

**К вопросу об алгоритмах кластеризации толпы в условиях чрезвычайных ситуаций<sup>1</sup>****Бекларян Армен Левонович***преподаватель кафедры бизнес-аналитики**Национального исследовательского университета «Высшая школа экономики»**E-mail: abeklaryan@hse.ru***Акопов Андраник Сумбатович***доктор технических наук, профессор кафедры бизнес-аналитики**Национального исследовательского университета «Высшая школа экономики»**E-mail: aakopov@hse.ru***Аннотация**

В работе развивается агентная модель поведения толпы на основе феноменологического подхода. С целью повышения качества проводимого кластерного анализа и для уменьшения размерности задачи был разработан модифицированный эволюционный алгоритм нечеткой кластеризации, учитывающий факт наличия препятствия на пути следования агента, а также текущее направление движения агента.

**1. Введение**

В процессе моделирования поведения толпы, и тем более при попытке моделирования чрезвычайной ситуации, сопряженной с паникой и другими сложными внутренними процессами, возникает проблема невозможности полного математического описания поведения индивида, поскольку действия и сама система принятия решений человека определяется большим количеством факторов, как физического, так и психо-социального характера. Также каждому человеку в разной степени свойственна иррациональность принимаемых решений. Все это обуславливает некоторую степень «усреднения» моделируемых процессов. Однако стоит отметить, что в основе подобного подхода лежит разумная гипотеза о применимости к толпе закона больших чисел: нетипичное поведение индивида не влияет на качественную картину поведения всей толпы.

**2. Анализ существующих моделей**

Ввиду неуклонного роста актуальности данной тематики, математическому моделированию поведения толпы уделено большое внимание в научной литературе и с каждым годом (видимо, вследствие роста вероятности возникновения ЧС по всему миру) подобный интерес только возрастает. Анализ существующих работ показал, что большая часть исследований следует отнести к трем ключевым подходам:

---

<sup>1</sup> Работа выполнена при финансовой поддержке Российского Фонда Фундаментальных Исследований, грант № 15-37-20265

1. *наивный*, основанный на психологических и социальных аспектах вопроса [6, 7, 21], рассматривающий толпу с фрейдистской точки зрения;
2. *молекулярный*, представленный в основном, так называемыми, моделями Хелбинга [2, 20], в основе которых лежит идея применения к толпе людей методов молекулярной динамики, где дается интерпретация психологическим и социальным факторам в форме потенциалов взаимодействия между молекулами-людьми;
3. *феноменологический*, предложенный, в частности, в работах Бекларяна(ст.)-Акопова [8, 9] (которые положены в основу предлагаемой модели), в рамках которого априори определяются состояния агентов с их характеристиками, правила взаимодействия агентов и правила принятия решений.

Класс агентных моделей представлен широким списком работ (например, [16-18]) и является самой естественной реализацией феноменологического подхода. Тем не менее, доля российских исследований в данной тематике непростительно мала. В результате учета имеющихся моделей и подходов, а также исследований по психологии толпы, были сформулированы основные априорные предположения, которые легли в основу предлагаемой агентной модели поведения при ЧС, среди которых стоит отметить:

- частичная или полная потеря ориентации в пространстве и во времени;
- высокая степень турбулентности толпы, т.е. наличие хаотичного движения во всех направлениях в условиях высокой плотности агентов;
- существенное замедление скорости передвижения при определенных условиях (ранение, уплотнение и т.д.);
- стремление к ближайшему выходу в случае нахождения выхода в пределах видимости;
- стремление к присоединению к ближайшей группе агентов (эффект притяжения толпы).

В предлагаемой модели реализуется концепция перехода от фиксированных значений ряда показателей, отражающих как геометрию помещения, так и физику процесса перемещения агентов, к представлению их в качестве управляющих параметров модели. В результате удастся построить гибкую, универсальную модель, позволяющую варьировать управляющими параметрами и, как следствие, калибровать модель с целью максимизации правдоподобия с реальными процессами. Также предлагаемая модель допускает дальнейшее усложнение во всех аспектах (геометрия помещения, механика взаимодействия агентов, характеристики самих агентов и т.д.) и введение новых уравнений связи и условий.

