B6030326 - Callum Simpson -
Project Proposal

Optimisation of energy consumption polices

Context: With the increased use of ICT systems there has been an increase in CO2 emissions. The amount produced by ICT is roughly the same amount produced by global air travel [7]. High CO2 levels will have disastrous impacts on the planet and this number is only going to increase as more ICT systems get added. Although there has been a big swing in looking for ways of creating sustainable clean energy there still is the large issue of energy wastage. According to a 2018 report by Department of Energy the USA commercial section had a wastage energy level of 35% [8].

Problem: In ICT systems energy is wasted when a device is left on and not in use. To combat this most devices, have an energy consumption policy put in place that helps by switching the device to an ideal energy-saving state when not in use. However, most polices are programmed for static use and don't account for device high and low usage peaks meaning they will be un-optimal at certain times. Printers are an example of one these devices and when printing is 10-16% of ICT related energy consumption in higher education [9], the small amount of wasted per device starts to add up.

Proposed approach: In my project I will be investigating how to create an optimal energy policy that helps reduce energy wastage on a set of printers. This is in hopes of finding ways to create optimal energy consumption policies that could be used to help reduce energy wastage in other devices, saving energy and money. Though there has been work looking into the ways we can optimise energy policies not many of them use data analysis to help them, that is where my dissertation shall be different. After researching current policies that are in use, I will be simulating them using a High-throughput computing (HTC) system against an analysed set of data to see which one is the most optimal. I will also modify policies and try to create new policies using reinforcement machine learning techniques.

The device I will be developing the optimal energy policy is the Konica Bizhub C368 printers [10].

Aim: Investigate how the use of data analyses can be used to create an optimal energy consumption policy that balances energy-performance trade-off.

The most optimal value will be the one that finds the best balance between low power consumption and causes a limited amount of delays to user (aka finding a policy that has the perfect amount of uptime and downtime). Finding this balance is important as if a user wants to use the printer, but the printer is in a sleep state then energy will be used moving it to an on state. If a user keeps needing to turn a printer on, then it would damage the printer in the long run and could potentially wastage more energy than leaving it on.

Objectives	
1: Research and summarise state of the art energy policies	To create an optimal energy policy, I must get an understanding of policies that are currently out there and how they work. I will be conducting some background reading to research a wide range of different policies that are in use. I am aiming to get around 6 base policies but if I find more then I will use more.
2: Perform data analyses to find trends in printer usage	I am to receive a set of anonymised data about the Konica bizhub C368 printers (i.e print times and usage). I am planning on using said data to help me get an understanding of realistic printer usage. Using statistical analysis techniques (or other techniques if appropriate), I will be looking for trends and peaks in the printers to create a map of printer usage.
3: Implement and simulate different policies	I have been granted accesses to an inhouse High-throughput computing simulation [5] that has been set up to evaluate different energy reduction policies across an HTC system. Using the printer model that I have developed in the second objective I will run different policies against the model. I will use the simulation to get a set of results to showing how optimal each policy is.

4: Modify policies to discover potential avenues for optimisation	Once I have simulated the polices that I discovered in objective one, I will use the data gotten in objective 3 to see which ones got the best results. Then I will then modify those polices and simulate them to refine them further. I will keep doing this to see how far I can refine them.
5: Analyse and compare different policies to find the optimal energy-performance trade-off	Once I have collected all my results that I have gotten throughout my investigation I will compare them to each other to find out which one is the most optimal. Again, the optimal value will be the one that finds the best balance between low power consumption and causes limited delays. My findings here will allow me to discuss in my report the changes we can make to current energy policies so that they become more optimal.

Background:

[1] Optimization of power consumption and device availability based on point process modelling of the request sequence
Summary: The aim of this paper is to find the optimal choice of the waiting policy that a device should respect before entering sleep mode to optimize power consumption. It also investigates using a Hidden Markov chain. It clearly explains many different types of policies, how they were used and how effective they are.
Relevant: This a research paper that investigates the different policies that can affect energy optimization. It clearly explains the different energy policies which helps my 1st objective. The conclusion of this paper dose lower power consumption but increases the amount of time the device would need to be turned on by 3-fold, which would damage the printer in the long run
Key Words: Hidden Markov chain modelling, Optimal time out, Power management, Renewal

[2] PRINTER TIME-OUT Patent No – US 8,230,248
Summary: A patent on a system and method for creating computing time-out for a device, like printers. It also that discusses a lot about the issues that a person can face when trying to optimise an energy policy based around a printer. It also discusses what type of things we should be looking for when optimising a printer policy.
Relevant: The patent has the some of the same authors as the 1st background resources and so share a lot of the same information. However, the patent discusses more about the different things that I would have to consider when creating/modifying energy policies. As a patent it also includes a step by step description of the method it used to create its optimisation policy. This will help me with me with my fourth objective. The patent also explains the different modes a printer could enter when trying to save energy and how they are linked together.
Key Words: Optimal time out, Power management, Statistical models for request processes

