8-1-2021

Predicting (Economic) Trends: Why Signature Method in Machine Learning

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Chon Van Le
International University of Ho-Chi-Minh City, lvchon@hcmiu.edu.vn

Comments:
Technical Report: UTEP-CS-21-69

Predicting (Economic) Trends: Why Signature Method in Machine Learning

Vladik Kreinovich and Chon Van Le

Abstract In many practical situations, we can predict the trend – i.e., how the system will change – but we cannot predict the exact timing of this change: this timing may depend on many unpredictable factors. For example, we may be sure that the economy will recover, but how fast it will recover may depend on the status of the pandemic, on the weather-affected agriculture input, etc. In such trend predictions, one of the most efficient methods is signature method, which is based on applying machine learning techniques to several special characteristics of the corresponding time series. In this paper, we provide an explanation for the empirical success of the signature method.

1 What Is Signature Method

Predictions are important. Prediction is one of the main objective of science, and economic predictions are one of the main objectives of econometrics. We want to predict what will happen to economy if we do not interfere. If we do not like this prediction, we need to decide what action to take to improve the economy – and for that, we need to be able to predict what will happen if we undertake different actions.

In some cases, we can predict the trend but not the timing. In some cases, we are able to predict exactly what will happen at different future moments of time – e.g., what will be the GDP next year. In other cases, there

Vladik Kreinovich
University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu

Chon Van Le
International University of Ho-Chi-Minh City, Vietnam, e-mail: lvchon@hcmiu.edu.vn

are too many factors affecting the situation: for example, in agriculture, a lot depends on current weather patterns. In many such situations, it is possible to predict the trend – but not the timing.

For example, we may predict that under an appropriate fiscal policy, the economy will improve, that at some future moment of time, the GDP will grow by 20% and the unemployment will decrease to half of the current value – but we cannot predict whether this will happen in 3 years or in 6 years. In such situations, what is important for prediction is also not when exactly different events happen in the past – the exact timing is too much affected by random events to be useful – but rather what was the state of economy at different moments in the past. What is important, e.g., is that when the crops decreased by 30%, the unemployment grew by 20% – this show how the country's economy depends on its agriculture sector – but it does not matter whether this happened 3 or 4 years ago.

Let us describe this situation in precise terms. The state of the economy at each moment of time t can be characterized by the values of several characteristics $x = (x_1, \ldots, x_n)$ at this moment of time. In these terms, what we know is how, in the past, the state of the economy changed, i.e., what were the values $x_i(t)$ of all these characteristics for all the moments of time t starting from the moment T_0 when we started recording these values to the current moment T. Based on this information, we want to predict how the state $x(t)$ will change in the future, for moments $t > T$.

This prediction should not depend on the exact duration of each state, only on the general trend. In other words, we should get the same prediction based on the actual values $x(t)$ and on the values $X(t) = x(\tau(t))$ for any increasing function $\tau(t)$.

- The processes may have been slower than they actually were – in this case, we may have $\tau(t) = c \cdot t$ for some $c < 1$.
- The processes may have been faster than they actually were – in this case, we may have $\tau(t) = c \cdot t$ for some $c > 1$.
- The processes may have been slower at some periods of time and faster and other – in this case, the dependence $\tau(t)$ is nonlinear.

In all these cases, whether we use the original records $x(t)$ or re-scaled records $X(t) = x(\tau(t))$, we should get the exact same predicted trend.

Additional requirement: predictions often depend only on changes, not on the initial state. Another reasonable assumption is that the trend's predictions should depend only on the relative changes, not on the actual initial state. For example, it is important to know that the GDP declined by 20% and the unemployment increased by 80%, but it does not matter that much whether we talk about a big country with a large population and large GDP or a smaller one with smaller population and a smaller GDP.

In precise terms, this means that what is important is not the actual values $x_i(t)$, but rather the ratios $x_i(t)/x_i(T_0)$ describing how these values changed.

