Romanuke V.V.
Khmelnitskiy National University,
Khmelnitskiy, Ukraine,
E-mail: romanukevadimv@gmail.com

EQUALLY - WEIGHTED COMPOSITIONS OF GAUSSIAN - NOISED - DATA - TRAINED TWO - LAYER PERCEPTRONS IN BOOSTING ENSEMBLES FOR HIGH - ACCURATE DISCONTINUOUS TRACKING OF WEAR STATES REGARDING STATISTICAL DATA INACCURACIES AND SHIFTS

UDC 539.375.6+539.538+519.237.8

Equally - weighted compositions of Gaussian - noised-data - trained two - layer perceptrons are studied in order to track metal wear states more accurately at the highest level of statistical data inaccuracies and shifts (noise). The noise range is modeled through four magnitudes characterizing ultimate jitters and shifts in wear influencing factors. Accuracy and variance gains of equally - weighted compositions seem to be increasing when noise intensities become lower. When boosting ensembles are composed from ordinary classifiers, high-accurate tracking fails. Only composing ensembles from a lot of the best optimized perceptrons, the accuracy improves by 1,5 % for the averaged tracking error rate and by 7,7 % for the tracking error rate at noise maximum. Here, the boosting appears to have its limit. But ensembles of equally-weighted compositions of perceptrons perform even better than ensembles of perceptrons weighted after training. And for ensuring high-accurate discontinuous tracking of wear states, we just need perceptrons trained by quite different backpropagation methods.

Key words: wear state tracking, statistical data, data inaccuracies, data shifts, accuracy, two-layer perceptron, boosting, boosting ensemble, tracking error rate.

Generation of statistical data inaccuracies and shifts in measuring metal wear states

While metal billets are processed, metal processing tools are worn within a range of metal wear states (MWS) being specific for the type of metal and peculiarities of processing. The range is ordered by wear accumulation. While metal details are run, the range is wider due to that intensity of wear is much lesser.

MWS are measured (scored) via wear influencing factors (WIF), including direct (speed, pressure, temperature, etc.) and indirect (time duration) ones. Measuring MWS is needful for controlling wear process and prevention of underuse and overuse of both tools and details. The measurement is a ground for tracking MWS along the whole life cycle. The tracker is usually stated after arranging the huge statistical data (HSD). However, these data are inaccurate as wear is very stochastic process. Additionally, MWS measurements may have instrumental and systematic bias errors which shift WIF being matched to the corresponding MWS. That is why statistical data inaccuracies and shifts (SDIS) in measuring MWS hinder in effective application of continuous differential-equations-based-models for tracking MWS and forecasting them. And using stochastic differential equations doesn't help much as expectance evaluation is not so reliable and volatility is generated very high. Thus discontinuous models are preferable, though they require HSD to set reliable statistical correspondence [1] in the form of finite statistical data set (FSDS) including each of $N \in \mathbb{N} \setminus \{1\}$ wear states.

FSDS is $F_L = \left\{\mathbf{X}_j,\, w_j\right\}_{j=1}^{L}$ by $L \in \mathbb{N} \setminus \left\{\overline{1,\, N-1}\right\}$ and the wear state $w_j \in \left\{\overline{1,\, N}\right\}$ labeled as $\mathbf{X}_j = \left[\, x_i^{\langle j \rangle} \,\right]_{1 \times Q} \in X$ for $Q \in \mathbb{N}$ WIF. The range of MWS is presumed to be wholly sampled into those WIF, and so $\left\{w_j\right\}_{j=1}^{L} \cap \left\{\overline{1,\, N}\right\} = \left\{\overline{1,\, N}\right\}$ by the s-th state w_{j_s} labeled as the pure representative (PR) $\mathbf{X}_{j_s}$. SDIS are generated as a result of the wear is influenced with innumerable multitude of factors, what lets treat those inaccuracies as normal noise. Due to that, the classifier can be two-layer perceptron with nonlinear transfer functions (2LPNLTF) [1]. This is a universal classifier, performing greatly on Gaussian - noised data (GND). Nonetheless, possible shifts in FSDS and shifts in input data make single 2LPNLTF classifier noneffective at higher intensities of the shift noise. But compositions of GND-trained 2LPNLTF may improve accuracy of tracking MWS [2]. Therefore, boosting ensembles of GND-trained 2LPNLTF are to be tried exhaustively.

