Volume 3, Issue 3, March 2018 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

A Survey on Optimization of Network Configuration on Real Time Data Using Data Mining Approach

Anamika D. Patil
Department of Computer Engineering
Gokhale Education Society's
R. H. Sapat College of Engineering
Management Studies and Research,
Nashik-5

Chandrakant R. Barde
Department of Computer Engineering
Gokhale Education Society's
R. H. Sapat College of Engineering
Management Studies and Research,
Nashik-5

*Abstract-: **Conventional network optimization technology in actively change the network configurations in light of networks congestion proportion, decreases rate, protection holes and soon, main to sub-ultimate user studies. The data mining method is acquainted with expect the useful resource limit construct absolutely in light of totally of historic dimension facts. To walk around the dynamic distribution of user demand and application request, a weighted k-Nearest Neighbors (k-NN) model is proposed to predict periodic characteristics of network traffics by denoting different temporal and spatial patterns of radio resource margins. This paper discusses the topic based on past research paper and also studies the network optimization.***

Keywords: Data Mining, K-NN, Radio resource margin, Resource reconfiguration

I. Introduction

The major network features, considering the cellular network as an example, are developing dramatically, such as the increasing demands for high data rate service and densely distributed traffics. Hence, network operators are facing the great challenges on how to improve the network capacity and minimize the coverage hole by network configuration Optimization approaches in an efficient way [2].To be specific, in order to capture the dynamic characteristics of radio resource demands, a weighted k-NN model has been proposed based on a largescale historical data set from cellular operators networks. And the Mixed Genetic and Cross Validation algorithm are used to determine the most accurate input time series data and k value for the proposed k-NN model. Furthermore, the optimization algorithm is put forward to reconfigure radio resources over the whole network based on the predicted traffic load.

As we know that, traditional network optimization approaches passively adjust the network parameters which are based on network's traffic ratio, drop-off rate, coverage holes, are no longer applicable in the dynamic changing wireless network environment. Besides, due to the number of services and human activates, the radio resource shortage and service demand are distributed in different areas and change drastically over different time slots. That means part of the network may be overloaded due to the radio resource shortage. Today, with recent developments of data mining techniques, the statistical learning methods can be utilized to model the large data set of users and then analyze the structure as well as pattern of the data set. Due to the astounding abilities in analyzing the data from different perspectives to obtain the valuable network information, the data mining technology has been paid much attention for its potential applications in both fields like academia and industry.

II. Literature Survey

Advancement in technology and development of numerous data mining techniques, the statistical learning methods can be utilized to model the large data set of users, analyze the structure and pattern of the data set. Due to the astonishing abilities in analyzing the data from diverse perspectives to obtain the useful network information, the data mining technology has been paid much attention for its potential applications in both fields like academia and industry. The various existing researches have proposed data mining technologies for traffic estimation and network optimization.

In this paper [2], As understanding the characteristics of spectrum utilization is necessary in providing guidelines for resource allocation. So, in this paper author give, a detailed measurement analysis of spectrum efficiency with data collected from tens of thousands of base stations during fifteen months. Also examine the characteristics of radio resource margins (RM) extensively by including its temporal skewness, diurnal patterns, weekly periodicity and spatial skewness. Main findings include that radio resources are not utilized efficiently both temporally and spatially and radio RM and traffic load show strong weekly periodicity which is conventional. Inspired by the inefficient utilization of radio resources, author then devise an optimization method for dynamic radio resources reconfiguration and the final experimental results show and prove that it improves radio resources utilization efficiency and also traffic load balance significantly.

In this paper [3] author explain, At the present time, the explosively on the increase demands for mobile services which bring both challenges and opportunities to wireless networks that giving birth to 5G-fifth generation mobile networks. The different features and requirements of different services are assorted in 5G. The management and coordination among heterogeneous networks, applications, user demands need the 5G network to be open and flexible to ensure that network resources are allocated with high efficiency. For fulfillment of these requirements, wireless network virtualization is used to amalgamate heterogeneous wireless networks and coordinate network resources. So in this paper author proposed a model of wireless network virtualization which consist of three planes- the data plane, the cognitive plane and the control plane. Also, a novel control signaling scheme has been designed to support the proposed model. From the implementation view of network virtualization, a hierarchical control scheme based on cell-clustering has been used with dynamically optimized efficiency of resource utilization. In this paper, two use cases have been analyzed to demonstrate how the schemes work under the proposed model to improve resource efficiency and the user experience.Rapid growth of wireless data service is approaching against the boundary of communication network's processing power and so the pervasive and exponentially increasing data traffic present impending challenges to all the aspects of the wireless system design like spectrum efficiency, computing capabilities and front haul/back haul link capacity.

