

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ И СОБЛЮДЕНИЕ ПРИНЦИПА СОГЛАСОВАННОСТИ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

А.С.Дулесов

Представлены теоретические аспекты возможности использования нейронных сетей в задачах прогнозирования. Дано понятие принципа согласованности результатов и модель его соблюдения.

The theoretical aspects of the opportunity of the neuron network usage in the problems of prognostication are represented. The notion of the principal of the observation results coordination and model of its observance are given.

ВВЕДЕНИЕ

В практике прогнозирования одной из основных задач ставилась возможность на основе имеющейся информации построить (оценить) точно или приближенно зависимость влияния как внешних, так и внутренних факторов на характеристики объекта. Процесс такой оценки является моделированием и в большинстве случаев опирается на наличие статистических данных и теорию прогнозирования.

1. ОСНОВНЫЕ ПОНЯТИЯ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ

Под объектом исследования, в каком-то смысле однородном (например, предприятие, технологический процесс, отрасль и т.д.), будем понимать реально существующий объект, который в некоторый фиксированные моменты времени может занимать определенное состояние. Поведение объекта описывается через свойства (например, "мощность", "скорость", "стабильность" и др.), отражаемые через характеристики.

Конечной целью наблюдения за объектом можно считать получение информации относительно некоторого конечного набора X_1, X_2, \dots, X_n его характеристик. Зависимость между характеристиками Y и X объекта моделируется обычно в виде функции $Y = f(X)$. В данном равенстве функция f , также может рассматриваться как характеристика, но уже не объекта, а совокупности объектов. Тогда под f следует понимать некоторый класс функций, объединенных какими-либо свойствами (свойством линейности или свойством принадлежности к модели одного и того же объекта). По сути, мы имеем дело с задачей возможно более точной оценки функции f по данным наблюдений (обычно неполным или неточным) отдельных объектов. В конечном итоге речь идет о построении "истинной" зависимости между показателями (числовыми характеристиками) объекта и рассчитанной зависимостью,

найденной в ходе решения поставленной задачи.

В данной работе представлены возможности моделирования зависимости $Y = f(X)$ с использованием нейронной сети прямого распространения, а также дана оценка эффективности его применения.

О методах построения зависимости $Y = f(X)$ можно прочесть в любом учебнике по статистике. Основным инструментом здесь принято считать метод наименьших квадратов. Однако дела обстоят не так просто, как может показаться на первый взгляд. Так в [1] обобщен опыт в сфере разработки и совершенствования моделей построения статистических зависимостей и изложен принцип максимальной согласованности. Учитывая всю сложность задачи, обусловленную наличием различного рода ошибок, возникает вопрос: способна ли нейронная сеть дать согласованную оценку, как наилучшую из приемлемых.

Наибольшее распространение в практике построения зависимостей получили методы, позволяющие строить зависимости вида:

$$Y = f(X_1, \dots, X_n, \xi).$$

Здесь характеристика объекта Y зависит от характеристик (факторов) X_1, \dots, X_n с некоторой долей уверенности известные исследователю и "прочих факторов" $X_{n+1}, \dots, X_{n+2}, \dots$, агрегируемых в один неопределенный фактор ("оценку наблюдения"). Искомые зависимости относятся (чаще всего) к классу параметрических, т.е. имеют вид известной функции $Y = f(X, a)$ от объясняющей (векторной) переменной X и конечномерного вектора неизвестных параметров a .

Механизм построения зависимости $Y = f(X, a)$ с помощью аналитических методов достаточно хорошо проработан. Если вести речь об аналогичной задаче, но в качестве инструмента использовать нейронные сети, то, прежде всего, следует задаться вопросом: возможно ли представление функции многих переменных с помощью функции одного или нескольких переменных. Оказалось, что, рассматривая класс непрерывных функций, разрешение данного вопроса возможно. Обращая внимание на функцию $Y = f(X, a)$, следует говорить о точном ее воспроизведении (аппроксимации). Следовательно, решение задачи направлено на вычисление функций, имеющих наилучшее приближение к истинной. Теорема Вейерштрасса утверждает, что непрерывную функцию нескольких пере-

менных $f(X_1, \dots, X_n)$ на замкнутом ограниченном множестве Ω можно равномерно приблизить последовательностью полиномов: для любого положительно малого значения ε существует такой многочлен $F(X_1, \dots, X_n)$, что

$$\max_{\Omega} |f(X_1, \dots, X_n) - F(X_1, \dots, X_n)| < \varepsilon.$$

О возможностях представления функций нескольких переменных функцией одной переменной сказано в [2]. Будем следовать логике рассуждений, представленных в данной работе.

