机器学习、人工智能与深度学习

王静远

北京航空航天大学

背景知识

大数据时代

- 海量数据的存在

新浪微博 weibo.com

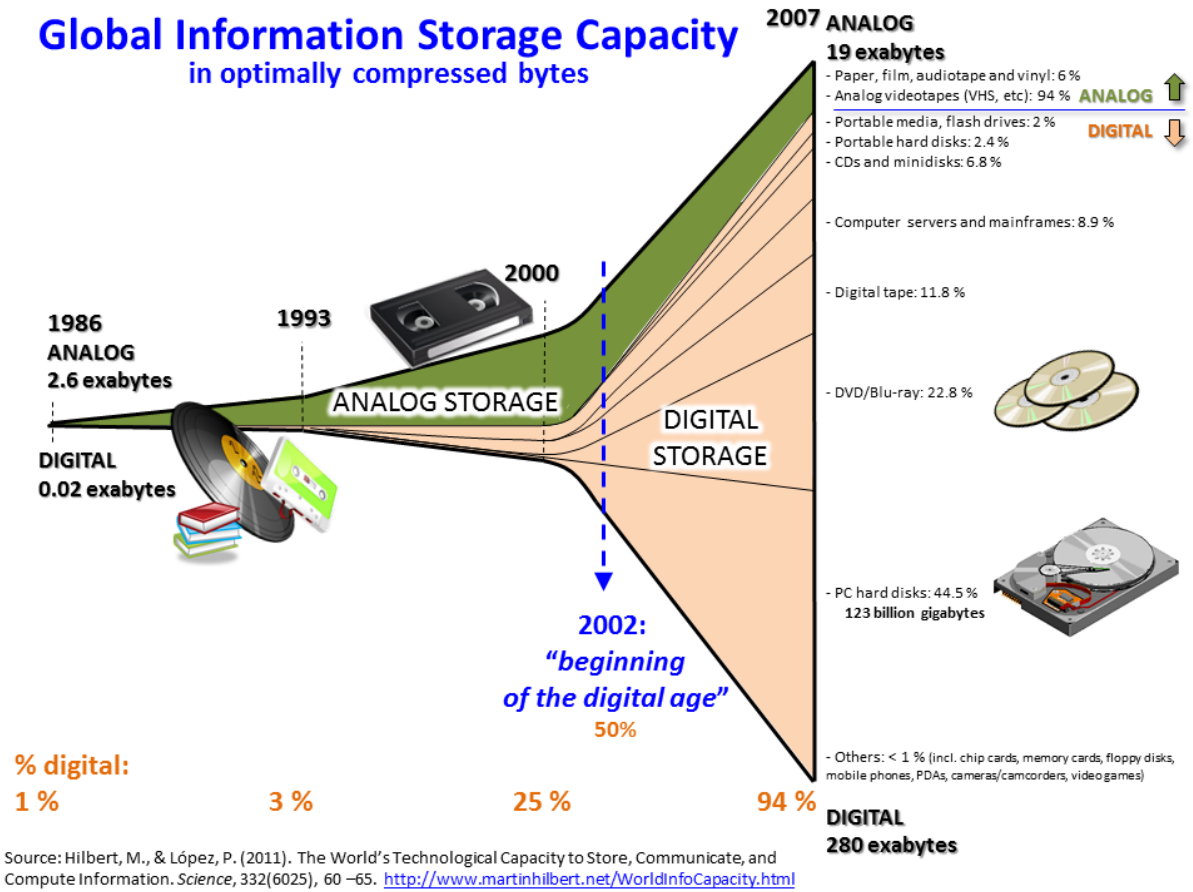
3亿用户,每天上亿条微博.

巡天望远镜,每年收集

600万G字节数据

2015年全球移动终端产生的数据量

6300PB

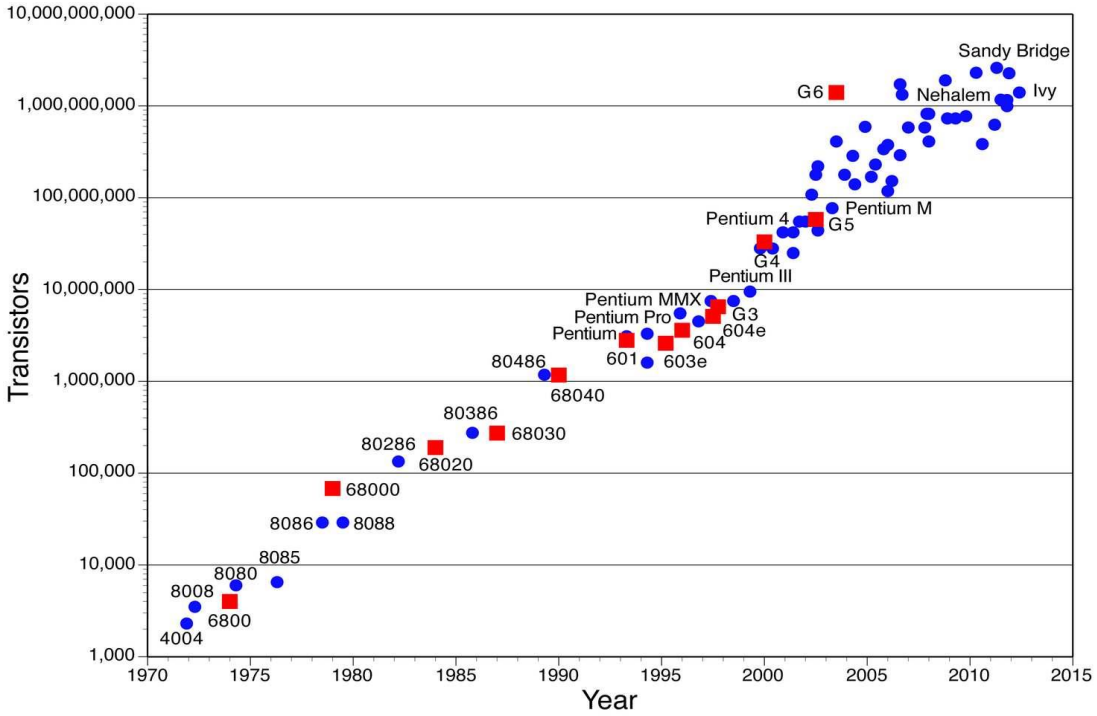


Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

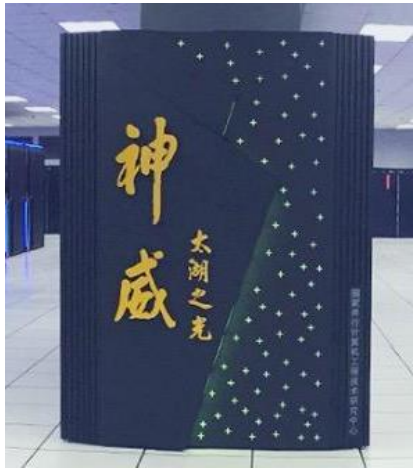
超强的计算能力

- 计算能力的增强

摩尔定律



峰值计算速度54.9PFLOPS



峰值计算速度125.436PFlops

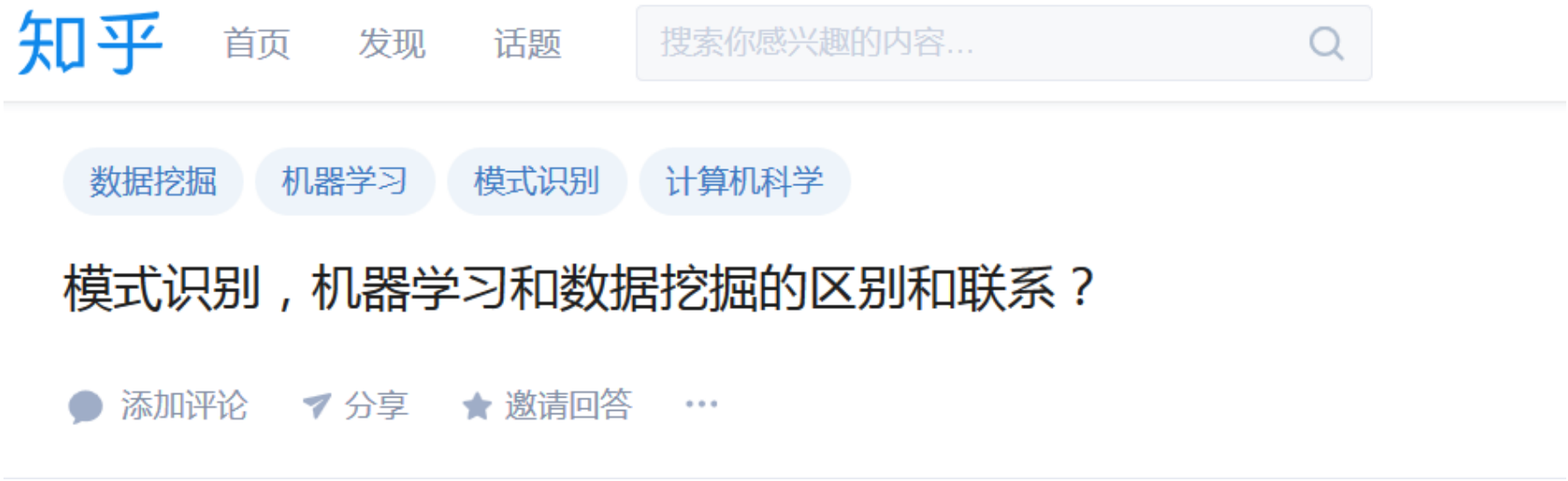
2015年5月，“天河二号”上成功进行了3万亿粒子数中微子和暗物质的宇宙学N体数值模拟，揭示了宇宙大爆炸1600万年之后至今约137亿年的漫长演化进程。

第四范式

- 第一范式：实验科学
  - 几千年前的科学，以记录和描述自然现象为主，其典型案例如钻木取火；
- 第二范式：理论科学
  - 数百年前，科学家们开始利用模型归纳总结过去记录的现象，其典型案例如牛顿三定律、麦克斯韦方程组、相对论等；
- 第三范式：计算科学
  - 过去数十年，科学计算机的出现，诞生了“计算科学”，对复杂现象进行模拟仿真，推演出越来越多复杂的现象，其典型案例如模拟核试验、天气预报等；
- 第四范式：数据密集型科学
  - 今天以及未来科学的发展趋势是，随着数据量的高速增长，计算机将不仅仅能做模拟仿真，还能进行分析总结，得到理论。也就是说，过去由牛顿、 爱因斯坦等科学家从事的工作，未来可以由计算机来做。

智能计算

机器学习　模式识别　数据挖掘　人工智能



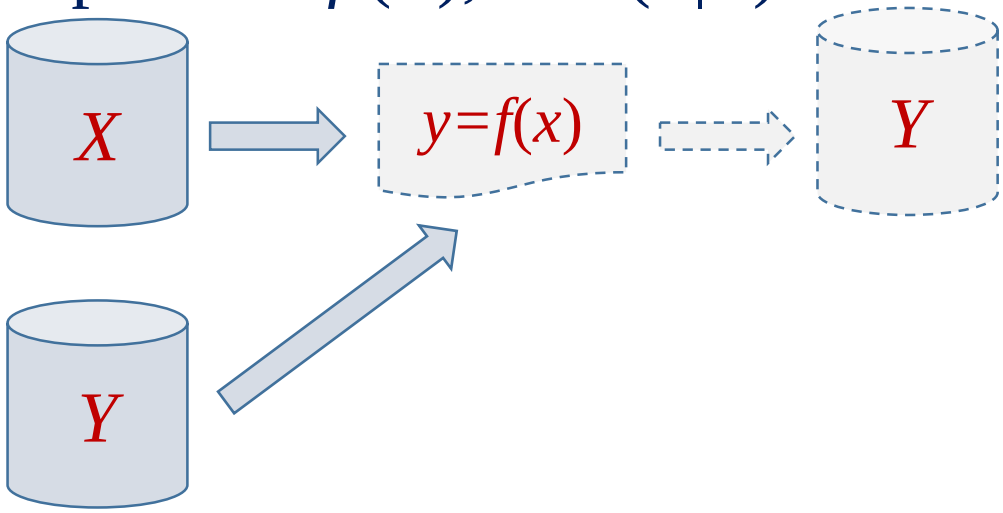
机器学习

- Tom Mitchell的机器学习(1997)对信息论中的一些概念有详细的解释,其中定义机器学习时提到，“**机器学习研究的算法是能够使用经验自动改进性能**”。（Machine Learning is the study of computer algorithms that improve automatically through experience.）
- Alpaydin（2004）提出自己对机器学习的定义，“**机器学习是用数据或以往的经验，以此优化计算机程序的性能标准。**”（Machine learning is programming computers to optimize a performance criterion using example data or past experience.）

机器学习

有监督学习

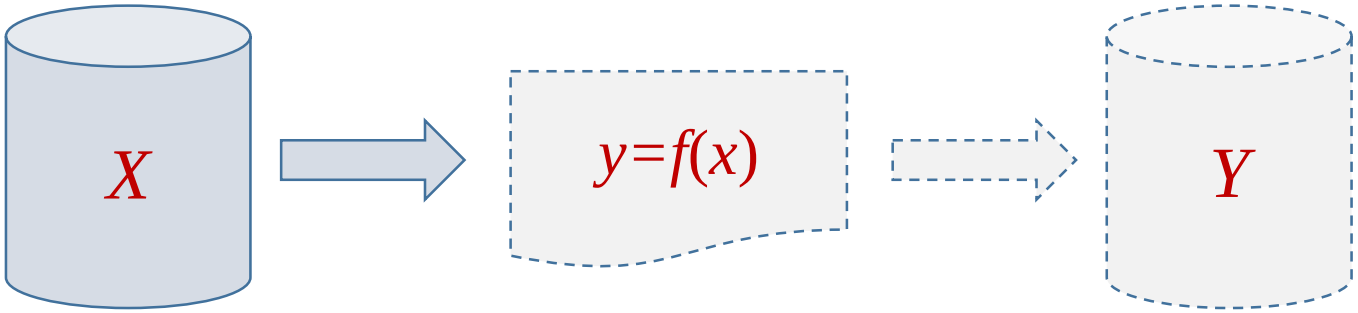
- 分类问题
- Input: training set $\underline{X}$, $\underline{Y}$; testing set X
- Output: $Y = f(X)$, or $P(Y|X)$



机器学习

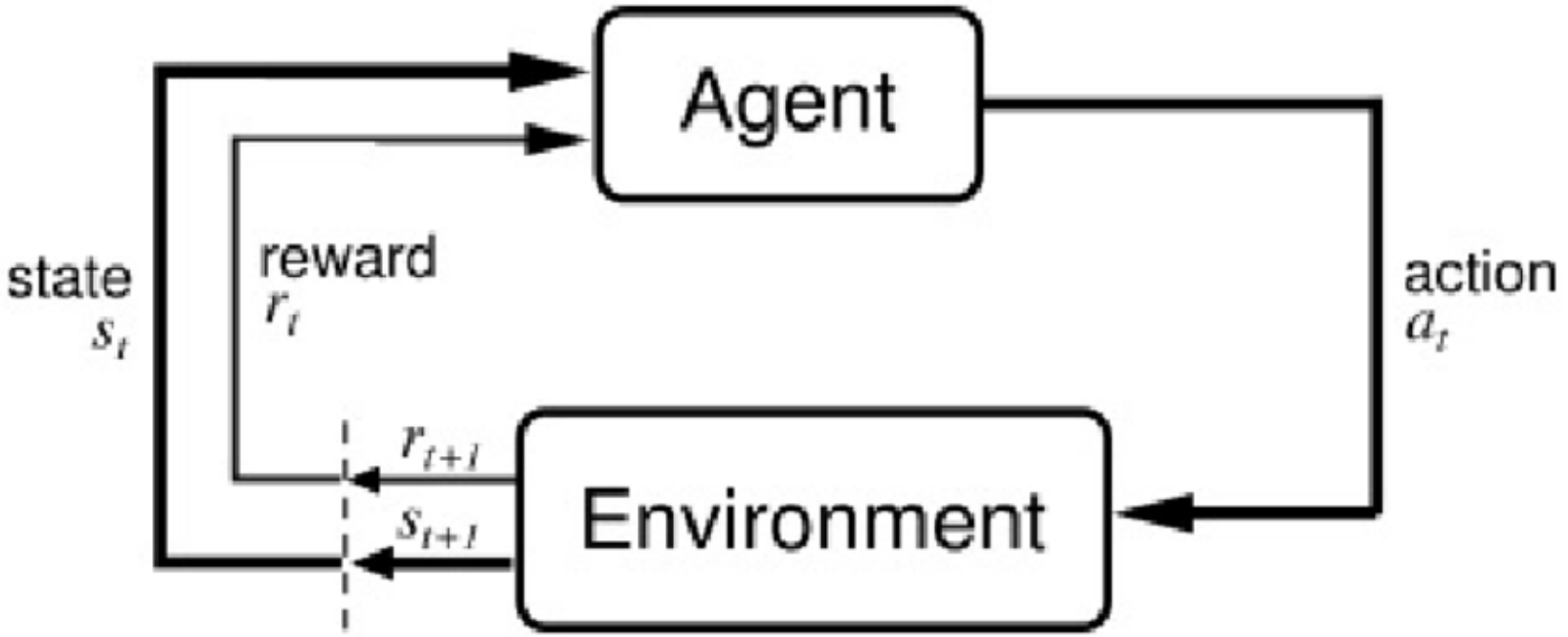
无监督学习

- 聚类问题
- Input: X
- Output: $Y = f(X)$, or $P(Y|X)$



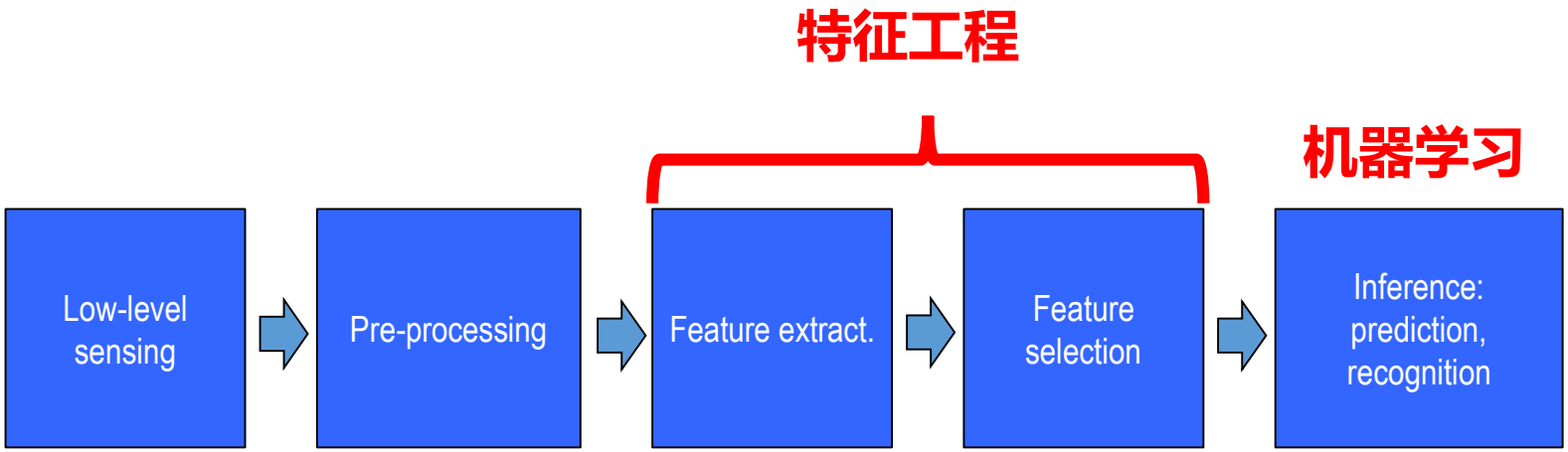
机器学习

强化学习



模式识别

- 模式识别（英语：Pattern Recognition），就是通过计算机用数学技术方法来研究模式的自动处理和判读。

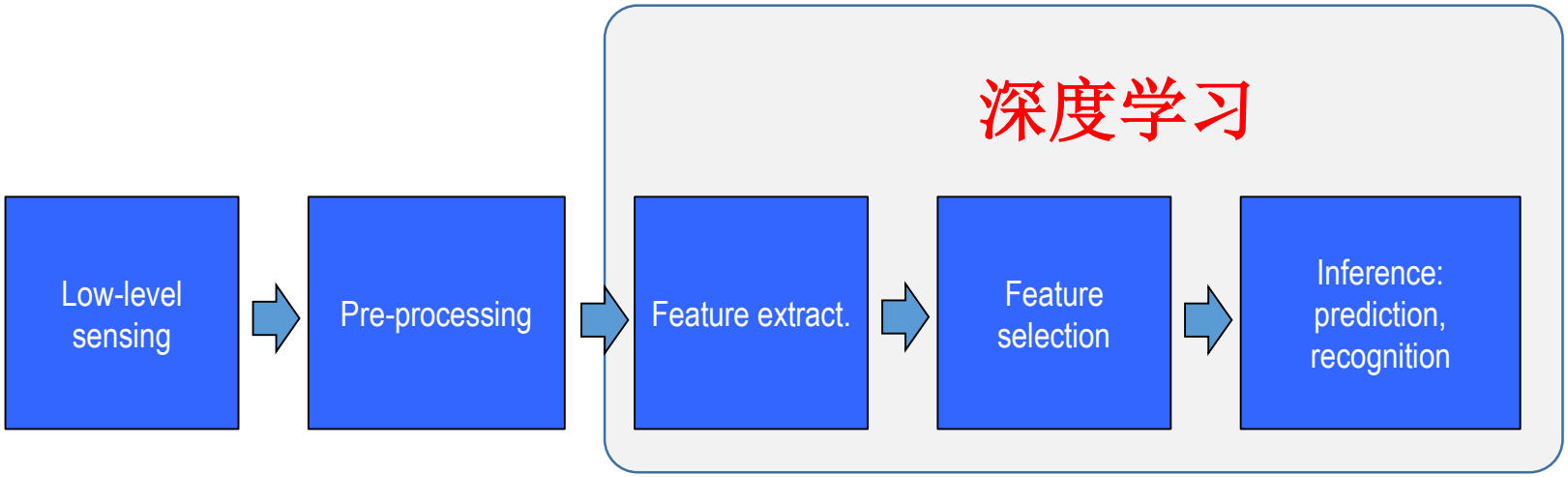


特征工程


现有的移动互联网用户行为数据
移动App
位置
搜索
浏览
女性
起居：晚12点-早7点
喜欢瑜伽/常慢跑
居住地:北京
家有孩子 幼儿期
80后 白领
常看电影
常去星巴克
喜欢 兰蔻
常去上海
爱打扮
宾馆：中高档
炒股
中移动4g用户低流量
使用银行：工行
喜欢海淘 常手机支付
自有住房/还贷中
喜欢做菜
爱看美剧
关注时尚
关注可穿戴设备
在学车
行为属性+基础标签 概率性描述
作息规律
注重品质
生活健康
爱尝试新事物
小资
宅人 0%
32-35岁：60%

模式识别

- 模式识别（英语：Pattern Recognition），就是通过计算机用数学技术方法来研究模式的自动处理和判读。



什么是数据挖掘

- 在大型数据存储中，自动地发现有用信息的过程
  - 探查大型数据集，发现先前**未知**的**有用**信息
  - 或是预测未来观测结果

如：预测某新客户是否会在一家商场消费100元以上？

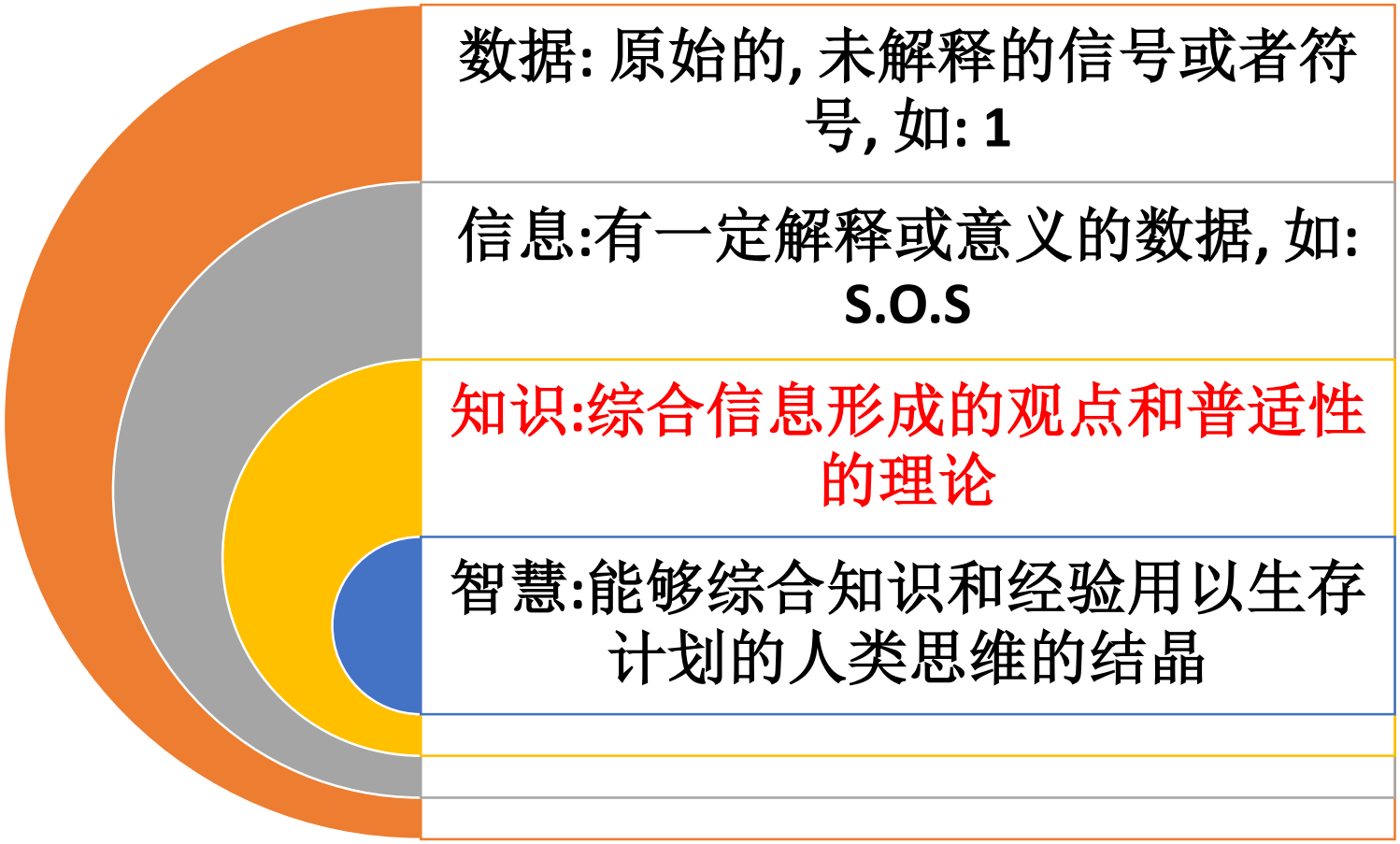
- 更严谨的表述
  - 数据挖掘就是**从大量的、不完全的、有噪声的、模糊的、随机的**实际应用数据中，提取**隐含在其中的、人们事先不知道的**、但又是**潜在有用**的信息和知识的过程。

什么（不）是数据挖掘

- 非数据挖掘
  - 从电话簿查找电话号码
  - 从Web中查找信息“数据挖掘”
  - 获得职工的平均资薪
  - ….
- 数据挖掘
  - 某插班生应该读几年级?
  - 买哪只股票更可能挣钱?
  - 怎么才能多卖化妆品?
  - 海量文档该如何归类?
  - 行驶车辆如何预警?
  - 广告如何派送更好?
  - ….

