КОЛМОГОРОВСКАЯ СЛОЖНОСТЬ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Подготовили
Досова Анна
Лемнёв Вадим

Шенноновская энтропия. Проблемы применения к индивидуальным объектам

- Энтропия — мера неопределенности некоторой системы, например, какого-либо эксперимента, который может иметь разные исходы.

- Недостатки подхода Шеннона:

  Если применить энтропию Шеннона к текстам, то выходит, что количество информации в тексте зависит только от частот символов, но не зависит от их порядка.

  При таком подходе получается, что два текста: исходный и отсортированный по символам – содержат одинаковое количество информации.

Новизна теории сложности Колмогорова

- В начале 1960-х гг. Колмогоров, Соломонов, Левин и другие ученые сформулировали способ измерения количества информации в конкретных объектах (строках), а не случайных величинах.

- Основная идея теории сложности Колмогорова в том, что сложность строки определяется длиной наикратчайшей компьютерной программы, способной ее выдать.

Колмогоровская сложность

- Пусть ***F*** – алгоритм, аргументами и результатами которого являются двоичные слова.
- Слово ***x*** является описанием слова ***y***, если ***F(x) = y***.
- Сложность слова ***y*** относительно ***F*** – длина его кратчайшего описания:

$$\mathbf{K_F(y) = \min\{|x|: F(x) = y\}.}$$

Колмогоровская сложность

- Пусть у нас есть два способа описания слова *F* и *U*.
- Сложности двух алгоритмов отличаются не более, чем на константу, зависящую от *F* и *U*, но не зависящую от слова *y:*

$$| K_F(y) - K_U(y) | \leq c_{F,U}$$

- Говорят, что *U* лучше *F,* если существует такая константа $c_{F,U}$, что для всех *y* выполняется следующее неравенство:

$$K_U(y) \leq K_F(y) + c_{F,U}$$

- При этом *U* является оптимальным способом описания слова *y.*
- Сложность K_U относительно *U* называют ***колмогоровской сложностью***.

Колмогоровская сложность

- Колмогоров и Левин доказали следующую формулу:

$$K(xy) = K(x) + K(y|x) + O(\log n)$$

- Разность *K(y) - K(y|x)* показывает, насколько знание слова ***x*** упрощает описание слова ***y***.
- Понятие условной сложности позволяет ответить на вопрос:

Сколько новой информации в ДНК одного организма по сравнению с ДНК другого?

- Аналогичным образом можно узнать, *какой процент информации был потерян при переводе текста на другой язык*. В этом случае нас интересует отношение:

$$\frac{K(\text{оригинал} \mid \text{перевод})}{K(\text{оригинал})}$$

Функция *K* является перечислимой сверху

- Функция K является перечислимой сверху, но не является вычислимой.

- Не существует никакой неограниченной вычислимой нижней оценки для функции K.

- **Теорема:**

Функция $\boldsymbol{K}$ перечислима сверху, причем выполняется следующее неравенство:

$$|\{\boldsymbol{y} \mid \boldsymbol{K}(\boldsymbol{y}) < \boldsymbol{n}\}| < \boldsymbol{2^n} \quad \text{при всех } \boldsymbol{n}.$$

Невычислимость Колмогоровской сложности.
Идея доказательства от противного. Парадоксы

- Не существует такого алгоритма, который по данному слову ***у*** определяет его колмогоровскую сложность.

- ***Парадокс интересных чисел*** связан с утверждением, что все числа интересные:

1 – это ***первое*** число. ***2*** – ***первое чётное*** число. ***3*** – ***первое нечётное простое*** число. ***4*** – интересное число, потому что ***4 = 2 × 2 и 4 = 2+2***.

В какой-то момент мы можем встретить число без интересных свойств. И мы можем назвать это число ***первым неинтересным номером*** – но это само по себе уже интересное свойство. В итоге неинтересные числа тоже оказываются интересными.

Невычислимость Колмогоровской сложности.
Идея доказательства от противного. Парадоксы

- ***Парадокс Берри*** связан с описанием больших чисел:

Чем больше слов мы используем, тем большее число мы можем описать. Рассмотрим число, описываемое следующей фразой:

«Самое маленькое число, которое нельзя описать меньше, чем пятнадцатью словами»

Для описания числа требуется ***15***, ***16*** или даже больше слов. Его нельзя описать ***12***, ***13*** или ***14*** словами. Однако, вот в чём проблема: приведённая выше фраза описывает это число при помощи ***10*** слов. Таким образом, приходим к противоречию.