УДК 621.317.799

С.И. Мельник

Харьковский национальный университет радиоэлектроники, Харьков, Украина

ТЕОРЕТИКО-ИНФОРМАЦИОННЫЙ ПОДХОД К АНАЛИЗУ ИЗМЕРИТЕЛЬНЫХ ДАННЫХ И ОЦЕНКЕ ИХ НЕОПРЕДЕЛЕННОСТИ

Для использования метода максимального правдоподобия, основанного на теореме Байеса в задаче фильтрации шума необходима априорная информация о статистических свойствах, как полезного сигнала, так и шума. Если же такая информация об объекте измерения (контроля) и шума не может быть формализована, как параметры вероятностного распределения, она теряется, как правило. В работе показано, что метод минимизации алгебраической сложности описания измеренных данных является обобщением теоремы Байеса на информацию любого вида. Показано, что широко известные методы решения обратных задач выделения полезного сигнала (минимизация СКО, оптимальная спектральная фильтрация и др.) можно интерпретировать, как его частные случаи. Даны иллюстрации применения этого метода, который в отдельных случаях дает улучшение точности реконструкции в несколько раз. Предложено рассматривать относительную сложность описания результатов измерения и найденного значения искомых параметров как меру их неопределенности.

Ключевые слова: *теорема Байеса, фильтрация шума, информационная сложность описания, разрешающая способность.*

Введение

В общефизическом аспекте задача обработки измерительной информации (ИИ) состоит в выделении из нее полезной информации о свойствах наблюдаемого объекта и является некорректной. Это связано с тем, что в условной записи процедуры измерения

$$\hat{A}\left(\vec{Z};\vec{N}\right)=\vec{U}\,, \qquad (1)$$

где $\vec{Z}$ – вектор искомых параметров; $\vec{N}$ – вектор остальных факторов (условного шума); $\vec{U}$ – вектор ИИ, одному значению $\vec{U}$ соответствует множество пар $\left\{\vec{Z};\vec{N}\right\}$. Задача обработки ИИ состоит в выборе оптимальной по выбранному критерию пары $\left\{\vec{Z};\vec{N}\right\}_{\text{opt}}$, которая, естественно, зависит от этого критерия. Существует множество разнообразных критериев выбора, как частных, предназначенных для решения конкретной задачи, так и общезначимых (СКО, максимальное правдоподобие, минимум функционала штрафных санкций и т.п.).

Данная работа посвящена анализу проблемы выбора критерия обработки ИИ. Мы утверждаем, что этот выбор полностью и однозначно определяется имеющейся априорной информацией (АИ) о параметрах процедуры измерения, свойствах полезного сигнала и условного шума (соответственно параметрах $\hat{A};\vec{Z};\vec{N}$), а также целью обработки ИИ. Например, минимизация ошибок первого или второго рода при контроле качества изделий требует различных критериев выбора и, соответственно, дает разные результаты измерения при одной и той же необработанной ИИ (некоторые из изделий будут признаны годными или бракованными для разных критериев).

Наиболее часто встречающейся целью обработки ИИ является повышение точности измерений. Одним из вариантов решения этой задачи является метод максимального правдоподобия (ММП), основанный на применении теоремы Байеса и минимизирующий вероятность возможной ошибки. Однако его применение требует априорных сведений о свойствах вероятностных распределений пар $\left\{\vec{Z};\vec{N}\right\}$. В большинстве случаев такая информация отсутствует и для применения ММП выдвигают статистические гипотезы, основанные на той или иной априорной информации. Мы утверждаем, что можно упростить процедуру, непосредственно учитывая АИ при выборе оптимальной пары $\left\{\vec{Z};\vec{N}\right\}_{\text{opt}}$ и не переходя к вероятностным гипотезам. Более того, оказывается, что различные критерии оптимизации выбора $\left\{\vec{Z};\vec{N}\right\}_{\text{opt}}$ являются частными случаями критерия минимальной сложности описания ИИ относительно имеющейся АИ.

