

Секція 11

Автоматизація та інформаційні технології

Алексеев М.А., д.т.н., профессор, Сироткина Е.И., аспирант

(Государственный ВУЗ "Национальный горный университет", г. Днепропетровск, Украина)

МЕТОДИКА ДИАГНОСТИКИ РАБОТЫ АППАРАТНО-ПРОГРАММНОГО КОМПЛЕКСА SCADA СИСТЕМ

В настоящее время во всем мире широкое распространение получают информационно-управляющие SCADA (Supervisory Control And Data Acquisition) системы [1, 2]. Данный класс систем применяется в энергетике, военной области, промышленности, в том числе горной области. Перечисленные области применения таких систем обуславливают повышенные требования к их надежности, живучести, отказоустойчивости и безопасности работы в процессе эксплуатации. Таким образом, актуальной является задача автоматической самодиагностики их работы в режиме реального времени с возможностью самовосстановления работоспособности после обратимых отказов.

SCADA система представляет собой распределенный многоуровневый и многозадачный аппаратно-программный комплекс (АПК), работающий в режиме реального времени, и является сложным и трудноформализуемым объектом диагностики с изменяющейся структурой и функциональностью в процессе «жизненного цикла». Структура и закономерности функционирования SCADA систем определяются совокупностью декларативных и процедуральных знаний, комплексное использование которых характерно для экспертных систем.

Для решения задачи автоматической самодиагностики работы АПК SCADA системы предлагается введение отдельной подсистемы диагностики, работающей в режиме реального времени на основе методологии экспертных систем [3].

Схема взаимодействия SCADA и экспертной системы приведена на рисунке 1.

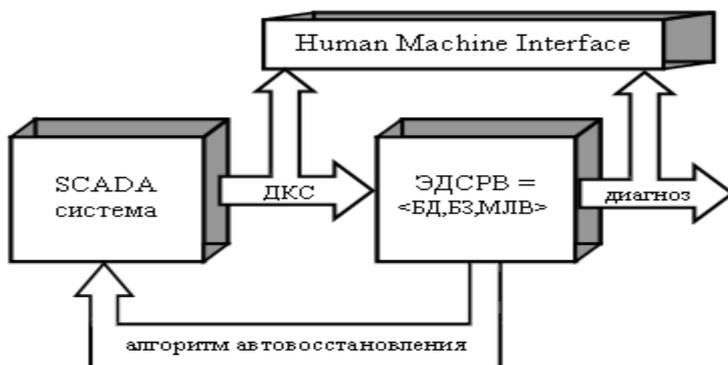


Рисунок 1 – Схема взаимодействия SCADA и ЭДСРВ

где ЭДСРВ – экспертная диагностическая система реального времени; БД – база данных, содержащая базовые компоненты (термины и факты предметной области, их свойства и связи между ними); БЗ – база знаний, содержащая правила конструирования вариантов структур из компонентов БД; МЛВ – машина логического вывода, устанавливающая последовательность применения правил из БЗ; ДКС – диагностические коды SCADA системы (коды ошибок ПО, исключения классов и функций, коды возвратов процессов, состояния неопределенности, т.е. частичное или полное отсутствие информации по истечении таймаута ожидания и т.д.), генерируемые в процессе ее работы; Human Machine Interface (HMI) – интерфейс "человек - машина".

Для анализа схемы взаимодействия SCADA системы и ЭДСРВ была применена методика объектно – классификационного моделирования (ОКМ) [4].

Диаграмма объектов предметной области представлена на рисунке 2.

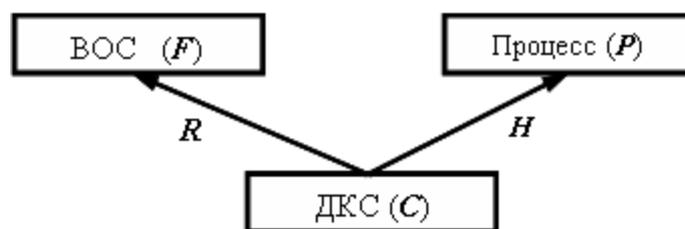


Рисунок 2 – Диаграмма объектов предметной области экспертной системы

где

ВОС – виды отказов системы, представленные множеством F ;

процесс – исполняемый программный модуль системы, работающий на системообразующем узле. Множество процессов P составляют программное обеспечение SCADA системы;

ДКС – диагностические коды системы, представленные множеством C ;

R – отношение соответствия диагностических кодов системы видам отказов (см. формулу 1);

H – отношение принадлежности диагностических кодов процессам (см. формулу 2).

$$R \subset C \times F = \{ (c, f) \mid c \in C, f \in F \} \quad (1)$$

$$H \subset C \times P = \{ (c, p) \mid c \in C, p \in P \} \quad (2)$$

В процессе постановки диагноза экспертная система выполняет следующие подзадачи:

- определение подмножества возможных видов отказов на основании текущей выборки ДКС, поступающей на вход экспертной системы при работе совокупности процессов;
- выявление закономерностей при анализе импликативных связей между ДКС, включая запреты на комбинации ДКС;
- определение минимально необходимого набора ДКС, позволяющих однозначно диагностировать вид отказа.

