

19. *Kharchenko A.M.* Adaptivnyy raschet funktsii dlya dinamicheskogo EM-algoritma [The adaptive function is calculated for the dynamic em algorithm], *Matematika* [Mathematics], 2015, pp. 134.
20. *Kureychik V.M., Kalanchuk S.A.* Obzor i sostoyanie problemy roevykh metodov optimizatsii [Review and state of the problem of swarm optimization methods], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie* [Informatics, computer science and engineering education], 2016, No. 1 (25), pp. 1-13.
21. *Karpenko A.P.* Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: ucheb. posobie [Modern search engine optimization algorithms. Nature-inspired algorithms: textbook]. Moscow: MGTU im. N.E. Bauman, 2014, 446 p.
22. *Yang X.S., Deb S.* Multiobjective cuckoo search for design optimization, *Comput. Oper. Res.*, 2013, No. 40 (6), pp. 1616-1624.
23. *Chifu V.R., Pop C.B., Salomie I., Niculici A.N.* Optimizing the semantic web service composition process using cuckoo search, *Intelligent Distributed Computing*, 2012, No. 5, pp. 93-102.
24. *Coelho L.S., Guerra F.A., Batistela N.J., Leite J.V.* Multiobjective cuckoo search algorithm based on duffings oscillator applied to jiles-atherton vector hysteresis parameters estimation, *IEEE Trans. Magn.*, 2013, No. 49 (5), pp. 1745.
25. Repozitoriy mashinnogo obucheniya. Available at: <http://www.ics.uci.edu/~mllearn/MLRepository.html> (accessed 24 June 2019).

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Гатчин.

Бова Виктория Викторовна – Южный федеральный университет; e-mail: vvbova@yandex.ru; 347928, г. Таганрог, Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

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

Bova Victoria Victorovna – Southern Federal University; e-mail: vvbova@yandex.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

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

УДК 004.6:004.8

DOI 10.23683/2311-3103-2019-4-102-114

Д.В. Балабанов, А.В. Ковтун, Ю.А. Кравченко

БУСТИНГ БИОИНСПИРИРОВАННЫХ АЛГОРИТМОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ИНТЕГРАЦИИ ДАННЫХ*

В настоящее время интеграция данных является актуальной проблемой. Интеграция данных может быть представлена на различных уровнях. Современные методы решения задач интеграции не могут решать семантическую проблему, к тому же они слишком сложны. На основе исследований методов, используемых на данный момент, можно сделать вывод, что самыми часто используемыми методами, для подходов к решению задачи неоднородности на семантическом уровне, используются такие эвристики, которые изменяют результирующую онтологию. В большинстве случаев, данные информационных систем представлены как объекты информации, которые в свою очередь формируют некую предметную область или ее часть, в тоже время к каждой части (области) относится ее собственная онтология. Исходя из этого, при решении задачи семантической неоднородности данных нужно привести определения предметных областей и взаимодействия их объектов. Таким образом можно построить взаимодействие информационных систем,

* Работа выполнена при поддержке РФФИ (проекты: № 19-07-00099, № 18-07-00055).

используя согласованную семантику предметной области. Рассматривается семантический уровень, на котором данные анализируются с точки зрения их семантических свойств, а также в аспекте единой онтологии. Для решения данного класса задач предлагается использовать эволюционные и биоинспирированные алгоритмы. Предлагается использовать бустинг, идея которого состоит в последовательном применении нескольких алгоритмов. В основе процесса интеграции информационных систем лежит задача оценки семантической близости объектов неоднородных онтологий, основанной на гибридной мере подобия, включающей в себя атрибутивную, таксономическую и реляционную. Предложена структура бустинга биоинспирированных алгоритмов для поиска концептов и оценки их семантической близости. Предлагается использовать динамическую вероятность выбора алгоритмов на основе полученных результатов.

Интеграция данных; семантическая близость; онтологии, концепты; неоднородные данные; бустинг; генетические и биоинспирированные алгоритмы.

D.V. Balabanov, A.V. Kovtun, Y.A. Kravchenko

BOOSTING OF BIOINSPIRED ALGORITHMS FOR SOLVING PROBLEM OF DATA INTEGRATION

Currently, the integration of data is an actual problem. Data integration can be presented at various levels. The existing methods for solving the problem of data integration are quite complex and still far from solving the problem of semantics. Analysis of the methods shows that when solving the problem of heterogeneity at the semantic level, methods are used which were based on the use of a single ontology of the upper level. Due to the fact that data in information systems usually represent information objects that model some parts of the domain, which in turn has its own ontology, in order to solve the problem of semantic heterogeneity of data, it is necessary to conform to the concepts of subject domains of interaction objects. In this case, the coordinated semantics of the subject area can be built interaction of information systems. The paper considers the semantic level at which data is analyzed in terms of their semantic properties, as well as in the aspect of a unified ontology. To resolve this class of problems, it is proposed to use evolutionary and, in particular, bioinspired algorithms. The paper proposes to use boosting, the idea of which is to apply several algorithms sequentially. The process of integration of information systems is reduced to solving the problem of estimating the semantic proximity of objects of inhomogeneous ontologies, based on a hybrid measure of similarity, including attribute, taxonomic and relational. The structure of boosting bioinspired algorithms for searching for concepts and evaluating their semantic proximity is proposed. It is proposed to use the dynamic probability of choosing algorithms based on the results obtained.

Data integration; semantic similarity; ontology; concepts; heterogeneous data; boosting; genetic and bioinspired algorithms.

