Tutorial: Explainable AI for Societal Event Predictions: Foundations, Methods, and Applications

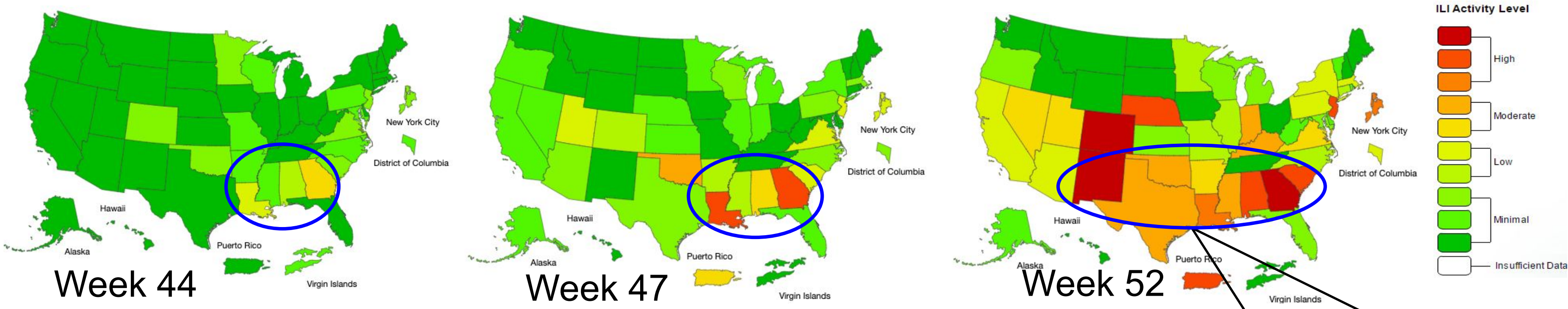
Songgaojun Deng, Yue Ning, and Huzefa Rangwala

Wednesday, February 3, 2021
8:30 am – 10:00 am (PST)

Roadmap

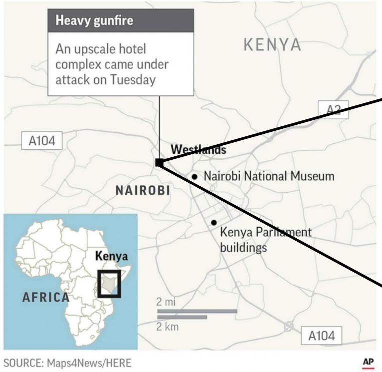
- Introduction and motivation
- Part 1: Precursor Identification for Interpretable Event Forecasting
- Part 2: Event Graphs for Interpretable Event Forecasting
- Conclusion and Future Directions

What are societal events?



influenza

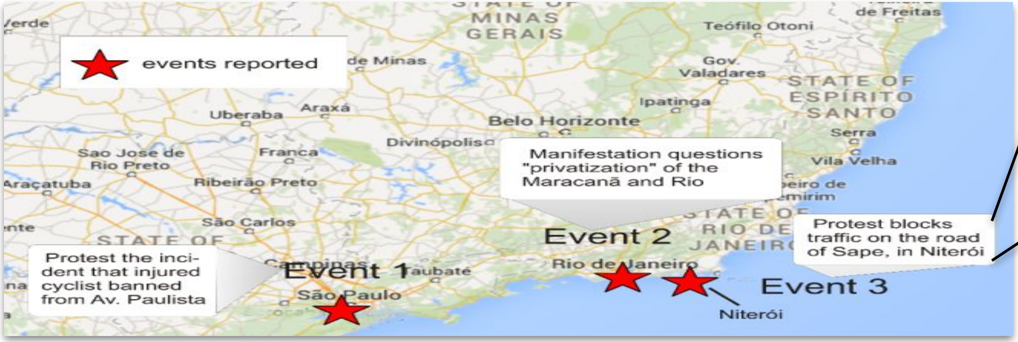
Epidemics outbreak during 2018-2019 in southern region



Terrorism events

Source: https://abcnews.go.com/International/high-end-complex-nairobi-attack-police/story?id=60387909

What are societal events?

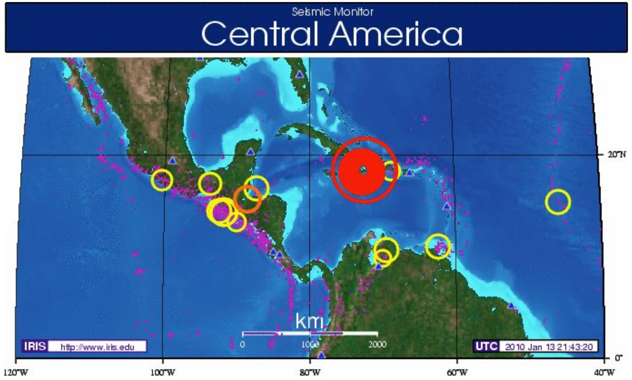


Civil unrest events on Mar 17, 2013 in Brazil

Protests

Economics crisis

https://medium.com/financial-times/has-the-yield-curve-predicted-the-next-us-downturn-b71a83ecfb74



Earthquake events

https://ds.iris.edu/ds/newsletter/vol12/no1/62/january-2010-m70-haiti-quake/

Societal Events

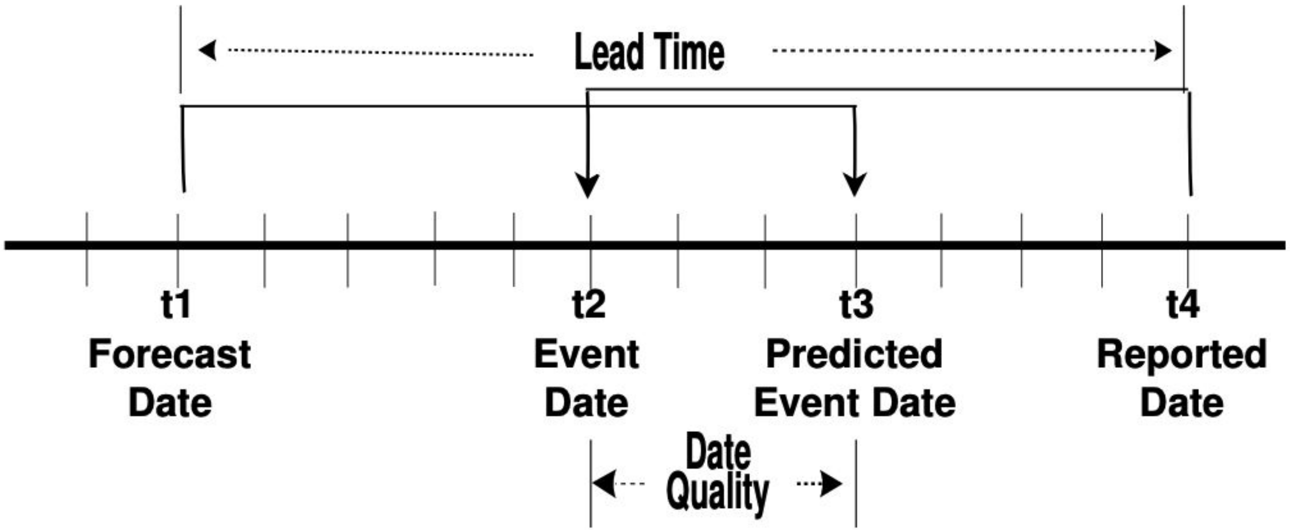
Societal Event Predictions

- The task is to predict the occurrence of events in the future with significant social impact.
- Underlying mechanism of societal events
  - Complex, dynamic, sparse
  - Hard to comprehensively model with limited data
  - Largely unknown

Social sensors → Model → Future events

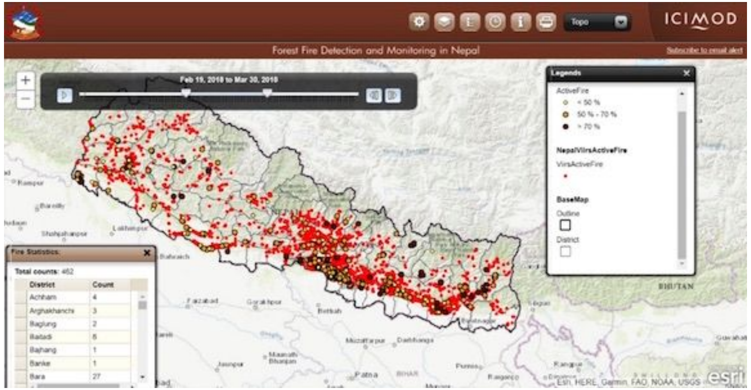
Build the forecaster driven by large historical data

Lead Time



N. Ramakrishnan et al. `Beating the news' with EMBERS: Forecasting Civil Unrest using Open Source Indicators, in Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'14), Aug 2014

Examples of Social Indicators

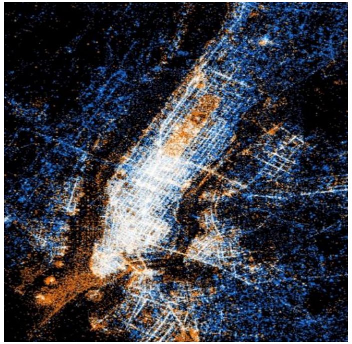


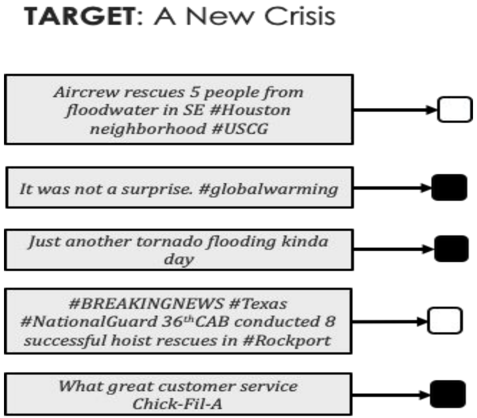
[Forest Fire Detection and monitoring in Nepal]

Global event encoding system

Characteristics of Social Indicators

- Ubiquitous
  - Every user/agent of social media/web/forum is a social sensor (citizen sensor)
  - They are everywhere observing the world all the time.
- Timeliness
  - 6,000 tweets every second.
  - 500 million tweets per day.
  - Usually beats the earliest official reports.
- Indicative and predictive signals



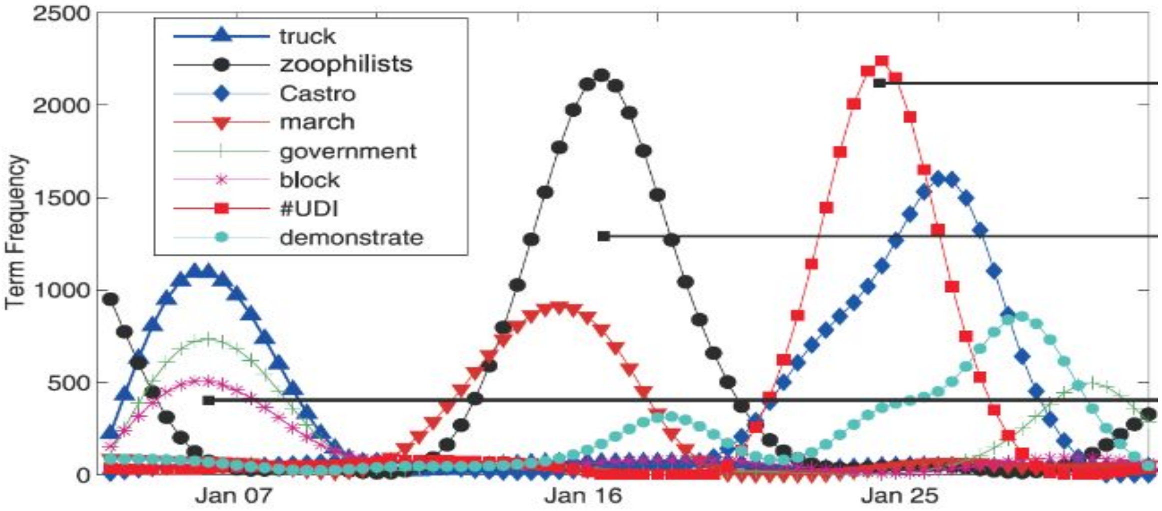


Explainable Event Predictions

- Social indicators can be general signals, features, and even distributions in open source data sets
- Precursor discovery refers to identifying specific examples or instances in the historical data given a prediction
- Explainable predictive models uncover significant features, graphs, documents for explaining prediction results.

Challenges in Explainable Event Predictions

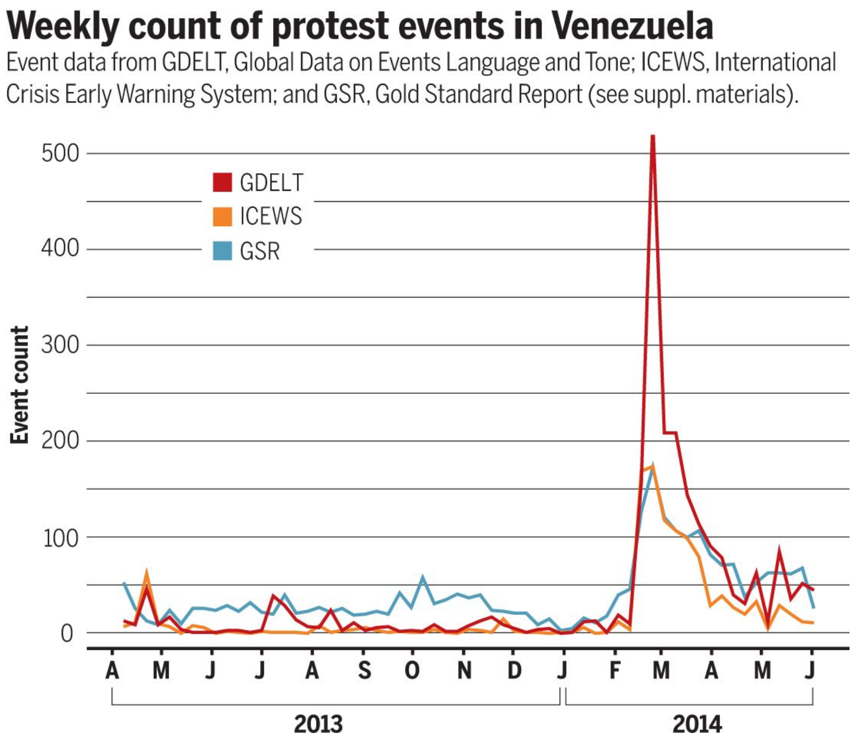
1. Dynamics
new #hashtags, abbreviations, new words



Source: Spatio-Temporal Event Forecasting and Precursor Identification. Tutorial in 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2019)

Challenges in Explainable Event Predictions

2. Multi-source unstructured data

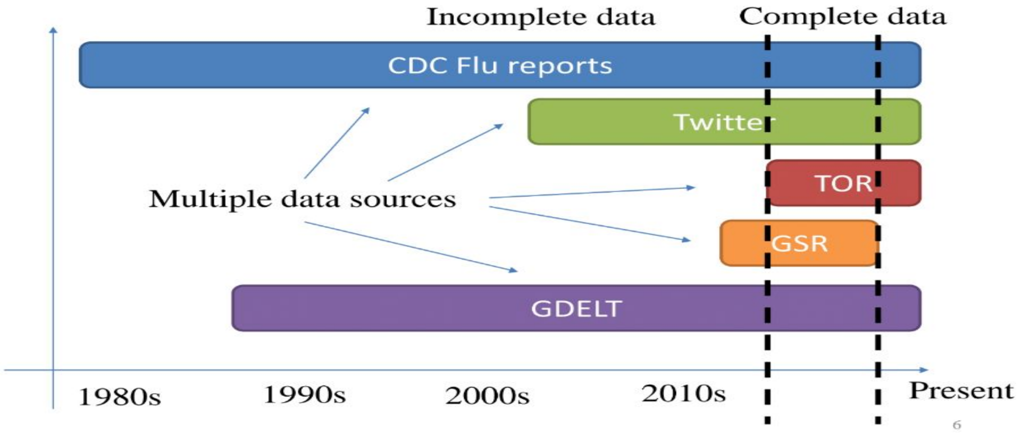


Source: Wei Wang, Ryan Kennedy, David Lazer, Naren Ramakrishnan. Growing pains for global monitoring of societal events. Science 30 Sep 2016. Vol. 353, Issue 6307, pp. 1502-1503

Challenges in Explainable Event Predictions

3. Data incompleteness

Reddits enable geo-info this year

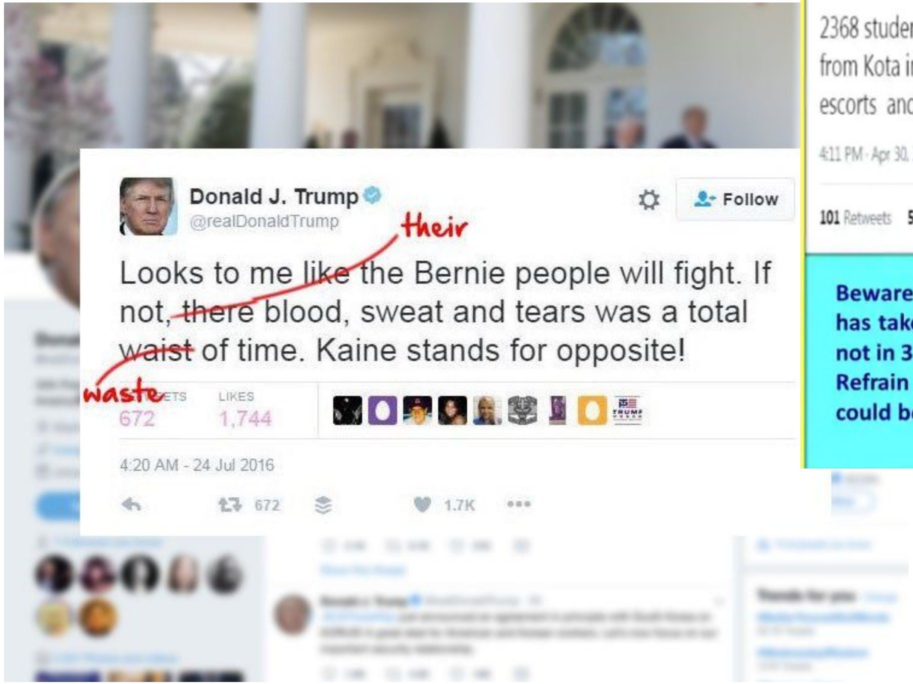


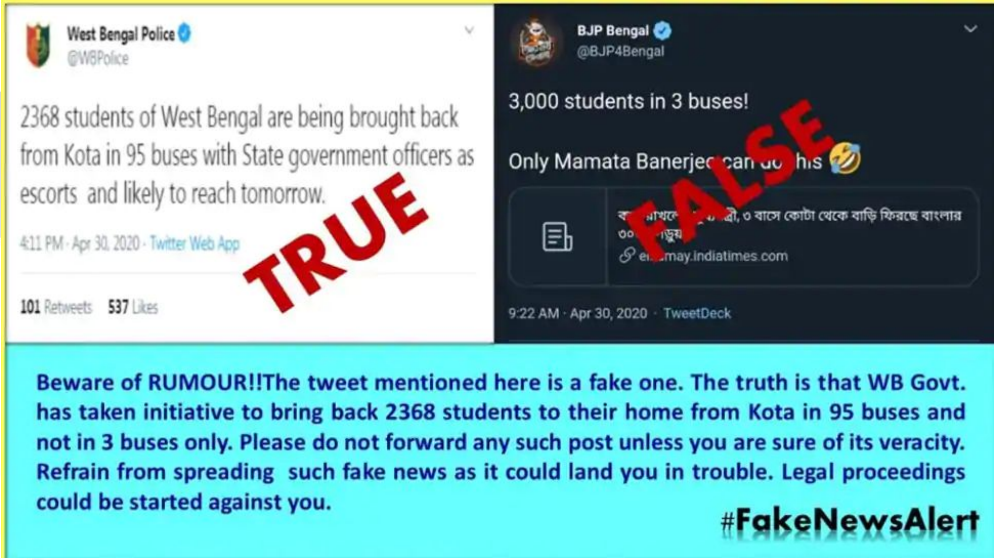
Source: Spatio-Temporal Event Forecasting and Precursor Identification. Tutorial in 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2019)

Challenges in Explainable Event Predictions

4. Noise in Data

typos, chit-chat, misinformation

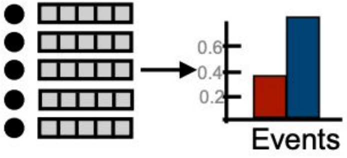




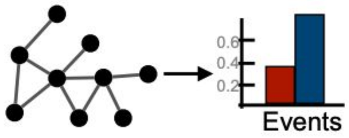
Challenges in Explainable Event Predictions

5. Heterogeneous data

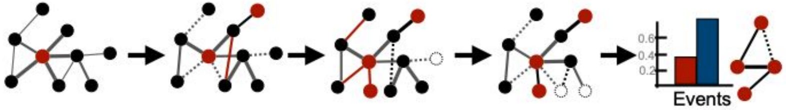
Need to specify forms of event explanations



