Faculdade de Computação

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal de Uberlândia

A Mobile Traffic Predictor Enhanced by Neighboring Transportation Data (MTP-NT)

Patrick Luiz de Araújo

Defesa de dissertação de mestrado em 21/11/2023

Contents

Introduction

5 billion

5G subscribers in 2028

19 GB/month

Data per month, per smartphone in 2028

100 exabytes

Data per quarter in 2028

Fonte: Ericsson Mobility Report, November 2022

Introduction

To

Allocate the maximum amount of users

Optimize network operability

Reach 5G QoS/QoE metrics[1]

- **1ms latency**
- **low energy consumption**
- **High coverage**

New 5G networks will count on

Core Network (CN) based on Virtual Network Functions (VNF) over a Network Function Virtualization (NFV)[2] topology

Cloud and edge computing[3]

Use Machine Learning (ML) and other predictive tools

- **Intelligent caching in network edge**
- **Cloud computing optimization**

[1] AGIWAL, M.; ROY, A.; SAXENA, N. Next generation 5g wireless networks: A comprehensive survey. IEEE Communications Surveys Tutorials, v. 18, n. 3, p. 1617–1655, 2016.
[2] Sun, Y. et al. Application of machine learning in wireless networks: Key techniques and open issues. IEEE Communications Surveys Tutorials, v. 21, n. 4, p. 3072–3108, 2019.
[3] ALAWE, I. et al. Improving traffic forecasting for 5g core network scalability: A machine learning approach. IEEE Network, v. 32, n. 6, p. 42–49, 2018.

Introduction

AI models advantages

- **Rely on historical data[1,2,3]**
- **Can be less complex than conventional approaches[4]**
- **Robust patterns and best overall performance[5]**

Limitants

1. **Enough data?**
2. **Pertinent information?**
3. **Response time**
4. **Return Over Investiment – ROI**

[1] Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: 2017 IEEE 25th International Conference on Network Protocols (ICNP). [S.l.: s.n.], 2017. p. 1–10
[2] Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. [S.l.: s.n.], 2017. p. 1–9.
[3] CHEN, X. et al. Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In: 2015 IEEE International Conference on Communications (ICC). [S.l.: s.n.], 2015. p. 3585–3591.
[4] SUN, H. et al. Learning to optimize: Training deep neural networks for wireless resource management. In: 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). [S.l.: s.n.], 2017. p. 1–6.
[5] Sun, Y. et al. Application of machine learning in wireless networks: Key techniques and open issues. IEEE Communications Surveys Tutorials, v. 21, n. 4, p. 3072–3108, 2019.

tl;dr

Enough Data	Innovation	Compatible responsiveness	Performance

Mobile Traffic Predictor Enhanced by Neighboring Transportation Data MTP-NT

City of Milan Dataset[1]

- Network usage
- Geolocalized tweets
- Weather
- Electricity
- News

[1] Barlacchi, G. et al. A multi-source dataset of urban life in the city of milan and the province of trentino. Sci Data 2, 2015.

Mobile Traffic Predictor Enhanced by Neighboring Transportation Data MTP-NT

City of Milan Dataset[1]

- Network usage
- Geolocalized tweets
- Weather
- Electricity
- News

Scalable public transport and neighboring data

Open source

[1] Barlacchi, G. et al. A multi-source dataset of urban life in the city of milan and the province of trentino. Sci Data 2, 2015.

Mobile Traffic Predictor Enhanced by Neighboring Transportation Data MTP-NT

City of Milan Dataset[1]

- Network usage
- Geolocalized tweets
- Weather
- Electricity
- News

Scalable public transport and neighboring data

Open source

Lightweight, adaptable and highly performant

[1] Barlacchi, G. et al. A multi-source dataset of urban life in the city of milan and the province of trentino. Sci Data 2, 2015.

Mobile Traffic Predictor Enhanced by Neighboring Transportation Data MTP-NT

City of Milan Dataset[1]

- Network usage
- Geolocalized tweets
- Weather
- Electricity
- News

Scalable public transport and neighboring data

Open source

Lightweight, adaptable and highly performant

State-of-art performance

[1] Barlacchi, G. et al. A multi-source dataset of urban life in the city of milan and the province of trentino. Sci Data 2, 2015.

Contents

Related work

Characterization

Metrics and mathematical characteristics of network usage

Prediction

Mathematical models to predict network traffic

Network Traffic Characterization

Regions of the city grouped based on network usage patterns[1]

Composition of trimodal distributions to describe the network traffic[2]

Sand temporal distribution of the network traffic results into extremely insufficient utilization of network resources[2]

Traffic was concentrated in some regions (city center) and peak hours[3]

[1] Xu, F. et al. Understanding mobile traffic patterns of large scale cellular towers in urban environment. IEEE/ACM Transactions on Networking, v. 25, n. 2, p. 1147–1161, 2017.

[2] WANG, H. et al. Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks. In: Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking. New York, NY, USA: Association for Computing Machinery, 2015. (HOTPOST '15), p. 19–24. ISBN 9781450335171. Disponível em: <https://doi.org/10.1145/2757513.2757518>.

[3] Gotzner, U.; Rathgeber, R. Spatial traffic distribution in cellular networks. In: VTC '98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No.98CH36151). [S.l.: s.n.], 1998. v. 3, p. 1994–1998 vol.3.

Characterization – Xu

Grouping of regions based on network usage patterns

- **Residential**
- **Transport**
- **Office**
- **Entertainment**
- **Comprehensive areas**

Clusterization of regions (5 clusters)

Human labelling of some regions to generalization

Characterization – Xu

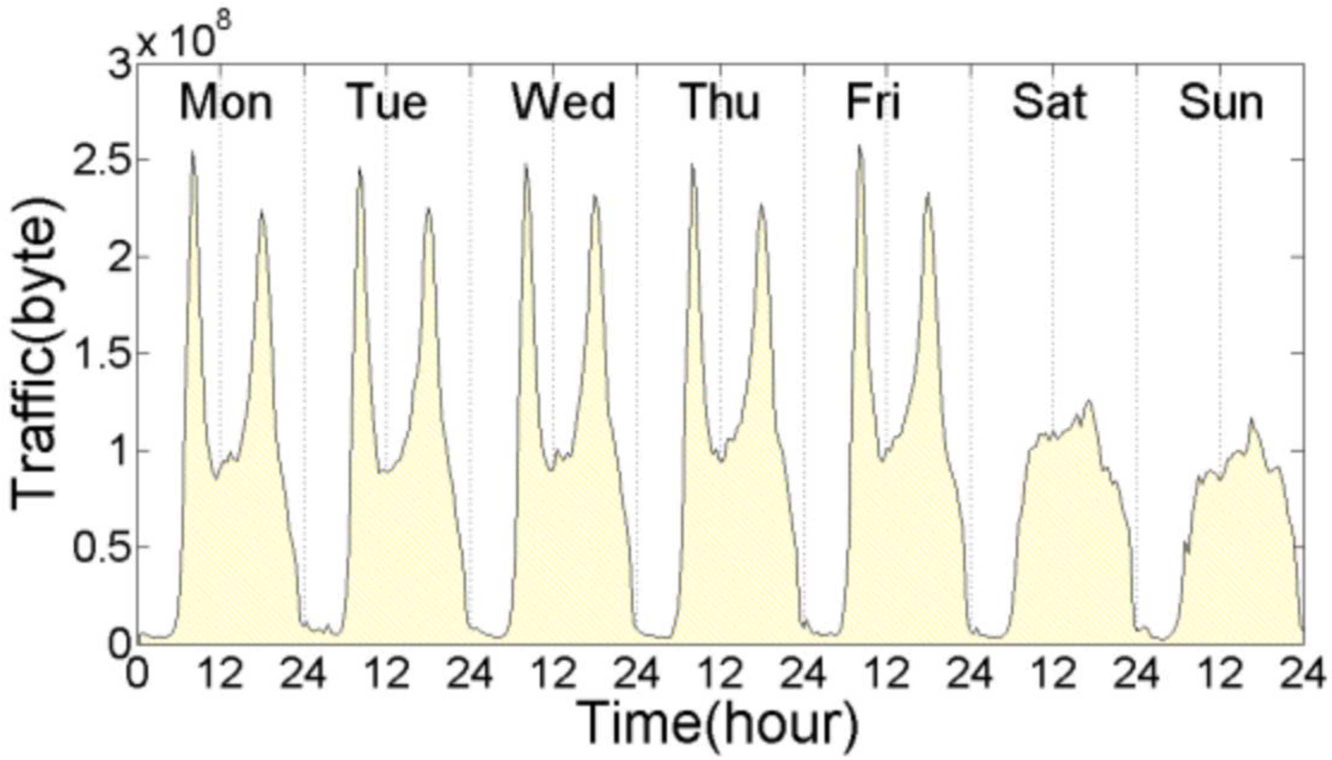
Weekday-Weekend Traffic Amount Ratio: ratio of weekdays and weekend traffic

Peak-Valley Features: ratio of maximum and minimum traffic registered as seen in the figure ahead

Time of Traffic Peak and Valley: time of the day of maximum and minimum network usage

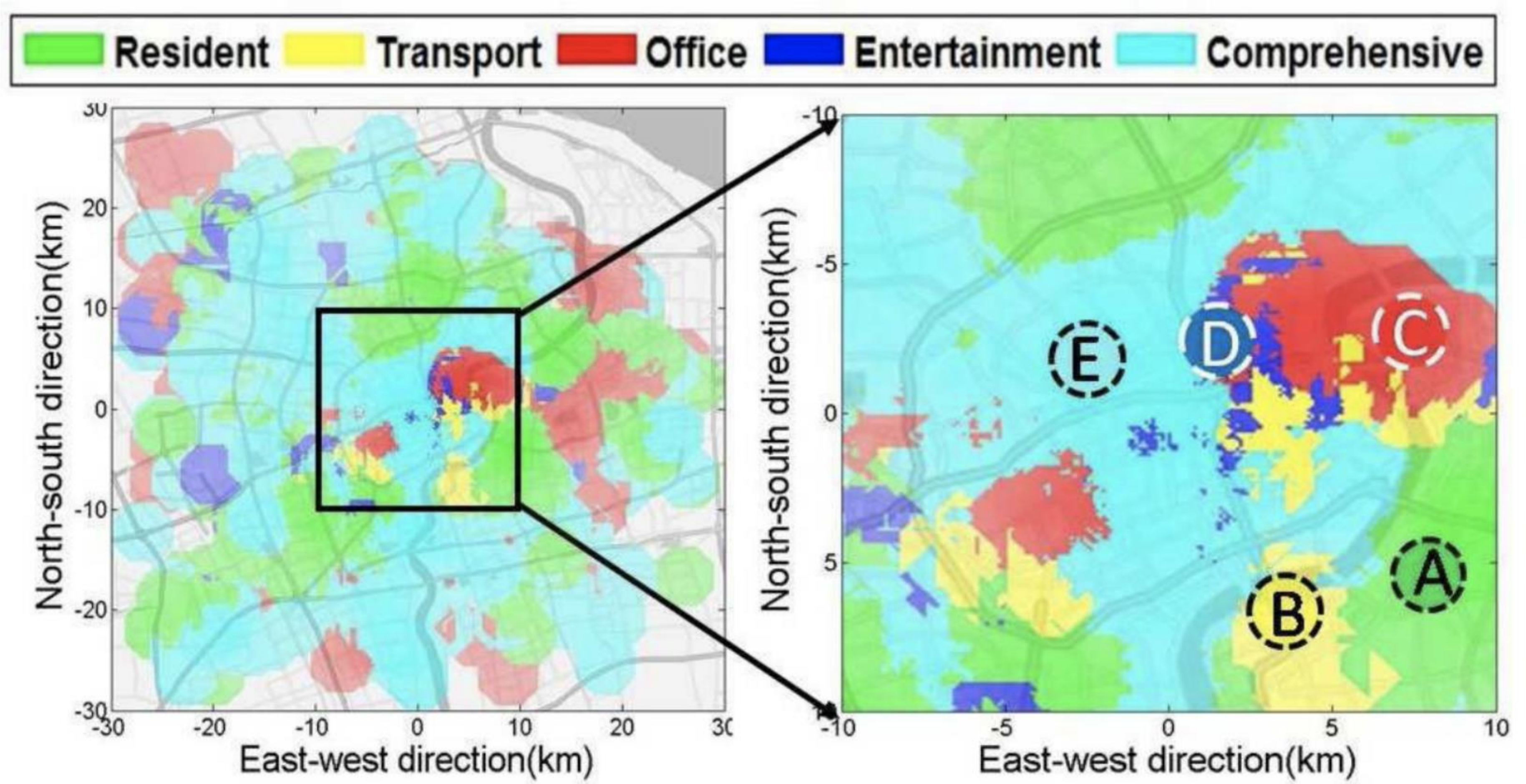
Discrete Fourier Transform (DFT)

Characterization – Xu



Source: Xu, F. et al. **Understanding mobile traffic patterns of large scale cellular towers in urban environment**. IEEE/ACM Transactions on Networking, v. 25, n. 2, p. 1147–1161, 2017.

Characterization – Xu



Source: Xu, F. et al. **Understanding mobile traffic patterns of large scale cellular towers in urban environment**. IEEE/ACM Transactions on Networking, v. 25, n. 2, p. 1147–1161, 2017.

Characterization – Wang 2015

380,000 Base Stations (BSs) in Shanghai

August 1 – August 31, 2014

10 minute samples of each Base Station (BS)

1.96 billion entires; 28PB (92TB per day, 7GB per BS on average)

Source: WANG, H. et al. **Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks**. In: Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking. New York, NY, USA: Association for Computing Machinery, 2015. (HOTPOST ’15), p. 19–24. ISBN 9781450335171. Disponível em: <https://doi.org/10.1145/2757513.2757518>.

Characterization – Wang 2015

Findings

Trimodal distribution

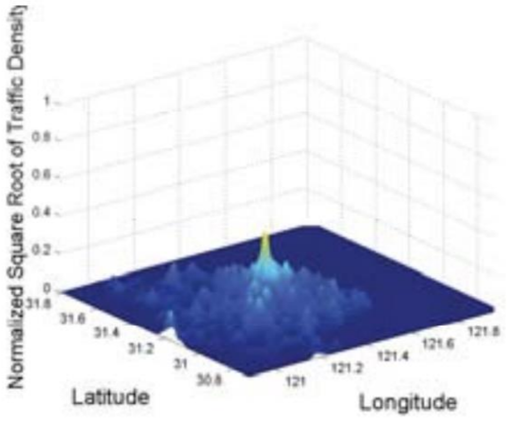
- **Compound-expontential**
- **Power-law distribution**
- **Exponential distribution**

R-square of 99%

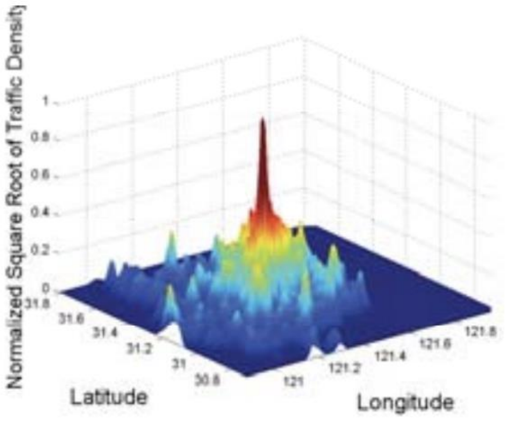
Characterization – Wang 2015

Findings

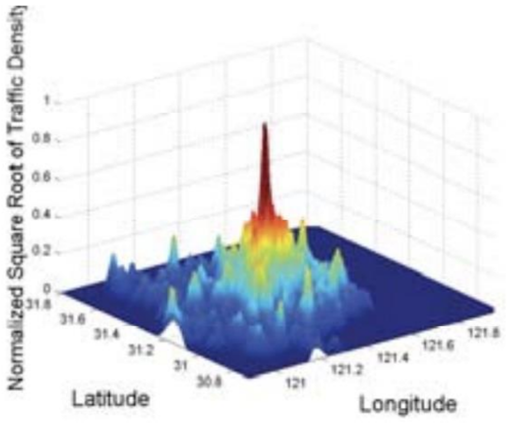
Spatial and temporal distribution of network traffic highly concentrated



