

Обробка інформації в складних організаційних системах

УДК 519.7

Кудхаир Абед Тамер

Харьковский национальный университет радиоэлектроники, Харьков

РАЗРАБОТКА ПРЕДИКАТНОЙ МОДЕЛИ ПРОТОТИПА СТРУКТУРИРОВАННОГО ОБЪЕКТА В БАЗЕ ЗНАНИЙ

Предложен подход к построению структурного описания сложных объектов на основе гибкого предикатного представления знаний. Основой подхода является формализация понятий предметной области и логических взаимосвязей между ними, что дает возможность наращивания этих знаний по мере появления дополнительной информации с учетом логических взаимосвязей как между элементами, так и между их свойствами.

Ключевые слова: алгебра конечных предикатов, предикат, структурированный объект, прототип, представление знаний.

Введение

Для эффективного использования систем сложных, основанных на знаниях процессов обработки информации необходимо структурировать обрабатываемую информацию и преобразовать ее в знания. Структурированное описание сложных объектов, отражающее взаимосвязи между их элементами, широко применяется в области интерпретации изображений, распознавания паттернов, обработке коллекций документов и иных сферах применения систем искусственного интеллекта [1].

Для представления знаний об объектах со сложной структурой используются графовые, а также предикатные модели. Они позволяют отобразить структурные элементы сложных объектов, связи между ними, а также свойства этих элементов [2]. Основное преимущество такого структурированного описания заключается в возможности создания максимально обобщающих прототипов в форме набора предикатов, которые однозначно интерпретируются подсистемой вывода интеллектуальной системы и позволяют решать задачи поиска, распознавания, а также data mining.

Существующие подходы к построению структурного описания сложных объектов основаны преимущественно на использовании обобщенных нагруженных графов отношений, которые не учитывают логические взаимосвязи между атрибутами составляющих таких объектов. Поэтому весьма актуальной является задача построения гибкого предикатного представления знаний о таких объектах, основанного на формализации понятий предметной области и логических взаимосвязей между ними, что дает возможность наращивания этих знаний по

мере появления дополнительной информации с учетом логических взаимосвязей как между элементами, так и между их свойствами.

Основная часть

Структурируемые объекты характеризуются следующими особенностями [3]:

- состоят из набора взаимосвязанных элементов, часто такого же типа, как и исходный объект;
- обладают сложной внутренней структурой;
- структура объекта отражает его внешние свойства, проявляющиеся в данной предметной области.

Первая особенность позволяет определить объект и часть его свойств через определения его элементов. Иными словами, первоначально мы определяем понятия для составляющих сложного объекта через их свойства. Затем мы получаем возможность определить свойства целого объекта на основе логических соотношений между свойствами его составляющих. Вторая особенность сложного структурируемого объекта связана с тем, что его структура влияет на свойства объекта. Изменение структуры приводит к изменению как внутренних, так и внешних характеристик объекта. Внутренние характеристики определяют порядок сочетания свойств элементов, составляющих сложный объект. Внешние характеристики проявляются в виде свойств в предметной области, формируя третью особенность таких объектов.

При построении моделей таких объектов обычно используют два основных подхода:

- формализация наличия (или отсутствия) тех или иных составляющих элементов объекта в форме дескриптора объекта [4];

– представление объекта в формате набора составляющих его элементов и взаимосвязей между этими элементами [5].

Недостатки первого подхода состоят в том, что дескриптивное описание не всегда позволяет достаточно точно описать структурированный объект без использования специфических знаний о предметной области. В результате мы получаем не общий, специализированный подход, требующий дополнительного преобразования модели объекта, что и затрудняет его практическое применение.

В отличие от первого подхода, второй подход позволяет описать объект с требуемой степенью точности без использования специализированных знаний. Недостаток данного подхода связан с тем, что необходимо представить набор элементов, между которыми существуют множественные связи и которые обладают дополнительными атрибутами.

Для преодоления этого недостатка необходимо представить унифицированное описание, которое позволяло бы единым образом представлять и структуру объекта и свойства его элементов.

Прототип структурируемого объекта в общем случае должен содержать две ключевых составляющих:

- набор элементов объекта, выраженный через их свойства;
- формализованную структуру объекта (возможно, многомерную).

При этом структура в общем случае может быть многомерной. Например, одним из измерений многомерной структуры может быть временное, что позволит объединить в одном прототипе различные модификации объекта, происходящие с течением времени.

Изложенное выше свидетельствует о важности объединения различных состояний одного класса структурированных объектов в одном прототипе – т.е. о построении многомерной модели, в частности формализации процесса изменения структурированного объекта.

Поэтому обобщенный граф G_i^* должен содержать информацию, которая обеспечила бы распознавание всех вариантов объекта во времени.

