

Метод оценки устойчивости нейронно-сетевых моделей¹.

Секерин А.Б. (sekerin@orel.ru)

Орловский государственный университет

Одним из наиболее широких направлений применения нейронных сетей в экономико-математическом моделировании является построение регрессионных моделей вида $Y = F(X_1, X_2, \dots, X_n)$, где Y – исследуемый показатель (объясняемая переменная), а X_1, X_2, \dots, X_n – объясняющие переменные (регрессоры) [1]. Как правило, для построения таких моделей используются сети, обучаемые на массивах статистических данных, отражающих наблюдаемую статистику зависимости переменной Y от переменных X_1, X_2, \dots, X_n (обучение с учителем). Основным преимуществом нейронных сетей по сравнению с классическими методами регрессионного анализа является принципиально более слабое требование к идентификации модели. В ходе применения классических методов регрессионного анализа в ходе идентификации модели необходимо достаточно точно указать вид искомой зависимости (например, $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$), задаваемый набором параметров, и дальнейший анализ фактически сводится к оценке значений этих параметров. Таким образом, в результате идентификации такой регрессионной модели значительно сужается класс функций $Y = F(X_1, X_2, \dots, X_n)$, в котором будет осуществляться поиск зависимости, наиболее согласующейся с данными наблюдений. Если модель идентифицирована неверно, то исследуемый класс не будет содержать искомую зависимость, т.е. последующий этап анализа будет заведомо обречен на неудачу. В отличие от методов классического регрессионного анализа, идентификация нейронно-сетевой модели состоит в выборе архитектуры нейронной сети,

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, региональный проект № 03-06-96424.

количества скрытых нейронов, а также передаточных функций. В этом случае существенного сужения исследуемого класса функций не происходит, поскольку класс зависимостей, выражаемый нейронными сетями заданной архитектуры, может быть весьма широким. В частности, на основе ставшей теперь классической теоремы А.Н. Колмогорова было установлено, что любую непрерывную функцию $f : [0,1]^n \rightarrow [0,1]$ можно аппроксимировать при помощи трехслойной нейронной сети, которая имеет n входных, $2n+1$ скрытых и один выходной нейрон [2, с. 61]. Одним из основных искомым свойств нейронно-сетевой модели является ее обобщающая способность. Суть данного свойства определяется на интуитивном уровне и состоит оно в том, что сеть должна показывать приблизительно одинаковые результаты на массиве данных, используемых для обучения, и на массиве данных, которые отличаются от обучающих данных, но достаточно близки к ним. Антиподом к обобщающему свойству нейронной сети является ее неустойчивость, обусловленная переходом процесса обучения в режим запоминания (memorize), образно называемый также «бабушкиным воспитанием». Такая сеть демонстрирует минимальную ошибку на обучающем массиве данных, но показывает высокий уровень ошибки при малых изменениях входных данных. Математическим аналогом такого режима является подгонка статистических данных полиномиальной зависимостью при условии, что степень полинома не ограничена. Для любого набора входных векторов X^1, X^2, \dots, X^N и выходных значений Y^1, Y^2, \dots, Y^N можно подобрать полином $P(X)$ достаточно высокой степени, такой, что для всех $j=1,2,\dots,N$ будут верны точные равенства $P(X^j) = Y^j$ (нулевое значение средней квадратической ошибки). При этом достаточно очевидно, что при незначительных изменениях массива ошибка может быть очень большой. Переход нейронной сети в режим запоминания, как правило, связан с «переопределенностью» сети, обусловленной чрезвычайно большим числом скрытых нейронов и, следовательно, большим количеством межнейронных соединений. В связи с этим возникает задача, каким образом можно проводить оценку

устойчивости нейронно-сетевой модели и ее обобщающей способности. С математической точки зрения устойчивость модели эквивалентна малым изменениям выходного параметра при малых изменениях входных параметров. Таким образом, речь идет о выполнении неравенства

$$|F(X) - F(X')| \leq A|X - X'|.$$

При этом оценка постоянной A может считаться оценкой уровня устойчивости модели. Проблема заключается в следующем. Для построения нейронно-сетевых регрессионных моделей чаще всего используется модель многослойного персептрона с гладкими передаточными функциями, а в качестве алгоритма обучения используется алгоритм обратного распространения ошибки (Back Propagation). В этом случае при любых значениях весов выходная функция нейронной сети $Y = F(X_1, X_2, \dots, X_n)$ является бесконечно дифференцируемой. Следовательно, для любого компакта K в \square^n существует постоянная $A(K)$, такая, что для $X, X' \in K$ выполняется неравенство

$$|F(X) - F(X')| \leq A(K)|X - X'|,$$

т.е. любая нейронная с гладкими передаточными функциями сеть обладает определенным уровнем устойчивости.

