

实验室交流周报告

——第七周

报告人：汪继龙

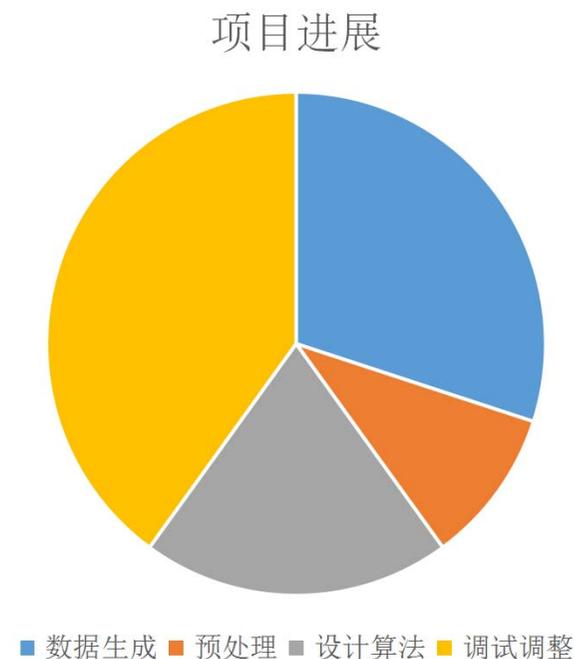
时 间：2018.04.21

本周工作

- 项目进展
- 《Deep Learning for IoT Big Data and Streaming Analytics: A Survey》
——Western Michigan University
- 流数据

一、项目进展

- 基于仿真模型，生成了五类目标的数据：
 - 歼十战斗机
 - JAS 39鹰狮战斗机
 - UH-1D直升机
 - 波音747-100 客机
 - 东风31型导弹
- 训练数据约 34000 条；
- 测试数据约 10900 条；
- 目前工作：预处理数据，根据数据特点设计不同的深度学习网络模型进行实验。



二、深度学习方法在IoT大数据和流数据中的应用

- 分析了 IoT 数据的关键特征和主要问题。
- 对前沿的深度学习方法及其在物联网领域对于大数据和流数据的应用进行了详细的总结。
- 作者对于目前应用了深度学习方法的少量 IoT 应用进行了介绍，并且对不同类型的深度神经网络在各种 IoT 领域的应用进行了概括和对比。
- 回顾了从资源受限设备到雾和云端的物联网层级上部署DL的最新方法和技术。
- 强调了深度学习与物联网应用成功结合所面临的挑战和未来的研究方向。

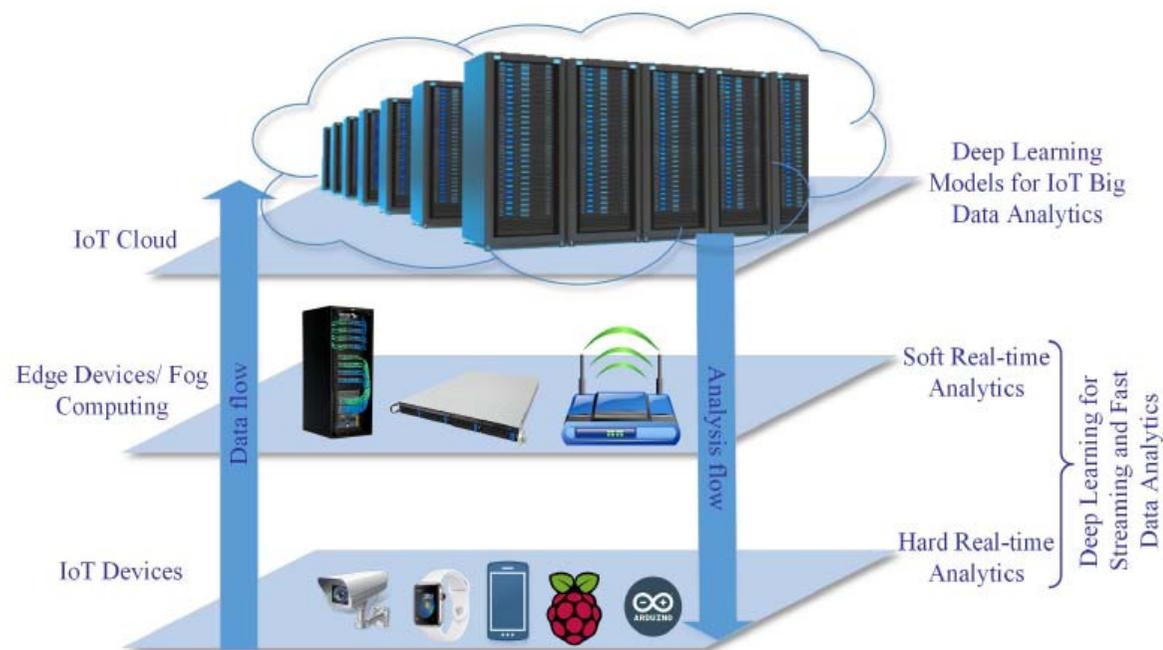


Fig. 1. IoT data generation at different levels and deep learning models to address their knowledge abstraction.

