ВЕРОЯТНОСТНЫЕ И КОЛМОГОРОВСКИЕ ЭКСТРАКТОРЫ ДЛЯ НЕСКОЛЬКИХ ИСТОЧНИКОВ

Д.В. Мусатов
Московский физико-технический институт (государственный университет), Долгопрудный, Россия;
Российская экономическая школа, Москва, Россия.

Аннотация

Экстракторы – это функции, производящие из «не очень хорошей» случайности «новую» и/или «лучшую» случайность. Есть два подхода к точным определениям: классический вероятностный и колмогоровский. Мы обсуждаем применение новых вероятностных конструкций для создания новых колмогоровских экстракторов.

Работа выполнена при поддержке гранта РФФИ №16-01-00362 и гранта RaCAF ANR-15-CE40-0016-01.

PROBABILISTIC AND KOLMOGOROV EXTRACTORS WITH MULTIPLE SOURCES

D.V. Musatov
Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia;
New Economic School, Moscow, Russia.

Введение

Вероятностные алгоритмы важны для решения многих теоретических и практических задач. Достаточно упомянуть проверку простоты, где вероятностные алгоритмы существенно быстрее детерминированных, задачу об эквивалентности арифметических схем, где полиномиальных детерминированных алгоритмов попросту неизвестно, или метод Монте-Карло подсчёта многомерных интегралов. Однако для успешной работы вероятностного алгоритма нужны равномерно распределённые случайные биты. Такие биты можно получить из некоторых физических или социальных процессов, таких как: колебание температуры процессора, точное время обращения к процедуре, белый шум в аудиокарте, измерение радиоактивного фона, колебания курсов валют и т.д. Полученные таким образом биты будут либо слишком дороги (как если механически кидать кубик или крутить рулетку), либо потенциально несвободны от смещённости и внутренних зависимостей.

Экстракторы – это конструкции, позволяющие почти полностью избавиться от любых несовершенств источника случайности. Они позволяют «извлечь» случайность из несовершенных источников, изготовить «лучшую» и/или «новую» случайность. Эта концепция формализуется двумя возможными способами: вероятностным и колмогоровским. В каждом случае

нужно указать, что является «контейнером» для случайности и как измеряется качество источника случайности до и после преобразования.

В первом случае случайность содержится в распределениях вероятности на двоичных словах некоторой длины. Используется две меры случайности: мин-энтропия, которая тем больше, чем меньше вероятность каждого конкретного слова, и статистическое расстояние до того или иного «хорошего» распределения. Во втором случае случайность содержится в конкретном двоичном слове. Мера случайности – близость колмогоровской сложности слова к его длине, т.е. невозможность эффективно заархивировать слово.

Оказывается, что эти два совершенно разных определения в некотором смысле эквивалентны: вероятностный экстрактор является колмогоровским со слегка худшими параметрами, и наоборот. Вероятностным методом можно показать существование экстракторов с почти оптимальными параметрами, однако в явных (полиномиально вычислимых) конструкциях такие параметры пока не были достигнуты. Более того, некоторые конструкции, созданные для вероятностных экстракторов, разрушались при переходе к колмогоровским. Однако в последние годы наметился существенный прогресс в построении явных конструкций вероятностных экстракторов. Такие конструкции потенциально могут быть переведены на колмогоровский язык, но лишь с использованием конкретики: общая конструкция всё ещё не универсальна. Кроме того, данные конструкции потенциально могут быть обобщены для построения колмогоровских экстракторов с ограничениями на вычислительные ресурсы.

Вероятностные экстракторы

В этом разделе сформулированы определения всех основных объектов, а также дан обзор известных результатов. Начнём с вероятностных экстракторов.

Определение. Пусть ξ – случайная величина, принимающая значения в $\{0,1\}^n$. Тогда *мин-энтропией* ξ называется величина $H_\infty(\xi) = -\log\left(\max_{x\in\{0,1\}^n} Prob\{\xi = x\}\right)$. Иными словами, мин-энтропия больше k, если любое конкретное значение принимается с вероятностью меньше 2^{-k}.

