References Record

---

time span: 20200928-20201015

Survey

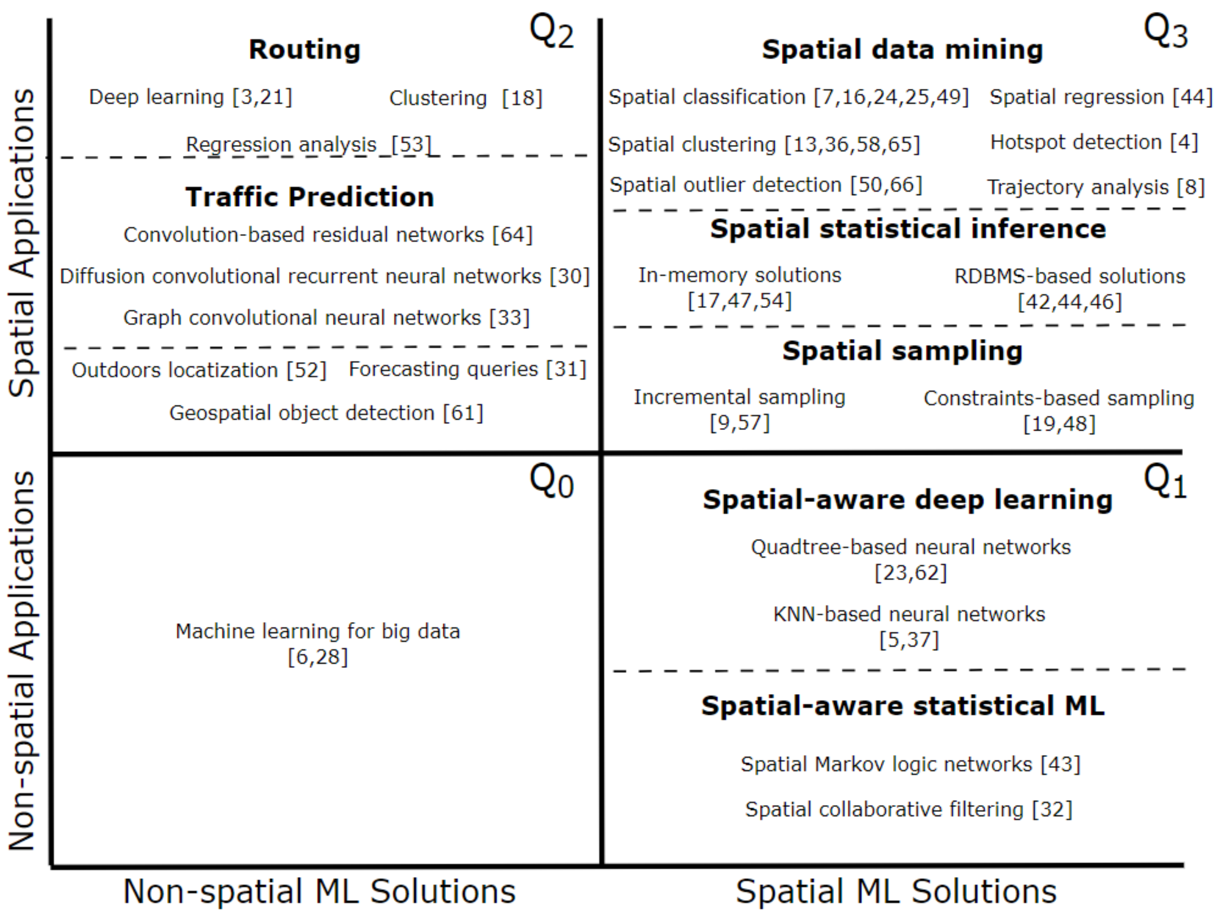
Machine Learning Meets Big Spatial Data

This article illustrates the structure of the whole spatial data region.

> *Abstract*—The proliferation in amounts of generated data has propelled the rise of scalable machine learning solutions to efficiently analyze and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, e.g., GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, drug addiction, among others. Consequently, major research efforts are exerted to support efficient analysis and intelligence inside these applications by either providing spatial extensions to existing machine learning solutions or building new solutions from scratch. In this 90-minutes tutorial, we comprehensively review the state-of-the-art work in the intersection of machine learning and big spatial data. We cover existing research efforts and challenges in three major areas of machine learning, namely, data analysis, deep learning and statistical inference. We also discuss the existing end-to-end systems, and highlight open problems and challenges for future research in this area.

The highlighted sentence gives a wider application than I've considered.

This article can be considered as a guide way application of spatial data or spatial itself usage. Spatial data is widely used in traffic field which expands to path finding, route recommendation and many other branches. So this article matches mainly.



