

Е.В. Бодянский, Е.А. Винокурова, Д.Д. Пелешко

Информационная технология кластеризации данных в условиях короткой обучающей выборки на основе ассоциативной вероятностной нейро-фаззи системы

Разработана гибридная нейро-фаззи система вычислительного интеллекта, позволяющая решать задачи вероятностной и нечеткой классификации данных, описываемых короткими выборками, в условиях пересекающихся классов произвольной формы.

The hybrid neuro-fuzzy system of computational intelligence, which allows to solve tasks of probabilistic and fuzzy data classification with short training set under overlapping arbitrary form classes is realized.

Розроблено гібридну нейро-фаззі систему обчислювального інтелекту, яка дозволяє розв'язувати задачі ймовірнісної і нечіткої класифікації даних, описані короткими вибірками, за умов перетинних класів довільної форми.

Введение. Задача классификации данных на основе парадигмы обучения с учителем – одна из основных в рамках научного направления, известного как интеллектуальный анализ данных (*Data Mining*), а для ее решения предложено множество подходов и алгоритмов [1–3] от сугубо эмпирических до строго математических. При этом *классические* алгоритмы классификации-распознавания образов опираются, как правило, на гипотезу выпуклости формируемых классов и их линейную делимость [4]. Поскольку при работе с реальными данными это предположение довольно часто не выполняется и классы обычно не выпуклые, традиционные методы теряют свою эффективность, а на первый план выступают искусственные нейронные сети благодаря прежде всего своим универсальным аппроксимирующим возможностям [5–9]. При этом нейронные сети также не панацея, поскольку нуждаются для своей настройки в больших объемах обучающих выборок в случае, если процесс обучения основывается на тех или иных оптимизационных процедурах. В ситуациях, когда объем обучающей выборки соизмерим с размерностью векторов признаков, обучение, основанное на памяти, не позволяет обеспечить нейронную сеть необходимым качеством классификации. В этом случае более предпочтительным представляется обучение, основанное на памяти и реализующее принцип «Нейроны в точках данных» [10, 11]. Одним из наиболее характерных образцов таких сетей считается вероятностная нейронная сеть, предложенная

в [12, 13]. Сеть подтвердила свою эффективность при решении целого ряда реальных задач классификации, диагностики, распознавания образов [5, 7, 8] благодаря прежде всего простоте обучения и высокому быстродействию. Вместе с тем обычная *PNN*, впрочем как и любая другая нейронная сеть, решает задачу классификации в рамках традиционного «четкого» подхода, когда каждый из анализируемых векторов признаков может быть отнесен только к одному единственному из возможных классов. Однако во множестве реальных задач, связанных с обработкой данных химических и физических экспериментов, медико-биологической информации, текстов и изображений, классифицируемый вектор признаков с различными уровнями вероятности или принадлежности может быть отнесен сразу к нескольким классам.

Для решения подобного типа задач наиболее эффективным представляется использование гибридных систем вычислительного интеллекта [14, 17] и, прежде всего, нейро-фаззи систем [18] и методов нечеткой классификации и кластеризации [19, 20]. При этом необходимо отметить, что как обучение нейро-фаззи систем, так и нечеткая классификация-кластеризация, базируются на тех или иных оптимизационных процедурах, что также выдвигает требование к достаточному объему обучающих выборок.

В связи с этим представляется целесообразной разработка гибридной нейро-фаззи системы вычислительного интеллекта, позволяющей решать задачи вероятностной и нечеткой клас-

сификации данных, описываемых короткими выборками в условиях пересекающихся классов произвольной формы.

Ассоциативная вероятностная нейро-фаззи система

В основу предлагаемой системы положена вероятностная нейронная сеть Шпехта и ее модификации, введенные в [21–23], а на рис. 1 приведена архитектура этой системы. Как известно, стандартная *PNN* решает задачу Байесовской классификации [4] путем восстановления плотностей распределения данных в каждом классе на основе Парзеновских ядерных оценок [24], а ее архитектура практически совпадает с радиально-базисной нейронной сетью (*RBFN*) [25] и обобщенной регрессионной нейронной сетью (*GRNN*) [26]. И хотя методы классификации по Байесу известны давно, их нейросетевая реализация позволила существенно упростить численную реализацию и повысить скорость решения задачи распознавания образов, классификации, диагностики и др. При этом следует отметить, что каждый предъявленный *PNN*-образ может быть отнесен только к одному классу с наибольшей плотностью распределения данных в окрестности этого образа.

