

## НОВЫЕ ПОДХОДЫ К ОЦЕНИВАНИЮ ПАРАМЕТРОВ В ГИББСОВСКИХ МОДЕЛЯХ РАСПОЗНАВАНИЯ ОБРАЗОВ

А.Н. Голодников, Т.П. Левашко, В.А. Пепеляев

Институт кибернетики им. В.М. Глушкова НАН Украины  
03187 Киев-187, ул. Академика В.М. Глушкова, 40

### Abstract

The paper considers approach to the image recognition problem that is based on random Gibbs fields. Estimation of parameters of Gibbs models. One of the possible approach to this problem is maximum entropy principle. According to this principle we should seek such distribution which is as non-informative as available prior information. the paper suggests another approach, which seeks such distributions, for which maximum and minimum values of probabilities of random object is attained.

### ВВЕДЕНИЕ

В задачах распознавания образов широкое распространение получил математический аппарат случайных полей [1]. В рамках этого подхода предполагается, что некий объект характеризуется двумя параметрами: наблюдаемым параметром - признаком объекта и скрытым параметром состоянием объекта. Тем или иным образом задана взаимосвязь между этими двумя параметрами, и на основании этого требуется построить стратегию распознавания, которая по известному наблюдению принимает разумное в определенном смысле решение о скрытом состоянии, в котором пребывает объект. Таким образом, задачи распознавания образов состоят в построении единой модели, объединяющей все параметры объекта, как наблюдаемые, так и скрытые. В качестве таких моделей широко используются гиббсовские случайные поля. Критическим в этих моделях является оценивание параметров гиббсовского распределения. В работе [1] для этой цели используется принцип максимума энтропии, согласно которому в качестве искомого гиббсовского распределения выбирается распределение вероятностей, которое было бы настолько неинформативным, насколько это возможно при условии наличия частичной информации. Такой подход является общепринятым в статистике, однако он является не единственно возможным. В настоящей статье предлагается альтернативный подход к выбору гиббсовского распределения. В основе его лежит очевидный факт, что в качестве «истинного» гиббсовского распределения может быть выбрано любое распределение, удовлетворяющее имеющейся априорной информации. В этой связи представляют особый интерес распределения при которых достигается максимальное или минимальное значения вероятностей наблюдаемого случайного объекта. В разделе 2 предлагаемый подход применяется для случая непараметрической формы гиббсовского распределения, а в разделе 3 - для случая однопараметрической формы.

## НЕПАРАМЕТРИЧЕСКАЯ ФОРМА РАСПРЕДЕЛЕНИЯ ГИББСА

Рассмотрим теперь следующую модель объектов [1]. Мы будем говорить, что объект описывается конечной совокупностью параметров, принимающих значения из конечного множества  $K$ . Пусть  $T$  - конечное множество индексов,  $t \in T$  - обозначение одного индекса, а  $k(t)$  - значение параметра с индексом  $t$ . Тогда полное описание наблюдаемого объекта - это функция  $k(t) : T \rightarrow K$ , которая для каждого параметра  $t$  указывает его значение  $k(t)$ . Обозначим  $K^t$  множество всех возможных описаний объекта, то есть множество всех возможных функций, определенных на  $T$  и принимающих значение из  $K$ . На множестве  $K^t$  всех возможных описаний задано распределение вероятностей  $p : K^t \rightarrow R$ , так, что  $p(\bar{k})$  - это вероятность описания  $\bar{k}$ . Это распределение вероятностей задано не явно, а посредством следующего определения.

На множестве  $T$  задана структура  $\mathfrak{S}$ , то есть совокупность подмножеств. Для каждого подмножества  $T'$  задано маргинальное распределение  $p_{T'}$  вероятностей частичных описаний  $\bar{k}_{T'} : T' \rightarrow K$ , так что  $p_{T'}(\bar{k}_{T'})$  - это маргинальная вероятность частичного описания  $\bar{k}_{T'}$ , то есть

$$p_{T'}(\bar{k}_{T'}) = \sum_{\bar{k} \in K(T', \bar{k}_{T'})} p(\bar{k}), \quad T' \in \mathfrak{S} \quad (1)$$

где  $K(T', \bar{k}_{T'})$  - множество тех описаний  $\bar{k} : T \rightarrow K$ , сужение которых на  $T'$  есть  $\bar{k}_{T'}$ . Система уравнений (1) не определяет вероятности  $p(\bar{k})$  единственным образом, так как при заданных маргинальных вероятностях  $p_{T'}(\bar{k}_{T'})$  она имеет множество решений относительно вероятностей  $p(\bar{k})$ . Обозначим это множество  $P$  и выберем из него такое распределение вероятностей  $p^*$  с наибольшей энтропией,

$$p^* = \arg \max_{p \in P} \left[ - \sum_{\bar{k} \in K^T} p(\bar{k}) \log p(\bar{k}) \right] \quad (2)$$

Формула (2) соответствует принципу максимума энтропии, согласно которому выбирается распределение вероятностей  $p^*$  (вероятности  $p(\bar{k})$ ), которое было бы настолько неинформативным, насколько это возможно при условии наличия частичной информации (1). Никакими иными преимуществами такой выбор по сравнению с любым другим выбором вероятностей  $p(\bar{k})$  из множества  $P$  не обладает. Любое распределение из множества  $P$  может быть с одинаковым основанием использовано вместо распределения  $p^*$ . Поэтому, варьируя распределения из множества  $P$ , можно построить диапазон  $(p_*(\bar{k}), p^*(\bar{k}))$  всех возможных значений вероятностей  $p(\bar{k})$  для фиксированного описания  $\bar{k}$ . Любое значение из диапазона  $(p_*(\bar{k}), p^*(\bar{k}))$  может быть использовано в качестве оценки вероятности  $p(\bar{k})$  реализации фиксированного описания  $\bar{k}$ .

