

Study on the Technology of Network Data Mining Based on Cloud Computing

Xiao-duan Wang, Li Wang and Zhen-tao Ni

Informatization Department, Hebei Academy Of Governance, Shijiazhuang, 050000, China

Keywords: Data mining techniques; Crowdsourcing; Technology of network; Internet of things; Cloud computing

Abstract. There are several things from the fusion of fusion technology. Therefore, in the cloud, things can be called a new paradigm. With computer operations, generalized computing, and related technologies more closely interdependent, a system has been formed. This system is a world of things. More and more people are now beginning to enjoy the presence of two worlds: the real world and the digital world, more and more people like this new way of life. In between these two worlds is constantly symbiotic interaction. Under the cloud storage conditions, one thing appears particularly important. It is to protect online users such as user name and password. Mining Association classification data has proven to be a clever way. This method seems to be effective in detecting phishing sites precision and handling potentially dangerous. It is also the solution to the problem of insecurity. In this paper, first part, in the cloud storage conditions, we introduce the relationship between the data mining and the internet of things. Then, we study the potential solutions. In the end, this paper presents findings: this paper presenting the solution is an effective way to protect the user's information security.

Introduction

The several years ago, there are some mobile devices and ubiquitous services those make a chance to people to communicate with other people wherever or whenever. The field of electronics and the wireless communication systems is the reason to the host of level to the advancement. Until now, the role that the devices played is could connect users through the internet without limitation. And interlink between the physical world and the cyber world is becoming expanding the opportunity, leading to the establishing of cyber physical systems.

In 2005, the international telecommunication union launched its first report on the subject, that the term of 'Internet of things' starts from the official room and has the relationship with the researchers industries and end-users. In the papers written by the authors, the international telecommunication union use the following sentence to describe its vision: 'the field of information and communication technologies has a new remember, a beautiful dimension: we could use the any channel what we want make it keep connect whatever the time, the place, if you are online. Connections will multiply and create an entirely new dynamic network of networks an Internet of Things'. Here the emphasis is on the fact that not only RFIDs, but also a high number of different objects – univocally addressable – constitute the underlying fabric of IoT.

The cooperators could run their business and trade each other and the host of the shopkeepers could offer or provide service or sale the goods, due to numerous organizations using the clouding services.

This paper is divided into different sections. Section 2 reveals a host of ways to data mining in the clouding service. In the clouding, data mining's access, the Apriori algorithm to scan the transaction database and the mathematic ways are setting forth in Section 3. Section 4 we discuss the key point in the data mining. Lastly, conclusions are described in Section5.

The access of data mining in the clouding computing

In this section, there are two important parts. One is revealing the data mining in the internet of things and the document of the phishing. Those are interpreted in the part. Another one illustrates the cycle about the phishing life.

Data mining.

In the Internet of things, the vision recently extends to merge two different concepts: (1) the Web 2.0, is expected to a large number of user interaction and (2) self-dependent, especially respect the individual participants may benefit. In particular, for Web 2.0 technology, its main advantage is the use of simplified and intuitive interface to enable users to provide Web contribution, regardless of their technical expertise. It is the interaction between the basic things and the user will be the future network of a core problem.

Internet access layer to realize data collection, extract the key data and the transmission of critical data to the data integration layer. Internet data access layer includes various sensing terminal, such as GPS sensor terminals, RFID sensor terminals, terminal video sensor, wireless sensor, etc. Use of these sensing terminal monitor real world objects, reaction monitoring the state of an object, and other information collected and sent to the corresponding computing nodes. To collect data including text, images, data and video data, etc. The compute nodes are sensing data preprocessing. It aims to extract the key data and transmitted to the data integration layer, the distributed space-time database for the Internet of things.

The central part in the whole process is data mining exploitation stage. Classification, clustering, association rules mining is data mining tasks may be divided. In many areas, most face recognition classification is that the disaster relief.

Ways to minimize un-security.

Associative classification in data mining is a promising method, can use special extraction from fishing and legal website, found. Email may be soft phishing attacks. It could be described, phishing attacks seem to be a real from the user group. Through a link in an email, asking them to change their personal information.

