Abstract: Layered Integration Approach for Multi-view Analysis of Temporal Data⋆ ⋆⋆

Michiel Dhont[1,2][0000−0002−5679−0991], Elena Tsiporkova[1], and Veselka Boeva[3][0000−0003−3128−191X]

[1] Sirris, Bd A. Reyerslaan 80, 1030 Brussels, Belgium
[2] Department of Electronics and Information Processing (ETRO), VUB, Belgium
[3] Blekinge Institute of Technology, Sweden
{michiel.dhont,elena.tsiporkova}@sirris.be, veselka.boeva@bth.se

Keywords: Data Integration · Data Mining · Temporal Data Clustering · Multi-view Learning.

1 Introduction

Mining data collected from continuous monitoring of industrial assets in the field allows to derive relevant insights about their operations and performance. Such complex real-world datasets are usually composed of heterogeneous subsets (or multi-views) of parameters, which should be considered explicitly during analysis in order to exploit fully the richness of the data. For instance, the performance of an industrial asset is impacted by a diverse set of factors e.g. operating modes concerned with the internal working of the asset and exogeneous factors such as weather conditions. However, it is not trivial to directly link or trace back certain performance to distinct operating modes due to the multitude of influencing factors, which are often also highly interdependent.

In addition, real-world datasets often originate from different sources, which may differ in period coverage, resolution, data quality, technical configuration, etc. Pooling multi-source datasets together, which is often done to increase statistical representativeness, requires standardization and normalization, which often leads to information loss and may mask source-specific features. For instance, mining for distinct operating modes is more appropriate to be pursued per asset, rather than pooling everything together, since not all assets might go through all operating modes. This implies that one might need to approach multi-source analysis in an incremental fashion rather than aiming for brute force integration of all the available data.

2 Proposed Approach

Classical data mining and analysis approaches have still some shortcomings in this aspect aiming at delivering a total integration solution at once. An alter-

⋆ This research was supported by the Brussels-Capital Region - Innoviris, and received funding from the Flemish Government (AI Research Program) and BitWind project.
⋆⋆ The full paper is accepted for the 5^{th} AALTD workshop and publication in LNAI.

native approach is to **exploit the multi-view nature of the data**. Some rewarding techniques of multi-view mining have been already proposed in the literature [1, 2]. However, they all were concerned with single-source datasets and dedicated to one specific mining approach (e.g. clustering or deep learning). This research provides a general analysis methodology, which focuses on the following key aspects: initial *individual analysis* per source in order to preserve the richness and the authenticity of each source; individual *mediation analysis* per source aiming at bringing the sources closer together; cross-source *integration analysis* aiming at leveraging analysis results across the sources without compromising their individual characteristics.

More concretely, the proposed approach consists of several distinctive layers: (i) select a suitable set (view) of parameters in order to identify characteristic behaviour within each individual source (ii) exploit an alternative set (view) of raw parameters (or high-level features) to derive some complementary representations (e.g. related to source performance) of the results obtained in the first layer with the aim to facilitate comparison and mediation across the different sources (iii) integrate those representations in an appropriate way, allowing to trace back similar cross-source performance to certain characteristic behaviour of the individual sources.

3 Implementation and Results

The validity and the potential of the proposed approach have been demonstrated on a real-world dataset of a fleet of wind turbines. We have been able to identify distinctive profiles of production performance and subsequently, have been able to establish an explicit link between those performance profiles and well characterised operating modes. Subsequently, distinctive performance profiles have been derived and associated with each operating mode, which enable converting the fleet data into powerful letter code suitable for more advanced mining.

4 Conclusion

We have proposed a novel data analysis approach that can be used for multi-view analysis and integration of heterogeneous real-world datasets originating from multiple sources. The validity and the potential of the proposed approach has been demonstrated on a real-world dataset of a fleet of wind turbines. The obtained results are very encouraging. The method is very efficient and robust in detecting characteristic operating modes across the fleet.

References

1. Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *ICDM*, volume 4, pages 19–26, 2004.
2. Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.