PENNSTATE

1855

Mobile Data Management Meets Deep Learning

Wang-Chien Lee

Intelligent Pervasive Data Access (iPDA) Group

Pennsylvania State University

wlee@cse.psu.edu

PENNSTATE

Vision of Ubiquitous Computing

- Ubiquitous computing names the third wave in computing, just now beginning. First were mainframes, each shared by lots of people. Now we are in the personal computing era, person and machine staring uneasily at each other across the desktop. Next comes ubiquitous computing, or the age of calm technology, when technology recedes into the background of our lives.

  -- by Mark Weiser

- The most profound technologies are those that *disappear*. They wave themselves into the fabric of everyday life until they are indistinguishable from it.

Party on Friday…

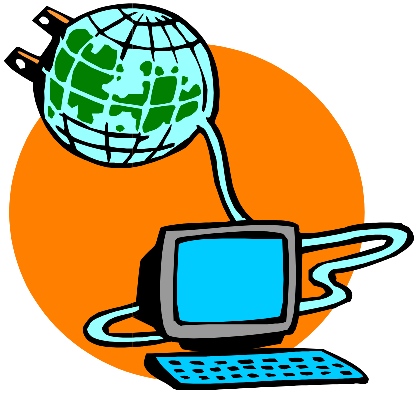
- Update Smart Phone's calendar with guests names.
- Make a note to order food from Dinner-on-Wheels.
- Update shopping list based on the guests drinking preferences.
- Don't forget to swipe that last can of beer's UPC/RFID label.
- The shopping list is always up-to-date.

Party on Friday…

- Approach a local supermarket
- AutoPC informs you that you are near a supermarket
- It informs you the soda and beer are on sale, and reminds you that your next appointment is in 1 hour.
- There is enough time based on the latest traffic report.

Party on Friday…

- TGIF…
- Smart Phone reminds you that you need to order food by noon.
- It downloads the Dinner-on-Wheels menu from the Web on your PC with the guests' preferences marked.
- It sends the shopping list to your CO-OP's PC.
- Everything will be delivered by the time you get home in the evening.



Mobile Data Management

- An important step proceeding the vision of Ubiquitous computing is *mobile computing*.
- The system and networking communities have Mobicom.
- There are needs for a forum to discuss and address research issues related to *data*, and other aspects…
- Prelude: *1998 Workshop on Mobile Data Access in Singapore.*
- Kick Off: *1999 International Conference on Mobile Data Management in Hong Kong.*

MDM Sessions – Early Years

1999	
	Wireless Networks and Communications
	Transaction Processing in Mobile Environments
	Ubiquitous Information Services
	Mobile Data Replication and Catching
	Mobility and Location Management
2001	
	Data Management Architectures
	Content Delivery
	Data Broadcasting
	Caching and Hoarding
	Coping with Movement
	Network and System issues
2002	
	Mobile and Disconnected Operation
	E-Commerce
	Data Allocation and Replication
	Moving Objects
	Location Management and Awareness

MDM Sessions – In Transition

2009	Location Data Management
	Mobile Peer-to-Peer Networks
	Embedded Devices and Applications
	Ad Hoc and Social Networks
	Sensor and Streaming Data Processing
	Location Based Services
	Mobile Data Dissemination and Access
	Location Privacy and Mining
	Mobile Peer-to-Peer Networks
2010	Localization and Location-Based Services
	GIS, Multimedia, and Storage
	Privacy and Trust Management
	Query Processing for Location-Based Services
	Wireless Networks
	Query Processing in Wireless Sensor Networks
	Moving Objects
2011	Location-Based Services and Query Optimization
	Moving Objects and Trajectories
	Mobility
	Personalization and Privacy
	Applications
	Vehicular and Mobile Networks
	Wireless Networks
	Pervasive Computing

MDM Sessions – Recent Years

2016	Information Management on Road Networks
	Query Processing and Information Search/Retrieval
	Smart City and Urban Applications
	Mining and Prediction for Streams and Moving Objects
	Social Media and Social Networks
	Ride Sharing, Road Networks and Routes
	Systems and Platforms
	Indexing and Querying: Road Networks, Moving Objects, and Trajectories
	Privacy and Security
2017	Location Services
	Mobile Data Processing
	Spatial+X Query Processing
	Ride Sharing and Recommendations
	Traffic Data Mining
	Connected Vehicles
	Localization and Traffic Analysis
	Trip Planning
	Trajectory Mining
2018	Trip Planning
	Data Mining and Machine Learning on Mobile Data 1
	Trajectory Mining
	Private Query Processing and Ride Sharing
	Mobile Data Processing
	Crowd Sourcing and LBSN

MDM Research Areas

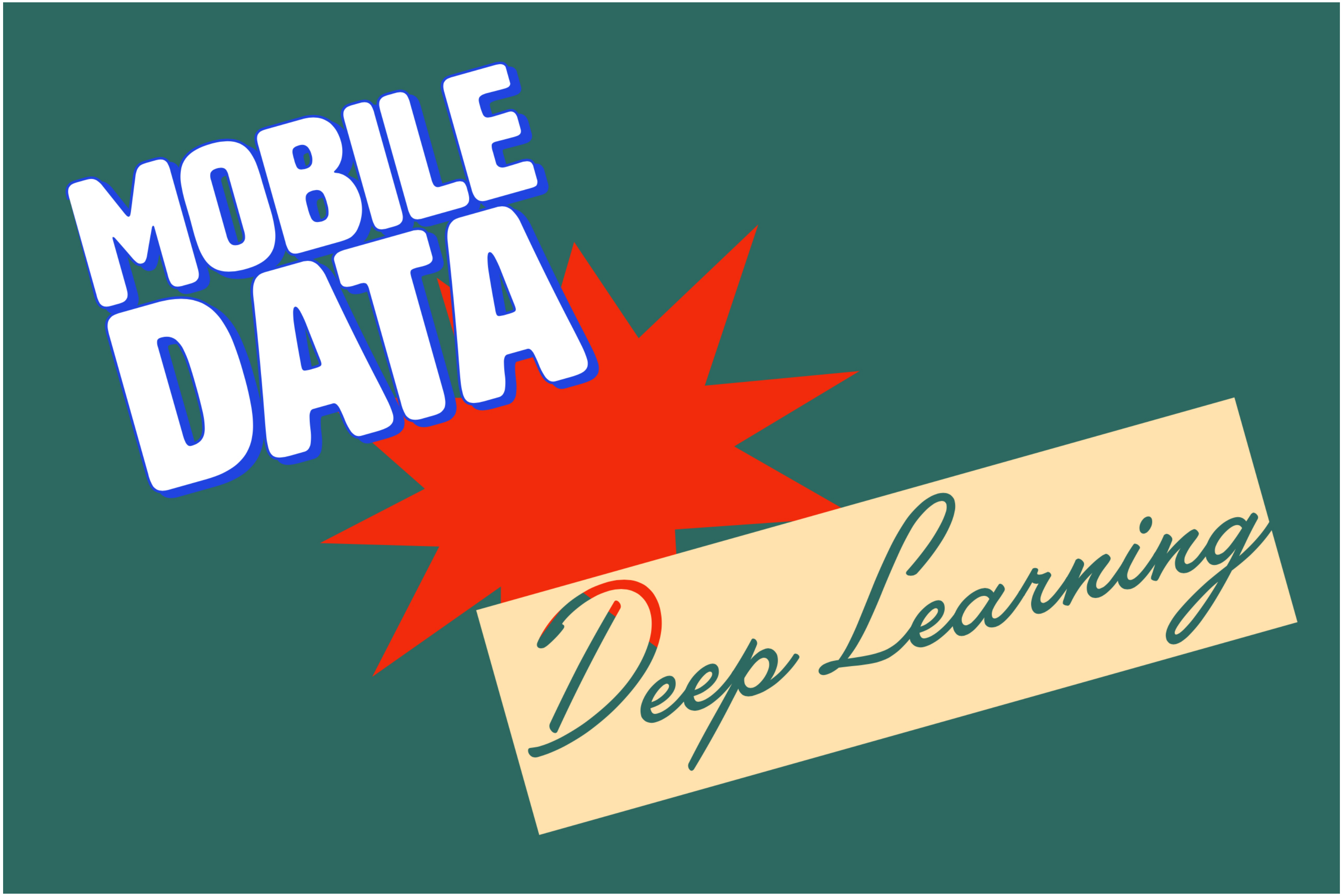
- Essential/Important Issues
  - Mobility and Location Management
  - Application, System and Network Issues
  - Mobile Data Processing, Query Processing
  - Privacy and Security
- Disappeared
  - Mobile Data Replication, Caching and Hoarding
  - Content Delivery, Data Broadcasting
- Emerging Topics
  - Smart City and Urban Applications, Trip Planning
  - Mining and Prediction for Streams and Moving Objects
  - Trajectory Mining, Traffic Data Mining, Ride Sharing and Recommendations

Ubiquitous Comp – Step Forward

- We are moving further towards the vision of Ubiquitous Computing
  - Abundant communication bandwidth
  - Abundant computing power
- Computing is becoming *Invisible*
  - Smart city, Smart building, Smart Vehicles
  - Smart watch, Smart Speakers, Smart applications
- We are in a process of *smartening* all the encounters in our daily life
  - Enabled by abundant data and machine learning, especially with the timely breakthrough of deep learning technology

