Event Prediction and Our Collected Dataset

2022-01-25 09:30AM (KST), 2022-01-24 07:30PM (EST)

Huigyu Yang
Sungkyunkwan University
huigyu@skku.edu

Presentation Outline

- Event Prediction
- Related Paper Review
- Our Dataset

Event Prediction (1/3)

- Events refer occurrences in specific locations, time, and semantics that affects real-world

- Event prediction is the capability to forecast possibility of the event occurrences in the future

- The event prediction is challenging task across various domains due to incompleteness of our knowledge regarding the mechanisms driving the event occurrences

Event Prediction (2/3)

- Recent efforts on the event prediction are aims to utilize deep/machine learning, data mining, or pattern recognition on the large amounts of historical events data

- However, the event prediction struggles to build models that can learn from various characteristics of the event data
  - Time, location, semantics (i.e., the unique features of the data)

- The event prediction models are not necessarily to be predicting all three domains of time, location, and semantics simultaneously, but a part of them should be predicted by the models

Event Prediction (3/3)

- This results the prediction models to address the following challenges:
- Complex dependencies among the prediction outputs
  - Not only dependencies between historical data and predicted result but also correlation of predicted results
  - The predicted events can influence the further events
- Heterogeneous multi-output predictions
  - Multiple features of event needs to be predicted including its time, location, duration, intensity, etc.

Relate Paper Brief Review (1/2)

- Web traffic and access-time of mobile users are predicted using Long Short-Term Memory (LSTM) and self-exciting point processes [1]

- The traffic patterns are mapped with the dependencies on time, location, and popularity of webpages

- The access events are modeled by self-exciting point processes and their intensities are utilized for access-time prediction

- The predictions of both traffic and access-time are done by LSTM

[1] A. Abdulrahman, et al., "Predicting Mobile Users Traffic and Access-Time Behavior Using Recurrent Neural Networks," *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021.

Relate Paper Brief Review (2/2)

- The dataset for this work was collected for a period of one month over 1000 cell tower locations
- Each log in the dataset contains the subscriber ID, session start and end time, cell tower ID, website ID, and traffic volume in kilobytes

TABLE I: Features used in the prediction model

ID	Feature
1	previous accessed website $[1, N_{website}]$
2	location of the access $[1, N_{location}]$
3	duration of the access (min)
4	time of the day (min)
5	is weekend? $[T, F]$
6	time since last access for each website $[1, N_{website}]$ (min)
7	average time between consecutive accesses (min)
8	website popularity $[1, N_{website}]$ $[0, 1]$

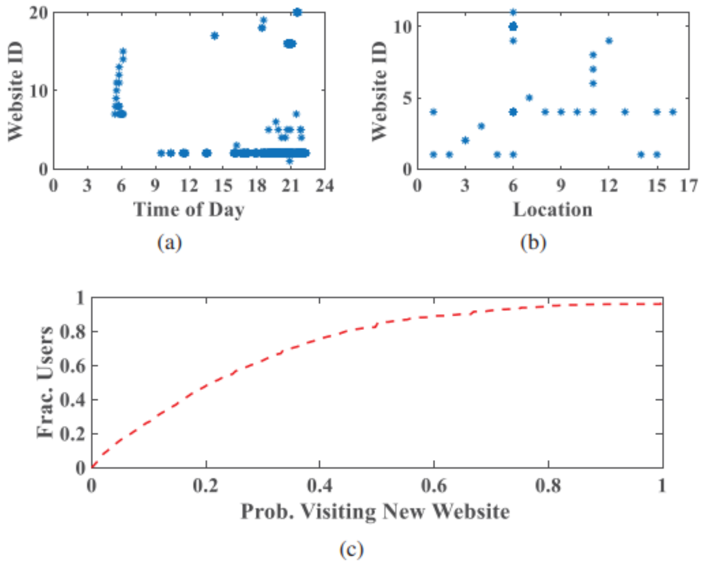
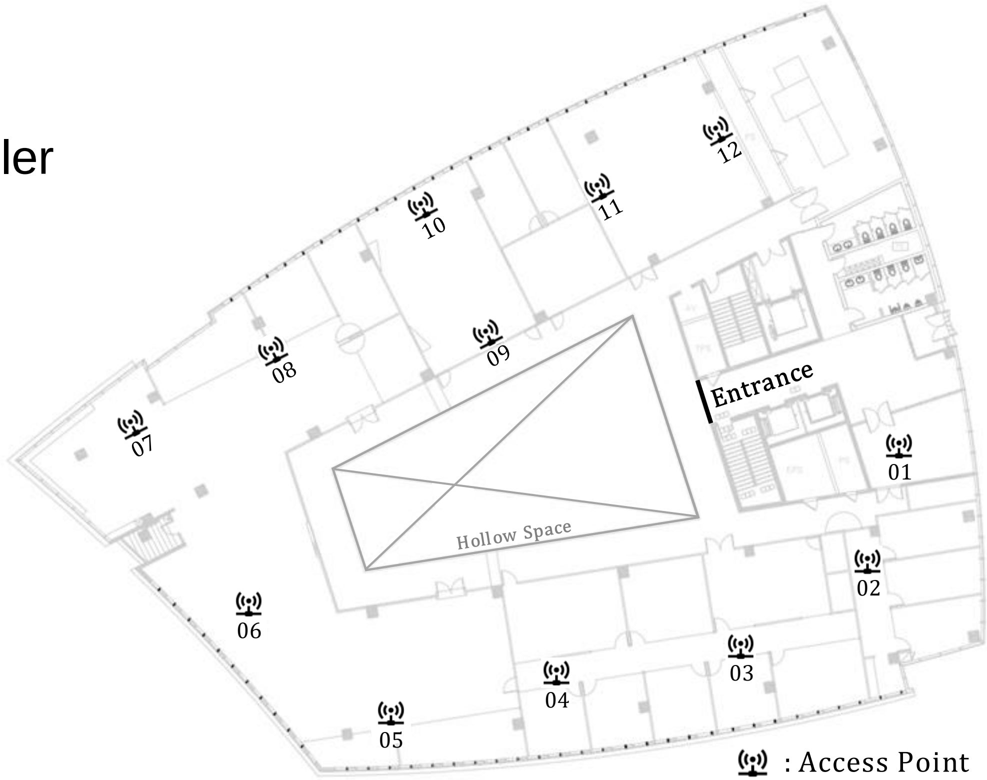


