

КОНЦЕПЦИЯ ФОРМИРОВАНИЯ КОДОГРАММЫ ДЛЯ МЕТОДА СТАТИСТИЧЕСКОГО КОДИРОВАНИЯ ДАННЫХ С ПРИМЕНЕНИЕМ КЛАСТЕРИЗАЦИИ ПО ПРИЗНАКУ КОЛИЧЕСТВА СЕРИЙ ЕДИНИЦ

Баранник В.В.¹, Тупица И.М.¹, Баранник В.В.²

¹Кафедра боевого применения и эксплуатации АСУ, Харьковский национальный университет Воздушных Сил имени Ивана Кожедуба, Харьков, Украина, E-mail: vybar_off@gmail.com

²Харьковский национальный университет радиозлектроники, Харьков, Украина, E-mail: valera462000@gmail.com

Аннотация – Разрабатывается концепция формирования кодограммы для метода статистического кодирования данных информационного ресурса с применением кластеризации по количественному признаку – признаку количества серий единиц. Инструментом кластеризации является внутренняя реструктуризация данных, которая заключается в выявлении закономерностей во внутренней структуре элементов сообщения. Анализируются требования, которые предъявляются к технологии формирования кодограммы для исследуемого метода статистического кодирования данных информационного ресурса.

Ключевые слова – кластеризация, признак количества серий единиц, кодограмма.

I. Введение

Использование современных алгоритмов сжатия данных информационного ресурса активно связано с понятием «реструктуризации» данных [1]. Под понятием реструктуризации данных подразумевается более выгодное представление кодируемых данных. В современных алгоритмах сжатия данных информационного ресурса активное применение нашли методы внешней реструктуризации. Однако анализ применения методов внешней реструктуризации данных показал, что они имеют ряд существенных недостатков, которые изложены в трудах [1 – 3].

Для устранения недостатков методов внешней реструктуризации данных информационного ресурса был предложен принципиально новый подход – внутренняя реструктуризация, которая заключается в выявлении закономерностей во внутренней структуре элементов сообщения по количественному признаку [4]. Это связано с тем, что применение внутренней реструктуризации данных имеет ряд преимуществ по сравнению с внешней, а именно:

- не нужно проводить никаких преобразований над данными информационного ресурса;
- сокращается время на обработку данных;
- дальнейшая кластеризация элементов сообщения позволяет повысить эффективность статистического кодирования с позиции повышения защиты и сокращения длины на представление информации.

В качестве количественного признака используется признак количества серий единиц (КСЕ). Внутренняя реструктуризация является инструментом для проведения кластеризации данных информационного ресурса.

В статье рассматриваются вопросы, связанные с разработкой концепции формирования кодограммы для метода статистического кодирования с применением кластеризации по признаку количества серий единиц. Проводится анализ требований, которые предъявляются к концепции формирования кодограммы.

II. Основная часть

Анализируя результаты формирования отдельных кодовых конструкций ℓ'_ξ , которые могут присваиваться элементам u_ξ сообщения $U(\theta)$ в процессе статистического коди-

рования в пределах множеств $U(\lambda_i)$, сформированных в процессе кластеризации по признаку λ_i КСЕ, можно сделать следующие выводы:

– кодовые конструкции ℓ'_ξ , которые присваиваются элементам u_ξ одного множества $U(\lambda_i)$, будут обладать свойством межмножественной префиксности. Это значит, что свойство префиксности соблюдается только для случая следующих подряд кодовых конструкций ℓ'_ξ , которые присваиваются элементам u_ξ одного множества $U(\lambda_i)$.

– для следующих подряд кодовых конструкций ℓ'_ξ , которые присваиваются элементам u_ξ разных множеств $U(\lambda_i)$, свойство префиксности не соблюдается.

– для однозначного декодирования кодовой последовательности $L(\theta)$ возникает необходимость в использовании идентификатора множества $U(\lambda_i)$ (разделителя), которому принадлежит тот или иной элемент u_ξ . Это позволит определить начальный разряд $q_{\xi,1}$ каждого кода ℓ'_ξ , при декодировании кодовой последовательности $L(\theta)$.

Анализ результатов статистического кодирования элементов u_ξ сообщения $U(\theta)$ в пределах множеств $U(\lambda_i)$, сформированных в процессе кластеризации элементов u_ξ по признаку λ_i КСЕ, свидетельствует о том, что при разработке концепции формирования кодограммы необходимо учитывать следующие факторы, которые оказывают значимое влияние как на формирование отдельных кодовых конструкций ℓ'_ξ , которые присваиваются элементам u_ξ сообщения $U(\theta)$, так и кодовой последовательности $L(\theta)$ в целом [3-5]:

– трансформация характера закона распределения вероятностей $P(u_\xi)$ появления элементов u_ξ сообщения $U(\theta)$ в результате применения кластеризации по количественному признаку λ_i ;

– трансформация стратегии позиционирования отдельных кодовых конструкций ℓ'_ξ в общей кодовой последовательности $L(\theta)$.

Это значит, что для однозначного декодирования, как отдельных кодовых конструкций ℓ'_ξ , так и кодовой последовательности $L(\theta)$ в целом возникает необходимость в использовании идентификаторов множеств $U(\lambda_i)$ (разделителей), которым принадлежит тот или иной элемент u_ξ .