Это обстоятельство существенно усложняет применение подхода, получившего наибольшее распространение в распознавании образов, состоящего в поиске такого

описания  $\bar{k} \in K^T$ , сужение которого на множества  $T'$  есть наблюдения  $\bar{k}_{T'}$  и которое обладает максимальной вероятностью. Одним из направлений преодоления отмеченной трудности является определение верхней и нижней границ диапазона  $(p_*(\bar{k}), p^*(\bar{k}))$  для каждого описания  $\bar{k} \in K^T$ , удовлетворяющего ограничениям (1).

Для решения этой задачи введем новые обозначения. Перенумеруем:

- все возможные описания  $\bar{k} \in K^T : \bar{k}_1, \bar{k}_2, \dots, \bar{k}_N$ ;
- все подмножества из заданной структуры  $\mathfrak{S} : T'_1, T'_2, \dots, T'_L$ ;
- все частичные описания, соответствующие фиксированному подмножеству  $T'_\ell : \bar{k}_{T'_\ell 1}, \bar{k}_{T'_\ell 2}, \dots, \bar{k}_{T'_\ell R_\ell}$ .

Введем матрицу  $B$ , с  $N$  столбцами и с  $\sum_{\ell=1}^L R_\ell$  строками. Строки матрицы  $B$  разбиты на  $L$  блоков;  $\ell$ -й блок соответствует  $\ell$ -му подмножеству  $T'_\ell$  и состоит из  $R_\ell$  строк, которые соответствуют частным описаниям  $\bar{k}_{T'_\ell 1}, \bar{k}_{T'_\ell 2}, \dots, \bar{k}_{T'_\ell R_\ell}$ . Столбец матрицы  $B$  будем обозначать индексом  $j$ , а строки - двумя индексами:  $\ell, r$ . Индекс  $\ell$  указывает на принадлежность строки к  $\ell$ -му блоку, а индекс  $r$  указывает порядковый номер строки внутри  $\ell$ -го блока. Столбец  $j$  соответствует  $j$ -му описанию. Определим элемент  $b_{j\ell r}$  матрицы  $B$  следующим образом:  $b_{j\ell r} = 1$ , если сужение  $j$ -го описания  $\bar{k}_j$  на множество  $T'_\ell$  есть  $\bar{k}_{T'_\ell r}$ ;  $b_{j\ell r} = 0$ , в противном случае. Обозначим  $d_{\ell r} = p_{T'_\ell}(\bar{k}_{T'_\ell r})$  маргинальную вероятность частичного описания  $\bar{k}_{T'_\ell r}$ . Пусть  $p(\bar{k}_j)$  обозначает вероятность реализации  $j$ -го описания  $\bar{k}_j$ .

Тогда, ограничения (1) на вероятности  $p(\bar{k}_1), p(\bar{k}_2), \dots, p(\bar{k}_N)$  реализации описаний  $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_N$  можно записать в следующем виде:

$$\sum_{j=1}^N b_{j\ell r} p(\bar{k}_j) = d_{\ell r}, \quad r = 1, \dots, R_\ell; \quad \ell = 1, \dots, L \quad (3)$$

Кроме того, должны выполняться следующие ограничения:

$$\sum_{j=1}^N p(\bar{k}_j) = 1, \quad (4)$$

и

$$p(\bar{k}_j) \geq 0, \quad j = 1, 2, \dots, N. \quad (5)$$

Задача поиска верхней границы диапазона  $(p_*(\bar{k}), p^*(\bar{k}))$  всех возможных значений вероятности  $p(\bar{k}_j)$  реализации фиксированного описания  $\bar{k}_j$  формулируется в виде следующей оптимизационной задачи:

$$p(\bar{k}_j) \rightarrow \max \quad (6)$$

при ограничениях (3)-(5).

Соответственно, задача поиска нижней границы диапазона  $(p_*(\bar{k}), p^*(\bar{k}))$  всех возможных значений вероятности  $p(\bar{k}_j)$  реализации фиксированного описания  $\bar{k}_j$  формулируется в виде следующей оптимизационной задачи:

$$p(\bar{k}_j) \rightarrow \min \quad (7)$$

при ограничениях (3)-(5).

Задачи (6), (3)-(5) и (7), (3)-(5) являются задачами линейного программирования с сильно разреженной матрицей линейных ограничений. Для отыскания нижней и верхней границ диапазонов возможных значений вероятностей реализации всех описаний необходимо решить  $N$  задач (6), (3)-(5) и (6), (3)-(5) при каждом фиксированном  $j$ .

