

**В.В. Марков, Ю.А. Кравченко, М.А. Кузьмина**

**РАЗВИТИЕ МЕТОДОВ СЕМАНТИЧЕСКОЙ ФИЛЬТРАЦИИ НА ОСНОВЕ РЕШЕНИЯ ЗАДАЧИ КЛАСТЕРИЗАЦИИ БИОИНСПИРИРОВАННЫМИ АЛГОРИТМАМИ\***

*Данная статья посвящена решению проблемы развития теоретических основ управления знаниями для создания междисциплинарных интеллектуальных информационных систем семантического поиска, способствующих развитию перспективных направлений информатики и информационных технологий, связанных с решением проблем интеграции знаний из неоднородных распределенных источников знаний, идентификации предметной области и концептуального моделирования, анализа и извлечения данных. Актуальность решения данной проблемы заключается в расширении прикладных областей применения разрабатываемых интеллектуальных информационных систем для целого ряда актуальных задач из класса NP-сложных. В предложенной статье рассмотрена задача кластеризации. С целью повышения эффективности и качества решения задач управления знаниями предлагаются модифицированные алгоритмы кластеризации, основанные на использовании биоинспирированных подходов. Разработанные модификации алгоритмов позволят повысить качество получаемых решений. Данные исследования в области семантического поиска составляют важное научное направление в рамках проблемы развития технологий управления знаниями. В статье рассмотрено решение задачи кластеризации при помощи биоинспирированных алгоритмов. Предложены два новых алгоритма кластеризации: комбинированный алгоритм на основе алгоритма пчелиного роя и алгоритма  $k$ -means, а также муравьиный алгоритм кластеризации. Для исследования полученных алгоритмов, их эффективности и качества результатов, были разработаны программные модули, моделирующие предложенные алгоритмы. Произведен анализ полученных результатов, сравнение с классическим алгоритмом, сделаны выводы об эффективности новых алгоритмов.*

*Семантическая фильтрация; кластеризация; биоинспирированные алгоритмы; управление знаниями; модификация, комбинирование.*

**V.V. Markov, Yu.A. Kravchenko, M.A. Kuzmina**

**DEVELOPMENT OF SEMANTIC FILTERING METHODS BASED ON SOLVING THE TASK OF CLUSTERING BY BIOINSPIRED ALGORITHMS**

*This article is devoted to solving the problems of theoretical knowledge management foundation development for the creation of multi-disciplinary intellectual information systems, semantic search, promoting the development of science promising areas and information technologies related to solving the problems of knowledge integration from heterogeneous distributed sources, the identification of the subject area and conceptual modeling, analysis and retrieval data. The urgency of solving this problem is to expand the application areas of application developed intelligent information systems for a range of class NP- complex urgent tasks. The proposed article considers the clustering problem. In order to improve the efficiency and quality of knowledge management solutions offered are the modified tasks clustering algorithms based on using bioinspired approaches. The developed modifications of the algorithms improve the quality of the solutions obtained. These studies in the field of semantic search constitute an important scientific direction within the framework of the problem of the knowledge management technologies development. The solution of the problem of clustering by means of using of bioinspired algorithms is considered herein. The following two new clustering algorithms: a combined algorithm based on the Artificial Bee Colony algorithm and the  $k$ -means algorithm, as well as the Ant clustering algorithm, are proposed. To study the algorithms obtained, their effectiveness and the results quality,*

\* Работа выполнена при поддержке РФФИ (проект № 16-07-00703).

*software modules that simulate the proposed algorithms have been developed. The analysis of the obtained results, comparison with the classical algorithm, as well as conclusions about the effectiveness of new algorithms. are presented.*

*Semantic filtering; clustering; bioinspired algorithms; knowledge management; modification; combining.*

**Введение.** Кластеризация данных – это процесс группировки данных в кластеры, при котором данные в одном кластере имеют высокую степень сходства, при этом максимально не похожи на данные из других кластеров [1–4]. Опишем значимость задачи кластеризации с точки зрения семантической фильтрации и идентификации. Особую сложность имеют задачи обработки семантического ядра в интеллектуальных информационных системах управления знаниями. Невозможно упорядочить «хаос» обрабатываемых поисковых запросов только лишь на основе отбора и анализа ключевых понятий и определений. Выходом из данной ситуации является группировка и кластеризация основных понятий и системы отношений между ними, а также множества зависимостей и закономерностей, построенных на основе данных отношений.