Для решения вышеуказанной проблемы необходимо разработать концепцию формирования кодограммы таким образом, чтобы она позволила однозначно декодировать кодовую последовательность $L(\theta)$. Для однозначного декодирования кодовой последовательности $L(\theta)$ необходи-

мо, чтобы кодовые конструкции ℓ'_ξ , которые присваиваются элементам u_ξ сообщения $U(\theta)$, несли в себе не только информацию про кодируемый элемент u_ξ , но и информацию про множество $U(\lambda_i)$, в пределах которого кодируется элемент u_ξ сообщения $U(\theta)$. Таким образом, структура

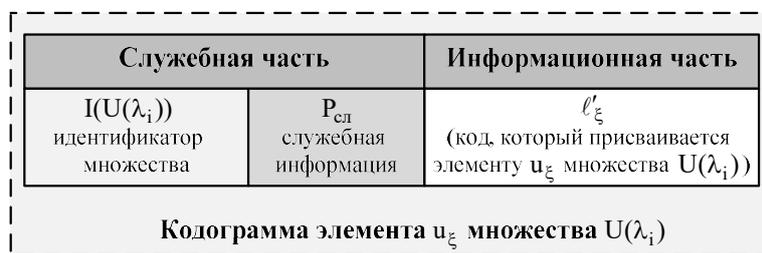


Рис. 1. Структурная схема кодограммы элемента u_ξ множества $U(\lambda_i)$

На рис.1. представлена в общем виде структурная схема кодограммы элемента u_ξ множества $U(\lambda_i)$, которая включает в себя две части:

1. информационная часть кодограммы - значение кода ℓ'_ξ , который присваивается элементу u_ξ множества $U(\lambda_i)$, сформированного в процессе кластеризации элементов u_ξ сообщения $U(\theta)$ по значению признака λ_i КСЕ. Эта часть кодограммы представляет собой семантическое содержание кодируемого элемента u_ξ , т.е. несет смысловую нагрузку.

2. служебная часть кодограммы, которая включает в себя две составляющие:

– идентификатор $I(U(\lambda_i))$ множества $U(\lambda_i)$, которому принадлежит кодируемый элемент u_ξ . Эта часть служебной составляющей кодограммы является инструментом для однозначного декодирования кодовой конструкции ℓ'_ξ , которая присваивается элементу u_ξ в процессе статистического кодирования в пределах сформированного в процессе кластеризации множества $U(\lambda_i)$.

Это значит, что в условиях отсутствия ошибок в процессе обработки и передачи кодов ℓ'_ξ их декодирование производится однозначно, значения элементов u_ξ сообщения $U(\theta)$ восстанавливаются без ошибок. Идентификатор $I(U(\lambda_i))$ множества $U(\lambda_i)$ оказывает значимое влияние на информационную часть ℓ'_ξ кодограммы. Это связано с тем, что при отсутствии информации о множестве $U(\lambda_i)$, которому принадлежит код ℓ'_ξ , присваиваемый элементу u_ξ , злоумышленнику не удастся, в случае несанкционированного доступа, безошибочно декодировать кодовую последовательность $L(\theta)$.

– служебная информация $P_{сл}$ – служебная информация о распределении значений вероятностей $P'(u_\xi)$ появления элементов u_ξ во множестве $U(\lambda_i)$.

Предложенная структура кодограммы, которая формируется в процессе статистического кодирования элемента u_ξ множества $U(\lambda_i)$, обеспечит однозначное декодирование кодовой последовательности $L(\theta)$ и позволит повысить эффективность статистического кодирования данных информационного ресурса с позиции повышения безопасности.

кода ℓ'_ξ , который присваивается элементу u_ξ множества $U(\lambda_i)$ должна состоять из двух частей – информационной и служебной. Структурная схема кодограммы элемента u_ξ сообщения $U(\theta)$ представлена на рис.1.

Идентификатор $I(U(\lambda_i))$ множества $U(\lambda_i)$, которому принадлежит тот или иной кодируемый элемент u_ξ , оказывает значимое влияние, как на служебную часть кодограммы, так и на кодограмму элемента u_ξ множества $U(\lambda_i)$ в целом. Это значит, что однозначное декодирование кода ℓ'_ξ , который присваивается элементу u_ξ множества $U(\lambda_i)$ возможно только при наличии идентификатора.

III. Выводы

1. Проанализированы требования, которые предъявляются к технологии формирования кодограммы для разработанного метода статистического кодирования данных информационного ресурса с применением кластеризации по признаку количества серий единиц.

2. Разработана концепция формирования кодограммы для метода статистического кодирования данных информационного ресурса с применением кластеризации по признаку количества серий единиц.

IV. Список литературы

- [1] Сэлмон Д. Сжатие данных, изображений и звука: Пер. с англ. В.В. Чепыжова. – Москва: Техносфера, 2004. – ISBN 5-94836-027-X.Р.М. – С.195-200.
- [2] Гонсалес Р., Вудс Р. Цифровая обработка изображений. – Москва: Техносфера, 2005. – 1072 с. – ISBN 5-94836-028-8Р.М.. – С.719-727.
- [3] Barannik V., I. Tupitsya, S. Shulgin, S. Sidchenko and V. Larin. The application for internal restructuring the data in the entropy coding process to enhance the information resource security // 2016 IEEE East-West Design & Test Symposium (EWDTS), Yerevan, Armenia, 2016, pp. 561-565. DOI:10.1109/EWDTS.2016.7807749
- [4] Баранник В.В. Двухиерархическая схема статистического кодирования с кластеризацией для повышения безопасности информационного ресурса / В.В. Баранник, И.М. Тупица, С.С. Шульгин // Тези доповідей учасників XIII міжнародної науково-технічної конференції «ABIA-2017», 19 – 21 квітня 2017 р.: тези доп. – К., 2017. – С. 1.52-1.54.
- [5] Баранник В.В. Технологія кластеризації даних інформаційного ресурсу за кількісною ознакою ресурса / В.В. Баранник, І.М. Тупиця, В.В. Баранник, А.Д. Сорокун // Науковий часопис № 4(40) – К., 2018. – С. 398-404. DOI: 10.18372/2310-5461.40.13264