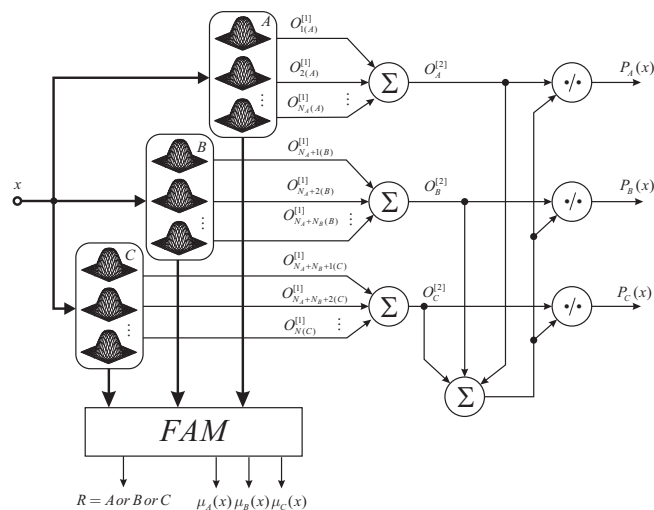


Рис. 1. Ассоциативная вероятностная нейро-фаззи система

Предлагаемая система состоит из двух частей: четырехслойной архитектуры с прямой передачей информации, решающей задачу вероятностной классификации, и нечеткой ассоциативной памяти (*FAM*) [27–29], в основу ко-

торой положена автоассоциативная эволюционирующая память на основе нечетких базисных функций [30], решающей задачу нечеткой классификации.

Часть рассматриваемой архитектуры, ответственная за вероятностную классификацию, содержит четыре последовательных слоя. *Первый скрытый слой* – это слой образов, полностью совпадающий с первым скрытым слоем *PNN* и *GRNN*, *второй скрытый слой* – слой локальных сумматоров, *третий скрытый слой* содержит единственный сумматор, общий для всех нейронов сети и, наконец, *выходной слой* образован делителями, число которых равно числу классов в обучающей выборке.

Исходной информацией для синтеза классифицирующей нейро-фаззи системы служит обучающая выборка многомерных наблюдений $x(1), x(2), \dots, x(N)$, образованная n -мерными векторами признаков с известной классификацией, при этом место конкретного образа в исходном наборе значения не имеет, т.е. фактически исходная информация задается таблицей *объект-свойство*, принятой в стандартном *DataMining*. При этом предполагается, что исходная информация предварительно обработана, чаще всего путем покомпонентного центрирования относительно среднего и последующего нормирования на гиперсферу

$$\|x(j)\| = 1, \quad \forall j = 1, 2, \dots, N. \quad (1)$$

В дальнейшем без потери общности будем полагать, что в обучающей выборке приведены представители трех классов *A*, *B* и *C*. Предполагается также, что в выборке N_A наблюдений относится к классу *A*, N_B наблюдений – к классу *B* и N_C наблюдений – к классу *C*, т.е.

$$N_A + N_B + N_C = N.$$

Следуя концепции, нейроны в точках данных, число нейронов в слое образов составляет N , а каждый блок этого слоя, обозначенный *A*, *B* и *C*, содержит N_A , N_B , N_C нейронов соответственно. При этом поступающие на обработку наблюдения $x(j)$ определяют координаты

наты центров w_j ядерных активационных функций так, что

$$w_{ji} = x_i(j); j = 1, 2, \dots, N; i = 1, 2, \dots, n,$$

или в векторной форме

$$w_j = x(j) = (x_1(j), x_2(j), \dots, x_n(j))^T.$$

Одновременно с установкой центров активационных функций слоя образов в нечеткой ассоциативной памяти вычисляются центры сформированных классов

$$w_R = \frac{1}{N_R} \sum_{l=1}^{N_R} x(l); x(l) \in R; R = A \text{ или } B \text{ или } C$$

и их радиусы

$$r_R = \max_l \|x(l) - w_R\|; x(l) \in R.$$

На этом обучение нейро-фаззи системы заканчивается.

При подаче на вход системы неклассифицированного наблюдения $x(k)$, $k = N+1, N+2, \dots$, на выходах нейронов слоя образов появляются сигналы

$$o_j^{[1]}(k) = \exp\left(-\frac{\|x(k) - w_j\|^2}{2\sigma^2}\right), \quad (2)$$

где $j = 1(A), 2(A), \dots, N_A(A), N_A+1(B), \dots, N_A+N_B(B), N_A+N_B+1(C), \dots, N(C)$, σ – параметр ширины активационной функции.

