

УДК 004.891

М.П. Силич, С.В. Аксенов

Генерация наименований нечетких кластеров в пространстве разнородных признаков для зонирования территориальных объектов по внешним условиям

Предлагается метод генерации наименований на естественном языке для групп территориальных объектов с приблизительно похожими внешними условиями. Группы формируются путем нечеткой кластеризации значений неоднородных индикаторов, характеризующих внешний фактор. Пространство значений индикаторов разбивается на зоны, сопоставляемые базовым термам, заданным для каждого индикатора. Наименование того или иного кластера генерируется по эвристическим правилам в зависимости от его расположения в пространстве индикаторов путем комбинирования базовых термов и связок.

Ключевые слова: нечеткая кластеризация, наименования кластеров, генерирование термов, зонирование территорий.

doi: 10.21293/1818-0442-2018-21-4-81-86

Сравнительный анализ уровня развития некоторой сферы общественной жизни территориальных объектов (субъектов РФ, муниципальных образований, поселений, отдельных зданий, сооружений) позволяет классифицировать территории по уровням либо расставить их по рейтингу. Как правило, используется индикативный подход, при котором оценки (в виде рангов, баллов или лингвистических значений) выводятся на основе значений индикаторов для сравниваемых территориальных объектов (ТО) [1, 2]. Полученные оценки являются основой для принятия решений по исправлению ситуации в проблемных регионах [3, 4].

Однако неоднородность природно-климатических, социально-экономических, инфраструктурных и других внешних факторов снижает адекватность оценок, поскольку регионы изначально находятся в различных условиях [5, 6]. Зонирование территорий по внешним факторам, влияющим на исследуемую сферу, дает возможность сравнивать ТО в рамках групп с приблизительно похожими внешними условиями. Зонирование может осуществляться различными способами. Например, в [7] для зонирования территорий по признакам, влияющим на энергосбережение, предлагается использовать метод группировки на основе средневзвешенных показателей по множеству признаков или поэтапной группировки по степени близости значений признаков.

Другой подход состоит в применении методов нечеткой кластеризации, например, нечетких с-средних (FCM) или Густаффсона–Кесселя (GK) [8, 9]. Они позволяют учитывать неоднозначность в определении границ кластеров. Немаловажно, что методы кластеризации не используют субъективные знания эксперта и выполняются автоматически. Пользователю предъявляются результаты в виде степеней принадлежности объектов кластеризации нечетким кластерам. В качестве идентификаторов кластеров, как правило, используются порядковые номера. Однако желательно, чтобы наименования кластеров несли смысловую нагрузку и были выражены в тер-

минах естественного языка. Это облегчает пользователю интерпретацию результатов зонирования территорий на семантическом уровне. Например, наименованиями зон, полученных в результате кластеризации регионов по таким индикаторам, характеризующим климатические условия, как влажность воздуха и средняя годовая температура воздуха, могут быть «сухой холодный климат», «влажный теплый климат», «сухой умеренный климат» и т.д.

Основной сложностью является то, что заранее не известно каково будет расположение кластеров. Их количество также может варьироваться в ходе итерационного поиска наилучшего разбиения, и сколько в итоге будет получено кластеров – не известно. Даже после получения результатов кластеризации подбор наименований может вызывать затруднения, особенно в случае множества индикаторов, поскольку визуализировать кластеры в многомерном пространстве не всегда возможно [10–12].

Задача подбора наименований кластеров может рассматриваться как задача формирования термов лингвистической переменной [13]. Кластеры при этом представляют собой дискретно заданные многомерные функции принадлежности. В большинстве работ, посвященных вопросам генерации значений лингвистической переменной (например, в [14–16]), предполагается, что функции принадлежности термов являются одномерными и задаются аналитически или графически. В [17] авторами данной статьи был предложен метод генерации наименований кластеров, основная идея которого состоит в сравнении границ «идеальных» кластеров, сопоставленных термам базового терм-множества, и реальных кластеров, полученных в результате кластеризации.

