

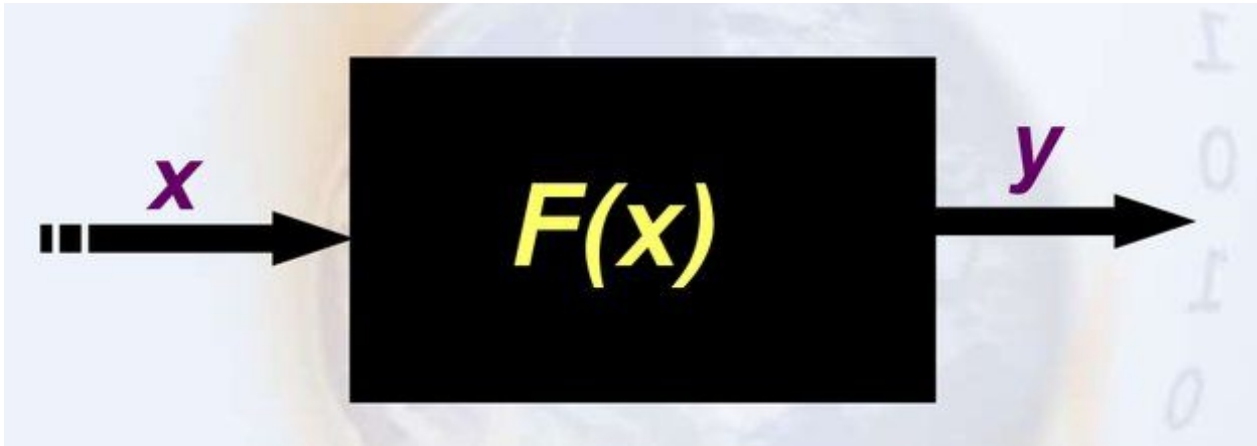
Моделирование

Лекция 2

2018

Построение модели

$$E = \Phi(y, x, a, \xi)$$



Два основных этапа моделирования

1. Построение модели
2. Применение модели

Две цели моделирования

1. Для анализа (напр. в задаче распознавания образов)
2. Для синтеза (напр. в задаче оптимизации)

Этапы построения модели

1. ТСиСА: уяснение природы исследуемой системы, входов и выходов. Содержательное описание системы и ее архитектуры. Можно исследовать всю систему, а можно отдельный объект системы (входы, выходы, внутреннюю структуру, процесс, поведение). **Цель моделирования формулируется в отношении объекта исследования;**

2. Формализация

- **формальное описание входов, выходов объекта исследования** (количественное или качественное моделирование входов и выходов), в зависимости от целей моделирования и исследовательских вопросов,
- **формальное описание зависимости выходов от входов (модель $Y=H(X)$),**
- **формальное описание зависимости переменных от времени (модель $Y=F(T)$),**
- **формальное описание критериев эффективности и целевой функции** в зависимости от целей моделирования;

Цели моделирования

1. Модели анализа. Получить новое знание о свойствах моделируемой системы:
 1. О природе ее входов и выходов
 2. О зависимостях, отражающих поведение системы
 3. О зависимостях в структуре элементов и процессов
2. Модели синтеза. Получить новое знание, улучшающее моделируемую систему - задача типа «Как сделать, чтобы...»
 1. Оптимальные значения параметров системы
 2. Оптимальную структуру системы

Таким образом, в качестве объектов моделирования в системах могут быть рассмотрены

- статические сущности (свойства, классы),
- динамические сущности (тенденции, поведение) и
- сущности, выражающие зависимости, в виде семантически различных структур.

Исследовательские вопросы

1. О природе и однородности входных и выходных данных
2. О методике и параметрах кластеризации
3. О результатах кластеризации
4. Об объектах в кластерах
 - В чем они схожи
 - В чем они различны



- Известно, что задачи анализа развития сложных слабоструктурированных процессов по временным рядам успешно решаются экспертами на основе предметно-ориентированных знаний и экспертных шкал.