数据挖掘的核心任务是**知识发现**

- Knowledge Discovery in Database （KDD）



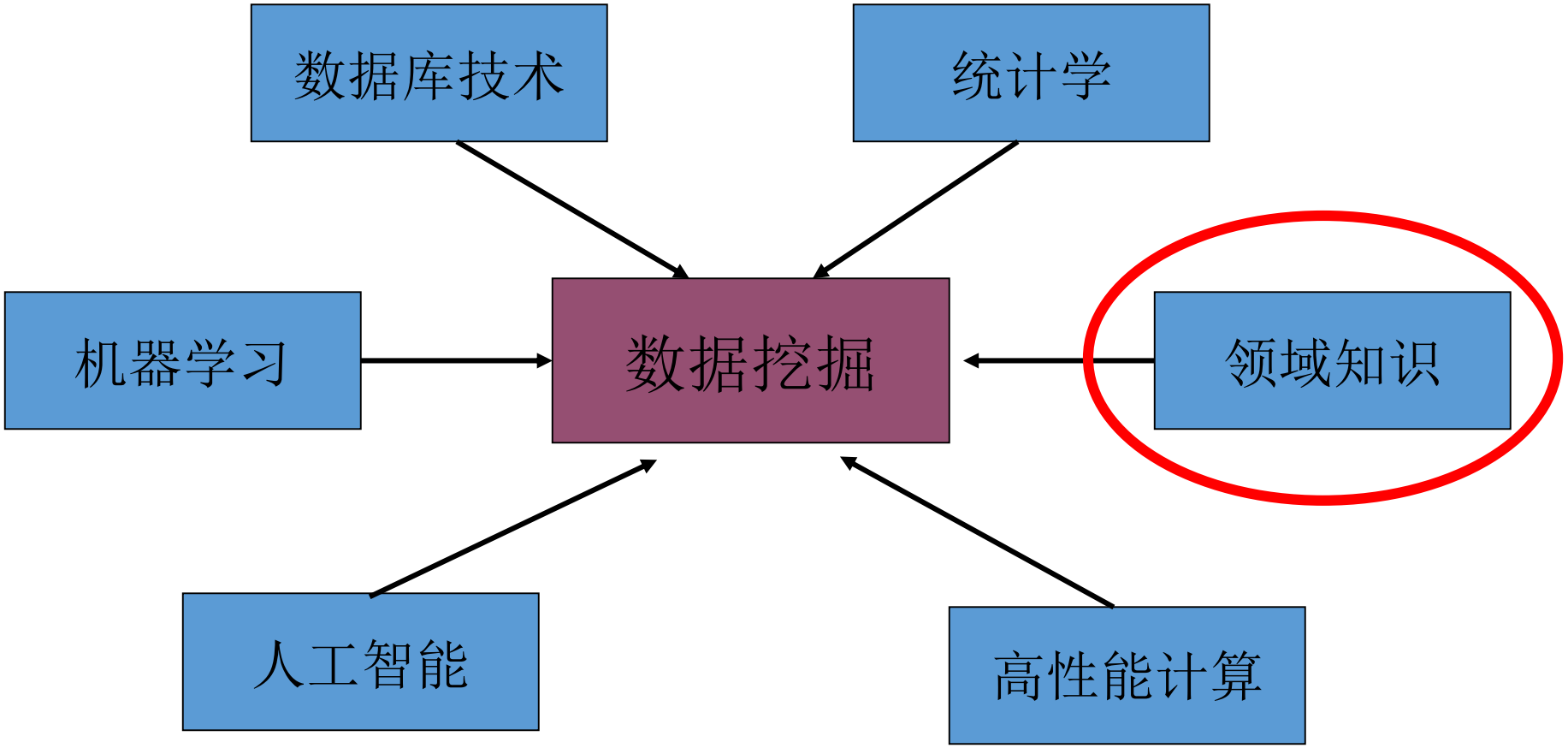
一句话总结数据挖掘

从数据中获取知识！

为决策（应用）提供支持！

数据挖掘和其它课程的关系

- 数据挖掘是多学科交叉的产物



人工智能?

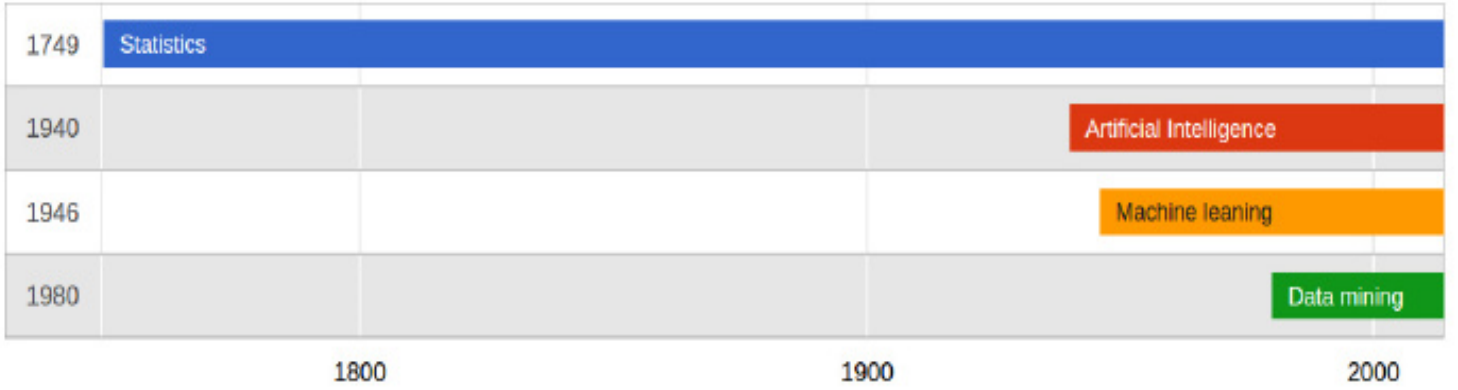
- 人工智能（Artificial Intelligence），英文缩写为AI。它是研究、开发用于**模拟、延伸和扩展人的智能**的理论、方法、技术及应用系统的一门新的技术科学。

统计学——1749年

人工智能——1940年

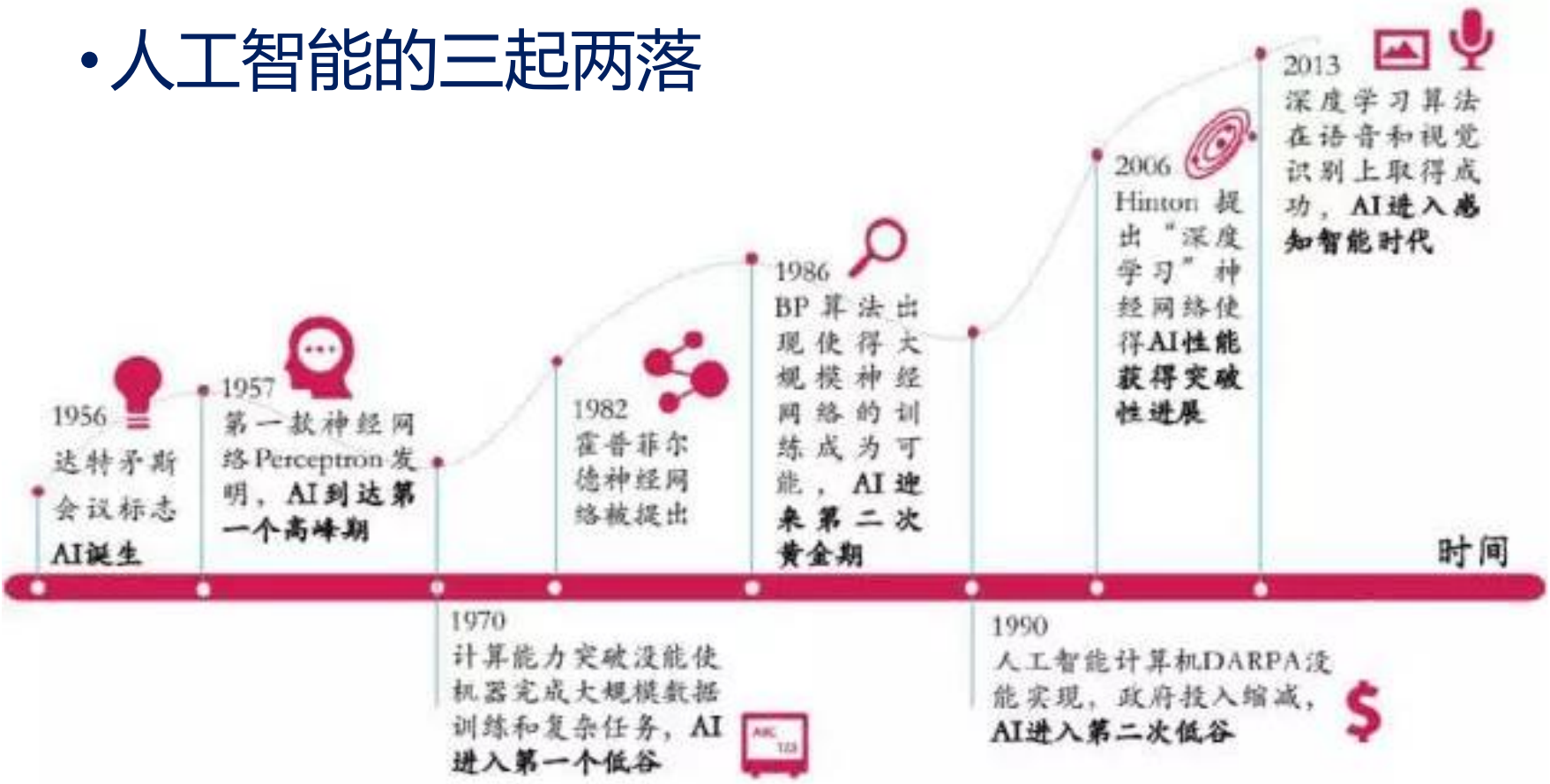
机器学习——1946年

数据挖掘——1980年



人工智能?

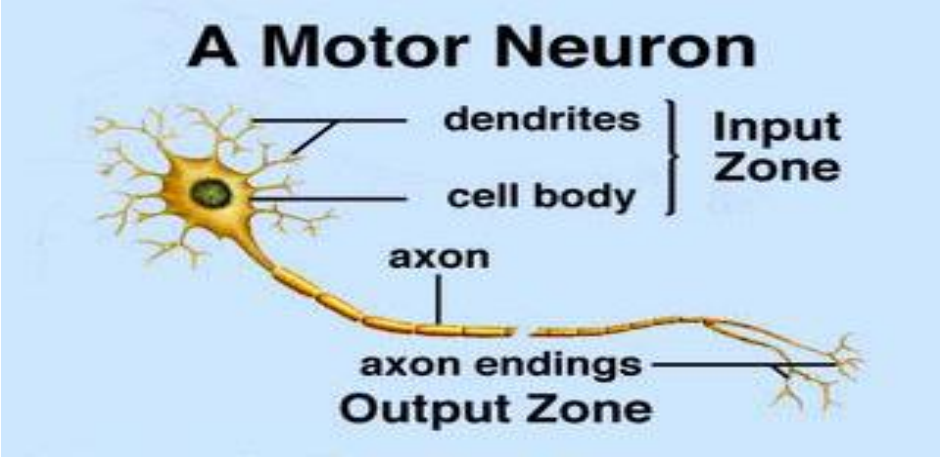
- 人工智能的三起两落



深度学习

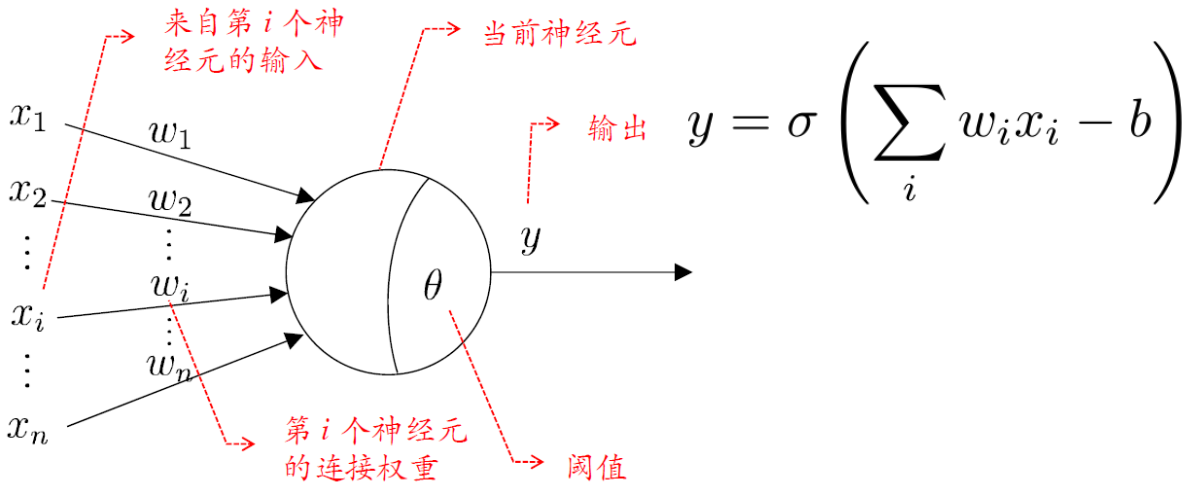
神经元模型

- 生物神经元
  - 神经元由两个部分构成：Input Zone (树突和胞体)，Output Zone（轴突和轴突末端）。在两个神经元之间，输出细胞的轴突末端和输入细胞的树突相连接。
  - 每个神经元有两种状态：“兴奋”与“抑制”。处于“兴奋”态时，神经元发出输出脉冲，并由轴突末端传递给其他神经元。
  - 神经元平时处于“抑制”状态，其树突接受其它“兴奋”态神经元传来的兴奋电位。如果输入兴奋电位总量超过某个阈值，神经元会被激发进入兴奋状态，发出输出脉冲，并由突触传递给其他神经元。



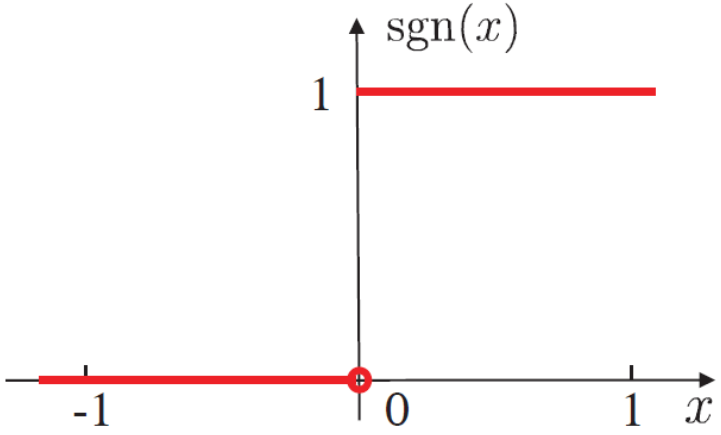
神经元模型

- **输入**：来自其它 n 个神经元传递过来的输入信号
- **处理**：输入信号通过带权重的连接进行传递，神经元接受到总输入值将与神经元的阈值进行比较
- **输出**：通过激活函数的处理以得到输出



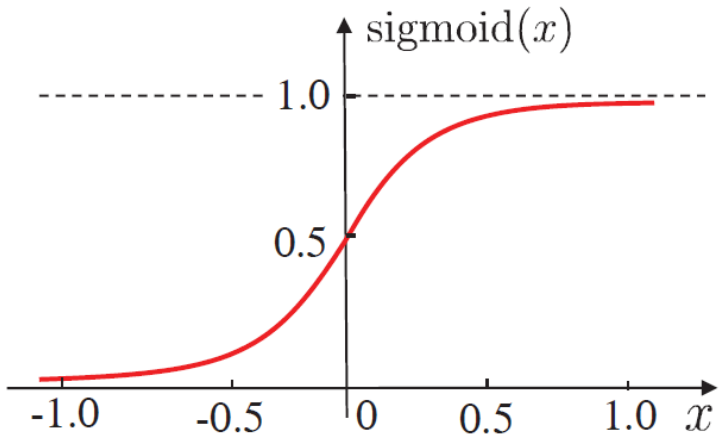
神经元模型

- 理想激活函数是阶跃函数, 0表示抑制神经元而1表示激活神经元
- 阶跃函数具有不连续、不光滑等不好的性质, 常用的是 Sigmoid 函数



$$\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

(a) 阶跃函数



$$\mathrm{sigmoid}(x) = \frac{1}{1+e^{-x}}$$

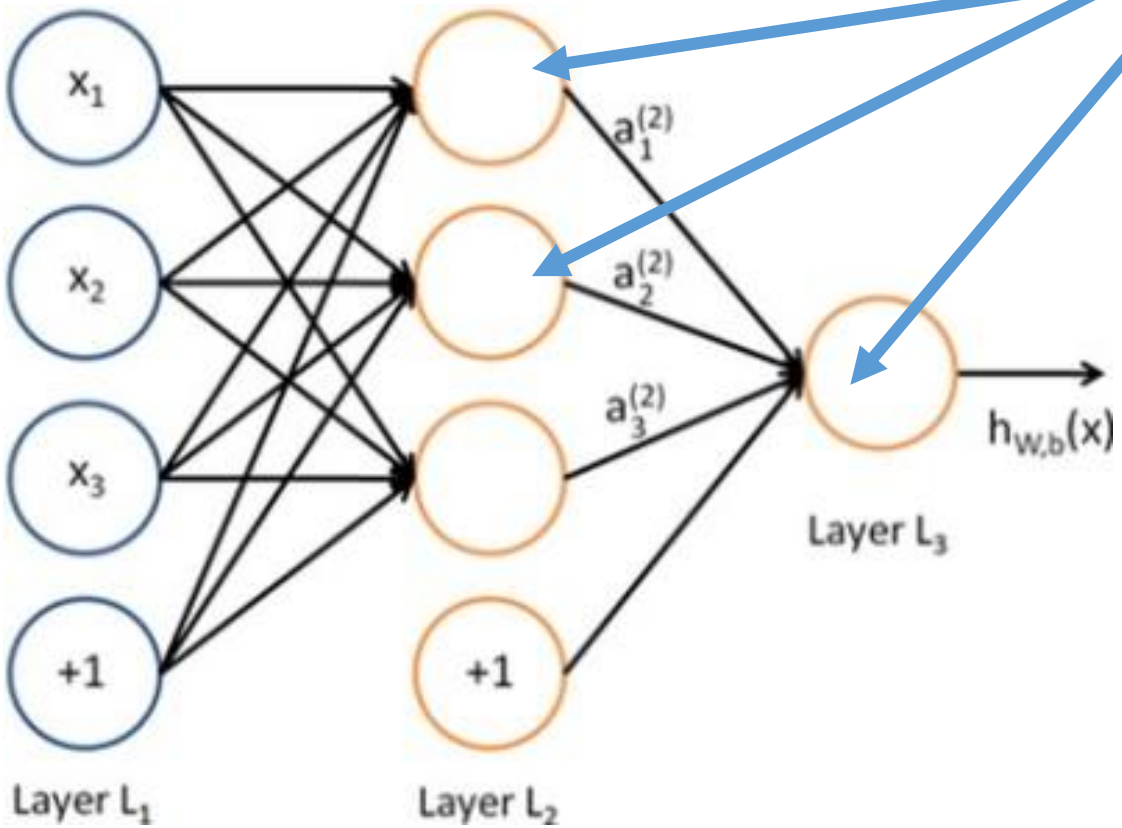
(b) Sigmoid 函数

神经网络模型

- 多层感知机（MLP）

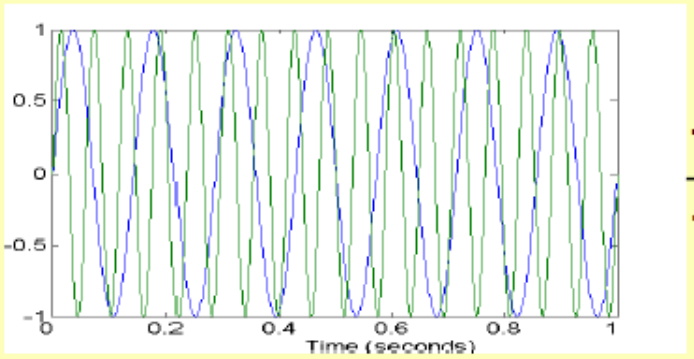
$$y = \sigma\left(\sum_i w_i x_i - b\right)$$

