УДК 004.852:681.514 **DOI** 10.15622/ia.2020.19.6.2

А.А. БАТЕНКОВ, К.А. БАТЕНКОВ, А.Г. БОГАЧЕВ, В.В. МИШИН

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КЛАССИФИКАТОРА ОБЪЕКТОВ НА ОСНОВЕ БАЙЕСОВСКОГО ПОДХОДА

Батенков А.А., Батенков К.А., Богачев А.Г., Мишин В.В. **Математическая модель классификатора объектов на основе байесовского подхода.**

Аннотация. Утверждается, что первостепенное значение в решении задачи классификации занимают: нахождение условий разбиения генеральной совокупности на классы, определение качества такого расслоения и верификация модели классификатора. Рассмотрена математическая модель нерандомизированного классификатора признаков, полученных без учителя, когда априори не задается число классов, а лишь устанавливается его верхняя граница. Математическая модель приведена в виде постановки минимаксной условной экстремальной задачи и представляет собой задачу поиска матрицы принадлежности объектов к какому-либо классу. В основе разработки классификатора признаков находится синтез двумерной плотности вероятностей в пространстве координат: классы – объекты. С помощью обобщенных функций вероятностная задача поиска минимума Байесовского риска сведена к детерминированной задаче на множестве нерандомизированных классификаторов. Вместе с тем использование специально введенных ограничений фиксирует нерандомизированные правила принятия решений и погружает целочисленную задачу нелинейного программирования в общую непрерывную нелинейную задачу. Для корректного синтеза классификатора необходимы дисперсионная кривая изотропной выборки и характеристики качества классификации в зависимости от суммарной внутриклассовой и межклассовой дисперсии. Задача классификации может быть интерпретирована как частная задача теории катастроф. В условиях ограниченных исходных данных найден минимаксный функционал, отражающий качество классификации при квадратичной функции потерь. Математическая модель представлена в виде задачи целочисленного нелинейного программирования и приведена с помощью полиномиальных ограничений к виду общей задачи нелинейного непрерывного программирования. Найдены необходимые условия расслоения на классы. Эти условия могут быть использованы как достаточные при проверке гипотезы о существовании классов.

Ключевые слова: нерандомизированный классификатор признаков, верхняя граница числа классов, минимакс, условная экстремальная задача, целочисленная задача нелинейного программирования

1. Введение. Выясним цели классификации, которые могут различаться в зависимости от особенностей конкретной прикладной области знаний, и место задачи классификации в общей теории системного анализа [1-3]. Известно, что под моделью понимается некоторый материальный или мысленно представляемый объект (образ объекта), который в процессе изучения замещает объект-оригинал, сохраняя некоторые важные для данного исследования типичные его свойства. Модель – это всегда упрощенное изображение реального объекта, процесса или явления. Тогда модель можно определить как некоторое сжатое изображение исходных данных. В случае их дискретного множества такую модель можно интерпретировать как замену более мощного множества менее мощным множеством, что в общем случае соответствует задаче классификации. В связи с этим будем рассматривать классификатор как совокупность моделей (по числу классов) представления выхода некоторого источника исходных данных, где каждый класс представлен своей эталонной моделью [3, 4].

В рамках такой концепции системного анализа (синтеза) решение задачи классификации занимает центральное место и позволяет достичь следующих результатов:

1. Разбить (расслоить) множество объектов на подмножества (классы) схожих объектов для дальнейшей раздельной обработки данных.

2. Установить наличие однородности (изотропности) исходных данных, если расслоение на классы невозможно (например, в случае выборки из непрерывного многомерного равномерного распределения).

3. Сократить объем хранимых данных, оставив по одному представителю от каждого класса (кластера), то есть решить задачу сжатия данных (построить векторный квантователь).

4. Выделить нетипичные объекты, которые не подходят ни к одному из кластеров, а именно выявить уникальные (новые по отношению к существующим) модели.

5. Построить иерархию множества объектов.

В первом и втором случаях особое внимание обращают на идентификацию числа кластеров. В третьем случае важнее обеспечить высокую степень сходства объектов внутри каждого кластера, а число классов выбирают исходя из условий функционирования системы. В четвертом случае наибольший интерес представляют отдельные объекты, не вписывающиеся ни в один из кластеров, а потому они могут описывать новые знания об интересующей проблеме. В пятом случае создают вложенные модели классификаторов, где значимость приобретает степень согласованности критериев качества классификации на различных уровнях древообразной иерархической структуры.

Анализ предполагаемых результатов и приоритетных направлений их достижения позволяет сделать вывод, что первостепенное значение в решении задачи классификации занимают: нахождение условий разбиения генеральной совокупности на классы, определение качества такого расслоения и верификация модели классификатора, что является целью исследования.

2. Условия правильной классификации. Известно [5-7], что разбиение множества на классы – это действие распределения объектов по классам на основе сходства объектов внутри класса и их отличия от объектов других классов.

Условия правильной классификации [8].

1. Подмножества (классы) попарно не пересекаются.

2. Объединение всех подмножеств (классов) совпадает с исходным множеством.

Необходимо отметить, что условия правильной классификации могут характеризовать построение классификатора по обучающей выборке (на этапе обучения), но не соответствуют условиям правильной классификации по контролирующей выборке (на этапе контроля), так как могут быть применены рандомизированные правила принятия решений, и обучающее множество не тождественно контролирующему множеству. Кроме того, неизвестно, как разделить предоставленный статистический материал на обучающую и контролирующую выборки. Обозначим еще ряд нерешенных вопросов и проблем в автоматической классификации без учителя:

1. Как учесть при классификации неточность измерений признаков?

2. Насколько формализовано понятие класса?

3. Почему классы не должны пересекаться?

4. Допускает ли исходное множество расслоение на классы и сколько их?

5. При каких условиях классификатор становится векторным квантователем? [9-11].

Чтобы ответить хотя бы на некоторые из поставленных вопросов, разрабатываемая математическая модель должна быть достаточно общей, а значит, иметь как можно больший набор стратегий игры с природой. В этой связи рассмотрена задача поиска количественного функционала (критерия) и его параметров, отражающих качество классификации объектов без учителя.

3. Байесовский подход при построении математической модели классификатора. Для построения математической модели классификатора применим Байесовский подход [12-15], который используется во многих теориях (принятия решений, оценивания параметров, распознавания образов и т.д.).

Пусть $\mathbf{Y}_n = (y_{1n}, y_{2n}, \ldots, y_{Ln})^{\mathrm{T}}$, $\forall n = \overline{1, N}$ – дискретное множество из N векторов в пространстве размерности L. На физическом уровне это вводит N векторов признаков, подлежащих классификации, которые характеризуются дескриптором из L компонент. Далее будем ассоциировать N с числом объектов, подлежащих классификации. Использование Байесовского подхода предполагает задание вероятностной меры в пространстве признаков. Будем считать, что известен вероятностный ряд появления n-го объекта для классификации:

$P(n)$, $\forall\, n=\overline{1,N}$. Если признаки всех объектов различны, то можно считать ряд равновероятным:

$$P(n)=\frac{1}{N},\ \forall\, n=\overline{1,N}.$$

Далее, пусть $\mathbf{Z}_k=(z_{1k},z_{2k},\ldots,z_{LK})^{\mathrm{T}},\forall k=\overline{1,K}$ – дискретное множество из K эталонных векторов, каждый из которых представляет центроид одного из K классов, и $\mathbf{Z}_k^{'}=(z_{1k}^{'},z_{2k}^{'},\ldots,z_{LK}^{'})^{T}$, $\forall\, k=\overline{1,K}$ – дискретное множество, тождественное эталонному множеству. Эти множества не известны в задаче классификации и могут быть определены после анализа решения исходной задачи.

На рисунке 1 представлена модель процедуры оценки [12], которую в дальнейшем будем использовать для формирования математической модели классификатора объектов.

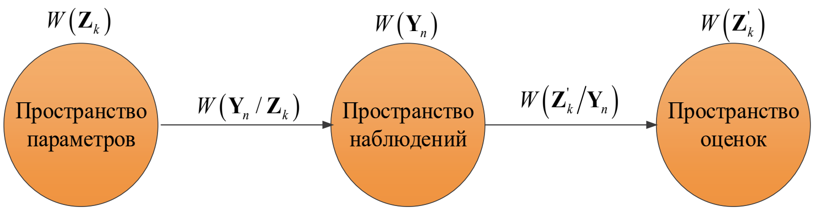

Рис. 1. Модель процедуры оценки

Применим формулу Байеса [12] для задачи классификации:

$$W\left(\mathbf{Z}_{\mathbf{k}}^{'}/\mathbf{Y}_{\mathbf{n}}\right)=\frac{W\left(\mathbf{Z}_{\mathbf{k}}\right)W\left(\mathbf{Y}_{\mathbf{n}}/\mathbf{Z}_{\mathbf{k}}\right)}{W\left(\mathbf{Y}_{\mathbf{n}}\right)},k=1,\ldots K;n=1,\ldots N;$$

где $W\left(\mathbf{Z}_k^{'}/\mathbf{Y}_{\mathbf{n}}\right)$ – апостериорная плотность вероятности эталонной модели $\mathbf{Z}_k$ при наблюдении объекта $\mathbf{Y}_n$; $W\left(\mathbf{Z}_k\right)$ – априорная плотность вероятности эталонной модели $\mathbf{Z}_k$; $W\left(\mathbf{Y}_n/\mathbf{Z}_k\right)$ – функция правдоподобия наблюдения $\mathbf{Y}_n$ при реализации эталонной модели $\mathbf{Z}_k$; $W\left(\mathbf{Y}_n\right)$ – плотность вероятности наблюдений $\mathbf{Y}_n$.

В основе байесовского подхода к решению задачи классификации лежит разложение дискретной двумерной плотности вероятности (в пространстве эталоны – признаки) на безусловные и условные дискретные плотности:

$$W(\mathbf{Z}_k,\mathbf{Y}_n)=W(\mathbf{Z}_k/\mathbf{Y}_n)W(\mathbf{Y}_n)=W(\mathbf{Z}_k)W(\mathbf{Y}_n/\mathbf{Z}_k),$$
$$k=1,\ldots,K,\ n=1,\ldots,N.$$

В наиболее общем виде представим ее с помощью смеси плотностей вероятности непрерывных случайных величин на основе обобщенных функций:

$$W(\mathbf{Z},\mathbf{Y})=\sum_{k=1}^{K}\sum_{n=1}^{N}x_{k,n}\,\delta(\mathbf{Z}-\mathbf{Z}_k)\delta(\mathbf{Y}-\mathbf{Y}_n),$$

где $x_{k,n}$ – вероятность назначения в k-й кластер n-го объекта; $\delta(\mathbf{Z})$ – обобщенная δ-функция [13].