Количественные модели входов и выходов

- Детерминированные:
 - В виде одного значения
 - в виде векторов значений одного типа
 - В виде матрицы значений одного типа
- Вероятностные, допускающие стохастическую неопределенность
 - В виде функции распределения вероятности для совокупности наблюдений (выборки из генеральной совокупности наблюдений) и ее характеристик:
 - среднее, медиана,
 - размах
 - вариация (среднеквадратическое отклонение или дисперсия)
 - В виде стохастических временных рядов, цифровых сигналов

Количественные модели входов и выходов

- Нечеткие, допускающие лингвистическую неопределенность
 - В виде лингвистической переменной для совокупности наблюдений X и ее характеристик:
 - количество термов,
 - виды функции распределения нечеткости термов и
 - ее носители в виде интервалов.
 - В виде нечетких временных рядов
- С неизвестной природой неопределенности в значениях
 - В виде множества однотипных значений, сформированных в процессе эволюции и адаптации к внешней среде

Трансформация – преобразование данных

- Число-Число: Стандартизация (standardization) или нормирование (normalization) приводит значения всех переменных к единому диапазону значений.
- Число-Число: Фильтрация(сглаживание) сигнала для получения сигнала с заданными свойствами.

Трансформация данных

- Categorical to Numeric Data: Binarization
- Text to Numeric Data
- Time Series to Discrete Sequence Data
- Time Series to Numeric Data
- Time Series to Text Data
- Text to DataBase
- Text to Time series
- Discrete Sequence to Numeric Data
- Graphs to Numeric Data

Стандартизация – приведение к единой шкале!

- деление исходных данных на среднеквадратичное отклонение соответствующих переменных;
- вычисление Z-вклада или стандартизованного вклада
- Логарифмирование
- Оценивание по специальной шкале

Трансформация

- Число – текст. Применяется теория нечетких множеств и нечетких шкал, например, ACL-шкала.
- Текст – число. Используются частоты события и вычисления его веса, например алгоритм TFIDF (Term Frequency Inverse Document Frequency)

Трансформация

- Последовательность чисел (временной ряд) – текст. Применяется теория нечетких тенденций и темпоральный анализ.
- Последовательность чисел – число. Применяется усреднение и выявление усредненных характеристик и среднеквадратичных отклонений

Стохастические модели процессов

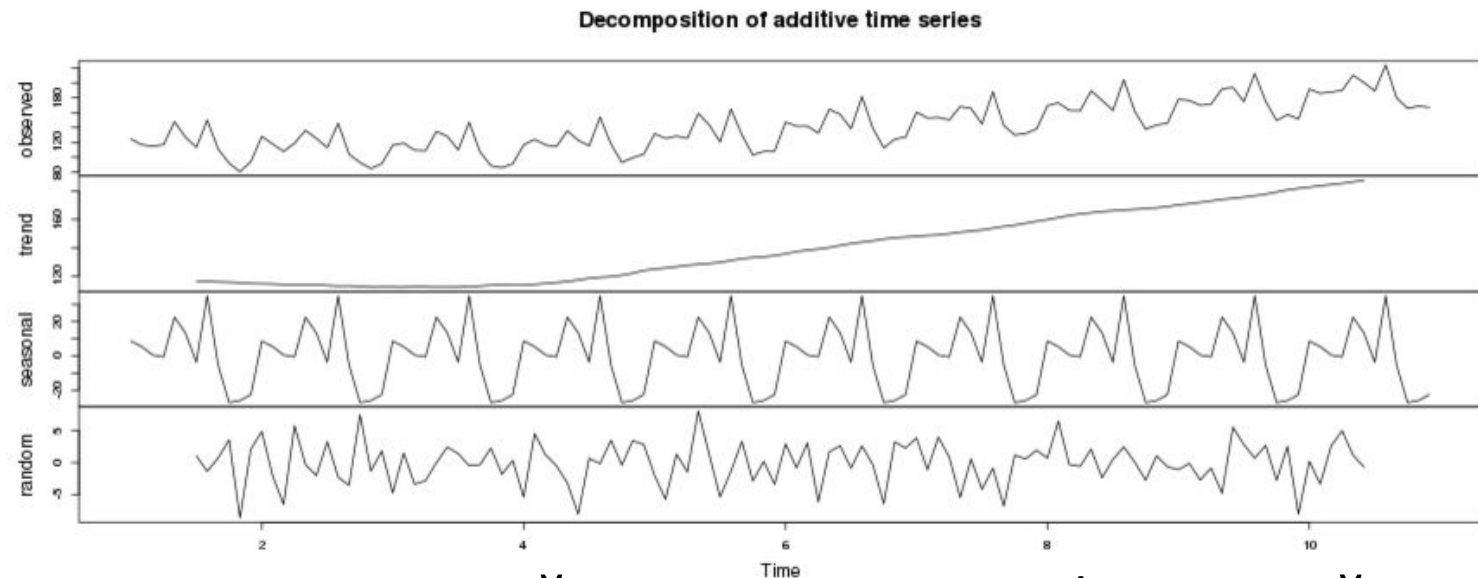
- Стохастические процессы, задающие зависимость между четким временем и случайной величиной с известным законом распределения вероятностей.
- Стохастические процессы указанного вида описывают поведение процесса в условиях риска и неопределенности, то есть случайные и недетерминированные ее изменения.

