Проблемы бионики
1998. Вып. 49. С. 99—108

ХТУРЭ

УДК 007.681.518.2

М.Ф. ВОРОНОЙ

ЗАДАНИЕ ТОПОЛОГИИ НА МНОГОАСПЕКТНЫХ КЛАССИФИКАЦИОННЫХ СТРУКТУРАХ ПРИ МОДЕЛИРОВАНИИ КОНЦЕПТУАЛЬНЫХ ЗНАНИЙ

Введение

Современное состояние рынка программных продуктов свидетельствует о наличии широкого спроса на системы, использующие знание-ориентированные технологии. Это обусловлено тем, что к кругу пользователей средств обработки информации для решения повседневных задач примыкает все большее число людей, не обладающих специальными техническими знаниями. В результате формируется спрос на программные продукты, которые, с одной стороны, снимают с пользователя значительную часть рутинной работы, а с другой — обеспечивают управление процессами обработки со стороны пользователя в виде метафор, близких к пользовательскому непрофессиональному уровню навыков и знаний.

Характеризуя данные требования, можно сказать, что средства обработки информации должны обладать интеллектуальным ядром, обеспечивающим представление, приобретение и обработку сложноструктурированной информации об окружающей действительности. В наиболее общем представлении данное интеллектуальное ядро рассматривается как база знаний и способы (которые зависят от целей реализации интеллектуальной компоненты) манипулирования знаниями, имеющимися или поступающими из окружающего мира.

Рассмотрим особенности формализации представления знаний, соответствующих запросам современных информационных технологий.

Считается общепризнанным, что одной из основных задач, решаемых при построении базы знаний, является моделирование понятийного аппарата. Так, в работе [1] утверждается, что большая часть наших знаний о реальном мире относится к понятийным. Поэтому считается, что те структуры данных, которые позволяют представлять этот тип знаний, наиболее пригодны для моделирования знаний. Данная парадигма моделирования понятий (иногда применяется синоним "концепт") и взаимосвязей между ними активно используется во многих современных системах представления знаний.

Исследование и разработка теоретических основ и принципов построения понятийной модели предметной области показывают необходимость наличия в ее основе иерархической (в первую очередь, родовидовой [2]) классификации понятий.

Основной задачей, решающейся в данной работе, является описание элементов классификационных структур с использованием аппарата топологии. Понятия рассматриваются как подмножества семантического пространст ва моделируемой проблемной области, на основании чего открывается воз можность описания формализмов для оперирования семантикой.

Элементы классификационных структур

В качестве средства моделирования концептуальных знаний проблем ных областей предложены классификационные структуры, адаптированны для одновременного решения задач моделирования знаний и вывода на них Для того чтобы описание данных структур было целостным, далее приведень как общепринятые сведения о подобных средствах моделирования знаний, та и специфические, необходимые для решения задач вывода.

Предлагаемая модель знаний представлена многоаспектной классифи кационной структурой. Объектами классификации являются понятия про блемной области. С понятием, отраженным в классификации, ассоциировань некоторые его "свойства". Под последними будем понимать любые признаки которые можно вычленить у рассматриваемого объекта, описываемого поня тием. Будем считать, что объект классификации обладает определенным свой ством, если с понятием, соответствующим объекту, ассоциировано данно свойство.

На всем множестве свойств, выделенных в проблемной области у рас сматриваемых объектов, введем понятие "тип свойства". Для пояснения ег смысла рассмотрим следующий пример. Пусть представлены два объекта обладающие соответствующими множествами свойств:

объект 1:{"красное", "гладкое"}, объект 2: {"синее", "шершавое"}.

Здесь выделены два типа свойств: свойство цвета (включающее свойст ва "красное" и "синее") и шероховатости (включающее "гладкое" "шершавое").

Тип свойства — это множество, в которое объединены признаки одно природы. С введением типов свойств все их множество оказывается разбиты на классы эквивалентности:

$$\forall i,j: \quad TS_i \cap TS_j = \varnothing, i \neq j; i,j = \overline{1,n},$$

где $TS_{i,j}$ — типы свойств некоторой проблемной области, содержащей n раз личных типов.

