УДК 519.72

В.В. Савченко

Нижегородский государственный лингвистический университет

А.В. Савченко

Нижегородский государственный технический университет

Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов

Ставится и решается задача распознавания случайных дискретных объектов по конечным выборкам наблюдений за их состояниями на основе принципа минимального информационного рассогласования распределений в метрике Кульбака-Лейблера. Предложен асимптотически оптимальный алгоритм с обнаружением и отклонением сомнительных решений. Рассмотрен пример его применения в задаче лингвистического анализа текстов.

Случайная величина, распределение вероятностей, выборка наблюдений, задача статистической классификации, распознавание образов, критерий минимума информационного рассогласования

Введение. Во многих областях прикладных исследований, таких как задачи технической и медицинской диагностики, лингвистического анализа и другие, где вся доступная информация заключена в конечных выборках наблюдений**,** возникает необходимость по возможности точно ответить на вопрос: насколько существенно отличаются свойства анализируемых объектов между собой? В рамках универсального статистического подхода указанная задача обычно сводится к задаче их статистической классификации. Принцип минимума информационного рассогласования (МИР) является одним из наиболее эффективных инструментов для ее решения. Это было показано, в частности, в работах [1,2] в предположении о гауссовом распределении результатов наблюдений. Ниже дается обобщение принципа МИР на задачи с произвольными распределениями вероятностей для дискретного стационарного случая.

1. Задача статистической классификации. Пусть $\mathbf{P}_r = \left\{p_{r,i} = P\left(X = a_i\right), i = \overline{1,N}\right\}$, $r = \overline{1,R}$, -- набор альтернативных распределений вероятности (*P*) некоторой случайной дискретной величины *X*, определенной на множестве различных состояний объекта анали-

за $\{a_i\}$ при очевидных ограничениях $\sum_{i=1}^{N} p_{r,i} = 1 \forall\ r \le R$ (условие нормировки) и $p_{r,i} \neq 0 \forall r \le R, i \le N$ (условие регулярности). И пусть $X_1, X_2, ..., X_n$ -- *n*-выборка независимых наблюдений (*n*>*N*), распределенных по закону $\mathbf{P} \subset \{\mathbf{P}_r\}$ из множества заданных альтернатив, какому именно – неизвестно. Задача распознавания объекта X сводится в таком случае к проверке *R* статистических гипотез о его законе распределения:

$$W_r: \qquad \mathbf{P} = \mathbf{P}_r, \quad r = \overline{1,R}. \tag{1}$$

Рассмотрим сначала эту задачу в наиболее простом варианте дихотомии ($R = 2$): проверяется гипотеза

$$W_1: \ \mathbf{P} = \mathbf{P}_1$$

против альтернативы

$$W_2: \ \mathbf{P} = \mathbf{P}_2.$$

Воспользуемся для ее решения классическим критерием максимального правдоподобия. При этом заключение в пользу гипотезы W_1 будем принимать по выборке $\mathbf{X} = \{X_j\}$ из условия

$$W_1(\mathbf{X}): \qquad \lambda(\mathbf{X}) = \frac{p(\mathbf{X}/W_1)}{p(\mathbf{X}/W_2)} > 1, \tag{2}$$

где $p(\mathbf{X}/W_r)$ -- функция правдоподобия *r* -ой гипотезы W_r, *r* = 1;2.

2. Синтез оптимального алгоритма. Учитывая независимость наблюдений $\{X_j\}$ в совокупности, с использованием аппарата абстрактных дельта-функций Дирака $\delta(\cdot)$ получим систему очевидных равенств

$$p(\mathbf{X}/W_r) = \prod_{j=1}^{n} \sum_{i=1}^{N} p_{r,i} \delta(X_j - a_i), \ r = 1;2.$$

А подставляя их в выражение (2), будем иметь

$$W_1(\mathbf{X}):\qquad \lambda(\mathbf{X})=\frac{\prod_{j=1}^{n}\sum_{i=1}^{N}p_{1,i}\delta(X_j-a_i)}{\prod_{j=1}^{n}\sum_{i=1}^{N}p_{2,i}\delta(X_j-a_i)}=\prod_{j=1}^{n}\sum_{i=1}^{N}p_{1,i}\Big/p_{2,i}\,I(X_j-a_i)>1, \tag{3}$$

где $I(X_j-a_i)$- функция единичного скачка при равенстве $X_j=a_i$.

