УДК 004.032.26

БОДЯНСЬКИЙ Є.В., д.т.н., професор,
ДЕЙНЕКО А.О., м.н.с.,
ДЕЙНЕКО Ж.В., к.т.н.,
ШАЛАМОВ М.О., аспірант (Харківський національний університет радіоелектроніки)

Адаптивне навчання нейронної мережі опорних векторів найменших квадратів

У роботі запропоновано адаптивний метод навчання нейронної мережі опорних векторів найменших квадратів (LS-SVM) незмінної архітектури. Особливістю цього методу є те, що мінімізація критерія емпіричного ризику відбувається на «ковзному вікні» фіксованої розмірності, що суттєво спрощує чисельну реалізацію процедури та дозволяє обробити інформацію, що генерується нестаціонарними об'єктами.

Ключові слова: *нейронні мережі, нейронна мережа опорних векторів найменших квадратів, критерій емпіричного ризику, «ковзне вікно»*

Вступ

Для розв'язання широкого класу задач обробки інформації, ідентифікації систем і об'єктів керування, насамперед, суттєво нелінійних та функціонуючих за умов структурної й параметричної невизначеності, широке розповсюдження отримали штучні нейронні мережі (ШНМ), завдяки, насамперед, універсальним апроксимуючим властивостям і здатності до навчання. При цьому, кажучи про навчання, слідураховувати можливість як настроювання синаптичних ваг і параметрів активаційних функцій, так і архітектури мережі в цілому. Ці властивості характерні для досить нового класу ШНМ – еволюційних коннекціоністських систем [1], що мають низку переваг перед традиційними нейромережами.

В ситуаціях, коли необхідна обробка інформації в on-line режимі при послідовному надходженні на вхід системи нових даних, на перший план виходить питання швидкості процесу навчання, що суттєво обмежує клас ШНМ, придатних для роботи в цьому режимі. З погляду оптимізації по швидкості процесу навчання досить перспективними є ШНМ, що базуються на ядерних (радіально-базисних, потенційних, дзвонуватих) функціях активації, найбільш характерними представниками яких є радіально-базисні нейронні мережі (RBFN), узагальнені регресійні нейронні мережі (GRNN) та машини опорних векторів (SVM) [2].

Постановка задачі дослідження

Машина опорних векторів (SVM) є своєрідним нейромережевим гібридом, що об'єднує в собі навчання, яке базується як на оптимізації, так і на пам'яті, а також реалізує метод мінімізації емпіричного ризику. Ключовим поняттям при синтезі цієї мережі є опорні вектори, що формують невелику підмножину найбільш інформативних векторів даних, виділених у процесі навчання. Машини опорних векторів є найбільш ефективними нейромережами за умов малих вибірок даних, що забезпечують високу якість апроксимації.

В той же час використання SVM-нейромереж обмежується тим, що їх навчання може бути реалізоване лише у пакетному режимі, оскільки кількість нейронів у такій мережі повністю визначається обсягом навчальної вибірки. Саме цей факт суттєво обмежує їх використання у задачах Dynamic Data Mining.

Тому основною задачею, що розглядається та вирішується у цій статті, є синтез адаптивного методу навчання нейронної мережі опорних векторів найменших квадратів, який дозволяє опрацьовувати інформацію в on-line режимі при фіксованій архітектурі та послідовному надходженню даних на обробку.

Основна частина

Адаптивний метод навчання нейронної мережі LS-SVM. Основним недоліком звичайних SVM є чисельна громіздкість процедури визначення синаптичних ваг, що зводиться до задачі нелінійного програмування з обмеженнями-нерівностями, кількість яких визначається об'ємом оброблюваної вибірки. Більш ефективними з цього погляду є машини опорних векторів, що базуються на методі найменших квадратів (LS-SVM) [3] та навчаються за допомогою методів квадратичного програмування, однак так або інакше це навчання відбувається в пакетному режимі.

