УДК 519.7:004.8

Е.В. Бодянский, В.В. Волкова, Е.В. Махиборода

Харьковский национальный университет радиоэлектроники, Харьков

НЕЙРОННАЯ СЕТЬ Т. КОХОНЕНА С НЕЧЕТКИМ ВЫВОДОМ И АЛГОРИТМ ЕЕ САМООБУЧЕНИЯ

Предлагается алгоритм обучения для самоорганизующихся карт (SOM), который позволяет ускорить процесс обработки информации благодаря рациональному выбору параметра величины шага обучения, обладает возможностью работы в случае неизвестного количества кластеров, а также при их пересечении, что достигается благодаря использованию процедуры нечеткого вывода, определяющей степени принадлежности классифицируемого наблюдения ко всем кластерам.

Ключевые слова: *кластеризация, самообучение, конкуренция, синаптическая адаптация, кооперация, нечеткий вывод.*

Введение

В задачах интеллектуального анализа данных, связанных с кластеризацией, диагностикой, сжатием информации и т.п. широкое распространение получили самоорганизующиеся карты, введенные Т. Кохоненом [1].

Свойства самоорганизации SOM связаны с тем, что настройка синаптических весов происходит без внешнего обучающего сигнала, т.е. в режиме самообучения, при этом каждый поступающий образ вызывает настройку тех или иных параметров.

В случае пересекающихся и невыпуклых кластеров решение о принадлежности входного образа к одному из кластеров, принимаемое в соответствии с правилом «победитель получает все», может давать неточную кластеризацию, так как некоторые образы могут принадлежать одновременно к нескольким кластерам с определенной степенью принадлежности. В связи с этим целесообразно обеспечить SOM свойства и возможности нечеткой кластеризации [2].

В [3, 4] была предложена нечеткая самоорганизующаяся карта, в которой нейроны SOM заменены нечеткими правилами и множествами. Эта сеть показала достаточно высокую эффективность, однако свойства ее обучения, особенно в реальном времени, были в значительной мере утеряны. В [5, 6] была предложена нечеткая сеть Кохонена для кластеризации, являющаяся, по сути, схемой реализации C-means метода кластеризации.

Постановка задачи. Целью настоящей работы является синтез алгоритма самообучения, позволяющего повысить быстродействие обработки информации, улучшить качество кластеризации в случае пересекающихся кластеров, а также обеспечить возможность работы в случае отсутствия априорной информации о количестве кластеров.

Процедура самообучения, описанная ниже, базируется на принципах конкурентного обучения. Общепринятая процедура самоорганизации состоит из трех основных этапов: конкуренции, кооперации и синаптической адаптации [7].

Конкуренция и синаптическая адаптация

Рассмотрим SOM, содержащую n рецепторов и m нейронов в слое Кохонена, каждый из которых характеризуется собственным n-мерным вектором синаптических весов w_j, $j=1,2...,m$, при этом все нейроны в процессе самообучения-кластеризации на свои входы получают входной вектор-образ $x(k)$ (здесь $k=1,2...$ – либо номер образа, либо номер такта текущего дискретного времени) и генерируют на своих выходах сигналы $y_j(k)=w_j^T(k)x(k)$, $j=1,2...,m$, зависящие от текущих значений векторов $w_j(k)$, настраиваемых с помощью принятого алгоритма самообучения на определенную область пространства X_j. Близкие в смысле принятой метрики входные векторы могут активировать либо один и тот же нейрон w_j, либо пару нейронов-соседей.

Процесс конкуренции начинается с анализа образа $x(k)$, поступающего с рецепторного слоя на все нейроны слоя Кохонена. Для каждого из нейронов вычисляется расстояние

$$D(w_j(k),x(k))=\left\|x(k)-w_j(k)\right\|, \quad (1)$$

при этом, если входы и синаптические веса нормированы так, что

$$\left\|x(k)\right\|=\left\|w_j(k)\right\|=1, \quad (2)$$

а в качестве расстояния используется евклидова метрика, то мерой близости между векторами $x(k)$ и $w_j(k)$ может служить скалярное произведение

$$\begin{gathered} w_j^T(k)x(k)=y_j(k)= \\ \cos(w_j^T(k),x(k))=\cos\theta_j(k). \end{gathered} \quad (3)$$

Далее определяется нейрон-победитель, «ближайший» по входному образу, такой, что

$$D(w^*(k), x(k)) = \min D(w_j(k), x(k)), \tag{4}$$

после чего (временно пропуская этап кооперации) можно настроить синаптические веса сети с помощью правила самообучения «победитель получает все»:

$$w^*(k+1) = w^*(k) + \eta(k)(x(k) - w^*(k)), \tag{5}$$

где $w^*(k)$ – нейрон- победитель, а $\eta(k)$ – параметр, определяющий величину шага обучения, который обычно выбирается эмпирически и уменьшается со временем.