二、IoT数据特征

	特点
大数据	4V: Volume (数据体量大)、Variety (数据类型繁多)、Velocity (处理速度快)、Value (价值密度低)
IoT大数据	6V: Volume (数据体量大)、Variety (数据类型繁多)、Velocity (处理速度快)、Value (价值密度低)、Veracity (真实性): 真实性是指质量, 一致性, 和数据的可信性, 有真实性的数据才能进行准确的分析 Variability (易变性): 这个属性是指数据流的速率不同
IoT快速流数据	<u>顺序</u> 、 <u>大量</u> 、 <u>快速</u> 、 <u>实时</u> 、 <u>多源</u>

二、深度模型总结

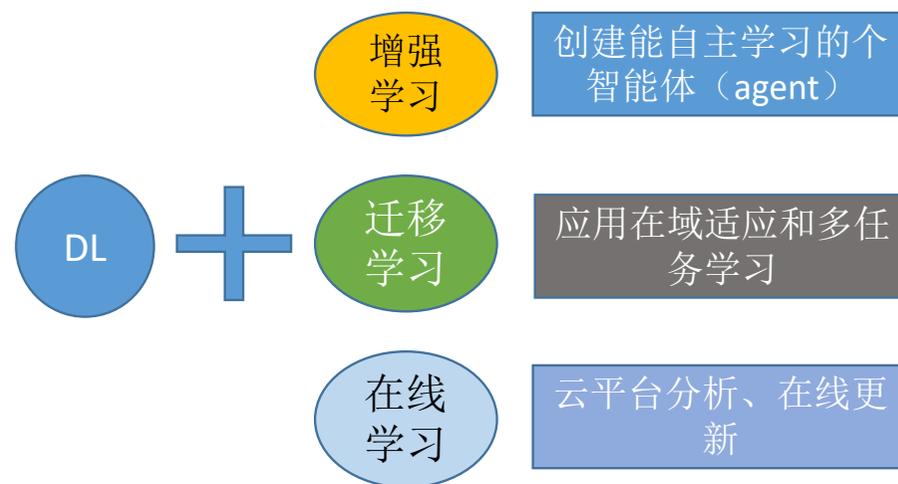
Model	Category	Learning model	Typical input data	Characteristics
AE	Generative	Unsupervised	Various	<ul style="list-style-type: none"> • Suitable for feature extraction, dimensionality reduction • Same number of input and output units • The output reconstructs input data • Works with unlabeled data
RNN	Discriminative	Supervised	Serial, time-series	<ul style="list-style-type: none"> • Processes sequences of data through internal memory • Useful in IoT applications with time-dependent data
RBM	Generative	Unsupervised, Supervised	Various	<ul style="list-style-type: none"> • Suitable for feature extraction, dimensionality reduction, and classification • Expensive training procedure
DBN	Generative	Unsupervised, Supervised	Various	<ul style="list-style-type: none"> • Suitable for hierarchical features discovery • Greedy training of the network layer by layer
LSTM	Discriminative	Supervised	Serial, time-series, long time dependent data	<ul style="list-style-type: none"> • Good performance with data of long time lag • Access to memory cell is protected by gates
CNN	Discriminative	Supervised	2-D (image, sound, etc.)	<ul style="list-style-type: none"> • Convolution layers take biggest part of computations • Less connection compared to DNNs. • Needs a large training dataset for visual tasks.
VAE	Generative	Semi-supervised	Various	<ul style="list-style-type: none"> • A class of Auto-encoders • Suitable for scarcity of labeled data
GAN	Hybrid	Semi-supervised	Various	<ul style="list-style-type: none"> • Suitable for noisy data • Composed of two networks: one generator and one discriminator
Ladder Net	Hybrid	Semi-supervised	Various	<ul style="list-style-type: none"> • Suitable for noisy data • Composed of three networks: two encoders and one decoder

