

Системы нечеткого логического вывода в кредитном скоринге

Е. С. Волкова¹, В. Б. Гисин²

Финансовый университет при Правительстве Российской Федерации

¹evolkova@fa.ru, ²vgisin@fa.ru

Аннотация. Приводится обзор основных методов теории нечетких множеств, применяемых в кредитном скоринге. Основное внимание уделено методам, использующим системы нечеткого логического вывода, которые представляются наиболее перспективными.

Ключевые слова: кредитный скоринг; нечеткий логический вывод; машинное обучение

I. ВВЕДЕНИЕ

Большинство работ, в которых для решения задач кредитного скоринга применяются методы нечеткой логики, посвящено разработке эффективных алгоритмов, их апробации и сравнению с другими алгоритмами. Можно отметить работы, в которых основой является метод опорных векторов, а нечеткие множества играют вспомогательную роль, см. [1], [2], [3]. Особый интерес представляют исследования, в которых задача кредитного скоринга решается на основе нечеткого вывода. Формирование систем правил может проводиться экспертным путем [4, 5, 6, 7], или на основе обучения [8], [9]. Сложности построения систем с обучением связаны с извлечением знаний, направленных на построение правил вывода. Для извлечения знаний используются различного рода сети: нейронные сети, персептроны, карты Кохонена [10–17]. В работах [18–22] получило развитие направление построения систем вывода, в которых обучение происходит с помощью генетических и эволюционных алгоритмов. При сравнении метода опорных векторов и нейронных сетей и в случае использования нечеткости нейронные сети показывают лучший результат. Перспективной представляется разработка комбинированных моделей, в которых методы нечеткой логики сочетаются с классическим методом опорных векторов. Преимущество применения нечетких методов заключается в том, что они позволяют формулировать правила принятия решений эксплицитно, в лингвистических терминах, при этом обеспечивая приемлемую эффективность.

II. НЕЧЕТКИЙ ВЫВОД

Системы нечеткого вывода строятся следующим образом. Пусть X – множество посылок, Y – множество заключений. И пусть $R \subseteq X \times Y$ – нечеткое отношение такое, что $R(x, y)$ – оценка истинности утверждения «из x следует y », где $x \in X$, $y \in Y$. Пусть A – нечеткое множество посылок, B – соответствующее ему множество

заключений. Тогда нечеткое правило вывода «если на входе получено x , на выходе получится y », определяется соотношением

$$B(y) = \max_x (\min(A(x), R(x, y))),$$

где $A(x), B(y)$ – соответствующие функции принадлежности. На основе подобных правил строятся системы нечеткого вывода Мамдани. Система Мамдани – это система правил (R_j) вида

$$(R_j) \text{ ЕСЛИ } (P_{j1}(x_1) \text{ И } \dots \text{ И } P_{jn}(x_n)) \text{ ТО } Q_j(y),$$

где P_{jk} – нечеткий предикат, задающий нечеткое множество A_{jk} , Q_j – нечеткий предикат, задающий нечеткое множество B_j . Если на вход системы поступает набор значений (x_1, \dots, x_n) истинность посылки правила (R_j) вычисляется по формуле $\gamma_j = \min_k A_{jk}(x_k)$. Применение правила (R_j) дает на выходе нечеткое множество нечеткое множество B_j , срезанное на уровне истинности посылки правила и имеющее функцию принадлежности $\min(\gamma_j, B_j(y))$.

После применения всех правил (R_j) , составляющих систему вывода, на выходе получается нечеткое множество с функцией принадлежности $\max_j \min(\gamma_j, B_j(y))$.

Для получения конкретного выходного значения y используется процедура дефазификации. Таким образом, в итоге, по входному набору значений (x_1, \dots, x_n) строится выходное значение $y^* = f(x_1, \dots, x_n)$, функционально зависящее от входа.

Зависимость входных и выходных значений в модели Такаджи–Сугено описывается линейными уравнениями:

$$(R_j) \text{ ЕСЛИ } (P_{j1}(x_1) \text{ И } \dots \text{ И } P_{jn}(x_n)) \text{ ТО } y = f_i(x_1, \dots, x_n),$$

где P_{jk} – нечеткий предикат, задающий нечеткое множество A_{jk} , а f_i – линейная функция.

В моделях кредитного скоринга чаще используются модели Мамдани, которые более приспособлены для категориальных решений.

