Исаев Андрей Станиславович
канд. техн. наук, доцент
Новомосковский институт (филиал) ФГБОУ ВО «Российский химико-технологический университет им. Д.И. Менделеева»
г. Новомосковск, Тульская область

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ АНАЛИЗА ЧАСТОТНЫХ H-РАСПРЕДЕЛЕНИЙ

Аннотация*: статья посвящена исследованию общих закономерностей видового разнообразия сложных систем (ценозов). Предложены алгоритмически и реализованы программно (Matlab) процедуры получения количественных оценок H-распределения в различных формах. Исходным материалом для построения моделей являются устойчивые словосочетания в актуальной нормативной литературе электроэнергетики.*

Ключевые слова*: видовое разнообразие, ценоз, математическое моделирование, ранговый анализ, семиотика, H-распределение, Matlab.*

Введение. Выделение электрики, включающей в себя наряду с классической электротехникой и закономерности построения системы электроснабжения, приводит к необходимости к модернизации понятийного и терминологического аппарата. Для этого актуально формирование профессионального словаря электрики как отрасли науки. С этой целью на первом этапе проанализирована соответствующая учебная, нормативная и справочная литература. Ранее установлена принципиальная тождественность видового разнообразия в биологии, экономике, лингвистике, теоретической физике [1].

Частотный анализ повторяемости слов в тексте используется для определения оптимального объема словаря, а ранее применялся в учебном процессе при оценке корректности выполнения учебных работ на основании общих закономерностей и соотношений между массовым и серийным, уникальным и повсеместным [2]. Известно, что построение текста основано на распределении словоформ негауссового типа, когда средняя повторяемость дефиниции не яв-

ляется наиболее вероятной, распределение повторяемости частот характеризуется высокой степенью асимметрии и эксцесса, ростом дисперсии при увеличении объема выборки [3].

При анализе частотных распределений в ранговой форме используется зависимость:

$$\Lambda(r) = \frac{B}{r^{\beta}}, \quad (1)$$

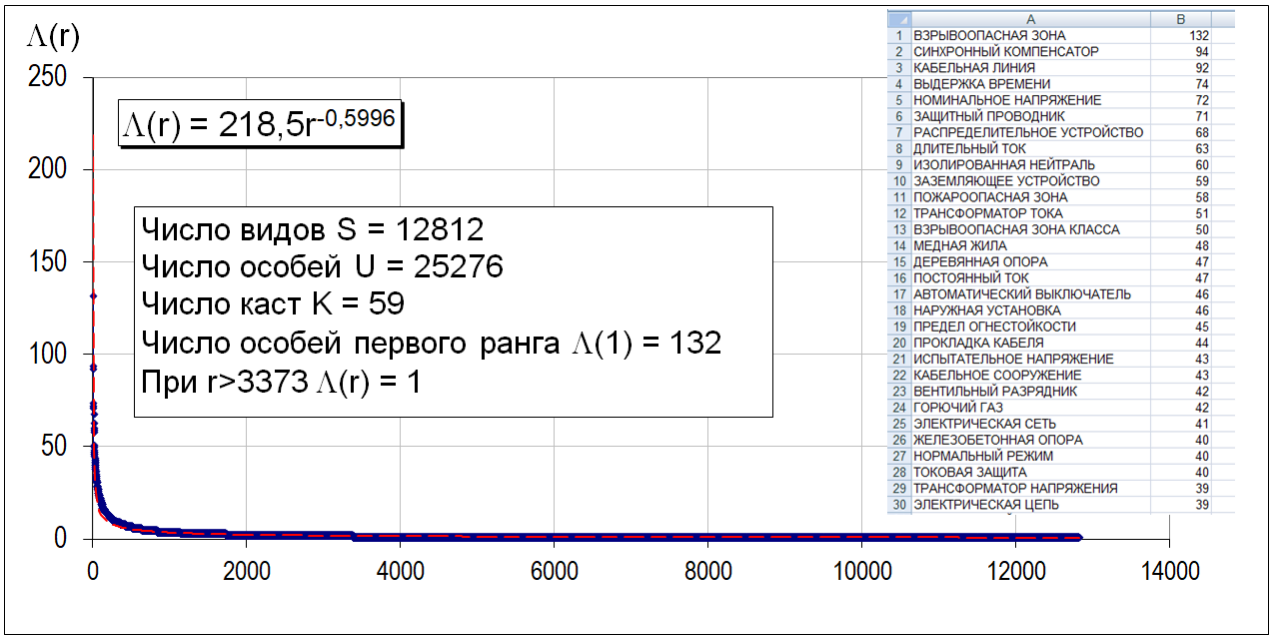
где r – ранг, $r = \overline{1,S}$; S – число видов (объем словаря); Λ (r) – количество слов ранга r (их сумма определяет число особей – объем текста U); B, β – коэффициенты аппроксимации.

