

ОЦЕНКА КРЕДИТОСПОСОБНОСТИ ЗАЕМЩИКА В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Наталия Шовгун

Abstract: The problem of assessing the creditworthiness of the borrower is considered. The application of fuzzy neural networks for this problem solution, TSK and recurrent TSK was suggested. The experimental investigations of application of this networks for our task were carried out and comparison with classical methods was performed.

Keywords: fuzzy neural networks, credit rating, fuzzy logic

ACM Classification Keywords: H.4 Information systems applications - H.4.2. Types of Systems Decision Support

Введение

Основным видом деятельности банка является выдача кредитов. Соответственно риски связанные с возможным неисполнением своих обязательств заемщиком, являются одними из основных банковских рисков. Основным банковским продуктом является потребительское кредитование физических лиц. Для минимизации потерь банка необходим тщательный отбор заемщиков. Для эффективной оценки кредитоспособности заемщиков в банках существуют модели и методы принятия решений по кредитным заявкам. В анкете заемщика содержится множество данных, которые могут быть неполными, разнородными, недостоверными, а также нести описательный характер. В такой ситуации целесообразно использовать методы нечеткой логики, которые могут работать как с качественными, так и с количественными характеристиками. В работе применение нечеткого контроллера Мамдани обобщается на случай оценки кредитоспособности физических лиц и предлагается подход к решению задачи на основе применения нечеткой нейронной сети (ННС) TSK и рекуррентной ННС TSK и проведено сравнение полученных результатов со стандартными методами

Постановка задачи оценки кредитоспособности заемщика в условиях неопределенности

Каждая кредитная заявка задается вектором $X = \{X_1, X_2, \dots, X_i, \dots, X_N\}$, где X_i - некоторым образом формализованные данные из анкеты заемщика и параметры кредита. По данному вектору необходимо принять решение о выдаче кредита, то есть отнести заемщика к категории «хорошие» или к категории «плохие».

Предлагаемые методы основываются на алгоритме нечеткого логического вывода, этапами которого являются:

1. Определение множества входных переменных (вектор заявки): $\{X_1, X_2, \dots, X_i, \dots, X_M\}$;
2. Определение множества выходных переменных (категории заемщика): $D = \{D_1, D_2, \dots, D_i, \dots, D_M\}$;

3. Формирование базового терм-множества с соответствующими функциями принадлежности каждого термина: $A = \{a_1, a_2, \dots, a_i\}$;
4. Формирование конечного множества нечетких правил;
5. Нахождение четкого значения для каждой лингвистической переменной выхода.

Методы на основе нечеткой логики. Нечеткий алгоритм Мамдани и Сугено

В основе нечеткого контроллера Мамдани лежит нечеткий алгоритм Мамдани [Зайченко, 2008]. Рассмотрим его.

Пускай базу знаний составляют два нечетких правила:

П₁: если x это A_1 и y это B_1 , то z это C_1 ,

П₂: если x это A_2 и y это B_2 , то z это C_2 ,

где x и y – входные переменные, z – переменная выхода, $A_1, A_2, B_1, B_2, C_1, C_2$ – некоторые заданные функции принадлежности (ФП), при этом четкое значение z необходимо определить на основе данной информации и четких значений x, y . Этапами алгоритма являются:

1. Введение четкости. Находим степень истинности для предпосылок каждого правила: $A_1(x_0), A_2(x_0), B_1(y_0), B_2(y_0)$.
2. Логический вывод. Находим уровни «отсечения» для предпосылок каждого из правил:

$$\alpha_1 = A_1(x_0) \cap B_1(y_0),$$

$$\alpha_2 = A_2(x_0) \cap B_2(y_0).$$

Далее находим «усеченные» функции принадлежности:

$$C'_1 = (\alpha_1 \cap C_1(z)),$$

$$C'_2 = (\alpha_2 \cap C_2(z)).$$

3. Композиция. Находим объединение найденных усеченных функций принадлежности и получаем итоговое нечеткое подмножество для переменной вывода с функцией принадлежности:

$$\mu_z = C(z) = C'_1(z) \cup C'_2(z) = (\alpha_1 \cap C_1(z)) \cup (\alpha_2 \cap C_2(z)).$$

4. Приведение к четкости.

