INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHINE LEARNING

ADAPTIVE LINEAR MODELS FOR REGRESSION IN EVOLVING DATA STREAMS

Dr. Natalia V. Smirnova
Department of Computer Science, Moscow Institute of Physics and Technology (MIPT), Russia

Elena Baranova
Department of Applied Mathematics, Saint Petersburg State University, Russia

Published Date: 14 December 2024 // Page no.:- 1-5

ABSTRACT

Regression analysis in data streams presents unique challenges due to the continuous, potentially infinite nature of the data and the phenomenon of concept drift, where the underlying data distribution or the relationship between variables changes over time. Traditional static regression models are ill-equipped to handle such dynamic environments. Adaptive linear filtering techniques offer a powerful paradigm for regression in data streams, allowing models to evolve and adjust to changing patterns. This article explores the application of linear adaptive filtering methods for regression tasks in data stream settings. We discuss the fundamental principles of adaptive filtering, common algorithms like Recursive Least Squares (RLS) and its variants, and their suitability for handling concept drift. By reviewing relevant literature on data stream mining, adaptive learning, and regression techniques, we highlight the advantages of using adaptive linear models, including their computational efficiency, ability to track changing relationships, and theoretical foundations in signal processing. While acknowledging limitations such as sensitivity to parameter choices and potential issues with non-linear relationships, this article argues that linear adaptive filtering provides a robust and efficient foundation for performing regression in dynamic data stream environments, serving as a crucial component in more complex adaptive learning systems.

Keywords: Adaptive Filtering, Linear Regression, Data Streams, Concept Drift, Recursive Least Squares, Online Learning, Evolving Data, Machine Learning.

INTRODUCTION

The proliferation of sensors, interconnected devices, and online services has led to the generation of vast amounts of data in the form of continuous, high-speed streams [5, 12]. Analyzing these data streams in real-time or near real-time is essential for various applications, including fraud detection, network monitoring, financial analysis, and predictive maintenance [3, 5]. Regression, the task of predicting a continuous target variable based on a set of input features, is a fundamental analytical technique required in many of these streaming scenarios [9, 11].

Traditional regression models, such as standard linear regression or batch-trained machine learning models, operate under the assumption that the data distribution is static [31]. However, data streams are inherently dynamic. The relationships between variables can change over time, a phenomenon known as concept drift [8]. For example, the factors influencing traffic volume on a highway [15] or the predictors of energy consumption in a building can evolve due to external factors, seasonality, or underlying system changes. Static models trained on historical data quickly become outdated and perform poorly in the presence of drift.

Adaptive learning techniques are designed to address this challenge by allowing models to continuously update and adjust as new data arrives [3, 8]. Adaptive filtering, a field with deep roots in signal processing, provides a powerful framework for developing models that can learn and track time-varying parameters [13, 19, 29]. Linear adaptive filters, in particular, offer computationally efficient methods for estimating the coefficients of a linear model that changes over time [13, 25, 29].

While more complex adaptive methods, including ensemble techniques [9, 10, 11], adaptive decision trees [16, 17, 18], and adaptive instance-based methods [21, 26], have been developed for data streams, linear adaptive filtering provides a foundational and often highly effective approach for regression tasks, especially when the underlying relationships are approximately linear or can be linearized. Its efficiency makes it suitable for high-speed data environments [5].

This article explores the application of linear adaptive filtering techniques for performing regression in evolving data streams. We will delve into the core concepts, discuss prominent algorithms, and highlight their advantages and considerations in the context of concept drift. By drawing upon literature from adaptive filtering, data stream mining, and machine learning, we aim to demonstrate the relevance and utility of this approach for building dynamic regression models capable of handling the challenges of continuous, changing data.

2. Methods

The application of linear adaptive filtering for regression in data streams involves formulating the regression problem within an adaptive framework and employing algorithms that can continuously update model parameters as new data instances arrive. This section outlines the core methodology.

2.1. Problem Formulation

In a data stream regression setting, we receive a sequence of data instances (xt,yt) over time t=1,2,…, where xt is a vector of input features at time t, and yt is the corresponding continuous target variable. The goal is to learn a model f(xt) that predicts yt, where the relationship between xt and yt may change over time due to concept drift [8]. A linear adaptive filter assumes a linear relationship:

$$y^\wedge t = wtTxt$$

where y^t is the predicted value of yt, xt is the input feature vector (potentially including a bias term), and wt is the vector of model coefficients (weights) that are updated at each time step t.

2.2. Adaptive Filtering Algorithms

The core of the method lies in the adaptive algorithm used to update the weight vector wt. These algorithms aim to minimize an error criterion, typically the squared prediction error (yt−y^t)2, as each new data instance becomes available.