В настоящей работе сформулирован и доказан обобщенный аналог теоремы Байеса, приведены примеры его реализации при обработке ИИ. Показано, что в тех случаях, когда большая часть АИ не может быть формализована на языке математической статистики, применение обобщенного критерия позволяет в несколько раз увеличить точность измерений (разрешающую способность в приведенном примере). Приведены аргументы за принятие в качестве неопределенности результатов измерений относительной сложности описания ИИ и АИ.

Теорема Байеса в приложении к задаче реконструкции изображений

В этом разделе мы будем иллюстрировать рассматриваемую методику на примере задачи реконструкции изображений в сканирующей микроскопии при наличии аддитивного шума [1]. Схема таких измерений приведена на рис.1.

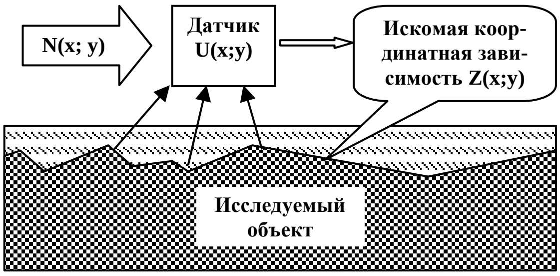

Рис. 1. Схема сканирующей микроскопии

В этом случае общая формула (1) принимает вид

$$\hat{A}\left[Z(x,y)\right] + N(x,y) = U(x,y), \qquad (2)$$

где $Z(x,y)$ – искомое планарное распределение исследуемого параметра объекта; $U(x,y)$ – двумерный массив измеренных данных; $\hat{A}$ – оператор преобразования, который может содержать случайные слагаемые (шумы) $N(x,y)$. В ряде методов (оптическая микроскопия, магнитная микроскопия и т.п.) близлежащие неоднородности вносят независимый вклад в измеряемый сигнал и можно записать (2) в виде свертки:

$$U(x,y) = A(x,y) * Z(x,y) + N(x,y). \qquad (3)$$

Если известно априорное распределение вероятности $\rho\left[Z_i(x,y)\right]$, то условная вероятность получения результата $U(x,y)$ при независимом аддитивном шуме рассчитывается как:

$$\rho\left(U_j / Z_i\right) = \rho\left(N\right) = \rho_N(U_j - \hat{A}(Z_i)), \qquad (4)$$

где $\rho(N)$ – вероятностное распределение шумов. В приложении к задаче реконструкции теорема Байеса формулируется следующим образом**:** Наиболее правдоподобным при измеренном U_j будет такое $Z_i(x,y)$, для которого произведение $\rho\left(Z_i\right)\cdot\rho_N(U_j - \hat{A}(Z_i))$ максимально. Но распределения вероятности $Z_i(x,y)$ может быть неизвестным или вообще не существовать (объект в единичном экземпляре). В этом случае теорема Байеса оказывается неприменима.

Обобщение теоремы Байеса для единичных объектов

Обратим внимание на то, что

$$\ln\left[\rho(a_i;b_j)\right] = \ln\left[\rho(a_i)\right] + \ln\left[\rho(b_j/a_i)\right] = \\ = \ln\left[\rho(b_j)\right] + \ln\left[\rho(a_i/b_j)\right].$$

Как было показано Шенноном, если событие a_i распределено с вероятностью $\rho(a_i)$, то $S(a_i) = \ln\left[\rho(a_i)\right] + C_0$ – (энтропия) единичного события a_i. Она же характеризует *<u>сложность описания</u>* события – минимальную длину программы (в битах), позволяющую записать объект a_i.