#### ПАРАМЕТРИЧЕСКАЯ ФОРМА РАСПРЕДЕЛЕНИЯ ГИББСА

Рассмотрим теперь одно-параметрическую форму распределения Гиббса, предложенную в работах [2], [3].

$$p(\bar{k}) = \frac{e^{-\frac{1}{\theta} \sum_{\chi \in \Omega} V_{\chi}(\bar{k}_{\chi})}}{\sum_{\bar{k} \in K^T} e^{-\frac{1}{\theta} \sum_{\chi \in \Omega} V_{\chi}(\bar{k}_{\chi})}} \quad (8)$$

где  $\theta$  - параметр распределения Гиббса, который требуется оценить;  $\chi$  обозначает клику,  $\chi \in \Omega$ ;  $\Omega$  - множество клик;  $\bar{k}_{\chi}$  - частичное описание, соответствующее клике  $\chi$ ; функции  $V_{\chi}$  (гиббсовские потенциалы) удовлетворяют ограничению

$$|V_{\chi}(\bar{k}_{\chi})| < \infty \quad (9)$$

Для того, чтобы полностью определить распределение Гиббса на случайном поле, необходимо оценить параметр  $\theta$ . Статистические методы, которые широко используются для оценки параметров распределений, базируются на двух основных подходах: методах классической выборочной теории и процедурах Байесовского оценивания. Классический подход обеспечивает получение удовлетворительной оценки в тех случаях, когда имеется достаточное количество необходимой информации. Однако выборочные методы не пригодны для обработки малых выборок, с которыми приходится часто иметь дело в задачах образного мышления. Использование в таких случаях классических выборочных методов приводит к неточности оценивания и к снижению степени доверия к полученным результатам. Один из способов преодоления проблемы малых выборок состоит в использовании не только информации, относящейся к конкретному объекту исследования, но и в привлечении к анализу дополнительной информации из различных источников. К сожалению, классические методы не могут одновременно использовать текущую и дополнительную информацию при проведении статистических выводов.

В отличие от методов выборочной теории, Байесовский подход допускает естественное объединение данных из различных источников для получения оценок параметров распределения. При этом каждый источник трактуется как выборка из одной популяции. Если при классическом подходе параметры функции распределения считаются константами, которые нужно оценить, то при Байесовском подходе параметры рассматриваются как случайные величины, которые описываются некоторой априорной функцией распределения.

К основным элементам Байесовского подхода относятся выборочная модель  $f(x_i|\theta)$  - условная плотность распределения или условное распределение случайной величины  $X_i$ , которая зависит от параметра  $\theta$ ; априорная модель  $H(\theta)$  - априорное распределение параметра  $\theta$  с плотностью  $h(\theta)$ ; апостериорная модель  $G(\theta|x)$  - апостериорное распределение параметра  $\theta$ , при условии, что  $x$  - фиксировано, с плотностью  $g(\theta|x)$  и функция потерь  $L(\theta, \delta)$  для оценки  $\delta$ . В нашем случае выборочной моделью является распределение Гиббса (8), которое зависит от неизвестного параметра  $\theta$ .

Априорная плотность распределения  $h(\theta)$  аккумулирует в себе все, что известно о параметре  $\theta$  до момента получения эмпирических данных. Априорная информация, трансформированная в априорную функцию распределения, может иметь как объективный, так и субъективный характер. К объективной априорной информации можно отнести информацию, полученную, при исследовании аналогичных объектов в предыдущие моменты времени. К субъективной информации можно отнести экспертные оценки, которые базируются на собственном опыте специалистов.

На практике, как правило, имеющейся априорной информации не достаточно для точного определения априорной функции распределения. В этих условиях проблема выбора априорной функции распределения и вычисления Байесовских оценок становится нетривиальной.

В случаях, когда полностью отсутствует априорная информация относительно параметра  $\theta$ , или имеется незначительное количество такой информации, используют неинформативную априорную функцию распределения. В основе такого выбора лежит «принцип недостаточной причины», в соответствии с которым распределение на конечном множестве возможных значений параметра  $\theta$  должно быть равномерным, если отсутствуют какие-либо основания считать одно значение параметра более вероятным, чем другое. Этот принцип впервые был сформулирован Джеффрисом (Jeffreys H.) в работе [4].

Более распространенной является ситуация, когда в наличие имеется некоторая информация о неизвестном параметре  $\theta$ , но ее недостаточно для точного определения априорной функции распределения. В таких ситуациях в процессе трансформации этой неполной априорной информации в априорную функцию распределения всегда появляется опасность внесения дополнительной субъективной информации, которая не содержится в имеющейся априорной информации.

Наиболее распространенный метод использует априорную информацию для выбора конкретной функции распределения из некоторого подобранного класса функций распределения. В соответствии с этим методом, сначала выбирается класс функций распределения, а затем из него выбирается конкретная функция распределения путем подбора значений ее параметров таким образом, чтобы она соответствовала априорным значениям моментов или квантилей.

Для того, чтобы избежать внесения дополнительного субъективизма при выборе априорной функции распределения, в работах [5], [6], [7] было предложено воспользоваться принципом максимума энтропии. В соответствии с этим принципом априорная плотность распределения  $h(\theta)$  параметра  $\theta$  выбирается таким образом, чтобы максимизировать значение энтропии Шенона-Джейнеса (Shannon-Jaynes) [7]:

$$I(h) = - \int h(\theta) \ln(h(\theta)) d\theta \rightarrow \max \quad (10)$$

при ограничении:

$$S(h(\theta)) = 0, \quad (11)$$

$S(h(\theta))$  - функциональное ограничение на  $h(\theta)$ , которое вытекает из имеющейся априорной информации. Принцип максимума энтропии обеспечивает выбор априорной функции распределения, которая была бы настолько неинформативной, насколько это возможно при условии наличия частичной априорной информации.

