ICWSM International AAAI Conference on Web and Social Media

上海科技大学 ShanghaiTech University

Where Did the President Visit Last Week? Detecting Celebrity Trips from News Articles

Kai Peng*, **Ying Zhang***, Shuai Ling, Zhaoru Ke, and Haipeng Zhang†

彭凯*、**章颖***、凌帅、柯昭如、张海鹏†

ShanghaiTech University

上海科技大学

June 05, 2024

Paper

Code&Dataset

Me!😍

Where did Donald Trump visit on 2017-01-26?

When and **where** a **celebrity** appears means a lot

When and where a celebrity appears means a lot

- Understand Policy Tendency [Doherty2009]

Trump held campaign rally in Phoenix

- Arizona is an important swing state
- Phoenix is the largest city of Arizona

When and where a celebrity appears means a lot

- Analyze International Relations [Cavari2019]

Merkel attended EU meetings in Brussels

- Brussels is the headquarters of the EU
- Trip to Brussels bolstered Germany-EU ties

Related Work

- Existing work analyzes the travel patterns of celebrity [Doherty2009], [Deville2014], [Toplak2017]
- Available trip data is scarce
  - Limited Volume and Coverage - POTUS
  - Low Update Frequency – Once a Year (MOFA of Japan)
  - Time Range - 1906 to 2016
- There lacks work focused on extracting celebrity trip data

Problem Statement

- Celebrities' travel information often mentioned in news
- Our Goal: Extract celebrity trips from news articles

Where did Donald Trump visit on 2017-01-26?

News related to Trump
on 2017-01-26

Trip Information

Donald Trump

2017-01-26

Philadelphia

上海科技大学
ShanghaiTech University

A Plausible Intuition

Person-Related Event Detection

- Person-Related Event Detection (PRED)
    - Using social media (tweets)
    - Focus on life event[Yen2021]: marriage and graduation etc.
- Traditional ML-based PRED[Dickinson2015, Khodabakhsh2018]
    - Manual feature engineering/Ignore the order of words
- Deep Learning PRED
    - Use CNN[Nguyen2017] and Bi-LSTM[Yen2018] to model single document/sentence
    - Capture semantics within short sequence

Challenges

- Trip descriptions scattered within and across articles
  - Average of 23 articles per Trump trip

··· Donald Trump visited ··· during a retreat in Philadelphia ··· Washington D.C. ··· Trump ··· Mike Pence

- Irrelevant celebrities and locations in the news
- Implicit trip: celebrities\locations are not directly mentioned by names

Graph Neural Network

- Graph Neural Network (GNNs)
  - Used to capture long-range dependency on text tasks[Peng2018]

→ Capture the scattered semantics in our scenario (Challenge 1)

→ Text Graph → Aggregating Semantics → Graph Pooling → Classify

- Common pooling methods are global pooling, which summarizes the entire graph

Challenges

- Trip descriptions scattered within and across articles
- Irrelevant celebrities and locations in the news
  - Even worse when examining multiple articles

··· Donald Trump visited ··· during a retreat in Philadelphia ··· Washington D.C. ··· Trump ··· Mike Pence ··· Chicago ··· Jim Kenney ···