Breakthroughs of Deep Learning

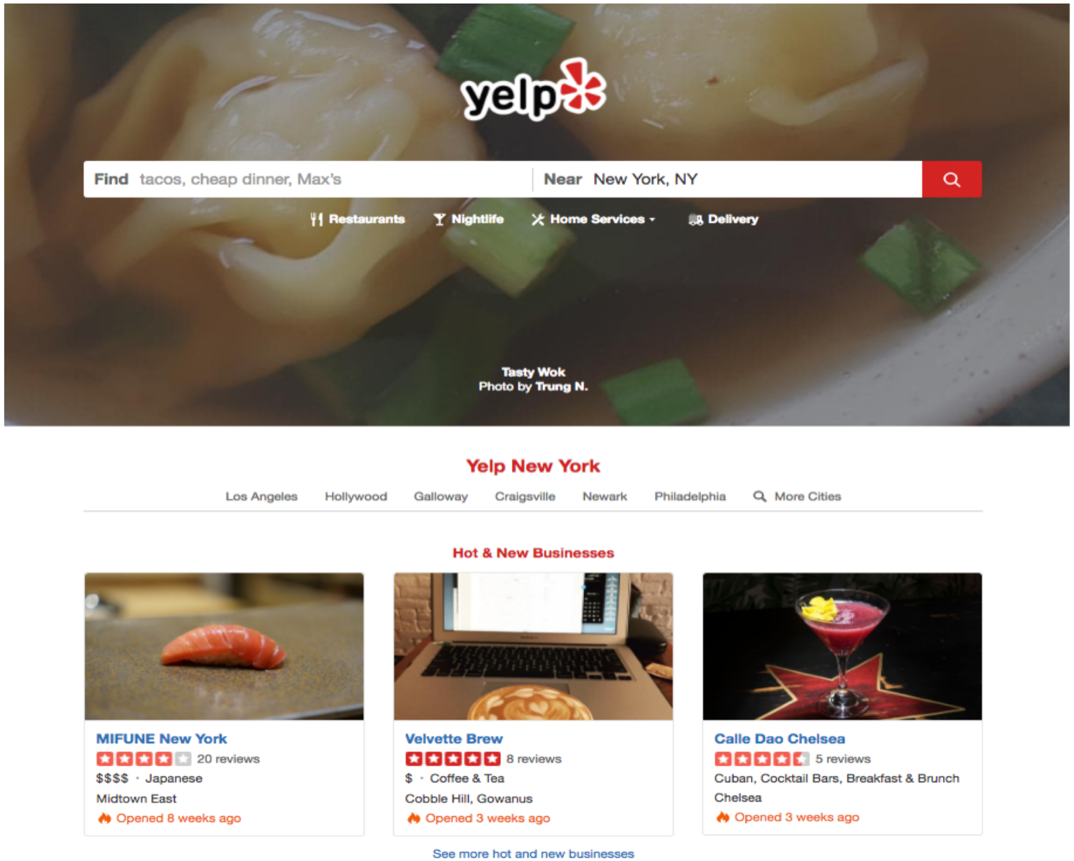
- In *2012*, AlexNet achieved 16% error rate in image classification on ImageNet. Then, VGG, GoogleNet, ResNet further improves to 7.3%, 6.7%, 3.5% compared with human average error 5%.
- In *2014*, DeepFace identifies faces with 97.35% accuracy, competitive with human performance.
- In *2016*, AlphaGo defeats a World Champ Lee Sedol (4:1) and is awarded an honorary 9-dan title.
- Models are proposed to various NLP apps, e.g., Word2Vec, Seq2Seq, Transformer. In *2018*, BERT obtains state-of-the-art results on 11 NLP tasks, described as the “Imagenet moment for NLP”.

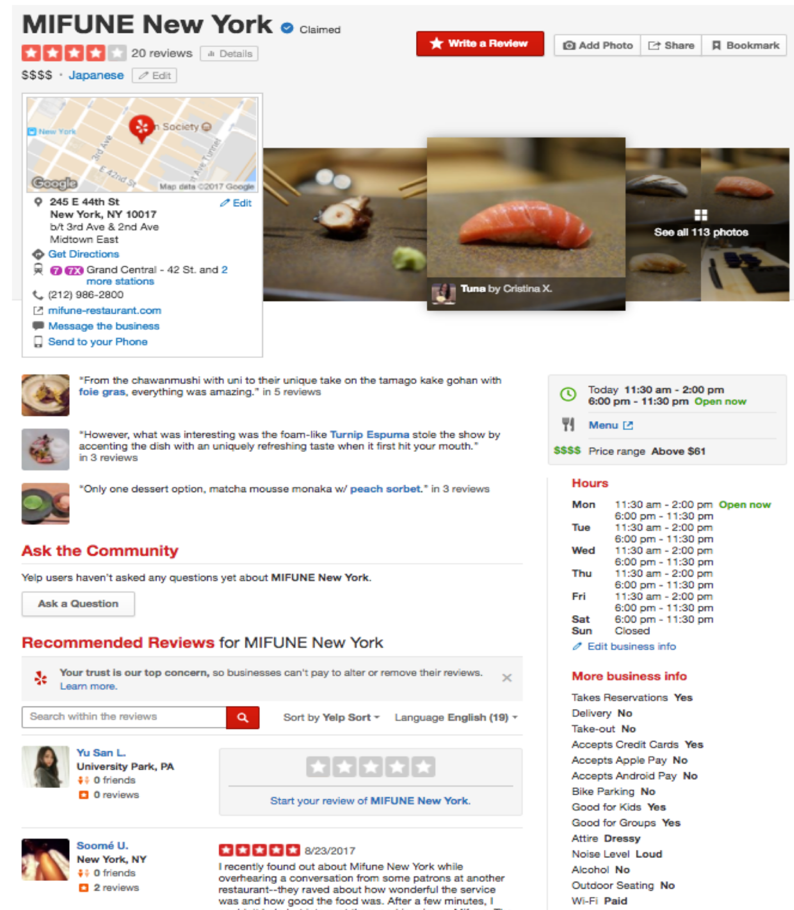

MOBILE
DATA
Deep Learning

Potential Research

- Location Based Social Networks
  - Network representation learning
- Trajectory Mining
  - Trajectory representation learning
  - Travel time estimation
- Intelligent Transportation Systems
  - Traffic Incident Inference
  - Traffic forecast
  - Traffic Sign Recognition

Location-Based Social Networks





Yelp Dataset

1M users

- user_id, name, review_count, yelping_since, friends, useful, funny, cool, fans, elite, average_stars, compliment_hot, compliment_more, compliment_profile, compliment_cute, compliment_list, compliment_note, compliment_plain, compliment_cool,compliment_funny, compliment_writer, compliment_photos

946K tips

- user_id, business_id, text, likes

144K restaurants

- business_id, name, neighborhood, address, city, state, postal_code, lng, lat, stars, review_count, is_open, attributes: [parking, payments, ...], categories: [tags], hours

125K check-ins

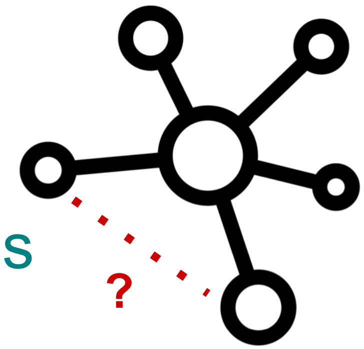
- business_id, time: [(time, count)]

4.1M reviews

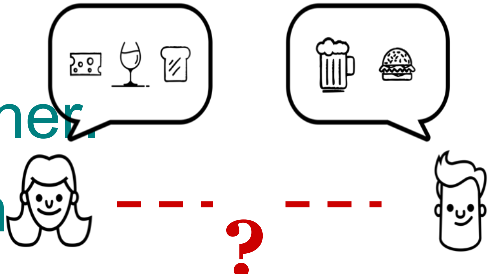
- review_id, user_id, business_id, star, date, text, useful, funny, cool

Functionality

- Restaurant search:
  - Given a restaurant, recommend *similar* restaurants
  - Formulate as k-nearest neighbor (KNN) search problem
- Personalized restaurant recommendation:
  - Given a user, recommend restaurants of her interests
  - Formulate as a link prediction problem



- Restaurant categorization:
  - Given a restaurant, classify it into categories
  - Formulate as a classification problem
- Friendship recommendation:
  - Given a user, recommend new friends to her
  - Formulated as a similarity search problem