При выборе априорной функции распределения исследователи очень часто руководствуются соображениями простоты математических выкладок, при помощи которых вычисляются Байесовские оценки. Такие свойства присущи сопряженным априорным функциям распределения. Сопряженной априорной плотностью распределения  $h(\theta)$  для данной выборочной плотности  $f(x|\theta)$  является такая плотность распределения, для которой как априорная плотность распределения,  $h(\theta)$ , так и апостериорная плотность распределения,  $g(\theta|x)$ , принадлежат одному и тому же классу распределений [8].

При использовании сопряженной функции распределения в качестве априорной возникает проблема подбора ее параметров. Очевидно, что в случаях, когда используется субъективная информация, эксперт не в состоянии дать удовлетворительную оценку непосредственно параметров априорной функции распределения. Однако он может достаточно адекватно оценить квантили априорной функции распределения. Во многих случаях эти квантили используются для выбора конкретной функции распределения из определенного класса сопряженных функций распределения. Такой выбор легко можно сделать, когда квантили и параметры функции распределения связаны между собой при помощи простой аналитической формулы [9].

Для того, чтобы избежать произвола при выборе априорной функции распределения, базируясь на имеющейся частичной априорной информации, целесообразно использовать  $\Gamma$ -минимаксный подход, который впервые был предложен Робинсом

в работе [10] и развит Бергером в работе [11]. В рамках  $\Gamma$  – минимаксного подхода берется во внимание тот факт, что при использовании Байесовских процедур на практике приходится иметь дело не с одной априорной функцией распределения, а с целым классом  $\Gamma$  таких функций. Причем каждая функция распределения из этого класса имеет одинаковые основания быть выбранной в качестве априорной.

Если для каждой функции распределения из класса  $\Gamma$  вычислить значения некоторой апостериорной функции (например, апостериорный риск, математическое ожидание и другие), то мы получим целый диапазон таких значений. Этот диапазон используется как мера робастности [12], [13], [14]. Узкий диапазон означает, что значения апостериорной функции существенно не изменяется при любом выборе априорной функции распределения из указанного класса, т.е. она является робастной по отношению к выбору априорной функции распределения. И, наоборот, когда этот диапазон - широкий, то значения апостериорной функции являются чувствительными к выбору априорной функции распределения, т.е. она является робастной. В последнем случае необходимо уделить особое внимание наименее благоприятной априорной функции распределения в классе  $\Gamma$  и использовать  $\Gamma$ -минимаксный подход для поиска робастной оценки.

В рамках  $\Gamma$ -минимаксного подхода выбирается оценка, которая минимизирует супремум целевого функционала по функциям распределения, которые принадлежат классу  $\Gamma$  [12]. Если в наличии имеется незначительный объем априорной информации, то класс  $\Gamma$  априорных функций распределения - очень большой и полученные при помощи этого метода результаты близки к результатам, полученным при минимаксном подходе. В другом граничном случае, когда в наличии имеется достаточная априорная информация, класс  $\Gamma$  может состоять только из одной априорной функции распределения и  $\Gamma$ -минимаксный подход совпадает с обычным Байесовским подходом.

Естественным выбором целевого функционала может быть:

- 1) Байесовский риск

$$r(H, \delta(x)) = \int_{\Theta} \int_X L(\theta, \delta(x)) f(x|\theta) dx dH(\theta)$$

- 2) апостериорный риск

$$\phi_H(\delta(x)) = \int_{\Theta} L(\theta, \delta(x)) dG(\theta|x) = \frac{\int_{\Theta} L(\theta, \delta(x)) f(x|\theta) dH(\theta)}{\int_{\Theta} f(x|\theta) dH(\theta)}$$

- 3) математическое ожидание параметра  $\theta$  при фиксированных выборочных данных  $x$

$$\int_{\Theta} \theta dG(\theta|x) = \frac{\int_{\Theta} \theta f(x|\theta) dH(\theta)}{\int_{\Theta} f(x|\theta) dH(\theta)}$$

С точки зрения статистика, априорная функция распределения, которая максимизирует Байесовский риск или апостериорный риск, является наименее предпочтительной функцией распределения в классе  $\Gamma$ , так как она соответствует наихудшей точности оценки. Для любой другой функции распределения из класса  $\Gamma$  точность соответствующей оценки может быть только лучшей.

Следуя консервативному подходу, естественно предположить, что на практике всегда реализуется наихудшая в  $\Gamma$  априорная функция распределения. Поэтому оценка, которая минимизирует Байесовский риск или апостериорный Байесовский риск при наихудшей априорной функции распределения является робастной.

В случае, когда используется квадратичная функция потерь, целевой функционал (3) совпадает с Байесовской оценкой  $\hat{\theta}_H$ . Если вычислить значения Байесовской оценки  $\hat{\theta}_H$  для любой априорной функции распределения  $H \in \Gamma$ , то мы получим диапазон возможных значений  $(\theta_*, \theta^*)$ .

