TweeTime:

A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter

Jeniya Tabassum
Alan Ritter
Wei Xu

THE OHIO STATE UNIVERSITY

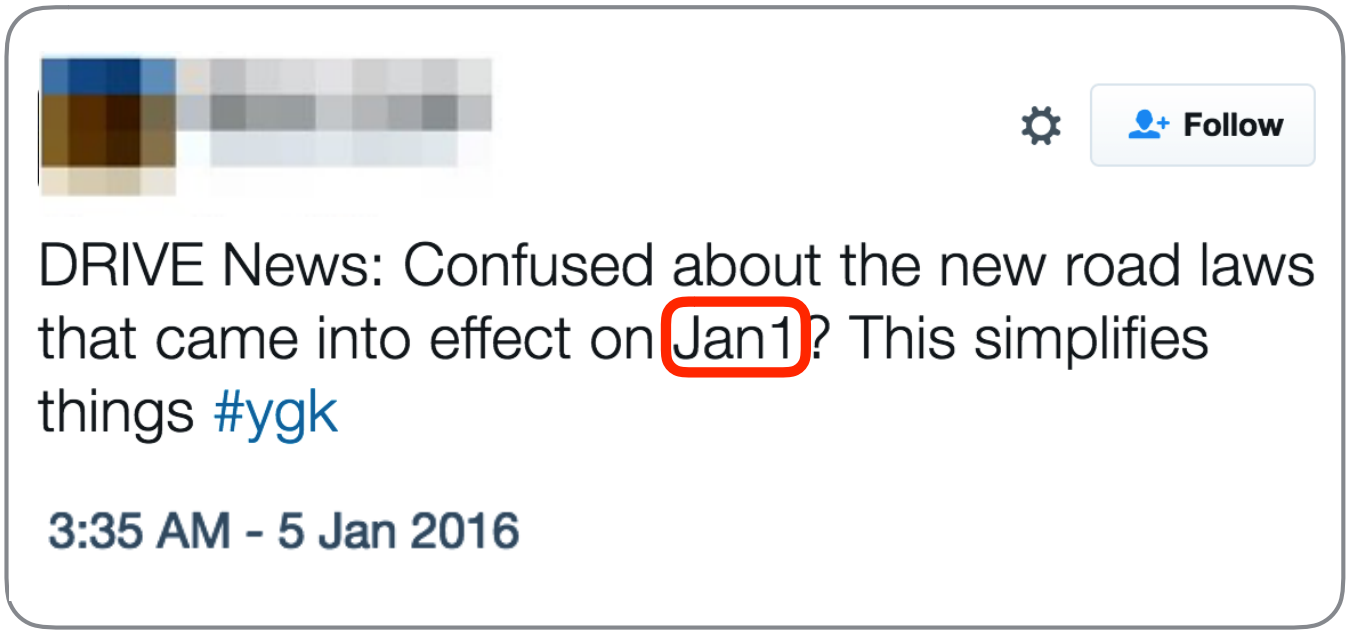
TweeTime:

A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter

Jeniya Tabassum
Alan Ritter
Wei Xu

The Ohio State University

Time Resolution from Tweets



Time Resolution from Tweets

Challenge: Diversity

Tomorrow

2m 2ma 2mar 2mara 2maro
2marrow 2mor 2mora 2moro
2morow 2morr 2morro 2morrow
2moz 2mr 2mro 2mrrw 2mrw 2mw
tmmrw tmo tmoro tmorrow tmoz tmr
tmro tmrow tmrrow tmrrw tmrw
tmrww tmw tomaro tomarow
tomarro tomarrow tomm tommarow
tommarrow tommoro tommorow
tommorrow tommorw tommrow
tomo tomolo tomoro tomorow
tomorro tomorrw tomoz tomrw tomz

Rule-Based
- HeidelTime [Strötgen & Gertz, 2013]
- SUTime [Chang & Manning, 2012]
- TempEX [Mani & Wilson, 2000]

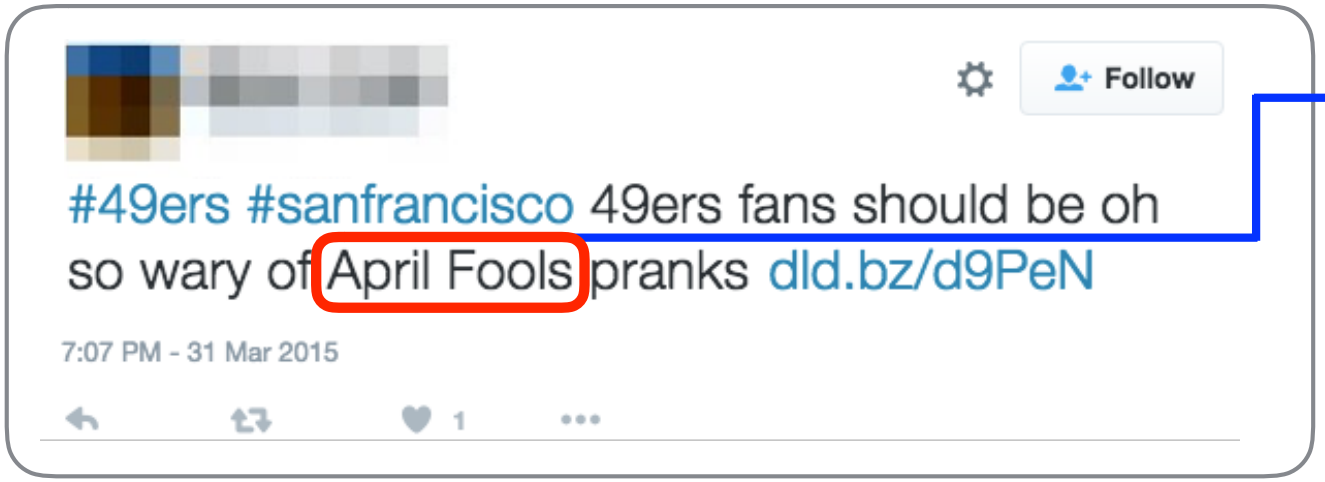
Previous **rule based systems** do not handle noisy text.

Previous **supervised systems** perform poorly due to domain mismatch.

Supervised — UWTime [Lee et al., 2014]

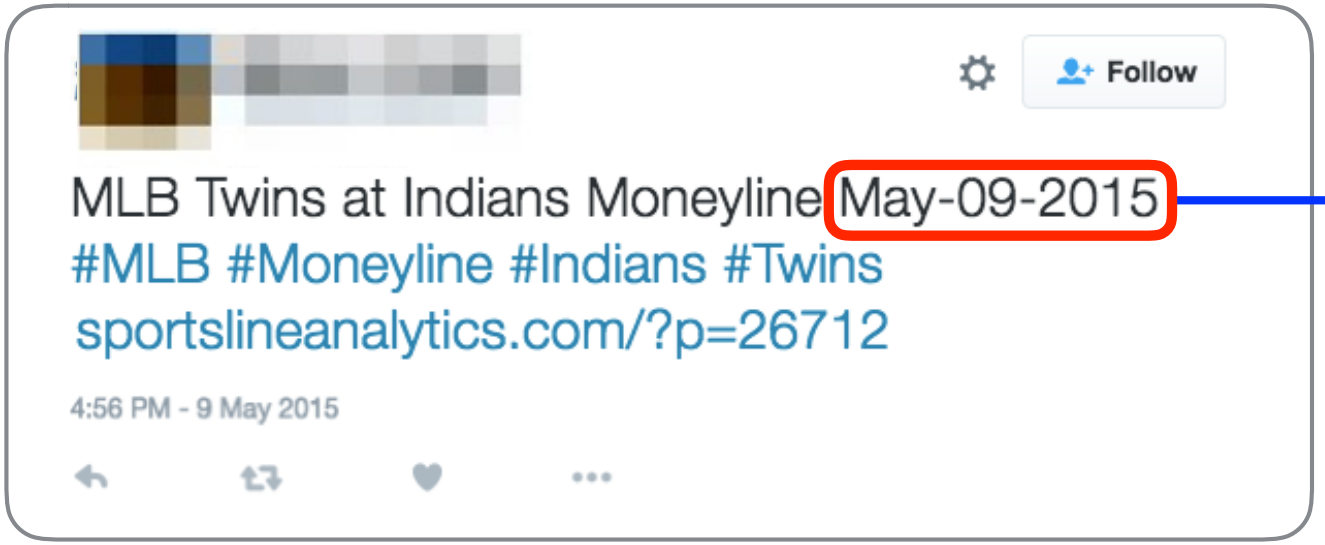
Requires Human Labels — ParsingTime [Angeli et al., 2012/2013]

Social Media is Hard



No Temporal Expressions

Missing Rule



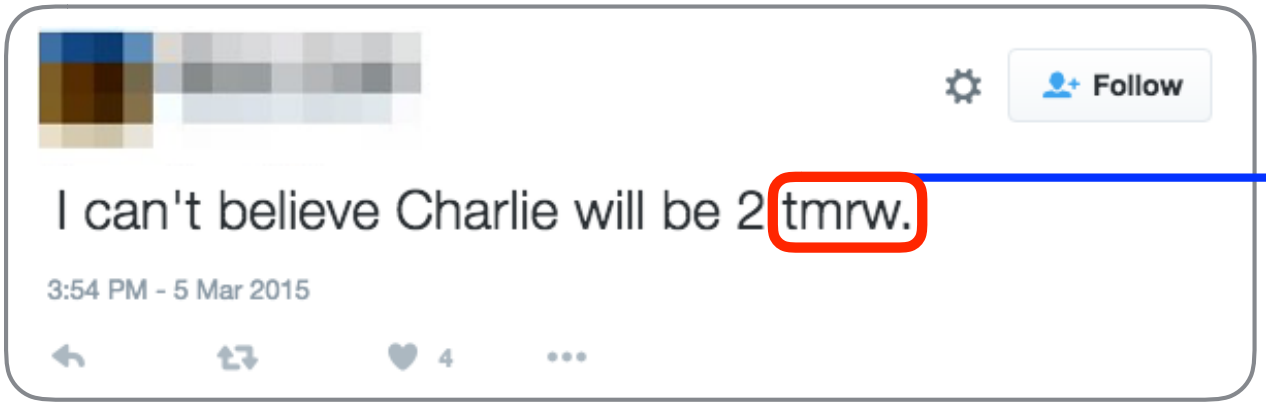
No Temporal Expressions

Tokenization

Social Media is Hard

No Temporal Expressions

Hashtag



No Temporal Expressions

Spelling Variation

Date Resolution in Social Media

Event Extraction

Disease Outbreaks

[Kanhabua et al., 2012]

Cyber Security

Victim	Date
spamhaus	2013/03/18
soca	2011/06/20
etrade	2012/01/05
interpol	2012/02/29
ustream	2012/05/09

[Ritter et al., 2015]

[Chang et al., 2016]

Date Resolution in Social Media

Event Extraction

Disease Outbreaks

[Kanhabua et al., 2012]

Cyber Security

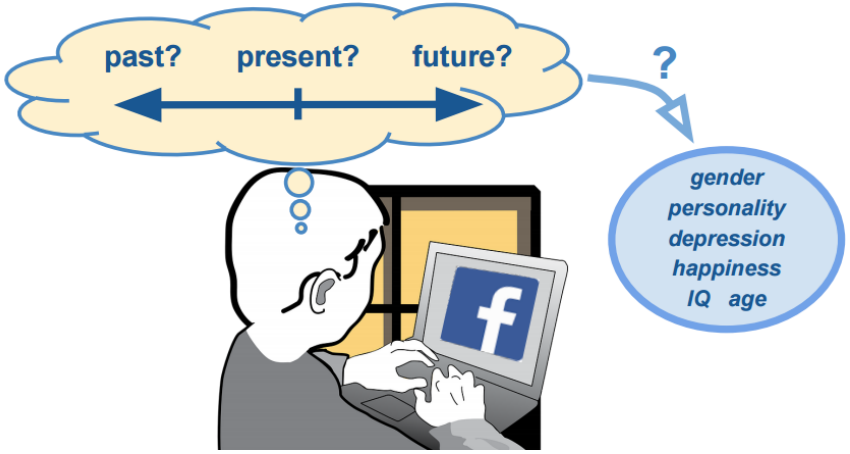
Victim	Date
spamhaus	2013/03/18
soca	2011/06/20
etrade	2012/01/05
interpol	2012/02/29
ustream	2012/05/09

