2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Study of Learning Entropy for Novelty Detection in Lung Tumor Motion Prediction for Target Tracking Radiation Therapy

Ivo Bukovsky, Noriyasu Homma, Matous Cejnek, Kei Ichiji

***Abstract*—This paper presents recently introduced concept of Learning Entropy (LE) for time series and recalls the practical form of its evaluation in real time. Then, a technique that estimates the increased risk of prediction inaccuracy of adaptive predictors in real time using LE is introduced. On simulation examples using artificial signal and real respiratory time series, it is shown that LE can be used to evaluate the actual validity of the adaptive predicting model of time series in real time. The introduced technique is discussed as a potential approach to the improvement of accuracy of lung tumor tracking radiation therapy.**

I. Introduction

Adaptive techniques of time series prediction can be successfully applied to respiratory time series prediction [1], in particular, to the prediction of the lung tumor motion to compensate a time delay inherent in a control system that re-positions radiation beam for target tracking radiation therapy (TRT) [2][3]. Therefore, adaptive predictive techniques, including also other ones than neural-network approaches (e.g. [4][5]), can be used for improving accuracy of TRT, because the dosage of the radiation beam can be synchronized with abdomen motion according to the real time prediction of the lung tumor respiratory position. Thus, the radiation beam shall more precisely aim to the tumor position and the radiation to surrounding healthy tissue would be minimized. Even in a supine position of a patient, his or her respiration time series are naturally nonstationary due to the complex physiological dynamics and because of unpredictable varying perturbations.

This naturally makes the difference between theoretical models and real data, e.g., such as between chaotic Mackey-Glass equation [6] (that can be adaptively well precisely predicted even without retraining [3][17]) and real respiratory time series (whose prediction error converges just temporarily and real–time retraining is necessary [1]–[5]). Thus, the prediction of real respiratory time series can not be very precise for a sufficiently long and continuous interval of time. One natural option for improving the accuracy of TRT can be to radiate the tumor only during the intervals when the predictive model is expected to be accurate, i.e., when the model is temporarily well re-trained and input data are not recognized as unusual. If a new measured sample of data is acquired and recognized as novel, i.e. the sample of data is incoherent to the temporary dynamics, then the temporarily retrained model naturally loses its credit for accuracy and the radiation beam shall be switched off to minimize the risk of radiating the healthy tissue. Thus novelty detection methods can be involved in respiratory time series prediction for TRT.

In order to review recent works on novelty detection in computational intelligence area, we shall first distinguish between statistical approaches [7] and neural–network approaches [10]. Regarding the statistical novelty detection approaches in time series, Sample Entropy of Pincus [8] and its extensions [9] have to be mentioned here. In this paper, we focus on the utilization of learning–system novelty detection approaches that clearly include neural networks in general [11]. Learning–systems based novelty detection can include fault detection approaches [12][13], nonstationary changes [14], recently encode-decode detection approach [15], or concept drifts detection utilizing the incremental learning techniques [16]. On the one hand, it was shown that neural networks can learn the contemporary governing laws of respiratory motion [1][3][5]. On the other hand, it is well known that neural networks can learn governing laws of nonlinear dynamical systems. Then, because the lung tumor motion series are nonstationary yet nonlinear and bounded by a governing law, the neural networks are a promising tool for novelty detection in respiratory time series via learning systems approach. Contrary to statistical approaches, neural networks can provide us with cognitive understanding to otherwise statistically complex behavior if the one can be adaptively learned and that can be used to improve novelty detection in complex time series. Recently, the Learning Entropy (LE) was introduced as a novel, cognitive, and non--probabilistic novelty detection approach for incrementally learning systems [17].

In this paper, we review a class of polynomial neural architectures as adaptive predictors and their use for time series prediction. Then, we recall the practical calculation of LE and we propose a method of using LE for real-time estimation of the prediction accuracy. We discuss that as a possible approach to improvement of TRT, where also the radiation beam intensity can be synchronized with estimated prediction accuracy, i.e. with LE.

Ivo Bukovsky and Matous Cejnek are with Czech Technical University in Prague, Department of Instrumentation and Control Engineering, Technicka 4, 166 07, Prague, Czech Republic, email: {ivo.bukovsky;matous.cejnek}@fs.cvut.cz.

Noriyasu Homma and Kei Ichiji are with Tohoku University, Japan. Homma is with the Department of Radiological Imaging and Informatics, Graduate School of Medicine, email: homma@ieee.org; Ichiji is with the Yoshizawa Laboratory, Graduate School of Engineering, email: ichiji@yoshizawa.ecei.tohoku.ac.jp.

This work has been partially supported by grants SGS12/177/OHK2/3T/12, by the Rectorate fund of CTU in Prague, JSPS KAKENHI Grant Number 25293258, and the Kayamori Foundation of Informational Science Advancement.