Fig. 1: Dataset analysis. (a) Absolute time of access; (b) Location of access; (c) Repeatability of websites visits

[1] A. Abdulrahman, et al., "Predicting Mobile Users Traffic and Access-Time Behavior Using Recurrent Neural Networks," *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021.

Our Dataset (1/2)

- Our mobility dataset is collected from wireless network
  - Intelligent ICT Convergence Research Center
- Cisco Access Point (AP)
  - 12 APs with proprietary controller
  - It has logging functionality
- Roaming Log Message
  - Time of occurrence
  - Terminal id
  - Source AP number
  - Destination AP number

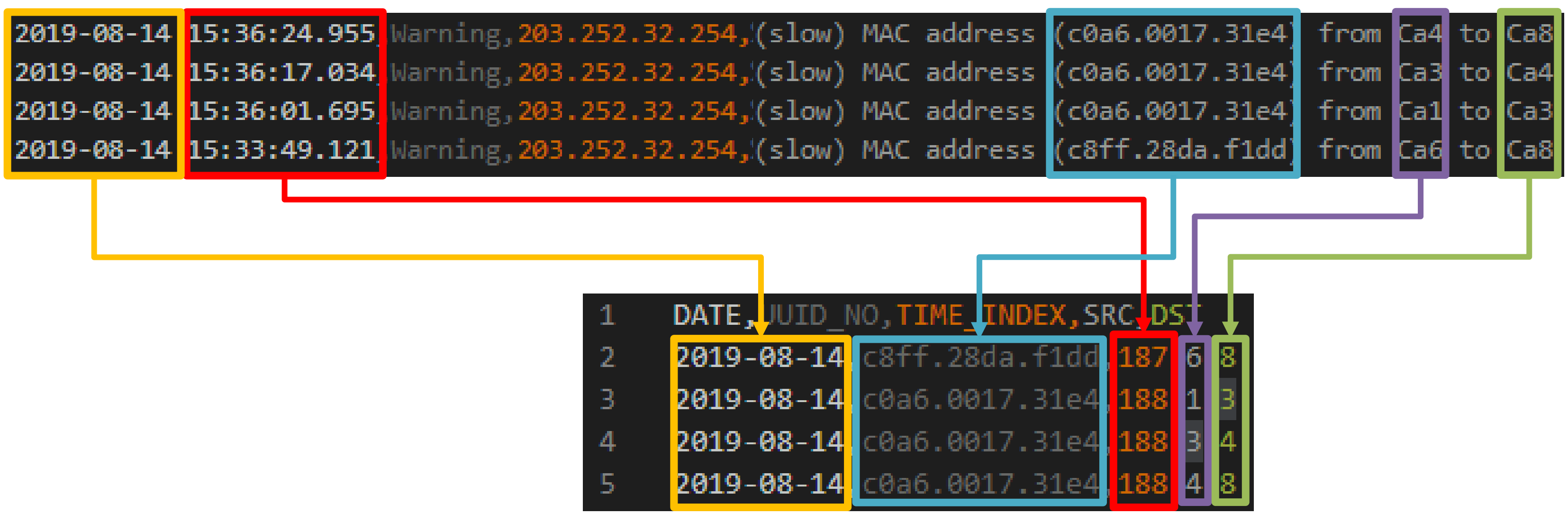


Our Dataset (2/2)