Перетворення, що реалізується машиною опорних векторів, може бути записане у вигляді

$$\hat{y}^{SV}(x) = \left(w_x^{SV}\right)^T \varphi^{SV}(x) + w_0^{SV}, \tag{1}$$

де $w_x^{SV} = \left(w_1^{SV}, ..., w_l^{SV}, ..., w_{h_{SV}}^{SV}\right)^T$,

$\varphi^{SV}(x) = (\varphi_1^{SV}(x), ..., \varphi_{h_{SV}}^{SV}(x))^T$, а її навчання (у випадку LS-SVM) зводиться до одночасного встановлення центрів активаційних функцій у точках навчальної вибірки $x(l), l = 1, 2, ..., k$ подібно GRNN [4] й оптимізації квадратичного критерію

$$E^{SV}(k) = \frac{1}{2}\left(w_x^{SV}\right)^T w_x^{SV} + \frac{\gamma}{2}\sum_k e^2(k)$$

за наявності системи з $h_{SV} = k$ лінійних обмежень-рівностей:

$$\begin{cases} y(1) = \left(w_x^{SV}\right)^T \varphi^{SV}(x(1)) + w_0^{SV} + e(1), \\ \vdots \\ y(k) = \left(w_x^{SV}\right)^T \varphi^{SV}(x(k)) + w_0^{SV} + e(k), \end{cases}$$

де $\gamma > 0$ – параметр регуляризації,

$e(l) = y(l) - \hat{y}^{SV}(x(l)),\ l = 1, 2, ..., k.$

У пакетному режимі задача навчання LS-SVM зводиться до знаходження сідлової точки функції Лагранжа

$$\begin{gathered} L(w_x^{SV}, w_0^{SV}, e(k), \lambda(k)) = E^{SV}(k) + \sum_k \lambda(k)(y(k) - \left(w_x^{SV}\right)^T \varphi^{SV}(x(k)) - \\ -w_0^{SV} - e(k)) = \frac{1}{2}\left(w_x^{SV}\right)^T w_x^{SV} + \frac{\gamma}{2}\sum_k e^2(k) + \sum_k \lambda(k)(y(k) - \left(w_x^{SV}\right)^T \varphi^{SV}(x(k)) - w_0^{SV} - e(k)), \end{gathered} \tag{2}$$

при цьому крім власне синаптичних ваг w_x^{SV} і w_0^{SV} повинні бути знайдені й k невизначених множників Лагранжа $\lambda(l)$.

Система рівнянь Куна-Таккера для лагранжіана (2) може бути записана у вигляді

$$\begin{cases} \nabla_{w_x^{SV}} L = w_x^{SV} - \sum_k \lambda(k)\varphi^{SV}(x(k)) = \vec{0}_k, \\ \dfrac{\partial L}{\partial w_0^{SV}} = -\sum_k \lambda(k) = 0, \\ \dfrac{\partial L}{\partial e(l)} = \gamma e(l) - \lambda(l) = 0, \\ \dfrac{\partial L}{\partial \lambda(l)} = y(l) - \left(w_x^{SV}\right)^T \varphi(x(l)) - w_0^{SV} - e(l) = 0 \end{cases}$$

(тут $\vec{0}_k$ – $(k \times 1)$ – вектор, утворений нулями), або

$$\begin{cases} w_x^{SV} = \sum_k \lambda(k)\varphi^{SV}(x(k)), \\ \sum_k \lambda(k) = 0, \\ \lambda(l) = \gamma e(l), \\ y(l) - w_x^{SV\,T}\varphi(x(l)) - w_0^{SV} - e(l) = 0. \end{cases} \tag{3}$$

З першого рівняння системи (3) випливає, що синаптичні ваги цілком залежать від значень невизначених множників Лагранжа, у зв'язку з чим навчання LS-SVM зводиться до їхнього визначення, при цьому система (3) може бути переписана в компактному вигляді

$$\begin{pmatrix} 0 & I_k^T \\ I_k & \Omega(k) + \gamma^{-1} I_{k,k} \end{pmatrix} \begin{pmatrix} w_0^{SV} \\ \Lambda(k) \end{pmatrix} = \begin{pmatrix} 0 \\ Y(k) \end{pmatrix}, \tag{4}$$