[3] Probabilistic Latent Clustering of Device Usage
Summary: A Paper talking about "an application of Probabilistic Latent Semantics to the problem of device usage analysis in an infrastructure in which multiple users have access to a shared pool of devices delivering different kinds of service and service levels." It discusses how to build a probabilistic latent class and how to group individual data into usage patterns.
Relevant: This document covers how data analysis can be used in a scenario that is like the area I will be looking into (a network of printers). In the document it discusses a model that was made to help create a model of printing behaviours. It also goes in-depth on how the model was made. The model created in this report is like what I want to create in objective 2 so I will be using this to help my model of realistic printer usage.
Key Words: Probabilistic Latent Clustering, Network device, Data analysis

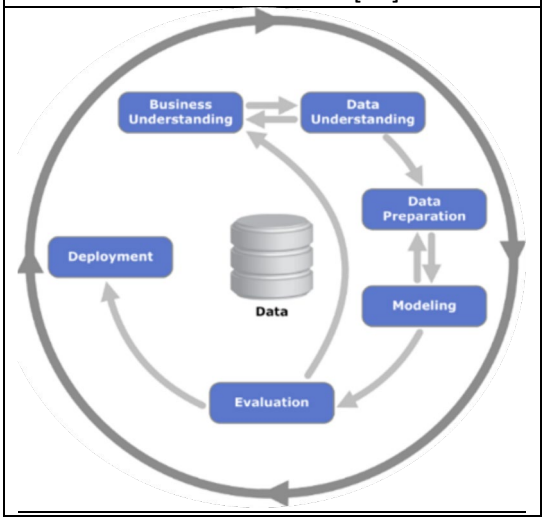
[4] Policy Optimization for Dynamic Power Management
Summary: A Paper talking about "Dynamic power management schemes (also called policies) and how they reduce the power consumption of complex electronic systems by trading off performance for power in a controlled fashion, taking system workload into account." It discusses an energy policy tool that uses an advanced LP solver. It also discusses how the tool being used to optimise a wide range of low-end devices.
Relevant: It includes a mathematical framework for the formulation and solution of the policy optimization problem based on the stochastic model of power-managed devices and workloads. What is good about this solution is that the tool runs and creates policy for a range of devices which is the big picture of what I am aiming for. The methods in this paper will help me create a method of creating more optimal energy policies that may a bit more generic (not tied to printers).
Key Words: Energy conservation, energy management, optimization methods

[5] HTC-Sim: a trace-driven simulation framework for energy consumption in high-throughput computing systems
Summary: A paper discussing a High Throughput Computing (HTC) simulation that allows the evaluation of different energy reduction policies.
Relevant: This the paper made about the HTC that I will be using in my 3rd objective to simulate policies. As I will be using it for a big part of my dissertation, I wanted to gain a understanding of how it works.
Key Words: Simulation, Energy, High-Throughput Computing

[6] A Bayesian Framework for Reinforcement Learning
Summary: A paper discussing a framework that can be used for reinforcement learning.
Relevant: As I going to be using reinforcement machine learning to help create an optimal policy, I decided that I am going to need to get a understanding of how one works. This recourse explains how one works and provided the framework to start creating one.
Key Words: Reinforcement Learning, Bayesian

Work plan

CRISP-DM model [11]



My plan is to complete my dissertation by following the CRISP-DM model. The Cross-industry standard process for data mining [12] is one of the most widely used analytical models that it used in industry and following it will help me create an optimal energy policy. This is because the CRISP-DM is a cycle model where the sequence of steps isn't rigid and what I do next will depend on what my exploratory data analysis tells me. The CRISP-DM will allow me to loop back and try different things. This is good for my project as each iteration will provide me with a new set of data that will allow me to do further tests to find the most optimal energy policy.

The Business understanding step focuses on the creation and understanding the of aims and objectives of the given project at hand. I have set it so this step is where I do all my research but that may change as I might do more research throughout my first few cycles. This will be working towards my 1st objective.

Then I will onto move Data understanding phase in which I will be analysing the data I have been given to identifying data quality problems (so I can try and get more data). If I find that I don't have

enough data, then this may be an issue as time and resources will have to be spent getting more. I will also try and detect interesting peaks in the information (i.e key printer usage times).

Data preparation will see me cleaning data and constructing the final dataset (data that will be fed into the modelling tool(s)) from the initial raw data. This is the step where the realistic printer model will be made. Completion of this step will mean my 2nd objective is fulfilled.

In the Modelling phase I will be modelling the different policies that I have come across. It possible that one of my policies has specific data requirements, if so, I will go back to the data preparation step to get the needed data. I will then test each of the policies in a simulation of realistic printer usage to see how they hold up. The simulation is new to me so it may take a bit of time for me to get a developed understanding of it. As it in house I can ask those who developed it for help and advice, this should help reduce the overall time needed to learn the system. My first cycle will be with base energy polices which will fulfil my 3rd objective. I would be also trying to implement some reinforcement learning in this step. Reinforcement learning is new to me so it may take me a while to get it to work, though I have done research on how it works which I hope will help me.