Несложно видеть, что такое обучение в процессе настройки минимизирует критерий

$$E_j^K = \sum_k E_j^K(k) = \frac{1}{2}\sum_{p=1}^{k}\left\|x(p) - w_j\right\|^2, \tag{6}$$

т.е. фактически в качестве оценки синаптических весов используется среднее арифметическое

$$w_j(k) = \frac{1}{k}\sum_{p=1}^{k} x(p). \tag{7}$$

Таким образом, (5) фактически является процедурой стохастической аппроксимации [8], а коэффициент $\eta(k)$ должен выбираться в соответствии с условиями А. Дворецкого [9], хотя при этом выбор $\eta(k) = 1/k$ приводит к явно заниженной скорости сходимости.

Требованию монотонного уменьшения параметра величины шага поиска соответствует введенная в [10] процедура

$$\eta(k) = r^{-1}(k),\ r(k) = \alpha r(k-1) + \|x(k)\|^2; \tag{8}$$
$$0 \le \alpha \le 1,$$

при этом $\alpha = 1$ приобретает форму, предложенную в [11] для адаптивного алгоритма с повышенной скоростью сходимости.

Варьируя фактором забывания α, несложно обеспечить достаточно широкий интервал изменения шага поиска

$$\frac{1}{k} \le \eta(k) \le 1.$$

Кооперация и синаптическая адаптация

Одной из особенностей SOM является возможное наличие этапа кооперации в процессе самоорганизации, когда нейрон-победитель определяет локальную область топологического соседства, в которой обучается не только он сам, но и его ближайшее окружение. Эта топологическая область определяется функцией соседства $\phi(j,l)$, зависящей от расстояния $D(w^*(k), w_l(k))$ между победителем $w^*(k)$ и любым из нейронов $w_l(k)$, $l = 1,2...,m$ и параметра σ, задающего ее ширину.

Обычно $\phi(j,l)$ – это ядерная функция, симметричная относительно максимума в точке $w^*(k) = w_{l^*}(k)(\phi(l^*, k) = 1)$ и убывающая с ростом расстояния $D(w^*(k), w_l(k))$. Наиболее часто в качестве функций соседства используются гаусиан, «Mexican Hat», косинусообразная и им подобные функции.

Использование функции соседства приводит к правилу обучения

$$w_l(k+1) = w_l(k) + \eta(k)\phi(l,k)(x(k) - w_l(k));$$
$$l = 1,2...,m, \tag{9}$$

минимизирующему критерий

$$E_j^K = \sum_{p=1}^{k} E_l^K(p) = \frac{1}{2}\sum_{p=1}^{k}\phi(l,p)\left\|x(p) - w_l\right\|^2 \tag{10}$$

и реализующему принцип «победитель получает больше». Анализ сходимости процессов самообучения, проведенный в [12 – 14], показал, что в процессе настройки синаптических весов должен уменьшаться не только шаг поиска $\eta(k)$, но и параметр, определяющий ширину функций соседства, которая таким образом становится зависимой от времени. Так для наиболее популярной гауссовской функции соседства

$$\phi(j,l) = \exp\left(-\frac{\left\|w^*(k) - w_l(k)\right\|^2}{\sigma^2}\right) \tag{11}$$

в [13, 14] было предложено для настройки параметра ширины σ использовать выражение

$$\sigma(k) = \beta\sigma(k-1),\quad 0 < \beta < 1, \tag{12}$$

где β – скалярный параметр, определяющий скорость уменьшения влияния нейрона-победителя на свое окружение.

