Wireless Big Data Analysis: A Machine Learning Perspective

Dongfeng Yuan

Shandong Provincial Key Lab. of Wireless Communication Technologies
Shandong University, Jinan, China
Email: dfyuan@sdu.edu.cn

Oct. 12, 2017

Contents

Introduction

➢ **5G mobile communication will be the beginning of a full scale Internet of Everything.**

Introduction

➢ 5G Vision: A Unifying Connectivity Fabric

5G KPIs

1000x data volume

1000x
higher mobile
data volumes

50/500 B devices

10-100x
higher number of
connected devices

Up to 10Gbps

10-100x
typical end-user
data rates

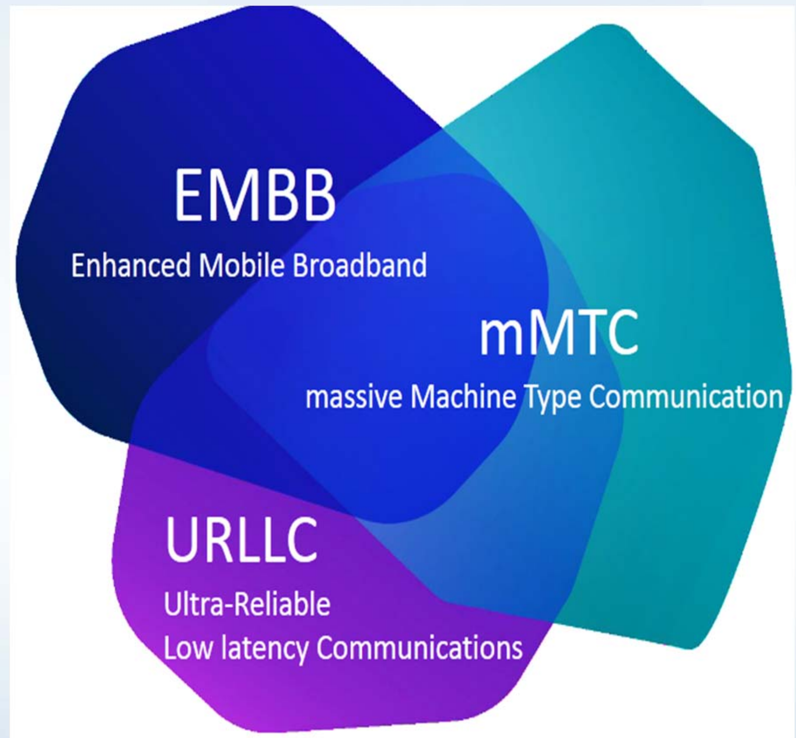
Few ms E2E

5x
lower latency

10 years

10x
longer battery life
for low-power devices

5G Scenarios



Introduction

➢ Enhanced Mobile Broadband

Ushering in the next era of immersive experiences and hyper-connectivity

3D/UHD video telepresence

Tactile Internet

Demanding conditions (venues)

Broadband to the home

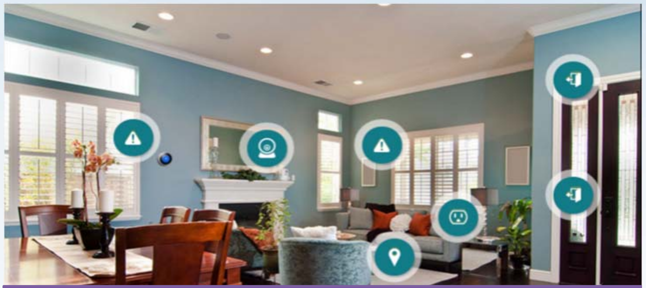
Extreme throughput **multi-gigabits per second**	**Ultra-low latency** **down to 1ms e2e latency**	**Uniform experience** **with much more capacity**

Introduction

➢ Massive Machine Type Communications

Optimizing to connect anything, anywhere with efficient, low cost communications

Smart cities



Smart homes

Wearables

Object tracking

Power efficient **Multi-year battery life**	**Low complexity** **Low device and network cost**	**Long range** **Deep coverage**

Introduction

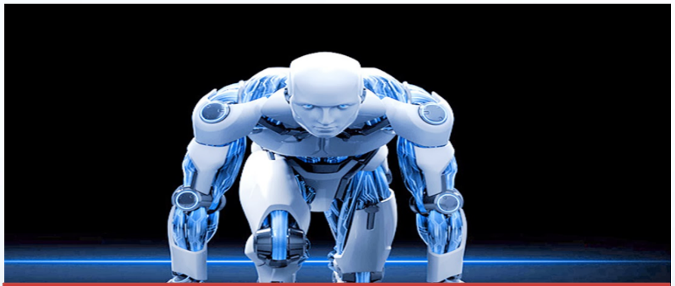
➢ Ultra-reliable, ultra-low latency communication

Enabling new Mission-critical Control Services

Autonomous vehicles

Medical

Industrial automation

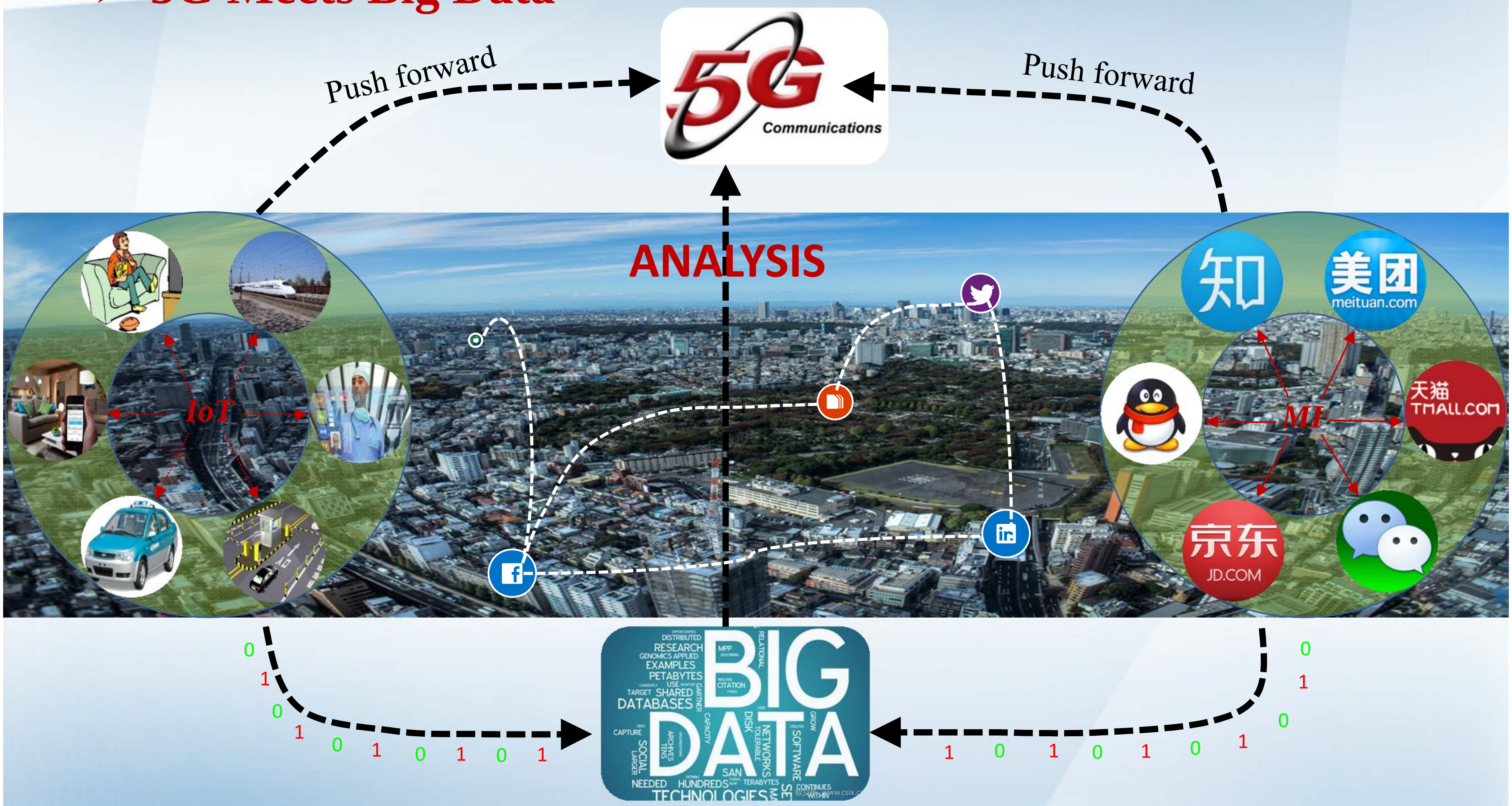


Robotics

High reliability **Extremely low loss rate**	**Ultra-low latency** **Down to 1ms e2e latency**	**High availability** **Multiple links for failure tolerance & mobility**

