

УДК 633.1:631.559

ББК 42.112

Т-32

Темиров Астемир Алиевич, аспирант кафедры прикладной математики федерального государственного бюджетного образовательного учреждения высшего образования «Финансовый университет при Правительстве Российской Федерации»; 125993, г. Москва, Ленинградский проспект, 49; e-mail: academy@fa.ru

**АЛГОРИТМЫ ЛИНЕЙНОГО КЛЕТОЧНОГО АВТОМАТА
ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ ЗЕРНОВЫХ**
(рецензирована)

Статья посвящена этапам алгоритма линейного клеточного автомата для прогнозирования временных рядов. Моделирование структуры клеточного автомата и его обучение осуществляется с помощью генетического алгоритма. Результаты прогнозирования представляются в виде лингвистического и числового нечеткого множества. Операция дефазификации выводит прогноз в виде одного четкого числа.

Ключевые слова: *нелинейная динамика, временные ряды, прогнозирование, нечеткое множество, клеточные автоматы, генетические алгоритмы*

Temirov Astemir Aliyevich, post graduate student of the Department of Applied Informatics of FSBEI HE "Financial University under the Government of the Russian Federation", 125993, Moscow, 49 Leningradsky Avenue; e-mail: academy@fa.ru

**ALGORITHMS OF THE LINEAR CELLULAR AUTOMATIC MACHINE
FOR THE FORECASTING GRAIN PRODUCTIVITY**
(reviewed)

The article is devoted to the stages of the algorithm of the linear cellular automatic machine for forecasting temporary ranks. Modeling of the structure of the cellular machine and its training is carried out by means of genetic algorithm. The results of the forecasting are presented in the form of linguistic and numerical fuzzy set. Operation of defuzzification makes the forecast in the form of one accurate number.

Keywords: *nonlinear dynamics, temporary ranks, forecasting, indistinct set, cellular machines, genetic algorithms.*

В настоящее время в связи с развитием вычислительной техники и прикладного программного обеспечения получило развитие новые прогнозные модели на базе искусственного интеллекта, такие как нейронные сети, клеточные автоматы с различными радиусами, генетические алгоритмы, нечеткие системы и др., которые на выходе дают хорошие приближения [4, 5, 6].

В статье представлены этапы алгоритма клеточно-автоматной прогнозной модели на примере временного ряда (ВР) урожайности «зерновых всего» в России. Графическое представление этого временного ряда в виде гистограммы на рисунке 1.

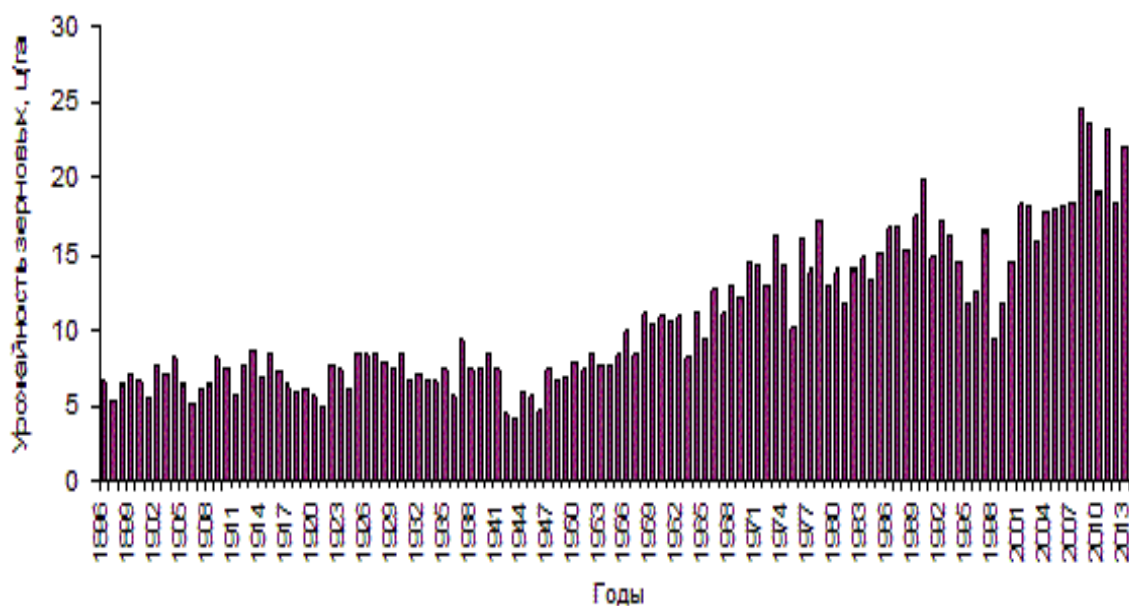


Рисунок 1. Гистограмма временного ряда урожайности «зерновых всего» в России за период времени 1896-2014 гг.