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

II. Applied Methods

In the following, bold symbols stand for vector and matrices, y stands for real time series, $\tilde{y}$ denotes neural output, and discrete time index k is used when necessary; constant sampling is considered. Meaning of other notations is explained at their first appearance.

A. Lung Tumor Motion Prediction by HONU

Higher-order nonlinear neural units (HONU) [19] has been shown as promising polynomial neural architectures to predict chaotic time series [17] and real signals including respiratory time series [3]. Linear predictors, i.e., linear neural units (LNUs) are considered to be the first–order neural units. The second–order neural unit is called the quadratic neural unit (QNU) and the third–order one can be called the cubic neural unit (CNU). The general long–vector form of higher-order neural units [3][19] can be expressed as follows

$$\tilde{y}(k+p) = \mathbf{w} \cdot \mathbf{colx}\,, \qquad (1)$$

where $\tilde{y}(k+p)$ is the predicted value of lung-tumor position at prediction horizon of p samples, $\mathbf{w}$ denotes the row vector of all neural weights, and $\mathbf{colx}$ stands for the column vector of polynomial terms including the neural inputs and feedbacks. Particularly for LNUs, the predictor's output is calculated as

$$\tilde{y}(k+p) = \sum_{i=0}^{n} w_i x_i = \mathbf{w} \cdot \mathbf{x};\ \text{for LNU } \mathbf{x} = \mathbf{colx}\,, \qquad (2)$$

where n is the length of signal history at predictor's input so

$$\mathbf{x} = \left[1 = x_0 \ \ y(k) \ \ y(k-1) \ \ \ldots \ \ y(k-n+1)\right]^T, \qquad (3)$$

where x_0=1 allows for neural bias in case of LNU and it also allows LNU be subset of HONUs in case of higher orders. Then, QNU can be expressed as follows

$$\tilde{y}(k+p) = \sum_{i=0}^{n} \sum_{j=i}^{n} w_{i,j} \cdot x_i \cdot x_j = \mathbf{w} \cdot \mathbf{colx}\,, \qquad (4)$$

where the all quadratic polynomial terms are in $\mathbf{colx}$ as

$$\mathbf{colx} = \left[\left\{x_i \cdot x_j \ ;\ i = 0 \ldots n,\ j = i \ldots n\right\}\right], \qquad (5)$$

and the weight vector becomes as follows

$$\mathbf{w} = \left[\left\{w_{i,j} \ ;\ i = 0 \ldots n,\ j = i \ldots n\right\}\right]. \qquad (6)$$

Similarly the long vectors can be extended for CNU and higher orders, while the simple vector notation of HONU (1) can be maintained. The long-vector representation of HONU (1) also clearly highlights the fact that HONUs are nonlinear mappings but linear in parameters. This has important connotations to the learning of HONUs as their optimization is a linear problem and thus there are no local minima for a given training data [19]. Furthemore, HONUs in form (1) can directly adopt learning improvement techniques that are applicable to linear filters as the normalized gradient descent (GD) [20] and other improvements [21]–[23] can also be adopted and studied for HONU. Using the above long-vector (flattened) representation of QNU, the classical GD learning rule for weight updates in the next computational step is as

$$\mathbf{\Delta w}(k+1) = -\frac{1}{2}\mu \frac{\partial e(k+p)^2}{\partial \mathbf{w}};\ \ \textbf{where}\ e = y - \tilde{y}\,, \qquad (7)$$

and for QNU with normalized learning rate μ it yields

$$\mathbf{\Delta w}(k+1) = \frac{\mu}{1 + \left\|\mathbf{colx}\right\|} \cdot e(k+p) \cdot \mathbf{colx}^T\,, \qquad (8)$$

where the analogy to the normalized gradient descent of linear adaptive filters as in [21] is apparent.

B. Learning Entropy

The LE, in particular the Approximated Individual Sample Learning Entropy (AISLE) [17], is a multiscale measure [18] that is based on the cognitive capabilities of incrementally learning systems where the weight updates are incrementally updated as follows

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \Delta\mathbf{w}(k)\,. \qquad (9)$$

