

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

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

ShanghaiTech University

{pengkai, zhangying12022, lingshuai, kezhr, zhanghp}@shanghaitech.edu.cn

Abstract

Celebrities' whereabouts are of pervasive importance. For instance, where politicians go, how often they visit, and who they meet, come with profound geopolitical and economic implications. Although news articles contain travel information of celebrities, it is not possible to perform large-scale and network-wise analysis due to the lack of automatic itinerary detection tools. To design such tools, we have to overcome difficulties from the heterogeneity among news articles: 1) One single article can be noisy, with irrelevant people and locations, especially when the articles are long. 2) Though it may be helpful if we consider multiple articles together to determine a particular trip, the key semantics are still scattered across different articles intertwined with various noises, making it hard to aggregate them effectively. 3) Over 20% of the articles refer to celebrity trips indirectly, instead of using the exact celebrity names or location names, leading to large portions of trips escaping regular detecting algorithms. We model text content across articles related to each candidate location as a graph to better associate essential information and cancel out the noises. Besides, we design a special pooling layer based on attention mechanism and node similarity, reducing irrelevant information from longer articles. To make up the missing information resulted from indirect mentions, we construct knowledge sub-graphs for named entities (person, organization, facility, etc.). Specifically, we dynamically update embeddings of event entities like the G7 summit from news descriptions since the properties (date and location) of the event change each time, which is not captured by pre-trained event representations. The proposed CeleTrip jointly trains these modules, which outperforms all baseline models and achieves 82.53% in the F1 metric. By open-sourcing the first tool and a carefully curated dataset for such a new task, we hope to facilitate relevant research in celebrity itinerary mining as well as the social and political analysis built upon the extracted trips.

Introduction

Celebrity itineraries have received extensive attention in social science, with applications to politics (Brace and Hinckley 1993; Goldsmith, Horiuchi, and Matush 2021), economics (Boianovsky 2018), arts (Badham 2017), and cul-

tural history (Schich et al. 2014). Specifically, researchers analyze both domestic and international situations using the travels of politicians. Ostrander and Rider (2019) point out that the president is more likely to travel overseas when there is a split government. Furthermore, itineraries of artists also serve the government's decision-making. For example, based on the mobility of artists, Toplak (2017) demonstrates that administrative barriers have impeded population mobility within the EU compared to the past.

Despite its value, the scarcity of available trip data prevents relevant studies from being extended to large-scale and network-wise. Though official websites may be an authoritative and accurate source for such studies (Ostrander and Rider 2019), they are infrequently updated and the coverage is limited. For instance, the Japanese Ministry of Foreign Affairs¹ releases the diplomatic blue book only once a year. Therefore, it is necessary to more efficiently obtain celebrity itineraries.

Online news contains a wealth of information related to celebrities, including their trips (Piskorski et al. 2020), thereby serving as a convenient source for acquiring copious celebrity itineraries. Our task is to find the time and locations of the trips made by given celebrities, from news articles. A reasonable intuition would be first narrowing down the search space to the articles containing the target celebrity's name, associated with the target date, further recognizing place names with off-the-shelf Named Entity Recognition (NER) tools (Lingad, Karimi, and Yin 2013; Li et al. 2020) as candidate locations, and choosing the most frequent one as the visited place. However, as shown in Figure 1, interfering information, including locations visited by other celebrities, always scatters around the travel destination, potentially impeding the task. According to our statistics, more than 50% of the locations mentioned in the news are not actually visited by the target celebrities. Even after sorting the locations by frequency, the results remain unfavorable (see Experimental Results). As an important refinement, the textual context of locations should be considered to distinguish travel locations from others. Figure 1 again demonstrates an intuitive example – words like *visited* (in the bottom article) signal the presence of Donald Trump in *Philadelphia*, while *Washington D.C.* without such words seems a less likely des-

*These authors contributed equally.

[†]Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹<https://www.mofa.go.jp/mofaj/gaiko/bluebook/index.html>

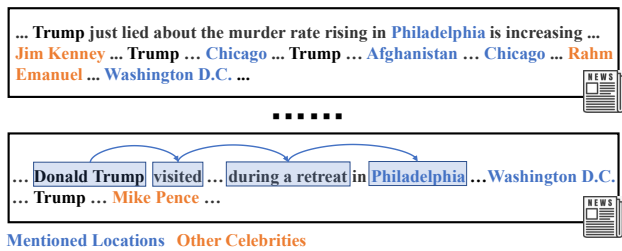


Figure 1: Example of extracting trip information of Donald Trump from news articles on 2017-01-26.

tinuation.

Besides the current task, textual descriptions are also utilized in solving detection problems, with the most relevant one being Person-Related Event Detection (PRED). Instead of mining public news reports on celebrities, PRED aims to detect predefined personal life events from social media by classifying user-posted text, especially tweets, into categories such as having children (Dickinson et al. 2015) or visiting (Yen, Huang, and Chen 2019). Though we can also model our task by classifying the aforementioned candidate locations into positives (actually visited) and negatives (actually not visited) given their textual contexts, adopting the PRED methods can be problematic. First, PRED methods classify one (short) tweet at a time (Yen, Huang, and Chen 2019) and they become less effective when dealing with multiple long articles. In our scenario, the mentions of celebrities and locations are scattered in much longer text, usually across different articles. Regarding the example in Figure 1, we need sequential combinations of *Trump*, *visit*, and *Philadelphia* that are far apart, as well as information about other candidate locations from various articles, to infer Trump’s itinerary for that day. Traditional bag-of-words based PRED approaches (Akbari et al. 2016) ignore the words’ sequence and associations, while methods that build upon convolutional neural networks (CNNs) (Nguyen et al. 2017) and recurrent neural networks (RNNs) (Yen, Huang, and Chen 2018) focus on local semantics and short sequence (Peng et al. 2019), overlooking long-range cross-document dependency. To incorporate all these, representing each candidate location in the binary classification problem with a graph of its contextual words, the documents mentioning it, and their associations, may be a more suitable choice.

The second major difference between our task and PRED is that we need to deal with more irrelevant celebrities and locations, since a single news article is much longer and the noises multiply when multiple articles are considered. The PRED task assumes the events are related to the specific user who posts the tweet, without delving into the identities of event participants (Wu, Koh, and Chen 2019; Yen, Huang, and Chen 2019). While in our scenario, we have to avoid confusing itineraries of other celebrities with the one of the target celebrity, by focusing on measuring how relevant each candidate location is to the target celebrity on its corresponding graph.

Another challenge faced in our scenario is the so-called “implicit trips,” in which the target celebrities and/or their visited locations are not directly mentioned by names, and these cases account for nearly 20% of all trips in our dataset. For example, the sentence “...Taylor Swift did something bad for the opening of the 2018 American Music Awards...from the Microsoft Theater...” does not directly reveal Taylor’s geographic location despite indicating her appearance at an event (2018 American Music Awards) and a particular entity (Microsoft Theater). Besides, for another common kind of implicit trips, news articles refer to politicians by their job titles, instead of their names. Background information of celebrities and other well-known entities can always be found in knowledge bases such as Wikidata (Vrandečić and Krötzsch 2014), while for specific events that are often absent in the knowledge bases, news articles published around the target date that cover these events may be a better source of information, especially when mining the latest itineraries.