[Ritter et al., 2015]

[Chang et al., 2016]

Social Science

Temporal Orientation



[Schwartz et al., 2015]

No existing temporal resolver for social media

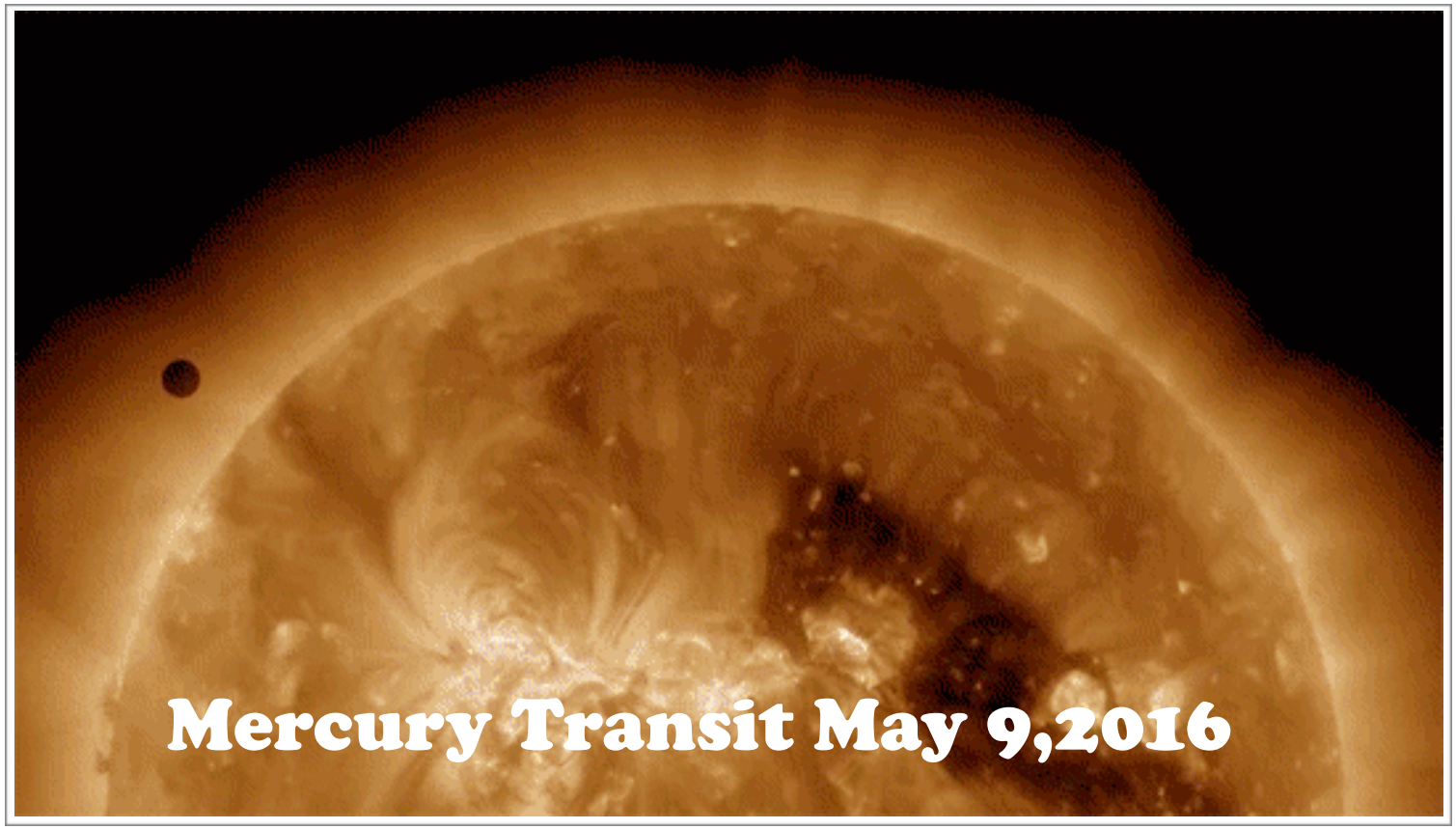
TweeTime

can handle noisy Twitter Text

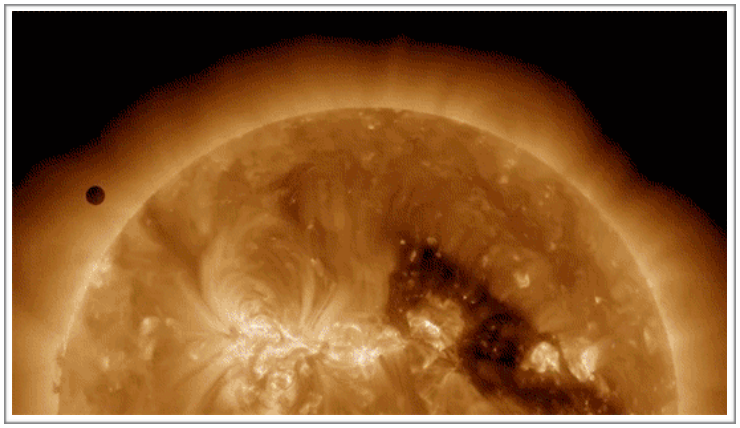
Distant Supervision

(no human labels needed!)

Distant Supervision Assumption



Distant Supervision Assumption



8 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

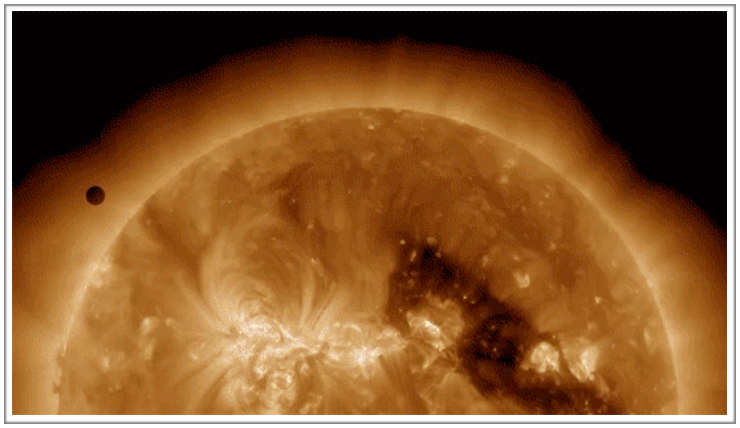
10 May

I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday! May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption



8 May

9 May

10 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather

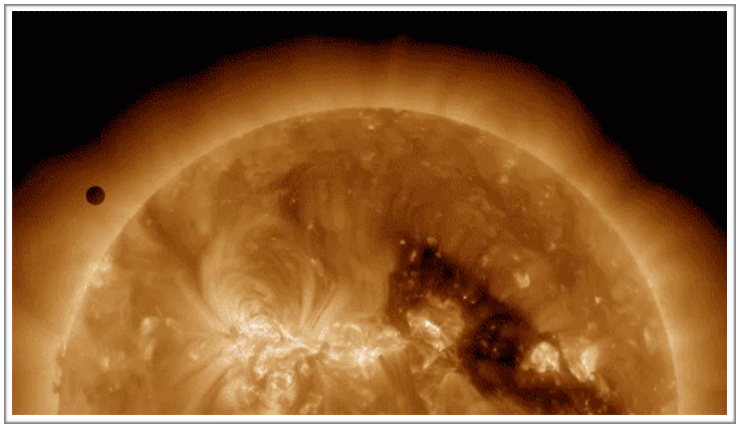
3:54 PM - 9 May 2016

I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption



8 May

9 May

10 May

Follow

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Follow

Mercury Transit 2morrow starting at 6:00 AM
Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

Follow

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

Follow

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

Follow

I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Follow

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Tweets posted near an event that mention a key entity are likely to **contain time expressions** referring to the event's date.

Contribution

- First distant Supervision approach for date resolution
  - Most prior work on relation extraction

    [Mintz et al., 2009b; Riedel et al., 2010; Hoffmann et al., 2011]

- Novel **multiple-instance-learning tagging model**
  - Learns word-level tags using only sentence labels

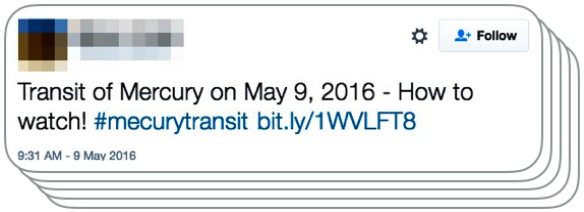
- State of the art results on social media domain
  - 17% increase in F-score over SUTime

System Overview

[EVENT EXTRACTOR]

NER

TempEx



(simple time expressions)

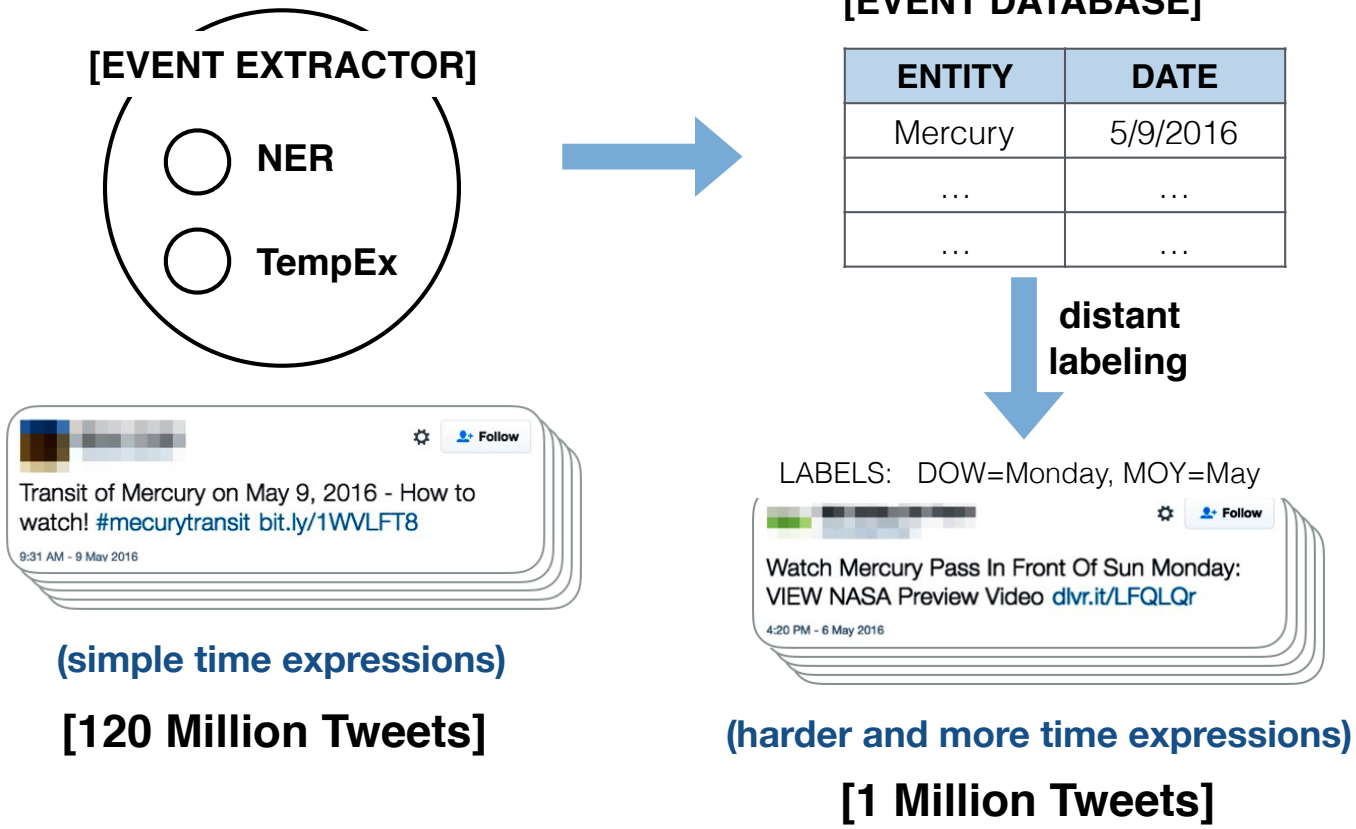
[120 Million Tweets]

