

# Об одном методе персонализации поиска информации

Рыжов А.П., Огородников Н.М.

В работе рассматривается задача персонализации нечётких пользовательских понятий с помощью действия модификаторов на данные нечёткие множества. Исследован и модифицирован существующий алгоритм персонализации, для новой его версии получена оценка числа выполняемых итераций.

**Ключевые слова:** нечёткие множества, персонализация.

## 1. Введение

За последние несколько лет произошло качественное изменение использования информации людьми в решении повседневных задач. Количество накопленной информации в доступном для электронной обработки виде изменяется зеттабайтами (триллион гигабайт). Так, в недавно проведенном исследовании ведущих компаний Seagate ([www.seagate.com](http://www.seagate.com)) и IDC (<https://www.idc.com>) [1] подсчитано, что в 2016 году объем данных измерялся 16 Збайт, а к 2025 году этот показатель увеличится до 163 Збайт. Там же утверждается, что к 2025 году около 20% всей информации будут играть критически важную роль в повседневной жизни, а примерно 10% этих данных будут "сверхкритичными". Кроме того, прогнозируется, что в 2025 году почти 20% генерируемых данных будут представлять собой информацию, получаемую в режиме реального времени. Количество активных SIM-карт, используемых в телефонах, смартфонах и планшетах, впервые в истории превысило число живущих на Земле людей в 2014 году и продолжает расти [2]. Количество сервисов, используемых в повседневной жизни, также растет: трудно представить себе современного человека, не использующего поиск в интернет, новостные агрегаторы, социальные сети, электронные карты, заказ билетов, и многие другие ставшие привычными сервисы.

В сложившейся ситуации способы взаимодействия пользователя с информационными ресурсами также меняются. Одним из признанных трендов таких изменений является персонализация [4-15]. Компании видят существенное влияние персонализации на многие сферы бизнеса от маркетинга и продаж до оптимизации операционной деятельности и принятия стратегических решений. Это справедливо для многих индустрий от массовой торговли до индустрии моды. В этой связи разработка и исследование моделей и алгоритмов персонализации видится востребованной задачей.

Впервые идея алгоритма персонализации была предложена в работе [16], в [23] дана ее детализация и рассмотрены варианты оптимизации. Примеры использования таких моделей в компьютерном обучении, оптимизации новостных лет, социальных сетях представлены в работах [17-22].

Базовой для многих моделей персонализации является задача поиска информации. В общем виде задача поиска информации для человеко-компьютерных систем рассмотрена в [24]. В [25] предложен и изучен один из возможных алгоритмов персонализации на основе итерационного уточнения функций принадлежности. В качестве механизма такого уточнения предлагалось использовать сдвиг функций принадлежности, однако вопрос определения шага для такого сдвига остался открытым, что не позволяет использовать алгоритм в практических задачах. В данной работе предлагается алгоритм, свободный от такого недостатка.

Работа организована следующим образом. В разделе 2 приведена формализация задачи поиска информации. Алгоритм персонализации на основе известного алгоритма нечеткой кластеризации *c-means* представлен в разделе 3, его обобщение для любого эмпирического распределения объектов в универсальном множестве описано в разделе 4. В заключении (раздел 5) подведены краткие итоги исследования и сформулированы связанные с ними задачи.

## 2. Формализация задачи

Рассмотрим конечное множество объектов, каждый из которых характеризуется некоторым числовым параметром. Например, это могут быть товары с ценой в качестве характеристики. Важно лишь то, что характеристика объекта принимает значения из ограниченного подмножества числовой прямой, таким образом, считаем, что на множестве объектов задан порядок. Будем называть всё множество объектов базой данных

или универсальным множеством и обозначим его за  $U$ . Считаем также, что объекты бывают (условно) *большими* и *маленькими*, то есть, характеризуются своей величиной, а их характеристика описывается словами естественного языка. Это порождает нечёткость, поскольку понятия *большой* и *маленький* не поддаются строгому определению, комфортно каждому человеку. По этим соображениям рассмотрим нечёткие множества *больших* и *маленьких* объектов с функциями принадлежности  $\mu_1$  и  $\mu_2$  соответственно, определёнными на универсальном множестве  $U$ . Ясно, что для разных людей (разных пользователей базы данных) функции  $\mu_1$  и  $\mu_2$  будут различаться в силу того, что каждый по-своему определяет *большие* и *маленькие* объекты. Далее речь пойдёт о том, как подобрать набор объектов из базы, комфортный конкретному пользователю, другими словами, как построить подходящие ему функции принадлежности  $\mu_1$  и  $\mu_2$ .