In this paper, we present **CeleTrip**, the first data-driven framework for automatically detecting celebrity itineraries from news articles. Before CeleTrip does its job, we find news articles related to the target celebrity and the particular date, and extract candidate locations from them. CeleTrip (see Figure 2) will determine the actually visited locations through binary classification. As mentioned before, utilizing the textual context of locations poses a challenge of integrating semantics that are far apart. Previous work shows that graph neural networks can capture long-distance dependencies (Sahu et al. 2019) and help information aggregation through message-passing (Kipf and Welling 2016). To this end, we construct a **Word-Article graph** for each location ((a) in Figure 2), which takes its corresponding articles and words in them as nodes to encode the location’s context. To uncover the “implicit trips,” external sources of knowledge – Wikidata and news articles, need to be included. Due to their obvious heterogeneity, we decide not to mix them directly with the Word-Article graphs. Instead, we use an **Entity Sub-graph** for each named entity in the candidate locations’ relevant news articles to embed knowledge from Wikidata ((b) in Figure 2), while event-relevant information is represented by event embeddings learned from event-related sentences in news articles ((c) in Figure 2). The resulting entity embeddings and event embeddings, together with these location embeddings extracted from Word-Article graphs, are used to represent corresponding nodes in the **Trip Graph** of candidate locations, and their relevant entities and events that depict the target celebrity’s possible trips on a particular day ((d) in Figure 2). In this way, CeleTrip can jointly model the context of locations with external knowledge from a comprehensive perspective. Moreover, we propose an Oriented Pooling combining node similarity to capture the semantic relation between the candidate location and the target celebrity, reducing the distraction from irrelevant celebrities in the Word-Article graph, as opposed to the regular max pooling in graph convolutional networks (GCNs) (Deng, Rangwala, and Ning 2019).

We summarize our contributions as follows.

- We propose a new yet important task of celebrity

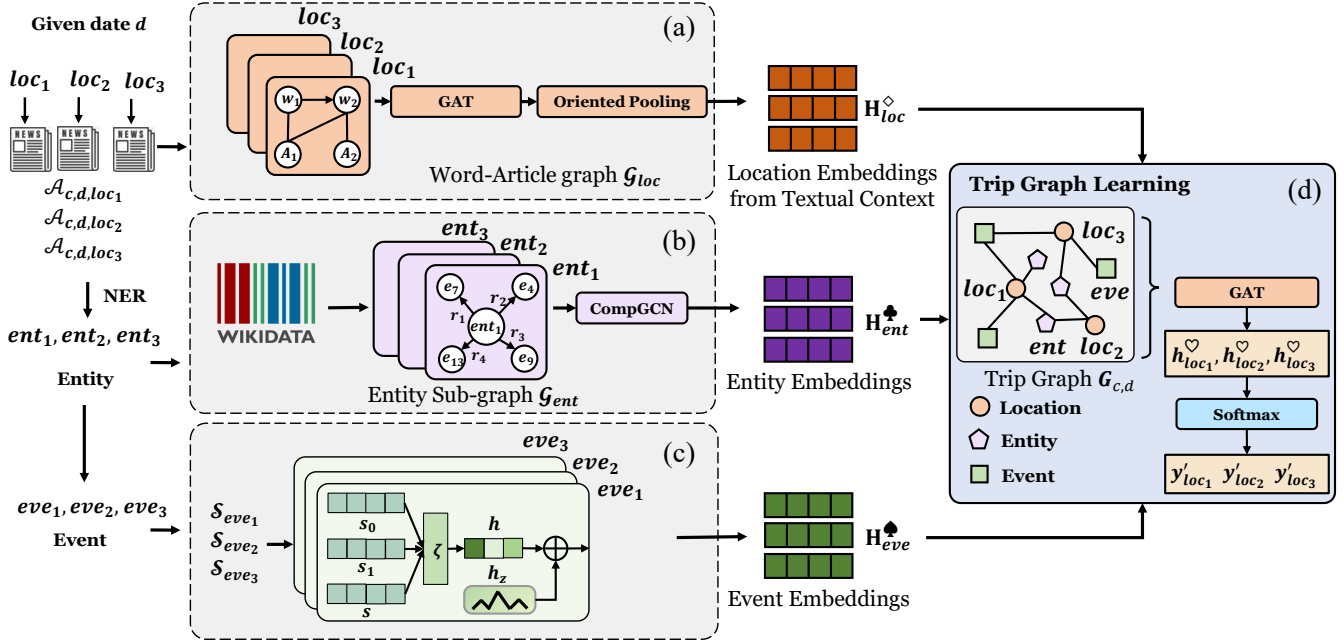


Figure 2: Our framework learns the overall representations of candidate locations and classifies them in Trip Graph, where the textual description of each candidate location is incorporated by the Location Embedding Learning, the knowledge of related entities is supplied by Entity Embedding Learning, and the information of related events is obtained by Event Embedding Learning.

itinerary detection, construct the largest dataset known so far, and develop a first framework, **CeleTrip**, for this task. **CeleTrip** mines news articles and outperforms all baselines with an F1 score of 82.53% on the dataset.

- We design Word-Article graphs to capture trip-related information scattered among different articles and propose a new graph pooling technique to refine the graph and focus on nodes of interest.
- We introduce a simple but effective method to update the representations of event entities based on related news and the attention mechanism.
- To facilitate relevant research, we release our dataset that includes 5,687 trips. Besides, we open-source our framework, and the assisting time/location extraction tools².

Problem Statement

Before we formalize the core problem, we describe the heuristic which is briefly mentioned in Introduction. Since we want to find the locations that celebrity c has visited on date d from news articles, we first narrow down our search to the articles published on date d or containing d in the text, and select these mentioning c 's name to be $\mathcal{A}_{c,d}$, which are the articles related to c and d . From these articles, we extract all mentioned locations as a set of candidate locations, denoted as $Loc_{c,d}$. The task now becomes: for each candidate location loc in $Loc_{c,d}$, decide whether it is actually visited

by celebrity c on date d , given articles $\mathcal{A}_{c,d}$ related to c and d .

We train a model f with trainable parameters Θ to classify $loc \in Loc_{c,d}$ into binary category $y \in \{0, 1\}$, with $y = 1$ indicating that c has visited loc on date d , and $y = 0$ being the opposite:

$$f : \{\mathcal{A}, loc, c, d, \Theta\} \rightarrow y, \quad (1)$$

As a side note, to better assist the core task, we develop our own tools for extracting time and locations, as described in the last subsection of Method.

Method

We elaborate **CeleTrip** (see Figure 2) in this section. As described in Introduction and Problem Statement, for any target celebrity c on date d , the framework learns the overall representations $(h_{loc_1}^{\heartsuit}, h_{loc_2}^{\heartsuit}, h_{loc_3}^{\heartsuit})$ of the candidate locations and classifies them in (d) **Trip Graph Learning**, by incorporating location embeddings from textual descriptions obtained in (a) **Location Embedding Learning**, embeddings of the locations' related entities from a knowledge base supplied by (b) **Entity Embedding Learning**, and the event embeddings for these event entities provided by (c) **Event Embedding Learning**. The Trip Graph in (d) models c 's possible trips on date d – candidate locations are connected to certain types of co-occurring entities (e.g. person, facility, and event) detected by a Named Entity Recognition (NER) tool and for these event entities, they are represented as event nodes instead, which augment event infor-

²<https://github.com/ZhangDataLab/CeleTrip>

mation from relevant news articles. In the Trip Graph, candidate locations can be indirectly linked if they share the same related entities/events. Additionally, we describe our tools for extracting time and locations at the end of this section.

Location Embedding Learning

Word-Article Graph As discussed in Introduction, it is important to aggregate scattered travel descriptions of the celebrity from various news articles. We draw inspirations from word graphs designed by Deng, Rangwala, and Ning (2019) and extend them to include articles as a Word-Article graph \mathcal{G}_{loc} ((a) in Figure 2), which allows us to capture the semantic relations across multiple articles. The Word-Article graph is constructed as follows.