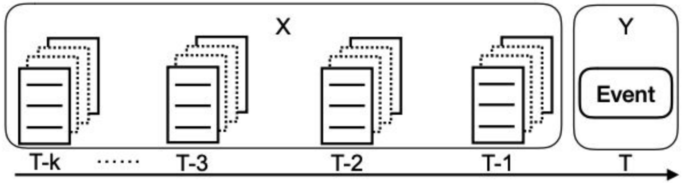
(a) Non-Graph Features



(b) Static Graph



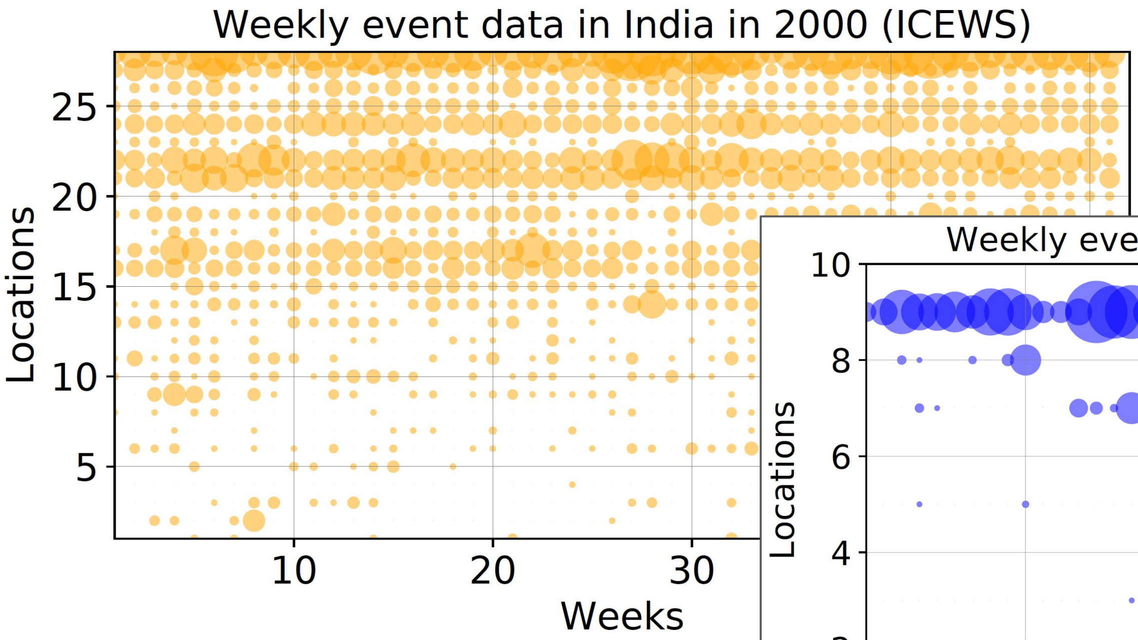
(c) Dynamic Context Graph

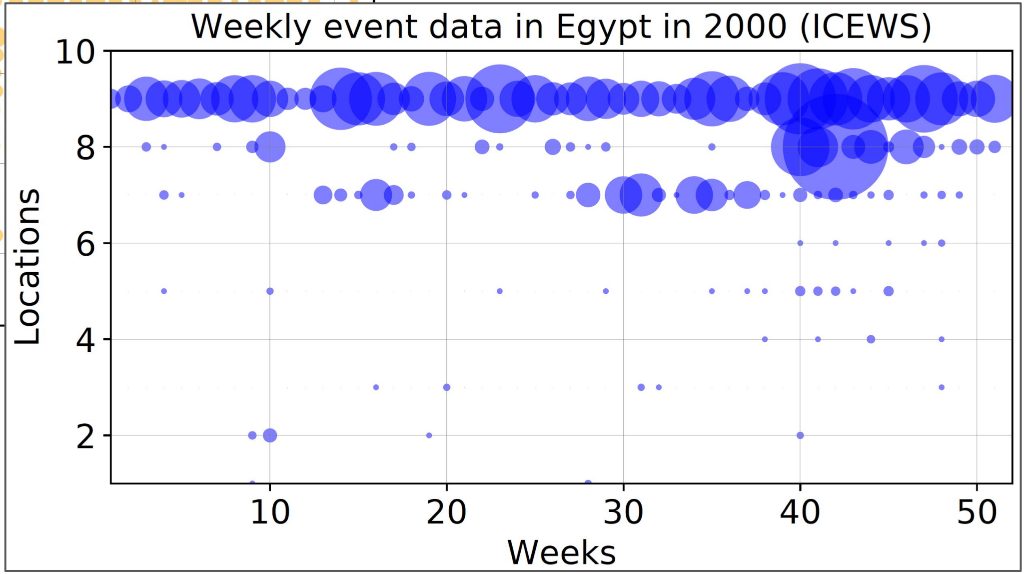


(d) Raw Input Data

Challenges in Explainable Event Predictions

6. Sparsity in spatio-temporal features





Other challenges

- Dependencies among events, e.g., spatial dependencies
- Lack of labeled data, cannot afford to label massive data

- Model interpretability – societal events are influential
- Lack mechanism models

Comparisons with Event Modeling Tasks

Event detection

- Historical or Ongoing events
- Discover anomaly
- Model types
  - Unsupervised learning
- Relevant techniques
  - Anomaly detection
  - Outlier detection
  - Change detection
  - Motif discovery

Event forecasting

- Future events
- Discover the mapping
- Model types
  - Supervised learning
  - Self-supervised learning
  - Semi-supervised learning
- Relevant techniques
  - Autoregressive
  - Markov chain
  - Classification

Explainable discovery

- Future events
- Discover the mapping
- Model types
  - Supervised learning
  - Self-supervised learning
  - Semi-supervised learning
- Relevant techniques
  - Multi-instance learning
  - Multi-task learning
  - Classification
  - Deep learning
  - Causal inference

Comparisons with Spatial Prediction

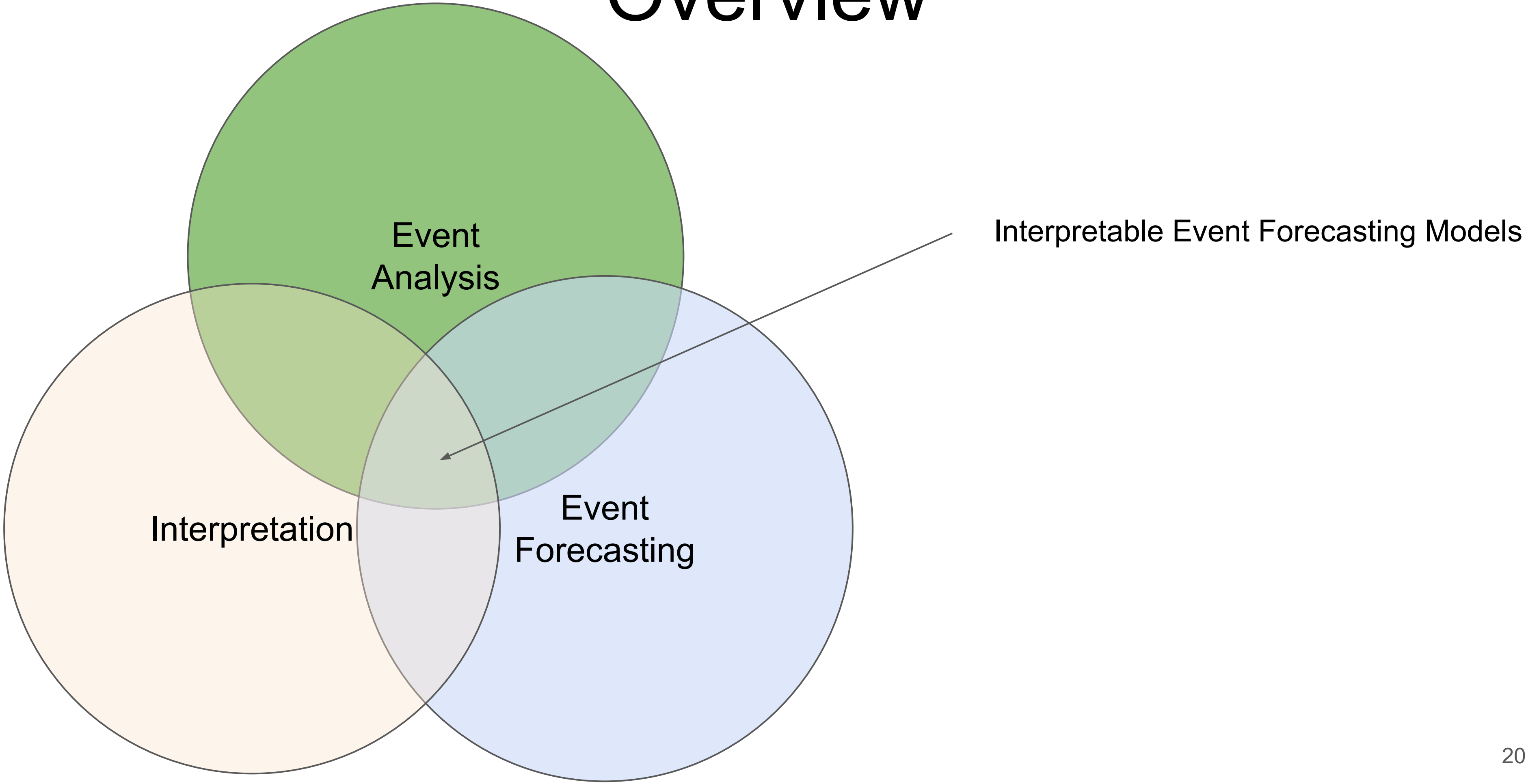
Prediction v.s. Forecasting:

- "Forecasting": Must be variable in the future.
- "Prediction": Not necessarily variable in the future.

- Spatial Prediction
  - Dependent variable
    - No need be in the future
    - Usually continuous values –"index"
  - Must have spatial dimension

- **Event Forecasting**
  - **Dependent variable**
    - **Must be in the future**
    - **Usually discrete values – "event"**
  - **No need be in spatial dimension**

Overview


Event
Analysis
Interpretation
Event
Forecasting
Interpretable Event Forecasting Models

Part 1: Precursor Identification in Spatio-Temporal Event Forecasting

2014 Venezuelan National Students Protest

major protests began with student marches led by opposition leaders in 38 cities.

Feb. 12

2014 Venezuelan National Students Protest

Opposition Leader, López, called upon students to peacefully protest.

major protests began with student marches led by opposition leaders in 38 cities.

Feb. 1

Feb. 12

2014 Venezuelan National Students Protest

López, alongside María Corina Machado launched a campaign to remove Maduro from office.

Opposition Leader, López, called upon students to peacefully protest.

major protests began with student marches led by opposition leaders in 38 cities.

Jan. 23

Feb. 1

Feb. 12

2014 Venezuelan National Students Protest

Murder of former Miss Venezuela, Monica Spear.

Former presidential candidate Henrique Capriles shook the hand of President Maduro

Attempted rape of a young student on a university campus in San Cristóbal

The harsh police response to their initial protest

January

López, alongside María Corina Machado launched a campaign to remove Maduro from office.

Jan. 23

Opposition Leader, López, called upon students to peacefully protest.

Feb. 1

major protests began with student marches led by opposition leaders in 38 cities.

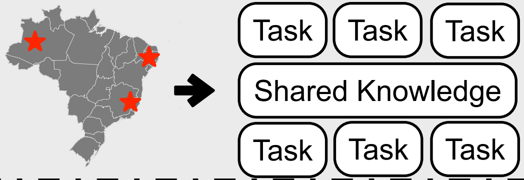
Feb. 12

If social scientists need to do this a lot ……

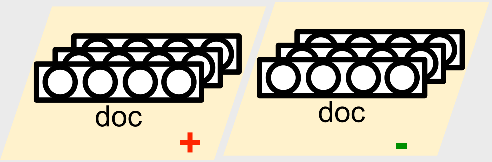
http://phdcomics.com/

The Big Picture

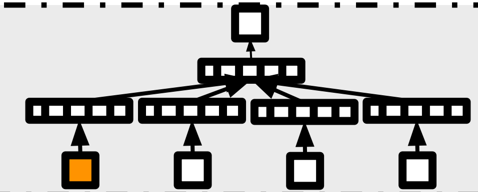
Multi-Task Learning
Relationships between locations;
Spatio-temporal event progression;



Multi-Instance Learning
Label propagation from bag to individual;
Temporal constraints between bags;

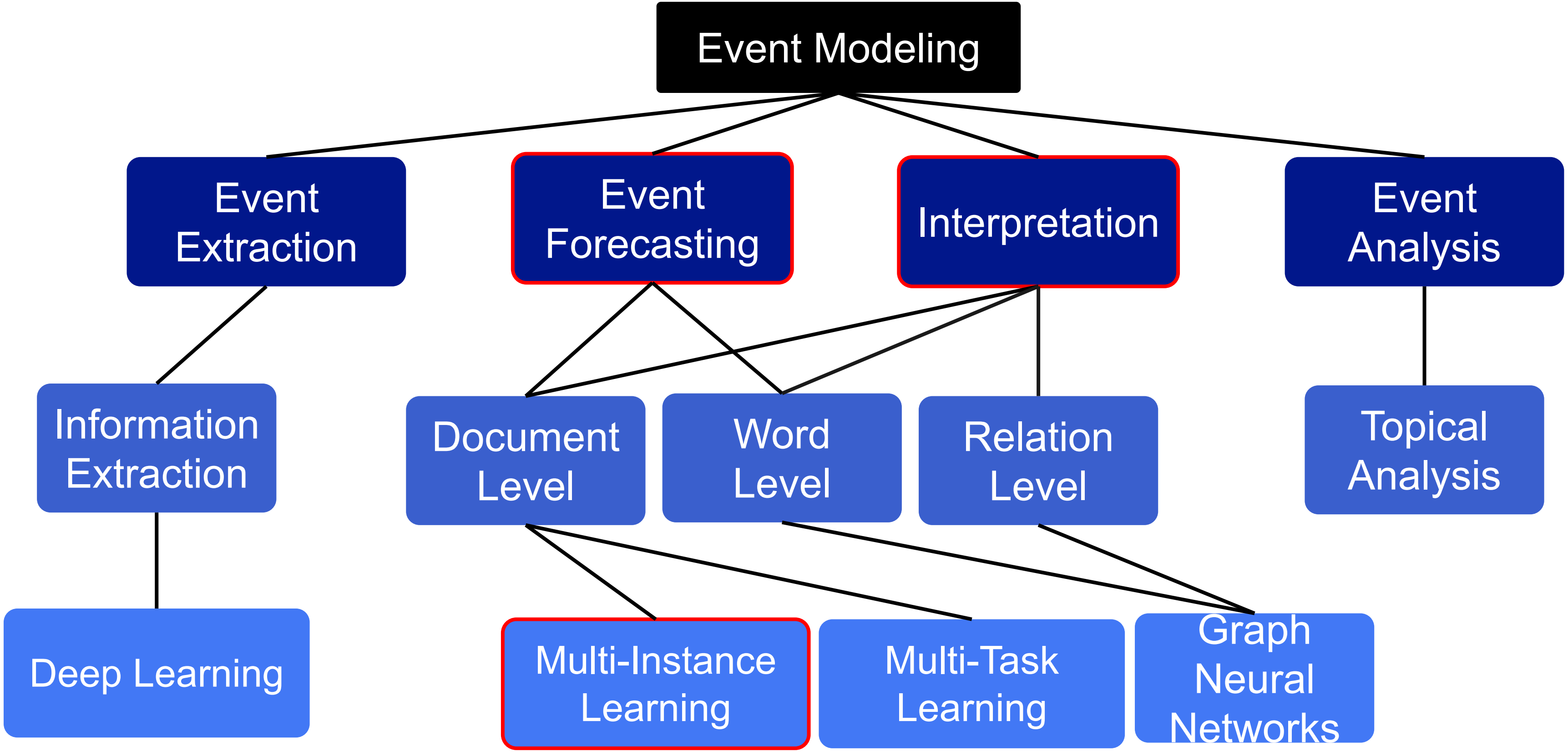


Representation Learning
Embeddings; word2vec; doc2vec; etc.



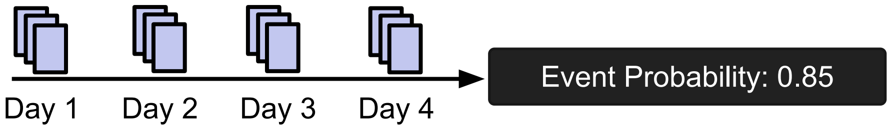
Social Indicators
News, blogs, social media, images, videos, time series, etc.

$$(X_1, X_2, ..., X_t) \rightarrow Y_{t+1}$$


Event Modeling
Event Extraction
Event Forecasting
Interpretation
Event Analysis
Information Extraction
Document Level
Word Level
Relation Level
Topical Analysis
Deep Learning
Multi-Instance Learning
Multi-Task Learning
Graph Neural Networks

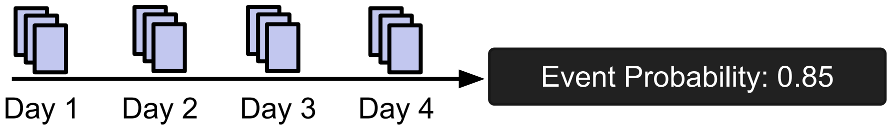
Precursor Discovery

- What is Precursor Discovery in Event Forecasting?
  - Forecast the occurrence of event of interest using historical data

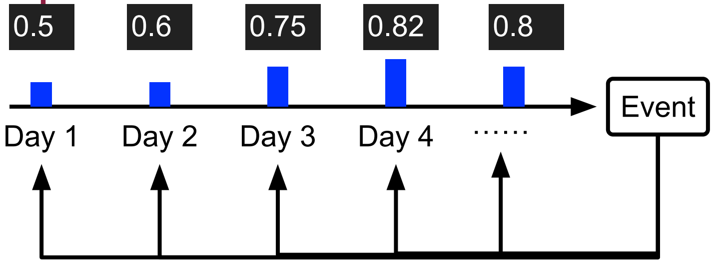


Precursor Discovery

- What is Precursor Discovery in Event Forecasting?
  - Forecast the occurrence of event of interest using historical data

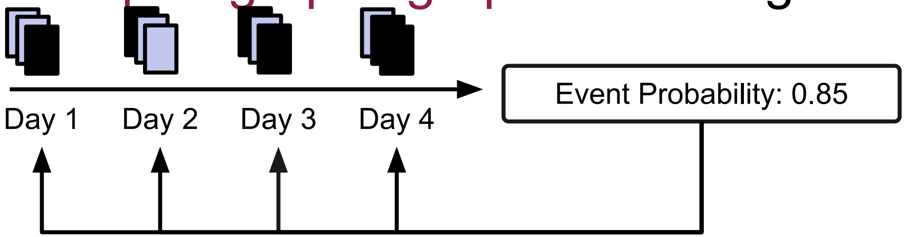


  - Predict days of importance before an event



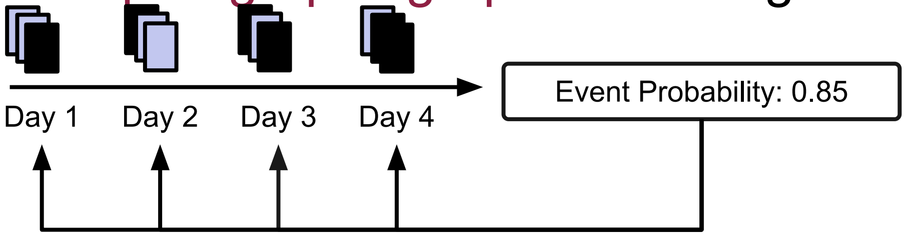
Precursor Discovery

- What is Precursor Discovery in Event Forecasting?
  - Identify key docs/paragraphs/graphs from large-scale input

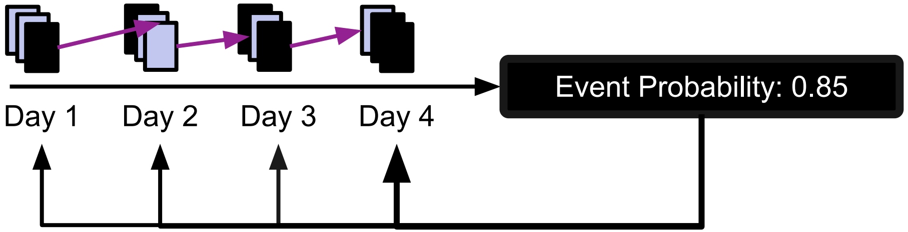


Precursor Discovery

- What is Precursor Discovery in Event Forecasting?
  - Identify key docs/paragraphs/graphs from large-scale input



  - Formalize precursor storylines



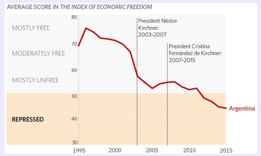
Precursor Storyline

Agenda developed for the migrant perspective

July 28

The Economic Convocation Argentina organized a demonstration

July 29



July 30: Standard of Poverty

Unresolved debts in Argentina

July 31

Aug. 1: International Court of Justice verdict on Argentine debt crisis

Article on government power

Aug. 3

Lear layoffs caused workers to block highway

Aug. 5

Aug. 7: Workers demand for better job opportunities

Existing Methods

- Existing approaches for event forecasting (when), examples:
  - Lasso [Zhao *et al*, TKDE17];
  - Fusion Method [Ramakrishnan *et al*, KDD14];
  - Multi-Task Learning [Zhao *et al*, KDD15];
  - Generative model [Zhao *et al*, SDM15];

➡ Limitations:

- Focus on prediction performance, lack of explanation
- Unable to provide structured evidence

Existing Methods

- Existing approaches for identifying precursors (why), examples:
  - Storytelling [Hossain *et al*, KDD12];
  - Combinational mixed Poisson process [Rong *et al*, KDD15];

➡ Limitations:

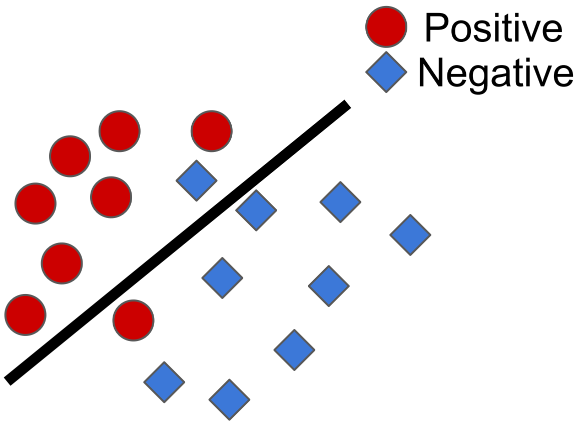
- Dependent on observed event sequence (time series, sequential)
- Lack of predictive value

Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning [Ning et al. KDD16]

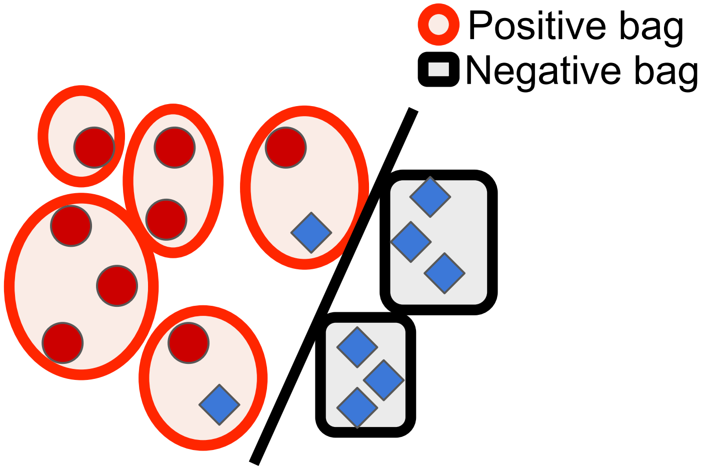
- The proposed method: a nested Multi-Instance Learning framework
  - Solve the above problems together (when & why)
  - Significantly reduce time of manual inspection of specialists/scientists
  - Generate storylines of indicators while predicting events of interest

Multi-Instance Learning

Supervised Learning

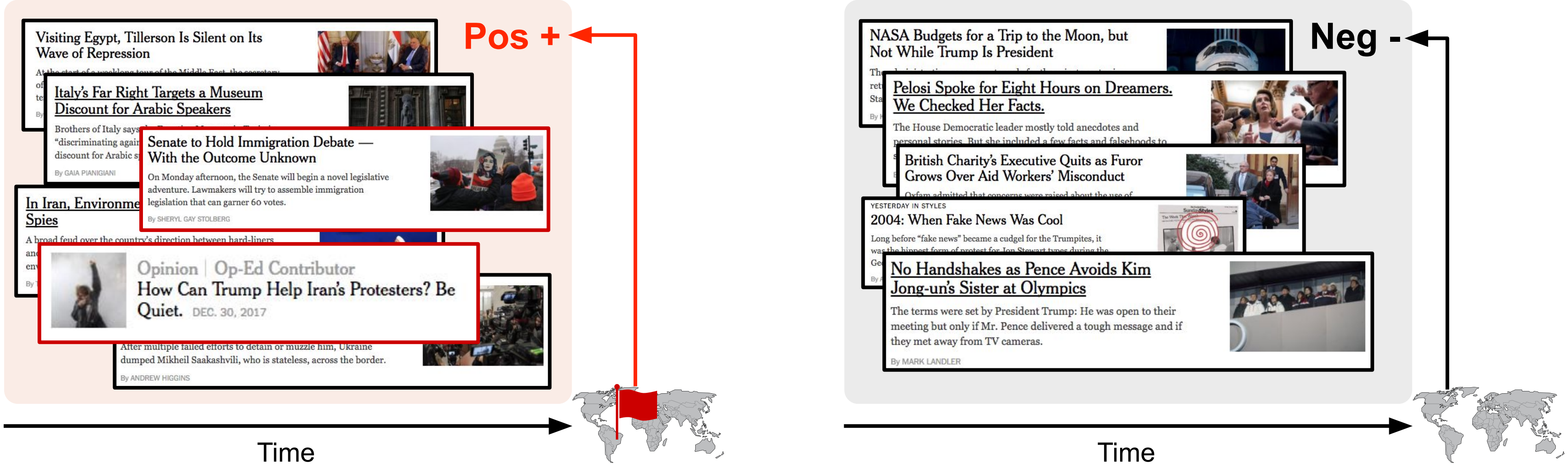


Multi-Instance Learning (MIL)



- Incomplete knowledge about labels in training data
- Propagate bag level supervision to individuals

Event Forecasting in Multi-Instance Learning

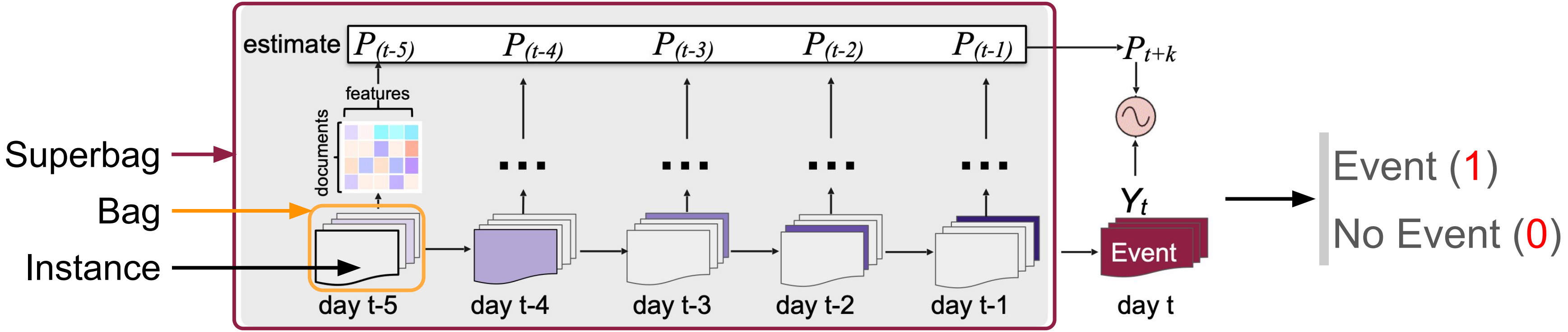


Event: Student Protest
Location: San Paulo
Time: 2014-05-01

Event: None
Location: San Paulo
Time: 2014-05-10

System Overview

- Target Prediction Label, Y



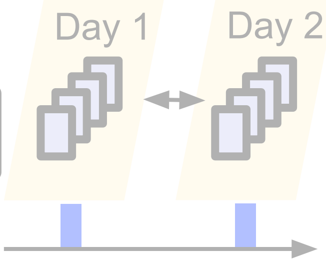
- *Nested Multiple Instance Learning*

- Each news article: *Instance*
- A group of news articles for a day: *Bag*
- A sequential collection of bags: *Super-Bag*
- Label is only associated at the *Super-Bag* Level
- Probabilistic Estimate for every *News Article* (Instance) and *Day* (Bag)

Nested MIL Objective Function

Reduce classification error

Control the probabilities of consecutive days

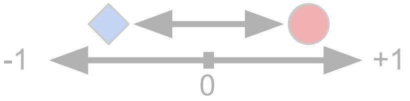


$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} \boxed{f(\mathbb{S}, Y, \mathbf{w})} + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^{t} \boxed{g(\mathcal{X}_i,\mathcal{X}_{i-1},\mathbf{w})}$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^{t} \frac{1}{n_i}\sum_{j=1}^{n_i} \boxed{h(\mathbf{x}_{ij},\mathbf{w})} + \lambda \boxed{R(\mathbf{w})}$$

Control the margin of instance probabilities

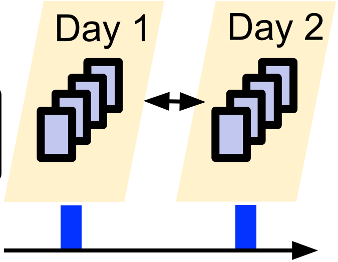
Avoid overfitting



Nested MIL Objective Function

Reduce classification error

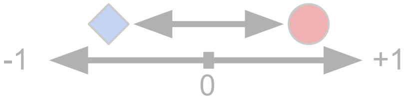
Control the probabilities of consecutive days



$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} \boxed{f(\mathbb{S}, Y, \mathbf{w})} + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^{t} \boxed{g(\mathcal{X}_i, \mathcal{X}_{i-1}, \mathbf{w})}$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^{t} \frac{1}{n_i}\sum_{j=1}^{n_i} \boxed{h(\mathbf{x}_{ij}, \mathbf{w})} + \lambda \boxed{R(\mathbf{w})}$$

Control the margin of instance probabilities

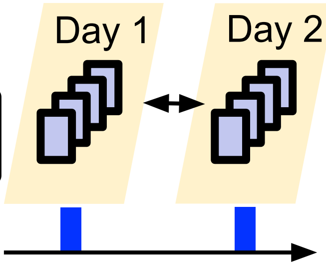


Avoid overfitting

Nested MIL Objective Function

Reduce classification error

Control the probabilities of consecutive days

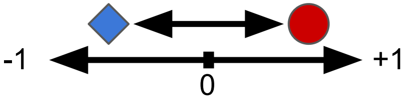


$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} \boxed{f(\mathbb{S}, Y, \mathbf{w})} + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^{t} \boxed{g(\mathcal{X}_i,\mathcal{X}_{i-1},\mathbf{w})}$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^{t}\frac{1}{n_i}\sum_{j=1}^{n_i}\boxed{h(\mathbf{x}_{ij},\mathbf{w})} + \lambda \boxed{R(\mathbf{w})}$$

Control the margin of instance probabilities

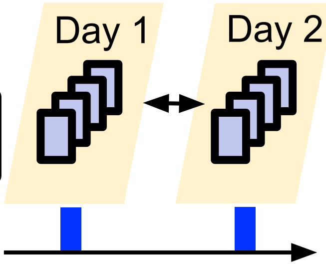
Avoid overfitting



Nested MIL Objective Function

Reduce classification error

Control the probabilities of consecutive days

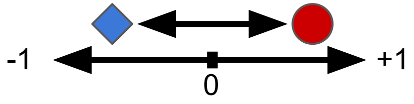


$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} \boxed{f(\mathbb{S}, Y, \mathbf{w})} + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^{t} \boxed{g(\mathcal{X}_i, \mathcal{X}_{i-1}, \mathbf{w})}$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^{t} \frac{1}{n_i}\sum_{j=1}^{n_i} \boxed{h(\mathbf{x}_{ij}, \mathbf{w})} + \lambda \boxed{R(\mathbf{w})}$$

Control the margin of instance probabilities

Avoid overfitting

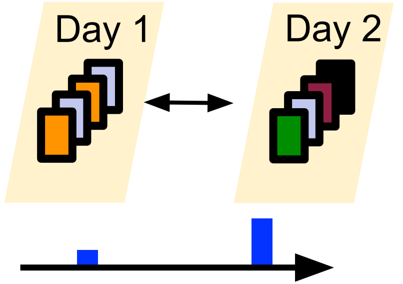


Nested MIL-Delta Objective Function

$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} f(\mathbb{S}, Y, \mathbf{w}) + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^{t} \boxed{g(\mathcal{X}_i, \mathcal{X}_{i-1}, \mathbf{w})}$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^{t}\frac{1}{n_i}\sum_{j=1}^{n_i} h(\mathbf{x}_{ij}, \mathbf{w}) + \lambda R(\mathbf{w})$$

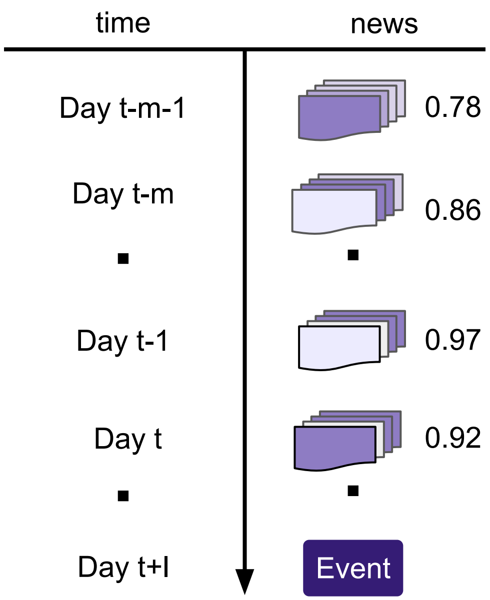
$$g(\mathcal{X}_i, \mathcal{X}_{i-1}, \mathbf{w}) = \underbrace{\Delta(\mathcal{X}_i, \mathcal{X}_{i-1})}_{\text{Cross-bag similarity}}(P_i - P_{i-1})^2$$



Precursor Discovery in Nested MIL

procedure PD-nMIL
  Input: $\mathcal{S} = \{(\mathbb{S}_r, Y_r)\}_{r \in n^+}, \mathcal{M}$
  Output: $\{(ps_r, Y_r)\}_{r \in n^+}$
  for super bag (S_r, Y_r) **do**
    $ps_r = []$
    for t = 1,2,...,h(history days) **do**
      $y_t = []$
      for $\mathbf{x}_{tm} \in \mathcal{X}_t$ **do**
        $\hat{y}_{tm} = \sigma(\hat{\mathbf{w}}\mathbf{x}_{tm})$
        if $\hat{y}_{tm} > \tau$ **then**
          $y_t \leftarrow (m, \hat{y}_{tm})$
      sort(y_t) by $\hat{y}_{tm}$ in descending order
      $ps_r \leftarrow$ m where m in top(y_t)
  return $\{(ps_r, Y_r)\}_{r \in n^+}$

<u>Selection of precursors based on their estimated probabilities</u>

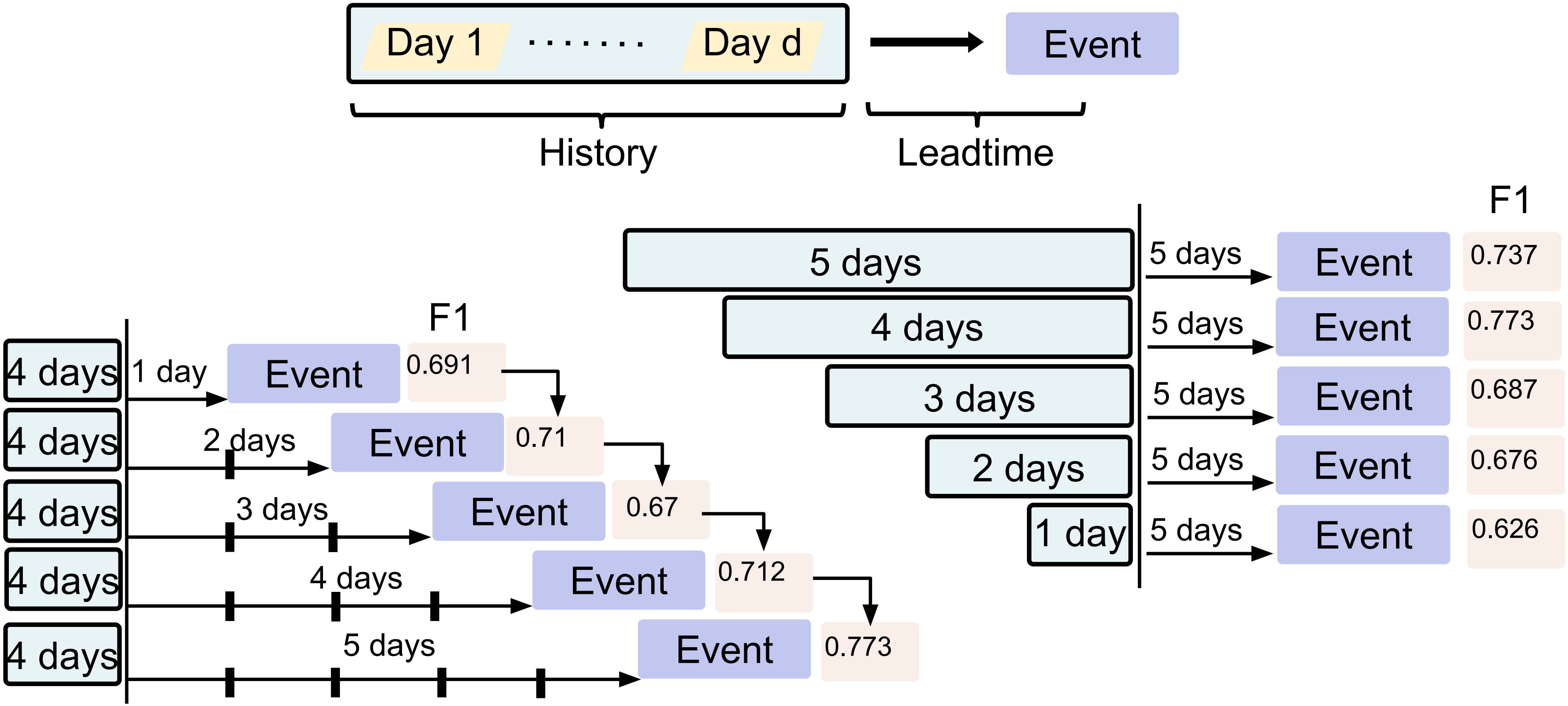


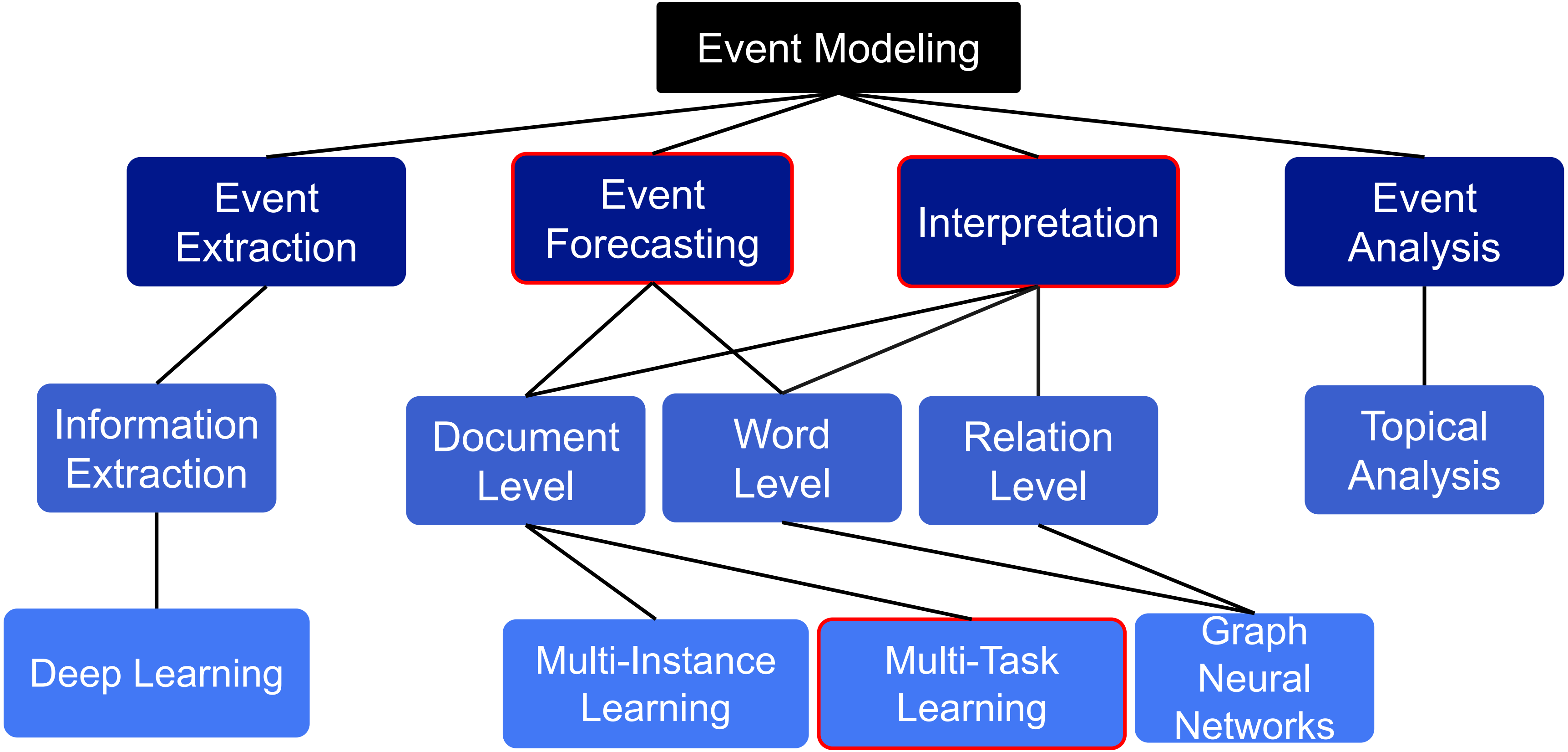
Predictive Performance

	Argentina		Brazil		Mexico	
	Acc	F-1	Acc	F-1	Acc	F-1
SVM	0.611($\pm$0.034)	0.406($\pm$0.072)	0.693($\pm$0.040)	0.598($\pm$0.067)	0.844($\pm$0.062)	0.814($\pm$0.091)
MI-SVM	0.676($\pm$0.026)	0.659($\pm$0.036)	0.693($\pm$0.040)	0.503($\pm$0.087)	0.880($\pm$0.025)	0.853($\pm$0.040)
rMIL-NOR	0.330($\pm$0.040)	0.411($\pm$0.092)	0.505($\pm$0.012)	0.661($\pm$0.018)	0.499($\pm$0.009)	0.655($\pm$0.025)
rMIL-AVG	0.644($\pm$0.032)	0.584 ($\pm$0.055)	0.509($\pm$0.011)	0.513($\pm$0.064)	0.785($\pm$0.038)	0.768($\pm$0.064)
GICF	0.589($\pm$0.058)	0.624($\pm$0.048)	0.650($\pm$0.055)	0.649 ($\pm$0.031)	0.770($\pm$0.041)	0.703($\pm$0.056)
nMIL	**0.709**($\pm$0.036)	0.702($\pm$0.047)	**0.723**($\pm$0.039)	0.686($\pm$0.055)	**0.898**($\pm$0.031)	**0.902**($\pm$0.030)
nMIL-Δ	0.708($\pm$0.039)	**0.714**($\pm$0.034)	0.705($\pm$0.048)	**0.698**($\pm$0.045)	0.861($\pm$0.014)	0.868($\pm$0.014)
nMIL-Ω	0.687($\pm$0.038)	0.680($\pm$0.045)	0.713($\pm$0.028)	0.687($\pm$0.038)	0.871($\pm$0.013)	0.879($\pm$0.014)

