Crowdsourcing Appliance Labels for Energy Disaggregation

Alex Shyr, Abhay Gupta, Vivek Garud
Bidgely Inc.
Sunnyvale, CA 94086
{alex, abhay, vivek}@bidgely.com

***Abstract*—This paper presents a novel and scalable solution to the prevalent problem of a lack of labels in the domain of energy disaggregation. The system detects consistently recurring appliance patterns, intelligently generates inquiries to the user and incorporates user responses in the disaggregation pipeline. We showcase a few system designs as well as live results obtained via our approach.**

I. Introduction

Energy disaggregation has received an increasing amount of attention in recent years. With the growing market adoption of smart meters and home-area network (HAN) devices, the availability of high-resolution consumption data is no longer the limiting factor in non-intrusive load monitoring (NILM) research. Rather, the amount of labeled and annotated datasets has lagged behind and is the biggest bottleneck in advancing the research.

Previously in both academic and industry settings, labeled datasets are collected by measuring plug-level loads in a few wired-up home. As the number of appliances in the home grows, collecting ground-truth labels become more laborious and expensive. This is prohibitively unscalable. The data is also static, as it does not adapt to changing user behavior or new appliances. It is also important to note that all ground truth collected using plug-level loads involves an inherent bias as these users are often quite energy savvy.

An alternative approach is to rely on the users for the labeled information. Services such as the Amazon mechanical turk and Google's 1800-GOOG-411 service are implemented with a similar intention. One such naïve system can pose a question to the user every time any appliance turns on (whenever a significant change in consumption level occurs). This unintelligent mechanism results in a myriad of unprioritized inquiries, as well as an undesirable user experience. Fundamentally, there is no notion of appliance pattern, and the above process is incapable of detecting a "session" of appliance usage.

In this paper, we propose and demonstrate a solution which poses intelligent questions to the user, and robustly robustly incorporates the user input into the disaggregation pipeline. Without the aid of prior knowledge or previously available user information, the system can adapt to each user's consumption patterns and gradually discover the existing appliances. It also opens up ground truth collection from a userbase that is as yet untapped.

II. System Design

A. Problem Definition

The scenario of interest is one in which the data is flowing with sufficiently high granularity and relatively low delay. The desired system should accomplish the following:

- Generate inquiries for consistently recurring appliances
- Generate inquiries for appliances with clean runs, as to not cause user confusion
- Ingest and start detecting the appliances

On top of these technical requirements, the final user experience will also need to be non-intrusive, non-tedious and intuitive. Section II.C will be devoted to addresses these issues.

The energy consumption data stream of a typical day in Europe is shown in Fig. 1, along with the annotated sessions for a few common household appliances.

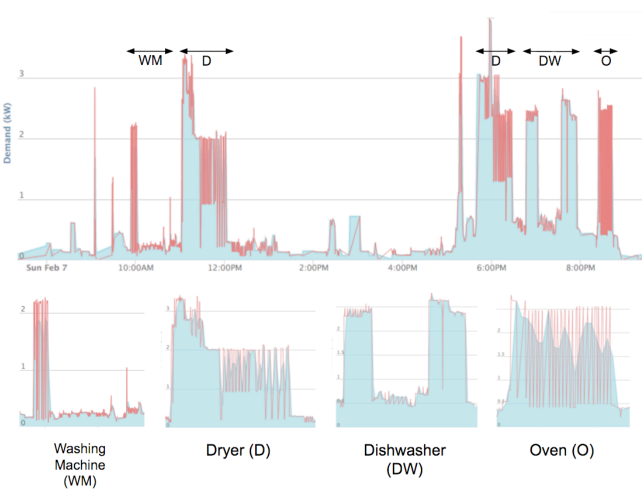

Fig. 1. (Above) Typical daily energy consumption stream, with annotated sessions of common household appliances. (Below) Zoomed-in appliance patterns

As seen in the appliance patterns, most appliances are in the similar amplitude range, between 2kW and 2.5kW. A simple transient-based system which raises an inquiry to the user when a significant surge in power is observed will have difficulty figuring out the appliance sessions. The resulting user experience will have an overwhelming number of inquiries, as well as significant confusion when soliciting confirmations.

The situation is further complicated in cases of overlapping appliance usage, as shown in Figure 2.

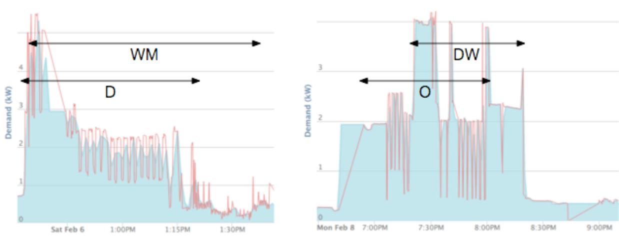

Fig. 2. Examples of concurrent appliance usage, resulting in overlapping patterns. (Left) Dryer and Washing Machine (Right) Oven and Dishwasher

The desired system should determine whether each appliance session is clean and not run concurrently with other appliances. The overall system design is shown in Figure 3.

