

ТЕМАТИЧЕСКАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ СМЕСЕВЫХ МОДЕЛЕЙ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ

И.А. Колтунов, Я.Г. Великая

Введение

Применение смесевых моделей вероятностных распределений для классификации объектов, распознавания образов и обнаружения целей имеет ряд особенностей по сравнению с десятками других методов классификации, распознавания и обнаружения. Эти особенности (некоторые из них являются уникальными) делают смесевые модели самым эффективным современным инструментом автоматического компьютерного распознавания.

1. Смесевая модель для изображений, получаемых дистанционным зондированием земных образований, является естественной физической моделью статистического описания данных, а не математической аппроксимацией сложного распределения набором простых компонент.

2. В рамках одной модели (смесевой модели гауссовых распределений) решаются все известные задачи тематической обработки изображений: сегментация, распознавание с обучением, обнаружение объектов, выбор распознающих признаков, оптимизация алгоритма классификации.

3. Смесевая модель дает так называемое полное семейство алгоритмов распознавания, включающее от простейших методов до как угодно сложных, из которых выбирается нужный оптимальный метод и алгоритм, наилучшим образом решающий поставленную задачу. Такими же свойствами обладают и другие известные семейства алгоритмов. Например, алгебраические семейства [1], семейство разделяющих поверхностей [2], нейронные сети (большинство прикладных работ). Однако ограничительные особенности первых двух семейств [1,2] делают их проблематичными для практических применений. Нейронные сети, эффективно работая на объектах обучающей выборки, оставляют много эвристических загадок для прогноза, не гарантируя надежности и достоверности классификации каждому вновь предъявляемому объекту [3].

4. В настоящее время только смесевая модель, поставляя именно статистическое семейство алгоритмов, наделяет каждый классифицируемый объект количественной характеристикой надежности.

5. Смесевая модель позволяет каждому объекту оставаться в условиях субъективной неуверенности и объективной неопределенности.

§ 1. Формулировки и методы решения задач тематической обработки

Назовем многомерным изображением векторную функцию от двух переменных, заданную в узлах прямоугольной сетки. Оцифрованный по прямоугольной сетке многоканальный снимок участка земной поверхности также является многомерным изображением. При дистанционном зондировании Земли значения функции, заданной в узлах прямоугольной сетки, представляют собой векторы интенсивностей излучения соответствующих элементарных участков земной поверхности в установленных спектральных диапазонах.

Будем считать, что значение вектора интенсивностей для одного элемента изображения является реализацией векторной случайной величины, распределение которой определяется физическими процессами, протекающими на поверхности участка Земли, соответствующего этому элементу.

1.1. Три постановки задачи классификации

Пусть генеральная совокупность всех возможных наблюдений случайного вектора ξ имеет плотность распределения $P(\xi)$, а наблюдения класса k распределены с плотностью $P_k(\xi) = P(\xi/k)$. Тогда вероятность $p(k/x)$ принадлежности наблюдения x классу k определяется равенством [4]:

$$p(k/x) = \frac{p_k P_k(x)}{P(x)}$$

где p_k - априорная вероятность класса k .

Плотность распределения $P(\xi)$ случайной величины ξ представляется параметрическим семейством функций заданного вида с неизвестным вектором параметров α :

$$P(\xi) = f(\xi, \alpha),$$

причем распределению каждого класса соответствует определенное значение вектора α . Вероятности принадлежности наблюдений классам находятся по формуле

$$p(k/x) = \frac{p_k f(x, \alpha_k)}{f(x, \alpha)} \quad (1)$$

Вид семейства функций для аппроксимации плотностей выбирается с учетом априорной информации о распределениях наблюдаемых величин и должен удовлетворять требованию полноты в функциональном пространстве распределений, т.е. каждое возможное распределение может быть с любой наперед заданной точностью описано распределением из выбранного семейства.

а) Классификация данных с обучением. Вероятности принадлежности наблюдения классам находятся в предположении полноты пространства гипотез, т.е. каждое наблюдение принадлежит к одному из C заданных классов. Значения прогнозных вероятностей определяются формулой Байеса

$$p(k/x) = \frac{p_k \cdot f(x, \alpha_k)}{\sum_{j=1}^c p_j \cdot f(x, \alpha_j)}, \quad \sum_{k=1}^c p(k/x) = 1 \quad (2)$$

Параметры распределений $\alpha_k, k = 1, 2, \dots, c$ оцениваются с помощью обучающей выборки $\{x^{(i)}\}_{i=1}^n$. Для каждого элемента выборки заданы отношения принадлежности $\rho_{ki}, k = 1, 2, \dots, c$ каждому из классов. При вероятностной постановке задачи распознавания отношения принадлежности определяются вероятностями принадлежности

$$\rho_{ki} = \rho(k / x^{(i)}); \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots, c$$

$$\sum_{k=1}^c \rho_{ki} = 1; \quad \rho_k = \sum_{i=1}^n \rho_{ki} > 0,$$

где ρ_k - объем обучающей выборки класса k .

Тогда, применяя формулу полной вероятности, можно считать $x^{(i)}$ выборочным значением случайной величины с плотностью распределения

$$P(x^{(i)}) = \sum_{k=1}^c \rho_{ki} \cdot f(x^{(i)}, \alpha_k) \quad (3)$$

б) Автоматическая классификация данных. Вероятности принадлежности наблюдения каждому классу определяются формулой Байеса (2). Решается задача разделения смеси распределений с плотностями $f(x, \alpha_k)$, $k = 1, 2, \dots, c$. Параметры смеси p_k, α_k , $k = 1, 2, \dots, c$ находятся с помощью исходной выборки $\{x^{(i)}\}_{i=1}^n$ элементы $x^{(i)}$ которой распределены с плотностью

$$P(x^{(i)}) = \sum_{k=1}^c p_k f(x^{(i)}, \alpha_k) \quad (4)$$

в) Классификация данных с частичным обучением. Распознавание производится в предположении непротиворечивости гипотез о принадлежности наблюдения разным классам, т.е. каждое наблюдение может принадлежать нескольким классам, одному из классов или не входить ни в один из заданных классов. В этом случае решается существенно другая задача распознавания - не разделение элементов на классы, а выделение элементов заданных классов. Такая ситуация характерна при выделении объектов на фоне шумов, поиске полезных ископаемых различных типов на одном полигоне и т.п.