Однако данный метод применим только в случае использования однородных признаков кластеризации, т.к. интерпретация значений индикаторов осуществляется на основе одного общего набора базовых термов, отражающих уровень развития оцениваемого фактора (например, «низкий», «средний», «высокий»). Оценки по всем индикаторам при

этом должны быть согласованы. В случае же зонирования территориальных объектов по внешним условиям очень часто используются разнородные признаки, не коррелирующие друг с другом. Еще один недостаток существующего метода состоит в том, что хотя и используется алгоритм нечеткой кластеризации, степени принадлежности объектов кластерам не учитываются, т.е. от нечетких кластеров осуществляется переход к четким. Поэтому возникла потребность в разработке нового метода автоматической генерации наименований нечетких кластеров, устранившего указанные выше недостатки и предназначенног для оценки внешних факторов при зонировании территориальных объектов по множеству разнородных индикаторов.

Метод генерации наименований кластеров

Имеется множество территориальных объектов $O = \{o_k\}$, которые необходимо распределить по группам, содержащим объекты с приблизительно одинаковым состоянием некоторого внешнего фактора. Состояние фактора оценивается по множеству индикаторов p_j – измеримых показателей, выступающих в роли признаков кластеризации. Например, для фактора, характеризующего погодные условия в некотором заданном году, индикаторами могут быть такие показатели, как «среднегодовая температура воздуха», «продолжительность отопительного периода», «среднегодовая влажность воздуха», «среднегодовая скорость ветра» и др. Каждый объект o_k характеризуется своим набором значений индикаторов. Это могут быть данные статистики или синтетические показатели, вычисленные на основе данных статистики.

Нечеткая кластеризация позволяет определить нечеткое покрытие: $\Im(O) = \{O_l | O_l \subseteq O\}, l=1, L$, которое задается матрицей $\|\mu_{O_l}(o_k)\|$ значений функций принадлежности объектов $o_k \in O$ нечетким кластерам O_l .

Кластеризацию территориальных образований предлагается выполнять с помощью алгоритма Густаффсона–Кесселя [9]. Он имеет преимущества над наиболее популярным методом нечеткой кластеризации FCM, так как позволяет находить нечеткие кластеры гиперэллипсоидной формы, более точно отражающие распределение объектов по сравнению с кластерами в форме гиперсфер, выделяемых FCM.

Наименование кластера формируется в соответствии с шаблоном, задаваемым пользователем. Шаблон включает в себя произвольный текст и совокупности базовых термов для каждого индикатора. Базовые термы позволяют описать типовые состояния внешнего фактора (той его составляющей, которая характеризуется соответствующим индикатором) на качественном уровне. Например, для фактора «климат», характеризуемого индикаторами «влажность воздуха» и «средняя температура воздуха», шаблон может быть таким: «< сухой, влажный > < холодный, умеренный, теплый > климат».

Базовые термы перечисляются в шаблоне в порядке увеличения значений индикатора: первым указывается терм, соответствующий области низких значений, последним – терм, соответствующий области высоких значений.

Чтобы сгенерировать наименование T_l кластера O_l , необходимо определить множество термов T_l^j по каждому из индикаторов p_j и подставить их в шаблон. Терм T_l^j по отдельному индикатору может либо совпадать с одним из базовых термов, либо быть составным, сгенерированным на основе базовых термов с использованием различных связок.

Обозначим множество базовых термов для индикатора p_j через $\{B_m^j\}$. Термы ранжированы в порядке возрастания свойства, измеряемого индикатором: $B_1^j < B_2^j < \dots < B_M^j$. Общее количество базовых термов M не должно быть большим. В большинстве случаев целесообразно задать $M = 3$.