Введение. Разработка методов интеграции информационных ресурсов является одной из наиболее актуальных проблем в сфере информационных систем. Актуальность обусловлена с том числе тем, что источники данных обладают различными свойствами, являющимися существенными при выборе методов интеграции данных, исходя из этого они могут поддерживать представление данных в форматах различных моделей, а так же разделяются на статические, динамические и т.п.

Интеграция данных – процесс компоновки информации из различных информационных систем (ИС), в том числе неоднородных, что имеет большое значение, установки соответствия информации между различными источниками (mapping), синхронизация одинаковых информационных объектов в различных ИС.

Интеграция данных – есть процесс, в котором данные объединяются из различных однородных и разнородных источников, чтобы в дальнейшем получить их в одном виде (представлении), в широком смысле – обмен (синхронизация) данных различных ИС. Существует три основных уровня интеграции данных, таких как физический, логический и семантический уровни [1–5]. Физическая интеграция данных – самая простая из этих трех и определяется как представление раз-

личных данных в одном виде (формате) для дальнейшего их физического представления. В свою очередь для логического уровня интеграции данных необходим доступ к данным из различных разнородных информационных систем, предполагающий общее представление и описание гетерогенных источников, а также берет во внимание как различные поведенческие, так и структурные свойства.

Наконец, семантический уровень рассматривает данные с точки зрения их семантических свойств, а также в аспекте единой онтологии.

Цель представленной работы – повышение эффективности алгоритмов, решающих задачу интеграции данных, и корректности результирующей онтологии на основе развития теории эволюционного и биоинспирированного поиска, а также применения процедур бустинга биоинспирированных алгоритмов, позволяющих устранить недостатки существующих аналогов. Объектом исследования являются информационные системы, предметом – генетические и биоинспирированные алгоритмы, а также процедуры бустинга.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести аналитический обзор основных подходов к решению проблем интеграции данных, проанализировать применимость генетических и биоинспирированных алгоритмов для решения задачи интеграции данных.

2. Построить математическую и онтологическую модели решаемой задачи, описать понятие концепта онтологии.

3. Определить область допустимых значений и привести обобщенную формулу для вычисления целевой функции.

4. Привести схему интеграции данных нескольких информационных систем.

5. Разработать структуру бустинга биоинспирированных алгоритмов для решения задачи интеграции неоднородных информационных систем. Данная структура направлена на определение комплексной оценки семантической близости концептов неоднородных онтологий. Основным преимуществом использования бустинга биоинспирированных алгоритмов для решения задачи интеграции данных являются устранение субъективности, понятий и их описаний в онтологии, а также избавление от зависимости различных экспертных оценок. Бустинг также позволяет выделять ключевые концепты, необходимые для построения результирующей онтологии.

6. Ввести динамическую вероятность применения биоинспирированных алгоритмов, для повышения качества решений;

7. Провести экспериментальные исследования, подтверждающие эффективность предложенных моделей и процедур.

Проведем аналитический обзор состояния исследований по выбранной тематике и дадим постановку задачи.

1. Аналитический обзор и постановка задачи интеграции данных. Источниками данных для интеграции являются традиционные базы данных, а также репозитории, web-сайты, структурированные данные в различных форматах [1, 2]. Принимая во внимание выбранные подходы и требования к системе, создавая систему интеграции, формируется следующая группа задач:

- ◆ создание архитектуры системы интеграции данных;
- ◆ разработка методов отображения моделей данных и построение отображений в интегрирующую модель, для каждой отдельно взятой модели, которые поддерживаются различными источниками данных;
- ◆ метаданные, содержащиеся в источниках, и их интеграция;
- ◆ решение проблемы неоднородности, характерной для различных источников данных;
- ◆ создание механизмов сравнения семантических весов для семантической интеграции.

Все перечисленные задачи имеют значительную сложность с точки зрения применения классических методов интеллектуального анализа данных.

Одним из перспективных направлений для решения данного класса задач является использование эволюционных и биоинспирированных алгоритмов [6–12]. Использование эвристик алгоритмов данного класса позволит получить оптимальные значения параметров и обеспечит высокое качество интеграции данных.

В процессе поиска оптимального решения биоинспирированные алгоритмы имитируют коллективное поведение какого-либо вида животных, насекомых или растений [11]. Алгоритмы, вдохновленные живой природой, показывают высокую эффективность в задачах, в которых нахождение глобального оптимума невозможно или требует больших временных затрат. Такие алгоритмы позволяют найти квазиоптимальное решение за приемлемое время [3].

Еще одной перспективной группой методов выступают ансамбли алгоритмов. Ансамбль состоит из нескольких взаимосвязанных алгоритмов, работающих последовательно и параллельно [4]. Результатом работы ансамбля выступает интегрированный набор решений. Популярным представителем ансамблей можно назвать бустинг, идея которого состоит в последовательном применении нескольких алгоритмов, причем каждый последующий алгоритм имеет доступ к решениям, полученным в процессе работы предыдущего алгоритма [13–15].

В условиях неоднородности источников данных, генетический алгоритм, алгоритм роя светлячков и алгоритм пчелиной колонии могут показывать различную эффективность, более того, она может варьироваться в зависимости от стадии интеграции. Таким образом, полагаем, что последовательное применение нескольких алгоритмов, с небольшим числом итераций, способно показать повышение качества получаемых решений, с незначительным увеличением временных затрат.