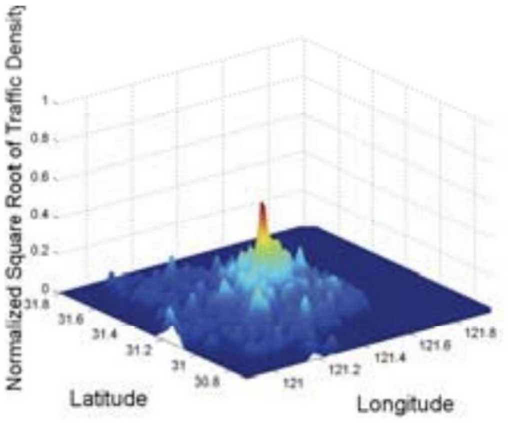
(a) 4AM



(b) 10AM



(c) 4PM



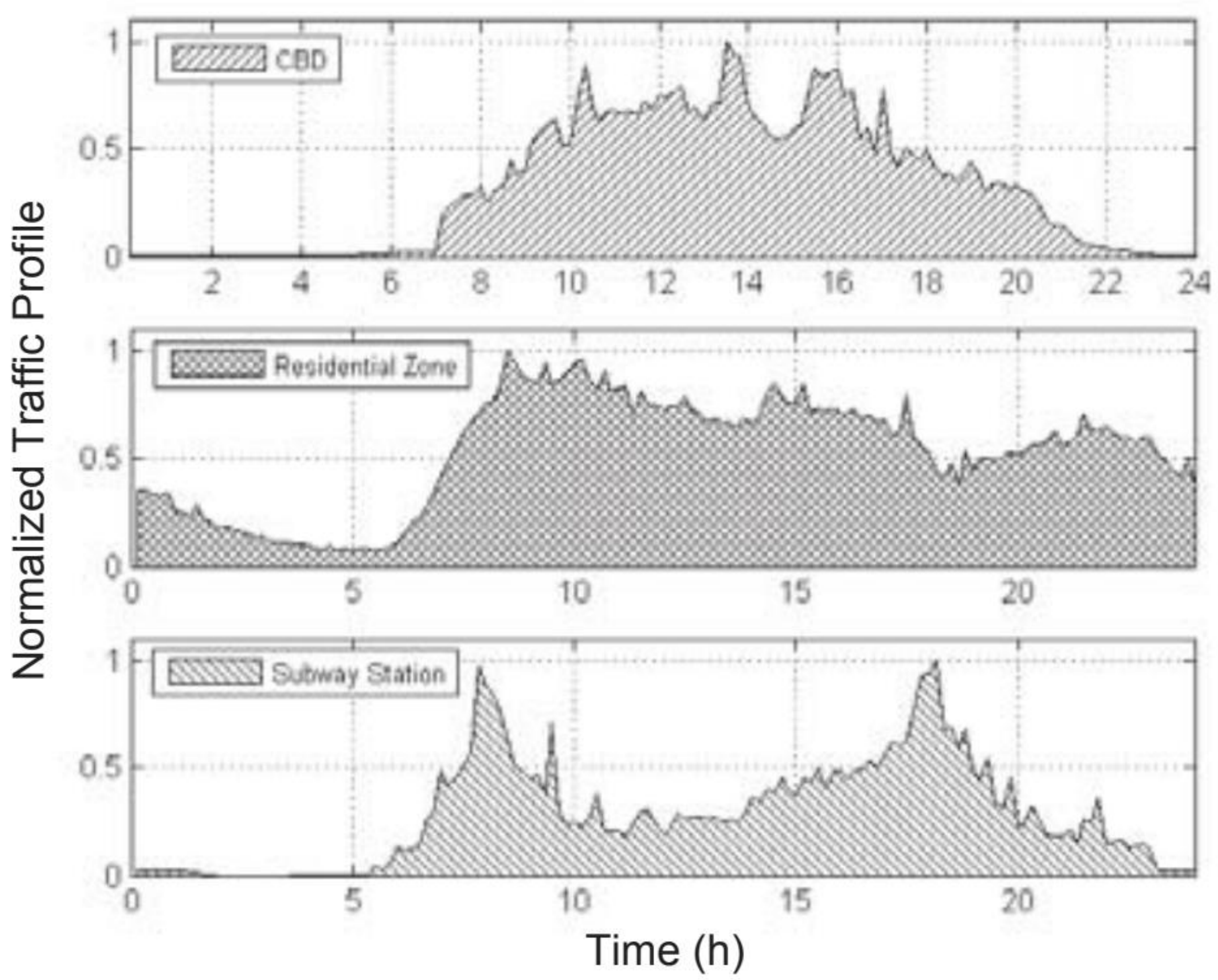
(d) 10PM

Source: WANG, H. et al. **Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks**. In: Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking. New York, NY, USA: Association for Computing Machinery, 2015. (HOTPOST '15), p. 19–24. ISBN 9781450335171. Disponível em: <https://doi.org/10.1145/2757513.2757518>.

Characterization – Wang 2015

Findings

Irregular usage depending on the urban ecology



Source: WANG, H. et al. **Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks**. In: Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking. New York, NY, USA: Association for Computing Machinery, 2015. (HOTPOST ’15), p. 19–24. ISBN 9781450335171. Disponível em: <https://doi.org/10.1145/2757513.2757518>.

Characterization – Gotzner

Network traffic in Berlin

May 1996 – August 1997

Findings

- **Traffic concentrated in city centre**
- **Congestions in peak hours**
- **Network resources not fully used in other points**
- **The infrastructure reaches the maximum capacity before all processing power request**

Source: Gotzner, U.; Rathgeber, R. **Spatial traffic distribution in cellular networks**. In: VTC'98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No.98CH36151). [S.l.: s.n.], 1998. v. 3, p. 1994–1998 vol.3.

Network Traffic Prediction

Pure Time Series Function (TSF) and Non-TSF problem[1]

1989

Multi Layer Perceptrion (MLP) to traffic perdiction[2]

Single Layer Perceptron (SLP) to traffic perdiction[3]

1991

MLP to network traffic forecasting[4]

2017

Recurrent Neural Network and usage of global + local autoencoders[5]

In-tower and inter-tower traffic analysis through a Graph Neural Network (GNN)[6]

2021

AutoRegressive Moving Average (ARIMA) and Neural Network model training with Simulated Annealing (SA)[7]

[1] BOUTABA, R. et al. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. Journal of Internet Services and Applications, Springer, v. 9, n. 1, p. 1–99, 2018.

[2] CYBENKO, G. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, Springer, v. 2, n. 4, p. 303–314, 1989.[3] Funahashi, 1989

[4] HORNIK, K. Approximation capabilities of multilayer feedforward networks. Neural networks, Elsevier, v. 4, n. 2, p. 251–257, 1991.[5] Wang, 2017a

[5] FUNAHASHI, K.-I. On the approximate realization of continuous mappings by neural networks. Neural networks, Elsevier, v. 2, n. 3, p. 183–192, 1989.

[6] Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: 2017 IEEE 25th International Conference on Network Protocols (ICNP). [S.l.: s.n.], 2017. p. 1–10.

[7] YANG, H. et al. A network traffic forecasting method based on sa optimized arima–bp neural network. Computer Networks, v. 193, p. 108102, 2021. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/ S1389128621001821>.

Prediction – Wang 2017a

2844 Base Stations (BSs) in Suzhou

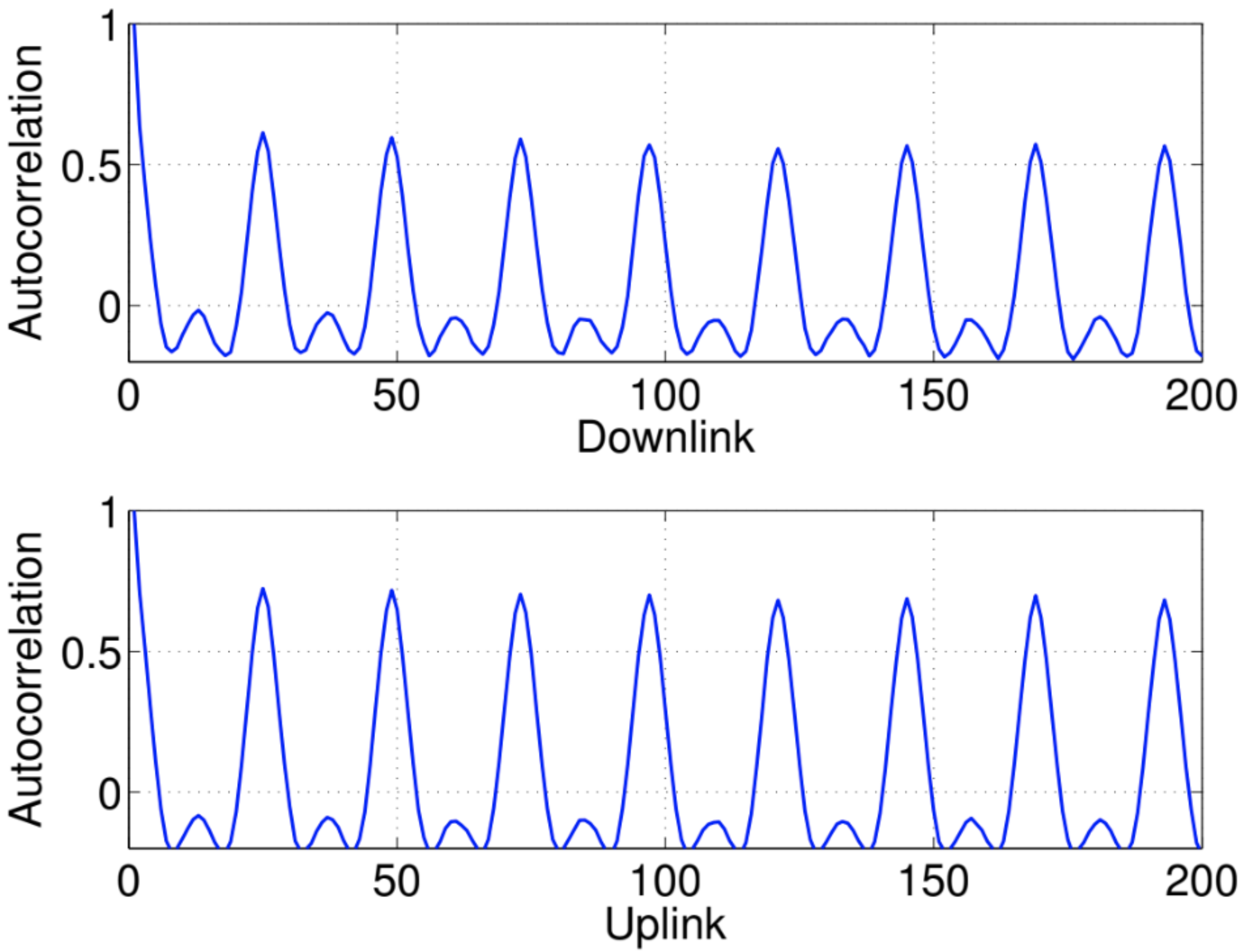
$500m^2$x $500m^2$

Uses the neighborhood concept

LSTM Cells paired with Global and Local Autoencoders

Source: Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p .1–9.

Prediction – Wang 2017a



Source: Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p. 1–9.

Prediction – Wang 2017a

Table 1 – Spatial correlation of a arrange of 7 BSs

	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6	Cell 7
Cell 1	1.000	0.167	0.435	0.130	0.040	0.341	0.307
Cell 2	0.396	1.000	0.338	0.129	0.084	0.310	0.222
Cell 3	0.345	0.541	1.000	0.159	0.162	0.697	0.536
Cell 4	0.437	0.439	0.458	1.000	0.104	0.131	0.114
Cell 5	0.360	0.471	0.492	0.508	1.000	0.163	0.080
Cell 6	0.286	0.491	0.550	0.432	0.535	1.000	0.603
Cell 7	0.284	0.506	0.526	0.459	0.535	0.577	1.000

Source: Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p. 1–9.

Prediction – Wang 2017a

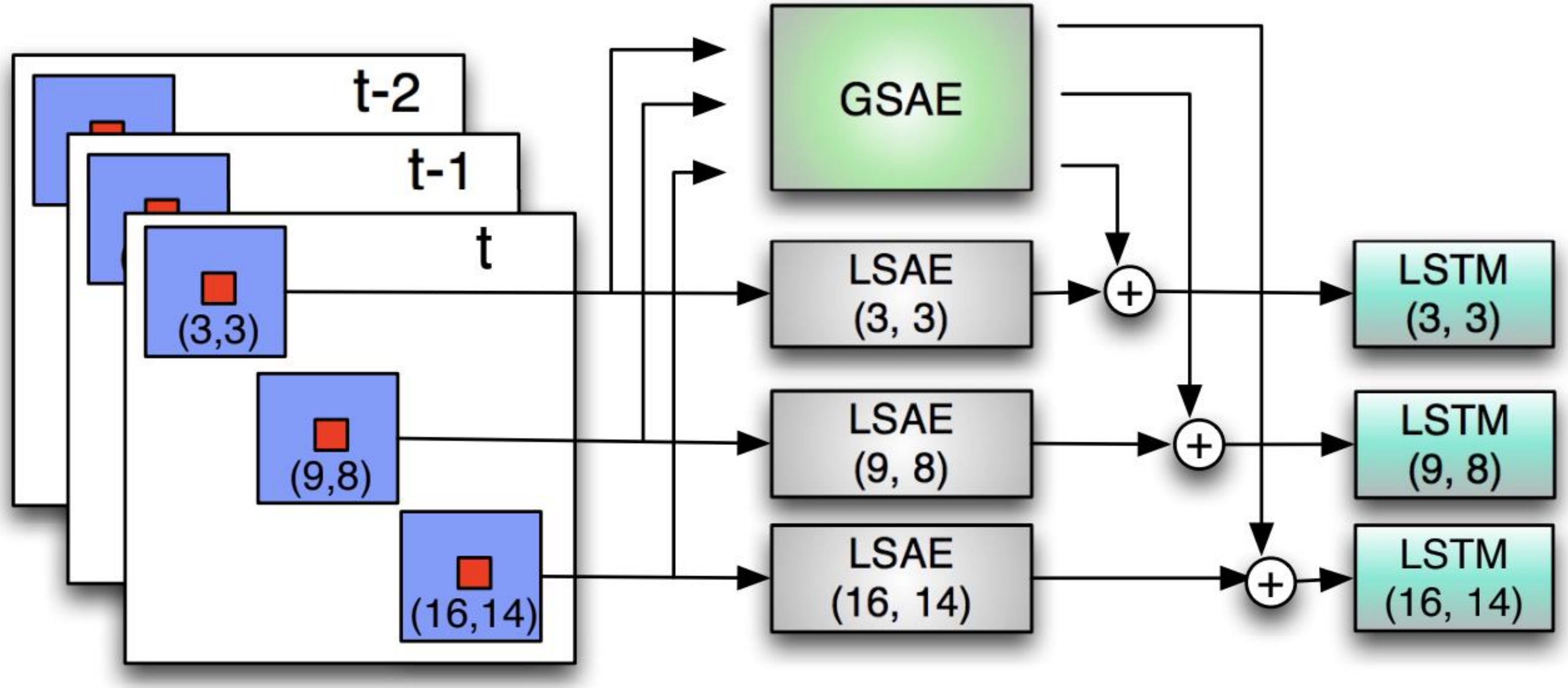
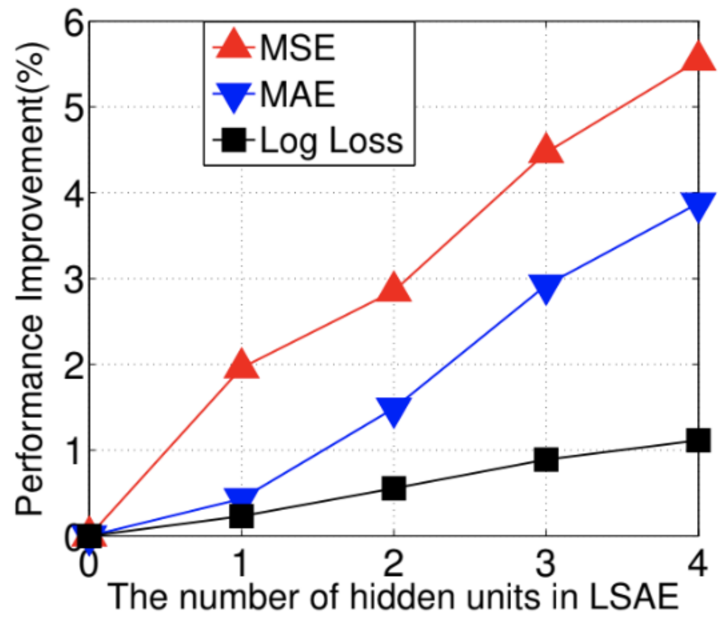
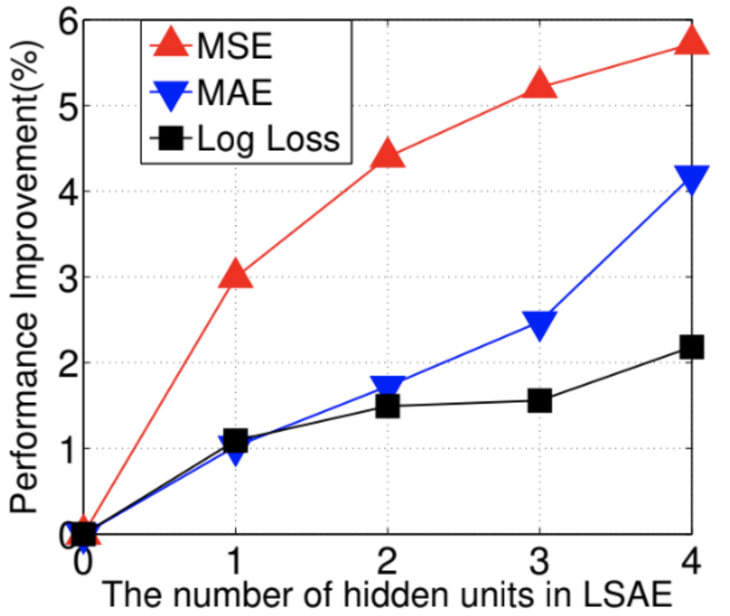


Fig. 3. The proposed deep learning model

Source: Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p. 1–9.

