

Нейросетевой подход к задаче оценки размера взноса при страховании гражданской ответственности владельцев транспортных средств

Приведены результаты экспериментальных исследований аппроксимационных свойств радиально-базисных нейронных сетей для прогнозирования размера страхового взноса в автостраховании.

The results of the experimental investigations of radial-basis neural network's approximating properties for the prediction of the size of the insurance premium in the vehicle insurance are given.

Подано результати експериментальних досліджень апроксимаційних властивостей радіально-базисних нейронних мереж для прогнозування розміру страхового внеску в автострахованні.

Введение. Обязательное страхование автогражданской ответственности (ОСАГО) получило широкое распространение во всех цивилизованных странах. Как известно, оно обязательно для автовладельца, так как эта форма страхования обеспечивает возмещение ущерба потерпевшим. Объектом такого страхования выступает ответственность страхователя по закону перед третьими лицами, которым может быть причинен ущерб в силу несчастного случая или профессиональной ошибки.

Определение размера страхового взноса, который принято называть страховой премией, т.е. суммы, которую должен оплатить владелец транспортного средства (ТС) при покупке страхового полиса ОСАГО, – достаточно трудная задача для сотрудников страхового бизнеса, так как при ее расчете должно быть учтено много факторов, связанных как с водителем (водителями) ТС, так и с самим ТС, а зависимость эта точно не установлена.

В настоящее время существуют различные достаточно эффективные методы и алгоритмы восстановления функциональных зависимостей по экспериментальным данным, включая метод группового учета аргументов (МГУА), предложенный А.Г. Ивахненко для построения математических моделей сложных систем и изложенный в монографиях и периодических изданиях [1–4].

В последнее десятилетие внимание исследователей привлекают методы построения слож-

ных моделей на базе искусственных нейронных сетей. Такие сети становятся хорошими аппроксиматорами неизвестных нелинейных функций [5–7]. Это позволяет применять их во многих экономико-финансовых областях знаний. В частности, в работе [8] на базе искусственной нейронной сети с так называемыми сигмоидальными активационными функциями была поставлена и решена задача прогнозирования стоимости объектов недвижимости.

Судя по доступным литературным источникам, какие-либо работы по построению математических моделей, которые могли бы связать размер страхового взноса ОСАГО с определенными характеристиками водителей и ТС, ранее не встречались.

Цель данной статьи – исследование возможности оценки страховым агентом размера взноса при оформлении автовладельцем страхового полиса ОСАГО на основе нейросетевой системы радиально-базисного типа [5, 6].

Характеристика исходных данных

Известно, что существует зависимость между размером C страхового взноса и следующими переменными: типом x_1 договора; статусом x_2 страхователя; резидентностью x_3 страхователя; количеством x_4 водителей, включаемых в страховой договор; минимальным водительским стажем x_5 ; типом x_6 ТС; зоной x_7 страхования.

Далее приведено описание переменных, которые могут быть использованы при определении размера страхового взноса ОСАГО.

Тип договора (x_1). Договор страхования ОСАГО может быть заключен на таких условиях:

- страхование ответственности за вред, причиненный жизни, здоровью, имуществу третьих лиц вследствие эксплуатации ТС, указанного в договоре страхования, любым лицом, эксплуатирующим его на законных основаниях (договор типа I);

- страхование ответственности за вред, причиненный жизни, здоровью, имуществу третьих лиц вследствие эксплуатации любого ТС или одного из ТС, указанных в договоре, лицом, указанным в договоре страхования (договор типа II);

- страхование ответственности за вред, причиненный жизни, здоровью, имуществу третьих лиц вследствие эксплуатации ТС, указанного в договоре страхования, лицом, указанным в договоре страхования, или одним из лиц, указанных в договоре (договор типа III).

Статус страхователя (x_2). Юридическое или физическое лицо.

Резидентность страхователя (x_3). Является страхователем гражданином Украины или нет.

Количество водителей (x_4). Максимальное количество лиц, которые будут управлять ТС на протяжении действия договора страхования.

Минимальный водительский стаж (x_4). Если ТС будут управлять несколько лиц, то учитывается наименьший водительский стаж среди этих лиц.

