

КОДИРОВАНИЕ МЯГКИХ ЦИКЛИЧЕСКИХ ДАННЫХ ПРИ ПОМОЩИ ПСЕВДО-НЕЧЕТКОЙ МЕРЫ

Новикова С.В., Кремлева Э.Ш., Валитова Н.Л.

Казанский национальный исследовательский технический университет имени
А.Н. Туполева – КАИ, г. Казань

Поступила в редакцию 10.07.2019, после переработки 05.09.2019.

В работе предложен новый численный метод, позволяющий представлять числовые переменные, обладающие рядом качественных, циклически повторяющихся признаков, в виде вектора, элементы которого отражают степень соответствия значения переменной каждому из таких признаков. Для кодирования качественных признаков предложено использовать специальные функции, обладающие рядом заданных свойств. Разработанная процедура названа псевдо-нечетким кодированием. В результате полученный кодирующий вектор способен адекватно представлять исходную переменную, отражая ее качественные свойства. Приводятся результаты вычислительных экспериментов, подтверждающих эффективность предложенного метода.

Ключевые слова: циклические данные, мягкие вычисления, кодирование, нечеткая мера, кластерный анализ.

Вестник ТвГУ. Серия: Прикладная математика. 2019. № 3. С. 90–101.
<https://doi.org/10.26456/vtprm542>

Введение

В математике хорошо известно понятие «периодические процессы» – функции, значения которых повторяются через равные интервалы значений аргумента [1]. При этом и аргумент, и значение функции выражаются количественно, числом. Таковыми, например, являются основные тригонометрические функции ($f(x) = \sin(x)$, $f(x) = \cos(x)$, $f(x) = \operatorname{tg}(x)$), функция взятия дробной части вещественного числа ($f(x) = x - [x]$, где $[x]$ – целая часть числа x), функция Дирихле и т.д.

Однако на практике существуют задачи, где значением процесса выступает не число, а некоторое качественное понятие, свойство аргумента [2]. Ярким примером может служить связь конкретной даты со временем года, к которому она относится: здесь аргументом периодической функции будут являться даты (количественные данные), а значением функции – название сезона (качественное значение). При этом, очевидно, что качественные значения функции являются периодическими.

Аналогично качественное понятие «время дня» является периодическим свойством количественного параметра- времени.

Назовем количественные данные, *качественные свойства* которых повторяются с заданной периодичностью, *мягкими циклическими данными*.

Сложность вычислений с такими данными заключается в невозможности применения к ним стандартных арифметических операций напрямую. Так, например, стандартная операция определения разности между моментами времени: 23ч. 45 мин. и 00ч. 15 мин. покажет расстояние в 23 часа 30 минут, хотя очевидно, что реальное расстояние составляет 30 минут, и оба момента времени близки качественному определению «полночь». Аналогично при оперировании с качественными понятиями времен года: при анализе расстояний между тремя датами: 5 января, 15 марта и 23 декабря прямые вычисления разности покажут, что даты 5 января и 15 марта гораздо «ближе» друг к другу, чем даты 5 января и 23 декабря, хотя с точки зрения времени года обе даты относятся к понятию «зима» и реально отстоят друг от друга на 12 дней.

Необходимость корректных расчетов с мягкими циклическими данными требует разработки специального численного метода определения истинных расстояний между ними, адекватных с точки зрения близости их качественных свойств.

1. Нечеткое кодирование

Для вычисления расстояний между двумя мягкими циклическими наборами данных предлагается предварительно закодировать их качественные характеристики количественными значениями. Чем «ближе» два набора x с качественной точки зрения, тем меньше должны отличаться их закодированные количественные представления y :

$$f^{\text{качество}}(x) = y. \quad (1)$$

Для решения подобной задачи как нельзя лучше подходит аппарат нечеткой логики [3]. Аналогично тому, как в нечеткой логике лингвистические переменные представляются в виде нечетких множеств с заданными функциями принадлежности, для кодирования мягких циклических данных предлагается использовать специальную функцию-кодировщик, преобразующую одно четкое значение – количественное выражение циклической переменной – в другое, отражающее степень выраженности качественного признака для представленной переменной.

В отличие от классической нечеткой функции принадлежности, функция-кодировщик может принимать любые вещественные значения, и должна быть симметрична относительно заданного центра, в котором она достигает своего максимума. Так, для понятия «время суток» функция, кодирующая свойство «ночь» достигает своего максимума в значении «00ч. 00мин.». При удалении от центра значение функции-кодировщика симметрично уменьшается.

