

A world map where the landmasses are filled with a dense collection of small, white, stylized icons representing various concepts like technology, communication, and science. The background is a solid blue color.

人工智能、机器学习 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



3亿用户,每天
上亿条微博.



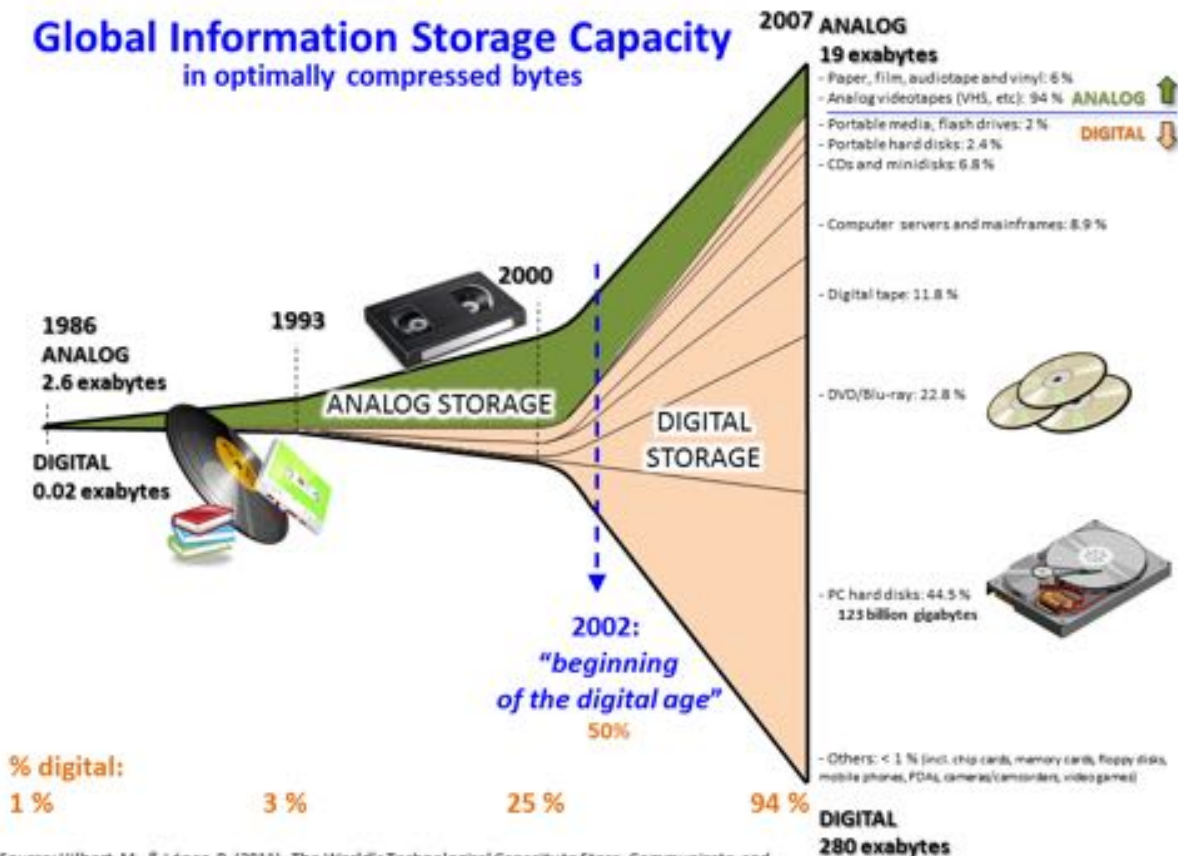
巡天望远镜,每年收集
600万GB数据



2015年全球移动终端
产生的数据量

6300PB

Global Information Storage Capacity in optimally compressed bytes



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>

大数据时代

• 大数据：规模大、变化快、种类杂

社交类应用

- **Facebook**：用户规模超过10亿，每天新增数据量10TB
- **四大微博(新浪，腾讯、搜狐和网易)**：用户8亿多，每天新增微博超过2亿条，图片2000万张

搜索类应用

- **百度**：每天新增日志数据量近1PB，数据总量近1000PB
- **Google**：每天新处理数据总量已超过20PB

图灵奖得主Jim Gray和IDC报告

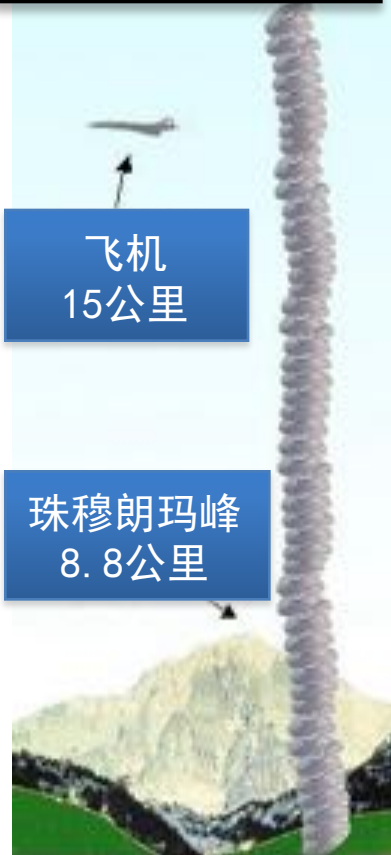
- 数据每18月翻一番，过去数据是确定的，当前伴随人机物融合，**网络信息空间大数据呈现多样性和异构性**
- **IDC报告**：全球数据2009年0.8 ZB，2012年 2.7 ZB，预计2020年达35ZB



1PB数据: DVD
存储**约25公里**
 $1ZB=1PB \times 10^6$

飞机
15公里

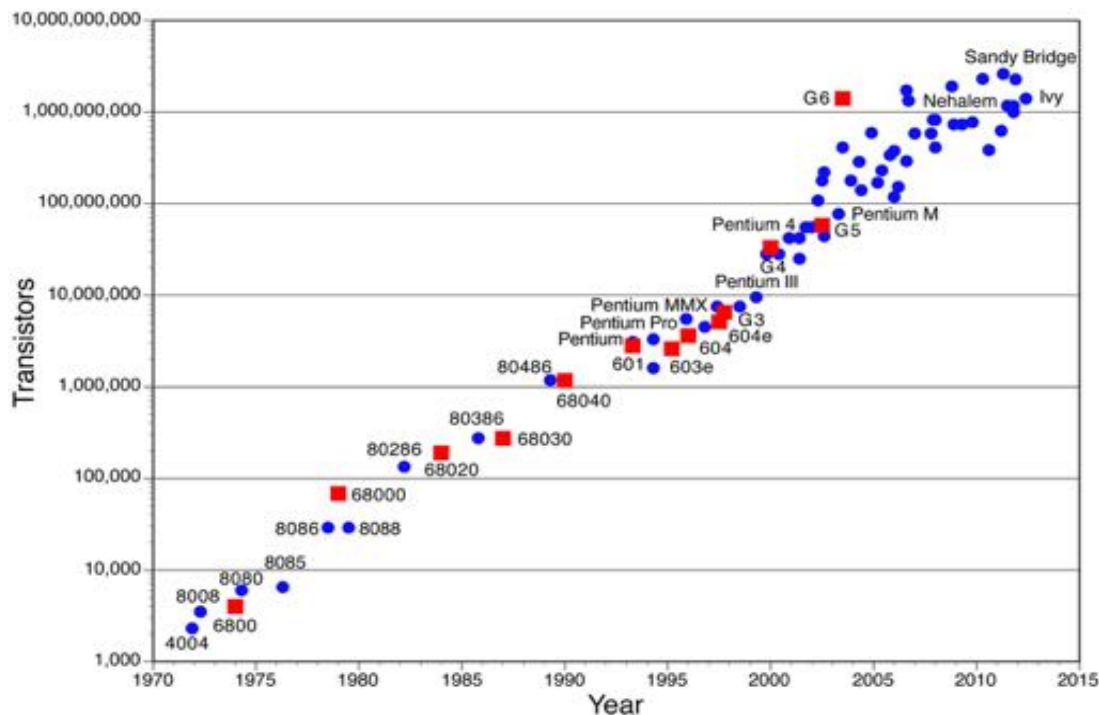
珠穆朗玛峰
8.8公里



大计算时代

• 计算能力的增强

摩尔定律



峰值计算速度54.9 PFLOPS



峰值计算速度125.436 PFLOPS

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。

知识发现的第四范式

- 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、中国古代的四大发明；

- 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。

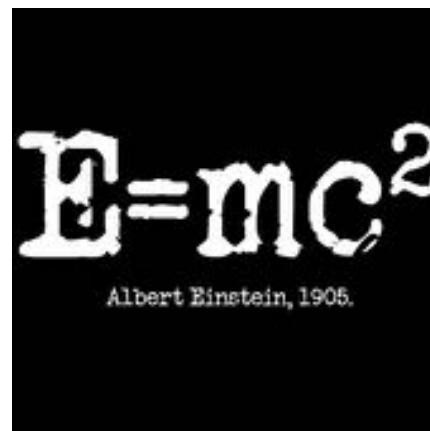
知识发现的第四范式

实验科学



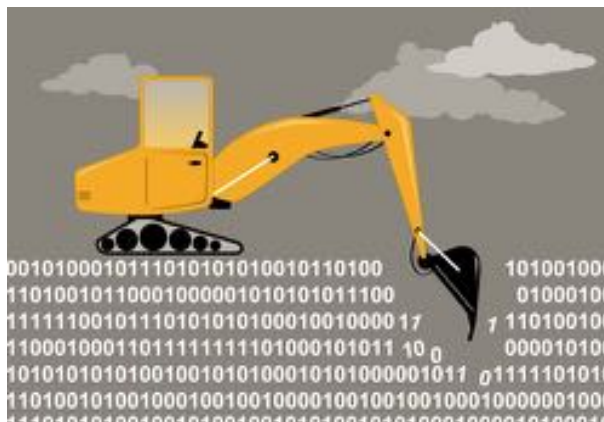
由人从自然现象（数据）中总结规律。

理论科学



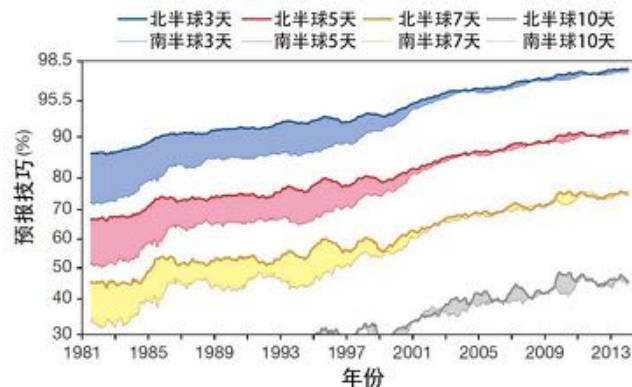
由人将规律总结成为数学模型。

数据科学



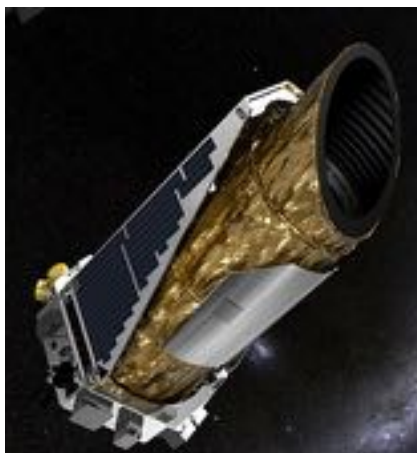
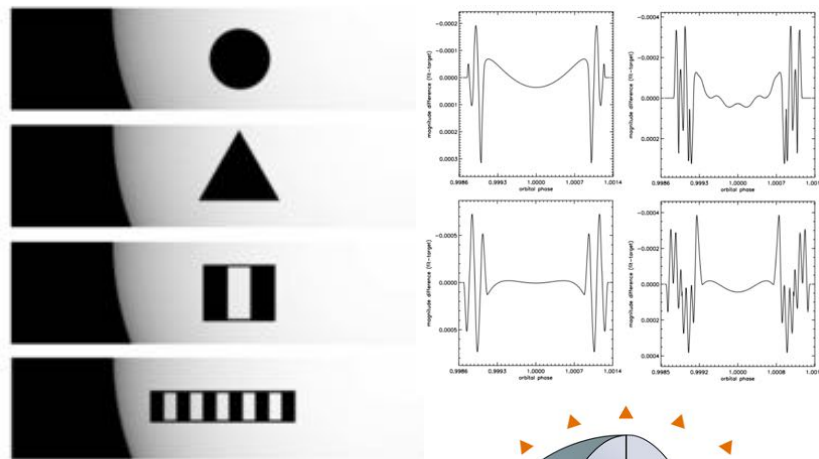
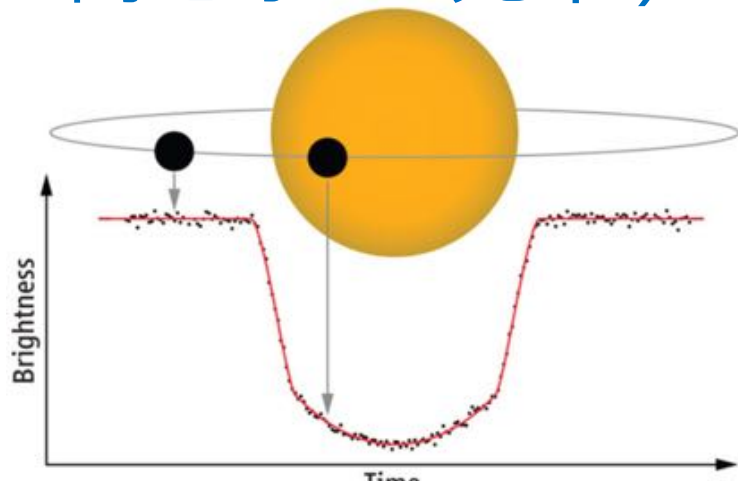
使用机器完成从规律总结到复杂预测的过程。

计算科学



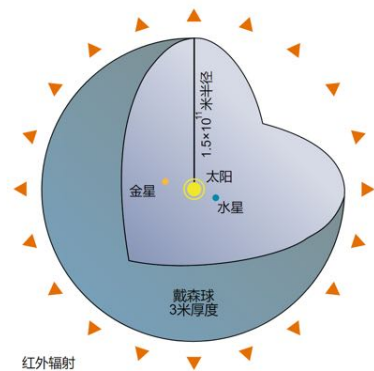
使用机器运行人类总结的规律实现复杂预测。

地球兄弟：Kepler-78b（距离地球400光年）



巡天望远镜, 每年收集
600万GB数据

KIC 8462852
1500光年外
疑似戴森球



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

知乎

首页

发现

话题

搜索你感兴趣的内容...



数据挖掘

机器学习

模式识别

计算机科学

模式识别，机器学习和数据挖掘的区别和联系？

● 添加评论

🔗 分享

★ 邀请回答



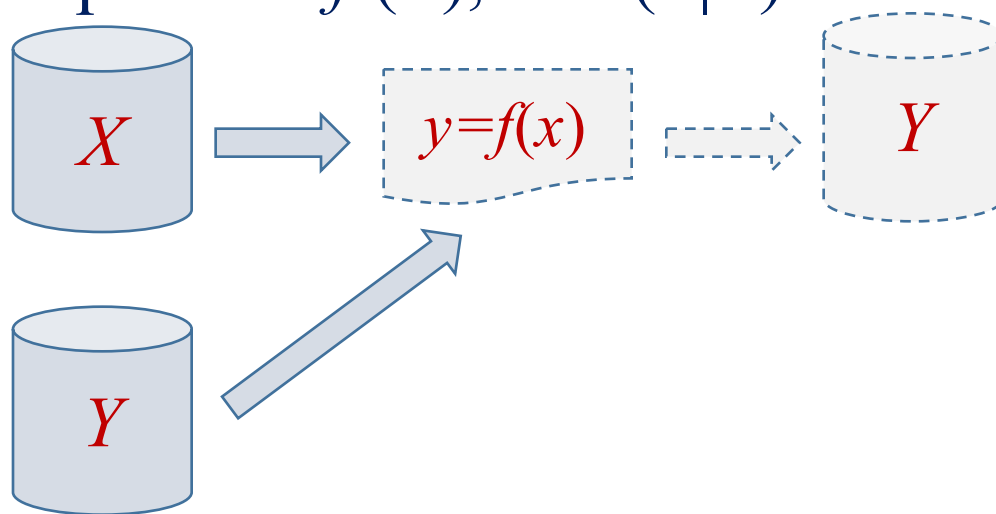
机器学习

- 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

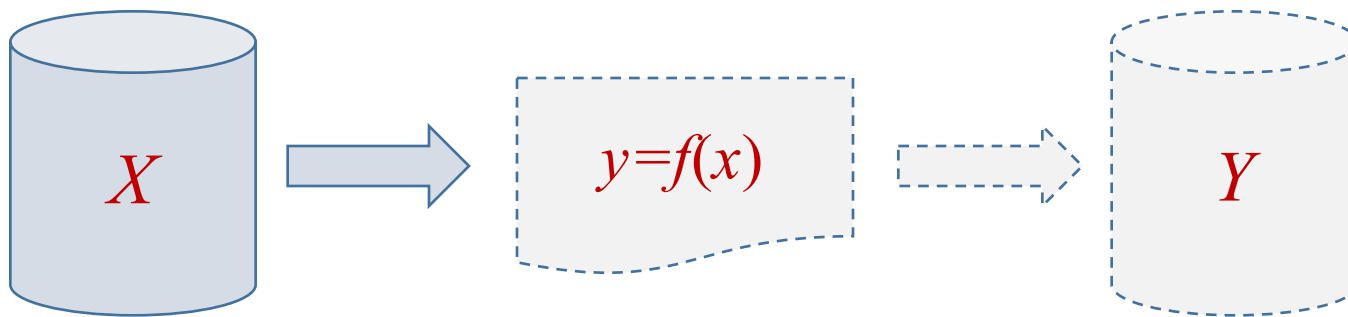
- 分类问题
- Input: training set \underline{X} , \underline{Y} ; testing set X
- Output: $Y = f(X)$, or $P(Y|X)$



机器学习

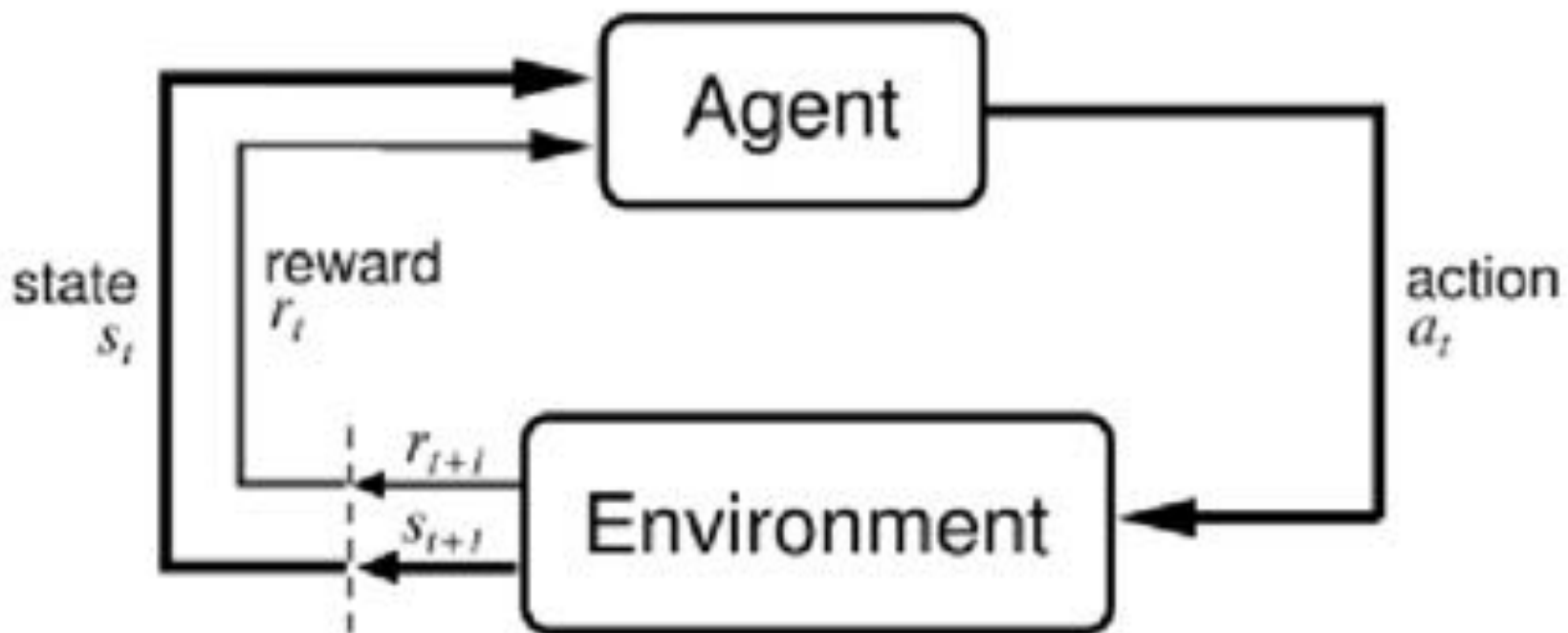
- 聚类问题
- Input: X
- Output: $Y = f(X)$, or $P(Y|X)$