Известно, что клеточный автомат работает с памятью и представляет собой систему клеток, в каждой из которых хранится информация. Характер изменения информации в каждый момент времени программируется с помощью генетического алгоритма [4, 5, 7, 8], который в свою очередь определяет взаимодействие клеток со всеми ближайшими соседями. Применение генетического алгоритма [7, 8] позволяет получить наилучшую конфигурацию клеточного автомата, и, следовательно, наилучшую прогнозную модель. Трудоемкость генетического алгоритма [6, 7, 8] обычно определяется глубиной памяти не только временного ряда, но и клеточного автомата.

Шаг 1. Для целей создания базиса памяти КА исходный числовой ВР преобразовывается в лингвистический временной ряд (ЛВР). Каждому числовому значению уровня ВР ставится в соответствие лингвистическая переменная, которая в теории нечеткого множества называется термом [1, 2, 3, 6]. Совокупность термов образует терм-множество как минимум из 3-х элементов $T = \{\hat{I}, \tilde{N}, \hat{A}\}$, где $H(C, B)$ означают соответственно низкий, средний и высокий уровень урожайности. Существуют различные способы преобразования числового ВР в ЛВР [6, 9]: метод огибающих ломаных, метод трендовых коридоров, метод кластеризации, метод равных интервалов и т.д. Выбор подходящего способа обычно осуществляется экспертным путем или же методом «мозгового штурма», при этом опираясь в основном на тенденцию динамики исследуемого ВР. Для рассматриваемого временного ряда в результате применения метода «огибающих ломаных» [6] получен ЛВР, представленный табл. 1, в которой отражены численные значения урожайности (ц/га) «зерновых всего» в России и их лингвистическая оценка.

Схематично модель памяти КА можно представить в виде двудольных ориентированных графов (см. рис. 2), в котором вершины первой доли представляют l -конфигурации определенной длины, имеющие место в ЛВР, а вершины второй доли – исходы $H(C, B)$. Дугам приписаны веса, означающие частоту перехода данной l -конфигурации в соответствующие термы $H, C, B \in T$.

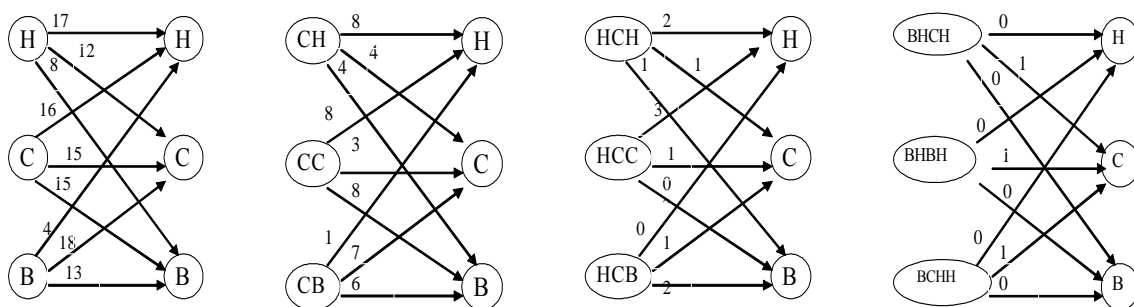


Рисунок 2. Типовые двудольные графы памяти клеточного автомата с переходом из 1-ой, 2-х, 3-х и 4-х конфигураций в состояния $H(C, B)$