- Implicit trip: celebrities\locations are not directly mentioned by names

Word-Article Graph for Modeling Text

- Word-Article Graph

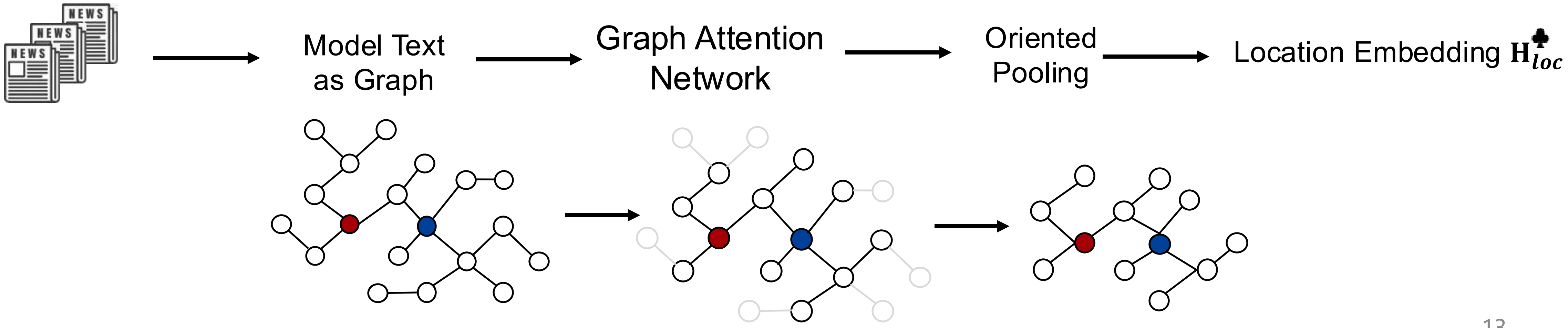
  → Challenge 1: Descriptions of trip are scattered

  - Model text as graph

- Oriented Pooling for Graph

  → Challenge 2: Interfering celebrities and locations in text

  - Sample sub-graph based on node similarity (target celebrity\location)



Challenges

- Trip descriptions scattered within and across articles
- Irrelevant celebrities and locations in the news
- Implicit trip: celebrities\locations are not directly mentioned by names

… **President of the United States** visited … during a retreat in **Philadelphia** …
Washington D.C. … Trump … **Mike Pence**

→ Introducing relationships of knowledge entity

- Pre-trained embeddings can not capture the latest status of specific event entity

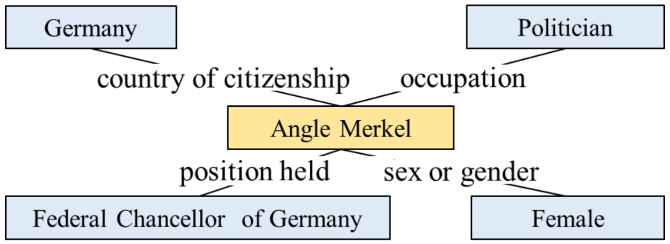
→ Introducing the sentences from recent news

Entity Sub-graph for Introducing Knowledge

- Learning Entity Relationships
  - Construct sub-graph for each entity

$$\mathrm{h_e} = \sigma\Big(\sum_{(r,u)\exists(e,r,u)\in\mathcal{G}_e} W_e \phi(h_u, \ell_r)\Big)$$

Entity $\boldsymbol{e_1}$ → Entity Graph $\mathcal{G}_e$ → **CompGCN**[Vashishth2020] → Entity Embedding $\mathbf{H}_{ent}^{\clubsuit}$



