Machine Learning Algorithm-Based Minimisation of Network Traffic in Mobile Cloud Computing

Praveena Akki and V. Vijayarajan

Abstract Mobile cloud computing is an emerging technology where mobile device is integrated with cloud computing. It has many applications such as social network, online shopping, Flickr, Picasa. Besides these applications, it suffers from network traffic issue. The demand of the users has been increasing day by day but due to the limited density on base stations, it has become an overhead to the network service providers to provide the service. In this paper, we have applied machine learning techniques on the preprocessed data to classify client requests and generated rules to accept or to discard a client request. We aimed to minimize network traffic. We have applied J48, Naïve Bayes, Multi-Boosting AB, Simple Logistic Regression, Random Forest. It is observed that Random Forest has highest accuracy rate of 86.36% compared with other algorithms.

Keywords Network traffic · Machine learning techniques · Preprocessing

1 Introduction

Mobile handsets are the biggest inventions of the early 1990s. They have changed human life. Users have migrated to portable computers such as smart phones from desktop computers due to the advancement in the technology. These advancements have made the users to access online services from anywhere at any time. Cloud computing is an emerging technology which provides different services like applications, infrastructure, and platform to the users on demand. In the traditional mobile computing, mobile devices had limited computing and storage capabilities. Applications were built and managed by the developers. But users prefer to access the services

Praveena Akki · V. Vijayarajan (✉)
School of Computer Science and Engineering,
Vellore Institute of Technology,Vellore 632014, Tamil Nadu, India
e-mail: vijayarajan.v@vit.ac.in

Praveena Akki
e-mail: praveena.akki2016@vitstudent.ac.in

A. J. Kulkarni et al. (eds.), *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 828, https://doi.org/10.1007/978-981-13-1610-4_58

at anytime from anywhere, store the data, share the data with their friends. To meet the requirements of the users, new technology called mobile cloud computing was introduced, cloud computing in combination with mobile devices over the network.

1.1 Features of Mobile Cloud Computing

The description of features of mobile cloud computing is shown in Fig. 1 Resource Management: Mobile clouds can enable resource provisioning and demand provisioning automatically. The resources are network resources, computing resources, mobile device resources.
Security and Privacy: This feature includes security capabilities, technologies, process. This feature protects the data, devices, and network from unauthorized access and damage.
Scalability: This feature relies on cloud, network, and mobile scalability.
Convenience: This feature is designed for end users to provide access to the cloud resources at any time and from anywhere.
Connectivity: This feature chooses different well-designed APIs, protocols, standards to provide secure and easy connection between different networks and third-party applications.

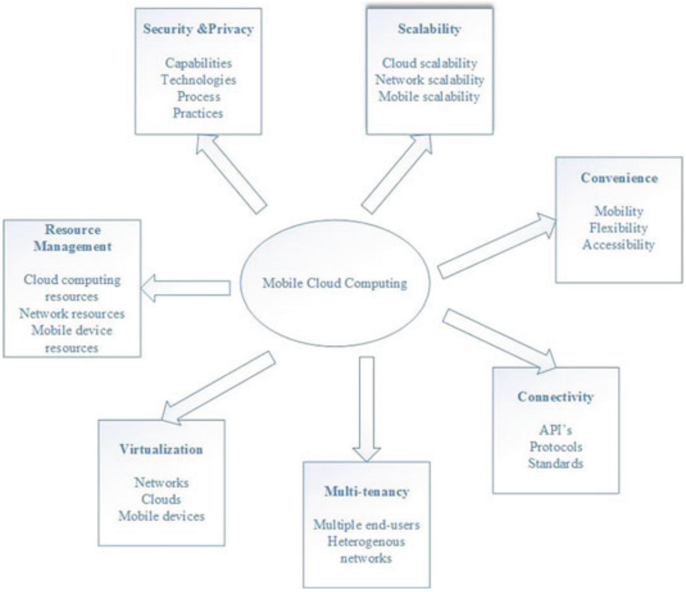

Fig. 1 Features of mobile cloud computing

Virtualization: Mainly three types of virtualization are used in mobile cloud computing namely network virtualization, cloud virtualization, mobile devices virtualization.