[10,000 events]

[EVENT DATABASE]

ENTITY	DATE
Mercury	5/9/2016
…	…
…	…

System Overview



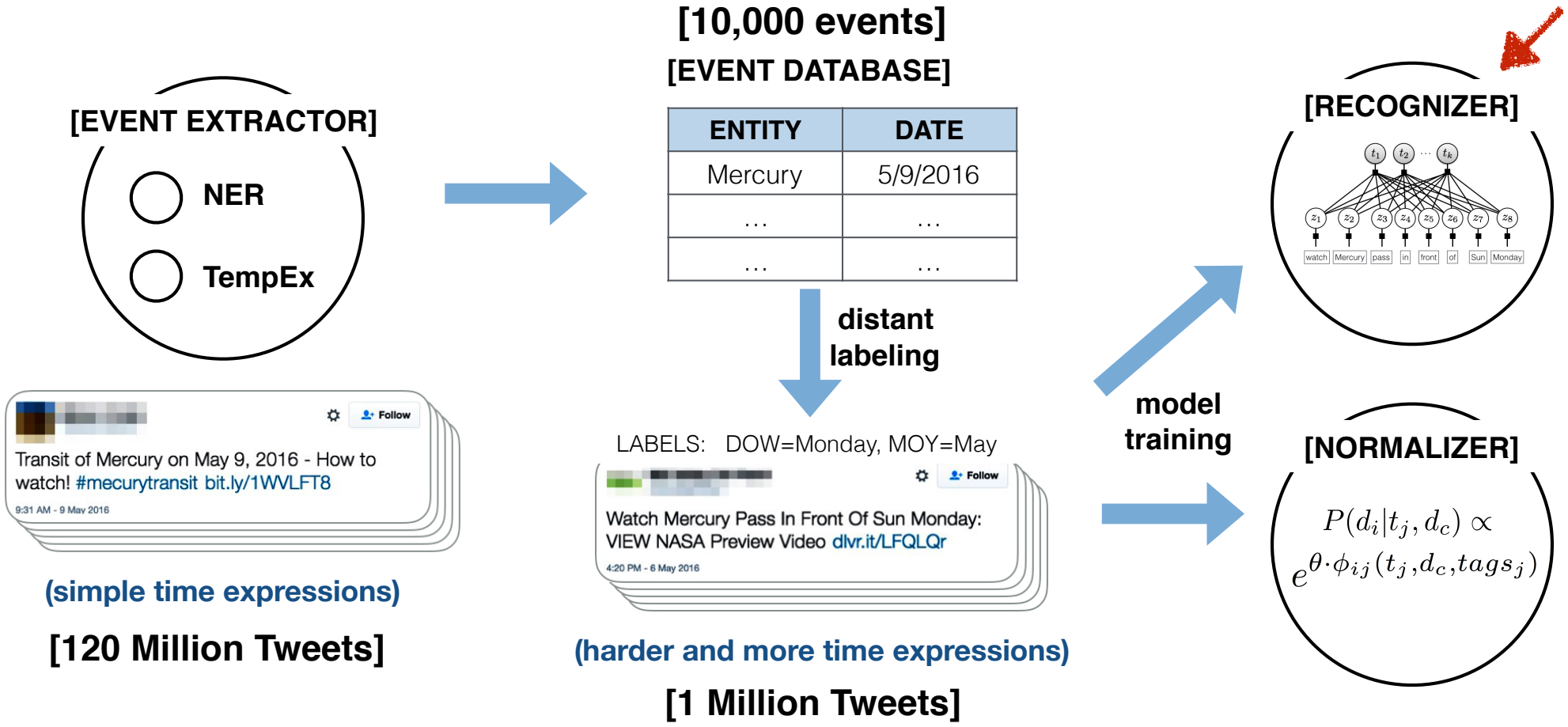
(simple time expressions)

[120 Million Tweets]

(harder and more time expressions)

[1 Million Tweets]

System Overview



Temporal Recognizer

[Event Database]

ENTITY	DATE
Mercury	5/9/2016
…	…

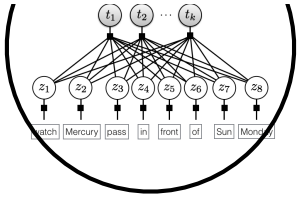
LABELS: Day of Week, Day of Month, Month, Timeline

< Monday, 9, May, Future >

[1 Million Tweets]

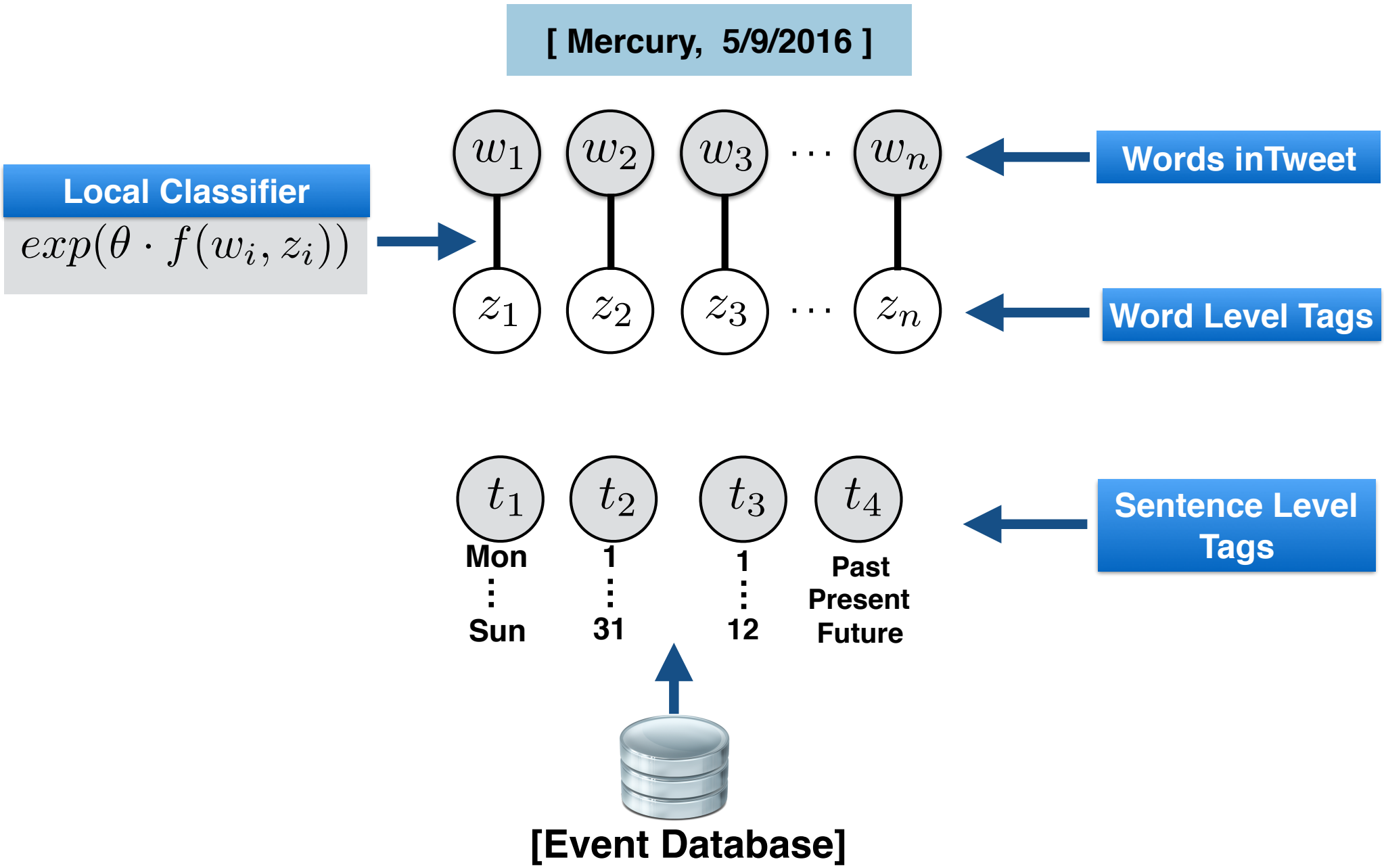
[automatically labeled]

[RECOGNIZER]

