

УДК 621.391.1:519.27

© 1999 г. Н.А. Кузнецов, В.А. Любецкий

КОМПЬЮТЕРНАЯ ЛОГИКА В ИНФОРМАЦИОННЫХ ПРОЦЕССАХ

Информационные процессы в технических системах изучаются достаточно давно. С этой целью был создан математический аппарат, включающий теорию релейно-контактных схем и других дискретных преобразователей информации, теорию различных типов автоматов, вероятностную теорию передачи сигналов по каналу, шенноновскую теорию информации (с опорой на численное измерение количества информации), теорию кодирования, теорию надежности, многие разделы теории вероятностей и т.д. При этом возникли такие фундаментальные понятия, как обратная связь, оптимальное управление техническим объектом, количество информации (энтропия), канал связи, пропускная способность, оптимальная передача информации, точность воспроизведения, кодирование-декодирование (в том числе с открытым ключом), распознавание образов и т.д. Эти понятия вышли далеко за пределы собственно теории информационных процессов в технических системах; в том числе их пытались (часто необоснованно) применять для описания информационных процессов в живых системах. В [1] ставятся задачи: развитие концепции информационных взаимодействий, возникающих в ходе процессов определенного типа в живых, смешанных и сложных искусственных системах; формирование научного языка, адекватного данному типу процессов; развитие в этом направлении математического аппарата.

Кажется, что некоторые задачи, связанные с биохимией и взаимодействием клеток, как и некоторые задачи, связанные с целенаправленным поведением, нуждаются в изучении определенных, свойственных этим задачам, процессов; эти задачи могут дать фактический материал для определения соответствующих процессов, для развития языка процессов этого типа и языка информационных взаимодействий, характерных для этих процессов.

Прежде чем продолжить обсуждение, приведем два примера упомянутых задач.

Задача 1. *Поиск регулона по известной его части.* Дан полный геном бактерии (т.е. слово длины около 3000000 символов в четырехбуквенном алфавите, соответствующем нуклеотидам). В этом слове помечены гены — подслова длины примерно 1000, а перед ними помечены участки — подслова длины примерно 15–20, называемые операторами, которые включают/выключают соответствующие гены. Как хорошо известно, каждый ген кодирует один белок: трем нуклеотидам (кодону) соответствует аминокислота, и около 300 аминокислот образуют белок.

Будем рассматривать белки, которые несут функцию фермента определенной химической реакции. Тогда каждому белку соответствует ребро в карте (графе) метаболических путей данной бактерии (ребро — химическая реакция, вершина — соответствующее химическое вещество). Будем считать заданным переход от белков к ребрам в карте метаболических путей.

Регулон — это группа генов, которые одновременно включаются для синтеза белков (или для его прекращения) с целью протекания единого связанного каскада химических реакций (идущих одна за другой и образующих путь длины порядка 20 ребер или небольшой связный подграф в карте метаболических путей). Обычно регулон содержит около 30 генов. Каждому регулону соответствует отдельный белок (называемый *фактором транскрипции*, у данного генома их около 100), который, чувствуя изменение внешних условий (а именно, концентраций разных веществ во внешней среде), узнает операторы данного регулона, связывается с ними и, тем самым, дает сигнал с целью одновременного включения (выключения) синтеза ферментов для протекания вышеупомянутого каскада химических реакций.

Первая часть задачи состоит в поиске новых операторов данного регулона по части уже известных операторов этого регулона. Для этого среди всех (заранее помеченных) операторов (бактерии) пытались вести поиск новых операторов данного регулона путем сравнения их с “образцом”, составленным по списку уже известных операторов этого регулона.

Вторая часть задачи состоит в том, чтобы по изученному регулону (т.е. такому, в котором известно около 10 генов и их операторов, но не все они) одной бактерии найти аналогичный регулон у другой *близкой* бактерии. “Близость” определяется рядом обстоятельств, среди которых главным является наличие гомологичного белка — фактора транскрипции.

Для решения второй части задачи по известным операторам также пытались составлять “усредненный образец”, который затем использовался для поиска среди всех операторов другой бактерии.

Недостатки такого рода подходов к решению подобных задач очевидны: обучающая выборка из 10–15 примеров очень мала; незнание механизмов ДНК-белкового взаимодействия приводит к тому, что неизвестно, как выбирать признаки для сравнения (строить “образец”). Часто гены и операторы (регуляторные области) бывают не помечены в ДНК, что еще более затрудняет поиск новых операторов в регулоне или аналогичного регулона у другой близкой бактерии; сам “критерий близости” также не вполне адекватен.