In this paper [4] author discuss the challenges and opportunities in the design of scalable wireless systems to grip such a "bigdata" era. Also, in this paper author gives review in two phases i.e. in first phase the state-of-the-art networking architectures and signal processing techniques adaptable for managing the bigdata traffic in wireless networks and in second phase instead of viewing mobile bigdata as an unwanted encumber, introduce methods to capitalize from the enormous data traffic, for building a bigdata-aware wireless network with better wireless service quality and also new mobile applications. Finally, also for wireless communications in the mobile bigdata era highlight several promising future research directions.

In this paper [5] author conduct a detailed measurement analysis to investigate the spatial characteristics of network resource usage using a large-scale data set collected `in situ' in a nationwide 3G cellular data network in which the data set spans over thousands of base stations. Also, characterize the spatial correlation in radio resource usage using different statistical techniques. This analysis highlights existence of significant spatial correlation that varies during the day, middle of the day peak and at middle of night waning. Also, there is use the notions of spectral clustering to show how base stations can be clustered and how correlated they are in conditions of radio resource usage. It shows spatially connected clusters are produced and a few clusters exist when clustered optimally. Finally, with the concept of Granger causality, understand the underlying functional connectivity and flow of influence in the network. This show that generally one-third of neighboring base station pairs exhibit statistically significant Granger causality and long causal paths exist in the network. The result of this observation can lead to enlargement of new techniques for network monitoring and resource management in future cellular data networks.Most presented studies of spectrum usage have been performed by actively sensing the energy levels in specific RF bands including cellular bands. In this paper, author provides a unique, complementary analysis of cellular primary usage by analyzing a dataset collected inside a cellular network operator. One of the primary aspects of this dataset is its scale which consists of data collected over three weeks at hundreds of base stations. Then dissect this data along assorted dimensions to characterize and model primary usage as well as understand its temporal and spatial variations. This analysis shows several outcomes that are relevant if dynamic spectrum access (DSA) approaches are to be deployed for cellular frequency bands. For some instance, it finds that call durations show significant deviations from the often used exponential distribution, which makes call-based modeling more problematical. he also demonstrate that a random walk process, which does not use call durations, can often be used for modeling the aggregate cell capacity. Finally highlight some applications of desired results to develop secondary usage of licensed spectrum [6].

This paper [7] give detailed measurement analysis of network resource usage and subscriber behavior using a large-scale data set collected inside a nationwide 3G cellular data network. Today all over the world a million subscribers, data set tracks close to over thousands of base stations. Also in this paper analyze individual subscriber behaviors and then observe a significant variation in network handling among subscribers. This paper characterizes subscriber mobility and temporal activity patterns and identifies their relation to traffic amount. From this author conclude how efficiently radio resources are used by different subscribers as well as in different applications. Also analyze the network traffic from the point of view of the base stations and find significant temporal and spatial variations in different parts of the network, while the aggregated behavior appears predictable. Finally, results show an important insight into network-wide resource usage with implications in pricing, protocol design with resource and spectrum management.

Today, Internet traffic dynamics in large cellular networks is an important task for network design, troubleshooting, performance evaluation and optimization. So in this paper [8] author discussed and present the results from review of study, which is based upon a week-long aggregated flow level mobile device traffic data collected from a major cellular operator's core network. Also In this study they measure and characterize the spatial and temporal dynamics of mobile Internet traffic. Then they distinguish this study from other related work by conducting the measurement at a larger scale and exploring mobile data traffic patterns along two new dimensions i.e. device types and applications which generate traffic patterns. So on the basis of findings of measurement analysis; author proposed a Zipf-like model to capture the volume distribution of application traffic and also a Markov model to capture the volume dynamics of aggregate Internet traffic. Then further customize these models for different device types using an unsupervised clustering algorithm for improvement in prediction accuracy.

This paper [9] proposes a load-balancing scheme for overlapping wireless LAN cells. Agents running in each access point broadcast periodically the local load level via the Ethernet backbone and determine whether the access point is overloaded, balanced or under-loaded by comparing it with the received reports. The load metric is the access point throughput and overloaded access points force the handoff of some stations to balance the load. Only the under-loaded access points accept the roaming stations in minimizing the number of handoffs. The author shows that experimental evaluation that our balancing scheme increases the total wireless network throughput and decreases the cell delay.

