# Weakly Supervised Extraction of Computer Security Events from Twitter

**Alan Ritter**, Evan Wright, William Casey, Tom Mitchell

The Ohio State University

Carnegie Mellon

# Natural Language Processing on Newswire

- Lots of previous work:
  - MUC & ACE competitions
  - Timebank
  - Penn Treebank
  - Topic Detection and Tracking
  - Etc…

# Microblogs vs. Newswire

- **Opportunities:**
  - Fresher Information
  - Broader Coverage
- **Challenges:**
  - Irrelevant and redundant messages
  - Unreliable information

# Microblogs vs. Newswire

- **Opportunities:**
  - Fresher Information
  - Broader Coverage
- **Challenges:**
  - Irrelevant and redundant messages
  - Unreliable information

**Information Overload**

# 1 May 2011

3:58 PM EST

**Sohaib Athar**
@ReallyVirtual

Follow

Helicopter hovering above Abbottabad at 1AM (is a rare event).

10:24 PM EST

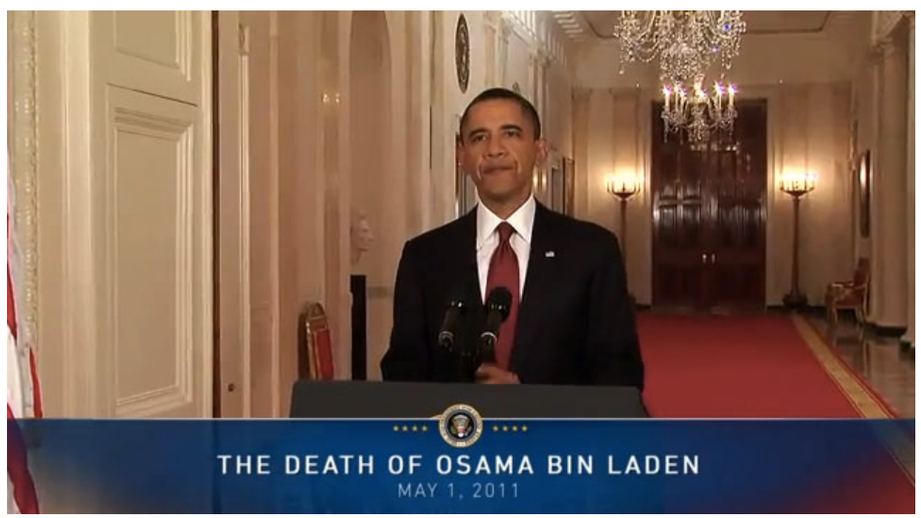
**Keith Urbahn**
@keithurbahn

Follow

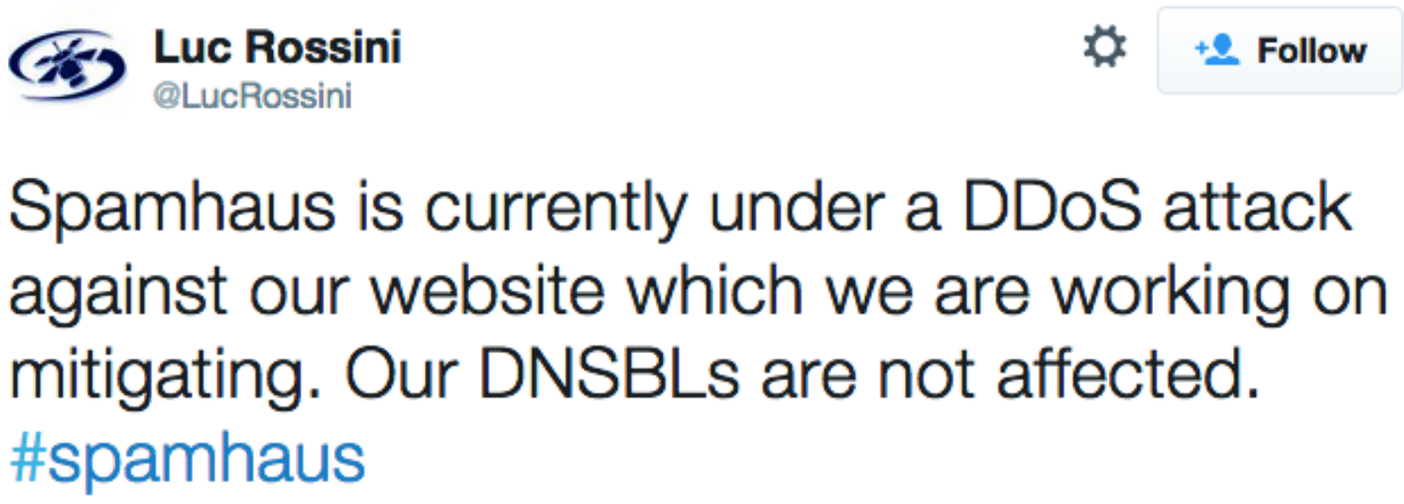
So I'm told by a reputable person they have killed Osama Bin Laden. Hot damn.

