2018 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 26-28, 2018

COMBINED ONTOLOGY-DRIVEN AND MACHINE LEARNING APPROACH TO MONITORING OF BUILDING ENERGY CONSUMPTION

Parastoo Delgoshaei[1], Mohammad Heidarinejad[2] and Mark A. Austin[3]
[1]National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA
[2]Illinois Institute of Technology, Chicago, IL, USA
[3]University of Maryland, College Park, MD, USA

Abstract

A number of studies within the building domain have independently considered knowledge representation and reasoning (KRR) and machine learning (ML). This paper explores opportunities for how these artificial intelligence (AI) technologies can be combined to provide synergic assistance in intelligent building systems. A case study is used to demonstrate the construction of semantic knowledge from weather and utility ontologies. That knowledge is used as a semantically annotated training set in clustering algorithms that identify consumption patterns, and the results of clustering are stored in ontologies for further inference. This paper presents a roadmap to intelligent building monitoring techniques that uses both historical data and the underlying semantic knowledge.

Introduction

Problem Statement. In 2016 the building sector in the U.S. consumed approximately 40% of the total energy consumption (EIA), suggesting that with advances in technology over the past few decades, positive opportunities exist. One opportunity resides in addressing challenges associated with buildings-to-grid integration (EERE 2018). First, there is currently an unprecedented amount of metered building energy and sensor data that are not being effectively used to optimize energy performance of buildings. Second, building energy management systems (BEMS) are used in only approximately 40% of commercial buildings (typically large commercial buildings), but there is an interest in utilizing sensors and data storage in all buildings. Present-day expectations are that deployment of sensors in support of connected devices and systems will grow 80% annually (Jiron, 2018). Current buildings under-utilize the available sensor data and do not efficiently integrate with the smart grid. Our work is driven by the belief that these limitations stand in the way of improvements to energy consumption in the building sector.

Scope and Objectives. This paper takes a first step toward enabling modern technologies by using judicious combinations of knowledge representation and reasoning (KRR) and machine learning (ML) to provide new approaches in management of energy in buildings. We envision that KRR and ML formalisms will work side-by-side, providing complementary and supportive roles in the collection and processing of data, identification of events and automated decision making.

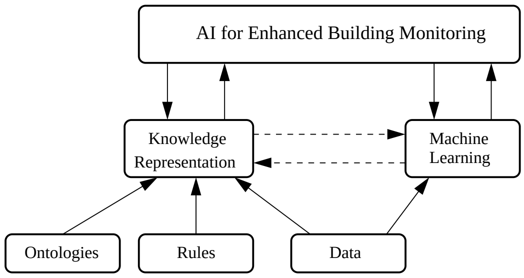

Figure 1: AI for enhanced energy monitoring.

Figure 1 is a high-level architectural schematic of the KRR/ML components and their interactions. The semantic (KRR) side is defined by domain and meta-domain data, ontologies and rules that can dynamically respond to events. The KRR encapsulates the data and represents it with reduced dimensionality. The ML side will classify data into collections and learn about cause-and-effect relationships embedded in the data. ML can also be developed for the identification of anomalies (faults) in system performance, thereby acting as a trigger for the activation of KRR diagnostic procedures. With this approach, the knowledge of cause-and-effect relationships embedded in the data revealed by ML algorithms can provide verification for the effectiveness of rules developed for KRR.

The design of a path from what "seems like a good idea" to a "prototype implementation" requires that three important challenges be addressed: (i) gathering of knowledge about the physical building environment, (ii) design of an efficient means for dealing with large volumes of heterogeneous streaming data (e.g., occupancy, utility, weather, building architecture, sensor and equipment), and (iii) use of the knowledge and data as a source of semantically annotated training sets in machine learning algorithms. Since the goals of the current study are just to demonstrate that the KRR-ML interaction is advantageous, we deliberately keep the data analysis simple.

Background

This section presents related work in knowledge modeling with ontologies and rules, and supervised and unsupervised machine learning.