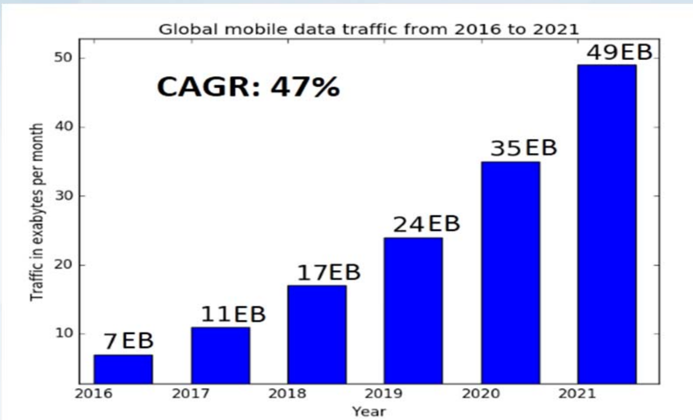
Introduction

➢ 5G Meets Big Data

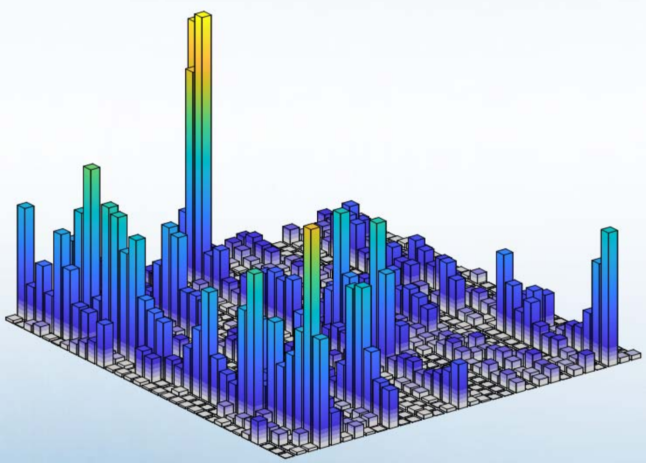


Introduction

➢ Characteristics of Wireless Big Data

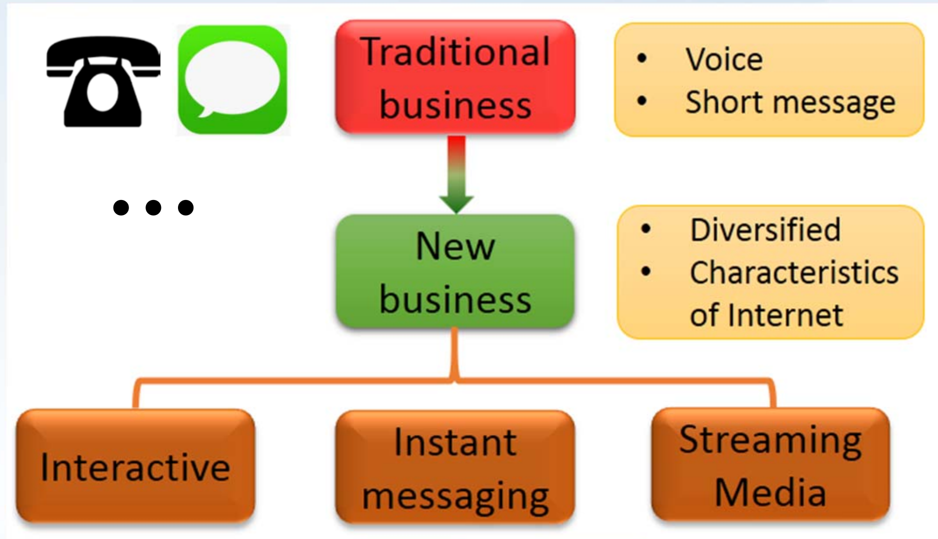


Volume of data



Uneven in the time domain and space domain

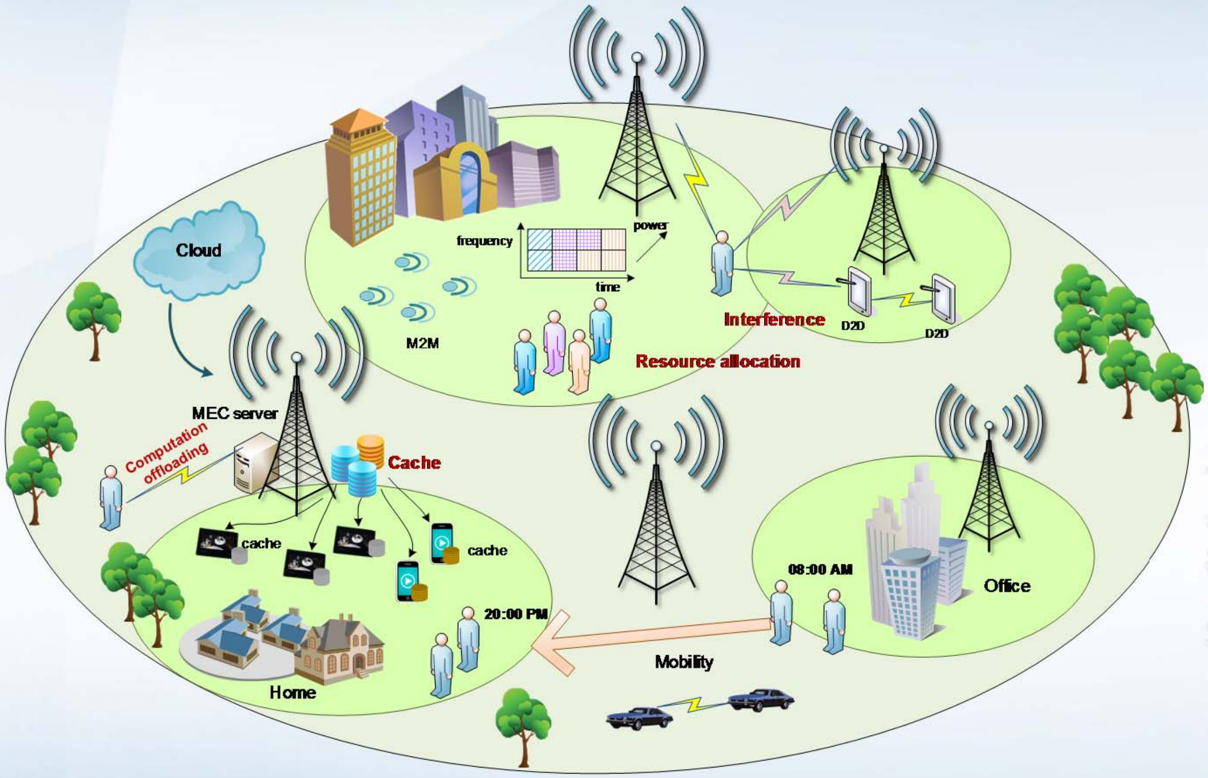
Business transform



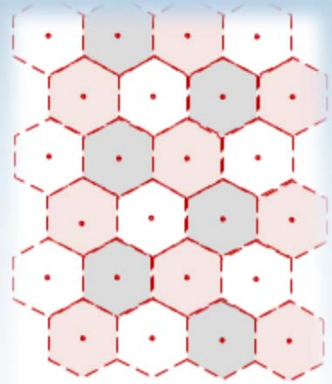
• • •

Introduction

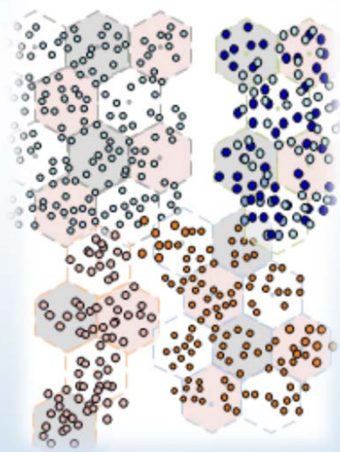
➢ Future Network Architecture



Ultra-dense and heterogeneous



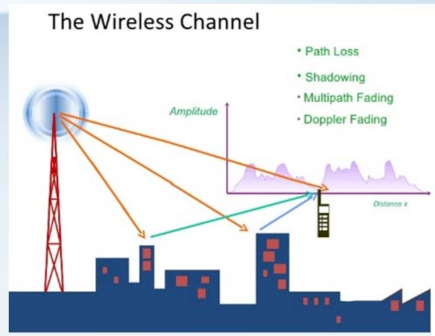
Bee Celluar Network: Coverage

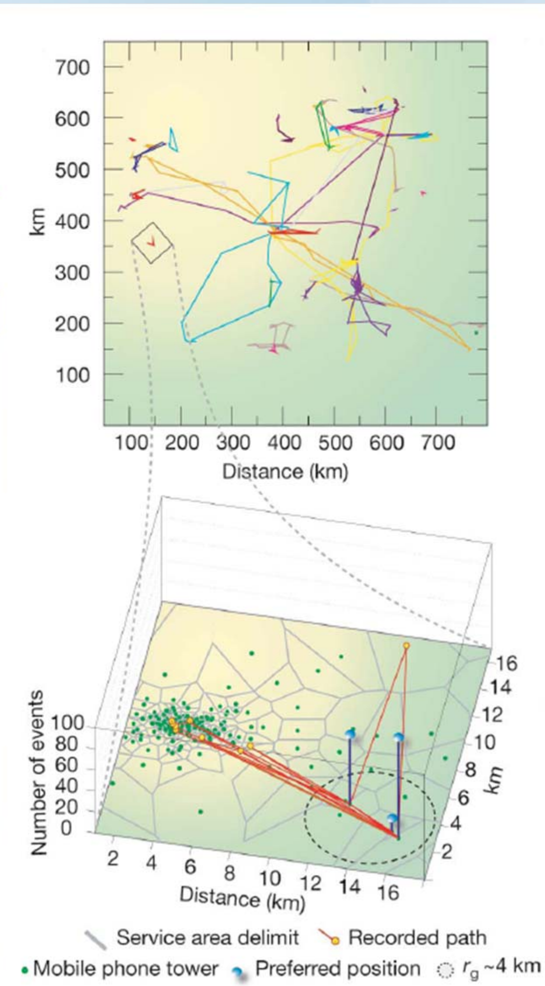


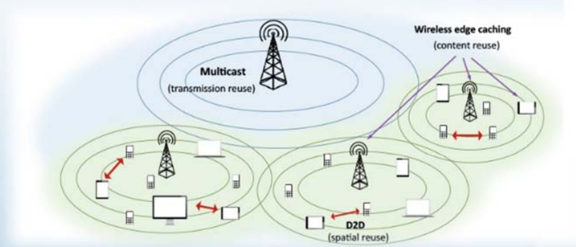
Bee Group Network: User behavior, intelligent

Introduction

- Channel analysis
  - Modeling
  - Identification
- Resource management
  - Accurate resource allocation
  - Scalable network capacity optimization
- Wireless caching
  - Automatic popular content detection
  - Social network analysis
- User mobility pattern analysis
  - Spatial-temporal distribution
  - Location prediction
  - Frequently pattern mining







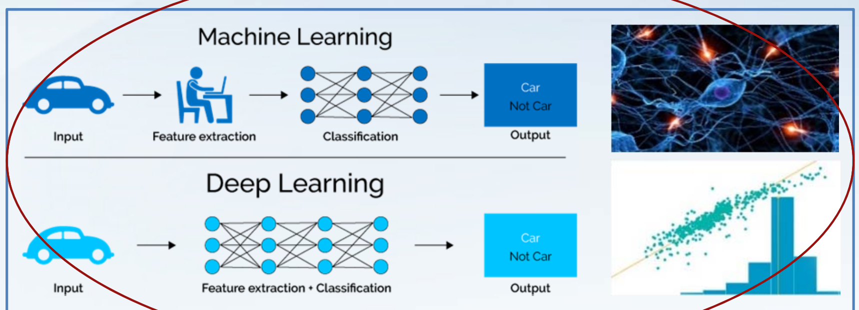
Introduction

➢ Key components in wireless big data analysis

➢ Clustering
- K-Means, DBSCAN

➢ Association analysis
- Apriori

➢ Bagging and Boost
- AdaBoost

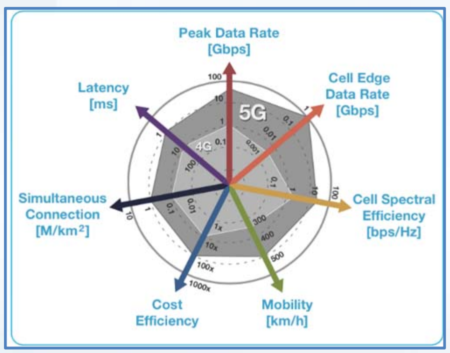


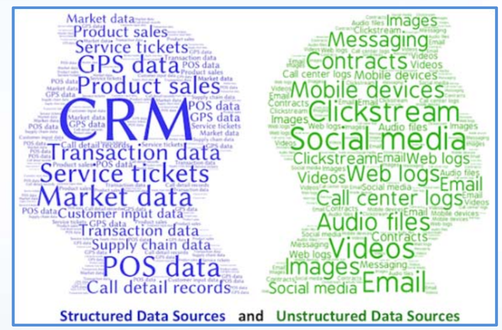
➢ Classification
- Decision tree, CART

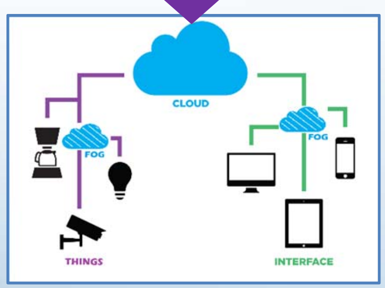
➢ Link mining
- Pagerank

➢ Statistical learning
- EM, SVM

CNN, RNN, LSTM, GCN







➢ Faster transmission rate

➢ Higher network capacity

➢ Stronger robustness

➢ Cloud computing
- Hadoop, Spark, Storm
- SDN, NFV, Network slicing

➢ Fog computing

➢ Machine Learning

Unknown target function f: x---> y

(ideal formula)

Training examples

D: $(x_1, y_1), \ldots, (x_N, y_N)$

(historical records)

Learning algorithm

A

Final hypothesis

$g \approx f$

(learned formula to be used)

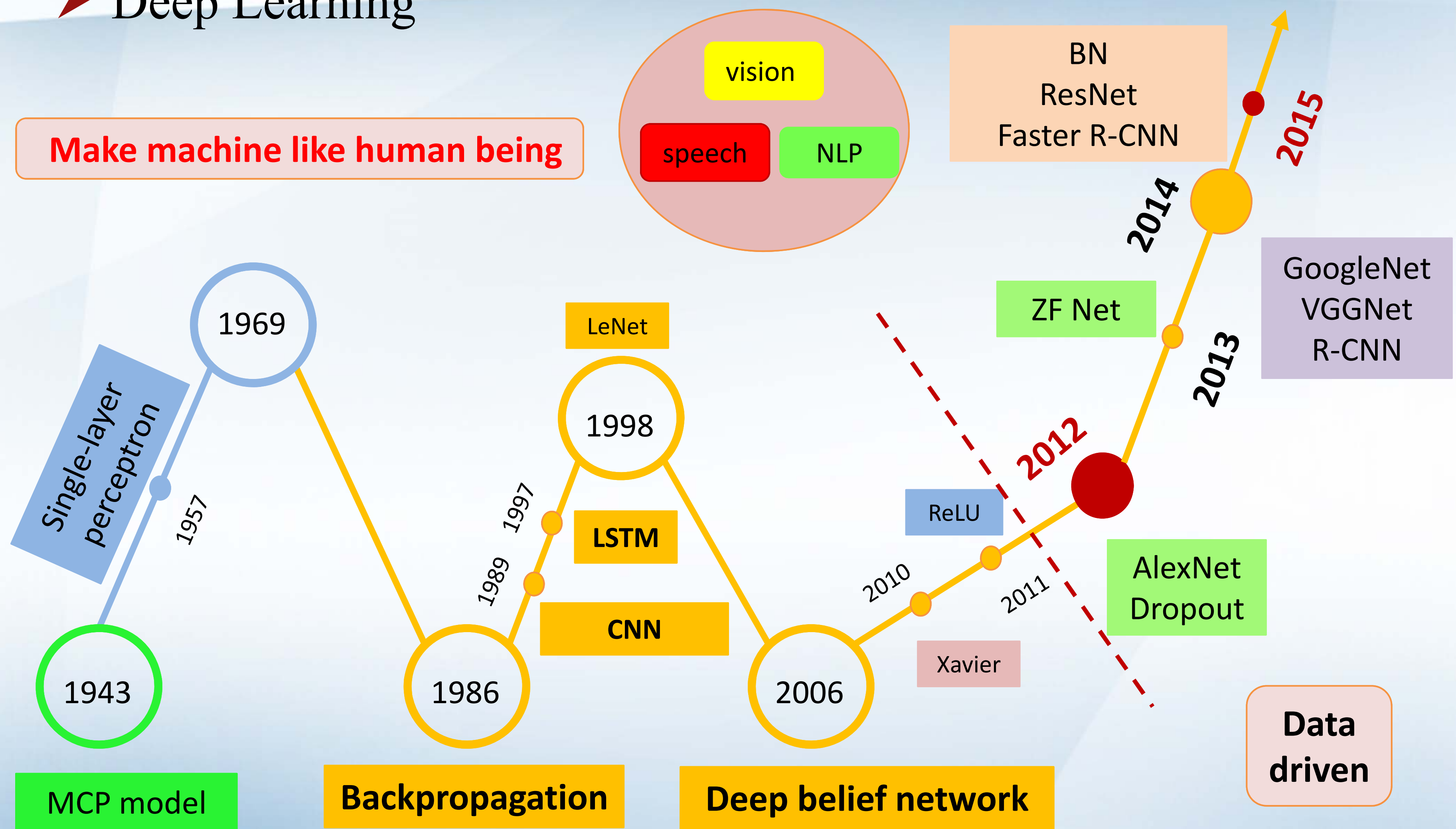
Hypothesis set

H

(set of candidate formula)