Under normal circumstances, these two methods, to combat phishing attacks most technology belongs to the blacklist and heuristic based. Anti-phishing technology success depends on identify phishing site and think deeply about an acceptable time. Blacklist is considered to be malicious and collect the technology is already in use, such as the user's list of URLs. Another method is based on the comparison of rules by collecting a series of different features, exposure classification algorithm. All they need to work methods and data mining.

Clouding computing.

In cloud computing platform, we know in the form of C/S (client/sever) application software will be less and less, instead, would be a large number of B/S (browser/sever) in the form of programs, then the user only needs to have a network FFL web-browsing and meet the data transmission network bandwidth can be fast and the use of cloud computing provides various services, which will greatly promote the development of application services such as Web 2.0.

How to deal with huge amounts of sensory data is also the focus of the Internet of things research problem, one is very important to solve the case is the cloud computing technology, based on the existing system of the Internet to provide points of space and time sharing hardware services to users. In a cloud computing platform, the data is stored in concentrated to a safe area, so people don't have to worry about suffering from virus infection, and data loss. In addition to data centralization, the cloud computing platform was centralized computing service ability, thus can greatly reduce equipment requirements for the client, so that its use is more convenient. This can be

understood as the future of cloud computing technology will achieve with data and service as the center, instead of the traditional application centered.

Technology network process and the mining model

Interface with production facilities: one of the most important data that a complete computer system through the facilities set point input, input process value, formulation and mixing, and procedures. Therefore, interface and facilities, such as machine, the machine group, IIASMP and production lines, is indispensable. A IIASMP not only receive data from the device, and the data sending instructions. This task is represented by an intelligent assembly, integration of some heterogeneous devices, facilitate data exchange, and perform the task.

Crowdsourcing method can be used to complete various types of data mining classification task, such as clustering, semi - supervised learning, association rule mining.

Database is dependent on the development of the relationship between the data basis of data mining. Establish system analysis data sets, important work is to find an efficient way. Generally speaking, the establishment of the necessary data set is usually a collection of database tables, become together, connection and transformation.

Technology network process.

Network communication is Internet information transmission, the basis of general network communication is done by some kind of agreement between different network nodes communication and exchanges, and only the same network protocol of computer communication and the exchange of information. In fact this way of thinking is the result of communication between people, for human to realize the communication between each other must use the same language and specifications. From a professional point of view to define, the network communication protocol is to realize the communication between different computers in a network must abide by the contract. It is primarily for information transmission rate and transmission format, transmission structure, transmission control step, and error control performance normalize constraints, so using the same protocol, network node can communicate with each other.

The combination of the technology and the cloud computing.

Cloud computing in terms of resource utilization has greater advantages than the traditional Internet data center. The traditional methods are relatively low, the adopted technology of the Internet in the Internet traffic will because of time different and different, this makes a significant portion of the resources are available, resulting in low average utilization rate of resources. On the contrary is efficient cloud computing platform is provided to the user the flexibility of a scalable server, it can be allocated according to the needs of different users and release resources. Furthermore, it is huge in cloud computing platform, large users of the data, so the average utilization rate increase a lot. From the analysis on the level, cloud computing technology could recognize to the realization of the Internet greatly reduces the operating costs.

The model of the minning.

Internet of things data mining model based on cloud computing architecture is divided into five layers, respectively is Internet of things data access layer, data integration, data mining platform layer, business layer and interactive control. Greedy algorithm traverses the performed on each task in each storage cost, establish a light path and calculation of the price, by comparing the final total cost, choose the optimal storage solution. How to unify between storage cost and optical path is, of course, the problem of this solution, the specific conversion can set different generations according to the needs of actual network value.