Заметим, что в силу условия (1), выражение (2) может быть представлено в более простой форме

$$o_j^{[1]}(k) = \exp\left(\frac{w_j^T x(k) - 1}{\sigma^2}\right) = \exp\left(\frac{\cos(w_j, x(k)) - 1}{\sigma^2}\right),$$

а поскольку $-1 \leq \cos(w_j, x(k)) \leq 1$, значения выходных сигналов первого скрытого слоя определяются неравенствами

$$-\frac{2}{\sigma^2} \leq \frac{\cos(w_j, x(k)) - 1}{\sigma^2} \leq 0, \\ \exp(-2\sigma^{-2}) \leq o_j^{[1]}(k) \leq 1.$$

Во втором скрытом слое локальных сумматоров (по одному на каждый класс) вычисляются суммы выходных сигналов первого скрытого слоя

$$\begin{cases} 0 < o_A^{[2]}(k) = \sum_{l=1(A)}^{N_A(A)} o_l^{[1]}(k) < N_A, \\ 0 < o_B^{[2]}(k) = \sum_{l=N_A+1(B)}^{N_A+N_B(B)} o_l^{[1]}(k) < N_B, \\ 0 < o_C^{[2]}(k) = \sum_{l=N_A+N_B+1(C)}^{N(C)} o_l^{[1]}(k) < N_C, \end{cases}$$

которые затем подаются на общий сумматор третьего скрытого слоя, вычисляющий скалярное значение $O_A^{[2]}(k) + O_B^{[2]}(k) + O_C^{[2]}(k)$, поступающее далее на входы делителей выходного слоя. Именно в выходном (четвертом) слое системы вычисляются вероятности принадлежности образа $x(k)$ к конкретному классу

$$\begin{cases} 0 \leq P_A(x(k)) = \frac{O_A^{[2]}(k)}{O_A^{[2]}(k) + O_B^{[2]}(k) + O_C^{[2]}(k)} \leq 1, \\ 0 \leq P_B(x(k)) = \frac{O_B^{[2]}(k)}{O_A^{[2]}(k) + O_B^{[2]}(k) + O_C^{[2]}(k)} \leq 1, \\ 0 \leq P_C(x(k)) = \frac{O_C^{[2]}(k)}{O_A^{[2]}(k) + O_B^{[2]}(k) + O_C^{[2]}(k)} \leq 1. \end{cases} \quad (3)$$

Одновременно с вычислением вероятностей (3) *FAM* вычисляет уровни принадлежности [19, 20] в нечетком смысле того же образа тем же классам:

$$\begin{cases} 0 \leq \mu_A(x(k)) = \frac{\|x(k) - w_A\|^{-2}}{\|x(k) - w_A\|^{-2} + \|x(k) - w_B\|^{-2} + \|x(k) - w_C\|^{-2}} \leq 1, \\ 0 \leq \mu_B(x(k)) = \frac{\|x(k) - w_B\|^{-2}}{\|x(k) - w_A\|^{-2} + \|x(k) - w_B\|^{-2} + \|x(k) - w_C\|^{-2}} \leq 1, \\ 0 \leq \mu_C(x(k)) = \frac{\|x(k) - w_C\|^{-2}}{\|x(k) - w_A\|^{-2} + \|x(k) - w_B\|^{-2} + \|x(k) - w_C\|^{-2}} \leq 1. \end{cases} \quad (4)$$

При этом на выходе *FAM* появляется ассоциированный с $x(k)$ конкретный класс R с максимальным уровнем принадлежности $\mu_R(x(k))$ и все значения рассчитанных принадлежностей.

Рассмотренная система содержит единственный свободный параметр ширины активационных функций σ . Этот параметр выбирается либо из сугубо эмпирических соображений в интервале от нуля до единицы, либо может быть рассчитан аналитически [21], либо настраивается с помощью рекуррентной оптимизационной

процедуры [23], что увеличивает требуемый объем обучающей выборки.

Заключение. Предложенная классифицирующая нейро-фаззи система, построенная на основе вероятностной нейронной сети и нечеткой ассоциативной памяти, предназначена для работы в условиях пересекающихся невыпуклых классов и короткой обучающей выборки. Особенность этой системы – высокое быстродействие и простота численной реализации. Имитационное моделирование подтверждает эффективность предложенного подхода.