Если вероятность не задана, то формула Шеннона неприменима, но и в этом случае следует ожидать:

$$S(U_i;Z_j) = S(U_i) + S(Z_j/U_i) = S(Z_j) + S(U_i/Z_j). \qquad (5)$$

Колмогоров доказал [2], что равенство $S(U_i) - S(U_i/Z_j) \approx S(Z_j) - S(Z_j/U_i)$ выполняется с точностью до $\log S(U_i:Z_j)$. Для массива данных 1 Мб погрешность составляет не более 20 байт, т.е. 0,002%). Другими словами это означает, что количества информация, содержащаяся в Z_j об объекте U_i и наоборот – приблизительно равны. Тогда:

$$S(Z_j/U_i) \approx S(Z_j) + S(U_i/Z_j) - S(U_i) = \\ = S(Z_j) + S(N) - S(U_i). \qquad (6)$$

Так как $S(U_i) = \text{const}$**,** отсюда следует, что

для нахождения максимально правдоподобного (в обобщенном смысле) $Z_j(x,y)$ необходимо минимизировать $S(Z_j) + S(N) = S(Z_j) + S_N(U_i - Z_j)$ при заданном $U_i(x,y)$

Эта формулировка и является обобщением теоремы Байеса для единичных объектов.

Алгоритмы расчета и минимизации сложности

Для применения обобщенной теоремы Байеса необходимо рассчитать сложность входящих в нее слагаемых. В общем случае таких алгоритмов не существует в силу теоремы о невычислимости алгебраической сложности. Однако в частных задачах такие алгоритмы удается разработать с учетом априорных сведений о сигнале и шумах [3].

<u>Сложность описания шумов</u> определяется теми закономерностями, которые им присущи. Если это только статистическое распределение $\rho_N\left[N(x;y)\right]$, то сложность описания задается формулой Шеннона

$$S\left[N(x,y)\right] = n\int \rho_N\left[N(x;y)\right]\log_2 \rho_N\left[N(x;y)\right]dxdy,$$

где n – число пикселей в отсканированном массиве. В случае нормального распределения эта формула сводится к сумме квадратов отклонений. Таким образом, метод минимизации СКО при поиске решения является частным случаем, когда первым слагаемым $S(Z_j)$ в суммарной сложности описания ИИ пренебрегают и распределение аддитивного шума нормально.

Сложность описания решения $S(Z_j)$ зависит от выбора способа описания. Далее мы приведем несколько вариантов алгоритма вычисления сложности при различной априорной информации о сигнале:

1. Естественный способ (отсутствие АИ):

$$S_{\text{åñò}}(Z_j) \approx \sum_{i;j} \log_2 \left[\frac{Z(x_i; y_j)}{\delta Z} \right],$$

где δZ – дискретность (точность) задания сигнала.

2. Для непрерывных зависимостей:

$$S_{\text{í åï ð}}(Z_j) \approx \sum_{i;j} \log_2 \left[\frac{Z(x_i; y_j) - Z(x_{i-1}; y_j)}{\delta Z} \right] \le S_{\text{åñò}}(Z_j)$$

3. Для гладких зависимостей:

$$S_{\text{ãëàä}}(Z_j) \approx$$
$$\approx \sum_{i;j} \log_2 \left[\frac{Z(x_{i-1}; y_j) - 2Z(x_i; y_j) + Z(x_{i-1}; y_j)}{\delta Z^2} \right] \le$$
$$\le S_{\text{í åï ð}}(Z_j).$$

В общем случае находится такой порядок описания (при фиксированном δZ), при котором на следующем этапе сложность описания увеличивается.

<u>*Пример*</u> расчета сложности:

Пусть необработанная ИИ представлена в виде:

<u>**0 1 3 5 6 4 8 5 3 1 4 6 9 7 5 4 2 1**</u>

Для приведенной выше последовательности алгоритмов описания получим следующие последовательности:

1 2 2 1 –2 4 –3 –2 –2 3 2 3 –2 –2 –1 –2 –1 –1
1 0 –1 –3 6 –7 1 0 5 –1 1 –5 0 1 –1 1 0 2
–1 –1 –1 9 –13 8 –1 5 –6 2 –6 5 1 –2 2 –1 –1 2

Расчет сложности каждой из них по Шеннону дает:

$$\mathbf{S_1 > S_2 < S_3 < S_4} \qquad \mathbf{S_{min} = S_2}.$$

Таким образом, сравнение сложности, рассчитанной разными алгоритмами, позволяет выявить закономерности, присутствующие в ИИ. Это эквивалентно «слепому» методу реконструкции, когда параметры оператора $\hat{A}$ заранее неизвестны. В приведенном примере из результатов расчета сложности описания можно сделать вывод, что исследуемая функция непрерывная, но не гладкая.