We ran out of the road) the size of the said price. N is total number of tasks, we will store business into a data tasks and n - 1 storage tasks. T1 refers to not take any cost, different storage tasks can't execute simultaneously on the same memory. So we need to set up n - 1 the light path is used to implement n - 1 storage tasks. These variables do not exist in the database: they will change the query to obtain. The program of the method is as follows:

<pre> Determine the task list and initialize From T2 to T1 into a list LIST Sequential scheduling over the list For each task Ti belong to the LIST do Sct cost_Min (t Tj) = infinity Sct temp_j = Mji, temp_k =0 For each or do If cop (arj) >= dem() Then Calculate the CoA(Ti ,ARj) Compute shortest route Pk by OPSE Calculate the Cot () If CoA (Ti ,ARj) +Cot (Pk) < cost_Min() Then </pre>	<pre> Else Continue Endif Else Neglect the cost End if End for If cost_min (tj) not belong to the infinity Then Scheduling T to Ar temp _j Setup the light path temp _ j Else exit End if End for </pre>
---	--

These are pseudo code algorithm. Extend algorithm to a data mining model with transformation entities.

The improve Apriori Algorithm.

Association rule mining is one of the most active research methods in data mining and Apriori algorithm is a kind of classic generated by the proposed project set the Boolean association rules mining algorithm of frequent item-sets. The Apriori algorithm is the discovery of association rules can be divided into two steps: first by iteration, find out the database to all support is not lower than the setting threshold item-sets, namely the frequent item-sets, and then use retrieved from frequent item-sets to structure can satisfy the rules of minimum confidence.

Map/Reduce is a huge amounts of data can be distributed programming model of parallel computing. Map/Reduce programming model to the operation of the data are divided into two steps: the Map (Map) and Reduce (specifications), Map operation is the input form, such as (key, value) of the specified key/value pair for operation, to generate a set of intermediate key/value pair. Reduce operation is on the middle of the Map operation output key/value pair statute consolidated

The key problem: the security

Safety is the main problem in the development of Internet technology. Safety performance is more complex, and can divide into many levels. Current development of Internet technology, the security problem of customer service is the most has the interoperability. Through the analysis of existing instances and summarizing, the author thinks that main security issues can be according to the structure of the Internet of things, divided into the following several aspects to solve one by one.

The problems of Network transport layer security. the main function of network transmission layer is to ensure that proper data transmission effectively, so the transport layer security problems

have to be attached great importance to, and are not malicious capture or modify the data and information.

Conclusions

Data mining is an important part of the Internet of things application, aiming at the characteristics of Internet data, this paper proposes a model of the Internet of things data mining based on cloud computing. Among them, the data mining platform is the key to the entire model. Priority is to design a variety of Internet oriented parallel data mining algorithm, the various data mining tasks to complete the Internet of things the Internet of data mining model based on cloud computing architecture is divided into five layers, respectively is the Internet of things data access layer, data integration, data mining platform layer, business layer and interactive control.

References

- [1] M.Conti, S.K.Das, C. Bisdikian, M. Kumar, L.M. Ni, A. Passarella, G. Roussos, G. Tröster, G. Tsudik, F. Zambonelli, Looking ahead in pervasive computing: challenges and opportunities in the era of cyber-physical convergence, *Pervez. Mob. Comput.* 8 (1) (2012) 2–21.
- [2] K. Gama, L. Touseau, D. Donsez, Combining heterogeneous service technologies for building an internet of things middleware, *Comp.Commun.* 35 (4) (2012) 405–417.
- [3] Arnon Rosenthal, Peter Mork. Cloud computing: A new business paradigm for biomedical information sharing [J]. *Future Generation Computer Systems*, 2010, 26(7): 947- 970.
- [4] Zhikui Chen, Haozhe Wang. A Context- Aware Routing Protocol on Internet of Things Based on Sea computing Model [J].*Journal of Computers*, 2012, 7 (1): 96- 105.
- [5] Bin Liu, Shu- Gui Cao. Machine Learning and Cybernetics [C]. International Conference on Machine Learning and Cybernetics. Guilin, China, 2011. 40- 44.
- [6] HongYuehua, Xu Shuang, Wu Huajian. Studyon distributed data mining model in Wireless Sensor Networks [C]. International Conference on Intelligent Computing and Integrated Systems. Guilin, China, 2010. 866- 869.
- [7] M.Lovbjerg, T.Krink. Extending particle swarm op'mizxrs with self-organized critically. *Proceeding IEEE Conference on Evolutionary Computation*, 2002: 1588-1593.
- [8] A.Ratnaweera, S.K.Halgamuge, H.C.Watson. Self-Organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Trans.Evol.Comput.*, 2004, 8(3): 240—255.