Prediction – Wang 2017a





Source: Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p. 1–9.

Related work – Wang 2017b

5929 Base Stations (BSs), 1.5 million users

In-tower and inter-tower traffic

Source: Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: **2017 IEEE 25th International Conference on Network Protocols (ICNP)**. [S.l.: s.n.], 2017. p. 1–10.

Related work – Wang 2017b

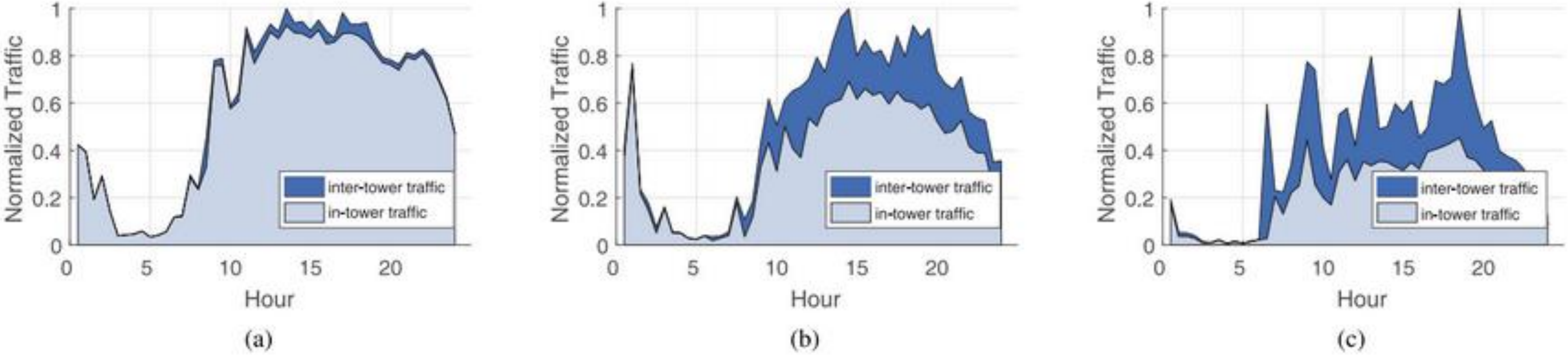


Fig. 5. An illustration of three typical in-tower and inter-tower cell tower data traffic characteristics. (a) In-tower traffic dominant, collected from a cell tower in a residential area; (b) inter-tower traffic consistently notable during the daytime, collected from a cell tower in a shopping mall; and (c) inter-tower traffic surges at certain times, collected from a cell tower in a transit station.

Source: Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: **2017 IEEE 25th International Conference on Network Protocols (ICNP)**. [S.l.: s.n.], 2017. p. 1–10.

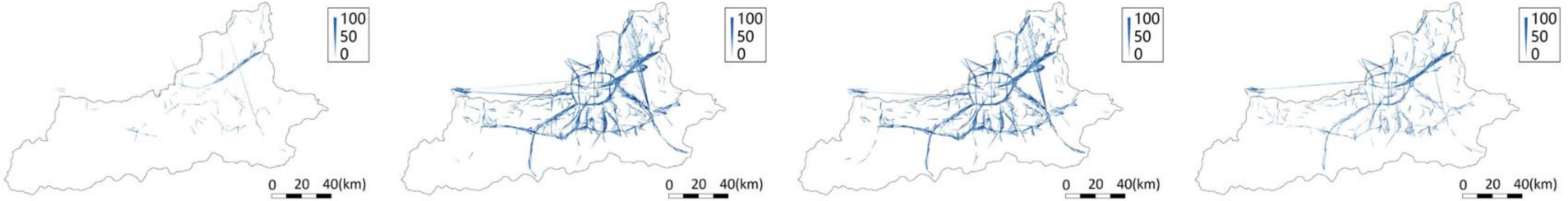
Related work – Wang 2017b

Temporal correlations between physically distant towers

Graph Neural Network (GNN) architecture

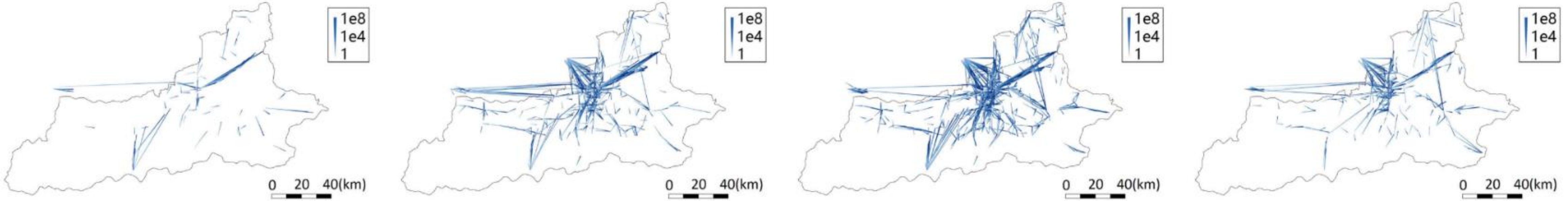
Source: Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: **2017 IEEE 25th International Conference on Network Protocols (ICNP)**. [S.l.: s.n.], 2017. p. 1–10.

Related work – Wang 2017b



(a) 04:00 (b) 10:00 (c) 16:00 (d) 00:00

Fig. 7. Distribution of user mobility at different times of a day. An edge with gradient color from white to dark blue represents the direction of user mobility between a pair of cell towers. The line width of the edge shows the intensity of user mobility.



(a) 04:00 (b) 10:00 (c) 16:00 (d) 00:00

Fig. 8. Distribution of inter-tower traffic volume at different times of the same day. The line width of each edge represents the volume of inter-tower traffic between a pair of cell towers, aggregated by half an hour.

Source: Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: **2017 IEEE 25th International Conference on Network Protocols (ICNP)**. [S.l.: s.n.], 2017. p. 1–10.

Related work – Summary

Ref.	Method	Dataset availability	Source code available
(Wang et al., 2017a)	Autoencoders	X	X
(Wang et al., 2017b)	Graph Neural Networks	X	X
(Sciancalepore et al., 2017)	Holt Winters	No information	X
(ALAWE et al., 2018)	Deep Learning	X	X
(YANG et al., 2021)	ARIMA and Neural Network	✓	X

Related work – Summary

Ref.	LSTM	Time series	Grid arrange	Neighborhood concept	Spatial modelling	Residual/aperiodic events	Network traffic consumption characterization
Wang et al., 2015							✓
Wang et al., 2017a	✓		✓	✓	✓		
Wang et al., 2017b		✓		✓	✓	✓	
YANG et al., 2021		✓					
Gotzner; Rathgeber, 1998							✓
BOUTABA et al., 2018							✓

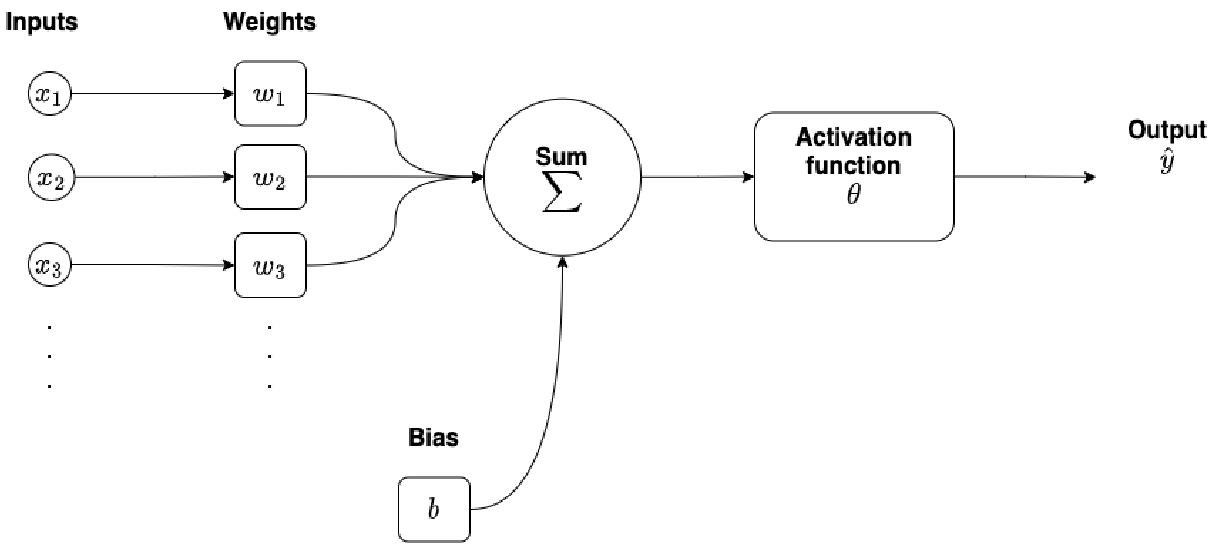
Contents

Theoretical fundamentation – Neural networks development and training

$$Z = X \times W + b$$

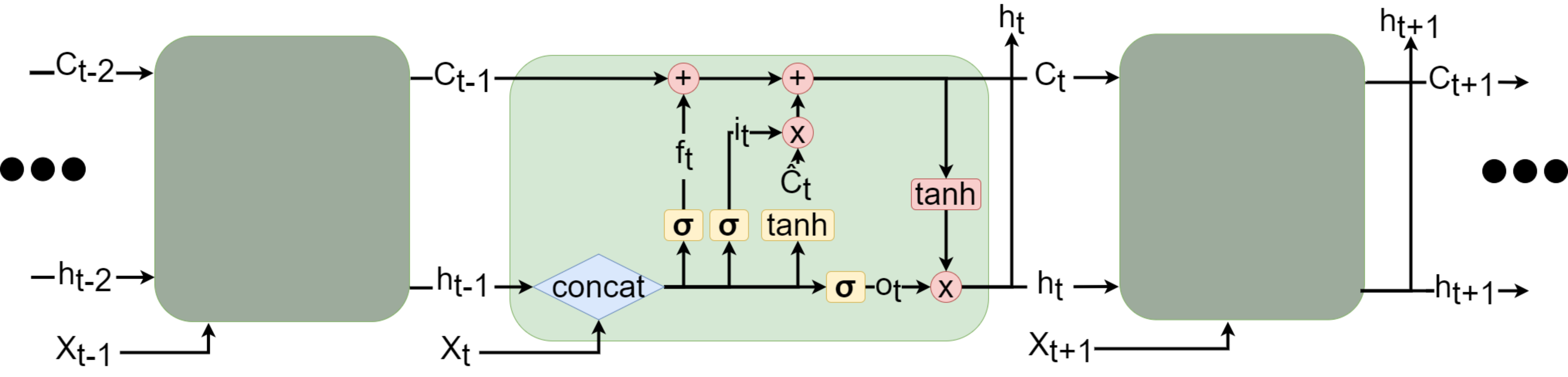
$$\hat{Y} = \theta(Z)$$

Common activation functions:
- Sigmoid: output between 0 and 1;
- Hyperbolic tangent: output -1 and 1;
- Rectified Linear Unit (ReLU)



Source: the author

Theoretical fundamentation – LSTM



Source: the author

Theoretical fundamentation – Feature selection

Mainly used to dimensionality reduction

Pearson correlation coefficient

- **Between -1 and 1**
- **Strength of the relationship between two variables**

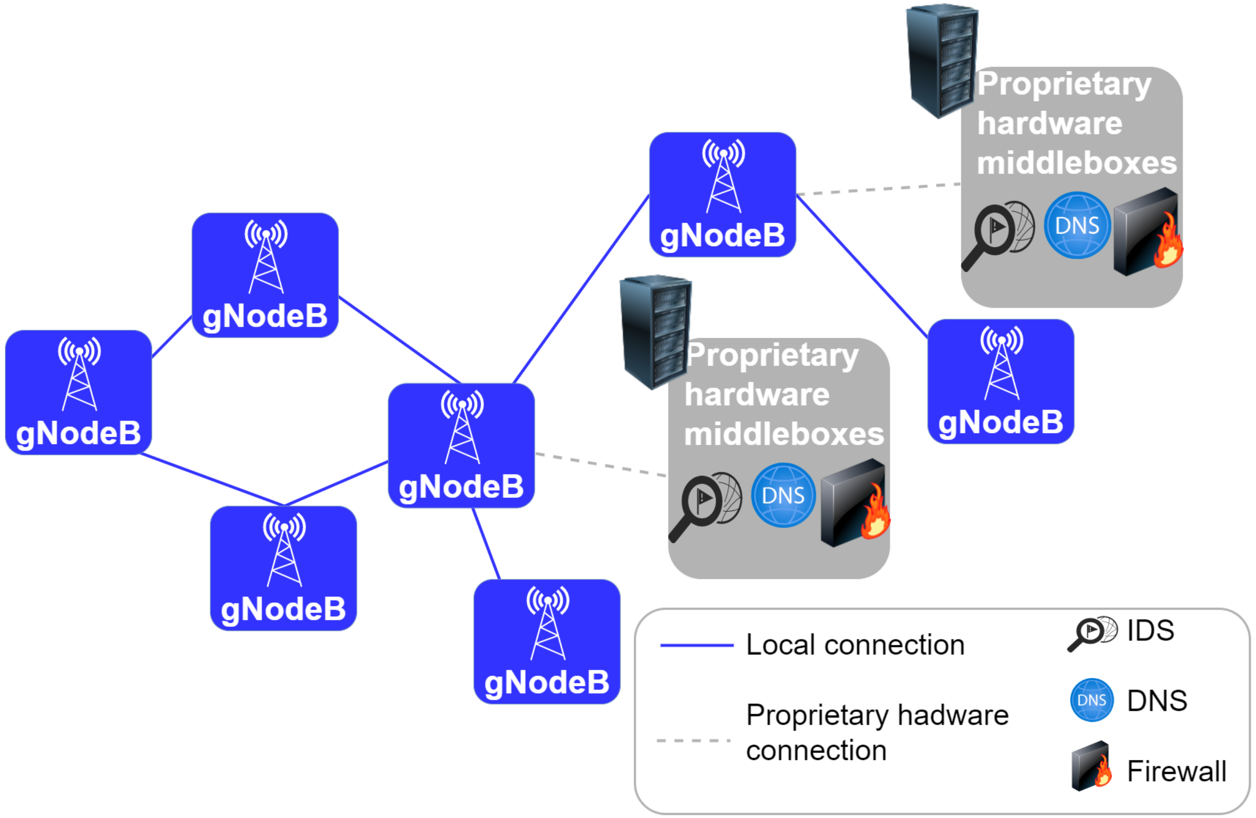
f-value

Distance based algorithm
- **All transport regions inside 20 Moore distance***

***Later explained**

Contents

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure



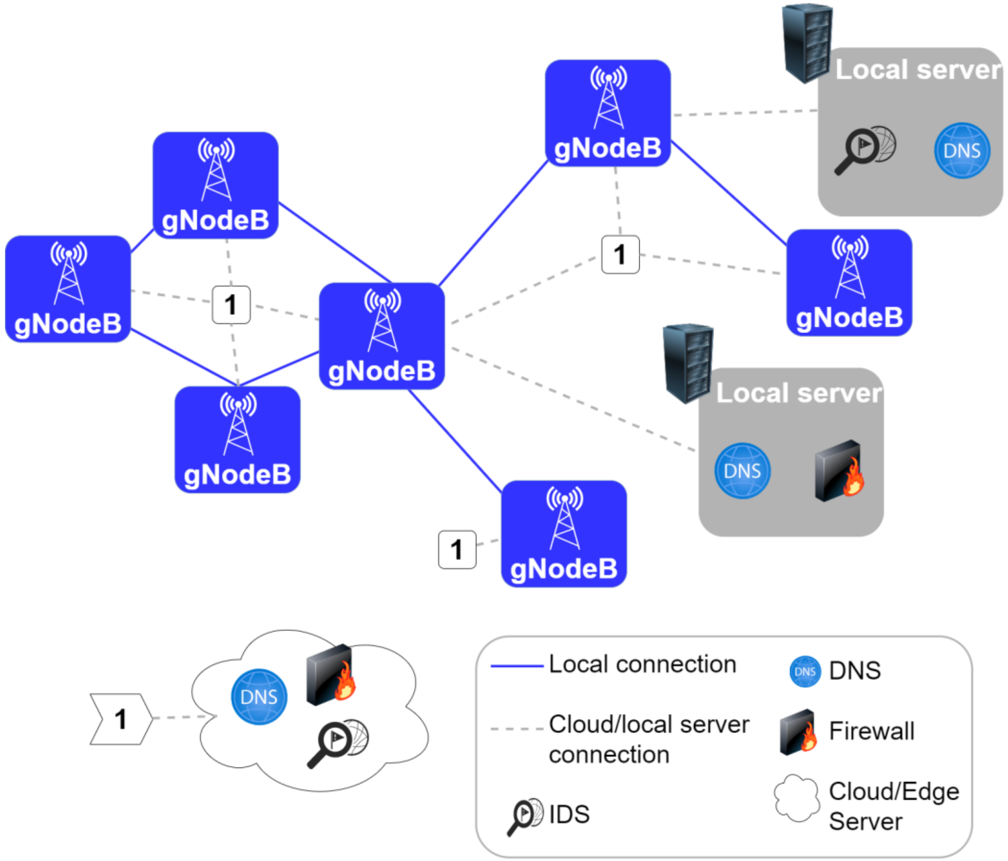
Source: the author

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure

Hardware middleboxes become software functions

Rely on dedicated hardware and/or cloud

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure



Source: the author

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure

Main advantages according to European Telecommunications Standards (ETSI)[1]

1. **NFV as a service[2]**
2. **Virtualization of Core Network (CN) and Base Stations (BSs)[3]**
3. **Virtualization of home environment[4]**
4. **Virtualization of CDNs[5,6]**

[1] ETSI, N. Network Function Virtualisation Use Cases. [S.l.]: European Telecommunications Standards Institute Sophia-Antipolis, France, 2013.