Кластеризация запросов – объединение ключевых понятий по их семантическому смыслу в контексте рассматриваемой предметной области или сегмента междисциплинарного информационного пространства для идентификации основного и дополнительного смысла исследуемого информационного ресурса. Объективное определение семантического ядра и его кластеризация обеспечит релевантное построение системы неявных зависимостей и закономерностей на множестве отношений между элементами знаний в распределенных разнородных информационных ресурсах.

Кластеризация широко используется как в качестве отдельного инструмента анализа, так и как один из этапов предварительной обработки данных перед использованием других аналитических методов. Несмотря на важность задачи кластеризации, в виду ее сложности многие проблемы остаются нерешенными, в том числе проблема застревания алгоритмов кластеризации вблизи локальных оптимальных решений.

В последнее время широкую известность получили биоинспирированные алгоритмы. Биоинспирированные алгоритмы – это инновационная категория искусственного интеллекта, группа метаэвристических алгоритмов, вдохновленных интеллектуальным поведением групп насекомых или животных в природе: муравьиных колоний, птичьих стай, пчелиных колоний, бактерий и т.д. [2–10]. Эти алгоритмы набирают популярность благодаря их способности к решению задач, где нахождение оптимальных решений классическими алгоритмами (k-means, c-means) неэффективно или невозможно.

В данной работе предлагаются новые алгоритмы, позволяющие применить инновационные биоинспирированные алгоритмы к задаче кластеризации. Целью данной работы является разработать улучшенные алгоритмы кластеризации и решить проблему застревания решений вблизи локальных оптимумов, тем самым повысив качество кластеризации, за счет применения биоинспирированных алгоритмов.

**1. Комбинированный алгоритм на основе k-means и алгоритма пчелиной колонии.** Алгоритм k-means является достаточно быстрым из-за его простоты и небольшого числа итераций. Однако зависимость от изначальных центроидов и застревание в локальном оптимуме – серьезные проблемы алгоритма. С другой стороны, алгоритм пчелиной колонии выполняет глобальный поиск во всем пространстве решений. Если дано достаточно времени, алгоритм пчелиной колонии может выдавать глобальные качественные результаты. В данной главе предлагает-

ся новый комбинированный алгоритм, использующий достоинства этих двух алгоритмов. Новый алгоритм не зависит от начальных центроидов и может избежать попадания в локальное оптимальное решение.

Алгоритм пчелиной колонии (Artificial Bee Colony Algorithm, ABC) моделирует пространство поиска как среду кормления, где каждая точка в поисковом пространстве соответствует источнику пищи (решению), который пчелы могут использовать. Пригодность решения представляется как количество нектара в источнике пищи. В рамках этого алгоритма, в пчелиной колонии существуют три вида пчел: рабочие пчелы, пчелы-наблюдатели и пчелы-разведчики. Рабочие пчелы отвечают за конкретные источники пищи, которые они изучили до этого, и дают качественную информацию о пищевых источниках пчелам-наблюдателям. Пчелы-наблюдатели принимают информацию об источниках пищи, а затем выбирают источник для изучения и использования в зависимости от этой информации о качестве нектара. Чем больше нектара содержит источник пищи, тем больше вероятность, что пчелы-наблюдатели выберут именно его [3, 11–15].

