Transfer Information Energy: A Quantitative Causality Indicator Between Time Series

Angel Caţaron[1] and Răzvan Andonie[1,2](✉)

[1] Electronics and Computers Department, Transilvania University, Braşov, Romania
cataron@unitbv.ro
[2] Computer Science Department, Central Washington University, Ellensburg, WA, USA
andonie@cwu.edu

Abstract. We introduce an information-theoretical approach for analyzing cause-effect relationships between time series. Rather than using the Transfer Entropy (TE), we define and apply the Transfer Information Energy (TIE), which is based on Onicescu's Information Energy. The TIE can substitute the TE for detecting cause-effect relationships between time series. The advantage of using the TIE is computational: we can obtain similar results, but faster. To illustrate, we compare the TIE and the TE in a machine learning application. We analyze time series of stock market indexes, with the goal to infer causal relationships between them (i.e., how they influence each other).

1 Introduction

Causal analysis is not merely a search for statistical correlations, but an investigation of cause-effect relationships. Although, in general, statistical analysis cannot distinguish genuine causation from spurious covariation in every conceivable case, this is still possible in many cases [15]. Causality is usually posed using two alternative scenarios: the Granger causality and the information-theoretical approach (based on the Kullback-Leibler divergence or the TE).

The Granger[1] causality test [5] is a statistical hypothesis test for determining whether one time series is useful in forecasting another. According to Granger, causality could be reflected by measuring the ability of predicting the future values of a time series using past values of another time series. The Granger test is based on linear regression modeling of stochastic processes. More complex extensions to nonlinear cases exist, but these extensions are more difficult to apply in practice [6].

The TE, introduced by Schreiber [17], has been used to quantify the statistical coherence between time-series. It is able to distinguish driving and responding elements and to detect asymmetry in the interaction of time-series. For instance, in the financial market, based on the TE concept, Kwon and Oh [12] found that the amount of information flow from index to stock is larger than from stock to

[1] Clive Granger, recipient of the 2003 Nobel Prize in Economics.

A. Lintas et al. (Eds.): ICANN 2017, Part II, LNCS 10614, pp. 512–519, 2017.
https://doi.org/10.1007/978-3-319-68612-7_58

index. It indicates that the market index plays a role of major driving force to individual stock. Barnett *et al.* proved that Granger causality and TE causality measure are equivalent for time series which have a Gaussian distribution [1]. Hlaváčková-Schindler [8] generalized this result.

Our main contribution is an information-theoretical approach for analyzing cause-effect relationships between time series. Rather than using the relatively well-known Kullback-Leibler divergence and the TE (both based on a measure of uncertainty - the Shannon entropy), we introduce the TIE, which is based on a measure of certainty - the Onicescu Information Energy (IE) [14]. In general, any monotonically growing and continuous probability function can be considered as a measure of certainty and the IE is such a function. The IE is a special case of Van der Lubbe *et al.* certainty measure [18] and was interpreted by several authors as a measure of expected commonness, a measure of average certainty, or as a measure of concentration, and is not related to physical energy. We claim that the TIE can substitute the TE for detecting cause-effect relationships between time series, with the advantage of being faster to compute.

An hot application area of causal relationships is finance. Most investors in the stock market consider various indexes to be important sources of basic information that can be used to analyze and predict the market perspectives. We may be interested in the correlation (and beyond that, causality as well) between two time series such as a market/bench index and an individual stock/ETF products. An ETF (Exchange Traded Fund), is a marketable security that tracks an index, a commodity, bonds, or a basket of assets like an index fund. In our application, we compare the TIE and the TE in a machine learning application, analyzing time series of stock market indexes with the goal to infer how they influence each other.

The paper is organized as follows. First, we refer to previous work (Sect. 2). Section 3 introduces the TIE. The financial application is presented in Sect. 4. The paper is concluded in Sect. 5.

2 Related Work: TE for Financial Time-Series

An overview of causality detection based on information-theoretic approaches in time series analysis can be found in [9]. Most of the information-theoretic approaches in time series analysis are based on the TE. The recent literature on TE applications is rich.