2 Related Work

In [2] soft air, a new SDN architecture for 5G has been proposed to overcome the limitations of software-defined networking (SDN) and network function virtualization (NFV). The software-defined network has been introduced for data center networks helps to deploy new apps and services. NFV abstracts network functionalities and implements them in software. The different aspects in wireless cellular network have been analyzed quantitively. To overcome limitations, soft air architecture was proposed. The scalability was improved by using high-performance controllers and network traffic was optimized. In [3], the resources will be dynamically allocated to the users by network virtualization. Network virtualization provides more efficient resource usage. The virtualization enables reliability and overcome the ossification of the Internet. How to assign the incoming service request to the server has become a challenge in mobile cloud computing. Migrating the service request can provide better utilization of resources; but it involves operational cost. In this paper, online migration algorithms have been used and the costs were compared without migration. In [1], various network centric parameters were analyzed such as traffic load, mobility speed, transfer type, energy consumed on migration, total execution time, end-to-end packed delay, packet delivery ratio. It was concluded that the transfer time, energy consumed on migration, and execution time increase with the traffic load.

By optimizing running states' size and applications, the network transfer time can be reduced. In [8], Rate Adaptive Topology-Aware Heuristic (RA-TAH) algorithm, an optimization algorithm, was proposed to minimize energy consumption. The network will be automatically adjusted according to the utilization. The performance of the algorithm is maximized when the utilization is high. A utilization threshold is defined as 0.7. If the load crossed threshold then another switch will be turned on, so the load will be balanced. In [12], a range-free centroid localization algorithm was proposed. Malguki algorithm was used to find location of an unknown node in the network. But its suffers from localization error. In traditional Malguki algorithm, the node is selected randomly, but in the proposed algorithm the unknown node is computed based on centroid of anchor points. The algorithm is simple, and the computed node is like the real node. This method is useful where GPS does not work to estimate nodes position. In [11], greedy algorithm was used for finding optimal service area. Greedy algorithm selects a tracking area and assigns to service area. This process will be repeated until a traffic load of service does not exceed service gateway capacity.

Limitation of greedy algorithm is due to its inefficiency since it does not consider the cost. To overcome the limitation, repeated greedy algorithm was proposed. It minimizes the number of Service Gateways (S-GW) relocations. The advantage is it can be launched periodically when there is change in traffic load. It can scale up or down the virtual service gateways. In [6], a dynamic programming is used to select the efficient communications between mobile devices and cloud servers. The large problems are divided into smaller sub-problems to solve whole problem and the best solution is selected. The cost function of each cloudlet is computed and assigns the devices to the cloudlet having minimum energy cost. In [4], the current research activities and further research challenges were presented. MANET paradigm and protocols were discussed. Mainly four adhoc networking paradigms, mesh, opportunistic, vehicular, and sensor networks, were discussed. The effect of increasing smartphones in everyday life on network traffic was discussed. In [7], a new mobile cloud computing framework, mobicloud was proposed which is used to enhance communication by addressing secure routing, risk management, trust management issues in the network. The development of new applications with enhanced processing power is the advantage of mobicloud. In [9], intelligent access schemes and heterogenous access management schemes were proposed inorder to improve quality of mobile connectivity and service availability to the users when required. In [5], an optimization approach, VM planner, was proposed to minimize the network power cost without affecting network performance. It also controls traffic flow routing by turning off as many unnecessary network elements to minimize power consumption. In [10], offloading strategies were proposed to minimize power consumption in femtocells. Femtocell is a small base station which is used to expand mobile network capacity. But the power consumption increases as the number of mobile users increases which leads to poor quality of service. The proposed offloading strategies have minimized energy consumption from 12.27–15.46%.

3 Proposed Architecture

The needs of mobile users have been growing exponentially, for example, Facebook, twitter, Picasa, eBay, amazon. Facebook, twitter have been using to share activity of users. The Picasa, Flickr are used for sharing photos. The eBay, amazon is used for online shopping. In the past era, people used to purchase items from shops in person. Due to the advancement of mobile cloud computing technology, people are preferring online transactions now. This has increased the demand for mobile data capacity. Mobile network operators carry large amount of internet traffic. It was observed that more than half of the mobile data traffic constitutes video streaming, and the remaining is images and text. Due to the limitation on the density of mobile base stations the network traffic has become a bottleneck. Many techniques were proposed to reduce the network traffic.