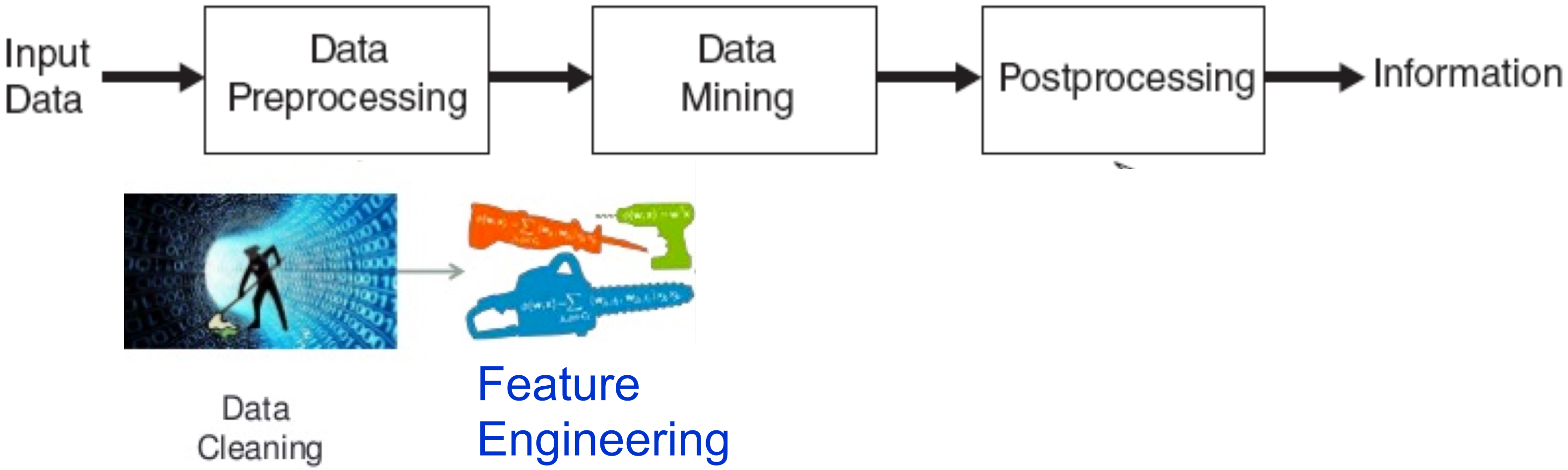
Data Mining on Network Data

Many applications of location based social network data and service functionality are formulated as classical data mining tasks:

- Node classification
  - Predict the type of a given node
- Link prediction
  - Predict whether two nodes are linked
- Clustering/Community detection
  - Identify densely linked clusters of nodes
- Similarity search
  - How similar/relevant are two nodes?
  - How similar are two (sub)networks

Automatic Feature Engineering

- Network data analytics often involve prediction tasks over nodes/edges. To achieve good performance, feature engineering is essential but <u>labor-intensive</u>.

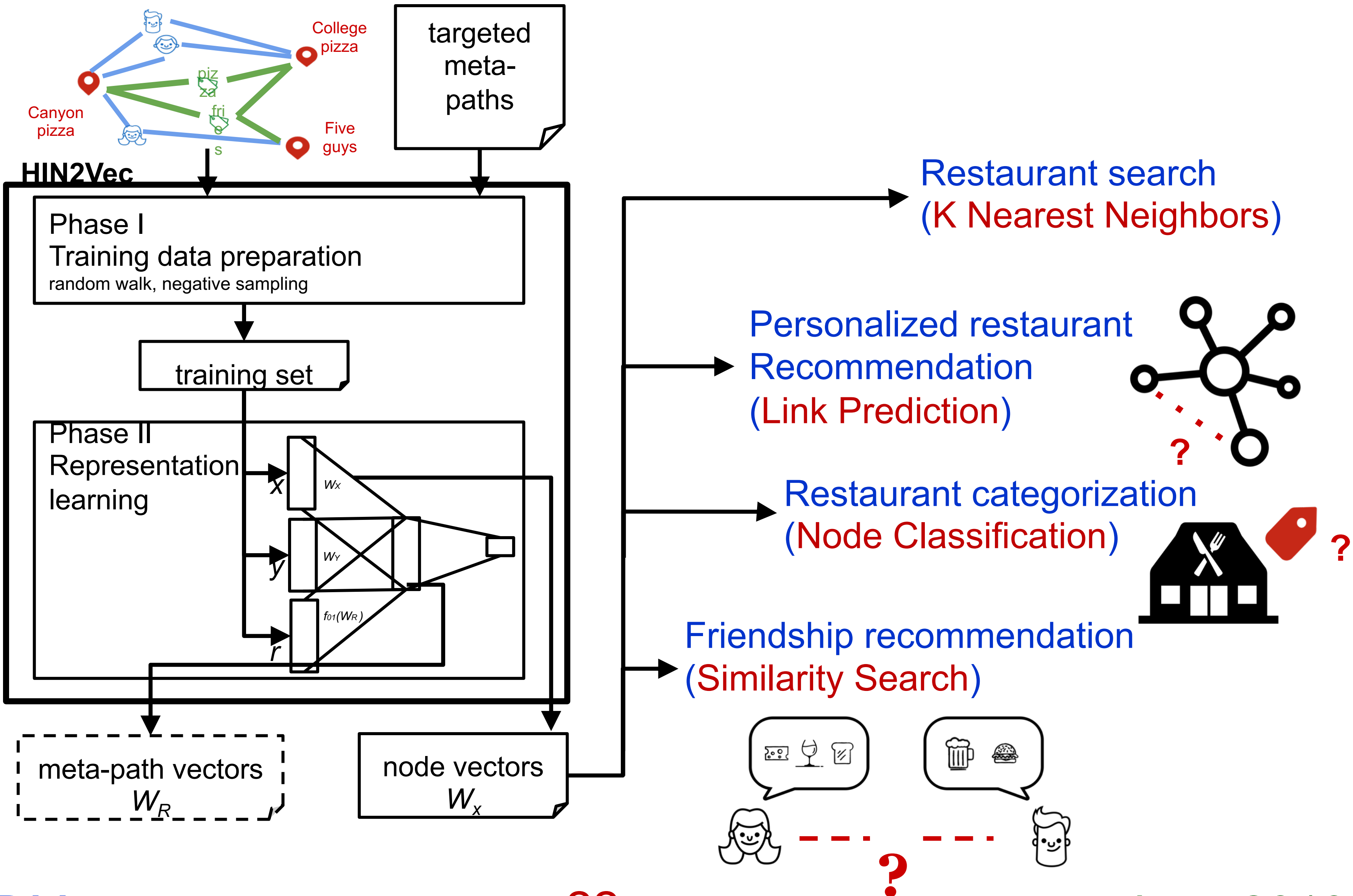


- Open problem: Efficient and automatic feature learning
  - Ideally, the learned features are task-independent!

HIN2Vec (Fu et al, CIKM'17)

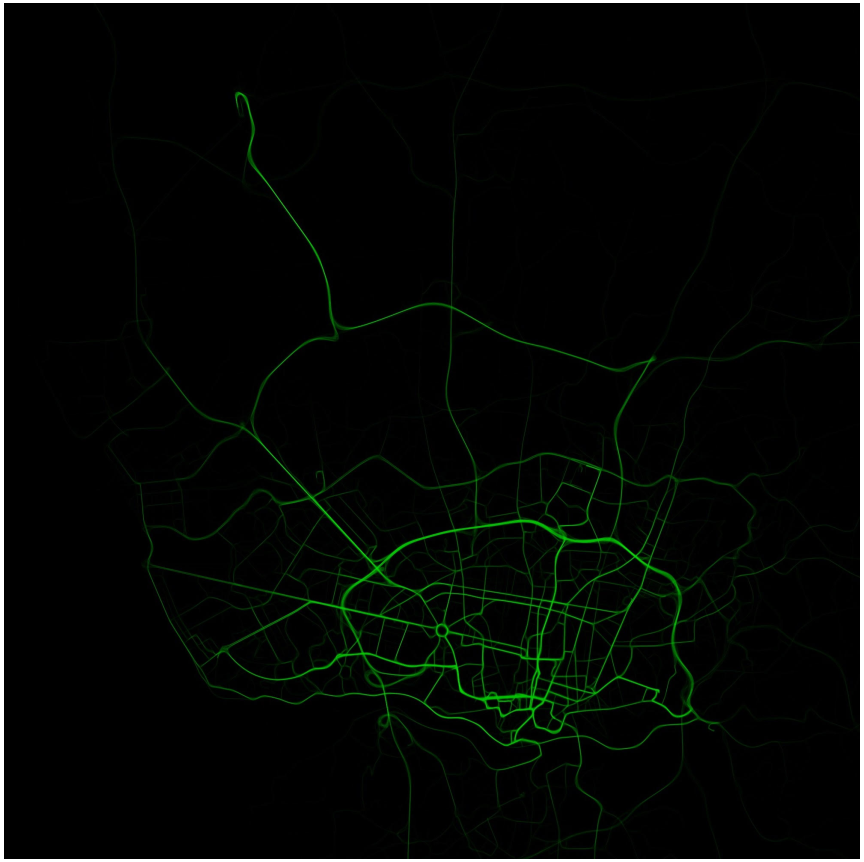
- To support a variety of LBSN applications, HIN2Vec automatically generates latent embeddings with inherent properties to serve as input features.
- HIN2Vec considers *heterogeneous data*
- HIN2Vec distinguishes the *different relationships* between nodes, and thus preserves more precise information
- HIN2Vec learns *meaningful representations* by encoding the rich information embedded in meta-paths and network structure.
  - Nodes with strong relationships are close to each other.
  - Relationship vectors provide analytical insights

HIN2Vec Framework



Trajectory Mining

- Many trajectory datasets made available publicly.
- Applications
  - Search for similar trajectories
  - Trajectory clustering
  - Travel time estimation
- Learned trajectory representations may be used for some applications.



