

Деревья решений для индукции лингвистических правил

А. А. Олейник¹, Е. А. Гофман²

Анотация – The problem of linguistic rules induction is considered. A method of decision trees synthesis linguistic rules induction is suggested.

Ключевые слова – деревья решений, искусственный интеллект, CART, индукция правил, лингвистические термины.

I. ВВЕДЕНИЕ

В настоящее время экспертные системы, основанные на лингвистических правилах [1], успешно применяются в различных прикладных областях, связанных с медицинским и техническим диагностированием, финансовым менеджментом, распознаванием образов, управлением компьютерными сетями и технологическими процессами и т.п. Поскольку базы правил в экспертных системах часто характеризуются большим объемом, то актуальной является задача индукции правил, суть которой заключается в том, чтобы на основе начального набора правил сформировать новую базу правил меньшего объема, которая в достаточной мере представляла бы начальную базу и была бы менее избыточной.

Существуют различные методы индукции правил [2, 3], которые, однако, предполагают анализ качества правил по отдельности, не рассматривая и не учитывая качество всей базы в целом, что иногда приводит к получению неприемлемых баз нечетких правил. Поэтому актуальной является разработка новых методов индукции правил, которые учитывали бы качество всей базы знаний, а не только отдельных правил. Для решения данной задачи предлагается создавать деревья решения, которые бы после их построения переводились в лингвистические правила. Выбор деревьев решений обосновывается их возможностью выявлять ненаблюдаемые связи внутри рассматриваемых процессов и систем.

II. ДЕРЕВЬЯ РЕШЕНИЙ ДЛЯ ИНДУКЦИИ ПРАВИЛ

Существующие методы построения деревьев решений [4] не учитывают особенностей рассматриваемой задачи. В связи с этим предлагается новый метод построения деревьев решений для индукции правил.

Как в базовом методе построения деревьев решений, так и в разрабатываемом методе идентификации деревьев решений основным этапом является разбиение дерева (рост дерева). С целью разбиения дерева решений для каждого возможного разбиения рассчитывается оценочная функция:

$$Q(f, v) = p_{f,v} g(p_{f,v}^1) + (1 - p_{f,v}) g(p_{f,v}^2), \quad (1)$$

где $p_{f,v}$ – вероятность того, что x относится к терму v ; $g(p)$ – модифицированная энтропия для вероятности отнесения выходной переменной y к рассматриваемому классу при условии, что x относится или не относится к терму v ($p_{f,v}^1$ и $p_{f,v}^2$, соответственно):

$$g(p) = -r(p) \ln(r(p)) - (1 - r(p)) \ln(1 - r(p)), \quad (2)$$

где $r(p)$ преобразует оценку вероятности:

$$r(p) = \begin{cases} \frac{1}{2}(1 + \sqrt{2p - 1}), & \text{если } p > 0,5; \\ \frac{1}{2}(1 - \sqrt{1 - 2p}), & \text{если } p < 0,5. \end{cases} \quad (3)$$

Таким образом, чем ближе значение вероятности к 0,5, тем выше модифицированное значение, а чем дальше от 0,5, тем значение ниже.

Оценочная функция рассчитывается для всех возможных разбиений и выбирается разбиение с наименьшим значением оценочной функции. Разбиение начинается от корневого узла и продолжается до тех пор, пока не возникнет ситуация, когда невозможно произвести новое разбиение.

III. ВЫВОДЫ

Разработанный метод идентификации деревьев решений для индукции нечетких правил позволяет выполнять преобразование и объединение лингвистических правил, что обеспечивает возможность разработки экспертных систем на основании более логических прозрачных и простых баз лингвистических правил.

Научная новизна работы заключается в том, что разработан новый метод построения деревьев решений, позволяющий выполнять индукцию лингвистических правил, что достигается за счет введения дополнительных функций преобразования.

ПЕРЕЧЕНЬ ССЫЛОК

- [1] T. Terano, K. Asai, M. Sugeno, "Fuzzy Systems Theory and Its Applications," Academic Press, San Diego, 1992, 268 p.
- [2] R.E. Bellman, L.A. Zadeh, "Decision-Making in Fuzzy Environment," *Management Science*, vol. 4, pp. 141–160, 1970.
- [3] Y. Yuan, M. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets and Systems*, vol. 69, pp. 125–139, 1995.
- [4] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, "Classification and regression trees," Wadsworth & Brooks, California, 1984, 368 p.

^{1,2}Запорожский национальный технический университет, ул. Жуковского, 64, Запорожье, 69063, УКРАИНА, E-mail: olejnikaa@gmail.com