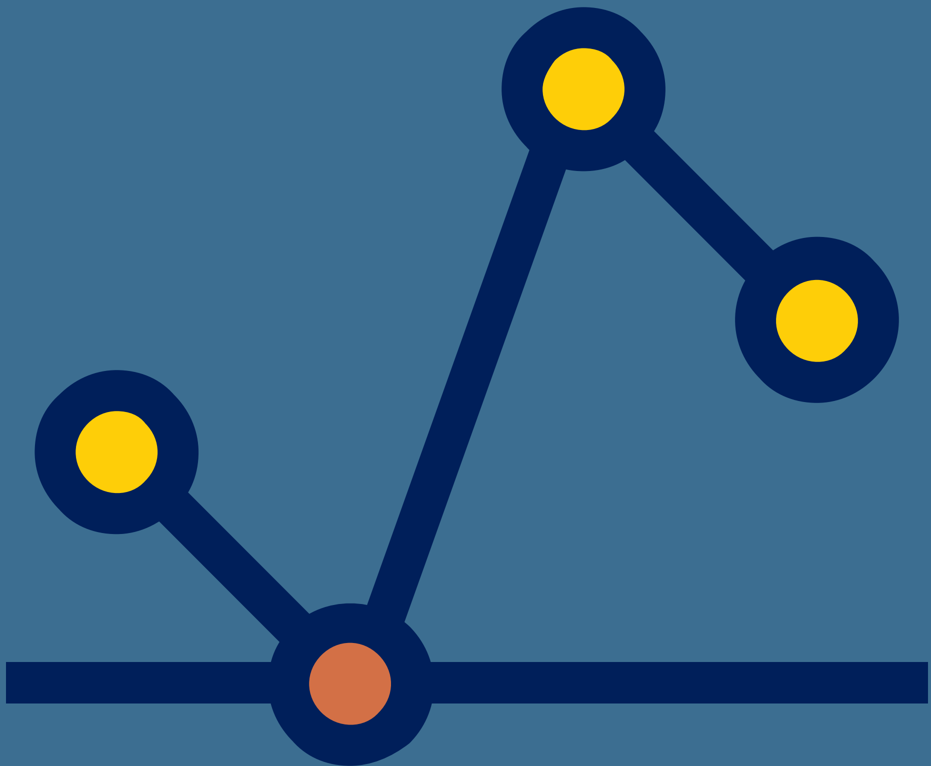
Edited by

Paola Cerchiello · Arianna Agosto
Silvia Osmetti · Alessandro Spelta

Proceedings of the Statistics and Data Science Conference



PaviaUniversityPress

Copertina: Cristina Bernasconi, Milano

Copyright © 2023 EGEA S.p.A.
Via Salasco, 5 - 20136 Milano
Tel. 02/5836.5751 - Fax 02/5836.5753
egea.edizioni@unibocconi.it - www.egeaeditore.it

Quest'opera è rilasciata nei termini della Creative Commons Attribution 4.0 International Licence (CC BY-NC-SA 4.0), eccetto dove diversamente indicato, che impone l'attribuzione della paternità dell'opera e ne esclude l'utilizzo a scopi commerciali. Sono consentite le opere derivate purché si applichi una licenza identica all'originale. Il testo completo è disponibile alla pagina web https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it.

Date le caratteristiche di Internet, l'Editore non è responsabile per eventuali variazioni di indirizzi e contenuti dei siti Internet menzionati.

Pavia University Press
info@paviauniversitypress.it – www.paviauniversitypress.it

Prima edizione: maggio 2023
ISBN volume 978-88-6952-170-6

Preface

The development of large-scale data analysis and statistical learning methods for data science is gaining more and more interest, not only among statisticians, but also among computer scientists, mathematicians, computational physicists, economists, and, in general, all experts in different fields of knowledge who are interested in extracting insight from data.
Cross-fertilization between the different scientific communities is becoming crucial for progressing and developing new methods and tools in data science.
In this respect, the Statistics & Data Science group of the Italian Statistical Society has organized an international conference held in Pavia on the 27 and 28 of April 2023, attended by over 70 researchers from different scientific fields.
A collection of the presented papers is available in the present Proceedings showing a huge variety of approaches, methods, and data-driven problems, always tackled according to a rigorous and robust scientific paradigm.