11:35 PM EST

Presidential Address Broadcast



# Long tail of Smaller Events



# Information Overload

# Information Overload

Information Extraction

7/4/2014 **Phishing Attack**

Victim: Bitcoins Reserve

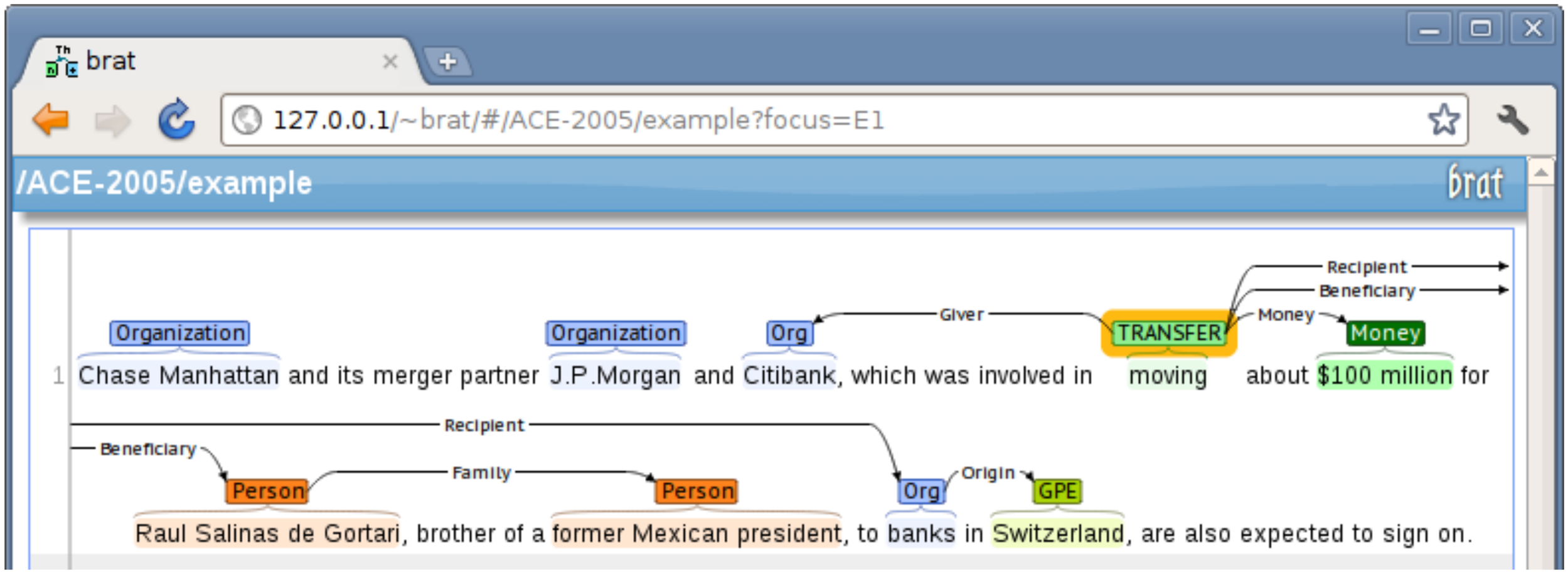
4/25/2015 **Account Hijacking**

Victim: Tesla

5/16/2015 **DDOS**

Victim: PSN

# Traditional Information Extraction



# Weakly Supervised Learning

Distant Supervision

**Structured Data**

**Unlabeled Text**

Information Extraction

# Relation Extraction

## e.g. [Mintz et. al. 2009]

Freebase

| Person | Birth Location |
|---|---|
| **Barack Obama** | **Honolulu** |
| Mitt Romney | Detroit |
| Albert Einstein | Ulm |
| Nikola Tesla | Smiljan |
| … | … |

**(Albert Einstein, Ulm)**

**(Mitt Romney, Detroit)**

**(Barack Obama, Honolulu)**

"**Barack Obama** was born on August 4, 1961 at … in the city of **Honolulu** …"

**"**Birth notices for **Barack Obama** were published in the **Honolulu** Advertiser…"

**"Born** in **Honolulu**, **Barack Obama** went on to become**…"**

⋮

# Relation Extraction

## e.g. [Mintz et. al. 2009]

Freebase

| Person | Birth Location |
|---|---|
| **Barack Obama** | **Honolulu** |
| Mitt Romney | Detroit |
| Albert Einstein | Ulm |
| Nikola Tesla | Smiljan |
| … | … |

**(Albert Einstein, Ulm)** +

**(Mitt Romney, Detroit)** +

**(Barack Obama, Honolulu)** +

"**Barack Obama** was born on August 4, 1961 at … in the city of **Honolulu** …"

**"**Birth notices for **Barack Obama** were published in the **Honolulu** Advertiser…"

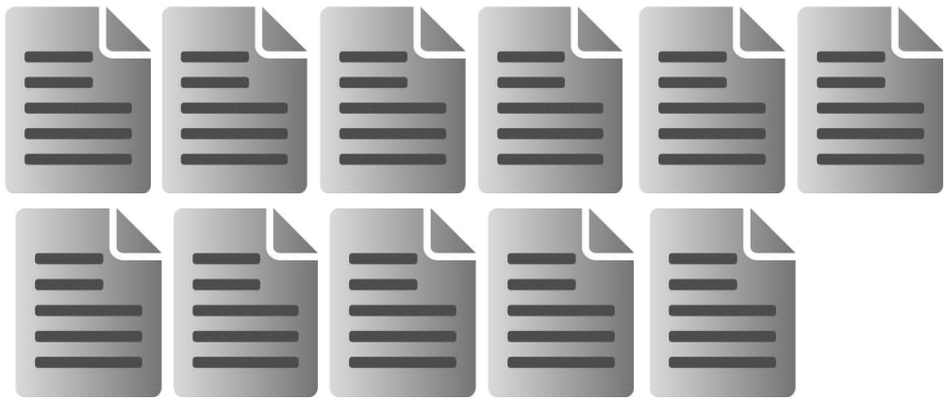
**"Born** in **Honolulu**, **Barack Obama** went on to become**…"**

⋮

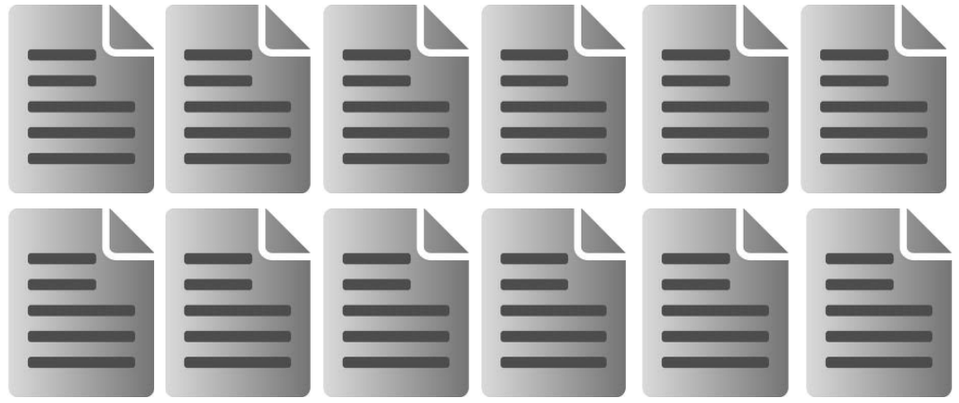
# Distant Supervision: Events vs. Relations

## Relations:

HEADQUARTEREDIN(Microsoft, Redmond)



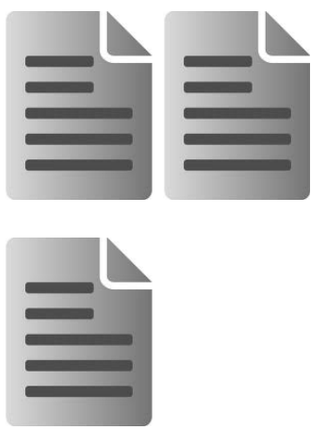
BORNIN(Barack Obama, Honolulu)



## Events:

TERRORISTATTACK(Quetta, Feb. 16 2011)

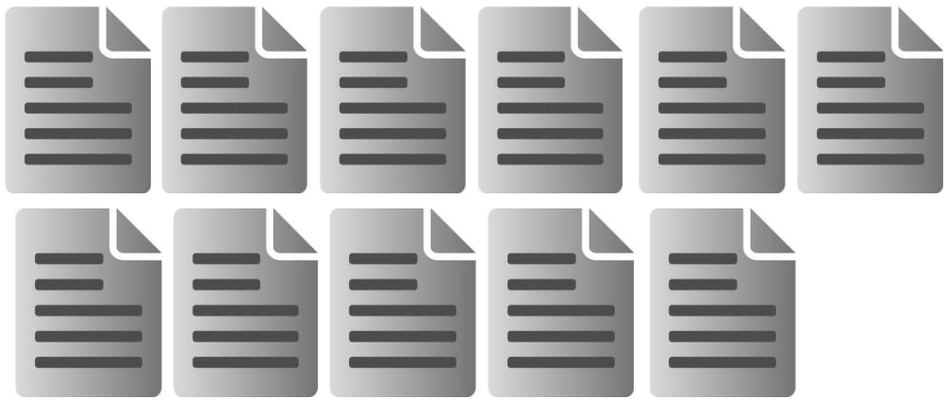
PRODUCTRELEASE(Galaxy S 6, Mar. 22 2015)



