

TiDE: Revolutionizing Long-Term Time Series Forecasting with Simple MLP Architectures



Philippe Dagher · [Follow](#)

9 min read · Sep 18, 2023



178



1



The realm of long-term time series forecasting is fraught with challenges. It's an essential endeavor across numerous industries, from predicting stock prices and consumer demands to anticipating energy consumption and beyond. At its core, the primary objective is to project future values based on historical data. This sounds simple in theory, but the multifaceted nature of time series data, replete with trends, seasonality, and sometimes unexpected shocks, makes this task immensely intricate.

Historically, a suite of methods has been employed to grapple with this challenge. Traditional statistical methods like ARIMA (AutoRegressive Integrated Moving Average) or Exponential Smoothing have been mainstays in this domain. Their strength lies in their foundation, deeply rooted in statistical theory, which often ensures robustness. However, they sometimes falter when dealing with non-linear patterns or external influencing factors called covariates.

The rise of neural networks brought a breath of fresh air to the forecasting landscape. Transformer models, particularly, with their self-attention mechanisms, seemed promising and showed substantial improvements on various benchmarks. They thrived on their ability to capture long-term dependencies and intricate relationships in the data. However, their complexity came at a cost. They were computationally expensive, memory-intensive, and often lacked the interpretability that traditional models offered.

Despite these advancements, an ideal solution for long-term time series forecasting remained elusive. The tug-of-war between simplicity and performance, between interpretability and computational efficiency, continued. It's within this backdrop that TiDE emerges as a potential game-changer.

As we progress through this post, we will delve deep into TiDE's architecture, its strengths, its theoretical underpinnings, and how it compares to existing models in the forecasting arena. Stay with us on this journey through the intricacies of time series forecasting and the innovations TiDE brings to the table.

Introducing TiDE

Enter TiDE, or Time-series Dense Encoder, a beacon of hope in the vast sea of time series forecasting models. At its core, TiDE is a representation of simplicity and efficiency fused with state-of-the-art performance. While the

world was tilting towards increasingly complex neural network models, TiDE took a step back, rooting itself in the foundational principle that sometimes, less is indeed more.

The architecture of TiDE is an embodiment of this philosophy. It leverages a straightforward MLP (Multi-Layer Perceptron)-based encoder-decoder model. For those uninitiated, MLPs are feedforward neural networks comprised of multiple layers of nodes, or “neurons”, where each layer is fully connected to the next. They have been around for decades, and while they may lack the flashiness of some newer models, their power and versatility have withstood the test of time.

In the case of TiDE, the encoder takes in past time series data and associated covariates, processing and transforming this information into a condensed hidden representation. This hidden representation, rich with patterns and nuances of the past data, is then handed over to the decoder. The decoder’s job is to unravel this representation, mapping it into future predictions.

The choice of an MLP-based encoder-decoder setup might raise eyebrows, especially when more intricate models like Transformers have been making waves. However, this is precisely where TiDE stands out. Instead of getting entangled in the web of complexity, it challenges the status quo, demonstrating that for the task of long-term forecasting, simple MLPs might just be the dark horse many overlooked.

As we delve further, we’ll uncover how exactly TiDE manages to match or even exceed the performance of its more sophisticated counterparts, and why this breakthrough might just be the paradigm shift the forecasting community has been yearning for.

How Does TiDE Work?

At the heart of TiDE lies its encoder-decoder architecture, a model structure that has proven effective across a range of deep learning applications, from machine translation to image captioning. For time series forecasting, this setup effectively captures the nuances of past data and translates them into actionable insights about the future.

Encoding the Past

TiDE begins its magic by first encoding the past of a time series along with any associated covariates. These covariates could be any external factors or indicators that influence the time series in question. The encoding process is executed using dense MLPs. Each layer in this MLP serves as a transformation, capturing increasingly abstract representations of the input data. By the end of the encoding phase, the model distills the past time series and covariates into a dense hidden representation, a vector filled with learned features that best describe the data's patterns and relationships.

Decoding into the Future

Once the past is neatly packaged into this hidden representation, the decoding process commences. This phase is responsible for predicting future values based on the extracted features. Another set of dense MLPs is used for this purpose, taking the hidden representation and generating a series of future predictions. But TiDE doesn't stop there.

Temporal Decoder — The Game Changer

A unique component of TiDE is the introduction of a temporal decoder. While the primary decoder produces a basic forecast, the temporal decoder refines these predictions by adapting them to future covariates. This is crucial because, in real-world scenarios, external influences can cause drastic shifts in time series data. The temporal decoder ensures that these potential future changes are considered, allowing for more accurate and adaptable forecasts.

The Significance of Residual Connections

Another noteworthy aspect of TiDE's architecture is the use of residual connections. These are essentially shortcuts that bypass one or more layers in the neural network. By doing so, they facilitate smoother gradient flow during training, making it easier for the model to learn and reducing the risk of vanishing gradient problems. Residual connections also introduce a form of model regularization, potentially preventing overfitting.

A testament to the effectiveness of both the temporal decoder and residual connections is the series of ablation studies conducted in the paper. These studies systematically removed or altered certain features of the model to

gauge their impact on performance. The findings were clear: both the temporal decoder and residual connections significantly contribute to TiDE's impressive forecasting capabilities.

In essence, TiDE is not just about simplicity but also about judiciously chosen innovations that maximize forecasting accuracy. It's a marriage of well-understood deep learning practices with new approaches tailored for the intricacies of time series forecasting.

