А.А. Зуенко, Б.А. Кулик, А.Я. Фридман

Унификация обработки данных и знаний на основе общей теории многоместных отношений

Аннотация. В работе исследуются возможности применения нового математического аппарата – алгебры кортежей (АК), которая реализует общую теорию многоместных отношений. Показано, что все основные структуры данных и знаний могут быть выражены в виде АК-объектов, матричные свойства которых хорошо изучены и пригодны для ускорения интеллектуальных процедур. Описана одна из возможных модификаций АК – алгебра условных кортежей.

Ключевые слова: алгебра кортежей, обработка данных и знаний, матричные свойства отношений.

Введение

В настоящее время при разработке интеллектуальных систем возникают проблемы, вызванные принципиально различными подходами к построению баз данных (БД) и баз знаний (БЗ). В основе проектирования БЗ лежит математическая система, которая имеет несколько названий: формальный подход, аксиоматический метод, символическая логика, теория формальных систем (ТФС). Представление и обработка данных соответствует процедурному (алгебраическому) подходу, который, в настоящее время считается плохо приспособленным для логического анализа.

Различия в подходах к построению БД и БЗ вызывают трудности при их сопряжении в рамках единой программной системы. Проблема сопряжения БД и БЗ была обозначена еще на конференции IJCAI'95 (The International Joint Conference on Artificial Intelligence in Montreal, Canada on August 19 through 25, 1995) и становится все более актуальной ввиду возрастающей необходимости интеллектуализации систем управления БД (семантические интерфейсы БД, дедуктивные БД и т.д.). Поэтому требуется единая методологическая основа для обработки данных и знаний. С точки зрения авторов такое единство может быть достигнуто в рамках алгебраического подхода, если принять в качестве базового понятие «многоместное отношение». Это дает возможность представить многие системы данных и знаний не только в виде некоторого искусственного языка, но и в виде совокупности отношений с различными схемами, над которыми можно производить определенные операции, аналогичные операциям алгебры множеств.

В представленной работе рассматривается алгебра кортежей (АК) [1], которая предназначена для решения комплекса упомянутых проблем [2-4]. Она использует в качестве основной структуры не элементарный кортеж, а декартово произведение множеств и реализует общую теорию многоместных отношений, в рамках которой, в частности, исследуются матричные свойства отношений [5]. В отличие от теории бинарных отношений (графы, семантические сети, онтологии и т.д.) и теории многоместных отношений, в основе которой лежит реляционная алгебра, АК обобщает операции алгебры множеств на случай, когда отношения заданы в различных схемах. Она позволяет формализовать широкий круг логических задач (модифицируемые заключения, моделирование графов и семантических сетей, экспертных правил и т.д.) [2-4, 6, 7]. Здесь мы остановимся на том, каким образом в терминах АК могут быть представлены основные структуры данных и знаний.

Использование "компактного" представления отношений в АК дает возможность применять алгебраический подход не только в системах управления БД, но и для систем БЗ, поскольку во многих случаях позволяет снизить трудоемкость (а иногда и вычислительную сложность) процедур обработки знаний. С точки зрения интеллектуализации БД алгебра кортежей может рассматриваться как расширение реляционной алгебры на задачу обработки знаний.

1. Основы алгебры кортежей

1.1. Основные термины и структуры

В отличие от реляционной алгебры, применяющейся для формализации БД, АК позволяет использовать все средства логического моделирования и анализа систем, которые входят в математическую логику. Свойства АК основаны на свойствах декартова (прямого) произведения множеств и соответствуют основополагающим законам математической логики. В АК при получении промежуточных результатов нет необходимости представлять структуры как множества элементарных кортежей. Все операции выполняются с компонентами атрибутов (множествами) или кортежами компонент.

Определение 1. ***Алгебра кортежей*** – это алгебраическая система, носителем которой является произвольная совокупность многоместных отношений, выраженных в специфических структурах (элементарный кортеж, C-кортеж, C-система, D-кортеж, D-система), называемых ***АК-объектами***. Таким образом, помимо элементарного кортежа в АК определены еще четыре структуры, с помощью которых сжато представляются множества элементарных кортежей.

Имена АК-объектов содержат собственно имя, которое иногда сопровождается заключенной в прямые скобки последовательностью имен атрибутов, формирующих схему отношения, в которой задан этот АК-объект. Например, если введен элементарный кортеж $T[XYZ] = (a, b, c)$, то подразумевается, что T – имя элементарного кортежа (a, b, c); X, Y, Z – имена ***атрибутов***; $[XYZ]$ – ***схема отношения*** (т.е. пространство атрибутов); $a\in X$, $b\in Y$ и $c\in Z$. Множество всех значений атрибута называется ***доменом***. В математической логике доменам атрибутов соответствуют области определения переменных, в информационных системах – шкалы признаков. В дальнейшем атрибуты будем обозначать прописными латинскими буквами, возможно, с индексами, а их значения – строчными латинскими буквами. Множество разных атрибутов, представляющих один и тот же домен, называется ***сортом***. Структуры, заданные в одной и той же схеме отношения, называются ***однотипными***.

В основе АК лежит понятие ***гибкого универсума.*** Если рассматривать пространство признаков $\boldsymbol{S}$ с атрибутами X_i, т.е. $\boldsymbol{S}= X_1\times X_2\times\ldots\times X_n$, то гибкий универсум будет представлять собой всевозможные проекции – подпространства, в которых используются только некоторые атрибуты из множества атрибутов, образующих $\boldsymbol{S}$. Такие подпространства соответствуют ***частным универсумам***. Схема каждого отношения задает определенный частный универсум.

Определение 2. ***Элементарный кортеж*** – это последовательность элементов, каждый из которых принадлежит домену соответствующего атрибута из схемы отношения. Пример элементарного кортежа приведен выше.

Определение 3. ***C-кортежем*** называется заданный в определенной схеме отношения кортеж множеств (***компонент***), каждое из которых является подмножеством домена соответствующего атрибута.

