УДК 004.8
ББК 32.813

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ НА ОСНОВЕ КОЛОНОК ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ

Чесноков А. М.[1]
(*ФГБУН Институт проблем управления им. В.А.Трапезниникова РАН, Москва*)

Рассматриваются интеллектуальные системы на основе колонок в условиях неполной информации, когда по тем или иным причинам на вход системы поступает только часть исходного образа. Приводится решение прямой и обратной задачи. Показана связь между способностью системы работать при неполной информации и прогнозом.

Ключевые слова: искусственный интеллект, интеллектуальные системы на основе колонок, колонка, неполная информация, прогноз.

1. Введение

Интеллектуальные системы на основе колонок представляют собой системы, рассматриваемые в рамках следующей модели (более подробно см. [1, 2]).

Имеется пусть и очень большое, но *конечное множество имен* U, предназначенных для наименования объектов произвольной природы. Не ограничивая общности, считается, что множество имен U является подмножеством множества целых чисел. В множестве имен U выделяются непересекающиеся подмножества, получившие название *областей имен*. В реальных предметных областях причины, приводящие к необходимости выделения областей имен, могут быть совершенно различ-

---

[1] *Александр Михайлович Чесноков, старший научный сотрудник, кандидат технических наук* (*alex-ches@yandex.ru*).

ными. Например, это может быть связано с типизацией. Одной из важнейших причин является необходимость обеспечить отсутствие случайных совпадений имен в различных частях большой системы и тем самым исключить возникновение соответствующих ошибок. Для рассматриваемой модели не важны исходные причины, которые требуют выделения областей имен. Главное, они существуют, и это нашло отражение в модели.

Любое конечное множество имен, принадлежащих тем или иным областям имен, называется *образом*.

Образы любого множества образов P можно перенумеровать, используя для этого имена некоторой области имен U':

$P = \{p_i \mid i \in U'\}$,

где $|U'| = |P|$; $|\cdot|$ – мощность множества.

Упорядоченная пара (i, p_i) получила название *колонки*. Колонка обозначается как $(i \mid p_i)$, где i – имя колонки, p_i – образ, содержащийся в колонке. Также используется обозначение $i \to p_i$. В этом случае говорится, что имя колонки i является *ссылкой* или *указателем* на содержащийся в колонке образ p_i. В свою очередь, про сам образ в колонке p_i будет говориться, что он имеет имя i.

Имя i, которое еще не использовалось для наименования образов, называется *чистым* или *пустым* именем. Его можно представить как колонку, имеющую пустой образ, т.е. колонку вида $(i \mid \varnothing)$ или $i \to \varnothing$.

В образы колонок могут входить имена других колонок, а также чистые имена. Таким образом, можно считать, что в образе одной колонки содержатся имена других колонок, каждое из которых служит указателем на соответствующий образ, возможно, пустой. В результате образуется показанная на рис. 1 сложная структура колонок (для наглядности на рисунке дублируются имена колонок).

Индексом называется любое конечное множество колонок. Состав любого индекса может меняться за счет добавления или удаления колонок. Эти операции называются сложением и вычитанием индексов и обозначаются через + и –.

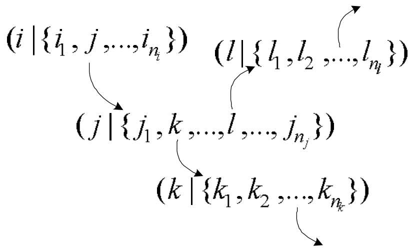

Рис. 1. Структура из колонок

Индекс может быть представлен в виде таблицы, состоящей из вертикальных колонок (столбцов) переменной высоты. В нижней строке таблицы (под чертой) – имена колонок. Над именем каждой колонки перечислены все имена, входящие в образ колонки. По умолчанию считается, что имена колонок и имена в образах принадлежат различным областям имен. В качестве простейшего примера на рис. 2 приведен индекс *A*, состоящий из трех колонок (1 | {1, 3}), (2 | {2, 3, 4}), (3 | {4, 5}).