Ontology Modeling and Representation. The work in (Lee E.A. 2003), presented an extension to the World Wide Web (WWW) named the Semantic Web. The Semantic Web is capable of automatically handling Web data without human knowledge interpretation and input. In the context of the Semantic Web, data retrieval is based on semantic relationships between data categories and classes, not just numeric values. This improves knowledge sharing and integration of homogenous data sources. In KRR, an ontology is a key element to formally and explicitly describe the main concepts or classes of the domain that stores data for and the relationships between those concepts. One common language to describe ontologies is Web Ontology Language (OWL). Other languages, such as Semantic Web Rule Language (SWRL), can express rules and logic in a model. Rules are mechanisms to infer implicit knowledge based on explicit facts in the ontology. Semantic Web technologies have been adopted in research efforts in computer science over the past two decades (Liu X., Li Z. and Jiang S. 2016).

Supervised and Unsupervised Machine Learning. ML techniques learn about a system's behavior and support decision-making and predictions. These methods are being used to solve complex engineering applications that entail a large number of independent parameters and nonlinear interdependencies that cannot be easily modeled from first principles. For our purposes, understanding building energy consumption patterns is among these applications, where outdoor weather, building occupancy, and performance of mechanical systems are the features that influence the building energy performance.

ML algorithms can use supervised or unsupervised learning. Supervised learning typically encompasses two steps: (i) training and (ii) predicting. Datasets are also divided into training datasets and testing datasets. The training step allows identification of the decision model that provides the dependency of the target (predicted variable) on the features (impacting variables). In the next step, the decision model is applied to the testing datasets; the effectiveness of the prediction performance of the model can then be calculated. Supervised learning such as the nearest neighboring algorithm, requires labeled datasets (e.g., the data are labeled with the correct answer), a process that can be very expensive. This algorithm makes predictions on new data points based on their proximity to the points in the training set.

In contrast to supervised learning, the goals of unsupervised learning such as the K means clustering algorithm are to model and identify the underlying structure or patterns in a dataset when no correct answers (labels) are provided. Semi-supervised learning methods fall between the strategies of supervised and unsupervised learning and employ combinations of labeled and unlabeled data. First, the unsupervised method is used to identify patterns and then supervised learning is used to draw the best predictions for the unlabeled data using the labels generated by unsupervised learning. The prediction decision model is then tested on labeled data. This semi-supervised technique can address a wide range of engineering applications including building energy performance, and procedures needed for fault detection and diagnostic analysis of building equipment.

Machine Learning for Building Energy. Recently, data scientists and engineers have applied machine learning techniques to a variety of problems ranging from fraud protection to online advertising. Applications in building performance include mechanical system controls, fault detection and diagnostics, and building energy monitoring. The K-means clustering algorithm is widely used to assess the energy performance of buildings (Miller C., Nagy Z. and Schlueter A. 2018). This unsupervised technique categorizes data into subgroups that share similarities. (Heidarinejad, M., Dahlhausen, M., McMahon, S., Pyke, C. and Srebric, J. 2014) used K-means cluster analysis to classify buildings in general into high, medium, and low energy intensity in order to identify similarities and differences in the key variables contributing to the energy use patterns. As a case for the supervised learning, (Valgaev O. and Kupzog F. 2016) developed a model that is parameterized automatically and provides a forecast using only historic building load measurements as an input based on K-nearest neighbor approach. It was used on a large sample of simulated mixed-usage buildings of different sizes. Their results show that the model accuracy is superior to the forecast obtained using individual load profiles created for each building.

An important consideration, in addition to energy consumption, is the cost of electricity. A flat electricity rate is common in residential buildings, but utility providers are beginning to offer programs that provide a balance between the demand and supply of electricity. Time-of-use

