

К-сложность в контексте новейших информационных технологий

© Авторы, 2012

С. П. Левашкин

к.ф.-м.н., Центр Компьютерных Исследований Национального Политехнического Института, Мехико

E-mail: sergei@levashkin.com

В. В. Александров

д.т.н., профессор, зав. лабораторией «Автоматизация научных исследований», СПИИРАН,

засл. деятель науки и техники РФ

E-mail: alexandr@iias.spb.su

Рассмотрены аспекты использования понятия Колмогоровской сложности в рамках новейших информационных технологий.

Ключевые слова: *К-сложность, алгоритмическая теория информатики, теория информации, количественная мера информации, энтропия.*

The paper considers the aspect of Kolmogorov's complexity notion usage within recent information technologies field.

Keywords: *K-complexity, algorithmic theory of informatics, information theory, quantitative measure of information, entropy.*

Введение

В 1965 г. в первом номере журнала «Проблемы передачи информации» появилась титульная статья «Три подхода к определению понятия количества информации» [1], написанная величайшим советским математиком Андреем Николаевичем Колмогоровым. Таким образом, появились на свет алгоритмическая теория информации и Колмогоровская сложность. Интересно отметить, что по существу это была последняя «серьезная» публикация Колмогорова. Дальше последовал период уточнений некоторых его более ранних работ и написание учебников по математике для средней школы. Кстати, создание учебников – задача титаническая, как раз по масштабам Колмогоровской личности. Впрочем, Андрей Николаевич за простые задачи в своей жизни и не брался. Масштабность решаемых А.Н. Колмогоровым задач наиболее ярко можно проиллюстрировать на примере Колмогоровской сложности.

Колмогоровская сложность

Проблема определения сложности некоторого объекта относится к фундаментальным и древнейшим научным проблемам. Над ее решением размышляли лучшие умы, начиная с древнейших философов и заканчивая современными представителями «новейших» наук, какой является, например, информатика. Кроме того, следует отметить, что эта проблема является одной из труднейших для формализации. Ведь, решая ее, необходимо ответить на вопрос: При заданном объекте, что является мерой его сложности? Другими словами: Какие объекты являются «сложными», какие «простыми», а какие «средними»? Хотелось бы обратить внимание на тот факт, что данная проблема имеет философский, метафизический, а также естественно-научный аспекты, а диапазон приложений – от космических тел до обычных чисел. Итак, какой объект считать сложным?

Колмогоров дает удивительный ответ на этот вопрос! Грубо говоря, объект X , представленный в виде конечной бинарной последовательности $B(X)$ имеет сложность $K(X)$, равную длине программы, которая воспроизводит эту последовательность и останавливается. Прежде чем обсудить основные детали Колмогоровской теории и ее перспективы для развития современных компьютерных технологий, заметим что развитие науки имеет свою логику и вряд ли это определение было бы возможно без появления первых компьютеров. Поясим этот тезис.

1. К моменту появления Колмогоровского определения стало ясно (во многом благодаря работам Шеннона), что наиболее удобный, компактный и реализуемый способ описания любых объектов – бинарный (т.е. в виде конечных последовательностей 0 и 1).

2. Именно в таком виде эти объекты комфортно путешествуют по каналам связи, подчиняясь при этом определенным законам, открытым Шенноном.

3. Появление компьютеров привело к бурному развитию программирования в начале–середине 1960-х гг. Если еще раз взглянуть на определения Колмогорова, то становится ясно, что ему удалось удивительно компактно, элегантно и естественно соединить воедино описанные выше обстоятельства. Повторимся, что наука имеет свою внутреннюю логику развития, и именно: Колмогоров в своих работах, как никто другой, умел эту логику отразить. И ярчайшее тому подтверждение – Колмогоровская сложность. По нашему разумению, К-сложность стала достойным завершением блестящей научной карьеры Колмогорова. Но что же предшествовало появлению алгоритмической теории информации?

