

19. Gladkov L.A., Kureychik V.V., Kureychik V.M. Geneticheskie algoritmy [Genetic algorithm]. Moscow: Fizmatlit, 2010.
20. Kureychik V.V., Glushchenko A.E., Kureychik L.V. Programmy kompleks kombinirovannogo poiska dlya resheniya zadachi trekhmernoy upakovki [Combined search software package for solving the problem of three-dimensional packaging] *Tr. II Vserossiyskoy nauchno-tehnicheskoy konferentsii Fundamental'nye i prikladnye aspekty komp'yuternykh tekhnologiy i informatsionnoy bezopasnosti* [Proceedings of the II all-Russian scientific and technical conference Fundamental and applied aspects of computer technologies and information security]. Taganrog: Izd-vo YuFU, 2016, pp. 216-220.
21. Zhukov L.A., Korchevskaya O.V. Metod ploskostey: chislennyy eksperiment dlya zadach dvukh i trekhmernoy ortogonal'noy upakovki [Method of planes: numerical experiment for two- and three-dimensional orthogonal packing problems], *Informatsionnye tekhnologii* [Information technologies], 2008, No. 11, pp. 41-45.

Статью рекомендовала к опубликованию д.т.н., профессор Л.С. Лисицына.

Курейчик Владимир Викторович – Южный федеральный университет; e-mail: vkur@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; зав. кафедрой САПР; д.т.н.; профессор.

Глушенко Александр Евгеньевич – Таганрогская таможня; e-mail: alex-14-93@mail.ru; 347939, г. Таганрог, ул. Чучева, д. 46-2, кв. 29; тел.: 88634371651; Уполномоченный по особо важным делам отдела административных расследований.

Kureichik Vladimir Victorovich – Southern Federal University; e-mail: vkur@sfedu.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; head of CAD department; dr. of eng. sc.; professor.

Glushchenko Alexander Evgenyevich – Taganrog customs; e-mail: alex-14-93@mail.ru; 46-2, Chucheva st., apt. 29th, Taganrog, 347939, Russia; phone: +78634371651; Commissioner for Special Cases, Division of Administrative Investigations.

УДК 004.89

DOI 10.18522/2311-3103-2020-2-16-28

Д.Ю. Кравченко, Ю.А. Кравченко, В.В. Марков

ГИБРИДНЫЙ БИОИНСПИРИРОВАННЫЙ АЛГОРИТМ ОТОБРАЖЕНИЯ ОНТОЛОГИЙ В ЗАДАЧАХ ИЗВЛЕЧЕНИЯ И УПРАВЛЕНИЯ ЗНАНИЯМИ*

Статья посвящена решению задачи отображения онтологических моделей в процессах извлечения и управления знаниями. Актуальность и значимость данной задачи обусловлены необходимостью сохранения достоверности и исключения избыточности знаний при интеграции (объединении) структурированных информационных источников различного происхождения. Близость и непротиворечивость понятийной семантики объединенного ресурса при проводимом отображении является основным критерием эффективности предложенных решений. В статье рассмотрены проблемы выбора соответствующих задаче подходов решения, сохраняющих семантику при отображении концептов. Обоснована стратегия выбора биоинспирированного моделирования. Проанализированы аспекты эффективности применения различных децентрализованных биоинспирированных методов. Определены причины необходимости проведения гибридизации. Предложено решать задачу отображения онтологических моделей с применением биоинспирированного алгоритма, построенного на основе гибридизации оптимизационных механизмов алгоритмов бактериального и кукушкиного поиска. Проведенная гибридизация данных алгоритмов позволила объединить их основные преимущества: последовательный бактериальный поиск, обеспе-

* Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов №№ 19-07-00099 и 18-07-00050.

чивающий детальное исследование локальных областей, и значительное число глобальных перемещений агента-кукушки при реализации полетов Леви. Для оценки эффективности предложенного гибридного биоинспирированного алгоритма разработан программный продукт и проведены эксперименты по отображению онтологий разного размера. Каждый концепт любой онтологии имеет определенный набор атрибутов, являющийся семантическим вектором признаков. Степень сходства семантических векторов сравниваемых концептов отображаемых онтологий является критерием их интеграции. Для повышения качества процесса отображения введена новая кодировка решений. Полученные количественные оценки демонстрируют экономию времени при решении задач относительно большой размерности (от 500000 вершин онтографа) не менее 13 %. Временная сложность разработанного гибридного алгоритма составляет $O(n^2)$. Описанные исследования имеют высокий уровень теоретической и практической значимости и напрямую связаны с решением классических задач искусственного интеллекта, направленных на поиск скрытых зависимостей и закономерностей на множестве элементов знаний.

Онтология; гибридные алгоритмы; роевые методы; семантический вектор; биоинспирированный поиск; оптимизация; интеллектуальные агенты.

