ТЕХНОЛОГИИ И СИСТЕМЫ ДЛЯ РЕШЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ РАСПОЗНАВАНИЯ ОБРАЗОВ

С. В. Абламейко, В. В. Краснопрошин, В. А. Образцов

Белорусский государственный университет
Минск, Беларусь
E-mail: (ablameyko, krasnoproshin, obraztsov)@bsu.by

Рассмотрены принципы построения технологии для решения задачи распознавания образов с обучением. Описаны некоторые компьютерные системы и приведены примеры их использования для решения конкретных практических задач медицинской диагностики, обработки изображений.

Ключевые слова: распознавание образов, технология построения компьютерных систем, практические задачи.

Распознавание образов является одним из актуальных направлений информатики, связанным с автоматизацией процессов обработки и использования информации. Несмотря на серьезные успехи в решении практических задач, методология распознавания остается по сути «эвристической». Даже на уровне теории трудно построить суждение об обоснованности моделей, методов и алгоритмов, что не позволяет говорить о правильности полученных результатов. Это существенно отражается и на принципах построения компьютерных систем, ориентированных на решение практических задач.

Начнем с формулировки стандартной задачи распознавания образов [1]:

на множестве объектов $\boldsymbol{X}$ *произвольной природы задано некоторое (возможно и бесконечное) число подмножеств (классов)* $\boldsymbol{X_1},...,\boldsymbol{X_l}$. *Задана также некоторая информация о классах* $\boldsymbol{I_0}(\boldsymbol{X_1},...,\boldsymbol{X_l})$. *Требуется, используя только информацию* $\boldsymbol{I_0}$, *указать алгоритм* $\boldsymbol{A}$, *который определен на всем множестве* $\boldsymbol{X}$ *и результат работы которого для каждого* $\boldsymbol{x} \in \boldsymbol{X}$ *можно интерпретировать в терминах принадлежности классам* $\boldsymbol{X_i}$ $(i \in \mathbb{N})$.

Информация $\boldsymbol{I_0}$ задает некоторую выборку $\boldsymbol{X^0}$, удовлетворяющую условиям: $|\boldsymbol{X^0}| < +\infty$, $\boldsymbol{X^0} \subset \boldsymbol{X}$. Кроме того, важен способ представления выборки. В этом смысле возможны два варианта: выборка задается **прецедентами** или с использованием **принципа свертки**. Обозначим соответствующие задачи через $\mathbf{Z_1}$ и $\mathbf{Z_2}$.

Рассмотрим понятие технологии, которое возникает вследствие необходимости автоматизации процессов обработки информации при решении задачи (рис. 1). Под **технологией** будем понимать набор инструментов и приемов, направленных на построение компьютерной системы для автоматизации процесса решения задачи. Технология должна обеспечивать всю необходимую функциональность во взаимодействии объектов, представленных на рис. 1. В данной трактовке уточняется стандартное понимание информационной технологии [2], которое описывает роль компьютерной системы.

