Transaction of Azerbaijan National Academy of Sciences, Series of Physical-Technical and Mathematical Sciences: Informatics and Control Problems, Vol. XXXVI, No.3, 2016
www.icp.az/2016/3-09.pdf

УДК 519.7

Р.И. ДАВУДОВА

ОБ ОДНОМ ЭВОЛЮЦИОННОМ ПОДХОДЕ К КЛАСТЕРНОМУ АНАЛИЗУ

Рассматривается задача построения оптимальной автоматической классификации объектов, характеризующихся количественными признаками. Работа посвящена применению эволюционных алгоритмов (ЭА) в задачах кластерного анализа. Для решения рассмотренной задачи кластерного анализа предложен прямой подход, базирующийся на эволюционных вычислениях по принципу Ламарка. Оценена вычислительная сложность предложенного алгоритма.

Ключевые слова: функционал, качество, эволюция, ламаркизм, кроссовер, мутация, популяция, особь, потомок

1. Введение. Кластерный анализ является методом разбиения выборки объектов и явлений на группы по каким-либо критериям, другими словами, цель кластерного анализа – типологическая группировка совокупностей массовых явлений на основе множества признаков. Классификация объектов позволяет разделить исходный набор исследуемых объектов на группы объектов, таким образом, чтобы каждый объект был более схож с объектами из своей группы, чем с объектами других групп. При больших размерах выборки задача построения оптимальной классификации становится трудоемкой. Для повышения эффективности и качества решения трудоемких и труднорешаемых задач искусственного интеллекта и функциональной оптимизации предлагается новый метод и алгоритм, основанные на использовании эволюционных и генетических методов поиска. Концепция эволюционных вычислений основана на некоторых формализованных принципах естественного эволюционного процесса биологических организмов.

В настоящей статье рассматривается задача кластерного анализа объектов, характеризующихся количественными признаками. В [1, 2] для решения рассмотренной задачи были использованы методы Хука-Дживса, Розенброка и их модификации, исходя из природы поставленных задач. Предложенный алгоритм в [3, 4, 5] оптимизирует начальную классификацию отнесением набора объектов из одного класса в другой.

В общем случае, в задаче кластерного анализа функционал качества классификации является ступенчатым и не обладает свойством унимодальности и дифференцируемости, что исключает возможность применения классических методов оптимизации. Использование указанных методов оптимизации порождает определенные проблемы, связанные со сходимостью итерационных процессов. Поэтому разработка более эффективных методов и алгоритмов решения кластерного анализа является актуальным вопросом.

В работах [6-9] для решения задачи АК предлагается алгоритмический подход, состоящий из двух эвристических алгоритмов – динамического алгоритма, входящего в семейство алгоритмов вычисления оценок (АВО) [10] и самоорганизующегося генетического алгоритма (СГА), который выбирает приоритетную схему многородительской репродукции и эффективного числа родителей [11]. А алгоритм, предложенный в [12, 13], состоит из АВО и эволюционного алгоритма, базирующегося на ламаркизме [14, с.29-32].

Рассмотренная задача кластерного анализа решается в рамках эволюционных вычислений.

2. Постановка задачи. В работе рассматривается задача кластерного анализа объектов, описанных количественными признаками [12, с.74; 13, с.15-16].

Пусть дано множество допустимых однотипных сложных объектов

$$X = \{x_1, x_2, \dots, x_m\}\ , \tag{2.1}$$

состояния которых описываются набором некоторых числовых признаков

$$T = \|x_{ij}\|;\ x_{ij} \in M_j \subset R : i = \overline{1,m};\ j = \overline{1,n} \tag{2.2}$$

Здесь T – прямоугольная матрица, x_{ij} j-й признак i-го объекта, M_j – ограниченная область изменения j-го признака. Таблицу T называют признаковой матрицей. Каждый объект x_i из (2.1) можно определить отображением $M_1 \times M_2 \times \ldots \times M_n = M^n \subset R^n$,, где R^n – пространство n-мерных векторов действительных чисел.

Множество объектов (2.1) считается допустимым, если признаки (2.2) определены областью значений M_j, т.е. для $\forall\ i\ \ x_{ij} \in M_j$, где $i = \overline{1,m};\ j = \overline{1,n}$. Однотипными понимаются системы, описание состояний которых дано в одном и том же признаковом пространстве.