C-кортеж представляет собой множество элементарных кортежей – это множество можно перечислить, если вычислить декартово произведение компонент C-кортежа. Для изображения C-кортежей используются прямые скобки. Например, $R[XYZ] = [A\ B\ C]$ означает, что $A\subseteq X$, $B\subseteq Y$, $C\subseteq Z$ и $R[XYZ] = A\times B\times C$.

Определение 4. C-системой называется множество однотипных C-кортежей, которые записываются в виде матрицы, ограниченной прямыми скобками. В этой матрице строками являются содержащиеся в ней C-кортежи.

C-система представляет собой множество элементарных кортежей, которое равно объединению множеств элементарных кортежей, содержащихся в соответствующих C-кортежах. Например, C-систему

$$Q[XYZ] = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}$$

можно представить как множество элементарных кортежей, вычисляемое по формуле: $Q[XYZ] = (A_1\times B_1\times C_1) \cup (A_2\times B_2\times C_2)$.

Чтобы соединить в единой алгебраической системе, изоморфной алгебре множеств, отношения, заданные в различных проекциях, в АК вводятся ***фиктивные атрибуты***, при формировании которых используются ***фиктивные компоненты***. Они бывают двух видов, одна из них – ***полная компонента*** – используется в C-кортежах и обозначается “*”. Значением этой фиктивной компоненты является домен соответствующего атрибута. Например, C-кортеж $Q[YZ]=[\{f,g\}\ \{a,c\}]$ при условии, что домен атрибута X известен (здесь он равен множеству $\{a, b, c, d\}$), записывается в схеме отношения $[XYZ]$ в виде C-кортежа $[\,*\ \{f,g\}\ \{a,c\}]$. Поскольку фиктивная компонента в Q отвечает атрибуту с доменом X, то справедливо равенство $[\,*\ \{f,g\}\ \{a,c\}] = [\{a,b,c,d\}\ \{f,g\}\ \{a,c\}]$. Другая фиктивная компонента “$\varnothing$” – ***пустое множество*** – используется в D-кортежах, представленных ниже.

С-кортеж, у которого имеется хотя бы одна пустая компонента, пуст. В АК, если речь идет о моделях исчисления высказываний или предикатов, это утверждение принято в качестве аксиомы, для которой имеется соответствующая интерпретация, основанная на свойствах декартовых произведений.

Далее будет показано, что за счет использования фиктивных компонент (фиктивных атрибутов) в АК имеется возможность приводить отношения с различными схемами к одному типу, что позволяет применять к ним операции алгебры множеств. Существенное отличие предлагаемого метода ввода фиктивных атрибутов от известных состоит в том, что новые данные вводятся в многоместные отношения не поэлементно, а как множества, что существенно уменьшает трудоемкость вычислительных алгоритмов и объем памяти для представления структур.

Операции (пересечение, объединение, дополнение) и проверки соотношений включения или равенства между определенными ранее классами АК-объектов осуществляются на основании теорем 1–6. Их формулировки в терминах АК соответствуют известным свойствам декартовых произведений, поэтому доказательства теорем опущены.

Пусть заданы однотипные C-кортежи $P=[P_1\ P_2\ \ldots\ P_n]$ и $Q = [Q_1\ Q_2\ \ldots\ Q_n]$.

Теорема 1. $P \cap Q = [P_1 \cap Q_1\ \ P_2 \cap Q_2\ \ \ldots\ \ P_n \cap Q_n]$.

Примеры: $[\{b, d\}\ \{f, h\}\ \{a, b\}] \cap [\,*\ \ \{f, g\}\ \{a, c\}] = [\{b, d\}\ \{f\}\ \{a\}]$;

$[\{b, d\}\ \{f, h\}\ \{a, b\}] \cap [\,*\ \ \{g\}\ \{a, c\}] = [\{b, d\}\ \varnothing\ \{a\}] = \varnothing$.

Теорема 2. $P \subseteq Q$, если и только если $P_i \subseteq Q_i$ для всех $i = 1, 2, \ldots, n$.

Теорема 3. $P \cup Q \subseteq [P_1 \cup Q_1\ \ P_2 \cup Q_2\ \ \ldots\ \ P_n \cup Q_n]$, причем равенство возможно лишь в двух случаях:

(i) $P \subseteq Q$ или $Q \subseteq P$;

(ii) $P_i = Q_i$ для всех соответствующих пар компонент, за исключением одной пары.

Заметим, что в алгебре кортежей в соответствии с Определением 4 во всех случаях справедливо равенство:

$$P \cup Q = \begin{bmatrix} P_1 & P_2 & \ldots & P_n \\ Q_1 & Q_2 & \ldots & Q_n \end{bmatrix}.$$

Теорема 4. Пересечение двух однотипных C-систем равно C-системе, содержащей все непустые пересечения каждого C-кортежа первой C-системы с каждым C-кортежем второй C-системы.

Пример. Пусть в пространстве $\boldsymbol{S} = X{\times}Y{\times}Z(X = \{a, b, c, d\},\ Y = \{f, g, h\},\ Z=\{a, b, c\})$ заданы C-системы:

$$R_1[XYZ] = \begin{bmatrix} \{a,b,d\} & \{f,h\} & \{b\} \\ \{b,c\} & * & \{a,c\} \end{bmatrix},$$

$$R_2[XYZ] = \begin{bmatrix} \{a,d\} & * & \{b,c\} \\ \{b,d\} & \{f,h\} & \{a,c\} \\ \{b,c\} & \{g\} & \{b\} \end{bmatrix}.$$

Нужно вычислить их пересечение. Оценим количество операций сравнения элементов при пересечении отношений в двух случаях: 1) отношения представлены через элементарные кортежи; 2) отношения выражены в виде C-систем. В первом случае имеем: количество операций = 18×22×3 = 1188, поскольку отношение R_1 содержит 18 элементарных кортежей, а отношение R_2 – 22 кортежа, причем каждый из элементарных кортежей содержит по 3 элемента. Во втором случае пересечение отношений R_1 и R_2 вычисляется как результат пересечения всех пар C-кортежей, содержащихся в этих C-системах:

