М.В. УЛЬЯНОВ[1] (д.т.н.), **Ю.Г. СМЕТАНИН**[2] (д.ф-м.н.),
([1] Институт проблем управления им. В.А. Трапезникова РАН)
([2] Вычислительный центр РАН имени А.А. Дородницына)

РЕКОНСТРУКЦИЯ СЛОВ ПО КОНЕЧНОМУ МУЛЬТИМНОЖЕСТВУ ПОДСЛОВ В ГИПОТЕЗЕ СДВИГА 1

Разрабатываемый авторами подход к исследованию временных рядов связан с использованием подхода символьного кодирования и последующим применением методов символической динамики для анализа особенностей и характеристик полученных слов.

Подход символьного кодирования состоит собственно в том, что в целях исследования временных рядов осуществляется кодирование значений наблюдаемой величины в некотором алфавите, символы которого именуют полусегменты значений ряда в порядке их возрастания. Если наблюдения ведутся в равных временных отсчетах, то полученное кодированием по именам полусегментов слово доставляет корректное описание временного ряда. При этом отдельной задачей является выбор количества полусегментов кодирования и их ширины.

Пусть, например, в алфавите $\Sigma = \{A, B, C, D, E, F\}$: A есть имя полусегмента наименьших значений, а F — наибольших. В случае, если наблюдаемый процесс характеризуется резкими выбросами значений наблюдаемой величины (до уровня F) относительно базального уровня (A, B) за один дискрет времени, равно как и резкими спадами (от F до B), то получаемые кодовые слова временного ряда не будут содержать подслов CDE и EDC. Если при этом исходные данные представляют собой разрозненные фрагменты наблюдений, полученные из различных источников со сдвигом во времени, то в аспекте исследования временных рядов возникает дополнительная задача реконструкции ряда по этим фрагментарным наблюдениям. В постановке символической динамики — это задача реконструкции слова по известному набору подслов без запретов, а в предположении об особенностях поведения временного ряда — задача реконструкции с запретами. Обсуждению подходов к решению данных задач и посвящен настоящий доклад.

Формальные постановки задач имеет следующий вид. В постановке I мы считаем заданными: длину подслова — k, число подслов — m и исходное мультимножество слов $V(L_k, m)$ над алфавитом $\Sigma = \{0, 1\}$, рассматриваемое как базис реконструкции — т.е. как мультимножество подслов сдвига 1 относительного некоторого неизвестного слова w. Кроме того, в постановке II задано единственное запрещенное слово ω для реконструируемых слов w. Оператор $SH1(w, k)$ порождает мультимножество подслов со сдвигом 1 длины k по слову w.

Постановка I. *Постановка задачи реконструкции без запретов*

<u>Содержательно</u>: В условиях гипотезы сдвига 1 относительно мультимножества $V(L_k, m)$: возможно ли выполнить реконструкцию слова w в принципе, и если эта задача имеет решение, то как выглядят возможные реконструкции?

<u>Формально</u>: Введем в рассмотрение множество

$$W = \{w \mid |w| = m + k - 1, \quad V(L_k, m) = SH1(w, k)\}, \tag{1}$$

при этом равенство понимается как равенство мультимножеств (равны как элементы, так и их кратности); тогда, если $|W| = 0$, то решения нет и реконструкция невозможна, если $|W| \geq 1$, то решение есть и реконструкция возможна. В дальнейшем возможные решения задачи в постановке I будем называть простыми решениями задачи реконструкции.

Постановка II. *Постановка задачи реконструкции с запретом*

<u>Содержательно</u>: В условиях гипотезы сдвига 1 относительно мультимножества $V(L_k, m)$: возможна ли в принципе реконструкция такого слова или слов w, которые не содержит слова ω в качестве подслова, и каковы эти реконструируемые слова?

Формально: Введем в рассмотрение, аналогично постановке I, множество

$$W \setminus \omega = \Omega(W, \omega),$$

где W определено в соответствии с (1), тогда, рассматривая совместно множества W и $W \setminus \omega$ можно выделить следующие случаи:

Случай 1: $|W| = 0$ — множество простых решений задачи реконструкции пусто, и очевидно $|W \setminus \omega| = 0$, тогда множество $V(L_k, m)$ в принципе не является реконструирующим;

Случай 2: $|W| \geq 1$, но при этом $|W \setminus \omega| = 0$ — решения в постановке II нет, все простые реконструкции содержат запрещенное слово, тем самым реконструкция на базе $V(L_k, m)$ без запрещенного слова ω невозможна;

Случай 3: $|W| \geq 1$ и $|W| = |W \setminus \omega|$ — решение в постановке II есть и все простые реконструкции не содержат запрещенного слова — мы говорим в этом случае, что мультимножество $V(L_k, m)$ доставляет полное решение задачи реконструкции в постановке II, которое совпадает с простым решением в постановке I. Заметим, что этот случай равноположен реконструкции w из пространства сдвигов, определяемого запрещенным словом ω;

Случай 4: $|W| \geq 1$ $|W| > |W \setminus \omega|$ — решение в постановке II есть, но есть такие простые реконструкции, которые содержат запрещенное слово — мы говорим, что мультимножество $V(L_k, m)$ доставляет частичное решение задачи реконструкции в постановке II. В последнем случае представляет интерес получение точного числа решений $M_\omega = |W \setminus \omega|$, равно как и самих решений задачи — множества слов $W \setminus \omega$.

Для решения поставленных задач авторами предложен основанный на операторном подходе формализм описания операций над словами, порожденными конечным алфавитом. На основе этого формализма дается содержательная формулировка постановок I и II задачи реконструкции слов в гипотезе сдвига 1.

Предложенное авторами решение основано на процедуре специальной разметки вершин и дуг в мультиорграфе де Брейна, которая позволила свести решение задачи реконструкции к задаче поиска всех эйлеровых путей или циклов в таком мультиорграфе. В терминах предложенного формализма сформулированы и доказаны леммы о существовании и о реконструкции слов по подсловам в гипотезе сдвига 1. В результате адаптации операции символьного умножения матриц к особенностям рассматриваемой задачи введена специальная операция символьного умножения дуг, позволяющая строить эйлеровы пути в мультиорграфе де Брейна.

При рассмотрении задачи реконструкции при наличии запрещенного слова предложенный операторный подход расширен на формализм описания операций с запрещенными словами. Рассмотрены четыре возможных случая решения задачи реконструкции в постановке II, специально выделены и исследованы тривиальные случаи. Предложен вариант полного решения задачи, основанный на решении постановки I путем перечисления всех эйлеровых путей или циклов в мультиорграфе де Брейна с применением специальных алгебраических операций умножения матриц смежности. Дополнительно предложен вариант частичного решения, основанный на введенной авторами процедуре редукции и последующем исследовании редуцированного мультиорграфа де Брейна. Предложенный вариант позволяет эффективно по трудоемкости идентифицировать случай отсутствия запрещенного слова в множестве слов реконструкции. Этот случай представляет определенный интерес, поскольку он соответствует пространству сдвигов в символической динамике. В этом аспекте авторы отмечают, что для задачи проверки, определяет ли заданное мультимножество подслов пространство сдвигов, заданное определенными запретами, предложенная процедура редукции и последующего исследования мультиорграфа де Брейна доставляет полное решение.

Исследование поддержано грантами РФФИ № 13-07-00516, № 15-07-04112

E-mail: muljanov@mail.ru