Подобными свойствами обладают, например, симметричные треугольные (2) и трапецевидные (3) функции, из непрерывных – гауссианы (4), и т.д. [4-5]

$$f^{\text{качество}}(x) = \begin{cases} 1 - \frac{b-x}{b-a}, & a \leq x \leq b, \\ 1 - \frac{x-2b+a}{b-a}, & b \leq x \leq 2b-a, \\ 0, & \text{в остальных случаях,} \end{cases} \quad (2)$$

$$f_{\text{качество}}(x) = \begin{cases} 1 - \frac{b-x}{b-a}, & a \leq x \leq b, \\ 1, & b \leq x \leq c, \\ 1 - \frac{x-c}{b-a}, & c \leq x \leq b+c-a, \\ 0, & \text{в остальных случаях,} \end{cases} \quad (3)$$

$$f_{\text{качество}}(x) = \exp \left[- \left(\frac{x-m}{\sigma} \right)^2 \right]. \quad (4)$$

Назовем функцию кодирования *псевдо-нечеткой мерой*, а сам способ представления мягких циклических данных – *псевдо-нечетким кодированием*.

Метод псевдо-нечеткого кодирования, таким образом, состоит в вычислении для мягкой циклической переменной псевдо-нечеткой меры. Если циклическая переменная теоретически может относиться к нескольким качественным понятиям, псевдо-нечеткая мера рассчитывается отдельно для каждого из них, и затем из полученных величин составляется вектор закодированных признаков. Полученный вектор будет представлять мягкую циклическую переменную в целом.

2. Численный метод псевдо-нечеткого кодирования

Пусть x – мягкая циклическая переменная, обладающая набором из s качественных признаков $A = (a_1, a_2, \dots, a_s)$. Численный метод кодирования заключается в следующем:

1. Для каждого качественного признака a_i ($i = 1, s$) задать симметричную функцию-кодировщик $f^{a_i}(x)$:
 - (а) определить числовое значение c_i , в наибольшей степени отвечающее качественному признаку a_i ;
 - (б) определить числовое значение b_i , – максимальное отклонение от значения c_i , при котором качественное свойство a_i сохраняется;
 - (в) положить $m_i = c_i$, $\sigma_i = \frac{b_i}{3}$;
 - (г) задать симметричную функцию-кодировщик $f^{a_i}(x)$ в виде гауссиана:

$$f^{a_i}(x) = \exp \left[- \left(\frac{x-m_i}{\sigma_i} \right)^2 \right].$$

2. Вычислить значения функций $f^{a_i}(x)$, $i = \overline{1, n}$ для заданного значения переменной x .
3. Сформировать вектор $Y = (y_1, y_2, \dots, y_s)$, где $y_i = f^{a_i}(x)$.
4. Считать вектор Y закодированным представлением мягкой циклической переменной x .

Используя представление Y вместо значения x , возможно проводить все вычисления с парами мягких циклических переменных при условии одинаковой размерности кодирующих векторов.

3. Применение метода псевдонечеткого кодирования в задачах кластеризации

Практическое применение разработанного метода рассмотрим на примере задачи кластеризации данных, состоящей в группировке многомерных объектов-векторов в непересекающиеся подмножества в том или ином смысле однородных данных [6-8]. Под однородностью как правило понимается близость векторов в смысле некоторой метрики, чаще всего Евклидовой. В этом случае мера близости вычисляется как расстояние между векторами $X^1 = (x_1^1, x_2^1, \dots, x_n^1)$, $X^2 = (x_1^2, x_2^2, \dots, x_n^2)$ по формуле:

$$d(X^1, X^2) = \sqrt{\sum_{i=1}^n (x_i^1 - x_i^2)^2}. \quad (5)$$

Если часть элементов x_k векторов X кластеризуемых данных является мягкими циклическими значениями, то встает проблема адекватного вычисления расстояний между такими объектами для последующего вычисления расстояний между векторами исходного множества.