Каждому базовому терму B_m^j сопоставляется зона $[a_m^j, c_m^j]$ – интервал значений индикатора. Весь диапазон значений индикатора от минимального до максимального (по всем оцениваемым объектам) разбивается на M интервалов. Причем границы интервалов могут перекрываться. Для двух соседних термов B_m^j и B_{m+1}^j , имеющих соответственно ранги m и $m+1$, границы зон должны удовлетворять следующим условиям:

$$a_m^j < a_{m+1}^j, c_m^j < c_{m+1}^j, a_{m+1}^j \leq c_m^j,$$

где a_m^j, a_{m+1}^j – нижнее границы зон для термов B_m^j и B_{m+1}^j , c_m^j, c_{m+1}^j – верхние границы зон для соответствующих термов.

На рис. 1 двумерное пространство значений двух индикаторов разбито на зоны, соответствующие трем базовым термам по индикатору p_1 и трем базовым термам по индикатору p_2 .

Для удобства можно нормировать значения индикатора и использовать универсальную шкалу действительных чисел на интервале [0; 1].

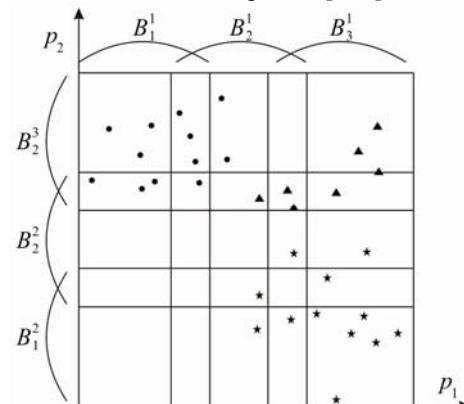


Рис. 1. Выделение зон, соответствующих базовым термам

Наименование кластера O_l по индикатору p_j определяется в зависимости от расположения его

проекции на ось индикатора – какие зоны и в какой мере охватывает данная проекция. Необходимо выделить в данном кластере подмножества $O_m^j (O_m^j \subseteq O_l)$, соответствующие каждому из базовых термов B_m^j . Значения индикатора для объектов подмножества O_m^j находятся в границах зоны, соответствующей базовому терму

$$O_m^j = \{o_k \in O_l \mid a_m^j \leq x_k^j \leq c_m^j\},$$

где x_k^j – значение индикатора p_j для объекта o_k , a_m^j , c_m^j – соответственно нижняя и верхняя границы зоны для терма B_m^j .

Для каждого базового терма определим вес его зоны по формуле

$$z_m^j = \frac{\sum_{o_k \in O_m^j} \mu_{O_l}(o_k)}{\sum_{o_k \in O_l} \mu_{O_l}(o_k)}.$$

В зависимости от соотношения величин z_m^j ($m=1, M$) определяется терм T_l^j – наименование l -го кластера по j -му индикатору. В названии фигурируют те базовые термы, веса зон которых существенно больше весов остальных термов. Обозначим подмножество таких термов через $\{B_m^j\}^*$. Для любого $B_m^j \in \{B_m^j\}^*$ вес z_m^j его зоны либо является максимальным, либо отличается от максимального на некоторую пороговую величину ϵ (например, 0,1): $z_{\max}^j - z_m^j \leq \epsilon$, где z_{\max}^j – максимальный вес зоны ($z_{\max}^j = \max_m z_m^j$).

Рассмотрим рекомендации по генерации составного терма на основе базовых.

Если веса зон всех базовых термов, кроме одного, невелики ($|\{B_m^j\}^*| = 1$), то в качестве наименования кластера используется базовый терм $B_m^j \in \{B_m^j\}^*$, вес зоны которого максимальен: $T_l^j = B_m^j$.