В предлагаемом алгоритме каждый источник пищи в среде поиска представляет собой набор центроидов, т.е. источник пищи представляет собой одно возможное решение для кластеризации, а положение  $x_i$  строится как:

$$x_i = (C_{i1}, C_{i2}, \dots, C_{iK}), \quad (1)$$

где  $K$  – количество кластеров,  $C_{ij}$  –  $j$ -й центроид  $i$ -го источника пищи. Процедуру предложенного алгоритма можно описать следующим образом:

1. Произвольно инициализировать позиции источников пищи (наборов центроидов) и использовать алгоритм  $k$ -means для завершения задачи кластеризации для всех подготовленных позиций и вычислить значения пригодности каждой группы центроидов;
2. Найти новые источники пищи и обновить место источников пищи рабочими пчелами. Применить алгоритм  $k$ -means и принцип жадного выбора для оценки новых значений пригодности, затем сравнить их с начальными. Лучшие источники пищи будут доставлены пчелам-наблюдателям;

Чтобы измерить качество кластеризации источника пищи, необходимо определить функцию критерия качества кластеризации. В предлагаемом алгоритме в качестве критерия кластеризации и функции расчета пригодности каждого источника пищи (функции fitness) используется расстояние по Хеммингу, определяемое по формуле (2):

$$d_p(x_i, x_j) = \sum_{t=1}^m |x_{it} - x_{jt}|. \quad (2)$$

1. Вычислить значения вероятности для источников пищи и обновить их место в соответствии с вероятностными значениями, вычисленными пчелами-наблюдателями. Снова применить алгоритм  $k$ -means и принцип жадного выбора для завершения кластеризации, оценки новых значений пригодности и сравнения их с оригинальными для их обновлениями;

2. Проверить счетчик проб источников пищи и создать новый источник пищи (набор центроидов) в поисковом пространстве, для которого увеличить значение параметра «лимит».

Графически алгоритм представлен на рис. 1.

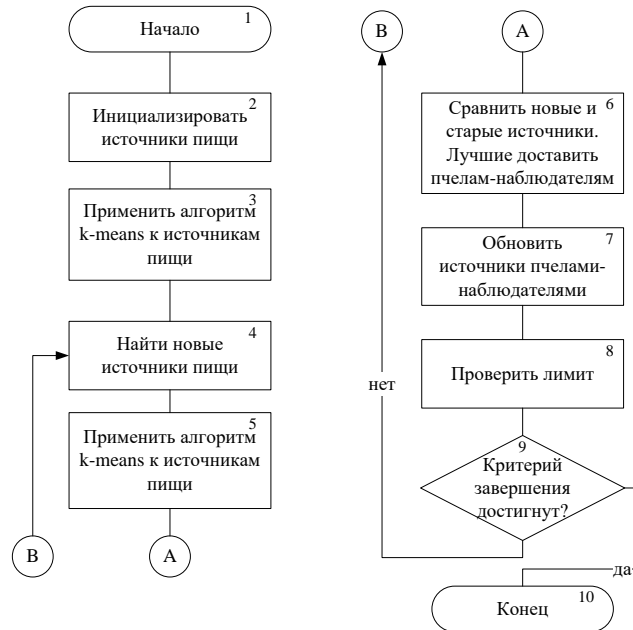


Рис. 1. Комбинированный пчелиный алгоритм кластеризации

Таким образом, в результате работы алгоритма на выход получаем оптимальный источник пищи, который представляет собой набор центроидов. Данный алгоритм повышает точность кластеризации за счет того, что изначальные центроиды генерируются не случайно, а отбираются на основе объективных оценок, с каждой итерацией становясь более качественными.

**2. Муравьиный алгоритм кластеризации.** Кластеризация с помощью муравьиного алгоритма – это распределенный процесс. В базовой модели муравьи представлены простыми агентами, которые случайным образом перемещаются в среде, представляющей собой квадратную сетку. Первоначально каждый объект данных (вектор) случайным образом распределяется по двумерному пространству. Элементы данных, которые разбросаны в этой среде, могут быть подобраны, перемещены и сброшены агентами с некоторой вероятностью [4, 16–20].

Предложенный алгоритм использует сетку с ячейками  $s*s$  для кластеризации, но искомые кластеры к ней не привязаны. Здесь  $k$  муравьев ищут  $k$  кластеров. Муравьи собирают объекты в своем списке грузов, поэтому нет необходимости в процессе сброса ноши. Количество ячеек должно быть больше, чем количество всех объектов. В начале кластеризации мы генерируем муравьев, в количестве равном количеству кластеров. Каждый муравей должен искать объекты одного кластера. Все муравьи движутся с разной скоростью по сетке. Максимальная скорость для всех муравьев определяется как:

$$v_{max} = \frac{s}{3}, \quad (3)$$

где  $s$  – сторона сетки.