Модель временного ряда

$$y_t = f_t + \psi_t + \xi_t$$

- y_t – заданный временной ряд, f_t – тренд-цикл временного ряда, ψ_t – сезонность, ξ_t – случайная компонента.
- Модель тренда :
 1. $y_t = f(t) + u_t$, $y_t = at + u_t$ [Андерсон, 1976]

Анализ
поведения
путем
декомпозиции
и ВР



Динамические свойства выхода Y (верхний график):

- Наличие восходящего тренда
- Наличие сезонности

$$\begin{cases} x(t) = x_{tr}(t) + s(t), \\ s(t) = x_{sea}(t) + a(t) \\ a(t) = x_{ARMA}(t) + g(t), \\ g(t) = x_{GARCH}(t) + e(t); \end{cases} \quad t = t_1, \dots, t_n$$

Нечеткие модели входов и выходов

Л. Заде предложил по аналогии с теорией вероятности использовать функцию в качестве математической модели лингвистической неопределенности

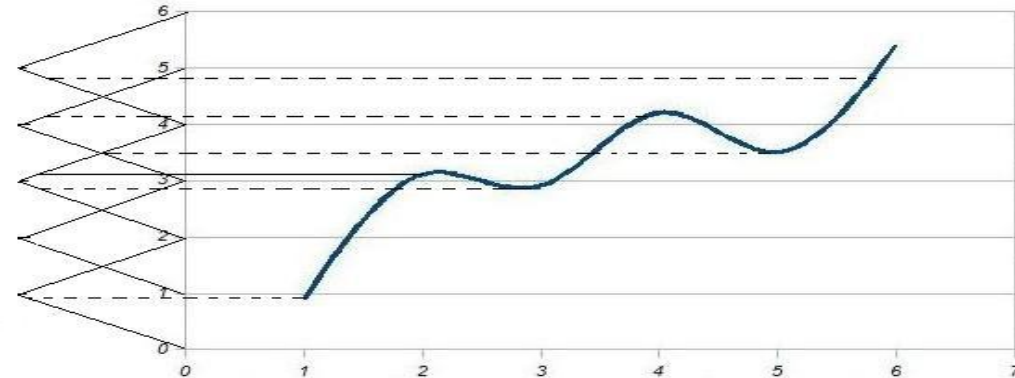
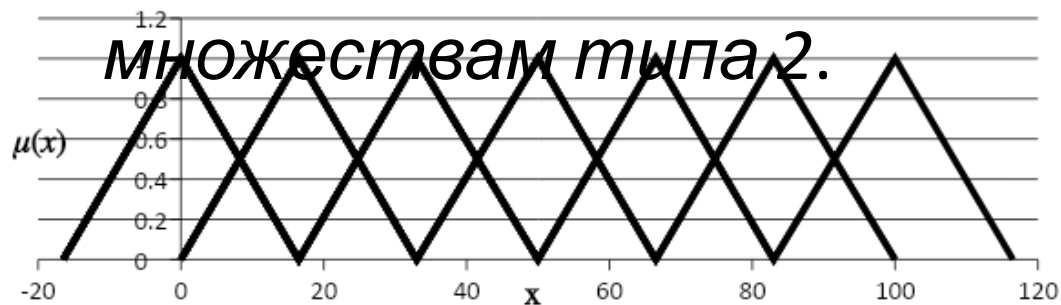
$$Y = \mu(x, B),$$