Prefetch is the technique where the proxy server retrieves the data from cloud server in advance based on the previous users log data. But not all the data retrieved from the server will be used by the user. Some of it will be discarded. Clients interact with server through API. The data returned by backend service should be understandable form. Mostly XML is used for its simplicity and interoperability, but it suffers with higher encoding overheads and repetitive nature of start–end tags, numerical values, binary data encoded in tags as text messages. This contributes to the increased size of messages across the network. Techniques have been proposed to eliminate the data redundancy but detecting the unnecessary tags and eliminating them increases additional overhead. Mobile clients can access cloud storage services and from everywhere. The proposed method aimed to reduce the network traffic. It predicts the client requests based on machine learning techniques. The data was collected from access log data, IRCache. Then the data was preprocessed to eliminate missing values and irrelevant information. The data was filtered by selecting frequently accessed values based on IP address and URLs. The data was analyzed by applying machine learning algorithms. The algorithm with good accuracy was selected to apply in mobile cloud computing environment. The machine learning algorithm predicts whether to accept the client request or to discard the request (Fig. 2).

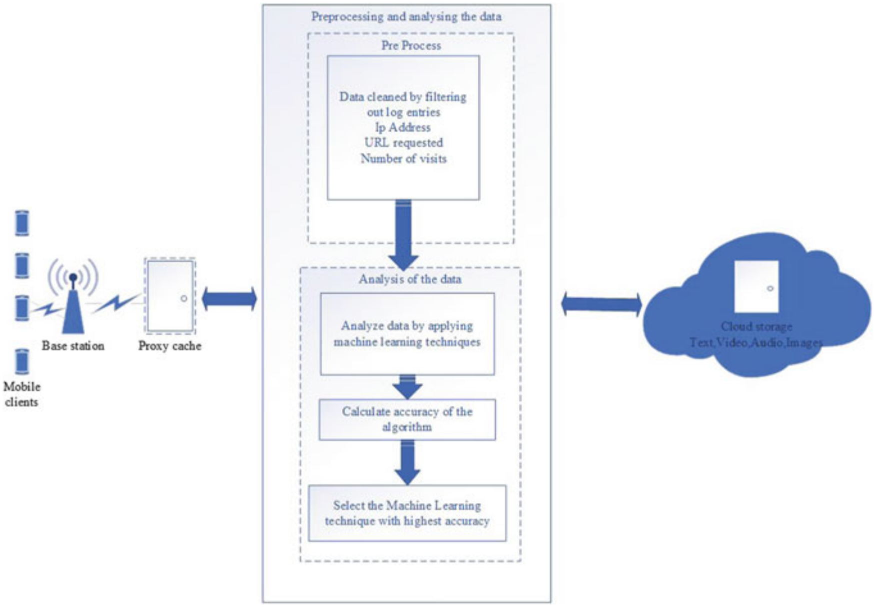

Fig. 2 Proposed architecture

3.1 Data Collection

Access log data set was taken in the proposed scheme from a proxy CC. Squid was used to collect log data based on cloud computing log data.The various fields in the client requests are
1157689312.049500610.105.21.199TCP_HIT/2001357
GET http://google.com/image.gif

1. Time stamp of the request
2. Time required to process the request in msec
3. IP address of the mobile client
4. Cache Hit/Miss
5. Requested data size in bytes
6. Type of method
7. Requested URL
8. Content type

3.2 Preprocessing

It is important to preprocess the data before analyzing. Preprocessing is used to clean and normalize the data to improve the quality of data. In the proposed method, mainly three attributes were taken into concern. IP address of the client, number of visits, and URL requested. Different IP addresses represent different users, and frequent requests were filtered out by analyzing the three attributes (Tables 1 and 2).

1. Remove entries with request methods except GET and POST.
2. Remove the records which have status code other than 200.
3. Remove uncatchable requests which have missing fields.

3.3 Preprocessing Access Log Dataset

Total number of records before preprocessing: 10000
Total number of records after preprocessing: 7635

Table 1 Before preprocessing

Time stamp	Elapsed time	IP address	Log tag and HTTP code	Request method	URL
1157689320	2864	10.105.21.199	TCP_MISS/304	GET	http://www.goonernews.com/
1157689359	1982	10.105.33.214	TCP_MISS/200	POST	http://shttp.msg.yahoo.com/notify/
1157689379	4	10.105.33.214	TCP_IMS_HIT/304	GET	http://a1568.g.akamai.net/7/1568/1600/20040405222807/radio.launch.yahoo.com/radio/common_radio/resources/images/t.gif
1157689388	60	10.105.21.199	TCP_HIT/200	GET	http://us.i1.yimg.com/us.yimg.com/i/us/pim/dclient/d/img/liam_ball_1.gif
1157689388	60	10.105.21.199	TCP_HIT/200	GET	http://us.i1.yimg.com/us.yimg.com/i/us/pim/dclient/d/img/liam_ball_1.gif
1157689409	15888	10.105.37.180	TCP_MISS/200	POST	http://gateway.messenger.hotmail.com/gateway/gateway.dll?
1157689437	752	10.105.37.180	TCP_REFRESH_HIT/200	GET	http://eur.i1.yimg.com/eur.yimg.com/i/fr/hp/sunsh.jpg
1157689446	429	10.105.37.184	TCP_HIT/200	GET	http://eur.i1.yimg.com/java.europe.yahoo.com/eu/hp/fu/jsbase003.js