2. ПРЕДСТАВЛЕНИЕ ФУНКЦИЙ С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ

Кроме аппроксимации функций многочленами широкое применение в настоящее время находят способы приближения функций многих переменных с помощью линейных операций и суперпозиций функций одного или нескольких переменных. Опорой здесь являются нейронные сети. Из всего многообразия существующих ныне моделей нейронной сети возьмем за основу сеть прямого распространения (персептрон). Имеющая многослойную структуру, состоящую из нейронов, она способна решать задачи оценивания и аппроксимации функций по нескольким точкам (примерам) и по цепочке "вход-выход". В дальнейшем с их помощью задача прогноза сводится к получению неизвестных значений на выходе сети по известным входам.

Нейронные сети позволяют вычислять функции одного переменного и суперпозиции (функции от функций). Возможности нейронных сетей прямого распространения сколь угодно точно воспроизводить $Y = f(X)$ и их принцип работы представлен в [3].

Архитектура нейронной сети, в общем виде, состоит из входного, выходного (один выходной нейрон) и нескольких внутренних слоев. На вход подаются факторы X_1, \dots, X_n , на выходе - получают Y . С выходного нейрона снимается сигнал, представляющий собой вычисляемые этой сетью функции от входных сигналов. В общем виде, она представляется как

$$Y(X, a) = \Phi_l \sum_{i=1}^k \left(\Phi_{l-1, i} \left(\dots \left(\sum_{i=1}^k \left(\Phi_{1i} \left(\sum_{j=1}^n (a_{1ij} X_j) \right) \right) \dots \right) \right) \right),$$

$$(j = \overline{1, n}), (i = \overline{1, k}).$$

Здесь Σ - адаптивный линейный сумматор, вычисляющий скалярное произведение вектора входного сигнала (на первом слое - X) на вектор параметров (вес синапса) a ; Φ - нелинейный преобразователь, преобразующий выходной сигнал сумматора и являющийся функцией нескольких переменных $\Phi = (a, X_1, \dots, X_n)$; n - число переменных; k - число функций Φ ; l - число слоев в нейронной сети.

Возвращаясь к вопросу о возможностях нейронной сети как аппарата для построения зависимости $Y = f(X)$, следует обратиться к обобщенной аппроксимационной теореме Стоуна. Ее трактовка является утверждением об универсальных аппроксимационных свойствах любой нелинейности [3]: "с помощью линейных операций и каскадного соединения можно из произвольных нелинейных элементов получать любой требуемый результат с любой первоначально заданной точностью".

3. ОЦЕНКА И СОГЛАСОВАННОСТЬ ЗАВИСИМОСТЕЙ

Точность оценки неизвестных зависимостей нередко оценивается следующим образом. Пусть истинные значения показателей Y и X связаны зависимостью $Y = f(X)$, а по наблюдаемым значениям (X, Y) построена зависимость $Y = \Phi(X)$. Точность восстановления истинного значения может быть охарактеризована показателем отклонения U_t . Обычно пользуются абсолютным отклонением $U_t = |Y_t - \Phi(X_t)|$ и относительными отклонениями $U_t = |Y_t / \Phi(X_t) - 1|$ или $U_t = |\Phi(X_t) / Y_t - 1|$. Использование этих отклонений оправдано в случаях, когда объясняемая характеристика Y наблюдается с ошибками. Однако определение U_t теряет свое предназначение, если ошибки присущи наблюдениям объясняющих переменных. Надо отдать должное тому, что оценка точности истинного значения с помощью U_t не является единственной. Далее в работе будет представлен алгоритм оценки адекватности полученной зависимости с истинной. При любом способе оценки неизвестной зависимости приходится учитывать требования согласованности полученной зависимости с имеющейся информацией. Особая роль здесь отводится наличию различного рода ошибок. Их разнородность, трудность формализации, накладывают на процесс исследования дополнительные затраты. Выделяют систематические и случайные ошибки. Систематические ошибки постоянны при наблюдениях всех значений показателей, их можно предугадать и в последующем свести к минимуму. Что касается случайных ошибок, вызванных проявлением случайных факторов, то они колеблются и эти колебания трудно предсказуемы. Понятие "случайная ошибка" относительно, поскольку в зависимости от постановки задачи исследования она может иметь "временный" характер. Например, на некоторых интервалах времени влияние отдельных факторов имеет "вялотекущий" процесс и характеризуется как случайная ошибка. Однако под действием сторонних возмущений параметр претерпевает резкие (скачкообразные) изменения и тем самым переходит из разряда "случайных" в разряд значимых. Решение задачи о согласованности, чаще всего, опирается на "общеизвестный" метод наименьших квадратов. Но он нуждается в содержательном обосновании исходя из условий поставленной задачи.

