

ОБРАБОТКА ИНФОРМАЦИИ

УДК 621.372(075)

А.С. Гуменюк, Е.В. Морозенко, И.Н. Родионов

ФОРМАЛИЗАЦИЯ АНАЛИЗА СТРОЯ ЗНАКОВЫХ ЦЕПЕЙ

В настоящее время не используются адекватные математические средства для анализа порядка следования событий в цепях конечной длины. Поэтому затруднены или невозможны измерение и сравнение формы сигналов, композиции музыкальных текстов, порядка элементов в длинных молекулярных цепях, взаимного расположения элементов в массивах данных, представляющих цепи событий разной природы. Объектом исследования в данной работе являются знаковые последовательности. Сформулированы выражения для числовых характеристик, которые описывают оригинальный порядок элементов цепи.

Ключевые слова: *строй элементов, интервал, числовые характеристики строя цепи, энтропия, информация.*

Для исследования текстов и знаковых цепей разработано и используется множество специальных подходов, процедур и моделей, среди которых можно отметить спектральный, корреляционный, фрактальный, статистический подходы, марковские цепи, потоки заявок. Однако почти не уделяется внимания закономерностям «конкретного расположения всех знаков или слов, составляющих отдельную символическую последовательность». На наш взгляд, такое положение, в некоторой степени, объясняется отсутствием формализма для выделенного абстрактного объекта и названного «строем или построением цепи» [1, 2]. Следует отметить, что разные по природе последовательности событий с равномошным составом могут иметь один и тот же оригинальный строй. С другой стороны, очевидно, что мультимножество событий некоторой мощности, может быть основой для построения последовательностей, имеющих разный строй. В данной работе рассматривается подход, который предназначен для формального анализа построения отдельного текста, знаковой последовательности или цепи сообщений. Подчеркнем, что здесь не рассматриваются методы исследования локальной структуры символических последовательностей.

1. Формализация строя цепи

Рассмотрим множество знаковых последовательностей конечной длины. Выделим подмножество кортежей с одинаковыми наборами чисел вхождения символов собственных алфавитов. Так как имеется в виду неоднократное вхождение некоторых компонентов, то все множество кортежей, составленных на основе конкретного алфавита, – это комбинации типа «перестановки с повторениями», каждая из которых отличается оригинальным взаимным расположением компонентов, собственной их композицией или построением.

Кортеж, в котором знаки заменены натуральными числами, как это показано на рис. 1, назовем порядком следования элементов, построением или строем цепи [1, 2]. В результате таких замен все равнозначные по составу знаковые последовательности, имеющие одинаковое взаимное расположение компонентов (композицию), будут отображены одной и той же последовательностью натуральных чисел, которая представляет оригинальный для данного подмножества текстов строю цепи (порядок следования элементов).

<i>Q</i>	<i>R</i>	<i>D</i>	<i>&</i>	<i>V</i>	<i>Y</i>	<i>S</i>	<i>S</i>	<i>&</i>	<i>S</i>	<i>D</i>	<i>D</i>	<i>S</i>	<i>Q</i>	<i>S</i>	<i>D</i>	<i>&</i>	знаковые последовательности
<i>W</i>	<i>T</i>	<i>C</i>	<i>G</i>	<i>H</i>	<i>U</i>	<i>R</i>	<i>R</i>	<i>G</i>	<i>R</i>	<i>C</i>	<i>C</i>	<i>R</i>	<i>W</i>	<i>R</i>	<i>C</i>	<i>G</i>	
T	Y	M	H	T	O	S	S	H	S	M	M	S	T	S	M	H	
<i>U</i>	<i>I</i>	<i>L</i>	<i>J</i>	<i>W</i>	<i>O</i>	<i>Y</i>	<i>Y</i>	<i>J</i>	<i>Y</i>	<i>L</i>	<i>L</i>	<i>Y</i>	<i>U</i>	<i>Y</i>	<i>L</i>	<i>J</i>	
V	N	A	B	J	K	T	T	B	T	A	A	T	V	T	A	B	
1	2	3	4	5	6	7	7	4	7	3	3	7	1	7	3	4	строй цепи

Рис. 1. Примеры знаковых последовательностей, имеющих одинаковый строй цепи

Строй цепи событий (сообщений, знаков и т.д.) определен как кортеж (упорядоченное множество), в котором каждому компоненту данной цепи поставлено в соответствие натуральное число, причем идентичные по выбранному признаку компоненты отображены одним и тем же числом. Самый первый компонент такого кортежа – единица, а все остальные первые встречные разные натуральные числа (представляющие вместе с единицей алфавит строя) возрастают на единицу.