逻辑回归



输入层　　隐藏层　　输出层

自动化的特征学习方法

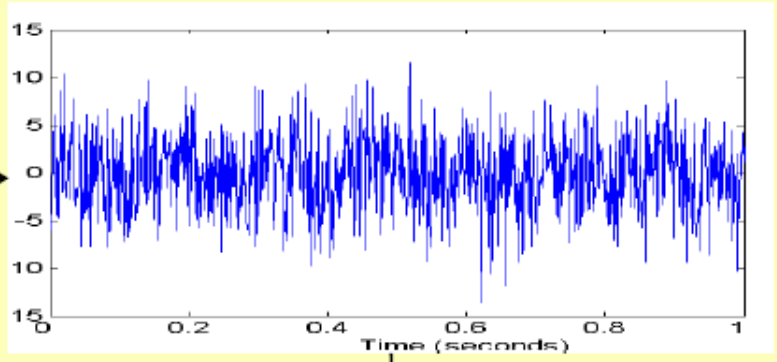


Two Sine Waves

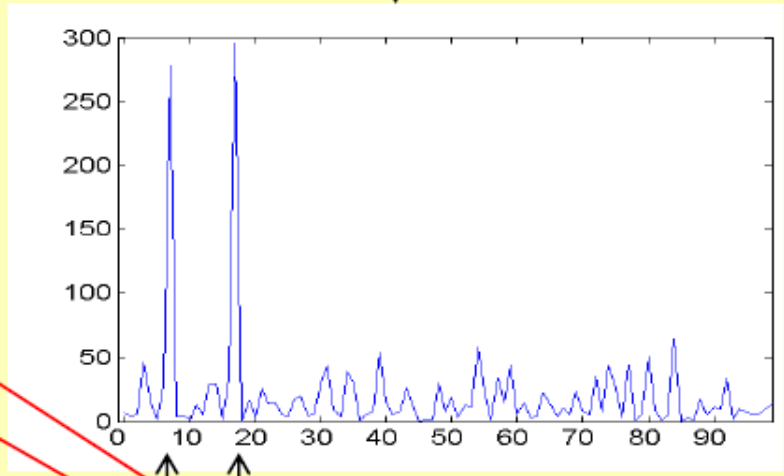
（7个/s + 17个/s）

Two Sine Waves

+ Noise



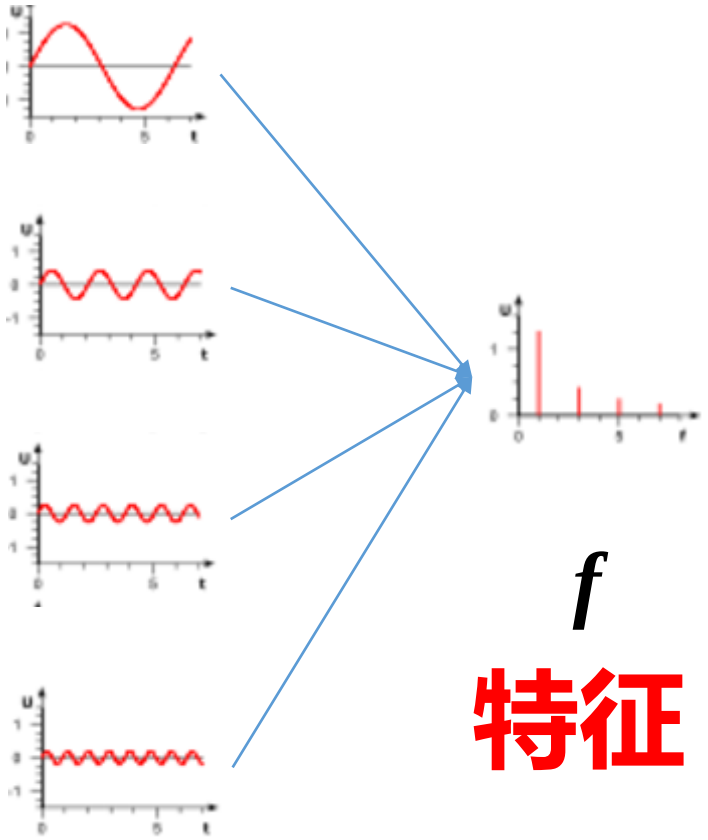
傅里叶变换



7 17

自动化的特征学习方法

字典

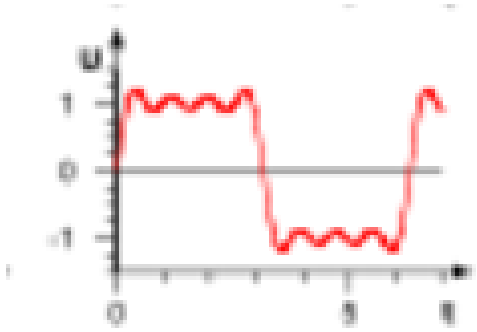


d

f

特征

$\approx$

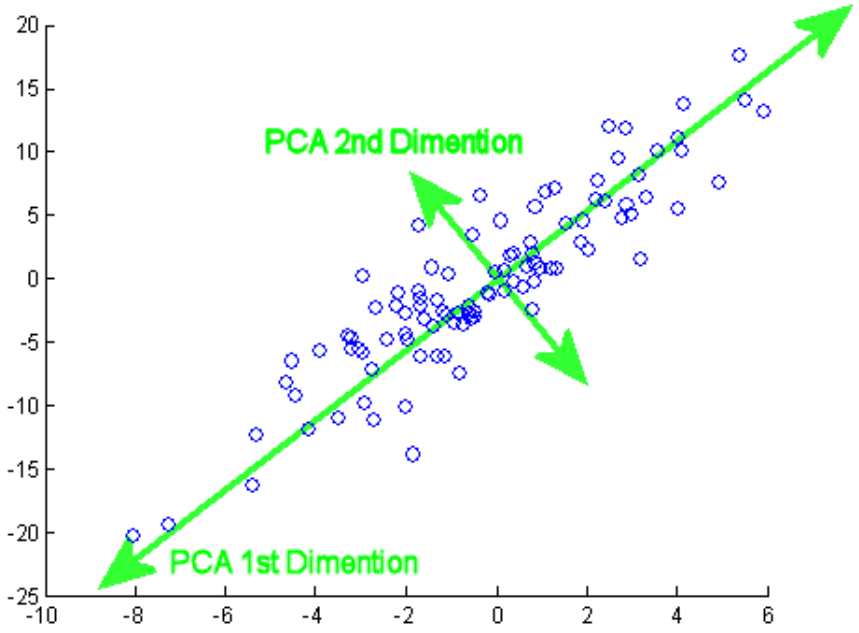


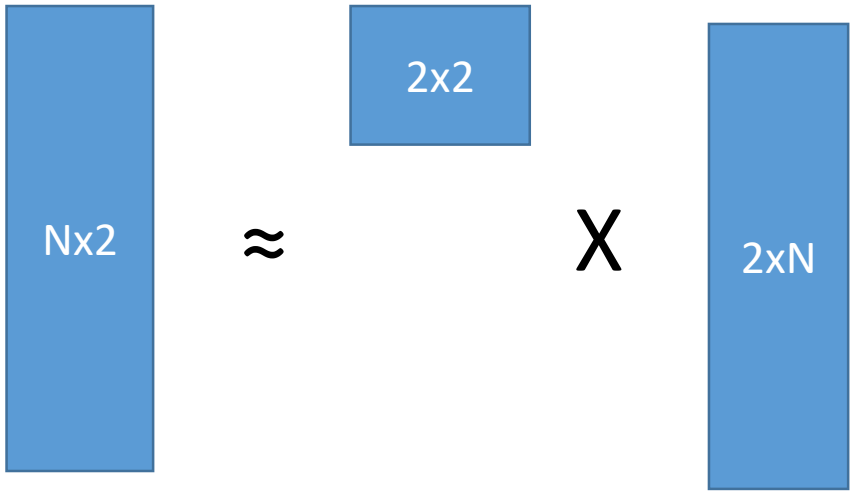
x

对象

自动化的特征学习方法

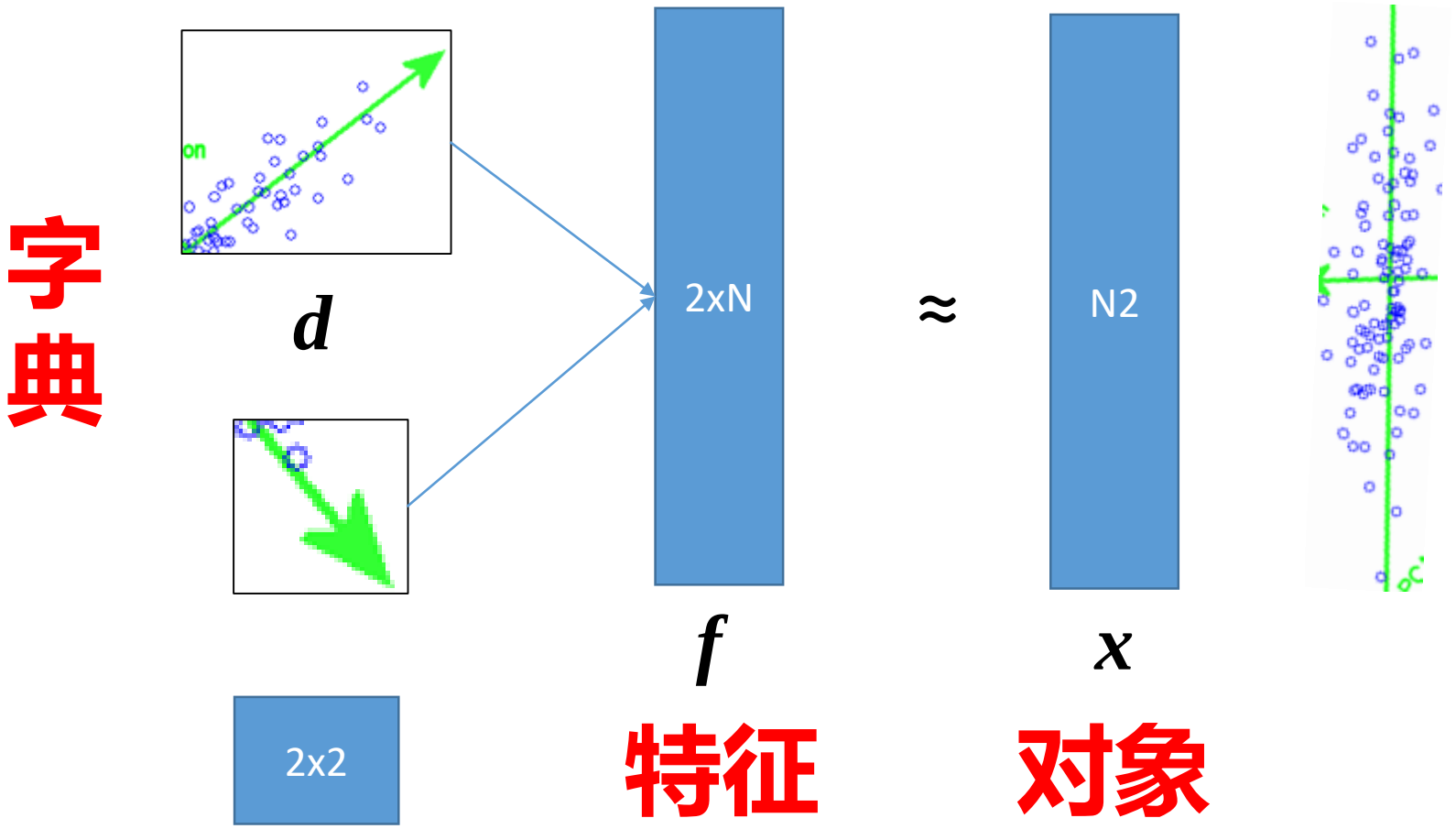
- PCA 主成分分析





自动化的特征学习方法

- PCA 主成分分析



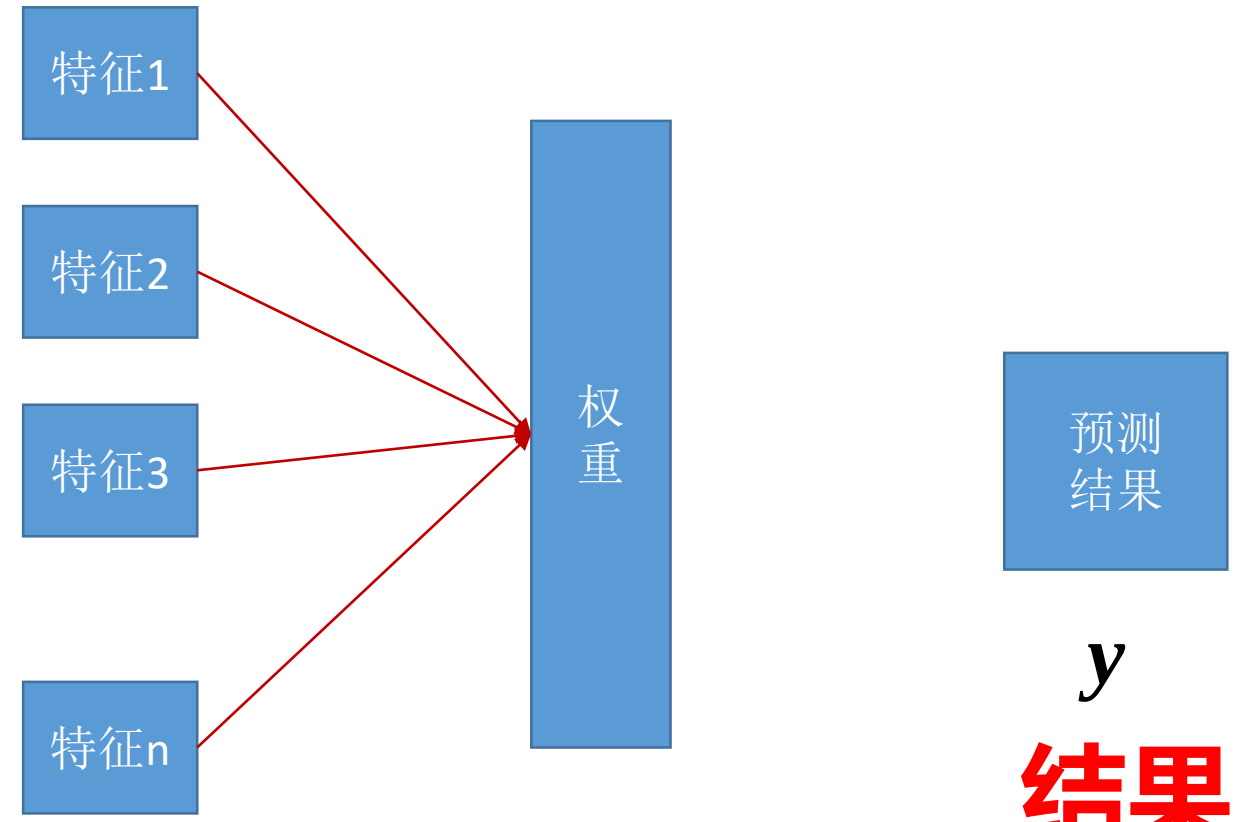
自动化的特征学习方法

- 将对象分解成为一组**“字典”**的线性组合
- 线性组合的权重是一种有效的**特征**
- 由**正交**字典所产生的特征是一种非常有效的特征
  - 信息冗余最小
  - 稀疏性最强

自动化的特征学习方法存在的问题

- **固定字典：**傅立叶变换、小波变换等
  - 优点：字典已经确定，因此特征的物理含义较为明确
  - 缺点：固定的字典不是对所有的数据集都适用，固定的字典产生的特征不是对有所的问题都有效
- **自学习字典：**PCA、ICA等
  - 有点：字典通过数据产生，对数据的适应性较好
  - 缺点：字典的学习目标不明确，对问题的针对性不强

基于线性/逻辑回归的预测



f

特征

w

权重

y

结果

结合起来看

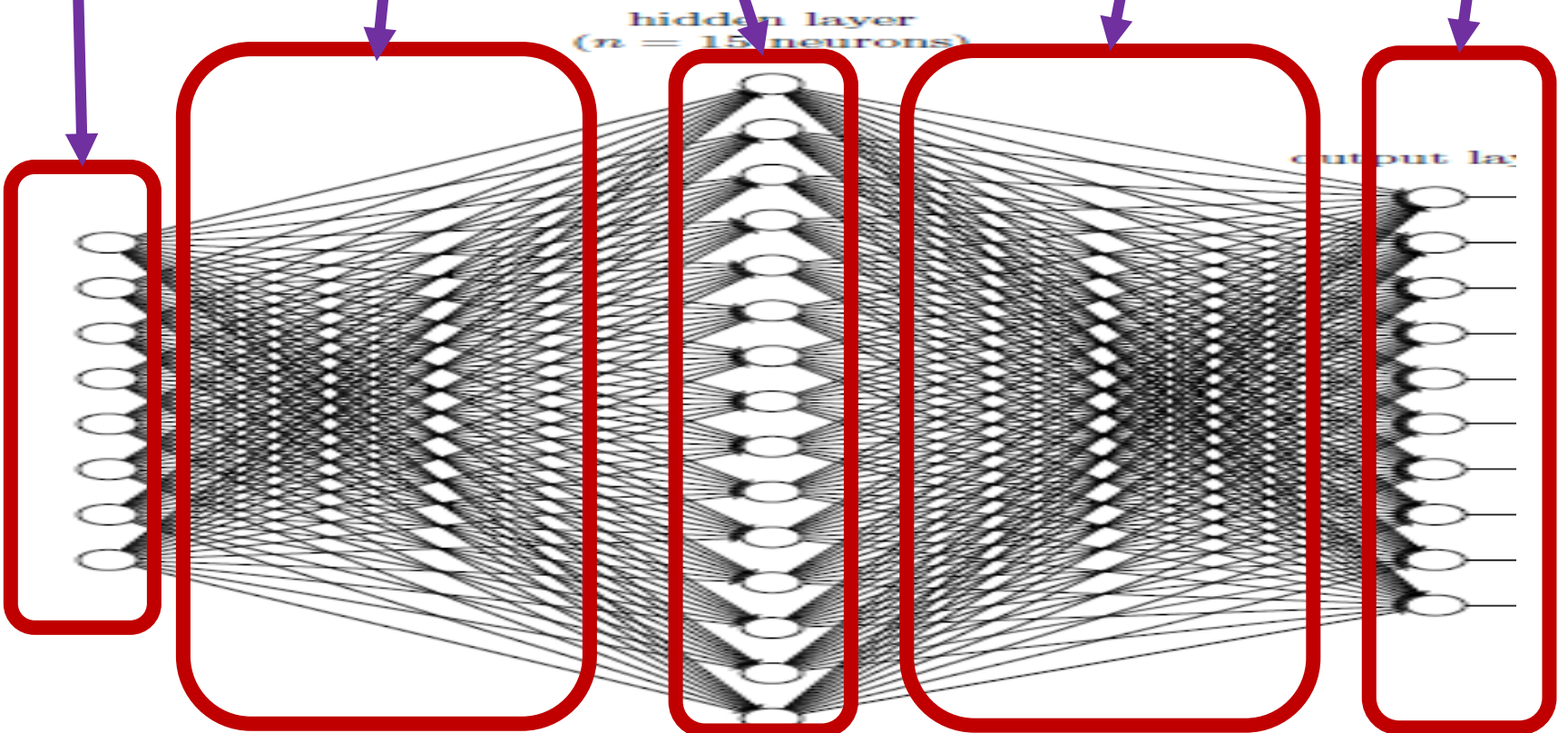
字典 X 特征 ≈ 数据

数据 X 字典$^{-1}$ ≈ 特征

特征 X 权重 ≈ 结果

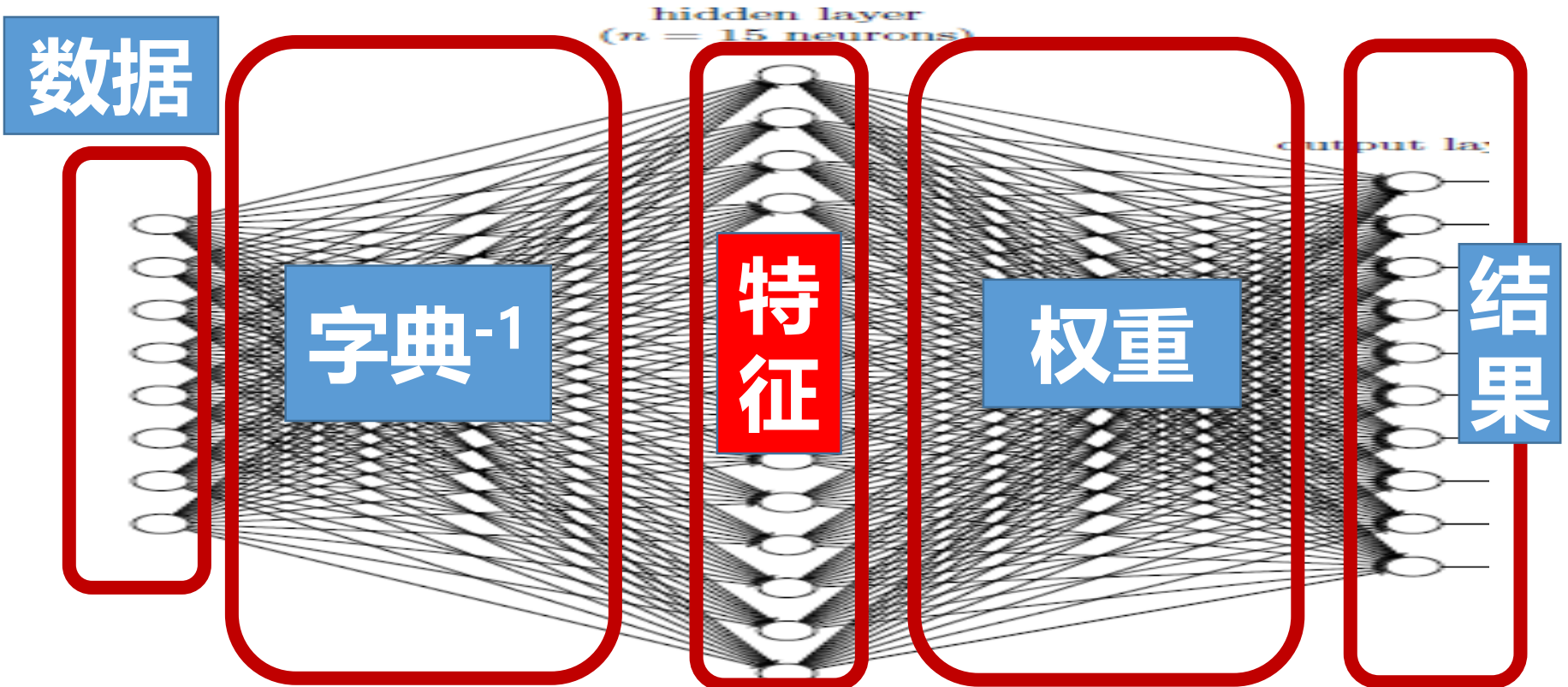
数据 X 字典$^{-1}$ ≈ 特征

特征 X 权重 ≈ 结果

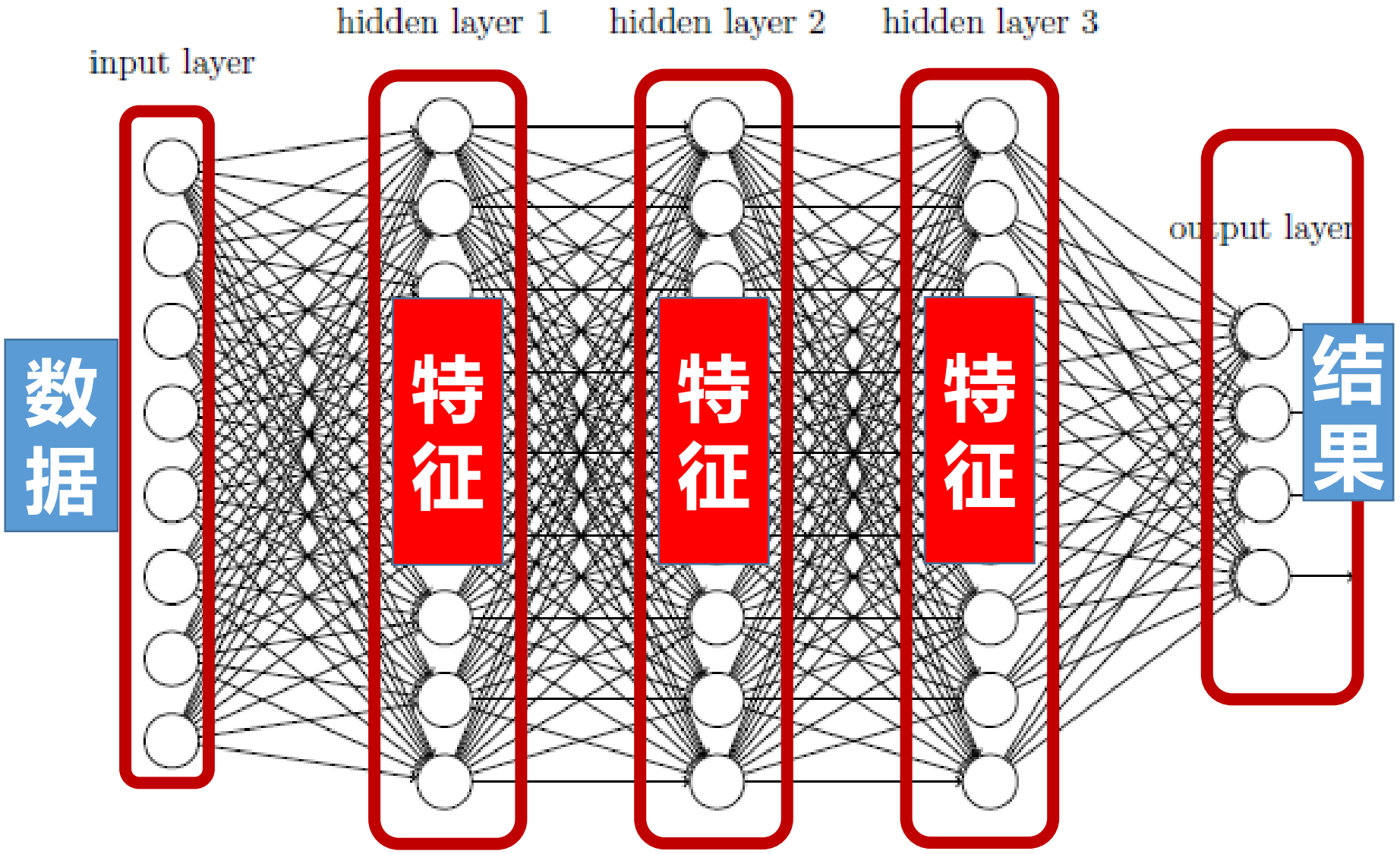


Representation Learning

- 表征学习
  - 特征和字典都通过学习获得，对数据有很强的**适应性**
  - 学习的目标是优化一个预测问题，具有很强的**针对性**



多层神经网络


hidden layer 1
hidden layer 2
hidden layer 3
input layer
output layer
数据
特征
特征
特征
结果

分层的特征提取

马云

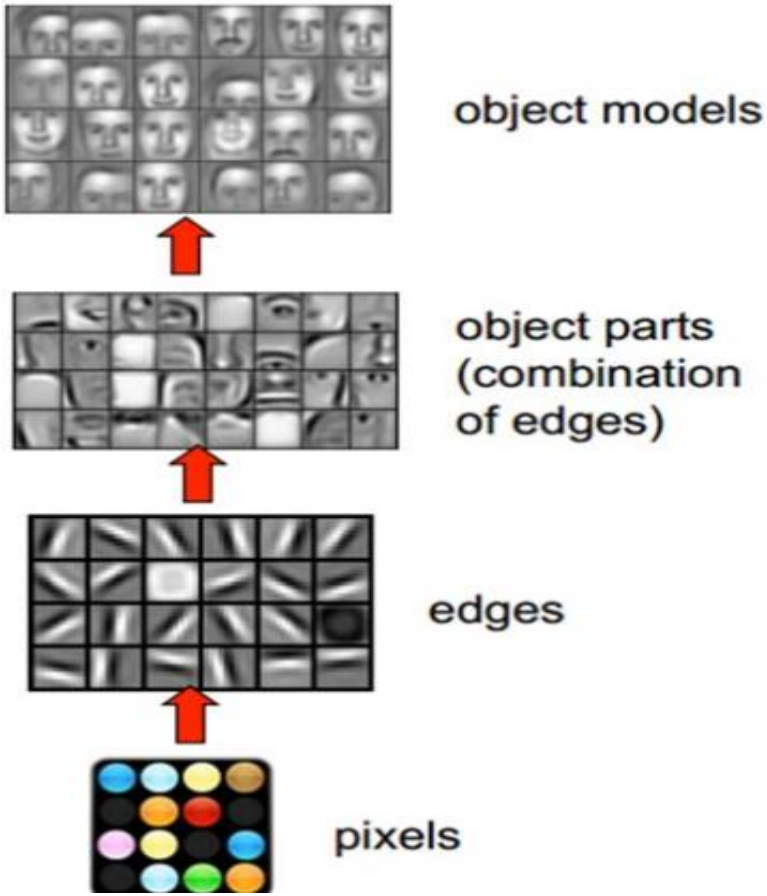
- **输入**数据：像素
- **初层**特征：边缘

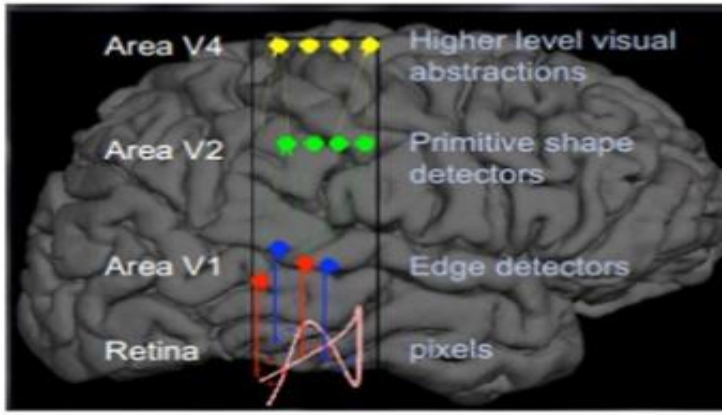
- **中层**特征：形状

- **高层**特征：物体部件

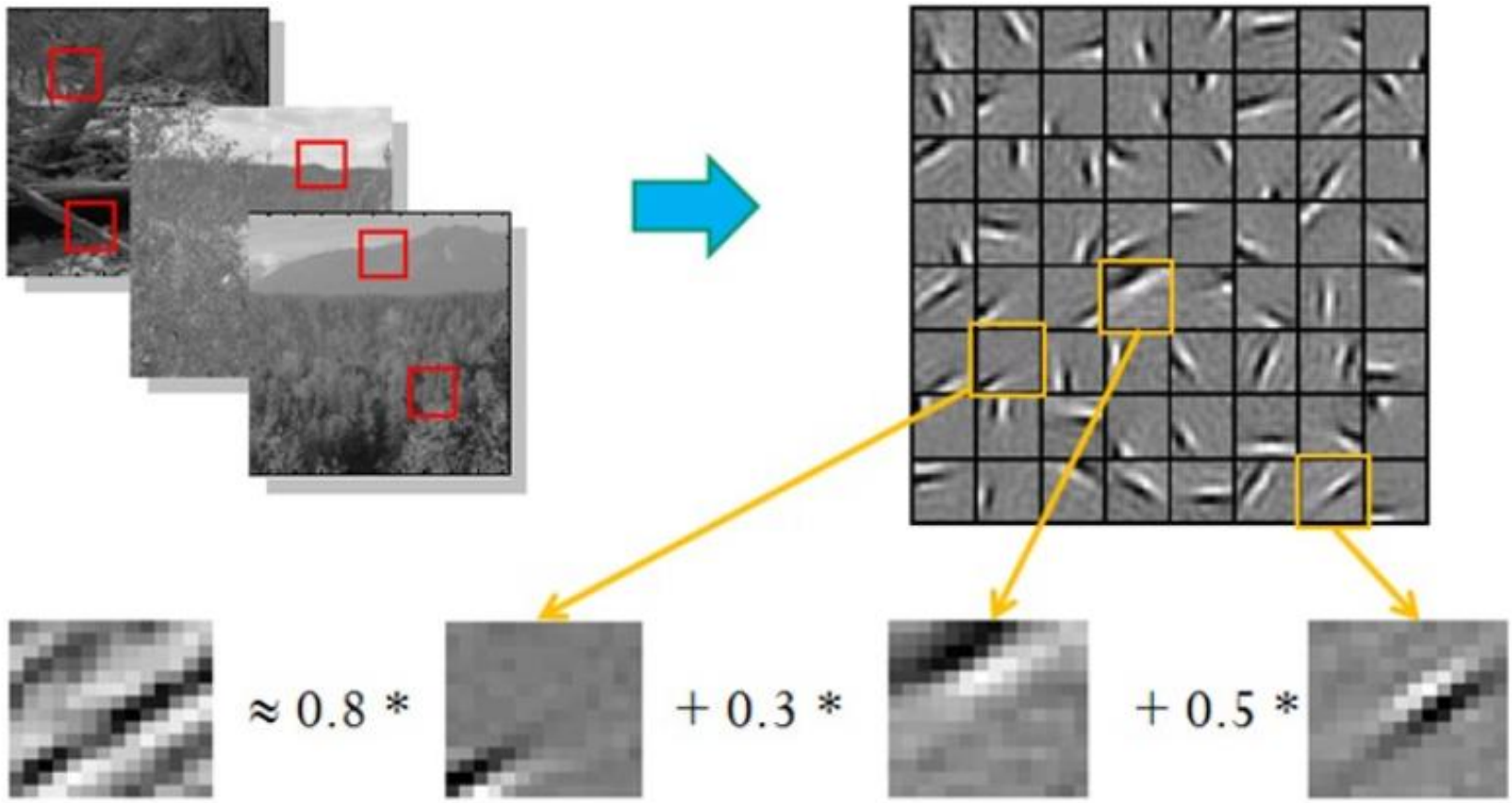
分层的特征提取

- 模仿了人脑的特征抽象过程





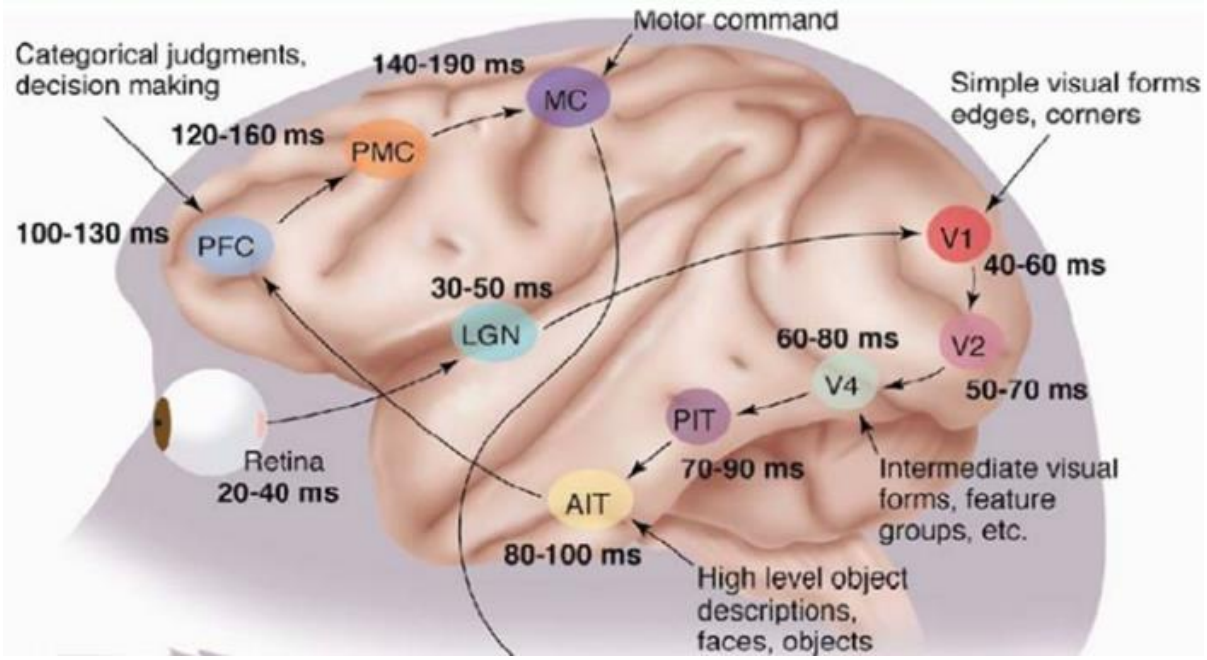
深层特征由浅层特征组合而成



$[a_1, \ldots, a_{64}] = [0, 0, \ldots, 0, \mathbf{0.8}, 0, \ldots, 0, \mathbf{0.3}, 0, \ldots, 0, \mathbf{0.5}, 0]$
(feature representation)

深度学习的优势

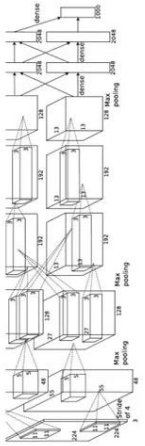
- 人脑视觉机理
  - 人的视觉系统的信息处理是分级的
  - 高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图
  - 抽象层面越高，存在的可能猜测就越少，就越利于分类



“深度”竞赛

http://cs231n.stanford.edu/slides/winter1516_lecture8.pdf

8 layers



16.4%

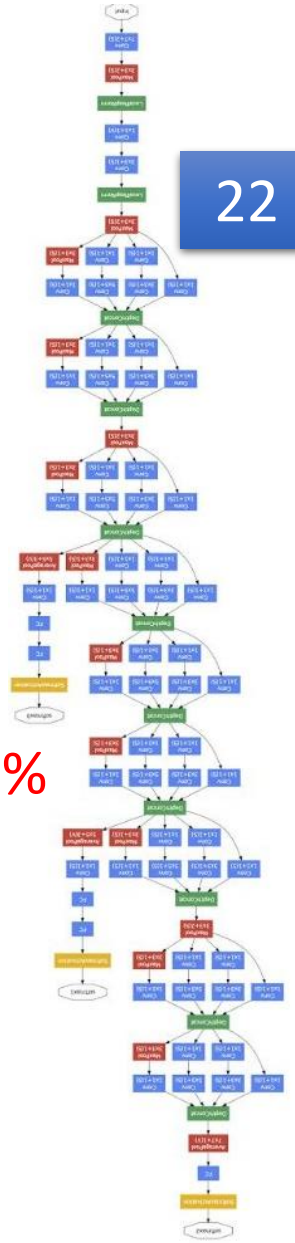
AlexNet (2012)

19 layers

7.3%

VGG (2014)

