Event Prediction:
Dataset Related Paper Review

2022-03-09 09:30AM (KST), 2022-03-08 07:30PM (EST)

Huigyu Yang
Sungkyunkwan University
huigyu@skku.edu

Presentation Outline

- Citybike Dataset
- Dataset Related Keywords
- Crowd Flow Prediction

Citybike Dataset (1/2)

■ GPS and Check-in based log dataset: Citi Bike

[16]	New York City bikes	-	from 2013	New York City, USA	[116, 184, 187, 222]	Crowd Flow Pred.	BikeNYCData
[17]	Washington DC bikes	-	from 2010	Washington DC, USA	[184]	Crowd Flow Pred.	BikeWashington

Citi Bike Trip Histories

We publish downloadable files of Citi Bike trip data. The data includes:

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID
- End station name
- End station ID
- Start latitude
- Start longitude
- End latitude
- End Longitude
- Member or casual ride

Citybike Dataset (2/2)

■ GPS and Check-in based log dataset: Citi Bike

[16]	New York City bikes	- from 2013	New York City, USA	[116, 184, 187, 222]	Crowd Flow Pred.	BikeNYCData
[17]	Washington DC bikes	- from 2010	Washington DC, USA	[184]	Crowd Flow Pred.	BikeWashington

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
2	1589851B36BB0B5C	classic_bike	2022-01-07 12:56	2022-01-07 13:01	Adam Clayton Powell Blvd & W 126 St	7738.04	Frederick Douglass Blvd & W 139 St	7876.07	40.80949535	-73.94776493	40.81900582	-73.94476891	member
3	4C0BB6BD8AFCA917	classic_bike	2022-01-06 16:01	2022-01-06 16:06	Adam Clayton Powell Blvd & W 126 St	7738.04	Frederick Douglass Blvd & W 139 St	7876.07	40.80949535	-73.94776493	40.81900582	-73.94476891	member
4	765572ACD0D65972	classic_bike	2022-01-31 15:31	2022-01-31 15:36	E 56 St & Madison Ave	6732.01	E 48 St & 5 Ave	6626.01	40.761573	-73.972628	40.75724568	-73.97805914	member
5	86E8E7C4791EA81D	classic_bike	2022-01-21 17:38	2022-01-21 17:46	3 Ave & E 100 St	7414.17	E 85 St & 3 Ave	7212.05	40.7877214	-73.94728331	40.77801203	-73.95407149	member
6	D3B80E976AC4DBCF	classic_bike	2022-01-23 18:37	2022-01-23 18:43	E 88 St & Park Ave	7293.1	E 85 St & 3 Ave	7212.05	40.7814107	-73.95595908	40.77801203	-73.95407149	member
7	283D619884B13025	classic_bike	2022-01-13 12:17	2022-01-13 12:20	E 85 St & York Ave	7146.04	E 85 St & 3 Ave	7212.05	40.77536905	-73.94803392	40.77801203	-73.95407149	member
8	CC02A6C3FA1F2083	classic_bike	2022-01-11 9:09	2022-01-11 9:20	Broadway & Madison St	4483.1	Suydam St & Broadway	4689.03	40.68822	-73.91966	40.69544	-73.93223	member
9	968470449EEB57C0	classic_bike	2022-01-09 12:33	2022-01-09 12:50	Broadway & W 58 St	6948.1	E 85 St & 3 Ave	7212.05	40.76695317	-73.98169333	40.77801203	-73.95407149	member
10	16C601B498A35DBF	classic_bike	2022-01-14 8:28	2022-01-14 8:33	Lenox Ave & W 111 St	7602.05	E 102 St & Park Ave	7488.24	40.7987859	-73.9523	40.7904828	-73.95033068	member
11	7E477AA47C849CE5	classic_bike	2022-01-19 13:27	2022-01-19 13:40	Hancock St & Bedford Ave	4255.05	Suydam St & Broadway	4689.03	40.68216564	-73.95399026	40.69544	-73.93223	member
12	4DC494327327DD7E	classic_bike	2022-01-31 13:17	2022-01-31 13:28	E 97 St & Madison Ave	7393.09	E 85 St & 3 Ave	7212.05	40.787801	-73.953559	40.77801203	-73.95407149	member
13	2FF6406688A54185	classic_bike	2022-01-27 12:45	2022-01-27 12:53	E 97 St & Madison Ave	7393.09	E 85 St & 3 Ave	7212.05	40.787801	-73.953559	40.77801203	-73.95407149	member
14	7BA60209DC4F5607	classic_bike	2022-01-04 6:54	2022-01-04 6:58	5 Ave & E 135 St	7769.06	Park Ave & E 124 St	7682.01	40.812191	-73.937838	40.8045555	-73.9396861	member
15	4E4ADEA1887ACAC3	classic_bike	2022-01-04 18:38	2022-01-04 18:48	E 91 St & 2 Ave	7286.01	Park Ave & E 124 St	7682.01	40.78115276	-73.94963041	40.8045555	-73.9396861	member
16	2DB47657BEC9445B	classic_bike	2022-01-09 14:11	2022-01-09 14:16	E 91 St & 2 Ave	7286.01	E 85 St & 3 Ave	7212.05	40.78115276	-73.94963041	40.77801203	-73.95407149	member
17	B0536F12DE7E5311	classic_bike	2022-01-19 8:44	2022-01-19 8:52	5 Ave & E 135 St	7769.06	Park Ave & E 124 St	7682.01	40.812191	-73.937838	40.8045555	-73.9396861	member
18	6D9D2C2222D7E33A	classic_bike	2022-01-04 13:59	2022-01-04 14:05	5 Ave & E 135 St	7769.06	Park Ave & E 124 St	7682.01	40.812191	-73.937838	40.8045555	-73.9396861	member
19	CD49F9CD64CD6D89	classic_bike	2022-01-23 12:24	2022-01-23 12:28	E 91 St & 2 Ave	7286.01	E 85 St & 3 Ave	7212.05	40.78115276	-73.94963041	40.77801203	-73.95407149	member
20	280A334FE7464EB4	classic_bike	2022-01-20 18:29	2022-01-20 19:20	E 91 St & 2 Ave	7286.01	E 102 St & Park Ave	7488.24	40.78115276	-73.94963041	40.7904828	-73.95033068	member
21	3C96570C77108EA3	classic_bike	2022-01-28 6:28	2022-01-28 6:33	W 41 St & 8 Ave	6602.03	E 48 St & 5 Ave	6626.01	40.75640548	-73.9900262	40.75724568	-73.97805914	member
22	D95AD5F2DAD2C0A6	classic_bike	2022-01-25 8:21	2022-01-25 8:40	Central Park W & W 91 St	7453.01	E 48 St & 5 Ave	6626.01	40.78866499	-73.96680057	40.75724568	-73.97805914	member
23	9C086BDA0F37CB0E	classic_bike	2022-01-02 8:49	2022-01-02 8:52	E 82 St & East End Ave	7049.04	E 85 St & 3 Ave	7212.05	40.7724607	-73.9468208	40.77801203	-73.95407149	member
24	8DCFD9368A79B545	classic_bike	2022-01-03 8:14	2022-01-03 8:18	E 82 St & East End Ave	7049.04	E 85 St & 3 Ave	7212.05	40.7724607	-73.9468208	40.77801203	-73.95407149	member
25	7B7B5A23EEF32F3E	classic_bike	2022-01-06 8:40	2022-01-06 8:50	Old Broadway & W 133 St	7881.09	Park Ave & E 124 St	7682.01	40.818212	-73.955277	40.8045555	-73.9396861	member
26	2DAEA88A2670038E	classic_bike	2022-01-24 15:41	2022-01-24 15:49	3 Ave & E 100 St	7414.17	E 85 St & 3 Ave	7212.05	40.7877214	-73.94728331	40.77801203	-73.95407149	member
27	12BAFFB86A19D6DA	classic_bike	2022-01-23 17:13	2022-01-23 17:20	3 Ave & E 100 St	7414.17	E 85 St & 3 Ave	7212.05	40.7877214	-73.94728331	40.77801203	-73.95407149	member

Dataset Related Keywords

- Crow Flow
- Spatiotemporal
- Flow prediction
- Traffic prediction
- Incomplete Data

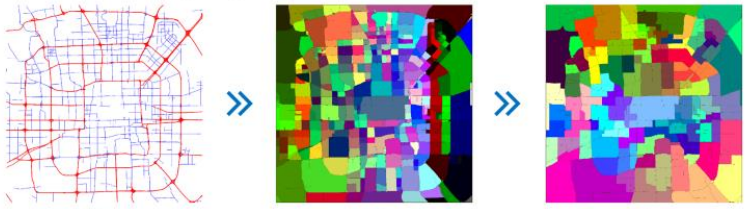
Crowd Flow Prediction Model (1/7)

- Predicting Citywide Crowd Flows in Irregular Regions Using Multi-view Graph Convolutional Networks [1]
  - Highly cited paper that has used the citybike dataset
- The paper aims to predict the crowd flow through Multi-View Graph Convolution Network (MVGCN)