Шаг движения каждого муравья генерируется случайным образом в интервале между 1 и  $v_{max}$  перед началом каждого движения по формуле:

$$step = 1 + (v_{max} - 1) * \beta, \quad (4)$$

где  $\beta$  – случайный коэффициент на отрезке  $[0, 1]$ .

Через каждые 200 итераций  $v_{max}$  уменьшается на 1. Функция схожести объектов определяется по формуле:

$$f(o_i) = \max \left[ 0, \frac{1}{\sigma^2} \sum_{o_j \in L} \left( 1 - \frac{d(o_i, o_j)}{1 + \frac{v-1}{v_{max}}} \right) \right], \quad (5)$$

где  $L$  – это список грузов, принадлежащих муравью, с возможностью добавить новый элемент или удалить из него существующий,  $\sigma$  – это размер  $L$ ,  $d(o_i, o_j)$  – расстояние по Хеммингу между объектами,  $v$  – скорость муравья,  $v_{max}$  – максимально допустимая скорость.

Функция вычисления вероятности подбора:

$$P_p = 1 - \text{Sigmoid}(f(o_i)). \quad (6)$$

Вероятность сброса объекта вычисляется по формуле:

$$P_p = \text{Sigmoid}(f(o_i)), \quad (7)$$

где  $f(o_i)$  определяется по формуле (5), а функция *Sigmoid* определяется как:

$$\text{Sigmoid}(x) = \frac{1 - e^{-x}}{1 + e^{-x}}. \quad (8)$$

Всякий раз, когда муравей находится в занятой ячейке, он пытается подобрать предмет и вставить его в свой список грузов в соответствии с формулой вставки. Аналогично, если муравей находится в пустой ячейке, он выполняет поиск в своем списке грузов и пробует сбросить один из своих грузов в соответствии с формулой удаления. Для повышения точности, на каждой итерации каждый муравей пытается объединить свои грузы с грузами в списках других муравьев. Для этого каждый муравей проверяет все свои грузы, используя функцию вероятности вставки, с целью вставить свои грузы в список другого муравья. Таким образом, неверно сгруппированные данные могут быть возвращены обратно в их правильные кластеры и ошибка кластеризации уменьшится. Это очень важный процесс, который предотвращает попадание муравьев в локальный оптимум.

Алгоритм можно представить в следующем виде:

1. Инициализировать число кластеров и сгенерировать соответствующее число муравьев;
  2. Распределить муравьев и объекты кластеризации случайно на сетке;
  3. Для каждого муравья:
    - 3.1 сгенерировать случайный шаг по формуле и переместить муравья на заданный шаг;
    - 3.2 если в новом положении муравья обнаружен груз (объект), вычислить вероятность подбора согласно формуле (6). Если полученная вероятность больше случайно сгенерированной вероятности, то добавить объект в лист грузов муравья;
    - 3.3 если в новом положении муравья груза не оказалось, то для всех грузов в списке муравья вычислить вероятность сброса груза согласно формуле (7). Если полученная вероятность больше случайно сгенерированной вероятности, то удалить объект из листа грузов муравья;
  4. Обновить грузы в списках муравьев. Для этого для каждого груза в списке каждого муравья вычислить вероятность вставки в список каждого муравья, отличного от текущего. Если вероятность вставки, вычисленная по формуле (6), превышает случайно сгенерированную вероятность, вставить объект в список выбранного муравья и удалить его из списка текущего;
  5. Пока не достигнуто максимальное число итераций, повторять шаги 3 и 4.
- Графически алгоритм представлен на рис. 2.



Рис. 2. Комбинированный муравьиный алгоритм кластеризации

В результате работы алгоритма на выходе получаем списки грузов каждого муравья. Каждый груз является объектом набора данных, а каждый муравей – отдельным кластером. Данный алгоритм повышает точность кластеризации за счет обмена муравьев объектами из списка грузов. Таким образом, объект, затерявшийся на плоскости и не встретивший муравья, к кластеру которого он принадлежит, имеет все шансы попасть в правильный кластер.