[2] RANKOTHGE, W. et al. Towards making network function virtualization a cloud computing service. In: IEEE. 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). [S.l.], 2015. p. 89–97.

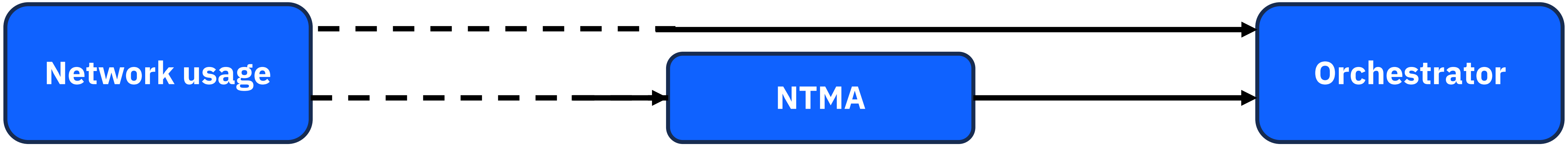
[3] BASTA, A. et al. Applying nfv and sdn to lte mobile core gateways, the functions placement problem. In: Proceedings of the 4th workshop on All things cellular: operations, applications, & challenges. [S.l.: s.n.], 2014. p. 33–38.

[4] BRONSTEIN, Z.; SHRAGA, E. Nfv virtualisation of the home environment. In: IEEE. 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC). [S.l.], 2014. p. 899–904.

[5] MANGILI, M.; MARTIGNON, F.; CAPONE, A. Stochastic planning for content delivery: Unveiling the benefits of network functions virtualization. In: IEEE. 2014 IEEE 22nd International Conference on Network Protocols. [S.l.], 2014. p. 344–349.

[6] KIM, T.; LEE, B. Scalable cdn service poc over distributed cloud management platform. In: IEEE. 2014 International Conference on Information and Communication Technology Convergence (ICTC). [S.l.], 2014. p. 832–833.

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure



Network Traffic Monitoring and Analysis

Also a NFV

Problems

- **Volume**
- **Velocity**

Problems

- **Volume**
- **Velocity**

Preliminaries on data Collection for MTP-NT – The predictive model in the 5G Infrastructure

How does it flow?

How to store?

Is there a open standard?

Preliminaries on data Collection for MTP-NT – Data flow

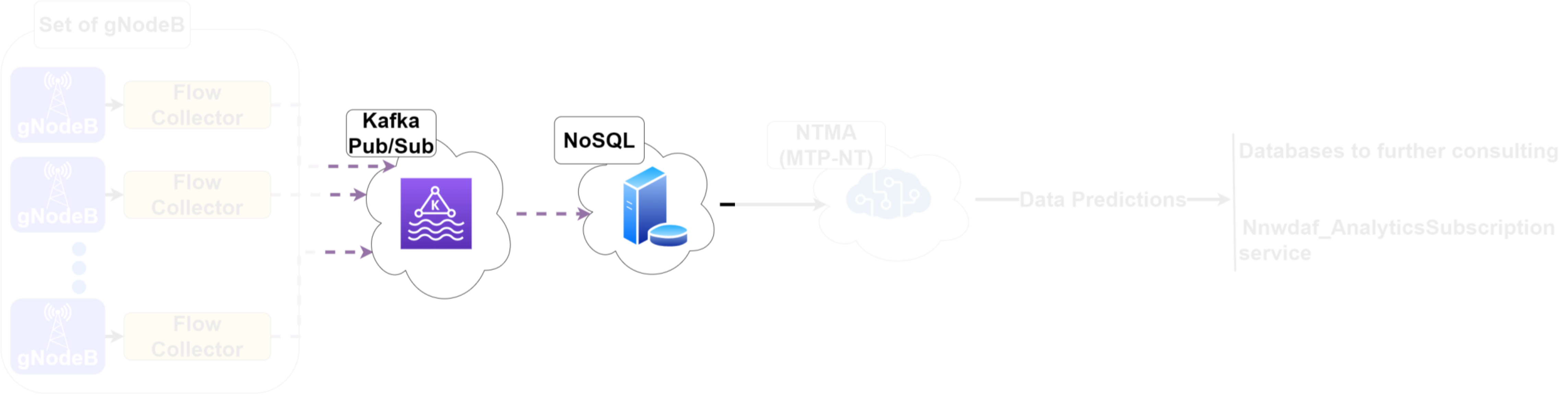
Flow collectors/Network exporters and collectors

Radio Access Network (RAN) layer

Network Data Analytics Function (NWDAF)[1]

[1] 3GPP. Architecture enhancements for 5G System (5GS) to support network data analytics services. [S.l.], 2022. Version 17.5.0. Disponível em: <<https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3579>>.

Preliminaries on data Collection for MTP-NT – Data flow



NoSQL[1,2]

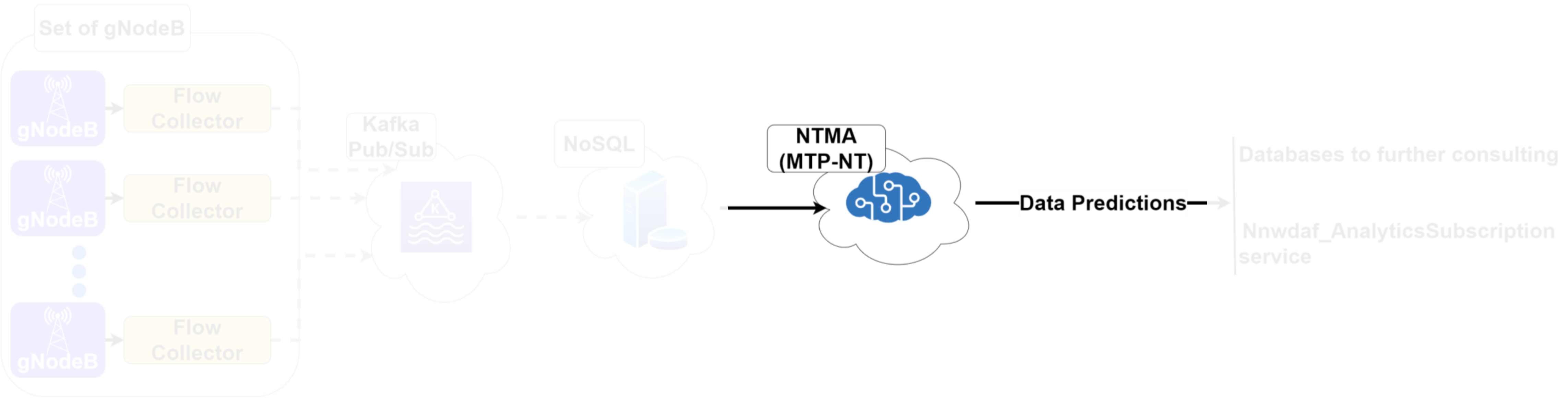
Less performance penalties with large datasets

Popular open source solutions

[1] HAN, J. et al. Survey on nosql database. In: IEEE. 2011 6th international conference on pervasive computing and applications. [S.l.], 2011. p. 363–366.

[2] D'ALCONZO, A. et al. A survey on big data for network traffic monitoring and analysis. IEEE Transactions on Network and Service Management, IEEE, v. 16, n. 3, p. 800–813, 2019.

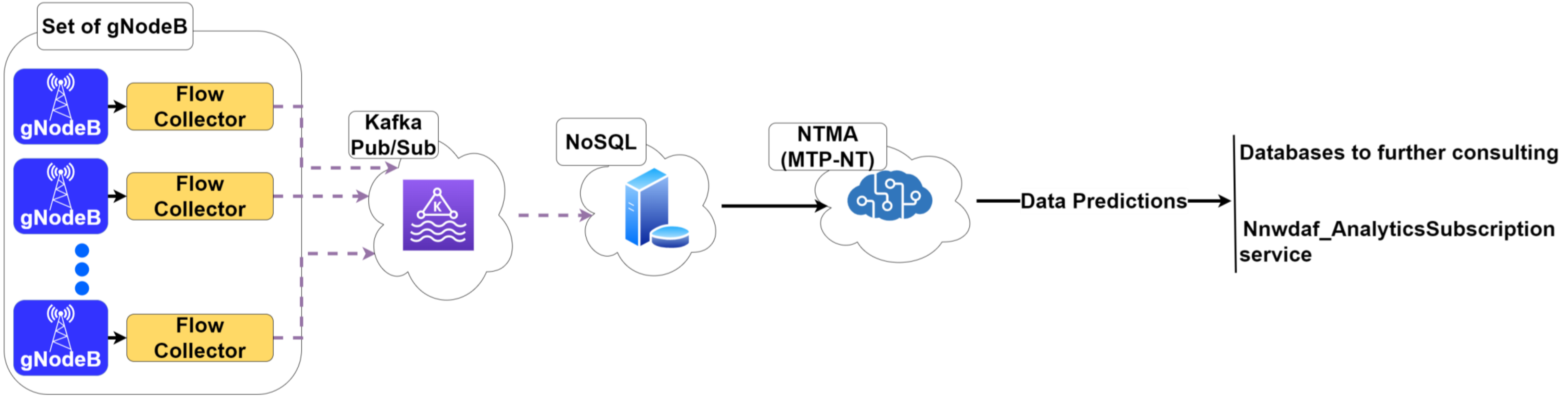
Preliminaries on data Collection for MTP-NT – Data flow



The data is pulled by the MTP-NT, which can be seen as a NTMA

Preliminaries on data Collection for MTP-NT – Data flow

Preliminaries on data Collection for MTP-NT – Data flow



Source: the author

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Database to Milan and Trento from November 1st to December 31st of 2013[1]

1. **Grid (Telecom Italia)**
2. **Social Pulse (Spazio Dati, DEIB)**
3. **Telecommunications (Telecom Italia)**
4. **Precipitations (Metereotrentino, ARPA)**
5. **Weather (ARPA)**
6. **Electricity (SET Distribuizione SPA)**
7. **News (Citynews)**

[1] Barlacchi, G. et al. A multi-source dataset of urban life in the city of milan and the province of trentino. Sci Data 2, 2015.

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Telecommunications dataset from Milan

- **10,000 zonal regions in a 100x100 regular grid**
- **Each regions is a square of 0.06km^2**
- **One log at every 10 minutes in every region**
- **Call Detail Records (CDRs)**

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Anonymization of data

- **GDPR compliant**
- **Protects the real infrastructure capabilities**
- **Each parameter f is multiplied by a anonymization constant k_f**

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Square id	Time Interval	Country code	SMS-in activity	SMS-out activity	Call-in activity	Call-out activity	Internet traffic activity
1	1383606E+6	0	1.7873E-3	NaN	NaN	NaN	NaN
1	1383606E+6	33	NaN	NaN	NaN	NaN	2.6137E-2
1	1383606E+6	39	8.8512E-2	1.4195E-1	1.0804E-1	2.73E-2	9.2032
10	1383606E+6	33	NaN	NaN	NaN	NaN	2.8653E-2
10	1383606E+6	39	6.7480E-2	1.0631E-1	5.9175E-2	1.0174E-2	5.7891

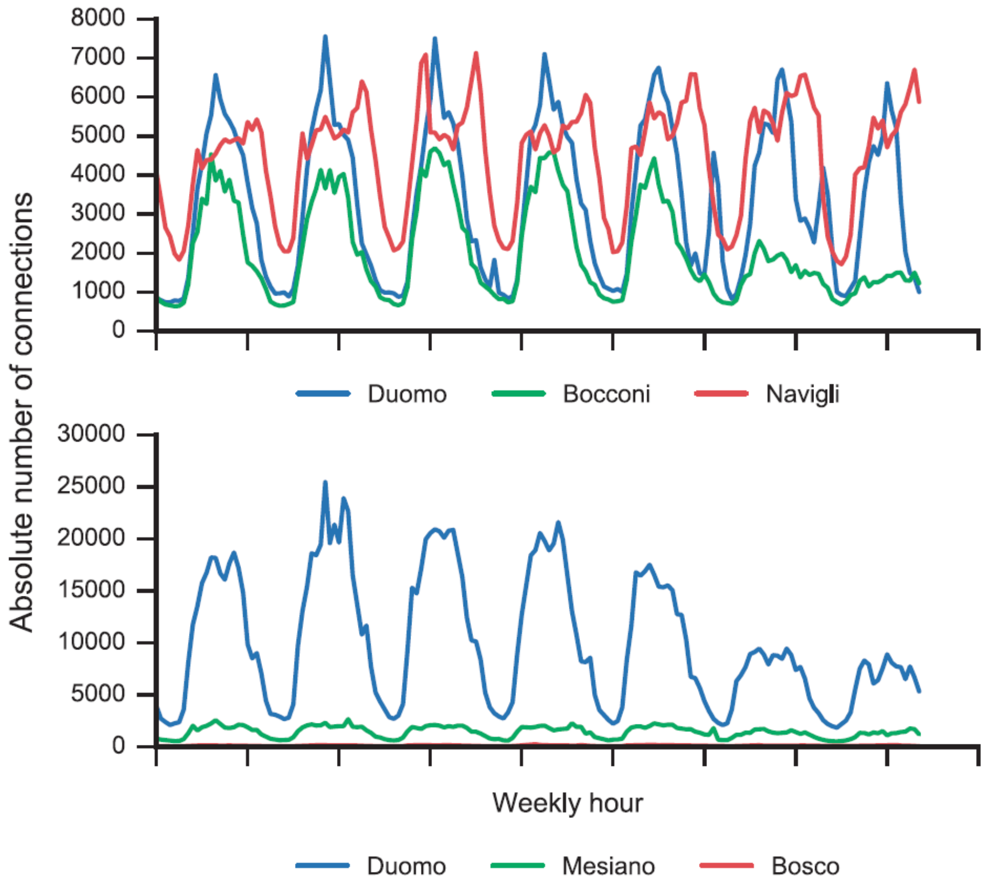
Square id	Time Interval	Country code	SMS-in activity	SMS-out activity	Call-in activity	Call-out activity	Internet traffic activity
1	1383606E+6	72	9.0299E-2	1.4195E-1	1.0804E-1	2.73E-1	9.2294
10	1383606E+6	72	6.7480E-2	1.0631E-1	5.9175E-2	1.0174E-2	5.8178

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Square id	Time Interval	Country code	SMS-in activity	SMS-out activity	Call-in activity	Call-out activity	Internet traffic activity
1	1383606E+6	0	1.7873E-3	NaN	NaN	NaN	NaN
1	1383606E+6	33	NaN	NaN	NaN	NaN	2.6137E-2
1	1383606E+6	39	8.8512E-2	1.4195E-1	1.0804E-1	2.73E-2	9.2032
10	1383606E+6	33	NaN	NaN	NaN	NaN	2.8653E-2
10	1383606E+6	39	6.7480E-2	1.0631E-1	5.9175E-2	1.0174E-2	5.7891

Square id	Time Interval	Country code	SMS-in activity	SMS-out activity	Call-in activity	Call-out activity	Internet traffic activity
1	1383606E+6	72	9.0299E-2	1.4195E-1	1.0804E-1	2.73E-1	9.2294
10	1383606E+6	72	6.7480E-2	1.0631E-1	5.9175E-2	1.0174E-2	5.8178

Preliminaries on data Collection for MTP-NT – Dataset used in this work



Source: Barlacchi

Preliminaries on data Collection for MTP-NT – Dataset used in this work

Sudden changes in network usage can make predictions inaccurate[1]

Despite changes, patterns can be identified and models could be developed[2]

Traffic “hubs” can be a good source of information to traffic prediction[3,4]

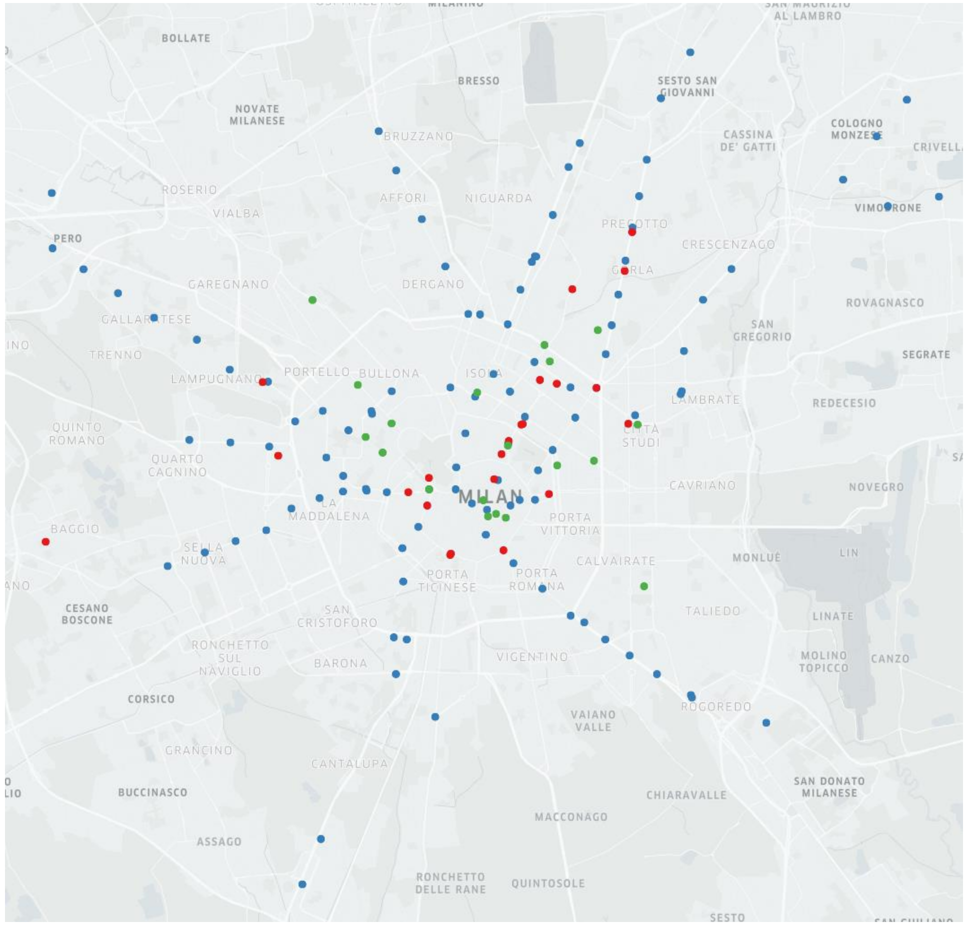
[1 D’ALCONZO, A. et al . A survey on big data for network traffic monitoring and analysis. IEEE Transactions on Network and Service Management, v. 16, n. 3, p. 800–813, 2019.

