Applied Natural Language Processing

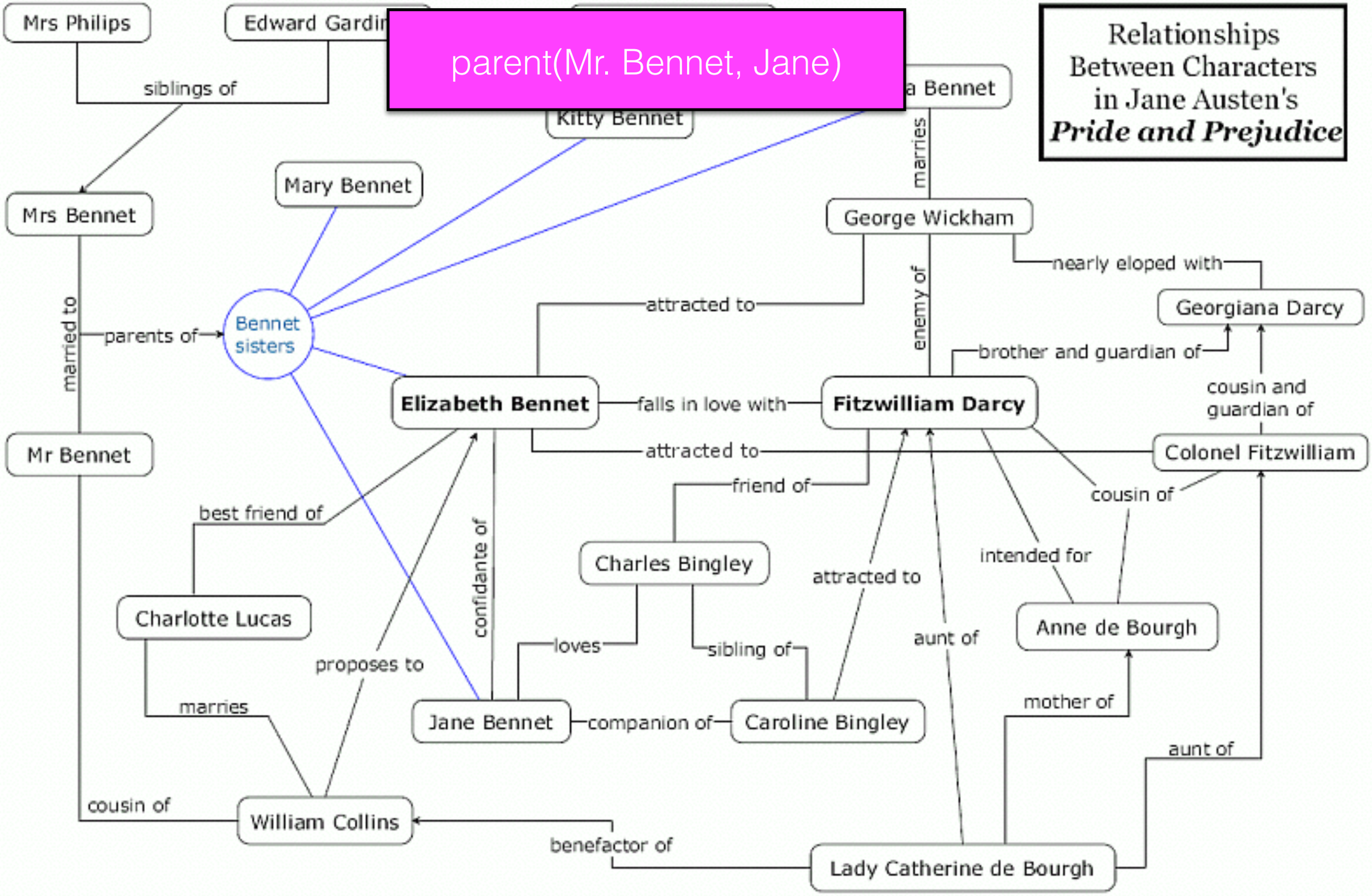
Info 256

Lecture 25: Information Extraction 2 (April 23, 2019)

Masha Belyi, UC Berkeley

Information Extraction

- Named entity recognition
- Relation extraction
- Entity linking
- Event detection
- Event coreference
- Extra-propositional information (veridicality, hedging)



https://en.wikipedia.org/wiki/Pride_and_Prejudice

Information Extraction

- Named entity recognition
- Relation extraction
- **Entity linking**
- Event detection
- Event coreference
- Extra-propositional information (veridicality, hedging)

Entity linking

Michael	Jordan	can	dunk	from	the	free	throw	line
B-PER	I-PER							

?

