

TweeTime:

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

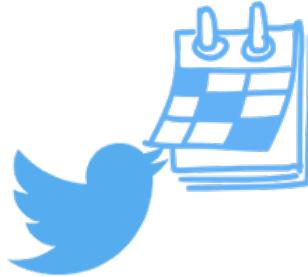
Jeniya Tabassum

Alan Ritter

Wei Xu



THE OHIO STATE UNIVERSITY



TweeTime:

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

Jeniya Tabassum

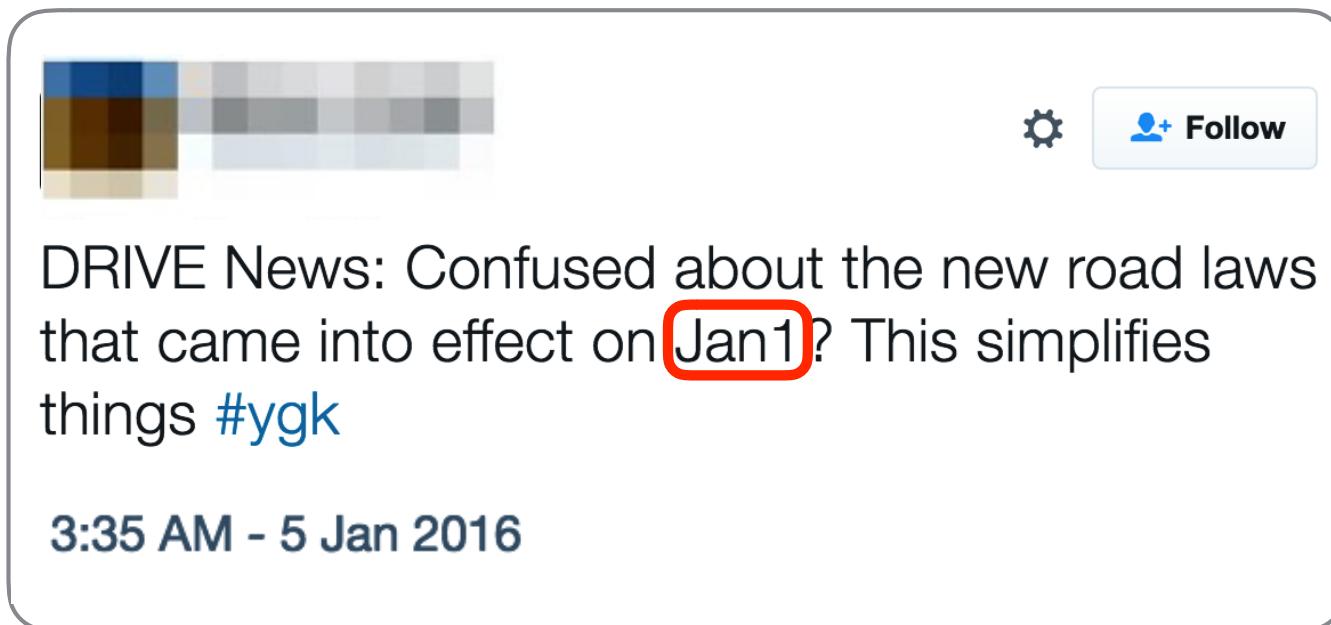
Alan Ritter

Wei Xu



THE OHIO STATE UNIVERSITY

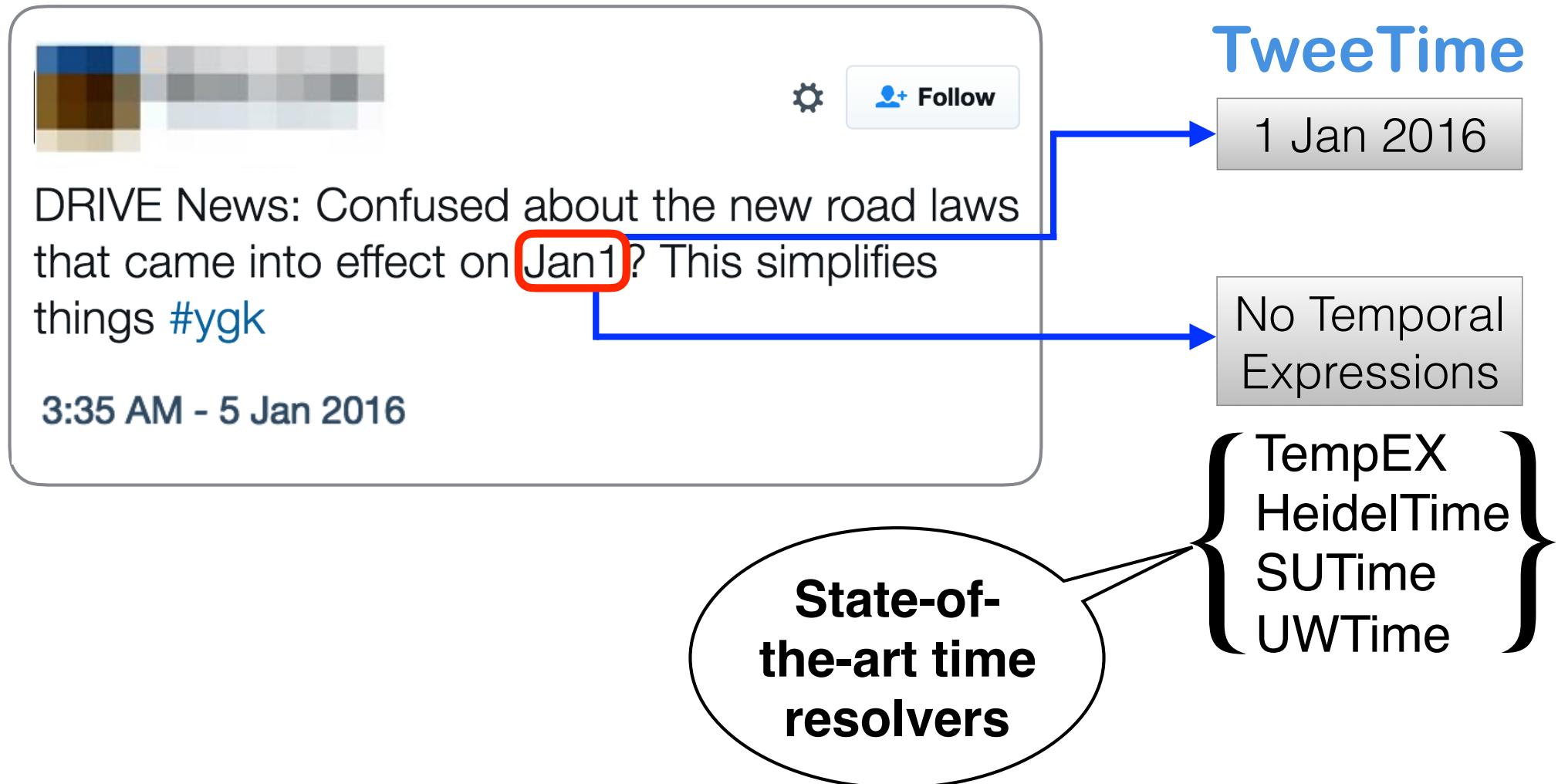
Time Resolution from Tweets



 DRIVE News: Confused about the new road laws that came into effect on Jan1? This simplifies things #ygk

3:35 AM - 5 Jan 2016

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 tomorrow tomm tommarow
tommarrow tommoro tommorow
tommorrow tommorw tommrow
tomo tomolo tomoro tomorrow
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

A screenshot of a Twitter post. The profile picture is blurred. The tweet text is: "#49ers #sanfrancisco 49ers fans should be oh so wary of April Fools pranks dld.bz/d9PeN". A blue box highlights the phrase "so wary of April Fools pranks". A red box highlights the date "April Fools". Below the tweet are standard Twitter interaction icons: reply, retweet, like (1), and more.

No Temporal Expressions

Missing Rule

A screenshot of a Twitter post. The profile picture is blurred. The tweet text is: "MLB Twins at Indians Moneyline May-09-2015 #MLB #Moneyline #Indians #Twins sportslineanalytics.com/?p=26712". A red box highlights the date "May-09-2015". Below the tweet are standard Twitter interaction icons: reply, retweet, like, and more.

No Temporal Expressions

Tokenization

Social Media is Hard



Day11....chilling out in #friends house #shoes
#hottubtimemachine #tv #saturday #night
instagram.com/p/1WSRWlqjyq/

1:33 PM - 11 Apr 2015

Follow

No Temporal Expressions

Hashtag

This block displays a tweet from a user (@jessica_williams) at 1:33 PM on April 11, 2015. The tweet contains several hashtags: #friends, #shoes, #hottubtimemachine, #tv, #saturday, and #night. The word '#saturday' is highlighted with a red box. A blue arrow points from this red box to a gray callout box labeled 'No Temporal Expressions'. Another blue arrow points from the same red box to a red callout box labeled 'Hashtag'.



I can't believe Charlie will be 2 tmrw.

