REINFORCEMENT LEARNING IN DATA CENTER MANAGEMENT

ENVIRONMENTAL CONSTRUCTION OF BIG DATA INITIATIVES

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Cameron S. Dorsch

August 5, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed Cameron S. Dorsch Date: 8/5/2020

Approved: Catherine D. Baritaud Date: August 9, 2020
Catherine D. Baritaud, Department of Engineering and Society

Approved: Patricia C. Click Date: August 8, 2020
Patricia C. Click, Department of Engineering and Society

In the digital age, trillions of bytes are collected, processed, and stored every day to fuel the operations of businesses and organizations. Passive and active data collection transformed the global economy and continue to guide innovation and optimization in the modern world. Financial expert J. Manyika (2011) predicted early in the boom of big data that a retailer using big data to track customer preferences and system inefficiencies "could increase its operating margin by more than 60," (p. 2) in a report on the transformations coming as a result of Big Data systems. Manyika saw potential benefits to businesses from Big Data practices in four areas: the improvement of organizational systems, the customization of services, the automation of manual systems, and increasing transparency (Manyika, 2011, p. 5). Today, businesses and organizations of every type use data collection on ever increasing scales to increase their margins and optimize their operations. In the wake of its successes and growing ubiquity, Big Data is now also praised as the new modern tool for improvement in tracking sustainability and improving wasteful practices. In 2014, a UN Independent Expert Advisory Group (IEAG) released a new environmental report centered on the importance of Big Data in tracking and improving sustainability on a global scale (United Nations IEAG, 2014, p. 2). Big Data is at the forefront of the modern push for sustainability and improvement, however, its own implications to sustainability must be scrutinized if it is to be the path forward.

AREAS FOR RESEARCH AND IMPROVEMENT

In applications such as these, Big Data is often seen as an instrument of the "common good", since it is used as a tool for the betterment and optimization of existing systems (Lucivero, 2019, p. 1). One of the key recommendations of the aforementioned UN report is to "Share [Big Data] technology and innovations for the common good" (United Nations IEAG, 2014, p. 3). Data initiatives are often presented to communities and users as the cutting-edge modern solution to improving services, but in the case of sustainability, it is very rarely discussed how Big Data initiatives threaten sustainable practices themselves. Lucivero, a senior researcher in ethics and data, finds it especially worrying how Big Data is often thought of as a "clean oil" or "limitless resource". The hardware systems that supported the data revolution have improved in efficiency and carbon footprint, but power use of data centers still accounts for three percent of annual electricity sales in the U.S., or about 70 billion kWh, equivalent to the annual electricity use of over 6 million households (Shehabi et al., 2018, p. 4). For scale, the United States Environmental Protection Agency's (U.S. EPA) data tool for carbon equivalency estimates this power use to be equivalent to 49.5 million metric tons of CO_2. This is the CO_2 output of approximately 54 billion pounds of coal, or the carbon offset of 800,000 tree saplings grown for 10 years (U.S. EPA, 2019). According to the taxonomy discussed by Pohl et al. (2018), direct, first order environmental effects of the hardware that supports Big Data are not the only environmental risks with data systems either. Big Data initiatives can also have high order environmental effects, such as the effects of the applications and uses of Big Data (Pohl et al., 2018, p. 699). Even if Big Data initiatives are praised for tracking and increasing sustainability, their own growth and continued success inherently means more hardware, data centers, and power usage.

Big Data systems should not be seen as a clean, unlimited resource, but one that requires ongoing improvement in limiting wasteful practices and improving efficiency. With Big Data established as a tool for modern success, the volume of data we collect and store is only set to increase in the near future. Since the onset of the internet, researchers have improved the efficiency and sustainability of data storage systems consistently. These technical improvements, along changes to the private data center market, have stifled the increase of data center power usage so far, but researchers such as Shehabi et al. (2018) worry that, “The growth of data center electricity use beyond 2020, however, is uncertain as the modeled trends indicate efficiency measures of the past my not be enough for the data center demand of the future,” (p. 9). Continued improvement to Big Data systems is necessary if data collection and demand continues to grow exponentially (Shehabi et al., 2018, p. 8). These systems must be improved technically and be regulated in order to reduce the increase in their power consumption that comes with an increase in demand for data services.