	A	
	4	
3	3	5
1	2	4
1	2	3

Рис. 2

Интеллектуальная система на основе колонок представляет собой один или несколько индексов, и работающий с ними механизм (машина колонок), который, получая информацию о внешним мире в виде образов, формирует новые колонки, изменяет уже существующие, удаляет ненужные и выполняет другие необходимые операции (рис. 3).

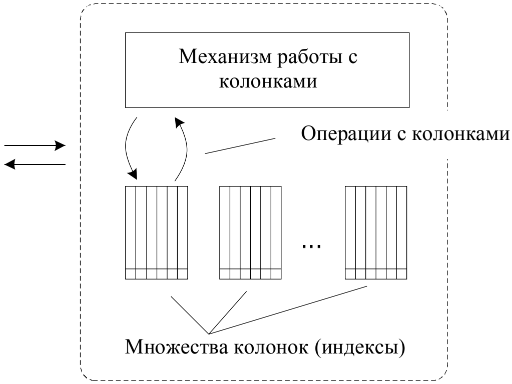

Рис. 3. Система на основе колонок

Знания в рассматриваемых системах представлены с помощью колонок, а в основе процесса накопления знаний лежит запоминание новых образов под определенными именами. При этом *элементарными базовыми задачами*, без которых невозможно функционирование системы, очевидно, являются *прямая задача* – по образу получить его имя, и *обратная задача* – по имени получить соответствующий образ.

В [1] при решении прямой и обратной задачи предполагалось, что все рассматриваемые образы представлены полностью. Однако в реальных условиях системе часто приходится работать при неполной информации, когда по тем или иным причинам рассматриваемые образы представлены лишь частично (рис. 4).

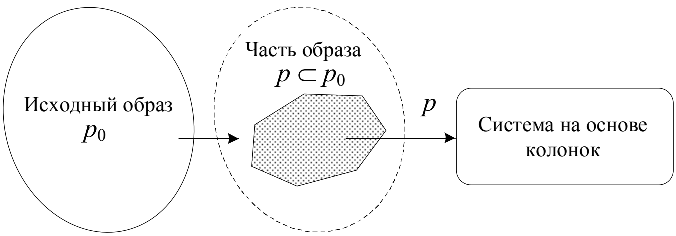

Рис. 4. Часть исходного образа на входе системы

Базовые задачи служат основой решения всех остальных задач [1, 2]. Поэтому для того чтобы система могла работать в условиях реального мира, она должна решать прямую и обратную задачи при неполной информации. Этому и посвящена данная работа.

В следующем разделе показывается существование решения базовых задач в условиях неполноты информации для любых типов образов. Затем для образов в виде конечных неупорядоченных множеств и образов в виде конечных последовательностей или векторов рассматривается метод пересечений. Наконец, в заключение показывается связь между способностью системы работать при неполной информации и прогнозом.

2. Решение прямой и обратной задачи при неполной информации

Рассматриваемые элементарные базовые задачи являются простейшими, при решении которых фрагменты каких-то неизвестных образов не запоминаются, так как в данном случае это не имеет смысла. Все, что требуется, – это по неполному образу p (рис. 4) определить, что из себя может представлять полный образ, т.е. указать имена всех известных системе образов, частью которых является образ p.

В дальнейшем будем полагать, что у системы имеется информация, с помощью которой она может различать полные и неполные образы. Наиболее просто это обеспечивается за счет замены недостающих элементов образа специальными служебными именами. Очевидным признаком того, что образ является полным, является отсутствие в его составе указанных служебных имен. Для образов в виде конечных неупорядоченных множеств число недостающих элементов образа можно заменить числом элементов n_0 исходного образа p_0. Признаком полноты в этом случае будет совпадение мощности входного образа p с полученной системой величиной n_0.

Постановка базовых задач при неполной информации выглядит следующим образом.