де $\Lambda(k) = (\lambda(1), ..., \lambda(l), ..., \lambda(k))^T$,

$I_k - (k \times 1)$ – вектор, утворений одиницями,

$I_{k,k} - (k \times k)$ – одинична матриця,

$i = 1, 2, ..., k,\quad j = 1, 2, ..., k,$

$Y(k) = \left(y(1), ..., \ y(l), ..., \ y(k)\right)^T$,

$\Omega(k) = \left\{\Omega_{ij} = \left(\varphi^{SV}\right)^T (x(i))\varphi^{SV}(x(j)) = K(x(i), x(j))\right\}$,

$K = (x\ (i), x\ (j))$ – деяка ядерна функція, що задовільняє умовам теореми Мерсера [2], найчастіше той же ґавсіан

$$K(x(i), x(j)) = \exp\left(-\frac{\left\|x(i) - x(j)\right\|^2}{2\sigma^2}\right). \tag{5}$$

При цьому перетворення (1), що реалізується машиною опорних векторів, може бути переписане у формі

$$\hat{y}^{SV}(x) = \sum_k \lambda(k) K(x, x(k)) + w_0^{SV},$$

а його параметри $\lambda(k)$, w_0^{SV} можуть бути знайдені безпосередньо з (4) у вигляді

$$\begin{pmatrix} w_0^{SV} \\ \Lambda(k) \end{pmatrix} = \begin{pmatrix} 0 & I_k^T \\ I_k & \Omega(k)+\gamma^{-1}I_{k,k} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k) \end{pmatrix} = P^{SV}(k) \begin{pmatrix} 0 \\ Y(k) \end{pmatrix}. \quad (6)$$

Зрозуміло, що адаптивне навчання SVM може бути організоване на основі чисельно простої процедури обернення матриці в правій частині системи (6).

Переписуючи (6) для $(k+1)$ – го моменту часу у вигляді

$$\begin{pmatrix} w_0^{SV} \\ \Lambda(k+1) \end{pmatrix} = \begin{pmatrix} 0 & I_{k+1}^T \\ I_k & \Omega(k+1)+\gamma^{-1}I_{k+1,k+1} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k+1) \end{pmatrix} = \begin{pmatrix} \begin{matrix} 0 & I_k^T \\ I_k & \Omega(k)+\gamma^{-1}I_{k,k} \end{matrix} & \begin{matrix} 1 \\ K(x(1),x(k+1)) \\ K(x(2),x(k+1)) \\ \vdots \\ K(x(k),x(k+1)) \end{matrix} \\ 1K(x(1),x(k+1))\ldots K(x(k),x(k+1)) & 1+\gamma^{-1} \end{pmatrix} =$$

$$= \begin{pmatrix} 0 \\ Y(k) \\ y(k+1) \end{pmatrix} = \begin{pmatrix} (P^{SV}(k))^{-1} & \vec{K}(x(i),x(k+1)) \\ \vec{K}^T(x(i),x(k+1)) & 1+\gamma^{-1} \end{pmatrix}^{-1} \times \begin{pmatrix} \vec{Y}(k) \\ y(k+1) \end{pmatrix}$$