Для решения возникшей проблемы необходимо закодировать каждую мягкую циклическую переменную вектором признаков согласно разработанному выше методу псевдо-нечеткого кодирования. При этом каждой такой переменной x_k будет соответствовать вектор:

$$Y^k = (y_1^k, y_2^k, \dots, y_s^k). \quad (6)$$

В результате получим расширенный вектор кластеризуемых данных вида:

$$Z = (x_1, x_2, \dots, x_{k-1}, y_1^k, y_2^k, \dots, y_s^k, x_{k+1}, \dots, x_n) = (z_1, \dots, z_p), \quad (7)$$

где s – количество качественных признаков, n – размерность исходного вектора данных, $p = n + s - 1$.

Далее необходимо подвергнуть кластеризации набор расширенных векторов Z одним из кластеризующих алгоритмов. Полученные в результате кластеры будут отражать структуру исходного множества объектов X .

В качестве примера применения метода псевдо-нечеткого кодирования для кластеризации рассмотрим кластеризацию данных о замере концентраций загрязняющих веществ в водоемах Республики Татарстан в период с 03 января 2014 по 17.10.2017 года. За охваченный период было произведено 391 измерение в 15 точках рек Волга, Казанка, Меша, Кама, Свияга, Сулица, а также на акватории Куйбышевского водохранилища [9]. Набор измеряемых данных содержал 32 значимых параметра, характеризующих качество поверхностных вод (соли металлов, взвешенные вещества, загрязнения нефтепродуктами и фенолом, биологические загрязнения и проч.), а также дату забора каждой пробы (Табл.1). В качестве практической задачи необходимо сгруппировать вектора данных о загрязнениях

в однородные кластеры для выявления зон со схожими проблемами и выработки для них адекватных типовых рекомендаций по природоохране.

Таким образом, исследуемые вектора имели вид:

$$X^j = (x_1^j, x_2^j, \dots, x_n^j), n = 32, j = \overline{1, 391}.$$

Таблица 1: Физический смысл компонентов кластеризуемых векторов

| Комп-та | Физический смысл | Комп-та | Физический смысл |
|----------|---|----------|--|
| x_1 | Координаты точки отбора проб воды | x_{17} | Нефтепродукты |
| x_2 | Дата проведения анализа | x_{18} | Никель |
| x_3 | Синтетические поверхностно-активные вещества (СПАВ) | x_{19} | Нитраты |
| x_4 | Ион алюминия | x_{20} | Нитриты |
| x_5 | Аммоний ион | x_{21} | Минерализация |
| x_6 | Биологическое потребление кислорода за 5 дней (БПК ₅) | x_{22} | Прозрачность |
| x_7 | Взвешенные вещества | x_{23} | Свинец |
| x_8 | Гидрокарбонаты | x_{24} | Сульфаты |
| x_9 | Железо | x_{25} | Температура |
| x_{10} | Жесткость | x_{26} | Фенол |
| x_{11} | Кальций | x_{27} | Фосфат ион |
| x_{12} | Кислород раствор | x_{28} | Фосфор общий |
| x_{13} | Магний | x_{29} | Химическое потребление кислорода (ХПК) |
| x_{14} | Марганец | x_{30} | Хлориды |
| x_{15} | Медь | x_{31} | Цинк |
| x_{16} | Натрий+калий | x_{32} | Эл. проводность |

Как отмечают специалисты предметной области, дата забора пробы имеет важное значение для определения характеристик загрязнения, так как на развитие многих загрязняющих факторов напрямую влияет сезонность [10]. Так уровень насыщения воды кислородом в летние месяцы стабильно выше, чем зимой, поэтому относить некоторую пробу воды, например, к пробам, обедненным кислородом, нужно с учетом фактора сезонности (одно и то же значение будет нормальным для зимы, но аномально низким для лета). Это же относится и к некоторым другим загрязняющим параметрам.

Таким образом, компонента x_2 – дата забора пробы, является важной для определения схожих групп векторов, однако «прямое» ее использование даст некорректный результат – разница между заборами проб 03.01.2014 и 20.01.2016 окажется равной двум годам и 17 дням, то есть согласно вычислениям, по компоненте x_2 два вектора будут сильно различаться. Однако совершенно очевидно, что

исходя из понятия «сезонности» два забора проб произошли примерно в одно и то же время – в середине зимы.

В связи с этим компоненту x_2 следует рассматривать как мягкую циклическую переменную с четырьмя возможными качественными признаками:

$$x_2 \in A = (\text{«зима», «весна», «лето», «осень»}).$$

Таким образом, после кодирования единственную переменную x_2 будем представлять закодированный вектор из 4-х компонент: $Y^2 = (y_1^2, y_2^2, y_3^2, y_4^2)$.

