

УДК 004.82

П.И. Сагайда

Донбасская государственная машиностроительная академия, г. Краматорск

**ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОБЛЕМНОЙ ОБЛАСТИ
«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ
ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ»****Аннотация**

Сагайда П.И. Онтологическое моделирование проблемной области «Интеллектуальный анализ данных для поддержки принятия решений». Разработана онтология объектов и задач проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений», которая позволяет определить направление теоретических исследований представления знаний и моделирования процессов обработки данных. Процесс принятия решения проанализирован и представлен в виде онтологии концептов с точки зрения задач, которые могут быть решены (прогноз, классификация, кластеризация, выбор), а также с точки зрения предмета выбора и результата выбора. Рассмотрены разновидности инструментов искусственного интеллекта, в том числе результаты гибридизации отдельных подходов, и возможные результаты их применения.

Ключевые слова: интеллектуальный анализ данных, поддержка принятия решений, проблемное пространство, онтология, методы искусственного интеллект.

Постановка задачи в общем виде и ее связь с важными научными и практическими заданиями.

Интеллектуальный анализ данных (ИАД) используется в современной практике при создании и обработке хранилищ данных с историческими и оперативными сведениями, в том числе хранимыми в различных форматах и представлениях. Такие хранилища создаются для анализа и подтверждения гипотез о зависимостях, существующих между поведением объектов в предметной области (ПО) и между параметрами отдельных объектов, в случаях, позволяющих сформулировать и применить такие модели зависимостей для прогнозирования и поддержки принятия решений.

ИАД подразумевает прежде всего использование интеллекта аналитика для проведения предварительной обработки данных, представления и просмотра (осмысления, интерпретации результатов), построения неформализованных, часто лишь интуитивно обоснованных моделей в ходе аналитической деятельности человека. Такой подход реализуется в системах отчетности на основе запросов к базам данных, в частности, в OLAP-системах [1]. Кроме того, в ходе ИАД возможно привлечение всего алгоритмического арсенала методов математической статистики и машинного обучения, позволяющих аналитику, с использованием своих умений и навыков как инженера по знаниям и специалиста в области обработки данных, применить для выявления существенных зависимостей различные подходы в определенной последовательности к одному и тому же, либо последовательно трансформируемому набору данных о предыстории развития ПО. При этом упор делается не на общую статистическую обработку набора данных с использованием концепции среднего значения, а на обработку отдельных подвыборок (кластеров) для нахождения шаблонов поведения данных либо объектов [2].

Для поддержки принятия решений важную роль играют форма и качество построения моделей предметной области, которые позволяют выполнить прогноз результата принятия решения (исполнения выбранного управляющего воздействия). Многообразие

математических моделей как форм представления знаний о ПО не позволяет в данной работе представить сколько-нибудь полную онтологию возможных моделей и математического аппарата для ИАД. Упомянем лишь модели, традиционно получаемые в ходе решения таких общепринятых задач ИАД, как регрессия, классификация и выявление ассоциаций [1,2], — регрессионных зависимостей, классификационных и ассоциативных правил как наиболее общих форм представления моделей функционирования ПО для решения задач ИАД. При этом регрессионную зависимость можно рассматривать как частный случай классификационного правила, для которого кроме функциональной зависимости можно указать и интервальные оценки зависимых переменных или их принадлежность множеству, а ассоциативное правило — как обобщение классификационного правила на случай рассмотрения в его условной и утверждающей частях нескольких объектов, участвующих в выборке.

Процесс решения (выработки управляющих воздействий) состоит из различных этапов, на каждом из которых решается множество подзадач, при этом их решение обычно носит параллельный и итерационный характер. Для некоторых подзадач, таких как вычисление закона оптимального управления для системы с известными передаточными функциями звеньев, получение критериальных оценок альтернатив, моделирование предпочтений лица, принимающего решения, выбор оптимального решения и др., существуют достаточно хорошо проработанные на сегодняшний день подходы к их моделированию — большинство таких подходов основано на математической теории принятия решений. С другой стороны, такие подзадачи, как анализ факторов, характеризующих моделируемую ситуацию, разработка прогноза ее развития, синтез и отбор альтернатив и управляющих стратегий, решаются в основном на приближенном, качественном уровне, с помощью интуиции и нестрогих рассуждений. Основная сложность, возникающая при построении моделей таких задач, состоит в том, что аналитическое описание либо статистическое наблюдение зависимостей между входными и выходными параметрами затруднено, а зачастую невозможно, и приходится прибегать к субъективным моделям, основанным на экспертной информации, обрабатываемой с привлечением различных формальных логик, эвристик и интуитивных подходов аналитика [3,4]. На этом этапе хорошим решением является применение онтологического подхода для моделирования концептов и связей ПО.