Предложенный подход к диагностике работы АПК SCADA с применением отдельной подсистемы диагностики, реализованной на основе ЭДСРВ, является предпосылкой формирования универсальных подсистем диагностики, не привязанных к конкретной реализации SCADA системы и стыкуемых с ней по множеству ДКС. Накопление «экспертного опыта» в процессе обучения ЭС увеличит пространство решений достижимости автовосстановления для обратимых отказов SCADA систем.

Перечень ссылок

1. The Fundamentals of SCADA. Bentley Systems, Incorporated. 2004. — 18 p.
2. Ken Barnes, Briam Johnson, Reva Nickelson. Review of Supervisory Control and Data Acquisition (SCADA) Systems. Idaho National Engineering and Environmental Laboratory, 2004. – 40 p.
3. Джексон П. Введение в экспертные системы. — 3-е изд. — М.: Вильямс, 2001. — 624 с.
4. Устенко А.С. Основы математического моделирования и алгоритмизации процессов функционирования сложных систем. — М.: БИНОМ, 2000. — 250с.

Вовк А.Н., студент гр. КМ-08

Научный руководитель: Сарычева Л.В.

(ГВУЗ Национальный горный университет, г. Днепропетровск, Украина)

РАЗРАБОТКА И ИССЛЕДОВАНИЕ ПРОГНОЗНЫХ МОДЕЛЕЙ ПОКАЗАТЕЛЕЙ БАНКОВ С ИСПОЛЬЗОВАНИЕМ ДРОБНОГО ИНТЕГРО-ДИФФЕРЕНЦИРОВАНИЯ

Задача прогнозирования банковских показателей является актуальной. Для получения количественного прогноза используются разнообразные методы, в том числе классические методы прогнозирования временных рядов.

Объект исследования: показатели банков

Цель работы: построение интегро-дифференциальной модели дробной размерности для прогнозирования банковских показателей

Дробное интегро-дифференцирование определяет объединённый оператор интегрирования-дифференцирования, порядок которого – произвольное вещественное или комплексное число. Наиболее употребительно интегро-дифференцирование Римана-Лиувилля, определяемое формулами [1]:

$$I_{a,t}^{\beta} f(t) = \frac{1}{\Gamma(\beta)} \int_a^t \frac{f(\tau) d\tau}{(t-\tau)^{1-\beta}},$$

$$D_{a,t}^{\beta} f(t) = \frac{d^n}{dt^n} I_{a,t}^{\beta} f(t) = \frac{1}{\Gamma(n-\beta)} \frac{d^n}{dt^n} \int_a^t \frac{f(\tau) d\tau}{(t-\tau)^{\beta-n+1}},$$

где $I_{a,t}^{\beta}$, $D_{a,t}^{\beta}$ – интегральный и дифференциальный оператор β -го порядка соответственно; $\beta, a \in \mathbb{R}$, $n-1 < \beta < n$.

Хотя математический аппарат дробного интегро-дифференцирования довольно развит [2], его использование для создания математических моделей систем начато сравнительно недавно [3]. Дробная производная, в отличие от производной целого порядка, – это нелокальная характеристика функции, она зависит не только от поведения функции в окрестности рассматриваемой точки t , но и от принимаемых ею значений на всем интервале (a, t) . Так как банковские показатели обладают долговременной памятью, то в данной работе предлагается использовать математический аппарат дробного интегро-дифференцирования для построения прогнозной модели показателей банков.

Исходные данные представлены в виде временных рядов :

$$y_i^s = y^s(t_i), t_i = i\Delta t, i = 1, 2, \dots, N, ,$$

где Δt – интервал выборки, N – длина ряда.

Построение прогнозной модели включает: 1) постулирование общего класса моделей; 2) структурную идентификацию конкретных подклассов выбранного типа моделей; 3) оценку параметров идентифицированных моделей; 4) проверку адекватности модели.