В связи с этим, предлагается считать модель устойчивой, если ее уровень устойчивости сравним с аналогичным уровнем устойчивости исходных данных, использовавшихся для обучения сети. Смысл такого подхода состоит в следующем. Обучение нейронной сети проводится на основе гипотезы, что набор статистических данных (X^i, Y^i) , где $X^i \in \square^n$ представляет собой таблицу значений искомой функции $Y=F(X)$. Поскольку, процесс обучения сети состоит в нахождении функции, наилучшим образом аппроксимирующей искомую зависимость, то оценивающая зависимость (функция) должна обладать количественными характеристиками, близкими к соответствующим характеристикам искомой функции. В частности, уровень устойчивости оценивающей функции должен

быть сравним с оценкой этого уровня искомой функции. Оценку уровня устойчивости искомой функции можно получить на основе исходных данных.

Рассмотрим следующий алгоритм. Пусть (X^i, Y^i) , $i=1,2,\dots,N$ – таблица исходных статистических данных. Для каждого фиксированного индекса i находим входной вектор X^j , ближайший к X^i относительно евклидовой метрики:

$$\min_{k \neq i} |X^i - X^k| = |X^i - X^j|$$

где

$$|X^i - X^k| = \sqrt{|X_1^i - X_1^k|^2 + |X_2^i - X_2^k|^2 + \dots + |X_n^i - X_n^k|^2}$$

Номер j зависит при этом от i , т.е. $j=j(i)$. Пусть $\varepsilon_i = |X^i - X^{j(i)}|$. Положим

$$K_i = \frac{|Y^i - Y^{j(i)}|}{\varepsilon_i}. \quad (1)$$

Пусть теперь $Y = F(X)$ нейронная сеть, обученная на массиве данных (X^i, Y^i) , $i=1,2,\dots,N$. Введем в рассмотрение случайную величину, являющуюся сравнительной характеристикой устойчивости исходных данных и нейронной сети. Для каждого i выберем случайный единичный вектор $\omega_i \in \square^n$. Положим

$$X^i[\varepsilon_i, \omega_i] = X^i + \varepsilon_i \omega_i, \quad \hat{Y}^i = F(X^i), \quad \hat{Y}^i[\varepsilon_i, \omega_i] = F(X^i[\varepsilon_i, \omega_i]).$$

Пусть далее

$$\hat{K}_i = \frac{|\hat{Y}^i - \hat{Y}^i[\varepsilon_i, \omega_i]|}{\varepsilon_i}, \quad (2)$$

Точки $\hat{X}^i[\varepsilon_i, \omega_i]$ и $X^{j(i)}$ отстоят от точки X^i на равном расстоянии ε_i , поэтому величины K_i и \hat{K}_i , задаваемые соответственно формулами (1) и (2) должны быть в каком-то смысле близки, т.е. их отношение должно быть близко к 1. Тогда набор чисел

$$Z_i = \frac{\widehat{K}_i}{K_i}$$

Можно считать набором наблюдаемых значений случайной величины Z . Устойчивость нейронно-сетевой модели $Y=F(X)$ будем считать соответствующей устойчивости исходных данных, если распределение случайной величины $\ln Z$ близко к нормальному распределению с нулевым математическим ожиданием (т.е. распределение величины Z является логнормальным). При этом возможна ситуация, когда $E(\ln Z) \leq 0$. Это означает, что нейронно-сетевая модель более устойчива, чем исходные данные. В качестве обобщающего показателя сравнительной устойчивости нейронно-сетевой модели может использоваться вероятность $P(\ln Z < 0)$, т.е. вероятность того, что устойчивость модели будет не ниже, чем устойчивость исходных данных.

Таким образом, оценка соответствия уровня устойчивости нейронно-сетевой модели должна проводиться по следующей схеме:

- а) Построение эмпирического распределения величины $\ln Z$;
- б) Проверка гипотезы о нормальности распределения величины $\ln Z$;
- в) Оценка вероятности $P(\ln Z < 0)$.

ЛИТЕРАТУРА

1. Галушкин А.И. Теория нейронных сетей / Сер. Нейрокомпьютеры и их применение. Кн. 1. – М.: ИПРЖР, 2000. – 416 с.
2. Головки В.А. Нейронные сети: обучение, организация и применение. Кн.4: Учебн. пособие для вузов / Общая ред. А.И. Галушкина. – М. ИЖПР, 2001. – 256 с.