3:54 PM - 5 Mar 2015

Follow

No Temporal Expressions

Spelling Variation

This block displays a tweet from a user (@jessica_williams) at 3:54 PM on March 5, 2015. The tweet reads 'I can't believe Charlie will be 2 tmrw.' The word 'tmrw.' is highlighted with a red box. A blue arrow points from this red box to a gray callout box labeled 'No Temporal Expressions'. Another blue arrow points from the same red box to a red callout box labeled '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]

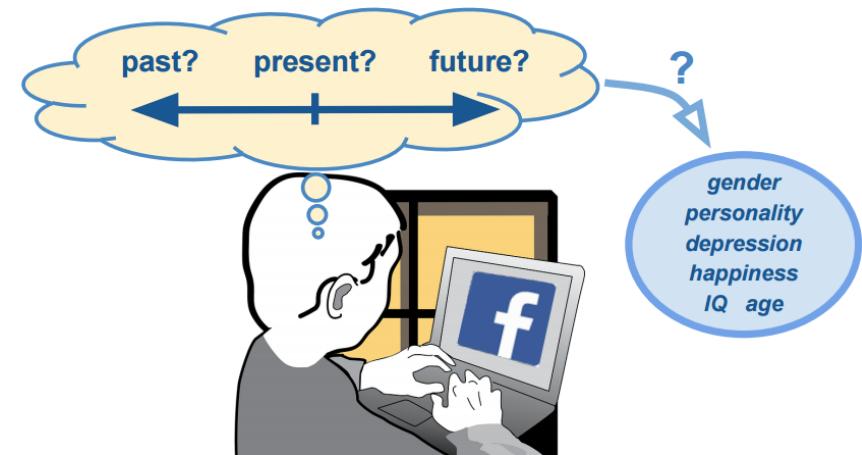
Victim	Date
spamhaus	2013/03/18
soca	2011/06/20
etrade	2012/01/05
interpol	2012/02/29
ustream	2012/05/09

[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

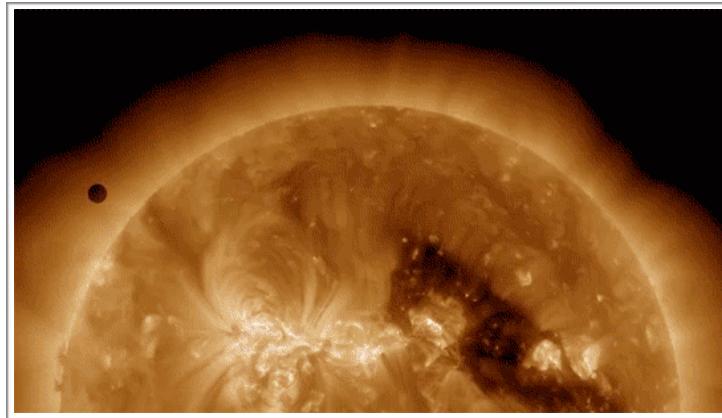
(no human labels needed!)

Distant Supervision Assumption



Mercury Transit May 9, 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

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!

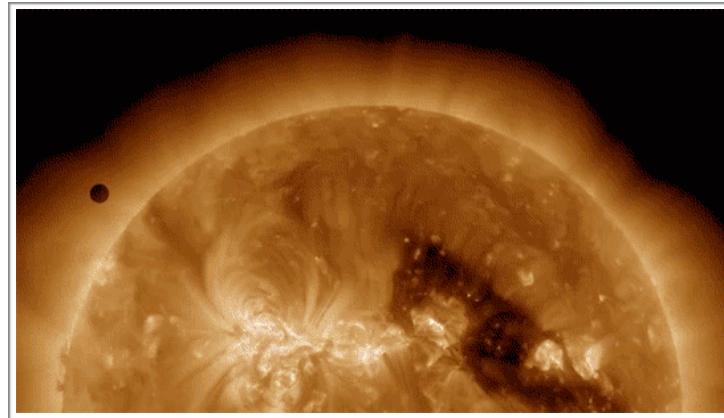
3:17 PM - 9 May 2016



[Follow](#)

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

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

3:54 PM - 9 May 2016



Follow

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



Follow

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



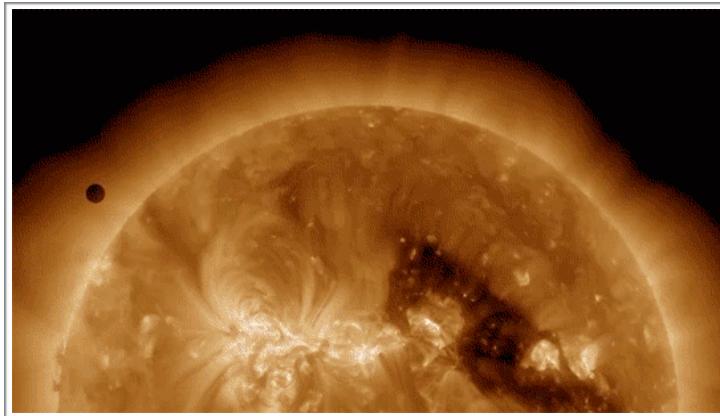
Follow

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



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

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7iaxf4rfC

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

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

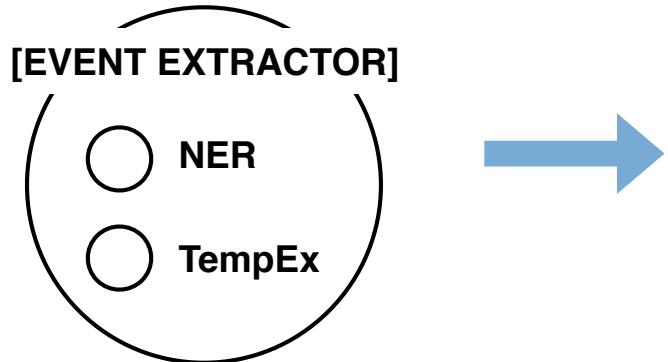
Distant Supervision Assumption

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



[10,000 events]
[EVENT DATABASE]

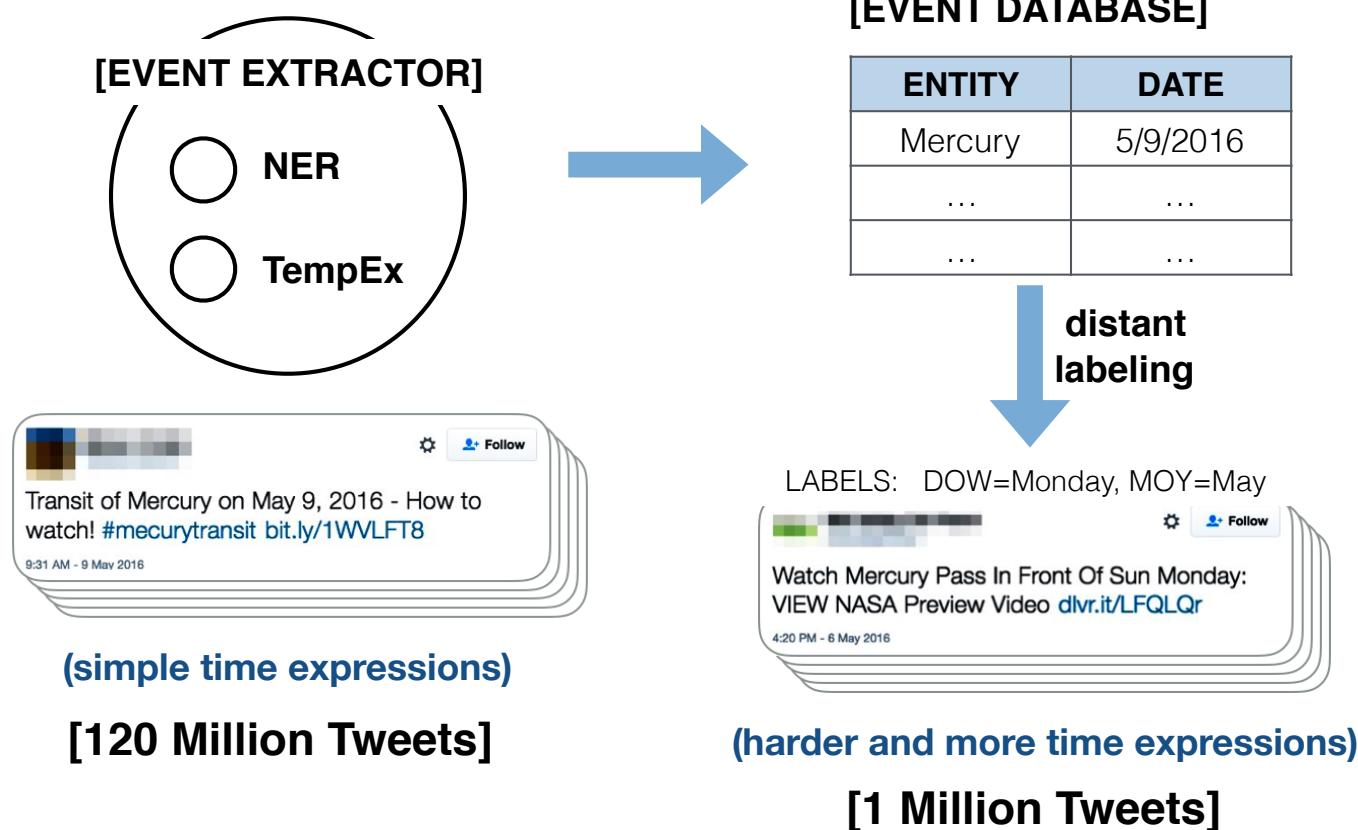
ENTITY	DATE
Mercury	5/9/2016
...	...
...	...