В ННС TSK используется нечеткий алгоритм Сугено с таким набором правил:

П₁: если x это A_1 и y это B_1 , то $z = a_1x + b_1y$,

П₂: если x это A_2 и y это B_2 , то $z = a_2x + b_2y$.

Этапы алгоритма имеют вид:

1. Введение нечеткости, как в алгоритме Мамдани.
2. Нечеткий вывод. Находим $\alpha_1 = A_1(x_0) \cap B_1(y_0)$, $\alpha_2 = A_2(x_0) \cap B_2(y_0)$ и индивидуальные выходы правил: $\dot{z}_1 = a_1x_0 + b_1y_0$, $\dot{z}_2 = a_2x_0 + b_2y_0$.
3. Определение четкого значения переменной вывода:

$$z_0 = \frac{\alpha_1 \dot{z}_1 + \alpha_2 \dot{z}_2}{\alpha_1 + \alpha_2}.$$

Нечеткая нейронная сеть TSK и рекуррентная ННС TSK

В сети TSK (Takagi, Sugeno, Kang'a) используется такая база правил [Зайченко, 2008]:

$$R_1: \text{если } x_1 \text{ это } A_1^{(1)}; x_2 \text{ это } A_2^{(2)}; \dots; x_n \text{ это } A_n^{(1)}, \text{ то } y_1 = p_{10} + \sum_{j=1}^N p_{1j} x_j,$$

$$R_M: \text{если } x_1 \text{ это } A_1^{(M)}; x_2 \text{ это } A_2^{(M)}; \dots; x_n \text{ это } A_n^{(M)}, \text{ то } y_M = p_{M0} + \sum_{j=1}^N p_{Mj} x_j,$$

где $A_i^{(k)}$ - значение лингвистической переменной x_i (характеристики заемщика) для правила R_k с

$$\text{функцией принадлежности } \mu_A^{(k)}(x_i) = \frac{1}{1 + \left(\frac{x_i - c_i^{(k)}}{\sigma_i^{(k)}}\right)^{2b_i^{(k)}}}.$$

В ННС TSK выделяют 5 слоев:

1. Первый слой выполняет раздельную фаззификацию каждой переменной x_i , $i=1,2,\dots,N$, определяя для каждого k -го правила вывода значение ФП $\mu_A^{(k)}(x_i)$ в соответствии с функцией фаззификации, например колоколообразной. Это – параметрический слой с параметрами $c_j^{(k)}$, $\sigma_j^{(k)}$, $b_j^{(k)}$, которые подлежат настройке, в процессе обучения сети.

2. Второй слой выполняет агрегирование отдельных переменных x_i , определяя результирующую степень принадлежности w_k вектора X условиям k -го правила, где

$$w_k = \mu_A^{(k)}(x) = \prod_{j=1}^N \frac{1}{1 + \left(\frac{x_j - c_j^{(k)}}{\sigma_j^{(k)}}\right)^{2b_j^{(k)}}}.$$

3. В третьем слое рассчитываются значения $y_k(x) = p_{k0} + \sum_{j=1}^N p_{kj} x_j$, а также производится умножение функций $y_k(x)$ и w_k , сформированных в предыдущем слое. Это параметрический слой с параметрами p_{k0} , p_{kj} , которые подлежат настройке.

4. В четвертом слое рассчитывается взвешенная сумма сигналов $y_k(x)$ и сумма весов $\sum_{k=1}^M w_k$.

5. В пятом слое веса подлежат нормализации и вычисляется выходной сигнал $y(x) = \frac{\sum_{k=1}^M w_k y_k(x)}{\sum_{k=1}^M w_k}$.

В рекуррентной ННС TSK [Jaung, 1999] выходы каждого правила подаются обратно на вход сети вместе с входными переменными x_i , таким образом, запоминается прошлое значение выхода каждого правила. Опишем слой за слоем рекуррентную ННС TSK:

1. Как и в ННС TSK, первый слой выполняет раздельную фаззификацию каждой переменной x_i $i=1,2,\dots,N$, определяя для каждого k -го правила вывода значение ФП $\mu_A^{(k)}(x_i)$ в соответствии с функцией фаззификации, например колоколообразной. Рассчитывается значение ФП внешней переменной h_k , которая соответствует выходу k -го правила. ФП имеет вид сигмоидной функции. Это параметрический слой с параметрами $c_j^{(k)}$, $\sigma_j^{(k)}$, $b_j^{(k)}$, которые подлежат настройке.