**3. Результаты экспериментальных исследований.** Для каждого из алгоритмов проводились тесты на выборках данных, находящихся в свободном доступе. Были выбраны следующие (самые популярные) наборы: «Ирисы», «Соевые бобы», «Дрожжи». Это реальные данные, представляющие собой статистику, собранную в разные годы.

Эксперимент заключался в следующем: для каждого из предложенных алгоритмов был выполнен запуск соответствующей программы на каждом из трех тестовых наборов данных, собраны результаты работы программы. Затем был запущен программный код алгоритма  $k$ -means и также собраны результаты. После этого собранные данные были оценены и проанализированы, были сделаны выводы об эффективности и точности разработанных алгоритмов.

Результаты по параметрам среднеквадратического отклонения, точности, F-меры и времени представленные в табл. 1–6 для трех наборов данных: «Ирисы», «Соевые бобы» и «Дрожжи».

Таблица 1

**Результаты выполнения эксперимента для набора данных «Ирисы»  
и комбинированного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	10,518	0,889	87,33 %	0,1
Предлагаемый алгоритм	0	0,897	89,25 %	23,7

Таблица 2

**Результаты выполнения эксперимента для набора данных «Соевые бобы»  
и комбинированного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	718,65	0,715	70,2 %	0,2
Предлагаемый алгоритм	188,13	0,73	75,47 %	35,8

Таблица 3

**Результаты выполнения эксперимента для набора данных «Дрожжи»  
и комбинированного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	51,32	0,402	69,71 %	0,13
Предлагаемый алгоритм	3,41	0,411	71,32 %	30,48

Таблица 4

**Результаты выполнения эксперимента для набора данных «Ирисы»  
и муравьиного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	10,52	0,888	87,32 %	0,1
Предлагаемый алгоритм	0	0,91	90,1 %	28,9

Таблица 5

**Результаты выполнения эксперимента для набора данных «Соевые бобы»  
и муравьиного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	718,68	0,715	70,23 %	0,2
Предлагаемый алгоритм	190,5	0,728	76,3 %	40,4

Таблица 6

**Результаты выполнения эксперимента для набора данных «Дрожжи»  
и муравьиного алгоритма**

Алгоритм	Среднеквадратическое отклонение	F-measure	Точность	Время (сек)
k-means	51,32	0,4	69,7 %	0,14
Предлагаемый алгоритм	3,4	0,426	71,51 %	31,2

Как можно заметить из полученных результатов, разработанные алгоритмы достигают лучших результатов по всем параметрам качества кластеризации: среднеквадратическому отклонению, точности и *F-measure*. Однако, с точки зрения вычислительных затрат эти алгоритмы требуют время вычисления, значительно превышающее время вычислений с помощью *k-means*. Такой результат обусловлен статистическим поведением всех биоинспирированных алгоритмов оптимизации.

**Заключение.** В данной статье рассмотрены два модифицированных алгоритма кластеризации на основе биоинспирированных алгоритмов. В первом разделе был рассмотрен комбинированный алгоритм решения проблемы кластеризации, основанный на сочетании биоинспирированного алгоритма пчелиной колонии и классического алгоритма *k-means*, а также были описаны ожидаемые преимущества, шаги выполнения и входные данные нового алгоритма. Во втором разделе приводится усовершенствованный алгоритм кластеризации на основе муравьиного алгоритма, и также описываются ожидаемые результаты. Были проведены эксперименты, представлены и проанализированы их результаты. Полученные результаты показывают, что оба разработанных алгоритма имеют преимущества перед классическим *k-means* по оценке качества кластеризации, но значительно проигрывают по времени выполнения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Чубукова И.А. Data Mining: учеб. пособие. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2006. – 327 с.
2. Гладков Л.А., Курейчик В.М., Курейчик В.В., Сороколетов П.В. Биоинспирированные методы в оптимизации: монография. – М.: Физмалит, 2009. – 384 с.
3. Анон М.Ф., Катыева Я.В., Михаличук В.И. Алгоритмы роя пчел и частиц в задаче обеспечения надежности по постепенным отказам // Наука и образование. – 2015. – № 1. – С. 144-157.
4. Сороколетов П.В. Методы и алгоритмы принятия решений на основе бионического поиска: монография. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – 352 с.
5. Kravchenko Y.A., Kureichik V.V. Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8<sup>th</sup> IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”. 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.
6. Gladkov L.A., Gladkova N.V., Legebokov A.A. Organization of knowledge management based on hybrid intelligent methods // Advances in Intelligent Systems and Computing. – 2015. – Vol. 349. – P. 107-112.
7. 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.
8. Курейчик В.М., Курейчик В.В. Эволюционные, синергетические и гомеостатические стратегии в искусственном интеллекте: состояние и перспективы // Новости искусственного интеллекта. – 2000. – № 3. – С. 39-67.