В соответствии с принятой в теории нечётких множеств терминологией [21] рассмотрим лингвистическую переменную и её терм-множество. Оно состоит из базовых терминов и их комбинаций с модификаторами. В нашем случае есть два базовых термина - *большие* и *маленькие* объекты. На данном этапе ограничимся четырьмя модификаторами - *значительно больше*, *слегка больше*, *значительно меньше* и *слегка меньше*, однако, ниже рассмотрим и произвольное их количество. Так как данные модификаторы представляют собой конструкции естественного языка, будем считать, что именно они используются в запросах пользователя к базе данных.

Теперь можно переформулировать задачу следующим образом: пусть у нас есть конечное универсальное множество объектов  $U$ , требуется задать на нём начальные функции принадлежности  $\mu_1$  и  $\mu_2$ , отвечающие *большим* и *маленьким* объектам, и сформулировать правило их изменения под действием каждого из модификаторов.

В работе [26] описан возможный вариант алгоритма построения и модификации функций  $\mu_1$  и  $\mu_2$ , основывающийся на нечёткой кластеризации *c-means*. Мы сначала рассмотрим этот алгоритм подробнее, а затем предложим способ его улучшения.

### 3. Алгоритм персонализации на основе *c-means*

При данном подходе к универсальному множеству  $U$  применяется алгоритм нечёткой кластеризации *c-means* с количеством кластеров, равным двум. Следовательно, для каждого объекта  $x$  из  $U$  определены  $\mu_1(x)$  и

$\mu_2(x)$  - коэффициенты его принадлежности нечётким множествам *больших* и *маленьких* объектов. Затем, для всех  $x$ , не меньших центра кластера  $c_1$ ,  $\mu_1(x)$  полагается равным 1, а  $\mu_2(x)$  - равным 0, а для всех  $x$ , не больших центра кластера  $c_2$ , значение  $\mu_1(x)$  берётся равным 0, а  $\mu_2(x)$  равным 1. Теперь функция  $\mu_1(x)$  является неубывающей на всём  $U$ , а функция  $\mu_2(x)$  - невозрастающей, что согласуется с порядком, определённым на элементах множества  $U$ . Множество тех  $x$ , для которых  $\mu_1(x) = 1$ , обозначим через  $U_1$ , а множество тех  $x$ , для которых  $\mu_2(x) = 1$  - через  $U_2$ . Построенные функции принадлежности изображены на картинке ниже.

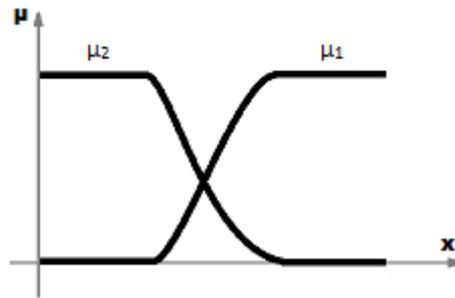


Рис.1. Функции принадлежности  $\mu_1$  и  $\mu_2$

Сам же алгоритм персонализации или определения действия модификаторов опирался на простое соображение: если при данной функции принадлежности множества *больших* объектов мы хотим получить функцию принадлежности множества объектов, которые *значительно больше*, то нас точно не интересуют объекты из  $U_2$ , эталонные *маленькие* объекты. Значит, при определении действия модификатора *значительно больше* будем полагать новые функции  $\mu_1(x)$  и  $\mu_2(x)$  равными 0 и 1 соответственно для всех  $x$  из  $U_2$ , а для остальных  $x$  снова применяем нечёткую кластеризацию и переопределяем коэффициенты принадлежности для всех  $x \geq c_1$  и всех  $x \leq c_2$ , как делали это ранее. Иными словами, чтобы определить модификатор *значительно больше*, надо удалить из рассмотрения все эталонные *маленькие* объекты. Аналогично, для модификатора *значительно меньше* отбрасываем все эталонные *большие* объекты. Для модификатора *слегка больше* будем отбрасывать не все объекты из  $U_2$ , а только те, что не больше его медианы; для *слегка меньше* - те, что не меньше медианы множества  $U_1$ .