(тут $\vec{K}^T(x(i),\ x(k+1)) = (1, K(x(1),\ x(k+1)),\ldots,\ K(x(k),\ x(k+1)))^T$, $\vec{Y}^T(k) = (0, Y^T(k))$) і, застосовуючи формулу Фробеніуса обернення блокових матриць [5], приходимо до простого виразу для розрахунків матриці $P^{SV}(k+1)$ [6]:

$$P^{SV}(k+1) = \begin{pmatrix} \begin{matrix} P^{SV}(k)+(P^{SV}(k)\vec{K}(x(i), \\ x(k+1))\vec{K}^T(x(i),x(k+1))P^{SV}) \\ (1+\gamma^{-1}-\vec{K}^T(x(i),x(k+1)) \\ P^{SV}(k)\vec{K}(x(i),x(k+1)))^{-1} \end{matrix} & \begin{matrix} -(P^{SV}(k)\vec{K}(x(i),x(k+1))) \\ (1+\gamma^{-1}-\vec{K}^T(x(i), \\ x(k+1))P^{SV}(k)\vec{K}(x(i), \\ x(k+1)))^{-1} \end{matrix} \\ \\ \begin{matrix} -(\vec{K}^T(x(i),x(k+1))P^{SV}(k)) \\ (1+\gamma^{-1}-\vec{K}^T(x(i),x(k+1)) \\ P^{SV}(k)\vec{K}(x(i),x(k+1)))^{-1} \end{matrix} & \begin{matrix} (1+\gamma^{-1}-\vec{K}^T(x(k), \\ x(k+1))P^{SV}(k)\vec{K}(x(i), \\ x(k+1)))^{-1} \end{matrix} \end{pmatrix}. \quad (7)$$

Зрозуміло, що при великих обсягах навчальної вибірки, обернення $(k\times k)$ – матриць набагато зручніше робити, використовуючи формулу (7).

Навчання нейронної мережі LS-SVM на «ковзному вікні». Слід пам’ятати, що із зростанням навчальної вибірки, збільшується і кількість нейронів у нейронній мережі, що рано чи пізно призведе до виникнення ефекту «прокльону розмірності». Тому є доцільним, особливо у випадках коли об’єкт, що генерує дані, є нестаціонарним, організувати обробку інформації на «ковзному вікні», що містить s останніх спостережень, що, у свою чергу, призведе до того, що і нейронна мережа буде утворена s вузлами.

У цьому випадку, вводячи у розгляд $(s\times s)$ – матрицю ядерних функцій

$$\Omega(k,s) = \{\Omega_{ij} = \varphi^{SVT}\ (x(i)\varphi^{SV}(x(j)) = K(x(i),x(j),s)),$$
$$i = k-s+1,\ k-s+2,\ \ldots,k;\ j = k-s+1,\ k-s+2,\ \ldots,k\},$$

можна записати перетворення, що реалізується нейромережею з фіксованою кількістю нейронів у вигляді

$$\hat{y}^{SV}(x,k,s) = \sum_{l=k-s+1}^{k} \lambda(l,s)\times K(x,x(l),s) + w_0^{SV}(k,s).$$

Параметри, цього перетворення $\lambda(l,s)$, $w_0^{SV}(k,s)$ можуть бути знайдені шляхом вирішення матричного рівняння типу (6)

$$\begin{pmatrix} w_0^{SV} \\ \Lambda(k,s) \end{pmatrix} = \begin{pmatrix} 0 & I_s^T \\ I_s & \Omega(k,s)+\gamma^{-1}I_{s,s} \end{pmatrix}^{-1} \times \begin{pmatrix} 0 \\ Y(k,s) \end{pmatrix} = P^{SV}(k,s)\begin{pmatrix} 0 \\ Y(k,s) \end{pmatrix}, \qquad (8)$$

де $\Lambda(k,s) = \left(\lambda(k-s+1,s),\ \lambda(k-s+2,s),\ ...,\ \lambda(k,s)\right)^T$, $Y(k,s) = \left(y(k-s+1)\ \ y(k-s+2),...,\ y(k)\right)^T$.

