УДК 519.6

В. И. ГРИЦЮК, канд. техн. наук, доц., ХНУРЭ

МОДИФИЦИРОВАННЫЕ УСТОЙЧИВЫЕ МЕТОДЫ ВЫБОРА МОДЕЛИ

Рассматривается алгоритм МНК с весовым параметром, определяемым выбранным пробным решением. Исследуются методы помехоустойчивого оценивания, обладающие численной устойчивостью. Определены асимптотические свойства полученных оценок. Предлагаются численно устойчивые методы помехоустойчивого выбора модели.

Ключевые слова: регуляризация, численная устойчивость, помехоустойчивый выбор модели.

Введение. Основная проблема, возникающая при решении задач восстановления регрессионной зависимости – проблема мультиколлинеарности. Мультиколлинеарность проявляется в сильной корреляции между двумя или более признаками, что затрудняет оценивание параметров модели. На практике встречаются случаи частичной мультиколлинеарности, когда имеется высокая степень корреляции между некоторыми признаками. Тогда решение получить можно, однако оценки параметров модели и их дисперсии могут быть неустойчивыми. Увеличиваются дисперсии оценок и абсолютные значения регрессионных параметров, что усложняет их интерпритацию. Основными методами устранения мультиколлинеарности является либо выбор признаков, либо введение ограничений на параметры [4].

При обработке информации, как правило, считают, что показатели подчиняются нормальному распределению. Однако практика обработки информации показывает, что показатели не так часто подчиняются теоретическому нормальному распределению. Наблюдаются отклонения как односторонние, так и двухсторонние, когда “хвосты” дифференциального закона оказываются более тяжелыми, чем предполагается, исходя из данных таблиц нормального распределения. Иногда статистическая информация, подлежащая обработке, по данному показателю представляет собой смесь нескольких законов распределения с разными дисперсиями. Наблюдаются смеси основного нормального распределения с распределением других видов. Встречаются случаи, когда из-за малого объёма выборки невозможно достаточно точно определить вид закона, засоряющего распределения. Кроме того, известно, что при применении метода наименьших квадратов небольшое число грубых ошибок может существенно исказить значения характеристик распределения. Исследованием и развитием робастных методов занимались П. Хубер, Д. Тьюки. В настоящее время необходима разработка таких методов обработки информации, которые были бы менее чувствительны к виду закона распределения и влиянию небольшого числа больших случайных отклонений.

Целью настоящей работы является создание методов помехоустойчивого выбора регрессионной модели, обеспечивающих получение наиболее адекватной и наименее мультиколлинеарной модели.

Робастная проверка линейных гипотез. При регуляризации параметры модели находят из минимизации функционала [1,2]

$$\beta^* = \arg\min_{\beta}\left(\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{n}\beta^j x_j^i\right)^2 + \lambda\|\beta\|^2\right) \quad (1)$$

Пусть $\hat{\beta}$ является М-оценкой, в качестве модифицированных весов примем

$$\omega_i = \frac{\psi\left[\left(\overline{b_i} - \overline{a}_i\hat{\beta}_1\right)/\hat{\sigma}\right]}{\left(\overline{b_i} - \overline{a}_i\hat{\beta}_1\right)/\hat{\sigma}}, \;\; i = 1, 2, \ldots M, \;\; M = m + n, \quad (2)$$

где $\overline{\mathrm{a}}_{\mathrm{i}}$ - i - ая строка матрицы $\overline{\mathrm{A}}$.

Помехоустойчивая мера разброса - медиана ненулевых отклонений [3] выражается

$$\hat{\sigma} = \frac{1}{0.675} Med_i\left(|r_i|\,|\; r_i \neq 0\right). \quad (3)$$

Для больших m мы можем выразить для $\hat{\beta}$

$$D(\hat{\beta}) \approx N_{k_1}(\beta, \hat{\upsilon}\;\; (\overline{\mathrm{A}}^T\overline{\mathrm{A}})^{-1}), \quad (4)$$

$$\hat{\upsilon} = \hat{\sigma}^2\,\frac{ave_i\{\psi(r_i/\hat{\sigma})^2\}}{[ave_i\{\psi'(r_i/\hat{\sigma})\}]^2}\,\frac{M}{M - k_1}, \quad (5)$$

где k_1 – ранг матрицы В.

В случае интервала и проверки для одного коэффициента β_{i} получим

$$\hat{\sigma}^2_{\beta_i} = \hat{\upsilon}\;\;\left(\overline{\mathrm{A}}^T\overline{\mathrm{A}}\right)^{-1}_{ii}, \quad (6)$$

где ii означает i-ый диагональный элемент матрицы $\left(\overline{\mathrm{A}}^T\overline{\mathrm{A}}\right)^{-1}$.