Возникает закономерное предположение о том, что, поскольку свойств образуют класс, возможно построение целой классификации свойств про блемной области. Реализации данной идеи препятствует ряд проблем. В част ности, должны быть решены вопросы о месте классификации свойств и соот несении ее с классификацией объектов, о вычленении у свойств оснований дл их классифицирования (т.е., по сути, для определения свойств свойства). Од нако в рамках данной работы подобные вопросы не затрагиваются. Их рас смотрению, например, посвящена работа [3]. Для дальнейшего изложения достаточно отметить следующее:

– каждый объект представляется множеством типов свойств;

– в одном типе свойств у некоторого объекта может быть более чем одно свойство, например: "бело-голубое небо";

– объект описывается как множество свойств, принадлежащих одному или нескольким типам.

Одним из важнейших понятий для семантических сетей, в частности для классификаций, является понятие отношений, задаваемых между объектами. Отношения для классификаций, как правило, двуместны и направленны, т.е. для каждого отношения строго заданы выходной и входной объекты.

Выходной объект, по отношению к входному, будем называть надобъектом, или надсистемой (super-system)[1], если речь идет о системах, отраженных в понятиях проблемной области. Входной объект, по отношению к выходному, будем называть подобъектом, или подсистемой (sub-system)[3], если имеется в виду система, отраженная в понятиях проблемной области.

Учитывая особенности описания в данной работе объектов классификации, будем рассматривать отношения как оперирующие над свойствами объекта множества отображений [3].

Столь краткое описание отношений не раскрывает их сути и важности для моделирования знаний, однако по мере раскрытия содержания это будет сделано.

При данном описании элементов классификации определим некоторые ее свойства.

1. С каждой классификацией можно связать направленный граф, узлы которого соответствуют объектам, а дуги — отношениям между объектами, причем дуга направлена от выходного к входному объекту.

2. Классификационные структуры не могут содержать циклических структур, поскольку иначе появляются описания объектов через самих себя, что делает невозможным их рассмотрение применительно к детерминированным системам, использующим моделирование понятий. Тем не менее у одного подобъекта возможно наличие двух и более надобъектов.

3. Наличие совокупности объектов и отношений между ними позволяет рассматривать классификацию как разновидность семантических сетей[4]. При введении такой трактовки необходимо разъяснить смысл понятия семантики по отношению к объектам. Для семантических сетей вообще данное понятие охватывает очень широкий круг интерпретаций, однако для описываемых когнитивных структур определим это следующим образом. Семантикой объекта *O* будем считать: множество принадлежащих ему свойств, любые признаки, которые ставятся ему в соответствие посредством отношений с

другими объектами, и имена объектов, рассматриваемые как признаки особого рода [5].

Подводя итог, отметим, что семантика объекта включает в себя: описанные выше свойства объектов, поставленное ему в соответствие множество свойств объектов посредством отношений и имена объектов. Отметим, что множество всех имен также образует свой тип свойств и с введением подобного рассмотрения появляется возможность поддержки механизма синонимов на всей семантической сети. В случае синонимии объект обладает несколькими свойствами в типе "имя".

Проанализируем все множество семантик проблемной области с теоретико-множественной точки зрения. Очевидно, что все указанное множество определяет некоторое пространство S. Так же, как и множество свойств, множество семантик можно разбить на подмножества типов, объединяющих семантики разной природы. Это разбиение тоже образует классы эквивалентности:

$$\forall i, j : TS_i \cap TS_j = \varnothing, i \neq j; i, j = \overline{1, k},$$

где k — количество типов семантик, заданных на классификации проблемной области.

Тогда все семантическое пространство описывается так:

$$\mathbf{S} = \bigcup_{i=1}^{k} TS_i .$$

Для пространства, разбитого на классы эквивалентности по типам, уточним основные теоретико-множественные операции между объектами. Прежде всего, отметим, что на исследуемой проблемной области будем рассматривать лишь конечное и фиксированное число типов семантик.

Для каждого объекта классификации разобьем все множество его семантик на два подмножества.

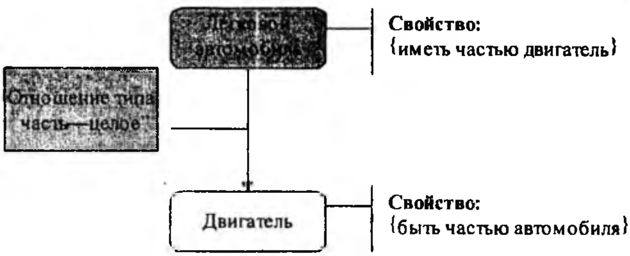
Будем называть Ω-семантикой (омега-семантикой) объекта O_i такое подмножество его семантик, которое существует только в данном объекте, а в связанных с ним объектах воспроизводится только за счет отображений посредством отношений.