Целью статьи является использование онтологического подхода для введения классификации объектов, процессов и задач обработки данных и построения моделей проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений».

Известно, что под онтологией объектов предметной области понимается кортеж четырёх множеств:

$$O^o = \langle X, R, F, A(D, R_s) \rangle,$$

где $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ — конечное множество концептов (понятий) заданной ПО; $R = \{r_1, r_2, \dots, r_k, \dots, r_m\}$, $R : x_1 \times x_1 \times \dots \times x_m$ — конечное множество семантически значимых отношений между концептами ПО, которые определяют тип взаимосвязи между понятиями; $F : X \times R$ — конечное множество функций интерпретации, заданных на концептах и/или отношениях; A — конечное множество аксиом, которые используются для записи всегда истинных высказываний (определений и ограничений).

Построение онтографов с учетом вида онтологии (онтологии задач обработки данных, онтологии объектов, данные о которых обрабатываются, либо онтологии процессов обработки данных) необходимо выполнять с соблюдением следующих требований: ясности (отсутствие смешения декларативного и процедурного аспектов), связности (логические аксиомы начального набора понятий должны быть непротиворечивыми), расширяемости

(изменение совместно используемых словарей понятий без нарушения целостности системы), минимальных обязательств (онтология отображает концептуальную структуру проблемной области, относительно стабильную на протяжении жизненного цикла системы поддержки принятия решений) [5,6]. С использованных рассмотренных выше подходов к онтологическому моделированию построена онтология объектов проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений» (фрагмент), показывающая место самих онтологических моделей в данном пространстве (рисунок 1).