Wikification!

Entity linking

- Task: Given a database of candidate referents, identify the correct referent for a mention in context.

Text	True wikipedia page
Hornets owner **Michael Jordan** thinks having one or two "superteams" is a detriment to the NBA because the other 28 teams "are going to be garbage."	`wiki/Michael_Jordan`
In 2001, **Michael Jordan** and others resigned from the Editorial Board of *Machine Learning*.	`wiki/Michael_I._Jordan`
The stars are aligning for leading man **Michael Jordan**, who just signed on for a new film, according to Variety.	`wiki/Michael_B._Jordan`
Michael Jordan played in 1,072 regular-season games in his 15-season career	`wiki/Michael_Jordan`

Michael Jordan (disambiguation)

From Wikipedia, the free encyclopedia

Michael Jordan (born 1963) is an American basketball player.

Michael or **Mike Jordan** may also refer to:

People [edit]

Sports [edit]

- Michael Jordan (footballer) (born 1986), English goalkeeper
- Mike Jordan (racing driver) (born 1958), English racing driver
- Mike Jordan (baseball, born 1863) (1863–1940), baseball player
- Mike Jordan (cornerback) (born 1992), American football cornerback
- Michael-Hakim Jordan (born 1977), American professional basketball player
- Michal Jordán (born 1990), Czech ice hockey player

Other people [edit]

- Michael B. Jordan (born 1987), American actor
- Michael Jordan (insolvency baron) (born 1931), English businessman
- Michael Jordan (Irish politician), Irish Farmers' Party TD from Wexford, 1927–1932
- Michael I. Jordan (born 1956), American researcher in machine learning and artificial intelligence
- Michael H. Jordan (1936–2010), American executive for CBS, PepsiCo, Westinghouse
- Michael Jordan (mycologist), English mycologist

Learning to rank

- Entity linking is often cast as a learning to rank problem: given a mention x, some set of candidate entities $\mathcal{y}$(x) for that mention, and context c, select the highest scoring entity from that set.

$$\hat{y} = \arg \max_{y \in \mathcal{Y}(x)} \Psi(y, x, c)$$

Some scoring function over the mention x, candidate y, and context c

Eisenstein 2018

Learning to rank

Michael	Jordan	can	dunk	from	the	free	throw	line
B-PER	I-PER							

1

3

1

Learning to rank

Some scoring function over the mention x, candidate y, and context c

$$\Psi(y, x, c)$$

$$\Psi(y, x, c) = f(x, y, c)^{\top}\beta$$

feature = f(x,y,c)
string similarity between x and y
popularity of y
NER type(x) = type(y)
cosine similarity between c and Wikipedia page for y

Learning to rank

- We learn the parameters of the scoring function by minimizing the **pairwise** ranking loss

$$\ell(\hat{y}, y, x, c) = \max\left(0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1\right)$$

other candidates

true candidate

Eisenstein 2018, Joachims 2002

Learning to rank

$$\ell(\hat{y}, y, x, c) = \max\left(0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1\right)$$

We suffer some loss if the predicted entity has a higher score than the true entity

$$\ell(\hat{y}, y, x, c) = \max\left(0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1\right)$$

You can't have a negative loss (if the true entity scores way higher than the predicted entity)

$$\ell(\hat{y}, y, x, c) = \max\left(0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1\right)$$

The true entity needs to score at least some constant margin better than the prediction; beyond that the higher score doesn't matter.