где Y – результат вычисления функции, выражающий меру неопределенности (нечеткости) для конкретного объекта X .

μ – непрерывная функция

$$\underline{\underline{\mu}}: X \rightarrow [0, 1].$$

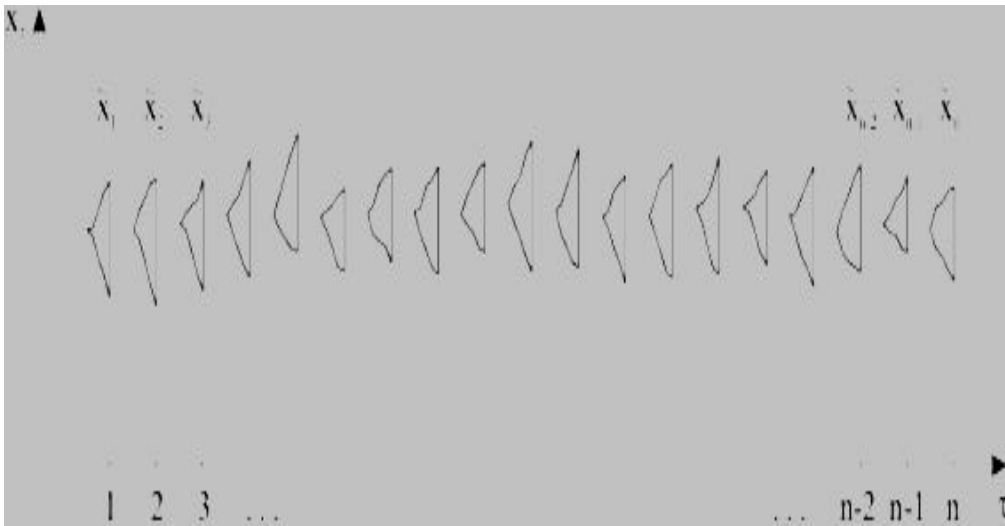
- В том случае, если значения функции принадлежности нечеткого множества представлены точными числовыми значениями, такие нечеткие множества относят к *нечетким множествам типа 1*.
- Если значения функции принадлежности нечеткого множества моделируются другими нечеткими множествами, то такое нечеткое множество относят к *нечетким множествам типа 2*.



Нечеткий временной ряд как модель процесса

$$\tilde{x}_i = \textit{Fuzzy}(\mu_{\tilde{x}_i}(w), x_i)$$

- Предположим, что задан процесс, состояния которого описываются n значениями одной переменной.
- В результате наблюдения получен временной ряд этой переменной, он преобразован в нечеткий



Нечеткие процессы

1. Нечеткие процессы. Нечеткие процессы представляют группу зависимостей между категориями x_i , t_i (время – значение), каждая из которых может быть задана нечеткими значениями. Рассмотрим их как разные классы отношений «время» – «значение».

а. Отношение «Четкое время t » – «Нечеткая переменная $\tilde{x}$ ».

