

УДК 004.9:[510.23+510.67]

П.И. Сагайда (канд. техн. наук, доц.)Донецкий национальный технический университет, г. Красноармейск
кафедра электронной техники
E-mail: PaulSagayda@ukr.net**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КОМПЬЮТЕРИЗИРОВАННЫХ
ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ
ДАННЫХ НА ОСНОВЕ ТЕОРИИ КАТЕГОРИЙ**

Разработана категория информационных систем для интеллектуальной обработки данных, выполнено ее представление в виде коммутативной диаграммы, что позволило построить математическую модель обобщенных процессов получения и преобразования данных в таких системах на высоком уровне абстракции. Определены минимально необходимые преобразования данных и моделей, перечень и последовательность которых определяется топологическими свойствами категории на основе объективных критериев в процессе субъективного концептуального моделирования. Для формализации моделей различных видов и на основе различного математического аппарата применены методы теории моделей. Выполнено типовое представление модели, которую необходимо извлечь из данных на основе их интеллектуальной обработки.

Ключевые слова: информационная система, интеллектуальная обработка данных, математическая модель, теория категорий, теория моделей.

Общая постановка проблемы. Организационно-технические системы (ОТС), функционирующие в различных предметных областях и на различных этапах проектирования, создания и распределения изделий и услуг, сформированные в виде производственных участков и технологических линий, конструкторских и планово-финансовых бюро, учебных заведений и подсистем массового обслуживания, – являются сложными системами, состоящими из подсистем как организационного, так и технического (технологического) характера [1 – 3]. Эти подсистемы имеют различную природу функционирования и различные методы взаимодействия и обмена материальными потоками и потоками данных, что затрудняет аналитическое моделирование их работы и оптимизацию для достижения системного эффекта. Кроме того, характерным для таких систем является вхождение в их состав активных элементов, – людей, которые являются интеллектуальными агентами и лицами, принимающими решения по организации и реализации производственной деятельности [2,3].

В настоящее время обеспечение эффективной работы ОТС невозможно без широкого внедрения компьютеризированных информационных систем (КИС), которые организованы на основе баз и хранилищ данных со структурой, близкой к оптимальной. Такие КИС должны обеспечивать не только оперативный учет данных о деятельности ОТС, а и, в большей мере, информационную поддержку плановой, проектной, конструкторской и других видов деятельности, связанных с принятием решений и генерацией управляющих воздействий на элементы ОТС, для достижения требуемого системного эффекта и минимизации затрат ресурсов. Эффективность информационной поддержки должна обеспечиваться за счет предоставления активным элементам ОТС (аналитикам и лицам, принимающим решения) моделей функционирования предметных областей в виде математических зависимостей, правил, деревьев решений, тенденций развития, которые извлекаются из накопленных данных в рамках корпоративных или глобальных информационных сетей.

В настоящее время, в условиях быстрого развития таких направлений интеллектуальной обработки данных (ИОД), как On-Line Analysis Processing, или OLAP (на основе визуализации агрегированных данных), Data Mining и Data Science (на основе методов математической статистики, Machine Learning и искусственного интеллекта) [4], недостаточно развиты теоретические основы создания информационных систем для ИОД. Выбор и использование методов и алгоритмов обработки, исследуемых параметров все еще требуют высокой квалификации привлекаемых аналитиков и инженеров по знаниям, результаты обработки данных не приводят к извлечению зависимостей достаточной специфичности и полезности для агентов ОТС. Накопленные в данной области знания, как теоретического, так и экспериментального характера, требуют соответствующей инженерии, организации в виде, доступном для автоматизации их обработки и модификации в ходе использования. Необходимо разработка принципов организации КИС для ИОД с интегрированными базами знаний и хранилищами данных, создаваемыми и используемыми на основе онтологического подхода [5] и других методов инженерии знаний, что позволит повысить эффективность как применения баз знаний, так и работы КИС с точки зрения точности и достоверности прогнозирования и поддержки принятия решений.

При этом важная научно-техническая проблема заключается в следующем: имеется большое количество разнородных математических и алгоритмических методов извлечения скрытых зависимостей из данных, которые необходимы для прогнозирования и поддержки принятия решений в ОТС; эти методы базируются на слабо совмещаемом математическом и операционном аппарате; особенности и порядок применения этих методов являются частью слабо формализованных знаний экспертов и аналитиков; автоматизация при обработке данных обеспечивается на низком уровне абстракции; практически каждая КИС для ИОД разрабатывается для решения конкретной задачи анализа данных, отсутствует универсальная технология решения широкого круга задач и адаптируемости баз знаний к изменениям условий обработки данных и свойств объектов анализа.

