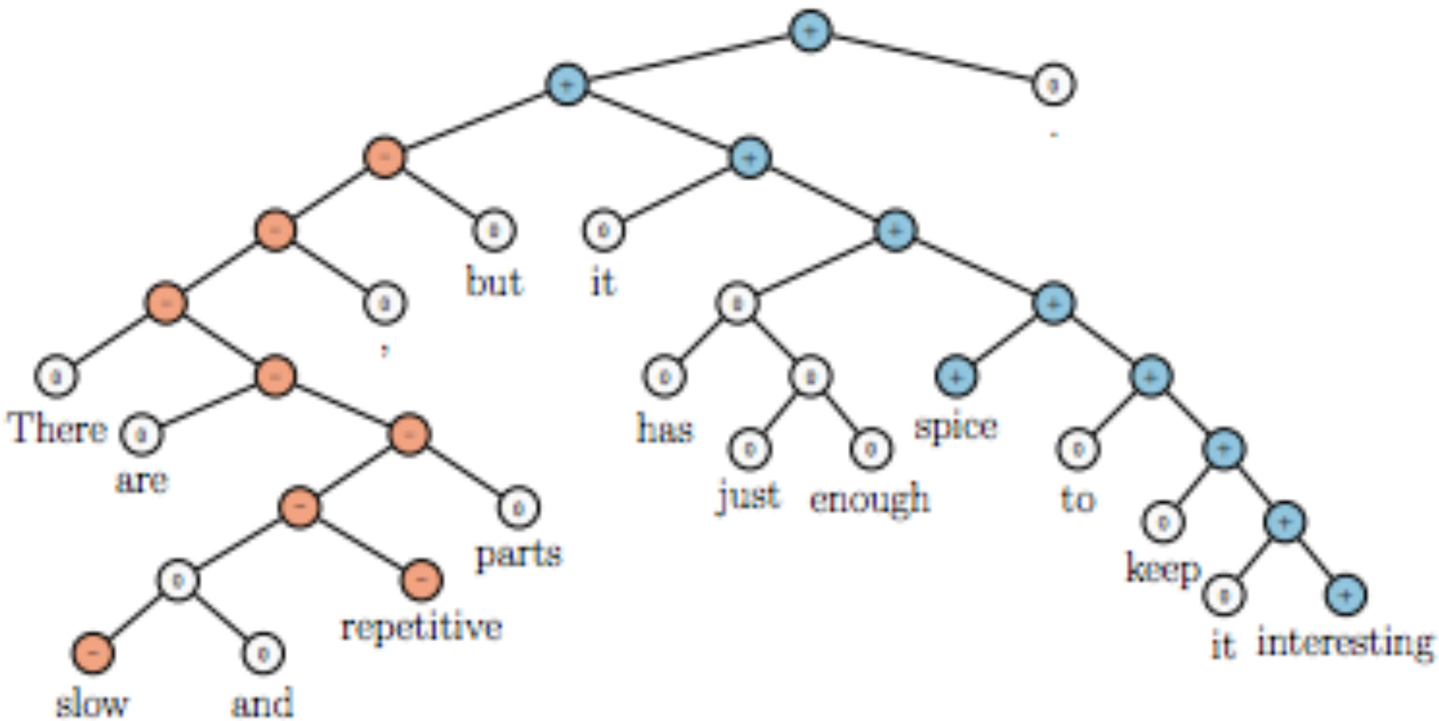


Figure 7: Example of correct prediction for contrastive conjunction X *but* Y.