Определение. Пусть ξ и ζ – случайные величины, принимающие значения в $\{0,1\}^n$. Тогда статистическим расстоянием между ξ и ζ называется величина $dist(\xi,\zeta) = \max_{\{S\subset\{0,1\}^n\}}|Prob\{\xi\in S\} - Prob\{\zeta\in S\}| = \frac{1}{2}\sum_{x\in\{0,1\}^n}|Prob\{\xi = x\} - Prob\{\zeta = x\}|$.

Определение. Функция $Ext\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ называется (k,ε)-экстрактором с двумя источниками, если для любых независимых случайных величин ξ и ζ, принимающих значения на $\{0,1\}^n$ и имеющих мин-энтропию не меньше k, статистическое расстояние от случайной величины $Ext(\xi,\zeta)$ до

равномерно распределённой на $\{0,1\}^m$ величины U_m не превосходит ε. Аналогично определяются экстракторы для большего числа источников.

Данное определение отвечает интуитивному понятию «лучшей» случайности: результат гораздо ближе к равномерному распределению, чем исходная величина. Для «новой» случайности нужно усиленное понятие: распределения близки, даже если известно значение одной из исходных случайных величие.

Интуитивно ясно, что случайности не может «стать больше», чем было изначально, т.е. стать больше, чем $2k$. Вероятностным методом можно показать, что такие оптимальные экстракторы действительно существуют, однако явные (полиномиально вычислимые) конструкции оптимальных экстракторов до сих пор не известны. В следующей таблице сравниваются разные конструкции экстракторов для двух источников:

Табл.1. Конструкции экстракторов

Число источников	**Мин-энтропия k**	**Длина выхода m**	**Ошибка ε**	**Ссылка**
2	$0.51n$	$\Theta(n)$	$2^{-\Omega(n)}$	[5]
$poly\left(\frac{1}{\delta}\right)$	δn	$\Theta(n)$	$2^{-\Omega(n)}$	[1]
3	δn	$\Theta(1)$	$O(1)$	[2]
2	$0.51n; k = polylog(n)$	$\Theta(k)$	$2^{-\Omega(k)}$	[13]
2	$0.499n$	$\Theta(n)$	$2^{-\Omega(n)}$	[3]
3	$0.01n; k = polylog(n)$	$\Theta(k)$	$2^{-k^{\Omega(1)}}$	[12]
3	$n^{0.51}$	$\Theta(k)$	$k^{-\Omega(1)}$	[8]
3	$polylog(n)$	$\Theta(k)$	$2^{-k^{\Omega(1)}}$	[9]
2	$polylog(n)$	1	$n^{-\Omega(1)}$	[4]
2	$polylog(n)$	$k^{\Omega(1)}$	$n^{-\Omega(1)}$	[10]
2	$\log(n)$	$\Theta(k)$	$2^{-\Omega(k)}$	Конечная цель

Колмогоровские экстракторы

Начнём с определений сложности, которые мы используем для колмогоровских экстракторов.

Определение. Условной колмогоровской сложностью $K(x|y)$ слова x относительно слова y называется длина кратчайшей программы, которая на входе y возвращает x. Безусловной колмогоровской сложностью $K(x)$ называется сложность с пустым условием.

Предполагается, что программа записана в некотором Тьюринг-полном языке программирования. Теорема Колмогорова-Соломонова утверждает, что изменение этого языка меняет сложность любого слова лишь на константу.

Определение. Зависимостью между двумя словами x и y называется величина $dep(x,y) = K(x) + K(y) - K(x,y)$.

Теорема о симметрии информации гласит, что эта же величина с точностью до логарифмического слагаемого равна $K(x) - K(x|y)$ и $K(y) - K(y|x)$.

