ПОСТРОЕНИЕ ЛОГИКО-ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ ЦЕПЕЙ МАРКОВА

Светлана Неделько

***Abstract**: The method of logic and probabilistic models constructing for multivariate heterogeneous time series is offered. There are some important properties of these models, e.g. universality. In this paper also discussed the logic and probabilistic models distinctive features in comparison with hidden Markov processes. The early proposed time series forecasting algorithm is tested on applied task.*

***Keywords**: multivariate heterogeneous time series, pattern recognition, classification, deciding functions, logic and probabilistic models.*

***ACM Classification Keywords**: G.3 Probability and statistics: time series analysis, Markov processes, multivariate statistics, nonparametric statistics; G.1.6. Numerical analysis: optimization.*

***Conference**: The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009*

Введение

Задачи анализа и прогнозирования многомерных разнотипных временных рядов в настоящее время представляют большой интерес для исследования. В зависимости от предположений о прогнозируемой функции, постановки задач делятся на статистические (вероятностные) и детерминированные. В вероятностной постановке задач прогнозирования и идентификации модели математической моделью временного ряда $Z(t)$ выступает случайный процесс.

Среди процессов с дискретным временем большой интерес для задач прогнозирования представляют стационарные процессы с конечной длиной значимой предыстории, т. е. такие процессы, для которых распределение величин в момент t не зависит от t и всецело определяется реализовавшимися значениями на предыдущих d моментах времени. Случайные процессы с такими свойствами называются *марковскими процессами*. Процессы рассмотренного типа удобны для анализа тем, что их реализацию можно свести к обычной таблице данных, после чего воспользоваться методами статистического анализа.

При $d = 1$ временной ряд с дискретным множеством значений (состояний) называется цепью Маркова, которая полностью определяется матрицей переходных вероятностей $p_{ij} = P\left(z(t) = j / z(t-1) = i\right)$.

Вероятности можно непосредственно оценить по выборочной реализации как частоты соответствующих переходов. В случае $d > 1$ матрица переходов становится многомерной, и ее непосредственное оценивание по частотам уже при небольших d требует большой длины обучения. В этом случае необходимо использовать специализированные методы прогнозирования. Существенно расширяет область применимости марковских цепей подход, основанный на использовании скрытых марковских процессов.

Скрытые марковские модели [Baum, 1970] строятся обычно в рамках параметрического подхода, при этом вид условных распределений на наблюдаемых переменных для скрытых состояний задается эвристически (экспертно). Для подбора параметров модели используется, в основном, критерий максимума правдоподобия.

Другим подходом к аппроксимации процесса, задающего временной ряд, выступает построение логико-вероятностных моделей [Lbov, Nedel'ko, 2001]. В отличие от скрытых марковских моделей, здесь задается лишь класс разбиений, области которых и рассматриваются в качестве состояний, являющихся наблюдаемыми. При построении логико-вероятностных моделей можно использовать различные варианты [Неделько, 2008] критериев информативности.

Постановка задачи

Имеется n-мерный временной ряд $v = \left\{ z^t \,\middle|\, t = \overline{1, N} \right\}$, $z^t = \left(z_1^t, ..., z_n^t \right)$, $z_j^t \in Z_j$. Здесь Z_j – множество допустимых значений j-й переменной ряда, $Z = \prod_{j=1}^{n} Z_j$. В наборе переменных $Z_1, ..., Z_n$ могут быть одновременно переменные разных типов.

Пусть ряд является реализацией случайного n-мерного процесса $z(t)$ с дискретным временем, который задается переходной (условной) вероятностной мерой $P : \Lambda \times Z^d \to [0,1]$, где Λ – σ-алгебра подмножеств из Z, а d – длина предыстории, которая определяет распределение в заданный момент.

Требуется на основе имеющихся данных v построить прогноз временного ряда в моменты времени $t > N$ в соответствии с некоторым критерием. Также можно ставить задачу аппроксимации условного распределения, т. е. построения модели описывающего ряд процесса. Найденная модель используется для анализа ряда, а также прогнозирования его значений.

Введем обозначения $X \equiv Z^d$, $Y \equiv Z$, причем X будем использовать для обозначения пространства предысторий, а Y — для пространства значений в момент времени, для которого делается прогноз. Тогда переходную меру можно записать как $P\,[Z/z(t-1), z(t-2), ..., z(t-d)] \equiv P\,[Y/x]$.