После несложных преобразований из (3) окончательно получаем

$$W_1(\mathbf{X}):\qquad \lambda(\mathbf{X})=\prod_{j=1}^{n}\sum_{i=1}^{N}p_{1,i}\Big/p_{2,i}\,I(X_j-a_i)=\prod_{i=1}^{N}\left(p_{1,i}\Big/p_{2,i}\right)^{nw_i}>1,$$

или, что эквивалентно,

$$W_1(\mathbf{X}):\qquad \prod_{i=1}^{N}\left(p_{1,i}\Big/p_{2,i}\right)^{w_i}>1. \tag{4}$$

Здесь введено обозначение: $w_i=\sum_{j=1}^{n}I(X_j-a_i)\Big/n$ -- относительная частота появления *i*-го состояния объекта $a_i, i=\overline{1,N}$, в серии из *n* наблюдений.

Определение 1. Последовательность (или ряд) $\hat{\mathbf{P}}=\{w_i\}$ определяет эмпирический закон распределения состояний рассматриваемого объекта по результатам *n* независимых наблюдений.

Утверждение 1. В условиях введенных выше ограничений на объекты статистической классификации из (1) критерий максимального правдоподобия (МП) эквивалентен критерию минимума величины информационного рассогласования (в метрике Кульбака-Лейблера) эмпирического распределения $\hat{\mathbf{P}}$ относительно рассматриваемых гипотез $\mathbf{P}_1$ и $\mathbf{P}_2$.

Доказательство. В соответствии с определением величины информационного рассогласования распределений по Кульбаку-Лейблеру [3] можно записать $\rho(\hat{\mathbf{P}}/\mathbf{P}_r)=\sum_{i=1}^{N}w_i\log(w_i/p_{r,i})$, r = 1;2. Тогда, следуя критерию МИР [1], получим решающее правило

$$W_1(\mathbf{X}):\qquad \lambda(\mathbf{X})=\rho(\hat{\mathbf{P}}/\mathbf{P}_2)-\rho(\hat{\mathbf{P}}/\mathbf{P}_1)>0, \tag{5}$$

которое сводится к проверке условия

$$W_1(\mathbf{X}):\qquad \sum_{i=1}^{N}w_i\log(w_i/p_{2,i})>\sum_{i=1}^{N}w_i\log(w_i/p_{1,i}).$$

Путем простых его преобразований приходим к выражению для решающей статистики вида

$$\lambda(\mathbf{X}) = \log \prod_{i=1}^{n} \left(p_{1,i} \Big/ w_i \right)^{w_i} \Big/ \left(p_{2,i} \Big/ w_i \right)^{w_i} .$$

Нетрудно убедиться, что полученный результат ничем не отличается от критерия МП (4), что и требовалось доказать.

Распространяя по индукции решающее правило (5) на произвольное число альтернатив $R \geq 2$ в задаче распознавания (1), приходим к следующему выводу.

Утверждение 2. Оптимальный алгоритм распознавания *R* случайных дискретных объектов общего вида реализует принцип минимума информационного рассогласования эмпирического распределения $\hat{\mathbf{P}}$ относительно всех рассматриваемых альтернатив:

$$W_\nu(\mathbf{X}): \qquad \sum_{i=1}^{N} w_i \log\left(w_i \Big/ p_{r,i} \right) \Big|_{r=\nu} = \min_r, r = \overline{1,R}, \nu \leq R, \tag{6}$$

Это стандартная формулировка критерия МИР в многоальтернативном варианте [1…3]. В данном случае она охватывает все мыслимое многообразие дискретных распределений $\{\mathbf{P}_r\}$ при несущественных для практики ограничениях на их нормировку и регулярность.