The Statistics & Data Science group

Contents

The structural behavior of Santa Maria del Fiore Dome: an analysis with machine learning techniques

Il comportamento strutturale della Cupola di Santa Maria del Fiore: un'analisi con tecniche di machine learning

Stefano Masini and Silvia Bacci and Fabrizio Cipollini and Bruno Bertaccini

Abstract The Brunelleschi's Dome overlooking the cathedral of Santa Maria del Fiore in Florence is a symbol of the Italian Renaissance. Because of the presence of numerous cracks distributed on its entire surface, the Dome is subjected to a continuous monitoring activity that relies, among others, on electronic sensors, mainly deformometers, to measure the movements of the cracks, and thermometers, to measure the masonry temperatures. These instruments are active since more than 30 years and take measures more times a day, thus producing a huge amount of data. In this contribution, we aim at applying some machine learning techniques (i) to describe the overall movement of Dome surface through a suitable synthesis of the measures of the sensors and (ii) to make medium- and long-term predictions about the evolution of the Dome.

Abstract *La Cupola del Brunelleschi sovrastante la cattedrale di Santa Maria del Fiore a Firenze è un simbolo del Rinascimento italiano. A causa della presenza di numerose crepe distribuite sull'intera superficie, la Cupola è sottoposta a una continua attività di monitoraggio che si basa, tra gli altri, su sensori elettronici, principalmente deformometri per misurare i movimenti delle crepe e termometri per misurare la temperatura dei muri. Questi strumenti sono attivi da oltre 30 anni e rilevano le misure più volte al giorno, producendo così un'enorme mole di dati.*

Stefano Masini
Dept. of Computer Science, University of Pisa, Largo B. Pontecorvo 3, I-56127 Pisa, e-mail: stefano.masini@unifi.it

Silvia Bacci
Dept. of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Viale Morgagni 59, I-50134 Firenze e-mail: silvia.bacci@unifi.it

Fabrizio Cipollini
Dept. of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Viale Morgagni 59, I-50134 Firenze e-mail: fabrizio.cipollini@unifi.it

Bruno Bertaccini
Dept. of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Viale Morgagni 59, I-50134 Firenze e-mail: bruno.bertaccini@unifi.it

In questo contributo, il nostro scopo è l'applicazione di alcune tecniche di machine learning per (i) descrivere il movimento complessivo della Cupola tramite un'opportuna sintesi delle misure dei sensori e (ii) fare previsioni a medio e lungo termini riguardo all'evoluzione della Cupola.

Key words: Artificial Intelligence, Cultural heritage preservation, Dimensionality reduction techniques, Forecasting, Multivariate time series data, Sensor data

1 Introduction

The cathedral of Santa Maria del Fiore in Florence (IT) with its Dome is one of the most famous buildings of the Italian Reinassance. The Dome was built by Filippo Brunelleschi in the period 1420-1436 adopting a special technique (with bricks disposed as an "herringbone pattern") that allowed setting up the construction site without shoring. The result was impressive: nowadays, the Dome is still of the largest masonry domes in the world, weighing more than 43,000 tons. Unfortunately, from the beginning some cracks appeared on the surface of the Dome, thus the building has always been subject to careful monitoring.

The monitoring system of Brunelleschi's Dome is made up of a multiplicity of instruments, such as piezometers, plumb lines, tele-coordinometers, thermometers, and mechanical and electronic deformometers. In particular, in 1987 were installed several electronic deformometers devoted to measuring the movements of the single cracks at least four times a day. Thus, a huge amount of data has been accumulated since the late of 1980s. The complex nature of relations among variables (mainly, movements of cracks and seasonal and daily changes of the masonry temperatures) together with the limits of the computational resources and competencies available in the scientific community have meant that to date these data have not yet been subjected to a systematic study. Indeed, the analyses carried out in previous works usually focused on a single device or a limited set of them [1, 4, 2]; a more recent work [3] took into account the entire set of electronic deformometers, but limited to a one-year period.

In this contribution, we aim at applying some machine learning techniques (i) to describe the overall movement of Dome surface through a suitable synthesis of the measures of the sensors and (ii) to make medium- and long-term predictions about the evolution of the Dome.

Section 2 provides some more details on data, Section 3 describes the machine learning methods used in the analysis, Section 4 illustrates some preliminary results, and Section 5 concludes with some final remarks.