Table 2 After preprocessing

Time stamp	Elapsed time	IP address	Log tag and HTTP code	Request method	URL
1157689359	1982	10.105.33.214	TCP_MISS/200	POST	http://shttp.msg.yahoo.com/notify/
1157689379	4	10.105.33.214	TCP_IMS_HIT/304	GET	http://a1568.g.akamai.net/7/1568/1600/20040405222807/radio.launch.yahoo.com/radio/common_radio/resources/images/t.gif
1157689388	60	10.105.21.199	TCP_HIT/200	GET	http://us.i1.yimg.com/us.yimg.com/i/us/pim/dclient/d/img/liam_ball_1.gif
1157689409	15888	10.105.37.180	TCP_MISS/200	POST	http://gateway.messenger.hotmail.com/gateway/gateway.dll?
1157689446	429	10.105.37.184	TCP_HIT/200	GET	http://eur.i1.yimg.com/java.europe.yahoo.com/eu/hp/fu/jsbase003.js

3.4 Algorithm

```
Result: Preprocessed data
Pc=client IP address;
 R=URL requests;
while i = 0 to total do
    count=1;
    if count==limit then
        delete the record;
        if log_tag! =200 then
            delete the record;
            if Request method! = GET || Request method ! = POST then
                delete the record;
            else
                Save the record in the log file and repeat the same for all the
                 records
            end
        else
            Save the record in the log file
        end
    end
    repeat the same for all the records;
end
```

Algorithm 1: Algorithm for Data preprocessing

3.5 Machine Learning Algorithms

Machine learning algorithms predict outcomes without being explicitly programmed. The algorithm receives input data and use some statistical analysis to predict the output within an acceptable range. After preprocessing the data, machine learning algorithms were applied. These algorithms are mainly classified into supervised and unsupervised. In the proposed method, J48, Random Forest, Naive Bayes, Multi-Boost AB, Simple Logistic Regression were applied to analyze the data and compared them with one another to identify highest accuracy.

$$\text{Random Forest Prediction } s = \frac{1}{K}\sum_{k-1}^{k} K\text{th tree response}$$

Using decision tree, it will generate rules, but the drawback is it is greedy. This problem can be overcome by using Random Forest. The main advantage of using Random Forest is the overfitting problem which can be eliminated and it is applicable to regression. It is also capable of handling missing values and categorical values. The accuracy of Random Forest is high because the number of trees in the forest is high.

4 Simulation Results

The input access log data was collected from real cloud. The file is converted to arff or csv format. The simulation was done by WEKA tool. Different IP address represent different users. The number of cross-validation iterations is set with 10, which are the number of folds. The experiment type of performance was set as classification. The performance evaluation is based on classifier evaluation which is based on classifier evaluation metrics. The accuracy is measured as,

$$\text{Accuracy} = \frac{\text{Number of Correct Data}}{\text{Total Data}} * 100 \tag{1}$$

Precision is the models ability to discern whether the data is relevant from the returned population and recall is the models ability to select relevant documents from the population. F-measure is the harmonic mean of precision and recall. These performance metrics have been calculated based on Eqs. 1–5, the results are shown in Tables 3 and 4 and the accuracy of machine learning algorithms are shown Fig. 3.

