

АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ АНАЛИЗА СООБЩЕНИЙ

В. В. Однобоков

Псковский политехнический институт
Санкт-Петербургского государственного технического университета
180680, Псков, ул. Л. Толстого, д.4
centrtrst@ppi.psc.ru

УДК 681.3

В. В. Однобоков. **Алгоритмы интеллектуальной поддержки анализа сообщений** // Труды СПИИРАН. Вып. 1, т. 3. — СПб: СПИИРАН, 2003.

Аннотация. Предлагаются новые программные решения, учитывающие реальные возможности создания обучающихся и аналитических систем. Обсуждаются возможности создания интеллектуальной поддержки принятия решений на основе использования предикатных построений. — Библ. 4 назв.

UDC 681.3

V. V. Odnobokov. **Algorithms of intellectual support of the message analysis** // SPIIRAS Proceedings. Issue 1, v. 3. — SPb: SPIIRAS, 2003.

Abstract. The new program solutions using real possibilities to create self-instructing and analytical systems are offered. The possibilities to create intellectual support to decision making on the basis of the predicate construction usage, are discussed. — Bibl. 4 items.

1. Постановка задачи интеллектуальной поддержки анализа текстовых сообщений

Значительная часть проблем системного анализа и обработки информации сводится к разработкам процедур интеллектуальной поддержки принятия управляющих решений и, в том числе, семантическому контролю входных сообщений. Методы и алгоритмы такой поддержки могут быть весьма различны, хотя, конечно, алгоритмы всегда связаны с принципиальными особенностями используемых методов. Особенно интересной представляется разработка такого рода алгоритмов, когда избранный метод представления входных сообщений позволяет контролировать их непосредственно на входе приемного устройства.

В этом случае нашей первичной задачей является создание модели виртуального объекта — процесса обучения, получения и усвоения знания некоторым техническим или живым (биологическим) устройством. Достаточно ясно, что интерес представляет только процесс обучения (анализа сообщений) с использованием естественного контекстно-зависимого (КЗ) языка. В противном случае система является чисто программной, и обработка сообщений сводится к достаточно тривиальным процедурам.

Обязательным требованием к построению алгоритмов обеспечения такого рода модели является возможно полное сохранение семантики естественного языка во внутримодельных связях, на котором только и может существовать модель, сколько-нибудь адекватная реальному процессу обучения сложной системы.

Это тем более необходимо, что мы фактически поставили перед собой цель создания модели активного обучающегося объекта, причем активного на уровне оценки непротиворечивости получаемого им потока входных сообщений сравнительно с ранее полученным знанием. В составе такой модели явно выделяются несколько алгоритмических составляющих. Это:

- собственно модель обучающегося устройства, представленная как алгоритмическая инструкция его поведения;
- алгоритм представление предметной области, соответствующий уровню необходимой модели обучающегося устройства;
- алгоритм преобразователя естественного языка в язык машинного представления предметной области;
- модель априорного знания;
- модель учителя (алгоритм поведения источников обучающих сообщений);
- модель (алгоритм) контроля процесса обучения.

Достаточно полное описание моделей и сопутствующих алгоритмов приводится в [4], а здесь мы рассмотрим только основные моменты.

Вполне естественно, для моделирования обучающейся системы с поставленным условием сохранения семантики ее взаимодействия с внешними сообщениями, выбрать какой-либо вариант логического моделирования, например предикатное или сетевое описание самого обучающегося с использованием одного из стандартных моделирующих подходов, давно известных для таких представлений. Мы делаем здесь выбор в пользу предикатного варианта, прежде всего потому, что сегодня имеется продуктивная возможность использования так называемой двухбазовой динамической модели [1].

Обучающаяся система может адекватно моделироваться только как процесс взаимодействия с внешними потоками сообщений от различных источников, обычно называемых внешним знанием. Такой процесс в интеллектуальных системах является процессом управления. Соответственно в основу модели необходимо положить алгоритмы, связанные с представлением процессов управления, вполне эквивалентных процессам «передачи знания». В этом случае обучение как процесс, в котором нам надо сохранить семантику, вполне можно моделировать на достаточно высоком уровне интеллектуального взаимодействия.

