Proceedings
Theory of Biomathematics and Its Applications XVI

Evaluation of periodicity behind time series data.

Hirotaka Kanazawa [1*] • Junichi Takagi [2,3*]

[1]International Institute of Advanced Studies, [2]National Institute of Polar Research,
[3]Graduate School of Informatics, Kyoto University
[*]Both authors contributed to this study equally.

Abstract

To detect a causality has been one of the most fundamental interests in natural science. In these days, Empirical Dynamic Modeling (EDM), especially Convergent Cross Mapping (CCM), is focused on as a new method detecting a causality. Despite of these many researches by CCM, it seems that there are less theoretical studies about CCM. In this proceedings, we would like to give some considerations of mathematical potential which CCM may has. As one suggestion, we propose the idea of *convergent skill*, which reflects a changing neighbourhood between time series data ,and by which we may be able to evaluate for linearity or periodicity behind ones.

1 Background

To detect a causality has been one of the most fundamental interests in natural science. Therefore, numerous attempts have been made to detect a causality from a time series data. In these days, Empirical Dynamic Modeling (EDM)[1], especially Convergent Cross Mapping (CCM), is focused on as a new method detecting a causality and there are some package to use this method[2, 3, 4] . Because of this, CCM is regarded as an algorithm to detect a causality in general. It is, of course, not incorrect but insufficient. CCM consists in many mathematical concepts and theories, so there are possibilities that we can know mathematically the properties which a time-series data has in addition to a causality. In addition to this, considering the fact that mathematical science is interested in the existence of a causality and, at the same time, also interested in the mathematical model behind a time series data or in predicting the future using the model, we needed to study CCM itself as one mathematical object. In other words, it seems that CCM needs to be studied not merely as an algorithm to detect a causality but as a mathematical theory by own itself.

In this proceeding, we would like to give some considerations of mathematical potential which CCM may has. This argument will offer the key to mathematical understanding of CCM. Before entering into the argument, we shall call some attention. We follow the terms, notations and mathematical conditions to common convention in CCM, especially, Sugihara et al. (2012). In this proceedings, therefore, we use the almost all notations without giving to the definitions.

2 Convergent skill

Given that the time series data $\boldsymbol{X}(=\{x(t)\})$ affects $\boldsymbol{Y}(=\{y(t)\})$, there is a causality of $x(t)$ and $y(t)$. In this case, there is a mapping between manifold $\boldsymbol{M}_X$ and $\boldsymbol{M}_Y$. Let us denote this mapping as ϕ. For simplicity, we assume that the function ϕ is C^∞-class and differomorphic.

$$\begin{array}{ccc} \boldsymbol{M}_X & \xrightarrow{\phi} & \boldsymbol{M}_Y \\ \Cup & & \Cup \\ \underline{\boldsymbol{x}}(t) & \longmapsto & \underline{\boldsymbol{y}}(t) \end{array}$$

Let's consider that neighbourhood at a given point of $\underline{\boldsymbol{y}}(t)$. In the argument of Simplex projection, this neighbourhood is called *simplex* [5, 6]. When dimension of time-delay coordinate is E, we can denote this simplex is $\{\underline{\boldsymbol{y}}(t_1), \underline{\boldsymbol{y}}(t_2), \cdots \underline{\boldsymbol{y}}(t_{E+1})\}$. According to the common convention of CCM the distance between $\underline{\boldsymbol{y}}(t)$ and $\underline{\boldsymbol{y}}(t_{E+1})$ is the most large. Let's denote this length as d and call the neighbourhood of $\underline{\boldsymbol{y}}(t)$ as *d-simplex*. Here, we shall discuss the case that ϕ^{-1} is monotonic increase or decrease. In this case, $\phi^{-1}(\underline{\boldsymbol{y}}(t_{E+1}))$ is the most distant from $\phi^{-1}(\underline{\boldsymbol{y}}(t))$. We denote this distance as D and the mapped neighbourhood as *D-simplex*. In this case, we obtain

$$x(t_i) = x(t) + \phi^{-1}(d_i) \leq x(t) + D$$

here, d_i express the distance between $\underline{\boldsymbol{y}}(t)$ and $\underline{\boldsymbol{y}}(t_i)$.

At CCM, the prediction of $x(t)$ is described as $\hat{x}(t)$ and given as

$$\hat{x}(t) = \sum_i w_i x(t_i)$$

$$w_i = \frac{e^{d_i/d1}}{\sum_j e^{d_j/d1}}$$

We obtain the following result by adapting the inequality to the equation above.

$$\hat{x}(t) = \sum_i w_i \{x(t) + \phi^{-1}(d_i)\} \leq x(t) + D$$

Of course, D is changeable with time t, so it is kind to denote as $D(t)$. We shall use the expression $D(t)$ from here.