Для принятого терм-множества $T = \{H, C, B\}$ возможное количество различных l -конфигураций, $l = 1, 2, \dots, m$, содержащихся в ЛВР составляет теоретически $\sum_{l=1}^m 3^l = 3^1 + 3^2 + \dots + 3^m$. Для исследуемого ЛВР, представленного в таб. 1 количество M_l , $l = \overline{1, 6}$ всех таких попарно различных l -конфигураций теоретически составило $\sum_{l=1}^6 3^l = 3^1 + 3^2 + 3^3 + 3^4 + 3^5 + 3^6 = 1092$, а количество реальных l -конфигураций – 285, из них: $M_1 = 3$, $M_2 = 9$, $M_3 = 26$, $M_4 = 56$, $M_5 = 86$, $M_6 = 105$, что составляет всего 10% от числа теоретически возможных конфигураций. Для лингвистического временного ряда исследуемого ВР итоговая статистика переходов l -конфигураций, $l = \overline{1, 6}$ в состояния $H(C, B)$ представлена табл. 2. Из таблицы 2 видно, что все конфигурации длины $l = 6$ демонстрируют 100% наличие памяти. Например, представленный на рис. 8 типовой ориентированный граф для конфигурации длины $l = 4$ демонстрирует однозначные переходы в состояния $H(C, B)$, тем самым, демонстрируя полную память. Формирование памяти КА завершается вычислением эмпирических значений частностей всех конфигураций имеющих место в ЛВР.

Таблица 1 - Числовые урожайности временного ряда «зерновых всего» в России и их лингвистическая оценка

Годы	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905
Урожайность, ц/га	6,5	5,4	6,3	7,1	6,6	5,5	7,7	7	8,2	6,4
Лингвистическая	С	Н	Н	С	С	Н	В	С	В	С
Годы	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915
Урожайность, ц/га	5,2	6,2	6,4	8,3	7,6	5,8	7,8	8,7	6,9	8,6
Лингвистическая	Н	Н	С	В	В	Н	В	В	С	В
Годы	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925
Урожайность, ц/га	7,2	6,4	6	6,2	5,7	5	7,7	7,3	6,2	8,6
Лингвистическая	С	С	Н	Н	Н	Н	В	С	С	В
Годы	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935
Урожайность, ц/га	8,4	8,6	7,9	7,5	8,5	6,7	7	6,7	6,5	7,3
Лингвистическая	В	В	В	В	В	С	С	С	С	С
Годы	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945
Урожайность, ц/га	5,7	9,3	7,3	7,5	8,6	7,3	4,4	4,2	6	5,6
Лингвистическая	Н	В	С	С	В	С	Н	Н	Н	Н
Годы	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955
Урожайность, ц/га	4,6	7,3	6,7	6,9	7,9	7,4	8,6	7,8	7,7	8,4
Лингвистическая	Н	С	Н	Н	С	С	С	С	Н	С
Годы	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Урожайность, ц/га	9,9	8,4	11,1	10,4	10,9	10,7	10,9	8,3	11,4	9,5
Лингвистическая	В	С	В	В	В	С	С	Н	С	Н
Годы	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
Урожайность, ц/га	12,7	11,2	13	12,2	14,4	14,2	13	16,3	14,2	10,1
Лингвистическая	В	С	В	С	В	В	С	В	С	Н
Годы	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Урожайность, ц/га	16,2	13,9	17,1	13,1	13,9	11,7	14,1	14,7	13,3	15,1
Лингвистическая	В	С	В	Н	С	Н	С	С	Н	С
Годы	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Урожайность, ц/га	16,6	16,8	15,3	17,5	19,9	14,7	17,2	16,3	14,4	11,6
Лингвистическая	В	В	С	В	В	С	В	С	С	Н
Годы	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Урожайность, ц/га	12,5	16,5	9,4	11,7	14,4	18,3	18,2	15,9	17,9	18
Лингвистическая	Н	С	Н	Н	Н	С	С	Н	Н	Н
Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Урожайность, ц/га	18,1	18,4	24,6	23,6	19	23,3	19,3	23,1	26,2	
Лингвистическая	Н	Н	В	В	Н	В	Н	С	В	

Шаг 3. На данном шаге осуществляется процесс получения лингвистических прогнозных значений на основе статистики переходов всех конфигураций в состояния H, C и B .