Непротиворечивость гипотез о принадлежности ведет к необходимости формирования обучающих выборок по каждому классу в отдельности с условием четкого определения принадлежности элементов выборки соответствующему классу. Значения вероятностей $p(k / x)$ определяются формулой (1). Параметры распределения всей совокупности данных определяются по исходной выборке $\{x^{(i)}\}_{i=1}^n$, класса k - по обучающей выборке $\{x^{(i)}\}_{i=1}^{n_k}$, причем одно наблюдение может входить в обучающие выборки разных классов. Отношения принадлежности ρ_{ki} принимают два значения: $\rho_{ki} \in \{0; 1\}$; $i = 1, \dots, n_k$; $k = 1, \dots, c$

1.2. Смесевые модели для описания плотностей распределения элементов каждого класса

При классификации с обучением (полным или частичным) плотность распределения $f(x, \alpha_k)$ элементов одного класса может быть найдена в виде многокомпонентной функции

$$f(x, \alpha_k) = \sum_{i=1}^{L_k} r_{ki} \cdot g(x, \alpha_{ki}) \quad (5)$$

Число L_k компонент $g(x, \alpha_{ki}), i = 1 \dots L_k$ в распределениях, отвечающих разным классам, в общем случае различно.

При классификации с самообучением плотность распределения $f(x, \alpha_k), k = 1, \dots, c$ не может быть многокомпонентной, так как нет априорных данных (обучающей выборки), позволяющих объединить несколько компонент в один класс. Заметим, что компоненты распределения могут быть и полимодальными, но количество мод и их "форма" должны быть заданы видом аппроксимирующей функции. Одна из основных проблем распознавания состоит в определении сложности аппроксимирующей модели для плотности распределения - нахождении числа компонент в распределениях при классификации с обучением и нахождении числа классов при автоматической классификации.

1.3. Оценка параметров распределения

Традиционным способом оценивания параметров распределений по обучающим выборкам является метод максимального правдоподобия. Для выборки $\{x^{(i)}\}_{i=1}^n$, плотность распределения (3), (4) элементов которой представлена в параметрическом виде $P(x^{(i)}) = f(x^{(i)}, \alpha)$, строится функция правдоподобия $l(\alpha)$:

$$l(\alpha) = \sum_{i=1}^n l_n[f(x^{(i)}, \alpha)] \quad (6)$$

Максимуму этой функции соответствуют "наиболее вероятные" значения параметров α , которые называются оценками максимального правдоподобия (ОМП).

Метод требует независимости выборки и ее однородности (т.е. каждое измерение заслуживает одинаковой степени доверия). Условие независимости для большинства реальных измерений не выполняется. Например, при дистанционных исследованиях земной поверхности и анализе радиосигналов в текущем времени существенная информация об объектах заложена в текстуре (пространственной или временной корреляции) получаемых данных. Учесть такого рода информацию можно путем формирования текстурных признаков для каждого участка изображения. Обычно с этой целью либо проводится частотный анализ изображений (расчет коэффициентов Фурье, Уолша и т.п.), либо текстурные признаки формируются экспертным методом. Использование полученных признаков в качестве исходных данных дает возможность существенно уменьшить потери информации при классификации в предположении независимости наблюдений.

1.4. Оценка точности в определении параметров распределений

Точность определения параметров распределений, а следовательно и прогнозных вероятностей принадлежности существенно зависит от многих факторов: объема обучающих выборок, размерности признакового пространства, сложности аппроксимирующей модели и т.д. Следовательно, результатом классификации должны быть не только прогнозные вероятности принадлежности, но и оценки для ошибок вычисления этих вероятностей. Асимптотическую оценку ковариационной матрицы T для ОМП параметров распределения можно получить, используя информационную матрицу Фишера I [5,6,7]

$$T \approx I^{-1} \quad (7)$$

Если генеральная совокупность наблюдений случайной величины ξ имеет плотность распределения $f(\xi, \alpha)$ и по n независимым наблюдениям найден вектор $\hat{\alpha}$ ОМП параметров α , то элемент матрицы Фишера приближенно определяется интегралом

$$I_{ij} \approx n \int \frac{f_{\alpha_i}(\xi, \hat{\alpha}) \cdot f_{\alpha_j}(\xi, \hat{\alpha})}{f(\xi, \hat{\alpha})} d\xi, \quad i, j = 1, \dots, M, \quad (8)$$

где α_i и α_j – компоненты вектора α , M – размерность вектора α , $f_{\alpha_i}(\xi, \hat{\alpha})$ – производная от плотности по параметру α_i в точке $\hat{\alpha}$.

Интегрирование ведется по области значений $\{\xi \cdot f(\xi, \hat{\alpha}) > 0\}$, приближенность определения I_{ij} объясняется подстановкой в подынтегральную функцию ОМП параметров вместо их истинных значений.

Оценки ошибок в определении параметров используются для выбора числа компонент $L_k, k = 1, \dots, c$ в распределениях (5) при классификации с обучением и числа классов в смеси распределений (4) при автоматической классификации, а также для нахождения оценок точности вычисления прогнозных вероятностей принадлежности.

Ошибки вычисления вероятностей $p(k/x)$ оцениваются в предположении линейности изменения вероятности как функции (1), (2) от параметров $\alpha_k, k = 1, \dots, c$ в окрестности значений α_k , полученных методом максимального правдоподобия.

1.5. Критерий качества классификации

Определим разделяющую силу пространства признаков $X(x_1, \dots, x_m)$ как средневзвешенное значение вероятности правильной классификации всех возможных наблюдений

$$R(X) = \sum_{k=1}^c w_k \int p(k, \xi) \cdot p(\xi / k) d\xi, \quad \sum_{k=1}^c w_k = 1 \quad (9)$$

где $w_k, k = 1, \dots, c$ - весовые коэффициенты, пропорциональные ценам ошибок классификации элементов каждого класса.