Более подробное описание такого алгоритма и некоторые его свойства можно найти в [26], здесь же нас будет интересовать его поведение. Для этого обозначим модификатор *значительно больше* через  $R$ , модификатор *значительно меньше* через  $L$ , модификатор *слегка больше* через  $r$ , а *слегка меньше* через  $l$ . Центры кластеров, получившиеся после применения модификаторов, будем обозначать, приписывая к  $c_1$  или  $c_2$  слева обозначение соответствующего модификатора (например,  $Rc_1$  обозначает центр кластера *больших* объектов после применения модификатора *значительно больше*). Исследуем, что будет происходить при чередовании модификаторов  $R$  и  $L$ . Сначала применяем  $R$ . Тогда  $Rc_1 > c_1$  и  $Rc_2 > c_2$ , так как мы отбросили все объекты, не большие  $c_2$ . Далее действует  $L$ .  $LRc_1 < c_1$  и  $LRc_2 < c_2$ , ведь  $c_1$  и  $c_2$  были получены без отбрасывания объектов, а для построения  $LRc_1$  и  $LRc_2$  были отброшены объекты, большие, чем  $Rc_1$ . Теперь снова применяем  $R$  и получаем, что  $c_1 < RLRc_1 < Rc_1$  и  $c_2 < RLRc_2 < Rc_2$ . Рассуждая точно так же дальше, можно увидеть, что центр кластера  $c_1$  смещается следующим образом:  $c_1 \rightarrow c_1 < Rc_1 \rightarrow LRc_1 < c_1 < Rc_1 \rightarrow RLRc_1 < Rc_1 \rightarrow LRRLc_1 < LRc_1 < c_1 < RLRc_1 < Rc_1 \rightarrow RLRRLc_1 < LRc_1 < c_1 < RLRRLc_1 < RLRc_1 < Rc_1 \rightarrow \dots$  Центр кластера  $c_2$  смещается аналогично. Если же начать чередование модификаторов с  $L$ , то будет "симметричная" ситуация:  $c_1 \rightarrow Lc_1 < c_1 \rightarrow Lc_1 < c_1 < RLc_1 \rightarrow Lc_1 < LRLc_1 < c_1 < RLc_1 \rightarrow Lc_1 < LRLc_1 < c_1 < RLc_1 < RLRLc_1 \rightarrow Lc_1 < LRLc_1 < LRLRLc_1 < c_1 < RLc_1 < RLRLc_1 \rightarrow \dots$  Центр кластера  $c_2$  смещается аналогично. Так как в базе конечное число объектов, то на определённом шаге мы получим, что в обоих кластерах единичную принадлежность имеют те же самые объекты, что и при  $c_1$  и  $c_2$  в качестве центров кластеров. Причём, если чередование модификаторов начинать с  $R$ , то эта ситуация возникнет после какого-то применения  $R$ , а если начинать с  $L$  - после какого-то применения  $L$ . Аналогичные рассуждения можно провести, заменив  $R$  на  $r$ , а  $L$  на  $l$  на каждом шаге. Однако если на каких-то шагах заменять  $R$  на  $r$ , а на каких-то нет, то смещения центров кластеров могут отличаться от приведённых выше. Например, рассмотрим такую последовательность модификаторов:  $rLR$ . Тогда  $c_2 < rc_2$ , далее,  $Lrc_2 < c_2 < rc_2$ , а положение центра кластера  $R Lrc_2$  уже невозможно определить, ибо мы не знаем, каких объектов больше: меньших, чем  $Lrc_2$ , или меньших, чем медиана множества  $U_2$  на первом шаге. Таким образом, выполняется следующая теорема.

**Теорема 1.** При чередовании модификаторов  $R$  и  $L$  (или  $r$  и  $l$ ) за конечное число шагов получим такие же эталонные объекты в обеих кластерах, как и до применения модификаторов.

С одной стороны, хорошо, что, чередуя модификаторы, пользователь в конечном итоге получит то же, что и до их применения, но, с другой стороны, можно несколько раз применить  $R$ , а потом один раз  $L$  и получить центры кластеров левее, чем начальные  $c_1$  и  $c_2$ . Другими словами, запрашивая несколько раз объекты *больше*, а потом один раз *меньше*, получаем объекты, меньшие, чем были изначально. Это недостаток алгоритма, поэтому его необходимо модифицировать.

