

CLUSTERED BY FEATURE LINEAR REGRESSION ON DATA WITH REAL VALUE

The authors offer a linear regression modification, based on preset clustering by char. value. Approbation of this method with some real-valued features databases showed highly better results, in comparison with the classical linear regression. The fuzzy sets method was applied to generate discrete features from the real ones.

Keywords: data mining, linear regression, clustering, fuzzy sets.

© Таскин А. С., Миркес Е. М., 2012

УДК 621.391.825

В. Н. Тяпкин, Д. Д. Дмитриев, В. Г. Коннов, А. Н. Фомин

МЕТОД ОПРЕДЕЛЕНИЯ ВЕКТОРА СПЕКТРАЛЬНЫХ КОЭФФИЦИЕНТОВ ПО КРИТЕРИЮ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ

Рассмотрена пространственная обработка сигнала в навигационной аппаратуре потребителя спутниковых радионавигационных систем для повышения ее помехоустойчивости. Приведен расчет оптимального весового коэффициента фазированной антенной решетки на основе критерия максимума правдоподобия.

Ключевые слова: навигационная аппаратура потребителей, помехоустойчивость, фазированные антенные решетки, весовой коэффициент.

Радиоприемная аппаратура спутниковых радионавигационных систем, как и любая другая радиоаппаратура, подвержена воздействию радиопомех естественного и искусственного происхождения. Исследованию направления повышения защиты от радиопомех посвящено множество работ как в отечественной, так и в зарубежной печати. Для реальных условий в последние несколько десятилетий специалистами разных стран разработано множество методов и устройств адаптации к сигнально-помеховой обстановке. Однако накопленные к настоящему времени достижения теории в существующих средствах навигации реализованы далеко не полностью. При достаточно большом разнообразии практически используемых систем защиты большинство из них по структуре и (или) параметрам отличаются от теоретически оптимальных для соответствующих условий. К настоящему времени в радиолокации и навигации с появлением и бурным развитием цифровой элементной базы (программируемых логических интегральных схем, сигнальных процессоров и др.) положение радикально изменилось. Появилась реальная возможность создавать не только надежные, высокоточные, экономичные и малогабаритные цифровые эквиваленты существующих аналоговых систем защиты от помех, но и принципиально новые системы с теоретически существенно более высокой эффективностью.

Наиболее перспективными в настоящее время признаются методы борьбы с помехами на основе пространственной селекции, обеспечиваемой при помощи многоэлементных фазированных антенных решеток (ФАР) [1; 2].

Пространственная селекция сигналов и активных помех основана на использовании отличий в угловых направлениях их воздействия на приемную аппаратуру. В результате пространственной селекции обеспечивается защита от активных помех, действующих по боковым лепесткам диаграммы направленности антенны (ДНА).

ФАР с цифровой реализацией ДНА обеспечивает ее формирование с нулями в направлении помехи и максимумами в направлении на навигационный космический аппарат (НКА). Поскольку весовые коэффициенты для формирования ДНА рассчитываются в программной части, имеется возможность применения различных алгоритмов определения коэффициентов и, соответственно, ДНА в зависимости от поставленной задачи и условий функционирования [3].

Формирование максимумов ДНА в направлении на источник сигнала повышает отношение сигнал/помеха, что повышает помехоустойчивость, а общий рост чувствительности обеспечивает прием НКА в сложных условиях, например, в лесу.

В ФАР можно сравнительно просто реализовать специальные виды амплитудно-фазового распределения (АФР) в раскрыве и применять различные методы обработки сигналов, принятых отдельными излучателями антенны. Это позволяет получать низкие уровни боковых лепестков ДНА, извлекать больше информации из проходящих к антенне радиоволн от различных источников, использовать адаптивные алгоритмы оптимальной пространственно-временной фильтрации сигналов на фоне помех, что в целом улучшает характеристики навигационной аппаратуры потребителя.

Наиболее сложной операцией оптимального многоканального обнаружителя является оценка корреляционной матрицы помех, ее обращение и вычисление весового вектора, но с помощью схемы M -элементной ФАР этот процесс можно реализовать (см. рисунок). Вместе с тем именно обратная корреляционная матрица помех содержит всю информацию об угловых положениях источников помех и излучаемой ими спектральной плотности мощности помехи.

