International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 29051-29060

http://www.ripublication.com

Internet of Things: A Vision Through Data Mining, Cloud Computing and Semantic Technologies

[*1]Sini Raj P , [*2]Y Reethu Sasira, [*3]L Revathi, and [*4] Su Rubaasri
[*]Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, India.
India.
Email Id –[1]p_siniraj@cb.amrita.edu, [2] reethusasira@gmail.com, [3]revathinani95@gmail.com, [4]rubaasri6@gmail.com.

Abstract

Connecting to different people and objects around us in our surrounding for various purposes is an important aspect in every individual's day-to-day life. This connectivity is focused by *Internet of Things (IOT),* which ultimately aims at establishing connectivity between objects regardless of time and location. This paper presents an idea of how Internet of Things happen from the view of three important and emerging fields of Computer Science namely, *Data Mining, Cloud Computing and Semantic Technology.* Development of wireless technologies such as Radio-Frequency Identification(RFID) , wireless sensors and actuators, Wi-fi, Bluetooth etc. has propelled the advancement of IOT and hence combined applications from different fields paves way for effective use of technology and also helps in betterment of human life. The paper is intended to present the usage of various techniques of above mentioned areas to bring together different objects or things around us so as to enable connectivity among them and hence support accessibility to various things to satisfy daily needs.

Keywords: Internet of Things, Cloud Computing, Data Mining, Semantic Web.

Introduction

In our day-to-day life there is need forhuge amount of data to be exchanged and processed for various purposes. This exchange of data thus forms the network which in turn paves way to a paradigm known as the *Internet of Things(IOT).* The main goal of networking has been to interconnect people anywhere and at anytime. The next revolution is to establish connection between objects or "things" so as to create a smart environment and also to provide accessibility to anything from anywhere and at

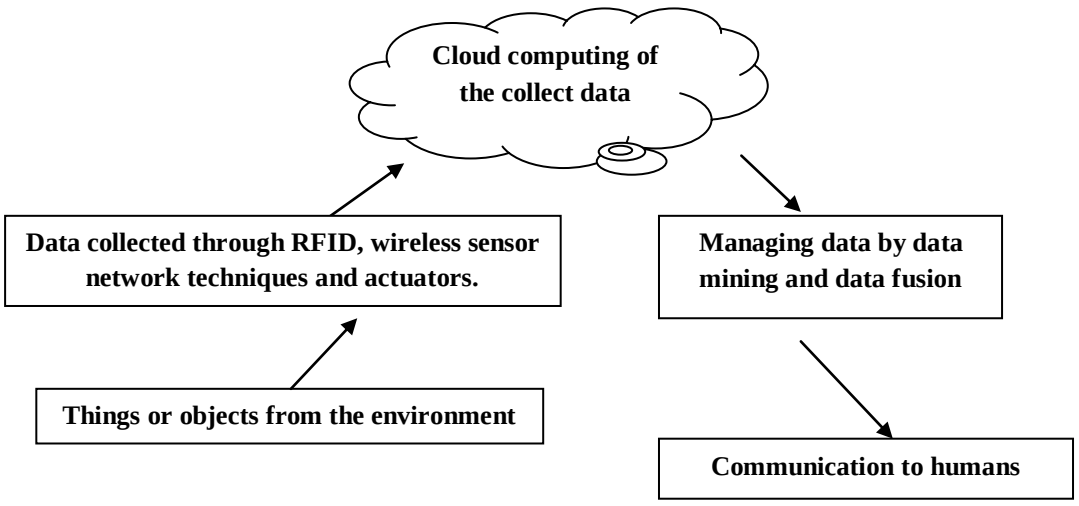
anytime. In this paradigm many objects surrounding us in the environment will be on the network in one form or the other, interact and communicate to other objects, machines, environment, infrastructure to make human life easier. The term Internet of Things was coined by Kevin Ashton in 1999 in the context of supply chain management, but the definition has been extended to a wide range of applications such as intelligent transportation of goods , healthcare, utilities, enhanced and assisted living, etc. The advancement in wireless technology such as Bluetooth, Wi-Fi, Radio-Frequency Identification (RFID), telephonic data services, embedded sensor and actuator nodes, etc., facilitates information transfer to actuate, command and control things from anywhere and at anytime. They are recognised as the **"Atomic Components that link the real world with the digital world".**

As the full deployment of IOT has to start from the augmentation in the **Things' intelligence** , IOT is responsible for the evolution of the new concept called **Spime. Spime** is defined as an object that can be tracked through space and time throughout its lifetime and that will be enhanceable, sustainable and uniquely identifiable.