$[\{a, b, d\}\ \{f, h\}\ \{b\}] \cap [\{a, d\}\ *\ \{b, c\}] = [\{a, d\}\ \{f, h\}\ \{b\}]$;

$[\{a, b, d\}\ \{f, h\}\ \{b\}] \cap [\{b, d\}\ \{f, h\}\ \{a, c\}] = \varnothing$;

$[\{a, b, d\}\ \{f, h\}\ \{b\}] \cap [\{b, c\}\ \{g\}\ \{b\}] = \varnothing$;

$[\{b, c\} * \{a, c\}] \cap [\{a, d\} * \{b, c\}] = \varnothing$;
$[\{b, c\} * \{a, c\}] \cap [\{b, d\}\ \{f, h\}\ \{a, c\}] = [\{b\}\ \{f, h\}\ \{a, c\}]$;
$[\{b, c\} * \{a, c\}] \cap [\{b, c\}\ \{g\}\ \{b\}] = \varnothing$.

Затем из непустых C-кортежей формируем итоговую C-систему:

$$R_1 \cap R_2 = \begin{bmatrix} \{a,d\} & \{f,h\} & \{b\} \\ \{b\} & \{f,h\} & \{a,c\} \end{bmatrix}.$$

В этом случае получим следующую оценку: количество операций = (3×2+2×3+1×2) + (3×2+2×2+1×2) + (3×2+2×1+1×1) + (2×2+3×3+2×2) + (2×2+3×2+2×2) + (2×2+3×1+2×1) = 75. Каждое слагаемое, заключенное в скобки, характеризует пересечение определенной пары C-кортежей и вычисляется как сумма произведений мощностей их соответствующих компонент.

Теорема 5. Объединение двух однотипных C-систем равно C-системе, содержащей все C-кортежи операндов.

В некоторых случаях после вычисления объединения C-систем можно сократить общее число кортежей в полученной C-системе, если использовать условия (i) или (ii) в Теореме 3.

Чтобы выразить алгоритмы вычисления дополнений АК-объектов, введем определение дополнения для отдельно взятой компоненты.

Определение 5. Для любой компоненты P_j АК-объекта ее ***дополнение*** ($\overline{P_j}$) определяется как дополнение относительно домена соответствующего ей атрибута. Например, если задан C-кортеж $R[XYZ] = [A\ B\ C]$, то $\overline{A} = X \backslash A$, $\overline{B} = Y \backslash B$ и $\overline{C} = Z \backslash C$.

Теорема 6. Для произвольного C-кортежа $P = [P_1\ P_2\ \ldots\ P_n]$

$$\overline{P} = \begin{bmatrix} \overline{P_1} & * & \ldots & * \\ * & \overline{P_2} & \ldots & * \\ \ldots & \ldots & \ldots & \ldots \\ * & * & \ldots & \overline{P_n} \end{bmatrix}.$$

В получившейся C-системе $\overline{P}$ размерности $n \times n$ все компоненты, кроме диагональных, – фиктивные. Такие C-системы будем называть ***диагональными***.

Рассмотрим пример. Пусть в пространстве $\boldsymbol{S}$, упомянутом выше, задан C-кортеж $T = [\{b, d\}\ \{f, h\}\ \{a, b\}]$, тогда

$$\overline{T} = \begin{bmatrix} X \setminus \{b,d\} & * & * \\ * & Y \setminus \{f,h\} & * \\ * & * & Z \setminus \{a,b\} \end{bmatrix} = \begin{bmatrix} \{a,c\} & * & * \\ * & \{g\} & * \\ * & * & \{c\} \end{bmatrix}.$$

Диагональные C-системы можно представить как один кортеж множеств, используя для обозначения *перевернутые* прямые скобки. Тогда получается равенство

$$\overline{T} =]\{a, c\}\ \ \{g\}\ \ \{c\}[.$$

Такое «сокращенное» представление диагональной C-системы образует новую структуру АК, которая называется D-кортежем.

Определение 6. ***D-кортеж*** – это кортеж-компонент, ограниченный перевернутыми прямыми скобками, равный диагональной C-системе, у которой диагональные компоненты равны соответствующим компонентам D-кортежа.

Дополнение C-кортежа можно непосредственно записать как D-кортеж. Например, если $T_1 = [\{b, d\} * \{a, b\}]$, то $\overline{T_1} =]\{a, c\}\ \varnothing\ \{c\}[$. В D-кортежах константа “$\varnothing$” – фиктивная компонента.

Такая структура не только позволяет компактно отобразить диагональные C-системы, но самостоятельно используется в некоторых операциях и поисковых запросах. Термины C-кортеж и D-кортеж выбраны не случайно: если представить компоненты этих кортежей предикатами, то C-кортеж соответствует конъюнкции, а D-кортеж – дизъюнкции предикатов.

Используя D-кортежи, можно сформировать еще одну структуру АК – D-систему.

Определение 7. ***D-система*** – структура, состоящая из множества однотипных D-кортежей, равная пересечению множеств элементарных кортежей, содержащихся в D-кортежах.

Изображение D-системы аналогично изображению C-системы, только вместо обычных прямых скобок используются перевернутые.

Соотношения между C-объектами (C-кортежи и C-системы) и D-объектами (D-кортеж и D-системы) соответствуют законам двойственности де Моргана. В силу этого они названы ***альтернативными классами***. Используя преобразования АК-объектов в альтернативные классы, можно обеспечить полноту всех теоретико-множественных операций с АК-объектами и проверок соотношений между ними, не прибегая к представлению АК-объектов как множеств элементарных кортежей.

Как уже упоминалось, в АК можно непосредственно выполнять операции алгебры множеств с АК-объектами, только если они однотипны (т.е. имеют одну и ту же схему отношения). Для того, чтобы производить теоретико-множественные операции над многоместными отношениями, сформированными в различных схемах, их необходимо привести к одной схеме. Для этого в состав операций АК добавлено 5 операций с атрибутами:

1) переименование атрибутов;
2) перестановка атрибутов и соответствующих им столбцов в АК-объектах;
3) обращение АК-объектов;
4) добавление нового фиктивного атрибута (+*Atr*);
5) элиминация атрибута из АК-объекта (–*Atr*).