1. Класс рассматриваемых моделей имеет вид:

$$\hat{y} = f(D_{a,t}^{\alpha_i} y, D_{b,T}^{\beta_i} Y),$$

где $y(t)$, $Y(t)$ – показатели банков в виде временных рядов, локальный и трендовый соответственно, α_i, β_i – степени производной, $i = 1, 2, \dots, m$, $\alpha_i, \beta_i \in \mathbb{R}$.

Для численного дифференцирования дробного порядка используется формула [1]:

$$D_{0,x}^{\alpha} y(t) \approx \frac{x^{-\alpha} n^{\alpha}}{\Gamma(-\alpha)} \sum_{k=0}^{n-1} \frac{\Gamma(k-\alpha)}{k!} y_k(x), x > 0,$$

где $y_n(x) = y(0)$, $y_{n-1}(x) = y(x/n)$, ..., $y_k(x) = y(x - kx/n)$, ..., $y_0(x) = y(x)$.

2. Принимается линейная структура модели со свободным членом:

$$\hat{y}(t+1) = \lambda_0 + \sum_{i=1}^m (\lambda_{i,y} D_{a,t}^{\alpha_i} y(t) + \lambda_{i,Y} D_{b,T}^{\beta_i} Y(t)).$$

В качестве метода перебора моделей используется метод группового учета аргументов [4].

3. Оценка параметров моделей производится по методу наименьших квадратов. Для оценки качества получаемых моделей применяется критерий остаточной суммы квадратов. При этом выбирается модель минимальной сложности.

4. Проверка адекватности модели проводится на реальных данных

Результаты интегро-дифференциального прогнозирования банковских показателей позволяют сделать прогноз с удовлетворительной точностью, во многих случаях лучше, чем полученный другими методами прогнозирования.

Список использованных источников

1. Васильев В.В., Симак Л.А. Дробное исчисление и аппроксимационные методы в моделировании динамических систем. – Киев. НАН Украины. – 2008. – 256 с.
2. Самко С.Г., Килбас А.А., Маричев О.И. Интегралы и производные дробного порядка. – Минск: Наука и техника. – 1987. – 688 с.
3. Нахушев А.М.. Дробное исчисление и его применение. – М.: ФИЗМАТЛИТ, 2003. – 272 с.
4. Ивахненко А.Г., Юрачковский Ю.П. Моделирование сложных систем по экспериментальным данным. – М.: Радио и связь. - 1987. – 120 с.

Войтюк Д.П., магистр гр. КМ-08-1м

Научный руководитель Сарычева Л.В., проф. каф. ГИС,

(ГВУЗ «Национальный Горный Университет», г. Днепропетровск)

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ОЦЕНКИ СТОИМОСТИ ОБЪЕКТА НЕДВИЖИМОСТИ

Стоимостный вывод эксперта-оценщика – трудоемкий процесс. Данное решение на сегодняшний день не автоматизировано, поэтому разработка и реализация программы оценки стоимости объекта недвижимости – актуальная задача.

Оценка стоимости объекта недвижимости – процесс определения его рыночной стоимости или отдельных прав (например, права аренды, права пользования и т.д.).

Существует три подхода (метода) к оценке любого вида недвижимости, используемых профессиональными оценщиками: сравнительный (или «рыночный») подход, затратный подход, доходный подход.

Сравнительный подход – совокупность методов оценки стоимости объекта оценки, основанных на сравнении объекта оценки с объектами-аналогами, в отношении которых имеется информация о ценах. Стоимость объекта недвижимости при этом подходе равна средневзвешенной цене по реализованным аналогам с учетом поправок:

$$V = P_k \times \sum_b D(b),$$

где P_k – продажная цена аналогичного объекта недвижимости;

$\sum_b D(b)$ – сумма поправочных корректировок.

Согласно затратному подходу, общая стоимость объекта недвижимости (S) определяется как сумма стоимости участка земли (C) и восстановительной стоимости (стоимости замещения или воспроизводства) объекта недвижимости (R) за минусом накопленного износа (W):

$$S = C + R - W$$

Доходный подход основан на определении стоимости недвижимости путем расчета дисконтированного потока доходов. При расчёте используются две методики: капитализации доходов и дисконтирования денежных потоков. Общая формула вычисления стоимости недвижимости:

$$C = \frac{I}{K},$$

где C – стоимость недвижимости; I – ожидаемый доход от оцениваемой недвижимости.

Цель работы – разработка автоматизированной системы для создания отчетов об оценке стоимости объекта недвижимости на основе доходного, сравнительного и затратного подходов.

Любая оценка недвижимости начинается с изучения соответствующих документов. Входными данными являются: документы права собственности на объект недвижимости; технический паспорт; информация о месторасположении; данные аналогов; паспортные данные и идентификационный код заказчика. Выходной информацией является документ (вывод об оценке) с рассчитанной стоимостью объекта оценки.