Далее определим маргинальные плотности вероятностей следующим образом:

$$W(\mathbf{Z}/\mathbf{Y})=\sum_{k=1}^{K}\frac{x_{k,n}}{\sum_{i=1}^{K}x_{i,n}}\delta(\mathbf{Z}-\mathbf{Z}_k)1(\mathbf{Y}-\mathbf{Y}_n),\ \forall n=1,\ldots,N.$$

Вывод выражения для апостериорной плотности вероятности принадлежности n-ого объекта k-ому классу приведен в приложении 1.

Аналогично апостериорной плотности получим формулировку функции правдоподобия:

$$W(\mathbf{Y}/\mathbf{Z})=\sum_{n=1}^{N}\frac{x_{k,n}}{\sum_{s=1}^{N}x_{k,s}}1(\mathbf{Z}-\mathbf{Z}_k)\delta(\mathbf{Y}-\mathbf{Y}_n),\ \ \forall k=1,\ldots,K,$$

$$W(\mathbf{Y})=\sum_{n=1}^{N}\delta(\mathbf{Y}-\mathbf{Y}_n)\sum_{k=1}^{K}x_{k,n}.$$

Здесь введена специальная разрывная $1(\mathbf{Z})$ -функция:

$$1(\mathbf{Z})=\begin{cases}1,\mathbf{Z}=\mathbf{0},\\0,\mathbf{Z}\neq\mathbf{0}.\end{cases}$$

Ее основные свойства приведены в приложении 2 [16, 17]. Условия нормировки для введенных плотностей показаны в приложении 3.

При таком подходе решение задачи классификации сводится к формированию двумерной плотности вероятностей $W\left(\mathbf{Z},\mathbf{Y}\right)$, а именно к нахождению матрицы **X** вероятностей назначений, размера $K \times N$, объектов в кластеры при ограничениях, представляющих собой условия нормировки. Однако обязательными дополнительными ограничениями должно быть то, что все представленные объекты необходимо классифицировать, то есть каждый объект должен принадлежать хотя бы одному классу, и то, что класс может не содержать объектов (быть пустым), так как априори неизвестно число классов. Если число классов известно, то это порождает задачу векторного квантования.

4. Матричная модель классификатора. Заметим, что искомая двумерная дискретная плотность распределения вероятностей состоит из трех компонент: матрицы **X**, множества признаков и множества эталонных векторов $\mathbf{Z}_k^{'} = (z_{1k}^{'}, z_{2k}^{'}, \ldots, z_{LK}^{'})^T$, $\forall\, k = \overline{1, K}$. Элементы множеств признаков и эталонных векторов связаны между собой с помощью матрицы **X** – вероятностей назначений в *k*-й кластер *n*-го объекта. Тогда **X** может рассматриваться как искомый двумерный ряд распределения дискретных случайных величин $\left\{k = \overline{1, K}\right\}$ *и* $\left\{n = \overline{1, N}\right\}$, и следовательно, допустимо получение любой маргинальной плотности вероятности, применяемой при байесовском подходе. Таким образом, если имеется матрица **X** с элементами $x_{k,n}$, удовлетворяющих условиям:

$$x_{k,n} \geq 0,\ \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n} = 1,\ \forall\ k = \overline{1,K},\ n = \overline{1,N},$$

то можно найти маргинальные плотности, опуская обозначения соответствующих δ, 1, ε-функций, то есть определить маргинальные ряды распределения:

$$P\left(k,n\right) = x_{k,n}, P(k/n) = \frac{x_{k,n}}{\sum_{s=1}^{K} x_{s,n}}, P\left(k\right) = \sum_{n=1}^{N} x_{k,n},$$

$$P\left(n/k\right) = \frac{x_{k,n}}{\sum_{s=1}^{N} x_{k,s}}, P\left(n\right) = \sum_{k=1}^{K} x_{k,n}, \forall k = \overline{1,K}, n = \overline{1,N}.$$

Для упрощения записи в матричном виде маргинальные ряды представим в виде:

$$P\left(k,n\right)=\mathbf{X}, P(k/n)=\mathbf{X}\left[\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_{k}\right)\right]^{-1};$$
$$P\left(k\right)=\mathbf{X}\mathbf{I}_{n}, P\left(n/k\right)=\mathbf{X}^{\mathrm{T}}\left[\operatorname{diag}\left(\mathbf{X}\mathbf{I}_{n}\right)\right]^{-1},$$
$$P\left(n\right)=\mathbf{X}^{\mathrm{T}}\mathbf{I}_{k}, \forall k=\overline{1,K}, n=\overline{1,N}.$$

где $\operatorname{diag}\left(\mathbf{x}\right)$ – оператор диагонализации вектора **x**.

Однако поиск неизвестной матрицы вероятностей назначений **X** значительно усложняет задачу классификации вследствие большого количества ограничений на ее элементы [18]. Поэтому будем искать решение в виде матрицы назначений **X'** с булевыми элементами. При этом сама матрица назначений **X'** естественно взаимосвязана с матрицей вероятностей назначений **X** [19, 20]. Введение булевой матрицы назначений возможно виду того, что классификатор нерандомизированный.

Для этого определим связь условного маргинального ряда вероятностей $P(k/n)=\mathbf{X}'$ и двумерного ряда вероятностей $P\left(k,n\right)=\mathbf{X}$ в матричном виде:

$$\mathbf{X}'\mathbf{G}=\mathbf{X}, \mathbf{G}=\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_{k}\right), \mathbf{X}'=\mathbf{X}\left[\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_{k}\right)\right]^{-1}, \mathbf{X}'\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_{k}\right)=\mathbf{X},$$

в скалярном виде:

$$x_{k,n}=x'_{k,n}b_{n}, k=\overline{1,K}, n=\overline{1,N}, b_{n}=\sum_{m=1}^{K}x_{m,n}, x_{k,n}=x'_{k,n}\sum_{m=1}^{K}x_{m,n},$$

где $\mathbf{I}_{k}=\underbrace{\begin{bmatrix}1 & \cdots & 1 & \cdots & 1\end{bmatrix}}_{k}^{\mathrm{T}}$ – единичный вектор размера $K\times 1$.

В задаче классификации будем использовать следующие условия нормировки, которые порождают ограничения в создаваемой математической модели:

$$\mathbf{I}_{k}^{\mathrm{T}}\mathbf{X}\mathbf{I}_{n}=1; \quad (1)$$

$$\mathbf{I}_{k}^{\mathrm{T}}\mathbf{X}=\mathbf{I}_{n}^{\mathrm{T}}; \quad (2)$$

$$x'_{k,n} = \{0;1\}, \forall k = \overline{1,K}, n = \overline{1,N}, \mathbf{X}' \bullet \mathbf{X}' - \mathbf{X}' = \mathbf{0}; \tag{3}$$

$$\begin{aligned} & x_{k,n} \geq 0, \forall k = \overline{1,K}, n = \overline{1,N}; \\ & \mathbf{X} - \mathbf{Q} \bullet \mathbf{Q} = \mathbf{0}, \end{aligned} \tag{4}$$

где $\mathbf{A} \bullet \mathbf{B}$ – произведение Адамара матриц **A** и **B**; **Q** – матрица дополнительных переменных размера $K \times N$, необходимых для перехода от ограничений неравенств (на неотрицательность переменных) к ограничениям типа равенств; **0** – нулевая матрица размера $K \times N$.

Ограничение (1) означает, что матрица **X** – представляет собой дискретный двумерный ряд распределения вероятностей.

Ограничения (2), наложенные на апостериорную плотность вероятности, обеспечивают классификацию каждого объекта, то есть их принадлежность одному или нескольким классам с определенной вероятностью. Тогда составляющие:

$$\mathbf{X}' = \mathbf{X}\left[\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_k\right)\right]^{-1}$$

интерпретируются как условные вероятности принадлежности n-го объекта k-му классу. Если эти компоненты определяются как элементы булевого множества {0; 1}, то классификатор оказывается нерандомизированным.

Ограничения (3) задают детерминированное множество решающих правил, то есть разрабатываемая математическая модель является нерандомизированным классификатором, и погружают дискретную булеву задачу в непрерывную.

Ограничения (4) позволяют интерпретировать составляющие матрицы **X** как вероятности.

Совокупность ограничений (2) и (3) обуславливает разбиение [21, 22] множества из «N» объектов на «K» классов, то есть вероятность принадлежности к k-му классу при условии наблюдения объекта с номером n. Однако условная плотность вероятности не определяет полностью основу математической модели классификатора – двумерную плотность вероятности, которая имеет большее число степеней свободы [23, 24]. В условиях равновероятного ряда появления объектов для классификации нетрудно увидеть, что ограничения (1) и (4) являются избыточными и **X'** однозначно определяет **X**.

Очень важно отметить, что ограничения на функцию правдоподобия не вводятся, что обеспечивает возможное существование пустых кластеров. Поэтому при автоматической классификации можно задавать лишь верхнюю границу числа классов либо осуществлять полное разбиение объектов на неупорядоченные классы, так как возможное число классов всегда определяется числом объектов: $1 \le K \le N$.

Тогда рассматриваемая математическая модель должна содержать постановку задач определения:

– сходства объектов внутри класса;

– отличия объектов класса от объектов других классов;

– числа классов;

– оптимального разбиения множества объектов на классы;

– гипотетического (представительного) объекта внутри каждого класса.

5. Межклассовая метрика. Для того чтобы сформулировать такую задачу, необходимо ввести метрику (расстояние) между и внутри множеств $\mathbf{Y}_n, \forall n = \overline{1, N},$ и $\mathbf{Z}_k, \forall k = \overline{1, K}.$

Эта метрика должна удовлетворять следующим требованиям:

– обладать известными свойствами метрики в смысле теории метрических пространств;

– отражать физическую сущность функции потерь при решении задачи минимизации Байесовского риска;

– иметь не высокую вычислительную сложность;

– учитывать статистические свойства результатов измерений признаков;

– использовать полученный классификатор для решения последующих задач (например, применение нелинейных функционалов для синтеза разделяющих многомерных поверхностей за счет увеличения размерности пространства [25- 27]).

Конструирование такой метрики представляет собой самостоятельную научную задачу, которой будут посвящены дальнейшие исследования. Поэтому здесь мы ограничимся математической моделью классификатора признаков в смысле экстремальной задачи с ограничениями при использовании обобщенной метрики эвклидова пространства [28-30]:

$$\rho(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^{\mathrm{T}} \mathbf{W} (\mathbf{x} - \mathbf{y})},$$

где $\mathbf{W}$ – положительно определенная квадратная матрица размера $(L \times L)$.

6. Условия построения классификатора. Разрабатываемая математическая модель классификатора признаков основана на следующих утверждениях, рассматриваемых сначала как некоторые гипотезы, которые в последующем могут быть доказаны.

