

ПОСТРОЕНИЕ РАЗМЫТОЙ ФУНКЦИИ ПРОСТРАНСТВЕННОГО ПОДОБИЯ ДЛЯ ГИС-ОРИЕНТИРОВАННЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

УДК 004.986

СКОРИК Сергей Николаевич

аспирант кафедры информационных технологий Херсонского национального технического университета.

Научные интересы: облачные системы и сервисы, методы и модели обеспечения безопасности, системы реального времени.

e-mail:skorik1992@gmail.com,

ШЕРСТЮК Владимир Григорьевич

д.т.н., профессор кафедры информационных технологий Херсонского национального технического университета.

Научные интересы: методы и модели поддержки принятия решений в реальном времени, принятие решений на основе прецедентов, мультиагентные системы, комбинированные логические системы представления знаний.

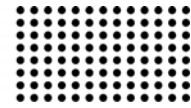
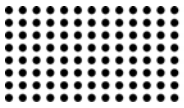
e-mail:v_sherstyuk@bigmir.net,

ВВЕДЕНИЕ

В процессе принятия управленческих решений, лицу принимающему решение (ЛПР), зачастую приходится учитывать значительное количество показателей, критериев и факторов. Принятие решения, как правило, сводится к обоснованному выбору из некоторого множества возможных альтернатив. Таким образом, принять «правильное» решение – значит выбрать такую альтернативу из числа возможных, которая с учетом заданных критериев, факторов и требований будет в максимальной степени способствовать достижению поставленной цели [1]. Для хорошо структурированных предметных областей в настоящее время уже имеется широкий выбор апробированных моделей и методов, с помощью которых может быть частично автоматизирован процесс принятия решений на различ-

ных его этапах и фазах, в том числе за счет реализации систем поддержки принятия решений (СППР) [2].

В то же время, существуют слабо структурированные предметные области, включающие множество взаимосвязанных процессов, развивающихся в пространстве и времени, для которых необходимо решать различные трудноформализуемые задачи поддержки принятия решений, например, задачи управления совместным движением множества динамических объектов, задачи локализации и предотвращения чрезвычайных ситуаций, поисково-спасательные задачи и др. Как правило, в них взаимодействующие процессы являются распределенными на некоторой территории, а для решения задач поддержки принятия решений используют геоинформационные системы (ГИС), содержащие модель тер-



ритории. Как и в любых других управленческих задачах, здесь существуют разного рода неопределенности, связанные с противоречивостью критериев, неполнотой знаний о проблеме, невозможностью количественного измерения влияния факторов-показателей. В то же время, на рассматриваемой территории наблюдается одновременное воздействие значительного числа факторов и событий, в том числе имеющих стохастическую природу, а развивающиеся при этом процессы, как правило, нелинейны, нестационарны и быстротечны. Следовательно, построение СППР для рассматриваемого класса предметных областей представляет собой сложную и нетривиальную задачу, которая существенно усложняется неточностью, неполнотой и противоречивостью исходной информации, значительной территориальной распределенностью событий и дефицитом времени [3]. Для эффективного решения подобных задач целесообразно использовать ГИС-системы в совокупности с программными модулями, выполняющими многовариантные сценарные и целевые расчеты на основе оптимизационных, имитационных либо интеллектуальных моделей с учетом субъективных предпочтений пользователя. В то же время, вопросы разработки интеллектуальных СППР на основе геолокационных и геоинформационных технологий в настоящее время проработаны недостаточно, таким образом, исследование моделей и методов поддержки принятия решений на основе ГИС для территориальных систем с множественными взаимодействующими быстротечными пространственно-распределенными процессами, является актуальной научно-технической задачей.

ПОСТАНОВКА ЗАДАЧИ

При решении трудноформализуемых задач в слабоструктурированных динамических пространственно-распределенных предметных областях зачастую невозможно построить математические модели взаимодействующих процессов с необходимыми полнотой и адекватностью, а имитационные модели рассматриваемых стохастических процессов не имеют приемлемого уровня достоверности вследствие отсутствия репрезентативной выборки. Следовательно, для построения СППР требуется использовать интеллектуальные модели и методы.