Если образ p полный, то решается обычная прямая задача [1]. Система пытается определить имя i образа p. Если ей это удается, то имя i является решением прямой задачи. В противном случае образ p является новым и система его запоминает под некоторым именем, которое и является в данном случае решением прямой задачи.

Если же образ p представляет собой лишь часть исходного образа p_0, то системе необходимо определить имена всех известных ей образов, частью которых является образ p (рис. 5). Если ей это удается, то множество имен таких образов является решением прямой задачи. В противном случае рассматриваемый образ p представляет собой часть какого-то неизвестного системе образа. Как уже говорилось, запоминать его не имеет смысла.

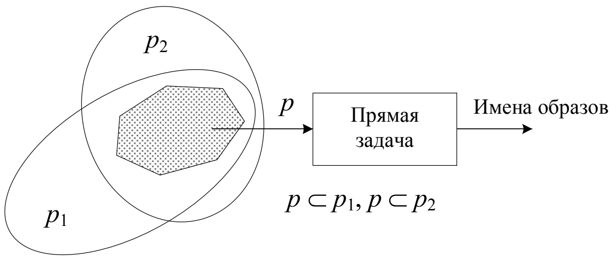

Рис. 5. К постановке прямой задачи

Обратная задача в условиях неполной информации остается без изменений – необходимо по имени образа i получить сам образ.

2.1. МЕТОД РЕШЕНИЯ БАЗОВЫХ ЗАДАЧ НА ОСНОВЕ ПОЭЛЕМЕНТНОГО СРАВНЕНИЯ

Для прямой и обратной задачи существует общий метод решения, применимый к любым типам образов. Это метод, использующий поэлементное сравнение [1].

Известные системе образы хранятся в индексе A, который представляет собой множество колонок вида $(i \mid p_i)$, где p_i – образ, известный под именем i. В исходном состоянии $A = \varnothing$.

Пусть p – произвольный полный образ, для которого надо решить прямую задачу. Он поэлементно сравнивается с образами всех колонок индекса A. Если найден совпадающий образ p_i, то его имя, т.е. имя колонки $(i \mid p_i) \in A$, является именем образа p и решением прямой задачи. В противном случае выбирается некоторое чистое имя i_p из соответствующей области имен и выполняется сложение $A + (i_p \mid p)$, т.е. к индексу A добавляется колонка $(i_p \mid p)$. Если полный образ p снова появится на входе системы, то поэлементное сравнение даст для него имя i_p.

Пусть теперь образ p представляет собой лишь часть некоторого исходного образа p_0. Решением прямой задачи в этом случае будет множество имен $\eta(p)$, состоящее из имен всех известных системе полных образов, частью которых является образ p. Перед началом сравнения устанавливается $\eta(p) = \varnothing$. Поступивший неполный образ p поэлементно сравнивается с образами колонок индекса A. Если все элементы образа p совпадают с соответствующими элементами образа колонки $(i_k \mid p_k) \in A$, то имя i_k добавляется в множество $\eta(p)$. После того как поэлементное сравнение с колонками индекса A закончено, множество $\eta(p)$ представляет собой решение прямой задачи для образа p. Если после всех сравнений окажется, что $\eta(p) = \varnothing$, то образ p – это часть какого-то неизвестного системе образа.

Обратная задача решается точно так же, как и в случае полной информации. Если имеется имя i, для которого необходимо решить обратную задачу, то соответствующий образ равен образу p_i колонки $(i \mid p_i) \in A$. Если колонки с таким именем не существует, то i – чистое имя.

Очевидно, для образов тех или иных типов могут существовать более эффективные методы решения прямой и обратной задачи при неполной информации. В частности, более эффективным является метод пересечений [1], который далее рассматривается для образов в виде неупорядоченных конечных множеств и образов в виде конечных последовательностей или векторов.