D.Yu. Kravchenko, Yu.A. Kravchenko, V.V. Markov

HYBRID BIOINSPIRED ALGORITHM FOR ONTOLOGIES MAPPING IN THE TASKS OF EXTRACTION AND KNOWLEDGE MANAGEMENT

The article is devoted to solving the problem of mapping ontological models in the processes of extracting and knowledge management. The relevance and significance of this task are due to the need to maintain reliability and eliminate redundancy of knowledge during the integration (unification) of various origins structured information sources. The proximity and consistency of the conceptual semantics of the combined resource during the mapping is the main criterion for the effectiveness of the proposed solutions. The article considers the problems of choosing appropriate solution approaches that preserve semantics when displaying concepts. The strategy of choosing bio-inspired modeling is substantiated. The aspects of the effectiveness of various decentralized bio-inspired methods are analyzed. The reasons for the need for hybridization are identified. The paper proposes to solve the problem of mapping ontological models using a bio-inspired algorithm based on hybridization of bacterial and cuckoo search algorithms optimization mechanisms. The hybridization of these algorithms allowed us to combine their main advantages: a consistent bacterial search that provides a detailed study of local areas, and a significant number of the cuckoo agent during the implementation global movements of Levy flights. To evaluate the effectiveness of the proposed hybrid bio-inspired algorithm, a software product was developed and experiments were performed on the mapping of different sizes ontologies. Each concept of any ontology has a certain set of attributes, which is a semantic vector of attributes. The degree of the semantic vectors similarity for the compared concepts of displayed ontologies is a criterion for their integration. To improve the quality of the display process, a new encoding of solutions has been introduced. The quantitative estimates obtained demonstrate time savings in solving problems of relatively large dimension (from 500,000 ontograph vertices) of at least 13 %. The time complexity of the developed hybrid algorithm is $O(n^2)$. The described studies have a high level of theoretical and practical significance and are directly related to the solution of classical problems of artificial intelligence aimed at finding hidden dependencies and patterns on a multitude of knowledge elements.

Ontology; hybrid algorithms; swarm methods; semantic vector; bio-inspired search; optimization; intelligent agents.

Введение. Проблема эффективного отображения онтологических структур имеет прямое отношение к решению следующих классических задач искусственного интеллекта: классификации, кластеризации, структурирования, интеграции и семантического поиска информации и знаний. Под отображением будем понимать процесс установления соответствий между понятиями (концептами) нескольких онтологических структур. Данная проблема является оптимизационной. Основным

критерием отображения концептов онтологий установим семантическую близость подразумеваемой понятийной семантики [1–5]. При соблюдении требований использования хорошо структурированных онтологий, точно и однозначно интерпретирующих семантику контекста ресурса, требуется создание формализованных интеллектуальных методов и алгоритмов отображения, позволяющих при сохранении высокого уровня обоснованности имеющихся системно значимых отношений обнаружить обоснованные скрытые зависимости и закономерности на множестве элементов информации и знаний. Подобные интеллектуальные методы и алгоритмы обладают способностью нахождения определенной эквивалентности в понятийной семантике, независимо от сходств и различий в имеющейся структуре отображаемых онтологий.

По мнению авторов, в наибольшей мере такими качествами обладают роевые популяционные биоинспирированные алгоритмы с децентрализованной схемой управления, что создает предпосылки для беспрецедентного параллелизма в проводимых вычислительных процедурах. В качестве основной сложности при решении проблемы отображения онтологий следует отметить наличие в большинстве рассматриваемых случаев области поиска большого объема, что требует одновременно с интенсификацией поиска в локальных фрагментах информационного пространства, также обеспечивать достаточное количество глобальных перемещений интеллектуальных агентов для обеспечения выхода из локальных оптимумов. Большинство известных биоинспирированных алгоритмов обладают только одной хорошо развитой особенностью из двух перечисленных [6–10]. Устранением данного недостатка является решение задачи гибридизации механизмов поиска различных роевых алгоритмов. В представленной работе для гибридизации выбраны алгоритмы бактериальной оптимизации (Bacterial Foraging Optimization, BFO) [11], предложенного Пассино (Passino) в 2002 году и кукушкиного поиска (Cuckoo Search, CS) [12], разработанного Янгом (Yang) и Дебом (Deb) в 2009 году. Для повышения эффективности получаемых решений оба алгоритма были модифицированы в соответствии с требованиями решаемой проблемы. Модифицированный бактериальный алгоритм предлагает реализацию наиболее последовательного поиска с детальным исследованием локальных областей, а модифицированный алгоритм, инспирированный поведением кукушек в процессе гнездового паразитизма, имеет достаточную вероятность глобальных перемещений.

1. Методы отображения онтологий. Рассмотрим основные методы отображения онтологических структур. Выбор среди данных методов производится не из соображений эффективного поиска совпадений в именах или структурных схем понятий, основными критериями является семантическая близость как мера сходства и непротиворечивость с точки зрения контекста.

Одним из таких методов является *отношение уточнения спецификаций сущностей онтологий*. Метод использует формальный критерий корректности отображения спецификаций онтологических концептов, соответствующих абстрактным типам данных. Установленное на основе данного критерия отношение уточнения спецификаций указывает на то, что уточняющая спецификация без ограничений применяется вместо уточняемой без потери качества имеющихся связей. Доказательство уточнения спецификаций поддерживают автоматический и интерактивный режимы в зависимости от сложности информационной модели.

Отметим, что одним из наиболее эффективных вариантов реализации процедуры уточнения спецификаций служит отношение поглощения, которое устанавливается на экстенционалах концептов [13, 14]. Это означает, что все экземпляры класса поглощаемого концепта являются одновременно экземплярами поглощающего. Учитывая тенденции развития онтологических моделей в настоящее время, в

большинстве выстраиваемых рассуждений достаточно применения процедуры поглощения. Отношение уточнения формируется между концептами как при построении одной онтологии, так и при интеграции (отображении) разных. Данная процедура обеспечивает корректность отображения концептов одной онтологии в другую, так как спецификации онтологических концептов отражают и сохраняют понятийную семантику и однозначность контекста.