Задача поиска робастной Байесовской оценки при фиксированных выборочных данных  $x$  сводится к стохастической минимаксной задаче, у которой внутренней является задача максимизации дробно-линейного функционала в пространстве функций распределения, принадлежащих классу  $\Gamma$ :

$$\min_{\delta} \sup_{H \in \Gamma} \phi_H(\delta, x) = \min_{\delta} \sup_{H \in \Gamma} \int_{\Theta} L(\theta, \delta) dG(\theta|x) = \min_{\delta} \sup_{H \in \Gamma} \frac{\int_{\Theta} L(\theta, \delta) f(x|\theta) dH(\theta)}{\int_{\Theta} f(x|\theta) dH(\theta)}$$

Для того, чтобы исследовать чувствительность Байесовских оценок к выбору априорной функции распределения из класса  $\Gamma$ , необходимо научиться отыскивать нижнюю и верхнюю границы диапазона возможных значений целевого функционала. В соответствии с Байесовским подходом, если  $H(\theta)$  является истинной априорной функцией распределения, то качество Байесовской точечной оценки  $\hat{\theta}_H(x)$  измеряется в терминах Байесовского риска  $r(H, \hat{\theta}_H(x))$  или апостериорного риска  $\phi(H, \hat{\theta}_H(x))$ . Рассмотрим теперь случай, когда вместо одной априорной функции распределения  $H(\theta)$  в наличии имеется класс  $\Gamma$  таких функций распределения. В этом случае значения  $r(H, \hat{\theta}_H(x))$  или  $\phi(H, \hat{\theta}_H(x))$  уже не могут использоваться для характеристики качества Байесовской точечной оценки  $\hat{\theta}_H(x)$ . Для этой цели более адекватным является диапазон  $(r_*(\hat{\theta}_H(x)), r^*(\hat{\theta}_H(x)))$  возможных значений Байесовского риска или диапазон  $(\phi_*(\hat{\theta}_H(x)), \phi^*(\hat{\theta}_H(x)))$  возможных значений апостериорного риска, которые вычисляются для всех априорных функций распределения из класса  $\Gamma$ .

Задача вычисления нижней  $r_*(\hat{\theta}_H(x))$  или верхней  $r^*(\hat{\theta}_H(x))$  границ диапазона возможных значений Байесовского риска для точечной оценки  $\hat{\theta}_H(x)$  сводится к задаче минимизации или максимизации линейного целевого функционала по функциям распределения из класса  $\Gamma$ :

$$r_*(\hat{\theta}_H(x)) = \min_{G \in \Gamma} \int_{\Theta} \int_X L(\theta, \hat{\theta}_H(x)) f(x|\theta) dx dG(\theta)$$



и

$$r^*(\hat{\theta}_H(x)) = \max_{G \in \Gamma} \int_{\Theta} \int_X L(\theta, \hat{\theta}_H(x)) f(x|\theta) dx dG(\theta)$$

Задача вычисления нижней  $\phi_*(\hat{\theta}_H(x))$  или верхней  $\phi^*(\hat{\theta}_H(x))$  границ диапазона возможных значений апостериорного риска для точечной оценки  $\hat{\theta}_H(x)$  сводится к задаче минимизации или максимизации дробно-линейного целевого функционала по функциям распределения из класса  $\Gamma$ :

$$\phi_*(\hat{\theta}_H(x)) = \min_{G \in \Gamma} \frac{\int_{\Theta} L(\theta, \hat{\theta}_H(x)) f(x|\theta) dG(\theta)}{\int_{\Theta} f(x|\theta) dG(\theta)}$$

$$\phi^*(\hat{\theta}_H(x)) = \max_{G \in \Gamma} \frac{\int_{\Theta} L(\theta, \hat{\theta}_H(x)) f(x|\theta) dG(\theta)}{\int_{\Theta} f(x|\theta) dG(\theta)}$$

При оценивании параметра гиббсовского распределения важно знать, насколько широким является диапазон  $(\theta_*, \theta^*)$  возможных значений апостериорного математического ожидания параметра  $\theta$  при фиксированных выборочных данных  $x$ , полученных для любой априорной функции распределения из класса  $\Gamma$ . Задача вычисления нижней  $\theta_*$  или верхней  $\theta^*$  границ сводится к следующим задачам минимизации или максимизации дробно-линейного целевого функционала по функциям распределения из класса  $\Gamma$ :

$$\inf_{H \in \Gamma} \frac{\int_{\Theta} \theta f(x|\theta) dH(\theta)}{\int_{\Theta} f(x|\theta) dH(\theta)}$$

или

$$\sup_{H \in \Gamma} \frac{\int_{\Theta} \theta f(x|\theta) dH(\theta)}{\int_{\Theta} f(x|\theta) dH(\theta)}$$

Будем рассматривать класс функций распределения, который можно описать при помощи следующих линейных ограничений:

$$\int_{\Theta} dH(\theta) = 1, \quad (12)$$

$$\int_{\Theta} f_i(\theta) dH(\theta) \leq d_i, \quad i = 1, 2, \dots, m_1, \quad \int_{\Theta} g_j(\theta) dH(\theta) = \mu_j, \quad j = 1, 2, \dots, m_2 \quad (13)$$

где  $f_i(\theta)$ ,  $i = 1, 2, \dots, m_1$ ,  $g_j(\theta)$ ,  $j = 1, 2, \dots, m_2$ , могут быть разрывными.

Если параметрическое пространство  $\Theta = [a, b]$  - одномерное, и функции  $g_j(\theta) = \theta^j$ ,  $j = 1, 2, \dots, m_2$ , то этот класс состоит из всех функций распределения, которые имеют заданные ограничения на моменты. Если  $f_i(\theta)$ ,  $i = 1, 2, \dots, m_1$  - индикативные функции для некоторых интервалов параметрического пространства  $\Theta = [a, b]$ , то этот класс состоит из всех функций распределения, которые имеют заданные ограничения на квантили.