Таблица 2 - Итоговая статистика переходов l – конфигураций, где $l = \overline{1,6}$
для ЛВР в состоянии $H(C, B)$

l – конфигурации	Всего конфигураций, шт.	Из них переходов:			Память		
		1-знач- ных, шт.	2-знач- ных, шт.	3-знач- ных, шт.	полная, %	частичная, %	нет памяти, %
1	3	-	-	3	-	-	100%
2	9	-	1	8	-	11%	89%
3	26	5	12	9	19%	46%	35%
4	56	30	23	3	54%	41%	5%
5	86	65	21	-	76%	24%	-
6	105	105	-	-	100%	-	-

Прогноз представляется в виде нечеткого лингвистического множества (НЛМ) вида $U_{n+1} = \{(H; \mu_H), (C; \mu_C), (B; \mu_B)\}$, где H, C, B – лингвистические переменные с соответствующими функциями принадлежности μ_H, μ_C, μ_B из интервала от 0 до 1, удовлетворяющие условию $\mu_H + \mu_C + \mu_B = 1$.

В виду того, что значение термина уже известно H, C, B , нахождение прогнозного значения сводится к вычислению значений функции принадлежности μ_H, μ_C, μ_B . Они в свою очередь вычисляются через значения частоты переходов, получаемых для различных l -конфигураций в отрезке ЛВР $u_{n-m+1}u_{n-m+2} \dots u_n$, где $m = 6$. Применительно к рассматриваемому ВР (см.рис. 1) берем базовый отрезок длины 6, состоящий из сочетаний H, C, B образующих конфигурацию $BHVBHCB$, термины которых соответствуют годам (см. табл.1): 2009, 2010, 2011, 2012, 2013, 2014.

Рассматриваем последовательно следующие l -конфигурации из этого отрезка $BHVBHCB$: $B; CB; HCC; BHCB; HBHCB; BHVBHCB$. Раскладываем этот отрезок $BHVBHCB$ на двудольные графы типа «звезда» [6], как это представлено на рис. 3. Следует отметить, что такое разложение прекратилось на конфигурации длины 4 ($BHCB$), т.к. клеточный автомат на ней уже продемонстрировал полную память типа 0,0,1. Ребрам графа приписаны веса, означающие количество переходов из l -конфигурации в состояния H, C, B .

Опираясь на графическое представление частоты переходов, найдем эмпирические значения частоты переходов, как отношение частоты переходов из данной конфигурации в соответствующие H, C, B к общему числу переходов:

Для конфигурации длины $l = 1$ составило $w_1(B \rightarrow H) = \frac{4}{32}, w_1(B \rightarrow C) = \frac{15}{32}, w_1(B \rightarrow B) = \frac{13}{32}$; для $l = 2$: $w_2(CB \rightarrow H) = \frac{1}{14}, w_2(CB \rightarrow C) = \frac{7}{14}, w_2(CB \rightarrow B) = \frac{6}{14}$; для $l = 3$: $w_3(HCB \rightarrow H) = \frac{0}{2}, w_3(HCB \rightarrow C) = \frac{1}{2}, w_3(HCB \rightarrow B) = \frac{1}{2}$; для конфигурации $l = 4$: $w_4(BHCB \rightarrow H) = 0, w_4(BHCB \rightarrow C) = 0, w_4(BHCB \rightarrow B) = 1$. Конфигурация длины 4 демонстрирует наличие памяти, в силу чего процесс вычисления частоты можно прекратить. На основании выше вычисленных значений частоты, определяем ненормированные значения функции принадлежности: $\mu'_H = \frac{4}{32} + \frac{1}{14} + \frac{0}{2} + 0 = 0,21$; $\mu'_C = \frac{15}{32} + \frac{7}{14} + \frac{1}{2} + 0 = 1,47$; $\mu'_B = \frac{13}{32} + \frac{6}{14} + \frac{1}{2} + 1 = 2,34$ и их сумму, обозначенную как $\sigma = 0,21 + 1,47 + 2,34 = 4,02$. Затем, осуществляя операцию нормирования, получим искомое значение функции принадлежности для каждого термина из терм-множества $T = \{H, C, B\}$: $\mu_H = \frac{\mu'_H}{\sigma} = 0,06, \mu_C = \frac{\mu'_C}{\sigma} = 0,45, \mu_B = \frac{\mu'_B}{\sigma} = 0,49$. Результат записываем в виде нечеткого лингвистического множества (НЛМ) $U_{n+1} = \{(H; 0,06), (C; 0,45), (B; 0,49)\}$.