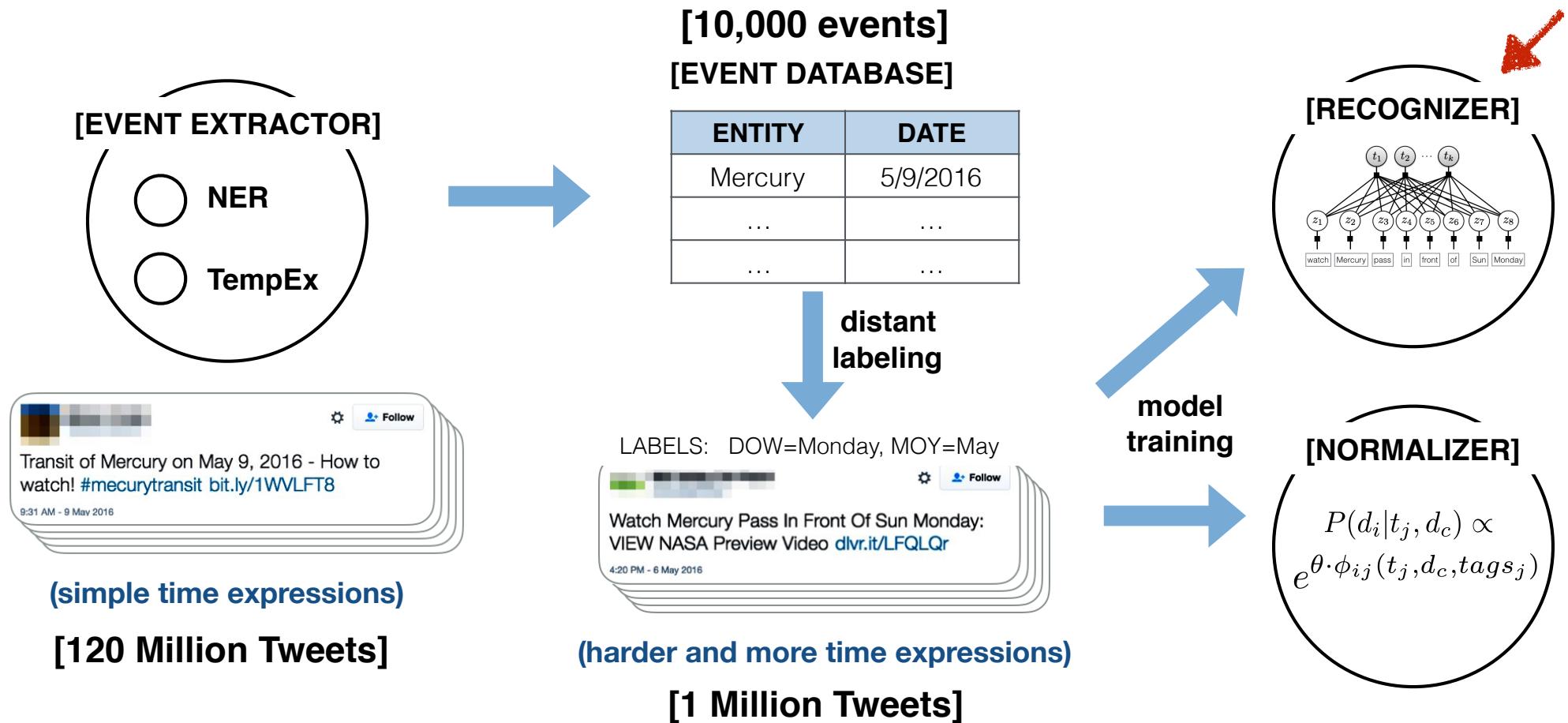
(simple time expressions)

[120 Million Tweets]

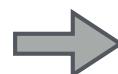
System Overview



System Overview



Temporal Recognizer



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

< Monday, 9, May, Future >

[Event Database]

ENTITY	DATE
Mercury	5/9/2016
...	...



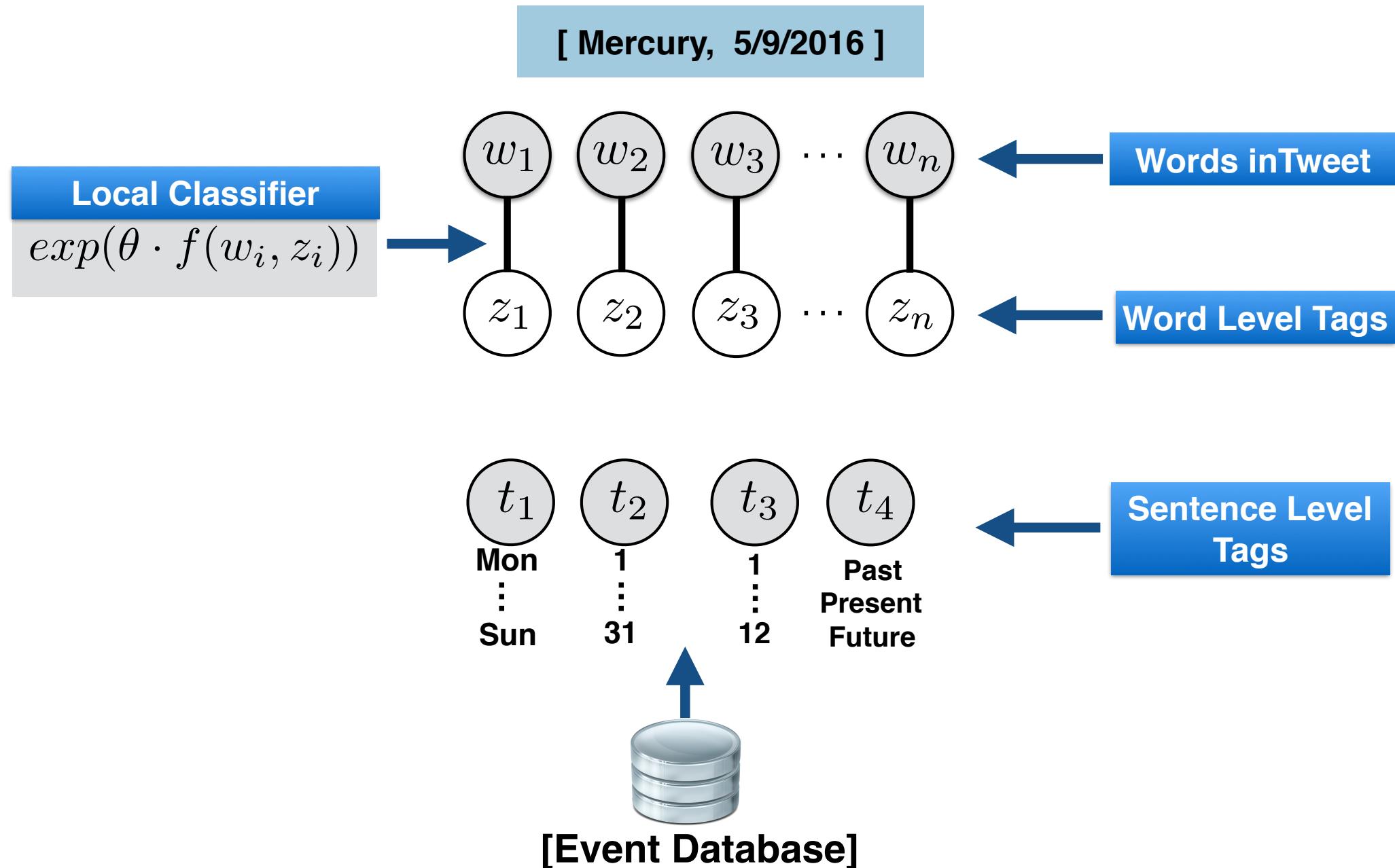
[1 Million Tweets]

[automatically labeled]