> Fig. 1. Landscape of Machine Learning for Big Spatial Data.

The figure above illustrates the crossover zone of machine learning and spatial data. And it uses landscape to outline the area.

The following job is to explore each field with the figure as a index and exert to find out a promising branch for study propellant.

> learning solution, whether the application is spatial or not. We mainly focus on the three quarters Q_1, Q_2, and Q_3 in Figure 1 because they cover the spatial dimension in the machine learning solutions and/or the big data applications. We skip the quarter Q_0 as it is already covered by previous SIGMOD tutorials about the techniques and challenges in machine learning for big data in general [6], [28]. Also note

The focus of the figure is 1,2 and 3 while 0 attaches to other fields closely.

Q2

Routing

Deep learning

RoadTracer: Automatic Extraction of Road Networks From Aerial Images

> **Abstract**
>
> *Mapping road networks is currently both expensive and labor-intensive. High-resolution aerial imagery provides a promising avenue to automatically infer a road network. Prior work uses convolutional neural networks (CNNs) to detect which pixels belong to a road (segmentation), and then uses complex post-processing heuristics to infer graph connectivity. We show that these segmentation methods have high error rates because noisy CNN outputs are difficult to correct. We propose RoadTracer, a new method to automatically construct accurate road network maps from aerial images. RoadTracer uses an iterative search process guided by a CNN-based decision function to derive the road network graph directly from the output of the CNN. We compare our approach with a segmentation method on fifteen cities, and find that at a 5% error rate, RoadTracer correctly captures 45% more junctions across these cities.*

It uses the CNN to segment a aerial picture for constructing the road. With improvement, this work traces to locate the roads without noisy which decreases the error rate to 5%.

route correction

Stochastic Weight Completion for Road Networks Using Graph Convolutional Networks

> **Abstract:**
> Innovations in transportation, such as mobility-on-demand services and autonomous driving, call for high-resolution routing that relies on an accurate representation of travel time throughout the underlying road network. Specifically, the travel time of a road-network edge is modeled as a time-varying distribution that captures the variability of traffic over time and the fact that different drivers may traverse the same edge at the same time at different speeds. Such stochastic weights may be extracted from data sources such as GPS and loop detector data. However, even very large data sources are incapable of covering all edges of a road network at all times. Yet, high-resolution routing needs stochastic weights for all edges. We solve the problem of filling in the missing weights. To achieve that, we provide techniques capable of estimating stochastic edge weights for all edges from traffic data that covers only a fraction of all edges. We propose a generic learning framework called Graph Convolutional Weight Completion (GCWC) that exploits the topology of a road network graph and the correlations of weights among adjacent edges to estimate stochastic weights for all edges. Next, we incorporate contextual information into GCWC to further improve accuracy. Empirical studies using loop detector data from a highway toll gate network and GPS data from a large city offer insight into the design properties of GCWC and its effectiveness.

Creating a new data type exerts to make a higher-resolution road map which can be resolved for GPS and others' location.

route preciser

Clustering

Learning to Route with Sparse Trajectory Sets

> by any trajectories. For example, trajectory T_1 visited A and then J, X, Y, and B_3 before reaching B. If routing from A to B is requested, the path $A \rightarrow J \rightarrow X \rightarrow Y \rightarrow B_3 \rightarrow B$, as captured by trajectory T_1, can be recommended directly. The challenge is to enable routing for (s, d) pairs that are not connected by trajectories, e.g., (A_1, B_2) and (H, F).