В данных обозначениях вероятность события $E_Y \in \Lambda$, $E_Y \subseteq Z$, записывается как

$$P\,\left(z(t) \in E_Y / z(t-1), z(t-2), ..., z(t-d) \right) \equiv P\left(E_Y / x \right) \equiv P\,[Y/x](E_Y).$$

Заметим, что круглые скобки используются для указания аргумента функции, а квадратные — как часть обозначения меры.

Во введенных обозначениях прогнозируемые значения есть $y_j(t) = z_j(t),\; j = 1, ..., n$, а переменные, используемые для прогнозирования,

$$x_j(t) = z_j(t-1),\; x_{j+n}(t) = z_j(t-2),\; ...,\; x_{j+n(d-1)}(t) = z_j(t-d),\; j = 1, ..., n.$$

Размерность пространства прогнозирующих переменных есть $m = nd$.

Логико-вероятностная модель

В работе будем прогнозировать многомерный разнотипный временной ряд в классе логических решающих функций, что сводится к задаче построения логико-вероятностной модели ряда. Пусть одновариантная решающая функция имеет вид $f : X \to Y$, многовариантное решающее правило представляется в виде $f : X \to S$, где решение $s \in S$ представляет собой множество пар

$s = \left\{ \left(E_Y^k, \tilde{p}_k \right) \,\middle|\, k = \overline{1, M} \right\}$, $\tilde{p}_k$ – оценка условной вероятности $P\left(E_Y^k / x \right)$, $E_Y^k \subseteq Y$.

Тогда логико-вероятностной моделью является многовариантное решающее правило $f_L = \left\{ \left(E_X^l, s_l \right) \middle| \ l = \overline{1, L} \right\}$, где $E_X^l \subseteq X$ образуют разбиение $\alpha_L = \left\{ E_X^l \ \middle| \ l = \overline{1, L} \right\}$ пространства X.

Решающую функцию можно строить на основе восстановления условного распределения либо в заданном классе в соответствии с эмпирическим критерием K. Используемый в данной работе подход предполагает частичную аппроксимацию распределений. Аппроксимирующей моделью случайного процесса будем называть случайный процесс, заданный на сигма-алгебре, являющейся подмножеством исходной. Таким образом, модель определяет вероятности не всех событий. Более того, практически мы будем рассматривать только модели, определяющие вероятности лишь на множествах конечной алгебры, порожденной множествами некоторого конечного разбиения пространства переменных. Вообще, говоря о сигма-алгебре, мы везде подразумеваем, что в частном случае она может быть конечной.

Определение. Носителем модели назовем совокупность таких множеств ее σ-алгебры, которые не включают в себя других элементов σ-алгебры.

Определение. Сложностью модели назовем мощность ее носителя.

Определение. Класс моделей Φ называется универсальным, если для любого $P[Z]$ и любого $\varepsilon > 0$ найдется модель $\overline{P}[Z] \in \Phi$, для которой $K(P[Z]) - K(\overline{P}[Z]) < \varepsilon$.

Содержательно, универсальность означает, что моделями класса можно сколь угодно точно аппроксимировать любое распределение (на изначально фиксированной σ-алгебре), при этом под точностью аппроксимации понимается близость значений критерия информативности.

Если пространство переменных непрерывно, то универсальный класс должен содержать, очевидно, бесконечное число моделей сколь угодно большой сложности. Однако, при этом он не обязан содержать моделей бесконечной сложности.

Определение. Класс моделей Φ назовем замкнутым, если для любых двух его моделей существует модель, носитель которой содержит пересечение носителей этих двух моделей.

Гипотеза. Если класс моделей Φ является замкнутым и лебегово замыкание объединения σ-алгебр моделей класса дает исходную σ-алгебру, то такой класс является универсальным.

Содержательно здесь утверждается, что универсальность класса моделей определяется тем, что модели в совокупности порождают ту же σ-алгебру событий, на которой задана вероятностная мера. Из данного утверждения следует, что класс логико-вероятностных моделей является универсальным.

Заметим, что сформулированное предположение тесно связано с универсальностью класса логических решающих функций [Лбов, Старцева, 1999], а также со свойством universal consistency решающих функций.

