Applied Energy 132 (2014) 1–14

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Dynamic classification system in large-scale supervision of energy efficiency in buildings

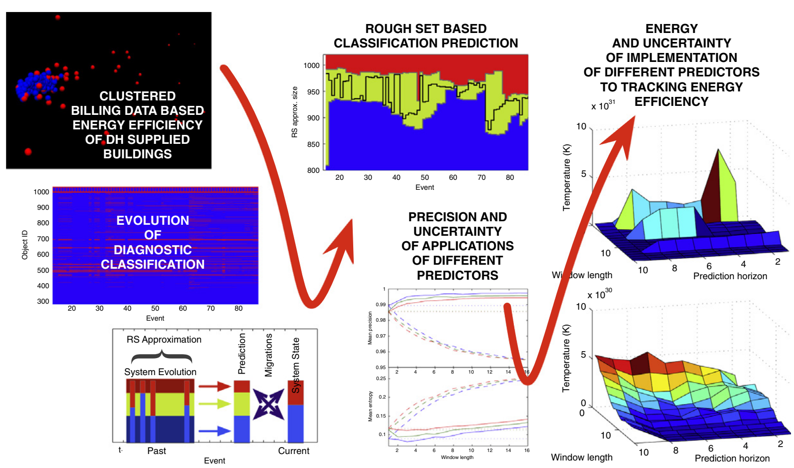
S. Kiluk *

AGH-University of Science and Technology, Faculty of Management, Gramatyka 10, 30-067 Krakow, Poland

HIGHLIGHTS

- Rough set approximation of classification improves energy efficiency prediction.
- Dynamic features of diagnostic classification allow for its precise prediction.
- Indiscernibility in large population enhances identification of process features.
- Diagnostic information can be refined by dynamic references to local neighbourhood.
- We introduce data exploration validation based on system dynamics and uncertainty.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 16 October 2013
Received in revised form 19 June 2014
Accepted 20 June 2014

Keywords:
District heating
FDD
Knowledge dynamics
Information entropy
Rough sets

ABSTRACT

Data mining and knowledge discovery applied to the billing data provide the diagnostic instruments for the evaluation of energy use in buildings connected to a district heating network. To ensure the validity of an algorithm-based classification system, the dynamic properties of a sequence of partitions for consecutive detected events were investigated. The information regarding the dynamic properties of the classification system refers to the similarities between the supervised objects and migrations that originate from the changes in the building energy use and loss similarity to their neighbourhood and thus represents the refinement of knowledge. In this study, we demonstrate that algorithm-based diagnostic knowledge has dynamic properties that can be exploited with a rough set predictor to evaluate whether the implementation of classification for supervision of energy use aligns with the dynamics of changes of district heating-supplied building properties. Moreover, we demonstrate the refinement of the current knowledge with the previous findings and we present the creation of predictive diagnostic systems based on knowledge dynamics with a satisfactory level of classification errors, even for non-stationary data.

1. Introduction

Building energy use is an important instrument for the evaluation of district heating (DH) operation, especially when energy efficiency is considered. The current development of metering systems

* Tel.: +48 12 617 4507; fax: +48 12 636 7005.
E-mail address: skiluk@agh.edu.pl

http://dx.doi.org/10.1016/j.apenergy.2014.06.054
0306-2619/

opens the possibility for intensive data exploration and the introduction of a variety of strategies for the acquisition and validation of diagnostic knowledge. Fault detection diagnostics (FDD) of DH-supplied buildings has remained a focus of research and development for a notably long time. Although progress has been made concerning the technical level of relevant data selection and modelling non-linearities [1–4], there is still a lack of reliable and scalable instruments for the evaluation of building energy use as influenced by the complex impact of occupants behaviour and building maintenance policy. One of the reasons is that the algorithms used for energy consumption-based diagnostics in buildings supplied from a DH network and based on the billing data require up-to-date knowledge following the dynamic development observed in supervised objects related both to technical and socioeconomic determinants.

It is difficult to include explicitly the different factors that affect the properties of heating energy consumption [5,6] in the diagnostic models [7,8]; thus, operator-oriented supervision structures are required [9–11] for centralised maintenance and service of DH-supplied buildings. Usually, the tasks involved in the FDD are based on the identification of the models for residual analysis [12,13] or pattern recognition and expert-based reasoning [14–16].

