

удк 007.5

Г. С. Осипов

Работы исследовательского центра искусственного интеллекта

Аннотация. В статье дана краткая характеристика существующих тенденций в искусственном интеллекте, описаны основные работы в этой области, выполняемые в исследовательском центре искусственного интеллекта. В частности, работы в области разработки и реализации инструментальных средств, методов семантического метапоиска и методов анализа текстов.

Ключевые слова и фразы: Динамические системы; моделирование поведения; динамические модели, основанные на знаниях; анализ полуструктурированной информации; искусственный интеллект.

Введение

В комплексе научных дисциплин под названием «Искусственный интеллект» можно выделить сегодня две тенденции: первая — продолжение теоретических и экспериментальных исследований с целью создания новых методов и технологий и вторая — желание применить уже наработанный инструментарий к решению полезных прикладных задач.

Основные работы первого направления можно кратко обозначить так:

- (1) обучение;
- (2) рассуждения;
- (3) поведение.

Перечисление работ второго направления занимает большее место.

Работы выполнены при поддержке РФФИ (гранты № 03-01-00853, № 02-01-00308, № 03-01-00808), Минпромнауки России (контракт № 37.011.11.0020), Программы ОИВТС РАН «Фундаментальные основы информационных технологий и систем», Программы президиума РАН «Математическое моделирование и интеллектуальные системы».

Уже существующие технологии искусственного интеллекта легли в основу систем управления силовыми установками, систем ориентации и навигации космических аппаратов, управления беспилотными воздушными транспортными средствами; решаются задачи планирования поведения коалиций и коллективов искусственных устройств; созданы и эксплуатируются автономные мобильные платформы и роботехнические устройства (робокраны, робокары и др.); выполняется диагностика оборудования, работающего в непрерывном производственном цикле (например, турбин); создаются автономные мобильные средства ведения боевых операций; разработаны системы моделирования бизнес-процессов на основе систем бизнес-правил, системы управления сборочным производством и испытаниями сложных технических систем, интеллектуальные системы управления большими телекоммуникационными сетями, интеллектуальные агенты, предназначенные для Web-поиска и электронной коммерции; системы интеллектуального анализа данных и анализа документов; интеллектуальные системы мониторинга среды; реализован ряд систем определения источников загрязнения и прогнозирования последствий, мониторинга, анализа и прогнозирования социальной напряженности, социальных, межэтнических и иных конфликтов, анализа заболеваемости в регионах и определения возможных техногенных причин; системы управления лечением; интеллектуальные атласы автомобильных дорог и др.

Из приведенного далеко не полного перечня видно, что приложения носят весьма различный и достаточно солидный характер.

В работах исследовательского центра искусственного интеллекта присутствуют обе составляющие, однако, создание новых инструментальных средств и основанных на них технологий является приоритетным.

Работы по созданию новых технологий, с одной стороны, позволяют находить непосредственное применение новым моделям и методам, если таковые возникают, с другой стороны — находятся на минимальном расстоянии от прикладных задач и, когда возникает потребность в них, позволяют плавно к ним переходить.

1. Динамические интеллектуальные системы

Работы по созданию новых технологий и соответствующих инструментальных средств включают как теоретическую, так и прикладную стадии.

Известно, что в свое время в ИЦИИ была создана первая отечественная технология для построения экспертных систем — SIMER + MIR [1]. Эта технология нашла применение при создании большого класса прикладных систем, однако ей было присуще одно существенное ограничение — в ней не было динамики.

Таким образом, из области её возможного применения выпадал довольно большой класс задач, а именно, задачи моделирования поведения.

Для решения таких задач требуется привлечение арсенала динамических систем.

Однако, характер управляемых объектов в нашем случае имеет ряд существенных особенностей, связанных с тем, что речь обычно идет об объектах, параметры состояния которых могут описываться переменными различной природы — например, количественными, логическими и лингвистическими, а состояния не имеют исчерпывающего априорного описания. Собственно, такие системы и являются предметом исследования в искусственном интеллекте. Добавим, что законы поведения таких систем либо не имеют исчерпывающего аналитического описания, либо не имеют никакого аналитического описания, но могут быть описаны некоторой совокупностью экспертных или эмпирических знаний, например, множеством экспертных правил или, например, семантическими сетями или сетями Петри.

Такие постановки часто возникают при попытках моделирования поведения сложных технических, экологических, социальных, политических, этнических и иных систем.

1.1. Теоретические результаты. Предложен и исследован новый класс динамических моделей — динамические модели, основанные на знаниях (ДМЗ), динамика которых описывается экспертными и эмпирическими знаниями; исследованы вопросы устойчивости ДМЗ и компенсации возмущений [2,3]. Исследованы архитектуры баз знаний и установлена связь архитектурных особенностей БЗ с множествами достижимых состояний ДМЗ; построены алгоритмы вычисления траекторий ДМЗ, рассмотрены задачи идентификации и наблюдаемости [4,5].

