Sustainability in Energy and Buildings: Research Advances ISSN 2054-3743
Vol. 10. Sustainability in Energy and Buildings (SEB-22) Short Papers pp.33-38 : seb22s-006

Intelligent Approaches for Fault Detection and Diagnosis in Building Substations: A Position Paper

Jonne van Dreven[*1,2], **Veselka Boeva**[*1], **Shahrooz Abghari**[*1], **Håkan Grahn**[*1], **Jad Al Koussa** [†2], and **Emilia Motoasca**[†2]

[1]*Department of Computer Science, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden*
[2]*Unit Energy Technology, Flemish Institute for Technological Research, Genk, Belgium*

Abstract

This paper discusses the present state of knowledge, methods, and challenges in using intelligent approaches for fault detection and diagnosis (FDD) in building substations associated with district heating (DH) networks. We briefly overview recent advances in the field of automatic FDD in DH from the perspectives of the used approaches and the need for well-defined labeled data. We describe three main steps of fault handling of building substations along with identified challenges and knowledge gaps to be covered by new scientific research. Furthermore, we describe the main phases of the planned research to address the formulated research questions and the outlined hypotheses.

1 Introduction

In Europe, more than 50% of the final energy consumption comes from the heating and cooling of buildings [1]. Decarbonization of the heating and cooling sector is essential and will contribute to the ambitious EU climate goals — the EU aims to be climate-neutral by 2050 by reducing greenhouse gas emissions to net-zero [2]. Collective heating and cooling solutions will play a vital role in the energy transition.

District heating (DH) networks utilize heat generated in a production unit to supply thermal energy to consumers. Figure 1 depicts a schematic overview of a DH network. Using steam or hot water, the heat is transported from the production unit to the buildings through a network of insulated pipes called the

* {jonne.van.dreven, veselka.boeva, shahrooz.abghari, hakan.grahn}@bth.se
† {emilia.motoasca, jad.alkoussa}@vito.be

Presented at KES-SEB-2022 14-16 September 2022

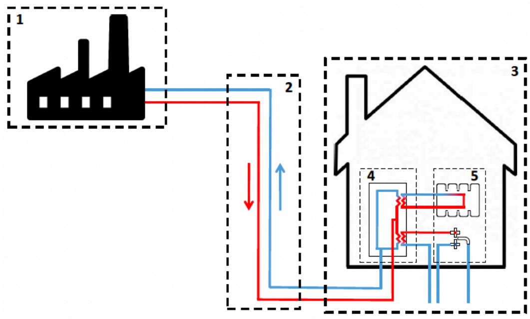

Figure 1: A schematic overview of a DH network, where (1) is the production unit, (2) is the distribution network, (3) is the customer installation, (4) is the substation, and (5) is the building's heating system. Note that (1) and (2) represents the primary side while (5) the secondary side. The image is taken from [3].

distribution network. The DH substation is an interface between the primary side — the production unit and distribution network — and the secondary side — the building's heating system, which is in most cases a hydraulically separated circuit.

An ongoing trend in DH is digitalization [4], introducing opportunities for applying data-driven approaches to make the DH networks more efficient. A significant challenge in automatic fault handling of faulty substations and sustaining optimal operation of current DH generations is slow, inaccurate, and labor-intensive. For example, Månsson et al. [5] estimate that 43% of substations and secondary systems in Sweden are performing sub-optimal; therefore, significantly impacting the efficiency of the DH network, resulting in increases in heat losses, energy consumption, and operational costs [6].

A substation that is performing sub-optimal, i.e., behaves not as intended, e.g., due to faulty components in the substations or the heating system of buildings, affects the operation of the entire DH network. Some DH utilities (DH network operators) may use manual analysis to monitor their DH networks. However, it is often labor-intensive, time-consuming, and expensive, thus, impractical. Therefore, there is an increasing interest in automatic fault detection and diagnosis (FDD) of substations as it can improve efficiency and reduce operational costs. Intelligent techniques provided by machine learning (ML) and artificial intelligence (AI) are promising candidates for automating FDD tasks. Data mining and ML algorithms can be applied for modeling, analyzing, and understanding of DH substations' behavior, which will further lead to developing accurate approaches for automatic FDD.

2 Present state of knowledge

There has been an increase in research publications related to automatic FDD in DH, more than doubling in the last four years [7]. The vast majority of the studies (72%) were mainly concerned with forecasting to plan the operation of the production units and to predict the optimal, intended operation of the substations. Only nine out of 179 studies address FDD in DH, suggesting that the topic is in its infancy. Consequently, the same study suggests that one of the main technical challenges is the availability of labeled datasets. Specifically, there is a need for well-defined datasets that contain labeled faults and fault-free instances, i.e., faults of different types, both in the substations and in the buildings' heating systems.