无监督学习

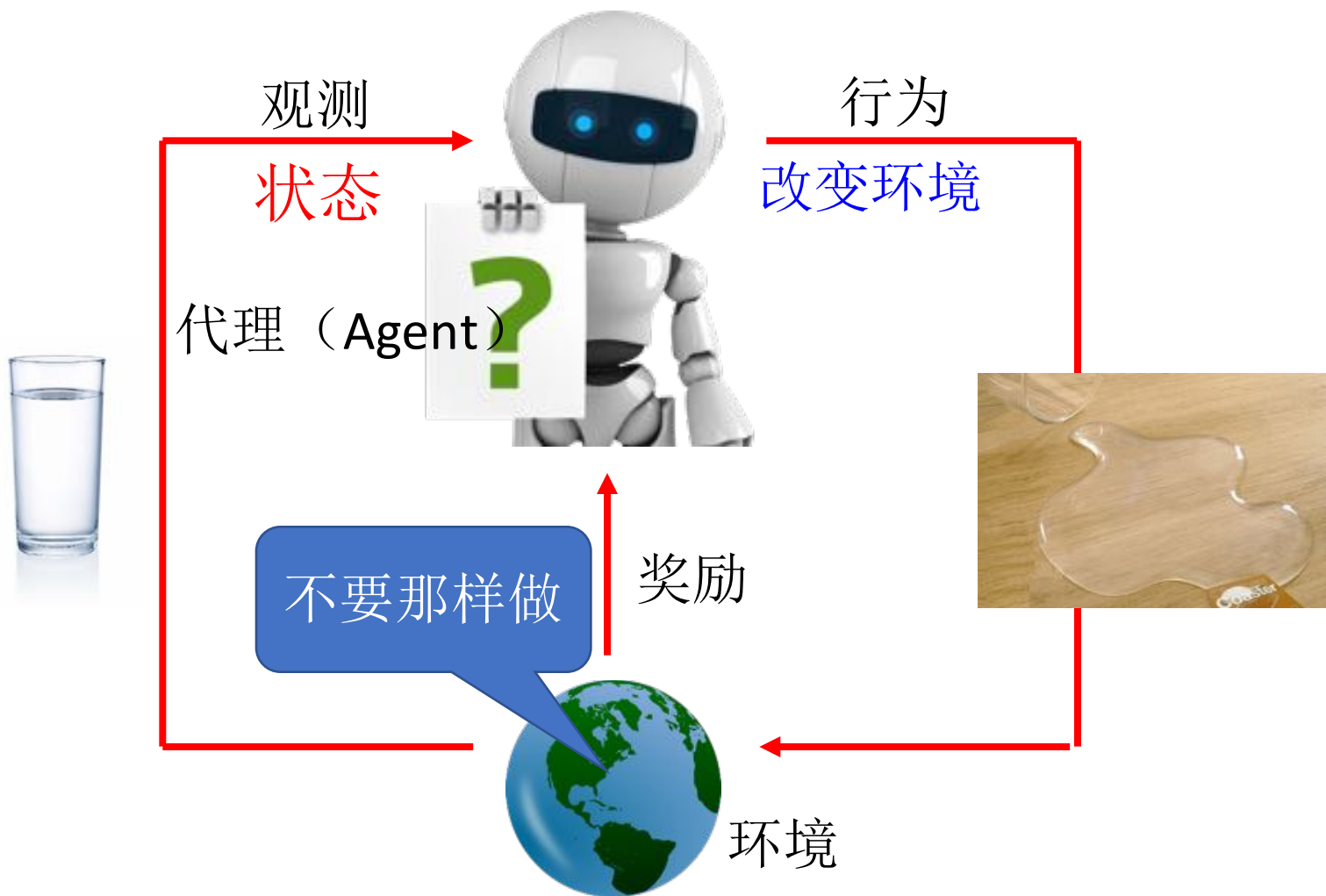


机器学习

强化学习



强化学习的场景

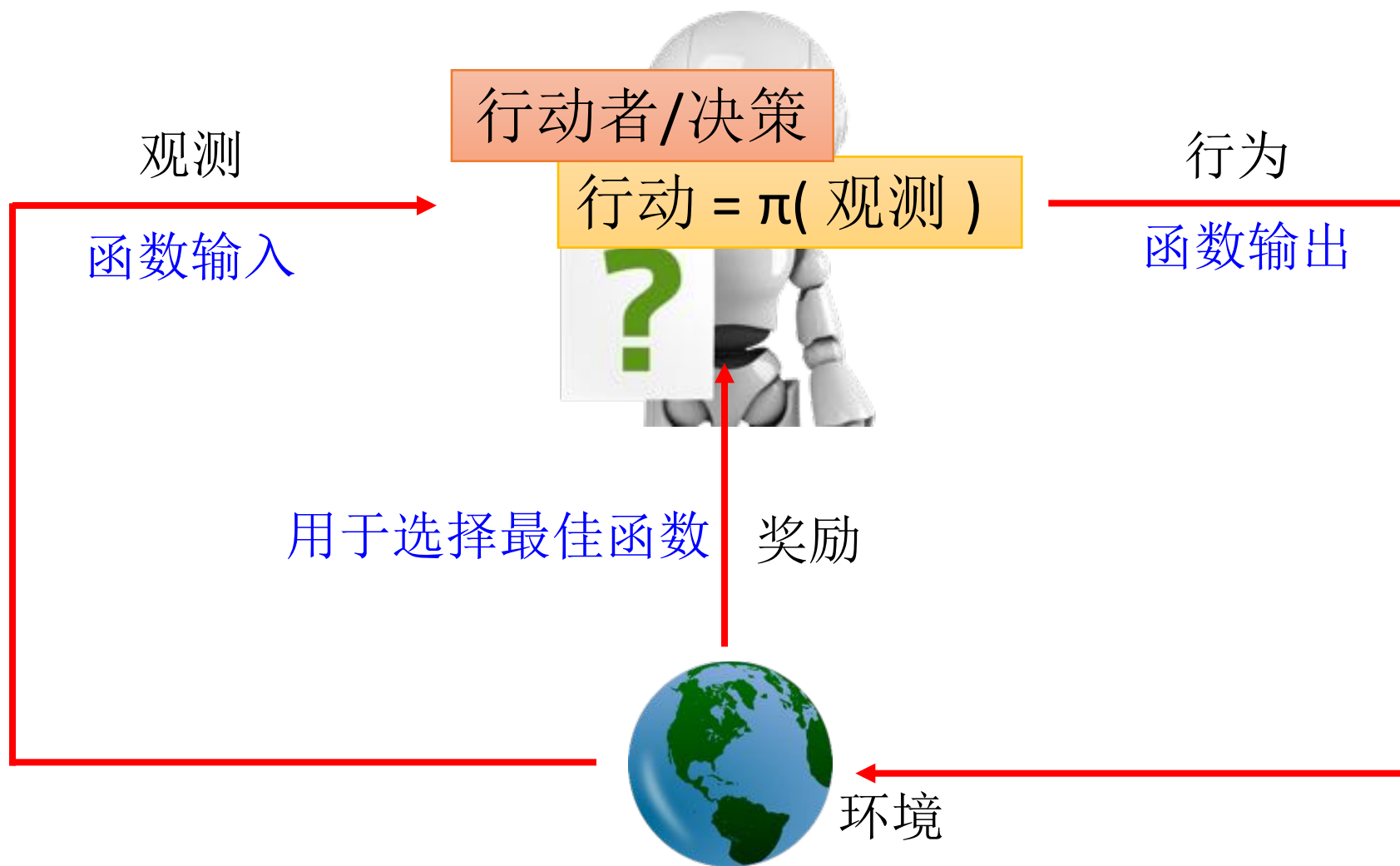


强化学习的场景

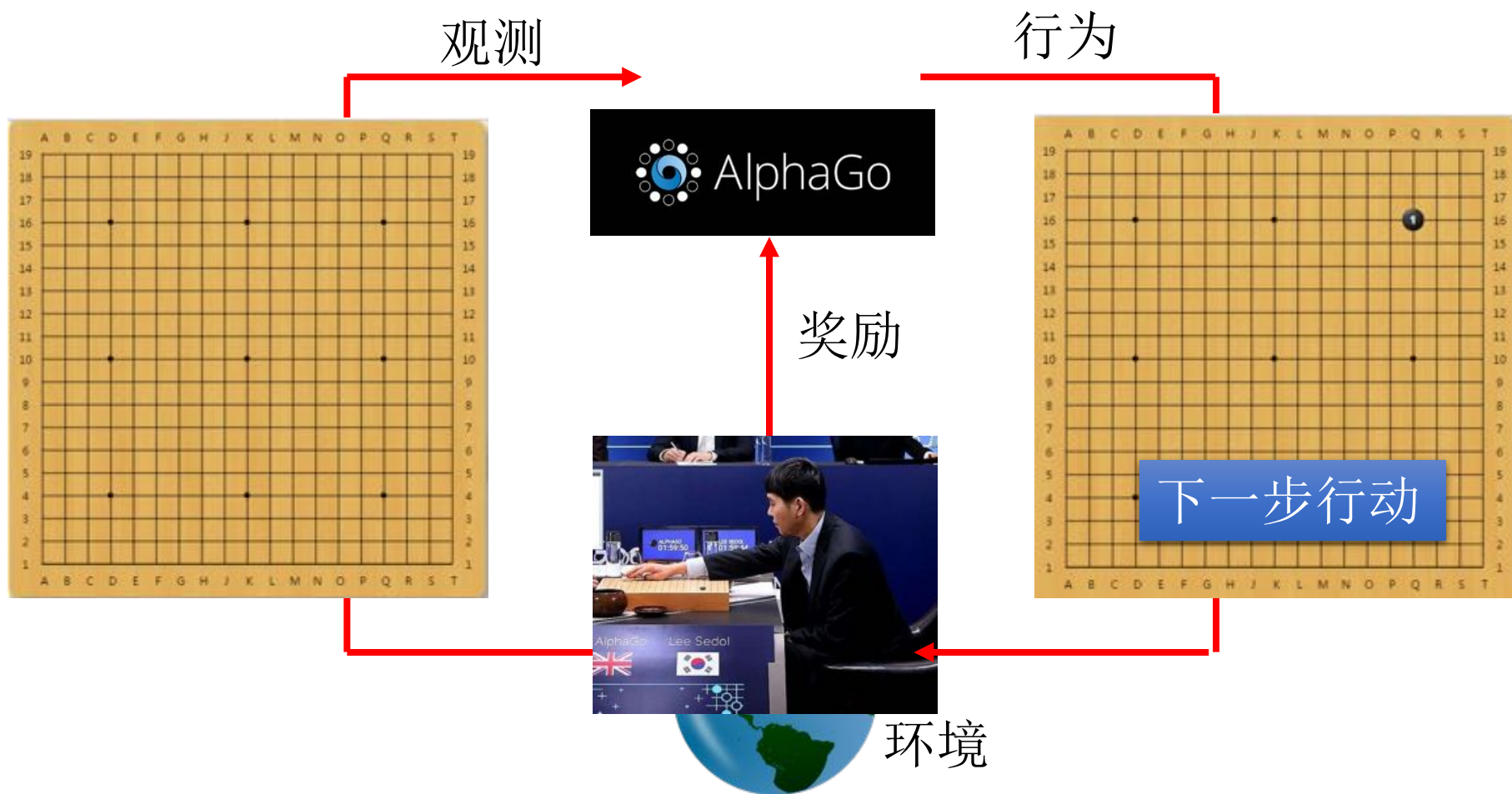
代理学会采取行动使回报的期望最大化。



机器学习 \approx 寻找一个函数

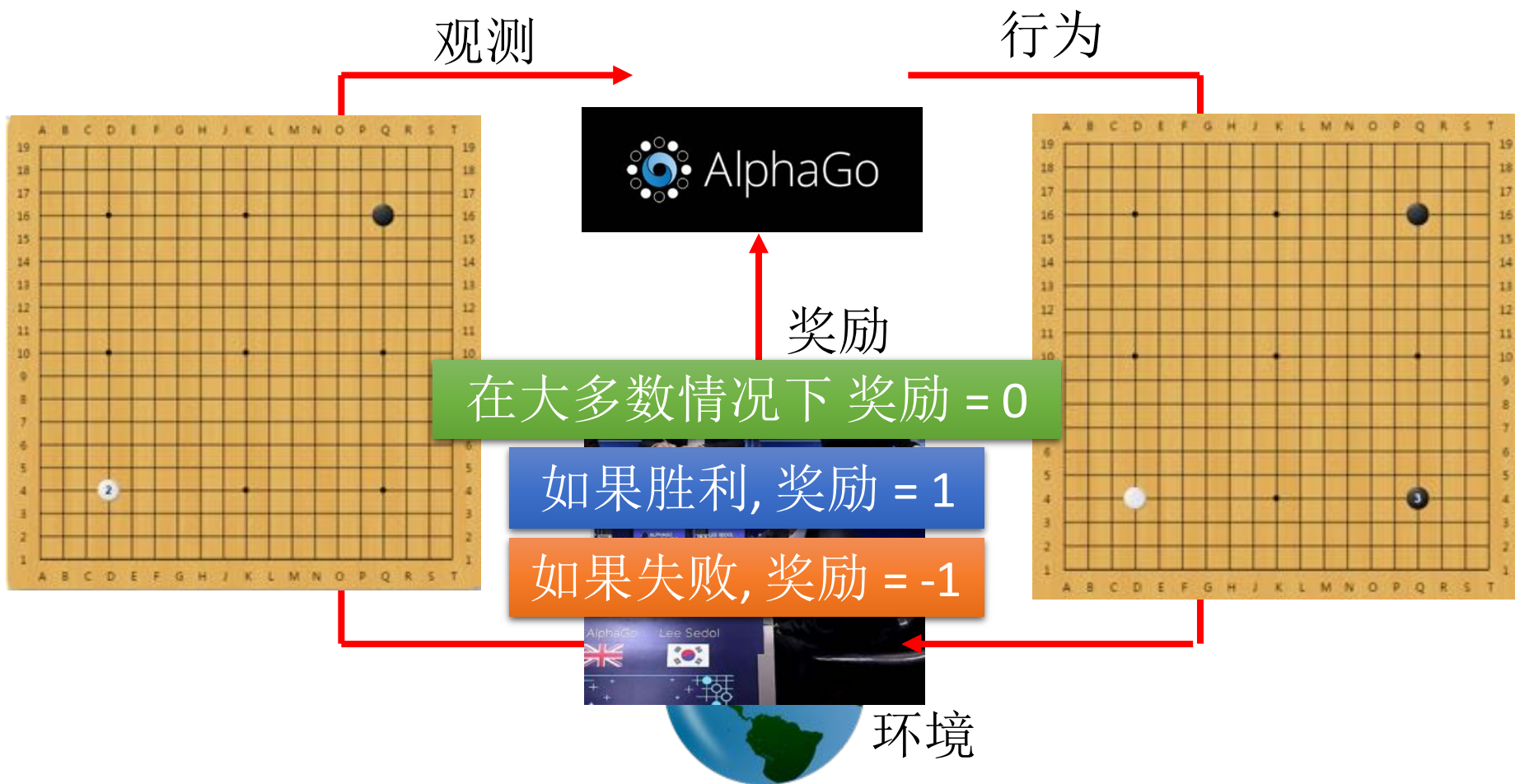


学习下围棋



学习下围棋

代理学会采取行动使回报的期望最大化。



学习下围棋

• 监督学习:

跟随老师学习



下一步:
“5-5”



下一步:
“3-3”

• 强化学习

从经验中学习

走一步



..... 继续走多步.....



胜利!

(两个代理一起对弈.)

Alpha Go 包含 监督学习 + 强化学习.

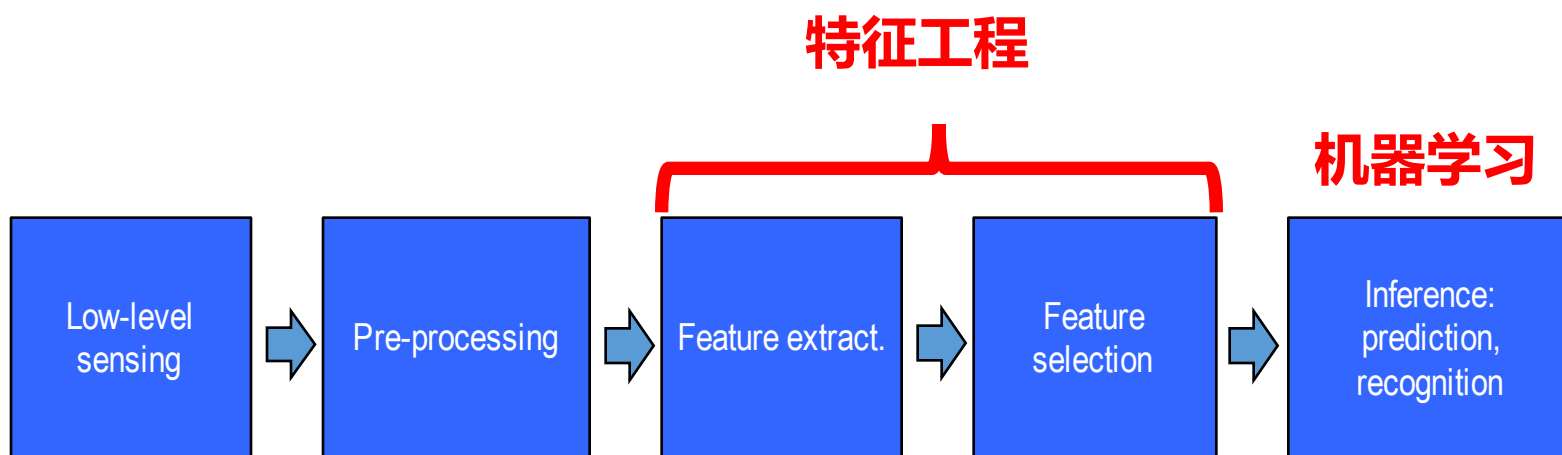
深度强化学习



Deep Reinforcement Learning: $AI = RL + DL$

模式识别

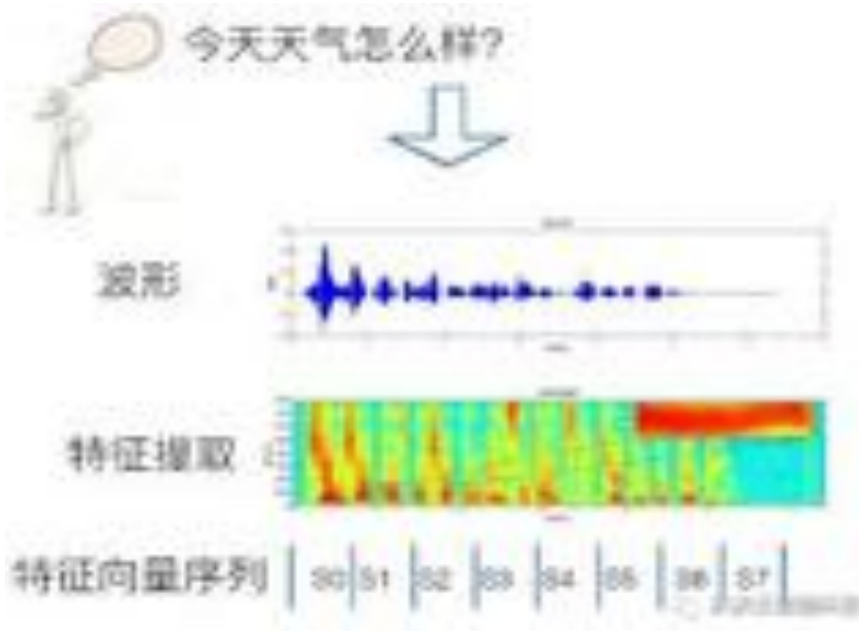
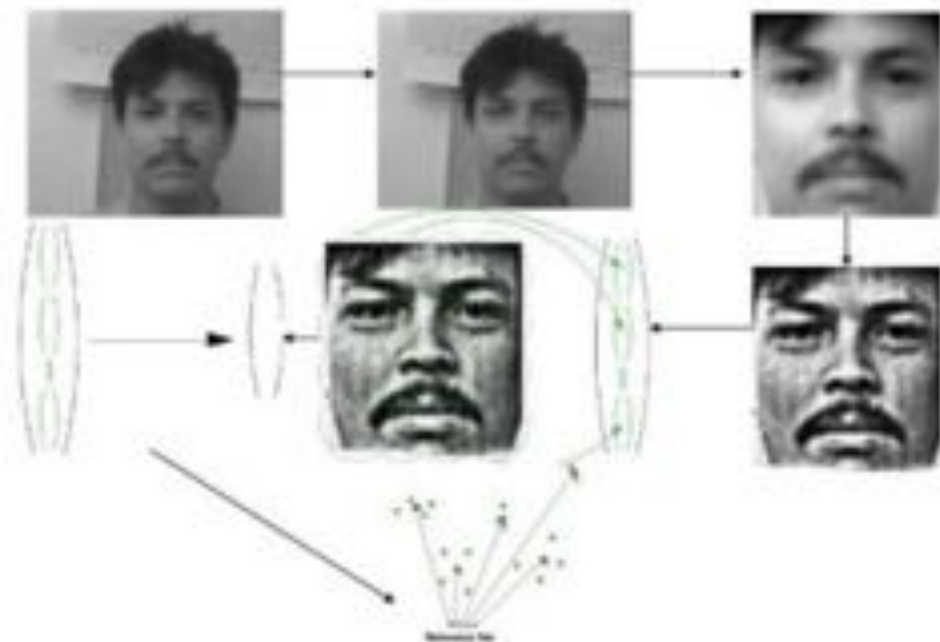
- 模式识别（英语：Pattern Recognition），就是通过计算机用数学技术方法来研究模式的自动处理和判读。



模式识别

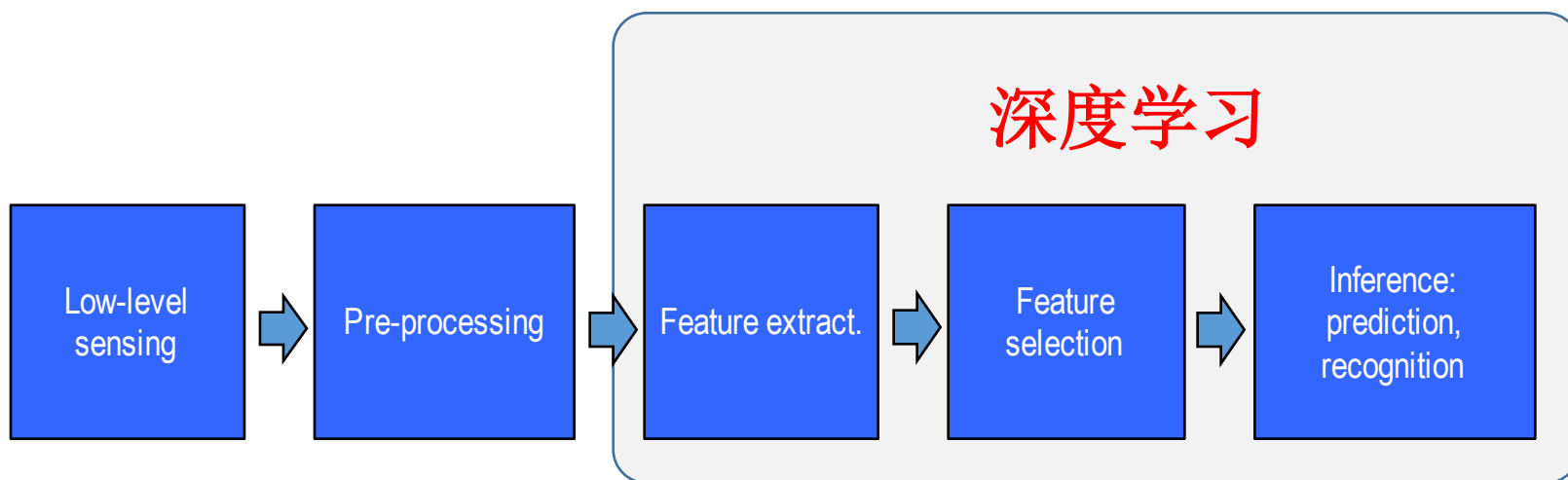
• 模式识别的任务

- 计算机视觉
 - 医学影像分析
 - 光学文字识别
- 自然语言处理
 - 语音识别
 - 手写识别
- 生物特征识别
 - 人脸识别
 - 指纹识别
 - 虹膜识别
- 文件分类
- 互联网搜索引擎
- 信用评分
- 测绘学
 - 摄影测量与遥感学



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

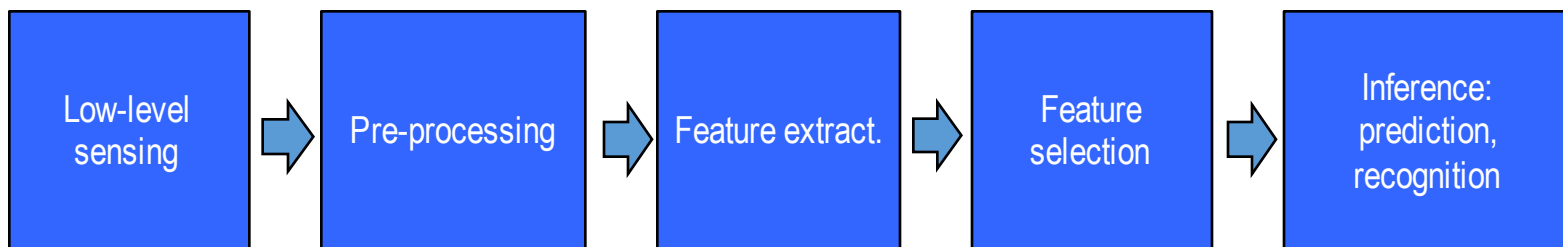
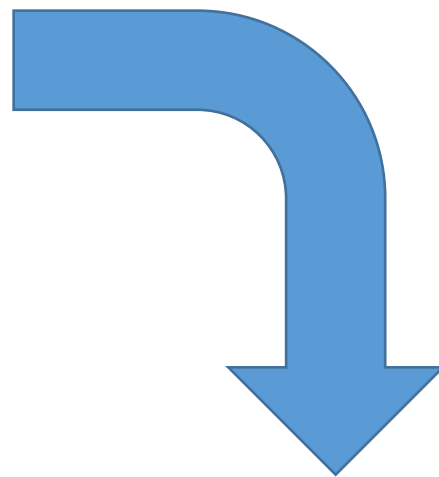
如：预测某新客户是否会在一家商场消费100元以上？

- 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。

什么是数据挖掘

• 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？
-



什么是数据挖掘

国务院关于印发打赢蓝天保卫战三年行动计划的通知

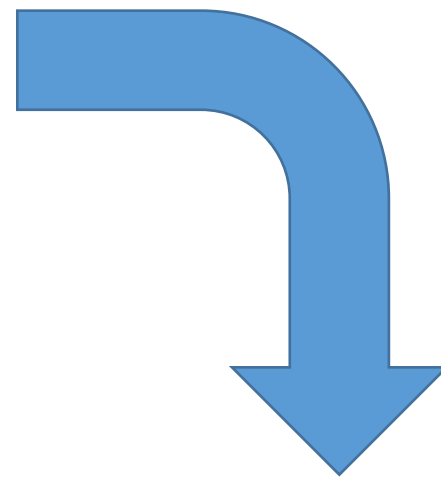
国发〔2018〕22号

各省、自治区、直辖市人民政府，国务院各部委，各直属机构：
现将《打赢蓝天保卫战三年行动计划》印发给你们，请认真贯彻执行。

国务院

2018年6月27日

(此件公开发布)

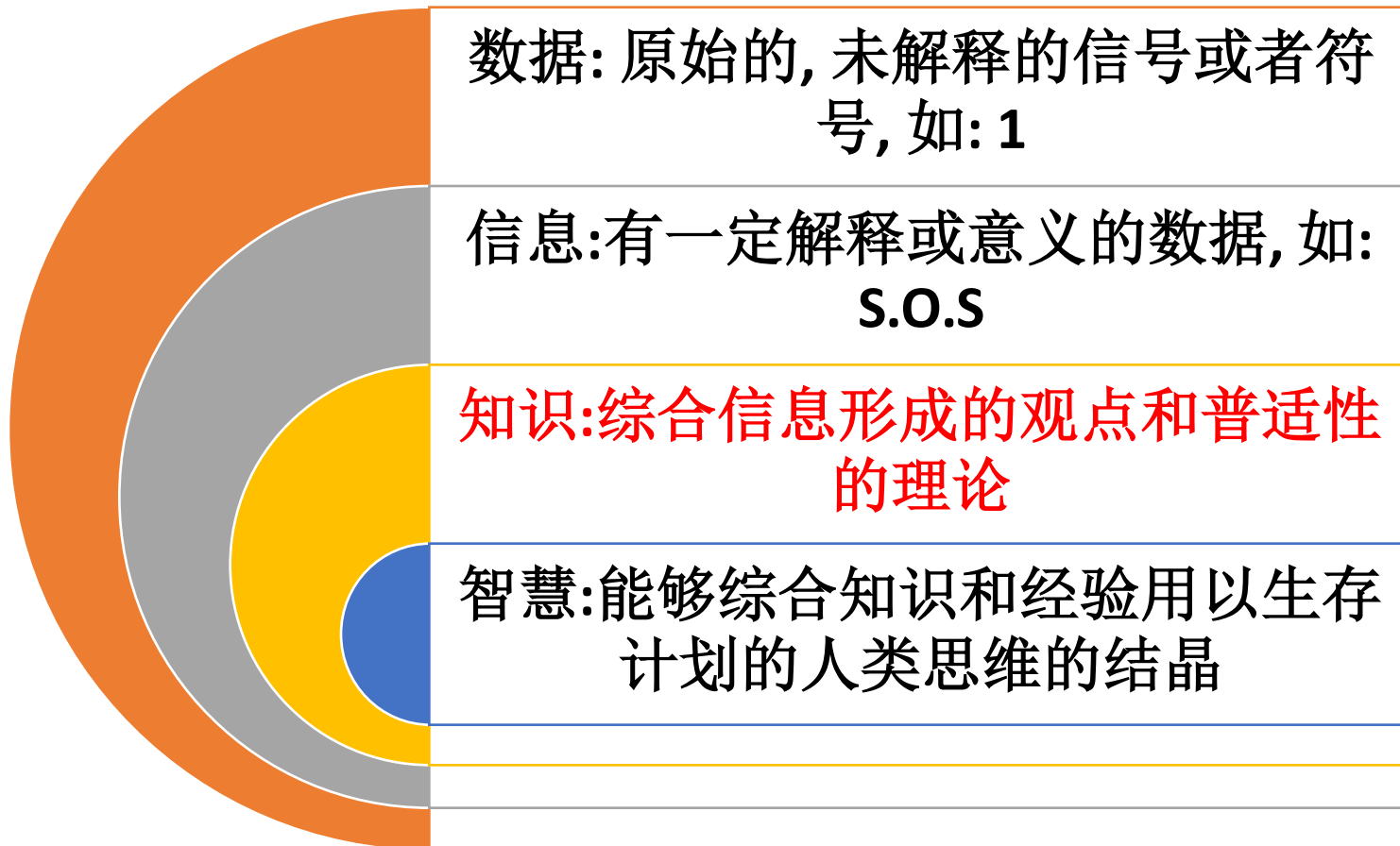


PM2.5预测



数据挖掘的核心任务是**知识发现**

- Knowledge Discovery in Database (KDD)

