ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 58 **2022** **Вып. 3**

УДК 621.391 : 517.938 : 519.766

© 2022 г. **М.Л. Бланк**

ВОССТАНАВЛИВАЕМЫЙ ФОРМАЛЬНЫЙ ЯЗЫК

Изучается задача восстановления искаженных произвольно длинных сообщений, записанных на некотором динамически заданном формальном языке. Получены необходимые и достаточные условия на задание языка для наличия допустимого сообщения в окрестности искаженного при условии, что локальные искажения происходят редко.

Ключевые слова: кодирование, формальный язык, динамическая система, отслеживание псевдотраекторий.

DOI: 10.31857/S055529232203007X, **EDN:** EAQKEE

§ 1. Введение

Один из основных вопросов теории кодирования состоит в том, как записать сообщение в такой форме, чтобы искажение некоторой части записи не помешало восстановлению сообщения в исходной форме (см., например, [1–4]). Мы попробуем взглянуть на эту задачу с несколько неожиданной точки зрения: какими свойствами должен обладать формальный язык, на котором пишется сообщение[1] для того, чтобы после "разумных" искажений (уже не являющееся допустимым) исходное сообщение могло быть хотя бы приблизительно восстановлено. Чуть более формально это означает, что для любого "разумно" искаженного сообщения произвольно большой длины (уже не принадлежащего нашему формальному языку) нашлось бы близкое к нему допустимое сообщение. Довольно близкая задача, но с разрешением пропусков кусков сообщений и их перестановками, изучалась, например, в [5].

Под формальным языком мы будем понимать следующее. Рассмотрим ориентированный граф G с конечным множеством вершин $\mathcal{A} := \{1, 2, \ldots, r\}$ (алфавитом языка), $r > 1$, и дискретной метрикой

$$\rho(a, b) := \begin{cases} 0 & \text{при } a = b, \\ 1 & \text{в противном случае,} \end{cases}$$

где $a, b \in \mathcal{A}$. Используемую нами терминологию, связанную с теорией графов, можно найти, например, в [6].

Определение 1. Любой бесконечный в обе стороны путь в ориентированном графе G будем называть *допустимым сообщением*, а объединение всех допустимых сообщений назовем *формальным языком*.

Не теряя общности, мы полагаем, что граф G не содержит "стоков" (вершин, из которых нельзя выйти) и "источников" (вершин, в которые нельзя попасть). В противном случае этого можно добиться модификацией (уменьшением) списка вершин.

[1] Сообщения, принадлежащие нашему формальному языку, назовем допустимыми.

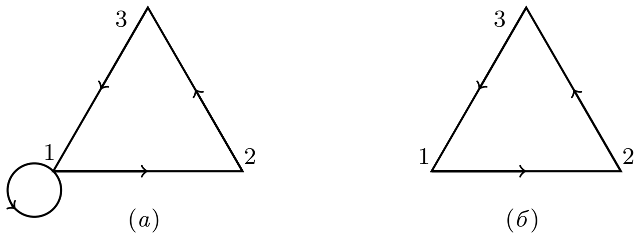

Рис. 1. Примеры графов, порождающих (*а*) восстанавливаемые и (*б*) невосстанавливаемые формальные языки

Заметим, что более сложные грамматические правила, связанные с запретом переходов между выделенными “словами” из нескольких букв, нетрудно реализовать при увеличении размера алфавита.

Определение 2. *Возмущением* допустимого сообщения $\vec{v} := \{v_t\}_{t\in\mathbb{Z}}$ в момент времени t_0 назовем двустороннюю последовательность вершин $\vec{w} := \{w_t\}_{t\in\mathbb{Z}}$, совпадающую с $\vec{v}$ до момента времени t_0, а начиная с t_0 стартующую из произвольной вершины графа G, такую что при любом $t > t_0$ ребро $(w_{t-1}, w_t) \in G$.