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Error Rate} = \frac{FP + FN}{TP + FP + FN + TN} \tag{4}$$

$$\text{F-Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

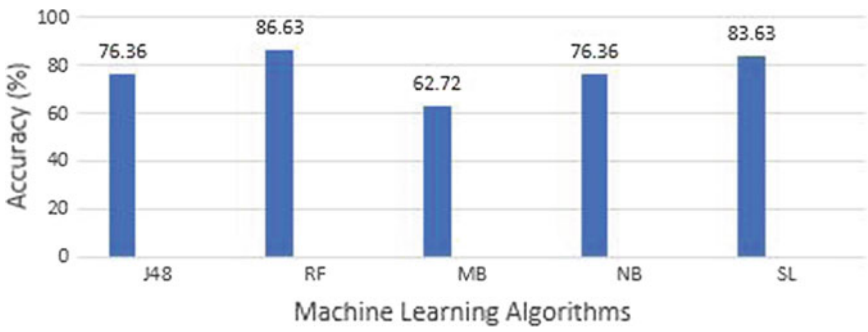

Fig. 3 Accuracy of machine learning algorithms

Table 3 Performance Metrics

Performance metrics/algorithms	J48	Random Forest	Multi-Boost AB	Naive Bayes	Simple Logistic
TP-rate	0.764	0.836	0.627	0.764	0.836
FP-rate	0.198	0.178	0.396	0.079	0.106
Precision	0.753	0.823	0.407	0.783	0.814
Recall	0.764	0.836	0.627	0.764	0.836
F-measure	0.741	0.816	0.491	0.764	0.822
ROC	0.869	0.968	0.639	0.871	0.913

Table 4 Accuracy of machine learning algorithms

Algorithm	Accuracy in %
J48	76.363
Random Forest	86.363
Multi-Boost AB	62.72
Naive Bayes	76.36
Simple Logistic	83.63

5 Conclusion

This paper presented an efficient machine learning technique for mobile cloud computing. The frequent requests from the same clients were identified and removed by applying preprocessing techniques. Then, the machine learning techniques were used to predict whether to accept the client request or to discard. Due to the limitation on the density of base stations, network traffic has become a bottleneck. By using the proposed method, the network traffic has been reduced therefore improving quality of service.

References

1. Ahmed E, Akhunzada A, Whaiduzzaman M, Gani A, Ab Hamid SH, Buyya R (2015) Network-centric performance analysis of runtime application migration in mobile cloud computing. Simul Modell Practice Theory 50:42–56
2. Akyildiz IF, Lin SC, Wang P (2015) Wireless software-defined networks (w-sdns) and network function virtualization (nfv) for 5g cellular systems: An overview and qualitative evaluation. Comput Networks 93:66–79
3. Arora D, Bienkowski M, Feldmann A, Schaffrath G, Schmid S (2011) Online strategies for intra and inter provider service migration in virtual networks. In: Proceedings of the 5th international conference on Principles, systems and applications of IP telecommunications, ACM 10 (2011)
4. Conti M, Boldrini C, Kanhere SS, Mingozzi E, Pagani E, Ruiz PM, Younis M (2015) From manet to people-centric networking: milestones and open research challenges. Comput Commun 71:1–21

5. Fang W, Liang X, Li S, Chiaraviglio L, Xiong N (2013) Vmplanner: optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers. Comput Networks 57(1):179–196
6. Gai K, Qiu M, Zhao H, Tao L, Zong Z (2016) Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. J Network Comput Appl 59:46–54
7. Huang D, Zhang X, Kang M, Luo J (2010) Mobicloud: building secure cloud framework for mobile computing and communication. In: Service Oriented System Engineering (SOSE). In: 2010 Fifth IEEE international symposium on, IEEE, pp 27–34 (2010)
8. Huu TN, Ngoc NP, Thu HT, Ngoc TT, Minh DN, Tai HN, Quynh TN, Hock D, Schwartz C et al (2013) Modeling and experimenting combined smart sleep and power scaling algorithms in energy-aware data center networks. Simul Modell Practice Theory 39:20–40
9. Klein, A, Mannweiler, C, Schneider, J, Schotten, HD (2010) Access schemes for mobile cloud computing. In: 2010 Eleventh International Conference on mobile data management (MDM), IEEE, pp 387–392 (2010)
10. Mukherjee A, Gupta P, De D (2014) Mobile cloud computing based energy efficient offloading strategies for femtocell network. In: Applications and innovations in mobile computing (AIMoC), 2014, IEEE, pp 28–35 (2014)
11. Taleb T, Ksentini A (2013) Gateway relocation avoidance-aware network function placement in carrier cloud. In: Proceedings of the 16th ACM international conference on modeling, analysis & simulation of wireless and mobile systems, ACM, pp 341–346 (2013)
12. Wang Y, Jin Q, Ma J (2013) Integration of range-based and range-free localization algorithms in wireless sensor networks for mobile clouds. In: Green computing and communications (GreenCom), 2013 IEEE and internet of things (iThings/CPSCom), IEEE international conference on and IEEE cyber, physical and social computing, IEEE pp 957–961 (2013)