Given the celebrity c , date d , and the candidate location loc , we first obtain a subset of news articles $\mathcal{A}_{c,d,loc}$ from $\mathcal{A}_{c,d}$ by selecting articles that contain location loc . We take both articles and words in $\mathcal{A}_{c,d,loc}$ as nodes in the Word-Article graph \mathcal{G}_{loc} , and initialize them using TF-IDF vectors and Word2Vec, respectively. To capture dependencies between words that are far apart, we use a sliding window to construct edges between words within sentences (Deng, Rangwala, and Ning 2019). To aggregate semantic information across different articles, we add edges between articles and words based on their co-occurrence within the articles (Wang et al. 2020a). Meanwhile, we also record the word node subscripts of words that appear in the names of the location loc and celebrity c during Word-Article graph construction as the location word node lw and celebrity word node cw , which will be used later in **Oriented Pooling**.

To identify important words from lengthy text descriptions, we use the graph attention network (GAT) (Velickovic et al. 2018), which employs an attention mechanism during graph aggregation, to selectively gather information from nearby nodes.

Given the Word-Article graph \mathcal{G}_{loc} , its adjacency matrix $\mathbf{B}_{loc} \in \mathbb{R}^{(\iota+\tau) \times (\iota+\tau)}$, its word feature matrix $\mathbf{X}_{loc}^W \in \mathbb{R}^{\iota \times F_W}$ and the article feature matrix $\mathbf{X}_{loc}^A \in \mathbb{R}^{\tau \times F_A}$, where ι is the number of words and τ is the number of articles, we project the article and word nodes onto the same feature dimension using two independent linear layers with identical output dimension (Eq. (2)- (3)).

$$\mathbf{H}_{loc}^W = \mathbf{W}_W \mathbf{X}_{loc}^W + b_W, \quad (2)$$

$$\mathbf{H}_{loc}^A = \mathbf{W}_A \mathbf{X}_{loc}^A + b_A, \quad (3)$$

where \mathbf{W}_* is a learnable weight matrix. Hence, we get the node feature matrix $\mathbf{H}_{loc} = \mathbf{H}_{loc}^W \oplus \mathbf{H}_{loc}^A$ by concatenating the word feature matrix and the article feature matrix.

Then we use GAT to update the node representation. Since the module would be repeatedly utilized in the model, we omit the superscripts indicating layer number for brevity.

$$\mathbf{H}_{loc} = \text{GAT}(\mathbf{H}_{loc}, \mathcal{G}_{loc}). \quad (4)$$

Oriented Pooling After updating node representations in the Word-Article graph, we design Oriented Pooling to focus on the semantics of the target celebrity and candidate

location, thus reducing the distraction from irrelevant information. Oriented Pooling refines the Word-Article graph through the relevance of each node to the location word node lw and the celebrity word node cw .

In Oriented Pooling, we first refer to SAGPool (Lee, Lee, and Kang 2019) to calculate the self-attention scores s_{attn} on each node using graph convolution. To capture the semantic relation between the target celebrity and candidate location, we compute the similarity s_{sim} between each node and the target nodes (location word lw and celebrity word cw). Here we obtain the embeddings of target nodes by node subscripts recorded during Word-Article graph construction. After combining s_{attn} and s_{sim} with additive attention, we get the oriented score s_{ori} .

$$s_{loc} = \text{similarity}(\mathbf{H}_{loc}, \mathbf{h}_{lw}), \quad (5)$$

$$s_c = \text{similarity}(\mathbf{H}_{loc}, \mathbf{h}_{cw}), \quad (6)$$

$$s_{sim} = s_{loc} \odot s_c, \quad (7)$$

$$s_{ori} = \tanh(\mathbf{W}_\alpha [s_{attn} \oplus s_{sim}]), \quad (8)$$

where \mathbf{h}_{lw} and \mathbf{h}_{cw} are the embeddings of node lw and node cw , $\text{similarity}(\ast)$ is the function to compute node similarity, \odot is the element-wise product, and \oplus means ‘‘concatenate’’.

Then we refine the graph using the oriented score. The pooling ratio $\epsilon \in (0, 1)$ is a hyperparameter that determines the ratio of nodes to be saved. We select $\lceil \epsilon N \rceil$ nodes with the highest oriented score. N is the number of nodes in the Word-Article graph.

$$idx = \text{top-rank}(s_{ori}, \lceil \epsilon N \rceil), \quad (9)$$

$$\mathbf{H}_{loc} = \mathbf{H}_{loc}[idx]; \mathbf{B}_{loc} = \mathbf{B}_{loc}[idx], \quad (10)$$

$$\mathbf{h}_{loc}^\diamond = \text{Maxpooling}(\mathbf{B}_{loc}, \mathbf{H}_{loc}), \quad (11)$$

where $\text{top-rank}(\ast)$ is the function to return the subscripts of the selected nodes and \mathbf{B} is the adjacency matrix. We operate GAT and Oriented Pooling iteratively in the location embedding learning module, and finally obtain the location embedding $\mathbf{h}_{loc}^\diamond$ of the candidate location loc by max pooling.

After performing the location embedding learning module for each candidate location, we get the location node feature matrix $\mathbf{H}_{loc}^\diamond = [\mathbf{h}_{loc_1}^\diamond, \dots, \mathbf{h}_{loc_k}^\diamond]^T \in \mathbb{R}^{k \times F}$, where $k = |\text{Loc}_{c,d}|$ is the number of candidate locations on date d for the celebrity c , and F is the feature dimension.

Entity Embedding Learning

As previously discussed, implicit trips pose another challenge in celebrity trip detection. To utilize the background knowledge of celebrities and other named entities, we first employ SpaCy³ to recognize entities of specific types, including person, nationalities or religious or political groups (NORP), facility and organization, from $\mathcal{A}_{c,d}$. Then we obtain the structured information and relationships of entities from Wikidata embeddings provided by OpenKE (Han et al. 2018) through a strict text matching method such that we do not introduce noise in this process.

³<https://spacy.io/>

Given the entity ent and its embedding h_{ent} initialized by Word2Vec, we construct an Entity Sub-graph \mathcal{G}_{ent} specific to ent based on the relation data from OpenKE ((b) in Figure 2). Nodes in \mathcal{G}_{ent} are ent and external entities connected to ent through the data provided by OpenKE. These relations are represented as edges in the graph. To learn the entity embedding vector, we adopt the CompGCN (Vashishth et al. 2020) layer, which is the state-of-the-art model for learning entity and relation embeddings on graphs.

$$\mathbf{h}_{ent}^- = \sigma\left(\sum_{(r,u)\exists(ent,r,u)\in\mathcal{G}_{ent}} \mathbf{W}_{es}\phi(\mathbf{h}_u, \ell_r)\right), \quad (12)$$

$$\ell_r = \mathbf{W}_{edge} \cdot \ell_r. \quad (13)$$

where \mathbf{h}_u and ℓ_r denote features for node u and relation type r in the \mathcal{G}_{ent} , respectively. \mathbf{W}_{es} is the weight parameter for aggregating node and edge features, and $\phi(\cdot)$ is the multiplication operator. $\sigma(\cdot)$ is the activation function. \mathbf{W}_{edge} is the weight parameter for updating the relation type embedding.

After aggregating the features of nodes and edges, we obtain the semantic feature \mathbf{h}_{ent}^- from the sub-graph. Then we concatenate the initialized embedding of entity \mathbf{h}_{ent} with the semantic feature \mathbf{h}_{ent}^- , and obtain the updated embedding of entity ent through a non-linear transformation:

$$\mathbf{h}_{ent}^\clubsuit = \sigma(\mathbf{h}_{ent} \oplus \mathbf{h}_{ent}^-). \quad (14)$$

The entity embeddings $\mathbf{H}_{ent}^\clubsuit = [\mathbf{h}_{ent_1}^\clubsuit, \dots, \mathbf{h}_{ent_n}^\clubsuit]^T \in \mathbb{R}^{n \times F}$ are obtained. n is the number of entities.

Event Embedding Learning

Events also provide important information for detecting implicit trips, since celebrity trips are often associated with specific events (Doherty 2009; Ostrander and Rider 2019) and our analysis on Trip Dataset gives consistent evidences (see Distinguishing Ability of Events). However, pre-trained embeddings from external knowledge bases cover limited events. To capture the current states of events, we propose a module that integrates contemporaneous event-related news articles to obtain the event embeddings.