Определение. Функция $KExt: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ называется (k, d)-колмогоровским экстрактором, если для любых слов x и y, таких что $K(x) > k$, $K(y) > k$ и $dep(x, y) < d$, выполнено $K\big(KExt(x, y)\big) > m - d - O(\log n)$. Если к тому же $K(KExt(x, y)|x) > m - d - O(\log n)$ и $K(KExt(x, y)|y) > m - d - O(\log n)$, то экстрактор называется усиленным.

Комбинаторным подсчётом нетрудно показать, что избавиться от вычитаемого $d + O(\log n)$ нельзя. Однако эта точная граница достигается.

Теорема. (Зиманд, [14, 15]) Существуют колмогоровские экстракторы для $m = 2k - O(\log n)$, а также усиленные колмогоровские экстракторы для $m = k - O(\log n)$.

Несмотря на то, что два типа экстракторов кажутся несвязанными друг с другом, на самом деле они в некотором смысле эквивалентны.

Теорема. [6, 7] Функция, являющаяся вероятностным экстрактором с параметрами k и ε, также является колмогоровским экстрактором с параметрами $k + O(\log n)$ и $d = \log\frac{1}{\varepsilon} + O(\log n)$.

Теорема. [7] Функция, являющаяся колмогоровским экстрактором с параметрами k и d, также является вероятностным «почти экстрактором» с параметрами $k' > k$ и $\varepsilon = \frac{1}{2^{k'-k}}$. «Почти экстрактор» означает, что результирующее распределение ε-близко не к равномерному, а к распределению с мин-энтропией $m - O(\log n)$.

Обсуждение

Конструкция Зиманда строится на основе вероятностного метода, поэтому не даёт явного способа построения колмогоровских экстракторов. До сих пор не было известно явных конструкций колмогоровских экстракторов для достаточно маленьких сложностей. Новые конструкции Чаттопадхьяи-Цукермана и Ли извлекают полилогарифмическую сложность, но лишь при логарифмической зависимости между источниками: исходные экстракторы дают полиномиальную, а не экспоненциальную ошибку. Возможно, более глубокое и специфичное применение новых конструкций позволит усилить результат в колмогоровском случае, но пока это остаётся предметом будущих исследований.

Однако сила новых конструкций заключается не только в возможности явного построения колмогоровских экстракторов. Полиномиальные алгоритмы позволяют распространить идею извлечения случайности на колмогоровскую сложность с ограничением на ресурсы. Сложность $K^{s.t}(x|y)$ слова x относительно слова y с ограничением на память s и время t определяется как длина кратчайшей программы, которая на входе y

возвращает x и при этом работает не дольше t шагов и использует не больше s ячеек памяти. Ранее был известен результат для полиномиальной памяти:

Теорема. [11] Для любых k, d и достаточно большого полинома s существует функция $KExt: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{2k}$, вычислимая на полиномиальной памяти, такая что для любых слов x и y, таких что $K^s(x) > k$, $K^s(y) > k$ и $K^{O(s)}(x,y) > K^s(x) + K^s(y) - d$, выполнено $K^s\big(KExt(x,y)\big) > 2k - d - O(\log n)$.

Новые конструкции позволяют распространить этот результат на сложность с ограничением на время. Правда, пока что получается доказать только для трёх источников, специфической модели вычислений и полилогарифмической точности.

Определение. Недетерминированной колмогоровской сложностью $CN^t(x|y)$ называется длина кратчайшей программы, которая на любой паре (y,z) возвращает либо x, либо специальный символ ошибки $\perp$, при этом всегда работает не дольше t шагов.

Теорема. Существует вычислимая за полиномиальное время функция $KExt: (\{0,1\}^n)^3 \to \{0,1\}^{3k}$, такая что для любых слов x, y и z, таких что $K^{poly(n)}(x) > k$, $K^{poly(n)}(y) > k$, $K^{poly(n)}(z) > k$ и $CN^{poly(n)}(x,y,z) > K^{poly(n)}(x) + K^{poly(n)}(y) + K^{poly(n)}(z) - d$, выполнено $K^{poly(n)}\big(KExt(x,y,z)\big) > 3k - d - O(\log^3 n)$.