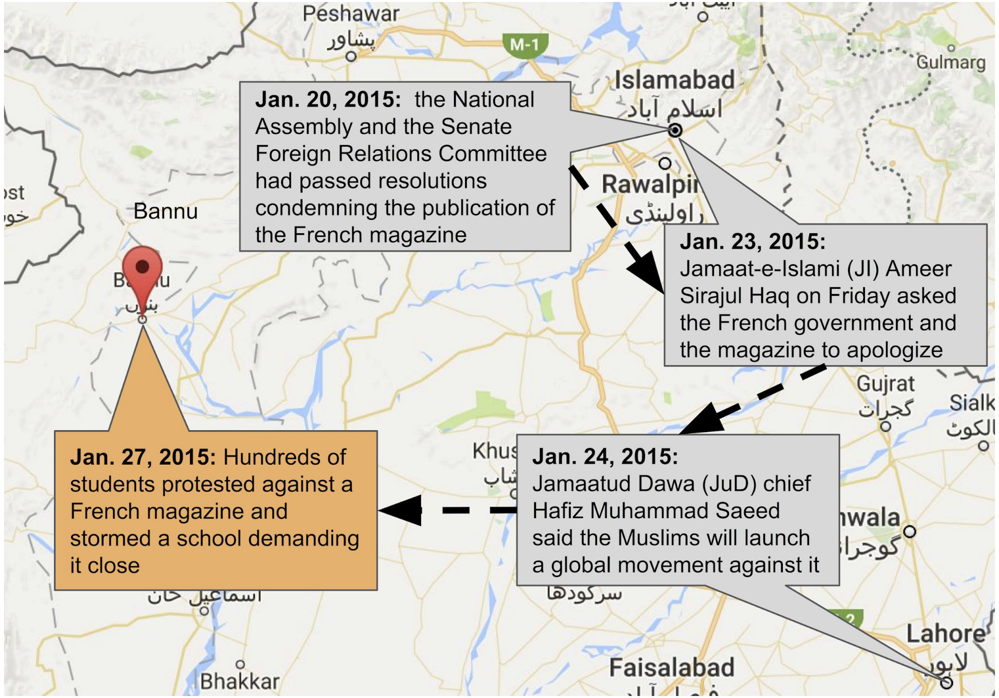
1. Nested structure models: nMIL, nMIL-Delta, nMIL-Omega
2. The averaged daily estimates help predict events of interest
3. Effect of time accumulation > a single input

How Early can NMIL Forecast?



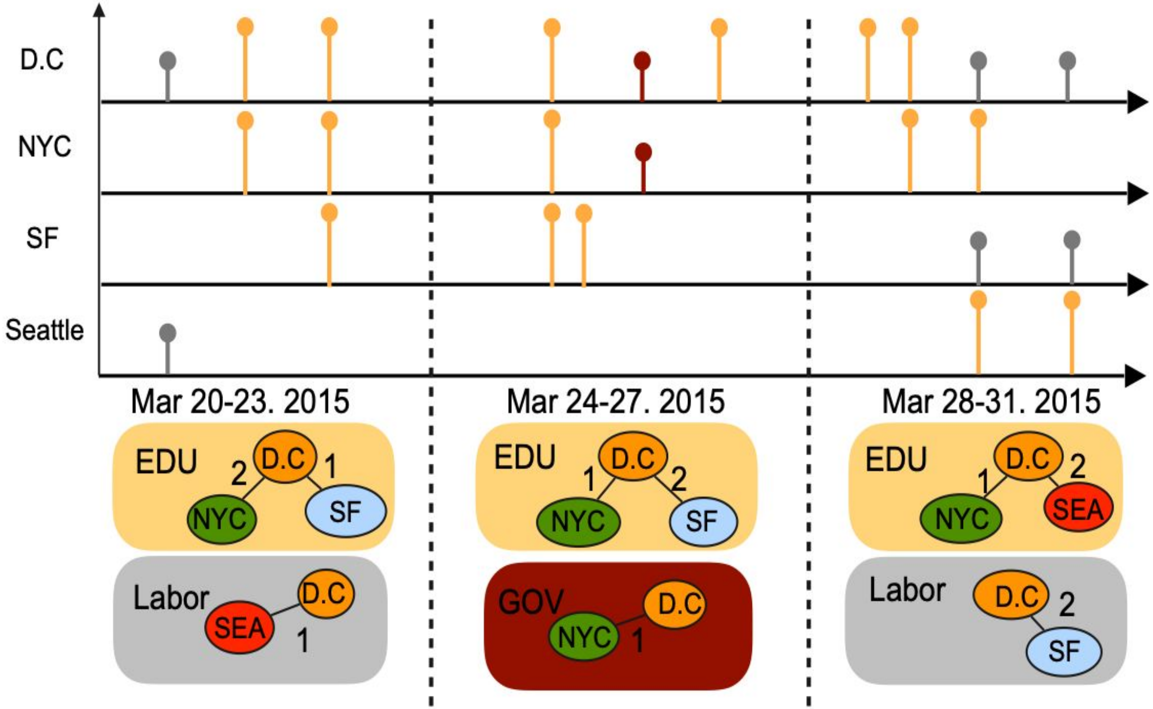

Event Modeling
Event Extraction
Event Forecasting
Interpretation
Event Analysis
Information Extraction
Document Level
Word Level
Relation Level
Topical Analysis
Deep Learning
Multi-Instance Learning
Multi-Task Learning
Graph Neural Networks

STAPLE: Spatio-Temporal Precursor Learning for Event Forecasting [Ning et al. SDM18]



Event, Geolocation, Time

STAPLE: Spatio-Temporal Precursor Learning for Event Forecasting [Ning et al. SDM18]



Event: Labor Protest
Location: D.C.
Time: 2015-04-01

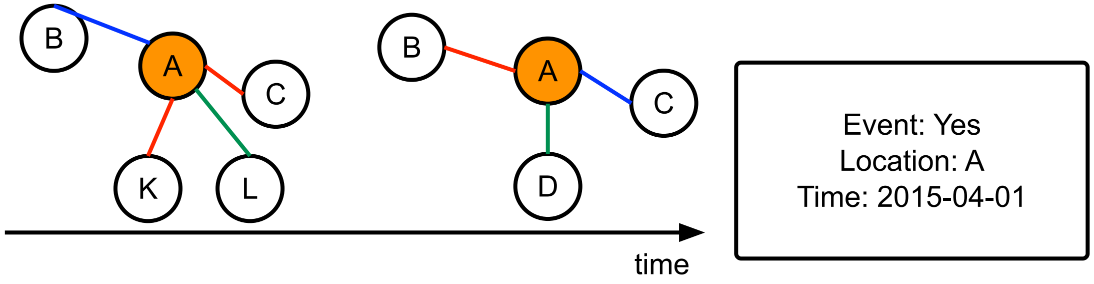
STAPLE: objective function

STAPLE: explicitly enforces pairs of cities with similar event patterns in the past to learn similar model vectors

all tasks share common features

$$\min_{\Theta} \sum_{k \in K} \Big(\underbrace{\frac{N_k}{N} \mathcal{L}(\theta^k)}_{\text{Multi-Instance Loss}} + \underbrace{\boxed{\frac{\lambda_1}{2} \sum_{i}^{N_k} \sum_{l \in \mathcal{G}_t} a_{k,l}^{t_i} \left(\theta^k - \theta^l\right)^2}}_{\text{Spatio-Temporal Constraints}} + \underbrace{\boxed{\frac{\lambda_2}{2} ||\hat{\theta} - \theta^k||_2^2}}_{\text{Global averaging}} + \underbrace{\frac{\lambda_3}{2} ||\theta^k||_2^2}_{L2} \Big)$$

STAPLE: spatio-temporal constraints

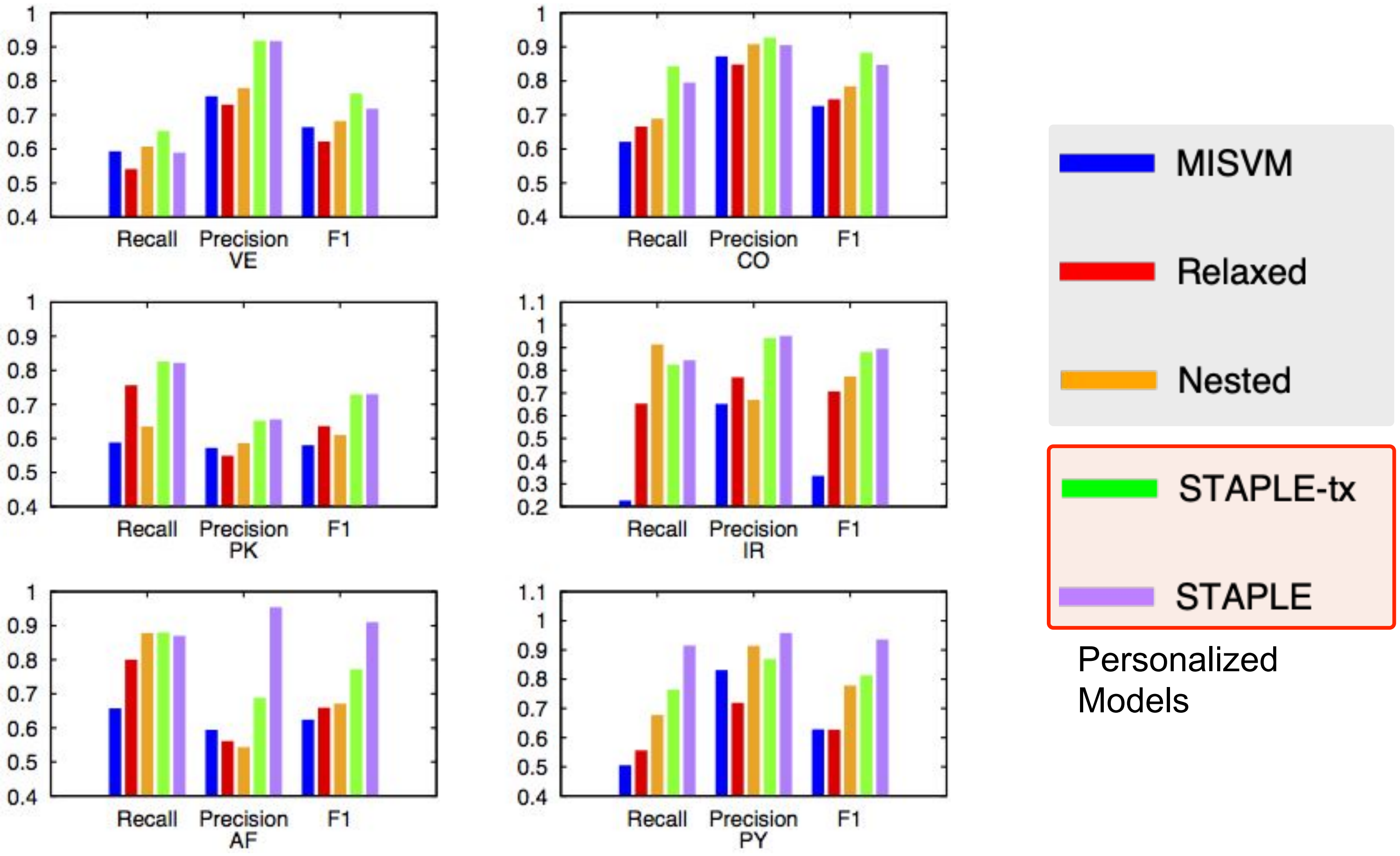


$$a_{k,l}^{t_i} = \left(\sum_c \sum_{t=t_i-H}^{t_i} \boxed{\min(E_t^k(c), E_t^l(c))} \right)' + \left(\boxed{\frac{1}{\mathrm{dist}(k,l)}} \right)'$$

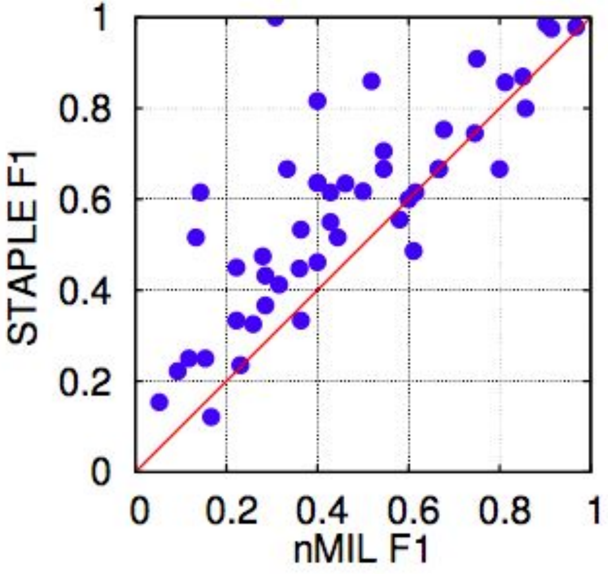
Similar event patterns in the past, similar models

Closer geolocations, similar models

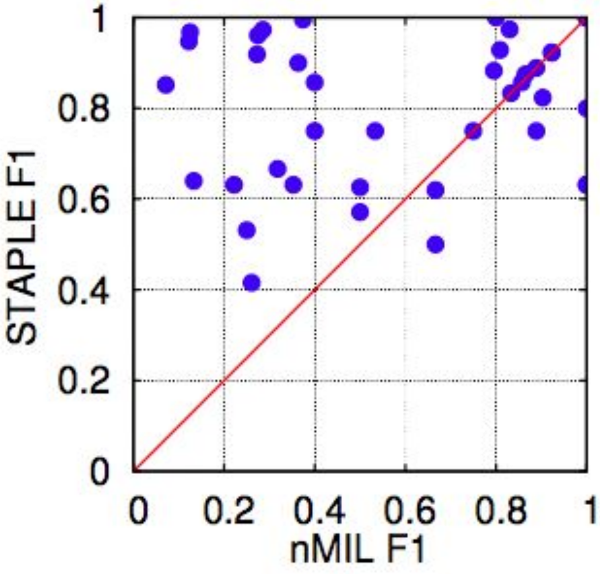
STAPLE:Event Prediction Performance



City-level Prediction Performance



(a) ICEWS



(b) GSR

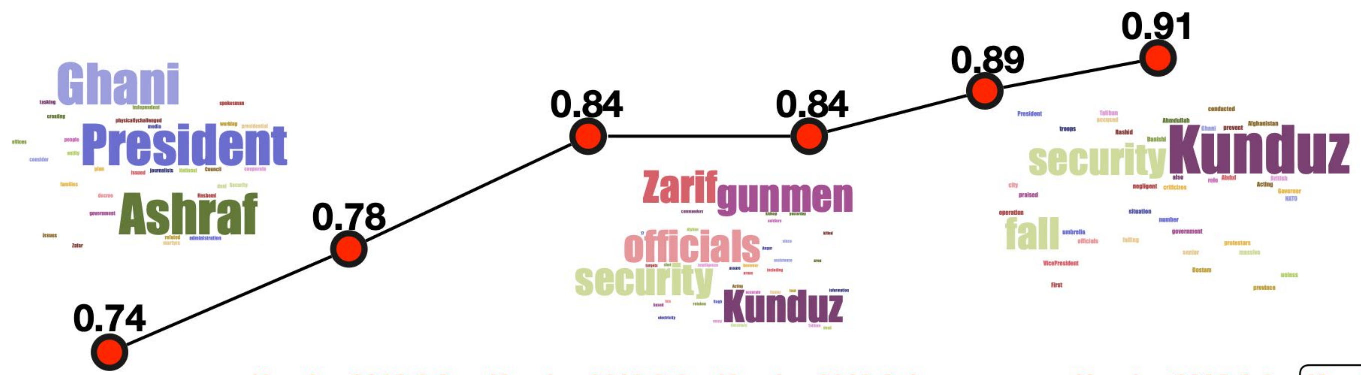
Security-related protest

KUNDUZ RESIDENTS STAGE PROTEST AGAINST MOUNTING INSECURITY

February 7, 2016 · Afghanistan · 13 Views

http://www.kabultribune.com/index.php/2016/02/07/kunduz-residents-stage-protest-against-mounting-insecurity/

Security-related protest - precursors



Kabul 2016-1-31
The presidential spokesman said the National Security Council was working on a plan

Kunduz 2016-2-2
Governor Zarif addressed speech on an accident

Kunduz 2016-2-3
Senior security officials say they killed 17 Taliban

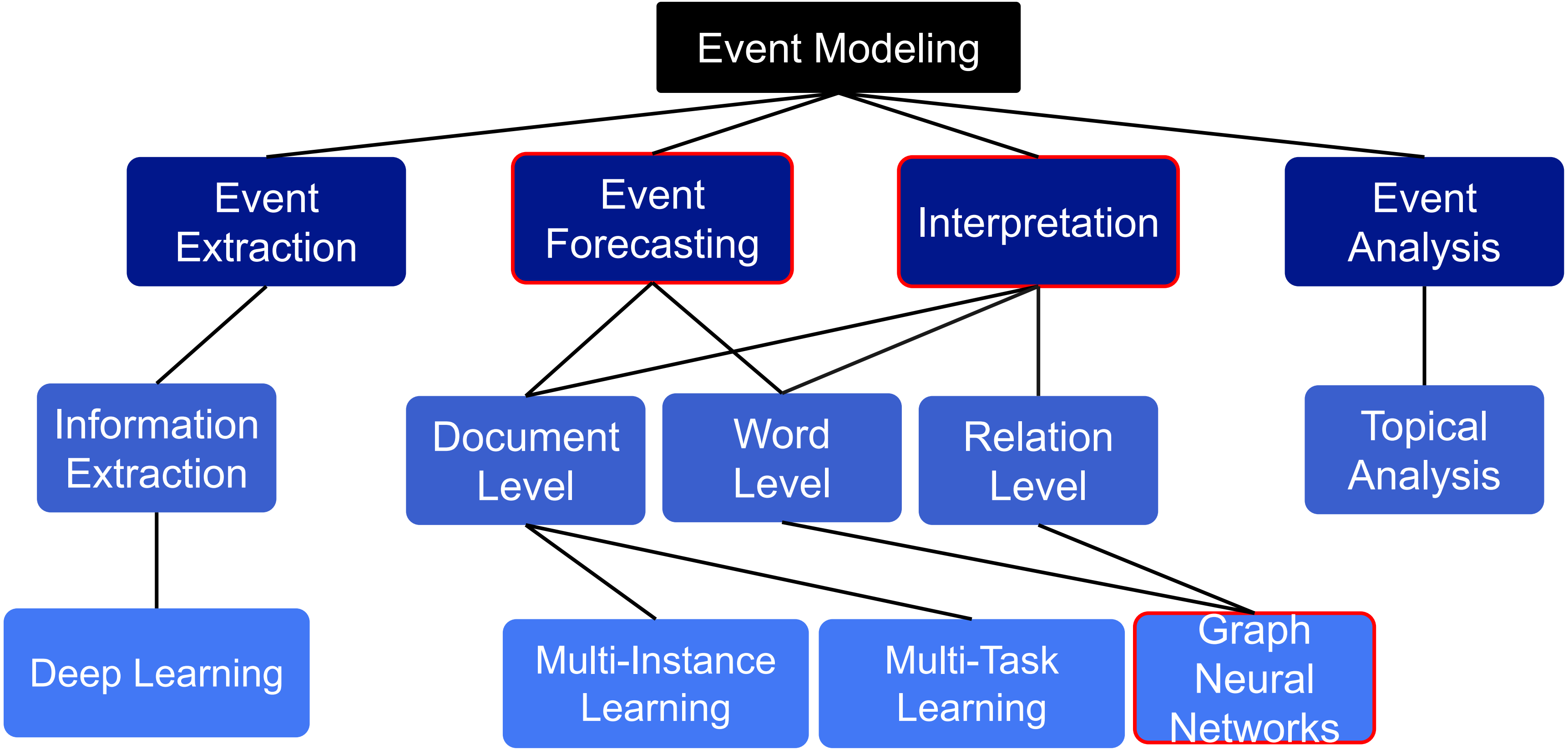
Kunduz 2016-2-4
First Vice-President criticizes the government for failing to prevent the fall of Kunduz

Kabul 2016-2-5
President Ghani praised the role of British troops

Kunduz 2016-2-6
Acting Kunduz Governor said Kuduz needs operation

Kunduz 2016-2-7
Protesters accused senior security officials of being negligent about the security situation

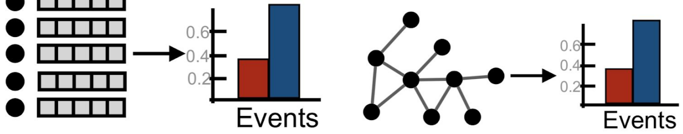
Time


Event Modeling
Event Extraction
Event Forecasting
Interpretation
Event Analysis
Information Extraction
Document Level
Word Level
Relation Level
Topical Analysis
Deep Learning
Multi-Instance Learning
Multi-Task Learning
Graph Neural Networks

Part 2: Event Graphs for Interpretable Event Forecasting

Learning Dynamic Context Graphs for Predicting Social Events [Deng et al. KDD19]

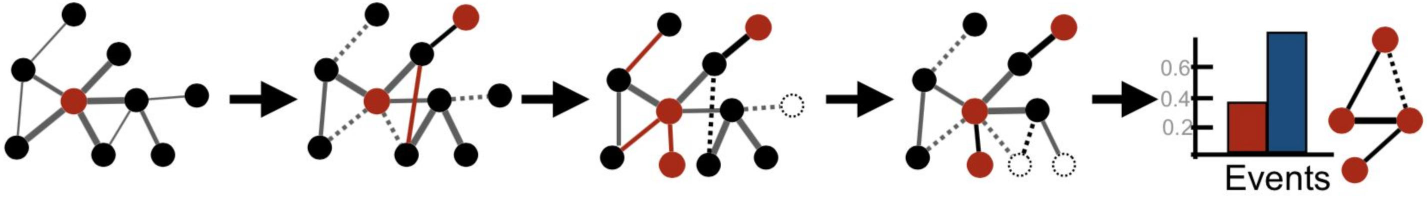
- Motivation
  - From the perspective of human analysts and policy makers, forecasting algorithms should
    - not only make accurate predictions
    - but also provide supporting evidence/clue
- Challenges
  - uncertainty of context structure and formulation
  - high dimensional features
  - adaptation of features over time



(a) Non-Graph Features (b) Static Graph

- **Model contextual information** for event forecasting

Learning Dynamic Context Graphs for Predicting Social Events [S. Deng et al. KDD19]



- Develop a novel graph-based model for predicting events
- Design a mechanism that encodes the dynamic graph structure of words from past input documents to forecast future events.
- Propose a temporal encoding module to alleviate the problem that pre-trained semantic features usually cannot reflect contextual changes over time.

