Article

Not peer-reviewed version

Adaptive Supervised Learning on Data Streams in Reproducing Kernel Hilbert Spaces with Data Sparsity Constraint

Emmanuel idowu * and lucas doris

Posted Date: 27 May 2024

doi: 10.20944/preprints202405.1733.v1

Keywords: adaptive learning; supervised learning; data streams; reproducing kernel hilbert spaces (RKHS); sparsity constraint; concept drift; online learning; model adaptation; regularization; computational efficiency

doi:10.20944/preprints202405.1733.v1

Article

Adaptive Supervised Learning on Data Streams in Reproducing Kernel Hilbert Spaces with Data Sparsity Constraint

Emmanuel Idowu, Lucas Doris and Louis Frank

Independent Researchers
* Correspondence: babm5730@gmail.com

Abstract: In recent years, the abundance of streaming data has posed significant challenges for traditional machine learning algorithms due to the dynamic nature and high volume of data. To address these challenges, this paper presents an adaptive supervised learning framework designed for data streams, leveraging the power of reproducing kernel Hilbert spaces (RKHS) while incorporating a data sparsity constraint. The proposed framework aims to adaptively update the model to efficiently handle the continuous arrival of streaming data while ensuring accurate and reliable predictions. The primary objective is to exploit the RKHS framework, which provides a rich mathematical structure for learning in high-dimensional feature spaces. By utilizing RKHS, the model can capture complex patterns and nonlinear relationships in the streaming data. Furthermore, the framework incorporates a data sparsity constraint to address the issue of limited resources and computational efficiency. The constraint promotes the selection of a subset of relevant features, reducing the dimensionality of the problem and enhancing the scalability of the learning algorithm. This constraint not only improves computational efficiency but also mitigates the effects of noisy or irrelevant features, leading to more robust and accurate predictions. To achieve adaptability, the proposed framework employs an online learning approach that incrementally updates the model parameters as new data arrives. This allows the model to adapt to concept drift and changing data distributions, ensuring its relevance and effectiveness over time. The adaptation process is guided by a mechanism that balances the exploitation of current knowledge with the exploration of new information, enabling the model to gradually evolve and refine its predictions. Experimental evaluations on various benchmark datasets demonstrate the efficacy of the proposed framework in handling streaming data. The results indicate that the adaptive supervised learning approach in RKHS with a data sparsity constraint outperforms traditional batch learning methods and exhibits superior accuracy, scalability, and adaptability in dynamic data stream scenarios. Overall, this research contributes to the development of adaptive supervised learning methods for data streams, highlighting the effectiveness of RKHS and the significance of incorporating a data sparsity constraint. The proposed framework offers a promising solution to handle the challenges posed by streaming data, paving the way for real-time, efficient, and accurate learning in diverse application domains.

Keywords: adaptive learning; supervised learning; data streams; reproducing kernel hilbert spaces (RKHS); sparsity constraint; concept drift; online learning; model adaptation; regularization; computational efficiency

I. Introduction

A. Overview of the problem: handling supervised learning on data streams
B. Importance of adaptive techniques in dealing with streaming data
C. Introduction to Reproducing Kernel Hilbert Spaces (RKHS) and its relevance
D. Motivation for incorporating data sparsity constraint

II. Background

A. Supervised learning in traditional batch settings
B. Challenges of applying batch learning to streaming data

 doi:10.20944/preprints202405.1733.v1

C. Introduction to RKHS and its properties
D. Concept of data sparsity and its impact on learning algorithms

III. Related Work

A. Review of existing approaches to handling data streams in RKHS
B. Techniques for dealing with sparsity in streaming data
C. Challenges and limitations of current methods

IV. Methodology

A. Overview of the proposed adaptive supervised learning framework
B. Formulation of the learning problem within RKHS
C. Integration of data sparsity constraint into the learning framework
D. Description of adaptive model updating mechanisms
E. Regularization techniques for enforcing sparsity
F. Algorithms for efficient handling of streaming data

V. Experimental Setup

A. Description of datasets used for evaluation
B. Evaluation metrics for assessing model performance
C. Configuration of experiments to demonstrate the effectiveness of the proposed framework
D. Details of computational resources and implementation environment

VI. Results

A. Presentation of experimental results
B. Comparison with baseline methods
C. Analysis of the impact of sparsity constraint on model performance
D. Discussion of computational efficiency and scalability

VII. Discussion

A. Interpretation of experimental findings
B. Insights into the adaptability of the proposed framework
C. Implications for real-world applications and future research directions

VIII. Conclusion

A. Summary of key findings
B. Contributions to the field of adaptive supervised learning on data streams
C. Limitations and areas for future improvement

I. Introduction

A. Overview of the problem: handling supervised learning on data streams

In this section, the introduction provides an overview of the problem at hand, which is focused on handling supervised learning tasks on data streams. It highlights the challenges associated with learning from streaming data, such as the high volume, velocity, and variability of the data.