Вместе с тем, оперативное проектирование КИС для ИОД для узкой проблемной области, как и построение более универсальных КИС для решения классов задач по обработке данных, обеспечение высокого качества функционирования таких систем, требуют разработки общих принципов и паттернов проектирования, в особенности для систем, основанных на знаниях. Такая разработка, в свою очередь, должна быть основана на эффективной математической модели процесса функционирования КИС для ИОД, которая должна включать в себя все аспекты – от получения данных на основе измерительных процедур до организации и использования хранилищ данных и процесса извлечения зависимостей (знаний) из данных и проверки их адекватности.

Постановка задач исследования. Целью работы является повышение качества и оперативности проектирования и реализации компьютеризированных информационных систем для интеллектуальной обработки данных, организованных с использованием методов инженерии знаний, за счет разработки математической модели функционирования таких систем.

Для максимального обобщения на данном (концептуальном) этапе моделирования будем использовать в качестве математического аппарата теорию категорий [6, 7] и теорию моделей [8]. Такой подход позволит с единых математических позиций представить данные, знания, процессы и инструменты их обработки и преобразования. Преимуществом подхода теории категорий к моделированию по сравнению с теоретико-множественным подходом, является анализ не объектов, а процессов их преобразований, т.е. сетей взаимосвязей между группами и пространствами [9, 10], что является важным с точки зрения функционирования КИС для ИОД. При этом результаты внутреннего теоретико-множественного описания объектов не теряются, а интерпретируются в новом качестве и с новыми связями [11]. С помощью теории категорий, на основе топологических шаблонов отношений между объектами, докажем допустимость и необходимость преобразований в информационном канале КИС.

Некоторые коммутативные диаграммы теории категорий как топологические шаблоны для концептуального математического моделирования. Рассмотрим некоторые определения, которые далее будут использованы при моделировании. Для категории \mathbf{A} с объектами X и Y объект $X \times Y$ с двумя морфизмами $X \xleftarrow{p_1} X \times Y \xrightarrow{p_2} Y$ называют произведением (*product*), если для любого данного объекта Z и морфизмов $x : Z \rightarrow X$, $y : Z \rightarrow Y$ есть уникальный морфизм $\alpha : Z \rightarrow X \times Y$ такой, что $p_1 \circ \alpha = x$ и $p_2 \circ \alpha = y$, где \circ - композиция морфизмов. В теории категорий существование и уникальность морфизма α доказана, и, соответственно, получаем коммутативную диаграмму, приведенную на рис. 1.

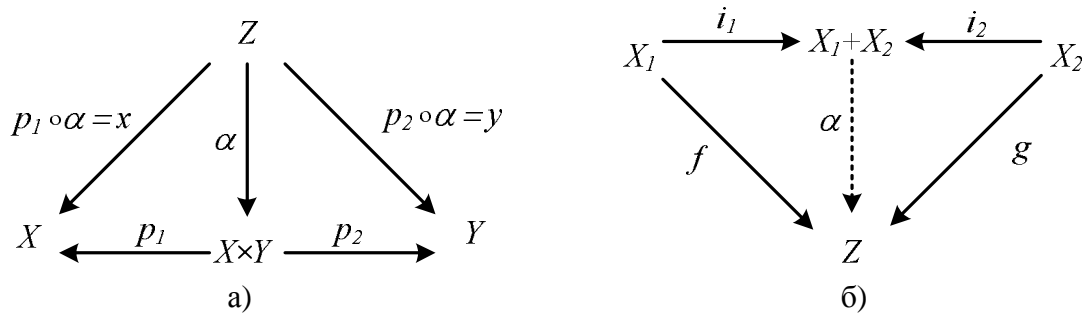


Рисунок 1 – Коммутативная диаграмма для морфизмов *product* (а) и *co-product* (б)

Для категории множеств \mathbf{SET} морфизм α обозначим как набор (x, y) подобъектов категории (т.е. элементов множеств X и Y), т.е. для этой категории *product* соответствует декартову произведению объектов.