Оценивая полученные зависимости с результатами наблюдений, должен соблюдаться общий принцип макси-

мальной согласованности. В соответствии с ним оценка зависимости должна быть наилучшей с точки зрения критерия, отражающего степень ее "согласованности" с имеющейся информацией. Центральным при этом становится вопрос о разумном выборе измерителя "степени согласованности". В числе одного из них можно предложить подход касающейся оценки статистических гипотез.

4. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ О СОГЛАСОВАННОСТИ

В качестве примера согласованности сопоставим два метода, регрессионный и нейросетевой, при помощи которых строятся и воспроизводятся зависимости соответственного вида $Y = f(X, a)$ и $Y = f(X)$. Выбор этих методов не является случайным, поскольку первый широко применяется, и второй находит свое применение в задачах техники, экономики и т.д. Последующий подбор критерия оценки должен явиться дополнением к общей части процесса "согласованности".

Регрессионная зависимость может быть как линейной, так и нелинейной. Нейросетевая модель позволяет лишь получать конечные результаты после ее успешного обучения. Поэтому будем принимать во внимание значения параметров Y рассматриваемых моделей. При этом полагаем, что ошибки являются независимыми нормально распределенными случайными величинами с нулевыми средними значениями и одинаковой дисперсией.

В качестве гипотезы о "согласованности" примем при заданном уровне значимости α , гипотезу о сопоставлении отклонений с весьма малым значением U_0 , т.е.

$H_0: U_t \leq U_0$. Проверка данной гипотезы осуществляется исходя из известного закона распределения Стьюдента, поскольку среднее квадратическое отклонение случайных переменных неизвестно. Предположение о подчиненности случайной величины U_t нормальному закону распределения позволяет воспользоваться статистической характеристикой

$$t_{\text{расч}} = \frac{\bar{U}_t - U_0}{S} \sqrt{T-1},$$

где $\bar{U}_t = \frac{1}{T} \sum_t U_t$ - среднее отклонение; $S = \frac{1}{T-1} \sum_t (U_t - \bar{U})^2$ -

среднее выборочное квадратическое отклонение. Согласно распределению Стьюдента (определяется квантиль $t_{\alpha\nu}$, где $\nu = T-1$) при условии $|t_{\text{расч}}| \leq t_{\alpha\nu}$ гипотеза H_0 принимается, в противном случае - отвергается.

Выдвигая гипотезу H_0 и проверяя ее, можно судить о

степени "согласованности" модели с желаемыми результатами.

Выделение такого критерия в качестве меры согласованности обусловлено еще и тем, что он чувствителен к изменениям оцениваемых параметров, т.е. к "выбору общего вида строящейся (воспроизводящейся зависимости)". В этой связи полезно сослаться на указанные в [4] принципы выбора общего вида зависимости:

- максимальное использование априорной информации об анализируемой зависимости;
- предварительный анализ структуры исходных данных;
- возможность упростить искомую зависимость;
- построенная зависимость должна быть, возможно, более устойчива по отношению к той совокупности исходных данных, на основании которых она оценена.

Последний принцип тесно связан с возможностями использования нейронных сетей. Реализация этого принципа предусматривает координацию оценок искомой зависимости, полученных на основании различных комбинаций "подвыборок" факторов воздействия из общего числа наблюдений. Неудачная комбинация выделенных факторов и "подвыборок" приведет к необходимости в отклонении гипотезы H_0 и/или (в худшем случае) к невозможности обучаться. Дать четкие рекомендации по формированию требующихся "подвыборок" в общем случае трудно. Однако можно использовать подход, когда на вход нейронной сети первоначально вводятся более значимые факторы X . После того, как сеть обучится, выполняется проверка на выполнение H_0 . Если в ходе анализа последняя будет отвергнута, то следует добавить очередной (уже менее значимый, чем предыдущие) фактор. Выбор факторов, их уровня значимости - сложная задача и во многом зависит от опыта, "знаний и интуиции исследователя".

ЗАКЛЮЧЕНИЕ

Практика расчетов показала, что нейронные сети прямого распространения являются эффективным инструментом анализа и предсказания поведения объектов исследования. Можно ожидать, что в недалеком будущем реализованные на базе нейронных сетей алгоритмы анализа и прогноза займут достойное место в системе построения зависимостей.

ПЕРЕЧЕНЬ ССЫЛОК

1. Клейнер Г.Б., Смоляк С.А. Экономические зависимости: Принципы и методы построения. - М.: Наука 2000. - 104 с.
2. Горбань А.Н. Функции многих переменных и нейронные сети // Соросовский образовательный журнал, № 12, 1998. С. 105-112.
3. Горбань А.Н., Росснев А.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. - 256 с.
4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справ. Изд. М.: Финансы и статистика, 1985. - 185 с.