



GGE6900 Research methods

Reading Skills Assignment

Part 2

Professor Monica Wachowicz

Nasrin Eshraghi Ivvari

October 2019

Table of Contents

Introduction	3
Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring	6
Untraceable Sensor Movement in Distributed IoT Infrastructure	8
Big data analytics: a survey.....	4
BIG Data collection: analysis and processing of Efficient IoT based sensor	10
T-Patterns Revisited: Mining for Temporal Patterns in Sensor Data.....	11

Introduction

Find the most useful articles for you to understand the research challenges in your field. (Minimum 5).

- Find the most cited articles on your research topic. (Minimum 5).
- Upload the articles to D2L. Bring them to the classroom.
- Answer the following questions for each article that you have found and carefully read:
 - Has the author formulated a problem/issue?
 - Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?
 - Could the problem have been approached more effectively from another perspective?
 - What is the author's theoretical framework (e.g., engineering, computer science, geography)?

As far as I searched, most of article in my research area are not as most of cited papers and because this area is very new topic there are few journals are available.

For finding the most cited article, I searched google scholar but I figure out that appropriate way is going through web of science. In the top of the web page, there is a filter named times cited. I applied that to find papers.

Big data analytics: a survey

Authors: Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos

Publication date: 2015

Cited by 295

❖ Has the author formulated a problem/issue?

Yes, the era with big data comes and the traditional data analysis cannot handle this huge amount of data. The most challengeable issue is the efficiency of analyzing big data and finding mining algorithms.

In this paper, these two issues are considered.

❖ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?

Based on the data analytics part, in the first phase which called data input, the gathering, selecting, preprocessing and transferring are in this phase but in my point of view, the preprocessing phase has a different role. Then, the author can give a separate phase to it and introduce some methods in this phase like cleaning and filtering.

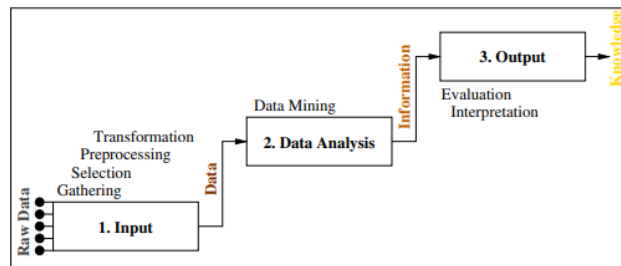


Figure 3 the process of knowledge discovery in databases

❖ Could the problem have been approached more effectively from another perspective?

One of the issues which this paper is considered is memory cost for big data but the other very important metric is time and how to reduce the time for analyzing data in a data science area. The author mentions the sampling method but not specifically introduced.

❖ What is the author's theoretical framework (e.g., engineering, computer science, geography)?

(Computer science)First, in the data analytics part there is a quick introduction to data analytics. As far as analyzing the large scale and handling input data are exist in data analytics there are some methods presented to increase the computational time. These methods listed below:

Problem	Method
Clustering	BIRCH
	DBSCAN
	Incremental DBSCAN
	RKM
	TKM
Classification	SLIQ
	TLAESA
	FastNN
	SFFS
	GPU-based SVM
Association rules	CLOSET
	FP-tree
	CHARM
	MAFIA
	FAST
Sequential patterns	SPADE
	CloSpan
	PrefixSpan
	SPAM
	ISE

Second, this paper discussed big data analytics. Big data is massive, high dimensional, complex, noisy and unstructured.

This problem causes to increase the importance of preprocessing level. Various solutions presented like Hadoop, NvidiaSUDA and so on. The most useful platform and framework is cloud computing.

Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring

Authors: Ming Jina, Shichao Liub, Stefano Schiavon, Costas Spanos

Publication date: 2018

Cited by126

❖ Has the author formulated a problem/issue?

Yes, indoor environmental quality is an approach which collects environmental measurement based on the sensor's networks but they are limited in scale and adaptability to infrastructure cost and maintenance.

Based on the above issue, an automated mobile sensing system is proposed.

Based on this system another issue is come up which is data collection is sparse in the joint temporal and spatial domain and cannot use directly from IEQ evaluation.

To deal with this one, the algorithm for Spatio-temporal interpolation is developed.

❖ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?

Yes, based on the table of sensing modules of ESP, the author wrote down all required parameters with price, module, and performance.