Formally, given the event entity eve recognized by SpaCy, the framework retrieves sentences \mathcal{S}_{eve} containing the event eve from news articles published on the date d . We introduce an attention mechanism that differentiates the importance of sentences by assigning varying weights to their content and aggregates the feature of sentences based on weights, since not all sentences contribute equally to the status of events.

Specifically, we encode the i -th sentence in \mathcal{S}_{eve} and get the sentence vector \mathbf{v}_i by Word2Vec. Afterwards, we derive the attention values by feeding the sentence vector \mathbf{v}_i into an attention layer, followed by obtaining the normalized attention weight λ_i through a softmax function. Finally, the representation \mathbf{h}_{eve} is obtained as a weighted sum of each sentence vector, and the attention layer can be trained end-to-end while gradually learning to assign more attention to

reliable and informative sentences.

$$\mu_i = \text{Sigmoid}(\mathbf{W}_{eve}\mathbf{v}_i + b_{eve}), \quad (15)$$

$$\lambda_i = \frac{\exp(\zeta_i \mu_i)}{\sum_j \exp(\zeta_j \mu_j)}, \quad (16)$$

$$\tilde{\mathbf{h}}_{eve} = \sum_i \lambda_i \mathbf{v}_i, \quad (17)$$

where ζ_i is the parameter for \mathbf{v}_i and \mathbf{W}_{eve} is the parameter matrix. Intuitively, the number of news articles related to an event increases when the event is ongoing, thus reflecting the state of the event. Therefore, we add a vector representing daily counts of news articles that contain the event. When the framework detects event eve on date d , it counts the number of articles containing event eve from $d - Q$ to $d + Q$ to construct the vector $\mathbf{z} \in \mathbb{R}^{1 \times (2Q+1)}$, where Q is the date offset. Then, the framework applies a linear layer to project the vector to a low-dimensional space $\mathbf{h}_z = \psi_1(\mathbf{z})$ and fuses it with $\tilde{\mathbf{h}}_{eve}$ to get the event status embedding $\mathbf{h}_{eve}^\spadesuit$.

$$\mathbf{h}_{eve}^\spadesuit = \sigma(\psi_2(\mathbf{h}_z \oplus \tilde{\mathbf{h}}_{eve})) + \mathbf{h}_z, \quad (18)$$

where $\sigma(\cdot)$ is activation function and $\psi_*(\cdot)$ is linear layer. In addition, we design a skip connection of the article-amount embedding \mathbf{h}_z to allow its gradients to flow in the network. Based on the above operation, we get the event embeddings $\mathbf{H}_{eve}^\spadesuit = [\mathbf{h}_{eve_1}^\spadesuit, \dots, \mathbf{h}_{eve_m}^\spadesuit]^T \in \mathbb{R}^{m \times F}$ (m is the number of events) for all events.

Trip Graph Learning

To perform trip detection from a comprehensive perspective and jointly train all modules, the framework consolidates the embeddings of locations, entities, and events from the same date in one graph. With the final representations learned from this module, locations can then be classified.

Trip graph construction Formally, we define the Trip Graph of the celebrity c on date d as $\mathcal{G}_{c,d} = (\mathcal{V}, \mathcal{E})$ ((d) in Figure 2), where \mathcal{V} consists of candidate locations $Loc_{c,d} = \{loc_1, \dots, loc_k\}$, entities $Ent = \{ent_1, \dots, ent_n\}$ and events $Eve = \{eve_1, \dots, eve_m\}$. We build an edge between the location node loc_k and the entity node ent_n (the event node eve_m), if entity ent_n (events eve_m) appears in \mathcal{A}_{c,d,loc_k} .

Trip graph aggregation Given the location embeddings $\mathbf{H}_{loc}^\diamond$, entity embeddings $\mathbf{H}_{ent}^\clubsuit$, event embeddings $\mathbf{H}_{eve}^\spadesuit$, and the Trip Graph $\mathcal{G}_{c,d}$, the framework jointly trains the embeddings of locations, entities, and events in the Trip Graph aggregation module. The Trip Graph aggregation module propagates the gradient back to the location embedding learning module, the entity embedding learning module, and the event embedding learning module.

Specifically, we concatenate embeddings of locations, entities, and events, and obtain the feature matrix $\hat{\mathbf{H}}_{c,d} \in \mathbb{R}^{(k+n+m) \times F}$ of the Trip Graph $\mathcal{G}_{c,d}$.

$$\hat{\mathbf{H}}_{c,d} = \mathbf{H}_{loc}^\diamond \oplus \mathbf{H}_{ent}^\clubsuit \oplus \mathbf{H}_{eve}^\spadesuit. \quad (19)$$

To capture the various influences of entities and events on trip detection, the framework uses GAT to aggregate the information of the Trip Graph.

$$\mathbf{H}_{c,d}^\heartsuit = \text{GAT}(\hat{\mathbf{H}}_{c,d}, \mathbf{G}_{c,d}). \quad (20)$$

Afterwards, our framework derives the feature matrix $\mathbf{H}_{c,d}^\heartsuit$ of the Trip Graph $\mathbf{G}_{c,d}$ and applies the softmax function to the location nodes’ features to estimate the probability of each location being the celebrity trip location.

$$y'_i = \text{Softmax}(\mathbf{h}_{loc_i}^\heartsuit). \quad (21)$$

The framework compares the prediction result with the ground truth and optimizes via the binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{|Loc_{c,d}|} \sum_i^{|Loc_{c,d}|} (y_i \log(y'_i) + (1 - y_i) \log(1 - y'_i)). \quad (22)$$

where $|Loc_{c,d}|$ is the number of candidate locations on date d , and y_i is the ground truth label of i -th location.

Extraction Tools for Dates and Locations

As mentioned in Problem Statement, we need to extract dates and locations from text. Existing time extraction tools, such as SUTIME (Chang and Manning 2012), are general-purposed for various time information, while we only care about dates. As a result, these tools are quite slow and may not archive the best performance (see Evaluation of Extraction Tools). We therefore develop a rule-based **time extraction tool** with regular expressions to handle absolute dates (e.g., “July 16, 2022”) and relative dates (e.g., “yesterday”).

Regarding locations, we can potentially use the results from NER tools. Again, as general-purpose tools, they often produce coarse results (e.g., “North Korea’s”) instead of “North Korea”) and locations with containment relationships (e.g., “New York” v.s. “United States”) that affect further modeling work. Thus, we develop a **location extraction tool** that integrates GeoNames⁴, a geographic database, to get high-quality location names, and uses administrative-level information from OpenStreetMap⁵ to resolve containment relationship issue.

Experiments

Data Collection We construct a ground truth trip dataset from the Wikipedia pages that list celebrity trips⁶. In total, there are 5,687 trips with dates and locations made by 26 politicians and 24 artists, from Jan 2016 to Feb 2021.

We then collect the URLs of news articles published during this period of time from the GDELT project⁷ and use Newspaper3k⁸ to acquire the news content of the URLs.

⁴<https://www.geonames.org/>

⁵<https://www.openstreetmap.org/>

⁶https://en.wikipedia.org/wiki/List_of_international_prime_ministerial_trips_made_by_Theresa_May The Wiki page for Theresa May’s trips, as an example.

