

## ИСПОЛЬЗОВАНИЕ МОДЕЛИ СУГЕНО ДЛЯ ПРОГНОЗИРОВАНИЯ МЕТЕОРОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ

М. Г. Матвеев\*, В. В. Михайлов\*\*, М. Е. Семенов\*\*

\* Воронежский государственный университет

\*\* Воронежский военный авиационный инженерный университет

Поступила в редакцию 12.03.2011 г.

**Аннотация.** Предложен подход к построению статистических моделей прогнозирования с использованием модели нечеткой логики Сугено, апробированный на временных рядах метеорологических показателей.

**Ключевые слова:** система нечетких продукционных правил Сугено, классы эквивалентности и толерантности, векторная авторегрессия.

**Abstract.** The approach to construction of statistical models of forecasting with use of model of fuzzy logic Sugeno approved on time numbers of meteorological indicators is offered.

**Key words:** System in fuzzy logic rules of Sugeno, equivalence and tolerance classes, vector autoregress.

### ВВЕДЕНИЕ

Метеорологическое прогнозирование достигло определенных успехов прежде всего за счет численного моделирования, опирающегося на огромное количество статистической информации о значениях метеорологических параметров. Для прогнозирования могут быть использованы модели в виде систем уравнений, отражающих фундаментальные законы сохранения и феноменологические зависимости потоков от сил их вызывающих, основанные на учете всех известных закономерностей процессов, протекающих в атмосфере [1]. Как правило, параметры таких моделей идентифицируются на основе имеющихся ансамблей статистических данных. Другой подход к моделированию атмосферных процессов основан на использовании принципа «черного ящика» и сводится к составлению систем регрессионных зависимостей заранее определенных типов, например, [2,3]. В последнее время широкое распространение получили модели в виде искусственных нейронных сетей, демонстрирующие неплохие результаты прогнозов, в том числе в метеорологии [3].

В любом случае при моделировании возникает задача параметрической идентификации или обучения искусственной нейронной сети, исходной информацией для которой является статистический ансамбль данных, отвечающих

определенным требованиям. К наиболее важным требованиям относятся достаточно большой объем статистической выборки и сохранение стабильности основных условий протекания процессов, то есть выполнение условия воспроизводимости результатов многократных наблюдений.

Как правило, такие требования в условиях временных рядов метеорологических наблюдений являются трудновыполнимыми. Условия протекания процессов могут за непродолжительный срок кардинально измениться и существенно ограничить объем статистически однородной выборки. Одним из подходов к решению этой проблемы являются модели онлайн-адаптации, которые позволяют подстраиваться под изменяющиеся условия протекания процессов [3,4]. Адаптивные модели способны показывать хорошие результаты, но только в случаях относительно медленного изменения условий, что не всегда характерно для атмосферных процессов. Другим подходом к решению этой проблемы могло бы быть формулирование границ подобластей статистических данных, в которых основные условия протекания процессов сохраняются, объединение этих областей и построение соответствующих моделей для каждого объединения. Однако здесь возникают трудности формулировки четких границ таких подобластей. Более или менее обосновано можно указать лишь центр такой области, удаление от которого приводит к сни-

жению роли основных условий рассматриваемой подобласти и возрастанию влияния условий смежной подобласти. Если в качестве прогнозных моделей используются регрессионные зависимости, то модель демонстрирует наилучшие результаты в точке, соответствующей среднему выборочному значению факторов. Отклонение от этой точки неизбежно снижает качество модели. В таких случаях адекватным математическим аппаратом может быть теория нечетких множеств и основанные на ней модели нечеткого логического вывода Сугено [5], которые рассматриваются как универсальный аппроксиматор.

## 1. СХЕМА МОДЕЛИРОВАНИЯ

Предлагаемый подход к решению задачи прогнозирования можно отобразить следующей последовательностью действий:

1. Определяются классы (подобласти) статистических данных,  $K_r \in X_1 \times \dots \times X_n$ , где  $X_i$  – диапазон изменения  $i$ -го параметра условий протекания процесса. Для этого выбирается вектор признаков (условий), по которым тот или иной элемент выборки может быть отнесен к определенному классу. Компоненты вектора признаков  $\bar{\pi}$  рассматриваются как функции от измеряемых метеорологических величин, представленных вектором  $\bar{x}$ .