Так при разбиении множества на классы цели являются противоречивыми. С одной стороны, при увеличении числа классов сходство объектов внутри класса увеличивается (расстояние между объектами одного класса уменьшается), а с другой стороны, уменьшается межклассовое различие (расстояние между представителями разных классов уменьшается). Поэтому обоснование числа классов проводится эвристически, без достаточного научного обоснования. Кроме того, множество может вообще не расслаиваться на классы, вследствие недостаточного статистического материала и (или) неправильного отбора признаков, и такие ситуации должны быть обнаружены. Поэтому концептуально выдвигается лишь гипотеза о существовании расслоения исходного множества на классы с помощью разрабатываемой математической модели. Эта гипотеза на заключительном этапе подлежит проверке (верификации), а именно: допускает ли исследуемое пространство признаков решение задачи классификации? В связи с этим рассмотрение достаточных условий правильной классификации является актуальной задачей.

Кроме того, конструирование функционала качества классификации предполагает рассмотрение двух граничных условий: 1) кластер является единственным; 2) число классов равно объему выборки. В первом случае межклассовая дисперсия минимальная (равна нулю), а внутриклассовая дисперсия максимальная. Во втором случае наоборот: межклассовая дисперсия максимальная, а суммарная внутриклассовая дисперсия минимальная (равна нулю). Кроме того, если суммарная внутриклассовая дисперсия может быть выбрана для количественного описания сходства объектов внутри классов, то межклассовая дисперсия недостаточно описывает отличие объектов разных классов, так как она определяется лишь координатами эталонов каждого класса и не учитывает «тонкого» распределения объектов внутри класса.

В противоположность задаче классификации в задаче векторного квантования основное внимание уделяется минимизации суммарной внутриклассовой дисперсии, а значит, все классы должны быть не пустыми, и конструирование функционала качества векторного квантователя не вызывает затруднений. Этой задаче будут посвящены последующие работы.

Интересно рассмотрение задачи классификации с позиций теории информации [9]. Известно [9], что дифференциальная энтропия монотонно возрастает с возрастанием дисперсии генеральной совокупности и достигает своего максимального значения для установленных плотностей вероятностей случайной величины в зависимости от ее области определения. Поэтому можно высказать следующие гипотезы, что:

1. При расслоении выборки на классы ее суммарная внутриклассовая дисперсия должна быть меньше максимально возможной, то есть, когда нет расслоения при заданном числе классов.

2. Чтобы идентифицировать расслоения, должна быть некоторая гипотетическая выборка (образец), в которой потенциально невозможно выделить кластеры (например, выборка из многомерного равномерного распределения [31, 32]);

3. Проверка предположения о наличии расслоения на классы может быть проведена путем сравнения рабочего и гипотетического множеств (например, с помощью расстояния Хаусдорфа [33], использования информационной меры Ку́льбака – Ле́йблера [34] и т. д.).

4. Система замкнута [1], а значит ее энтропия и, следовательно, дисперсия не зависят от способа разбиения выборки на классы, то есть:

$$D_m\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)+D_v\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)=\mathrm{const}\left(\mathbf{Y}\right),$$

где $\mathbf{X}$ – матрица размера $K\times N$ назначений объектов в кластеры; K – число классов; N – число классифицируемых объектов; D_m – межклассовая дисперсия (дисперсия эталонов); D_v – суммарная по всем классам внутриклассовая дисперсия; $\mathbf{Y}$ – матрица координат размера $L\times N$ распознаваемого множества объектов; L – размерность пространства признаков; $\mathbf{Z}$ – матрица координат размера $L\times K$ представителей классов.

5. Существует канонический вид гипотетической и рабочей выборок, основанный на свойстве замкнутости классификатора.

Эти гипотезы наталкивают на комплексное решение проблемы классификации:

– определение дисперсионных последовательностей (кривых), образуемых различными разбиениями гипотетической (образцовой) изотропной выборки;

– приведение рабочей выборки к каноническому виду;

– проверка гипотезы о расслоении рабочей выборки на кластеры;

– синтез классификатора;

– верификация полученной математической модели.

В условиях, когда гипотеза о расслоении рабочей выборки на классы подтверждается (проверке этой гипотезы будут посвящены последующие исследования), основной задачей можно считать задачу синтеза классификатора по некоторому функционалу качества.

Функционал качества сформируем в общем виде (по принципу гарантированного результата) без априорного задания числа классов (оно естественно ограничено сверху объемом выборки) и эталонов на множестве нерандомизированных классификаторов, задаваемых как множество всех неупорядоченных разбиений исходного множества объектов [35, 36]. При этом каждое разбиение (классификатор) будем характеризовать двумя параметрами: межклассовой дисперсией и суммарной внутриклассовой дисперсией.

7. Дисперсии разбиений. Известен теоретико-множественный подход к постановке задачи классификации [5]. Однако его применение совместно с математическим программированием [37] невозможно. Поэтому математическую модель представим как некоторую условную экстремальную задачу в формальном матричном виде.

Для этого конкретизируем межклассовую дисперсию.

Следуя Байесовскому подходу, будем рассматривать оценки $\{\mathbf{Z'}\}$ множества эталонов $\{\mathbf{Z}\}$ при заданной функции стоимости (потерь), наблюдая множество признаков $\{\mathbf{Y}\}$. Так как, в ходе решения задачи классификации необходимо определить: 1) априорное множество эталонов $\{\mathbf{Z}\}$; 2) правило принятия решения о принадлежности объекта к классу: 3) апостериорное множество оценок $\{\mathbf{Z'}\}$, то естественно положить, что оценки должны совпадать с эталонами $\{\mathbf{Z}'\} \equiv \{\mathbf{Z}\}$ и также должны быть найдены в результате решения задачи классификации.

Известно [9], что в случае квадратичной функции потерь, а также в достаточно широком классе других функций стоимости, оптимальная оценка есть среднее значение апостериорной плотности вероятности признаков класса. Тогда применительно к задаче детерминированной классификации с необходимыми плотностями вероятностей:

$$W(\mathbf{Z},\mathbf{Y}) = W(\mathbf{Z}/\mathbf{Y})W(\mathbf{Y}) =$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} x_{k,n} \frac{\delta(\mathbf{Z}-\mathbf{Z}_k)1(\mathbf{Z}-\mathbf{Y}_n)}{\sum_{s=1}^{K} x_{s,n}} \delta(\mathbf{Y}-\mathbf{Y}_n)\sum_{s=1}^{K} x_{s,n} =$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} x_{k,n}\, \delta(\mathbf{Z}-\mathbf{Z}_k)1(\mathbf{Y}-\mathbf{Y}_n)\delta(\mathbf{Y}-\mathbf{Y}_n);$$

$$W(\mathbf{Y}) = \sum_{n=1}^{N} \delta(\mathbf{Y}-\mathbf{Y}_n)\sum_{k=1}^{K} x_{k,n};$$

$$W(\mathbf{Z}/\mathbf{Y}) = \sum_{k=1}^{K} \frac{x_{k,n}}{\sum_{s=1}^{K} x_{s,n}} \delta(\mathbf{Z}-\mathbf{Z}_k)1(\mathbf{Y}-\mathbf{Y}_n), \forall\, n = 1,\ldots,N,$$

составим матрицу оптимальных оценок $\mathbf{Z'}$ размера $(L \times N)$, которая может быть выражена как:

$$\mathbf{Z}_n' = \int\limits_{\Omega_\mathbf{Z}} \mathbf{Z} W\left(\mathbf{Z}/\mathbf{Y}\right) d\mathbf{Z} = \mathbf{Z}\mathbf{X}_n' 1\left(\mathbf{Y} - \mathbf{Y}_n\right), \forall n = \overline{1,N};$$

$$\mathbf{Z}' = \left[\mathbf{Z}_1', \quad \mathbf{Z}_2', \quad \cdots, \quad \mathbf{Z}_N'\right] = \mathbf{Z}\mathbf{X}',$$

где $\mathbf{Z}_n'$ – оценка $\mathbf{Z}_n$ при наблюдении вектора признаков $\mathbf{Y}_n$; $\mathbf{Z}_n' \in \left\{\mathbf{Z}_k = (z_{1k}, z_{2k}, \ldots, z_{LK})^\mathrm{T}\right\}, \forall k = \overline{1,K}$; $\mathbf{Z}$ – формируемое эталонное множество – матрица размера $\left(L \times K\right)$; $\mathbf{X} = \mathbf{X}' \mathrm{diag}\left(\mathbf{X}^\mathrm{T}\mathbf{I}_k\right)$ – матрица размера $\left(K \times N\right)$ апостериорных вероятностей эталонов при наблюдении вектора признаков $\mathbf{Y}_n$; $\mathbf{X}_n'$ – n-й столбец матрицы $\mathbf{X}'$.

Тогда можно найти среднее значение на множестве оценок:

$$\mathbf{Z}_s' = \mathbf{Z}\mathbf{X}'\mathbf{X}^\mathrm{T}\mathbf{I}_k,$$

и определить дисперсию оценок как

$$D_m\left(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}\right) = \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n} \int\limits_{\Omega_\mathbf{Y}}\int\limits_{\Omega_\mathbf{Z}} \Big[\left(\mathbf{Z}' - \mathbf{Z}_s'\right)^\mathrm{T} \times$$

$$\times \mathbf{W}\left(\mathbf{Z}' - \mathbf{Z}_s'\right)\Big] \delta\left(\mathbf{Z}' - \mathbf{Z}_n'\right) \delta\left(\mathbf{Y} - \mathbf{Y}_n\right) d\mathbf{Z}\, d\mathbf{Y} =$$

$$= \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n} \left(\mathbf{Z}_n' - \mathbf{Z}_s'\right)^\mathrm{T} \mathbf{W}\left(\mathbf{Z}_n' - \mathbf{Z}_s'\right).$$

Эталоны $\mathbf{Z}_k$ будем искать как матрицу координат эталонов $\mathbf{Z}$ размера $L \times K$. Тогда после подстановки оценок и их среднего, используя фильтрующее свойство δ-функции, получим:

$$D_m\left(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}\right) = \mathrm{tr}\left\{\left(\mathbf{Z}\mathbf{X}' - \mathbf{I}_n^\mathrm{T} \otimes \left\{\mathbf{Z}\mathbf{X}'\mathbf{X}^\mathrm{T}\mathbf{I}_k\right\}\right)^\mathrm{T} \mathbf{W} \times\right.$$

$$\left.\times \left(\mathbf{Z}\mathbf{X}' - \mathbf{I}_n^\mathrm{T} \otimes \left\{\mathbf{Z}\mathbf{X}'\mathbf{X}^\mathrm{T}\mathbf{I}_k\right\}\right) \mathrm{diag}\left(\mathbf{X}^\mathrm{T}\mathbf{I}_k\right)\right\},$$

где $\mathrm{tr}\left(\mathbf{A}\right)$ – след квадратной матрицы $\mathbf{A}$; $\mathbf{A} \otimes \mathbf{B}$ – произведение Кронекера матриц **A** и **B**.

Отметим, что здесь под межклассовой дисперсией понимается не дисперсия эталонного множества, мощность которого равна K, а дисперсия рабочей выборки после замены каждого объекта всех классов его эталонным значением для класса, которому он принадлежит, а значит, мощность преобразованного множества равна N.