Кодирующие функции зададим в виде гауссианов с параметрами:

а) Зима: $c_1 = 15$ января, $b_1 = 60$ дней

$$y_1^2 = f^{\text{Зима}}(x) = \exp \left[- \left(\frac{x - 15.01}{20} \right)^2 \right].$$

б) Весна: $c_2 = 15$ апреля, $b_2 = 60$ дней

$$y_2^2 = f^{\text{Весна}}(x) = \exp \left[- \left(\frac{x - 15.04}{20} \right)^2 \right].$$

в) Лето: $c_3 = 15$ июля, $b_3 = 60$ дней

$$y_3^2 = f^{\text{Лето}}(x) = \exp \left[- \left(\frac{x - 15.07}{20} \right)^2 \right].$$

г) Осень: $c_4 = 15$ октября, $b_4 = 60$ дней

$$y_4^2 = f^{\text{Осень}}(x) = \exp \left[- \left(\frac{x - 15.10}{20} \right)^2 \right].$$

Результат преобразования данных для четырех качественных признаков:

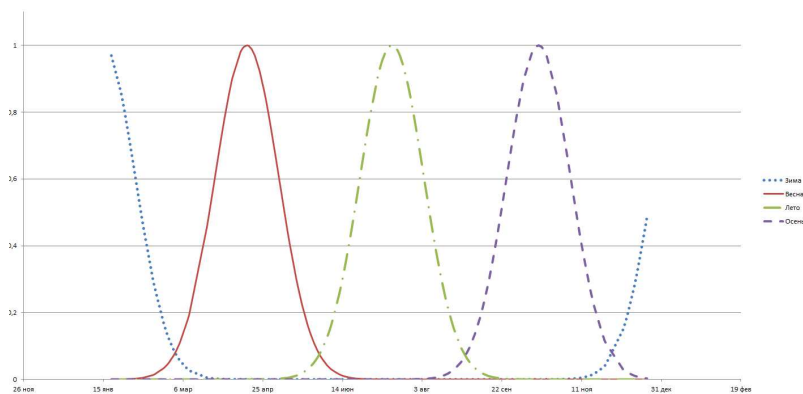


Рис. 1: Графическое представление кодирующих функций для мягкой циклической переменной x_2

Таблица 2: Пример вычисления элементов кодирующих векторов для различных дат забора проб (четких значений переменной x_2)

| Дата анализа x_2 | Зима y_1^2 | Весна y_2^2 | Лето y_3^2 | Осень y_4^2 |
|-----------------------|-----------------|------------------|-----------------|------------------|
| 28 января 2015 г. | 0,8095716 | 0,000604 | 4,77E-16 | 2,01E-37 |
| 1 февраля 2016 г. | 0,6968048 | 0,00128 | 2,51E-15 | 2,65E-36 |
| 2 февраля 2015 г. | 0,6669768 | 0,001534 | 3,77E-15 | 5,01E-36 |
| 3 февраля 2014 г. | 0,6368316 | 0,001834 | 5,66E-15 | 9,47E-36 |
| ... | ... | ... | ... | ... |
| 4 марта 2014 г. | 0,0561348 | 0,110251 | 2,5E-10 | 3,29E-28 |
| 9 марта 2015 г. | 0,0298596 | 0,18064 | 1,28E-09 | 5,31E-27 |
| 10 марта 2015 г. | 0,0261214 | 0,197899 | 1,75E-09 | 9,19E-27 |
| 21 марта 2014 г. | 0,0050861 | 0,457833 | 4,96E-08 | 3,26E-24 |
| ... | ... | ... | ... | ... |
| 2 мая 2016 г. | 6,091E-07 | 0,696805 | 0,001065 | 1,1E-15 |
| 10 мая 2016 г. | 6,616E-08 | 0,457833 | 0,004318 | 2,8E-14 |
| 12 мая 2015 г. | 3,704E-08 | 0,402021 | 0,005976 | 6,15E-14 |
| 12 мая 2015 г. | 3,704E-08 | 0,402021 | 0,005976 | 6,15E-14 |
| 16 мая 2016 г. | 1,127E-08 | 0,300818 | 0,011109 | 2,87E-13 |
| ... | ... | ... | ... | ... |
| 23 мая 2014 г. | 1,275E-09 | 0,164474 | 0,02986 | 3,86E-12 |
| 24 мая 2016 г. | 9,25E-10 | 0,149382 | 0,034047 | 5,53E-12 |
| ... | ... | ... | ... | ... |

Таким образом, для кластеризации предъявляются расширенные за счет псевдонечеткого кодирования 32-х мерных векторов $X^j = (x_1^j, x_2^j \dots x_n^j)$ 36-тимерные вектора:

$$Z^j = (z_1^j, z_2^j \dots z_{n+4}^j), \quad (8)$$

где $(z_2^j, z_3^j, z_4^j, z_5^j) = Y^{2j}$.