Критерий информативности

Одним из критериев информативности, используемых для построения логико-вероятностной модели, является критерий на основе дивергенции, который может быть записан в виде

$$K_C\left(P[X,Y]\right) = \int \ln \frac{dP[Y/x]}{dP[Y]} \, dP[X,Y] = \int \ln \frac{dP[X,Y]}{dP[X]dP[Y]} \, dP[X,Y].$$

Известно, что дивергенция является неотрицательной величиной и равна нулю, только если распределения совпадают. Видно, что выражение критерия свелось к дивергенции между совместным распределением и произведением маргинальных распределений, что является величиной, характеризующей степень зависимости между X и Y. Заметим, что марковский случайный процесс полностью определяется вероятностной мерой $P[X,Y]$.

Под $\mu[Z]$ будем понимать меру Лебега, если Z – непрерывное пространство, и считающую меру, если Z – дискретно. Для разнотипного Z мера $\mu[Z]$ будет определяться как сумма лебеговых мер непрерывных компонент множества.

Свойство. Значения дивергенции лежат в диапазоне $[0,+\infty]$, в частности, существуют марковские случайные процессы, для которых $K_C(P[X,Y]) = +\infty$.

Рассмотрим следующий пример. Пусть $Z=[0,1]$, $X=Y=Z$. Зададим случайный процесс через плотности

$$\frac{dP[X]}{d\mu[X]} = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}, \qquad \frac{dP[Y/x]}{d\mu[Y]} = \begin{cases} 0{,}5, & y \in [0,1],\ y \neq x \\ 0, & y \notin [0,1] \end{cases}, \qquad P(y = x/x) = 0{,}5.$$

При данном процессе значение единственной переменной в следующий момент времени с вероятностью $0{,}5$ остается прежним, а с вероятностью $0{,}5$ изменяется на случайное значение из интервала $[0,\ 1]$. Для построенного примера $K_C(P[X,Y]) = +\infty$.

Таким образом, дивергенция не просто неограничена, но может принимать бесконечные значения.

Определение. Энтропией меры $P[Z]$ будем называть величину $H(P[Z]) = -\int \ln \frac{dP[Z]}{d\mu[Z]}\, dP[Z]$.

Дивергенцию можно легко выразить через энтропию

$$K_C(P[X,Y]) = H(P[X]) + H(P[Y]) - H(P[X,Y]).$$

Здесь мы предполагаем конечность всех энтропий.

Связь дивергенции и правдоподобия

Пусть $\varphi(y) = \frac{dP[Y]}{d\mu[Y]}$ – безусловная плотность вероятности, $\varphi(y/x) = \frac{dP[Y/x]}{d\mu[Y]}$ – условная плотность вероятности, а $\varphi(x,y) = \frac{dP[X,Y]}{d\mu[X,Y]}$ – совместная плотность вероятности, и $\nu = (z^1,\ldots,z^N)$ – выборка, представляющая собой реализацию случайного процесса. Для простоты рассмотрим случай $d=1$.

Функция правдоподобия выборки есть

$$\pi(\nu) = \ln\varphi(z^1) + \sum_{i=2}^{N} \ln\varphi(z^i/z^{i-1}) = \sum_{i=2}^{N} \ln\varphi(z^{i-1},z^i) - \sum_{i=2}^{N} \ln\varphi(z^i).$$

Пусть $\widetilde{P}[Y]$ – эмпирическая (выборочная) мера (каждой выборочной точке, кроме первой, приписано значение $\frac{1}{N-1}$), а $\widetilde{P}[X,Y]$ – эмпирическая (выборочная) мера на парах, составленных из соседних выборочных значений (каждой паре приписано значение $\frac{1}{N-1}$). Тогда функция правдоподобия может быть записана в виде

$$\pi(\nu) = \int \ln\varphi(x,y)\, d\widetilde{P}[X,Y] - \int \ln\varphi(y)\, d\widetilde{P}[Y].$$

При достаточно больших объемах выборки эмпирическую меру можно приближенно заменить вероятностной мерой (в соответствии с которой получена выборка). Тогда в правой части получим разность энтропий

$$\pi(\nu) \approx H(P[Y]) - H(P[X,Y]).$$

Нетрудно убедиться, что это выражение получается и в случае $d > 1$.

Заметим, что полученное выражение отличается от $K_C(P[X,Y])$ на величину $H(P[X])$.

Данное отличие очень существенно. Если дивергенция определяется исключительно вероятностными мерами, то значение правдоподобия зависит от выбора меры μ. Так, например, при умножении меры μ на константу A к величине правдоподобия прибавится слагаемое $-\ln A$. Таким образом, абсолютное значение правдоподобия не имеет содержательного смысла, а имеет смысл лишь отношение правдоподобия (или разность логарифмов).

