

Т. М. КОСОВСКАЯ

Санкт-Петербургский государственный университет, Санкт-Петербург

ФОРМУЛЫ ИСЧИСЛЕНИЯ ПРЕДИКАТОВ ДЛЯ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ О СЛОЖНЫХ СТРУКТУРИРОВАННЫХ ОБЪЕКТАХ

Описывается подход к решению задач Искусственного Интеллекта, связанных с исследованием сложных структурированных объектов (ССО), основанный на использовании языка исчисления предикатов (ИП). Рассматривается решение следующих задач: описание сложного структурированного объекта, распознавание сложного структурированного объекта; выделение общих свойств объектов; построение логико-предикатных распознающих сетей; вычисление метрики в пространстве описаний объектов; построение нечётких предикатных сетей; построение онтологии.

Введение. При решении многих задач искусственного интеллекта используются логические описания исследуемых объектов. При этом, как правило, объект описывается с помощью строки значений бинарных или конечнозначных признаков, характеризующих свойства этого объекта. При этом решение большинства задач при таком описании объектов имеет полиномиальную вычислительную сложность

Для сложных структурированных объектов (ССО), которые являются совокупностью элементов, обладающих заданными свойствами, и находящихся в заданных отношениях (быть может многоместных), такой подход не вполне удобен. В работе изложен подход, основанный на использовании языка исчисления предикатов (ИП). Такой подход был изложен ещё в книге Н. Нильсона [1], однако необходимость использования знаний о предикатных формулах и экспоненциальная вычислительная сложность решения задач с таким описанием, отпугивает многих исследователей. Ниже будет показано, что реальное время решения одной и той же задачи при описаниях в виде строки значений признаков и в виде формулы исчисления предикатов практически совпадает, так как длина соответствующей строки экспоненциальна по сравнению с длиной записи формулы.

В предлагаемом докладе описываются различные задачи, которые могут быть решены при описании ССО средствами языка ИП. Приводятся оценки вычислительной сложности задач. Для некоторых задач сравнивается их вычислительная сложность при разных описаниях объекта.

Описание сложного структурированного объекта. Пусть исследуемый объект ω задан множеством своих элементов $\omega = \{\omega_1, \dots, \omega_i\}$, на котором задан набор предикатов p_1, \dots, p_n , задающих свойства элементов или отношения между ними. Описанием ССО ω называется элементарная конъюнкция всех атомарных формул с предикатами p_1, \dots, p_n , истинных для ω [2]. В дальнейшем описание ССО будем обозначать посредством $S(\omega)$.

Отметим, что если только один элемент ω_j обладает свойством p_i , то достаточно записать $p_i(\omega_j)$. При описании с помощью бинарной строки необходимо для этого свойства выделить t позиций и только на j -ом месте поставить 1, а на всех остальных местах записать 0. То есть если имеется r свойств, то в описании объекта для записи этих свойств с помощью бинарной строке следует выделить rt позиций, в большинстве из которых будет записан 0.

Если же p_i – задаёт k -местное отношение, то для него в описании с помощью бинарной строки следует выделить t^k позиций, в большинстве из которых будет записан 0. Поскольку k может быть достаточно велики (сравнимо с t), то длина записи бинарной строки может экспоненциально зависеть от длины записи элементарной конъюнкции, задающей описание ССО.

Распознавание сложного структурированного объекта. Под задачей распознавания здесь будет пониматься задача отнесения ССО к тому или иному классу объектов, общие свойства которых заданы с помощью элементарной конъюнкции $A(x)$, где x – список переменных. Для ССО можно сформулировать 3 задачи распознавания [2]. Задача классификации: Удовлетворяет ли объект ω формуле $A(x)$? Задача идентификации: Найти часть объекта ω формуле $A(x)$? Задача анализа: Найти и идентифицировать все части объекта ω , удовлетворяющие формулам $A_1(x_1), \dots, A_K(x_K)$. Эти задачи могут быть записаны следующими формулами:

$$S(\omega) \Rightarrow \exists \Pi(\omega) A(\Pi(\omega)) \quad (1)$$

$$S(\omega) \Rightarrow \exists x_{\neq} A(x) \quad (2)$$

$$S(\omega) \Rightarrow \bigvee_{j=1}^K \exists x_{\neq} A_j(x) \quad (3)$$

где $\Pi(\omega)$ – перестановка элементов из ω , $\exists x_{\neq}$ означает «существует ли набор различных значений для списка x ?».