22 layers

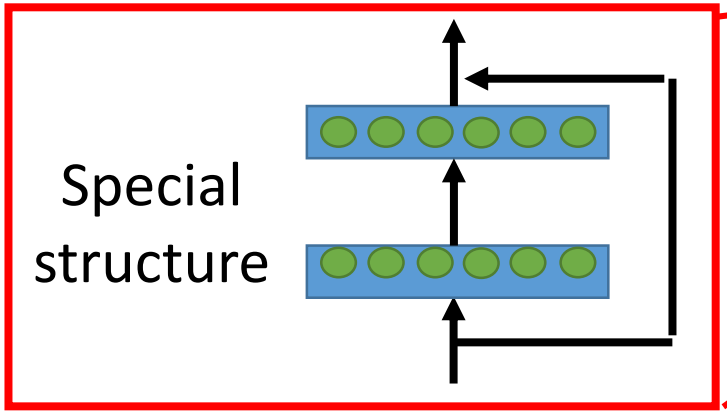


6.7%

GoogleNet (2014)

“深度” 竞赛

目瞪狗呆



152 layers

3.57%

162 layers

16.4%

AlexNet
(2012)

7.3%

VGG
(2014)

6.7%

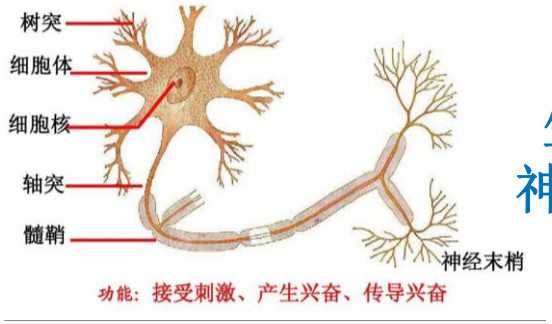
GoogleNet
(2014)

Residual Net
(2015)

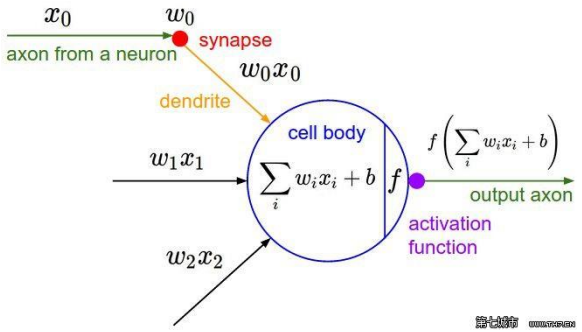
迪拜
哈利法塔

神经网络模型

- 不同动物神经元数量的对比



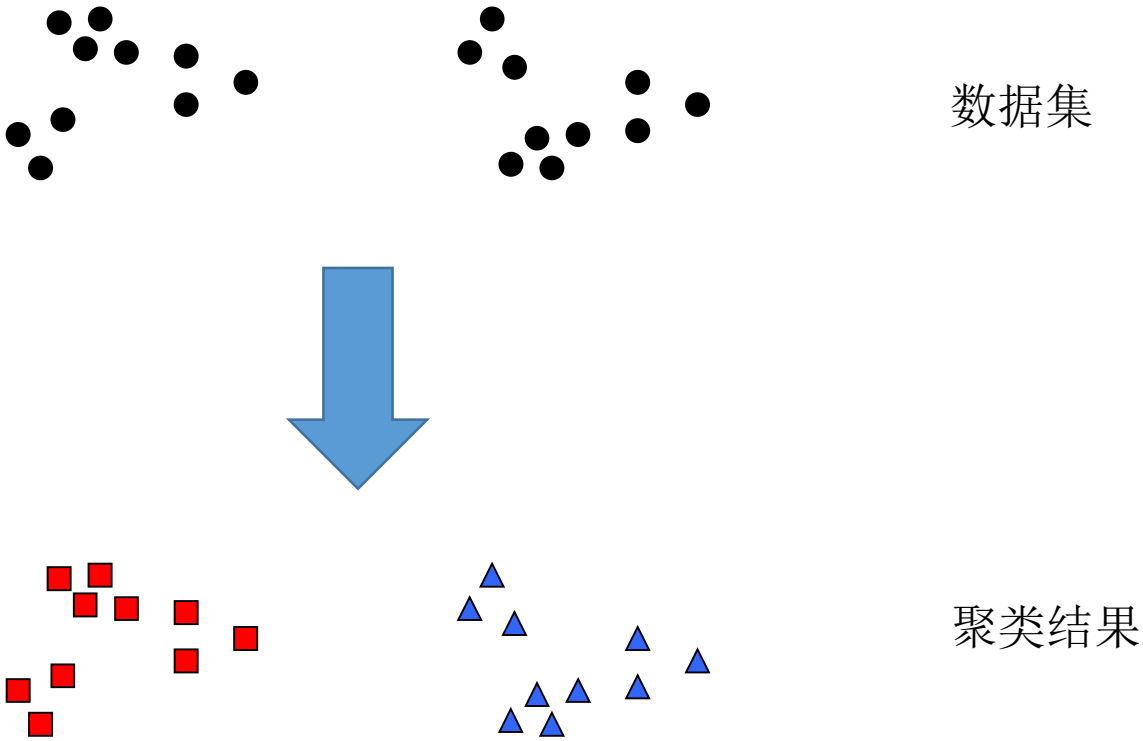
生物
神经元



信息
神经元

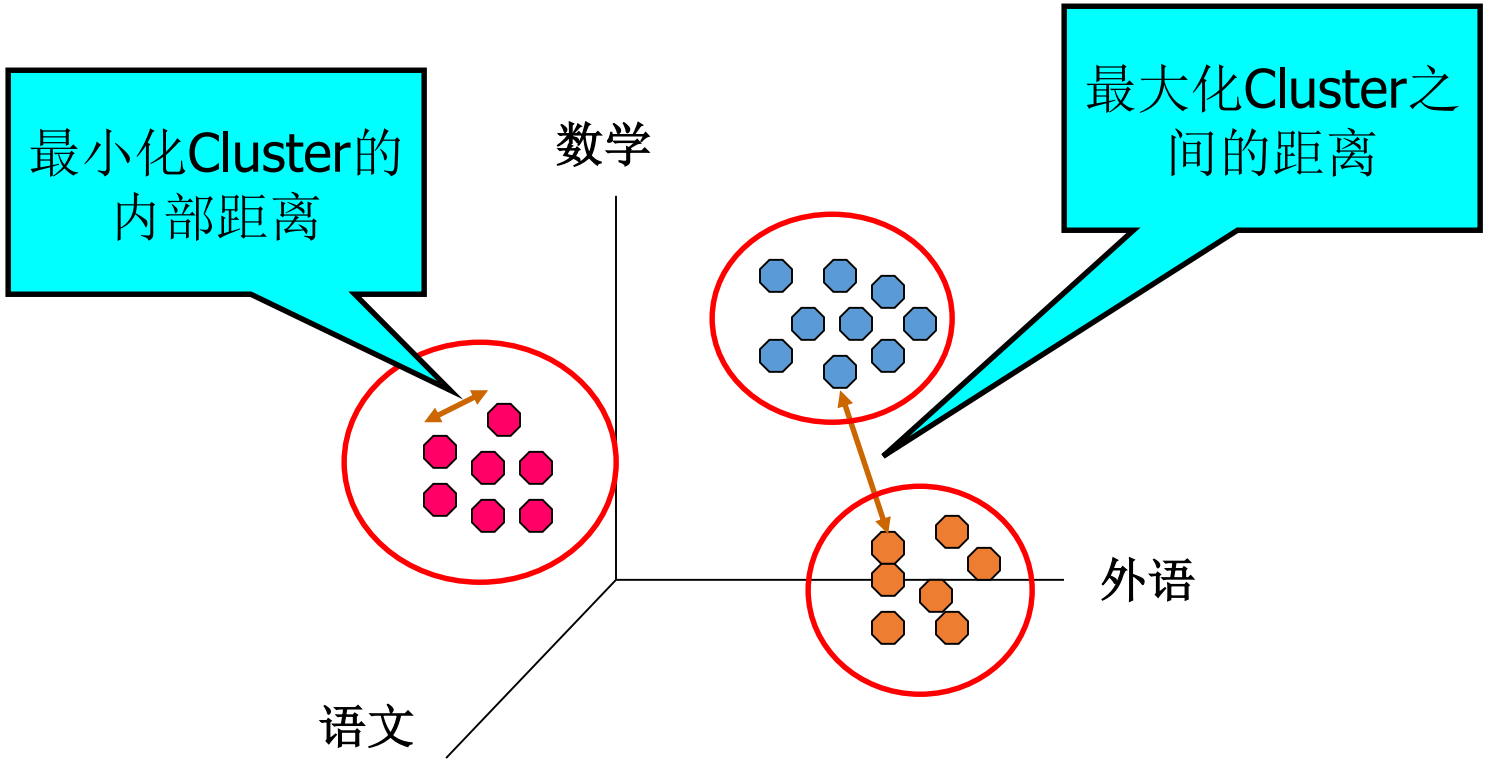
城市科学当中的应用

无监督学习：聚类问题 Clustering



无监督学习：聚类问题 Clustering

- 核心思想
  - 对给定对象集寻找一种分组方式，使得**组内**的各个对象尽可能的相似，**组间**的对象差异尽可能的大。



Motivation

Animal Mobility

Fish School

Sheep Flock

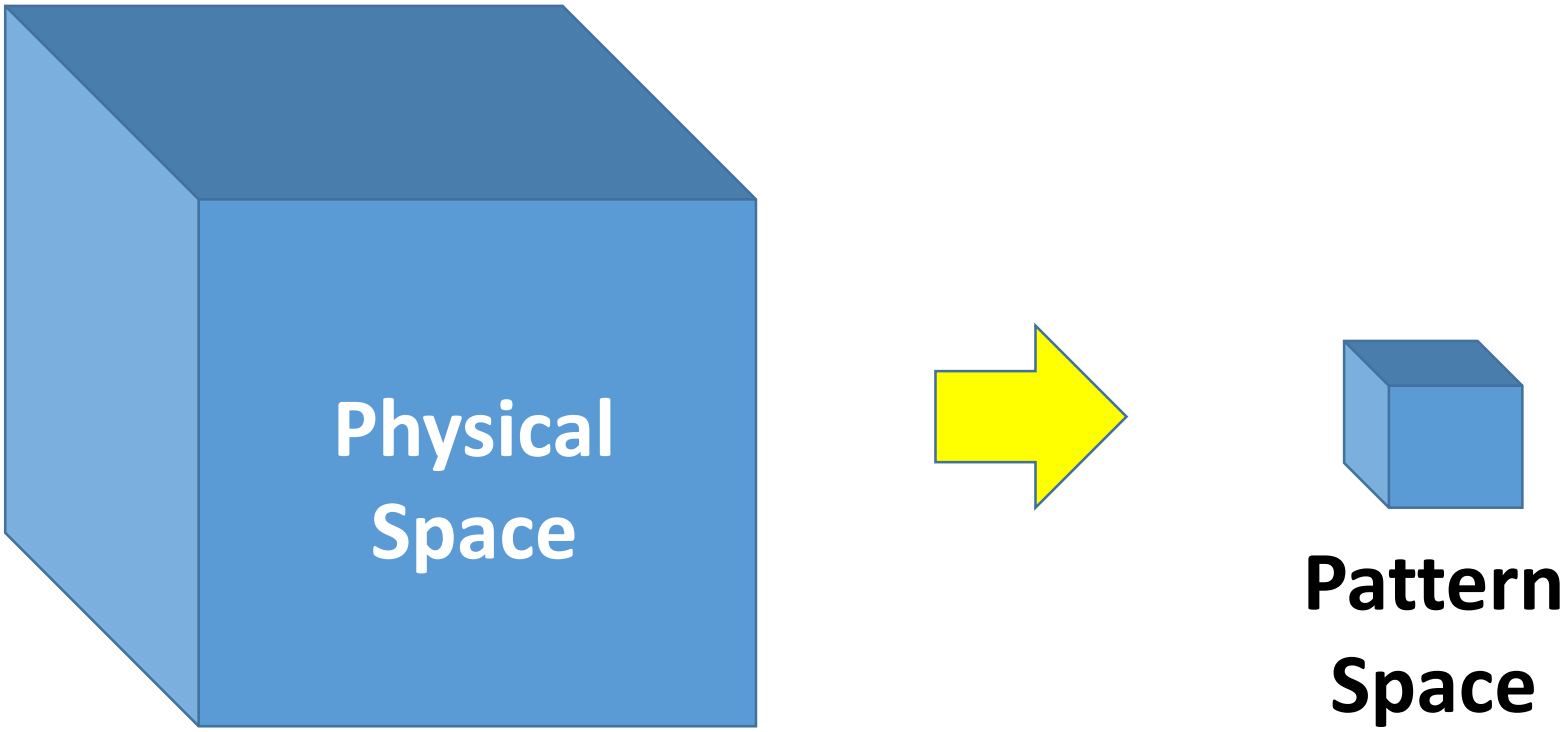
Animal Behavior

Handclap

Traffic Tidal

All contain patterns

Motivation



Our goal

Input:

Urban big data

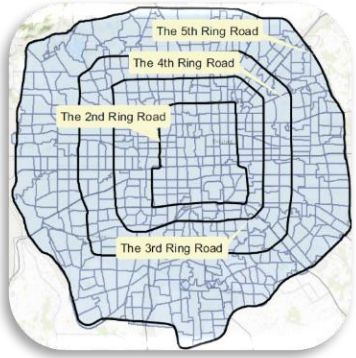
- Output: spatial-temporal patterns
  - Which areas are in the same spatial patterns? (**Community**)
  - What temporal patterns exist in urban traffic? (**Rhythm**)
  - How traffic occurs between spatial patterns with different temporal patterns? (**Relations between community and rhythm**)

Our Works

Map road

Input:

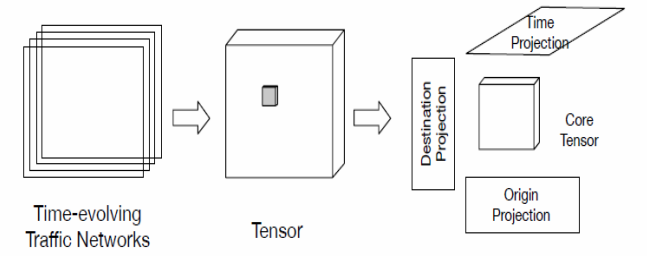
Taxi GPS in Beijing



Beijing Map

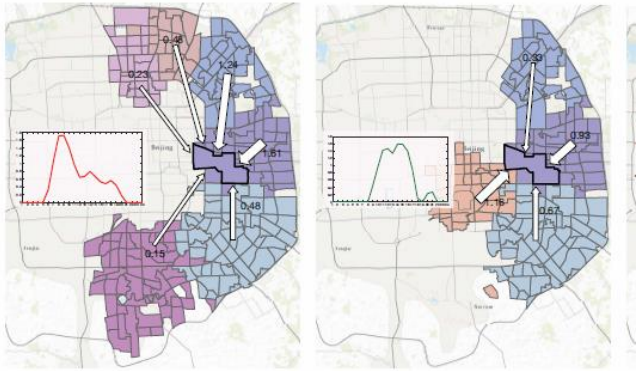
Beijing POI

Tools:



Regularized Non-negative Tensor Decomposition

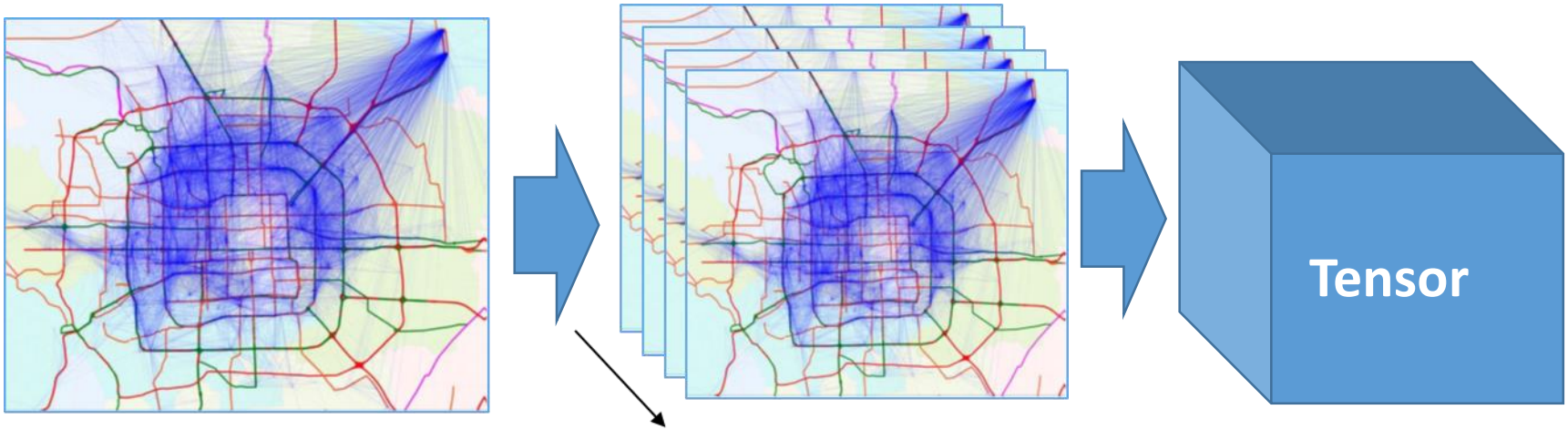
Output:



Spatio-temporal patterns

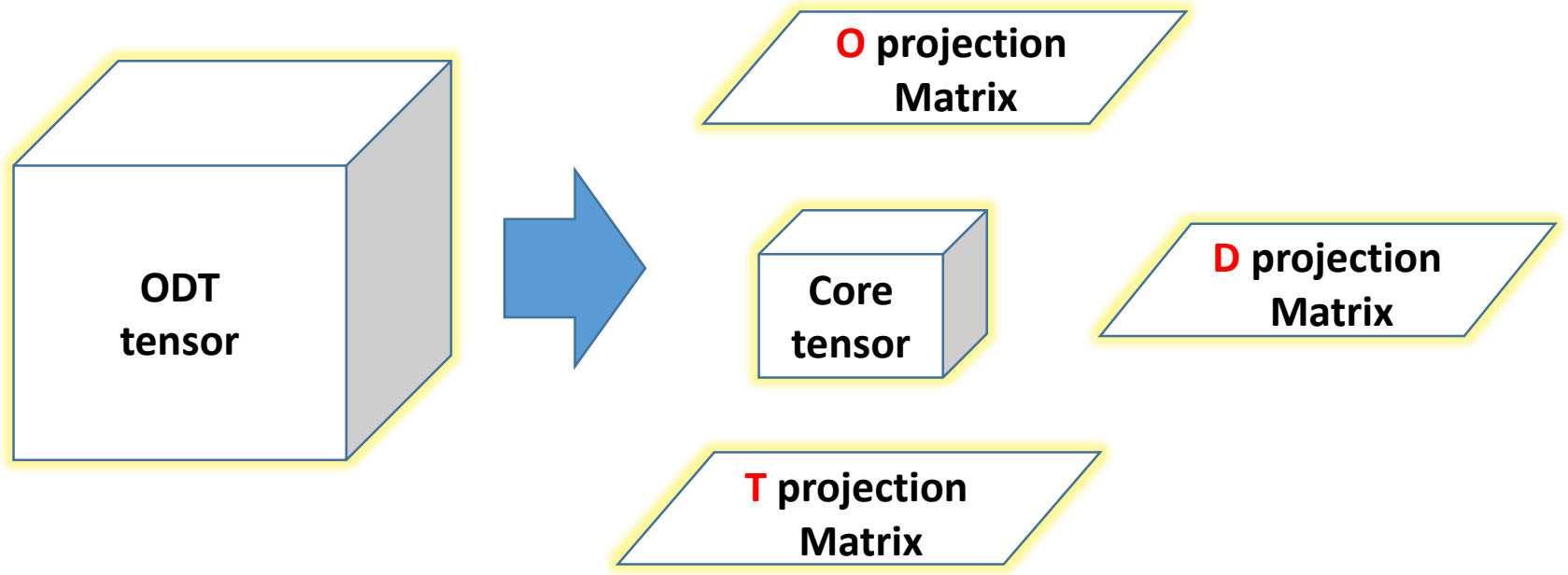
Problem definition

- **O**rigin-**D**estination-**T**ime tensor
  - (*i*, *j*, *k*)-th element: the traffic volume from *i*-th origin zone to *j*-th destination zone in *k*-th time.

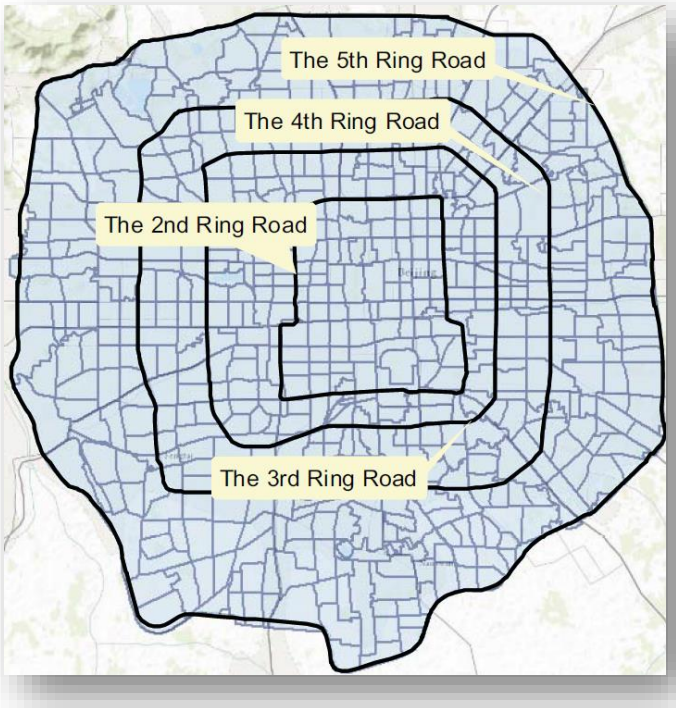


Problem definition

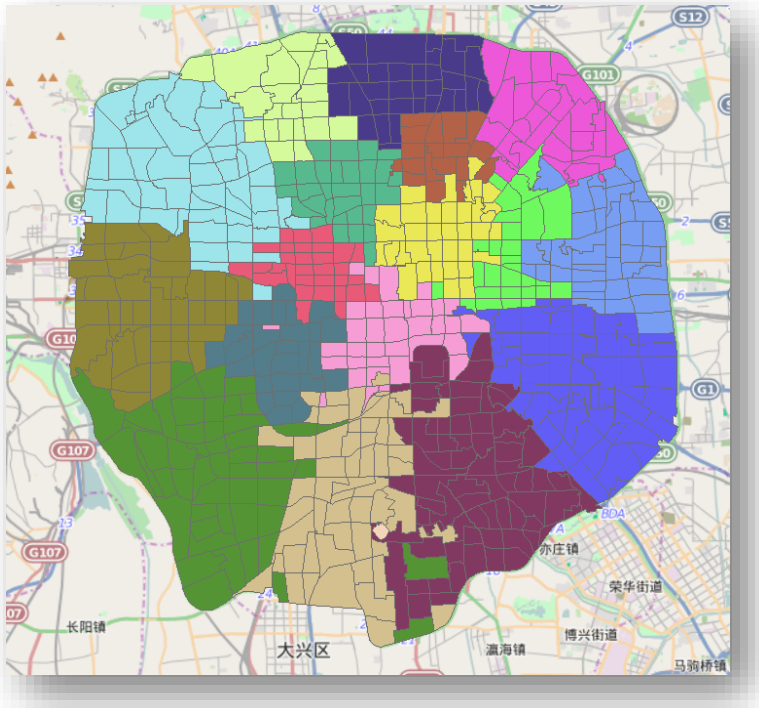
- Decompose the tensor as three projection matrixes and a core tensor



Spatial pattern projection

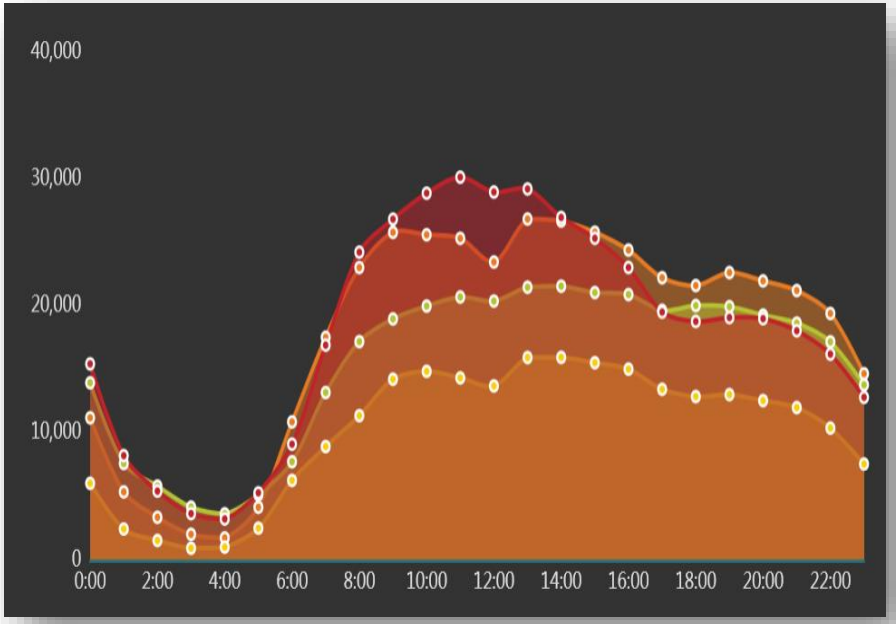


Urban zones

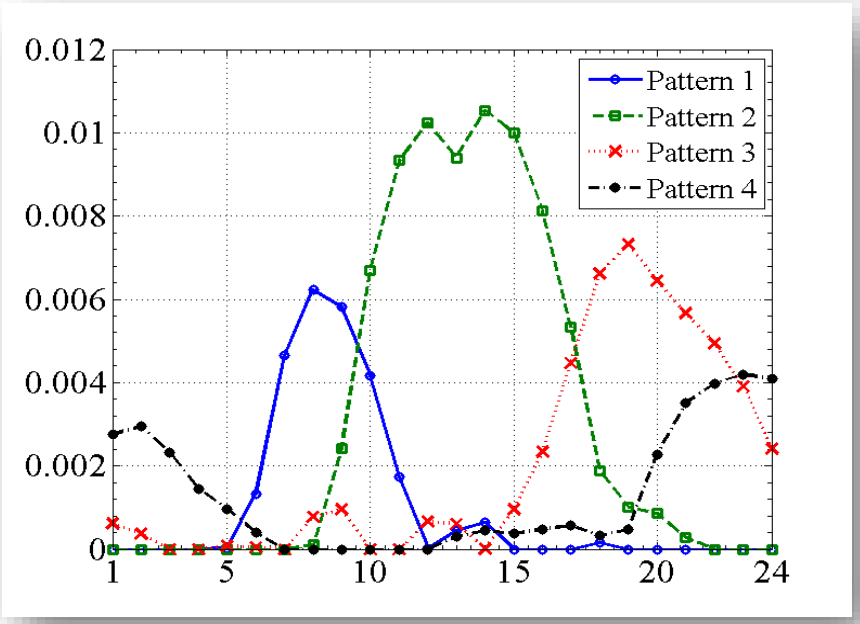


Urban communities

Temporal pattern projection



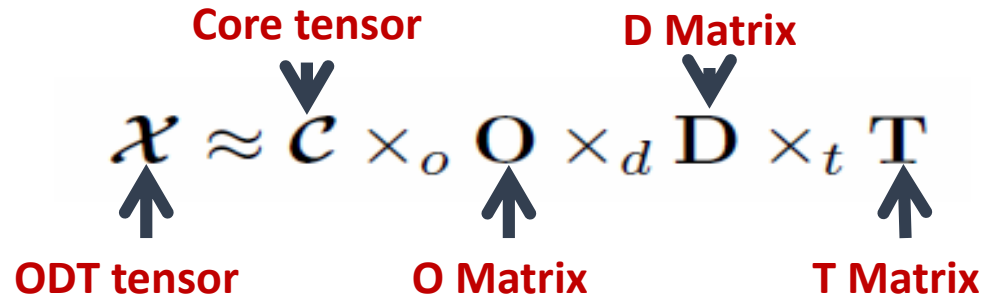
Urban traffic



Urban rhythm

Regularized Non-negative Tucker Decomposition

- Basic Tucker decomposition



  - **Objective function (1)**

$$\min \quad \mathcal{J}_1 = \|\mathcal{X} - \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2$$

- Challenge
  - Values in the tensor is very sparse (only 8% non-zero elements)
  - Urban traffic patterns have close relations with urban context, such as POI.