Функционал $R(X)$ неубывающий: $\forall X, Y: X \subset Y, R(X) < R(Y)$, т.е. разделяющая сила пространства признаков растет с увеличением размерности. При оценивании вероятностей $p(k/\xi)$ на основании выборочных наблюдений и приближенном расчете интеграла в формуле (9) суммированием по элементам той же обучающей выборки значение оценки \hat{R} функционала R получается завышенным. Величина этого так называемого "оптимистического смещения" определяется точностью нахождения вероятностей принадлежности. Существующие в настоящее время методы устранения смещения с использованием контрольных наблюдений имеют ряд серьезных проблем: использование части априорной информации для контроля ведет к уменьшению обучающей выборки; в зависимости от случайного разбиения тестовых данных на обучение и контроль могут быть получены разные оценки качества классификации; методы типа "скользящего экзамена" вызывают необходимость многократного определения параметров распределений. Кроме того, даже при получении несмещенных оценок для $p(k/\xi)$ функционал \hat{R} не может полностью характеризовать качество проведенной классификация - исследователя обычно больше интересует не средняя вероятность правильной классификации (или средняя вероятность ошибки, если задано решающее правило), а значения $p(k/\xi), k = 1, \dots, c$ в каждой наблюдаемой точке ξ , которые могут быть вычислены со значительными ошибками. Поэтому в функционале качества классификации предлагается использовать нижние доверительные границы $p_n(k, \xi)$ для оценок $p(k, \xi)$ вероятностей принадлежности $p(k/\xi)$

$$p_n(k, \xi) = p(k, \xi) - \gamma \cdot \Delta p(k, \xi) \quad (10)$$

где $\Delta p(k, \xi)$ - оценка среднеквадратичной ошибки вычисления $p(k/\xi)$, γ - коэффициент учета ошибки, определяющийся необходимой степенью доверия к найденным значениям вероятностей.

Учет ошибок вычисления вероятностей принадлежности позволяет получить ($\gamma = 1$) оценку снизу для средней прогнозной вероятности правильной классификации при любом «оптимистическом смещении».

1.6. Определение признаков для классификации

Качество классификации, основанной на оценивании параметров распределения обучающей выборки, в значительной степени зависит от размерности пространства признаков, в котором проводится классификация. Действительно, при заданном объеме обучающей выборки с ростом размерности пространства признаков увеличиваются ошибки в оценках параметров, а следовательно, если "расстояния" между классами увеличились незначительно, снижается качество классификации. Различные наборы исходных признаков обычно не равноценны по своим диагностическим способностям, поэтому выбор признакового пространства - одна из основных проблем распознавания.

В задачах распознавания с большим числом классов оптимальное пространство признаков следует выбирать по каждому классу в отдельности. Разделяющая сила пространства X при выделении класса k в этом случае определяется функционалом (9) с отличным от нуля значением w_k и нулевыми $w_i, i \neq k$. Выбор единого пространства для выделения всех классов $w_i > 0, i = 1, \dots, c$ в общем случае приводит к

менее высокому качеству классификации. Если число классов равно двум, а задача распознавания решается в предположении полноты пространства гипотез, оба варианта дают одинаковый результат.

Рассмотрим два подхода к проблеме выбора оптимального пространства признаков. Первый состоит в использовании для классификации набора признаков из множества исходных. Такой подход универсален (не зависит от вида функций, аппроксимирующих плотности распределения) и применяется как при классификации наблюдений, так и в задачах регрессионного анализа. В общем случае для отыскания оптимального по некоторому критерию набора признаков необходим полный перебор всех подмножеств исходных признаков. В экспериментах с большим числом признаков такой перебор невозможен (методы типа ветвей и границ обычно не дают значительного сокращения числа проверяемых подмножеств), поэтому существующие методы выбора признаков либо локально оптимальны, либо дают оптимальное решение при существенных ограничениях на распознающие свойства признаков.

В настоящей работе предлагается комбинированный метод: сначала выбирается оптимальный набор признаков в предположении независимости вклада признаков в функционал качества классификации, а затем этот набор уточняется пошаговым перебором. Предположение о независимости состоит в том, что для наборов одинаковой величины (одного уровня на дереве перебора) каждый признак вносит в значение функционала качества вклад, не зависящий от остальных признаков, входящих в набор. Тогда, обозначив вклад (информативность) признака с номером i переменной y_i , можно для каждого набора записать уравнение с неизвестными значениями y_i , $i = 1, \dots, M$ (M - количество исходных признаков)

$$\sum_{i=1}^M a_i y_i = F, \quad (11)$$

где F - значение функционала качества, a_i принимает значение либо 1 (если признак с номером i входит в набор), либо 0 (признак не входит в набор). Вектор информативностей признаков находится решением системы таких уравнений. Набор, составленный из наиболее информативных признаков, используется в качестве начального приближения для пошагового перебора.

Второй подход к проблеме выбора пространства признаков состоит в поиске оптимального набора функций от всех исходных признаков. Такой подход может дать существенно лучшие результаты, чем первый, однако он основан на использовании информации о виде распределений, описывающих классы, и требует более сложного математического аппарата. В частности, представляется перспективным поиск линейных комбинаций исходных признаков, разделяющих классы с нормальными или полинормальными распределениями [8].

1.7. Обработка наблюдений с дискретными распределениями

Описанная методика распознавания предназначена для обработки наблюдений, распределения которых непрерывны; классификация данных по дискретно распределенным признакам должна проводиться специальными методами. Существенные сложности возникают при обработке, если часть признаков имеет непрерывное распределение, а часть - дискретное. При аппроксимации плотностей распределения многокомпонентными функциями возникает необходимость преобразования дискретных признаков в непрерывные [9]. Такое преобразование выполняется случайной раскачкой обучающих выборок. Если амплитуда раскачки

значительно меньше минимальной разницы между двумя соседними значениями дискретного признака, качество классификации не изменится, несмотря на неточность вычисления параметров распределения. При этом результаты распознавания практически не зависят от закона распределения, по которому производится раскачка. Метод «раскачки» может быть использован при обработке качественных признаков, если предварительно каждому значению такого признака поставить в соответствие числовой символ.

