

1 Создание символьной последовательности

- Скачивание значений свеч с сайта fnam.
- Соединение нескольких файлов в один для большей последовательности.
- Рассчёт для каждой свечи $s(i) = x_2(i) - x_1(i)$, где x_2 - значение в конце свечи, x_1 - в начале.
- Сортировка числовой последовательности, при этом сохраняя исходную.
- Деление отсортированной последовательности на 20 равных интервалов.
- Поиск наиболее часто повторяющегося числа. Если его больше, чем N/l , где N - длина s , l - размер алфавита, то сдвиг границ интервала.
- Присваивание каждому интервалу буквы.
- Возврат к s , замена каждого числа на соответствующую букву.

2 Создание набора случайных последовательностей

- Копировать символьную последовательность.
- Создать random engine.
- Shuffle копированной последовательности.
- Повторить 200 раз.

3 Создание набора случайных матриц

- Создать матрицу размером $l * n$. Здесь n - длина предполагаемого периода.
- Заполнить матрицу: на каждую позицию $m(i, j)$ равновероятно поставить 0 или 1 ($i = 1, \dots, l; j = 1, \dots, n$).

- Далее, если матрица первая, просто поместить её в набор.
- Если эта матрица не первая, то для каждой имеющейся в наборе матрицы рассчитать "меру различия" между сгенерированной матрицей и каждой другой из набора по формуле $x(M_1, M_2) = \sqrt{4I(M_1, M_2)} - \sqrt{2df - 1}$.
Информация $I(M_1, M_2)$ находится по формуле: $I(M_1, M_2) = \sum_{j=1}^n I_j(M_1, M_2)$, где

$$I_j(M_1, M_2) = \sum_{i=1}^l m_1(i, j) \ln(m_1(i, j)) + \sum_{i=1}^l m_2(i, j) \ln(m_2(i, j)) -$$

$$- \sum_{i=1}^l (m_1(i, j) + m_2(i, j)) \ln(m_1(i, j) + m_2(i, j)) +$$

$$+ (s_1(j) + s_2(j)) \ln(s_1(j) + s_2(j)) - s_1(j) \ln(s_1(j)) - s_2(j) \ln(s_2(j)),$$
где $s_k(j) = \sum_{i=1}^l m_k(i, j)$, а $m(i, j)$ - элемент матрицы на (i, j) позиции.
Величина df , в свою очередь, равна $(n - 1)(l - 1)$, так как среди I_j ровно $j - 1$.
- Если данная мера больше 1 для всех матриц из набора, то матрица добавляется в набор. Если хотя бы для одной меньше - то нет.
- Заполнить набор таким образом 10^8 матрицами.

4 Создание модифицированной случайной матрицы и расчет функции подоби́я

- Создадим m' . Для этого сначала найдём для матрицы m значения $A = \sum_{i=1}^{20} \sum_{j=1}^n m(i, j)^2$ и $B = \sum_{i=1}^{20} \sum_{j=1}^n m(i, j) * p(i)$, где $p(i) = n(i)/N$, где N - длина последовательности, $n(i)$ - количество i -го символа в s .
- Построим матрицу m' . Для этого сначала найдем проекцию на "гиперплоскость". Получим $m_1(i, j) = m(i, j) - D * \frac{p(i)}{\sqrt{\sum_i n * p(i)^2}}$, где $D = \frac{\langle m(i, j), p(i) \rangle - B}{\sqrt{\sum_i n * p(i)^2}}$.
- Затем найдем проекцию центра на эту плоскость. Она равна (для всех $m(i, j) = 0$): $m_2(i, j) = \frac{-B * p(i)}{\sum_i n * p(i)^2}$.
- Далее рассчитаем точку пересечения прямой, проходящей через эти две точки с "гиперсферой" в параметрическом виде: $l_{ij}(t) = m_1(i, j) + t * (m_2(i, j) - m_1(i, j))$ для любого $m = 1, \dots, 20 * n$.

- Найдём точку пересечения прямой с гиперсферой. Для этого подставим координаты прямой в уравнение гиперсферы: $\sum_{i=1}^{20} \sum_{j=1}^n (m_1(i, j) + t * (m_2(i, j) - m_1(i, j)))^2 = A$. Это квадратное уравнение относительно t . Решив его, возьмём меньший действительный корень. Это будет $t = t_0$.
- Найдём матрицу m' : $m'(i, j) = m_1(i, j) + t_0 * (m_2(i, j) - m_1(i, j))$.
- Рассчитаем для m' итеративно $F(i, j) = \max \begin{cases} 0 \\ F(i - 1, j - 1) + m'(s(i), k) \\ F(i, j - 1) - d \\ F(i - 1, j) - d \end{cases}$
где $s(i)$ - номер буквы на i -й позиции в S , i, j изменяются от 1 до N , $k = j - n * \text{int}((j - 0.1)/n)$, d - штраф за вставку/делецию. Размерность F будет $N * N$, где N - длина последовательности.
- Найдём в матрице F максимальный элемент $\max F$ и его координаты i, j .
- Если $i_m - i_0$ не входит в интервал (540, 660) перевыберем B и повторим вычисление m' . Если разность больше интервала, то уменьшаем B . Если меньше, то увеличиваем. При этом B должно лежать в интервале $(-15, 0)$.
- Создадим матрицу F' . Для этого из каждой клетки матрицы F в F' вписать, откуда он был получен. То есть для $0 \rightarrow (i, j)$, для $F(i - 1, j - 1) + m'(s(i), k) \rightarrow (i - 1, j - 1)$, для $F(i, j - 1) - d \rightarrow (i, j - 1)$, а для $F(i - 1, j) - d \rightarrow (i - 1, j)$.
- Выравниваем последовательность S с последовательностью индексов, соответствующих столбцам матрицы следующим образом. Идя от максимального значения $\max F$ по диагонали, выравнивать символ алфавита напрямую. Идя вверх - удаляем символ (устанавливаем * в последовательность S). Идя влево - добавляем символ (устанавливаем * в последовательность индексов).

5 Поиск максимального $\max F$

- Найти значение $\max(m')$ при котором среди всех случайных матриц для S достигается максимум F при выполнении всех условий.

- Прodelать процедуру поиска $\max F$ для всех матриц, "мера различия" которых с матрицей $m - > m'$ (пункт 3) находится от 0 до 1.
- Выбрать максимальное из всех $\max F$ и соответствующую этому $\max F$ матрицу T .

6 Поиск спектра

- Прodelать вышеперечисленные шаги (4 – 5) для набора случайных последовательностей из (2) и полученной матрицей T . Получить спектр значений $\{\max F_i\}$.
- Получить значение $Z(n) = \frac{\overline{\max F(n)} - \max F_s(n)}{\sqrt{D(\max F(n))}}$, где n - длина периода, а s обозначает исходную последовательность.
- Рассчитать шагами (3 – 5) значение Z для различных длин периодов.