На самом деле, нам как раз и нужно сравнивать модели, поэтому разность логарифмов правдоподобия была бы подходящей, если бы ее можно было вычислить во всех практических ситуациях. Однако для нахождения отношения правдоподобия необходимо, чтобы сравниваемые модели определяли вероятности на одном и том же множестве событий. При этом разные логико-вероятностные модели содержат оценки вероятностей для разных разбиений (совокупностей областей), и вовсе не обязаны включать в себя оценки вероятностей для областей из пересечения разбиений. Поэтому критерий правдоподобия, в отличие от дивергенции, практически не пригоден для оценивания качества логико-вероятностных моделей.

Заметим, что указанное обстоятельство является одним из наиболее существенных отличий рассматриваемого подхода, основанного на построении логико-вероятностных моделей, от подхода, связанного с построением скрытых марковских моделей, в котором требуется, чтобы модели были сравнимы по правдоподобию.

Метод прогнозирования

Зафиксируем некоторое разбиение $\lambda = \left\{E^{\omega} \subseteq Z \;\middle|\; \omega = \overline{1,k}\right\}$, $\bigcup\limits_{\omega=1}^{k} E^{\omega} = Z$, $\omega \neq \varpi \Rightarrow E^{\omega} \cap E^{\varpi} = \varnothing$, пространства Z. Теперь исходному многомерному ряду *v* можно сопоставить одномерную символьную последовательность $w = \left\{ \omega^t \;\middle|\; z^t \in E^{\omega^t},\ t = \overline{1,N} \right\}$. Случайному процессу $z(t)$ будет соответствовать процесс $\omega(t)$, переходные вероятности для которого обозначим

$$p_{\omega_0|\omega_1,\omega_2,\dots,\omega_d} = P\big(\omega(t) = \omega_0 / \omega(t-1) = \omega_1,\ \dots,\ \omega(t-d) = \omega_d\big).$$

Критерий качества, основанный на дивергенции, вводится следующим образом:

$$K_C(\lambda) = \sum_{\omega_0=1}^{k} \dots \sum_{\omega_d=1}^{k} p_{\omega_0\omega_1\dots\omega_d} \ln \frac{p_{\omega_0|\omega_1,\omega_2,\dots,\omega_d}}{p_{\omega_0}}.$$

Приближенное к оптимальному разбиение λ пространства переменных ряда Z ищется алгоритмом направленного поиска LRP [Лбов, Старцева, 1999]. Оценивается матрица переходных вероятностей между состояниями, представляющими собой области из разбиения. Критерием качества построенного решения является мера информативности матрицы переходных вероятностей между состояниями.

Задача прогнозирования состояния ионосферы

Ionosphere — массив данных, содержащий результаты электромагнитного зондирования ионосферы, представленный в коллекции задач UCI (University of California, Irvine) Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets.html). Имеется 351 объект, характеризующийся 34 переменными. Каждая из 17 пар переменных есть действительная Z_1 и мнимая Z_2 части комплексной величины, соответствующей некоторой характеристике объекта. Известно, что для 225 объектов характеристики соответствуют хорошему прохождению определенного сигнала через ионосферу, а для 126 объектов плохому прохождению этого сигнала. Будем обозначать эти случаи как классы 1 и 0. Целью является прогнозирование класса (0 или 1), т.е. ставится задача классификации.

