УДК 519

© 2013 г. **О.С. Амосов**, д-р техн. наук,
С.Г. Баена
(Комсомольский-на-Амуре государственный технический университет)

БАЙЕСОВСКОЕ ОЦЕНИВАНИЕ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ С РАДИАЛЬНЫМИ БАЗИСНЫМИ ФУНКЦИЯМИ

Дано решение задачи байесовского оптимального нелинейного оценивания случайных векторов с использованием нейронной сети с радиальными базисными функциями. Показано, что эта нейронная сеть при соответствующем выборе критерия, используемого для ее обучения, обеспечивает получение оценок, близких по своим свойствам к оценкам, получаемым с помощью байесовского алгоритма. Рассматривается иллюстрирующий пример.

Ключевые слова: оптимальное нелинейное оценивание, байесовский подход, нейронная сеть с радиальными базисными функциями, среда MatLab.

Введение

В последнее время для решения сложных в вычислительном отношении задач нелинейного оптимального оценивания исследуется возможность применения искусственных нейронных сетей (НС) [1 – 7]. Из всех архитектур НС наибольшее внимание было уделено применению хорошо изученных многослойных сетей прямого распространения [3 – 8]. Обращено внимание на то, что наиболее важным, но и наименее проработанным вопросом в процедуре решения задачи оценивания с использованием НС представляется вопрос выбора типа НС, обеспечивающего точность, близкую к точности оптимального в классическом смысле байесовского алгоритма [3 – 5]. При этом без должного внимания остались нейронные сети с радиальными базисными функциями (РБФ), хотя они имеют ряд очевидных преимуществ перед указанными выше сетями [7, 8]: имеют всего один промежуточный слой, избавляя разработчика от решения вопроса о числе слоев; весовые коэффициенты линейного выходного слоя сети оптимизируются с помощью методов линейной оптимизации, что сокращает время обучения.

Поэтому представляется целесообразным оценить возможность и эффективность применения нейронных сетей с РБФ для решения задач оптимального оценивания и фильтрации, что и сделано в предлагаемой статье.

Постановка задачи байесовского оценивания

Необходимо оценить n-мерный вектор $\mathbf{x} = [x_1 \quad ... \quad x_n]^T$ по m-мерным измерениям $\mathbf{y} = [y_1 \quad ... \quad y_m]^T$, которые могут быть записаны, аналогично [3–5, 9], следующим образом:

$$\mathbf{y} = \mathbf{s}(\mathbf{x}) + \mathbf{v}, \qquad (1)$$

где $\mathbf{s}(\mathbf{x}) = [s_1(\mathbf{x}) \ \ \ldots \ \ s_m(\mathbf{x})]^T$ – m-мерная в общем случае нелинейная вектор-функция векторного аргумента, которая обычно считается известной; $\mathbf{v} = [v_1 \ \ \ldots \ \ v_m]^T$ – случайный вектор, передающий наличие ошибок измерения.

Отличительная особенность байесовской постановки задачи оценивания заключается в том, что априорная информация об оцениваемом векторе $\mathbf{x}$ и используемых измерениях $\mathbf{y}$ задается либо в виде их совместной функции плотности распределения вероятностей (ФПРВ) $f(\mathbf{x},\mathbf{y})$, что характерно для традиционной постановки [9], либо в виде согласованного набора реализаций

$$\{(\mathbf{y}^{(j)}, \mathbf{x}^{(j)})\}_{j=1}^{N}, \qquad (2)$$

что характерно для постановки задачи оценивания с использованием настраиваемых, обучаемых систем, – например, нейронных сетей [3–5]. Согласованность здесь понимается в том смысле, что они представляют независимые между собой реализации составного случайного вектора с одной и той же функцией плотности распределения вероятностей $f(\mathbf{x},\mathbf{y})$.