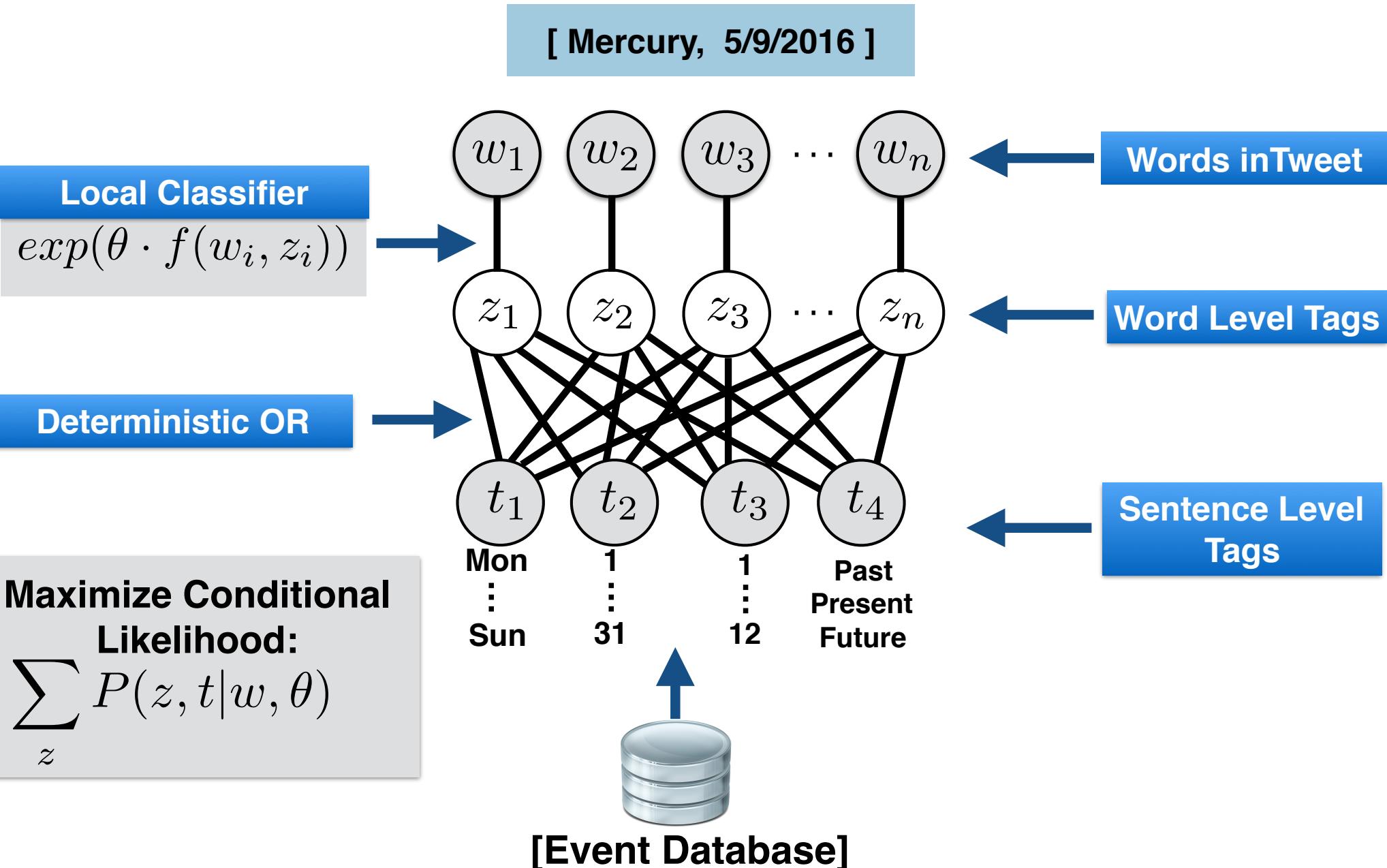


Word	Happy	Friday!	#Fri	yeay	#TGIF
Tag	NA	FRI	FRI	NA	FRI

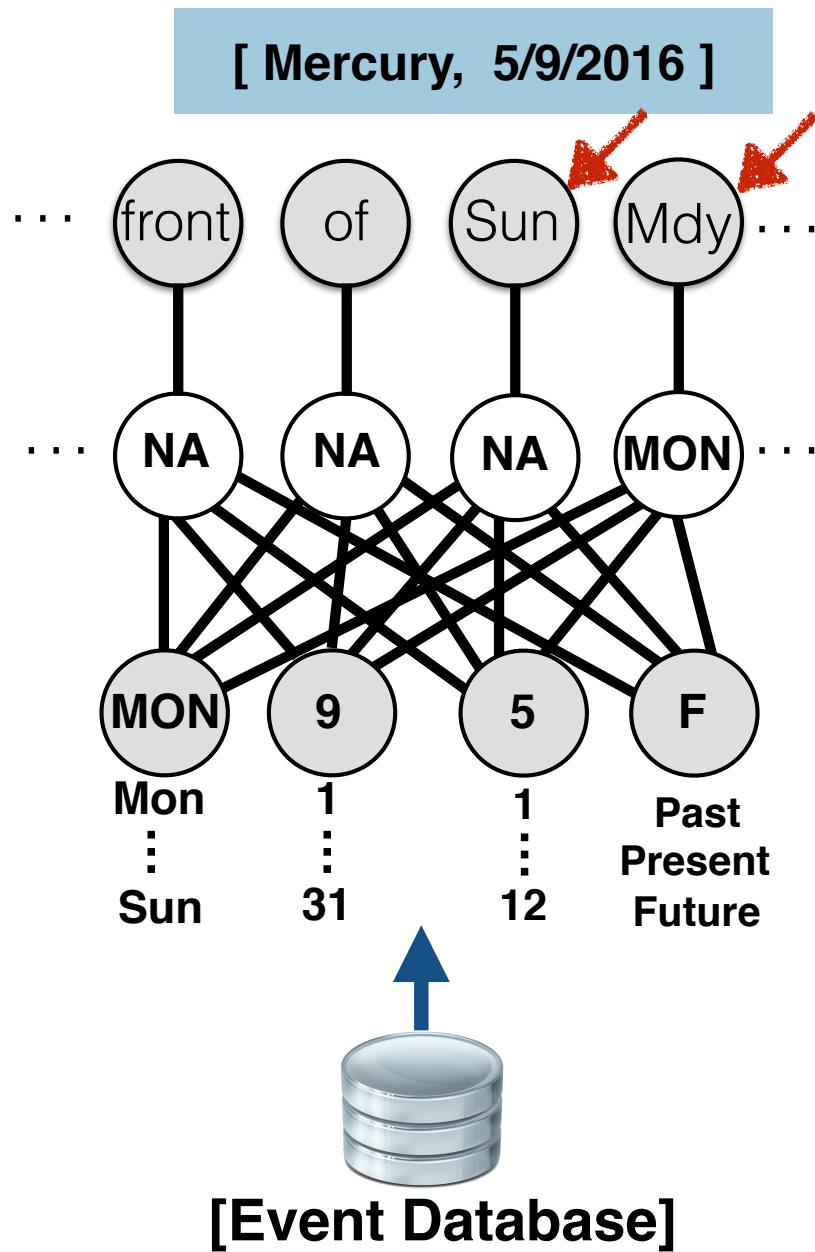
MultiT Tagger



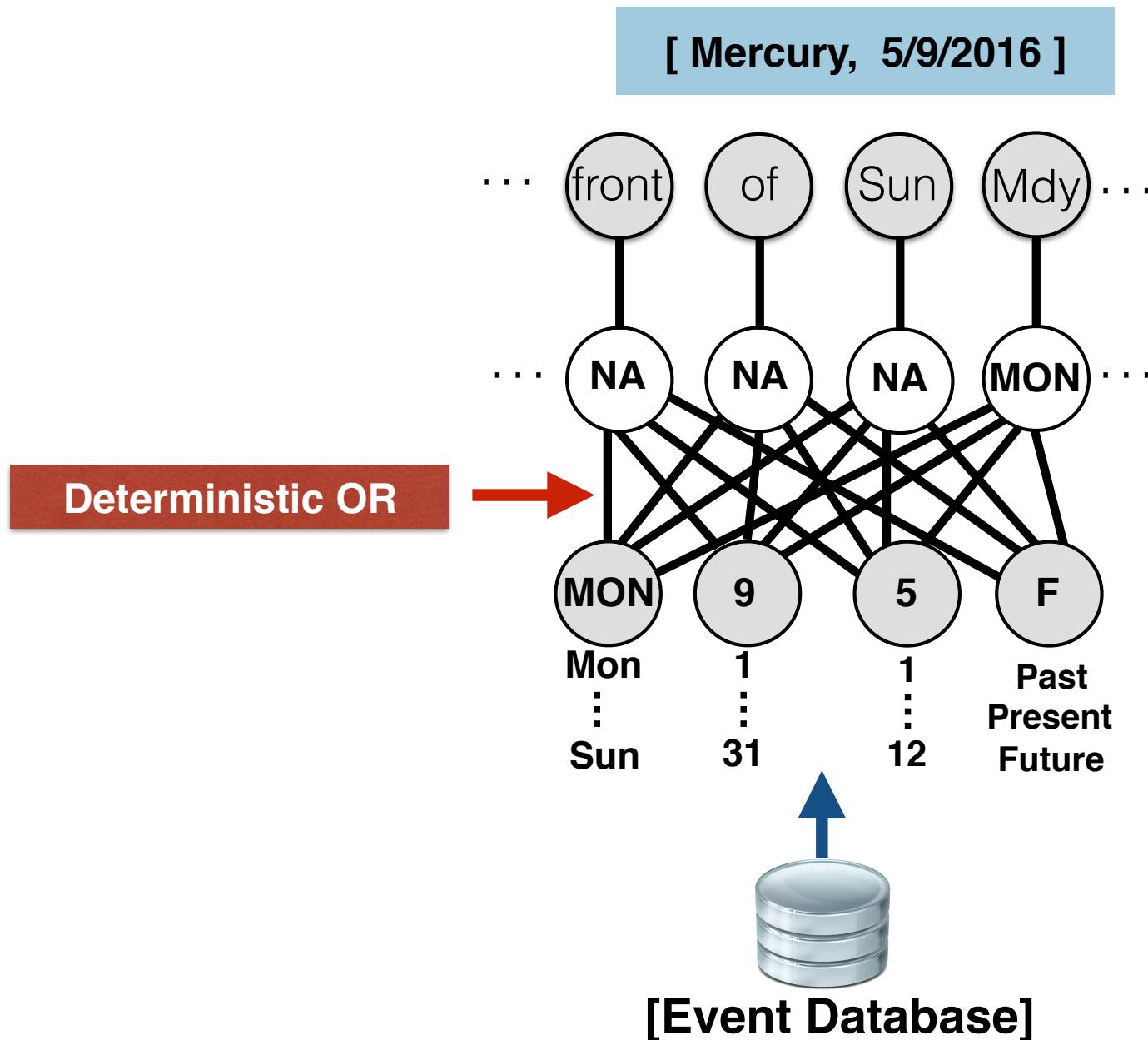
MultiT Tagger



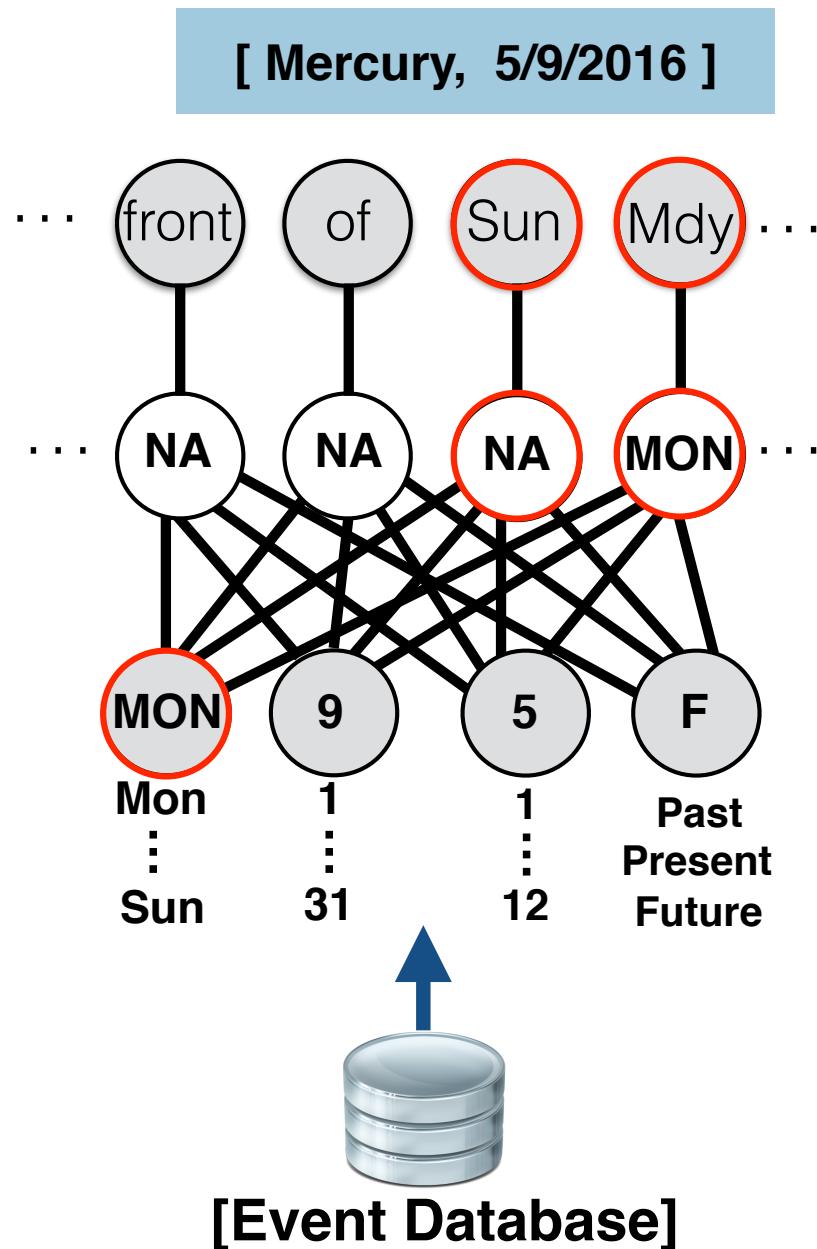
MultiT Tagger



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:

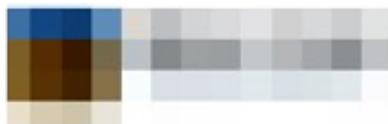
$$\nabla P(t|w) = \sum_z P(z|w, t; \theta) \cdot f(z, w) - \sum_{t,z} P(t, z|w; \theta) \cdot f(z, w)$$

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

Assumptions: MultiT Tagger

Sentence Level Tags \in *Word Level Tags*

Assumptions: MultiT Tagger



Follow

TOMORROW: Mercury will cross the face of the Sun. Watch from 9am-2pm Monday, May 9 at the Planetarium. FREE. ow.ly/WNpbC

11:30 AM - 8 May 2016