В зависимости от типа генерирующих правил принято различать два типа нечетких систем: дескриптивные и аппроксимационные. Если в дескриптивных нечетких системах нечеткие множества и их функции принадлежности ассоциированы с терминами, близкими к естественному языку, то в аппроксимационных системах каждое правило использует свои локально определяемые функции принадлежности.

Для обучения нечетких систем и формирования базы знаний используются импульсные алгоритмы, генетические алгоритмы, эволюционные алгоритмы и их комбинацию. Использование импульсных алгоритмов обучения практически не зависит от того, какая строится система: дескриптивная или аппроксимационная. В то же время тип системы существенно влияет на генетические алгоритмы и тип алгоритма оптимизации. В дескриптивной системе число правил вывода конечно, поиск оптимального правила можно рассматривать как решение задачи дискретной оптимизации, для решения которой применяются генетические алгоритмы. В аппроксимационной системе формируются линейные функции, непрерывно зависящие от параметров. Для решения задачи оптимизации в этом случае используются эволюционные стратегии.

III. СИСТЕМЫ НЕЧЕТКИХ ПРАВИЛ В МОДЕЛЯХ КРЕДИТНОГО СКОРИНГА

Дескриптивные правила вывода близки к правилам вывода в системах Такаджи–Сугено и имеют вид:

$$(R_j) \text{ ЕСЛИ } (Q_{j1}(x_1) \text{ И } \dots \text{ И } Q_{jn}(x_n)) \text{ ТО } Class = c_j,$$

где c_j – метка класса (например «Good» для заемщика, которому кредит может быть одобрен). В качестве $Q_{jk}(x_k)$ допускаются дизъюнкции предикатов $P_{jk1}(x_k), \dots, P_{jkl}(x_k)$, которые соответствуют нечетким множествам, выделенным на области допустимых значений переменной x_k . Правилам (R_j) при настройке системы приписывается их значимость в ходе обучения w_j .

Список переменных определяется заранее. При построении систем кредитного скоринга в качестве переменных берутся, например, такие величины, как месячный доход (непрерывная величина), остаток на счете (непрерывная величина), количество иждивенцев (дискретная величина), назначение кредита (номинальная величина), срок (непрерывная величина), возраст (непрерывная величина), профессия (номинальная величина) и т.п.

Величина $R_j(x) = \min_k Q_{jk}(x_k)$ называется величиной активации правила (R_j) на входе x . Каждое правило (R_j) голосует за своего кандидата c_j количеством голосов, равным величине активации правила. Решение (выбор подходящего класса) определяется большинством голосов.

Несмотря на то, что заключительная процедура выглядит достаточно произвольно, во всех работах принимается именно она, или ее модификации. В силу теоремы Эрроу, такая процедура может привести к противоречивым решениям. Избежать противоречий можно, используя процедуры нечеткого коллективного принятия решений [23].

Правило обычно оценивается по трем критериям: частота; степень покрытия; согласованность. Частота определяется как доля примеров, в которых отнесение к указанному классу было выполнено по рассматриваемому правилу, к числу всех примеров с отнесением к этому классу. Степень покрытия оценивает долю примеров, в

которых отнесение к указанному классу было выполнено по рассматриваемому правилу, в обучающей выборке в целом. Согласованность оценивает разницу между правильной и неправильной классификацией по заданному правилу на обучающей выборке. Три указанных критерия сводятся в один агрегированный критерий. Окончательное формирование базы правил происходит путем отбора заданного числа правил по агрегированному критерию.

Аппроксимационные правила имеют более простой вид, чем дескриптивные:

$$(R_j) \text{ ЕСЛИ } (P_{j1}(x_1) \text{ И } \dots \text{ И } P_{jn}(x_n)) \text{ ТО } Class = c_j,$$

где P_{jk} – предикат, который определяет трапецевидное нечеткое множество A_{jk} на множестве значений переменной x_k . Нечеткое множество A_{jk} задается четверкой чисел $(a_{jk}, d_{jk1}, d_{jk2}, d_{jk3})$, $d_{jk1}, d_{jk2}, d_{jk3} \geq 0$, так, что функция принадлежности множества A_{jk} кусочно-линейна, $[a_{jk}; a_{jk} + d_{jk1} + d_{jk2} + d_{jk3}]$ – носитель множества A_{jk} , а $[a_{jk} + d_{jk1}; a_{jk} + d_{jk1} + d_{jk2}]$ – множество модальных значений.