Определение 1. Подмножества $K_i, i = \overline{1,l}$ набора объектов X, удовлетворяющие следующим условиям, называются классами:

1) $\forall\ i,j\ \ (i \neq j) \rightarrow (K_i \cap\ K_j = \emptyset)$;

2) $\bigcup_{i=1}^{l} K_i = X$

3) Для $x_q \in X\ \ \left((x_q \in K_i) \rightarrow\rightarrow \neg\exists\ j \neq i\ \ \left(\left(D(x_q, K_j) > D(x_q, K_i)\right)\right)\right)$, где $D(x_q, K_j)$ – мера сходства объекта x_q и класса K_j.

4) Для $x_p, x_q \in X\left((x_p, x_q \in K_i;\ x_t \in K_j) \rightarrow\rightarrow \neg\exists\ j \neq i\ \left(D(x_q, x_t) > D(x_q, x_p)\right)\right.$, и $\left.D(x_p, x_t) > D(x_q, x_p)\right)$, где $D(x_q, x_t)$ – мера сходства объектов x_q и x_t.

Предполагается, что количество классов *l* до классификации является известным.

Таким образом, задача кластерного анализа состоит в том, чтобы по описанию допустимых объектов (2.2) вычислить значения предикатов $P_q(x_i)$ "$x_i \in K_q$", $q = \overline{1,l}$, где $P_q(x_i) \in \{0,1,\Delta\}$, ,т.е.

$$P_q(x_i) = \begin{cases} 1, & \text{если } x_i \in K_q, \\ 0, & \text{если } x_i \notin K_q, \\ \Delta, & \text{если алгоритм отказывается от принад} - \\ & \text{лежности объекта } x_i \text{ к классам } K_q, q = \overline{1,l}. \end{cases}$$

3. Методы решения. В данной статье рассмотренная задача решается в рамках эволюционных вычислений [14-16].

Эволюционная теория утверждает, что каждый биологический вид (биологическая популяция) целенаправленно развивается и изменяется в течение нескольких поколений. История эволюционных вычислений началась с разработки ряда различных независимых моделей эволюционного процесса. В отличие от эволюции, происходящей в природе, эволюционные алгоритмы только моделируют те процессы в популяции, которые являются существенными для развития.

Основной механизм эволюции Дарвина – естественный отбор. По принципу Чарльза Дарвина биологические популяции развиваются в течение нескольких поколений, подчиняясь законам естественного отбора и "выживает наиболее приспособленный". Генетические алгоритмы основаны на процессах биологических организмов: биологические популяции развиваются в течение нескольких поколений, подчиняясь законам естественного отбора, то есть по принципу "выживает наиболее приспособленный". Основной идеей Ламарка было то, что организмы изменяются под воздействием окружающей среды и условий их жизнедеятельности. Главное отличие от дарвинской теории состояло в том, что по Ламарку виды могут изменяться в течение своей жизни, а не только на генетическом уровне. По теории эволюции Ламарка характерные особенности организма, приобретенные в результате его адаптации на протяжении срока жизни этого организма, передаются его потомкам.

В работе [14, с.13] дается такая классификация моделей эволюции, на которых базируются эволюционные алгоритмы:

модель эволюции Ч. Дарвина – процесс, посредством которого особи некоторой популяции, имеющие более высокое функциональное значение (с сильными признаками), получают большую возможность для воспроизведения потомков, чем «слабые» особи. Такой механизм часто называют методом «выживания сильнейших»;

ламаркизм или модель эволюции Ж. Ламарка. Им предложена теория, основанная на предположении, что характеристики, приобретенные особью (организмом) в течение жизни, наследуются его потомками. В отличие от простого генетического алгоритма данная модель оказывается наиболее эффективной, когда популяция имеет тенденцию сходимости в область локального оптимума;

сальтационизм (модель эволюции де Фриза). В основе этой модели лежит моделирование социальных и географических катастроф, приводящих к резкому изменению видов и популяций. Эволюция, таким образом, представляет собой последовательность скачков в развитии популяции без предварительного накопления количественных изменений в эволюционных процессах;

