Sentiment Analysis

Announcements

- Homework 3 due at 2:30pm next Tuesday.

From Core NLP to Applications

CORE NLP
- Parsing
- POS tagging
- Semantics

APPLICATIONS
- Sentiment
- Summarization
- Information Extraction
- Machine Translation

Today

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

Do embeddings handle negation?

- 1: not 1.0000000000000004
- 2: n't 0.8595728019811346
- 3: but 0.839545755064721
- 4: did 0.8378272618764329
- 5: would 0.8187187243474063
- 6: should 0.8147740055059252
- 7: if 0.8116091330796058
- 8: because 0.7987450091713499
- 9: they 0.7944962528430977
- 10: be 0.791361002418091
- 11: could 0.7894321710724349
- 12: never 0.7860447682817786
- 13: any 0.7842654035407371
- 14: even 0.776876305477035

Do embeddings handle negations?

Yes

No

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

Antonyms and Synonyms: embedding for “hot”

- 1: hot 0.9999999999999996
- 2: cool 0.6137860693915586
- 3: hottest 0.5816319703075693
- 4: heat 0.5266680665228453
- 5: warm 0.51671900202736
- 6: cold 0.5093751774671291
- 7: chili 0.4624909077143189
- 8: dry 0.4613872561048547
- 9: heated 0.45721498258314297
- 10: bubbling 0.4534137094122158
- 11: hotter 0.4529186992101415
- 12: spots 0.4416197356093728
- 13: boiling 0.44035866405318447
- 14: billboard 0.4340849896360003
- 15: soft 0.4268572343642097
- 16: temperature 0.42600188687018437
- 17: wet 0.4198371006642362
- 18: chocolate 0.41844508951954273
- 19: water 0.4174513613786725
- 20: temperatures 0.4160490504977998
- 21: drink 0.41476813237122767
- 22: stove 0.41431353491608697

e embedding for "hot" distinguish it from "

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

What is sentiment?

- Expression of positive or negative opinions
- .. Towards a topic, person, event, entity
- .. Towards an aspect

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
- 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

Additional Examples

- Authorities are only too aware that Kashgar is 4,000 kilometres (2,500 miles) from Beijing but only a tenth of the distance from the Pakistani border.
- Taiwan-made products stood a good chance of becoming even more competitive thanks to wider access to overseas markets and lower costs for material imports, he said.
- "March appears to be a more reasonable estimate while earlier admission cannot be entirely ruled out," according to Chen, also Taiwan's chief WTO negotiator.
- friday evening plans were great, but saturday's plans didnt go as expected -- i went dancing & it was an ok club, but terribly crowded :-(
- WHY THE HELL DO YOU GUYS ALL HAVE MRS. KENNEDY! SHES A FUCKING DOUCHE
- AT&T was okay but whenever they do something nice in the name of customer service it seems like a favor, while T-Mobile makes that a normal everyday thin

Examples from Rosenthal 2014

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 forthcoming

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.

Wordnet

- Proper nouns (e.g. Britney Spears) are automatically marked as objective
- Words that do not exist in the DAL are looked up in Wordnet
- Compute the average of the DAL scores of all the synonyms of the first sense
- If there are no synonyms, look at the hypernym

Wiktionary

- Wiktionary is a free content dictionary
  - http://www.wiktionary.org
- If a word does not appear in the DAL or Wordnet look it up in Wiktionary
- Compute the average of the DAL scores for each word in the definition that has its own Wiktionary page

Verb

LOL (*third-person singular simple present* **LOLs**, *present participle* **LOLing**, *simple past and past participle* **LOLed**, **LOLd** *or* **LOL'd**)

1. To laugh out loud.

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

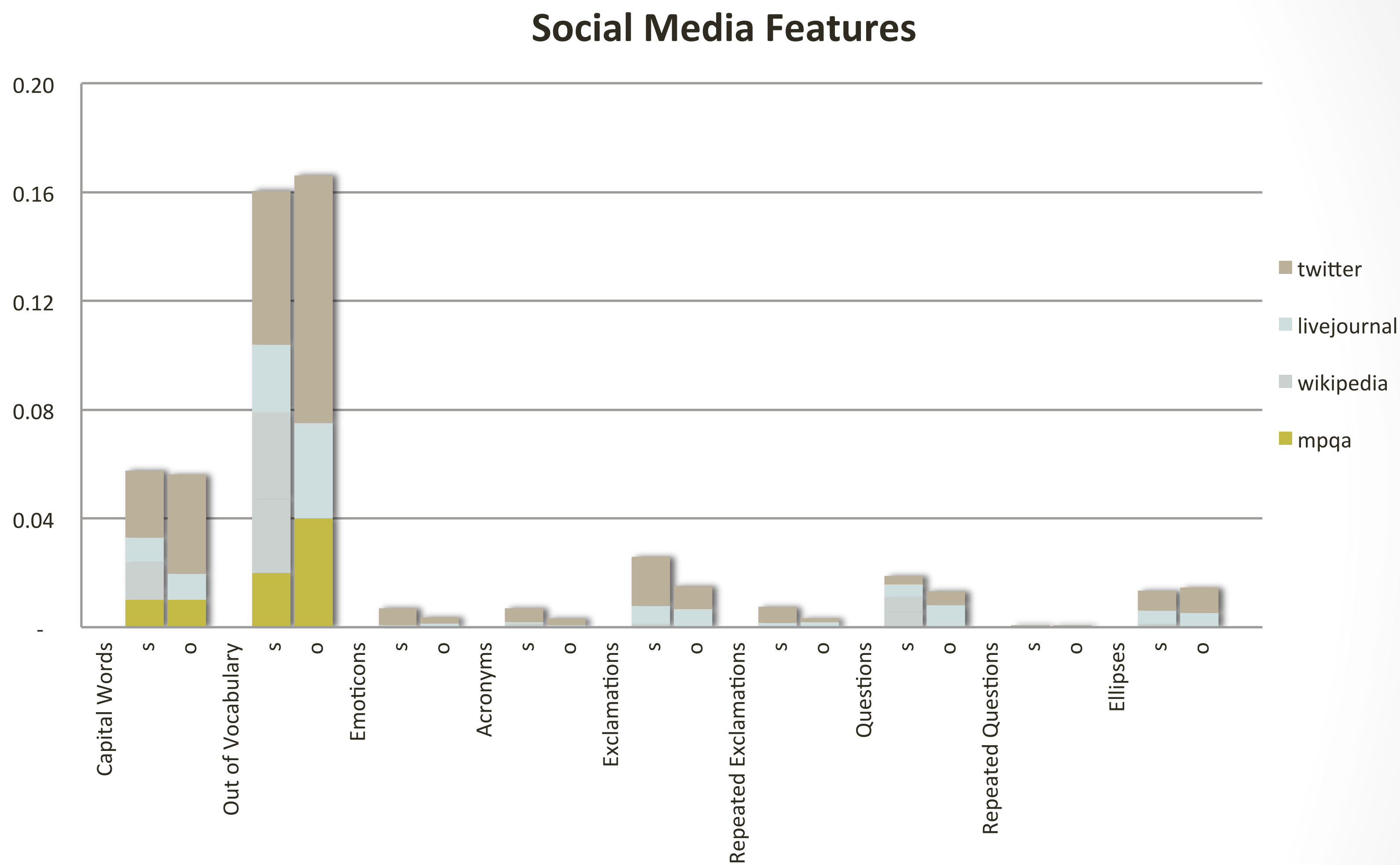
Syntactic Features

- 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	...