Также для категории \mathbf{A} определена сумма $X_1 + X_2$ (*co-product*) двух объектов X_1 и X_2 с двумя морфизмами-инъекциями $X_1 \xrightarrow{i_1} X_1 + X_2 \xleftarrow{i_2} X_2$ такими, что для любого данного объекта Z и двух морфизмов $f : X_1 \rightarrow Z$ и $g : X_2 \rightarrow Z$ есть уникальный морфизм $\alpha : X_1 + X_2 \rightarrow Z$, что дает коммутативную диаграмму, представленную на рис. 1. Существует также уникальный морфизм $\nabla : X + X \rightarrow X$, называемый *co-diagonal*, порождающий коммутативную диаграмму, приведенную на рис. 2.

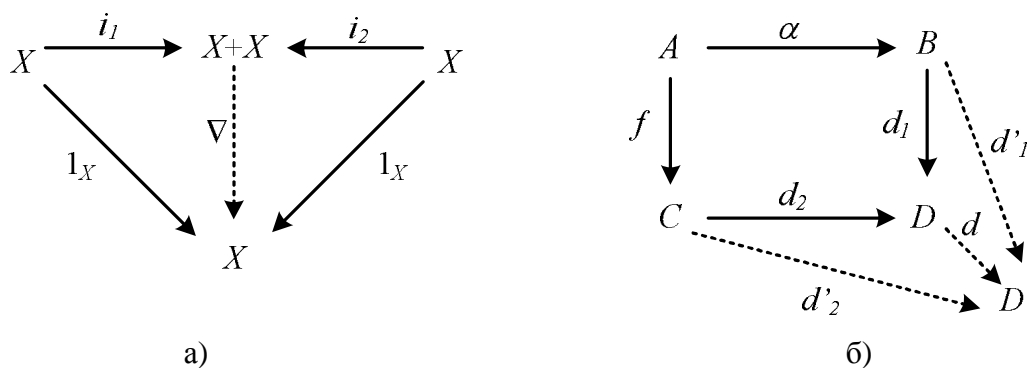


Рисунок 2 – Коммутативная диаграмма для морфизма *co-diagonal* (а) и комплекса *pushout* (б)

Комплекс *pushout*, или *ко-декартов квадрат*, для двух морфизмов $\alpha : A \rightarrow B$ и $f : A \rightarrow C$ есть объект D , и два морфизма $d_1 : B \rightarrow D$ и $d_2 : C \rightarrow D$, таких, что порождают коммутативную диаграмму, приведенную на рисунке 2, и для всех объектов D' с морфизмами $d'_1 : B \rightarrow D'$ и $d'_2 : C \rightarrow D'$ существует уникальный морфизм $d' : D \rightarrow D'$.

Концептуальное моделирование функционирования КИС для ИОД на основе коммутативных диаграмм теории категорий. С учетом необходимости соблюдать вышеуказанные структурные и топологические зависимости, и на основании результатов проведенного ранее системного анализа, было выполнено определение категории **CISforKDD** для предметной области данного исследования, как объектов, так и морфизмов между ними, что позволило определить необходимые преобразования между объектами для достижения целей функционирования КИС для ИОД. Подход в рамках теории категорий позволил абстрагироваться от внутренней структуры отдельных объектов категории **CISforKDD** и физического смысла морфизмов, отображающих объекты друг на друга, и позволил определить минимально необходимые преобразования данных и моделей, перечень и последовательность которых определяется топологическими свойствами категории.

Введем обозначения для объектов категории: X_{area} – «Проблемная область КИС (инициальный объект категории)», X_{method} – «Метод измерения», X_{device} – «Прибор измерения», X_{signal} – «Измерительный сигнал», $X_{message}$ – «Сообщение («сырые» данные)», X_{scale} – «Формат представления (шкала)», X_{result} – «Результат измерения», $X_{warehouse}$ – «Хранилище данных», X_{select} – «Выборка данных», $X_{features}$ – «Набор признаков (столбцов)», X_{order} – «Информативность признаков для теории (порядок)», X_{theory} – «Теория на основе данных», X_{model} – «Модель на основе данных», $X_{negative}$ – «Отрицательная диаграмма», $X_{positive}$ – «Положительная диаграмма», $X_{relevant}$ – «Оценка релевантности модели», $X_{forecast}$ – «Прогноз», $X_{solution}$ – «Решение», $X_{display}$ – «Отображение (терминальный объект категории)».