В соответствии с теоретико-множественным определением вектора назовем такой специфически сформированный (организованный) кортеж «вектор строя».

Разложим полную неоднородную символьную последовательность (без свободных мест на ее позиции) на m неполных «однородных» кортежей, в которых заняты только некоторые места данной позиции одинаковыми знаками. Аналогом однородной последовательности является однородный поток событий, рассматриваемый в теории массового обслуживания. Вообще разложение цепи может осуществляться по разным правилам. Так на рис. 2 представлена декомпозиция строя неоднородной знаковой цепи на неполные однородные. Разложение на неполные «разнородные» цепи, когда отдельные места данной позиции заняты только разными знаками, осуществляется по следующему правилу: при просмотре цепи от ее начала в состав первой разнородной цепи выбираются все первые вхождения каждого знака из алфавита, при втором – все вторые вхождения и т.д.

В случае правильно выполненных декомпозиций полученные множества однородных (разнородных) последовательностей будут несовместными (т.е. не содержат занятых мест с одинаковыми номерами на позиции). Композиция всех неполных однородных (разнородных) кортежей дает исходную полную неоднородную знаковую последовательность.

Определим «интервал» как расстояние от выделенного в цепи компонента до другого, ближайшего, отмеченного в направлении просмотра (рис. 2); по другому – это модуль разности номеров мест двух выделенных компонентов на позиции кортежа.

Пусть первое считывание текста осуществляется отличным от обычного (чтение подряд) способом с самого начала до конца таким образом, что выбираются только элементы строя с номером «1»; при этом последний интервал определяется до знака «финиш» (возможен и другой вариант – определение первого интервала от начала текста – «старта»). Интервалы данной однородной последовательности

разместим в соответствии с номерами считываемых элементов в первой строке матрицы. Далее, при втором просмотре строя текста, аналогично выберем элементы с номером «2» и разместим вектор интервалов, соответствующий другой однородной последовательности, во второй строке матрицы. В каждой следующей строке помещается вектор интервалов «новой» при очередном просмотре однородной последовательности. Одиночные знаки, слова или сообщения будут представлены всего одним интервалом (до финиша), который размещается в крайнем столбце соответствующей строки матрицы. Число столбцов $n_{j \max}$ в «матрице интервалов» однородных цепей равно числу вхождений самого частого знака (или слова) текста. Незанятые интервалами элементы матрицы заполним нулями. Число строк m равно мощности собственного алфавита или словаря текста.



Рис. 2. Декомпозиция строя неоднородной знаковой цепи на неполные однородные

Пусть считывание текста осуществляется вторым (вышеописанным) способом по разнородным цепям. В результате получим матрицу интервалов разнородных цепей, в которой число столбцов равно m , а число строк – $n_{j \max}$.

2. Числовые характеристики строя цепи

Используем понятие однородной знаковой последовательности и ее векторное отображение в виде соответствующей строки матрицы интервалов для определения некоторых числовых характеристик строения текста, которые для компактности представлены в табл. 1 [3].

В таблице обозначены: j – номер знака в алфавите или номер однородной цепи; Δ_{ij} – интервал между i -м и $(i+1)$ -м вхождениями знака j в однородной цепи; $\log \Delta_{ij}$ – удаленность $(i+1)$ -го вхождения знака j относительно его i -го вхождения; n_j – число вхождения знака j в цепи; m – мощность алфавита знаков цепи; n – длина полной знаковой цепи, равная числу всех ее позиций; Δ_{aj} , Δ_{gj} – соответственно средний арифметический и средний геометрический интервалы строя j -й однородной цепи; V_j – абсолютный объем строя j -й однородной цепи; G_j – глубина расположения строя j -й однородной цепи; $g_j = \log \Delta_{gj}$ – средняя удаленность знака j в строе однородной цепи; V – абсолютный объем строя на основе однородных цепей; G – глубина расположения строя на основе однородных цепей; Δ_g – средний геометрический интервал строя на основе однородных цепей; g – средняя удаленность любого элемента строя на основе однородных цепей; D – число описатель-

ных информаций, используемых (по Мазуру [4]) для определения некоторого знака цепи; I – число идентифицирующих информаций, используемых (по Мазуру) для дихотомического распознавания отдельного знака; τ_j – периодичность (следования знаков) строя j -й однородной цепи; r – регулярность (следования любого элемента) строя на основе однородных цепей.