In this paper [10], author explains and analyzes the Erlang capacity of a code-division multiple-access system supporting voice and delay-tolerant data services. The novelty here is to consider the characteristic of delay-tolerant traffic in terms of a delay confidence. The delay confidence is the probability that a new data call is accepted within the maximum tolerable delay without being blocked. In this case, the Erlang capacity is confined not only by the required blocking probability of voice call but also by the required delay confidence of data call. For the performance analysis, develop a 2-D Markov model, which is based on the First-Come- First-Serve service (FCFS) discipline, and further present a numerical procedure to analyze the Erlang capacity. According to this procedure make a balance between the Erlang capacities with respect to the blocking probability of voice call and with respect to the delay confidence of data call, in order to accommodate extra Erlang capacity. Finally, demonstrate the balancing by properly selecting the size of the designated queue for data

- **Three contributions are summarized in the following**.

1. Temporal and Spatial Skewness Study: A novel framework is developed to make a full use of data collected from tens of thousands of base stations and the statistical learning methods are utilized to analyze temporal and spatial characteristics of radio resource margins. The cells lacking radio resources are distributed from 7 o'clock to 23 o'clock, instead of centering in one or two hours. And only less than 30% of the cells are over-loaded or short of radio resources, while over 70% of the cells are sufficient in radio resources. These results show that the inefficiency in resource utilizations, which inspires us to propose a network-wide radio resource reconfiguration framework to make a better use of radio resource margins.

2. Data Mining Based Dynamic Resource Reconfiguration Framework: By using data mining techniques, the pattern and characteristic beneath the data can be captured and a dynamic and active radio resource reconfiguration framework is proposed in this paper. Thus, operators can find the problems before the network break-down and actively optimize the network to improve the quality of service experience. Results show that a higher resource utilization and a better load balance over the whole network can be achieved by using the proposed framework.

3. Clustering Cells to Reduce Time Complexity: The traffic load dynamics are similar in some cells which can be characterized by the same model but are different in other cells. Therefore, different models are used to characterize the dynamics of traffic load in different cells. The cells are first divided into several clusters to reduce the number of models and the time complexity.

III. PROPOSED SYSTEM

The major network features, considering the cellular network as an example, are developing dramatically, such as the increasing demands for high data rate service and densely distributed traffics. Hence, network operators are facing the great challenges on how to improve the network capacity and minimize the coverage hole by network configuration optimization approaches in an efficient way [2]. Traditional network optimization approaches, which passively adjust the network parameters based on network's congestion ratio, drop off rate, coverage holes and etc., are no longer applicable in the dynamic changing wireless network environment. Besides, due to the variety of services and human activates, the radio resource shortage and service demand are distributed unevenly in different areas and fluctuate drastically over different time periods.

Earlier network optimization technologies passively adjust the network configurations based on network's congestion ratio, drop-off rate, coverage holes etc. So to optimize the network configurations by obtaining the accurate network status, user demand and application request distribution based on the real time data.

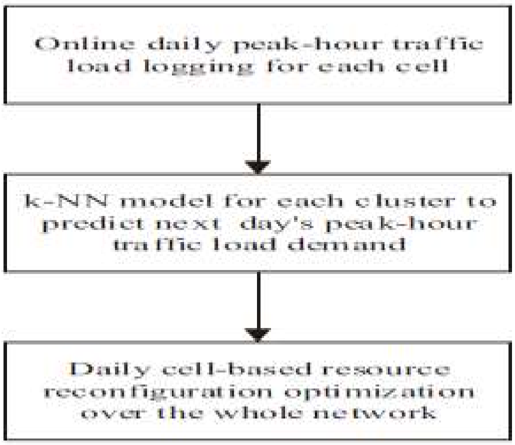

Fig.1.System Architecture

To construct a superior utilize of radio asset constrains transiently and spatially in excess of complete network, the resource reconfiguration and the elements of the traffic load in every cell are very important as well as necessary, the dynamic resource re-configuration frame has two major parts as shown in Fig. 1. Initially, the everyday top hour traffic load is recorded. After that, the past records with day to day patterns are utilized to instruct the k-NN model which is able to forecast the traffic load for the after that day. Lastly, the day by day top hour traffic load is anticipated via utilizing the k-NN model, and the radio resource are reconfigured every day in excess of the entire network through utilizing the optimization algorithms to assemble resource difficulty in every cell. During performance the resource re-configuration operation in cellular networks is able to achieve remotely by utilizing the software product in [15].