В процессе обучения SOM можно вообще отказаться от этапа конкуренции, настраивая синаптические веса сети в зависимости от их близости к текущему предъявляемому образу $x(k)$. В этом случае функция соседства $\phi(l)$ привязывается не к нейрону-победителю $w^*(k)$, а к вектору $x(k)$, при этом в точке $x(k) = w_j^T(k)\phi(j) = 1$, а с ростом $D(w_l(k), x(k))\phi(l)$ убывает подобно обычной ядерной функции.

В качестве основы такой функции может быть использована радиально-базисная конструкция типа «поднятый косинус» [15]:

$$\phi(\theta) = \begin{cases} \dfrac{1 + \cos\theta}{2}, \text{если} - \pi \le 0 \le \pi; \\ 0, \text{в\_противном\_случае}, \end{cases} \tag{13}$$

приводящая к функции соседства

$$\phi_l(k) = \frac{1 + \cos(w_l(k), x(k))}{2} = \\ = \frac{1 + \cos\theta_l(k)}{2} = \frac{1}{2} + \frac{1}{2} y_l(k). \tag{14}$$

Алгоритм самообучения в этом случае приобретает вид

$$w_l(k+1) = w_l(k) + \eta(k)\frac{1 + y_l(k)}{2}; \\ (x(k) - w_l(k)) = w_l(k) + \frac{1}{2}\eta(k); \\ (x(k) - w_l(k)) + \frac{1}{2}\eta(k) y_l(k); \\ (x(k) - w_l(k)) \tag{15}$$

и состоит из двух частей, первая из которых $\eta(k)(x(k) - w_l(k))$ соответствует правилу самообучения Т. Кохонена, а вторая – $\frac{1}{2}\eta(k) y_l(k)(x(k) - w_l(k))$ – правилу обучения входной звезды С. Гроссберга [16], минимизирующему при выполнении условия (2) критерий

$$E_l{}^G = \sum_k E_l{}^G(k) = \frac{1}{2}\sum_{p=1}^{k}(w_l{}^T x(p))^2 = \\ = \frac{1}{2}\sum_{p=1}^{k} y_l{}^2(p). \tag{16}$$

Таким образом, алгоритм самообучения (15) в процессе своей работы оптимизирует аддитивную целевую функцию

$$E_l = \frac{1}{2}E_l{}^K + \frac{1}{2}E_l{}^G. \tag{17}$$

Возможности SOM могут быть расширены, если вместо целевой функции использовать более общую конструкцию

$$E_l = \gamma E_l{}^K + (1-\gamma)E_l{}^G, \quad 0 \le \gamma \le 1, \tag{18}$$

приводящую к алгоритму настройки

$$w_l(k+1) = w_l(k) + \eta(k)(\gamma(x(k) - w_l(k)) + \\ +(1-\gamma) y_l(k)(x(k) - w_l(k))) \tag{19}$$

с функцией соседства

$$y(l) = \gamma + (1-\gamma)\cos\theta_l(k). \tag{20}$$

Варьируя параметром γ, можно реализовать ту или иную форму компромисса между обучением по Кохонену и обучением по Гроссбергу.

С тем, чтобы обеспечить сходимость процедуры (19), необходимо предусмотреть возможность уменьшения ширины функции соседства $y_l(k)$ по ходу процесса обучения. Вводя в (19) параметр ширины $\sigma(k)$, приходим к окончательной форме комбинированного алгоритма настройки SOM:

$$w_l(k+1) = w_l(k) + \eta(k)\delta_{[-1,1]}(\theta_l(k)\sigma^{-1}(k)) \\ (x(k) - w_l(k)) \cdot (\gamma + (1-\gamma)\cos(\pi\theta_l(k)\sigma^{-1}(k))), \\ \eta(k) = r^{-1}(k), \\ r(k) = \alpha r(k-1) + \|x(k)\|^2, \; 0 \le \alpha \le 1, \\ \sigma(k) = \beta\sigma(k-1), \quad 0 < \beta < 1, \tag{21}$$

где функция $\delta_{[a,b]}(x)$ определена как

$$\delta_{[a,b]}(x) = 1, \text{если\_} x \in [a,b] i; \\ 0, \text{в\_противном\_случае.}$$

В случае, если задана обучающая выборка $x(1), x(2), ..., x(N)$ (с классификацией или без нее), в результате процесса самоорганизации будет получен набор синаптических весов $w_l(N+1) = \hat{w}_l$, $l = 1, 2, ..., m$, определяющий координаты центроидов сформированных кластеров.