Помимо формальных методов отображения онтологий значительной популярностью обладают методы, определяющие эквивалентность и семантические различия концептов на основании метаинформации. Реализацией такого метода является применение *метаонтологии*, связанной с обеими интегрируемыми онтологиями [1–5, 13, 14]. В случае, когда отображаемые онтологии построены на основе разных онтологических структур, применение обобщенной метамодели является наиболее эффективным. Метаонтология более абстрактна и содержит в себе большинство разновидностей понятий, составляющих исследуемую предметную область. Метаонтология используется в качестве подложки для отображаемых онтологий, что позволяет унифицировать результирующую структуру. Принцип создания данной подложки следующий. Каждый концепт отображаемых онтологий вместе с предикатами и отношениями становится экземпляром некоторого концепта метаонтологии. При отсутствии семантически близких концептов создаются так называемые служебные концепты как подпонятия метаонтологии, задающие необходимую семантику в терминах концептов метаонтологии. В итоге элементы спецификаций отображаемых онтологий будут распределены по классам, определяемым концептами метаонтологии с учетом семантики и контекста ресурса.

Предлагаемый принцип создания подложки на метаонтологии позволяет распределить спецификации по разным уровням иерархии классов, что делает их независимыми друг от друга. Сформированные в терминах метаонтологии спецификации используются:

- ◆ для проверки корректности отображения концептов;
- ◆ при семантическом поиске релевантных концептов.

При использовании данного метода поглощающий и поглощаемый концепты должны быть в одном классе, соответствующем концепту метаонтологии или служебному концепту. Также возможно в качестве подложки использовать общую онтологию верхнего уровня, что является более сложной задачей, так как требуется интеграция каждой онтологии в общую, что приводит к изменению начальных версий онтологических моделей при совместимости их концептов с онтологией верхнего уровня.

Оценка фундаментальных метасвойств лежит в основе следующего метода отображения онтологий. Любая онтология включает в себя множество свойств и отношений между концептами. Набор фундаментальных метасвойств позволяет оценить концепты или отношения по следующим критериям:

- 1) существенность – неотъемлемость свойства сущности;
- 2) строгость – принадлежность существенного свойства сущности в любом контексте или предметной области;
- 3) идентификация – является ли утверждение об объекте идентифицирующим его свойством;
- 4) собственная идентификация – несет ли эту идентификацию само утверждение, либо оно наследует ее из других свойств или сущностей;
- 5) неизменность – может ли свойство меняться во времени;
- 6) зависимость – может ли сущность существовать без других;
- 7) постоянство – как долго сущность остается таковой;
- 8) объединение – существование экземпляров контекста как целых сущностей в отношении часть-целое.

Обратим внимание на то, что формальное определение перечисленных метасвойств требует введения некоторых ограничений. Одни метасвойства совместимы, другие обладают признаком ортогональности и исключают друг друга. Для свойства W , поглощающего свойство V , верны следующие ограничения:

- ◆ если W строгое для любых сущностей, то V также строгое для любых сущностей;
- ◆ если W несет критерий идентификации сущностей, то и V также;
- ◆ если W несет критерий объединения, то и V также;
- ◆ если W не несет объединение, то и V также;
- ◆ всякая сущность должна быть значением наиболее общего свойства, несущего его идентификацию.

Поясним, что при построении последовательности поглощений между концептами существуют случаи наследования метасвойств более специфичных концептов, поэтому наличие конфликтных ситуаций в данной иерархии приведет к некорректному отображению. Описанные ограничения также выполняются для отношений, обнаруженных между концептами различных онтологий, что позволяет выявлять семантические конфликты, полученные при отображении онтологий. На присутствие конфликта указывает противоречие метасвойств связанных концептов, в этом случае отображение принято считать некорректным. Исследование метасвойств дает возможность строить более сложные рассуждения, которые необходимы для повышения эффективности решения задачи отображения онтологий.

В иерархии свойств концептов учитывают систему классов, в которой родовые концепты идентифицируют объекты реального мира, создавая при этом типы, видовые – описывают категории объектов, ролевые – распределяют роли. Причем, ролевые концепты входят в подкласс родовых, а видовые могут быть подклассами как ролевых, так и родовых концептов [1–5, 13, 14]. В метасвойствах отношений любая часть является неотъемлемой, что позволяет однозначно идентифицировать целое по части, а в случае инвариантности целого, части идентифицируется по нему, все эти особенности применяются для корректного отображения онтологий.

Описанный метод имеет связь с предыдущим, использующим общую онтологию верхнего уровня для проведения процедуры отображения. Фундаментальные концепты общей онтологии имеют вполне определенные наборы значений метасвойств. Совместное применение указанных методов дает дополнительные возможности для повышения эффективности решения поставленной задачи.

В заключении проведенного в данном пункте обзора рассмотрим метод отображения онтологий на основе экземпляров экстенционалов понятий. В данном методе релевантность определяется на основе принадлежности экземпляров одним и тем же концептам. В качестве экземпляров рассматриваются: объекты-сущности реального мира; модели реального мира; классифицированные в онтологии данные. Проверка корректности обнаруженных отношений проводится с применением экстенциональной составляющей связанных концептов онтологий. Если существует хотя бы один пример экземпляра, в котором сущности не принадлежат одновременно классам эквивалентным концептам отображаемых онтологий, тогда возникает конфликт, ставящий под сомнение корректность обнаруженной связи между концептами. Описанный экстенциональный метод проверки отображения онтологий «по образцу» реализуется формальным образом.

Реализация всех вышеперечисленных методов возможна только при наличии эффективных автоматических средств оценки семантической близости концептов, как меры сходства, позволяющей утверждать об эквивалентности рассматриваемых понятий и отношений. Создание таких средств на основе алгоритмов биоинспирированного поиска с децентрализованной схемой управления является основной целью данной работы.