Рассмотрим вычисление весового вектора на основе критерия максимума правдоподобия.

Принятый сигнал в общем случае описывается векторной функцией, учитывающей его значения во времени и пространстве.

Помеховые колебания воздействуют в каждом из каналов дискретной приемной аппаратуры, поэтому их будем описывать вектором столбцом:

$$n(t) = \|n_1(t), n_2(t), n_3(t) \dots n_M(t)\|^T, \quad (1)$$

где M – число элементов приемной аппаратуры.

При выделении полезного сигнала на фоне помех используются различия между ними. Математически эти различия лучше учитывать зависимостью сигналов и помех от параметров, которые могут быть пространственными, временными и частотными.

С учетом сказанного, принимаемое колебание, в общем случае, представим в виде выражения

$$y(t) = x(t, \alpha, \beta) + n(t, \nu), \quad (2)$$

где $x(t, \alpha, \beta)$ – вектор принимаемого полезного сигнала с параметрами α и β ; α – вектор информативных параметров (фаза, время задержки, частота Доплера и т. д); β – вектор неинформативных параметров, вызванных

флюктуациями сигнала; $n(t, \nu)$ – вектор помеховых колебаний; ν – вектор параметров помехи.

Относительно помехи будем полагать, что она представляет собой векторный случайный стационарный процесс.

Нормальный закон распределения помеховых колебаний в большинстве случаев физически оправдан, поскольку в сравнительно узкополосных частотных трактах приемников помеха нормализуется.

Тогда можно записать:

$$P_n(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-y^2 / 2\sigma^2), \quad (3)$$

где σ^2 – дисперсия помехи.

Полагаем, что сигнал и помеха воздействуют на вход многоканальной M -элементной ФАР, расположенной в одном или нескольких пунктах приема. Снимаемая с выходов M -элементной ФАР совокупность напряжений описывается функцией времени $y_1(t), y_2(t), \dots, y_M(t)$ и образуют вектор-столбец $y(t) = \|y_1(t), y_2(t), y_3(t) \dots y_M(t)\|^T$. При этом одноканальный прием ($M = 1$) рассматривается как частный случай многоканального.

Дискретизацию принимаемых колебаний, в том числе и помеховых, можно производить на радио- и видеочастоте.

Дискретизация на радиочастоте обычно осуществляется с целью упрощения закона распределения помехи. В этом смысле такая дискретизация представляет собой своеобразный математический прием, облегчающей задачу синтеза аналоговых устройств обработки. При пониженной радиочастоте эта дискретизация имеет также целевое назначение при синтезе устройств дискретной (цифровой) обработки.