In other words, the predictions should remain the same whether we use the values $x_i(t)$ or the values $c_i \cdot v_i(t)$ for some constants c_i. In many economic situations, it is convenient to use logarithms $v_i(t) = \ln(x_i(t))$ of the actual values. The logarithm $\ln(c_i \cdot x_i(t))$ of each re-scaled value is equal to the sum $\ln(x_i(t)) + \ln(c_i) = v_i + C_i$, where $C_i \stackrel{\text{def}}{=} \ln(c_i)$. In these terms, we should predict the same trends whether we use the original dependence $v_i(t)$ or the re-scaled dependence $v_i(t) + C_i$.

Signature method: a brief description. In such situations, it turned out to be very efficient to replace the original description $v_i(t)$ with the so-called *signature*, i.e., with the sequence of the values

$$s_{i_1} = \int \dot{v}_{i_1}(t_1)\,\mathrm{d}t_1,$$

where, as usual, $\dot{v}_i(t)$ indicates the derivative,

$$s_{i_1,i_2} = \int_{T_0 \le t_1 \le t_2 \le T} \dot{v}_{i_1}(t_1) \cdot \dot{v}_{i_2}(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2,$$

$$\ldots$$

$$s_{i_1,\ldots,i_k} = \int_{T_0 \le t_1 \le \ldots \le t_k \le T} \dot{v}_{i_1}(t_1) \cdot \ldots \cdot \dot{v}_{i_k}(t_k)\,\mathrm{d}t_1\,\ldots\,\mathrm{d}t_k, \qquad (1)$$

$$\ldots;$$

This idea especially useful when we use machine learning: trend predictions based on the signature are much more accurate than if we apply deep learning to the actual record $v_i(t)$; see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20].

Known fact: signature has the desired invariance properties. If we add the same constant vector C_i to all the values $v_i(t)$, the derivatives will not change, and thus, the signature values will not change. Hence, the signature values remain the same whether we use the original dependence $v_i(t)$ or the re-scaled dependence $v_i(t) + C_i$.

Similarly, since $\dot{v}_i(t) \cdot \mathrm{d}t_i = \mathrm{d}v_i(t_i)$, each signature value can be represented in the following equivalent form

$$s_{i_1,\ldots,i_k} = \int_{T_0 \le t_1 \le \ldots \le t_k \le T} \mathrm{d}v_{i_1}(t_1)\,\ldots\,\mathrm{d}v_{i_k}(t_k). \qquad (2)$$

From this expression, it is clear that this value does not change if we re-scale time, i.e., replace the original dependence $v_i(t)$ with the re-scaled dependence $v_i(\tau(t))$.

But why signature? A natural question is: why signature and not other characteristics? In this paper, we provide a possible explanation of why it makes sense to use signature.

2 Why Signature: An Explanation

What we want: reminder. We want to find characteristics depending on all the values $v_i(t)$ for all i and t, characteristics that will be invariant under re-scaling of time $t \mapsto \tau(t)$ and under re-scaling of the values $v_i(t) \mapsto v_i(t) + C_i$.

How to achieve independence with respect to re-scaling of values $v_i(t)$. Independence on re-scaling of values can be achieved if we only consider dependence on the derivatives $\dot{v}_i(t)$.

Indeed, derivatives do not change under such re-scaling. Vice versa, once we know the derivatives, we can reconstruct the differences

$$v_i(t) - v_i(T_0) = \int_{T_0}^{t} \dot{v}_i(s)\, ds$$

and thus, indeed reconstruct the value $v_i(t)$ modulo such re-scaling.

Resulting reformulation of the problem. So, to make sure that our characteristics do not change under re-scaling of values $v_i(t)$, it makes sense to consider characteristics depending on all the values $\dot{v}_i(t)$ for all i and t, characteristics that will be invariant under re-scaling of time $t \mapsto \tau(t)$.