Теперь введем и рассмотрим суммарную внутриклассовую (усредненную на множестве признаков класса) дисперсию, которая в общем случае задается выражением:

$$D_v\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)=\sum_{k=1}^{K}\int\limits_{\Omega_{\mathbf{Y}_k}}\int\limits_{\Omega_{\mathbf{Z}_k}}\left(\mathbf{Y}-\mathbf{Z}\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{Y}-\mathbf{Z}\right)W\left(\mathbf{Z},\mathbf{Y}\right)d\mathbf{Z}\,d\mathbf{Y};$$

$$\Omega_{\mathbf{Y}}=\bigcup_k\Omega_{\mathbf{Y}_k},\bigcap_k\Omega_{\mathbf{Y}_k}=\theta,\Omega_{\mathbf{Z}}=\bigcup_k\Omega_{\mathbf{Z}_k},\bigcap_k\Omega_{\mathbf{Z}_k}=\theta,\operatorname{card}\Omega_{\mathbf{Z}_k}=1,\forall k=\overline{1,K};$$

$$D_v\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)=\sum_{k=1}^{K}\sum_{n=1}^{N}x_{k,n}\int\limits_{\Omega_{\mathbf{Y}}}\int\limits_{\Omega_{\mathbf{Z}}}\left(\mathbf{Y}-\mathbf{Z}'\right)^{\mathrm{T}}\mathbf{W}\times$$

$$\times\left(\mathbf{Y}-\mathbf{Z}'\right)\delta\left(\mathbf{Z}'-\mathbf{Z}'_n\right)1\left(\mathbf{Y}-\mathbf{Y}_n\right)\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)d\mathbf{Y}\,d\mathbf{Z}'.$$

Используя фильтрующее свойство δ-функции окончательно получим:

$$D_v\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)=\sum_{k=1}^{K}\sum_{n=1}^{N}x_{k,n}\left(\mathbf{Y}_n-\mathbf{Z}'_n\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{Y}_n-\mathbf{Z}'_n\right),$$

где $\mathbf{Z}_n$ – n-й столбец матрицы $\mathbf{Z}'$.

Суммарную внутриклассовую дисперсию определим с помощью следующей конструкции:

$$D_v\left(K,\mathbf{X},\mathbf{Y},\mathbf{Z}\right)=\mathbf{I}_k^{\mathrm{T}}\mathbf{X}\operatorname{diag}\left[\left(\mathbf{Y}-\mathbf{Z}\mathbf{X}'\right)^{\mathrm{T}}\mathbf{W}\left(\mathbf{Y}-\mathbf{Z}\mathbf{X}'\right)\right]\mathbf{I}_n.$$

Для сокращения размерности пространства решений учтем наличие связи между искомыми переменными и ограничениями задачи. Так:

$$\operatorname{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{I}_k\right)=\frac{1}{N}\mathbf{I},\mathbf{X}=\frac{1}{N}\mathbf{X}',\mathbf{X}'^{\mathrm{T}}\mathbf{I}_k=\mathbf{I}_n,$$

где $\mathbf{I}$ – единичная матрица.

С этой же целью будем искать координаты центроидов как взвешенное среднее векторов, находящихся в кластере:

$$\mathbf{Z}=\mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left(\operatorname{diag}\left(\mathbf{X}'\mathbf{I}_n\right)\right)^{-1}.$$

Полученные выражения для дисперсий содержат явную зависимость от предполагаемого числа классов K, которая выражается числом строк в матрице назначений $\mathbf{X}'$. Для исключения этой зависимости определим эту матрицу для максимально возможного числа классов размером $(N \times N)$.

Тогда выражения для дисперсий записываются в виде:

$$D_m(\mathbf{X}') = \frac{1}{N}\operatorname{tr}\Bigg\{\Bigg(\mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left[\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right]^{-1}\mathbf{X}' - \\ -\frac{1}{N}\mathbf{I}_n^{\mathrm{T}} \otimes \left\{\mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left[\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right]^{-1}\mathbf{X}'\mathbf{I}_n\right\}\Bigg)^{\mathrm{T}} \times \\ \times\mathbf{W}\Bigg(\mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left[\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right]^{-1}\mathbf{X}' - \\ -\frac{1}{N}\mathbf{I}_n^{\mathrm{T}} \otimes \left\{\mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left[\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right]^{-1}\mathbf{X}'\mathbf{I}_n\right\}\Bigg)\Bigg\}, \tag{5}$$

$$D_v(\mathbf{X}') = \frac{1}{N}\mathbf{I}_k^{\mathrm{T}}\mathbf{X}' \times \operatorname{diag}\Big\{\Big[\mathbf{Y} - \mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left(\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right)^{-1}\mathbf{X}'\Big]^{\mathrm{T}}\mathbf{W} \times \\ \times\Big[\mathbf{Y} - \mathbf{W}\mathbf{Y}\mathbf{X}'^{\mathrm{T}}\left(\operatorname{diag}(\mathbf{X}'\mathbf{I}_n)\right)^{-1}\mathbf{X}'\Big]\Big\}\mathbf{I}_n. \tag{6}$$

Максимальные значения (5) и (6) используем для определения канонического вида рабочей выборки, для чего произведем нормировку матрицы признаков $\mathbf{Y}$ по правилу:

$$\mathbf{Y}_s = \frac{\mathbf{Y}}{\sqrt{s}},$$

где $s = D_m(\mathbf{I}) = D_v(\mathbf{I}_n)$, что означает, что максимальное значение межклассовой дисперсии наблюдается, когда число классов равно числу объектов классификации, а максимальное значение суммарной внутриклассовой дисперсии – если класс один. Эти значения одинаковы и равны выборочной дисперсии исходной выборки, что несложно показать (для упрощения $\mathbf{W} = \mathbf{I}$):

$$s = \max\left\{D_m(\mathbf{X}')\right\} = \frac{1}{N}\operatorname{tr}\left\{\left(\mathbf{Y} - \frac{1}{N}\mathbf{I}_n^{\mathrm{T}} \otimes \{\mathbf{Y}\mathbf{I}_n\}\right)^{\mathrm{T}}\left(\mathbf{Y} - \frac{1}{N}\mathbf{I}_n^{\mathrm{T}} \otimes \{\mathbf{Y}\mathbf{I}_n\}\right)\right\},$$

которое получено из (5), и

$$s = \max\left\{Dv\left(\mathbf{X}'\right)\right\} = \frac{1}{N}\mathbf{I}_n^{\mathrm{T}}\,\mathrm{diag}\left[\left(\mathbf{Y} - \frac{1}{N}\mathbf{Y}\,\mathbf{I}_n\,\mathbf{I}_n^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{Y} - \frac{1}{N}\mathbf{Y}\,\mathbf{I}_n\,\mathbf{I}_n^{\mathrm{T}}\right)\right]\mathbf{I}_n,$$

определяемое из (6). Нетрудно заметить, что эти выражения для масштабирующей константы эквивалентны.

8. Функционал качества классификатора. Теперь введем некоторый функционал качества классификатора:

$$\Phi\left(\mathbf{Y}_r, \mathbf{H}_s, \mathbf{X}'\right) \xrightarrow[\mathbf{X}']{} extr, \tag{7}$$

где $\mathbf{Y}_r$ – рабочая выборка канонического вида; $\mathbf{H}_s$ – гипотетическая (образцовая) выборка канонического вида; $\mathbf{X}'$ – искомое разбиение рабочей выборки на классы.

Для конструирования функционала будем использовать зависимости дисперсий канонических рабочей и гипотетической выборок от разбиений на классы: $D_{v_s}\left(\mathbf{X}_i', \mathbf{H}_s\right)$ – суммарная внутриклассовая дисперсия изотропной канонической выборки; $D_{v_s} = D_{v_s}\left(\mathbf{Y}_r, \mathbf{X}'\right)$ – суммарная внутриклассовая дисперсия канонической рабочей выборки.

Чтобы получить формализованное описание функционала (7), были проведены исследования зависимости межклассовой и суммарной внутриклассовой дисперсий от неупорядоченных разбиений путем их полного перебора на различных рабочих выборках. По сути, функционал (7) должен ввести отношение порядка [1] на множестве всех возможных неупорядоченных разбиений, определяемых числом Белла [38] и обладающих следующими свойствами:

– число классов в разбиении не фиксируется и ограничено сверху числом объектов N, что позволяет избавиться от явного задания числа кластеров в задаче классификации;

– условия (2) и (3) формально задают искомые разбиения (область определения функционала (7)), хотя могут описывать и упорядоченные разбиения при дальнейшем развитии задачи классификации;

– полный перебор всех разбиений осуществляется формированием матриц назначений $\mathbf{X}'$ путем использования известного рекуррентного свойства [35] (необходимые свойства разбиений приведены в приложении 4);

– для идентификации оптимального расслоения должна быть некоторая гипотетическая изотропная выборка (образец), в которой потенциально невозможно выделить кластеры;

– значения признаков объектов изотропной выборки ограничены сверху и снизу и образуют компактное [33] множество, чему удовлетворяет выборка из непрерывного многомерного равномерного распределения [31] (его определение приведено в приложении 5).

Для того чтобы здесь не конкретизировать область *S,* приведем изотропную выборку к каноническому виду с помощью вышеприведенных масштабирующих констант *S*. Это должно обеспечить некоторый стандарт дисперсионной последовательности (кривой), определяемой различными разбиениями гипотетической (образцовой) изотропной выборки (рис. 2). Эта последовательность не должна зависеть от размерности признакового пространства и числа объектов классификации, если оно велико. Конечно, этот факт нуждается в строгом математическом доказательстве. Приведем лишь экспериментальное подтверждение существования такой последовательности при малом числе объектов и для двумерного признакового пространства, так как вычислительная сложность экспериментов экспоненциально возрастает с ростом числа объектов и увеличением размерности пространства признаков.

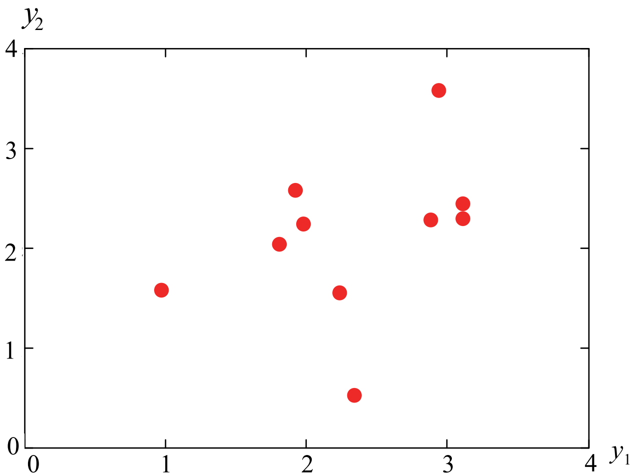

Рис. 2. Корреляционное поле изотропной выборки

На рисунке 3 изображены зависимости суммарной внутриклассовой дисперсии D_{v_S} от разбиений изотропных канонических выборок различных объемов. При этом разбиения упорядочены по возрастанию дисперсии, а значит, образуют монотонную неубывающую последовательность. Значения по оси абсцисс не имеют физического смысла, а лишь образуют порядковую шкалу *v* на множестве разбиений.