Таблица 1

Числовые характеристики строя на основе однородных цепей

(1) $V_j = \prod_{i=1}^{n_j} \Delta_{ij}$	(2) $\Delta_{gj} = \sqrt[n_j]{V_j}$	(3) $G_j = \sum_{i=1}^{n_j} \log_2 \Delta_{ij}$
(4) $V = \prod_{j=1}^m V_j$	(5) $\Delta_g = \sqrt[n]{V}$	(6) $G = \sum_{j=1}^m G_j$
(7) $\Delta_{aj} = \frac{n}{n_j} = \frac{1}{P_j}$	(8) $\tau_j = \frac{\Delta_{gj}}{\Delta_{aj}}$	(9) $r = \frac{\Delta_g}{D} \leq 1$
(10) $\Delta_g = \prod_{j=1}^m \Delta_{gj}^{\frac{n_j}{n}}$	(11) $g = \sum_{j=1}^m \frac{n_j}{n} \log \Delta_{gj}$	(12) $g = \log \Delta_g$
(13) $D = \prod_{j=1}^m \Delta_{aj}^{\frac{n_j}{n}}$	(14) $I = \sum_{j=1}^m \frac{n_j}{n} \log \Delta_{aj}$	(15) $H = -\sum_{j=1}^m P_j \log P_j$

Множества величин $\{G_j\}$ и $\{g_j\}$ являются соответственно распределением глубин и распределением средних удаленностей всех однородных последовательностей для строя отдельной цепи. При анализе и описании строя данной цепи распределение средних удаленностей дополняет обычное частотное распределение знаков $\{n_j\}$ комплексным распределением строя вида $\{<n_j, g_j>\}$.

Легко определяются аналогичные по форме, но отличные по содержанию, числовые характеристики строя данной цепи, с использованием векторов интервалов, представленных строками матрицы для разнородных цепей. В таком случае формулы (1) – (15) определяются при выделении знаков j в i -х разнородных цепях и представляют числовые характеристики строя на основе разнородных цепей.

3. Числовые характеристики для «предельных» моделей строя цепи

Для «регулярной» знаковой цепи, в которой все интервалы каждой однородной цепи равны $\Delta_{ij} = \Delta_{aj} = n/n_j$, числовые характеристики строя цепи (10) и (11) принимают максимальные значения (см. (16), (17)) и записываются формулами Мазура (13) и (14), представляющими соответственно числа описательных и идентифицирующих информаций в цепи сообщений.

$$\Delta_g = \Delta_{g \max} = D; \quad (16)$$

$$g = g_{\max} = I; \quad (17)$$

$$D > \Delta_g; \quad (18)$$

$$H > g = \log \Delta_g. \quad (19)$$

Для бесконечной знаковой цепи ($n \rightarrow \infty$) формула Мазура (14), в которой $\Delta_{aj} = (n/n_j) \rightarrow (1/P_j)$, принимает вид формулы К.Шеннона (15) для энтропии или количества информации. Такая информация используется только для дихотомии-

ческой идентификации (но не для описания) отдельных сообщений. Для текстов и других нерегулярных последовательностей формулы Мазура и Шеннона дают оценку строя только «сверху», так как в этих случаях имеют место неравенства (18) и (19). Соответственно числовые характеристики строя на основе однородных цепей принимают минимальные значения для «сплошных» последовательностей, в которых все одинаковые элементы расположены подряд.

Таким образом, формулы для среднего геометрического интервала и средней удаленности знаковой цепи обобщают формулы Мазура и Шеннона, так как, в отличие от последних, при описании строя данной цепи учитывают не только мощность состава, но и взаимное расположение ее компонентов (знаков, слов).