Для морфизмов, отображающих один объект категории на другой, введем обозначения в виде $\mu_{\beta-\varepsilon}$, где β – первый символ нижнего индекса инициального объекта, участвующего в морфизме, а ε – первый символ нижнего индекса объекта-образа. При этом смысл введенных отображений с точки зрения функционирования КИС для ИОД, там, где они понятны из результатов предыдущего системного анализа, специально оговаривать не будем. Коммутативная диаграмма для категории **CISforKDD**, без обозначения вводимых морфизмов, приведена на рисунке 3. На диаграмме выделены топологические шаблоны с необходимо присутствующими им морфизмами. Полученная модель позволяет пополнять как компонент объектов категории, так и компонент морфизмов категории, расширяя и углубляя понимание происходящих в КИС для ИОД процессов.

В данный момент на этой модели определены следующие свойства морфизмов. В категории **CISforKDD** определены произведения (products), первое из которых описывает процессы измерения и получения сведений о проблемной области работы КИС для ИОД:

$$X_{method} \xleftarrow{\mu_{s-m}} X_{signal} = X_{method} \times X_{device} \xrightarrow{\mu_{s-d}} X_{device}, \langle \mu_{a-m}; \mu_{a-d} \rangle = \mu_{a-s},$$

где морфизм μ_{a-s} является эпиморфизмом и описывает отображение объекта «Проблемная область КИС» на объект «Измерительный сигнал» (сюръекцию, т.к. элементы множества подобъектов проблемной области могут быть представлены различными наборами измеренных сигналов).

Такое отображение $\mu_{a-s} : X_{area} \rightarrow X_{signal}$ в терминах теории измерений можно называть протоколом измерительной процедуры. Процесс получения измерительных сигналов,

параметров в формируемой теории (гипотезе) о взаимозависимости параметров проблемной области.

Приложение теории моделей для описания процесса ИОД. Следующим по порядку обсуждения фрагментом разработанной категории являются ее объекты и морфизмы, определяемые в теории моделей. Теория моделей как часть формальной логической теории имеет важные приложения [8,14]. Например, некоторые части алгебры могут рассматриваться как части теории моделей. Различие между алгебрами и теорией моделей состоит в том, что в теории моделей внимание уделяется также роли языка, в алгебре же языком часто пренебрегают. Использование в данном исследовании теории моделей позволило обобщенно представить любую математическую обработку массива данных в хранилище КИС. Массив частично агрегированных данных в КИС представляет собой решетку подмножеств на упорядоченном по отношению включения множестве, которое для наглядности можно представить многомерным параллелепипедом в признаковом пространстве проблемной области. Обобщение на основе теории моделей включает и интеллектуальную обработку данных, так как она связана с упорядочением цепочек моделей и определением эквивалентности конструируемых теорий и их моделей. Выполним такое обобщение для извлекаемых из данных моделей с помощью понятий *языка, теории, модели, фильтра* [14].

Любую теорию, которая описывает предполагаемые зависимости между признаками (параметрами) проблемной области, можно описать с помощью многосортной алгебры (многосортной предикатной логики), базирующейся на понятии многосортного языка \mathbf{J} , с помощью которого можно описывать поведение сложных отношений между объектами различных типов. В синтаксисе языка эти различные типы объектов представлены переменными различных сортов.

Введем абстрактное множество сортов Ψ . Любая последовательность $\langle t_1, \dots, t_n \rangle$ сортов из Ψ называется типом. Формальный язык \mathbf{J} многосортной логики предикатов состоит из переменных x, \dots различных сортов $i \in \Psi$, констант u, \dots различных сортов $i \in \Psi$, предикатных символов P, Q, \dots различных типов, и функциональных символов f, g, \dots различных типов. Чтобы учитывать, какие сорта переменных могут занимать данное место в формуле, введем тип $\langle i \rangle$, определяемый сортом $i \in \Psi$. Тогда структура \mathbf{D} языка \mathbf{J} многосортной логики предикатов определяется так:

$$\mathbf{D} = \langle \{D_i / i \in \Psi\}, P_D, \dots, f_D, \dots, u_D, \dots \rangle,$$

где D_i – непустые подмножества, отвечающие сортам $i \in \Psi$. В рамках такого представления модели, которые мы должны получить в ходе интеллектуальной обработки данных, являются структурами \mathbf{D} для данного языка \mathbf{J} . Посредством их мы интерпретируем термы так, что каждому терму t присваивается элемент $\mathbf{D}(t) \in D$ и каждой формуле A присваивается значение истинности $\mathbf{D}(A) \in [0,1]$ (в общем случае для нечеткой логики). Свойство компактности, обоснованное в теории моделей, позволяет характеризовать бесконечный объект средствами конечных объектов, что позволяет значительно упростить различные вероятностные методы.

