



Ссылка на статью:

// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ. Электрон. журн. 2019, № 1, с. 7-19.

Поступила: 06.05.2019

Окончательный вариант: 10.06.2019

© УлГУ

УДК 621.122:658(075)

Кластеризация динамических объектов наблюдения

Бобров К.А.¹, Чекал Е.Г.^{1,*}, Чичев А.А.¹

[*checal@mail.ru](mailto:checal@mail.ru)

¹УлГУ, Ульяновск, Россия

Описывается модификация компоненты Кластеризатор при кластеризации динамических объектов наблюдения, описанной авторами ранее. Данная модификация снижает трудоемкость Кластеризатора.

Ключевые слова: кластеризация динамических объектов.

Введение

Кластеризация данных (Data clustering или группирование) — одна из важнейших задач, используемых в системах интеллектуального анализа данных. Результатом кластерного анализа являются группы объектов такие, что объекты, находящиеся в одной группе, более подобны между собой, нежели объекты из разных групп.

Можно выделить следующие цели кластеризации:

- упрощение дальнейшей обработки данных: разбиение объектов по группам позволяет работать с каждой из них по отдельности и применять свои методы анализа;
- сжатие данных: в некоторых задачах полная выборка данных является избыточной и чтобы ее сократить достаточно выбрать из каждой группы одного типичного представителя.
- выявление нестандартных объектов: объекты с девиантным набором признаков могут представлять особый интерес в ряде задач;
- задачи таксономии: в данном случае говорят об иерархической кластеризации, результатом которой является древообразная иерархическая структура;
- визуализация данных: визуализация остается одним из основных средств поддержки принятия решений, стимулирующим интуицию человека.

Первые работы, описывающие методы кластерного анализа относятся к концу 30-х годов. Считается, что термин «кластерный анализ» первым в употребление ввёл американский психолог из университета Беркли Роберт Трайон (Robert C. Tryon) в 1939. Однако активный интерес к данной теме пришёлся на период 60-80 гг. Импульсом для разработки многих кластерных методов послужила книга «Начала численной таксономии», опубликованная биологами Робертом Сокэлом и Петером Снитом. В это же время в СССР возникает несколько школ мирового значения, разрабатываются прикладные и теоретические работы [1]. В настоящее время актуальность работ в этой области подтверждается многочисленными исследованиями [например, 2-6].

Особое место занимают алгоритмы кластеризации динамических объектов наблюдения для систем управления надводным судном.

1. Особенности исходных данных

Объекты наблюдения, прежде всего воздушные, надводного судна характеризуются количественными и качественными свойствами, причем для кластеризации более важными являются первые, а вторые – являются уточняющими признаками. Значения количественных свойств могут быть константами, функциями времени и других параметров. Эти значения задаются дискретно через неравномерные промежутки времени. Начиная с любого момента времени, данные по одним объектам могут отсутствовать, и, наоборот, появиться по новым объектам.

Погрешности определения значений свойств в каждый момент времени не известны, но могут быть заданы технические (паспортные) характеристики источников обнаружения объектов (дисперсии по углу расхождения гирокомпасов судов, дальности, пеленгу, углу места, либо прямоугольным координатам).

Информация об объектах наблюдения может передаваться с иных подвижных источников информации.

Наблюдаемые объекты динамичны и могут управляться человеком, поэтому их поведение трудно предсказуемо.

Количество объектов достаточно велико для того, чтобы выделением и сопровождением групп занимался вручную оператор в системах жесткого реального времени.

Отсутствуют какие-либо оценки размеров и количества кластеров, но пространственное размещение объектов явно неоднородно.

2. Требования к кластер-процедуре

Процедура кластеризации должна удовлетворять следующим требованиям:

- отсутствие управляющих параметров;
- инвариантность относительно возможных преобразований координат;
- кластеризация должна осуществляться и в условиях отсутствия накопленных данных;
- стабильность выявленных кластеров в течении экспертно задаваемого времени;

- минимальные вычислительные затраты.