Introduction

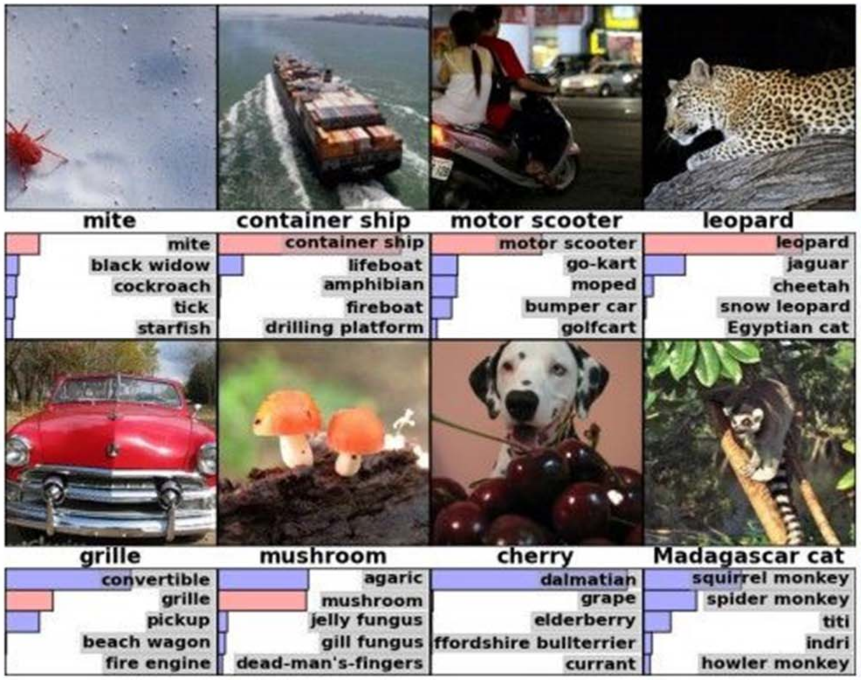
- Deep Learning



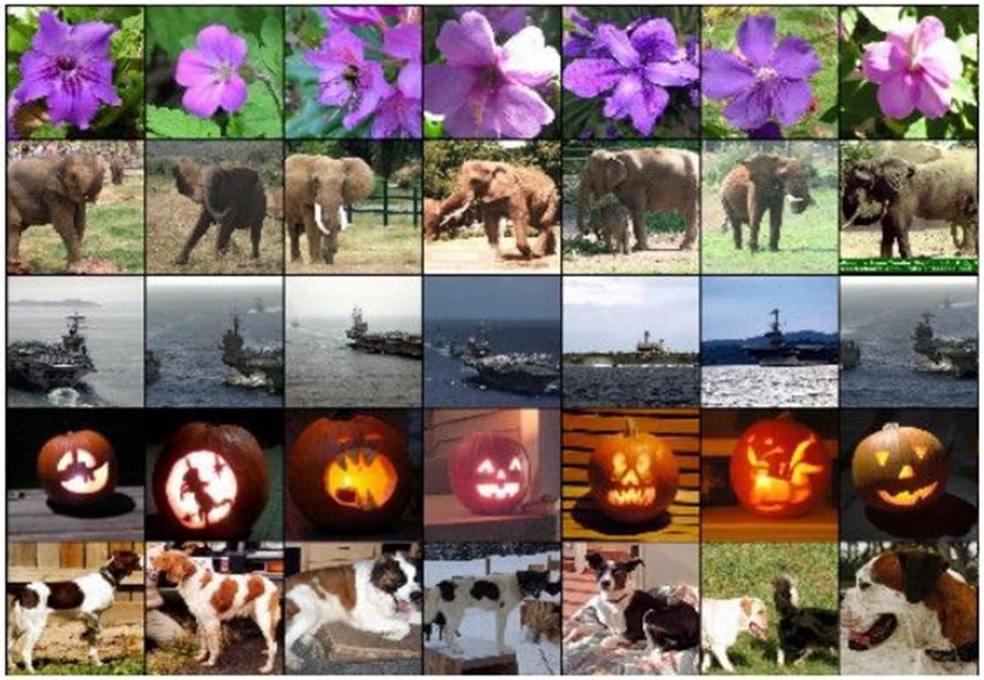
Introduction

- Applications

Classification



Retrieval



[Krizhevsky 2012]

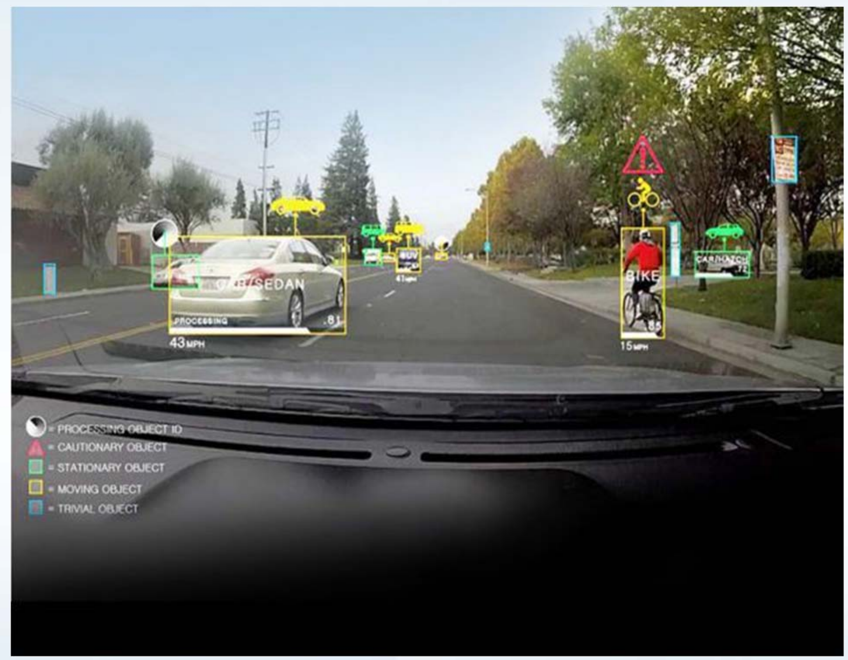
Introduction

- Applications

Image Captioning

[Vinyals et al., 2015]

Self-driving



Introduction

Machine Learning in wireless communication

Supervised learning

Regression model, KNN, SVM, Bayesian learning

- channel identification
- traffic prediction
- massive MIMO channel estimation/detection
- user location/behavior learning/classification

Unsupervised learning

Clustering algorithm, PCA, ICA

- MTC devices clustering
- small cell clustering
- device-to-device user clustering;
- HetNet clustering
- Signal dimension reduction

Reinforcement learning

MDP, POMDP, Q-learning, multi-armed bandit

- decision making under unknown network conditions
- energy modeling in energy harvesting
- HetNet selection/association

Contents

Case 1: CNN based Wireless Channel Identification

3V of Wireless Channel Data

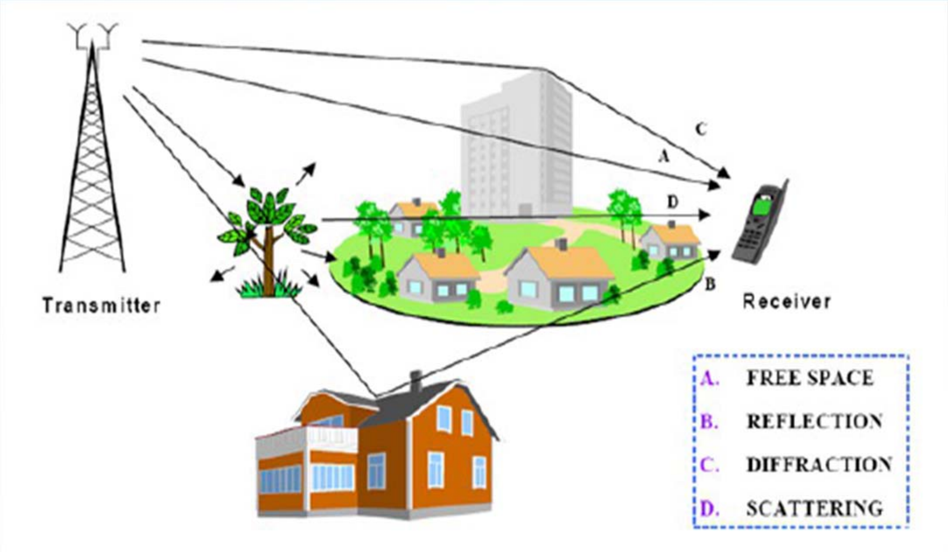
Big Data		Wireless Channel Data
• Volume	→	32 Gbyte (massive MIMO:32×56 antennas,100 MHz bandwidth)
• Variety	→	vary as frequencies, bandwidths and scenarios.
• Value	→	discovering new phenomenon, extracting new channel characteristics, supporting the accurate modeling of radio wave propagation etc.

Wireless channel data analysis with big data technology will support the study of all aspects of wireless channel

Case 1: CNN based Wireless Channel Identification

➢ Why we do channel identification?

- Reflection
- Diffraction
- Scattering



- **Solve the problems of multipath interference in process of wireless communication**
- **Decide which relevant wireless channel features should be used**

http://www.ee.iitm.ac.in/~giri/pdfs/EE5141/Jakes-Simulation.pdf

Case 1: CNN based Wireless Channel Identification

Wireless channel data description

- The dataset consists of **50000 samples**
- Each scenario contains **10000 samples**
- Each entry in one example represents **CIR(channel impulse response)**
- **Sampling 2 points** per half wavelength

LoS(line of sight)

Scenarios	Data format
InH(Indoor hotspot) A2	[2, 2, 19, 60]
UMi(urban micro) B1	[2, 2, 16, 60]
SMa(suburban macro) C1	[2, 2, 19, 60]
UMa(urban macro) C2	[2, 2, 16, 60]
RMa(Rural macro) D1	[2, 2, 14, 60]

NLoS(none line of sight)

Scenarios	Data format
InH(Indoor hotspot) A2	[2, 2, 23, 60]
UMi(urban micro) B1	[2, 2, 23, 60]
SMa(suburban macro) C1	[2, 2, 18, 60]
UMa(urban macro) C2	[2, 2, 24, 60]
RMa(Rural macro) D1	[2, 2, 14, 60]

Case 1: CNN based Wireless Channel Identification

Wireless channel data description

[2, 2, 19, 60]

- Number of transmitting antennas
- Number of receiving antennas
- Clusters of multi-path
- Number of samples

val(:,:,1,1) =

0.2855 + 0.1234i	-0.0894 - 0.2485i
-0.1368 + 0.3137i	0.2789 - 0.0944i

val(:,:,2,1) =

-0.1081 + 0.2683i	0.3160 - 0.0893i
-0.2489 - 0.0640i	0.1154 + 0.2746i

val(:,:,1,2) =

0.3295 + 0.0248i	-0.1754 - 0.2220i
-0.0301 + 0.3537i	0.2453 - 0.1843i

val(:,:,2,2) =

0.0063 + 0.2719i	0.2434 - 0.1956i
-0.2345 + 0.0382i	0.2011 + 0.1936i

Case 1: CNN based Wireless Channel Identification

➢ Wireless channel data description

Unify the input size

- Fixed size: 4560(76*60, LoS) , 5760(96*60, NLos)

Change the data type

- Use the modulus to replace the complex number

Create the label and split the data

one-hot coding, label mapping

InH(Indoor hotspot) A2	[1, 0, 0, 0, 0]
UMi(urban micro) B1	[0, 1, 0, 0, 0]
SMa(suburban macro) C1	[0, 0, 1, 0, 0]
UMa(urban macro) C2	[0, 0, 0, 1, 0]
RMa(Rural macro) D1	[0, 0, 0, 0, 1]

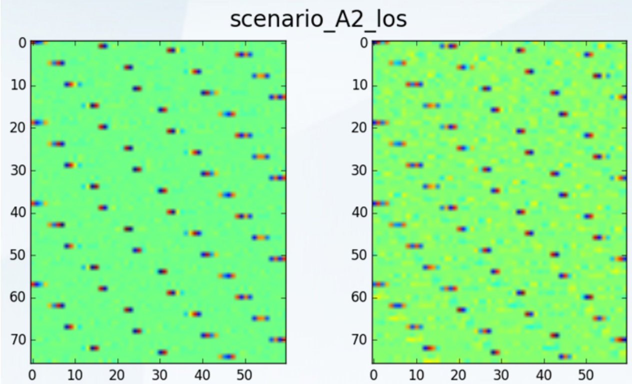
Total data : 50000

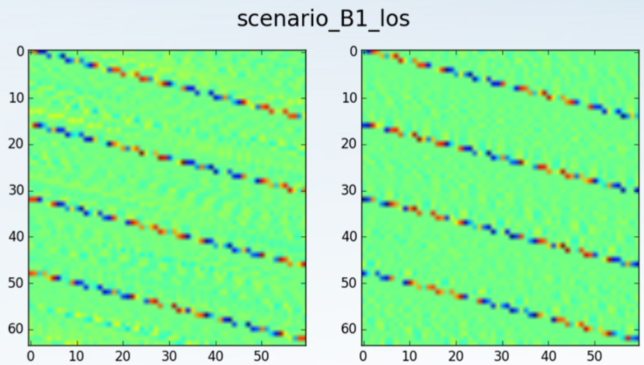
shuffle ↓ split

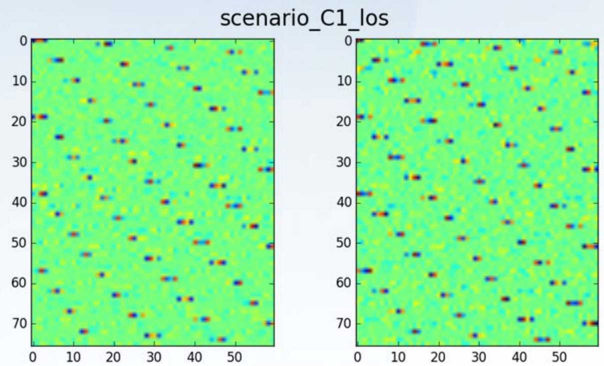
Training set:	80% of the total data
Test set:	20% of the total data