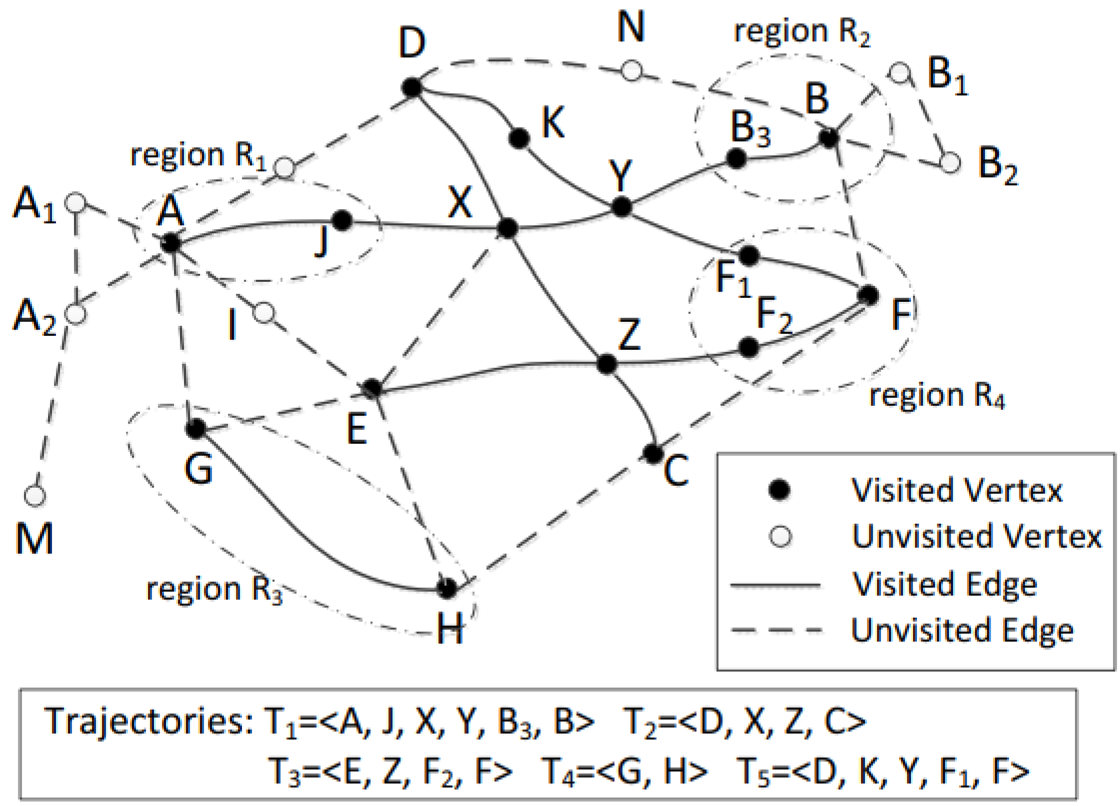


Fig. 1: Motivating Example

maintain sets of trajectories to cluster out a new path finding way.

regression analysis

W-edge: weighing the edges of the road network

> **ABSTRACT**
>
> Understanding link travel times (LTT) has received significant attention in transportation and spatial computing literature but they often remain behind closed doors, primarily because the data used for capturing them is considered confidential. Consequently, free and open maps such as OpenStreetMap (OSM) or TIGER, while being remarkably accurate in capturing geometry and topology of the road network are oblivious to actual travel times. Without LTTs computing the optimal routes or estimated time of arrival is challenging and prone to substantial errors. In this work we set to enrich the underlying map information with LTT by using a most basic data about urban trajectories, which also becomes increasingly available for public use: set of origin/destination location/timestamp pairs. Our system, `W-edge` utilizes such basic trip information to calculate LTT to each individual road segment, effectively assigning a weight to individual edges of the underlying road network. We demonstrate that using appropriately trained edge weights, the errors in estimating travel times are up to 60% lower than the errors observed in OSRM or GraphHopper, two prominent OSM-based, traffic-oblivious, routing engines.

This work uses time consumption on trajectories in LTT to estimate a closer time to the real world.

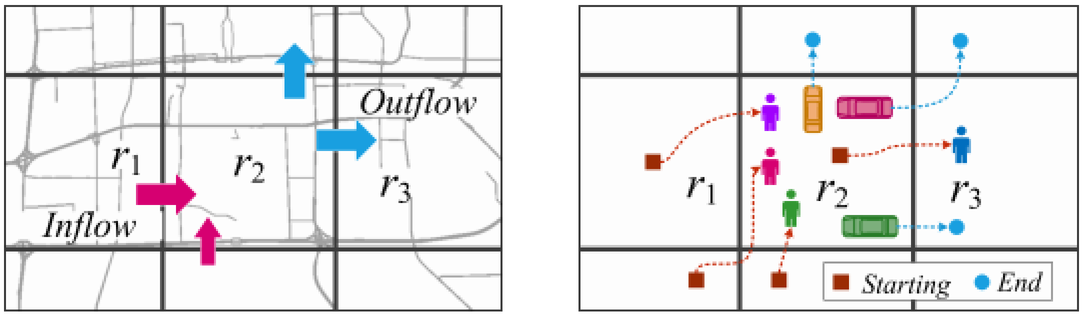
Traffic Prediction

Convolution-based residual networks

Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction

> fic, events, and weather. We propose a deep-learning-based approach, called ST-ResNet, to *collectively* forecast the inflow and outflow of crowds in each and every region of a city. We design an end-to-end structure of ST-ResNet based

Trying to follow the crowd and predict how the trend flows, this article uses convolution networks to separate the complex problem.



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

Figure 1: Crowd flows in a region

So the network implies topology structure.