§ 2. Описание основных алгоритмов тематической обработки изображений

2.1. Смесь нормальных распределений в задачах классификации

Алгоритм классификации данных основан на изложенной в § 1 методике и состоит из двух этапов: обучение и распознавание. На стадии обучения по заданной выборке восстанавливаются параметры распределений искоемых классов, на стадии распознавания проводятся классификация наблюдений в соответствии с найденными параметрами. При автоматической классификации данных выборки на обоих этапах совпадают.

В качестве параметрического семейства функций для аппроксимации плотности распределения выборки рассмотрим полинормальное распределение.

$$f(x) = \sum_{l=1}^L r_l \frac{|G^{(l)}|^{1/2}}{(2\pi)^{m/2}} \cdot \exp\left[-\frac{1}{2}(x - a^{(l)})^x \cdot G^{(l)}(x - a^{(l)})\right] \quad (12)$$

с неизвестными параметрами: r_l - весовые коэффициенты при компонентах распределения; $a^{(l)}$ - векторы средних значений нормальных компонент; $G^{(l)}$ - матрицы, обратные к ковариационным матрицам компонент.

Схема автоматической классификации включает: расчет ОМП параметров смеси нормальных распределений с выбором числа классов (компонент смеси) на основании точности определения параметров; вычисление оценок вероятностей принадлежности наблюдений классам; оценку ошибок вычисления этих вероятностей; оценку качества проведенной классификации.

Схема классификации с обучением (полным или частичным) содержит: расчет ОМП параметров полинормальных распределений каждого класса с выбором числа компонент в распределениях либо по критерию точности аппроксимации обучающих выборок, либо на основании качества классификации тестовых данных; вычисление оценок вероятностей принадлежности наблюдений классам; оценку ошибок вычисления этих вероятностей; оценку качества проведенной классификации; выбор диагностических признаков из множества исходных.

2.2. Расчет вероятностей принадлежности наблюдений определенным классам

Задача автоматической классификаций данных состоит в разделении смеси нормальных распределений (12), причем каждому классу соответствует одна из компонент смеси. Оценка $p(k, x)$ вероятности принадлежности наблюдения x классу k находится, по формуле Байеса (2), в которой $p_k = r_k$; $f(x, \alpha_k) = N(x, a^{(k)}, G^{(k)})$:

$$p(k, x) = \frac{r_k \cdot N_k(x)}{\sum_{l=1}^c r_l \cdot N_l(x)} \quad (13)$$

где $N_k(x) = N(x, \hat{a}^{(k)}, \hat{G}^{(k)})$ – оценка плотности нормального закона распределения с параметрами $a^{(k)}, G^{(k)}$, определенными методом максимального правдоподобия.

При классификации с обучением (полным или частичным) распределение наблюдений каждого класса считается полинормальным (12).

Формула Байеса для определения прогнозных вероятностей в предположении полноты пространства гипотез принимает вид

$$p(k, x) = \frac{p_k \cdot \sum_{l=1}^{L_k} r_{kl} \cdot N_{kl}(x)}{\sum_{j=1}^c p_j \sum_{l=1}^{L_j} r_{jl} \cdot N_{jl}(x)} \quad (14)$$

где L_j - количество компонент в распределении класса с номером j .

Если априорные вероятности $p_k, k=1, \dots, c$ неизвестны, они полагаются либо равными, либо пропорциональными суммарным априорным вероятностям элементов обучающей выборки.

$$p_k = \rho_k / n = \frac{1}{n} \sum_{i=1}^n \rho_{ki}$$

После проведения классификации всех наблюдений значения $p_k, k=1, \dots, c$, можно уточнить на основании полученных апостериорных вероятностей, а затем повторить классификацию.

Оценки прогнозных вероятностей в предположении непротиворечивости гипотез о принадлежности находятся из равенств (1),(12)

$$p(k, x) = \frac{p_k \cdot \sum_{l=1}^{L_k} r_{kl} \cdot N_{kl}(x)}{\sum_{l=1}^L r_l \cdot N_l(x)}, \quad (15)$$

где L - количество компонент в распределении всей исходной выборки. Априорные вероятности $p_k, k=1, \dots, c$ полагаются пропорциональными объемам обучающих

выборок: $p_k = n_k / n$. По результатам классификации выборки могут быть дополнены новыми наблюдениями с целью уточнения априорных вероятностей.

2.3. Расчет ОМП параметров смеси гауссовых распределений

Параметры полинормальных распределений, описывающих каждый класс, в задаче распознавания с обучением находятся методом максимального правдоподобия по выборке $\{x^{(i)}\}_{i=1}^n$. Плотность распределения элемента $x^{(i)}$ определяется формулами (3), (12). Из необходимых условий экстремума функции правдоподобия (6) получены уравнения относительно параметров $r_{kl}, a_i^{(k,l)}, G_{ij}^{(k,l)}$, $k = 1, \dots, c; l = 1, \dots, L_k; i = 1, \dots, m; j = 1, \dots, i$, записанные в удобном для итераций виде.

$$\begin{aligned}\tilde{r}_{kl} &= \frac{1}{n} \sum_{v=1}^n \omega_v^{(k,l)}, \\ \tilde{a}_i^{(k,l)} &= \left(\sum_{v=1}^n \omega_v^{(k,l)} \cdot x_i^{(v)} \right) / \left(\sum_{v=1}^n \omega_v^{(k,l)} \right), \\ \tilde{H}_{ij}^{(k,l)} &= \tilde{H}_{ji}^{(k,l)} = \left[\sum_{v=1}^n \omega_v^{(k,l)} (x_i^{(v)} - a_i^{(k,l)}) \cdot (x_j^{(v)} - a_j^{(k,l)}) \right] / \sum_{v=1}^n \omega_v^{(k,l)}, \\ \tilde{G}^{(k,l)} &= [\tilde{H}^{(k,l)}]^{-1},\end{aligned}\tag{16}$$