- Least Mean Squares (LMS): A fundamental adaptive filtering algorithm [29]. The weight vector is updated iteratively in the direction that reduces the instantaneous squared error:

$$wt + 1 = wt + \mu(yt - y^\wedge t)xt$$

where μ is the learning rate, a small positive constant that controls the step size of the adaptation. LMS is computationally simple but can be slow to converge and sensitive to the learning rate choice.

- Recursive Least Squares (RLS): RLS algorithms aim to minimize the sum of squared errors over a window of past data, often using exponential weighting to give more importance to recent data [4, 13, 20, 25]. The standard RLS update equations involve matrix operations, including matrix inversion or its inverse update using the matrix inversion lemma [27]:

$$Pt = \lambda 1(Pt - 1 - 1\lambda + xtTPt - 1xtPt - 1xtxtTPt - 1)$$

$$kt = Ptxt$$

$$wt = wt - 1 + kt(yt - y^\wedge t)$$

where λ is the forgetting factor (0<λ≤1), controlling the exponential weighting of past data. A smaller λ makes the algorithm more adaptive to recent changes but also more sensitive to noise. Pt is related to the inverse of the input correlation matrix. RLS is generally faster to converge and more effective at tracking concept drift than LMS but is computationally more expensive due to matrix operations.

Exponentially Weighted Least Squares (EWLS): This is closely related to RLS with a forgetting factor, explicitly minimizing the exponentially weighted sum of squared errors [4, 20]. The recursive updates are equivalent to RLS with exponential forgetting.

Variants and Extensions: Numerous variants of LMS and RLS exist to improve performance, robustness, or computational efficiency, such as Normalized LMS (NLMS) or approximations for large-scale data [13].

2.3. Handling Concept Drift

Adaptive linear filters inherently handle concept drift by continuously updating their parameters [8]. The forgetting factor λ in RLS (or the learning rate μ in LMS) plays a crucial role in the model's ability to adapt. A smaller λ (or larger μ) allows the model to quickly adjust to new concepts but can lead to instability or over-sensitivity to noise. Conversely, a larger λ (or smaller μ) provides smoother estimates but makes the model slower to adapt to significant changes.

While basic adaptive filters continuously update, more sophisticated drift detection mechanisms can be integrated [8, 22]. These mechanisms monitor the model's performance (e.g., prediction error) and signal when a significant change (drift) is detected. Upon detecting drift, the adaptive filter's parameters (like λ or μ) could be adjusted, or the model could be partially or fully reset. However, standard linear adaptive filters like EWLS/RLS are often used without explicit drift detection, relying solely on the forgetting factor for continuous adaptation.

2.4. Evaluation

Evaluating adaptive regression models on data streams requires specific methodologies [7]. Unlike batch learning where a single test set is used, evaluation in data streams is typically performed using a prequential (or interleaved test-then-train) approach. Each incoming data instance is first used to test the current model's performance (e.g., calculating the squared error), and then it is used to update the model. Performance metrics, such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE), are calculated incrementally over time or over a sliding window to assess the model's performance as it adapts to the stream. Cumulative metrics can also be used to evaluate overall performance.

3. Results (Expected Outcomes and Properties)

Applying linear adaptive filtering techniques for regression in data streams is expected to yield models with specific properties and performance characteristics, particularly in comparison to static linear models. Based on the principles of adaptive filtering and existing literature, the following results and outcomes are anticipated:

3.1. Ability to Track Changing Linear Relationships

The primary expected result is that linear adaptive filters, especially those employing exponential forgetting like RLS/EWLS [4, 13, 20, 25], will effectively track changes in the underlying linear relationship between input features and the target variable in a data stream. As the true coefficients of the linear model drift over time, the adaptive filter's estimated coefficients wt are expected to converge towards the current optimal values, provided the rate of adaptation (controlled by λ or μ) is appropriately matched to the rate of concept drift [8]. This contrasts sharply with static linear regression, where the model parameters remain fixed after initial training, leading to degraded performance in the presence of drift.

3.2. Continuous Learning and Adaptation

Linear adaptive filters facilitate continuous learning [3, 8, 12]. With each new data instance, the model parameters are updated. This ensures that the model is always learning from the most recent data, making it suitable for environments where data arrives sequentially and the underlying patterns are non-stationary. This is a fundamental advantage over batch learning methods that require periodic retraining on accumulated data, which can be computationally expensive and introduce latency.