To address these needs, a state-of-the-art technical paper will examine the potential of reinforcement learning systems in reducing the power usage of data center systems. This technical paper will examine both benefits and drawbacks, and study how accessible these technologies may be to the cheaper, less funded colocation data center markets. Additionally, an STS advocacy paper will apply the Social Construction of Technology to examine the absence of awareness of the environmental risks of Big Data (Bijker, Bönig, & Oost, 1984). This paper will examine how the scale of Big Data systems affects the ability of individuals to impart environmental views on to its construction, as well as social and regulatory paths to creating a balanced perception of Big Data.

REINFORCEMENT LEARNING SYSTEMS IN DATA CENTER MANAGEMENT

Constantly changing workloads and environment experienced by data center systems require dynamic management of cooling systems to optimize both power usage and performance. Systems must balance the need to provide fast, reliable data access with the desire to limit power consumption, which ultimately determines the profit margin of data services. Traditional task scheduling (TS) algorithms lack the scalability and adaptability needed to further improve these management systems, and because of this, machine learning, specifically reinforcement learning (RL) models, have been of recent focus in research.

RL MODELS AND ADAPTABILITY

Reinforcement learning models are machine learning models which instead of minimizing error with respect to some true value, maximize reward received from acting in some environment. The model receives an input from the environment, produces an action back to the environment, and gets some amount of reward in return representing how good the action was. Over many iterations, the model can maximize reward with respect to its action choice, and therefore the model can be incentivized to solve any problem in which a reward function can be well defined. One such model created by Silver et al. (2018) can play Chess, Shogi, and Go at superhuman levels of skill, having learned only by simulating play against itself. RL has been recognized for its ability to be applied to such a variety of problems and produce consistent results.

In research to develop more efficient resource management systems, RL models are praised for the adaptability to the range of use patterns experienced by data centers. In research by Cheng, Li, and Nazarian (2018) to construct an RL management system, data use patterns which occur over the span of a day as well as patterns that occur over months mean that "adaptability and self-learning capacity of the energy and electric cost reduction method are required" (p. 129) for improving management systems (Cheng, Li, & Nazarian, 2018). These systems respond to use patterns that traditional models could not have adapted to, and also perform better when applied to data centers with thousands of servers. Rolik et al. (2018), researchers who built a similar model, stated that RL models are especially good at "tasks with a compromise between long-term and short-term penalties" (p. 238), such as using more power vs.

breaking a service level agreement (SLA) in a resource management system. Overall, RL models represent a potentially lucrative technology in solving problems of compromise or optimization in which there is no known ideal solution.

THE STATE OF RL AND BIG DATA

A state-of-the-art report written during STS 4600 will synthesize current and relevant research on the application of RL models to data center resource management. This research will investigate in what ways RL models have been tested in both simulated environments and real data centers to gauge their viability in improving the energy efficiency of data centers. Models such as those built by Cheng, Li, & Nazarian and Rolik et al. will be compared and contrasted to show what type of systems show the most potential for success, and also what the shortcomings of RL resource management systems may be. The results of RL research will be used to extrapolate to the decoupling of energy use from data demand increase, to gauge whether or not RL systems could be the needed improvement to safely limit data center electricity use in coming years. This work will be completed and presented in an academic report by December 2020.

ENVIRONMENTAL CONCERNS IN THE CONSTRUCTION OF BIG DATA

Much of the time Big Data initiatives are pitched to the public, emphasis is placed on their advantage as the new frontier for innovation. However, very rarely is there any environmentally conscious pitch of Big Data systems. As Lucivero (2019) points out, there are even eco-friendly biases in the metaphor of the naming of the digital “cloud” and cloud computing, the practice of decentralized storing and running of processes over many servers. The average perception of Big Data systems lacks proper environmental context that accounts for both how systems may aid in tracking sustainability as well as the environmental risks of an increased dependence on Big Data.

RISKS

The overall environmental impact of Big Data systems is relatively unknown and challenging to measure. High order effects of the deployment of data initiatives are hard to predict and even more challenging to measure. Environmental researchers Pohl et al. (2019) found that life cycle analysis (LCA), though typically more product focused, can be productive in measuring higher order effects of technology replacement and optimization. However, they remark that "not only were user-related ICT effects less frequently integrated into the assessment, but many behavioral effects, … were not taken into account," concluding that current metrics of the total environmental impact of Information and Communication Technologies (ICT) are relatively uncertain. As shown in Figure 1 below, a multitude of high order effects can increase the environmental impact of a technology. Some of these include rebound effects, in which the reduction in scarcity of a resource or commodity increases consumption, or substitution effects, the positive or negative effect of replacing another