Gain in accuracy of tracking MWS at higher intensities of SDIS using boosting ensembles

Intensities of SDIS become higher when either the metal life cycle has run near its half or MWS are measured with poor-calibrated instrument. The accuracy of tracking MWS is tracking error rate (TER) being the percentage ratio of the classifier's incorrect responses to the classifier's total inputs, although the percentage of the classifier's correct responses among its total inputs is implied by TER.

Based on a boosting technique stated in [2], ensemble of three 2LPNLTF solved a problem of tracking 24 MWS with 16 WIF in that the mean ensemble TER was 6,82 % at the highest level of SDIS in every state. In term of TER, the averaged gain (relative decrement of TER or something else) of the boosting exceeded 50 %, and variance of wear states' TER became more than 50 % lower with the ensemble.

For the same problem, the classifier was optimized in [3] for improving accuracy of the tracking. Having increased number of 2LPNLTF within the ensemble up to 30, the averaged gain with the optimized ensemble became about 56 % in respect of the best ensemble of three 2LPNLTF. Similarly, variance of TER over 24 MWS became about 53 % lower. However, nearly the same results were registered when the ensemble was composed without training, but just setting the weight of every 2LPNLTF to one thirtieth. The explanation is that all those 2LPNLTF were roughly similarly-trained GND classifiers, without focusing on specific SDIS. Hence, equally-weighted compositions of $B \in \mathbb{N} \setminus \left\{\overline{1, 30}\right\}$ GND-trained 2LPNLTF are believed to track MWS in that problem more accurately.

The goal of composing a boosting ensemble of GND-trained 2LPNLTF for high-accurate tracking

Taken the spoken problem of tracking 24 MWS with 16 WIF from [2, 3], there should be composed a boosting ensemble of equally-weighted GND-trained 2LPNLTF for high-accurate tracking at the highest level of SDIS. The result is expected to be better than in [3]. In any case, evaluation of the ratio between TER for single 2LPNLTF and TER for the ensemble is going to be plotted. This implies both TER on average and TER at the highest level of SDIS. Similar plots of the boosting gain will be done for the variance of wear states' TER.

Accuracy and variance gains of equally-weighted compositions of GND-trained 2LPNLTF

In general statement, the boosting ensemble as B 2LPNLTF equally-weighted composition output is

$$\tilde{s}_* \in \arg\max_{s=\overline{1,\,N}} \tilde{m}_s \quad \text{by} \quad \tilde{m}_s = \frac{1}{B}\sum_{\alpha=1}^{B} m_s(\alpha) \tag{1}$$

on the forecasted MWS

$$s_* \in \arg\max_{s=\overline{1,\,N}} m_s(\alpha) \tag{2}$$

by the α -th 2LPNLTF [3], giving the value $m_s(\alpha)$ in its s -th output neuron. A 2LPNLTF itself is trained with training sets

$$\left\{\mathbf{Y} = \left[y_{is}\right]_{Q\times N} : y_{is} = x_i^{\langle j_s\rangle}\right\} \text{ and } \left\{\left\langle \left\{\mathbf{Y}\right\}_{r=1}^{R}, \left\{\tilde{\mathbf{Y}}_h\right\}_{h=1}^{H}\right\rangle : \tilde{\mathbf{Y}}_h = \mathbf{Y} + \sigma_h \cdot \mathbf{\Xi} + \mu \cdot \sigma_h \cdot \mathbf{\Theta},\ \sigma_h = h\sigma_0 H^{-1}\right.$$
$$\left.\forall h = \overline{1, H},\ H \in \mathbb{N},\ \sigma_0 > 0,\ \mathbf{\Xi} = \left[\xi_{is}\right]_{Q\times N},\ \xi_{is} \in \mathbf{N}\,(0, 1),\ \mu > 0,\ \mathbf{\Theta} = \left[\theta_{is}\right]_{Q\times N},\ \theta_{is} = \zeta_s \in \mathbf{N}\,(0, 1)\right\} \tag{3}$$