Наиболее близкими в части системы принятия решений, к предлагаемой в данной работе модели, являются агентные модели, представленные в работах [13] и [10]. В обеих

указанных моделях каждому агенту сопоставляется некий сектор обзора, в рамках которого агент выбирает дальнейший вектор своего движения на основе функции полезности, которая, в свою очередь, отражает такие стремления, как избегать столкновения с другими агентами, не менять направление своего движения, прямолинейность движения к выходу, максимизация скорости движения в случае паники и пр. В предлагаемой модели функция полезности расширена дополнительными факторами и учитывает такие характеристики, как состояние агента, состояние ситуации и др.

С формальным описанием модели, множества ее параметров, правилами взаимодействия агентов, а также с реализацией в среде AnyLogic можно ознакомиться в работах [3-5]. В рамках данной статьи дается описание разработанного эволюционного алгоритма нечеткой кластеризации и приводятся полученные результаты кластерного анализа.

### 3. Алгоритм эволюционной кластеризации динамики толпы

С целью выявления устойчивых групп агентов в основную модель был импортирован авторский программный пакет динамической кластеризации. В каждый момент модельного времени на множестве всех агентов проводится кластеризация одновременно по трем алгоритмам [1, 12, 19]:

1. Метод k-средних (k-means++).
2. Иерархический метод.
3. Метод неточной кластеризации C-средних (Fuzzy C-means).

Ввиду того, что для указанных алгоритмов в качестве входного параметра выступает количество кластеров, чье значение, с учетом частоты проведения анализа и вариативных количества агентов и геометрии активного пространства, не поддается никакой оценке, то для вычисления оптимального количества кластеров использовался известный критерий компактности [14]:

$$S(U, c) = \frac{\frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n \mu_{ij}^2 \|x_k - v_i\|^2}{\min_{i,j} \|v_i - v_j\|^2}, \quad (1)$$

$$c = \arg \min_c \left\{ \min_{\Omega_c} S(U, c) \right\}, \quad (2)$$

где  $U$  – множество всех агентов,  $i, j$  – индекс кластеров,  $k$  – индекс агентов в кластере,  $\mu_{ij}$  – вероятность принадлежности  $k$ -ого агента  $i$ -ому кластеру,  $x_k$  – координата  $k$ -ого агента,  $v_i$  – координата центра  $i$ -ого кластера,  $\Omega_c$  – всевозможные разбиения  $U$  на  $c$  кластеров.

С учетом высокой степени перестроения кластеров (распад имеющихся кластеров и соединения агентов в новые), а также максимального «правдоподобия», в качестве базового алгоритма был выбран алгоритм нечеткой кластеризации.

Алгоритм нечеткой кластеризации Fuzzy C-means является широко используемым подходом к проблеме категоризации наблюдаемых объектов по признакам. В отличие от четких алгоритмов, алгоритм Fuzzy C-means не относит объект однозначно к какому-нибудь кластеру, а сопоставляет каждому кластеру вероятность отнесения к нему наблюдаемых объектов, формируя, так называемую, матрицу принадлежности. Таким образом, на каждом шаге алгоритм каждый объект одновременно относится ко всем кластерам.

Впервые данный алгоритм был предложен в работе [22], далее развит в работе [15]. Окончательная форма была представлена в работе [11], а доказательство сходимости – в монографии [12]. Ниже идет описание шагов алгоритма.

Исходной информацией для кластеризации является матрица наблюдений  $l \times n$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{l1} & \cdots & x_{ln} \end{pmatrix},$$

где  $l$  – число объектов,  $n$  – число признаков (наблюдений) для каждого объекта (в нашем случае – координаты).

Задача кластеризации состоит в разбиении множества объектов на группы (кластеры) «похожих» между собой объектов. В  $n$ -мерном метрическом пространстве признаков мерой «сходства» двух объектов будем считать расстояние между ними.

Кластерная структура задаётся матрицей принадлежности ( $c \times l$  матрица):

$$M = \begin{pmatrix} m_{11} & \cdots & m_{1l} \\ \vdots & \ddots & \vdots \\ m_{c1} & \cdots & m_{cl} \end{pmatrix},$$

где  $m_{ij}$  – степень принадлежности  $j$ -ого элемента  $i$ -ому кластеру.