Приложение реализовано на объектно-ориентированном языке программирования C# на платформе .Net Framework. Данный продукт производит создание отчетов об оценке недвижимости исходя из данных введенных пользователем (оценщиком).

В программе реализовано три метода для оценки недвижимости (сравнительный, затратный, доходный). В зависимости от выбранного пользователем сценария, выполняется соответствующий подход.

Так как предмет деятельности является узконаправленным, сфера информационных технологий не задействовалась в оценке в полной мере. Поэтому в данном продукте произведена автоматизация процесса сбора данных аналогичных объектов недвижимости. Для этого написана специальная, со специфическими алгоритмами, библиотека для просмотра и анализа интернет-страниц.

Данные статистики и аналогов продаж актуальны согласно Национальному стандарту №1 Украины, поэтому обновление данных производится раз в пол года – еще одна функция библиотеки анализа интернет-страниц.

Произведено усовершенствование алгоритма вычисления стоимости объекта недвижимости, улучшение направлено на вычисление оптимальной цены без появления так называемого выброса, или скачка стоимости в большую, или меньшую сторону относительно средней цены по аналогичным объектам.

Итоговый отчет создается на основе шаблона (готового документа). Технология формирования отчета основана на поиске меток в документе и замене их на соответствующий текст или числовое значение. В программе реализован класс с массивом меток. Эти метки сравниваются с теми, которые в документе. Если метка найдена, производится замена на соответствующее значение.

Разработанные программы апробированы на реальных данных, составлено 15 отчетов по оценке объектов недвижимости в разных районах города Днепропетровск. Данные отражают действительность цен за квадратный метр. Это свидетельствует о корректной работоспособности приложения.

Внедрение автоматизированной системы оценки стоимости объекта недвижимости позволяет повысить эффективность работы оценщика недвижимости более чем в десять раз.

Глущенко О.С., студентка гр.ТТмм-09-1, Трубицин М.Н., к.т.н., доцент,
(Государственное ВУЗ «Национальный горный университет», г. Днепропетровск,
Украина),

Шишко С.В., ассистент

(Государственное ВУЗ «Украинский государственный химико-технологический университет», г. Днепропетровск, Украина)

РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ МЕТОДОМ ПОТЕНЦИАЛОВ С ИСПОЛЬЗОВАНИЕМ ГРАФИЧЕСКИХ ВОЗМОЖНОСТЕЙ MathCad

Продолжительность решения транспортной задачи, как задачи линейного программирования (ЗЛП), методом потенциалов, в первую очередь, зависит от количества итераций, смысл которых состоит в переходе от некоторого исходного плана перевозок к улучшенному (по значению целевой функции). Таким образом, в процессе решения необходимо построить большое количество таблиц-планов, в клетках которых одновременно отобразить тарифы, объемы перевозок, псевдостоимости и некоторые другие обозначения. Графические возможности пакета MathCad достаточно удобны для производства таких действий.

В связи с этим определим **цель** настоящей работы: разработка алгоритма решения ЗЛП с поэтапным графическим сопровождением и возможностью сравнения итерационных процессов для выбора рационального опорного плана.

Идея работы заключается в специальном формировании матриц MathCad для необходимого нанесения любого количества линий-графиков в табличном оформлении планов (итераций).

В качестве примера рассмотрим ЗЛП 3x5, исходные данные которой и результат начальной итерации приведены на рис.1.

		1,10	2,50	3,70	1,30	2,10	
		A_1	A_2	A_3	A_4	A_5	Q_i
0	B_1	0 ^{4,30}	73 ^{2,50}	0 ^{2,80}	45 ^{1,30}	55 ^{2,10}	173
1	B_2	104 ^{2,10}	49 ^{3,50}	9 ^{4,70}	0 ^{3,80}	0 ^{3,20}	153
1,40	B_3	0 ^{4,60}	0 ^{4,90}	194 ^{2,30}	0 ^{3,30}	0 ^{2,10}	194
	Q_j	104	113	203	45	55	520

Рис. 1. Пример построения первого опорного плана (метод минимального элемента) с нанесенными потенциалами (количество линий-графиков – 7, по цвету линий)

На этом рисунке план проверен на оптимальность, при возникшей необходимости от клетки с максимальным потенциалом (отмечается мигающим квадратиком-сигналом – анимация) строится цикл перемещения объемов, таким образом получен новый улуч-

шенный план, рис.2. Проверка последнего плана на оптимальность дает положительный результат при значении целевой функции $z(X)=1195$ у.е.