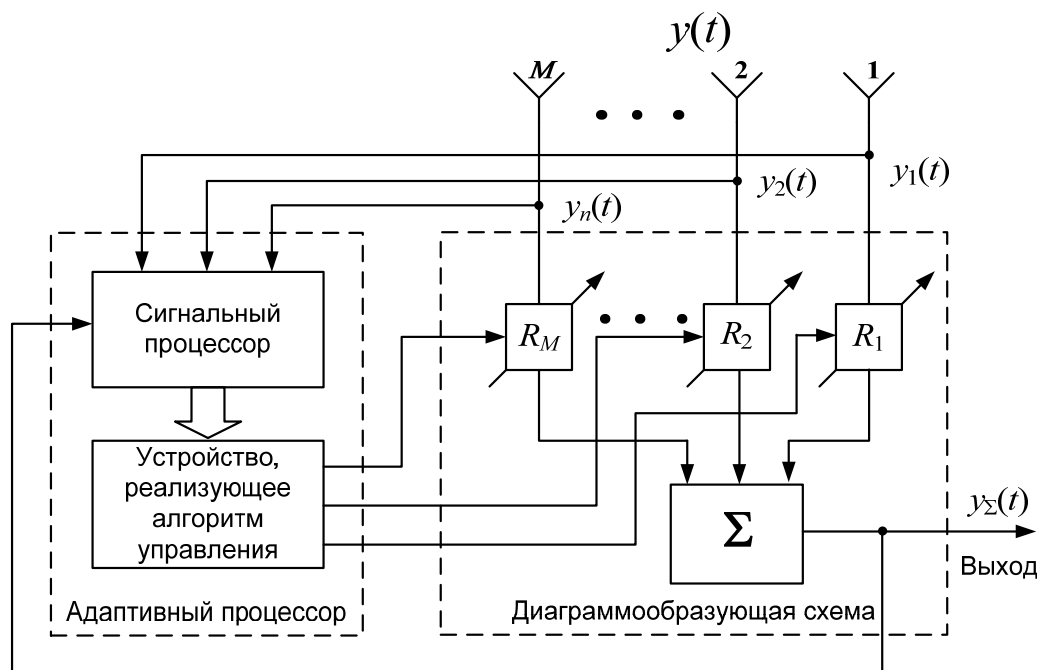


Схема M -элементной ФАР

Особенностью дискретизации на радиочастоте f_0 является относительно малый интервал дискретизации Δt , равный примерно половине периода радиочастотного заполнения $T_0 = 1/f_0$, т. е. $\Delta t \approx 1/2f_0 = T_0/2$. Дискретные значения радиопомехи y_i являются случайными числами, которые удобно представить в виде вектора столбца $y = \|y(i\Delta t)\| = \|y(i)\|$, $i = \overline{1, L}$, где L определяет длину интервала наблюдения T , а именно: $L = T/\Delta t$.

В случае многоканального приема вектор принимаемых колебаний записывается в следующем виде: $\dot{Y} = \|y_i(t)\| = \|y_k^i\|$, элементами вектора являются значения функции $y_i(t)$ в дискретные моменты времени $t_k = k\Delta t$, $k = \overline{1, L}$. Общее количество дискрет составит $m = ML$. Пронумеровав эти дискреты от 1 до m , принимаемому колебанию ФАР будет соответствовать многомерная плотность распределения, представленная следующим выражением:

$$P_n(\dot{Y}\dot{Y}^*) = (4\pi)^{-m} |\Phi^{-1}| \exp(-\dot{Y}^{*T} \Phi^{-1} \dot{Y} / 2), \quad (4)$$

где

$$\dot{Y} = \|\dot{Y}_l\|, \quad l = 1, m \quad \Phi = \|\Phi_{ik}\| = M \|\dot{Y}_i \dot{Y}_k^* / 2\|. \quad (5)$$

Необходимую для расчета отношения правдоподобия условную плотность вероятности $P_{cn}(\dot{Y}\dot{Y}^*)$ можно получить по известной плотности вероятности $P_n(\dot{Y}\dot{Y}^*)$. При наличии сигнала выборка Y характеризуется ненулевым математическим ожиданием:

$$M(Y) = M[X(\alpha)] + M[n(v)] = M[X(\alpha)] = X(\alpha).$$

В этом случае каждая дискрета y_i выборки Y получает приращение $x_i(\alpha)$, а само распределение $P_{cn}(\dot{Y}\dot{Y}^*)$ смещается на величину математического ожидания $X(\alpha)$.

В результате имеем

$$P_{cn}(\dot{Y}\dot{Y}^*) = P_n[\dot{Y} - \dot{X}(\alpha), \dot{Y}^* - \dot{X}^*(\alpha)]. \quad (7)$$

Составляя логарифм отношения правдоподобия $\ln l = \ln[P_{cn}(\dot{Y}\dot{Y}^*) / P_n(\dot{Y}\dot{Y}^*)]$ и подставляя в него (4) и (6) получаем

$$\begin{aligned} \ln l(\alpha) = & -[\dot{Y} - \dot{X}(\alpha)]^{*T} \Phi^{-1} [\dot{Y} - \dot{X}(\alpha)] / 2 + \\ & + \dot{Y}^{*T} \Phi^{-1} \dot{Y} / 2 = \dot{Y}^{*T} \Phi^{-1} \dot{X}(\alpha) / 2 + \\ & + \dot{X}(\alpha)^{*T} \Phi^{-1} \dot{Y} / 2 - \dot{X}(\alpha)^{*T} \Phi^{-1} \dot{X}(\alpha) / 2. \end{aligned} \quad (8)$$