2. Второй слой выполняет агрегирование отдельных переменных x_i , определяя результирующую степень принадлежности w_k вектора X условиям k -го правила, где

$$w_k = \mu_A^{(k)}(x) = \frac{1}{1 + \exp(-h_k)} \prod_{j=1}^N \frac{1}{1 + \left(\frac{x_j - c_j^{(k)}}{\sigma_j^{(k)}}\right)^{2b_j^{(k)}}}.$$

3. В третьем слое рассчитываются значения $y_k(x) = p_{k0} + \sum_{j=1}^N p_{kj}x_j + p_{kN+1}h_k$, а также производится умножение функций $y_k(x)$ и w_k , сформированных в предыдущем слое. Это параметрический слой с параметрами p_{k0} , p_{kj} , которые подлежат настройке.

4. В четвертом слое вычисляются суммы выходов правил $h_k = \sum_{j=1}^M \mu_A^{(k)} v_{kj}$, которые затем передаются в первый слой. Настройке подлежат параметры v_{kj} .

5. В пятом рассчитывается взвешенная сумма сигналов $y_k(x)$ и сумма весов $\sum_{k=1}^M w_k$.

6. В шестом слое веса подлежат нормализации и вычисляется выходной сигнал $y(x) = \frac{\sum_{k=1}^M w_k y_k(x)}{\sum_{k=1}^M w_k}$.

При настройке параметров ФП $c_j^{(k)}$, $\sigma_j^{(k)}$, $b_j^{(k)}$ и h_k использовался алгоритм Resilient Propagation для сокращения процесса обучения сети.

Параметры p_{k0} , p_{kj} находятся в результате решения системы из L линейных уравнений, при фиксации

прочих параметров. Система имеет вид $d^l = \frac{\sum_{k=1}^M w_k y_k(x^l)}{\sum_{k=1}^M w_k}$ при фиксации прочих параметров, где d^l -

желаемый выход сети, $l=1,\dots,L$.

Экспериментальные исследования

Для оценки качества предложенных методов оценки кредитоспособности заемщика, важным показателем является общая интенсивность ошибок (ИО), которая определяется отношением числа неверных классификаций к общему числу случаев и наличие ошибок первого и второго рода. Ошибкой второго рода является отнесение заемщика к «плохим», хотя на самом деле он «хороший», а ошибкой первого рода является отнесение заемщика к «хорошим», хотя на самом деле он «плохой». Также для сравнения разных методов рассчитывается СКО. Задача оценки кредитоспособности заемщика является задачей бинарной классификации. Так как на выходе сети формируется непрерывное значение, то важна точка разделения примеров на положительные и отрицательные. В результате проведенных экспериментов пороговое значение выбрано 0.55. Зависимость интенсивности ошибок от порогового значения приведена на графике 1.