Традиционно выделение конкретных признаков обеспечивается за счет свойства специализации графа G_i^* .

Покажем, что специализация не позволяет рассматривать структурированный объект как процесс. Действительно, при традиционном подходе в графе G_i^* мы объединяем все свойства структурированного объекта. Специализированный граф G_i^{**} содержит подмножества элементов графа G_i^* :

$$\forall G_i^{**} \exists G_i^* : V_i^{**} \subseteq V_i^*, E_i^{**} \subseteq E_i^*, \\ A_{V_i}^{**} \subseteq A_{V_i}^*, A_{E_i}^{**} \subseteq A_{E_i}^* \quad (1)$$

где V_i^{**}, V_i^* – множества вершин графов G_i^{**} и G_i^* соответственно;

E_i^{**}, E_i^* – множества дуг графов G_i^{**} и G_i^* соответственно;

$A_{V_i}^{**}, A_{V_i}^*, A_{E_i}^{**}, A_{E_i}^*$ – наборы атрибутов для вершин и ребер графов G_i^{**} и G_i^* соответственно.

Объединив специализированные графы G_i^{**} , мы, на первый взгляд, получаем возможность описать различные варианты одного и того же объекта. Однако в данной постановке традиционные подходы к представлению структурированных объектов, в частности обобщенный нагруженный граф отношений имеют ограниченные возможности:

- граф G_i^* обладает классической «плоской» структурой, т.е. состоит из вершин, связывающих их дуг, а также атрибутов вершин и дуг.
- для атрибутов графа доступны только две теоретико-множественные операции:
- объединения подмножеств значений атрибутов;
- объединения типов атрибутов.

Первое свойство графа G_i^* не позволяет изменить количество используемых при описании объекта вершин и дуг в зависимости от изменения объекта во времени.

Проблему пытаются решить через наборы атрибутов, обобщая свойства вершин либо дуг. В предельном случае используется атрибут, обобщающий все возможные свойства текущего элемента графа.

Изменение структурных взаимосвязей целесообразно формализовать на основе объединения в прототипе возможных изменений структуры графа, а атрибуты использовать для описания свойств вершин и дуг.

Второе свойство графа G_i^* не позволяет отразить различные отношения между атрибутами. Граф G_i^* лишь объединяет свойства атрибутов. Однако между атрибутами могут быть сложные логические зависимости.

Обобщая изложенные положения, можем утверждать, что для описания структурированных объектов необходимо выполнить следующие действия:

- описать элементы объекта;
- формализовать взаимосвязи между элементами;
- задать атрибуты элементов и взаимосвязей;
- описать логические взаимосвязи между атрибутами.

Для получения унифицированного описания целесообразно использовать предложенные предикатные формализмы понятий.

Элементы объекта, а также атрибуты описываются на основе определенных ранее понятий (или наборов понятий). Понятийное представление структурированных объектов важно при решении рассмотренной ранее проблематики формализации изменений объекта во времени. Вне зависимости от их свойств, в основе формализации объектов лежит ответ на вопрос: какие понятия позволяют описать эти объекты? Взаимосвязи между понятиями должны отражать взаимосвязи между элементами объекта. В противном случае введенные понятия не будут соответствовать предметной области, для которой формируется база знаний.

В работе предлагается для построения единого формального описания структуры графа G_i^* и свойств его вершин и дуг при построении прототипов M_i структурированных объектов использовать предикатную модель понятий. При таком подходе граф G_i^* представляется системой взаимосвязанных понятий. Наборы свойств вершин и дуг графа G_i^* формализуются в виде предикатов понятий:

$$C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), \quad (2)$$

$$x_{i1}^p \in A_{i1}^{p*}, x_{i2}^p \in A_{i2}^{p*}, \dots, x_{in}^p \in A_{in}^{p*}, p = \overline{1, k}.$$

где $C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p)$ - в общем случае многоместный предикат понятия C_i^p (от англ. «concept»), характеризующего p свойств $x_{i1}^p, x_{i2}^p, \dots, x_{in}^p$ p -го понятия, связанные с p -й вершиной V_i^{p*} графа G_i^* . Переменные $x_{i1}^p, x_{i2}^p, \dots, x_{in}^p$ предиката C_i^p определены на множествах $A_{i1}^{p*}, A_{i2}^{p*}, \dots, A_{in}^{p*}$, содержащих соответственно свойства понятия C_i^p , графически представленные вершиной V_i^{p*} .