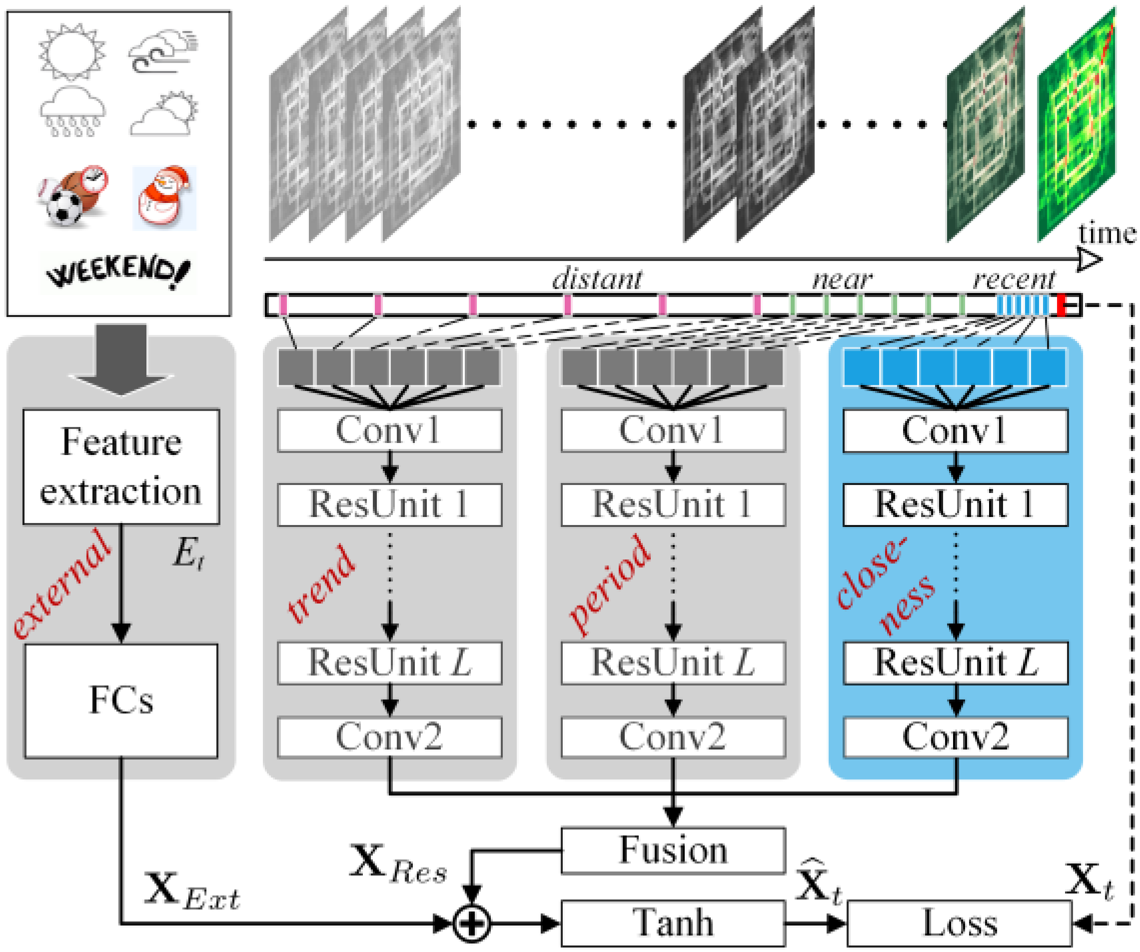


Figure 3: ST-ResNet architecture. Conv: Convolution; ResUnit: Residual Unit; FC: Fully-connected.

Diffusion convolutional recurrent neural networks

DIFFUSION CONVOLUTIONAL RECURRENT NEURAL NETWORK: DATA-DRIVEN TRAFFIC FORECASTING

> ABSTRACT
>
> Spatiotemporal forecasting has various applications in neuroscience, climate and transportation domain. Traffic forecasting is one canonical example of such learning task. The task is challenging due to (1) complex spatial dependency on road networks, (2) non-linear temporal dynamics with changing road conditions and (3) inherent difficulty of long-term forecasting. To address these challenges, we propose to model the traffic flow as a diffusion process on a directed graph and introduce *Diffusion Convolutional Recurrent Neural Network* (DCRNN), a deep learning framework for traffic forecasting that incorporates both spatial and temporal dependency in the traffic flow. Specifically, DCRNN captures the spatial dependency using bidirectional random walks on the graph, and the temporal dependency using the encoder-decoder architecture with scheduled sampling. We evaluate the framework on two real-world large scale road network traffic datasets and observe consistent improvement of 12% - 15% over state-of-the-art baselines.

traffic forecasting is a seemingly promising aspect.