	A_1	A_2	A_3	A_4	A_5	Q_i
B_1	$\begin{matrix} 4,30 \\ 0 \end{matrix}$	$\begin{matrix} 2,50 \\ 64 \end{matrix}$	$\begin{matrix} 2,80 \\ 9 \end{matrix}$	$\begin{matrix} 1,30 \\ 45 \end{matrix}$	$\begin{matrix} 2,10 \\ 55 \end{matrix}$	173
B_2	$\begin{matrix} 2,10 \\ 104 \end{matrix}$	$\begin{matrix} 3,50 \\ 49 \end{matrix}$	$\begin{matrix} 4,70 \\ 0 \end{matrix}$	$\begin{matrix} 3,80 \\ 0 \end{matrix}$	$\begin{matrix} 3,20 \\ 0 \end{matrix}$	153
B_3	$\begin{matrix} 4,60 \\ 0 \end{matrix}$	$\begin{matrix} 4,90 \\ 0 \end{matrix}$	$\begin{matrix} 2,30 \\ 194 \end{matrix}$	$\begin{matrix} 3,30 \\ 0 \end{matrix}$	$\begin{matrix} 2,10 \\ 0 \end{matrix}$	194
Q_j	104	113	203	45	55	520

Рис.2. Оптимальный план перевозок – решение транспортной задачи.

Отметим, что подключение нелинейных алгоритмов встроенной MathCad-функцией Minimize, рис.3, дает решение только близкое к оптимальному (разница – 8 у.е.). Поэтому, при использовании известных программ решения ЗЛП (TORA, OPTIMAL, MERKS и др.) необходима обязательная проверка полученного экстремума на глобальность, что наглядно можно сделать при помощи разработанной программы.

<p>Матрица стоимостей перевозок</p> $Cst := \begin{pmatrix} 43 & 25 & 28 & 13 & 21 \\ 21 & 35 & 47 & 38 & 32 \\ 46 & 49 & 23 & 33 & 21 \end{pmatrix} \cdot \frac{1}{10} = \begin{pmatrix} 4,3 & 2,5 & 2,8 & 1,3 & 2,1 \\ 2,1 & 3,5 & 4,7 & 3,8 & 3,2 \\ 4,6 & 4,9 & 2,3 & 3,3 & 2,1 \end{pmatrix}$ <p>Вектор запасов поставщиков</p> $bq = \begin{pmatrix} 173 \\ 153 \\ 194 \end{pmatrix}$ <p>Строка спроса потребителей</p> $aq := (104 \ 113 \ 203 \ 45 \ 55)$ <p>$m := \text{rows}(Cst) = 3$ $n := \text{cols}(Cst) = 5$ $i = 1..m$ $j = 1..n$</p> <p>Проверка "закрытости" ЗЛП</p> $\sum_{j=1}^n (aq^T)_j = 520 \quad \sum_{i=1}^m bq_i = 520$ <p>Начальное приближение объемов перевозки</p> $x_{i,j} := \left[\frac{bq_i}{n} + \frac{(aq^T)_j}{m} \right] \cdot \frac{1}{2}$ $x = \begin{pmatrix} 34,633 & 36,133 & 51,133 & 24,8 & 26,467 \\ 32,633 & 34,133 & 49,133 & 22,8 & 24,467 \\ 36,733 & 38,233 & 53,233 & 26,9 & 28,567 \end{pmatrix}$ <p style="text-align: right;">страница 1</p>	<p>Целевая функция</p> $z(x) := \sum_{i=1}^m \sum_{j=1}^n (Cst_{i,j} \cdot x_{i,j})$ <p>Ее значение при начальном приближении</p> $z(x) = 1,682 \times 10^3$ <p>Вспомогательные единичные вектора-сомножители</p> $E_{m_1} = 1 \quad E_{n_1} = 1$ <p>Given</p> $x \cdot E_n = bq \quad E_m^T \cdot x = aq \quad x \geq 0$ <p>Блок решения</p> $X := \text{Minimize}(z, x)$ <p>поставлен нелинейный алгоритм</p> $X = \begin{pmatrix} 0 & 73 & 0 & 45 & 55 \\ 104 & 40 & 9 & 0 & 0 \\ 0 & 0 & 194 & 0 & 0 \end{pmatrix} \quad z(X) = 1,203 \times 10^3$ <p style="text-align: right;">страница 2</p>
---	---

Рис. 3. MathCad-документ решения ЗЛП встроенной подпрограммы Minimize.

ВЫВОДЫ

Проведенные расчеты показали, что достижение решения ЗЛП по опорным планам, построенным с помощью методов минимальной стоимости, северо-западного угла и двойного предпочтения достигается за наименьшее количество итераций (2) по методу минимальной стоимости.