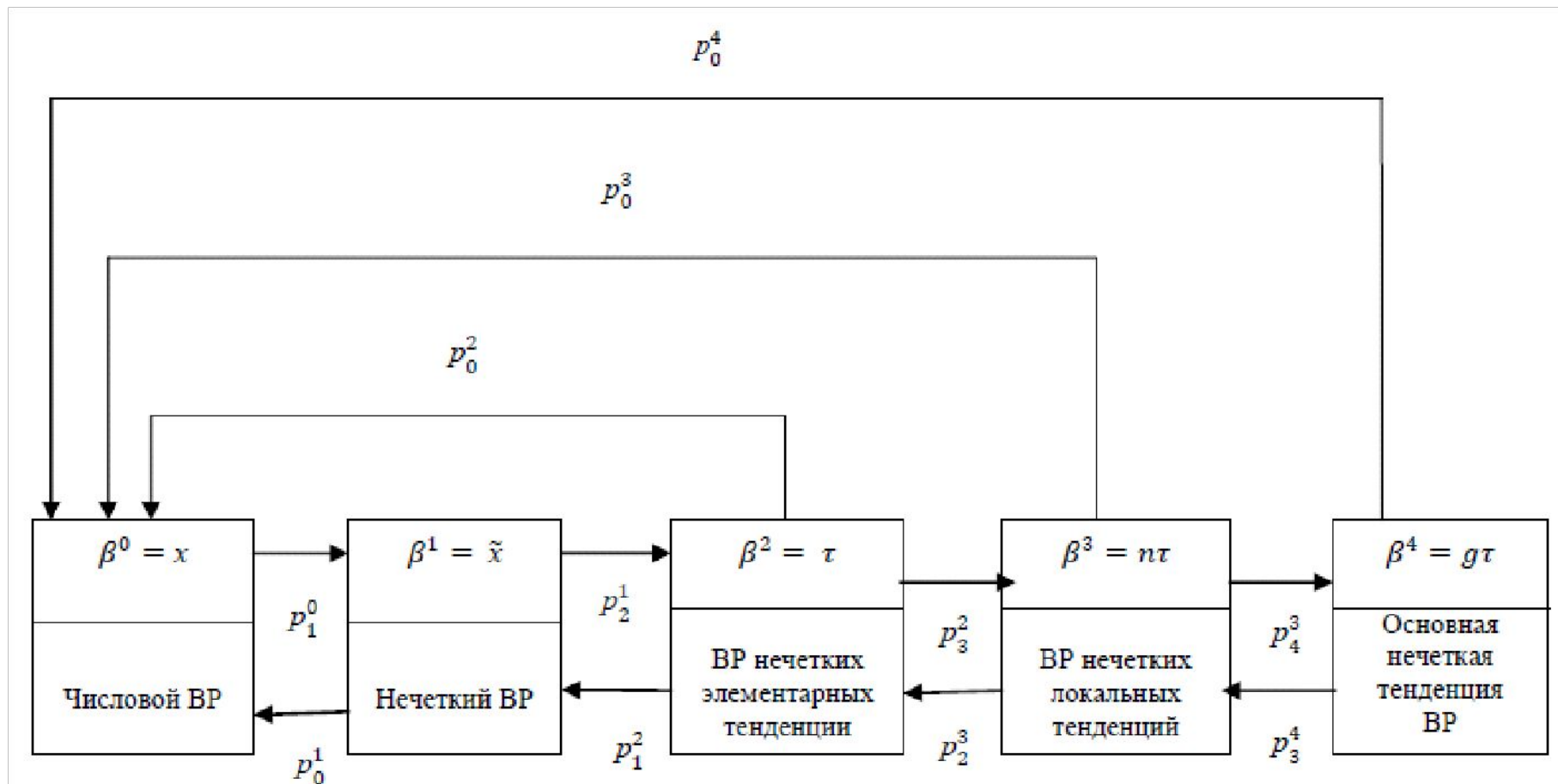
В этом отношении каждое нечеткое значение $\tilde{x}_i$ представляется нечетким термом, определенным на некотором базовом множестве и функцией принадлежности, выражающей количественную зависимость между элементами базового множества и нечетким значением $\tilde{x}_i$. Это отношение соответствует представлению нечеткого ВР в виде $\tilde{x}_i = Fuzzy(x_i, t_i)$.