Модель \mathbf{D} является элементарной подмоделью модели \mathbf{D}' , $\mathbf{D} < \mathbf{D}'$ если каждая формула A , выводимая из \mathbf{D}' , также выводима из \mathbf{D} [14]. Расширением модели \mathbf{D} является такая модель \mathbf{D}_D , что:

$$\mathbf{D}_D = \langle \mathbf{D}, \{d / d \in \} \rangle,$$

т.е. мы добавляем в \mathbf{D} константы $\mathbf{d} \in \mathbf{J}(\mathbf{D})$, являющиеся именами для всех элементов D . Диаграммой модели \mathbf{D} называется множество $\Delta_{\mathbf{D}}$ атомарных утверждений B или их отрицаний $\neg B$, $B \in F_{\mathbf{J}(\mathbf{D})}$, верных в \mathbf{D}_D . Подмножество неотрицаемых утверждений называют положительной диаграммой модели, а подмножество отрицаний – отрицательной диаграммой модели. Множество моделей $\{\mathbf{D}_\alpha / \alpha < \xi\}$, где ξ - некоторое ординальное число, называется элементарной цепочкой моделей, если $\mathbf{D}_\alpha \prec \mathbf{D}_\beta$ для любого $\alpha < \beta < \xi$.

Важным методом в теории моделей является конструкция моделей с помощью фильтров [8,14]. Фильтр может быть определен как подмножество решётки, удовлетворяющее определённым условиям. Взятие такого подмножества для множества моделей, которые строятся по результатам обработки данных в КИС для ИОД, отражает происходящие при интеллектуальной обработке данных процессы, связанные с извлечением моделей различных видов и на основе различного математического аппарата из одних и тех же выборок данных и наборов признаков, а также из различных их расширений и подмножеств. Кроме того, возможно выполнение цепочек преобразований получаемых моделей друг в друга (например, в ходе последовательного выполнения факторного анализа данных, корреляционного анализа, регрессионного и т.п.).

Возьмем множество полученных в ходе ИОД моделей $\{\mathbf{D}_i / i \in \mathbf{I}\}$ языка \mathbf{J} , где \mathbf{I} – некоторое индексное множество и $F \in \mathbf{F}$ - фильтр. Новая модель \mathbf{D}_F строится следующим образом. Возьмем декартово произведение носителей D_i из соответствующих моделей \mathbf{D}_i , $i \in \mathbf{I}$:

$$D = \prod_{i \in \mathbf{I}} D_i .$$

Для элементов $d = \langle (d)_i / i \in \mathbf{I} \rangle$, где $(d)_i$ является i -й компонентой $d \in D$, и фильтра $F \in \mathbf{F}$ определено, что: $d \sim d' \Leftrightarrow \{i \in \mathbf{I} / (d)_i = (d')_i\} = F$, где \sim - отношение эквивалентности в D . Таким образом, фильтр, например, обеспечивает группирование кортежей в хранилище данных по эквивалентности значений в отдельных полях. Получаем фактор-множество (т.е. множество всех классов эквивалентности заданного множества) для носителя модели:

$$D_F = \prod_F D_i = D / \sim .$$

Окончательно новая модель \mathbf{D}_F , полученная в результате применения фильтра $F \in \mathbf{F}$, определяется как структура языка \mathbf{J} :

$$\mathbf{D}_F = \prod_F \mathbf{D}_i = \langle D_F, P_{D_F}, \dots, f_{D_F}, \dots, u_{D_F}, \dots \rangle .$$

Типовое представление модели, извлекаемой из массива данных, с использованием теории категорий. Рассмотрим с помощью теории категорий типовое представление модели, которую необходимо извлечь из данных на основе их интеллектуальной обработки. Для этого рассмотрим пространство вещественных чисел \mathbf{R} , на котором в общем случае искомая модель должна быть определена, и соответствующие морфизмы для категории **SET** в виде всех функций из \mathbf{R} в \mathbf{R} [15]. Эта категория обладает *co-product* (суммой) со строгой ассоциативностью. Пусть в общем случае мы строим модель функционирования проблемной области в виде $\text{if } x \leq \alpha \text{ then } y = h_1(x) \text{ else } y = h_2(x)$. Вначале определим следующий морфизм, который позволяет осуществить тестирование значения независимой переменной x :

$$\begin{aligned}
 test_{x>\alpha} : \mathbf{R} &\rightarrow \mathbf{R} + \mathbf{R} \\
 x &\mapsto (x,0) \quad \text{if } x \leq \alpha \\
 x &\mapsto (x,1) \quad \text{if } x > \alpha.
 \end{aligned}$$