Далее рассмотрим эти операции.

1.2. Операции с атрибутами, обобщенные операции, операция композиции

Переименование атрибутов допускается только для атрибутов, относящихся к одному сорту. Эта операция используется при замене переменных, в частности, в алгоритмах вычисления транзитивного замыкания графа.

Перестановка атрибутов – операция, при выполнении которой в матрице АК-объекта меняются местами столбцы, а в схеме отношения соответственно изменяется порядок атрибутов.

При выполнении этой операции содержание отношения не изменяется. Она необходима для того, чтобы привести АК-объекты с одинаковыми, но расположенными в разном порядке атрибутами к такому виду, при котором можно выполнять операции алгебры множеств.

Например, *C*-система

$$P[XYZ] = \begin{bmatrix} \{a,b,d\} & \{f,h\} & \{b\} \\ \{b,c\} & * & \{a,c\} \end{bmatrix}$$

при перестановке атрибутов преобразуется в *C*-систему

$$P[YXZ] = \begin{bmatrix} \{f,h\} & \{a,b,d\} & \{b\} \\ * & \{b,c\} & \{a,c\} \end{bmatrix}.$$

Обращение АК-объектов. В случае *бинарных отношений* перестановка столбцов без перестановки атрибутов позволяет получить отношение, *обратное* исходному.

Добавление фиктивного атрибута осуществляется в том случае, если добавляемый атрибут отсутствует в схеме отношения АК-объекта (АК-объекты с дублирующимися атрибутами допустимы, но здесь не рассматриваются). При выполнении этой операции одновременно в схему отношения добавляется имя нового атрибута, а в структуру добавляется на соответствующем месте новый столбец с фиктивными компонентами. При этом в *C*-кортежи и в *C*-системы добавляются фиктивные компоненты “$*$”, а в *D*-кортежи и *D*-системы – фиктивные компоненты “$\varnothing$”.

Элиминация атрибута осуществляется так: из АК-объекта удаляется столбец, а из его схемы отношения – соответствующий атрибут.

Семантика операций “+*Atr*” и “–*Atr*” будет изложена ниже в Разделе 3.

Назовем операции и отношения алгебры множеств с АК-объектами с предварительным добавлением недостающих фиктивных атрибутов ***обобщенными операциями и отношениями*** алгебры множеств в АК и обозначим их соответственно $\cap_G$, $\cup_G$, $\setminus_G$, $\subseteq_G$, $=_G$ и т.д. Первые две операции полностью соответствуют логическим операциям $\wedge$ и $\vee$. Отношение $\subseteq_G$ в АК соответствует ***отношению выводимости*** в исчислении предикатов. Отношение $=_G$ означает, что структуры равны при условии, если они приведены к одной схеме отношения путем добавления атрибутов.

Через перечисленные базовые операции могут быть выражены все стандартные действия над отношениями, в частности ***операция композиции отношений***.

В АК операция композиции существенно упрощается, и ее можно вычислить без попарного сравнения всех элементарных кортежей по формуле:

$$R_1[\boldsymbol{YZ}] \circ R_2[\boldsymbol{XY}] = -\boldsymbol{Y}(R_1 \cap_G R_2) = -\boldsymbol{Y}(+\boldsymbol{X}(R_1) \cap +\boldsymbol{Z}(R_2)), \quad (1)$$

при условии, что ($R_1 \cap_G R_2$) является *C*-кортежем или *C*-системой.

Рассмотрим пример. Пусть в пространстве $\boldsymbol{S}$ заданы АК-объекты:

$$R_1[YZ] = \begin{bmatrix} \{f\} & \{a,b\} \\ \{g,h\} & \{a,c\} \end{bmatrix};$$

$$R_2[XY] = \begin{bmatrix} \{a\} & \{g,h\} \\ \{b,c\} & \{f\} \end{bmatrix}.$$

Сначала вычислим обобщенное пересечение этих отношений:

$$R_1 \cap_G R_2 = \begin{bmatrix} * & \{f\} & \{a,b\} \\ * & \{g,h\} & \{a,c\} \end{bmatrix} \cap \begin{bmatrix} \{a\} & \{g,h\} & * \\ \{b,c\} & \{f\} & * \end{bmatrix} =$$

$$= \begin{bmatrix} \{b,c\} & \{f\} & \{a,b\} \\ \{a\} & \{g,h\} & \{a,c\} \end{bmatrix}.$$

Далее элиминируем атрибут Y, и по формуле (1) имеем:

$$R_1 \circ R_2 = \begin{bmatrix} \{b,c\} & \{a,b\} \\ \{a\} & \{a,c\} \end{bmatrix}.$$

Рассмотрим некоторые примеры того, каким образом известные математические структуры могут быть представлены с помощью АК-объектов и введенных операций.

2. Представление данных и знаний структурами АК

2.1. Графы

Обычно графы и сети представлены в компьютерах как списковые структуры. В системах искусственного интеллекта логический вывод на графах и семантических сетях реализуется как алгоритм поиска достижимых вершин или построения транзитивного замыкания графа. В то же время эти алгоритмы недостаточно эффективны и плохо поддаются распараллеливанию. Рассмотрим, как выражаются графы в структурах АК. В качестве примера используем граф, изображенный на Рис. 1.

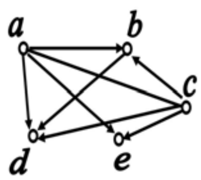

Рис. 1. Пример графа

Этот граф можно записать как C-систему:

$$G[XY] = \begin{bmatrix} \{a\} & \{b,c,d,e\} \\ \{b\} & \{d\} \\ \{c\} & \{a,b,d,e\} \end{bmatrix},$$

которая изоморфна матрице смежности графа.

