

ФОРМИРОВАНИЕ БАЗЫ ЗНАНИЙ НА ОСНОВЕ ВЫДЕЛЕНИЯ ТИПОВЫХ СОСТОЯНИЙ СЛОЖНОЙ СИСТЕМЫ

© 2020 Т. М. Леденева[✉], М. А. Сергиенко, Е. А. Тихомирова

*Воронежский государственный университет
Университетская пл., 1, 394018 Воронеж, Российская Федерация*

Аннотация. В данной статье представлен подход для формирования базы знаний, описывающей поведение сложной системы. Для того, чтобы описать это поведение вводится система показателей. Предполагается, что в результате их наблюдения формируются временные ряды. На основе кусочно-линейной аппроксимации выделяются такие временные промежутки, внутри которых линейные тренды временных рядов не изменяются. Данные промежутки определяют некоторое состояние сложной системы. Для формального описания состояний используются кодовые векторы, которые формируются на основе лингвистической шкалы. Ее градации определяют базисные направления линейных трендов. Каждому базисному направлению соответствует целочисленный код. Близость угла наклона линейного тренда к базисному направлению определяется с помощью функции принадлежности. Для выделения типовых состояний предлагается использовать кластерную процедуру. Анализ подходящих методов позволил выделить в качестве такой процедуры метод декомпозиционного дерева. Его преимуществом является то, что он позволяет сгенерировать все возможные разбиения заданного множества состояний. На данном этапе возникает проблема выбора оптимального разбиения. В данной статье под оптимальным подразумевается такое разбиение, которое содержит как можно больше классов, встречающихся в декомпозиционном дереве. Такие классы проявляют устойчивость в некотором смысле. Оптимальному разбиению соответствует определенный уровень декомпозиционного дерева, а классам разбиения — типовые состояния сложной системы. В рамках предположения, что показатели системы зависят от некоторого множества факторов, формируется база продукционных правил. Заключение данных правил содержат термы или функции, которые соответствуют факторам. Предложенный подход апробирован в среде FuzzyClips для анализа инвестиционного портфеля.

Ключевые слова: временной ряд, линейный тренд, типовое состояние, продукционное правило.

ВВЕДЕНИЕ

Разработка интеллектуальных информационных систем и экспертных систем обуславливает актуальность моделей и методов, позволяющих на основе анализа наблюдаемых данных в форме временного ряда построить компоненты знаний для принятия решений в сложных системах. Задача моделирования временного ряда заключается в выявлении зависимости соответствующего

ему показателя или параметра от времени на основе данных наблюдения, при этом предполагается, что данная зависимость сохранится на ограниченном отрезке времени в будущем. В настоящее время можно выделить несколько основных подходов к моделированию временных рядов. Статистический подход [1] основывается на восстановлении зависимости в форме статистической модели, которая включает систематическую и случайную составляющие, при этом систематическая составляющая в общем случае является комбинацией трендовой, периодической и сезонной компонент, а также (при необходимости)

✉ Леденева Татьяна Михайловна
ledeneva-tm@yandex.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

авторегрессионной компоненты. Моделирование временных рядов в рамках нейросетевого подхода сводится к задаче наилучшей аппроксимации нелинейной функции от многих переменных, параметрической моделью которой служит нейронная сеть [2–3]. Возможность обучения — одно из главных преимуществ нейронных сетей, причем в процессе обучения нейронная сеть способна выявлять сложные нелинейные зависимости и выполнять обобщение. Нечеткое моделирование временных рядов базируется на приближенном описании поведения временного ряда с помощью лингвистических переменных и *если-то-правил* (продукционных правил), описывающих зависимость соответствующего показателя от времени [4]. В [5] предложен подход для формирования различных типов продукционных правил на основе кластеризации наблюдаемых данных.

Цель статьи заключается в представлении подхода для формирования базы знаний нечеткой системы прогнозирования состояния сложной системы на основе временных рядов, описывающих ее поведение на заданном временном промежутке.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть в моменты времени $t = \underline{0}, \dots, T$ наблюдаются n показателей P_i ($i = \overline{1, n}$), характеризующих состояние (или функционирование) некоторой сложной системы, и тогда в промежутке $[0, \dots, T]$ поведение системы в каждый момент времени описывается совокупностью из n временных рядов. Предположим, что на основе кусочно-линейной аппроксимации можно выделить сегменты временных рядов, которым соответствуют линейные тренды. На $[0, T]$ определим точки τ_i , в которых хотя бы для одного временного ряда изменяется направление тренда. В результате получим промежутки $T_i = [\tau_i, \tau_{i+1}]$ ($i = \overline{1, m}$), которые образуют разбиение области наблюдения $[0, T]$, при этом на каждом промежутке направление любого тренда не изменяется. Под *состоянием* сложной системы S_i ($i = \overline{1, m}$) будем понимать совокупность линейных трендов на промежутке T_i . Заме-

тим, что таких состояний может быть достаточно много, поэтому целесообразно выделить некоторые типовые состояния, которые предполагают принятие управленческих решений определенного типа. Формирование множества типовых состояний можно осуществить различными способами, например, «объединяя» в одно типовое состояние несколько близких по поведению показателей состояний, или привлекая экспертные оценки. Если множество типовых состояний определено, то, оценивая конкретную ситуацию текущего момента, можно построить прогнозную оценку поведения показателей системы или других характеристик, являющихся функциями от этих показателей. Таким образом, для реализации подхода необходимо решить следующие проблемы: разработать способ для описания состояний и типовых состояний сложной системы; разработать процедуру для кластеризации/классификации состояний системы на промежутке $[0, T]$ с целью выявления типовых состояний. Для прогнозирования состояний сложной системы можно использовать нечеткую продукционную систему, база знаний которой в форме совокупности *если-то-правил* описывает типовые состояния.

2. МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Дескрипторы для описания трендов и формирование кодового вектора

После кусочно-линейной аппроксимации временной ряд некоторого показателя P_i в промежутке $[0, T]$ представляется ломаной. Пусть $\tau_j^i \in [0, T]$ — точка, в которой направление тренда показателя P_i изменяется как-либо образом, $M_i = \{\tau_j^i\}_{j=\overline{1, n_i}}$ — множество таких точек для показателя P_i , n_i — их количество. $\bigcup_{i=1}^n M_i = \{T_1, \dots, T_N\}$ — это множество точек, расположенных в промежутке $[0, T]$, таких, что в каждой точке T_i тренд хотя бы одного из показателей изменяется. Заметим, что в промежутке $[T_i, T_{i+1}]$ направления трендов всех показателей неизменны.

Пусть α — угол наклона тренда. Сформируем оси D_i базисных направления трендов,

задав их соответствующими углами. Каждому базисному направлению поставим в соответствие код. В табл. 1 перечислены базисные направления трендов, соответствующие им углы φ_i угла наклона оси базисного направления D_i , а также коды базисных направлений.

Таблица 1. Шкала линейных трендов
[Table 1. Linear trend scale]

Код	Угол наклона тренда	Градации
9	$4\pi/8 - \alpha_0$	Очень быстро возрастает (VQI)
8	$3\pi/8$	Быстро возрастает (QI)
7	$\pi/4$	Возрастает (I)
6	$\pi/8$	Медленно возрастает (SI)
5	0	Постоянно (C)
4	$-(\pi/8)$	Медленно убывает (SD)
3	$-(\pi/4)$	Убывает (D)
2	$-(3\pi/8)$	Быстро убывает (QD)
1	$-(4\pi/8 - \alpha_0)$	Очень быстро убывает (VQD)

На основе табл. 1 строится лингвистическая шкала переменной *Возрастание-Убывание* линейного тренда с учетом угла α . Данная шкала включает 9 термов в форме симметричных треугольных нечетких чисел с модальным значением, равным углу базисного направления, и коэффициентом неопределенности $\pi/8$. По сути, каждому терму соответствует базисное направление D_i , а сама лингвистическая шкала имеет вид

$$LingScale = \{VQI, QI, I, SI, C, SD, D, QD, VQD\}.$$

Зафиксируем промежуток T_i . Каждому показателю P_j ($j = 1, n$) на данном промежутке соответствует линейный тренд LT_j^i с углом наклона $\varphi_{LT_j^i}$. Для определения базисного направления, которому в наибольшей степени

соответствует линейный тренд LT_j^i , введем функцию принадлежности, при этом ее тип должен быть таким, чтобы учитывать отклонение угла $\varphi_{LT_j^i}$ от угла, который соответствует базисному направлению. Подходящие типы таких функций принадлежности перечислены, например, в [6]. Тогда анализируя полученные значения функций принадлежности, припишем показателю P_j базисное направление D_{k^*} , которое соответствует индексу $k^* = \arg \max_k \left\{ \mu_{D_k} \left(\varphi_{LT_j^i} \right) \right\}$.

Так как каждому базисному направлению соответствует некоторый код, то каждому состоянию S_i системы на промежутке T_i можно поставить в соответствие вектор $CodV_i = (c_j^i)_{j=1,n}$, где c_j^i — код базисного направления для тренда LT_j^i на промежутке T_i . Данный вектор будем называть *кодовым*.

2.1. Выявление типовых состояний

В промежутке $[0, T]$ эволюция системы описывается множеством кодовых векторов $\{CodV_i\}_{i=1,m}$, некоторые из которых могут оказаться близкими в некотором смысле. Объединяя в кластеры такие векторы, можно построить множество типовых состояний сложной системы. В настоящее время существует большое количество различных методов кластеризации, однако в данном случае подходящими являются метрические методы [7]. Одним из перспективных, на наш взгляд, является метод декомпозиционного дерева, идея которого изложена в [6], а затем развита в [8, 9]. Основными преимуществами метода декомпозиционного дерева являются следующие:

- возможность получения множества различных разбиений, что позволяет за счет привлечения дополнительной информации о ситуации принятия решений или уточнения критериев оптимальности выбрать наиболее подходящий вариант разбиения;
- декомпозиционное дерево отражает последовательный процесс разбиения заданного множества объектов на классы, причем для каждого объекта можно проследить его эволюцию от ситуации, когда все объекты попадают в один класс, до ситуации, когда данный объект образует тривиальный класс;

– каждому уровню можно сопоставить определенный порог классификации, а за счет анализа получившегося разбиения подобрать оптимальное значение порога;

– использование параметрических нечетких операций позволяет осуществить настройку метода на конкретного пользователя или конкретную проблему.

Основываясь на [8], изложим основные шаги метода декомпозиционного дерева.

1. Выбрав подходящую функцию расстояния и определяя ее значения для каждой пары объектов, построить матрицу расстояний, которая определяет нечеткое антирефлексивное и симметричное отношение R , которое является несходством.

2. От отношения несходства R перейти к отношению сходства \bar{R} с помощью стандартного отрицания, при этом данное отношение обладает рефлексивностью и симметричностью.

3. Определить транзитивное замыкание \hat{R} отношения сходства, при этом рефлексивность и симметричность сохраняются, но добавится транзитивность, поэтому \hat{R} является подобием или нечетким отношением эквивалентности.

4. Построить декомпозицию отношения подобия \hat{R} на обычные отношения эквивалентности \hat{R}_α , где $\alpha \in (0, 1]$, а затем для каждого отношения эквивалентности найти классы эквивалентности, образующие соответствующие разбиения.

5. Построить декомпозиционное дерево, отражающее процесс разбиения исходного множества на все более мелкие классы, вплоть до тривиальных, включающих единственный элемент.

Для оценки близости векторов при формировании отношения несходства можно использовать стандартные функции расстояния, которые обобщаются формулой

$$\rho(x^i, x^j) = \sqrt[p]{\left(\sum_{k=1}^n |x_k^i - x_k^j|^p\right)},$$

где x^i, x^j — компоненты векторов, p — настраиваемый параметр.