В данной работе задачу интеграции данных ИС предлагается рассматривать как интеграцию онтологий разнородных предметных областей. Основной целью задачи является обеспечение обобщенного представления данных.

Системная интеграция нескольких онтологий, в общем виде, может быть представлена формулой:

$$O^U = \bigcup_{i=1}^N O_i,$$

где \bigcup – знак концептуального объединения. Данный знак подразумевает что исходные онтологические графы будут системно интегрированы, причем будут учтены их связи и области определений $O_i (i = \overline{1, N})$.

Для корректной интеграции, в условиях неоднородности информационных систем, требуется найти чем похожи онтологии, лежащие в основе систем и чем они отличаются. Кроме того, также имеет место необходимость согласования спецификаций, в том числе и неоднородных, онтологий, включая и преобразование информации. В этом случае, различные ИС начинают работать в рамках единой предметной области, эта совместная работа происходит на семантическом уровне.

Зачастую разрабатывая алгоритмы решающие задачу интеграции, целевая функция получается со множеством экстремумов и сочетающая в себе множество параметров. В работе предлагается искать оптимум целевой функции используя стохастическую оптимизацию, которая представлена такими алгоритмами как: генетический алгоритм, алгоритм колонии пчел, алгоритм имитирующий поведение светлячков, рой частиц или колонии муравьев. Данные алгоритмы производят поиск субоптимального решения по всей площади пространства поиска. Основным критерием в работе, является качество получаемых итоговых решений при минимальном увеличении временной сложности.

2. Математическая модель и метод оценки семантической близости. Под термином «онтология» принято понимать некую формальную спецификацию концептуализации, обозначенную в определенном контексте предметной области [1, 2]. Термин концептуализация подразумевает представление предметной области, путем рассмотрения и описания концептов множеств некоторой предметной области, а также описание связей концептов. [6].

Концепт онтологии ИС – это некоторая единица знаний, которая может быть идентифицирована по имени. Она характеризуется следующим образом:

$$C_i = (N_i, T_i), \quad (1)$$

где N_i – идентификатор (уникальное имя) i -го концепта; T_i – тип i -го концепта (абстрактный, представимый, либо составной).

Модель онтологии в формальном виде предлагается представлять следующим образом:

$$O_f = \langle P, C, A, R \rangle, \quad (2)$$

где P – множество объектов (понятий); C – множество классов (экземпляров понятий); A – множество атрибутов; R – множество отношений [1, 2].

В связи с применением подхода с использованием множества концептов, представим онтологию информационной системы в виде:

$$O = \langle C, A, D, R_c, R \rangle, \quad (3)$$

где C – множество концептов; A – множество атрибутов; R_c – отношение концепта со множеством его атрибутов; R – множество отношений между концептами; D – словарь терминов, включающий характеристики объектов.

Тогда информационную систему, которая использует онтологию O , можно представить в следующем виде:

$$U = \langle O, U_i, R_u, R_r \rangle, \quad (4)$$

где U_i – множество элементов информационной системы; R_u – отношение элемента ИС с его концептом; R_r – связь между элементами ИС.

Построим математическую модель оценки семантической близости. Различные информационные системы должны соответствовать общей структурной схеме и учитывать всевозможные семантические зависимости, необходимые для редактирования информационных систем и для построения однородного формата, в котором будут представлены данные из нескольких ИС. Для осуществления этого нужно найти что у этих информационных систем общего и в чем различия. Поэтому требуется определить семантическую однородность, и зависимости концептов онтологий [8]. Таким образом, целью интеграции можно назвать сведение множества онтологий разнородных ИС к какой-то общей онтологии.

Опираясь на вышесказанное, сделаем вывод о том, что зависимость между различными онтологиями ИС есть множество зависимостей между концептами, входящими в онтологии ИС. Обозначим множество зависимостей различных концептов в онтологии в виде $D = \langle D_1, \dots, D_n \rangle$.

Для представления модели интеграции, учитывая вышеприведенные онтологии, предлагается использовать кортеж длинны четыре:

$$M = \langle O, U, D, mapping \rangle, \quad (5)$$

где O – онтология ИС; U – ИС; D семантические зависимости онтологий; $mapping$ – представление отображения онтологий.

Рассмотрим предложенный подход к оценке семантической близости онтологий ИС. Наиболее популярными методами семантической интеграции онтологий считаются те, оценка близости в которых определяется на уровне элементов онтологии.

Одной из основных проблем мер в структурах, которые построены на онтологиях является симметричность. Как показывают результаты, для таких отношений как иерархические и ассоциативные, меры близости могут быть не симметричны [6, 9]. Исходя из этого нужно выявить семантическую близость для таких объектов, что не имеют иерархической связи, а также у которых отсутствуют проблемно-специфические связи. Опираясь на вышесказанное можно с уверенностью сказать, что выявление данной близости является важной и актуальной задачей.

В данной работе, для численной оценки семантической близости отдельных онтологических концептов, предлагается использовать эволюционный подход и бустинг. Поиск решений с помощью генетических и биоинспирированных методов, и дальнейших анализ их семантической близости позволят найти субоптимальные решения за полиномиальное время [14–20]. В меру близости включена оценка критериев подобия концептов онтологии. Эта оценка строится из трех основных мер: атрибутивная, таксономическая и реляционная меры. Атрибутивная мера подразумевает сопоставление значений и атрибутов понятий (концептов).