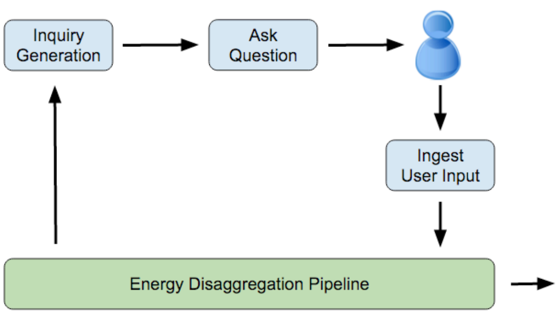

Fig. 3. Overall system implementation to crowdsource appliance labels

B. Metrics

In order to evaluate the performance of the end-to-end system, we measure how quickly it adapts to a home without any prior knowledge, and how accurate the final appliance detections are. Specifically, for each appliance, we measure

- Number of instances until first inquiry generation
- Number of instances until first appliance detection, after the user has responded to a few inquiries and the input have been ingested
- F1 score (a combination of the recall and precision metrics) of appliance detection

C. Asking Intelligent Questions

The user experience is paramount to the crowdsourcing process, and the questions posed to the users need to be carefully crafted. Here are some of the issues that need to be addressed:

- *Limited memory*. Users often forget which appliances they started, let alone when. The inquiries therefore must be timely.
- *Inevitable mistakes*. The system must robustly deal with unavoidable erroneous input or unsure answers that the users provide. [3]
- *Intuitive context*. Most users remember their usage events in context, rather than in absolute terms (e.g. "running dryer after washing machine", or "turning on heater in after waking up").

A few mobile app designs are shown in Figure 4. The users are prompted to respond to the questions with a limited set of options.

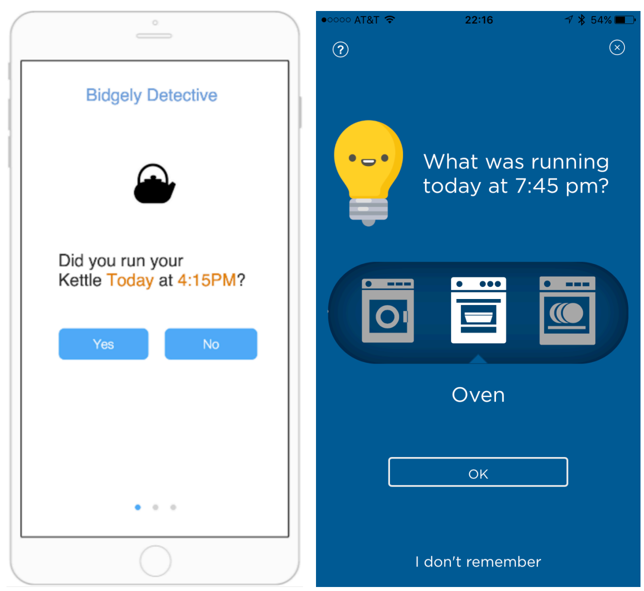

Fig. 4. Some mobile designs for inquiry generation

III. Experiment and Results

To backtest our design, we included 60 homes with one month of data and existing ground-truth labels, many having multiple appliances. User input to the generated inquiries is simulated using the labels (assuming complete correctness). The results are shown in Table I.

TABLE I. SIMULATION RESULTS

# Instances until Inquiry	# Instances until Detection	Detection F1 Score
3.4 ± 1.2	6.3 ± 2.6	0.77 ± 0.16

We further presented the solution to live users on the Bidgely platform and we received a significant amount of ground-truth labels. The wide variety of patterns for the common household appliance list which have been confirmed by the user is shown in Figures 5-8.