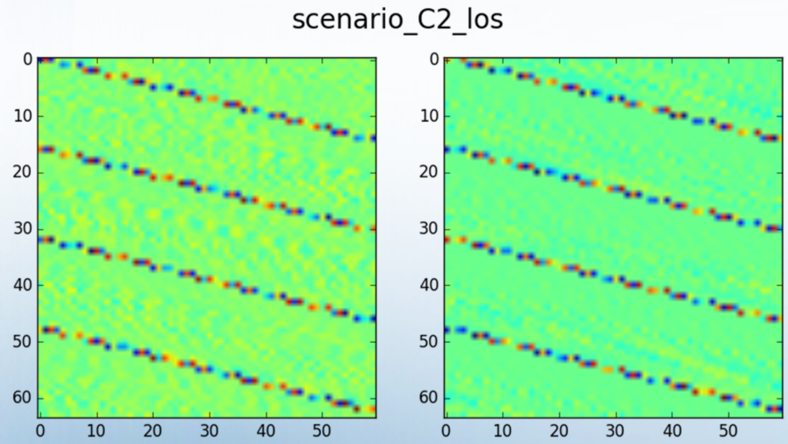
Case 1: CNN based Wireless Channel Identification

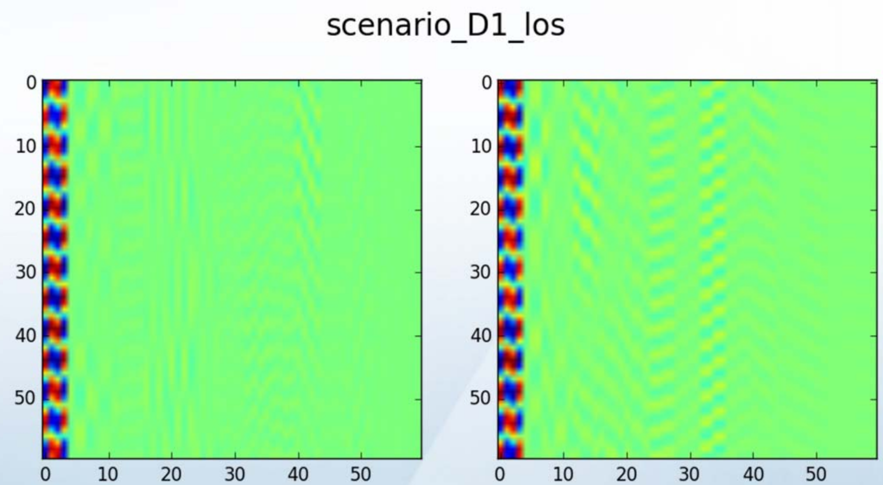
➢ Data visualization(LoS)





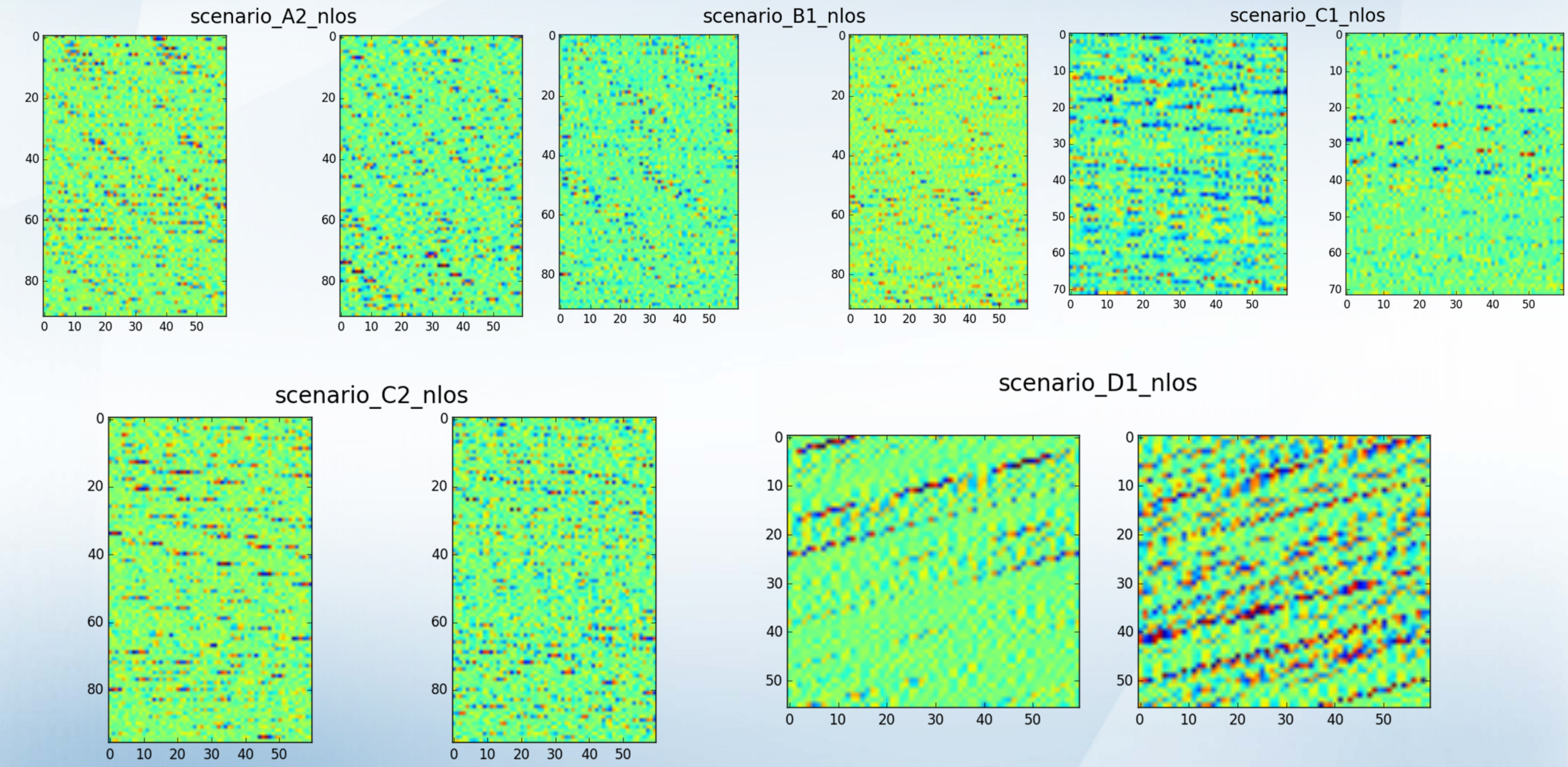






Case 1: CNN based Wireless Channel Identification

➢ Data visualization(NLoS)



Case 1: CNN based Wireless Channel Identification

Experimental Results

Input
Conv3-32
Maxpool
Conv3-64
Maxpool
Conv3-64
maxpool
Fc-1024
soft-max

Training configuration

- Batch size: 64
- Optimization algorithm: Adam
- Total iterations: 20000
- Dropout: training(0.5), testing(1)
- Batch normalization: used in Conv layer

Results

Training accuracy	**1**
Testing accuracy	**1**

Robustness of the model

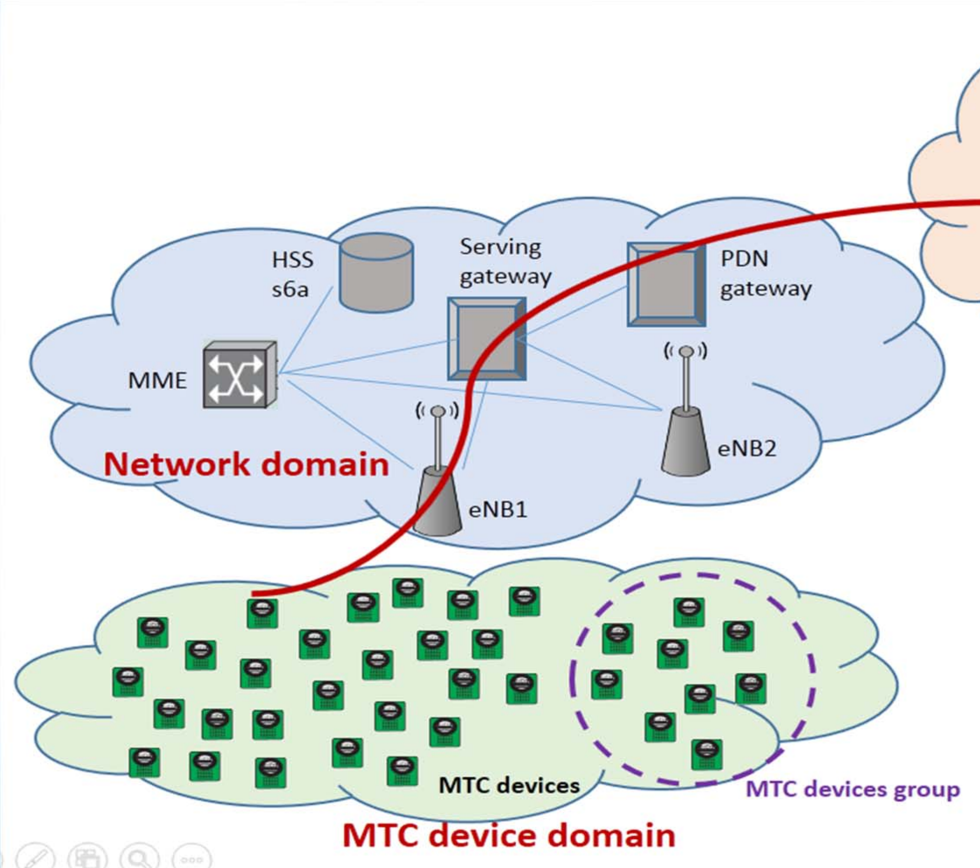
accuracy	**0.9856**

- Test other 2000 scenarios' samples(each scenario contains 400)

Contents

Case 2: Clustering based Transmission-efficient MTC

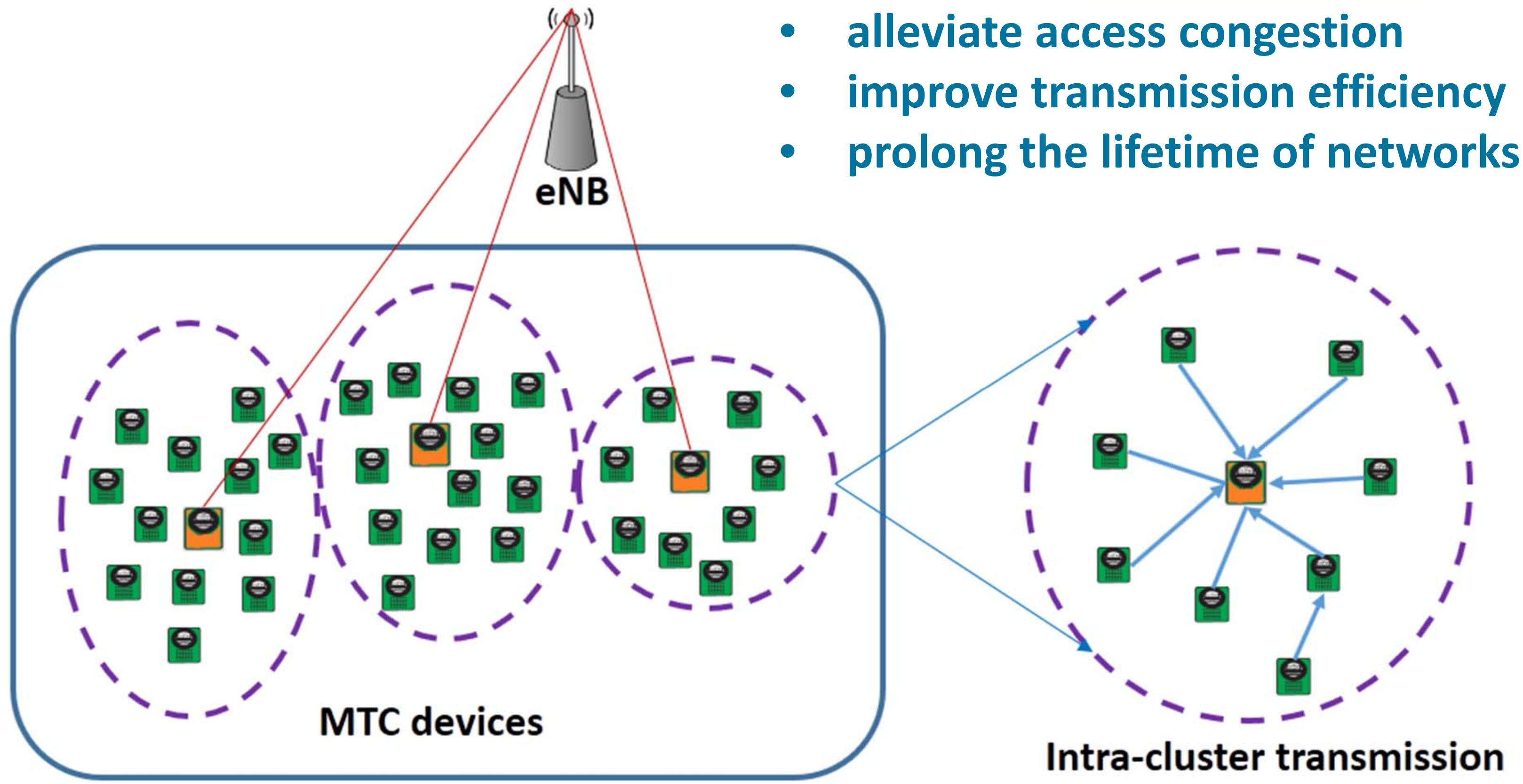
➢ Massive machine type communications (mMTC)



- **Large number of nodes, e.g. up to 300000 devices in a single cell.**
- Small packets / Low user data rates
- Uplink-dominated transmissions.
- Mixed traffic models with period and event-driven traffic.
- Low complexity and battery constrained MTC devices.