2.2. МЕТОД ПЕРЕСЕЧЕНИЙ ДЛЯ ОБРАЗОВ В ВИДЕ КОНЕЧНЫХ НЕУПОРЯДОЧЕННЫХ МНОЖЕСТВ

Будем полагать, что образы представляют собой конечные неупорядоченные множества имен вида $p = \{i_1, \ldots, i_m\}$, $m \geq 1$, где $i_k \in U'$, U' – некоторая область имен. Очевидно, любой такой образ $p \in P$, где P – множество всех подмножеств множества U', исключая пустое.

Кроме того, будем полагать, что для каждого входного образа p известна *истинная мощность* n_0, т.е. мощность, которую имеет полный образ p_0. Так как $p \subset p_0$, то $|p| \leq n_0$, и признаком полноты образа p является равенство $|p| = n_0$.

Для решения прямой и обратной задачи будут использоваться индексы A, B и заданная в виде множества пар функция $n(i)$ [1], которая в данном случае будет содержать мощности известных системе *полных* образов.

В начале работы системы $A = \varnothing$, $B = \varnothing$ и $n(i) = \varnothing$.

Для $\forall p \in P$ обозначим через $\eta(p)$ пересечение образов колонок $(k \mid a_k) \in A$ для всех имен $k \in p$, т.е. $\eta(p) = \bigcap\limits_{k \in p} a_k$. Если в индексе A отсутствует колонка с именем k, то считается, что ее образ $a_k = \varnothing$, т.е. отсутствующая колонка заменяется пустой колонкой $(k \mid \varnothing)$.

Пусть $p \in P$ – некоторый *полный* образ на входе системы, т.е. $|p| = n_0$. Для него базовые задачи решается обычным образом [1].

Если пересечение $\eta(p) = \varnothing$ или если $\eta(p) \neq \varnothing$ и $n(i) \neq |p|$ для $\forall i \in \eta(p)$, то образ p является новым и его необходимо запомнить под некоторым именем. Для этого система выбирает любое чистое имя i из множества $U'' \backslash U_p$, где U'' – область имен для наименования образов, U_p – множество имен всех известных образов. Затем выполняется сложение $A + (p \mid \{i\})$, т.е. к индексу A добавляется $|p|$ колонок $(k \mid \{i\})$ для всех имен $k \in p$. Кроме того, выполняется сложение $B + (i \mid p)$, а в определение функции $n(i)$ добавляется пара $(i, |p|)$. Имя i является решением прямой задачи.

Таким образом, запоминание системой l неизвестных полных образов можно представить в виде цепочки сложений:

$$A = \sum_{k=1}^{l}(p_k \mid \{i_k\}),\ B = \sum_{k=1}^{l}(i_k \mid p_k),\ n(i) = \bigcup_{k=1}^{l}(i_k, \mid p_k \mid),$$

где i_k – имя образа p_k.

Если же для входного образа p пересечение $\eta(p) \neq \varnothing$ и имеется по крайней мере одно имя $i \in \eta(p)$, для которого $n(i) = |p|$, то такое имя является единственным, представляет собой имя образа p и является решением прямой задачи [1].

Пусть теперь $|p| < n_0$, т.е. p – *неполный* образ. Как уже говорилось, в этом случае будет определяться только множество имен известных системе образов, частью которых является входной образ p. Очевидно, для любого непустого подмножества p, входящего в некоторый известный образ, пересечение $\eta(p) \neq \varnothing$. Следовательно, если $\eta(p) = \varnothing$, то на входе образ, являющееся подмножеством неизвестного образа.

Если $\eta(p) \neq \varnothing$ и $n(i) \neq n_0$ для $\forall i \in \eta(p)$, то образ p также является частью неизвестного образа.

Наконец, пусть $\eta(p) \neq \varnothing$ и существует по крайней мере одно имя $i \in \eta(p)$ такое, что $n(i) = n_0$. Обозначим через S_0 подмножество $S_0 = \{i \in \eta(p) \mid n(i) = n_0\}$. Очевидно, множество имен S_0, является решением прямой задачи, так как для $\forall i \in S_0$ образ, известный под этим именем, содержит p в качестве подмножества и имеет мощность, равную мощности n_0 исходного образа p_0.