Таким образом, сформулированные выше задачи оценивания сводятся к задачам оптимизации линейного и дробно-линейного функционалов в классе функций распределения, удовлетворяющих линейным ограничениям.

В общем виде задача оптимизации линейного функционала при линейных ограничениях на функции распределения формулируется в виде:

$$\varphi(H) = \int_X f_0(x) dH(x) \rightarrow \inf, \quad (14)$$

при ограничениях

$$\int_X dH(x) = 1, \quad (15)$$

$$\psi_i(H) = \int_X f_i(x) dH(x) \leq a_i, \quad i = 1, 2, \dots, m, \quad (16)$$

где  $X$  - компактное множество в пространстве  $R^n$ .

Задача оптимизации дробно-линейного функционала при линейных ограничениях на функции распределения формулируется в виде:

$$F(H) = \frac{\int_X g_1(x) dH(x)}{\int_X g_2(x) dH(x)} \rightarrow \inf \quad (17)$$

при ограничениях (15)-(16).

Пусть  $K_m(X)$  обозначает множество функций распределения, которые удовлетворяют ограничениям (15)-(16). Предположим, что функции  $f_\nu(x)$ ,  $\nu = 0, 1, \dots, m$ , имеют конечное число поверхностей разрыва  $A_1, A_2, \dots, A_r$ , т.е., если  $A$  - поверхность разрыва функции  $f_\nu(x)$ , то  $\forall x' \in A$  и  $\forall \varepsilon > 0 \exists \delta > 0$  такие, что  $\forall x \in U_\delta(x') \cap A$  выполняется неравенство

$$|f_\nu(x') - f_\nu(x)| < \varepsilon,$$

где  $U_\delta(x') = \{x : \|x - x'\| < \delta\}$ .

Если  $X \subset R^1$ , то поверхность разрыва совпадает с точкой разрыва.

#### АЛГОРИТМ РЕШЕНИЯ ЛИНЕЙНОЙ ЗАДАЧИ

Идея алгоритма состоит в сведении начальной задачи бесконечномерного программирования к последовательности конечномерных задач. На каждой итерации будем искать решение задачи (15)-(16) на множестве ступенчатых функций распределения, точки роста которых принадлежат конечному множеству. Количество точек этого множества возрастает от итерации.

На  $s$ -ой итерации строим разбиение  $R_s^i = (x_{1,i}^s, x_{2,i}^s, \dots, x_{i_s,i}^s)$  множеств  $A_i$ ,  $i = 0, 1, \dots, r$  так что  $A_i \subset \cup_{j=1}^{i_s} U_{\lambda_s}(x_{j,i}^s)$ ,  $R_s^i \subset R_{s+1}^i$ ,  $i = 0, 1, \dots, r$ ;  $\lambda_s \rightarrow 0$  при  $s \rightarrow \infty$ , где  $A_0 = X \setminus (\cup_{i=1}^r A_i)$ ;  $U_{\lambda_s}(x_{j,i}^s) = \{x : \|x - x_{j,i}^s\| < \lambda_s\}$ .

Пусть  $R_s = \cup_{i=0}^r R_s^i$ ,  $R_s = (x_1^s, x_2^s, \dots, x_{n_s}^s)$ , где  $n_s = (r+1) \cdot \ell_s$ ,  $x_{(j+1)\ell_s+i}^s = x_{i,j}^s$ ,  $i = 1, 2, \dots, \ell_s$ ;  $j = 0, 1, \dots, r$ . Тогда решаем следующую задачу линейного программирования:

$$\sum_{j=1}^{n_s} f_0(x_j) p_j \rightarrow \min \quad (18)$$

при ограничениях

$$\sum_{j=1}^{n_s} p_j = 1, \quad (19)$$

$$\sum_{j=1}^{n_s} f_i(x_j) p_j \leq a_i, \quad i = 1, 2, \dots, m, \quad (20)$$

$$p_j \geq 0, \quad x_j \in R_s, \quad j = 1, 2, \dots, n_s.$$

Пусть  $p^s = (p_1^s, p_2^s, \dots, p_{n_s}^s)$  - решение задачи (18)-(20). Поскольку число базисных переменных в задаче (18)-(20) равно  $m+1$ , то не более, чем  $m+1$  компонента вектора  $p^s$  может быть ненулевой. Без потери общности, можно предположить, что первые  $m+1$  компоненты  $p_1^s, p_2^s, \dots, p_{m+1}^s$  - ненулевые. Следовательно, решение задачи (18)-(20) можно представить в виде:

$$H_s(x) = (x_1^s, \dots, x_{m+1}^s, p_1^s, \dots, p_{m+1}^s) \quad (21)$$

где базисная переменная  $p_i^s$  соответствует точке  $x_i^s$  и соответствующему базисному столбцу

$$(1, f_1(x_i^s), \dots, f_m(x_i^s))'.$$

С другой стороны, запись (21) можно рассматривать как обозначение функции распределения, где ступенчатая функция  $H_s(x)$  имеет точку роста  $p_i^s$  в точке  $x_i^s$ ,  $i = 1, 2, \dots, m$ . Решение  $H_s(x)$ , найденное на  $s$ -ой итерации, будет использоваться как начальные базисные переменные при решении задачи (15)-(17) на  $s+1$ -ой итерации.

В работе [14] была исследована сходимость этого алгоритма и были доказаны следующие теоремы.