При объединении видов с одинаковой частотой встречаемости в касты ранговое распределение преобразуется к видовому:

$$\Omega(x) = \frac{A}{x^{1+\alpha}}, \quad (2)$$

где x – непрерывный аналог численности вида; Ω (x) – количество видов с численностью x; A, α – коэффициенты аппроксимации.

В качестве эмпирического материала рассмотрим современную нормативную документацию в области электроэнергетики. На рис. 1 приведено распределение профессиональных устойчивых словосочетаний актуальной редакции ПУЭ [4]. Наиболее распространенным является «взрывоопасная зона», объем текста составляет U = 25276, словаря – U = 12812.



	A	B
1	ВЗРЫВООПАСНАЯ ЗОНА	132
2	СИНХРОННЫЙ КОМПЕНСАТОР	94
3	КАБЕЛЬНАЯ ЛИНИЯ	92
4	ВЫДЕРЖКА ВРЕМЕНИ	74
5	НОМИНАЛЬНОЕ НАПРЯЖЕНИЕ	72
6	ЗАЩИТНЫЙ ПРОВОДНИК	71
7	РАСПРЕДЕЛИТЕЛЬНОЕ УСТРОЙСТВО	68
8	ДЛИТЕЛЬНЫЙ ТОК	63
9	ИЗОЛИРОВАННАЯ НЕЙТРАЛЬ	60
10	ЗАЗЕМЛЯЮЩЕЕ УСТРОЙСТВО	59
11	ПОЖАРООПАСНАЯ ЗОНА	58
12	ТРАНСФОРМАТОР ТОКА	51
13	ВЗРЫВООПАСНАЯ ЗОНА КЛАССА	50
14	МЕДНАЯ ЖИЛА	48
15	ДЕРЕВЯННАЯ ОПОРА	47
16	ПОСТОЯННЫЙ ТОК	47
17	АВТОМАТИЧЕСКИЙ ВЫКЛЮЧАТЕЛЬ	46
18	НАРУЖНАЯ УСТАНОВКА	46
19	ПРЕДЕЛ ОГНЕСТОЙКОСТИ	45
20	ПРОКЛАДКА КАБЕЛЯ	44
21	ИСПЫТАТЕЛЬНОЕ НАПРЯЖЕНИЕ	43
22	КАБЕЛЬНОЕ СООРУЖЕНИЕ	43
23	ВЕНТИЛЬНЫЙ РАЗРЯДНИК	42
24	ГОРЮЧИЙ ГАЗ	42
25	ЭЛЕКТРИЧЕСКАЯ СЕТЬ	41
26	ЖЕЛЕЗОБЕТОННАЯ ОПОРА	40
27	НОРМАЛЬНЫЙ РЕЖИМ	40
28	ТОКОВАЯ ЗАЩИТА	40
29	ТРАНСФОРМАТОР НАПРЯЖЕНИЯ	39
30	ЭЛЕКТРИЧЕСКАЯ ЦЕПЬ	39

Рис. 1. Ранговое распределение словосочетаний ПУЭ

Развитие информационных технологий позволяет использовать не языки программирования, а современные пакеты прикладной математики. В [5] показана программная реализация в Matlab, позволяющая получать количественные оценки устойчивых распределений негауссового типа, не прибегая в явном виде к параметрическим методам (прежде всего, речь идет о методе наименьших квадратов).

Методы и результаты. Исходные данные считаны из файла формата MS Excel функцией *xlsread*. Для ранжирования используется функция *sort* (параметр *descend* соответствует сортировке по убыванию, по умолчанию массив сортируется по возрастанию). Получение массива видов (из массива особей), массива каст (из массива видов) получено функцией *unique*, частотные характеристики – *hist*, число особей (видов, каст) как размерность массива – *length*. Построение графиков выполнено функцией *plot* с соответствующими настройками (опция *figure* перед *plot* необходима для вывода изображения в новом окне, по умолчанию – закрывается предыдущее) (результаты приведены на рис. 2).