(a) Inflow and outflow (b) Measurement of flows

Fig. 1. Crowd flows in an irregular region

<Irregular Regions>



(a) (b) (c)

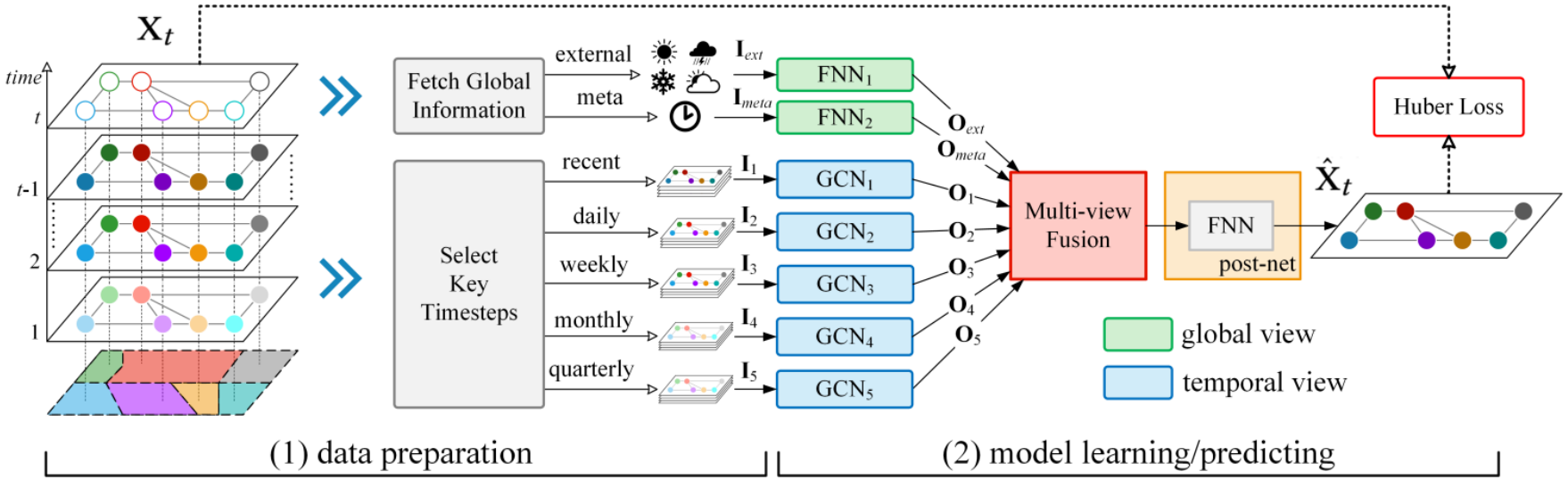
Fig. 4. (a) Road network in Beijing; (b) Regions after map segmentation; (c) Regions after map clustering.

- The crowd flows are measured by the number of bikes running on the roads

[1] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang and Y. Zheng, "Predicting Citywide Crowd Flows in Irregular Regions Using Multi-View Graph Convolutional Networks," in *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2020.3008774.

Crowd Flow Prediction Model (2/7)

■ The architecture of MVGCN and major components are shown below:



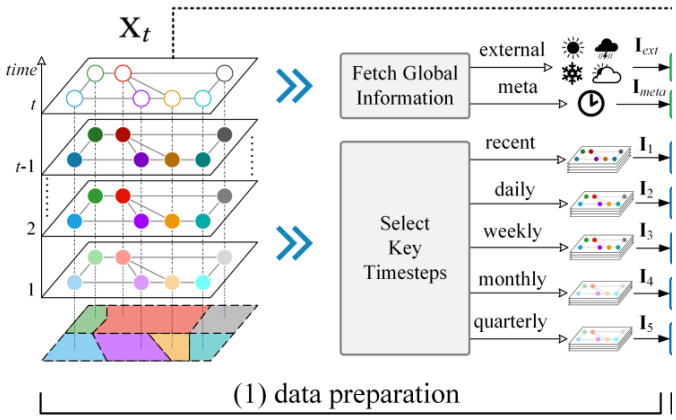
★ Fully-connected Neural Network (FNN)

★ Graph Convolutional Network(GCN)

► Composed of two stage: data preparation and model training(and inference)