Кроме функций расстояния для формирования отношения сходства/несходства ис-

пользуются также специальные функции подобия, индексы сходства/несходства. Транзитивное замыкание реализуется с помощью операции нечеткой композиции. В [9] исследуются свойства $(\max - T)$ - и $(\min - S)$ -композиций и соответствующие типы транзитивности. Здесь T и S — треугольные норма и конорма, которые моделируют соответственно операции пересечения и объединения нечетких множеств [6, 8].

Таким образом, применяя кластерную процедуру (метод декомпозиционного дерева), получим разбиение множества состояний системы на классы близких по поведению состояний. Для выбора подходящего разбиения используется дополнительная информация или привлекается эксперт.

Пусть на основе анализа поведения системы выявлены классы Cl_1, \dots, Cl_N , при этом каждому классу соответствует определенное поведение показателей P_1, \dots, P_n , а, следовательно, определенная совокупность кодовых векторов. Определим типовое состояние системы, которое соответствует классу $Cl_r = \{S_{i_1}, \dots, S_{i_{n_r}}\}$, где каждое состояние S_{i_k} определяется кодовым вектором $(c_1^{i_k}, \dots, c_n^{i_k})$, следующим образом:

$$Type_r = \left(\frac{1}{n_r} \sum_{l=1}^{n_r} c_1^{i_l}, \dots, \frac{1}{n_r} \sum_{l=1}^{n_r} c_n^{i_l} \right),$$

т. е. типовое состояние определяется некоторым «усредненным» поведением системы в данном классе по каждому показателю.

Предположим, что существует универсальное множество параметров $Z = \{Z_j\}_{j=1, \dots, h}$, от которых зависят наблюдаемые показатели P_1, \dots, P_n , причем каждый показатель в зависимости от поведения зависит от «своего» набора параметров. Пусть на временном промежутке T_i показатель P_j зависит от множества параметров $Z_{ij} \subseteq Z$, причем эта зависимость с учетом типа поведения описывается функцией $\varphi_{ij}(c_j^i, z_{ij})$, где z_{ij} — перечень переменных, соответствующих параметрам из множества Z_{ij} . В частном случае функция φ_{ij} может быть константой.

Пусть типовое состояние системы определяется классом $Cl_r = (S_{i_1}, \dots, S_{i_{n_r}})$, где состоянию S_{i_k} соответствует кодовый вектор $(c_1^{i_k}, \dots, c_n^{i_k})$, $c_j^{i_k}$ — код, определяющий поведе-

ние показателя P_j на промежутке T_{i_k} , причем данному поведению соответствует функция регрессии $\varphi_{i_k j}(c_j^{i_k}, z_{i_k j})$. Функцию для класса Cl_r определим с помощью операции среднего арифметического в виде

$$Val_r = \left(\underbrace{\frac{1}{n_r} \sum_{k=1}^{n_r} \varphi_{i_k 1}}_{V_{r1}}, \dots, \underbrace{\frac{1}{n_r} \sum_{l=1}^{n_r} \varphi_{i_k n}}_{V_m} \right) = (V_{r1}, \dots, V_m).$$

Заметим, что помимо среднего арифметического могут использоваться другие функции осреднения, в том числе, операции взвешенного агрегирования [10].

Таким образом, каждому типовому состоянию системы, которое определяется кластером Cl_r , будет поставлен в соответствие вектор $Val_r = (V_{r1}, \dots, V_m)$.

Заметим, что каждое типовое состояние системы позволяет учитывать весь комплекс факторов, действующих на совокупность рассматриваемых показателей.

3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

3.1. Иллюстративный пример

Рассмотрим задачу формирования портфеля ценных бумаг. Портфель — это совокупность различных инвестиционных инструментов (акции, облигации, депозитные и сберегательные сертификаты и т. п.), которые собраны воедино для достижения конкретной инвестиционной цели вкладчика. Инвестиции в ценные бумаги всегда связаны с ответом на вопрос: вкладывать деньги в ценные бумаги с большей доходностью и большим риском, или довольствоваться меньшей доходностью, но и меньшим риском. Главная цель при формировании портфеля состоит в достижении оптимального сочетания между риском и доходностью. Иначе говоря, соответствующий набор инвестиционных инструментов призван снизить до минимума риск потерь для инвестора при заданном уровне доходности или обеспечить максимальную ожидаемую доходность при заданном уровне риска. Заметим, что компоненты портфеля определенным образом взаимодействуют, так что риски и доходности ценных бумаг могут изменяться

в зависимости от комбинации. Задача об оптимальном портфеле заключается в том, чтобы определить какая доля портфеля должна быть отведена каждой инвестиции так, чтобы ожидаемая доходность и уровень риска соответствовали целям инвестора.

3.1. Выявление типовых ситуаций и формирование фрагментов базы знаний

Рассмотрим данные, представленные временными рядами котировок акций трех компаний P_1 , P_2 и P_3 во временном промежутке 16.09.2015–16.09.2019. Кусочно-линейная аппроксимация временных рядов позволяет выявить линейные тренды на отдельных промежутках T_1, \dots, T_{21} (рис. 1).

С помощью шкалы, определяющей базисные направления, сопоставим каждому промежутку T_i вектор, компоненты которого определяют коды базисных направлений, в наибольшей степени соответствующие трендам P_1, P_2, P_3 . Кодовые векторы состояний представлены в табл. 2.

Для классификации состояний и выявления типовых состояний воспользуемся методом декомпозиционного дерева. Построим матрицу относительных евклидовых расстояний, которую будем рассматривать как матрицу нечеткого отношения несходства R , от которой перейдем к матрице сходства \bar{R} , используя функцию стандартного отрицания. Затем на основе операции максиминного транзитивного замыкания [8] построим матрицу отношения подобия \hat{R} . Фрагмент данной матрицы представлен в табл. 3.