В ситуации, когда имеется несколько базовых термов $B_{m1}^j, B_{m2}^j, \dots \in \{B_m^j\}^*$, веса зон которых близки к максимальному весу, наименование терма формируется с помощью связки «или»:

$$T_l^j = \langle B_{m1}^j \text{ или } B_{m2}^j \text{ или ... } \rangle.$$

Если при этом веса зон всех базовых термов приблизительно равны, т.е. $|\{B_m^j\}^*| = M$, то вместо комбинации термов со связкой «или» можно использовать терм «любое». Если же в подмножество $\{B_m^j\}^*$ существенных термов попали все базовые термы, кроме одного ($|\{B_m^j\}^*| = M - 1$), то альтернативой является применение связки «не»:

$$T_l^j = \langle \text{не } B_{mi}^j \rangle, \text{ где } B_{mi}^j \notin \{B_m^j\}^*.$$

В случае, когда существенными были признаны веса зон двух соседних базовых термов B_m^j и B_{m+1}^j ,

можно выполнить более тонкий анализ, позволяющий при определенном условии заменить связку «или» в наименовании кластера на связку «между». Для этого зоны обоих термов делятся пополам: для B_m^j будут получены интервалы $[a_m^j, b_m^j]$ и $[b_m^j, c_m^j]$, для B_{m+1}^j – интервалы $[a_{m+1}^j, b_{m+1}^j]$ и $[b_{m+1}^j, c_{m+1}^j]$, где $b_m^j = a_m^j + (c_m^j - a_m^j)/2$, $b_{m+1}^j = a_{m+1}^j + (c_{m+1}^j - a_{m+1}^j)/2$.

Определяются веса полученных полузон. Если для базового терма B_m^j вес верхней полузоны $[b_m^j, c_m^j]$ существенно (больше, чем на пороговую величину ϵ) превышает вес нижней полузоны $[a_m^j, b_m^j]$, а для терма B_{m+1}^j вес нижней полузоны $[a_{m+1}^j, b_{m+1}^j]$ существенно превышает вес верхней полузоны $[b_{m+1}^j, c_{m+1}^j]$, то наименование терма формируется с помощью связки «между»:

$$T_l^j = \langle \text{между } B_{m1}^j \text{ и } B_{m+1}^j \rangle.$$

Окончательное наименование кластера определяется путем подстановки в шаблон сгенерированных термов для каждого из индикаторов.

Пример генерирования наименований кластеров

Рассмотрим в качестве примера формирование наименований кластеров субъектов РФ, выделенных по такому внешнему фактору, как «потребность в тепловой энергии», влияющему на сравнительную оценку регионов по уровню энергоэффективности в сфере теплопотребления. Это фактор, который наряду с климатическими условиями необходимо учитывать при интерпретации показателей уровня потребления тепловой энергии в регионе. Например, одно и то же значение удельного потребления тепловой энергии для субъекта РФ, в котором доля централизованного теплоснабжения невелика и в структуре отраслей преобладает сельское хозяйство, может быть оценено как «высокий уровень», а для субъекта, в котором высока доля централизованного теплоснабжения и в структуре отраслей преобладает обрабатывающая промышленность, – как «средний уровень» или даже «низкий уровень». Типологическая группировка территорий позволяет учесть различия во внешних условиях, т.к. сравнение регионов по уровню энергоэффективности выполняется в рамках групп с похожими условиями.

На роль признаков кластеризации были выбраны два индикатора: p_1 – «удельный вес общей площади жилья, оборудованной отоплением» и p_2 – «нормативная теплоэнергомкость». Индикатор p_2 рассчитывается как сумма по всем отраслям произведенений нормативного потребления тепловой энергии в отрасли на долю соответствующей отрасли в экономике субъекта РФ. В качестве норматива использовалось среднее по России потребление тепловой энергии в отрасли на 1 тыс. руб. валовой добавленной стоимости, полученной в данной отрасли. Источником данных явились сайты Федеральной службы государственной статистики [18, 19]. Значения индикаторов нормировались.

Кластеризация данных выполнялась с помощью алгоритма Густаффсона–Кесселя. Было выделено четыре нечетких кластера. Результаты кластеризации представлены на рис. 2.