Neural learning to rank

Parameters measuring the compatibility of the candidate and mention

Parameters measuring the compatibility of the candidate and context

$$\Psi(y, x, c) = v_y^\top \Theta^{(x,y)} x + v_y^\top \Theta^{(y,c)} c$$

Embedding for candidate

Embedding for mention

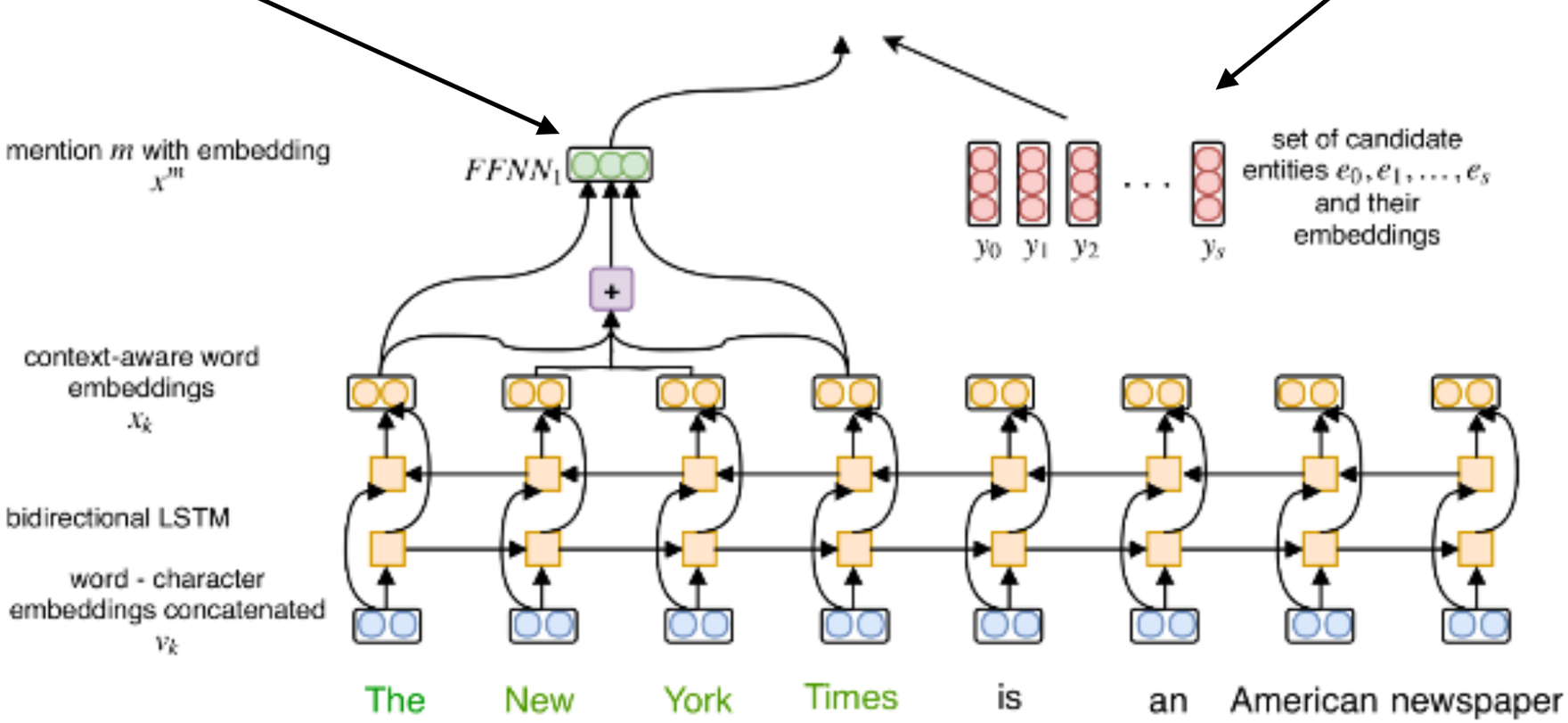
Embedding for context

Neural learning to rank

Embedding for mention

Embedding for candidate

$$s(y, x, c) = v_y^\top \Theta^{(x,y)} x + y^\top c$$



Sil et al., 2018

Information Extraction

- Named entity recognition
- Relation extraction
- Entity linking
- **Event detection**
- Event coreference
- Extra-propositional information (veridicality, hedging)