Поскольку дефицит времени на принятие решений не позволяет использовать традиционные, основанные на переборе, интеллектуальные модели, а значительная неопределенность и противоречивость исходных данных, факторов и критериев ограничивает применение классических интеллектуальных методов, во многих случаях в основу построения СППР может быть положен интуитивно-эвристический подход, не требующий наличия формальных моделей и методов. Так, при наличии достаточного уровня компетентности ретроспективной выборки может быть использован сценарно-прецедентный подход, позволяющий скомбинировать эвристический способ поиска опорного решения для текущей проблемной ситуации с многовариантной его адаптацией на основе множества уместных сценариев [4].

С другой стороны, накопление достаточного уровня компетентности требует учета чрезвычайно больших объемов данных, обработка которых традиционными способами практически невыполнима, что создает проблему информаци-

онной сложности. В современных условиях решение данной проблемы может быть основано на использовании облачных вычислений, где вычислительные возможности могут быть быстро и гибко масштабированы под решаемые задачи. В этом случае сточки зрения СППР доступные вычислительные возможности выглядят ничем неограниченными и могут быть расширены в любом количестве и в любое время.

Таким образом, для разработки СППР, предназначенных для решения трудноформализуемых задач в слабоструктурированных динамических пространственно-распределенных предметных областях, необходимо использовать: а) геоинформационные системы, с помощью которых решается вопрос привязки объектов и процессов предметной области к конкретным участкам пространства; б) сценарно-прецедентные интеллектуальные системы, позволяющие решать трудноформализуемые задачи управления на основе накопленной компетентности (опыта) и логики "здорового смысла"; в) облачные технологии, с помощью которых существенно снижается информационная сложность решения задачи накопления компетентности в виде прецедентов. Чтобы связать воедино сценарно-прецедентную и геоинформационную системы, обеспечив возможность поиска опорных решений в привязке к пространственному положению объектов, характеризующему сложившуюся проблемную ситуацию, необходимо как минимум задать в сценарно-прецедентной системе функцию пространственного подобия. С учетом того, что геолокационная информация наблюдений об объектах характеризуется неполнотой и неточностью, функция пространствен-

ного подобия должна носить размытый характер.

Цель данной статьи состоит в построении функции пространственного подобия, использующей предоставляемые ГИС пространственные метрики, которая на основе находящихся в облачной системе неточных и неполных геолокационных данных множества наблюдаемых объектов будет искать схожие конфигурации с точки зрения их географической близости.

АНАЛИЗ ПОСЛЕДНИХ ИССЛЕДОВАНИЙ И ПУБЛИКАЦИЙ

При наличии некоего облачного хранилища прецедентов \mathcal{M} требуется решить проблему выбора подходящего прецедента $e_j \in \mathcal{M}$, $j \in [1, m]$, уместного для принятия решения в текущей (проблемной) ситуации s_i . Логично искать уместный прецедент в той области пространства, где находятся решения сходных (близких) проблем, исходя из некоего отношения подобия.

Для нахождения степени близости некоей ситуации $s_j \in e_j$ к ситуации s_i и, соответственно, оценки близости решения r_j к искомому, на основе выбранного отношения подобия строится функция подобия ζ между прецедентами и выводится соответствующая ей мера подобия SIM .

Поиск близких относительно заданной функции подобия ζ описаний ситуаций s_j основан на сравнении значений их параметров (числовых, логических, строковых и т.д.). Существует ряд известных методов и различных их модификаций.

Метод ближайшего соседа (*NN*, *NearestNeighbor*) [5] – наиболее часто используемый метод нахождения подобных

прецедентов. Различные модификации метода ближайшего соседа широко применяются для решения многих задач классификации, кластеризации, регрессии и распознавания образов, где решение выбирается на основе нескольких ближайших точек (соседей), а не одной (метод k ближайших соседей, k -NN) [6].