TE measures the directionality of a variable with respect to time base on the probability density function (PDF). For two discrete stationary processes I and J, TE relates k previous samples of process I and l previous samples of process J and is defined as follows [17]:

$$TE_{J \to I} = \sum_{t=1}^{n-1} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \, log \frac{p(i_{t+1}|i_t^{(k)}, j_t^{(l)})}{p(i_{t+1}|i_t^{(k)})}, \tag{1}$$

where i_t and j_t are the discrete states at time t of I and J, respectively; $i_t^{(k)}$ and $j_t^{(k)}$ are the k and l dimensional delay vectors of time series I and J, respectively.

$T_{J \to I}$ measures the extend to which time series J influences time series I. The TE is asymmetric under the exchange of i_t and j_t, and provides information regarding the direction of interaction between the two time series. In fact, the TE is an equivalent expression for the conditional mutual information [9].

Accurate estimation of entropy-based measures is notoriously difficult and there is no consensus on an optimal way for estimating TE from a dataset [4]. Schreiber proposed the TE using correlation integrals [17]. The histogram estimation approach with fixed partitioning is the most widely used. This method is simple and efficient, but not scalable for more than three scalars. It also has another drawback - it is sensible to the size of bins used. Since estimating the TE reduces to the non-parametric entropy estimation, other entropy estimation methods have been also used for computing the TE [4,7,19]: kernel density estimation methods, nearest-neighbor, Parzen, neural networks, etc.

Without intending to be exhaustive, we mention two papers which describe time-series information flow analysis with TE. Other recent results can be found in [3,13].

Kwon and Yang [11] computed the information flow between 25 stock markets to determine which market serves as a source of information for global stock indexes. They analyzed the daily time series for the period of 2000 to 2007 using TE in order to examine the information flow between stock markets and identify the hub. They concluded that the American and European markets are strongly clustered and they are able to be regarded as one economic region, while Asia/Pacific markets are economically separated from American and European market cluster. Therefore, they could infer that American and European stock markets fluctuate in tune with a common deriving mechanism. The considerable quantity of the TE from American and European market cluster to the Asia/Pacific markets is the strong evidence that there is an asymmetry of information flow between the deriving mechanisms.

Sandoval [16] used the stocks of the 197 largest companies in the world, in terms of market capitalization, in the financial area, from 2003 to 2012. He studied the causal relationships between them using TE. He could assess which companies influence others according to sub-areas of the financial sector. He also analyzed the exchange of information between those stocks and the network formed by them based on this measure, verifying that they cluster mainly according to countries of origin, and then by industry and sub-industry.

3 Transfer Information Energy

The information entropy of a discrete random variable I with possible values $\{i_1, i_2, \ldots, i_n\}$ is the expected value of the information content of I [2], $H(I) = -\sum_{t=1}^{n} p(i_t)\, log\, p(i_t)$, whereas the IE is the expected value of the probabilities of the possible values of I [14], $IE(I) = \sum_{t=1}^{n} p(i_t) \cdot p(i_t)$.

We define the TIE:

$$TIE_{J \to I} = \sum_{t=1}^{n-1} p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \left(p(i_{t+1} | i_t^{(k)}, j_t^{(l)}) - p(i_{t+1} | i_t^{(k)}) \right), \qquad (2)$$

to quantify the increase in certainty (energy) of process I, knowing k previous samples of process I and l previous samples of process J. Like the TE, the TIE is non-symmetric and measures cause-effect relationships between time series I and J. For computational reasons, we take $k = l = 1$.

Both (1) and (2) can be rewritten substituting the conditional probabilities:

$$TE_{J \to I} = \sum_{i_{t+1}, i_t, j_t} p(i_{t+1}, i_t, j_t) \log \frac{p(i_{t+1}, i_t, j_t) p(i_t)}{p(i_{t+1}, i_t) p(i_t, j_t)}, \tag{3}$$

$$TIE_{J \to I} = \sum_{i_{t+1}, i_t, j_t} p(i_{t+1}, i_t, j_t) \left(\frac{p(i_{t+1}, i_t, j_t)}{p(i_t, j_t)} - \frac{p(i_{t+1}, i_t)}{p(i_t)} \right). \tag{4}$$

Comparing formulas (3) and (4), we observe that for TE we have 4 multiplications/divisions and one logarithm, whereas for TIE we have 3 multiplications/divisions and 1 subtraction. Considering all operations equivalent, the TIE is theoretically 20% faster, which is obviously a rough theoretical estimate.