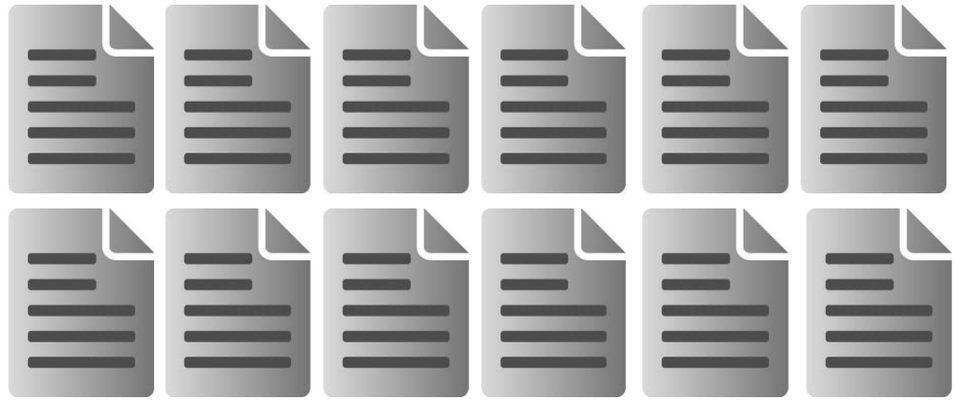
# Distant Supervision: Events vs. Relations

## Relations:

HEADQUARTEREDIN(Microsoft, Redmond)

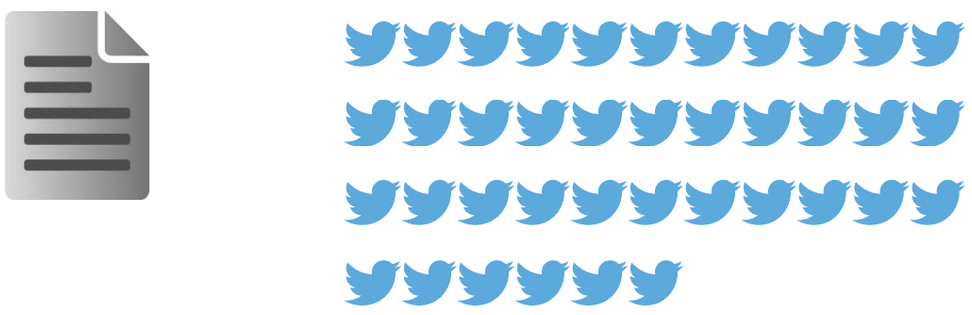


BORNIN(Barack Obama, Honolulu)

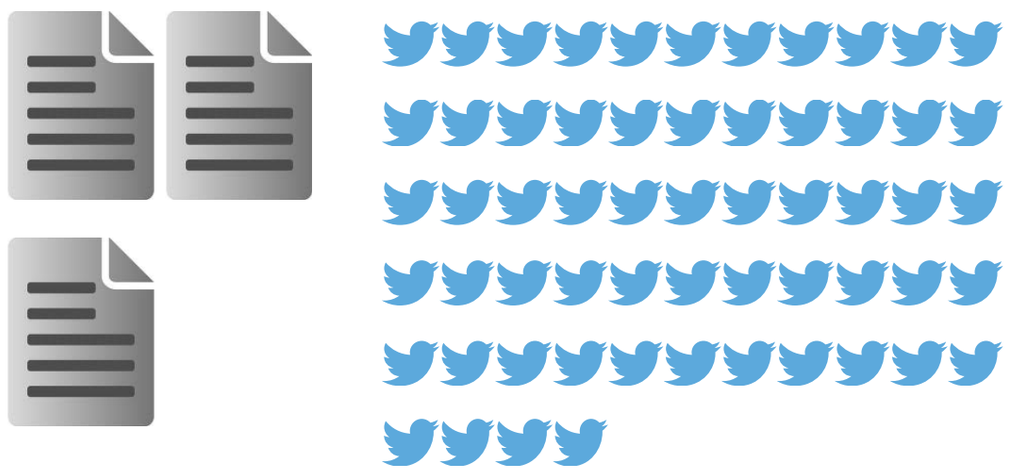


## Events:

TERRORISTATTACK(Quetta, Feb. 16 2011)



PRODUCTRELEASE(Galaxy S 6, Mar. 22 2015)