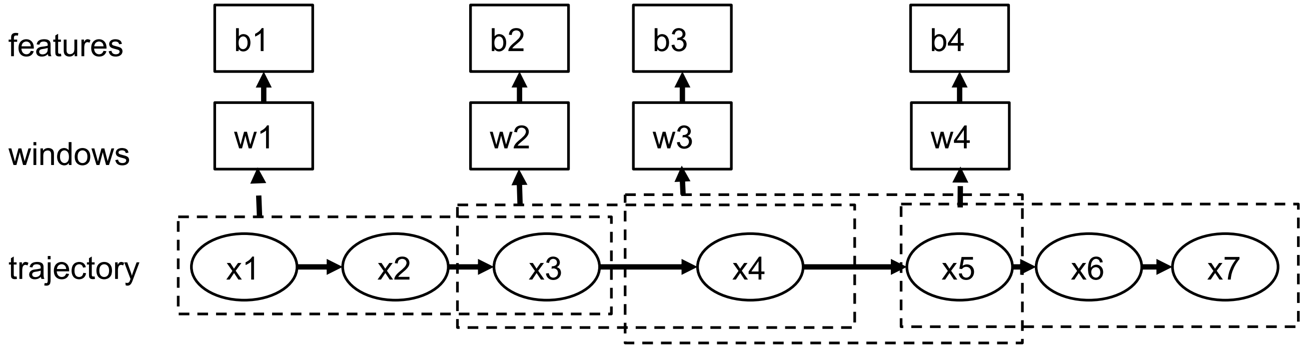
Porto taxi data, Taxi Service Trajectory Prediction Challenge@ ECML/PKDD 2015, contains 1.7 million taxi trajectories of 442 taxis in Porto, Portugal over 19 months.

Trajectory Representation Learning

- Trajectory Clustering
  - To learn trajectory embeddings by capturing mobile users' moving behaviors for trajectory clustering applications.
  - Yao, Di, et al. Trajectory clustering via deep representation learning, 2017 international joint conference on neural networks (IJCNN), 2017
- Trajectory Similarity Computation
  - To learn trajectory embeddings by capturing mobile users' moving behaviors for trajectory similarity computation.
  - X. Li, et al., Deep Representation Learning for Trajectory Similarity Computation, International Conference on Data Engineering (ICDE). 2018.

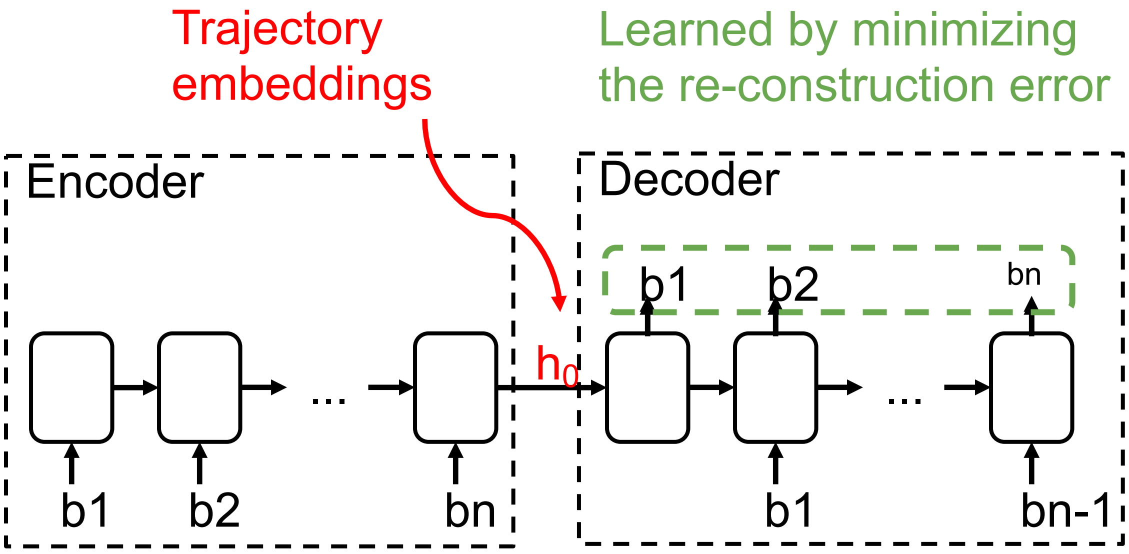
Trajectory2Vec

- Trajectory Preprocessing Layer
  - It applies existing techniques for data cleaning by filtering low-quality sample points
- Moving Behavior Feature Extraction Layer
  - It applies a sliding window to transform a raw trajectory as a sequence of windows containing sample points.
  - Generate a number of features (e.g., time interval, moving distance, change of speed, etc) for each window.



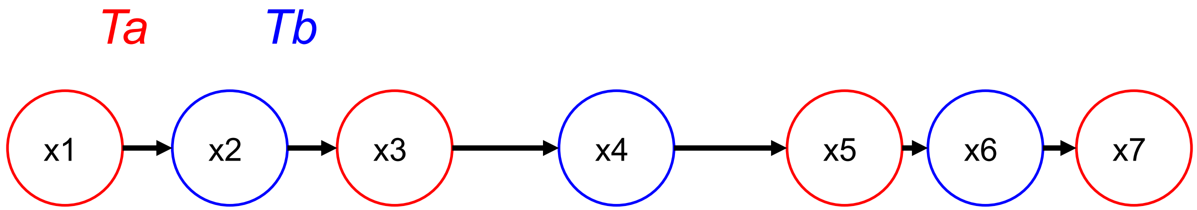
Seq2Seq Auto-encoder

- It applies Seq2Seq model to encode a trajectory (transformed as B={b1, b1, ...}) into a low-dimensional vectors which in turn is decoded back to the original B.



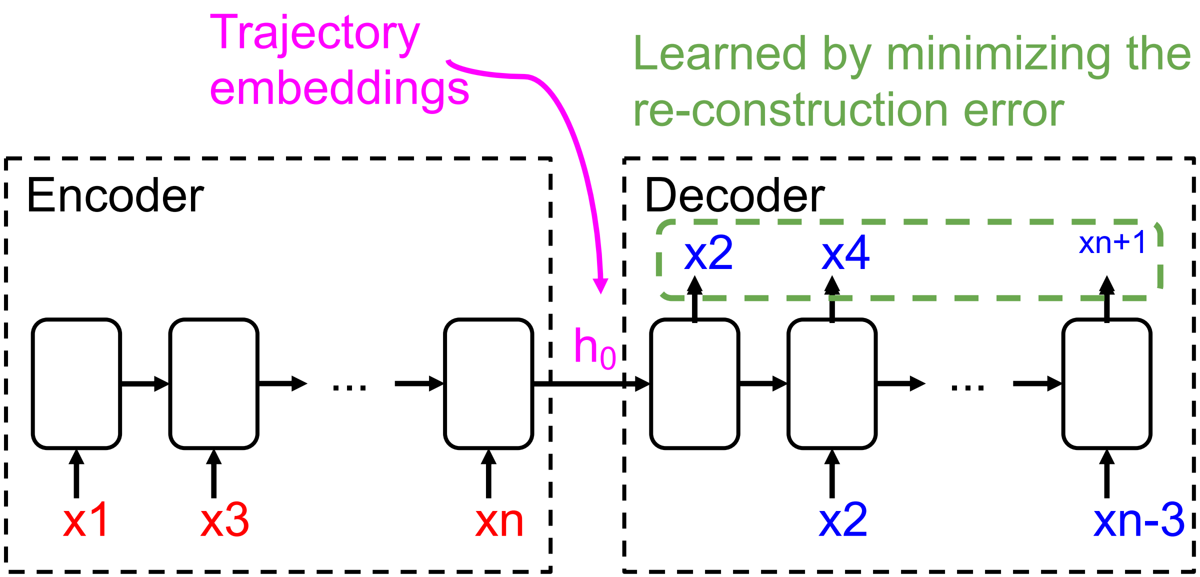
T2Vec - Data preprocessing

- For low sampling rate
  - For a trajectory T, t2vec splits it interleavingly to Ta and Tb (like downsampling)
  - Then, the proposed RNN-based encoder-decoder aims to encode Ta into a low-dimensional vector which is used to decode Tb
- For noisy data
  - It randomly adds more noises to sample data



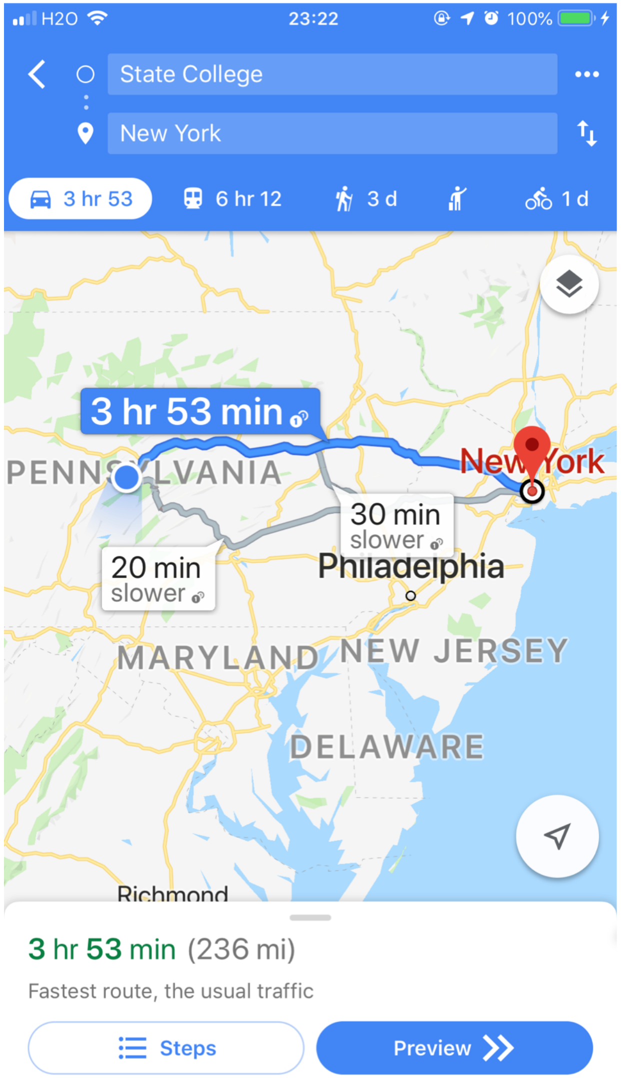
T2Vwec - Seq2Seq Auto-encoder

- Apply Seq2Seq model to encode Ta into a low-dimensional vector and then decode in turn to Tb