Обратная задача решается обычным образом [1]. Если входное имя $i \in U_p$, то соответствующий образ равен образу b_i колонки $(i \mid b_i) \in B$. В противном случае имя i – это чистое имя.

ЗАМЕЧАНИЕ. Если в постановке задачи оговаривается наличие у системы лишь признака полноты входного образа (истинная мощность n_0 неизвестна), то задача будет решаться с меньшей определенностью. Для неполного входного образа p множество $S_0 = \eta(p)$, т.е. в него будут включены вообще все известные образы, содержащие образ p как подмножество.

ПРИМЕР. Пусть имеются показанные на рис. 6 индексы A, B и функция $n(i)$:

A

	4	4	
3	3	3	4
1	2	1	2
1	2	3	4

$n(i)$

i	1	2	3	4
n	2	2	3	3

B

		3	4
3	4	2	3
1	2	1	2
1	2	3	4

Рис. 6

Предположим, что на вход системы поступил образ $p = \{1, 3\}$. Пересечение $\eta(p) = \{1, 3\}$. Если $n_0 = 2$, то p – полный образ. Так как $n(1) = |p|$, то p – образ, известный под именем 1. Если $n_0 = 3$, то p – неполный образ. Так как $n(3) = n_0$, то $S_0 = \{3\}$ и p – это часть образа, известного под именем 3.

Пусть теперь на входе появился образ $p = \{2, 3\}$ и $n_0 = 2$, т.е. p – полный образ. Так как $\eta(p) = \{3, 4\}$ и $n(3) = n(4) \neq |\, p \,|$, то это неизвестный новый образ. После запоминания его под именем 5 будем иметь (рис. 7):

A

	5	5	
	4	4	
3	3	3	4
1	2	1	2
1	2	3	4

$n(i)$

i	1	2	3	4	5
n	2	2	3	3	2

B

		3	4	
3	4	2	3	3
1	2	1	2	2
1	2	3	4	5

Рис. 7

Пусть опять на вход поступил образ $p = \{2, 3\}$. Пересечение $\eta(p) = \{3, 4, 5\}$. Если $n_0 = 2$, то p – полный образ, известный под именем 5, так как $n(5) = |p|$. Если $n_0 = 3$, то на входе неполный образ. Так как $n(3) = n(4) = n_0$, то $S_0 = \{3, 4\}$, т.е. образ p представляет собой подмножество, являющееся частью образов по имени 3 и 4. Если же $n_0 = 4$, то $S_0 = \varnothing$, и образ p – часть некоторого неизвестного образа, состоящего из четырех имен.

Наконец, пусть решается обратная задача для имени $i = 3$. Так как $i \in U_p$, то образ p_i равен образу колонки 3 индекса B, т.е. $p_i = \{1, 2, 3\}$.

2.3. МЕТОД ПЕРЕСЕЧЕНИЙ ДЛЯ ОБРАЗОВ В ВИДЕ КОНЕЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ИЛИ ВЕКТОРОВ

Рассмотрим теперь прямую и обратную задачи при неполной информации для образов в виде конечных последовательностей или векторов вида $(i_1, i_2, \ldots, i_m)$. В условиях неполноты информации координаты входного образа p могут быть известны лишь частично и содержать специальное служебное имя i_0 или просто 0, которое интерпретируется как отсутствие данных о соответствующей координате. Например, $(0, i_2, \ldots, i_m)$ или $(0, i_2, \ldots, i_{m-1}, 0)$. В связи с этим будем полагать, что любой входной образ p принадлежит множеству $P = \bigcup_{k=1}^{n} P^k$, где $P^k = U_1^0 \times \ldots \times U_k^0$, $U_k^0 = U_k \bigcup \{0\}$, U_k – область имен k-й координаты. Через $|p|$ будет обозначаться число координат (размерность) образа p.