> In this work, we represent the pair-wise spatial correlations between traffic sensors using a directed graph whose nodes are sensors and edge weights denote proximity between the sensor pairs measured by the road network distance. We model the dynamics of the traffic flow as a diffusion process and propose the *diffusion convolution* operation to capture the spatial dependency. We further propose *Diffusion Convolutional Recurrent Neural Network* (DCRNN) that integrates *diffusion convolution*, the *sequence to sequence* architecture and the *scheduled sampling* technique. When evaluated on real-world traffic datasets, DCRNN consistently outperforms state-of-the-art traffic forecasting baselines by a large margin. In summary:
>
> - We study the traffic forecasting problem and model the spatial dependency of traffic as a diffusion process on a directed graph. We propose *diffusion convolution*, which has an intuitive interpretation and can be computed efficiently.
> - We propose *Diffusion Convolutional Recurrent Neural Network* (DCRNN), a holistic approach that captures both spatial and temporal dependencies among time series using *diffusion convolution* and the sequence to sequence learning framework together with scheduled sampling. DCRNN is not limited to transportation and is readily applicable to other spatiotemporal forecasting tasks.
> - We conducted extensive experiments on two large-scale real-world datasets, and the proposed approach obtains significant improvement over state-of-the-art baseline methods.

With necessity, traffic prediction requires flow model, abstract path model and a data template to construct a algorithm.

Graph convolutional neural networks

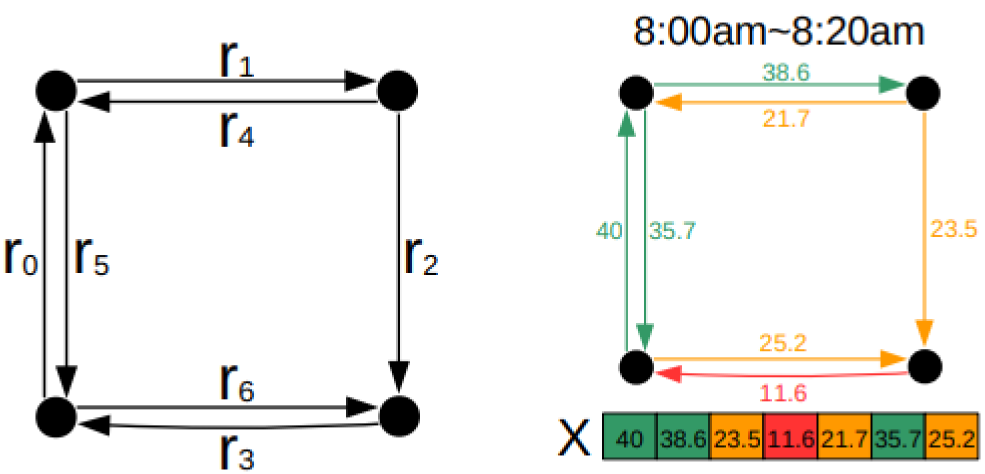
LC-RNN: A Deep Learning Model for Traffic Speed Prediction

Zhongjian Lv[1], Jiajie Xu[1,2,3*], Kai Zheng[4*], Hongzhi Yin[5*], Pengpeng Zhao[1], Xiaofang Zhou[5,1]
[1] School of Computer Science and Technology, Soochow University, China

first thing first

Abstract

Traffic speed prediction is known as an important but challenging problem. In this paper, we propose a novel model, called LC-RNN, to achieve more accurate traffic speed prediction than existing solutions. It takes advantage of both RNN and CNN models by a rational integration of them, so as to learn more meaningful time-series patterns that can adapt to the traffic dynamics of surrounding areas. Furthermore, since traffic evolution is restricted by the underlying road network, a network embedded convolution structure is proposed to capture topology aware features. The fusion with other information, including periodicity and context factors, is also considered to further improve accuracy. Extensive experiments on two real datasets demonstrate that our proposed LC-RNN outperforms seven well-known existing methods.