Вспомним определение понятия интеллекта в его классическом понимании. Как известно, под интеллектом понимается способность мышления, рационального познания. Конструктивно, исходя из существующих определений субъекта и объекта, это определение реализуется в тройке *субъект — управляющее действие — объект (СУО)* или (в общем случае) в последовательности *субъект — управляющее действие — субъект — управляющее действие ... — объект* [2].

Процесс обучения в интеллектуальной системе такого рода возникает, когда в процессе познания (обучения!) две указанные ее части периодически меняются ролями: субъект (учитель) переходит в состояние объекта (обучаемого), объект (обучающийся) становится ответным источником активности — субъектом.

Это утверждение надо понимать следующим образом. В процессе обучения непрерывно требуется подтверждение — субъект должен становиться объектом для восприятия правильности контекстного понимания его сообщения. Без такой организации общения нельзя получить даже директивно-командного стиля обучения в сколько-нибудь сложной системе, ибо учителю, для планирования следующей команды, надо, как минимум, быть уверенным в правильном контекстном понимании предыдущей.

Исходя из сказанного, можно сделать следующие выводы.

- Интеллект инженерно существует (проявляется как алгоритм) как атрибут субъектно-объектной пары, как процесс субъектно-объектного взаимодей-

ствия обучающегося и источника знания, в котором считается, что источник занимает доминирующее положение. Соответственно модель такого взаимодействия структурно должна быть эквивалентна модели интеллектуальной системы, отличаясь от нее именно фактором доминирования одной из сторон.

- Все обучающиеся системы в процессе обучения образуют системы управления по определению субъекта и объекта, без информационного взаимодействия которых они не существуют. Модель обучающейся системы должна содержать в себе факторы, учитывающие особенности именно информационного взаимодействия — информационного управления, возникающего в процессе обучения.

Итак, в концептуальном смысле под обучением понимается направление, связанное с организацией специального варианта управления в сложных системах и обеспечивающее самостоятельное принятие ими управляющих решений по обмену и восприятию информации. Следовательно, адекватной моделью можно признать ту, которая сможет представить обучаемого в процессе обучения — процессе сложного двухстороннего взаимодействия с источником знания.

Достаточно ясно, что указанная выше тройка СУО является двухместным предикатом, изменяющимся в процессе его восприятия на входе системы при сравнении с уже имеющимся в обучающейся системе знанием — набором ранее приходивших предикатов вместе со всеми следующими из них логическими построениями.

2. Представление проблемной области на основе языка предикатов

На основе введенной трактовки исчисления предикатов как СУО может быть рассмотрена последовательность проведения работ по преобразованию некоторого входного проблемно-ориентированного языка (ПОЯ) в язык предикатов, т.е. практика сведения КЗ ПОЯ к машинно-представимому языку, ради чего, собственно говоря, мы и рассматриваем предикаты.

Путем анализа значительного количества текстов на ПОЯ некоторой области науки или практической деятельности можно выделить и стандартизировать ограниченную, но достаточную группу отношений R и правил построения на их основе логических выводов, практически выполняющихся на всем информационном массиве. Правильность текущих формализмов (текущих соглашений о правилах проведения логических выводов) определяется на основе их проверки на всем объеме имеющегося текста после его преобразования в простые синтагмы [2].

По мере поступления на вход системы новых текстов ведется проверка логических выводов, следующих из них по имеющемуся набору правил вывода. При расхождении правил логики производства выводов в источнике текстов и сообщений и в системе, принимающей эти информационные посылки, в последней включаются механизмы изменения правил логических выводов.

Ясно, что фактически здесь речь идет об организации субъектно-объектного взаимодействия, о ситуации, когда структура взаимосвязей данных в предметной области источника информации начинает отличаться от таковой структуры в предметной области системы — приемника информации.

Такой процесс принципиально может быть организован в полностью автоматическом режиме, ибо он всегда связан с согласованием текущих формализмов в уже имеющейся системе и в новой (присущей новому сообщению) предметной области в процедуре СУО.

Рассмотрим три разные бинарные отношения, взяв за образец пример из известной публикации [3]:

«Реализует» R_P

Формально-логические свойства: $AR_P A$; $AR_P B \Rightarrow BR_P A$; если $AR_P B$ и $BR_P C$ то $AR_P C$.

«Включает» R_B

Формально-логические свойства: $A R_B A$; $AR_B B \Rightarrow BR_B^{-1} A$; $AR_B B \neq BR_B A$; если $AR_B B$ и $BR_B C$ то $AR_B C$; если $AR_B B$ и $BR_B A$ то $AR_P B$.