CONCLUSION

This paper proposed to apply data mining techniques with regard to historical measurement statistics for the active network configurations. In this paper we cover Network Optimization Approaches for different attributes. So by considering different optimization model or algorithm, the data rate, traffic problem is reduced by reconfiguring radio resources over the whole network, is presented this review. Also these reviews illustrate how the dynamic wireless environments for communication overcome the statistical network configuration.

REFERENCE

1. Z. Y. Feng, J. Min, X. Yan, Y. Gao, Q. X. Zhang and Y. Zhang. Characterizing and Exploiting Temporal-Spatial Radio Resource Margins in Cellular Networks. in Proc. of IEEE VTC, 2014.
2. Z. Y. Feng, C. Qiu, Z. B. Feng, Z. Q. Wei, W. Li and P. Zhang. An effective approach to 5G: Wireless network virtualization. IEEE Communications Magazine, 53(12):53–59, Dec. 2015.
3. S. Bi, R. Zhang, Z. Ding and S. G. Cui. Wireless Communications in the Era of Big Data. IEEE Communications Magazine, 53(10):190-199, Oct. 2015.
4 M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. Characterizing geospatial dynamics of application usage in a 3G cellular data network. in Proc. of IEEE INFOCOM, 2012.
5. M. Z. Shafiq, L. Ji, A. X. Liu, S. V. J. Pang, and J. Wang. A first look at cellular network performance during crowded events. in Proc. of ACM SIGMETRICS, 2013.
6. U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding spatial relationships in resource usage in cellular data networks. in Proc. of IEEE INFOCOM, 2012.
7. D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary users in cellular networks: A large-scale measurement study. in Proc. of IEEE DySPAN, 2008.
8. U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding traffic dynamics in cellular data networks. in Proc. of IEEE INFOCOM, 2011.
9. M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang. Characterizing and modeling internet traffic dynamics of cellular devices. in Proc. of ACM SIGMETRICS, 2011.
10. H. Velayos, V. Aleo, and G. Karlsson. Load balancing in overlapping wireless lan cells. in Proc. of IEEE ICC, 2004.
11.Y. Bejerano, S. J. Han, and L. Li. Fairness and load balancing in wireless lans using association control. in Proc. of ACM MobiCom, 2004.
12. A. Pillekeit, F. Derakhshan, E. Jugl, and A. Mitschele-Thiel. Force-based load balancing in co-located umts/gsm networks. in Proc. of IEEE VTC, 2004.
13. H. Chen and D. D. Yao. Fundamentals of Queuing Networks: Performance, Asymptotics, and Optimization. Springer, 2001.
14. I. Koo, J. R. Yang, and K. Kim. Erlang Capacity Analysis of CDMA Systems Supporting Voice and Delay-Tolerant Data Services Under the Delay Constraint. IEEE Transcations on Vehicular Technology Magazine, 56(4):2375-2385, Jul. 2007.
15. H. Technologies. BSC6900 GSM Initial Configuration Guide(V 900R013C00 03). 2009.
16. C. M. Bishop. Neural networks for pattern recognition. Oxford University Press, 1995.
17. A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse. Methodology for long-term prediction of time series. Neurocomputing, 70(16):2861- 2869, October 2007.
18. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. in Proc. of IJCAI, 1995.
19. B. Efron and R. Tibshirani. An Introduction to the Bootstrap. Chapman & Hall/CRC, 1993.
20. D. E. Goldberg and J. H. Holland. Genetic algorithms and machine learning. Machine learning, 3(2):95-99, October 1988.
21. D. E. Goldberg. The Design of Innovation: Lessons from and for Competent Genetic Algorithms. Springer, 2002.
22. J. D. Schaffer, R. A. Caruana, L. J. Eshelman, and R. Das. A study of control parameters affecting online performance of genetic algorithms for function optimization. in Proc. of Genetic Algorithms, 1989.
23. T. Segaran. Programming collective intelligence: building smart web 2.0 applications. O'Reilly Media, 2007.
24. D. Whitley. A genetic algorithm tutorial. Statistics and Computing, 4(2):65-85, June 1994.
25.S. Hartmann. A competitive genetic algorithm for resource-constrained project scheduling. Naval Research Logistics (NRL), 45(7):733-750, October 1998.
26. J. MacQueen. Some methods for classification and analysis of multivariate observations. in Fifth Berkeley Symposium on Math Statistics and Probability, 1967.