Отметим, что матрица принадлежности должна удовлетворять следующим условиям:

1.  $m_{ij} \in [0,1], i \in \overline{1,c}, j \in \overline{1,l}$ ,
2.  $\sum_{i=1}^c m_{ij} = 1, j \in \overline{1,l}$ , то есть каждый объект должен быть распределён между всеми кластерами,
3.  $0 < \sum_{j=1}^l m_{ij} < l$ , то есть ни один кластер не должен быть пустым или содержать все элементы.

Для оценки качества разбиения используется критерий разброса, показывающий сумму расстояний от объектов до центров кластеров с соответствующими степенями принадлежности:

$$J = \sum_{i=1}^c \sum_{j=1}^l (m_{ij})^w d(v_i, x_j),$$

где  $d(v_i, x_j)$  – евклидово расстояние между  $j$ -м объектом  $x_j = (x_{j1}, x_{j2}, \dots, x_{jn})$  и  $i$ -м центром кластера  $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$ ,  $w \in (1, \infty)$  – экспоненциальный вес, определяющий нечёткость, размытость кластеров.

$$V = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{c1} & \cdots & v_{cn} \end{pmatrix},$$

–  $c \times n$  матрица координат центров кластеров, элементы которой вычисляются по формуле

$$v_{ik} = \frac{\sum_{j=1}^l (m_{ij})^w x_{jk}}{\sum_{j=1}^l (m_{ij})^w}, k = \overline{1, n}.$$

Задачей является нахождение матрицы  $M$ , минимизирующей критерий  $J$ . Для этого используется алгоритм нечётких  $C$ -средних, в основе которого лежит метод множителей Лагранжа. Он позволяет найти локальный оптимум, поэтому для различных запусков могут получиться разные результаты.

На первом шаге матрица принадлежностей  $M$ , удовлетворяющая условиям 1-3, генерируется случайным образом. Далее запускается итерационный процесс вычисления центров кластеров и пересчёта элементов матрицы степеней принадлежности:

$$m_{ij} = \frac{1}{(d_{ij})^{\frac{2}{w-1}} \sum_{k=1}^c \frac{1}{(d_{kj})^{\frac{2}{w-1}}}} \text{ при } d_{ij} > 0 \text{ и } m_{kj} = \begin{cases} 1, & k = i \\ 0, & k \neq j \end{cases} \text{ при } d_{ij} = 0,$$

где  $d_{ij} = d(v_i, x_j)$  для  $i \in \overline{1, c}, j \in \overline{1, l}$ .

Вычисления продолжаются до тех пор, пока изменение матрицы  $M$ , характеризующееся величиной  $\|M - M^*\|^2$ , где  $M^*$  – матрица на предыдущей итерации, не станет меньше заранее заданного параметра остановки  $\varepsilon$ .

Остановимся на выборе значения  $w$  – экспоненциального веса. Чем больше это значение, тем матрица принадлежности более размазанная и при  $w \rightarrow \infty$  элементы примут вид  $m_{ij} = 1/c$ , что является плохим решением, так как все объекты с одинаковой степенью распределены по всем кластерам. При  $w = 1$  алгоритм Fuzzy  $C$ -means вырождается в обычный  $k$ -means. Теоретически обоснованного правила выбора веса пока не существует, и обычно устанавливают  $w = 2$ .

Схема работы стандартного алгоритма Fuzzy  $C$ -means представлена в левой части на рис.1.

Необходимо отметить ключевые недостатки данного алгоритма:

- Высокая степень зависимости результирующего разбиения от начальной матрицы принадлежности.
- Необходимость априорного задания числа кластеров.

- Отсутствие учета специфики предметной области (динамика толпы).