Тип транспортного средства (x_6) (табл. 1).

Таблица 1. Типы ТС

Легковые автомобили	
Объем двигателя в см ³	Тип ТС
До 1600 см ³	B1
1600 ... 2000 см ³	B2
2000 ... 3000 см ³	B3
Более 3000 см ³	B4
Прицепы к легковым автомобилям	F
Автобусы	
С количеством мест до 20 чел.	D1
С количеством мест более 20 чел.	D2
Грузовые автомобили	
Грузоподъемностью до 2 т	C1
Грузоподъемностью более 2 т	C2
Прицепы к грузовым автомобилям	E

Зона страхования (x_7). В зависимости от количества населения населенный пункт, в котором страхуется ТС, имеет свою зону (табл. 2).

Таблица 2. Зона страхования

Количество населения	Зона
Менее 100 тыс. чел.	5
Более 100 тыс. чел. и менее 500 тыс. чел.	4
Более 500 тыс. чел. и менее 1 млн. чел.	3
Более 1 млн. чел.	2

Задача исследования

Из изложенного следует, что между размером C страхового взноса и семью независимыми переменными $x = [x_1, \dots, x_7]^T$, упорядоченный набор которых образует семимерный вектор, существует определенная функциональная зависимость:

$$C = F(x). \quad (1)$$

Здесь и в дальнейшем T – знак транспонирования. Составляющие вектора x могут принимать значения (табл. 3).

Таблица 3. Возможные значения составляющих вектора x

Переменная	Значения
x_1	1, 2, 3
x_2	1, 2
x_3	1, 2
x_4	1, 3, 5
x_5	0, 1, 3, 10
x_6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
x_7	2, 3, 4, 5

Хотя зависимость (1) в явном виде априори неизвестна (неизвестен оператор $F: \mathbf{R}^7 \rightarrow \mathbf{R}$, формально отображающий семимерное евклидово пространство векторов x на одномерном евклидовом пространстве скалярных величин C), но зато имеется множество L пар

$$\underbrace{\{x^{(1)}, C^{(1)}\}, \dots, \{x^{(L)}, C^{(L)}\}}_L \quad (2)$$

каждая из которых определяется набором значений переменных $x_1^{(i)}, \dots, x_7^{(i)}$ (табл. 3) вместе со страховым взносом $C^{(i)}$, соответствующим данному i -му набору ($i = 1, \dots, L$).

Следуя работе [9], аппроксимируем $F(x)$ функцией

$$\hat{C} = F_0(x, w), \quad (3)$$

зависящей не только от вектора x , но и от некоторого пока не известного вектора w . В этом выражении \hat{C} приобретает смысл прогнозируемого страхового взноса.

Качество аппроксимации будем оценивать среднеквадратичной ошибкой, которая в силу (3) зависит от вектора w , и с учетом (1) определяется выражением

$$\bar{E}_{\text{кв}} = \sqrt{\frac{1}{L} [(C^{(1)} - \hat{C}^{(1)})^2 + \dots + (C^{(L)} - \hat{C}^{(L)})^2]}, \quad (4)$$

которая в силу (1) и (3) зависит от вектора w и определяется выражением

$$\bar{E}_{\text{кв}}(w) = \sqrt{\frac{1}{L} \sum_{i=1}^L [F(x^{(i)}) - F_0(x^{(i)}, w)]^2}. \quad (5)$$

Процесс обучения нейросетевой системы (3) сводится, как известно [7], к использованию множества пар (2), позволяющему найти такой вектор $w = w^*$, при котором $\bar{E}_{\text{кв}}(w^*)$ достигает минимума:

$$\bar{E}_{\text{кв}}(w^*) = \min_w \bar{E}_{\text{кв}}(w). \quad (6)$$

Согласно (6) с учетом выражения (5) искомый вектор w^* определяется как

$$w^* = \arg \min_w \sum_{i=1}^L [F(x^{(i)}) - F_0(x^{(i)}, w)]^2. \quad (7)$$

Задача исследования состояла в том, чтобы на конкретных реальных данных проверить возможность обучения нейронной сети радиально-базисного типа для достижения приемлемой среднеквадратичной ошибки прогнозирования (4).