Effectively analysing and understanding the required knowledge from the massive amount of data is hectic and the best way to handle such a large amount of data produced by IOT, in an effective, accurate and secure way is done by combining data mining, cloud computing and semantic techniques, so as to obtain desired results from the fetched data. Hence this paper presents an overview of how the ideologies of IOT are implemented from the perspective of data mining, cloud computing and semantic fields enabling technology at every individual's doorstep and better use of available services .The following figure shows the process of IOT.

The Process of IOT:

The paper constitutes five sections. In the remainder of the paper, the concepts of IOT realised from cloud computing point of view is presented in section 2, the data mining techniques that are useful in arriving at capabilities of IOT ideas is summarised in section 3, and semantic technology that is helpful to implement IOT paradigm is given in section 4, and the consolidated summary of the concepts is given in section 5.



Internet of Things In The View of Cloud Computing

The ultimate vision of IOT is to deliver ubiquitous computing that can be commoditized in a manner similar to utilities such as water, electricity, gas, etc. Thus cloud computing, that supports any IT service that can be consumed as utility and that which can be delivered through the network can be combined to form a framework with ubiquitous sensing, which not only allows flexibility in investments in a logical manner but also is highly scalable. This proves to be beneficial as there are no upfront commitments, on demand access, good pricing, simplified application acceleration and scalability, efficient resource allocation, energy efficiency and seamless creation and the use of third-party services. The cloud integrates required data by providing scalable storage, necessary computation time and other tools to build new businesses or environments.

This perspective includes different aspects such as

1. Infrastructure
2. Development platforms
3. Application and services

These aspects are delivered as services, i.e Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) . The remainder of this section focuses on two services namely Amazon Elastic Cloud and Microsoft Azure which are taken as instances to explain how IOT visions are established through cloud comluting concepts.

IAAS

Iaas is centered around a model of service delivery that provides a predefined, standardized infrastructure specifically optimized for customer's application. Customers owner and manage their applications while hosting operations and infrastructure management are taken by Iaas providers. The different layered components necessary for setting the requitred environment are computer hardware ,computer networks , internet connectivity, platform virtualisation environment for client – specified virtual machines ,service level agreements and utility computing billing.

Resources such as software, network equipment, servers, data center space etc., are provided as fully outsourced service. Similar to a utility company that bills a customer, this service is also billed on monthly basis usually, thus provisioning the fundamental infrastructure to facilitate the requirements of IOT. The customer is charged only for the resources consumed on pay-per-use basis.

Amazon Web Services

An example for product that delivers Iaas service is Amazon Web Services (AWS) which is a collection of web services. They provide storage, compute and more advanced services and are popular for their Iaas service and elastic cloud service, i.e Amazon's Elastic Cloud (Amazon's EC2). EC2 presents a true virtual environment for computing purposes thereby allowing usage of web-based interface to obtain and manage services needed to launch a variety of operating systems. Clients can load the OS environments with their customized applications and manage their network access

permissions to run as many or as few systems as needed. EC2 is used by creating Amazon Machine Image (AMI) wherein all required configuration settings are made available. Once the AMI is created it is uploaded into Amazon S3 which is a repository that provides authentic, secure and quick access to the client AMI. EC2 is also used to run instances. EC2 instances, which represent virtual machines are created by default with the kernel and disk associated to the AMI.

Benefits

First and foremost, Amazon EC2 provides financial benefits to the customers. Owing to Amazon's broad scale and huge customer base, the cost for setting up and running an operation becomes inexpensive for individual customer as it is being shared by many customers. Customers pay low price for the services they actually consume. EC2 provides dynamic scalability by enabling the users to increase or decrease capacity within a few minutes span of time. Users are allowed to invoke as many instances as possible simultaneously. Also users have total control of their instances EC2 gives the user flexibility to configure memory, CPU and instance storage as preferred by the user. There are other Amazon Web Services such as Amazon Simple Storage Service, Amazon Simple DB, Amazon Simple Queue Service, Amazon Cloud Front. EC2 provides integration with other AWS to facilitate complete computing, processing and storage over vast range of applications. It also enables usage in different geographical areas, i.e any number of instances can be situated in disparate locations. EC2 locations consists of Regions and Availability zones. Regions consists of one or more Availability zones while Availability zones are distinct areas that are protected from failures in other Availability zones, provisioning low cost, low-latency network connectivity to other Availability zones in the same region. United States and Europe are the two regions which currently possess Amazon EC2.