- Существует множество определений понятия «информация». От самых общих возвышенно-философских типа «информация есть отражение реального мира» или «информация есть всеобщее свойство материи и мера организации систем» до сугубо практических «информация есть сведения, являющиеся объектом сбора, преобразования, хранения и передачи». Так или иначе, теория информации имеет дело с отображениями предметов или явлений в виде символов, образов. Символы могут быть самыми разнообразными, такими, например, как последовательность электромагнитных импульсов, поступающая со спутника связи, устная и письменная речь, телевизионное изображение, генетический код, записывающий наследуемые свойства в биологических клетках. *Создание оптимальной системы символов, отражающих свойства объектов (кодирование), получение сведений об объектах по свойствам символов (декодирование) – вот типичные задачи теории информации.*
- В большей степени в России теорию информации считают одним из разделов кибернетики. Во введении к первому изданию книги «Кибернетика», вышедшему в свет в 1948 г., Норберт Винер пишет: «... нам пришлось разработать статистическую теорию количества информации. В этой теории за единицу количества информации принимается количество информации, передаваемое при одном выборе между равновероятными альтернативами. Такая мысль возникла почти одновременно у нескольких авторов, в том числе у статистика Р. А. Фишера, у д-ра К. Шеннона из Белловских телефонных лабораторий и у автора настоящей книги. При этом Р.А. Фишер исходил из классической статистической теории, К. Шеннон – из проблемы кодирования информации, автор настоящей книги – из проблемы сообщения и шумов в электрических фильтрах. Следует отметить, что некоторые мои изыскания в этом направлении связаны с более ранней работой А.Н. Колмогорова в России, хотя значительная часть моей работы была сделана до того, как я обратился к трудам русской школы».
- Интересные работы американского ученого К. Шеннона были опубликованы в 1948 г. С помощью методов теории вероятностей К. Шеннон искал оптимальные пути кодирования и декодирования информации в целях ее передачи и хранения. Вполне естественно, что первые публикации К. Шеннона, написанные на уровне «физической строгости», привлекли внимание А.Н. Колмогорова. В предисловии к русскому переводу этих работ академик писал: «Значение работ Шеннона для чистой математики не сразу было достаточно оценено. Мне вспоминается, что еще на международном съезде математиков в Амстердаме (1954) мои американские коллеги, специалисты по теории вероятностей, считали мой интерес к работам Шеннона несколько преувеличенным, так как это более техника, чем математика. Сейчас такие мнения вряд ли нуждаются в опровержении. Правда, строгое математическое «обоснование» своих идей Шеннон в сколько-нибудь трудных случаях предоставил своим продолжателям. Однако его математическая интуиция изумительно точна...»¹.

¹ В этом весь Колмогоров! С одной стороны, «физический уровень строгости», а с другой, «...его математическая интуиция изумительно точна...». Именно Колмогоров с учениками довел результаты Шеннона до строгого математического «обоснования» в трудных случаях. Шеннон получил в их лице достойных продолжателей ...

Первой важной задачей новой теории было нахождение количественной меры информации, т.е. численной оценки «информативности» сообщения. Разрабатывая эту идею, т.е. характеризуя неопределенность источников сообщений, К. Шеннон воспользовался термином «энтропия» [3]. Понятие энтропии позволяло решить много важных задач, относящихся к передаче и хранению дискретных сообщений. Однако все попытки перенести его на случай непрерывных сигналов оказались безуспешными. *«Далее я настаиваю на той идее, что основным понятием, допускающим обобщение на совершенно произвольные непрерывные сообщения и сигналы, является не непосредственно понятие энтропии, а понятие количества информации $I(x, h)$ в случайном объекте x относительно объекта h »,* – эта фраза из доклада А.Н. Колмогорова «Теория передачи информации» на заседании АН СССР в 1956 г. и следующие за ней формальные математические построения определили новый подход к алгоритмической теории информатики [3].

В начале 1960-х гг. А.Н. Колмогоров решил реконструировать фундамент теории информации. Об этом периоде своей работы он писал так: *«...занятия совсем общими полуфилософскими размышлениями у меня самого заняли больше времени и энергии, чем, может быть, кажется издали. В такой выработке совсем общих взглядов итог усилий заключается не в формулировке точно фиксированных «результатов», а в общей перестройке собственного сознания и размещения всего в надлежащей перспективе. Поэтому потом оказывается, что как бы и ничего не открыл «нового», а потратил много сил и времени».* Академик лукавит. Он открыл, всего-навсего, новое направление в теории информации. С появлением статьи [1] родилась *алгоритмическая теория информатики, в основе которой лежало понятие колмогоровской сложности конечного объекта.*

Три подхода

А.Н. Колмогоров буквально на десяти страницах текста в абсолютно ясной форме подвел итог десятилетиям многочисленных исследований понятия «количество информации». Он выделил и описал три основных подхода к решению проблемы: комбинаторный; вероятностный; алгоритмический. В последнем случае он дал приведенное выше определение K -сложности и глубоко исследовал ее основные свойства. Кроме того, он дал также определение относительной $K(X, Y)$ сложности одного объекта X относительно другого Y . Главный же вывод статьи абсолютно парадоксален: *«Теория информации должна предшествовать теории вероятностей, а не опираться на нее».* И это говорит человек, превративший теорию вероятностей, бывшую на протяжении веков игрой в кости и карты, в строгую математическую науку! Тем не менее, строгое построение теории вероятностей на базе теории информации остается по-прежнему делом будущего и представляется очень интересной проблемой.