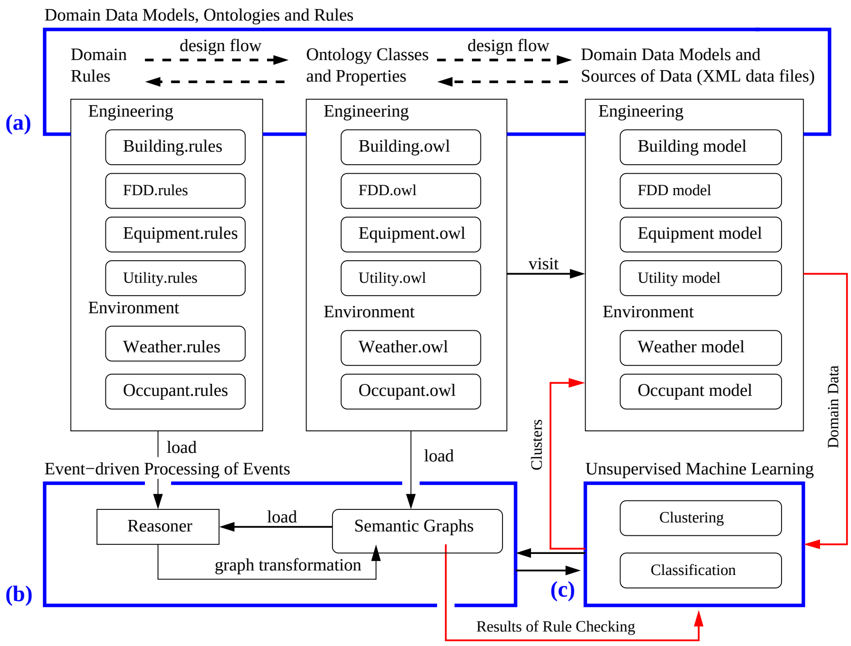

Figure 2: Proposed architecture for knowledge-assisted machine learning approach to building monitoring and management.

(TOU) programs calculate the cost of electricity based on the time of day and date. A common TOU program offers summer and winter rates, and each season could have up to three rates: on-peak, mid-peak, and off-peak. Another example is a voluntary rebate program in which the resident allows the utility provider to cycle the resident's central air conditioner or heat pump over short intervals on selected summer days. These days are often called peak savings days and only occur for a short period of time during the summer. We anticipate that machine learning techniques will provide valuable insight on the impact of complex utility rate structures on the electricity consumption of residential buildings.

Proposed Methodology

Figure 2 is the proposed architectural schematic for a combined semantic modeling and machine learning approach to building energy monitoring. The semantic modeling and machine learning blocks and their interactions serve the following purposes:

Block (a). On the semantic modeling side of the problem, data, ontologies and rules are placed on an equal footing and are developed for a multiplicity of domains. The goal is to keep the ontologies small and use object property relationships and rules to link sources of data needed for multi-domain reasoning. A key benefit in co-developing the data models with ontologies and rules for a domain is that it represents domain knowledge and supports rule-based decision making. Rules can be developed to infer new knowledge by combining and reasoning with data from multiple domains to help with cross-domain decision making.

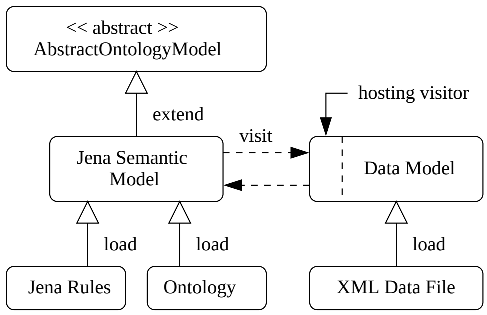

Figure 3: Schematic for ontologies visiting data models.

The ontologies and rules are implemented in Jena (Apache Jena: 2016), and the data models in Java. The right-hand side (c) of Figure 2 deals with data sources and models across the participating domains (e.g., building occupant, utility, weather, building architecture, fault detection and diagnostics). The semantic graph models will be populated with individuals (i.e., instances of real-world data) by visiting (a mechanism software design pattern: see Figure 3) the relevant data models and gathering the data and object properties relevant to the application at hand. For example, the occupant emulator will provide data to the Occupant ontology; the weather server and data model will add instances to the Weather ontology.