Генетический код, описывающий правило (R_j) , имеет следующую структуру: числовой вектор размерности $4n$ для описания множеств A_{jk} , битовая строка длины n – индикаторы включенности переменных, указатель класса. Пусть z_i , $i = 1, \dots, 4n$, – координаты числового вектора. Они подвергаются мутациям, стандартным для эволюционных алгоритмов. Шаг мутации определяется вектором $(\sigma_i)_{i=1, \dots, 4n}$ (также являющимся составной частью хромосомы). Мутации описываются следующими уравнениями:

$$z_i(t+1) = z_i(t) + N_i(0, \sigma_i);$$

$$\sigma_i(t+1) = \sigma_i(t) \exp(\tau' N_i(0, 1) + \tau N_i(0, 1)).$$

Здесь параметр τ' ускоряет или замедляет общую скорость мутаций, а параметр τ – скорость мутаций одного «гена». Значение σ_i берется достаточно малым, и при $d_{jkl} \leq 0$ мутация не происходит.

IV. МЕТОД ОПОРНЫХ ВЕКТОРОВ И НЕЧЕТКОСТЬ

Первые публикации, посвященные применению метода опорных векторов для решения задач кредитного скоринга, появились около десяти лет назад [24, 25]. В дальнейших исследованиях метод совершенствовался, но оставались проблемы, которые снижали эффективность алгоритмов. В [26] описан нечеткий метод опорных векторов, который вскоре был применен для оценки кредитного риска в [27]. Впоследствии нечеткий метод опорных векторов стал применяться и непосредственно для решения задач кредитного скоринга.

Нечеткий метод опорных векторов состоит в следующем. В отличие от обычного метода нечетких векторов, обучающее множество состоит из троек вида (x_i, y_i, s_i) , $i = 1, 2, \dots, M$, где $x_i \in X \subseteq \mathbf{R}^n$, $y_i \in \{-1; 1\}$, а $s_i \in [0; 1]$ – мера принадлежности заявителя с параметрами x_i классу y_i .

Система кредитного скоринга получается в результате решения оптимизационной задачи:

$$\begin{aligned} \|w\|^2 + C(s_1\gamma_1 + s_2\gamma_2 + \dots + s_M\gamma_M) &\rightarrow \min; \\ y_i(\langle w, x_i \rangle - w_0) &\geq 1 - \gamma_i, i = 1, 2, \dots, M; \\ \gamma_i &\geq 0, i = 1, 2, \dots, M. \end{aligned}$$

В. ЗАКЛЮЧЕНИЕ

В работе приводится обзор мягких вычислений, применяемых в моделях кредитного скоринга. К таким вычислениям относятся вычисления, связанные с нейронными сетями, генетическими алгоритмами и их разновидностями и вычисления, основанные на системах нечеткого вывода. Системы, основанные на нечетком выводе, допускают естественную содержательную интерпретацию. Комбинирование классических методов и мягких вычислений приводит к построению моделей с достаточной эффективностью.