Уточнено понятие плана целенаправленного поведения динамической модели, основанной на знаниях, предложен и исследован эффективный метод автоматического синтеза планов и установлены

условия существования планов; исследовано целенаправленное поведение в условиях динамической смены целевых состояний [6, 7].

Выполнен ряд исследований сетей Петри, в частности, исследованы формальные модели распределенных систем с динамической структурой, основанные на сетях Петри и системах конечных автоматов [8].

Определен формализм сообществ взаимодействующих автоматов. Исследована выразительность данной модели. Доказано, что сообщества взаимодействующих автоматов и недетерминированные автоматы с магазинной памятью не сравнимы по выразительности, а именно, классы их языков пересекаются, но ни один из них не вкладывается в другой. Для взаимодействующих автоматов доказана также разрешимость проблем останова и достижимости. Продолжены исследования семантических свойств вложенных сетей Петри. Для многоуровневых сетей Петри доказана разрешимость проблемы потенциальной живости данного перехода при данной начальной разметке и неразрешима проблема живости сети в целом [9–11].

Выполнены работы по проблемам формального анализа корректности программных систем; изучен класс логических исчислений, называемых нечёткими модальными логиками [12]. Описана семантика указанных исчислений в классе нечетких моделей Крипке [13].

1.2. Инструментальные средства. На основе перечисленных результатов разработана архитектура динамических систем, основанных на знаниях, и реализованы соответствующие инструментальные программные средства, включающие: базу знаний, средства динамического целеуказания, механизм динамического планирования, механизм анализа текущего состояния и механизм управления системой [14]. Для повышения эффективности работы программистов и экспертов созданы текстовый и графический интерфейсы. Работа системы проверена на ряде модельных задач [15].

База знаний хранит и предоставляет методы доступа к когнитивным структурам. Для хранения когнитивных структур используется гибридный метод представления знаний на основе правил, семантических сетей и фреймов. Архитектура базы знаний позволяет корректно представлять классы объектов, ситуаций, процессов предметной области, а также существенные взаимосвязи между ними.

В инструментарии представлены средства решения задач в предметных областях со сложной динамикой и структурой. Средства динамического целеуказания позволяют устанавливать необходимые целевые состояния интеллектуальной системы в зависимости от конкретной ситуации. Механизм интеллектуального динамического планирования прогнозирует изменение состояния предметной области, учитывая всевозможные управления в будущих состояниях, и вырабатывает план действий для достижения цели. Механизм анализа и пополнения описания состояния на основе текущего состояния предметной области выводит дополнительные факты, замыкая тем самым описание состояние предметной области. Механизм управления изменяет основные параметры процесса управления в соответствии с текущими целями и выработанным планом. Работа интеллектуальных систем в условиях динамичной внешней среды подразумевает своевременную реакцию на изменение параметров среды, что накладывает временные ограничения на функционирование механизмов анализа, управления, планирования и средства динамического целеуказания. Инструментальные средства программно реализованы, как для фон-неймановской, так и для кластерной архитектуры вычислительных средств.

2. Точный поиск в локальных, глобальных вычислительных сетях и базах данных

В настоящее время в связи с развитием глобальных сетей и роста объемов циркулирующей в них информации существенно возросла важность проблем её релевантного поиска и анализа [16]. Однако, хорошо известно, что применяемые в существующих поисковых системах методы не позволяют достичь высокой полноты и точности поиска.

Детальное обсуждение возникающих здесь проблем и путей их решения приведено в одной из статей настоящего сборника, поэтому мы кратко остановимся лишь на основных идеях, приведших к появлению новой системы поиска информации, обеспечивающей высокое качество поиска. Основное соображение состояло в том, что высокого качества поиска (главным образом, точности) можно достичь, сопоставляя ситуации, описываемые в запросе и искомым документах. Отсюда немедленно следовало, что запрос должен быть задан не

списком ключевых слов, а фразой на языке предметной области. Таким образом, первой задачей явилась задача анализа фразы на естественном языке, а именно, такого анализа, который был бы достаточен для извлечения из фразы описания ситуации. При этом не было желания привязывать процесс к какой бы то ни было предметной области.

Из этих двух посылок следует, что анализ запроса (и фраз текста) должен опираться, главным образом, на лингвистическую (т.е. морфологическую и синтаксическую) и некоторую семантическую информацию. При этом мы были крайне осторожны в привлечении семантической информации, чтобы избежать привязки системы к какой-либо предметной области. Экспериментально было установлено, что для целей поиска достаточно описания ситуации в виде ролевого фрейма. С лингвистической точки зрения это означало использование семантических падежей [17, 18], вместе с некоторыми дополнительными отношениями [19], а в качестве минимальной лексической единицы, служащей для заполнения валентностей в падежных (ролевых) структурах, была выбрана именная синтаксема [17, 18].