(a) A road network

(b) A speed vector at 8:00am - 8:20am

Figure 1: Road network and speed vector

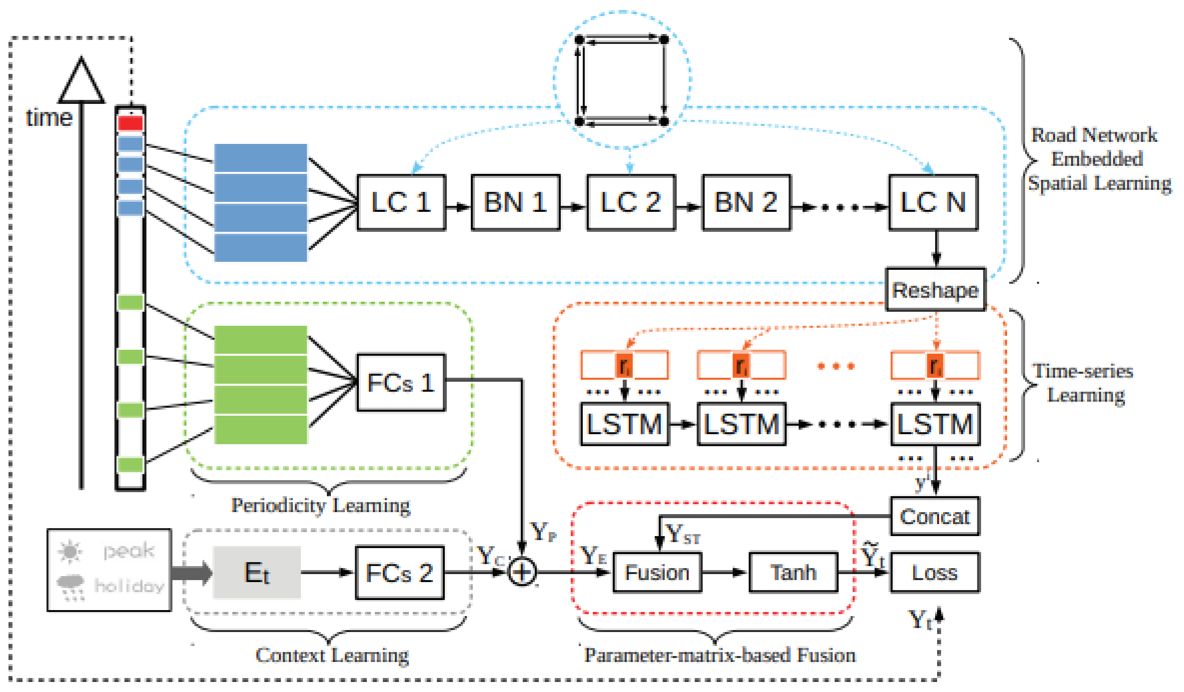


Figure 2: The architecture of LC-RNN. LC:Look-up Convolution; BN:Batch Normalization; FC:Fully-connected; LSTM:Long short-term memory

Traffic speed prediction (TSP) means to predict the future speed of each road segment based on historical observations.

Unclassified

Outdoor localization

DeepLoc: A Ubiquitous Accurate and Low-Overhead Outdoor Cellular Localization System

>
> ## ABSTRACT
>
> Recent years have witnessed fast growth in outdoor location-based services. While GPS is considered a ubiquitous localization system, it is not supported by low-end phones, requires direct line of sight to the satellites, and can drain the phone battery quickly.
>
> In this paper, we propose *DeepLoc*: a deep learning-based outdoor localization system that obtains GPS-like localization accuracy without its limitations. In particular, *DeepLoc* leverages the ubiquitous cellular signals received from the different cell towers heard by the mobile device as hints to localize it. To do that, crowd-sensed geo-tagged received signal strength information coming from different cell towers is used to train a deep model that is used to infer the user's position. As part of *DeepLoc* design, we introduce modules to address a number of practical challenges including scaling the data collection to large areas, handling the inherent noise in the cellular signal and geo-tagged data, as well as providing enough data that is required for deep learning models with low-overhead.
>
> We implemented *DeepLoc* on different Android devices. Evaluation results in realistic urban and rural environments show that *DeepLoc* can achieve a median localization accuracy within 18.8m in urban areas and within 15.7m in rural areas. This accuracy outperforms the state-of-the-art cellular-based systems by more than 470% and comes with 330% savings in power compared to the GPS. This highlights the promise of *DeepLoc* as a ubiquitous accurate and low-overhead localization system.
>

Auxiliary localization system

Forecasting queries

Exploiting Spatiotemporal Patterns for Accurate Air Quality Forecasting using Deep Learning

> and Beijing and compare our forecasting results to other existing approaches. The experiments show our GC-DCRNN model consistently outperforms other air quality forecasting methods (section 4).
>
> The main contribution of this paper is that we present the GC-DCRNN model that jointly manipulates the spatial and temporal dependencies in location-dependent air quality time series data for forecasting. We utilize an automatic approach to describe the spatial dependency by considering the similarity of the built environment with regard to air quality. We construct a geo-context based graph that enables the DCRNN model to handle the spatial correlations by automatically selecting important geographic feature types that

A approach to apply spatial data into wider usage. But without aerodynamic, the work speaks weakly.

Horizon	Hybrid Model [34]	**GC-DCRNN**
h=1	8.44	**8.10**
h=2	13.07	**12.15**
h=3	16.62	**15.26**
h=4	19.45	**17.85**
h=5	21.83	**20.13**
h=6	23.86	**22.05**
h=7-12	28.19	**25.45**
h=13-24	34.41	**31.43**

Additionally its result shows no remarkable predomination.