Будем называть Ψ-семантикой (пси-семантикой) объекта O_i такое подмножество его семантик, которое отображается посредством отношений из других объектов.

Другими словами, Ω-семантика некоторого объекта O_i — это подмножество, с одной стороны, уникально определяющее объект в классификации (при условии, что Ψ-семантика не рассматривается), а с другой — определяющее, у какого объекта данный набор семантик встречается впервые. В свою очередь, Ψ-семантика всегда отражает тот факт, что объект O_i, обладающий ею, связан набором отношений с другими объектами, обладающими

Ω-семантикой. Таким образом, не существует Ψ-семантики самостоятельно без Ω-семантики или в отрыве от классификации.

Следует обратить внимание на то, что в обоих определениях намеренно не указывается, связан ли рассматриваемый объект с надобъектами или с подобъектами. Дело в том, что хотя отношения на классификации и направленные, но семантика, отражаемая ими, может распространяться в обе стороны. Для примера, представленного на рисунке, указаны свойства, полученные в результате рассмотрения связи между объектами. Хотя пример упрощен, он отражает возможность появления свойств как у надобъекта (за счет наличия подобъекта), так и у подобъекта (за счет связи с надобъектом).



При неправильном понимании определения Ω-семантики может сложиться мнение о том, что ее возникновение не связано с введением отношений. Однако (по крайней мере, для данного примера) семантики {иметь частью двигатель} и {быть частью автомобиля} полностью соответствуют определению
Ω-семантики.

Рассмотрим основные свойства объектов, вытекающие из разбиения их множества семантик на два подмножества.

Свойства Ω- и Ψ- семантик

Очевидно, что любой объект в классификации будет определяться объединением его Ω- и Ψ-семантик. Такое объединение будем называть полной семантикой объекта:

$$O_i = \Omega_i \cup \Psi_i .$$

Следует отметить, что возможно существование объектов в классификации, для которых Ω-семантика — пустое множество. Такое состояние свидетельствует о наличии развитой и информативной классификации, для которой удалось полностью описать некоторый объект через содержание других.

Для дальнейшего рассмотрения свойств семантик зафиксируем два объекта O_i и O_j, такие, что $j \neq i$. С их помощью опишем три свойства:

I. Пересечение любых двух Ω-семантик пусто:

$$\Omega_i \cap \Omega_j = \varnothing .$$

Данное свойство позволяет сделать математическую запись ранее ука занной "уникальности" для каждой Ω-семантики произвольного объекта:

II. Пересечение Ω- и Ψ- семантики любого объекта всегда пусто:

$$\Omega_i \cap \Psi_i = \varnothing .$$

Этим определяется определяется тот факт, что для любого объекта пере сечение его Ω- и Ψ- семантики всегда пусто.

Д о к а з а т е л ь с т в о. Допустим обратное, т.е.

$$\Omega_i \cap \Psi_i \neq \varnothing .$$

Тогда, по определению Ψ-семантики, для объекта O_i существует неко торый объект O_j, такой что

$$\Omega_j \cap \Psi_i \neq \varnothing .$$

Поскольку любая Ψ-семантика состоит из Ω-семантик других объектов то, без ограничения общности, можно предположить, что

$$\Omega_i \cap \Omega_j \neq \varnothing .$$

Однако это невозможно ввиду определения Ω-семантики.

III. Если для двух разных объектов существует пересечение их Ω- и Ψ семантик:

$$\Omega_i \cap \Psi_j \neq \varnothing ,$$

то эти объекты непосредственно связаны между собой некоторым отношени ем R:

$$O_i \xrightarrow{R} O_j$$

либо они связаны транзитивно через множество других объектов:

$$O_i \xrightarrow{R_1} \alpha_1 \xrightarrow{R_2} \alpha_2 \xrightarrow{R_3} \cdots \alpha_n \xrightarrow{R_{n+1}} O_j ,$$

где n — количество объектов α_q, $q = \overline{1, n}$, связанных между собой отношения ми $\{R_p\}_{p=2}^{n+1}$ и находящихся между объектами O_i и O_j. Далее указанную транзитивность будем обозначать так : $O_i \xrightarrow{+} O_j$.

