

Ю. Р. ЦОЙ

Национальный исследовательский Томский политехнический университет
Томский государственный университет систем
управления и радиоэлектроники
yurytsoy@gmail.com

НЕЙРОЭВОЛЮЦИОННЫЙ МЕТОД ПСЕВДО-ГЛАВНЫХ КОМПОНЕНТ*

В данной статье рассматривается применение нейроэволюционного подхода к сокращению размерности пространства признаков путем приближенного вычисления собственных векторов матрицы автоковариации обучающих данных. В силу этой особенности алгоритм назван методом псевдо-главных компонент. В алгоритме используются оригинальные процедуры оценивания особей и скрещивания. Экспериментально показано, что неточные значения главных компонент позволяют сократить размерность описания данных и получить точность классификации, сопоставимую с результатами традиционного обучения нейронной сети. Алгоритм может быть эффективно распараллелен, в отличие от численных методов решения полной проблемы собственных значений и от алгоритмов Хеббовского типа, реализующих метод главных компонент.

Введение

Метод главных компонент (МГК) является одним из наиболее популярных методов сокращения размерности пространства признаков для решения задач распознавания образов [1]. В этом методе используется простая и эффективная идея о том, что существуют линейные подпространства в пространстве признаков, которые не вносят существенного информационного вклада в имеющийся массив данных, и поэтому могут быть отброшены без значительной потери информации. МГК можно рассматривать как алгоритм сжатия с потерями, поскольку отбрасываемые линейные подпространства часто содержат некоторую полезную информационную составляющую.

Технически в МГК требуется найти собственные значения и векторы для матрицы автоковариации S обучающих данных. Для этого применя-

* Данная работа выполнена при поддержке РФФИ, проекты № 09-08-00309-а, 11-07-00027-а.

ются численные методы, при этом, учитывая, что матрица \mathbf{C} – симметричная, можно использовать специализированные более эффективные с вычислительной точки зрения методы [2]. Однако одним из недостатков применения классических численных методов к решению проблемы собственных значений является их слабая распараллеливаемость.

В то же время можно предположить, что не так уж важно найти «главные» собственные векторы, имеющие наибольшие собственные значения, как отбросить малоинформативные векторы, которым соответствуют малые собственные значения. В данной статье предлагается нейрорволюционный подход для удаления линейных подпространств, вносящих малый вклад в описание обучающих данных. Предлагаемый метод отличается от известных нейросетевых алгоритмов [3] для реализации МГК, основывающихся на идее Хеббовского обучения тем, что в нем не рассматриваются корреляции между выходами нейронов. Поскольку в представленном методе не производится поиск точных координат собственных векторов, будем называть его Метод псевдо-Главных Компонент (МпГК).

Идея метода и описание алгоритма

Как было обозначено выше, представленный метод удаляет малоинформативные линейные подпространства пространства признаков. Для начала отметим, что МГК может быть сформулирован как поиск линейных подпространств, ортогональная проекция обучающих данных на которые имеет максимальную дисперсию. Другими словами, линейные подпространства с малой дисперсией проекций точек данных не удовлетворяют МГК. Вопрос заключается в том, что можно ли отбросить малоинформативные подпространства, если они определены с некоторой погрешностью, без ущерба для поиска главных компонент? Положительный ответ возможен благодаря следующему утверждению

Утверждение 1: Пусть $\mathbf{X}_i = \{\mathbf{X}_i, i=1, \dots, N\}$, $\mathbf{X}_i \in R^n$ набор точек данных и $\mathbf{Q}_i = \{\mathbf{q}_i, i=1, \dots, n\}$ ортогональный базис в R^n . Обозначим проекцию точек данных из \mathbf{X} на координатный вектор \mathbf{q}_i как $proj_{\mathbf{q}_i}(\mathbf{X})$, а дисперсию проекций через $Var(proj_{\mathbf{q}_i}(\mathbf{X}))$. Тогда сумма по всем координатам

$$\sum_{i=1, \dots, n} Var(proj_{\mathbf{q}_i}(\mathbf{X})), \quad (1)$$

постоянна и не зависит от выбора \mathbf{Q} .

Доказательство. Утверждение доказывается путем раскрытия скобок в выражении (1), и после ряда преобразований можно показать, что результирующее выражение не содержит \mathbf{Q} .

Поскольку матрица автоковариации симметрична, то ее собственные векторы, отыскиваемые с помощью МГК, ортогональны, и их можно использовать в качестве координатных векторов некоторого ортогонального базиса. Таким образом, данное утверждение гласит, что т.к. сумма проекций постоянна, то если столбцы матрицы $\hat{\mathbf{Q}}$ являются оценками собственных векторов, то малоинформативные столбцы $\hat{\mathbf{Q}}$ будут обладать еще меньшим вкладом в описании данных, когда координаты оценок «главных» собственных векторов будут уточнены. Поэтому можно отбросить собственные векторы, если соответствующие им собственные значения не будут удовлетворять некоторому критерию, без ущерба для уточнения координат более «важных» собственных векторов. Помимо всего прочего такой подход может иметь преимущество с вычислительной точки зрения, т.к. при отбрасывании собственных векторов размерность задачи уменьшается.

Другими словами, утверждение можно трактовать так, что суммарная дисперсия проекций на координатные векторы является подобием конечного ресурса, который «распределяется» по координатным векторам. Если дисперсию проекций на какой-либо вектор увеличивается, то должен быть вектор (или векторы), дисперсия проекций на которые уменьшается. Это и является основанием для предположения, что малоинформативные собственные векторы не будут «улучшены» (в смысле дисперсии проекций) на последующих итерациях, и рассматривать их не имеет смысла.

В качестве критерия для отбрасывания малоинформативных собственных векторов (линейных подпространств) в данной статье будет рассматриваться следующее неравенство (предполагается, что собственные векторы отсортированы в порядке убывания дисперсии):

$$\frac{\text{Var}(\text{proj}_{\hat{\mathbf{q}}_i}(\mathbf{X}))}{\text{Var}(\text{proj}_{\hat{\mathbf{q}}_i}(\mathbf{X}))} < \tau, \quad (2)$$

где $\hat{\mathbf{q}}_i$ – оценка i -го собственного вектора, τ – пороговое значение, чем оно меньше, тем больше координатных векторов будут отброшены. Типичными значениями для τ могут быть 5, 10, 20, ...

Рассмотрим искусственную нейронную сеть (ИНС) без скрытых нейронов, с нулевыми смещениями и линейными функциями активации. В

такой сети сигнал выходного нейрона равен $y = \mathbf{w}^T \mathbf{x}$. Полагая $\|\mathbf{w}\| = 1$, видно, что y – координата проекции точки с координатами \mathbf{x} на вектор \mathbf{w} . Тогда ИНС с n выходными нейронами, веса которых представляют ортогональные координатные векторы некоторого базиса в \mathcal{X}^n , для данного вектора \mathbf{x} входных сигналов будет вычислять координаты проекций на координатные векторы. Таким образом можно провести параллель между весами связей выходного нейрона и координатным вектором ортогонального базиса. Далее эти понятия будут использоваться как синонимы.

Вышеприведенные рассуждения приводят к построению следующего нейроэволюционного алгоритма:

1. **Инициализация** случайной популяции, каждая особь является потенциальным решением для МПГК.
2. **Оценка** каждой особи с использованием следующей функции приспособленности:

$$f = \alpha \sum_i \text{Var}(\text{proj}_{\hat{\mathbf{q}}_i}(\mathbf{X})) \rightarrow \max$$

$$\alpha = (\hat{\mathbf{q}}_0^T \mathbf{r})^2,$$

$$\mathbf{r} = \frac{\mathbf{C}\hat{\mathbf{q}}_0}{\|\mathbf{C}\hat{\mathbf{q}}_0\|},$$

Коэффициент α необходим для оценки, насколько наиболее «значимый» координатный вектор близок к собственному вектору. Если $\hat{\mathbf{q}}_0$ – нормированный собственный вектор, то

$$\alpha = \left(\frac{\hat{\mathbf{q}}_0^T \mathbf{C}\hat{\mathbf{q}}_0}{\|\mathbf{C}\hat{\mathbf{q}}_0\|} \right)^2 = \left(\frac{\lambda \hat{\mathbf{q}}_0^T \hat{\mathbf{q}}_0}{\|\lambda \hat{\mathbf{q}}_0\|} \right)^2 = 1, \text{ в противном случае } \alpha < 1.$$

- a. Вычисление вектора средних дисперсий $\overline{\text{var}}$ проекций на оценки собственных векторов.
 - b. Начиная с последнего элемента $\overline{\text{var}}$ проверить, удовлетворяется ли критерий (2). При положительном ответе необходимо удалить соответствующие гены из хромосом всех особей популяции и перейти к следующему элементу. В противном случае переход на **Шаг 3**.
3. **Селекция.**
4. **Скрещивание и Мутация.**

5. Если выполнены условия останова алгоритма, то перейти на **Шаг 6**, иначе перейти на **Шаг 2**.
6. **Вернуть** лучшее найденное решение.

Благодаря Шагу 2b имеется возможность удаления выходных нейронов в обучаемой нейронной сети, таким образом, реализуется сокращение размерности пространства признаков. В алгоритме используются специальные процедуры для оценки приспособленности и скрещивания, описываемые далее.

Вычисление приспособленности

Вычисление приспособленности для i -й особи может быть описано следующим алгоритмом:

1. Задать веса ИНС с использованием генов особи.
2. Применить ортогонализацию Грама-Шмидта к векторам весов выходных нейронов ИНС.
3. Вычислить выходные сигналы ИНС для каждого примера из обучающего множества.
4. Вычислить дисперсию выходных сигналов ИНС.
5. Отсортировать выходные нейроны ИНС по убыванию дисперсии.
6. Скопировать полученный вектор весов ИНС обратно в хромосому.

Сортировка выходных нейронов ИНС по дисперсиям сигналов необходима для упорядочивания координатных векторов базиса, задаваемых весами связей ИНС. Таким образом, нейроны сортируются по убыванию вклада в описание обучающих данных. Отметим, что оценка каждой особи может происходить независимо и поэтому поддается распараллеливанию.

Скрещивание

В предлагаемом алгоритме также используется оператор кроссинговера, в котором рассматривается линейная аппроксимация градиента целевой функции.

Пусть необходимо скрестить i -ю и j -ю особь, представляющих комбинации всех весов ИНС (\mathbf{w}_i и \mathbf{w}_j соответственно). Предположим, что i -я особь – более приспособленная. Каждый вектор весов связей разбивается

на N_O непересекающихся частей, где N_O – количество выходных сигналов ИНС:

$$\mathbf{w}_i = \{\mathbf{w}_i^{(k)}, k = 1, \dots, N_O\}.$$

Разбиение производится таким образом, чтобы k -я часть соответствовала весам k -го выходного нейрона. Каждой части $\mathbf{w}_i^{(k)}$ поставлен в соответствие ее вес $v_i^{(k)}$, равный дисперсии проекций примеров обучающих данных на соответствующий координатный вектор.

Скрещивание особей осуществляется по полученным частям. В результате получается один потомок:

$$\mathbf{c}^{(k)} = \mathbf{w}_i^{(k)} + \frac{|v_i^{(k)} - v_j^{(k)}|}{\|\mathbf{w}_i^{(k)} - \mathbf{w}_j^{(k)}\|} (\mathbf{w}_i^{(k)} - \mathbf{w}_j^{(k)}), \quad (3)$$

где $\mathbf{c}^{(k)}$ – k -я часть хромосомы потомка, $\|\mathbf{x}\|$ – евклидова норма вектора \mathbf{x} .

Дробь в (3) можно использовать как аппроксимацию градиента веса k -й части, таким образом, выражение (3) может быть рассмотрено как линейная аппроксимация обновления k -й части для ее улучшения при движении с началом в точке $\mathbf{w}_i^{(k)}$.

Описание экспериментов

Двумя основными задачами экспериментов являются:

1. Важно определить, возможно ли эффективное сокращение размерности пространства признаков.
2. Поскольку МпГК не направлен на вычисление точных координат главных компонент, то необходимо выяснить, как этот факт влияет на точность классификации.

Тестирование предлагаемого метода МпГК будет осуществлено с использованием нескольких задач классификации из набора Proben1 [4], а именно (в скобках указана размерность пространства признаков): *cancer1* (9), *card1* (51), *diabetes1* (8), *glass1* (9), *heart1* (35), *horse1* (58), *soybean1* (82), *thyroid1* (21).

Нейроэволюционный подход используется для сокращения размерности пространства признаков, а полученные преобразованные описания объектов из обучающей выборки применяются для традиционного обучения ИНС с помощью градиентного алгоритма.

Будут рассматриваться следующие значения τ для критерия (2) удаления неинформативных линейных подпространств: 5, 10, 15, 20.

Во время каждого запуска находится решение для МпГК, для которого градиентным методом обучается 10 ИНС без скрытых слоев. Из этих ИНС выбирается одна (победитель), на основании ошибки классификации на проверочном множестве. Выбранная ИНС будет использована при тестировании. Для каждой задачи осуществляется 10 запусков, по результатам которых определяется средняя точность классификации, а также средняя размерность сокращенного пространства признаков.

Длительность нейроразвивающегося поиска решения для МпГК составляет 50 поколений, а ИНС для классификации обучается с применением алгоритма RPROP на протяжении 1000 эпох. Во избежание переобучения ИНС используется ранняя остановка с 10% порогом.

Результаты экспериментов

Значения ошибок классификации, полученные на тестовом множестве, для каждой задачи приведены в табл. 1. Также в табл. 1 в скобках приведено среднее количество выходов ИНС, получаемой в результате предлагаемого метода МпГК для оценки размерности преобразованного признакового пространства.

Таблица 1

Ошибки классификации (%) для различных значений τ

| Задача | $\tau = 5$ | $\tau = 10$ | $\tau = 15$ | $\tau = 20$ |
|---------------|--------------|--------------|--------------|--------------|
| cancer1 (9) | 2,30 (1) | 2,82 (1,2) | 1,78 (4,6) | 1,84 (6,3) |
| card1 (51) | 16,28 (28,5) | 15,41 (50,7) | 15,64 (51) | 15,76 (51) |
| diabetes1 (8) | 24,95 (7,6) | 25,00 (8) | 25,00 (8) | 25,00 (8) |
| glass1 (9) | 36,23 (5,5) | 33,02 (6,7) | 32,07 (7,9) | 32,26 (8,4) |
| heart1 (35) | 21,13 (22,3) | 19,91 (31,5) | 20,00 (34,2) | 20,04 (35) |
| horse1 (58) | 28,79 (35,3) | 29,23 (57,7) | 30,66 (58) | 29,56 (58) |
| soybean1 (82) | 50,65 (3,3) | 20,47 (11,7) | 11,94 (26,1) | 10,47 (37,3) |
| thyroid1 (21) | 7,19 (8,9) | 6,03 (16,3) | 5,87 (18) | 5,92 (19,8) |

В разных задачах изменение параметра τ оказывало различное влияние на количество признаков по МпГК. Тем не менее, из полученных результатов видно, что рекомендуемое значение для τ не должно быть меньше 15.

Для сравнения в табл. 2 содержатся лучшие значения ошибок классификации для рассматриваемых задач, полученные с использованием других методов. Результаты из статьи [4] получены при традиционном обу-

чении ИНС с использованием ручного подбора структуры и алгоритма RPROP; результаты для обучения ИНС с использованием только генетического алгоритма при фиксированной структуре взяты из статьи [5]; а данные из исследования [6] получены для ИНС обученных с применением градиентного метода и алгоритмов контрастирования.

Таблица 2

Значения ошибки классификации (%), полученные для других подходов на наборе Proben1

| Задача | Лучшие значения ошибки классификации, % | | | МпГК, $\tau = 15$ |
|----------------------|---|-------|------|----------------------|
| | [4] | [5] | [6] | |
| <i>cancer1</i> (9) | 1,38 | 1,24 | 1,1 | 1,78 (4,6) |
| <i>card1</i> (51) | 14,05 | 14,27 | 13,7 | 15,64 (51) |
| <i>diabetes1</i> (8) | 24,10 | 23,70 | 20,8 | 25,00 (8) |
| <i>glass1</i> (9) | 32,7 | 47,62 | 30,2 | 32,07 (7,9) |
| <i>heart1</i> (35) | 19,72 | 21,87 | 18,5 | 20,00 (34,2) |
| <i>horse1</i> (58) | 29,19 | 26,44 | 26,9 | 30,66 (58) |
| <i>soybean1</i> (82) | 9,06 | 8,47 | N/A | 11,94 (26,1) |
| <i>thyroid1</i> (21) | 2,32 | 6,12 | 5,7 | 5,87 (18) |

Сравнение результатов показывает, что для многих рассматриваемых задач снижение размерности пространства признаков, полученной с помощью МпГК, дает точность классификации, сопоставимую с ИНС, обученных более «традиционными» методами. При этом в ряде задач наблюдается существенное сокращение количества признаков (задачи *cancer1* и *soybean1*).

На рис. 1 и 2 показаны примеры изменения средней размерности пространства признаков (количество оставшихся компонент МпГК) и дисперсии проекций на оценки наиболее информативных собственных векторов в зависимости от количества поколений. Данные усреднены по 100 запускам.

На основании полученных результатов можно сделать следующий весьма интересный вывод:

Существуют случаи, когда нет необходимости определять точные координаты собственных векторов матрицы автоковариации обучающих данных для снижения размерности пространства признаков в задаче классификации.

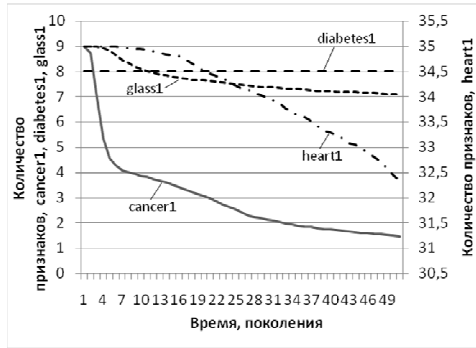


Рис. 1. Изменение средней размерности пространства признаков для задач *cancer1*, *diabetes1*, *glass1* и *heart1* при $\tau = 10$.

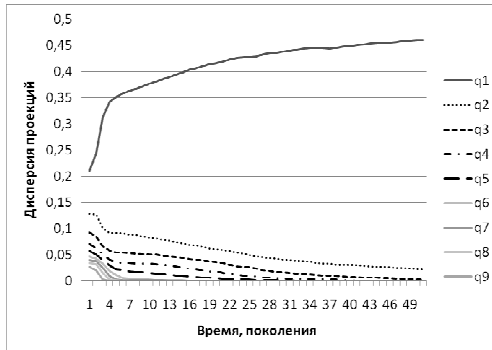


Рис. 2. Изменение средней дисперсии проекции данных на оценки собственных векторов для задачи *cancer1*, $\tau = 10$.

Представляет интерес вопрос о том, какова цена за использование приближенных главных компонент. В случае, если отбрасываемые линейные подпространства были определены точно, преобразованные описания объектов будут находиться в «корректном» линейном подпространстве, и поэтому существует линейное преобразование, которое переводит оставшиеся оценки собственных векторов к точным решениям МГК. Однако если сокращение размерности было произведено с погрешностью, то существует ошибка в определении значений признаков объектов. Измерение этой ошибки представляется нетривиальной задачей, поскольку точные значения собственных векторов полагаются неизвестными. Эта задача

требует отдельного изучения, и позволит глубже понять свойства МпГК, а также разработать более эффективные методы сокращения размерности.

Заключение

В статье представлен оригинальный способ сокращения размерности пространства признаков путем реализации метода псевдо-главных компонент. Метод основан на применении нейроэволюционного алгоритма для обучения ИНС с линейными функциями активации без скрытых слоев. В алгоритме используются специальные процедуры для вычисления приспособленности и скрещивания особей.

Результаты исследования показали, что несмотря на то, что сокращение размерности пространства признаков было совершено с некоторой ошибкой, полученная точность классификации сравнима с результатами обучения ИНС, обученных более «традиционными» способами.

Интересным вопросом для будущих исследований является распараллеливание алгоритма МпГК и анализ ошибки при вычислении собственных векторов. Также необходимо сравнить полученные результаты с результатами классификации при использовании классического МГК, при идентичном изменении размерности пространства признаков.

Список литературы

1. Gorban A., Kegl B., Wunsch D., Zinovyev A. (Eds.) Principal Manifolds for Data Visualisation and Dimension Reduction. LNCSE 58. Berlin, Heidelberg, New York: Springer, 2007.
2. Golub G.H., Van Loan C.F. Matrix Computations. Second Edition. Baltimore, London: The John Hopkins University Press, 1989.
3. Oja E. Simplified neuron model as a principal component analyzer // Journal of Mathematical Biology. 1982. Vol. 15. No. 3. P. 267-273.
4. Prechelt L. PROBEN1 – a set of neural network benchmark problems and benchmarking rules. Technical Report 21/94. Universitat Karlsruhe, Karlsruhe, Germany, 1994.
5. Barrera F.Y. Búsqueda de la estructura óptima de redes neurales con Algoritmos Genéticos y Simulated Annealing. Verificación con el benchmark PROBEN1 // Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. 2007. Vol. 11. No. 34. P. 41-61.
6. Weman N. Empirical Investigation of the Effect of Pruning Artificial Neural Networks With Respect to Increased Generalization Ability. Linköpings universitet (Sweden), 2010.