Paulo Moura, William Liaw

Analyzing and Extending Time Series Kernels based on Nonlinear Vector AutoRegressive Delay Embeddings

This report is based on the paper (FELICE; GOULERMAS; GUSEV, 2023), published at NeurIPS 2023.

Palaiseau
2024

Paulo Moura, William Liaw

Analyzing and Extending Time Series Kernels based on Nonlinear Vector AutoRegressive Delay Embeddings

This report is based on the paper (FELICE; GOULERMAS; GUSEV, 2023), published at NeurIPS 2023.

Advisor: Prof. Florence D'Alché

Palaiseau
2024

ABSTRACT

In this work we study the...

Keywords: KEYWORD. KEYWORD.

CONTENTS

1 Introduction **1**

2 Analysis **3**

2.1 Novelty and Contributions 3

2.2 Methodology 4

2.3 Comparison to State-of-the-Art 5

2.4 Strengths 5

2.5 Weaknesses 6

3 Methodology **7**

3.1 Experimental setup 7

3.1.1 Datasets 7

3.1.1.1 Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) 7

3.1.1.2 Berlin Database of Emotional Speech 8

3.2 Evaluation Metrics 9

4 Results **10**

5 Conclusion **11**

References **12**

1 INTRODUCTION

In recent years, kernel methods have become a cornerstone of machine learning, especially for tasks involving non-linear data structures. Kernels enable linear algorithms to operate in transformed feature spaces, allowing for the efficient handling of complex relationships within data. Time-series data, with its inherent sequential dependencies, is a particularly challenging domain for machine learning, demanding specialized methods to capture temporal patterns and underlying dynamics, and arguably one of the most important data types in the moders era (HAMILTON, 1994; STROGATZ, 2018; ZHANG, 2017; ZEROUAL et al., 2020). As traditional kernel approaches often struggle with the complexity of time-series data, recent research has explored innovative kernel designs tailored specifically for this data type.

The selected paper, “Time Series Kernels based on Nonlinear Vector AutoRegressive Delay Embeddings” (FELICE; GOULERMAS; GUSEV, 2023), addresses a major challenge in time-series kernel design by introducing a new kernel based on Nonlinear Vector AutoRegressive (NVAR) delay embeddings. The proposed NVAR kernel draws on the principles of reservoir computing (RC), adapting them into a more interpretable and computationally efficient framework (BOLLT, 2021). Unlike standard RC-based kernels, which rely on recurrent structures and complex hyperparameter tuning, the NVAR kernel leverages non-recursive embeddings. This approach reduces the dependency on recurrent hyperparameters, making it better suited for classification tasks with small datasets, where deep learning techniques may not be feasible.

This project aims to conduct a comprehensive analysis of the NVAR kernel’s methodology, positioning, and effectiveness in the broader landscape of kernel machines. By comparing the NVAR kernel with benchmarks in a select few datasets, we seek to evaluate its advantages and limitations in terms of both accuracy and computational efficiency. Furthermore, this study involves implementing and testing the NVAR kernel on a new dataset to assess its practical applicability and performance in real-world scenarios.

This report is structured as follows: First, we present an in-depth analysis of the

NVAR kernel and its unique contributions to time-series classification. We then describe the methodology, experimental setup, followed by the results of testing the kernel on distinct datasets and a critical evaluation of the method's strengths and weaknesses. Finally, we discuss potential future directions for kernel design in time-series analysis and conclude with key insights drawn from this study.

2 ANALYSIS

2.1 Novelty and Contributions

The paper introduces a novel approach to time-series kernel design using NVAR. This method stands out for its integration of kernel design techniques, RC principles and the NVAR framework, a combination that enhances both interpretability and efficiency in time-series data analysis. Traditional RC-based kernels rely heavily on recurrent structures that are complex and sensitive to hyperparameter tuning, often demanding high computational resources to achieve optimal performance. In contrast, the NVAR kernel circumvents the need for recurrence by structuring embeddings as non-recursive transformations of the input data: time-delays and nonlinear functionals, such as products (GAUTHIER et al., 2021).

This approach not only simplifies the model architecture but also reduces the dependency on interpretatively opaque hyperparameters, making it easier to apply and understand. Thus, it represents a significant advancement in kernel design by developing an NVAR-based kernel suitable for both univariate (UTS) and multivariate time-series (MTS) data, with parameter settings guided by simple heuristics. Experiments across diverse datasets show that this NVAR kernel achieves accuracy comparable to the state-of-the-art (SOTA) in time-series classification while offering a substantial improvement in computational efficiency. This balance between accuracy and speed highlights its practicality for real-world applications, especially where computational resources are limited.