- Currently, the Pangyo dataset is in two different formats
- Pangyo original raw data

```
2019-08-14 15:36:24.955,Warning,203.252.32.254,(slow) MAC address (c0a6.0017.31e4) from Ca4 to Ca8
2019-08-14 15:36:17.034,Warning,203.252.32.254,(slow) MAC address (c0a6.0017.31e4) from Ca3 to Ca4
2019-08-14 15:36:01.695,Warning,203.252.32.254,(slow) MAC address (c0a6.0017.31e4) from Ca1 to Ca3
2019-08-14 15:33:49.121,Warning,203.252.32.254,(slow) MAC address (c8ff.28da.f1dd) from Ca6 to Ca8
```

- Converted raw data within timeslot



Time index refers that index of a day when 24 hours is divided into 5 mins

Appendix: result of the related paper

- The result of traffic activity (*i.e.*, next webpage access) predictions
  - Maximum 66% accuracy

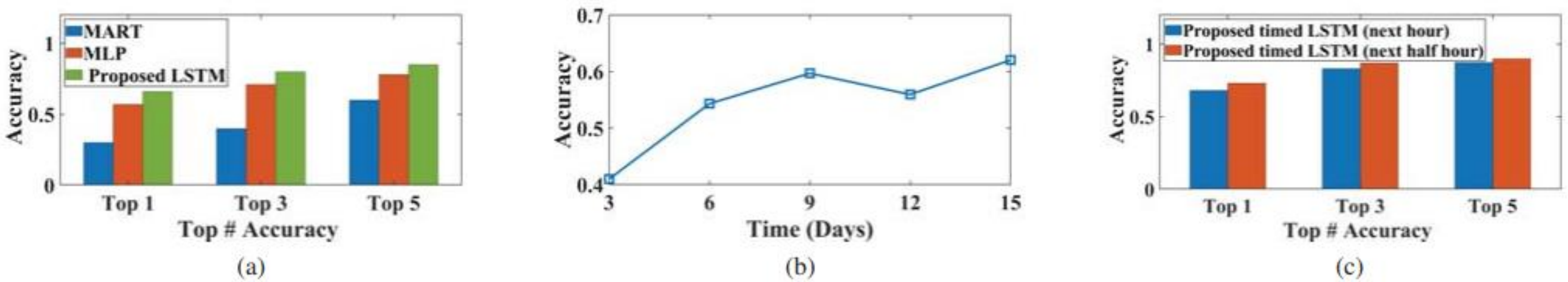


Fig. 6: Performance comparison of various user web-access modeling methods. (a) Proposed LSTM versus MART and MLP; (b) Proposed LSTM performance over time; (c) Proposed LSTM performance for both the next hour and next half hour.

Appendix: result of the related paper

- The result of access-time predictions

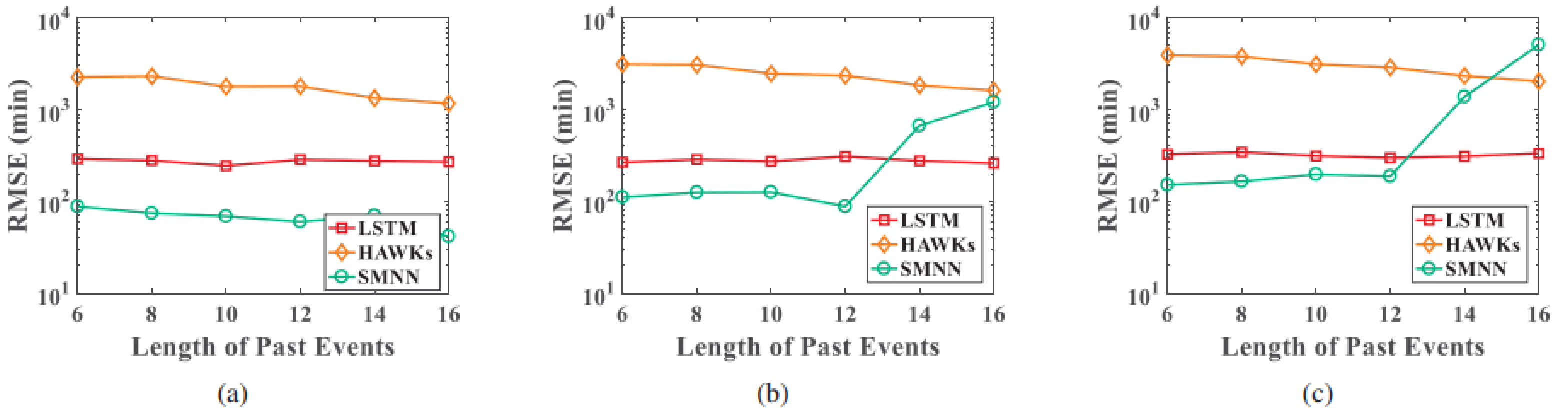


Fig. 7: Performance comparison of predicting access-time. (a) the next 1*st* access-time; (b) the next 2*nd* access-time; (c) the next 3*rd* access-time