In the evaluation step, I will evaluate my findings. To fully evaluate it I will need to go back to the business understanding step to make sure my results fully match the requirements. If the results have not met the requirements or If I feel like I could refine them in some way, then I will loop through every step again to get a set of results that match meets my aim. Each loop through this will towards my 5th objective and the data I have achieved in this step will be used to help decided where I will go next, linking to my 4th objective.

For my dissertation the deployment step will be writing up my findings.

Future plans: Because I am using the reactive CRISP-DM model it's difficult to set what time I will be done with each step. This is because I react to the data, so my order of steps is always changing. However I have set myself some key deadline outside the CRISP-model (and alongside the other hard-set deadline set by the course) that I am planning to follow so that I can finish my deadline on time.

I am currently waiting for the data to be anonymised, this is so the data protection act [13] is met. Once that is done, I will start working towards completing objective 2 which will allow me to start my simulations. Over the Christmas exam period I will mostly likely slow down the amount of work I do on my dissertation to focus on my exams.

At a minimum I would like to have at least one cycle done by the 1st of March. Though I would like to have a lot more cycles done by the time I am finished its likely that the first cycle will take a while as it involves setting up a lot of things (creation of the map of realistic printer usage and setting up the initial policies). I would at least like to have one done by that time so I have something to discuss in my poster and demo.

I want to complete the final cycle on the 7th of April. This is so I have enough time to write my finding which I am hoping to get done by the 18th of April (I will be also be writing my dissertation throughout the year).

What has been done so far: Background research around current energy saving policies has been done and I have collected info on basic and advanced policies, meeting my 1st objective. I have studied the maths behind them to get an understanding of how they work and to see how I can implement them in java which will help me in objective 3. I have investigated the ways that polices can be made and potential ways they could be modified. I have also looked over the documentation for the HTC to get a better understanding of the simulation and how I am going to use it, helping me further prepare for objective 3.

References:

[1] - Durand, J., Girard, S., Ciriza, V. and Donini, L. (2013). Optimization of power consumption and device availability based on point process modelling of the request sequence. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 62, No. 2 , pp. 151-165 [Accessed 21 Oct. 2019].

[2] - Dance, C., Ciriza, V. and Donini, L. (2012). Printer Time out. US 8,230,248 B2. [online] https://patentimages.storage.googleapis.com/0e/0b/a1/68ff4f0f49072c/US8230248.pdf [Accessed 26 Oct. 2019].

[3] - Andreoli, J. and Bouchard, G. (2005). Probabilistic Latent Clustering of Device Usage. [online] Citeseerx.ist.psu.edu. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.9423&rep=rep1&type=pdf [Accessed 5 Nov. 2019].

[4] - Benini, L., Bogliolo, A., Paleologo, G. and De Micheli, G. (1999). Policy Optimization for Dynamic Power Management. IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, VOL. 18, NO. 6 [Accessed 5 Nov. 2019].

[5] - Forshaw, M., McGough, A. and Thomas, N. (2015). HTC-Sim: a trace-driven simulation framework for energy consumption in high-throughput computing systems. CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE, pp 1-32 [Accessed 10 Nov. 2019].

[6] - Strens, M. (2000). A Bayesian Framework for Reinforcement Learning. Stanford Universit: Proceeedings of the Seventeenth International Conference on Machine Learning. [Accessed 13 Nov. 2019].

[7] - Fettweis, G. and Zimmermann, E. (2008). ICT ENERGY CONSUMPTION – TRENDS AND CHALLENGES. The 11th International Symposium on Wireless Personal Multimedia Communications [Accessed 27 Nov. 2019].

[8] – Lawrence Livermore National Laboratory Energy Flow Charts (2018). [online] Flowcharts.llnl.gov [online] Available at: https://flowcharts.llnl.gov/commodities/energy [Accessed 24 Nov. 2019].

[9] - James, P. and Hopkinson, L. (2009). Sustainable ICT in Further and Higher Education. [online] Ictliteracy.info. Available at: http://www.ictliteracy.info/rf.pdf/rptgreenictv1.pdf [Accessed 21 Oct. 2019].

[10] - Konicaminolta.eu. (n.d.). bizhub C368 Multifunctional Office Printer. [online] Available at: https://www.konicaminolta.eu/eu-en/hardware/discontinued-products/bizhub-c368 [Accessed 24 Nov. 2019].

[11] - Jensen, K. (2012). A diagram showing the relationship between the different phases of CRISP-DM and illustrates the recursive nature of a data mining project. [image] Available at: https://commons.wikimedia.org/w/index.php?curid=24930610 [Accessed 11 Nov. 2019].

[12] - Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Shearer, C. and Wirth, R. (2019). CRISP-DM 1.0. pp 13-15 [Accessed 11 Nov. 2019].

[13] - Legislation.gov.uk. (2019). Data Protection Act 2018. [online] Available at: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [Accessed 1 Dec. 2019].