Метод NN позволяет вычислить степень сходства текущей проблемной ситуации s_j и прецедентов e_j из \mathcal{M} , основываясь на покоординатном измерении степени совпадения значений параметров $x_1, \dots, x_m \in s_j$, определяющих прецедент e_j . Наиболее часто используются количественные метрики (Евклидова, Канберрова, Манхэттенская и др.), иногда также номинальные (Хэмминга) и смешанные (Журавлева). Эти и другие (Жаккара, Гауэра, Воронина, Миркина и т.д.) метрики подробно рассмотрены в [7, 8].

Основными преимуществами метода ближайшего соседа являются его простота реализации и универсальность в смысле независимости от специфики конкретной ПО. К существенным недостаткам метода можно отнести сложность выбора метрики для определения степени сходства и прямую зависимость требуемых вычислительных ресурсов от размера ХП, а также неэффективность при работе с неполными и плохо определенными (так называемыми «зашумленными») исходными данными [9]. В то же время, метод NN (k -NN) может быть реализован весьма эффективно [10].

Использование метода ближайшего соседа для построения функции пространственного подобию в СППР рассматриваемого класса требует учета неполноты и неточности геолокационных измерений, производимых с помощью ГИС. Одним из возможных решений мо-

жет являться использование технологии интервального «размывания» [11], позволяющей создать окрестности различной геометрической формы с размытыми границами для каждой рассматриваемой точки пространства.

Рассмотрим способ построения размытой функции пространственного подобию подробнее.

ОСНОВНОЙ МАТЕРИАЛ

Для определения степени сходства на множестве входных параметров $\{x_1, \dots, x_m\}$, используемых при описании прецедентов и проблемной ситуации, можно построить m -мерное пространство $M = X_1 \times \dots \times X_i \times \dots \times X_m$. Затем можно ввести метрику d на пространстве M , образовав метрическое пространство $M = (M, d)$.

Теперь на пространстве M в соответствии с выбранной метрикой d можно определять расстояние D от целевой точки Z_I , описываемой в пространстве M координатами $Z_I = (x_{1I}, \dots, x_{iI}, \dots, x_{mI})$, соответствующей проблемной ситуации s_j , до точек $\{Z_1, \dots, Z_q\} \in M$, представляющих прецеденты $\{e_1, \dots, e_q\} \in \mathcal{M}$, и выбирать ближайшие из них (исходя из $J = \arg \min_D (D(X_I, X_i))$) к целевой точке X_I .

Таким образом, в рамках метрики d могут быть определены наиболее близкие к целевой среди всех точек пространства M , описывающих прецеденты ХП \mathcal{M} .

На рис. 1 показана целевая точка I и соседние точки a, b, c на трехмерном пространстве $M = X_1 \times X_2 \times X_3$.

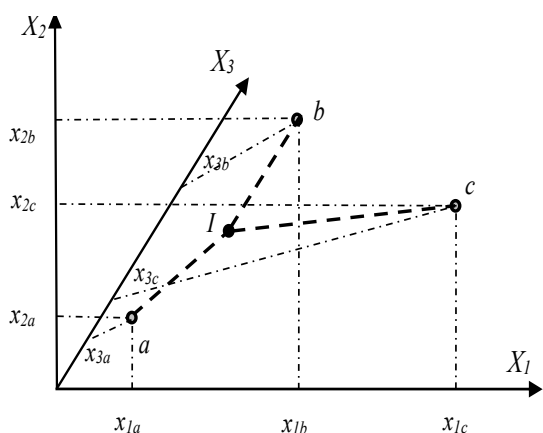


Рисунок 1 – Поиск ближайших соседних точек в трехмерном пространстве $M = X_1 \times X_2 \times X_3$

Простейшим способом найти ближайшего соседа из $\{a, b, c\}$ является прямое измерение расстояний:

$$\begin{aligned} D_{al} &= \sqrt{|x_{1l} - x_{1a}|^2 + |x_{2l} - x_{2a}|^2 + |x_{3l} - x_{3a}|^2}, \\ D_{bl} &= \sqrt{|x_{1l} - x_{1b}|^2 + |x_{2l} - x_{2b}|^2 + |x_{3l} - x_{3b}|^2}, \\ D_{cl} &= \sqrt{|x_{1l} - x_{1c}|^2 + |x_{2l} - x_{2c}|^2 + |x_{3l} - x_{3c}|^2}, \end{aligned} \quad (1)$$

и выбор минимального из них

$$D_{\min} = \min \{D_{al}, D_{bl}, D_{cl}\}. \quad (2)$$

Поскольку $D_{al} < D_{bl} < D_{cl}$, ближайшим соседом целевой точки I на рис. 1 является точка a .