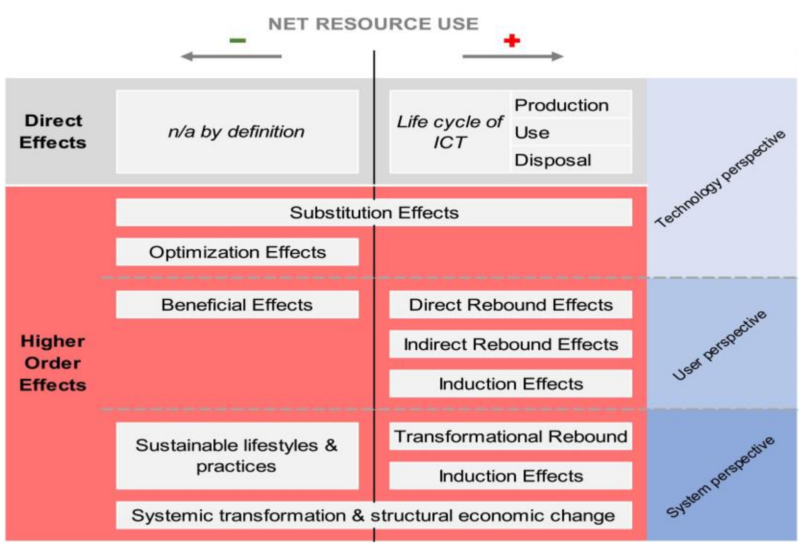

Figure 1: The High Order Effects of ICT. A depiction of environmentally friendly and dangerous effects of ICT separated by order and perspective. (Pohl, 2019).

technology (Pohl et al., 2019, p. 4). Lucivero (2019) also raises concern for the lack of real measurement methods of the overall effects of ICT technologies on the environment, concluding that more research must be done to establish common measurement practices. However, if our current understanding of Big Data's impact on the environment is so vague, how are initiatives always pitched as being the newest eco-friendly project?

THE STATE OF THE ENVIRONMENT AND BIG DATA

In practice, data storage systems only present more environmental risk than most digital technologies when deployed at massive scale. Due to this, individuals are unlikely to impart environmental concerns on to Big Data initiatives, and much of the construction and public perception is created by large technology companies and organizations. Oftentimes these companies present Big Data along with pushes for sustainability because of the benefits in tracking and analysis it provides. However, an environmentally balanced view is unlikely to be presented when the success of data initiatives is at risk, creating a one-sided perception of Big Data as a whole. My research will focus on how this relationship has developed, and how individuals can affect the construction of data systems.

An advocacy paper to be completed in STS 4600 will present research on how environmental concerns are excluded from the social construction of Big Data initiatives. The Social Construction of Technology (SCOT) seeks to identify how social and cultural values are imbued on an artefact through its design, deployment and use. Every decision made in SCOT is considered a social, moral, political, and cultural decision to expose how a technology has been shaped by the people around it (Bijker, Bönig, & Oost, 1984). This paper will analyze the systems in which Big Data initiatives are deployed to determine who holds environmental values, and how those values do not end up imparted on the final object. As shown below in

Figure 2, this analysis attempts to identify gatekeepers to the construction of Big Data that block environmental values from being imparted onto the technology. These gatekeepers will be examined in order to identify social paths, such as community participation in determining data collection procedures, through which individuals and consumers can impart environmental values on data systems. In addition to this, the paper will examine how large groups who control design, deployment, and advertising can be encouraged to impart environmental values onto the technology, as oftentimes their constructions are focused only on benefits to sustainability. The goal of this research will be to determine how Big Data can be more accurately represented and created to encompass a responsible view of its relationship to sustainability.