⁷<https://www.gdeltproject.org/>

⁸<https://github.com/codelucas/newspaper>

Data Processing and Labeling To obtain the dataset for training and testing in the format described in the Problem Statement, we further process the collected data. As a pre-step, we first perform NER on the news articles using SpaCy. For each ground truth trip, we extract candidate locations of the corresponding celebrity from news articles that contain the celebrity name and relate to the trip date (published on the date, or mentioning the date in text) by our extraction tool. For each candidate location, if it is actually visited on that date according to Wikipedia, it will be labelled positive and vice versa. For each negative/positive location, its articles are news articles that contain the location name among the above news articles. As a result, we identify 2,404 locations as “positive locations” and 9,707 locations as “negative locations”. This approach allows us to construct the **Trip Dataset** for the task, with each row recording the celebrity name, location name, date, article IDs, and the label. Table 1 shows a sample from this dataset.

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

Table 1: A sample from the Trip Dataset.

Train/Test Split With July 1, 2019 as the split date, we partition the Trip Dataset into training (66.5%) and testing (33.5%) to avoid potential data leakage, since SpaCy is trained on data prior to July 2019. Then we randomly sample 10% from the training set as the validation set to optimize the hyperparameters. Table 2 lists the details of the training and test sets.

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: Training and test sets from the Trip Dataset.

Distinguishing Ability of Events We wonder whether the positive locations are associated with different events, compared with negative ones, as suggested by some studies (Doherty 2009; Ostrander and Rider 2019). If so, event information may be helpful for distinguishing positive locations from negative ones. In Table 3, we list the top 10 most frequent events extracted by SpaCy from articles of positive samples and negative samples respectively, and see how they differ. We can see that 6 events out of 10 for positive samples are international meetings of leaders (highlighted in Table 3), while for the negative ones, the events seem a mixture of protests (“Women’s March”), sports events (e.g., “Olympics”), social gatherings (“Unite the Right”), and disasters (e.g., “Hurricane Harvey” and “Sandy Hook”). This hints that event information might contain clues for classifying the samples.

Implementation Details To prepare the news articles for further processing, we remove stop words and perform stem-

Positive	Negative
G20	Hurricane Harvey
World Economic Forum	Women’s March
G7	Unite the Right
Brexit	Republican Nat’l Convention
NATO Summit	Democratic Nat’l Convention
Hurricane Harvey	Olympics
APEC Summit	Super Tuesday
Women’s March	Super Bowl
World Cup	Hurricane Maria
Belt and Road Forum	Sandy Hook

Table 3: Frequent events associated with positive and negative samples.

ming by NLTK⁹. In the location embedding learning module and event embedding learning module, to get original representations of words and sentences, we train an unsupervised CBOW Word2Vec (Mikolov et al. 2013) with the dimension of 100 on the entire news dataset.

We describe the parameters for the four components:

- Location Embedding Learning:** During the construction of Word-Article graphs, we set the size of the sliding window to 15 and aggregate the graph with two GAT layers. Additionally, the dimension of TF-IDF vectors used to initialize article features is 1,000. In Oriented Pooling, we use cosine similarity as the similarity function in Eq. (5) and Eq. (6).
- Entity Embedding Learning:** We utilize pre-trained Wikidata embeddings provided by OpenKE (Han et al. 2018) to construct the Entity Sub-graphs, with entities and relations dimensions set to 50. One CompGCN layer is employed in Entity Sub-graphs.
- Event Embedding Learning:** The date offset Q is set to 7.
- Trip Graph Learning:** We set the input dimension F of Trip Graph to 128 and the number of GAT layers to 2.

To train CeleTrip, we set the learning rate to 0.001 and use the Adaptive Moment Estimation (Adam) optimizer (Kingma and Ba 2015). Meanwhile, we adopt an early stop strategy to avoid overfitting.

Evaluation Metrics We use Accuracy (Acc), Precision (P), Recall (R), and F1 score for evaluation. Precision measures the proportion of the locations predicted as positive that are actually visited by the target celebrities, while Recall indicates the ratio of the actually visited locations that are correctly classified. F1 score, as the harmonious average of Precision and Recall, is an important overall metric in our scenario.

Baselines Here we introduce seven baselines. The first two (LocFre and LocJaccard) are based on the intuition described in Introduction which relies on the frequency of each candidate location to make the decisions. The next four models, LR (TF-IDF), CNN and Bi-LSTM, and GCN, as mentioned in Introduction, are often used in PRED and other

⁹<https://www.nltk.org/>

Detection tasks. Among them, the graph-based GCN may bring improvements, since it has the ability to capture long-distance dependencies. But as discussed before, GCN employs max pooling to read out the graph, compared with our model CeleTrip. We need to mention that Oriented Pooling in our model uses the features of celebrity/location word nodes during graph pooling, and this information is not originally included in the input to these four models. Therefore, as a fair comparison, we concatenate the celebrity name and location name embeddings with the output vectors of these four models, before feeding them to the classifying modules.

Another important aspect that we want to evaluate is the inclusion of external knowledge through entity embeddings and event embeddings. The aforementioned four models by their original designs do not incorporate such information, and therefore we alternatively replace the location embedding learning module in our framework with each of these models, to see how the knowledge and the joint learning scheme would affect their performance. For the same purpose, we introduce CeleTrip_{w/o kn}, without modules for external knowledge and the accompanying joint learning scheme, to compare against the full CeleTrip.

- **LocFre:** It chooses the location mentioned most frequently in the news as the trip location.
- **LocJaccard:** LocFre favors the locations that are generally popular across all news articles and this can be a bias. To normalize the frequency by popularity, LocJaccard calculates the Jaccard coefficient between articles containing candidate location and articles containing target celebrity, as a possible improvement over LocFre.
- **LR (TFIDF) (Dickinson et al. 2015):** The researchers use TF-IDF vectors as the representations of news articles and employs Logistic Regression as the classifier for a PRED task. We directly adopt it in our experiments.
- **CNN (Nguyen et al. 2017):** In their study of detecting crisis from personal tweets, the model learns local semantic features of articles. The article feature vectors yielded by the CNN model are then concatenated together as the input to the classifier. In our experiments, we use the softmax classifier for all deep learning models.
- **Bi-LSTM (Yen, Huang, and Chen 2019):** For the PRED task, the researchers use Bi-LSTM, a sequential learning model to captures the continuous semantic features of Chinese characters in tweets. In our task, we treat English words as the Chinese characters in their study.
- **GCN (Deng, Rangwala, and Ning 2019):** In the event prediction task, the researchers model articles as word graphs and aggregate the information in the word graph by graph convolutional network (GCN). In the experiments, we construct the same Word-Article graphs instead of the original word graphs.
- **CeleTrip_{w/o kn}:** This model employs the proposed Word-Article graph to model location context and obtains the representation of the graph via Oriented Pooling.

Experimental Results

Prediction Performance The results averaged over 5 runs are presented in Table 4. Among the models, CeleTrip exhibits the best overall performance (F1=82.53%), followed by CeleTrip_{w/o kn}, the version of our model without external knowledge (entity and event embeddings).

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 _{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

Table 4: Performance comparison on the test set.

As the best model among these without external knowledge (upper part of the table), CeleTrip_{w/o kn} outperforms by large margins, suggesting the effectiveness of graph structure with attention mechanism. The GCN model which is also graph-based, however, does not perform so well, only surpassing the two frequency-based methods. A possible explanation is that words should not be treated equally in our scenario – indicative words need to have higher weights. This is further supported by the interesting observation that the LR (TFIDF) method, has the highest Precision (88.64%) and it gives higher weights to indicative words relevant to trips, such as “visit” or “arrive”. On the other hand, LR (TFIDF) has the lowest Recall, and it may be related to the fact that it does not model relationships between words, therefore missing meaningful word dependencies/combinations.

We notice that CeleTrip_{w/o kn} outperforms CNN and Bi-LSTM, which focus on local semantics, by 10.36% and 7.18% in F1, highlighting the necessity of capturing long-distance dependencies in our scenario. In our task, descriptions of a trip sometimes scatter in different sentences or articles, making it important to consider word relationships that are far apart.