В соответствии с теоремой декомпозиции [6] разложим матрицу \hat{R} на уровни для тех значений матрицы отношения подобия, которые встречаются в таблице, а именно $\alpha = 0.86, 0.88, 0.89, 0.9, 0.93, 0.95$. Заметим, что каждому α -срезу нечеткого отношения подобия будет соответствовать свое отношение эквивалентности, а, следовательно, и свое разбиение на классы эквивалентности. Данный процесс можно представить в виде декомпозиционного дерева (рис. 2), каждый уровень которого соответствует разбиению исходного множе-

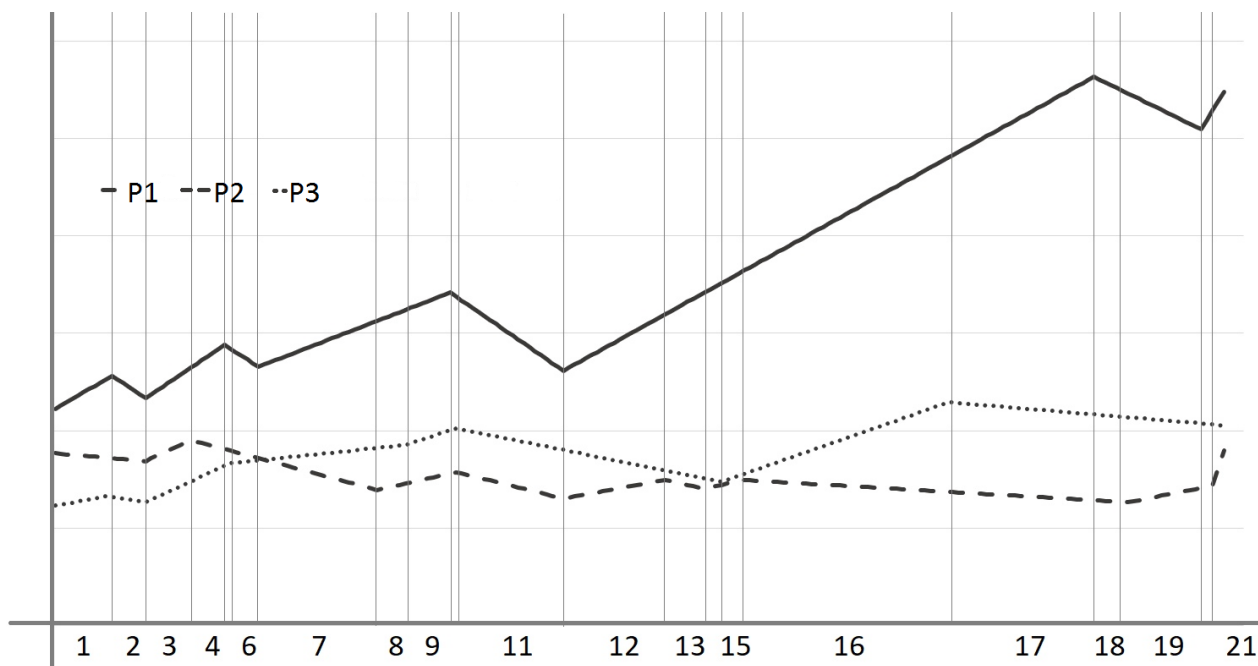


Рис. 1. Кусочно-линейная аппроксимация показателей P_1, P_2, P_3
 [Fig. 1 Piecewise linear approximation of indicators]

Таблица 2. Кодовые векторы
 [Table 2. Code vectors]

S_1	S_2	S_3	S_4	S_5	S_6	S_7
(7,5,6)	(3,5,4)	(7,6,7)	(7,4,7)	(3,4,7)	(3,4,6)	(6,4,6)
S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
(6,6,6)	(6,6,7)	(2,6,7)	(2,3,3)	(7,6,3)	(7,3,3)	(7,6,3)
S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}	S_{21}
(7,6,7)	(7,5,7)	(7,5,4)	(2,5,4)	(2,7,4)	(9,7,4)	(8,9,4)

ства состояний на классы состояний, близких по поведению показателей.

Под *оптимальным* будем подразумевать такое разбиение, которое содержит как можно больше классов, сохраняющихся при переходе от уровня к уровню, т. е. таких, которые в определенном смысле проявляют устойчивость. Построим таблицу, в которой для каждого класса Cl_i перечислим количество h_i включающих его уровней, при этом классы упорядочиваются по невозрастанию величин h_i . Двигаясь по таблице, попытаемся подобрать оптимальное разбиение, включая каждый раз такие классы, которые не изменяют уже сделанный выбор, при этом, находясь на фиксированном уровне целесообразно выбирать максимальный по мощности класс, содержащий фиксированный элемент.

С уменьшением значения h_i необходимо стремиться выбирать максимальный по мощности класс, содержащий элементы, еще не распределенные по классам. В табл. 4 представлен процесс выбора оптимального разбиения. После выбора классов 1–6 в таблице остаются классы 9, 10 и 15. На данном шаге целесообразно выбрать класс 15, поскольку остальные классы допускают объединение при минимальном значении $h_i = 1$.

Таким образом, оптимальным является разбиение: $\{10\}, \{11\}, \{13\}, \{20\}, \{21\}, \{5,6\}, \{19\}, \{2,18\}, \{12,14\}, \{7\}, \{17\}, \{1,3,4,8,9,15,16\}$. В декомпозиционном дереве данному разбиению соответствует $\alpha = 0.95$. Типовые ситуации, соответствующие каждому классу, представлены в табл. 5.

