Навстречу
75-летнему
юбилею ТулГУ

ИЗВЕСТИЯ

ТУЛЬСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА

Серия

Вычислительная техника.
Информационные технологии.
Системы управления

Выпуск 4

2005

МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННЫХ И УПРАВЛЯЮЩИХ СИСТЕМ

УДК 621

Е.О. Аверченков, Г.Ф. Филаретов (Москва, МИКМ)

ПОСТРОЕНИЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ПРЕДСКАЗАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

Рассматриваются вопросы построения моделей дрейфа с использованием искусственных нейронных сетей. Предлагается новый метод построения нейросетевых моделей дрейфа для решения задач экстраполяции. На его основе реализуется идея построения многоальтернативных моделей предсказания.

Решается задача построения моделей временных рядов, содержащих неизвестную детерминированную компоненту - аддитивный дрейф:

$$y_t = x_t + f(t), \qquad (1)$$

где y^t - наблюдаемые значения временного ряда, x^t - случайная компонента с постоянным математическим ожиданием, f(t) - дрейф или тренд, относительно которого предполагается, что он представляет собой некоторую детерминированную (и априори неизвестную) функцию дискретного времени.

Если предположить, что имеет место весьма распространенная на практике задача определения модели тренда, справедливой для некоторого временного отрезка (на множестве наблюдений, содержащем N отчетов), то в этом случае по сути решается задача оценивания функции *f{t)* на фоне случайной помехи x^t. Для ее решения чаще всего используются различные статистические процедуры [1], а в последнее время и алгоритмы, основанные на применении искусственных нейронных сетей (ИНС). В нейросетевом варианте, как правило, фигурируют классические многослойные персептроны с сигмоидальными функциями активации, обучаемые при помощи стандартных градиентных методов обучения типа *backpropagation, qiuckpropagation* [2]. Все эти методы обучения минимизируют дисперсию ошибки предсказания значений обучающей выборки на один шаг вперед. По окончании процесса обучения, когда дисперсия остатков предсказания минимальна (для выбранной нейросетевой структуры и обучающей выборки длины *N),* преобразование *вход-выход,* осуществляемое ИНС, может служить некоторой аппроксимацией функции дрейфа *fit),* а остатки предсказания будут являться оценками значений стохастической компоненты x^t.

Получаемые с помощью указанного нейросетевого алгоритма модели дрейфа достаточно хорошо решают, как правило, только задачи интерполяционного типа. Если же попытаться использовать их для предсказания будущих значений тренда, т. е. для задач экстраполяции, то в этом случае они оказываются мало пригодными. Это во многом связано с тем, что сигмоидальные функции активации естественным образом ограничивают возможный диапазон изменения выходного сигнала ИНС и, кроме того, вносят дополнительные нелинейные искажения в прогнозируемые значения.

Для улучшения экстраполяционных свойств конструируемых с помощью ИНС моделей рекомендуется применять в качестве выходного элемента нейронной сети элемент с линейной активационной функцией, которому не присущ эффект насыщения. Однако это усовершенствование лишь немного улучшает экстраполяционные свойства нейросетевых моделей тренда. Далее предлагается иной подход к решению данной задачи.

Предлагаемый подход основан на предположении о том, что исследователю известна некоторая дополнительная априорная информация о свойствах трендовой составляющей процесса, имеющая самый общий характер и выражающаяся в терминах, описывающих аналитические свойства функции *fit),* таких как *fit)* - функция "возрастающая", "убывающая", "выпуклая", "вогнутая", «с точкой перегиба», «с наличием экстремума» и т.п. [3]. Считается также, что аналитические свойства функции $f(t)$, имеющие место на интервале, используемом для обучения ИНС, сохраняют свою силу и для интервала экстраполяции. Очевидно, что вряд ли имеет смысл задаваться слишком сложным видом функции $f(t)$. Поэтому в дальнейшем будут использоваться четыре основные модели:

- модель 1: $f(t)$ - монотонно возрастающая (убывающая) функция времени;

- модель 2: $f(t)$- выпуклая (вогнутая) функция времени;

- модель 3: $f(t)$ - функция времени с наличием точки перегиба;

- модель 4: $f(t)$ - функция с наличием экстремума.

Если такого рода априорная информация имеется и можно с той или иной степенью уверенности остановиться на одной из указанных моделей $f(t)$, то тогда эту информацию следует учесть при построении экстраполирующей ИНС с помощью соответствующим образом выбранного критерия обучения.

Рассмотрим основные моменты предлагаемого метода построения экстраполирующей ИНС. Пусть y^t, $t = 1, N$ - элементы обучающей выборки, полученные экспериментально. Введём **дополнительные пробные точки,** которые вместе с элементами обучающей выборки будут использоваться в процедуре обучения для контроля качества получаемой аппроксимирующей зависимости. Часть из них в количестве M^1 разместим равномерно в промежутках между точками обучающей выборки в количестве L точек, т. е. их общее количество M^1 в этом случае равно, очевидно $M^1 = (N-1)L$. Вторую часть в количестве M^2 разместим равномерно на интервале экстраполяции. Тогда суммарное количество проверочных точек, с помощью которых можно оценить качество обучения ИНС, равно $N + M$, где $M = M^1 + M^2$.

Сформулируем новый критерий обучения ИНС в виде

$$\sum_{t=1}^{N}[y_t - z(t)]^2 + \eta d = \min_{\overline{\mathbf{w}}}, \qquad (2)$$

где *z(t)* **-** функция, реализуемая нейронной сетью и аппроксимирующая искомую трендовую составляющую; **W-** матрица весовых коэффициентов ИНС, *г/* - некоторая константа (77 > 0) ; ***d*** - число нарушений условия, определяемого из априорной информации, на множестве всех ***(N + M)*** проверочных точек.

