

УДК 004.8

ВСТРОЕННЫЙ МЕТОД ОТБОРА ИНФОРМАТИВНЫХ ПРИЗНАКОВ НА ОСНОВЕ ДВУХЭТАПНОЙ СХЕМЫ ОБУЧЕНИЯ НЕЙРОСЕТЕВОГО НЕЧЕТКОГО КЛАССИФИКАТОРА

© Новоселова Н.А., Том И.Э.

Объединенный институт проблем информатики НАН Беларуси
ул. Сурганова 6, г. Минск, Беларусь, 220012

E-MAIL: tom@newman.bas-net.by

Abstract. The paper describes the embedded method of informative feature selection, based on two-stage training scheme of neural-fuzzy classifier. The method provides the possibility of separate feature set selection for every class of data, retaining the maximal classification accuracy, which is attained by applying modified technique of classifier parameter tuning.

ВВЕДЕНИЕ

Для решения классификационных, регрессионных задач в медицинской области и в других сферах, где необходимо построение не только достаточно точной модели, но и обеспечение ее интерпретируемости, т.е. объяснение полученного решения, стали широко применяться нейросетевые нечеткие классификационные и регрессионные модели. Необходимость реализации автоматического построения нейросетевых нечетких моделей на основе имеющегося набора экспериментальных данных обусловила *постановку следующей проблемы*: разработать алгоритмы предобработки данных и обучения модели, которые позволяют не только настроить ее параметры, но и определить структуру модели. Основная идея предобработки состоит в сокращении избыточных признаков и признаков, которые не несут существенной информации для решения конкретной классификационной задачи. Извлечение как можно большего количества информации из имеющихся данных при использовании меньшего количества признаков способствует не только сокращению вычислительных затрат, но и построению модели решаемой задачи, которая имеет более высокую обобщающую способность. *Согласно проведенному анализу последних достижений и публикаций* все существующие методы отбора информативных признаков подразделяются на три категории: упаковочные, фильтровочные и встроенные [1]–[4], основное отличие которых заключается в способе оценки отобранного подмножества признаков. Особый интерес представляют собой встроенные методы, в которых процесс отбора признаков осуществляется внутри самого обучающего алгоритма. Например, методы основанные на разбиениях многомерного признакового пространства и методы по принципу «разделяй и властвуй» неявно отбирают наиболее релевантные решаемой задаче признаки для включения их в поддерево решений.

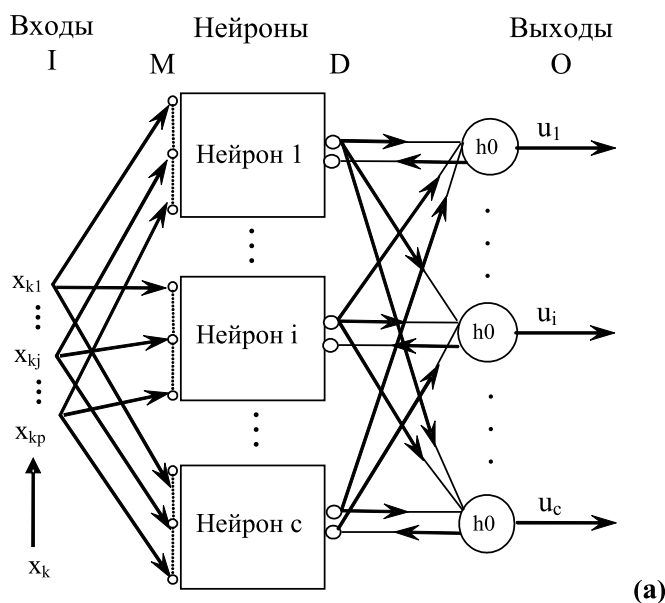
До настоящего времени в алгоритмах обучения нейросетевых нечетких моделей не реализована возможность отбора признаков, которая позволила бы динамически формировать структуру модели на основе обучающего набора входных-выходных данных. В результате разработки такого алгоритма возможным является

одновременное осуществление отбора наиболее информативных признаков и получение высокой точности классификации или прогнозирования (в зависимости от решаемой задачи).