Решение задачи байесовского оценивания

Для *традиционной постановки* требуется [3 – 5, 9], располагая значением вектора $\mathbf{y}$ (1), найти оценку $\tilde{\mathbf{x}}(\mathbf{y})$, минимизирующую критерий

$$J = M[(\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{y}))^T(\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{y}))] = M\|(\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{y}))\|^2 = \int\int\|(\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{y}))\|^2 f(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}, \quad (3)$$

где M – знак математического ожидания, соответствующий плотности $f(x,y)$. Здесь и далее интегралы понимаются как многократные, с бесконечными пределами. Известно, что оптимальная в среднеквадратическом смысле оценка, минимизирующая критерий (3), определяется как [9]

$$\tilde{\mathbf{x}}(\mathbf{y}) = \int \mathbf{x} f(\mathbf{x}/\mathbf{y})d\mathbf{x}, \qquad (4)$$

где $f(\mathbf{x}/\mathbf{y}) = f(\mathbf{x},\mathbf{y})/f(\mathbf{y})$ – условная (апостериорная) функция плотности распределения вероятности вектора $\mathbf{x}$, $f(\mathbf{y}) = \int f(\mathbf{x},\mathbf{y})d\mathbf{x}$.

Для *обучаемых систем* требуется, располагая значением вектора $\mathbf{y}$ и обучающей выборкой $\{(\mathbf{y}^{(j)},\mathbf{x}^{(j)})\}_{j=1}^{N}$, найти оценку $\tilde{\mathbf{x}}(\mathbf{y})$, минимизирующую критерий вида

$$\tilde{J} = \frac{1}{N}\sum_{j=1}^{N}\left\|\mathbf{x}^{(j)} - \tilde{\mathbf{x}}(\mathbf{y}^{(j)})\right\|^2 .$$

Приближенное решение этой задачи может быть найдено, если ввести класс зависящих от параметров функций, используемых для вычисления оценки $\tilde{\mathbf{x}}(\mathbf{y},\tilde{\mathbf{W}})$

[3 – 5] и найти значения параметров $\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^*$, обеспечивающих минимум функции эмпирического риска $\widetilde{J}^*(\widetilde{\mathbf{W}})$ [3–5, 10]:

$$\widetilde{J}^*(\widetilde{\mathbf{W}}) = \frac{1}{N}\sum_{j=1}^{N}\left\|\mathbf{x}^{(j)} - \widetilde{\mathbf{x}}(\mathbf{y}^{(j)}, \widetilde{\mathbf{W}})\right\|^2, \qquad (5)$$

где $\widetilde{\mathbf{W}}$ – вектор или матрица, определяющая набор параметров, конкретизирующих функцию $\widetilde{\mathbf{x}}(\mathbf{y}, \widetilde{\mathbf{W}})$ в выбранном классе.

Таким образом, проблема получения алгоритма оценивания сводится к процедуре нахождения параметров $\widetilde{\mathbf{W}}$, отыскиваемых в результате минимизации критерия (5), формируемого с использованием данных обучающей выборки (2).

Такой подход в полной мере соответствует решению задач с помощью обучаемых систем, настраиваемых в режиме обучения «с учителем». К обучаемым системам можно отнести искусственные нейронные сети, нечеткие системы, системы, построенные на использовании вейвлет-преобразований (вейвлет-системы) и их объединение.

В случае нейросетевого подхода задачу оценивания вне зависимости от топологии НС можно рассматривать как задачу нахождения неизвестного и реализуемого с использованием НС отображения:

$$\widetilde{\mathbf{x}}(\mathbf{y}, \widetilde{\mathbf{W}}) \equiv \widetilde{\mathbf{x}}^{NN}(\mathbf{y}, \mathbf{W}) = \mathbf{K}^{NN}(\mathbf{y}, \widetilde{\mathbf{W}}),$$

где $\mathbf{K}^{NN}(\mathbf{y}, \widetilde{\mathbf{W}})$ – специальная, в общем случае нелинейная функция; $\widetilde{\mathbf{W}}$ – определяет массив смещений и весовых коэффициентов НС.

Режимы работы нейронной сети

Для байесовского оценивания с помощью НС существуют два основных режима работы [4, 5]. Первый из них – это режим обучения с учителем, второй – штатный режим оценивания в реальном времени.