#### 4. Алгоритм персонализации на основе эмпирического распределения объектов в универсальном множестве

Пример с цепочкой модификаторов  $rLR$  показывает, что мы не можем предсказать поведение алгоритма, так как не знаем, сколько объектов на данной итерации содержат множества  $U_1$  и  $U_2$ . А из ситуации, описанной после теоремы 1, можно извлечь вывод, что необходимо каким-то образом ограничивать сдвиги функций принадлежности на каждом запросе. Эти соображения наталкивают на идею отказаться от нечёткой кластеризации в пользу привязки нечётких множеств к эмпирическому распределению объектов и не использовать на данной итерации объекты, которые были отброшены на предыдущих шагах.

Для этого рассмотрим квантили эмпирического распределения объектов в универсальном множестве  $U$ . Напомним, как для эмпирических распределений определяется квантиль уровня  $\alpha$ . Пусть  $x_0, \dots, x_{N-1}$  - все объекты из  $U$ . Составим вариационный ряд  $V_0 < V_1 < \dots < V_{N-1} = V_N$ . Мы добавляем  $V_N$ , чтобы формулы ниже были верны и для квантиля уровня 1, а также считаем, что все объекты различны. Определим число  $K = \lfloor \alpha \cdot (N-1) \rfloor$  (где  $\lfloor \cdot \rfloor$  - округление вниз до ближайшего целого числа) и сравним  $K + 1$  с  $\alpha \cdot N$ .

- 1) если  $K + 1 < \alpha \cdot N$ , то  $x_\alpha = V_{K+1}$ ;
- 2) если  $K + 1 > \alpha \cdot N$ , то  $x_\alpha = V_K$ ;
- 3)  $K + 1 = \alpha \cdot N$ , то  $x_\alpha = (V_K + V_{K+1})/2$ .

Теперь приступим к построению функций принадлежности. Идея, которая поможет нам это осуществить, заключается в том, чтобы работать

не с центрами кластеров, как ранее, а с "границами" множеств эталонных объектов. Выберем два квантиля эмпирического распределения объектов,  $x_{\alpha_1} > x_{\alpha_2}$ , причём, с точки зрения дисбаланса выделения из множества  $U$  эталонных *маленьких* и эталонных *больших* объектов будет разумным взять эти квантили такими, что  $\alpha_1 + \alpha_2 = 1$ . Для всех  $x \geq x_{\alpha_1}$  положим значения функций принадлежности  $\mu_1(x) = 1$ ,  $\mu_2(x) = 0$ , а для всех  $x \leq x_{\alpha_2}$   $\mu_1(x) = 0$ , а  $\mu_2(x) = 1$ . Числа  $x_{\alpha_1}$  и  $x_{\alpha_2}$  назовём границами множеств эталонов  $U_1$  и  $U_2$  соответственно. Для остальных  $x$  линейно построим функции  $\mu_1$  и  $\mu_2$ , то есть для  $x \in (x_{\alpha_2}; x_{\alpha_1})$

$$\mu_1(x) = \frac{x - x_{\alpha_2}}{x_{\alpha_1} - x_{\alpha_2}}, \quad \mu_2(x) = \frac{x - x_{\alpha_1}}{x_{\alpha_2} - x_{\alpha_1}}.$$

Таким образом, выбор конкретных значений  $\alpha_1$  и  $\alpha_2$  приведёт нас к явно заданным функциям принадлежности. В частности, рассмотрим  $\alpha_1 = 0.75$  и  $\alpha_2 = 0.25$ . Тогда верна следующая лемма.

**Лемма 1.** *При  $\alpha_1 = 0.75$ ,  $\alpha_2 = 0.25$  количество объектов в каждом из множеств  $U_1$ ,  $U_2$  равно  $\lceil N/4 \rceil$ , где  $\lceil \cdot \rceil$  - округление вверх до ближайшего целого числа.*