How can we describe general characteristics. Most dependencies are smooth. There are seemingly non-smooth processes like phase transition, but in reality, they are smooth too: just the time scale becomes different. Similarly in economics, most characteristics smoothly change with time. Sometimes the changes are fast and this seem abrupt and discontinuous, but in reality, they are reasonably smooth.

In general, a sufficiently smooth function $b = f(a_1, \ldots, a_n)$ of n inputs a_i can be described by its Taylor series:

$$b = b_0 + \sum_{i_1=1}^{n} b_{i_1} \cdot a_{i_1} + \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} b_{i_1,i_2} \cdot a_{i_1} \cdot a_{i_2} + \ldots +$$

$$\sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} b_{i_1,\ldots,i_k} \cdot a_{i_1} \cdot \ldots \cdot a_{i_k} + \ldots$$

In our case, the unknowns a_i are the values $\dot{v}_i(t)$ corresponding to different values of i and t. Theoretically, there are infinitely many moments of time t. However, of course, in practice, we only have values $v_i(t)$ corresponding to finitely many moments of time $t_1 < \ldots < t_m$. In this case, as approximations to derivatives, we have finite differences

$$\dot{v}_i(t_\ell) \approx \delta v_i(t_\ell) \stackrel{\text{def}}{=} \frac{v_i(t_{\ell+1}) - v_i(t_\ell)}{\Delta t_\ell}, \quad (3)$$

where we denoted

$$\Delta t_\ell \stackrel{\text{def}}{=} t_{\ell+1} - t_\ell. \tag{4}$$

By definition of the derivative, when Δt_ℓ tends to 0, the finite difference $\delta v_i(t_\ell)$ tends to the derivative. Thus, when the differences Δt_ℓ are sufficiently small – i.e., when the moments t_ℓ are sufficiently close to each other – we can safely assume, for all practical purposes, that the differences $\delta v_i(t_\ell)$ are equal to the corresponding derivatives.

In terms of these variables $\delta v_i(t_\ell)$ corresponding to difference values of i and ℓ, the general Taylor series expansion of a characteristic

$$s(\delta v_1(t_1), \ldots, \delta v_1(t_m), \delta v_2(t_1), \ldots)$$

takes the form

$$s = S_0 + \sum_{i_1=1}^{n} S_{i_1} + \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} S_{i_1,i_2} + \ldots + \sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} S_{i_1,\ldots,i_k} + \ldots, \tag{5}$$

where we denoted

$$S_{i_1,\ldots,i_k} \stackrel{\text{def}}{=} \sum_{\ell_1=1}^{m} \ldots \sum_{\ell_k=1}^{m} b_{i_1,\ldots,i_k,\ell_1,\ldots,\ell_k} \cdot \delta v_{i_1}(t_{\ell_1}) \cdot \ldots \cdot \delta v_{i_k}(t_{\ell_k}). \tag{6}$$

Dividing and multiplying each terms in the sum by the product $\Delta t_{\ell_1} \cdot \ldots \cdot \Delta t_{\ell_k}$, we conclude that

$$S_{i_1,\ldots,i_k} =$$

$$\sum_{\ell_1=1}^{m} \ldots \sum_{\ell_k=1}^{m} B_{i_1,\ldots,i_k}(t_{\ell_1}, \ldots, t_{\ell_k}) \cdot \delta v_{i_1}(t_{\ell_1}) \cdot \ldots \delta v_{i_k}(t_{\ell_k}) \cdot \Delta t_{\ell_1} \cdot \ldots \cdot \Delta t_{\ell_k}, \tag{7}$$

where we denoted

$$B_{i_1,\ldots,i_k}(t_{\ell_1}, \ldots, t_{\ell_k}) \stackrel{\text{def}}{=} \frac{b_{i_1,\ldots,i_k,\ell_1,\ldots,\ell_k}}{\Delta t_{\ell_1} \cdot \ldots \cdot \Delta t_{\ell_k}}. \tag{8}$$