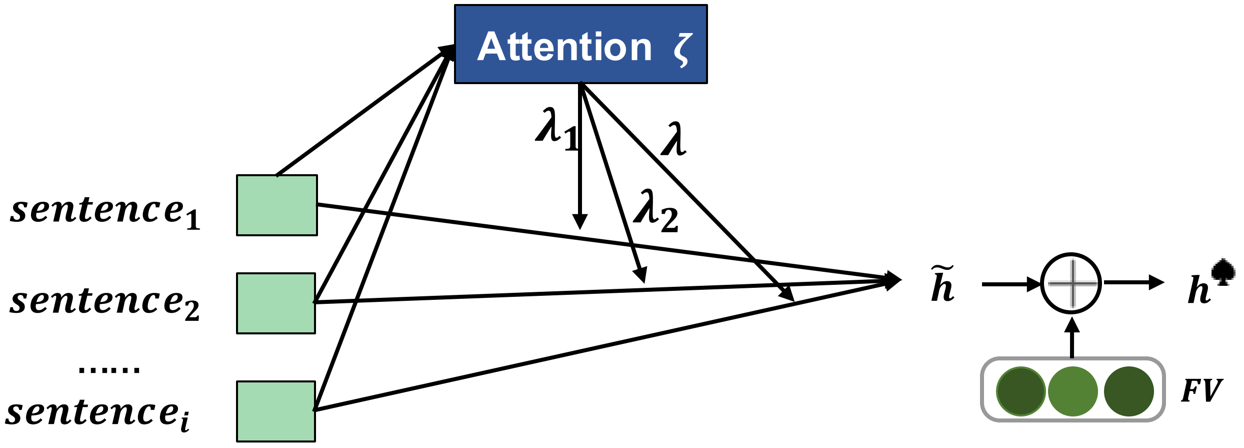
上海科技大学
ShanghaiTech University

Updating the Status of Event Entity

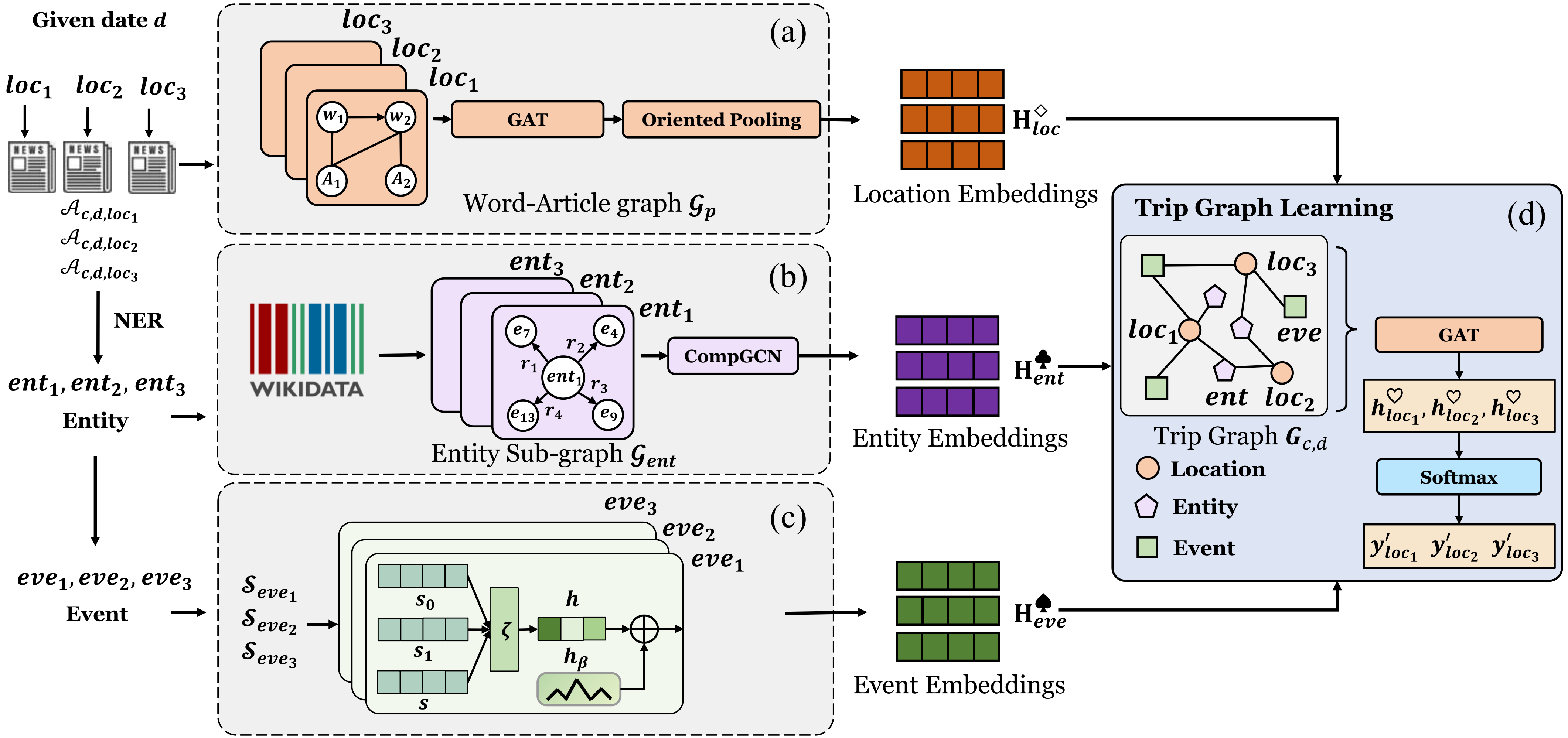
- Capturing the latest status of event
    - Sentences from news published around the target date

··· the 60th GRAMMY Awards will take place at New York City's Madison Square Garden on Sunday, Jan. 28, 2018.