Иными словами, возмущенное сообщение состоит из части допустимого сообщения $\vec{v}$ до момента времени $t_0 - 1$ и части некоторого другого допустимого сообщения с момента времени t_0. В терминах блуждания по графу G это означает, что по времени от $-\infty$ до t_0-1 мы следуем по какому-то пути на графе G, в момент t_0 происходит перескок в произвольную вершину графа, а после этого мы вновь следуем пути на графе G.

Напомним, что плотностью последовательности целых чисел $\{t_i\}_{i\in\mathbb{Z}}$ называется

$$\limsup_{n\to\infty} \frac{\#\{i \in \mathbb{Z} :\ |t_i| \leqslant n\}}{2n+1}.$$

Определение 3. *ε-искаженным* сообщением в моменты времени $\ldots < t_{-1} < t_0 < t_1 < \ldots$ назовем результат применения последовательности возмущений при условии, что плотность последовательности моментов возмущения не превосходит значения $\varepsilon > 0$.

Таким образом, низкая плотность последовательности моментов возмущения связана с тем, что они происходят редко по времени. Здесь важно отметить, что в силу отсутствия “стоков” и “источников” как сами сообщения, так и результат их возмущения описываются двусторонне бесконечными последовательностями.

Определение 4. Будем говорить, что произвольная последовательность вершин $\vec{w}$ графа G *отслеживается в среднем* допустимым сообщением $\vec{v}$ с точностью δ, если

$$\limsup_{n\to\infty} \frac{1}{2n+1} \sum_{k=-n}^{n} \rho(w_k, v_k) \leqslant \delta. \qquad (1)$$

Определение 5. Будем говорить, что формальный язык, построенный по ориентированному графу G, является *восстанавливаемым*, если $\forall\delta > 0\ \exists\varepsilon > 0$, такое что любое ε-искаженное сообщение отслеживается с точностью δ.

Наш основной результат дает необходимые и достаточные условия восстанавливаемости формального языка, построенного по ориентированному графу G.

Обозначим через π матрицу переходов в графе G, т.е. $\pi_{ij} = 1$, если ребро (i, j) принадлежит G, и $\pi_{ij} = 0$ в противном случае.

Т е о р е м а 1. *Формальный язык, построенный по ориентированному графу G, является восстанавливаемым тогда и только тогда, когда найдется такое натуральное N, что $\pi^N > 0$ (т.е. все элементы матрицы π^N строго положительны).*

§ 2. Доказательство теоремы 1

С точки зрения теории динамических систем последовательное построение допустимого пути на графе G можно описать с помощью многозначного отображения $T\colon \mathcal{A} \to 2^{\mathcal{A}}$, ставящего в соответствие элементу $a \in \mathcal{A}$ все вершины графа G, в которые можно из него попасть за один шаг. Поэтому задача проверки восстанавливаемости языка сводится к известной в теории динамических систем задаче отслеживания псевдотраекторий (траекторий возмущенного отображения) – см., например, [7; 8; 9, раздел 6; 10, глава 6]. Однако, во-первых, на сегодня нет решения задачи отслеживания для многозначных отображений, а во-вторых, поскольку фазовое пространство $\mathcal{A}$ дискретно, возмущения отображения ни в каком смысле не являются малыми.

Поэтому мы пойдем другим путем. Рассмотрим отображение левого сдвига σ, действующее в пространстве двусторонних последовательностей $\vec{X} := \mathcal{A}^{\mathbb{Z}}$ с элементами из алфавита $\mathcal{A}$, т.е. $(\sigma\vec{x})_i := x_{i+1}\ \ \forall \vec{x} \in \vec{X}$, $i \in \mathbb{Z}$.

О п р е д е л е н и е 6. *Топологической марковской цепью* (subshift of finite type) с матрицей переходов π называется ограничение левого сдвига σ на инвариантное множество *допустимых* последовательностей $\vec{X}_\pi$, определяемых следующим условием: $\vec{x} \in \vec{X}_\pi$ тогда и только тогда, когда $\pi_{x_i x_j} > 0\ \ \forall i \in \mathbb{Z}$.