Задача (1) полиномиально эквивалентна открытой задаче «Изоморфизм графов», т. е. GI-полна [3], а задачи (2) и (3) NP-полны [2]. Оценки числа шагов известных алгоритмов экспоненциально зависят от длины записи правых частей формул.

Выделение общих свойств объектов. При различных способах описания ССО понятие общего свойства объектов формализуется существенно разными способами. В частности, при описании объектов с помощью конечнозначных строк общим свойством объектов является общая подстрока их описаний, нахождение которой линейно (на устройствах с прямым доступом) относительно длины строк. При описании ССО на языке ИП в качестве общего свойства объектов не может выступать общая подформула их описаний, т. к. эти подформулы могут иметь существенно разные аргументы. Это привело к введению понятия изоморфизма предикатных формул (изоморфизма элементарных конъюнкций предикатных формул) [4].

Определение. Две элементарные конъюнкции предикатных формул называются изоморфными, если они совпадают с точностью до имён аргументов и порядка следования литералов.

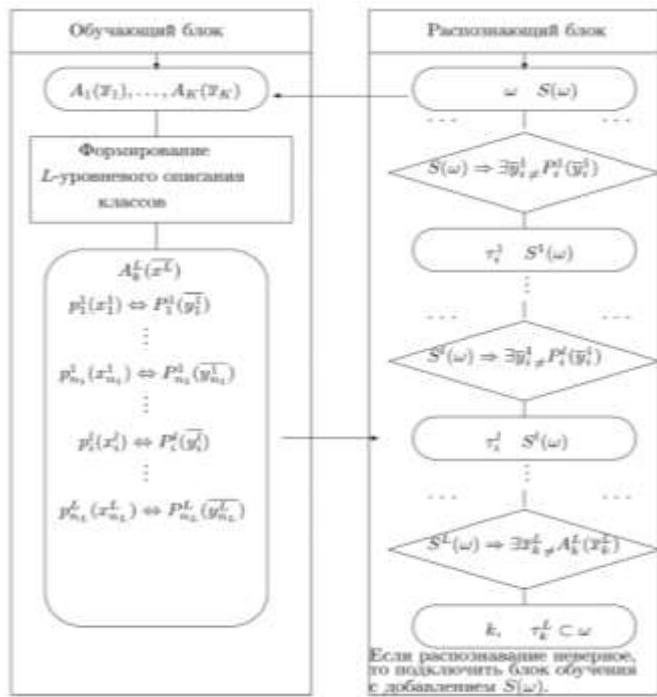


Рис. 1. Схема логико-предикатной распознающей сети

Следует отметить, что задача проверки формул на изоморфизм GI-полна. Если же требуется не только проверить, что переименование аргументов возможно, но и найти соответствующее взаимно-однозначное соответствие, то она GI-трудна [3]. В работе [5] предложен и исследован эвристический полиномиальный по времени алгоритм проверки формул на изоморфизм. В случае отрицательного ответа формулы действительно не изоморфны. Численные эксперименты показали, что в случае положительного ответа для 99,5 % пар формул они изоморфны.

Построение логико-предикатных распознающих сетей. Использование понятия изоморфизма предикатных формул позволило ввести понятие многоуровневого описания классов объектов. В процессе построения такого описания последовательно выделяются общие свойства объектов одного класса

и формулы, соответствующие одному общему свойству, заменяются на атомарную формулу с новым предикатом [5].

Решение задачи распознавания сводится к последовательной проверке формул вида (2) с меньшей длиной записи правой части, что существенно уменьшает вычислительную сложность решения задач распознавания.

На основе многоуровневого описания предложено понятие логико-предикатной распознающей сети, представленной на рис. 1.