Подход, который предполагается развить в связи с этой задачей, состоит в следующем. Одновременно рассматривается несколько близких бактерий, в одной из которых интересующий нас регулон изучен; по этому регулону составляется (как именно — отдельный вопрос) “слабое” правило (а именно, отбирающее больше генов, чем нужно), которое применяется ко всем этим бактериям; в результате в каждой из них отбирается около 200 генов, которые образуют предполагаемый (заведомо избыточный) “регулон”; затем в заранее известных картах метаболических путей этих бактерий одновременно запускается поиск аналогичного для всех этих бактерий химически осмысленного каскада реакций с ферментами из этого “регулона”; и в этом процессе отбираются “востребованные” гены в уже искомый регулон.

Примечание. Этот алгоритм похож на алгоритмы построения сложных рекурсивно перечислимых множеств и предельных структур, которые упоминаются дальше. ▲

Задача 2. Достижение цели информационно взаимодействующей системой. Задача о достижении цели интеллектуальной системой допускает много разных математических постановок. В одной из них система сохраняет неизменный *мотив* поведения в течение некоторого отрезка времени. На основе мотива вырабатывается текущая *цель* поведения. Сначала цель поведения формулируется в неопределенной форме как задача построения предельной структуры, в которой выполнялся бы заранее фиксированный набор свойств. Эти свойства содержат символы операций, отношений и объектов, которые получают свои значения только в случае достижения цели, т.е. в случае успешного завершения динамического процесса построения этой предельной структуры. Формально говоря, каждое из этих свойств является

высказыванием о неопределенных объектах в языке, содержащем символы операций и отношений, смысл которых до конца не фиксирован; достижение цели означает успешное построение полной семантики языка, в которой эти свойства будут выполняться.

Примечание. Предельная структура и полная семантика взаимно определяют друг друга; алгоритм построения полной семантики похож на алгоритмы построения предельных (генерических) структур в теоремах опускания типов.

Отметим ряд понятий, общих для обеих задач. Можно говорить о максимально полной карте метаболических путей, и тогда карты метаболических путей бактерий описываются как фрагменты аксиоматики, задающей эту полную карту. Происходит эволюция (динамика) и самой карты бактерии (точнее, штамма, “популяции” бактерий); например, у многих бактерий появились новые куски карт, связанные с приобретенной ими способностью к удалению пенициллина. В зависимости от текущей ситуации (наличия глюкозы, фруктозы и т.п.) активизируется определенная часть карты данной бактерии. В случае интеллектуальной системы это соответствует текущей *внешней ситуации* и текущей *реакции*. Понятия *памяти* и *модели мира* также находят отражение в обеих задачах. (См. далее замечание об аналогии естественного и генетического языков.) ▲

Взаимодействие в технических системах обычно описывается на естественном языке или на языке соответствующего раздела математики; этот последний язык (например, язык дифференциальных уравнений или язык теории вероятностей) по сути является опять-таки специализированной частью естественного языка. Напротив, упомянутые выше задачи требуют описаний на структурированных и, как правило, формализованных языках. “Структурированность” языка означает выделение в нем алфавита (знаков, фонем), связок (логико-лингвистического типа), слов (морфем, формул), иногда еще и предложений (фраз), текстов, а самое главное — наличие в нем явно задаваемой семантики (смысла) фраз. Все эти составляющие языка должны быть формализованы и доступны эффективной компьютерной обработке. В таком случае говорят о формализованном языке. Естественный язык и его специализированные части (подобные языку медицины или языку дифференциальных уравнений), конечно, не являются формализованными языками. Естественный язык — очень сложная информационная система (включающая взаимодействующие процессы внутри самого языка). В то же время естественный язык отражает, по крайней мере, некоторые процессы, происходящие в интеллектуальных системах. Формализованные языки являются как раз одним из подходов к изучению естественного или близкого к нему языка. Другим важным языком является язык построения белков в клетке (говоря шире, язык связанных с ДНК биохимических процессов в клетке). Он называется генетическим языком. Можно надеяться использовать формализованные языки также и при описании генетического языка. Общепризнанным наблюдением является определенное содержательное соответствие между естественным и генетическим языками. Напомним, что при этом соответствуют: фонема — нуклеотиду (или аминокислоте), морфема — сайту кооперативного связывания белков (и т.п., т.е. единице, имеющей значение, но функционально несамостоятельной), слово — промотору (или терминатору и т.п), предложение — регулону (и набору регулируемых областей регулона), текст — молекуле ДНК (см. [2]). Поскольку в содержательных терминах можно проследить соответствие этих языков, кажется возможным проследить его и на уровне соответствующих им формальных языков; затем разработать единый формальный язык для описания, по крайней мере, некоторых процессов, характерных для обоих случаев.