The neural weight increments $\mathbf{\Delta w}$ are the key quantities for evaluating LE. In particular, evaluation of LE for every newly measured sample at sample number k, where weight increments $\mathbf{\Delta w}$ were evaluated by normalized gradient descent, has been shown in [17]. However, the LE is not limited to GD, but it can be principally calculated for any learning rule where the weight updates follow the equation (9). Regardless of the particular algorithm of calculating $\mathbf{\Delta w}$, the LE is evaluated as quantification of unusual behavior of weight updates $\mathbf{\Delta w}$. Then, the Individual Sample Learning Entropy (ISLE) is calculated as follows. At every update sample k in (9), each single-weight update $\Delta w_{i,j}$ is checked for unusually large magnitude as

$$N(\alpha) = \sum_{i=0}^{n} \sum_{j=0}^{n} f(\Delta w_{i,j}, \alpha) \qquad (10)$$

where N is the count of unusually large magnitudes of weight updates for a given detection sensitivity α, and the condition of exceeding contemporary update magnitude is evaluated as follows

$$f(\Delta w_{i,j}, \alpha) = \begin{cases} 1, & \left(\left|\Delta w_{i,j}\right| > \alpha \cdot \overline{\Delta w_{i,j}}\right) \\ 0, & \text{otherwise} \end{cases} \qquad (11)$$

where the contemporary average of weight-update absolute values $\overline{\left|\Delta w_{i,j}\right|}$ is calculated over a pragmatically chosen number of recent samples and a modification for quasi-periodical signals has been proposed in [17] as well. Also, N corresponds to the count of markers of the Adaptation Plot (AP) in [17]. Because the proper parameter of sensitivity detection α can not be automatically identified, then the multi-scale approach is adopted as N is evaluated over the whole range of its values. Then, E_A denotes AISLE and it is finally evaluated at every weight update as

$$E_A(k) = \frac{1}{n_w \cdot n_\alpha} \sum_i \left\{N(\alpha_i)\right\};\ \alpha = \alpha_1, \alpha_2, \ldots, \alpha_{n_\alpha}\,, \qquad (12)$$

where E_A is thus normalized measure $E_A \in \langle 0\,, 1 \rangle$. n_w is the number of all weights in $\mathbf{w}$, and n_α is the number of detection sensitivities. In the end, the AISLE evaluated via (10)-(12) is

considered to be the first-order LE because conditions (10), resp. (11), are calculated with the first difference of neural weights while the higher orders of LE would be evaluated using higher-order weight differences (such as weight acceleration etc., please see [17] for more details).

III. LE AND HIGHER RISK OF PREDICTION ERROR

This section introduces a technique that estimates the chance if individually predicted samples can be predicted accurately and thus if the radiation beam shall be maintained or temporarily suppressed in real time. Second order HONU (QNU) is here studied as the real-time adaptive predictor of lung tumor motion time series. Particularly the QNU shows promising results for lung tumor motion time series prediction [3] and it is also very suitable for LE because HONUs are linear in parameters [19]. The in-parameter linearity frees HONUs from the local minima issue for a given training data set [19], so the incremental learning updates are not affected due to local-minima-perturbed error surface (as in case of multilayer perceptron neural networks, e.g.); thus, the GD weight updates of a predictor can reflect novelty in each sample of measured data, which is the core idea of LE (though not limited to GD learning in principle [17]). Because the novelty detection in each sample of data by LE is evaluated by the prediction learning algorithm itself, it is principally suitable to estimate the risk of low accuracy of the predictor with every new measured sample of data.

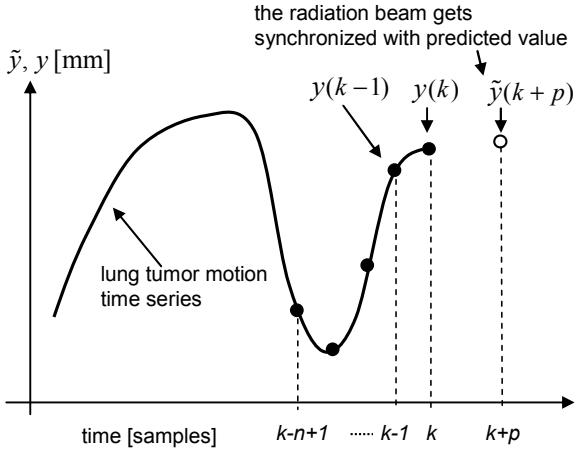

Fig. 1: The novel possible improvement of TRT consists in using the LE to estimate the risk of prediction inaccuracy in real time and thus in adapting also the intensity of the radiation beam (usually only repositioning is considered for synchronization with the prediction $\tilde{y}(k+p)$).

The use of the LE for the possible enhancement of the radiation therapy is principally sketched through Fig. 1 and rules (13) or (14), where we propose two possible criteria for estimating the prediction accuracy and thus possibly for improving the accuracy of TRT. The first criterion is the **Actual Individual Sample Criterion** (**AISC**) (13) for estimating the increased risk of inaccurate prediction in real-time and it is given as follows

$$\begin{aligned} &if: E_A(k) > \tilde{E}_A \\ &\Rightarrow \text{novel data, risk of less accurate prediction} \\ &else: \\ &\Rightarrow \text{not novel data, not increased risk} \end{aligned} \quad \textbf{(13)}$$

where k denotes the reference time index of the last measured sample, $\tilde{E}_A$ is the LE bias deciding the unusual learning effort of the adaptive model to the last measured sample of data; thus, AISC indicates the higher risk of possible prediction inaccuracy of a future lung tumor position.