SM features tend to be very rare. The frequency for each feature is less than 1% per dataset

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%

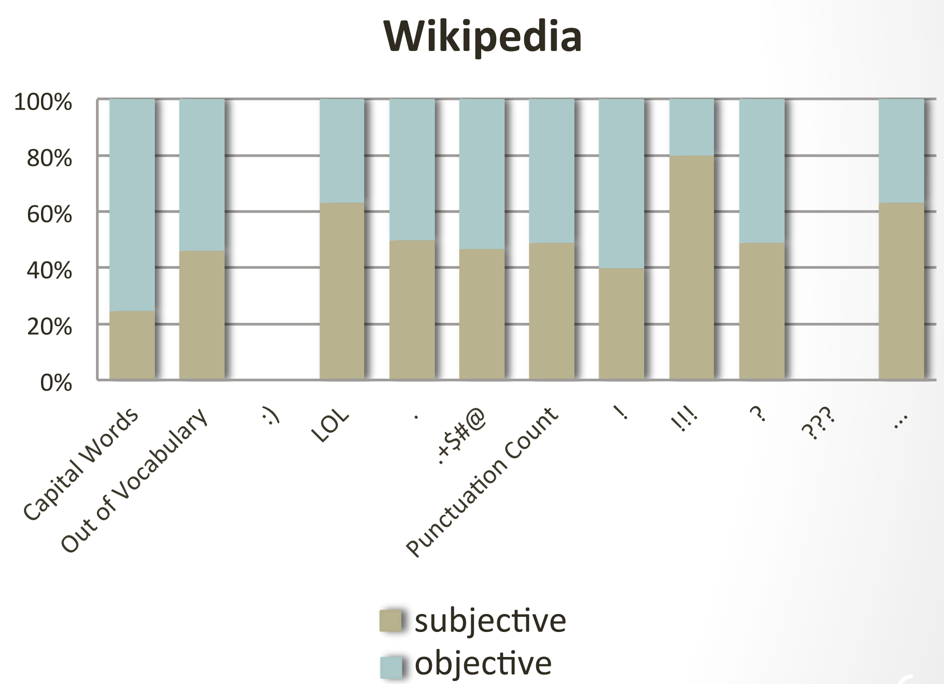
Balanced

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 in Weka
- 10 runs of 10-fold cross-validation
- Statistical significance using the t-test with p = .001

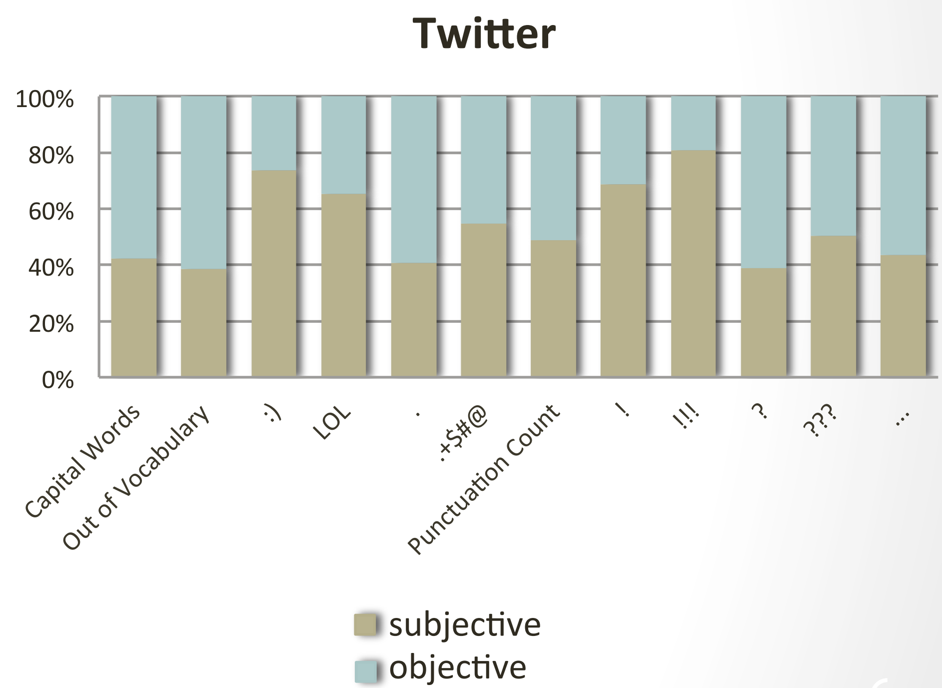
Social Media Error Analysis

- 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



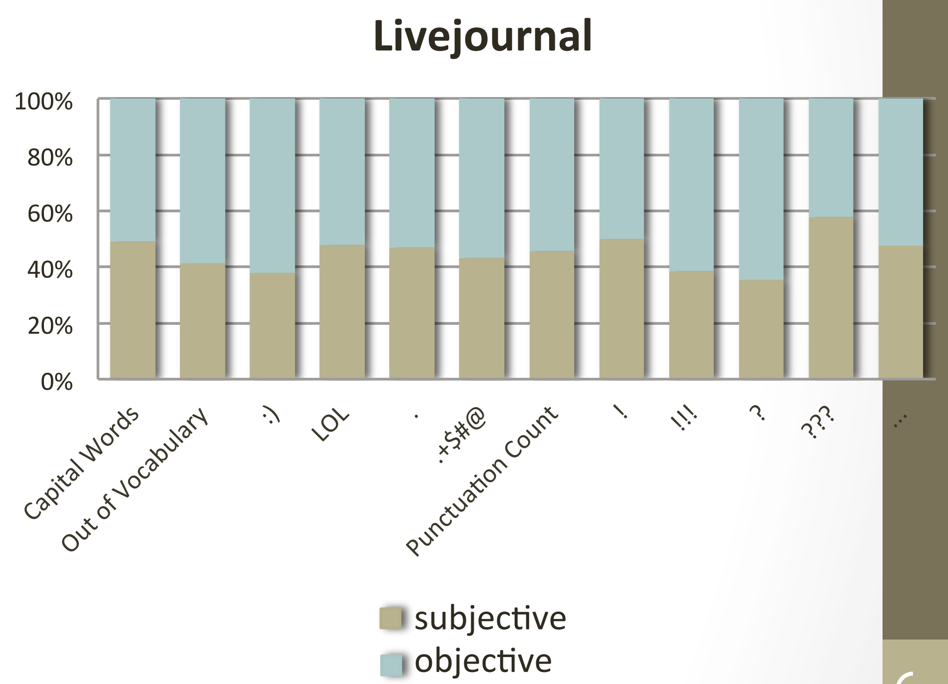
Social Media Error Analysis

- Twitter
  - ellipses do 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.



Cross-Genre Classification

Testing

Training

	Twitter	LiveJournal	MPQA	Wikipedia
Twitter		71.6%	62.1%	76.9%
LiveJournal	82.5%		65.4%	80.9%
MPQA	75.6%	69.3%		71.2%
Wikipedia	82.4%	76.7%	62.4%	

This chart displays the best results for each experiment

Cross-Genre Classification

Testing

Training

	Twitter	LiveJournal	MPQA	Wikipedia
Twitter		71.6%	**62.1%**	76.9%
LiveJournal	82.5%		**65.4%**	80.9%
MPQA	75.6%	69.3%		71.2%
Wikipedia	82.4%	76.7%	**62.4%**	

- The online genres do not do well in predicting MPQA sentences

Cross-Genre Classification

Testing

Training

	Twitter	LiveJournal	MPQA	Wikipedia
Twitter		71.6%	62.1%	76.9%
LiveJournal	**82.5%**		65.4%	**80.9%**
MPQA	75.6%	69.3%		71.2%
Wikipedia	**82.4%**	76.7%	62.4%	

- LiveJournal training data does a good job of predicting the other online genres
- Wikipedia training data does a good job of predicting Twitter

Cross-Genre Classification

Testing

Training

	Twitter	LiveJournal	MPQA	Wikipedia
Twitter		71.6%	62.1%	**76.9%**
LiveJournal	82.5%		65.4%	80.9%
MPQA	75.6%	69.3%		71.2%
Wikipedia	82.4%	**76.7%**	62.4%	

- Twitter training data does a decent job of predicting Wikipedia
- Wikipedia training data does a decent job of predicting LiveJournal

Cross-Genre Classification

Testing

Training

	Twitter	LiveJournal	MPQA	Wikipedia
Twitter		**71.6%**	62.1%	76.9%
LiveJournal	82.5%		65.4%	80.9%
MPQA	**75.6%**	**69.3%**		**71.2%**
Wikipedia	82.4%	76.7%	62.4%	

- In general, using the MPQA as training does not perform well
- Using Twitter as training does not perform well in predicting LiveJournal sentences

Neural Network Approaches to Sentiment

- Goldberg:
- Take a standard RNN such as shown in class last time
- 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.