Целью настоящей работы является разработка встроенного метода отбора информативных признаков на основе двухэтапной схемы обучения нейросетевого нечеткого классификатора (детектора). Метод обеспечивает возможность отбора отдельных подмножеств признаков для каждого из классов данных при сохранении максимальной точности классификации, что достигается за счет использования модифицированного способа настройки параметров, базирующегося на алгоритме пропорционального векторного квантования [5].

1. НЕЙРОСЕТЕВОЙ НЕЧЕТКИЙ КЛАССИФИКАТОР

Нейросетевой нечеткий классификатор, называемый далее для краткости нечетким детектором, изображен на рисунке 1. Нечеткий детектор состоит из четырех слоев: входного слоя (I), количество элементов которого равно размерности p входного пространства признаков; слоя сравнения (M), определяющего расстояния между входным вектором $x_k = (x_{k1}, x_{k2}, \dots, x_{kp})$, ($k = 1, \dots, N$; N – количество объектов данных) и вектором m_i ($i = 1, \dots, c$; c – количество кластеров или нейронов выходного слоя), который задает центр кластера, соответствующего классу и называется вектором памяти; детекторного слоя (D), количество элементов которого равно количеству классов или кластеров в данных; выходного слоя (O), который рассчитывает нормализованные значения принадлежности входного вектора данных к каждому из классов. Элементы детекторного слоя соединены двунаправленными связями с входным и выходным слоями, а его весовые коэффициенты ω_{ij} определяют степень важности признака j для класса i .



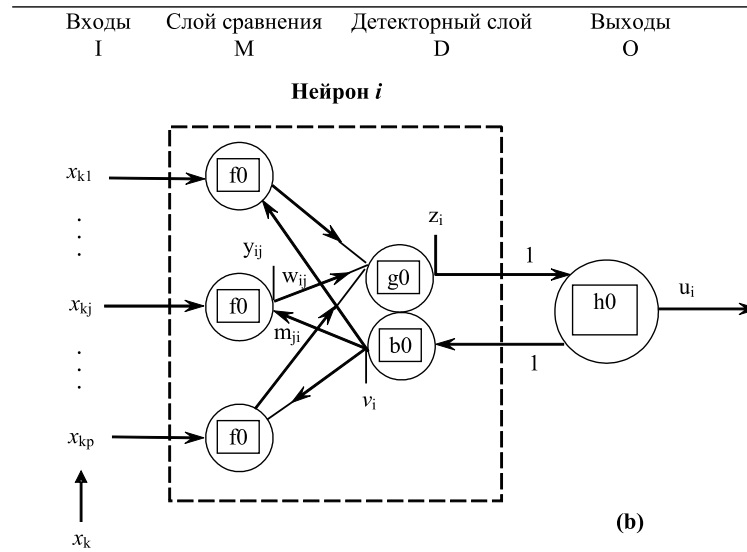


Рис. 1. (а) схематическое нейросетевое представление нечеткого детектора; (б) структура и соединения i -го нейрона

В детекторном слое имеется два типа узлов: прямые и обратные. Каждый прямой узел получает на вход значения от p узлов слоя сравнения M , связанных с ним посредством взвешенных соединений $\{\omega_{ij}\}$. Каждый обратный узел получает на вход значение от узлов выходного слоя O , соединенного с детекторным слоем соединениями с фиксированным весом равным единице. Выходной слой O позволяет получить окончательную классификацию каждого входного вектора путем нормализации выходных значений всех узлов детекторного слоя D .