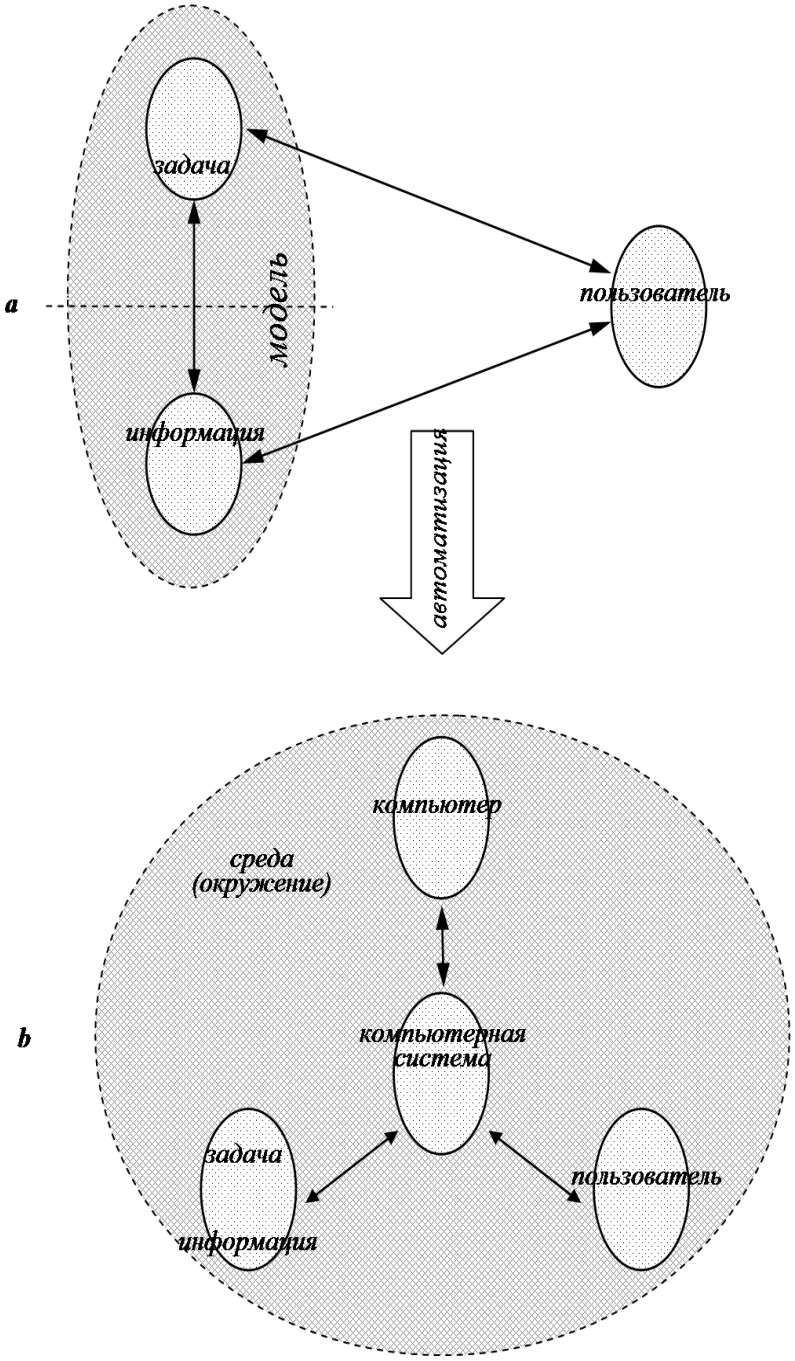

Рис.1. Схема взаимодействия объектов в процессе решения задачи

Приведем основные результаты, опубликованные в работах [3–7]. В области информационных технологий можно сформулировать несколько идей, положенных в основу построения систем. Описанные в виде принципов, они и определяют искомое понятие технологии. Следствием любой технологии является архитектура системы, на основе которой можно строить компьютерную систему. Последняя чаще всего не связана с предметной областью, и может использоваться для решения различных прикладных задач. Опишем технологию в форме тезиса.

Тезис 1. *Для решения задачи* $\mathbf{Z}_1$ *технология может быть построена на следующих принципах:*

*Принцип 1 (**принцип отделимости**). Данные и знания должны быть отделены от средств манипулирования информацией.*

*Принцип 2 (**принцип структурирования**). Задача* $\mathbf{Z_1}$ *должна быть подходящим образом структурирована, сведена к множеству неприводимых задач. Результат решения* $\mathbf{Z_1}$ *представляет суперпозицию результатов решения подходящих неприводимых задач.*

*Принцип 3 (**принцип модульности**). Структура системы должна быть модульной. Каждый модуль отвечает за решение отдельной неприводимой задачи.*

*Принцип 4 (**принцип инвариантности**). Компьютерная система, построенная на базе принципов 1–3, может решать любую задачу* $\mathbf{Z_1}$ *и в пределах особенностей данной постановки не зависит от предметной области (характера информации).*

На базе описанной технологии, с учетом особенностей систем программирования и вычислительной техники 80-х гг. XX в., была построена компьютерная система ПАРУС (Полигон Алгоритмов Распознавания с Управляющей Системой). Архитектура данной системы приведена на рис. 2.