The study by Abghari et al. [8] present a robust data-driven approach to detect sub-optimally performing substations. Månsson et al. [5] present a statistical approach for fault detection in DH customer installations. The study claims that out of 3,000 substations accounted in this work, 1,273 (43%) were performing sub-optimal. The results deviate from others, e.g., [9] which estimates 75% perform sub-optimal due to the lack of a clear definition of a well-performing substation.

Bode et al. [10] present an ML fault detection approach trained on experimental data, which contains simulated faults of heat pump systems. The study mentions that real-world datasets do not contain much information on faults, as technicians mainly document severe faults; thus, the study has considered most real-world data as fault-free and used oversampling to balance the dataset.

Zhan et al. [11] present a comprehensive literature review of 195 studies in FDD for building energy systems. They point out some interesting studies which have used support vector machines (SVM)-based FDD methods—some achieving 95% accuracy. Overall, data-driven methods have generally higher accuracy, at the cost of needing a large amount of labeled data to train the models. Knowledge-driven methods have higher generalization capabilities and do not need much data. However, they heavily rely on expert knowledge, and adjustments are always necessary.

Wang et al. [12] present an approach to fault diagnosis for heating, ventilation, and air-conditioning and refrigeration (HVAC&R) systems. They successfully present a discrete Bayesian network (DBN) to deal with uncertainty, common in most engineering scenarios. Li et al. [13] propose a two-level fault detection approach using a convolutional neural network (CNN) on simulated data of a DH system. The algorithm successfully classified most system faults and sub-faults.

Lei et al. [14] present a review of ML application to machine fault diagnosis in mechanical systems. The study states that deep learning (DL) has been successful over the years but relies on two assumptions: 1) a large amount of balanced data and 2) a labeled data set. Overall, these assumptions are unrealistic as faults occur less often, and labeled data are quite rare as manual labeling is expensive.

3 Methods and challenges

Automatic fault handling of building substations typically consists of three steps: (i) Fault detection, (ii) Fault diagnosis, and (iii) Fault correction. Figure 2 presents the steps for automatic fault handling in building substations.

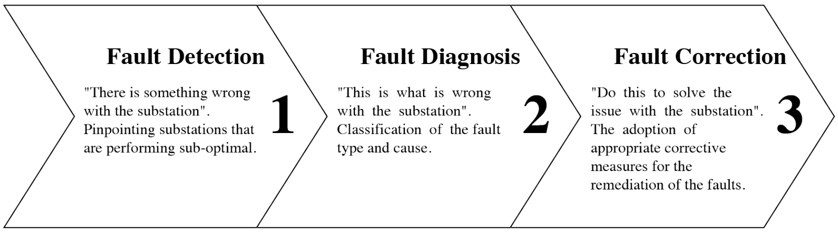

Figure 2: Three steps to automatic fault handling of DH building substations.

Most of the previous work, as discussed in Section 2, has mainly been in fault detection, e.g., see [5], [8], [15]. Our ambition in the planned research is to focus on developing intelligent solutions for automatic FDD of building substations. The results will assist DH utilities in maintaining, monitoring, and improving the DH systems. We aim to automatically diagnose faulty substations by developing a methodology for the analysis, evaluation, and classification of faults by using ML and AI techniques. Specifically, our research aims at:

- Collecting, validating, analyzing, and generating relevant labeled data to enable fault diagnosis and improve existing fault detection methods.
- Developing data-centric FDD models and algorithms.
- Evaluating and improving the FDD-developed models and algorithms.
- Implementing and validating the FDD-developed models and algorithms in real-world scenarios.

One of the main technical challenges will be the availability of labeled data [7]. While data used for billing purposes may be abundant, there seems to be a lack of known ground truth, e.g., labeled data of known faults. Another major technical challenge is the wide variety of substations, end-user heat delivery systems, and heat demand profiles, e.g., a hospital may operate 24/7 with a constant heat demand while an office building may only require heat during the working days. Also, heat meters measure all delivered heat for space heating (SH) and domestic hot water (DHW) preparation, adding noise to the data. Separating SH and DHW usage is not trivial.

To design and develop successful FDD in DH substations, we have separated the research into three phases: exploration, experimentation, and implementation.

The *exploration* phase will investigate the existing data and the DH field. We will conduct a systematic literature review (SLR) on the application of ML in FDD of DH building substations, perform explanatory data analysis (EDA) on existing real-world data, and conduct lab emulations to generate labeled data. Introducing faults manually into the substations may extend the dataset with ground truth information on faults. Additional sensors may be placed to gather more data to provide richer information on the equipment behavior to ensure higher classification performance. To increase the reliability of the lab experiments, domain experts will be involved in the whole process.