где $\omega_v^{(k,l)}$ - вероятность принадлежности наблюдения $x^{(v)}$ компоненте с номером l класса k :

$$\omega_v^{(k,l)} = \frac{\rho_{kv} \cdot r_{kl} \cdot N_{kl}(x^{(v)})}{\sum_{j=1}^c \rho_{jv} \sum_{i=1}^{L_j} r_{ji} N_{ji}(x^{(v)})}$$

В случае $\rho_{kv} \in \{0;1\}; k = 1, \dots, c; v = 1, \dots, n$ ОМП параметров распределений находятся для каждого класса в отдельности по обучающим выборкам $\{x^{(v)}\}_{v=1}^{n_k}, k = 1, \dots, c$, причем $\{x^{(v)}\}_{v=1}^{n_k} \subset \{x^{(v)}\}_{v=1}^n; \{x^{(v)}\}_{v=1}^{n_k} = \{x^{(v)} : \rho_{kv} = 1\}$. Итерационные формулы для случая $\rho_{kv} \in \{0;1\}$ приведены в книге [10]. Тем же методом находятся ОМП параметров смеси (4) в задаче автоматической классификации данных, а также ОМП параметров распределений исходной выборки и обучающих выборок каждого класса при распознавании с частичным обучением.

Сходимость итерационного процесса существенно зависит от начальных приближений параметров. Предлагаются два способа формирования начальных приближений: для случая заданного числа компонент в распределении и при последовательном увеличении числа компонент. Первый способ основан на принципе максимальной разнесенности математических ожиданий компонент в области значений

элементов выборки. Начальные приближения весовых коэффициентов $r_l, l = 1, \dots, L$, при компонентах полагаются равными $\frac{1}{L}$, ковариационные матрицы $H^{(l)}$ выбираются одинаковыми и диагональными, причем их диагональные элементы полагаются пропорциональными дисперсиям выборки по соответствующим признакам. Векторы a^l средних помещаются в вершинах m -мерного параллелепипеда (m - размерность пространства признаков), центр которого расположен в точке среднего значения выборки, а длины сторон, направленных вдоль осей координат, равны удвоенным среднеквадратичным отклонениям выборки по соответствующим признакам. Если L нечетно, в качестве начального приближения $a_{(l)}^{(L)}$ выбирается вектор среднего значения выборки. Остальные векторы a располагаются в попарно симметричных относительно центра вершинах параллелепипеда. Если вершин больше, чем компонент, в качестве начальных приближений выбираются наиболее разнесенные из них, в противном случае ($L > 2^m + 1$) строятся два или больше параллелепипедов, центры которых совпадают, а длины сторон $d_i^{(k)}$ определяются равенством

$$d_i^{(k)} = \frac{L \cdot k \cdot \sigma_i}{(n + 1)}; \quad k = 1, \dots, n; \quad i = 1, \dots, m,$$

где n - необходимое количество параллелепипедов, σ_i - среднеквадратичное отклонение выборки по признаку с номером i .

При таком выборе начальных приближений метод максимального правдоподобия позволяет оценить параметры 5-6-компонентного полинормального распределения в двумерном признаковом пространстве. С ростом размерности пространства признаков число компонент в распределении может быть увеличено.

Второй способ задания начальных приближений позволяет оценить параметры более сложных распределений, однако он требует и более длительных расчетов. Для определения параметров L -компонентного ($L = 2, 3, \dots$) распределения предварительно должны быть построены параметры распределения в предположении $L-1$ компонент. Затем производится попытка разбить одну из этих компонент на две. При этом ковариационные матрицы обеих новых компонент полагаются равными матрице выбранной компоненты, а векторы средних разносятся поочередно вдоль каждой из осей выбранной компоненты, начиная, например, с главной оси. В качестве начальных значений для остальных $L-2$ компоненты выбираются параметры, полученные на предыдущем шаге. Параметры всех L компонент уточняются в итерационном процессе. Перебор прекращается, если все компоненты построенного распределения оказались значимыми. В противном случае делается заключение, что L -компонентное распределение по данной обучающей выборке не может быть построено.

Такой способ формирования начальных приближений параметров удобен при автоматическом выборе числа компонент в распределении.

2.4. Расчет ошибок в ОМП параметров распределений

Ковариационная матрица ОМП параметров распределения определяется приближенными равенствами (7), (8). В случае полинормального распределения с невырожденными компонентами информационные количества Фишера (элементы информационной матрицы I) всегда существуют, так как подынтегральные функции в

(8) ограничены и быстро ($\approx e^{-x^2}$) стремятся к нулю при росте модуля аргумента. Учитывая, что количество экстремумов функций определяется только числом компонент в распределении, можно использовать для расчета интегралов (8) метод Монте-Карло. В качестве случайной выборки удобно использовать ту же выборку, по которой определялись параметры распределений.

Для параметров $\alpha = \{r_l, a^{(l)}, G^{(l)}\}_{l=1}^L$ распределения (12), полученных методом максимального правдоподобия, имеем следующую оценку информационной матрицы I :

$$\hat{I} = \sum_{v=1}^n [\underset{\alpha}{\text{grad}} \ln f(x^{(v)}, \hat{\alpha})] \cdot [\underset{\alpha}{\text{grad}} \ln f(x^{(v)}, \hat{\alpha})]^T \quad (17)$$

Вектор $\underset{\alpha}{\text{grad}} f(x^{(v)}, \hat{\alpha})$ содержат производные трех типов;

$$\begin{aligned} f_{a_i}^{(l)}(x^{(v)}) &= \sum_{j=1}^m G_{ij}^{(l)} (x_j^{(v)} - a_j^{(l)}) \cdot r_l \cdot N_l(x^{(v)}), \\ f_{G_{ij}}^{(l)}(x^{(v)}) &= \frac{1}{1 + \delta_{ij}} [H_{ij}^{(l)} - (x_i^{(v)} - a_i^{(l)})(x_j^{(v)} - a_j^{(l)}) \cdot r_l \cdot N_l(x^{(v)})], \\ f_{r_l}(x^{(v)}) &= N_l(x^{(v)}), \end{aligned} \quad (18)$$

где $H^{(l)} = (G^{(l)})^{-1}$, $\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$.

Оценки для квадратов ошибок ОМП параметров находятся на диагонали матрицы T , обратной к информационной матрице I .