Second, we introduce the **Actual Input Vector Criterion** (**AIVC**) (14) as follows

$$\begin{aligned} &if: \max\{E_A(k-n+1:k)\} > \tilde{E}_A \\ &\Rightarrow \text{novel data, risk of less accurate prediction} \\ &else: \\ &\Rightarrow \text{not novel data, not increased risk} \end{aligned} \quad \textbf{(14)}$$

where n is the length of input vector $\mathbf{x}$ in (3), i.e., n is here the number of recently measured samples used to predict next value. The criterion (14) requires all of these recent values be usual (i.e. of low LE), so the model knows input data and the risk of increased prediction error is not detected. This section introduced a new real-time technique that employs the novelty detection using LE for estimating the risk of inaccurate prediction with QNU and with normalized GD adaptation via the single–sample based criteria (AISC) (13) or via the vector–based criteria (AIVC) (14).

IV. EXPERIMENTAL ANALYSIS

This section demonstrates the introduced real-time method of prediction–accuracy estimation on artificial data and then on real respiratory time series, in particular, on a lung-tumor motion time series.

A. Artificial Signal with Perturbations

As on the artificial signal, we demonstrate the method on a perturbed sinus (AC voltage like) time series that is given as follows

$$y(t) = A \cdot \sin(2\pi \cdot f \cdot t + \phi) \,, \quad \textbf{(15)}$$

with initial amplitude A=230 [V], initial frequency f=60 Hz, and initial zero phase shift $\phi = 0$. Sampling is chosen as 17 samples per period, i.e. $t \leftarrow k \cdot \Delta t$ where $\Delta t = \frac{1}{17 \cdot f}$. The perturbations to the signal (15) are introduced either as three large perturbations or as three small perturbations as follows:

- at k = 101, amplitude offset of +230 [V] or +10 [V],
- for k > 150, frequency change to 55 [Hz] or 59.5 [Hz],
- for k > 200, phase shift +1 [Rad] or +0.1 [Rad].

The adaptive predictor used for this signal is QNU that is incrementally adapted by normalized GD (as derived via (1)–(9)) starting from random initial weights without pre-training.