модель К. Поппера, который рассматривал эволюцию как развивающуюся иерархическую систему гибких механизмов управления, в которых мутация интерпретируется как метод случайных проб и ошибок, а отбор, как один из способов управления с помощью устранения ошибок при взаимодействии с внешней средой;

синтетическая теория эволюции, описанная Н. Дубининым (попытка интеграции различных моделей эволюции, в том числе Ч. Дарвина, Ж. Ламарка и де Фриза). Ее кардинальным положением является признание стохастичности процессов мутации и больших резервов рекомбинационной изменчивости. Условия внешней среды – не только факторы исключения неприспособленных индивидов из популяции, но и формирующие особенности самой синтетической теории эволюции.

Для решения рассмотренной задачи кластерного анализа предлагается подход, базирующийся на эволюционных вычислениях по принципу Ламарка [14, с.29-32].

Меры сходств объектов, объекта к классу и классов выбирается следующим образом [17, с.218-219; 18]:

Таблица 1

Мера сходства объектов x_p и x_q	$d(x_p, x_q) = \sqrt[s]{\frac{1}{n}\sum_{t=1}^{n}(x_{pt} - x_{qt})^{s}}$–взвешенное $l_s -$ норма (расстояние Минковского порядка s) между точками x_p и x_q в многомерном признаковом пространстве
Мера сходства объекта $x_p \in X$ и класса (множества объектов) $K_i \in K$	$D(x_p, K_i) = \sqrt[s]{\frac{1}{n}\sum_{t=1}^{n}(x_{pt} - \overline{x_{it}})^{s}}$– взвешенное $l_s -$норма между объектом x_p и «средним объектом» класса K_i в многомерном признаковом пространстве

Мера сходства классов (множества объектов) $K_i, K_j \in K$	$D(K_i, K_j) = \sqrt[s]{\frac{1}{n} \sum_{t=1}^{n} (\overline{x_{it}} - \overline{x_{jt}})^{s}}$ – взвешенное l_s –норма между «средними объектами» классов K_i и K_j в многомерном признаковом пространстве

А критерий качества классификации, как в “центроидном” методе, выбирается в виде максимизации меры сходства групп внутри классификации [17, с.221]:

$$D(K) = \sum_{\substack{K_i, K_j \in K \\ K_i \neq K_j \\ \cup_{s=1}^{l} K_s = K}} D(K_i, K_j) = \sum_{\substack{K_i, K_j \in K \\ K_i \neq K_j \\ \cup_{s=1}^{l} K_s = K}} \sqrt[s]{\frac{1}{n} \sum_{t=1}^{n} (\overline{x_{it}} - \overline{x_{jt}})^{s}} \rightarrow max \qquad (3.1)$$

где $\overline{x_i}$ – «центральный» объект класса K_i.

Определение 2. Классификация, полученная из решения задачи (3.1) и удовлетворяющая условиям 1) - 4) определения 1, называется оптимальной.

В данной работе, для решения поставленной задачи кластерного анализа, применяется итеративный подход, базирующийся на эволюционном принципе Ж.Ламарка

Кратко изложим модифицированный эволюционный алгоритм по принципу Ламарка [12, с.29-32; 13, с.17]:

Шаг 0 : Положить $k=0$ и перейти к шагу 1.

Шаг 1 : Генерируется случайная начальная популяция. Перейти к шагу 2.

Шаг 2 : Вычисляется функция приспособленности**.** Перейти к шагу 3.

Шаг 3 : Случайно выбираются из популяции $к\_m$ особей (индивидов) и d_i битов для многоточечной мутации. Перейти к шагу 4.

Шаг 4 : Проводится многоточечная мутация. Полученные особи останутся в популяции (заменяя своих «родителей»). Перейти к шагу 5.

Шаг 5 : Случайно выбирается первый член пары скрещивания, а второй и третий член будут самые близкие особи к нему. Перейти к шагу 6.

Шаг 6 : Проводится операция трехточечного кроссинговера между тремя выбранными особями (инбридинг). Полученные потомки заменяют в популяции своих родителей. Перейти к шагу 7.

Шаг 7 : Выбираются максимально далекие особи. Перейти к шагу 8.