Определение степени схожести концептов, а также поиск понятий и их положение в онтологии принадлежит таксономической мере. Кроме того, определение наименьшего возможного пути можно принять за некоторое количество концептов в совокупности находящимися между исследуемыми, исходя из этого наименьшее количество этих концептов, является признаком наибольшего из сходства.

Следующая мера – реляционная. Такая мера показывает отношения концептов. При использовании данного подхода и сопоставлении двух концептов, представляется возможным вычислить, так называемую, гибридную меру семантической близости в неоднородных онтологиях. По своей природе, гибридные меры есть ни что иное как сборные меры, т.е. они гибридизированы из уже приведенных (известных) мер оценки и критериев определения подобия концептов. Такие меры могут быть применены в ходе автоматической оценки полученных результатов отображения. Подобные данные лишены расхождений и в тоже время имеют повышенную точность.

Данная формула формально описывает численную оценку семантической близости:

$$M(c_i, c_j) = \sum_{q=1}^n w_q M^q(c_i, c_j), \quad (6)$$

где M^q – q -я мера близости для выбранного критерия; вес каждой меры сходства w_q является весом меры сходства, с ее помощью можно определить важность критерия; сумма всех весов равна 1; n – кол-во мер близости.

Для определения меры близости понятий, с учетом их положения, представлены $M^T(c_i, c_j)$, $M^R(c, c_j)$, $M^A(c, c_j)$ соответственно. Основой для определения близости являются их отношения, сопоставление их атрибутов и значений.

В таксономической мере близости учтены все иерархические характеристики двух понятий $M^T(c, c_j)$. В ней происходит сравнение всех элементов концептов. $PL_p(c_i)$ и $PL_p(c)$. Ниже приведена формула которая определяет общие элементы:

$$M^T(c_i, c_j) = 1, \text{ если } c_i = c_j \frac{|PL_p(c_i) \cap PL_p(c_j)|}{|PL_p(c_i) \cup PL_p(c_j)|}, \text{ если } c_i \neq c_j, \quad (7)$$

где $PL_p(C_i)$ – множество элементов концепта C_i .

В основе реляционной оценки близости лежит предположение: при нахождении одинаковых отношений R_1, R_2, R_3 у двух концептов, некоторым третьим концептом, они автоматически становятся более похожими, в сравнении с концептами, которые таких отношений не имеют. $C_r(c_i) = \{c_j \in C \vee R_1(c_i, c_j) \vee R_2(c_i, c_j) \vee R_3(c_i, c_j) \vee c_j = c_i\}$ – множество, объединяющее концепты с отношениями R_1, R_2, R_3 .

Ассоциативность концептов можно определить следующим образом:

$$R_A(c_i) = \{c_i: c_i \in C_r(c_i)\}. \quad (8)$$

Ассоциативная мера близости для концептов из множеств $R_A(c_j)$ и $R_A(c_i)$ будет иметь вид:

$$M_{R_a}(R_a(c_i), R_a(c_j)) = \sum_{c_i \in R_A(c_i), c_i \in R_A(c_j)} M^T(c_i, c_j). \quad (9)$$

$M^R(c_i, c_j)$ – реляционная мера близости. С ее помощью можно с ее помощью можно оценить насколько один концепт подобен другому, на основании схожести концептов множества $C_r(c_i)$:

$$M^R(c_i, c_j) = 1, \text{ если } c_i = c_j \frac{M_{R_a}(R_a(c_i), R_a(c_j))}{|R_a(c_i) \cup R_a(c_j)|}, \text{ если } c_i \neq c_j. \quad (10)$$

$M^A(c_i, c_j)$ – атрибутивная мера близости. В ней учтены такие аксиомы как: аксиома независимости и разрешимости. Она определяется следующей формулой:

$$M^A(c_i, c_j) = \frac{|A^{c_i} \cap A^{c_j}|}{|A^{c_i} \cup A^{c_j}|}, \quad (11)$$

где A^{c_i} и A^{c_j} множество атрибутов концептов c_i, c_j соответственно.

Таким образом согласовав реляционную, атрибутивную и таксономическую меры подобия можно предложить следующую формулу, с помощью которой можно оценить семантическую близость $M(c_i, c_j)$ объектов знаний c_i и c_j неоднородных онтологий O и O' :

$$M(c_i, c_j) = t \cdot M^T(c_i, c_j) + r \cdot M^R(c, c_j) + a \cdot M^A(c, c_j), \quad (12)$$

где t, r, a – коэффициенты, с помощью которых определяется степень важности мер и критериев близости.

$$M^T(c_i, c_j), M^R(c, c_j), M^A(c, c_j).$$

В работе предлагается использовать t, r, a – весовые коэффициенты, которые позволят управлять процессом вычисления семантической близости разнородных концептов. В соответствии с формулой (12) в задаче оценки семантической близости присутствует ряд ограничений:

$$\begin{aligned} \min_{t,r,a} f(\bar{x}), \bar{x} &= (t, r, a) \in F \subseteq S \\ t, r, a &\in [0,1] \\ t + r + a &= 1, \end{aligned}$$

где \bar{x} – допустимое решение, F – область допустимых значений, S – область поиска.

Отсюда следует, что целевая функция будет выглядеть так:

$$f_{t,r,a} = \sum k_{j \in O_i} k_{j \in O_2} (t \cdot M^T(c_i, c_j) + r \cdot M^R(c, c_j) + a \cdot M^A(c, c_j) - 1)^2. \quad (13)$$

Следующим этапом будет рассмотрение обобщенной архитектуры подсистемы интеграции данных, в которой применен бустинг биоинспирированных алгоритмов.