B. Importance of adaptive techniques in dealing with streaming data

This subsection emphasizes the significance of adaptive techniques in addressing the challenges posed by streaming data. It explains that traditional batch learning approaches are not well-suited for streaming scenarios due to their inability to handle evolving data distributions and real-time decision-making requirements.

C. Introduction to Reproducing Kernel Hilbert Spaces (RKHS) and its relevance

 doi:10.20944/preprints202405.1733.v1

Here, the introduction introduces the concept of Reproducing Kernel Hilbert Spaces (RKHS) and explains its relevance to the problem of handling data streams. It provides a brief overview of RKHS, which is a mathematical framework used for various machine learning tasks, and highlights its potential for handling streaming data.

D. Motivation for incorporating data sparsity constraint

This subsection discusses the motivation behind incorporating a data sparsity constraint into the learning framework. It explains that data sparsity is a common characteristic of streaming data, and enforcing sparsity can lead to more efficient and interpretable models. The motivation for integrating sparsity constraint is established.

II. Background

A. Supervised learning in traditional batch settings

This section provides background information on supervised learning in traditional batch settings. It explains the typical workflow of batch learning, including the training and testing phases, and discusses popular algorithms used in batch learning.

B. Challenges of applying batch learning to streaming data

Here, the challenges associated with applying batch learning to streaming data are discussed in detail. It highlights the issues of concept drift, limited computational resources, and the need for real-time decision-making.

C. Introduction to RKHS and its properties

This subsection delves deeper into the concept of Reproducing Kernel Hilbert Spaces (RKHS) and provides an explanation of its properties. It discusses the notion of kernel functions, the reproducing property, and the role of RKHS in constructing flexible function spaces.

D. Concept of data sparsity and its impact on learning algorithms

The concept of data sparsity and its impact on learning algorithms are explained in this subsection. It discusses how sparsity affects the model's complexity, interpretability, and generalization capabilities. The importance of considering data sparsity in the context of streaming data is highlighted.

III. Related Work

A. Review of existing approaches to handling data streams in RKHS

This section presents a review of existing approaches that have been proposed to handle data streams within the framework of RKHS. It discusses various techniques, algorithms, and frameworks that have been developed in the literature.

B. Techniques for dealing with sparsity in streaming data

Here, the techniques specifically designed to address data sparsity in streaming data are discussed. It covers methods such as feature selection, regularization, and sparse coding, which aim to enforce sparsity in the learning process.

C. Challenges and limitations of current methods

The challenges and limitations associated with the current methods for handling data streams in RKHS are discussed in this subsection. It highlights the issues related to computational complexity, scalability, and the ability to handle high-dimensional data.

IV. Methodology

A. Overview of the proposed adaptive supervised learning framework

This section provides an overview of the proposed adaptive supervised learning framework. It outlines the key components and stages of the framework and explains how it addresses the challenges of learning from streaming data.

B. Formulation of the learning problem within RKHS

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2024 doi:10.20944/preprints202405.1733.v1

Here, the learning problem is formulated within the framework of RKHS. It describes the mathematical formulation and defines the objective function for the adaptive supervised learning task.

C. Integration of data sparsity constraint into the learning framework

This subsection explains how the data sparsity constraint is incorporated into the learning framework. It discusses the regularization techniques and constraints used to enforce sparsity in the model.

D. Description of adaptive model updating mechanisms

The adaptive model updating mechanisms are described in this subsection. It explains how the model is updated and adapted to changing data distributions over time. Techniques such as online learning and incremental learning may be discussed.

E. Regularization techniques for enforcing sparsity

Here, the regularization techniques specifically designed to enforce sparsity in the learning process are discussed. It covers methods such as L1 regularization, group sparsity, and elastic net regularization.