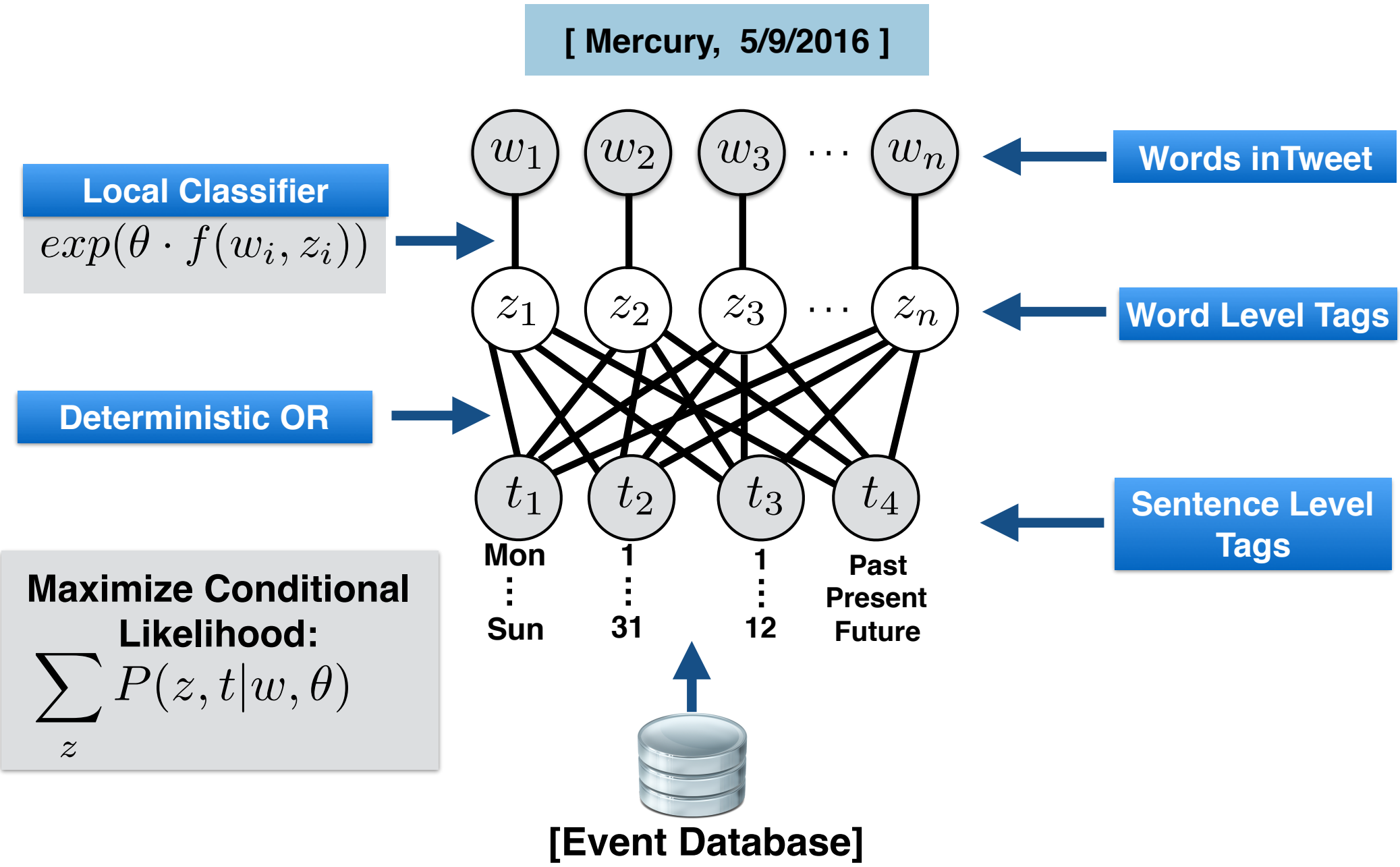


Word	Happy	Friday!	#Fri	yey	#TGIF
Tag	NA	FRI	FRI	NA	FRI

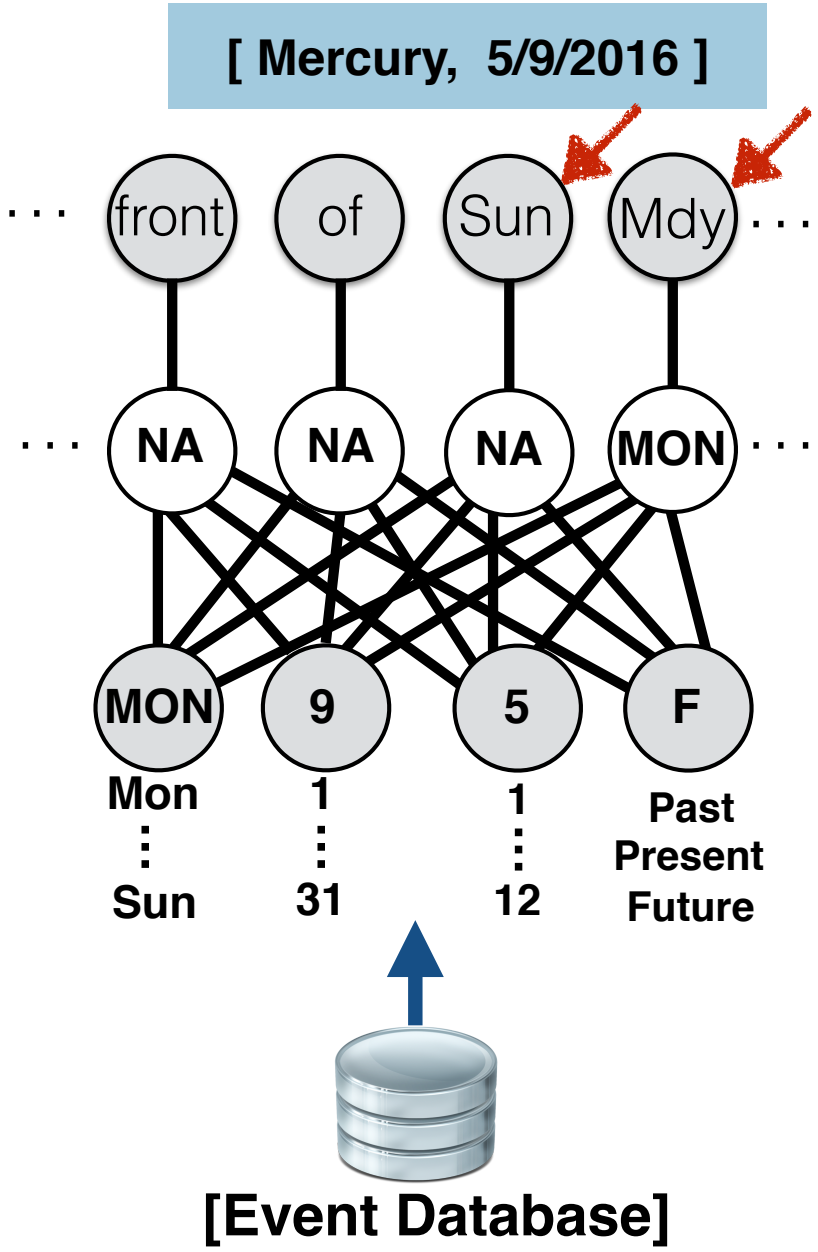
MultiT Tagger



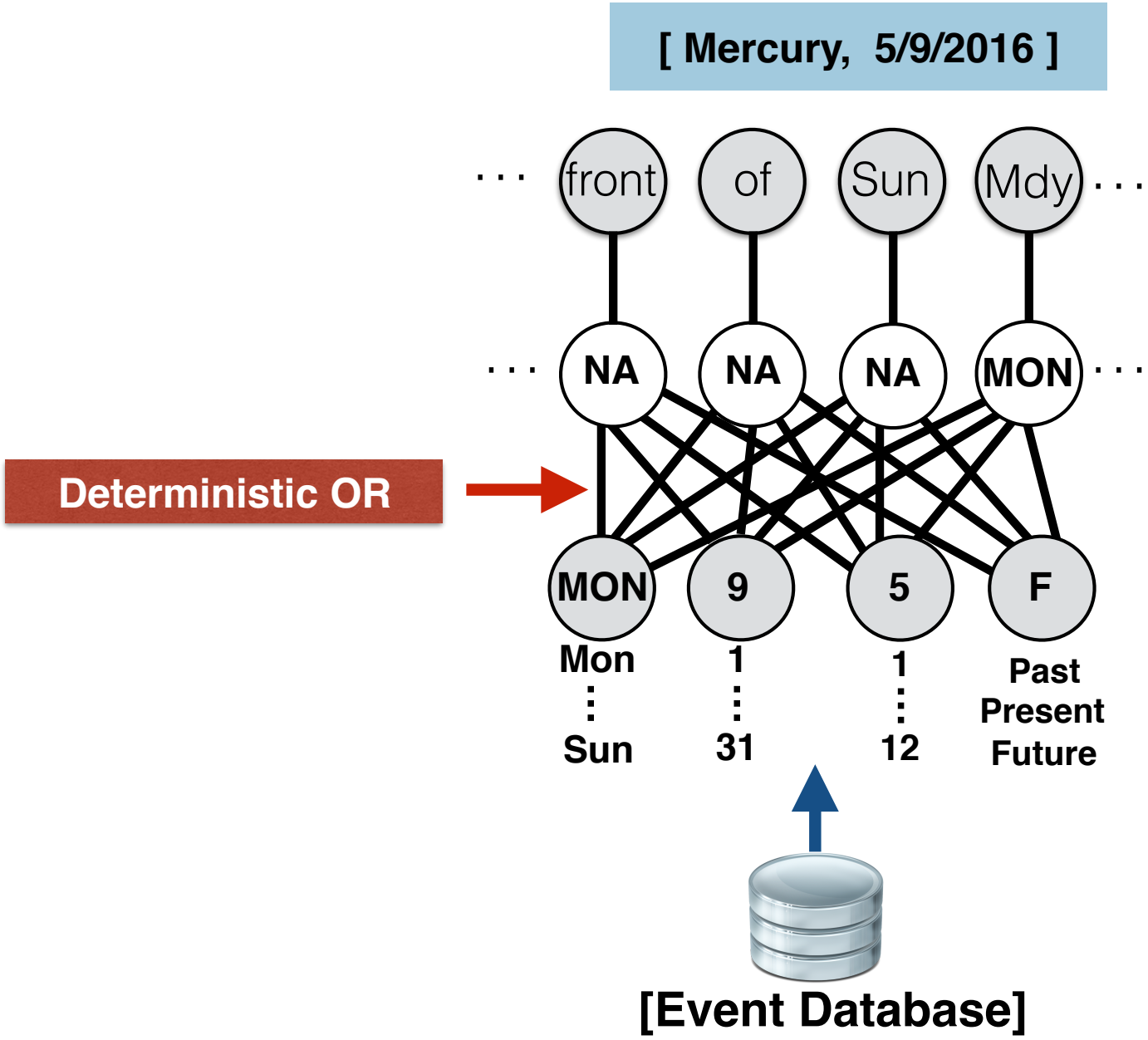
MultiT Tagger



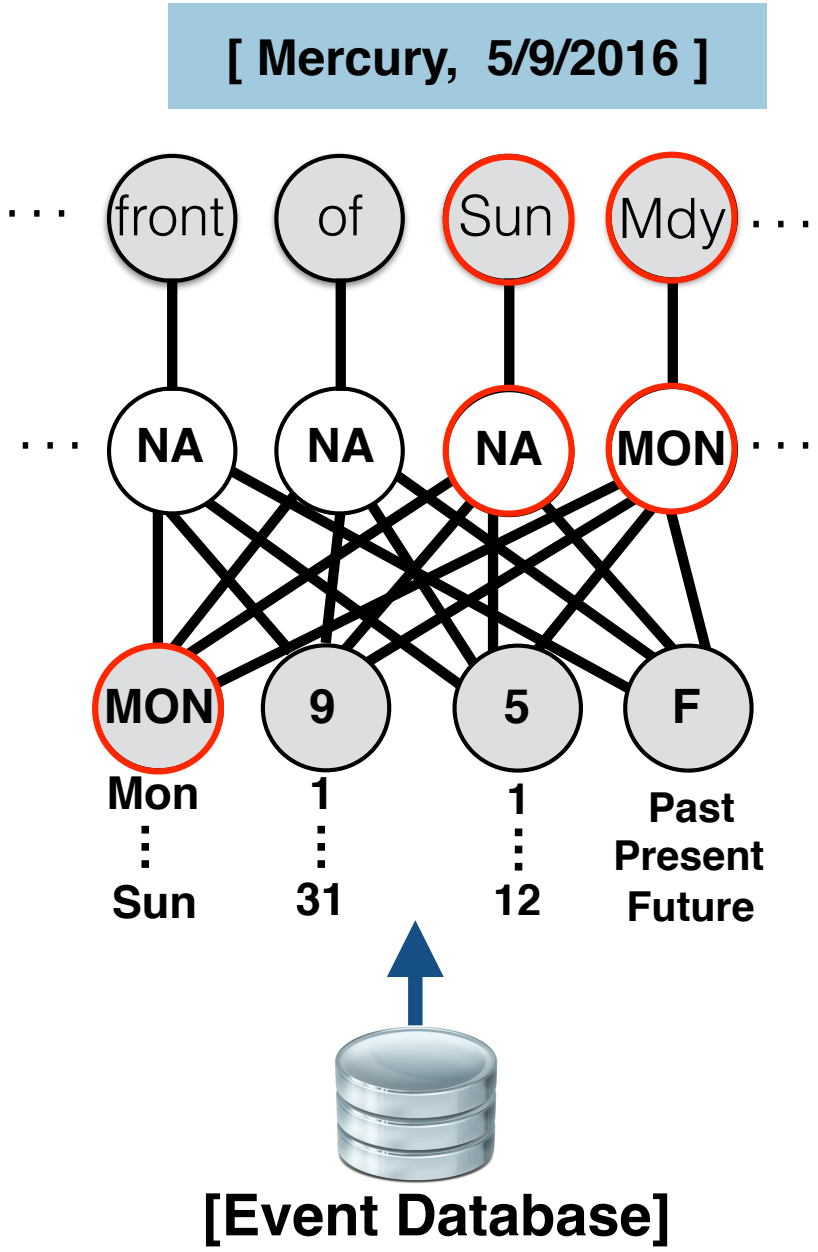
MultiT Tagger



MultiT Tagger



MultiT Tagger



MultiT : Learning

Latent-Variable Perceptron (MAP-based learning)

