Т. М. КОСОВСКАЯ
Санкт-Петербургский государственный университет, Санкт-Петербург

ВЫДЕЛЕНИЕ ИЗОМОРФНЫХ ПОДФОРМУЛ КАК СРЕДСТВО ДЛЯ СОЗДАНИЯ ЛОГИЧЕСКОЙ ОНТОЛОГИИ

Рассматривается способ построения логической онтологии для сложных структурированных объектов, описания которых заданы на языке исчисления предикатов. Кратко описывается алгоритм выделения подмножеств объектов, обладающих одинаковыми свойствами, и установления их взаимного расположения. Для построения такой онтологии использован алгоритм выделения наибольшей подформулы, изоморфной подформулам всех описаний объектов из класса.

Введение. Понятие онтологии как раздела философии существует с античных времён. При этом в философии под онтологией понимают абстракции, принимающие во внимание одни признаки предметов (или отношения между частями этих предметов) и игнорирующие другие. С развитием вычислительной техники и информатики этот термин был перенесён на построение подробной формализации некоторой области знаний, учитывающей взаимосвязь и структуру исследуемых объектов. Создание онтологии предметной области является одним из направлений представления знаний для использования их в системах искусственного интеллекта и извлечения знаний из баз данных [2, 3].

Для математика онтология – это ориентированный граф, вершины которого определяют некоторые понятия (классы объектов). Если дуга (N_1, N_2) соединяет два понятия N_1 и N_2, то все объекты, входящие в понятие N_2, входят также и в понятие N_1.

Под логической онтологией обычно понимают онтологию над сложными структурированными объектами. Элементы таких объектов обладают заранее заданными признаками и находятся в заранее заданных отношениях. При этом для всех объектов, входящих в одно понятие, найдётся подмножество его элементов, удовлетворяющее одной и той же формуле.

Удобным языком для описания таких объектов является язык исчисления предикатов [4]. В [5] описан алгоритм выделения формулы, изоморфной подформулам описаний множества объектов. Этот алгоритм является базовым для создания логической онтологии.

В настоящей работе рассматривается следующая процедура построения логической онтологии.

Дано множество сложных структурированных объектов, описания которых представлены в виде элементарных конъюнкций предикатных формул, задающих свойства элементов объектов и отношения между ними. Это множество задаёт исходное понятие для логической онтологии.

Для каждого множества объектов, задающего уже построенное понятие онтологии, требуется выделить максимальные формулы, являющиеся (с точностью до имён переменных) подформулами описаний группы объектов из этого понятия. Множества объектов, удовлетворяющие этим подформулам, задают новые понятия логической онтологии, являющиеся дочерними по отношению к тому, из которого они получены.

Предлагаемый доклад посвящен описанию основных определений и изложению схемы алгоритма, позволяющих реализовать только что описанную процедуру построения логической онтологии.

Основные определения. Прежде, чем дать постановку задачи, необходимо сформулировать основные определения.

Определение 1. Две элементарные конъюнкции атомарных формул исчисления предикатов $P(a_1,...,a_m)$ и $Q(b_1,...,b_m)$ называются **изоморфными**, если существуют такая элементарная конъюнкция $R(x_1,...,x_m)$ и подстановки аргументов $a_{i1},..., a_{im}$ и $b_{j1},..., b_{jm}$ формул $P(a_1,...,a_m)$ и $Q(b_1,...,b_m)$ соответственно вместо всех вхождений переменных $x_1,...,x_m$ формулы $R(x_1,...,x_m)$, что результаты этих подстановок $R(a_{i1},..., a_{im})$ и $R(b_{j1},..., b_{jm})$ совпадают с формулами $P(a_1,...,a_m)$ и $Q(b_1,...,b_m)$ соответственно с точностью до порядка литералов.

По сути дела для изоморфных формул существуют такие перестановки их аргументов, что они задают одно и то же отношение между своими аргументами.

Определение 2. Элементарная конъюнкция называется **максимальной формулой, изоморфной подформулам двух заданных элементарных конъюнкций**, если она изоморфна некоторым подформулам этих элементарных конъюнкций, но после добавления в неё любого литерала она не изоморфна ни одной подформуле хоть одной из них.