В соответствии с терминологией, описанной в предыдущем параграфе, множество допустимых последовательностей совпадает с множеством допустимых сообщений. Как и ранее, мы предполагаем отсутствие “стоков” и “источников”. Поэтому допустимые последовательности являются двусторонне бесконечными.

О п р е д е л е н и е 7. Для упорядоченной пары допустимых последовательностей $\vec{x}, \vec{y}$ под *возмущением* в момент времени t будем понимать такую новую последовательность $\vec{z} \in \vec{X}$ (вообще говоря, не лежащую в $\vec{X}_\pi$), что

$$z_i = x_i\ \ \forall i < t, \quad z_i = y_i\ \ \forall i \geqslant t.$$

О п р е д е л е н и е 8. *Псевдотраекторией* $\widetilde{y}$ левого сдвига σ назовем траекторию возмущенной системы с произвольным набором моментов возмущения $\ldots < t_{-1} < t_0 < t_1 < \ldots$ и выбором самих возмущений.

Определим расстояние между последовательностями:

$$\vec{\rho}(\vec{w}, \vec{v}) := \limsup_{n\to\infty} \sum_{k=-n}^{n} 2^{-|k|} \rho(w_k, v_k). \tag{2}$$

О п р е д е л е н и е 9. Будем говорить, что псевдотраектория $\widetilde{y} := \{\vec{y}_n\}_{n\in\mathbb{Z}}$ *отслеживается в среднем* с точностью δ траекторией $\{\sigma^n\vec{x}\}_{n\in\mathbb{Z}}$ точки $\vec{x} \in \vec{X}$, если

$$\limsup_{n\to\infty} \frac{1}{2n+1} \sum_{k=-n}^{n} \vec{\rho}(\vec{y}_k, \sigma^k\vec{x}) \leqslant \delta. \tag{3}$$

Замечание 1. Заметим, что при $\vec{w} := \vec{y}$, $\vec{v} := \vec{x}$ левые части неравенств (1) и (3) согласованы в том смысле, что первая из них не превышает второй.

Аналогично предыдущему параграфу низкая плотность моментов возмущения соответствует редким возмущениям.

Теперь мы уже находимся в рамках теории однозначных динамических систем, и для анализа возможности отслеживания можно воспользоваться следующим недавно полученным результатом [11].

Определение 10. Отображение σ метрического пространства $(\vec{X}, \vec{\rho})$ в себя удовлетворяет условию *склеивания со скоростью* $\varphi \colon \mathbb{Z} \to \mathbb{R}$, если для любой пары односторонних последовательностей элементов $\vec{X}$ вида $\{\sigma^k \vec{x}\}_{k<0}$ и $\{\sigma^k \vec{y}\}_{k \geqslant 0}$ найдется такая последовательность $\vec{z} \in \vec{X}_\pi$, что

$$\vec{\rho}(\sigma^k \vec{x}, \sigma^k \vec{z}) \leqslant \varphi(k) \cdot \vec{\rho}(\vec{x}, \vec{y}) \;\; \forall k < 0, \quad \vec{\rho}(\sigma^k \vec{y}, \sigma^k \vec{z}) \leqslant \varphi(k) \cdot \vec{\rho}(\vec{x}, \vec{y}) \;\; \forall k \geqslant 0.$$

Другими словами, траектория точки $\vec{z}$ одновременно отслеживает траекторию назад точки $\vec{x}$ и траекторию вперед точки $\vec{y}$ со скоростью, контролируемой функцией φ.

Теорема 2 [11]. *Если возмущение ε-редко, то при выполнении условия склеивания с суммируемой функцией φ любая псевдотраектория отслеживается в среднем с точностью $C\varepsilon$, где $C < \infty$ не зависит от выбора возмущений.*

Доказательство. Для применения этого результата при проверке достаточности условий теоремы 1 нужно доказать выполнение условия склеивания с суммируемой точностью аппроксимации φ для отображения сдвига. Согласно условию π^N – положительная матрица. Поэтому за время N мы можем перейти из любого элемента алфавита $\mathcal{A}$ в любой другой. Таким образом, полагая функцию φ равной индикаторной функции целочисленного отрезка $[-N, N]$, получаем требуемый результат.