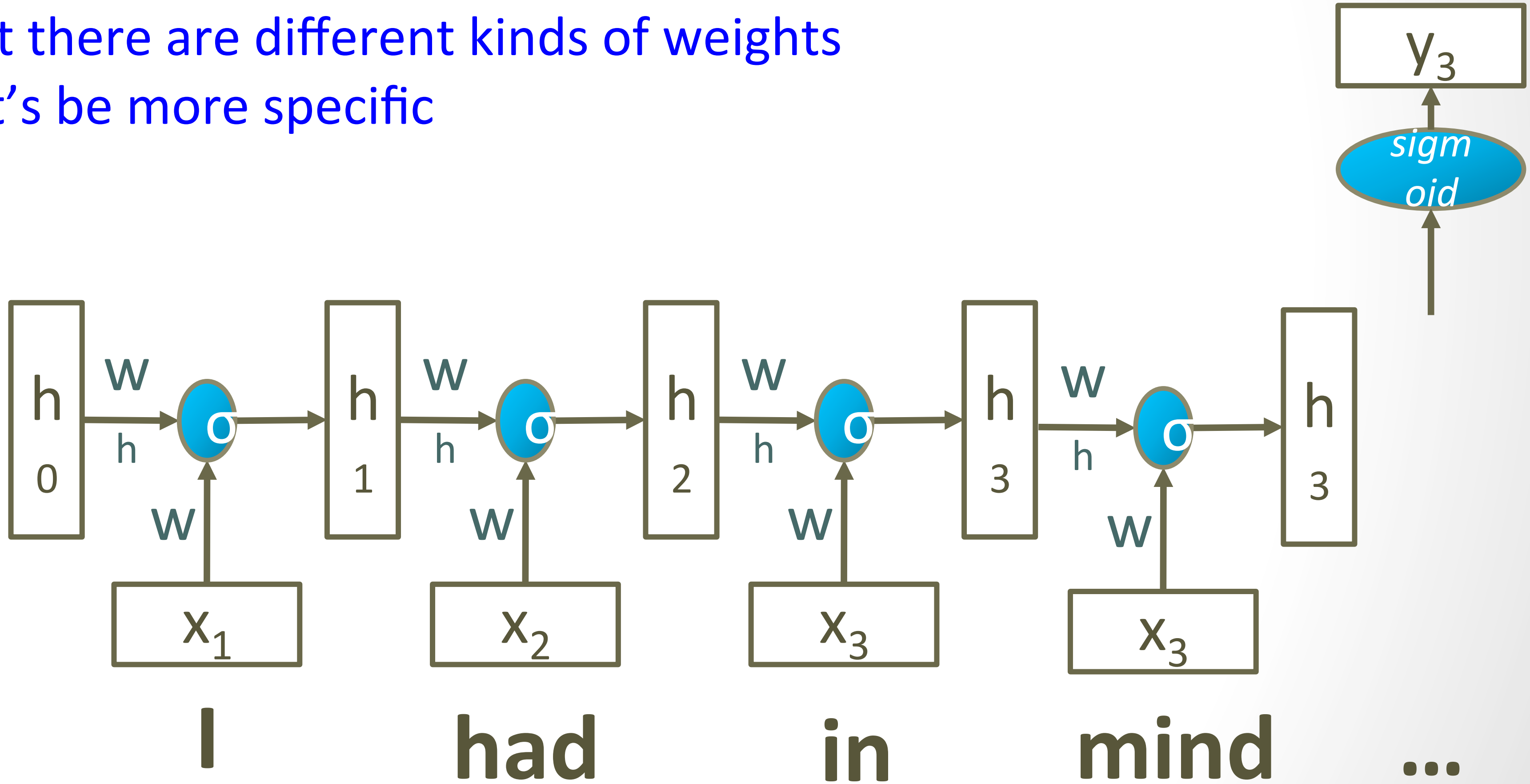
Example

- 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

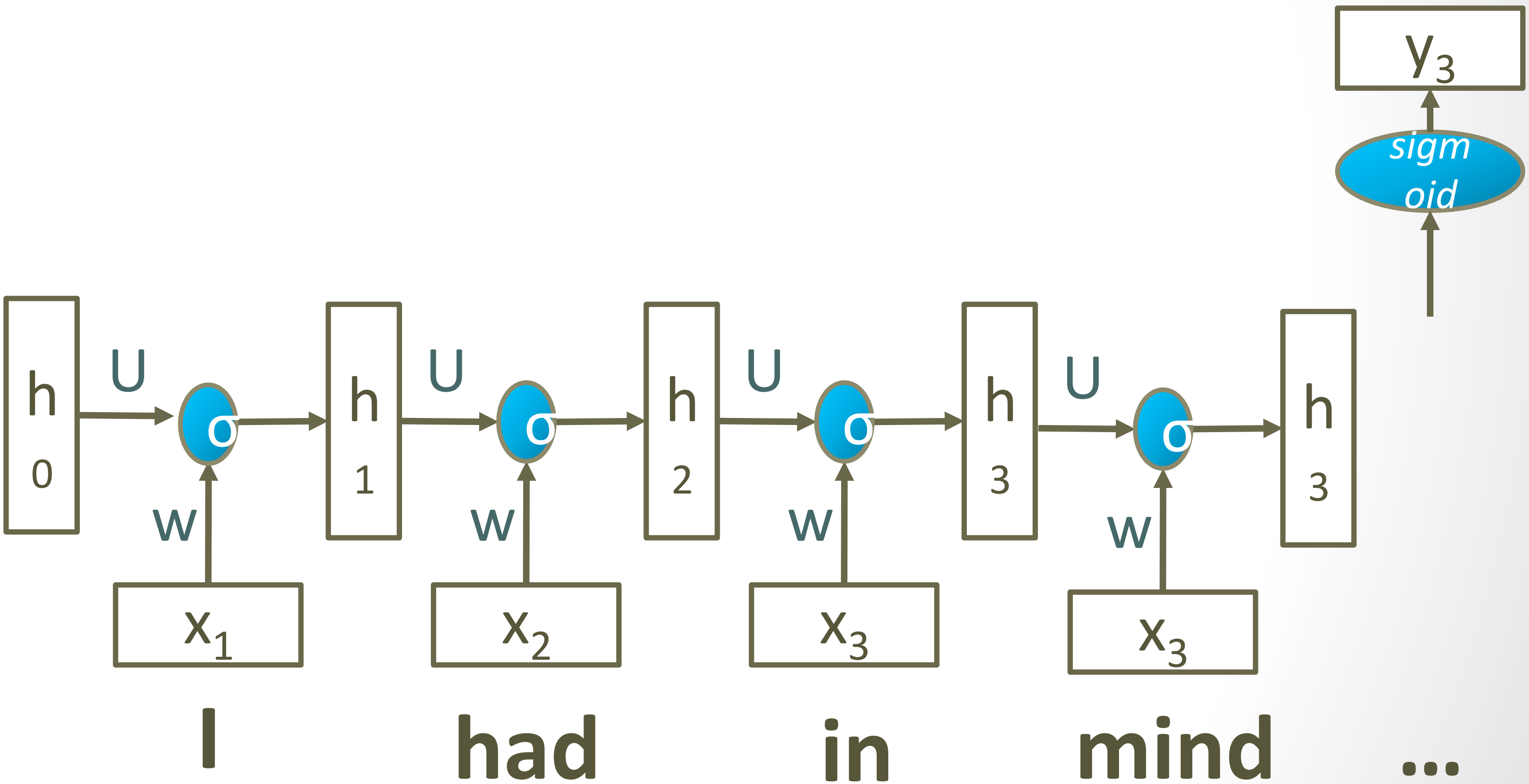
- *I had in mind your facts, Buddy, not hers.*