Структура нейросетевой системы

При построении нейросетевой системы как радиально-базисной сети оценка стоимости \hat{C} страхового взноса будет определяться как

$$\hat{C} = w^{(2)T} y^{(1)} + b^{(2)}. \quad (8)$$

Здесь $b^{(2)}$ – смещение второго (выходного) слоя, $w^{(2)T}$ – L -мерный вектор весов этого слоя, $y^{(1)} = [y_1^{(1)}, \dots, y_L^{(1)}]^T$ – L -мерный вектор, компонентами которого служат выходные сигналы скрытого слоя нейронной сети $y_i^{(1)}$ ($i = 1, \dots, n$). Каждый i -й сигнал скрытого слоя определяется так:

$$y_i^{(1)} = \varphi(r_i) \quad (i = 1, \dots, L). \quad (9)$$

В этом выражении

$$\varphi(r_i) = \exp(-r_i^2) \quad (10)$$

– так называемая радиально-базисная функция [5], зависящая от переменной

$$r_i = b^{(1)} d(x, w_i^{(1)}), \quad (i = 1, \dots, L) \quad (11)$$

в которой коэффициент $b_i^{(1)}$ [7] находится по формуле

$$b_i^{(1)} = \frac{\sqrt{L}}{d_{\text{max}}^2}, \quad (12)$$

где

$$d_{\text{max}} := \max_{i,j=1,\dots,L} \|w_i^{(1)} - w_j^{(1)}\| \quad (13)$$

обозначает максимальное расстояние между всеми L -векторами скрытого слоя;

$$d(x, w_i^{(1)}) = \|x - w_i^{(1)}\| \quad (14)$$

имеет смысл расстояния между входным вектором $x = [x_1, \dots, x_7]^T$ и i -м вектором весов $w_i^{(1)}$ скрытого слоя.

На рис. 1 изображена радиально-базисная функция $\varphi(r)$ вида (10).

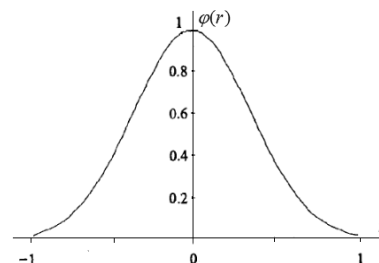


Рис. 1. График радиально-базисной активационной функции с центром в начале координат

Особенностью нейросетевой системы радиально-базисного типа есть то, что в качестве i -го вектора весов скрытого слоя $w(i)$ принимается i -й входной вектор $x(i)$ [7]:

$$w_i^{(1)} = x^{(i)} \quad \text{для всех } i. \quad (15)$$

Структурная схема нейросетевой системы показана на рис. 2, где операция определения расстояний $d(x, w_i^{(1)})$ в скрытом слое символически обозначена как $\|\cdot\|$, а операция определения скалярного произведения векторов $w^{(2)T}$ и $y^{(1)}$ как $\langle \cdot, \cdot \rangle$.

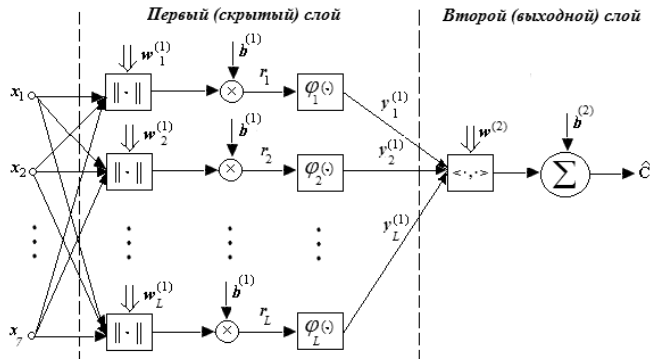


Рис. 2. Двухслойная радиально-базисная нейронная сеть

Поскольку в силу соотношений (9) – (11) $w_i^{(1)}, b_i^{(1)}, w_i^{(2)} \in R^7$, а $b^{(2)} \in R$, то можно заключить, что вектор w , фигурирующий в выражении (3) для аппроксимирующей функции $F_0(\cdot, \cdot)$, будет представлять собой $9L + 1$ -мерный вектор вида

$$w = \underbrace{[w_1^{(1)}, \dots, w_L^{(1)}, b^{(1)}, w^{(2)}, b^{(2)}]^T}_{9L+1}. \quad (16)$$