Шаг 8 : Проводится операция трехточечного кроссинговера между тремя выбранными особями (аутбридинг). Полученные потомки заменяют в популяции своих родителей. Перейти к шагу 9.

Шаг 9 : Вычисляется целевая функция – функционал качества классификации.

Шаг 10 : Проверить условия для завершения алгоритма. Если условие удовлетворяется, то остановиться, иначе положить $k=k+1$ и перейти к шагу 3.

В статье понимается под геном – объект, хромосомом – класс, индивидом или особью – классификация. И популяция состоит из множества случайных классификаций.

Вычислительная сложность алгоритма. Был проведен сравнительный анализ вычислительной сложности предложенного алгоритма ($\mathcal{A}5$) с алгоритмами $\mathcal{A}1$, $\mathcal{A}2$, $\mathcal{A}3$, $\mathcal{A}4$ и метода К-средних, где алгоритм $\mathcal{A}1$ уточняет начальную классификацию перераспределением между классами по одному объекту [4], алгоритм $\mathcal{A}2$ оптимизирует начальную классификацию отнесением набора объектов из одного класса в другой [3], алгоритм $\mathcal{A}3$, состоит из суперпозиций АВО и СГА (АВО – алгоритм вычисления оценок, СГА - самоорганизующийся генетический алгоритм) [5, 8], а алгоритм $\mathcal{A}4$ состоит из АВО и эволюционного алгоритма, базирующегося на ламаркизме [12, 13]. Результаты проведенного анализа представлены в таблице 2.

Таблица 2

№	Алгоритм	Вычислительная сложность
1.	$\mathcal{A}1$	$O(m \cdot n \cdot l^2)$
2.	$\mathcal{A}2$	$O(m^2 \cdot n \cdot l^2)$
3.	$\mathcal{A}3$	$O(m^3 \cdot n^2 \cdot l \cdot p^2)$
4.	$\mathcal{A}4$	$O(m^2 \cdot n^2 \cdot l \cdot p^2)$
5.	Метод K-средних	$O(m \cdot n \cdot l^2)$
6.	$\mathcal{A}5$	$O(m \cdot n \cdot l \cdot p)$

Здесь $m,\ n,\ l -$ количество объектов, признаков и классов соответственно, а $p -$ размер популяции.

***Численная реализация* алгоритма.** Для численной реализации предложенного алгоритма составлен MatLab M-файл. С целью построения оптимальной классификации скважин по отношению к солеобразованию проводились вычислительные эксперименты с данными 58 скважин месторождений Грязевой Сопки и Песчаного, которые описываются нижеследующими 16 числовыми признаками:

1. Среднесуточный дебит нефти ($T/\text{сут}$);
2. Процент обводненности (%);
3. Среднесуточная вода (T);
4. Пластовое давление ($ат$);
5. Забойное давление ($ат$);
6. Разность температур – забойной и устьевой (C^O);
7. Температура устья (C^O);
8. Температура забойная (C^O);
9. Содержание Cl^- ($мг{-}экв./л$);
10. Содержание SO_4^- ($мг{-}экв./л$);
11. Содержание HCO_3 ($мг{-}экв./л$);
12. Содержание CO_3 ($мг{-}экв./л$);
13. Содержание HK ($мг{-}экв./л$);
14. Содержание Ca ($мг{-}экв./л$);
15. Содержание Mg ($мг{-}экв./л$);
16. Содержание $Na + K$ *(% вес)*.

Эксплуатационные показатели скважин Грязевой Сопки и Песчаного показаны в таблицах 3 и 4 соответственно.

Transaction of Azerbaijan National Academy of Sciences, Series of Physical-Technical and Mathematical Sciences: Informatics and Control Problems, Vol. XXXVI, No.3, 2016
www.icp.az/2016/3-09.pdf