Graph Convolutional Networks

[kipf and welling ICLR17]

Main idea: Pass messages between pairs of nodes

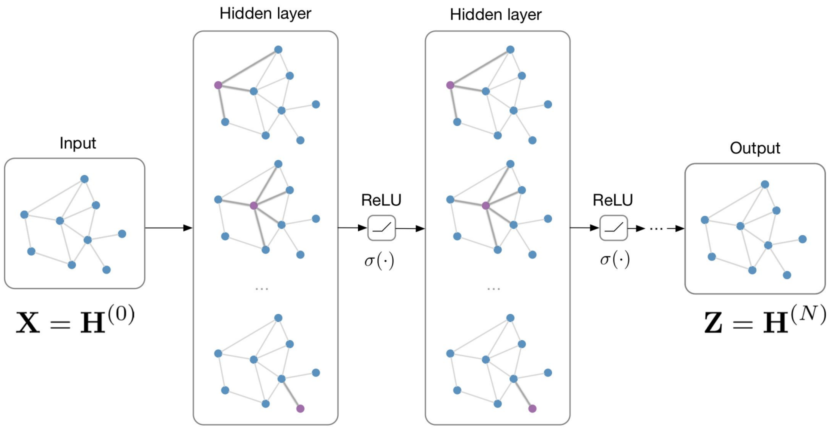
Graph: $G = (\mathcal{V}, \mathcal{E})$

$\mathcal{V}$: Set of nodes $\{v_i\}$, $|\mathcal{V}| = N$

$\mathcal{E}$: Set of edges $\{(v_i, v_j)\}$

Notation: $\mathcal{G} = (\mathbf{A}, \mathbf{X})$

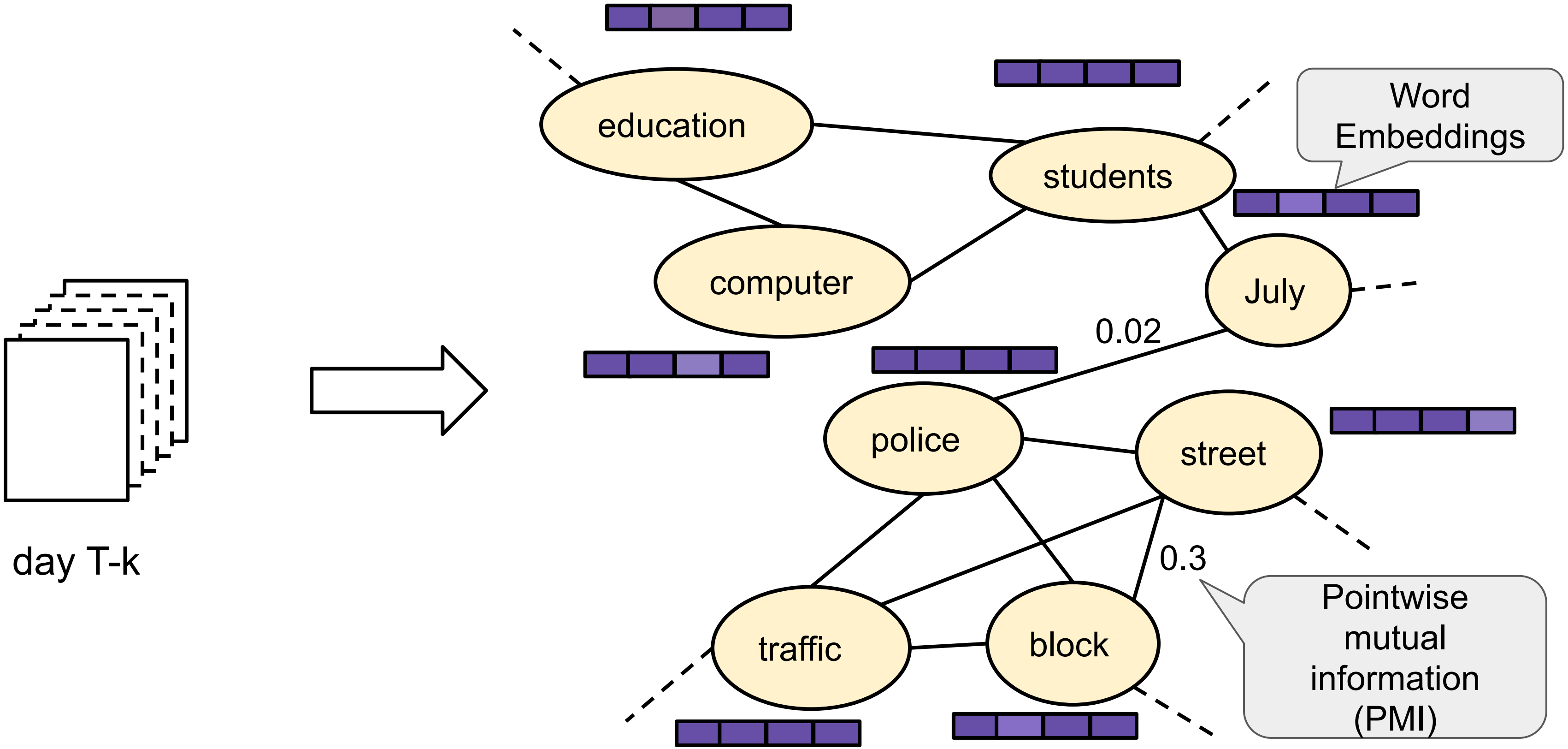
- Adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$
- Feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$



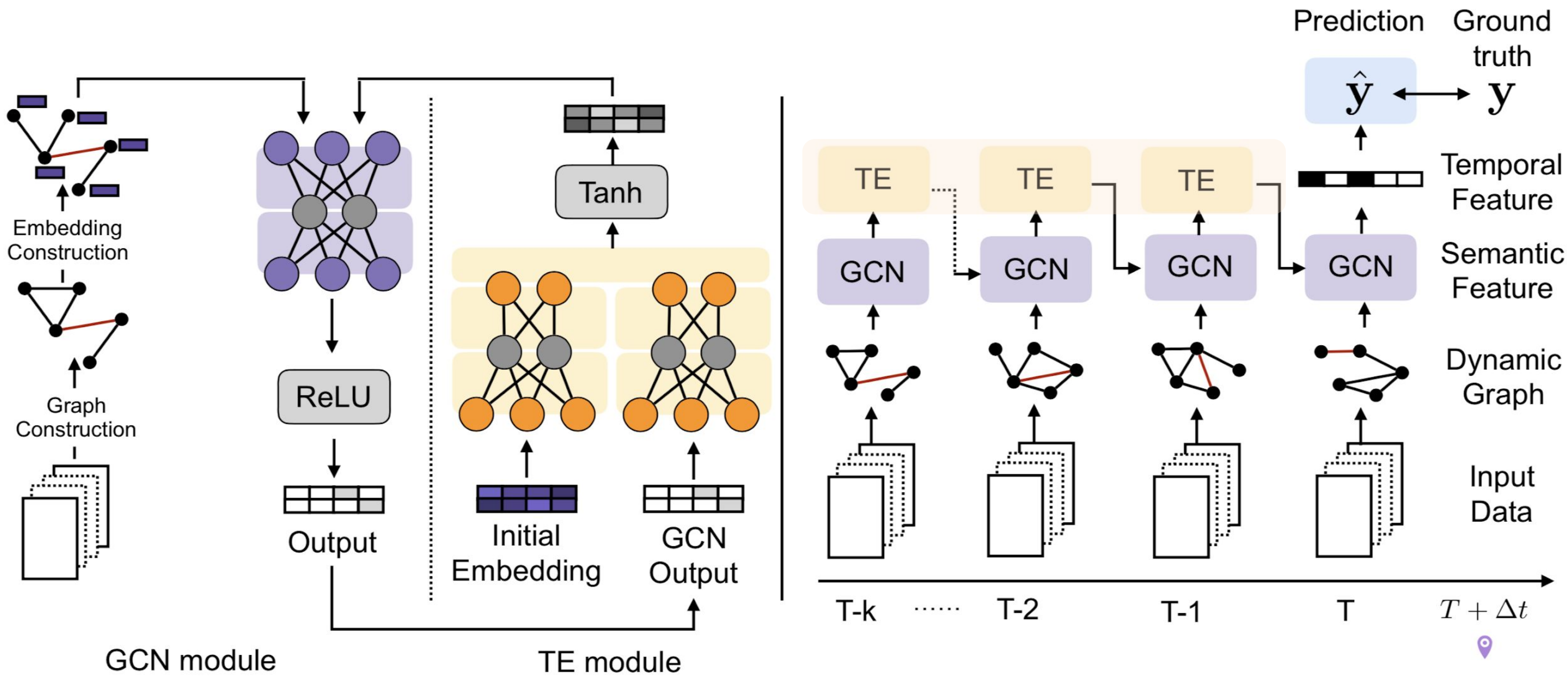
$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

Source: https://tkipf.github.io/graph-convolutional-networks/

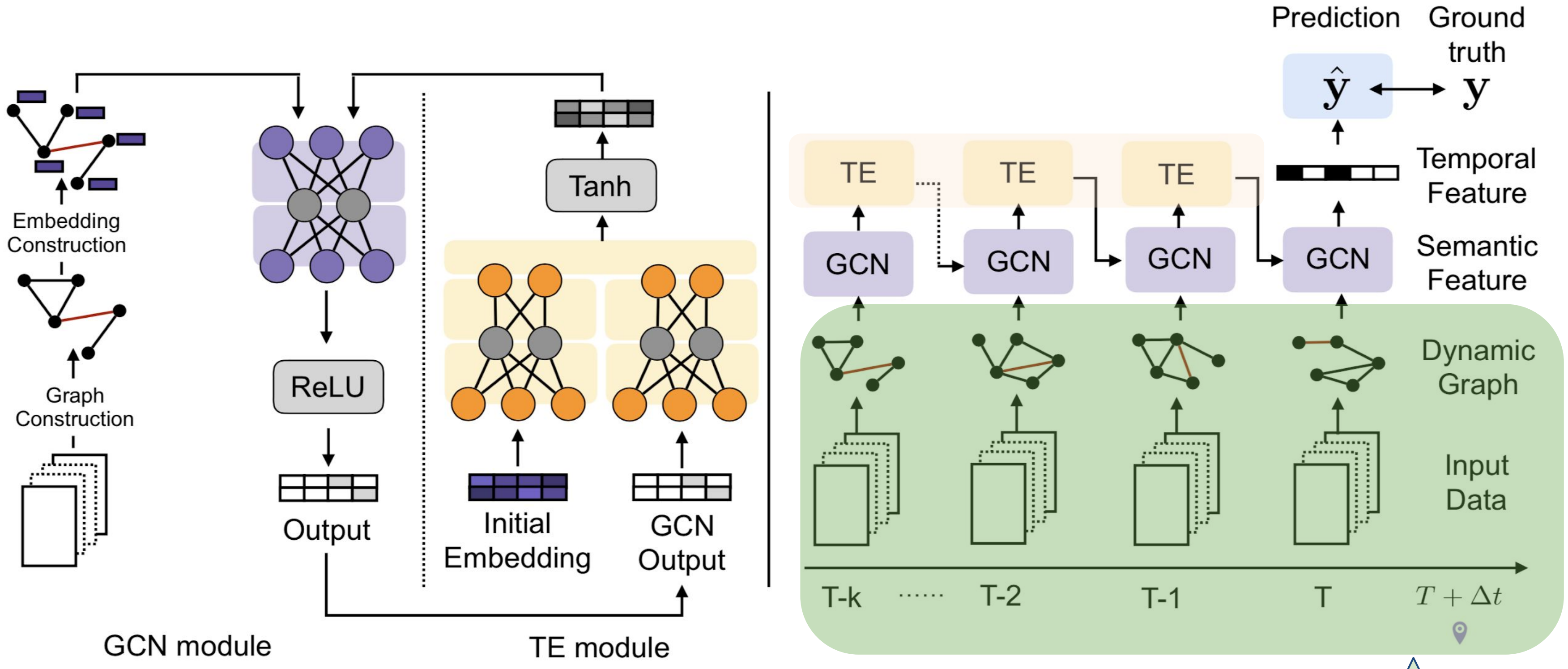
Encoding documents into graphs



DynamicGCN: model framework

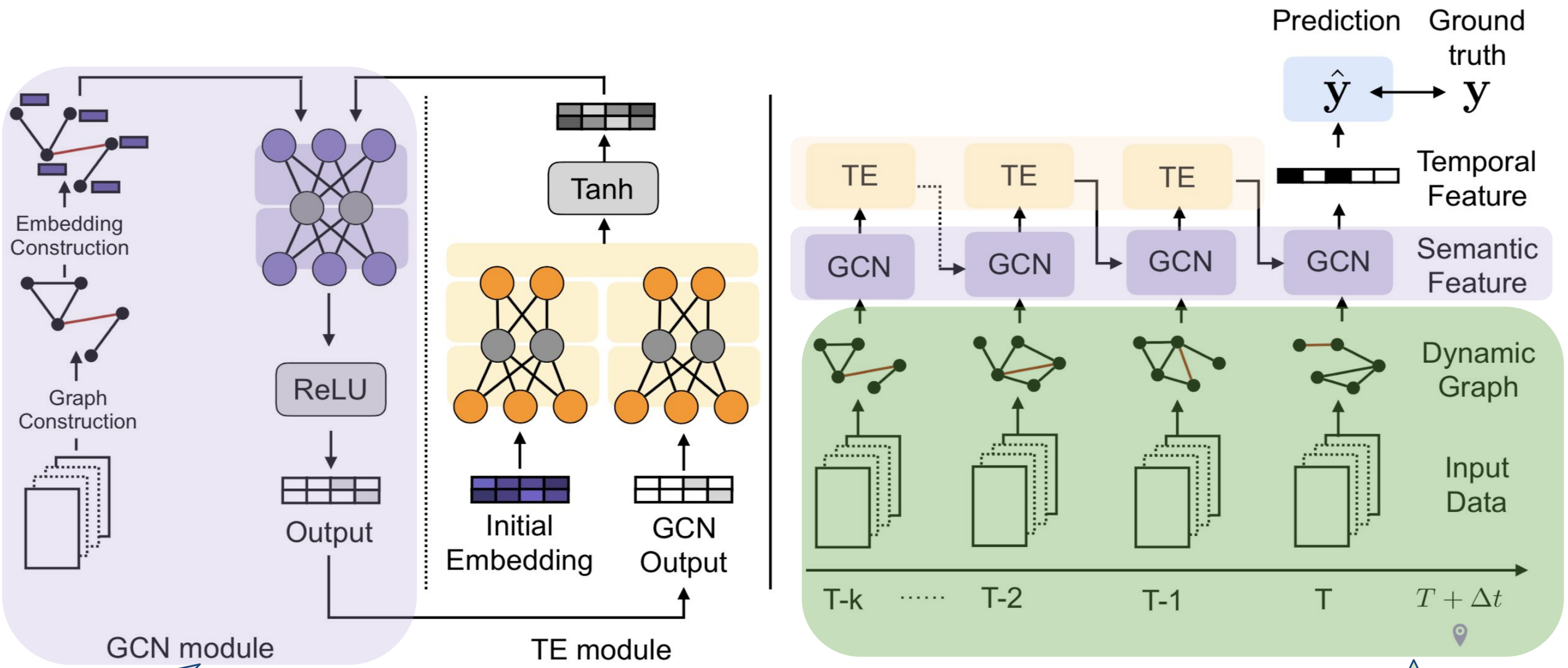


DynamicGCN: model framework



Encode data

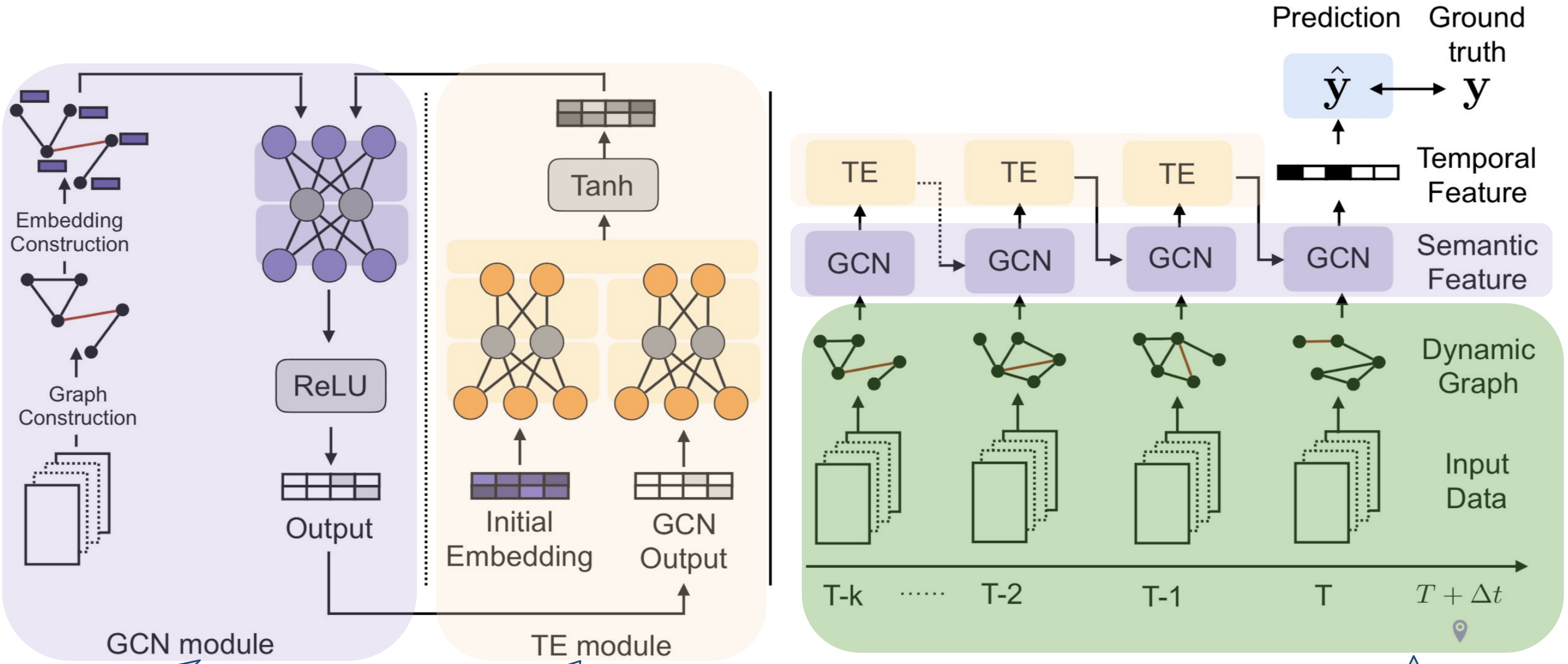
DynamicGCN: model framework



Learn the node representation by involving the semantic information from neighbors