The histogram estimation of TE and TIE between two time series can be computed in three steps: *(a)* Transformation of the continuous valued time series into series with discrete values by binning; the result is a sequence of tokens selected from an alphabet with as many symbols as the number of bins; *(b)* Evaluation of the probabilities $p(i_{t+1}, i_t, j_t)$, $p(i_t)$, $p(i_{t+1}, i_t)$, and $p(i_t, j_t)$, for all i_t and j_t; and *(c)* Computation of TE and TIE by using Eqs. (3) and (4).

4 Transfer Energy Between Financial Time Series

We illustrate with all details the estimation of TI and TIE on a real-world data set, to make the procedure reproducible.

Table 1. The 20 stock market indexes, obtained from the finance.yahoo.com web site. We estimate the TE and TIE of all pairs from the 20 stock market symbols. Each symbol represents a time series of daily closing prices recorded between Jan. 3, 2000–Feb. 14, 2017.

Americas	Asia/Pacific	Europe
1: MERV	8: AORD	16: ATX
2: BVSP	9: SSEC	17: BFX
3: GSPTSE	10: HSI	18: GDAXI
4: MXX	11: BSESN	19: AEX
5: GSPC	12: JKSE	20: SSMI
6: DJA	13: N255	
7: DJI	14: KS11	
	15: TWII	

For 20 stock market indexes from Americas, Asia/Pacific and Europe (Table 1), we estimate the TE and TIE for all pairs. The working days of the markets across the world may vary from one country to another. Therefore, the time series are aligned by time stamp and the missing values are replaced with the previous available ones. We estimate TE and TIE as follows:

(a) *Discretization: binning the time series*

We slice the domain limited by the minimum and maximum values from the whole data set into equally sized intervals which are then labeled by assigning a symbol to each of them. The result is a sequence of characters, for which we compute the probabilities needed in Eqs. (3) and (4).

When the binning is applied on the first log-returns of stocks, the narrow bins provide more information content, thus a higher value of entropy H then the large bins. Nevertheless, the correlation between the two choices of binning is high in general, reflecting an important similarity of the approaches [16]. In general, for shorter time series it is advisable to use larger bins in order to avoid the excessive fragmentation (and thus very low or uniform probabilities of symbols). We use 24 bins, noting that the binning strategy is less relevant in our case, since we are not interested in absolute values for TE and TIE, but in their relative values (for comparison). Fig. 1 depicts the binning and Table 2 shows a numerical example of binning based on the first values of the DJI and HSI stock indexes.

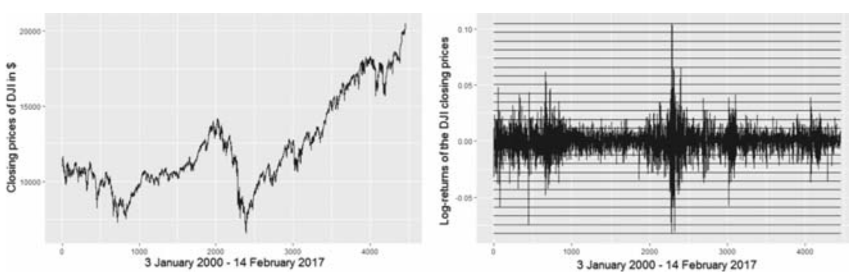

Fig. 1. Binning the time series. The left graph presents the raw values of the DJI stock ranging between Jan. 3, 2000–Feb. 14, 2017. On the right side, we represent the log-returns of closing prices and the slicing of the values domain, with 24 equal intervals between minimum and maximum values. Each interval is labeled with a symbol (a letter).

(b) *Compute the marginal and joint probabilities in Eqs. (3) and (4)*

We denote by TE_t the term under the sum sign in (3) and by TIE_t the term under the sum sign in Eq. (4). The next step is to evaluate TE_t and TIE_t by counting the number of each occurrence (Table 2). The string obtained by binning the log-returns of the DJI stock starts with the symbols "g l m n l j k k m k ...". Therefore, $p(i_1) = 0.00673$ is the probability of occurrence of symbol "g", $p(i_2) = 0.21054$ is the probability of occurrence of symbol "l",

Table 2. Illustration of the step by step calculation of TE and TIE. Binning the log-returns of the DJI values is subject to slicing the values interval, limited by −0.082 and 0.105. The limits of log-returns of HSI are −0.135 and 0.134. The probabilities are the relative frequencies of symbols or combination of symbols, while TE and TIE can be calculated from the intermediary values TE_i and TIE_i, which are obtained from the probabilities listed on column t_i.