9. Bova V.V., Kureichik V.V., Lezhebokov A.A. The integrated model of representation of problem-oriented knowledge in information systems // 8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 – Conference Proceedings 8. – 2014. С. 7035923. – P. 111-114.
10. Бова В.В., Курейчик В.В. Интегрированная подсистема гибридного и комбинированного поиска в задачах проектирования и управления // Известия ЮФУ. Технические науки. – 2010. – № 12 (113). – С. 37-43.
11. Зайцев А.А., Курейчик В.В., Полупанов А.А. Обзор эволюционных методов оптимизации на основе роевого интеллекта // Известия ЮФУ. Технические науки. – 2010. – № 12 (113). – С. 7-12.
12. Курейчик В.В., Курейчик Вл.Вл. Биоинспирированный поиск при проектировании и управлении // Известия ЮФУ. Технические науки. – 2012. – № 11 (136). – С. 178-183.
13. Гладков Л.А., Курейчик В.М., Курейчик В.В. Генетические алгоритмы. – М.: Физматлит, 2006. – 320 с.
14. Родзин С.И., Курейчик В.В. Состояние, проблемы и перспективы развития биоэвристик // Программные системы и вычислительные методы. – 2016. – № 2. – С. 158-172.
15. Родзин С.И., Курейчик В.В. Теоретические вопросы и современные проблемы развития когнитивных биоинспирированных алгоритмов оптимизации // Кибернетика и программирование. – 2017. – № 3. – С. 51-79.
16. Курейчик В.В., Родзин С.И. О правилах представления решений в эволюционных алгоритмах // Известия ЮФУ. Технические науки. – 2010. – № 7 (108). – С. 13-21.
17. Запорожец Д.Ю., Кудяев А.Ю., Лежебоков А.А. Многоуровневый алгоритм решения задачи параметрической оптимизации на основе биоинспирированных эвристик // Известия Кабардино-Балкарского научного центра РАН. – 2013. – № 4 (54). – С. 21-28.
18. Курейчик В.М. Особенности построения систем поддержки принятия решений // Известия ЮФУ. Технические науки. – 2012. – № 7 (132). – С. 92-98.
19. Курейчик В.М., Кажаров А.А. Использование шаблонных решений в муравьиных алгоритмах // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 11-17.
20. Gladkov, L.A., Kravchenko Y.A., Kureichik V.V. Evolutionary Algorithm for Extremal Subsets Comprehension in Graphs // World Applied Sciences Journal. – 2013. – Vol. 27 (9). – P. 1212-1217.