Транзитивное замыкание можно построить с помощью следующей последовательности операций:

$$G^+ = G \cup G^2 \cup G^3 \cup \ldots \cup G^k, \text{ где } k \le n.$$

Практически во всех случаях преобразования конечного графа G в граф G^+ эта операция заканчивается прежде, чем будет получено последняя «степень» G^k. Критерием более раннего завершения этой операции является отсутствие в очередной «степени» графа не встречавшихся ранее дуг. В качестве примера продемонстрируем, как в АК выполняется построение второй степени графа. Фактически требуется вычислить композицию графа с самим собой, т.е. $G^2[XY] = G[XY] \circ G[YZ] = -Y(G[XY] \cap_G G[YZ])$, где атрибуты X, Y первого операнда переименованы во втором операнде в атрибуты Y, Z.

Сначала производим пересечение:

$$G[XY] \cap_G G[YZ] = \begin{bmatrix} \{a\} & \{b,c,d,e\} & * \\ \{b\} & \{d\} & * \\ \{c\} & \{a,b,d,e\} & * \end{bmatrix} \cap$$

$$\begin{bmatrix} * & \{a\} & \{b,c,d,e\} \\ * & \{b\} & \{d\} \\ * & \{c\} & \{a,b,d,e\} \end{bmatrix} = \begin{bmatrix} \{a\} & \{b\} & \{d\} \\ \{a\} & \{c\} & \{a,b,d,e\} \\ \{c\} & \{a\} & \{b,c,d,e\} \\ \{c\} & \{b\} & \{d\} \end{bmatrix}.$$

Затем элиминируем атрибут Y и после сокращения по Теореме 3 получим:

$$G^2 = -Y(G[XY] \cap_G G[YZ]) = \begin{bmatrix} \{a\} & \{a,b,d,e\} \\ \{c\} & \{b,c,d,e\} \end{bmatrix}.$$

Видно, что в основе этих действий лежит операция обобщенного пересечения, трудоемкость которой, как уже было показано, в АК существенно снижена.

2.2. Семантические сети

Покажем, как в АК осуществляется вывод на семантических сетях [4, 6]. Каждую семантическую сеть можно представить как совокупность бинарных отношений. Правила вывода в семантической сети записываются в виде продукций, в левой части которых содержится некоторый фрагмент семантической сети, а в правой части – фрагмент, который заменяет левую часть при срабатывании правила. Допустим, из наличия в исходной семантической сети (Рис. 2) отношений R_1 и R_2 следует, что между областью отправления отношения R_1 (вершина K) и областью прибытия R_2 (вершина N) должна

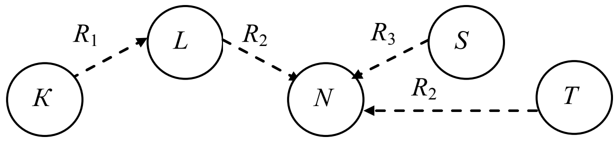

Рис. 2. Исходная семантическая сеть

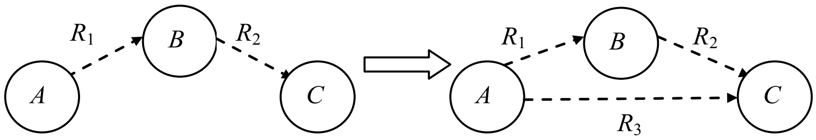

Рис. 3. Пример правила преобразования сети

существовать дополнительная связь R_3. Соответствующее правило изображено на Рис. 3, где A, B, C – переменные, значениями которых могут быть вершины представленной семантической сети.

Эту сеть на языке АК можно записать в виде совокупности C-систем: $R_1[XY] = [\{K\},\{L\}]$, $R_2[YW] = [\{L,T\},\{N\}]$, $R_3[X,W] = [\{S\},\{N\}]$.

Процедура применения правила, показанного на рис. 3, сводится к следующему. Поскольку в правиле производится добавление новой связи без удаления каких-либо других, то необходимо только вычислить $R_1[XY] \circ R_2[YW] = [\{K\},\{L\}] \circ [\{L,T\},\{N\}]=[\{K\},\{N\}]$, а затем добавить полученный кортеж в C-систему, которая сопоставляется отношению R_3:

$R_3[X,W] = R_3[X,W] \cup (R_1[XY] \circ R_2[YW]) = [\{S\},\{N\}] \cup [\{K\},\{N\}] = [\{S,K\},\{N\}]$.

После выполнения преобразований семантическая сеть примет вид:

$R_1[XY] = [\{K\},\{L\}]$, $R_2[YW] = [\{L,T\},\{N\}]$, $R_3[X,W] = [\{S,K\},\{N\}]$.

2.3. Соответствие между АК и исчислением предикатов

В исчислении предикатов C-кортежу в тривиальном случае (когда отдельные атрибуты не соотносятся с многоместными отношениями) соответствует ***конъюнкция*** одноместных предикатов с разными переменными. Например, логической формуле

$$H = P_1(x) \wedge P_2(y) \wedge P_3(z)$$

соответствует C-кортеж $P[XYZ] = [P_1\ P_2\ P_3]$, где $P_1 \subseteq X$; $P_2 \subseteq Y$; $P_3 \subseteq Z$.

Отрицанию формулы H (***дизъюнкции*** одноместных предикатов)

$$\neg H = \neg P_1(x) \vee \neg P_2(y) \vee \neg P_3(z)$$

соответствует D-кортеж $\overline{P} =]\,\overline{P_1}\ \overline{P_2}\ \overline{P_3}\,[$.

Элементарный кортеж, входящий в состав непустого АК-объекта, соответствует ***выполняющей подстановке*** логической формулы.

АК-объект, равный любому частному универсуму, соответствует общезначимой формуле или ***тавтологии***, непустой АК-объект – ***выполнимой формуле***, а пустой АК-объект – ***тождественно ложной формуле***.

Домены атрибутов в АК могут быть любыми, не обязательно равными друг другу множествами. Это означает, что структуры АК изоморфны формулам ***многосортного исчисления предикатов***.

Далее покажем, каким образом представляются кванторы средствами АК (Табл. 1).