Дискуссия и заключение

В связи с бурным развитием информационных технологий, наблюдаемых в последние десятилетия, возникает вопрос о месте исследований Колмогорова, так или иначе связанных с этими технологиями. Разумеется, K -сложность и алгоритмическая теория информатики все еще ждут новых импульсов для своего развития, прежде всего практического. Пожалуй, первым, кто обратил внимание на возможность практического применения Колмогоровских теорий в инфокоммуникационной среде, был В.В. Александров [4] (дальнейшее развитие сформулированных в [4] идей читатель может найти в [5]). На сайте его лаборатории [6], которую он создал и возглавляет много лет, читаем, что приоритетные и реализованные направления научных исследований суть:

- 1) программируемая технология по А.Н. Колмогорову (ПТК) для организации виртуальной полосы передачи данных;
- 2) информационная безопасность на основе программируемой технологии Колмогорова.

Вряд ли какая-либо другая лаборатория мира занимается подобными разработками или даже мечтает об этом! Попробуем в дискуссионном ключе разобраться «почему?»².

1. Оказалось, что K -сложность алгоритмически не реализуема. Другими словами, для произвольной бинарной последовательности X , вообще говоря, не существует конструктивного способа построения сложности $K(X)$. Этот факт отмечал и сам Колмогоров.

2. Развитие алгоритмической теории информации пошло по «теоретическому» пути, что сводилось к получению различных оценок для $K(X)$ в как можно более «общих» ситуациях, часто довольно искусственных (за исключением, пожалуй, работ известных школ в Голландии, Финляндии и Англии). Это, в конце концов, превратилось в своего рода спортивное состязание между математиками. При этом инженерам и информатикам стало практически невозможно пробиться сквозь сложные математические конструкции. Соответственно стал угасать интерес среди практиков к теории Колмогорова.

3. Вместо намеченных А.Н. Колмогоровым программ развития данного направления (разбросанных, правда, по различным его публикациям), развивались направления конъюнктурные («grantable») на данный момент. К таким направлениям следует отнести, например, многочисленные попытки «измерить» информативное содержание данных, организованных в тахонимии типа WordNet, используя K -сложность. Особым усердием здесь отличаются «специалисты» по естественным языкам.

Следует напомнить, что в 1960 г. А.Н. Колмогоров организовал при кафедре «Теория вероятностей» (МГУ) статистическую лабораторию. Одним из направлений стало применение математических методов в языкознании. Изучая сложность прозаических и поэтических литературных текстов, А.Н. Колмогоров пришел к выводу о необходимости разложения энтропии языка H в сумму слагаемых: $H = h_1 + h_2$, где h_1 – степень гибкости в выражении одной и той же мысли разными способами; h_2 – информационная емкость языка, т.е. количество разных мыслей, которые могут быть изложены текстом одной и той же длины. Важно, что обе эти величины поддаются достаточно точному измерению. Соотношение между h_1 и h_2 позволило формально, с точки зрения теории информации, проанализировать различия между прозой и поэзией. Все глубже погружаясь в поэзию, академик вычислил вероятность того, что пара наугад взятых слов рифмуется, предложил количественные оценки сложности рифм, определил затраты энтропии на строку «Евгения Онегина».

Другой яркий пример развития Колмогоровского подхода, – работа [7]. В ней, в частности, проводится анализ информационных особенностей самой загадочной Пушкинской поэмы «Медный всадник». Акцент делается на семантические особенности этого гениального произведения, а анализ базируется на разработанных авторами новых математических методах. Прекрасный пример проверки гармонии алгеброй! Думается, что Колмогоров одобрил бы такой подход.

«Стих только тогда убедителен, когда проверен математической (или музыкальной, что тоже) формулой. Проверять буду не я», – писала Марина Цветаева.