3. Анализ эффективности. Эффективность синтезированного алгоритма (6) может быть охарактеризована набором условных вероятностей перепутывания каждого ν-го объекта анализа с каждым *r*-м объектом:

$$\alpha_{\nu,r} = P(W_r(X) \mid W_\nu) = P\{(\rho(\hat{\mathbf{P}} / \mathbf{P}_\nu) > \rho(\hat{\mathbf{P}} / \mathbf{P}_r) \mid W_\nu\}, \quad \nu \neq r = \overline{1;R}\,. \tag{7}$$

К сожалению, строгий анализ выражения (7) в общем случае наталкивается на непреодолимые трудности вычислительного характера. Поэтому упростим задачу, сводя ее к асимптотическому случаю $n \to \infty$. Рассмотрим отдельно каждый из элементов под знаком вероятности $P\{\cdot\}$ в (7). Известно [3], что при справедливости гипотезы W_ν статистика $\rho(\hat{\mathbf{P}} / \mathbf{P}_\nu) = \sum_{i=1}^{N} w_i \log\left(w_i \Big/ p_{\nu,i} \right) = (2n)^{-1} \chi_{N-1}^{\,2}$ с точностью до константы $(2n)^{-1}$ асимптотически распределена как $\chi_{N-1}^{\,2}$-случайная величина Пирсона с (*N*-1)-ой степенью свободы. С другой стороны, при том же условии выполняется асимптотическое равенство $\rho(\hat{\mathbf{P}} / \mathbf{P}_r)|_{W_\nu} \xrightarrow{\text{п.н.}} \rho(\mathbf{P}_\nu / \mathbf{P}_r)$, где *п.н.* – это сходимость с вероятностью 1 или «почти наверное» к величине информационного рассогласования

$$\rho(\mathbf{P}_\nu / \mathbf{P}_r) = \sum_{i=1}^{N} p_{\nu,i} \log(\; p_{\nu,i} \big/ p_{r,i} \;)$$

между ν-м и *r*-м объектами анализа. На основании сказанного из (7) в асимптотике будем иметь

$$\alpha_{\nu,r} = P(\chi_{N-1}{}^2 > 2n\rho(\mathbf{P}_\nu / \mathbf{P}_r)), \quad \nu \neq r = \overline{1;R}. \tag{8}$$

Чем больше «расстояние» между двумя объектами в теоретико-информационном смысле, тем сложнее их перепутать при распознавании по имеющейся выборке наблюдений. При увеличении объёма выборки *n* до бесконечности вероятность ошибочных решений в соответствии с (8) в пределе уменьшается до нуля, что означает оптимальное решение поставленной задачи (1) в асимптотике. Сделанный вывод имеет непосредственное отношение к задаче распознавания образов как обобщению задачи статистической классификации на случай априорной неопределенности в отношении анализируемых распределений [4].

4. Задача распознавания образов. Наиболее общая формулировка большинства задач статистической обработки информации может быть дана в терминах теории распознавания образов: требуется отнести предъявляемый объект анализа **X** (в нашем случае – выборку объема *n*) к одному из $R > 1$ классов, строго говоря, заранее точно не определенных. Каждый класс характеризуется тем, что принадлежащие ему объекты обладают некоей общностью или сходством. То общее, что объединяет объекты в класс, и называют *образом*. Иными словами, каждый конкретный объект наблюдения мы задаем в виде его вполне определенного образа, т.е. некоторого набора устойчивых признаков $\mathbf{P}_r$, $r = \overline{1,R}$. В таком случае решение рассматриваемой задачи (1) сводится к установлению отношения эквивалентности между соответствующим набором признаков объекта наблюдения X и одного из *R* образов в базе априорных данных.