Таким образом, основной акцент был сделан на формировании семантического (в указанном смысле) представления запроса и построении семантического поискового представления документа (в дальнейшем — поискового образа документа). Поисковый образ документа — это индекс пар роль, именная синтаксема; именная синтаксема = предлог + падеж следующего за ним существительного. Релевантность (релевантность 1) определяется на основе сопоставления поисковых образов документа и запроса.

Второй способ вычисления релевантности основан на сравнении семантических сетей запроса и фрагмента документа, содержащего ключевые слова запроса. Фрагменты передаются для обработки модулю семантического анализа, который строит их поисковый образ. В данном случае поисковый образ — это индекс троек тип семантической связи, 1-я именная синтаксема, 2-я именная синтаксема. Далее, построенный таким же способом, образ запроса сравнивается с образами фрагментов документа. На основе этого вычисляется релевантность 2.

Система, построенная на описанных принципах, демонстрирует достаточно высокую точность поиска, близкую к 90% [20].

3. Задачи анализа текстов

Еще одно из направлений, развиваемое в работах Центра — анализ полуструктурированной информации. В данном случае речь идет о задачах обработки текстовой информации. Некоторые виды такой обработки в последние годы приобрели высокую значимость, вызванную резким увеличением количества нуждающейся в обработке информации.

К числу таких работ относятся задачи автоматической классификации текстов и задачи извлечения из неструктурированных текстов информации в структурированной форме.

3.1. Классификация текстов. Задачи классификации текстов решались с применением достаточно ограниченного арсенала лингвистических средств — морфологического анализа, лемматизации, элементов синтаксического анализа, направленного, главным образом, на распознавание именных групп и некоторых иных лингвистических механизмов.

Основным в этих задачах являлся механизм формирования образов классов — взвешенных векторов ключевых слов, соответствующих каждому из классов. Этот вектор формировался в результате обучения для каждого набора классов на основе соответствующей обучающей выборке.

Результатом работ этого направления явились системы классификации текстов КЛАСТЕР и АКТИС (на фон-неймановской и кластерной архитектуре, соответственно), демонстрирующие точность классификации 87%.

3.2. Извлечение данных. С лингвистической (как и программистской) точки зрения — задача, значительно более сложная, нежели задача классификации.

Прежде всего отметим, что разработка программных средств, ориентированных на извлечение данных только лишь одного определенного сорта, во многих отношениях неэффективна. Независимо от того, какого рода данные требуется извлечь из текста, выполняется во многом пересекающийся код. Это означает, что программные средства для решения такого рода задач должны иметь характер инструмента, с большими или меньшими усилиями адаптируемого под

конкретную задачу. Такие программные средства основаны на многоуровневом анализе текста и используют механизм извлечения информации, основанный на правилах. Последующие уровни анализа текста используют результаты, полученные на предыдущих этапах [21].

На каждом уровне лингвистический процессор предоставляет дополнительную информацию о тексте. Процесс получения дополнительной информации при анализе текста называется аннотированием. Любая информация о тексте (фрагменте текста) представляется в виде аннотации. Аннотация сопоставляется фрагменту текста, принадлежит классу аннотаций и обладает атрибутами. На основании аннотаций алгоритм извлечения информации выполняет поиск фрагментов текста, релевантных цели. Структурное описание релевантных фрагментов выполняется на языке описания правил. Правила компилируются в конечные автоматы и применяются к порожденному на ранних этапах анализа текста графу аннотаций [22].

Эти работы привели к реализации системы извлечения информации ИНЭКС. Подробные описания всех названных систем приведены в настоящем сборнике.

Список литературы

- [1] Osipov G. S., Gukova S. M., Komarov S. I., Kurshev E. P. Expert system tools for badly structured fields // Artificial Intelligence and Information-Control Systems of Robots-87 / ред. Plander I.: Elsevier Science Publishers B.V. (North-Holland), 1987. ↑1
- [2] Osipov G. *Dynamics in Integrated Knowledge-Based Systems* // Proceedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference, Gaithersburg, MD, USA.— Т. 5, 1998, с. 24–28. ↑1.1
- [3] Осипов Г. С. *Динамика в системах, основанных на знаниях* // Теория и системы управления. Известия АН, № 5, с. 24–28. ↑1.1
- [4] Лебедева Т. Г., Осипов Г. С. *Архитектура и управляемость дискретных динамических систем, основанных на знаниях* // Теория и системы управления. Известия АН, № 5, с. 37–43. ↑1.1
- [5] Osipov G. *Attainable Sets and Knowledge Base Architecture in Discrete Dynamic Knowledge-based Systems* // Proc. of the Workshop «Applied Semiotics: Control Problems (ASC 2000)» 2000. 14-th European Conference of Artificial Intelligence, Berlin (ECAI 2000), 2000, с. 39–43. ↑1.1

