

УДК 004.8;004.032.26

ОБ АДАПТИВНОМ УВЕЛИЧЕНИИ РАЗМЕРНОСТИ ПРОСТРАНСТВА ПРИЗНАКОВ

Ю.Р. Цой (*yurytsoy@gmail.com*)

Томский политехнический университет, Томск

В работе описываются результаты исследования, направленного на анализ влияния размерности пространства признаков на точность нейросетевой классификации. Первые результаты показывают, что, несмотря на адаптивное преобразование входного пространства, осуществляемое с использованием нейроэволюционного подхода, возможно как повышение точности классификации, так и ее снижение. Дается обсуждение результатов и возможный план будущих исследований.

Введение

Разделимость описаний объектов (далее «разделимость объектов») из разных классов является одним из требований для успешного решения задачи классификации. Однако в подавляющем большинстве задач требование разделимости не выполняется в силу неоднозначности данных, ошибок измерения, недостаточности информации и других причин.

В наиболее сильной формулировке требование разделимости совпадает с гипотезой компактности, в соответствии с которой расстояние между любыми объектами из одного класса должно быть меньше, чем расстояние между любыми объектами разных классов. Однако гипотеза компактности при решении реальных задач также, как правило, не выполняется, и более того, ее принятие без веских на то оснований может приводить к ошибочным заключениям при классификации объектов, например, при использовании метода ближайших соседей, и решении задач кластеризации.

В большинстве случаев, говоря о разделимости объектов из разных классов, рассматривают базовый вариант: количество классов ограничено двумя (дихотомия), поскольку он проще с точки зрения анализа и может быть легко расширен на случай большего числа классов.

Один из основных результатов, касающихся разделимости объектов, получен в работе Ковера [Cover, 1965], в которой показано, что вероятность неоднозначного обобщения для обучаемого классификатора

(т.е. вероятность неверной классификации объекта находящегося близко к разделяющей поверхности) уменьшается с ростом размерности пространства. Другими словами с ростом размерности пространства увеличивается вероятность построения корректной процедуры классификации. Там же показано, что линейная разделяемость является более сильным свойством, чем сферическая разделяемость, а последняя, в свою очередь, сильнее квадратичной.

Можно показать, что если два множества являются линейно разделяемыми в \mathbf{R}^N , то существует такая цепочка преобразований из \mathbf{R}^N в \mathbf{R}^1 , в результате которой образы этих множеств будут также линейно разделяемыми. Однако алгоритм построения такой цепочки преобразований может быть достаточно трудоемок и сопоставим по вычислительной сложности с отысканием оптимальной разделяющей гиперплоскости [Yogananda et al., 2007].

Увеличение размерности пространства признаков размерности N , которое для вектора $\mathbf{x} \in \mathbf{R}^N$ осуществляется с использованием матрицы \mathbf{W} размерностью $M \times N$, $M > N$, путем преобразования $\mathbf{W}\mathbf{x}$, в реальности не способствует увеличению размерности входного пространства, несмотря на увеличение размерности вектора признаков. Это происходит потому, что образы линейного преобразования с оператором \mathbf{W} находятся в линейном подпространстве размерности $rank\mathbf{W}$, который не превышает N . Одним из выходов является нелинейное преобразование компонент вектора-образа.

Отметим, что ввиду большого разнообразия задач и особенностей данных создать универсальный алгоритм преобразования входного пространства для лучшей разделяемости описаний объектов из разных классов не представляется возможным.

Можно сделать предположение, что разделяемость двух и более множеств может быть косвенно оценена по результатам обучения классификатора, поскольку, если множества разделяемы, то объекты, им принадлежащие являются различимыми, и вероятность успешного обучения выше. Среди известных методов преобразования входного пространства для большей различимости объектов из разных классов отметим машины опорных векторов [Vapnik, 1998], нейросетевые методы классификации с применением карты Кохонена [Suganthan, 2001], Echo State Networks [Jaeger et al., 2004] и так называемые «deep-learning» архитектуры нейронных сетей [Bengio et al., 2007].

В данной работе рассматривается исследование применения эволюционного подхода для преобразования входного пространства признаков таким образом, чтобы повысить вероятность успешного обучения искусственной нейронной сети (ИНС).

1. Описание экспериментов

Будем рассматривать применение эволюционного подхода для этапа обработки входных векторов таким образом, чтобы изменить размерность входного пространства для обучения ИНС.

Для этого эволюционный алгоритм будет использоваться для настройки весов искусственной нейронной сети без скрытых слоев. Количество нейронов в выходном слое определяется как αn_l , где α – константа, n_l – количество входных нейронов. Будем обозначать такую ИНС как ИНС-1.

Оценка приспособленности f_i для i -й особи производится следующим образом: $f_i = \max\{e_j, j = 1, 2, 3\}$, где e_j – ошибка обучения ИНС по алгоритму RPROP (Resilient PROPagation) на протяжении 50 эпох. Используется максимальная ошибка по трем попыткам обучения ИНС для того чтобы уменьшить вероятность переобучения нейронной сети. Для реализации алгоритма RPROP использовалась библиотека Encog¹.

Выходные сигналы обученной ИНС-1 используются в качестве входных сигналов для ИНС-2, обучаемой по алгоритму RPROP. ИНС-2 также не имеет скрытых слоев и имеет αn_l входных нейронов. В целом совокупность ИНС-1 и ИНС-2 можно рассматривать как одну нейронную сеть с одним скрытым слоем, содержащим αn_l нейронов. Обучение такой сети разбивается на два этапа:

1. Нейроэволюционное обучение ИНС-1. В качестве решения выбирается ИНС-1 с наименьшей приспособленностью, встреченной за все время эволюции.
2. Обучение ИНС-2 с использованием в качестве входных сигналов выходных сигналов ИНС-1, полученной на этапе 1.

2. Результаты

Целью экспериментов является проверка влияния эволюционной настройки преобразования входных векторов на результат решения задачи классификации. Рассматривается ряд задач из набора Proben1 [Prechelt, 1994], а именно: cancer1, card1, diabetes1, glass1, heart1, horse1.

В качестве эволюционного алгоритма используется вещественный генетический алгоритм с турнирной селекцией, BLX-alpha кроссинговером, гауссовской мутацией и элитаризмом. Размер популяции равен 20 особям; вероятность кроссинговера и мутации соответственно 0,7 и 0,01; размер турнира – 4 особи.

¹ <http://www.heatonresearch.com/encog>

Функции активации в ИНС-1 и ИНС-2 – сигмоидные, вида: $y = (1 + \exp(-aS))^{-1}$, где a – константа, $S = \mathbf{w}^T \mathbf{x}$ – взвешенная сумма входных сигналов нейрона.

Результаты экспериментов для ИНС-1: средняя и среднеквадратическое отклонение (СКО) точности классификации на тестовом множестве по 10 ИНС-2 для данной ИНС-1, представлены в табл. 1. Жирным шрифтом выделены лучшие результаты. Производилось сравнение с «базовыми» результатами для набора Proben1 из статьи [Prechelt, 1994], полученными при ручной настройке структуры ИНС, обучаемой по алгоритму RPROP.

Табл. 1. Результаты тестирования ИНС: средняя ошибка классификации на тестовом множестве (в скобках – СКО), – при длительности эволюции ИНС-1 10 поколений

Задача	Длительность обучения ИНС-2, эпох					Результаты из [Prechelt, 1994]
	100	200	300	400	500	
cancer1	2,36 (0,18)	2,70 (0,28)	2,64 (0,30)	2,64 (0,30)	2,59 (0,30)	1,38 (0,49)
card1	10,17 (0,63)	11,51 (0,66)	11,74 (0,77)	11,51 (0,86)	11,86 (0,74)	14,05 (1,03)
diabetes1	22,34 (1,05)	21,98 (0,59)	21,93 (0,46)	21,51 (0,43)	21,41 (0,16)	24,10 (1,91)
glass1	28,68 (1,73)	26,98 (1,55)	26,98 (1,79)	26,41 (0)	26,60 (0,60)	32,70 (5,34)
heart1	21,00 (1,26)	21,87 (0,98)	21,70 (0,69)	21,43 (0,58)	21,26 (0,52)	19,72 (0,96)
horse1	36,15 (3,38)	37,25 (3,04)	37,25 (1,90)	38,02 (2,85)	36,48 (1,85)	29,19 (2,62)

Результаты экспериментов показывают, что в ряде случаев (задачи card1, diabetes1, glass1) удается добиться повышения точности классификации, а в других случаях (задачи cancer1, heart1, horse1) точность классификации падает. При этом если в задачах heart1 и horse1 «отставание» рассматриваемого подхода небольшое (хотя и статистически значимое, с уровнем значимости 0,05), то для задачи horse1 разница в точности классификации весьма существенная и составляет 7-9 процентов в зависимости от длительности обучения ИНС-2.

Дополнительное исследование зависимости ошибки классификации от длительности обучения ИНС-1 (10, 20, 30, 40 и 50 поколений), при 100 эпохах обучения ИНС-2, показали (табл. 2), что для задачи heart1 точность классификации с ростом количества поколений обучения ИНС-1 повышается, а затем начинает уменьшаться. Для задач cancer1 и horse1 эта закономерность также наблюдается, но улучшений недостаточно, чтобы показать результаты, превосходящие таковые из [Prechelt, 1994].

Табл. 2. Зависимость средней и СКО ошибки классификации на тестовом множестве от длительности обучения ИНС-1. Длительность обучения ИНС-2 равнялась 100 эпохам

Задача	Длительность обучения ИНС-1, поколения					Результаты из [Prechelt, 1994]
	10	20	30	40	50	
cancer1	2,36 (0,18)	2,76 (0,53)	2,24 (0,18)	3,05 (0,39)	2,41 (0,24)	1,38 (0,49)
card1	10,17 (0,63)	11,16 (0,86)	11,16 (0,90)	10,99 (1,14)	11,40 (0,68)	14,05 (1,03)
diabetes1	22,34 (1,05)	23,39 (0,57)	22,08 (0,56)	22,81 (0,77)	23,45 (0,74)	24,10 (1,91)
glass1	28,68 (1,73)	35,28 (1,55)	34,34 (1,73)	33,40 (2,68)	36,23 (1,95)	32,70 (5,34)
heart1	21 (1,26)	17,74 (1,02)	18,65 (1,29)	18,87 (0,80)	17,87 (0,95)	19,72 (0,96)
horse1	36,15 (3,38)	39,89 (3,20)	34,40 (3,84)	40,44 (4,88)	40,44 (3,83)	29,19 (2,62)

Отметим, что для задач, в которых были показаны сравнительно хорошие результаты, в ряде случаев имеется тенденция к снижению точности классификации с ростом числа поколений для обучения ИНС-1, что, вероятно, можно объяснить переобучением этой сети.

Исследование зависимости результатов классификации от размеров выходного слоя ИНС-1, количества t_1 поколений обучения ИНС-1 и эпох t_2 для обучения ИНС-2 (табл. 3), выявило, что увеличение α приводит к повышению точности классификации, в частности, во многих случаях результаты улучшились при $\alpha = 2,5$. Однако в ряде случаев этого улучшения недостаточно, чтобы показать результат, превосходящий «традиционное» обучение ИНС.

Табл. 3. Зависимость средней и СКО ошибки классификации на тестовом множестве для задач cancer1 и horse1 от размеров выходной слоя ИНС-1. Длительность обучения ИНС-2 равнялась 100 эпохам

Задача	Значение α					
	0,5	1,0	1,5	2,0	2,5	3,0
cancer1, $t_1=10, t_2=100$	3,33 (0,53)	3,91 (0,24)	5,80 (0,42)	2,36 (0,18)	2,07 (0,73)	2,59 (0,49)
cancer1, $t_1=20, t_2=100$	3,56 (0,24)	2,36 (0,33)	2,87 (0,27)	2,76 (0,53)	2,07 (0,30)	2,53 (0,30)
cancer1, $t_1=10, t_2=200$	3,98 (0,18)	3,39 (0,18)	6,32 (0,27)	2,70 (0,28)	2,36 (0,33)	3,74 (0,62)
cancer1, $t_1=20, t_2=200$	3,16 (0,30)	2,99 (0,24)	2,64 (0,40)	2,36 (0,33)	1,78 (0,18)	2,30 (0)
horse1, $t_1=10, t_2=100$	35,16 (1,94)	42,18 (2,21)	34,84 (3,66)	36,15 (3,38)	34,18 (3,29)	36,04 (2,12)
horse1, $t_1=20, t_2=100$	39,89 (1,94)	42,31 (1,49)	36,15 (2,95)	39,89 (3,20)	37,25 (4,62)	40,55 (2,61)
horse1, $t_1=10, t_2=200$	38,68 (1,78)	40,22 (1,73)	36,15 (2,16)	37,25 (3,04)	36,04 (3,06)	38,68 (4,17)
horse1, $t_1=20, t_2=200$	39,01 (2,50)	40,22 (3,12)	36,15 (3,17)	41,54 (2,19)	36,92 (4,12)	40,33 (1,72)

Таким образом, в 2 задачах из 6 рассматриваемых не удалось добиться улучшения результатов классификации. И если для задачи cancer1 во многих случаях разница точности классификации с результатом из [Prechelt, 1994] в абсолютном выражении была небольшой, то для задачи horse1, она иногда превышала 10 %.

3. Заключение

Результаты исследований показали, что возможно улучшение точности классификации при преобразовании пространства входных признаков, с помощью ИНС, обучаемой эволюционным алгоритмом.

Отметим, что наилучшие результаты для рассматриваемого подхода достигались, как правило, при небольшом количестве поколений и эпох для обучения соответственно ИНС-1 и ИНС-2. С ростом значений этих параметров в ряде случаев наблюдалось падение точности классификации.

Полученные в проведенном исследовании результаты неполны и оставляют немало открытых вопросов. В частности:

1. Результаты получены по однократному обучению ИНС-1 и поэтому могут не являться объективными.
2. Как изменяются результаты, если добавить скрытые слои в ИНС-1 и ИНС-2 и рассмотреть большие диапазоны длительности обучения и значения коэффициента α ?

В [Cover, 1965] показано, что разделяющая емкость дискриминантной функции порядка k для случайных образов пропорциональна C_k^{n+k} , поэтому можно предположить, что использование функций активации порядка выше 1 способно дать большую различимость объектов из разных классов. Однако, следует заметить, что среди всех возможных дихотомий только малая часть может оказаться «полезной».

В целом, можно сделать вывод, что увеличение размерности входного пространства даже в случае адаптации преобразования не гарантирует повышение точности классификации и в некоторых случаях способно существенно снизить результаты нейросетевой классификации.

Благодарности. Работа выполнена при финансовой поддержке РФФИ (проект № 09-08-00309-а) и гранта фонда «Научный потенциал».

Список литературы

- [**Bengio et al., 2007**] Bengio Y., LeCun Y. Scaling Learning Algorithms towards AI / In Bottou L., Chapelle O., De Coste D., and Weston J. (Eds): Large-Scale Kernel Machines. – MITPress, 2007.
- [**Cover, 1965**] Cover T.M. Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition // IEEE Transactions on Electronic Computers. 1965. EC Vol. 14(3). P. 326-334.
- [**Jaeger et al., 2004**] Jaeger H., Haas H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication // Science. 2004. Vol. 304 (5667). P. 78 – 80.
- [**Prechelt, 1994**] Prechelt L. PROBEN1 – a set of neural network benchmark problems and benchmarking rules. Technical Report 21/94. – Fakultat fur Informatik, Universitat Karlsruhe, Karlsruhe, Germany, 1994.
- [**Suganthan, 2001**] Suganthan P. N. Pattern classification using multiple hierarchical overlapped self-organising maps // Pattern Recognition. 2001. Vol. 34(11). P. 2173-2179.
- [**Vapnik, 1998**] Vapnik V. Statistical Learning Theory. – New York: Wiley-Interscience, 1998.
- [**Yogananda et al., 2007**] Yogananda A. P., M Narasimha M., Lakshmi G. A fast linear separability test by projection of positive points on subspaces // Proc. of the 24-th International Conference on Machine Learning, Corvallis, OR. 2007. P. 713-720.