З надходженням $(k+1)-$го спостереження, воно повинне бути враховано у матриці $\Omega(k+1,s)$ і одночасно з цим з неї повинно бути виключене спостереження, що надійшло у момент часу $k-s+1$. При цьому

$$\begin{pmatrix} w_0^{SV}(k+1,s) \\ \Lambda(k+1,s) \end{pmatrix} = \begin{pmatrix} 0 & I_s^T \\ I_s & \Omega(k+1,s)+\gamma^{-1}I_{s,s} \end{pmatrix}^{-1} \times \begin{pmatrix} 0 \\ Y(k+1,s) \end{pmatrix} = P^{SV}(k+1,s)\begin{pmatrix} 0 \\ Y(k+1,s) \end{pmatrix}, \qquad (9)$$

де $\Lambda(k+1,s) = \left(\lambda(k-s+2,s),\ \lambda(k-s+3,s),\ ...,\ \lambda(k+1,s)\right)^T$,

$Y(k+1,s) = \left(y(k-s+2),\ y(k-s+3),\ ...,\ y(k+1)\right)^T$.

Нескладно бачити, що співвідношення (8), (9) за своєю сутністю є on-line алгоритмом навчання нейронної мережі фіксованої архітектури.

Висновки

Запропоновано адаптивний метод навчання нейронної мережі опорних векторів найменших квадратів (LS-SVM) незмінної архітектури. Особливістю цього методу є те, що мінімізація критерію емпіричного ризику відбувається на «ковзному вікні» фіксованої розмірності, що суттєво спрощує чисельну реалізацію процедури та дозволяє обробити інформацію, що генерується нестаціонарними об'єктами.

Список використаної літератури

1. Kasabov N. Evolving Connectionist Systems. / Kasabov N. –London: Springer – Verlag, 2003 – 307 p.
2. Haykin S. Neural Networks. A Comprehensive Foundation. / Haykin S. – Upper Saddle River, N.J.: Prentice Hall, Inc., 1999. – 842 p.
3. Suykens J.A.K. Least Squares Support Vector Machines. / Suykens J.A.K., Gestel T.V., Brabanter J.D., Moor B.D., Vandewalle J. – Singapore: World Scientific, 2002. - 294 p.
4. Specht D. F. A general regression neural network /Specht D. F. // IEEE Trans. on Neural Networks – 1991. – 2. – P. 568-576.
5. Гантмахер Ф.Р. Теория матриц. / Гантмахер Ф.Р. – Москва: Наука. – 1988 – 552 с.
6. Бодянський, Є.В. Еволюційна нейронна мережа з ядерними функціями активації та адаптивний алгоритм її навчання / Є.В. Бодянський, А.О. Дейнеко, Н.О. Тесленко // Наукові праці: науково-методичний журнал. – Миколаїв: Вид-во ЧДУ ім. Петра Могили – 2011. – Вип. 130. – Т. 143. – С. 71-78.

Бодянский Е.В., Дейнеко А.А., Дейнеко Ж.В., Шаламов М.А. Адаптивное обучение нейронной сети опорных векторов наименьших квадратов. В работе предложен адаптивный метод обучения нейронной сети опорных векторов наименьших квадратов (LS-SVM) неизменной архитектуры. Особенностью этого метода является то, что минимизация критерия эмпирического риска происходит на «скользящем окне» фиксированной размерности, что существенно упрощает численную реализацию процедуры и позволяет обработать информацию, генерируемую нестационарными объектами.

Ключевые слова: нейронные сети, нейронная сеть опорных векторов наименьших квадратов, критерий эмпирического риска, «скользящее окно».

Yevgeniy V. Bodyanskiy, Anastasiia O. Deineko, Zhanna V. Deineko, Maksym O. Shalamov. **Adaptive learning of least-squares support vector machine.** In this paper an adaptive learning method of least-squares support vector machine (LS-SVM) with unchanging architecture is proposed. Feature of this method is that the minimization criterion of empirical risk occurs at the "sliding window" of fixed dimension essentially simplifies numerical implementation of the procedure and allows to process information generated by non-stationary objects.

Key words: neural networks, least-squares support vector machine, empirical risk criterion, "sliding window".

Рецензент д.т.н., професор Вінокурова О.А. (ХНУРЕ)

Поступила 23.03.2015г.