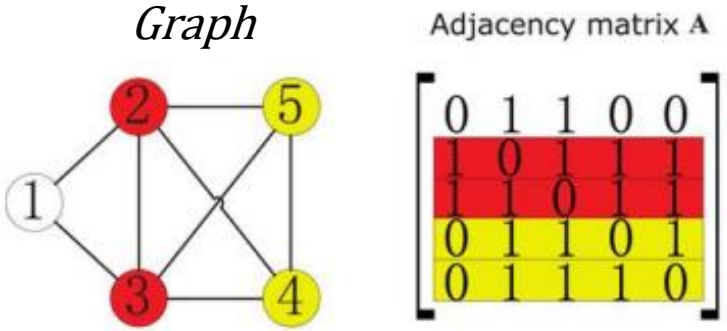
Crowd Flow Prediction Model (3/7)

- Data preparation
  - Weather, time of the day, and period are used as major factor in the process
- Those factors are processed in global and temporal views
- Global view includes meteological data of previous timestep, time of the day and data of the week
- Temporal view is composed of temporal closeness, period, and trend
  - Recent, daily, weekly, monthly, and quarterly timesteps
  - Multi-view of 5 key timesteps



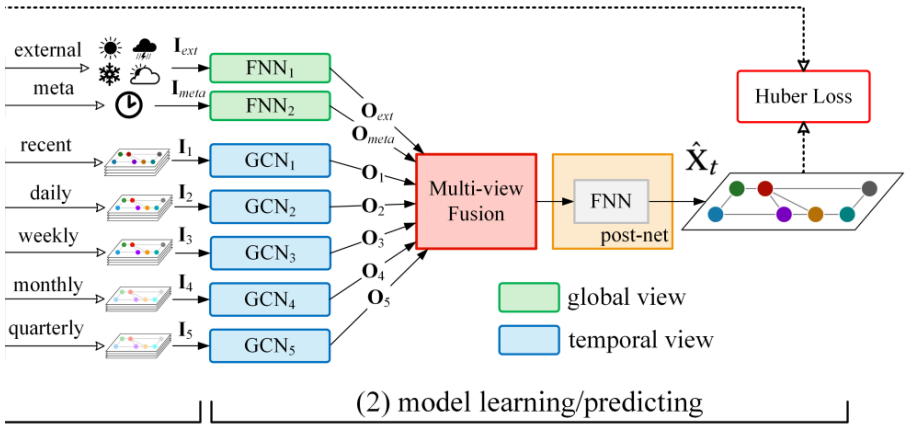
Crowd Flow Prediction Model (4/7)

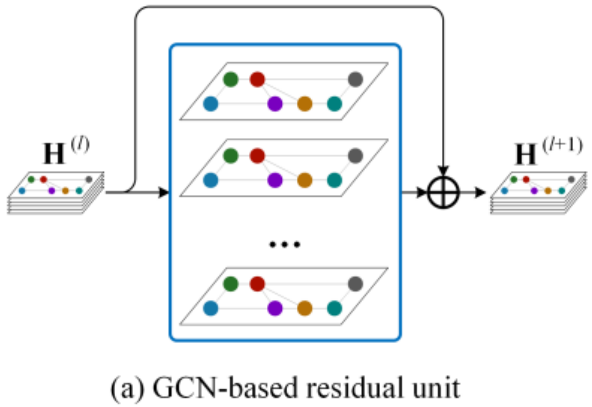
- Graph Convolutional Network in MVGCN
  - Convolutional computation with graph-represented data
  - Extract the spatial features from structured data
- Graph representation of the crowd flow
  - Vertex of the graph refers geospatial positions (*i.e.*, nodes)
  - Edge between two vertex has the value of region-wise transition flows



Crowd Flow Prediction Model (5/7)

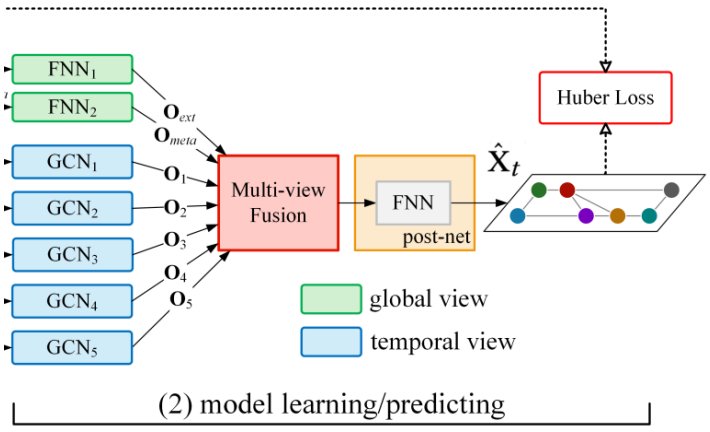
- Graph Convolutional Network models 5 different views of the data
  - Time-varying spatial correlations
  - Output of 5 GCNs are denoted O_1, O_2, O_3, O_4, O_5
- Each GCN uses residual connection to establish deeper connection
  - To capture multi-hop(*i.e.*, node distances) spatial correlations