Процедура обучения

Процедура сводится к нахождению псевдорешения системы уравнений вида [6, п. 5.13.1]

$$w^{(2)T} y_i^{(1)} + b^{(2)} = C^{(i)}, \quad i = 1, \dots, L \quad (17)$$

относительно неизвестных вектора $w^{(2)}$ весов выходного слоя и смещения $b^{(2)}$, где

$$y_i^{(1)} = [y_{i1}^{(1)}, \dots, y_{iL}^{(1)}]^T \quad (18)$$

– вектор выходных сигналов первого (скрытого) слоя, появляющихся при предъявлении на вход нейросети i -го вектора $x^{(i)} = [x_1^{(i)}, \dots, x_7^{(i)}]^T$.

(Система уравнений (18) несовместна.)

Учитывая выражение (18), получаем псевдорешение $w = [w^{(2)T}, b^{(2)}]^T$ системы (17) в виде

$$w = Y^+ c, \quad (19)$$

где $c = [c^{(1)}, \dots, c^{(L)}]^T$, а Y^+ – матрица, псевдообратная $L \times (L + 1)$ -матрице

$$Y = \begin{pmatrix} y_{11}^{(1)} & \dots & y_{1L}^{(1)} & 1 \\ \cdot & \cdot & \cdot & \cdot \\ y_{L1}^{(1)} & \dots & y_{LL}^{(1)} & 1 \end{pmatrix}.$$

Выбор вектора в виде (19) согласно [9] гарантирует решение поставленной задачи оптимизации (5), в которой вектор w^* определяется как $w^* = w$.

Экспериментальные результаты

Для экспериментального исследования возможностей обучения нейросетей и оценки эффективного прогнозирования размера страхового взноса в среде *MATLAB* была разработана прикладная программа, позволяющая после завершения процесса обучения реализовать функциональные преобразования вида (8).

При проведении модельных экспериментов исходная информация о размерах страховых взносов ОСАГО была представлена набором $L = 320$ пар (2). Заданная точность прогнозирования была принята равной $\bar{E}_{\text{зад}} = 0,022$ тыс. грн.

Как и при реализации алгоритма МГУА [5, с. 124], имеющийся набор указанных пар (2) был разделен на две группы, а именно на обучающее множество, содержащее $L_{\text{об}}$ таких пар, и проверочное множество $L_{\text{пров}}$ пар, т.е.

$$L = L_{\text{об}} + L_{\text{пров}}. \quad (20)$$

Было взято $L_{\text{об}} = 300$, $L_{\text{пров}} = 20$.

Для сравнения результатов работы сетей разных архитектур была создана другая нейронная сеть, использующая на том же обучающем множестве алгоритм обратного распространения ошибки [8]. Процесс обучения для сети второго типа прекращался на той итерации, когда впервые выполнялось требование

$$\bar{E}(n) \leq E_{\text{зад}}, \quad (21)$$

где $n = 1, 2, \dots$ – порядковый номер итерации, а $\bar{E}(n)$ – соответствующая среднеквадратичная ошибка прогнозирования.

Поведение величины $\bar{E}(n)$ в процессе обучения показано на графике (рис. 3).

Результаты обучения нейросетей приведены в табл. 4.

В обоих случаях для проверки точности аппроксимации после завершения обучения на вход сети подавались образцы, не участвующие при ее обучении, и оценивалась точность аппроксимации.