Инвариантность алгоритма к преобразованию координат важна, поскольку информация об объектах наблюдения может передаваться с иных подвижных источников информации. Ввиду расхождения гироскопов надводных судов и погрешностей определения местонахождения судна для приведения координат объектов к системе координат надводного судна, принимающего информацию, требуется параллельный перенос и поворот координатных осей [6].

По результатам кластеризации требуется определить характеристики групп. Результат должен быть предсказуем и согласовываться с интуитивным представлением человека о группах.

Учитывая особенности исходных данных и требования к кластер-процедуре, применение известных алгоритмов кластеризации [1-6] нецелесообразно прежде всего потому, что они рассматривают статические объекты, и используют управляющие параметры, которые в реальной быстро меняющейся обстановке определить сложно.

Разработанный в [7] алгоритм кластеризации динамических объектов наблюдения учитывает эти особенности. Согласно [1, 7], процесс кластеризации можно представить как совокупность следующих компонент: множество допустимых разбиений совокупности объектов на непересекающиеся кластеры S , множество описаний кластеров L , кластеризатор K , множество порций объектов P , поступающих на кластеризатор, генератор порций объектов G , дескриптор D . В работе [7], в виду динамичности рассматриваемых объектов, были добавлены стабилизатор T , идентификатор I , синтезатор обобщенных характеристик кластеров C . Однако алгоритм кластеризатора K в работе [7] оказался достаточно трудоемок.

3. Кластеризация объектов по проекциям

Исходя из особенностей требований предметной области предлагается реализовать алгоритм кластеризатора по проекциям. В его основе лежит наблюдение, что кластер, существующий в пространстве определенной размерности, гарантированно существует во всех его подпространствах меньшей размерности.

Алгоритм сначала выделяет кластеры в одномерном пространстве по координате X . Далее для каждого выделенного кластера производится сегментация в одномерном пространстве по координате Y , и далее, при необходимости по третьей координате Z . Если после выполнения этих шагов множество объектов действительно было разбито на кластеры, то рекурсивно запускается новый процесс для каждого нового кластера и он продолжается до тех пор, пока кластеры можно разделять.

Два объекта разносятся по разным кластерам в случае, если расстояние между ними больше определенного значения R . Это значение предлагается рассчитывать по формуле, полученной эмпирическим путем:

$$R = \frac{\max(\{ix : i \in S\}) - \min(\{ix : i \in S\}) + \max(\{iy : i \in S\}) - \min(\{iy : i \in S\})}{2 * (|S| - 1)},$$

где

i_x, i_y – значение координат x и y объекта i ,

S – множество объектов разделяемого кластера,

$|S|$ – мощность множества S .

Коэффициент 2 в знаменателе указывает на размерность пространства, для трехмерного случая коэффициент равен 3. Таким образом, R представляет собой число близкое к среднему расстоянию по проекциям между объектами разделяемого кластера. Первоначально разделяемый кластер представляет собой всю совокупность объектов.

Как показывают эксперименты, для ситуации, в которой объекты образуют явно выраженные группы, глубина дерева работы рекурсивного алгоритма равна двум. Уже на первом уровне алгоритм выявляет группы, а на втором лишь проверяется то, что дальнейшее разбиение этих групп не нужно.