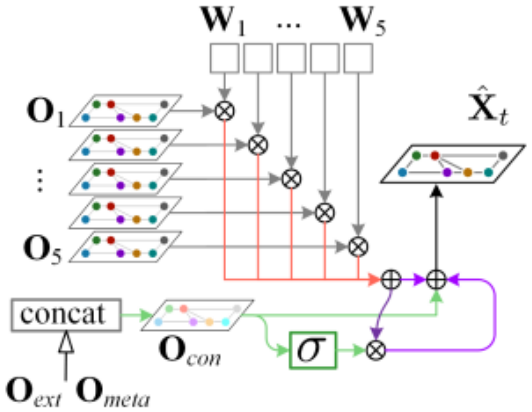




Crowd Flow Prediction Model (6/7)

- Multi-view fusion module
  - $W_1, W_2, W_3, W_4, W5$ are the learnable parameters that control the influences from 5 different views





(b) Multi-view fusion module

Crowd Flow Prediction Model (7/7)

■ Results: RMSE and MAE of the predicted crowd flow graph

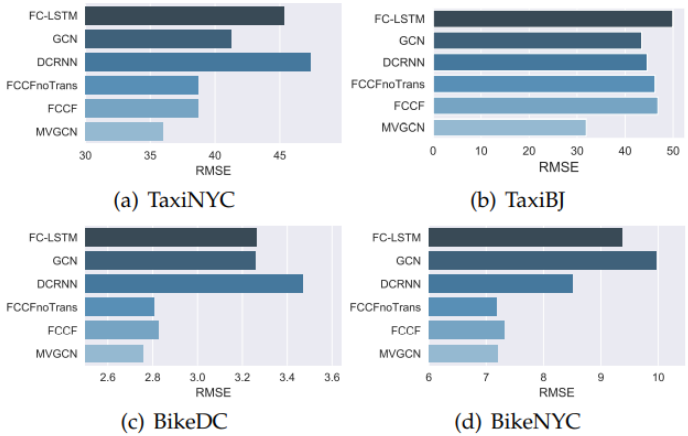


Fig. 7. RMSE comparisons on *sudden changes* in the four datasets.

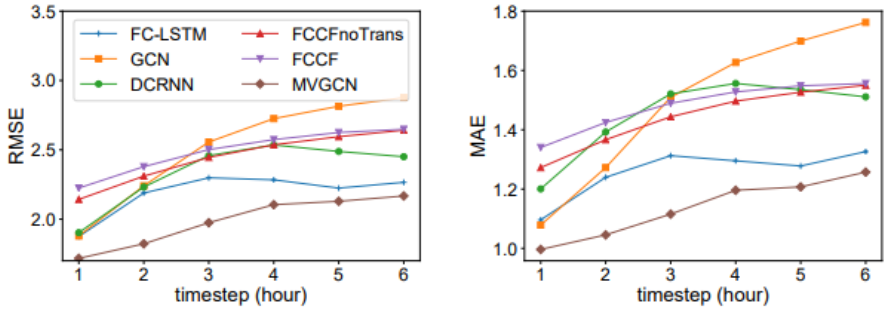


Fig. 9. Step-wise comparisons on the BikeDC test set.

Dataset	Metric	HA	VAR	GBRT	FC-LSTM	GCN	DCRNN	FCCFnoTrans	FCCF	ST-MGCN	**MVGCN**
TaxiNYC	RMSE	101.54	30.78	83.71	27.82	26.52	25.50	26.02	26.00	23.53	**23.15**
	MAE	33.02	11.21	23.46	11.25	11.12	11.20	9.25	**9.24**	9.52	9.40
TaxiBJ	RMSE	38.77	18.79	33.89	19.04	17.38	16.44	18.70	18.42	16.30	**14.37**
	MAE	22.89	11.38	20.34	11.86	10.60	9.68	10.74	10.44	10.18	**9.11**
BikeDC	RMSE	2.61	1.95	3.46	1.88	1.88	1.90	2.22	2.14	-	**1.72**
	MAE	1.48	1.20	1.98	1.10	1.08	1.20	1.34	1.27	-	**1.00**
BikeNYC	RMSE	6.77	4.21	8.57	4.66	5.06	4.35	4.41	4.19	-	**4.15**
	MAE	4.00	2.71	5.17	2.78	2.85	2.90	2.79	2.65	-	**2.60**