Здесь использовались выборки объемом: $N = 7, 8, 9, 10,$ для которых числа Белла равны: $B = 877, 4140, 21147, 115975$ соответственно.

Представленные графики позволяют заметить особенности последовательности:

– дисперсия изменяется дискретно плавно без резких скачков с ростом объема выборки;

– при увеличении числа классифицируемых объектов кривая, отображающая последовательность, стремится к некоторому предельному положению;

– при $N \to \infty$ дисперсионная кривая изотропной канонической выборки может носить характер устойчивой закономерности.

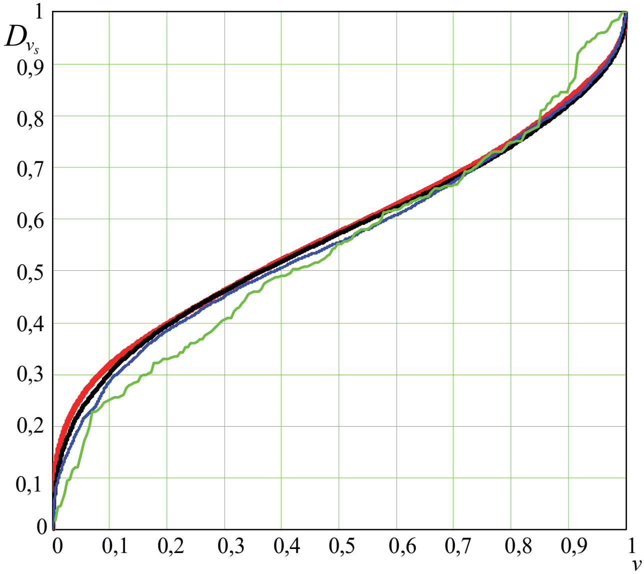

Рис. 3. Дисперсионные кривые изотропной выборки

9. Дисперсионные кривые. Теперь рассмотрим поведение дисперсионной кривой анизотропной выборки, то есть выборки с возможным расслоением на классы (рис. 4). Из большого количества экспериментов по исследованию анизотропных выборок в работе обсуждается один из них (наиболее представительный). Выявленные свойства дисперсионной кривой устойчиво сохраняются и в других экспериментах.

На рисунке 5 изображены зависимости суммарной внутриклассовой дисперсии D_{v_s} и межклассовой дисперсии D_{m_s} от разбиений изотропной (рис. 2) и анизотропной (рис. 4) канонических выборок объема $N = 10$. При этом разбиения упорядочены по уменьшению межклассовой дисперсии, а значит, образуют монотонную невозрастающую последовательность. Здесь же на графике изображена сумма внутриклассовой и межклассовой дисперсий выборок канонического вида. Эта сумма не зависит от разбиения исходной рабочей совокупности и равна ее общей дисперсии, а после приведения к каноническому виду соответствует единичному значению.

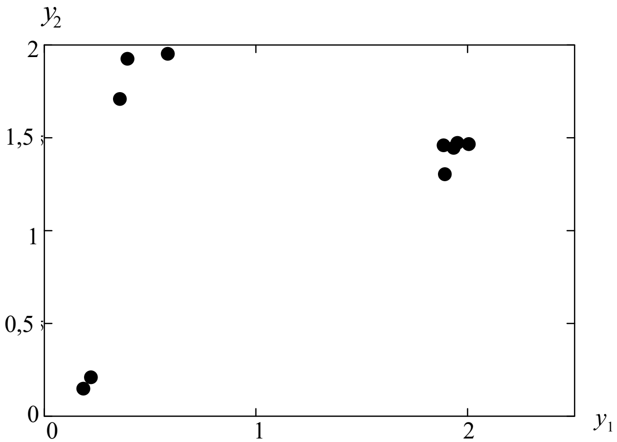

Рис. 4. Корреляционное поле анизотропной выборки

Значения по оси абсцисс на рисунке 5, как и ранее на рисунке 3, не имеют физического смысла, а лишь образуют порядковую шкалу v на множестве разбиений.

Установлено, что дисперсионные кривые изотропной и анизотропной выборок отличаются друг от друга тем больше, чем больше неоднородность анизотропной выборки. При дальнейших исследованиях это свойство может быть использовано для проверки гипотезы о расслоении статистического материала на классы.

Дискретные зависимости на рисунке 5 аппроксимированы кусочно-линейными непрерывными кривыми. Поэтому детализируем их с точностью до дискретных отсчетов. Особый интерес представляет условие, когда число классов значительно меньше, чем объем выборки: $K << N$ при малых (больших) значениях суммарной внутриклассовой дисперсии (межклассовой дисперсии). Для этих условий представлены зависимости на рисунках 6 и 7 суммарной внутриклассовой D_{v_S} и межклассовой D_{m_S} дисперсий соответственно.

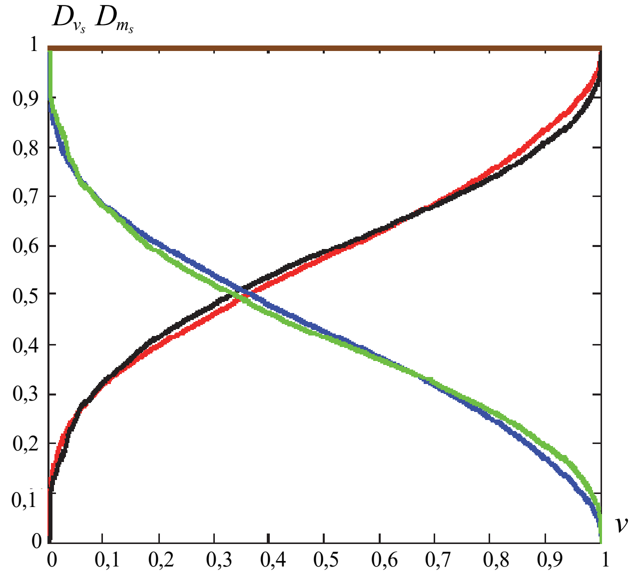

Рис. 5. Дисперсионные кривые изотропной и анизотропной выборок

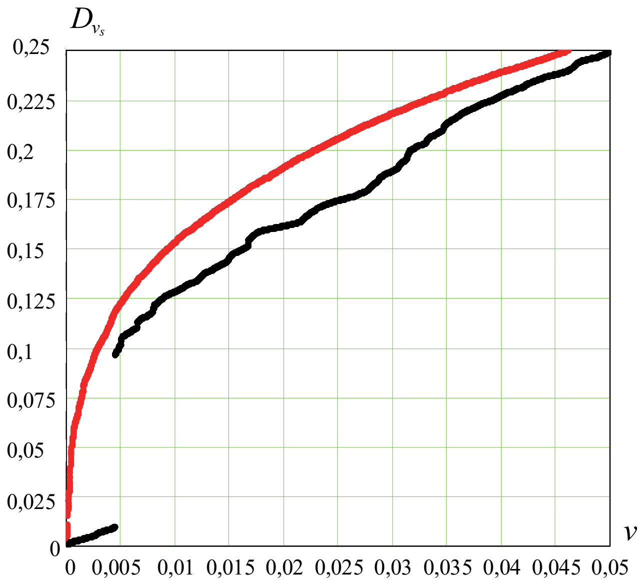

Рис. 6. Внутриклассовые дисперсионные последовательности изотропной и анизотропной выборок

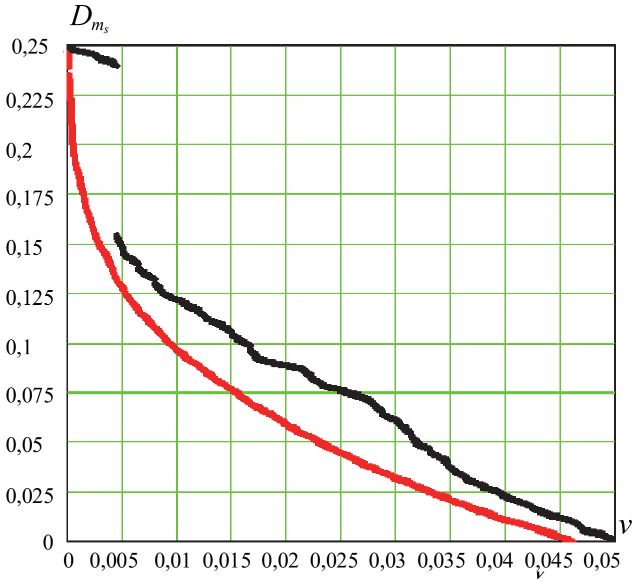

Рис. 7. Межклассовые дисперсионные последовательности изотропной и анизотропной выборок

В обоих случаях расчета дисперсий по (5) и (6) на графиках наблюдаются разрывы области допустимых значений функционалов для анизотропной выборки. Используем это свойство для формирования функционала качества классификации (7).

10. Задача классификации. Так многочисленные исследования (эксперименты) показали, что при наличии классов наблюдается неоднородность выборки. Эта неоднородность проявляется в виде существенных скачков межклассовой дисперсии (суммарной внутриклассовой дисперсии) разбиений, упорядоченных по их убыванию (возрастанию). Наибольший выброс идентифицирует наилучшее разбиение на классы. Значит надо искать минимум (максимум) производной какой-либо из дисперсий на множестве всех возможных разбиений. Однако производной в классическом смысле здесь не существует, так как производная слева не равна производной справа и, кроме того, функционал определен на дискретном множестве [39, 40]. По сути, это означает, что происходит качественное изменение количественной характеристики разбиения при небольшом изменении его параметров. Тогда задачу классификации можно отнести к задачам теории катастроф.

Важно отметить, что в случае изотропных выборок катастрофы нет, и она наблюдается, когда существует некоторая организация в системе, генерирующей анизотропную выборку. Поэтому конструкция функционала (7) должна учитывать, в какой мере решение задачи классификации отлично от наиболее близкого решения при замене анизотропной выборки изотропной. Это можно сформулировать в виде следующей минимаксной задачи:

$$\min_{\mathbf{X}_a^{'}} \left| D_{v_S}\left(\mathbf{X}_i^{'}, \mathbf{H}_s\right) - D_{v_S}\left(\mathbf{X}_a^{'}, \mathbf{Y}_r\right) \right| \to \max_{\mathbf{X}_i^{'}}, \tag{8}$$

где $D_{v_S}\left(\mathbf{X}_i^{'}, \mathbf{H}_s\right)$ – суммарная внутриклассовая дисперсия изотропной канонической выборки; $D_{v_S}\left(\mathbf{X}_a^{'}, \mathbf{Y}_r\right)$ – суммарная внутриклассовая дисперсия анизотропной канонической выборки; $\mathbf{X}_i^{'}$ – матрица назначений объектов в классы изотропной канонической выборки; $\mathbf{X}_a^{'}$ – матрица назначений объектов в классы анизотропной канонической выборки; $\mathbf{H}_s$ – матрица признаков объектов изотропной канонической выборки; $\mathbf{Y}_r$ – матрица признаков объектов анизотропной канонической выборки.