В этом выражении первые два слагаемые образуют сумму комплексно-сопряженных величин. После транспонирования и комплексного сопряжения первого слагаемого последнее приводится к виду $\dot{X}(\alpha)^{*T} \Phi^{-1} \dot{Y} / 2$. При этом было учтено свойство эрмитовости обратной матрицы Φ^{-1} , а именно $(\Phi^{-1})^{*T} = \Phi^{-1}$. В результате сумма первых двух слагаемых (8) будет равна

$$\zeta = \text{Re } \dot{Z}(\alpha), \quad (9)$$

где

$$\dot{Z}(\alpha) = \dot{Z} = \dot{X}(\alpha)^{*T} \Phi^{-1} \dot{Y} = \dot{Y}^T (\Phi^{-1})^* \dot{X}^*(\alpha). \quad (10)$$

Вводя обозначение

$$q^2(\alpha) = \dot{X}(\alpha)^{*T} \Phi^{-1} \dot{X}(\alpha) / 2,$$

соотношение (8) окончательно можно записать в виде

$$\ln l(\alpha) = \zeta(\alpha) - q^2(\alpha) / 2.$$

Первое слагаемое в (12) представляет собой реальную часть комплексной весовой суммы:

$$\dot{Z}(\alpha) = \sum_{i,k} \dot{Y}_i (\Phi^{-1})_{ik}^* X_k^*(\alpha),$$

зависящей от дискреты принимаемого колебания y . Параметр $q^2(\alpha)$ в (12) по своей природе является чисто вещественным и непосредственно от принимаемой реализации не зависит. Он пропорционален энергии полезного сигнала на входе устройства обработки.

Из сказанного следует, что отношение правдоподобия и его логарифм в первую очередь определяется весовой суммой $\zeta = \text{Re } \dot{Z}(\alpha)$, вычисление которой определяет первый и главный этап обработки принимаемого колебания y . По весовой сумме $\zeta(\alpha)$ вычисляются те существенные операции, которые необходимы для извлечения из принимаемой реализации y всей информации о параметрах принятого сигнала. Поэтому $\zeta(\alpha)$ наряду с $\ln l(\alpha)$ также является достаточной статистикой. Вычисление комплексной весовой суммы предусматривает предварительное формирование весового вектора:

$$\dot{R}(\alpha) = \Phi^{-1} \dot{X}(\alpha). \quad (14)$$

с последующим образованием m -элементной весовой суммы

$$\dot{Z}(\alpha) = \dot{Y}^T \dot{R}^*(\alpha) = \sum_i \dot{Y}_i \dot{R}_i^*. \quad (15)$$

Данное выражение является откликом адаптивной ФАР. Если рассматривать другие критерии эффективности, такие как отношение сигнал-шум, минимальной средней квадратической ошибки, минимум дисперсии шума и другие, то полученные результаты по поиску оптимального весового вектора отличаются лишь постоянным множителем.

Таким образом, получено аналитическое выражение для вычисления весового вектора ФАР, а тот факт, что все решения для оптимального весового вектора в адаптивных ФАР при использовании различных критериев эффективности сводятся (с точностью до постоянного множителя) к винеровскому решению, выбор критерия эффективности принципиального значения не имеет. Более важным является выбор алгоритма управления, предназначенного для постройки ДНА, поскольку он непосредственно влияет на скорость переходного процесса и на сложность технической реализации системы в целом.

Библиографические ссылки

1. ГЛОНАСС. Принципы построения и функционирования / под ред. А. И. Перова, В. Н. Харисова. 4-е изд., перераб. и доп. М.: Радиотехника, 2010.
2. Сетевые спутниковые радионавигационные системы / В. С. Шебшаевич, П. П. Дмитриев, Н. В. Иванцевич и др. М.: Радио и связь, 1993.
3. Тяпкин В. Н., Лубкин И. А. Использование рекуррентных адаптивных алгоритмов для решения задачи подавления активно-шумовых помех в системах спутниковой связи // Вестник СибГАУ. 2010. Вып. 2 (28). С. 39–43.