2.5. Определение числа классов при автоматической классификации

Алгоритм определения числа классов основан на оценке значимости апостериорных вероятностей классов. В начале число классов L полагается равным единице. Находятся ОМП параметров распределения (12) в предположении одной компоненты. Затем L полагается равным двум, находятся ОМП параметров $(\tilde{r}, \tilde{a}, \tilde{G})$ смеси двух нормальных распределений и оценка ковариационной матрицы ОМП параметров. Для каждого класса проверяется отношение

$$\Delta \tilde{r}_l / \tilde{r}_l < x, \quad l = 1, \dots, L$$

где $\Delta \tilde{r}_l$ – среднеквадратичная ошибка вычисления r_l , x – заданное пороговое значение.

Если все L неравенств выполняются, параметры r, a, G уточняются, число классов увеличивается на единицу и процесс повторяется. В противном случае

выбирается полученное на предыдущем шаге максимальное количество значимых классов и находятся соответствующие ему оценки параметров смеси нормальных распределений.

Заметим, что при расчетах с числом классов, меньшим истинного, оценки r_l получаются завышенными (обучающая выборка делится на меньшее число классов), а оценки Δr_l - заниженными (строится псевдоинформационная матрица (16) для меньшего числа параметров). Понятно, что такие смещения оценок r_l и Δr_l не могут привести к неправильному нахождению числа классов.

Аналогичным способом может быть выбрано число компонент в полинормальных распределениях каждого класса при распознавании с обучением.

2.6. Расчет ошибок в оценках вероятностей принадлежности

Пусть α - вектор параметров распределений, использованных при оценке вероятности принадлежности $p(k/x)$, $\hat{\alpha}$ - вектор ОМП параметров α . Тогда, считая вероятность $p(k/x)$ функцией от параметров α и разложив функцию $p(\alpha)$ в ряд Тейлора в окрестности точки $\hat{\alpha}$, получим

$$p(\alpha) = p(\hat{\alpha}) + \left[\underset{\alpha}{\text{grad}} p(\hat{\alpha}) \right] (\alpha - \hat{\alpha}) + O[(\alpha - \hat{\alpha})^T (\alpha - \hat{\alpha})]$$

Отбрасываем члены второго порядка малости, обозначаем $\delta_p = p(\alpha) - p(\hat{\alpha})$, возводим δ_p в квадрат и усредняем по α

$$(\delta_p)^2 = \left[\underset{\alpha}{\text{grad}} p(\hat{\alpha}) \right]^T \cdot \overline{(\alpha - \hat{\alpha})(\alpha - \hat{\alpha})^T} \cdot \underset{\alpha}{\text{grad}} p(\hat{\alpha})$$

Обозначив среднеквадратичную ошибку вычисления вероятности через Δp и, используя оценку T для ковариационной матрицы ОМП параметров α , получаем

$$(\Delta p)^2 = \left[\underset{\alpha}{\text{grad}} p(\hat{\alpha}) \right]^T \cdot T \cdot \underset{\alpha}{\text{grad}} p(\hat{\alpha}) \quad (19)$$

При классификации с обучением вектор α состоит из параметров $\alpha_k, k = 1, \dots, c$, описывающих распределения каждого класса. В случае независимости обучающих выборок ($\rho_{kl} = \{0;1\}$) матрица T блочно-диагональна, причем каждый блок $T_k, k = 1, \dots, c$ является ковариационной матрицей ОМП параметров распределения соответствующего класса. Используя (14),(19), получаем оценку $\Delta p(k, x)$ ошибки вычисления вероятности $p(k/x)$

$$[\Delta p(k, x)]^2 = S_k^T \cdot T_k \cdot S_k - \sum_{i=1}^c S_{ki}^T \cdot T \cdot S_{ki}, \quad (20)$$

где
$$S_k = [p_k \cdot \underset{\alpha_k}{grad} f(x, \hat{\alpha}_k)] / \sum_{j=1}^c p_j f(x, \hat{\alpha}_j),$$

$$S_{ki} = [p_k p_i f(x, \hat{\alpha}_k) \cdot \underset{\alpha_i}{grad} f(x, \hat{\alpha}_i)] / [\sum_{j=1}^c p_j f(x, \hat{\alpha}_j)]^2.$$

Векторы $\underset{\alpha_k}{grad} f(x, \hat{\alpha}_k)$ состоят из производных вида (18), где $l = 1, 2, \dots, L; i = 1, 2, \dots, m; j = 1, \dots, i$.

Оценка $\Delta p(k, x)$ ошибки вычисления вероятности $p(k/x)$ в задаче автоматической классификации данных находится с использованием (13), (19). Вектор $\underset{\alpha}{grad} p(k, x, \hat{\alpha})$ состоит из C векторов $\underset{\alpha}{grad} p(k, x, \hat{\alpha})$, где $\alpha = \{\alpha_l\}_{l=1}^c$

$$\underset{\alpha_l}{grad} p(k, x, \hat{\alpha}) - \underset{\alpha_l}{grad} f(x, \hat{\alpha}) [\delta_{kl} \sum_{i=1}^c r_i N_i(x) - r_k N_k(x)] / [\sum_{i=1}^c r_i N_i(x)]^2$$

Векторы $\underset{\alpha_l}{grad} f(x, \alpha)$ содержат производные вида (18), где $i = 1, \dots, m; j = 1, \dots, i$.

При распознавании с частичным обучением точность вычисления параметров распределения всей совокупности наблюдений обычно намного выше, чем параметров каждого класса, вследствие существенно большего объема выборки в целом по сравнению с обучающей выборкой одного класса. Оценка $\Delta p(k, x)$ ошибки вычисления вероятности $p(k/x)$ находится из (15), (19) с использованием блочной структуры матрицы T , причем параметры в знаменателе формулы (15) считаются фиксированными

$$\Delta p(k, x)^2 = S_k^T \cdot T \cdot S_k$$

где
$$S_k = [p_k \cdot \underset{\alpha_k}{grad} f(x, \hat{\alpha}_k)] / [f(x, \hat{\alpha})]^2$$

Вектор $\underset{\alpha_k}{grad} f(x, \alpha_k)$ состоит из производных вида (18), где $l = 1, \dots, L; i = 1, \dots, m; j = 1, \dots, i$.

