

Big Data for Internet of Things: A Survey

Mouzhi Ge , Hind Bangui , Barbora Buhnova , Masaryk University

FGCS 2018

报告人：刘聪

目录

『CONTENT』

- ▷ 第一部分 『背景介绍』
- ▷ 第二部分 『物联网域』
- ▷ 第三部分 『大数据流程和生命周期』
- ▷ 第四部分 『物联网域与大数据』
- ▷ 第五部分 『总结』

第一部分



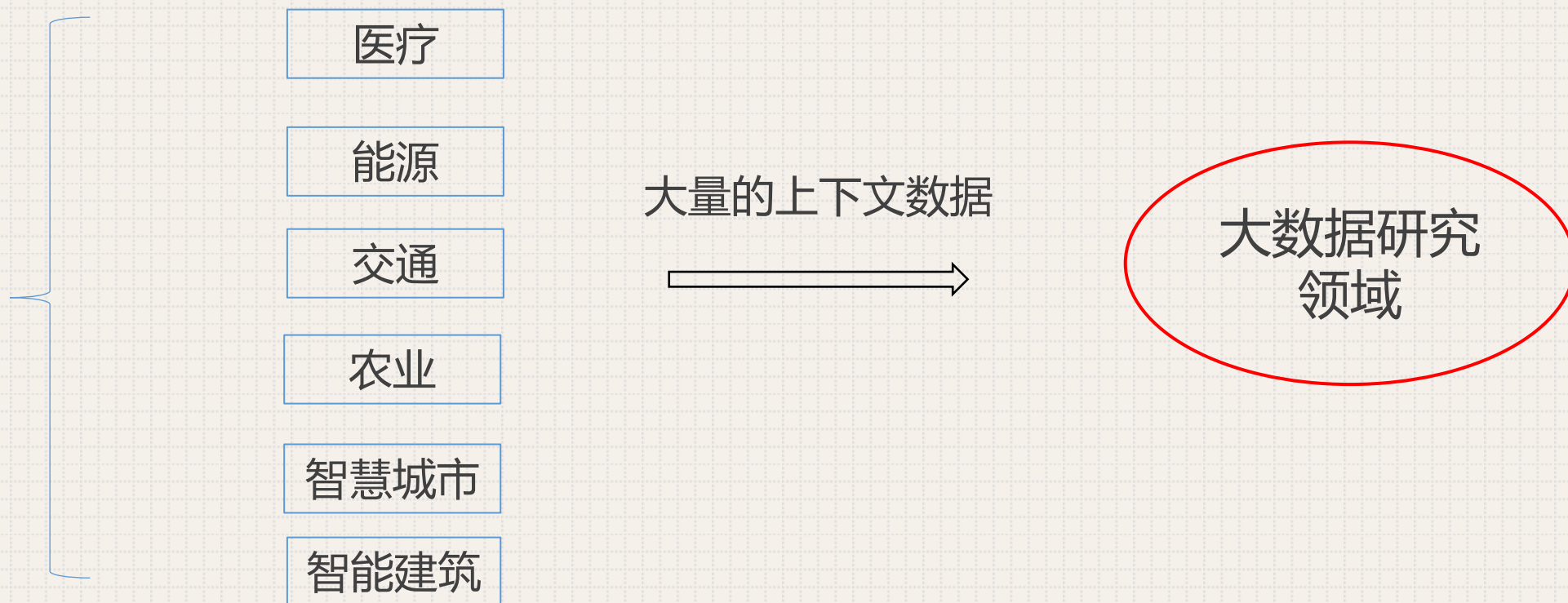
『背景介绍』



1

背景介绍

物联网（IoT）是当今时代最具有前景的技术之一，大量的物联网应用被开发出来并部署在不同的领域。



1 背景介绍

- 一方面，物联网在不同领域的应用产生的数据量 and 数据类型都在显著增加，
- 与此同时，大数据技术也在为物联网提供新的解决方案



1

背景介绍

现有的工作

- 把握了两个领域的研究趋势
- 总结了最新的大数据技术或物联网技术
- 很少有研究结合两个领域，这就缩小了不同应用场景中共享知识的空间。

本文的贡献

- 分析了物联网中最先进的大数据研究
- 对比并表明哪些大数据技术可以用于其他IoT领域
- 对每个物联网域中的大数据研究进行分析和解释
- 提出了一个概念框架来指导研究人员和从业人员在不同IoT领域中选择合适的大数据技术。