Geo-spatial object detection

An Unsupervised Augmentation Framework for Deep Learning based Geospatial Object Detection: A Summary of Results

> Given a remote sensing dataset, we aim to automatically generate catalogs of geospatial objects (e.g., buildings, vehicles, farm fields) by leveraging deep learning frameworks for object detection. The detection of an object O is modeled by the **Minimum Bounding Rectangle (MBR)** of O, which is the smallest rectangle (i.e., angle-aware) that entirely covers O. Fig. 1 shows an example of an input image and the MBR output.
>
> 
>
> (a) Input image (b) MBR output
>
> **Figure 1: Example input and MBR output.**

So in my view, it matches CV better.

Q3

Spatial data mining

Spatial classification

Spatial classification

> **Abstract**
>
> The aim of a spatial classification is to position the units on a spatial network and to give simultaneously a set of structured classes of these units "compatible" with the network. We introduce the basic needed definitions: compatibility between a classification structure and a tessellation, (m, k)-networks as a case of tessellation, convex, maximal and connected subsets in such networks, spatial pyramids and spatial hierarchies. As like Robinsonian dissimilarities induced by indexed pyramids generalize ultrametrics induced by indexed hierarchies we show that a new kind of dissimilarity called "Yadidean" induced by spatial pyramids generalize Robinsonian dissimilarities. We focus on spatial pyramids where each class is a convex for a grid, and we show that there are several one-to-one correspondences with different kinds of Yadidean dissimilarities. These new results produce also, as a special case, several one-to-one correspondences between spatial hierarchies (resp. standard indexed pyramids) and Yadidean ultrametrics (resp. Robinsonian) dissimilarities. Qualities of spatial pyramids and their supremum under a given dissimilarity are considered. We give a constructive algorithm for convex spatial pyramids illustrated by an example. We show finally by a simple example that spatial pyramids on symbolic data can produce a geometrical representation of conceptual lattices of "symbolic objects".
> © 2007 Elsevier B.V. All rights reserved.
>
> *Keywords:* Pyramidal clustering; Spatial classification; Symbolic data analysis; Conceptual Lattices; Kohonen mapping

It gives a basic definition of **Spatial Classification** while this article is widely quoted by many spatial articles. Yet I am new to it and expecting a further study.

Spatial Big Data Science: Classification Techniques for Earth Observation Imagery.

> **Abstract** This chapter overviews earth observation imagery big data and its general classification methods. We introduce different types of earth observation imagery big data and their societal applications. We also summarize some general classification algorithms. Open computational challenges are also identified in this area.

This book gives an overview of spatial data usage.

> **Part II Classification of Earth Observation Imagery Big Data**
>
>

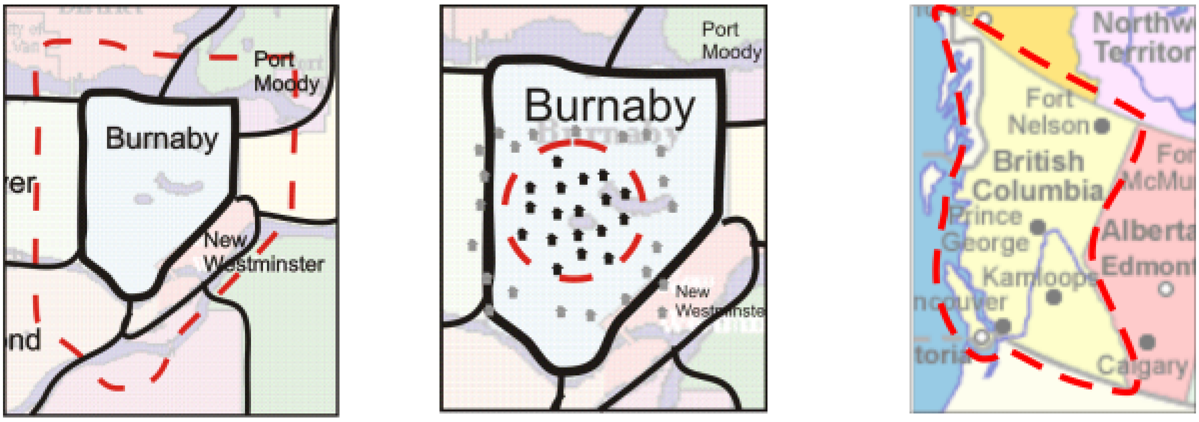
This part is recommended as the attachment to machine learning.

The two contents above worthy study in detail later.