2.7. Расчет функционала качества классификации и определение числа компонент в распределениях по критерию качества классификации

Функционал качества распознавания строится по результатам классификации элементов обучающих выборок в случае $\rho_{k_l} \in \{0;1\}$. Для каждого наблюдения ξ вычисляется (10) нижняя доверительная граница $p_n(k, \xi)$ оценки вероятности правильной классификации $p(k/\xi)$. Качество распознавания определяется как средневзвешенное значение оценок $p_n(k, \xi)$ по всем элементам обучающих выборок $\{x^{(v)}\}_{v=1}^{n_k}, k = 1, \dots, c$:

$$F = \sum_{k=1}^c v_k \sum_{v=1}^{n_k} [p(k, x^{(v)}) - \gamma \cdot \Delta p(k, x^{(v)})] / \sum_{k=1}^c v_k \cdot n_k \quad (21)$$

где v_k - цена ошибки распознавания элемента класса k , соответствующая значению

$$w_k \text{ в формуле (9): } w_k = v_k \cdot n_k / \sum_{i=1}^c v_i n_i.$$

Функционал F служит критерием качества для выбора диагностических признаков и определения числа компонент в распределениях при классификации с обучением (полным ила частичным).

В задаче автоматической классификации данных качеством распознавания назовем четкость разбиения выборки на классы (максимум вероятности отнесения элементов выборки к классам) с учетом ошибок вычисления вероятностей

$$F = \frac{1}{n} \sum_{v=1}^n \max_k [p(k, x^{(v)}) - \gamma \cdot \Delta p(k, x^{(v)})]$$

Выбор числа компонент $(L_k, k = 1, \dots, c)$ при распознавании начинается с предположения нормальности распределения элементов каждого класса. Находятся параметры распределений $r_k, a^{(k)}, G^{(k)}, k = 1, \dots, c$ и функционал качества классификации F . Число компонент L_1 в распределении первого класса увеличивается на единицу, вычисляется функционал качества F . Если он меньше исходного, число компонент L_2 в распределении первого класса считается равным единице и дальше не изменяется. Затем увеличивается на единицу число компонент в распределении второго класса и т.д. Если добавление второй компоненты в некоторых классах привело к повышению качества классификации ($n < c$), то в этих классах ($i_k = 0$) наращивание числа компонент продолжается. Процесс останавливается в том случае, когда добавление компоненты в распределение любого класса ведет к уменьшению функционала качества ($n = c$).

Этот алгоритм позволяет выбрать более простые модели распределений в случае хорошо разделимых классов сложной структуры, что приводит к сокращению расчетов при высоком качестве классификации наблюдений.

§ 3. Практика применения смесевых моделей для классификации, распознавания и обнаружения

По методам и алгоритмам, изложенным в § 1 и § 2, были разработаны и испытаны на сотнях модельных задач компьютерные комплексы программ распознавания образов, классификации объектов и обнаружения целей.

Первый комплекс программ распознавания РОСА был создан в ФТИНТ АН УССР (г.Харьков) [11]. С помощью этого программного комплекса на основе самолетных данных, полученных профильным сверхпроводниковым ИК-радиометром [12] проводилось геологическое картирование и велись поиски полезных ископаемых на территории Якутии, Украины, Казахстана, Карелии, Чехии, Тянь-Шаня. Комплекс РОСА был передан в эксплуатацию в следующие харьковские организации: в ЦКБ «Протон» (классификация радиосигналов), в Институт терапии (диагностика инфаркта миокарда), в Институт радиологии (онкологические заболевания), в ПО «Карбонат» (химические технологии), а также в ИГН АН УССР (г.Киев), в Лабораторию Аэрометодов (г.Ленинград), в ЦГЭ (г.Москва), в ЦКБ «Прогресс» (г.Самара).

Второй комплекс программ распознавания смесевых моделей был разработан для персонального компьютера Пентиум на языке СИ, в израильском концерне «Авиационная Промышленность Израиля» [13]. Этим программным комплексом решались задачи геологического картирования на территории Израиля, задачи дистанционного обнаружения искусственных объектов – замаскированных, подземных и наземных – в различных условиях видимости. Приборные данные были предоставлены иерусалимским геологическим институтом (гиперспектральная самолетная съемка) и авиационным заводом «Ташан» (многоканальная инфракрасная камера).

Третий комплекс программ (в системе MATLAB) был разработан в Тель-Авивском университете [14]. Этим программным комплексом решались различные задачи дистанционного зондирования Земли на основе гиперспектральных и многоканальных тепловых и оптических данных [15].

Заключение

1. Дальнейшее совершенствование методов, алгоритмов и программ

Развитие и совершенствование методов смесового моделирования будет продолжаться в трех направлениях:

- А) вычислительно-математическом;
- Б) физико-методическом;
- В) съемочно-аппаратном.

По всем перечисленным пунктам имеются конструктивные идеи и соображения, ожидающие компьютерной и приборной реализации.

A1. Ускорение сходимости EM-алгоритма оценки параметров смеси.

A2. Ускорение вычислений нижних доверительных границ для байесовских вероятностей принадлежности объектов классам.

A3. Существенное ускорение расчетов информационной матрицы Фишера и вместе с ней критерия значимости компонент смеси.

A4. Учет текстуры сигналов, полей и сцен методом параллельного переноса данных без расчета текстурных признаков.

B1. Обобщение смесей распределений компонентами негауссовского типа.

B2. Совершенствование методов восстановления стереоизображений.

B3. Развитие динамической модели распознавания, учитывающей изменчивость внешних условий среды в промежутке времени между обучением и распознаванием.

B4. Создание специального участка местности (методического полигона) для обучения с применением динамической модели распознавания.

B5. Оптимизация обучающих и разведочных экспериментов.

В1. Использование многопроцессорных компьютеров для опознавания целей методами параллельных алгоритмов.

В2. Постановка распознающих систем на борту носителя аппаратуры для тематической обработки данных в реальном времени съемок.

