Spatial Data Quality in the IoT Era
Management and Exploitation

Huan Li, Bo Tang, Hua Lu, Muhammad Aamir Cheema, Christian S. Jensen

Team Presentation

Christian S. Jensen

Professor at Aalborg University

https://www.cs.aau.dk/~csj

Background and challenges of Spatial IoT Data (SID)

Hua Lu

Professor at Roskilde University

http://luhua.ruc.dk/

General picture of SID quality aspects

Huan Li

Assist. Prof. at Aalborg University

https://homes.cs.aau.dk/~lihuan

Techniques for SID quality management

Bo Tang

Assist. Prof. at SuSTech, China

http://acm.sustech.edu.cn/btang

Techniques for exploitation of low-quality SID

Muhammad Aamir Cheema

Assoc. Prof. at Monash U

http://www.aamircheema.com/

Trends and future directions

1.
INTRODUCTION

By Christian S. Jensen

Outline

▷ Internet of Things (IoT)

▷ IoT applications

▷ IoT statistics

▷ Data challenges

▷ Tutorial outline

Internet of Things (IoT) - I

▷ The IoT encompasses physical objects ("things") with sensing and data processing and communication capabilities.

▷ The naming suggests that the IoT is the Internet with people replaced by, or generalized to, things, but it goes further.

▷ Enabled by a convergence of technologies, just like the PC and the Internet.

Internet of Things (IoT) - II

▷ The IoT is an integral element of the ongoing digitalization of industrial and societal processes.

▷ Things
  ○ Vehicles, containers, farm animals, …

▷ Sensor data
  ○ Timestamped location samples, status information, e.g., battery level, temperature, vibration, …

▷ Data collection and analysis
  ○ Offers insight and enables purposeful action

IoT Applications I

▷ Asset tracking

▷ Waste management

▷ Utilities

▷ Predictive maintenance

▷ Healthcare

▷ Smart cities

▷ Smart agriculture, farming

IoT Applications II

▷ Defibrillators
  o Placed at locations outside hospitals to save lives, but many are defective; IoT monitoring saves lives

▷ Fill-level Sensors
  o Enables less frequent emptying of garbage, recycling
  o Enables more effective pest control

▷ Sowing and weeding robots
  o Solar-powered and geo-located robots reduce emissions

IoT Statistics [Techjury, 2022]

▷ Size
  - ○ 5.8B automotive and enterprise devices in 2020; 64B devices by 2025 (uncertain!)

▷ Value
  - ○ Potentially USD 4-11T by 2025; USD .3-1.7T in 2020; the main driver is cost savings

▷ It remains challenging to create value from IoT data, in part due to quality issues

IoT Data Challenges

▷ Increasingly massive data (80ZB by 2025)

▷ Much of it is spatially referenced: spatial, including trajectory, and spatiotemporal data.
   o We call this Spatial IoT Data (SID)

▷ Data quality challenges
   o Due to (i) limited device capabilities, (ii) inherent decentralization, and (iii) heterogeneity, e.g., positioning
   o Desired qualities: accurate and reliable; comprehensive and informative; and easy to use

Tutorial Outline

Part II: SID quality framework (Hua)
- Quality dimension, issues; means of resolving issues

Part III: SID quality management (Huan)
- Location refinement, uncertainty elimination, outlier removal, fault correction, data integration and reduction

Part IV: Exploitation of low-quality SID (Bo)
- Queries, analyses, decision making tasks

Part V: Trends and future directions (Muhammad)
- Trends, open issues and directions

Disclaimer, etc.

▷ Although the tutorial aims to cover key aspects within its scope, it is necessarily incomplete

▷ A tutorial web page is available:
https://msca-malot.github.io/sigmod22-tutorial/

References

▷ [Telenor, 2022] https://www.telenor.dk/erhverv/internet-of-things/

▷ [Techjury, 2022] https://techjury.net/blog/internet-of-things-statistics/#gref

▷ [Li et al., 2023] Spatial Data Quality in the Internet of Things: Management, Exploitation, and Prospects. *Comp. Surv*.

▷ See also the tutorial paper.