2. Постановка задачи. Используем следующую модель онтологической структуры $O = \langle P, V, R, C \rangle$, где P – множество концептов; V – множество экземпляров концептов; R – множество предикатов; C – множество значимых отношений. Для проведения процедуры отображения построим две онтологии O_1 и O_2 . $O_1 = \langle P^1, V^1, R^1, C^1 \rangle$ и $O_2 = \langle P^2, V^2, R^2, C^2 \rangle$. Примем равным n количество концептов в онтологии O_1 , и равным m – в онтологии O_2 . Для оценки релевантности концептов применим эквивалентный тип семантической близости [1–5], в которой $mapping(P^1) = P^2$, if $sim(P^1, P^2) \geq d$, где d – пороговое значение сходства предикатов, достижения которого достаточно для отображения концепта P^1 в онтологию O_2 . Совпадение не ниже порогового значения d предикатов понятия P_i^1 онтологии O_1 с предикатами понятия P_j^2 онтологии O_2 указывает на эквивалентность множеств R^1, R^2 .

Итоговая онтологическая структура, полученная после отображения онтологии O_1 в онтологию O_2 , примет следующий вид:

$$Z = \langle O, H^O, S, U \rangle, \quad (1)$$

где $O = \langle P, V, R, C \rangle$ – обобщенная онтология; H^O – интегрированная интеллектуальная информационная система с онтологией O ; S – множество значений эквивалентной семантической близости; U – изменение информационной системы в виде отображения.

Целевой функцией поставленной задачи является следующее выражение:

$$f(R^1, R^2) \rightarrow \max, \quad (2)$$

где R^1, R^2 – множества предикатов концептов онтологий O_1 и O_2 . Каждое множество предикатов состоит из подмножеств конкретных атрибутов, принадлежащих каждому из понятий двух онтологий.

$$R^1 = \{R_1^1; R_2^1; \dots; R_i^1; \dots; R_n^1\}, \quad (3)$$

где $i = \overline{1, n}$, n – количество концептов в O_1 .

$$R^2 = \{R_1^2; R_2^2; \dots; R_j^2; \dots; R_m^2\}, \quad (4)$$

где $j = \overline{1, m}$, m – количество концептов в O_2 .

Решением задачи оценки эквивалентной семантической близости является следующее выражение:

$$f(R^1, R^2) = \sum_{i=1}^n \sum_{j=1}^m R_i^1 \cap R_j^2 \geq d, \quad (5)$$

где $d > 0$, d – пороговое значение, равенство или превышение которого означает присутствие эквивалентной семантической близости между подмножествами предикатов, принадлежащих концептам P_i^1 и P_j^2 . Чем выше показатель d , тем более значимой является эквивалентная семантическая близость между рассматриваемыми концептами.

Таким образом, при решении задачи отображения большего числа онтологий требуется найти такую пару онтологий O_a и O_b , для которых функция f примет наибольшее значение.

2. Разработка гибридного биоинспирированного алгоритма отображения онтологий. Оценка эквивалентной семантической близости является трансвычислительной задачей, так как в общем случае потребуются полный перебор для проверки всех оценок попарных сравнений предикатов концептов отображаемых онтологий. Вычислительная сложность задачи требует применения методов искусственного интеллекта.

В работе задача отображения онтологий решена с применением нового биоинспирированного алгоритма, созданного на основе гибридизации оптимизационных механизмов алгоритмов бактериального и кукушкиного поиска. Первый алго-

ритм позволяет интенсифицировать поиск в локальных областях, а второй дает возможность реализовывать глобальные перемещения интеллектуальных агентов для выхода из локальных оптимумов.

Отметим, что поставленная задача решается одновременно в двух информационных пространствах, первой и второй онтологий, что использовано авторами для дифференцирования процедур локальных и глобальных перемещений. Для этого каждая онтология получает определенный статус. Для верификации поиска предусмотрена возможность смены статусов онтологий. При первом запуске процедуры отображения условно будем считать онтологию с номером 1 – информационным пространством для реализации глобальных перемещений агентов на основе реализации полетов кукушки по закону Леви, а онтологию под номером 2 – информационным пространством для осуществления последовательного поиска на основе передвижений агентов-бактерий. Правило смены статуса онтологий описан ниже.

Популяционное децентрализованное поведение бактерий стало основой канонического бактериального алгоритма поисковой оптимизации (Bacterial Foraging Optimization, BFO), предложенного Пассино (Passino) в 2002 году [11, 15–17]. В представленной работе модификация данного метода коснулась структуры используемой области поиска, разделенной на две части с изменяемыми статусами в соответствии с онтографами двух отображаемых онтологий, а также привела к реализации нового гибридного асимметричного механизма перемещений популяции агентов, основанного на чередовании локальных последовательных передвижений агентов-бактерий с глобальными перемещениями агентов-кукушек. Алгоритм поиска кукушки (Cuckoo Search, CS) был разработан Янгом (Yang) и Дебом (Deb) в 2009 году [12, 15–17]. Основное достоинство алгоритма – незначительное количество свободных параметров, что делает метод более универсальным и эффективным по сравнению с другими роевыми методами оптимизации. Отличием использованного в работе модифицированного метода поиска кукушки от канонического является применение одного онтографа в качестве статичного пространства популяции подкладываемых яиц кукушки, а второго – динамичного пространства перемещений популяции агентов кукушек по чужим гнездам в поиске сходства подкладываемого яйца с оригиналом. Данная особенность также реализована с помощью периодических смен статуса онтологий во время поиска.