for PR and SDIS correspondingly, where $\left\{x_i^{\langle j_s\rangle} \in [0;\, 1]\right\}_{i=1}^{Q}$ and $R \in \mathbb{N} \cup \{0\}$, and $\mathbf{N}\,(0, 1)$ is the infinite set of standard normal variate's values. Term $\sigma_h \cdot \mathbf{\Xi}$ in (3) is responsible for modeling jitter inaccuracies and omissions in statistical data or measurements. And term $\mu \cdot \sigma_h \cdot \mathbf{\Theta}$ in (3) models WIF shifts in every state. Coefficient σ_0 characterizes ultimate jitters and μ is ratio between the suspected jitters and WIF shifts [1]. These ones were put to $\sigma_0 = 0.25$ and $\mu = 1.5$ whilst the problem of tracking 24 MWS was studied in [2, 3].

Denote by $\tilde{\rho}(\alpha)$ the averaged TER of the α -th 2LPNLTF along with its TER $\rho(\sigma_H;\, \alpha)$ at SDIS maximum. For the ensemble of B 2LPNLTF, $\tilde{\rho}_B$ is the averaged TER and $\rho_B(\sigma_H)$ is TER at SDIS maximum. Henceforward, the accuracy gains are the averaged TER gain and the gain of TER at SDIS maximum

$$\tilde{g}_{\text{TER}}(B) = \left(\tilde{\rho}_B\right)^{-1} \cdot \min_{\alpha=\overline{1,\,B}} \tilde{\rho}(\alpha) \quad \text{and} \quad g_{\text{TER}}(\sigma_H;\, B) = \left(\rho_B(\sigma_H)\right)^{-1} \cdot \min_{\alpha=\overline{1,\,B}} \rho(\sigma_H;\, \alpha) \tag{4}$$

respectively. For seeing the boosting gain for the variance of wear states' TER, denote by $\tilde{v}(\alpha, N)$ the averaged variance of N wear states' TER of the α -th 2LPNLTF along with its variance $v(\sigma_H;\, \alpha,\, N)$ at SDIS maximum. For the ensemble of B 2LPNLTF, $\tilde{v}_B(N)$ is the averaged variance of N wear states' TER and $v_B(\sigma_H;\, N)$ is variance at SDIS maximum. Henceforward, the variance gains are the averaged variance gain and the gain of variance at SDIS maximum

$$\tilde{g}_{\mathrm{var}}\left(B,\, N\right)=\left(\tilde{v}_{B}\left(N\right)\right)^{-1}\cdot\min_{\alpha=\overline{1,\, B}}\tilde{v}\left(\alpha,\, N\right) \quad \text{and} \quad g_{\mathrm{var}}\left(\sigma_{H};\, B,\, N\right)=\left(v_{B}\left(\sigma_{H};\, N\right)\right)^{-1}\cdot\min_{\alpha=\overline{1,\, B}}v\left(\sigma_{H};\, \alpha,\, N\right) \quad (5)$$

respectively. The accuracy gains (4) for the problem of tracking 24 MWS with 16 WIF are shown in Figure 1 and Figure 2, where $R=1$, $H=10$, and SDIS are modeled by

$$\sigma_{0}\in\left\{0.15,\, 0.2,\, 0.25,\, 0.3\right\} \quad \text{and} \quad \mu\in\left\{1,\, 1.25,\, 1.5,\, 2\right\}. \quad (6)$$

The variance gains (5) for the spoken problem are shown in Figure 3 and Figure 4, plotted by (6). The increment of SDIS intensity according to (6) is designated with dots, circles, squares, and diamonds, respectively.