Прямая задача при неполной информации для образов $p \in P$ формулируются следующим образом. Если образ p не имеет координат, содержащих имя 0, т.е. является полным образом, то прямая задача решаются обычным образом [1]. Если же образ p содержит координаты с именем 0, то необходимо указать имена всех известных системе образов той же размерности, с соответствующими координатами которых совпадают все ненулевые координаты образа p. Обратная задача остается без изменений.

Для решения прямой и обратной задачи будут использоваться индекс $A = \{A_1, A_2, \ldots, A_n\}$, где A_k – индекс для k-й координаты, индекс B для решения обратной задачи и заданная с помощью множества пар функция $m(i)$ [1], которая в данном случае содержит размерности известных системе *полных* образов.

Пусть $p = (i_1, i_2, \ldots, i_m) \in P$ – произвольный образ. Обозначим через $\eta(p)$ пересечение $\eta(p) = \bigcap_{\substack{k=1 \\ i_k \neq 0}}^{m} a_{i_k}$, где i_k – имя, являющееся k-й координатой образа $p = (i_1, i_2, \ldots, i_m)$, a_{i_k} – образ колонки $(i_k \mid a_{i_k}) \in A_k$. Очевидно, если образ p не содержит нулевых координат, то $\eta(p)$ – это обычное покоординатное пересечение, и для полных образов может без всяких изменений использоваться метод пересечений [1].

Итак, пусть появившийся на входе образ p не содержит нулевых координат и является полным.

Если $\eta(p) = \varnothing$ или если $\eta(p) \neq \varnothing$ и $|p| \neq m(i)$ для $\forall i \in \eta(p)$, то образ p – неизвестный полный образ и его необходимо запомнить под некоторым именем. Для этого система выбирает любое чистое имя i из множества $U''\backslash U_p$, где U'' – область имен для наименования образов, U_p – множество имен всех известных образов. Затем выполняется покоординатное сложение:

$$A + (p \mid \{i\}) = \{A_1 + (i_1 \mid \{i\}), A_2 + (i_2 \mid \{i\}),\ \ldots,\ A_m + (i_m \mid \{i\})\}.$$

где i_k – имя, являющееся k-й координатой входного образа $p = (i_1, i_2, \ldots, i_m)$. Кроме того, с индексом B складывается колонка $(i \mid p)$, а в определение функции $m(i)$ добавляется пара $(i, |p|)$. Имя i является решением прямой задачи.

Таким образом, запоминание системой l неизвестных образов можно представить в виде цепочки сложений:

$$A = \sum_{k=1}^{l} (p_k \mid \{i_{p_k}\}),\ B = \sum_{k=1}^{l} (i_{p_k} \mid p_k),\ m(i) = \bigcup_{k=1}^{l} (i_{p_k}, |p_k|),$$

где i_{p_k} – имя образа p_k.

Если $\eta(p) \neq \varnothing$ и существует имя $i \in \eta(p)$, такое, что $|p| = m(i)$, то такое имя является единственным, представляет собой имя образа p и является решением прямой задачи [1].

Рассмотрим теперь случай с неполной информацией, когда у образа p имеются координаты, содержащие имя 0.

Прежде всего необходимо заметить, что для любого непустого подмножества координат любого известного системе образа пересечение $\eta(p) \neq \varnothing$. Поэтому если для *неполного* образа p пересечение $\eta(p) = \varnothing$, то неполный образ p является частью неизвестного образа.

Если $\eta(p) \neq \varnothing$ и для $\forall i \in \eta(p)$ имеет место $|p| \neq m(i)$, то неполный образ p также представляет собой часть неизвестного образа.