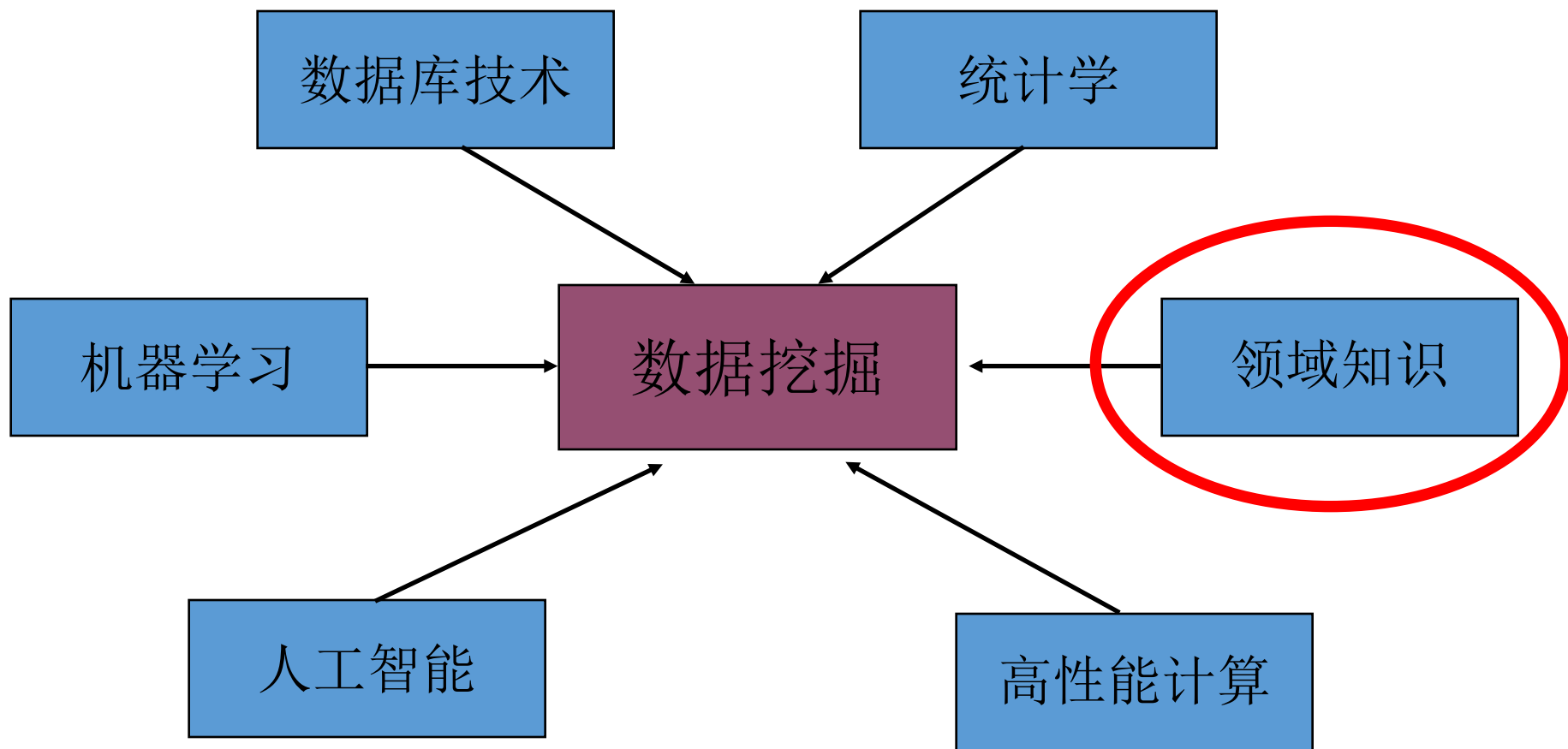


一句话总结数据挖掘

**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？

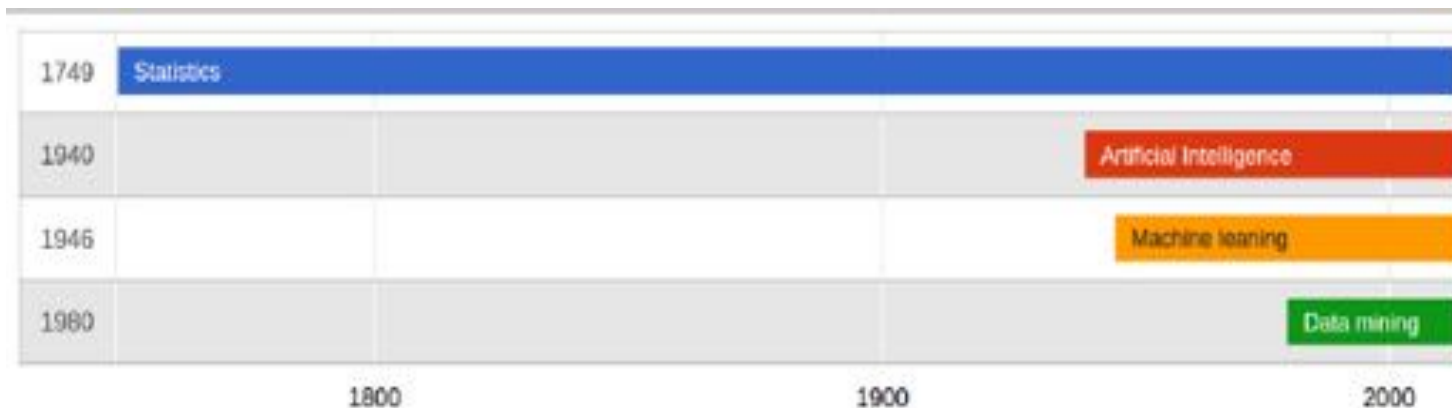
- 人工智能（Artificial Intelligence），英文缩写为AI。它是研究、开发用于**模拟、延伸和扩展人的智能**的理论、方法、技术及应用系统的一门新的技术科学。

统计学——1749年

人工智能——1940年

机器学习——1946年

数据挖掘——1980年



人工智能涉及多个学科



人工智能？

• 人工智能的三起两落

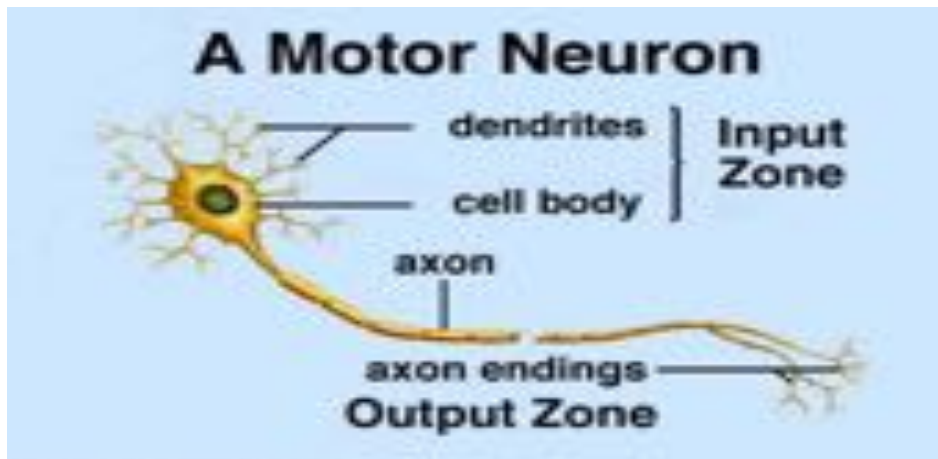


深度学习

神经元模型

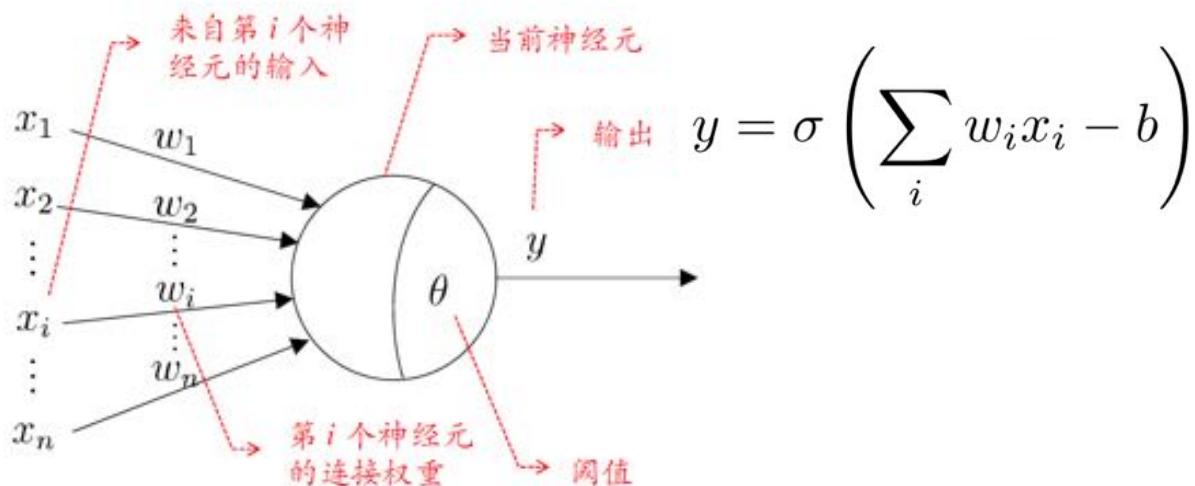
- 生物神经元

- 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。



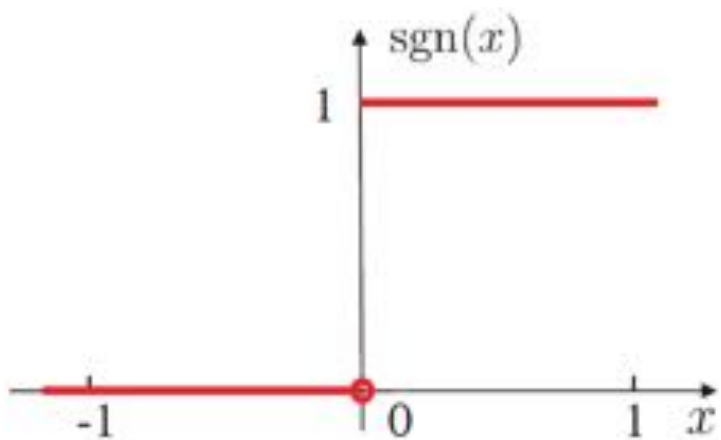
神经元模型

- **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



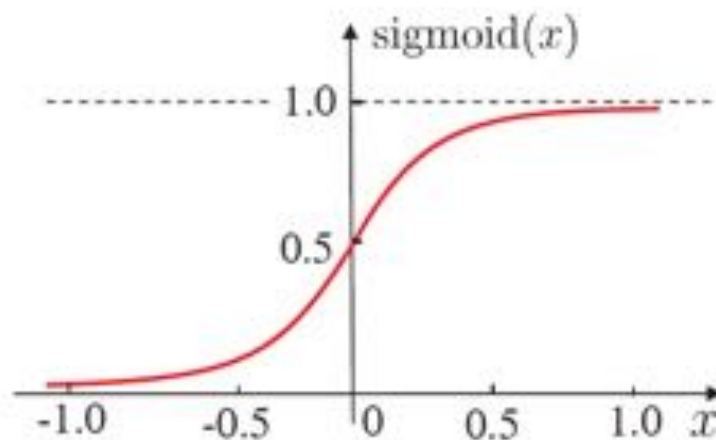
神经元模型

- 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



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

(a) 阶跃函数



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

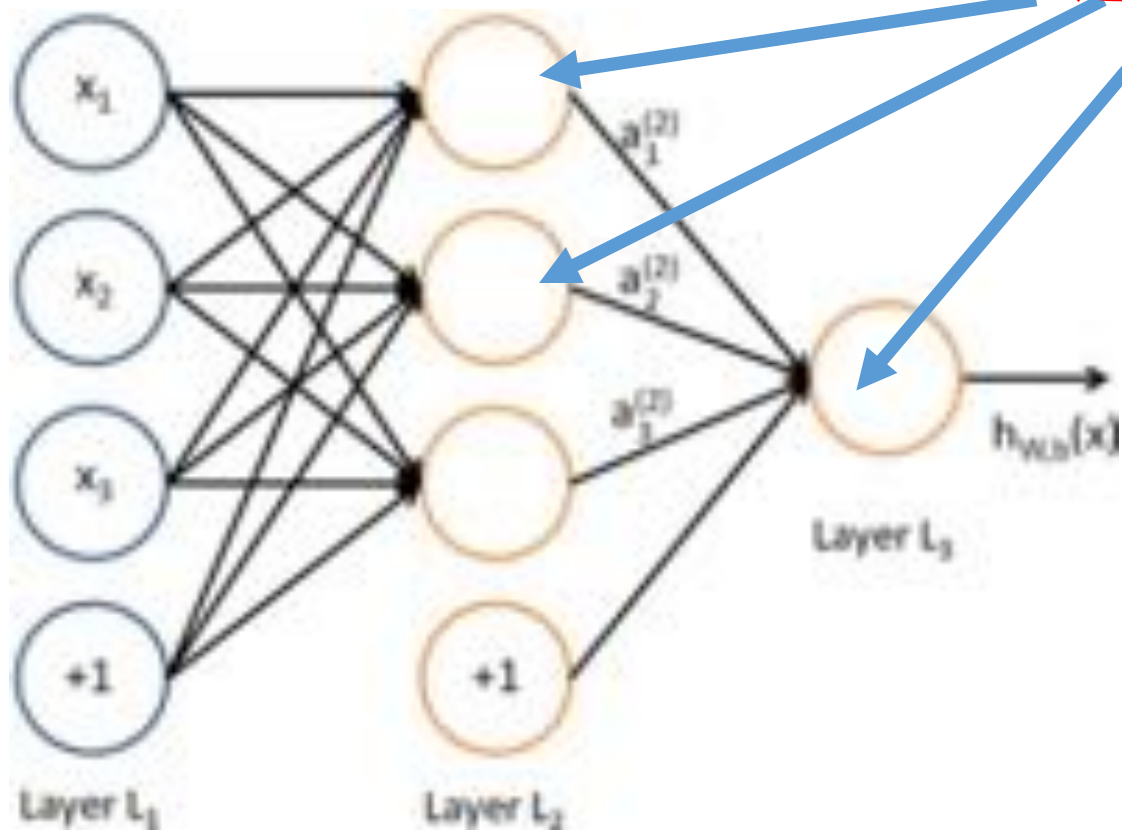
(b) Sigmoid 函数

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

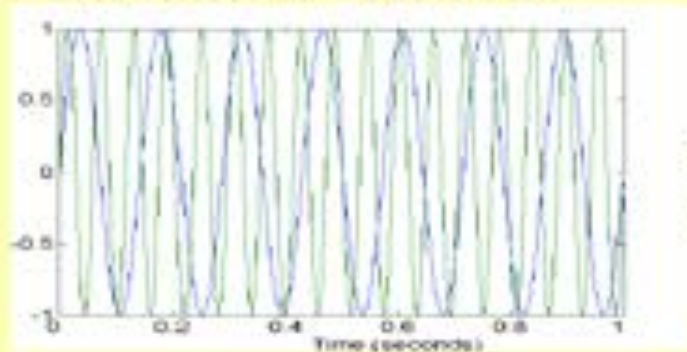


输入层

隐藏层

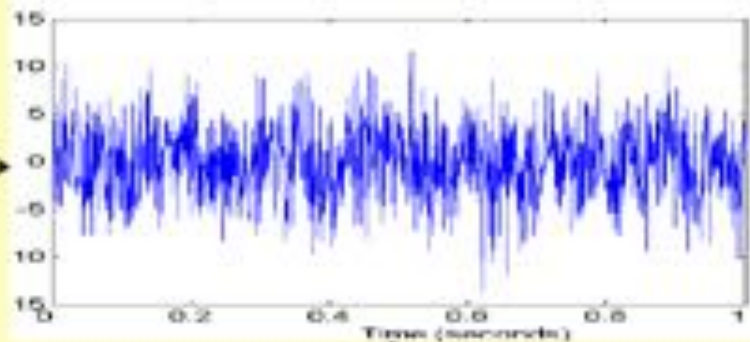
输出层

自动化的特征学习方法

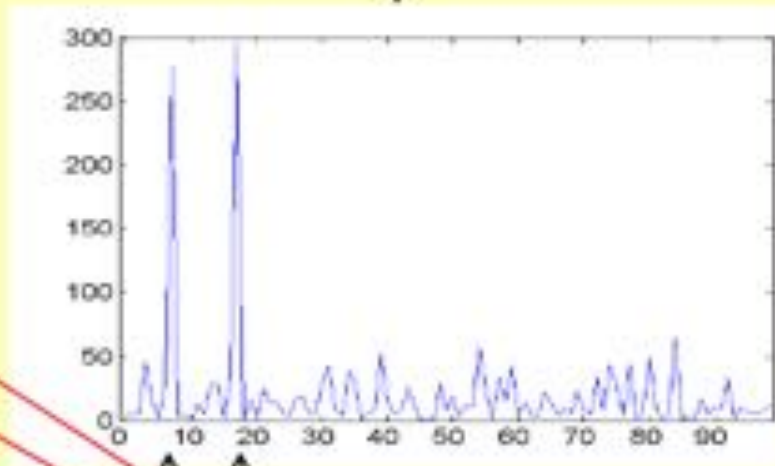


Two Sine Waves
(7个/s + 17个/s)

Two Sine Waves
+ Noise



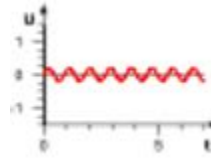
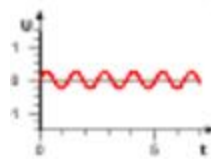
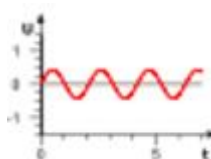
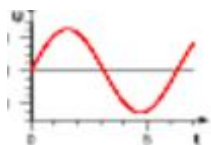
傅里叶变换



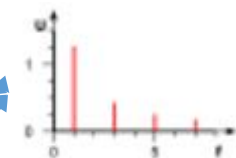
7 17

自动化的特征学习方法

字典



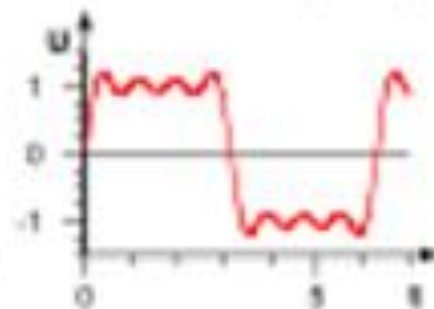
d



f

特征

\approx

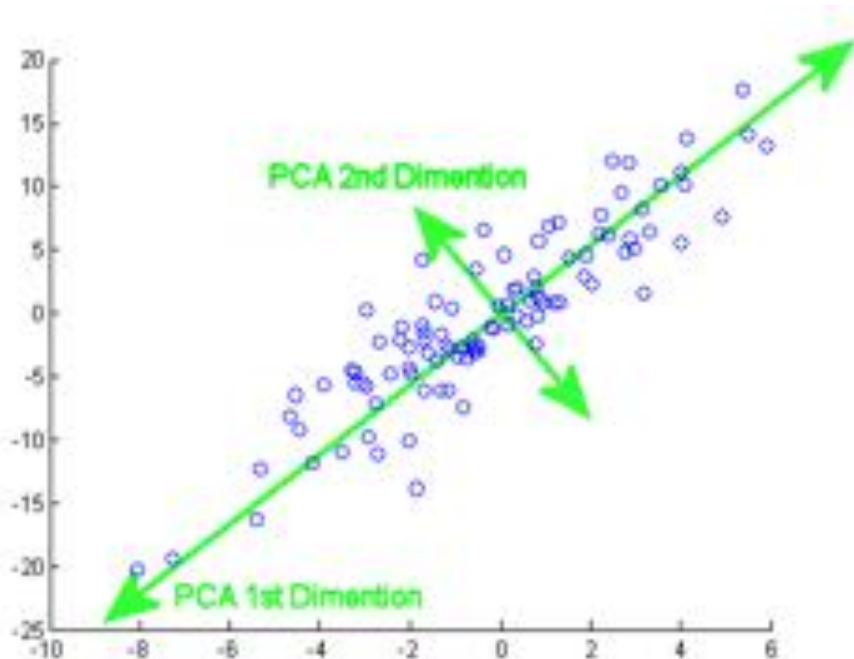


x

对象

自动化的特征学习方法

- PCA 主成分分析



\approx



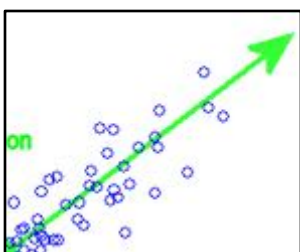
\times



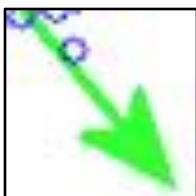
自动化的特征学习方法

- PCA 主成分分析

字典



d



2×2



$2 \times N$

f

特征

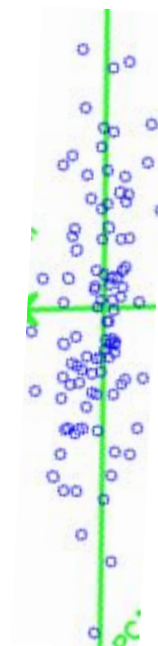
\approx



$N \times 2$

x

对象



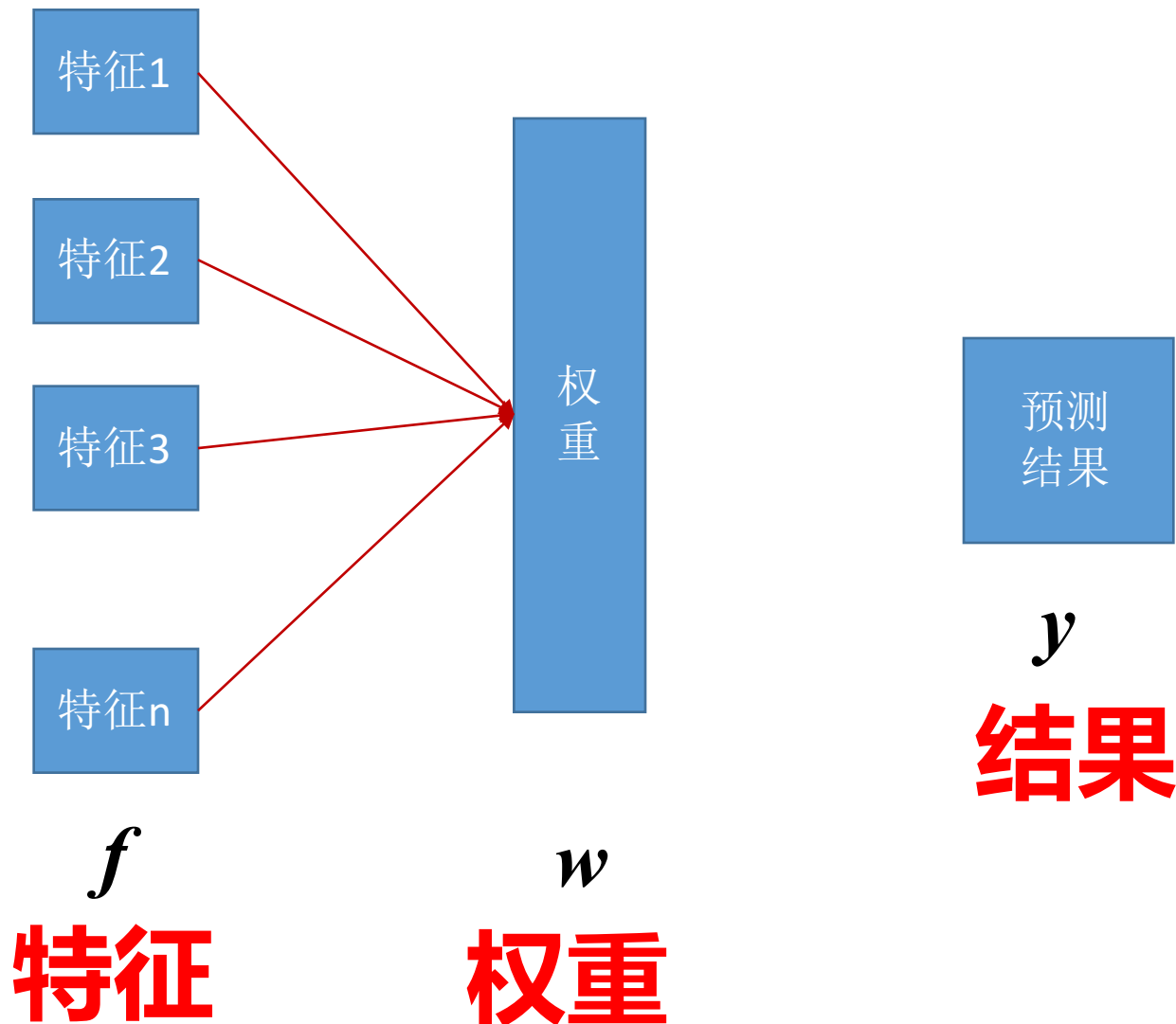
自动化的特征学习方法

- 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



结合起来看


$$\text{字典} \times \text{特征} \approx \text{数据}$$

$$\text{数据} \times \text{字典}^{-1} \approx \text{特征}$$

$$\text{特征} \times \text{权重} \approx \text{结果}$$

数据

\times

字典⁻¹

\approx

特征

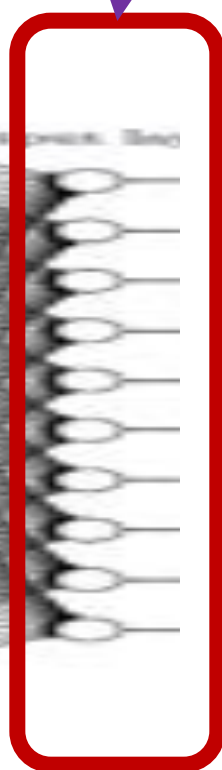
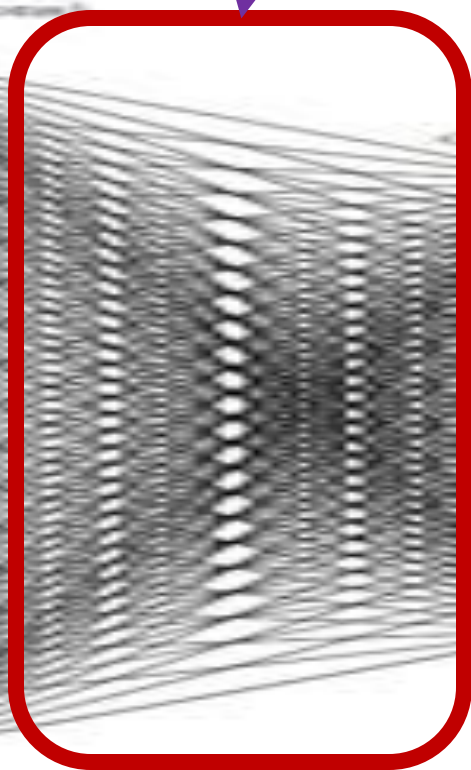
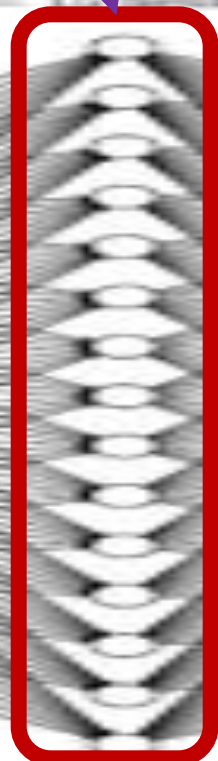
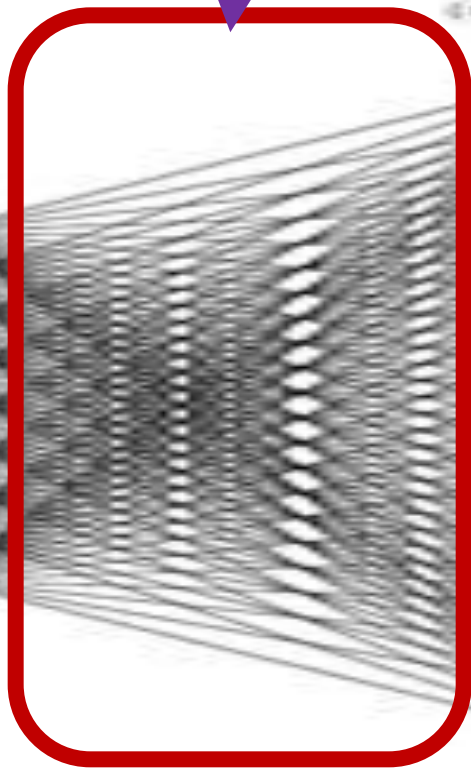
特征