Block (b). The second purpose of semantic modeling is to provide support for the executable processing of events associated with features of the building domain and energy consumption. As illustrated in the center and left-hand sides of Figure 2, the domain ontologies and rules (e.g., Weather.owl and Weather.rules; Utility.owl and Utility.rules) are loaded into the semantic model and general purpose reasoner. Semantic graph representations can listen (Listener software design pattern) for incoming events – this, in turn, triggers the execution of rules and, if needed, transformation of the semantic graph. The result is semantic graphs that dynamically adapt to the consequences of incoming data and events.

Block (c). This study uses K-means cluster analysis and the nearest neighboring algorithm on residential building data to classify and predict building electricity energy consumption. Our prototype software uses TensorFlow (Abadi M., Barham P., et al. 2016), an open source Python library, for machine learning purposes. We employ JPype (JPype 2018) to integrate Apache Jena (Java) (Apache Jena: 2016) with TensorFlow (Python). It is important to note that this framework allows us to use semantic information as features in learning algorithms (e.g., the results of the knowledge inference such as the category of utility rate On-Peak, Off-Peak, or the occupancy schedule).

Data flows among blocks (a)-(b)-(c). Interactions between the blocks for semantic model representation, executable processing, and machine learning are defined by two streams of data: (1) the machine learning block uses the inference results in the ontologies, and (2) results from machine learning techniques are fed back into the ontolgoies for further inferencing.

Case Study Problems

Problem Description. The case study exercises a subset of the vision described by Figure 2, and involves a coupling of semantic modeling and reasoning, and machine learning for the weather, utility, and occupant domains. This case problem uses hourly data on electricity usage (obtained from the utility company in 2016) for one bedroom residential buildings located in Maryland for a four-month winter utility season (January 1 through April 30) and one month of the summer utility (May 1 through May 31) season. Data is provided for the rate category, the weather condition (hot, cold), occupancy schedule, and dew point.

Domain Ontologies. Figures 4 and 5 show fragments of the classes, data and object properties in the Weather and Utility ontologies. Current and future forecast weather data was collected from an online server (WeatherAPI) and stored in Weather ontology. The utility ontology employs temporal concepts such as time interval and time instant (Petnga L. and Austin M.A. 2013) to represent utility tariff scheduling. In this ontology, a season is a concept that begins and ends with time instants and may have one or more sub-intervals associated with the tiered rates; e.g., on-peak, off-peak, and mid-peak rates. At any point in time, temporal reasoning procedures are capable of determining whether or not a peak interval is in effect, and identifying the associated rate for that season and time of the day.

Jena Rules for Domain Ontologies. The occupancy schedule is defined as an inference rule in KRR shown in Figure 6. This example shows how the knowledge about the occupants' presence during a weekday can be derived from temporal reasoning and semantic facts, i.e., time of day and the calendar (weekday, weekend, holiday). In this setting, the semantic terms such as holiday and weekend are used in the occupancy schedule inference rule, the values of which are computed from Jena boolean built-in functions named isWeekend() and isHoliday(). Figures 7 and 8 represent sample Jena rules associated with weather and utility ontologies, respectively. The former presents a semantic rule for the identification of a frost (i.e., observed temperature is below 0 °C) condition. In the latter, UtilityRule01 identifies the season that "time-of-the-use" belongs to and UtilityRule02, identifies if that time-of-use belongs to the on-peak tier.

Results

Figure 9 shows the distribution of actual, obtained from utility provider, hourly consumption target values compared with predicted values. As shown in the graph, the shape of the predicted distribution is similar to the actual distribution. Figure 11 shows results of the classification analysis using the k-nearest neighbor algorithm to predict electricity consumption. The features for this analysis were based on outdoor temperature, solar radiation, and wind speed from raw data, and occupancy obtained from KRR. The algorithm used 90 % of annual electricity for training and 10 % for testing and k = 4. The training set is composed of 700 data points and with the batch size of 100, 7 mean square errors (MSE) were computed as 0.033, 0.041, 0.054, 0.036, 0.021, 0.027, 0.021. The predicted values are close to the actual values except when predicting unusually high consumption, $>$ 0.5 kWh. Overall, based on Figure 9 distribution, the data is mainly concentrated in the low energy consumption range, less than 0.2 kWh. This observation and the good prediction on low range results from Figure 11 suggest that the model works well for these data in the study.