Regularized Non-negative Tucker Decomposition

- Model urban contextual information
  - Assumption: areas with similar POI information should have the similar urban structure
  - Construct area-area similarity matrix

$$W_{ij} = \frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|}$$

- **Objective fu**

min

$$\mathcal{J}_2 = \|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2$$

$$\mathcal{J}_3 = \|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2$$

id	POI category	id	POI category
1	food & beverage Service	8	education and culture
2	hotel	9	business building
3	scenic spot	10	residence
4	finance & insurance	11	living service
5	corporate business	12	sports & entertainments
6	shopping service	13	medical care
7	transportation facilities	14	government agencies

Regularized Non-negative Tucker Decomposition

- Make patterns more explainable
  - Non-negative constraints
  - L1 regul
    - Core tensor: Only keep strong interactions
    - Projection matrix: make each cluster more meaningful and enhance uniqueness

$$s.t. \quad \mathcal{C} \geq 0, \mathbf{O} \geq 0, \mathbf{D} \geq 0, \mathbf{T} \geq 0$$

$$+ \gamma \|\mathcal{C}\|_1 + \delta \|\mathbf{O}\|_1 + \epsilon \|\mathbf{D}\|_1 + \varepsilon \|\mathbf{T}\|_1$$

Regularized Non-negative Tucker Decomposition

- Final objection function

$$\mathcal{J} = \|\mathcal{X} - \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2$$

Model traffic information

$$+ \alpha\|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 + \beta\|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2$$

Model urban contextual information (POI)

$$+ \gamma\|\mathcal{C}\|_1 + \delta\|\mathbf{O}\|_1 + \epsilon\|\mathbf{D}\|_1 + \varepsilon\|\mathbf{T}\|_1$$

Make results sparse

$$s.t. \quad \mathcal{C} \geq 0, \mathbf{O} \geq 0, \mathbf{D} \geq 0, \mathbf{T} \geq 0$$

Make results non-negative

- Human mobility and urban context are jointly optimized !

Optimization

- Block Coordinate Descent (BCD)
- Alternating Proximal Gradient (APG)

$$\frac{\partial \mathcal{J}}{\partial \mathcal{C}} = 2\left(\mathcal{C} \times_o (\mathbf{O}^\top \mathbf{O}) \times_d (\mathbf{D}^\top \mathbf{D}) \times_t (\mathbf{T}^\top \mathbf{T}) - \mathcal{X} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right)$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{O}} = 2\Big(\mathbf{O}\left(\mathcal{C} \times_d (\mathbf{D}^\top \mathbf{D}) \times_t (\mathbf{T}^\top \mathbf{T})\right)_{(o)} \mathcal{C}_{(o)}^\top - \left(\mathcal{X} \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right)_{(o)} \mathcal{C}_{(o)}^\top - \alpha\left(\mathbf{W} - \mathbf{O}\mathbf{O}^\top\right)\mathbf{O}\Big)$$

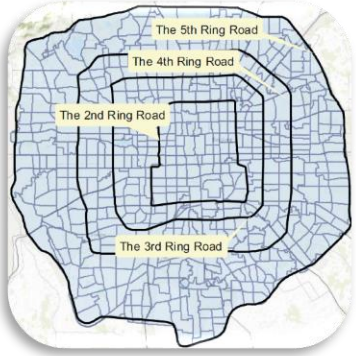
$$\frac{\partial \mathcal{J}}{\partial \mathbf{D}} = 2\Big(\mathbf{D}\left(\mathcal{C} \times_o (\mathbf{O}^\top \mathbf{O}) \times_t (\mathbf{T}^\top \mathbf{T})\right)_{(d)} \mathcal{C}_{(d)}^\top - \left(\mathcal{X} \times_o \mathbf{O}^\top \times_t \mathbf{T}^\top\right)_{(d)} \mathcal{C}_{(d)}^\top - \beta\left(\mathbf{W} - \mathbf{D}\mathbf{D}^\top\right)\mathbf{D}\Big)$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{T}} = 2\left(\mathbf{T}\left(\mathcal{C} \times_o (\mathbf{O}^\top \mathbf{O}) \times_d (\mathbf{D}^\top \mathbf{D})\right)_{(t)} \mathcal{C}_{(t)}^\top - \left(\mathcal{X} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top\right)_{(t)} \mathcal{C}_{(t)}^\top\right)$$

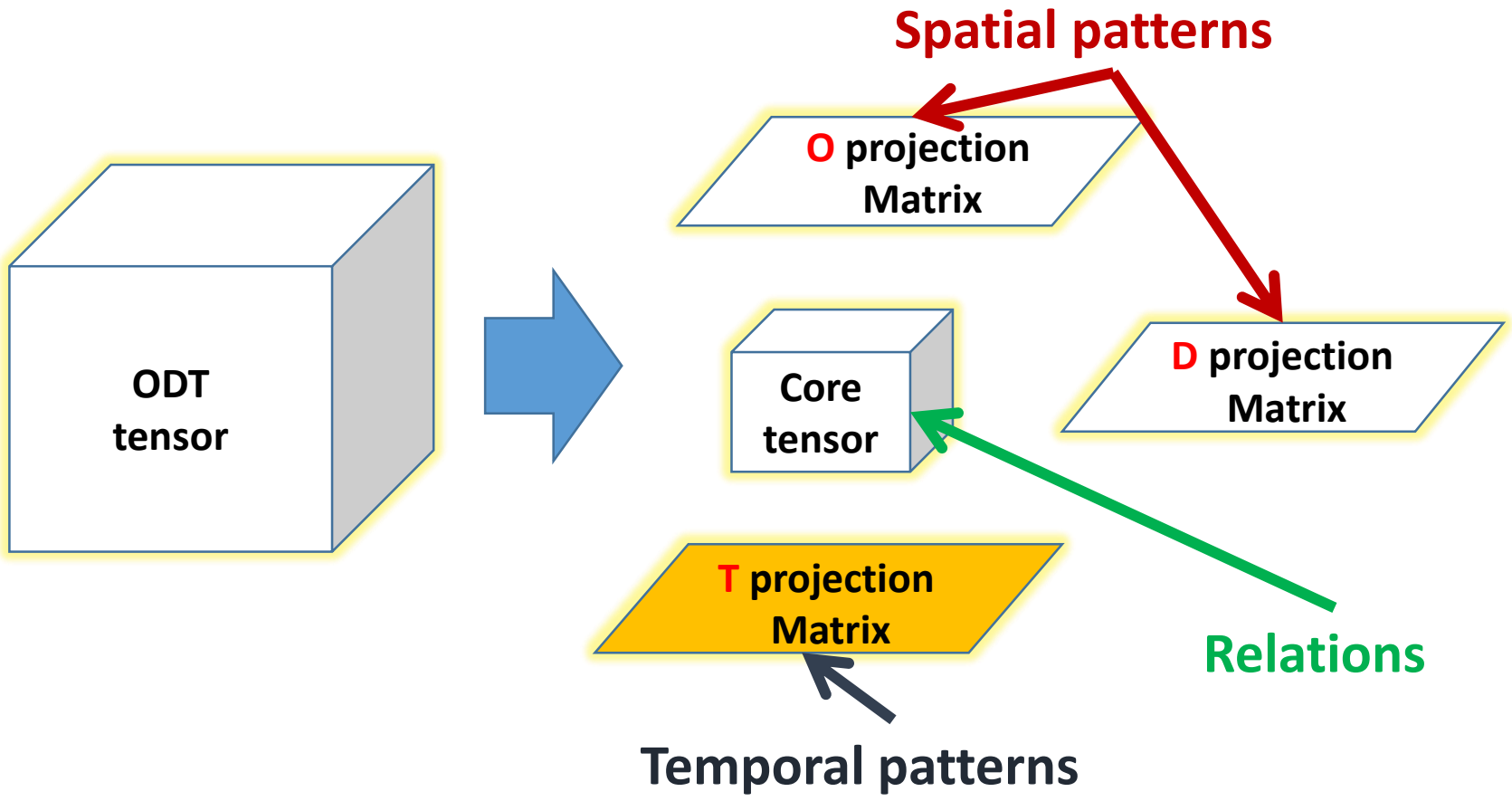
Experiments

Data

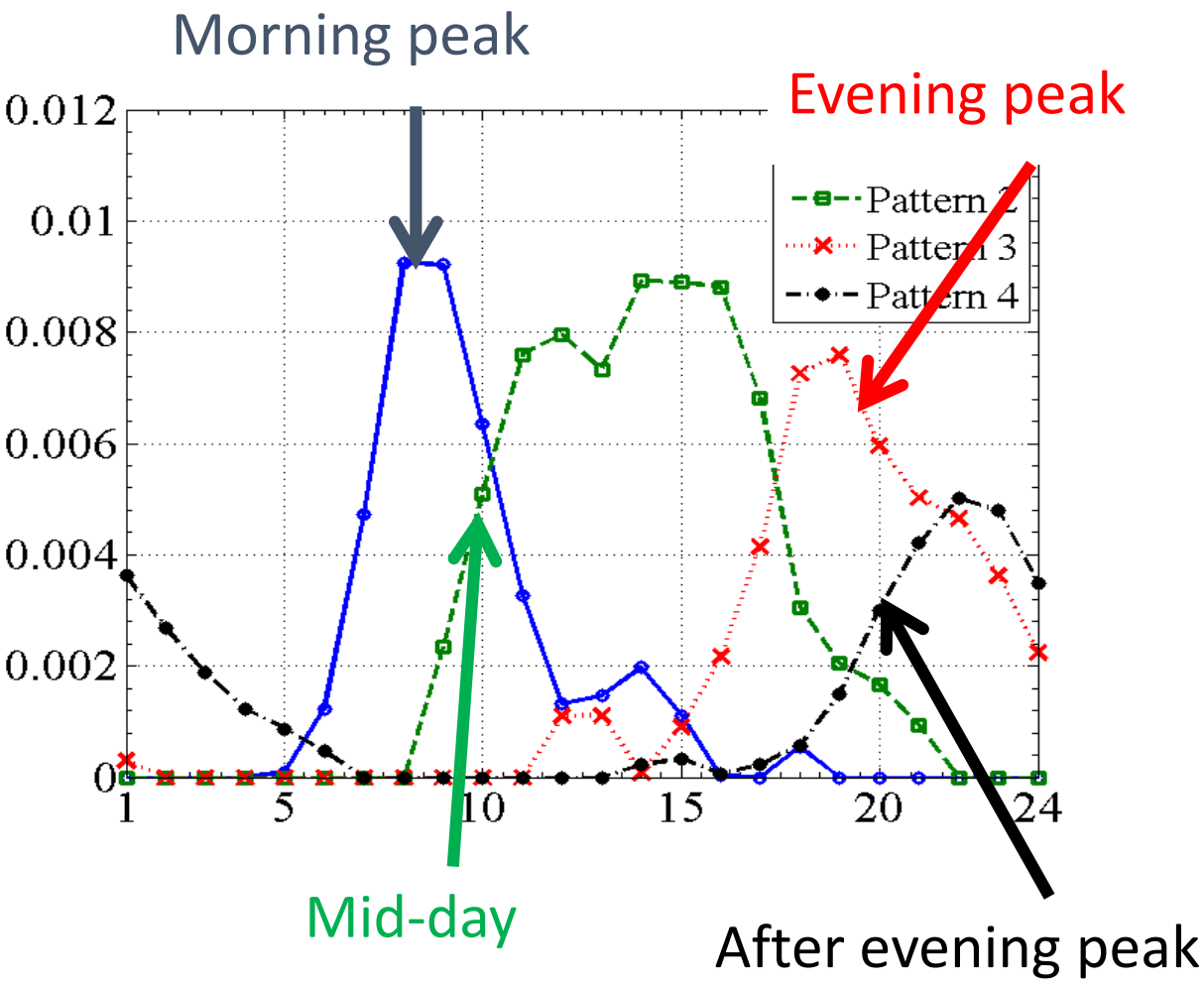
- Beijing GPS data
  - GPS trajectory of 20000 Beijing taxies, 2008 and 2012.
- Beijing traffic analysis zones map
  - 600 zones in the 5th ring road in Beijing
- Position of Interesting (POI) in Beijing
  - 380,000 points