В результате кластеризации расширенного набора данных при помощи нейронной самообучающейся сети Кохонена было выделено 4 компактные группы (Рис. 2).

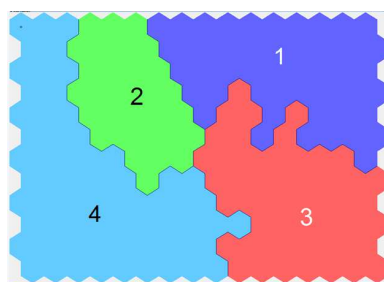


Рис. 2: Выделенные кластеры для расширенного набора данных

Затем в выделенных кластерах на основе раскраски SOM-карт было проанализировано распределение векторов данных по сезонам. Здесь оттенки красного символизируют высокие значения показателя, а синего – низкие. Остальные цвета представляют промежуточные числовые значения (Рис. 3).

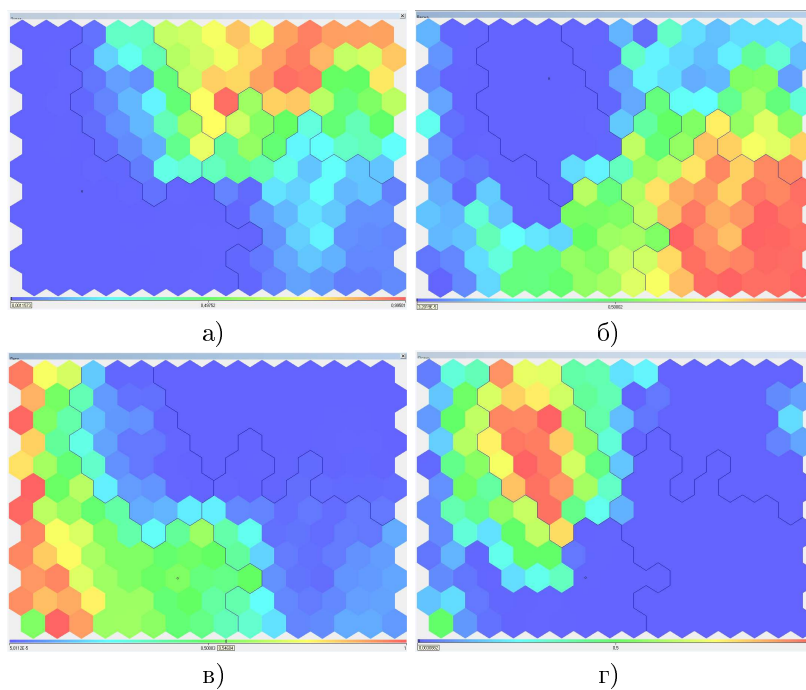


Рис. 3: Распределение данных о загрязнении водоемов по сезонам: а) зима, б) весна, в) лето, г) осень

Хорошо прослеживается сильное влияние сезонов на распределение данных: в кластер №1 попали в основном замеры, сделанные зимой, в кластер №2 – осенью, кластер №3 – весной, и кластер №4 – летом. Автоматическая группировка правильно отслеживает близость сезонных характеристик замеров, что говорит об эффективности применения метода псевдонечеткого кодирования для замены псевдонечетких данных кодирующими векторами.

Заключение

Разработанный численный метод псевдо-нечеткого кодирования позволяет представлять числовые переменные, обладающие рядом качественных, циклически повторяющихся признаков, в виде вектора, элементы которого отражают степень соответствия значения переменной каждому из таких признаков. В результате полученный кодирующий вектор способен адекватно представлять исходную переменную, отражая ее качественные свойства. Проведенные вычислительные эксперименты доказали эффективность предложенного метода для реальных вычислений.