Нечеткие процессы

а. Отношение «Нечеткое время $\tilde{t}$ » – «четкая переменная x ». В этом случае можно предположить, что четкое значение времени было предварительно преобразовано в нечеткие значения $\tilde{t}_i = Fuzzy_t(x_i, t_i)$. Идентификация нечетких тенденций для указанного отношения может проводиться двояким образом: по отношению к нечеткому времени $\tau_i = Tend(\tilde{x}_i, \tilde{t}_i)$ и по отношению к приближенному моменту времени, полученному в результате его дефаззификации $\tau_i = Tend(\tilde{x}_i, t'_i)$.

б. Отношение «Нечеткое время $\tilde{t}$ » – «нечеткая переменная $\tilde{x}$ ». Данное отношение может использовать без предварительных преобразований алгоритм $\tau_i = Tend(\tilde{x}_i, \tilde{t}_i)$.

Декомпозиция Нечеткого ВР



Модели
преобразования
объектов НВР

p_1^r	Вид перехода	Функционал	Результат вычисления
p_1^0	$r = 0 \rightarrow r = 1$	$Fuzzy: X \rightarrow \tilde{X}$	$\tilde{x}_t = Fuzzy(x_t)$
p_3^1	$r = 1 \rightarrow r = 0$	$deFuzzy: \tilde{X} \rightarrow X$	$x_t = deFuzzy(\tilde{x}_t)$
p_2^1	$r = 1 \rightarrow r = 2$	$TTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{V},$ $RTend: \tilde{X} \times \tilde{X} \rightarrow \tilde{A}$	$\tilde{v}_t = TTend(\tilde{x}_{t-1}, \tilde{x}_t)$ $\tilde{a}_t = RTend(\tilde{x}_{t-1}, \tilde{x}_t)$
p_1^2	$r = 2 \rightarrow r = 1$	$deFuzzyTend: \tilde{X} \times \tilde{\Xi} \rightarrow \tilde{X}$	$\tilde{x}_t = deFuzzyTend(\tilde{x}_{t-1}, \tilde{v}_t, \tilde{a}_t)$
p_3^2	$r = 2 \rightarrow r = 0$	$deFuzzy: \tilde{V} \rightarrow V,$ $deFuzzy: \tilde{A} \rightarrow A$ $deETend: \tilde{X} \times \tilde{\Xi} \rightarrow X$	$v_t = deFuzzy(\tilde{v}_t)$ $a_t = deFuzzy(\tilde{a}_t)$ $x_t = deETend(\tilde{x}_{t-1}, \tilde{v}_t, \tilde{a}_t)$
p_3^2	$r = 2 \rightarrow r = 3$	$STend: \tilde{\Xi} \times \tilde{\Xi} \rightarrow N\tilde{\Xi}$	$n\tau_s = STend(\tau_{1s}, \tau_{2s}, \dots, \tau_{ms})$
p_2^3	$r = 3 \rightarrow r = 2$	$deSTend:$ $N\tilde{\Xi} \rightarrow \tau_{1s}, \tau_{2s}, \dots, \tau_{ms}$	$(\tau_{1s}, \tau_{2s}, \dots, \tau_{ms}) = deSTend(n\tau_s)$
p_3^3	$r = 3 \rightarrow r = 0$	$deFuzzy: N\tilde{\Xi} \rightarrow X$	$x_t = deFuzzy(n\tau_s).$
p_4^3	$r = 3 \rightarrow r = 4$	$GTend: N\tilde{\Xi} \times N\tilde{\Xi} \rightarrow G\tilde{\Xi}$	$g\tilde{\tau} = GTend(n\tau_{1s}, n\tau_{2s}, \dots, n\tau_{ms})$
p_3^4	$r = 4 \rightarrow r = 0$	$deFuzzy: G\tilde{\Xi} \rightarrow X$	$x_t = deFuzzy(g\tilde{\tau}_q).$

Методологические основы нечеткого моделирования ВР

- Состояния динамического процесса, порождающего ВР, представляются нечеткими множествами, введенными Заде.
- Идентификация модели процесса в виде $f: A \times B$ на основе нечеткой импликации $R: A \rightarrow B$ и нечетких конечно-разностных уравнений (Song, Chissom, 1993)
- Выражение модели в форме базы нечетких правил «Если – То»
- Нечеткий логический вывод по базе нечетких правил, как численный алгоритм приближения нелинейных зависимостей