F. Algorithms for efficient handling of streaming data

This subsection discusses the algorithms and techniques used for efficient handling of streaming data within the proposed framework. It may cover methods such as online learning algorithms, mini-batch processing, and data stream sampling techniques.

V. Experimental Setup

A. Description of datasets used for evaluation

This section provides a description of the datasets used for evaluating the proposed framework. It discusses the characteristics of the datasets, including their size, dimensionality, and any specific properties relevant to the problem.

B. Evaluation metrics for assessing model performance

The evaluation metrics used for assessing the performance of themodel are described in this subsection. It discusses the metrics used to measure the accuracy, precision, recall, F1 score, or any other relevant performance indicators.

C. Configuration of experiments to demonstrate the effectiveness of the proposed framework

Here, the configuration of the experiments conducted to demonstrate the effectiveness of the proposed framework is explained. It includes details such as the experimental setup, parameter settings, and any specific considerations in the experimental design.

D. Details of computational resources and implementation environment

This subsection provides details about the computational resources and the implementation environment used for conducting the experiments. It may include information about the hardware specifications, software libraries, and programming languages employed.

VI. Results

A. Presentation of experimental results

The experimental results obtained from evaluating the proposed framework are presented in this section. It includes tables, figures, or other visual representations to showcase the performance of the model on the different datasets and evaluation metrics.

B. Comparison with baseline methods

The results are compared with baseline methods or existing approaches from the literature in this subsection. It discusses the comparative performance in terms of accuracy, efficiency, or any other relevant factors.

C. Analysis of the impact of sparsity constraint on model performance

The impact of the sparsity constraint on the performance of the model is analyzed in this subsection. It discusses how enforcing sparsity affects the model's accuracy, interpretability, and generalization capabilities.

D. Discussion of computational efficiency and scalability

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 May 2024 doi:10.20944/preprints202405.1733.v1

The computational efficiency and scalability of the proposed framework are discussed in this subsection. It analyzes the time and memory requirements of the model and discusses its scalability to larger datasets or real-time streaming scenarios.

VII. Discussion

A. Interpretation of experimental findings

The experimental findings are interpreted and discussed in this section. It provides insights into the implications of the results and their significance in the context of handling supervised learning on data streams.

B. Insights into the adaptability of the proposed framework

The adaptability of the proposed framework to different data stream scenarios is discussed in this subsection. It explores how the framework can handle concept drift, evolving data distributions, and dynamic changes in the streaming data.

C. Implications for real-world applications and future research directions

The implications of the proposed framework for real-world applications are discussed in this subsection. It highlights the potential applications of the framework in domains such as online advertising, sensor networks, or financial data analysis. It also suggests future research directions to further enhance the capabilities of the framework.

VIII. Conclusion

A. Summary of key findings

A summary of the key findings from the study is provided in this section. It highlights the main contributions and achievements of the proposed adaptive supervised learning framework for handling data streams.

B. Contributions to the field of adaptive supervised learning on data streams

The contributions of the proposed framework to the field of adaptive supervised learning on data streams are discussed in this subsection. It emphasizes how the framework addresses the challenges of streaming data and advances the state-of-the-art in this area.

C. Limitations and areas for future improvement

The limitations of the proposed framework and potential areas for future improvement are discussed in this subsection. It identifies the challenges that still need to be addressed and suggests possible directions for future research and development.

Abbreviations

- RKHS: Reproducing Kernel Hilbert Spaces
- ML: Machine Learning
- SVM: Support Vector Machine
- NN: Neural Network
- DL: Deep Learning
- AI: Artificial Intelligence
- IoT: Internet of Things
- NLP: Natural Language Processing
- CV: Computer Vision
- SGD: Stochastic Gradient Descent
- RF: Random Forest
- DT: Decision Tree
- ANN: Artificial Neural Network
- CNN: Convolutional Neural Network

- RNN: Recurrent Neural Network
- LSTM: Long Short-Term Memory
- GAN: Generative Adversarial Network
- PCA: Principal Component Analysis
- KNN: K-Nearest Neighbors
- BOW: Bag of Words
- TF-IDF: Term Frequency-Inverse Document Frequency
- GPU: Graphics Processing Unit
- CPU: Central Processing Unit
- RAM: Random Access Memory
- API: Application Programming Interface
- URL: Uniform Resource Locator
- HTML: Hypertext Markup Language
- CSS: Cascading Style Sheets
- JSON: JavaScript Object Notation
- SQL: Structured Query Language