Полагаем важным отметить взаимосвязь характеристик строя цепи, полученных на основе однородных и разнородных цепей. Так, для рассмотренных выше моделей регулярной и сплошной последовательностей, характеристики строя на основе разнородных цепей принимают соответственно минимальное и максимальное значения.

4. Числовые характеристики строя текста для литературных произведений

Проиллюстрируем чувствительность описанных выше характеристик к изменению взаимного расположения элементов последовательности на примере литературных текстов.

В табл. 2 представлены результаты обработки строя произведения А.С. Пушкина «Капитанская дочка» и его трех модификаций в виде энтропийных характеристик и числовых характеристик строя текста. Проведенные эксперименты показали, что даже малые изменения строя текста (модификация 2) приводят к изменениям значений числовых характеристик строя: для глубины G во втором знаке после запятой, для удаленности g и регулярности r – в шестом. Отметим, что данные характеристики позволяют только определять факт изменения взаимного расположения знаков в цепи, без его локализации.

При автоматической сегментации текстов в качестве элементарных единиц использовались слова, нормированные путем отбрасывания окончания. Критерием качества сегментации являлась степень совпадения мощностей составов, полученных автоматически и вручную в работах Ю.К. Орлова [5].

Таблица 2

Энтропийные и числовые характеристики строя повести А.С. Пушкина «Капитанская дочка»

	n	m	H_s	G	H	g	r
1	29312	4789	290585.01	253104.07	9.913517	8.634847	0.412207
2	29312	4789	290585.01	253104.03	9.913517	8.634845	0.412206
3	29312	4789	290585.01	263790.75	9.913517	8.999375	0.530715
4	29312	4789	290585.01	63944.14	9.913517	2.181500	0.004704

В таблице обозначены: H_s – количество информации в тексте; вычисляется в предположении статистической независимости слов путем перемножения H и n ; 1 – строй оригинального текста; 2 – модификация строя с небольшими изменениями (переставлены два смежных слова); 3 – модификация строя с большими изменениями (со случайными перестановками всех слов); 4 – модификация строя со словами, упорядоченными по алфавиту.

Так же было проведено исследование 56 литературных произведений, представляющих три направления: русская художественная классика (Пушкин (6), Тургенев (6), Гоголь (7)), фантастика (Фармер (6), Муркок (5), Желязны (6)) и философские тексты (Ницше (6), Шопенгауэр (5), Фрейд (9)). В табл. 3, приведены числовые характеристики строя текстов трех авторов, представляющих разные направления. Тексты упорядочены по возрастанию значений средней удаленности.

Первые исследования небольшой выборки текстов разных авторов и направлений показали, что каждый текст характеризуется уникальным значением рассмотренных числовых характеристик, которые представляют только взаимное расположение слов без учета их морфологии. Исследованные литературные произведения, упорядоченные по величине g или r , группируются в различимые таксоны.

Таблица 3

Характеристики строя исследованных текстов

№	Автор	Произведение	m	n	G	g	r
1	Шопенгауэр	О политике	1818	8363	65655,1	7,851	0,385
2	Шопенгауэр	О жизни	1712	5740	45541,4	7,934	0,410
3	Шопенгауэр	О смерти	2824	17853	144301,9	8,083	0,429
4	Шопенгауэр	О философии	1711	5495	44611,8	8,119	0,436
5	Тургенев	Муму	2363	8510	69871,3	8,211	0,415
6	Тургенев	Ася	2953	13835	115179,9	8,325	0,447
7	Тургенев	Рудин	5019	35498	300182,9	8,456	0,403
8	Фармер	Несколько слов	3040	11288	95830,8	8,490	0,416
9	Тургенев	Вешние воды	6020	38718	338561,2	8,744	0,403
10	Тургенев	Дворянское гнездо	6480	46795	409538,6	8,752	0,411
11	Тургенев	Отцы и дети	6964	54790	480344,1	8,767	0,410
12	Шопенгауэр	Афоризмы	6581	54276	477795,4	8,803	0,402
13	Фармер	Создатель	7669	58617	519919,2	8,870	0,405
14	Фармер	В тела вернитесь	8314	69902	620965,8	8,883	0,404
15	Фармер	Врата	6512	44450	396971,7	8,931	0,407
16	Фармер	Пассажир	5960	23536	216793,4	9,211	0,390
17	Фармер	Гоблин	5030	19641	182877,1	9,311	0,442

5. Характеристики зависимости однородных цепей

В работе представлены, введенные в [6, 7], характеристики зависимости пары однородных цепей (j -й и l -й), взятых из состава данной неоднородной цепи. Также, приведены вычисленные значения одного из данных коэффициентов для некоторых слов в составе литературных произведений.