Таблица 3. Фрагмент матрицы нечеткого отношения подобия $S_1 - S_{11}$
 [Table 3. Fragment of matrix of a fuzzy similarity relation $S_1 - S_{11}$]

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
S_1	1	0.86	0.95	0.95	0.86	0.86	0.93	0.95	0.95	0.86
S_2	0.86	1	0.86	0.86	0.89	0.89	0.86	0.86	0.86	0.89
S_3	0.95	0.86	1	0.95	0.86	0.86	0.93	0.95	0.95	0.86
S_4	0.95	0.86	0.95	1	0.86	0.86	0.93	0.95	0.95	0.86
S_5	0.86	0.89	0.86	0.86	1	0.95	0.86	0.86	0.86	0.89
S_6	0.86	0.89	0.86	0.86	0.95	1	0.86	0.86	0.86	0.89
S_7	0.93	0.86	0.93	0.93	0.86	0.86	1	0.93	0.93	0.86
S_8	0.95	0.86	0.95	0.95	0.86	0.86	0.93	1	0.95	0.86
S_9	0.95	0.86	0.95	0.95	0.86	0.86	0.93	0.95	1	0.86
S_{10}	0.86	0.89	0.86	0.86	0.89	0.89	0.86	0.86	0.86	1
S_{11}	0.86	0.89	0.86	0.86	0.89	0.89	0.86	0.86	0.86	0.89
S_{12}	0.9	0.86	0.9	0.9	0.86	0.86	0.9	0.9	0.9	0.86

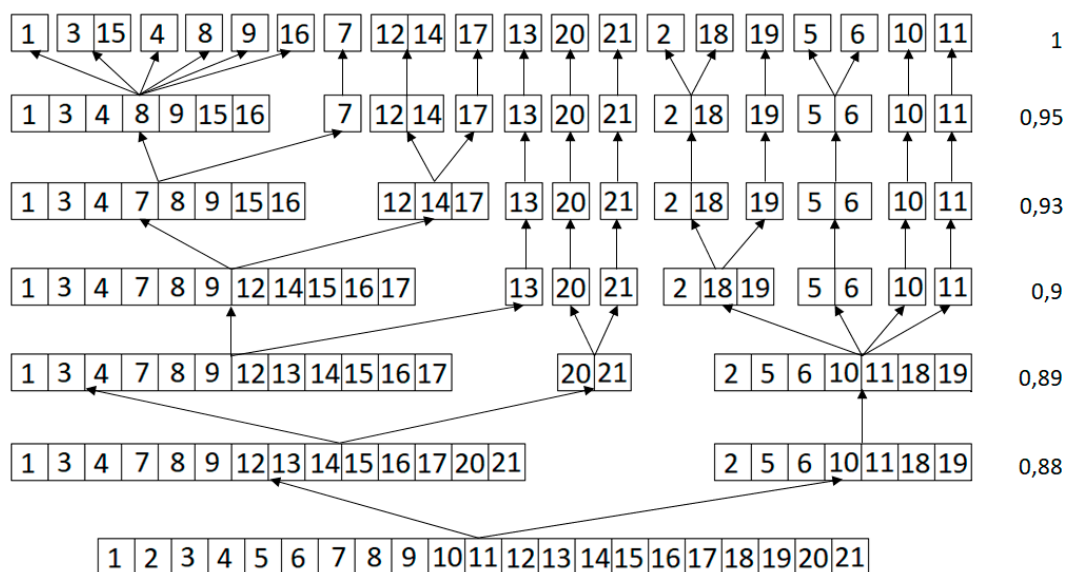


Рис. 2. Деконпозиционное дерево
 [Fig. 2. Decomposition tree]

Пусть каждому типовому состоянию Cl_r соответствует r -й инвестиционный портфель, характеризующийся следующими параметрами: x_j^r — долей инвестиций, помещенных в каждый из видов активов; m_j^r — ставкой доходности по каждому из активов. С каждым портфелем свяжем функцию полезности $u_r = \varphi(x^r, m^r, \dots)$, где x^r — вектор портфельных весов; m^r — вектор доходностей. В общем случае, с учетом вектора порт-

фельных весов x , полезность можно измерять различными способами: как внутреннюю норму доходности проекта с инвестициями x ; чистую приведенную стоимость проекта портфельных вложений x , приращение стоимости портфеля после дополнительного рефинансирования в размере x проекта; приращение нормы прибыли после дополнительного рефинансирования в размере x проекта; как степень достижения какой-либо цели в

Таблица 4. Иллюстрация процедуры выбора оптимального разбиения
 [Table 4. Illustration of the procedure for choosing the optimal partition]

№	Кластер Cl_i	h_i	Процесс выбора	
			Выбор/Запрет	Комментарий для запрета
1	{10}, {11}, {13}, {20}, {21}	4	+	
2	{5, 6}	3	+	
3	{19}	3	+	
4	{2, 18}	2	+	
5	{12, 14}	2	+	
6	{7}, {17}	2	+	
7	{2, 5, 6, 10, 11, 18, 19}	2	⊗	при выборе класса 1
8	{12, 14, 17}	1	⊗	при выборе класса 5
9	{3, 15}	1		
10	{1}, {4}, {8}, {9}, {16}	1		
11	{20, 21}	1	⊗	при выборе класса 1
12	{5}, {6}	1	⊗	при выборе класса 2
13	{2}, {18}	1	⊗	при выборе класса 4
14	{2, 18, 19}	1	⊗	при выборе класса 3
15	{1, 3, 4, 8, 9, 15, 16}	1	+	
16	{1, 3, 4, 7, 8, 9, 15, 16}	1	⊗	при выборе класса 6
17	{1, 3, 4, 7, 8, 9, 12, 14, 15, 16, 17}	1	⊗	при выборе класса 5
18	{1, 3, 4, 7, 8, 9, 12, 13, 14, 15, 16, 17}	1	⊗	при выборе класса 5
19	{1, 3, 4, 7, 8, 9, 12, 13, 14, 15, 16, 17, 20, 21}	1	⊗	при выборе класса 1

зависимости от инвестиций x . Характер роста функции полезности может быть использован для выяснения степени отношения инвестора к риску, а также для решения различных задач, связанных с анализом портфеля.