2 Data

In the following we focus on data coming from the 57 electronic deformometers. A deformometer is a sensor installed on the walls across a crack to measure the changes of its width: at installation the instrument is set on value 0, so that positive measures denote a dilatation of the masonry structure and, then, a shrink of the crack, while negative measures refer to a contraction of the walls and, then, a widening of the crack. Deformometers are allocated on the entire surface of the Dome, with a major concentration on those sectors where there is a major presence of cracks. Here we consider the measurements of the complete set of 57 deformometers collected from 1997 to 2017.

Together with the measures of the deformometers, we also take into account the measures of the 47 masonry thermometers installed upon the Dome, as previous studies [1, 4, 2, 3] outlined a strong association between temperatures and movements of the cracks.

To account for gaps and outliers present in the data due to blackouts that periodically put electronic sensors out of action, producing anomalous oscillations, full scale values, or missing observations, we have to pre-treat data. For this aim, we followed the approach proposed in [2], based on the estimation of a quadratic-sinusoidal regression model per each sensor, thus obtaining a complete data matrix.

3 Methods

The first part of the analysis aims at synthesizing the measures of the entire set of sensors to describe the overall behavior of the Dome (and not of its single cracks). This typical problem of dimensionality reduction is addressed through the Kernel Principal Component Analysis (KPCA) [5].

Compared to traditional PCA, which combines observations in a linear way, KPCA allows us to make a non-linear projection of the observations preserving the relative distances between data points. In KPCA we use a function (kernel) to map the data from the original space in a new high-dimensional features space in order to verify whether, in the new space, the data are linearly separable. The algorithm requires to set the kernel type (linear, polynomial, gaussian rbf or sigmoid) and the gamma parameter (which is a space regularization parameter). In order to find their best combination, we use ScikitLearn's GridSearchCV with cross-validation function and, since KPCA is an unsupervised learning algorithm, we use the distance between the original point and the pre-image calculated on the new high-dimensional feature space as reconstruction error.

The principal components resulting from the application of the KPCA as well as the series of masonry temperatures are then used as inputs in a subsequent analysis aimed at providing predictions of the movements of the Dome at medium- and long-term. For this aim, we exploit the performance of some recurrent and convolutional neural network models [6], typically adopted for the prediction of multivariate time series data.

The above neural networks are used to make predictions of a certain number of steps (days) in the future (multiple-step forecasting). The model is trained using a sliding window of consecutive days (the further the future is, the wider the window), with the mean squared error as loss function and the mean absolute error as metric. The best performing model is a network composed of an initial convolutional layer with 40 (6x6) convolutional filters, followed by a bidirectional layer [8] with 20 Gated Recurrent Units (GRU) [7] and 2 more consecutive hidden layers.

4 Results

The results of the KPCA executed on the entire series of measures are displayed in Figure 1, where the first two principal components are plotted with points related to observations differently coloured according to the season.

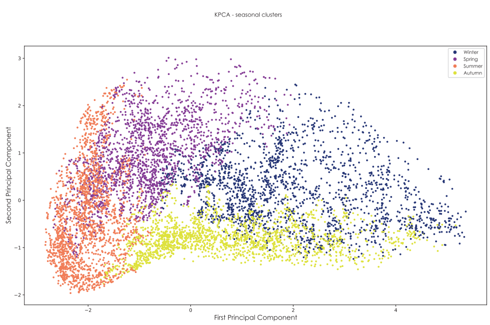


Fig. 1 Results of KPCA (best model: {'gamma': 0.1, 'kernel': 'poly'}): seasonal clustering of sensor data

The relation between observed cracks and seasonality emerges clearly from the figure. Namely, clusters of points relating to winter and summer seasons are well separated.

In light of these results, we execute again the KPCA on separate sets of observations, according to the location of the deformometers. We distinguish the deformometers into eight groups, corresponding to the eight slice webs that characterize the surface of the Dome, easily distinguishable with the naked eye thanks to the white marble cords. For the sake of clarity, the webs are numbered counterclockwise starting from the web that faces the nave (see [2] for the planning of the Dome and its webs). Figure 2 shows the trend of the first principal components for each web (top panel: odd webs; bottom panel: even webs), together with the trend of the daily average masonry temperatures (central panel). Note that the figure refers to a one-year time window, but the trend repeats with the same pattern throughout the entire period of observation (i.e., 1997-2017).