Gradient of Conditional Log-Likelihood:

$$\nabla P(t|w) = \sum_{z} P(z|w,t;\theta) \cdot f(z,w) \; - \; \sum_{t,z} P(t,z|w;\theta) \cdot f(z,w)$$

MultiT : Learning

Latent-Variable Perceptron (MAP-based learning)

Gradient of Conditional Log-Likelihood:

$$\nabla P(t|w) = \sum_{z} P(z|w,t;\theta) \cdot f(z,w) \; - \sum_{t,z} P(t,z|w;\theta) \cdot f(z,w)$$

$$\nabla P(t|w) = \max_{P(z|w,t;\theta)} f(z,w) \; - \max_{P(t,z|w;\theta)} f(z,w)$$

Weighted Edge Cover Problem

MultiT : Learning

Latent-Variable Perceptron (MAP-based learning)

Gradient of Conditional Log-Likelihood:

$$\nabla P(t|w) = \sum_{z} P(z|w,t;\theta) \cdot f(z,w) \; - \sum_{t,z} P(t,z|w;\theta) \cdot f(z,w)$$

$$\nabla P(t|w) = \max_{P(z|w,t;\theta)} f(z,w) \; - \max_{P(t,z|w;\theta)} f(z,w)$$

Assumptions: MultiT Tagger

$$Sentence\ Level\ Tags \in Word\ Level\ Tags$$

Assumptions: MultiT Tagger

Sentence Level Tags:

TL = Future
MOY= May
DOM=9
DOW= Mon

$$Sentence\ Level\ Tags \in Word\ Level\ Tags$$

Assumptions: MultiT Tagger

Sentence Level Tags:

TL = Future

MOY= May

DOM=9

DOW= Mon

$$Sentence\ Level\ Tags \in Word\ Level\ Tags$$

$$Word\ Level\ Tags \in Sentence\ Level\ Tags$$

Missing Data Problem

Sentence Level Tags:

TL = Future
MOY= May
DOM=9
DOW= Mon

Missing Data Problem

Sentence Level Tags:

- ✅ TL = Future
- ❌ MOY= May
- ❌ DOM=9
- ✅ DOW= Mon

Missing Data Problem

Sentence Level Tags:

- ✅ TL = Future
- ❌ MOY= May
- ❌ DOM=9
- ✅ DOW= Mon

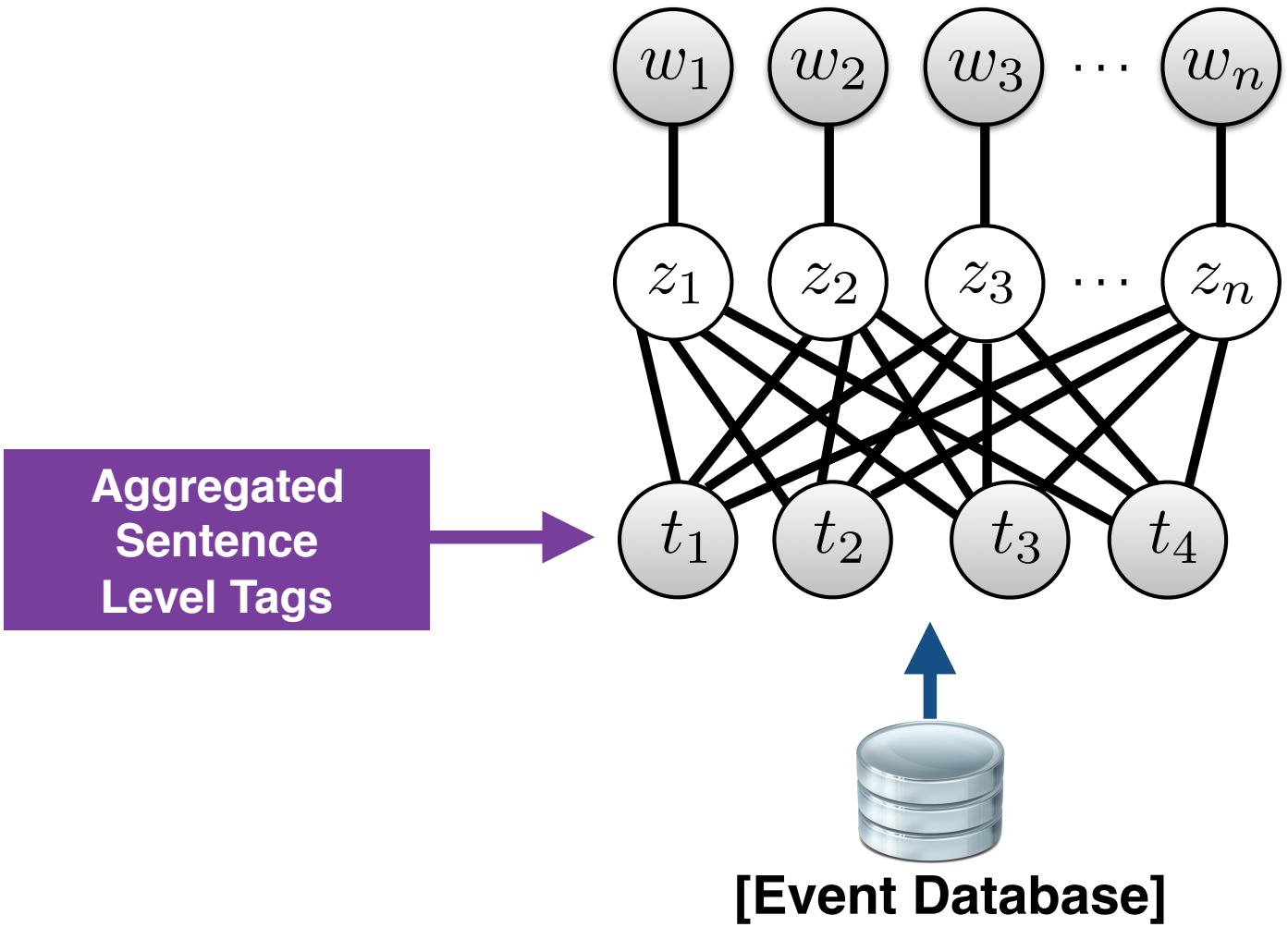
$$Sentence\ Level\ Tag \notin Word\ Level\ Tag$$

$$Word\ Level\ Tag \notin Sentence\ Level\ Tag$$

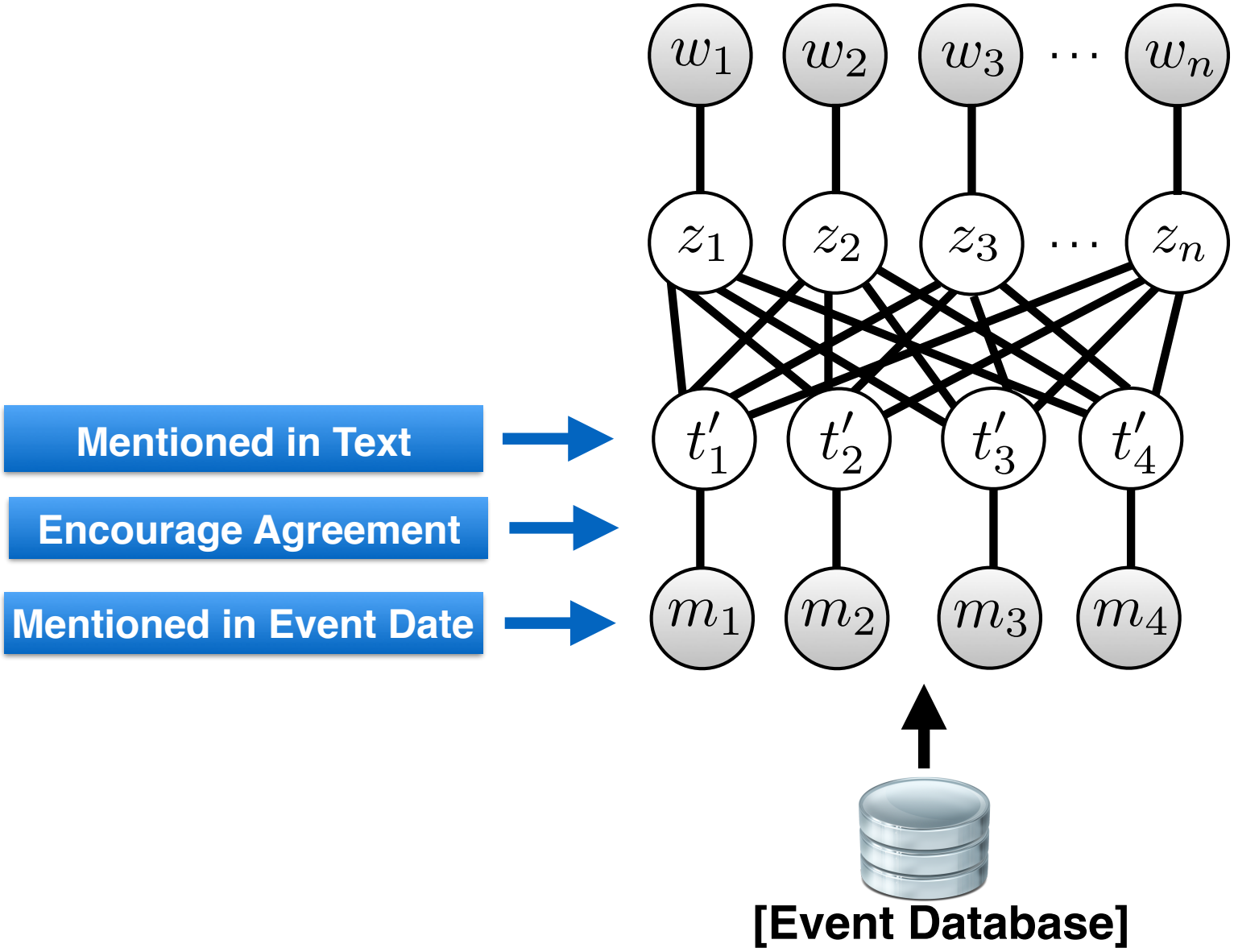
Solution:

Multiple Instance Learning Missing Data Model

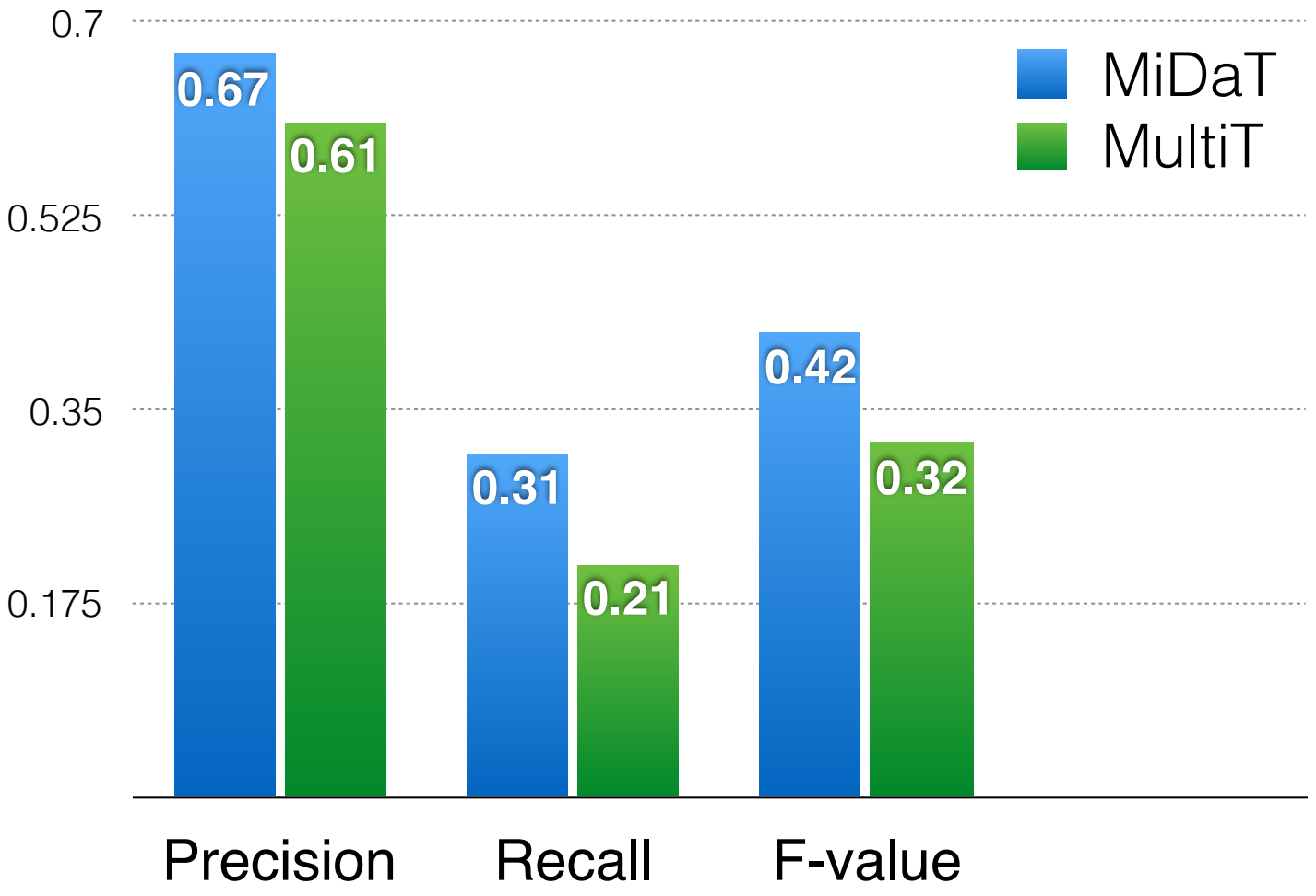
Missing Data Extension



MiDaT Tagger [Extension of MultiT]



MultiT Vs MiDaT [Intrinsic evaluation]



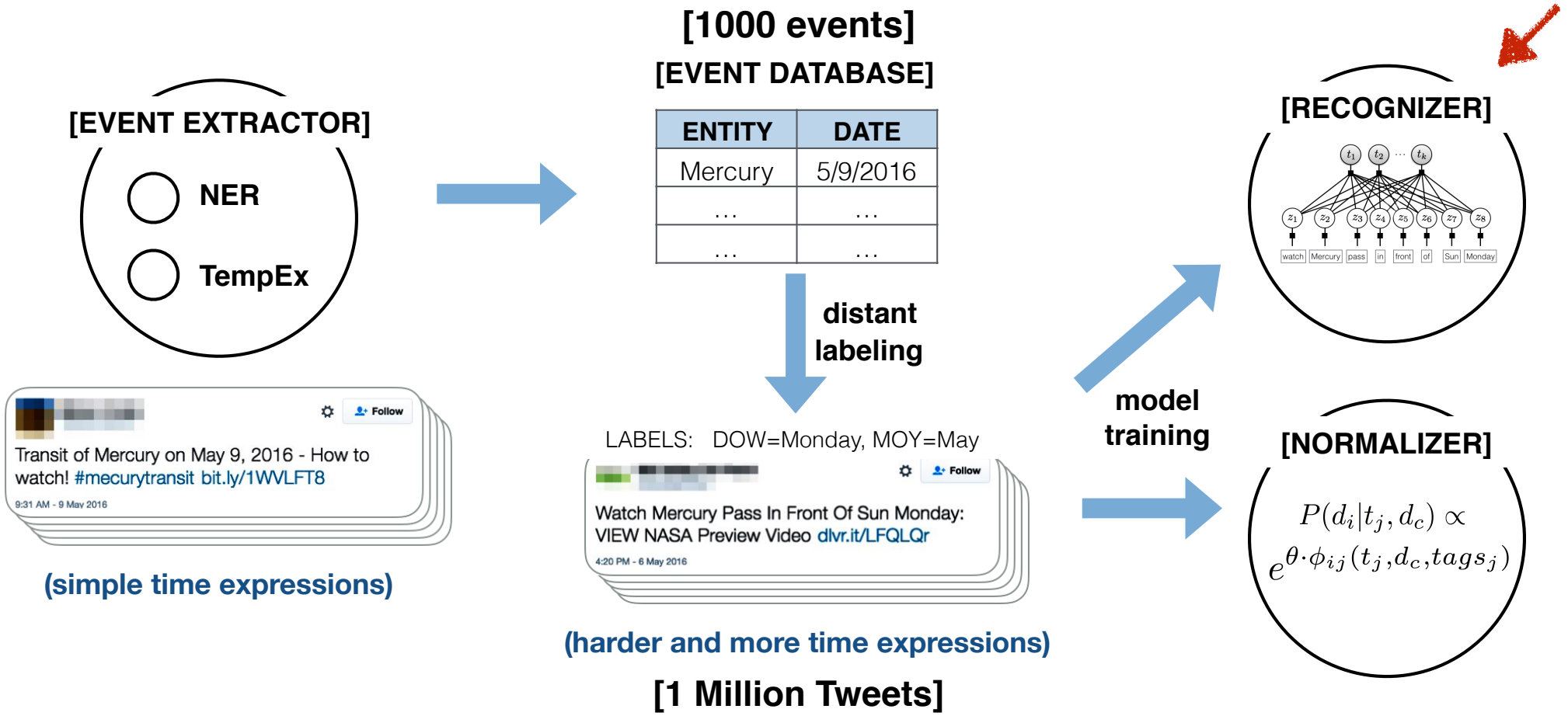
Example Tags

Word	Im	Hella	excited	for	tomorrow
Tag	NA	NA	Future	NA	Future

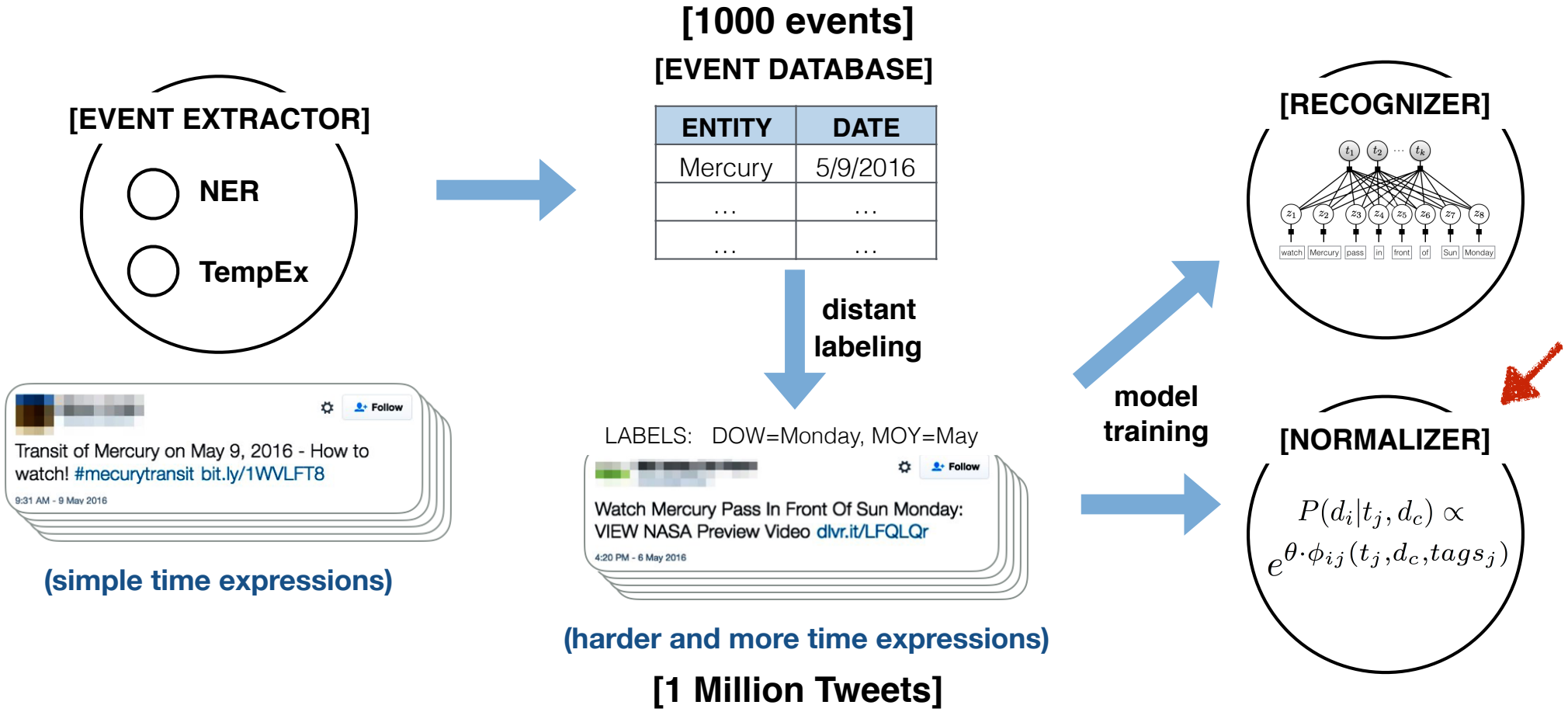
Word	Thnks	for	a	Christmas	party	on	fri
Tag	NA	NA	NA	Dec	NA	NA	FRI

[Not always matches human intuition]

System Overview



System Overview



Temporal Normalizer

[Event Database]

LABEL: May 9 2016

[$\frac{1}{2}$ Million Tweets]

[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

Temporal Normalizer

[Event Database]

LABEL: May 9 2016

[$\frac{1}{2}$ Million Tweets]

[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

Temporal Normalizer

[Event Database]

LABEL: May 9 2016

[$\frac{1}{2}$ Million Tweets]

Creation Date: 11/1

EMNLP starts 4m tmrw!