PAAS:

Paas is an outcome of SaaS application delivery model. This service model enables to support complete life cycle of building and delivering web applications and services available from the internet without any software downloads or installation. Paas developers mainly focus on web-based developments rather than concerning the operating system which is being used. Hence anyone with a connection to Internet can build powerful and useful applications for global use thus making IOT possible across the globe. Paas services include developing, testing, deploying, hosting and managing applications to support the application development life cycle. It is also useful in integrating with web services and databases, i.e it offers creation of combination of web browsers called mashups and also accessibility to databases.It enables reuse of services maintained inside private networks. The traditional On-Premises model of building and running applications was found to be complex with high pricing and involved several risk factors. Hence the new cloud model facilitates new capabilities to new makers through the web browsers.

Microsoft Azure

Microsoft Azure is a cloud operating system that provides developers with a collection of services for building application with cloud technology. Azure is built on Microsoft data centers' infrastructure. The Azure platform consists of a foundation layer and set of developer services that are used for building scalable applications. The services include compute, storage, networking and identity management that have been tied together by middleware known as App Fabric. App Fabric is a comprehensive middleware that is helpful in integrating existing cloud services. It offers collection of services such as communication, authentication and authorisation, and data access. Azure provides scalable runtime environment for web applications and distributed applications in general. The different services are explained below.

Compute Services :

This type of service is the central component of Microsoft Azure and they are delivered by means of abstraction of roles. A role is defined as a runtime environment that is customized for particular compute task. At present there are three different roles namely,

i) Web Role: This role is designed for implementation of scalable web applications and they represent the units of deployment of web applications within the Azure infrastructure. And are being hosted by IIS 7 web browser, which is a component of Azure infrastructure.
ii) Worker Role : This role is designed for hosting general-compute sevices on Azure. They can either be used for fast provision of compute power or to host services that do not interact with the external world through HTTP.
iii) Virtual Machine Role : This role is designed to allow developers fully control the computing stack of their compute service. It is based on Windows Hyper-V virtualization technology.

Storage Services:

Azure provides various storage solutions with more durability and redundant option in comparison with local storage. These services can be accessed by multiple users at the same time from any location thus enabling a general solution for storage. The following are the services offered.

i) Blobs : Binary Large Objects (BLOB) allows storage of huge amounts of data and this service is optimal for storing . There are two types of Blob namely, Block Blobs and Page Blobs.
ii) Azure Drive :A Page Blob mounted on an NTFS tree is called Azure Drive.
iii) Tables : Tables consist of a semi-structured storage solution, enabling users to store information in the form of entities having a collection of properties. Entities are stored as rows into the table.
iv) Queues : In order to prevent loss of messages and to ensure processing of messages Queue storage type is used to allow applications to communicate by exchanging messages through durable queues.

Network Services

An important area for these services is the connectivity of applications. Windows Azure Virtual Network that includes Windows Azure Connect and Windows Azure Traffic Manager is the network service offered by Azure. It facilitates easy setting up of IP-based network connectivity and is particularly used in VM roles. It provides three different load-balancing strategies to the developers, namely, Performance, Round-Robin and Failover. Windows Azure Content Delivery Network (CDN) is the content delivery network that improves content delivery capabilities of Windows Azure Storage and other Microsoft services. Azure platform can also be deployed as an appliance on third-party centers which include Windows Azure, SQL Azure and Microsoft-specified configuration of network, storage, and server hardware. Apart from above mentioned services, software is offered as a service (SaaS) wherein software is distributed to the clients from the network. Many types of software like email software, human resources, IT security, vedio conferencing, web analytics etc., are made available for users to be purchased and installed in their devices at anytime from anywhere so as to achieve the aims of IOT. The service ensures availability of applications at all locations across the globe. It provides streamlined administration, automated update and patch management services, data compatability across the enterprise, enterprise-wide collaboration, global accessibility. There are different products available for this purpose for instance SalesForce.com and Force.com is a Software-as-a-Service solution that allows prototyping CRM applications. Cloud Computing thus integrates various services to ultimately rendor computing as a utility to the customers thus facilitating the concepts of IOT come true.