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 the word x_i

the matrix of embeddings for all words

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

what is h?

the previous state

the hidden layer

weights for the previous state

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

What is x_t * W

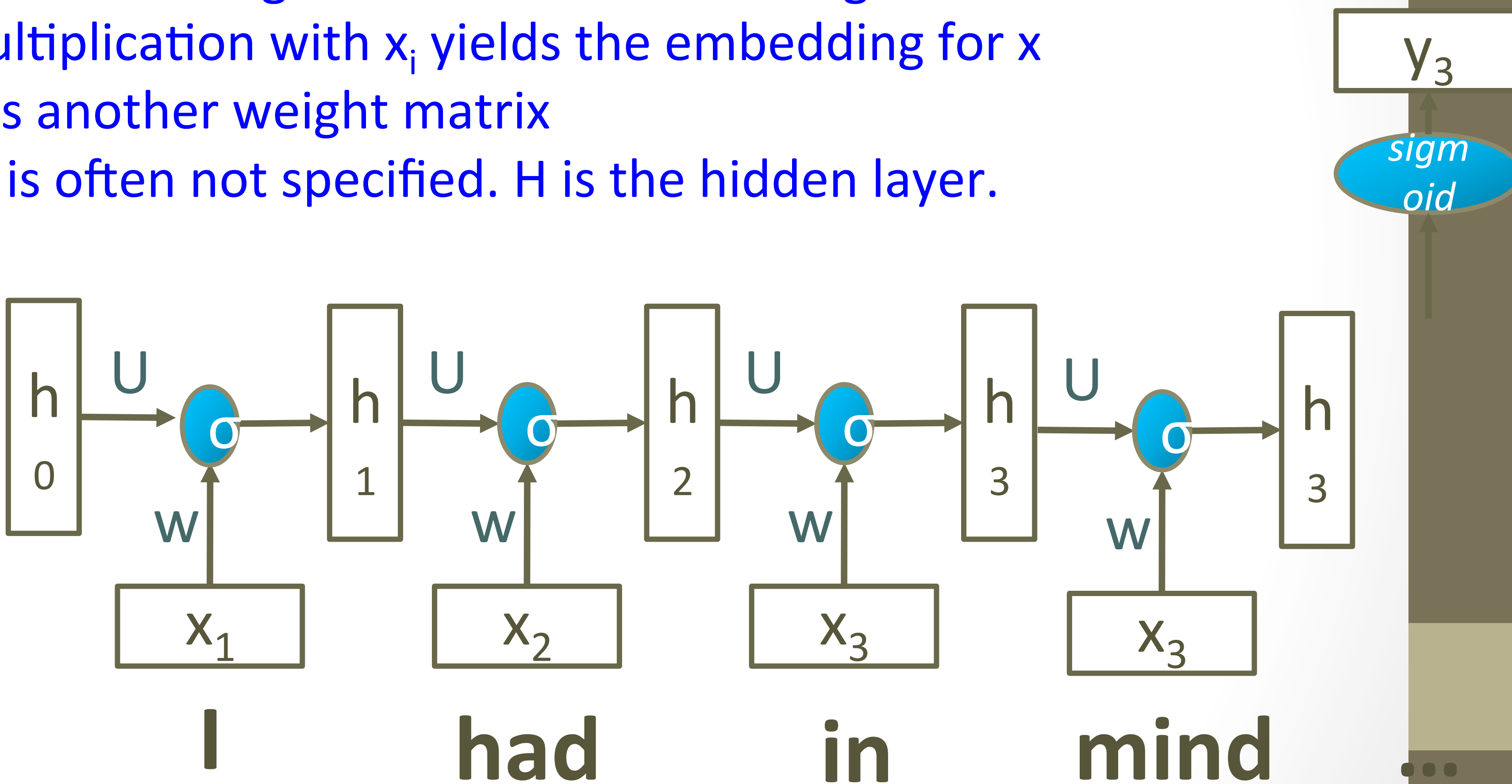
the hidden layer

the word embedding for x_t

the preivous state

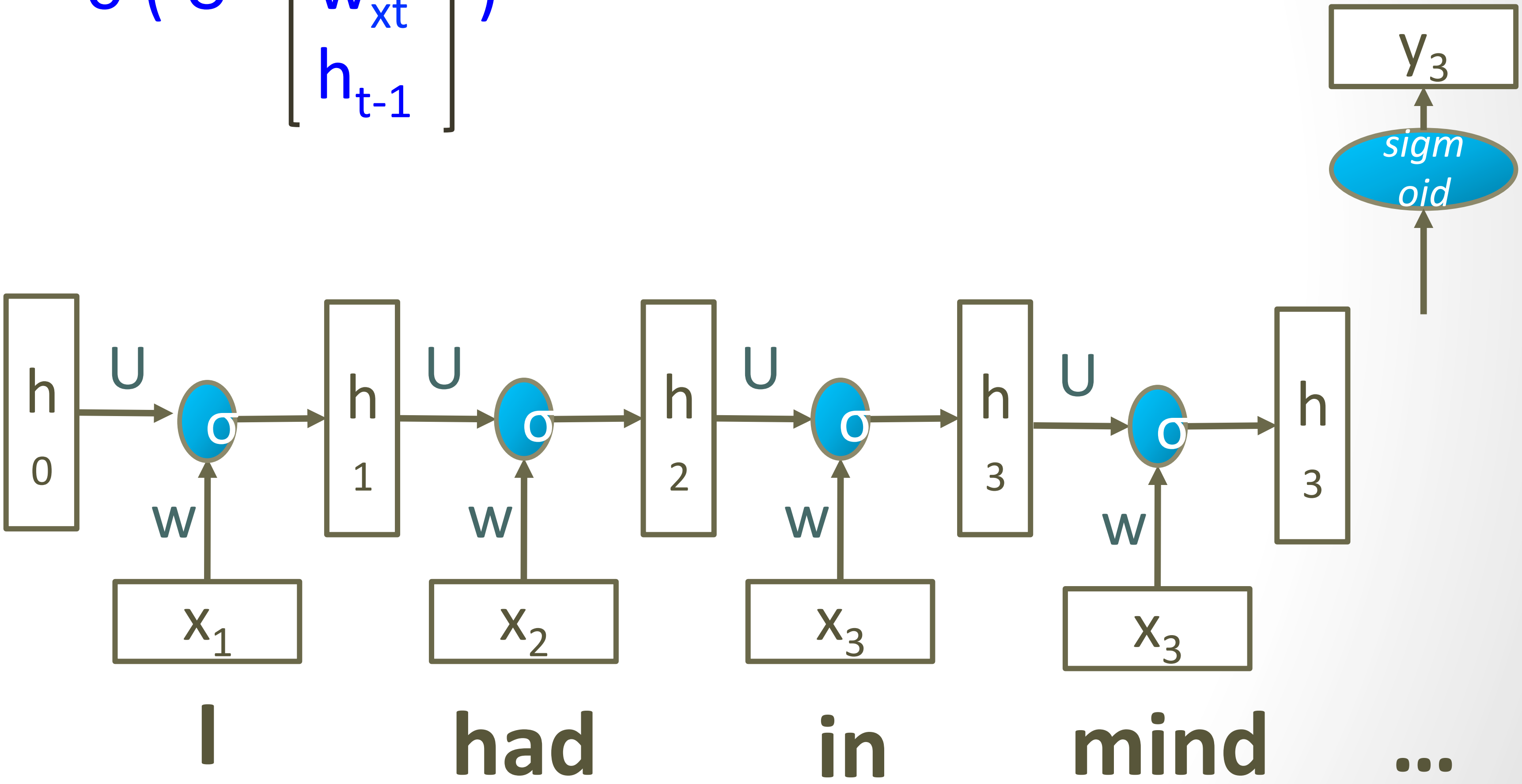
RNN – I had in mind your facts, buddy, not hers.