«Пересекает» R_Π

Формально-логические свойства: неверно, что $AR_\Pi A$; $AR_\Pi B \Rightarrow BR_\Pi^{-1} A$; $AR_\Pi B \neq BR_\Pi A$; вообще говоря, неверно, что если $AR_\Pi B$, а $BR_\Pi C$, то $AR_\Pi C$; если $AR_\Pi B$, то $BR_{до} A$; если $AR_\Pi B$, то $A'R_{дв}$, где $R_{до}$ — отношение «локализуется раньше», $R_{д}$ — отношение «дислоцирует», A' — характеризует деформацию тела, обозначенного термином A .

С точки зрения математической логики приведенное выше можно рассматривать как сумму формально-логических и эмпирических свойств выделенных бинарных отношений. При их анализе становится ясно, что многие из них способны порождать новые отношения. Поэтому ключевым моментом разработки языка интеллектуальной системы является исследование правил преобразования бинарных отношений данных в набор синтагм, составляющих информационный массив.

Представляется логически непродуктивными надежды на достаточность разового выделения «чисто математических» свойств отношений и эмпирик. Их текущее по времени и объему массива изменение достаточно вероятно следует из их указанного выше общего свойства порождения новых отношений. Поэтому мы будем продолжать придерживаться терминологии «логики с изменяющимися отношениями» и «изменения формально-логических и эмпирических правил».

При расширении источников поступления информации, что эквивалентно расширению объема предметной области, установленные соглашения могут изменяться за счет выявления и коррекции тех формально-логических правил, которые начинают давать ошибки логического вывода.

В открытой системе с внешним потребителем информации (человеком, например) критерием правильности логических выводов и необходимости их текущих изменений является, естественно, согласие или несогласие потребителя информации с утверждениями, полученными в системе путем логического вывода. В полностью автоматической (интеллектуальной) системе таким критерием может являться только базовый набор правил проблемной области, нарушение которых по какой-либо причине недопустимо, а все остальные выводы эквивалентны формированию «собственной точки зрения» интеллектуальной системы.

Последнее как раз и подтверждает, что в некотором, пусть еще и достаточно примитивном смысле, мы начинаем иметь дело с целеустремленной системой, имеющей свой взгляд на логику проблемной области, т.е. с системой, способной вступать в субъектно-объектное взаимодействие.

Итак, при формировании информационного массива в виде наборов двуместных предикатов, может быть получена достаточно интересная система представления естественного языка человека в компьютере с «некоторыми надеждами» на сохранение контекстной зависимости после всех преобразований, пригодная как для практических выводов, так и для решения вопросов *автоматического* анализа текстов на ПОЯ. Действительно, предикатное представление ПОЯ с механизмом модификации текущих логических формализмов с одной стороны не уничтожает его КЗ сущности, по крайней мере, в смысле обеспечения возможности построения множества формальных моделей, соответствующего множеству возможных контекстов, а с другой является подходящим механизмом разборки записи на ПОЯ на термины и отношения. Следовательно, мы получаем аппарат построения информационной базы в составе базы данных и базы знания.

Такого рода процесс естественен для любой серьезной информационной системы и не несет никакой интеллектуальной нагрузки до тех пор, пока мы не выведем эту «половинку» интеллекта на уровень интеллектуальной системы СУО. Вот тогда мы получим инженерную возможность автоматической выработки в системе изменений в законах формально-логического вывода под действием входного потока информации на ПОЯ, оценки их «глобальной» правильности с нашей точки зрения.

3. Модель источников знания и априорного знания

Как же мы можем представить себе процедуру преобразования текста ЕЯ в их предикатное представление? В настоящее время это самое сложное устройство всей моделирующей системы. Его задачей является автоматизация сведения контекстно-зависимого входного языка, заданного текстовыми сообщениями к предикатам первого (далее - второго) порядка.

За основу эвристической модели преобразования примем ручную процедуру, изначально предложенную в [3]. Изначально эта модель сводилась к следующим преобразованиям:

- перевод текста ЕЯ в граф синтагматической структуры текста с последующим его представлением в виде некоторого массива двуместных предикатов;
- простановка указателей связи при терминах графа текста, что необходимо, ибо одно и то же слово может встречаться в графе несколько раз в различных употреблениях;
- дополнение последовательности простых синтагм на основе их формальных преобразований. Сюда относятся и преобразования по предварительно выбранным для отношений формально-логическим законам, и эвристические дополнения этих правил и раскрытие терминов «вниз» по их родовидовым связям.