Табл.1. Представление кванторов операциями АК

Операции алгебры кортежей	Кванторы исчисления предикатов
$+Y(R)$ для АК-объекта $R[X_1X_2 \ldots X_n]$	$\forall y(R)$
$-X(R)$ для C-системы $R[\ldots X\ldots]$ (*не содержащей «∅» в атрибуте X*)	$\exists x(P)$
$-X(R)$ для D-системы $R[\ldots X\ldots]$ (*не содержащей «*» в атрибуте X*)	$\forall x(P)$

Например, пусть заданы C-система и ее дополнение, выраженное как D-система:

$$Q[XYZ] = \begin{bmatrix} \{a,b,d\} & \{f,h\} & \{b\} \\ \{b,c\} & * & \{a,c\} \end{bmatrix}$$

$$\text{и } \overline{Q}\,[XYZ] = \left]\begin{matrix} \{c\} & \{g\} & \{a,c\} \\ \{a,d\} & \varnothing & \{b\} \end{matrix}\right[.$$

Тогда

$$\exists x(Q[XYZ]) = \begin{bmatrix} \{f,h\} & \{b\} \\ * & \{a,c\} \end{bmatrix}$$

$$\text{и } \forall x(\overline{Q}\,[XYZ]) = \left]\begin{matrix} \{g\} & \{a,c\} \\ \varnothing & \{b\} \end{matrix}\right[.$$

Одной из наиболее важных задач в исчислении предикатов является задача «Выполнимость конъюнктивной нормальной формы (КНФ)». В АК *D*-системы изоморфны КНФ, а задача о выполнимости логической формулы сводится к определению непустоты *D*-системы. Известно, что задача о выполнимости КНФ относится по вычислительной сложности к классу NP-полных, т.е. в общем случае решается с помощью алгоритмов экспоненциальной сложности. Однако имеется немало частных случаев этой задачи, которые решаются за полиномиальное время. Известные частные случаи могут быть интерпретированы и в структурах АК, но здесь появляются дополнительные средства снижения трудоемкости, а в некоторых случаях и вычислительной сложности алгоритмов. Это достигается, в основном, за счет использования матричных свойств АК-объектов, подробно описанных в [5]. В АК выявлены новые структурные и статистические классы КНФ с полиномиально распознаваемым свойством выполнимости. На настоящий момент наиболее исследован алгоритм распознавания выполнимости КНФ в рамках исчисления высказываний. Доказано, что если КНФ выражена как *D*-система, где содержатся только три равномерно распределенные (с вероятностью 1/3) символа: 1, 0 и $\varnothing$, то КНФ разрешима в среднем за полиномиальное время, причем время решения не превышает в среднем полинома 3-й степени от числа конъюнктов.

Подводя итог вышесказанному, особо подчеркнем, что структуры АК полиномиально сводимы к логическим структурам, из чего следует, что ***вычислительная сложность алгоритмов на структурах АК полностью соответствует вычислительной сложности алгоритмов решения типовых задач, заданных на логических структурах.*** Кроме того, в АК за счет "компактного" представления отношений и использования их матричных свойств, появляется возможность дополнительно снизить трудоемкость вычислительных операций при выполнении интеллектуальных процедур.

3. Расширение алгебры кортежей – алгебра условных кортежей

Хотя структуры АК позволяют отображать многоместные предикаты, но компонентами этих структур являются обычные множества, которые соответствуют одноместным предикатам. Если в качестве значений атрибутов в АК-объектах рассматривать многоместные отношения, то зачастую возникает необходимость дополнить базовые определения и операции АК.

Далее описывается одно из возможных расширений АК – алгебра условных кортежей (АУК), позволяющая использовать в качестве значений компонент бинарные отношения и служащая для моделирования логических формул над элементарными двуместными предикатами [8].

Пусть $\boldsymbol{A} = \{A_1, A_2, \ldots, A_n\}$ – множество атрибутов, называемых простыми, а $\boldsymbol{C} = \{A_1, A_2, \ldots, A_n, A_1 \times A_2,\ A_1 \times A_3, \ldots\}$ – множество, состоящее из ***простых атрибутов*** и их всевозможных попарных декартовых произведений (***сложных атрибутов***).

АУК задается следующим образом:

$$\Lambda = <\{R_{Ki}[\boldsymbol{Q}_i]\}, \cup, \cap, +Atr, -Atr, Pr, \leftrightarrow Atr, \quad >,$$

где: $R_{Ki}[\boldsymbol{Q}_i]$ – некоторый АУК-объект (условный *C*-кортеж или условная *C*-система); $\boldsymbol{Q}_i$ – схема отношения, представляющая собой некоторое множество простых и сложных атрибутов; – операция наполнения АУК-объектов, остальные операции аналогичны операциям АК.

Условным С-кортежем называется заданный в определенной схеме отношения кортеж множеств (компонент), каждое из которых является либо подмножеством домена соответствующего простого атрибута, либо подмножеством бинарных отношений, которые могут быть заданы на соответствующем сложном атрибуте. В качестве бинарных отношений (двуместных предикатов) выступают: «<» (меньше), «>» (больше), «=» (равно), «≠» (не равно). Компоненты простых атрибутов формируются из открытых интервалов вида (a, b) и дискретных значений. В описании компонент сложных атрибутов используются лишь имена бинарных отношений, а конкретные множества значений формируются в результате операции наполнения. Условный *С*-кортеж будем называть ***элементарным***, если каждая его компонента либо является полной, либо содержит только одно из возможных значений. Элементарные условные *С*-кортежи соответствуют логическим формулам, содержащим только связки «и» между двуместными предикатами. Условные *С*-системы учитывают связки «или».

Условной С-системой называется множество однотипных условных C-кортежей, которые записываются в виде матрицы, ограниченной прямыми скобками.

Условные C-кортежи и C-системы будем называть ***АУК-объектами***. Чтобы отличать АУК-объекты от АК-объектов, в их именах явно указывается множество сложных атрибутов. Например, $R_{\boldsymbol{K}}[\boldsymbol{M}]$ означает, что $\boldsymbol{K}$ – множество сложных атрибутов.

Операции ***добавления***, ***элиминации***, ***проекции***, ***перестановки*** атрибутов, а также операции ***пересечения*** и ***объединения*** условных C-кортежей и условных C-систем аналогичны соответствующим операциям АК.