IOT from the perspective of Data Mining

As the name implies , data mining is the process of exploration and analysis of large amount of data in order to discover new patterns and relationships within the data.The fundamental activities of data mining are

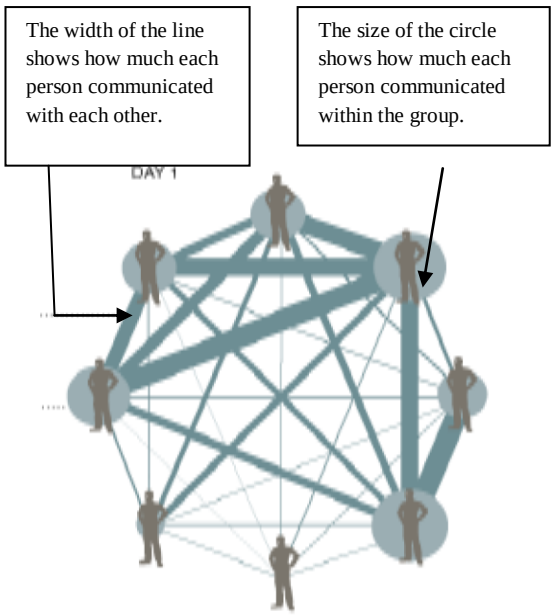
1) Classification
2) Estimation
3) Prediction
4) Affinity grouping (Association rules)
5) Clustering
6) Description and visualisation

Data mining is also known as KKD (Knowledge in the Discovery of Data). There are two types of data mining - directed data mining and undirected data mining .Classification , estimation and prediction are examples of directed data mining , while affinity grouping , clustering and description and visualisation are examples of undirected data mining. Data mining plays an important role in the internet of things and can be referred as the building block of IoT.

Data mining is accomplished by building models which uses an algortihm to act on a set of data. The process of appllying the model to new data is known as **scoring**.

Example:
Eight people working on a task, had their communications tracked for seven days by collective intelligence techniques. The pattern of their interaction was monitored , at the end of each day. By the end of the week, their ineraction had increased greatly and the group dynamic was much more successful.

Looking at the data collected, we compare the amount of interactions between different persons and the interaction between each person and the group, and then group them to different levels like low , medium , high etc. and then we register the time for each grouped activity. This process is called data clustering.



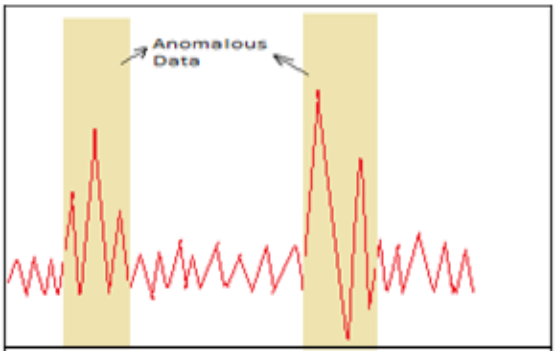
Data minning helps to discover new patterns from large stores of data with the help of algorithms to analyse the data and evaluate the probability of future events.

Therefore the key features of data mining are:

- Discovery of patterns
- Prediction of the outcomes
- Creation of actionable information
- Focus on databases

The most important applications of data mining are data anomaly detection, data clustering , data classification , feature selection and time series prediction.

Data Anomaly Detection
Data anomaly detection helps to identify the data that is different from the other data. It helps to detect the problem and unusual behaviour in the data.



Data anomaly detection technique assists in discovering new phenomena and improves the analysis of data objects. From the above example , the data shows how many persons never interacted in the group at all and how many communicated for very long hours and by featuring the data mining techniques , each and every member of the group may recieve a notification of the people who have not made any interaction. Thus anomaly detection system alerts us , of any unusual behaviour of the data.

Clustering of Data

Data Clustering refers to natural grouping of data with similar features and characteristics. No prior knowledge is required , since the data will be grouped , just by their specific values without any supervised learning. Thus data clustering indirectly paves the way for the discovery of new patterns.

Data Classification

The collected data is revised carefully and then they are classified according to the specific features based on their similarities and differences . Once the whole data is classified , the additional data is then added to the corresponding levels. Thus data classification determines the overall activity at the end.

Feature selection in IoT

Feature selection is a term that is most often used in data mining.It focusseson the tools and techniques available for reducing the size of the collected data for better processing and analysis.The attributes are selected on the basis of their utility.The prime idea of feature selection is to keep only required details and eliminate all the unnecessary data. This prevents the wastage of memory space.Thus feature selection helps to solve the twin problems of having too much data that is of less value or less data that is of high value.

Time Series Prediction

As its name implies , it is the prediction of the future data for particular inervals of time. Time series prediction is mainly applied in the field of weather forecasting and also in meteorology. The clustered and classified data is revised carefully for the estimation of the future data.