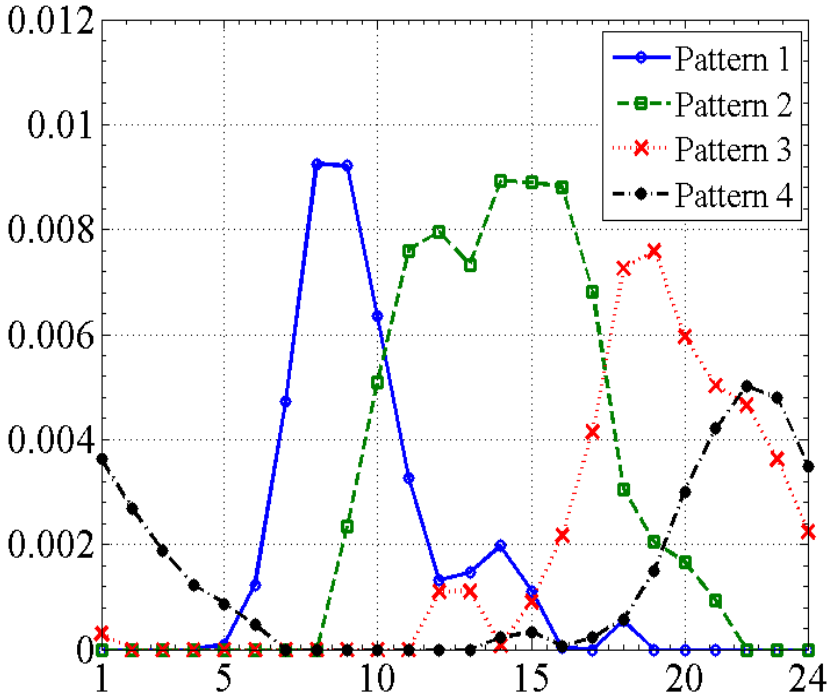
Beijing show case



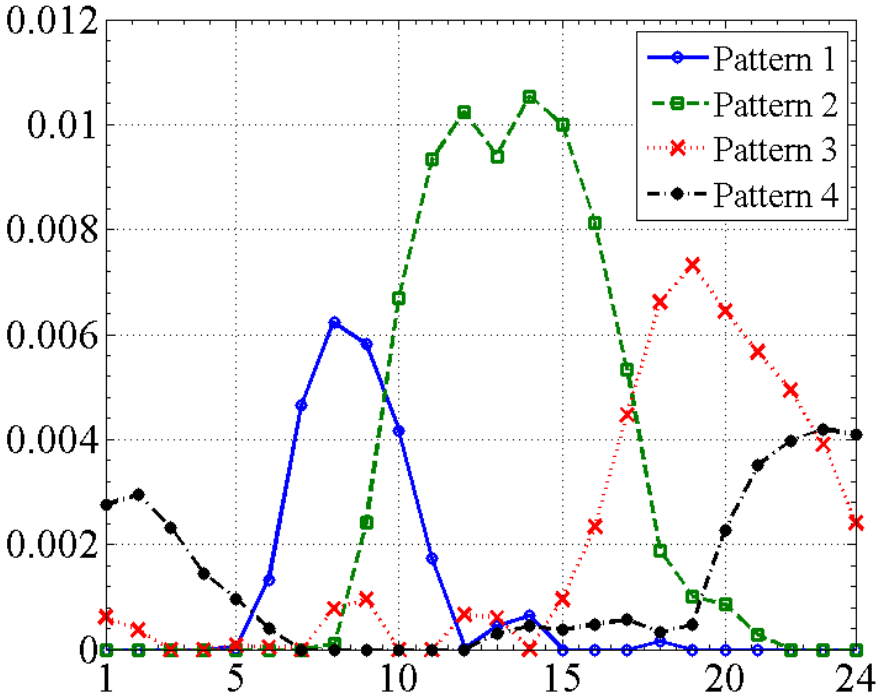
Temporal patterns



Temporal patterns

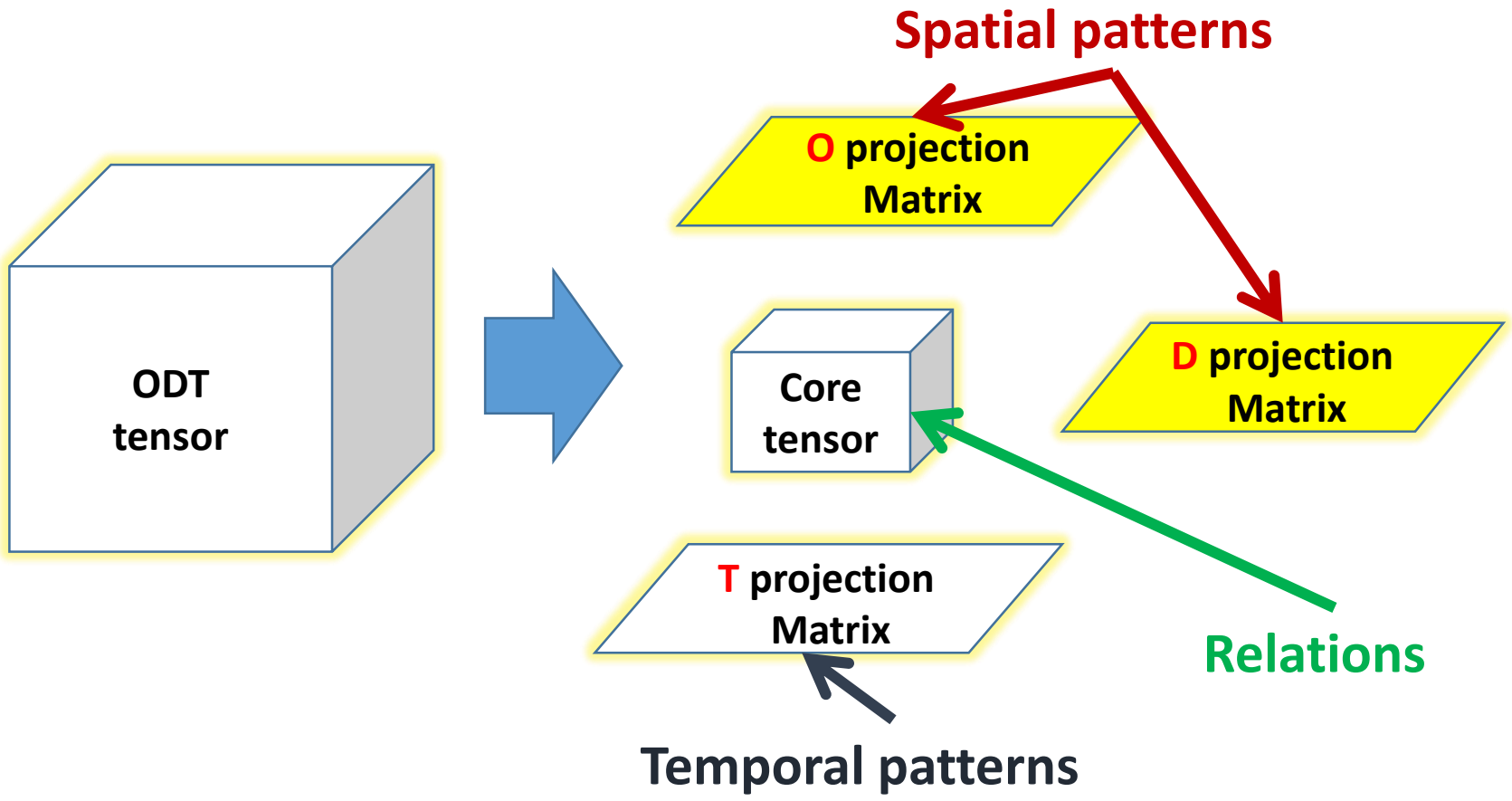


2008

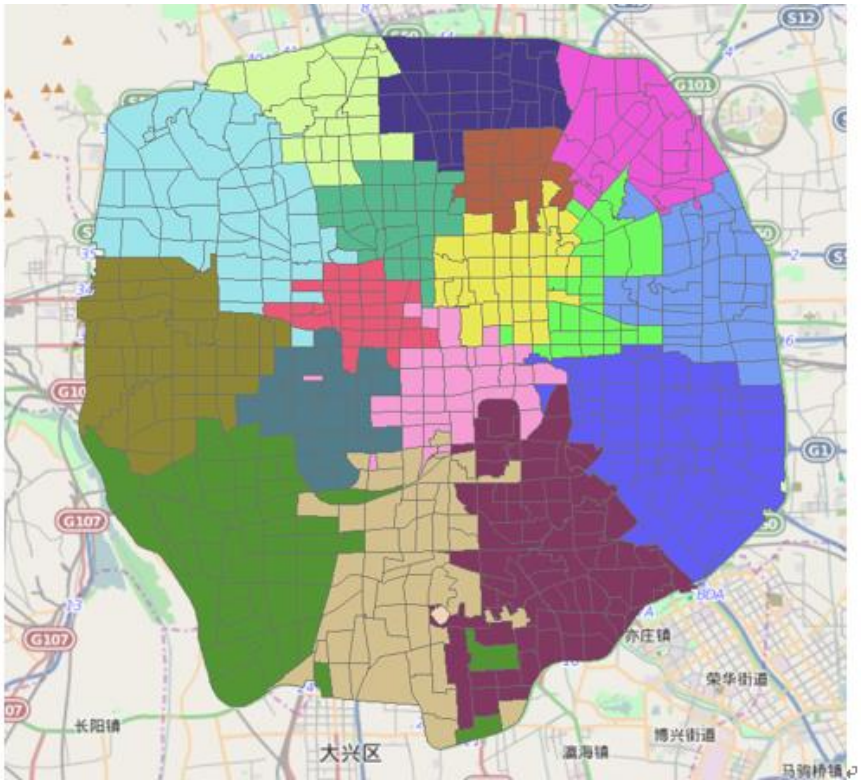


2012

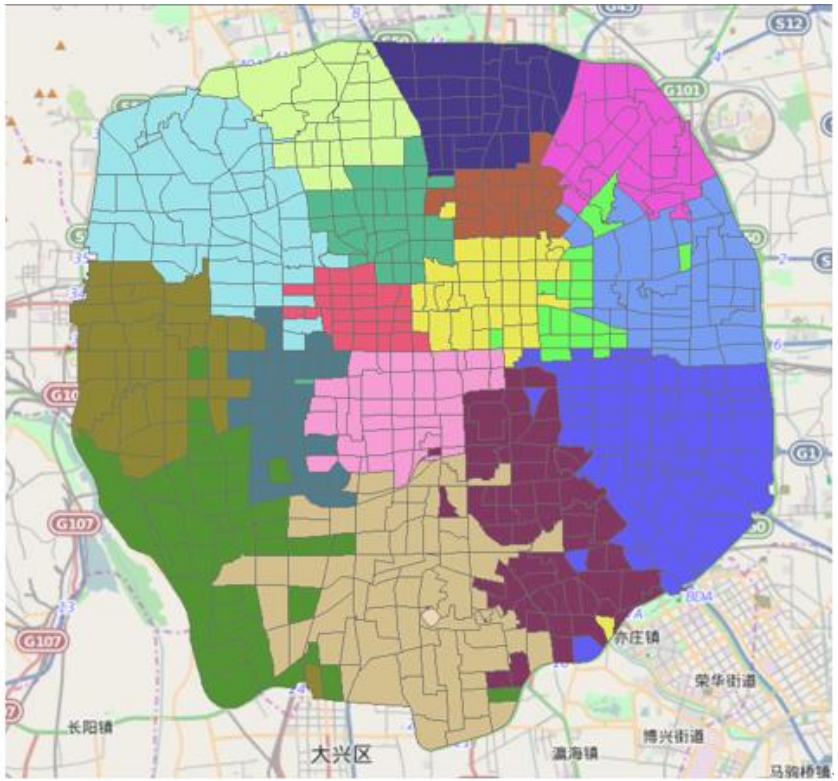
Beijing show case



Spatial patterns

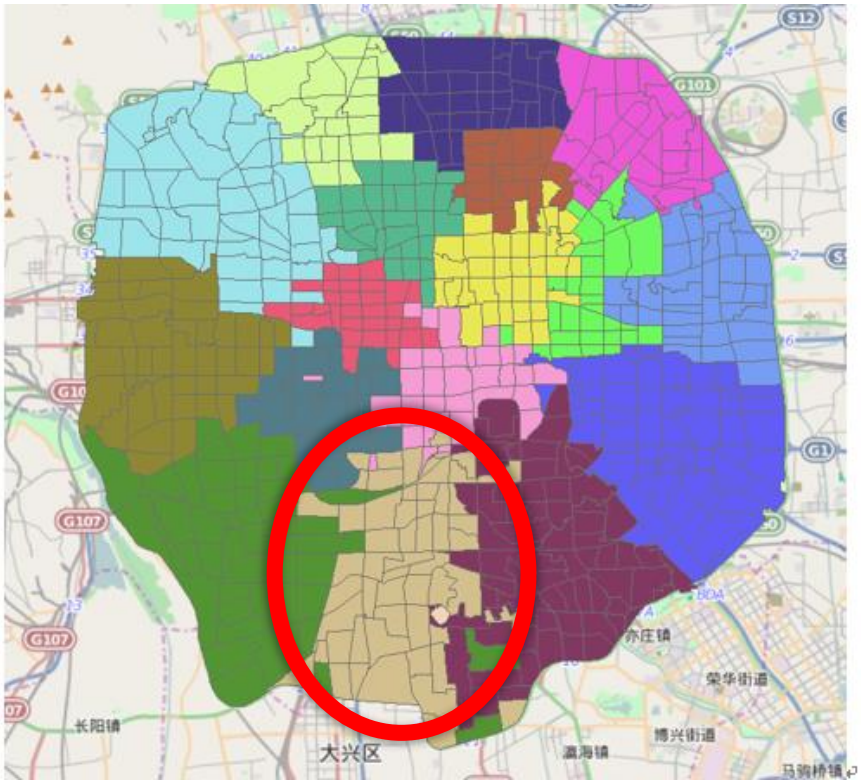


2008

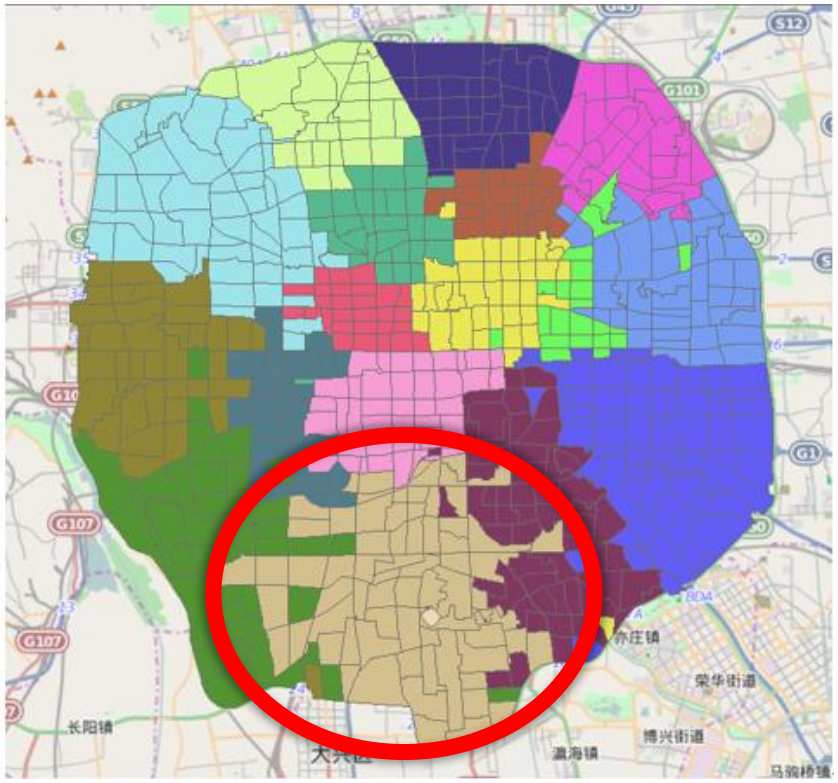


2012

Spatial patterns

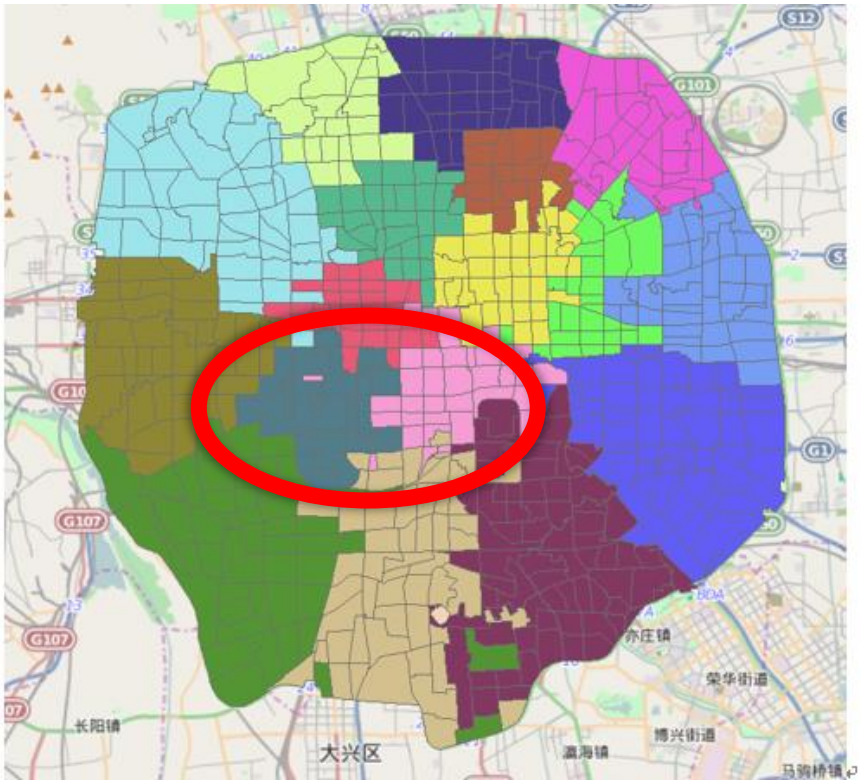


2008

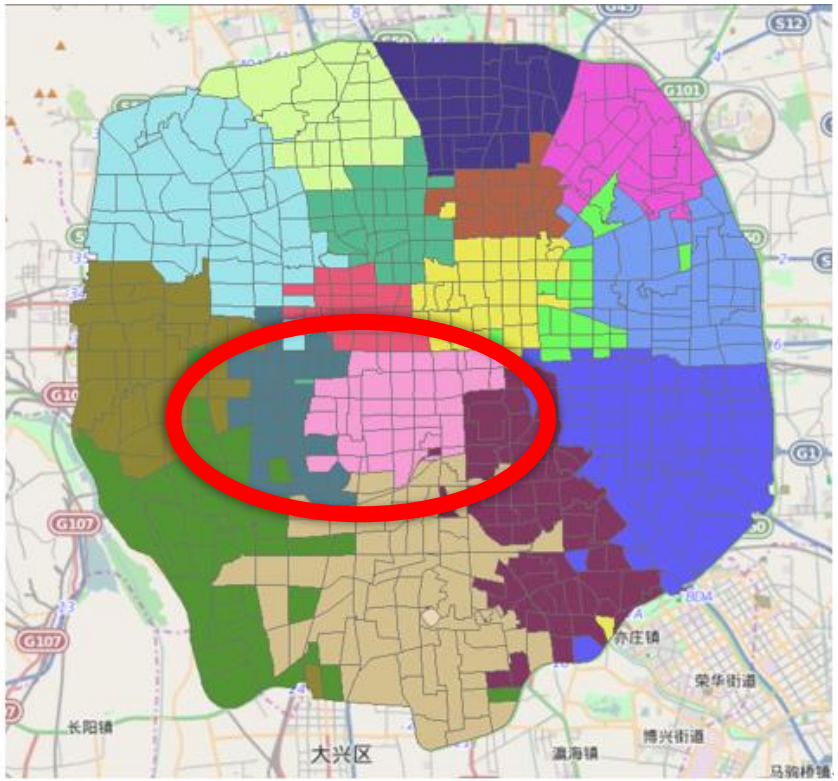


2012

Spatial patterns

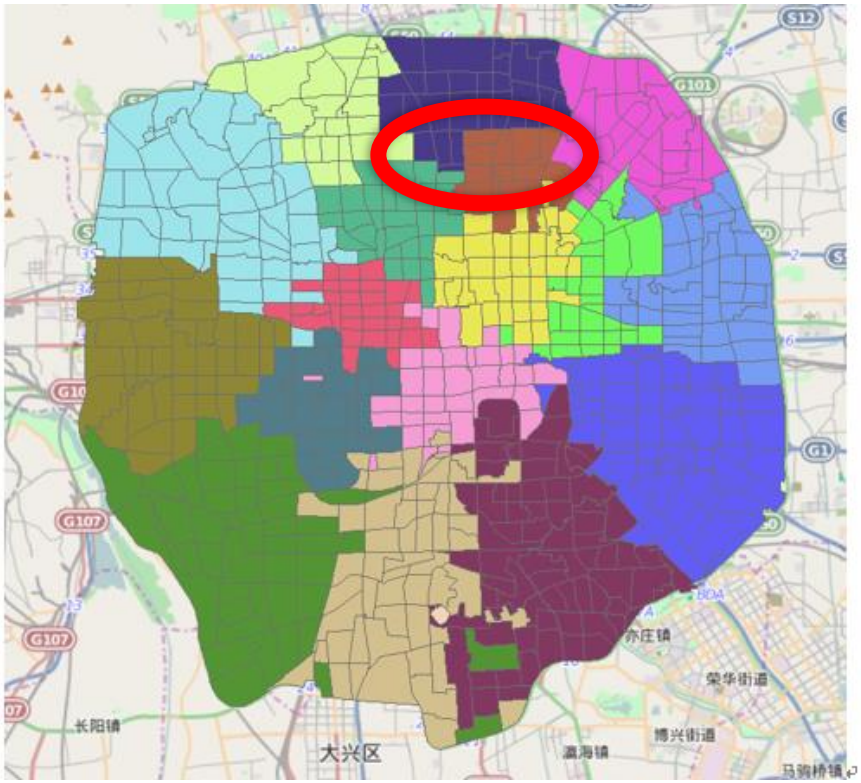


2008

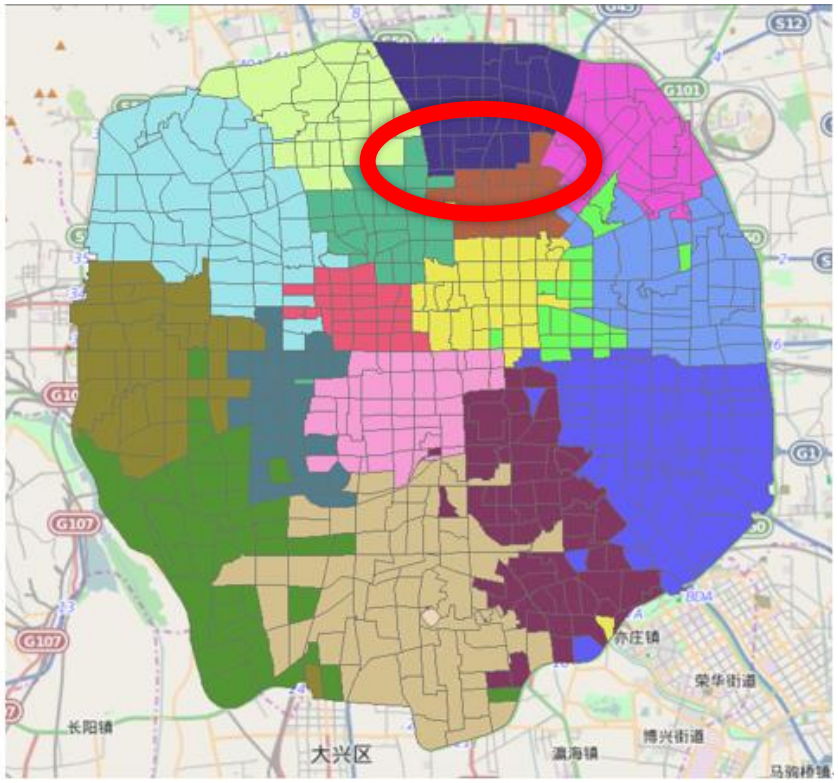


2012

Spatial patterns

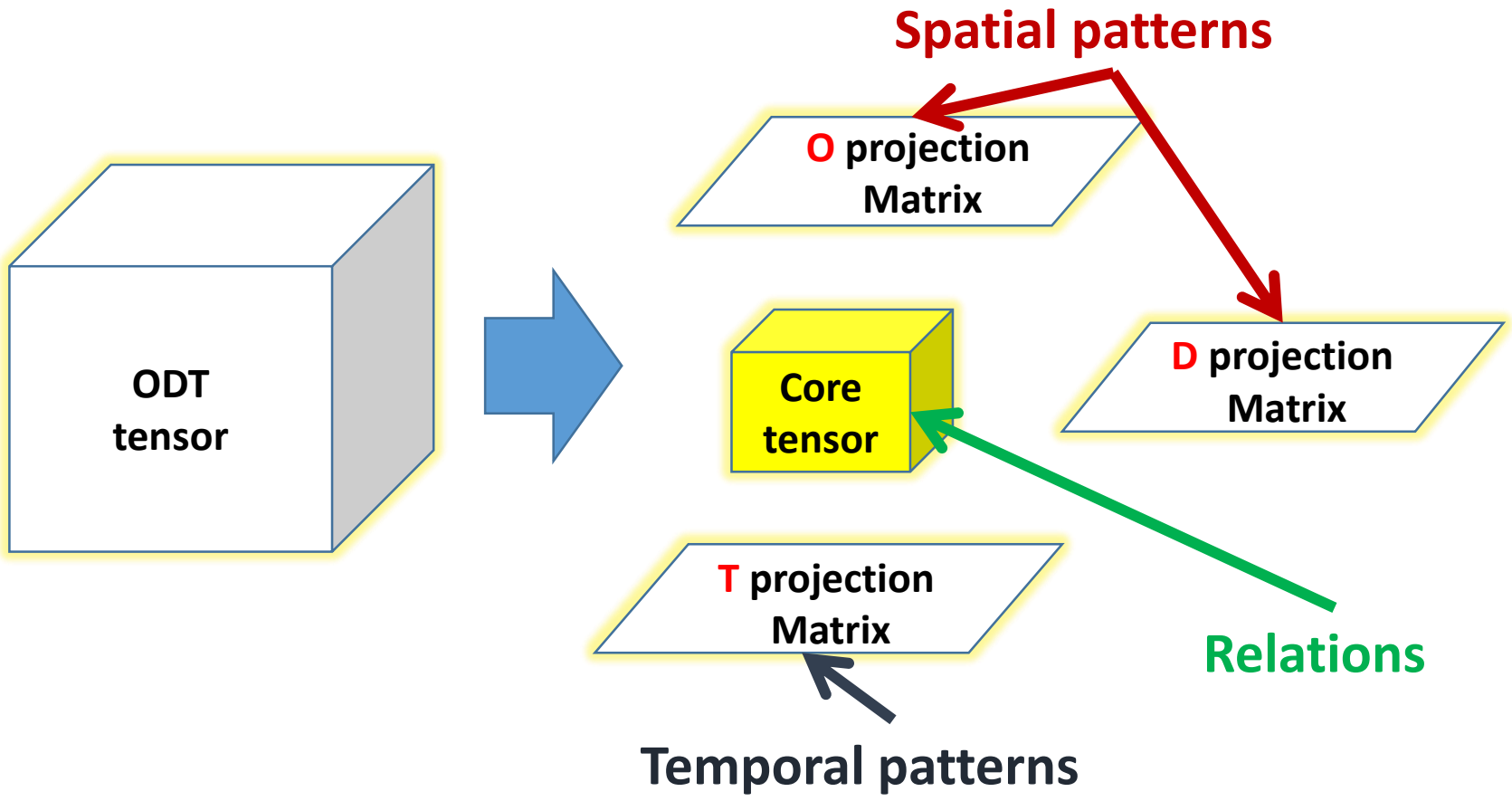


2008

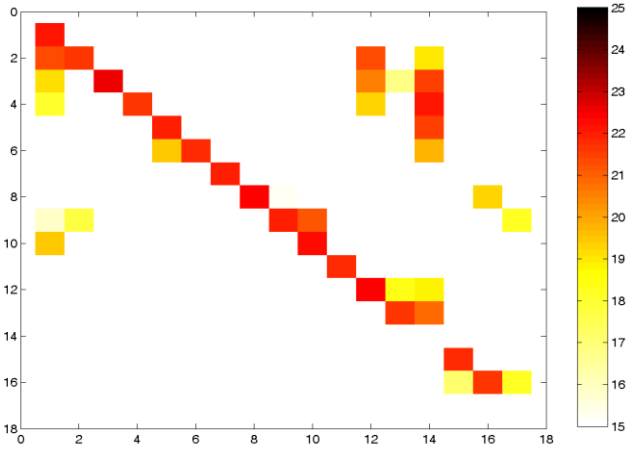


2012

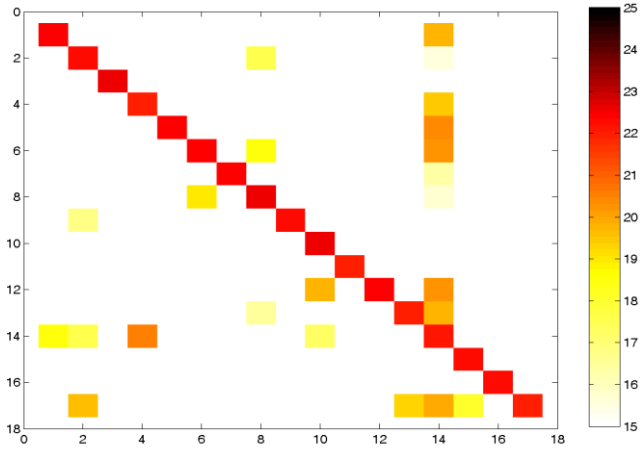
Beijing show case



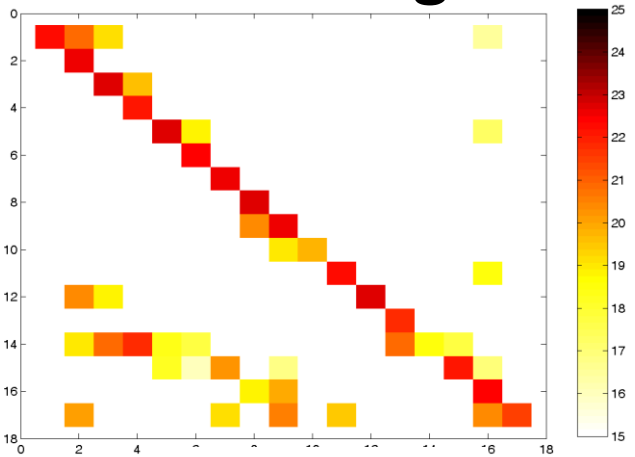
Core tensor



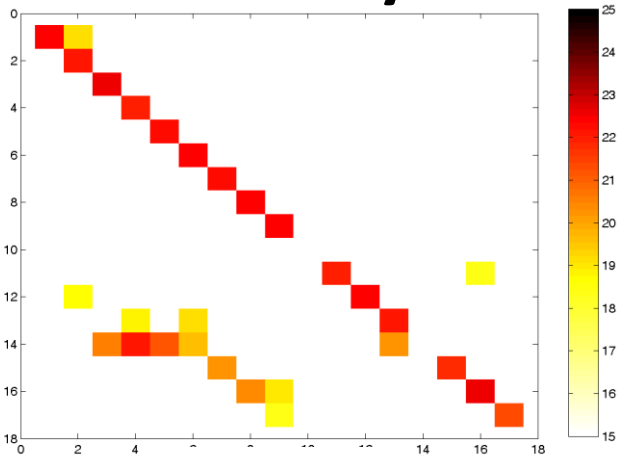
Morning



Midday

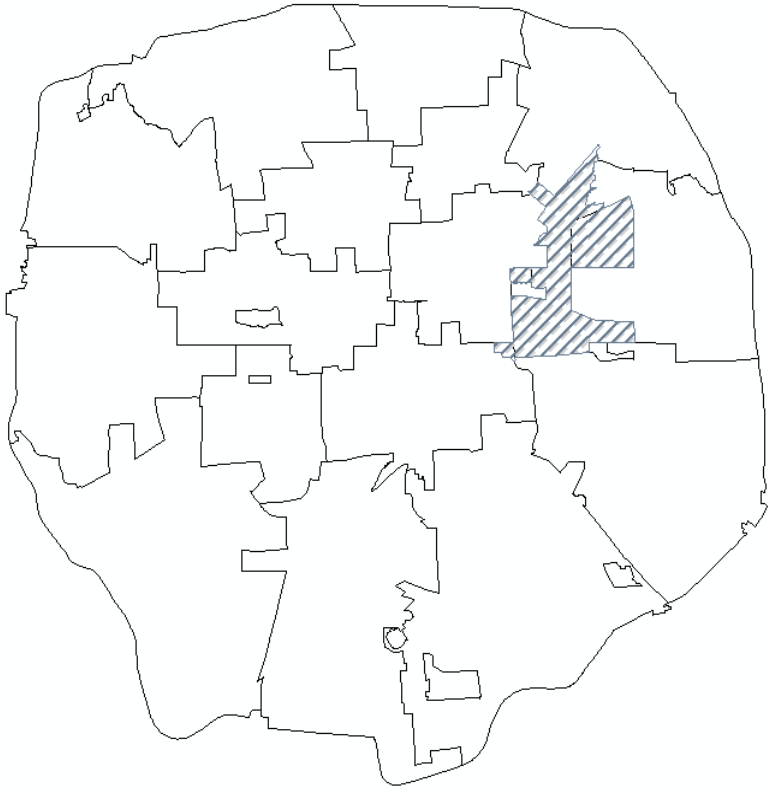


Evening

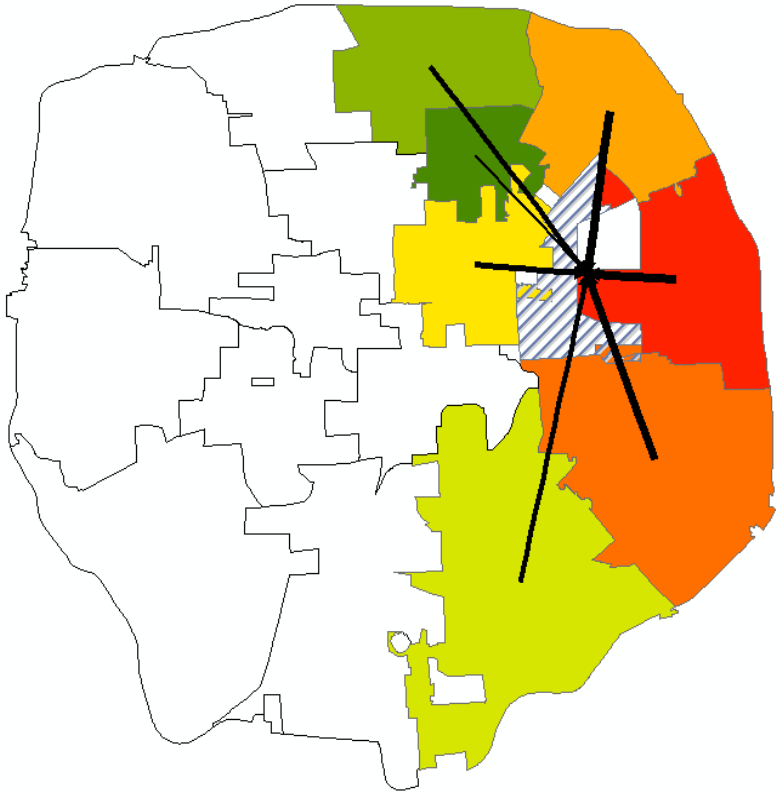


Night

Morning

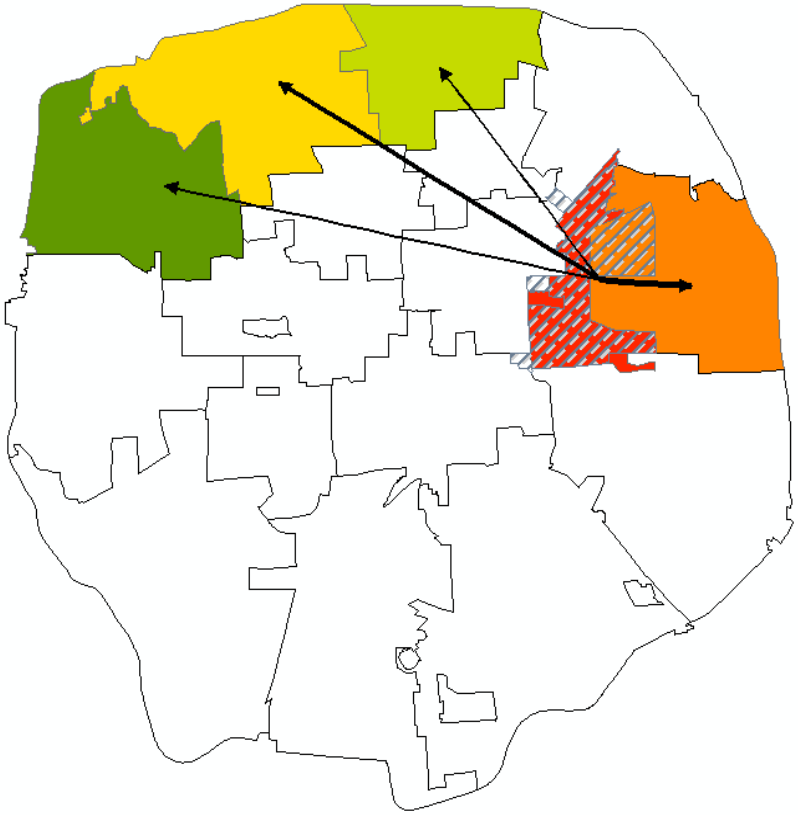


Origin

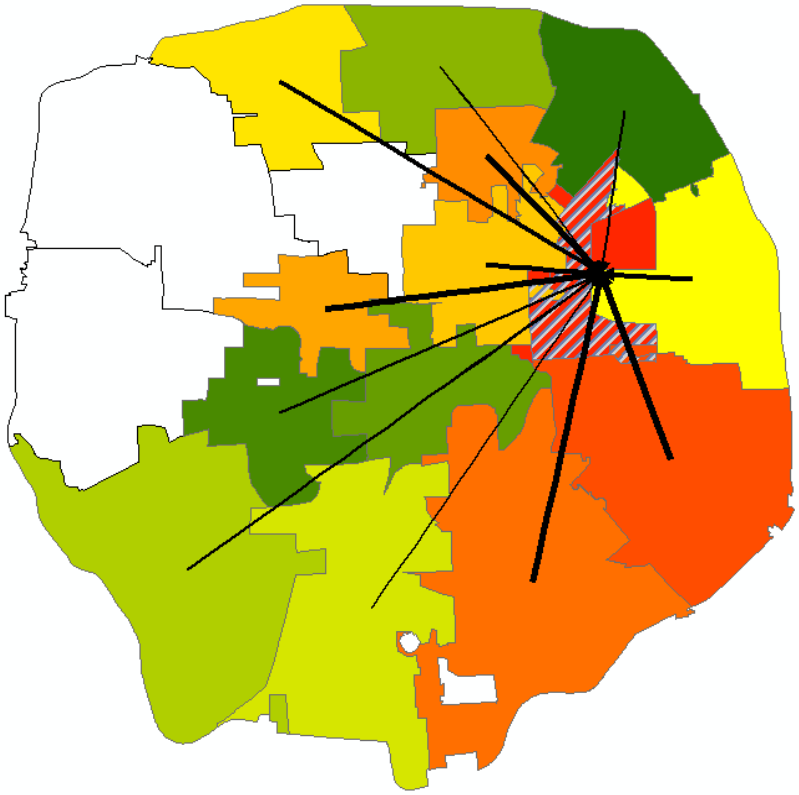


Destination

Midday

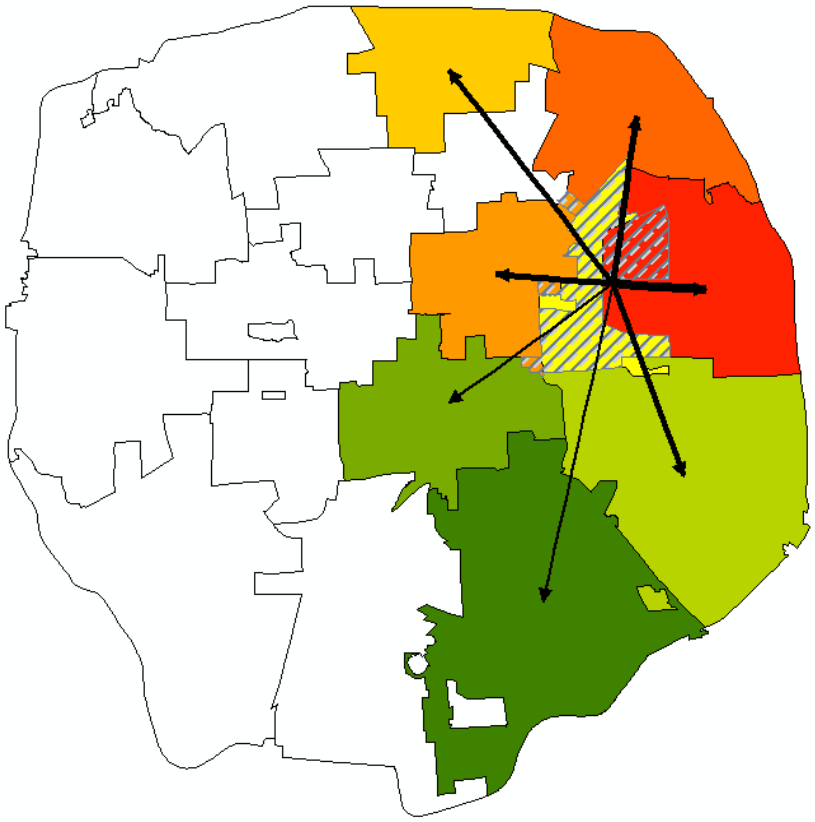


Origin

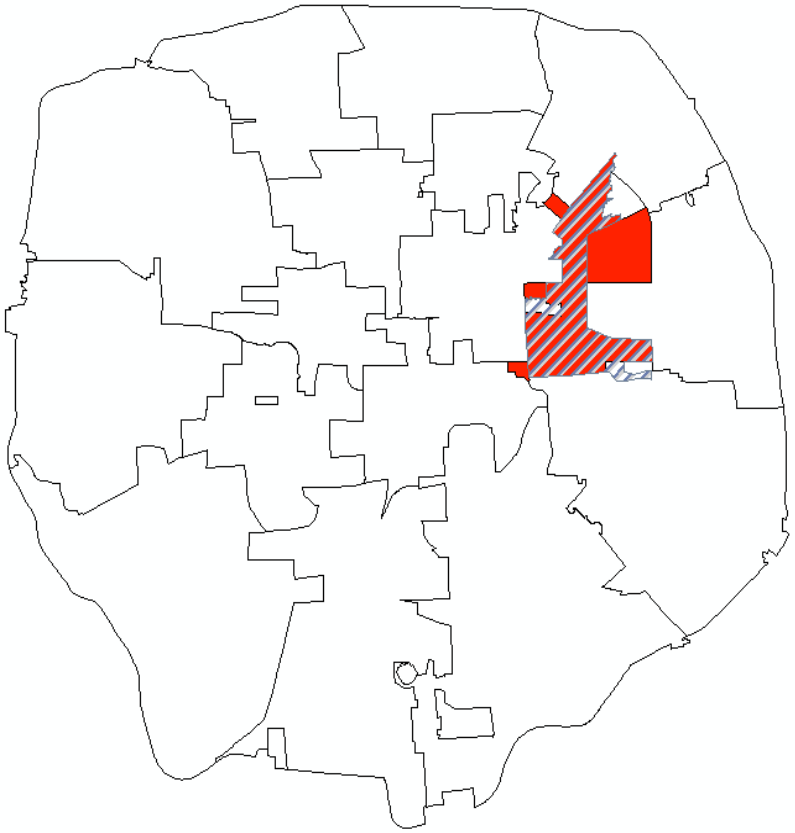


Destination

Evening

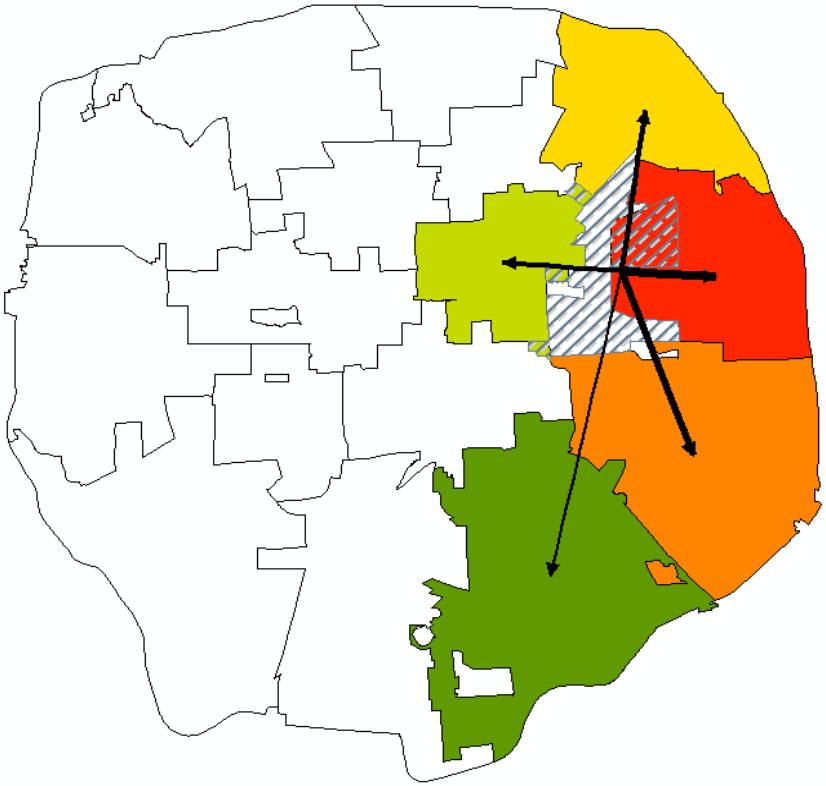


Origin

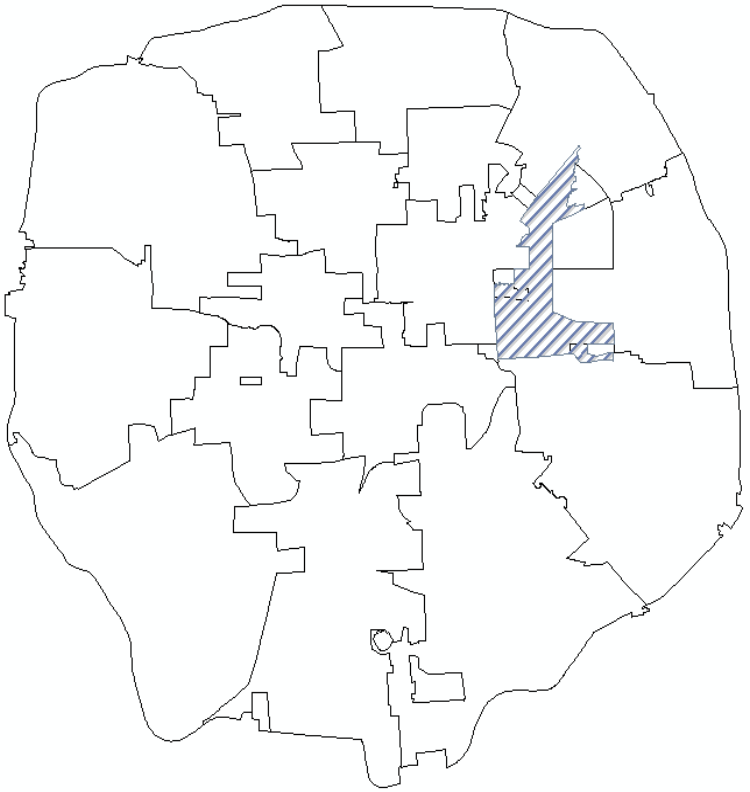


Destination

Night

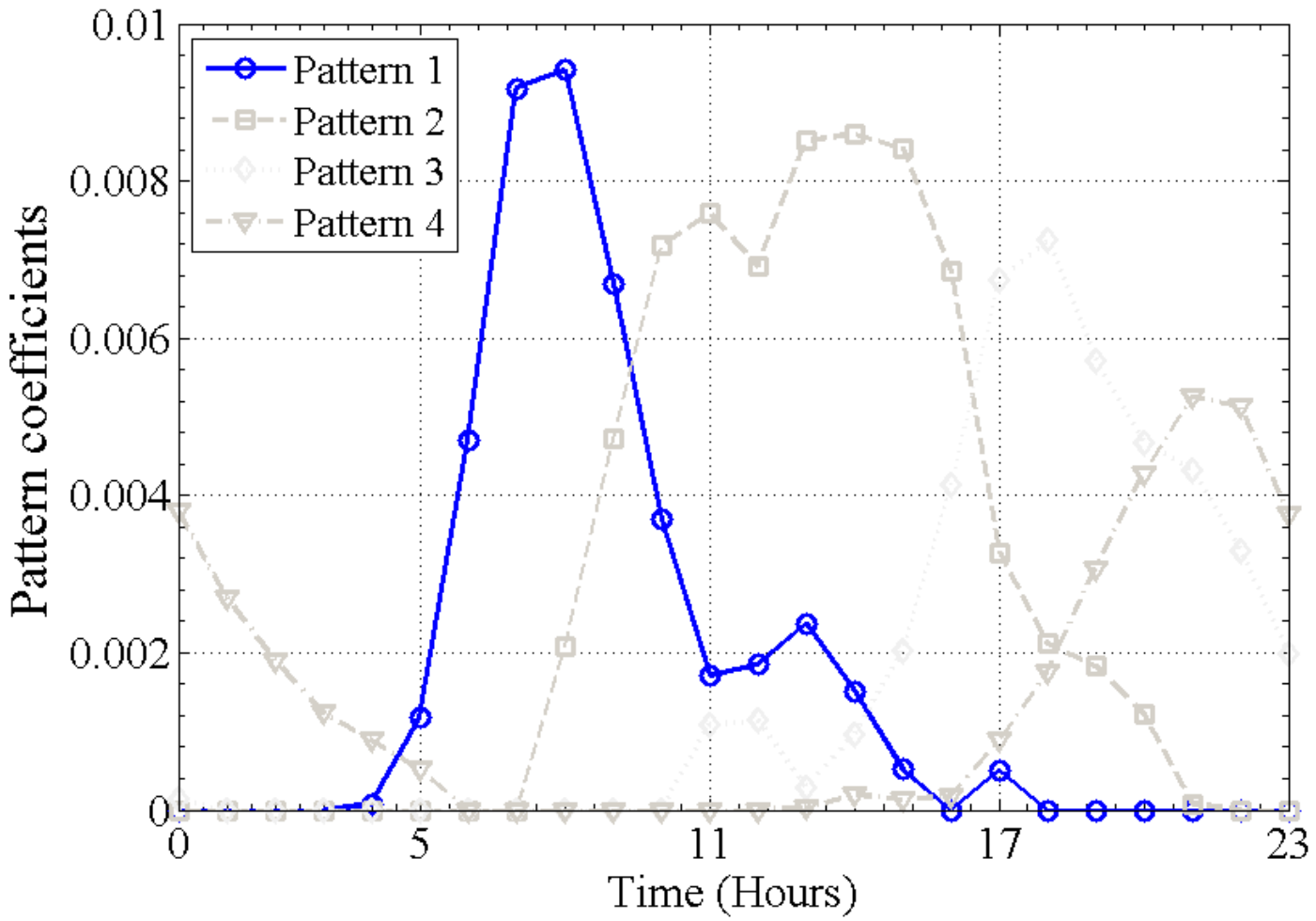


Origin

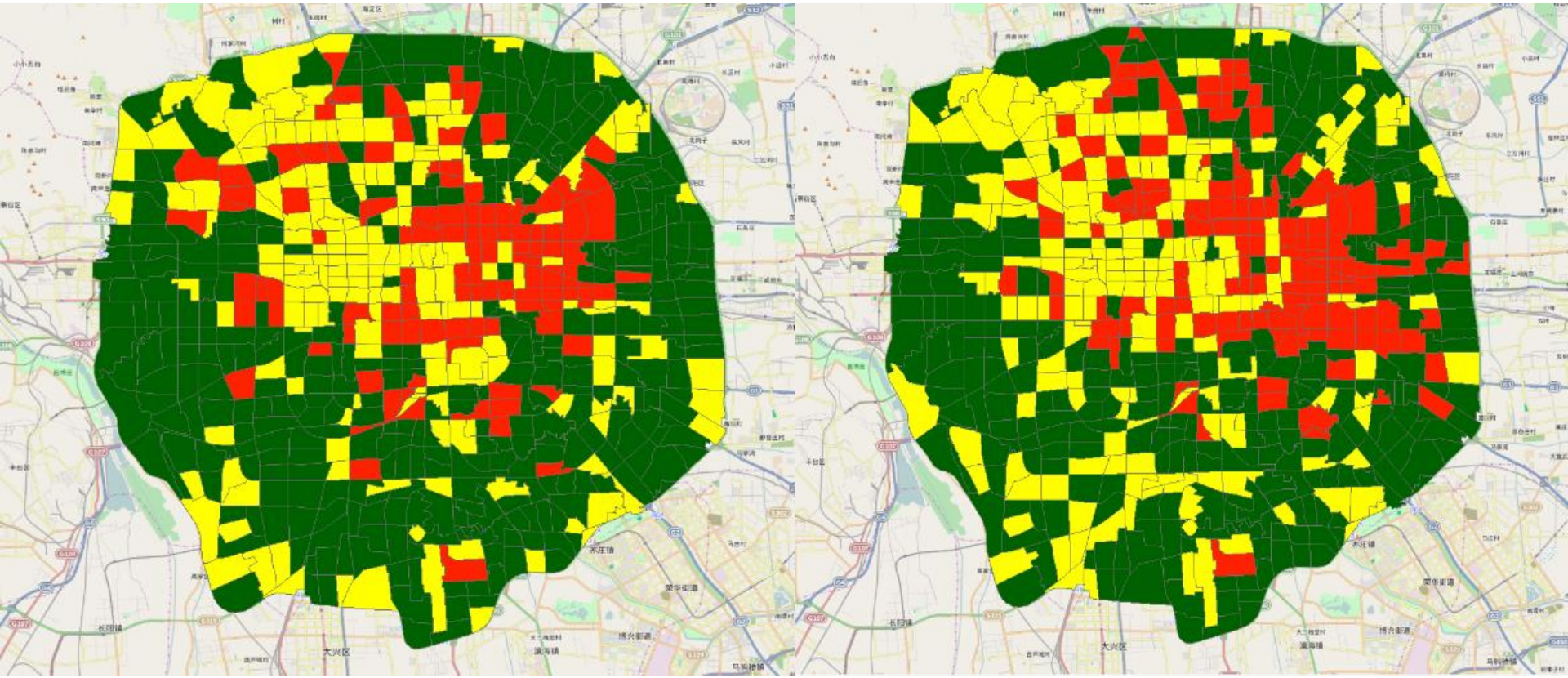


Destination

Function of urban zones



Function of urban zones



2008 2012

Where is the center of Beijing?

研究二：基于矩阵分解的服务点客流预测

研究任务：

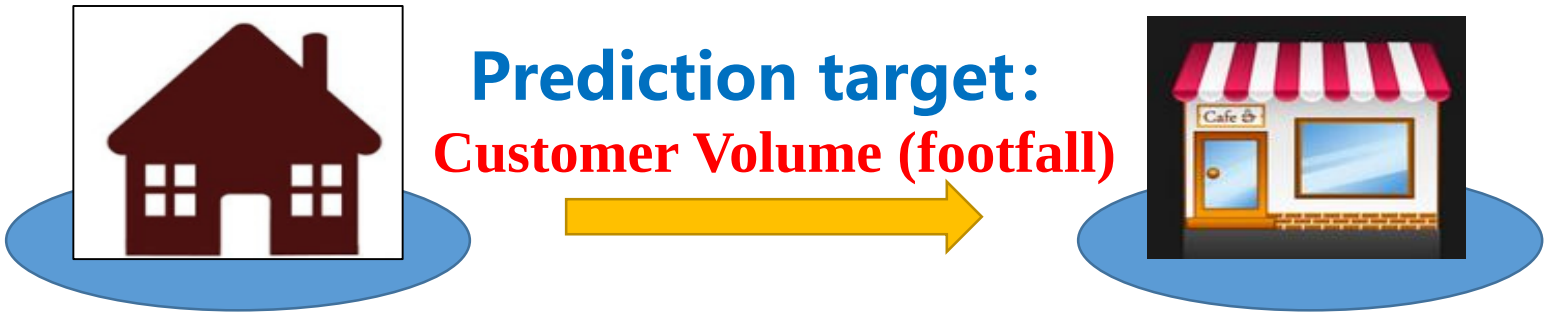
- 预测城市中一个服务型兴趣点（医院、饭店）的客流量

核心挑战：

- 如何识别影响客流量的各类因素并实现融合建模?

Application Scenarios

- The problem definition



- Implicit and explicit knowledges in footfall prediction
  - Explicit knowledges: **geographical correlations**
  - Implicit knowledges: **latent correlations**

- Ideas: Mining implicit and explicit knowledges with an fusion model framework.
  - Explicit knowledges: **linear regression models**
  - Implicit knowledges: **matrix factorization models**

Explicit knowledge: Geographical Context

- Geographical Relation
  - The geographic relationship between service and customer points
    - Geographical distance
    - Number of service/ customer points around customer/ service points
- Geographical Similarity
  - The footfall similarity among customer-service point pairs that are geographically close.
    - The average footfall to a service/customer point from the customer/service points nearest to a customer/service point
- Social Geography Features
  - Social connections between customer and service points
    - The traffic flow intensity from a customer point and a service point
    - Whether a customer point and a service point are in the same administrative region

Explicit Correlations

- Geographical Regression
  - The linear regression model

$$x_{ij} = \mathbf{w}^\top \mathbf{a}_{ij} + b$$

  - A tensor/matrix expression of LR model

$$\mathbf{X} = \mathcal{A} \times_k \mathbf{w} + \mathbf{E}_2$$

  - The objective function for footfall prediction

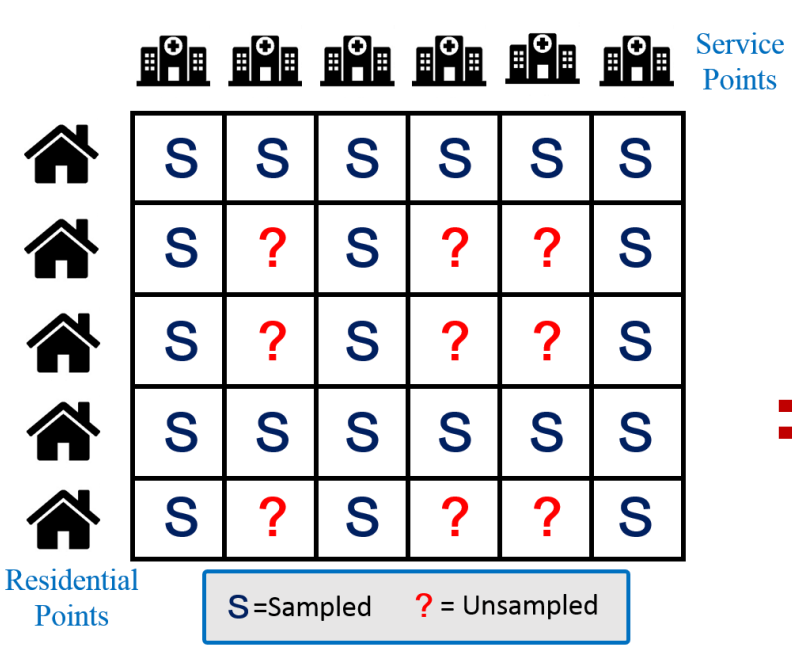
