

Биномиальное сжатие бернуллиевского источника информации

Костель С.В., *ассист.*; Кобзар А.М., *студ.*
Сумский государственный университет, г. Сумы

Бернуллиевский источник (БИ) информации порождает символы из некоторого алфавита A с независимыми вероятностями их появления. Невысокая сложность такой модели обуславливает широкое ее использование в задачах сжатия данных. Метод биномиального нумерационного сжатия (БНС) информации был описан в работе [1]. Недостатком его является ориентированность на сжатие двоичной информации. Поэтому актуальной является задача разработки метода БНС для сжатия недвоичной информации, порожденной БИ.

Для решения данной задачи можно воспользоваться моделью векторного представления сообщений [2]. Даная модель для каждого символа в сжимаемом сообщении ставит в соответствие двоичный вектор. Вектор содержит единицы в тех позициях, которые соответствуют позициям символа в сообщении. Полученное множество двоичных векторов сжимается по методу БНС. В результате сжатия двоичный вектор представляется в виде информации о числе единиц в векторе (количестве символов в сообщении) и номера биномиального числа. Сжатое сообщение, порожденное БИ, будет представлять собой множество уникальных символов, содержащихся в сообщении, и соответствующее им множество сжатых двоичных сообщений.

К достоинствам биномиального сжатия данных, порожденных БИ, можно отнести низкую сложность реализации и достаточно высокую эффективность сжатия сообщений при незначительной длине самих сообщений. Сжатые данные являются независимыми друг от друга, что позволяет выполнять операции поиска в сжатом виде и частичного восстановления сжатой информации. Предложенный метод сжатия можно использовать для сжатия информации и базах и хранилищах данных.

1. И.А. Кулик, С.В. Костель, *АСУ и приборы авт.*, No149, 66 (2009).
2. И.А. Кулик, В.В. Гриненко, С.В. Костель, Е.М. Скордина, *Вестник СумГУ*, No3, 60 (2013).