Research Statement

Nipun Batra

February 25, 2018

Overall Interest

My research interest can be summarised as **sustainable computing at scale using machine learning**. Computational sustainability focuses on developing computational models, methods and tools to support policy makers in developing more effective policies for sustainable development related to biodiversity, climate, environment, urban design, transportation, buildings and others.[1] Example applications[2] include structural health monitoring for bridges, deploying sensor network for data-driven farming. To achieve scale for a computational sustainability task, one can develop and deploy a large number of cheap sensors. Instead, my work uses machine learning to achieve the same effect as a dense sensor deployment, using a smaller number of (existing) sensors.

The Billion Building Challenge

In this statement, I will be discussing my sustainable computing work in the context of buildings. I propose "the billion building challenge" - enabling smart building functionality for a billion buildings across the world. The "billion" number comes from the fact that we can ballpark the number of buildings in the world to be in that range. Buildings are complex cyber-physical systems (CPS) and have a profound impact on occupant health, productivity, comfort and energy consumption. Smart building technology is aimed at improving several aspects of building operation by collecting and analysing sensor data to carry out informed control. Despite the potential benefits, smart building technologies have not been widely adopted, since they require expensive instrumentation in each building. Prior work had mostly focused on using novel sensors and intelligence for fine tune control aimed at optimising building efficiency. However, we argued that such heavy and often expensive instrumentation can only be scaled to a small number of buildings limiting the potential impact and would not suffice for the billion building challenge.

The Energy Breakdown Problem

I now discuss the billion building challenge in the context of the problem of energy breakdown - breaking down the household total energy measured at a single point into constituent appliances. Studies indicate that energy breakdown feedback can help occupants make informed decisions about their energy consumption with a potential of saving up to 15% energy savings [1]. To put this number into context, the estimated savings by replacing all the incandescent bulbs with efficient LEDs is about 2%. Most prior works required hardware (high-frequency smart meters, among others) to be installed in each home. This hardware would cost more than 500$ per home, and would thus be prohibitively expensive to scale.

Approach: Collaborative Sensing

Our aim was to estimate the energy breakdown of a home based without installing any sensor in the home using techniques we call "collaborative sensing" - reconstructing the sensor data of one home based on sensor data collected in other homes. The main challenge for this approach is that every home is unique; different homes can have highly varying appliance energy consumption and no two homes will produce exactly the same energy breakdown. However, we hypothesised that this data is sparse, i.e. much of the variation in energy breakdown across homes occurs along a relatively small number of dimensions. Intuitively, this sparsity stems from the fact common design patterns create a common and repeating structure in homes. The goal of collaborative sensing is

[1] https://cs.stanford.edu/~ermon/cs325/

[2] http://www.compsust.net

to exploit this sparsity by creating low-dimensional models of homes based on trends in their monthly energy bills, square footage, climate zone, and other easily available information. This model can then be used to reconstruct a home's energy breakdown based on homes which already have energy breakdown.

All homes have household aggregate energy consumption data available in the form of monthly bills, and additional features like household area. In addition, some homes have energy breakdown data. We formulate our sparse learning method or learning a low subspace representation of appliance energy consumption as a matrix factorisation (MF) problem (rows corresponding to appliances and columns to homes) used commonly in the recommendation space.

We found that our MF based solution [2, 3] was not only more scalable than prior literature, but also more accurate than the state-of-the-art source separation techniques which leverage high frequency minutely data. Our MF based solution was able to automatically infer useful domain properties, such as the relation of air conditioning usage to physical phenomena such as degree days.

While we had shown that we can provide an accurate energy breakdown without installing sensors in a test home, our MF approach would not likely scale to a billion buildings. The fundamental challenge is that our MF approach still required homes with energy breakdown to be from the same region as a test home. To mitigate this limitation, we proposed a transfer learning approach where the goal would be to require minimal homes with an energy breakdown in a target region of interest by re-using the homes with energy breakdown from a source region.

Our main hypothesis was that energy consumption has attributes that vary across regions (like weather, types of homes, etc.) and once we have factored those out, we are left with region independent properties. Using this insight, in our transfer learning method [4] we only need to re-learn the region independent properties in a new test region. Our modelling approach was similar to the MF approach described above. However, we presented a custom tensor design suited to our transfer learning settings. Our evaluation showed that using our transfer learning approach; we can estimate the energy breakdown of Austin homes from San Diego homes and vice versa, more accurately than just using data from the test region of interest. Our transfer learning approach was especially useful when we used a small number of homes from the target region.

I believe our collaborative sensing work is an important first step towards the realisation of the billion buildings challenge. While the above statement focused on scaling energy breakdown to homes without sensors, and has been published at premier machine learning/data science conferences [2, 3, 4], I have also worked on several systems buildings and algorithms for providing the most value to homes where we sensors are available, which has been published at community-specific CPS conferences [5, 6, 7, 8]. Apart from scaling sustainable computing, I have worked on the themes of making sustainable computing comparable and utility-driven. I built an open source toolkit for benchmarking energy breakdown algorithms [5], which has garnered more than 150 citations and has been adopted as the standard in the community for comparable research. I carried out the first dense home deployment in the context of developing countries [6] and presented a system [8] and several insights specific to the Indian scenario. I presented algorithms for deriving actionable insights from high-resolution energy breakdown data [7], which instead of just a pie-chart energy breakdown, would provide feedback like - "your fridge is defrosting too much wasting 12% fridge energy". My work has received the best PhD presentation award at a premier sensor networks conference [9], and a best demo award. Besides academic impact, my work has been well received as evidenced by us being a finalist in a large nation wide challenge.