Таким образом, для входного вектора промежуточные выходы отдельных нейронных слоев (см. рисунок 1) рассчитываются следующим образом:

$$y_{ij} = (x_{kj} - m_{ji}), \quad j = 1, \dots, p,$$

$$z_i = \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^p \omega_{ij}^2 (x_{kj} - m_{ji})^2 \right], \quad (1.1)$$

$$u_i = z_i / \sum_{j=1}^c z_j. \quad (1.2)$$

Обучение нечеткого детектора осуществляется в два этапа, позволяя параллельно с настройкой центров кластеров, реализовать отбор информативных признаков для классификации.

На первом этапе осуществляется настройка векторов памяти m_i . Настройка векторов памяти выполняется с использованием алгоритма обучения без учителя, а именно, алгоритма пропорционального векторного квантования [5]:

$$\Delta m_i = \alpha_t u_i(x_k)(x_k - m_i), \quad (1.3)$$

где x_k представляет собой k -й входной вектор.

На втором этапе обучаются веса ω_{ij} , представляющие собой степень влияния признака j на определение кластера i . Для того чтобы найти правило для настройки ω_{ij} вводится следующая функция ошибки:

$$E = \frac{1}{2} \sum_{k=1}^N \sum_{i=1}^c (u_i(x_k) - d_i)^2, \quad (1.4)$$

где d_i – желаемое значение выхода i -го узла выходного слоя. Таким образом, весовые коэффициенты настраиваются с использованием алгоритма обучения с учителем. На основе правила дифференцирования по цепочке и с использованием выражений (1.1), (1.2), (1.4) для настройки ω_{ij} используется следующее выражение:

$$\Delta\omega_{ij} = \frac{\beta}{\sigma^2 s^2} (u_i(x_k) - d_i) \times \left(\sum_{l=1}^c z_l - z_i \right) \cdot \omega_{ij} z_i (x_{kj} - m_{ji})^2, \quad (1.5)$$

где $\beta > 0$ – коэффициент обучения. Значение $\omega_{ij} = 0$ означает, что j -й признак не является информативным для определения i -го кластера; значение $\omega_{ij} = 1$ означает, что j -й признак наиболее информативный для определения i -го кластера.

Формальное описание алгоритма обучения представляется следующим образом:

1. Зафиксировать $\sigma > 0, \alpha \in [0, 1], \beta > 0, \varepsilon > 0$ и T – максимальное количество итераций.
2. Инициализировать $\{m_i(0)\}$, используя c случайным образом выбранных объекта данных из $\{x_k\} (k = 1, \dots, N)$, и $\omega_{ij}(0) = 1$ для любого i и j .
3. Для $t = 1, \dots, T$; для $k = 1, \dots, N$
 - a) вычислить $\{u_i\}$ используя (1.2),
 - b) вычислить $\{m_i(t)\}$ используя (1.3),
 - c) модифицировать $\{\omega_{ij}(t)\}$ используя (1.5),
 - d) перейти к следующему шагу $k + 1$.
4. Вычислить E по формуле (1.4).
5. Если $E < \varepsilon$ или $t > T$, то остановить алгоритм, иначе перейти к следующему шагу $t + 1$.

Таким образом, согласно предлагаемому и описанному выше алгоритму, настройка весовых коэффициентов ω_{ij} осуществляется параллельно с настройкой векторов памяти m_i с использованием обучающего набора данных. В ходе реализации алгоритма обучения осуществляется отбор наиболее информативных признаков для каждого из классов объектов данных, что позволяет выделить подмножество признаков для построения окончательного варианта классификатора. В дальнейшем классификатор может быть представлен в виде набора нечетких правил, где функции принадлежности нечетких множеств представлены гауссовыми функциями вида (1.1).

2. ЭКСПЕРИМЕНТАЛЬНАЯ ПРОВЕРКА ПРЕДЛОЖЕННОГО НЕЙРОСЕТЕВОГО НЕЧЕТКОГО КЛАССИФИКАТОРА

Предложенный метод отбора информативных признаков с одновременным построением классификатора протестирован на наборе данных *Iris* из международного архива данных по машинному обучению. Набор данных состоит из 150 объектов, каждый из которых описан значениями 4-х признаков с непрерывными значениями: длина чашелистика (x_1), ширина чашелистика (x_2), длина лепестка (x_3), ширина лепестка (x_4). Объекты данных относятся к трем классам: тип сестон (s_1), цветоизменяющий (s_2) и виргинский (s_3) типы.