Аналогичный функционал может быть записан и для межклассовой дисперсии:

$$\min_{\mathbf{X}_a^{'}} \left| D_{m_S}\left(\mathbf{X}_i^{'}, \mathbf{H}_s\right) - D_{m_S}\left(\mathbf{X}_a^{'}, \mathbf{Y}_r\right) \right| \to \max_{\mathbf{X}_i^{'}}.$$

Поскольку дисперсионная кривая изотропной выборки уникальна, она может быть табулирована или аппроксимирована с заданной точностью с помощью некоторой функции $f_i\left(x\right)$. Теперь задачу (8) можно записать как:

$$\min_{\mathbf{X}_a^{'}} \left| f_i\left(x\right) - D_{v_S}\left(\mathbf{X}_a^{'}, \mathbf{Y}_r\right) \right| \to \max_{x}. \tag{9}$$

Тогда, математическую модель можно представить как задачу нахождения матрицы назначений $\mathbf{X}_a^{'}$ и вспомогательной переменной x, удовлетворяющих (10) при известных матрицах признаков объектов $\mathbf{Y}_r$, обобщенной метрики эвклидова пространства $\mathbf{W}$ и эталонной дисперсионной кривой $f_i\left(x\right)$, при ограничениях на компоненты матрицы назначений (2), (3). Вычисление (9) производится с помощью выражения (6).

Задачи (3), (2), (6), (9) можно классифицировать как задачи целочисленного нелинейного программирования на булевом множестве решений, которое задается с помощью нелинейных уравнений. В этой задаче неявно осуществляется оптимизация по числу классов, так как на строки матрицы $\mathbf{X}_a^{'}$ не накладывается никаких дополнительных ограничений, кроме булевости переменных. Это может приводить к наличию нулевых (пустых) кластеров вследствие ограничения снизу и сверху числа классов ($1 \leq K \leq N$), а также вследствие предлагаемой минимаксной математической модели задачи классификации. Разработке алгоритма решения поставленной задачи будут посвящены последующие работы.

Решение задач (9), (6), (2), (3) определяет наилучшее разбиение множества объектов на классы, а по достигнутому значению функционала (9) можно судить о качестве классификации.

В целом можно выделить следующие основные этапы решения задачи классификации:

1. Определение эталонной дисперсионной кривой изотропной выборки.
2. Нормирование рабочей выборки относительно ее общей дисперсии, то есть приведение рабочей выборки к каноническому виду.
3. Проверка гипотезы о расслоении на классы, а именно: можно выделить классы или нет?
4. Решение задачи синтеза классификатора, то есть поиск оптимального разбиения в случае истинности гипотезы о расслоении на классы.
5. Верификация полученного решения и определение качества классификации.

11. Заключение. В рамках общей задачи теории принятия решений при байесовском подходе задача классификации признаков объектов с точки зрения разбиения множества на подмножества в условиях правильной классификации представляет собой модель системы распознавания, где функция правдоподобия определяется в ходе решения задачи, а класс правил распознавания образован на множестве известных признаков объектов, что определяет теоретическую значимость результатов работы. Тогда задача классификации сводится к задаче определения вероятностей принадлежности исследуемых объектов к некоторым классам. При этом постулируется тождественность эталонного множества и множества оценок классов.

В результате разработана математическая модель классификатора признаков, которая предполагает синтез двумерной плотности вероятностей в пространстве координат: классы – объекты. Использование обобщенных функций позволило свести вероятностную задачу поиска минимума байесовского риска к детерминированной задаче на множестве нерандомизированных правил принятия решений при помощи специально введенных ограничений.

В подобных условиях для корректного синтеза классификатора необходима дисперсионная кривая изотропной выборки, а также использование суммарной внутриклассовой и межклассовых дисперсий для характеристики качества классификации, что предполагает интерпретацию задачи классификации как частной задачи теории катастроф. Ограниченность исходных данных предопределила поиск минимаксного функционала, отражающего качество классификации при квадратичной функции потерь. В итоге разработанную математическую модель оказалось возможным рассматривать в форме задачи целочисленного нелинейного программирования, а последующее использование полиномиальных ограничений позволило ее преобразовать к виду общей задачи нелинейного непрерывного программирования, что обуславливает практическую значимость работы, как способа решения NP – полных комбинаторных задач. В дополнение к этому найдены необходимые условия расслоения на классы, которые могут быть использованы как достаточные при проверке гипотезы о существовании классов.

Дальнейшие исследования в области разработки моделей классификаторов должны быть направлены на синтез динамических классификаторов [41], для чего необходимо:

– разделить представленный статистический материал на обучающую и контролирующую выборки;

– найти эффективный алгоритм получения решения задачи классификации в рамках представленной математической модели;

– выполнить верификацию разработанной модели классификатора и по возможности проверку его адекватности;

– предусмотреть процедуру обучения классификатора, как в признаковом пространстве объектов для снижения его размерности (отбора информативных признаков), так и в пространстве самих объектов для расслоения их на классы;

– определить, при каких условиях задачу синтеза классификатора нужно решать заново в динамике изменения признаков объектов.

Литература

1. *Блауберг И.В.* Проблемы методологии системного исследования // М.: Мысль. 1970. 454 с.
2. *He H.* A Deep Research in Classic Classification Network // IOP Conf. Series: Materials Science and Engineering. 2020. 740 p.
3. *Wehrmann J., Cerri R., Barros R.C.* Hierarchical Multi-Label Classification Networks // Proceedings of the 35th International Conference on Machine Learning, Stockholm (PMLR 80). 2018. pp. 5075–5084.
4. *Li P., Wang D., Wang L., Lu H.* Visual Tracking by Dynamic Matching-Classification Network Switching // Pattern Recognition. 2020. pp. 107419.

5. *Айвазян С.А.* Прикладная статистика. Классификация и снижение размерности // М.: Финансы и статистика. 1989. 606 с.
6. *Pontone S., Grimaldi G.* What is the Technique without a Proper Classification? // Int J Gastroenterol Disord Ther. 2015. pp.113.
7. *Nwafor G., Onwukwe C.* On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis // American Journal of Applied Mathematics and Statistics. 2014. pp. 394–397.
8. *Мандель И.Д.* Кластерный анализ // М.: Финансы и статистика. 1988. 176 с.
9. *Галлагер Р.Дж.* Теория информации и надежная связь // М.: Сов. Радио. 1974. 719 с.
10. *Xiaotong L. et al.* Learning a Deep Vector Quantization Network for Image Compression // IEEE WESCANEX 93 Communications, Computers and Power in the Modern Environment – Conference Proceedings. 1993. pp. 299–312.
11. *Janabi Samaher A.-J., Abed S.M.* Digital Video Scenes Recognition using M_{ijn}-EA and Learning Vector Quantization Network // Journal of Babylon University. vol. 9. no. 24. 2016. pp. 2362–2373.
12. *Ван Трис Г.Л.* Теория обнаружения, оценок и модуляции // М.: Сов. радио. 1972. 744 с.
13. *Батенков К.А.* Синтез детерминированных нелинейных дискретных отображений непрерывных каналов связи // Труды СПИИРАН. 2016. № 2(45). С. 75–101.
14. *Kipping D.* A Bayesian Approach to the Simulation Argument // Universe. 2020. vol. 6. no. 8. pp. 109.
15. *Mihnea A., John H.* A Bayesian Approach for Asset Allocation // International Journal of Statistics and Probability. 2020. vol. 4. pp. 1–14.
16. *Тихонов В.И.* Марковские процессы // М.: Сов. радио. 1977. 488 с.
17. *Schlegel M., White A., Patterson A., White M.* General Value Function Networks. 2018. URL: arxiv.org/pdf/1807.06763v1.pdf (дата обращения: 21.10.2020).
18. *Батенков К.А.* Точные и граничные оценки вероятностей связности сетей связи на основе метода полного перебора типовых состояний // Труды СПИИРАН. 2019. Т. 18. № 5. С. 1093–1118.
19. *Taketoshi Y., Masahiro I.* Extended responsibility assignment matrix (ERAM) suitable for a cross functional project // EDULEARN17 Proceedings.2017. pp. 4825–4834.
20. *Qingge J., Haoqiang Y.* Online Multiple Object Tracking with Reid Feature Extraction Network and Similarity Matrix Network // Journal of Physics: Conference Series, Volume 1544, 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP). 2020. pp. 20–22.
21. *Эндрюс Г.* Теория разбиений // М.: Наука. 1982. 255 с.
22. *Menchaca B. et al.* Technique for setting network communication parameters : publ. no. WO/2011/008515, publ. date 20.01.2011, int. appl. No. PCT/US2010/040298, int. fil. date 29.06.2010.
23. *Rahman, H., Sheikh, N.U., Saleh Al-Qahtani, H., Hazra, T.K.* Partitioned network with Adaptive Mobile Sinks // 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2019. pp. 1098–1103.
24. *Chalupa D.* Partitioning Networks into Cliques: A Randomized Heuristic Approach // Information Sciences and Technologies Bulletin of ACM Slovakia. 2014. vol. 6. pp. 1–8.
25. *Ланнэ А.А.* Нелинейные динамические системы: синтез, оптимизация, идентификация // Л.: ВАС. 1985. 240 с.
26. *Kostrikov I., Bruna J., Panozzo D., Zorin D.* Surface Networks. 2017. URL: arxiv.org/pdf/1705.10819.pdf (дата обращения: 21.10.2020).
27. *Pinghua G., Jieping Y., Changshui Z.* Robust multi-task feature learning // KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. August 2012. pp. 895–903.