Наконец, пусть $\eta(p) \neq \varnothing$ и существует по крайней мере одно имя $i \in \eta(p)$ такое, что $m(i) = |\,p\,|$. Обозначим через S_0 подмножество $S_0 = \{i \in \eta(p) \,|\, m(i) = |\,p\,|\}$. Очевидно, множество имен S_0 является решением прямой задачи, так как для $\forall i \in S_0$ образ, известный под этим именем, имеет ту же размерность, что и образ p, и у них совпадают ненулевые координаты.

Обратная задача решается обычным образом [1]. Если имя $i \in U_p$, то соответствующий образ равен образу b_i колонки $(i \mid b_i) \in B$. В противном случае имя i – это чистое имя.

ПРИМЕР. Пусть для $n = 3$ имеются показанные на рис. 8 индексы A, B и функция $m(i)$:

A_1

		5
4		2
1		3
1	2	3

A_2

5		
2		
3	4	1
1	2	3

A_3

	4	
	3	2
1	2	3

B

	3	2	2	
3	1	1	2	1
1	3	3	1	3
1	2	3	4	5

$m(i)$

i	1	2	3	4	5
m	2	3	3	3	2

Рис. 8

Если на вход пришел полный образ $p = (3, 1)$, то пересечение $\eta(p)$ будет равно пересечению образов колонки 3 индекса A_1

и колонки 1 индекса A_2, т.е. $\eta(p) = \{2, 3, 5\}$, причем $m(5) = 2$. Следовательно, на входе образ, известный под именем 5.

Пусть теперь на входе неполный образ $p = (3, 1, 0)$. Пересечение $\eta(p)$ образов колонки 3 индекса A_1 и колонки 1 индекса A_2 опять будет равно $\{2, 3, 5\}$, причем $m(2) = m(3) = |p|$. Следовательно, $S_0 = \{2, 3\}$, т.е. неполный образ p ненулевыми координатами совпадает с имеющими ту же размерность образами, известными системе под именем 2 и 3. Так как, очевидно, имена $2, 3 \in U_p$, то решая обратную задачу, получим, что указанные образы равны образам колонок 2 и 3 индекса B, т.е. это образы (3, 1, 3) и (3, 1, 2).

3. Заключение

Приведенные в работе методы решения базовых задач представляют собой более общий вариант методов [1], который включает в себя возможность системы работать при неполной информации. Важнейшим следствием из этого является то, что способность системы работать при неполной информации одновременно означает способность системы к прогнозу, так как по имеющейся к настоящему моменту части временной последовательности система может восстановить варианты развития событий в будущем. Причем и имеющаяся к настоящему моменту часть последовательности может быть известна лишь частично, т.е. речь идет о прогнозе при неполной информации. Таким образом, элементарный базовый прогноз является внутренне присущим свойством системы, способной решать базовые задачи в условиях неполной информации. Это демонстрирует одну из основных особенностей модели на основе колонок – универсальность внутренних механизмов, когда один и тот же механизм служит различным целям.

Литература

1. ЧЕСНОКОВ А.М. *Интеллектуальные системы на основе колонок* // Управление большими системами. – 2013. – № 46. – С. 118–146.
2. ЧЕСНОКОВ А.М. *Введение в общую теорию колонок*. – М.: ИПУ РАН, 2012. – 141 с.

COLUMNS-BASED INTELLIGENT SYSTEMS UNDER INCOMPLETE INFORMATION

Alexander Chesnokov, Institute of Control Sciences of RAS, Moscow, Cand. Sc. (alex-ches@yandex.ru).

Abstract: *The paper considers columns-based intelligent systems working under incomplete information, when only a part of the whole input image is received. We provide solutions to both direct and inverse problems. We also reveal the relation between system's ability to work under incomplete information and predicting ability of the system.*

Keywords: artificial intelligence, columns-based intelligent systems, column, incomplete information, prediction.

Статья представлена к публикации членом редакционной коллегии Н.Н. Непейводой

Поступила в редакцию 24.02.2014.
Опубликована 31.07.2014.