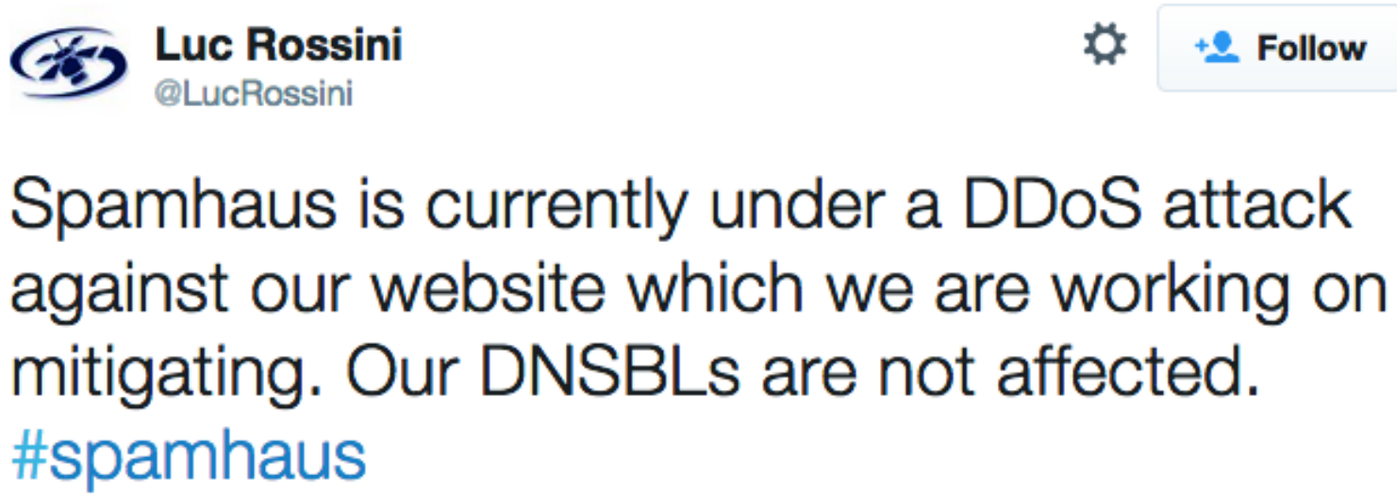


# Cybersecurity Events

**Denial of Service**

Account Hijacking

Data Breach



# Cybersecurity Events

Denial of Service

**Account Hijacking**

Data Breach


# Cybersecurity Events

Denial of Service

Account Hijacking

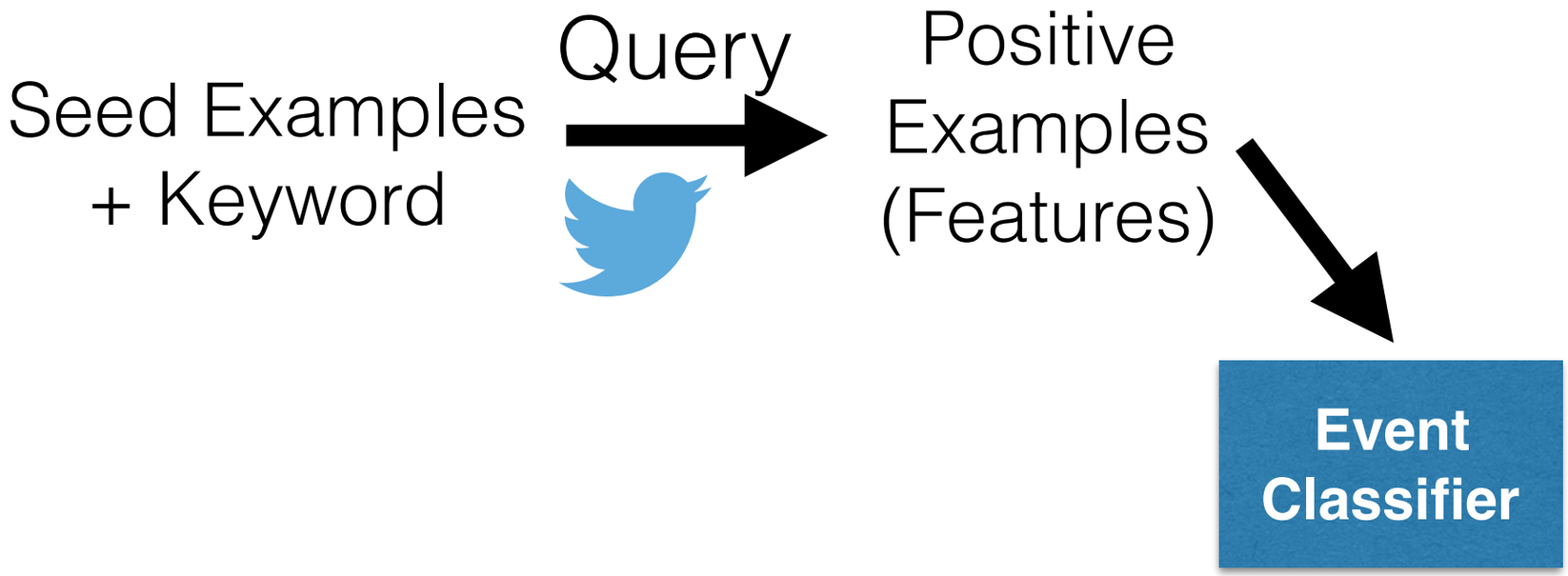
**Data Breach**

# System Overview

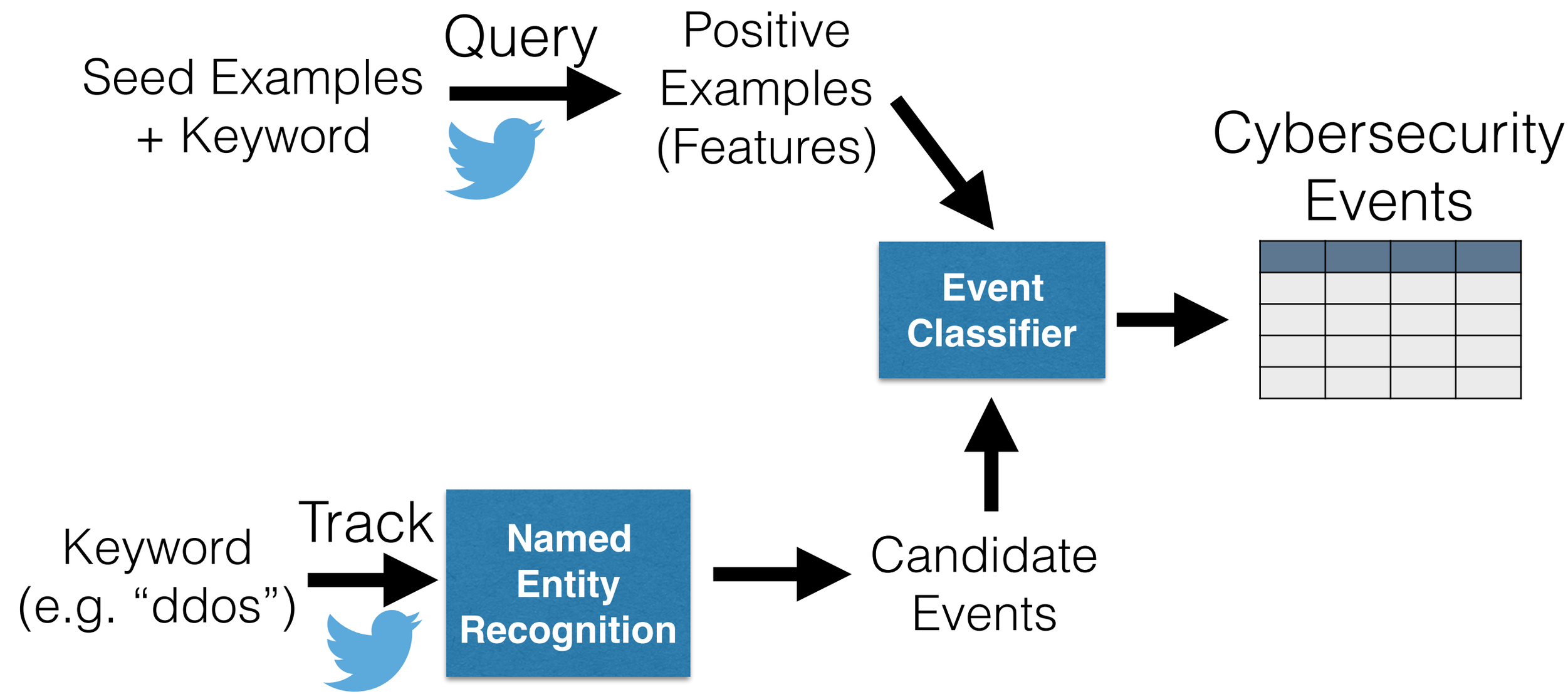
Seed Examples
+ Keyword

**(Associated Press, 4/23/2013)**

# System Overview



# System Overview



# Account Hijacking Seed Events

| Victim | Date |
|---|---|
| associated press | 2013/04/23 |
| reuters | 2012/08/05 |
| us marines | 2013/09/03 |
| sarah palin | 2008/09/18 |
| mitt romney | 2012/06/05 |
| cnn | 2014/01/23 |
| justin bieber | 2012/03/27 |
| mutunga | 2013/09/27 |
| yes scotland | 2013/08/20 |
| zuckerberg | 2013/08/18 |

associated press hacked since:2013-4-23 until:2013-4-24

**Tip:** use operators for advanced search.

Search

**CyG US** @CyG_US · Apr 23
**Associated Press** Twitter account **hacked**, false report of explosions, Obama injury at White House sent out bit.ly/11LLrhk
Expand Reply Retweet Favorite ••• More