Encode data

DynamicGCN: model framework

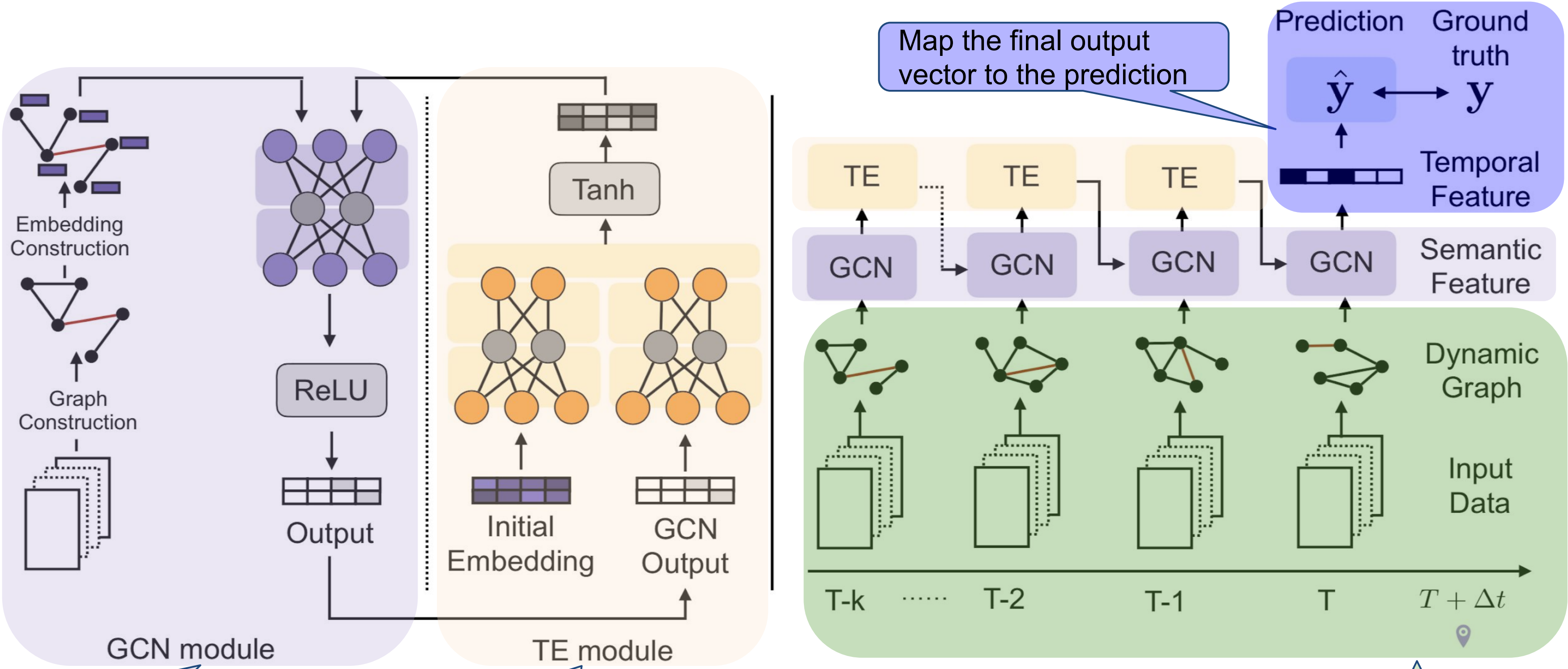


Learn the node representation by involving the semantic information from neighbors

Integrate both the semantic information and node embedding

Encode data

DynamicGCN: model framework



Learn the node representation by involving the semantic information from neighbors

Integrate both the semantic information and node embedding

Encode data

DynamicGCN: experimental evaluation

	Thailand		Egypt		India		Russia	
	F1	Rec.	F1	Rec.	F1	Rec.	F1	Rec.
LR-Count	0.77	0.713	0.794	0.747	0.618	0.559	0.739	0.721
LR-word	0.715	0.634	0.78	0.751	0.543	0.433	0.705	0.689
LR-NGram	0.7293	0.6535	0.761	0.7039	0.552	0.441	0.714	0.714
GCN	0.761	0.758	0.849	0.816	0.653	0.627	0.784	**0.826**
nMIL	0.73	0.661	0.723	0.797	0.628	**0.719**	0.76	0.769
GCN+GRU	<u>0.782</u>	0.769	0.85	0.825	<u>0.655</u>	0.621	0.787	<u>0.809</u>
GCN+LSTM	0.781	<u>0.77</u>	<u>0.851</u>	<u>0.827</u>	0.649	0.614	0.786	0.791
GCN+RNN	0.757	0.755	<u>0.851</u>	0.82	0.642	0.602	<u>0.787</u>	<u>0.809</u>
Ours	**0.797**	**0.773**	**0.862**	**0.829**	**0.669**	<u>0.627</u>	**0.804**	0.799

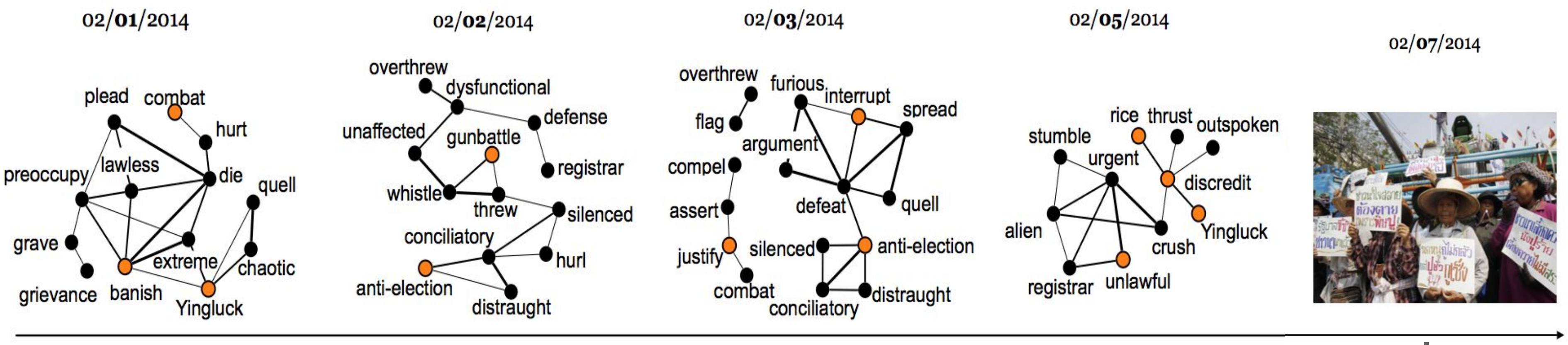
Non temporal: LR-Count, LR-word, LR-NGram, GCN

Temporal: nMIL, GCN+GRU, GCN+LSTM, GCN+RNN, Ours

Data:
Integrated Crisis Early Warning System (ICEWS) Dataverse

DynamicGCN: A Case Study

Context subgraphs generated from the train model.



Violence grips Thai capital on eve of vote called by Yingluck.

Thailand started voting.Voters blocked by anti-election groups squared off with scuffles and hurled objects.

Election Commission asked the national police chief to maintain law and order. Thai Protests Disrupt Vote.

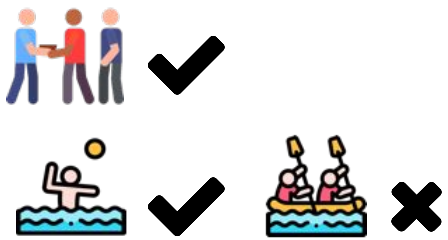
Yingluck's former commerce ministers were suspected of being involved in improper rice deals.

A possible fraud involving **rice** traders and some **politicians.**

The election related to **Yingluck** was ever **interrupted**

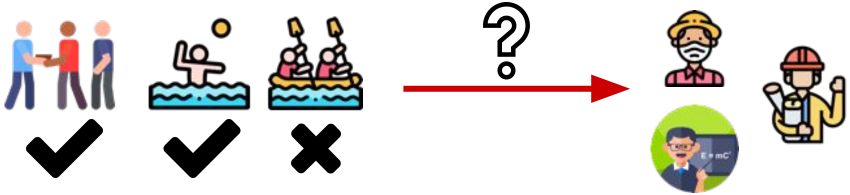
Dynamic Knowledge Graph based Multi-Event Forecasting [Deng et al. KDD20]

- Existing approaches for predicting:
  - an event type [Deng *et al*. KDD19];
  - an event subtype [Gao *et al*, AAAI19];

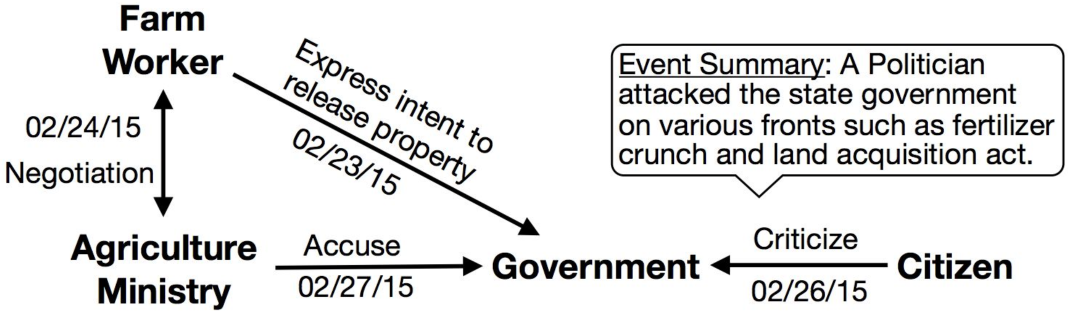


➡ Limitations:

- Unable to identify concurrent events of multiple types,
- and event participants/actors.



Dynamic Knowledge Graph based Multi-Event Forecasting [Deng et al. KDD20]

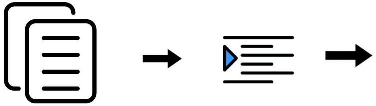


- Design a novel multi-event multi-actor forecasting framework
- Introduce an encoding method for integrating both dynamic event graphs and text data into graph-based relational features.
- Propose a context-aware embedding fusion method, which incorporates attention to fuse semantic features of words with entities and event types.

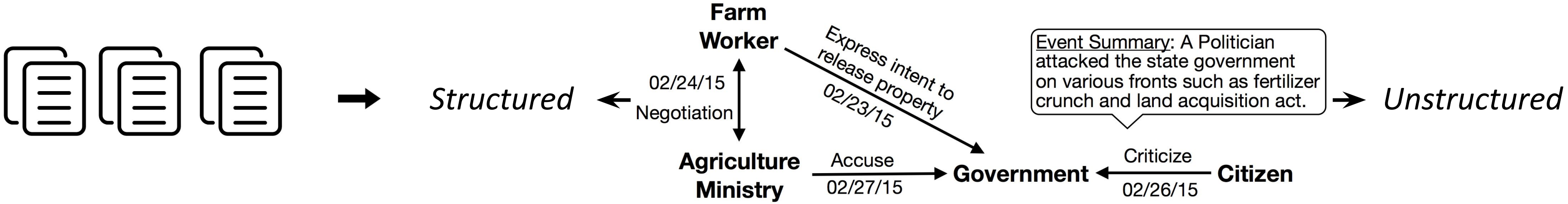
Problem Formulation

- Problems
  - Multi-event prediction $\mathbb{P}(\mathbf{y}_t | X_{1:t-1})$
  - Multi-actor prediction $\mathbb{P}(\mathbf{a}_t | y_t, X_{1:t-1})$

- Event data



	Structured	**Unstructured**
	(Citizen, Criticize, Government, 02/26/2015)	"A politician attacked… "
	Subject *Event type* *Object* *Time*	*Event summary*



Methodology

- Data Encoding
  - Event graph (built from structured data)
    - Multi-relational, directed graph with time-stamped edges
    - Node: entity
    - Edge: event type

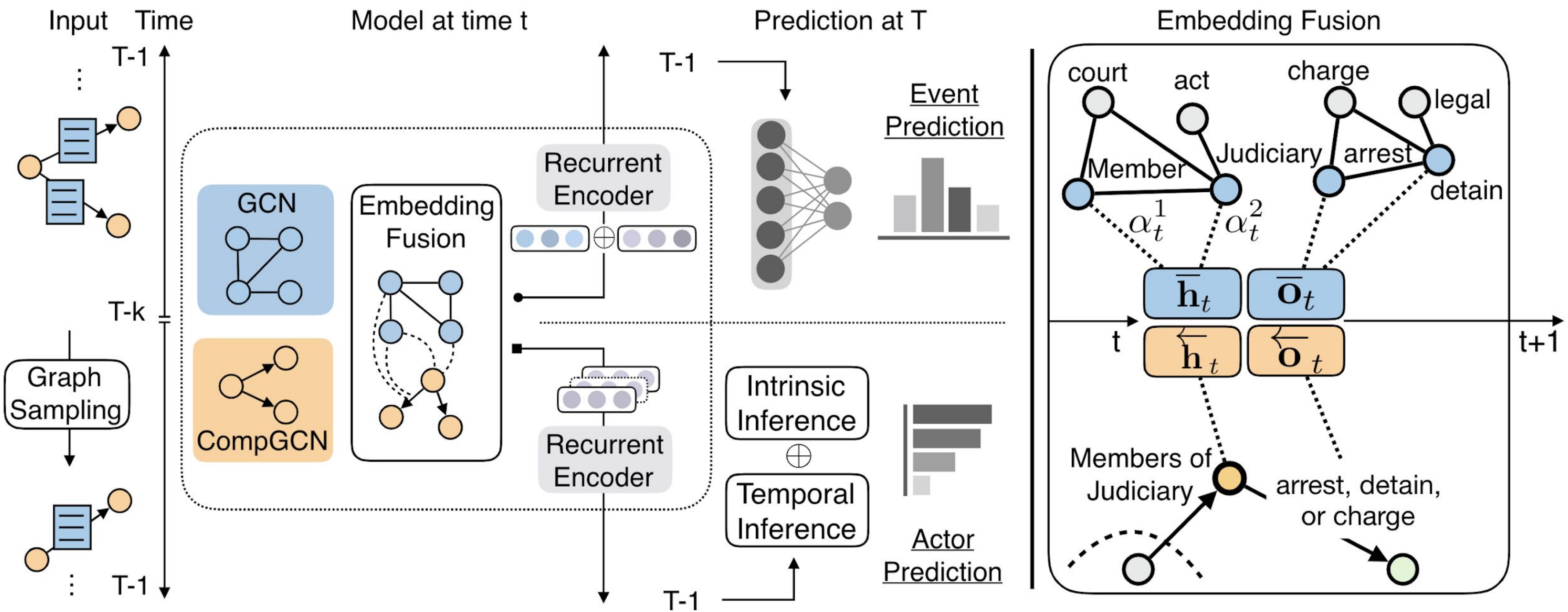
$$\text{citizen} \xrightarrow[\text{day t}]{\text{criticize}} \text{government}$$

  - Word graph (built from unstructured data)
    - Undirected and weighted graph
    - Node: word
    - Edge weight: PMI

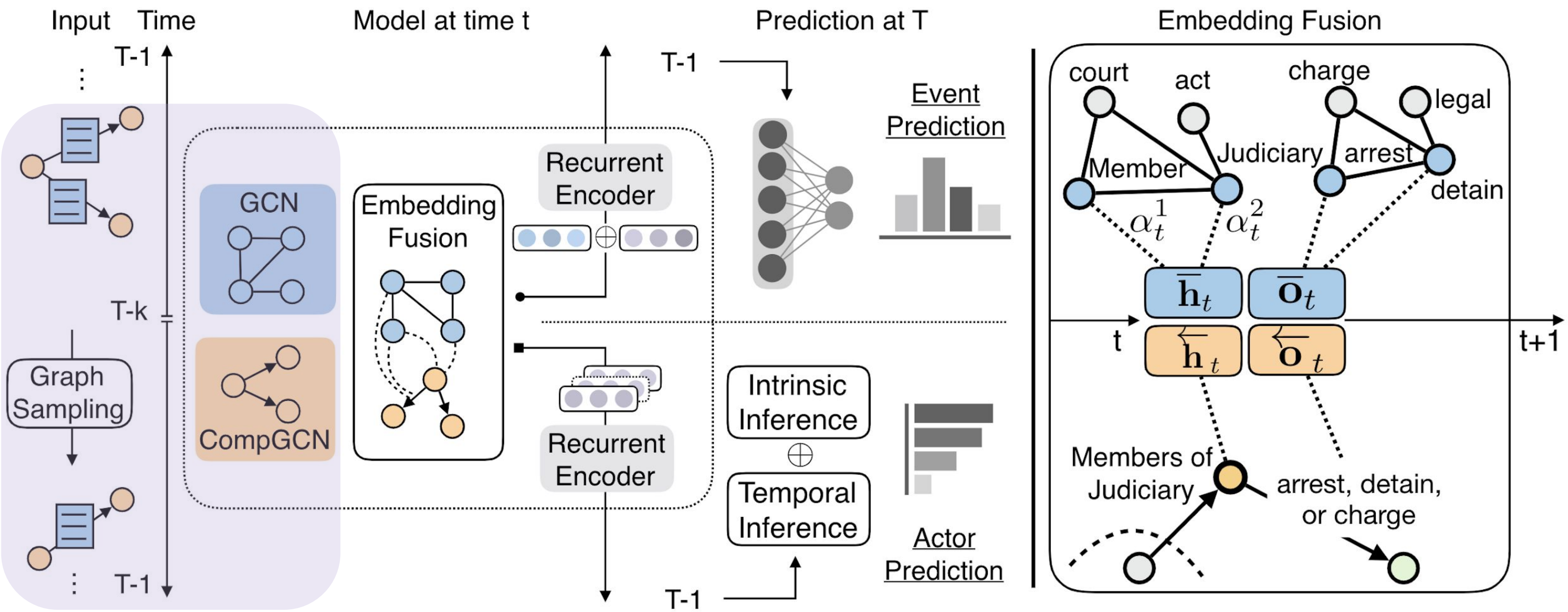
$$\text{attack} \overset{0.3}{\text{———}} \text{farmer}$$

- Model Training

Glean: Model Framework

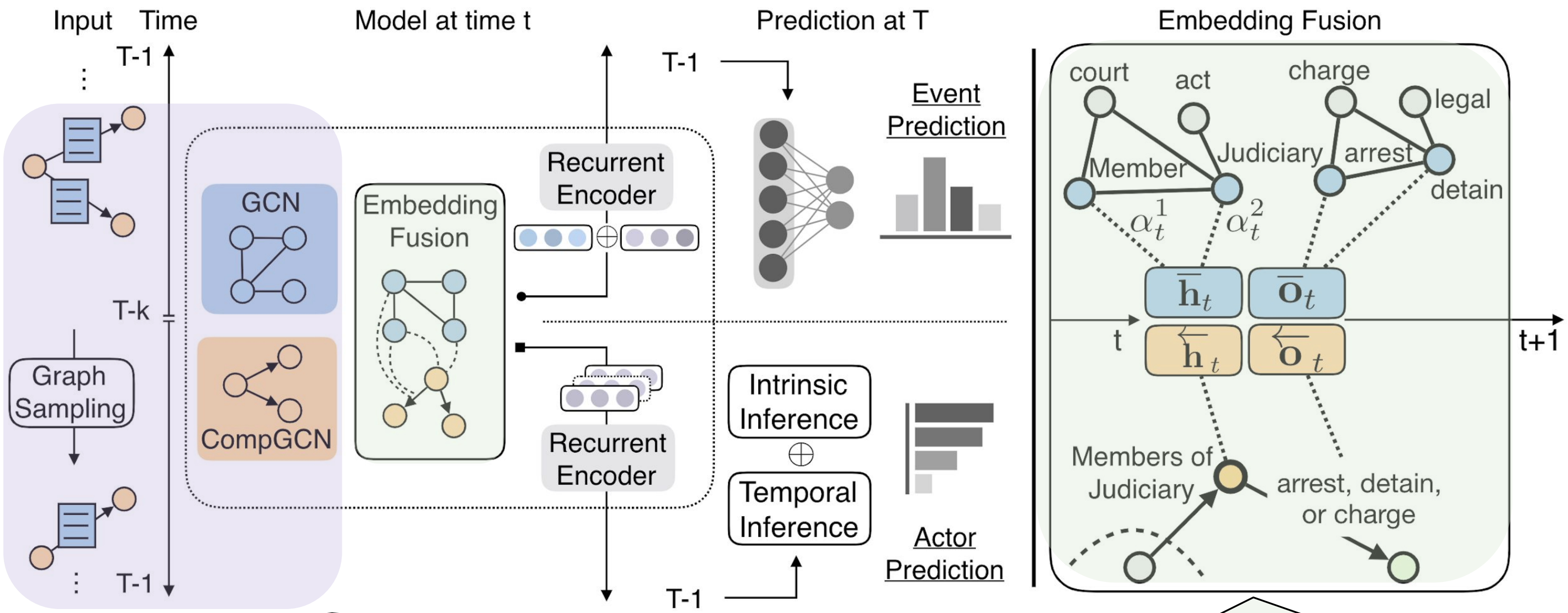


Glean: Model Framework



Encoding data into graphs;
Learn semantic and relational embedding

Glean: Model Framework



Encoding data into graphs;
Learn semantic and relational embedding

Fuse different embeddings

Context-aware Embedding Fusion

An event: (**Citizen**, *Criticizes*, **Government**, 02/26/2015) "A Politician attacked the state government on various fronts such as <u>fertilizer crunch</u> and <u>land acquisition act</u>."

Context-aware Embedding Fusion

An event: (**Citizen**, *Criticizes*, **Government**, 02/26/2015) "A Politician attacked the state government on various fronts such as <u>fertilizer crunch</u> and <u>land acquisition act</u>."

Relational embedding learned from event graphs.