Связи между вершинами онтографов зададим множествами отношений C^1 и C^2 онтологий O_1 и O_2 соответственно. Вершины онтографов содержат значения предикатов концептов, представленных в виде подмножеств $R_1^1; R_2^1; \dots; R_i^1; \dots; R_N^1$ для онтологии O_1 , и подмножества $R_1^2; R_2^2; \dots; R_j^2; \dots; R_M^2$ для онтологии O_2 . Превышение порогового значения d показателем семантической близости между предикатами концептов P_i^1 и P_j^2 указывает на их эквивалентность [18–20]. Пара сравниваемых концептов с номерами a и b , где $a \in [1: N]$, $b \in [1: M]$ для каждого интеллектуального агента выбирается случайным образом на каждой итерации работы алгоритма.

На начальном этапе работы алгоритма подмножество предикатов концептов 1-ой онтологии будет статичным, а сама 1-ая онтология будет иметь статус константы, напротив, подмножество предикатов концептов 2-ой онтологии станет динамичной составляющей поиска, а сама 2-ая онтология станет координатным пространством для передвижений агентов (рис. 1). Динамика поиска во 2-ой онтологии реализована на основе бактериальной оптимизации, которая продолжается пока удастся получить прирост фитнес-функции. В случае отсутствия прогресса в бактериальном поиске происходит смена статусов онтологий. Текущие вершины-координаты бактерий во 2-ой онтологии назначаются константами, а поиск экви-

валентных им концептов возобновляется в пространстве 1-ой онтологии на основе перелетов агентов-кукушек. Перелеты продолжаются до увеличения значения фитнес-функции, либо до достижения их максимального количества, после чего поиск в 1-ой онтологии продолжается на основе механизма бактериальной оптимизации, который применяется до следующей смены статуса онтологий по уже описанному выше сценарию.

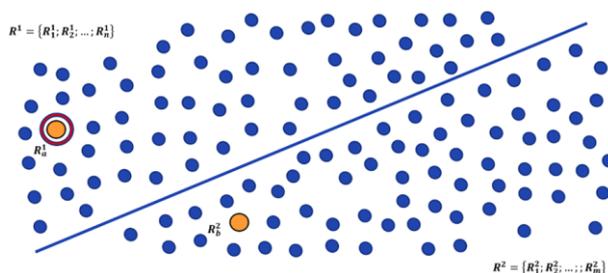


Рис. 1. Модель разделенного на подмножества пространства решений

Свободными параметрами алгоритма являются T, T^r, T^l – общие числа шагов хемотаксиса (перелетов), репродукции и ликвидации соответственно, а $|S|$ – четное число агентов в популяции. Координаты при локальном поиске $X_{i,r,l} = X_{i,r,l}(t)$ бактерии $s_i \in S$ на t -м шаге хемотаксиса, r -м шаге репродукции и l -м шаге ликвидации определяются направлением вектора «плавания» $V_i(\epsilon)$, где ϵ – мощность окрестности целевой вершины. Этот параметр отличает данный механизм поиска при бактериальной оптимизации от канонического и позволяет повысить вероятность обнаружения траектории передвижения, дающей наибольший прирост фитнес-функции, благодаря выбору по ходу движения бактерии вершины с наибольшей мощностью окрестности (рис. 2).

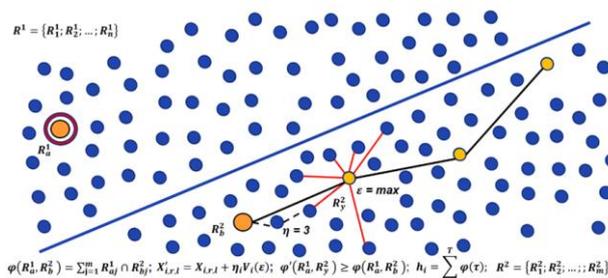


Рис. 2. Передвижение бактерии с учетом мощности окрестности целевой вершины

Выражение фитнес-функции имеет следующий канонический вид:

$$\varphi_{i,r,l} = \varphi_{i,r,l}(T). \quad (6)$$

Новые координаты $X'_{i,r,l}$ агента-бактерии вычисляются по следующему выражению:

$$X'_{i,r,l} = X_{i,r,l} + \eta_i V_i(\epsilon), \quad (7)$$

где $\eta \geq 1$, задает величину шага передвижения бактерии (свободный параметр алгоритма). При увеличении значений фитнес-функции направление вектора «плавания» остается неизменным $V'_i(\epsilon) = V_i(\epsilon)$. При уменьшении значений фитнес-

функции агент-бактерия совершает «кувырок», возвращается в предыдущую вершину и продолжает поиск, кардинальным образом меняя направление вектора своего движения (рис. 3).