\times

权重

\approx

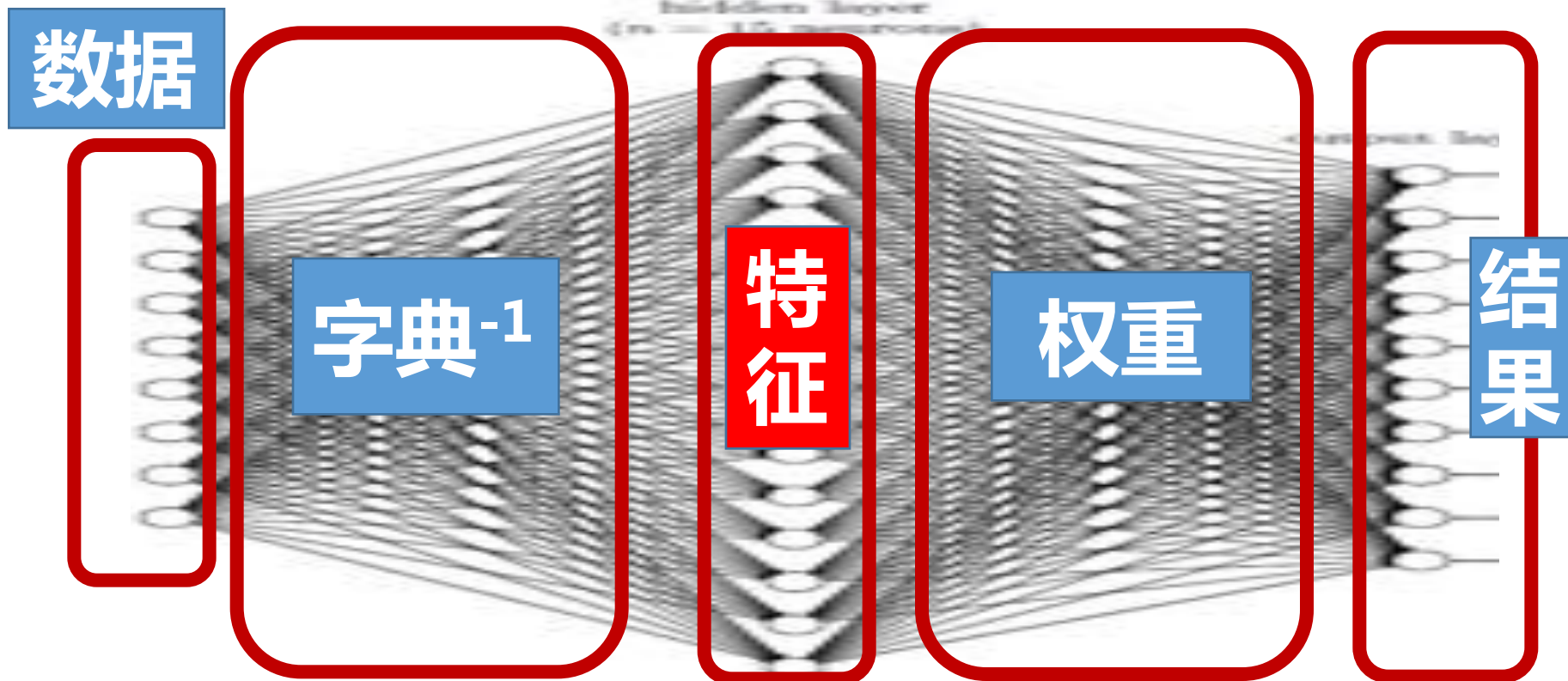
结果



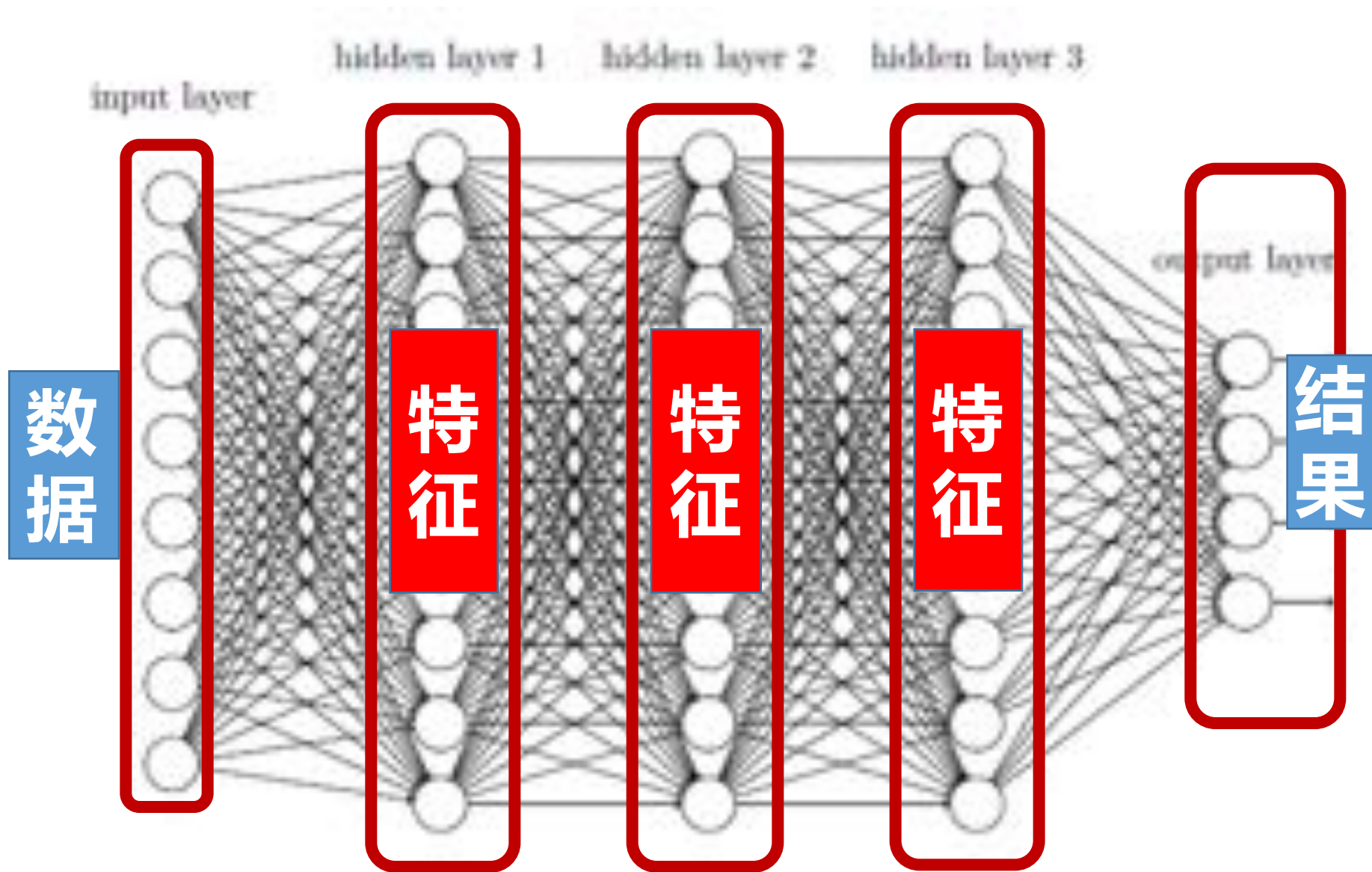
Representation Learning

• 表征学习

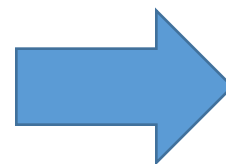
- 特征和字典都通过学习获得，对数据有很强的**适应性**
- 学习的目标是优化一个预测问题，具有很强的**针对性**



多层神经网络

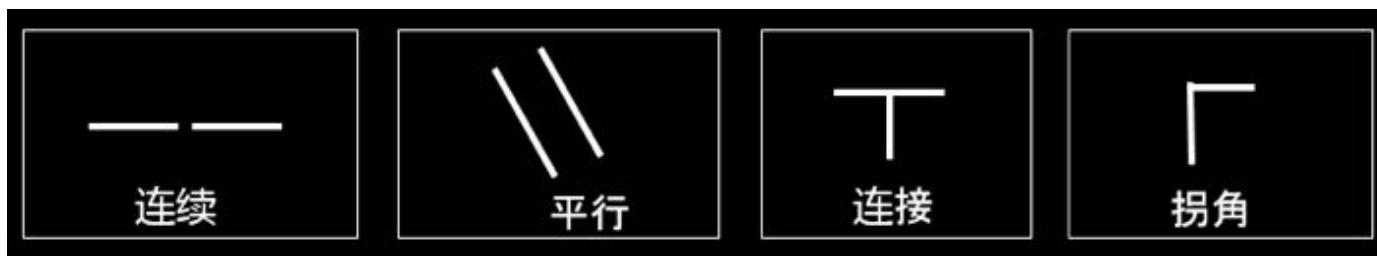


分层的特征提取



马云

- **输入**数据：像素
- **初层**特征：边缘



- **中层**特征：形状

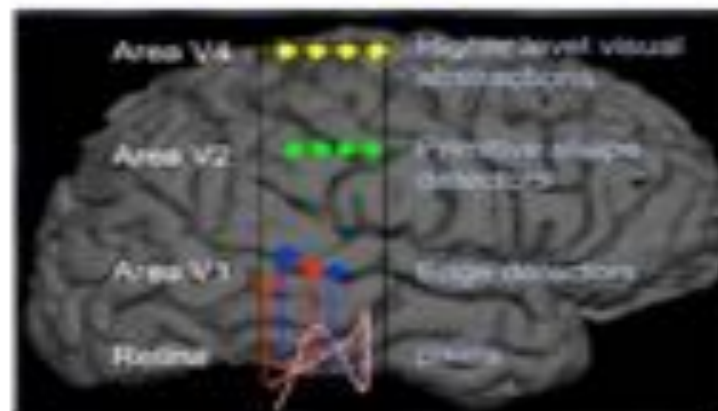
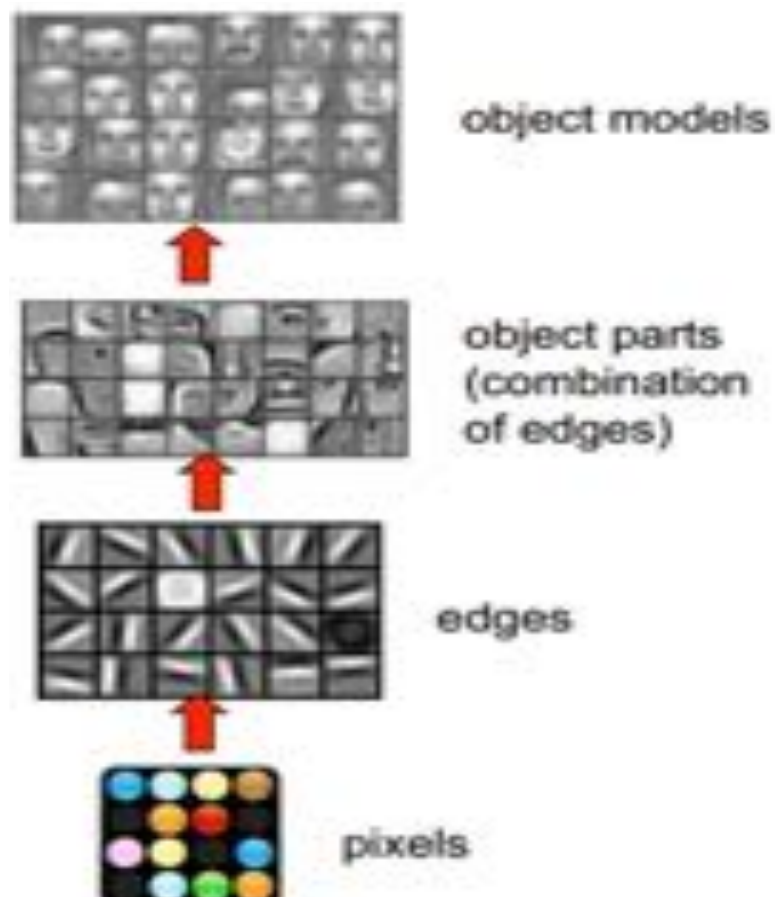


- **高层**特征：物体部件

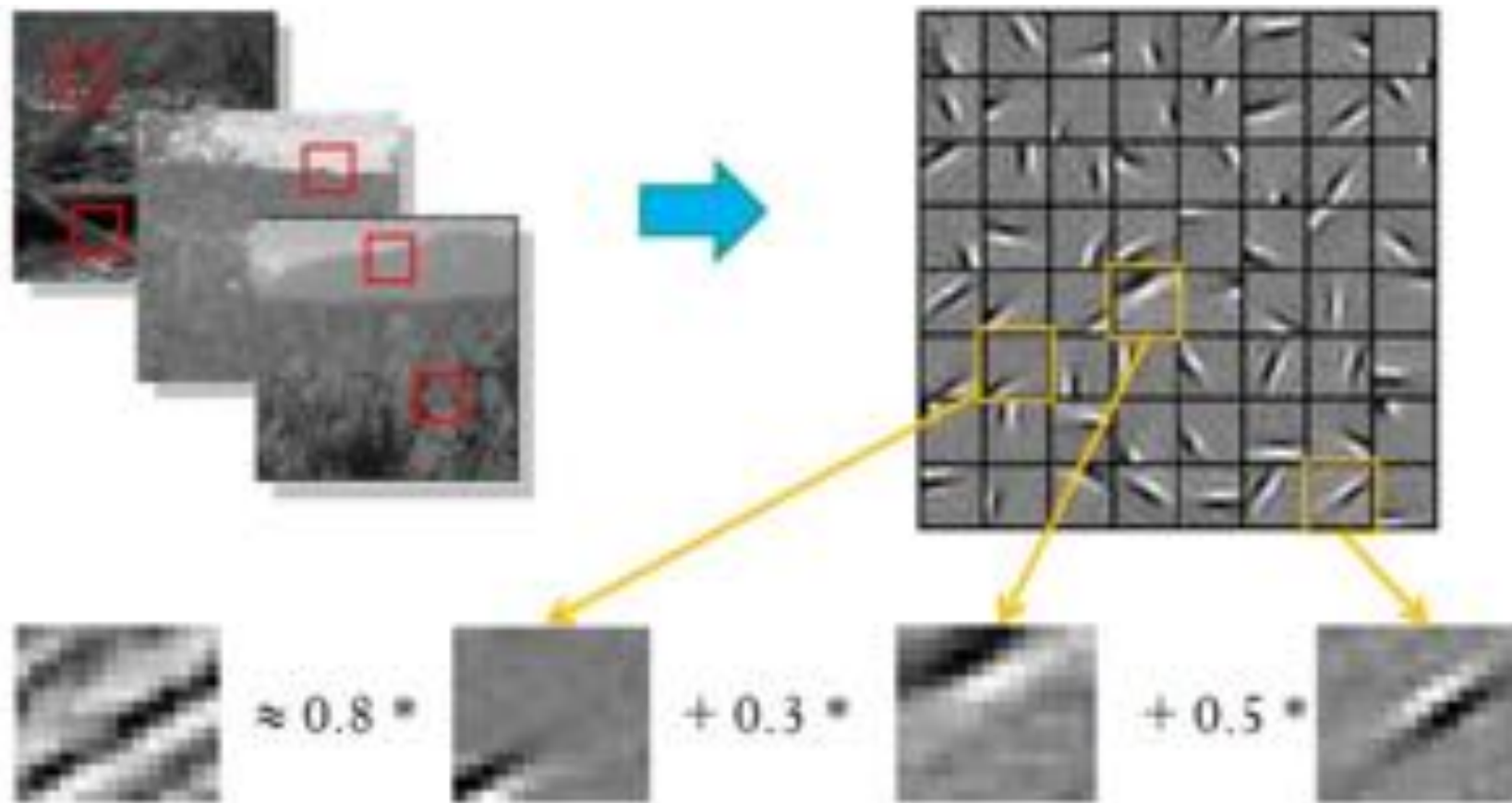


分层的特征提取

- 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

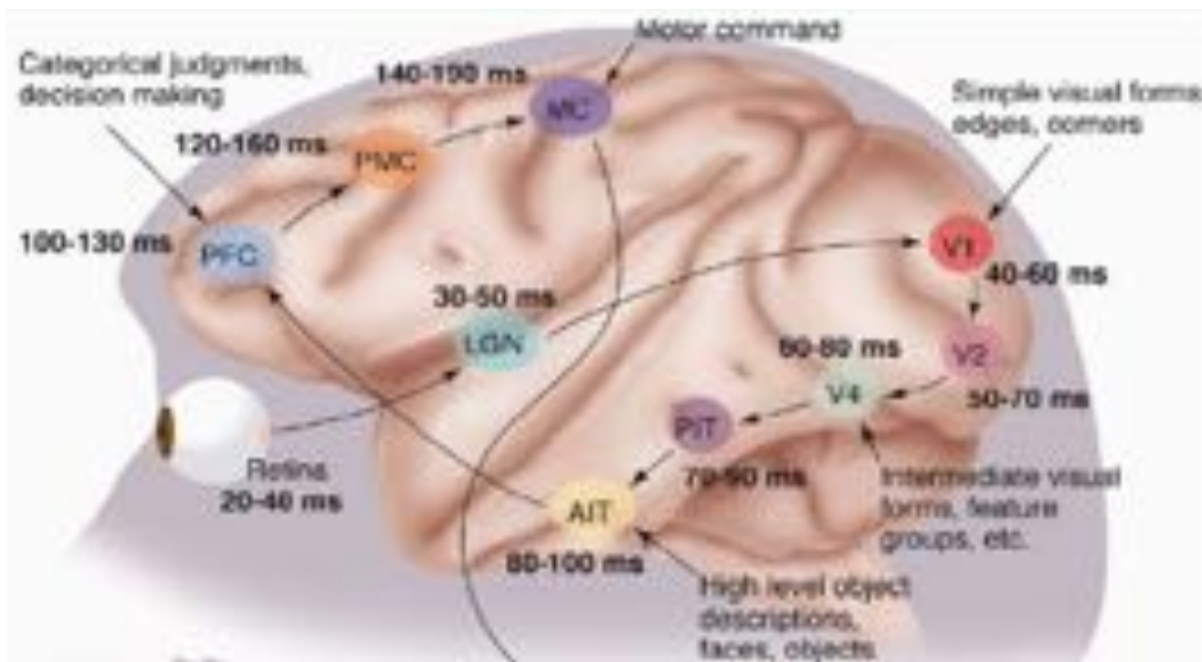


$[a_1, \dots, a_{64}] = [0, 0, \dots, 0, \mathbf{0.8}, 0, \dots, 0, \mathbf{0.3}, 0, \dots, 0, \mathbf{0.5}, 0]$
(feature representation)

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

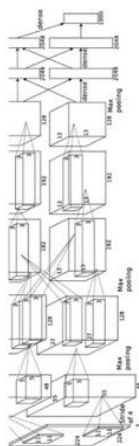


“深度” 竞赛

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)

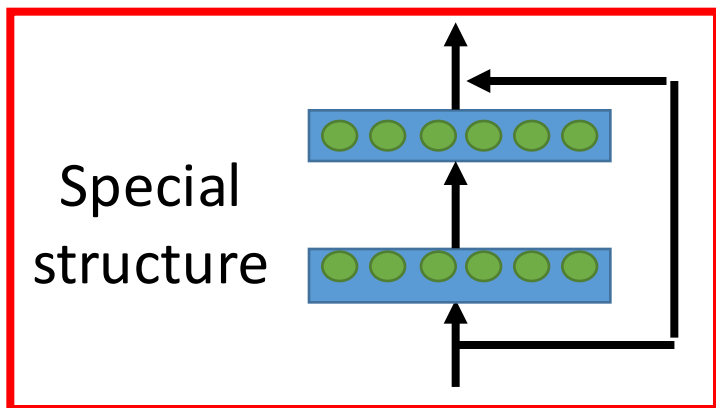
“深度” 竞赛



目瞪口呆

152 layers

162 layers



3.57%

16.4%

AlexNet
(2012)

7.3%

VGG
(2014)

6.7%

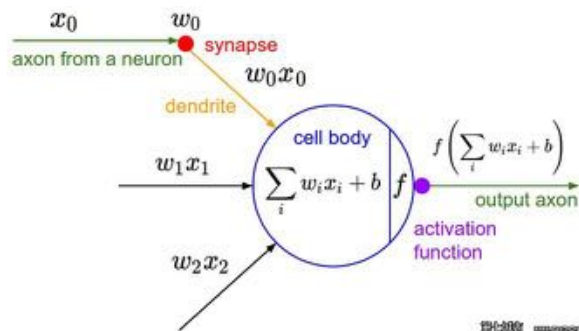
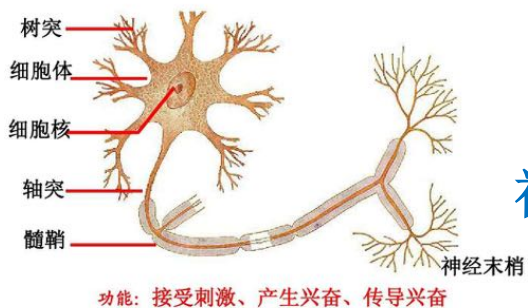
GoogleNet
(2014)

Residual Net
(2015)

迪拜
哈利法塔

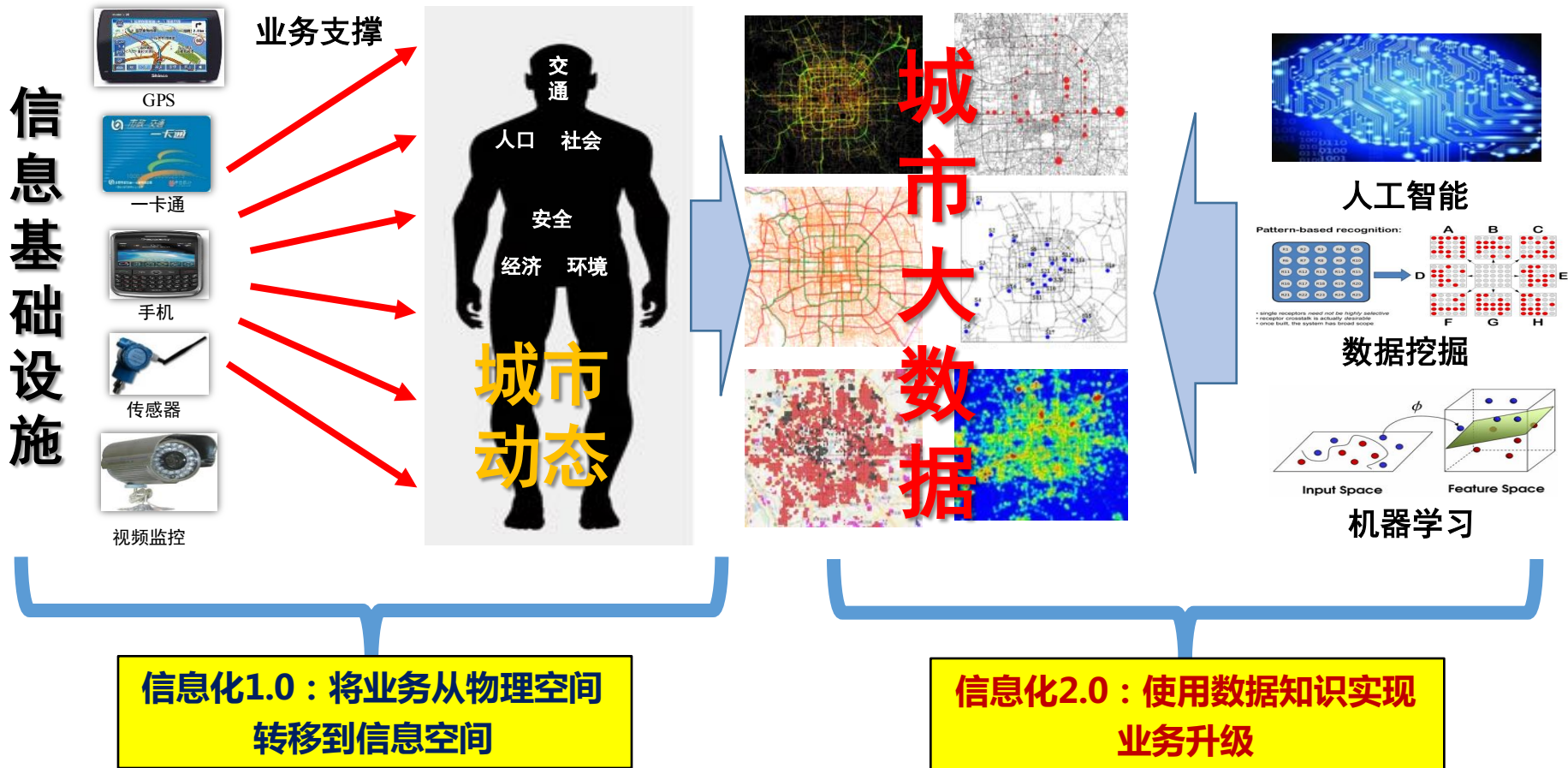
神经网络模型

• 不同动物神经元数量的对比

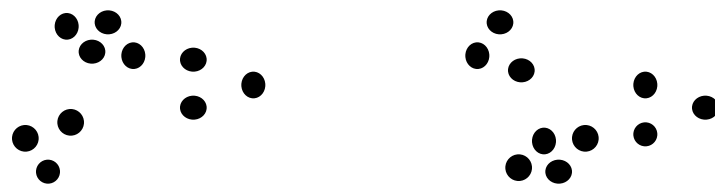


城市科学当中的应用

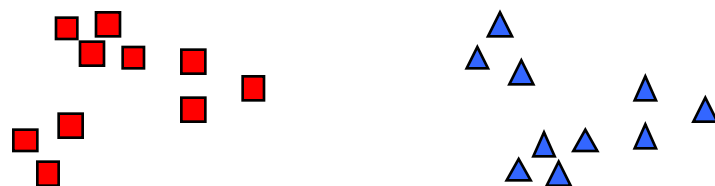
智慧城市—城市信息化2.0



无监督学习：聚类问题 Clustering



数据集

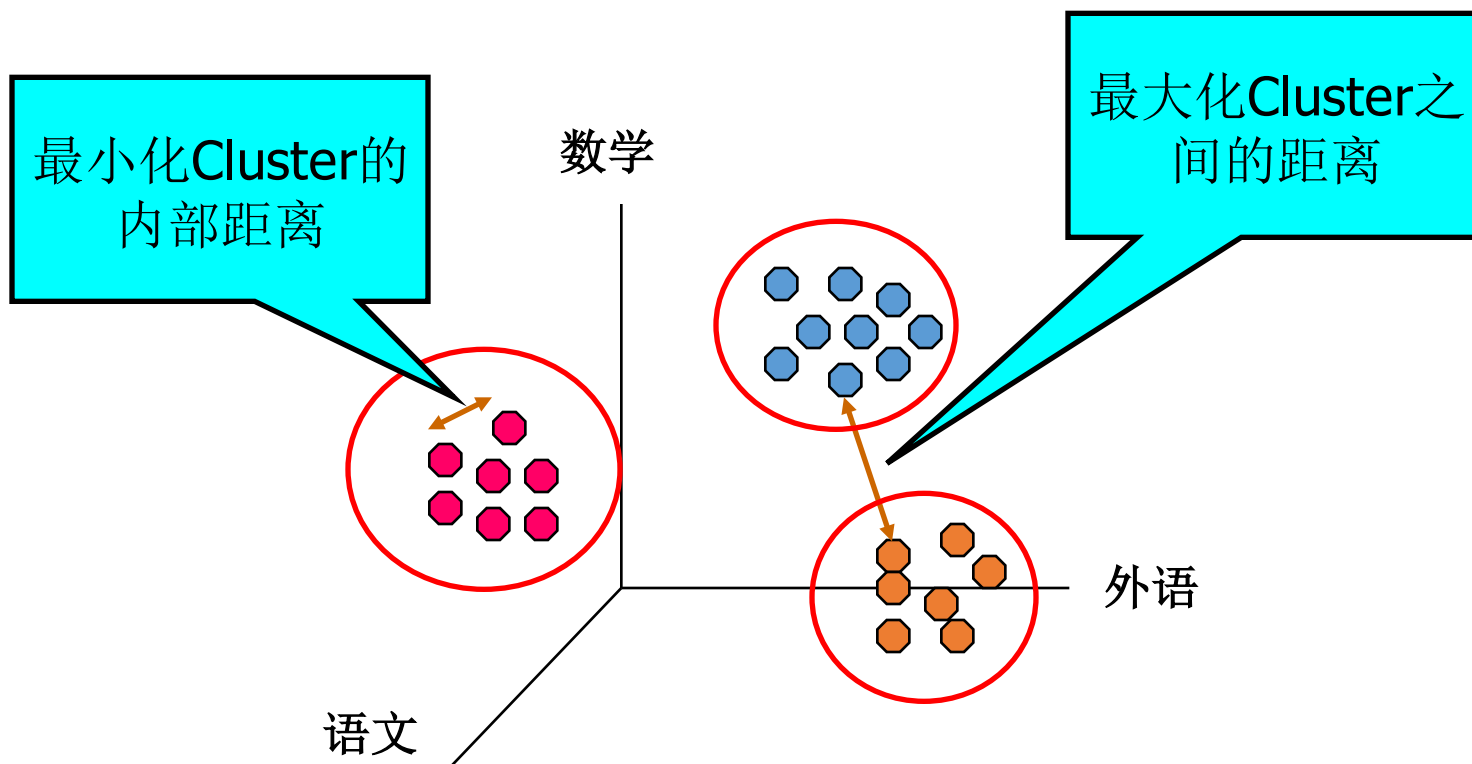


聚类结果

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**



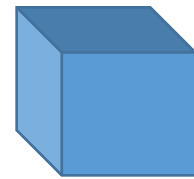
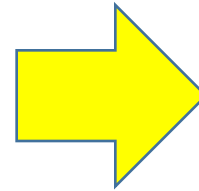
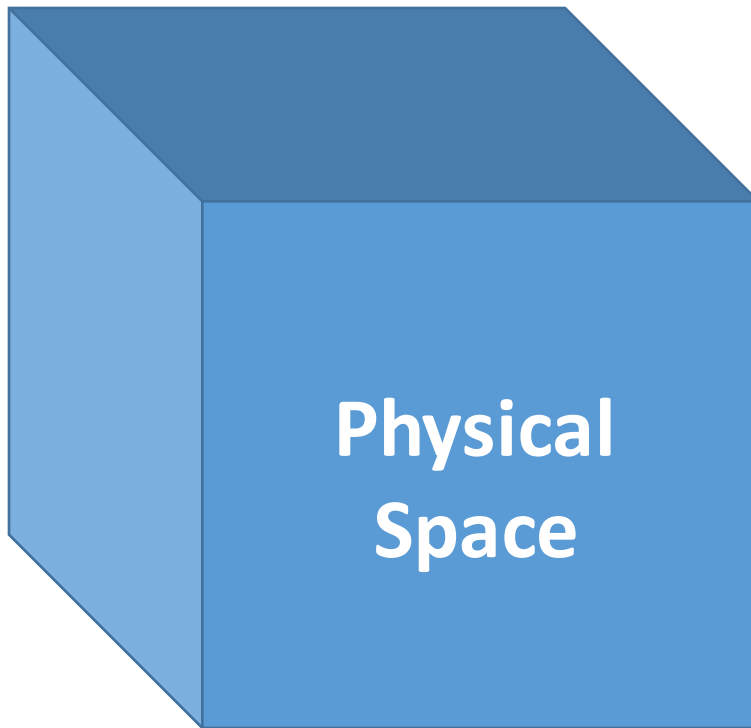
Handclap