3. Обобщенная архитектура интеграции данных. Использование онтологии при решении задачи интеграции данных по требованию, позволяет разрабатывать приложения, которые работают с информацией на семантическом уровне. Помимо прочего такой подход лишен множества недостатков других методов. На рис. 1 приведена схема интеграции информационных систем.

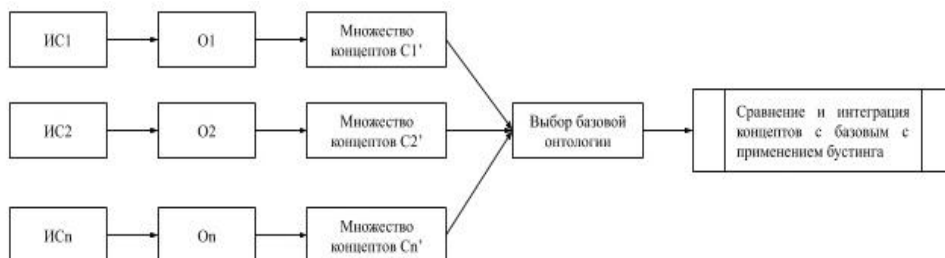


Рис. 1. Схема интеграции нескольких ИС

Данную схему можно разделить на несколько частей:

- ◆ представление онтологии ИС;
- ◆ представление множества концептов онтологий;
- ◆ выбор базовой онтологии;
- ◆ сравнение и интеграция концептов с применением бустинга (подробно рассмотрена на рис. 2).

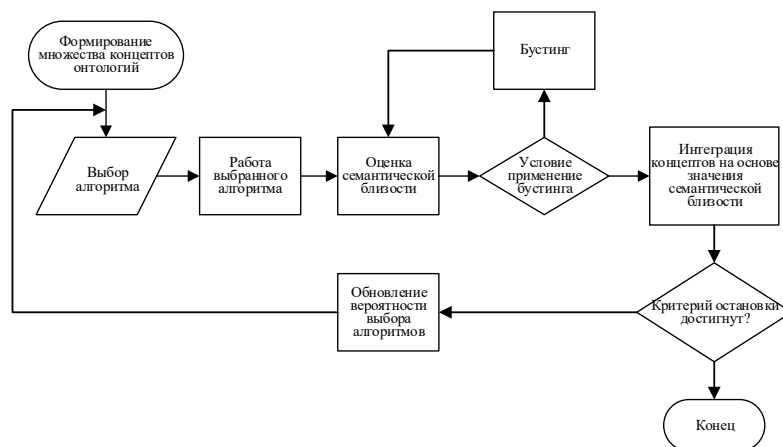


Рис. 2. Структура процедуры бустинга биоинспирированных алгоритмов

Опишем основные шаги процедуры бустинга:

1. Формирование множества концептов онтологий каждой ИС, инициализация множества вероятностей выбора каждого отдельного алгоритма;
2. Запуск одного из алгоритмов интеграции;
3. Процесс поиска субоптимального решения выбранным алгоритмом;
4. Семантическая оценка полученного решения;
5. При невыполнении критерия, применение бустинга. Последующий алгоритм будет оперировать в том числе решениями, полученными предшественником;
6. Интеграция концептов элементов с помощью выбранного алгоритма и оценка семантической интеграции по формуле (6);

7. Обновление вероятностей выбора алгоритмов в связи с полученными результатами;

8. Закончить цикл, если критерий остановки достигнут, в противном случае перейти к пункту 2.

Предполагается, что процедура бустинга повысит качество решений, получаемых за полиномиальное время. Таким образом в алгоритме интеграции устранены множество недостатков и представлена возможность разрабатывать интегрированные ИС, которые оперируют информацией на семантическом уровне.

4. Экспериментальные исследования. На основе сравнительного анализа с такой популярной системой как QOM [18], а также системой PROMPT [17] проводилась оценка эффективности представленного алгоритма. Приведенные выше системы поддерживают как отображение, так и процессы объединения различных онтологий ИС.

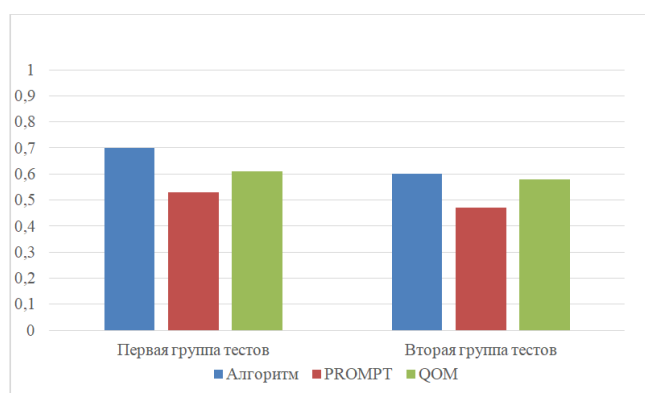


Рис. 3. Оценка точности разных подходов к оценке семантической близости

Полученные результаты позволяют сказать, что созданный алгоритм показывает повышение точности и улучшение качества при сопоставлении онтологий.