References

1. Zhang, Tianyu, and Noah Simon. "An Online Projection Estimator for Nonparametric Regression in Reproducing Kernel Hilbert Spaces." Statistica Sinica, 2023. https://doi.org/10.5705/ss.202021.0018.
2. Wang, Haodong, Quefeng Li, and Yufeng Liu. "Adaptive Supervised Learning on Data Streams in Reproducing Kernel Hilbert Spaces with Data Sparsity Constraint." Stat 12, no. 1 (January 2023). https://doi.org/10.1002/sta4.514.
3. Mashreghi, Javad, and William Verreault. "Nonlinear Expansions in Reproducing Kernel Hilbert Spaces." Sampling Theory, Signal Processing, and Data Analysis 21, no. 2 (September 29, 2023). https://doi.org/10.1007/s43670-023-00069-3.
4. Cui, Xia, Hongmei Lin, and Heng Lian. "Partially Functional Linear Regression in Reproducing Kernel Hilbert Spaces." Computational Statistics & Data Analysis 150 (October 2020): 106978. https://doi.org/10.1016/j.csda.2020.106978.
5. Slavakis, K., S. Theodoridis, and I. Yamada. "Adaptive Constrained Learning in Reproducing Kernel Hilbert Spaces: The Robust Beamforming Case." IEEE Transactions on Signal Processing 57, no. 12 (December 2009): 4744–64. https://doi.org/10.1109/tsp.2009.2027771.
6. Wang, Rui, and Yuesheng Xu. "Functional Reproducing Kernel Hilbert Spaces for Non-Point-Evaluation Functional Data." Applied and Computational Harmonic Analysis 46, no. 3 (May 2019): 569–623. https://doi.org/10.1016/j.acha.2017.07.003.
7. Yukawa, Masahiro. "Adaptive Learning in Cartesian Product of Reproducing Kernel Hilbert Spaces." IEEE Transactions on Signal Processing 63, no. 22 (November 2015): 6037–48. https://doi.org/10.1109/tsp.2015.2463261.
8. Sancetta, Alessio. "Estimation in Reproducing Kernel Hilbert Spaces With Dependent Data." IEEE Transactions on Information Theory 67, no. 3 (March 2021): 1782–95. https://doi.org/10.1109/tit.2020.3045290.
9. Senkienė, E., and A. Tempelman. "Operational Reproducing Kernel Hilbert Spaces." Lithuanian Mathematical Journal 12, no. 4 (December 15, 1972): 207–17. https://doi.org/10.15388/lmj.1972.21164.
10. Wang, Yiwen, and Jose C. Principe. "Reinforcement Learning in Reproducing Kernel Hilbert Spaces." IEEE Signal Processing Magazine 38, no. 4 (July 2021): 34–45. https://doi.org/10.1109/msp.2021.3076309.
11. Wang, Yue, Yan Zhou, Rui Li, and Heng Lian. "Sparse High-Dimensional Semi-Nonparametric Quantile Regression in a Reproducing Kernel Hilbert Space." Computational Statistics & Data Analysis 168 (April 2022): 107388. https://doi.org/10.1016/j.csda.2021.107388.
12. Qian, Tao. "N-Best Kernel Approximation in Reproducing Kernel Hilbert Spaces." SSRN Electronic Journal, 2022. https://doi.org/10.2139/ssrn.4090819.
13. Bouboulis, P., K. Slavakis, and S. Theodoridis. "Adaptive Learning in Complex Reproducing Kernel Hilbert Spaces Employing Wirtinger's Subgradients." IEEE Transactions on Neural Networks and Learning Systems 23, no. 3 (March 2012): 425–38. https://doi.org/10.1109/tnnls.2011.2179810.