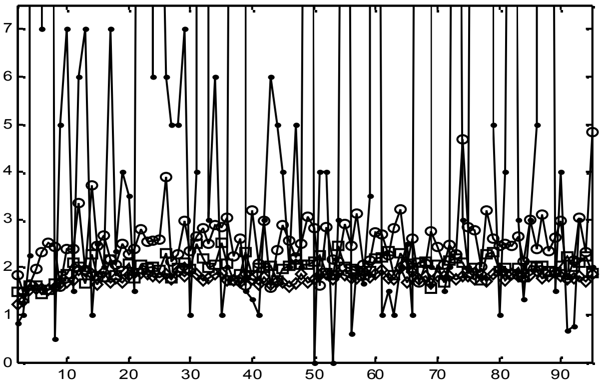

Fig. 1 – The averaged TER gain

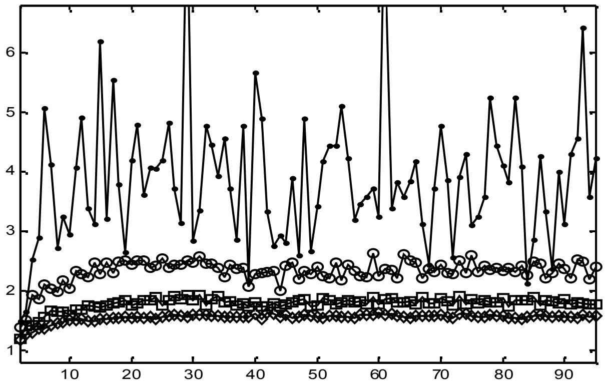

Fig. 2 – The gain of TER at SDIS maximum

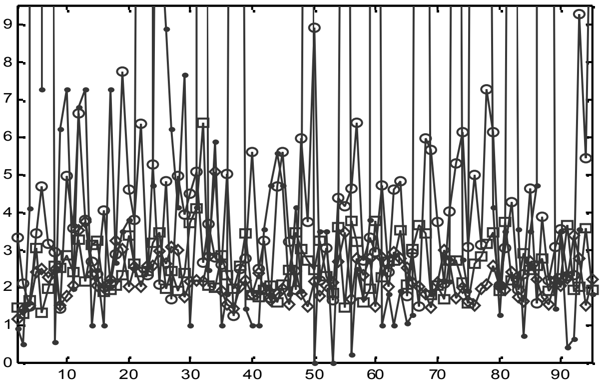

Fig. 3 – The gain of averaged variance of 24 wear states' TER

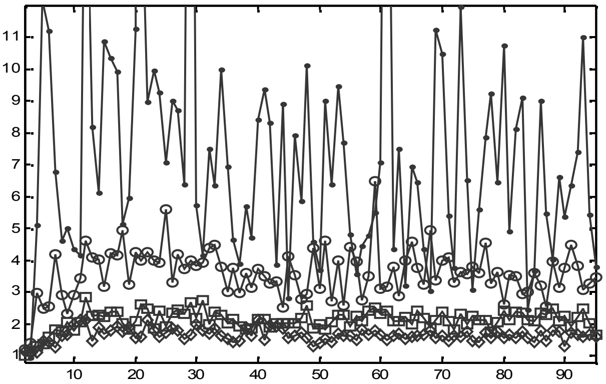

Fig. 4 – The gain of variance at SDIS maximum

Figures 1 - 4 show clearly that those gains unexpectedly are about those ones for 30 GND-trained 2LPNLTF weighted within the ensemble after training [3]. It means that tracking MWS more accurately fails when any $B\in\mathbb{N}\setminus\left\{\overline{1,\, 30}\right\}$ GND-trained 2LPNLTF are taken either into equally-weighted composition or into ensemble to be trained. Notwithstanding the fail, high-accurate discontinuous tracking (HADCT) is available if ensemble is composed of GND-trained 2LPNLTF of higher accuracy (HA2LPNLTF). Surely, such HA2LPNLTF occur rarer than an ordinary GND-trained 2LPNLTF. Among 17818 ordinary GND-trained 2LPNLTF with 70 neurons in hidden layer by $R=1$ and $H=18$, there happened to be found 60 HA2LPNLTF performing at $\tilde{\rho}\left(\alpha\right)\leqslant 1{,}17$ for the problem by $\sigma_{0}=0{,}25$ and $\mu=1{,}5$, and the equally-weighted composition of these HA2LPNLTF reinforced by resume-training performs at $\tilde{\rho}_{60}<0{,}62$, $\rho_{60}\left(\sigma_{20}\right)<4{,}43$, $\tilde{v}_{60}\left(24\right)<0{,}14$, $v_{60}\left(\sigma_{20};\, 24\right)<5{,}4$. Thus, the corresponding gains remain roughly as those ones in Figures 1 - 4, although HADCT is realizable.