[2] PIROZMAND, P. et al. Human mobility in opportunistic networks: Characteristics, models and prediction methods. Journal of Network and Computer Applications, Elsevier, v. 42, p. 45–58, 2014.

[3] WANG, H. et al. Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks. In: Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking. New York, NY, USA: Association for Computing Machinery, 2015. (HOTPOST ’15), p. 19–24. ISBN 9781450335171. Disponível em: <https://doi.org/10.1145/2757513.2757518>.

[4] Wang, X. et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis. In: 2017 IEEE 25th International Conference on Network Protocols (ICNP). [S.l.: s.n.], 2017. p. 1–10.

Preliminaries on data Collection for MTP-NT – Dataset used in this work



Source: the author

Contents

Framework structure and fundamentation – Mathematical formalization of MTP-NT operations

10,000 regions

1 CDR sample for each 10 minutes, from November 1st to December 31st of 2013. Total: 8928 traffic samples for each region

$x^{m,n}(t)$

Region m, n for m, n between 0 and 99

Time period t between $13832604E+5$ to $13886166E+45$

Framework structure and fundamentation – Mathematical formalization of MTP-NT operations

Moore neighborhood

$$|m - m'| \ \leq d, |n - n'| \leq d$$

For a region $x^{m,n}$ the group of neighbors N within a degree d of distance is the group $N(x^{m,n})_d$

$$N(x^{m,n})_d = \left\{x^{m,n} \big| |m - m'| \ \leq d, |n - n'| \leq d \ \forall m, n, m', n' \ \in \{0, 1, \ldots, 99\}\right\}$$

Framework structure and fundamentation – Mathematical formalization of MTP-NT operations

$N(x^{m,n})_d$ **Scales with the increment of moore neighborhoods**

Table 5 – Number of regions and data samples in a 24 hour interval with increasing neighborhoods.

Neighborhoods	1	2	3	4	5
Total regions	9	25	49	81	121
Samples in 24 hours	1,296	3,600	7,056	11,664	17,424

Source: the author

Framework structure and fundamentation – Mathematical formalization of MTP-NT operations

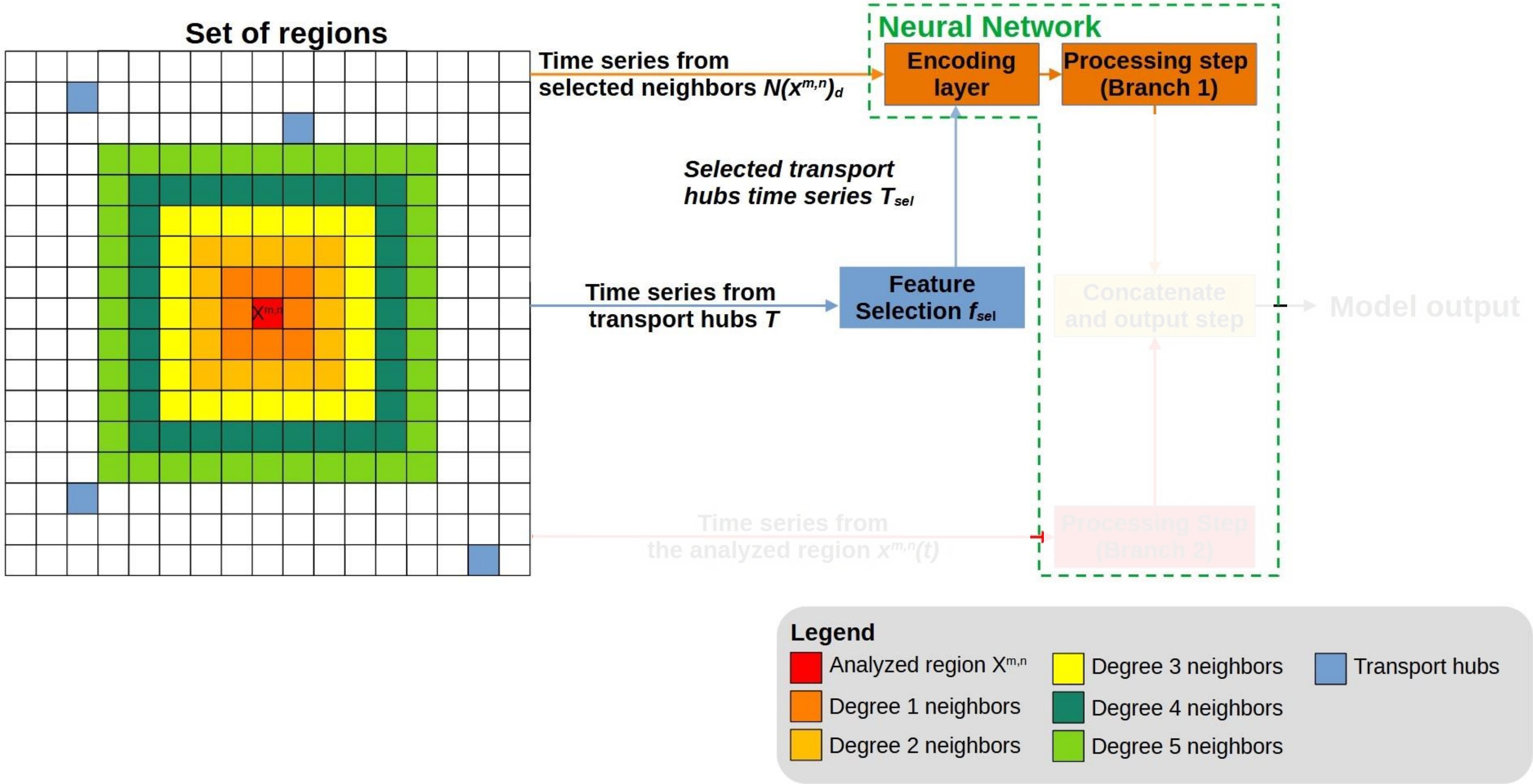
T: Set of transport hubs

f: Feature selection to choose the most pertinent transport hubs to the given region

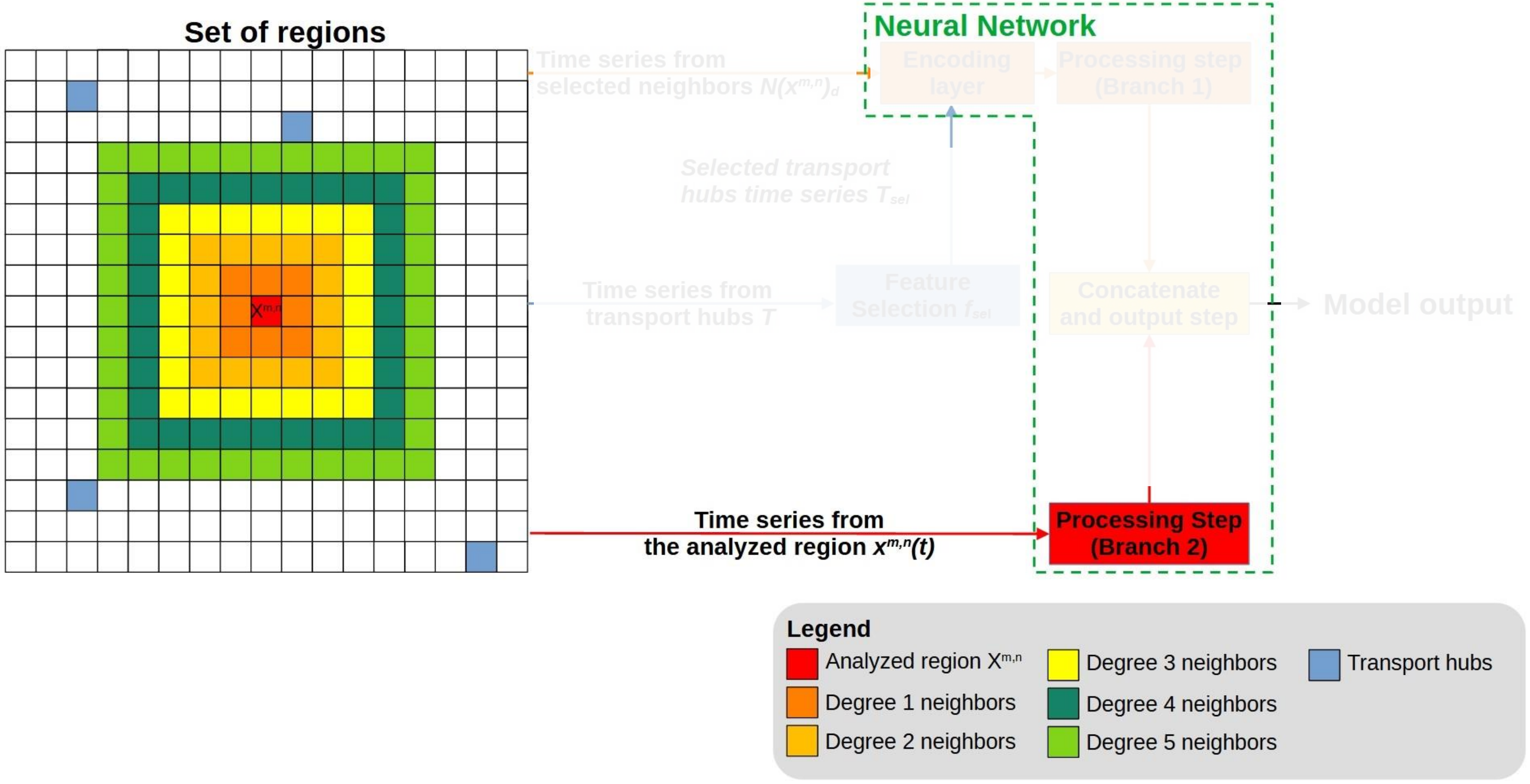
$T_{sel}(x^{m,n})$: Group of transport hubs selected to a given region

$$\boldsymbol{T_{sel}(x^{m,n}) = \{x^{m',n'} | f_{sel}(x^{m,n}(t), x^{m',n'}(t)) \forall x^{m',n'} \in T\}}$$

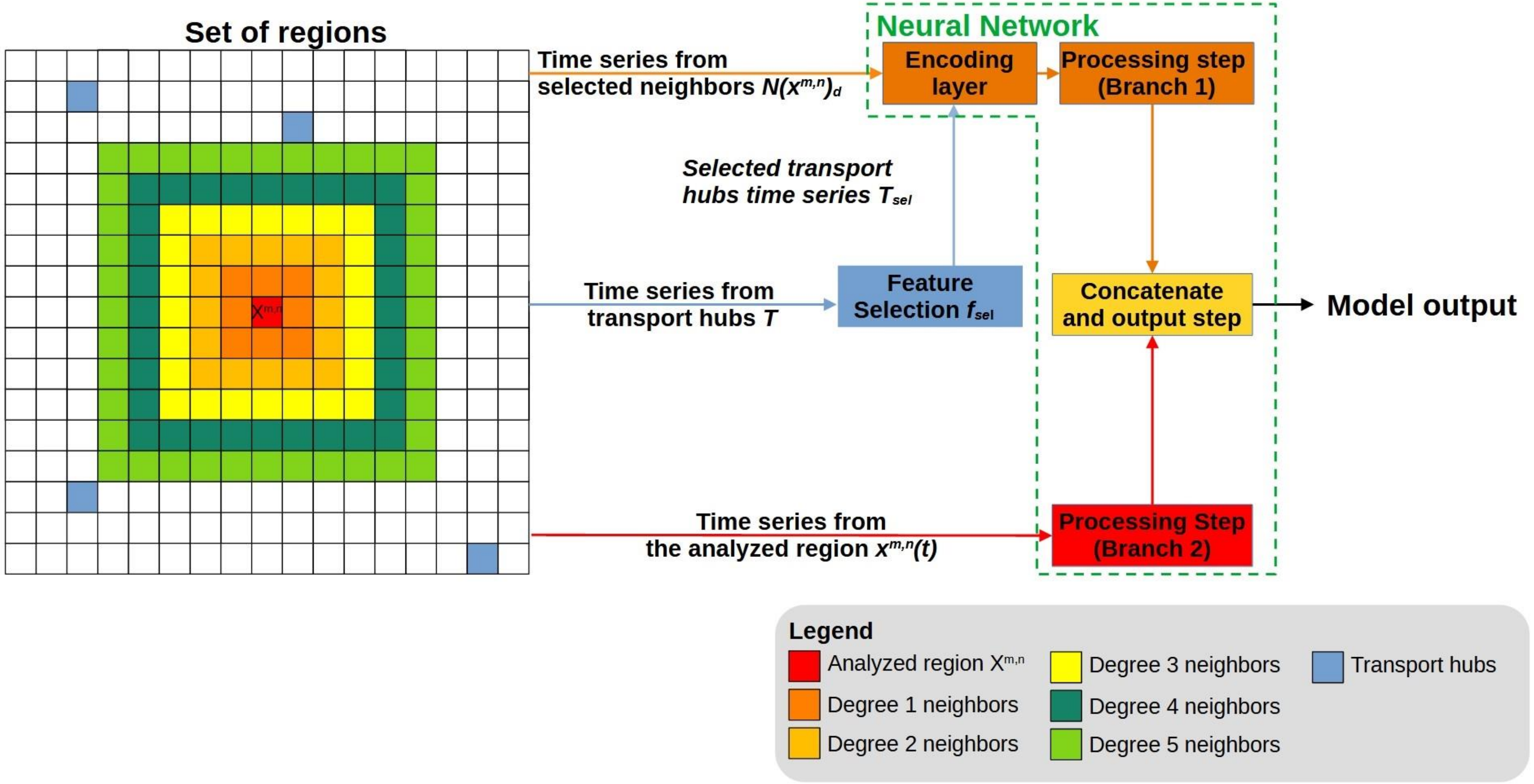
Framework structure and fundamentation – Mathematical formalization of MTP-NT operations



Framework structure and fundamentation – Mathematical formalization of MTP-NT operations



Framework structure and fundamentation – Mathematical formalization of MTP-NT operations