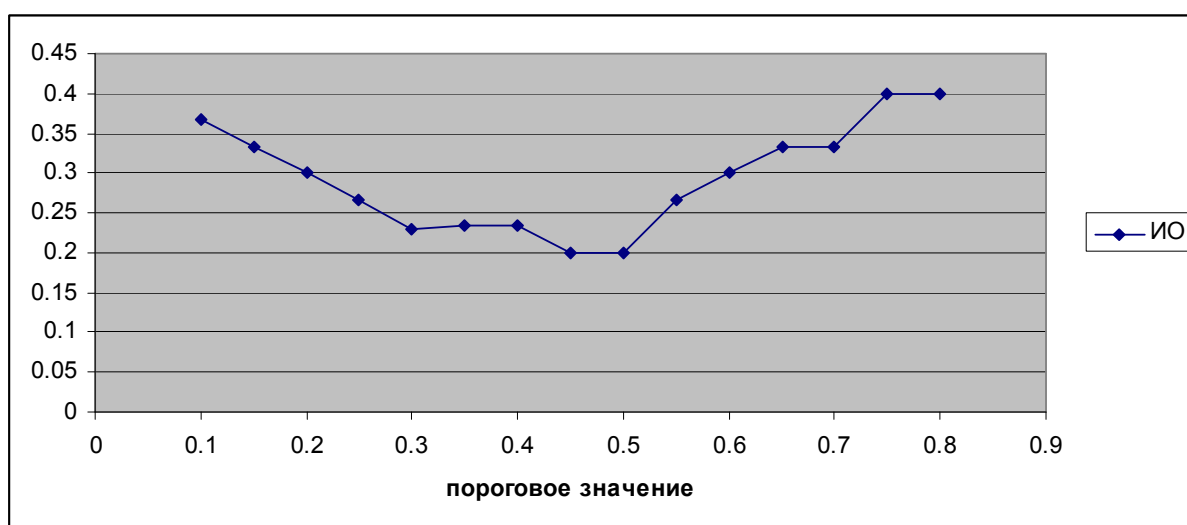


График 1. Интенсивность ошибок

Как видно из графика, при увеличении порогового значения до 0.55 ИО падает, а при дальнейшем его увеличении, возрастает.

Для анализа кредитоспособности заемщика с помощью предложенных методов использовалась выборка данных одного из украинских банков, содержащая 500 кредитных заявок за 2010г. Для сравнения приведены наиболее популярные в оценке кредитоспособности заемщика логит и пробит модели, а также сеть Байеса (наивный байесовский классификатор). Результаты оценки кредитоспособности заемщика можно увидеть в Табл.1.

Табл.1 Результаты оценки кредитоспособности разными методами

Метод	СКО	ИО, %
Логит модель	0,1142	26,6
Пробит модель	0,1142	26,6
ННС TSK	0.1066	20
Рекуррентная ННС TSK	0.1214	23,3
НК Мамдани	0.1075	21,6
Сеть Байеса	-	23,3

Соотношение ошибок первого и второго рода для разных методов показаны на графиках 2-3.