Case 2: Clustering based Transmission-efficient MTC

➢ Why Clustering

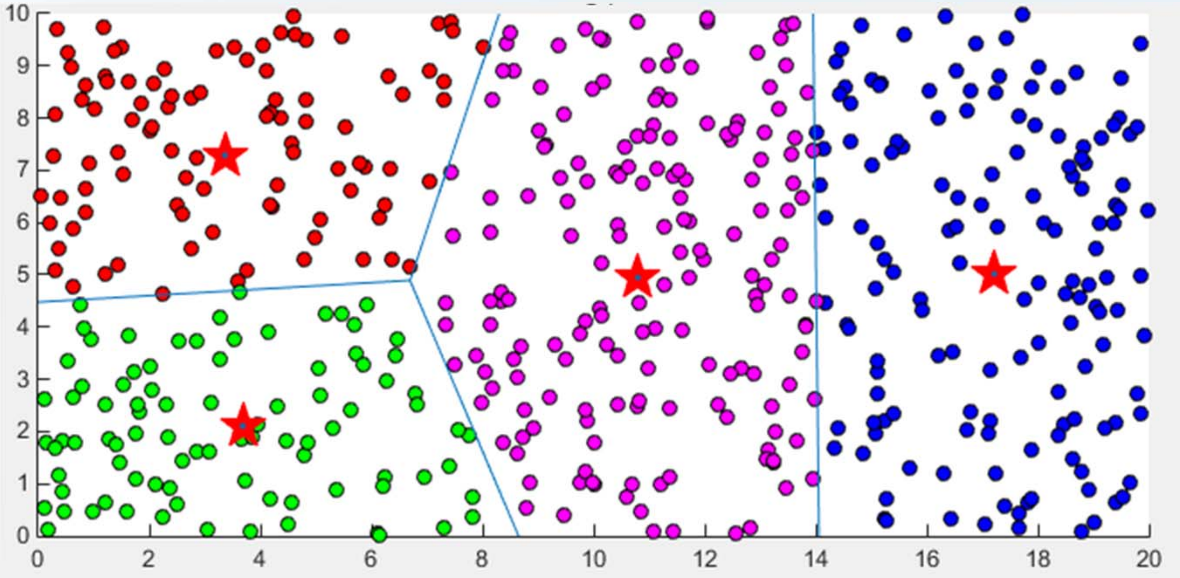


- alleviate access congestion
- improve transmission efficiency
- prolong the lifetime of networks

Case 2: Clustering based Transmission-efficient MTC

➢ Clustering of Machine Nodes

Machine nodes are grouped into clusters. Each cluster has a leader, namely cluster head (CH).



- Which machine nodes can be grouped as a cluster?
- Which machine node can be cluster head in each cluster?
- How to reduce the number of transmissions and the energy consumption?

Case 2: Clustering based Transmission-efficient MTC

➢ Traditional Methods

LEACH
and its variations
in WSN

MAC design
in cellular-based
M2M networks

- Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol.
- Random access channel (RACH) in LTE-A

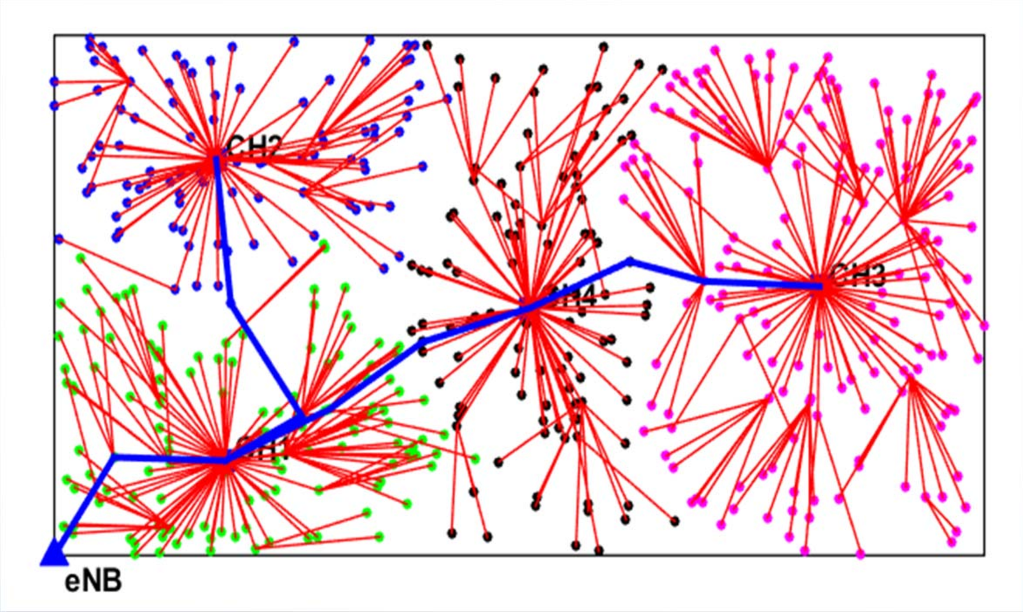
Cons:

- ➢ Using single hop in cluster, not applicable to large scale networks
- ➢ Might lead to large number of clusters
- ➢ Few discussion on optimal CH selection and optimal number of clusters
- ➢ Focus on Energy-efficiency and collision avoidance, seldom on transmission-efficiency

Case 2: Clustering based Transmission-efficient MTC

➢ System Model

mMTC can make use of LTE-M as its backbone network. The system consists of one eNB and *N* machine nodes for data collection.



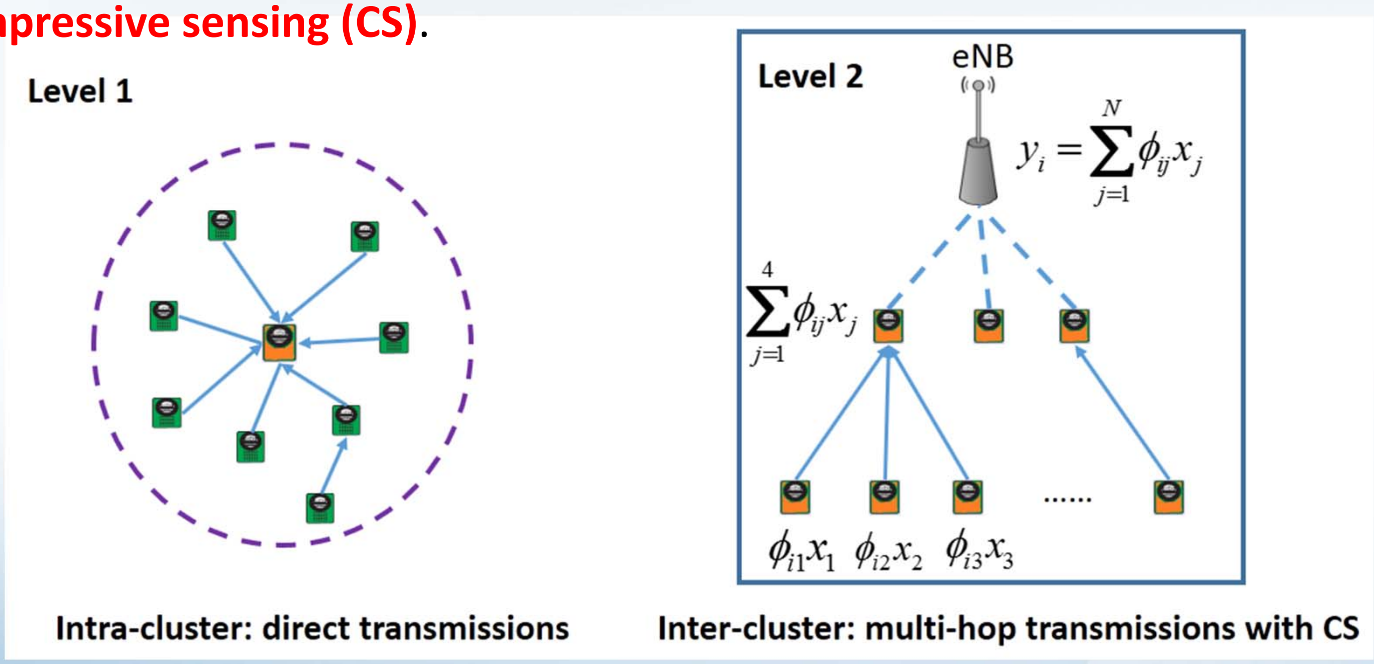
- Multi-hop networks
- Uniformly and independently distributed
- Fixed transmission power and rate.
- Nodes are grouped into clusters.

Fig. A multi-hop network in an regular field, where the machine nodes form 4 clusters.

Case 2: Clustering based Transmission-efficient MTC

➢ Two Levels of Transmission

- **Intra-cluster transmission:**
  Within a cluster, nodes transmit data to the CH directly to save energy.
- **Inter-cluster transmission:**
  A backbone tree spanning all CHs is constructed to transmit data to the eNB using **compressive sensing (CS)**.



Case 2: Clustering based Transmission-efficient MTC

➢ Compressive Sensing for Big Data Transmission

Compressive sensing (CS) can reduce the volume of data transmissions.

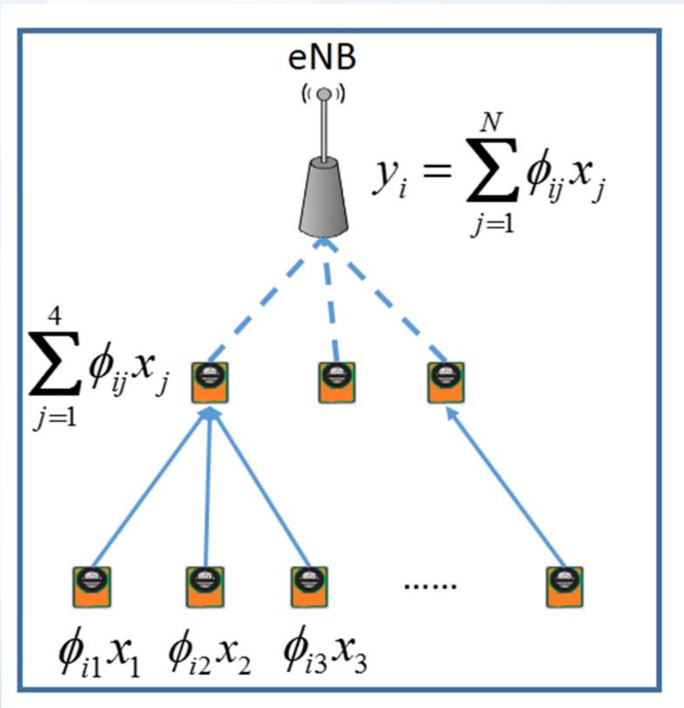


Fig. Data collection with CS in Inter-cluster transmission.

*Assume **x** is k-sparse in the Ψ domain.*

$$\boldsymbol{x} = \psi \boldsymbol{s}$$

Then:

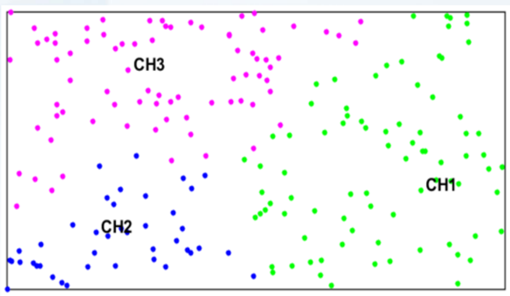
$$\boldsymbol{y} = \Phi \boldsymbol{x}$$

Small number of projections of $\boldsymbol{x}$ to the eNB	Measurement matrix	Original data collected from nodes

Case 2: Clustering based Transmission-efficient MTC

➢ Clustering based Transmission-efficient Scheme

Given a Network
G=<V,E>



Select initial cluster heads and divide the machine nodes into C clusters.