Traffic Tidal

All contain patterns

Motivation



**Pattern
Space**

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

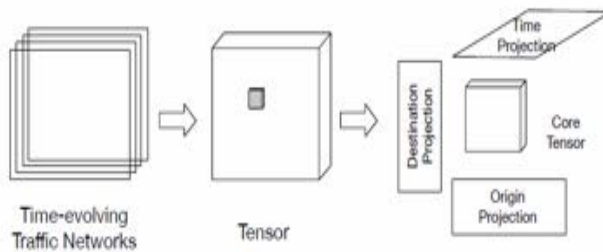


Beijing Map

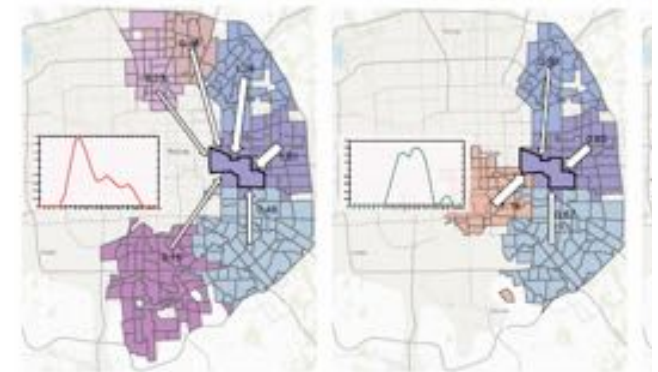


Beijing POI

Tools:



Output:

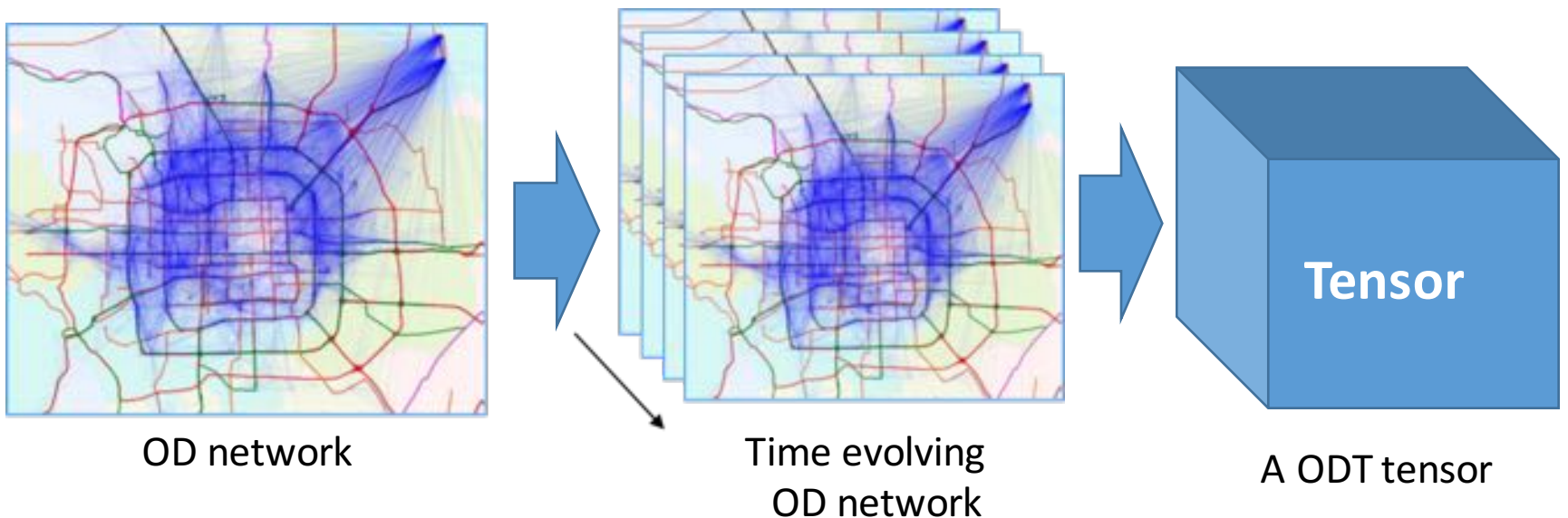


Spatio-temporal patterns

Regularized Non-negative Tensor Decomposition

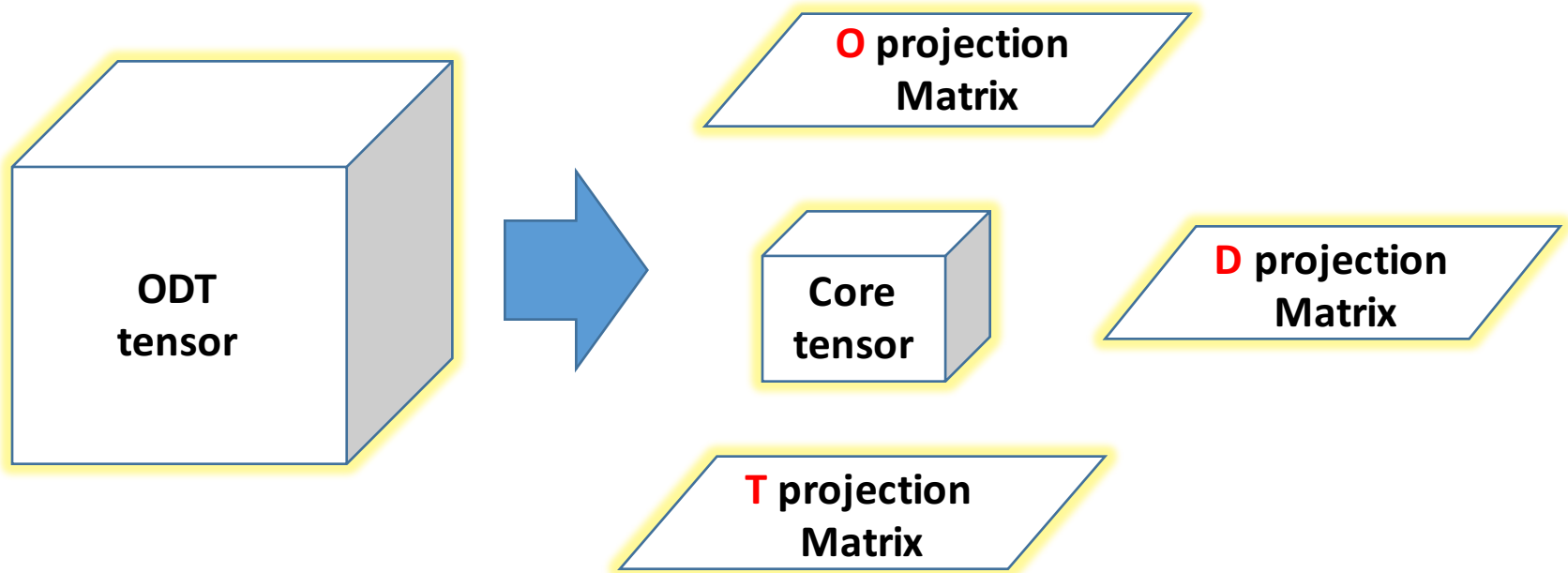
Problem definition

- **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.

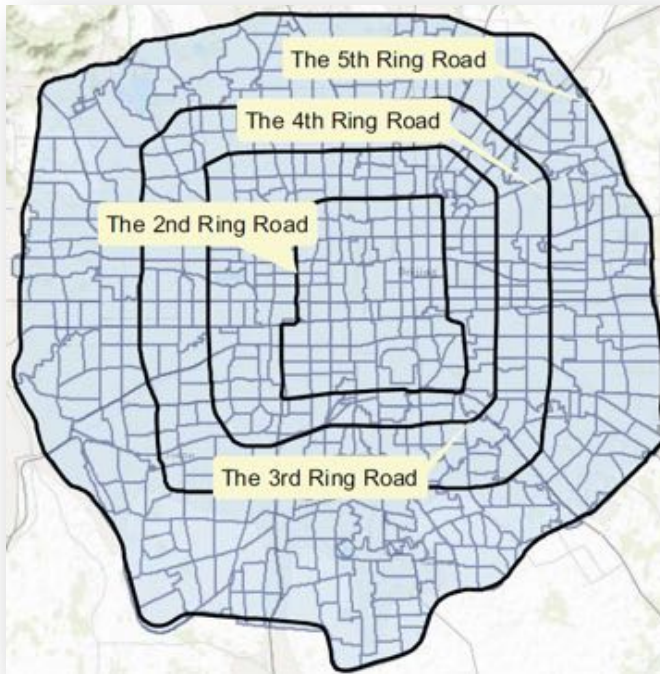


Problem definition

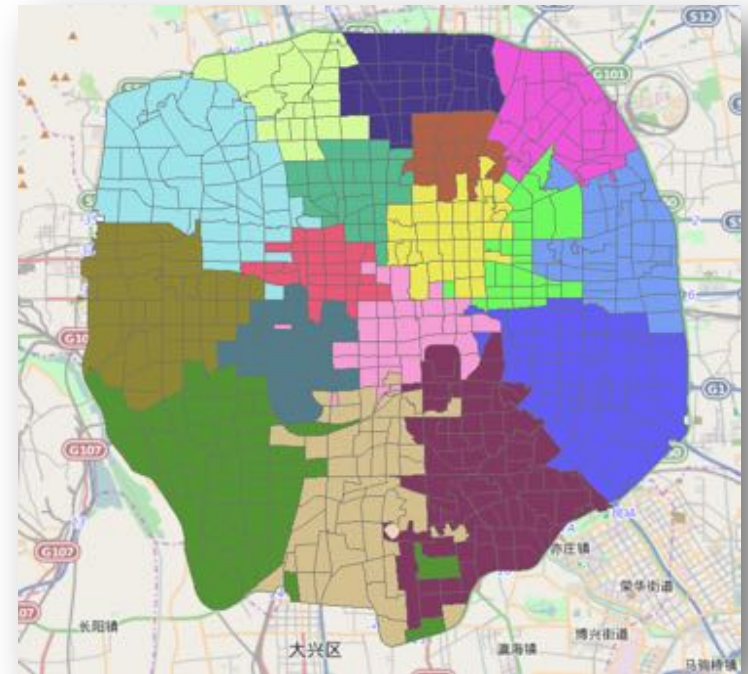
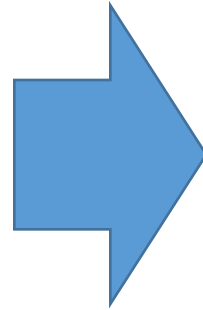
- Decompose the tensor as three projection matrixes and a core tensor



Spatial pattern projection

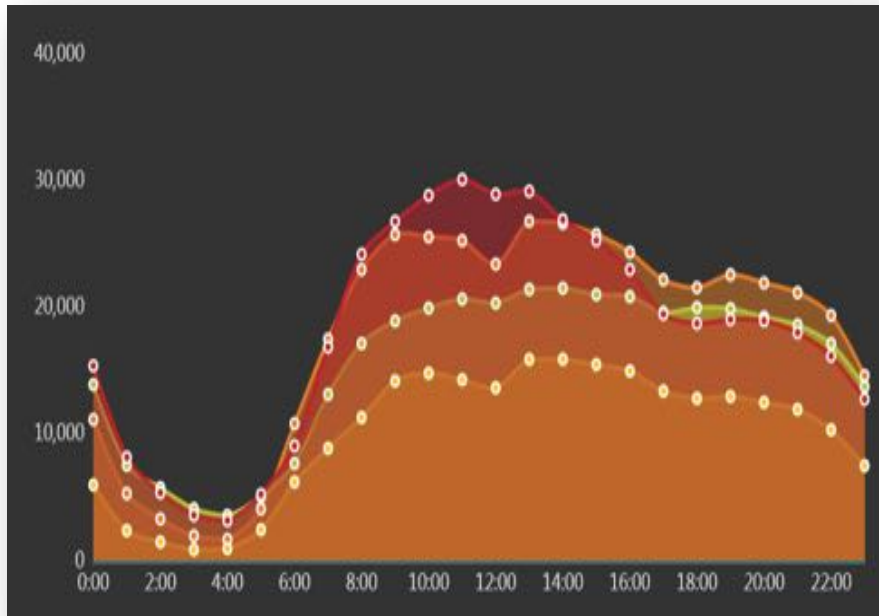


Urban zones

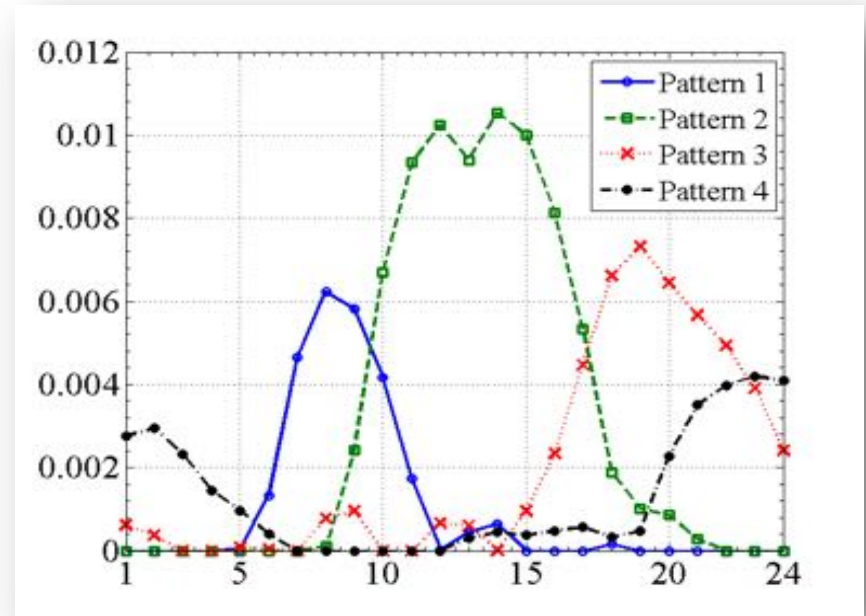


Urban communities

Temporal pattern projection



Urban traffic



Urban rhythm

Regularized Non-negative Tucker Decomposition

- Basic Tucker decomposition

$$\mathcal{X} \approx \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}$$

Core tensor
D Matrix
O Matrix
T Matrix

- Objective function (\mathcal{J}_1)

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

- 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

• **Objective function** $W_{ij} = \frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|}$

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 $s.t. \mathbf{C} \geq 0, \mathbf{O} \geq 0, \mathbf{D} \geq 0, \mathbf{T} \geq 0$
 - Core tensor: Only keep strong interactions
 - Projection matrix: make each cluster more meaningful and enhance uniqueness

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

Regularized Non-negative Tucker Decomposition

- Final objection function

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

Model traffic information

Model urban contextual information (POI)

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

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

Make results sparse

Make results non-negative

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

- Human mobility and urban context are jointly optimized !

Optimization

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

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

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

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

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



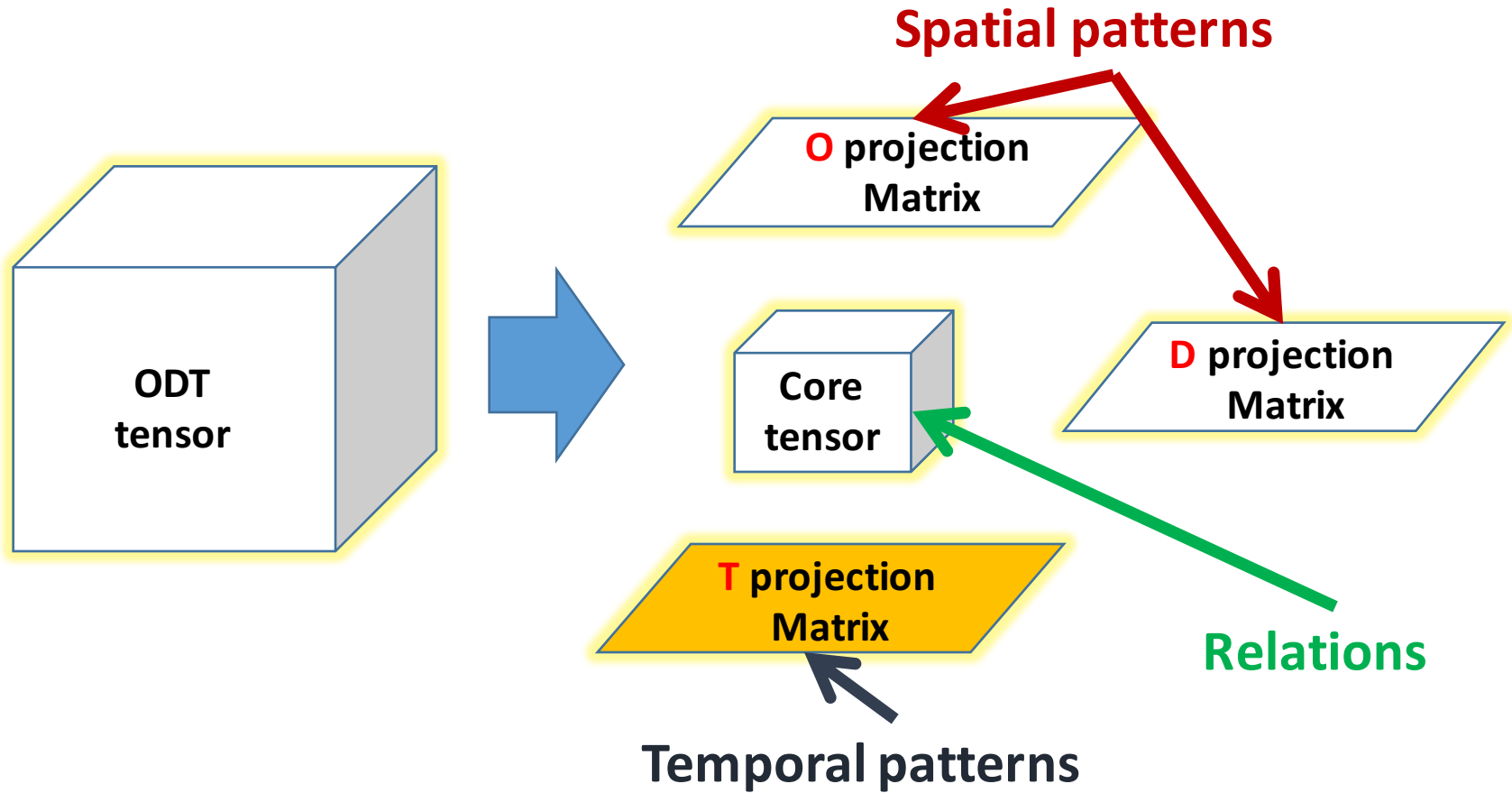
Experiments

Data

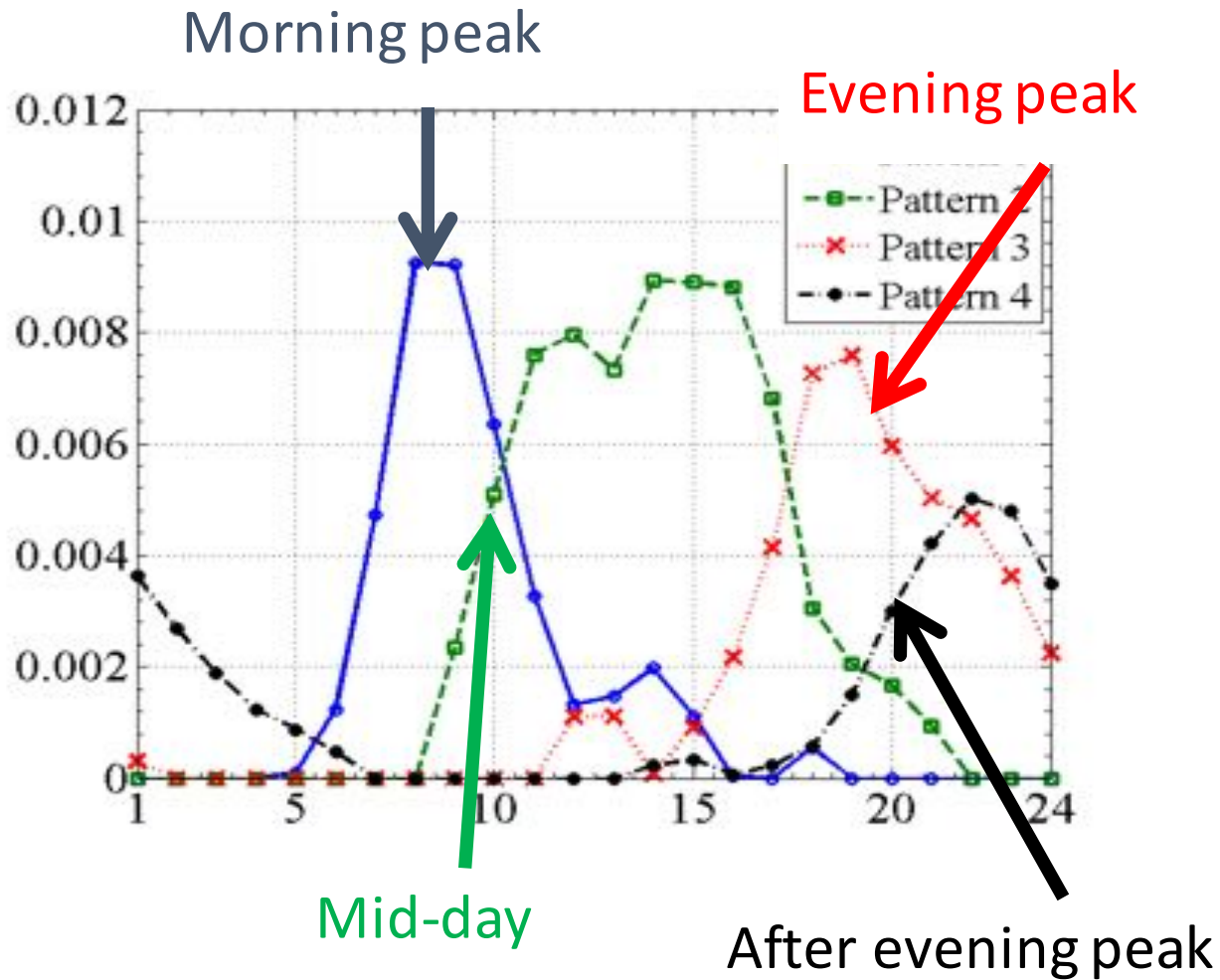
- Beijing GPS data
 - GPS trajectory of 20000 Beijing taxis, 2008 and 2012.
- Beijing traffic analysis zones map
 - 600 zones in the 5th ring road in Beijing
- Position of Interesting (POI) in Beijing
 - 380,000 points