Source: the author

MTP-NT's framework architecture

Branch 1

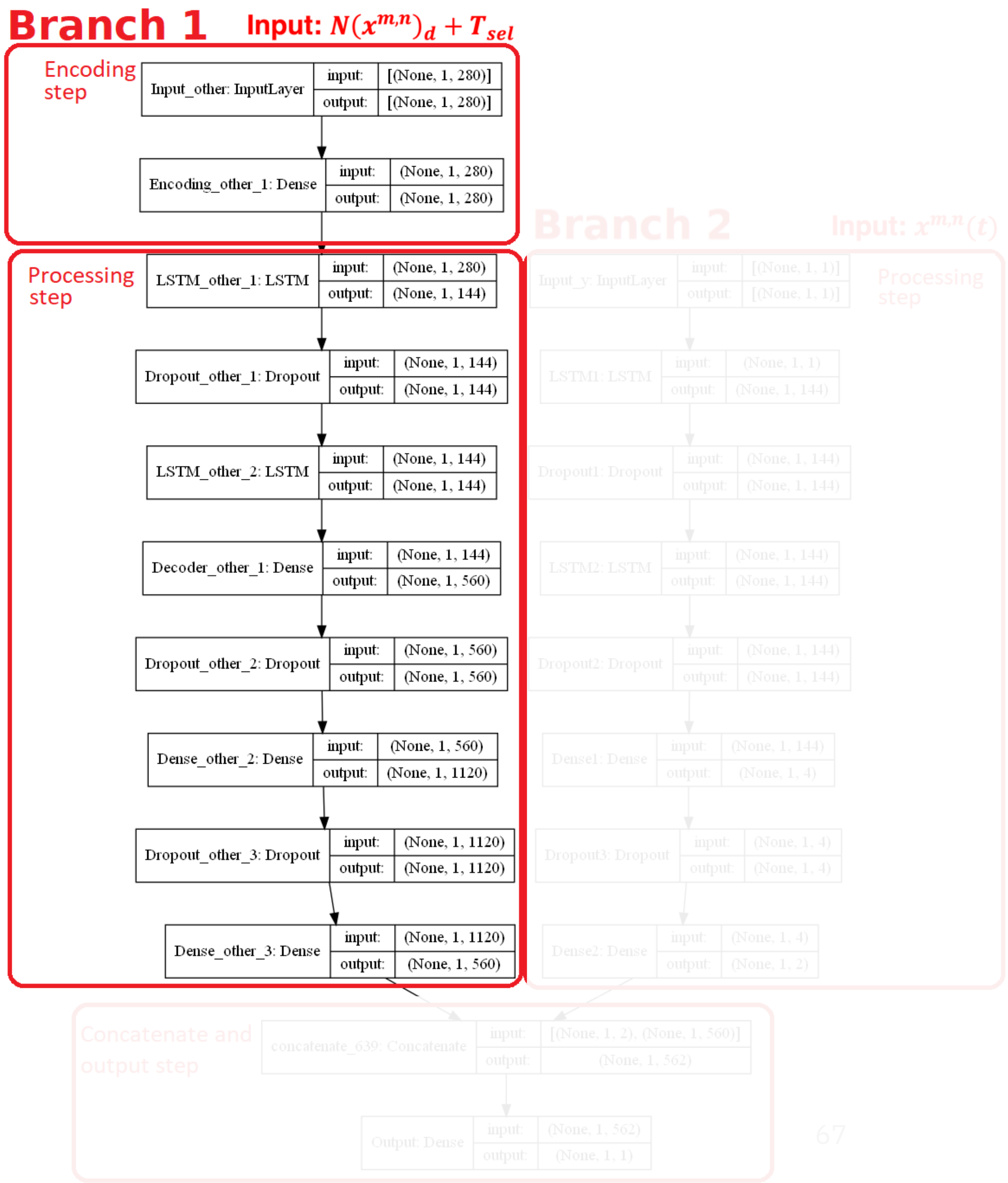
1. **Decode the input and begin the compression**
   $Input\_other$; $Encoding\_other\_1$

2. **Temporal relations**
   $LSTM\_other\_1$; $Dropout\_other\_1$

3. **Temporal relations**
   $LSTM\_other\_2$; $Decoder\_other\_1$; $Dropout\_other\_2$

4. **General purpose correlations and reduce overfitting**
   $Dense\_other\_2$; $Dropout\_other\_3$

5. **Concatenation of Branches**
   $Dense\_other\_3$



MTP-NT's framework architecture

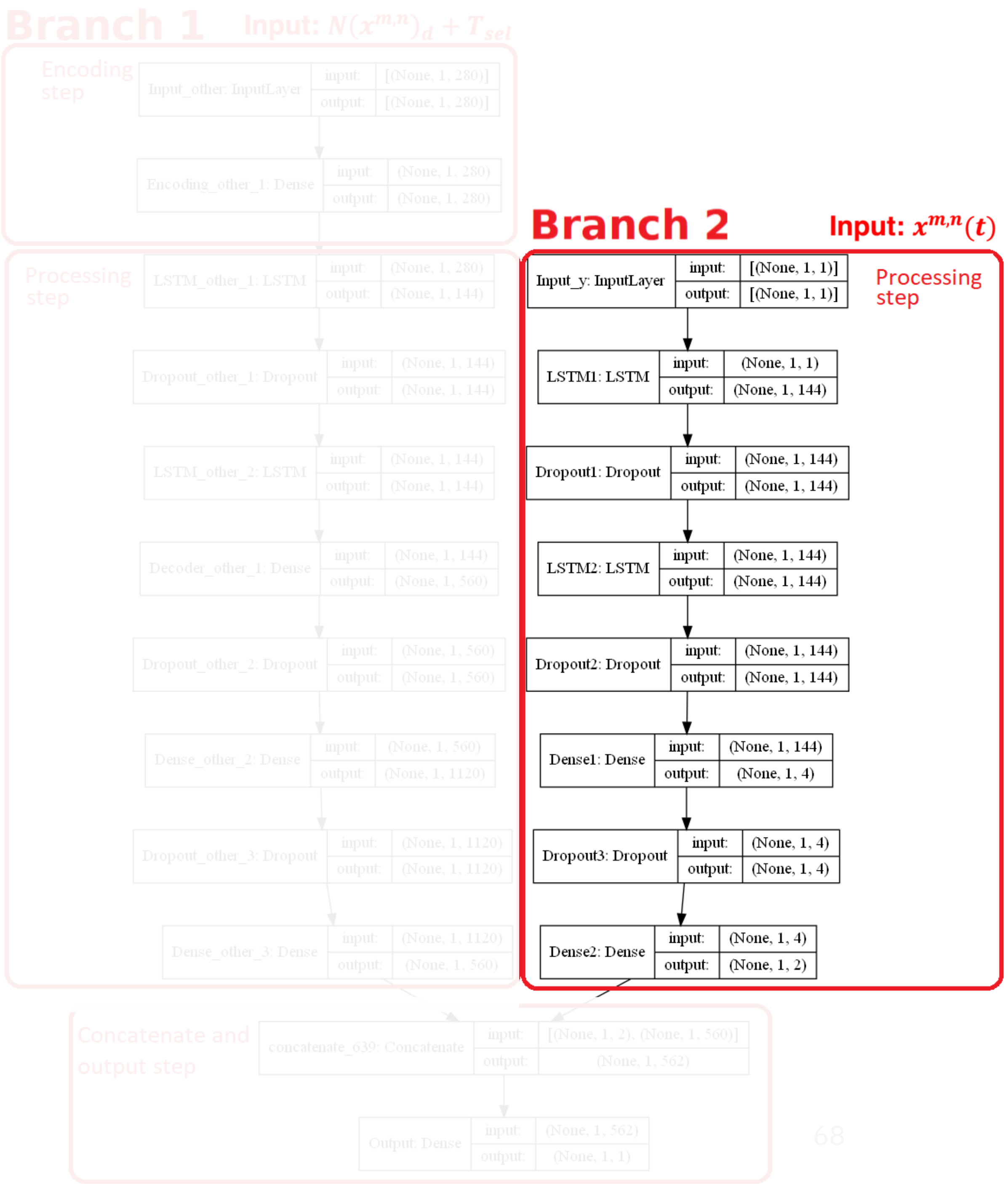
Branch 2

1. **Input layer**
   $Input\_y$

2. **Temporal relations**
   $LSTM1$; $Dropout1$
   $LSTM2$;$Dropout2$

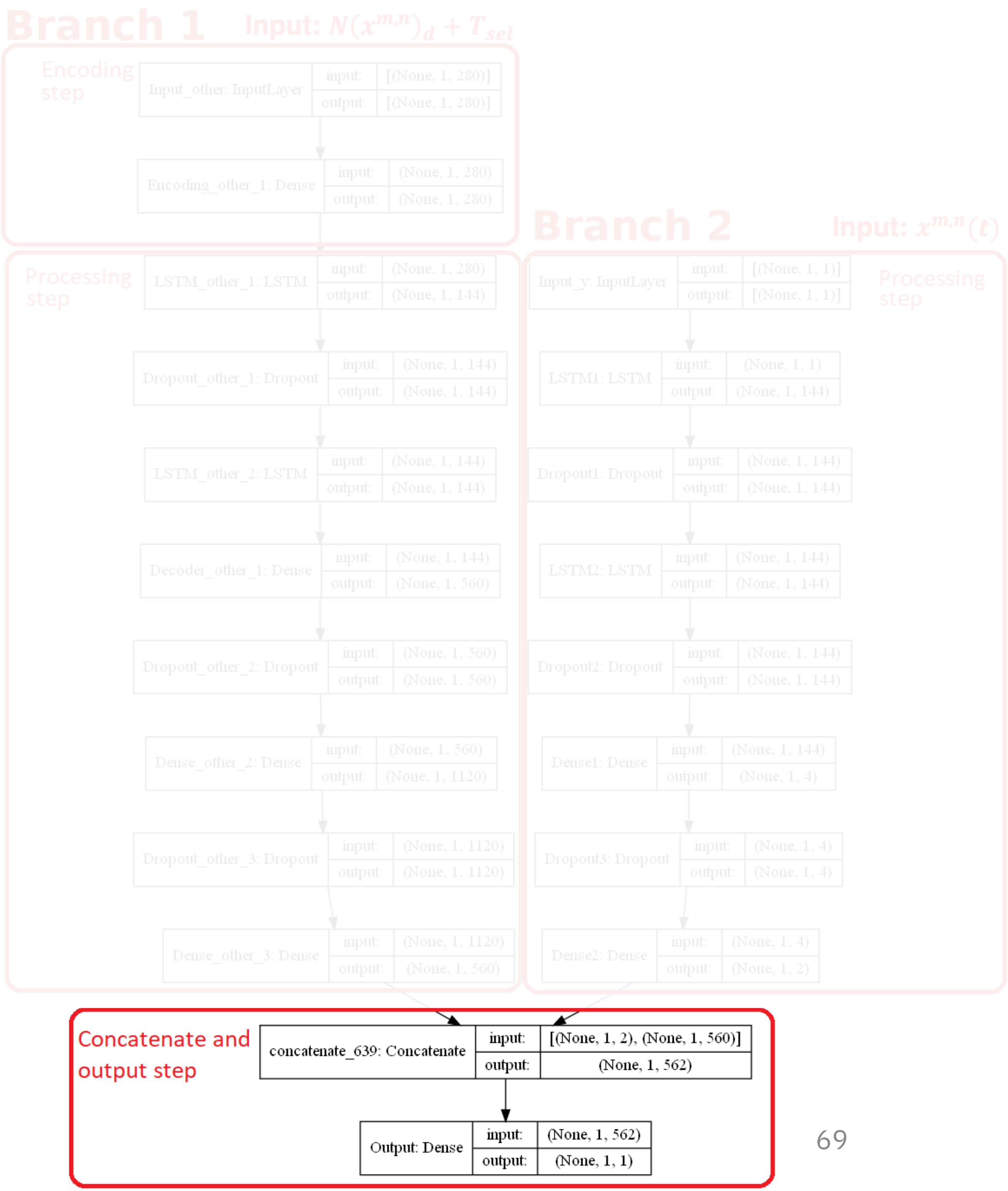
3. **General purpose correlations and reduce overfitting**
   $Dense1$
   $Dropout3$; $Dense2$

