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

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

UDC 539.375.6+539.538+519.237.8

There is a trial of optimization for improving accuracy in tracking metal tool wear states discontinuously, when the states' finite set has been statistically tied to the set of representative wear influencing factors. Range of wear states is presumed to be wholly sampled into those factors. The tracker is a static model based on boosting ensemble of two-layer perceptrons with nonlinear transfer functions. It successfully regards statistical data inaccuracies and shifts in a problem of tracking 24 wear states featured with 16 wear influencing factors. Having increased number of classifiers within the ensemble up to 30, the averaged gain with the optimized ensemble is about 56 % in respect of the best ensemble of three classifiers. Similarly, variance of tracking error rate over 24 wear states is about 53 % lower. Nearly the same results are registered when the ensemble is composed without training, but just setting every classifier's weight to one thirtieth. To get the perfected accuracy more, such equally-weighted compositions shall be investigated in the sequel.

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

Tracking wear states regarding statistical data inaccuracies and shifts

Metal processing is an inseparable part of heavy industry. For rationalized usage of billets and tools preventing their underuse and overuse, metal wear states are tracked and forecasted. At this, unavoidable statistical data inaccuracies and shifts (SDIS) of wear influencing factors (WIF) should be regarded because of the high stochasticity of the wear process. In tracking wear states regarding SDIS, the tracking accuracy is improved either with refinement of continuous models or accumulating additional statistics for discontinuous tracking models based on statistical correspondence [1, 2]. The statistical correspondence approach [1] looks well-promising inasmuch as regarding SDIS is possible with manipulating huge statistics anyway. Nonetheless, the high stochasticity of the wear process provoking the spoken SDIS allows using universal classifiers which perform greatly on Gaussian-noised data (GND). And namely GND are specificity of wear influenced with a great deal of factors (which, upon the whole, are innumerable). This lets treat those noises as normal.

Approaches to wear state tracking accuracy improvement and the gain

With universal classifiers of GND, there are two major approaches to improve their accuracy. They are training process optimization and boosting. Based on a boosting technique stated in [3] for ensemble of three learners, where everyone is two-layer perceptron with nonlinear transfer functions (2LPNLTF), the averaged gain of the boosting in tracking 24 wear states with 16 WIF exceeded 50 %. This gain is defined with the tracking error rate (TER) indicating virtually the percentage of the classifier's correct responses among the total inputs. At the highest level of SDIS in every state, the mean ensemble TER was 6.82 %, and the averaged TER varied between 0.96 % and 1.12 %. And with the ensemble, moreover, variance of wear states' TER became more than 50 % lower.

However, the accuracy gains were reached at the raw parameters of the boosting training process (BTP). The rule for redistributing weights of the training samples was naive as well. Thus, those parameters may be optimized in order to get the accuracy perfected.

The article goal and tasks

Taken the example of tracking 24 wear states with 16 WIF from [3], the 2LPNLTF-boosting gain is going to be improved. The improvement is the statistical approximation accuracy increment. For this, two raw parameters of BTP are swept within their ranges to see the lowest TER. Along with that, the linear function rule for redistributing weights of the training samples will be adjusted through a set of nonlinear functions. Eventually, we are to evaluate the new gain after testing the optimized 2LPNLTF-boosting.

Tracking wear states with boosting ensemble of 2LPNLTF

In general, there are $Q \in \mathbb{N}$ WIF and $N \in \mathbb{N} \setminus \{1\}$ wear states. The forecasted state by 2LPNLTF is

$$s_* \in \arg\max_{s=\overline{1,\,N}} v_s\,, \quad v_s = \left(1+\exp\left[-\left(\sum_{k=1}^{S_{\mathrm{HL}}} u_{ks}\cdot\left(1+\exp\left[-\left(\sum_{i=1}^{Q} x_i a_{ik} + h_k\right)\right]\right)^{-1} + b_s\right)\right]\right)^{-1} \tag{1}$$