Результаты могут быть сохранены в виде файлов (например, функцией записи в файл матрицы A – *writematrix*(A)) или переданы в другую программы стандартными средствами Windows.

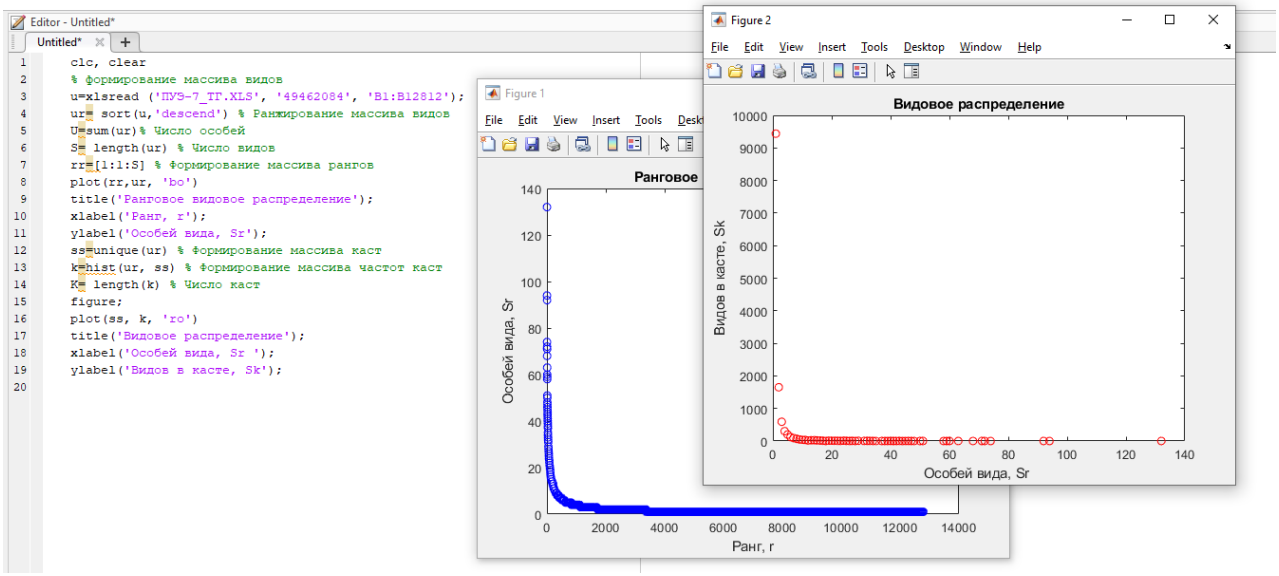


Рис. 2. Результаты частотного анализа словосочетаний ПУЭ

Учитывая недостатки линеаризованных методов, оценка видового распределения выполнена решением оптимизационной задачи. Расчет коэффициентов зависимости выполнен с помощью *nlinnfit* (нелинейная регрессия), вид зависимости задан непосредственно в виде пользовательской функции f=@(x, u). Оценка точности аппроксимации выполнена по критерию *MAPE* (средняя относительная погрешность) – рис. 3. Помимо этого может использоваться функция *lsqcurvefit* (оптимизация параметрических нелинейных функций), она приведена как комментарий, результаты ее использования для данной задачи тождественны *nlinnfit*.