Sentence Level Tags:

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

Sentence Level Tags ∈ Word Level Tags

Assumptions: MultiT Tagger



Sentence Level Tags:

TL = Future ←
MOY= May
DOM=9
DOW= Mon

Sentence Level Tags ∈ *Word Level Tags*
Word Level Tags ∈ *Sentence Level Tags*

Missing Data Problem



Follow

Watch Mercury Pass In Front Of Sun Monday:
VIEW NASA Preview Video dlvr.it/LFQLQr

4:20 PM - 6 May 2016

Sentence Level Tags:

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

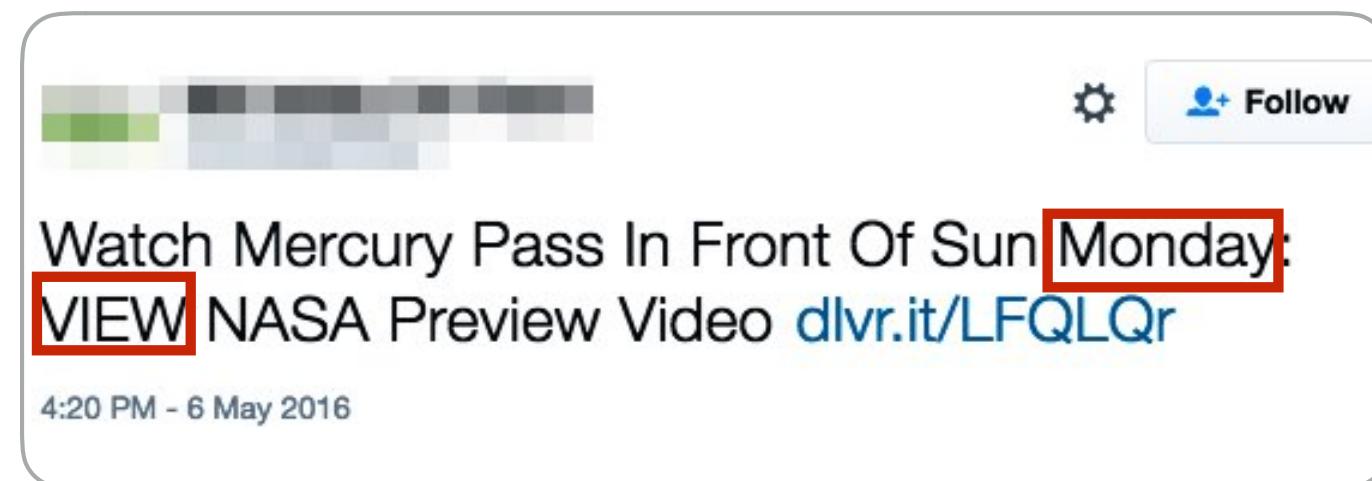
Missing Data Problem



Sentence Level Tags:

- ✓ TL = Future
- ✗ MOY= May
- ✗ DOM=9
- ✓ DOW= Mon

Missing Data Problem



Sentence Level Tags:

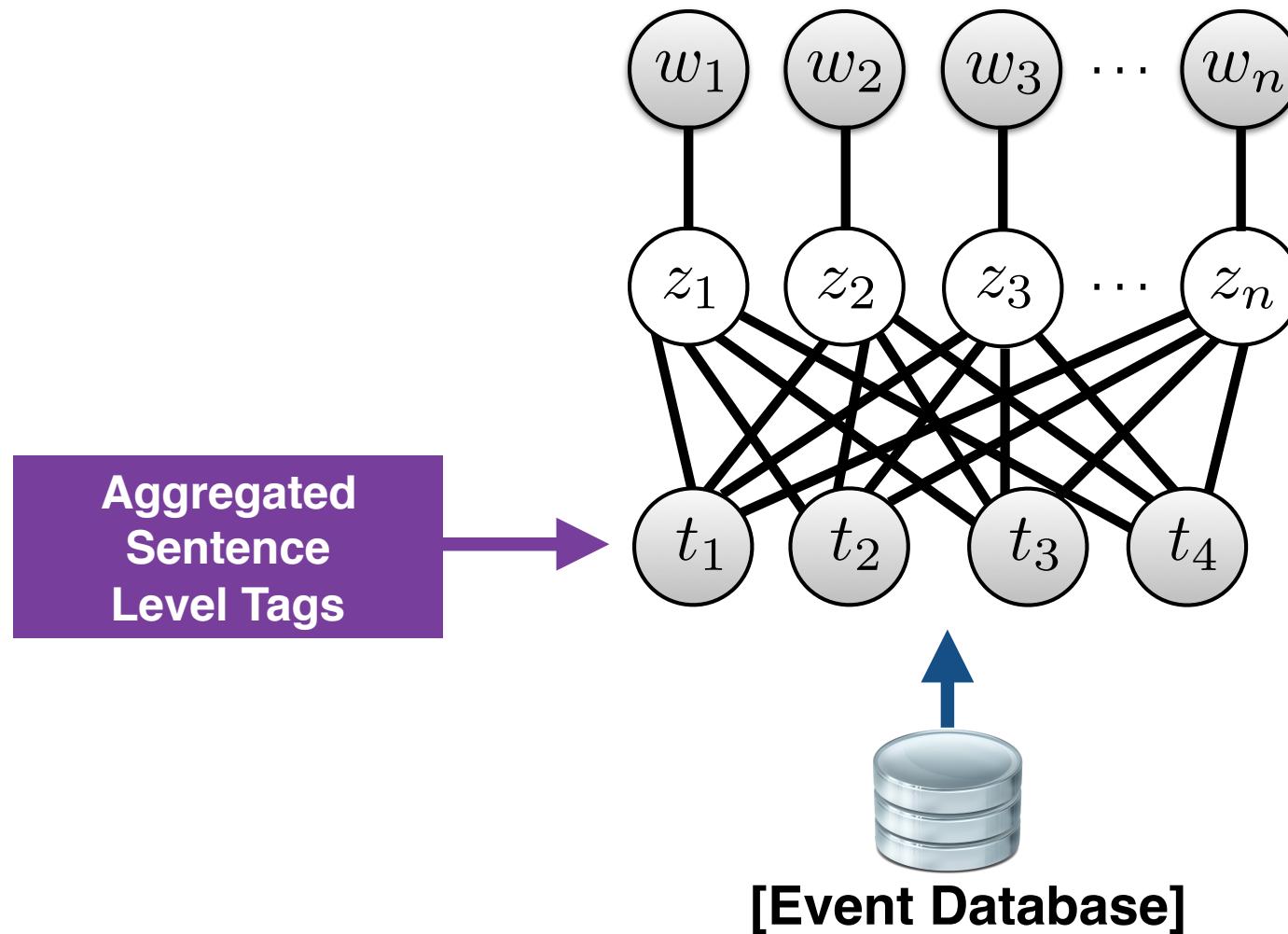
- ✓ TL = Future
- ✗ MOY= May
- ✗ DOM=9
- ✓ DOW= Mon