На рис. 3 представлено несколько диаграмм цепей; в полной неоднородной цепи выделена «бинарно-однородная» цепь (далее бинарная), в состав которой включены только те пары разных знаков j и l сравниваемых однородных цепей, которые определяются знаками l , «ближайшими справа» относительно знаков j . Предполагается, что наличие бинарной цепи свидетельствует о такой причинно-следственной связи цепей, при которой хотя бы некоторые элементы (следствия) l -й цепи расположены вслед за событиями (причинами) j -й цепи на определенных интервалах. Применительно к реальным текстам в диаграммах закрашенные по-разному элементы представляют разные слова.

Пространственная связь такого типа названа «причинной зависимостью» событий одной l -й цепи от другой j -й цепи.

Обозначим: $\Delta(l/j)_i$ – интервал между i -ми ближайшими справа знаками l по отношению к знакам j ; $n(l/j)$ – число пар знаков j и l , связанных интервалами $\Delta(l/j)_i$. На диаграмме: $\Delta(l/j)_1 = \Delta(l/j)_3 = 1$; $\Delta(l/j)_2 = 2$; $\Delta(l/j)_4 = 3$. Определим **среднее геометрическое** значение $\Delta(l/j)_g$ всех $n(l/j)$ установленных интервалов (бинарной цепи) $\Delta(l/j)_i$ между «смежными» элементами двух однородных цепей.

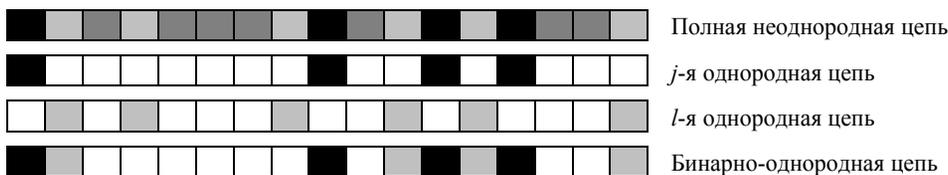


Рис. 3. Диаграммы цепей

Интервалы только между теми знаками выделенной однородной цепи l , которые являются «ближайшими справа» относительно знаков цепи j , обозначены $\Delta(l/j)_i$. В примере: $\Delta(l/j)_1 = 9$; $\Delta(l/j)_2 = 2$; $\Delta(l/j)_3 = 4$; $\Delta(l/j)_4 = 1$.

Определим среднее геометрическое значение $\Delta(l/j)_g$ интервалов $\Delta(l/j)_i$ между выделенными элементами l_j в данной однородной цепи. При условии $\Delta(l/j)_g > \Delta(l/j)_g$, разность $\nu(l/j) = 1 - \Delta(l/j)_g / \Delta(l/j)_g$ представляет «избыточность» l -й цепи, зависимой от j -й. Некоторые элементы такой цепи «связаны справа» с элементами j -й цепи. В противном случае, при $\Delta(l/j)_g \leq \Delta(l/j)_g$, данная разность свидетельствует об отсутствии избыточности.