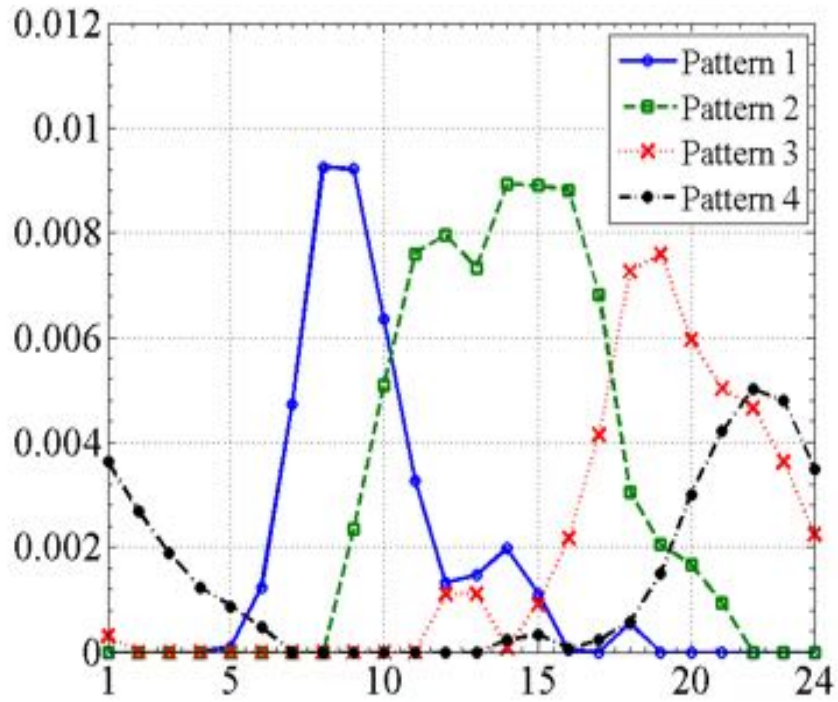
Beijing show case



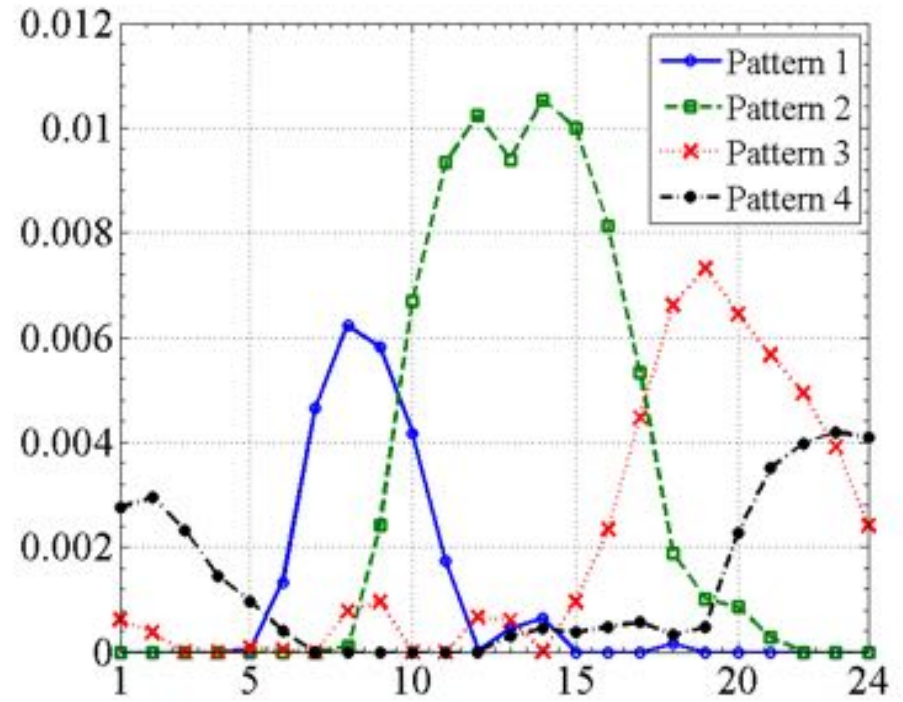
Temporal patterns



Temporal patterns

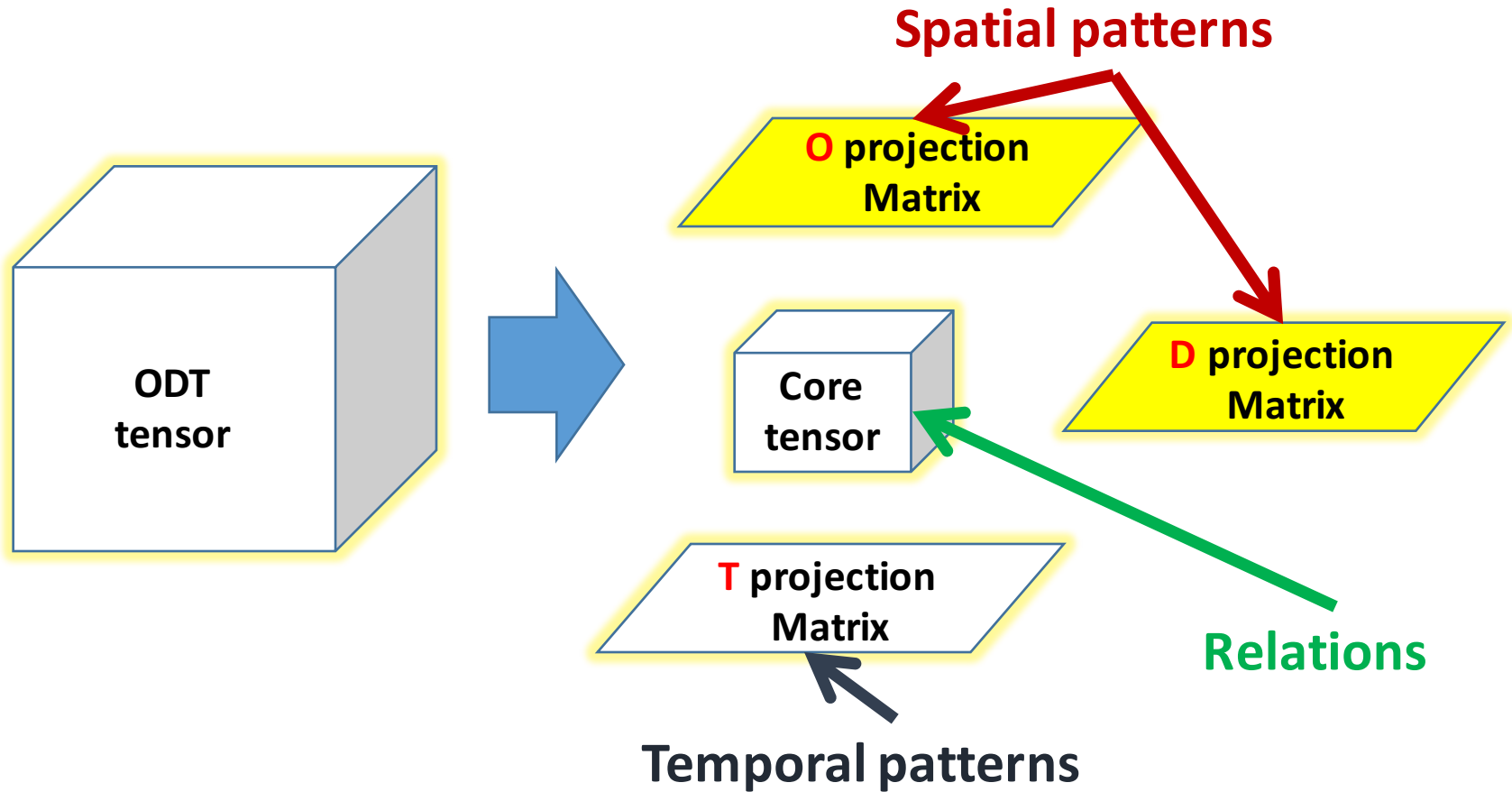


2008

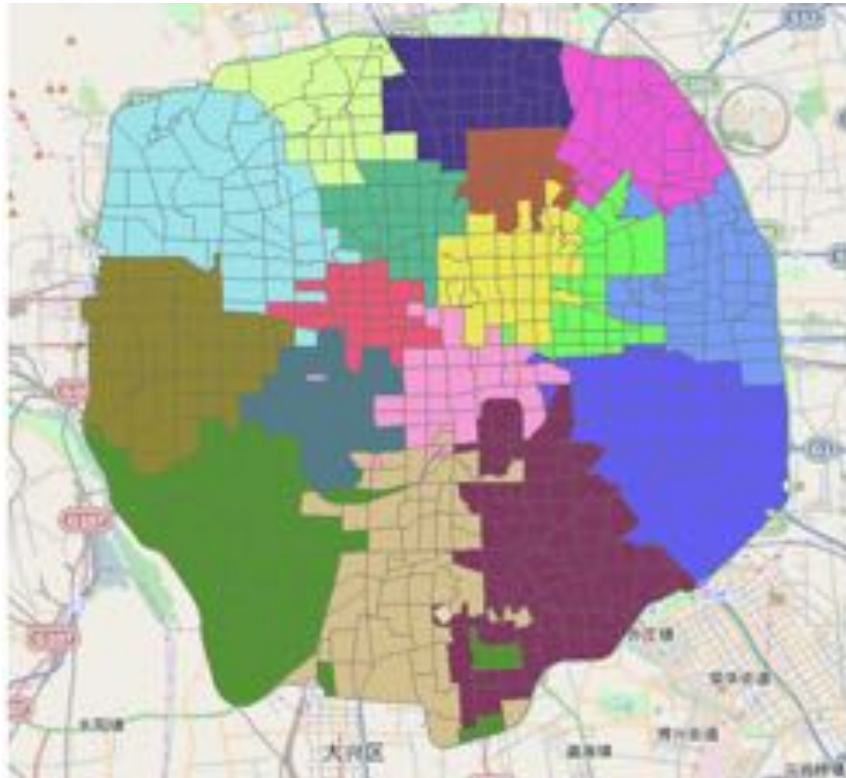


2012

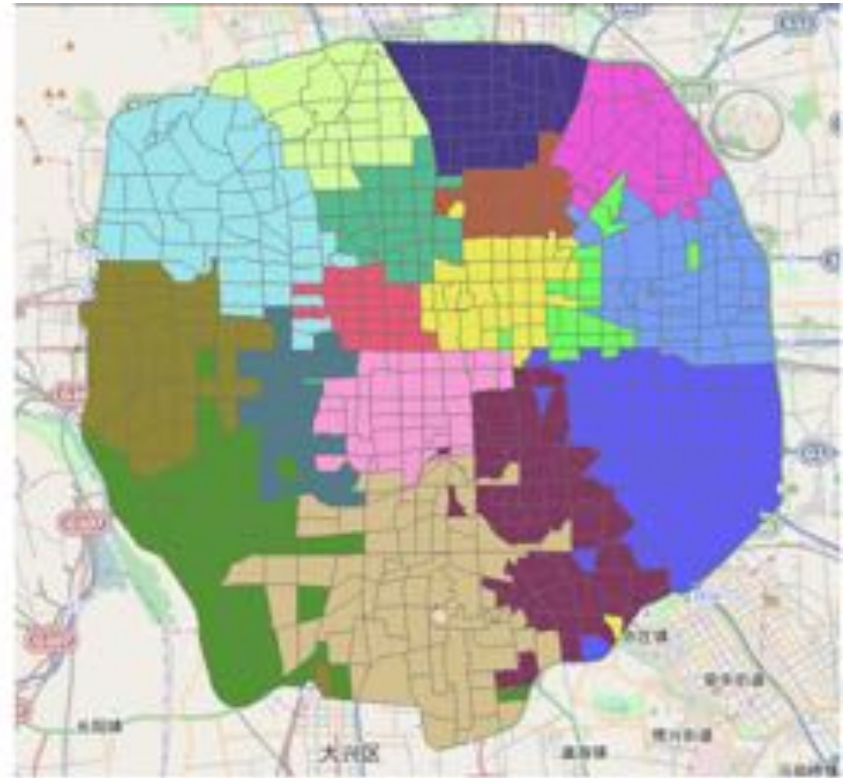
Beijing show case



Spatial patterns

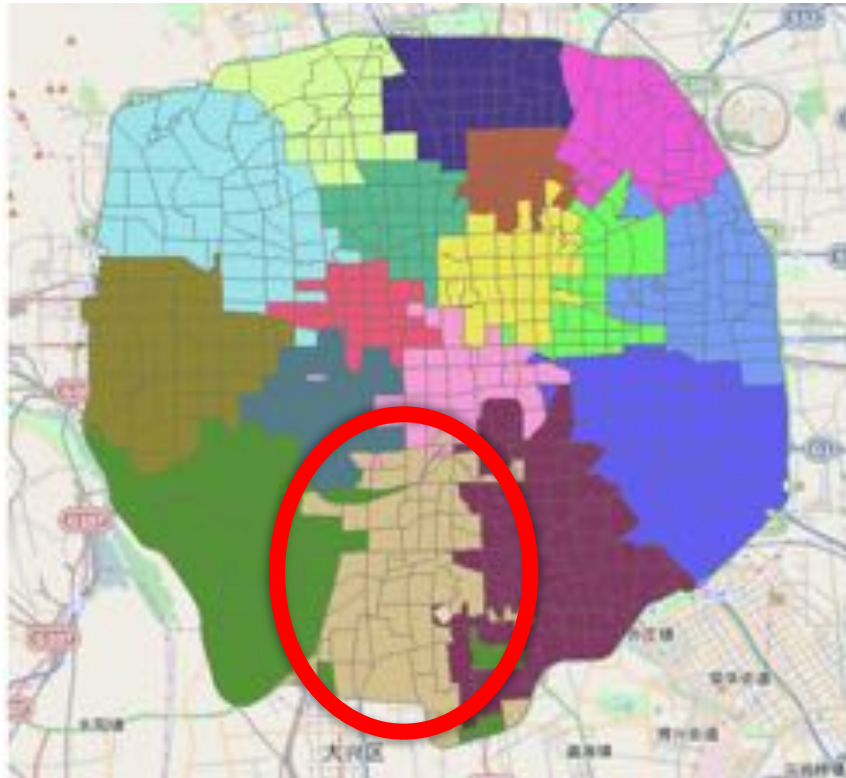


2008

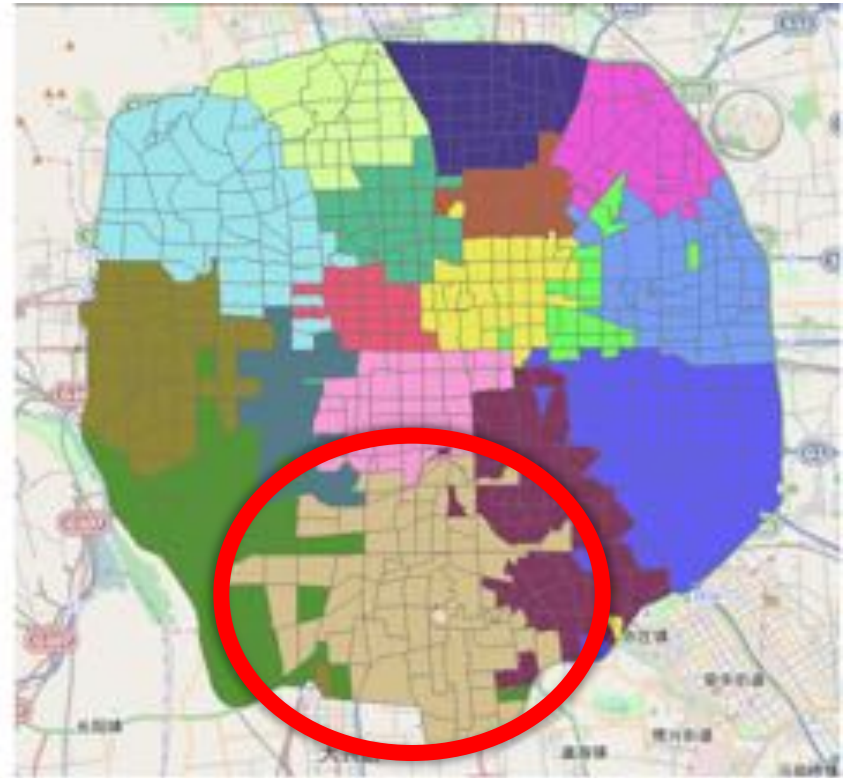


2012

Spatial patterns

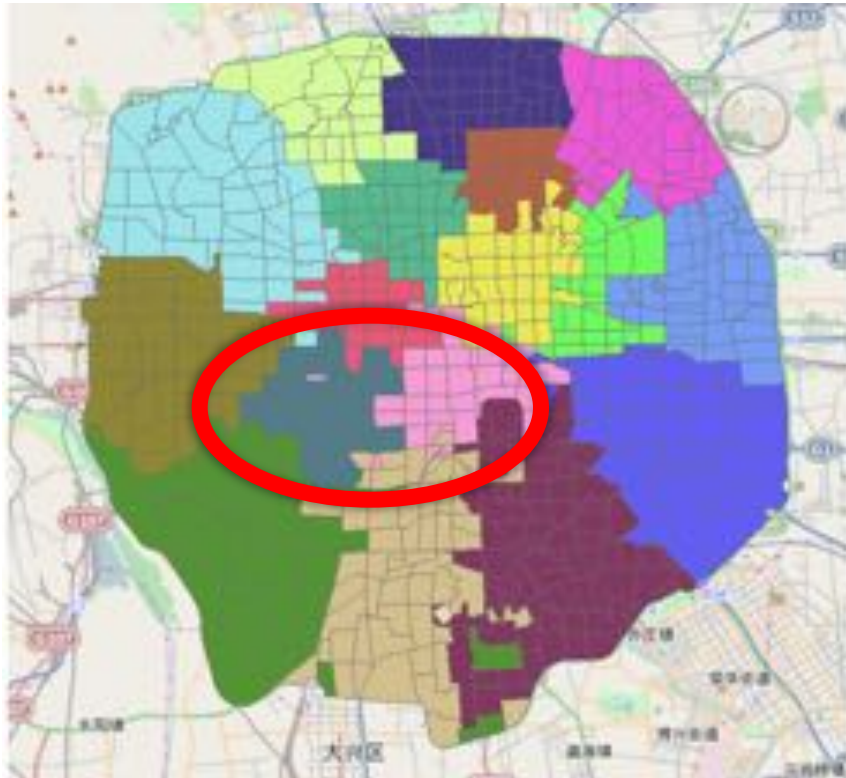


2008

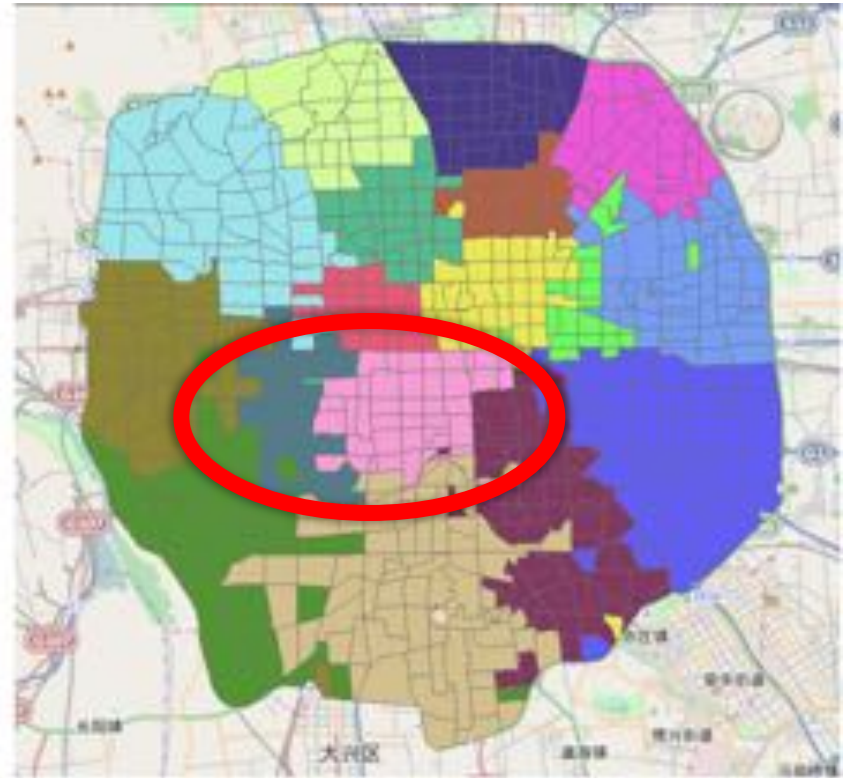


2012

Spatial patterns

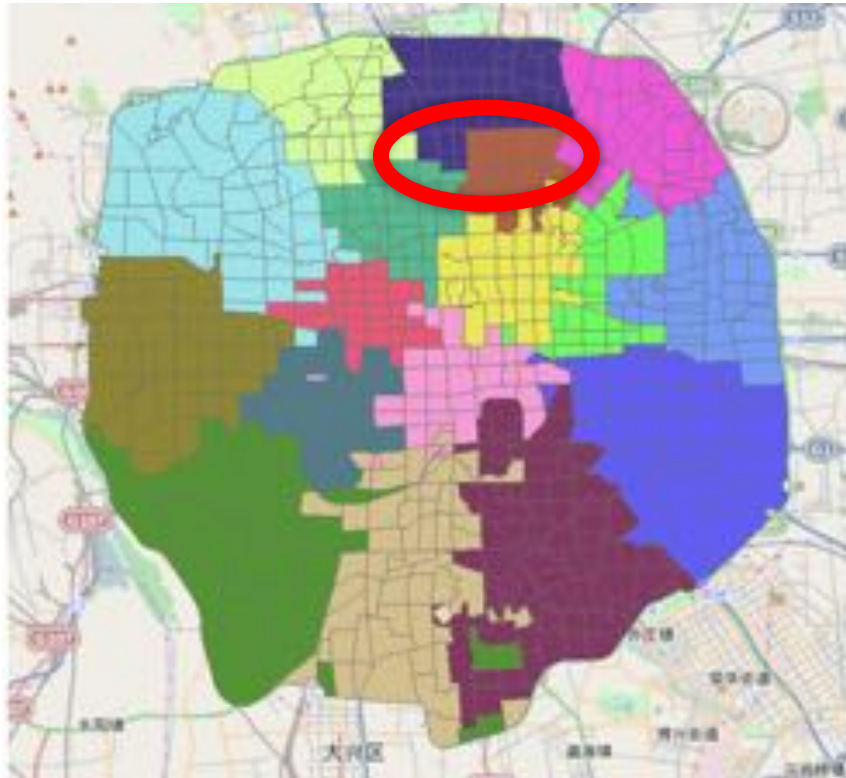


2008

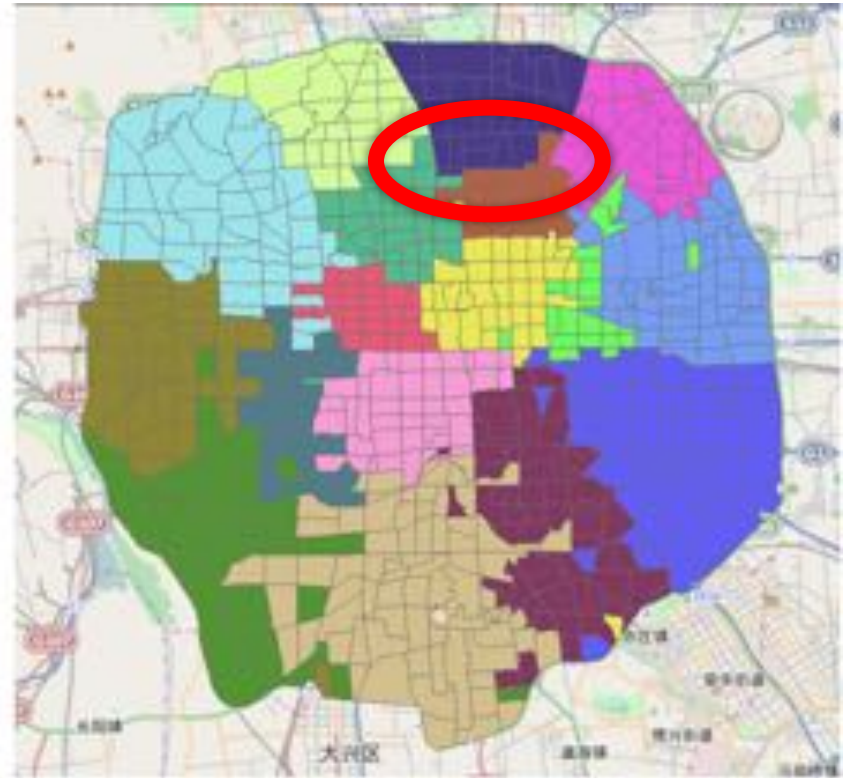


2012

Spatial patterns

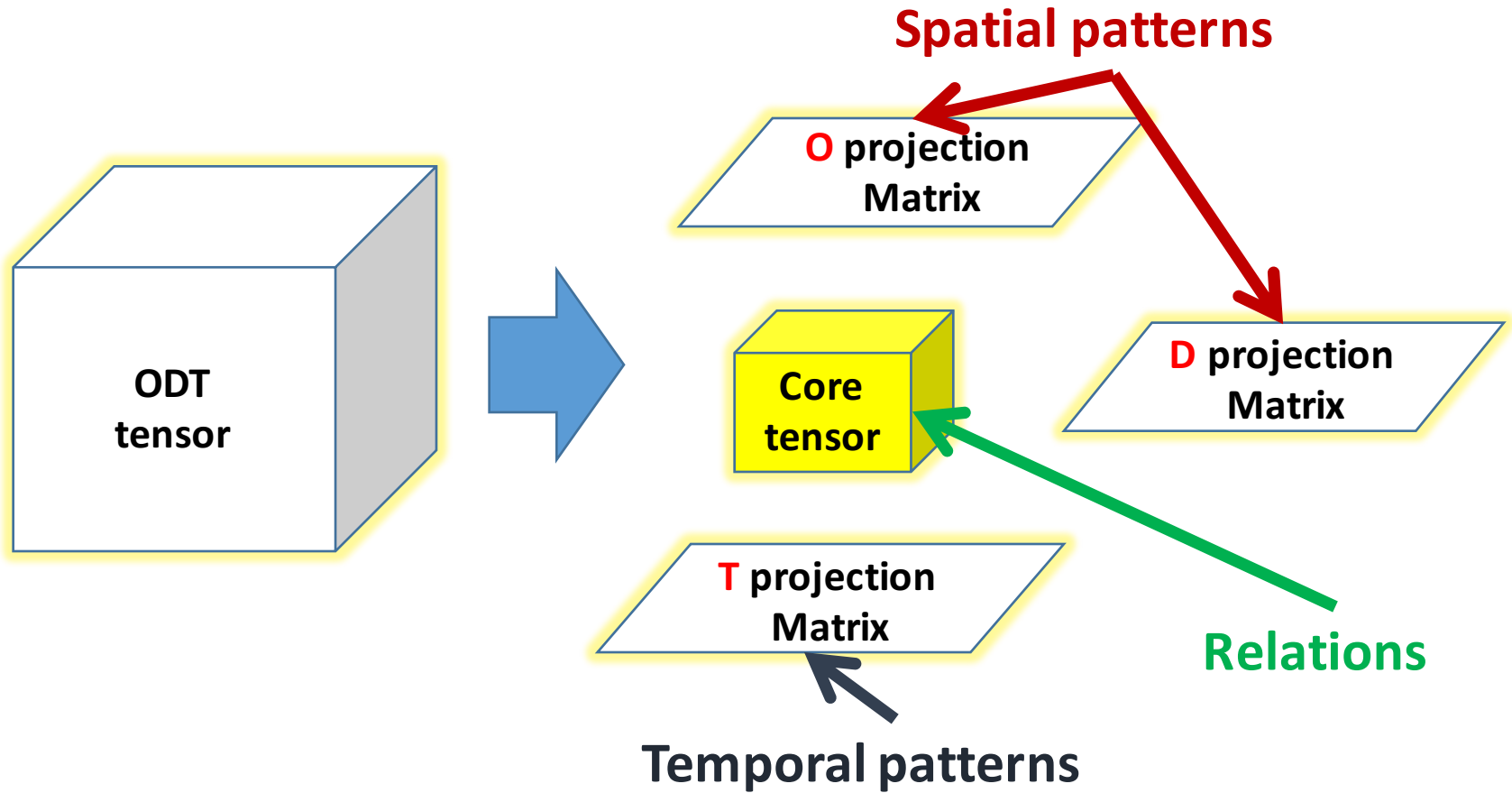


2008

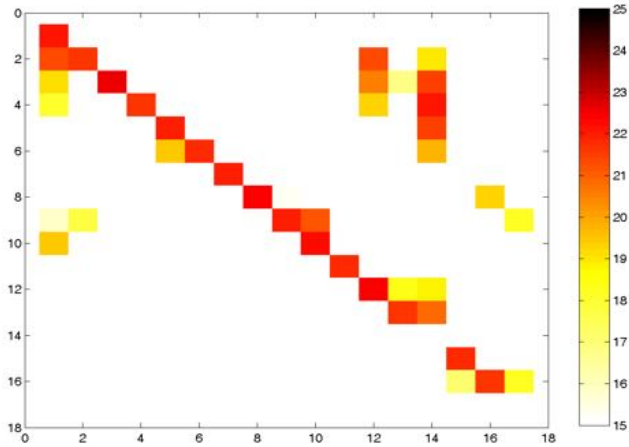


2012

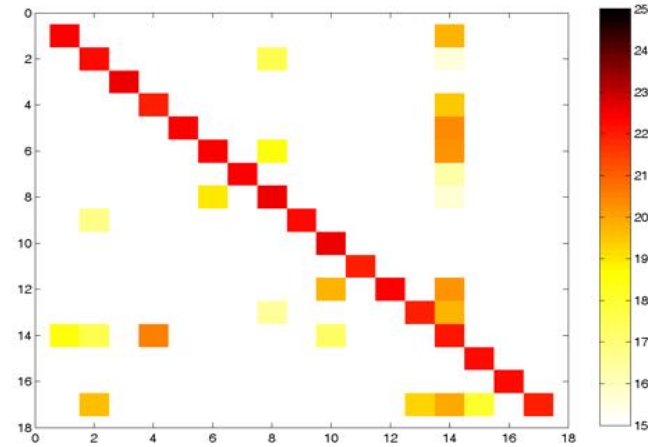