Определение 3. **Наибольшим общим свойством** элементов из множества N называется наибольшая по количеству литералов формула, для которой в каждом описании объекта из N имеется изоморфная подформула.

Постановка задачи.

Дано множество N_0 составных структурированных объектов вида $\omega=\{\omega_1,\dots,\omega_t\}$, на элементах которых заданы предикаты $p_1,\dots, p_n$, задающие свойства его элементов и отношения между ними. Описание объекта $S(\omega)$ – элементарная конъюнкция всех истинных на ω атомарных формул с предикатами $p_1,\dots, p_n$.

Требуется выделить из множества N_0 (возможно пересекающиеся) подмножества $N_1,\dots,N_m$ и построить ориентированный граф, у которого вершина с нулевым заходом помечена множеством N_0 и соответствует множеству неизоморфных формул, каждая из которых изоморфна некоторым описаниям объектов из N_0.

Если из вершины N_k выходят дуги к вершинам, помеченным множествами $N_{k1},\dots,N_{kr}$, то

- N_k является объединением множеств $N_{k1},\dots,N_{kr}$;
- для каждого $i=1,\dots,r$ каждая формула в описании объектов из N_{ki} изоморфна одной из формул в описании N_k, т.е. объекты из N_{ki} обладают всеми свойствами, общими для всех объектов из N_k;
- если $i \neq j$, то наибольшие общие свойства объектов из N_{ki} и N_{kj} различны.

Алгоритм решения задачи. Для решения этой задачи можно воспользоваться алгоритмом выделения максимальной формулы, изоморфной подформулам двух элементарных конъюнкций [5, 6]. Ниже посредством $\bar{x}$ обозначается список переменных в формуле.

1. Для каждой пары объектов ω^i и ω^j $(i \neq j)$ из множества N_0 выделяем максимальную формулу $Q^1_{ij}(\bar{x}^1_{ij})$, изоморфную подформулам элементарных конъюнкций $S(\omega^i)$ и $S(\omega^j)$. Если среди выделенных максимальных формул имеются изоморфные, то оставляем только одну из них и перенумеровываем эти формулы без повторений $P^1_1(\bar{x}^1_1), \dots, P^1_{n1}(\bar{x}^1_{n1})$. Из вершины, помеченной N_0, проводим ориентированные ребра к вершинам, помеченным $P^1_1(\bar{x}^1_1), \dots, P^1_{n1}(\bar{x}^1_{n1})$.
2. Повторяем при $l = 2, \dots, L$ (процесс завершится, так как на каждой итерации длины подформул уменьшаются) выделение максимальных формул, изоморфных подформулам для пар $P^{l-1}_i(\bar{x}^{l-1}_i)$ и $P^{l-1}_j(\bar{x}^{l-1}_j)$, получив формулы $Q^l_{ij}(\bar{x}^l_{ij})$.
   - Если среди выделенных максимальных формул имеются изоморфные, то оставляем только одну из них и перенумеровываем эти формулы без повторений $P^l_1(\bar{x}^1_1), \dots, P^l_{nl}(\bar{x}^l_{nl})$.
   - Из вершин, помеченных $P^{l-1}_i(\bar{x}^{l-1}_i)$ и $P^{l-1}_j(\bar{x}^{l-1}_j)$, проводим ориентированные рёбра к вершине, помеченной формулой $P^l_k(\bar{x}^l_k)$, изоморфной подформулам формул $P^{l-1}_i(\bar{x}^{l-1}_i)$ и $P^{l-1}_j(\bar{x}^{l-1}_j)$.
   - Если среди выделенных максимальных формул имеется изоморфная ранее выделенной, то вершину, соответствующую ранее выделенной максимальной формуле, «склеиваем» с вновь выделенной (вместе с рёбрами, ицидентными этой вершине).