W are the weights: the word embedding maatrix multiplication with x_i yields the embedding for x
U is another weight matrix
H_0 is often not specified. H is the hidden layer.



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

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



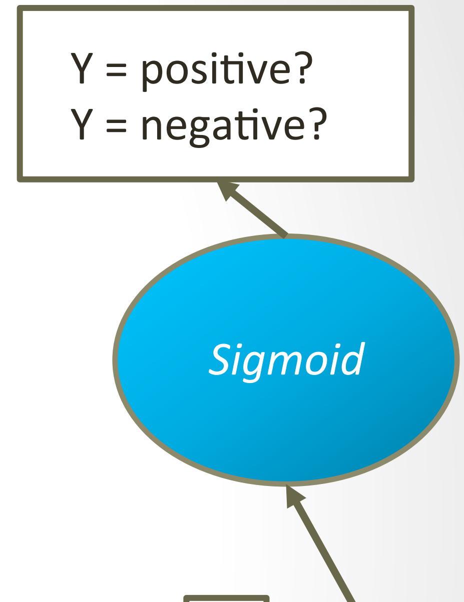
RNN – I had in mind your facts, buddy, not hers.

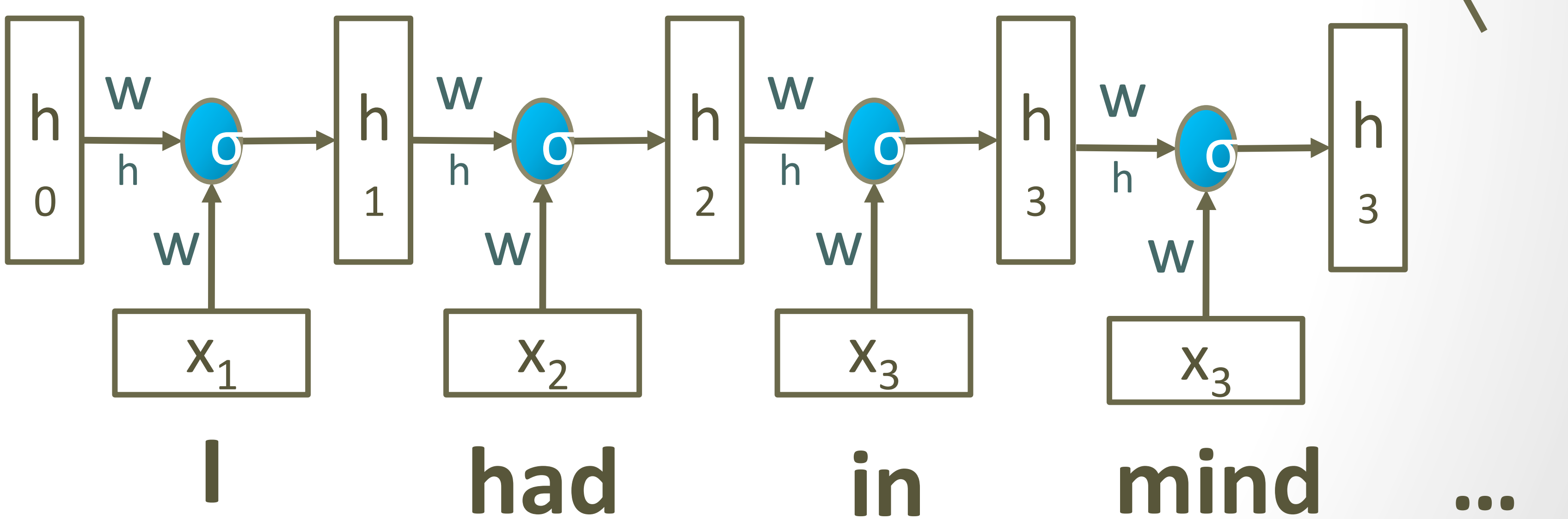
Final embedding run through the sigmoid function -> [0,1]

1 = positive

0= negative

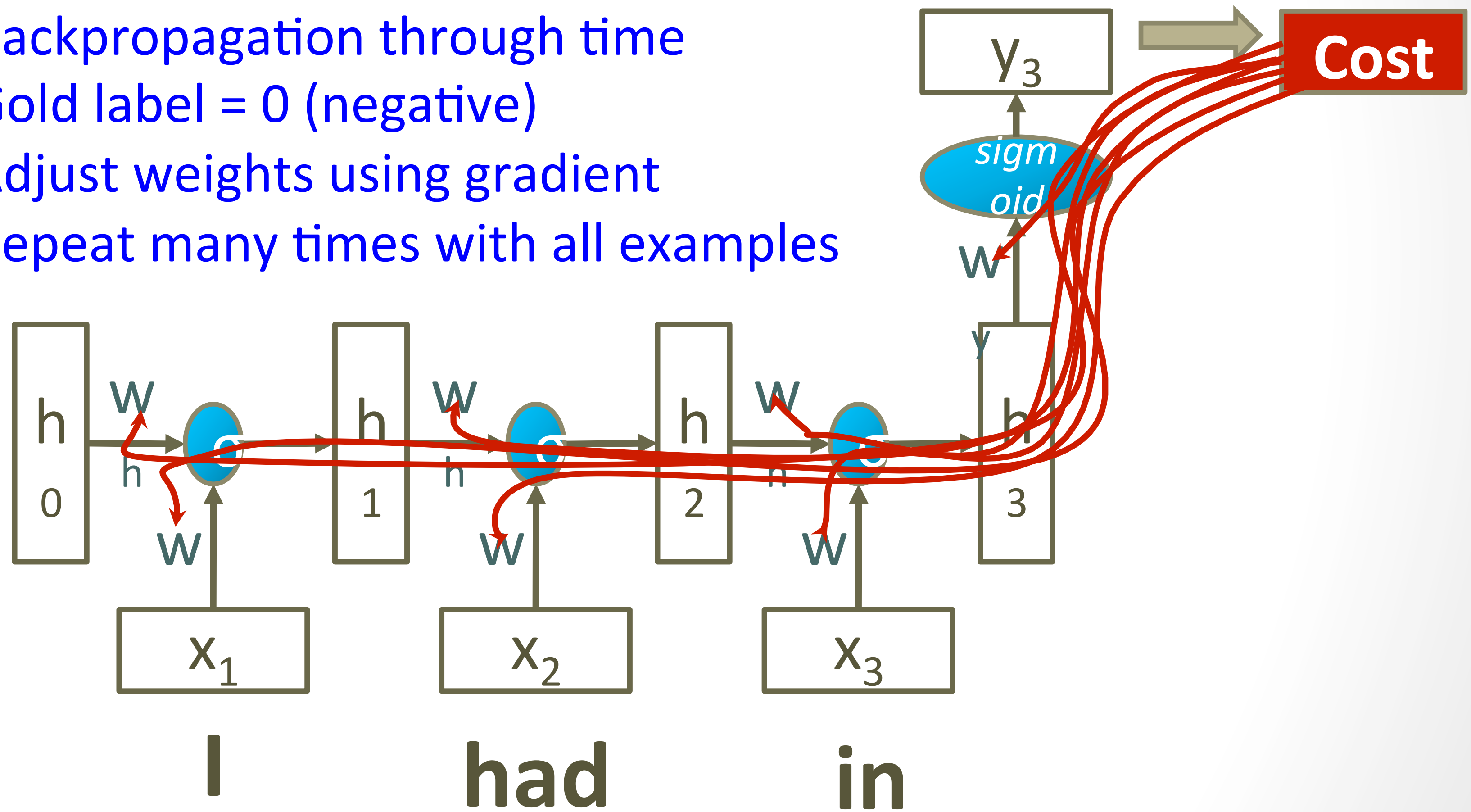
Often final h is used as word embedding for the sentence