Если точка I с координатами $Z_l = (x_{1l}, \dots, x_{il}, \dots, x_{ml})$ в заданном пространстве M является отображением описания проблемной ситуации s_j , и ее положение по координате x_{il} определяется значением параметра $x_i \in \{x_1, \dots, x_m\}$, представленный метод можно использовать для поиска ближайших в пространстве M прецедентов.

Метод NN позволяет найти один – самый близкий по критерию J к проблем-

ной ситуации s_j – прецедент e_j . Метод k - NN позволяет найти множество ближайших к s_j прецедентов $\{e_{j1}, \dots, e_{jk}\}$ мощностью k , вследствие этого он является более устойчивым, поскольку позволяет «сгладить» отдельные выбросы и случайный шум, всегда присутствующие в хранилище прецедентов.

Для определения значения степени близости $SIM^i(Z_j, Z_l)$ по параметру x_i необходимо в выбранной метрике d определить максимальное расстояние d_{\max}^i , используя границы области допустимых значений параметра x_i $X_{i\min}$ и $X_{i\max}$:

$$d_{\max}^i = d(X_{i\max} - X_{i\min}). \quad (3)$$

Тогда можно вычислить

$$SIM^i(Z_j, Z_l) = 1 - \frac{d_{jl}^i}{d_{\max}^i}. \quad (4)$$

Для повышения результативности метода NN (k - NN) учитывают определенные зависимости между параметрами ситуаций, например, каждому параметру x_i может быть назначен вес w_i [200], учитывающий его относительную ценность при проведении покоординатного сравнения описаний ситуаций, такой что

$$\sum_{i=1}^m w_i = 1. \quad (5)$$

Тогда степень близости прецедентов e_j и e_l по всем признакам можно вычислить, используя обобщенную формулу вида:

$$SIM(e_j, e_l) = \sum_{i=1}^m w_i \cdot SIM^i(Z_j, Z_l). \quad (6)$$

Соответственно, после вычисления степеней близости к проблемной ситуации s_j все прецеденты $e_i \in \mathcal{M}$ выстраиваются в единый ранжированный список по убыванию степени близости:

$$e_s = (e_c, \dots, e_l, \dots, e_p),$$

$$SIM(s_c, s_l) > \dots > SIM(s_l, s_l) > \dots > SIM(s_p, s_l) \quad (7)$$

При работе с неполными и плохо определенными исходными данными вычисление степени близости встречает затруднения [10]. Преодолеть их можно, во-первых, «размывая» понятие близости интервалами расстояний, и во-вторых, индексируя результаты вычислений для полученных размытых областей. Будем считать, что пространственное положение точки относительно другой точки может быть описано с помощью пеленга (угла, отсчитываемого от координатных осей) и расстояния. При наблюдении в реальных условиях и угол, и расстояние могут измеряться с погрешностями, что вызывает неточность.

Рассмотрим это с точки зрения СППР, основанной на ГИС. Одной из наиболее частых задач них является поиск объектов, которые наиболее подобны запросу. В таких случаях используется поиск ближайшей точки. Для нахождения наиболее близкой географической точки необходимо сопоставлять координаты заданной позиции и соседних точек и исходя из полученной выборки отбирать наиболее близлежащую точку к текущей. Такой поиск реализуется с помощью пространственных метрик подобия.

Допустим, в некотором пространстве имеется некая точка O , относительно которой необходимо найти ближайшую точку среди точек A_1, \dots, A_n (рис. 2). Для нахождения расстояния необходимо знать угол α до точки и расстояние L . Таким образом получаем точку A_n координатами (α_n, L_n) .