4. **Concatenation of Branches**
   $Dense2$

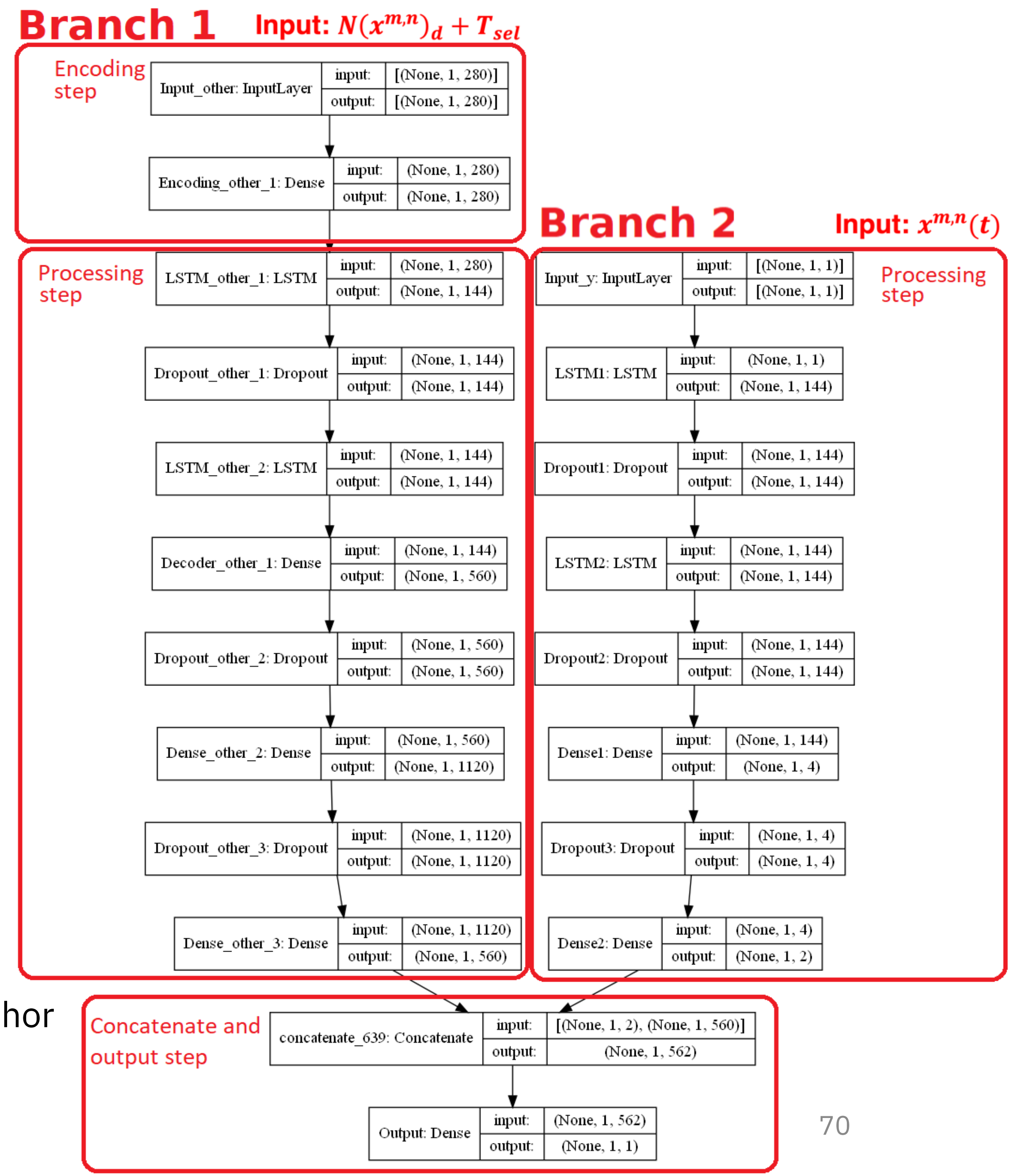


MTP-NT’s framework architecture

Concatenate and output step



MTP-NT’s framework architecture



Source: the author

Contents

Experimental results - Setup

Moore regions: 1; 1~2; 1~3; ... ; 1~5

Experimental results - Setup

Moore regions: 1; 1~2; 1~3; ... ; 1~5

Feature selection algorithms

- **F-test**
- **Pearson correlation coeficient**
- **Moore test**

Experimental results - Setup

Moore regions: 1; 1~2; 1~3; ... ; 1~5

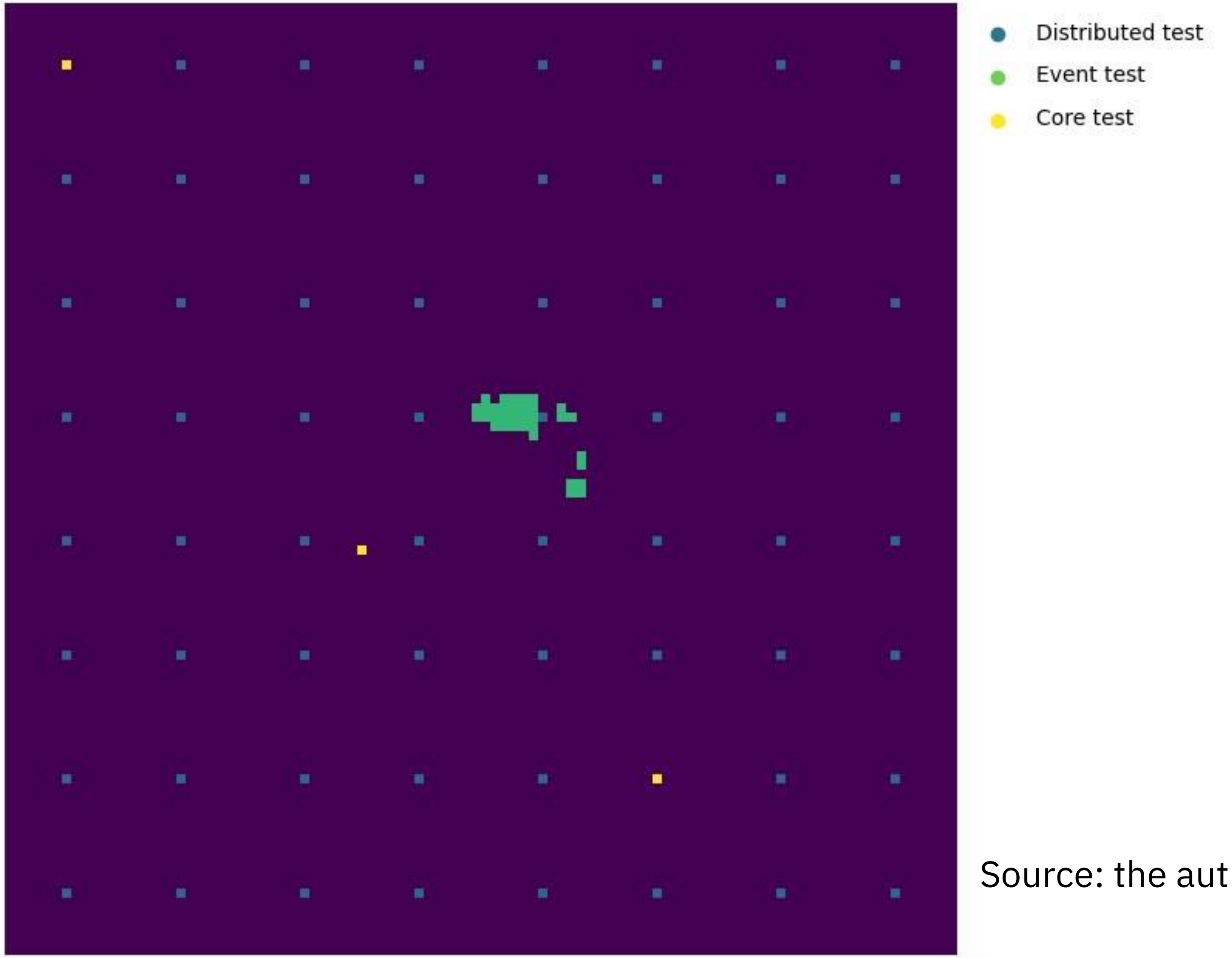
Feature selection algorithms

- **F-test**
- **Pearson correlation coeficient**
- **Moore test**

Tests

- **Distributed tests**
- **Core tests**
- **Event regions**

Experimental results - Setup



Source: the author

Experimental results - Setup

Moore regions: 1; 1~2; 1~3; ... ; 1~5

Feature selection algorithms

- **F-test**
- **Pearson correlation coeficient**
- **Moore test**

Tests

- **Distributed tests**
- **Core tests**
- **Event regions**

Variations

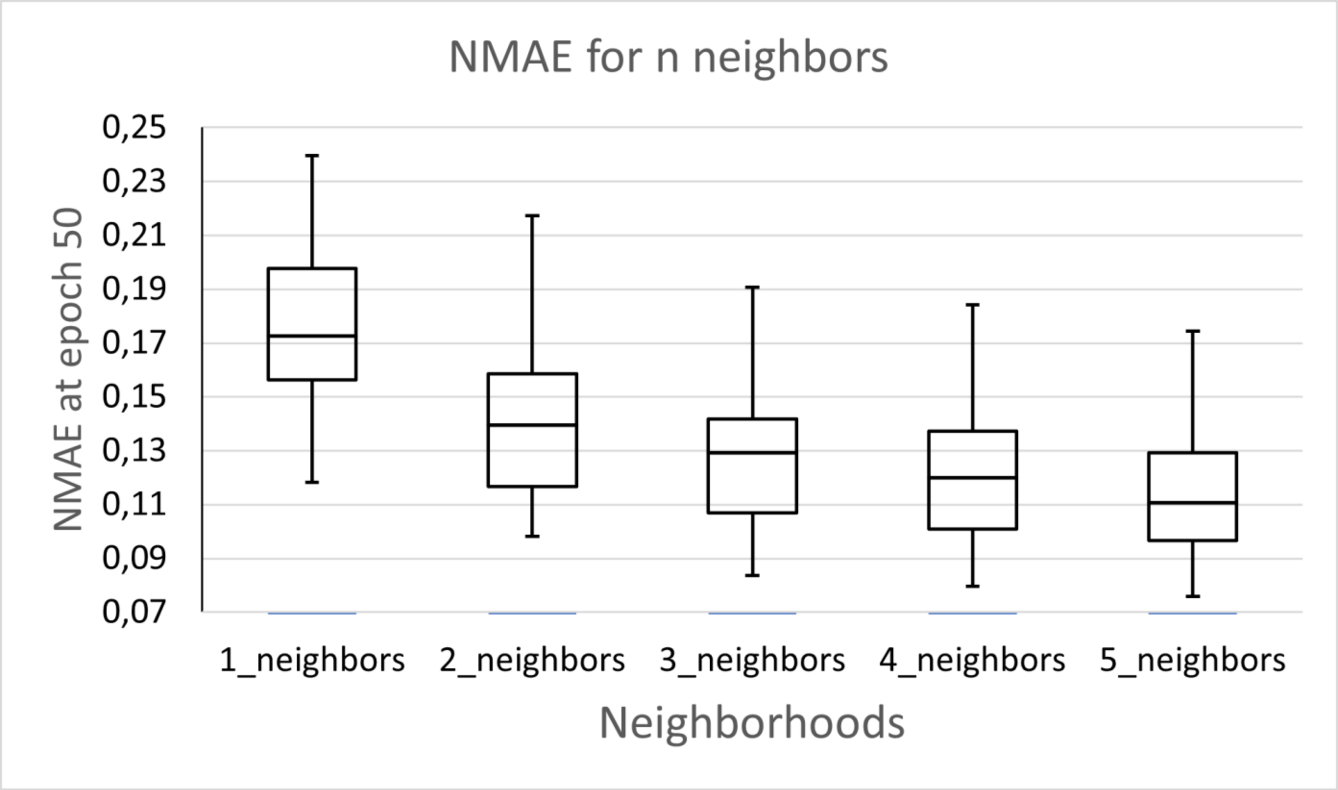
- **With transport hubs**
- **Without transport hubs**

Experimental results - Setup

Evaluation metric

$$NMAE(y, \hat{y}) = \frac{\sum |\hat{y} - y|}{\sum y} \tag{12}$$

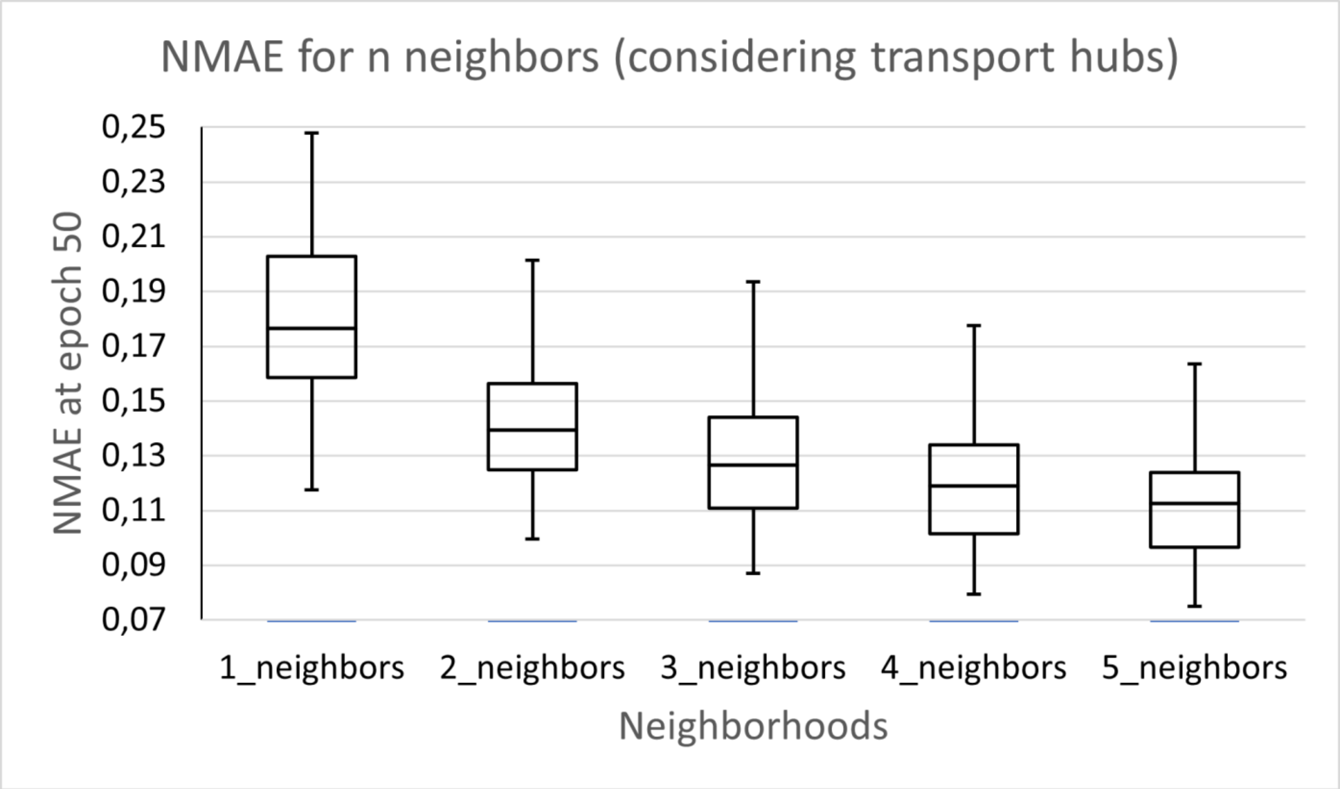
Experimental results - Results



Source: the author

Descendent error

Experimental results - Results

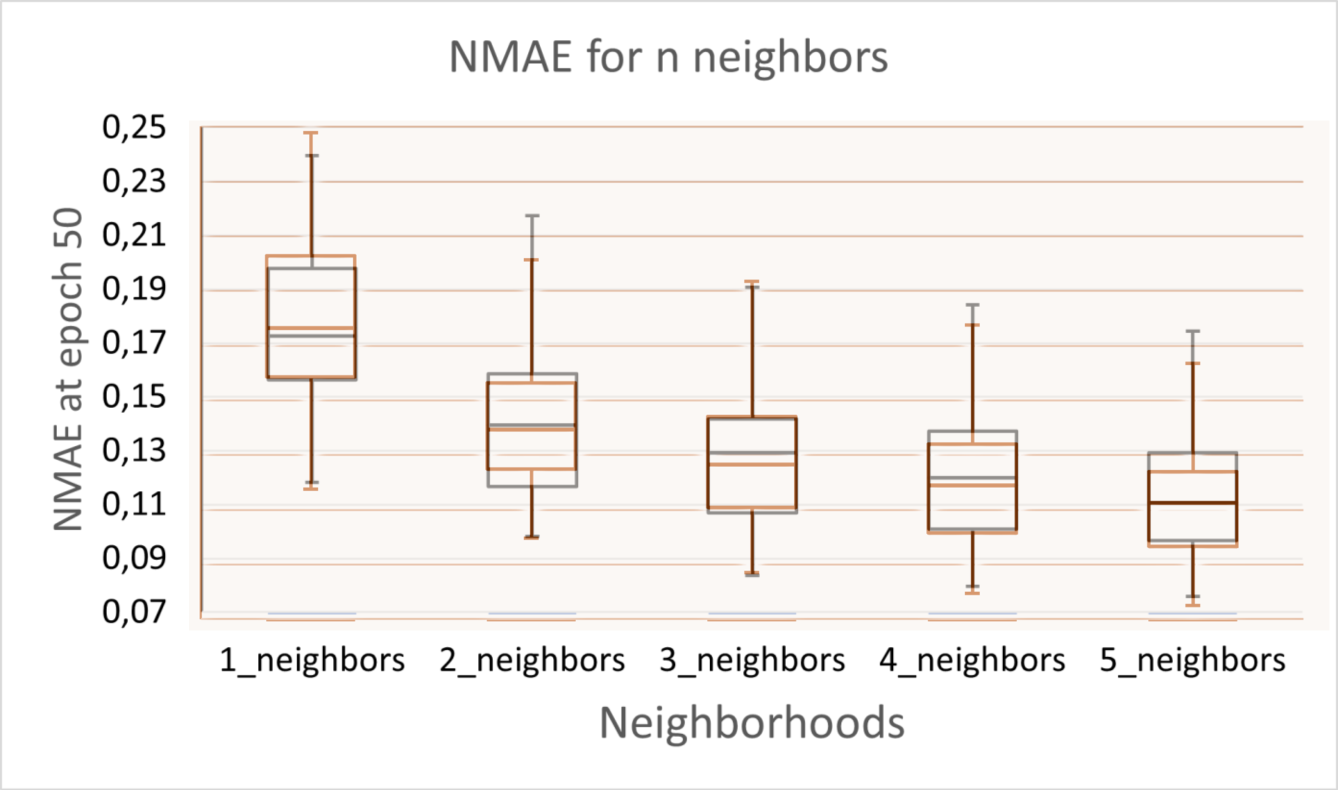


Source: the author

Better overall performance, specially in:

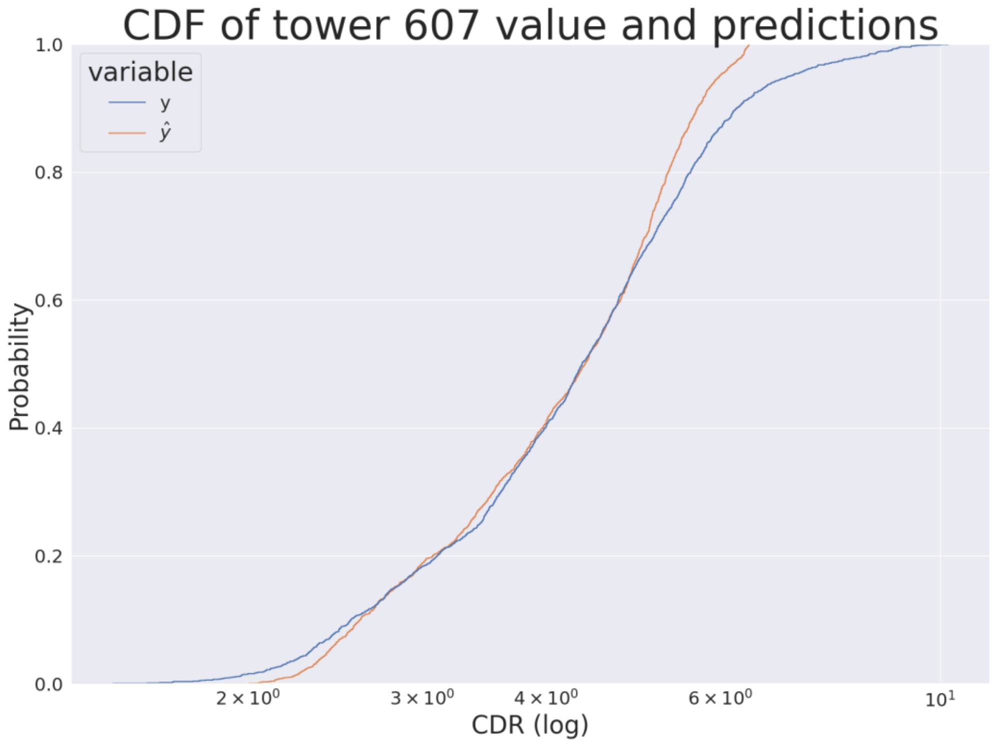
- **Outliers**
- **Lower Moore distance models**

Experimental results - Results



Source: the author

Experimental results - Results



Source: the author

Considerations of transport hubs

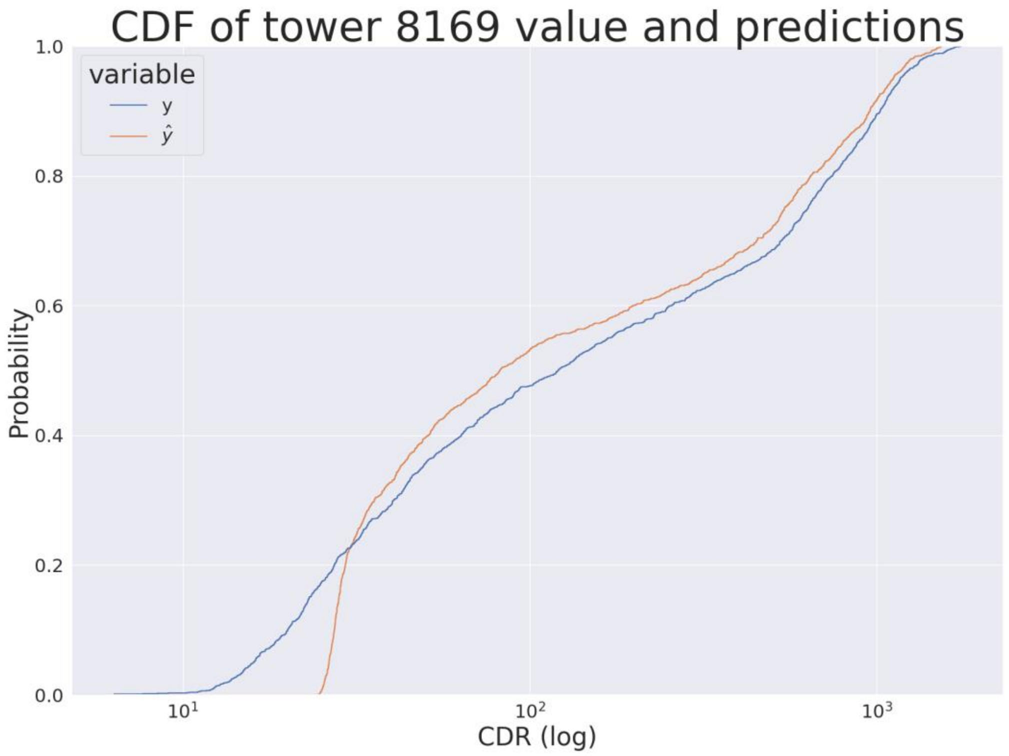
- **Can be important to anticipate fast changing demand from non-seasonal events[1]**
- **More important in central regions of the city**

Region 607 (near Vigano)

- **Normal tests:** $NMAE = 12\%$
- **Transport hubs tests:** $NMAE = 11\%$

[1] Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. [S.l.: s.n.], 2017. p. 1–9.