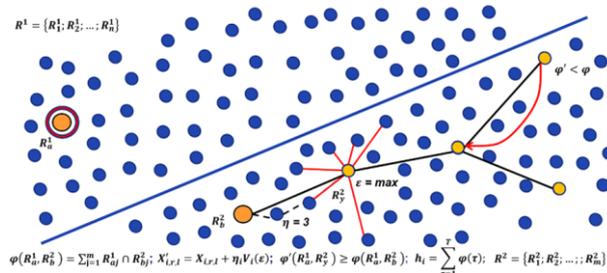


Рис. 3. «Кувырок» бактерии в условиях снижения значений фитнес-функции

Интенсификация процесса локального поиска достигается применением процедуры репродукции, которая позволяет на заданном шаге r воспроизвести копии наиболее удачных агентов бактерий, составляющих половину всей популяции, уничтожив при этом вторую половину безуспешных. Решение о репродукции основано на высоком показателе здоровья бактерии, вычисляемом по следующему выражению:

$$h_i = \sum_{\tau=1}^T \varphi_{i,r,l}(\tau), i \in [1: |S|]. \quad (8)$$

Например, если успешная бактерия $s_j, j \in [1: |S|]$ имеет координаты $X_{j,r,l}$, тогда после процедуры репродукции возникнет бактерия s_k , причем, $k = \frac{|S|}{2} + j$, $X_{j,r+1,l} = X_{j,r,l}$, $X_{k,r+1,l} = X_{j,r,l}$. При этом, общее количество бактерий в колонии останется неизменным.

Процедура репродукции совершается многократно. Максимальное количество репродукций задается свободным параметром алгоритма Z . Достижение данного значения запускает процедуру ликвидации, при которой случайное число агентов $w < |S|$ (свободный параметр алгоритма) уничтожается, а вместо них в случайных координатах возникает такое же количество новых агентов, что позволяет алгоритму выйти из локальных оптимумов.

В случае отсутствия прогресса в поиске какого-либо агента при достижении или превышении им показателя половины шагов до репродукции $r \geq T/2$, статус онтологий меняется (рис. 4).

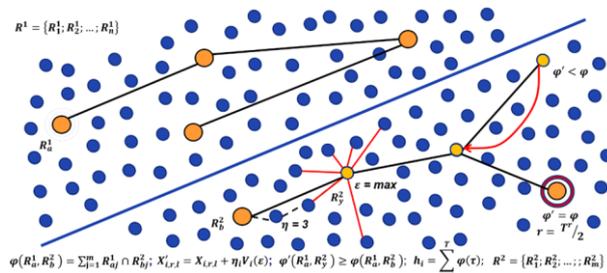


Рис. 4. Смена статусов онтологий, организация «перелетов»

Текущая координата бактерии замораживается и становится константой, процедура поиска переводится в информационное пространство другой онтологии, где в последующем производятся процедуры «перелетов» кукушки (рис. 5), что

обеспечивает глобальные перемещения. При обнаружении вершин дающих прирост фитнес-функции, «перелеты» прекращаются, агент-кукушка заменяется на агента-бактерию, и поиск по заданным выше правилам возобновляется в пространстве онтологии, получившей новый статус динамичной. Данный процесс продолжается до очередной смены статуса.

Обобщенная схема гибридного биоинспирированного алгоритма отображения онтологий, разработанного на основе алгоритмов бактериального и кукушкиного поиска, представлена на рис. 5. Основные модифицированные процедуры созданного гибридного алгоритма выделены на рис. 5 фоном.

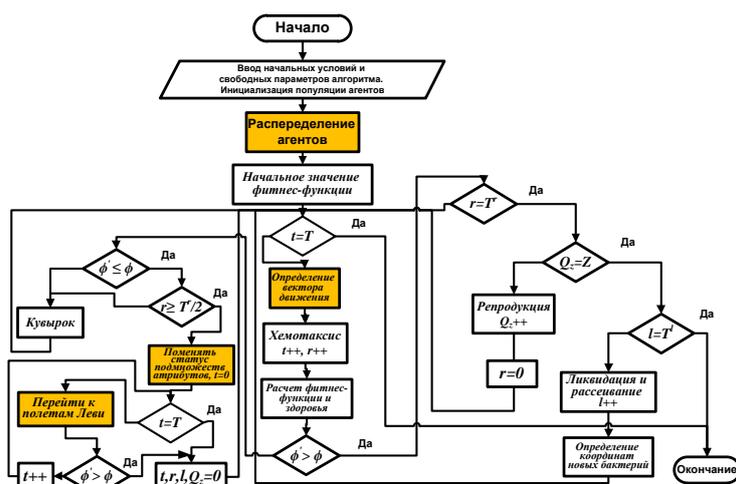


Рис. 5. Обобщенная схема гибридного биоинспирированного алгоритма отображения онтологий

Алгоритм, инспирированный поведением кукушек в процессе гнездового паразитизма, отличается от известных аналогов более эффективным способом определения траектории передвижения агента в пространстве поиска. Кукушка свои передвижения осуществляет в виде полетов Леви (Levy flights), реализуемых по следующей канонической формуле (9) с тензорным произведением вектора размера шагов и случайного вектора независимых вещественных чисел.

$$X' = X + U \otimes L_{|X|}(\lambda), \quad (9)$$

где $U = (u_a, a \in [1: |X|])$ – вектор размера шагов по соответствующим компонентам вектора координат местонахождения агента X ; $L_{|X|}(\lambda) = (|X| \times 1)$ – вектор независимых вещественных случайных чисел, распределенных по закону Леви [12, 21–25]. Данный способ обеспечивает высокую вероятность малых и достаточную вероятность больших перелетов агента-кукушки, что позволяет интенсифицировать поиск в локальных областях, с одной стороны, и реализовать выход из локальных оптимумов, с другой.

3. Экспериментальные исследования. Для подтверждения эффективности разработанного гибридного биоинспирированного алгоритма отображения онтологий был создан программный модуль, позволяющий сравнить результаты решения поставленной задачи, полученные с применением предложенного и канонических роевых алгоритмов. В качестве значений предикатов применены усредненные абстрактные показатели, полученные случайным образом с заданной вероятностью.