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

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

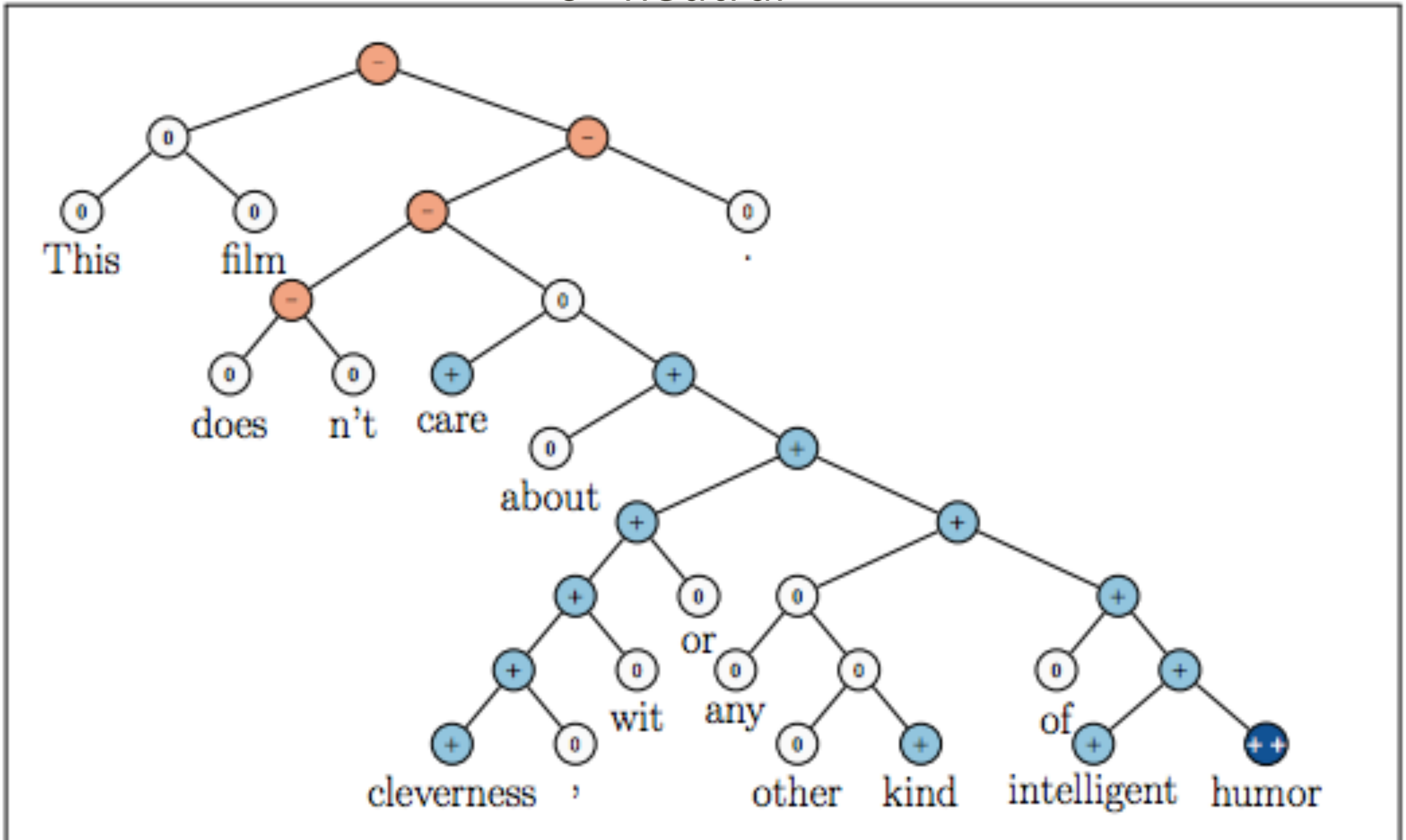
- Goal
  - To predict sentiment at the sentence or phrase level
  - Capture effect of negation and conjunctions
  - Sentiment Treebank
  - Recursive Neural Tensor Network

Sentiment Treebank

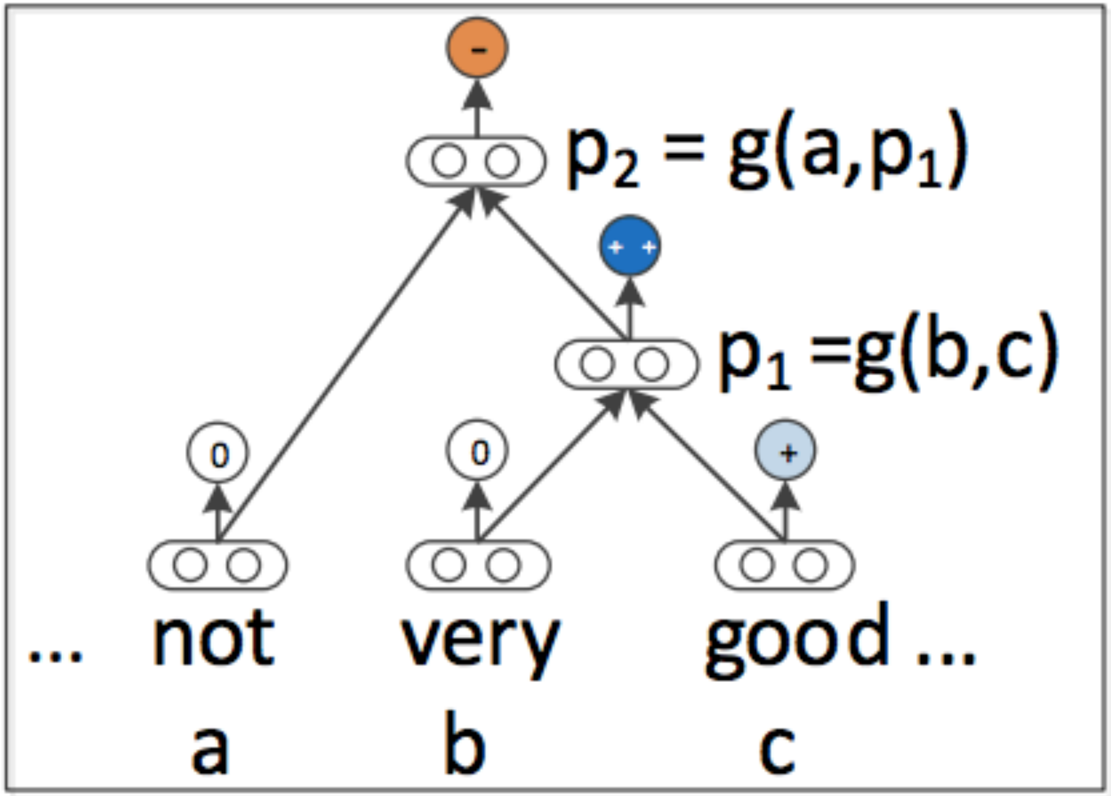
- Movie review excerpts from rottentomatoes.com (Pang & Lee 2005)
  - 10,662 sentences
- Parsed by Stanford Parser (Klein & Manning 2003)
  - 215,154 phrases
- Each phrase labeled for sentiment using Amazon Mechanical Turk (AMT)
  - 5 classes emerge: negative, somewhat negative, neutral, somewhat positive,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