**Ambling Freest** @amblingfreest · Apr 23
**Associated Press**' Twitter **hacked**, fake tweet sends stocks plunging - Los Angeles Times: **Associated Press**' Twit... bit.ly/12c1jwf
Expand Reply Retweet Favorite ••• More

**Al Stefanelli** @Stefanelli · Apr 23
**Associated Press** Twitter Accounts **Hacked**, Fake White House Bomb Report Posted wp.me/p3oa9W-AD
Expand Reply Retweet Favorite ••• More

**DudeYouCrazy** @DudeYouCrazy · Apr 23
Wait, so was the **Associated Press** also **hacked** when its final football poll had Notre Dame ranked fourth?
Expand Reply Retweet Favorite ••• More

**BiGxGh.Com** @bigxghdotcom · Apr 23
#MissedThisOnOMGGhana: **Associated Press**' Twitter Account **Hacked**! bit.ly/11LBaS9
Expand Reply Retweet Favorite ••• More

# Features

- Binary features extracted from “bag-of-tweets”
- Context words surrounding the victim of the attack
- Context words surrounding keyword

# Features

- Binary features extracted from “bag-of-tweets”
- Context words surrounding the victim of the attack
- Context words surrounding keyword

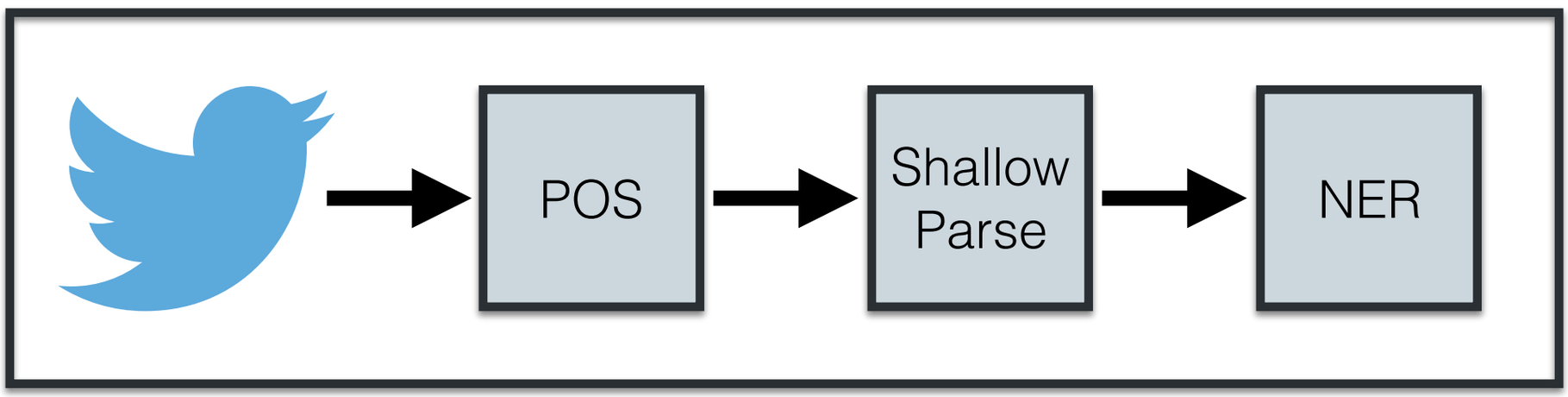
| Sample Feature | Event Category |
|---|---|
| **security breach** | Data Breach |
| **X** admits | Data Breach |
| **X** data breach affecting | Data Breach |
| DT **ddos** attack | DDoS |
| NNP NNP hit IN **ddos** | DDoS |
| **X** getting ddos’d | DDoS |
| PRP hacked **X** POS account | Account Hijacking |
| POS email was **hacked** | Account Hijacking |
| **X** POS NNP account | Account Hijacking |

# Gathering Unlabeled / Test Data

- Track a keyword associated with the event (e.g. **hacked, DDOS, breach**)

- Extract named entities (using Twitter-tuned NLP tools)

**[Ritter, et. al. EMNLP 2011a]**



# Not account hijacking…

**Gavin Caney** @GavinCaney · Mar 28
PENALTY to Lynn as @JakeJonesx11 is **hacked** down by Hutchinson. 1-0 Blyth but for how long? #EDPsport

# Not a Data Breach…

# Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

- [Mann and McCallum]

$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

# KL Divergence

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p}\log\frac{\tilde{p}}{\hat{p}_\theta} + (1-\tilde{p})\log\frac{1-\tilde{p}}{1-\hat{p}_\theta}$$

# KL Divergence

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p}\log\frac{\tilde{p}}{\hat{p}_\theta} + (1-\tilde{p})\log\frac{1-\tilde{p}}{1-\hat{p}_\theta}$$

User-provided target expectation

# KL Divergence

Empirical expectation on unlabeled examples

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p}\log\frac{\tilde{p}}{\hat{p}_\theta} + (1-\tilde{p})\log\frac{1-\tilde{p}}{1-\hat{p}_\theta}$$

User-provided target expectation

# KL-Div Gradient

$$\frac{\partial}{\partial \theta_k} D(\tilde{p}||\hat{p}_\theta) =$$

$$\frac{1}{N}\left(\frac{1-\tilde{p}}{1-\hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta}\right) \sum_{i=1}^{N} p_\theta(y_i = 1|x_i)(1 - p_\theta(y_i = 1|x_i))x_{i,k}$$

# KL-Div Gradient

$$\frac{\partial}{\partial \theta_k} D(\tilde{p}||\hat{p}_\theta) =$$

$$\frac{1}{N}\left(\frac{1-\tilde{p}}{1-\hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta}\right) \sum_{i=1}^{N} p_\theta(y_i = 1|x_i)(1 - p_\theta(y_i = 1|x_i))x_{i,k}$$

No Change if $\tilde{p} = \hat{p}_\theta$

Otherwise push weights up or down

# KL-Div Gradient

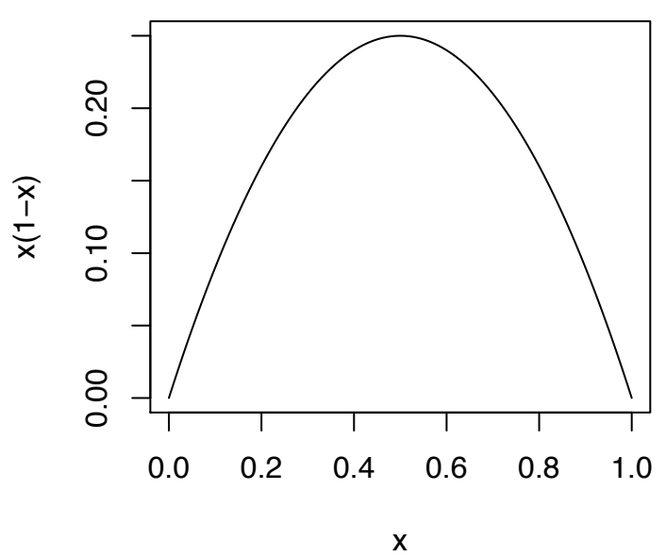
$$\frac{\partial}{\partial \theta_k} D(\tilde{p}||\hat{p}_\theta) =$$