- [6] Виноградов А. Н. *Интеллектуальная система с динамически изменяющейся целью* // Труды конгресса ICAI'2001, 3-8 сентября 2001: Физматлит, 2001, с. 339–348. ↑1.1
- [7] Виноградов А. Н., Жилякова Л. Ю., Осипов Г. С. *Динамические интеллектуальные системы. I. Представление знаний и основные алгоритмы* // Теория и системы управления. Известия АН, № 6, с. 119–127. ↑1.1
- [8] Ломазова И. А. *Вложенные сети Петри: моделирование и анализ распределенных систем с объектной структурой*. Монография. — М.: Научный мир, 2004. ↑1.1
- [9] Bashkin V. A., Lomazova I. A. *Petri nets and resource bisimulation*. — Т. 56, №2, 2003, с. 101–114. ↑1.1
- [10] Башкин В. А., Ломазова И. А. *Бисимуляция ресурсов в сетях Петри* // Теория и системы управления. Известия АН, № 4, с. 115–123. ↑
- [11] Ломазова И. А. *Сети Петри и анализ поведенческих свойств распределенных систем*. Учебное пособие. — Ярославль, 2002. ↑1.1
- [12] Миронов А. М. *Нечёткие модальные логики* // Фундаментальная и прикладная математика. — 9, вып.1, с. 201–230. ↑1.1
- [13] Миронов А. М. *Каноническая реализация совокупностей морфизмов реакции для автоматов в категориях* // Математика, информатика: теория и практика. Сборник трудов, посвященный 10-летию Университета города Переславля ред. А.К. Айламазян. — Переславль-Залесский: Издательство УГП, 2003, с. 65–88. ↑1.1
- [14] Тарханов Т. С. *Архитектура и ядро комплекса инструментальных программных средств для создания динамических интеллектуальных систем* Прикладная математика и компьютерное моделирование, № 1, 2003, с. 9–13. ↑1.2
- [15] Виноградов А. Н., Лебедева Т. Г. *Применение методов искусственного интеллекта в социально-экономических задачах* Прикладная математика и компьютерное моделирование. — Т. 2 №1, 2003, с. 39–43. ↑1.2
- [16] Куршев Е. П., Осипов Г. С., Рябков О. В., Самбу Е. И., Соловьева Н. В., Трофимов И. В. *Интеллектуальная метапоисковая система* // Труды международного семинара Диалог'2002 «Компьютерная лингвистика и интеллектуальные технологии». — Т. 2. — М.: Наука, 2002, с. 320–330. ↑2
- [17] Золотова Г. А. *Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса*. — М.: Эдиториал УРСС, 2001, с. 201–230. ↑2
- [18] Золотова Г. А., Онипенко Н. К., Сидорова М. Ю. *Коммуникативная грамматика русского языка*. — М.: Изд. МГУ, 1998. ↑2
- [19] Осипов Г. С. *Приобретение знаний интеллектуальными системами: Основы теории и технологии*. — М.: Наука, Физматлит, 1997. ↑2
- [20] Осипов Г. С., Куршев Е. П., Кормалев Д. А., Трофимов И. В., Рябков О. В., Тихомиров И. А. *Семантический поиск в среде интернет*. Препринт. — Переславль-Залесский: ИПС РАН, 2003. ↑2
- [21] Кормалев Д. А., Куршев Е. П., Сулейманова Е. А., Трофимов И. В. *Извлечение данных из текста. Анализ ситуаций ньюсмейкинга* // Труды Восьмой национальной конференции по искусственному интеллекту с международным участием КИИ-2002. — М.: Физматлит, 2002, с. 199–206. ↑3.2

- [22] Кормалев Д. А., Куршев Е. П., Осипов Г. С., Сулейманова Е. А., Трофимов И. В. Методы поиска и анализа информации. Автоматическое извлечение данных. Препринт. — Переславль-Залесский: ИПС РАН, 2003. ↑[3.2](#)

ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ИПС РАН

G. S. Osipov. *Activity of AI Research Centre.* (in russian.)

ABSTRACT. This paper presents brief characteristics of the existing directions in the field of artificial intelligence, describes the principals in this area that are carried out by Artificial Intelligence Research Center. In particular, the works on development and implementation of AI-tools, methods of semantic meta search and methods of the text analysis are also presented.