Разработанная программа позволяет получить решение ЗЛП в наглядном виде, и может служить удобным учебным пособием, иллюстрирующим этапы и пункты получения решения транспортной задачи методом потенциалов.

Улучшение начальных опорных планов и определение их приоритетов возможно только с привлечением процесса «взвешивания» полученных решений.

Калиниченко Е.С., магистр гр. ГИС-08-1м

(ГВУЗ «Национальный горный университет», г. Днепропетровск, Украина)

ИЗУЧЕНИЕ КОЛИЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ ГОРОДСКИХ ОСТРОВОВ ТЕПЛА ПО ДАННЫМ О ПОВЕРХНОСТНОЙ ТЕМПЕРАТУРЕ TERRA/MODIS

"Городской тепловой остров" (Urban Heat Islands, UHI) - феномен изменения температуры городских территорий по сравнению с сельскими районами [1]. Воздействие UHI на окружающую природную среду определяется влиянием размещения техногенных объектов на распределение температур городских территорий, отличающееся от распределения температур прилегающих территорий. Для изучения UHI в исследованиях широкого пространственного охвата используются данные Terra/MODIS.

Сенсор MODIS спутника Terra осуществляет постоянную съемку поверхности Земли с периодом от 1 до 2 дней, регистрируя данные в 36 спектральных каналах. Для исследования теплового состояния земной поверхности используются данные каналов 8-36 MODIS в спектральном диапазоне 0,41-14,39 мкм с пространственным разрешением 1000 м.

Разновременные данные о приповерхностной температуре Terra/MODIS позволяют производить расчет количественных показателей UHI для изучения динамики изменения теплового состояния исследуемых территорий. Перечень таких показателей включает [2]:

- различие температур городских и сельских территорий;
- различие температур центральных частей городских территорий и сельских районов;
- различие температур городских и прочих территорий;
- различие температур город – вода;
- площадь UHI;
- интенсивность UHI и др.

Процесс изучения UHI включает следующие этапы:

- формирование выборки разновременных данных спутниковых наблюдений;
- определение температуры земной поверхности и формирование контуров городских тепловых островов;
- расчет характеристик UHI и оценка динамики изменения теплового состояния исследуемых территорий.

Следует учесть, что на временную динамику изменения поверхностных UHI оказывают существенное влияние топографические, климатические характеристики и местные погодные условия, что требует проведения дополнительных исследований с использованием многовременных данных климатических наблюдений.

Сравнение значений количественных характеристик UHI, рассчитанных для территорий крупных промышленных мегаполисов в различные моменты времени позволит произвести их ранжирование по степени воздействия теплового загрязнения на окружающую природную среду.

Перечень источников

1. Грищенко, М.Ю. Применение тепловых снимков системы ETM+ для изучения теплового состояния Москвы / М.Ю. Грищенко // Современные проблемы дистанционного зондирования Земли из космоса. – 2012. – Т.9. - №4. – С.95-101.
2. Schwarz, N. Exploring indicators for quantifying surface urban heat islands of European cities with MODIS land surface temperatures / N. Schwarz, S. Lautenbach, R. Seppelt // Remote Sensing of Environment. – 2011. – Vol. 115. – P. 3175-3186.

Константинова Д. В., магистр гр. КМ-08-1м,
(ГВУЗ «Национальный Горный Университет», г. Днепрпетровск)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ НЕКОТОРЫХ АЛГОРИТМОВ КЛАССИФИКАЦИИ ТЕКСТУР ГЕОИЗОБРАЖЕНИЙ

Геоизображение - пространственно-временная масштабная генерализованная модель земных (планетных) объектов или процессов, представленная в графической образной форме.

Под классификацией понимают процесс отнесения элементов изображения (пикселей) к конечному числу классов на основе значений их яркостей.

Целью работы является сравнительный анализ, заключающийся в оценке эффективности некоторых алгоритмов классификации текстур геоизображений - локального бинарного шаблона (ЛБШ), BIRCH, сети Кохонена, CLOPE и MST.

Критерии сравнения: общая точность (процент правильно классифицированных пикселей), средняя точность (среднее значение точности по всем строкам и среднее значение точности по всем столбцам), комбинированная точность (среднее арифметическое из общей точности и средней точности по строкам, среднее арифметическое из общей точности и средней точности по столбцам).

ЛБШ представляет собой описание окрестности пикселя изображения в двоичной форме. Оператор ЛБШ использует восемь пикселей окрестности, принимая центральный пиксель в качестве порога. Пиксели, имеющие значения больше, чем центральный пиксель (или равное ему), принимают значения «0». Таким образом, получается восьмиразрядный бинарный код, описывающий окрестность пикселя.