Необходимо отметить, что Ω-семантики всех объектов классификации разбивают все семантическое пространство S классифицируемой проблемной области на классы эквивалентности*, т.е.:

$$S = \bigcup_{i=1}^{n} \Omega_i \, ,$$

где n — количество объектов в классификации, обладающих Ω-семантикой.

Рассмотренный перечень свойств классификаций далеко не исчерпывающий, и по мере раскрытия содержания работы он будет пополняться. Однако изложенного материала уже достаточно для формализации классификации в терминах топологии.

Основные определения топологии

Пусть X — произвольное множество и T — совокупность его подмножеств (называемых открытыми). Последние удовлетворяют следующим трем условиям[4]:

1. Пустое множество $\varnothing \in T$, $X \in T$.
2. Пересечение двух множеств из T принадлежит T.
3. Объединение любой совокупности множеств из T принадлежит T.

В этом случае пара (X, T) называется топологией, а множество T — топологической структурой.

Приведем основные определения топологических структур.

Множество X с выделенной топологической структурой T называется топологическим пространством (X,T).

Множество B открытых множеств топологического пространства называется его базой (базой открытых множеств), если каждое открытое множество этого пространства является объединением множеств из B [7]. Очень часто, вводя во множестве X топологию, указывают явно лишь некоторую ее базу. Очевидно при этом, что множество B подмножеств множества X тогда и только тогда может служить базой открытых множеств некоторой топологии на X, когда пересечение любых двух множеств из B является объединением множеств из B [6].

Топологическое пространство X называется хаусдорфовым (или отделимым), если любые две его точки p и q обладают непересекающимися окрестностями[7].

Подмножество C топологического пространства X называется замкнутым, если открыто множество $X \backslash C$. Здесь и далее '\' — операция дополнения, означающая множество элементов X, не содержащихся в C. Множества $\varnothing$ и X

---

* Особо следует проанализировать свойства классификации, для которой класс эквивалентности типов семантик и класс Ω-семантик совпадают.

замкнуты. Объединение любых двух замкнутых множеств замкнуто. Пересечение любой совокупности замкнутых множеств замкнуто[8].

Для любого подмножества Y топологического пространства X можно рассмотреть наименьшее замкнутое множество, содержащее Y; оно обозначается через $\bar{Y}$ и называется замыканием Y. Другими словами, $\bar{Y} = \bigcap_j F_j$, где F_j — семейство всех замкнутых множеств, содержащих Y.

Если топология T совпадает с множеством $\mathfrak{I}(X)$ всех подмножеств X, то в этом случае задается топология на X, называемая дискретной.

Сравнение топологий [7]

Пусть на одном и том же носителе X заданы две топологии τ_1 и τ_2. Тем самым определены два топологических пространства:

$$T_1 = (X, \tau_1) \; ; \; T_2 = (X, \tau_2).$$

Топология τ_1 сильнее топологии τ_2, если система множества τ_2 содержится в τ_1. О топологии τ_2 говорят, что она слабее, чем τ_1. В совокупность всех возможных топологий множества X естественным образом вводится частичная упорядоченность (топология τ_2 предшествует τ_1, если она слабее, чем τ_1). В этой совокупности топологий есть максимальный элемент — топология, в которой все множества открыты, и минимальный — топология, в которой открыты только все X и $\varnothing$.

Введение топологии на многоаспектных классификационных структурах

В качестве носителя топологии X рассмотрим объединение всех Ω-семантик проблемной области:

$$X = \bigcup_{i=1}^{n} \Omega_i ,$$

где n — количество объектов, обладающих Ω-семантикой.

Зададим топологию T следующим образом:

$$T = \left\{ \bigcup_{i \in \{0,1\}^n} A_i \right\},$$

где n — количество объектов, обладающих Ω-семантикой; i — некоторый вектор, $i = (i_1, i_2, \ldots, i_n) \in \{0, 1\}^n$; A_i — всевозможные комбинации объединений Ω- семантик на X,

$$A_{i_k} = \begin{cases} \Omega_k, j_k = 1; \\ \varnothing, j_k = 0; \end{cases}$$

$$A_i = \bigcup_{k=1}^{n} A_{i_k}.$$

Для введенных обозначений пара (***X*, *T***) является топологическим пространством

Проверим выполнение трех аксиом топологии.

1. Пустое множество ∅ и само семантическое пространство ***S*** принадлежат ***T***.