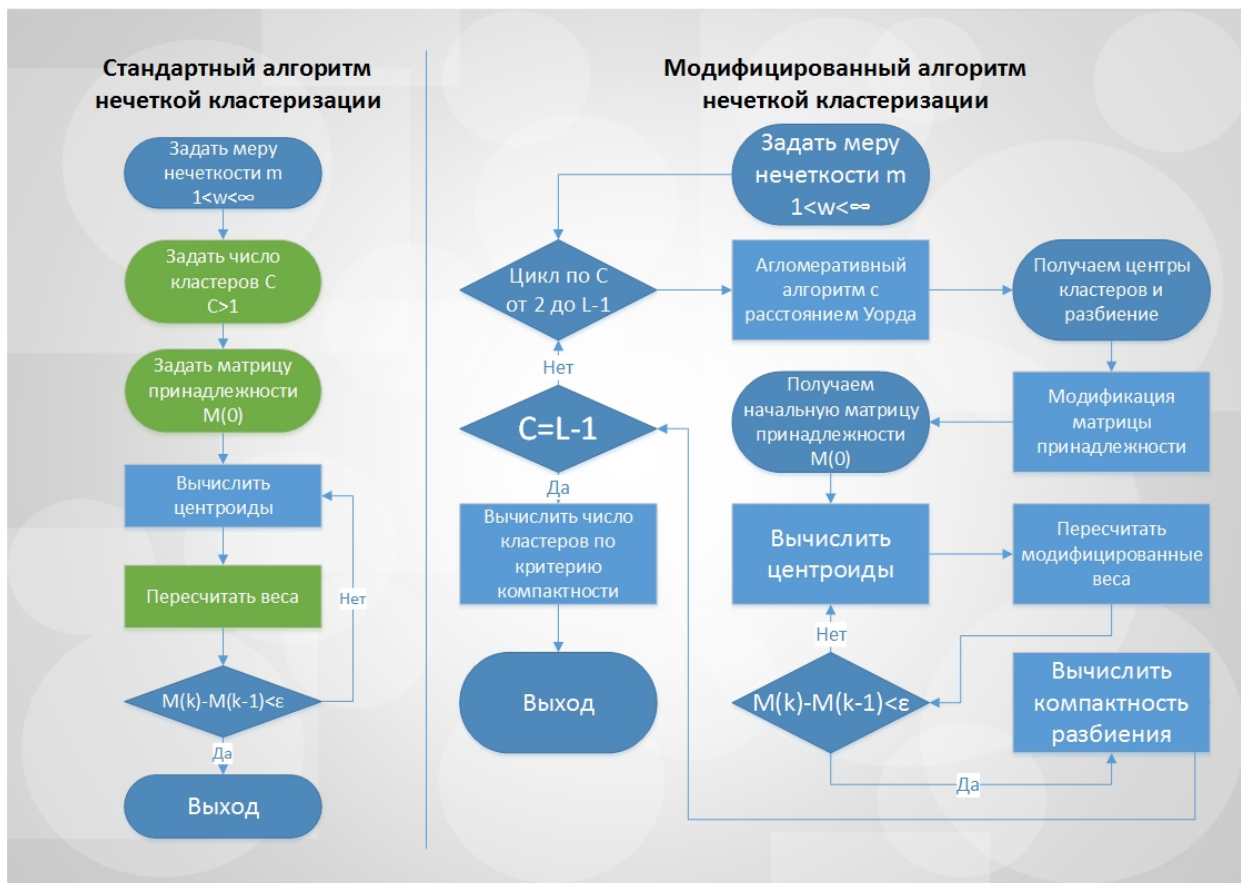


Рисунок 1. Схема алгоритмов нечеткой кластеризации.

С целью уменьшения влияния перечисленных недостатков был разработан и имплементирован в модель модифицированный (эволюционный) алгоритм нечеткой кластеризации. Схема работы эволюционного алгоритма Fuzzy C-means представлена в правой части на рис. 1.

Также, как и в стандартном алгоритме кластеризации, в начале экспертно задается экспоненциальный вес  $w$ . Далее осуществляется цикл по количеству кластеров от 2 до  $N - 1$ . В рамках каждого шага цикла осуществляется работа иерархического агломеративного алгоритма кластеризации с расстоянием Уорда [23]

$$R(P, S) = \frac{|P||S|}{|P| + |S|} \rho^2 \left( \sum_{p \in P} \frac{p}{|P|}, \sum_{s \in S} \frac{s}{|S|} \right),$$

где  $p \in P$  – элементы кластера  $P$ ,  $s \in S$  – элементы кластера  $S$ ,  $|\cdot|$  – число элементов в кластере,  $\rho(\cdot, \cdot)$  – длина наименьшего пути с учетом наличия препятствий. Выбор именно расстояния Уорда в качестве оценки меры принадлежности к кластеру обусловлен тем фактом, что данное расстояние обладает свойствами монотонности, растяжения и редуцивности.