Example

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



Recursive Neural Models



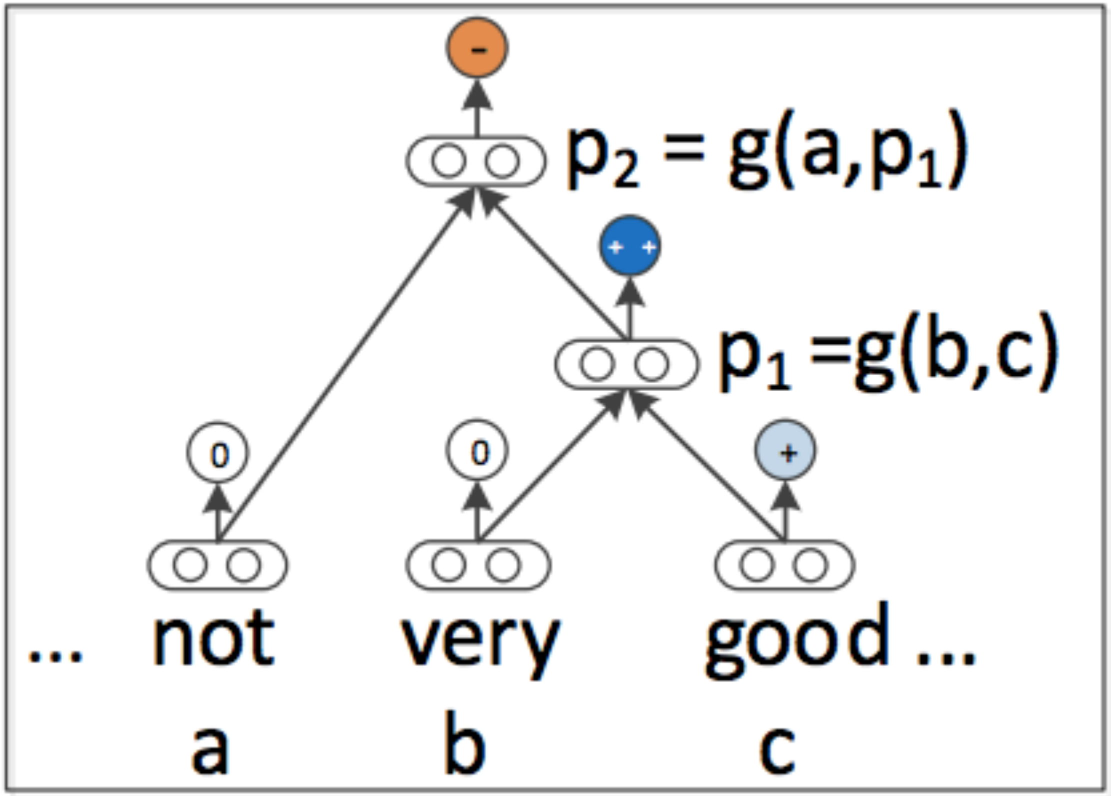
RNN: Recursive Neural Network

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

W are the weights to learn

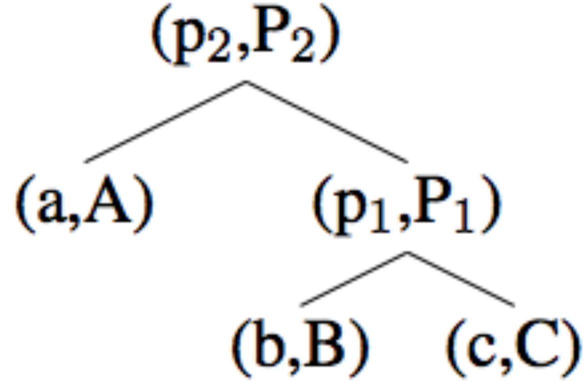
W ε $\mathbb{R}^{d\times 2d}$

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?

- $P_1 = f\left([a\ b] \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}\right)$

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

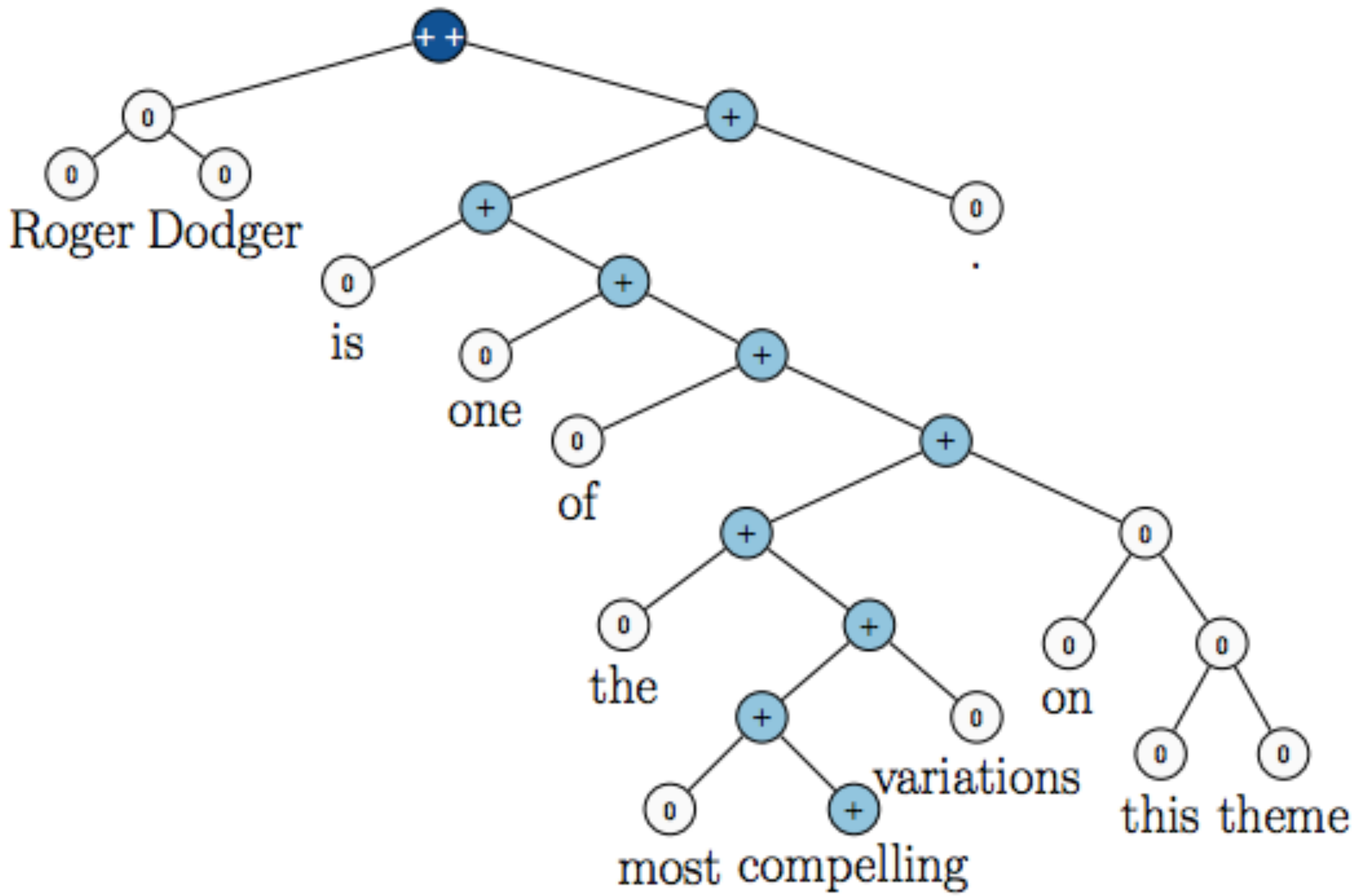
$$= f\left(u_1 aa + u_2 ab + u_3 ab + u_4 bb\right)$$