В нашем случае для каждого класса введем некоторую количественную характеристику полезности $\mu \in [0, 1]$ (табл. 6).

На основе данной информации может быть сгенерирована база знаний, которая содержит продукционные правила следующего вида:

R_i : **если** состояние портфеля есть Cl_i ,

то полезность есть μ_i .

Заметим, что выбор ретроспективных данных в качестве обучающего множества играет существенную роль для формирования множества типовых состояний и, в ко-

нечном итоге, для получения качественной базы знаний.

3.2. Описание реализации нечеткой продукционной системы

На основе полученной базы знаний с помощью FuzzyCLIPS [11] была разработана нечеткая продукционная система для оценки полезности инвестиционного портфеля. База знаний в этой системе представляется в виде двух компонент: база фактов, которая представляет исходное состояние проблемы; база правил, которая содержит набор нечетких продукционных правил, преобразующих факт в решение.

Машина логического вывода циклически сопоставляет факты с правилами и выясня-

Таблица 5. Описание типовых состояний
[Table 5. Description of typical states]

Класс	Кодовый вектор состояния	Описание состояния
$Cl_1 = \left\{ S_1, S_3, S_4, S_8, S_9, S_{15}, S_{16} \right\}$	$Type_1 = (7, 5, 7)$	P_1 возрастает и P_2 постоянно и P_3 возрастает
$Cl_2 = \{ S_{12}, S_{14} \}$	$Type_2 = (7, 6, 3)$	P_1 возрастает и P_2 медленно возрастает и P_3 убывает
$Cl_3 = \{ S_2, S_{18} \}$	$Type_3 = (3, 5, 4)$	P_1 убывает и P_2 постоянно и P_3 медленно убывает
$Cl_4 = \{ S_5, S_6 \}$	$Type_4 = (3, 4, 7)$	P_1 убывает и P_2 медленно убывает и P_3 возрастает
$Cl_5 = \{ S_7 \}$	$Type_5 = (6, 4, 6)$	P_1 медленно возрастает и P_2 медленно убывает и P_3 медленно возрастает
$Cl_6 = \{ S_{10} \}$	$Type_6 = (2, 6, 7)$	P_1 быстро убывает и P_2 медленно возрастает и P_3 возрастает
$Cl_7 = \{ S_{11} \}$	$Type_7 = (2, 3, 3)$	P_1 быстро убывает и P_2 убывает и P_3 убывает
$Cl_8 = \{ S_{13} \}$	$Type_8 = (7, 3, 3)$	P_1 возрастает и P_2 убывает и P_3 убывает
$Cl_9 = \{ S_{17} \}$	$Type_9 = (7, 5, 4)$	P_1 возрастает и P_2 постоянно и P_3 медленно убывает
$Cl_{10} = \{ S_{19} \}$	$Type_{10} = (2, 7, 4)$	P_1 быстро убывает и P_2 возрастает и P_3 медленно убывает
$Cl_{11} = \{ S_{20} \}$	$Type_{11} = (9, 7, 4)$	P_1 очень быстро возрастает и P_2 возрастает и P_3 медленно убывает
$Cl_{12} = \{ S_{21} \}$	$Type_{12} = (8, 9, 4)$	P_1 быстро возрастает и P_2 очень быстро возрастает и P_3 медленно убывает

Таблица 6. Полезность портфеля для типовых состояний
[Table 6. Portfolio utility for typical states]

Cl_1	Cl_2	Cl_3	Cl_4	Cl_5	Cl_6
0.11	0.89	0.62	0.099	0.02	0.012
Cl_7	Cl_8	Cl_9	Cl_{10}	Cl_{11}	Cl_{12}
0.206	0.019	0.153	0.075	0.14	0.135

ет, какие из правил можно активировать, при этом каждая итерация включает следующие шаги: сопоставление фактов и правил; выбор правила, подлежащего активации; выполнение действий, предписанных правилом.

В качестве входного значения для нечеткой системы выступает инвестиционный портфель с оценкой котировок акций P_1, P_2, P_3 на интересующий временной период, а в качестве выходного значения — его полезность. Оценка котировок задается в градусах из промежутка $[0;180]$ — углы наклона прямых, проведенных из точки S_i в S_{i+1} применительно к кусочно-линейной аппроксимации показателей.

Инвестиционный портфель с оценкой котировок акций определяется с помощью неупорядоченного факта investment-portfolio (Листинг 1).

Шкала линейных трендов определяется в виде лингвистической переменной trend, а ее значения представлены нечеткими треугольными числами.

Каждому факту investment-portfolio НПС ставит в соответствие нечеткий неупорядоченный факт fuzzy-investment-portfolio (листинг 2).

Фашификация четких фактов investment-portfolio в нечеткие факты fuzzy-investment-portfolio реализуется с помощью от-

Листинг 1

```
(deftemplate investment-portfolio
  ; Уникальный идентификатор состояния
  (slot time-state-id)
  ; Оценка котировок в градусах
  (slot P1 (type NUMBER) (range 0 180) (default 0))
  (slot P2 (type NUMBER) (range 0 180) (default 0))
  (slot P3 (type NUMBER) (range 0 180) (default 0))
  ; Признак фазификации факта
  (slot fuzzify (type SYMBOL) (allowed-symbols yes no)
                (default no))
).
```

Листинг 2

```
(deftemplate fuzzy-investment-portfolio
  ; Уникальный идентификатор состояния на временной промежуток,
  ; соответствующий time-state-id у investment-portfolio
  (slot time-state-id)
  ; Нечеткие значения котировок типа trend
  (slot P1 (type FUZZY-VALUE trend))
  (slot P2 (type FUZZY-VALUE trend))
  (slot P3 (type FUZZY-VALUE trend))
  ; средняя доходность портфеля
  (slot profit (default NONE))
).
```