Elect new cluster head

Construct a backbone tree that links all CHs to the eNB

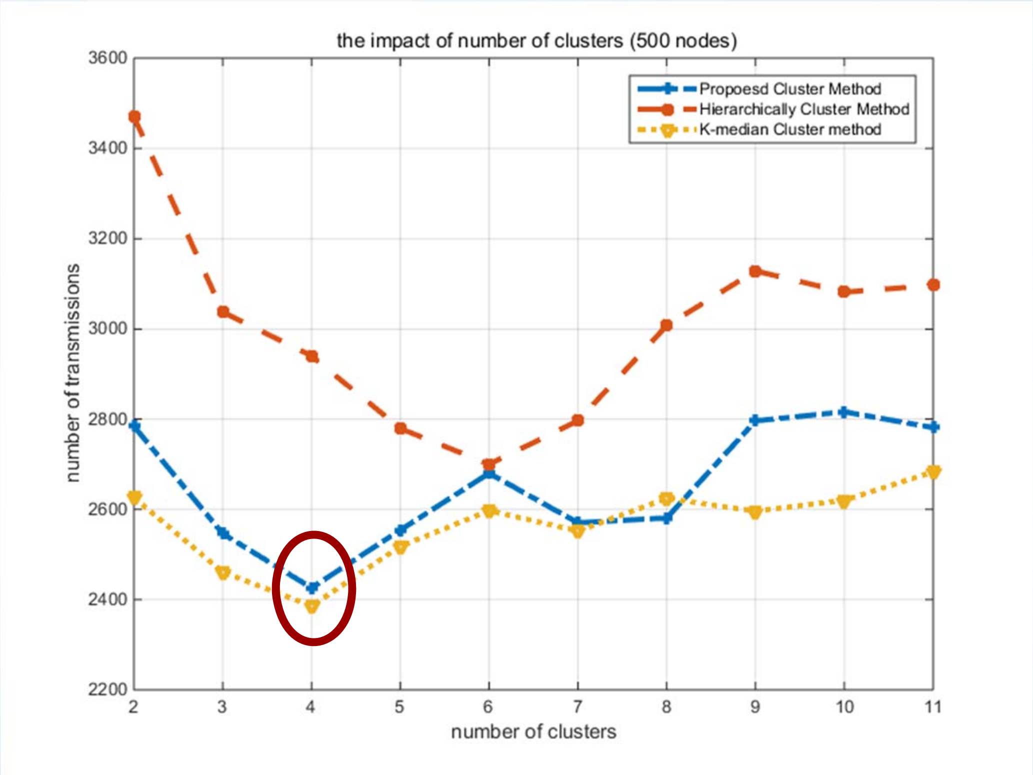
Clustering

Transmit data from CHs to eNB

Compressive sensing

Case 2: Clustering based Transmission-efficient MTC

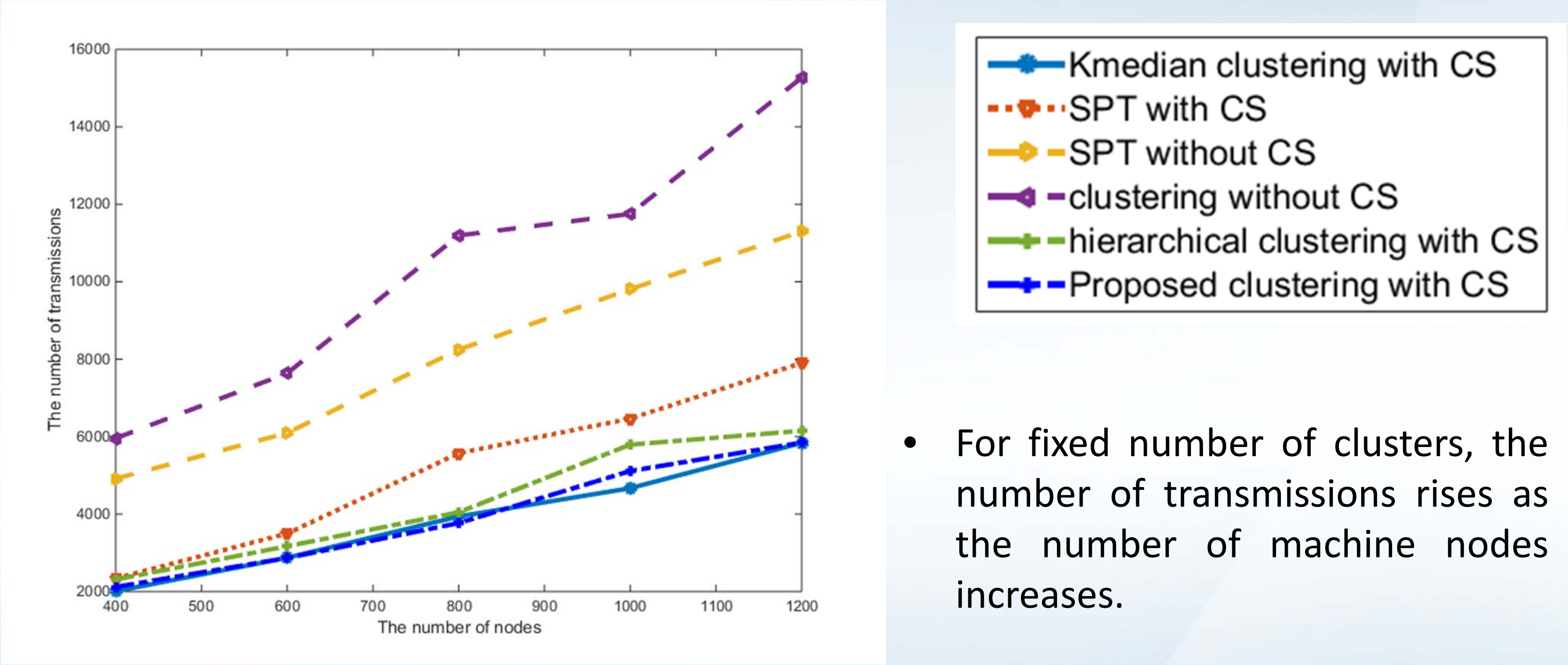
➢ Experimental Results (1/4)



- For a clustering method, there is always an optimal number of clusters that minimizes the data transmissions.

Case 2: Clustering based Transmission-efficient MTC

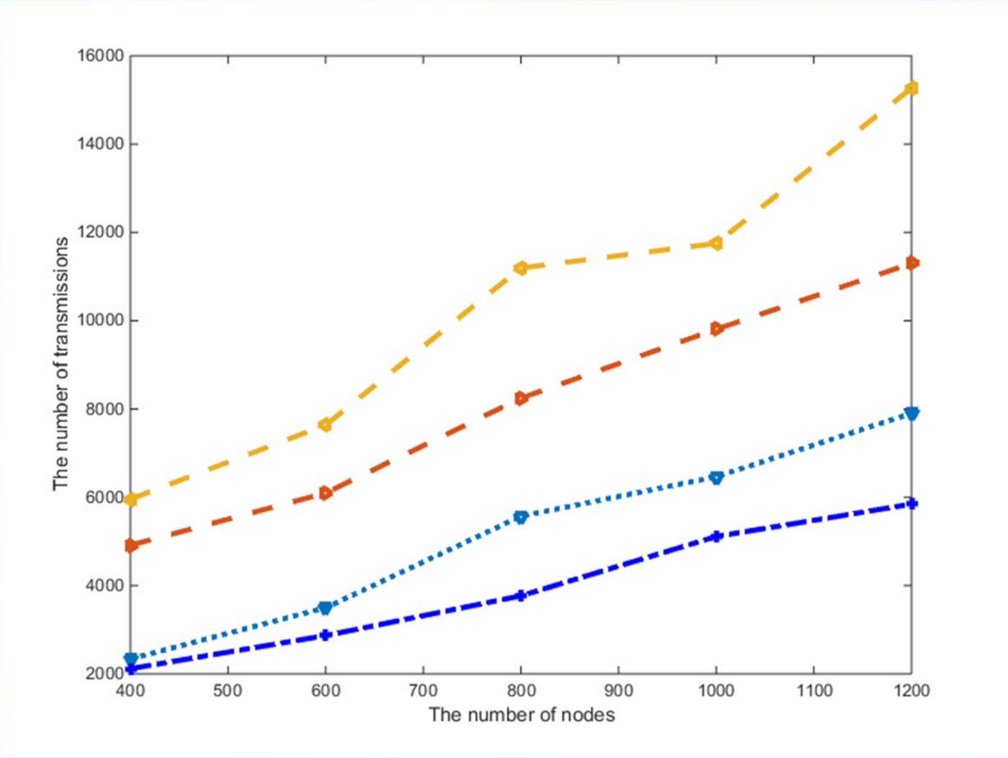
➢ Experimental Results (2/4)

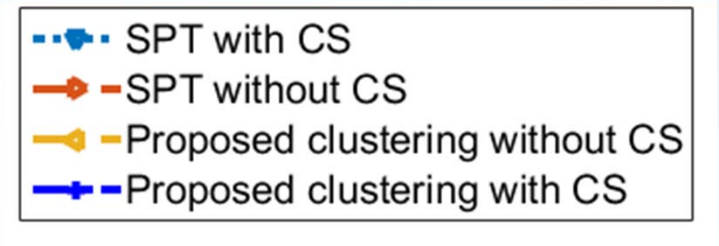


- For fixed number of clusters, the number of transmissions rises as the number of machine nodes increases.

Case 2: Clustering based Transmission-efficient MTC

➢ Experimental Results (3/4)

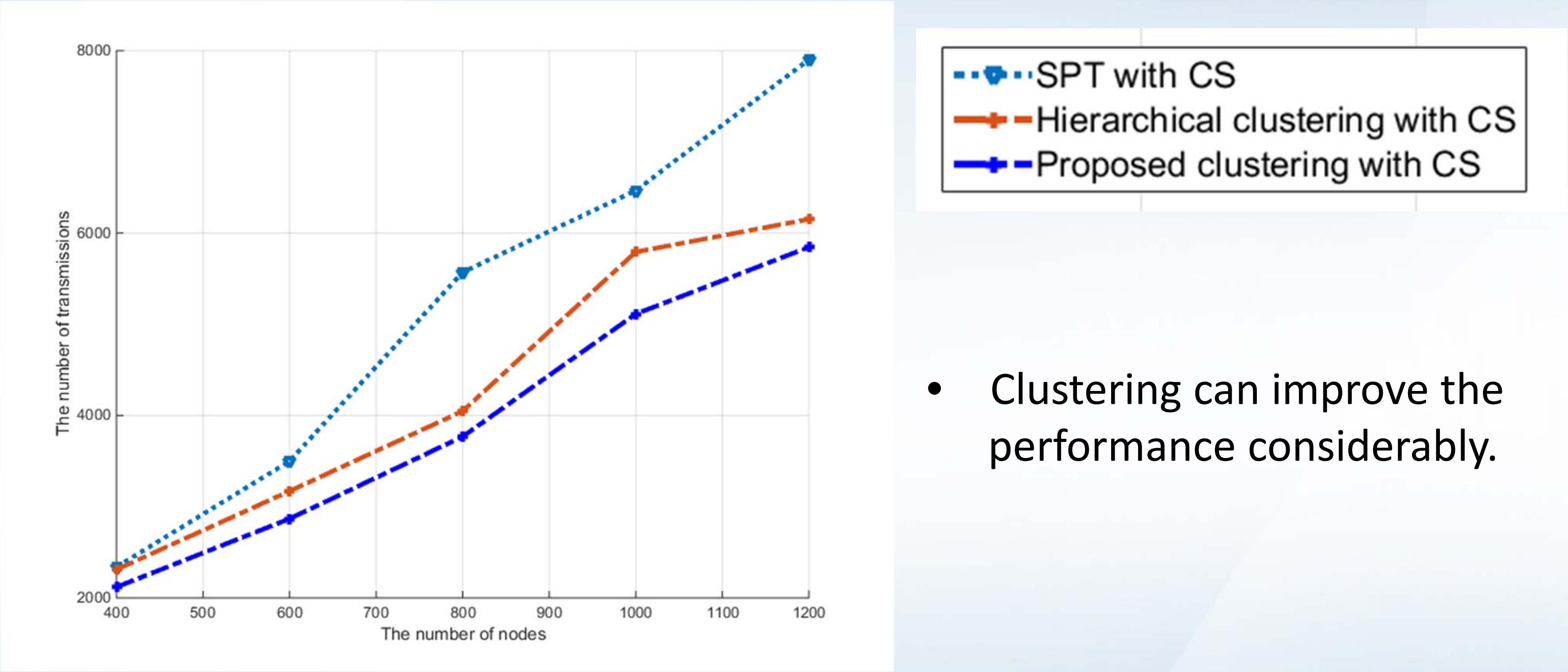




- Compressive sensing can reduce the number of transmission dramatically.

Case 2: Clustering based Transmission-efficient MTC

➢ Experimental Results (4/4)



- Clustering can improve the performance considerably.

Contents

Case 3: DenseNet for Wireless Traffic Prediction

- Why wireless traffic prediction?
  - It plays a very important role in the future communication networks!

Improve network management:
Dynamic network congestion control

Reduce OPEX:
Accurate radio resource purchase

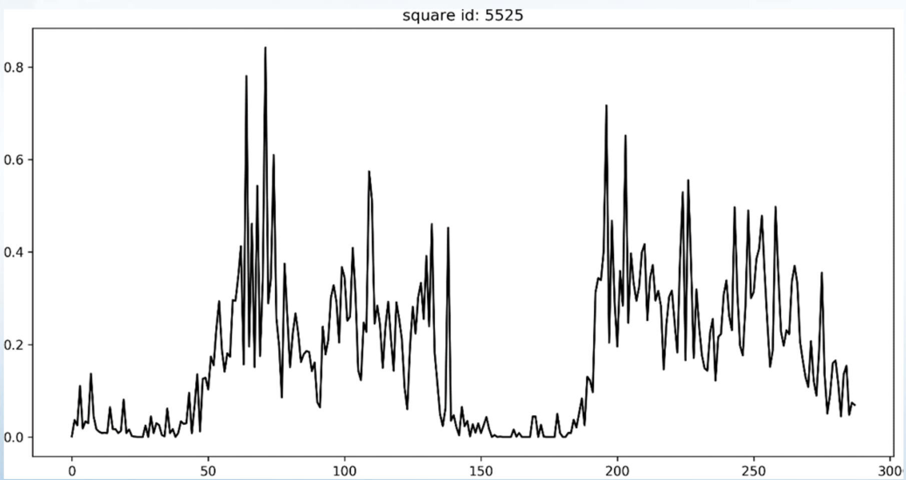
Enhance energy efficiency:
Intelligent BS ON-OFF

Case 3: DenseNet for Wireless Traffic Prediction

Traditional Methods

- Traffic prediction is essentially a time series analysis problem.
- ARIMA model (Autoregressive Integrated Moving Average)
  - The evolving variable of interest is regressed on its own lagged values.
  - The regression error is a linear combination of error terms.