$$\mathcal{J}_2 = \frac{1}{\sigma_{X2}^2} \|\mathbf{Y} \odot (\mathbf{X} - \mathcal{A} \times_k \mathbf{w})\|_F^2 + \frac{1}{\sigma_{W1}^2} \|\mathbf{w}\|_2^2$$

Motivation: Implicit v.s. Explicit Knowledge

- **Explicit Knowledge**
  - Correlation with a known reason
    - For example: twins are looked like each other.
- **Implicit Knowledge**
  - Correlation with unknown reasons
    - For example: doppelganger

Implicit Correlations

- Probabilistic Matrix Factorization



=

Residential Points -Hidden Space

×

Service Points -Hidden Space

- Objective function:

$$\mathcal{J}_1 = \frac{1}{\sigma_{X1}^2} \left\| \mathbf{Y} \odot (\mathbf{X} - \mathbf{S}^\top \mathbf{C}) \right\|_F^2 + \frac{1}{\sigma_S^2} \|\mathbf{S}\|_1 \frac{1}{\sigma_R^2} \|\mathbf{C}\|_1$$

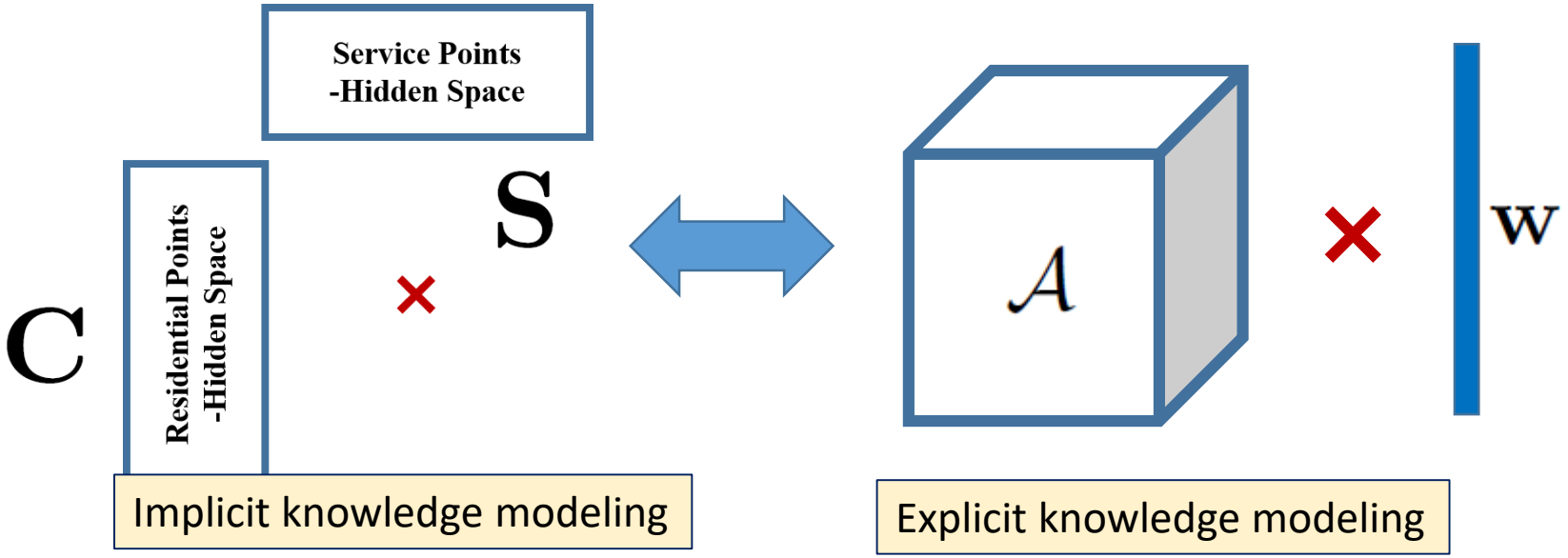
Modeling Unobserved Volumes

- The objective function

$$\mathcal{J}_3 = \frac{1}{\sigma_{W2}^2} \left\| \overline{\mathbf{Y}} \odot (\mathcal{A} \times_k \mathbf{w} - \mathbf{S}^\top \mathbf{C}) \right\|_F^2$$

Unobserved Volumes

- Calibrating implicit knowledge modeling with explicit knowledge modeling



Model and Inference

- **GR-NMF:** Integrated Model for Footfall Prediction

$$\min \mathcal{J} = \left\| \mathbf{Y} \odot (\mathbf{X} - \mathbf{S}^\top \mathbf{C}) \right\|_F^2 \quad \textbf{Implicit Correlations}$$

$$+ \alpha \left\| \mathbf{Y} \odot (\mathbf{X} - \mathcal{A} \times_k \mathbf{w}) \right\|_F^2 \quad \textbf{Explicit Correlations}$$

$$+ \beta \left\| \overline{\mathbf{Y}} \odot (\mathcal{A} \times_k \mathbf{w} - \mathbf{S}^\top \mathbf{C}) \right\|_F^2 \quad \textbf{Unobserved Volumes}$$

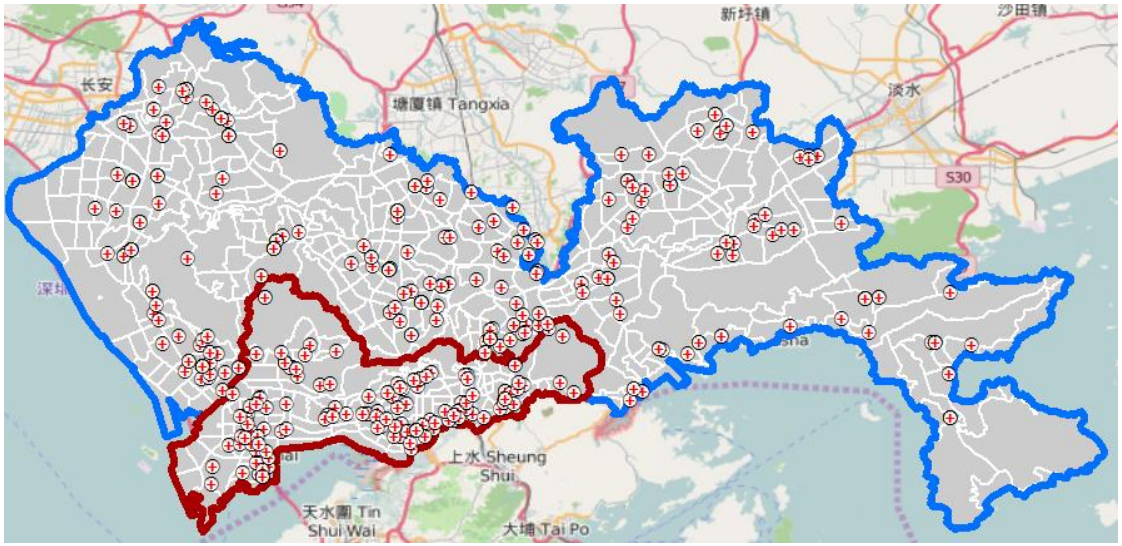
$$+ \gamma \left\| \mathbf{w} \right\|_2^2 + \delta \left\| \mathbf{S} \right\|_1 + \zeta \left\| \mathbf{C} \right\|_1 \quad \textbf{Sparsity Factors}$$

$$s.t.\ \mathbf{S} \geq 0, \mathbf{C} \geq 0, \mathbf{w} \geq 0, \quad \textbf{Non-negativity Constraints}$$

- **Inference**
  - Alternating Proximal Gradient Descent (APGD)

Experiments

- Experiments setup
  - **Data set:** collected from the public hospital system of Shenzhen, a major city in southern China
  - **Service points:** 321 public hospitals of Shenzhen
  - **Customer points:** 1343 residential zones of Shenzhen
  - **Time range:** January to December, 2014



A 321×1343 patients volume matrix $\boldsymbol{X}$, with a high sparsity (the ratio of zero elements) equal to 94.87%.

Experiments

- Baselines
  - Linear Regression (LR):

$$\min \sum_{y_{ij}=1} (x_{ij} - \mathbf{w}^\top \mathbf{a}_{ij})^2.$$

  - Singular Value Decomposition (SVD):

$$\min \|\mathbf{Y} \odot (\mathbf{X} - \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)\|_F^2.$$

  - Basic Non-Negative Matrix Factorization (bNMF):

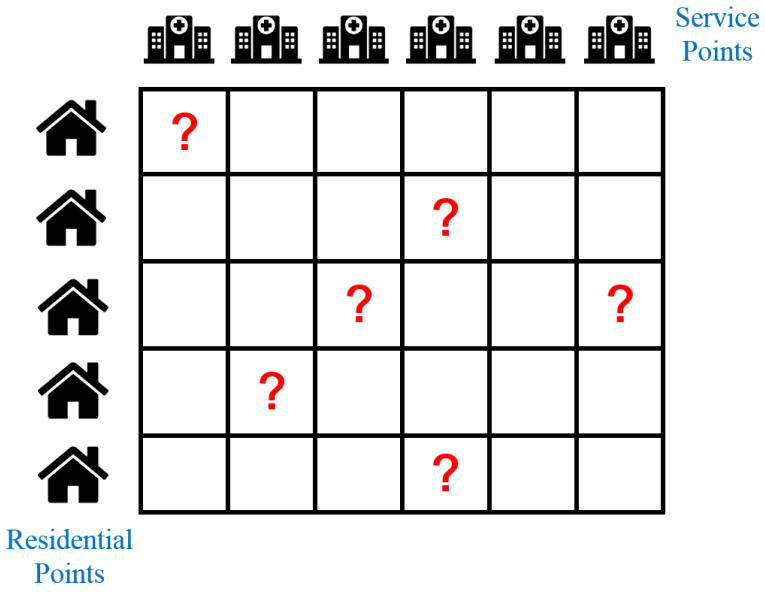
$$\min \|\mathbf{Y} \odot (\mathbf{X} - \mathbf{S}^\top \mathbf{C})\|_F^2,$$
$$s.t.\ \mathbf{S} \geq 0, \mathbf{C} \geq 0.$$

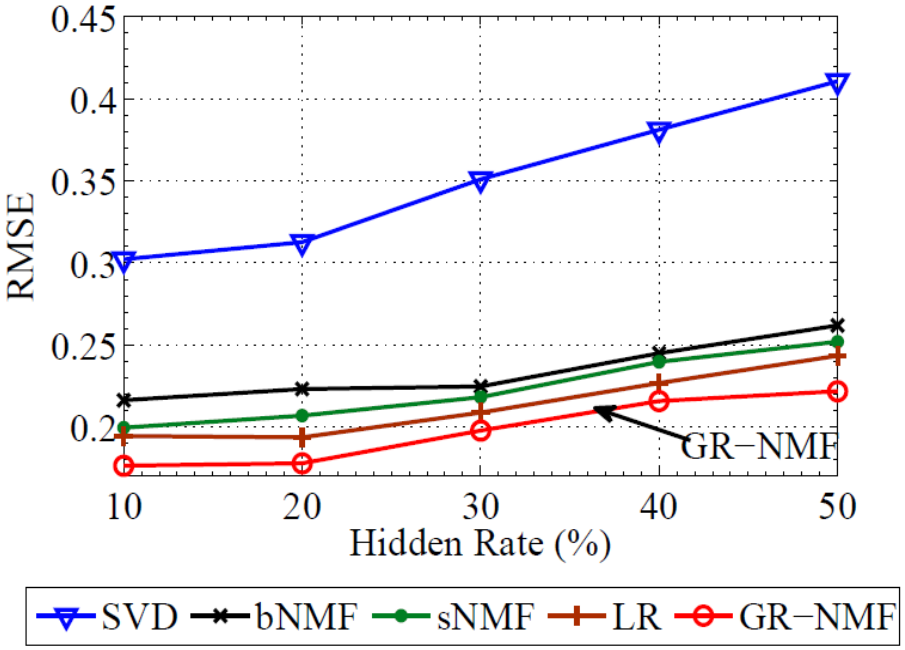
  - Sparse Non-Negative Matrix Factorization (sNMF):

$$\min \|\mathbf{Y} \odot (\mathbf{X} - \mathbf{S}^\top \mathbf{C})\|_F^2 + \alpha\|\mathbf{S}\|_1 + \beta\|\mathbf{C}\|_1,$$
$$s.t.\ \mathbf{S} \geq 0, \mathbf{C} \geq 0.$$

Experiments

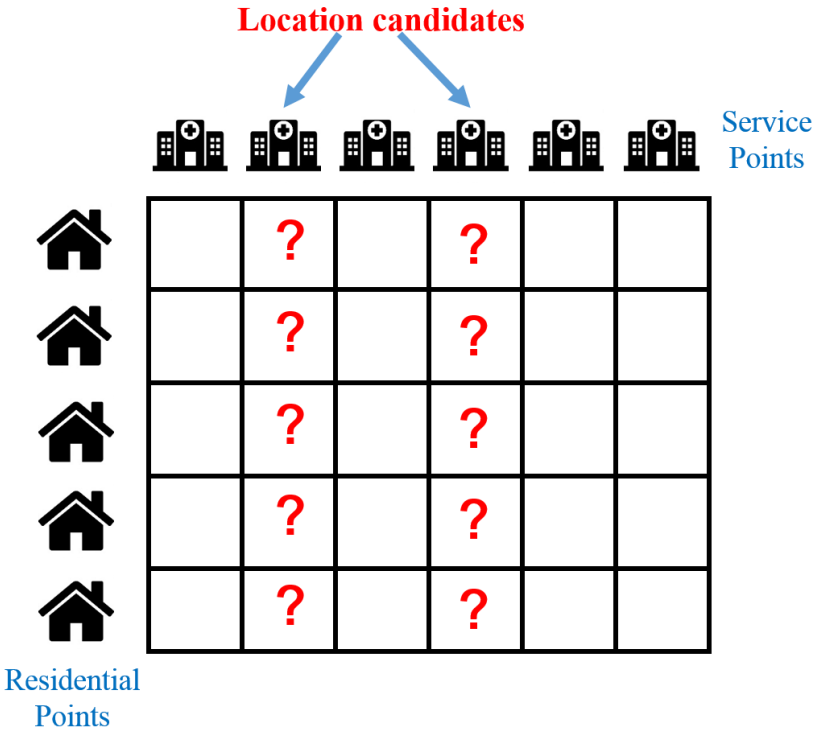
- The general scenario
  - Randomly hide 10% to 50% samples of the footfall matrix.