Проблема состоит в том, что каждому конкретному образу обычно присуща известная вариативность, т.е. изменчивость его признаков от одного образца наблюдения к другому, которая носит к тому же неопределенный, случайный характер. Обычно решение данной проблемы связывают со статистическим подходом, когда в роли каждого образа $\mathbf{P}_r$ выступает соответствующий закон распределения повторной выборки наблюдений из некоторой гипотетической генеральной совокупности. Задача установления отношения эквивалентности общего вида (1) переходит в таком случае в задачу проверки статистических гипотез о неизвестном законе распределения. Но здесь возникает новое препятствие, а именно: проблема встречных гипотез в отношении вида каждого из R альтернативных распределений. Радикальное средство для ее преодоления – это восста-

новление неизвестных законов $\mathbf{P}_r$, $r=\overline{1,R}$, в процессе предварительного обучения. Указанное обучение осуществляется, например, по повторным выборкам $\mathbf{X}_r=\{X_j^{(r)}, j=\overline{1,n_r}\}$ конечных объемов $n_r, r=\overline{1,R}$, принадлежность которых к тому или иному классу (образу) заранее точно известна.

Утверждение 3. При априорной неопределенности в отношении гипотетических распределений в задаче (1) асимптотически оптимальный (состоятельный) критерий многоальтернативного распознавания случайных дискретных объектов сводится к критерию МИР между эмпирическими законами распределения наблюдаемого объекта *X*, с одной стороны, и каждой из *R* его альтернатив, с другой, т.е.

$$W_\nu(\mathbf{X}):\quad \rho\left(\hat{\mathbf{P}}/\hat{\mathbf{P}}_r\right)_{r=\nu}=\sum_{i=1}^{N} w_i \log\left(w_i\big/\theta_{r,i}\right)\Big|_{r=\nu}=\min_r, r=\overline{1,R}, \nu\le R, \qquad (9)$$

где $\theta_{r,i}=\sum_{j=1}^{n_r} I\left(X^{(r)}{}_j-a_i\right)\Big/ n_r, i=\overline{1,N}$ -- оценка *r*-ого распределения по *r*-ой обучающей выборке $\left\{X^{(r)}{}_j\right\}$; $\hat{\mathbf{P}}_r=\left\{\theta_{r,i}\right\}$, $r=\overline{1,R}$.

Доказательство сказанного автоматически вытекает из свойства сильной состоятельности оценок вероятностей $\forall r\le R:\theta_{r,i}\xrightarrow{п.н.} p_{r,i}$, $i\le N$ [4,5], а также из рассмотренных выше асимптотических свойств (8) критерия МИР.

Еще более общий вывод формулируется следующим образом.

Утверждение 4. Асимптотическая оптимальность критерия МИР в формулировке (9) распространяется на задачи (1) любой размерности *m*>1.

Доказательство следует из очевидной справедливости утверждений 1…3 при замене в них всех скалярных на соответствующие *m*-мерные (векторные) величины: $\vec{X}_j=\left(X_{j,l}, l=\overline{1,m}\right)$, $\vec{X}^{(r)}{}_j=\left(X^{(r)}{}_{j,l}, l=\overline{1,m}\right)$, $\vec{a}_i=\left(a_{i,l},\ l=\overline{1,m}\right)$, $r\le R, j\le n, i\le N$.

Отметим в заключение, что в соответствии с утверждением 1 строгим эквивалентом (9) на множестве всех допустимых решений задачи (1) в условиях априорной неопределенности является алгоритм

$$W_\nu(\mathbf{X}):\qquad \sum_{i=1}^{N} w_i \log\theta_{r,i}\,|_{r=\nu}=\max_r, r=\overline{1,R}, \nu\le R, \qquad (10)$$

реализующий критерий МП (4) в адаптивном варианте [4].

Сопоставляя выражения (9) и (10) между собой, приходим к выводу, что критерий МИР – это одна из возможных интерпретаций классического критерия МП в задачах распознавания дискретных объектов общего вида. Причем, во многих своих приложениях такая интерпретация имеет ряд важных для пользователя преимуществ. Одно из них мы уже

использовали выше при выводе выражения для ошибки распознавания (8) в терминах теоретико-информационного подхода. Второе и, по-видимому, наиболее важное преимущество критерия МИР в формулировке (9) затрагивает актуальнейшую проблему контроля надежности принимаемых решений в задачах распознавания образов. Проиллюстрируем эту проблему на следующем примере из практики лингвистического анализа.