V. N. Tyapkin, D. D. Dmitriev, V. G. Konnov, A. N. Fomin

METHOD OF DETERMINATION OF SPECTRUM FACTORS VECTOR DUE TO LIKELIHOOD RATIO CRITERION

The authors consider spatial signal processing in consumer navigation equipment of satellite navigation systems for to improve its noise immunity. The calculations of the optimal weight ratio of phased-array antenna, based on maximum likelihood criterion, are presented.

Keywords: consumer navigation equipment, noise immunity, phased-array antenna, weight ratio.

© Тяпкин В. Н., Дмитриев Д. Д., Коннов В. Г., Фомин А. Н., 2012

УДК 314.7:51

Е. Ю. Федорова

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МИГРАЦИИ В КРАСНОЯРСКОМ КРАЕ

Проведен статистический анализ данных по миграции населения в Красноярском крае за период с января 1998 г. по август 2010 г. Построены уравнения на основе регрессионного, авторегрессионного анализа и моделирования периодических колебаний.

Ключевые слова: миграция, прогнозирование, регрессия, авторегрессия, периодические колебания.

Прогнозирование изменения численности населения является важнейшей задачей демографии, так как численность населения влияет на экономическое развитие как региона, так и страны в целом. Росстат при расчете прогнозов численности населения, миграционного прироста, продолжительности жизни и других показателей применяет компонентный метод (возрастных передвижек) [1, с. 495–496]. Данный метод использует показатели только текущего расчетного периода, что не обеспечивает точность прогноза, а только определяет тенденцию, исходя из ситуации, сложившейся в текущем периоде. Но для того чтобы проводить краткосрочные и долгосрочные экономические расчеты, осуществлять планирование инвестиций, развивать те или иные экономические и социальные сферы регионов страны, необходимо прогноз численности населения осуществлять как можно точнее.

Миграция населения является одним из основных факторов динамики численности и структуры населения России и отдельных ее регионов. Правительство РФ выделяет в Концепции демографического развития РФ на период до 2015 г. [2] следующие основные задачи в области миграции и расселения:

– регулирование миграционных потоков в целях создания действенных механизмов замещения естественной убыли населения Российской Федерации;

– повышение эффективности использования миграционных потоков путем достижения соответствия их объемов, направлений и состава интересам социально-экономического развития Российской Федерации;

– обеспечение интеграции мигрантов в российское общество и формирование толерантного к ним отношения.

В интервью директора Института демографии Высшей школы экономики Анатолий Григорьевич Вишневым сказал: «Миграция сегодня – одна из самых острых проблем. Для многих природа роста миграции непонятна, и наша задача сделать видимыми те глубокие процессы, которые лежат в ее основе и которые уже нельзя изменить. И, соответственно, необходимо решать совершенно иные задачи – размышлять не о том, как остановить процесс, а о том, как его направить в «правильное» русло и контролировать» [3].

Анализ данных по иммиграции. Объектом статистического анализа являются абсолютные показатели иммиграции населения по Красноярскому краю. Временной ряд состоит из 152 значений внешней для региона иммиграции [4]. Данные представлены ежемесячно и охватывают период с января 1998 г. по август 2010 г. Наиболее часто встречающееся значение прибывших мигрантов равно 1 885. Среднее значение прибывших человек составляет 2 010. Значение 1 924 человек приходится на середину временного ряда прибывших мигрантов и делит его на две равные по числу единиц части. Всего на территорию Красноярского края с января 1998 г. по август 2010 г. прибыло 305 516 человек. При разбивке данных по годам (рис. 1) прослеживаются следующие тенденции: во второй половине года, по сравнению с первой, наблюдается подъем иммиграции, пик иммиграции приходится на сентябрь–октябрь, минимальное количество прибывших в основном приходится на январь и май. Это указывает на то, что иммиграционная активность населения совпадает с экономической и социальной активностью общества в осенний период.