Отметим, что в конце 70-х годов XX века, когда разрабатывался этот подход, в распоряжении его авторов была машина класса БЭСМ-6, значительно уступающая по своим возможностям сегодняшней персональной машине. Процедуры упомянутой выше сложности в те годы представлялись невыполнимыми, прежде всего, по масштабам сложности необходимой для этого базы данных и недостаточностью разработки алгоритмов автоматического разбора текстов на естественном языке, что тоже определялось возможностями вычислительной техники.

Сегодня такие задачи, не имеющие точного математического решения, как задачи «ввода семантики в машину», все более успешно решаются на уровне множественных эвристических правил, накапливаемых в машине по мере разбора все большего числа реальных текстов единой тематической направленности.

Практически неограниченные вычислительные возможности современной техники относительно 70-х годов сняли ограничения на объемы накапливаемого материала и быстродействия. Остались ограничения просто на время реальной работы исследователей, на качество выделения эвристик, т.е. не машина стала определять «количество» введенной семантики (эвристик анализа текстов, заменяющих семантику), а экономические соображения, связанные с пониманием тем или иным заказчиком целесообразности проведения таких работ.

Однако некоторые коммерческие фирмы не смотря ни на что, финансируют такие разработки, понимая, что реальных успех в анализе текстов может быть достигнут не только на основе совершенствования формальной логики, но, скорее на использовании огромных мощностей современных машин.

Действительно, необходимость создания коммерчески полезных продуктов типа семантических переводчиков, языковых анализаторов для целей коррекции грамматики текста, информационно-поисковых аналитических систем на естественном языке, систем семантического контроля за потоками сообщений и другие аналогичные работы привели к новым «силовым» решениям теоретически неразрешенных проблем. Во многом эти решения можно иллюстрировать их сходством с заменой, например, устройства перемножения на матрицу уже готовых результатов, что в ряде случаев дает существенный выигрыш в быстродействии.

Конечно, такие системы, по всей видимости, никогда не смогут стать «безошибочными», но постепенно, по мере проведения такого рода «эвристического обучения» и этот фактор становится приемлемым. Как известно, далеко не каждый человек ныне способен понять и разобрать смысл поступающей к нему информации. Последняя становится все более многоплановой и ее полная интерпретация доступна далеко не каждому. Успех обучения системы «с собственным мнением» зависит от непротиворечивости входной информации. Такого рода система может рассматриваться одновременно и как потребитель результатов предикатного преобразования текстов, и как стенд для проверки качества (непротиворечивости) такой процедуры.

Как известно, трактовка полученных сообщений опирается на их интерпретацию с учетом ранее накопленного знания, в том числе и знания, полученного «в других тематических полях». Не зря же учебный процесс строится не в произвольном порядке, но в выверенной поколениями учащихся последовательности изучения отдельных предметов. Кроме всего прочего это означает, что в той или иной степени обучающегося надо подготовить к восприятию нового тематического поля.

В связи с нереальностью представления полного цикла обучения в модельном варианте, для целей снабжения модели «начальным знанием» можно использовать некоторое количество специальных словарей терминов и их сочетаний, словарей возможных терминов–отношений, родовидовых словарей и другое. В педагогической литературе такое априорное знание в некоторой степени соответствует термину «умение».

Литература

- [1] *Лачинов В. М., Поляков А. О.* Информодинамика или Путь к Миру открытых систем. — СПб.: Изд. СПбГТУ, 1999. — 432 с.
- [2] *Ерофеев А. А., Поляков А. О.* Интеллектуальные системы управления. — СПб.: Изд. СПбГТУ, 1999. — 264 с.
- [3] *Булкин Г. А., Григорьев В. В., Поляков А. О.* Организация геологической автоматизированной информационно-логической системы на основе языка предикатов. В кн.: Алгоритмические модели автоматизации исследований. — М.: Наука, 1980. — с. 221-235
- [4] *Однобоков В. В.* Проблемы разработки активных обучающихся и анализирующих систем. — СПб.: Изд-во СПбГТУ, 2002. — 96 с.