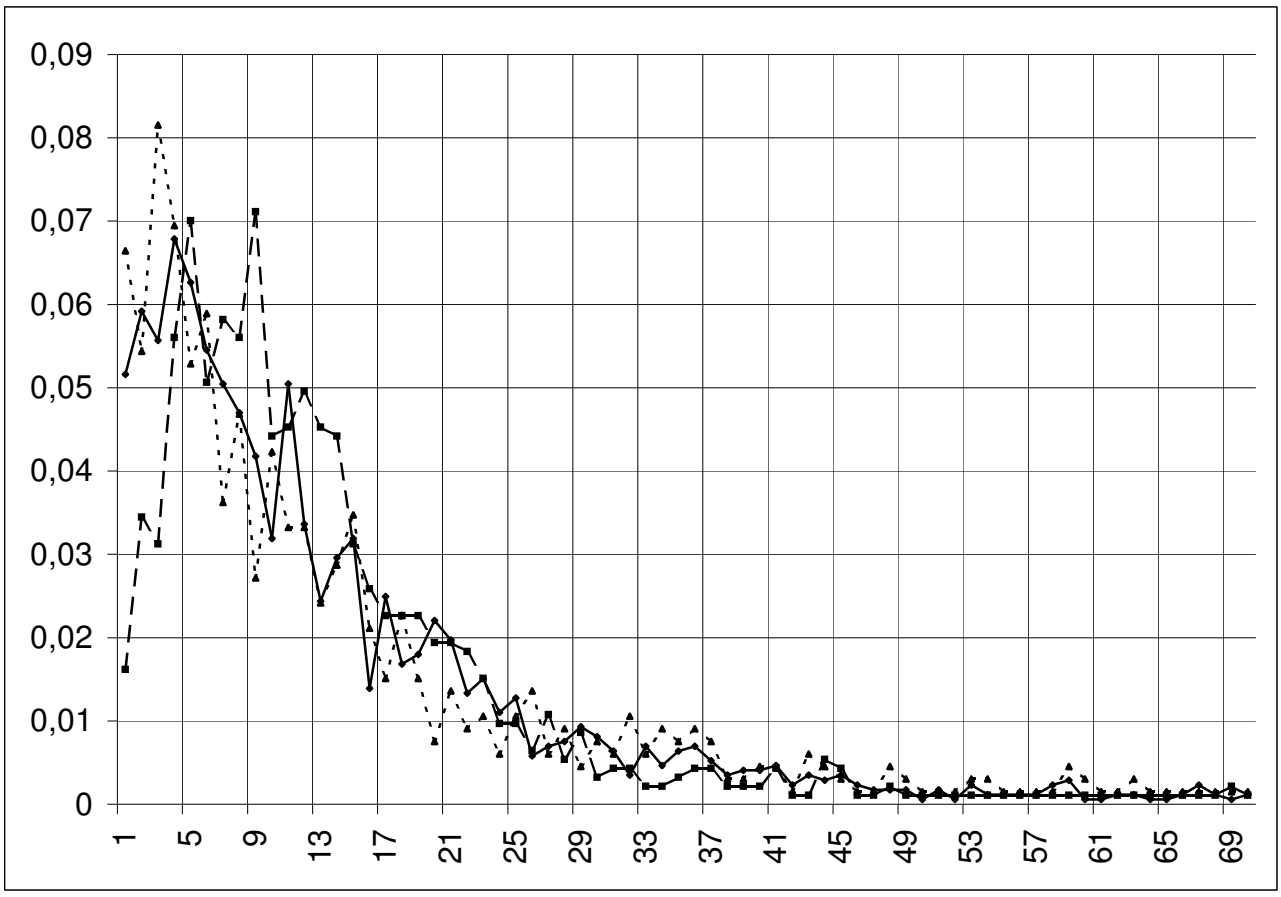
5. Пример применения. В процессе проведенных исследований были рассмотрены два текста: роман «Идиот» Ф. Достоевского и повесть «Судьба человека» М. Шолохова. В качестве показателя их стилистических особенностей, весьма условного, но, вместе с тем, информативного и наглядного, была выбрана сложность или длина каждого отдельного предложения в обоих произведениях, измеряемая количеством используемых в нем слов. В рамках применяемого статистического подхода обозначим ее как случайную величину *X* у Достоевского и *Y* у Шолохова. Никакими априорными данными об их распределениях вероятностей $\mathbf{P}_1$ и $\mathbf{P}_2$ мы не располагали. Поэтому для анализа были сначала получены эмпирические распределения $\hat{\mathbf{P}}_1$ и $\hat{\mathbf{P}}_2$ по выборкам $\mathbf{X}_r = \{X_j^{(r)}, j = \overline{1, n_r}\}$, r=1;2, объемов $n_1 = 1650$ и $n_2 = 860$ (определяются числом последовательных предложений в текстах «*Идиот*» и «*Судьба человека*» объемов 150 000 и 70 000 знаков соответственно). При этом применялась стандартная процедура вычислений с регуляризацией вида