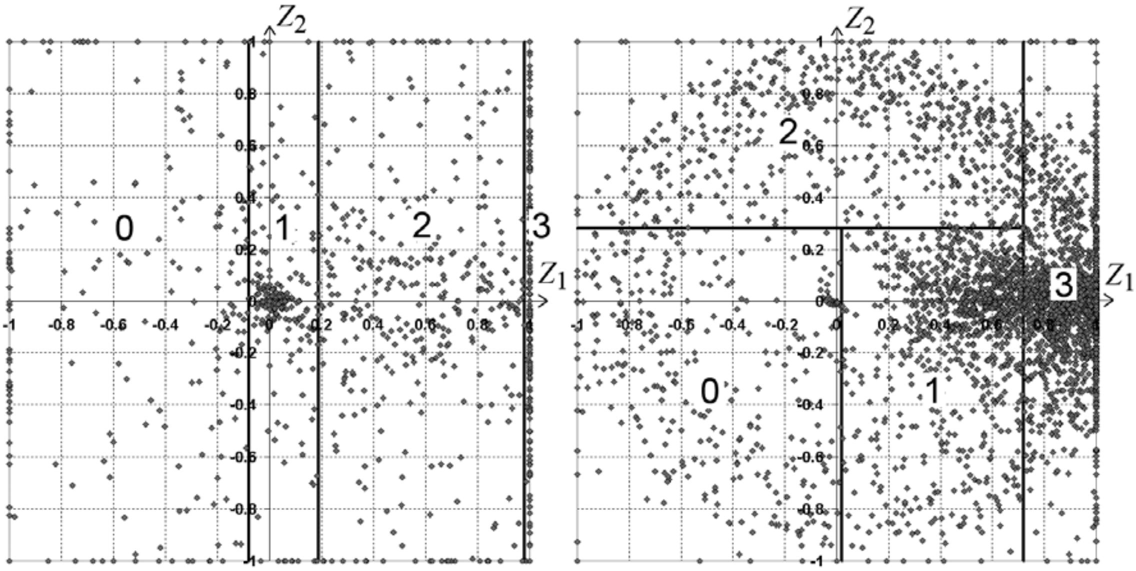

Рис. 1. Области наиболее информативного разбиения в пространстве значений характеристик сигнала для классов 0 – слева, и 1 – справа.

На основе предварительного анализа массива данных было сделано предположение, что указанные переменные соответствуют значениям одной физической величины, измеренной в последовательные моменты времени, т.е. что строки исходной таблицы можно интерпретировать как отрезки двумерного временного ряда длиной 17.

Метод классификации предлагается следующий. Временные ряды каждого класса объединяются в один временной ряд, для которого предложенным выше методом находятся закономерности. На рисунке 1 представлены объединенные временные ряды для обоих классов (отображены только точки, соответствующие значениям ряда, при этом траектории движения не показаны ввиду громоздкости), а также области разбиения пространства характеристик сигнала для моделей по каждому классу отдельно.

В таблице 1 приведены матрицы оценок переходных вероятностей между областями для классов 0 и 1 (левая и правая части таблицы). При этом в первой строке и первом столбце указаны оценки априорных вероятностей нахождения в соответствующих состояниях.

Табл. 1. Оценки переходных и априорных вероятностей для классов 0 – слева, и 1 – справа.

p_ω		0,18	0,34	0,16	0,32
	ω	**0**	**1**	**2**	**3**
0,18	**0**	0,33	0,27	0,12	0,27
0,34	**1**	0,14	0,57	0,07	0,22
0,16	**2**	0,15	0,16	0,50	0,19
0,32	**3**	0,14	0,23	0,11	0,52

p_ω		0,08	0,24	0,15	0,52
	ω	**0**	**1**	**2**	**3**
0,08	**0**	0,78	0,12	0,00	0,09
0,24	**1**	0,01	0,74	0,03	0,21
0,15	**2**	0,10	0,03	0,78	0,09
0,52	**3**	0,00	0,09	0,05	0,86

Для каждой из 351 отдельной реализации вычислялось правдоподобие по отношению к обеим моделям. В пространстве полученных значений правдоподобия строилось решающая функция (правило классификации). При этом ошибка на скользящем экзамене составила 11,4%. В случае априорной классификации (объектам приписывается преобладающий класс 1) ошибка составляет 0,36. Это говорит о том, что построенные логико-вероятностные модели действительно являются информативными и отражают свойства, специфичные для классов.