Beijing show case



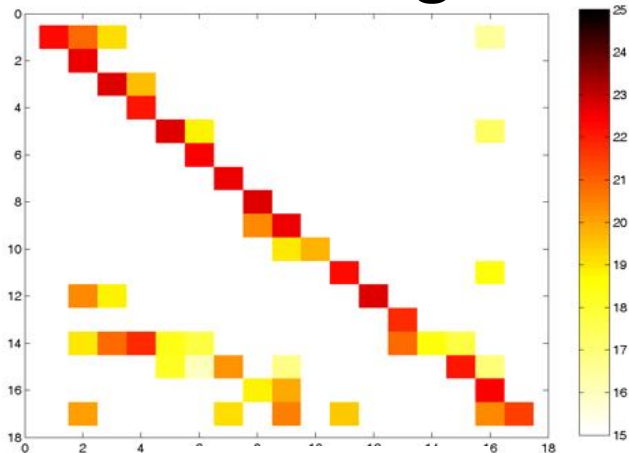
Core tensor



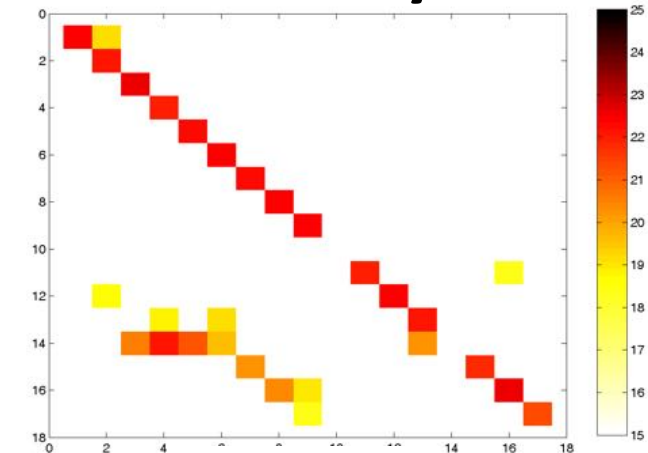
Morning



Midday

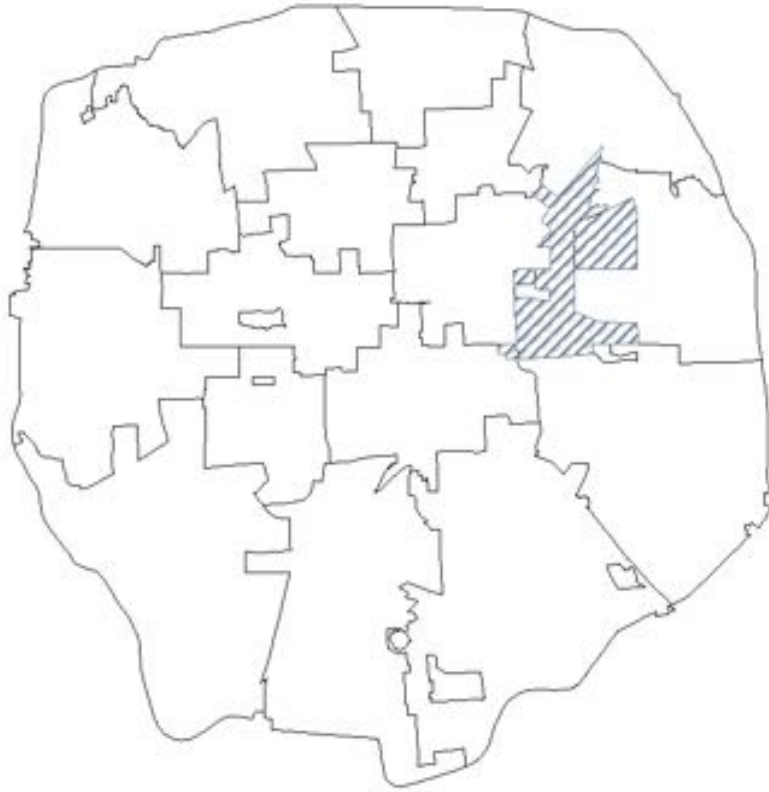


Evening

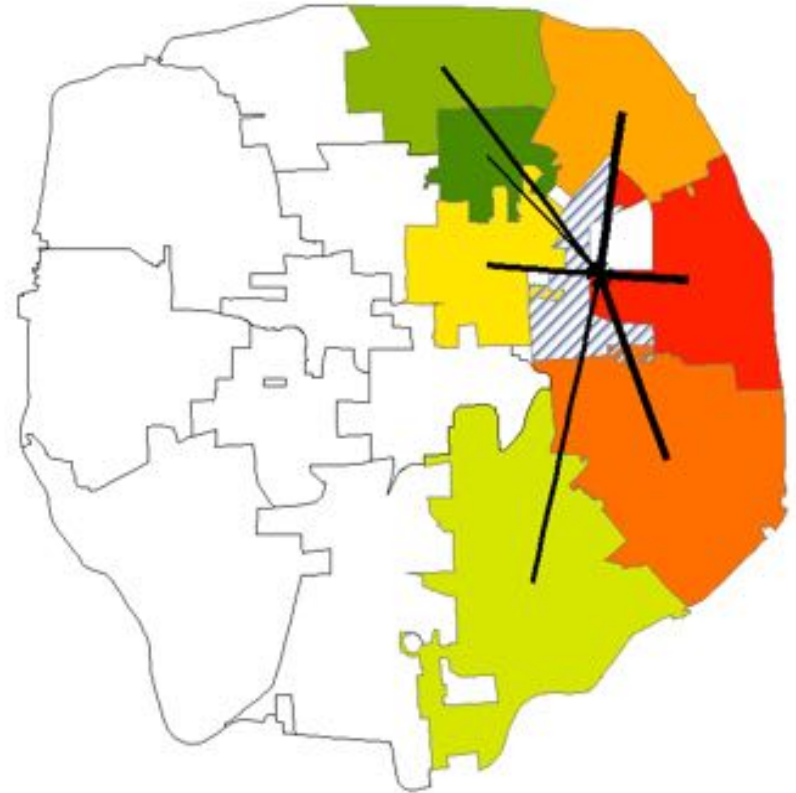


Night

Morning

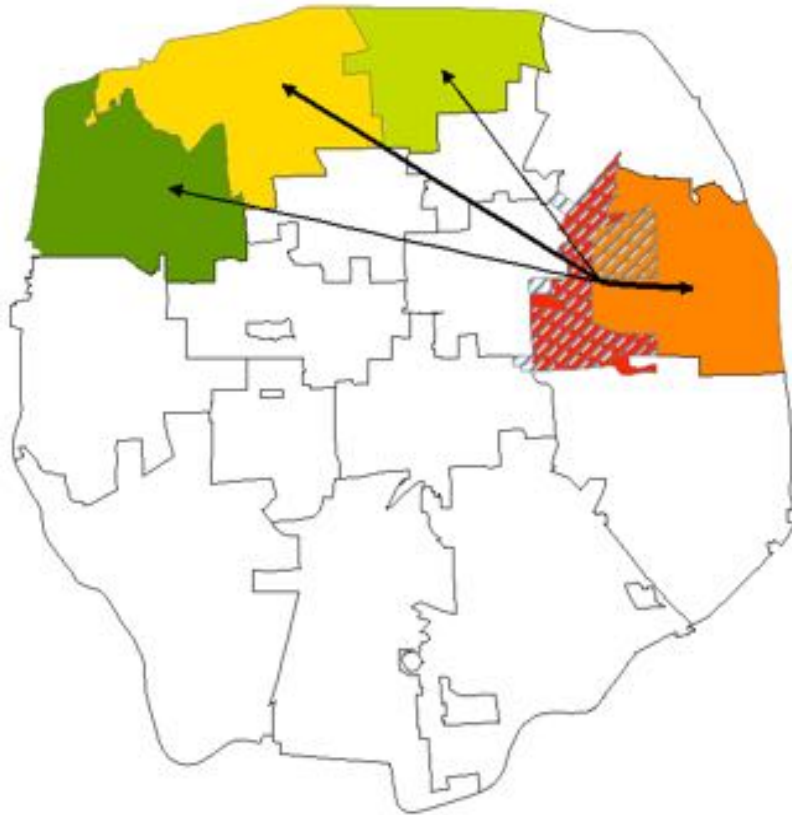


Origin

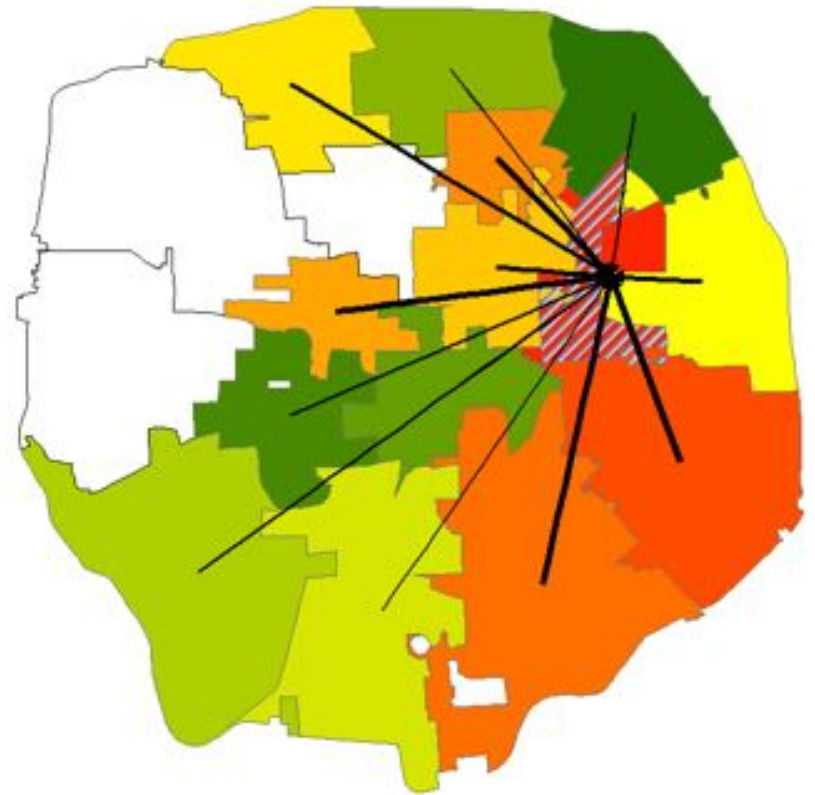


Destination

Midday

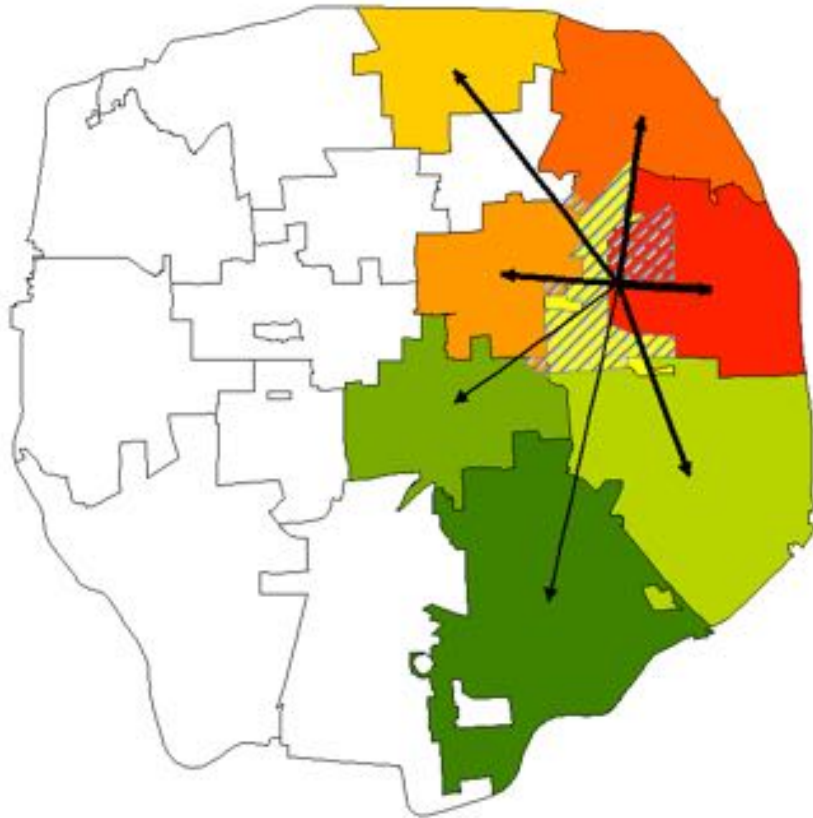


Origin

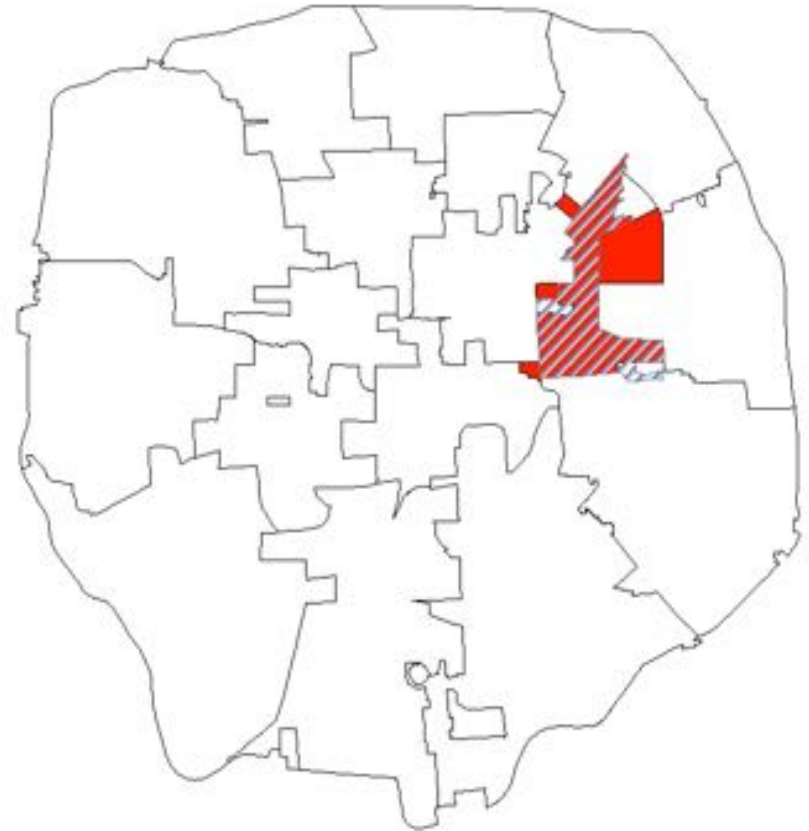


Destination

Evening

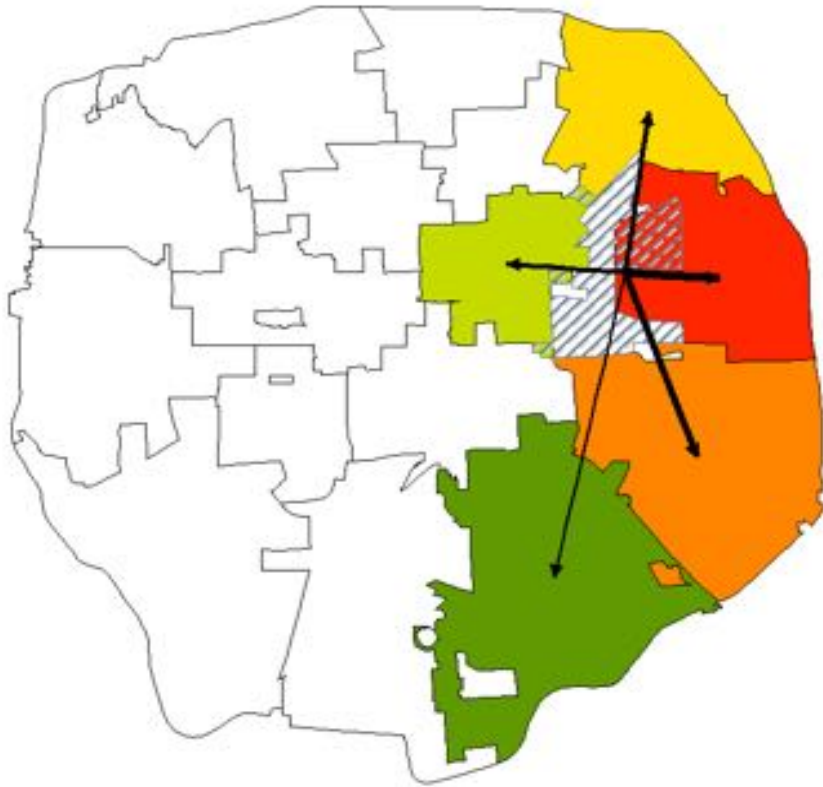


Origin

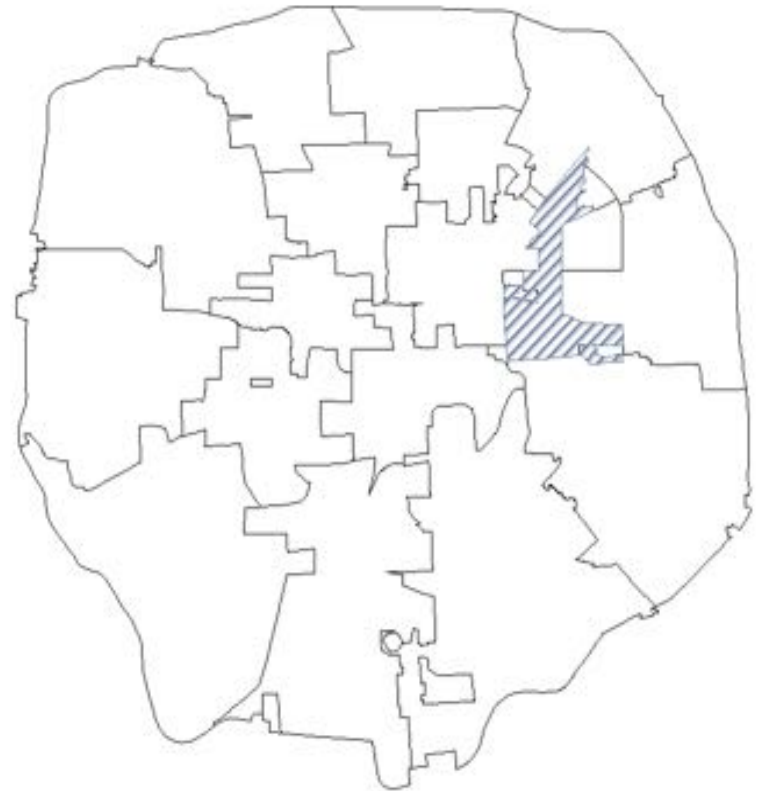


Destination

Night

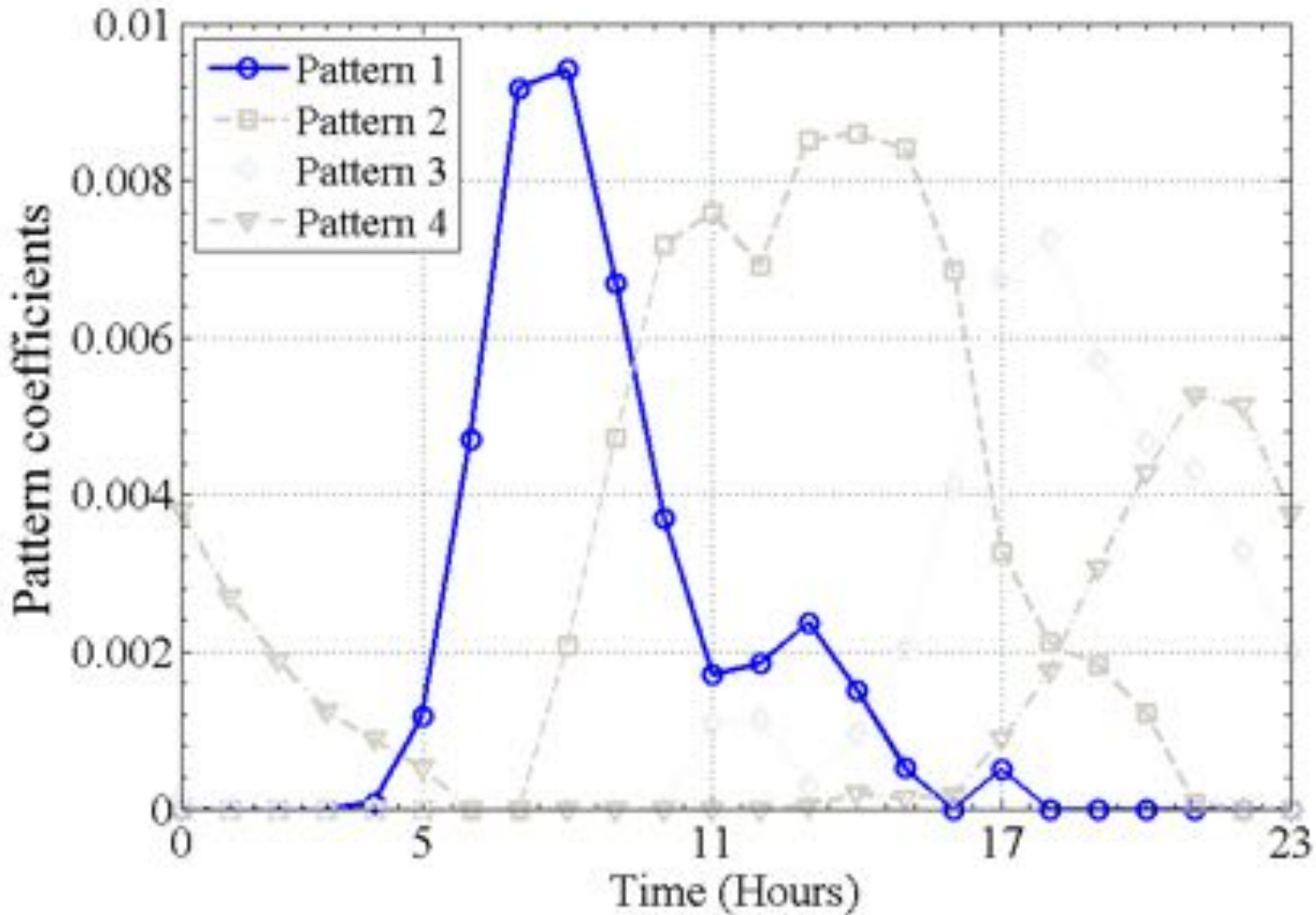


Origin

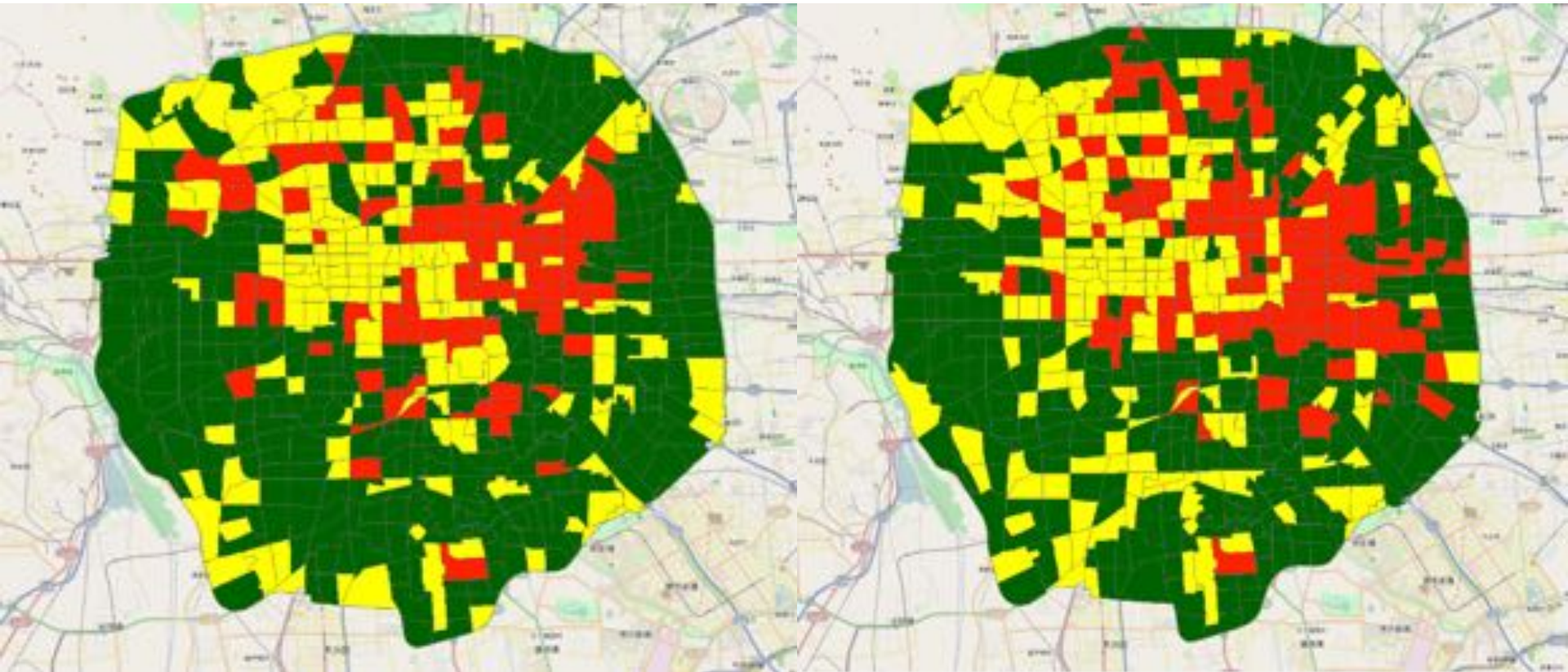


Destination

Function of urban zones



Function of urban zones



2008

2012

Where is the center of Beijing?