Semantic Technologies For Internet of Things

The rapid increase in the network enabled devices and sensors used in today's world is changing the information communication network , services and its applications in various fields. Within the next decade millions of devices will generate huge amount of real world data for applications and services in various fields like smart homes, healthcare, automotives and logistics. The solutions and the related technologies that allow us to combine the real world data and the services into the information networking technology are often described under the term of "Internet of Things".

The main idea behind the interconnection of devices (i.e sensors), collecting the data and processing it is to enable the humans and machines better understand their surrounding environment. The data collected by different sensors or devices is usually 'multi modal' (sound, light, temperature etc) and diverse in nature (quality of data is time dependent and also depends on the location , at times the data can be noisy and incomplete). This ambiguity, diversity makes the job of processing, combining and interpreting the real world data a very tough and challenging task. This implies that enabling inter-operability among the 'things' if IOT is the basic requirement and this can be achieved by applying " Semantic Technologies " to IOT. This enhances the inter-operability among various sources - data providers – dconsumers. At different levels of IOT the data annotation (i.e. attaching a comment or explanation to the original data) and semantic analysis of the description can be used .

The list of technologies developed in the Semantic Web like

- Ontologies
- Semantic annotation
- Linked data
- Semantic web service can be used as solutions for the purpose realizing the IOT.

Semantic web of things is a new field that combines IOT and semantic web technologies. It aims at changing the current web data which mostly consists of unstructed or semi-structured documents into a webdata.

IOT Data Abstraction Ad Access

Data abstraction is concerned in a way how the physical world data is being integrated. Recently, ontologies such as W3C's and SSN ontology have been developed to provide a number of constructs to formally describe not only snsor resources. Data access in IOT can be implemented at low-level by the use of low-level programming language and operationg systems . It is obvious that heterogeneity of the devices ans sensors made make data access a tough task.

In this paper , we saw the main reason behind applying semantic technologies to IOT. They are many other fields in IOT to which semantics contribute and the reasearchers will continue studying the use of semantic technologies in IOT. Also it requires support from other fields like Data mining, Cloud Computing, Service computing to enhance the processing and interpretation of semantic data in IOT domain.

Discussion

Scientific inventions have led to a great revolution in human life and one among them is the concept of IOT wherein the devices or physical objects are connected to the network. It is like transforming a fictious idea into a real one. We get a chance to access information about dthe objects from anywhere in this world through IOT. Data mining , cloud computing and semantics play a vital role in achieving IOT. The procedures of data mining , cloud computing and semantics in IOT are succinctly explained in this paper. Thus in a nutshell , IOT enables a communication with the things around us with the help of the above explained process.

Conclusion

Scientific inventions have led to a great revolution in human life and one among them is the concept of IOT wherein the devices or physical objects are connected to the network. It is like transforming a fictious idea into a real one. We get a chance to access informaion about the objects from anywhere in this world through IOT. Data mining , cloud computing and semantics play a vital role in achieving IOT. The procedures of data mining , cloud computing and semantics in IOT are succinctly explained in this paper. Thus in a nutshell , IOT enables a communication with the things around us with the help of the above explained process.

References:

[1] Atzori L, Iera A , Morabito G. The Internet Of Things: A Survey, Computer Networks. vol. 54. June 2010. p.2787-2805.

[2] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IOT) : A vision architectural elements, and future directions. Future Generation Computer Systems, vol.29, 2013 Jan. p.1645-1660.

[3] Buyya R, Vecchiola C , Selvi S T. Mastering Cloud Computing. McGraw Hill Education Private Limited, New Delhi.

[4] Sosinsky B, Cloud Computing Bible. Wiley India Private Limited, 2013.New Delhi.

[5] Rittinghouse J W, Ransome J S, Cloud Computing Implementation, Management and Security. CRC Press, London. New York.

[6] Berry M J A, Linoff G F. Mastering Data Mining. Wiley Computing Publishing, Canada.

[7] A.K.Pujari, Data Mining Techniques, Universities Press, Hyderabad, India, 2001.

[8] Yan Z ,Yang L T. Fusion - An aide to Data Mining in Internet of Things, Information Fusion. vol. 23.2014 Aug. p.1-2,

[9] Barnaghi P, Henon C , Taylor K. Semantic Technologies for the Internet of Things : early progress and back to the future. International Journal on Semantic Web and Information Systems, IGI Global, 2012 Sep.

[10] Pat Doody, Andrew Shields. Mining Network Relationships in the Internet of Things. [Internet]. Available from: http://www.iot-butler.eu/wp-content/uploads/2012/11/Mining-Network-Relationships-in-the-Internet-of-Things-x.pdf.