by $\mathbf{X}=\left[x_i\right]_{1\times Q}\in X\subset\mathbb{R}^Q$ and S_{HL} neurons in 2LPNLTF hidden layer. 2LPNLTF coefficients in matrices $\left[a_{ik}\right]_{Q\times S_{\mathrm{SHL}}}$, $\left[u_{ks}\right]_{S_{\mathrm{SHL}}\times N}$, $\left[h_k\right]_{1\times S_{\mathrm{SHL}}}$, $\left[b_s\right]_{1\times N}$, particularly, in [2, 3] were determined by three methods: "traingda" ($\alpha=1$), "traingdx" ($\alpha=2$), "trainscg" ($\alpha=3$). Initially, finite statistical data set (FSDS) is $\left\{\mathbf{X}_j,\, w_j\right\}_{j=1}^{L}$ by $L\in\mathbb{N}\backslash\left\{\overline{1,\,N-1}\right\}$ and the wear state $w_j\in\left\{\overline{1,\,N}\right\}$ labeled as $\mathbf{X}_j=\left[x_i^{\langle j\rangle}\right]_{1\times Q}\in X$. All possible wear states are represented in FSDS: $\left\{w_j\right\}_{j=1}^{L}\bigcap\left\{\overline{1,\,N}\right\}=\left\{\overline{1,\,N}\right\}$ by the s-th state w_{j_s} labeled as the pure representative (PR) $\mathbf{X}_{j_s}$. The 2LPNLTF training sets are $\left\{\mathbf{Y}=\left[y_{is}\right]_{Q\times N} : y_{is}=x_i^{\langle j_s\rangle}\right\}$ and

$$\left\{\left\langle\left\{\mathbf{Y}\right\}_{r=1}^{R},\left\{\tilde{\mathbf{Y}}_h\right\}_{h=1}^{H}\right\rangle : \tilde{\mathbf{Y}}_h=\mathbf{Y}+\sigma_h\cdot\boldsymbol{\Xi}+\mu\cdot\sigma_h\cdot\boldsymbol{\Theta},\ \sigma_h=h\sigma_0 H^{-1}\ \forall\, h=\overline{1,\,H},\ H\in\mathbb{N}\,,\right.$$

$$\left.\sigma_0>0,\ \boldsymbol{\Xi}=\left[\xi_{is}\right]_{Q\times N},\ \xi_{is}\in\mathbf{N}\ (0,1),\ \mu>0,\ \boldsymbol{\Theta}=\left[\theta_{is}\right]_{Q\times N},\ \theta_{is}=\zeta_s\in\mathbf{N}\ (0,1)\right\} \tag{2}$$

for PR and SDIS correspondingly, where $R\in\mathbb{N}\bigcup\{0\}$ and $\mathbf{N}\ (0,1)$ is the infinite set of standard normal variate's values. Denoting the α-th 2LPNLTF value v_s in (1) as $v_s(\alpha)$, the boosted classifier output is

$$s_*\in\arg\max_{s=\overline{1,\,N}}\tilde{v}_s \quad \text{by} \quad \tilde{v}_s=\sum_{\alpha=1}^{B}\beta(\alpha)v_s(\alpha) \tag{3}$$

for $B\in\mathbb{N}\backslash\{1\}$ 2LPNLTF in the ensemble, where weights $\left\{\beta(\alpha)\right\}_{\alpha=1}^{B}$ are determined as follows. In training the ensemble, the set (2) is re-generated for T times, forming FSDS of $M=(R+H)\cdot N\cdot T$ training samples. These samples have the weights in $\mathbf{D}(q)=\left[d_\tau(q)\right]_{1\times M}$ at the q-th iteration of BTP by $q=\overline{1,\,q_0}$ at some final iteration number q_0, where $d_\tau(1)=M^{-1}\ \ \forall\,\tau=\overline{1,\,M}$. Matrix $\mathbf{A}=\left[\overline{a}_{\alpha\tau}\right]_{B\times M}$ is of flags of classifiers' correct responses, where $\overline{a}_{\alpha\tau}=1$ is the correct classification of τ-th sample by the α-th 2LPNLTF, otherwise $\overline{a}_{\alpha\tau}=0$. The weighted errors are in $\mathbf{E}(q)=\left[\eta_\alpha(q)\right]_{B\times M}$, where the α-th classifier's weighted error

$$\eta_\alpha(q)=\sum_{\tau=1}^{M}d_\tau(q)\cdot\left(1-\overline{a}_{\alpha\tau}\right),\ \alpha=\overline{1,\,B}\,. \tag{4}$$