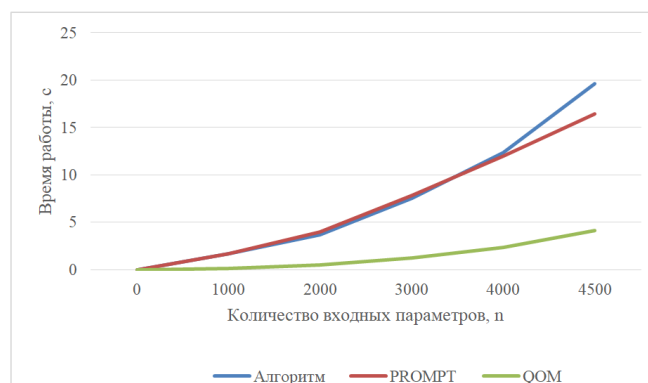


Рис. 4. График временной сложности

Результаты экспериментальных исследований позволили определить зависимость времени работы алгоритма от размера входных параметров n . При небольшом количестве входных параметров, время работы сопоставимо с временем работы аналогов. При увеличении n разница во времени с конкурентами, компенсируется качеством получаемых решений.

Заключение. В результате проведенной работы была предложена математическая модель, схема интеграции данных нескольких информационных систем, и структура процедуры бустинга биоинспирированных алгоритмов, были проведены экспериментальные исследования для подтверждения эффективности полученных результатов.

Предложенные алгоритм интеграции в условиях неоднородности ИС и математическая модель способны обеспечить повышение эффективности в процессе интеграции, в сравнении с известными подходами. Представленная математическая модель интеграции учитывает спецификации онтологий и предлагает измерять близость концептов сопоставляемых онтологий. Бустинг, структура которого была представлена в работе, позволяет выделять ключевые концепты, необходимые для построения результирующей онтологии. Это позволяет строить результирующую онтологию, устраняя субъективность различных понятий в онтологиях, а также избавляться от различных точек зрения создателей онтологий.

Экспериментальные исследования предложенной процедуры бустинга биоинспирированных алгоритмов для решения задач интеграции данных показали повышение эффективности получаемых решений в сравнении с аналогами. Алгоритм становится эффективнее с ростом числа входных параметров. Временная сложность не превышает полиномиальную.

В дальнейшем для развития проведенных исследований планируется применение при решении задачи интеграции данных модифицированных биоинспирированных алгоритмов и базы предикатов. Также необходима разработка нового подхода к формированию методов оценки семантической близости и программного комплекса для решения задачи интеграции данных с использованием бустинга модифицированных биоинспирированных алгоритмов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Бажанова А.И., Мартыненко Т.В.* Исследование применения онтологических моделей для семантического поиска // Матер. II научно-практической конференции студентов, аспирантов и молодых ученых. – Донецк: ДонНТУ, 2011. – С. 244–248.
2. *Вагин В.Н., Михайлов И.С.* Разработка метода интеграции информационных систем на основе метамоделирования и онтологии предметной области // Программные продукты и системы. – 2008. – С. 22–26.
3. *Курейчик В.В., Курейчик Вл.Вл.* Биоинспирированный поиск при проектировании и управлении // Известия ЮФУ. Технические науки. – 2012. – № 11 (136). – С. 178–183.
4. *Родзин С.И., Курейчик В.В.* Теоретические вопросы и современные проблемы развития когнитивных биоинспирированных алгоритмов оптимизации // Кибернетика и программирование. – 2017. – № 3. – С. 51–79.
5. *Тузовский А.Ф.* Работа с онтологиями в системе управления знаниями организации // Сб. тезисов докладов второй международной конференции по когнитивной науке CogSci-2006. – СПб.: СПбГУ, 2006. – Т. 2. – С. 581–583.
6. *Батоврин В.К., Козаловский М.Р., Королев А.С., Петров А.Б.* Разработка понятийной схемы (онтологии) для обеспечения единой семантики в среде открытой системы интеграции разнородных данных // Телематика-2006: Матер. Всероссийской научно-методической конференции. – СПб.: Изд-во СПбГУ, 2006. – С. 90–91.
7. *Тузовский А.Ф., Козлов С.В.* Работа с онтологиями в системе управления знаниями организации // Сб. тезисов докладов второй международной конференции по когнитивной науке CogSci-2006 – Санкт-Петербург, 9–13 июня 2006. – СПб.: СПбГУ, 2006. – Т. 2. – С. 581–583.
8. *Тузовский А.Ф.* Онтолого-семантические модели в корпоративных системах управления знаниями: дисс. ... д-ра техн. наук. – Томск: Томский политехнический университет, 2007.
9. *Скворцов Н.А.* Вопросы согласования неоднородных онтологических моделей и онтологических контекстов // Онтологическое моделирование. – М.: ИПИ РАН, 2008. – С. 149–166.