[NORMALIZER]

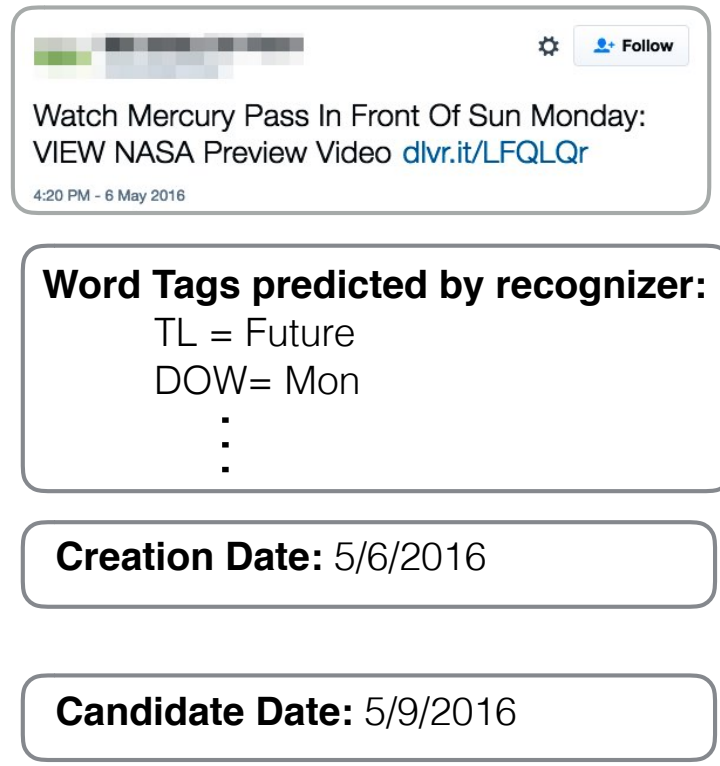
$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

D_{-10}	...	D_{-1}	D_0	D_{+1}	...	D_{+10}	$None$
10/22		10/31	11/1	11/2		11/11	NA

Temporal Normalizer: Features

$P($ **Candidate Date:** 5/9/2016 $| \cdots)$ $\propto f$



$\cdot \theta$

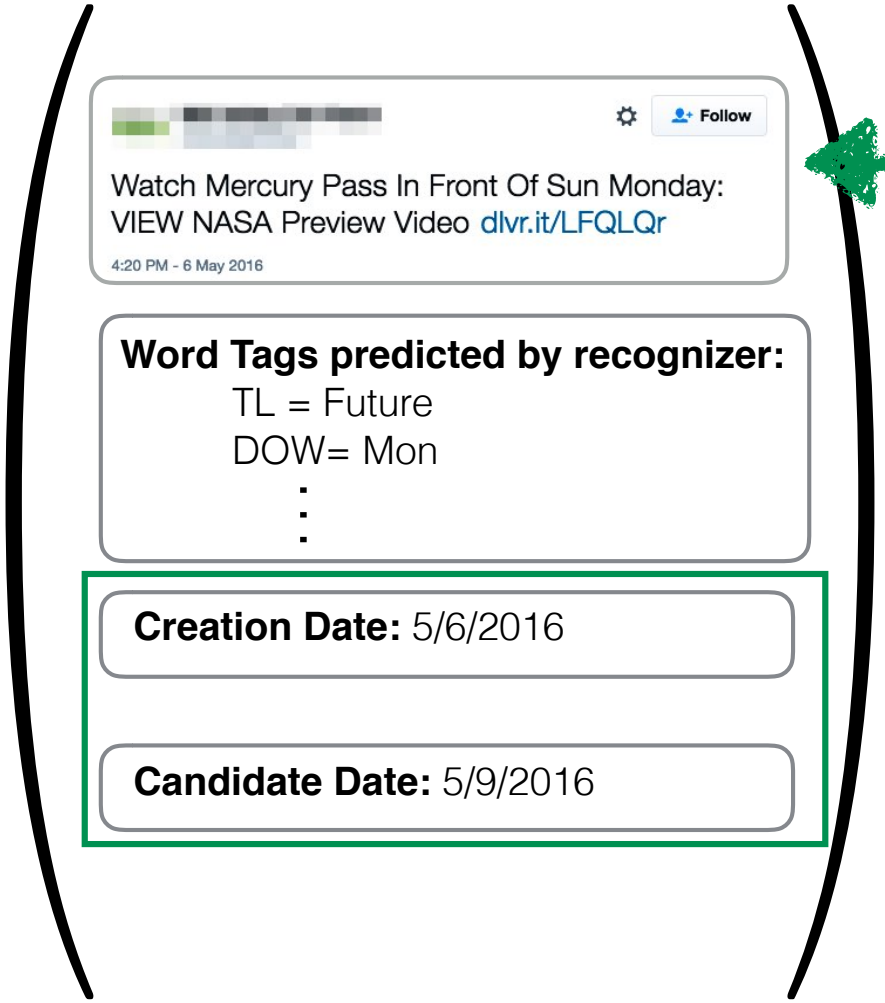
Lexical Features

Tag Features

Time Difference Features

Temporal Normalizer: Features

$$P\left(\boxed{\textbf{Candidate Date: } 5/9/2016 \;\middle|\; \cdots}\right) \propto f$$



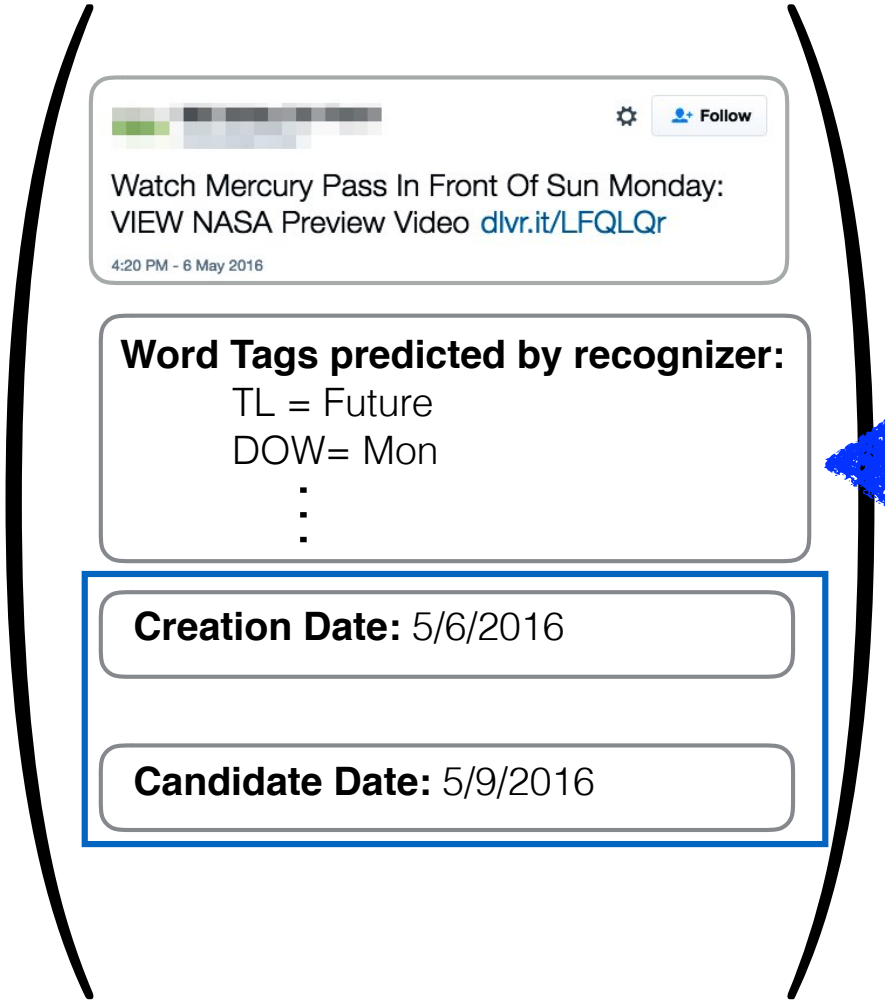
$$\cdot\,\theta$$

Lexical Features

$$w = \text{“Monday”} \wedge \text{candidate date} \in Monday$$

Temporal Normalizer: Features

$P($ **Candidate Date:** 5/9/2016 $\mid \cdots) \propto f$



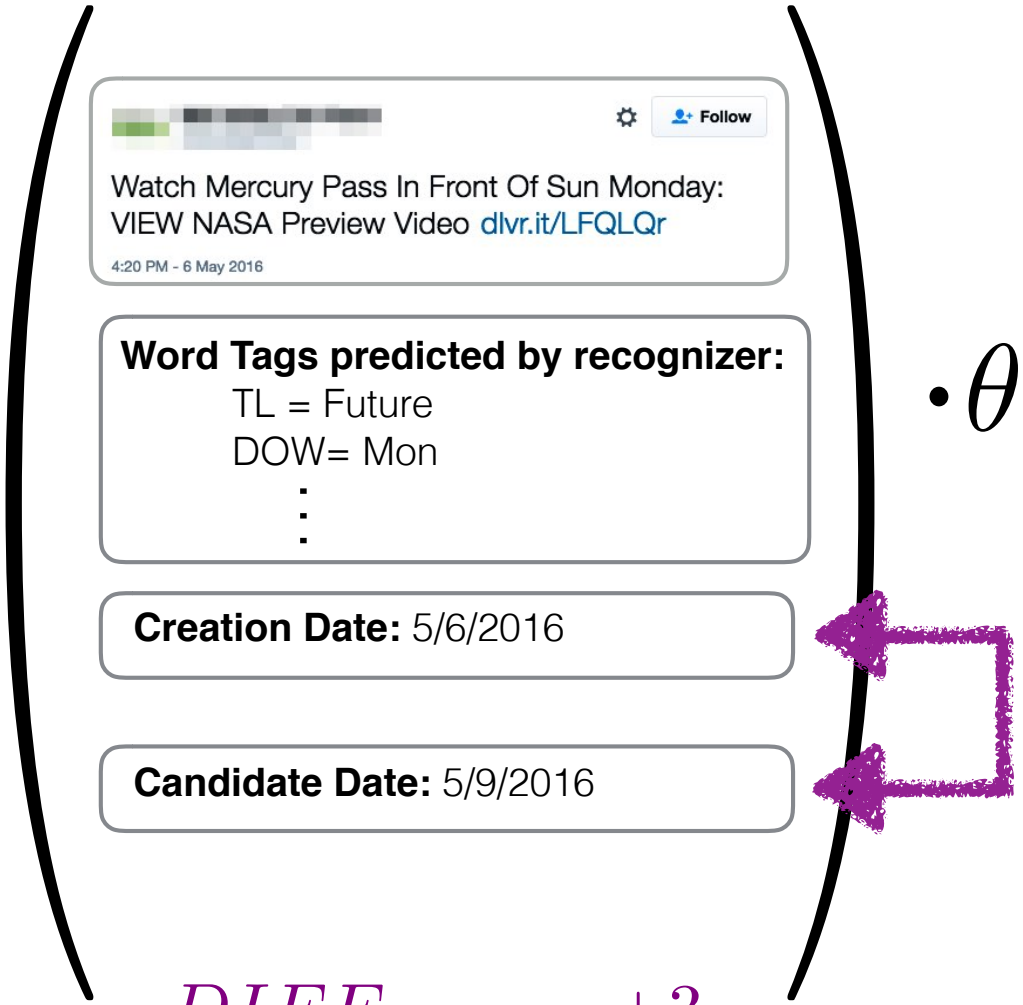
$\cdot\,\theta$

Tag Features

$t = \textsc{Future} \wedge \text{candidate date} \in future$

Temporal Normalizer: Features

$$P\left(\boxed{\textbf{Candidate Date: } 5/9/2016 \;\middle|\; \cdots}\right) \propto f$$