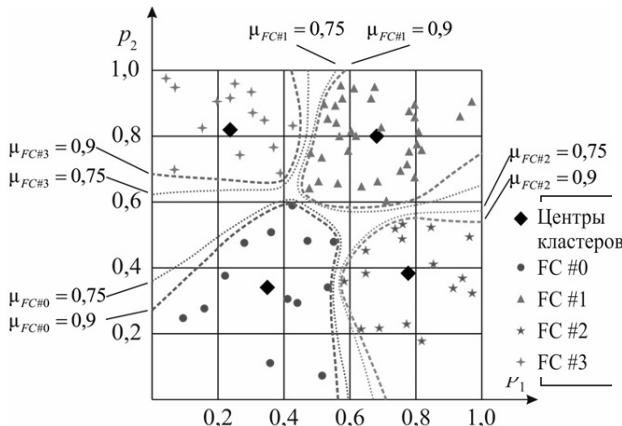


Рис. 2. Кластеры субъектов РФ по индикаторам «Удельный вес общей площади жилья, оборудованной отоплением» и «Нормативная теплоэнергоемкость»

Для наглядности на рисунке каждый субъект РФ отнесен лишь к одному кластеру – тому, степень принадлежности к которому максимальна.

Формирование наименований кластеров по фактору «потребность в тепловой энергии» осуществлялось на основе шаблона: «<низкая, средняя, высокая> в производственной сфере и <низкая, средняя, высокая> в жилищной сфере». Первая оценка в шаблоне определяется по индикатору p_1 , вторая – по индикатору p_2 . Таким образом, в качестве базовых для обоих индикаторов использовались термы: B_1^j = «низкая», B_2^j = «средняя», B_3^j = «высокая».

Границы зон базовых термов для обоих индикаторов были определены следующим образом: B_1^j = [0; 0,4], B_2^j = [0,3; 0,7], B_3^j = [0,6; 1]. Для зоны каждого базового терма обоих индикаторов были вычислены веса. Результаты приведены в таблице.

Веса зон базовых термов

Кластеры	Веса зон термов индикатора p_1			Веса зон термов индикатора p_2		
	B_1^1	B_2^1	B_3^1	B_1^2	B_2^2	B_3^2
Кластер FC#0	0,58	0,68	0,05	0,53	0,68	0,07
Кластер FC#1	0,03	0,58	0,56	0,01	0,23	0,93
Кластер FC#2	0,02	0,36	0,86	0,37	0,74	0,07
Кластер FC#3	0,82	0,35	0,01	0,02	0,12	0,93

Для кластера FC#0 в подмножество $\{B_m^1\}^*$ базовых термов, веса зон которых по индикатору p_1 были определены как существенные, попали термы B_1^1 («низкая») и B_2^1 («средняя») при условии, что $\varepsilon = 0,1$. Соответственно был сгенерирован терм T_0^1 = «низкая или средняя», который может быть заменен на терм «невысокая». Анализ весов полузон термов показал, что связка «между» в данном случае не может использоваться. Наименованием кластера по индикатору p_2 стал терм T_0^2 = «средняя», т.к. в

подмножество $\{B_m^2\}^*$ существенных базовых термов вошел только один терм – B_2^2 . В результате подстановки в шаблон было сгенерировано следующее название кластера FC#0, являющееся оценкой состояния фактора «потребность в тепловой энергии» для соответствующей группы субъектов РФ: «невысокая в производственной сфере и средняя в жилищной сфере».

Таким же образом были определены названия остальных кластеров: для FC#1 – «ненизкая в производственной сфере и высокая в жилищной сфере», для FC#2 – «высокая в производственной сфере и средняя в жилищной сфере», для FC#3 – «низкая в производственной сфере и высокая в жилищной сфере».