Нечеткий вывод

Обученная SOM может быть использована для классификации вновь поступающих наблюдений $x(p)$, $p = N+1, N+2, ...$ В случае пересекающихся кластеров решение типа «победитель получает все» может оказаться некорректным, в связи с чем целесообразно ввести оценку уровня принадлежности наблюдений каждому из имеющихся кластеров на основе функции принадлежности [17]:

$$\mu_{\hat{w}l}(x(p)) = \frac{1 + \cos(\hat{w}_l, x(p))}{2}, \tag{22}$$

ограниченной интервалом $0 \le \mu_{\hat{w}l}(x) \le 1$, а, следовательно, отвечающей требованиям к нечеткой функции принадлежности.

Для оценки принадлежности предъявленного вектора $x(p)$ каждому из кластеров используется нормированное выражение

$$\mu_{\hat{w}l}(x(p)) = \frac{y_l(x(p))}{\sum_{l=1}^{m} y_l(x(p))}. \tag{23}$$

Функция принадлежности $\mu_{\hat{w}l}(x)$ является частным случаем квадратичной радиально-базисной функции [18, 19]:

$$\phi_l(x) = \max\left\{0, 1 - \frac{\|x - \hat{w}_l\|^2}{\sigma_l{}^2}\right\} = \\ = \max\left\{0, 1 - 2\frac{1 - \hat{w}_l{}^T x}{\sigma_l{}^2}\right\}. \tag{24}$$

В случае, если в обучающей выборке имеется классификация наблюдений, то для дальнейшей более точной настройки радиусов кластеров может быть использована простая градиентная процедура оптимизации, как, например, введена в [18 – 20]. Настройка радиусов кластеров будет выражаться в получении более точной нечеткой классификации наблюдений.

Выводы

Предложен алгоритм самообучения самоорганизующейся карты, отличающейся повышенным быстродействием и позволяющий осуществлять классификацию наблюдений в условиях пересекающихся кластеров. Важной особенностью алгоритма является возможность определения количества кластеров в данных при отсутствии априорной информации (при $\gamma = 0,45 \div 0,5$) в случаях раздельных и пересекающихся кластеров. Эксперименты подтверждают эффективность предложенного алгоритма.