Experimental results - Results



Source: the author

Considerations of transport hubs

- **Can be important to anticipate fast changing demand from non-seasonal events[1]**
- **More important in central regions of the city**

Region 8169 (mall near Parco Nord Milano)

- **Normal tests: $NMAE = 13\%$**
- **Transport hubs tests: $NMAE = 11\%$**

[1] Wang, J. et al. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. [S.l.: s.n.], 2017. p. 1–9.

Experimental results - Results

How can MTP-NT perform in a different scenario?

Table 6 – NMAE in tests with 10-minute and 1-hour observations, varying the usage of transport hubs.

Window size	NMAE with transport data	NMAE without transport data
10 minutes	0.1120	0.1100
1 hour	0.1355	0.1441

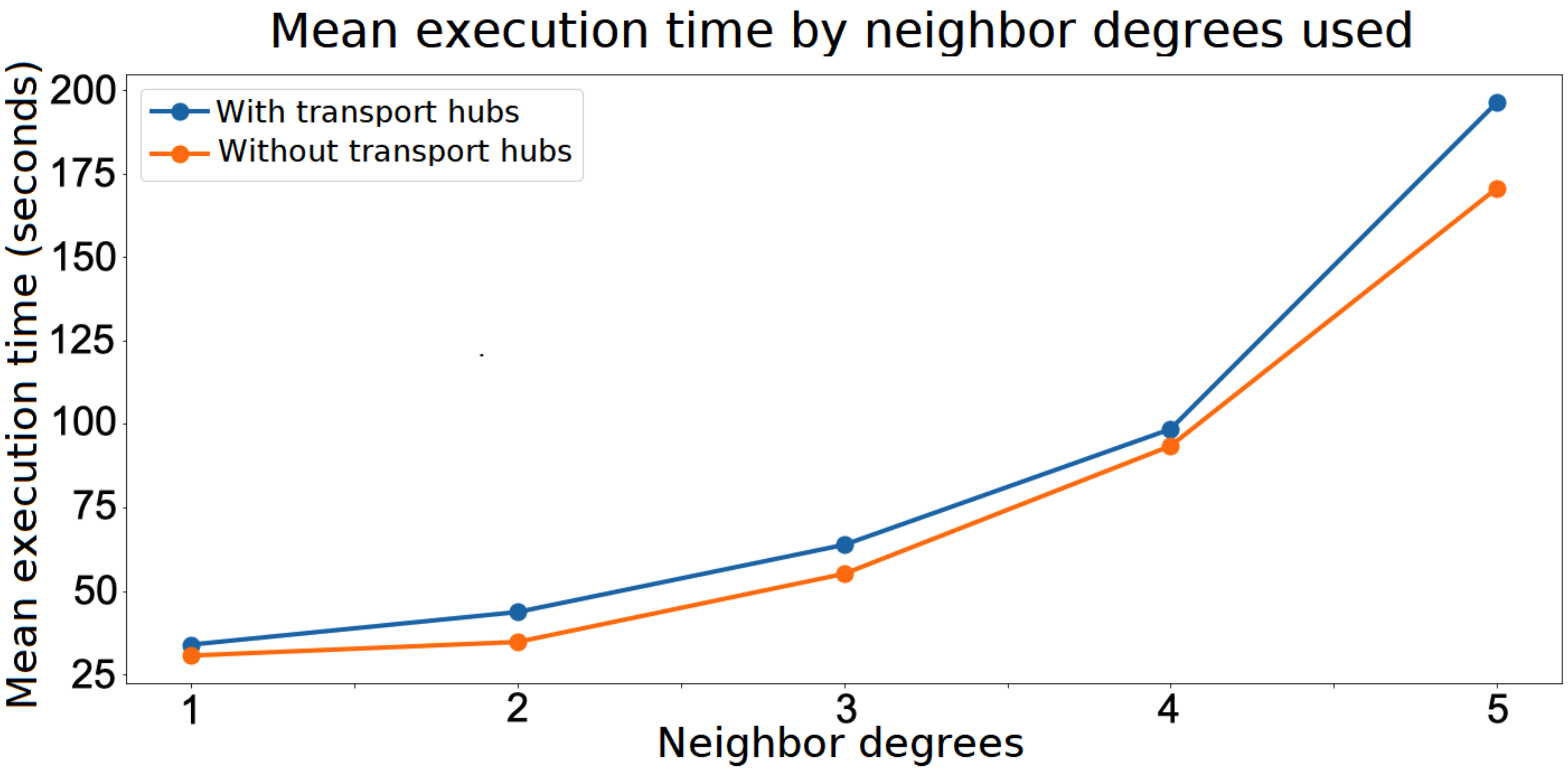
Source: the author

Experimental results – Execution Time Evaluation

Growth of regions with the maximum Moore distance d considered

$$N_d = N_{d+1} + 8(d+1)$$

Experimental results – Execution Time Evaluation



Source: the author

Experimental results – Performance Analysis

Auto Regressive Integrated Moving Average - ARIMA

- **Autoregression: self correlation (p)**

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p}$$

- **Integration: stationary time series (d)**

$$Z_t = Y_{t+1} - Y_t \qquad \ldots d = 1$$
$$Q_t = Z_{t+1} - Z_t \qquad \ldots d = 2$$
$$\ldots$$

- **Moving Average: relation between error of previous samples and the actual (q)**

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \ldots + \omega_q \varepsilon_{t-q} + \varepsilon_t$$

Experimental results – Performance Analysis

Auto Regressive Integrated Moving Average – ARIMA

Best model: p=36;d=1;q=0

Experimental results – Performance Analysis

Holt-Winters (HW) model: aditive trend

Prophet: daily and weenkly seazonality

LSTM model

- **Standard scaler**
- **128-cell LSTM layer + Dropout (10%)**
- **128-cell LSTM layer + Dropout (10%)**
- **Dense layer**

Experimental results – Performance Analysis

Table 6 – NMAE among different benchmarking techniques in Distributed, Core and Event tests.

	Distributed test	Core test	Event test
ARIMA	51.00	65.03	60.014
HW	11.78	9.34	15.16
LSTM	57.03	54.35	67.06
Prophet	61.00	94.66	178.65
MTP-NT	**11.47**	**8.22**	**11.62**

Source: the author

Experimental results – Performance Analysis

Table 6 – NMAE among different benchmarking techniques in Distributed, Core and Event tests.

	Distributed test	Core test	Event test
ARIMA	51.00	65.03	60.014
HW	11.78	9.34	15.16
LSTM	57.03	54.35	67.06
Prophet	61.00	94.66	178.65
MTP-NT	**11.47**	**8.22**	**11.62**

Source: the author

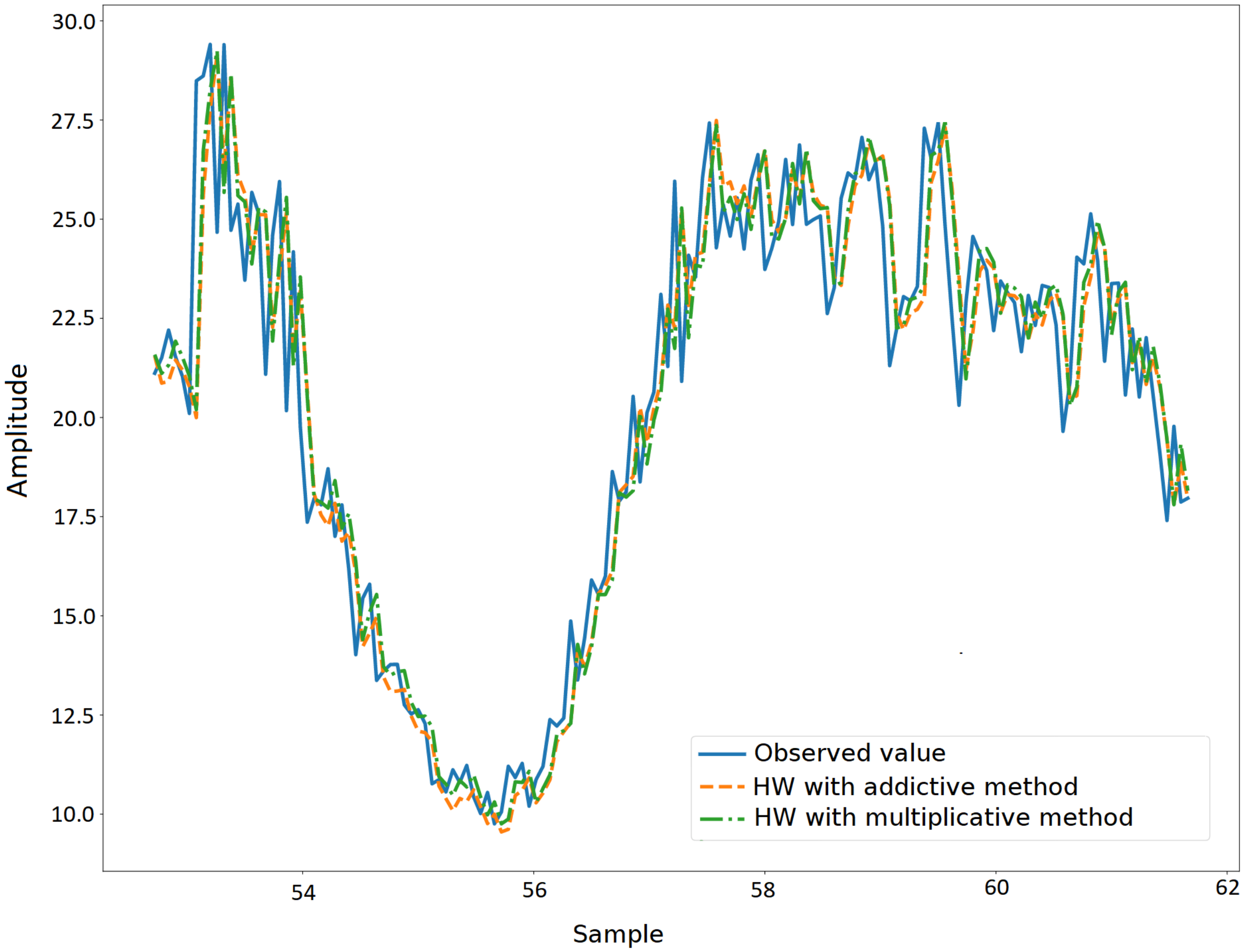
(Wang 2017b) 45% drop in error compared to LSTM

MTP-NT: 77%

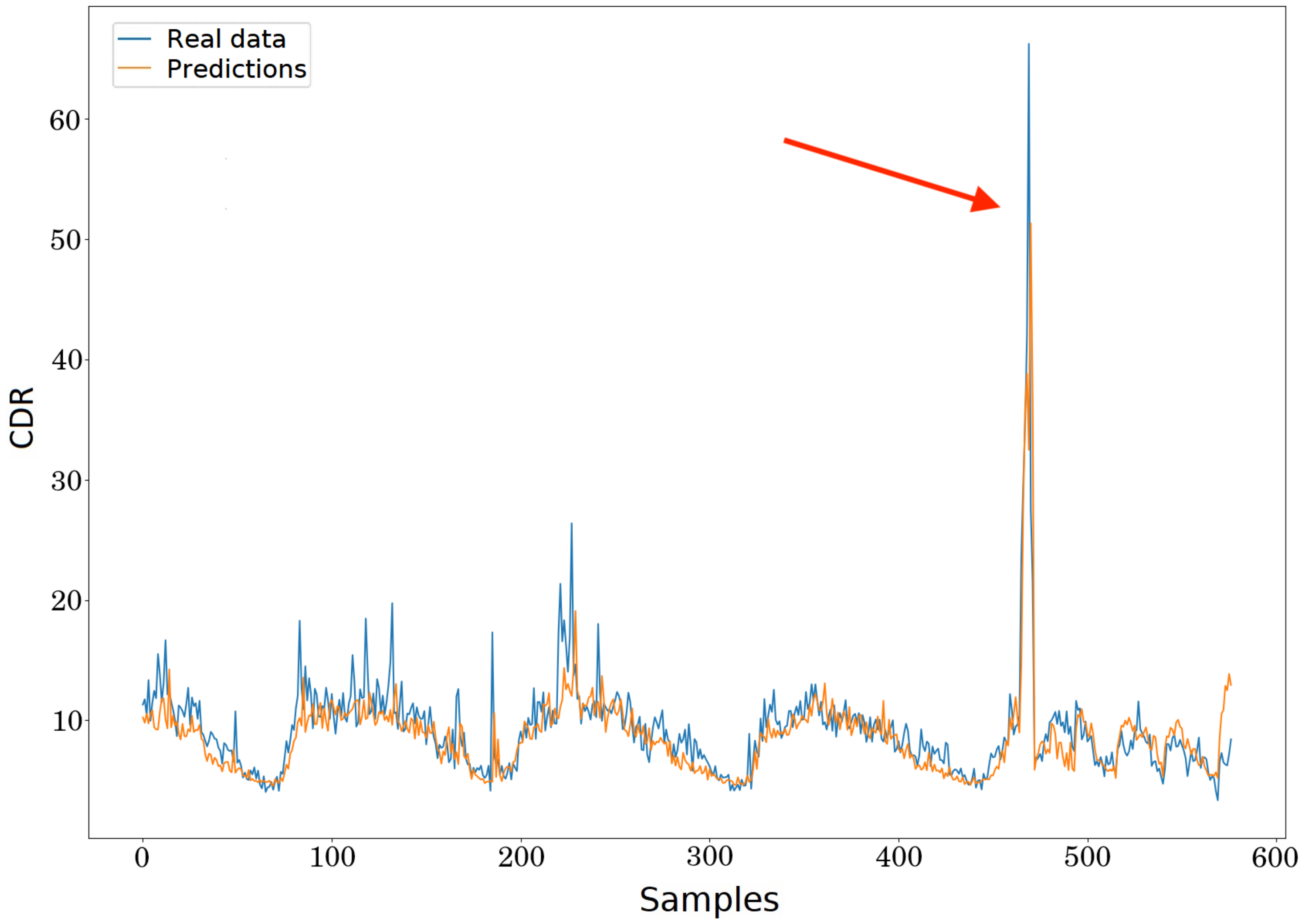
(Wang 2017b) 62% drop in error compared to HW

MTP-NT: similar

Experimental results – Performance Analysis



Experimental results – Performance Analysis



Contents

Scenario in 5G

Data driven technologies in mobile networks

Scenario in 5G

Data driven technologies in mobile networks

Large urban centers

- Dynamic scenarios
- Patterns and irregular factors
- Strict QoS/QoE metrics

Scenario in 5G

Data driven technologies in mobile networks

Large urban centers

- Dynamic scenarios
- Patterns and irregular factors
- Strict QoS/QoE metrics

Cloud based and compatible with 3GPP architecture purposals

Scenario in 5G

Data driven technologies in mobile networks

Large urban centers

- Dynamic scenarios
- Patterns and irregular factors
- Strict QoS/QoE metrics

Cloud based architecture; 3GPP architecture purposal

GDPR and other major privacy policies

MTP-NT

Public transport hubs

As a way to bring new information to the model

Improve the coverage of aperiodic events

MTP-NT

Public transport hubs

As a way to bring new information to the model

Improve the coverage of aperiodic events

Performance aligned with state of art research

MTP-NT

Public transport hubs

As a way to bring new information to the model

Improve the coverage of aperiodic events

Performance aligned with state of art research

Performance and accuracy tradeoff

The model is flexible and allows a configuration to best fit the accuracy-model size best scenario

MTP-NT

Public transport hubs

As a way to bring new information to the model

Improve the coverage of aperiodic events

Performance aligned with state of art research

Performance and accuracy tradeoff

The model is flexible and allows a configuration to best fit the accuracy-model size best scenario

Multiple tests to ensure the performance

Distributed
Core
Event

1-hour compilling

MTP-NT

Open-source

Public transport hubs

As a way to bring new information to the model

Improve the coverage of aperiodic events

Performance aligned with state of art research

Performance and accuracy tradeoff

The model is flexible and allows a configuration to best fit the accuracy-model size best scenario

Multiple tests to ensure the performance

Distributed
Core
Event

1-hour compilling

Future work

More neighborhoods

Heavier models

GPUs and other advancements make cloud computing cheaper

Multi-region model

MTP-NT compiles a single region at a time

A new multi-region architecture can be better

Thanks!

A Mobile Traffic Predictor Enhanced by Neighboring Transportation Data (MTP-NT)

Patrick Luiz de Araújo

patrick@ufu.br

Prof. Dr. Rafael Pasquini

rafael.pasquini@ufu.br

Prof. Dr. Murillo Guimarães Carneiro

mgcarneiro@ufu.br