Заключение

Сравнительный анализ территорий, имеющих существенные различия во внешних условиях, требует дифференцированного подхода, заключающегося в выполнении сравнения в рамках групп территориальных объектов с приблизительно похожими условиями. Использование методов нечеткой кластеризации позволяет выделять группы объектов на основании значений множества разнообразных признаков без использования субъективных суждений экспертов. Для решения возникающей при этом проблемы подбора наименований кластеров, отражающих семантику результатов зонирования территорий, и был предложен описанный в данной работе метод. С его помощью наименования на естественном языке формируются автоматически путем комбинирования базовых термов и различных связок. Наименования подбираются в зависимости от расположения кластеров в пространстве значений индикаторов, причем с учетом «размытости» границ кластеров. Метод применим для случаев, когда в качестве признаков кластеризации выступают разнородные индикаторы, для семантической интерпретации значений которых используются разные наборы базовых термов.

Литература

- Булетова Н.Е. Содержание оценки результатов социально-экономического развития регионов России с использованием методов системной диагностики / Н.Е. Булетова, О.Г. Евстифеева, Е.В. Кособокова // Экономика и предпринимательство. – 2017. – № 1. – С. 996–1004.
- Крыжановская О.А. Структурно-динамические индикаторы эффективного управления развитием территорий: европейский опыт // Теория и практика сервиса. – 2017. – № 2. – С. 28–33.
- Орехова Е.А. Пространственный анализ и оценка угроз безопасности социально-экономического развития малых и средних городов / Е.А. Орехова, А.В. Плякин // Вестник Волгоград. гос. ун-та. Сер. 3: Экономика, экология. – 2016. – № 37(4). – С. 65–81.
- Development trends of the Russian regions / S.D. Valenov, A.R. Bakhtizin, Ye.V. Bukhvald, A.V. Kolchugina. // Экономика региона. – 2014. – № 39 (3). – С. 9–22.
- Myzin A.L. The modeling of national wealth of the Russia's regions / A.L. Myzin, A.I. Tatarkin // Экономика региона. – 2013. – № 4. – С. 53–65.

6. Криворотов В.В. Научно-методические основы исследования экономической безопасности в системе устойчиво-безопасного социально-экономического развития территорий / В.В. Криворотов, А.В. Калина, С.Е. Ерьпалов // Вестник Урал. фед. ун-та. – 2014. – № 5. – С. 121–132.
7. Иванченко О.Г. Методические положения зонирования территории регионов по признакам энергосбережения / О.Г. Иванченко, Л.А. Голованова // Вестник ТОГУ. – 2008. – № 2 (9). – С. 57–68.
8. Fuzzy cluster analysis: method for classification, data analysis, and image recognition / F. Hoepfner, F. Klawonn, R. Kruse, T. Runkler. – N.Y.: John Wiley & Sons, Inc., 1999. – 300 p.
9. Silich V.A. An Approach to Speed-up the Density-based Clustering via Gustafsson-Kessel Fuzzy Algorithm / V.A. Silich, M.P. Silich, S.V. Axyonov // Proceedings of 6th International Conference on Computer and Electrical Engineering (ICCEE 2013). – Paris: ETP Press, 2013. – PP. 167–172.
10. Sevilla-Villanueva B. A methodology to discover and understand complex patterns: Interpreted Integrative Multiview Clustering (I2MC) / B. Sevilla-Villanueva, K. Gibert, M. Sanchez-Marre // Pattern Recognition Letters. Pattern Recognition Techniques in Data Mining. – 2017. – Vol. 93. – P. 85–94.
11. Feyereisl J. STORM – A Novel Information Fusion and Cluster Interpretation Technique / J. Feyereisl, U. Aickelin // Proceedings of the 10th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 09), Lecture Notes in Computer Science 5788, Burgos, Spain. – 2009. – P. 208–218.
12. Cornforth D.J. Cluster evaluation, description, and interpretation for serious games: player profiling in Minecraft. / D.J. Cornforth, M.T.P. Adam // Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement. – Springer, 2015. – P. 135–155.
13. De Carvalho F. Fuzzy clustering of interval-valued data with City-Block and Hausdorff distances / F. De Carvalho, E. Simones // Neurocomputing. – 2017. – Vol. 266. – P. 259–273.
14. Борсук С.П. Свойства модификаторов составных термов лингвистических переменных // Електроніка та системи управління. – 2012. – № 3. – С. 152–157.
15. Румбешт В.В. Расширение аппарата теории нечетких множеств и лингвистической переменной для представления экспертных знаний // Научные ведомости Белгородского государственного университета. Сер.: история. Политология. Экономика. Информатика. – 2012. – Т. 24, № 19-1. – С. 141–148.
16. Ефремов А.А. Новые операции над нечеткими числами и интервалами // Доклады ТУСУР. – 2013. – № 1(27). – С. 95–99.
17. Метод генерирования наименований кластеров территориальных образований при оценке уровня их развития / М.П. Силич, Е.И. Грибков, В.А. Силич, С.В. Аксенов // Автоматизация и современные технологии. – 2014. – № 11. – С. 34–40.
18. Единая межведомственная информационно-статистическая система [Электронный ресурс]. – URL: <http://www.fedstat.ru/indicator/data.do>, свободный (дата обращения: 25.05.2017).
19. Официальный интернет-портал Федеральной службы государственной статистики [Электронный ресурс]. – URL: <http://www.gks.ru/>, свободный (дата обращения: 25.05.2017).