10. Курейчик В.В., Курейчик В.М., Сороколетов П.В. Анализ и обзор моделей эволюции // Известия РАН. Теория и системы управления. – 2007. – № 5. – С. 114-126.
11. Родзин С.И., Курейчик В.В. Состояние, проблемы и перспективы развития биоэвристик // Программные системы и вычислительные методы. – 2016. – № 2. – С. 158-172.
12. Зайцев А.А., Курейчик В.В., Полуанов А.А. Обзор эволюционных методов оптимизации на основе роевого интеллекта // Известия ЮФУ. Технические науки. – 2010. – № 12 (113). – С. 7-12.
13. Mayr A., Binder H., Gefeller O., Schmid M. The Evolution of Boosting Algorithms – From Machine Learning to Statistical Modelling // Methods Inf. Med. – 2014. – Vol. 53. – P. 419-427.
14. Freund Y. and Schapire R. Experiments with a new boosting algorithm // In Thirteenth International Conference on Machine Learning. Bari, Italy, 1996. – P. 148-156.
15. Freund Y. and Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting // Journal of Computer and System Sciences. – 1997. – Vol. 55 (1). – P. 119-139.
16. Курейчик В.М., Курейчик В.В., Родзин С.И. Модели параллелизма эволюционных вычислений // Вестник Ростовского государственного университета путей сообщения. – 2011. – № 3 (43). – С. 93-97.
17. Noy N., Musen M. The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. – Stanford Medical Informatics, Stanford University, 2003.
18. Ehrig, Marc and Staab, Steffen QOM – Quick Ontology Mapping. in S.A. McIlraith et al. (Eds.): ISWC 2004, LNCS 3298. – 2004. – P. 683-697.
19. Курейчик В.М., Курейчик В.В., Родзин С.И., Гладков Л.А. Основы теории эволюционных вычислений. – Ростов-на-Дону: ЮФУ, 2010. – 222 с.
20. Курейчик В.В., Бова В.В., Курейчик Вл.Вл. Комбинированный поиск при проектировании // Образовательные ресурсы и технологии. – 2014. – № 2 (5). – С. 90-94.

REFERENCES

1. Bazhanova A.I. Martynenko T.V. Issledovanie primeneniya ontologicheskikh modeley dlya semanticheskogo poiska [Research of application of ontological models for semantic search], *Mater. II nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh* [Materials of the II scientific-practical conference of students, postgraduates and young scientists]. Donetsk: DonNTU, 2011, pp. 244–248.
2. Vagin V.N., Mikhaylov I.S. Razrabotka metoda integratsii informatsionnykh sistem na osnove metamodelirovaniya i ontologii predmetnoy oblasti [Development of information systems integration method based on metamodeling and domain ontology], *Programmnye produkty i sistemy* [Software products and systems], 2008, pp. 22-26.
3. Kureychik V.V., Kureychik V.V. Bioinspirirovannyi poisk pri proektirovani i upravlenii [Biospherology search in the design and management], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 11 (136), pp. 178-183.
4. Rodzin S.I., Kureychik V.V. Teoreticheskie voprosy i sovremennye problemy razvitiya kognitivnykh bioinspirirovannykh algoritmov optimizatsii [Theoretical questions and modern problems of development of cognitive bioinspired optimization algorithms], *Kibernetika i programmirovaniye* [Cybernetics and programming], 2017, No. 3, pp. 51-79.
5. Tuzovskiy A.F. Rabota s ontologiyami v sisteme upravleniya znaniyami organizatsii [Work with ontologies in the knowledge management system of the organization], *Sb. tezisov dokladov vtoroy mezhdunarodnoy konferentsiya po kognitivnoy nauke CogSci-2006* [Collection of abstracts of the second international conference on cognitive science CogSci-2006]. Saint Petersburg: SPbGU, 2006, Vol. 2, pp. 581-583.
6. Batovrin V.K., Kogalovskiy M.R., Korolev A.S., Petrov A.B. Razrabotka ponyatiynoy skhemy (ontologii) dlya obespecheniya edinoy semantiki v srede otkrytoy sistemy integratsii raznorodnykh dannykh [Development of a conceptual scheme (ontology) to provide a unified semantics in an open system of heterogeneous data integration], *Telematika-2006: materialy Vserossiyskoy nauchno-metodicheskoy konferentsii* [Telematics-2006: Proceedings of the all-Russian scientific and methodological conference]. Saint Petersburg: Izd-vo SPbGU, 2006, pp. 90-91.

7. *Tuzovskiy, A.F., Kozlov S.V.* Rabota s ontologiyami v sisteme upravleniya znaniyami organizatsii [Work with ontologies in the knowledge management system of the organization], *Sb. tezisev dokladov vtoroy mezhdunarodnoy konferentsiya po kognitivnoy nauke CogSci-2006 – Sankt-Peterburg, 9–13 iyunya 2006* [Collection of abstracts of the second international conference on cognitive science CogSci-2006-St. Petersburg, June 9-13, 2006]. Saint Petersburg: SPbGU, 2006, Vol. 2, pp. 581-583.
8. *Tuzovskiy A.F.* Ontologo-semanticheskie modeli v korporativnykh sistemakh upravleniya znaniyami: diss. ... d-ra tekhn. nauk [Ontological and semantic models in corporate knowledge management systems: dr. of eng. sc, diss.]. Tomsk: Tomskiy politekhnicheskii universitet, 2007.
9. *Skvortsov N.A.* Voprosy soglasovaniya neodnorodnykh ontologicheskikh modeley i ontologicheskikh kontekstov [Problems of coordination of heterogeneous ontological models and ontological contexts], *Ontologicheskoe modelirovanie* [Ontological modeling]. Moscow: IPI RAN, 2008, pp. 149-166.
10. *Kureychik V.V., Kureychik V.M., Sorokoletov P.V.* Analiz i obzor modeley evolyutsii [Analysis and review of models of evolution], *Izvestiya RAN. Teoriya i sistemy upravleniya* [Izvestiya RAS. Theory and control systems], 2007, No. 5, pp. 114-126.
11. *Rodzin S.I., Kureychik V.V.* Sostoyanie, problemy i perspektivy razvitiya bioevristik [Status, problems and prospects of development of biometric], *Programmnye sistemy i vychislitel'nye metody* [Software systems and computational methods], 2016, No. 2, pp. 158-172.
12. *Zaytsev A.A., Kureychik V.V., Polupanov A.A.* Obzor evolyutsionnykh metodov optimizatsii na osnove roevogo intellekta [Review of evolutionary optimization methods based on swarm intelligence], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12 (113), pp 7-12.
13. *Mayr A., Binder H., Gefeller O., Schmid M.* The Evolution of Boosting Algorithms – From Machine Learning to Statistical Modelling, *Methods Inf. Med.*, 2014, Vol. 53, pp. 419-427.
14. *Freund Y. and Schapire R.* Experiments with a new boosting algorithm, *In Thirteenth International Conference on Machine Learning, Bari, Italy, 1996*, pp. 148-156.
15. *Freund Y. and Schapire R.* A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 1997, Vol. 55 (1), pp. 119-139.
16. *Kureychik V.M., Kureychik V.V., Rodzin S.I.* Modeli parallelizma evolyutsionnykh vychisleniy [Models of parallelism of evolutionary calculations.], *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov state University of railway engineering], 2011, No. 3 (43), pp. 93-97.
17. *Noy N., Musen M.* The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. Stanford Medical Informatics, Stanford University, 2003.
18. Ehrig, Marc and Staab, Steffen QOM – Quick Ontology Mapping. in S.A. McIlraith et al. (Eds.): ISWC 2004, LNCS 3298, 2004, pp. 683-697.
19. *Kureychik V.M., Kureychik V.V., Rodzin S.I., Gladkov L.A.* Osnovy teorii evolyutsionnykh vychisleniy [Fundamentals of the theory of evolutionary computation]. Rostov-on-Don: YuFU, 2010, 222 p.
20. *Kureychik V.V., Bova V.V., Kureychik V.V.* Kombinirovannyi poisk pri proektirovanii [Combined search in design], *Obrazovatel'nye resursy i tekhnologii* [Educational resources and technologies], 2014, No. 2 (5), pp. 90-94.