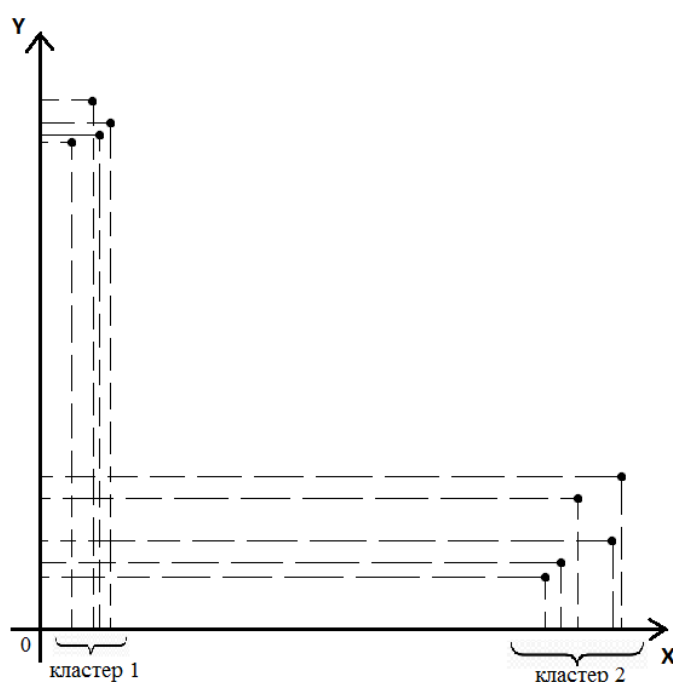


Рис. 1. Процесс кластеризации по проекциям

В примере, представленном на рис. 1, алгоритм выделяет два кластера по координате X . По координате Y ни один из кластеров уже не разбивается.

4. Примеры кластеризации объектов

В **примере 1** представлены неподвижные объекты, у которых отсутствует скорость. В таком случае функция экстраполяции никак не меняет значения координат. Таким образом, прогнозируемые координаты совпадают с текущими ($X = X_0, Y = Y_0$), и состав групп не меняется с течением времени (табл. 1). В результате получается четыре кластера (рис. 2).

Таблица 1

Пример 1. Недвижные объекты

Id	X	Y	X ₃	Y ₃	№ кластера
1	1231	100	1231	100	3
2	-80	-100	-80	-100	1
3	-70	-70	-70	-70	1
4	1200	140	1200	140	3
5	1100	120	1100	120	3
6	-150	-150	-150	-150	1
7	-99	-99	-99	-99	1
8	1170	120	1170	120	3
9	1199	1170	1199	1170	4
10	1160	1180	1160	1180	4
11	1201	1203	1201	1203	4
12	80	1240	80	1240	2
13	20	1200	20	1200	2
14	50	1175	50	1175	2

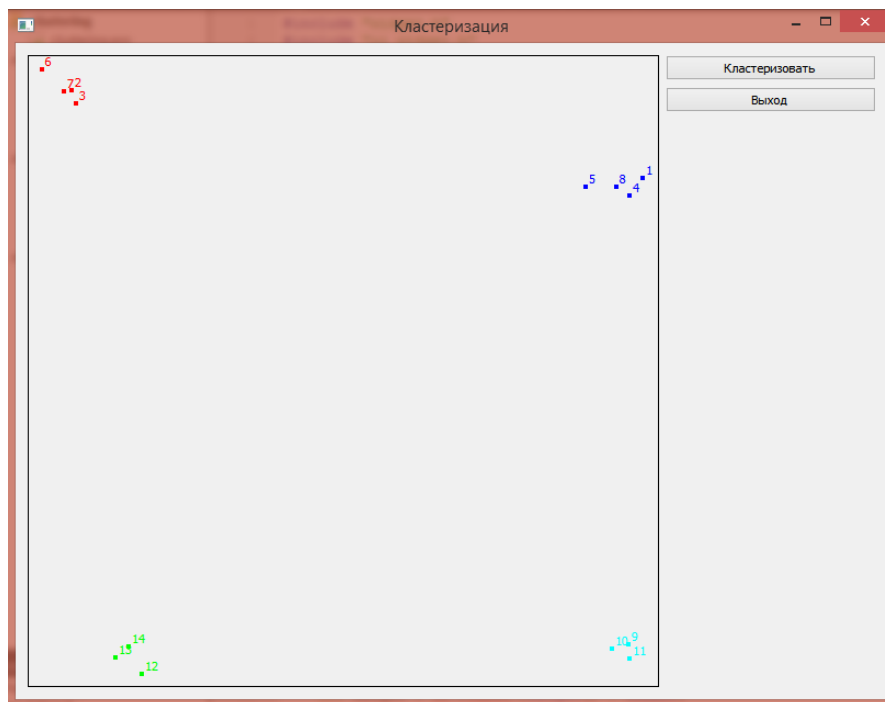


Рис. 2. Пример 1. Кластеризация неподвижных объектов

В **примере 2** объекты также неподвижны и прогнозируемые координаты совпадают с текущими (табл. 2). В результате получилось 3 кластера (рис. 3).