$$\theta_{r,i} = (n_r + N)^{-1}\left[\sum_{j=1}^{n_r} I\left(X^{(r)}{}_j - i\right) + 1\right], i = \overline{1, N}, r = 1,2, \qquad (11)$$

где объем *N* множества различных состояний обоих объектов $\{a_i = i \le N\}$ был установлен равным 70 согласно максимальной длине предложений в рассматриваемых текстах. Полученные кривые эмпирических распределений показаны на рисунке в виде двух графиков: сплошная и штриховая линии соответственно. Видно, что различия между ними не носят качественного характера. Их количественная характеристика: величина информационного рассогласования была рассчитана по формуле Кульбака-Лейблера: $\rho(\mathbf{P}_1 / \mathbf{P}_2) = \sum_{i=1}^{N} \theta_{1,i} \log(\theta_{1,i} / \theta_{2,i}) = 0{,}090$. На этом был завершен этап подготовки данных для распознавания.

На втором, заключительном этапе вычислений была сформирована независимая выборка наблюдений $\{X_j\}$ объема n =590 (предложений) по фрагменту (50 000 знаков) из текста "Идиот", который не пересекался с обучающим фрагментом. После этого по аналогии с выражением (11) был рассчитан эмпирический закон распределения $\hat{\mathbf{P}} = \{w_i\}$ --

представлен пунктирной линией на рисунке, а также соответствующие значения двух решающих статистик из критерия МИР (9).



В результате было установлено соотношение

$$\rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_{1}\right) = \sum_{i=1}^{N} w_i \log\left(w_i \big/ \theta_{1,i} \right) = 0{,}059 < \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_{2}\right) = \sum_{i=1}^{N} w_i \log\left(w_i \big/ \theta_{2,i} \right) = 0{,}176.$$

Таким образом, решение здесь принимается в пользу первой гипотезы, а именно: текста "Идиот". В нашем случае, мы знаем, это верно. Однако в общем случае неизбежно возникает вопрос: насколько можно доверять принятому решению? Казалось бы, ответом на него может служить расчет вероятности возможной ошибки по формуле (8). В данном случае она равна $\alpha_{1,2} = P(\chi_{69}^{2} > 2 \cdot 590\rho(\mathbf{P}_1 / \mathbf{P}_2)) = P(\chi_{69}^{2} > 106{,}2) = 0{,}003$ [5].

Много это, или мало? На первый взгляд, очень мало: из каждой тысячи независимых испытаний мы будем ошибаться в среднем только 3 раза. Но где гарантия, что проведенный нами эксперимент не взят как раз из этой тройки? Или, еще один вопрос: насколько устойчив этот результат по отношению к составу и объему выборочных данных? И, наконец, а сами априорные данные в какой мере точны и корректны? Вопросы подобного рода особенно актуальны в задачах распознавания образов, когда исследователи принци-

пиально ограничены в имеющихся информационных ресурсах. К сожалению, существующая теория не дает на такие вопросы строгих ответов. Поэтому чаще всего на практике исследователи полагаются почти исключительно на свой опыт и интуицию, а это, разумеется, далеко не всегда приводит к положительному результату. Логика рассуждений совершенно иного рода, отталкивающаяся от строгого информационного критерия МИР, реализуется в идее "обнаружения" ошибочных решений по признаку несоблюдения в сомнительных случаях минимальной решающей статистикой из (9) своих оптимальных асимптотических свойств (8).