Теоретические основы нечеткого моделирования

- Теорема нечеткой аппроксимации (Kosko, 1992)
- Теорема аппроксимации любой непрерывной функции с произвольной точностью нечеткой моделью и нечетким логическим выводом (Vang, 1992), (Kastro, 1995)



$$\bullet \left\{ \begin{array}{l} R_1: \text{ЕСЛИ } x_1 \text{ есть } A_{11} \dots \text{И} \dots x_n \text{ есть } A_{1n}, \text{ТО } y \text{ есть } \\ R_j: \text{ЕСЛИ } x_1 \text{ есть } A_{j1} \dots \text{И} \dots x_n \text{ есть } A_{jn}, \text{ТО } y \text{ есть } \\ \dots \\ R_m: \text{ЕСЛИ } x_1 \text{ есть } A_{m1} \dots \text{И} \dots x_n \text{ есть } A_{mn}, \text{ТО } y \text{ есть } \end{array} \right.$$

Меры точности моделей ВР

Формула расчета
$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \cdot 100 \%$
$SMAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{(y_i + \hat{y}_i) / 2} \right \cdot 100 \%$

Пример. Определите цель моделирования

- **Объектом исследования** является сложный технический объект – сервер предприятия
- Исходными данными для вычислительных экспериментов послужил массив данных $\{x_{\text{сеть}}, x_{\text{цп}}, x_{\text{оп}}\}$, представляющий собой значения параметров сервера: загрузка сети, процент загрузки процессора и процент загрузки оперативной памяти, измеренных в течение 90 дней.

Обобщенная регрессионно-нечеткая модель сервера

$$\tilde{S} = F(\tilde{Z}, A, \tilde{E})$$

● $\tilde{S}$ - лингвистическая оценка состояния сервера; $\tilde{E}$ - лингвистическая экспертная оценка состояния предприятия, являющаяся для сервера внешним фактором; $\tilde{Z}$ - лингвистическая оценка состояния отдельных характеристик сервера; A - набор правил для определения состояния сервера.

$$\tilde{Z} = F(\tilde{V}, R)$$

R - набор правил для определения состояния отдельных характеристик сервера, $\tilde{V}$ - лингвистическая оценка локальной тенденции характеристик сервера.

$$\tilde{V} = GTend(TTend(\tilde{X}))$$

$\tilde{X} = \{\tilde{x}_i\}, i = \overline{1, n}$ - множество нечетких значений характеристик сервера, получаемых по формуле:

$$\tilde{x}_i(t) = Fuzzy(x_i(t))$$

x_i - прогнозные значения характеристик сервера, получаемые по модели:

$$x_i(t) = f(t) + \varphi(t) + e(t)$$

- Нечеткое шкалирование – построение ACL-шкалы для каждой из характеристик.
 - $C_x = \{\tilde{X}, \tilde{V}, Fuzzy, TTend, GTend\}$
 - $\tilde{X} = \{\text{Хороший, Нормальный, Плохой}\}$
 - $\tilde{V} = \{\text{Падение, Стабильность, Рост}\}$
- Фаззификация прогнозов характеристик сервера.
 - $\tilde{x}_i(t) = Fuzzy(\hat{x}_i(t))$

Введены три лингвистические переменные

Нечеткие множества		Нечеткие тенденции		Нечеткие интенсивности	
1	Очень низкий	1	Рост	1	Нулевой
2	Низкий	2	Падение	2	Очень малый
3	Ниже среднего	3	Стабильность	3	Малый
4	Средний			4	Средний
5	Выше среднего			5	Больше среднего
6	Высокий			6	Большой
7	Очень высокий			7	Очень большой