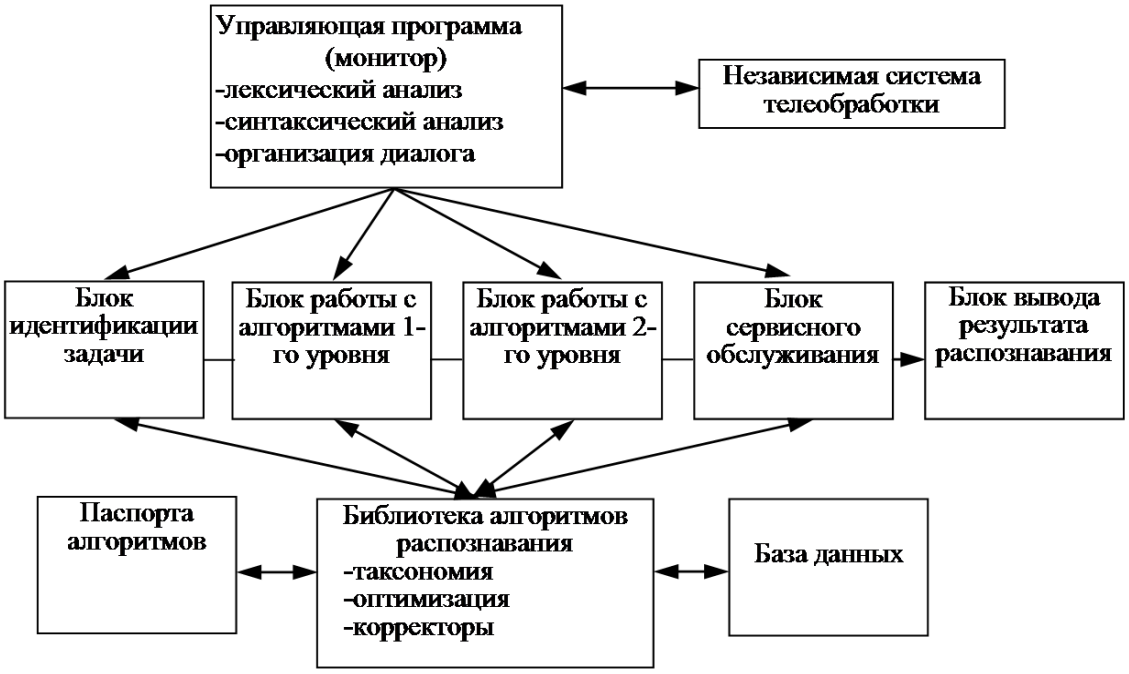

Рис.2. Архитектура компьютерной системы ПАРУС

В технологическом смысле процесс решения задачи в постановках $\mathbf{Z_1}$ и $\mathbf{Z_2}$ практически одинаков, но имеют небольшое различие на уровне архитектуры [8].

Сформулируем принцип, на базе которых может быть решена задача $\mathbf{Z_2}$.

Тезис 2. *Для решения задачи* $\mathbf{Z_2}$ *технология может быть построена на принципах 1–4 и следующем:*

*Принцип 5 (**принцип разделения**). Данные и знания должны быть разделены и эти виды информации можно использовать для получения результата независимо.*

Структура соответствующей системы показана на рис. 3.

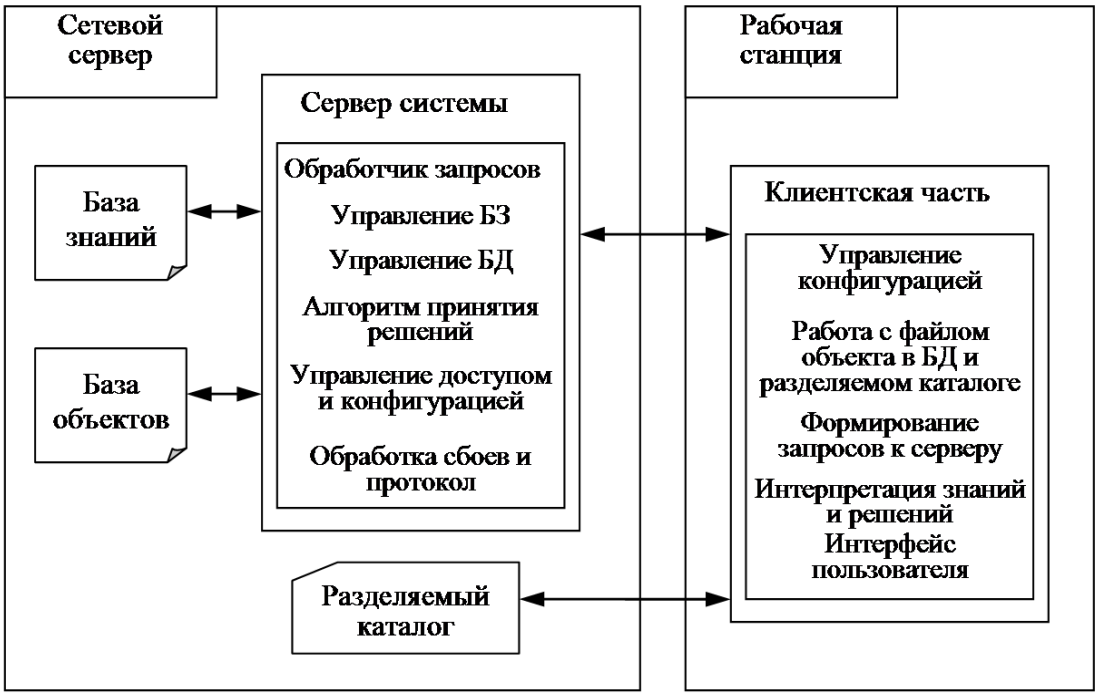

Рис.3. Структура компьютерной системы Орто-Эксперт

Приложения в интеллектуальном анализе данных. На базе описанных технологий были построены компьютерные системы, которые использовались для решения задач из области интеллектуального анализа данных. Опишем кратко суть задач и основное содержание полученных результатов. Более подробно с ними можно ознакомиться в цикле работ [3–8].