Additionally, we observe that simple frequency-based models have F1 scores of only 40%, indicating limited practicality despite some predictive effect. Conversely, models incorporating textual information significantly surpass naive approaches in all metrics, demonstrating the importance of textual description in trip detection.

Performance for Unseen Celebrities in Training Two politicians, “Boris Johnson” and “Joe Biden”, appear only in the test set, with 33 possible trips (16 positives and 17 negatives). All of these trips are correctly classified by CeleTrip,

indicating the model generalizes well for these two celebrities.

Effectiveness of External Knowledge The performance of all baseline models improves when combined with external knowledge. Specifically, the incorporation of external knowledge in LR (TFIDF), CNN, Bi-LSTM, GCN, and CeleTrip_{w/o kn} yields improvements of 3.85%, 6.10%, 3.75%, 4.52%, and 8.42% respectively in F1 score, among which CeleTrip has the largest gain.

Ablation Study

We remove certain component(s) in CeleTrip to validate their efficacy, with the results in Table 5. CeleTrip_{w/o ent} removes the entity embedding learning module and directly uses pre-trained entity (Word2Vec) embeddings. Similarly, CeleTrip_{w/o eve} gets rid of the event embedding learning module and uses pre-trained event embeddings constructed from Word2Vec embeddings (see construction details in Method). CeleTrip_{w/o ent&eve} is the ablation without both aforementioned modules and employs original entity and event embeddings as CeleTrip_{w/o ent} and CeleTrip_{w/o eve} do. Additionally, we replace Oriented Pooling in CeleTrip with max pooling and obtain CeleTrip_{w/o ori}.

In Table 5, as expected, CeleTrip has the best overall performance compared with all variants, and CeleTrip_{w/o ent&eve} is the worst, demonstrating the effectiveness of corresponding module designs. Compared with CeleTrip, CeleTrip_{w/o eve} has a slightly higher Precision (increased by 1.05%) but a much lower Recall (decreased by 5.98%). It is possible that by introducing event information from additional news articles, the Recall is greatly improved, but the accompanying noise affects Precision by a small margin. Moreover, we find that CeleTrip outperforms CeleTrip_{w/o eve} by 2.88% and CeleTrip_{w/o ent} by 4.75% in the F1 metric, indicating that the entity embedding learning module has a larger contribution, compared with the other module. It is worth noting that Oriented Pooling brings a large improvement of 5.73% in F1, and the enhancement may be attributed to the suppression of unrelated information.

Methods	F1 (%)	P (%)	R (%)	Acc (%)
CeleTrip _{w/o ent&eve}	76.68	84.20	70.39	92.77
CeleTrip _{w/o ori}	76.80	81.29	72.91	92.51
CeleTrip _{w/o ent}	77.78	84.38	72.13	93.00
CeleTrip _{w/o eve}	79.65	87.22	73.29	93.64
CeleTrip	82.53	86.17	79.27	94.30

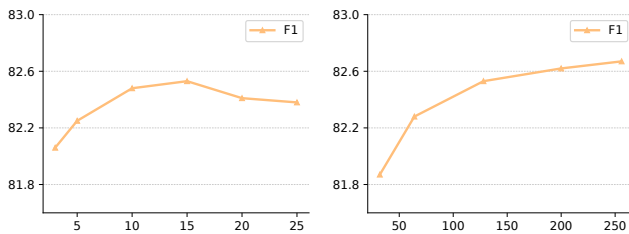
Table 5: Performance of different CeleTrip variants.

Parameter Sensitivity

Window Size for Word-Article Graph When building the Word-Article graph, we have a sliding window to decide which words should be connected. Figure 3(a) shows that 15 may be an ideal size – smaller windows may fail to capture long-distant dependencies while the ones larger than 15 may introduce more noise than the possible gains.

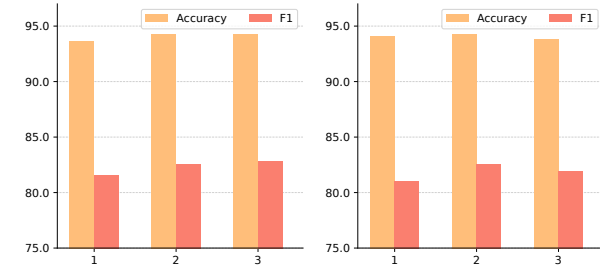
Node Dimension F in Trip Graph In Figure 3(b), the model performance improves with increased input dimensionality. Considering the fact that the improvement is marginal beyond $F = 128$, and the trade-off between performance and computational cost, we decide 128 to be a practical cut.

Layers of Graph Aggregators We vary the number of GAT layers from 1 to 3 for the Word-Article graph and the Trip Graph, and show the results in Figure 3(c) and Figure 3(d), respectively. It seems that 3 is a good choice for the Word-Article graph, while 2 may be ideal for the Trip Graph. Observing the increasing performance in the Word-Article graph, as the number of layers increases, we guess that for a large graph like the Word-Article graph, more layers may help encode more helpful information, and for a much smaller Trip Graph, too many layers may cause over-smoothing (Chen et al. 2020).



(a) Window Size

(b) Dimension F



(c) # of GATs in Word-Article graph (d) # of GATs for Trip Graph

Figure 3: Sensitivity analysis of different parameters. The horizontal coordinates indicate the values of parameters, and the vertical coordinates indicate the value of corresponding metrics.

Qualitative Evaluation

We examine what the event and entity embedding learning modules capture respectively. Besides, we perform an error analysis on incorrectly classified samples.

Interpretation of Event Embeddings The event embedding learning module learns the representations of the sentences that contain the events appearing in the context of candidate locations, during which each sentence receives an attention value. We manually examine top and bottom 500

sentences among 11,891 sentences ranked by their attention values, respectively. We find that in 85% of the top sentences, the events are actually related to the positive trips. As a comparison, in 96% of the bottom sentences, the corresponding events seem unrelated and are mostly about natural disasters and wars. As a demonstration, we show top 5 and bottom 5 sentences in Figure 4, and the events are highlighted.



Figure 4: Top 5 and bottom 5 sentences ranked by attention values from the event embedding learning module.

Celebrity Entity Embedding Visualization To see how the relationships between entities reflect the real world, we choose to study the celebrity entities by using t-SNE (Van der Maaten and Hinton 2008) to reduce the embedding dimensions and visualizing them in a 2-D space. In Figure 5, interestingly, politicians cluster together in the lower left diagonal half, while artists occupy the other half. The relatively close proximity between European politicians, compared with other regions in the world, may be attributed to frequent regional meetings within the EU. Specifically, Theresa May and Angela Merkel form the closest pair among all celebrities and this may be the result of frequent co-occurrences, given that the two had overlapped terms of office during which both focused on Brexit and other EU affairs. These highlight the model’s potential to identify regional political patterns.

The artists in the upper right diagonal half, on the other hand, are more scattered than politicians. The reason for this discrepancy may be that artists are more likely to attend solo activities/performances, while politicians often meet each other. This is further confirmed by our statistics – 50.15% of the politician trips are joined by others, and the ratio is only 13.04% for artists.

Error Analysis We examine the error cases to summarize their causes. In the evaluation, there are two types of errors, false positives (locations not actually visited but predicted as visited, associated with Precision) and false negatives (locations actually visited but predicted as not visited, associated with Recall). We sample 25% from each type, resulting in 21 false positives and 36 false negatives.

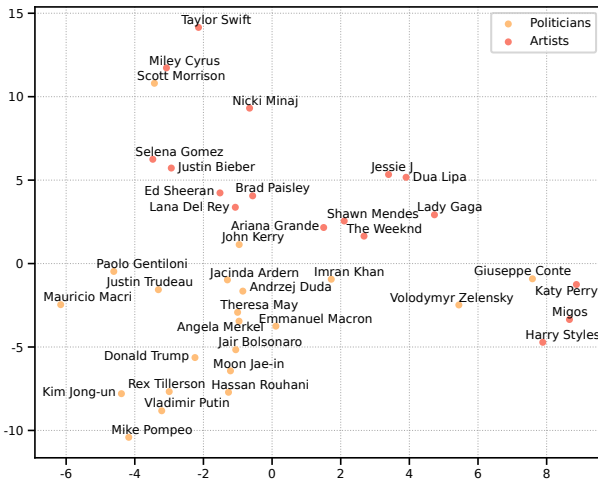


Figure 5: Visualization of the learned representations for celebrity entities, through t-SNE.