BIRCH является локальным, т. е. кластеризация выполняется по небольшим группам объектов без обработки всего набора данных. Он также использует естественную меру сходства между объектами, отражающую их близость в многомерном пространстве. Алгоритм выделяет кластеры в объекты, образующие группы, характеризующиеся наибольшей плотностью в многомерном пространстве. Удаленные от общей группы объекты, как правило, выделяются в отдельный кластер или удаляются из дальнейших расчетов. BIRCH использует в полном объеме доступную память для выявления субкластеров, что позволяет минимизировать затраты на ввод-вывод данных. Время работы линейно масштабируемое. Алгоритм обрабатывает весь набор данных по частям единоразово, без обращения к одним и тем же данным.

Назначением самоорганизующейся карты Кохонена является кластеризация многомерных векторов, разведочный анализ данных, обнаружение новых выявлений. Самоорганизующая карта Кохонена – нейронная однослойная сеть прямого распространения. При использовании карты входной вектор предъявляется на вход, после чего на выходе активизируется нейрон или группа нейронов, которые соответствуют тому или иному кластеру, полученному в процессе обучения сети. Используется универсальный аппроксиматор – нейронная сеть, обучение сети без учителя, самоорганизация сети, простота реализации, гарантированное получение ответа после прохождения данных по слоям.

Характеристика алгоритма CLOPE: высокие масштабируемость и скорость работы, а также качество кластеризации, что достигается использованием глобального критерия оптимизации на основе максимизации градиента высоты гистограммы кластеры. Во время работы алгоритм хранит в памяти небольшое количество информации по каждому кластеру и требует минимальное число сканирований набора данных. CLOPE автоматически подбирает количество кластеров, что регулируется одним единственным параметром – коэффициентом отталкивания.

Алгоритм MST, основанный на минимальных остовных деревьях, выделяет кластеры произвольной формы, в том числе кластеры выпуклой и вогнутой форм, выбирает из нескольких оптимальных решений наиболее оптимальное.

Алгоритм	Достоинства	Недостатки
ЛБШ	Экономия памяти, ЛБШ определяют только важные локальные текстуры, такие как, концы линий, края, углы, пятна	Сложная реализация
BIRCH	Двухступенчатая кластеризация, кластеризация больших объемов данных, работает на ограниченном объеме памяти, является локальным алгоритмом, может работать при одном сканировании входного набора данных, использует тот факт, что данные неодинаково распределены по пространству, и обрабатывает области с большой плотностью как единый кластер	Работа с только числовыми данными, хорошо выделяет только кластеры выпуклой или сферической формы, есть необходимость в задании пороговых значений
MST	Выделяет кластеры произвольной формы, в т. ч. Кластеры выпуклой и вогнутой форм, выбирает из нескольких оптимальных решений наиболее оптимальное	Чувствителен к выбросам
CLOPE	Высокие масштабируемость и скорость работы, а также качество кластеризации, что достигается использованием глобального критерия оптимизации на основе максимизации градиента высоты гистограммы кластера. Во время работы алгоритм хранит в памяти небольшое количество информации по каждому кластеру и требует минимальное число сканирований набора данных. CLOPE автоматически подбирает количество кластеров, причём это регулируется одним единственным параметром – коэффициентом отталкивания	Сложная реализация
НС Кохонена	Используется универсальный аппроксиматор – нейронная сеть, самоорганизация сети, простота реализации, гарантированное получение ответа после прохождения данных по слоям	Работа только с числовыми данными, минимизация размеров сети, необходимость задания количества кластеров

Табл. 1. Достоинства и недостатки сравниваемых алгоритмов

Крыжановская С.В., студентка гр.ТТмм-09-2, **Трубицин М.Н.**, к.т.н., доцент,
(Государственное ВУЗ «Национальный горный университет», г. Днепропетровск,
Украина),

Шишко С.В., ассистент

(Государственное ВУЗ «Украинский государственный химико-технологический университет», г. Днепропетровск, Украина),

РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ МЕТОДОМ ПРЯМОГО ПЕРЕБОРА ВЕРШИН

Известные точные методы решения задачи линейного программирования (ЗЛП) основаны на симплекс-методе, итерации которого подразумевают прохождение по ребру N-мерного многогранника (поверхность которого есть граница области определения целевой функции) от вершины с худшим значением целевой функции, к вершине с лучшим значением. Здесь $N = m+n$, где m и n – количество поставщиков и потребителей. Вершины этого N-мерного многогранника представляют собой возможные решения ЗЛП и их общее количество соизмеримо с величиной $1,5 \cdot 2^{N-1} = 24576$ (для рассматриваемого в данной работе случая $N=15$). Производство такого количества итераций не имеет смысла, даже для задачи с $m=3$ и $n=5$, очевидно, поэтому метод перебора вершин в ЗЛП до сих пор не рассматривается и не применяется.