2. С использованием статистических данных, принадлежащих  $r$ -му классу, строится прогнозная модель  $M_r(\bar{x})$ ; таких моделей может быть столько же, сколько выделено классов статистических данных.

3. Осуществляется переход к нечетким объектам, то есть фазификация исходной информации, в процессе которой классы  $K_r$  представляются как нечеткие подмножества. Здесь основной задачей является целесообразный выбор типа функции принадлежности элементов выборки определенному классу.

4. Строится система нечетких продукционных правил типа Сугено – «если вектор  $\bar{\pi}$  измеренных параметров условий процесса принадлежит некоторому классу со значением принадлежности  $\mu_{K_r}(\bar{\pi})$ , то прогнозное значение метеорологического показателя  $y^r$  может быть рассчитано по модели  $M_r$  со степенью обоснованности  $\mu_{K_r}(\bar{\pi})$ », которые в формальном виде можно записать нечеткого продукционного правила

$$\text{если } \bar{\pi} \in K_r, \text{ то } y^r = M_r(\bar{x}), \forall r. \quad (1)$$

5. Дефазифицированное значение показателя  $y$  рассчитывается по методу «центра тяжести»

$$y = \frac{\sum_r y^r \mu_{K_r}(\bar{\pi})}{\sum_r \mu_{K_r}(\bar{\pi})}. \quad (2)$$

## 2. ПОСТРОЕНИЕ КЛАССОВ ЭКВИВАЛЕНТНОСТИ СТАТИСТИЧЕСКИХ ДАННЫХ

Предложенный подход рассмотрим на примере скалярного поля температур  $x$ , представленного длинными временными рядами в узлах плоской сетки, параллельной поверхности земли на некоторой высоте (рис. 1). Дискретность сетки составляет  $2,5^\circ$  по широте и долготе. Временные ряды в узлах сетки коррелированы между собой. Коэффициент корреляции уменьшается по мере увеличения расстояния между узлами. Например, смежные временные ряды имеют парный коэффициент корреляции в пределах от 0,85 до 0,99. Для рядов удаленных друг от друга на 10–15 шагов сетки коэффициент парной корреляции становится статистически незначимым. Вычисление частного коэффициента корреляции показывает лишь очень незначительное снижение зависимости близких рядов. В такой ситуации анализ временного ряда следует проводить с учетом поведения смежных рядов, т.е. рассматривать многомерные временные ряды.

Для выделения классов  $K_r$  необходимо определить признаки этих классов. Одним из признаков может быть направление изменения температуры. В качестве второго признака может рассматриваться интенсивность этого изменения.

Направление изменения по каждой координатной оси определяется конечной разностью температур  $x$ . В показанной схеме две конечные разности по каждому направлению: по горизонтальному направлению  $\Delta_i = x_{i,j} - x_{i-1,j}$ ; и  $\Delta_{i+1} = x_{i+1,j} - x_{i,j}$ ; по вертикальному –  $\Delta_j = x_{i,j} - x_{i,j-1}$ ; и  $\Delta_{j+1} = x_{i,j+1} - x_{i,j}$ .

Начнем выделение классов с единственным признаком – направлением изменения температуры, при этом косвенно постараемся учесть и второй признак – интенсивность. Для этого сложим конечные разности по каждому направлению: горизонтальному –  $\delta_1 = \Delta_i + \Delta_{i+1}$  и вертикальному –  $\delta_2 = \Delta_j + \Delta_{j+1}$ . Результаты сложения будем рассматривать как координаты вектора направления, которые показывают, в какой

из четвертой схемы (1, 2, 3, 4), обозначенной на рис.2, находится этот вектор. Сложение конечных разностей позволяет учесть степень устойчивости направления изменения температуры. Если конечные разности по одному направлению имеют одинаковый знак, то это характеризует устойчивое изменение, если знаки разные, то направление меняется, а значит, вектор изменения приближается к границе четверти. Таким образом, выделяется 4 класса:  $K_1, K_2, K_3, K_4$ . Будем рассматривать эти классы как классы эквивалентности, определяемые таблицей 1.