В результате работы агломеративного алгоритма получена матрица принадлежности  $M'_0$

$$M'_0 = \begin{pmatrix} m'_{11} & \cdots & m'_{1l} \\ \vdots & \ddots & \vdots \\ m'_{c1} & \cdots & m'_{cl} \end{pmatrix}$$

где  $m'_{ij} \in \{0,1\}$  и в каждом столбце встречается ровно одна единица. Далее, чтобы получить начальную матрицу принадлежности для работы алгоритма нечеткой кластеризации проводится следующее преобразование

$$m_{ij} = \frac{m'_{ij} + \frac{\frac{A+1}{\alpha_{ij}} - 1 + \frac{\rho_j}{\rho_{ij}}}{A + \frac{\pi}{1 + \frac{1}{A}}}}{\sum_{i=1}^c m'_{ij} + \frac{\frac{A+1}{\alpha_{ij}} - 1 + \frac{\rho_j}{\rho_{ij}}}{A + \frac{\pi}{1 + \frac{1}{A}}}}, \quad (3)$$

где  $\rho_{ij}$  – расстояние от  $j$ -ого агента до  $i$ -ого кластера с учетом наличия препятствий,  $\rho_j$  – расстояние от  $j$ -ого агента до того кластера, которому он принадлежал по итогам работы агломеративного алгоритма,  $\alpha_{ij}$  – угол между направлением движения  $j$ -ого агента ( $\vec{d}_j$ ) и направлением от агента к центру  $i$ -ого кластера,  $A \in (0,1]$  – коэффициент компенсации (на сколько совпадение направления движения агента и направления от агента к центру кластера компенсирует дальность от кластера). Чем меньше значение  $A$ , тем более значимо совпадение направлений.

Таким образом, в результате преобразования (3) начальная матрица принадлежности формируется не случайно, а с учетом как метрической близости, так и динамики агентов.

Новый, модифицированный критерий разброса также учитывает динамику агентов:

$$J' = \sum_{i=1}^c \sum_{j=1}^l (m_{ij})^w e^{B \frac{\alpha_{ij}}{\pi}} d(v_i, x_j),$$

где  $B > 0$  – весовой коэффициент значимости направления движения агента.

Как следствие, соответствующая модификация распространяется и на формулы расчета центров кластеров и степеней принадлежности:

$$v_{ik} = \frac{\sum_{j=1}^l (m_{ij})^w e^{B \frac{\alpha_{ij}}{\pi}} x_{jk}}{\sum_{j=1}^l (m_{ij})^w e^{B \frac{\alpha_{ij}}{\pi}}}, \quad k = \overline{1, n},$$

$$m_{ij} = \frac{1}{\left(d_{ij}^2 e^{B \frac{\alpha_{ij}}{\pi}}\right)^{\frac{1}{w-1}} \sum_{k=1}^c \frac{1}{\left(d_{kj}^2 e^{B \frac{\alpha_{kj}}{\pi}}\right)^{\frac{1}{w-1}}}} \text{ при } d_{ij} > 0 \text{ и } m_{kj} = \begin{cases} 1, & k = i \\ 0, & k \neq j \end{cases} \text{ при } d_{ij} = 0.$$

По итогам работы модифицированного алгоритма нечеткой кластеризации вычисляется компактность разбиения по формулам (1)-(2). После завершения цикла по количеству кластеров выбирается то, разбиение, которое дало минимальную компактность.

#### 4. Результаты имитационного моделирования в AnyLogic

Сравнение качества кластеризации по результатам имитационного моделирования в среде AnyLogic при различных значениях управляющих параметров представлено в табл.1.

Таблица 1. Сравнение алгоритмов кластеризации.