Листинг 3

```
(defrule T1-rule
  (ready (value yes))
  ?fz-portfolio <- (fuzzy-investment-portfolio
                    (time-state-id ?time-state-id)
                    (P1 I) (P2 C) (P3 I)
                    (profit NONE))

  =>
  (bind ?profit 0.11)
  (modify ?fz-portfolio (profit ?profit))
  (printout t «Investment portfolio profit by « ?time-state-id «
              equal « ?profit crlf)
).
```

Листинг 4

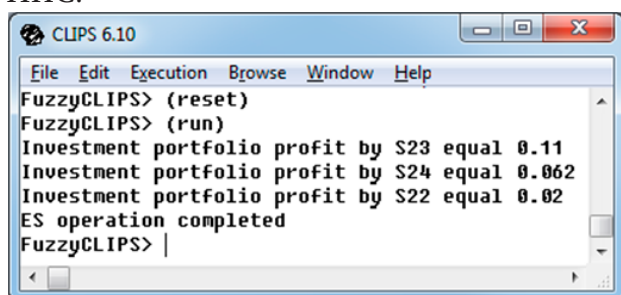
```
(deffacts start-facts
  ; значение левой/правой границ для треугольных нечетких чисел
  (delta 22.5)
  ; начальное значение статуса готовности
  (ready (value no))
  ; Инвест. портфель с id S22 и четкой оценкой котировок в
  ; градусах
  (investment-portfolio (time-state-id S22)
                       (P1 23) (P2 100) (P3 160))
  ; Инвест. портфель с id S23 и нечеткой оценкой котировок
  (fuzzy-investment-portfolio (time-state-id S23)
                              (P1 very I) (P2 C) (P3 I))
  ; Инвест. портфель с id S24 и нечеткой оценкой котировок
  (fuzzy-investment-portfolio (time-state-id S24)
                              (P1 I)
                              (P2 more-or-less C) (P3 very D))
).
```

дельного правила *fuzzify-rule*, вызывающей функцию *fuzzify-by-trend*, которая значению котировки в градусах ставит в соответствие нечеткое треугольное число.

Каждому правилу ставится в соответствие нечеткое продукционное правило (например, правила Листинг 3 содержит определение правила R_1).

Пусть в НПС поступает набор фактов, представленный в Листинге 4.

На рис. 3 представлен результат работы НПС.



```
CLIPS 6.10
File Edit Execution Browse Window Help
FuzzyCLIPS> (reset)
FuzzyCLIPS> (run)
Investment portfolio profit by S23 equal 0.11
Investment portfolio profit by S24 equal 0.062
Investment portfolio profit by S22 equal 0.02
ES operation completed
FuzzyCLIPS> |
```

Рис. 3. Результат работы нечеткой продукционной системы
[Fig. 3. The operating result of the fuzzy production system]

Разработанное приложение позволяет определять функцию полезности для заданного портфеля. Заметим, что заключение правила можно сформулировать иначе в зависимости от цели исследования и данных, которые задаются в качестве входной информации для каждого типового состояния.

ЗАКЛЮЧЕНИЕ

В данной статье представлен подход для формирования базы знаний в виде совокупности если-то-правил. Такая база знаний является ядром продукционных систем, предназначенных для решения различных задач в рамках процедур принятия решений в сложных системах. Формирование правил осуществляется на основе выявления типовых состояний системы. Исходная информация формируется в результате наблюдения основных системных показателей в форме временных рядов. Кусочно-линейная аппроксимация позволяет выявить тренды данных показателей, а на их основе выявить множество таких состояний системы, когда на соответ-

ствующем временном промежутке ни один из трендов не изменяется. В статье вводится понятие типовой ситуации для сложной системы. Для выявления типовых ситуаций предлагается использовать метод декомпозиционного дерева — кластерной процедуры, которая позволяет сгенерировать все возможные разбиения полученного множества ситуаций. Выбор оптимальной кластеризации можно осуществить либо с привлечением эксперта, либо на основе формальной процедуры, уточнив, что понимается под оптимальностью. Предложенный подход апробирован на основе задачи об оптимальном портфеле.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Draper, N. R. Applied Regression Analysis / N. R. Draper, H. Smith. – John Wiley – Sons, 1998. – 736 p.*
2. *Каширина, И. Л. О методах формирования нейросетевых ансамблей в задачах прогнозирования временных рядов / И. Л. Каширина // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2009. – № 2. – С. 116–119.*
3. *Ярушев, С. А. Модулярная модель прогнозирования временных рядов на основе нейро-нечетких сетей и когнитивного моделирования / С. А. Ярушев, А. Н. Аверкин, А. В. Федотова // Нечеткие системы и мягкие вычисления. – 2017. – №12(2). – С. 159–168.*
4. *Ярушкина, Н. Г. Интеллектуальный анализ временных рядов / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. – Ульяновск : УлГТУ, 2010. – 320 с.*
5. *Леденева, Т. М. Об одном подходе к формированию базы знаний для сегментации временных рядов / Т. М. Леденева, М. А. Сергиенко // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии, 2017. – № 3. – С. 156–164.*

6. Кофман, А. Введение в теорию нечетких множеств / А. Кофман. – Москва : Радио и связь, 1982. – 432 с.
7. Лекции по метрическим алгоритмам классификации. – Режим доступа: <http://www.scas.ru/frc/papers/voron04mpc.pdf>. – (Дата обращения: 4.03.2020).
8. Леденева, Т. М. Обработка нечеткой информации / Т. М. Леденева. – Воронеж : Изд-во ВГУ, 2006. – 233 с.
9. Каплиева, Н. А. Исследование различных типов транзитивности в приложении к нечеткой классификации / Н. А. Каплиева, Т. М. Леденева // Вестник Воронеж. гос. ун-та. Сер. Физика. Математика, 2006. – № 2. – С. 206–216.
10. Леденева, Т. М. Агрегирование информации в оценочных системах / Т. М. Леденева, С. Л. Подвальный // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии, 2016. – № 4. – С. 155–164.
11. Сергиенко, М. А. Разработка нечетких экспертных систем с помощью FuzzyCLIPS / М. А. Сергиенко // Актуальные проблемы прикладной математики, информатики и механики : сб. тр. Междунар. науч.-техн. конф. (Воронеж, 11–13 ноября, 2019г.) : электронный ресурс. – С. 323–325.