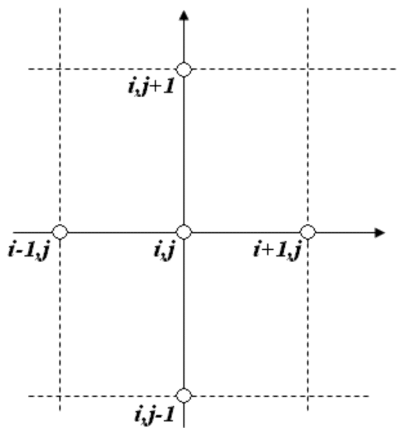


Рис. 1. Фрагмент сетки скалярного поля температур

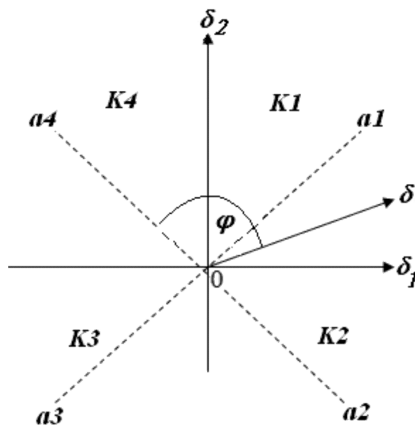


Рис. 2. Схема определения классов эквивалентности и толерантности

Таким образом, для каждой точки  $x^t_{i,j}$ , в момент времени  $t$ , с помощью введенных признаков можно установить соответствующий ей класс эквивалентности.

Таблица 1

Отношения, определяющие классы эквивалентности  $K_1, K_2, K_3, K_4$

$\delta_1 \geq 0, \delta_2 \geq 0$	$K_1$
$\delta_1 \geq 0, \delta_2 \leq 0$	$K_2$
$\delta_1 \leq 0, \delta_2 \leq 0$	$K_3$
$\delta_1 \leq 0, \delta_2 \geq 0$	$K_4$

### 3. АПРБАЦИЯ НА МОДЕЛИ ВЕКТОРНОЙ АВТОРЕГРЕССИИ

Следующая проблема – выбор типа модели прогнозирования,  $M_r$ , которая должна уметь адекватно описывать поведение нестационарного многомерного ряда при соблюдении условий выделенного класса исходных данных.

Будем считать, что изменение температуры происходит за счет некоторого внутреннего источника и конвективных процессов переноса воздушных масс. Если рассматривать процесс изменения температуры как марковский процесс, то можно считать, что температура в рассматриваемый момент времени в заданном узле сетки будет определяться температурой в предыдущий момент времени в этом же узле и температурами в предыдущий момент времени в смежных узлах. В первом приближении (с учетом взаимного влияния рядов только в смежных узлах) адекватным описанием такого представления является модель векторной авторегрессии первого порядка, представляющая собой систему разностных уравнений вида

$$x^t_{i,j} = \alpha_{i,j} x^t_{i,j} + \alpha_{i-1,j} x^t_{i-1,j} + \alpha_{i,j+1} x^t_{i,j+1} + \alpha_{i+1,j} x^t_{i+1,j} + \alpha_{i,j-1} x^t_{i,j-1} + \xi^t_{i,j}, \quad (3)$$

записанную для конечного значения индексных переменных  $i$  и  $j$ .

Систему уравнений (3) удобно представить в матричной форме

$$x^t = Ax^{t-1} + \xi^t,$$

где  $x^t \in R^{n \times m}$ ,  $A$  – пятидиагональная матрица размерности  $n \times m$ ,  $\xi^t$  –  $n \times m$ -вектор случайных величин с нулевым средним и конечной дисперсией.