Силич Мария Петровна

Д-р техн. наук, профессор каф. автоматизации обработки информации (АОИ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР) Ленина пр-т, д. 40, г. Томск, Россия, 634050
Тел.: +7 (382-2) 70-15-91
Эл. почта: mary.silich@yandex.ru

Аксенов Сергей Владимирович

Канд. техн. наук, доцент инженерной школы информационных технологий и робототехники (ИШИТР) Национального исследовательского Томского политехнического ун-та (НИ ТПУ) Ленина пр-т, д. 30, г. Томск, Россия, 634050
Тел.: +7 (382-2) 60-61-30
Эл. почта: axyonov@tpu.ru

Silich M.P., Axyonov S.V.

Generation of names of fuzzy clusters in the space of heterogeneous features to zone territories by external conditions

The article proposes a method to generate names in a natural language for clusters of territories with approximately similar external conditions. The clusters are formed by fuzzy clustering of heterogeneous indicators representing external factors. The method divides the indicator value space into regions that are mapped to the basic terms defined for each indicator. The name of a particular cluster is generated by heuristic rules depending on its location in the indicator value space by combining the basic terms and connectors.

Keywords: fuzzy clustering, names of the clusters, generation of terms, zoning of territories.

doi: 10.21293/1818-0442-2018-21-4-81-86

References

1. Buletova N.E., Evtseva O.G., Kosobokova E.V. Content assessment of the results of socio-economic development of the regions of Russia using the methods of system diagnostics. *Economy and entrepreneurship*, 2017, no. 1, pp. 996–1004 (in Russ.).
2. Kryzhanovskaya O.A. Structural and dynamic indicators of effective territorial development management: European experience. *Theory and practice of service*, 2017, no. 2, pp. 28–33 (in Russ.).
3. Orekhova E.A., Plyakin A.V. Spatial analysis and assessment of threats to the security of socio-economic development of small and medium-sized cities. *Volgograd State University Journal. Third Series: Economics, Ecology*, 2016, vol. 37, no. 4, pp. 65–81 (in Russ.).
4. Valentyey S.D., Bakhtizin A.R., Bukhvald Ye.V., Kolchugina A.V. Development trends of the Russian regions. *Regional Economics*, 2014, vol. 39, no. 3, pp. 9–22.
5. Myzin A.L., Tatarkin A.I. The modeling of national wealth of the Russia's regions. *Regional Economics*, 2013, no. 4, pp. 53–65.
6. Krivorotov V.V., Kalina A.V., Erypalov S.E. Research and methodological foundations of the study of economic security in the system of sustainable and safe socio-economic development of territories. *Ural Federal University Journal*, 2014, no. № 5, pp. 121–132 (in Russ.).
7. Ivanchenko O.G., Golovanova L.A. Methodical provisions of regional zoning on the basis of energy saving. *TOGU Journal*, 2008, vol. 2, no. 9, pp. 57–68 (in Russ.).