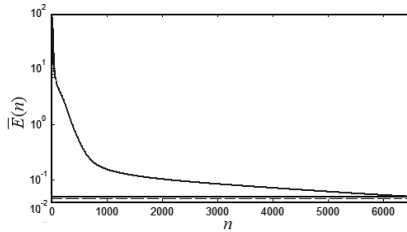


Рис. 3. Процесс обучения нейросети прямого распространения (сплошная линия) и радиально-базисной (прерывистая линия)

Таблица 4. Результаты обучения нейронных сетей двух типов

Архитектура нейронной сети	Продолжительность процесса обучения (в числе итераций)	Среднеквадратичная ошибка прогнозирования на проверочном множестве
Радиально-базисная	–	0,043
Прямого распространения	6470	0,069

Пример 1. Пусть требуется спрогнозировать размер страхового взноса ОСАГО при таких условиях для договора страхования: имеет тип III; страхователь – физическое лицо, резидент Украины; количество водителей – не более двух; минимальный стаж водителей – менее одного года; тип ТС – легковое с объемом двигателя до 1600 см³; зона страхования – с населением более 1 млн. чел.

Результаты прогнозирования сети приведены в табл. 5.

Таблица 5. Сравнительная оценка размера страхового взноса в примере 1

Реальный страховой взнос ОСАГО, который был определен Страховой компанией, грн.	400,50
Страховой взнос ОСАГО, спрогнозированный посредством радиально-базисной сети, грн.	396,28
Страховой взнос ОСАГО, спрогнозированный посредством сети прямого распространения, грн.	389,13

Пример 2. Пусть требуется спрогнозировать размер страхового взноса ОСАГО при таких условиях для договора страхования: имеет тип III; страхователь – юридическое лицо, резидент Украины; количество водителей – не более пяти; минимальный стаж водителей – более пяти лет; тип ТС – грузовое с грузоподъемностью более двух тонн; зона страхования – с населением до 100 тыс. чел.

Результаты прогнозирования сети приведены в табл. 6.

Таблица 6. Сравнительная оценка размера страхового взноса в примере 2

Реальный страховой взнос ОСАГО, который был определен Страховой компанией, грн.	630,40
Страховой взнос ОСАГО, спрогнозированный посредством радиально-базисной сети, грн.	634,92
Страховой взнос ОСАГО, спрогнозированный посредством сети прямого распространения, грн.	621,12

Заключение. Показана возможность прогнозирования размера страхового взноса ОСАГО с использованием нейронной сети радиально-базисного типа. Для оценки качества приведены результаты прогнозирования традиционной сетью прямого распространения со стандартным градиентным алгоритмом обучения. При относительно небольшой обучающей выборке использование нейронной сети радиально-базисного типа может быть целесообразной альтернативой использованию сети со стандартным градиентным алгоритмом обучения: ведь для ее обучения на выборке, состоящей из трехсот образцов, понадобились доли секунды, в то время как во втором случае при аналогичной обучающей выборке понадобилось 6470 итераций обучения и около 10 минут машинного времени.

Приведенные примеры демонстрируют возможность использования предложенной архитектуры нейросетевой системы для прогнозирования размера страхового взноса ОСАГО.

1. Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д. Принятие решений на основе самоорганизации. – М.: Сов. радио, 1976. – 280 с.
2. Зайченко Ю.П. Основы проектирования интеллектуальных систем. – К.: Дім «Слово», 2004. – 352 с.
3. Ивахненко А.Г., Степашко В.С. Помехоустойчивое моделирование. – К.: Наук. думка, 1985. – 300 с.
4. Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей // Автоматика. – 1981. – № 3. – С. 31–36.
5. Бодянский Е.В., Руденко О.Г. Искусственные нейронные сети: архитектуры, обучение, применения. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с.
6. Haykin S. Neural networks, a comprehensive foundation. – NY: Prentice Hall, 1999. – 1121 p.
7. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation // Parallel distributed processing: explorations in the microstructure of cognition. – Cambridge: MIT Press, 1986. – 1. – P. 318–362.
8. Скурихин В.И., Житецкий Л.С., Николаенко С.А. Прогнозирование стоимости объектов недвижимости с использованием нейросетевых систем // УСИМ. – 2011. – № 5. – С. 86–92.
9. Альберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. – М.: Наука, 1977. – 224 с.

Поступила 06.03.2013
Тел. для справок: +38 044 526-1154 (Киев)
© С.А. Николаенко, 2013