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_{p'} X_{t-p'} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

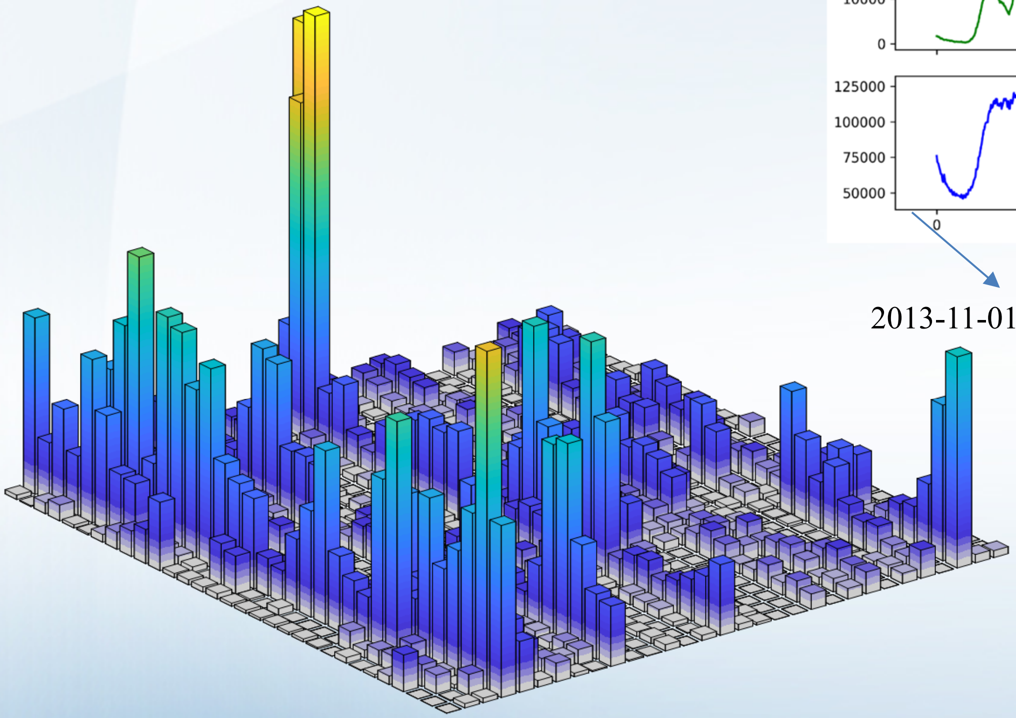


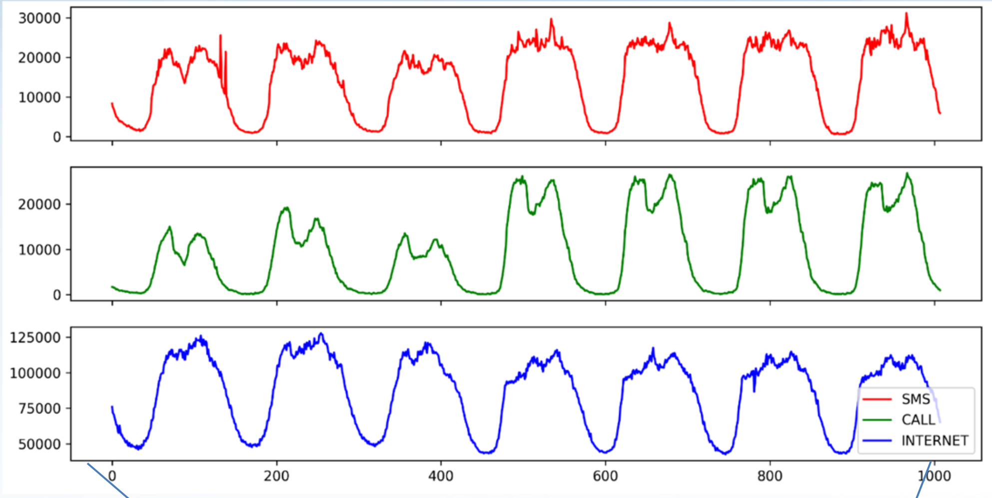
- ✕ Can only capture the linearity of the data
- ✕ The spatial information of other cells is not considered
- ✕ Works only for stationary data set

Case 3: DenseNet for Wireless Traffic Prediction

- Real traffic distribution

Spatially uneven





2013-11-01 00:00

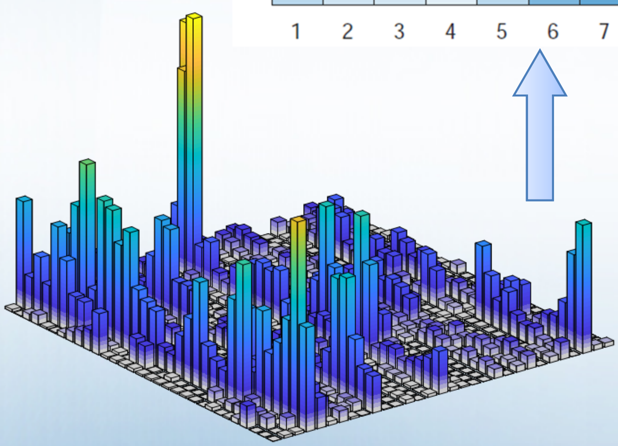
2013-11-08 00:00

Temporally periodicity

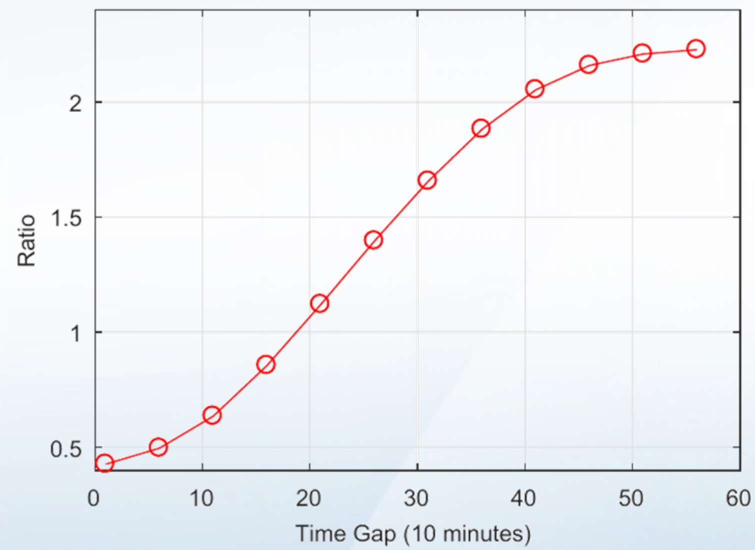
Case 3: DenseNet for Wireless Traffic Prediction

- Real traffic distribution

Spatial correlation

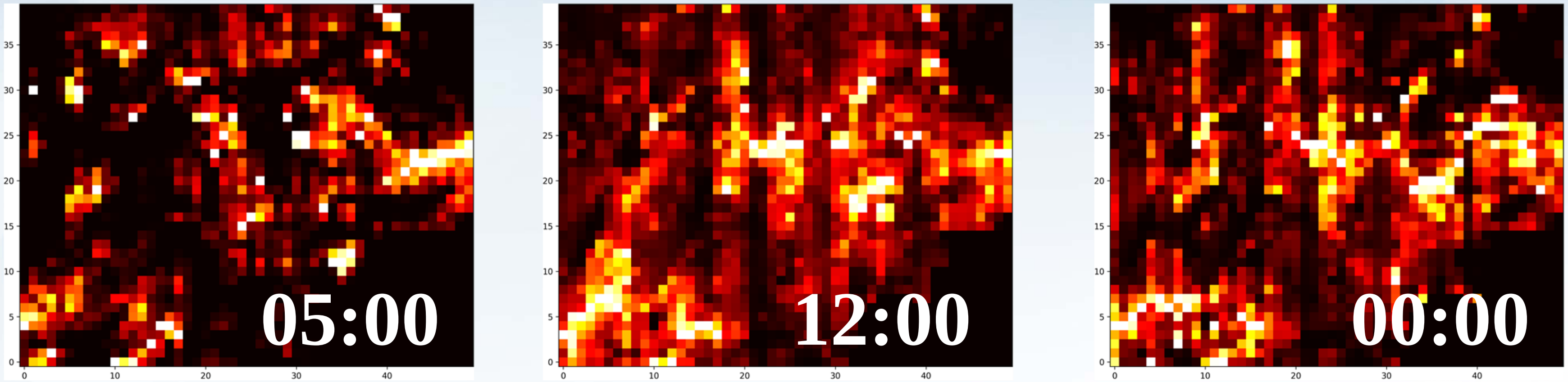


Temporal correlation

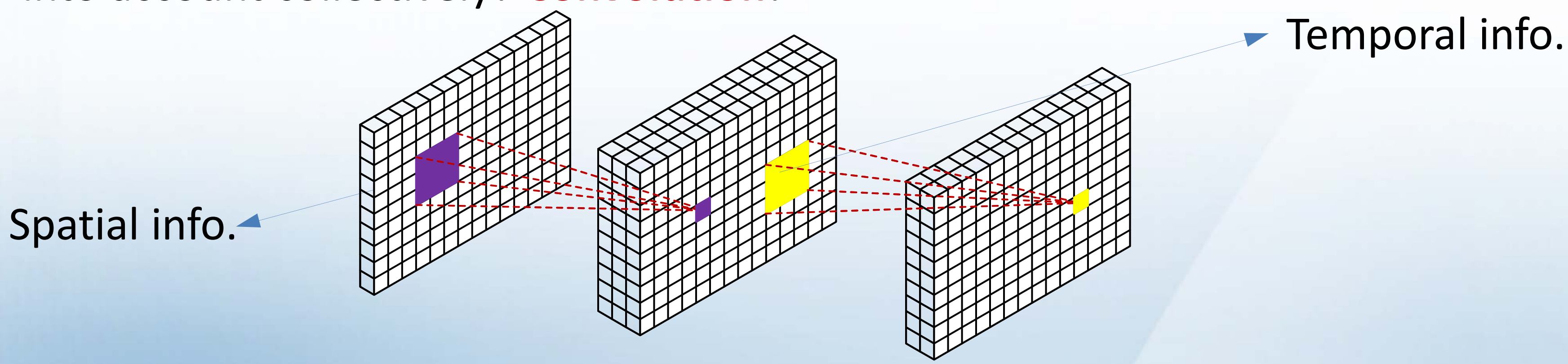


Case 3: DenseNet for Wireless Traffic Prediction

- Traffic data is image-like

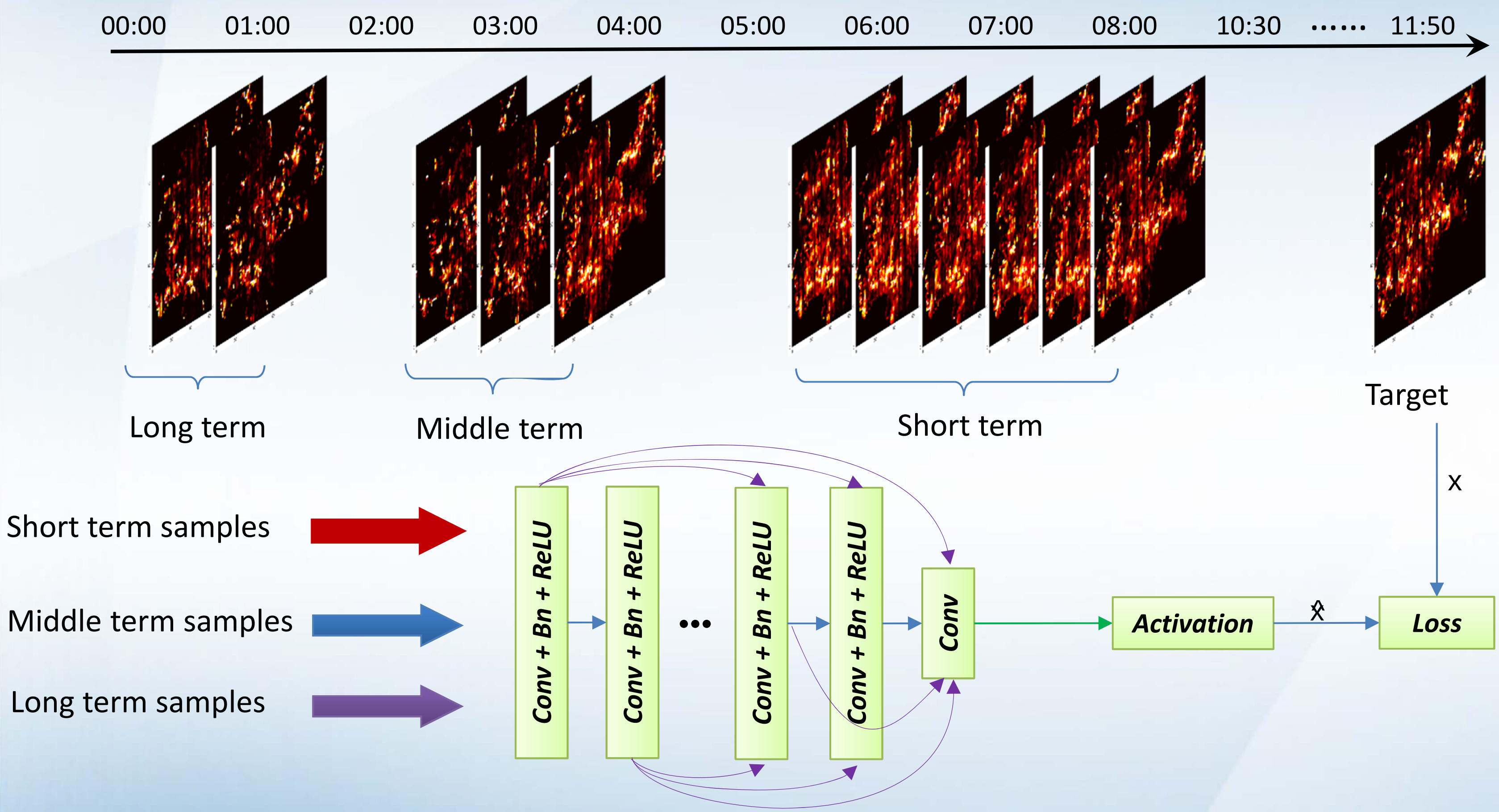


How to perform traffic prediction taking the spatial and temporal information into account collectively? **Convolution**!