3.3. Computational Efficiency (Relative to Complex Models)

Compared to more complex adaptive models for data streams, such as adaptive ensembles of non-linear models [9, 10, 11], adaptive decision trees [16, 17, 18], or adaptive instance-based methods [21, 26], linear adaptive filters like LMS and RLS offer relatively high computational efficiency per data instance. LMS is particularly lightweight, involving simple vector updates. While RLS involves matrix operations, its recursive nature and the use of the matrix inversion lemma [27] make it significantly more efficient than recomputing the least squares solution from scratch on a sliding window of data [4, 13, 20, 25]. This efficiency is crucial for processing high-speed data streams [5].

3.4. Predictable Performance in Linear or Approximately Linear Scenarios

In scenarios where the underlying relationship between variables is truly linear or can be reasonably approximated by a linear model that changes over time, linear adaptive filters are expected to provide accurate and stable predictions. Their performance is theoretically well-understood in such settings [13, 19].

3.5. Basis for More Complex Adaptive Systems

Linear adaptive filters can serve as foundational components within more complex adaptive learning systems for data streams. For example, they can be used as the base learners in adaptive ensembles [9, 10, 11] or as components within adaptive model trees [16, 17]. Their efficiency allows them to be combined effectively to handle more complex relationships or detect drift.

3.6. Sensitivity to Parameter Choice

A key result is that the performance of linear adaptive filters is highly sensitive to the choice of adaptation parameters, particularly the learning rate μ for LMS or the forgetting factor λ for RLS [4, 13, 29]. An inappropriate choice can lead to slow adaptation (missing drift) or instability (over-sensitivity to noise). Finding the optimal parameter often requires experimentation or the use of adaptive parameter tuning techniques.

3.7. Limitations with Strongly Non-Linear Relationships

While effective for linear or approximately linear relationships, linear adaptive filters will struggle to accurately model strongly non-linear relationships, even if those relationships are static or changing. In such cases, non-linear adaptive methods or transformations of the input features would be necessary [23].

These expected results highlight the strengths and limitations of linear adaptive filtering for regression in data streams. They provide a foundation for understanding when and how these techniques can be effectively applied.

4. DISCUSSION AND CONCLUSION

The challenge of performing regression in evolving data streams, where the underlying relationships between variables can change over time, necessitates the use of adaptive learning techniques. Linear adaptive filtering offers a powerful and efficient paradigm for addressing this challenge, providing models that can continuously adjust to new data and track concept drift.

As discussed, linear adaptive filters, particularly variants of Recursive Least Squares with exponential forgetting [4, 13, 20, 25], are expected to effectively track changes in linear or approximately linear relationships within data streams. Their ability to continuously update parameters as new data arrives ensures that the model remains relevant in dynamic environments, a critical advantage over static regression methods [31]. The computational efficiency of algorithms like LMS and RLS makes them well-suited for processing the high-speed, potentially infinite data volumes characteristic of data streams [5, 12].

Furthermore, linear adaptive filters can serve as fundamental building blocks for more sophisticated adaptive learning systems [9, 10, 11, 16, 17]. Their simplicity and efficiency allow them to be readily integrated into ensemble methods or hierarchical structures designed to handle more complex forms of concept drift or non-linear relationships.

However, it is crucial to acknowledge the limitations. The performance of linear adaptive filters is highly dependent on the appropriate selection of adaptation parameters (e.g., forgetting factor λ) [4, 13, 29]. Tuning these parameters can be challenging and often requires domain knowledge or empirical evaluation on representative data streams. Moreover, linear adaptive filters are inherently limited to modeling linear relationships. While feature engineering or transformations can sometimes linearize non-linear problems, strongly non-linear dynamics may require the use of non-linear adaptive methods [23, 26].

Future research in this area could explore methods for automatically tuning the adaptation parameters of linear adaptive filters in response to detected concept drift [8, 22]. Investigating hybrid approaches that combine the efficiency of linear adaptive filters with the modeling power of non-linear techniques, perhaps within an ensemble framework, could also be a fruitful direction [9, 10, 11]. Furthermore, applying these techniques to a wider variety of real-world data streams and analyzing their performance under different types and rates of concept drift would provide valuable empirical evidence of their effectiveness.

In conclusion, linear adaptive filtering provides a robust, efficient, and theoretically grounded approach for performing regression in evolving data streams. While not a panacea for all data stream regression problems, particularly those involving strongly non-linear relationships, its ability to continuously adapt and track changing linear patterns makes it an indispensable tool in the data stream mining toolkit. As data streams continue to grow in volume and complexity, adaptive linear models will remain a crucial component in the development of intelligent systems capable of learning and making predictions in dynamic, real-world environments.