	<b>Иерархический</b>	<b>k-means++</b>	<b>Fuzzy C-means</b>	<b>Modified Fuzzy C-means</b>
<i>Частота смены кластера</i>	6,3	4,2	1,8	1,4
<i>Компактность</i>	2,8	2,3	1,3	1,1
<i>Период двух кластеров</i>	2,7	3,1	1,7	1,5

Частота смены кластера – общее число смены кластера каждым агентом за весь период симуляции. Для нечетких алгоритмов считается смена основного кластера с порогом 0,6. Компактность – среднее значение критерия компактности за весь период симуляции. Период двух кластеров – отношение времени симуляции, в течение которого алгоритм определял два кластера, к общему времени симуляции.

Каждая оценка выражена относительным значением: значения каждого критерия нормировались относительно наименьшего, а затем усреднялись по числу прогонов.

Таким образом, определение кластеров является более точным по сравнению с другими известными методами кластеризации применительно к задаче кластеризации толпы. Кроме того, использование данной программной библиотеки позволяет перейти от динамики агентов к динамике (центров) кластеров, и, таким образом, принципиально уменьшить размерность данной задачи (зависящую от числа агентов), обеспечивая численное моделирование развития ЧС за полиномиальное время. Для этого, библиотека содержит специальную функцию, которая на основе данных о динамике агентов, сохраняемых в БД, формирует динамику центров кластеров непосредственно в имитационной модели AnyLogic, существенно сокращая, таким образом, время на проведение экспериментов с моделью.



В результате использования эволюционного алгоритма кластеризации для помещения длиной 110 метров, шириной 65 метров и количества агентов 150 человек удалось уменьшить время симуляции с одного часа до 8 минут при 20 прогонах модели (рис. 2 и 3).

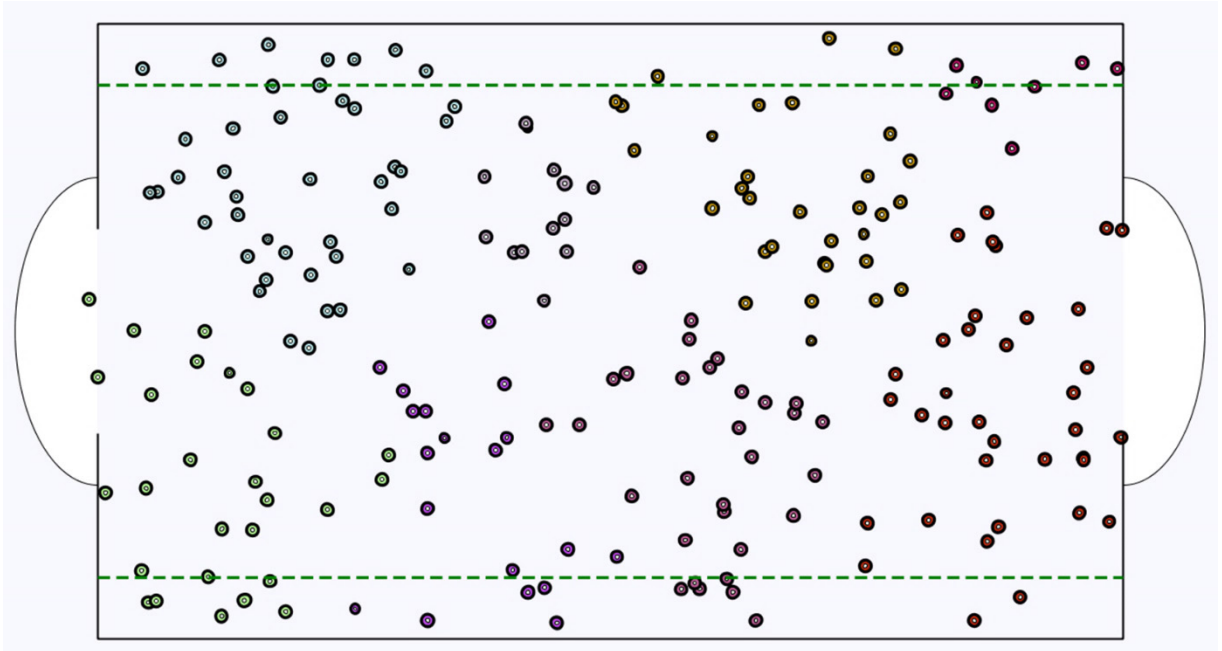


Рисунок 2. Динамика на уровне агентов.