CCM judges whether a causality exits or not using prediction skill ρ. This ρ is defined as Pearson correlation coefficient between $x(t)$ and $\hat{x}(t)$.

$$\rho = \frac{\sum_t (x(t) - \overline{x})(\hat{x}(t) - \overline{\hat{x}})}{\sqrt{\sum_t (x(t) - \overline{x})^2} \sqrt{\sum_t (\hat{x}(t) - \overline{\hat{x}})^2}}$$

From the result above, we can rearrange this equation as

$$\rho \leq \frac{\sum_t (x(t) - \overline{x})((x(t) - \overline{x}) + (D(t) - \overline{D}))}{\sqrt{\sum_t (x(t) - \overline{x})^2} \sqrt{\sum_t ((x(t) - \overline{x}) + (D(t) - \overline{D})^2}}$$

This shows that ρ has maximum value. Let us call this the maximum as *convergent skill*. We shall now look more carefully into the properties which this convergent skill has.
Here, we regard the right-hand side as inner production.

$$R.H.S \Leftrightarrow \langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}, \frac{\boldsymbol{u} + \boldsymbol{v}}{\|\boldsymbol{u} + \boldsymbol{v}\|} \rangle$$

Here, we denote each vectors as

$$\boldsymbol{v} = \begin{pmatrix} x(1) - \overline{x} \\ x(2) - \overline{x} \\ \vdots \\ x(L) - \overline{x} \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} D(1) - \overline{D} \\ D(2) - \overline{D} \\ \vdots \\ D(L) - \overline{D} \end{pmatrix}$$

L is data size or library data size and the set$\{1, 2, \cdots L\}$ includes t.
Because each $\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}, \frac{\boldsymbol{u} + \boldsymbol{v}}{\|\boldsymbol{u} + \boldsymbol{v}\|}$ are unit vectors, ρ equals one only if $\boldsymbol{v}$ equals zero. This means that if *convergent skill* equals one, $D(t)$ must be independent on time t. Therefore, *convergent skill* express how *D-simplex* (and also *d-simplex* in this case) depends on t.

3 Our forward study

The simple consideration above implies the possibility that convergent skill has strong relation with D-simplex: how prediction depend on local trajectory. According to S-maps, we can classify or evaluate linearity or perhaps periodicity behind time series data. In short, this method is quantified the degree of local dependency when one predict the future. From this idea, we can say that the value of prediction skill of CCM, at the least the value of convergent skill, may show that how linearity or periodicity are in the data. As mentioned in background, we take a profound interest in detecting causalities. This is because the modern natural science are not only purely interest in detecting fundamental principals behind the data but try to control nature depending human being's interest, or at the least predict the future exactly. Therefore, after knowing the existence of causalities, we may get to find how strongly causalities affect among varieties.

If the convergent skill shows how linearity or periodicity are in the data, we can know how nature is far from our hands by CCM. Chaos shows so great complex behaviour that we can't predict future, global property although we can control local one. From this points, we can classify more controlable or predictable phenomena from others with CCM's convergence value. This is the theoretical advantage of CCM which other detecting causality method doesn't have. This properties come from the fact that CCM does not only algorithm detecting causality but theory concerned with causality. Although there are some studies tried to improved CCM's theory mathematically[7, 8, 9], almost all researches concerning with CCM seems to go toward only practical applications to data analysis. These studies are, of course, very important, but at the same time, it is needed to study CCM as not only algorithm but mathematical or physical theory.

We have no space for no more than an indication of the possibility which convergent skill has. It is obvious that the proof is insufficient mathematically and many restriction exit, so there is room for more strictly argument on these points. We, however, think that our argument has also worth a mention in this proceeding.

References

[1] Chang CW, Ushio M, Hssieh CH (2017)Empirical dynamics modelingfor beginners. Ecological Reserch, 32: 785-796

[2] Sugihara G, May RM, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science, 338:496-500.

[3] Ye H, Deyle ER, Gilarranz LJ, Sugihara G (2015) Distinguishing time-delayed causal interactions using convergent cross mapping. Scientific reports, 5:14750.

[4] Ye H, Clark A, Deyle E, Sugihara G (2016) rEDM: an R package for empirical dynamic modeling and convergent cross-mapping.

[5] Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, 344:734-741.

[6] Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences, 348:477-495.

[7] Clark AT, Ye H, Isbell F, Deyle ER, Cowles J, Tilman GD, Sugihara G (2015) Spatial convergent cross mapping to detect causal relationships from short time series. Ecology, 96:1174 – 1181.

[8] Ye H, Sugihara G (2016) Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. Science, 353:922-925.

[9] Ushio M, Hsieh CH, Masuda R, Deyle ER, Ye H, Chang CW, Sugihara G, Kondoh M (2018) Fluctuating interaction network and time-varying stability of a natural fish community. Nature. 554:360-363.

International Institute of Advanced Studies
Kyoto 619-0225
Japan
E-mail address: kusanagi0013@gmail.com