第二部分

『物联网域』



- 物联网域指的是在某些环境中应用的物联网技术
- 不同的物联网域也有一些共同特征。

医疗



收集和分析实时医疗信息
云平台存储数据流
根据患者的健康状况信息给出最佳康复计划，减少诊断错误。

能源



主要是智能电网物联网
使用大数据技术实现智能和高效电能管理质量
整合可再生资源、储能等

交通运输




物联网传感器的数据可用于指导路线规划、监视，应急管理，交通控制等。
最大限度地减少交通事故的风险。

智能建筑



物联网设备会收集及时的敏感信息，监控公民的日常活动，实现如门禁，数字视频，入侵检测，火灾探测和报警等功能



智慧城市

通过提供各个领域的智能应用来改善公民的生活方式。如停车场，城市清洁，废物管理等



农业

物联网设备的监控可以在保护农产品免受昆虫伤害，通过数据分析还可以提高农业生产力并节约能源



工业

通过集成物联网技术，利用共享数据来提供客户喜欢的合适商品或自身系统功能的改进



军事

物联网应用可以提供各种军事情报来支持军事物流，监视和军事机器人等。

第三部分

『大数据流程和生命周期』



3

大数据流程和生命周期

- 大数据技术包括多种方法和技术，每种的应用目的都略微不同。

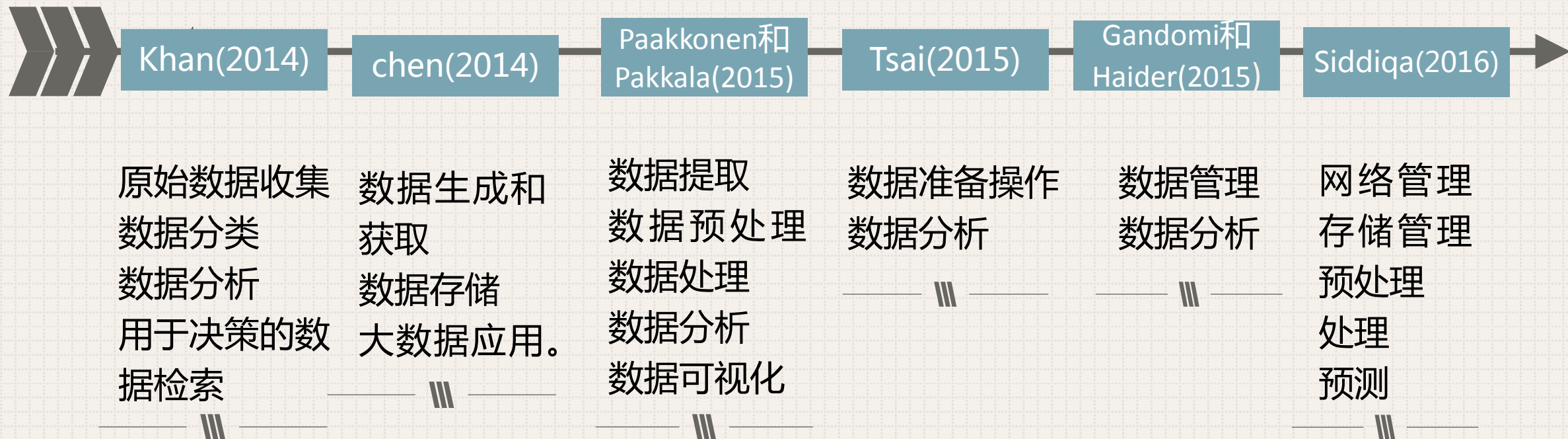
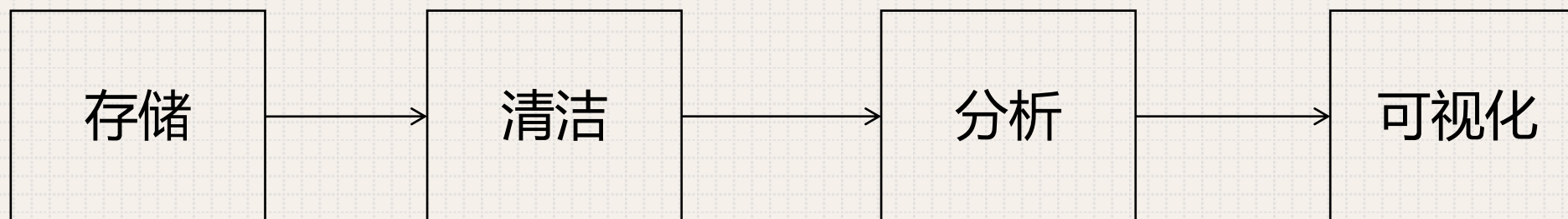