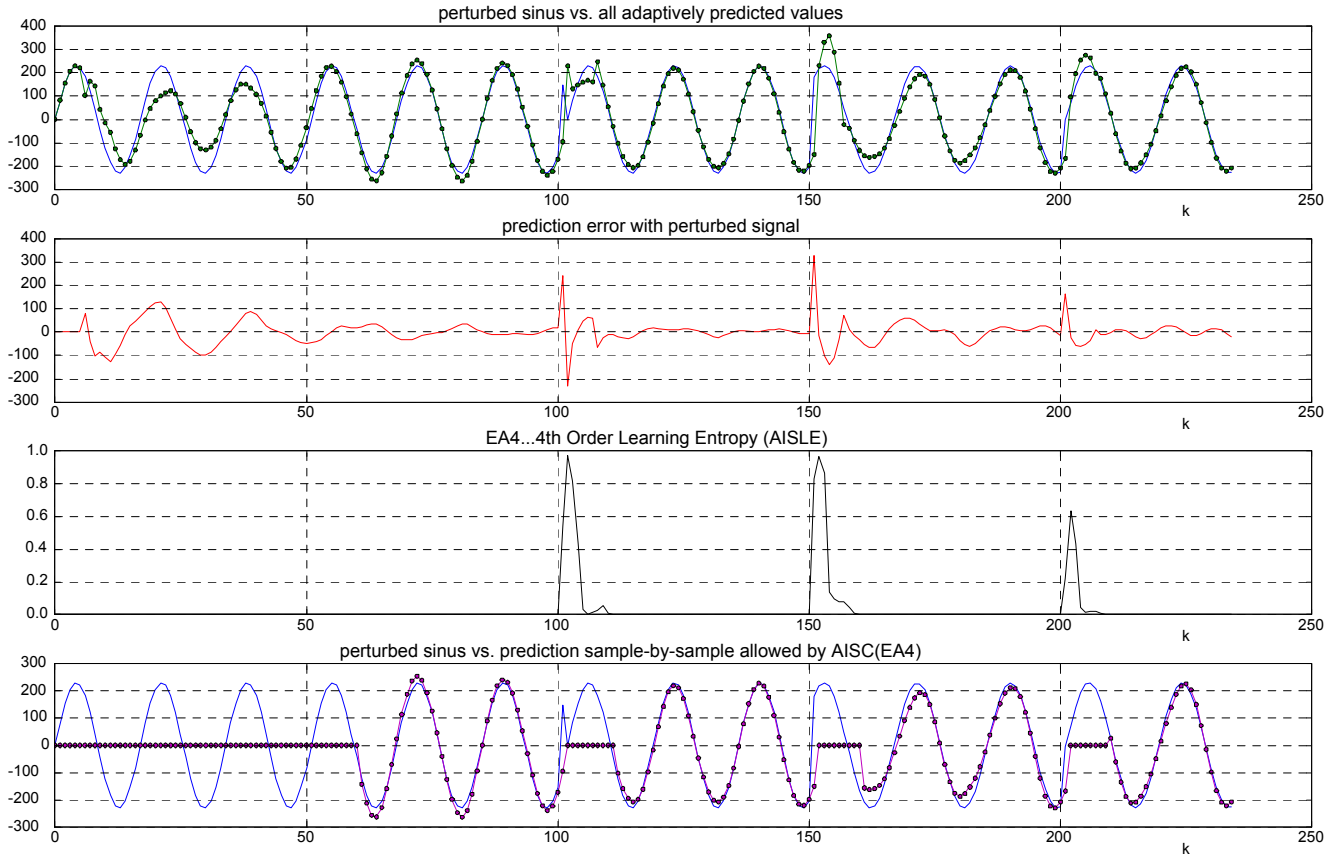

Fig. 2: Application of AISC criteria with 4th Order Learning Entropy during real-time prediction of sinusoidal signal (15) with three large perturbation perturbations at k=101,150,200; prediction is suppressed to 0 at samples when novelty of data is detected by increased LE (AISC not applied to first 60 samples, μ=1, n=6, $\tilde{E}_A = 10^{-5}$, perturbed sinus is in a blue color).

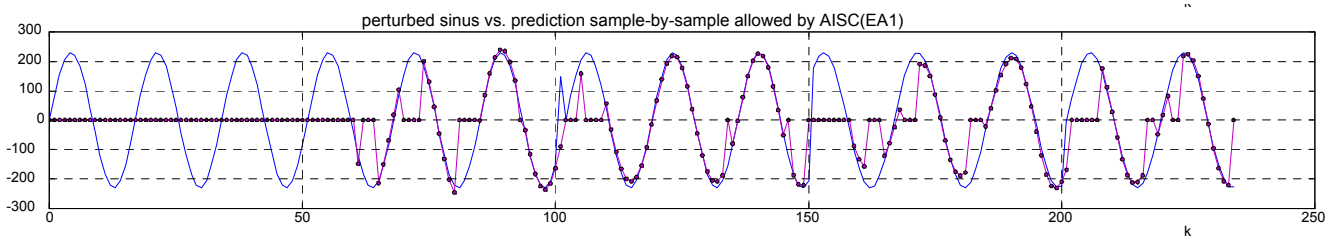

Fig. 3: Application of AISC criteria with 1st Order Learning Entropy during real-time prediction of sinusoidal signal (15) with three large perturbations at k=101,150,200; prediction is suppressed to 0 at samples when novelty of data is detected by increased LE (AISC not applied to first 60 samples, μ=1, n=6, $\tilde{E}_A = 10^{-5}$, perturbed sinus is in a blue color).

Fig. 2 represents the results of the introduced method of estimating the accuracy of adaptive real-time prediction with incremental learning (1)–(9). The first axes show artificial time series (15) with three large perturbations as described above. The occurrences of perturbations result in unusually large prediction error (2nd axes from top) and in drastic increase of 4th Order LE (3rd axes from top). The bottom axes in Fig. 2 then shows the superposition of the predicted values on the original perturbed signal, where the predicted values are set to zero if the AISC (13) holds for them, i.e., when the previous sample has increased LE. We can clearly see that the prediction is temporarily suppressed right after perturbations and prediction becomes allowed when LE decrease soon again. Thus the AISC estimates risk of imprecise predictions in a real time and most of risky-prediction samples can be excluded from use so it leads to the lower mean absolute error (MAE=21.7) of prediction processed with AISC (bottom axes in Fig. 2) than if all predicted data were used (top axes in Fig. 2, MAE=26.1). Fig. 3 then demonstrates the effect of using a lower Order of Learning Entropy as the estimation of increased risk of inaccuracy by AISC or AIVC becomes more intermittent for lower OLE; for the filtering effect of higher OLEs on sensitive novelty detection, please see [17]. On the other hand, there is MAE=15.41 of the prediction that comply to criterion AISC in Fig. 3 that is in fact lower than MAE achieved with 4th OLE yet the prediction validity is more intermittent, and also the number of effective samples, i.e., predicted samples that comply AISC is lower for Fig. 3. Understanding to these aspects need further research and exceeds the scope of this paper.