Although applications of the only recently developed Data Mining and Knowledge Discovery (DMKD) algorithms for the supervision of energy efficiency in buildings are highly promising, they require the validation of results with a large amount of external information [17]. However, collecting the required precise information on an individual building level would increase system reliability at the cost of scalability and affordability [18]. In contrast, for classification systems derived from a series of past algorithmically gained partitions of the problem, its evaluation can be addressed in the domain of dynamic system theory, where methods based on the analysis and exploration of system evolution have been established.

The dynamics of algorithm-based knowledge brings new possibilities for evaluation of the validity of the supervision system, but it has not been explored to date for industrial data collected for DH diagnostic tasks. However, the application of DMKD to the billing data from a DH distribution system allows the collection of sequences of diagnostic patterns and classifications of remote heating systems [19,20] making it possible to address the validation problem in the domain of the dynamic properties of DMKD findings.

The quality of the building energy consumption classification system is essential for a proper implementation of the output derived from data exploration and requires additional validation. Hypotheses for the desired properties of the resulting classification evolution include specific system dynamics and the uncertainty of the resulting information system. However, the precise construction of the measures of the dynamic content of data mining-based classification system itself is unknown.

The evolution of algorithmic classification of energy-consuming buildings is attributable to changes in building properties; thus, the classification relevance for supervisory tasks can be examined in the domain of system dynamics. The specific design of a multi-stage data exploration algorithm combined with a reasoning system algorithmically produces classifications of buildings by differentiating the level of energy consumption expressed in terms of energy consumption model coefficients [19]. Furthermore, model coefficient estimation errors are also taken into account when distinguishing objects because of the uncertainty of the identified relation between the outside temperature and heating energy use. Buildings with similar and certain energy use constitute a class opposed by another class containing buildings with excessive or uncertain energy consumption; thus, the classification system automatically divides the population into two subgroups of buildings with clear differences in the amount and uncertainty of energy use. Though the qualitative distinction between the classes is clear, the currently obtained borderline between classes might vary in time, influencing the results of the classification implementation. Because the static properties of the current classification have been demonstrated to be bound to the energy use [19], we focused on the dynamic aspects of the series of classifications to draw inferences about the relevance of the classifications for energy use tracking and the sustainability of the obtained results. Despite changing outside temperatures, energy consumption levels and corresponding model properties, we expect the obtained diagnostic knowledge not to change significantly and constitute dynamic references for the detection of unexpected changes of individual energy-use model properties. The observed non-stationarities of models coefficients estimated in moving windows were explored to verify their impact on the composition of classes. Furthermore, the relation between the variation of expected energy use in the series of classes and classification uncertainty changes was investigated with the originally developed post-implementation measure.

There is a demand for methods that would further introduce generalisation derived from the dynamics of knowledge in different areas, in which the nonstationarity of the models representing the current knowledge or statistical properties of the applied models is a limiting factor [21]. In the areas where DMKD use is established [22,23], additional generalisation brought about by the analysis of the dynamics of the discovered knowledge can be valuable in introducing new measures to describe the features of the investigated processes. The properties of the dynamic system can be investigated in terms of their ability to follow the observed changes and differentiate common factors constituting neighbourhood-level references for future reasoning regarding the objects. Furthermore, the dynamics of algorithm-based results provide new ways to validate the quality of the DMKD results and creates new prospects for the management of algorithmically gained knowledge.

In this paper, we adopt a strategy to validate the DMKD of DH billing data involving the investigation of the dynamic properties of the gained classification with predictors based on rough set (RS) approximations of past diagnostic classes. To aggregate information from heating processes, we employ simple energy consumption models and explore similarity relations between the supervised objects in terms of their sustainability during a long period of consecutive observations. When not changing, such similarity relations are called *steady* from the point of view of classification system evolution. Energy use similarity relation is expressed in terms of the statistical significance of differences of model parameters related to building energy efficiency. For efficient energy use, we refer to the delivery of functionality such as indoor climate with the lowest possible amount of energy and, as described in the paper, this concept is specifically useful in DH systems where climate conditioning and network operation are common.

Algorithmically gained building energy consumption clusters contain aggregated information about the similarities between buildings in a quantitative description of the amount and uncertainty of energy use implied by the outside temperature. To ensure the relevance of selected variables for energy consumption, the evaluation significance of differences between clusters in building size, distribution network operation and heating energy consumption was analysed. The meaning of such clusters in the evaluation of efficiency of energy use was established according to obtained statistics.

Classes with an established diagnostic meaning gained from DMKD allow for observation of object migrations, and thus, the RS indiscernibility of consecutive system states was exploited to