Time2Vec and the Nuances of Temporal Representation

Efstathios Kotsis, Michael Darmanis
Nektarios Christou, Vasileios Venieris

September 12, 2023

Time2Vec, as proposed by Borealis et al.[1], provides a vector representation of time for machine learning models. However, its application across various domains can be susceptible to misinterpretation. This essay provides a rough overview of the strengths, as well as the potential pitfalls, of Time2Vec.

[1] Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time, 2019

Introduction

In the presented paper, the authors introduce a unique learnable vector representation of time, eliminating the need for hand-crafted time representations. This representation employs a feed-forward layer with sine activations[2] to process time data. Given its vector nature, it seamlessly integrates with other deep neural network methodologies. The authors adeptly underscore the significance of time data across various domains and position their solution as a distinct approach compared to existing literature. Drawing inspiration from Fourier analysis, they substantiate their representation. To validate their claims, they employ both fabricated and real-world time series datasets, supplemented by ablation studies to justify their design choices.

[2] authors state that other periodic-like functions will yield similar results

Mathematical Framework

At its core, Time2Vec draws inspiration from the basic idea of the Fourier Series, using the frequency and phase terms as training parameters. Time2Vec's mathematical framework represents time through the following formulation

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0 \\ \mathcal{F}(\omega_i \tau + \varphi_i), & \text{if } 1 \leq i \leq k, \end{cases}$$

where $t2v(\tau)[i]$ denotes the representation of time τ for the i-th component, ω_i represents a frequency term for the i-th component, φ_i signifies a phase term for the i-th component and $\mathcal{F}$ is a periodic function encapsulating repeating patterns in time.

In essence, this formula captures both linear and periodic patterns. The linear term addresses trends, while the periodic function

captures recurring sequences in time, thereby producing a comprehensive time representation.

Incorporating Time2Vec in Predictive Modeling

Integrating Time2Vec can optimise models, enabling them to discern temporal patterns and overarching trends. The process begins with Time Embedding, transforming raw temporal data into intricate embeddings. This is followed by the Model Integration phase, where time-based embeddings are integrated into predictive models. Once integrated, the model generates predictions in the prediction phase. Post prediction, the model is fine-tuned in the training phase, and finally, optimisation is performed to enhance the model's accuracy and efficiency.

Strengths

One of the most compelling attributes of Time2Vec is its rich representation of time. It can transform mere timestamps into multi-dimensional vectors that capture both linear and periodic nuances of time. This is akin to how humans perceive time, not just as a linear progression but in terms of its cyclical nature, such as days, weeks, and years. Furthermore, its systematic way of encoding time offers versatility across different time series datasets, eliminating the need for handcrafted features that may vary between datasets. Additionally, Time2Vec's adaptability shines through its integration capabilities. It can seamlessly blend with various machine learning models, from transformers to LSTMs, enriching them with a deeper understanding of time and potentially enhancing their predictive accuracy.

Concerns and Clarifications

However, Time2Vec's application can raise concerns. While some studies[3] seem to use time-series-values as inputs rather than actual temporal information, Time2Vec seems more like feature representation enrichment rather than true positional encoding. The term "positional encoding" can lead to confusion, especially when architectures are layered in complex models like Transformers. Furthermore, if Time2Vec is applied to a time index, it should be a global one, as a local index might introduce inconsistencies.

Moreover, the paper offers very little that is new when compared to say Vaswani et al.[4,5] (see section 3.5). In addition, the authors compare to a baseline that seems to consist of passing time as a float.

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023

[5] the authors acknowledge this work several times

This seems like a very weak baseline; there are any number of other reasonable ways to encode this.

Last but not least, the paper's reference to using up to 64 sinusoids implies transforming a single time point into multiple features, capturing various cyclical patterns in time data. However, this approach appears more like an ad-hoc solution rather than a well-grounded explanation. The authors have removed their code[6], making it impossible to validate and understand the intricacies of their proposed process.

[6] the provided link in the paper is `https://github.com/borealisai/Time2Vec`

Conclusion

Time2Vec can potentially offer an innovative approach to representing time in machine learning models. Its ability to transform a single time point using multiple sinusoids provides a richer representation of time. Despite the potential of this work to be groundbreaking, however, it falls short in certain areas. It lacks rigorous theoretical motivation and a comprehensive analysis of the experimental results. A more in-depth exploration of the model's capabilities, through Fourier analysis and a comparative table showcasing test accuracy or recall against other state-of-the-art algorithms, would boost the its credibility. Also, clearer guidelines for its application would be greatly beneficial for its reproducibility and its overall acceptance.

Acknowledgements

This report was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. A template that can be used to format documents with this look and feel has been released under the permissive APACHE LICENSE 2.0, and can be found online at `https://tufte-latex.github.io/tufte-latex/`.

References

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time, 2019.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.