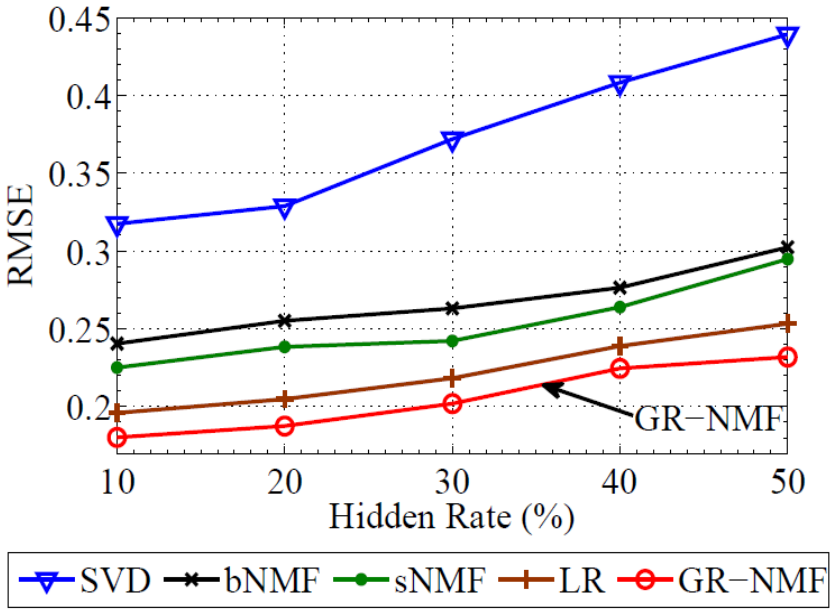




Experiments

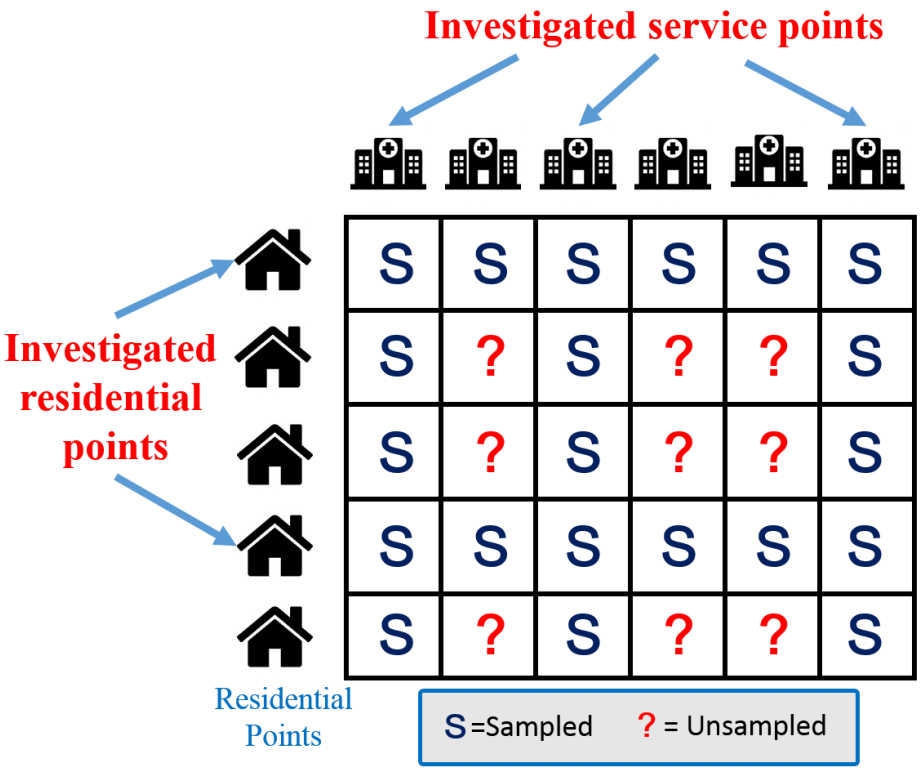
- The location selection scenario
  - Randomly hide 10% to 50% columns of the footfall matrix.
  - The hidden columns correspond to location candidates.

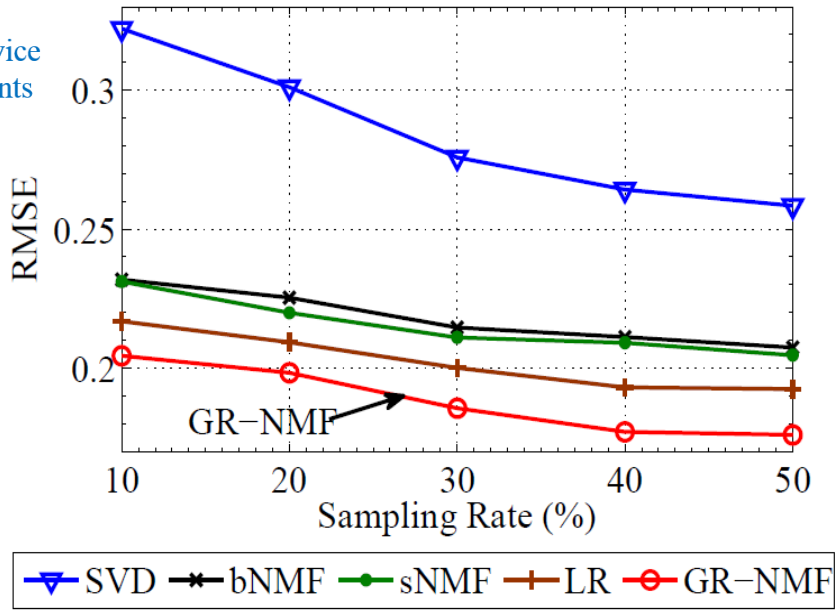




Experiments

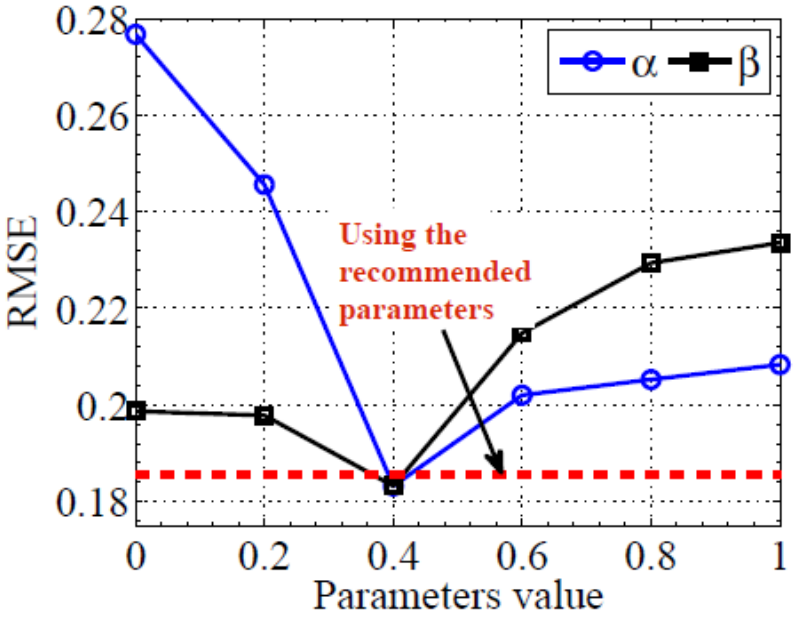
- The market investigation scenarios
  - Randomly sample 10%-50% rows and columns.
  - The sampled rows and columns are investigated service and residential points.



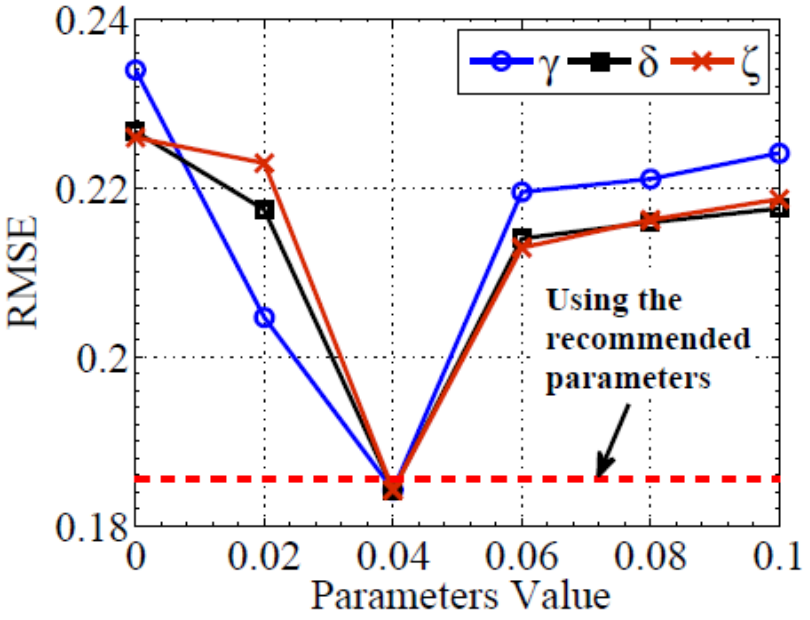


Experiments

- The setting of parameters
  - The performance of our approximate method is just slightly worse than the optimal performance of the traversal method.



(a) α and β



(b) γ, δ and ζ

Conclusions

- The model proposed by this study have following contributions:
  - GR-NMF is able to **jointly model implicit knowledge** hidden inside customer volumes **and explicit knowledge** expressed as geographical relations.
  - GR-NMF has a unified probabilistic interpretation, which makes **the model theoretically solid**.
  - Extensive experiments are conducted on **a real-life outpatient dataset** obtained from the Shenzhen city of China.
  - The results show that **GR-NMF outperforms competitive baselines** consistently in various application scenarios with different sampling rates.

研究三：基于多视角深度学习的城市外来人口识别

研究任务： 利用手机信令大数据识别城市的外来人口

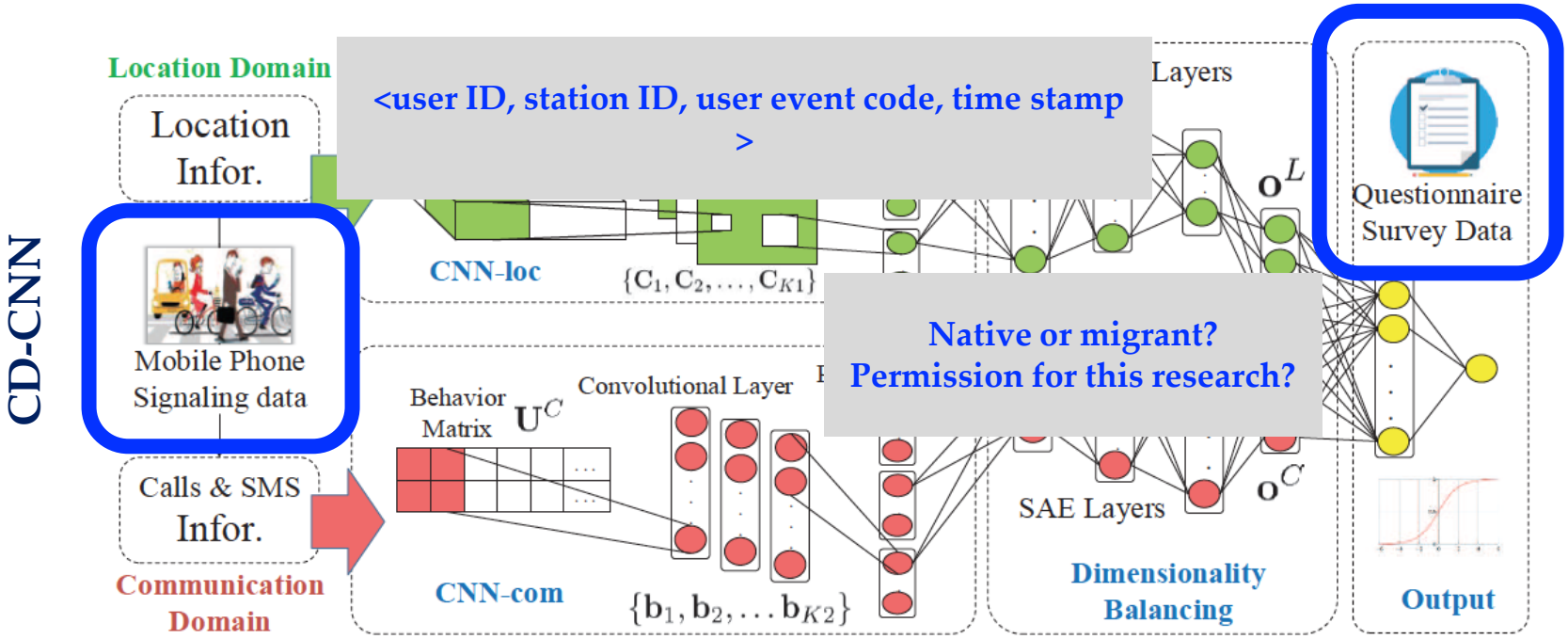
核心问题： 如何从信令数据中提取预测相关因子?

Native or Migrant: A Partially Supervised Cross-Domain Deep Learning Model for Citizenship Recognition. Submitted to ***IJCAI 2017***, with *J. Wang, et al.*

研究三：基于多视角深度学习的城市外来人口识别

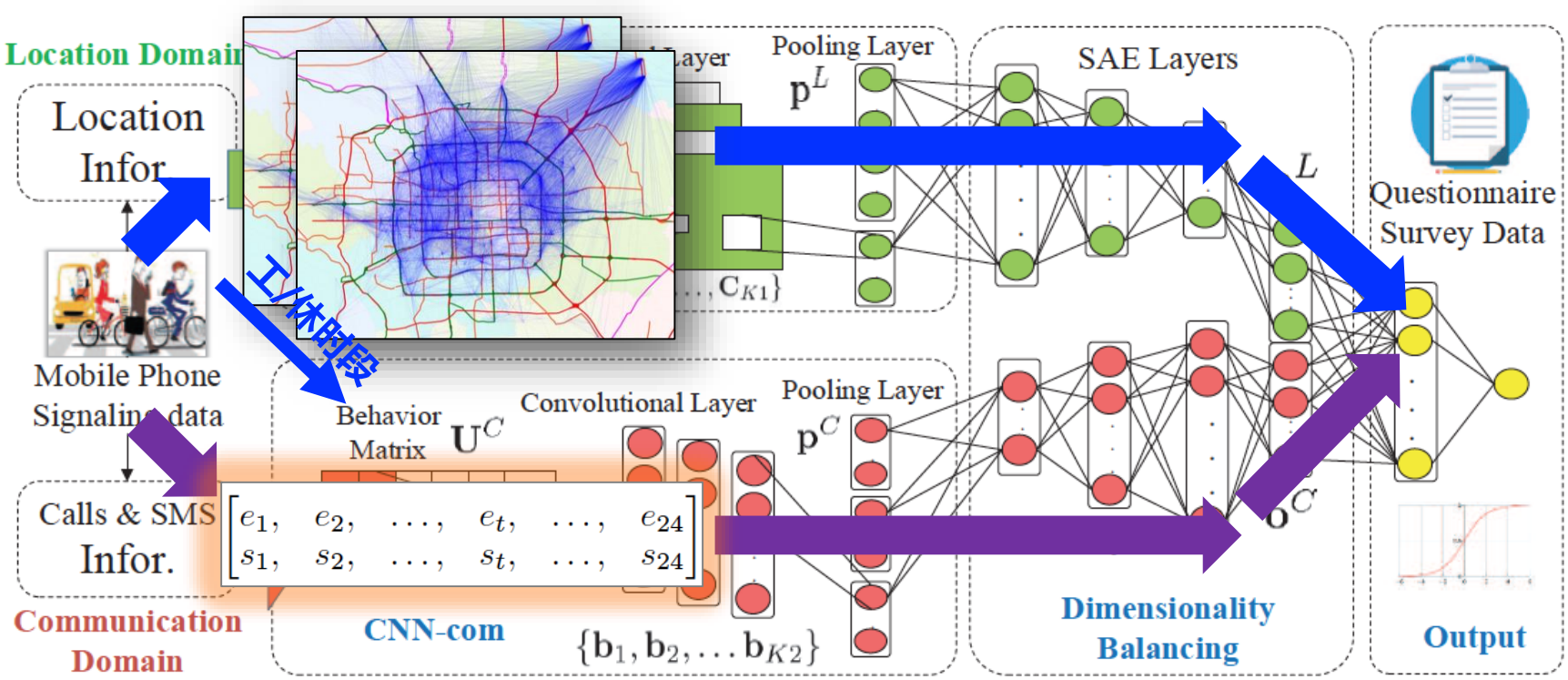
研究思路：

- 从信令数据中分别提取用户的**职住行为**（**空间信息**）和**通话行为**（**时间信息**）
- 利用**多视角**深度学习分别学习空、时特征，然后进行**特征融合建模**



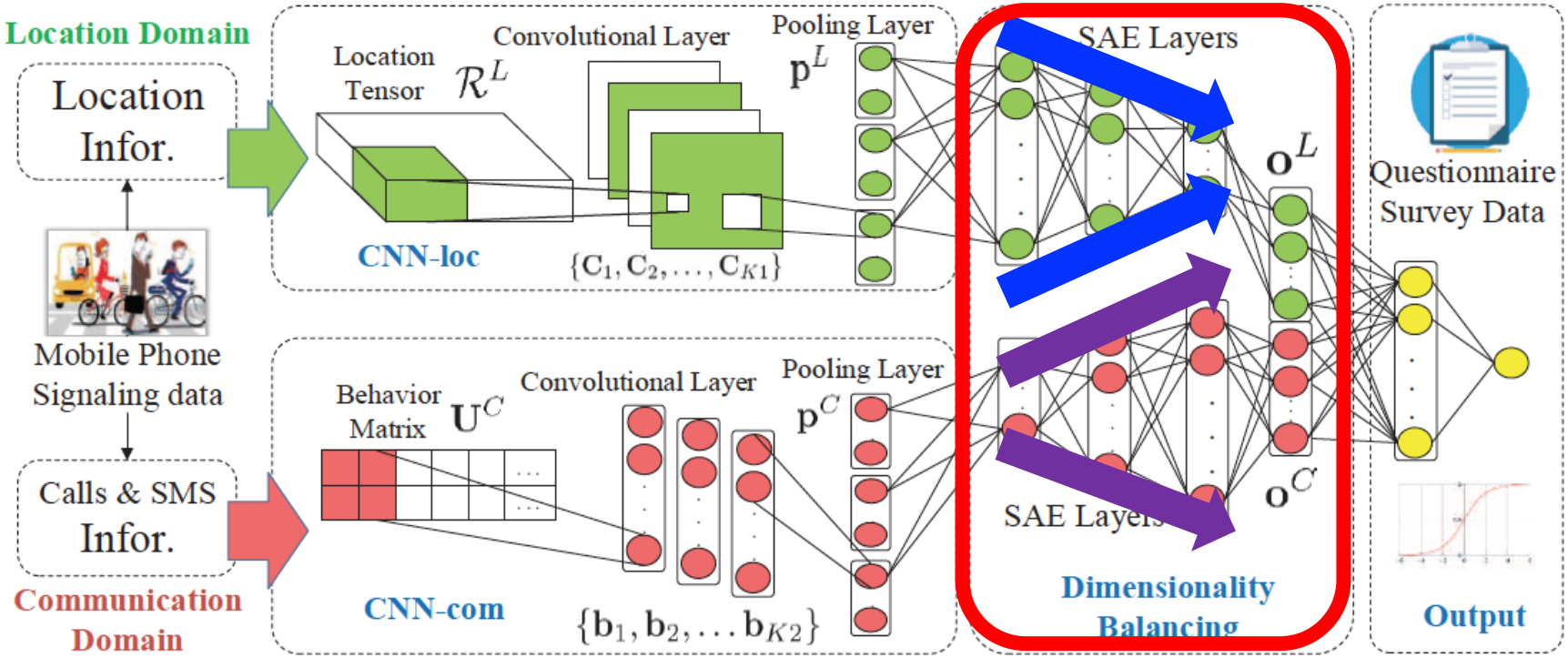
研究三：基于多视角深度学习的城市外来人口识别

核心贡献1：时-空多视角深度学习与特征融合建模



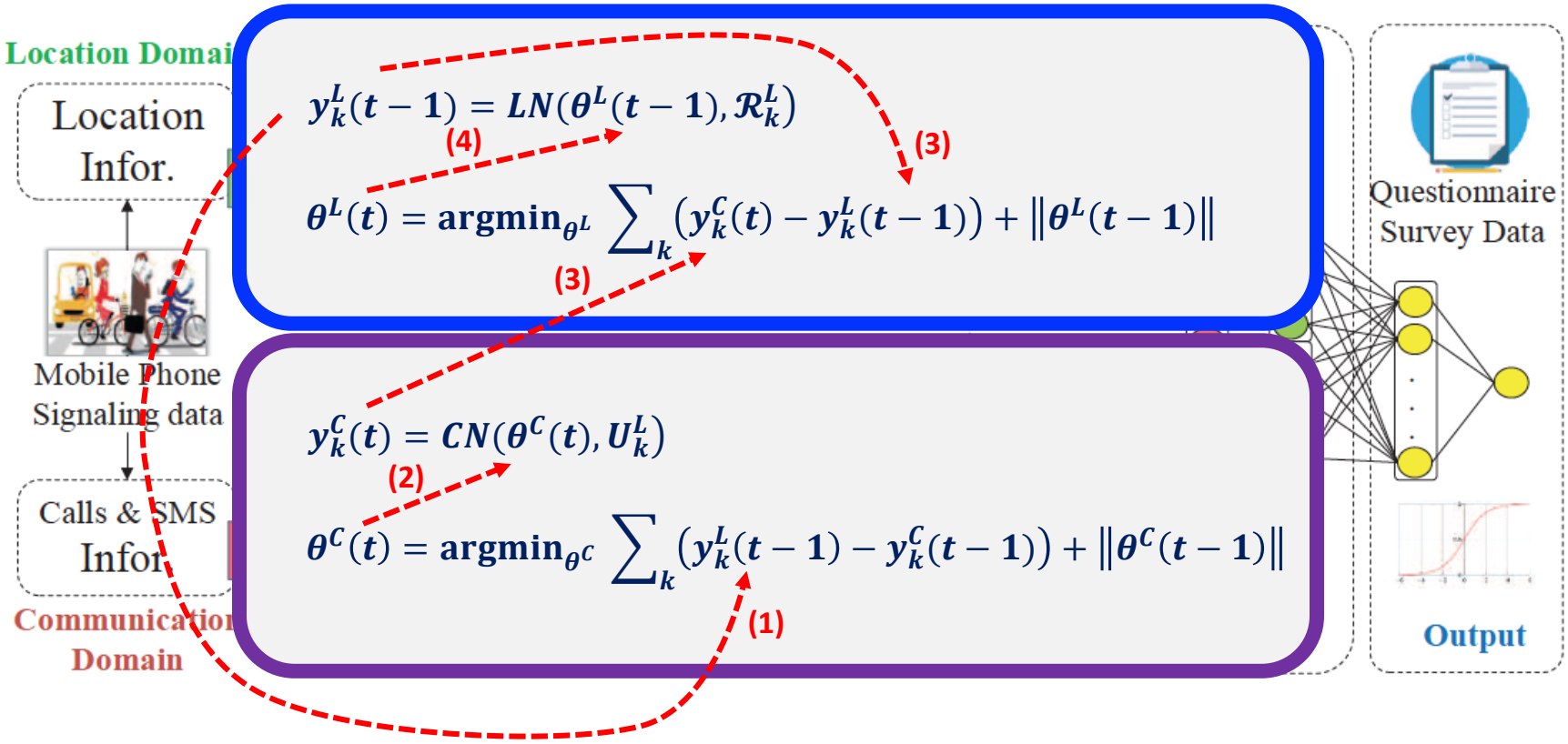
研究三：基于多视角深度学习的城市外来人口识别

核心贡献2：特征均衡处理



研究三：基于多视角深度学习的城市外来人口识别

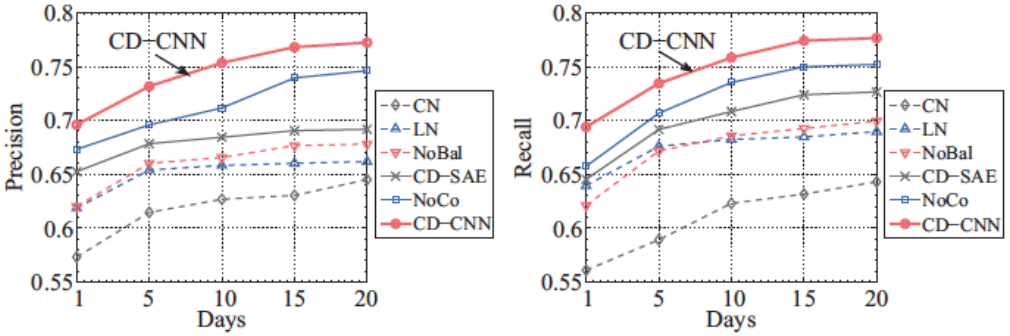
核心贡献3：基于跨领域co-training的部分监督学习



研究三：基于多视角深度学习的城市外来人口识别

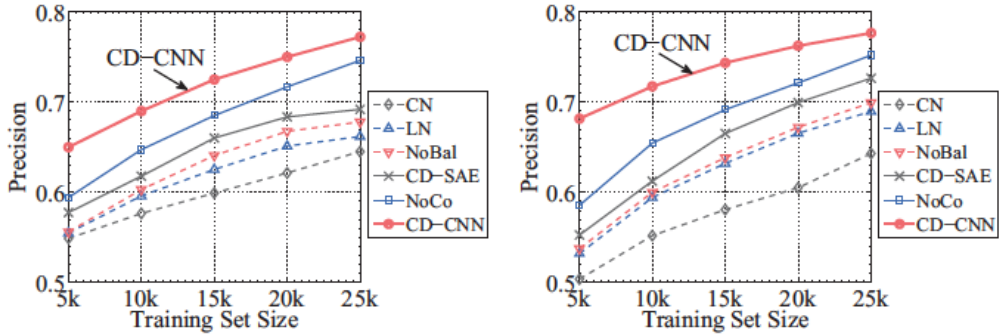
实验：依标签数据评价

- 城市：无锡，4787平方公里，600万人，有完善的工业生产体系和众多工业园区；被划分为10120个区块，3475个区块上有基站
- 信令：500万用户，2013.10 - 2014.03
- 问卷：3万人，本地/外地一半对一半
- 基线：5个



(a) Precision (b) Recall

Figure 2: Classification with varying data collecting days



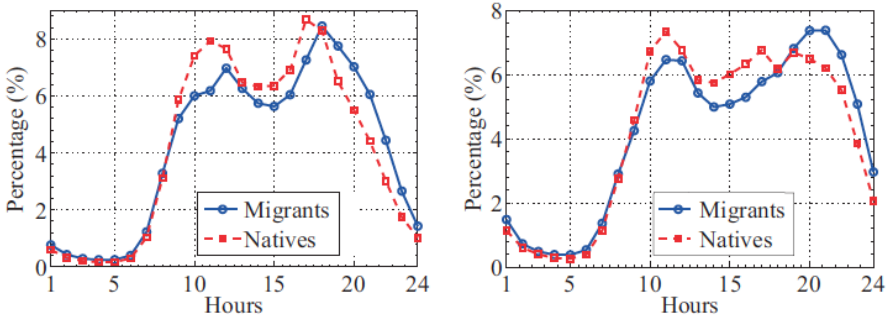
(a) Precision (b) Recall

Figure 3: Classification with varying sizes of labeled samples

研究三：基于多视角深度学习的城市外来人口识别

应用：无锡人口普查

- 总量：35%属于外来人口，与2014年无锡统计年鉴得到的32%极为接近！
- 通话模式：外来人口更晚
- 职住模式：外来人口集中于城中周



(a) Calls (b) Short Messages

Figure 4: Temporal distribution of resident communication behaviors.

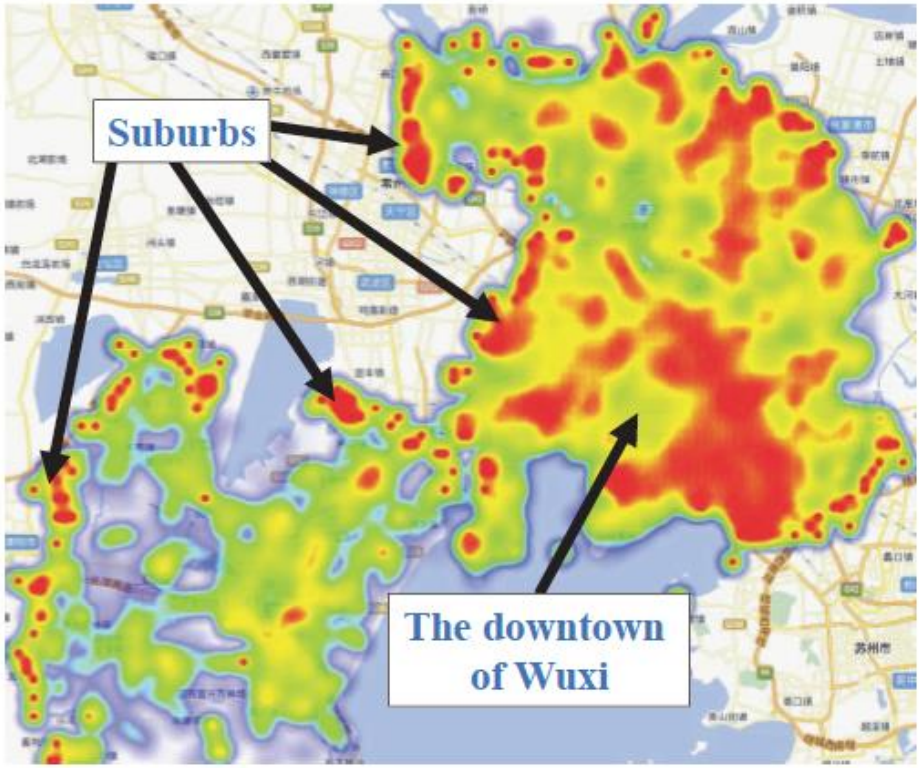


Figure 5: Residence distribution of migrants.

谢 谢

E-mail： jywang@buaa.edu.cn
Weibo： @王静远BUAA