Такая структура вершин V_i^{p*} предполагает иерархическое построение графа G_i^* , при котором каждая вершина может представлять собой подграф графа G_i^* . Дуги графа G_i^* формализуются в виде бинарных предикатов вида:

$$B_i^{pr*}(x_i^p, x_i^r), x_i^p = (x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), \quad (3)$$

$$x_i^r = (x_{i1}^r, x_{i2}^r, \dots, x_{im}^r),$$

$$x_{i1}^p \in A_{i1}^{p*}, x_{i2}^p \in A_{i2}^{p*}, \dots, x_{in}^p \in A_{in}^{p*},$$

$$x_{i1}^r \in A_{i1}^{r*}, x_{i2}^r \in A_{i2}^{r*}, \dots, x_{im}^r \in A_{im}^{r*},$$

где каждый предикат $B_i^{pr*}(x_i, x_j)$ (от англ. «branch») формализует дугу графа G_i^* , которая со-

единяет его вершины V_i^{p*} и V_i^{r*} с номерами p и r .

Предикат $B_i^{pr*}(x_i, x_j)$ содержит информацию о связях понятий C_i^p и C_i^r . Векторная переменная $x_i^r = (x_{i1}^r, x_{i2}^r, \dots, x_{im}^r)$ предиката C_i^r определена на множествах $A_{i1}^{r*}, A_{i2}^{r*}, \dots, A_{im}^{r*}$, содержащих m свойств вершины V_i^{r*} .

Построение предикатных моделей специализированных графов G_i^{**} заключается в сокращении областей определения переменных $x_{i1}^p, x_{i2}^p, \dots, x_{in}^p$. Свойства понятия C_i^p , характеризующего вершину V_i^{p*} графа G_i^{**} , будут иметь области определения $A_{i1}^{p**} \subseteq A_{i1}^{p*}, A_{i2}^{p**} \subseteq A_{i2}^{p*}, \dots, A_{in}^{p**} \subseteq A_{in}^{p*}$, т.е. в общем случае – подмножества свойств понятия C_i^p :

$$C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), \quad (4)$$

$$x_{i1}^p \in A_{i1}^{p**}, x_{i2}^p \in A_{i2}^{p**}, \dots, x_{in}^p \in A_{in}^{p**}, p = \overline{1, k}.$$

Дуги графа G_i^{**} представляются через отношения между понятиями:

$$B_i^{pr**}(x_i^p, x_i^r), x_i^p = (x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), \quad (5)$$

$$x_i^r = (x_{i1}^r, x_{i2}^r, \dots, x_{im}^r),$$

$$x_{i1}^p \in A_{i1}^{p**}, x_{i2}^p \in A_{i2}^{p**}, \dots, x_{in}^p \in A_{in}^{p**},$$

$$x_{i1}^r \in A_{i1}^{r**}, x_{i2}^r \in A_{i2}^{r**}, \dots, x_{im}^r \in A_{im}^{r**},$$

при этом предикаты понятий C_i^p и C_i^r остаются теми же, изменяются только области определения их переменных.

Общая модель прототипа структурированного объекта имеет вид системы предикатов:

$$\begin{cases} k \\ \bigwedge_{i=1}^k C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), \\ k \\ \bigwedge_{i,j=1}^k B_i^{pr**}(x_i^p, x_j^r), \end{cases} \quad (6)$$

где k - количество вершин V_i^{**} графа G_i^{**} и

$$x_i^p = (x_{i1}^p, x_{i2}^p, \dots, x_{in}^p), x_i^r = (x_{i1}^r, x_{i2}^r, \dots, x_{im}^r),$$

$$x_{i1}^p \in A_{i1}^{p**}, x_{i2}^p \in A_{i2}^{p**}, \dots, x_{in}^p \in A_{in}^{p**},$$

$$x_{i1}^r \in A_{i1}^{r**}, x_{i2}^r \in A_{i2}^{r**}, \dots, x_{im}^r \in A_{im}^{r**}.$$

Как видно из выражения (6), в отличие от существующих моделей в форме обобщенного нагруженного графа отношений, предлагаемая предикатная модель позволяет построить унифицированное описание структурируемого объекта.

Прототип использует единый формализм для представления как структурной составляющей объекта, так и свойства его элементов. Это позволяет рассматривать свойства элементов в качестве от-

дельных понятий и итеративно уточнять и дополнять описание объекта при его изменении во времени – т.е. создает условия для решения подзадачи адаптации.

Тогда изменение объекта во времени можно формализовать двумя способами:

- через изменение свойств вершин и дуг графа;
- через изменение структуры представляющего его графа;

Изменение структуры объекта в значительной степени можно отразить через изменение свойств его элементов. Для этого достаточно определить понятие отсутствия элемента, определив его через свойство не-существования, запрещающее наличие всех иных свойств.

$$\forall x_{il}^p \in A_{il}^{p**}, x_{il}^p \notin A_{il}^{p*} \setminus A_{il}^{p**} \quad (7)$$

$$C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p) = 1, l \in \overline{1, n}.$$

Поэтому в данной работе используется первый способ формализации изменений объекта.