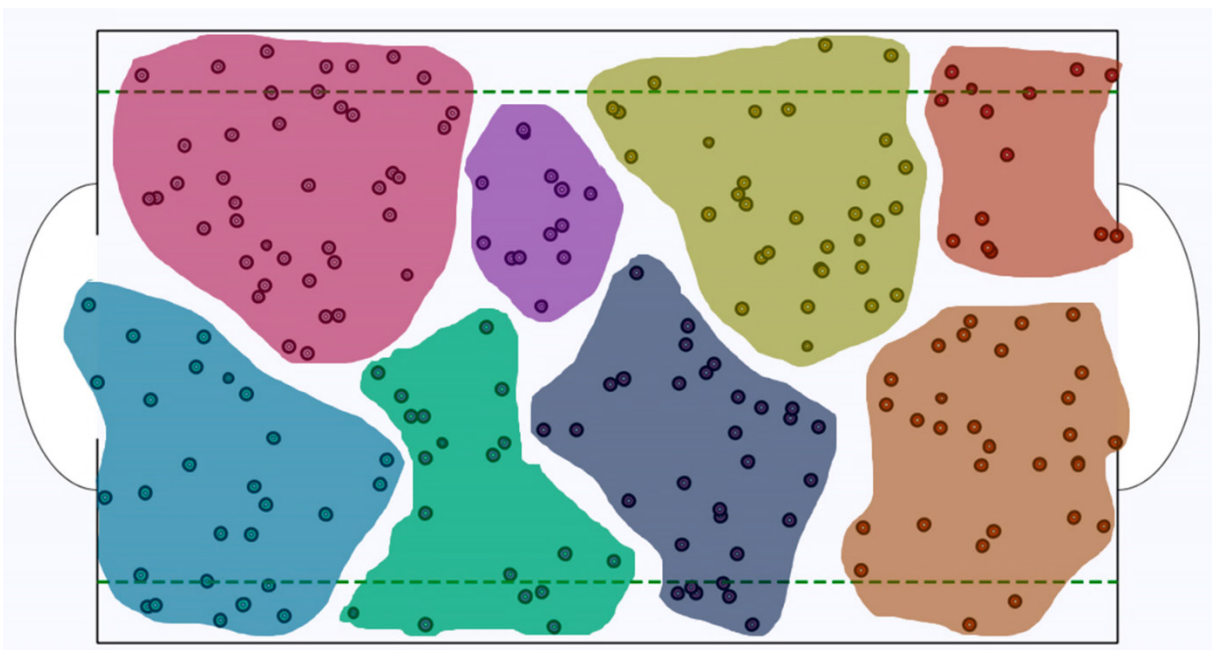


Рисунок 3. Динамика на уровне кластеров.

Также, как видно из табл. 2, за счет применения эволюционного алгоритма нечеткой кластеризации удастся повысить временную эффективность модели, при этом рост времени симуляции при кластерной динамике в зависимости от числа агентов и начального

распределения значительно ниже, чем при динамике модели на уровне агентов (приведена усредненная статистика по 100 прогонам для каждого случая).

Таблица 2. Сравнение эффективности работы модели.

<b>Ширина помещения, м</b>	<b>Длина помещения, м</b>	<b>Число агентов, чел</b>	<b>Начальное распределение</b>	<b>Тип динамики</b>	<b>Время симуляции, мин</b>
110	65	150	Равномерное	Агент	40
110	65	150	Равномерное	Кластер	8
110	65	150	Нормальное	Агент	55
110	65	150	Нормальное	Кластер	10
110	65	300	Равномерное	Агент	100
110	65	300	Равномерное	Кластер	15
110	65	300	Нормальное	Агент	112
110	65	300	Нормальное	Кластер	17
200	100	300	Равномерное	Агент	150
200	100	300	Равномерное	Кластер	21
200	100	300	Нормальное	Агент	170
200	100	300	Нормальное	Кластер	23

## 5. Заключение

В рамках введенной ранее агентной модели поведения толпы был разработан эволюционный алгоритм нечеткой кластеризации, учитывающий факт наличия препятствий на пути следования агента, с учетом текущего направления движения агента, позволяющий существенно улучшить точность идентификации таксонов (кластеров толпы). Проведенные эксперименты при различных значениях управляющих параметров модели показали качественное снижение размерности задачи и повышение временной эффективности модели.