Далее проведем последовательные преобразования с учетом изложенных выше определений и обозначений, примененных для рассматриваемой категории:

$$f : \mathbf{R} \rightarrow \mathbf{R}, \quad x \mapsto \begin{cases} h_1(x) & \text{if } x \leq \alpha \\ h_2(x) & \text{if } x > \alpha. \end{cases}$$

$$\mathbf{R} \xrightarrow{test_{x>\alpha}} \mathbf{R} + \mathbf{R}, \quad x \mapsto \begin{cases} (x,0) & \text{if } x \leq \alpha \\ (x,1) & \text{if } x > \alpha. \end{cases}$$

$$\diamond = ([h_1(x)] + [h_2(x)]): \mathbf{R} + \mathbf{R} \longrightarrow \mathbf{R} + \mathbf{R}, \quad \begin{cases} (x,0) & \mapsto (h_1(x),0) \\ (x,1) & \mapsto (h_2(x),1) \end{cases}$$

Для данного преобразования выполняется коммутативная диаграмма, приведенная на рисунке 4 и указывающая на альтернативное выполнение отображения с помощью морфизмов f или g .

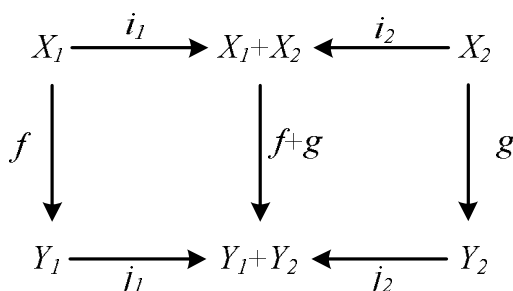


Рисунок 4 – Коммутативная диаграмма для морфизмов $f : X_1 \rightarrow Y_1$, $g : X_2 \rightarrow Y_2$ и уникального морфизма $\diamond = (f + g) : X_1 + X_2 \rightarrow Y_1 + Y_2$

где ∇ – обозначение коммутативной диаграммы морфизмов для комплекса co-diagonal, приведенного на рис. 2.

Аналогичным образом можно представить приложение теории категорий при формализации любой кусочно-определенной функции в рамках модели проблемной области, конструируемой по результатам интеллектуальной обработки данных. Применение в коммутативной диаграмме, приведенной на рисунке 3, комплекса морфизмов pushout концептуально представляет процесс отображения результатов проверки адекватности полученных моделей и прогнозирования с помощью этих моделей для поддержки принятия решений аналитиком, использующим КИС для ИОД. Использованный топологический шаблон представляет отображение результатов ИОД средствами человеко-машинного интерфейса КИС. Таким образом, изложение результатов исследования с применением математического аппарата теории категорий и теории моделей логически завершено.

Выводы.

1. Разработанная категория информационных систем для интеллектуальной обработки данных, ее представление в виде коммутативной диаграммы позволили выполнить модели-

рование обобщенных процессов получения и преобразования данных в компьютеризированных информационных системах на высоком уровне абстракции, в виде наиболее общего представления математической модели.

2. Подход в рамках теории категорий дал возможность абстрагироваться от внутренней структуры отдельных объектов разработанной категории и физического смысла морфизмов, отображающих объекты друг на друга, и позволил определить минимально необходимые преобразования данных и моделей, перечень и последовательность которых определяется топологическими свойствами категории. Это обеспечило введение объективных критериев в процесс субъективного, по сути, концептуального моделирования. Также на основе теории категорий было выполнено типовое представление модели, которую необходимо извлечь из данных на основе их интеллектуальной обработки.

3. Применение теории моделей в ходе моделирования компьютеризированных информационных систем для интеллектуальной обработки данных позволило в общем виде представить процессы, связанные с извлечением моделей различных видов и на основе различного математического аппарата из различных выборок данных и наборов признаков, а также для различных их расширений и подмножеств. Такой подход дал возможность обосновать допустимость выполнения цепочек преобразований получаемых моделей друг в друга, что для конкретных реализаций информационных систем обеспечивает последовательное и параллельное проведение анализа данных с использованием различного инструментария для поддержки принятия решений по управлению предметной областью.