28. *Гантмахер Ф.Р.* Теория матриц // М.: Физматлит. 2010. 559 с.
29. *Lee S., Song B.* Transformation of Non-Euclidean Space to Euclidean Space f.or Efficient Learning of Singular Vectors // IEEE Access. 2020. vol. 8. pp. 127074–127083.
30. *Bandyapadhyay S., Fomin F., Simonov K.* On Coresets for Fair Clustering in Metric and Euclidean Spaces and Their Applications. 2020. URL: arxiv.org/pdf/2007.10137.pdf (дата обращения: 21.10.2020).
31. *Королюк В.С. , Портенко Н.И. , Скороход А.В., Турбин А.Ф.* Справочник по теории вероятностей и математической статистике // М.: Наука. 1985. 640 с.
32. *Borzov A.B., Labunets L.V., Steshenko V.B.* Noncanonical Spectral Model of Multidimensional Uniform Random Fields // J. Comput. Syst. Sci. Int. 2017. vol. 57. pp. 874–889.
33. *Федоров В.В.* Численные методы максимина // М.: Наука. 1979. 278 с.
34. *Кловский Д.Д., Конторович В.Я., Широков С.М.* Статистическая теория связи // М.: Связь. 1974. 247 с.
35. *Kreher D.L.* Combinatorial algorithms : Generation, enumeration, a. search // CRC press, Cop. 1999. 329 p.
36. *Батенков А.А., Батенков К.А., Фокин А.Б.* Методы формирования множеств состояний телекоммуникационных сетей для различных мер связности // Труды СПИИРАН. 2020. № 3 (19). С. 644–673.
37. *Химмельблау Д.* Прикладное нелинейное программирование // М.: Мир. 1975. 534 с.
38. *Яблонский С.В.* Введение в дискретную математику : учебное пособие для студентов вузов, обучающихся по специальности «Прикладная математика» // М.: Высш. шк. 2006. 384 с.
39. *Батенков К.А.* Числовые характеристики структур сетей связи // Труды СПИИРАН. 2017. № 4(53). С. 5–28.
40. *Батенков А.А.,. Батенков К.А.* Анализ и синтез структур сетей связи по детерминированным показателям устойчивости // Труды СПИИРАН. 2018. № 3(58). С. 128–159.
41. *Nooka S.P. et al.* Adaptive hierarchical classification networks // 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. pp. 3578–3583.

Батенков Александр Александрович – д-р техн. наук, профессор, профессор кафедры, кафедра электроники, радиотехники и систем связи, Орловский государственный университет имени И.С. Тургенева (ОГУ). Область научных интересов: статистическая теория связи, модели и методы обработки сигналов, вычислительные алгоритмы. Число научных публикаций – 110. pustur@yandex.ru; ул. Комсомольская, 95, 302026, Орел, Россия; р.т.: +7-906-570-16-66.

Батенков Кирилл Александрович – д-р техн. наук, доцент, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: статистическая теория связи, модели и методы обработки сигналов, анализ качества сетей связи. Число научных публикаций – 152. pustur@yandex.ru; ул. Приборостроительная, 35, 302015, Орел, Россия; р.т.: 4862-54-97-63.

Богачёв Андрей Геннадьевич – канд. техн. наук, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: модели и методы обработки сигналов. Число научных публикаций – 26. pustur@yandex.ru; ул. Приборостроительная, 35, 302015, Орел, Россия; р.т.: 4862-54-97-63.

Мишин Владислав Владимирович – канд. техн. наук, доцент, заведующий кафедрой, кафедра электроники, радиотехники и систем связи, Орловский государственный университет имени И.С. Тургенева (ОГУ). Область научных интересов: вычислительные алгоритмы. Число научных публикаций – 106. pustur@yandex.ru; ул. Комсомольская, 95, 302026, Орел, Россия; р.т.: +7-906- 570-16-66.

ПРИЛОЖЕНИЕ 1.

Вывод выражения для апостериорной плотности вероятности принадлежности *n*-ого объекта *k*-ому классу:

$$W\left(\mathbf{Z}/\mathbf{Y}\right)=\frac{\sum_{k=1}^{K}\sum_{n=1}^{N}x_{k,n}\,\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)}{\sum_{n=1}^{N}\sum_{k=1}^{K}x_{k,n}\,\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)}=$$

$$=\frac{\sum_{k=1}^{K}\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)\sum_{n=1}^{N}x_{k,n}\,\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)}{\sum_{n=1}^{N}\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)\sum_{k=1}^{K}x_{k,n}}=$$

$$=\sum_{k=1}^{K}\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)\sum_{n=1}^{N}x_{k,n}\frac{\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)}{\sum_{n=1}^{N}\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)\sum_{k=1}^{K}x_{k,n}}=$$

$$=\sum_{k=1}^{K}\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)\sum_{n=1}^{N}x_{k,n}\frac{\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)}{\sum_{n=1}^{N}\delta\left(\mathbf{Y}-\mathbf{Y}_n\right)\sum_{k=1}^{K}x_{k,n}}=$$

$$=\sum_{k=1}^{K}x_{k,n}\frac{\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)1\left(\mathbf{Y}-\mathbf{Y}_n\right)}{\sum_{k=1}^{K}x_{k,n}},$$

$$W\left(\mathbf{Z}\right)=\sum_{k=1}^{K}\delta\left(\mathbf{Z}-\mathbf{Z}_k\right)\sum_{n=1}^{N}x_{k,n}.$$

ПРИЛОЖЕНИЕ 2.

Свойства специальной разрывной функции $1\left(\mathbf{Z}\right)$. По сути, эта функция определяет отношение

$$1\left(\mathbf{Z}\right)=\frac{\delta\left(\mathbf{Z}-\mathbf{Z}_i\right)}{\delta\left(\mathbf{Z}\right)}\ \textit{при}\ \mathbf{Z}_i\to\mathbf{Z};$$

и поэтому так же является обобщенной [16, 17] функцией.

В качестве приближения $1(\alpha, \mathbf{Z})$ -функции к идеальной предлагается использовать экспоненциальную функцию вида:

$$\varepsilon(\alpha, \mathbf{Z}) = e^{-\alpha\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)}, \alpha \to \infty,$$

где $\alpha -$ параметр настройки аппроксимации разрывной функции непрерывной дифференцируемой функцией.

На рисунке П2 изображена $\varepsilon(\alpha, x)$ при различных значениях α:

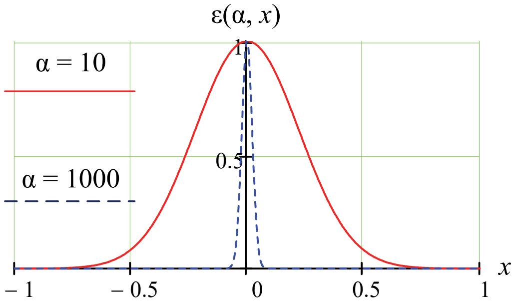

Рис. П2. График функции $\varepsilon(\alpha, x)$ при различных значениях α

ПРИЛОЖЕНИЕ 3.

Условия нормировки для введенных плотностей. Целесообразно их записать в виде:

$$\int\limits_{\Omega\mathbf{Z}}\int\limits_{\Omega\mathbf{Y}} W(\mathbf{Z},\mathbf{Y})\,d\mathbf{Z}\,d\mathbf{Y} = \int\limits_{\Omega\mathbf{Z}}\int\limits_{\Omega\mathbf{Y}} \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n}\,\delta(\mathbf{Z}-\mathbf{Z}_k)\,\delta(\mathbf{Y}-\mathbf{Y}_n)\,d\mathbf{Z}\,d\mathbf{Y} =$$

$$= \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n} = 1;$$

$$\int\limits_{\Omega_Z} W(\mathbf{Z}/\mathbf{Y})\,d\mathbf{Z} = \int\limits_{\Omega_Z} \sum_{k=1}^{K} \frac{x_{k,n}}{\sum_{s=1}^{K} x_{s,n}}\,\delta(\mathbf{Z}-\mathbf{Z}_k)\,1(\mathbf{Y}-\mathbf{Y}_n)\,d\mathbf{Z} =$$

$$= \sum_{k=1}^{K} \frac{x_{k,n}}{\sum_{s=1}^{K} x_{s,n}}\,1(\mathbf{Y}-\mathbf{Y}_n) = 1(\mathbf{Y}-\mathbf{Y}_n), \forall n = \overline{1, N};$$

$$\int_{\Omega_{\mathbf{Z}}} W(\mathbf{Z})\,d\mathbf{Z} = \int_{\Omega_{\mathbf{Z}}} \sum_{k=1}^{K}\sum_{n=1}^{N} x_{k,n}\,\delta(\mathbf{Z}-\mathbf{Z}_k)\,d\mathbf{Z} = 1;$$

$$\int_{\Omega_{\mathbf{Y}}} W(\mathbf{Y}/\mathbf{Z})\,d\mathbf{Y} = \int_{\Omega_{\mathbf{Y}}} \sum_{n=1}^{N} \frac{x_{k,n}}{\sum_{s=1}^{N} x_{k,s}} 1(\mathbf{Z}-\mathbf{Z}_k)\,\delta(\mathbf{Y}-\mathbf{Y}_n)\,d\mathbf{Y} =$$

$$= \sum_{n=1}^{N} \frac{x_{k,n}}{\sum_{s=1}^{N} x_{k,s}} 1(\mathbf{Z}-\mathbf{Z}_k) = 1(\mathbf{Z}-\mathbf{Z}_k), \forall k = \overline{1,K};$$

$$\int_{\Omega_{\mathbf{Y}}} W(\mathbf{Y})\,d\mathbf{Y} = \int_{\Omega_{\mathbf{Y}}} \sum_{n=1}^{N}\sum_{k=1}^{K} x_{k,n}\,\delta(\mathbf{Y}-\mathbf{Y}_n)\,d\mathbf{Y} = 1,$$

где $\mathbf{\Omega_X}$ – область определения переменных $\mathbf{X}$ (скаляра, вектора матрицы и т. п.)

ПРИЛОЖЕНИЕ 4.

Необходимые свойства разбиений. Разбиение Q_n множества $\{1, \ldots, n\}$ однозначно определяет разбиение Q_{n-1} множества $\{1, \ldots, n-1\}$, которое получается из Q_n после удаления элемента n (или пустого блока) из соответствующего блока. Также, если имеется разбиение $W = \{B_1, \ldots, B_k\}$ множества $\{1, \ldots, n-1\}$, то можно отыскать все разбиения Q_n множества $\{1, \ldots, n\}$, для которых $Q_{n-1} = W$, то есть такие разбиения:

$$B_1 \cup \{n\}, \ldots, B_k; \ldots, B_1, \ldots, B_k \cup \{n\}; B_1, \ldots, B_k, \{n\}.$$

ПРИЛОЖЕНИЕ 5.

Определение непрерывного многомерного равномерного распределения [31]. Пусть $S \subset R^n$ – борелевское множество с конечной лебеговой мерой $\lambda(S)$; говорят, что вектор $(\xi_1, \ldots, \xi_n)$ имеет равномерное распределение в области S, если плотность совместного распределения $f_{(\xi_1,\ldots,\xi_n)}(x_1, \ldots, x_n)$ постоянна в области S и равна нулю вне этой области:

$$f_{(\xi_1,\ldots,\xi_n)}(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\lambda(S)}, (x_1, \ldots, x_n) \in S, \\ 0, (x_1, \ldots, x_n) \notin S. \end{cases}$$

DOI 10.15622/ia.2020.19.6.2

A. BATENKOV, K. BATENKOV, A. BOGACHYOV, V. MISHIN

MATHEMATICAL MODEL OF OBJECT CLASSIFIER BASED ON BAYESIAN APPROACH

Batenkov A., Batenkov K., Bogachyov A., Mishin V. **Mathematical Model of Object Classifier based on Bayesian Approach**.