Таблица 3

№	№ скважина	Среднесуточный дебит нефти	Процент обводненности	Среднесуточная вода	Пластовое давление	Забойное давление	Температура забойная	Температура устья	Разность температур – забойной и	Cl^-	SO_4^-	HCO_3	CO_3	HK	Ca	Mg	$Na+K$
1.	11	3.88	58.34	5.45	40.8	65.9	30.	22.	42.	28.8	0.1	2.8	0.5	0.2	0.6	0.3	32.7
2.	318	0.120	62.5	0.2	53.2	81.7	63.	45.	18	21.51	0.1	6.8	-	0.5	0.6	0.4	28.1
3.	321	11.64	5.74	0.71	60.2	82.5	59.74	43.	16.74	18.5	0.2	6.	0.8	0.3	0.2	0.2	18.5
4.	362	0.129	60.34	0.193	85.7	96.5	59.7	41.	18.7	27.7	0.2	3.7	0.4	0.1	0.7	0.6	31.1
5.	460	0.096	66.67	0.195	48.9	72.3	45.8	32.	13.8	27.5	0.8	2.8	-	0.9	1.1	0.1	31.5
6.	1001	1.367	96.419	36.8	78.6	110.5	51.	42.	9.	29.9	0.2	4.5	-	-	0.6	0.4	33.9
7.	1003	1.962	90.229	18.115	71.4	102.9	37.9	26.	11.9	22.2	0.1	6.4	-	0.7	1.2	0.9	27.6
8.	1004	0.645	97.942	30.709	20.4	59.3	57.4	45.8	11.6	28.3	0.1	4.5	-	0.5	1.1	0.8	31.8
9.	1005	2.733	89.646	23.666	67.3	103.3	58.	34.3	15.7	28.2	0.1	3.6	-	0.1	0.8	0.2	31.5
10.	1008	0.12	50.01	0.12	27.55	42.1	43.1	34.6	8.5	18.7	0.1	5.4	0.5	0.2	0.5	0.5	24.3
11.	1014	0.45	97.51	17.68	40.8	63.1	53.	32.4	21.6	18.4	0.4	5.9	0.8	0.1	0.3	0.5	16.7
12.	1017	1.52	95.165	29.92	45.19	54.4	56.	38.3	11.7	21.2	0.1	5.8	1	0.2	0.3	0.8	27.4
13.	1018	0.065	60.12	0.0967	39.8	67.1	61.	42.7	18.3	17.7	0.6	4.3	0.9	0.5	0.5	0.6	23.3
14.	1023	1.161	90.348	10.870	81.6	1103	52.6	44.8	17.8	20.	0.1	5.2	1	0.1	0.5	0.5	25.6
15.	1030	7.321	8.889	0.7142	63.3	81.6	39.	33.5	5.5	17.5	0.1	5.9	0.9	0.1	0.4	0.4	24.1
16.	1032	1.774	71.204	4.387	73.9	114.6	45.8	39.	5.2	22.	0.2	3.9	1.4	0.2	0.3	0.6	26.9
17.	1036	0.571	99.216	72.285	96.7	153.6	51.6	46	5.6	21.5	0.1	5.6	0.9	-	0.5	0.5	27.4
18.	1092	2.32	65.21	4.35	46.9	55.7	48.65	35.	13.65	18.6	0.4	7.5	0.2	0.4	0.2	0.9	25.4
19.	1109	12.387	4.950	0.645	26.5	42.7	34.39	22.	12.39	23.4	-	4.4	1.	0.4	0.6	1.1	22.8