#### REFERENCES

1. Chubukova I.A. Data Mining: ucheb. posobie [Data Mining: tutorial]. Moscow: Internet-Universitet Informacionnykh Tekhnologiy; BINOM. Laboratoriya znaniy, 2006, 327 p.
2. Gladkov L.A., Kureichik V.M., Kureichik V.V., Sorokoletov P.V. Bioinspirovannyye metody v optimizatsii: monografiya [Bioinspired methods in optimization: monograph]. Moscow: Fizmalit, 2009, 384 p.
3. Anop M.F., Katueva Ya.V., Mikhailichuk V.I. Algoritmy roya pchel i chastic v zadache obespecheniya nadezhnosti po postepennym otkazam [Algorithms of swarm of bees and particles in the problem of reliability on gradual failures], *Nauka i obrazovanie* [Science and education], 2015, No. 1, pp. 144-157.
4. Sorokoletov P.V. Metody i algoritmy prinyatiya resheniy na osnove bionicheskogo poiska: monografiya [Methods and algorithms of decision-making based on bionic search: monograph]. Taganrog: Izd-vo TTI YuFU, 2009, 352 p.
5. Kravchenko Y.A., Kureichik V.V. Knowledge management based on multi-agent simulation in informational systems, *Conference proceedings. 8<sup>th</sup> IEEE International Conference "Application of Information and Communication Technologies – AICT 2014". 15-17 October 2014, Astana, Kazakhstan*, pp. 264-267.
6. Gladkov L.A., Gladkova N.V., Legebokov A.A. Organization of knowledge management based on hybrid intelligent methods, *Advances in Intelligent Systems and Computing*, 2015, Vol. 349, pp. 107-112.
7. Kureychik V.M., Kureichik V.V., Rodzin S.I. Modeli parallelizma evolyucionnykh vychisleniy. [Models of parallelism of evolutionary calculations], *Vestnik Rostovskogo gosudarstvennogo universiteta puty soobshcheniya* [Vestnik RGUPS], 2011, No. 3 (43), pp. 93-97.

8. Kureychik V.M., Kureychik V.V. Evolyucionnyye, sinergeticheskie i gomeostaticheskie strategii v iskusstvennom intellekte: sostoyanie i perspektivy [Evolutionary, synergetic and homeostatic strategies in artificial intelligence: state and prospects], *Novosti iskusstvennogo intellekta* [News of artificial intelligence], 2000, No. 3, pp. 39-67.
9. Bova V.V., Kureichik V.V., Lezhebokov A.A. The integrated model of representation of problem-oriented knowledge in information systems, *8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 – Conference Proceedings 8*, 2014. С. 7035923, pp. 111-114.
10. Bova V.V., Kureychik V.V. Integrirovannaya podsistema gibridnogo i kombinirovannogo poiska v zadachakh proektirovaniya i upravleniya [Integrated subsystem of the hybrid and combined search in problems of design and management], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12 (113), pp. 37-43.
11. Zaycev A.A., Kureychik V.V., Polupanov A.A. Obzor evolyucionnykh metodov optimizatsii na osnove roevogo intellekta [Overview of evolutionary optimization techniques based on swarm intelligence], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12 (113), pp. 7-12.
12. Kureychik V.V., Kureychik V.V. Bioinspirirovannyy poisk pri proektirovanii i upravlenii [Biospherology search in the design and management], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 11 (136), pp. 178-183.
13. Gladkov L.A., Kureychik V.M., Kureychik V.V. Geneticheskie algoritmy [Genetic algorithm]. Moscow: Fizmatlit, 2006, 320 p.
14. Rodzin S.I., Kureychik V.V. Sostoyanie, problemy i perspektivy razvitiya bioevristik [State, problems and prospects of bio-heuristics development], *Programmnye sistemy i vychislitel'nye metody* [Software systems and computational methods], 2016, No. 2, pp. 158-172.
15. Rodzin S.I., Kureychik V.V. Teoreticheskie voprosy i sovremennyye problemy razvitiya kognitivnykh bioinspirirovannykh algoritmov optimizatsii [Theoretical questions and contemporary problems of the development of cognitive bio-inspired algorithms for optimization], *Kibernetika i programmirovaniye* [Cybernetics and programming], 2017, No. 3, pp. 51-79.
16. Kureychik V.V., Rodzin S.I. O pravilakh predstavleniya resheniy v evolyucionnykh algoritmakh [About the rules for the submission of solutions in evolutionary algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 7 (108), pp. 13-21.
17. Zaporozhec D.Yu., Kudaev A.Yu., Lezhebokov A.A. Mnogourovnevyy algoritm resheniya zadachi parametricheskoy optimizatsii na osnove bioinspirirovannykh evristik [Multi-level algorithm for solving the problem of parametric optimization based on bioinspired heuristics], *Izvestiya Kabardino-Balkarskogo nauchnogo centra RAN* [Izvestiya of Kabardino-Balkar scientific center of RAS], 2013, No. 4 (54), pp. 21-28.
18. Kureychik V.M. Osobennosti postroeniya sistem podderzhki prinyatiya resheniy [Features of construction of systems of support of acceptance of decisions], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 7 (132), pp. 92-98.
19. Kureychik V.M., Kazharov A.A. Ispol'zovanie shablonykh resheniy v murav'inykh algoritmakh [The use of standard solutions in ant colony optimization algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 11-17.
20. Gladkov, L.A., Kravchenko Y.A., Kureichik V.V. Evolutionary Algorithm for Extremal Subsets Comprehension in Graphs, *World Applied Sciences Journal*, 2013, Vol. 27 (9), pp. 1212-1217.