В заключение отметим, что пока что даже наиболее продвинутые конструкции как вероятностных, так и колмогоровских экстракторов носят лишь теоретический характер. Во-первых, они работают только асимптотически, для больших длин. Во-вторых, они хоть и полиномиальны, но настолько сложны, что для тех длин, где они начинают работать, они работают уже слишком долго. Типичная конструкция состоит в многократном применении и комбинировании более простых. Возможно, получится сэкономить, если эти более простые использовать не как «чёрные ящики», а раскладывать на составляющие. Возможно также, что на практике будут хорошо работать упрощённые конструкции, основанные на похожих идеях, но без строго доказательства. Построение таких конструкций остаётся предметом будущих исследований.

Литература

1. Barak B. Extracting randomness using few independent sources / B. Barak, R. Impagliazzo, A. Wigderson // Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS). – 2004. – P. 384-393.
2. Barak B. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. / B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, A. Wigderson // Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC). – 2005. – P. 1-10.
3. Bourgain J. More on the sum-product phenomenon in prime fields and its applications / J. Bourgain // International Journal of Number Theory. – 2005. – Vol. 1. – P. 1–32.

4. Chattopadhyay E. Explicit two-source extractors and resilient functions / E. Chattopadhyay, D. Zuckerman // Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC). – 2016. – P. 670-683.
5. Chor B. Unbiased bits from sources of weak randomness and probabilistic communication complexity / B. Chor, O. Goldreich // SIAM Journal on Computing. – 1988. – Vol. 17, no. 2. – P. 230-261.
6. Fortnow L. Extracting Kolmogorov Complexity with Applications to Dimension Zero-One Laws / L. Fortnow, J. Hitchcock, A. Pavan, N.V. Vinodchandran, F. Wang // Information and Computation. – 2011. – Vol. 209, no. 4. – P. 627-636.
7. Hitchcock J. Kolmogorov Complexity in Randomness Extraction / J.M. Hitchcock, A. Pavan, N.V. Vinodchandran // ACM Transactions on Computation Theory. – 2011. – Vol. 3, no. 1. – P. 1:1-1:12.
8. Li X. Improved constructions of three source extractors / Xin Li // Proceedings of the 26th Annual IEEE Conference on Computational Complexity (CCC). – 2011. – P. 126-136.
9. Li X. Three-source extractors for polylogarithmic min-entropy / Xin Li // Proceedings of 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS). – 2015. – P. 863-882.
10. Li X. Improved constructions of two-source extractors [Электронный ресурс] / Xin Li // Режим доступа: https://arxiv.org/pdf/1508.01115.pdf (дата обращения: 10.10.2017)
11. Musatov D. V. On extracting space-bounded Kolmogorov complexity / D.V. Musatov // Theory of Computing Systems. – 2015. – Vol. 56, no. 4. – P. 643-661.
12. Rao A. Extractors for a constant number of polynomially small min-entropy independent sources / A. Rao // SIAM Journal on Computing. – 2009. – Vol. 39, no. 1. – P. 168-194.
13. Raz R. Extractors with weak random seeds / R. Raz // Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC). – 2005. – P. 11-20.
14. Zimand M. Extracting the Kolmogorov complexity of strings and sequences from sources with limited independence / M. Zimand // Proceedings the 26th Symposium on Theoretical Aspects of Computer Science (STACS). – 2009. – P. 697-708.
15. Zimand M. Impossibility of independence amplification in Kolmogorov complexity theory / M. Zimand // Proceedings of the 35th International Symposium on Mathematical Foundations of Computer Science (MFCS). – 2010. – LNCS, Vol. 6281. – P. 701-712.

Сведения об авторе

Мусатов Даниил Владимирович, кандидат физико-математических наук, без учёного звания, доцент кафедры дискретной математики, Московский физико-технический институт (государственный университет), научный сотрудник лаборатории исследований социальных отношений и многообразия общества, Российская экономическая школа, musatych@gmail.com, научные интересы: колмогоровская сложность, псевдослучайные конструкции, дерандомизация, теоретико-игровые модели.