В *режиме обучени,* с использованием обучающего множества (2) отыскивается зависимость вида $\widetilde{\mathbf{x}}^{NN(j)}(\mathbf{y}^{(j)}, \widetilde{\mathbf{W}})$ в соответствии с заданным критерием (5), т.е. отыскиваются оптимальные параметры $\widetilde{\mathbf{W}}^*$ нейронной сети.

В штатном режиме с использованием найденных на предыдущем режиме параметров $\widetilde{\mathbf{W}}^*$ отыскивается оценка вектора $\mathbf{x}$ по вектору измерений $\mathbf{y}$.

В данной статье рассмотрим использование нейронных сетей с радиальными базисными элементами.

Нейронная сеть с радиальными базисными элементами

Нейронная сеть с радиальными базисными функциями, сеть РБФ – Radial Basis Function Network (RBFN) – представляет собой двухслойную сеть без обратных связей, которая содержит единственный скрытый слой радиально симметричных шаблонных нейронов – шаблонный слой [7, 8].

Выходной сигнал шаблонного нейрона – это функция только от расстояния между входным n-вектором $\mathbf{u}$ и сохраненным центром $\mathbf{c}$:

$$h(\mathbf{u}) = \varphi\left(\frac{\|\mathbf{u}-\mathbf{c}\|}{\sigma}\right).$$

Обычно вектор **c** сохраняется в пространстве весов от входного слоя к слою шаблонов, а функция активации нейронов скрытого слоя, задающая способ измерения расстояния, $\varphi(s) = e^{-s^2}$ – функция Гаусса.

Выходной слой сети является линейным:

$$z_j = \sum_{k=1}^{m_h} w_{kj}\varphi\left(\frac{\|\mathbf{u}-\mathbf{c}_k\|}{\sigma_k}\right),\ j=\overline{1.m}, \tag{6}$$

где $\mathbf{c}_k$ – центры, количество которых равно m_h; σ_k – отклонения радиальных элементов.

Обучение сети РБФ происходит следующим образом. Сначала определяются центры и отклонения для радиальных элементов; после этого оптимизируются параметры w_{kj}, $j=\overline{1.m}$ линейного выходного слоя.

Пример решения задачи оценивания

В качестве примера компьютерного моделирования целесообразно остановить выбор на задаче оптимального нелинейного оценивания, которая была решена в работе [4] с использованием нелинейной нейронной сети прямого распространения – Feed Forward Neural Network (FFN), чтобы провести сопоставление последней и НС с РБФ.

Постановка задачи. Необходимо оценить равномерно распределенную на интервале $[0, b]$ случайную величину x по зашумленным измерениям вида

$$y_i = x + v_i,\quad i=\overline{1.l}, \tag{7}$$

в которых ошибки измерений v_i, $i=\overline{1.l}$ представляют собой независимые друг от друга и от x центрированные случайные величины, равномерно распределенные в интервале $[-a/2, a/2]$. В этом примере $\mathbf{x} \equiv x$, $\mathbf{y} \equiv [y_1 \quad \dots \quad y_i]^T$, $\mathbf{v} = [v_1 \quad \dots \quad v_i]^T$. Необходимо отметить, что апостериорная ф.п.р.в. $f(\mathbf{x}/\mathbf{y})$ здесь не является гауссовской, так как x и v_i, $i=\overline{1.l}$ – равномерно распределенные случайные переменные.

Оптимальное решение задачи подробно описано в работах [4], здесь же приведем для удобства читателя русский перевод.

<u>Оптимальная линейная оценка</u> $x^*(\mathbf{y})$ и соответствующая ей ковариация ошибок оценивания P_e^*:

$$x^*(\mathbf{y}) = \bar{x} + \mathbf{P}_{x\mathbf{y}}\mathbf{P}_{\mathbf{yy}}^{-1}[\mathbf{y}-\bar{\mathbf{y}}],\ P_e^* = P_{xx} - \mathbf{P}_{x\mathbf{y}}\mathbf{P}_{\mathbf{yy}}^{-1}\mathbf{P}_{\mathbf{y}x}, \tag{8}$$

где $\bar{x} = \dfrac{b}{2}$; $\bar{\mathbf{y}} = \dfrac{1}{2}[b \quad \dots \quad b]^T$; $P_{xx} = \sigma_x^2 = b^2/12$; $\mathbf{P}_{x\mathbf{y}} = \sigma_x^2\mathbf{H}^{\mathrm{T}}$; $\mathbf{P}_{\mathbf{yy}} = \sigma_x^2\mathbf{I}_i + r^2\mathbf{E}_i$.