2. Пересечение двух множеств из ***T*** принадлежит ***T***, поскольку любой объект на классификации описывается через объединение Ω- и Ψ- семантик, а Ψ-семантика, в свою очередь, является также объединением Ω-семантик. Следовательно, результатом операции объединения будет множество, состоящее из пересечений Ω-семантик.

3. Объединение любой совокупности множеств из ***T*** также принадлежит ***T***, поскольку в указанной топологии перечислены все возможные пересечения Ω- семантик пространства ***S***.

Из описания ***T*** вытекают два требования к классификации:

1. На классификации объектов должно существовать отражение наличия пустого множества ∅.

Содержательная интерпретация данного требования такова. В работе [3] имеется явное указание на единственную вершину естественной классификации с нулевым (пустым) содержанием. В прагматических (не обязательно существенных) классификациях конкретных проблемных областей следует рассматривать вершину с нулевым содержанием как объект, отражающий классифицируемую проблемную область. Поскольку для построения такой классификации инженер по знаниям должен зафиксировать тот уровень обобщений, выше которого рассмотрение не имеет содержательной ценности, то таковым уровнем и будет объект — “проблемная область” с пустой семантикой, соответствующей введенному требованию о пустом множестве.

2. Свойства без объектов не существуют, так как в противном случае свойство, не принадлежащее никакой Ω-семантике, окажется вне какого-либо объекта и, следовательно, вне рассматриваемой топологии.

Содержательная интерпретация данного требования заключается в том, что оно вполне естественно и отражает тот факт, что любое свойство обязательно проявляется каким-то объектом и, следовательно, может быть вписано в классификацию проблемной области.

Перечислим некоторые свойства введенной топологии.

1. Рассматриваемая топология не является в общем случае дискретной (исключая случай, когда каждый объект представляет собой пустое множество либо множество из одного элемента).

Доказательство. Выделим непустое подмножество A Ω-семантики некоторого объекта O_i т.е. $A \subset \Omega_i$. Поскольку $A \neq \Omega_i$; $A \neq \varnothing$; $\Omega_i \bigcap \Omega_j = \varnothing$ для любого $j \neq i$ (это равносильно тому, что A не принадлежит никакому объединению Ω-семантик), то $A \notin \boldsymbol{T}$, что противоречит определению дискретной топологии.

2. Любое подмножество рассматриваемой топологии замкнуто. Для доказательства этого следует показать, что любое множество A замкнуто ($A \in \boldsymbol{T}$).

Доказательство. Множества $\varnothing$, $\boldsymbol{X}$ замкнуты [8]. Множество Ω_i замкнуто, поскольку $\boldsymbol{X} / \Omega_i = \bigcup\limits_{j \neq i} \Omega_j$ открыто (объединение открытых множеств является открытым множеством [8]). И так как объединение любых двух замкнутых множеств замкнуто, делаем вывод, что любое множество $A \in \boldsymbol{T}$ замкнуто.

3. Топологическое пространство ($\boldsymbol{X}$, $\boldsymbol{T}$) в общем случае не хаусдорфово (исключая случай, когда каждый объект представляет собой пустое множество либо множество из одного элемента).

Доказательство. Достаточно указать две точки топологического пространства, любые окрестности которых имеют непустое пересечение.

Рассмотрение основных свойств введенной топологии позволяет перейти к анализу важных следствий, являющихся результатом изучения семантического пространства концептуальных моделей проблемных областей.

Список литературы: 1. *Dilger W.* Object-oriented knowledge representation an overview // J. New generation of computer systems. 1989. N 2. P. 339–363. 2. *Чупилина Е.И.* Место термина в лексико-семантической системе языка // Вопросы терминологии и лингвистической статистики. Воронеж, 1972. С. 25–31. 3. *Соловьева Е.А., Маторин С.И.* Методы моделирования и модели понятийных знаний // НТИ. Сер. 2. 1989. № 4. С. 2–8. 4. *Искусственный* интеллект: В 3 кн.: Справ. / Под ред. Д.А. Поспелова М.: Радио и связь, 1990. Кн. 2: Модели и методы. 304 с. 5. *Поспелов Д.А.* Ситуационное управление: теория и практика. М.: Наука, 1986. 288 с. 6. *Постников М.М.* Лекции по геометрии. Гладкие многообразия. М.: Наука, 1987. 480 с. 7. *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. М.: Наука, 1989. 624 с. 8. *Косневски Ч.* Начальный курс алгебраической топологии: Пер. с англ. М.: Мир, 1983. 304 с.

Поступила в редколлегию 09.04.98