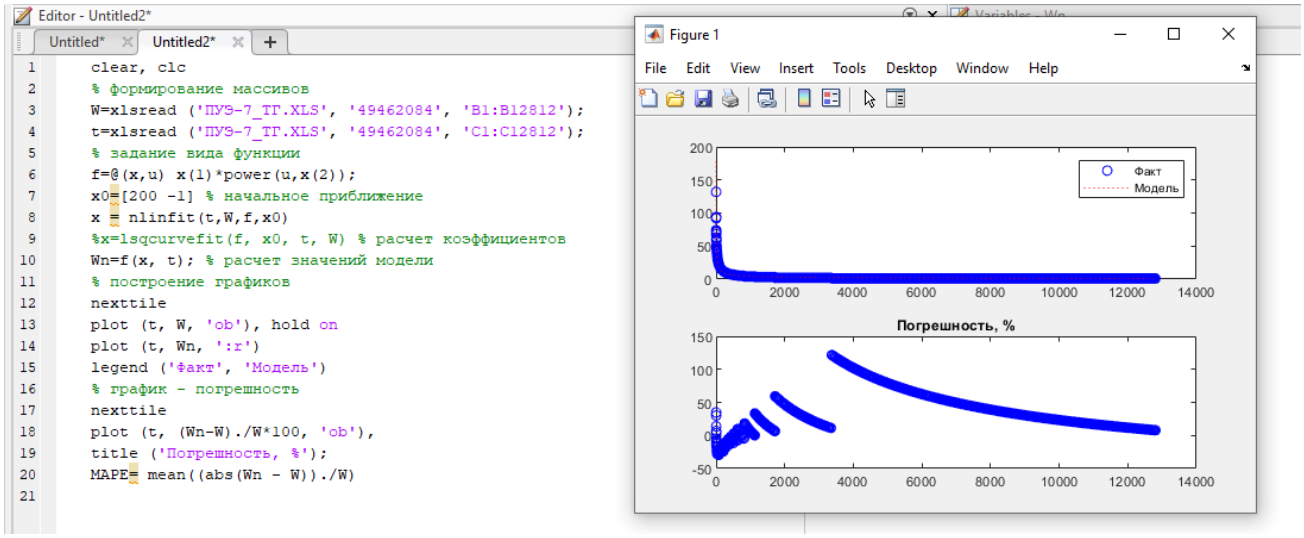


Рис. 3. Результаты формализации частотного анализа ПУЭ

Обсуждение. Работа не претендует на широкие выводы в сфере лингвистики (подобно [6], где выполнена модернизация распределения Ципфа, соответствующего (1) при равенстве рангового коэффициента единице). Решена более узкая задача – представлено инструментальное средство, позволяющее получать частотные H-распределения в ранговой и видовой формах.

Работы, основанные на методе рангового анализа (в частности, [7]), используют параметрическое распределение (а в качестве программного средства – MathCAD). Эта задача также решается в Matlab (функция *sort*), а использование программного средства инвариантно. Хотя представляется, что Matlab обладает большими функциональными возможностями.

Выводы. Показана возможность применения для анализа видового разнообразия пакетов прикладной математики (Matlab). Программный код приведен полностью (рис. 2). Корректность результатов подтверждается анализом эмпирического материала и соответствием известным теоретическим положениям [1]. Реализовано программно оценка параметров H-распределения (рис. 3), что позволяет, в частности, отказаться от однопараметрических моделей с фиксированной первой точкой.

Продолжение работы представляется в уточнении методов расчета количественных характеристик. Основным направлением является расчет параметров на основе нейросетевого алгоритма и необходимо использование методов оценок при несоответствии отклонений модельных значений от эмпирических нормальному закону.

Список литературы

1. Кудрин Б.И. Введение в технетику / Б.И. Кудрин. – 2-е изд., перераб. и доп. – Томск: Изд. Томского гос. ун-та, 1993. – 552 с.

2. Гурина Р.В. Ранговый анализ в оценке валидности олимпиадных заданий / Р.В. Гурина, Е.В. Морозова, В.В. Кошева // Профессиональное образование в современном мире. – 2020. – Т. 10. №4. – С. 4302–4309. – DOI 10.20913/2618-7515-2020-4-14. – EDN JYCZMB.

3. Ковригина Л.Ю. Негауссовое моделирование лексико-статистической структуры вариативного текста (на примере «Сказания о Мамаевом побоище»): специальность 10.02.21 «Прикладная и математическая лингвистика»: дис. ... канд. филол. наук / Ковригина Любовь Юрьевна. – 2015. – 356 с. – EDN MDVUBC.

4. Исаев А.С. Особенности формализации частотных H-распределений / А.С. Исаев, Н.А. Пряхина // Актуальные проблемы науки и образования: сборник материалов III Международной научно-практической конференции (Москва, 14 декабря 2023 года). – М.: Алеф, 2023. – С. 283–290. – DOI 10.26118/1453.2023.76.30.004. – EDN SWDHKI.

5. Правила устройства электроустановок. – М.: КНОРУС, 2010. – 488 с.

6. Маслов В.П. О законе Ципфа и ранговых распределениях в лингвистике и семиотике / В.П. Маслов, Т.В. Маслова // Математические заметки. – 2006. – Т. 80. №5. – С. 718–732. – EDN HVSUKD.

7. Гнатюк В.И. Определение потенциала энергосбережения объектов припортового электротехнического комплекса в рамках развития интеллектуальных энергетических систем / В.И. Гнатюк, О.Р. Кивчун, А.Я. Яфасов // Морские интеллектуальные технологии. – 2017. – №3–1 (37). – С. 142–148. – EDN XROPNZ.