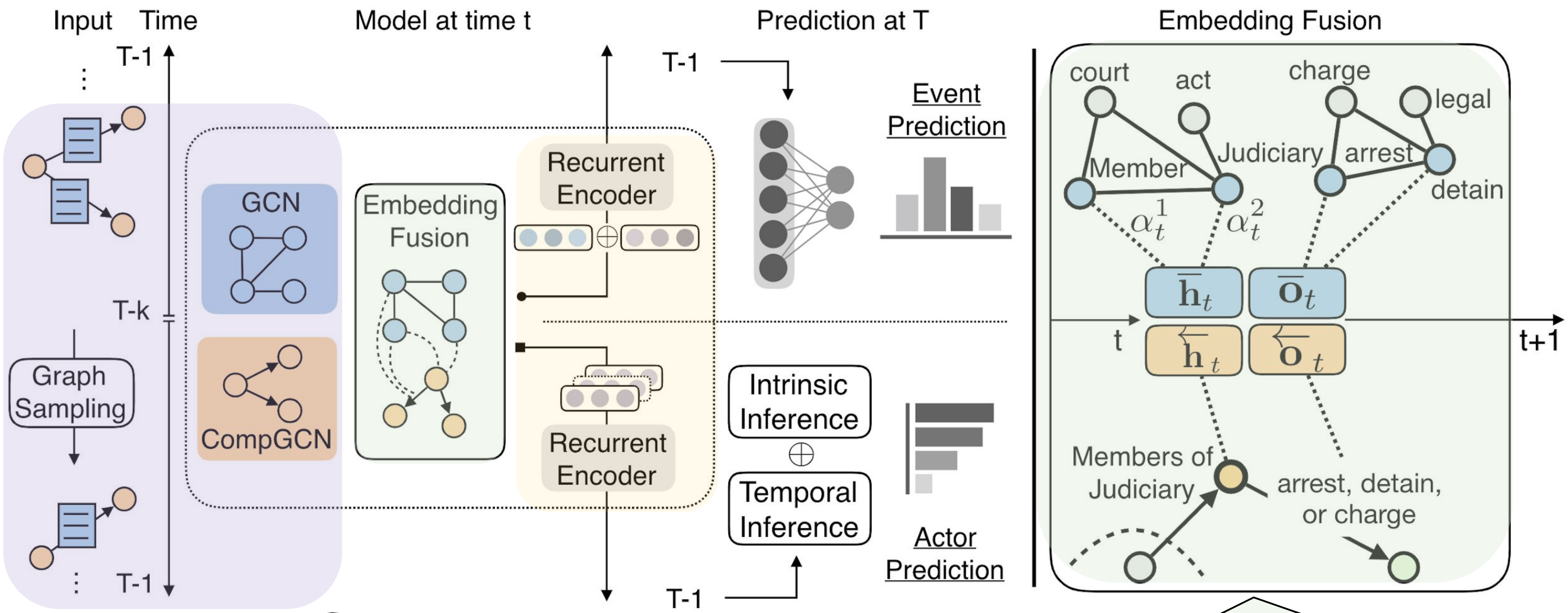
Semantic embedding learned from word graphs

$$\alpha_{t,(i,\omega)} = \frac{\exp\left(\mathrm{Attn}\left(\overleftarrow{\mathbf{h}}_{t,(i)}, \bar{\mathbf{h}}_{\omega}\right)\right)}{\exp\left(\sum_{\varphi \in \mathcal{W}_i} \mathrm{Attn}\left(\overleftarrow{\mathbf{h}}_{t,(i)}, \bar{\mathbf{h}}_{\varphi}\right)\right)} \in \mathbb{R}, \quad (7)$$

$$\mathbf{h}^{\spadesuit}_{t,(i)} = \tanh\left(\mathbf{W}_{\alpha}^{\top} \cdot \left[\underbrace{\overleftarrow{\mathbf{h}}_{t,(i)}}_{\text{rel.}}; \sum_{\omega \in \mathcal{W}_i} \underbrace{\alpha_{t,(i,\omega)} \bar{\mathbf{h}}_{\omega}}_{\text{semantic}}\right]\right) \in \mathbb{R}^d, \quad (8)$$

Fused embedding that enhances the information of entities and event types from words.

Glean: Model Framework



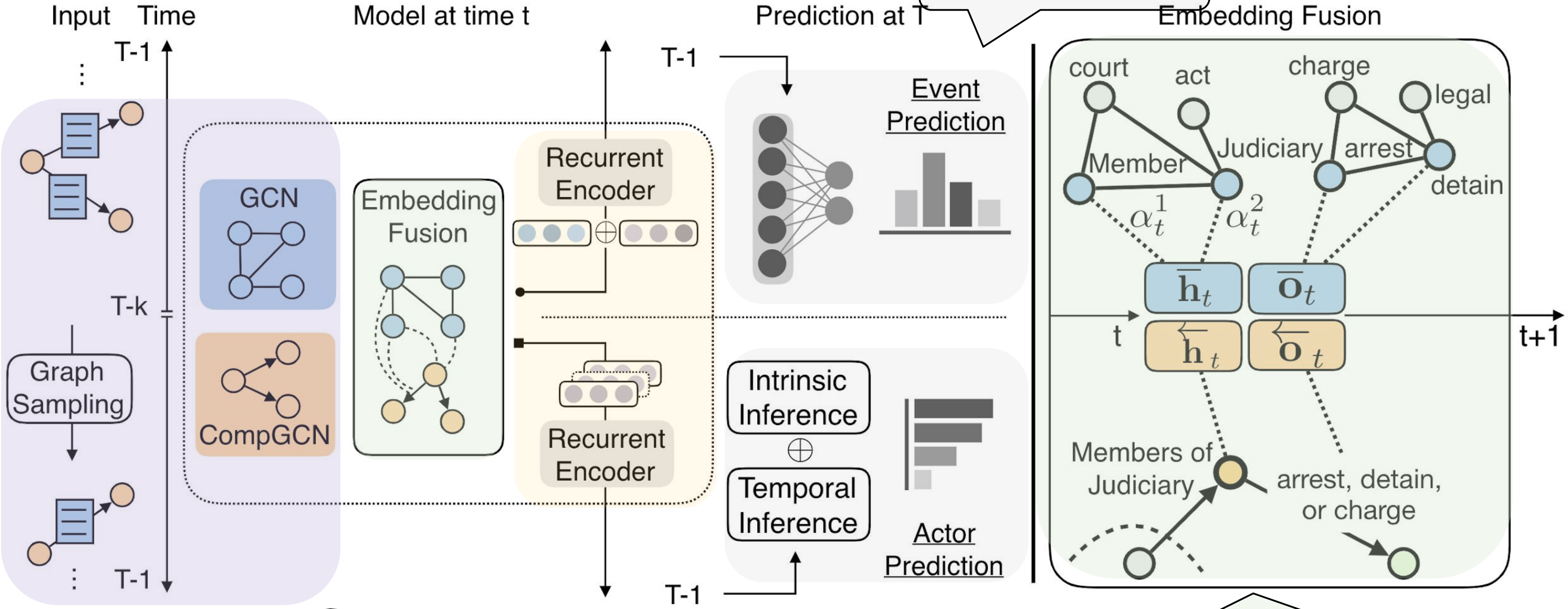
Encoding data into graphs; Learn semantic and relational embedding

The recurrent encoder models temporal information for final predictions.

Fuse different embeddings

Glean: Model Framework

Prediction model



Encoding data into graphs; Learn semantic and relational embedding

The recurrent encoder models temporal information for final predictions.

Fuse different embeddings

Glean: Experimental Evaluation

Multi-event prediction

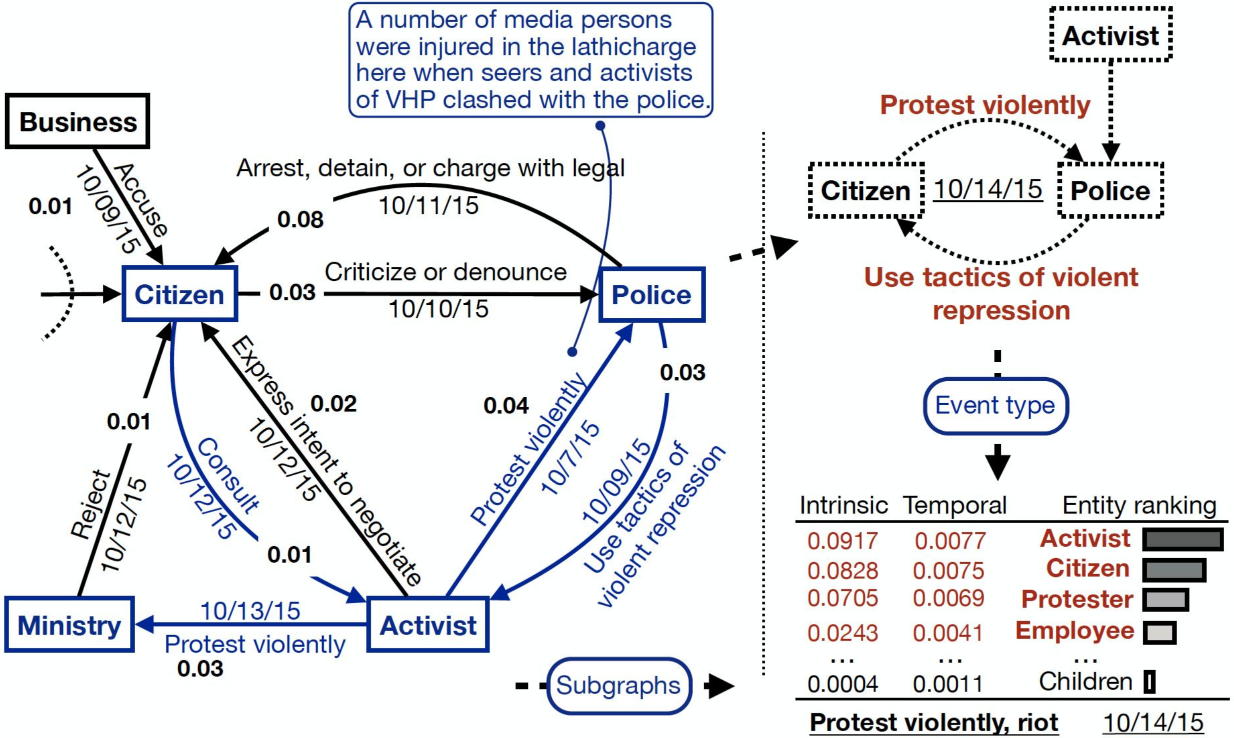
	India			Russia			Nigeria			Afghanistan			Iran		
	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall
DNN	52.49	54.65	56.38	53.81	58.44	62.61	53.54	60.64	67.70	55.77	61.80	68.14	57.54	61.85	66.19
MLkNN	52.33	54.27	55.77	51.38	55.29	58.62	26.92	28.10	28.97	45.43	48.10	50.35	53.86	56.68	59.01
BRkNN	50.36	53.05	56.00	47.46	51.53	56.64	42.48	47.28	52.45	49.89	54.98	61.52	48.56	52.24	56.77
MLARAM	33.68	33.93	34.10	25.67	26.27	26.71	41.78	45.56	48.80	33.84	34.66	35.26	27.46	27.71	27.88
DynGCN	41.80	42.57	43.19	52.81	56.77	60.14	46.27	54.65	54.65	50.05	53.97	57.75	54.22	56.93	59.21
T-GCN	60.73	64.14	67.20	56.36	61.86	67.66	56.06	63.88	72.19	60.04	67.82	76.93	61.65	67.35	73.77
RENET[1]	55.10	57.26	58.99	54.47	58.98	63.02	53.47	60.07	66.54	55.07	60.60	66.32	58.89	63.41	68.09
RENET[2]	58.44	61.46	64.18	55.85	60.86	65.66	56.44	64.37	72.82	60.58	68.47	77.75	61.66	67.24	73.52
Glean$_{-fusion}$	65.91	70.87	75.80	**58.92**	65.60	73.47	58.13	66.95	77.07	62.28	71.14	82.36	63.84	70.78	79.60
Glean	**66.69**	**71.95**	**77.31**	**58.92**	**65.64**	**73.57**	**58.76**	**68.13**	**79.49**	**62.48**	**71.43**	**82.84**	**64.12**	**71.25**	**80.46**
% relative gain	9.8%	10.9%	15.0%	4.5%	6.1%	8.7%	4.8%	5.8%	10.1%	3.1%	4.3%	6.5%	4.0%	5.8%	9.1%

Multi-actor prediction

	India			Russia			Nigeria			Afghanistan			Iran		
	H @ 1	3	10	1	3	10	1	3	10	1	3	10	1	3	10
DNN	2.09	11.01	33.87	1.46	9.72	36.40	5.10	17.06	**43.35**	8.55	17.42	35.32	10.71	19.48	26.50
RENET[3]	8.87	21.57	39.85	16.52	22.31	40.21	4.02	11.53	26.95	7.28	18.65	37.44	12.81	18.36	37.44
tRGCN	9.74	22.74	41.04	18.83	30.79	44.62	6.73	15.17	31.69	9.58	24.14	49.17	12.93	22.26	34.98
tCompGCN	9.62	21.91	40.53	18.27	30.20	44.79	6.50	14.95	31.06	9.64	23.67	49.04	12.79	21.43	34.88
Glean$_{-temp}$	13.39	24.50	43.68	18.24	31.15	43.27	6.16	14.41	26.98	9.21	22.27	47.03	11.01	17.96	29.87
Glean$_{-fusion}$	13.95	27.03	**45.73**	20.25	**34.64**	**48.10**	7.63	**18.06**	35.84	12.28	29.82	**56.89**	14.27	**24.41**	39.74
Glean	**14.01**	**27.17**	**45.73**	**20.49**	34.36	**48.10**	**7.66**	18.03	35.85	**12.29**	**30.04**	56.74	**14.31**	24.27	**39.75**
% relative gain	4.6%	10.9%	4.7%	8.8%	11.2%	7.4%	13.8%	5.9%	-	27.5%	24.4%	15.7%	10.7%	9.7%	6.2%

Glean: Case study 1

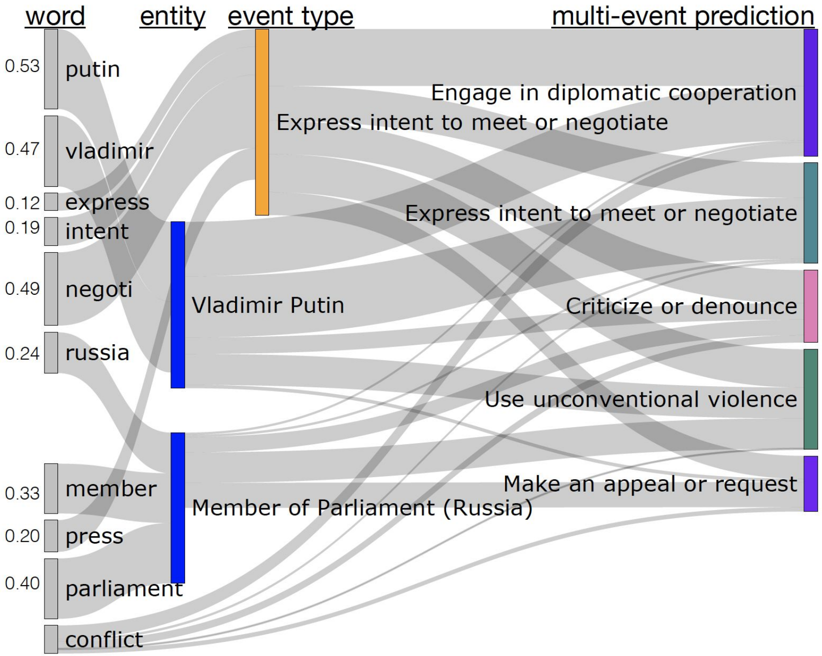
Identifying important historical events



- The red font indicates the model prediction.
- The blue part represents the subgraph sampled for actor prediction.

Glean: Case study 2

Identify Semantic Contexts and Feature Flows

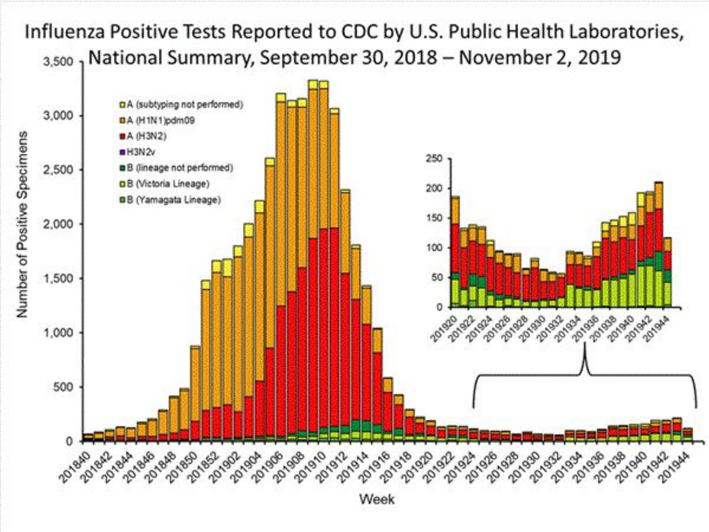


- ➢ Attention scores quantify the importance of words in contributing the entity/event type embedding for prediction
- ➢ *parliament* contributes more than *member* in *Member of Parliament*

Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction

[Deng et al. CIKM20]

- The proposed model:
  a graph-based deep learning framework with time series attributes for each node to study the spatio-temporal influence of long-term ILI predictions.



Existing work

- **Traditional causal models** [Bisset *et al*. ICS09]
  - compartmental models and agent-based models, employ disease progression mechanisms such as Susceptible-Infectious-Recovered (SIR) to capture the dynamics of ILI diseases.
- **Statistical models such as Autoregressive**
  - autoregressive (AR) and its variants (e.g., VAR)
- **Deep learning methods** [Venna et al. IEEE18]
  - recurrent neural networks, convolutional neural networks, etc.

Challenges

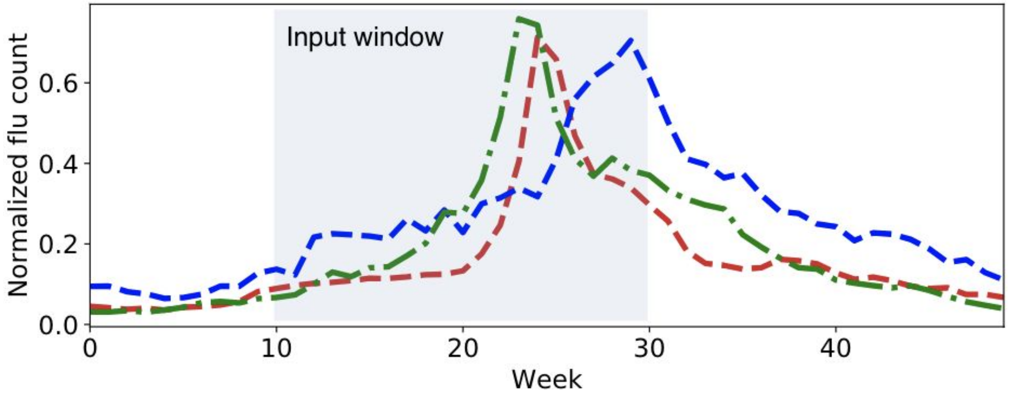
- The temporal dependency is hard to capture with short-term input data.
- The influence of other locations often changes over time.
- Adequate data sources are required to achieve decent performance.

Problem Formulation

A graph-based propagation model

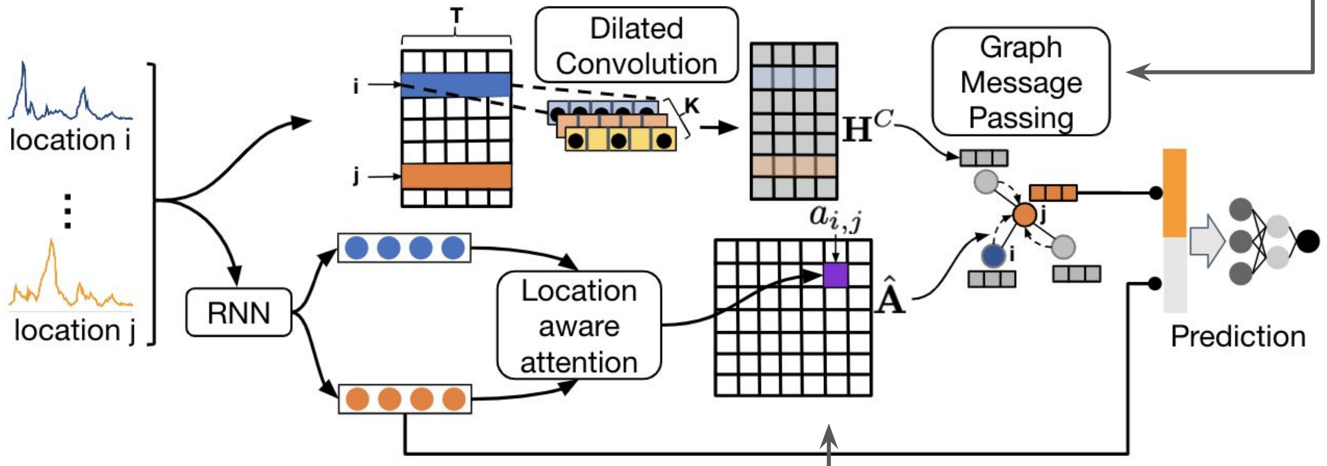
- N **locations.** Each location (e.g., a state) is a node; associated with a time series input for a window T
- **Input**: ILI patient counts for T weeks
- **To predict** the ILI patient counts at a future time point T+h

h refers to the horizon/lead time of the prediction



Cola-GNN: Model Framework

3. Graph Message Passing
   - use graph neural networks



2. Multi-Scale Dilated Convolution
   - encode local patterns with short-term and long-term trends by employing dilated convolution layers.