В результате проведенных комплексных исследований получена временная сложность (ВСА) предложенного гибридного алгоритма, которая составила $O(n^2)$, где n – количество анализируемых алгоритмом входных данных, которым в данном случае является число вершин онтографов (рис. 6). В представленном биоинспирированном алгоритме, построенном на основе гибридизации оптимизационных механизмов алгоритмов бактериального и кукушкиного поиска, предусмотрено два вложенных цикла прохождения по множествам предикатов концептов: первый, для реализации локального поиска, а второй, для осуществления глобальных перемещений. Таким образом, максимально возможное количество операций зависит от числа вершин онтографов N как $n * n$.

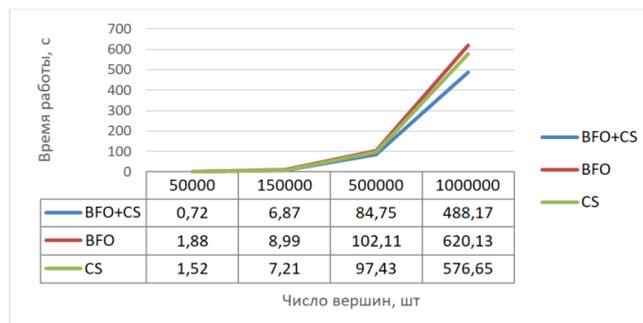


Рис. 6. Результаты сравнения скорости работы канонических алгоритмов бактериальной оптимизации (BFO) и кукушкиного поиска (CS) с предложенным гибридным алгоритмом (BFO + CS)

Экспериментальные исследования разработанного метода проводились на вычислительной системе с производительностью порядка 10^9 (гигафлопс) операций в секунду. В этом случае, например, при $n = 150000$ вершин, время выполнения алгоритма не превысило 7-ми секунд, при $n = 500000$ вершин – 90 секунд, при $n = 1000000$ вершин – 490 секунд, что наглядно иллюстрирует представленная на рисунке 6 полученная зависимость времени выполнения предложенного метода от числа входных вершин. Исследование скорости работы предложенного гибридного алгоритма биоинспирированной оптимизации для решения задачи отображения онтологий проводилось в сравнении с каноническими бактериальным и кукушкиным алгоритмами.

Полученные количественные оценки сравнения скорости работы канонических алгоритмов бактериальной оптимизации (BFO) и кукушкиного поиска (CS) с предложенным гибридным алгоритмом (BFO + CS) показывают, что разработанный гибридный алгоритм дает лучшие результаты по сравнению с каноническими. Отметим, что при относительно небольших размерностях задачи (от 50000 до 150000 вершин) время работы оцениваемых алгоритмов соизмеримо: от 0,72–6,87 секунды в предлагаемом гибридном алгоритме до 1,88–8,99 и 1,52–7,21 секунды соответственно в бактериальном и кукушкином алгоритмах по отдельности. При количестве вершин от 500000 до 1000000 разница в показателях времени работы становится более значительной. В целом, на указанных относительно больших размерностях задачи, алгоритм кукушкиного поиска быстрее бактериального алгоритма на 6–7 %, а разработанный гибридный алгоритм быстрее кукушкиного на 13–15 %. Эффективность предложенного гибридного алгоритма повышена за счет улучшения способа определения траектории передвижения агента в пространстве поиска, позволяющего обеспечить максимальное количество итераций локального поиска при достаточном количестве – глобального.

Заключение. В данной работе представлена разработка гибридного биоинспирированного алгоритма отображения онтологий. Предложено решать задачу отображения онтологических моделей с применением биоинспирированного алгоритма, построенного на основе гибридизации оптимизационных механизмов алгоритмов бактериального и кукушкиного поиска. Проведенная гибридизация данных алгоритмов позволила объединить их основные преимущества: последовательный бактериальный поиск, обеспечивающий детальное исследование локальных областей, и значительное число глобальных перемещений агента-кукушки, позволяющих алгоритму выходить из «локальных ям».