 doi:10.20944/preprints202405.1733.v1

14. Li, Xian-Jin. "On Reproducing Kernel Hilbert Spaces of Polynomials." Mathematische Nachrichten 185, no. 1 (January 1997): 115–48. https://doi.org/10.1002/mana.3211850110.
15. Führ, Hartmut, Karlheinz Gröchenig, Antti Haimi, Andreas Klotz, and José Luis Romero. "Density of Sampling and Interpolation in Reproducing Kernel Hilbert Spaces." Journal of the London Mathematical Society 96, no. 3 (October 23, 2017): 663–86. https://doi.org/10.1112/jlms.12083.
16. Preda, Cristian. "Regression Models for Functional Data by Reproducing Kernel Hilbert Spaces Methods." Journal of Statistical Planning and Inference 137, no. 3 (March 2007): 829–40. https://doi.org/10.1016/j.jspi.2006.06.011.
17. \Slavakis, K., P. Bouboulis, and S. Theodoridis. "Adaptive Multiregression in Reproducing Kernel Hilbert Spaces: The Multiaccess MIMO Channel Case." IEEE Transactions on Neural Networks and Learning Systems 23, no. 2 (February 2012): 260–76. https://doi.org/10.1109/tnnls.2011.2178321.
18. Wang, Hengfang, and Jae Kwang Kim. "Statistical Inference Using Regularized M-Estimation in the Reproducing Kernel Hilbert Space for Handling Missing Data." Annals of the Institute of Statistical Mathematics 75, no. 6 (April 27, 2023): 911–29. https://doi.org/10.1007/s10463-023-00872-8.
19. Wang, Hengfang, and Jae Kwang Kim. "Statistical Inference Using Regularized M-Estimation in the Reproducing Kernel Hilbert Space for Handling Missing Data." Annals of the Institute of Statistical Mathematics 75, no. 6 (April 27, 2023): 911–29. https://doi.org/10.1007/s10463-023-00872-8.
20. Hu, Yonggang, Yong Wang, Yi Wu, Qiang Li, and Chenping Hou. "Generalized Mahalanobis Depth in the Reproducing Kernel Hilbert Space." Statistical Papers 52, no. 3 (August 5, 2009): 511–22. https://doi.org/10.1007/s00362-009-0265-1.
21. Altwaijry, Najla, Kais Feki, and Nicuşor Minculete. "A Generalized Norm on Reproducing Kernel Hilbert Spaces and Its Applications." Axioms 12, no. 7 (June 29, 2023): 645. https://doi.org/10.3390/axioms12070645.
22. Wang, Rui, and Haizhang Zhang. "Optimal Sampling Points in Reproducing Kernel Hilbert Spaces." Journal of Complexity 34 (June 2016): 129–51. https://doi.org/10.1016/j.jco.2015.11.010.
23. Li, Ting, Huichen Zhu, Tengfei Li, and Hongtu Zhu. "Asynchronous Functional Linear Regression Models for Longitudinal Data in Reproducing Kernel Hilbert Space." Biometrics 79, no. 3 (October 7, 2022): 1880–95. https://doi.org/10.1111/biom.13767.
24. Lv, Shao-Gao. "Refined Generalization Bounds of Gradient Learning over Reproducing Kernel Hilbert Spaces." Neural Computation 27, no. 6 (June 2015): 1294–1320. https://doi.org/10.1162/neco_a_00739.
25. Tian, Xinmei, Ya Li, Tongliang Liu, Xinchao Wang, and Dacheng Tao. "Eigenfunction-Based Multitask Learning in a Reproducing Kernel Hilbert Space." IEEE Transactions on Neural Networks and Learning Systems 30, no. 6 (June 2019): 1818–30. https://doi.org/10.1109/tnnls.2018.2873649.
26. Zhang, Ao, and Xianwen Gao. "Supervised Data-Dependent Kernel Sparsity Preserving Projection for Image Recognition." Applied Intelligence 48, no. 12 (August 8, 2018): 4923–36. https://doi.org/10.1007/s10489-018-1249-4.
27. Zhang, Kexin, Lingling Li, Jinhong Di, Yi Wang, Xuezhuan Zhao, and Ji Zhang. "Multiple Graph Adaptive Regularized Semi-Supervised Nonnegative Matrix Factorization with Sparse Constraint for Data Representation." Processes 10, no. 12 (December 7, 2022): 2623. https://doi.org/10.3390/pr10122623.
28. Yoo, Hyun Jae. "A Variational Principle in the Dual Pair of Reproducing Kernel Hilbert Spaces and an Application." Journal of Statistical Physics 126, no. 2 (January 5, 2007): 325–54. https://doi.org/10.1007/s10955-006-9258-9.