#### Доказательство

Рассмотри сначала множество  $U_2$  и определяющее его число  $\alpha_2 = 0.25$ . В этом случае, следуя определению квантиля эмпирического распределения выше, имеем  $K = \lfloor (N-1)/4 \rfloor$ ,  $K+1 = \lfloor (N+3)/4 \rfloor$ . Сравним теперь  $\lfloor (N+3)/4 \rfloor$  и  $N/4$ . Нетрудно получить, что  $\lfloor (N+3)/4 \rfloor = N/4$  для  $N \equiv 0(4)$ , для остальных  $N$  имеет место неравенство  $\lfloor (N+3)/4 \rfloor > N/4$ . Рассмотрим случай с неравенством. Так как оно выполняется, то  $x_{0.25} = V_{\lfloor (N-1)/4 \rfloor}$ , а значит, множество  $U_2$  будет содержать ровно те объекты, которые не больше, чем  $V_{\lfloor (N-1)/4 \rfloor}$ , то есть объекты  $V_0, \dots, V_{\lfloor (N-1)/4 \rfloor}$ . При рассматриваемых  $N$  верно  $\lfloor (N-1)/4 \rfloor = \lceil N/4 \rceil - 1$ , поэтому первые  $\lceil N/4 \rceil$  членов вариационного ряда и только они попадут в  $U_2$ . Теперь обратимся к случаю с  $N \equiv 0(4)$ , и пусть  $N = 4m$ . Так как  $\lfloor (N+3)/4 \rfloor = N/4$ , то

$$x_{0.25} = \frac{V_{\lfloor (N-1)/4 \rfloor} + V_{\lfloor (N+3)/4 \rfloor}}{2} = \frac{V_{m-1} + V_m}{2}.$$

А поскольку  $V_{\lceil N/4 \rceil - 1} = V_{m-1}$ , по определению, меньше, чем  $x_{0.25}$ , а  $V_{\lceil N/4 \rceil} = V_{\lfloor (N+3)/4 \rfloor} = V_m > x_{0.25}$ , то  $U_2$  снова содержит ровно  $\lceil N/4 \rceil$  наименьших объектов универсального множества  $U$ .

Перейдём к рассмотрению множества  $U_1$  и квантиля  $x_{\alpha_1} = x_{0.75}$ . Здесь  $K = \lfloor (3N-3)/4 \rfloor$ ,  $K+1 = \lfloor (3N+1)/4 \rfloor$ . Сравняя  $K+1$  и  $3N/4$ , получаем три возможных случая:

1) При  $N = 4m$  имеем  $\lfloor (3N + 1)/4 \rfloor = 3N/4 = 3m$ . Значит,

$$x_{0.75} = \frac{V_{\lfloor (3N-3)/4 \rfloor} + V_{\lfloor (3N+1)/4 \rfloor}}{2} = \frac{V_{3m-1} + V_{3m}}{2}.$$

Заметим, что  $N - \lceil N/4 \rceil = 3m$ , и так как  $V_{3m-1} < x_{0.75} < V_{3m}$ , то  $\lceil N/4 \rceil$  последних объектов вариационного ряда  $V_0 < \dots < V_{N-1}$  и только они попадут в множество  $U_1$ .

2) Если  $N = 4m + 1$ , то  $\lfloor (3N + 1)/4 \rfloor > 3N/4$ , поэтому  $x_{0.75} = V_{\lfloor (3N-3)/4 \rfloor} = V_{3m}$ . Также,  $N - \lceil N/4 \rceil = 3m$ , значит, в этом случае множество  $U_1$  снова содержит  $\lceil N/4 \rceil$  наибольших объектов из  $U$  и только их.

3) Наконец,  $\lfloor (3N + 1)/4 \rfloor < 3N/4$  при  $N$ , равных  $4m + 2$  и  $4m + 3$ , тогда  $x_{0.75} = V_{\lfloor (3N+1)/4 \rfloor}$ . То есть,  $V_{\lfloor (3N+1)/4 \rfloor} = V_{3m+1}$ , а  $N - \lceil N/4 \rceil = 3m + 1$  для  $N = 4m + 2$ , и в этом случае получаем требуемое утверждение, а для  $N = 4m + 3$ , аналогично,  $V_{\lfloor (3N+1)/4 \rfloor} = V_{3m+2}$  и  $N - \lceil N/4 \rceil = 3m + 2$ , следовательно, и здесь  $\lceil N/4 \rceil$  наибольших объектов множества  $U$  образуют множество  $U_1$ .

Лемма 1 доказана.