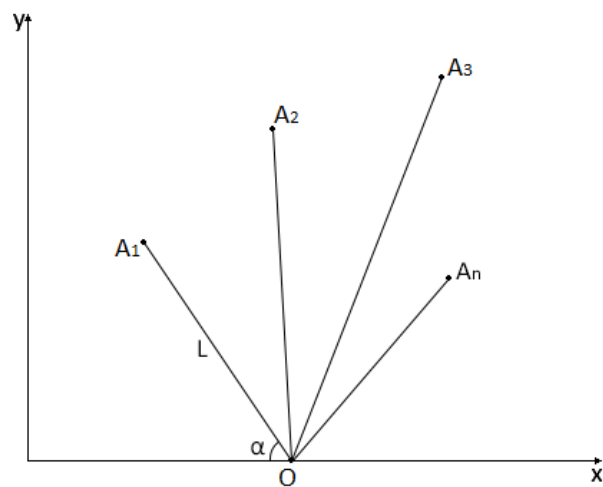


Рисунок 2—Определение точки с использованием угла и расстояния

Для более точного определения расстояния от точки O введем понятие размытия. Необходимость его использования заключается в том, что относительно угла и расстояния до точки невозможно точно узнать ближайшую координату.

Величину размытия будем определять через координаты минимального и максимального углов, а также минимальное и максимальное расстояние до некоего пятна, которое и определяет размытие-точки (рис. 3) в некоторую пространственную область, которую можно считать размытой.

$$\text{Получаем } (\alpha_{1\min}, \alpha_{1\max})(L_{1\min}, L_{1\max}).$$

Чем меньше его значение, тем точнее получается результат.

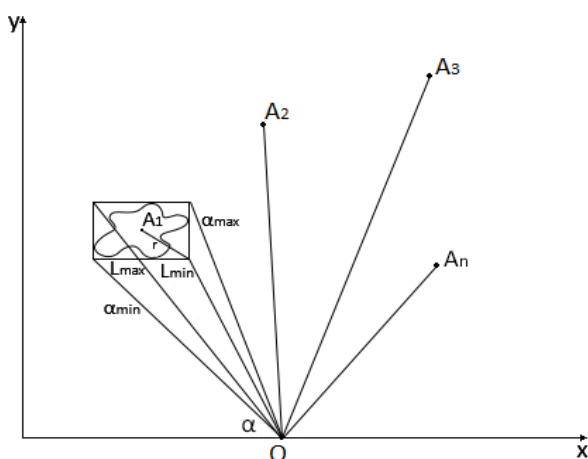


Рисунок 3– Определение величины размытия

Поисковый алгоритм для нахождения наименьшей величины размытия $\text{Search}(A, r, \max_r)$ может быть представлен следующим образом:

1. $r=0$;
2. $\text{flag} = \text{false}$;
3. initialize $lp[], rp[], oflag[]$;
4. while $\text{flag} == \text{false}$
5. $r = r + \Delta r$;
6. $\text{SearchO}(A; r)$;
7. End Search;

Для дальнейшего индексирования полученной информации могут быть использованы следующие структуры данных:

- квадродревцо (quadtree) – последовательное деление координатной плоскости на четыре части;

- R-дерево – иерархическая пространственная структура, каждый пространственный объект представлен минимальным охватывающим его прямоугольником со сторонами, параллельными координатным осям, каждый из которых в свою очередь служит дочерним не листовым;

- GRID – одноуровневая структура, сетка квадратов или прямоугольников.

В случае размытых областей наиболее эффективной структурой, позволяющей обрабатывать неточечные объекты при эффективном размещении данных во вторичной памяти, считаются R-деревья.

Эта структура данных разбивает пространство на множество иерархически вложенных и, возможно, пересекающихся прямоугольников (для двумерного пространства). В случае трехмерного или многомерного пространства это будут прямоугольные параллелепипеды (кубониды) или параллелотопы.

Для эффективности использования R-деревьев решающее значение имеет минимизация как их покрытия, так и перекрытия.

Перекрытие означает, что при запросе или вставке данных, раскрытия требует более чем одна ветка дерева (из-за того, что данные расщепляются на области, которые перекрываются).