Case 3: DenseNet for Wireless Traffic Prediction

➢ Network Architecture ST-DenseNet



Case 3: DenseNet for Wireless Traffic Prediction

➢ Experimental Results

Evaluation metric: $$RMSE = \sqrt{\frac{1}{z}\sum_i (x_i - \hat{x}_i)^2}$$

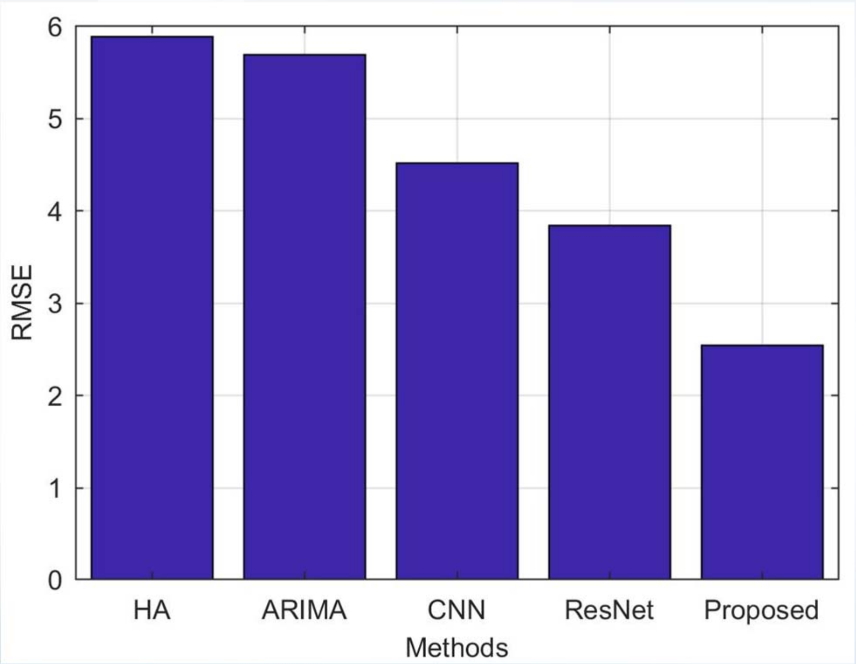
Parameter settings

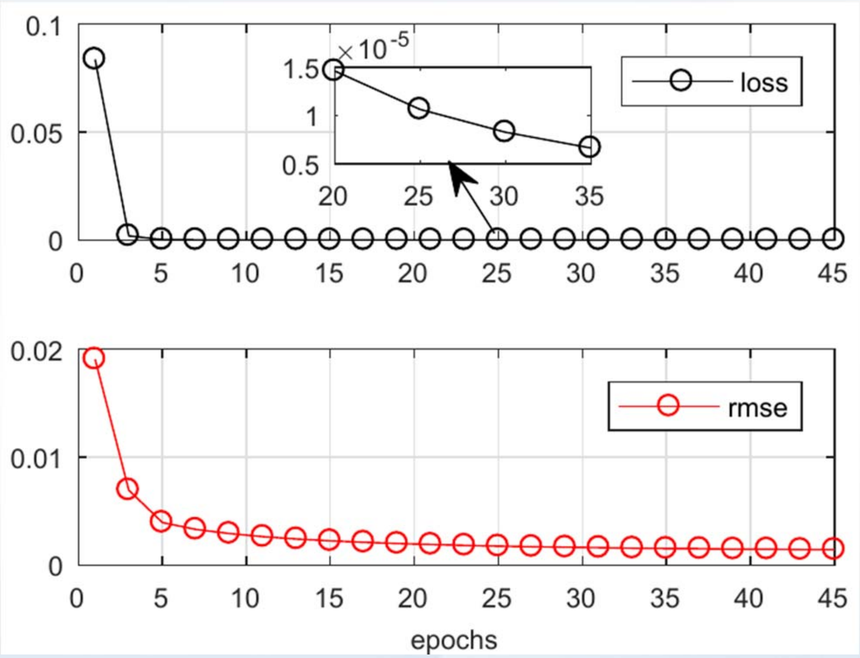
Parameter	**Value**	**Note**
Epochs	100	How many times we trained on the data
Learning rate	0.005	Learning rate of optimizer
Dropout	0.2	Drop a specific percent of connections between two layers
Filters	4	Number of feature maps
Growth rate	4	How many feature maps we add after one layer
Closeness size	3	Short term dependent
Period size	3	Middle term dependent
Trend size	1	Long term dependent

Case 3: DenseNet for Wireless Traffic Prediction

➢ Experimental Results

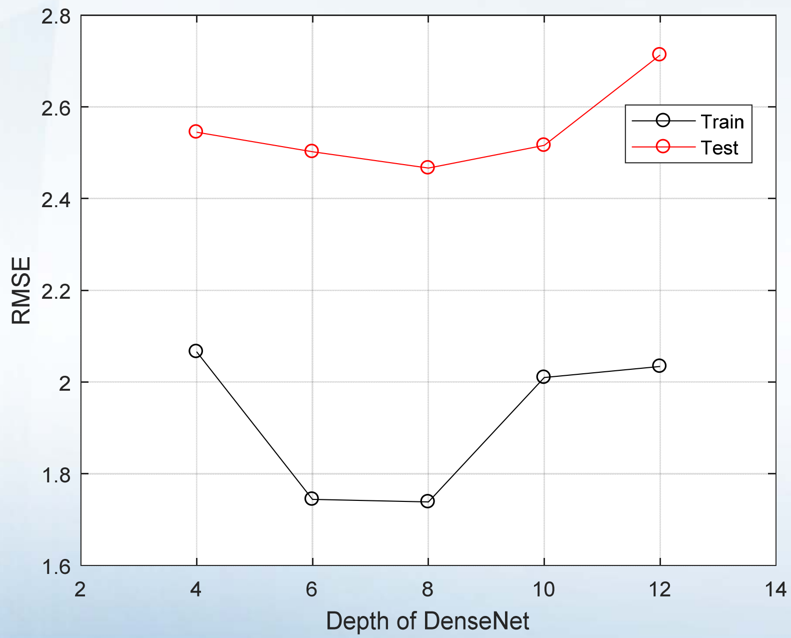
- To prevent over fitting, two measures are adopted in our scheme: early stopping and regularization

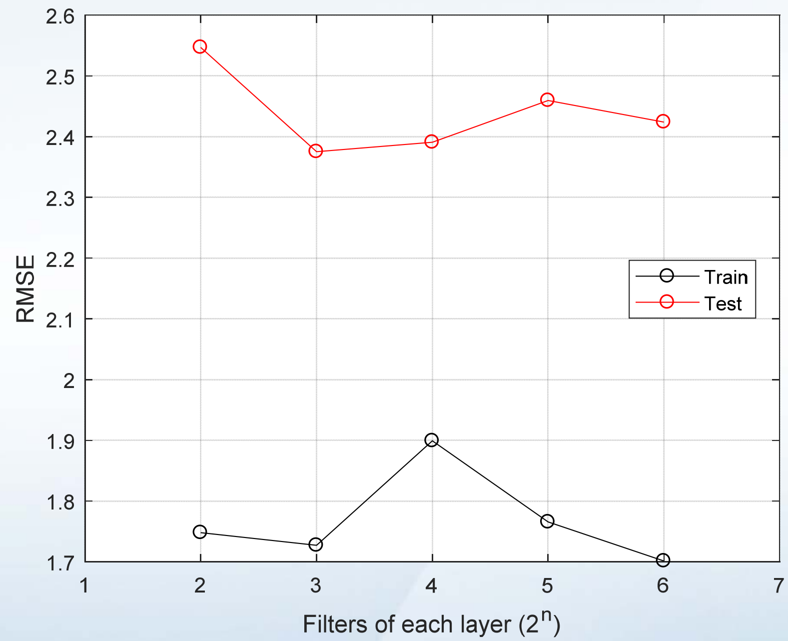




Case 3: DenseNet for Wireless Traffic Prediction

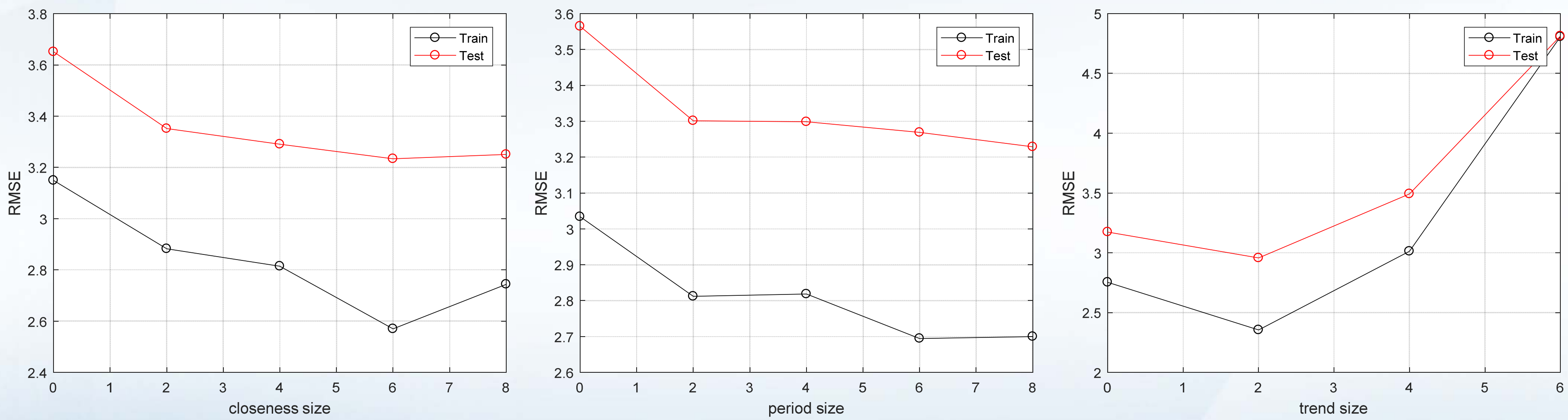
- Experimental Results: hyper-parameters (number of layers and number feature maps)
  - To a certain degree, the deeper the model, the better the performance
  - But too complex models will overfit the data and increase RMSE





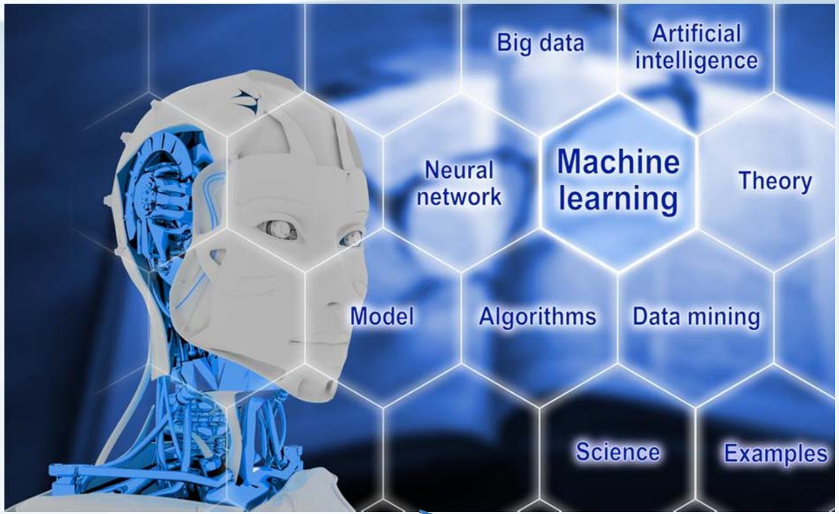
Case 3: DenseNet for Wireless Traffic Prediction

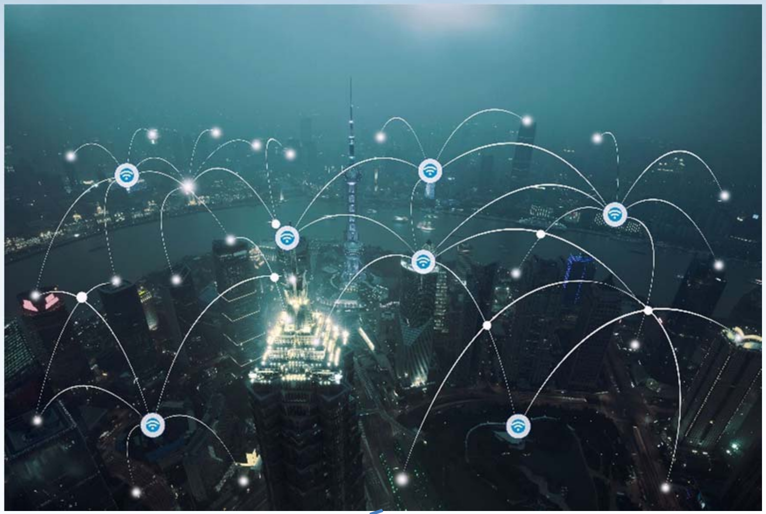
- Experimental Results: hyper-parameters (short, middle and long term dependency)
  - All three factors are beneficial to improve the prediction
  - They have different impacts on the performance



Contents

Conclusion





Smart & Intelligent

5G and Beyond

Thank you!