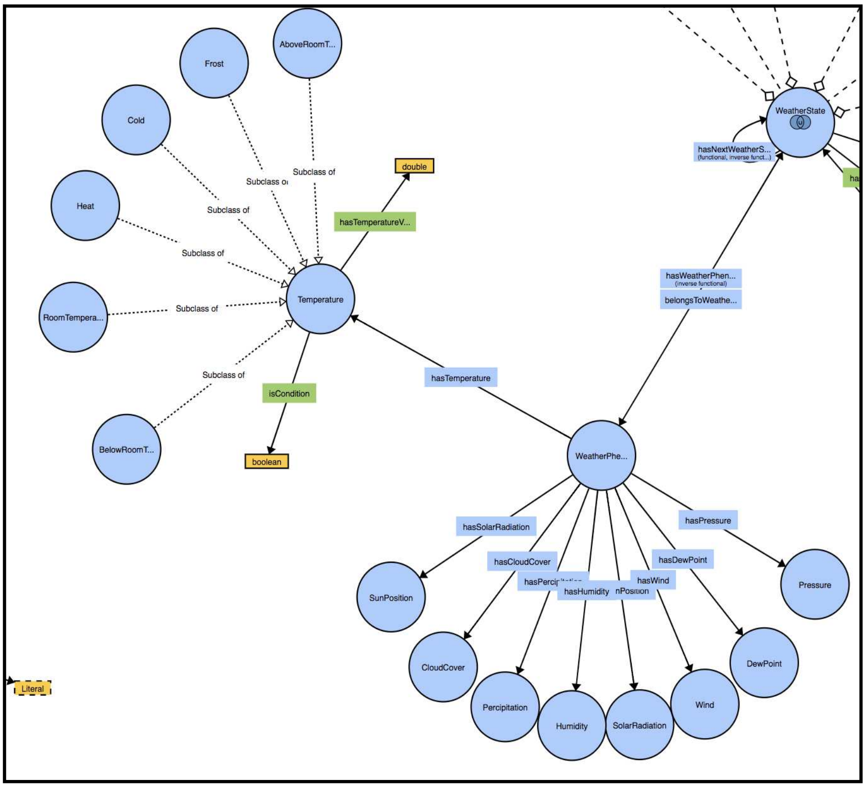

Figure 4: Partial view of weather ontology classes and properties (Source: Adapted from Staroch (Staroch P. 2013; Delgoshaei P. 2017)).

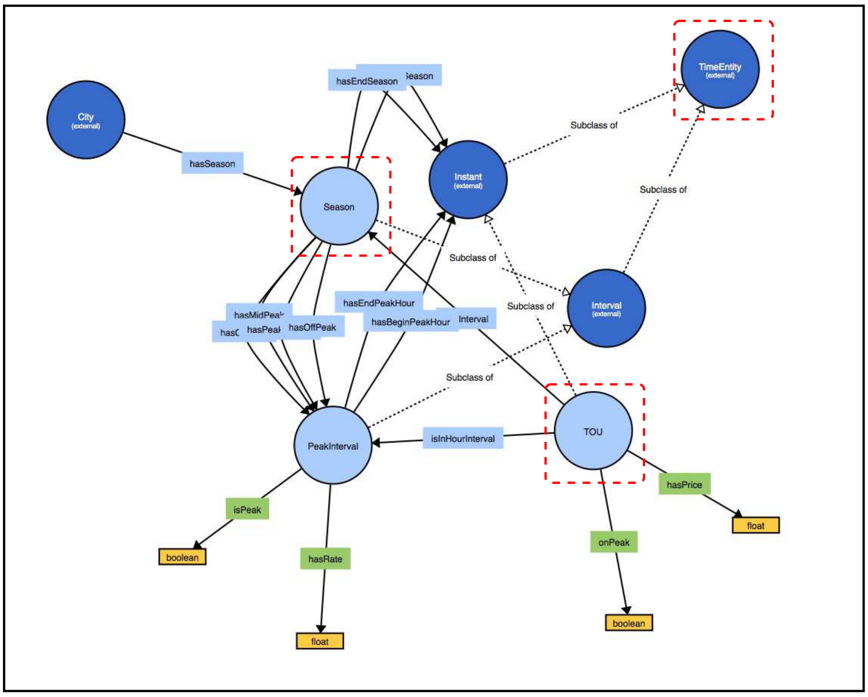

Figure 5: Schematic of utility ontology (Delgoshaei P. 2017).