С учётом избыточности $\nu(l/j)$ и величины $n(l/j)/n_l$ введен коэффициент частичной зависимости $K_1(l/j)$ l -й однородной цепи (от j -й цепи). Отметим, что отношение $n(l/j)/n_l$ представляет собой условную вероятность события, состоящего в появления пары знаков l и j , связанных причинной зависимостью, от появления знаков цепи l . Если все элементы l -й цепи связаны интервалами $\Delta(l/j)_i$ с элементами j -й однородной цепи $n(l/j) = n_l$, то оценка частичной зависимости становится оценкой полной зависимости и определяется равенством $K_1(l/j) = \nu(l/j)$. Если, кроме отмеченного, числа выделенных элементов сравниваемых однородных цепей равны $n_l = n_j$, то такие цепи считаются полностью взаимозависимыми. Причинная зависимость названа установленной, определенной или закономерной, если размеры всех интервалов $\Delta(l/j)_i$ бинарной цепи задаются определенным операционным преобразованием. В противном случае отмечается неопределенная причинная зависимость. Частный случай определенной зависимости назван фиксированной причинной зависимостью. При этом «следование за» в бинарной цепи представлено равными интервалами $\Delta(l/j)_i = \Delta(l/j)_g = const, i = 1, 2, \dots, n(l/j)$. Предельный случай определенной зависимости назван непосредственной причинной зависимостью; при этом бинарная цепь задается единичными интервалами $\Delta(l/j)_i = 1, i = 1, 2, \dots, n(l/j)$.

Также определена степень зависимости $K_2(l/j)$ одной цепи от другой, с учетом величины $2n(l/j)/(n_j + n_l)$, характеризующей «полноту её участия» в составе обеих однородных цепей.

Если не требуется учитывать индивидуальную зависимость l -й цепи от j -й, то вычисляется (средний геометрический) коэффициент взаимной зависимости $K_3(j, l)$.

Наличие «сильной» причинной зависимости может быть использовано для «сжатия» информации при описании взаимного расположения элементов массива данных.

В табл. 4 представлена причинная зависимость K_1 однородных последовательностей некоторых слов, рассматриваемых в составе повести «Пиковая дама» А.С. Пушкина. Полная матрица причинной зависимости слов имеет большой размер, поэтому были выбраны только некоторые пары, по признаку относительно большого значения коэффициента.

Для каждой пары однородных цепей j и l в таблицах приведены отношение $n(l/j)/n_j$, избыточность $v(l_j)$ и определяемый ими коэффициент частичной зависимости K_1 .

Таблица 4

Значения коэффициента причинной зависимости (K_1) для некоторых пар слов из текста повести А.С. Пушкина «Пиковая дама» (в первом столбце слова-причины)

	$\frac{n(l/j)}{n_j}$	$v(l_j)$	K_1									
	Выигрыш			Долг			Игроки			Деньги		
карты	0,900	0,957	0,862	0,858	0,820	0,703	1,000	0,681	0,681	0,243	0,837	0,204
	Вечер			Дом			Года			День		
время	0,750	0,854	0,640	0,800	0,849	0,679	0,715	0,841	0,601	0,750	0,797	0,598
	Игроки			Лица			Люди			Долг		
гости	0,625	0,847	0,530	0,556	0,923	0,513	0,715	0,628	0,449	0,715	0,644	0,460
	Люди			Ужас			Смотрит			Слово		
старуха	1,000	0,903	0,903	0,910	0,906	0,824	0,875	0,932	0,816	0,924	0,876	0,809

Как видно из представленных данных, коэффициент пространственной зависимости можно использовать для определения контекстуальной емкости понятий, придаваемой им автором соответствующего произведения.

6. Числовые характеристики строя для нуклеотидных цепей

Описанные выше формализмы апробированы при компьютерном исследовании рибосомальных РНК 18 организмов, представляющих два царства жизни – прокариота и эукариота [8]. Характеристика g полного неоднородного строя отдельной РНК вычислялась на основе четырех его характеристик однородных цепей: гуанина G, цитазина C, тимина T, аденина A.

Названия исследуемых организмов (упорядоченных по возрастанию величины средней удаленности g) получили следующие номера: 1 – *mus musculus domesticus*, 2 – *caiman crocodylus*, 3 – *canis familiaris*, 4 – *gallus gallus*, 5 – *amia calva*, 6 – *homo sapiens* (человек), 7 – *thermotoga thermarum*, 8 – *thermus thermophilus*, 9 – *ixodes persulcatus*, 10 – *ornithodoros moubata*, 11 – *pediculus humanus capitis*, 12 – *musca domestica*, 13 – *streptococcus pyogenes*, 14 – *bacillus anthracis*, 15 – *borrelia burgdorferi* B31, 16 – *candidatus nitros. maritimus*, 17 – *mycoplasma pneumoniae*, 18 – *neisseria gonorrhoeae*.