Список литературы

- [1] Далингер В.А., Симонженков С.Д. Теория функций действительного переменного. 2-е изд. М.: Издательство Юрайт, 2019. 242 с.
- [2] Тунакова Ю.А., Новикова С.В., Шагидуллина Р.А., Кремлева Э.Ш. Роль качественных оценок в задачах экологического управления в зонах действия полимерных производств // Вестник Казанского технологического университета. 2013. Т. 16, № 20. С. 276–279.
- [3] Ланге Ф. Нечеткая логика. СПб.: Страта, 2018. 116 с.
- [4] Катасёв А.С. Методы и алгоритмы формирования нечетких моделей оценки состояния объектов в условиях неопределенности // Вестник технологического университета. 2019. Т. 22, № 3. С. 138–147.
- [5] Данилаев Д.П., Емалетдинова Л.Ю. Нечеткая модель отбора профилей подготовки технических специалистов // Открытое образование. 2015. № 4. С. 28–32.
- [6] Гитис Л.Х. Статистическая классификация и кластерный анализ. М.: Изд-во Московского государственного горного университета, 2003. 158 с.
- [7] Новикова С.В., Емалетдинова Л.Ю., Максютин С.А. Методика анализа функционирования жилищно-коммунальной отрасли региона на основе банков данных и метода кластеризации // Вестник Казанского государственного технического университета им. А.Н. Туполева. 2008. № 1. С. 75–78.
- [8] Новикова С.В., Тунакова Ю.А., Габдрахманова Г.Н., Кузнецова О.Н. Кластерный анализ данных для интегральной характеристики безопасности питьевых вод для потребителей на территории г. Казани // Вестник Казанского технологического университета. 2016. Т. 19, № 19. С. 164–167.
- [9] Новикова С.В., Тунакова Ю.А., Кирпичников А.П., Шагидуллин А.Р., Валева В.С., Кремлева Э.Ш., Габдрахманова Г.Н. Использование инновационных информационных технологий для типизации поверхностных вод р. Волга и питающих ее малых рек // Вестник Казанского технологического университета. 2019. Т. 22, № 2. С. 150–154.
- [10] Комплексная оценка степени загрязненности воды реки Волга района г. Казани // Материалы XIV Международной научно-технической конференции “Наука, образование, производство в решении экологических проблем”, Экология-2018. Уфа: РИК УГАТУ, 2018. С. 130–136.

Образец цитирования

Новикова С.В., Кремлева Э.Ш., Валитова Н.Л. Кодирование мягких циклических данных при помощи псевдо-нечеткой меры // Вестник ТвГУ. Серия: Прикладная математика. 2019. № 3. С. 90–101. <https://doi.org/10.26456/vtprmk542>

Сведения об авторах**1. Новикова Светлана Владимировна**

профессор кафедры Прикладной математики и информатики Казанского национального исследовательского технического университета им А.Н. Туполева – КАИ.

Россия, 420015, г. Казань, ул. Большая Красная, д. 55, КНИТУ-КАИ.

E-mail: sweta72@bk.ru

2. Кремлева Эльмира Шамильевна

старший преподаватель кафедры прикладной математики и информатики Казанского национального исследовательского технического университета им А.Н. Туполева – КАИ.

Россия, 420111, г. Казань, ул. К. Маркса, д. 10, КНИТУ-КАИ.

E-mail: e-smile29.04@mail.ru

3. Валитова Наталья Львовна

доцент кафедры прикладной математики и информатики Казанского национального исследовательского технического университета им А.Н. Туполева – КАИ.

Россия, 420111, г. Казань, ул. К. Маркса, д. 10, КНИТУ-КАИ.

E-mail: valinata@rambler.ru

SOFT CYCLIC DATA ENCODING USING A QUASI-FUZZY MEASURE

Novikova Svetlana Vladimirovna

Professor at the Department of Applied Mathematics and Computer Science, Kazan National Research Technical University named after A.N. Tupolev – KAI
Russia, 420015, Kazan, 55 Bolshaya Krasnaya str., KNRTU-KAI.
E-mail: sweta72@bk.ru

Kremleva Elmira Shamilyevna

Senior Lecturer at the Department of Applied Mathematics and Computer Science, Kazan National Research Technical University named after A.N. Tupolev – KAI
Russia, 420111, Kazan, 10 Karla Marx str., KNRTU-KAI.
E-mail: e-smile29.04@mail.ru

Valitova Natalya Lvovna

Associate Professor at the Department of Applied Mathematics and Computer Science, Kazan National Research Technical University named after A.N. Tupolev – KAI
Russia, 420111, Kazan, 10 Karla Marx str., KNRTU-KAI.
E-mail: valinata@rambler.ru

Received 10.07.2019, revised 05.09.2019.