False positives are mainly caused by: (1) trips by other candidates (38%); and (2) absent actual dates (19%). Though Oriented Pooling is designed to put focus on target celebrities, trips of others still distract the model to some extent. On the other hand, actual dates of the trips may not appear in the articles, and other dates in the articles or the publishing dates may be mistakenly used instead.

The main error reasons for false negatives are: (1) implicit trips with very vague clues (43%); and (2) the target celebrity and the visited place have no apparent associations, though appearing in the same article (37%). For reason (1), we see cases such as campaigns and fund-raising events, indicating the target celebrities must have been somewhere, but the available information is not enough for the model to link to external knowledge. We demonstrate reason (2) by the following sentences from one article: “Some leaders, including German Chancellor **Angela Merkel**, see regulating AI as part of the EU’s ...” and “Von der Leyen said the **Brussels** summit deal, reached in the wee hours of Friday by 27 national EU leaders”. The association between Merkel and Brussels seems very insufficient for the model to classify it as a positive trip.

Evaluation of Extraction Tools

Time Extraction We compare our detection tool with SUTIME (Chang and Manning 2012), an advanced time recognition tool, on the TempEval dataset (UzZaman et al. 2012). Our tool achieves an accuracy of 91.1%, which is slightly higher than SUTIME (90.6%), and greatly reduces the processing time by over 80%, from 41 mins to 8 mins.

Location Extraction Our location extraction tool (SpaCy+GeoNames) is built upon the NER results from SpaCy, and further uses a geolocation database, GeoNames, to remove inaccurate location names and expand some abbreviated names to increase matches. The performance is evaluated on the New York and Christchurch datasets (Middleton, Middleton, and Modafferi 2013), and Table 6 shows

Methods	F1 (%)	P (%)	R (%)
New York Dataset			
SpaCy	68.02	60.36	77.91
SpaCy+GeoNames	82.38	86.66	78.50
Christchurch Dataset			
SpaCy	64.88	63.50	66.31
SpaCy+GeoNames	71.18	78.62	65.02

Table 6: Performance of location extraction strategies.

the large improvements in F1 scores and Precision.

Related Work

We discuss three threads of research: celebrity trip analysis, person-related event extraction and graph neural network.

Celebrity Trip Analysis

Celebrity trips have purposes that may extend to important implications. For instance, politicians travel to negotiate foreign policies and appeal to the public (Kernell 2006; Ostrander and Rider 2019), while artists make trips to promote their work (Hautala and Jauhiainen 2019). Based on the mobility (births and deaths) of 150,000 notable individuals, researchers carry out an impressive analysis to chart cultural history (Schich et al. 2014).

Among the quantitative studies, U.S. presidential travels have been extensively studied. Findings include that the presidents’ trips are influenced by world events and the level of U.S. involvement (Cavari and Ables 2019), and the purposes of early presidents’ domestic trips are to advance reelection (Sellers and Denton 2006), promote policies and meet with key leaders or groups (Doherty 2009).

As mentioned in Introduction, data scarcity is a common problem that prevents relevant studies from being extended to large-scale, fine-grained, and network-wise.

Person-Related Event Detection

In Introduction, we have briefly discussed Person-Related Event Detection (PRED), which classifies user-posted short text into predefined personal life events, while our classification task deals with much longer news articles. In a relatively early study (Dickinson et al. 2015), researchers detect events such as ‘Falling in Love’ and ‘Getting Married’ by traditional machine learning models (e.g., SVM) with manually designed bag-of-words features. As pointed out by Nguyen et al. (2017), feature engineering based methods cannot effectively model word dependencies. To improve this, deep learning models on text are consequentially used, including CNN (Nguyen et al. 2017; Wu, Koh, and Chen 2019) and Bi-LSTM (Yen, Huang, and Chen 2018, 2019). Comparing with social media text, key information in news articles is inevitably mixed with and scattered around much more irrelevant information. Therefore, graph-based models with the ability to capture long-distance semantic dependency (Yao, Mao, and Luo 2019), may be a suitable choice for our task.

Graph Neural Network

Recently, graph neural network has been widely used in tasks such as text classification (Alam, Joty, and Imran 2018; Yao, Mao, and Luo 2019), event prediction (Deng, Rangwala, and Ning 2019), and sentiment analysis (Wang et al. 2020b; Li et al. 2021) due to its ability to capture long-range structural relationships on word graphs converted from the text. In our task, a major difference/difficulty is that news articles often contain trips of other celebrities, which distracts the model from focusing on the target celebrity. This can be improved, regarding the graph pooling technique applied to read out the graphs. The aforementioned methods rely on global pooling which cannot effectively distinguish key information (related to the target celebrity) from distractions (Atwood and Towsley 2016; Simonovsky and Komodakis 2017) and this is later improved (Lee, Lee, and Kang 2019) by incorporating the attention mechanism that learns the importance automatically. In our scenario, rather than letting the attention mechanism to decide the importance all by itself, we further guide it to focus on the target celebrity, to suit our ultimate task.

Ethics Discussion and Broader Perspective

Our study uses public news and Wikipedia data, and we ensure that our data collection process does not infringe upon any privacy or confidentiality concerns. A potential ethical concern is the misuse to track individual users (non-celebrities). However, as our framework relies on detailed text descriptions and background knowledge related to the person, the risk would stem from breaches of personal information rather than the framework itself. Therefore, we believe that our work does not raise essential ethical issues. Regarding a broader perspective, our framework opens up the possibility to perform large-scale and network-wise studies on the travels and interactions of celebrities including politicians and artists, with implications in politics, economics, arts, and cultures.

Conclusion and Future Work

This paper proposes a new task of celebrity trip detection and introduces CeleTrip, a first framework that effectively extracts these trips from news articles, by modelling text and external knowledge as graphs. We construct the largest celebrity trip dataset known so far, on which our model outperforms other baselines. To facilitate relevant research on trip extraction as well as any analysis of celebrity trips, we open source the framework, the time and location extraction tools and the trip dataset. As a possible future step, we consider analyzing the patterns of the celebrity trips, given many historical trips, and extending the framework to various types of documents.

Acknowledgments

The authors thank the editors and anonymous reviewers for their invaluable assistance in improving the quality of the paper.

References

- Akbari, M.; Hu, X.; Liqiang, N.; and Chua, T.-S. 2016. From tweets to wellness: Wellness event detection from twitter streams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- Alam, F.; Joty, S.; and Imran, M. 2018. Graph based semi-supervised learning with convolution neural networks to classify crisis related tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.
- Atwood, J.; and Towsley, D. 2016. Diffusion-convolutional neural networks. *Advances in Neural Information Processing Systems*, 29.
- Badham, M. 2017. The Social life of artist residencies: Working with people and places not your own. *Seismopolite: Journal of Art and Politics*, 18: 1–6.
- Boianovsky, M. 2018. 2017 HES presidential address: Economists and their travels, or the time when JFK sent Douglass North on a mission to Brazil. *Journal of the History of Economic Thought*, 40(2): 149–177.
- Brace, P.; and Hinckley, B. 1993. Presidential activities from Truman through Reagan: Timing and impact. *The Journal of Politics*, 55(2): 382–398.
- Cavari, A.; and Ables, M. 2019. Going global: Assessing presidential foreign travel. In *Congress & the Presidency*, 306–329.
- Chang, A.; and Manning, C. D. 2012. SUTime: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, 3735–3740.
- Chen, D.; Lin, Y.; Li, W.; Li, P.; Zhou, J.; and Sun, X. 2020. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 3438–3445.
- Deng, S.; Rangwala, H.; and Ning, Y. 2019. Learning dynamic context graphs for predicting social events. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1007–1016.
- Dickinson, T.; Fernandez, M.; Thomas, L. A.; Mulholland, P.; Briggs, P.; and Alani, H. 2015. Identifying prominent life events on twitter. In *Proceedings of the 8th International Conference on Knowledge Capture*, 1–8.
- Doherty, B. J. 2009. POTUS on the road: International and domestic presidential travel, 1977-2005. *Presidential Studies Quarterly*, 322–346.
- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.
- Geburu, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Goldsmith, B. E.; Horiuchi, Y.; and Matush, K. 2021. Does public diplomacy sway foreign public opinion? Identifying the effect of high-level visits. *American Political Science Review*, 115(4): 1342–1357.