Минимизированное покрытие улучшает эффективность продуктивности отсечения, позволяя чаще исключать из поиска целые блоки, в частности для перекрестных запросов по областям.

Существуют разные типы R-деревьев, различающиеся только способом выбора конечного и деления переполненного узла.

В нашем случае наиболее эффективной структурой данных является R*-дерево (рис. 4). Данная структура является несколько более ресурсоемкой при первоначальном ее построении, но зато дает намного лучшие результаты при поисковых запросах по сравнению с оригинальной структурой R-дерева.

Эффективность R*-дерева можно обеспечить, применяя комбинацию алгоритма расщепления узлов и алгоритма принудительной повторной вставки при переполнении узлов.

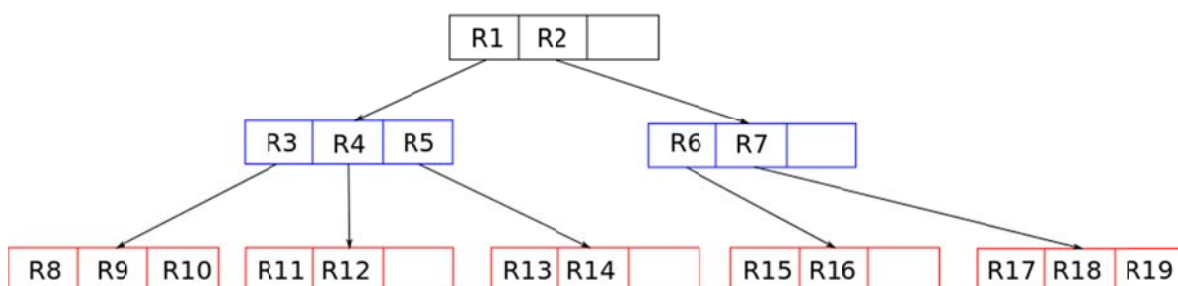
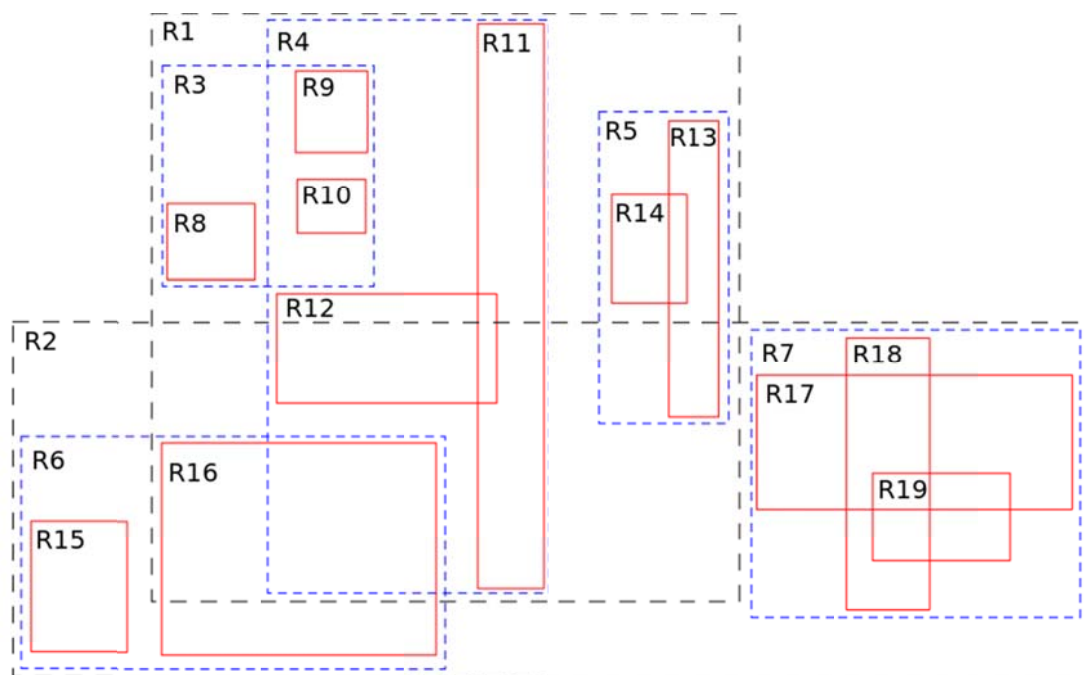