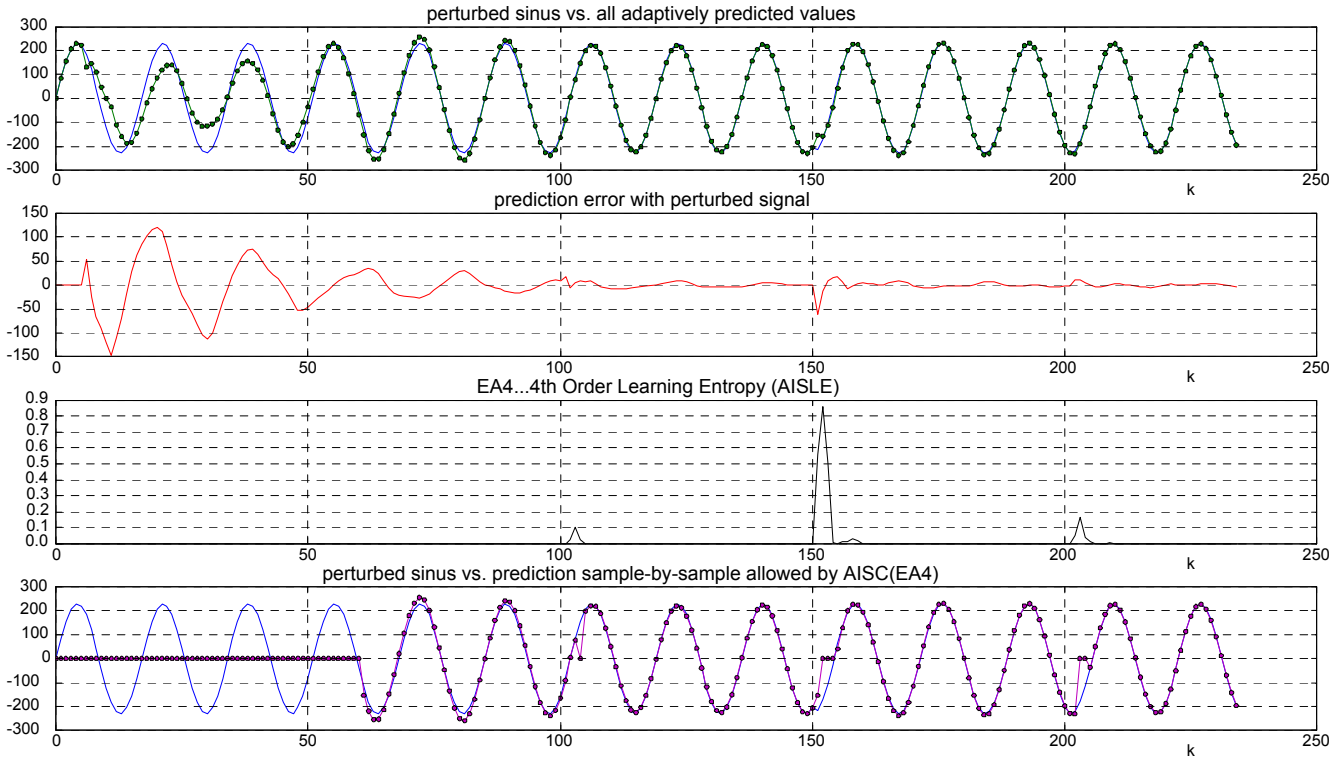

Fig. 4: Application of AISC criteria with 4th Order Learning Entropy during real-time prediction of sinusoidal signal (15) with three small perturbations at k=101,150,200; prediction is suppressed to 0 at samples when novelty of data is detected by increased LE (AISC not applied to first 60 samples, μ=1, n=6, $\tilde{E}_A = 0.02$, perturbed sinus is in a blue color) .

B. *Lung Tumor Motion Time Series*

The lung-tumor motion time series that is used in this analysis was acquired at Hokkaido University Hospital at 30 Hz sampling. QNU with normalized GD is again used as an adaptive predictor, this time it is with pre-training in 10 epochs on initial 60 samples. The prediction horizon is 1/3 second (time series was re-sampled to 3 samples per second).