Список литературы

1. *Kohonen T. Self-Organizing Maps / T. Kohonen. – Berlin: Springer-Verlag, 1995. – 362 p.*
2. *Bezdek J.C. Pattern Recognition with Fuzzy Objective Function Algorithms / J.C. Bezdek. – N.Y.: Plenum Press, 1981. – 272 p.*
3. *Vuorimaa P. Fuzzy self-organizing maps / P. Vuorimaa // Fuzzy Sets and Systems. – 1994. – 66. – P. 223-231.*
4. *Vuorimaa P. Use of fuzzy self-organizing maps in pattern recognition / P. Vuorimaa // In: Proc. 3-rd IEEE Int. Conf. Fuzzy Systems "FUZZ-IEEE`94". – Orlando, USA, 1994. – P. 798-801.*
5. *Tsao E. C.-K. Fuzzy Kohonen clustering networks / E. C.-K. Tsao, J.C. Bezdek, N.R. Pal // Pattern Recognition. – 1994. – 27. – P. 757-764.*
6. *Pascual-Marqui R.D. Smoothly Distributed Fuzzy c-Means: a New Self-Organizing map / R.D. Pascual-Marqui, A.D. Pascual-Montano, K. Kochi, J.M. Caraso // Pattern Recognition. – 2001. – 34. – P. 2395-2402.*
7. *Haykin S. Neural Networks. A Comprehensive Foundation / S. Haykin. – Upper Saddle Press, N.Y.: Prentice Hall, Inc., 1999. – 842 p.*
8. *Vazan M.T. Stochastic Approximation / M.T. Vazan. – Cambridge: AT The University Press, 1969.*
9. *Dvoretzky A. On stochastic approximation / A. Dvoretzky // In: Proc. 3-rd Berkeley Symp. Math. Statistics and Probability. – 1956. – 1. – P. 39-55.*
10. *Бодянский Е.В. Многошаговые оптимальные упредители многомерных нестационарных стохастических процессов / Е.В. Бодянский, И.П. Плисс, Т.В. Соловьева // Доклады АН УССР. – 1986- А. – № 12. – С. 47-49.*
11. *Goodwin G.C. A globally convergent adaptive predictor / G.C. Goodwin, P.J. Ramadge, P.E. Caines // Automatica. – 1981. – 17. – No. 1. – P. 135-140.*
12. *Cottrel M. A stochastic model of retinotopy: a self-organizing process / M. Cottrel, J. Fort // Biological Cybernetics. – 1986. – 53. – P. 405-411.*
13. *Ritter H. On the stationary state of the Kohonen self-organizing sensory mapping / H. Ritter, K. Shulten // Biological Cybernetics. – 1986. – 54. – P. 234-249.*
14. *Ritter H. Covergence properties of Kohonen`s topology conserving maps: fluctuations, stability, and climension selection / H. Ritter, K. Shulten // Biological Cybernetics. – 1988. – 60. – P. 59-71.*
15. *Schilling R.J. Approximation of Nonlinear Systems with Radial Basis Function Neural Networks / R.J. Schilling, J.J. Carroll, A.F. Al-Ajlouni // IEEE Trans. On Neural Networks. – 2001. – 12. – No. 1. – P. 1-15.*
16. *Grossberg S. Classical and instrumental learning by neural networks / S. Grossberg // In: Proc. "Progress in Theoretical Biology". – 3. – N.Y.: Academic Press, 1974. – P. 57-141.*
17. *Bodyanskiy Ye. Combined learning algorithm for a self-organizing map with fuzzy inference / Ye. Bodyanskiy, Ye. Yorshkov, V. Kolodyazhniy, A. Stephan // In "Computational Intelligence: Theory and Applications" – Ed. B. Reusch. – Berlin-Heidelberg: Springer, 2005. – P. 641-650.*
18. *Bodyanskiy Ye. Adaptive quadratic radial basic function for time series forecasting / Ye. Bodyanskiy, O. Chaplanov, V. Kolodyazhniy, P. Otto // In: Proc. East West Fuzzy Coll. 2002, 10-th Zittau Fuzzy Coll. – Zittau/Goerlitz: HS, 2002. – P. 164-172.*
19. *Bodyanskiy Ye. Probabilistic neuro-fuzzy network with nonconventional activation functions / Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy, J. Wernstedt // Lecture Notes in Artificial Intelligence. – Berlin-Heidelberg-New York: Springer, 2003. – V. 2773. – P. 973-979.*
20. *Bodyanskiy Ye. Resource-allocating probabilistic neuro-fuzzy network / Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy // In: Proc. Third Conf. of the European Society for Fuzzy Logic and Technology (EUSFLAT-2003), 10-12 September, 2003. – Zittau, Germany, 2003. – P. 392-395.*

Поступила в редколлегию 7.05.2009

Рецензент*:* д-р техн. наук, проф. В.А. Филатов, Харьковский национальный университет радиоэлектроники, Харьков.

НЕЙРОННА МЕРЕЖА Т.КОГОНЕНА З НЕЧІТКИМ ВИВЕДЕННЯМ ТА АЛГОРИТМ ЇЇ САМОНАВЧАННЯ

Є.В. Бодянський, В.В. Волкова, К.В. Махіборода

Пропонується алгоритм навчання для самоорганізувальних мап (SOM), що дозволяє прискорити процес обробки інформації завдяки раціональному вибору параметра величини кроку навчання, може працювати при невідомій кількості кластерів, а також якщо вони перетинаються. Це досягається завдяки використанню процедури нечіткого виведення, яка визначає ступінь належності спостереження, що класифікується, до всіх кластерів.

Ключові слова: *кластеризація, самонавчання, конкуренція, синаптична адаптація, кооперація, нечітке виведення.*

T.KOHONEN'S NEURAL NETWORK WITH FUZZY INFERENCE AND ITS SELF-LEARNING ALGORITHM

Ye.V. Bodyanskiy, V.V. Volkova, Ye.V. Mahiboroda

A learning algorithm for a self-organizing map (SOM) is proposed. The algorithm accelerates information processing due to the rational choice as the learning rate parameter, and can work when the number of clusters is unknown, as well as when the clusters are overlapping. This is achieved via the introduction of fuzzy inference that determines the level of membership of the classified pattern to each of the available classes.

Keywords: *clustering, self-training, competition, synaptic adaptation, co-operation, unclear conclusion.*