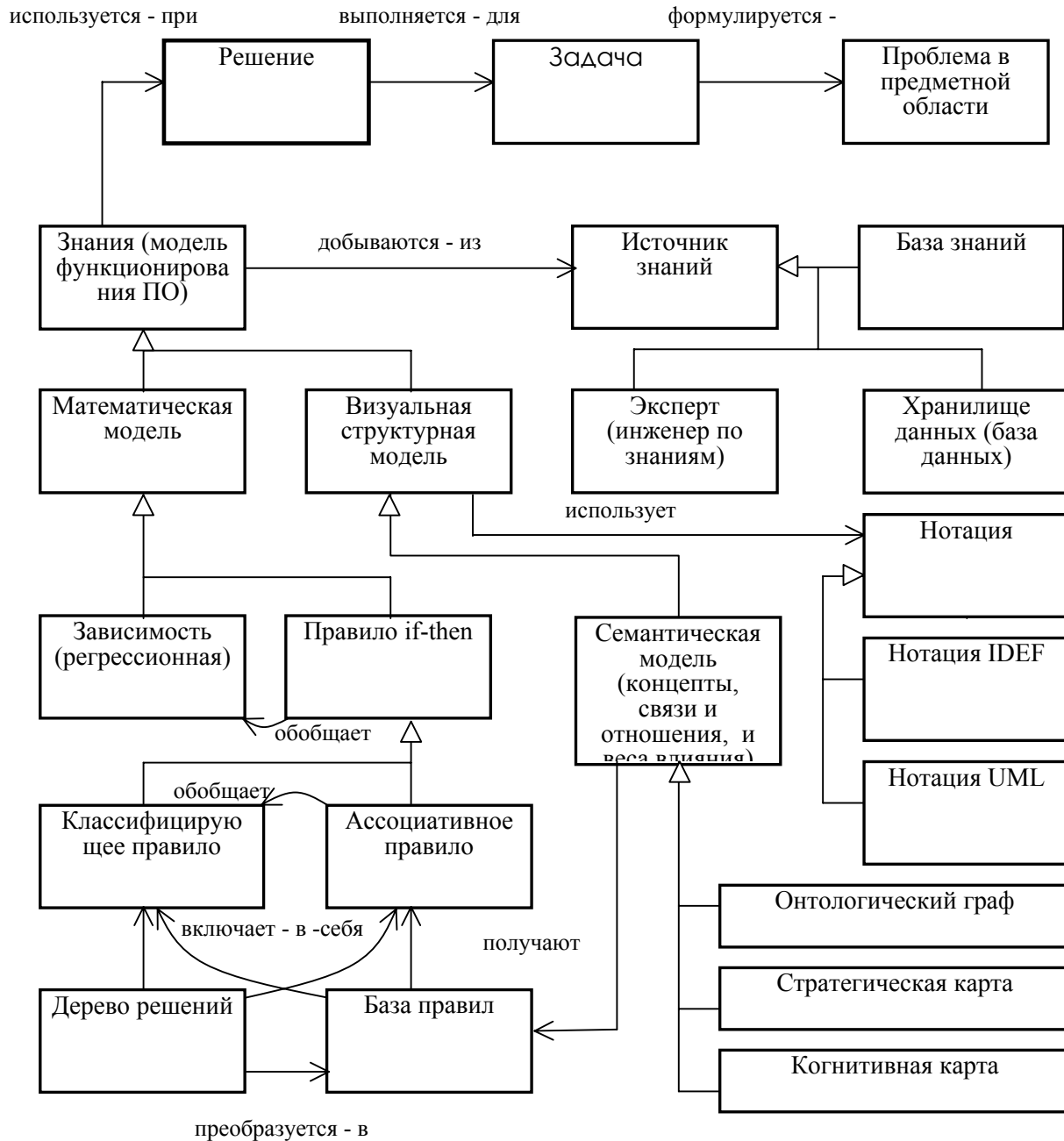


Рисунок 1 – Онтология объектов проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений» (фрагмент), показывающая место самих онтологических моделей в данном пространстве

На данном онтографе приведены основные объекты предметной области, участвующие в процессе добычи знаний и являющиеся его результатом, а также рассмотрена частичная классификация видов получаемых моделей. Особое место среди результатов

моделирования занимает семантическая модель, реализующая визуальный структурный подход к моделированию. В качестве таких моделей могут выступать графы, включающие не только концепты, их связи и отношения, но и данные о весах их взаимного влияния (для дальнейшего имитационного моделирования и поддержки принятия решений).