Леденева Татьяна Михайловна – д-р. техн. наук, проф., зав. кафедрой вычислительной математики и прикладных информационных технологий Воронежского государственного университета. E-mail: ledeneva-tm@yandex.ru
ORCID iD: <https://orcid.org/0000-0002-3944-2266>

Сергиенко Михаил Александрович – канд. техн. наук, доц. кафедры вычислительной математики и прикладных информационных технологий Воронежского государственного университета. E-mail: fers2003@list.ru
ORCID iD: <https://orcid.org/0000-0003-1322-9624>

Тихомирова Екатерина Александровна – студентка бакалавриата 4-го года обучения кафедры вычислительной математики и прикладных информационных технологий Воронежского государственного университета.
E-mail: katya246893@gmail.com
ORCID iD: <https://orcid.org/0000-0003-3141-9798>

FORMATION OF THE KNOWLEDGE BASE ON THE BASIS OF IDENTIFICATION OF TYPICAL STATES OF COMPLEX SYSTEM

© 2020 T. M. Ledeneva✉, M. A. Sergienko, E. A. Tichomirova

Voronezh State University
1, Universitetskaya square, 394018 Voronezh, Russian Federation

Annotation. This article presents the approach to form a knowledge base that describes the behaviour of a complex system. A scorecard is introduced to describe this behavior. It is assumed that as a result of their observation, time series are formed. Such time intervals, within which the linear trends of time series do not change, are identified based on the piecewise linear approximation. These intervals determine some state of a complex system. For a formal description of states, code vectors are used, which are formed based on a linguistic scale. Its gradations determine the base directions of linear trends. Each base direction corresponds to an integer code. The proximity of the slope of the linear trend to the base direction is determined using the membership function. It is proposed to use a cluster procedure to identify typical states. The analysis of suitable methods made it possible to separate the decomposition tree method as such a procedure. Its advantage is that it allows us to generate all possible partitions of a given set of states. At this stage, the problem of choosing the optimal partition arises. In this paper, optimal partition is a partition that contains as many classes as possible in the decomposition tree. Such classes exhibit stability in some sense. The optimal partition corresponds to a certain level of the decomposition tree, and the partition classes correspond to the typical states of a complex system. Under the assumption that the system indicators depend on some set of factors, a base of production rules is formed. The conclusions of these rules contain terms or functions that correspond to factors. The proposed approach is tested in the FuzzyClips environment to analyze the investment portfolio.

Keywords: time series, linear trend, typical state, production rule.

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. Draper N. R., Smith H. Applied Regression Analysis / John Wiley & Sons, 1998. 736 p.
2. Kashirina I. L. On the methods of formation of neural network ensembles in problems of time series forecasting. Bulletin of the Voronezh State University. Series: Systems Analysis and Information Technology. No 2. 2009. P. 116–119.
3. Yarushev S. A., Averkin A. N., Fedotova A. V. Modular model of time series forecasting based

on neuro-fuzzy networks and cognitive modeling. Fuzzy Systems and Soft Computing. 2017. No. 12(2). P. 159–168.

4. Yarushkina N. G., Afanasyeva T. V., Perfi-lyeva I. G. Intelligent analysis of time series. Ulyanovsk, UISTU. 2010. 320 p.

5. Ledeneva T. M., Sergienko M. A. On one approach to the formation of a knowledge base for time series segmentation. Bulletin of VSU. Series: System Analysis and Information Technology. No 3. 2017. P. 156–164.

6. Kofman A. Introduction to the theory of fuzzy sets. Moscow, Radio and communication; 1982. 432 p.

7. Lectures on metric classification algorithms. – Available at <http://www.ccas.ru/frc/papers/voron04mpc.pdf>. – (Date of access: 04.03.2020).

✉ Ledeneva Tatyana M.
ledeneva-tm@yandex.ru

8. *Ledeneva T. M.* Processing of fuzzy information. Voronezh, VSU publishing house. 2006. 233 p.
9. *Kaplieva N. A., Ledeneva T. M.* Study of various types of transitivity as applied to fuzzy classification. Bulletin of VSU. Series: Physics. Mathematics. No 2. 2006. P. 206–216.
10. *Ledeneva T. M., Podvalniy S. L.* Information Aggregation in Evaluation Systems: Bulletin of Voronezh State University. Series: System Analysis and Information Technology. No 4. 2016. P. 155–164.
11. *Sergienko M. A.* Development of fuzzy expert systems using FuzzyCLIPS. Contemporary problems of applied mathematics, computer science and mechanics: Proceedings of the International Scientific and Technical Conference (Voronezh, November 11–13, 2019): electronic source. P. 323–325.

Ledeneva Tatyana M. — Doctor of Technical Sciences, Professor, Chair of Computational Mathematics and Applied Information Technology, Voronezh State University. E-mail: ledeneva-tm@yandex.ru
ORCID iD: <https://orcid.org/0000-0002-3944-2266>

Sergienko Mikhail A. — Candidate of Technical Sciences, Associate Professor of the Department of Computational Mathematics and Applied Information Technology, Voronezh State University.
E-mail: fers2003@list.ru
ORCID iD: <https://orcid.org/0000-0003-1322-9624>

Tikhomirova Ekaterina A. — Student in the fourth year of bachelor's studies at the Department of Computational Mathematics and Applied Information Technology, Voronezh State University.
E-mail: katya246893@gmail.com
ORCID iD: <https://orcid.org/0000-0003-3141-9798>