Give more weight to uncertain cases

$$\frac{1}{N}\left(\frac{1-\tilde{p}}{1-\hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta}\right) \sum_{i=1}^{N} p_\theta(y_i = 1|x_i)(1 - p_\theta(y_i = 1|x_i))x_{i,k}$$

No Change if $\tilde{p} = \hat{p}_\theta$

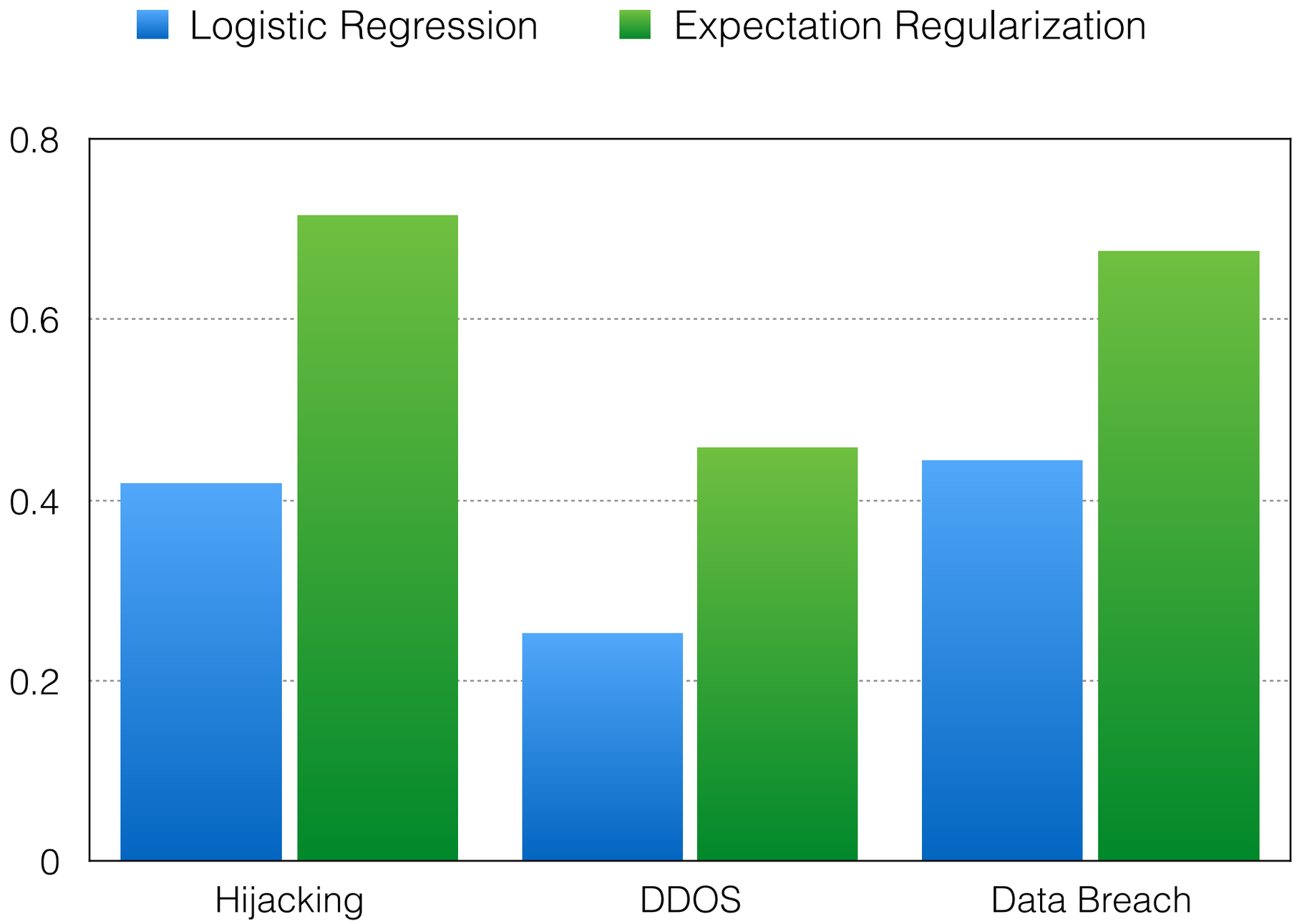
Otherwise push weights up or down



# Baselines

- Logistic Regression with Heuristic Negatives
  - used in lots of previous work
- One-class support vector machines [Schölkopf et. al.]
  - ignore unlabeled events
- Semi-supervised EM

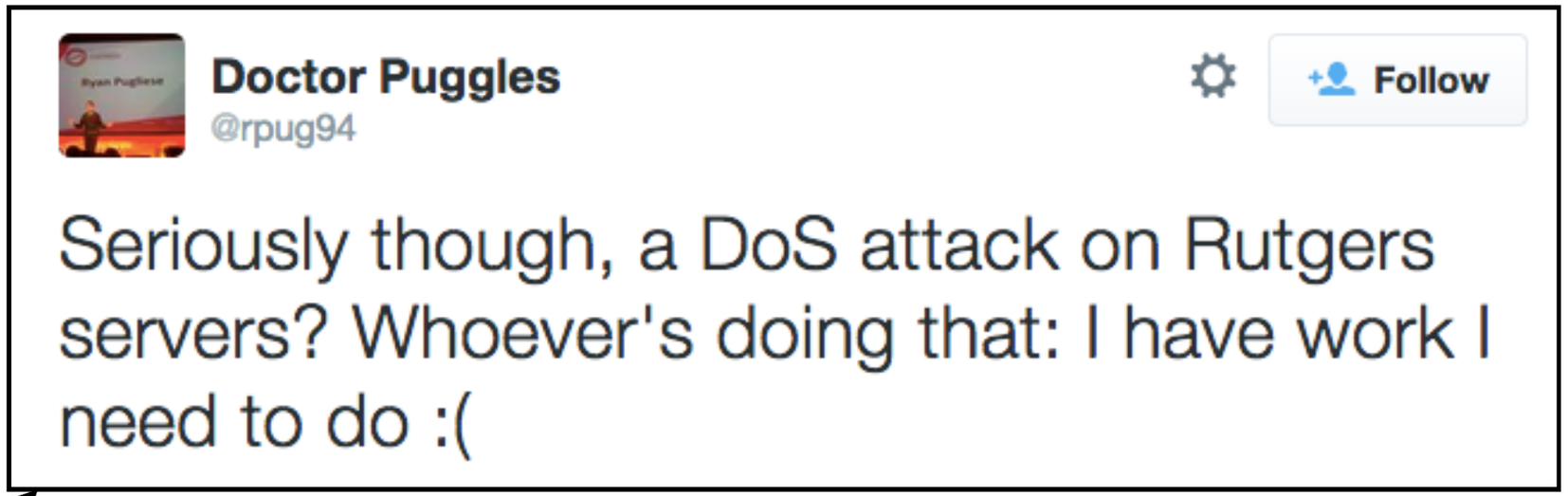
# AUC


Logistic Regression
Expectation Regularization
0.8
0.6
0.4
0.2
0
Hijacking
DDOS
Data Breach