In the article, the authors propose a new numerical method for representing quantitative variables that have several qualitative, cyclically repeating attributes. Such variables will be represented by a vector whose elements reflect the degree to which the value of the variable corresponds to each of the qualitative features. To encode quality features, the authors propose using special functions that have the specified properties. The developed procedure is called quasi-fuzzy coding. As a result, the coding vector is able to adequately represent the original variable, reflecting its qualitative properties. The article presents the results of computational experiments that confirm the effectiveness of the proposed method.

Keywords: cyclic data, soft calculations, coding, fuzzy measure, cluster analysis.

Citation

Novikova S.V., Kremleva E.Sh., Valitova N.L., “Soft cyclic data encoding using a quasi-fuzzy measure”, *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2019, № 3, 90–101 (in Russian).
<https://doi.org/10.26456/vtpmk542>

References

- [1] Dalinger V.A., Simonzhenkov S.D., *Teoriya funktsij dejstvitelnogo peremennogo [Theory of functions of a real variable]*, second edition, Urait Publ., Moscow, 2019 (in Russian), 242 pp.
- [2] Tunakova Yu.A., Novikova S.V., Shagidullina R.A., Kremleva E.Sh., “The role of qualitative assessments in the tasks of environmental management in the areas of polymer production”, *Vestnik Kazanskogo tekhnologicheskogo universiteta [Bulletin of Kazan Technological University]*, **16**:20 (2013), 276–279 (in Russian).
- [3] Lange F., *Nechetkaya logika [Fuzzy logic]*, Strata Publ., SPb., 2018 (in Russian), 116 pp.
- [4] Katasyov A.S., “Methods and algorithms for the formation of fuzzy models for assessing the state of objects in conditions of uncertainty”, *Vestnik Tekhnologicheskogo Universiteta [Bulletin of Technological University]*, **22**:3 (2019), 138–147 (in Russian).
- [5] Danilaev D.P., Emaletdinova L.Yu., “Fuzzy model for the selection of profiles for the training of technical specialists”, *Otkrytoe obrazovanie [Open education]*, 2015, № 4, 28–32 (in Russian).
- [6] Gitis L.Kh., *Statisticheskaya klassifikatsiya i klasternyj analiz [Statistical classification and cluster analysis]*, Publishing House of Moscow State Mining University, Moscow, 2003 (in Russian), 158 pp.
- [7] Novikova S.V., Emaletdinova L.Yu., Maksyutin S.A., “Methodology for analyzing the functioning of the housing and communal sector in the region based on data banks and the clustering method”, *Vestnik Kazanskogo Gosudarstvennogo Tekhnicheskogo Universiteta im. A.N. Tupoleva [Bulletin of Kazan State Technical University named after A.N. Tupolev]*, 2008, № 1, 75–78 (in Russian).
- [8] Novikova S.V., Tunakova Yu.A., Gabdrakhmanova G.N., Kuznetsova O.N., “Cluster data analysis for the integral characteristics of drinking water safety for consumers in the territory of Kazan”, *Vestnik Kazanskogo tekhnologicheskogo universiteta [Bulletin of Kazan Technological University]*, **19**:19 (2016), 164–167 (in Russian).
- [9] Novikova S.V., Tunakova Yu.A., Kirpichnikov A.P., Shagidullin A.R., Valeev V.S., Kremleva E.Sh., Gabdrakhmanova G.N., “Using innovative information technologies for typing surface waters Volga and small rivers feeding it”, *Vestnik Kazanskogo tekhnologicheskogo universiteta [Bulletin of Kazan Technological University]*, **22**:2 (2019), 150–154 (in Russian).
- [10] “Comprehensive assessment of the degree of water pollution in the Volga River district of Kazan”, *Materialy XIV Mezhdunarodnoj nauchno-tekhnicheskoy konferentsii “Nauka, obrazovanie, proizvodstvo v reshenii ekologicheskikh problem” [Materials of the XIV International Scientific and Technical Conference “Science, Education, Production in Solving Environmental Problems”]*, Ecology 2018, RIK UGATU Publ., Ufa, 2018, 130–136 (in Russian).