Теперь нам известны мощности множеств  $U_1$  и  $U_2$ . Перейдём к описанию алгоритма изменения функций принадлежности  $\mu_1$  и  $\mu_2$  под действием тех же модификаторов, что и ранее:  $R$ ,  $L$ ,  $r$  и  $l$ , для краткости используем их обозначения вместо названий. Как и ранее, для опеределения модификатора  $R$  положим новые значения коэффициентов принадлежности  $\mu_1(x) = 0$  и  $\mu_2(x) = 1$  для всех  $x$  из множества  $U_2$ . Для оставшихся  $x$  строим функции принадлежности по квантилям эмпирического распределения объектов множества  $U \setminus U_2$ . Аналогичным образом определяется модификатор  $L$ :  $\mu_1(x) = 1$ ,  $\mu_2(x) = 0$  для всех  $x$  из множества  $U_1$ , для оставшихся объектов строим коэффициенты принадлежности по квантилям их эмпирического распределения. Чтобы определить модификатор  $r$ , будем полагать  $\mu_1(x) = 0$ ,  $\mu_2(x) = 1$  только для тех  $x$  из  $U_2$ , которые не больше его медианы, для остальных объектов, как и в предыдущих случаях, строим функции. Наконец, для определения модификатора  $l$  положим  $\mu_1(x) = 1$ ,  $\mu_2(x) = 0$  для всех  $x$ , не меньших медианы множества  $U_1$ , коэффициенты принадлежности для остальных объектов получим на основе их эмпирического распределения. Таким образом, для определения модификаторов мы всегда отбрасываем часть объектов, полагая их коэффициенты принадлежности равными 0 или 1, а для оставшихся проводим построение функций принадлежности.

Далее, будем считать, что объекты исключаются из рассмотрения безвозвратно, то есть, будучи отброшены на какой-либо итерации алгоритма, они не используются и на всех последующих итерациях. Под  $U$  будем понимать не всё универсальное множество, а только те его объекты, которые не были отброшены до данной итерации. Исследуем теперь поведение алгоритма при многократных запросах пользователя. Из построения функций принадлежности  $\mu_1$  и  $\mu_2$  следует, что мощность множества всех находящихся в рассмотрении объектов уменьшается хотя бы на единицу при переходе к следующей итерации до тех пор, пока текущие множества  $U_1$  и  $U_2$  содержат хотя бы по одному объекту. Из Леммы 1 получаем, что мощность каждого из множеств  $U_1$  и  $U_2$  равна 1, если на момент данной итерации остались не отброшенными как максимум 4 объекта. С другой стороны, вспоминая семантику наших модификаторов, будет разумным утверждать, что в ситуации, когда  $U_1$  и  $U_2$  содержат по одному объекту, алгоритм построения новых функций принадлежности не имеет смысла, так как если есть всего один эталонный *большой* и всего один эталонный *маленький* объект, то нельзя говорить об объектах, которые *больше* или *меньше*. Поэтому в качестве критерия останова алгоритма (помимо удовлетворённости пользователя текущими коэффициентами принадлежности, то есть ситуацией, когда пользователь нашел удовлетворяющий его объект) определим порог мощности множества  $U$ , равный 4. Далее полагаем, что удовлетворённость результатами работы алгоритма не наступает никогда, чтобы исследовать его поведение в этом случае, который является наихудшим в смысле количества выполняемых итераций.

Обратимся к Лемме 1 и установим, как меняется мощность множества  $U$  при последовательных итерациях алгоритма. Рассмотрим сначала модификаторы  $R$  и  $L$ . После первой итерации множество  $U$  будет содержать  $N - \lceil N/4 \rceil = \lfloor 3N/4 \rfloor$  объектов. Обозначим за  $N_i$  мощность  $U$  после  $i$ -й итерации и заметим, что она таким же образом выражается через  $N_{i-1}$ :  $N_i = \lfloor 3N_{i-1}/4 \rfloor$ ,  $N_0 = N$ . Аналогичные рассуждения, правда, с чуть более громоздкими выкладками можно провести и для модификаторов  $l$  и  $r$ :

$$N_i = N_{i-1} - \lceil \frac{\lceil N_{i-1}/4 \rceil}{2} \rceil = \lfloor \frac{2N_{i-1} - \lceil N_{i-1}/4 \rceil}{2} \rfloor = \lfloor \frac{\lfloor 7N_{i-1}/4 \rfloor}{2} \rfloor, N_0 = N.$$

Далее, пусть через  $M$  итераций с модификаторами  $R$  и  $L$  мощность множества  $U$  стала не больше 4; это случится, ибо на каждой итерации она уменьшается хотя бы на 1. Также пусть через  $m$  итераций с модификаторами  $r$  и  $l$  мощность множества  $U$  стала не больше 4. Тогда заметим,

что при использовании всех четырёх модификаторов, мощность  $U$  будет не превышать 4 через число итераций  $M' : t \leq M' \leq M$ . Таким образом, выполняется следующая теорема.