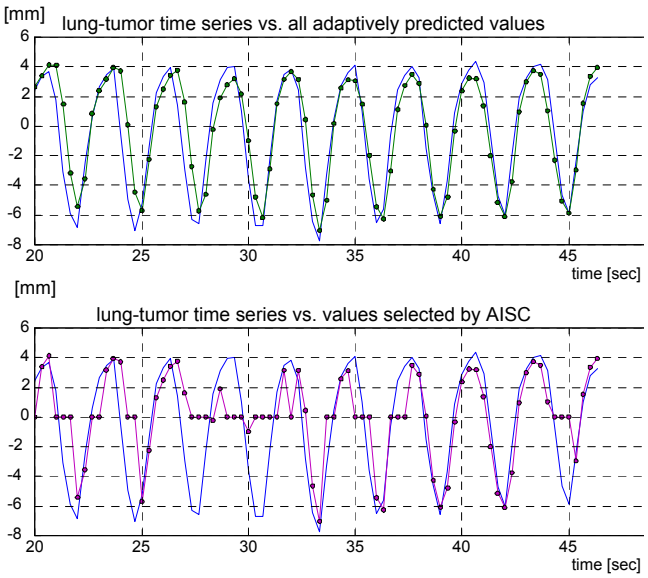

Fig. 5: The novel possible improvement of TRT consists in using the LE to estimate the risk of prediction inaccuracy in real time and thus in adapting also the intensity of the radiation beam (usually only repositioning is considered for synchronization with the prediction $\tilde{y}(k+p)$).

Fig. 5 demonstrates the use of the introduced method on the real data with 4th OLE. For all 350 samples of time series (+ 60 initial samples for pre-training), for 2nd Order of LE e.g., and for $\tilde{E}_A = 0.15$, the prediction accuracy improvements achieved for lung-tumor motion time series are shown in Table 1.

TABLE 1: PREDICTION ACCURACY IMPROVEMENT WITH USE OF LE

prediction accuracy estimation rule	MAE [mm]	RMSE [mm]	# of used samples
none (all data)	0.894	1.215	350
AISC	0.726	0.996	177
AIVC	0.735	0.941	176

V. DISCUSSION

During tracking radiation therapy of lung tumors, it is usually assumed that the radiation beam can be synchronized with real time predicted position to avoid the irradiation of healthy tissue. However, the accuracy of the prediction varies in time, so we have developed and introduced the fundaments of the technique that estimates the increased risk of inaccurate prediction. The TRT can be then adaptively controlled as the beam repositioning can be synchronized with the predicted tumor position plus the beam intensity can be perhaps instantly suppressed or timely synchronized with the novelty of input data and the beam can be retrieved when the novelty, i.e., LE decrease back again as soon as the learning system learns new governing law and the input data are not novel to the learning system. The radiation of the tumor might then become intermittent; however, it can be better synchronized with tumor motion that has lowered the risk of prediction inaccuracy. To verify the true accuracy of

the model of course, we have to wait for a prediction interval to check if the model had predicted accurately. Notice, the actual prediction error $e(k)$ itself is not necessary related to novelty detection by Learning Entropy [17] and $e(k)$ can still be considered as another indicator of novelty in data. Thus, both the LE and the prediction error together could estimate the chance for higher prediction accuracy and thus for intermittent, yet more accurate and timely beam focus during online lung tumor radiation. Employment of LE together with the prediction error in criteria AISC (13) and AIVC (14) would exceed the scope of this paper and would deserve more extensive investigation. Also, AIVC represents a stricter criterion than AISC and we are going to focus the more detailed research on these criteria in a close future. Importantly, the major weakness of the proposed method is the determination of the proper values of bias $\tilde{E}_A$, which is the typical single-scale weakness of the proposed approach. Currently, a proper selection of the bias can be resolved by an automated heuristic approach in a sliding window.

VI. Conclusions

Learning Entropy for novelty detection in time series was proposed as a method of real-time estimation of actual risk of increased prediction inaccuracy for incrementally learning predictive models. The approach detects a possible risk that newly predicted samples with the adaptive predictor might be inaccurate and LE is capable to detect novelty in data even if contemporary prediction error is of usual magnitude (for LE vs prediction error please see [17]). The experimental part demonstrated that the proposed theoretical method has potentials to increase the accuracy of the TRT if also the radiation beam intensity could be synchronized with the estimated prediction accuracy by the proposed criteria using also LE.

Acknowledgment

The authors would like to thank Madan M. Gupta from the University of Saskatchewan for introducing them to higher–order neural units and for continuous support and consultancy. The authors would also like to thank Witold Kinsner from the University of Manitoba for continuous support and consultancy on multi–scale analysis approaches.

References

[1] Kakar M., Nystrom H., Aarup L. R., Nottrup T. J. and Olsen D R Respiratory motion prediction by using the adaptive neuro fuzzy inference system (ANFIS) *Physics in Medicine and Biology* 50 4721-28, 2005.

[2] Isaksson M., Jalden J. and Murphy M. J. "On using and adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications", *Med. Phys*. 32 3801-9, 2005.

[3] Bukovsky, I., Ichiji, K., Homma, N., Yoshizawa, M.: "Testing Potentials of Dynamic Quadratic Neural Unit for Prediction of Lung Motion during Respiration for Tracking Radiation Therapy", *WCCI 2010, IEEE Int. Joint. Conf. on Neural Networks IJCNN*, Barcelona, Spain, 2010.