# DDOS

| Mar-28-2015 | | |
|---|---|---|
| china | Is China responsible for Massive DDoS Attack against GitHub? http://t.co/JDGYInB0Lv | 265 |
| rutgers | RT @MikeBrady35: That moment when your so frustrated with the bus system that you successfully ddos the absolute shit out of Rutgers servers | 10 |
| baidu | RT @jschauma: Some more info on the @GitHub DDoS, apparently by way of Baidu: http://t.co/J9HF2NijiY | 13 |
| github | Is China responsible for Massive DDoS Attack against GitHub? http://t.co/JDGYInB0Lv | 236 |
| ruwireless | RT @francesssss: this kid on reddit saying he's responsible for the DDOS attack on RUwireless. bruh, I'm just trying to do my work #iwillcu… | 3 |
| github ddos | RT This GitHub DDoS is serious. https://t.co/HRmqMiOIba https://t.co/PUzHbFrYWv http://t.co/59Z2O27qcz | 6 |

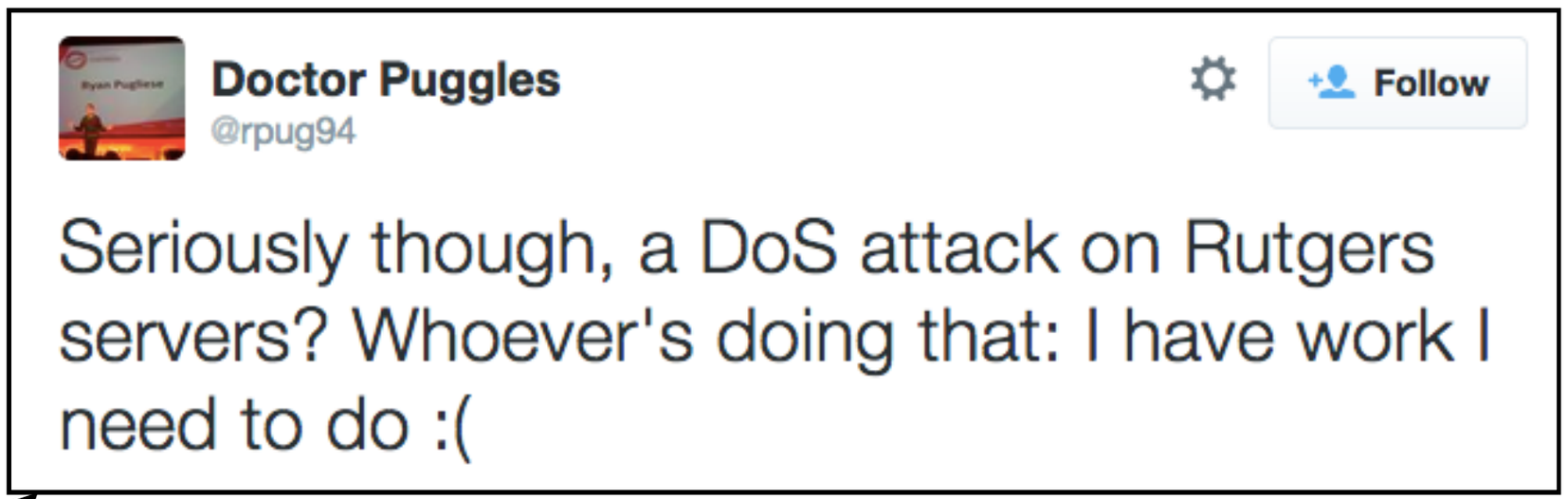
# DDOS




**Mar-28-2015**

| | | |
|---|---|---|
| china | Is China responsible for Massive DDoS Attack against GitHub? http://t.co/JDGYInB0Lv | 265 |
| rutgers | RT @MikeBrady35: That moment when your so frustrated with the bus system that you successfully ddos the absolute shit out of Rutgers servers | 10 |
| baidu | RT @jschauma: Some more info on the @GitHub DDoS, apparently by way of Baidu: http://t.co/J9HF2NijiY | 13 |
| github | Is China responsible for Massive DDoS Attack against GitHub? http://t.co/JDGYInB0Lv | 236 |
| ruwireless | RT @francesssss: this kid on reddit saying he's responsible for the DDOS attack on RUwireless. bruh, I'm just trying to do my work #iwillcu… | 3 |
| github ddos | RT This GitHub DDoS is serious. https://t.co/HRmqMiOlba https://t.co/PUzHbFrYWv http://t.co/59Z2O27qcz | 6 |

# DDOS




Mar-28-2015

| | | |
|---|---|---|
| china | Is China responsible for Massiv | 265 |
| rutgers | RT @MikeBrady35: That Rutgers servers | 10 |
| baidu | RT @is auma: Some more inf | 13 |
| github | Is China responsible for Massive DDoS Attack against GitHub? http://t.co/JDGYInB0Lv | 236 |
| ruwireless | RT @francesssss: this kid on reddit saying he's responsible for the DDOS attack on RUwireless. bruh, I'm just trying to do my work #iwillcu… | 3 |
| github ddos | RT This GitHub DDoS is serious. https://t.co/HRmqMiOlba https://t.co/PUzHbFrYWv http://t.co/59Z2O27qcz | 6 |

# Account Hijacking

**Mar-29-2015**

| | | |
|---|---|---|
| british airways | British Airways says some frequent flyer accounts hacked http://t.co/rsh1qOZ7WH | 210 |
| nigeria | "Nigeria: Election Website Hacked" #newspapers #feedly http://t.co/JngBLetfyn | 37 |
| slacker | Group Chatting Application Slacker Has Been Hacked - http://t.co/whBy5QdmCR > And I always thought slacker was for streaming music. #infosec | 9 |
| uk | RT @PotCoinFan: Hacking is cool if we do it just not you. - UK Government Authorized GHCQ to Hack Any Device https://t.co/D0AbsHG9nB via @H… | 11 |
| inec | RT @OmoOduaRere: INEC's Website get hacked by Nigeria Cyber hacking group on Election day #Nigeriadecides http://t.co/nn8mOhPulp | 61 |
| taylor swift | Taylor Swift Hacked, 6 Nudes for Sale? http://t.co/k6HhFe9nEa | 13 |
| inec website | INEC Website has been Hacked http://t.co/346VTVfVcD | 4 |
| darren | The people who think Darren was hacked are the same people who think he doesn't lie to his fans, wrote This Time for Lea, and likes vag. | 12 |

# Account Hijacking


Mar-29-2015

| | | |
|---|---|---|
| british airways | British Airways says some frequent flyer accounts hacked http://t.co/rsh1qOZ7WH | 210 |
| nigeria | "Nigeria: Election Website Hacked" #newspapers #feedly http://t.co/JngBLetfyn | 37 |
| slacker | Group Chatting Application Slacker Has Been Hacked - http://t.co/whBy5QdmCR > And I always thought slacker was for streaming music. #infosec | 9 |
| uk | RT @PotCoinFan: Hacking is cool if we do it just not you. - UK Government Authorized GHCQ to Hack Any Device https://t.co/D0AbsHG9nB via @H… | 11 |
| inec | RT @OmoOduaRere: INEC's Website get hacked by Nigeria Cyber hacking group on Election day #Nigeriadecides http://t.co/nn8mOhPulp | 61 |
| taylor swift | Taylor Swift Hacked, 6 Nudes for Sale? http://t.co/k6HhFe9nEa | 13 |
| inec website | INEC Website has been Hacked http://t.co/346VTVfVcD | 4 |
| darren | The people who think Darren was hacked are the same people who think he doesn't lie to his fans, wrote This Time for Lea, and likes vag. | 12 |