Здесь $\mathbf{H} = [1 \quad \dots \quad 1]^T$, $\mathbf{I}_i$ – квадратная матрица, составленная из 1; $\mathbf{E}_i$ – единичная матрица, $\sigma_x^2 = b^2/12$, $r^2 = a^2/12$.

Оптимальная нелинейная оценка (4) для данного примера может быть получена в явном виде [4]. Для того, чтобы объяснить это, введем область Ω, представляющую собой отрезок, формируемый в результате пересечения всех интервалов $[y_i - a/2, y_i + a/2], i = \overline{1.l}$, т.е.:

$$\Omega \equiv [d_1, d_2] = \bigcap_{i=1}^{l} [y_i - a/2, y_i + a/2].$$

Можно показать, что апостериорная плотность в рассматриваемом примере является равномерной на отрезке $[c_1, c_2]$, представляющим собой пересечение априорной области $[0, b]$ и области Ω, так что $c_1 = \max\{0, d_1\}$, $c_2 = \min\{b, d_2\}$. Тогда из этого следует, что

$$\tilde{x}(\mathbf{y}) = (c_2 + c_1)/2. \qquad (9)$$

Решение задачи с использованием НС. Будем считать, что при построении алгоритма в качестве априорной информации выступает набор пар (2) $\{(\mathbf{y}^{(j)}, x^{(j)})\}_{j=1}^{N}$. Тогда для решения задачи оценивания будем использовать НС с РБФ, в частности двухслойную сеть без обратных связей, которая содержит скрытый слой радиально симметричных шаблонных нейронов. В соответствии с (6)

$$\tilde{x}^{RBFN}(\mathbf{y}) = \sum_{k=1}^{m_h} w_{k1} \varphi\left(\frac{\|\mathbf{y} - \mathbf{c}_k\|}{\sigma_k}\right),$$

число которых m_h = 500 подобрано опытным путем. Выходной слой с одним нейроном (по числу оцениваемых переменных). Параметры y_k=1, $k = \overline{1.500}$.

Кроме того, приведем полученное в [4] решение задачи оценивания с использованием как линейной однослойной НС LinN с одним нейроном, с тождественной функцией активации $\psi(s) = s$ и i входами, так и нелинейной НС FFN – двухслойной НС с последовательными связями с i входами, с q нейронами в скрытом слое и одним нейроном в выходном слое.

Выход НС FFN может быть определен как [4]

$$\tilde{x}^{FFN}(\mathbf{y}) = \psi\left(\sum_{\mu=1}^{q}\left(w_{\mu 1}^{2} \varphi\left(\sum_{k=1}^{i}(w_{k\mu}^{1} y_k) + w_{0\mu}^{1}\right)\right) + w_{01}^{2}\right),$$

где $w_{0\mu}^{1}$, $w_{k\mu}^{1}$, $\mu = \overline{1.q}$, $k = \overline{1.i}$ – смещения и веса нейронов скрытого слоя; $w_{01}^{2}, w_{\mu 1}^{2}$, $\mu = \overline{1.q}$ – смещение и веса нейрона выходного слоя FFN; q – число нейронов скрытого слоя; $\varphi(s) = th\, s = \dfrac{e^{s} - e^{-s}}{e^{s} + e^{-s}}$ – активационная функция для нейронов скрытого слоя; $\psi(s) = s$ – активационная характеристика нейрона выходного слоя.

Среда моделирования. При проектировании НС с РБФ были использованы функции newrbe и newrb пакета Neural Networks среды Matlab R2011b (win64), автоматизирующие синтез структуры нейронной сети и ее обучение.