1. Directed Spatial Influence Learning
   - dynamically model the impact of one location on other locations during the epidemics.

Cola-GNN: Experimental Evaluation

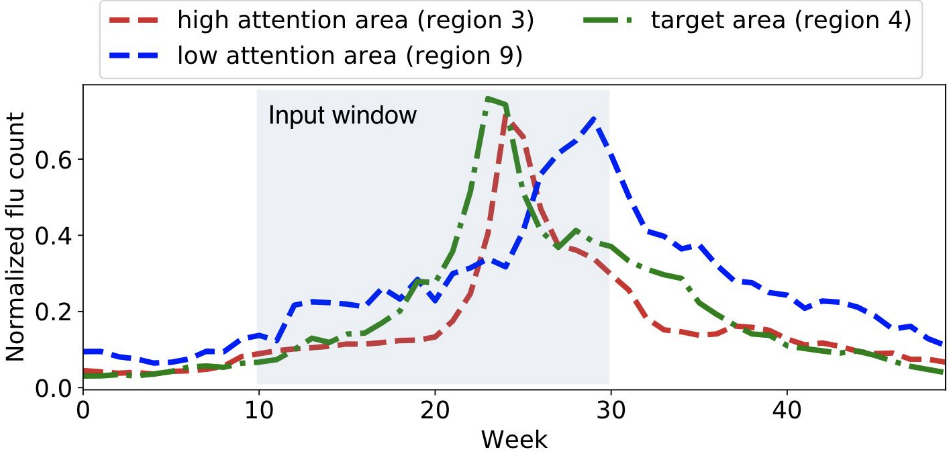
Short-term leadtime=2,3

long-term leadtime=5,10,15

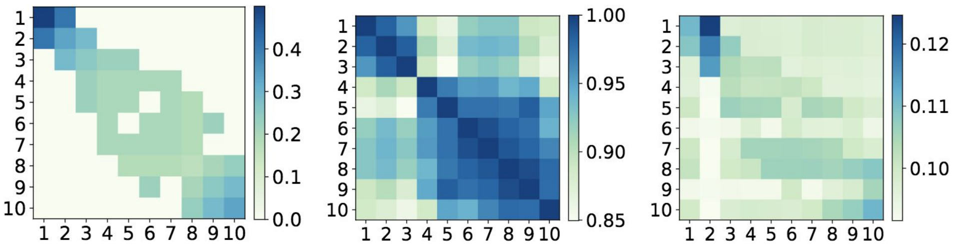
	Japan-Prefectures					US-Regions					US-States				
RMSE(↓)	2	3	5	10	15	2	3	5	10	15	2	3	5	10	15
GAR	1232	1628	1988	2065	2016	536	715	991	1377	1465	150	187	236	314	340
AR	1377	1705	2013	2107	2042	570	757	997	1330	1404	161	204	251	306	327
VAR	1361	1711	2025	1942	1899	741	870	1059	1270	1299	290	276	295	324	352
ARMA	1371	1703	2013	2105	2041	560	742	989	1322	1400	161	200	250	306	326
RNN	1001	1259	1376	1696	1629	513	689	896	1328	1434	149	181	217	274	315
LSTM	1052	1246	1335	1622	1649	507	688	975	1351	1477	150	180	213	276	307
RNN+Attn	1166	1572	1746	1612	1823	613	753	1065	1367	1368	152	186	234	315	334
DCRNN	1502	1769	2024	2019	1992	711	874	1127	1411	1434	165	209	244	299	298
CNNRNN-Res	1133	1550	1942	1865	1862	571	738	936	1233	1285	205	239	267	260	250
LSTNet	1133	1459	1883	1811	1884	554	801	998	1157	1231	199	249	299	292	292
ST-GCN	996	1115	1129	1541	1527	697	807	1038	1290	1286	189	209	256	289	292
Cola-GNN	**929**	**1051**	**1117**	**1372**	**1475**	**480**	**636**	**855**	**1134**	**1203**	**136**	**167**	**202**	**241**	**237**
% relative gain	6.7%	5.7%	1.1%	11.0%	3.4%	5.3%	7.6%	4.6%	2.0%	2.3%	8.7%	7.2%	5.2%	7.3%	5.2%
PCC(↑)	2	3	5	10	15	2	3	5	10	15	2	3	5	10	15
GAR	0.804	0.626	0.339	0.288	0.470	0.932	0.881	0.790	0.581	0.485	0.945	0.914	0.875	0.777	0.742
AR	0.752	0.579	0.310	0.238	0.483	0.927	0.878	0.792	0.612	0.527	0.940	0.909	0.863	0.773	0.723
VAR	0.754	0.585	0.300	0.426	0.474	0.859	0.797	0.685	0.508	0.467	0.765	0.790	0.758	0.709	0.653
ARMA	0.754	0.579	0.310	0.253	0.486	0.927	0.876	0.792	0.614	0.520	0.939	0.909	0.862	0.773	0.725
RNN	0.892	0.833	0.821	0.616	0.709	0.940	0.895	0.821	0.587	0.499	0.948	0.922	0.886	0.821	0.758
LSTM	0.896	0.873	0.853	0.681	0.695	0.943	0.895	0.812	0.586	0.488	0.948	0.922	0.889	0.820	0.771
RNN+Attn	0.850	0.668	0.590	0.741	0.522	0.887	0.859	0.752	0.554	0.552	0.947	0.922	0.884	0.780	0.739
DCRNN	0.697	0.537	0.292	0.342	0.525	0.897	0.849	0.760	0.604	0.558	0.941	0.886	0.886	0.829	0.837
CNNRNN-Res	0.852	0.673	0.380	0.438	0.467	0.920	0.862	0.782	0.552	0.485	0.904	0.860	0.822	0.820	0.847
LSTNet	0.846	0.728	0.432	0.518	0.515	0.935	0.868	0.746	0.609	0.533	0.913	0.850	0.759	0.760	0.802
ST-GCN	0.902	0.880	0.872	0.735	**0.773**	0.879	0.840	0.741	0.644	0.619	0.907	0.778	0.823	0.769	0.774
Cola-GNN	**0.915**	**0.901**	**0.890**	**0.813**	0.753	**0.946**	**0.909**	**0.835**	**0.717**	**0.639**	**0.955**	**0.933**	**0.897**	**0.822**	**0.856**
% relative gain	1.4%	2.4%	2.1%	9.7%	-	0.6%	1.6%	1.7%	10.2%	3.2%	0.7%	1.2%	0.9%	0.1%	1.1%

Cola-GNN: A Case Study

Predict the patient count for region 4



Comparison of predefined matrix and attention matrix.



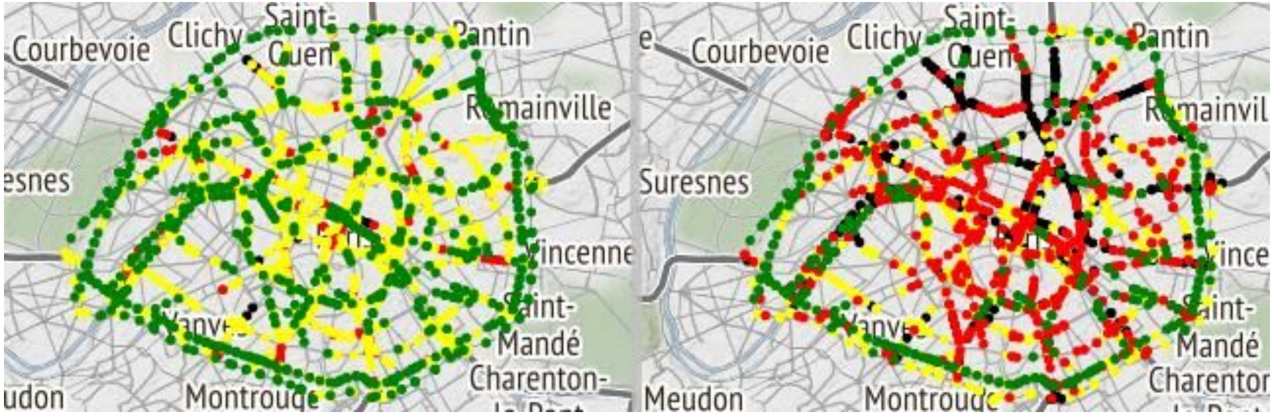
(a) Geolocation (b) Correlation (c) Attention

Applications Based on Temporal Event Modeling

Traffic Prediction

- Predict traffic volumes, utilising historical speed and volume data.

image source https://github.com/raphaelauv/Paris-Traffic-Prediction



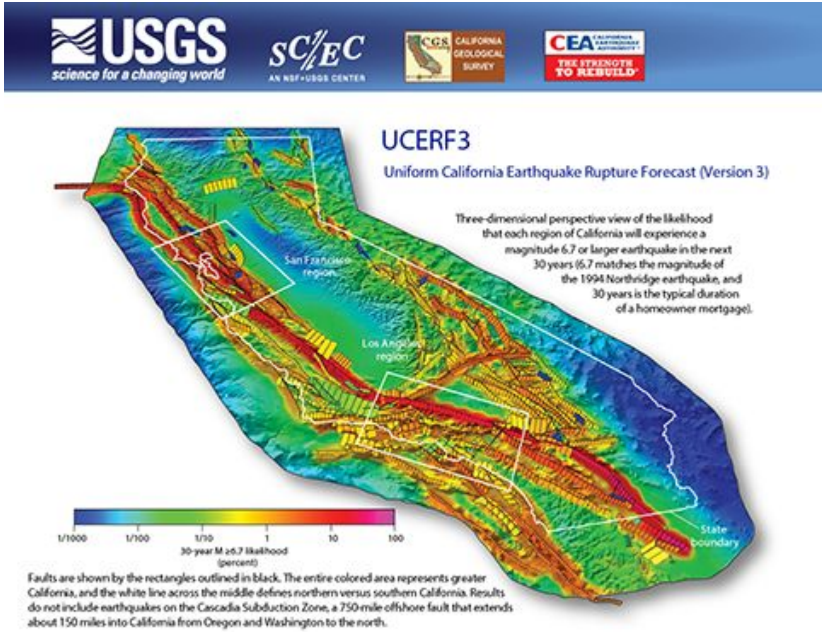
[1] Lv, Yisheng, et al. "Traffic flow prediction with big data: a deep learning approach." IEEE Transactions on Intelligent Transportation Systems 16.2 (2014): 865-873.
[2] Yao, Huaxiu, et al. "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.
[3] Guo, Shengnan, et al. "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.

Applications Based on Temporal Event Modeling

Earthquake Prediction

- Predict the time, location, and magnitude of future earthquakes.

The epidemic type aftershock sequence (ETAS) Model



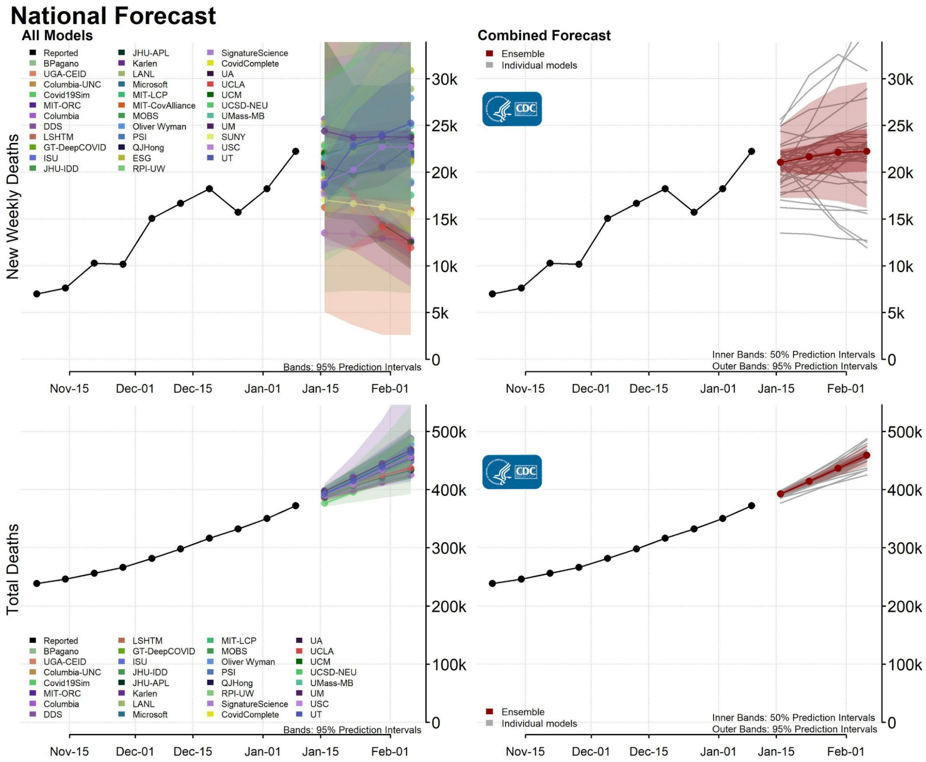
[1] http://www.wgcep.org/UCERF3
[2] Bansal, A.R., Dimri, V.P. & Babu, K.K. Epidemic type aftershock sequence (ETAS) modeling of northeastern Himalayan seismicity. J Seismol 17, 255–264 (2013). https://doi.org/10.1007/s10950-012-9314-7
[3] Sun, Han. A Data-driven Building Seismic Response Prediction Framework: from Simulation and Recordings to Statistical Learning. Diss. UCLA, 2019.

Applications Based on Temporal Event Modeling

Epidemic Prediction

- Predict the time, peak and intensity of an infectious disease at a certain location in the future.

Forecasting of new reported COVID-19 cases over next 4 weeks.

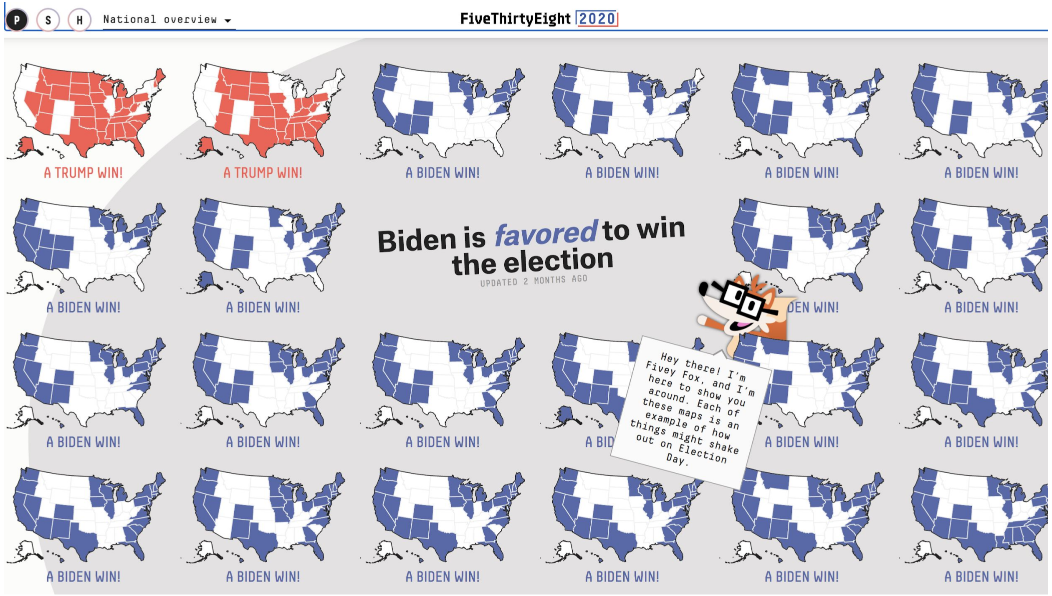


[1] https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

Applications Based on Temporal Event Modeling

Election Prediction

Forecasting the 2020 presidential election between President Donald Trump and Joe Biden by FiveThirtyEight

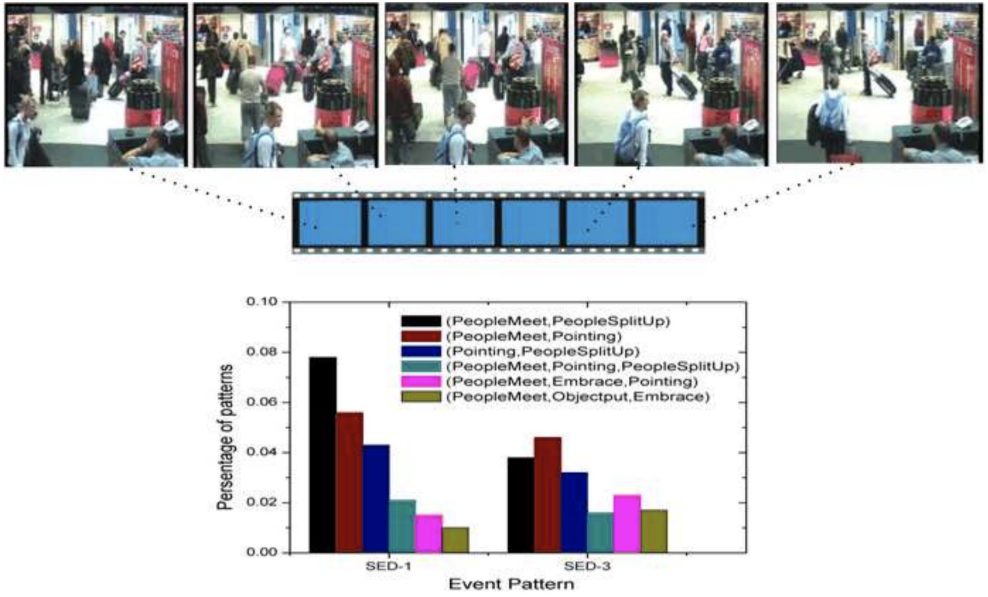


[1] https://projects.fivethirtyeight.com/2020-election-forecast/
[2] Dwi Prasetyo, Nugroho, and Claudia Hauff. "Twitter-based election prediction in the developing world." Proceedings of the 26th ACM Conference on Hypertext & Social Media. 2015.

Applications Based on Temporal Event Modeling

Video Event Detection

- Identifying the temporal range of an event in a video (i.e. when) and sometimes the location of the event as well (i.e. where).



[1] Cheng, Yu, et al. "Temporal sequence modeling for video event detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

Applications Based on Temporal Event Modeling

Civil Unrest Forecasting

- predicting the occurrence of a future protest within a target city using open source data.

EMBERS

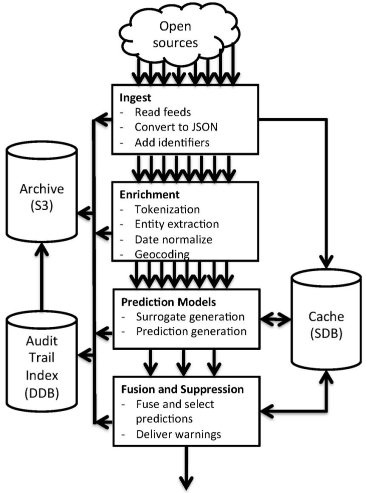
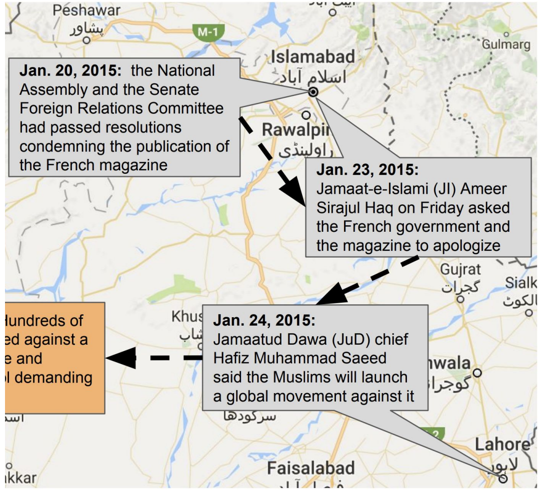


Figure 2: EMBERS system architecture



[1] Ramakrishnan, Naren, et al. "'Beating the news' with EMBERS: forecasting civil unrest using open source indicators." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014.
[2] Ning, Yue, et al. "STAPLE: Spatio-Temporal Precursor Learning for Event Forecasting." Proceedings of the 2018 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2018.

Conclusion and Future Directions

- **Introduction and motivation for event predictions**
  - Definitions and challenges
- **Precursor Identification for Interpretable Event Forecasting**
  - Representation Learning and Deep Learning
    - *to automatically encode raw input and learn hidden features*
  - Multi-Instance Learning
    - *Identify key characteristics in semi-supervised event modeling*
  - Multi-Task Learning
    - *to infer relationships across different tasks (locations)*
- **Event Graphs for Interpretable Event Forecasting**
  - Graph Neural Network
    - *To model associations among words in word graphs, entities in knowledge graphs, and locations in geographic networks*

Conclusion and Future Directions

Future directions

- ***Event modeling***
  - *Data integration for multiple sources*
  - *Learning hierarchies of spatial precursors*
  - *Semantic encoding and optimization*
- ***Precursor Identification***
  - *Transparent event forecasting model*
  - *Model interpretation*
  - *Causal interpretation*

Thank you

If you have any questions, feel free to contact:

Songgaojun Deng: sdeng4@stevens.edu
Yue Ning: yue.ning@stevens.edu
Huzefa Rangwala: rangwala@gmu.edu