Статью рекомендовал к опубликованию д.т.н., профессор В.И. Финаев.

**Марков Владимир Васильевич** – Южный федеральный университет; e-mail: vvmarkov@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

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

**Кузьмина Мария Анатольевна** – e-mail: kuzmina.maria.tti@gmail.com; кафедра систем автоматизированного проектирования; магистрант.

**Markov Vladimir Vasilievich** – Southern Federal University; e-mail: vvmarkov@sfnedu.ru; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

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

**Kuzmina Maria Anatolyevna** – e-mail: kuzmina.maria.tti@gmail.com; the department of computer aided design, graduate student.

УДК 002.53:004.89

DOI 10.23683/2311-3103-2018-4-185-197

**А.Н. Нацкевич, И.О. Курситыс**

### **КОМБИНИРОВАННЫЙ БИОИНСПИРИРОВАННЫЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАСТЕРИЗАЦИИ ДАННЫХ\***

*Статья посвящена решению одной из популярнейших задач интеллектуального анализа данных – задачи кластеризации. Кластеризация – объединение в группы схожих объектов – является одной из фундаментальных задач в области анализа данных и Data Mining. Список прикладных областей, где она применяется, широк: сегментация изображений, маркетинг, борьба с мошенничеством, прогнозирование, анализ текстов и многие другие. Решение данной задачи приобретает особую актуальность в условиях постоянно растущего объема генерируемых, передаваемых и обрабатываемых данных. Авторами исследована задача кластеризации, приведены постановка задачи, основные формулы для ее решения, а так же целевая функция. Проведен аналитический обзор существующих алгоритмов, таких как: алгоритмы иерархической кластеризации, квадратичной ошибки, алгоритмы k-means и c-means, алгоритмы, основанные на теории графов. Отмечены основные достоинства и недостатки рассмотренных алгоритмов. Предложено использовать методы биоинспирированного поиска для решения задачи кластеризации, обоснована актуальность применения биоинспирированных моделей и методов для решения NP-полных задач, к классу которых относится и исследуемая задача. Отмечен вклад ученых в решение данной проблемы – биоинспирированные алгоритмы, такие как метод роя частиц, муравьиный алгоритм, пчелиный алгоритм, алгоритм бактериальной оптимизации, алгоритм кукушки и многие другие подобные методы успешно применяются для решения задачи кластеризации. Предложен комбинированный биоинспирированный алгоритм, применяющий последовательно муравьиный алгоритм и алгоритм летучих мышей. Раскрыты основные идеи алгоритмов, приведены схемы решения задачи, кодирования решений. Реализован метод локального поиска для алгоритма летучих мышей. Проведены экспериментальные исследования на тестовых примерах (бенчмарках), которые доказывают эффективность разработанного алгоритма по сравнению с алгоритмами k-средних и генетическим алгоритмом. В ходе проведения экспериментов определена временная сложность разработанного комбинированного биоинспирированного алгоритма. Предложенную модель комбинированного решения задачи планируется в дальнейшем использовать для работы ранее разработанного бустинга алгоритмов, который работает с несколькими алгоритмами и позволяет найти лучшее решение из решений, полученных разными биоинспирированными алгоритмами.*

*Кластеризация; биоинспирированный алгоритм; муравьиный алгоритм; алгоритм летучих мышей; роевой алгоритмы; роевой интеллект; искусственный интеллект.*

---

\* Работа выполнена при поддержке РФФИ (проект № 16-07-00703)..