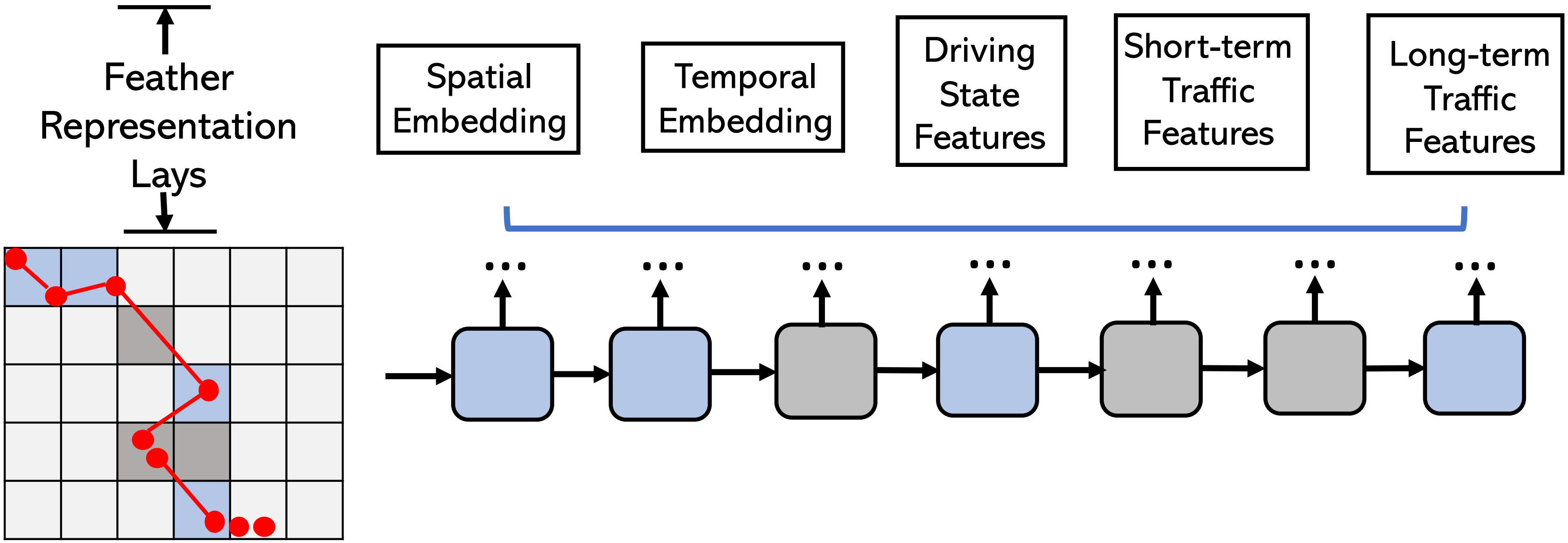
Travel Time Estimation

- Applications: Route planning, Navigation, Ridesharing and Traffic dispatching, etc.
  - H. Zhang, et al., Deeptravel: a neural network based travel time estimation model with auxiliary supervision, International Joint Conference on Artificial Intelligence (IJCAI-18).
  - D. Wang, et al., When Will You Arrive? Estimating Travel Time Based on Deep Neural Networks, AAAI Conference on Artificial Intelligence (AAAI-18).



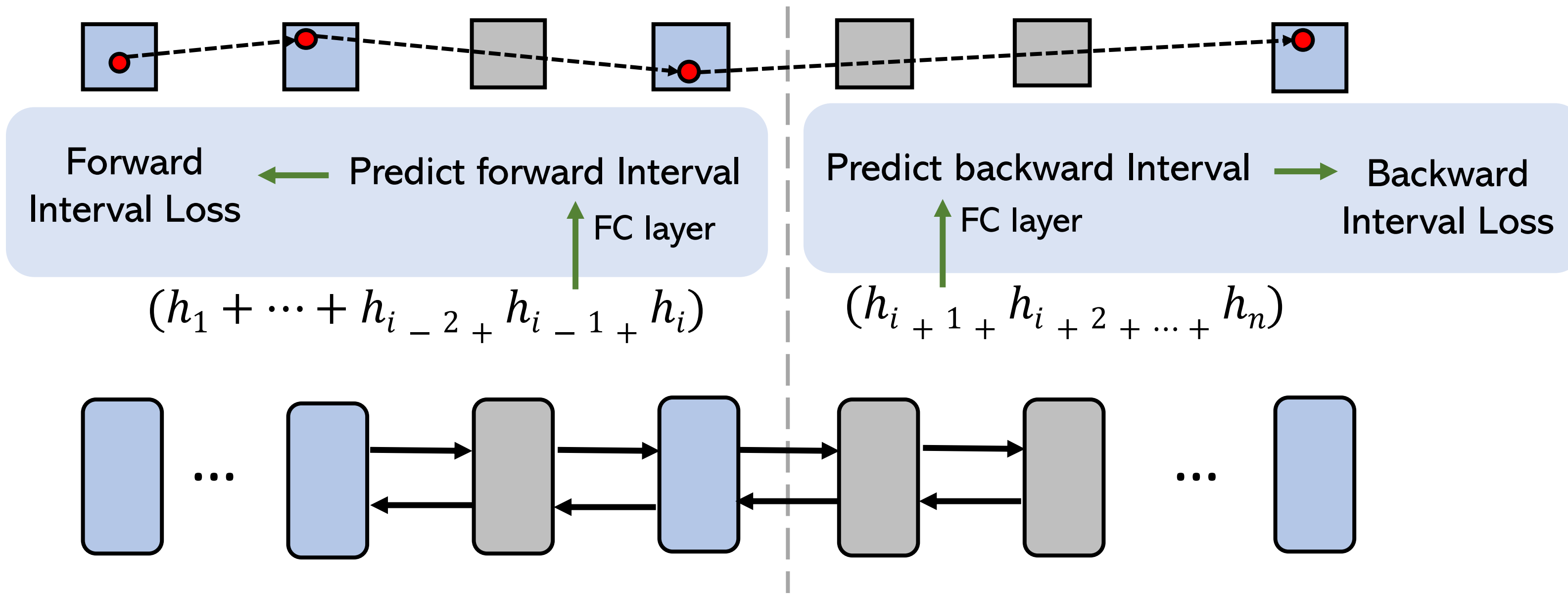
DeepTravel – Feature Extraction

- Partition a trajectory into a grid and map each GPS sample point into a grid cell.
- Extract features for each cell, including spatial and temporal embeddings, driving state features, short-term and long-term traffic features.



DeepTravel – Prediction

- The prediction layer consists of two parts.
  - BiLSTM: uses the extracted features to infer travel time
  - Dual loss: forces the model to learn by simultaneously predicting forward interval from the start point and backward interval from the destination to each intermediate GPS sample point.