[4] Bodyanskiy, Y., Kolodyazhniy, V.: "Real-Time Identification and Forecasting of Chaotic Time Series Using Hybrid Systems of Computational Intelligence", *StudFuzz* 187, Springer-Verlag, 439–480, 2006.

[5] Ichiji K, Homma N, Sakai M, Narita Y, Takai Y, Zhang X, Abe M, Sugita N, Yoshizawa M: A Time-Varying Seasonal Autoregressive Model-Based Prediction of Respiratory Motion for Tumor following Radiotherapy, *Computational and Mathematical Methods in Medicine* 390325, 2013 .

[6] Mackey, M.C.; Glass, L. Oscillation and Chaos in Physiological Control Systems. *Science,* Vol. 197, Issue 4300, pp.287-289. 1977.

[7] Markou, M.; Singh, S. Novelty detection: a review—part 1: statistical approaches, *Signal Processing*, Volume 83, Issue 12, 2003, pp. 2481-2497.

[8] Pincus, S.M. Approximate Entropy as a Measure of System Complexity. *Proc Natl Acad Sci*, USA 88: pp. 2297–2301. 1991.

[9] Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circul. Physiol*., 278, H2039–H2049, 2000.

[10] Markou, M.; Singh, S. Novelty detection: a review—part 2: neural network based approaches, *Signal Processing*, Volume 83, Issue 12, pp. 2499-2521. 2003.

[11] Stephen Marsland. Novelty detection in learning systems. *Neural Computing Surveys*, 3:157-195, 2003.

[12] Demetriou, M.A: Polycarpou, M.M.: Incipient Fault Diagnosis of Dynamical Systems Using Online Approximators. *IEEE Transactions on Automatic Control*, , Vol. 43, No. 11., 1998.

[13] Trunov, A.B.; Polycarpou, M.M. Automated fault diagnosis in nonlinear multivariable systems using a learning methodology. *IEEE Transactions on Neural Networks*. vol. 11, no. 1, pp.91–101, 2000.

[14] Alippi, C.; Roveri, M. Just-in-Time Adaptive Classifiers—Part I: Detecting Nonstationary Changes. *IEEE Transactions on Neural Networks*, Vol. 19, No. 7. 2008.

[15] C Alippi, L Bu, D Zhao:"Prior-Free Encode-Decode Change Detection Test to Inspect Datastreams for Concept Drift", *The 2013 International Joint Conference on Neural Networks (IJCNN),* Dalas, USA, 2013.

[16] Ryan Elwell, Robi Polikar: "Incremental Learning of Concept Drift in Nonstationary Environments*",IEEE Transactions on Neural Networks*, vol. 22, no. 10, pp. 1517-1531, 2011.

[17] Bukovsky, I.: *¨Learning Entropy: Multiscale Measure for Incremental Learning¨*, journal of *Entropy*, special issue on Dynamical Systems, ISSN 1099–4300, 2013, *15*(10), 4159-4187; doi:10.3390/e15104159.

[18] Kinsner, W. A unified approach to fractal dimensions. *Intern. J. Cognitive Informatics and Natural Intelligence*, vol. 1, no. 4, October-December pp. 26-46., 2007.

[19] Gupta, M., M., Bukovsky, I., Homma, N. , Solo M. G. A., Hou Z.-G.: "Fundamentals of Higher Order Neural Networks for Modeling and Simulation", in *Artificial Higher Order Neural Networks for Modeling and Simulation*, ed. M. Zhang, IGI Global, 2012.

[20] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood, Cliffs, NJ: Prentice-Hall, 1985.

[21] Mandic, D.P. A Generalised Normalised Gradient Descent Algorithm. *IEEE Signal Processing Letters* ,Vol. 11, No. 2, pp.115-118. 2004.

[22] Young-Seok Choi, Hyun-Chool Shin, Woo-Jin Song: "Robust Regularization for Normalized LMS Algorithms", *IEEE Transactions on Circuits and Systems—II:Express Briefs*, Vol. 53, No.8, 2006.

[23] Qing Song, Yilei Wu, Yeng Chai Soh: "Robust Adaptive Gradient-Descent Training Algorithm for Recurrent Neural Networks in Discrete Time Domain", *IEEE Transactions on Neural Networks*, Vol.19, No.11, pp. 1841-53, 2008.

[24] *Yoshizawa-Homma Lab*, Research Division on Advanced Information Technology, Tohoku University, Japan.

[25] Guido van Rossum; Jelke de Boer. Linking a Stub Generator (AIL) to a Prototyping Language (Python). *Spring 1991 EurOpen Conference Proceedings* (May 20-24, 1991), Tromso, Norway.