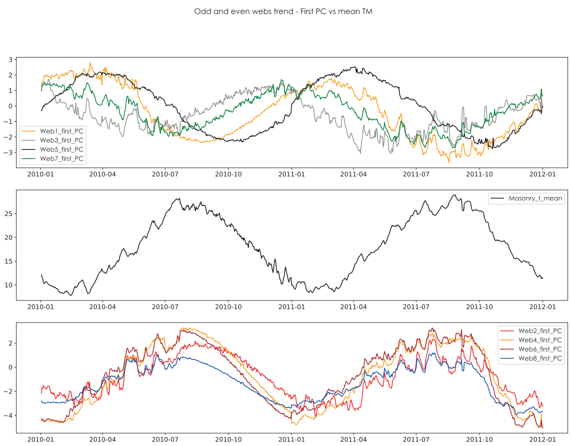


Fig. 2 First principal component of each web (top panel: odd webs, bottom panel: even webs) along a one-year window (January 1st, 2014 to January 1st, 2015)

Looking at Figure 2, we observe that movements of all webs follow a sinusoidal trend according to the temperature, with odd webs that move in the opposite direction with respect to even webs. These results provide evidence for a breathing mechanism of the entire Dome: when even webs shrink, odd webs widen, and vice-versa.

Finally, the first principal components obtained for each web through the KPCA are used in input in a neural network to make predictions. Figure 3 shows the next 100 days prediction results for web 2 with a window size equals to 300 ; results for other webs are similar.

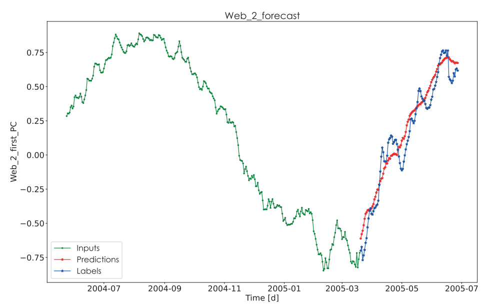


Fig. 3 Forecasting of web 2 trend (window size: 300 days, steps forward: 100 days)

5 Conclusions

The application of the machine learning techniques described in the contribution allowed us to achieve two important goals. First, we demonstrated the close correlation between the temperature and the behavior of the Dome over time, bringing to outlining a symmetry in the movements of the webs in even position and those in odd position. Second, we trained and tested some recurrent neural networks in order to predict the behavior of each web.

For the future, we will prosecute the work along the following main research lines. First, further variables will be taken into account, such as humidity, wind, solar exposition, and seismic measurements. Second, a software will be integrated in the current monitoring system to build a sort of "alarm system" in real-time.

Acknowledgements Authors thank the "Opera del Duomo Foundation" for having making available the data

References

1. Bartoli, G., Chiarugi, A., Gusella, V.: Monitoring systems on historic buildings: Brunelleschi Dome. Journal of Structural Engineering **122**, 663–673 (1996) doi: 10.1061/(ASCE)0733-9445(1996)122:6(663)
2. Bertaccini, B.: Santa Maria del Fiore Dome behavior: Statistical models for monitoring stability. International Journal of Architectural Heritage **9**, 25–37 (2015) doi: 10.1080/15583058.2013.774071
3. Bertaccini, B., Bacci, S., Crescenzi, F.: A Dynamic latent Variable Model for Monitoring the Santa Maria del Fiore Dome Behavior. In: Broy, M., Dener, E. (eds.) Lecture Notes in Computer Science - Computational Science and Its Applications ICCSA 2020, pp. 47-58. Springer professional (2020) doi: 10.1007/978-3-030-58811-3_4
4. Ottoni, F., Blasi, C., Coisson, E.: The crack pattern in Brunelleschi?s Dome in Florence: Damage evolution from historical to modern monitoring system analysis. Advanced Materials Research **133-134**, 53–64 (2010) doi: 10. 4028/www.scientific.net/AMR.133-134.53
5. Schölkopf B., Smola A., Müller K.: Kernel principal component analysis.Springer-Verlag Berlin Heidelberg 1997.; 7th International Conference on Artificial Neural Networks, ICANN 1997; doi: 10.1007/bfb0020217, isbn: 3540636315
6. Hochreiter S., Schmidhuber J. Long Short-Term Memory 1997. Neural Computation, nr.8, vol. 9, pp.1735-1780, doi: 10.1162/1997.9.8.1735
7. Cho K., Van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, 2014, doi: 10.48550/ARXIV.1406.1078
8. Schuster M., Paliwal, Kuldip K. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing. November 1997. Vol. 45, pp 2673–2681, biburl: https://www.bibsonomy.org/bibtex/26026f083db110838a6b62c2eedfec9e9/nilsd