Table 1
Important Big Data terms from the selected literature.

	Gandomi and Haider 2015 [72]	Paakkonen and Pakkala 2015 [73]	Khan et al. 2014 [74]	Tsai et al. 2015 [13]	Chen et al. 2014 [11]	Siddiqua et al. 2016 [75]	Appearance frequency
Acquisition	X				X		2
Aggregation	X						1
Annotation	X						1
Analytics	X	X	X	X	X		5
Classification						X	1
Cleaning or Cleansing	X	X	X			X	4
Compression		X					1
Clustering			X			X	2



➤ 不同物联网领域的大数据技术的分析

医疗领域中的大数据技术



Table 2

Classification of Big Data approaches applied in healthcare IoT.

Big Data	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Healthcare (35 papers)	Raspberry Pi [24], Cloud [26,27,76–78], HBase [6,79], Spark SQL [6], Hive [6], Cloudera Impala [6], Apache Parquest [6], Python-based [80], Mysql [81],	Integrity [76,77], Information Exploration [79], Anomaly Detection [80,81,84,85], Data Integration [79,82], Reasoning [82], Co-reference [82],	Semantic Analysis [24], Clustering [6,26,79–87], AHP Method [84], Classification [26, 27,79,80,82,84,87] Feature Extraction [27,77,88,89], MapReduce [6,79], Strom [6],	Data Visualization [6,79,80, 82][26,83–85] Interpretation [79,80,82] Data Lineage [76] [77] Visual Exploration [80], Visual Analytics [83]