	t_0	t_1	t_2	t_3
Closing prices of DJI	11357.51	10997.93	11122.65	11253.26
Log-returns of DJI		−0.0321	0.0112	0.0116
Binned log-returns of DJI		$i_1 : g$	$i_2 : l$	$i_3 : m$
Closing prices of HSI	17369.63	17072.82	15846.72	15153.23
Log-returns of HSI		-0.0172	-0.0745	-0.0447
Binned log-returns of HSI		$j_1 : k$	$j_2 : f$	$j_3 : i$
(i_{t+1}, i_t)		$(i_2, i_1) : lg$	$(i_3, i_2) : ml$	$(i_4, i_3) : nm$
(i_t, j_t)		$(i_1, j_1) : gk$	$(i_2, j_2) : lf$	$(i_3, j_3) : mi$
(i_{t+1}, i_t, j_t)		$(i_2, i_1, j_1) : lgk$	$(i_3, i_2, j_2) : mlf$	$(i_4, i_3, j_3) : nmi$
$p(i_t)$		$p(i_1)$	$p(i_2)$	$p(i_3)$
$p(i_{t+1}, i_t)$		$p(i_2, i_1)$	$p(i_3, i_2)$	$p(i_4, i_3)$
$p(i_t, j_t)$		$p(i_1, j_1)$	$p(i_2, j_2)$	$p(i_3, j_3)$
$p(i_{t+1}, i_t, j_t)$		$p(i_2, i_1, j_1)$	$p(i_3, i_2, j_2)$	$p(i_4, i_3, j_3)$
$TE = \sum(TE_t)$		TE_1	TE_2	TE_3
$TIE = \sum(TIE_t)$		TIE_1	TIE_2	TIE_3

etc. The probability $p(i_2, i_1) = 0.00179$ is the probability of the sequence "gl", $p(i_3, i_2) = 0.00942$ is the probability of "lm", etc. The string obtained by binning the log-returns of the HSI stock starts with the symbols "k f i n o m l l l m ...". We obtain the probability of "gk": $p(i_1, j_1) = 0.00224$, the probability of "gk": $p(i_2, j_2) = 0.00224$, etc. Next, $p(i_2, i_1, j_1) = 0.00067$ is the probability of "lgk", $p(i_3, i_2, j_2) = 0.00022$ is the probability of "mlf", etc. For an accurate estimation, a larger number of decimals is preferred.

(c) *Estimate TE and TIE*
We calculate TE_t and TIE_t. For the first step, $TE_1 = 0.000011$ and $TIE_1 = 0.0000022$, etc. Finally, we compute $TE = 47.76$ and $TIE = 17.85$, summing-up the partial results.

The results are summarized in the heatmaps (Fig. 2). The lighter shaded pixels are associated with a higher values of TE and TIE. We visually observe that the two heatmap correlate well. In fact, Pearson correlation coefficient is 0.973, showing a strong correlation.

Figure 3 illustrates the execution time for computing TIE and TI. For time series with more than 1,000 values, the execution time for TIE becomes clearly shorter. For an increasing number of values, the ratio TIE/TE of the executions times decreases.

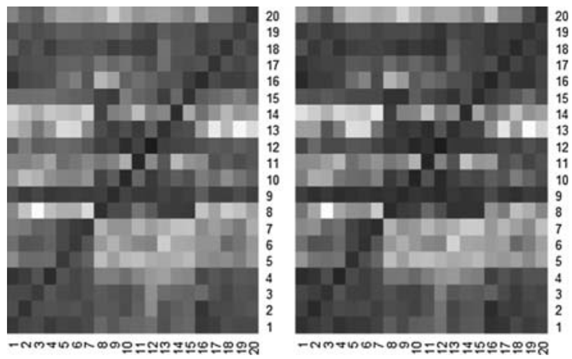

Fig. 2. The two heatmaps are calculated for TE (left) and TIE (right), between all combinations of the 20 stock indexes.