Таблица 4

№	№ скважина	Среднесуточный дебит нефти	Процент обводненности	Среднесуточная вода	Пластовое давление	Забойное давление	Температура забойная	Температура устья	Разность температур – забойной и устьевой	Cl^-	SO_4^-	HCO_3	CO_3	HK	Ca	Mg	$Na+K$
1.	2	6	62.5	10	13.4	35.2	52	35	17	25.9	5.1	0.4	2.2	2.1	0.4	0.4	26.4
2.	4	20.17	54.06	23.73	21.6	43.7	61	49	12	19.5	0.8	5.9	–	0.5	0.1	0.1	26.6
3.	5	16.52	71.93	42.32	17.3	40.5	57	30	27	25.9	1.6	–	–	–	–	–	–
4.	6	38.55	27.62	14.71	26.7	59.2	67	51	16	21.9	0.7	4.7	1.2	1.6	0.2	0.2	29.9
5.	7	0.13	75	0.4	12.5	48.5	79	58	21	65.9	0.7	–	–	–	–	–	–
6.	8	0.5	97.91	23.37	7.9	38.1	80	68	12	15.4	0.8	8.7	–	0.5	0.3	0.32	20.5
7.	9	12.19	10.85	1.48	7.9	19.2	68	57	11	124.8	1.2	0.3	–	–	17.9	1.5	106
8.	10	5.37	97.36	198	26	44	43	27	16	56.8	2.7	–	–	–	–	–	–
9.	11	4.16	98.46	266	23	45	48	39	9	10.4	3	2.1	0.9	2.7	0.3	0.3	18.8
10.	12	3	91.3	31.5	28	37.9	80	64	16	28.7	1.8	–	–	–	–	–	–
11.	16	0.06	85.71	0.39	64	42.9	67	53	14	151.2	0.5	0.5	–	–	11.2	17.1	123.9
12.	17	1.53	99.22	196.03	16.2	47.3	78	69	9	2.53	2.6	5.	–	0.4	0.7	0.7	32.6
13.	20	2.42	97.34	88.58	14.6	35.8	48	35	13	39.7	3.1	–	–	–	–	–	–
14.	22	0.79	97.45	30	11.9	32.5	44	29	15	74.3	1.1	2.2	3.4	0.6	0.1	0.1	81.4
15.	24	1.74	93.47	24.94	18.5	42.8	46	36	10	87.5	1.4	–	–	–	–	–	–
16.	25	38.33	38.14	23.63	19.5	45.8	55	43	12	50	2.5	8.9	–	0.4	0.4	0.4	64.4
17.	29	0.03	96.15	0.81	32.1	41.8	91	76	15	72.8	1.8	–	–	–	–	–	–
18.	30	24.61	68.92	54.58	56.2	67.4	94	78	16	74.8	0.6	5.8	–	0.4	0.1	0.1	81.4
19.	32	19.96	53.81	23.26	61.8	79.3	93	80	13	93.7	2.4	–	–	–	–	–	–
20.	34	8.68	50.81	8.96	46.8	59.3	83	78	15	61.3	0.9	6.9		0.7	0.1	0.1	68.8
21.	36	2.09	52.08	2.27	21.1	56.5	82	75	7	12.6	1.8	6.8	0.7	0.7	0.4	0.4	21.1
22.	38	1.37	83.6	6.97	37.4	56.9	72	62	10	52.9	1.6	–	–	–	–	–	–
23.	41	8.82	94.51	151.91	19.4	28.4	77	63	14	93.6	0.1	1.5		0.6	1.5	1.5	94.3
24.	43	8.39	57.65	11.42	35.3	52.9	68	59	9	69.2	2.5	–	–	–	–	–	–

25.	48	0.13	75	0.4	57.	64	62	48	14	55.5	0.4	8.8		1.4	0.1	0.1	65.2
26.	51	1.33	95.48	28.14	19.4	50.8	87.8	77	11	14.3	8.2	2.4	0.4	3.1	0.2	0.2	28.1
27.	52	1.33	95.48	28.14	17.	32	74	67	17	84.7	0.9	–	–	–	–	–	–
28.	53	0.08	83.33	0.42	10.5	36.9	71	60	11	19.7	1.7	5.2		0.8	0.2	0.2	27.2
29.	58	0.71	93.37	10.51	53.	64.6	69	57	12	93.7	2.4	–	–	–	–	–	–
30.	59	0.89	52.94	1.89	18.8	43.1	65	49	16	10.4	1.7	4.5	–	0.6	0.2	0.2	16.9
31.	62	43.52	70.67	104.84	25.3	43.8	70	57	13	49.6	1.1	–	–	–	–	–	–
32.	63	5.35	90.77	52.58	8.6	30.4	71	54	17	69.1	1.2	–	–	–	–	–	–
33.	65	52.1	20.01	13.03	16.6	40.6	72	60	12	47.7	0.9	9.6	–	0.5	0.1	0.1	58.6
34.	70	8.65	44.74	7.67	7.4	27.4	78	67	11	21.6	4.5	0.5	1.4	0.8	0.4	0.4	28.4
35.	71	1.5	56.36	1.05	17.1	32.5	93	78	15	23.3	1.4	3.7	1.3	1.1	0.2	0.2	30.6
36.	90	16.68	45.64	14.39	17.7	41.1	86	70	16	44.3	1.6	3.8	0.6	0.4	0.2	0.2	50.5
37.	102	12.35	47.1	11.45	11.5	26.9	97	85	12	29.2	5.5	1.	1.5	1.	0.2	0.2	38.
38.	121	21.73	45.3	18.27	15.2	37.1	84	60	16	93.4	0.1	2.7	–	0.4	0.8	0.8	85.8
39.	406	9.94	47.5	9.35	4.1	24.3	90	72	18	14.5	1.1	2.9	0.9	2.	0.4	0.4	24.1