From the perspective of RC, the NVAR kernel introduces a non-recursive model that circumvents the challenges of hyperparameter optimization typically faced with RC-based kernels. By forgoing recurrence, the NVAR kernel remains interpretable and simplifies the model architecture without sacrificing the quality of representation. This change not only enhances interpretability but also positions the NVAR kernel as a more accessible tool for practitioners who might otherwise face the complexities and resource demands of RC.

The paper also expands the use of NVAR beyond its traditional role in forecasting

chaotic, noise-free systems, demonstrating its applicability to real-world time-series data. By connecting the method to foundational principles in dynamical systems theory, including Takens' theorem and state-space reconstruction, the paper provides a theoretical basis for understanding the embeddings used in the NVAR kernel. This connection to established theory underscores the method's robustness and supports its potential as a versatile tool in machine learning for tasks requiring sophisticated temporal representations.

2.2 Methodology

The methodology of the NVAR kernel is structured around transforming time-series data using a series of lagged embeddings and nonlinear transformations, creating a high-dimensional representation that effectively captures the underlying temporal dynamics. Specifically, the kernel leverages delay embeddings, a technique rooted in dynamical systems theory, to form feature vectors that encapsulate the past states of the time-series data. These feature vectors are then mapped to a high-dimensional space, where a similarity measure is calculated to define the kernel.

This embedding process follows Takens' theorem, which suggests that a time-delayed version of a series can reveal its latent dynamics. In this framework, each time series is enriched with delayed copies and nonlinear combinations of the input, allowing the NVAR kernel to uncover complex patterns and dependencies that traditional kernels might overlook. The paper further simplifies the hyperparameter space by employing a heuristic-based approach for setting the lag length and polynomial order, avoiding the computational burden typically associated with exhaustive tuning.

The NVAR kernel also builds on the kernel trick, which allows the computation of inner products in a transformed feature space without explicitly computing the transformation. By integrating the NVAR embedding structure into the kernel trick, the paper effectively combines the advantages of kernel-based methods with those of RC, resulting in a versatile and efficient approach to time-series analysis.

2.3 Comparison to State-of-the-Art

The proposed NVAR kernel distinguishes itself from existing time-series similarity measures and kernel methods, particularly those based on RC. Traditional RC-based kernels, such as Echo State Networks (ESNs), rely on recurrent structures that introduce complexity in hyperparameter optimization and high sensitivity to initial conditions. These recurrent models often struggle with interpretability, as their performance heavily depends on tuning multiple hyperparameters related to the reservoir's size, connectivity, and spectral properties.

In comparison, the NVAR kernel simplifies the representation by using a non-recursive structure, significantly reducing the number of hyperparameters that require fine-tuning. Unlike elastic measures, which directly measure similarity in the input space and can overlook deeper temporal dependencies, the NVAR kernel captures underlying dynamics through delay embeddings, improving its ability to handle time distortions and shifts. Additionally, while model-based kernels such as the Time Cluster Kernel (TCK) can achieve high accuracy, they tend to be computationally intensive and may not scale well with larger datasets. The NVAR kernel, in contrast, offers a compromise between accuracy and computational efficiency, making it well-suited for applications where computational resources are limited.

2.4 Strengths

The strengths of the NVAR kernel are evident in its computational efficiency, scalability, and suitability for small datasets. The non-recursive nature of the kernel design eliminates the iterative computations required in recurrent models, allowing it to scale linearly with the number of time series and reducing overall computational cost. This efficiency makes the NVAR kernel particularly valuable in contexts where fast processing is essential, such as real-time monitoring or applications with limited computational resources.

Moreover, the simplicity of the NVAR kernel's hyperparameters contributes to its scalability and ease of use. By relying on a heuristic for setting lag and polynomial order, the kernel can be quickly adapted to a variety of datasets without extensive

optimization, a feature that is especially useful when working with small datasets where overfitting risks are higher. The interpretability of the NVAR approach also adds to its strengths, as the non-recursive embeddings make it easier to understand and analyze the extracted temporal features compared to traditional RC methods.

2.5 Weaknesses

Despite its advantages, the NVAR kernel has some limitations. One potential drawback is its reliance on heuristic-based hyperparameter settings, which may not always yield optimal results, particularly for highly complex datasets or data with irregular temporal patterns. While the heuristics provide a practical solution, they may oversimplify the selection of critical parameters such as lag size, leading to suboptimal embeddings in certain cases.