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

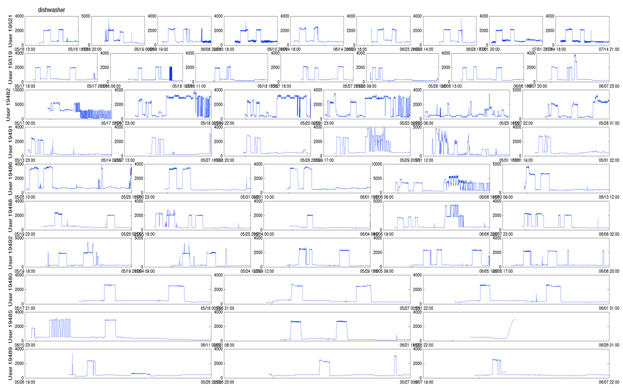

Fig. 5. Variety of dishwashers in the United Kingdom

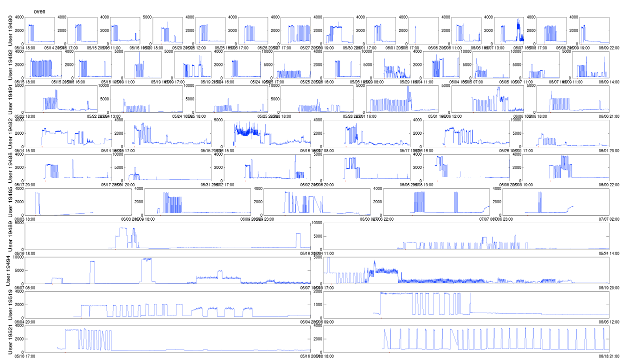

Fig. 6. Variety of ovens in the Netherlands

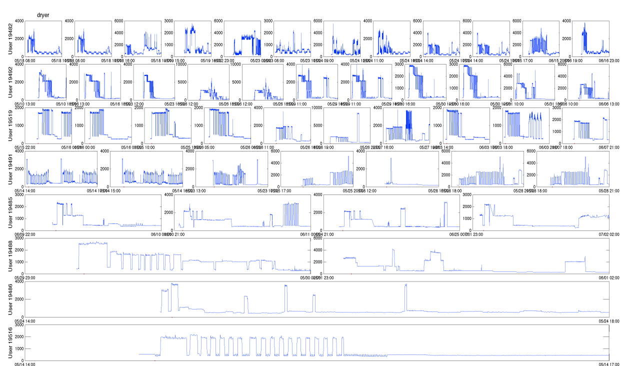

Fig. 7. Variety of dryers in Germany

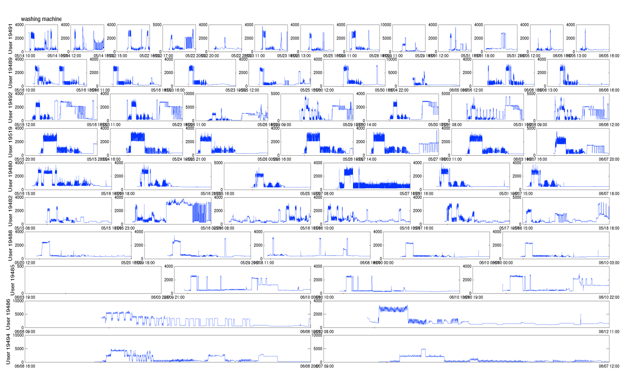

Fig. 8. Variety of washing machines in Germany

To demonstrate our solution's viability to crowdsource labels for appliances beyond the original list of {washing machine, dryer, dishwasher, oven}, patterns for two geography-specific appliances are shown in Figures 9-10.

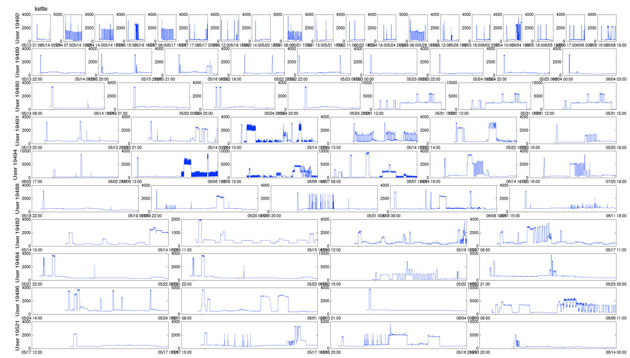

Fig. 9. Variety of kettles in the UK

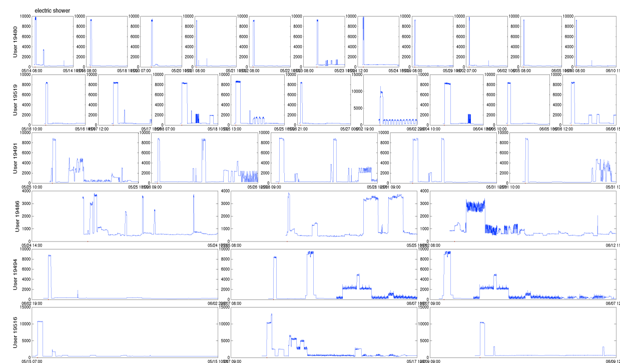

Fig. 10. Variety of electric showers in the UK

IV. FUTURE WORK

Based on the results we see, extending this technique to real time inquiry generation could provide much better results.

A number of gamification techniques can be applied to this approach because of its high user interactivity. [2]

V. CONCLUSIONS

The experiments prove that crowdsourcing appliance tags could be an efficient and inexpensive way to collect large amounts of ground truth. It is also apparent that there is an opportunity to improve a number of methods used to further enhance the benefits.

As a secondary benefit of this approach, we also note a significantly higher user interest and engagement level. This could be an end in itself that can help toward greater awareness and energy savings.

REFERENCES

[1] Parson, O. (2016) *Public Data Sets for NIALM.* Retrieved from http://blog.oliverparson.co.uk/2012/06/public-data-sets-for-nialm.html

[2] Frederick Grossberg, Mariel Wolfson, Susan Mazur-Stommen, Kate Farley, and Steven Nadel, “Gamified Energy Efficiency Programs,” American Council for an Energy-Efficient Economy. Washington, DC, Report Number B1501, February 2015.

[3] John Le, Andy Edmonds, Vaughn Hester, Lukas Biewald, “Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution”, SIGIR ’10, Geneva, Switzerland, July 2010.