На числовой оси (рис. 4) организмы расположены в соответствии со значениями их характеристики g – средней удаленности элементов в нуклеотидной цепи. Как видно, эукариоты (позвоночные) образовали группу организмов (слева на оси), имеющих наименьшие значения средней удаленности. Прокариоты (бактерии) образовали группу организмов (справа на оси), имеющих наибольшие значения средней удаленности. Эукариоты (беспозвоночные) расположены между этими крайними на оси группами, примыкая к бактериям. Факт выделения из группы прокариот (по характеристике g) двух бактерий (7 и 8), значительно отличающихся

ся условиями существования от других, исследуемых в данной работе, требует проведения дальнейших исследований.

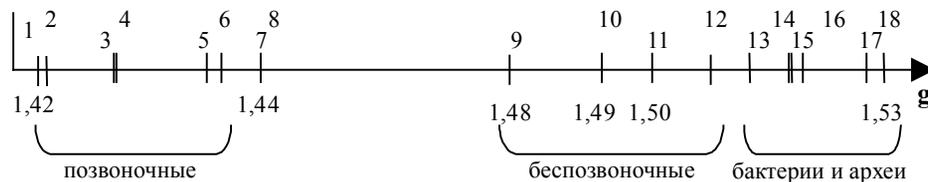


Рис. 4. Расположение номеров организмов на оси средней удалённости нуклеотидов

Заключение

В работе представлен инструментарий для описания и анализа нового абстрактного объекта, названного строем цепи событий. Вычисленные значения характеристик строя однозначно отображают оригинальные построения выбранных нуклеотидных цепей и литературных текстов. Это дает основания для проведения исследований больших множеств знаковых последовательностей разной природы с помощью средств анализа строя цепи.

ЛИТЕРАТУРА

1. Гуменюк А.С. О формализме «строение знаковой последовательности» и его использовании для анализа, сравнения, преобразования и распознавания лингвистических текстов // Четвертый сибирский конгресс по прикладной и индустриальной математике: Тез.докл. Новосибирск: Изд-во Института математики СО РАН, 2000. Ч. 3. С. 84–85.
2. Gumenyuk A., Kostyshin A., Simonova S. An approach to the research of the structure of linguistic and musical texts // *Glottometrics*. 3 (2002). Liideschtd RAM Verl. S. 61–89.
3. Гуменюк А.С., Богорад Д.М. Об исчислении взаимного расположения элементов в массивах данных // Труды РНТОРЭС им. Попова, серия: научная сессия, посвященная Дню радио. 2004. Вып. LIX. № 2. С. 114–116.
4. Мазур М. Качественная теория информации. М.: Мир, 1974. 249 с.
5. Орлов Ю.К. Частотные структуры конечных сообщений в некоторых естественных информационных системах: дис. ... канд. наук. Тбилиси: Изд-во Тбилисского университета, 1974.
6. Гуменюк А.С. О средствах анализа взаимного расположения компонентов знаковой последовательности. // Материалы III Международного технологического конгресса. Омск: ОмГТУ, 2005. Ч. II. С. 48–52.
7. Гуменюк А.С., Морозенко Е.В. О характеристике зависимости однородных цепей в составе нуклеотидной последовательности / Нейроинформатика, её приложения и анализ данных: Материалы XVII Всероссийского семинара, 2–4 октября 2009 г. / под ред. А.Н. Горбана, Е.М. Миркеса. Красноярск: ИВМ СО РАН, 2009. С. 53–56.
8. Гуменюк А.С., Шпынов С.Н., Морозенко Е.В., Родионов И.Н. Пример применения средств анализа для сравнения строя нуклеотидных последовательностей и кластеризации организмов // Нейроинформатика, ее приложения и анализ данных: Материалы XVI Всероссийского семинара / под ред. А.Н. Горбана, Е.М. Миркеса. Красноярск: ИВМ СО РАН, 2008. С. 166–170.

Гуменюк Александр Степанович

Морозенко Евгений Витальевич

Родионов Игорь Николаевич

Омский государственный технический университет

E-mail: gumas45@mail.ru;

brunet1500@rambler.ru; goruha@gmail.com

Поступила в редакцию 17 ноября 2010 г.