**Теорема 1.** *Предположим, что*

1. *Функции  $f_\nu(x)$  - полунепрерывны снизу и ограничены на множестве  $X$ ,  $\nu = 0, 1, \dots, m$ .*
2.  *$X$  - компактное множество в пространстве  $\mathbb{R}^n$ .*
3.  *$\exists \tilde{H}(x)$  такая, что выполняются такие ограничения:*

$$\int_X d\tilde{H}(x) = 1,$$

$$\psi_i(\tilde{H}(x)) = \int_X f_i(x) d\tilde{H}(x) < a_i,$$

$$i = 1, 2, \dots, m.$$

Тогда предел любой сходящейся подпоследовательности  $\{H_{s_k}\}$  последовательно-сти  $\{H_s\}$  принадлежит множеству  $X^*$ , где

$$X^* = \{H^*(x) : \varphi(H^*) = \min_{H(x) \in K_m(X)} \varphi(H)\}$$

и

$$\lim_{s \rightarrow \infty} \varphi(H_s) = \varphi(H^*)$$

Эта теорема была доказана в работе [15].

**Теорема 2.** *Предположим, что*

1. *Функции  $f_\nu(x)$  - полунепрерывны сверху и ограничены на множестве  $X$ ,  $\nu = 0, 1, \dots, m$ .*
2.  *$X$  - компактное множество в пространстве  $\mathbb{R}^n$ .*
3.  *$\exists \tilde{H}(x)$  такая, что выполняются такие ограничения:*

$$\int_X d\tilde{H}(x) = 1,$$

$$\psi_i(\tilde{H}(x)) = \int_X f_i(x) d\tilde{H}(x) < a_i,$$

$$i = 1, 2, \dots, m.$$

Тогда

$$\lim_{s \rightarrow \infty} \varphi(H_s) = \inf_{H(x) \in K_m(X)} \varphi(H).$$

Эта теорема была доказана в работе [15].

#### АЛГОРИТМ РЕШЕНИЯ ДРОБНО-ЛИНЕЙНОЙ ЗАДАЧИ

Рассмотрим следующую задачу оптимизации дробно-линейного функционала:

$$F(H) = \frac{\int_X g_1(x) dH(x)}{\int_X g_2(x) dH(x)} \rightarrow \inf_{H(x) \in K_m(x)}, \quad (22)$$

где  $K_m(X)$  - множество функций распределения, удовлетворяющих ограничениям (15)-(16).

Обозначим

$$J_1(H) = \int_X g_1(x) dH(x)$$

и

$$J_2(H) = \int_X g_2(x) dH(x)$$

Будем предполагать, что  $J_2(H) \geq \gamma > 0$  для всех функций распределения  $H(x) \in K_m(X)$ , где  $\gamma$  - фиксированное положительное число. Справедлива следующая теорема [15].

**Теорема 3.** *Задача (22) эквивалентна задаче поиска числа  $t_*$  такого, что выполняется следующее равенство:*

$$\inf_{H(x) \in K_m(x)} [J_1(H) - t_* J_2(H)] = 0 \quad (23)$$

Рассмотрим теперь сам алгоритм. На  $s$ -ой итерации строим разбиения  $R_s^i = (x_{1,i}^s, x_{2,i}^s, \dots, x_{i_s,i}^s)$  множеств  $A_i$ ,  $i = 0, 1, \dots, q$ , такие, что  $A_i \in \cup_{j=1}^{\ell_s} U_{r_s}(x_{j,i}^s)$ ,  $R_s^i \subset R_i^{s+1}$ ,  $i = 0, 1, \dots, q$ ;  $r_s \rightarrow 0$  при  $s \rightarrow \infty$ , где  $A_0 = X \setminus (\cup_{i=1}^q A_i)$ ;  $U_{r_s}(x_{j,i}^s) = \{x : \|x - x_{j,i}^s\| < r_s\}$ .

Пусть  $R_s = \cup_{i=0}^q R_s^i$ ,  $R_s = (x_1^s, x_2^s, \dots, x_{n_s}^s)$ , где  $n_s = (q+1) \cdot \ell_s$ ,  $x_{(j+1) \cdot \ell_s + i}^s = x_{i,j}^s$ ,  $i = 1, 2, \dots, \ell_s$ ;  $j = 0, 1, \dots, q$ .

Решаем следующую задачу линейного программирования:

$$\sum_{j=1}^{n_s} [g_1(x_j) - t_{s-1} g_2(x_j)] \rightarrow \min \quad (24)$$

при ограничениях

$$\sum_{j=1}^{n_s} p_j = 1, \quad (25)$$

$$\sum_{j=1}^{n_s} f_i(x_j) p_j \leq a_i, \quad i = 1, 2, \dots, m, \quad (26)$$

$$p_j \geq 0, \quad x_j \in R_s, \quad j = 1, 2, \dots, n_s.$$

Пусть функция распределения  $H_s(x)$  - решение этой задачи. Тогда вычисляем

$$t_s = \frac{J_1(H_s)}{J_2(H_s)} \quad (27)$$

и переходим к  $s+1$ -ой итерации.

Начальная аппроксимация вычисляется по формуле:

$$t_0 = \frac{J_1(H_0)}{J_2(H_0)},$$

где  $H_0$  - произвольная функция распределения из класса  $K_m(X)$ , которая имеет не более чем  $m+1$  точек роста.

В работе [15] была исследована сходимость этого алгоритма и была доказана следующая теорема.