For implementing the algorithm, they mentioned that the samples are located in multiple locations but did not give any information about the area and the condition.

The author first introduces a robotic base for environmental sensing platform which is the first step for this system

❖ Could the problem have been approached more effectively from another perspective?

Based on the methodology which they used, they just monitored temperature and humidity which they mentioned these are the only environmental parameters that can be measured but other parameters like motion didn't mention.

- ❖ What is the author's theoretical framework (e.g., engineering, computer science, geography)?

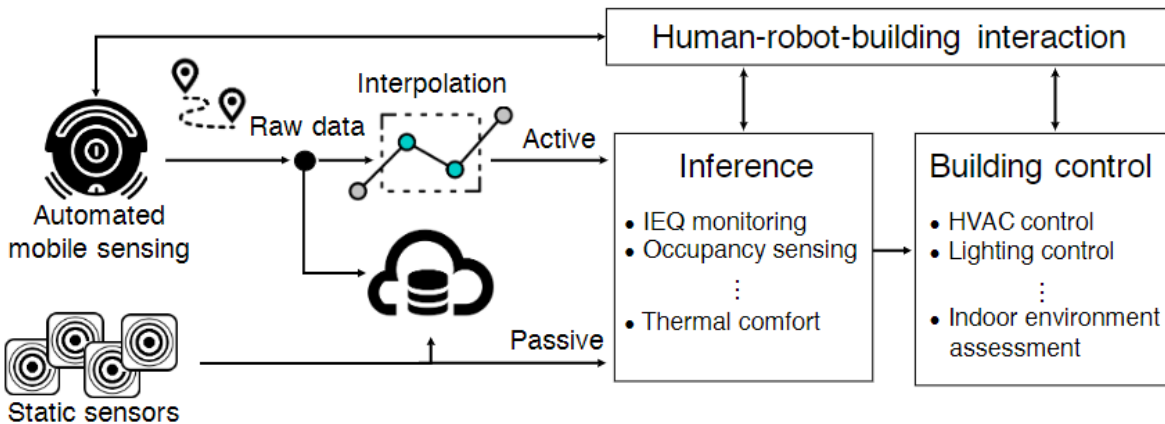


Figure 1 Automated mobile sensing system overview

The robot operating system runs the mapping and navigation algorithm. The first step, Indoor positioning without the detail of the floor map. The SLAM problem was solved based on the particle filter.

Secondly, navigation and collision avoidance, the robot finds the shortest path based on the current knowledge using a global planner.

Last but not least, the Spatio-temporal interpolation algorithm. This algorithm has three main steps; ST binning to reduce measurement errors, global trend extraction, to measure the global variation and local variation estimation, applied on unknown points.

Untraceable Sensor Movement in Distributed IoT Infrastructure

Authors: Prosanta Gope and Tzonelih Hwang

Publication date: 2015

Cited by 56

❖ Has the author formulated a problem/issue?

Yes, based on the article title some issues on movement are considered. The problem is some properties in IoT systems, such as sensor anonymity, sensor untraceability, resistance to replay attacks, cloning attacks should be considered.

On the other hand, one of the most important issues is the security and privacy of each sensor node in the IoT infrastructure which this paper focused on the distributed IoT nodes.

❖ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?

Yes, based on the figure1, exactly mentioned about project scale and components and how they are related together.

❖ Could the problem have been approached more effectively from another perspective?

There are other secure algorithms that exist which can use to for this authentication design. Maybe the can suit for this proposed research.

❖ What is the author's theoretical framework (e.g., engineering, computer science, geography)?

First the distributed IoT system architecture is proposed.

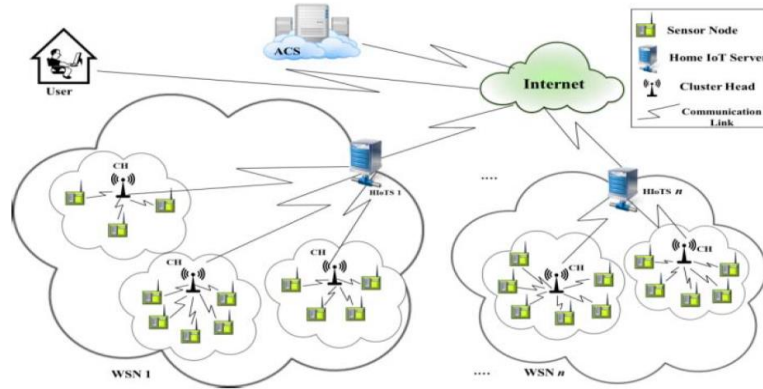


Figure 2. Proposed IoT architecture

As it is shown, it has four components: authenticated cloud server (ACS), two types of the network like cluster head and home IoT server, and sensor nodes. Based on this method, all sensor nodes need to registers on the network and all users register to ACS. For interacting networks, ASC can help for authentication to establish secure communication.