<u>Учет АИ при описании</u> $S(Z_j)$. Априорная информация позволяет ограничить множество возможных описаний ИИ, и тем самым уменьшить сложность ее описания. В зависимости от характера АИ применяется тот или иной метод обработки ИИ.

Примеры априорной информации:

Распределение $\rho\left[Z_i(x,y)\right]$. Алгоритм ее учета – оптимальная спектральная фильтрация по Винеру.

Ограничения на амплитуду и крутизну сигнала: алгоритм - стабилизирующий функционал вида

$$\Omega(z) = \int_a^b \left\{ z^2 + p \cdot (z')^2 \right\} ds.$$

Форма неоднородностей – специализированные алгоритмы описания.

Пример реализации информационного метода при реконструкции изображений

Описанная методика была использована при разработке алгоритма реконструкции поверхностных токов в методе СКВИД – микроскопии. Проводилось сравнение с методом регуляризации Тихонова. В результате применения метода минимизации сложности описания результатов сканирования удалось повысить разрешающую способность микроскопа в 2,3 раза. При учете априорной информации о форме токовых дорожек (прямоугольная) – еще в 4 раза. Таким образом, было получено общее улучшения качества восстановления на порядок. Приведем иллюстрацию численных данных и результатов реконструкции.

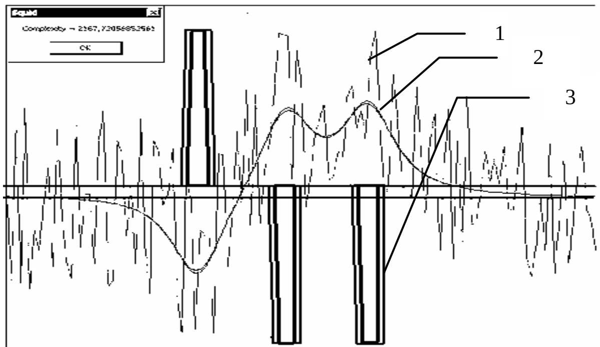

Рис. 2. Реконструкция распределения токов на основе информационного подхода: 1 – необработанный результат сканирования СКВИД-антенной; 2 – результат реконструкции полезной составляющей сигнала; 3 – результат реконструкции карты токов (априорная информация о силе тока позволяет уточнить ширину дорожек – более тонкая линия)

Визуально неразличимые на профиле исходных данных (1 на рис. 2) токи оказываются пространственно разрешены в результате реконструкции. При этом разрешающая способность, оцениваемая как точность реконструкции положения и ширины токовых дорожек, повышена на порядок. Сложность описания измеренного зашумленного сигнала (1 на рис. 2) при априорно заданном точном распределении токов – **2159** бит; минимальная сложность решения, найденного без учета АИ о характере токов (метод регуляризации Тихонова) – **2255** бит; сложность описания решения, найденного с учетом АИ - **2167** бит. Приближение к первому значению (разница уменьшилась на порядок) характеризует качество алгоритма восстановления истинного распределения токов в образце.