Вычисление метрики в пространстве описаний объектов. Использование понятия изоморфизма предикатных формул позволило ввести понятие метрики в пространстве описаний объектов. В [6] в качестве метрики в пространстве описаний ССО предложена следующая формула: $\rho(\omega^1, \omega^2) = (s^1 - s^{1,2}) + (s^2 - s^{1,2})$, где s^1 и s^2 – количества литералов в ω^1 и ω^2 соответственно, $s^{1,2}$ – количество литералов в наибольшей формуле, задающей общее свойство ω^1 и ω^2 .

Недостатком такой метрики является то, что для объектов с длинным описанием и достаточно большой формуле, задающей их общее свойство, расстояние может оказаться таким же, как для объектов с описанием небольшой длины и общим свойством из одного литерала.

Такая метрика может быть применена при решении задачи кластеризации. Кроме того, с её помощью может быть определена «степень похожести» объектов, определяемая посредством деления $\rho(\omega^1, \omega^2)$ на $s^1 + s^2$.

Построение нечётких предикатных сетей. Логико-предикатная распознающая сеть правильно распознаёт те, и только те объекты, на которых она была обучена. Введение «степени похожести» между описаниями ССО позволило определить нечёткие предикатные сети, в каждой ячейке которых вычисляется такая «степень похожести». В результате нечёткая предикатная сеть выдаёт «степень уверенности» в том, что объект распознан правильно.

Построение онтологии. Для математиков онтология является ориентированным графом, в корневой вершине которого находятся описания объектов. В каждой дочерней вершине находится подмножество описаний отцовской, обладающих некоторым общим свойством.

Выделение общих свойств ССО, основанное на понятии изоморфизма предикатных формул, позволило разработать алгоритм построения онтологии множества ССО [7].

Заключение. В работе изложен подход к исследованию сложных структурированных объектов, описанных с использованием предикатных формул. Указываются отличия при решении задач при таком подходе от подхода, в котором объекты описываются с помощью конечнозначных строк. Сравняется вычислительная сложность решения рассмотренных задач при разных подходах.

Исследование выполнено в рамках проекта Санкт-Петербургского государственного университета ID 93024916.

ЛИТЕРАТУРА

1. **Нильсон Н.** Искусственный интеллект. Методы поиска решений. М.: Мир, 1973. 279 с.
2. **Косовская Т.М.** Доказательства оценок числа шагов решения некоторых задач распознавания образов, имеющих логические описания. *Вестн. С.-Петербург.ун-та. Сер. 1.* 2007. Вып.(4) С. 82–90.
3. **Косовская Т.М., Косовский Н.Н.** Полиномиальная эквивалентность задач изоморфизм предикатных формул и изоморфизм графов. *Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия.* 2019. Т. 6 (64). Вып. 3. С. 430–439.
4. **Kosovskaya T.** Isomorphism of Predicate Formulas in Artificial Intelligence Problems. *International Journal "Information Theories and Applications"*. 2019. Vol. 26. № 3. P. 221–230.
5. **Косовская Т.М., Петров Д.А.** Выделение наибольшей общей подформулы предикатных формул для решения ряда задач искусственного интеллекта. *Вестник СПбГУ. Прикладная математика. Информатика. Процессы управления.* 2017. Т. 13. Вып. 3. С. 250–263.
6. **Kosovskaya T.** Distance between objects described by predicate formulas. *International Book Series. Information Science and Computing. Book 25. Mathematics of Distances and Applications.* ITHEA–Publisher, Sofia, 2012. P. 153–159.
7. **Kosovskaya T.** Isomorphism of predicate formulas as the base of logic ontology construction. *International Journal "Information Theories and Applications"*. 2020. Vol. 27. № 3. P. 248-254.

T.M.Kosovskaya (St. Petersburg State University, St. Petersburg)

Predicate calculus formulas for representing knowledge on complex structured objects

An approach to solving Artificial Intelligence problems related to the study of complex structured objects, based on the use of the predicate calculus language, is described. The solution of the following problems is considered: description of a complex structured object, recognition of a complex structured object; identification of common properties of objects; construction of logical-predicate recognition networks; calculation of metrics in the description space objects; construction of fuzzy predicate networks; construction of ontology. The differences in solving problems with this approach from the approach in which objects are described using finite-valued strings are indicated. The computational complexity of solving the considered problems with different approaches is compared.

Автор готов представить текст на английском языке для сборника материалов мультиконференции, который будет подан для индексирования в Scopus.