Результаты фаззификации значений выходных характеристик сервера

№	Значение	Степень принадлежности	Нечеткое значение	Нечеткая тенденция	Интенсивность
82	77,3	0,76041	Выше среднего	Рост	Очень малый
83	54,71	0,79087	Низкий	Падение	Средний
84	76,438	0,89196	Выше среднего	Рост	Средний
85	68,159	0,84419	Средний	Стабильность	Очень малый
86	64,029	0,78638	Ниже среднего	Падение	Очень малый
87	73,36	0,6382	Выше среднего	Рост	Малый
88	67,396	0,72763	Средний	Падение	Очень малый
89	65,366	0,58223	Ниже среднего	Стабильность	Очень малый
90	71,681	0,61818	Средний	Рост	Очень малый

Правила нечеткой модели

- $$\left\{ \begin{array}{l} R_1: \text{Если } \tilde{x}_i(t) \text{ есть } \tilde{X}_{i,1} \text{ и } \tilde{v}_{Gt,i} \text{ есть } \tilde{V}_{i,1}, \text{ то } z_i \text{ есть } \tilde{Z}_1 \\ R_2: \text{Если } \tilde{x}_i(t) \text{ есть } \tilde{X}_{i,2} \text{ и } \tilde{v}_{Gt,i} \text{ есть } \tilde{V}_{i,2}, \text{ то } z_i \text{ есть } \tilde{Z}_2 \\ \dots \\ R_k: \text{Если } \tilde{x}_i(t) \text{ есть } \tilde{X}_{i,k} \text{ и } \tilde{v}_{Gt,i} \text{ есть } \tilde{V}_{i,k}, \text{ то } z_i \text{ есть } \tilde{Z}_k \end{array} \right.$$

Нечеткое моделирование

Для всех технических параметров построены правила нечеткого вывода:

	Нечеткое значение	Общая тенденция	Решение по параметру
	Низкий	Падение	Недостаток загрузки
	Низкий	Рост	Норма
	Средний	Стабильность	Норма
	...	...	...
	Высокий	Рост	Превышение загрузки

Сформированы правила нечеткого вывода для прогноза состояния сервера

	Загрузка ЦП	Загрузка ОП	Загрузка по трафику	Экспертная оценка нагрузки	Состояние сервера
	Норма	Норма	Норма	Стабильность	Норма
	...	...	...	...	...
	Норма	Недостаток загрузки	Превышение нагрузки	Стабильность	Возможно превышение
	Норма	Недостаток загрузки	Превышение нагрузки	Снижение нагрузки	Норма
	...	...	...	...	...
	Недостаток загрузки	Норма	Норма	Снижение нагрузки	Возможен простой ресурсов

№	Загрузка ЦП	Загрузка ОП	Загрузка сети
82	Рост	Стабильность	Рост
83	Падение	Стабильность	Падение
84	Рост	Стабильность	Рост
85	Стабильность	Падение	Стабильность
86	Падение	Стабильность	Падение
87	Рост	Стабильность	Рост
88	Падение	Рост	Падение
89	Падение	Падение	Стабильность
90	Рост	Стабильность	Рост
ОТ	Рост	Стабильность	Стабильность

- Нечеткое моделирование всего сервера на основе предложенной методики

A_m : Если $\tilde{z}_1$ есть $\tilde{Z}$ и $\tilde{z}_2$ есть $\tilde{Z}$ и ... $\tilde{z}_n$ есть $\tilde{Z}$ и $E_{\text{пр}}$ есть $\tilde{E}$, то S есть $\tilde{S}$
 $\tilde{E} = \{\text{Уменьшение нагрузки, Стабильность, Увеличение нагрузки}\}$
 $\tilde{S} = \{\text{Простой, Норма, Превышение}\}$

						
Загрузка ЦП	Выше среднего	0,971	Рост	0,925	Превышение нагрузки	0,925
Загрузка ОП	Выше среднего	0,729	Рост	0,957	Превышение нагрузки	0,729
Загрузка сети	Выше среднего	0,618	Стабильность	0,985	Норма	0,618

	Загрузка ЦП	Загрузка ОП	Загрузка сети	Экспертная оценка нагрузки	Состояние сервера
Лингвистическая оценка	Превышение нагрузки	Превышение нагрузки	Норма	Рост	Превышение нагрузки
Степень принадлежности	0,925	0,729	0,618	1	0,618