



Review

Data quality in internet of things: A state-of-the-art survey

Aimad Karkouch^{a,*}, Hajar Mousannif^b, Hassan Al Moatassime^a, Thomas Noel^c^a OSER research team, Computer Science Department, FSTG, Cadi Ayyad University, Morocco^b LISI Laboratory, Computer Science Department, FSSM, Cadi Ayyad University, Morocco^c ICube Laboratory, University of Strasbourg, France

ARTICLE INFO

Article history:

Received 31 March 2016

Received in revised form

12 July 2016

Accepted 1 August 2016

Available online 3 August 2016

Keywords:

Internet of things

Data quality

Data cleaning

Outlier detection

ABSTRACT

In the Internet of Things (IoT), data gathered from a global-scale deployment of smart-things, are the base for making intelligent decisions and providing services. If data are of poor quality, decisions are likely to be unsound. Data quality (DQ) is crucial to gain user engagement and acceptance of the IoT paradigm and services. This paper aims at enhancing DQ in IoT by providing an overview of its state-of-the-art. Data properties and their new lifecycle in IoT are surveyed. The concept of DQ is defined and a set of generic and domain-specific DQ dimensions, fit for use in assessing IoT's DQ, are selected. IoT-related factors endangering the DQ and their impact on various DQ dimensions and on the overall DQ are exhaustively analyzed. DQ problems manifestations are discussed and their symptoms identified. Data outliers, as a major DQ problem manifestation, their underlying knowledge and their impact in the context of IoT and its applications are studied. Techniques for enhancing DQ are presented with a special focus on data cleaning techniques which are reviewed and compared using an extended taxonomy to outline their characteristics and their fitness for use for IoT. Finally, open challenges and possible future research directions are discussed.

© 2016 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	58
2. Data and DQ in IoT	59
2.1. A new data lifecycle	59
2.2. IoT data characteristics	59
2.3. Definition of DQ and DQ dimensions	60
2.4. DQ and DQ dimensions for IoT	60
2.5. DQ dimensions for IoT domain-specific applications	61
2.6. DQ dimensions trade-offs	61
3. Factors endangering IoT DQ and their impact	62
3.1. Factors affecting IoT DQ	62
3.2. Layered distribution of factors threatening DQ	62
3.3. Impact on IoT DQ	63
3.3.1. Impact of deployment scale, failing-dirty, vandalism and environment on DQ dimensions in the IoT	63
3.3.2. Impact of resources constraints, unreliable sensors, network, security vulnerability and privacy preservation processing on DQ dimensions in the IoT	64
3.3.3. Impact of data stream processing on DQ dimensions in the IoT	65
3.3.4. Impact of different IoT problems on ease of access, access security and interpretability DQ dimensions	66
4. Manifestation of DQ problems in IoT and their symptoms	66
4.1. DQ problems' manifestation classes	66
4.2. Symptoms of DQ dimensions difficulties associated with DQ problems classes	67

* Corresponding author.

E-mail addresses: aimad.karkouch@ced.uca.ac.ma (A. Karkouch), mousannif@uca.ma (H. Mousannif), Hassan.al.moatassime@gmail.com (H. Al Moatassime), noel@unistra.fr (T. Noel).