В таблице для разных моделей тренда указаны способы определения значения ***d*** **.**

Способы определения значения ***d***

Модель	Вид функции $f(t)$	d
Модель 1	Монотонно возрастающая	d равно числу нарушений условия: $z(i+1) > z(i), \quad i=\overline{1, N+M-1}$
	Монотонно убывающая	d равно числу нарушений условия: $z(i+1) < z(i), \quad i=\overline{1, N+M-1}$
Модель 2	Выпуклая	d равно числу нарушений условия: $\nabla(i+1) > \nabla(i)$, $i=\overline{1, N+M-2}$
	Вогнутая	d равно числу нарушений условия: $\nabla(i+1) < \nabla(i)$, $i=\overline{1, N+M-2}$
Модель 3	С точкой перегиба	$d = d^*$, где d^* равно числу нарушений условия: $\lvert\nabla(i) \quad \nabla(i \quad 1)\rvert \bullet \lvert\nabla(i+1) \quad \nabla(i)\rvert > 0$, $i = \overline{2, N+M-3}$
Модель 4	С точкой экстремума	$d = d^*$, где d^* равно числу нарушений условия: $\nabla(i+1) \bullet \nabla(i) > 0$, $i=\overline{1, N+M \quad 2}$
Примечание. В таблице использовано обозначение для приращений функции $z(t): \nabla(i) = z(i+1) \quad z(i)$,		

Применение критерия (2) делает невозможным использование при обучении ИНС традиционных градиентных методов. Наилучшим выходом из этой ситуации является переход к методам случайного поиска, что, однако, может существенно замедлить процесс обучения. Поэтому предлагается обучение нейросетевой модели проводить в два этапа. На первом этапе производится предварительное формирование модели при помощи традиционных методов обучения. Целью данного этапа является быстрое получение "прототипа" искомой модели без ярко выраженного эффекта"переобучения", обладающей достаточно малой ошибкой предсказания в области интерполяции, но, возможно, не удовлетворяющей наложенным на искомую модель ограничениям. На втором этапе производится дообучение нейросетевой модели с использованием критерия (2) и метода случайного поиска.

Для практического использования критерия (2) необходимо разумно выбрать коэффициент *г*. Очевидно, что в ходе процесса дообучения при помощи используемого критерия оба слагаемых, входящих в (2), должны вносить сравнимый вклад в значение критерия с тем, чтобы процедура обучения минимизировала ошибку предсказания данных и одновременно формировала модель с заданными аналитическими свойствами. Исходя из этого, начальное значение коэффициента ***г]*** предлагается выбирать как отношение величины суммарной квадратичной ошибки предсказания обучающей выборки в начале этапа дообучения к найденному по завершении первого этапа значению $d > 1$. Таким образом, первоначально коэффициент ***rj*** рассчитывается по формуле

$$\eta = \frac{1}{d}\sum_{t=1}^{N}[y_t - z(t)]^2 .$$

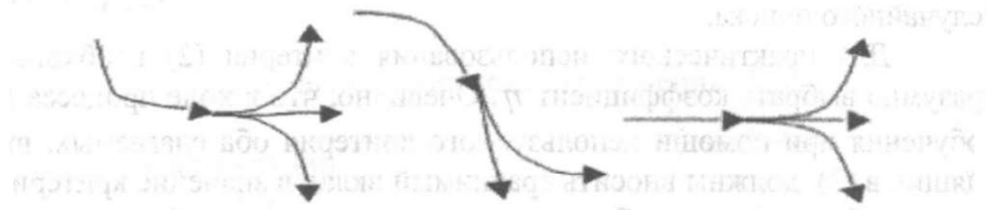
В ходе дообучения для повышения скорости сходимости процесса обучения возможна корректировка *г*.

Критерий (2) можно несколько видоизменить, если считать, что аналитические свойства функции *f{t)* для интервала интерполяции и интервала экстраполяции различны. Тогда он примет следующий вид:

$$\sum_{i=1}^{N}\left(y_t - z(t)\right)^2 + \eta_1 d_1 + \eta_2 d_2 \xrightarrow[\vec{W}]{} \min , \qquad (3)$$

где d^1 и d^2 - число нарушений заданных аналитических свойств функции *f(t)* для первого и второго интервала соответственно.

Поскольку априорная информация о свойствах трендовой составляющей зачастую носит чисто предположительный характер, предлагается строить не одну, а сразу несколько моделей предсказания («семейство моделей») Можно интерпретировать различные модели такого семейства, как оптимистический, пессимистический и прагматический варианты развития событий. На рисунке приведены возможные варианты моделей такого семейства.



Семейство моделей, отражающих возможное поведение функции дрейфа

По сути такое семейство может отражать различные сценарии развития тенденции, имеющей место в значениях наблюдаемого процесса: продолжающийся рост или снижение, изменение скорости изменения тренда (стремление к застою, к насыщению), достижение экстремума и изменение тенденции и т.п.

Библиографический список

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. - М.: ЮНИТИ, 1998. - 1022 с.

2. Уоссерман Ф. Нейрокомпьютерная техника: Теория и практика./ Пер. с англ. - М.: Мир, 1992 . - 240 с.

3. Filaretov G.F., Averchenkov E.O. Time series forecasting using neural networks, 45h International Scientific Colloquium, Ilmenau Technical University, 2000.