Остановимся подробнее на том, как осуществляется пересечение двух условных C-кортежей. Пересечение компонент простых атрибутов сводится к пересечению между собой интервалов и дискретных значений. Пересечение компонент сложных атрибутов регламентируется, исходя из свойств отношений «<», «>», «=», «≠»[8].

Проверки, производимые при пересечении, касаются в отдельности простых и в отдельности сложных атрибутов, они не учитывают того, как соотносятся между собой разные виды атрибутов. Кроме того, не анализируется сама возможность одновременно сформировать все бинарные отношения, описанные в элементарном кортеже, на основе значений компонент простых атрибутов. Для анализа таких соотношений предназначена операция ***наполнения*** АУК-объектов значениями.

Введем на C-кортежах вида $S[X,Y]=[A, B]$ следующие операторы: операторы взятия диагонали (Δ), взятия элементов, лежащих над ($\overline{\Delta}$) и под диагональю ($\underline{\Delta}$). Оператор $\Delta([A, B])$ возвращает $\Delta([A\cap B, A\cap B])$ и соответствует отношению «равно», т.е. Δ находит диагональ множества $A\cap B$ (подмножество диагонали универсума $X\times Y$). Операторы $\overline{\Delta}$ и $\underline{\Delta}$ (сопоставляются отношениям «меньше» и «больше») находят кортежи, которые принадлежат прямоугольнику $A\times B$ и лежат соответственно над и под диагональю универсума $X\times Y$. Отношение «не равно» может быть выражено либо как дополнение отношения «равно», либо как объединение отношений «меньше» и «больше». Операторы возвращают пустое множество, если их область определения не содержит необходимых элементов.

Экземпляром $R_{\boldsymbol{K}}[\boldsymbol{M}]$ АУК-объекта $R_{\boldsymbol{K}}[\boldsymbol{M}]$ называется C-система, которая получается в результате конкретизации всех бинарных отношений, производимой в соответствии со следующим алгоритмом.

<u>Алгоритм.</u> Наполнение условного
С-кортежа значениями.

Пусть имеется условный C-кортеж $R_{\boldsymbol{K}}[\boldsymbol{F}]$.

1. Раскладываем $R_{\boldsymbol{K}}[\boldsymbol{F}]$ в набор элементарных кортежей – $<R_1,\ldots, R_n>$.

2. Формируем множества атрибутов по следующему принципу: простые атрибуты X и Y включаются в одно множество $\boldsymbol{L}_i$, если они вместе образуют некоторый сложный атрибут из $\boldsymbol{K}$.

3. Выбираем очередной элементарный кортеж R_i из тех, что ранее не рассматривались. Если таких кортежей нет, переходим на шаг 9.

4. Выбираем очередное множество $\boldsymbol{L}_i$ из полученных на шаге 2 множеств атрибутов. Если нет ни одного множества, которое еще не анализировалось для кортежа R_i, то переходим к шагу 7.

5. Формируем упорядоченную по возрастанию последовательность, элементы которой берутся из компонент кортежа R_i, соответствующих атрибутам выбранного множества $\boldsymbol{L}_i$, и представляют собой либо дискретные значения, либо концы интервалов. На основе последовательности получаем множество атомов – непересекающихся интервалов и дискретных значений.

6. Компоненты кортежа R_i, которые соответствуют атрибутам из $\boldsymbol{L}_i$, выражаем через атомы. Переходим к шагу 4.

7. Полученный в результате предыдущих шагов (4-6) из R_i условный C-кортеж раскладываем в элементарные. Получаем некоторую условную C-систему.

8. Каждый предикат «=», «<», «>», «≠» в описании сложных компонент XY выражаем через операторы $\Delta([A, B])$, $\overline{\Delta}([A, B])$, и $\underline{\Delta}([A, B])$, где $A\subseteq X$, $B\subseteq Y$. Вычисляем эти выражения. Переходим к шагу 3.

9. Конец алгоритма.

Поясним алгоритм на примере. Пусть имеется логическая формула:

$$(x>3)\wedge(x<4)\wedge(y>2)\wedge(y<7)\wedge(z>5)\wedge(z<6)\wedge(y<x)\wedge(y>z) \quad (2)$$

и требуется установить выполнимость этой формулы на основе анализа ее структуры. В

АУК формуле (2) соответствует условный C-кортеж:

$$P_{XY,YZ}[X,Y,Z,YX,YZ] = \left[\{(3,4)\}, \{(2,7)\}, \{(5,6)\}, \{<\}, \{>\}\right],$$

а решение задачи выполнимости заключается в наполнении заданного кортежа значениями.

Покажем, как пустота условного C-кортежа $P_{XY,YZ}[X,Y,Z,YX,YZ]$ устанавливается с помощью представленного выше алгоритма.

Первый шаг алгоритма в этом случае можно пропустить, поскольку кортеж изначально является элементарным. Приступим к выполнению второго шага. В нашем примере можно сформировать только одно множество простых атрибутов – $\{X,Y,Z\}$, из которых образуются сложные атрибуты YX, YZ. Далее следуют шаги 3-7. Для кортежа $P_{XY,YZ}[X,Y,Z,YX,YZ]$ и множества атрибутов $\{X,Y,Z\}$ последовательность чисел будет выглядеть так: 2, 3, 4, 5, 6, 7. Следовательно, атомами являются: (2,3), (3,4), (4,5), (5,6), (6,7). $P_{XY,YZ}[X,Y,Z,YX,YZ]$ можно записать в виде:

$$\left[\{(3,4)\}, \{(2,3),(3,4),(4,5),(5,6),(6,7)\}, \{(5,6)\}, \{<\}, \{>\}\right].$$

Полученный условный C-кортеж раскладываем в элементарные кортежи.