1. *Han J., Kamber M.* Data Mining: Concepts and Techniques. – Amsterdam: Morgan Kaufman Publ., 2006. – 743 p.
2. *Zaki M.J., Meira Jr. W.M.* Data Mining and Analysis: Fundamental Concepts and Algorithms. – Cambridge University Press, 2014. – 624 p.
3. *Witten I.H., Frank E., Hall M.A.* Data Mining: Practical Machine Learning Tools and Techniques. – Amsterdam: Morgan Kaufmann Publ., 2011. – 664 p.
4. *Fukunaga K.* Introduction to Statistical Pattern Recognition. – N.Y.: Academic Press, 1972. – 368 p.
5. *Bishop C.M.* Neural Networks for Pattern Recognition. – Oxford: Clarendon Press, 1995. – 482 p.
6. *Rojas R.* Neural Networks. A Systematic Introduction. – Berlin: Springer Verlag, 1996. – 502 p.
7. *Looney C.* Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists. – N.Y.: Oxford University Press, 1997. – 480 p.
8. *Callan R.* The Essence of Neural Networks. – London: Prentice Hall Europe, 1999. – 288 p.
9. *Haykin S.* Neural Networks. A Comprehensive Foundation. – Upper Saddle River, N.J.: Prentice Hall, 1999. – 842 p.
10. *Pattern recognition using radial basis function networks / D.R. Zahirniak, R. Chapman, S.K. Rogers et al. // Proc. 6-th Ann. Aerospace Appl. of AI. – Dayton, OH, 1990. – P. 249–260.*
11. *Nelles O.* Nonlinear System Identification. – Berlin: Springer Verlag, 2001. – 785 p.
12. *Specht D.F.* Probabilistic neural networks // Neural Networks. – 1990. – N 3 (1). – P. 109–118.
13. *Specht D.F.* Probabilistic neural networks and polynomial ADALINE as complementary techniques for classification // IEEE Trans. on Neural Networks. – 1990. – N 1 (1). – P. 111–121.
14. *Rutkowski L.* Computational Intelligence. Methods and Techniques. – Berlin: Springer Verlag, 2008. – 514 p.
15. *Kroll A.* Computational Intelligence. Eine Einfuehrung in Probleme, Methoden und technische Anwendungen. – Muenchen: Oldenbourg Verlag, 2013. – 428 s.
16. *Computational Intelligence. A Methodological Introduction / R. Kruse, C. Borgelt, F. Klawonn et al. – London: Springer Verlag, 2013. – 490 p.*
17. *Du K.-L., Swamy M.N.S.* Neural Networks and Statistical Learning. – Springer Verlag, 2014. – 824 p.
18. *Jang J.-S.R., Sun C.-T., Mizutani E.* Neuro-Fuzzy and Soft Computing – Computational Approach to Learning and Machine Intelligence.– Upper Saddle River, N.J.: Prentice Hall, 1997. – 640 p.
19. *Bezdek J.C.* Pattern Recognition with Fuzzy Objective Function Algorithms. – Norwell: Kluwer Acad. Publ. MA, USA, 1981. – 256 p.
20. *Fuzzy Cluster Analysis Methods for Classification, Data Analysis and Image Recognition / F. Hoepfner, F. Klawonn, R. Kruse et al. – Chichester: JohnWiley & Sons, Ltd., 1999. – 289 p.*
21. *Bodyanskiy Ye., Shubkina O.* Semantic annotation of text documents using evolving neural network based on principle «Neurons at Data Points» // Proc. 4th Int. Workshop on Inductive Modelling «IWIM 2011», Kyiv, 2011. – P. 31–37.
22. *Bodyanskiy Ye., Shubkina O.* Semantic annotation of text documents using modified probabilistic neural network // Proc. 6th IEEE Int. Conf. on Intelligent Data Acquisition and Advanced Comp. Syst., 15–17 Sept. 2011, Prague. – P. 328–331.
23. *Bodyanskiy Ye., Pliss I., Volkova V.* Modified probabilistic neuro-fuzzy network for text documents proc. // Int. J. Computing. – 2012. – N 11 (4). – P. 391–396.
24. *Parzen E.* On the estimation of a probability density function and the model // Ann. Math. Statist. – 1962. – N 38. – P. 1065–1076.
25. *Moody J., Darken C.J.* Fast learning in networks of locally-tuned processing units // Neural Computing. – 1989. – N 1. – P. 281–299.
26. *Specht D.E.* A general regression neural network // IEEE Trans. on Neural Networks. – 1991. – N 6(2). – P. 568–576.
27. *Farrell J.A., Michel A.N.* A synthesis procedure for Hopfield's continuous-time associative memory // IEEE Trans. on Circuits Syst. – 1990. – N 37. – P. 877–884.
28. *Hassoun M.H., Watta P.B.* Associative memory networks // Handbook on Neural Computation. – Oxford: IOP Publ. Ltd. and Oxford University Press, 1997. – C.1.3:1 – C.1.3:14.
29. *Cios K.J., Pedrycz W.* Neural-fuzzy algorithms / Ibid. – C.1.3:1–C.1.3:7.
30. *Bodyanskiy Ye., Teslenko N.* Autoassociative memory evolving system based on fuzzy basis functions // Sci. J. of Riga Technical University «Computer Science, Information Technology and Management Sci.» – 2010. – N 44. – P. 9–14.

Поступила 18.06.2014

Тел. для справок: +38 067 399-9150 (Харьков)

E-mail: bodya@kture.kharkov.ua

© Е.В. Бодянский, Е.А. Винокурова, Д.Д. Пелешко, 2014