Сложность описания как мера неопределенности результатов измерения

Информационный подход к описанию результатов измерения позволяет предложить соответствующий критерий их неопределенности. В общефилософском смысле – неопределенность – это количество недостающей информации. В рамках теории сложности – это сложность описания точного значения измеряемых данных относительно полученных результатов измерения. То есть $S(Z_j / U_i)$ в принятых выше обозначениях. Однако, не зная точного значения измеряемой величины, мы не можем ее рассчитать точно. Поэтому для оценки неопределенности можно воспользоваться относительной сложностью восстановленного (неточного) значения $Z_{opt}(x, y)$. В случае аддитивного шума это не что иное, как его сложность. Таким образом, для **оценки неопределенности** результатов измерения получим выражение:

$$S(Z_{opt} / U_i) \approx S_N(U_i - Z_{opt}) . \quad (7)$$

В случае нормально распределенного шума это логарифм от СКО (с точностью до константы, зависящей от дискретности сигнала) относительно найденного $Z_{opt}(x, y)$.

Список литературы

1. *Гордиенко Ю.Е. Информационное направление повышения разрешающей способности микроволновой микроскопии / Ю.Е. Гордиенко, С.И. Мельник, Н.И. Слипченко // Радиотехника. – 2006. – № 147. – С. 157-163.*
2. *Kolmogoroff A.N. .Logical basis for information theory and probability theory / A.N. Kolmogoroff // IEEE Trans. Inform. Theory. – 1968. – Vol. IT-14. – P. 662-664.*
3. *Melnyk S. Thermal tomography on the basis of an information method [Электронный ресурс] / S. Melnyk. – Режим доступа к документу: qirt.gel.ulaval.ca/archives /qirt2004/papers/028.pdf.*

Поступила в редколлегию 16.07.2009

Рецензент: д-р техн. наук, проф. Ю.П. Мачехин, Харьковский национальный университет радиоэлектроники, Харьков.

ТЕОРЕТИКО-ІНФОРМАЦІЙНИЙ ПІДХІД ДО АНАЛІЗУ ВИМІРЮВАНИХ ДАНИХ ТА ОЦІНЦІ ЇХ НЕВИЗНАЧЕНОСТІ

С.І. Мельник

Для використання методу максимальної правдоподібності, що спирається на теорему Байєса, в задачі фільтрації шуму необхідна апріорна інформація що до статистичних властивостей. як корисного сигналу. так і шуму. Якщо така інформація про обєкт вимірювання (контроля) не може бути формалізована, як параметри розподілу вірогідності, вона губиться, як правило. В роботі показано, що метод мінімізації алгебраїчної складності опису вимірюваних даних є узагальненям теореми Баєсана інформацію будь якого типу. Показано, що широко відомі методи вирішення зворотних задач виділення корисного сигналу (мінімізація СКВ, оптимальна спектральна фільтрація та ін.)можна подати як йогоособливі випадки. Дани ілюстрації застосування цього методу, який в окремих випадках дає покращення точності реконструкції у декілька раз. Запропоновано розглядати відносну складність опису результатів вимірювань та знайденого значення шуканих параметрів як міру їх невизначеності.

Ключові слова: *теорема Баєса, фільтрація шуму, інформативна складність опису, разрізнювальна здатність.*

THE THEORETIC INFORMATION APPROACH TO THE ANALYSIS OF THE MEASURING DATA AND AN ESTIMATION OF THEIR UNCERTAINTY

S.I. Melnyk

The aprioristic information on statistical properties is necessary for using of a method of the maximum credibility based on the Bayes rule in a problem of a filtration of noise, both a useful signal, and noise. If such information on object of measurement (control) and noise cannot be formalized, as parameters of probability distribution, it is lost, as a rule. In the paper it is shown that the method of minimization of algebraic complexity of the description of the measured data is generalization of the Bayes rule on any kind information. It is shown that widely known methods of the decision of return problems of allocation of a useful signal (root-mean-square minimization, an optimum spectral filtration, etc.) it is possible to interpret, as its special cases. Illustrations of application of this method which on occasion gives improvement of accuracy of reconstruction several times are given. It is offered to consider relative complexity of the description of results of measurement and reconstructed values as the measure of their uncertainty.

Keywords: *Bayes rule, noise filtration, information complexity of the description, resolving power.*