Abstract. The paper claims that the primary importance in solving the classification problem is to find the conditions for dividing the General complexity into classes, determine the quality of such a bundle, and verify the classifier model. We consider a mathematical model of a non-randomized classifier of features obtained without a teacher, when the number of classes is not set a priori, but only its upper bound is set. The mathematical model is presented in the form of a statement of a minimax conditional extreme task, and it is a problem of searching for the matrix of belonging of objects to a class, and representative (reference) elements within each class. The development of the feature classifier is based on the synthesis of two-dimensional probability density in the coordinate space: classes-objects. Using generalized functions, the probabilistic problem of finding the minimum Bayesian risk is reduced to a deterministic problem on a set of non-randomized classifiers. At the same time, the use of specially introduced constraints fixes non-randomized decision rules and plunges the integer problem of nonlinear programming into a General continuous nonlinear problem. For correct synthesis of the classifier, the dispersion curve of the isotropic sample is necessary. It is necessary to use the total intra-class and inter-class variance to characterize the quality of classification. The classification problem can be interpreted as a particular problem of the theory of catastrophes. Under the conditions of limited initial data, a minimax functional was found that reflects the quality of classification for a quadratic loss function. The developed mathematical model is classified as an integer nonlinear programming problem. The model is given using polynomial constraints to the form of a General problem of nonlinear continuous programming. The necessary conditions for the bundle into classes are found. These conditions can be used as sufficient when testing the hypothesis about the existence of classes.

Keywords: Non-randomized Feature Classifier, Upper Bound of the Number of Classes, Minimax, Conditional Extreme Problem, Integer Problem of Nonlinear Programming

Batenkov Aleksandr – Ph.D., Dr.Sci., Professor, Professor, Department of Electronics, Radio Engineering and Communication Systems, Orel State University named after I.S. Turgenev (OSU). Research interests: statistical communication theory, models and methods of signal processing, computational algorithms. The number of publications – 110. pustur@yandex.ru; 95, Komsomolskaya str., 302026, Orel, Russia; office phone: +7-906-570-16-66.

Batenkov Kirill – Ph.D., Dr.Sci., Associate Professor, Employee, Academy of Federal Guard Service. Research interests: statistical communication theory, models and methods of signal processing, communications network quality analysis. The number of publications – 152. pustur@yandex.ru; 35, Priborostroitelnaya str., 302015, Orel, Russia; office phone: 4862-54-7-63.

Bogachev Andrey – Ph.D., Employee, Academy of Federal Guard Service. Research interests: models and methods of signal processing. The number of publications – 26. pustur@yandex.ru; 35, Priborostroitelnaya str., 302015, Orel, Russia; office phone: 4862-54-7-63.

Mishin Vladislav – Ph.D., Associate Professor, Head of a Department, Department of Electronics, Radio Engineering and Communication Systems, Orel State University named after I.S. Turgenev (OSU). Research interests: computational algorithms. The number of publications – 106. pustur@yandex.ru; 95, Komsomolskaya str., 302026, Orel, Russia; office phone: +7-906- 570-16-66.

References

1. Blauberg I.V. *Problemy metodologii sistemnogo issledovaniya* [Problems of system research methodology]. M.: Mysl. 1970. 454 p. (In Russ.).
2. He H. A Deep Research in Classic Classification Network. IOP Conf. Series: Materials Science and Engineering. 2020. 740 p.
3. Wehrmann J., Cerri R., Barros R.C. Hierarchical Multi-Label Classification Networks. Proceedings of the 35th International Conference on Machine Learning (PMLR 80). 2018. pp. 5075–5084.
4. Li P., Wang D., Wang L., Lu H. Visual Tracking by Dynamic Matching-Classification Network Switching. Pattern Recognition. 2020. pp. 107419.
5. Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika. Klassifikaciya isnizhenie razmernosti :Sprav. izd.* [Applied statistics. Classification and reduction of dimension: Reference ed.]. M.: Finansy i statistika. 1989. 606 p. (In Russ.).
6. Pontone S., Grimaldi G. What is the Technique without a Proper Classification?. *Int J Gastroenterol Disord Ther.* 2015. pp. 113.
7. Nwafor G., Onwukwe C. On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis. *American Journal of Applied Mathematics and Statistics.* 2014. pp. 394–397.
8. Mandel I.D. *Klasternyj analiz* [Cluster analysis]. M.: Finansy i statistika. 1988. 176 p. (In Russ.).
9. Gallager R.G. Information Theory and Reliable Communication. Springer. 1968. (Russ. ed.: Gallager R.G. Teoriya informacii i nadezhnaya svyaz. M.: Sov. Radio. 1974. 719 p.).
10. Xiaotong L. et al. Learning a Deep Vector Quantization Network for Image Compression. IEEE WESCANEX 93 Communications, Computers and Power in the Modern Environment – Conference Proceedings. 1993. pp. 299–312.
11. Janabi Samaher A.-J., Abed S.M. Digital Video Scenes Recognition using M_{ijn}-EA and Learning Vector Quantization Network. *Journal of Babylon University*. vol. 9. no. 24. 2016. pp. 2362–2373.
12. Van Trees H.L. Detection, estimation, and linear modulation theory, Hoboken, New Jersey: John Wiley & Sons, Inc. 1968. (Russ. ed.: Van Trees H.L. Teoriya obnaruzheniya, ocenok i modulyacii. M.: Sov. radio. 1972. 744 p.).
13. Batenkov K.A. [Synthesis of deterministic nonlinear discrete mappings of continuous communication channels]. *Trudy SPIIRAN – Proceedings of SPIIRAS.* 2016. vol. 2(45). pp. 75–101 (In Russ.).
14. Kipping D. A Bayesian Approach to the Simulation Argument. *Universe.* 2020. vol. 6. no. 8. pp. 109.
15. Mihnea A., John H. A Bayesian Approach for Asset Allocation. *International Journal of Statistics and Probability.* 2020. vol. 4. pp. 1–14.
16. Tihonov V.I., *Markovskie process* [Markov processes]. M.: Sov. radio. 1977. 488 p. (In Russ.).
17. Schlegel M., White A., Patterson A., White M. General Value Function Networks. 2018. Available at: arxiv.org/pdf/1807.06763v1.pdf (accessed: 21.10.2020).
18. Batenkov K. A. [Accurate and boundary estimates of the connectivity probabilities of communication networks based on the method of full iteration of typical States]. *Trudy SPIIRAN – Proceedings of SPIIRAS.* 2019. Issue 18. vol. 5. pp. 1093–1118 (In Russ.).
19. Taketoshi Y., Masahiro I. Extended responsibility assignment matrix (ERAM) suitable for a cross functional project. EDULEARN17 Proceedings.2017. pp. 4825–4834.
20. Qingge J., Haoqiang Y. Online Multiple Object Tracking with Reid Feature Extraction Network and Similarity Matrix Network. *Journal of Physics: Conference Series, Volume 1544, 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP).* 2020. pp. 20–22.

21. Endryus G. *Teoriya razbienij* [Theory of partitions]. M.: Nauka. 1982. 255 p.
22. Menchaca B. et al. Technique for setting network communication parameters : publ. no. WO/2011/008515, publ. date 20.01.2011, int. appl. No. PCT/US2010/040298, int. fil. date 29.06.2010.
23. Rahman, H., Sheikh, N.U., Saleh Al-Qahtani, H., Hazra, T.K. Partitioned network with Adaptive Mobile Sinks. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2019. pp. 1098–1103.
24. Chalupa D. Partitioning Networks into Cliques: A Randomized Heuristic Approach. *Information Sciences and Technologies Bulletin of ACM Slovakia*. 2014. vol. 6. pp. 1–8.
25. Lanne A.A. *Nelinejnye dinamicheskie sistemy: sintez, optimizaciya, identifikaciya* [Nonlinear dynamical systems: synthesis, optimization, identification]. L.: VAS. 1985. 240 p. (In Russ.).
26. Kostrikov I., Bruna J., Panozzo D., Zorin D. Surface Networks. 2017. Available at: arxiv.org/pdf/1705.10819.pdf (accessed: 21.10.2020).
27. Pinghua G., Jieping Y., Changshui Z. Robust multi-task feature learning. KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. August 2012. pp. 895–903.
28. Gantmaher F.R. *Teoriya matric* [Matrix Theory]. M.: Fizmatlit. 2010. 559 p. (In Russ.).
29. Lee S., Song B. Transformation of Non-Euclidean Space to Euclidean Space f.or Efficient Learning of Singular Vectors. *IEEE Access.* 2020. vol. 8. pp. 127074–127083.
30. Bandyapadhyay S., Fomin F., Simonov K. On Coresets for Fair Clustering in Metric and Euclidean Spaces and Their Applications. 2020. Available at: arxiv.org/pdf/2007.10137.pdf (accessed: 21.10.2020).
31. Korolyuk V.S., Portenko N.I., Skorohod A.V., Turbin A.F. *Spravochnik po teorii veroyatnostej i matematicheskoj statistike* [Handbook of probability theory and mathematical statistics]. M.: Nauka. 1985. 640 p. (In Russ.).
32. Borzov A.B., Labunets L.V., Steshenko V.B. Noncanonical Spectral Model of Multidimensional Uniform Random Fields. *J. Comput. Syst. Sci. Int*. 2017. vol. 57. pp. 874–889.
33. Fedorov V.V. *Chislennye metody maksimina* [Numerical methods of Maximin]. M.: Nauka. 1979. 278 p. (In Russ.).
34. Klovskij D.D., Kontorovich V.Ya., Shirokov S.M. *Modeli nepreryvnyh kanalov svyazi na osnove stohasticheskih differencial'nyh uravnenij* [Models of continuous communication channels based on stochastic differential equations]. M.: Svyaz. 1984. 247 p. (In Russ.).
35. Kreher D.L. Combinatorial algorithms: Generation, enumeration, a. search. CRC press, Cop. 1999. 329 p.
36. Batenkov K.A., Batenkov A.A., Fokin A.B. [Methods for generating sets of telecommunication network States for various connectivity measures]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2020. Issue 3(19). pp. 644–673 (In Russ.).
37. Himmelblau D. *Prikladnoe nelinejnoe programmirovanie* [Applied nonlinear programming]. M.: Mir. 1975. 534 p. (In Russ.).
38. Yablonskij S.V. *Vvedenie v diskretnuy umatematiku :uchebnoe posobie dly astudentov vuzov, obuchayushchihsya po special'nosti "Prikladnaya matematika"* [Introduction to discrete mathematics: a textbook for University students studying in the specialty "Applied mathematics"]. M.: Vyssh. shk. 2006. 384 p. (In Russ.).
39. Batenkov K.A. [Numerical characteristics of the structures of communication networks]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2017. vol. 4(53). pp. 5–28 (In Russ.).
40. Batenkov K.A., Batenkov A.A. [Analysis and synthesis of communication network structures based on deterministic stability indicators]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2018. vol. 3(58). pp. 128–159 (In Russ.).
41. Nooka S.P. et al. Adaptive hierarchical classification networks. 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. pp. 3578–3583.