

# Knowledge Discovery and Data Mining for Enhanced Sustainability of Physical Ecosystems

Manish Marwah<sup>1</sup>, Ratnesh Sharma<sup>1</sup>, Naren Ramakrishnan<sup>2</sup>, Cullen Bash<sup>1</sup>, Chandrakant Patel<sup>1</sup>

<sup>1</sup>*HP Labs, Palo Alto, CA*

<sup>2</sup>*Virginia Tech, Blacksburg, VA*

## I. PROBLEM ADDRESSED

In recent years, global warming, depleting conventional sources of energy and rising costs have led to increased focus on sustainable operation of physical ecosystems. A physical ecosystem here refers to a collection of entities and processes that consume physical resources and fulfill a functionality, e.g., an IT ecosystem could consist of data centers (including computing, power and cooling resources), client devices, networking infrastructure and print factories; a power ecosystem could consist of power plants, distribution infrastructure, and, where applicable, local power generation equipment such as solar panels, wind turbines or diesel generators. Typically, and more so recently due to the availability of inexpensive sensing devices, an ecosystem generates large volumes of data related to its physical and operational state, including environmental sensor data (e.g. temperature), operational state of systems and devices (e.g. utilization values), and, user demand (e.g. user requests sent to a data center, or load at a power distribution site). This data, available as real-time streams or from an archival store, contains important correlations, trends, and patterns that can be exploited for optimizing operations with respect to a sustainability metric, such as, energy consumption or carbon footprint. However, considering the large amount of data produced, manual inspection is virtually impossible and thus automated knowledge discovery and data mining techniques [1] are necessary to synthesize models for enabling sustainable end-to-end operation and management of physical ecosystems.

## II. KNOWLEDGE OF PRIOR WORK

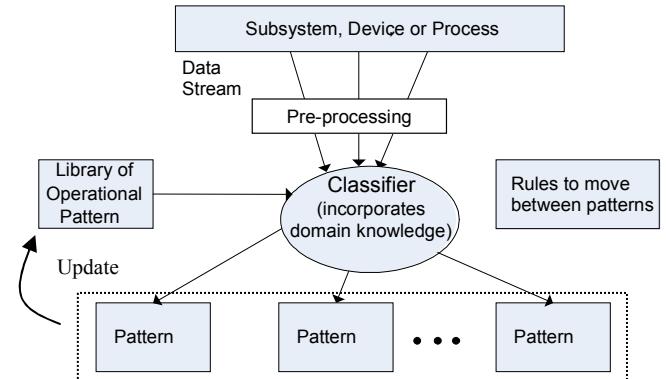
Improving sustainability of physical ecosystems is a challenging problem. For example, in an IT ecosystem [2], dynamically optimizing sustainability across power, cooling and IT subsystems of a data center is hard and knowledge discovery techniques can compliment “first principles” approaches such as CFD-based models [3].

## III. PROPOSED METHODOLOGY

In this paper, we propose a methodology for real-time characterization of the operational state of a subsystem, device or process in an ecosystem, with the goal of improving its sustainability. The objective is to re-describe the operational state, identified by a set of relevant sensor measurements and other parameters, in terms of known operational patterns. Each pattern is also associated with an operational efficiency attribute, which can incorporate a number of criteria, such as

TCO (total cost of ownership), and sustainability metrics such as exergy, power consumption and carbon footprint. This attribute quantifies the desirability of operating the system in that state. The aim is to monitor these patterns and determine if it is possible to transform the current operational state to another more efficient pattern. Domain knowledge required to move from one pattern to another, where feasible, can be encoded into actionable rules.

Figure 1 shows the main elements of the framework. It consists of: (1) a pre-processing module, which transforms the data, and includes interpolation of missing data and synchronization of different data streams; (2) a library of operational patterns constructed from historical data, with an efficiency attribute, characterizing its sustainability, estimated for each pattern; and (3) a classifier that classifies operational state as one of the known patterns.



**Figure 1: Framework to characterize the current operational state as a pattern.**

## IV. CONCLUSIONS

Physical ecosystems, such as an IT ecosystem, consist of multiple disparate entities that generate large volumes of data related to their environmental and operational state. Knowledge discovery and data mining techniques can be applied to identify the current state, and use knowledge derived from historical data to transform the system to a more sustainable operational state.

## REFERENCES

- [1] Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques”, 2nd edition, Morgan Kaufmann, 2006.
- [2] C. Bash, C. Patel, A. Shah, R. Sharma, “The Sustainable Information Technology Ecosystem”, *ITHERM’08*, June 2008, Orlando, FL.
- [3] M. Marwah et al., “Data analysis, Visualization and Knowledge Discovery in Sustainable Data Centers”, *Proc. of ACM Compute*, Jan 9-10, 2009, Bangalore, India.