Taking into account that, for the practical purposes, the differences $\delta v_{i_k}(t_{\ell_k})$ are equal to the corresponding derivatives $\dot{v}_i(t_\ell)$, we conclude that

$$S_{i_1,\ldots,i_k} =$$

$$\sum_{\ell_1=1}^{m} \ldots \sum_{\ell_k=1}^{m} B_{i_1,\ldots,i_k}(t_{\ell_1}, \ldots, t_{\ell_k}) \cdot \dot{v}_{i_1}(t_{\ell_1}) \cdot \ldots \dot{v}_{i_k}(t_{\ell_k}) \cdot \Delta t_{\ell_1} \cdot \ldots \cdot \Delta t_{\ell_k}. \tag{9}$$

The expression (9) is nothing else but the integral sum for the integral

$$I_{i_1,\ldots,i_k} =$$

$$\int_{t_1=T_0}^{T} \ldots \int_{t_k=T_0}^{T} B_{i_1,\ldots,i_k}(t_1,\ldots,t_k)\cdot \dot v_{i_1}(t_1)\cdot\ldots\dot v_{i_k}(t_k)\,\mathrm{d}t_1\,\ldots\,\mathrm{d}t_k. \quad (10)$$

When the differences Δt_ℓ are small, for practical purposes, the integral sum $S_{i_1,\ldots,i_k}$ is equal to the integral $I_{i_1,\ldots,i_k}$.

So, we conclude that a generic representation of a characteristic s has the form

$$s = S_0 + \sum_{i_1=1}^{n} I_{i_1} + \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} I_{i_1,i_2} + \ldots + \sum_{i_1=1}^{n}\ldots\sum_{i_k=1}^{n} I_{i_1,\ldots,i_k} + \ldots, \quad (11)$$

where the integrals $I_{i_1,\ldots,i_k}$ are determined by the formula (10).

When is the general expression (11) invariant under re-scaling of time? To come up with a description of all possible invariant characteristics, we need to find out which expressions (11) are invariant under re-scaling of time $t \mapsto \tau(t)$.

To find out which expressions (11) are thus invariant, let us first take into account – as we did in the previous section – that $\dot v_i(t)\cdot \mathrm{d}t_i = \mathrm{d}v_i(t)$. Then, the expression (10) takes the equivalent form

$$I_{i_1,\ldots,i_k} = \int_{t_1=T_0}^{T} \ldots \int_{t_k=T_0}^{T} B_{i_1,\ldots,i_k}(t_1,\ldots,t_k)\,\mathrm{d}v_{i_1}(t_1)\,\ldots\mathrm{d}v_{i_k}(t_k). \quad (12)$$

If we re-scale time, we get an expression

$$I^\tau_{i_1,\ldots,i_k} = \int_{t_1=T_0}^{T} \ldots \int_{t_k=T_0}^{T} B_{i_1,\ldots,i_k}(\tau(t_1),\ldots,\tau(t_k))\,\mathrm{d}v_{i_1}(t_1)\,\ldots\mathrm{d}v_{i_k}(t_k). \quad (13)$$

Invariance means that we should have $I_{i_1,\ldots,i_k} = I^\tau_{i_1,\ldots,i_k}$ for all possible functions $\dot v_i(t)$. This means that the coefficients $B_{i_1,\ldots,i_k}$ should be the same in both cases, i.e., that we should have

$$B_{i_1,\ldots,i_k}(t_1,\ldots,t_k) = B_{i_1,\ldots,i_k}(\tau(t_1),\ldots,\tau(t_k)) \quad (14)$$

for all possible increasing functions $\tau(t)$.

One can easily check that every two tuples $t_1 < \ldots < t_k$ and $t'_1 < \ldots < t'_k$ can be obtained from each other by some increasing function. Thus, for all such tuples, the value $B_{i_1,\ldots,i_k}(t_1,\ldots,t_k)$ is the same; we will denote it by $\mathbf{B}_{i_1,\ldots,i_k}$.