Starting from $q=1$, there are found the best 2LPNLTF $\alpha_*(q)$ and the minimal weighted error $\eta_*(q)$,

$$\alpha_*(q)\in\arg\min_{\alpha=\overline{1,\,B}}\eta_\alpha(q) \quad \text{and} \quad \eta_*(q)=\min_{\alpha=\overline{1,\,B}}\eta_\alpha(q) \tag{5}$$

respectively, letting learn the coefficient

$$\gamma(q)=1-\eta_*(q) \tag{6}$$

and calculate the new distribution $\mathbf{D}(q+1)$ of weights

$$d_\tau(q+1)=\tilde{d}_\tau\Bigg/\sum_{\upsilon=1}^{M}\tilde{d}_\upsilon \quad \text{by} \quad \tilde{d}_\tau=d_\tau(q)\cdot\exp\left[-\gamma(q)\left(2\cdot\overline{a}_{\alpha_*(q),\tau}-1\right)\right] \tag{7}$$

over M training samples. If $\eta_*(q)<1-N^{-1}$ then $\tilde{q}=q$ and $q=\tilde{q}+1$, and (4) — (7) are re-found. If $\eta_*(q)\geqslant 1-N^{-1}$ then $q_0=q$ and there are calculated the coefficients

$$\tilde{\gamma}(q)=\gamma(q)\Bigg/\sum_{p=1}^{q_0}\gamma(p) \quad \text{by} \quad q=\overline{1,\,q_0} \quad \text{for} \quad \beta(\alpha)=\sum_{q\in\left\{\overline{1,\,q_0}\right\},\ \alpha=\alpha_*(q)}\tilde{\gamma}(q)\quad \forall\,\alpha=\overline{1,\,B}\,. \tag{8}$$

For regarding SDIS modeled with $\sigma_0 = 0.25$ and $\mu = 1.5$ for $Q = 16$, $N = 24$, $\left\{ x_i^{\langle j_s \rangle} \in [0;\, 1] \right\}_{i=1}^{16}$, by $S_{\mathrm{HL}} = 45$ and $B = 3$ in [3], FSDS (2) for BTP was formed by $R = 1$, $H = 20$, $T = 100$. Denote the averaged TER by $\rho(H, T, g)$ for the power $g > 0$ to form a set of nonlinear functions $\gamma(q) = 1 - \left[\eta_*(q) \right]^g$ instead of (6). Thus, the problem is $\min_{H \in \mathbb{N}} \min_{T \in \mathbb{N}} \min_{g > 0} \rho(H, T, g)$ and to find H_*, T_*, g_*, which

$$\{H_*, T_*, g_*\} = \arg \min_{\{H \in \mathbb{N},\, T \in \mathbb{N},\, g > 0\}} \rho(H, T, g). \tag{9}$$

Having swept those three parameters of BTP within their reasonable ranges, there has been exposed that by $H \in \left\{ \overline{1, 40} \right\}$ and $T \in \left\{ \overline{1, 200} \right\}$ and $g \in (0; 10)$ TER remains nearly the same. At the most, the averaged value $\rho(H_*, T_*, g_*)$ doesn't seem to be significantly less than the registered TER in [3]. An important fact is that by $H \in \left\{ \overline{1, 5} \right\}$ and $T \in \left\{ \overline{1, 5} \right\}$ the inequality $\eta_*(q) \geqslant 1 - N^{-1}$ is never true for some BTP. Sometimes this gap can be dealt through setting up a best classifier weighted error tolerance (BCWET) $\varepsilon_{\mathrm{BCWET}} > 0$. Then the items (4) — (7) are re-found while $\eta_*(q) < 1 - N^{-1} - \varepsilon_{\mathrm{BCWET}}$. For this, BCWET may be taken as $\varepsilon_{\mathrm{BCWET}} \in \left\{ 0.1,\, N^{-1},\, 0.01 \right\}$ or other reasonable values. Nonetheless, here $H_* = T_* = 6$ and $g_* = 1$ because the greater values of H and T the longer BTP is.

However, when assembling much greater number of 2LPNLTF, TER becomes lower. For instance, it is expected $\rho(20, 100, 1) < 0.75$ with $B = 30$ classifiers. Amazingly enough, but here TER also isn't influenced much when H and T increase. And with $H = T = 30$ BTP convergence is ensured even for $\varepsilon_{\mathrm{BCWET}} = 0$.