## Список литературы

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
2. Аптуков А.М., Брацун Д.А., Люшнин А.В. Моделирование поведения паникующей толпы в многоуровневом разветвленном помещении // Компьютерные исследования и моделирование. 2013. Т. 5, № 3. С. 491-508.
3. Бекларян А.Л., Акопов А.С. Моделирование поведения толпы на основе интеллектуальной динамики взаимодействующих агентов // Бизнес-информатика. 2015. Т. 31, №. 1. С. 69-77.
4. Бекларян А. Л. Имитационная модель поведения толпы в среде разработки AnyLogic // Вестник Бурятского государственного университета. 2015. № 9. С. 40-53.

5. Бекларян А. Л. Фронт выхода в модели поведения толпы при чрезвычайных ситуациях // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2015. Т. 20. № 4. С. 851-856.
6. Лебон Г. Психология народов и масс. М.: Академический проект, 2011. 238 с.
7. Фрейд З. Психология масс и анализ человеческого «Я». М.: Азбука-Аттикус, 2013. 192 с.
8. Akopov A.S., Beklaryan L.A. An Agent Model of Crowd Behavior in Emergencies // Automation and Remote Control. 2015. Vol. 76, No. 10. P. 1817-1827.
9. Akopov A.S., Beklaryan L.A. Simulation of human crowd behavior in extreme situations // International Journal of Pure and Applied Mathematics. 2012. Vol. 79, No. 1. P. 121-138.
10. Antonini G., Bierlaire M., Weber M. Discrete choice models of pedestrian walking behavior // Transportation Research Part B. 2006. Vol. 40, No. 8. P. 667-687.
11. Bezdek C.J. Cluster Validity with Fuzzy Sets // Journal of Cybernetics. 1974. Vol. 3, No. 3. P. 58-73.
12. Bezdek C.J. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers Norwell, 1981. 256 p.
13. Cherif F., Chighoub R. Crowd simulation influenced by agent's sociopsychological state // Journal of computing. 2010. Vol. 2, No. 4. P. 48-54.
14. Chi Z., Yan H., Pham T. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition. World Scientific, 2006. 225 p.
15. Dunn J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters // Journal of Cybernetics. 1973. Vol. 3, No. 3. P. 32-57.
16. Farrahi K., Zia K., Sharpanskykh A. et al. Agent Perception Modeling for Movement in Crowds // 11th International Conference Advances on Practical Applications of Agents and Multi-Agent Systems. Springer Berlin Heidelberg. 2013. P. 73-84.
17. Golas A., Narain R., Lin M.C. Continuum modeling of crowd turbulence // Physical review E. 2014. Vol. 90, № 4. P. 042816.

18. Guy S.J., Kim S., Lin M.C., Manocha D. Simulating heterogeneous crowd behaviors using personality trait theory // SCA'11 Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. ACM New York. 2011. P. 43-52.
19. Hartigan J.A., Wong M.A. A K-Means Clustering Algorithm // Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979. Vol. 28, № 1. P. 100-108.
20. Helbing D., Farkas I., Vicsek T. Simulating dynamical features of escape panic // Nature. 2000. No. 407. P. 487-490.
21. Mintz A. Non-adaptive group behavior // Journal of Abnormal Psychology. 1951. Vol. 46, No. 2. P. 150-159.
22. Ruspini E.H. Numerical methods for fuzzy clustering // Information Sciences. 1970. Vol. 2, No. 3. P. 319–350.
23. Ward J.H. Hierarchical grouping to optimize an objective function // Journal of the American Statistical Association. 1963. Vol. 58, No. 301. P. 236–244.