快速实时深度学习结构

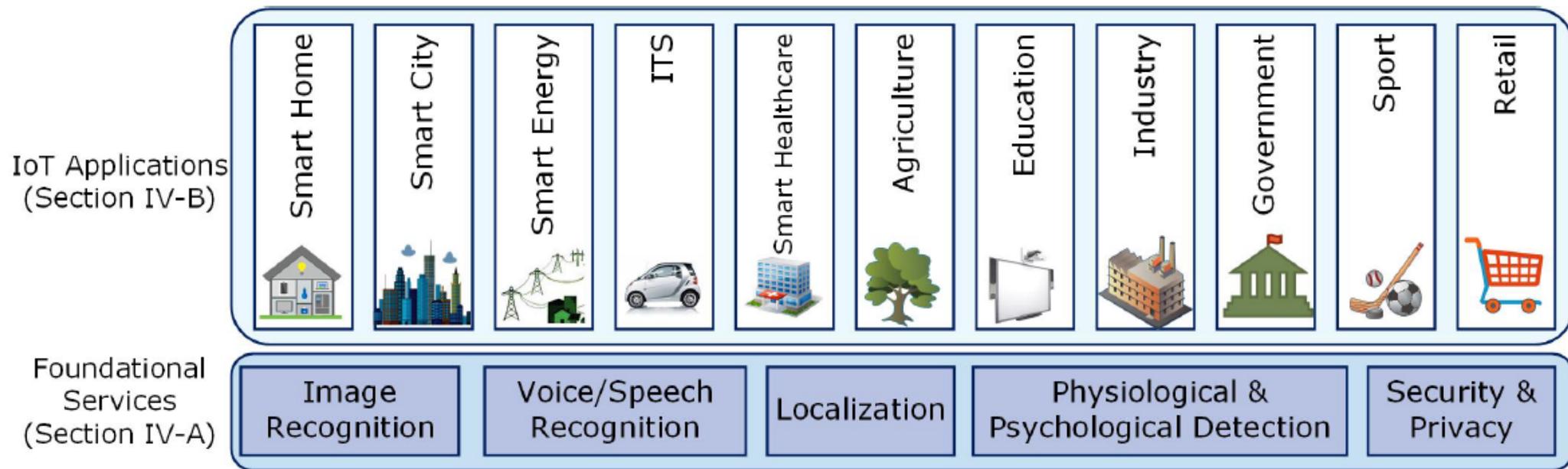
使用深度学习模型对数据流进行快速实时的处理仍在起步阶段。早期工作：

- OS-ELM, 将一个实时序列学习算法应用到单隐层的前馈神经网络。Ren 等人
- Faster-RCNN 在图片中的目标检测中达到了接近实时的速度。他们的目标检测框架的运行时间为 5-17fps。
- YOLO将目标检测的速度提高到 45fps, 以及更小版本的 YOLO, 速度更是达到了 155fps, 已经适用于智能相机。

深度学习与其他方法结合



二、深度学习在IoT领域中应用



二、IoT设备上的深度学习

- 将深度学习模型和方法部署在资源受限的设备上成了 IoT 领域的一个重要研究方向。
- 方法和技术：
 - 网络压缩：在资源受限设备上应用深度神经网络的方法之一是网络压缩，将密集的网络转化为一个稀疏的网络；
 - 近似计算：近似计算是实现在 IoT 设备上部署机器学习工具的另一种方法；
 - 加速器：设计特定的硬件和电路来优化 IoT 设备中深度学习模型的能量效率和内存占用是另一个活跃的研究方向；
 - 微处理器：开发具有强深度学习能力的小尺寸处理器也是研究热点；

二、深度学习带来的IoT挑战

(1) **缺少大型 IoT 数据集**: 缺乏可用的实际 IoT 应用大数据集将深度学习模型引入 IoT 的一个主要障碍, 因为深度学习需要更多的数据来实现更高的精度。

(2) **预处理**: 许多深度学习方法需要对数据进行预处理以产生更好的结果, 对于 IoT 应用, 预处理会更复杂, 因为系统处理的是来自不同数据源的数据, 可能有多种格式和分布, 而且还可能有数据丢失。

(3) **安全和隐私**: 深度学习训练模型也容易受到恶意攻击, 如虚假数据注入或对抗性样本输入, 其中 IoT 系统的许多功能或非功能性要求可能无法得到保证。

(4) **IoT 大数据” 6V “特性**: Volume (数据量) Variety (多样性) Velocity (速率) Veracity (可信度) Variability (可变性) Value (价值)

(5) **IoT 设备上的深度学习**: 在 IoT 设备上资源受限, 无法满足深度神经网络的需求。

(6) **深度学习自身局限**: 回归能力有待增强。

三、流数据

1. 百度百科：流数据（streaming data）是一组顺序、大量、快速、连续到达的数据序列,一般情况下,数据流可被视为一个随时间延续而无限增长的动态数据集合。应用于网络监控、传感器网络、航空航天、气象测控和金融服务等领域。
2. 维基百科：流数据是由不同数据源连续高速生成的数据。这些数据应该使用流处理技术进行递增式处理，而无需访问所有数据。另外，应该考虑到在数据中可能发生概念漂移，这意味着流的属性可能随时间而改变。
3. 流数据具有四个特点：
 - 1) 数据实时到达；
 - 2) 数据到达次序独立，不受应用系统所控制；
 - 3) 数据规模宏大且不能预知其最大值；
 - 4) 数据一经处理，除非特意保存，否则不能被再次取出处理，或者再次提取数据代价昂贵。