Optimization results and discussion

In a problem of 24 wear states tracking with universal GND classifiers, the boosting ensemble of 30 2LPNLTF is trained optimally under parameters $H_* = 30$ and $T_* = 40$ and $g_* = 1$. This is so because, partially, by $H = T = 20$ misconvergence was spotted. Although the inequality $\rho(60,\, 200,\, 1) < \rho(30,\, 40,\, 1)$ is expected, time of BTP is obviously shorter for lower H and T, and the difference between $\rho(60,\, 200,\, 1)$ and $\rho(30,\, 40,\, 1)$ is too small and unstable. The best ensemble has $\rho(30,\, 40,\, 1) < 0.63$ and it takes about 330 seconds to train it, i. e. to find the distribution $\left\{ \beta(\alpha) \right\}_{\alpha=1}^{30}$. Moreover, $\rho(30,\, 40,\, 1) < 4.8$ at the highest level of SDIS in every state, giving the 56 % gain in respect of the best ensemble with three 2LPNLTF. And variance of 24 wear states' TER has become 53 % lower when used 10 times greater number of 2LPNLTF.

The astounding event is that without factual training, but just setting $\beta(\alpha) = 30^{-1}$ $\forall\, \alpha = \overline{1,\, 30}$, TER close to optimal $\rho(30,\, 40,\, 1)$ is revealed. This is explained with that all 2LPNLTF are roughly similarly-trained GND classifiers, without focusing on specific SDIS. Such fact is a heuristic alternative to optimization of BTP.

Conclusion

Tracking wear states regarding SDIS at low TER is very essential for metal processing. With the optimized boosting 2LPNLTF ensemble as a wear state discontinuous tracking model, the accuracy is reached higher. In the presented example, just every twentieth state is tracked erroneously at the highest level of SDIS. On average, just a state from 158 is tracked erroneously. All what is needed is high-precision statistical correspondence of 24 wear states and 24 groups of 16-dimensional points labeled as 24 PR. The rest correspondence is modeled as FSDS (2) due to that wear is valued as GND. It's a way of real implementation of rationalized usage of billets and tools under controlling their wear. Generally, the optimized 2LPNLTF-boosting shall work in solving problems having various Q and N. A remarkable property of straight boosting under simply $\beta(\alpha) = B^{-1}$ $\forall\, \alpha = \overline{1,\, B}$ is going to be explored to perfect the tracker accuracy further.

References

1. Chungchoo C. On-line tool wear estimation in CNC turning operations using fuzzy neural network model / C. Chungchoo, D. Saini. International Journal of Machine Tools and Manufacture. 2002. Volume 42, Issue 1. P. 29 - 40.

2. 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. Problems of tribology. 2014. N. 3. P. 50 - 53.

3. 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. Problems of tribology. 2014. N. 4. P. 55 - 58.

Поступила в редакцію 09.02.2015

Романюк В. В. **Оптимізація параметрів навчання комітету бустингу двошарових персептронів для покращення точності у дискретній моделі відслідковування стану зносу з урахуванням похибок і зсувів у статистичних даних.**

Представляється спроба оптимізації для покращення точності дискретного відслідковування станів зносу металевого засобу, коли скінченна множина цих станів була статистично пов'язана з множиною репрезентативних факторів, що впливають на знос. Діапазон станів зносу вважається повністю розбитим за цими факторами. Відстежувачем є статична модель на основі комітету бустингу двошарових персептронів з нелінійними передавальними функціями. Вона успішно враховує похибки і зсуви у статистичних даних в задачі відслідковування 24 станів зносу з 16 факторами впливу на знос. Збільшивши кількість класифікаторів у комітеті до 30, усереднений виграш з оптимізованим комітетом складає близько 56 % по відношенню до найкращого комітету з трьох класифікаторів. Аналогічно дисперсія рівня помилок відслідковування по 24 станам зносу є майже на 53 % меншою. Приблизно такі самі результати зафіксовані тоді, коли комітет складається без навчання, а лише з прирівнюванням ваги кожного класифікатора до однієї тридцятої. Такі рівноважні композиції будуть досліджені у подальшому для того, щоб отримати ще більш вдосконалену точність.

Ключові слова: стан зносу, статистичні дані, похибки у даних, зсуви у даних, модель відслідковування, точність, двошаровий персептрон, бустинг, навчання комітету бустингу, оптимізація, рівень помилок відслідковування.