Таблица 2

Пример 2. Неподвижные объекты

Id	X	Y	X _г	Y _г	№ кластера
1	1231	100	1231	100	3
2	-80	-100	-80	-100	1
3	-70	-70	-70	-70	1
4	70	-30	70	-30	1
5	14	-120	14	-120	1
6	-150	-150	-150	-150	1
7	-99	-99	-99	-99	1
8	0	0	0	0	1
9	299	1170	299	1170	2
10	260	1180	260	1180	2
11	301	1203	301	1203	2
12	160	1140	160	1140	2
13	20	1200	20	1200	2
14	50	1175	50	1175	2
15	200	1099	200	1099	2
16	150	1150	150	1150	2
17	90	1100	90	1100	2

В **примере 3** все объекты движутся в одном направлении с одной скоростью. В данном случае функция экстраполяции сдвигает все объекты на одинаковую величину и поэтому расположение объектов относительно друг друга не меняется, а значит и состав групп тоже (табл. 3). В результате получается 4 кластера (рис. 4).

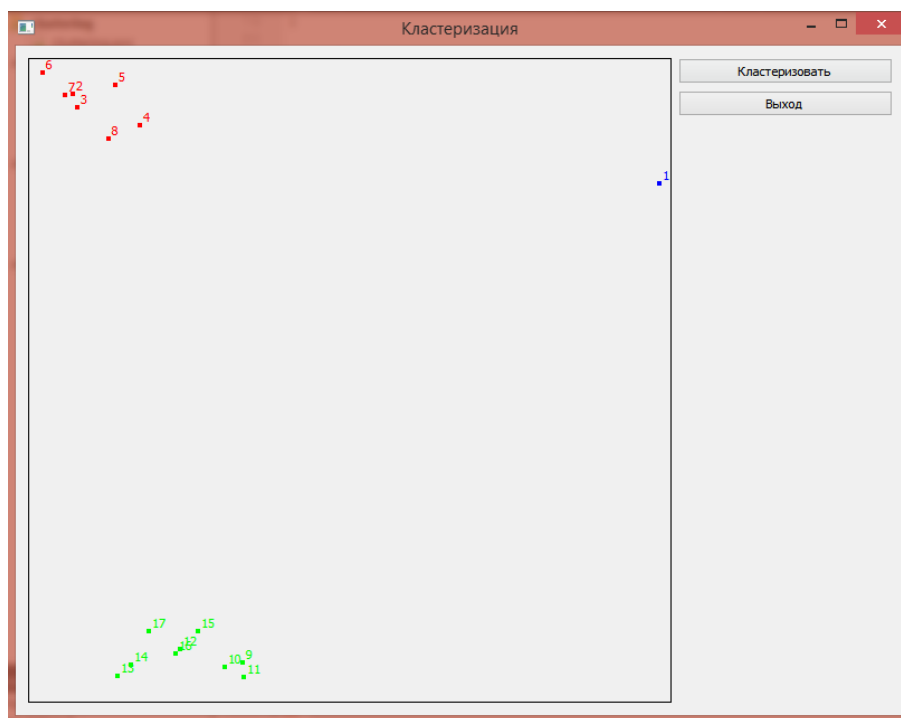


Рис. 3. Пример 2. Кластеризация неподвижных объектов

Таблица 3

Пример 3. Объекты с одинаковой скоростью и направлением

Id	X	Y	V_x	V_y	X_0	Y_0	№ кластера
1	50	666	90	150	590	1566	2
2	50	580	90	150	590	1480	2
3	90	1100	90	150	630	2000	3
4	50	1175	90	150	590	2075	3
5	-80	100	90	150	460	1000	1
6	-70	-70	90	150	470	830	1
7	20	1200	90	150	560	2100	3
8	111	0	90	150	651	900	1
9	14	-120	90	150	554	780	1
10	160	1140	90	150	700	2040	3
11	-150	-150	90	150	390	750	1
12	-99	-99	90	150	441	801	1
13	0	0	90	150	540	900	1