Функция newrbe реализует НС с РБФ с нулевой ошибкой на обучающей выборке, когда число шаблонных нейронов в точности равно числу примеров из обучающей выборки. Использование данной функции не дало удовлетворительного результата по точности.

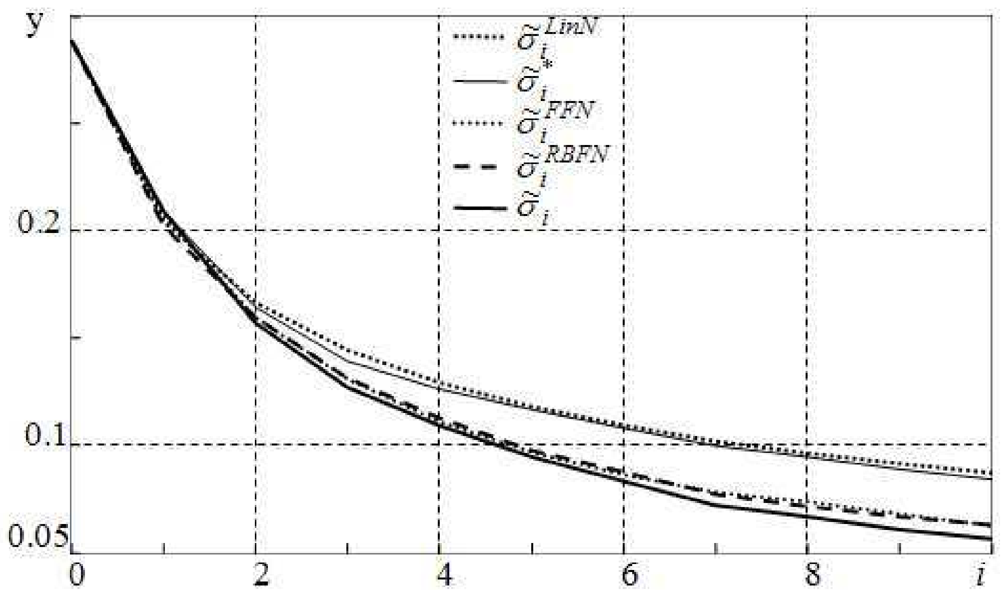
Функция newrb реализует обобщенную НС с РБФ, когда число шаблонных нейронов можно оптимизировать, при этом их количество будет значительно меньше числа примеров из обучающей выборки [7].

Для моделирования НС FFN была использована функция newff с использованием обучения по методу обратного распространения ошибки. Для моделирования линейной НС LinN использована функция newlin.

Ниже приведены полученные путем моделирования результаты, соответствующие линейному и нелинейному оптимальным и нейросетевым алгоритмам, для разного числа измерений i. При проведении моделирования принималось: $b=1$, $r^2=1$, $i=\overline{1.l}$, $l=10$. Вычисление оптимальной нелинейной оценки осуществлялось в соответствии с (9).

На рисунке представлены среднеквадратичные отклонения (СКО) ошибок оценивания: расчетное СКО $\tilde{\sigma}_i^* \approx \sqrt{P_e^*}$ – для линейных оптимальных оценок (8); выборочные СКО ошибок $\tilde{\sigma}_i$ – для нелинейных оптимальных оценок и $\tilde{\sigma}_i^{\eta}$, $\eta = LinN, FFN, RBFN$ – для нейросетевых оценок. Выборочные СКО были рассчитаны следующим образом:

$$\tilde{\sigma}_i \approx \sqrt{\frac{1}{N}\sum_{j=1}^{N}(e_i^{(j)})^2},\ e_i^{(j)} = x^{(j)} - \tilde{x}^{(j)}(\mathbf{y}^{(j)}),\ \tilde{\sigma}_i^{\eta} \approx \sqrt{\frac{1}{N}\sum_{j=1}^{N}(e_i^{\eta\,(j)})^2},\ e_i^{\eta\,(j)} = x^{(j)} - \tilde{x}^{\eta\,(j)}(\mathbf{y}^{(j)},\tilde{\mathbf{W}}).$$



Для обучения нейронных сетей в соответствии с (7) моделировались реализации $\{(\mathbf{y}^{(j)}, x^{(j)})\}_{j=1}^{N}$. При этом для получения приемлемых результатов по точности для линейной LinN и нелинейной НС FFN N было выбрано равным 3000, а для НС RBFN число реализаций N, полученное эмпирическим путем, оказалось порядка 20000.