Similarly, for any other ordering of the moments $t_{\pi(1)} < \ldots < t_{\pi(k)}$ corresponding for any permutation $\pi : \{1,\ldots,k\} \mapsto \{1,\ldots,k\}$, the value $B_{i_1,\ldots,i_k}(t_1,\ldots,t_k)$ depends only on this permutation and is the same for all tuples $(t_1,\ldots,t_k)$ for which this ordering is true. We will denote this common value by $\mathbf{B}_{i_{\pi(1)},\ldots,i_{\pi(k)}}$.

The whole domain of all possible tuples $(t_1,\ldots,t_k)$ – over which the integral (10) is computed – can be divided into sub-domains corresponding to different orders between t_i. For example, for $k = 2$, we divide the domain of all the pairs (t_1, t_2) into two sub-domains:

- the set of all the pairs for which $t_1 < t_2$ that corresponds to the identity permutation $\pi(i) = i$, and
- the set of all the pairs for which $t_2 < t_1$ that corresponds to swap $\pi(1) = 2$ and $\pi(2) = 1$.

On the sub-domain $D_{k,\pi}$ corresponding to the identity permutation π, the value $B_{i_1,\ldots,i_k}(t_1,\ldots,t_k)$ is a constant $\mathbf{B}_{i_1,\ldots,i_k}$, so the integral over this sub-domain has the form

$$\int_{D_{k,\pi}} B_{i_1,\ldots,i_k}(t_1,\ldots,t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \ldots \dot{v}_{i_k}(t_k)\, \mathrm{d}t_1 \ldots \mathrm{d}t_k =$$

$$\mathbf{B}_{i_1,\ldots,i_k} \cdot \int_{T_0 < t_1 < \ldots < t_k < T} \mathrm{d}v_1(t_1) \ldots \mathrm{d}v_{i_k}(t_k). \tag{14}$$

As usual, in the integration, the integral over measure-0 parts corresponding to possible equalities such as $t_1 = T_0$, $t_1 = t_2$, etc. is 0, so we can say that

$$\int_{D_{k,\pi}} B_{i_1,\ldots,i_k}(t_1,\ldots,t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \ldots \dot{v}_{i_k}(t_k)\, \mathrm{d}t_1 \ldots \mathrm{d}t_k =$$

$$\mathbf{B}_{i_1,\ldots,i_k} \cdot \int_{T_0 \le t_1 \le \ldots \le t_k \le T} \mathrm{d}v_1(t_1) \ldots \mathrm{d}v_{i_k}(t_k). \tag{15}$$

The integral in the right-hand side is exactly one of the signature values $s_{i_1,\ldots,i_k}$, so we get

$$\int_{D_{k,\pi}} B_{i_1,\ldots,i_k}(t_1,\ldots,t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \ldots \dot{v}_{i_k}(t_k)\, \mathrm{d}t_1 \ldots \mathrm{d}t_k =$$

$$\mathbf{B}_{i_1,\ldots,i_k} \cdot s_{i_1,\ldots,i_k}. \tag{16}$$

The integral $I_{i_1,\ldots,i_k}$ over the whole set of tuples $(t_1,\ldots,t_k)$ is equal to the sum of the integrals over all sub-domains corresponding to different permutations, so we get

$$I_{i_1,\ldots,i_k} = \sum_{\pi} \mathbf{B}_{i_{\pi(1)},\ldots,i_{\pi(k)}} \cdot s_{i_{\pi(1)},\ldots,i_{\pi(k)}}. \tag{17}$$

So, in the invariant case, the general expression (11) for a characteristic takes the form

$$s = S_0 + \sum_{i_1=1}^{n} \mathbf{B}_{i_1} \cdot s_{i_1} + \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \mathbf{B}_{i_1,i_2} \cdot s_{i_1,s_2} + \ldots +$$

$$\sum_{i_1=1}^{n} \dots \sum_{i_k=1}^{n} \mathbf{B}_{i_1,\dots,i_k} \cdot s_{i_1,\dots,i_k} + \dots, \tag{18}$$

i.e., it is a linear combination of the signature values.