DeepTTE – Model Architecture

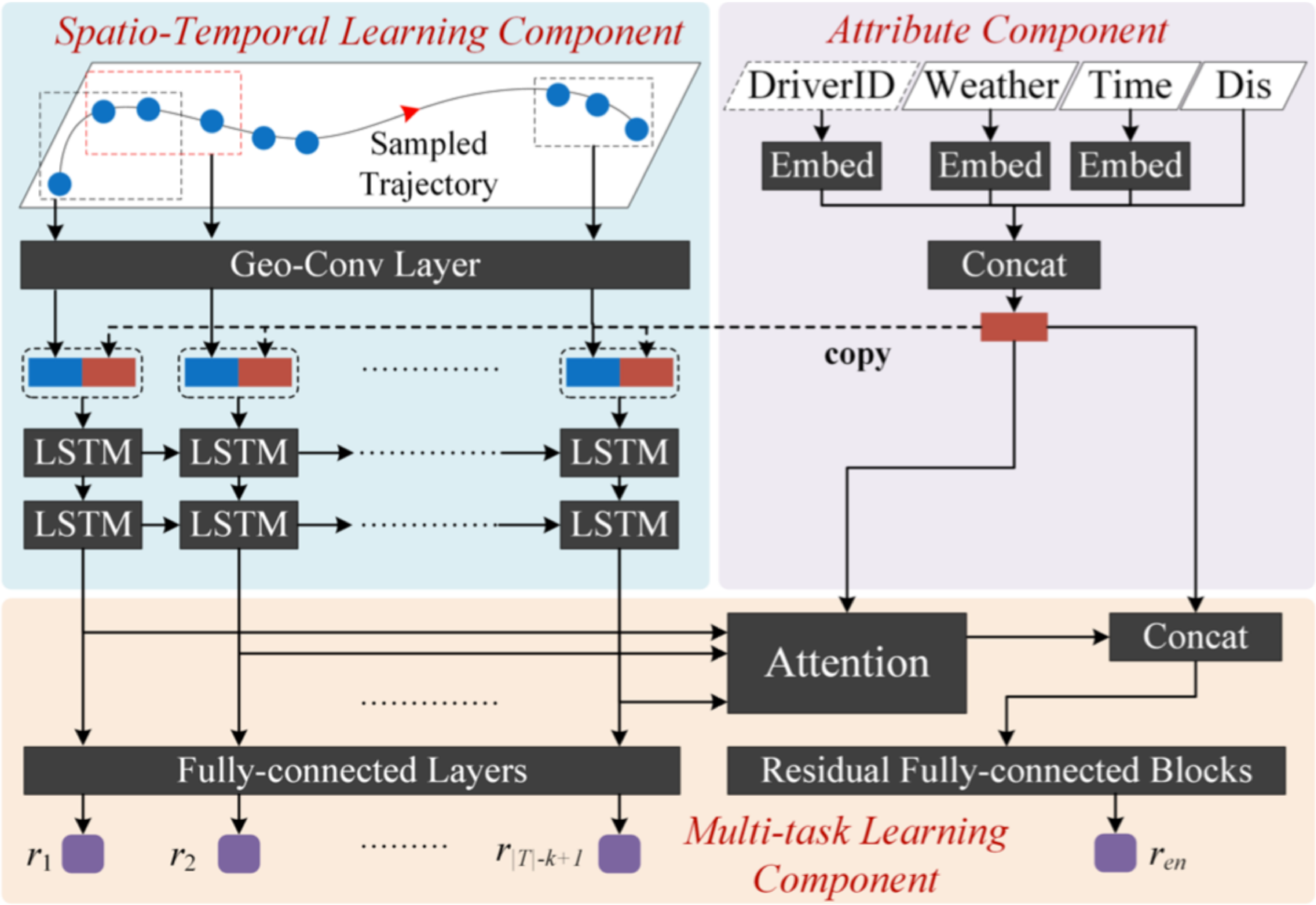
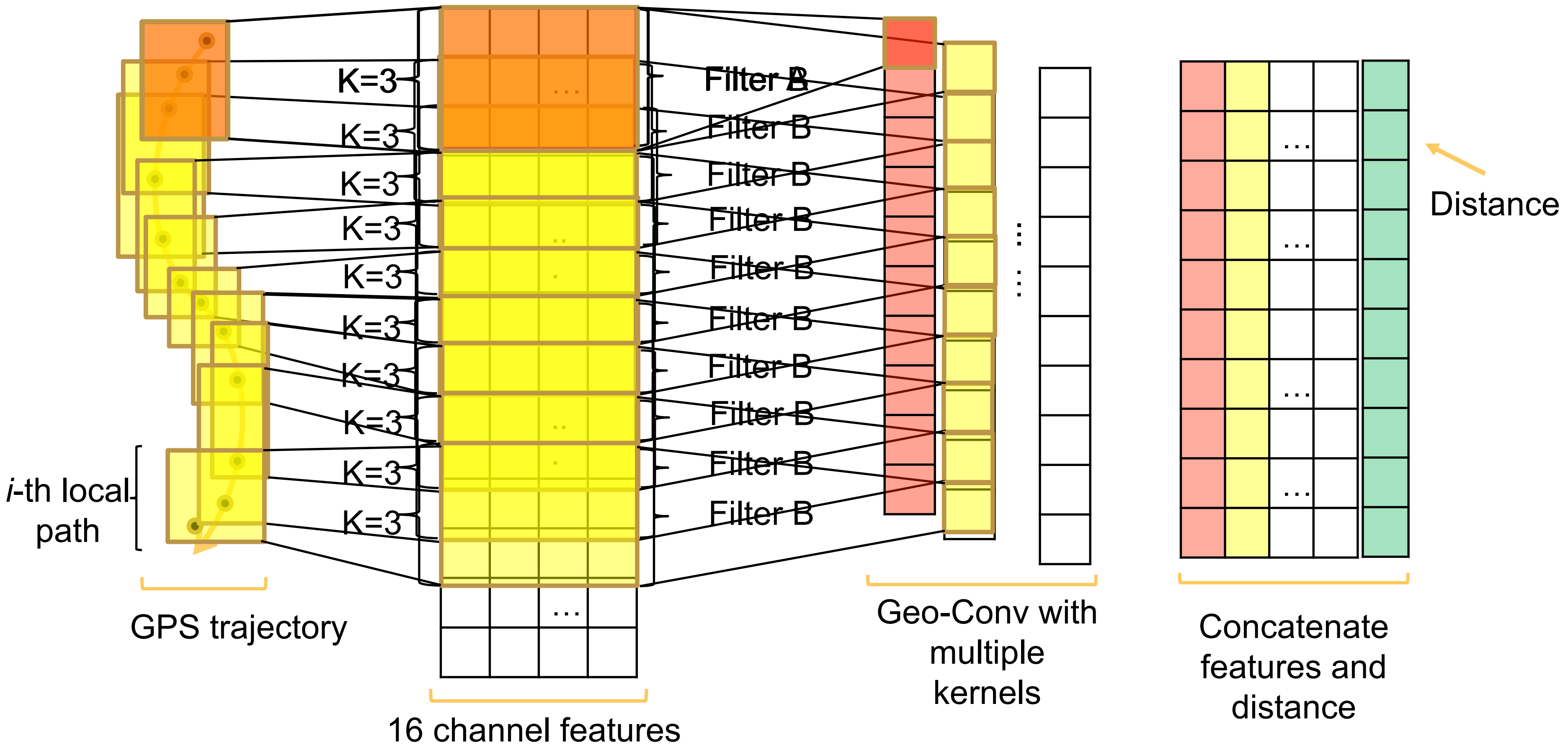
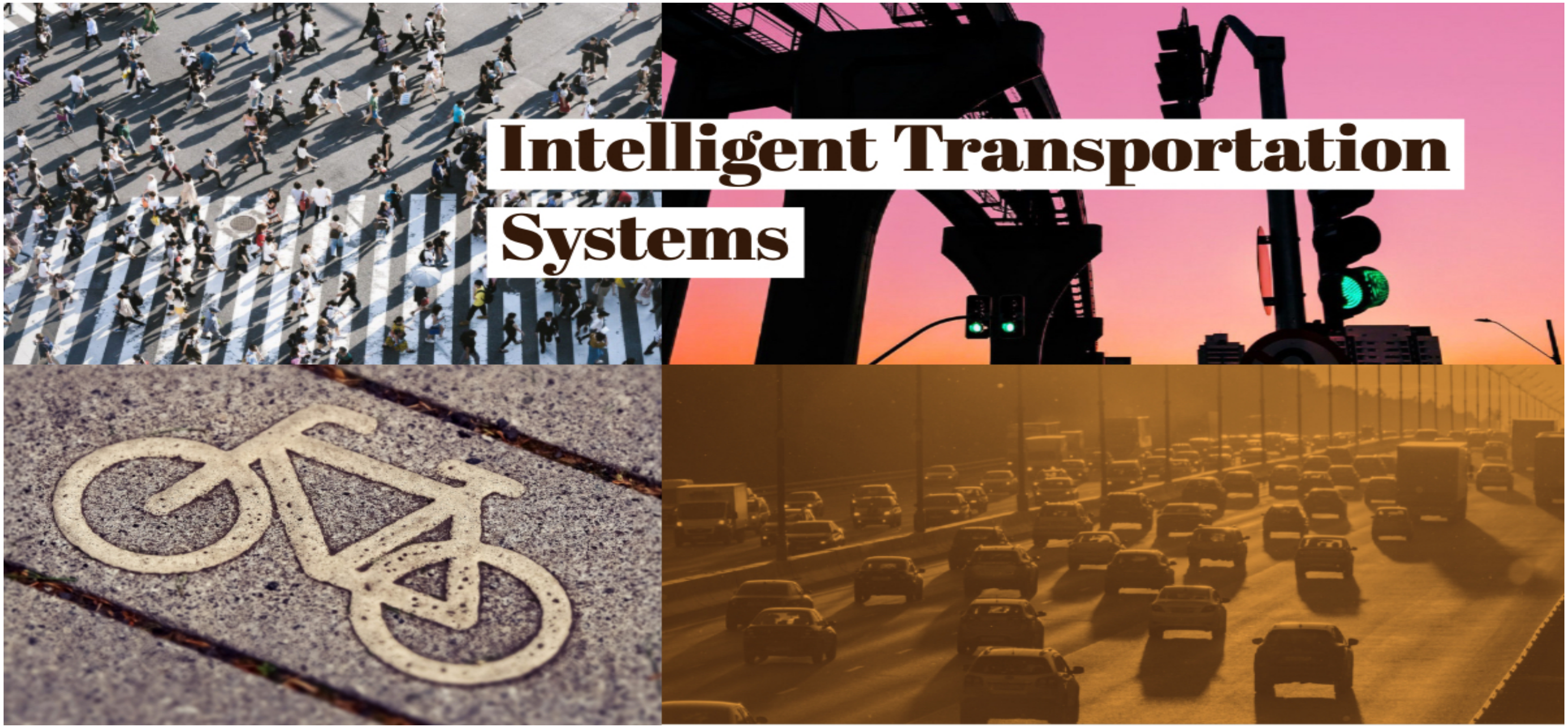