**Теорема 2.** Алгоритм персонализации функций принадлежности на основе эмпирического распределения объектов обладает свойствами:

- 1) При любых запросах пользователя он совершает конечное число итераций;
- 2) Если используются только модификаторы  $R$  и  $L$  и не наступает удовлетворённость результатами работы, то алгоритм остановится через  $M$  итераций, где  $M$  таково, что  $N_M = \lfloor 3N_{M-1}/4 \rfloor \leq 4$ ,  $N_0 = N$ ;
- 3) Если используются только модификаторы  $r$  и  $l$  и не наступает удовлетворённость результатами работы, то алгоритм остановится через  $t$  итераций, где  $t$  таково, что  $N_t = \lfloor \frac{\lfloor 7N_{t-1}/4 \rfloor}{2} \rfloor \leq 4$ ,  $N_0 = N$ ;
- 4) Если же используются все модификаторы  $R$ ,  $L$ ,  $r$  и  $l$  и не наступает удовлетворённость результатами работы, то алгоритм остановится через  $M'$  итераций, где  $M'$  таково, что  $t \leq M' \leq M$ .

## 5. Заключение

В данной работе была исследована задача персонализации нечётких пользовательских понятий на примере построения и модификации функций принадлежности двух нечётких множеств. Как результат можно выделить описание алгоритма персонализации и оценку количества его итераций в худшем случае.

Для последующего изучения перспективным выглядит вопрос о возможности алгоритма обнаружить в универсальном множестве заранее заданный объект. Также представляется интересным исследование произвольного числа функций принадлежности в данном контексте.

## Список литературы

- [1] David Reinsel, John Gantz, John Rydning. The digitization of the world: from edge to core. November 2018 (<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>)

- [2] В. Елистратов. Количество активных смартфонов и планшетов превысило население Земли. 7 октября 2014. <https://tjournal.ru/52596-more-phones-than-people>
- [3] The Biggest Marketing Trend Of 2017: Personalization. January 31, 2017. <https://fusiononemarketing.com/biggest-marketing-trend-2017-personalization/>
- [4] Dillon Baker. How Personalization Is Changing Content Marketing. March 31st, 2017, <https://contently.com/2017/03/31/personalization-changing-content-marketing/>
- [5] Jim Rund. HOW PERSONALIZATION WILL EVOLVE IN 2018, January 29, 2018, <http://adage.com/article/epsilon/personalization-evolve-2018/312048/>
- [6] Shep Hyken. Recommended Just For You: The Power Of Personalization. May 13, 2017, <https://www.forbes.com/sites/shephyken/2017/05/13/recommended-just-for-you-the-power-of-personalization/#613481826087>
- [7] Discover the Power of Personalization, <https://www.demandware.com/resources/power-of-personalized-shopping>
- [8] Heike Young. Welcome to 2018, The Year of Personalization. Here's 3 Things You Need to Know. DECEMBER 21, 2017, <https://www.demandware.com/blog/retail-intelligence/welcome-2018-year-personalization-heres-3-things-need-know>
- [9] Nigel Wixcey. Made-to-order: The rise of mass personalisation. The Deloitte Consumer Review. <https://www2.deloitte.com/tr/en/pages/consumer-business/articles/made-to-order-the-rise-of-mass-personalisation.html>
- [10] Andy Betts. A new era of personalization: The hyperconnected customer experience. January 23, 2018, <https://martechtoday.com/new-era-personalization-hyper-connected-customer-experience-209529>
- [11] Shauna Robinson. Personalization: The Next Big E-Learning Trend, <https://www.td.org/magazines/td-magazine/personalization-the-next-big-e-learning-trend>
- [12] Fashion in 2018. Getting Personal. BY BOF TEAM AND MCKINSEY & COMPANY, JANUARY 2, 2018, <https://www.businessoffashion.com/articles/intelligence/top-industry-trends-2018-4-getting-personal>