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

Кравченко Юрий Алексеевич – Южный федеральный университет; e-mail: yakravchenko@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Балабанов Дмитрий Валерьевич – e-mail: dbalabanov@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Ковтун Антон Владиславович – e-mail: anton.kovtun93@gmail.com. кафедра систем автоматизированного проектирования, аспирант.

Kravchenko Yuriy Alekseevich – Southern Federal University; e-mail: yakravchenko@sfedu.ru; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Balabanov Dmitryi Valerievich – e-mail: dbalabanov@sfedu.ru; the department of computer aided design; graduate student.

Kovtun Anton Vladislavovich – e-mail: anton.kovtun93@gmail.com; the department of computer aided design; graduate student.

УДК 004.023

DOI 10.23683/2311-3103-2019-4-114-126

С.Н. Щеглов

МЕТОДОЛОГИЯ ИСПОЛЬЗОВАНИЯ БИОИНСПИРИРОВАННЫХ МЕТОДОВ ДЛЯ ИНТЕГРИРОВАННОЙ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ НА ПРИМЕРЕ РЕШЕНИЯ ЗАДАЧИ ТРАНСПОРТНОГО ТИПА*

Приводится методология использования биоинспирированных методов для интегрированной обработки больших данных на примере решения задачи транспортного типа. Основное место среди прикладных задач транспортного типа, занимают задачи построения транспортных маршрутов, которые позволяют до минимума сократить пробег транспортных средств или минимизировать затраты на перевозку грузов. Маршрутизация перевозок – это наиболее совершенный способ организации потоков грузов с предприятий, оказывающий существенное влияние на ускорение оборота транспорта при рациональном и эффективном его использовании. Для данного класса комбинаторных задач, отсутствуют эффективные классические методы и алгоритмы решения. Эти задачи характеризуются конечным, но весьма большим числом возможных решений. Их можно поставить как задачи целочисленного программирования, но и в этом случае отсутствуют эффективные алгоритмы. Поэтому, разработка методов и алгоритмов для решения задач транспортного типа, осуществляющаяся на протяжении многих лет, является по-прежнему, актуальной проблемой. Осуществлен методологический анализ проблемы исследования. Анализ показал, что использование методов и алгоритмов последовательного и параллельного биоинспирированного поиска для решения рассматриваемого класса задач транспортного типа, является актуальной научной задачей, представляющей практический интерес. Приведена постановка задачи. Показана схема интегрированного поиска, которая позволяет распараллелить процесс нахождения приемлемого решения для задач большой размерности. Рассмотрена структурная схема биоинспирированного поиска для задачи об экстремальном пути. Представлены результаты вычислительных экспериментов. Результаты исследований позволяют сделать вывод о том, что временная сложность рассмотренных алгоритмов биоинспирированного поиска не выходит за пределы полиномиальной зависимости, и может быть выражена формулой: $O(N^2)$, где N – число вершин графа (размер решаемой задачи).

Алгоритм; метод; графовые модели; эвристика; структура; схема; эксперимент; исследование; генетический поиск; адаптация; искусственный интеллект; эволюция; решение.

S.N. Scheglov

METHODOLOGY OF USING BIOINSPIRED METHODS FOR INTEGRATED PROCESSING OF GREAT DATA ON THE EXAMPLE OF THE TRANSPORT TYPE PROBLEM SOLUTION

This paper presents a methodology for using bioinspired methods for integrated processing of big data using the example of solving the transport type problem. The main place among the applied problems of the transport type is given to the tasks of building transport facilities, which

* Работа выполнена при финансовой поддержке РФФИ (проект № 18-07-00055).