14	499	270	90	150	1039	1170	4
15	460	300	90	150	1000	1200	4
16	501	203	90	150	1041	1103	4
17	20	555	90	150	560	1455	2

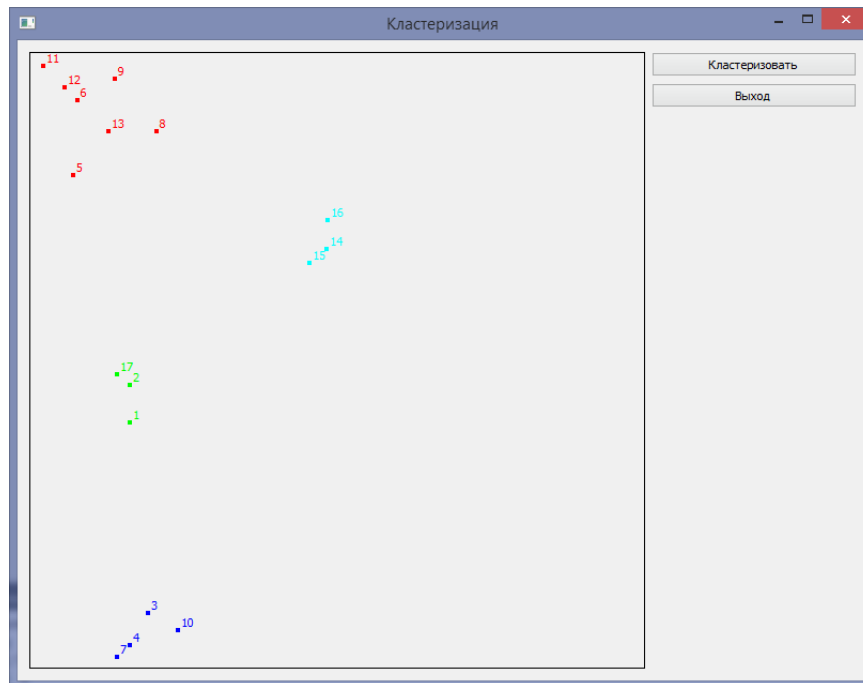


Рис. 4. Пример 3. Кластеризация объектов с одним направлением и скоростью

В **примере 4** представлены группы объектов, причем имеется две группы, которые в текущий момент времени имеют примерно одни и те же координаты, но разное направление движения. Так как на прогнозируемый момент времени они будут далеко друг от друга, функция кластеризации разбивает эти объекты по разным группам (табл. 4). В примере 4 это кластеры 2 и 3 (рис. 5).

Таблица 4

Пример 4. Пересечение двух групп

Id	X	Y	V_x	V_y	X_t	Y_t	№ кластера
1	1231	100	140	100	2071	700	4
2	-80	-100	140	100	760	500	1
3	-70	-70	140	100	770	530	1
4	70	-30	140	100	910	570	1
5	14	-120	140	100	854	480	1

6	-150	-150	140	100	690	450	1
7	-99	-99	140	100	741	501	1
8	0	0	140	100	840	600	1
9	20	1170	140	100	860	1770	2
10	160	1180	-140	100	-680	1780	3
11	201	1203	-140	100	-639	1803	3
12	160	1140	140	100	1000	1740	2
13	20	1200	140	100	860	1800	2
14	50	1175	140	100	890	1775	2
15	200	1099	-140	100	-640	1699	3
16	150	1150	130	110	930	1810	2
17	90	1100	150	100	990	1700	2