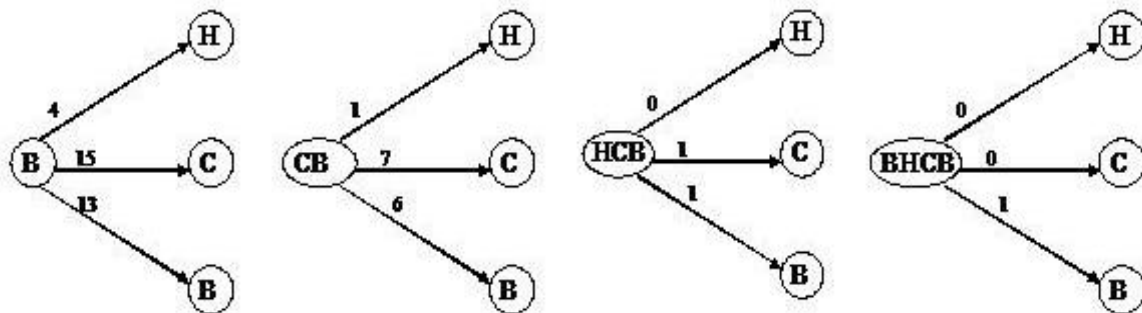


Рисунок 3. Двудольные ориентированные графы типа «звезда» для базового отрезка $BHVBHCB$

Так как прогноз осуществляется на лингвистическом уровне, поэтому определенно на качественном уровне можно охарактеризовать урожайность зерновых следующим образом: в 2015 году в России ожидается высокая урожайность со степенью истинности $\mu = 0,49$ и менее вероятно, что будет низкой.

Шаг 4. Получение числового прогноза осуществляется следующим образом: лингвистическим значениям из базового отрезка длины 6 (*ВНВНСВ*) ставится в соответствие среднее значение числовых данных из табл. 1:

$$H(\text{низкий}) = \frac{19,3+19,0}{2} = 19,15; C(\text{средний}) = 23,1;$$

$$B(\text{высокий}) = \frac{26,2+23,3+23,6}{3} = 24,37.$$

С учетом значений функции принадлежности μ_H, μ_C, μ_B , которые уже имеют место в НЛМ, получаем искомый прогноз в виде числового нечеткого множества (НЧМ) $Z_{n+1}^0 = \{(19,15;0,06), (23,1;0,45), (24,37;0,49)\}$. Чтобы представить урожайность зерновых в виде одного обычного числа применим к НЧМ Z_{n+1}^0 операцию дефазификации [1, 3] $Z_{n+1}^0 = 19,15 \cdot 0,06 + 23,1 \cdot 0,45 + 24,37 \cdot 0,49 = 24 \text{ ц/га}$. Таким образом, прогнозная урожайность «зерновых всего» в России на 2015 год составит ≈ 24 ц/га.

Шаг 5. Оценка валидации прогнозной модели осуществляется последовательным прогнозированием последних пяти известных уровней исходного ВР на рис.1. Результаты верификации и валидации прогнозной модели на базе клеточного автомата представлены в табл. 3. Средняя ошибка числового прогноза при этом составила $\approx 8,04\%$, а ошибка лингвистического прогноза составила 0%.

Таблица 3 - Результат валидации прогнозной модели на примере прогнозирования последних 5 уровней ВР Z

№ п/п	l-конфигурация	Исходные данные	Численный прогноз	Погрешность	Терм	H	μ_H	C	μ_C	B	μ_B
119	НВНС	26,2	22,09	15,70	В	19,3	0,295	23,1	0,165	23,3	0,54
118	НВН	23,1	19,62	15,08	С	19,3	0,153	18,25	0,608	23,3	0,239
117	НВВН В	19,3	20,22	4,79	Н	19	0,448	18,25	0,228	23,3	0,324
116	НВВН	23,3	22,31	4,26	В	19	0,115	18,25	0,206	24,1	0,679
115	ННВВ	19	18,93	0,38	Н	17,6	0,449	18,25	0,39	24,1	0,161
				$\approx 8,04\%$							

При анализе временных рядов используется множество различных методов и подходов, которые продолжают совершенствоваться и развиваться. Одним из таких методов является разработанная клеточно-автоматная прогнозная модель, которая экспериментально реализована на примере временного ряда ежегодных наблюдений урожайности «зерновых всего» в России с 1896 по 2014 гг.