Figure from DeepTTE paper

DeepTTE – Geo-Convolution



Intelligent Transportation Systems



- Traffic Incident Inference
- Traffic Forecast
- Traffic Sign Recognition

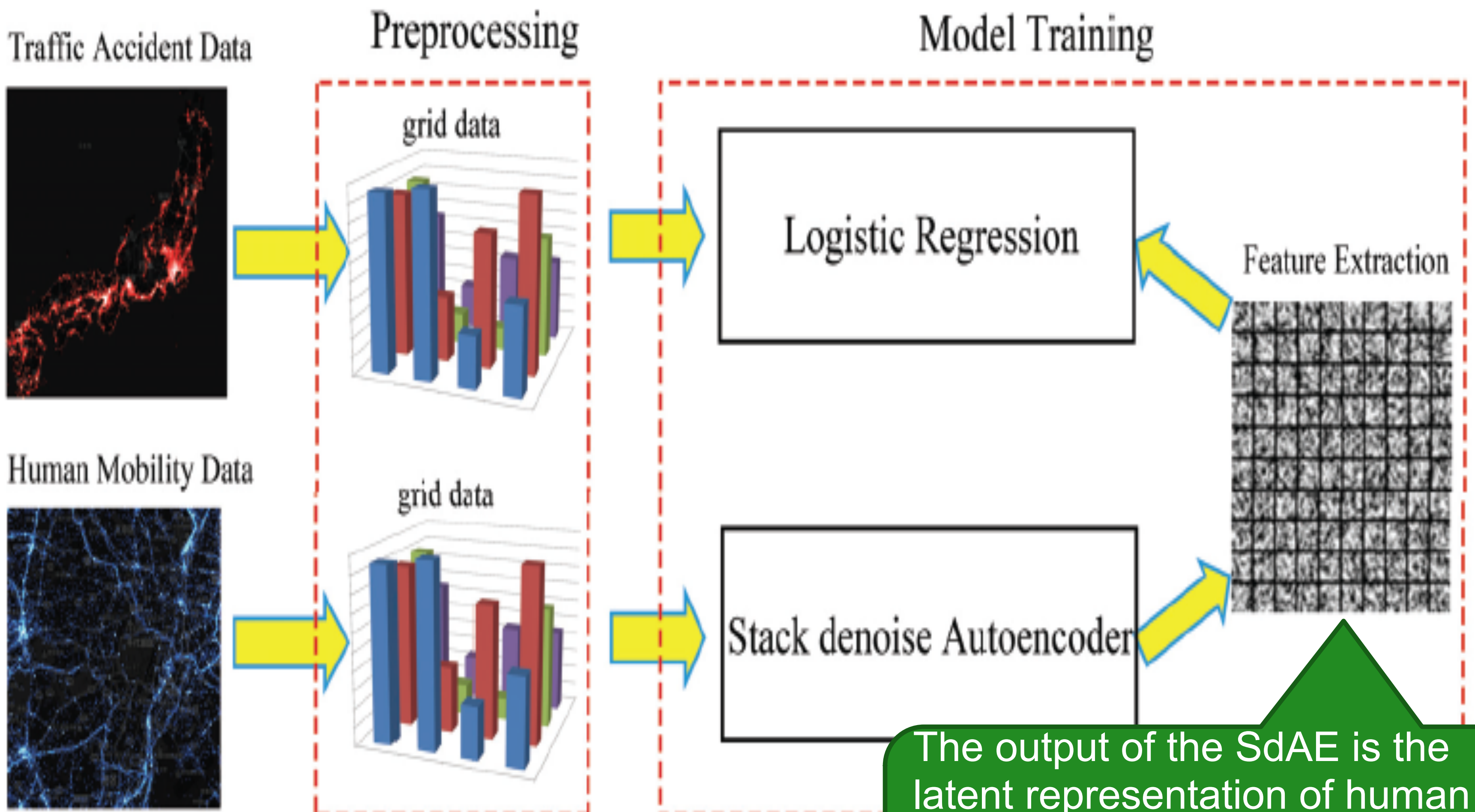
Traffic Incident Inference

Of all the systems with which people have to deal every day, road traffic systems are the most complex and dangerous.

World report on road traffic injury prevention, published by World Health Organization 2004.

- Estimate traffic accident risk by mining big and heterogeneous data.
- Q. Chen, et al., Learning deep representation from big and heterogeneous data for traffic accident inference’. AAAI, Toronto, Canada, 2016, pp. 338–344.

Accident Prediction Framework



The output of the SdAE is the latent representation of human mobility in each grid cell, which serve as input features to the classifier to predict accident risk

Traffic Forecast

- Predicting the most likely traffic measurements, e.g., speed, traffic flow, in the next H time steps, given previous M traffic observations.
- B. Yu, et al., Spatio-temporal graph convolutional neural network: a deep learning framework for traffic forecasting, International Joint Conference on Artificial Intelligence (IJCAI-18)

Graph-Structured Traffic Data

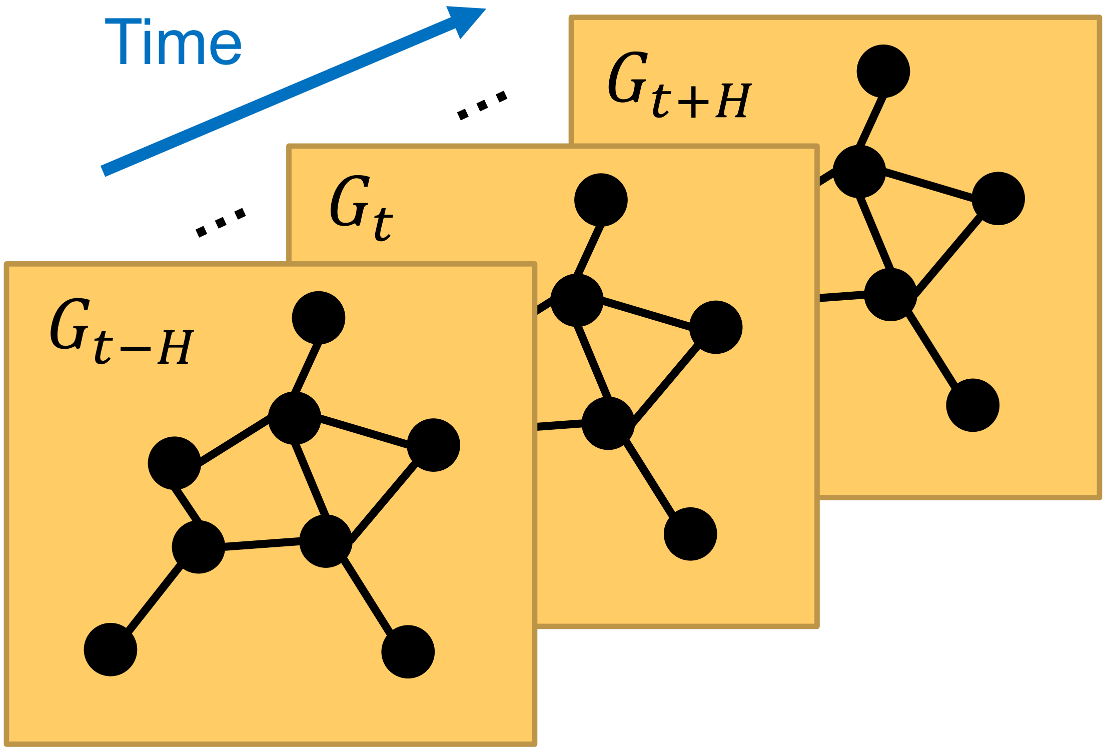
- To fully utilize spatial information, the traffic network is modeled by a general graph
- Temporal patterns of traffic flows are also important