В таблице 5 приведены результаты вычислений, проводимых с нормированнными данными исследуемых скважин:

Таблица 5

№ итер.	Объекты, входящие в кластеры	$n(K_i)$	$D(\widetilde{K})$
51.	I-{1,3,5,8,9,10,12,14,22,44,45,50,51,53,55,56,57,58} II – {18,25,30,32,33,35,37,38,39,40,42,43,46,49,52} III – {2,4,6,7,11,13,16,20,23,24,27,28,47,48,54} IV – {15,17,19,21,26,29,31,34,36,41}	18 15 15 10	181.527
58.	I – {2,3,4,5,6,8,9,12,14,17,18,20,23} II – {19,26}; III – {25,28,30,31,32,33,35,36,39,40,42,44,45,46,49, 51,52,53,55,56}; IV – {1,7,10,11,13,15,16,21,22,24,27,29,34,37,38,41, 43,47,48,50,54,57,58}.	13 2 20 23	427.122
59.	I – {1,2,3,4,5,6,7,9,11,12,13,14,15,16,17,18} II – {8,10,19,20,21,22,23,24,25,29,33,34,35,40,41, 43,45,46,47,49,51,52,53,54,55,56,58} III – {27,28,31,32,42,50} IV – {26,30,36,37,38,39,44,48,57}	16 27 6 9	722.9286

(где первые 19 скважин относятся к месторождению Грязевой Сопки, а остальные к Песчаному и $l=4$).

Задача кластерного анализа рассмотренных скважин решена также с методом К-средних и, построено следующее оптимальное решение задачи:

Объекты, входящие в кластеры	$n(K_i)$	$D(\widetilde{K})$
I – {1,2,3,4,5,6,7,9,11,12,13,14,15,16,17,18,44}	17	708.103
II – {8,10,19,20,21,22,23,25,29,35,40,45,47,49,52,53, 54,55,56,58}	20	
III – {27,28,31,32,42,50}	6	
IV – {24,26,30,33,34,36,37,38,39,41,43,46,48,51,57}	15	

4. Выводы. В работе предложен прямой подход, базирующийся на эволюционных вычислениях по принципу Ламарка для решения задачи кластерного анализа объектов, характеризующихся количественными признаками. Разработан программный продукт в виде MatLab M-файла, реализующий этот алгоритм. С целью построения оптимальной классификации 58 скважин месторождений Грязевой Сопки и Песчаного по отношению с солеобразованию проведены вычислительные эксперименты с нормированными числовыми признаками. Оценена вычислительная сложность алгоритма. Установлено, что этот алгоритм входит в класс эффективных алгоритмов P.