Event eve_1 ⟶ Attention Layer ⟶ Event Embedding $\mathbf{H}^{\spadesuit}_{eve}$



Joint Learning on One Graph



CeleTrip

Experiment

- Ground Truth Dataset
  - Label 2,000+ positive trips/9,000+ negative trips

Celebrity	Location	Date	Articles	Label
Donald Trump	Langley	01/21/2017	$\mathcal{A}_1$,$\mathcal{A}_2$,...	Positive

Table 1

  - Split the dataset by date: 66.5% and 33.5%

Trip Dataset	Positive	Negative	Period
Train	1,715	6,341	01/01/2016 - 06/30/2019
Test	689	3,366	07/01/2019 - 02/28/2021

Table 2

Experimental Result

- Baselines from Person-Related Event Extraction

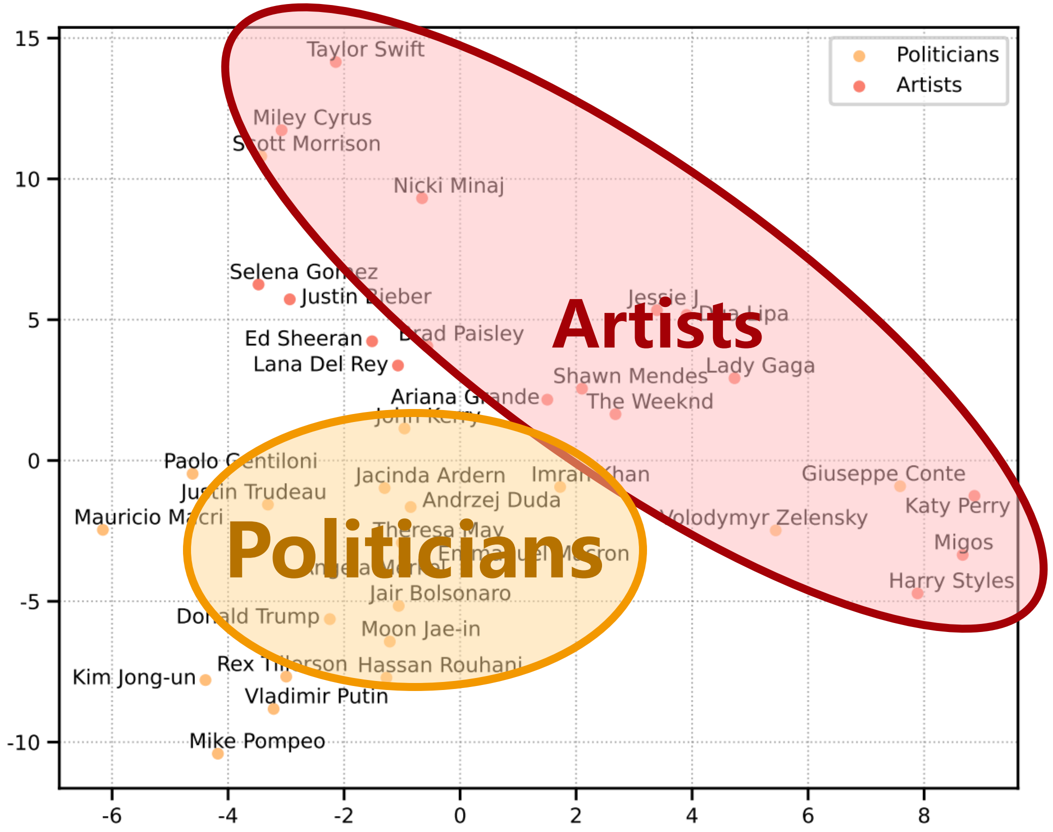
The actually visited locations that are correctly classified