Another limitation lies in the method's performance with high-dimensional datasets. As the number of dimensions in the input data increases, the kernel may face challenges due to the curse of dimensionality, potentially requiring adjustments to its feature selection strategy. Additionally, the NVAR kernel's dependence on delay embeddings may restrict its applicability in situations where the time-series data lacks well-defined temporal dependencies or exhibits chaotic behavior that is difficult to model with fixed embeddings.

Overall, while the NVAR kernel offers significant improvements over traditional RC-based kernels, further work could be done to refine its hyperparameter optimization process and to extend its applicability to a broader range of complex, high-dimensional datasets.

3 METHODOLOGY

3.1 Experimental setup

3.1.1 Datasets

Selecting appropriate datasets is pivotal in the development, training, and evaluation of emotion recognition models. In this study, we employ two widely recognized datasets commonly used in speech and audiovisual emotion classification: the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Berlin Database of Emotional Speech (Emo-DB). Both datasets feature emotionally expressive performances by professional actors, providing high-quality labeled data that is indispensable for benchmarking and enhancing the capabilities of emotion classification systems. The following subsections provide detailed descriptions of each dataset, highlighting their key features, structures, and the rationale for their use in this study. For the purposes of this study, all audio data from the datasets were downsampled to a sampling rate of 8 kHz.

3.1.1.1 Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) (LIVINGSTONE; RUSSO, 2018) is a widely recognized benchmark dataset utilized extensively in audiovisual emotion classification research (ANUSHA et al., 2021; VIMAL et al., 2021; ABDULLAH; AHMAD; HAN, 2020). The dataset consists of short audio and video recordings that feature both spoken and sung performances, enacted by a cohort of 24 actors (12 male and 12 female). Each recording is labeled with one of the following emotion categories: *angry*, *calm*, *disgust*, *fearful*, *happy*, *neutral*, *sad*, and *surprised*.

To promote consistency and reproducibility, each actor delivers two predefined phrases in English: “Kids are talking by the door” and “Dogs are sitting by the door.” Apart from the neutral category, all emotions are expressed at two distinct intensity levels (normal and strong), with each instance repeated twice. These structured variations

in emotional intensity, repetition, and diversity of vocal expressions make RAVDESS an invaluable asset for the development and validation of emotion recognition models in a wide range of applications.

For the audio-only subset of the dataset which we employ for further analysis in the present work, there are a total of 1440 speech recordings and 1012 song recordings. It is worth noting that the singing subset is slightly smaller, as one actor's data is missing, and the emotions *sad* and *surprised* are not included for singing performances.

Despite its favorable reception within the academic community, as demonstrated by its widespread adoption, evidence suggests that the application of RAVDESS in real-world scenarios may lead to underwhelming results (CHURAEV; SAVCHENKO, 2021). One possible explanation for this discrepancy is the issue of data leakage. Specifically, an overlap of similar samples between the training and validation sets may result in unintended information sharing, thereby artificially inflating performance metrics. This overestimation does not accurately reflect the generalizability and practical effectiveness of models when deployed in real-world environments.

3.1.1.2 Berlin Database of Emotional Speech

The Berlin Database of Emotional Speech (Emo-DB) (BURKHARDT et al., 2005), akin to RAVDESS, is a well-regarded dataset for speech emotion classification tasks (SINITH et al., 2015; KOTTI; KOTROPOULOS, 2008; YING; ZHANG, 2010). It comprises short spoken audio recordings performed by 10 professional actors (5 male and 5 female), each enacting various grammatical phrases in German, as detailed in Table 1. Each recording is annotated with one of the following emotion categories: *anger*, *anxiety/fear*, *boredom*, *disgust*, *happiness*, *neutral*, and *sadness*.

To ensure the quality and reliability of the dataset, these samples underwent evaluation by a significant number of listeners, who assessed the naturalness of the emotional expressions. In total, the dataset comprises 535 speech files.

Table 1: Grammatical phrases in the Emo-DB dataset

German	English
Der Lappen liegt auf dem Eisschrank.	The cloth is on the refrigerator.
Das will sie am Mittwoch abgeben.	She will deliver it on Wednesday.
Heute abend könnte ich es ihm sagen.	Tonight I could tell him.
Das schwarze Stück Papier befindet sich da oben neben dem Holzstück.	The black sheet of paper is located up there next to the piece of wood.
In sieben Stunden wird es soweit sein.	In seven hours it will be time.
Was sind denn das für Tüten, die da unter dem Tisch stehen?	What about the bags that are under the table?
Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter.	They just carried it upstairs and now they are going back down.
An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht.	On weekends, I now always went home and visited Agnes.
Ich will das eben wegbringen und dann mit Karl was trinken gehen.	I will just take this away and then go have a drink with Karl.
Die wird auf dem Platz sein, wo wir sie immer hinlegen.	It will be in the place where we always put it.