После обучения осуществлялась проверка. С этой целью дополнительно моделировалось еще $M=3000$ пар реализаций $\mathbf{y}^{(j)}, x^{(j)}$ для разных $i=\overline{1.l}$, $l=10$.

Как видно из рисунка, оценка обученной линейной НС и оптимальная линейная оценка совпадают, но при этом значительно отличаются от оптимальной нелинейной оценки. Оценки же обученных нелинейных НС FFN и НС RBFN совпадают между собой и близки к оптимальной нелинейной оценке.

Моделирование проводилось на компьютере Intel (R) Xeon (R) CPU X5680 3.33GHz (4 процессора), 20.0 ГБ ОЗУ, 64-разрядная операционная система, предоставленном факультетом компьютерных технологий ФГБОУ ВПО «Комсомольский-на-Амуре государственный технический университет».

Заключение

Точность оценивания с помощью нейронной сети с радиальными базисными элементами близка к точности оценивания с помощью нейронной сети прямого распространения и мало отличается от предельно достижимой точности оптимального нелинейного алгоритма. Применение сети РБФ с одним скрытым слоем снимает остроту вопроса проектирования структуры нейронной сети, что имеет место при использовании НС прямого распространения или рекуррентных НС. Однако сети с РБФ присущи недостатки, связанные с ее излишней громоздкостью. Представляется перспективным дальнейшее исследование эффективности применение нейронных сетей как с РБФ, так и НС прямого распространения для многомерных задач оптимального нелинейного оценивания.

ЛИТЕРАТУРА

1. *Lo J. T. H.* Synthetic approach to optimal filtering// IEEE Trans. Neural Networks. – 1994. – Vol. 5, N 5. – P.803–811.
2. *Haykin S. and P. Yee*. Optimum nonlinear filtering // IEEE Trans. On Signal Processing. – 1997. – Vol. 45(11). – P.2774–2786.
3. *Степанов О.А., Амосов О.С.* Байесовское оценивание с использованием нейронной сети // Авиакосмическое приборостроение. – 2004. – № 6. – С.46–55.
4. *Stepanov O.A., Amosov O.S.* Optimal estimation by using neural networks // Proc. of the16-th World Congress IFAC. – Prague, Czech Republic, 2005.
5. *Stepanov O.A., Amosov O.S.* The Comparison of the Monte-Carlo Method and Neural Networks Algorithms in Nonlinear Estimation Problems // Proc. of the 9th IFAC WORKSHOPS «Adaptation and Learning in Control and Signal Processing», Saint Peterburg, Russia, 2007.
6. *Amosov O.S.* Optimal and Adaptive Estimation Using On-Line Training Neural Networks // Proc. of the 2nd International Conference on Intelligent Control and Information Processing ICICIP. – Harbin, China, 2011. – P. 208–213.
7. *Хайкин С.* Нейронные сети: Полный курс. – Изд. 2-е. Пер. с англ. – М.: Изд. дом «Вильямс», 2006.
8. *Круглов В. В., Дли М. И., Голунов Р. Ю.* Нечеткая логика и искусственные нейронные сети.– М.: Изд-во физ.-мат. лит., 2001.
9. *Степанов О.А.* Применение теории нелинейной фильтрации в задачах обработки навигационной информации. – СПб.: ГНЦ РФ ЦНИИ «Электроприбор», 1998.
10. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов. – М.: Наука, 1974.

Статья представлена к публикации членом редколлегии Е.Л. Ереминым.

E-mail:
Амосов Олег Семенович – osa18@yandex.ru;
Баена Светлана Геннадьевна – svetlana.baena@yandex.ru.