A Multi-Relational Approach to Spatial Classification

> Spatial classification is the task of learning models to predict class labels based on the features of entities as well as the spatial relationships to other entities and their features. Spatial data can be represented as multi-relational data, however it presents novel challenges not present in multi-relational problems. One such problem is that spatial relationships are embedded in space, unknown a priori, and it is part of the algorithm's task to determine which relationships are important and what properties to consider. In order to determine when two entities are spatially related in an adaptive and non-parametric way, we propose a

This article presents a way of classification by study their geometry relationships.



a) around a city b) around a house c) around a province

Figure 3: 3 km buffer zone around different entity types

This figure illustrates a way to define an area.

Spatial Ensemble Learning for Heterogeneous Geographic Data with Class Ambiguity: A Summary of Results

cannot effectively minimize class ambiguity. In contrast, our spatial ensemble framework explicitly partitions input data in geographic space. Our approach first preprocesses data into homogeneous spatial patches and uses a greedy heuristic to allocate pairs of patches with high class ambiguity into different zones. Both theoretical analysis and experimental evaluations on two real world wetland mapping datasets show the feasibility of the proposed approach.

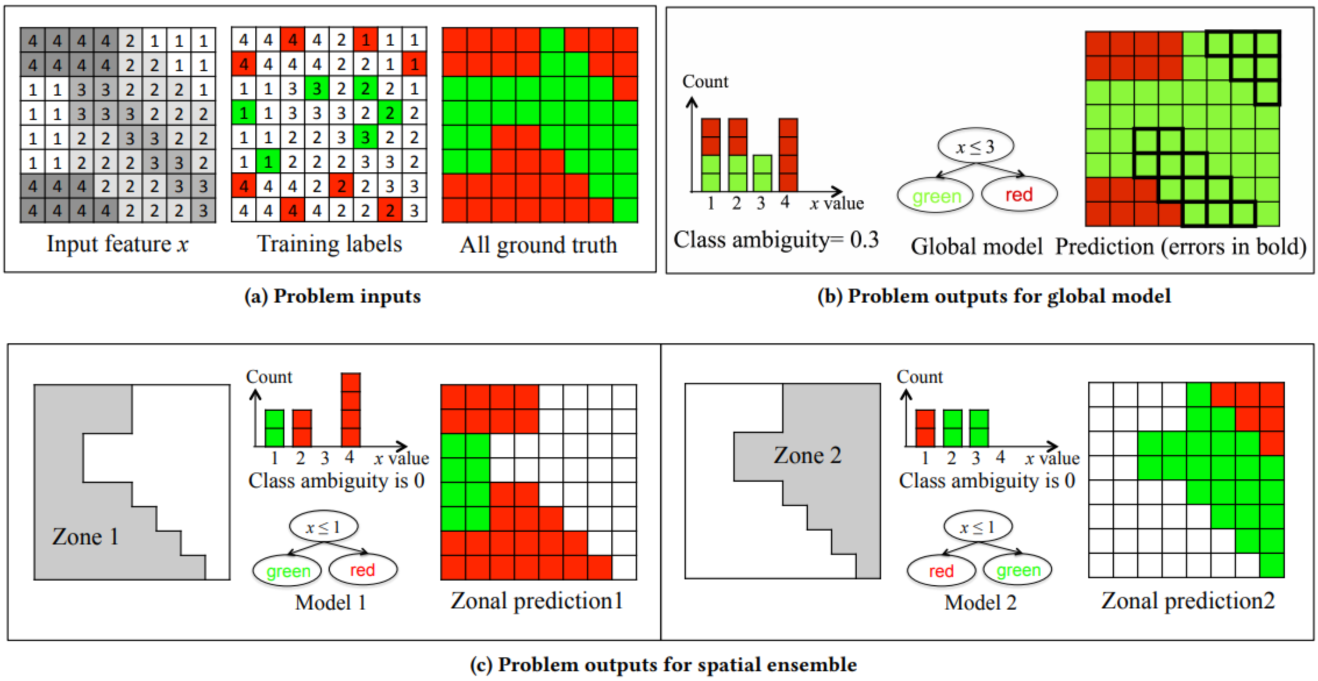


Figure 2: Illustrative example of problem inputs and outputs (best viewed in color)

This article applies a heuristic algorithm to rank the so-called truth score and classification an area.

Spatial regression

Spatial clustering

Hotpot detection

Spatial outliers detection

Trajectory analysis

Spatial statistical inference

In-memory sampling

Constrains-based sampling

Spatial sampling

Incremental sampling

Constrains-based sampling

Q1

Spatial-aware deep learning

Quad-tree-based neural networks

KNN-based neural networks

Spatial-aware statistical ML

Spatial Markov logic networks

Spatial collaborative filtering