Прототип должен объединять различные модификации объекта во времени. Реализация такого представления при предикатном подходе выполняется через объединение подмножеств свойств отличающихся понятий, при этом связи между ними для различных модификаций определяются полученными в результате объединения областей определения переменных предикатов C_i^p и B_i^{pr**} :

$$\forall x_{il}^p \in A_{il}^{p**} \cup A_{i2}^{p**} \cup \dots \cup A_{in}^{p**} \quad (8)$$

$$C_i^p(x_{i1}^p, x_{i2}^p, \dots, x_{in}^p) = 1, l \in \overline{1, n}.$$

$$\forall x_{il}^p \in A_{il}^{p**} \cup A_{i2}^{p**} \cup \dots \cup A_{in}^{p**},$$

$$\forall x_{iq}^r \in A_{il}^{r**} \cup A_{i2}^{r**} \cup \dots \cup A_{im}^{r**} \quad (9)$$

$$B_i^{pr**}(x_i^p, x_i^r) = 1, l \in \overline{1, n}, q \in \overline{1, m},$$

где l - некоторое свойство переменной x_i^p вершины V_i^{p**} , а q - некоторое свойство переменной x_i^r вершины V_i^{r**} графа G_i^{**} .

Таким образом, предлагаемая модель обеспечивает возможность постепенного наращивания прототипа на основе введения новых понятий и уточнения их свойств.

Выводы

Предложена предикатная модель представления прототипа структурированного объекта. Данная модель использует предикатную формализацию для описания как структурной составляющей объекта, так и свойства его элементов. Это позволяет рассматривать свойства элементов в качестве отдельных понятий, что обеспечивает возможность постепенного наращивания прототипа на основе введения новых понятий и уточнения их свойств. В частности, такой прототип может объединять различные модификации объекта во времени.

Список литературы

1. Posner M.I. Perceived distance and the classification of distorted patterns [Текст] / M.I. Posner, R. Goldsmith, K.E. Welton // *Journal of Experimental Psychology*. – 1967. – Vol. 73. – P. 28-38.
2. Neumann P.G. Visual prototype formation with discontinuous representation of dimensions of variability. [Текст] / P.G. Neumann // *Memory & Cognition*. – 1977. – 5(2). – P. 187-197.
3. Weskamp N. Efficient Algorithms for Robust Pattern Mining on Structured Objects with Applications to Structure-Based Drug Design [Текст] / Nils Weskamp // *Dissertation dem Fachbereich Mathematik und Informatik der Philipps-Universität Marburg, Marburg/Lahn*, 2006. – 177 p.
4. Raymond J.W. Effectiveness of graph-based and fingerprint-based similarity measures for virtual screening of 2d chemical structure databases [Текст] / J.W. Raymond, P. Willet // *Journal of Computer-Aided Molecular Design*. – 2002. – 16(1). – P. 59-71.
5. Holder L.B. Graph-Based Relational Learning: Current and Future Directions [Текст] / Lawrence B. Holder, Diane J. Cook // *SIGKDD Expl. Newsl.* – 2003. – 5(1). – P. 90-94.

Поступила в редколлегию 5.06.2015

Рецензент: д-р техн. наук, проф. С.Ф. Чалый, Харьковский национальный университет радиоэлектроники, Харьков.

РОЗРОБКА ПРЕДИКАТНОЇ МОДЕЛІ ПРОТОТИПУ СТРУКТУРОВАНОГО ОБ'ЄКТА В БАЗІ ЗНАНЬ

Кудхаир Абед Тамер

Запропонована предикатна модель представлення прототипу структурованого об'єкта. Дана модель використовує предикатну формалізацію для опису як структурної складової об'єкта, так і властивості його елементів. Це дозволяє розглядати властивості елементів в якості окремих понять, що забезпечує можливість поступового нарощування прототипу на основі введення нових понять і уточнення їх властивостей.

Ключові слова: алгебра скінченних предикатів, предикат, структурований об'єкт, прототип, подання знань.

DEVELOPMENT OF THE STRUCTURED OBJECTS PROTOTYPE PREDICATE MODEL IN THE KNOWLEDGE BASE

Kudhair Abed Tamer

A predicate model of representation of the prototype of a structured object. This model uses a predicate to describe how the formalization of the structural component of the object, and the properties of its elements. This allows us to consider the properties of the elements as separate concepts, enabling a gradual build-up of the prototype based on the introduction of new concepts and update their properties.

Keywords: algebra of finite predicate, the predicate, structured object, prototype, knowledge representation.