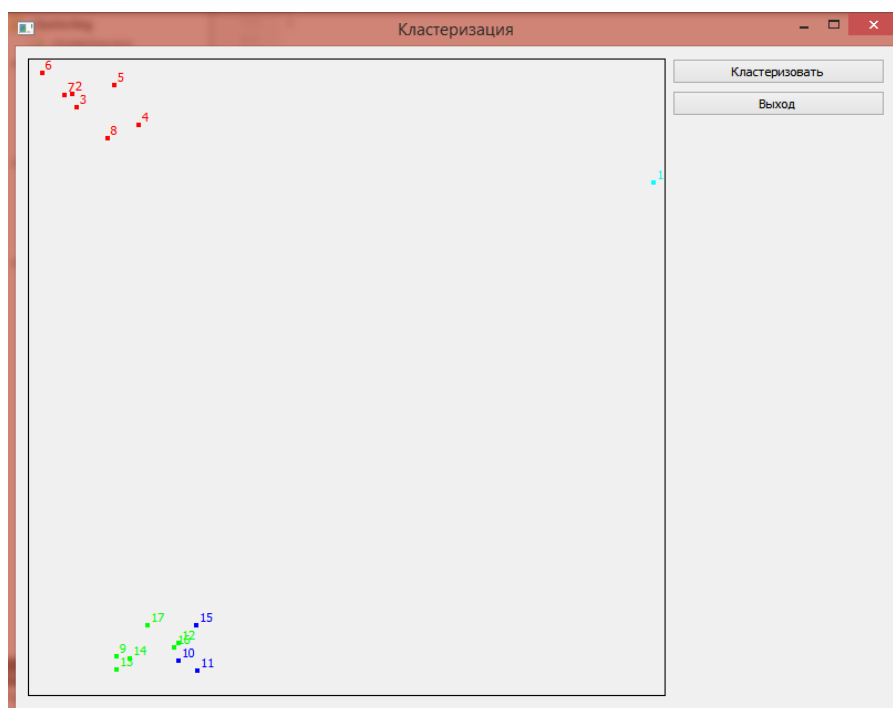


Рис. 5. Пример 4. Кластеризация объектов при пересечении групп с разной траекторией движения

В **примере 5** представлена еще одна ситуация, когда имеется две группы, которые располагаются очень близко друг к другу в текущий момент времени, но имеют разное направление движения. Так как на прогнозируемый момент времени они будут далеко друг от друга, функция кластеризации разбивает эти объекты по разным группам (таб. 5). В примере 5 это кластеры 2 и 3 (рис. 6).

Таблица 5

Пример 5. Пересечение двух групп

Id	X	Y	V_x	V_y	X_3	Y_3	№ кластера
1	1231	100	-100	-100	631	-500	3
2	1200	10	-100	-100	600	-590	3
3	1220	40	-100	-100	620	-560	3
4	1200	140	-100	-100	600	-460	3
5	1100	120	100	100	1700	720	2
6	1150	50	100	90	1750	590	2
7	1170	0	-100	-100	570	-600	3
8	1170	120	90	100	1710	720	2
9	1199	1170	-100	-100	599	570	4
10	1160	1180	-100	-100	560	580	4
11	1201	1203	-100	-100	601	603	4
12	80	1240	-100	-100	-520	575	1
13	20	1200	-100	-100	-580	600	1
14	50	1175	-100	-100	-550	575	1