**Теорема 4.** *Предположим, что*

1. Функции  $g_1(x)$ ,  $f_i(x)$ ,  $i = 1, \dots, m$  полунепрерывны снизу и  $g_2(x)$  - полунепрерывна сверху и ограничены на множестве  $X$ .
2.  $X$  - компактное множество в пространстве  $\mathbb{R}^n$ .
3.  $\exists \tilde{H}(x)$  такая, что выполняются соотношения:

$$\int_X d\tilde{H}(x) = 1,$$

$$\psi_i(\tilde{H}) = \int_X f_i(x) d\tilde{H}(x) < a_i,$$

$$i = 1, 2, \dots, m.$$

Тогда предел любой сходящейся подпоследовательности  $\{H_{s_s}\}$  последовательности  $\{H_s\}$  принадлежит множеству  $X^*$ , где

$$X^* = \{H^*(x) : F(H^*) = \min_{H(x) \in K_m(X)} F(H)\}$$

и

$$\lim_{s \rightarrow \infty} t_s = \min_{H(x) \in K_m(X)} \frac{\int_X g_1(x) dH(x)}{\int_X g_2(x) dH(x)}$$

#### ЗАКЛЮЧЕНИЕ

В статье был предложен новый подход к оцениванию параметров в гиббсовских моделях распознавания образов. Этот подход представляет собой альтернативу одному из известных способов нахождения гиббсовского распределения, основанного на принципе максимума энтропии, согласно которому в качестве искомого гиббсовского распределения выбирается распределение вероятностей, которое было настолько неинформативным, насколько это возможно при условии наличия частичной информации.

В основе предложенного подхода лежит очевидный факт, что в качестве «истинного» гиббсовского распределения может быть выбрано любое распределение, удовлетворяющее имеющейся априорной информации. В этой связи представляют особый интерес распределения, при которых достигается максимальное или минимальное значения вероятностей наблюдаемого случайного объекта. Предложены алгоритмы поиска максимального и минимального значений вероятностей наблюдаемого случайного объекта. Возможность определения нижней и верхней границ для вероятностей наблюдаемого случайного объекта имеет важное значение, поскольку расстояние между этими границами можно использовать как меру ошибки в задачах распознавания.

## СПИСОК ЛИТЕРАТУРЫ

1. *Schlesinger M.I. and Vaclav Hlavac*, Ten Lectures on Statistical and Structural Pattern Recognition. Computation Imaging and Vision. Kluwer Academic Publishers - Dordrecht. Boston. London. 2002, 520 p.
2. *Гимельфарб Г.Л., Залесный А.В.* Байесовская сегментация изображений с использованием моделей марковских полей, задаваемых гиббсовскими распределениями, Автоматизированные системы обработки изображений. Тез. докладов 3 Всес. конф. Ленинград 1989 г. с. 22-23.
3. *Залесный А.В.* Алгоритмы цифровой обработки изображений, описываемых марковскими случайными полями. Диссертация на соискание ученой степени канд. тех. наук, Киев, Институт кибернетики им. В.М Глушкова АН УССР, 1991.
4. *H.Jeffreys*, Theory of Probability, 1961, (3rd ed.), Clarendon Press, Oxford.
5. *E.T. Jaynes* «Prior Probabilities», IEEE Transactions on System Science and Cybernetics, Vol. SSC – 4, 1968, pp. 227-241.
6. *J.J. Deeley, M.S. Tierney, and W.J. Zimmer*, «On the Usefulness of the Maximum Entropy Principle in the Bayesian Estimation of Reliability», IEEE Transactions on Reliability, vol. R-19, 1970, pp. 110 – 115.
7. *V.P.Savchuk, H.F. Martz*, «Bayes Reliability Estimation Using Multiple Sources of Prior Information: Binomial Sampling», IEEE Transactions on Reliability, vol.43, 1994, pp. 138-144.
8. *H.F.Martz, R.A. Waller*, Bayesian Reliability Analysis. Krieger publishing company, Malabar, Florida, 1991.
9. *Mosleh, G.Apostolakis*, «Some Properties of Distributions Useful in the Study of Rare Events», IEEE Transactions on Reliability, vol. R-31, 1982, pp. 87-94.
10. *H. Robins*, «Asymptotically Sub-Minimax Solutions to Compound Statistical Decision Problems», In: Proc. Second Berkeley Symp. Math Stat. Probab., 1951, 1, University of California Press, Berkley.
11. *J.O. Berger*, «The Robust Bayesian Viewpoint (with discussion)», Robustness of Bayesian Analysis, (J.Kadane, ed.). Amsterdam: North Holl
12. *B. Vidakovic*, «Gamma-Minimax: A Paradigm for Conservative Robust Bayesians», In: Robustness of Bayesian Inference. Editors Rios and Ruggeri, Springer-Verlag, Lecture Notes in Statistics 152, 2000, pp. 241-259.
13. *F. Ruggeri, S. Sivaganesan*, «On a Global Sensitivity Measure for Bayesian Inference», Sankhya: The Indian Journal of Statistics, Vol. 62, 2000, Series A, Pt. 1, pp. 110-127.
14. *C. Carota, F. Ruggeri*, «Robust Bayesian analysis given priors on partition sets», Test, Vol. 3, 1994, No.2, pp. 73-86.
15. *A. Golodnikov, P. Knopov, P.Pardalos, S. Uryasev* «Optimization in the Space of Distribution Functions and Applications in the Bayes Analysis» in «Probabilistic Constrained optimization: Methodology and Applications», Kluwer Academic Publishers, 2000, p. 102-131.