Отметим, что при выполнении «склейки» невозможно возникновение «порочного круга», при котором вершина, полученная в результате «склейки», находится в ориентированном цикле. Действительно, пусть «склеиваются» вершины, соответствующие элементарным конъюнкциям $P^{l+1}_k(\bar{x}^{l+1}_k)$ и $P^{l'}_{k'}(\bar{x}^{l'}_{k'})$. Эти формулы изоморфны и, следовательно, имеют одинаковую длину. При похождении вдоль ориентированного пути длины элементарных конъюнкций убывают, т. к. формула, соответствующая следующей вершине, изоморфна подформуле предыдущей. Следовательно, невозможен ориентированный путь между «склеиваемыми» вершинами.

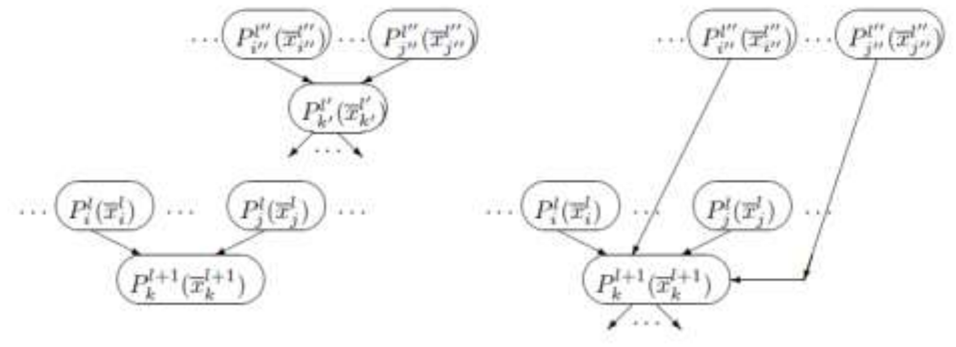

Рисунок. «Склеивание» вершин с изоморфными формулами $P^{l+1}_k(\bar{x}^{l+1}_k)$ и $P^{l'}_{k'}(\bar{x}^{l'}_{k'})$

Заключение. Предложенный алгоритм построения логической онтологии является экспоненциально сложным. Это обусловлено тем, что задача выделения максимальной формулы, изоморфной подформулам двух элементарных конъюнкций, NP-трудна. Однако, её использование, как и использование любой другой онтологии, позволяет достаточно быстро классифицировать исследуемый объект.

Интересно рассмотреть случай многозначных предикатов, описывающих исследуемые объекты. Перевод многозначных предикатов $q(\bar{x}) = a$ в бинарные может быть осуществлён посредством введения дополнительного бинарного предиката $p(\bar{x};a)$ с дополнительным аргументом a. Проблема состоит в том, что дополнительный аргумент, в отличие от аргументов $\bar{x}$, не является переменной для элемента объекта.

ЛИТЕРАТУРА

1. **Natalya F. Noy and Deborah L. McGuinness.** Ontology Development 101: A Guide to Creating Your Firt Ontology // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
2. **Бениаминов Е.М.** Некоторые проблемы широкого внедрения онтологий в IT и направления их решений. //Труды Симпозиума "ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ". М.:ИПИ РАН, 2008, с. 71-82.
3. **Дяченко О.О., Загорулько Ю.А.** Подход к коллективной разработке онтологий и баз знаний // Знания – Онтологии – Теории. Материалы Всероссийской конференции с международным участием. Изд-во: Институт математики им. С.Л. Соболева СО РАН, 2013. С. 141–149.
4. **Kosovskaya T.** Predicate Calculus as a Tool for AI Problems Solution: Algorithms and Their Complexity. Chapter 3 in: Intelligent System. Open access peer-reviewed Edited volume. 2018. Pp. 1–20. https://www.intechopen.com/books/intelligent-system/predicate-calculus-as-a-tool-for-ai-problems-solution-algorithms-and-their-complexity
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. **Косовская~Т.М.** Подход к решению задачи построения многоуровневого описания классов на языке исчисления предикатов // Труды СПИИРАН, 2014. №~3 (34). С. 204–217.

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

Isomorphous Subformules Isolation as a Means to Construct Logical Ontology

A method of constructing a logical ontology for complex structured objects, the descriptions of which are given in the predicate calculus language, is considered. An algorithm for extraction of subsets of objects with the same properties and establishing their relations is briefly described. To construct such an ontology, an algorithm for extraction the largest sub-formula isomorphic to sub-formulas of every class descriptions is used.