Обработка данных дистанционного зондирования Земли. Рассматривалась задача классификации подстилающих поверхностей Земли по данным космического спектрометрирования и распознавание нефтеносности месторождений по результатам изучения близлежащих водоносных пластов. Для сравнения были выбраны 11 эвристик: варианты алгоритмов вычисления оценок, ближайших соседей, потенциальных функций. По результатам экспериментов формировались наборы для построения модели алгоритмов распознавания [1] из «слабокоррелированных» алгоритмов. Далее исследовалась эффективность различных типов корректоров. Определялся объем обучения и его зависимость от числа алгоритмов в корректируемом наборе. Экспериментальные исследования позволили в целом сделать вывод, что двухуровневые схемы с корректором [8] являются разумным подходом к повышению качества и надежности работы систем распознавания.

Решались задачи обработки изображений и выделения требуемых объектов на снимках земной поверхности. Предложенные подходы позволили автоматизировать режим обновления снимков на основе их совмещения с цифровыми картами.

Обработка медицинских данных. На основе предложенных технологических и структурных решений, в результате практического взаимодействия со специалистами-медиками, была реализована серия компьютерных систем с общим названием – Орто-Эксперт (для поддержки процессов диагностики и лечения пациентов с нарушениями опорно-двигательного аппарата). Некоторые специализированные варианты

системы внедрены в Республиканском центре и областных диспансерах спортивной медицины. Их практическое использование обеспечило эффективное обследование, обоснованную диагностику и выбор соответствующего метода лечения, что повысило качество диагностики ортопедических заболеваний и травм и сократило сроки реабилитации спортсменов. Система использовалась в сборных командах республики по различным видам спорта на Пекинской олимпиаде. В 2011 г. на выставке инновационных технологий в Санкт-Петербурге система была награждена серебряной медалью как лучший инновационный продукт.

На основе предложенных технологий были также разработаны системы [5–7], в рамках которых решены проблемы анализа состояния и диагностики рака щитовидной железы по гистологическим, ультразвуковым и цитологическим изображениям клеток. Диагностики злокачественных новообразований органов средостения и забрюшинного пространства у детей на основе анализа изображений, полученных с помощью лучевых методов исследования и др.

Таким образом, предложенные технологии получили практическое применение при разработке компьютерных систем, ориентированных на решение важных народнохозяйственных задач.

ЛИТЕРАТУРА

1. *Журавлёв, Ю. И.* Избранные научные труды. М.: Магистр, 1998.
2. http://ru.wikipedia.org/wiki/Технология
3. *Krasnoproshin, V.* Solution of applied problems: formalization, methodology and justification / V. Krasnoproshin, V. Obraztsov, H. Vissia // Computational Intelligence in Business and Economics, London: World Scientific, 2010. Vol. 3. P. 57–64.
4. *Krasnoproshin, V. V* Problem of Solvalibility and Choice of Algorithms for Decision Making by Precedence / V. Krasnoproshin, V. Obraztsov // Pattern Recognition and Image Analysis. 2006. Vol. 16. № 2. P. 155–169.
5. *Абламейко, С. В* Обработка изображений: технология, методы, применение / С. В. Абламейко, Д. М. Лагуновский. Минск: Ин-т техн. кибернетики НАН Беларуси, 1999.
6. *Абламейко, С. В.* Медицинские информационные технологии и системы / С. В. Абламейко, В. В. Анищенко, В. А. Лапицкий, А. В. Тузиков. Минск: ОИПИ НАН Беларуси, 2007.
7. *Абламейко, С. В.* Обработка оптических изображений клеточных структур в медицине / С. В. Абламейко, А. М. Недзьведь. Минск: Объед. ин-т проблем информатики НАН Беларуси, 2005.
8. *Краснопрошин, В. В*. Проблема принятия решений по прецедентности: разрешимость и выбор алгоритмов / В. В. Краснопрошин, В. А. Образцов // Выбраныя навуковыя працы Беларускага дзярж. ун-та. Т. 6. Матэматыка, 2001. С. 285–312.