TiDE vs. Other Approaches

The vast landscape of time series forecasting models is marked by an array of sophisticated techniques, each vying for supremacy. Among the most celebrated in recent times is the Transformer model, renowned for its self-attention mechanism that allows it to weigh the importance of different time steps in the data. But how does TiDE, with its emphasis on simplicity, stack up against such a heavyweight?

Efficiency and Memory Benefits

One of the most striking advantages of TiDE over Transformer models lies in its sheer efficiency. In the arena of long-term forecasting, time is of the essence, both in terms of model training and inference. Here, TiDE takes the crown, proving to be 5–10 times faster than some of the best Transformer models out there. Such speed gains are not just about faster results but also about enabling more iterative experimentation, enhancing the overall research and application process.

Memory consumption is another crucial aspect, especially when dealing with extensive time series data. While Transformer models, with their intricate self-attention mechanisms, often scale quadratically with sequence length, leading to escalated memory requirements, TiDE exhibits linear scaling. This difference becomes particularly pronounced for longer sequences, where Transformers may stumble upon memory constraints, but TiDE continues unabated.

Performance Metrics — The Real Test

Beyond efficiency and memory, the ultimate litmus test for any forecasting model is its performance. On this front, TiDE doesn't merely compete but

often surpasses Transformer models. Particularly on the M5 forecasting dataset, which is a benchmark for evaluating forecasting methods using real-world retail data, TiDE demonstrates its prowess. It not only handles the covariates of this dataset adeptly but also manages to outpace DeepAR, a popular forecasting method, by a significant 20% margin on this benchmark.

One cannot also overlook the 10% reduction in Mean Squared Error (MSE) that TiDE achieves on the largest dataset in comparison to the best prior methods. Given that MSE is a measure of the average squared difference between the estimated values and the actual value, a 10% reduction is not just a statistical victory but a massive leap in forecasting accuracy.

The landscape of forecasting models is vast and varied. Transformers, with their self-attention mechanisms, had long held the limelight for their ability to capture intricate patterns. However, TiDE's emergence challenges this status quo, proving that with the right architecture and design choices, simpler models can not only compete but often surpass their more sophisticated counterparts. The evidence isn't just in theoretical design but in empirical results, showcasing TiDE as a formidable contender in the world of time series forecasting.

Theoretical Insights

In the world of data science and modeling, empirical performance is undoubtedly vital. However, for a deeper understanding and trust in any model, one must venture into its theoretical underpinnings. TiDE's success isn't merely an empirical marvel; it's undergirded by rigorous theoretical analysis, especially when considering its linear variant.

Linear Version of TiDE

The paper presents an intriguing exploration of a linear version of TiDE. But why linear? Linear models, in their simplicity, offer clearer analytical insights, acting as a bridge to understand more complex non-linear

mechanisms. By studying this stripped-down version, the researchers provided an invaluable window into the workings of the model.

Optimal Performance for Linear Dynamical Systems

Diving deeper, the linear TiDE was put to test against linear dynamical systems (LDS). For the uninitiated, LDS is a mathematical formulation that describes the evolution of a set of variables over time, driven by linear relationships. The big revelation? Under specific assumptions, the linear TiDE was found to be nearly optimal in its forecasting performance for such systems. This near-optimality isn't a trivial finding. It signifies that within the constraints of linear modeling and the stated assumptions, TiDE pushes the boundaries of forecasting accuracy, getting as close as possible to the "best" one might achieve.

TiDE's empirical success is undeniably impressive. Still, it's the theoretical insights, like the near-optimal performance of its linear version for LDS, that lend it a robust foundation. Such theoretical explorations offer a reassuring nod to the model's design and illuminate the pathways for future enhancements. It's a testament to the adage — to truly advance in the realm of complex systems, one often needs to revisit and understand the basics, in this case, linear dynamics.

Conclusion

In the ever-evolving landscape of time series forecasting, TiDE has emerged as a beacon of promise. As we navigate through the challenges and intricacies of predicting the future based on past data, this new approach underscores the significance of combining the power of modern machine learning with the elegance of simplicity.

TiDE's breakthrough is not just in its performance metrics, though they are undeniably impressive. It is in the underlying philosophy that simpler models, when designed with care and understanding, can not only compete with but even surpass their more complex counterparts. In an age where the

complexity of models often gets conflated with their efficacy, TiDE serves as a refreshing reminder that less can indeed be more.

This paradigm shift towards simpler models in the realm of time series forecasting doesn't just have computational and memory benefits; it might be the key to unlocking deeper insights and more reliable predictions. The transformative power of such simplicity has the potential to make advanced forecasting more accessible, more interpretable, and more effective.

However, as with all new methodologies, the real test of TiDE will be in its widespread adoption and its ability to stand the test of time. Encouragingly, the early signs are positive. Its accomplishments on benchmark datasets are a testament to its potential, and the growing interest in the research community suggests a bright future ahead.

In closing, TiDE's emergence in the forecasting domain beckons us to revisit and possibly rethink our reliance on increasingly intricate models. As we stand on the cusp of what might be a new era in time series modeling, the challenge and the opportunity lie in harnessing the power of simple, effective, and efficient models like TiDE. The future, it seems, might very well be simpler than we've often imagined.

Link to the original paper

<https://arxiv.org/pdf/2304.08424.pdf>

Data Science

Demand Forecasting

Vertex AI

Neural Networks

Encoder Decoder