Machine Learning Techniques in Network Optimization Problems

Bartosz Rodziewicz
nr indeksu: 226105
Politechnika Wrocławska

Adam Hyjek
nr indeksu: 234987
Politechnika Wrocławska

***Abstract*—The abstract goes here.**

I. Introduction

In recent years we can observe increasing amount of new devices connected to various networks. This trend is expected to increase further with 5G release and IoT concept gaining popularity. This causes increased network traffic leading to exhausting networks' capacity. To prevent that, there are numerous techniques of network optimization. Some of the most popular methods of network optimization discussed in the article are edge caching, real-time routing, load balancing, traffic shaping and data compression.

Edge caching is a single-sided storing of popular, recently viewed content allowing to reduce number of subsequent requests to external network. It allows not only to reduce external traffic, but also decreases response time significantly. Biggest problem in edge caching is predicting future demand for data and choosing what content already existing in cache should be replaced by incoming one. Most popular approaches are two simple algorithms: Least recently used(LRU) and Least frequently used(LFU)[1][2].

Real-time routing is a process where data is forwarded to it's destination based on current network condition, allowing route path to stay relevant even despite constant changes in network condition. There is a big variety in algorithms analyzing current network state such as RIP or EIGRP, but performing optimal routing based only on current network situation is a complex task. Routing results can be improved though by taking into account previous experiences to avoid some recurring problems[3][4].

Load balancing is a process of distributing network traffic across multiple servers to avoid overloading particular server leading to deterioration of service quality or even complete service degradation for some users. Typical algorithms used in load balancing problems are Round Robin algorithms and solutions based on current state's metrics,but just as in routing problem, making decision may be significantly improved by analyzing past circumstances and staying away from former issues[5][6].

Traffic shaping is an optimization technique which delays some of the packets to bring the overall traffic into the desired profile. It is used to optimize performance, increase usable bandwidth or improve latency for other applications and users. Since most of the network traffic is encrypted nowadays, the need to quickly and accurately classify Internet traffic for this purpose has been growing steadily[7]. This is a field where machine learning comes into play and provides a way to classify the encrypted network traffic.

Data compression is a subject that goes far beyond the network optimization problems. It is a process of encoding information using fewer bits than original representation. There are two types of data compression - lossy and lossless. The lossy compression reduce the amount of bits used to store information by removing less important information. The lossless compression reduces bits by identifying and eliminating statistical redundancy. There exists many different data compression algorithm, both lossy and lossless, and some of them use machine learning to increase their efficiency[8][9].

Machine learning is a subset of artificial intelligence. It is a study of algorithms that improve automatically over time. Machine learning algorithms use statistics to find patterns in massive amounts of data and make predictions based on that data. The more data the model has for training, the better the prediction are.

In this article we focus on machine learning usage in network optimization problems mentioned above. The article's goal is to point out increasing popularity of artificial intelligence methods applied to solving network optimization related issues and their success rate. The rest of this survey is organized into 5 sections dedicated to methods of network optimization and section containing our conclusion.

II. Edge Caching

III. Real-Time Routing

IV. Load Balancing

V. Traffic Shaping

VI. Data Compression

VII. Conlusions

Acknowledgment

References

[1] Z. Chang, L. Lei, Z. Zhou, S. Mao and T. Ristaniemi, "Learn to Cache: Machine Learning for Network Edge Caching in the Big Data Era," in IEEE Wireless Communications, vol. 25, no. 3, pp. 28-35, JUNE 2018, doi: 10.1109/MWC.2018.1700317.

[2] H. Zhu, Y. Cao, W. Wang, T. Jiang and S. Jin, "Deep Reinforcement Learning for Mobile Edge Caching: Review, New Features, and Open Issues," in IEEE Network, vol. 32, no. 6, pp. 50-57, November/December 2018, doi: 10.1109/MNET.2018.1800109.

[3] F. Tang et al., "On Removing Routing Protocol from Future Wireless Networks: A Real-time Deep Learning Approach for Intelligent Traffic Control," in IEEE Wireless Communications, vol. 25, no. 1, pp. 154-160, February 2018, doi: 10.1109/MWC.2017.1700244.

[4] S. Troia et al., "Machine-Learning-Assisted Routing in SDN-Based Optical Networks," 2018 European Conference on Optical Communication (ECOC), Rome, 2018, pp. 1-3, doi: 10.1109/ECOC.2018.8535437.

[5] R. A. Haidri, C. P. Katti and P. C. Saxena, "A load balancing strategy for Cloud Computing environment," 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014), Ajmer, 2014, pp. 636-641, doi: 10.1109/ICSPCT.2014.6884914.

[6] Anna Victoria C R Oikawa, Vinicius Freitas, Márcio Castro, Laércio Lima Pilla. Adaptive Load Balancing based on Machine Learning for Iterative Parallel Applications. 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Mar 2020, Västerås, Sweden. ff10.1109/PDP50117.2020.00021ff. ffhal-02570549f

[7] T. S. Tabatabaei, F. Karray and M. Kamel, "Early internet traffic recognition based on machine learning methods," 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, 2012, pp. 1-5, doi: 10.1109/CCECE.2012.6335034.

[8] F. H. Kingma, P. Abbeel and J. Ho, "Bit-Swap: Recursive Bits-Back Coding for Lossless Compression with Hierarchical Latent Variables" 2019.

[9] J. Park, H. Park and Y. Choi, "Data compression and prediction using machine learning for industrial IoT," 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 818-820, doi: 10.1109/ICOIN.2018.8343232.