Conclusion. It was known that signature values are invariant. What we have shown is that any other invariant characteristic is nothing else but a linear combination of signature values. In this sense, signature values is all that we can extract from the data, they provide full information about the inputs.

This explains why signature values are so successful – since they provide full information about the input.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

1. C. Amendola, P. Friz, and B. Sturmfels, *Varieties of Signature Tensors*, arXiv preprint arXiv:1804.08325, 2018.
2. I. P. Arribas, *Derivatives pricing using signature payoffs*, arXiv preprint arXiv:1809.09466, 2018.
3. I. P. Arribas, G. M. Goodwin, J. R. Geddes, T. Lyons, and K. E. Saunders, "A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder", *Translational Psychiatry*, 2018, Vol. 8, No. 1, Paper 274, https://doi.org/10.1038/s41398-018-0334-0
4. H. Boedihardjo, X. Geng, T. Lyons, and D. Yang, "The signature of a rough path: uniqueness", *Advances in Mathematics*, 2016, Vol. 293, pp. 720–737, https://doi.org/10.1016/j.aim.2016.02.011
5. H. Boedihardjo, H. Ni, and Z. Qian, "Uniqueness of signature for simple curves", *J. Funct. Anal.*, 2014, Vol. 267, No. 6, pp. 1778–1806.
6. K. T. Chen, "Integration of paths – A faithful representation of paths by non-commutative formal power series", *Transactions of the American Mathematical Society*, 1958, Vol. 89, No. 2, pp. 395–407, https://doi.org/10.2307/1993193
7. I. Chevyrev and A. Kormilitzin, *A primer on the signature method in machine learning*, arXiv preprint arXiv:1603.03788, 2016.
8. J. Field, L. G. Gyurkó, M. Kontkowski, and T. Lyons. *Extracting information from the signature of a financial data stream*, arXiv preprint arXiv:1307.7244, 2014.

9. X. Geng, *Reconstruction for the signature of a rough path*, arXiv preprint arXiv:1508.06890, 2015.
10. B. Hambly and T. Lyons, "Uniqueness for the signature of a path of bounded variation and the reduced path group", *Annals of Mathematics*, 2010, Vol. 171, pp. 109–167, https://doi.org/10.4007/annals.2010.171.109
11. A. Kormilitzin, *The signature method in machine learning*, https://github.com/kormilitzin/
12. T. Lyons, *Rough paths, Signatures and the modelling of functions on streams*, arXiv preprint arXiv:14054537, 2014.
13. T. Lyons and Z. Qian, *System Control and Rough Paths*, Oxford University Press, New York, 2007.
14. T. Lyons and W. Xu, *Hyperbolic development and inversion of signature*, arXiv preprint arXiv:1507.00286, 2015.
15. T. Lyons and W. Xu, *Inverting the signature of a path*, arXiv preprint arXiv:1406.7833, 2015.
16. P. J. Moore, T. J. Lyons, and J. Gallacher, "Using path signatures to predict a diagnosis of Alzheimer's disease", *PLoS One*, 2019, Vol. 14, No. 9, Paper e0222212, doi:10.1371/journal.pone.0222212
17. H. Ni, *A multi-dimensional stream and its signature representation*, arXiv preprint arXiv:1509.03346, 2015.
18. M. Pfeffer, A. Seigal, and B. Sturmfels, *Learning paths from signature tensors*, arXiv preprint arXiv:1809.01588, 2018.
19. N. Suguira and S. Hosoda, "Machine Learning Technique Using the Signature Method for Automated Quality Control of Argo Profiles", *Earth and Space Science*, 2020, Vol. 7, No. 9, Paper e2019EA001019, https://doi.org/10.1029/2019EA001019
20. W. Yang, T. Lyons, H. Ni, C. Schmid, and L. Jin, *Developing the Path Signature Methodology and its Application to Landmark-Based Human Action Recognition*, arXiv preprint arXiv:1707.03993, 2019.