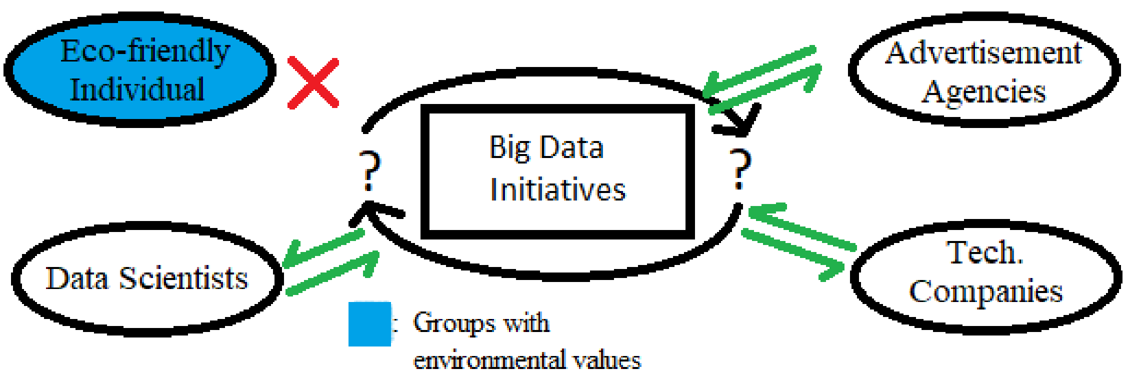

Figure 2: Social Construction of Big Data. Depiction of the lack of environmental values imparted on Big Data initiatives and the unknown gatekeeper. (Created by Dorsch, 2020)

A SUSTAINABLE DIGITAL WORLD

This research will examine both the technical opportunities to improve data center efficiency using RL as well as why and how environmental concerns regarding Big Data are absent from its deployment and construction. The exponential growth of data demand represents a serious threat to a sustainable future without further improvement to data center power use. We must adopt a more reasonable view of Big Data systems and continue improving them in order to create a sustainable world where Big Data is used to optimize, innovate, and improve systems.

WORKS CITED

Bijker, W., Bönig, J., Oost, E. (1984). The social construction of technological artefacts. *Journal for Science Research:* 3, 39-51. Retrieved from https://cris.maastrichtuniversity.nl/en/publications/the-social-construction-of-technological-artefacts-problems-and-p

Cheng, M., Li, J., Nazarian, S. (2018). DRL-Cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers. *2018 23rd Asia And South Pacific Design Automation Conference (ASP-DAC)*, 129-134. doi:10.1109/ASPDAC.2018.8297294

Dabbagh, M., Hamdaoui, B., Rayes, A. (2020). Peak power shaving for reduced electricity costs in cloud data centers: Opportunities and challenges. *IEEE Network*, 34(3), 148-153. doi:10.1109/MNET.001.1900329

Ding, Z., Tian, Y., Tang, M., Li, Y., Wang, Y.G., Zhou, C. (2019). Profile-guided three-phase virtual resource management for energy efficiency of data centers. *IEEE Transactions on Industrial Electronics*, 67(3), 2460 - 2468. doi:10.1109/TIE.2019.2902786

Islam, M.A., Mahmud, H., Ren; S., Wang, X. (2020). A carbon-aware incentive mechanism for greening colocation data centers. *IEEE Transactions on Cloud Computing*, 8(1), 4-16. doi:10.1109/TCC.2017.2767043

LaBrie, R.C., Steinke, G.H., Li, X., Cazier, J.A. (2018). Big data analytics sentiment: US-China reaction to data collection by business and government. *Technological Forecasting and Social Change*, 130, 45-55. doi:10.1016/j.techfore.2017.06.029

Lucivero, F. (2019). Big data, big waste? A reflection on the environmental sustainability of big data initiatives. *Science & Engineering Ethics*, *26*(2), 1009-1030. Retrieved from https://link-springer-com.proxy01.its.virginia.edu/article/10.1007/s11948-019-00171-7

Manyika, J. (2011). *Big Data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute.

Pohl, J., Hilty, H.M., Finkbeiner, M. (2018). How LCA contributes to the environmental assessment of higher order effects of ICT application: A review of different approaches. *Journal of Cleaner Productions*, 219, 698-712. doi:10.1016/j.jclepro.2019.02.018.

Rolik, O., Zharikov, E., Koval, A., Telenyk, S. (2018). Dynamic management of data center resources using reinforcement learning. *14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*: 237-244. doi:10.1109/TCSET.2018.8336194

Shehabi, A., Smith, S.J., Masanet, E., Koomey, J. (2018). Data center growth in the United States: decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12), 1. Retrieved from https://iopscience-iop-org.proxy01.its.virginia.edu/article/10.1088/1748-9326/aaec9c#erlaaec9cs4

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science:* 362:6419, 1140-1144. doi:10.1126/science.aar6404

United Nations Independent Expert Advisory Group (IEAG) on a Data Revolution for Sustainable Development (2014). *A world that counts: mobilising the data revolution for* *sustainable development.* United Nations. Retrieved from https://www.undatarevolution.org/report/

United States Environmental Protection Agency. (2019). *Greenhouse Gas Equivalencies Calculator* (20) [Data Set]. EPA. Retrieved from https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator