

УДК 681.518:004.93.1

ПОНИЖЕНИЕ РАЗМЕРНОСТИ ВХОДНЫХ ГРАФИЧЕСКИХ ДАННЫХ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ

М.С. Бабий, канд. техн. наук, доцент;
А.П. Чекалов, канд. техн. наук, доцент,
Сумский государственный университет, г. Сумы

Запропоновано алгоритм зниження розмірності графічних зображень на основі методу головних компонент. Розроблено програму перекладу графічних файлів у стиснутий формат з мінімальною втратою кількості інформації. Отримані файли використані під час навчання багатошарової нейронної мережі зворотного поширення з метою економії оперативної пам'яті і зменшення часу навчання. На прикладі ORL-бази зображень осіб підтверджена ефективність використання стиснутих файлів при використанні метричних способів розпізнавання.

Ключові слова: метод головних компонент, графічні зображення, зниження розмірності, розпізнавання.

Предложен алгоритм понижения размерности графических изображений на основе метода главных компонент. Разработана программа перевода графических файлов в сжатый формат с минимальной потерей количества информации. Полученные файлы использованы при обучении многослойной нейронной сети обратного распространения с целью экономии оперативной памяти и уменьшения времени обучения. На примере ORL-базы изображений лиц подтверждена эффективность использования сжатых файлов при использовании метрических способов распознавания.

Ключевые слова: метод главных компонент, графические изображения, понижение размерности, распознавание.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Распознавание графических изображений широко используется в различных областях науки, техники и общественной деятельности: физике, биологии, медицине, системах компьютерного зрения, обработке документов, военной области.

Основной проблемой при обработке изображений является большой объем, требуемый для хранения и обработки графической информации, что приводит к большим затратам оперативной памяти и замедляет время обработки.

Задачей настоящего исследования является разработка практических способов понижения размерности графических изображений с последующей проверкой результатов на конкретных технологиях распознавания.

В качестве объекта исследования взяты изображения лиц, которые обладают достаточно высокой изменчивостью для одного и того же индивидуума. Кроме того, системы распознавания по изображению лица имеют и практическую ценность, т.к. в отличие от других

биометрических систем не требует дорогостоящего специального оборудования, в простейшем случае достаточно обычного компьютера, снабженного Web-камерой.

РЕАЛИЗАЦИЯ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Одним из основных способов уменьшить размерность данных, потеряв при этом наименьшее количество информации, является метод главных компонент [1].

Пусть имеется обучающая выборка из K изображений высотой h , шириной w и размером $n=hw$ пикселей. Представим отдельное изображение с номером k как вектор $\mathbf{x}_k = (x_{k1}, \dots, x_{ki}, \dots, x_{kn})$ в n -мерном линейном пространстве. Будем считать его отдельной реализацией случайного вектора \mathbf{X} .

Ковариационная матрица вектора \mathbf{X} представляет собой матрицу \mathbf{C} размером $n \times n$ с элементами

$$C_{ij} = \text{cov}(X_i, X_j) = \mathbf{M}[(X_i - \mu_i)(X_j - \mu_j)] = \\ = \frac{1}{K} \sum_{k=1}^K [(x_{ki} - \mu_i)(x_{kj} - \mu_j)],$$

где

$$\mu_i = \mathbf{M}(X_i) = \frac{1}{K} \sum_{k=1}^K x_{ki}.$$

Матрица ковариации является обобщением дисперсии для многомерной случайной величины, а её след – скалярным выражением дисперсии. Метод главных компонент состоит в переходе к новому ортогональному базису, оси которого ориентированы по направлениям максимальной дисперсии набора входных данных. Первая ось нового базиса выбирается в направлении максимальной дисперсии, вторая ось – в направлении максимальной дисперсии в подпространстве, ортогональном первой оси, последняя ось будет иметь минимальную дисперсию. Последние координаты, соответствующие малым дисперсиям, обычно отбрасываются и таким образом осуществляется переход к подпространству меньшей размерности.

Суть метода можно пояснить на примере двух переменных, между которыми существует корреляция. Определив главную компоненту как направление, в котором наблюдается наибольший разброс данных, мы теряем минимум информации об отличии объектов друг от друга. Для нескольких переменных принцип определения главных компонент тот же.

Векторы главных компонент для задачи о поиске ортогональных проекций с наибольшим рассеянием – это ортонормированный набор собственных векторов ковариационной матрицы \mathbf{C} , расположенных в порядке убывания собственных значений. Ковариационная матрица в базисе из собственных векторов диагональная, и в этом базисе ковариация между различными координатами равна нулю.

Определим матрицу $\mathbf{A} = (\mathbf{x}_1 - \mathbf{M}(\mathbf{X}), \dots, \mathbf{x}_K - \mathbf{M}(\mathbf{X}))$ размером $n \times K$. Тогда ковариационную матрицу можно записать в виде

$$\mathbf{C} = \mathbf{M}[(\mathbf{X} - \mathbf{M}(\mathbf{X}))(\mathbf{X} - \mathbf{M}(\mathbf{X}))^T] = \\ = \frac{1}{K} \sum_{k=1}^K [(\mathbf{x}_k - \mathbf{M}(\mathbf{X}))(\mathbf{x}_k - \mathbf{M}(\mathbf{X}))^T] = \frac{1}{K} \mathbf{A}\mathbf{A}^T.$$

Нам необходимо найти собственные векторы матрицы \mathbf{C} . Они будут такими же, как и у матрицы $\mathbf{D} = \mathbf{K}\mathbf{C} = \mathbf{A}\mathbf{A}^T$, имеющей ту же размерность $n \times n$. Численное нахождение собственных векторов и значений такой матрицы из-за большого размера довольно затруднительно. Как отмечено в [2], вместо этого можно находить собственные векторы и значения матрицы $\mathbf{B} = \mathbf{A}^T\mathbf{A}$, имеющей размерность $K \times K$. В этом случае справедлив следующий вывод:

$$\mathbf{A}^T \mathbf{A} \mathbf{u}_k = \lambda_k \mathbf{u}_k \Rightarrow \mathbf{A} \mathbf{A}^T \mathbf{A} \mathbf{u}_k = \lambda_k \mathbf{A} \mathbf{u}_k \Rightarrow \mathbf{D} \mathbf{A} \mathbf{u}_k = \lambda_k \mathbf{A} \mathbf{u}_k$$

или $\mathbf{D} \mathbf{v}_k = \lambda_k \mathbf{v}_k$, где $\mathbf{v}_k = \mathbf{A} \mathbf{u}_k$, λ_k – собственные значения, $\mathbf{u}_k, \mathbf{v}_k$ – собственные векторы. Из вывода следует, что $\mathbf{A} \mathbf{A}^T$ и $\mathbf{A}^T \mathbf{A}$ имеют те же самые собственные значения, а их собственные векторы связаны формулой $\mathbf{v}_k = \mathbf{A} \mathbf{u}_k$. При этом K собственных значений матрицы $\mathbf{A}^T \mathbf{A}$ соответствуют K большим собственным значениям матрицы $\mathbf{A} \mathbf{A}^T$.

Собственные векторы и значения матрицы \mathbf{B} будем находить методом Якоби. Данный метод является более медленным, чем QR-алгоритм. Однако он всегда устойчив для действительных симметричных матриц и проще в реализации. А так как размер матрицы у нас теперь значительно меньше, то время выполнения не является критическим.

Метод состоит из последовательности ортогональных преобразований подобия исходной матрицы, называемых ротациями Якоби. Каждая ротация – это плоский поворот с целью обнуления одного из недиагональных элементов матрицы.

Базовая матрица ротации имеет вид

$$\mathbf{P}_{pq} = \begin{pmatrix} 1 & & & & & \\ & \dots & & & & \\ & & c & \dots & s & \\ & & \dots & 1 & \dots & \\ & & -s & \dots & c & \\ & & & & & \dots \\ & & & & & & 1 \end{pmatrix} .$$

Её диагональные элементы равны 1, за исключением двух в строках p и q , а недиагональные равны нулю, за исключением двух, равных s и $-s$. Величины c и s – соответственно косинус и синус угла ротации. Ротация преобразует матрицу \mathbf{B} по формуле $\mathbf{B} = \mathbf{P}_{pq}^T \mathbf{B} \mathbf{P}_{pq}$. При этом $\mathbf{P}_{pq}^T \mathbf{B}$ меняет только строки p и q матрицы \mathbf{B} , а $\mathbf{B} \mathbf{P}_{pq}$ только столбцы с номерами p и q . Матричным преобразованиям соответствуют формулы

$$\begin{aligned} b'_{rp} &= cb_{rp} - sb_{rq}, \quad b'_{rq} = cb_{rq} + sb_{rp}, \quad (r \neq p, r \neq q) \\ b'_{pp} &= c^2 b_{pp} + s^2 b_{qq} - 2csb_{pq}, \quad b'_{qq} = s^2 b_{pp} + c^2 b_{qq} + 2csb_{pq}, \\ b'_{pq} &= (c^2 - s^2)b_{pq} + cs(b_{pp} - b_{qq}). \end{aligned}$$

Угол ротации определяется из условия $b'_{pq} = 0$. После выполнения преобразований получается диагональная матрица $\mathbf{F} = \mathbf{U}^T \mathbf{B} \mathbf{U}$, где $\mathbf{U} = \mathbf{P}_1 \mathbf{P}_2 \mathbf{P}_3 \dots$ – произведение матриц ротации. Диагональные элементы матрицы \mathbf{F} являются собственными значениями исходной матрицы \mathbf{B} , а столбцы \mathbf{u}_i матрицы \mathbf{U} являются собственными векторами.

Выбираем L собственных векторов, соответствующих наибольшим собственным значениям. После этого по формуле $\mathbf{v}_l = \mathbf{A}\mathbf{u}_l$ находим L собственных векторов \mathbf{v}_l , которые после нормализации по формуле $\mathbf{e}_l = \frac{\mathbf{v}_l}{\|\mathbf{v}_l\|}$ становятся базисом для пространства главных компонент.

Проекции \mathbf{y}_k центрированных изображений, представленных векторами $\mathbf{x}_k - \mathbf{M}(\mathbf{X})$, в подпространство главных компонент вычисляются как произведение $\mathbf{y}_k = (\mathbf{x}_k - \mathbf{M}(\mathbf{X}))^T (\mathbf{e}_1, \dots, \mathbf{e}_L)$.

В соответствии с данным алгоритмом написана программа на языке Visual C++. Для облегчения работы с изображениями дополнительно использована библиотека компьютерного зрения OpenCV v2.0, скомпилированная для среды Visual Studio 2008.

ПРИМЕНЕНИЕ ДЛЯ НЕЙРОННОЙ СЕТИ

Изображения, сжатые по данной программе, были использованы как входные данные для обучения многослойной нейронной сети обратного распространения (рис.1).

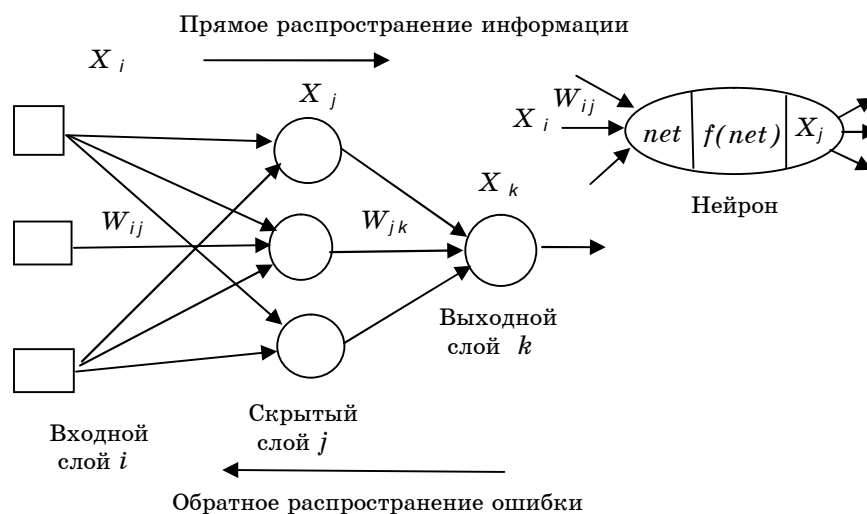


Рисунок 1 – Нейронная сеть обратного распространения

Нейроны в данной сети располагаются по слоям, причем элементы слоя n связаны только с узлами слоя $n+1$. При многослойной обработке сигнала ошибка на выходе сети может быть вызвана сложными процессами внутри нее, поэтому источник ошибки в выходном слое сети необходимо анализировать в комплексе. Вводится понятие поверхности ошибки, которая представляет кумулятивную ошибку на всем наборе данных, как функцию от весовых коэффициентов w сети. Каждая возможная конфигурация весов определяет точку поверхности ошибки. Имея определенную конфигурацию весов, с помощью градиентного метода можно найти направление на этой поверхности, вдоль которого происходит наиболее быстрое уменьшение функции ошибки. Настройку весов с учетом многослойной структуры нейронной сети удобнее всего выполнить способом обратного распространения ошибки. Для нейронов выходного слоя величина ошибки вычисляется как разность между ожидаемым t_k и реальным выходным значением x_k . Итерационные

формулы вычисления весовых коэффициентов для выходного слоя имеют вид

$$w_{jk} \leftarrow w_{jk} + \Delta w_{jk};$$

$$\Delta w_{jk} = c \delta_{ok} x_j; \quad \delta_{ok} = x_k(1 - x_k)(t_k - x_k),$$

где c – коэффициент скорости обучения.

Формулы вычисления весовых коэффициентов для узлов скрытого слоя вычисляются с использованием активационной функции, в качестве которой взята логистическая функция $f(net) = 1/(1 + \exp(-net))$, для

узла с индексом j $net_j = \sum_{i=1}^n w_{ij} x_i$. Соответствующие формулы вычисления

весовых коэффициентов для скрытого слоя имеют вид

$$w_{ij} \leftarrow w_{ij} + \Delta w_{ij};$$

$$\Delta w_{ij} = c \delta_{hj} x_i; \quad \delta_{hj} = x_j(1 - x_j) \sum_k w_{jk} \delta_{ok}.$$

Программа для обучения сети в качестве входных данных использует графические изображения в формате PGM. Каждый пиксель в этом формате занимает один байт и представляет собой градацию серого цвета. Соответственно для однообразного представления входных данных для нейронной сети в программе понижения размерности было предусмотрено преобразование компонент сжатого изображения в формат PGM.

Тестирование программы показало, что при обучении сети по исходному и по сжатому изображению отсутствует прямая связь между количеством эпох (проходов по всему обучающему множеству с последующей проверкой на контрольном множестве). Для сжатого изображения их может быть как меньше, так и больше. Тем не менее среднестатистическое время обучения для сжатого изображения значительно меньше.

ПРИМЕНЕНИЕ ДЛЯ МЕТРИЧЕСКИХ МЕТОДОВ РАСПОЗНАВАНИЯ

Под метрическими методами будем понимать методы распознавания, основанные на измерении расстояний между представителями классов в метрическом пространстве.

Для задачи распознавания была взята база данных ORL [3], содержащая 400 фотографий лиц в формате PGM, по 10 фотографий для каждого из 40 человек в соответствующих директориях от s1 до s40. Размер одного изображения 112x92 пикселей, и непосредственная работа с 10 тысячами признаков вызывает ряд вычислительных трудностей.

Выходные данные программы понижения размерности в целях сохранения точности не переводились в графический формат, а использовались их значения в вещественном формате.

В первом варианте тестирования в качестве обучающего набора были взяты 50 изображений 1.pgm – 5.pgm в директориях s1 – s10, а в качестве тестового – изображения 6.pgm – 10.pgm в тех же директориях. Принадлежность к классу определялась по ближайшему соседу, расстояние до которого минимально. В результате тестирования правильно были распознаны 49 изображений тестового набора, за исключением s10/10.pgm, которое было отнесено к классу s8.

Во втором варианте обучающий и тестовый набор были поменяны местами. В результате тестирования правильно были распознаны также 49 изображений тестового набора, за исключением s3/5.pgm, которое было отнесено к классу s5.

Естественно, что неправильное распознавание изображения может быть вызвано его реальными особенностями, а не погрешностью метода. Таким образом, тестирование показало высокую точность сжатия с минимальной потерей информации при переходе от 10304 признаков к 50. Таким образом, разработанная программа понижения размерности изображений может быть рекомендована для использования в задачах распознавания.

ВЫВОДЫ

1. 1 Предложен алгоритм и разработана программа понижения размерности графических изображений с минимальной потерей количества информации на основе метода главных компонент.

2. 2 Результаты работы программы использованы при обучении многослойной нейронной сети обратного распространения. Расчёты показали значительное уменьшение времени обучения сети при экономии оперативной памяти.

3. 3 На примере ORL-базы изображений лиц подтверждена практически полная сохранность количества информации при переходе от исходного к сжатому изображению.

Данная программа может быть рекомендована для сжатия входной графической информации в задачах распознавания.

SUMMARY

REDUCTION OF DIMENSIONALITY THE OUTPUT GRAPHIC DATA FOR RECOGNITION TASK

M.S. Babiy, O.P. Chekalov
Sumy State University, Sumy

There is offered algorithm of reduction the dimensionality for the graphic images on base of principal component analysis. Program of translation the graphic files to compressed format with minimal information loss is designed. The received files are used in learning the multilayer backpropagation neural network to reduce operative memory and running time. On example ORL database of faces is confirmed efficiency of the use the compressed files in metric recognition methods.

Key words: *principal component analysis, graphic images, reduction of dimensionality, recognition.*

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Л. Прикладная статистика. Классификация и снижение размерности / . –М.:Финансы и статистика, 1989.
2. Turk M., Pentland A. Eigenfaces for Recognition; Vision and Modeling Group, The Media Laboratory, Massachusetts Institute of Technology; September 1990.
3. The ORL Database of Faces.
http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.tar.Z.

Поступила в редакцию 30 апреля 2010 г.