Список использованной литературы

1. Месарович, М. Общая теория систем: математические основы / М. Месарович, Я. Такахара. – М.: Мир, 1978. – 311 с.
2. Лега, Ю.Г. Управління проектом в класі організаційно-технічних систем / Ю.Г. Лега, Т.О. Прокопенко, Ю.І. Урецька // Вісник ЧДТУ, 2014. – № 1. – С. 46 – 50.
3. Интеллектуальные системы управления организационно-техническими системами / [под ред. А. А. Большакова]. – М.: Горячая линия-Телеком, 2006. – 160 с.
4. Сагайда, П.И. Онтологическое моделирование проблемной области «Интеллектуальный анализ данных для поддержки принятия решений» / П.И. Сагайда // Наукові праці ДонНТУ, серія «Обчислювальна техніка та автоматизація», Донецьк: ДонНТУ, 2010. – Вип. 171. – С. 27 – 33.
5. Feilmayr, C. An analysis of ontologies and their success factors for application to business / C. Feilmayr, W. Wöß // Data & Knowledge Engineering. – 2016. – v. 101 – pp. 1 – 23.
6. Цаленко, М.С. Основы теории категорий / М.С. Цаленко, Е.Г. Шульгейфер. – М.: Наука, 1974. – 356 с.
7. Букур, И. Введение в теорию категорий и функторов / И. Букур, А. Деляну. – М.: Мир, 1972. – 259 с.
8. Chang, C.C. Model Theory / C.C. Chang, H.J. Keisler. – Amsterdam: North-Holland Publ. Co., 1973. – 482 p.
9. Левич, А.П. Язык категорий и функторов как архетип количественного и динамического описания Мира / А.П. Левич // Системы и модели: границы интерпретаций. – Томск: Изд-во Томского государственного педагогического университета, 2008. – С. 25 – 33.
10. Кондратьев, Г.В. Возможные применения теории категорий в информационных науках / Г.В. Кондратьев // Математические методы в естественных, технических и социальных науках, 2013. – № 1. – С. 339 – 345.
11. Ковалев С.П. Теоретико-категорные модели и методы проектирования больших информационно-управляющих систем: дис. ... доктора физико-математических наук: 05.13.17 – теоретические основы информатики / Ковалев Сергей Протасович. – Москва, 2013. – 279 с.
12. Adamek, J. Abstract and Concrete Categories. / J. Adamek, H. Herrlich, G. Strecker. – N.Y.:

Wiley and Sons, 1990. – 388 p.

13. Barr, M. *Category Theory for Computing Science* / M. Barr, C. Wells. – London: Prentice Hall, 1990. – 538 p.
14. Новак, В. Математические принципы нечеткой логики / В. Новак, И. Перфильева, И. Мочкорж; пер с англ. под ред. Аверкина А.Н. – М.: ФИЗМАТЛИТ, 2006. – 352 с.
15. Walter, R.F.C. *Categories and Computer Science* / R.F.C. Walter. – Cambridge: Cambridge Univ. Press., 1991. – 166 p.

References

1. Mesarovich, M. and Takahara Ja. (1978), *Obshhaja teorija sistem: matematicheskie osnovy* [General Systems Theory: mathematical foundations], Mir, Moscow, Russia.
2. Lega, Ju.G., Prokopenko T.O. and Urec'ka Ju.I (2014), “Project management in the class of technical and organization systems”, *Visnik ChDTU*, № 1, pp. 46-50.
3. Bol'shakov, A.A. (2006), *Intellektual'nye sistemy upravlenija organizacionno-tehnicheskimi sistemami* [Intelligent control system of organizational and technical systems] / [под ред. А. А. Большакова], Gorjachaja linija-Telekom, Moscow, Russia.
4. Sahaida, P.I. (2010), «Ontological modeling problem area "Knowledge data discovery for decision support"», *Naukovi praci DonNTU, ser. «Obchisljuval'na tehnika ta avtomatizacija»*, № 171, pp. 27 – 33.
5. Feilmayr, C. and Wöß W. (2016), “An analysis of ontologies and their success factors for application to business”, *Data & Knowledge Engineering*, № 101, pp. 1 – 23.
6. Calenko, M.S. and Shul'gejfer E.G. (1974), *Osnovy teorii kategorij* [Fundamentals of category theory], Nauka, Moscow, Russia.
7. Bukur, I. and Deljanu A. (1972), *Vvedenie v teoriju kategorij i funktorov* [Introduction to the theory of categories and functors], Mir, Moscow, Russia.
8. Chang, S.S. and Keisler H.J. (1973), *Model Theory*, North-Holland Publ. Co., Amsterdam, Holland.
9. Levich, A.P. (2008), “Language of categories and functors as the archetype of quantitative and dynamic descriptions of the world” in *Systems and Models: interpretation borders*, Izd-vo Tomskogo gosudarstvennogo pedagogicheskogo universiteta, Tomsk, Russia, pp. 25 – 33.
10. Kondrat'ev, G.V. (2013), “Possible applications of category theory in information sciences”, *Matematicheskie metody v estestvennyh, tehnichestkih i social'nyh naukah*, №.1, pp. 339 – 345.
11. Kovalev, S.P. (2013), A category-theoretic models and methods of designing large information management systems, D. of Sc. Thesis, Foundations of information science, Institute of Control Promlems, Moscow, Russia.
12. Adamek, J., Herrlich H. and Strecker G. (1990), *Abstract and Concrete Categories*, Wiley and Sons, NY, US.
13. Barr, M. and Wells C. (1990), *Category Theory for Computing Science*, Prentice Hall, London, UK.
14. Novak, V., Perfil'eva I. and Mochkorzh I. (2006), *Matematicheskie principy nechetkoj logiki* [Mathematical principles of fuzzy logic], Translated by Averkin, A.N., FIZMATLIT, Moscow, Russia.
15. Walter, R.F.C. (1991), *Categories and Computer Science*, Cambridge Univ. Press, Cambridge, UK.