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

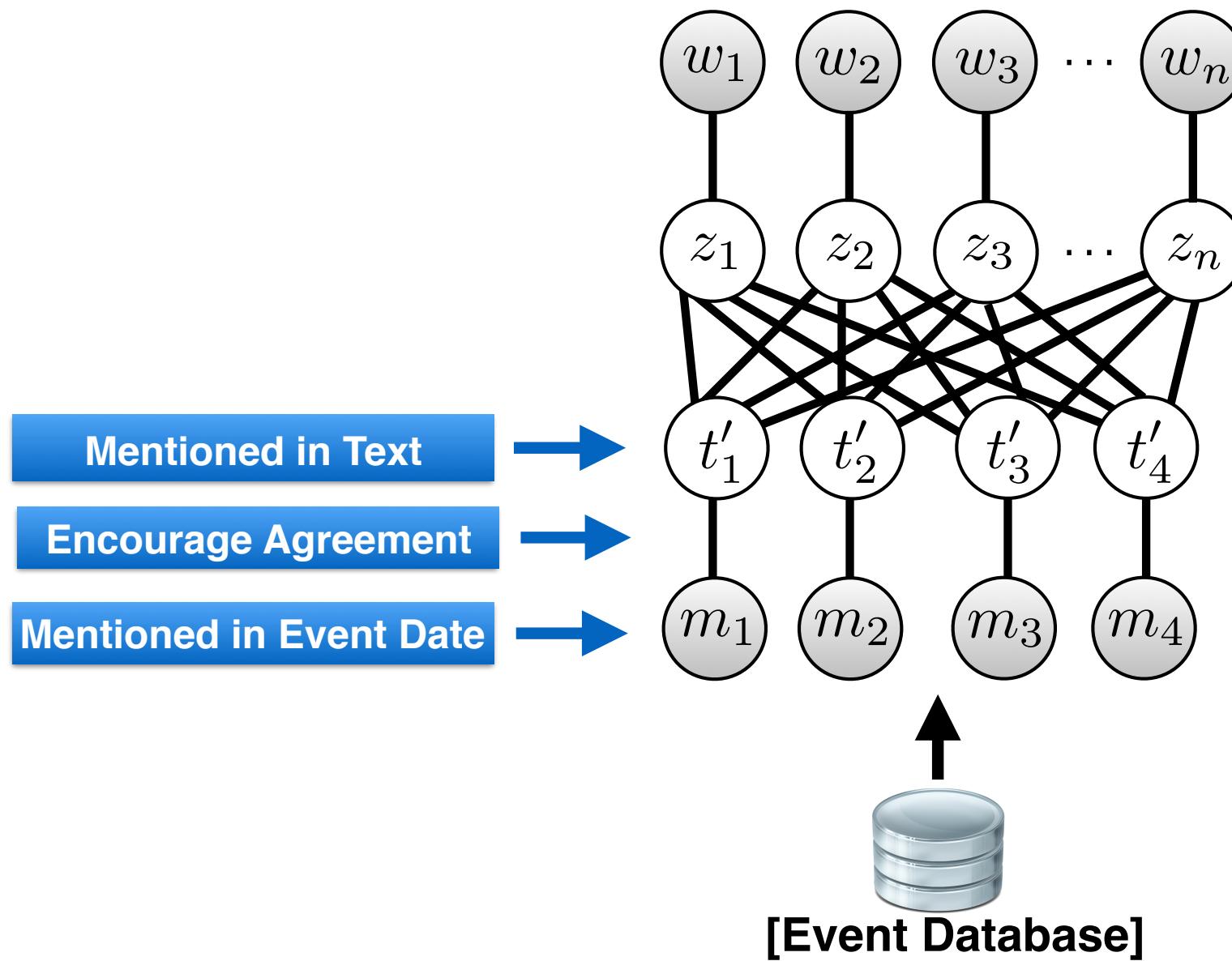
Solution:

Multiple Instance Learning Missing Data Model

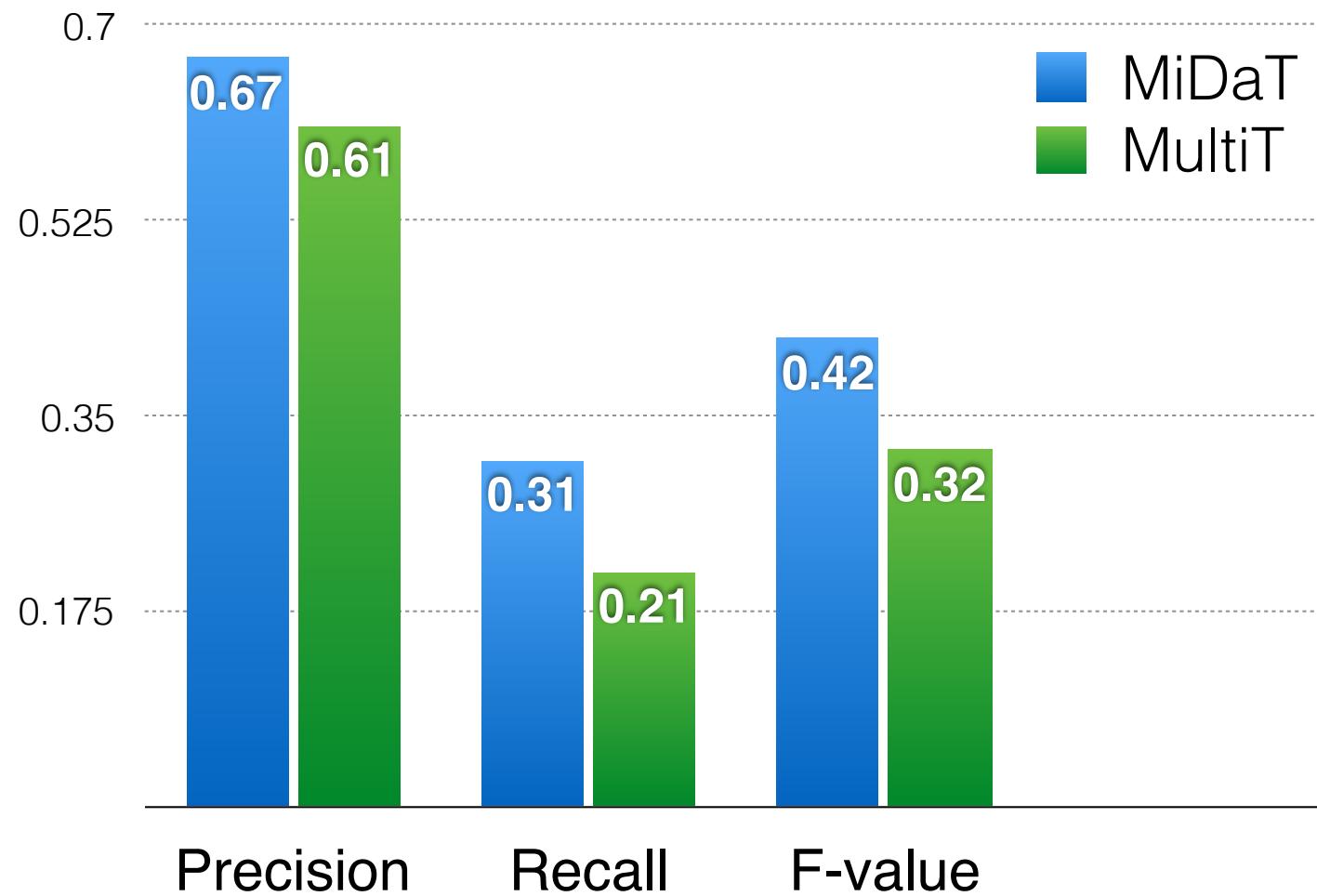
Missing Data Extension



MiDaT Tagger [Extension of MultiT]



MultiT Vs MiDaT [Intrinsic evaluation]