Source: Own authorship

3.2 Evaluation Metrics

Describe the metrics you used to evaluate performance (e.g., classification accuracy, execution time, etc.).

4 RESULTS

Present the results of your experiments, including any tables or graphs that compare the performance of the NVAR kernel with other methods.

Highlight the performance differences between the NVAR kernel and existing kernels (e.g., reservoir-based kernels or elastic measures like Dynamic Time Warping).

Interpret the results: Did the NVAR kernel perform as expected? How did it compare to other kernels in terms of accuracy, efficiency, and scalability?

Discuss any insights gained from the experiments, such as the effect of hyperparameter settings (e.g., number of lags, polynomial order).

5 CONCLUSION

In this work we have considered ...

REFERENCES

ABDULLAH, M.; AHMAD, M.; HAN, D. Facial expression recognition in videos: An cnn-lstm based model for video classification. In: **2020 International Conference on Electronics, Information, and Communication (ICEIC)**. [S.l.: s.n.], 2020. p. 1–3.

ANUSHA, R. et al. Speech emotion recognition using machine learning. In: **2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)**. [S.l.: s.n.], 2021. p. 1608–1612.

BOLLT, E. On explaining the surprising success of reservoir computing forecaster of chaos? the universal machine learning dynamical system with contrast to var and dmd. **Chaos an Interdisciplinary Journal of Nonlinear Science**, v. 31, n. 1, jan. 2021. Disponível em: <https://pubs.aip.org/aip/cha/article/31/1/013108/341924/On-explaining-the-surprising-success-of-reservoir>.

BURKHARDT, F. et al. A database of German emotional speech. In: . [S.l.: s.n.], 2005. v. 5, p. 1517–1520.

CHURAEV, E.; SAVCHENKO, A. V. Touching the limits of a dataset in video-based facial expression recognition. In: **2021 International Russian Automation Conference (RusAutoCon)**. [S.l.: s.n.], 2021. p. 633–638.

FELICE, G. D.; GOULERMAS, J. Y.; GUSEV, V. Time series kernels based on nonlinear vector autoregressive delay embeddings. In: **Thirty-seventh Conference on Neural Information Processing Systems**. [s.n.], 2023. Disponível em: <https://openreview.net/forum?id=UBUWFEwn7p>.

GAUTHIER, D. J. et al. Next generation reservoir computing. **CoRR**, abs/2106.07688, 2021. Disponível em: <https://arxiv.org/abs/2106.07688>.

HAMILTON, J. D. **Time Series analysis**. [s.n.], 1994. Disponível em: <https://doi.org/10.1515/9780691218632>.

KOTTI, M.; KOTROPOULOS, C. Gender classification in two emotional speech databases. In: **2008 19th International Conference on Pattern Recognition**. [S.l.: s.n.], 2008. p. 1–4.

LIVINGSTONE, S. R.; RUSSO, F. A. The Ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in North American English. **PLOS ONE**, Public Library of Science, v. 13, n. 5, p. 1–35, 05 2018. Disponível em: <https://doi.org/10.1371/journal.pone.0196391>.

SINITH, M. S. et al. Emotion recognition from audio signals using support vector machine. In: **2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)**. [S.l.: s.n.], 2015. p. 139–144.

STROGATZ, S. H. **Nonlinear dynamics and chaos**. [s.n.], 2018. Disponível em: <https://doi.org/10.1201/9780429492563>.

VIMAL, B. et al. Mfcc based audio classification using machine learning. In: **2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)**. [S.l.: s.n.], 2021. p. 1–4.

YING, S.; ZHANG, X. A study of zero-crossings with peak-amplitudes in speech emotion classification. In: **2010 First International Conference on Pervasive Computing, Signal Processing and Applications**. [S.l.: s.n.], 2010. p. 328–331.

ZEROUAL, A. et al. Deep learning methods for forecasting covid-19 time-series data: A comparative study. **Chaos Solitons & Fractals**, v. 140, p. 110121, jul. 2020. Disponível em: <https://doi.org/10.1016/j.chaos.2020.110121>.

ZHANG, Z. **Multivariate time series analysis in climate and environmental research**. [s.n.], 2017. Disponível em: <https://doi.org/10.1007/978-3-319-67340-0>.