Таким образом, для того чтобы описать поведение поля температур в прямоугольнике сетки  $n \times m$ , необходимо  $n \times m$  разностных уравнений вида (3), у которых оценке подлежат

параметры  $\alpha_{i,j}$ . При этом измерения должны производиться в узлах сетки ограниченной прямоугольником  $(n+1) \times (m+1)$ . Поскольку уравнение типа (3) может служить описанием только стационарных центрированных процессов, переменная  $x_{i,j}^t$  должна представлять собой либо элемент стационарного центрированного ряда, либо конечную разность заданного порядка исходной переменной – температуры.

Для проведения эксперимента с реальными данными взяты промежутки временных рядов, отвечающие условиям стационарности и центрированности. Такой подход позволит, по крайней мере, проверить гипотезу о возможности применения модели Сугено. В дальнейшем (за рамками данной статьи) необходимо будет вернуться к нестационарным рядам.

Рассматриваемые данные распределились по классам следующим образом: в классе  $K_3$  оказалось 49 уровней; в классе  $K_4$  – 24 уровня и в классе  $K_2$  – 3 уровня. Построим модели типа (3) по каждому классу, кроме  $K_2$  (в этом классе мало данных).

Рассмотрим единственную точку в узле  $i,j$ . Сначала построим обычную модель векторной авторегрессии вида (3), которую используем для сравнения с предлагаемым подходом

$$x_{i,j}^{t+1} = a_{i,j}x_{i,j}^t + a_{i-1,j}x_{i-1,j}^t + a_{i,j+1}x_{i,j+1}^t + a_{i+1,j}x_{i+1,j}^t + a_{i,j-1}x_{i,j-1}^t.$$

Идентификация этой модели по имеющейся статистике (76 временных уровней) дает характеристику модели –  $R^2 = 0,836$  и статистически значимые (только один из пяти значений  $t$ -статистики находится на критическом уровне) оценки параметров модели:

$$a_{i,j} = -3,61; a_{i-1,j} = 2,96; a_{i,j+1} = 0,22; a_{i+1,j} = 1,75; a_{i,j-1} = -0,50.$$

Несмотря на снижение размера выборок результаты моделирования по классам лучше, чем в общей выборке:

– модель  $M_3$  в классе  $K_3$  характеризуется коэффициентом детерминации  $R^2 = 0,846$  и безоговорочно статистически значимыми оценками параметров

$$a_{i,j} = -1,86; a_{i-1,j} = 1,50; a_{i,j+1} = 0,68; a_{i+1,j} = 0,56; a_{i,j-1} = -0,06,$$

– модель  $M_4$  в классе  $K_4$  характеризуется коэффициентом детерминации  $R^2 = 0,864$  и также безоговорочно значимыми оценками (несмотря на уменьшение объема выборки)

$$a_{i,j} = -5,30; a_{i-1,j} = 4,46; a_{i,j+1} = -0,19; a_{i+1,j} = 3,33; a_{i,j-1} = -1,36.$$

Это означает, что любая линейная комбинация вида (2) даст результат лучше, чем результат, полученный по неклассифицированной, исходной выборке.

Необходимо найти рациональные коэффициенты такой линейной комбинации, которые определяются значениями функций принадлежности  $\mu_{Kr}(\bar{x})$ . Для построения функции принадлежности вектора к классу рассмотрим схему на рис. 2.

Представляется рациональным следующее распределение степени принадлежности направления вектора  $\delta = (\delta_1, \delta_2)$  первой четверти схемы (класс  $K_1$ ): если направление совпадает с прямой  $(0, a_1)$ , то степень принадлежности этого направления равна единице. Отклонение направления  $\delta$  влево или вправо от  $(0, a_1)$  влечет симметричное уменьшение степени принадлежности примерно до 0,5; пока направление  $\delta$  остается в пределах первой четверти. Отклонение этого направления в четвертую четверть до прямой  $(0, a_4)$  и отклонение во вторую четверть до прямой  $(0, a_2)$  симметрично уменьшает степень принадлежности до нуля. Степень принадлежности вектора  $\delta$  другим четвертям схемы 2 определяется аналогично.