Jena Rules

```
[weekDayVacant: (?z rdf:type bld:Zone)(?z bld:hasSchedule ?sc) equal(?sc,"Schedule1")
                (?td rdf:type te:TOU) isWeekend(?td,?w) isHoliday(?td,?h)
                notEqual(?w,"false"^^xs:boolean ) notEqual(?h,"false"^^xs:boolean)
                (?td te:hasTimeValue ?time) le("16:59:59"^^xs:time,?time)
                ge("06:59:59"^^xs:time,?time)->(?s bld:isOccupied "false"^^xs:boolean)]}
```

Figure 6: A Jena rule to detect schedule of occupancy during weekdays that do not fall on a holiday.

Jena Rules

```
[WeatherRule01: (?t rdf:type we:Temperature) (?t we:hasTemperatureValue ?tv)
               lessThan(?tv,0) ->  (?t rdf:type we:Frost)
               (?t, we:isCondition, "true"^^xs:boolean) print(?tv,'FrostCondition')]
```

Figure 7: A Jena rule for weather ontology to detect frost (Delgoshaei P. 2017).

Jena Rules

```
[ UtilityRule01: (?interval rdf:type te:Season) (?interval te:endsAt ?end)
                 (?interval te:beginsAt ?begin) (?t rdf:type te:TOU) (?t te:hasTime ?time) lessThan(?begin,?time)
                 lessThan(?begin,?end) greaterThan(?end,?time) -> (?t te:isInInterval ?interval) ]

[ UtilityRule02: (?t te:isInHourInterval ?interval) (?interval te:isPeak ?peak) -> (?t te:onPeak ?peak) ]
```

Figure 8: Jena rules for the utility ontology to identify the season and the tiered rates(Delgoshaei P. 2017).

Figure 10 is a plot of clustering analysis on energy consumption based on inferred weather condition obtained from KRR side. The main categories of the weather condition are: Room, Below Room, Cold and Frost. These conditions are inferred based on the semantic rules depicted in Figure 7. The data represents the electricity consumption for the heating and the shoulder season, from January to May. The clusters also confirm the energy consumption as a function of weather conditions belong two these seasons. The red cluster containing the Room and Below conditions represents the shoulder season and the blue cluster represents the heating season. The centroids, the dataset average, are marked by "x" representing each cluster. The results show that the two clusters confirm the expected seasons of heating and shoulder seasons. Also, the heating cluster has a higher average consumption than the shoulder season. This is a potentially useful insight for utility companies on how to structure their electricity tariffs (two vs. three seasons) based on semantic labels (i.e., frost condition associated with the heating season) rather than temperature numeric values.

Figure 12 shows results of the cluster analysis for electricity consumption and its relationship to outdoor air temperature and tiered utility rate (1 represents on-peak and 0 represents off-peak). The tiered rates were obtained from semantic rules defined in Figure 8. This study considers two clusters to represent conditions during cold and shoulder seasons. The consumption during the peak hours in summer is not as well correlated with outdoor air temperature. This reflects the setpoint reset during unoccupied hours. However, during the off peak, occupied hours in winter, heating is correlated with the outdoor air temperature. Overall, the results of this figure suggesting that during (1) occupied, extreme weather conditions, there is a linear correlation between consumption and outdoor air temperature and (2) during moderate weather conditions, the energy consumption has little correlation with outdoor temperature and occupant behavior.