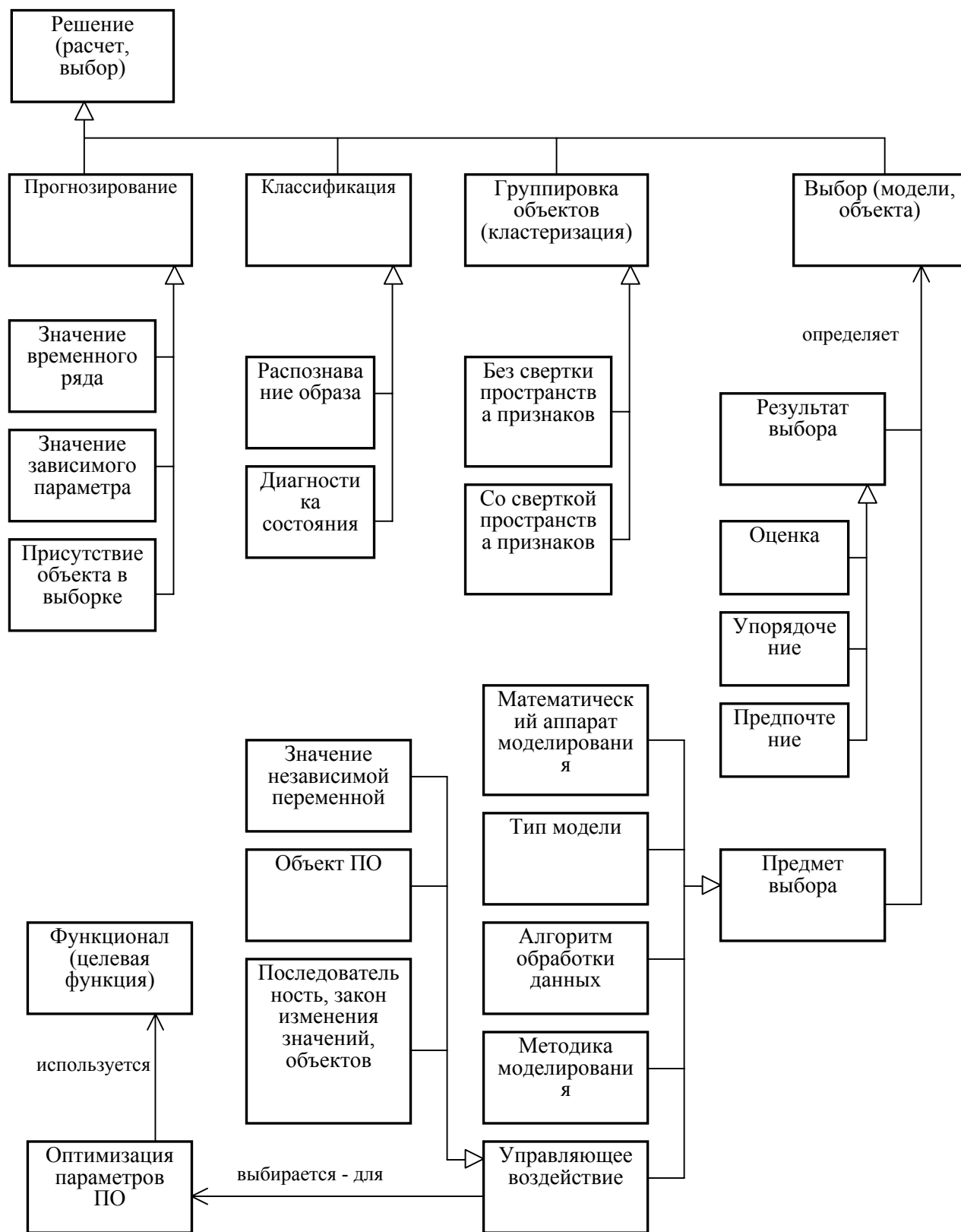


Рисунок 2 – Онтология задач предметной области «Поддержка принятия решений» (фрагмент)

Процесс принятия решения также был проанализирован и представлен в виде онтологии концептов с точки зрения задач, которые могут быть решены (прогноз, классификация, кластеризация, выбор), а также с точки зрения предмета выбора и результата выбора. Выполнено обобщение понятия управляющего воздействия, подклассами которого являются: значение независимой переменной (например, результат выработки уставки для исполнительного устройства); объект ПО (например, альтернатива, лучшая по какому-либо параметру или дающая экстремальное значение функционала); последовательность, или закон, изменения значений, объектов (алгоритм управления технологическим процессом, вариант упорядочения объектов и действий и т.п.).

Результат онтологического моделирования задач предметной области «Поддержка принятия решений» приведен на рисунке 2.

Вместе с тем анализ концептов и вид их взаимного влияния на рисунках 1-2 показывает насущную необходимость повышения интеллектуальности информационного обеспечения деятельности аналитиков, экспертов, специалистов. Как показали исследования, средства искусственного интеллекта обладают рядом преимуществ при их использовании в предпроектной (на этапах сбора, поиска и анализа данных) и проектной (на этапе принятия решений) деятельности.

Разновидности инструментов ИИ, в том числе результаты гибридизации отдельных подходов, подробно рассмотрены в [7].

Построим с помощью онтологического подхода классификацию процессов обработки данных и знаний с помощью методов ИИ. Результат такой классификации приведен на рисунке 3.



Рисунок 3 – Фрагмент иерархии подпроцессов для онтологии процессов проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений»

Упорядочение и подготовка данных должны проводиться, прежде всего, инженерами по знаниям, однако такие задачи, как фильтрация и восстановление данных, выявление выбросов и аномалий, представление данных в виде лингвистических термов и, наоборот, представление данных нечисловой природы в виде точных или интервальных оценок, свертка данных (уменьшение факторного пространства, интеграция отдельных оценок и

критериев) достаточно успешно выполняются с использованием методов ИИ: нейронных сетей с различной архитектурой и методами обучения, нечетких множеств и нечетких мер.

Формализация рассуждений в виде баз знаний с использованием лингвистических термов и функций принадлежности точечных оценок к рассматриваемым критериям и понятиям позволяет выполнять нечеткий вывод, логические и арифметические действия над нечеткими множествами, в том числе при решении задач в условиях действия НЕ-факторов (нечеткости, недоопределенности, неточности и т.д.).

В ходе обучения распознаванию образов (состояний) и реализации родственного подпроцесса «Выявление зависимостей между данными и моментами их появления» использование нейронных сетей как универсальных аппроксиматоров математических зависимостей позволяет получать хорошие результаты.