- Han, X.; Cao, S.; Lv, X.; Lin, Y.; Liu, Z.; Sun, M.; and Li, J. 2018. Openke: An open toolkit for knowledge embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 139–144.
- Hautala, J.; and Jauhiainen, J. S. 2019. Creativity-related mobilities of peripheral artists and scientists. *GeoJournal*, 84(2): 381–394.
- Kernell, S. 2006. *Going public: New strategies of presidential leadership*. Cq Press.
- Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.
- Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Lee, J.; Lee, I.; and Kang, J. 2019. Self-attention graph pooling. In *International Conference on Machine Learning*, 3734–3743. PMLR.
- Li, J.; Sun, A.; Han, J.; and Li, C. 2020. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1): 50–70.
- Li, R.; Chen, H.; Feng, F.; Ma, Z.; Wang, X.; and Hovy, E. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 6319–6329.
- Lingad, J.; Karimi, S.; and Yin, J. 2013. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd International Conference on World Wide Web*, 1017–1020.
- Middleton, S. E.; Middleton, L.; and Modafferi, S. 2013. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2): 9–17.
- Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Nguyen, D.; Al Mannai, K. A.; Joty, S.; Sajjad, H.; Imran, M.; and Mitra, P. 2017. Robust classification of crisis-related data on social networks using convolutional neural networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 632–635.
- Ostrander, I.; and Rider, T. J. 2019. Presidents abroad: The politics of personal diplomacy. *Political Research Quarterly*, 72(4): 835–848.
- Peng, H.; Li, J.; Gong, Q.; Song, Y.; Ning, Y.; Lai, K.; and Yu, P. S. 2019. Fine-grained event categorization with heterogeneous graph convolutional networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 3238–3245.
- Piskorski, J.; Zavarella, V.; Atkinson, M.; and Verile, M. 2020. Timelines: Entity-centric event extraction from online news. In *Text2Story@ ECIR*, 105–114.
- Sahu, S. K.; Christopoulou, F.; Miwa, M.; and Ananiadou, S. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4309–4316.
- Schich, M.; Song, C.; Ahn, Y.-Y.; Mirsky, A.; Martino, M.; Barabási, A.-L.; and Helbing, D. 2014. A network framework of cultural history. *Science*, 345(6196): 558–562.
- Sellers, P. J.; and Denton, L. M. 2006. Presidential visits and midterm senate elections. *Presidential Studies Quarterly*, 36(3): 410–432.
- Simonovsky, M.; and Komodakis, N. 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3693–3702.
- Toplak, K. 2017. Artists’ mobility in the EU: Between opportunities and impediments. *Dve domovini*, (46).
- UzZaman, N.; Llorens, H.; Allen, J.; Derczynski, L.; Verhagen, M.; and Pustejovsky, J. 2012. Tempeval-3: Evaluating events, time expressions, and temporal relations. *arXiv preprint arXiv:1206.5333*.
- Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Vashishth, S.; Sanyal, S.; Nitin, V.; and Talukdar, P. P. 2020. Composition-based multi-relational graph convolutional networks. In *International Conference on Learning Representations*.
- Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph attention networks. In *International Conference on Learning Representations*.
- Vrandečić, D.; and Krötzsch, M. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10): 78–85.
- Wang, D.; Liu, P.; Zheng, Y.; Qiu, X.; and Huang, X.-J. 2020a. Heterogeneous graph neural networks for extractive document summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6209–6219.
- Wang, K.; Shen, W.; Yang, Y.; Quan, X.; and Wang, R. 2020b. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3229–3238.
- Wu, P.; Koh, J.; and Chen, A. L. 2019. Event detection for exploring emotional upheavals of depressive people. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2086–2095.
- Yao, L.; Mao, C.; and Luo, Y. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7370–7377.
- Yen, A.-Z.; Huang, H.-H.; and Chen, H.-H. 2018. Detecting personal life events from twitter by multi-task lstm. In *Companion Proceedings of The Web Conference 2018*, 21–22.
- Yen, A.-Z.; Huang, H.-H.; and Chen, H.-H. 2019. Personal knowledge base construction from text-based lifelogs. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 185–194.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? [Yes, see the Introduction, Ethics Discussion and Broader Perspective, Conclusion and Future Work.](#)
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes, see the Abstract and Introduction.](#)
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? [Yes, see the Introduction, Experimental Results, Conclusion and Future Work.](#)
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? [Yes, see the Ethics Discussion and Broader Perspective.](#)
 - (e) Did you describe the limitations of your work? [Yes, see the Error Analysis.](#)
 - (f) Did you discuss any potential negative societal impacts of your work? [Yes, see the Ethics Discussion and Broader Perspective.](#)
 - (g) Did you discuss any potential misuse of your work? [Yes, see the Ethics Discussion and Broader Perspective.](#)
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? [Yes, see the Ethics Discussion and Broader Perspective. The data we use is public news and Wikipedia data.](#)
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes.](#)
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? [NA](#)
 - (b) Have you provided justifications for all theoretical results? [NA](#)
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? [NA](#)
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? [Yes, see the Experimental Results.](#)
 - (e) Did you address potential biases or limitations in your theoretical framework? [Yes, see the Experimental Results.](#)
 - (f) Have you related your theoretical results to the existing literature in social science? [NA](#)
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? [Yes, see the Introduction, Ethics Discussion and Broader Perspective, Conclusion and Future Work.](#)
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? [NA](#)
 - (b) Did you include complete proofs of all theoretical results? [NA](#)
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes, see the open-sourced repository <https://github.com/ZhangDataLab/CeleTrip>.](#)
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes, see the Experiments.](#)
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No, because our experiment is conducted using fixed random seed and controls all random numbers to ensure that our experimental results can be reproduced.](#)
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes, see the open-sourced repository <https://github.com/ZhangDataLab/CeleTrip>.](#)
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? [Yes, see the Experimental Results.](#)
 - (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? [Yes, see the Experimental Results.](#)
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
 - (a) If your work uses existing assets, did you cite the creators? [Yes, see the Method.](#)
 - (b) Did you mention the license of the assets? [No, because all the tools, algorithms and data we use are publicly available.](#)
 - (c) Did you include any new assets in the supplemental material or as a URL? [Yes, see the URLs in footnotes.](#)
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes, see the Ethics Discussion and Broader Perspective. The data we use is public news and Wikipedia data.](#)
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes, see the Ethics Discussion and Broader Perspective. The data we use is public news and Wikipedia data.](#)
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets

FAIR (see FORCE11 (2020))? [Yes, see the open-sourced repository https://github.com/ZhangDataLab/CeleTrip.](#)

- (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? [Yes, we follow the instructions and create a Datasheet.](#)
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
- (a) Did you include the full text of instructions given to participants and screenshots? *NA*
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? *NA*
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? *NA*
 - (d) Did you discuss how data is stored, shared, and de-identified? *NA*