СПИСОК ЛИТЕРАТУРЫ

- [1] Chaudhuri A. Modified fuzzy support vector machine for credit approval classification. // *AI Communications*. 2014. Т. 27, вып. 2. С. 189-211.
- [2] Credit scoring with an improved fuzzy support vector machine based on grey incidence analysis / B. Yi, J. Zhu. // *Proc. Int. Conf. on Grey Systems and Intelligent Services (GSIS)*, 2015 / IEEE, 2015. С. 173-178.
- [3] Shi J., Xu B. Credit Scoring by Fuzzy Support Vector Machines with a Novel Membership Function. // *Journal of Risk and Financial Management*. 2016. Т. 9(4), вып.13. С. 1-10. Available at: doi:10.3390/jrfm9040013.
- [4] On Consumer Credit Scoring Based on Multi-criteria Fuzzy Logic. / Xinhui C., Zhong Q. // *Proc. Int. Conf. Business Intelligence and Financial Engineering*, 2009, BIFE'09 / IEEE, 2009. С. 765-768.
- [5] Lukashevich N. S. The Credit Scoring System for Evaluating Personal Loans Based on the Fuzzy Sets Theory. // *World Applied Sciences Journal*. 2014. Т. 31, вып. 5. С. 840-845.
- [6] Gorchushkina N. N., Shin E. V. Reengineering of business process of crediting and application of fuzzy sets for classification of borrowers in credit scoring. // *Internet Journal of Science*, 2015. Т. 7, вып. 2 (27). С. 1-11. Available at: <http://dx.doi.org/10.15862/82EVN215>
- [7] Mammadli S. Fuzzy Logic Based Loan Evaluation System. // *Procedia Computer Science*. 2016. Т. 102. С. 495-499.
- [8] Lughofer E. *Evolving fuzzy systems-methodologies, advanced concepts and applications*. Berlin: Springer, 2011. 454 c. Available at: doi:10.1007/978-3-642-18087-3
- [9] Bosque G., del Campo I., Echanobe J. Fuzzy systems, neural networks and neuro-fuzzy systems: A vision on their hardware implementation and platforms over two decades. // *Engineering Applications of Artificial Intelligence*. 2014. Т. 32. С. 283-331. Available at: <https://doi.org/10.1016/j.engappai.2014.02.008>
- [10] Piramuthu S. Financial credit-risk evaluation with neural and neurofuzzy systems. // *European Journal of Operational Research*. 1999. Т. 112, вып. 2. С. 310-321.
- [11] Malhotra R., Malhotra D. K. Differentiating between good credits and bad credits using neuro-fuzzy systems. // *European journal of operational research*. 2002. Т. 136, вып. 1. С. 190-211.
- [12] Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. Using neural network rule extraction and decision tables for credit-risk evaluation. // *Management science*. 2003. Т. 49, вып. 3. С. 312-329.
- [13] Baesens, B., Van Gestel T., Viaene S., Stepanova M., Suykens J., & Vanthienen J. Benchmarking state-of-the-art classification algorithms for credit scoring. // *Journal of the operational research society*. 2003. Т. 54, вып. 6. С. 627-635.
- [14] Laha A. Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring. // *Advanced Engineering Informatics*. 2007. Т. 21, вып. 3. С. 281-291.
- [15] Jiao Y., Syau Y. R., Lee E. S. Modelling credit rating by fuzzy adaptive network. // *Mathematical and Computer Modelling*. 2007. Т. 45, вып. 5. С. 717-731.
- [16] Khashei M., Rezvan M. T., Hamadani A. Z., & Bijari M. A bilevel neural-based fuzzy classification approach for credit scoring problems. // *Complexity*. 2013. Т. 18, вып. 6. С. 46-57.
- [17] Intuitionistic Fuzzy Neural Network: The Case of Credit Scoring Using Text Information. / P. Hájek, V. Olej // *Engineering Applications of Neural Networks / Springer International Publishing*, 2015. С. 337-346.
- [18] Boosting a genetic fuzzy classifier / F. Hoffmann // *Proc. Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, 2001. / IEEE, 2001. Т. 3. С. 1564-1569.
- [19] Hoffmann F., Baesens B., Martens J., Put F., Vanthienen J. Comparing a genetic fuzzy and a neurofuzzy classifier for credit scoring. // *International Journal of Intelligent Systems*. 2002. Т. 17, вып. 11. С. 1067-1083.
- [20] Hoffmann F. Combining boosting and evolutionary algorithms for learning of fuzzy classification rules. // *Fuzzy Sets and Systems*. 2004. Т. 141, вып. 1. С. 47-58.
- [21] Hoffmann, F., Baesens, B., Mues, C., Van Gestel, T., & Vanthienen, J. Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. // *European Journal of Operational Research*. 2007. Т. 177, вып. 1. С. 540-555.
- [22] Del Jesus M. J., Hoffmann F., Navascués L. J., Sánchez L. Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms. // *IEEE Transactions on Fuzzy Systems*, 2004. Т. 12, вып. 3. С. 296-308.
- [23] Gibilisco M., M Gowen A., Albert K., Mordeson J., Wierman M., Clark T. *Fuzzy social choice theory*. NY: Springer, 2014. 186 c.
- [24] Van Gestel, T., Baesens, B., Suykens, J. A., Van den Poel, D., Baestaens, D. E., & Willekens, Bayesian kernel based classification for financial distress detection. // *European journal of operational research*, 2006. Т. 172, вып. 3. С. 979-1003.
- [25] Xiao W., Fei Q. A Study of Personal Credit Scoring Models on Support Vector Machine with Optimal Choice of Kernel Function Parameters. // *Systems Engineering-Theory & Practice*. 2006. Т. 26, вып. 10. С. 73-79.
- [26] Lin C. F., Wang S. D. Fuzzy support vector machines. // *IEEE transactions on neural networks*. 2002. Т. 13, вып. 2. С. 464-471.
- Wang Y., Wang S., Lai K. K. A new fuzzy support vector machine to evaluate credit risk. // *IEEE Transactions on Fuzzy Systems*. 2005. Т. 13, вып. 6. С. 820-831.