$G_t = (V_t, E, W)$

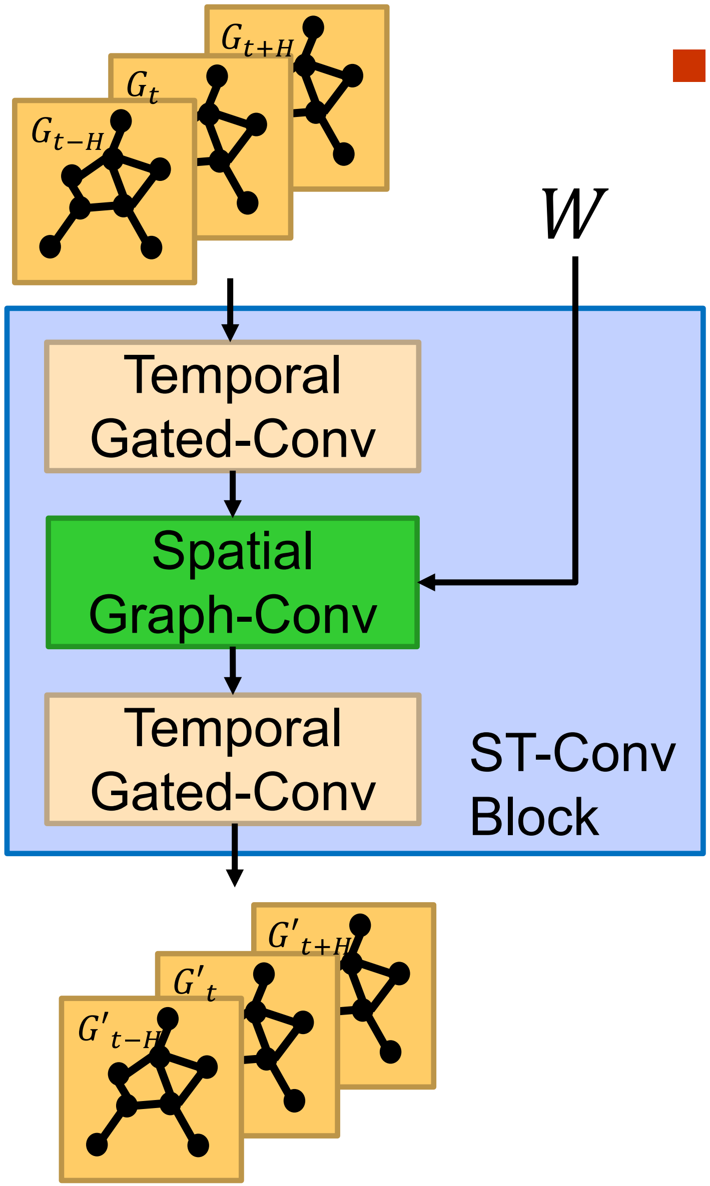
V_t: monitor stations

E: connectedness

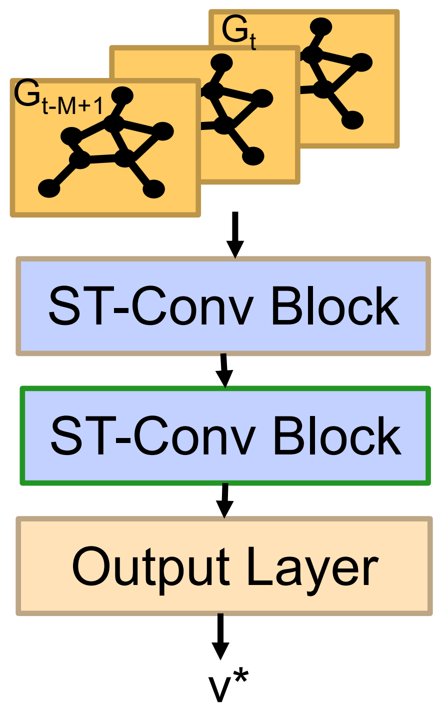
W: edge weights



STGCN Framework



- *ST-Conv Block* consists of:
  - *Graph Convolutional Network (GCN)* extracts meaningful spatial patterns and features
  - *Gated CNNs* captures temporal dynamic behaviors of traffic flow



Spatio-Temporal Graph Convolutional Networks

Traffic Sign Recognition

- Unmanned vehicle has attracted significant attention.
- Traffic sign recognition is an essential functionality for the upcoming unmanned vehicles.
- P. Sermanet and Y. LeCun, Traffic sign recognition with multi-scale convolutional networks, International Joint Conference on Neural Networks (IJCNN), 2011, pp. 2809–2813.
- J. Zhang, et al., A shallow network with combined pooling for fast traffic sign recognition. Information 8(2), 2017.

GTSRB Competition

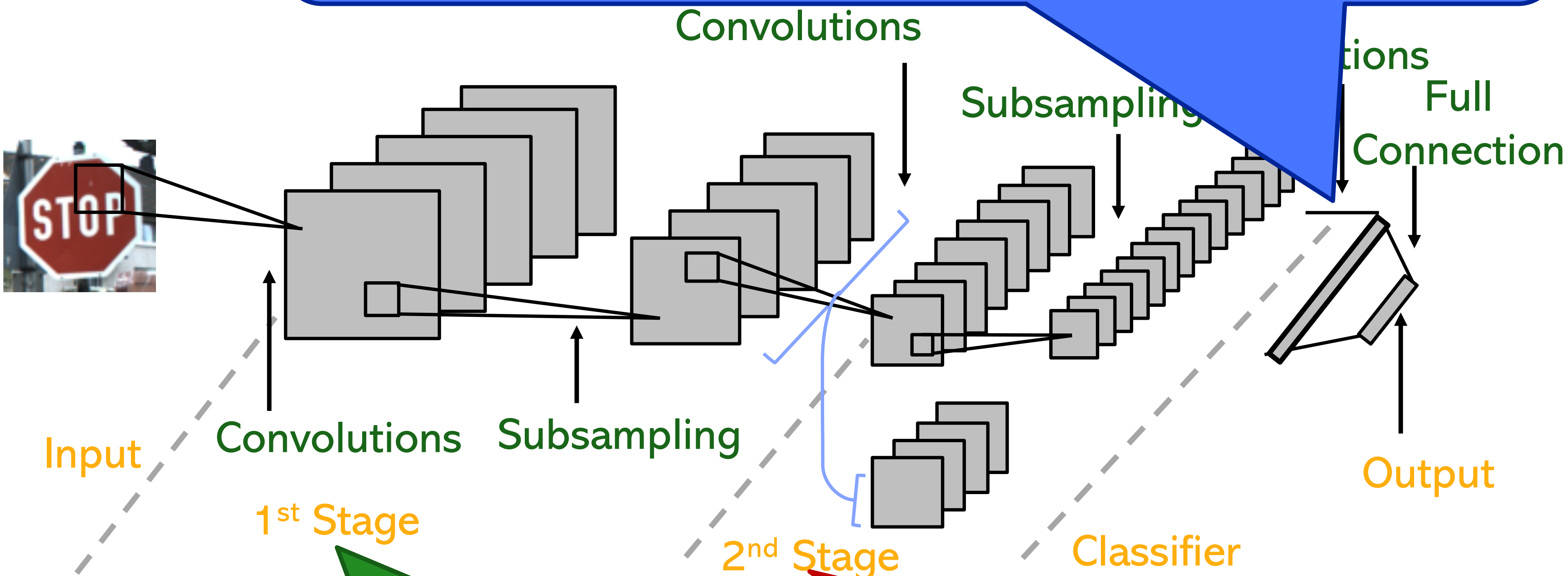
- The German Traffic Sign Recognition Benchmark

- The German Traffic Sign Detection Benchmark

- Challenging Examples

Two-stage ConvNet

The outputs of all the stages are fed to the classifier.
This allows the classifier to use not just high-level features but also pooled low-level features, which tend to be more local, less invariant, and more accurately encode local motifs.



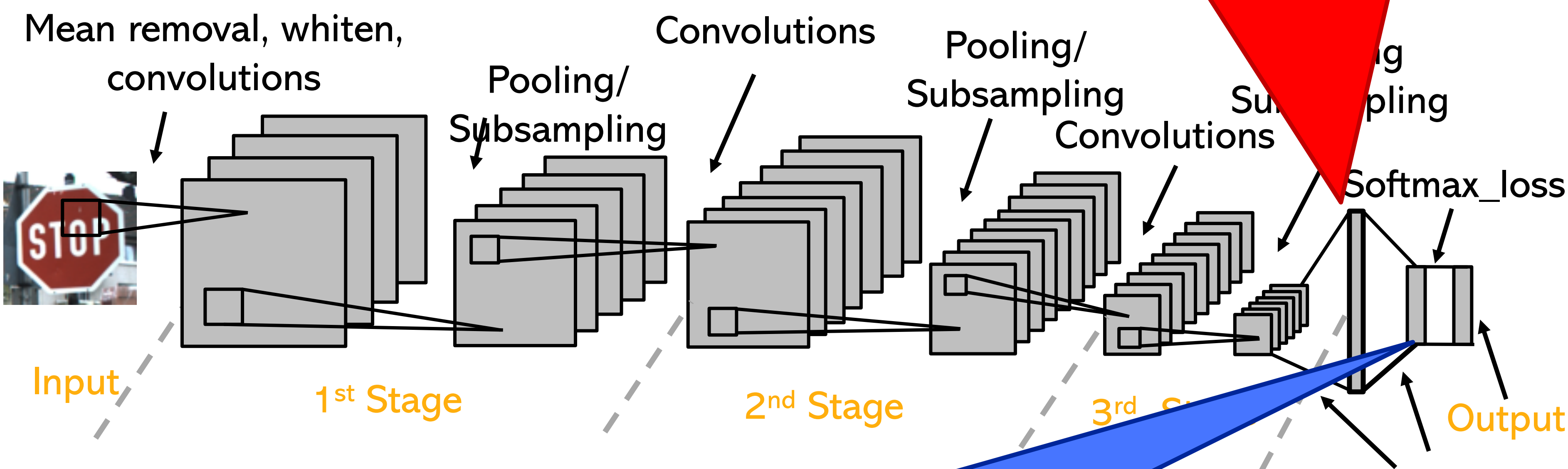
The first stage extracts “local” motifs with more precise details.

The 2nd stage extracts “global” and invariant shapes and structures.

Three-Stage Shallow CNNs

- Each stage includes a convolutional layer and a subsampling layer

The first fully connected layer is identical to a convolutional layer, aimed at reducing the dimensionality and preparing for the classification.



The second full-connected layer is similar to a single-hidden layer feedforward neural network (SLFN), and the output size is equal to the class number.

Conclusion

- We are moving a step further towards the vision of Ubiquitous Computing.
- Research in MDM has expanded from accessing mobile data, managing mobile data, to now smartening mobile applications.
- Recent breakthrough in deep learning technology brings opportunities and promises to many MDM research areas.
- Look forward to seeing blossom of research in the coming decade, when we celebrate the 30th Anniversary of MDM.