В таблице 1 приведены результаты двух экспериментов. В первом эксперименте для построения нейросетевого нечеткого классификатора были использованы все 4 признака, характеризующие объекты данных *Iris*. Из сравнения полученных результатов следует, что признаки x_1 и x_2 не влияют на формирование классов s_2 и s_3 ; для формирования класса s_1 все признаки информативны. Во втором эксперименте проверяется, насколько изменится точность классификатора, если при его построении использовать только признаки x_3 и x_4 , имеющие наибольшие значения весовых коэффициентов для большинства классов. Результаты, полученные во втором эксперименте, представлены в третьем столбце таблицы 1. Очевидно, что точность классификации при использовании двух признаков не изменилась, таким образом, эти признаки являются наиболее информативными для классификации, а признаками x_1 и x_2 можно пренебречь.

Таблица 1. Результаты классификации и отбора признаков путем построения нечеткого детектора

	Эксперимент 1 (признаки x_1, x_2, x_3, x_4)	Эксперимент 2 (признаки x_3, x_4)
Ошибка классификации	5	5
Степень важности признаков для класса s_1	(1.00, 1.00, 0.95, 1.00)	(1.00, 1.00)
Степень важности признаков для класса s_2	(0.00, 0.00, 1.00, 1.00)	(1.00, 1.00)
Степень важности признаков для класса s_3	(0.00, 0.00, 0.82, 0.97)	(0.82, 0.97)

Конечный набор нечетких правил, полученный путем построения нейросетевого нечеткого классификатора для всего набора данных *Iris* с использованием значений только двух признаков x_3 и x_4 , следующий:

R_1 : если *длина лепестка* «около 1.46» и *ширина лепестка* «около 0.25», то класс s_1 .

R_2 : если *длина лепестка* «около 4.29» и *ширина лепестка* «около 1.36», то класс s_2 .

R_3 : если *длина лепестка* «около 5.54» и *ширина лепестка* «около 2.00», то класс s_3 .

Нечеткие числа «около 1.46», «около 0.25» и т. д. представляют собой нечеткие множества, которые заданы гауссовыми функциями принадлежности с использованием выражения (1.1).

Проведен также сравнительный анализ результатов классификации набора данных *Iris* с другими семью известными в литературе методами. В таблице 2 показаны точность классификации и количество используемых правил для классификаторов, построенных другими методами, в том числе с использованием встроенного метода отбора признаков, предлагаемого в настоящей работе. Согласно таблице 2 с использованием предложенного нами метода отбора признаков с одновременным обучением нечеткого детектора была получена точность классификации, превышающая или сравнимая с другими рассмотренными методами за исключением классификаторов **FuGeNeSys** [6] и **SuPFuNIS** [7]. При этом необходимо отметить, что все классификаторы за исключением **NFS** при построении наборов правил использовали четыре признака, тогда как построенный нами классификатор является самым компактным – состоит всего из трех правил с двумя признаками.

Таблица 2. Сравнительный анализ точности и компактности классификаторов

Метод	Количество правил	Количество признаков	Точность классификации (%)
FuGeNeSys [6]	5	4	100
SuPFuNIS [7]	5	4	100
NEFCLASS [8]	7	4	96.7
ReFuNN [9]	9	4	95.3
EFuNN [10]	17	4	95.3
FeNe-I [11]	7	4	96.0
NFS [12]	5	2	96.7
Нечеткий детектор	3	2	96.7