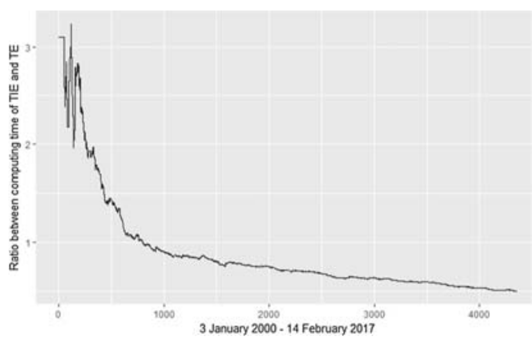

Fig. 3. Execution time. The graph shows the ratio TIE/TE of the executions times, for an increasing number of values. The time is computed for the DJI and HSI stocks ranging between Jan. 3, 2000–Feb. 14, 2017. The relative efficiency of TIE increases for larger time series. For 4357 points, the ratio is 0.49918.

5 Conclusions

According to our preliminary results, the TIE can substitute the TE for detecting cause-effect relationships between time series, with the advantage of a computational complexity reduction. This result is very interesting, since the TE is already a standard concept (Scheiber's paper [17] has at this moment 842 citations.

Even if its use as an information flow measure is debatable (see [10]), the TE can be used as a measure of the reduction in uncertainty about one time series given another. Symmetrically, the TIE may be viewed as a measure of the increase in certainty about one time series given another. It is an open problem if the TIE is an appropriate energy flow measure.

References

1. Barnett, L., Barrett, A.B., Seth, A.K.: Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett. **103**, 238701 (2009)
2. Borda, M.: Fundamentals in Information Theory and Coding. Springer, Heidelberg (2011). doi:10.1007/978-3-642-20347-3
3. Dimpfl, T., Peter, F.J.: Using transfer entropy to measure information flows between financial markets. SFB 649 Discussion Papers SFB649Dpp. 2012-051, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, August 2012
4. Gencaga, D., Knuth, K.H., Rossow, W.B.: A recipe for the estimation of information flow in a dynamical system. Entropy **17**(1), 438–470 (2015)
5. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), 424–438 (1969)
6. Guisan, M.C.: A comparison of causality tests applied to the bilateral relationship between consumption, GDP in the USA. Mexico. Int. J. Appl. Econom. Quant. Stud.: IJAEQS **1**(1, (1/3)), 115–130 (2004)
7. Hlaváčková-Schindler, K.: Causality in time series: its detection and quantification by means of information theory. In: Emmert-Streib, F., Dehmer, M. (eds.) Information Theory and Statistical Learning, pp. 183–207. Springer, Boston (2009). doi:10.1007/978-0-387-84816-7_8
8. Hlaváčková-Schindler, K.: Equivalence of granger causality and transfer entropy: a generalization. Appl. Math. Sci. **5**(73), 3637–3648 (2011)
9. Hlaváčková-Schindler, K., Palu, M., Vejmelka, M., Bhattacharya, J.: Causality detection based on information-theoretic approaches in time series analysis. Phys. Rep. **441**(1), 1–46 (2007)
10. James, R.G., Barnett, N., Crutchfield, J.P.: Information flows? A critique of transfer entropies. Phys. Rev. Lett. **116**(23), 238701 (2016)
11. Kwon, O., Yang, J.-S.: Information flow between stock indices. EPL (Europhys. Lett.) **82**(6), 68003 (2008)
12. Kwon, O., Oh, G.: Asymmetric information flow between market index and individual stocks in several stock markets. EPL (Europhys. Lett.) **97**(2), 28007 (2012)
13. Mao, X., Shang, P.: Transfer entropy between multivariate time series. Commun. Nonlinear Sci. Numer. Simul. **47**, 338–347 (2017)
14. Onicescu, O.: Theorie de l'information energie informationelle. C. R. Acad. Sci. Paris Ser. A–B **263**, 841–842 (1966)
15. Pearl, J.: Causality: Models, Reasoning and Inference, 2nd edn. Cambridge University Press, New York (2009)
16. Sandoval, L.: Structure of a global network of financial companies based on transfer entropy. Entropy **16**(8), 4443–4482 (2014)
17. Schreiber, T.: Measuring information transfer. Phys. Rev. Lett. **85**, 461–464 (2000)
18. van der Lubbe, J.C.A., Boxma, Y., Bockee, D.E.: A generalized class of certainty and information measures. Inf. Sci. **32**(3), 187–215 (1984)
19. Zhu, J., Bellanger, J.-J., Shu, H., Le Bouquin Jeannès, R.: Contribution to transfer entropy estimation via the k-nearest-neighbors approach. Entropy **17**(6), 4173–4201 (2015)