2. Перспективы использования смесевых моделей распознавания на примере изучения природных ресурсов Белгородской области

Белгородская область как известно [16] является наиболее богатой геологическими и агрономическими природными ресурсами областью российского черноземья. Кроме всемирно известных месторождений железных руд, в области большой перечень других полезных ископаемых: все необходимые стройматериалы, агрохимическое сырье, ювелирно-поделочные камни. Имеются определенные предпосылки добычи цветных металлов и даже углеводородного сырья.

Однако, по мнению ведущих специалистов к настоящему времени тщательно разведано не более 15% территории Белгородской области [17]. Предлагается продолжить геологоразведочные работы и провести агроресурсные исследования, сочетая дистанционные и наземные методы, по следующим пунктам:

А) Оконтуривание известных и доразведка новых месторождений полезных ископаемых.

А1. Получение спектрзональных и интегральных космических снимков в различных диапазонах излучения.

А2. Дополнение космической информации аэросъемками и наземными измерениями.

А3. Совместно с геологической службой области задание известных участков для компьютерного обучения.

А4. Создание распознающего программного комплекса.

А5. Обнаружение интересующих объектов и подтверждение в качестве теста известных месторождений.

Б) Решение различных задач землепользования дистанционными методами.

Б1. Изучение экологической обстановки в городе и области.

Б2. Агрономические проблемы: состав, сортность, урожайность и текущее состояние растительных культур.

Б3. Лесное и садовое хозяйство.

Б4. Проблемы урбанизации.

Б5. Ландшафтное картирование.

Б6. Поиск заранее указанных искусственных и естественных объектов.

Б7. Мониторинг атмосферы и водоемов.

Б8. Инженерно-геологическое картирование.

Библиографический список

9. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации. Проблемы кибернетики.-М.Наука.1978.-Вып. 33, стр. 5-68.

10. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. Статистические проблемы обучения.-М.Наука.1974.-415 стр.

11. Верхаген К., Дейн Р. и др. Распознавание образов: состояние и перспективы.-М. Радио и связь. 1985.-104 стр.

12. Колтунов И.А., Кондратьева Л.М., Монастырев А.П. Статистическая классификация наблюдений с полимодальными распределениями. Статистические проблемы управления. Вильнюс.1987. Вып.78, стр. 83-121.

13. Рао К.Р. Линейные статистические методы и их применение. М.Наука.1968, 548 стр.

14. Колтунов И.А. Обобщенные оценки параметров и их применение. Вычислительная математика и вычислительная техника.Харьков,1974.Вып. 5, стр. 102-113.

15. Колтунов И.А. О статистиках с минимальными дисперсиями. Теория вероятностей и ее применение. М. 1977, №3, стр.624-644.

16. Колтунов И.А., Чаркина Л.Я., Монастырев А.П. и др. Обработка изображений. Избранные методы и алгоритмы. Препринт 26-88.ФТИНТ. Харьков, 1988.

17. Малая Л.Т., Яблчанский Н.И, Рудинский О.О., Колтунов И.А., Кондратьева Л.М., Монастырев А.П., Чаркина Л.Я. Возможности использования комплекса программ распознавания образов РОСА для обработки медицинской информации. Харьков, 1987, 19 стр. Препринт ФТИНТ АН УССР, №12-87.

18. Дуда Р., Харт П. Распознавание образов и анализ сцен. М. Мир, 1976, 559 стр.

19. Колтунов И.А., Кондратьева Л.М., Кочарян Е.В. и др. Комплекс программ РОСА. Харьков, 1986, 48 стр. Препринт. ФТИНТ АН УССР, №9-86.

20. Веркин Б.И., Коноводченко В.Н. и др. Сверхпроводниковый ИК-радиометр. Тепловые приемники излучения. Л. Изд-во ГОИ, 1980, стр. 144-146

21. J.Koltunov, A.Maximov, I.Meitin et al. Determination of temperature and/or emissivity function of objects by remote sensing. International Patent WO 99/27336, 03.06.99.

22. Alexander Koltunov, Joseph Koltunov, and Eyal Ben-Dor Adaptive recognition under static and dynamic environment assumptions.- in Infrared Technology and Applications XXIX, Proc. SPIE v.5074, Aerosense - 2003 Symposium, (Orlando, FL), 21-25 April, 2003.

23. J.Koltunov, A.Koltunov Dynamic Detection Model and its applications: in Ifrared Technology and Applications XXIX. Proc SPIE 5074, Aerosense-2003. Symposium, (Orlando,FL), April, 2003.

24. Белгородоведение. Под. ред. В.А. Шаповалова. Белгород, 2002, 410 с.

25. Хрисанов В.А., Петин А.Н., Яковчук М.М. Геологическое строение и полезные ископаемые Белгородской области. – Белгород: Изд-во БелГУ, 2000, 250 с.

УДК 681.5.015.8

ЦИФРОВОЙ СПЕКТРАЛЬНЫЙ АНАЛИЗ ЧЁТНЫХ ФУНКЦИЙ

Ю.М. Перлов

Основные соотношения

Введём обозначения

$$F(R, x) = (2\pi R)^{-1} \left(\frac{\sin(Rx/2)}{\sin(x/2)} \right)^2 - \text{ядро Фейера};$$

$$\sigma_R(x, f) = \int_{-\pi}^{\pi} F(R, t) f(x+t) dt - \mathcal{R}\text{-я сумма Фейера функции } f(t) \text{ в точке } x.$$

Теорема 1.

Пусть функция $f(x)$, $x \in (-\infty, \infty)$ обладает свойствами:

$$\sum_{j=-\infty}^{\infty} f(j\Delta x) < \infty, \forall \Delta x > 0; \int_{-\infty}^{\infty} f(x) dx < \infty \quad (1)$$

тогда для $\forall t > 0$

$$\lim_{R \rightarrow \infty} t \int_{-\infty}^{\infty} F(R, xt) f(x) dx = \sum_{j=-\infty}^{\infty} f(jx_t), x_t = \frac{2\pi}{t} \quad (2)$$

Доказательство.

Вследствие (2) существует интеграл