第四部分

『物联网域与大数据』



物联网领域中大数据相关的发现



在云存储时，NoSQL比关系数据库更适用于存储IoT数据，但医疗和农业等传统行业中仍使用关系数据库，建议进行数据转换。



大数据清理包含两组关键字，一组是数据集成，它是数据分析的前提。另一组关键词是数据质量管理，用于处理低质量数据，比如数据完整性检查，数据损坏检测，冗余减少等。

物联网领域中大数据相关的发现



不同物联网域的数据在类型层面存在某些相似性，数据分析方法可以重用，例如决策树、神经网络、特征提取等。

聚类是物联网域中最重要的算法之一，同时也给研究带来挑战。



现阶段缺乏针对物联网的特定大数据可视化方法，也忽略了机器学习和数据挖掘的作用。

可视化分析是最常用的方法，但也尚未普及。

大数据技术中物联网领域相关的发现

领域	论文占比
医疗	25%
能源	17%
智慧城市	13%
农业	9%
交通	8%
工业	7%
军事	6%
智能建筑	5%

医疗领域相对成熟，吸引了大量研究人员

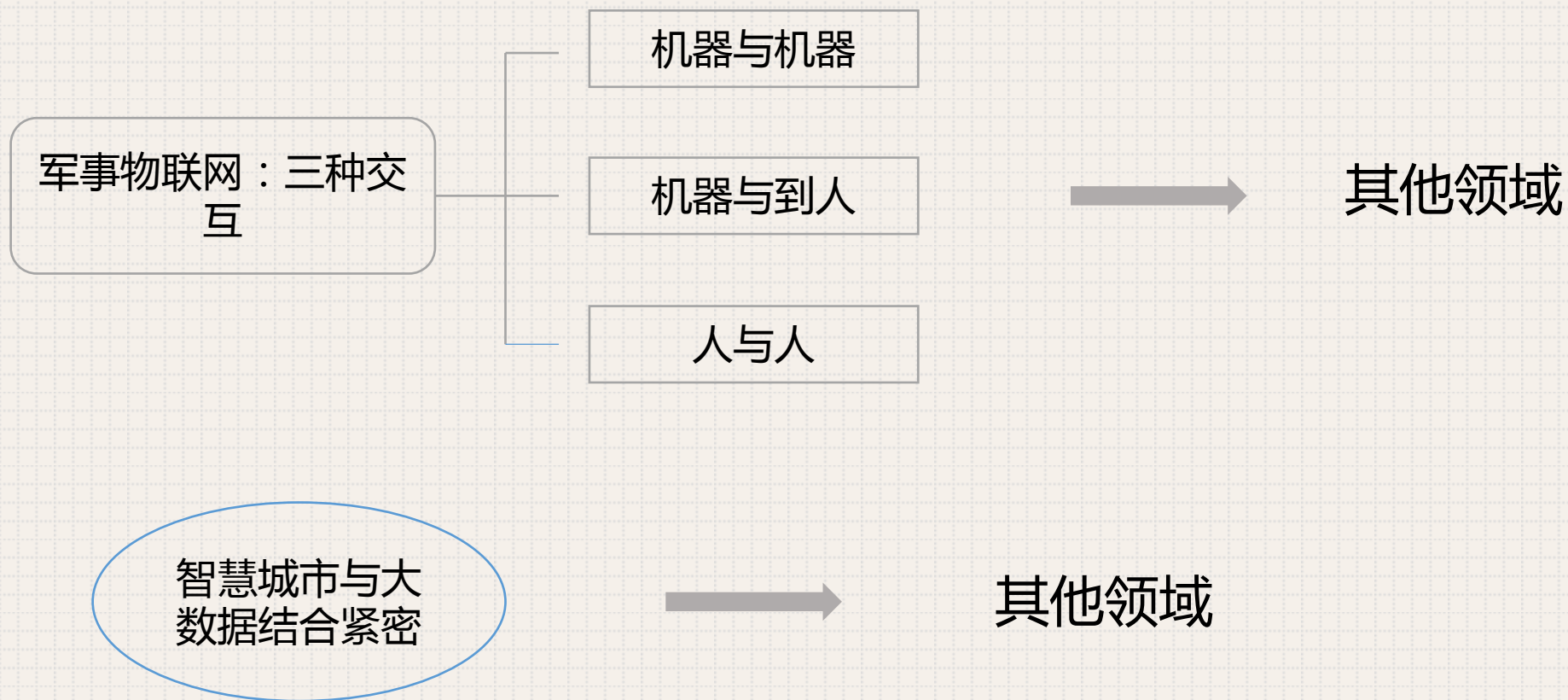
领域	NoSQL利用率
医疗	24%
能源	22%
运输	9%
农业	7%
智能建筑	7%
工业	4%
军事	7%

物联网域对特定的大数据技术普及程度不同

4

物联网域与大数据

大数据技术中物联网领域相关的发现



概念框架

=

“该领域占主导地位的技术/不同领域广泛使用的技术”

Conceptual framework of Big Data approaches applied in IoT.

Big Data	Storage	Cleaning, Cleansing	Analysis, Analytics	Visualization
Healthcare	Cloud	Anomaly Detection/Data Integration	Clustering	Data Visualization
Energy	Cloud	Outlier Detection/Data Integration	Clustering	Data Visualization
Transportation	Cloud	Data collection/Data Integration	Clustering	Data Visualization
Building automation	Cloud	Data aggregation/Data Integration	Clustering	Data Visualization
Smart Cities	Cloud	Anomaly Detection / Data Integration	Clustering	Data Visualization
Agriculture	Cloud	Data Integration	Classification / Clustering	Data Visualization
Industry	-/Cloud	Data Integration	-/Clustering	Data Presentation/Data Visualization
Military	Cloud	Data Quality/Data Integration	Data filtering/Clustering	-/Data Visualization

第五部分

『总结』



1. 对物联网和大数据领域的论文进行了广泛的调研，描述了一组典型物联网域的功能。
2. 指出了每个物联网域中使用了哪些大数据元素和如何应用于该领域。
3. 提出了一个概念框架，以指导从业人员选择合适的大数据技术，也为某一领域中如何使用来自其他领域的大数据技术提供了实际意义。

『Thanks』