Event Detection

- Goal: identify event triggers
- optional: classify them by type. E.g. *Life Movement, Transaction, Business, Conflict, Contact*

RELATION **directed**(m1, m2)

[The Big Sleep]$_{m1}$ is a 1946 film noir directed by [Howard Hawks]$_{m2}$, the first film version of Raymond Chandler's 1939 novel of the same name.

Event Detection

- Goal: identify event triggers
- optional: classify them by type. E.g. *Life Movement, Transaction, Business, Conflict, Contact*

EVENT **directed**

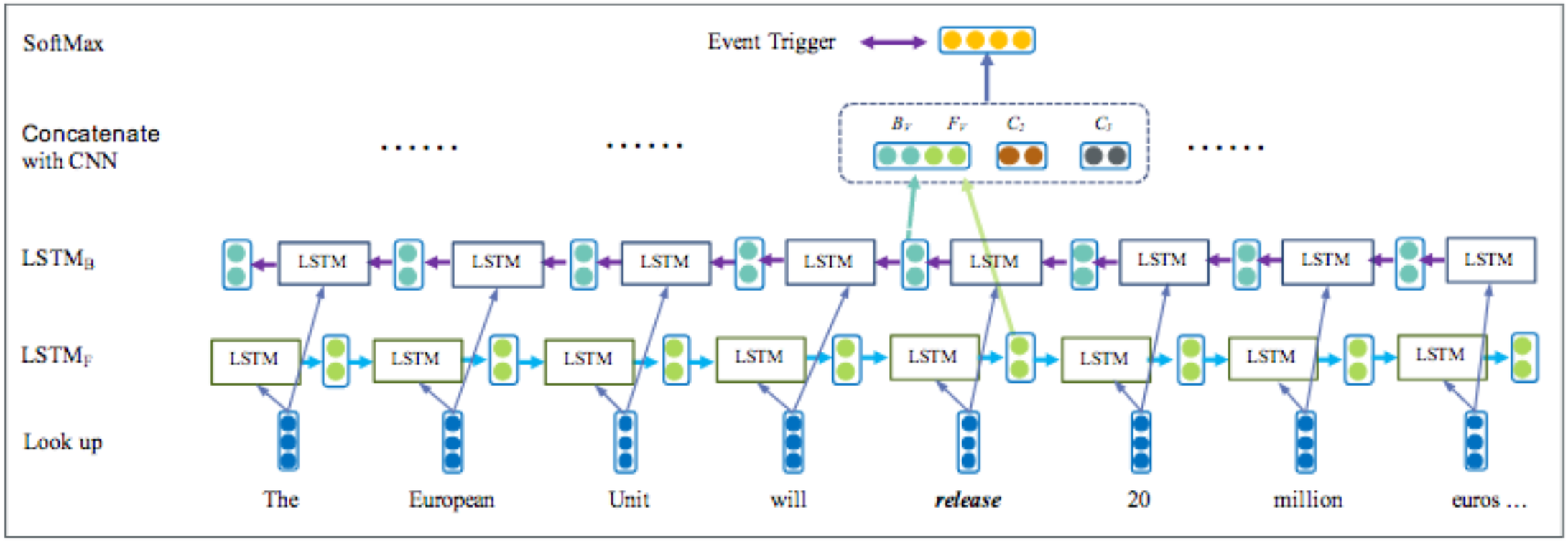
The Big Sleep is a 1946 film noir directed by Howard Hawks, the first film version of Raymond Chandler's 1939 novel of the same name.

Event Detection

- Event triggers can be verbs, nouns, adjectives, adverbs.

event	type
It **rained** last night.	verb
Her father is **retired.**	adjective
The **rioting** crowd approached the Capitol	modifier
The **attack** killed 7 and injured 20.	noun

Neural Event Detection



- Evaluation metric: Precision, Recall, Fscore over predicted vs. gold event labels.