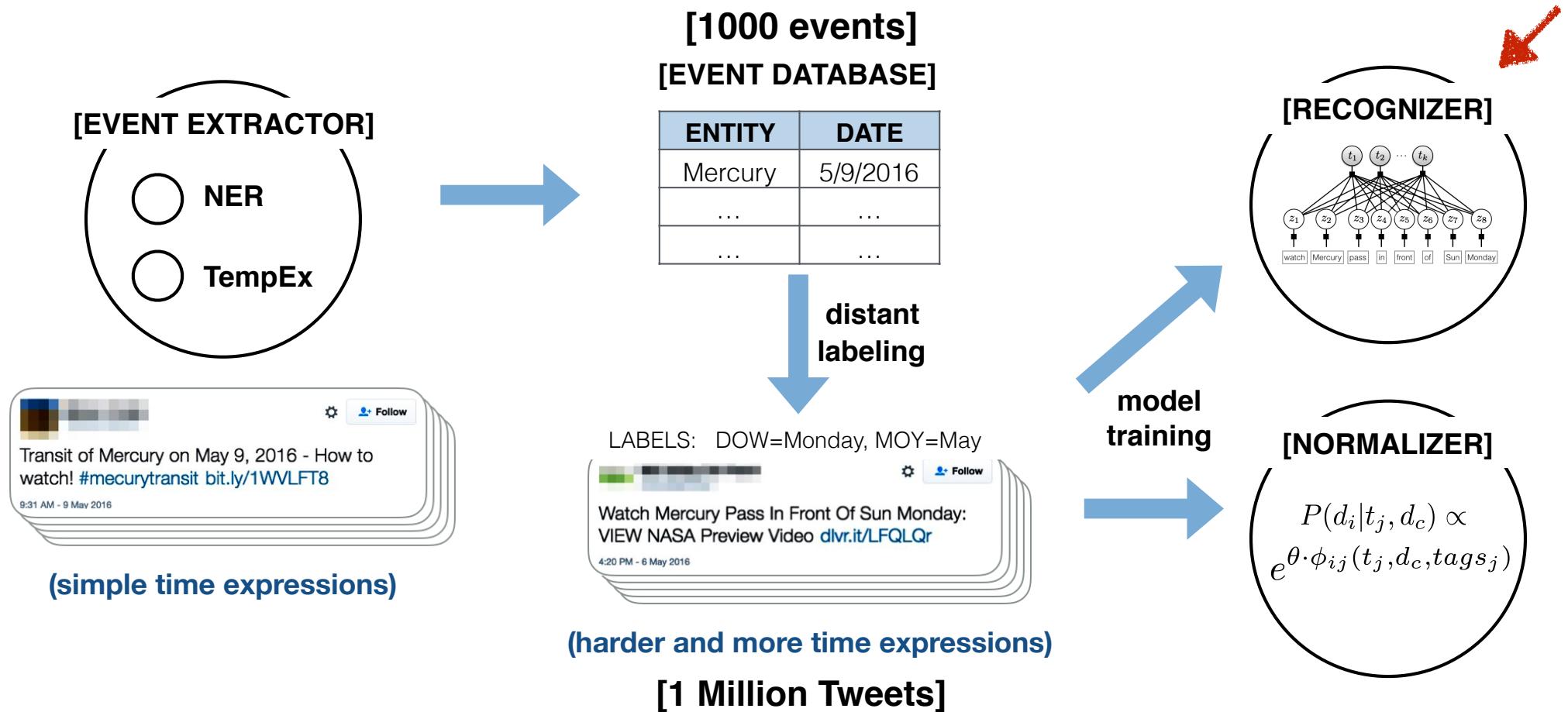


Example Tags

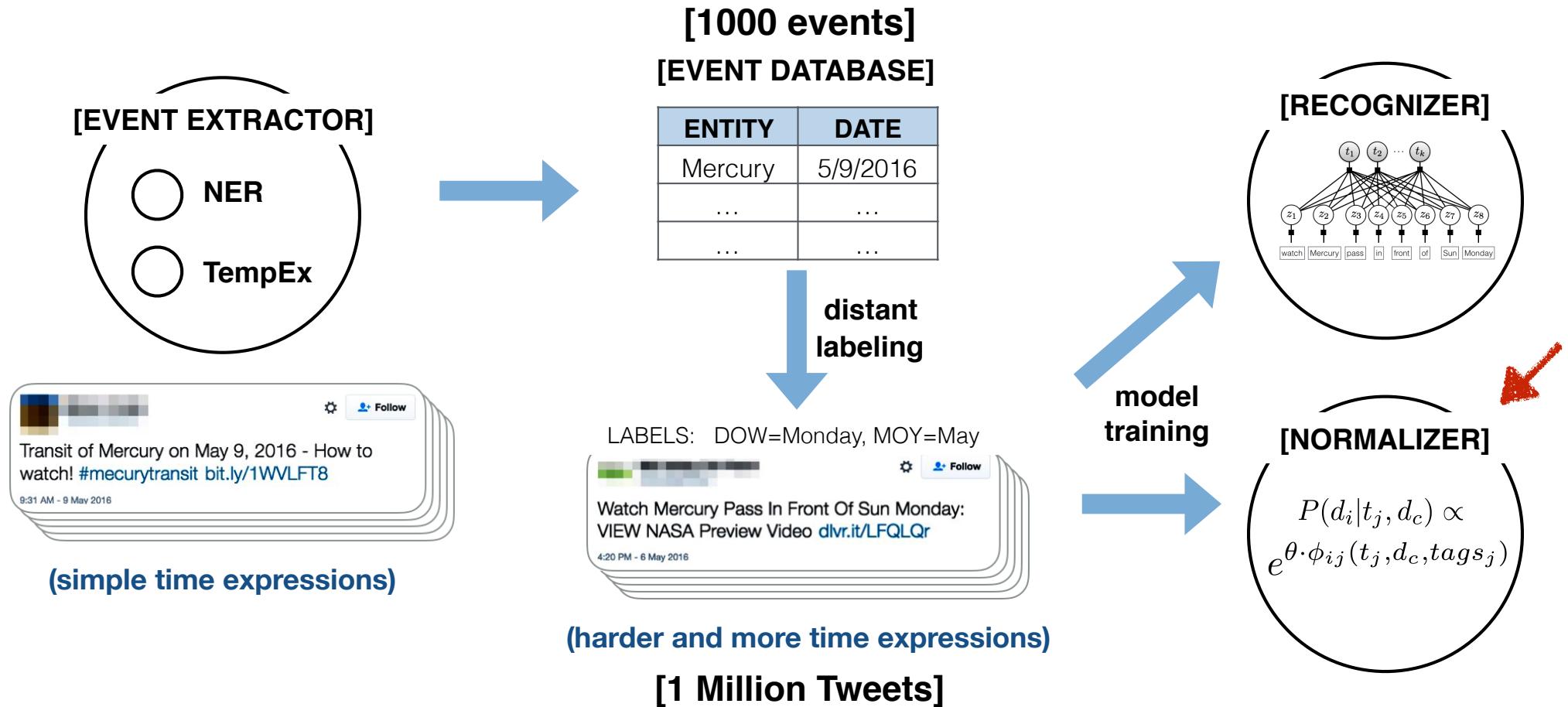
Word	Im	Hella	excited	for	tomorrow
Tag	NA	NA	Future	NA	Future
Word	Thnks	for	a	Christmas	party
Tag	NA	NA	NA	Dec	NA
Word					on fri
Tag					FRI

[Not always matches human intuition]

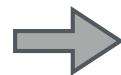
System Overview



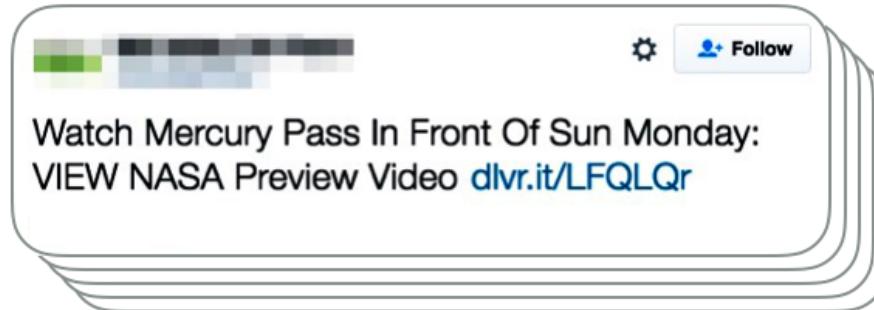
System Overview



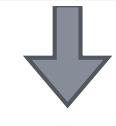
Temporal Normalizer



LABEL: May 9 2016



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



[NORMALIZER]

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

Temporal Normalizer



LABEL: May 9 2016



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

[NORMALIZER]

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

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

Temporal Normalizer



→ LABEL: May 9 2016

[Event Database]



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

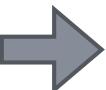
[NORMALIZER]

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

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

Creation Date: 11/1

EMNLP starts 4m tmrw!



D_{-10}

10/22

...

D_{-1}

10/31

D_0

11/1

D_{+1}

11/2

...