Discussion of the gains and findings

Having processed HSD for trying HADCT, the boosting appears to have its limit. Unexpectedly, but HADCT here is realizable only by using a lot of the best HA2LPNLTF. The gains (4) and (5) seem to be increasing when intensities of SDIS become lower. But they have their own limits, nearly corresponding to the circled polylines in Figures 1 - 4. Comparing the gains to those ones in [3], the accuracy improvement exists anyway. It is about 1,5 % for the averaged TER and 7,7 % for TER at SDIS maximum. And ensembles of equally-weighted compositions of GND-trained 2LPNLTF perform even better than ensembles of GND-trained 2LPNLTF weighted after training. For ensuring HADCT, we just need 2LPNLTF trained by quite different methods of backpropagation - “traingda”, “traingdx”, “traingdm”, “traingd”, “trainlm”, “trainscg”, etc.

References

1. Romanuke V.V. Wear state discontinuous tracking model as two-layer perceptron with nonlinear transfer functions being trained on an extended general totality regarding statistical data inaccuracies and shifts / V. V. Romanuke // Problems of tribology. – 2014. – N. 3. – P. 50-53.

2. Romanuke V.V. Accuracy improvement in wear state discontinuous tracking model regarding statistical data inaccuracies and shifts with boosting mini-ensemble of two-layer perceptrons / V.V. Romanuke // Problems of tribology. – 2014. – N. 4. – P. 55-58.

3. Romanuke V.V. Optimizing parameters of the two-layer perceptrons’ boosting ensemble training for accuracy improvement in wear state discontinuous tracking model regarding statistical data inaccuracies and shifts / V.V. Romanuke // Problems of tribology. – 2015. – N. 1. – P. 65-68.

UDC 539.375.6+539.538+519.237.8

Поступила в редакцію 15.04.2015

Романюк В.В. **Рівнозважені склади двошарових персептронів, навчених на даних з гаусовими шумами, у комітетах бустингу для високоточного дискретного відслідковування станів зносу з урахуванням похибок і зсувів у статистичних даних.**

Вивчаються рівнозважені склади двошарових персептронів, навчених на даних з гаусовими шумами, для того, щоб відслідковувати стани зносу металу більш точно за найвищого рівня похибок і зсувів у статистичних даних (шуму). Діапазон цього шуму моделюється за чотирма амплітудами, що характеризують граничні флуктуаційні похибки та зсуви у факторах впливу на знос. Виграші у точності та дисперсії цих рівнозважених складів здаються зростаючими при зменшенні інтенсивностей шуму. Коли комітети бустингу складаються зі звичайних класифікаторів, високоточне відслідковування не вдається. Лише при складанні комітетів зі значного числа найкращих оптимізованих персептронів точність покращується на 1,5 % для усередненого рівня помилок відслідковування і на 7,7 % для рівня помилок відслідковування за максимального шуму. Здається, бустинг тут має свою межу. Однак комітети рівнозважених складів персептронів функціонують навіть краще, ніж комітети персептронів, що зважуються після навчання. І для забезпечення високоточного дискретного відслідковування станів зносу нам необхідні просто персептрони, навчені за доволі різними методами зворотного поширення.

Ключові слова: відслідковування стану зносу, статистичні дані, похибки у даних, зсуви у даних, точність, двошаровий персептрон, бустинг, комітет бустингу, рівень помилок відслідковування.