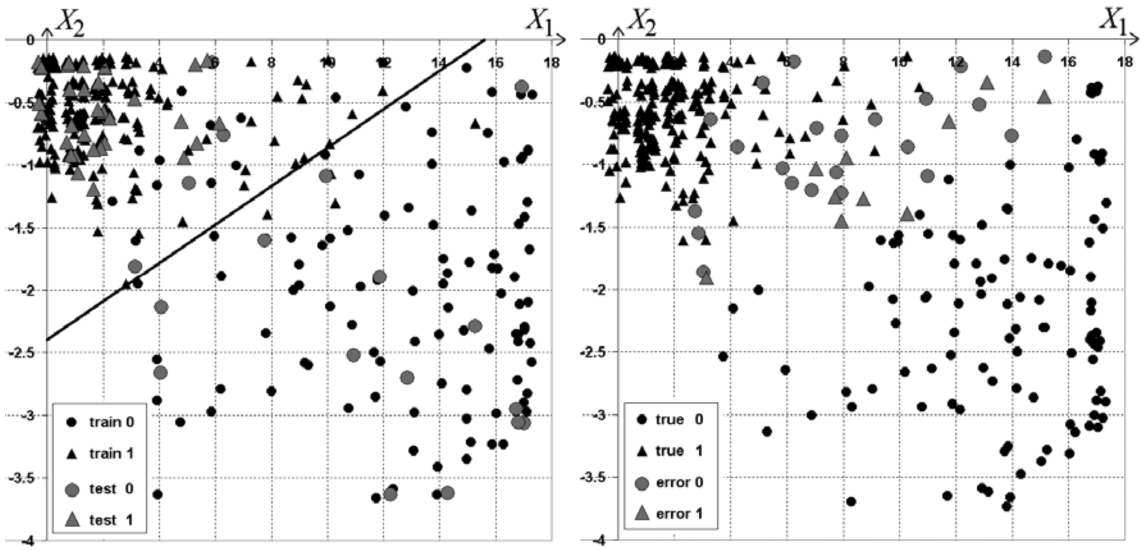

Рис.2. Реализации классов в пространстве значений правдоподобия и голосования по методу ближайшего соседа по подпространствам с выделением контрольной выборки (слева) и скользящего экзамена (справа).

Полученные значения качества классификации близки к значениям, достигаемым другими методами классификации, не интерпретирующими данные как временные ряды, т.е. использующими только «статические» свойства данных. Предложенный метод, наоборот, использует только «динамические» закономерности. Естественно ожидать, что комбинированное использование обоих типов информации может улучшить результат.

В качестве дополнительного метода был взят метод ближайшего соседа. Для этого метода при вычислении расстояния в исходном 34-мерном пространстве ошибка на скользящем экзамене составила 13%; для модификации метода с голосованием по 17 подпространствам ошибка на скользящем экзамене составила 13,7%. Таким образом, эти вариации метода можно считать равноценными по качеству. В случае k ближайших соседей ошибка получается более 13% и растет с ростом k.

Вариант метода ближайшего соседа с голосованием по 17 двумерным подпространствам позволяет сформировать признак X_1 – количество «голосов» за принадлежность объекта классу 0. Признак X_2 – величина правдоподобия объекта к логико-вероятностной модели, построенной по всем реализациям первого класса.

При построении линейного классификатора в пространстве (X_1, X_1) ошибка на скользящем экзамене составила 9,1%. Результаты представлены на рис. 2. На левой диаграмме изображены точки обучающей (train) и контрольной выборки (test) для нулевого и первого классов, а также линейная разделяющая функция. На правой диаграмме изображены объекты, правильно (true) и неправильно (error) классифицированные в процессе скользящего экзамена. При этом значения (X_1, X_1) для каждого объекта соответствуют модели, построенной без использования этого объекта для обучения.

Применение динамической модели, учитывающей упорядоченность данных, дает меньшую ошибку классификации по сравнению с методом ближайшего соседа. При этом одновременный учет статических и динамических свойств данных позволяет еще уменьшить этот показатель.

Заключение

В работе рассмотрен метод анализа многомерного разнотипного временного ряда, основанный на построении логико-вероятностной модели, представляющей собой марковскую цепь, заданную на состояниях, выбираемых по реализации в соответствии с критерием максимума информативности. На примере решения прикладной задачи продемонстрирована возможность использования метода для формирования признаков, позволяющих классифицировать временные последовательности.

Благодарности

Статья частично финансированна из проекта **ITHEA XXI** Института Информационных теории и Приложений FOI ITHEA и Консорциума FOI Bulgaria (www.ithea.org, www.foibg.com).

Литература

[Baum, 1970] L. Baum et. al. A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. Annals of Mathematical Statistics, 1970. 41. P. 164–171.

[Lbov, Nedel'ko, 2001] G.S. Lbov, V.M. Nedel'ko. A Maximum informativity criterion for the forecasting several variables of different types. // Computer data analysis and modeling. Robustness and computer intensive methods. Minsk, 2001, vol 2. P. 43–48.

[Неделько, 2008] С.В. Неделько. Исследование статистической устойчивости логико-вероятностных моделей временного ряда // Научный вестник НГТУ, 2008, №4(33), с. 43-52.

[Лбов, Старцева, 1999] Г.С. Лбов, Н.Г. Старцева. Логические решающие функции и вопросы статистической устойчивости решений. Институт математики СО РАН, Новосибирск, 1999, 211 с.

Информация об авторе

***Светлана Неделько** – ассистент кафедры Высшей математики НГТУ, 630092, Россия, г. Новосибирск, проспект К. Маркса, 20, e-mail: nedelko@math.nsc.ru*