REFERENCES

Almeida, E., Ferreira, C., Gama, J.: Adaptive model rules from data streams. In: ECML, pp. 480–492. Springer (2013)

Bifet, A., Gavalda, R.: Learning from time-changing data with adaptive windowing. In: SIAM international conference on data mining, pp. 443–448 (2007)

Bifet, A., Gavaldà, R.: Adaptive learning from evolving data streams. In: International Symposium on Intelligent Data Analysis, pp. 249–260. Springer (2009)

Campi, M.C.: Exponentially weighted least squares identification of time-varying systems with white disturbances. IEEE Transact. Signal Proc. 42(11), 2906–2914 (1994)

Domingos, P., Hulten, G.: Mining high-speed data streams. In: ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 71–80 (2000)

Fanaee-T, H., Gama, J.: Event labeling combining ensemble detectors and background knowledge. Pro. Artif. Intell. 2, 113–127 (2014)

Gama, J., Sebastiao, R., Rodrigues, P.P.: On evaluating stream learning algorithms. Mach. learn. 90, 317–346 (2013)

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM comput. Surv. (CSUR) 46(4), 1–37 (2014)

Gomes, H.M., Barddal, J.P., Ferreira, L.E.B., Bifet, A.: Adaptive random forests for data stream regression. In: ESANN (2018)

Gomes, H.M., Bifet, A., Read, J., Barddal, J.P., Enembreck, F., Pfahringer, B., Holmes, G., Abdessalem, T.: Adaptive random forests for evolving data stream classification. Machine Learning (2017)

Gomes, H.M., Montiel, J., Martiello Mastelini, S., Pfahringer, B., Bifet, A.: On ensemble techniques for data stream regression. In: 2020 International Joint Conference on Neural Networks (IJCNN). IEEE (2020)

Gomes, H.M., Read, J., Bifet, A., Barddal, J.P., Gama, J.: Machine learning for streaming data: state of the art, challenges, and opportunities. ACM SIGKDD Explorat. Newsletter 21(2), 6–22 (2019)

Haykin, S.S.: Adaptive filter theory. Pearson (2005)

Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. (1963). https://doi.org/10.1007/978-1-4612-0865-5_26

Hogue, J.: Metro interstate traffic volume (2019). https://doi.org/10.24432/C5X60B. https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume

Ikonomovska, E., Gama, J., Džeroski, S.: Learning model trees from evolving data streams. Data Min. know. Discov.

23(1), 128–168 (2011)

Ikonomovska, E., Gama, J., Džeroski, S.: Online tree-based ensembles and option trees for regression on evolving data streams. Neurocomputing (2015). https://doi.org/10.1007/978-1-4612-0865-5_26

Ikonomovska, E., Gama, J., Zenko, B., Dzeroski, S.: Speeding-up hoeffding-based regression trees with options. In: ICML, pp. 537–544. Citeseer (2011)

Ingle, V.K., Kogon, S.M., Manolakis, D.G.: Statistical and adaptive signal processing. Artech House, Cambridge, MA, USA (2005)

Li, T.H.: On exponentially weighted recursive least squares for estimating time-varying parameters and its application to computer workload forecasting. J. Stat. Theory Pract. 2(3), 339–354 (2008). https://doi.org/10.1080/15598608.2008.10411879

Losing, V., Hammer, B., Wersing, H.: Knn classifier with self adjusting memory for heterogeneous concept drift. In: 2016 IEEE 16th international conference on data mining (ICDM), pp. 291–300. IEEE (2016)

Mouss, H., Mouss, D., Mouss, N., Sefouhi, L.: Test of page-hinckley, an approach for fault detection in an agro-alimentary production system. In: Asian Control Conference, vol. 2 (2004)

Narendra, K., Gallman, P.: An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Transact. Automat. Control 11(3), 546–550 (1966). https://doi.org/10.1109/TAC.1966.1098387

Nash, W., Sellers, T., Talbot, S., Cawthorn, A., Ford, W.: Abalone. UCI Machine Learning Repository (1995). https://doi.org/10.24432/C55C7W

Sayed, A.H., Kailath, T.: A state-space approach to adaptive rls filtering. IEEE signal processing magazine 11(ARTICLE), 18–60 (1994)

Sun, Y., Pfahringer, B., Gomes, H.M., Bifet, A.: Soknl: a novel way of integrating k-nearest neighbours with adaptive random forest regression for data streams. Data Min. Know. Discov. 36(5), 2006–2032 (2022)

Tylavsky, D.J., Sohie, G.R.: Generalization of the matrix inversion lemma. Proc. IEEE 74(7), 1050–1052 (1986)

Weiss, S.M., Indurkhya, N.: Rule-based machine learning methods for functional prediction. J. Artif. Intell. Res. 3, 383–403 (1995)

Widrow, B., Stearns, S.D.: Adaptive signal processing. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 491 p. 1 (1985)

Wills, A., Schön, T.B., Ljung, L., Ninness, B.: Identification of Hammerstein-wiener models. Automatica 49(1), 70–81 (2013).