Для оценки эффективности предложенного метода разработан программный продукт, включающий в себя функции построения онтологий, генерации набора предикатов концептов и проведения оценки эквивалентной семантической близости. Проведенные экспериментальные исследования подтвердили эффективность разработки в сравнении с каноническими биоинспирированными алгоритмами бактериальной оптимизации и кукушкиного поиска. Полученные количественные оценки подтверждают снижение времени решения задачи отображения онтологий при помощи разработанного гибридного алгоритма (BFO + CS) не менее чем на 13 % по отношению в результатам работы канонических алгоритмов бактериальной оптимизации (BFO) и кукушкиного поиска (CS). При относительно небольших размерностях задачи (от 50000 до 150000 вершин в отображаемых онтологиях) время работы всех алгоритмов соизмеримо, наибольший прирост скорости решения достигается в задачах с размерностью не менее 500000 вершин. Временная сложность разработанного гибридного алгоритма составляет $O(n^2)$, где n – количество анализируемых методом входных данных, которым в данном случае является число вершин онтографов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Andreasen T., Knappe R., Bulskov H.* Domain specific similarity and retrieval, *11th Int. Fuzzy Systems Association World Congress*, 2016, Vol. 1, pp. 496-502.
2. *Castano S., Ferrara A., Montanelli S., Racca G.* Semantic information interoperability in open networked systems, *Proc. of the Int. Conf. SNW*. Paris, 2004, pp. 215-230.
3. *Kravchenko Yu.A., Kursitys I.O., Markov V.V.* Bioinspired Algorithm for Acquiring New Knowledge based on Information Resource Classification, *2019 International Russian Automation Conference (RusAutoCon)*.
4. *Haase P., Siebes R., Harmelen F.* Peer selection in peerto-peer networks with semantic topologies, *Proc. of Int. Conf. on Semantics in a Networked World*. Paris, 2004, pp. 108125.
5. *Maedche A., Zacharias V.* Clustering ontology-based metadata in the Semantic Web, *Proc. 6th European PKDD Conf. LNCS*. Berlin: Springer, 2002, Vol. 2431, pp. 348-360.
6. *Abraham A., Grosan G., Ramos V.* Swarm Intelligence in Data Mining. Berlin. Heidelberg: SpringerVerlag, 2006, 267 p.
7. *Sousa T., Silva A., Neves A.* Particle Swarm based Data Mining Algorithms for classification tasks, *Parallel Computing*, 2004, Vol. 30, Issue 5–6, pp. 767-783.
8. *Parsopoulos K.E., Vrahatis M.N.* Recent Approaches to Global Optimization Problems Through Particle Swarm Optimization, *Natural Computing*, 2002, No. 1 (2–3), pp. 235-306.
9. *Kravchenko Yu.A., Kravchenko D.Y., Kursitys I.O.* Architecture and method of integrating information and knowledge on the basis of the ontological structure, *Advances in Intelligent Systems and Computing. 1st International Conference of Artificial Intelligence, Medical Engineering, and Education, AIMEE 2017*. Moscow, 2018, Vol. 658, pp. 93-103.
10. *Kravchenko Yu.A., Kureichik V.V.* Knowledge management based on multi-agent simulation in informational systems, *Conference proceedings. 8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014". – 15-17 October 2014, Astana, Kazakhstan*, pp. 264-267.
11. *Arijit S.D., Sambarta B., Abraham D.A.* Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications, *Foundations of Computational Intelligence*. Springer, 2009, Vol. 3, pp. 23-55.

12. Payne R.B., Sorenson M.D., and Klitz K. The Cuckoos. Oxford University Press, 2005.
13. Kravchenko Yu.A., Kuliev E.V., Kulieva N.V., Kureichik V.V. Problem-oriented knowledge processing on the basis of hybrid approach, *Information technologies in science. Management, Social Sphere and Medicine (ITSMSSM 2016)*, pp. 510-513.
14. Martens D., De Backer M., Haesen R., Vanthienen J., Snoeck M. and Baesens B. Classification With Ant Colony Optimization, *IEEE Transactions on Evolutionary Computation*, 2007, Vol. 11, No. 5, pp. 651-665.
15. Falco I.D., Cioppa A.D., and Tarantino E. Evaluation of particle swarm optimization effectiveness in classification, *LNAI3849*, 2006, pp. 164-171.
16. Soliman O.S. and Adly A. Bio-inspired algorithm for classification association rules, *8th International Conference on Informatics and Systems (INFOS), Cairo, 2012*, pp. 154-160.
17. Bova V., Zaporozhets D., and Kureichik V. Integration and processing of problem-oriented knowledge based on evolutionary procedures, *Advances in Intelligent Systems and Computing*, 2016, Vol. 450, pp. 239-249.
18. Semenova A.V. and Kureichik V.M. Ensemble of classifiers for ontology enrichment, *Journal of Physics: Conference Series*, 2018, Vol. 1015, Issue 3, article id. 032123.
19. Kureychik V.M. Overview and problem state of ontology models development, *9th International Conference on Application of Information and Communication Technologies, AICT 2015 - Proceedings 9, 2015*, pp. 558-564.
20. Semenova A.V. and Kureychik V.M. Application of swarm intelligence for domain ontology alignment, *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (ITI'16)*, 2016, Vol. 1, pp. 261-270.
21. Bova V., Kureichik V. and Zaruba D. Heuristic approach to model of corporate knowledge construction in information and analytical systems, *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, 2016*, pp. 1-5.
22. Kureichik V., Zaporozhets D., and Zaruba D. Generation of bioinspired search procedures for optimization problems, *Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings*, 2016, Vol. 10.
23. Kar A.K. Bio inspired computing – A review of algorithms and scope of applications, *Expert Systems with Applications*, 2016, Vol. 59, pp. 20-32.
24. Zaporozhets D., Zaruba D., and Kulieva N. Parallel approach for bioinspired algorithms, *Journal of Physics: Conference Series Ser. "International Conference Information Technologies in Business and Industry 2018 - Enterprise Information Systems"*, 2018.
25. Bova V.V., Nuzhnov E.V., Kureichik V.V. The combined method of semantic similarity estimation of problem oriented knowledge on the basis of evolutionary procedures, *Advances in Intelligent Systems and Computing*, 2017, Vol. 573, pp. 74-83.

Статью рекомендовал к опубликованию к.т.н. С.Г. Буланов.

Кравченко Даниил Юрьевич – Южный федеральный университет; e-mail: kravchenkodanil122@gmail.com; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; студент.

Кравченко Юрий Алексеевич – e-mail: yakravchenko@sfedu.ru. кафедра систем автоматизированного проектирования; доцент.

Марков Владимир Васильевич – e-mail: vvmarkov@sfedu.ru; кафедра систем автоматизированного проектирования; доцент.

Kravchenko Daniil Yurievich – Southern Federal University; e-mail: kravchenkodanil122@gmail.com; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; student.

Kravchenko Yury Alekseevich – e-mail: yakravchenko@sfedu.ru; the department of computer aided design; associate professor.

Markov Vladimir Vasilievich – e-mail: vvmarkov@sfedu.ru; the department of computer aided design; associate professor.