Feng et al., 2016

Information Extraction

- Named entity recognition
- Relation extraction
- Entity linking
- Event detection
- **Event coreference**
- Extra-propositional information (veridicality, hedging)

Event Coreference

HP acquires Electronic Data Systems

Document 1

Hewlett-Packard is negotiating to buy technology services provider Electronic Data Systems.

With a market value of about $115 billion, HP could easily use its own stock to finance the purchase

If the deal is completed, it would be HP's biggest acquisition since it bought Compaq Computer Corp. for $19 billion in 2002.

Document 2

Industry sources have confirmed to eWEEK that Hewlett Packard will acquire ElectronicData Systems for about $13 billion

Bejan and Harabagiu, 2014

Event Coreference

- Similar to named entity coreference, we can train a binary classifier to predict the probability of coreference for each pair of event mentions.

- Features include similarity measures between event triggers and event arguments:

- Evaluation: B^3 Precision and Recall

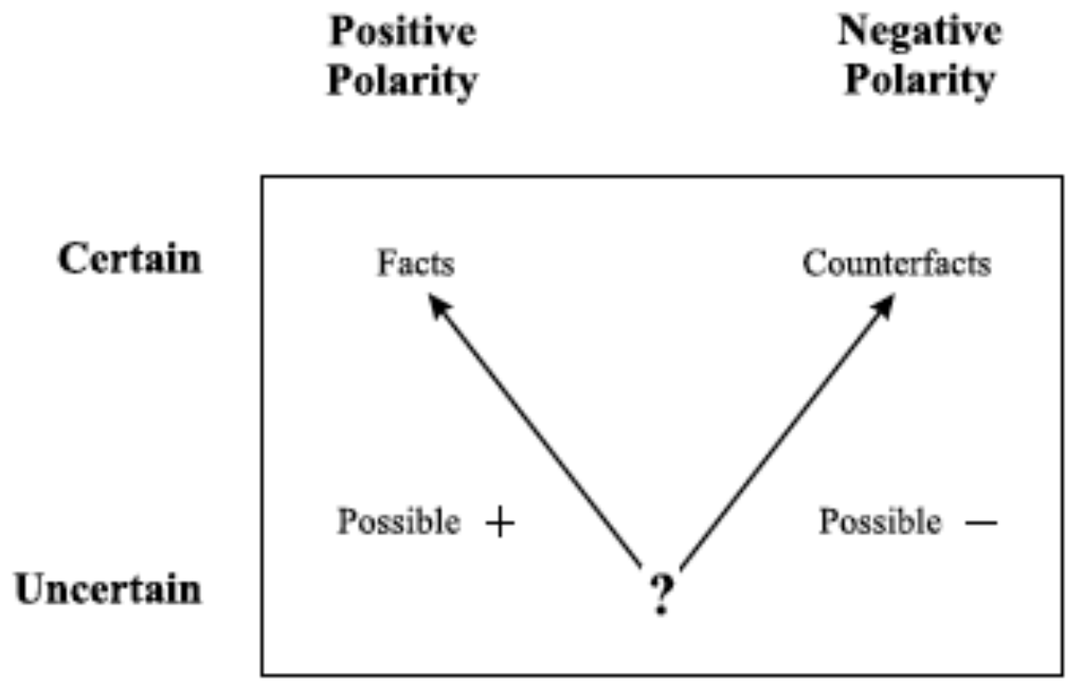
Information Extraction

- Named entity recognition
- Relation extraction
- Entity linking
- Event detection
- Event coreference
- **Extra-propositional information (veridicality, hedging)**

Factuality Detection

I locked the door

I think I locked the door



I forgot to lock the door

I possibly forgot to lock the door

Figure 1
The double range of factuality.

Sauri et al., 2012

Factuality Detection

(1) U.S. embassies and military installations around the world were ordered(3.0) to set(2.6) up barriers and tighten(2.6) security to prevent(1.8) easy access(-2.4) by unauthorized people.