По результатам проведенного фрактального анализа [7], для исследуемого ВР урожайности «зерновых всего» в России получено, что глубина его памяти находится в окрестности числа $\approx 5,4$. Оценка глубины памяти ВР с помощью КА на базе генетического алгоритма составила число 6 (шесть), что «почти» согласуется с результатами фрактального анализа. В силу этого, для достижения высокой точности прогнозного значения, в качестве базового отрезка, загружаемого в оперативную память КА, рассматриваются последние шесть уровней ВР и соответствующие им термы из ЛВР,

образующие в комплексе конфигурацию «ВНВНСВ».

Прогнозное значение представлено в виде нечеткого лингвистического множества, нечеткого числового множества и четкого числа после применения к последнему операции дефазификации. Аппарат теории нечеткого множества позволяет адекватно отражать нечеткость и неопределенность значений исходных данных в рассматриваемых математических моделях. Адекватность клеточно-автоматной прогнозной модели для прогнозирования ВР урожайности сельскохозяйственных культур, в частности, «зерновых» подтверждена. Ошибка прогноза составила $\approx 8,04\%$.

Литература:

1. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях. Тюмень: ТюмГУ, 2000. 352 с.
2. Андерсон Т.В. Статистический анализ временных рядов. Москва: Мир, 1976. 756 с.
3. Жиравок А.Н. Нечеткие множества и их использование для принятия решений // Сорровский образовательный журнал. 2001. Т. 7, №2. С. 109-115.
4. Курдюмов С.П., Малинецкий Г.Г., Потапов А.Б. Нестационарные структуры, динамический хаос, клеточные автоматы // Новое в синергетике. Загадки мира неравновесных структур. Москва: Наука, 1996. С. 95-164.
5. Нейман Дж. Теория самопроизводящихся автоматов. Москва: Мир, 1971. 378 с.
6. Перепелица В.А., Тебуева Ф.Б., Темирова Л.Г. Структурирование данных методами нелинейной динамики для двухуровневого моделирования. Ставрополь, 2006. 284 с.
7. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. Москва: Мир, 2000. 333 с.
8. Романов В.П. Интеллектуальные информационные системы в экономике: учебное пособие / под ред. проф. Н.П. Тихомирова. 2-е изд., стереотип. Москва: Экзамен, 2007. 496 с.
9. Сергеева Л.Н. Моделирование поведения экономических систем методами нелинейной динамики (теории хаоса). Запорожье: ЗГУ, 2002. 277 с.

Literature:

1. Altunin A.E., Semukhin M. V. Models and algorithms of decision-making in fuzzy conditions. Tyumen: TyumSU, 2000. 352 p.
2. Anderson T.V. Statistic analysis of temporary ranks. Moscow: World, 1976. 756 p.
3. Zhirabok A.N. Fuzzy sets and their use for decision-making// Sorovsky educational magazine. 2001. V. 7, No. 2. P. 109-115.
4. Kurdyumov S.P., Malinetsky G.G., Potapov A.B. Non-stationary structures, dynamic chaos, cellular machines//New in synergetics. Riddles of the world of nonequilibrium structures. Moscow: Science, 1996. P. 95-164.
5. Neumann J. The theory of the self-made machines. Moscow: World, 1971. 378 p.
6. Perepelitsa V.A., Tebuyeva F.B., Temirova L.G. Structuring data by methods of nonlinear dynamics for two-level modeling. Stavropol, 2006. 284 p.
7. Peters E. Chaos and order in the capital markets. A new analytical view on cycles, prices and variability of the market. Moscow: World, 2000. 333 p.
8. Romanov V.P. Intellectual information systems in economy: manual / ed. by prof. N. P. Tikhomirov. 2d ed., stereotype. Moscow: Examination, 2007. 496 p.
9. Sergeyeva L.N. Modeling of the behavior of economic systems by methods of nonlinear dynamics (theory of chaos). Zaporozhye: ZSU, 2002. 277 p.