D_{+10}

11/11

None

NA

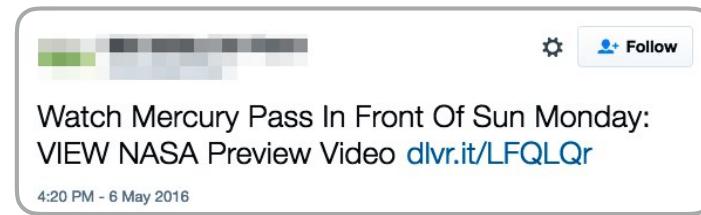
Temporal Normalizer: Features

$$P \left(\text{Candidate Date: } 5/9/2016 \mid \dots \right) \propto f$$

Lexical Features

Tag Features

Time Difference Features



Word Tags predicted by recognizer:
TL = Future
DOW= Mon
:

Creation Date: 5/6/2016

Candidate Date: 5/9/2016

• θ

Temporal Normalizer: Features

$$P \left(\text{Candidate Date: } 5/9/2016 \mid \dots \right) \propto f$$

Lexical Features

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



Word Tags predicted by recognizer:
TL = Future
DOW= Mon
⋮

Creation Date: 5/6/2016

Candidate Date: 5/9/2016

$\cdot \theta$

Temporal Normalizer: Features

$$P(\text{Candidate Date: } 5/9/2016 \mid \dots) \propto f$$

Tag Features



Word Tags predicted by recognizer:
TL = Future
DOW= Mon
⋮

Creation Date: 5/6/2016

Candidate Date: 5/9/2016

• θ

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

Temporal Normalizer: Features

$$P \left(\text{Candidate Date: } 5/9/2016 \mid \dots \right) \propto f$$

Time Difference Features



Word Tags predicted by recognizer:
TL = Future
DOW= Mon
⋮

Creation Date: 5/6/2016

Candidate Date: 5/9/2016

$$DIFF_{days} = +3$$

$$DIFF_{weeks} = +1$$

$$\cdot \theta$$

Evaluation

250 Tweet [2014-2015]

dev : 50
test : 200

Manually
Annotated



Temporal Resolver

[TweeTime, HeidelbergTime, UWTime, SUTime, TempEX]

Evaluation

250 Tweet [2014-2015]

dev : 50
test : 200

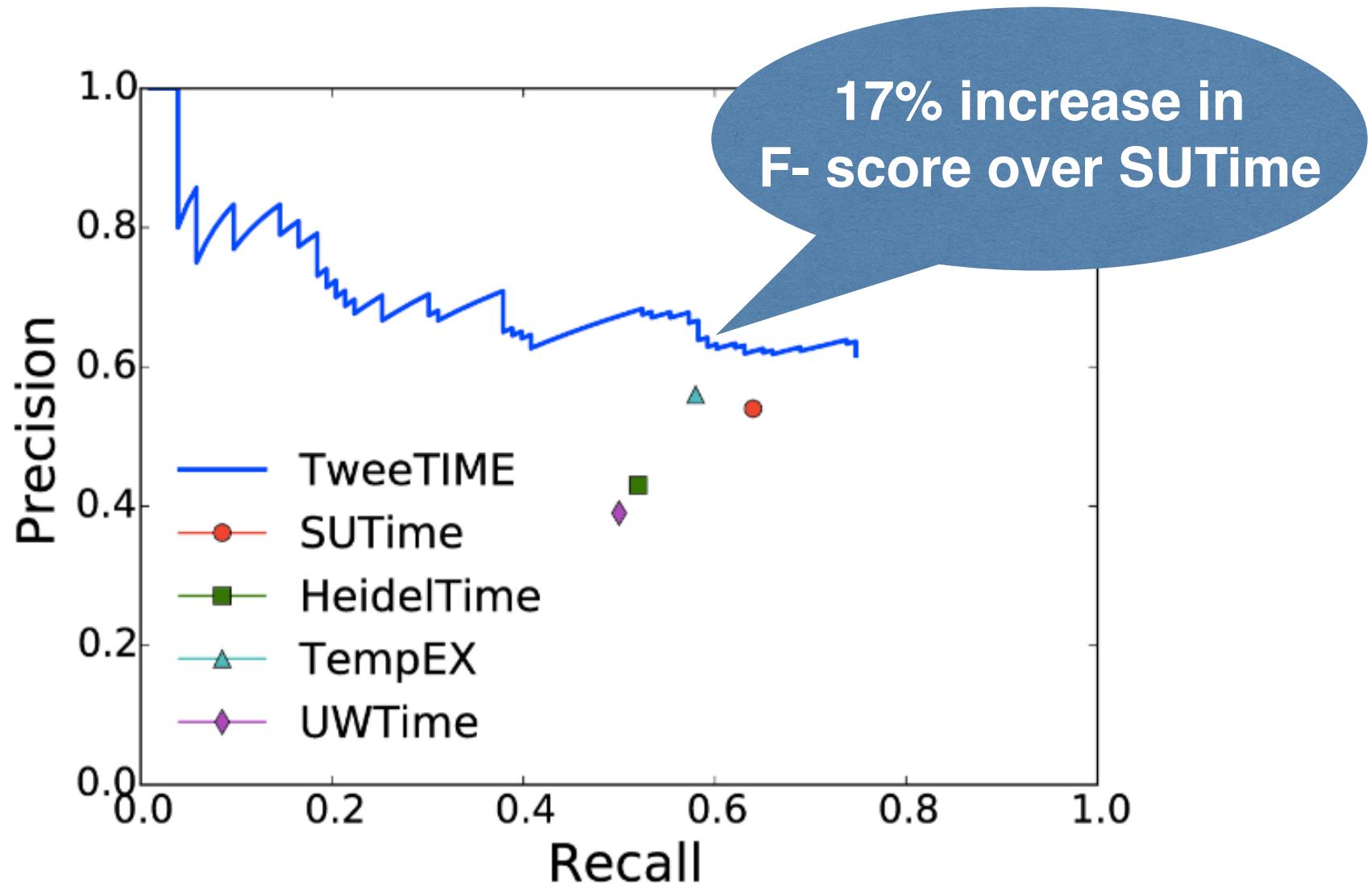
Manually
Annotated

Temporal Resolver

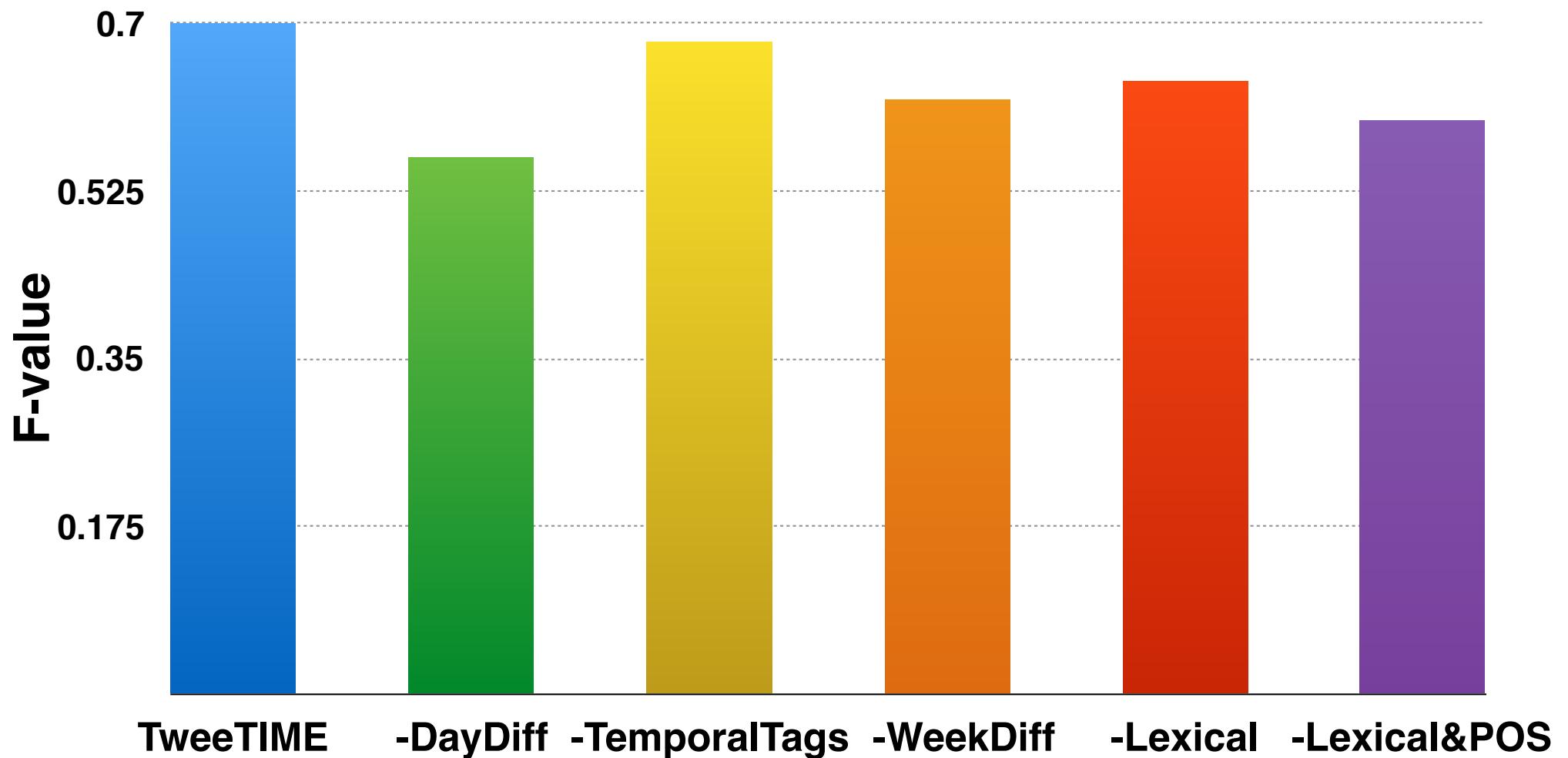
[**TweeTime**, HeidelTime, UWTime, SUTime, TempEX]

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>