Такой подход обеспечивает замену введенных классов эквивалентности на соответствующие классы толерантности статистических данных. Классу  $K_1$  соответствует полуплоскость справа от прямой  $a_2 - a_4$ , классу  $K_2$  – полуплоскость справа от прямой  $a_1 - a_3$ , классу  $K_3$  – полуплоскость слева от прямой  $a_2 - a_4$  и классу  $K_4$  – полуплоскость слева от прямой  $a_1 - a_3$  дальнейшие описания позволяют для новых классов толерантности оставить те же обозначения, что и для классов эквивалентности.

Тип функции принадлежности наиболее адекватный вышеприведенному распределению принадлежности вектора  $\delta$  классу толерантности выбран в классе тригонометрических функций  $f(\cos \phi)$ , где  $\phi$  – угол между границами полуплоскостей  $(a_1 - a_3)$  и  $(a_2 - a_4)$  и вектором  $\delta$  (см. рис.2).

$$\mu_{Kr}(\delta) = 1 - |\cos \phi| = 1 - \left| \frac{b_i \delta_1 + b_j \delta_2}{\sqrt{(\delta_1^2 + \delta_2^2)(b_i^2 + b_j^2)}} \right|, \quad (4)$$

$$r = 1, 2, 3, 4.$$

где  $b_i$  и  $b_j$  – векторы, определяемые классом толерантности; для класса  $K_1$  это вектор  $(-1, 1)$ ; для  $K_2 - (1, 1)$ ; для  $K_3 - (1, -1)$  и для  $K_4 - (-1, -1)$ .

Нечеткие продукционные правила вида (1), с учетом введенных классов толерантности, можно записать в виде следующей системы для каждого  $i, j$ -узла сетки

$$\begin{aligned} &\text{если } ((\delta_1 \leq 0) \wedge (\delta_2 \geq 0) \wedge (|\delta_1| \leq \delta_2)) \vee \\ &((\delta_1 \geq 0) \wedge (\delta_2 \geq 0)) \vee ((\delta_1 \geq 0) \wedge \\ &(\delta_2 \leq 0) \wedge (\delta_1 \geq |\delta_2|)), \text{ то } \mu_{K_1}(\delta) \wedge M_1; \end{aligned} \quad (5)$$

$$\begin{aligned} &\text{если } ((\delta_1 \geq 0) \wedge (\delta_2 \geq 0) \wedge (\delta_1 \leq \delta_2)) \vee \\ &((\delta_1 \geq 0) \wedge (\delta_2 \leq 0)) \vee ((\delta_1 \leq 0) \wedge (\delta_2 \leq 0) \wedge \\ &(|\delta_1| \leq |\delta_2|)), \text{ то } \mu_{K_2}(\delta) \wedge M_2; \end{aligned} \quad (6)$$

$$\begin{aligned} &\text{если } ((\delta_1 \leq 0) \wedge (\delta_2 \geq 0) \wedge (|\delta_1| \geq \delta_2)) \vee \\ &((\delta_1 \leq 0) \wedge (\delta_2 \leq 0)) \vee ((\delta_1 \geq 0) \wedge (\delta_2 \leq 0) \\ &\wedge (\delta_1 \leq |\delta_2|)), \text{ то } \mu_{K_3}(\delta) \wedge M_3; \end{aligned} \quad (7)$$

$$\begin{aligned} &\text{если } ((\delta_1 \leq 0) \wedge (\delta_2 \leq 0) \wedge (|\delta_1| \geq |\delta_2|)) \vee \\ &((\delta_1 \leq 0) \wedge (\delta_2 \geq 0)) \vee \\ &((\delta_1 \geq 0) \wedge (\delta_2 \geq 0) \wedge (\delta_1 \leq \delta_2)), \\ &\text{то } \mu_{K_4}(\delta) \wedge M_4, \end{aligned} \quad (8)$$

$M_r$  – векторная авторегрессия, полученная по статистике  $K_r$  класса эквивалентности

$$\begin{aligned} x_{i,j}^{t+1,r} = &a_{i,j}^r x_{i,j}^t + a_{i-1,j}^r x_{i-1,j}^t + a_{i,j+1}^r x_{i,j+1}^t + \\ &+ a_{i+1,j}^r x_{i+1,j}^t + a_{i,j-1}^r x_{i,j-1}^t, \end{aligned}$$

где  $a_{i,j}^r$  – оценки параметров  $\alpha_{i,j}^r$  в выражении (3).