6. Алгоритм с обнаружением ошибок. Предположим, что при принятии решения согласно алгоритму (9) произошла ошибка, т.е. принято решение $W_\gamma(\mathbf{X})$ при справедливости гипотезы W_ν. В соответствии с выражением (8) вероятность такого события $\alpha_{\nu,\gamma} << 1$ весьма мала, т.е. произошло довольно редкое событие. С другой стороны, если оно все же произошло, тогда (a posteriori) выполняется соотношение $\rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_{\nu}\right) > \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_{\gamma}\right) = \min_r \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_{r}\right)$. Вероятность такого события (*a priory)* равна $P(\chi_{N-1}^2 > 2n \min \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_r\right)) = \beta_{\nu,\gamma}$. При малости значения $\beta_{\nu,\gamma} << 1$ наше предположение об ошибке в решении $W_\gamma(\mathbf{X})$ подтверждается (точнее, не опровергается) нетипичным характером данной конкретной выборки **X**. В таком случае первоначальное решение $W_\gamma(\mathbf{X})$ вполне может быть отброшено из рассмотрения как сомнительное, и эксперимент повторяется с новыми (откорректированными) данными.

Напротив, если полученное значение $\beta_{\nu,\gamma}$ сравнимо с 1, наше первоначальное предположение об ошибке в решении $W_\gamma(\mathbf{X})$ по логике опровергается, и его поэтому следует отклонить как недостаточно обоснованное.

Таким образом, признаком отбраковки решения $W_\gamma(\mathbf{X})$ в общем случае может служить соотношение вида

$$P(\chi_{N-1}^2 > 2n \min \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_r\right)) \le \beta,$$

или, что эквивалентно,

$$\min_r \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_r\right) \ge \chi_{N-1,1-\beta}^2 / (2n) \quad . \tag{12}$$

Здесь β -- некоторый пороговый уровень (сверху) для априорной вероятности $\beta_{\nu,\gamma}$ решения $W_\gamma(\mathbf{X})$, а $\chi_{N-1,1-\beta}^2$ - квантиль распределения Пирсона с *N*-1 степенями свободы на

уровне значимости $1-\beta$. Например, при равенстве $\beta = 0{,}1$ в условиях примера из предыдущего раздела по таблицам χ^2 – распределения получаем $\chi_{69,0.9}{}^2 / 1180 \approx 0{,}076$.

Отметим, что предложенный алгоритм асимптотически инвариантен по отношению к распределениям вероятностей $\{\mathbf{P}_r\}$. В этом смысле здесь много общего с идеями последовательного анализа А. Вальда [6]. Отличия условия (12), причем принципиального характера, состоят в реализации упомянутых идей не на основе отношения правдоподобия, а на основе асимптотически оптимальной решающей статистики $\min_r \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_r\right)$ в метрике Кульбака-Лейблера [3].

В зависимости от объема множества внутренних состояний каждого объекта *N*, а также объема выборки наблюдений *n*>*N* в данном алгоритме варьируется ограничение сверху на допустимый уровень минимальной решающей статистики из (9) в задаче распознавания образов (1). Как результат, устраняется по крайней мере часть возникающих ошибок в принимаемых решениях и, следовательно, повышается их надежность. Поясним сказанное на нашем предыдущем примере (см. рисунок).