Заключение

1. Богатейшее наследие А.Н. Колмогорова срочно требует внимательного изучения и практического развития в применении к новейшим информационным технологиям. Мы не сомневаемся, что это наследие также богато по содержанию, как и сами эти технологии и безусловно будет способствовать их дальнейшему улучшению и развитию.
2. Напомним, что один из последних тезисов, оставленных нам А.Н. Колмогоровым, гласит, что *будущее за дискретной математикой, за компьютерными технологиями, за K -сложностью ...*

² Кроме K -сложности заслуживают также внимания работы Колмогорова (–Успенского) по теории автоматов, а также понятие машины Колмогорова (–Успенского), обобщающего понятие машины Тьюринга, и также не получивших, с нашей точки, достаточного практического применения. Сразу вспоминается знаменитый доклад А.Н. Колмогорова «Автоматы и жизнь». Однако в этой статье, из-за недостатка места, мы ограничимся только алгоритмической теорией информации...

ЛИТЕРАТУРА

1. Колмогоров А.Н. Три подхода к определению понятия количества информации // Проблемы передачи информации. 1965. Т. 1. № 1. С. 3–11.
2. Шеннон К.Е. Математическая теория связи // Bell System Technical Journal. 1948. V. 27. P. 379–423, 623–656.
3. Александров В.В., Кулешов С.В., Цветков О.В. Цифровая технология инфокоммуникации. Передача, хранение и семантический анализ текста, звука, видео. СПб.: Наука. 2008.
4. Александров В.В., Арсентьева А.В. Информация и развивающиеся структуры. Л.: ЛИИАН. 1984.
5. Levashkin S., Alexandrov V. Data semantic associative analysis and synthesis // Preprint 2012 г., 20 с. <http://levashkin.com/files/Levashkin-Alexandrov.pdf>
6. Веб-сайт Лаборатории Автоматизации Научных Исследований. СПИИРАН <http://sial.ias.spb.su/>
7. Александров В.В., Арсентьева А.В., Семенов А.И. Структурный анализ диалога // Препринт. 1983 г., 49 с. <http://sial.ias.spb.su/files/strandialog.pdf>

Поступила 2 апреля 2012 г.

K-complexity with in Recent Information Technologies Context

© Authors, 2012

S.P. Levashkin, V.V. Alexandrov

In 1965 the titular paper «Three approaches to the definition of the concept «quantity of information» by A.N. Kolmogorov has appeared in «The problems of information transmission», which opened a new direction as a result of many years of his research of information theory. Thus algorithmic theory of information and Kolmogorov's complexity had appeared. The problem of evaluation of object's complexity belongs to fundamental and ancient scientific problems and is most difficult to formalize. This problem has philosophic, metaphysic and natural-science aspects, and the range of its application varies from cosmic bodies to common numbers.

The first problem this the new theory had addressed was to define quantitative measure of information i.e. numeric estimation of message «informativity». The notion of entropy allowed to solve great variety of important problems related to transmission and storage of discrete data. But all attempts to extend this notion to continuous signals failed. A.N. Kolmogorov discovered new trend in information theory, he created algorithmic theory of information based on the notion of Kolmogorov's complexity.

Due to rapid growth of information technologies observed last 10 years the question of Kolmogorov's research place has emerged one way or the other related to this technologies. Of course, K-complexity and algorithmic theory of information are still waiting for new impulses for its development, firstly the practical. It is turned out that K-complexity can not be implemented algorithmically. Another words, for any binary sequence X the constructive way to obtain the complexity $K(X)$ does not exist. The development of algorithmic information theory has gone to «theoretical» way, which resulted in obtaining various estimations for $K(X)$ preferably for «common» situations most of which are fairly artificial. This led to the kind of sport competition between mathematicians. At the same time it has become almost impossible for engineers and informatics to overcome complex mathematical constructions. Which in turn resulted in interest loss for Kolmogorov's theory among the practitioners.

The rich Kolmogorov's legacy is urgently demands careful study and practical development in application to modern information technologies. We are in no doubt that this legacy is also rich in content as the new technologies and doubtfully will assist its further improvement and development.

НАУЧНАЯ ШКОЛА

«АЛГОРИТМИЧЕСКИЕ МОДЕЛИ ЦИФРОВОЙ ПРОГРАММИРУЕМОЙ ТЕХНОЛОГИИ РАЗВИВАЮЩИХСЯ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ»



Основатель школы:

Виктор Васильевич Александров –
д.т.н., профессор,
заслуженный деятель науки и техники РФ,
лауреат премии Правительства РФ
в области науки и техники