(2) Intel's most powerful computer chip has flaws that could delay(0.8) several computer makers' marketing efforts(2.6), but the "bugs" aren't expected(-2.6) to hurt(-2.0) Intel.

(3) President Bush on Tuesday said(3.0) the United States may extend(1.6) its naval quarantine(2.6) to Jordan's Red Sea port of Aqaba to shut(1.4) off Iraq's last unhindered trade route.

(4) He also said(3.0) of trade(-0.8) with Iraq: "There are no shipments at the moment."

Lee et al., 2015

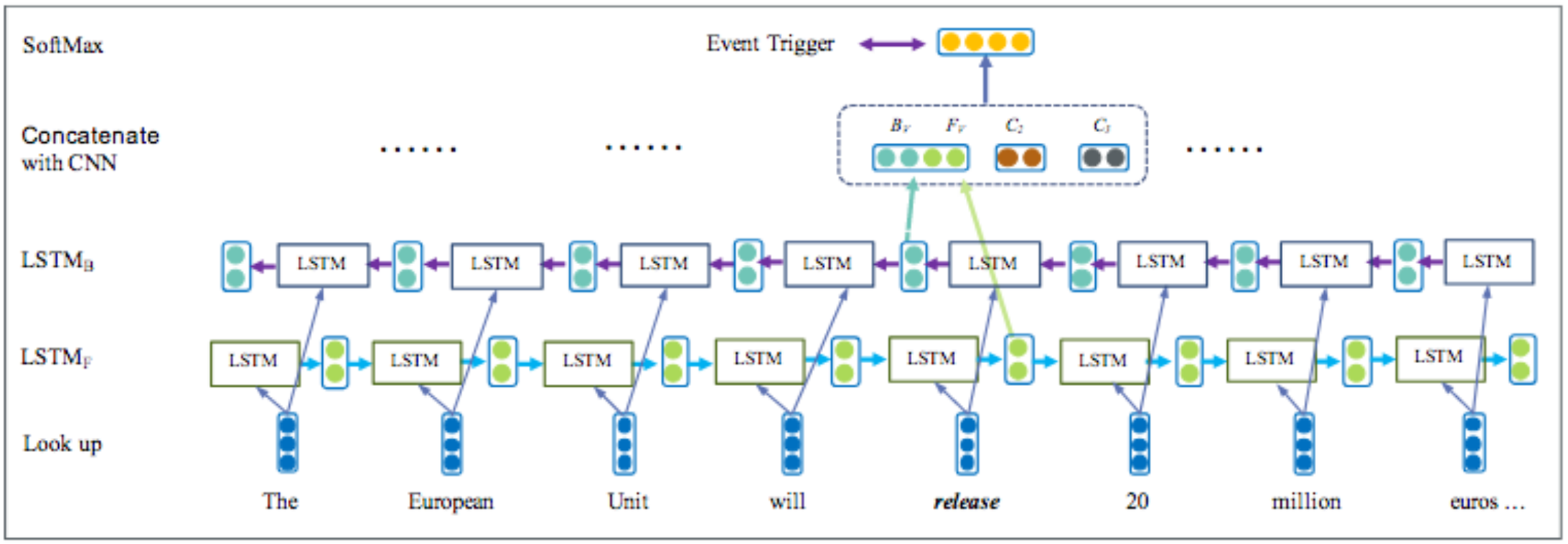
Factuality Features

$$\textbf{\textit{Score} = \textit{features}(x)}^{T} * \textbf{\textit{W}}$$

feature
token lemma, part of speech, dependency
token lemma, part of speech, dependency
presence hedge words (probably, possibly, maybe, …)
presence of implicative verbs (manage to vs. forget to)

Neural Factuality Detection

- Same architecture as event detection models, but predict a scalar value rather than a binary event indicator



Information Extraction

- Named entity recognition
- Relation extraction
- Entity linking
- Event detection
- Event coreference
- Extra-propositional information (veridicality, hedging)

Activity

- 16.ie/EntityLinking_TODO.ipynb