# Account Hijacking

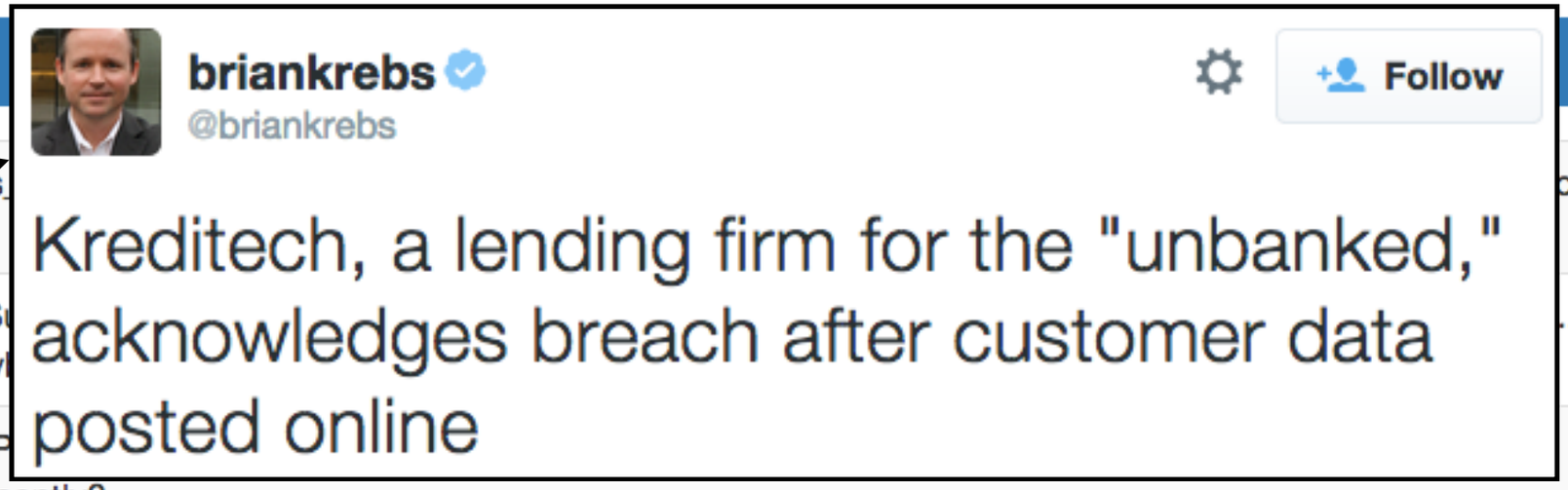

Mar-29-2015

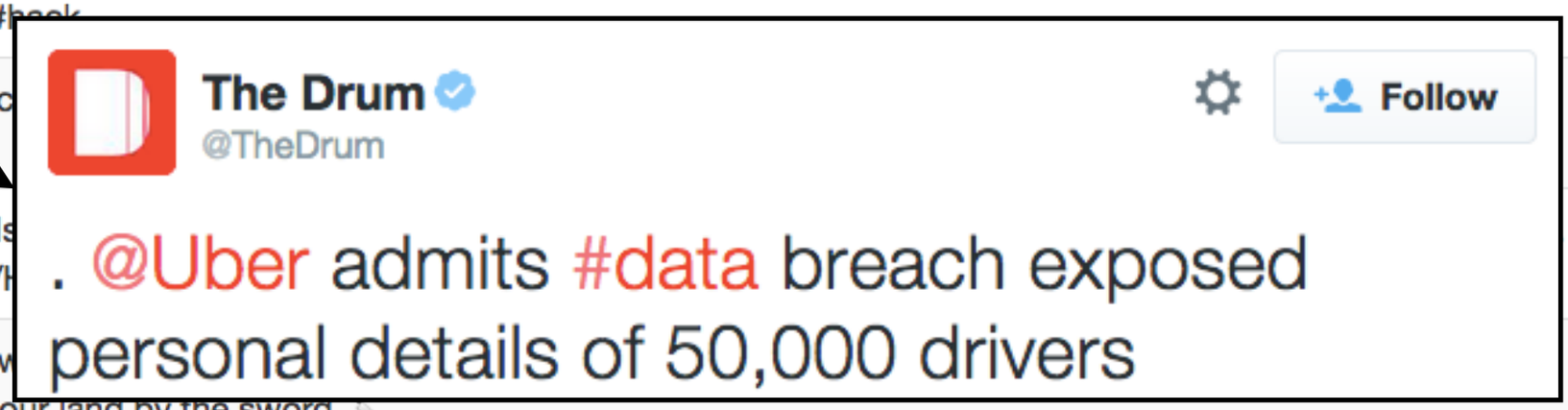
| | | |
|---|---|---|
| british airways | British Airways says some frequent flyer accounts hacked http://t.co/rsh1qOZ7WH | 210 |
| nigeria | "Nigeria: Election Website | 37 |
| slacker | Group Chatti[illegible] Appli[illegible]atio streaming music. #i[illegible] | 9 |
| uk | RT @PotCoinFan: Hacking https://t.co/D0AbsHG9nB | 11 |
| inec | RT @OmoOduaRere: INEC http://t.co/nn8mOhPulp | 61 |
| taylor swift | Taylor Swift Hacked, 6 Nu | 13 |
| inec website | INEC Website has been H | 4 |
| darren | The people who think Dar[illegible] likes vag. | 12 |

# Data Breach

Mar-28-2015

| | | |
|---|---|---|
| banks crabb badger davis elphick breach | RT @u…s_ | 4 |
| kreditech | Kreditech Su… http://t.co/vl… | 13 |
| uber | RT @EnverP… reported recently? | 13 |
| premera | RT @DMBisson: Premera faces class action suit over data #breach http://t.co/oiKLf01HkI via @modrnhealthcr #security #h… | 14 |
| provider reports security breach | …ck Attac… | 3 |
| google | Google Ads… http://t.co/H… | 21 |
| israel | Fav if you w… take back our land by the sword | 15 |
| amazon | Twitch, the video game streaming site Amazon bought for $970 million, has been hacked (AMZN) http://t.co/1YMfZPT2Cx #business | 7 |
| twitch | New hacking group group DDoS attacks Amazon's Twitch, US department websites \| http://t.co/jkzZV2JH28 | 15 |



# Contributions

- Microblogs are a valuable source of info for cybersecurity events
- Novel approach to event extraction with minimal supervision
  - Weakly supervised learning as a problem with positive and unlabeled data.
  - New approach based on expectation regularization
- Demo: http://securitytweets.org/

# Contributions

- Microblogs are a valuable source of info for cybersecurity events
- Novel approach to event extraction with minimal supervision
  - Weakly supervised learning as a problem with positive and unlabeled data.
  - New approach based on expectation regularization
- Demo: http://securitytweets.org/

**Thanks!**