Для каждого  $i, j$ -узла сетки прогноз рассчитывается по всем четырем моделям с обязательным предварительным обнулением всех значений  $\mu_{K_r}(\delta)$ .

Последним этапом расчета прогноза температуры является дефаззификация по формуле (2), которая в рассматриваемом случае принимает вид

$$x_{i,j}^{t+1} = \frac{\sum_r x_{i,j}^{t+1,r} \mu_{K_r}(\delta)}{\sum_r \mu_{K_r}(\delta)}. \quad (9)$$

**Матвеев Михаил Григорьевич** – д.т.н., профессор кафедры программирования и информационных технологий, Воронежский государственный университет. Тел. (473) 220-84-70. E-mail: mgmatveev@yandex.ru

Применительно к моделям  $K_3$  и  $K_4$  расчеты по формуле (9) показывают, что коэффициент детерминации по сравнению с классической моделью векторной авторегрессии (3) увеличивается до величины  $R^2 = 0,870$ , что подтверждает эффективность предложенного подхода к построению прогнозной модели.

## ЗАКЛЮЧЕНИЕ

Использование модели Сугено при анализе многомерных временных рядов позволяет корректно использовать статистическую информацию для прогнозного моделирования, что положительно сказывается на прогностических свойствах моделей этих рядов. Следует подчеркнуть, что модели прогнозирования могут выбираться не только в классе векторной авторегрессии. Например, аналогичное исследование целесообразно провести применительно к нейросетевым моделям. Кроме того, в качестве предметной области может выступать и экономика и ряд технических приложений. При этом в качестве объектов исследования должны рассматриваться признаки выделяемых классов и способы построения функций принадлежности.

## СПИСОК ЛИТЕРАТУРЫ

1. Матвеев Л.Т. Курс общей физики атмосферы. 2-е изд. Л.: Гидрометеиздат, 1984. – 687 с.
2. Матвеев М.Г. Параметрическая идентификация моделей векторной авторегрессии // Современная экономика: проблемы и решения, ВГУ, № 5 (5), 2010. С. 133–141.
3. Матвеев М.Г., Михайлов В.В. Управление организационно-технической системой в условиях метеорологической неопределенности: монография. – Воронеж: изд-во ВВАИУ, 2006. – 147 с.
4. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с.
5. Tacagi T., Sugeno M. Fuzzy identification of systems and its application to modeling and control // IEEE Trans. Systems, Man and Cybernetics, 1985, № 15 (1), pp. 116–132.

**Matveev Mikhail G.** – doctor of technical sciences, professor department of programming and information technologies, Voronezh State University. Tel. (473) 220-84-70. E-mail: mgmatveev@yandex.ru

**Михайлов Владимир Владимирович** – д.т.н.,  
начальник факультета метеорологии, Воронеж-  
ский военный авиационный инженерный уни-  
верситет. Тел. (473) 226-65-06.

**Семенов Михаил Евгеньевич** – д.ф.-м.н.,  
профессор кафедры теоретической гидромете-  
орологии, Воронежский военный авиационный  
инженерный университет. Тел. (473) 226-13-66.  
E-mail: mk150@mail.ru

**Mikhailov Vladimir V.** – doctor of technical  
sciences, the chief of faculty meteorology,  
Voronezh military aviation engineering university.  
Tel. (473) 226-65-06.

**Semenov Mikhail E.** – doctor of physics-math.  
sciences, professor department of theoretical  
hydrometeorology, Voronezh military aviation  
engineering university. Tel. (473) 226-13-66.  
E-mail: mk150@mail.ru