深度学习交通速度预测与拥堵分析

• 深度学习+智能交通= ?



深度学习：从大数据中获取智能

+ 计算机视觉=



+ 围棋 =



+ 智能交通 =

?

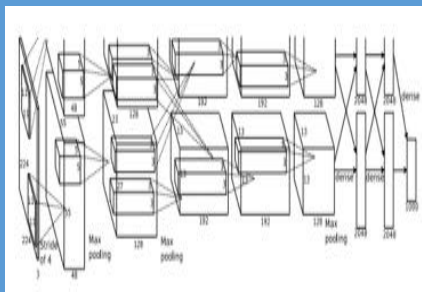
图像识别vs交通预测

图像识别问题



人脸照片

+



深度学习模型

输出结果

Obama

Hillary

Trump

Bush

.....

ImageNet
世界纪录

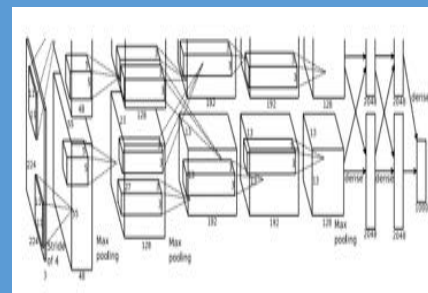
96.5%

交通预测问题



输入矩阵

+



深度学习模型

输出结果

高速



低速

深度学习交通速度预测与拥堵分析

研究任务：

- 实现精准的交通拥堵**预测**（服务驾驶）；**理解**交通拥堵的成因（服务规划）



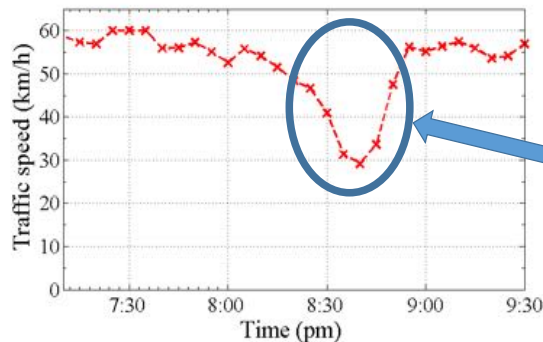
No-congestion



Congestion

主要挑战：

- 如何建模交通拥堵的**时空相关性**？
- 如何建模**突发事件**对于拥堵的影响？
- 如何寻找导致交通拥堵的**关键路段**？



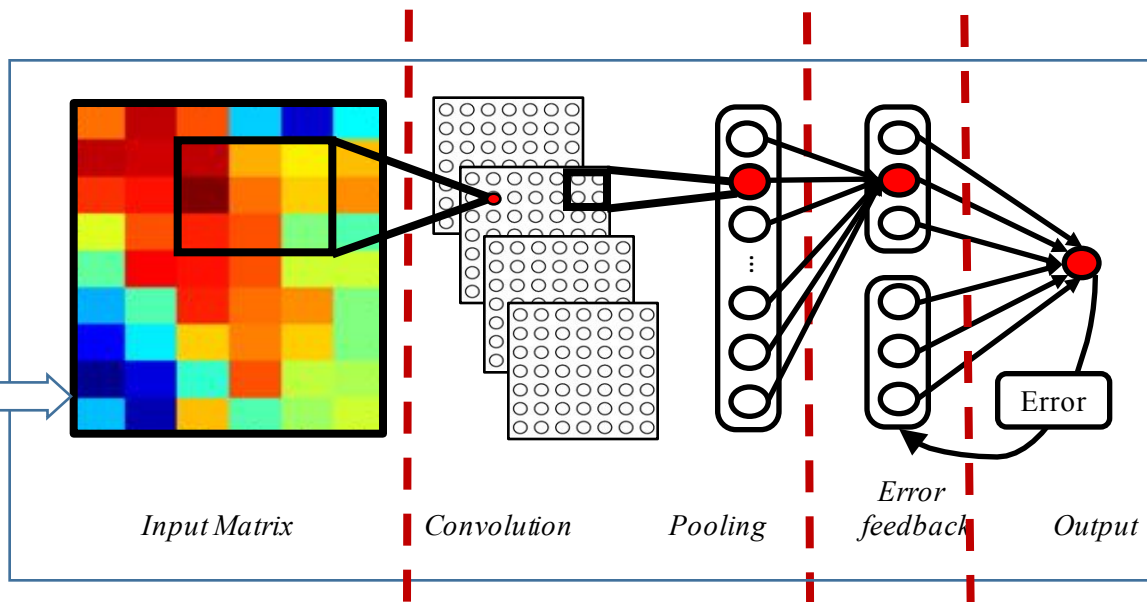
深度学习交通速度预测与拥堵分析

研究思路：

Field	Definition
ID	The unique ID of a taxi.
TIME	The sample time stamp of this record.
LON	The current longitude of the taxi.
LAT	The current latitude of the taxi.
DIR	The current driving direction of the taxi.
STATUS	Whether the taxi is carrying a passenger.

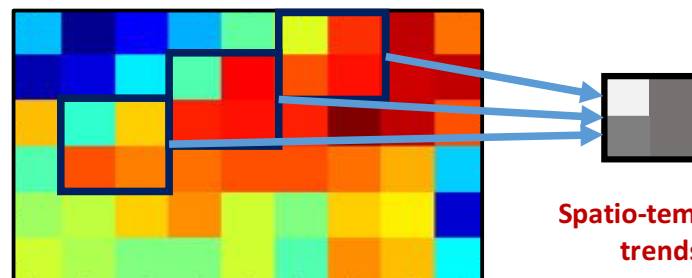
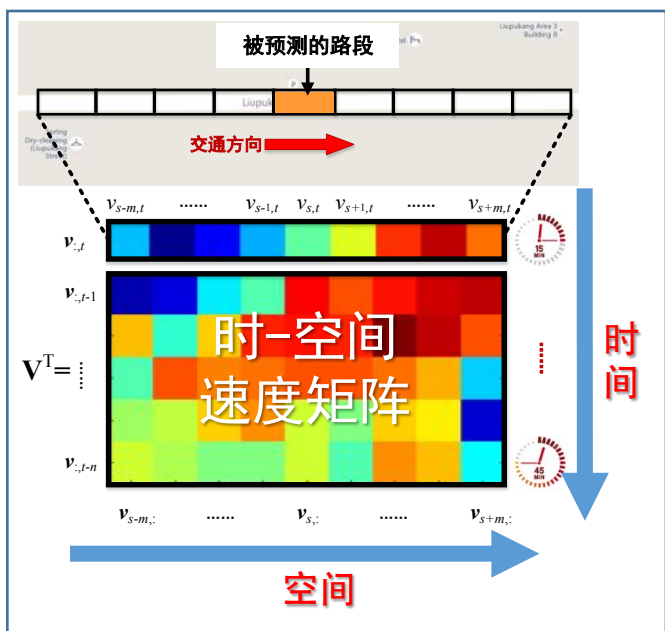


路况数据



深度学习交通速度预测与拥堵分析

核心贡献1：构建时空速度矩阵



there exist a number of local patterns (spatio-temporal trends) in the spatio-temporal input matrix

Convolution layer:

$$c_k^{p,q} = \sigma \left(b_k + \sum_{x=0}^i \sum_{y=0}^i w_k^{x,y} m^{p+x,q+y} \right)$$

using several filters to convolute the input matrix

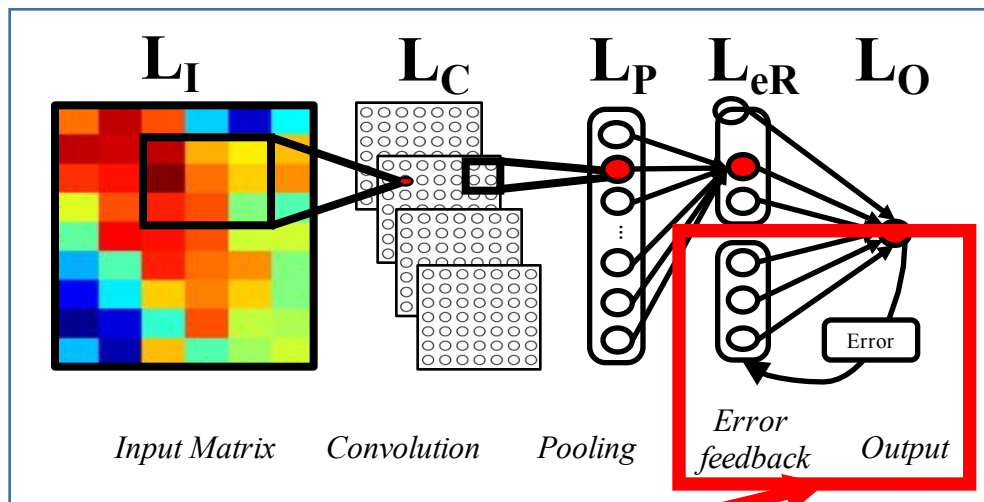
Pooling layer:

$$p_c = \frac{1}{N} \sum_p \sum_q c_k^{p,q}$$

using average pooling to down sampling the convolution neuro matrix

深度学习交通速度预测与拥堵分析

核心贡献2：采用带**误差反馈**的递归神经网络，应对交通突发事件



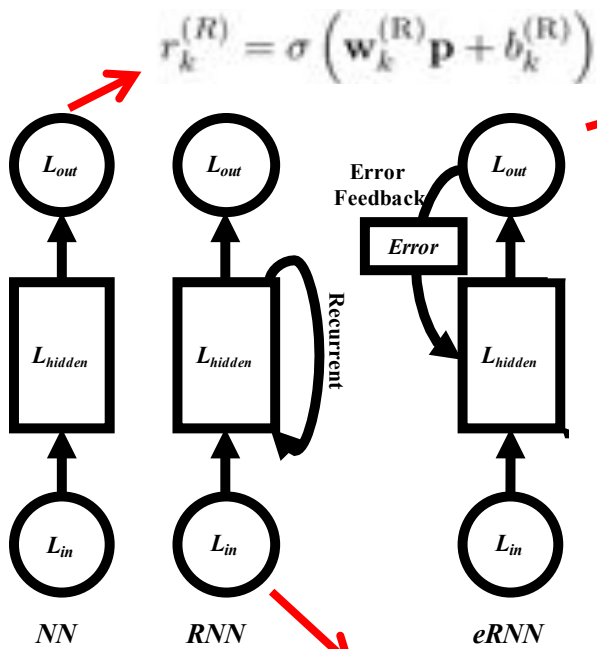
预测
误差



Error feedback recurrent neural network
eRNN带误差反馈的递归神经网络

深度学习交通速度预测与拥堵分析

- Error feedback recurrent neural network
带**误差反馈**的递归神经网络 (eRNN)



$$r_k^{(R)} = \sigma \left(\mathbf{w}_k^{(R)} \mathbf{p} + b_k^{(R)} \right)$$

eRNN隐藏层公式

$$r_k^{(E)}(t) = \text{sigmoid} \left(\mathbf{w}_k^{(E)} \mathbf{e}(t-1) + b_k^{(E)} \right)$$

$$\mathbf{e}(t) = [y(t-1) - o(t-1), \dots, y(t-l) - o(t-l)]$$

$$\mathbf{r} = [\mathbf{r}^{(R)}; \mathbf{r}^{(E)}]$$

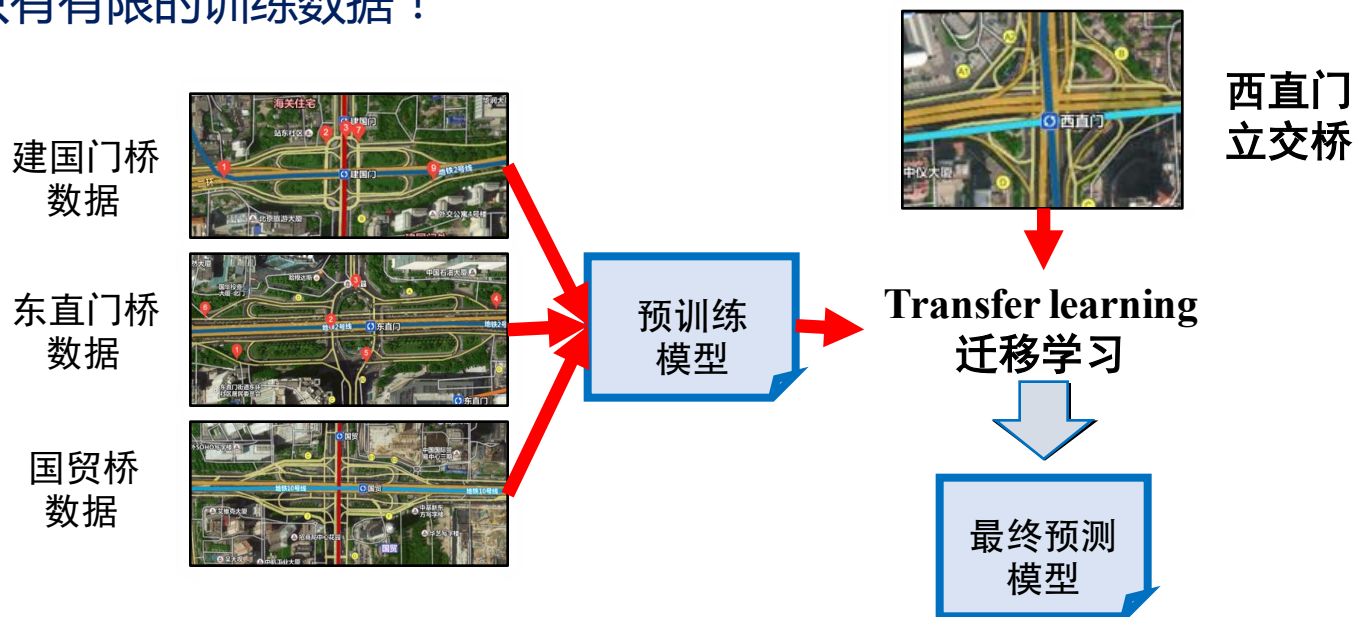
卷积信息子层+误差反馈子层

$$r_k^{(R)}(t) = \text{sigmoid} \left(\mathbf{w}_k^{(R)} \mathbf{p} + \tilde{\mathbf{w}}_k \mathbf{r}(t-1) + b_k^{(R)} \right)$$

深度学习交通速度预测与拥堵分析

核心贡献3：在模型训练中，使用**迁移学习**引入类似路段的数据知识，避免过度拟合

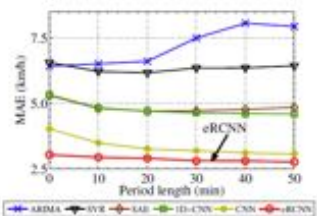
起因：每个road segment需要一个预测模型！甚至还要细分时段！
但只有有限的训练数据！



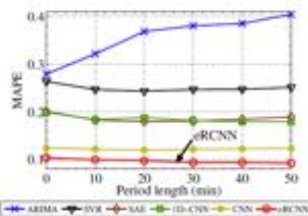
深度学习交通速度预测与拥堵分析

应用：北京二、三环车速预测（基于2013年11月工作日6万辆出租车数据；400米设为一个路段；每个路段的平均速度每隔5分钟更新一次）

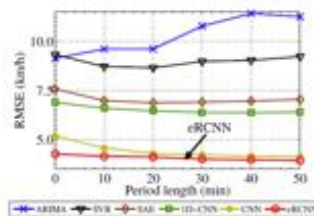
平均速度预测：二环



(a) MAE Comparison



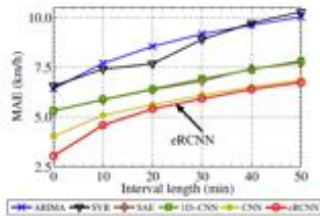
(b) MAPE Comparison



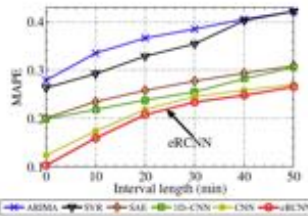
(c) RMSE Comparison

在不同场景和指标下，预测误差均小于现有算法。

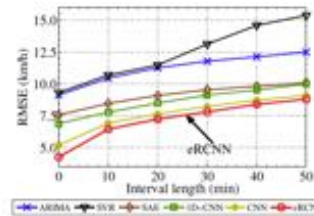
瞬时速度预测：二环



(a) MAE Comparison



(b) MAPE Comparison



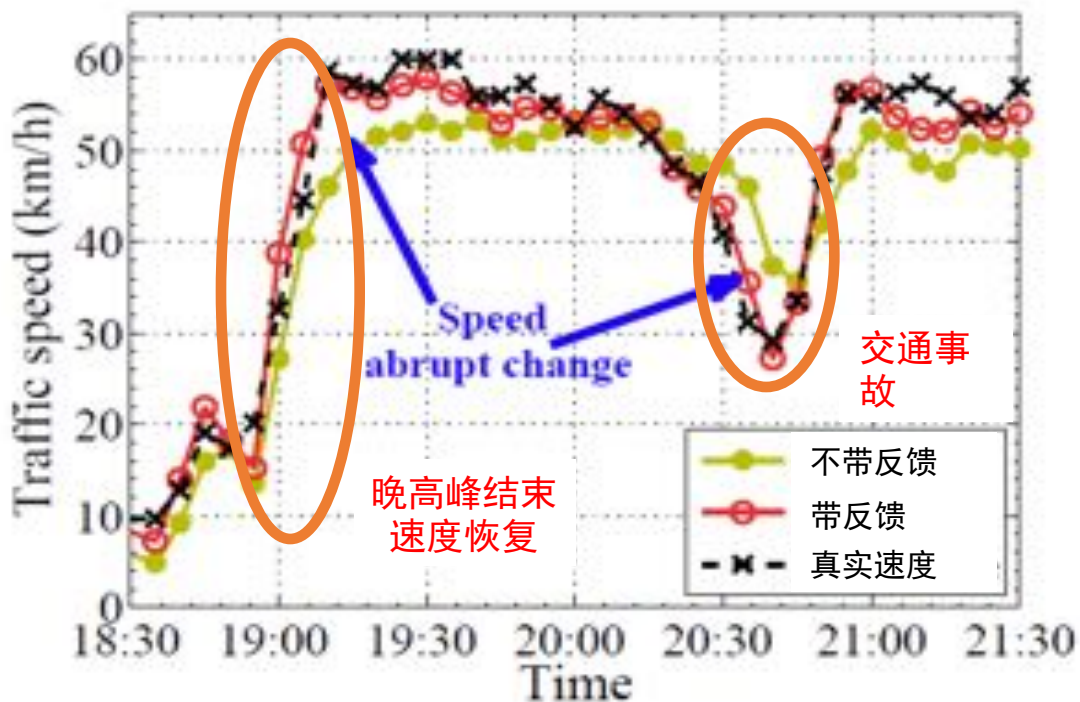
(c) RMSE Comparison

实用场景下预测误差 < 4km/h 接近GPS终端的采样误差。

ARIMA: Auto Regression Integrated Moving Average; SVR: Support Vector Regression; SAE: Stacked Auto Encoders; 1D-CNN: 1D Convolutional Neural Network; CNN: Convolutional Neural Network

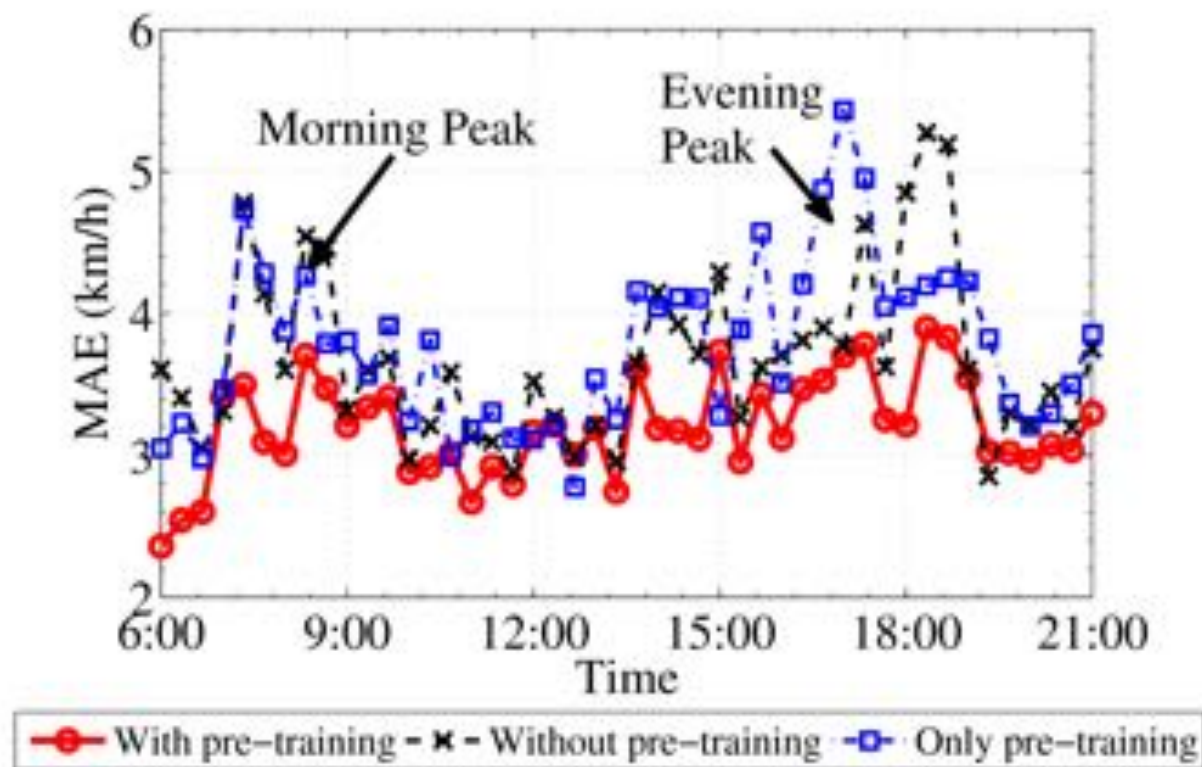
深度学习交通速度预测与拥堵分析

典型应用：准确预测突发事件造成的交通拥堵



深度学习交通速度预测与拥堵分析

典型应用：迁移学习避免过度拟合，提高预测精度



深度学习交通速度预测与拥堵分析

典型应用：评估道路影响力

$$Y_j = f \begin{pmatrix} x_1 \\ \text{道路1的速度} \\ x_2 \\ \text{道路2的速度} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ x_i \\ \text{道路 } i \text{ 的速度} \end{pmatrix}$$

被预测的
道路 j 的速度



$$\frac{\partial Y_j}{\partial x_i}$$

路段 i 对被预测
路段 j 的影响力。

$$\frac{1}{N} \sum_n \sum_j \frac{\partial Y_j}{\partial x_i}$$

路段 i 的重要
性：
对所有路段的
总体影响力。

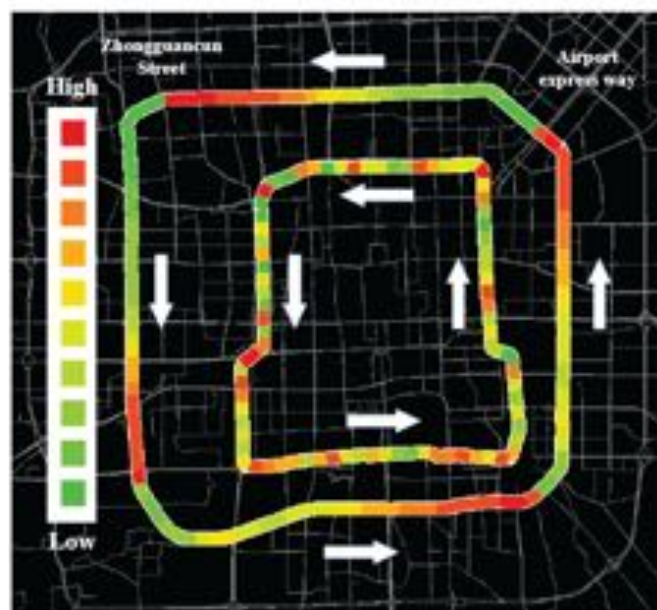
深度学习交通速度预测与拥堵分析

典型应用：评估道路影响力

对环路交通拥堵影响最大路段：**环路的拐弯与主要高速衔接处**



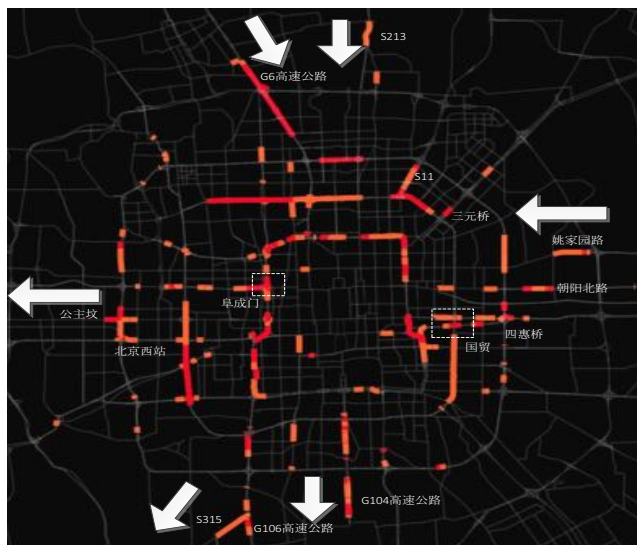
(a) The Inner Loop (clockwise)



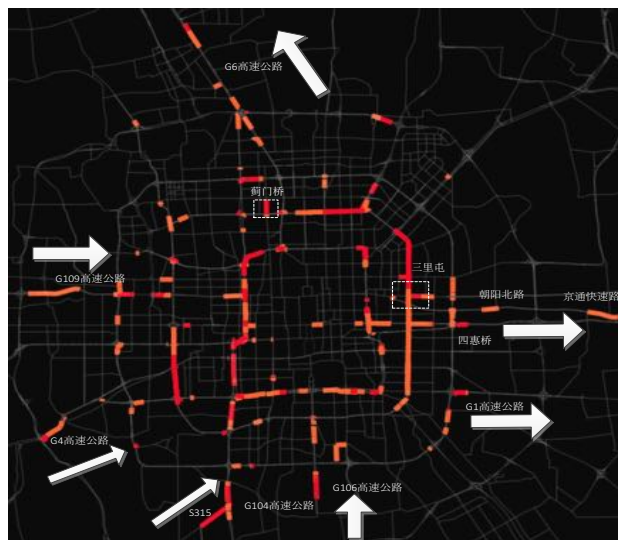
(b) The Outer Loop (anti-clockwise)

深度学习交通速度预测与拥堵分析

• 典型应用：评估道路影响力



第一类路段



第二类路段



全路网路段重要性分析,重要性高的路段主要分布在环路、立交桥及几个主要进出城道路。

谢 谢

E-mail: jywang@buaa.edu.cn

Weibo: @王静远BUAA