- [13] SPECIAL REPORT: PERSONALIZATION TECH TRENDS, <https://www.chiefmarketer.com/special-reports/special-report-personalization-tech-trends/>
- [14] Thijs Kuin . How to Start with Digital Marketing Personalization. 23 February, 2018, <https://www.accenture-insights.nl/en-us/articles/digital-market-personalization>
- [15] Natalie Mouradian. The Dieline's 2018 Trend Report: Brands Become Hyper-Personalized. February 27, 2018, <http://www.thedieline.com/blog/2018/2/26/brands-become-hyper-personalized>
- [16] Lyapin B. , Ryjov A. A Fuzzy Linguistic Interface for Data Bases in Nuclear Safety Problems. Fuzzy Logic and Intelligent Technologies in Nuclear Science. Proceedings of the 1st International FLINS Workshop, Mol, Belgium, September 14-16, 1994. Edited by Da Ruan, Pierre D'hondt, Paul Govaerts, Etienne E. Kerre, World Scientific. p. 212-215.
- [17] Рыжов А. П., Журавлев А. Д., Вахов А. Н., Кривцов В. В. Об одном подходе к персонификации обучения в рамках компьютерных обучающих систем. Интеллектуальные Системы. Теория и приложения. Т. 20, Вып. 3, 2016, с. 180-185.
- [18] A. Ryjov, A. Vakhov, V. Krivtsov, and A. Zhuravlev. Personalization and optimization of learning based on technology for evaluation and monitoring of complex processes: Uchi.ru case study. The 2016 International Conference on Computational Science & Computational Intelligence. Ed. by: Hamid R. Arabnia, Leonidas Deligiannidis, and Mary Yang. Las Vegas, Nevada, USA, 15-17 December 2016, pp. 378 - 381.
- [19] Рыжов А. П., Кривцов В.В., Журавлев А.Д. Некоторые задачи кластеризации и ранжирования для персонификации компьютерных обучающих систем. Дистанционные образовательные технологии: материалы I Всероссийской научно-практической интернет-конференции (г. Ялта, 19-23 сентября 2016 года) – Симферополь, ИТ «АРИАЛ», 2016. – С. 37-41.
- [20] Рыжов А.П., Новиков П.А. Об одной модели цифровых привычек. Интеллектуальные Системы Теория и приложения. Т. 21 Вып. 3, 2017, с. 91-102.
- [21] Рыжов А.П. Некоторые задачи оптимизации и персонификации социальных сетей. Saarbrucken, LAP, 2015, 88 с.

- [22] Ryjov A. Personalization of Social Networks: Adaptive Semantic Layer Approach. In: Social Networks: A Framework of Computational Intelligence. Witold Pedrycz and Shyi-Ming Chen (Eds.). Springer International Publishing Switzerland 2014, pp. 21-40.
- [23] Ryjov A. Personalization and Optimization of Information Retrieval: Adaptive Semantic Layer Approach. In: Zadeh L., Yager R., Shahbazova S., Reformat M., Kreinovich V. (eds) Recent Developments and the New Direction in Soft-Computing Foundations and Applications. Studies in Fuzziness and Soft Computing, vol 361, Springer, Cham, 2018, pp. 15-24.
- [24] Рыжов А.П. Модели поиска информации в нечеткой среде. Москва, Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 2004, 96 с. <http://www.intsys.msu.ru/staff/ryzhov/FuzzyRetrieval2010.htm>
- [25] Дмитриенко Г.С. О формализации и модификации понятий при нечетком поиске. Дипломная работа. Москва, МГУ им. М.В. Ломоносова, механико-математический факультет, кафедра математической теории интеллектуальных систем, 2009.
- [26] Ogorodnikov N. (2017) One Approach to the Description of Linguistic Uncertainties. In: Nguyen N., Papadopoulos G., Jędrzejowicz P., Trawiński B., Vossen G. (eds) Computational Collective Intelligence. ICCCI 2017. Lecture Notes in Computer Science, vol 10449. Springer, Cham. (ISBN:978-3-319-67077-5)

**On one method of information retrieval personalization  
Ryjov A.P., Ogorodnikov N.M.**

In this paper we consider the problem of fuzzy user concepts personalization using the action of modifiers on certain fuzzy sets. The existing algorithm of personalization was investigated and modified; for its new version the number of iterations performed was estimated.

*Keywords:* fuzzy sets, personalization.