Процесс принятия решений, представляющий в общем случае выбор из нескольких альтернатив, слабо формализуется при помощи средств многокритериальной оптимизации в условиях отсутствия четко определенной функции цели. Ряд авторов последовательно вводит в этот хорошо исследованный подход различные виды нечеткости и разрабатывает алгоритмы реализации нелинейной свертки частных решений и отдельных ранжировок анализируемых альтернатив.

В случаях, когда имеется возможность обучения программной системы на основе достаточного объема выборок сведений о поведении ПО, хорошо зарекомендовали себя гибридные нейронно-нечеткие системы, интегрирующие способность к обучению нейронных сетей с возможностями нечеткого вывода и наглядностью базы нечетких правил.

Процесс оптимизации работы ПО в условиях неопределенности и при недостаточно формализованных моделях также успешно проводится методами эволюционного моделирования и программирования, а также путем стохастической оптимизации (при помощи генетических и муравьиных алгоритмов).

Выводы.

Для поддержки принятия решений важную роль играют форма и качество построения моделей ПО, которые позволяют выполнить прогноз результата принятия решения (исполнения выбранного управляющего воздействия).

Онтология объектов проблемного пространства «Интеллектуальный анализ данных для поддержки принятия решений» позволила определить направление теоретических исследований представления знаний и моделирования процессов обработки данных.

Процесс принятия решения проанализирован и представлен в виде онтологии концептов с точки зрения задач, которые могут быть решены (прогноз, классификация, кластеризация, выбор), а также с точки зрения предмета выбора и результата выбора. Анализ концептов и вид их взаимного влияния показывает возможность повышения интеллектуальности информационного обеспечения деятельности аналитиков, экспертов, специалистов.

Рассмотрены разновидности инструментов искусственного интеллекта, в том числе результаты гибридизации отдельных подходов, и возможные результаты их применения.

Литература

1. Баргесян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. / А.А. Баргесян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – М.: BHV, 2008. – 384 с.
2. Дюк В.А. Data Mining: учебный курс. / В.А. Дюк, А.П. Самойленко – СПб: Питер, 2001. – 368 с.
3. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. - СПб.: Питер. 2001. - 384 с.

4. Трахтенгерц Э.А. Компьютерная поддержка формирования целей и стратегий. / Э.А. Трахтенгерц. // Сер. «Системы и проблемы управления». – М.: СИНТЕГ, 2005. – 224 с.
5. Палагин А.В., Петренко Н.Г. К вопросу системно-онтологической интеграции знаний предметной области // Математические машины и системы. – 2007. – №3,4. – С.63-75.
6. Палагин А.В. Системная интеграция средств компьютерной техники / А.В. Палагин, Ю.С. Яковлев – Винница: «УНІВЕРСУМ-Вінниця». – 2005. – 680 с.
7. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы. / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия – Телеком, 2007. – 452 с.

Abstract

Sagayda P.I. The Ontological design of problem area the «Intellectual analysis of data for support of making decision». *Ontology of objects and tasks of problem space is worked out the "Intellectual analysis of data for support of making decision", which allows to define direction theoretical discovery representations of knowledge and design of computational processes. The process of making decision is analysed and presented as ontology of concepts from the view point of tasks, which can be decided (prognosis, classification, clusterization, choice), and also from the view point of the choice subject and result of choice. The varieties of instruments of artificial intelligence are considered, including results of separate approaches hybridization, and possible results of their application.*

Keywords: *intellectual analysis of data, support of making decision, problem space, ontology, methods of artificial intelligence*

Анотація

Сагайда П.І. Онтологічне моделювання проблемної області «Інтелектуальний аналіз даних для підтримки прийняття рішень». *Розроблена онтологія об'єктів та задач проблемного простору "Інтелектуальний аналіз даних для підтримки прийняття рішень", яка дозволяє визначити напрям теоретичних досліджень представлення знань і моделювання процесів обробки даних. Процес прийняття рішення проаналізований і представлений у вигляді онтології концептів з точки зору завдань, які можуть бути вирішені (прогноз, класифікація, кластеризація, вибір), а також з точки зору предмета вибору і результату вибору. Розглянуті різновиди інструментів штучного інтелекту, у тому числі результати гібридизації окремих підходів, і можливі результати їх застосування.*

Ключові слова: *інтелектуальний аналіз даних, підтримка прийняття рішень, проблемний простір, онтологія, методи штучного інтелекту.*

Здано в редакцію:
15.04.2010р.

Рекомендовано до друку:
д.т.н, проф. Зорі А.А.