Discussion

A recent study (BAS 2018) indicates that the market for building automation system installation is expected to expand at 4.6 percent annual growth from 2016 through 2022. This trend is driven in part by the belief that energy efficiency of buildings can be improved through the use of data/information working alongside technologies for AI, big data management and cloud platforms. Implementation of this prototype requires considerations when integrating ML techniques with KRR models. The specific areas include: data exchange and synchronization

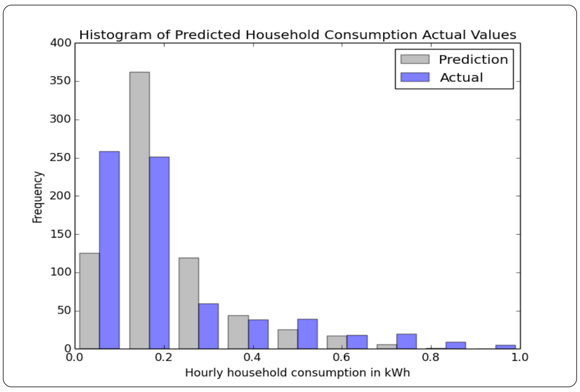

Figure 9: Distribution of predicted and actual energy consumption test data for k-NN (k=4).

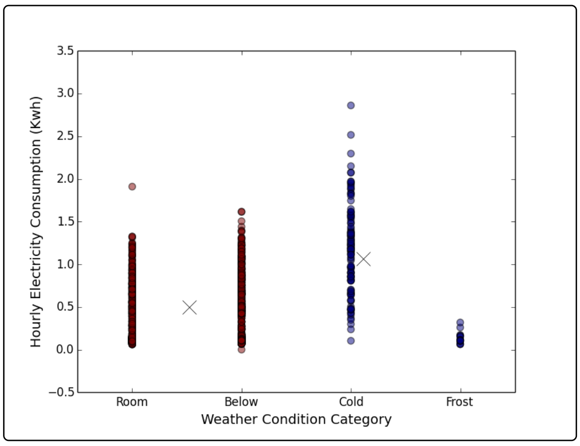

Figure 10: Hourly energy consumption as a function of weather condition category.

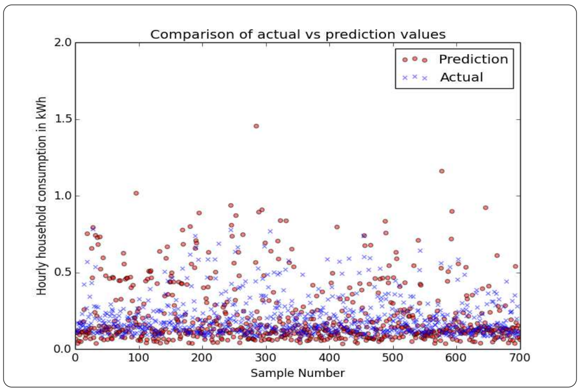

Figure 11: Predicted values and actual target values of energy consumption based on occupancy and weather conditions.

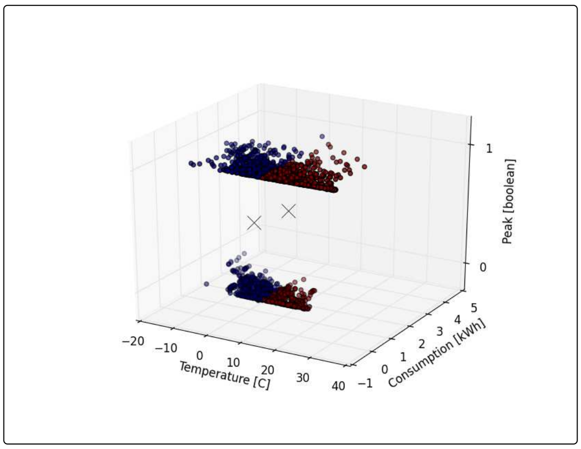

Figure 12: Cluster analysis for hourly electricity consumption based on outdoor air temperature and utility category.

of results between ML and KRR models, and handling the difference in processing time in the semantic domain (slow) and the ML algorithms (fast). This study demonstrates how collection of data-ontology-rules can be used to represent and reason with data from multiple domains and provide essential semantic knowledge to the learning algorithms.