На шаге 8 выполняем конкретизацию бинарных отношений условной C-системы $P_{XY,YZ}[X,Y,Z,YX,YZ]$, т.е. определяем, не являются ли пустыми результаты вычисления операторов $\Delta([A, B])$, $\overline{\Delta}([A, B])$, и $\underline{\Delta}([A, B])$. В итоге имеем:

$$P_{XY,YZ} = \begin{bmatrix} \{(3,4)\}, \{(2,3)\}, \{(5,6)\}, [\{(2,3)\}, \{(3,4)\}], \varnothing \\ \{(3,4)\}, \{(3,4)\}, \{(5,6)\}, \overline{\Delta}[\{(3,4)\}, \{(3,4)\}], \varnothing \\ \{(3,4)\}, \{(4,5)\}, \{(5,6)\}, \varnothing, \varnothing \\ \{(3,4)\}, \{(5,6)\}, \{(5,6)\}, \varnothing, \underline{\Delta}[\{(5,6)\}, \{(5,6)\}] \\ \{(3,4)\}, \{(6,7)\}, \{(5,6)\}, \varnothing, \{[(6,7)\}, \{(5,6)\}] \end{bmatrix} = \varnothing .$$

В результате операции наполнения выяснилось, что бинарные отношения, образующие компоненты сложных атрибутов исходного C-кортежа, не согласуются между собой по областям их определения.

Таким образом, с помощью АУК появляется возможность автоматизировать анализ составляющих логической формулы, содержащей элементарные двуместные предикаты, и делать выводы относительно ее выполнимости или невыполнимости (как в нашем примере).

Заключение

Для унификации обработки данных и знаний в статье предлагается использовать алгебраический подход на основе общей теории многоместных отношений. В качестве математической основы такого подхода принимается алгебра кортежей, которая с точки зрения универсальной алгебры является булевой структурой. За счет предложенных обобщенных операций и отношений аналитические возможности и области применения объектов алгебры кортежей существенно расширены по сравнению с математическими структурами, применяющимися в настоящее время при моделировании и анализе отношений, в частности, в теории бинарных отношений и реляционной алгебре.

Как показали исследования, применение методов алгебры кортежей к стандартным задачам логического анализа способствует ускорению процедур их решения, поскольку при выводе учитывается внутренняя структура обрабатываемых знаний, а не только осуществимость отдельных подстановок. Алгебра кортежей позволяет распараллеливать алгоритмы логического вывода, т.е. производить обработку знаний аналогично обработке табличных данных в реляционных СУБД. Изложенное обосновывает целесообразность применения алгебраического подхода для унификации процедур управления данными и обработки знаний в интеллектуальных системах.

Литература

1. Кулик, Б.А. Обобщенный подход к моделированию и анализу интеллектуальных систем на основе алгебры кортежей // Труды VI Международной конференции «Идентификация систем и задачи управления» SICPRO'07 (Москва, 29 января – 1 февраля 2007 г.). - С. 679-715.
2. Кулик, Б.А. Моделирование рассуждений на основе законов алгебры множеств // Труды Пятой национальной конференции по искусственному интеллекту, Казань, 7 - 11 октября 1996 г. Т. 1, -С. 58-61.
3. Кулик, Б.А. Вероятностная логика на основе алгебры кортежей // Известия РАН. Теория и системы управления. -2007. -№ 1. -С. 118-127.
4. Зуенко, А.А. Анализ корректности запросов к базам данных систем концептуального моделирования средствами алгебры кортежей / А.А. Зуенко, Б.А. Кулик, А.Я. Фридман // Искусственный интеллект. Интеллектуальные системы (ИИ-2009). Материалы X Международной научно-технической конференции. - Таганрог: Изд-во ТТИ ЮФУ, 2009. - С.86-88.

5. Кулик, Б.А. Новые классы КНФ с полиномиально распознаваемым свойством выполнимости // Автоматика и телемеханика. -1995. -№ 2. -С. 111-124.
6. B. Kulik, A. Fridman, A. Zuenko, Logical Analysis of Intelligence Systems by Algebraic Method, in: Sybernetics and Systems 2010: Proceedings of Twentieth European Meeting on Cybernetics and Systems Research (EMCSR 2010), 2010, Vienna, Austria, pp.198-203.
7. Кулик, Б.А. Управление логико-семантическим анализом на основе теории отношений / Б.А. Кулик, А.А. Зуенко, А.Я. Фридман // VIII Всероссийская школа-семинар «Прикладные проблемы управления макросистемами» (Апатиты, 29 марта – 2 апреля 2010 г.). Материалы докладов. – Апатиты: изд-во КНЦ РАН, 2008. - С. 23-24.
8. Зуенко, А.А. Развитие алгебры кортежей для логического анализа баз данных с использованием двуместных предикатов / А.А. Зуенко, А.Я. Фридман // Известия РАН. Теория и системы управления. -2009. -№2. -С.95-103.

Зуенко Александр Анатольевич. Научный сотрудник Института информатики и математического моделирования технологических процессов Кольского научного центра РАН. Преподаватель в Кольском филиале Петрозаводского государственного университета. Окончил Петрозаводский государственный университет в 2005 году. В 2009 году защитил кандидатскую диссертацию. Имеет 25 печатных работ. Области научных интересов: системы моделирования на основе открытой концептуальной модели предметной области, представление и обработка знаний.
E-mail: zuenko@iimm.kolasc.net.ru.

Кулик Борис Александрович. Работает в ИПМаш РАН. Преподаватель в Университете культуры и искусств. Окончил Ленинградский горный институт в 1963 году. Доктор физико-математических наук (2008 г.). Имеет 70 научных публикаций, в том числе 4 книги. Область научных интересов: исследования в области логики, математики и искусственного интеллекта. E-mail: ba-kulik@yandex.ru.

Фридман Александр Яковлевич. Заведующий лабораторией Института Информатики и Математического Моделирования Технологических Процессов КолНЦ РАН, профессор Кольского филиала Петрозаводского государственного университета. Окончил Ленинградский электротехнический институт в 1975 году. Доктор технических наук (2001 г.). Имеет 185 публикаций, включая одну монографию и 16 свидетельств об изобретении. Области научных интересов: моделирование комплексных технологий и их воздействия на окружающую среду, прикладные интеллектуализированные системы. E-mail: fridman@iimm.kolasc.net.ru.