Пусть

$$\hat{\gamma} = \mathrm{B}\hat{\beta}, \quad \hat{\Sigma}_{\gamma} = \mathrm{B}\hat{\Sigma}_{\hat{\beta}}\mathrm{B}^T = \hat{\upsilon}\ \mathrm{B}\left(\overline{\mathrm{A}}^T\overline{\mathrm{A}}\right)^{-1}\mathrm{B}^T \tag{7}$$

Тогда робастная типа Уальда проверка (WTT) определяется отклоненной областью

$$\{T_W > F_{\mathrm{q,\,M\text{-}k_1}}(1-\alpha)\}, \tag{8}$$

где

$$T_W = \frac{1}{q}(\hat{\gamma}-\gamma_0)^T\,\hat{\Sigma}_{\gamma}^{-1}(\hat{\gamma}-\gamma_0), \qquad \gamma = \gamma_0. \tag{9}$$

где $F_{n_1,n_2}(\delta)$ - δ квантиль F - распределения с n_1 и n_2 степенями свободы.

Пусть $\hat{\beta}_R$ - M – оценка, вычисленная при ограничениях $\gamma = \gamma_0$

$$\hat{\beta}_R = \arg\ \min_{\beta}\ \left\{\sum_{i=1}^{M}\rho\left(\frac{r_i\left(\hat{\beta}_R\right)}{\hat{\sigma}}\right)\ :\ \mathrm{B}\beta = \gamma_o\right\} \tag{10}$$

Робастная проверка типа отношения правдоподобия (LRTT) может быть определена статистикой с ограниченной функцией ρ.

$$T_L = \sum_{i=1}^{M}\rho\left(\frac{r_i(\hat{\beta}_R)}{\hat{\sigma}}\right) - \sum_{i=1}^{M}\rho\left(\frac{r_i\left(\hat{\beta}\right)}{\hat{\sigma}}\right). \tag{11}$$

Пусть

$$\xi = \frac{E\psi'(u/\sigma)}{E\psi(u/\sigma)^2}. \tag{12}$$

Тогда можно показать [Хэмпел и др.], что при соответствующих условиях регулярности, ξT_L сходится с гипотезой H_0 к распределению хи-квадрат с q степенями свободы. Так как ξ может быть оценена как

$$\hat{\xi} = \frac{ave_i\{\psi'(r_i(\hat{\beta})/\hat{\sigma})\}}{ave_i\{\psi(r_i(\hat{\beta})/\hat{\sigma})^2\}}, \tag{13}$$

аппроксимация LRTT для больших m имеет отклоненную область

$$\hat{\xi} T_L > \chi_q^2 (1-\alpha), \quad (14)$$

где $\chi_n^2(\delta)$ обозначает δ - квантиль хи – квадрат распределения с n степенями свободы.

Заключение. Таким образом, использование ограничений на параметры в задаче МНК, позволяющего определить необходимое значение весового параметра λ, а также применение методов помехоустойчивого выбора модели, позволяет создать методы, обеспечивающие помехоустойчивый выбор модели, обладающие численной устойчивостью. Применение различных моделей должно расширить область исследований.

Список літератури: **1**. *Грицюк В. И.* Модифицированный алгоритм НК и выбор модели [Текст] / В. И. Грицюк // Вестник национального технического университета "ХПИ". - 2004, №17, С.47-50. **2**. *Грицюк В. И.* Рекуррентный алгоритм идентификации модели, основанный на ортогональном разложении [Текст] / В. И. Грицюк // Радиоэлектроника и информатика. -1998г., №3, с. 46-47. **3**. *Holland P. W.* Robust regression using interactively reweghted least squares [Text] / P. W. Holland, R. E. Welsch // Commun Statist. -1977, v. A6.- p.813-828. **4**. *Afif, A. A.* Computer-aided multivariate analysis [Text] / A. A, Afifi, V. Clark, S. May // CRS Press,-2004.

Надійшла до редколегії 05.02.2013

УДК 519.6

Модифицированные устойчивые методы выбора модели. / В.И. Грицюк // Вісник НТУ «ХПІ». Серія: Автоматика та приладобудування. – Х. : НТУ «ХПІ», 2013. – № (). – С. – Бібліогр.: 4 назв.

Досліджується алгоритм МНК з вагомим параметром, визначеним вибраним пробним рішенням. Розглянуті методи завадостійкого оцінювання, які мають чисельну стійкість. Визначені асимптотичні властивості отриманих оцінок. Запропоновано чисельно стійкі методи завадостійкого вибору моделі.

Ключові слова: регуляризація, чисельна стійкість, завадостійкий вибір моделі.

The algorithm of the least squares method with weighting parameter, which determined selected test decision, is researched. The methods of robust estimation, which have numerical stability, are considered. The asymptotic properties of received estimates are determined. Numerical stable methods of robust model choice are offered.

Key words: regurarization, numerical stability, robust model choice.