8. Hoeppner F., Klawonn F., Kruse R., Runkler T. *Fuzzy cluster analysis: method for classification, data analysis, and image recognition*. NY, John Wiley & Sons, Inc., 1999, 300 p.
9. Silich V.A., Silich M.P., Axyonov S.V. An Approach to Speed-up the Density-based Clustering via Gustafsson-Kessel Fuzzy Algorithm. *Proceedings of 6th International Conference on Computer and Electrical Engineering (ICCEE 2013)*, Paris. ETP Press, 2013, pp. 167–172.
10. Sevilla-Villanueva B., Gibert K., Sanchez-Marre M. A methodology to discover and understand complex patterns: Interpreted Integrative Multiview Clustering (I2MC). *Pattern Recognition Letters. Pattern Recognition Techniques in Data Mining*, 2017, Vol. 93, pp. 85–94.
11. Feyereisl J., Aickelin U. STORM – A Novel Information Fusion and Cluster Interpretation Technique. *Proceedings of the 10th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 09), Lecture Notes in Computer Science* 5788, Burgos, Spain, 2009, pp. 208–218.
12. Cornforth D.J., Adam M.T.P. Cluster evaluation, description, and interpretation for serious games: player profiling in Minecraft. In book: *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, Springer, 2015, pp. 135–155.
13. De Carvalho F., Simões E. Fuzzy clustering of interval-valued data with City-Block and Hausdorff distances. *Neurocomputing*, 2017, Vol. 266, pp. 259–273.
14. Borsuk S.P. Properties of modifiers of compound linguistic variables terms. *Electronics and control systems*, 2012, no. 3, pp. 152–157 (in Russ.).
15. Rumbesht V.V. Expansion of the apparatus of the theory of fuzzy sets and linguistic variable to represent expert knowledge. *Research letters of Belgorod State University. Series: History, Political Science, Economics, and Computer Science*, 2012, Vol. 24, no. 19-1, pp. 141–148 (in Russ.).
16. Efremov A.A. New operations on fuzzy numbers and intervals. *Proceedings of TUSUR University*, 2013, vol. 1, no. 27, pp. 95–99. (in Russ.).
17. Silich M.P., Gribkov E.I., Silich V.A., Axyonov S.V. Method of generating names of clusters of territories in assessing their development levels. *Automation and modern technologies*, 2014, no. 11, pp. 34–40. (in Russ.).
18. *Edinaya mezhvedomstvennaya informacionno-statisticheskaya sistema* [Unified interdepartmental information and statistical system] Available at: <http://www.fedstat.ru/indicator/data.do>. (accessed May 25, 2017) (in Russ.).
19. *Oficial'nyj internet-portal Federal'noj sluzhby gosudarstvennoj statistiki* [Official Internet Portal of Federal State Statistics Service] Available at: <http://www.gks.ru/> (accessed May 25, 2017) (in Russ.).

Maria P. Silich

Doctor of Engineering, Professor,
Department of Data Processing Automation
Tomsk State University of Control Systems
and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 70-15-91
Email: mary.silich@yandex.ru

Sergey V. Axyonov

Candidate of Engineering Sciences, Assistant Professor,
School of Computer Science & Robotics,
Tomsk Polytechnic University (TPU)
30, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 60-61-30
Email: axyonov@tpu.ru