Введем искусственные искажения в выборочный ряд распределения $\hat{\mathbf{P}} = \{w_i\}$ путем его смещения влево на две позиции (пунктирная кривая на рисунке смещается вдоль оси абсцисс влево на 2 деления). В результате получаем, строго говоря, некорректно поставленную задачу: гипотетическую выборку $\{X_j\}$ мы должны отнести к одному из альтернативных распределений $\mathbf{P}_1$ или $\mathbf{P}_2$ при том, что в действительности данная выборка не принадлежит ни одному из них. Отметим, что это острейшая проблема для всех без исключения задач распознавания образов. Никто в них не дает никаких гарантий, что рассматриваемое множество образов исчерпывает или хотя бы охватывает наблюдаемый объект. Более того, даже в лучшем случае ввиду одной только проблемы априорной неопределенности указанное строгое соответствие между объектом наблюдения и множеством образов, настроенных каждый по обучающей выборке конечного объема $n_r < \infty$, устанавливается весьма не точно, или даже ошибочно (проблема встречных гипотез). Поэтому, строго следуя критерию МП и принимая при этом любое возможное решение, в указанном случае по сути всегда получаем ошибочный результат. В отличие от классического критерия (10) критерий МИР (9) в значительной мере защищен от данного недостатка. При выполнении в нем условия (12) решение вообще не принимается. В нашем примере с искаженным выборочным распределением $\hat{\mathbf{P}}^* = \{w^*{}_i = w_{i+2}, i \le N\}$ после ряда вычислений получаем

$$\min \rho\left(\hat{\mathbf{P}}^* / \hat{\mathbf{P}}_{r}\right) = \rho\left(\hat{\mathbf{P}} / \hat{\mathbf{P}}_1\right) = 0{,}092 > \chi_{69,0.9}^{2} / 1180 \approx 0{,}076 ,$$

т.е. условие (12) выполнено. Поэтому несмотря на установленное согласно (9) соотношению вида

$$\rho\left(\hat{\mathbf{P}}^* / \hat{\mathbf{P}}_{1}\right) = 0{,}092 < \rho\left(\hat{\mathbf{P}}^* / \hat{\mathbf{P}}_{2}\right) = 0{,}240$$

решение в пользу текста «Идиот» в данном случае не принимается, что точно отвечает существу рассмотренной ситуации. Отметим, что в отличие от традиционного подхода [4] решения здесь носят (R+1)-альтернативный характер -- по числу R альтернатив распределения $\mathbf{P}$ из общей формулировки задачи (1) плюс еще одно, а именно: «нет решения».

7. Заключение. Кажется, нет никакой разницы в том, по какому алгоритму (9) или (10) приняты наши решения $W_{\nu}(\mathbf{X})$ в задаче распознавания образов**,** если они совпадают друг с другом автоматически. Однако ситуация резко усложняется на практике при неизбежном стремлении исследователя проконтролировать и по возможности гарантировать достаточную степень надежности каждого принятого им решения. И здесь выясняется, что критерий МИР в своей формулировке (9) в отличие от критерия МП содержит дополнительную важную информацию именно в отношении надежности принимаемых решений. Сначала это было показано при анализе статистических характеристик качества (8) – весьма нетривиальная, как известно [4], задача. Еще более наглядно преимущества критерия МИР проявляются в правиле (12) отбраковки сомнительных с точки зрения своей надежности решений. Рассмотренный пример из практики лингвистического анализа служит хорошей иллюстрацией к сказанному. Поэтому можно утверждать, что при прочих равных условиях в задачах распознавания дискретных объектов общего вида (1) применение критерия МИР существенно более обоснованно по сравнению с классическим критерием МП. Сделанный вывод особенно важен и актуален в расчете на широкое многообразие сложных цифровых систем, которые бурно распространяются на практике одновременно с современной вычислительной техникой.

Библиографический список.

1. *Савченко В.В.* Различение случайных сигналов в частотной области //РЭ. Изд-е РАН. 1997. Т.42. №4. - С. 426-431.

2. *Акатьев Д.Ю., Савченко В.В*. Принцип минимального информационного рассогласования при различении случайных сигналов//. Известия вузов России. 2004. №1. - С. 12 - 17.

3. *Кульбак С.* Теория информации и статистика. М.: Наука, 1967. – 408 с.

4. *Боровков А.А*. Математическая статистика. Дополнительные главы. М.: Наука, 1984. – 615 с.

5. Таблицы по математической статистике/ П. Мюллер, П. Нойман, Р. Шторм. Пер. с нем./ под ред. В.М. Ивановой. М.: Финансы и статистика, 1982. – 278 с.

6. *Вальд А.* Последовательный анализ.: Пер. с англ./ под ред. Б.А.Севастьянова. М.: Физматгиз , 1960. -- 260 с.