Then, a lightweight anonymous authentication scheme was designed, which can guarantee various security issues related to the privacy of the sensor node like anonymity, untraceability, replay attacks, DoS attacks, etc. To design the lightweight authentication framework for IoT, we will use the lightweight cryptographic primitives like the hash function and bitwise exclusive-OR, where these cryptographic primitives cause less computational overhead and reasonably much less execution time as compared to other cryptographic primitives like asymmetric encryption/decryption, modulo operation etc., which is highly adequate for the tiny-powered sensor node.

BIG Data collection: analysis and processing of Efficient IoT based sensor

Shruthi BM, Pruthvi PR, Parashivamurthy BM, Priyanka KR

Publication date: 2017

Cited by 93

❖ Has the author formulated a problem/issue?

Yes, they tried to find a solution to collect and manage data which gathered from IoT- based sensors by a remote device which can be energy efficient in smart buildings. They served for technologies to figure out their common operations and combined their functionality to have a beneficial use.

Other challengeable subjects in this area can be QoS and security and how to protect these sensitive data.

❖ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?

The scope did not clarify exactly, but they want to implement such a system in the University of Mysuru. The author mention that the Contiki simulation is used to implement their proposed system. The experimental result of this simulator and data collected are shown with all detail needed.

❖ Could the problem have been approached more effectively from another perspective?

The topology which they used is hybrid. This topology is reliable and also managing clients are easier but on the other side, this is the most expensive one that can be implement. The number of nodes between each station can be equal to number of all machines and stations.

On the other hand, monitoring data is very crucial which this research did not pay attention to it.

❖ What is the author's theoretical framework (e.g., engineering, computer science, geography)?

First the architecture of the proposed system is achieved which shown the communication between users and sensors proposed, this connection which is defined as a topology.

This communication in the network is based on the IPv6 and NAMRTP real time protocol for remote transferring data from remote sources to the data base.

This theory was implemented on the Contiki operating system. This software is useable for collecting large amount of data.

T-Patterns Revisited: Mining for Temporal Patterns in Sensor Data

Albert Ali Salah 1, Eric Pauwels , Romain Tavenard and Theo Gevers

Publication date: 2010

Cited by 23

❖ Has the author formulated a problem/issue?

Yes, large amounts of simple sensors as opposed to a few complex sensors to monitor places and systems create a need for temporal pattern mining algorithms to work on such data. The existing method has several shortcomings. Using a Driving pattern from the movement of the object in the environment is challengeable. First of all, in the real world, action patterns by using multi factors like users or at the same time by using of single actors performing multiple actions.

On the other hand, the patterns are in a different time gap, and this time gap between events that have a large variation is different. Consequently, detecting these patterns becomes a very difficult issue, and most of the old pattern analysis methods are not usable.

❖ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?

Yes, the best thing I saw in this paper was about their new method and the previous one. First, the author introduces the first method completely with all mathematic formulation and then in the separate section in the methodology part, the new method is introduced and compared all differentiations and improvements.

❖ Could the problem have been approached more effectively from another perspective?

In my perspective, I think adding the data mining method to this approach is very different and can improve the performance.

❖ What is the author's theoretical framework (e.g., engineering, computer science, geography)?

The t-pattern(or better to say GMM T-pattern) method which is talked in this paper, symbolic time series is investigated that each of these symbols represents the event or activity with the principal aim of reducing possible relationships between each two

symbols and building a tree that makes of dependencies of these symbols in a hierarchical way.

But this paper introduces two modifications to t-pattern algorithms. This algorithm is more robust.

Repeating expounded in the preceding increases the risk of false-positive which means between sensors this issue is increasing the correlation. To avoid it there are some ways in which one of them is Bonferroni correction which in this research was defied and applied.