Формализованный язык рассматривается всегда в окружении ряда обслуживающих его алгоритмов (типа алгоритмов логического вывода или построения сложных перечислимых множеств). Среди этих алгоритмов особенно важными представляются те, которые обеспечивают возможность автоматической обработки семантики

языка. Среди них выделяются алгоритмы, обеспечивающие обработку семантики путем приписывания фразам (языка, адекватного рассматриваемому процессу) степени их достоверности (выполненности).

Упомянутого типа задачи отличаются тем, что к ним плохо применим математический аппарат, обычный при изучении информационных взаимодействий в технических системах. Кажется перспективным использовать в них соответствующим образом развитый аппарат конструктивной теории моделей, соединяя его с систематическим развитием аппарата семантического оценивания [3–6].

Перечислим некоторые желательные черты такого развития математического аппарата.

1) Умение строить отражающий динамические отношения, богатый по выразительным возможностям формальный язык (описывающий процессы данного типа).

2) Умение работать с переменными описаниями (аксиоматиками); в зависимости от хода процесса выделять бедные фрагменты богатого языка и динамически обрабатывать их.

Эффективно описанное множество фраз (формул) некоторого фиксированного языка называется текстом (аксиоматикой). Аксиоматику можно понимать как описание текущей ситуации. Отсюда, естественно, возникает представление о переменном (динамическом) описании: о последовательности аксиоматик, где n -й член последовательности описывает, например, вход системы в n -й момент времени ($n = 1, 2, 3, \dots$). Представим себе (в качестве примера), что система фиксирует описание предметов и их взаиморасположение (т.е. описание того геометрического мира, который она “видит”) в каждый n -й момент времени (скажем, на языке аналитической геометрии). Или вспомним, что, рассматривая художественную картину, мы в каждый момент времени обращаем внимание только на ее небольшую часть; аналогично, одномоментное внимание может быть обращено только на простую короткую фразу (формулу). Желание использовать богатый по выразительным возможностям язык расходится с желанием эффективно обрабатывать язык. Поэтому можно рассматривать бедные фрагменты богатого языка: в каждый n -й момент времени используется такой фрагмент. Он допускает эффективную компьютерную обработку, но за некоторый отрезок времени последовательные описания с помощью этих фрагментов языка дают в целом содержательное впечатление.

3) Умение работать с отношением типа “обстоятельство Φ выполняется в ситуации K с достоверностью u ” (если $u = 1$, то выполняется, а если $u = 0$, то не выполняется). Умение эффективно вычислять это важное для описания процессов отношение.

Отношение “обстоятельство (фраза, формула) Φ выполняется в ситуации K ” (зависящее от двух переменных Φ и K) играет фундаментальную роль в математике и вне ее (вспомним представление о контексте слова или об условном рефлексе Павлова). Обычно это отношение определяется как “формула Φ истинна в структуре K ” (что совершенно не эффективно) или как “формула Φ выводима из аксиоматики K ” (что для содержательных аксиоматик, как правило, приводит к алгоритмам, которые не могут быть эффективно, компьютерно реализованы). Можно надеяться, что отношение, упомянутое в начале этого пункта, будет вычисляться более эффективно.

В заключение повторим, что развитие математического аппарата было бы желательно проводить в тесной взаимосвязи с исследованием самих процессов.

СПИСОК ЛИТЕРАТУРЫ

1. Кузнецов Н.А. О развитии фундаментальных исследований по информационному взаимодействию в природе и обществе // Пробл. передачи информ. 1997. Т. 33. № 3. С. 111–112.

2. Гельфанд М.С. Коды генетического языка и естественный язык // Вопросы языкознания. 1990. № 6. С. 60–70.
3. Любецкий В.А. Оценки и пучки. О некоторых вопросах нестандартного анализа // УМН. 1989. Т. 44. № 4. С. 99–153.
4. Любецкий В.А. Об одном подходе к моделированию интеллектуальных систем // Пробл. передачи информ. 1993. Т. 29. № 3. С. 107–109.
5. Любецкий В.А. Теоремы переноса и алгебра модальных операторов // Алгебра и логика. 1997. Т. 36. № 3. С. 282–303.
6. Chagrov A., Zakharyashev M. Modal logic. Oxford: University Press, 1997.

Поступила в редакцию

26.10.98