Рисунок 4- Пример R-дерева

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

В работе рассмотрены вопросы построения размытой пространственной функции подобия, основанной на достаточно эффективном методе ближайшего соседа (k -NN). Предложенный способ «размывания» точек пространства введением интервальных значений углов (пеленгов) и расстояний (дистанций) позволяет создать вокруг каждой рассматриваемой точки пространства окрестность с размытыми границами, что позволяет учесть неточность геолокационных данных об объектах, поступающих от

геоинформационной системы. Для индексирования построенных таким образом размытых окрестностей точек пространства используются структуры R*-деревьев, что позволяет повысить эффективность обработки значительных объемов геолокационных данных. Размещение структур R*-деревьев в облачном хранилище благодаря оперативному масштабированию вычислительной мощности позволяет значительно снизить вычислительную сложность определения пространственного подобия в СППР, использующих сценарно-

прецедентные интеллектуальные системы в качестве источников информации и использующих геоинформационные системы.

ЛИТЕРАТУРА

1. Burstein, F. Handbook on Decision Support Systems / F. Burstein; C. W. Holsapple. – Berlin: Springer Verlag, 2008. – 846 p.
2. Kovalenko, I. Nekotorye principy postroeniya sistemy podderzhki prinjatija reshenij v proektah rekonstrukcii municipal'nyh sistem vodosnabzhenija / I. I. Kovalenko, V. K. Koshkin // Upravlinnija rozvitkom skladnih sistem. – 2014. – Vip. 19. – S. 39-43.
3. Singaevskij, P. Proektirovanie slozhnyh sistem upravlenija dinamičeskimi ob'ektami na vodnom transporte / P.S. Singaevskij, A.S. Judin // Materialy Mezhdunarodnoj konferencii po mjagkim vychislenijam i izmerenijam. – S.-Pb., 2015, T. 1, S. 304-311.
4. Sherstjuk, V. Osnovy teorii dinamičeskikh scenarno-precedentnyh intelektual'nyh sistem / V.G. Sherstjuk. – Herson: Feniks, 2012. – 476 s.
5. Aha, D. Learning to Catch: Applying Nearest Neighbor Algorithms to Dynamic Control Tasks / D.W. Aha, S.L. Salzberg // Selecting Models from Data: Artificial Intelligence and Statistics. – N.Y.: Springer-Verlag, 1993. – Pp.363-368.
6. Aha, D. Lazy learning / D.W. Aha. – Norwell: Kluwer Academic Publishers. – 1997. – 432 p.
7. Finnie, G. Similarity and metrics in case-based reasoning / G. Finnie, Z. Sun // International Journal of Intelligent Systems. – 2002. – Vol. 17. – №3. – Pp.273-287.
8. Principal Manifolds for Data Visualization and Dimension Reduction / A. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.). – Berlin-Heidelberg-N.Y.: Springer, 2007. – 334 p.
9. Aha, D. Feature Weighting for Lazy Learning Algorithms / D.W. Aha // Feature Extraction, Construction and Selection. – The Springer International Series in Engineering and Computer Science. – 1998. – Vol.453. – Pp.13-32.
10. Ling, C. Computing Optimal Attribute Weight Settings for Nearest Neighbor Algorithms / C.X. Ling, H. Wang // Artificial Intelligence Review. – 1997. – Vol.11. – №1-5. – Pp. 255-272.
11. Remezova, E. Nechetkie mnozhestva vtorogo porjadka: ponjatje, analiz i osobennosti primenenija / E.M. Remezova // Sovremennye problemy nauki i obrazovanija. – 2013. – №5, URL: <http://www.science-education.ru/ru/article/view?id=10506>.

Рецензент: д.т.н., проф. Рудакова А.В.,
Херсонский национальный технический университет.