Future research

In the future, I plan to work on solve pertinent and impactful computational sustainability and social problems using machine learning and data science. My long term dream is to help towards an India where everyone has access to clean air, water, energy, health and education. Impact in these problems can only be achieved at large scale. I plan to build on my foundation in machine learning and cyber-physical systems to address these problems at scale. I now describe three examples of problems that I would like to work in the future.

Estimating dense city-scale air quality data without additional sensors

The air pollution levels in cities like New Delhi have reached an alarming level, having a significant impact on the health and productivity of the population. Air quality can vary significantly even within a city. However, current

levels of instrumentation are not adequate for providing high fidelity information. The air-quality estimation problem has a few similar attributes to the energy breakdown problem.
Inspired by my existing work on scalable energy breakdown, the machine learning challenge here boils down to estimating the air quality of new regions using air quality data from already instrumented areas. An important distinction from the energy breakdown problem is that air pollution is continuous spatially, and this additional insight brings forth newer challenges and opportunities. Gaussian processes would be suited for such class of problems as they have been shown to work well for related problems [10, 11].

Developing minimally intrusive wearables/sensors for personal health monitoring

With the advent of smart watches, human activity and healthcare tracking studies are on the rise. The key challenge in this line of research is to identify markers of interest using proxy sensing methods, while being constrained by sparsity of sensor data in space and time. The sparsity in time arises due to battery limitations, and the spatial sparsity arises due to impedance of human comfort on extensive instrumentation. I now discuss one such example application that I will work on.
Sitting is the new smoking [12]. Poor posture-related health problems are on a rise, leading to significant reduction in productivity and life quality [13, 14]. ICT presents opportunities for early detection and prevention of such problems [15]. There are at least two ways in which this problem can be attacked - sensing the human body via minimally intrusive wearables or sensing objects of interest such as chairs, shoes. In both the approaches, the technical problem boils down to finding the optimal placement of sensors for an accurate identification of the underlying health marker, given a limited budget on the number of sensors and their battery life.

Optimising the utility distribution networks

The distribution of resources such as water and electricity can be viewed as complex networks [16]. A high proportion of the resources gets wasted due to high transmission and end-usage losses. I plan to study the problem of dynamically optimising these networks when considering the similarities with the vast literature of survey in computer networks. Optimisation can be studied at both the node level (end-use) or the edge level (transmission network). The key challenge here is to deploy sensors offering maximum utility at the minimum sensing and monetary cost, while being an accurate proxy of the phenomenon under study. Applications would include detecting transmission losses (lossy network nodes) and maintaining quality control (for example - water potability across the network).
Significant optimisation can also be achieved at the node or end-use level. Studies indicate that a large proportion of population has a limited understanding of resource consumption [1]. Overflowing water storage tanks are a common sight. It is highly likely that the population at large is unaware of the grim future if these sustainability problems are not addressed. This direction of work would be towards development of a mix of crowdsensing and low-cost ICT technologies for large-scale end-use resource monitoring. The availability of such rich information can greatly aid in policy making. The key challenge in such a project is to keep the per-home cost to be low and to get active and helpful citizen participation.

References

[1] Sarah Darby. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 2006.

[2] Nipun Batra, Hongning Wang, Amarjeet Singh, and Kamin Whitehouse. Matrix factorisation for scalable energy breakdown. In *AAAI*, pages 4467–4473, 2017.

[3] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. Gemello: Creating a detailed energy breakdown from just the monthly electricity bill. In *SIGKDD 2016*, 2016.

[4] Nipun Batra, Yiling Jia, Hongning Wang, Kamin Whitehouse, et al. Transferring decomposed tensors for scalable energy breakdown across regions. *AAAI*, 2018.

[5] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems*, Cambridge, UK, 2014.

[6] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.

[7] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. If you measure it, can you improve it? exploring the value of energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.

[8] Pandarasamy Arjunan, Nipun Batra, Haksoo Choi, Amarjeet Singh, Pushpendra Singh, and Mani B Srivastava. Sensoract: a privacy and security aware federated middleware for building management. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 80–87. ACM, 2012.

[9] Nipun Batra. Non intrusive load monitoring: Systems, metrics and use cases. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 501–502. ACM, 2015.

[10] Deepak Vasisht, Zerina Kapetanovic, Jongho Won, Xinxin Jin, Ranveer Chandra, Sudipta N Sinha, Ashish Kapoor, Madhusudhan Sudarshan, and Sean Stratman. Farmbeats: An iot platform for data-driven agriculture. In *NSDI*, pages 515–529, 2017.

[11] Sahil Garg, Amarjeet Singh, and Fabio Ramos. Learning non-stationary space-time models for environmental monitoring. In *AAAI*, 2012.

[12] Peter T Katzmarzyk, Timothy S Church, Cora L Craig, and Claude Bouchard. Sitting time and mortality from all causes, cardiovascular disease, and cancer. *Medicine & Science in Sports & Exercise*, 41(5):998–1005, 2009.

[13] Marjorie Woollacott and Anne Shumway-Cook. Attention and the control of posture and gait: a review of an emerging area of research. *Gait & posture*, 16(1):1–14, 2002.

[14] Etienne Grandjean and W Hünting. Ergonomics of postureâĂŤreview of various problems of standing and sitting posture. *Applied ergonomics*, 8(3):135–140, 1977.

[15] Erin Griffiths, T Scott Saponas, and AJ Brush. Health chair: implicitly sensing heart and respiratory rate. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 661–671. ACM, 2014.

[16] Gollakota Kaushik, Abinaya Manimaran, Arunchandar Vasan, Venkatesh Sarangan, and Anand Sivasubramaniam. Cracks under pressure? burst prediction in water networks using dynamic metrics. In *AAAI*, pages 4694–4700, 2017.