ЗАКЛЮЧЕНИЕ

В статье рассмотрен разработанный авторами вариант нейросетевого нечеткого классификатора (детектора) со специальными соединениями нейронов для запоминания образов классификации и для определения весов каждого из признаков. Основным результатом данной статьи является новый встроенный метод выделения наиболее информативных признаков для классификации. Отличительной положительной особенностью разработанного метода является возможность отбора отдельных подмножеств признаков для каждого из классов данных, что достигается за счет использования модифицированного способа обучения нейросетевого нечеткого классификатора, базирующегося на алгоритме пропорционального векторного квантования [5]. Модификация заключается в использовании двухэтапной процедуры обучения: на первом этапе настраиваются центры классов с использованием алгоритма пропорционального векторного квантования, на втором этапе настраиваются весовые коэффициенты отдельных признаков с использованием градиентного алгоритма на основе расчета среднеквадратичной ошибки классификации. Сравнение предложенного метода отбора признаков при построении классификатора многомерных данных

с наиболее близким по свойствам встроенным методом D.Chakraborty и N.Pal [12] показывает, что удастся повысить компактность набора нечетких классифицирующих правил при сохранении точности классификации объектов данных на уровне 96,7% для набора данных *Iris* из международного архива данных по машинному обучению <http://joc.pubs.informs.org/Supplements/Lee/iris.data>. Был проведен также сравнительный анализ предложенного метода с семью наиболее известными методами классификации многомерных данных, который показал преимущество нового метода в части компактности информативного множества признаков. Метод выделения информативных признаков и соответствующий алгоритм обучения нечеткого детектора реализованы в виде экспериментального варианта программного модуля **FDetector**, который будет использован в разрабатываемой авторами системе поддержки принятия решения медицинского назначения.

СПИСОК ЛИТЕРАТУРЫ

1. *Das S.* Filters, wrappers and a boosting-based hybrid for feature selection // In Proc. 18th Int. Conf. on Machine Learning. – Morgan Kaufmann Publishers Inc., USA. – 2001. – P. 74-81.
2. *Kohavi R., John G.* Wrappers for feature subset selection // Artificial Intelligence. – 1997. – №1-2. – P. 273-324.
3. *Blum A.I., Langley P.* Selection of relevant features and examples in machine learning // Artificial Intelligence. – 1997. – Vol.97, №1-2. – P. 245-271.
4. *Kohavi R., John G., Pfleger K.* Irrelevant features and the subset selection problem // In Proc. 11th Intl. Conf. on Machine Learning. – New Brunswick, NJ – 1994. – P. 121-129.
5. *Li R.-P., Mukaidono M.* Proportional learning vector quantization // Journal of Japan Society for Fuzzy Theory and Systems. – 1998. – Vol.10, №6. – P. 1129-1134.
6. *Russo M.* FuGeNeSys - a fuzzy genetic neural system for fuzzy modeling // IEEE Transactions on Fuzzy Systems. – 1998. – Vol.6. – P. 373-387.
7. *Kumar Paul, S.* Subsethood-product fuzzy neural inference system // IEEE Transactions on Neural Networks. – 2002. – Vol.13. – P. 578-599.
8. *Nauck D., Kruse R.* A neuro-fuzzy method to learn fuzzy classification rules from data // Fuzzy Sets and Systems. – 1997. – Vol.89. – P. 277-288.
9. *Kasabov N.* Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems // Fuzzy Sets and Systems. – 1996. – Vol.82. – P. 135-149.
10. *Kasabov N., Woodford B.* Rule insertion and rule extraction from evolving fuzzy neural networks: algorithms and application for building adaptive, intelligent expert systems // In Proc. IEEE International Conference on Fuzzy Systems FUZIEEE 99. – Seoul, Korea – 1999. – Vol.3. – P. 1406-1411.
11. *Halgamuge S., Glesner M.* Neural networks in designing fuzzy systems for real world applications // Fuzzy Sets and Systems. – 1994. – Vol.65. – P. 1-12.
12. *Chakraborty D., Pal N. R.* A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification // IEEE Transactions on Neural Networks. – 2004. – Vol.15, №6. – P. 110-123.

Статья поступила в редакцию 27.04.2008