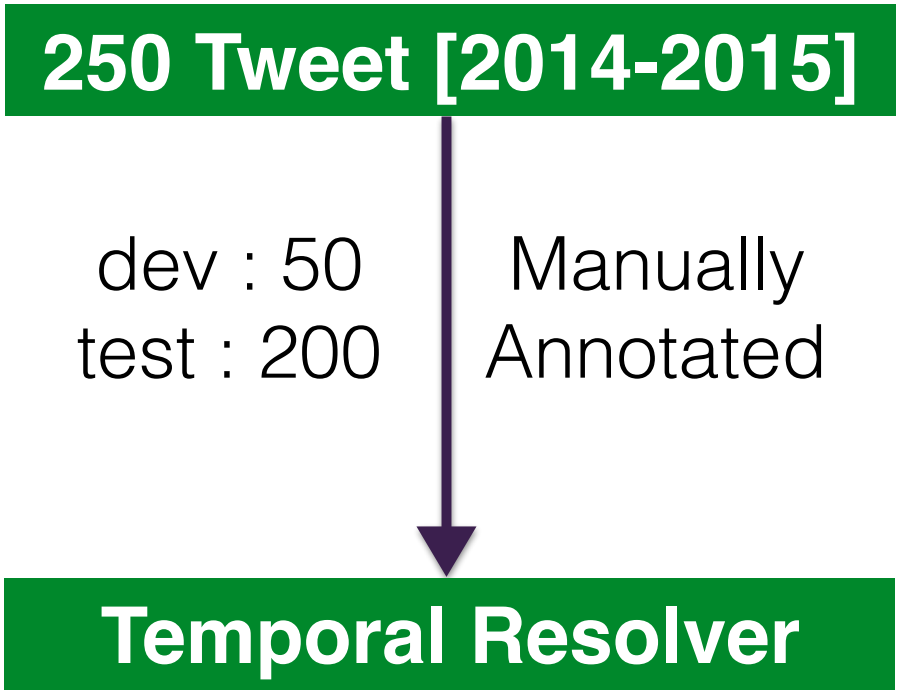


$$\cdot\,\theta$$

Time Difference Features

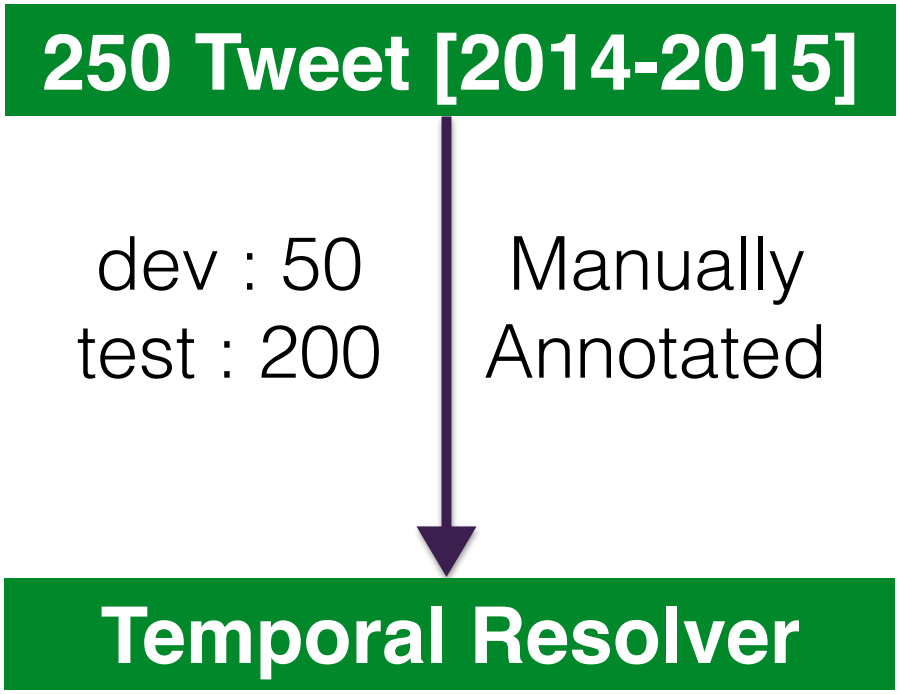
$$DIFF_{days} = +3$$
$$DIFF_{weeks} = +1$$

Evaluation



[TweeTime, HeidelTime, UWTime, SUTime, TempEX]

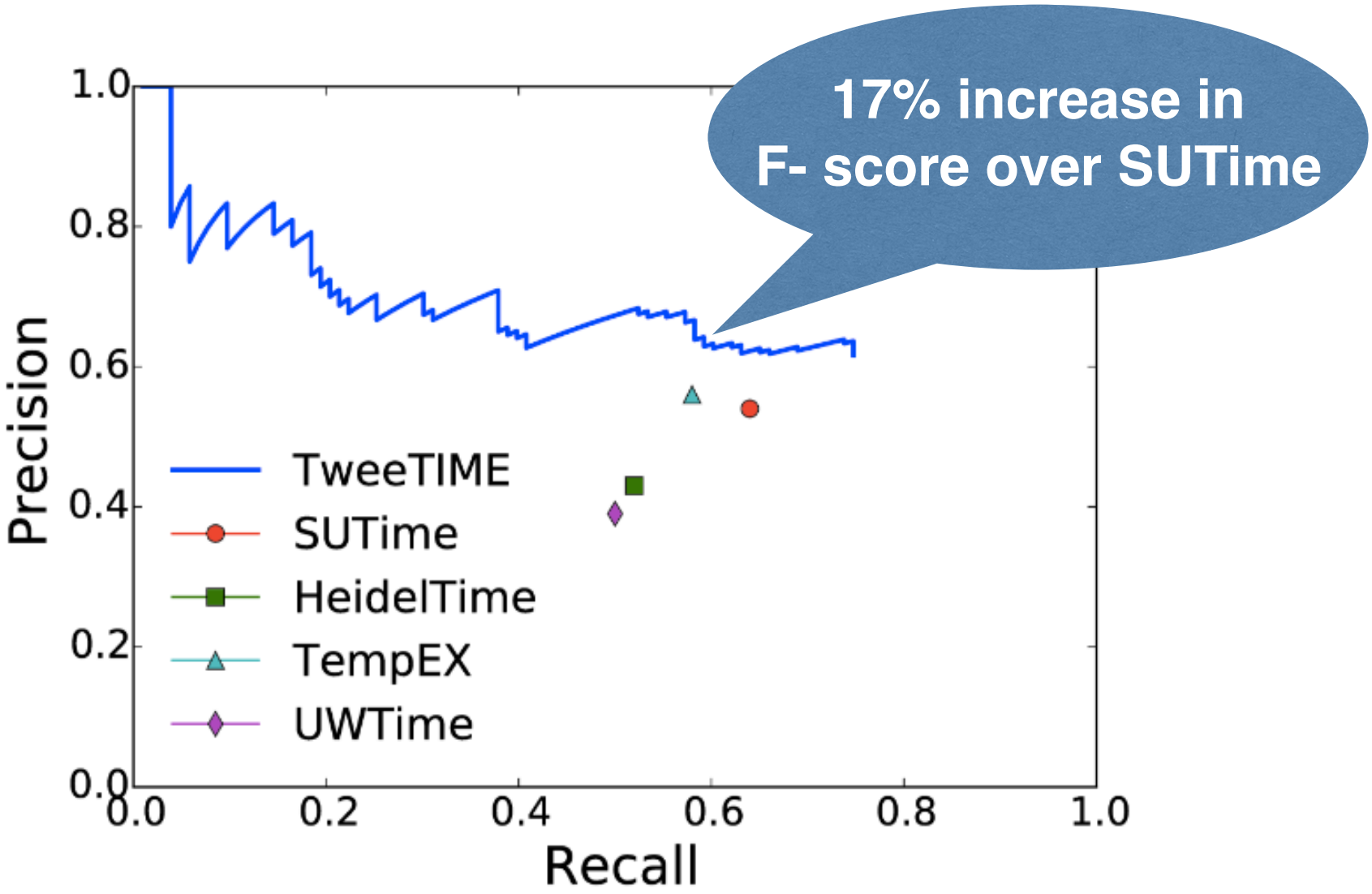
Evaluation



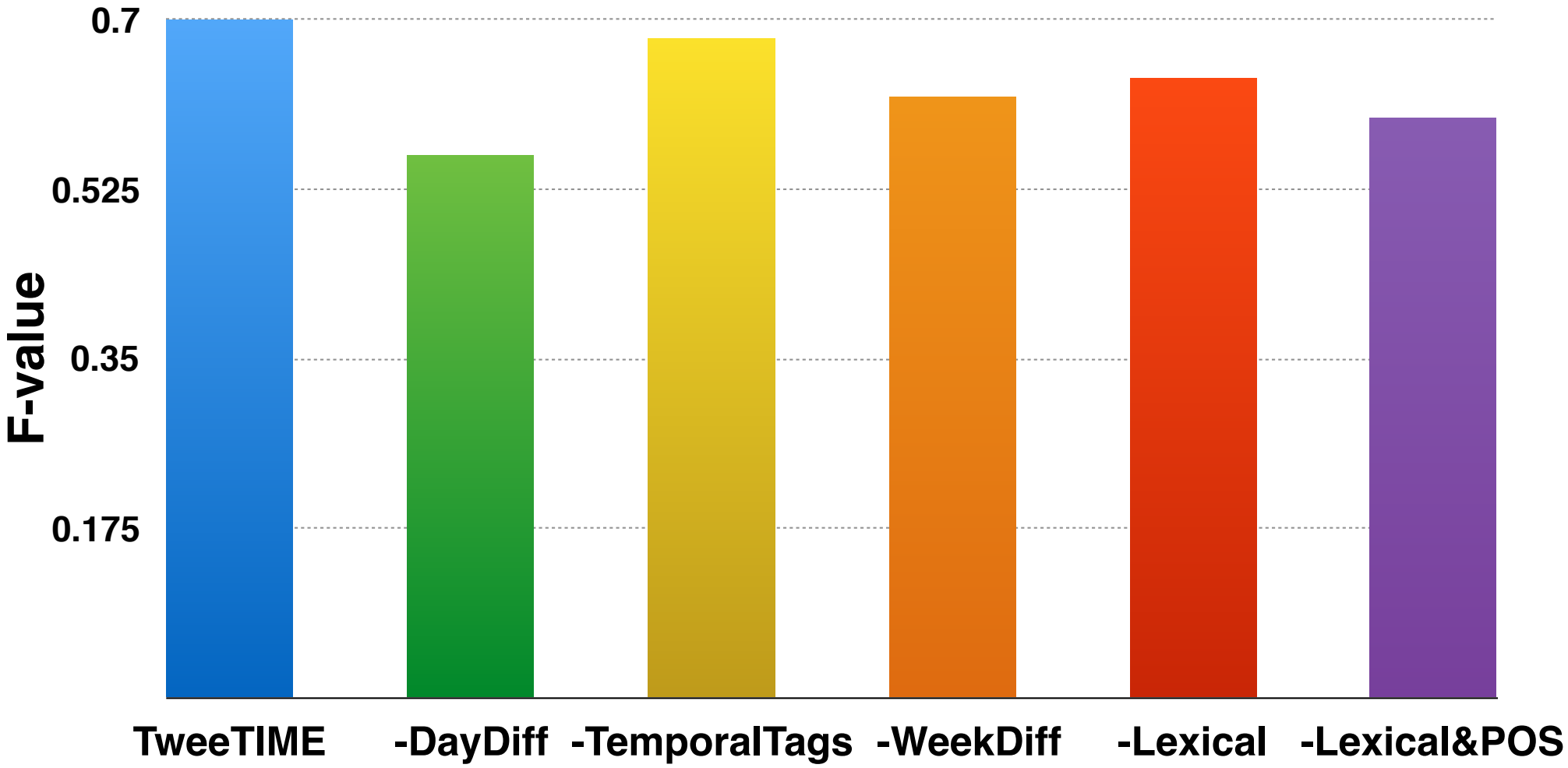
[**TweeTime**, HeidelTime, UWTime, SUTime, TempEX]

Train Set [2011-2012]

System Performance



Feature Ablation



Takeaways

- 1st and best date resolver for Twitter
- 1st use of distant supervision for time expressions (**no human labels needed**)
- Code will be available: ***https://github.com/jeniyat/TweeTime***

Takeaways

- 1st and best date resolver for Twitter
- 1st using distant supervision for time expressions (**no human labels needed**)
- Code will be available: ***https://github.com/jeniyat/TweeTime***