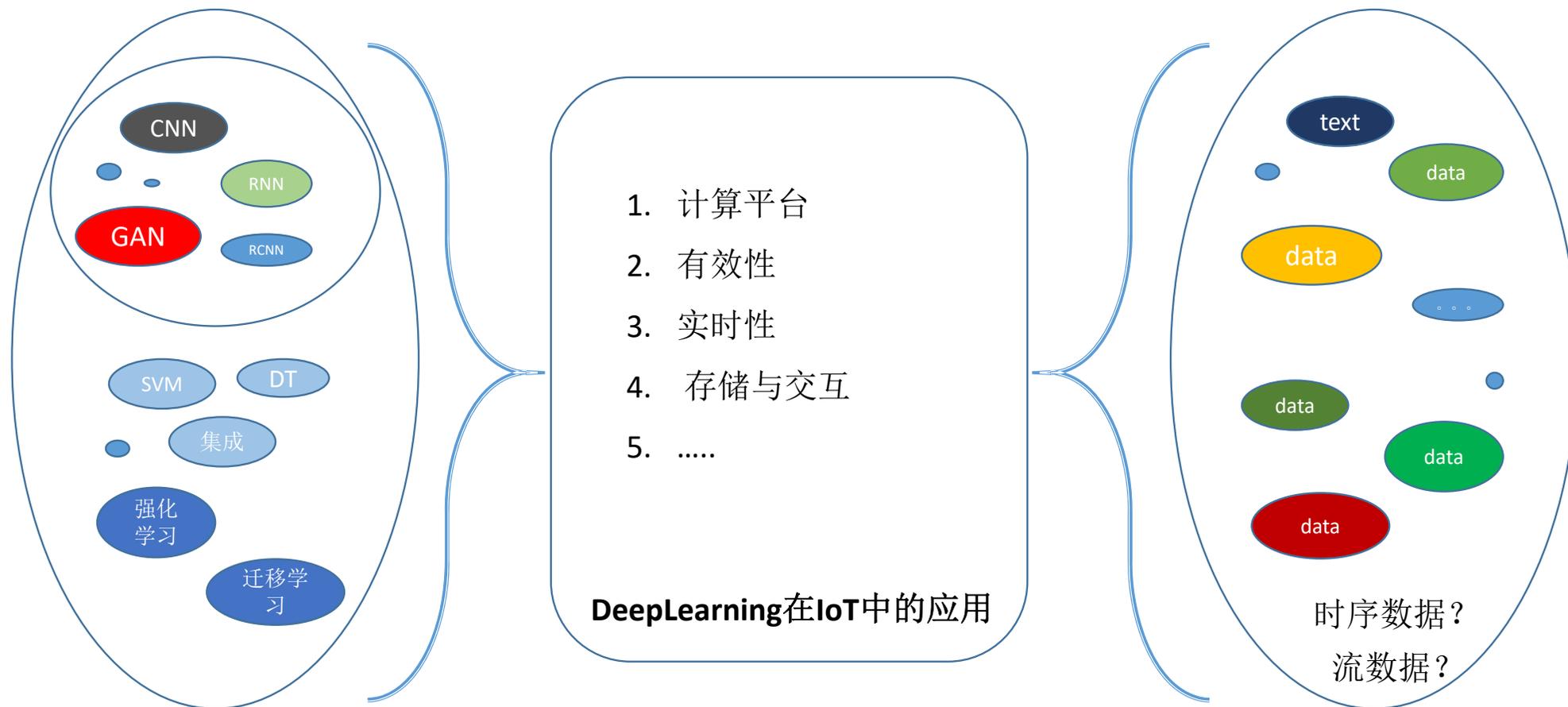
三、对比

流数据的特点在于：

- 数据一般具有时序性；
- 对实时性要求高；
- 数据的处理方式一般为并行处理和增量处理；
- 流式大数据处理框架：
 - Apache Storm
 - Apache Spark
 - Apache Samza

	特点
大数据	4V: Volume (数据体量大)、Variety (数据类型繁多)、Velocity (处理速度快)、Value (价值密度低)
IoT大数据	6V: Volume (数据体量大)、Variety (数据类型繁多)、Velocity (处理速度快)、Value (价值密度低)、Veracity (真实性)、Variability (易变性)
IoT快速流数据	顺序、大量、快速、实时、多源
时序数据 (经济计量学专用名词)	时间序列 (或称动态数列, time series) 是指将 <u>同一统计指标的数值按其发生的时间先后顺序排列而成的数列</u> 。时间序列分析的主要目的是根据已有的历史数据对未来进行 <u>预测</u> 。如股市、天气预测等。

研究方向



欢迎批评指正
谢谢大家！