Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (32nd cycle)

Machine learning techniques to forecast non-linear trends in smart environments

Alessandro Aliberti

* * * * * *

Supervisor
Prof. Macii Enrico

Doctoral Examination Committee:
Prof. Luciano Baresi, Referee, Politecnico di Milano
Prof. Marco Pau, Referee, RWTH Aachen University
Prof. Andrea Calimera, Politecnico di Torino
Prof. Silvia Chiusano, Politecnico di Torino
Prof. Ana-Maria Dumitrescu, Politehnica University of Bucharest

Politecnico di Torino
July 27, 2020

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .

Alessandro Aliberti
Turin, July 27, 2020

Summary

According to the World Health Organisation (WHO), air pollution is responsible for 7 million deaths every year, and 91% of the world population lives in places where air quality exceeds the limits mandated by WHO itself. In recent years, the research about energy waste and pollution reduction has gained a strong momentum, also pushed by European and national funding initiatives. The primary purpose of this large effort is to reduce the effects of greenhouse emission, climate change to head for a sustainable society. In this scenario, Information and Communication Technologies (ICT) play a key role in reducing energy consumption and moving forward to a more sustainable and smart society.

IoT devices are used in many contexts, to add smartness to cities, energy, and industrial processes. Their pervasiveness, combined with the recent development of machine learning techniques, allows collecting a large amount of data, enabling original opportunities to create innovative modelling and optimization approaches. Consequently, moving towards smart and sustainable energy use, my research activities have mainly focused on the design and the optimization of innovative machine learning methodologies, by exploiting primarily neural networks, for the forecasting of time-series in Smart City context. In this manuscript, I propose innovative and optimized stream data processing and machine learning methodologies, ranging from energy and environmental data and moving to data from CPS systems. I have designed and validated innovative modelling and control strategies in specific application case studies: i) Renewables, ii) Smart Building and iii) Smart Health.

In detail, I addressed the issues of prediction of GHI which is the energy component necessary for the development of photovoltaic energy, the thermal modelling of Smart Buildings and finally, I moved toward the person health by addressing the topic of blood glucose level prediction for Type I diabetic patients. In all the three cases, the studies were conducted following a bottom-up approach starting from the analysis and appropriate pre-processing of IoT data, designing neural models suitable for the type of dataset and its characteristics and finally comparing these new models with the methodologies of the literature.

In the context of renewables, I developed a methodology for photovoltaic predictions starting from the physical phenomenon of GHI. Then, I focused on the optimization of the methods for GHI forecasting in short- and mid-term. Lastly,

I also investigated how to properly exploit exogenous inputs (i.e. physical factors related to the phenomenon of solar radiation) to further improve GHI predictions. In the context of Smart Buildings, I developed a comprehensive methodology that enables thermal modelling in both new generation and historic buildings, by exploiting the possibility of creating a very reliable synthetic dataset based on BIM technology and real weather data (TMY). This configuration allowed me to develop hybrid neural models trained with synthetic data and able to exploit real data (i.e. provided by IoT devices installed in a real-world demonstrator) for indoor air-temperature predictions. Finally, by using Transfer Learning techniques, I was able to specialize the hybrid models on real data. Lastly, in the context of smart health, I addressed the problem of automated glucose level prediction leveraging multi-patient CGMS data. The aim was to learn a generalizable glucose level prediction model from a multi-patient training set, using this model to predict the future glucose values of a new patient. In practice, the objective is to create a device that can be purchased and is ready to use, without the need for initial tuning. Besides, I started to evaluate techniques to specialize this methodology by integrating real-time information to specialize the predictor system specifically on the single end-user.

This Ph.D. thesis has been typeset by means of the TeX-system facilities. The typesetting engine was pdfLaTeX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TeX-system installation.