Литература

1. Зенкин А.А., Зенкин А.И. Задача построения оптимальных классификаций. //В сб.: “Сборник работ по мат. кибернетике ВЦ АН СССР ”, М., 1981,с.20-33.
2. Керимов А.К. Построение оптимальных классификаций динамических объектов и прогнозирование состояний сложных систем. Автореферат дисс. на соис. уч.ст. к.ф.-м.н., Москва, 1985. 18 стр.
3. Давудова Р.И. Об одном алгоритме построения оптимальной классификации. //Известия НАНА. Сер. физ.-тех и мат.наук.Т.XXV, №3, 2005. стр.227-230.
4. Davudova R.İ. Klaster analizi məsələsinin həlli üçün yeni alqoritm, onun yiğilmasının tədqiqi. // AMEA-nın Məruzələri. №5, C. LXIV, 2008, səh.9-15.
5. Davudova R.İ. Avtomatik təsnifat məsələsinin bir həll üsulu haqqinda. //AMEA-nın Məruzələri. №6, C. LXIV, 2008, səh.20-25.
6. Керимов А.К., Давудова Р.И. Генетический алгоритм в задачах автоматической классификации. / Сборник докладов межд. конф. по мягким вычислениям и измерениям (SCM’2006). Санкт-Петербург, 26-29 июня 2006 г. Т.1. с.250- 252.
7. Ширинов К.А., Керимов А.К., Давудова Р.И. и др. Применение генетического алгоритма к задаче классификации нефтяных скважин по осложнениям. / Материалы межд. конф.“ Проблемы кибернетики и информатики (PCI 2006)”. Баку, 24-26 октября 2006 г. Том 1, стр.97-101.
8. Керимов А.К., Давудова Р.И. Применение самоорганизующегося генетического алгоритма в одной задаче оптимальной классификации. // Известия НАНА. Сер.физ.-тех. и мат. наук. Т.XXV, №3, 2008. стр.35-40.
9. Керимов А.К., Давудова Р.И. Генетический алгоритм в задачах автоматической классификации. / SCM 2006. IX Межд. конф. по мягким вычислениям и измерениям. 27- 29 июня, Санкт-Петербург.стр.250-252.
10. Журавлев Ю.И., Никифоров В.В. Алгоритмы распознавания, основанные на вычислении оценок. // “Кибернетика”, 1971, №3, Киев с.1-11.
11. Минкин Ю.И., Петров А.И. Самоорганизующийся генетический алгоритм.// Изв. АН. Теория и системы управления. М., 2001, №3, с.66-74.
12. Керимов А.К., Давудова Р.И. Эволюционный алгоритм для решения задачи автоматической классификации. //Искусственный интеллект и принятие решений. Институт системного анализа РАН. №4, 2009, стр. 74-79 (http://www.aidt.ru/)
13. Керимов А.К., Давудова Р.И. Программная реализация эволюционного алгоритма решения одной задачи автоматической классификации. // Международный журнал “Программные продукты и системы”. НИИ «ЦЕНТР-ПРОГРАММСИСТЕМ». № 1, 2010, стр. 15-18. (http://www.swsys.ru/)

14. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. – М., ФИЗМАТЛИТ, 2003, -432 стр.
15. J. Handl, J. Knowles, “An Evolutionary Approach to Multiobjective Clustering”, IEEE Trans. on Evolutionary Computation, Vol. 11, pp. 56-76, 2007.
16. M H Marghny, Rasha Abd M El-Aziz and Ahmed I Taloba. An Effective Evolutionary Clustering Algorithm: Hepatitis C Case Study. International Journal of Computer Applications 34(6):1-6, November 2011.
17. P.H.A. Sneath, R.R. Sokal. Numerical Taxonomy. The.Principles and Practice of Numerical Classification. San Francisco: W.H. Freeman, 1973. pp.573.
18. http://www.intuit.ru/studies/courses/6/6/lecture/182?page=3

UOT 519.7

R.İ. Davudova

Klaster analizi məsələlərində evolyusion hesablamalar

Ədədi əlamətləri ilə təsvir edilən obyektlərin optimal təsnifatının qurulması məsələsi nəzərdən keçirilmişdir. Məqalə klaster analizi məsələsinin həllində evolyusion hesablamaların tətbiqinə həsr edilmişdir. Baxılan məsələnin həlli üçün Lamarkın təkamül prinsipinə əsaslanan yeni yanaşma təklif edilmişdir.

Açar sözlər: keyfiyyət funksionalı, təkamül, lamarkizm, krossover, mutasiya, populyasiya, fərd, nəsil, autbridinq, inbridinq

R.I. Davudova

Evolutionary calculations in the cluster analysis problems

Automatic classification problems (ACP) of objects described by numerical attributes are investigated. The paper is devoted to the use of the evolutionary algorithms (EA) in cluster analysis problem. For the solution of the cluster analysis problem, a new approach is offered which is based on evolutionary calculations by Lamarck's principle

Keywords: functionality quality, evolution, Lamarckism, crossover, mutation, population, individual, child, outbreeding, inbreeding

Институт Систем Управления НАН Азербайджана Представлено 11.06.2015