The *experimentation* phase will focus on testing the applicability of ML algorithms, e.g., tree algorithms, artificial neural networks (ANN), deep neural networks (DNN), or SVM, on lab emulated data. This will lead to additional knowledge and a better understanding of the data and challenges in DH substations. Additional techniques to evaluate are semi-supervised learning, which makes it possible to augment a small labeled dataset to a larger unlabeled dataset, as well as transfer learning [16] that trains a model on data from one source domain and then adapt the model to data in another target domain. It could be an excellent technique to generalize lab-generated data (source) toward a real-world system (target). In this phase, novel or adapted versions of the existing transfer learning and semi-supervised learning algorithms will be developed to address the discussed challenges.

Finally, the *implementation* phase will focus on testing the algorithms in real-world scenarios in close collaboration with industry and DH experts.

4 Conclusions and planned research

Automatic fault detection and diagnosis in district heating of building substations contains several challenges. Namely, the limitations of the data are the primary concern of the planned research. Labeled data is scarce, and the few parameters heat meters collect result in a dataset with very few parameters to study. Consequently, the current data may contain noise due to domestic hot water preparation aside from space heating in some types of substations. The wide variety of substations, end-user heat delivery systems, and heat demand profiles pose significant challenges to our study.

Our research will focus on lab emulations to counter the lack of labeled data and the shortcomings of simulated synthetic data. We plan to explore various traditional ML methods to study the behavior of emulated data, create benchmarks for further studies and gain additional knowledge about fault diagnosis tasks. Furthermore, novel or adapted versions of the existing transfer learning and semi-supervised learning algorithms will be developed to address the discussed challenges. Finally, we will evaluate and further optimize the developed ML algorithms in real-world scenarios. By closer cooperation between relevant industrial and academic experts, it could lead to valuable insights and further improvements in the developed algorithms.

References

[1] Luca Ferrari, Stefano Morgione, Dominik Rutz, et al. “A comprehensive framework for District Energy systems upgrade”. In: *Energy Reports* 7 (2021), pp. 359–367.

[2] *2050 long-term strategy*. en. https://ec.europa.eu/clima/eu-action/climate-strategies-targets/2050-long-term-strategy_en. Accessed: 2022-1-9.

[3] Sara Månsson. "Spot the difference!: On the way towards automated fault handling in district heating buildings". PhD thesis. Lund University, 2021.

[4] Luca Angelino, Jack Kiruja, Nis Bertelsen, et al. *Integrating low-temperature renewables in district energy systems: Guidelines for policy makers*. en. https://irena.org/publications/2021/March/Integrating-low-temperature-renewables-in-district-energy-systems/. Accessed: 2022-1-31.

[5] Sara Månsson, Kristin Davidsson, Patrick Lauenburg, et al. "Automated statistical methods for fault detection in district heating customer installations". In: *Energies* 12.1 (2019), p. 113.

[6] Henrik Gadd and Sven Werner. "Fault detection in district heating substations". In: *Applied Energy* 157 (2015), pp. 51–59.

[7] Gideon Mbiydzenyuy, Sławomir Nowaczyk, Håkan Knutsson, et al. "Opportunities for machine learning in district heating". In: *Applied Sciences* 11.13 (2021), p. 6112.

[8] Shahrooz Abghari, Veselka Boeva, Jens Brage, et al. "Higher order mining for monitoring district heating substations". In: *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2019, pp. 382–391.

[9] Henrik Gadd. "To analyse measurements is to know!" PhD thesis. Lund University, 2014.

[10] Gerrit Bode, Simon Thul, Marc Baranski, et al. "Real-world application of machine-learning-based fault detection trained with experimental data". In: *Energy* 198 (2020), p. 117323.

[11] Yang Zhao, Tingting Li, Xuejun Zhang, et al. "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future". In: *Renewable and Sustainable Energy Reviews* 109 (2019), pp. 85–101.

[12] Yalan Wang, Zhiwei Wang, Suowei He, et al. "A practical chiller fault diagnosis method based on discrete Bayesian network". In: *International Journal of Refrigeration* 102 (2019), pp. 159–167.

[13] Mengshi Li, Weimin Deng, Kaishun Xiahou, et al. "A data-driven method for fault detection and isolation of the integrated energy-based district heating system". In: *IEEE Access* 8 (2020), pp. 23787–23801.

[14] Yaguo Lei, Bin Yang, Xinwei Jiang, et al. "Applications of machine learning to machine fault diagnosis: A review and roadmap". In: *Mechanical Systems and Signal Processing* 138 (2020), p. 106587.

[15] Shahrooz Abghari, Veselka Boeva, Jens Brage, et al. "Multi-view Clustering Analyses for District Heating Substations". In: *Proceedings of the 9th International Conference on Data Science, Technology and Applications (DATA)*. Vol. 1. SciTePress. 2020, pp. 158–168.

[16] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, et al. "A comprehensive survey on transfer learning". In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.