Results

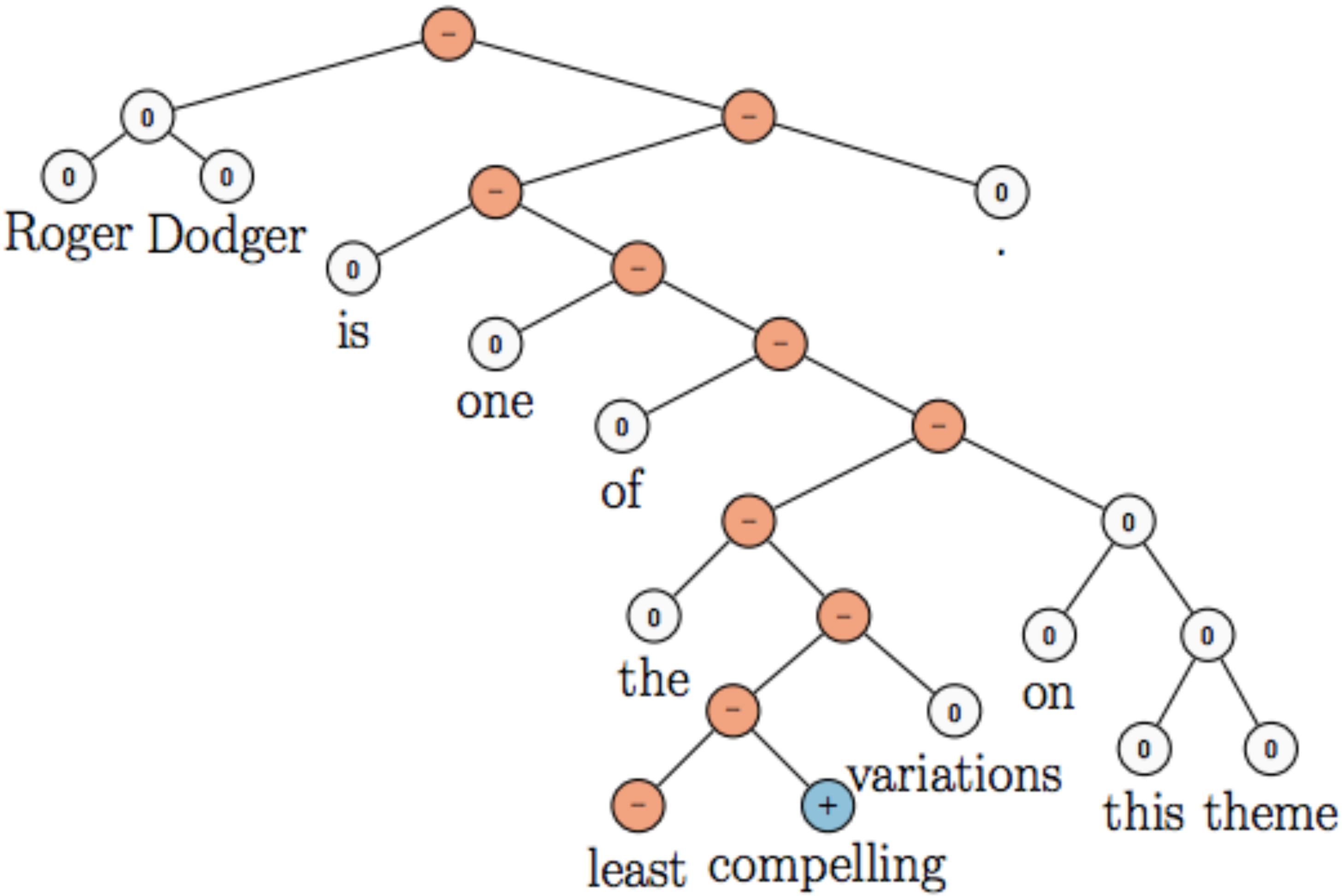
Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

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

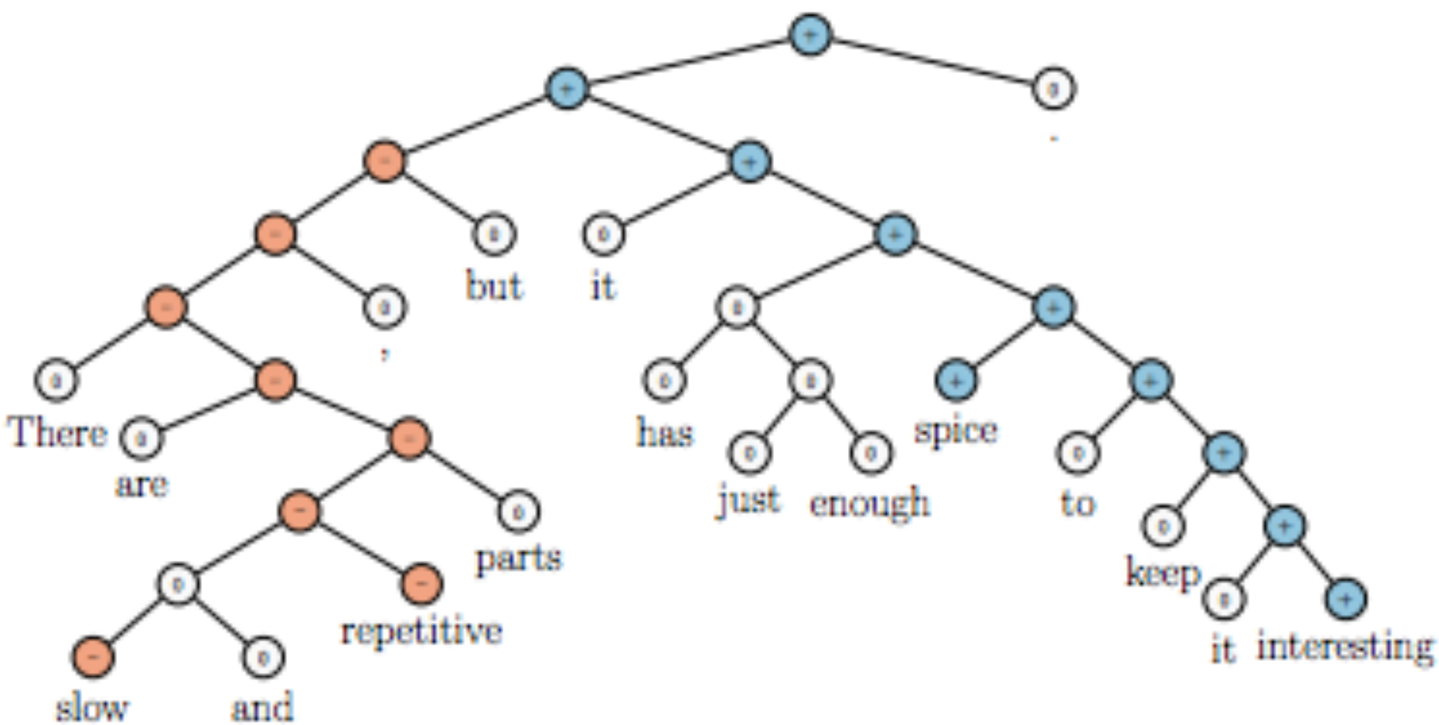


Figure 7: Example of correct prediction for contrastive conjunction *X but Y*.

Next Time

- Summarization