Поступила в редакцію:
26.04.2016

Рецензент:
д-р техн. наук, проф. Зори А.А.

П.И. Сагайда

ДВНЗ «Донецкий национальный технический университет»

Математичне моделювання комп'ютеризованих інформаційних систем для інтелектуальної обробки даних на основі теорії категорій. Розроблено категорію інформаційних систем для інтелектуальної обробки даних, виконано її подання у вигляді комутативної діаграми, що дозволило побудувати математичну модель узагальнених процесів одержання й перетворення даних у таких системах на високому рівні абстракції. Визначені мінімально необхідні перетворення даних і моделей, перелік і послідовність яких визначається топологічними властивостями категорії на основі об'єктивних критеріїв у процесі суб'єктивного концептуального моделювання. Для формалізації моделей різних видів і на основі різного математичного апарата застосовані методи теорії моделей. Виконано типове подання моделі, яку необхідно витягти з даних на основі їхньої інтелектуальної обробки.

Ключові слова: інформаційна система, інтелектуальна обробка даних, математична модель, теорія категорій, теорія моделей.

P.I. Sahaida

Donetsk National Technical University

Mathematical Modeling of Computerized Information Systems for Knowledge Data Discovery Based on the Category Theory. The aim is to improve the quality and efficiency of the design and implementation of computerized information systems for knowledge data discovery, organized using knowledge engineering methods, through the development of the mathematical functioning model of such systems. The developed category of information systems for knowledge data discovery, its representation in the form of a commutative diagram allowed performing generalized modeling of data reception and conversion processes into computerized information systems at a high level of abstraction, as the most general representation of a mathematical model. The approach in the framework of category theory made it possible to ignore the internal structure of individual objects in the developed category and the physical meaning of morphisms that map objects at each other, and allowed to determine the minimum necessary data conversion and model list and sequence that is determined by the topological properties of the category. This ensured the introduction of objective criteria to subjective process, in fact, conceptual modeling. Also based on category theory typically representation of the model was carried out, which must be drawn from the data on the basis of their intellectual processing. Application of the model theory in the modeling of computerized information systems for knowledge data discovery allowed in general view to represent occurring during knowledge data discovery processes associated with the extraction of models of various types and based on various mathematical apparatus from different data samples, and feature sets, as well as for a variety of their extensions and subsets. This approach made it possible to substantiate the admissibility of the implementation chain of transformations derived models into each other, which ensures consequent and parallel analysis of data using various tools to support decision-making on the subject area management for the concrete implementation of information systems.

Keywords: information system, knowledge data discovery, mathematical model, category theory, model theory.



Сагайда Павел Иванович, Украина, окончил Краматорский индустриальный институт, канд. техн. наук, доцент, докторант кафедры электронной техники. ГВУЗ «Донецкий национальный технический университет» (пл. Шибанкова, 2, г. Красноармейск, 85300, Украина). Основное направление научной деятельности – разработка, моделирование и исследование компьютеризированных информационных систем для интеллектуальной обработки данных.