The predicted trip location that are actually visited by celebrities

Methods	F1 (%)	P (%)	R (%)	Acc (%)
LocFre	37.34	39.06	35.76	81.96
LocJaccard	40.65	40.59	40.53	85.37
LR (TFIDF)	64.65	**88.64**	52.83	90.69
CNN	63.75	66.46	61.25	89.72
Bi-LSTM	66.93	71.95	62.55	90.26
GCN	60.18	60.71	59.65	86.58
CeleTrip $_{\text{w/o kn}}$	**74.11**	78.89	**69.88**	**92.04**
With External Knowledge				
LR (TFIDF)	68.50	84.14	57.76	91.29
CNN	69.85	79.08	62.55	90.83
Bi-LSTM	70.68	81.71	62.26	91.22
GCN	64.70	71.94	58.78	89.10
CeleTrip	**82.53**	**86.17**	**79.27**	**94.30**

Performance Evaluation

Learnt Representations of Celebrities

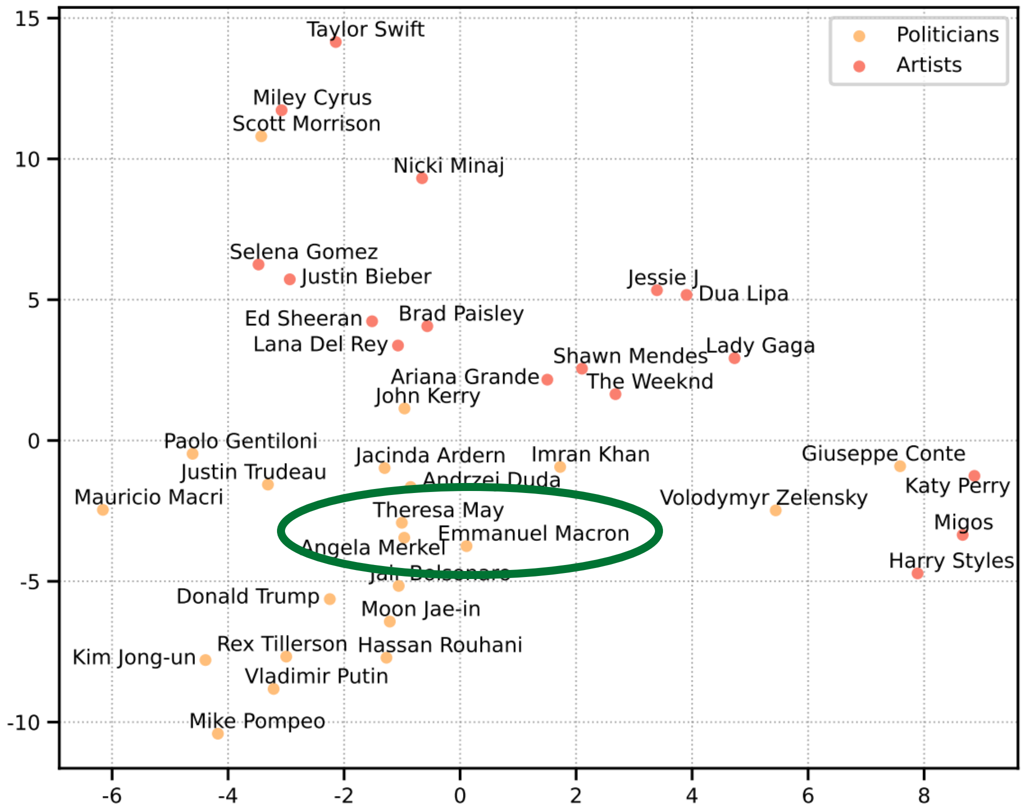


Visualization of the representations for celebrity entities (t-SNE)

- Artists in the upper-right
  - Sparse - Solo Concert
- Politicians in the lower-left
  - Clustered - International Conference

上海科技大学
ShanghaiTech University

Learnt Representations of Celebrities



Visualization of the representations for celebrity entities (t-SNE)

- Merkel, May and Macron are closer
  - Guess：Regional Meetings
- Learnt representations are suitable for trip detection task

Summary

- Propose a **task** of celebrity itinerary detection, and our framework, **CeleTrip**, for this task
- Design **Word-Article graph** and **Oriented Pooling** to refine the graph and focus on nodes of interest
- Release our labeled dataset and **open-source** our framework

Paper

Code&Dataset

Thanks.