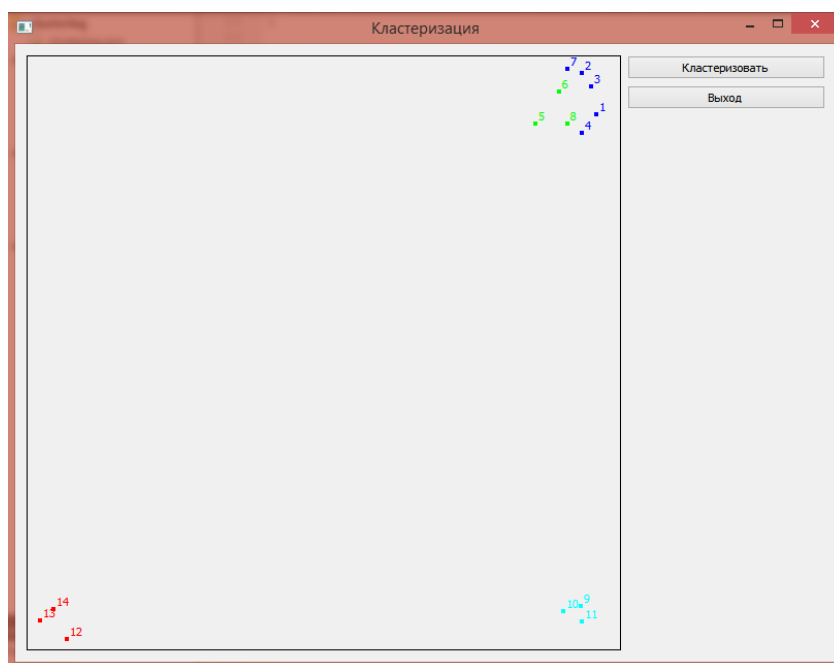


Рис. 6. Пример 5. Кластеризация объектов при пересечении групп с разной траекторией движения

В **примере 6** объекты имеют случайные координаты и случайный вектор скорости (таб. 6). В данном случае кластеризация не имеет смысла и каждый объект находится в отдельном кластере (рис. 7).

Таблица 6

Пример 6. Каждый объект движется в случайном направлении

Id	X	Y	V_x	V_y	X_3	Y_3	№ кластера
1	1231	100	-100	-100	631	-500	9
2	1200	1378	-100	-1	600	1372	5
3	500	500	0	0	500	500	3
4	123	140	-100	-100	-477	-460	6
5	1100	120	50	-150	1400	-780	1
6	150	1250	100	90	750	1790	4
7	1170	1111	80	10	1650	1171	7
8	170	1280	90	-90	710	740	8
9	300	700	-100	100	-300	1300	2

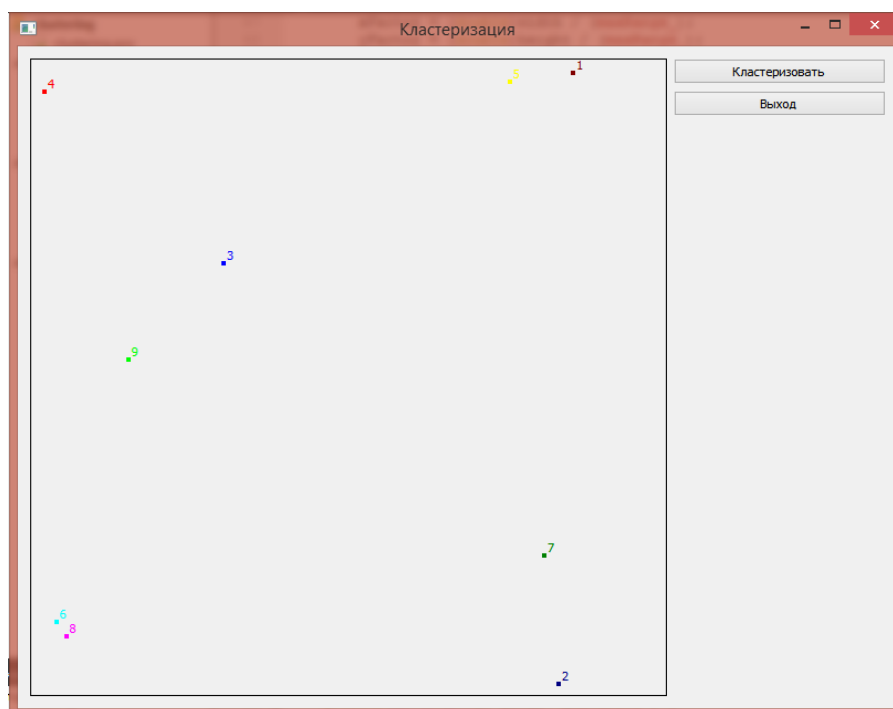


Рис. 7. Пример 6. Кластеризация объектов при движении в случайном направлении

В **примере 7** на рис. 8 алгоритм не обнаружил ни одного кластера. Однако он может быть применим, если провести предварительный поворот координатной плоскости, при котором одна из осей координат будет параллельна прямой, образуемой объектами.