Остается проверить необходимость. Предположим, что условие теоремы не выполнено и $\nexists N : \pi^N := \big(\pi_{ij}^{(n)}\big) > 0$. Из этого следует, что

$$\forall n \in \mathbb{Z}_+ \;\; \exists i_n, j_n : \; \pi_{i_n, j_n}^{(n)} = 0.$$

Действительно, если бы это было не так, то нашлось бы такое $k \in \mathbb{Z}_+$, что $\pi^k > 0$. Но в этом случае $\pi^n > 0 \;\; \forall n > k$, что противоречит предположению.

Как мы уже отмечали, матрица переходов π индуцирует многозначное отображение алфавита в себя по формуле $\pi a := \{b \in \mathcal{A} : \; \pi_{ab} > 0\}$. В этих терминах из отсутствия строгой положительности π и конечности $\mathcal{A}$ следует, что при некотором $M \in \mathbb{Z}_+$ имеется разбиение $\mathcal{A} := \bigsqcup\limits_i \mathcal{A}_i$ на непустые π^M-инвариантные подмножества, т.е. $(\pi^M)^{-1} \mathcal{A}_i = \mathcal{A}_i \;\; \forall i$. Из этого следует, что при x_0, y_0, принадлежащих различным элементам этого разбиения, соответствующие допустимые последовательности $\vec{x}, \vec{y}$ не пересекаются, т.е. $x_i \neq y_i \;\; \forall i \in \mathbb{Z}$. Поэтому не существует их "склейки" с суммируемой функцией φ.

С другой стороны, отслеживание в среднем в рассматриваемой постановке эквивалентно тому, что отслеживающая траектория может отличаться от отслеживаемой лишь на конечном временном интервале. В противном случае в силу дискретности метрики ρ усредненное расстояние (левая часть формулы (3)) не может быть малой величиной. Поэтому здесь отслеживание в среднем эквивалентно выполнению условия склеивания с суммируемой функцией φ. ▲

Автор благодарен рецензенту за конструктивные и полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Марков А.А.* Введение в теорию кодирования. М.: Наука, 1982.
2. *Хэмминг Р.У.* Теория кодирования и теория информации. М.: Радио и связь, 1983.

3. *Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979.
4. *Ромащенко А.Е., Румянцев А.Ю., Шень А.* Заметки по теории кодирования. М.: МЦНМО, 2017.
5. *Elishco O., Barg A.* Recoverable Systems, `arXiv:2010.00589v2 [cs.IT]`, 2022.
6. *Bollobás B.* Modern Graph Theory. New York: Springer, 1998.
7. *Аносов Д.В.* Об одном классе инвариантных множеств гладких динамических систем // Тр. V Междунар. конф. по нелинейным колебаниям (25 августа – 4 сентября 1969 г.). Т. 2: Качественные методы. Киев: Ин-т матем. АН Украины, 1970. С. 39–45.
8. *Blank M.* Metric Properties of ε-Trajectories of Dynamical Systems with Stochastic Behaviour // Ergodic Theory Dynam. Systems. 1988. V. 8. № 3. P. 365–378. `https://doi.org/10.1017/S014338570000451X`
9. *Blank M.* Discreteness and Continuity in Problems of Chaotic Dynamics. Providence, R.I.: Amer. Math. Soc., 1997.
10. *Katok A., Hasselblatt B.* Introduction to the Modern Theory of Dynamical Systems. Cambridge: Cambridge Univ. Press, 1995.
11. *Blank M.* Average Shadowing and Gluing Property, `arXiv:2202.13407v1 [math.DS]`, 2022.

Бланк Михаил Львович
Институт проблем передачи информации
им. А.А. Харкевича РАН, Москва
Национальный исследовательский университет
"Высшая школа экономики", Москва
`blank@iitp.ru`

Поступила в редакцию
23.03.2022
После доработки
11.06.2022
Принята к публикации
11.06.2022