Conclusion

This paper describes an approach to monitor building energy consumptions by integrating machine learning techniques with mechanisms for semantic knowledge representation and reasoning. This work implements a supervised learning algorithm, nearest neighbor, to predict the electricity consumption based on raw data such as solar radiation, outdoor temperature, and wind speed, as well as knowledge data such as occupancy inferred by semantic rules. We also integrated the semantic knowledge in weather conditions (i.e., frost, above room temperature, below room temperature) integrated to K-means clustering algorithm to identify the electricity consumption seasons (i.e., heating, shoulder, cooling). Our long-term vision is that this framework will be used to couple semantic and machine learning techniques for buildings-to-grid integration.

Acknowledgement

The authors would like to acknowledge National Research Council Research Associates Program as a funding source for Parastoo Delgoshaei.

References

Abadi M., Barham P., et al. 2016. "TensorFlow: A System for Large-Scale Machine Learning."

Apache Jena:. 2016. An Open Source Java framework for building Semantic Web and Linked Data Applications. For details, see https://jena.apache.org/.

BAS. 2018. BAS Automation Systems. For details, see https://internetofbusiness.com/ demand-building-automation-systems-continues-rise-says-study/.

Delgoshaei P. 2017. "Semantic Models and Reasoning for Building System Operations: Focus on Knowledge-Based Control and Fault Detection for HVAC." Ph.D. diss. in Civil Engineering, University of Maryland, College Park, MD.

EERE. 2018. Buildings and Grid helping Commercial Buildings get Smarter, For details, see https://energy.gov/eere/buildings/articles/ buildings-and-grid-helping-commercial-buildings-get-smarter.

EIA. See https://www.eia.gov/tools/faqs/, 2016.

Heidarinejad, M., Dahlhausen, M., McMahon, S., Pyke, C. and Srebric, J. 2014. "Cluster Analysis of Simulated Energy Use for LEED Certified U.S. Office Buildings." *Energy and Buildings* 85:86 – 97.

Jiron, 2018. Jiron A., Buldings and the Grid: Helping Commercial Buildings Get Smarter: For details, see: https://www.energy.gov/eere/buildings/articles/ buildings-and-grid-helping-commercial-buildings-get-smarter, February, 2018.

JPype. 2018. Bridging the Worlds of Python and Java. For details, see http://jpype.sourceforge.net.

Lee E.A. 2003. Model-Driven Development – From Object-Oriented Design to Actor-Oriented Design, Presentation at Workshop for Software Engineering for Embedded Systems, From Requirements to Implementation, Chicago,.

Liu X., Li Z. and Jiang S. 2016. "Ontology-Based Representation and Reasoning in Building Construction Cost Estimation in China." *Future Internet* 8, no. 3.

Miller C., Nagy Z. and Schlueter A. 2018. "A Review of Unsupervised Statistical Learning and Visual Analytics Techniques Applied to Performance Analysis of Non-Residential Buildings." *Renewable and Sustainable Energy Reviews* 81:1365 – 1377.

Petnga L. and Austin M.A. 2013. "Ontologies of Time and Time-based Reasoning for MBSE of Cyber-Physical Systems." *Procedia Computer Science* 16:403 – 412. Conference on Systems Engineering Research (CSER 2013).

Staroch P. 2013. A Weather Ontology for Predictive Control in Smart Homes. M.S. Thesis in Software Engineering and Internet Computing, Vienna University of Technology.

Valgaev O. and Kupzog F. 2016. "Building Power Demand Forecasting Using K-Nearest Neighbors Model - Initial Approach ." pp. 1055 – 1060. 2016 IEEE PES Asia-Pacific Power and Energy Conference - Xi'an - China.

WeatherAPI. Online Open Weather API. See https://openweathermap.org/api (Accessed September 14, 2017).

Nomenclature

AI	Artificial Intelliegence
BAS	Building Automation Systems
OWL	Web Ontology Language
FDD	Fault Detection and Diagnostics
KRR	Knowledge Representation and Reasoning
ML	Machine Learning
WWW	World Wide Web
SWRL	Semantic Web Rule Language
TOU	Time of Use