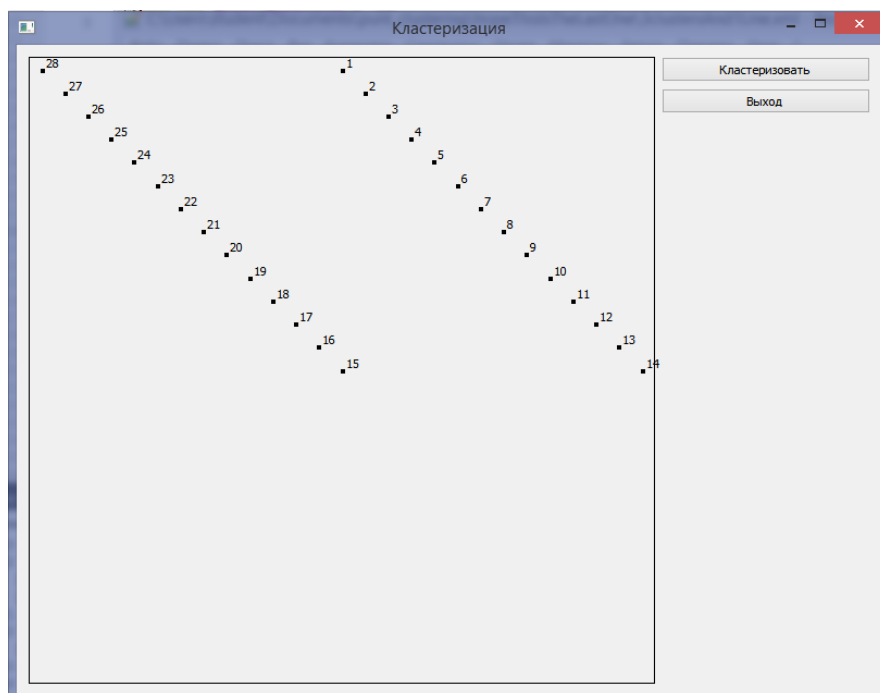


Рис. 8. Пример 7. Кластеризация объектов специального расположения

Заключение

Предлагаемая модификация алгоритма кластеризатора динамических объектов наблюдения позволяет уменьшить временные характеристики процесса обработки, причем, при отсутствии управляющих параметров и образцов кластеров.

Список литературы

1. Айвазян С.А. *Прикладная статистика: Классификация и снижение размерности: Справочное издание* / С. А. Айвазян [и др.] / под ред. С. А. Айвазяна. М.: Финансы и статистика, 1989.
2. Белим С.В., Кутлуниин П. Е. Выделение контуров на изображениях с помощью алгоритма кластеризации. // *Компьютерная оптика*, 2015, т. 39, №1, с. 119-124.
3. Егоров А. В., Куприянова, Н. И. Особенности методов кластеризации данных // *Известия Южного федерального университета. Технические науки*. 2011, №11, с.174-178.
4. Куприянова, Н. И. Концептуальная модель кластеризации данных // *Известия Южного федерального университета. Технические науки*. 2012, №4, с. 256-260.

5. Миркин Б. Г. *Методы кластер-анализа для поддержки принятия решений*. М.: Изд. дом Национального исследовательского университета «Высшая школа экономики», 2011.
6. Пестунов И. А., Синявский Ю. Н. Алгоритмы кластеризации в задачах сегментации спутниковых изображений // *Вестник КемГУ*, 2012, № 4, с. 110-125.
7. Чекал Е. Г., Чичев А. А. Алгоритм автоматической классификации динамических объектов наблюдения. // *Интегрированные АСУ. Корабельные комплексы и системы*. – 2012, №1(27), с. 87-94.