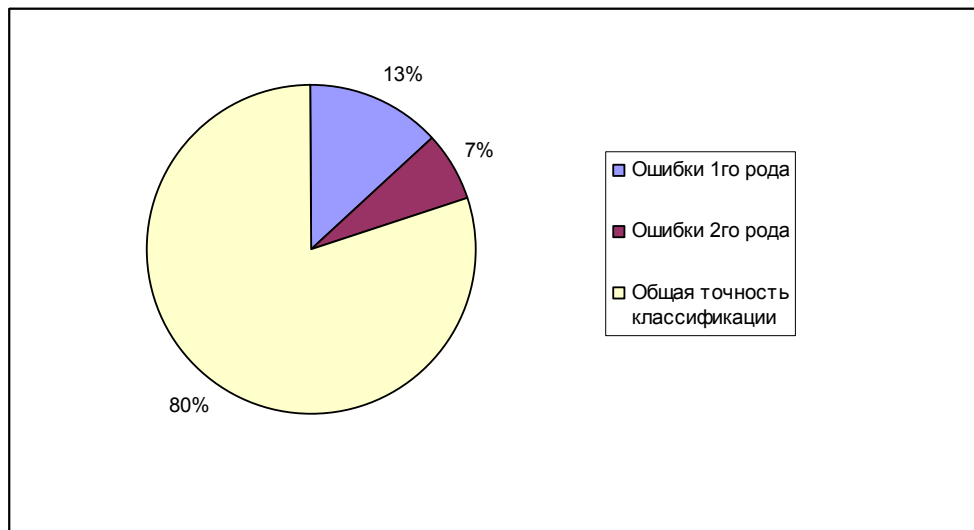


График 2. Результаты классификации с помощью ННС TSK

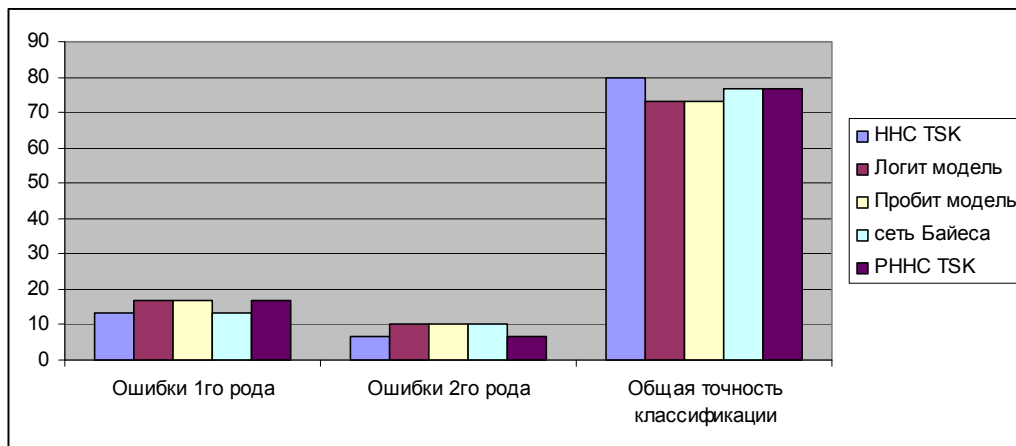


График 3. Результаты классификации разными методами

Как видно из графика количество ошибок первого рода больше, чем ошибок второго рода, а это потенциальные потери для банка. При увеличении порогового значения до 0,7 увеличивается ИО, но уменьшается значение ошибок 1го рода, график 4.

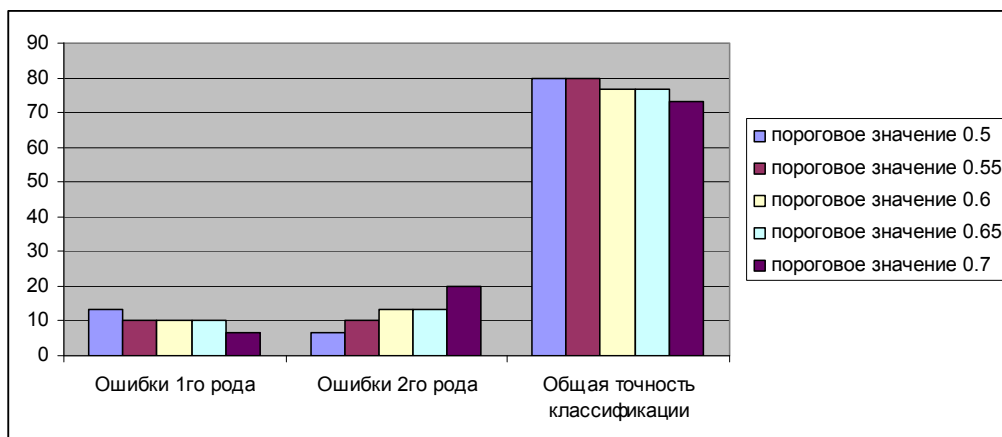


График 4. Количество ошибок в зависимости от порогового значения, ННС TSK

Выводы

Предложенный подход к оценке кредитоспособности заемщика с помощью нечеткого контроллера Мамдани и нечетких нейронных сетей показывает лучшие результаты, чем стандартные методы. В ходе экспериментов выбрано оптимальное пороговое значение 0.5 для разделения заемщиков на «хорошие/плохие». Для снижения потерь банка, за счет уменьшения количества ошибок первого рода необходимо повысить пороговое значение, если же банк проводит политику по привлечению все новых клиентов необходимо снизить порог отсека. К достоинствам метода стоит отнести автоматическую настройку параметров ФП и соответственно базы правил без привлечения экспертов.

Благодарности

Зайченко Юрию Петровичу за выбор направления исследований и предоставление данных для исследований.

"The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua)."

Литература

[Зайченко, 2008] Зайченко Ю.П. Оценка кредитных банковских рисков с использованием нечеткой логики// Intelligent Information and Engineering Systems. – 2008. - №13 – с. 190-200

[Зайченко, 2008] Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. – К.: Издательский дом «Слово», 2008. – 334с.

[Juang, 1999] CF Juang and CT Lin A recurrent self-organizing neural fuzzy inference network// IEEE transaction of neural networks, Volume 10, Number 4, Pages 828–845, July 1999

Информация об авторе

Шовгун Наталия Виталиевна, аспирантка Национального технического университета Украины «КПИ», адрес электронной почты: shovgun@gmail.com