В связи с выше изложенным, **целью** настоящей работы является демонстрация применения метода прямого перебора вершин и определение его целесообразности для ЗЛП. **Идея** работы состоит в выделении различных множеств вершин, которые заранее не могут являться решением ЗЛП. Задача решалась с помощью MathCad, как наиболее доступным, наглядным и воспринимаемым студентами пакетом. Постановка задачи показана на рис.1.

$$\begin{array}{l}
 \begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \begin{array}{l} \text{поставщики} \\ \text{потребители} \end{array} \\
 C = \begin{array}{c} \text{тарифы} \\ \begin{pmatrix} 4.3 & 2.5 & 2.8 & 1.3 & 2.1 \\ 2.1 & 3.5 & 4.7 & 3.8 & 3.2 \\ 4.6 & 4.9 & 2.3 & 3.3 & 2.1 \end{pmatrix} \\
 \text{объемы } a = (104 \ 113 \ 203 \ 45 \ 55) \\
 \text{объемы } b = \begin{pmatrix} 173 \\ 153 \\ 194 \end{pmatrix} \\
 \text{Целевая функция} \\
 z(x) := \sum_{i=1}^m \sum_{j=1}^n (C_{i,j} \cdot x_{i,j}) \\
 \text{найти } (X_1 \dots X_{15})^T = (x_{11} \ x_{21} \ x_{31} \ x_{12} \ x_{22} \ x_{32} \ x_{13} \ x_{23} \ x_{33} \ x_{14} \ x_{24} \ x_{34} \ x_{15} \ x_{25} \ x_{35})^T \quad z \rightarrow \min
 \end{array}
 \end{array}$$

Рис.1. Постановка задачи, коллаж из MathCad-документа.

Первая часть описания области определения целевой функции представляет собой $m+n-1=7$ гиперплоскостей, составляющих баланс спроса и предложений. В связи с закрытостью ЗЛП отбрасываем последнее уравнения (объем - b_3). Для определения координат вершин $X_1 \dots X_{15}$ (как пересечения 15 гиперплоскостей) дополним систему линейных уравнений $m \cdot n - (m+n-1) = 8$ нулевыми (координатными) гиперплоскостями $X_\alpha = 0, X_\beta = 0 \dots X_\omega = 0$, где $\alpha, \beta \dots \omega$ представляют, в общем случае, некоторое сочетание без повторений 8 цифр из 15. Всего таких сочетаний (условимся, что в дальнейшем цифры расположены по возрастающей), а следовательно, вершин области и решений может быть

$$C_{m \cdot n}^{m \cdot n - [m+n-1]} = C_{15}^8 = \frac{15!}{7! \cdot 8!} = 6435$$

При построении всего списка сочетаний использовался лексикографический метод. Для этого в текущем сочетании брался самый правый элемент, не достигший ещё своего наибольшего значения; он увеличивался на 1, а всем последующим элементам присваивались наименьшие значения. Сочетания, включающие цифры 1,2,3, а также 4,5,6 и т.д. не дают решения в силу вырожденности составной матрицы системы линейных уравнений, $\det(A/A_{\text{соч}})=0$, рис. 2.

}	A=	A/A соч =
		A _{соч} =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
6	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
7	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

	1
1	104
2	113
3	203
4	45
5	55
6	173
7	153
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0

Рис. Пример формирования системы при $x_1 = \dots = x_8 = 0$

Отбросив их из рассмотрения, у нас останется 2025 вершин. На следующем шаге (в силу положительности перевозок ЗЛП) можно отбросить вершины, у которых хотя бы одна координата отрицательна. В итоге остается 222 вершины.

Считаем целесообразным отбросить и вершины, в которые включен максимальный тариф, имеем $222 - 100 = 122$. Поступив аналогично с обязательным присутствием минимального тарифа, получим $122 - 72 = 50$. На последнем этапе применим метод двойного предпочтения, отбрасывая «двойной» максимум и оставляя «двойной» минимум в результате получим

$$50 - 30 - 10 = 10 \text{ вершин.}$$

Показанная стратегия отбрасывания «ненужных» вершин является перспективной, и конечно же нуждается в более строгом математическом описании выделения всех перечисленных множеств таких вершин. Наиболее громоздкие действия, связанные с построением всего набора сочетаний S_n^k предполагается в дальнейшем упростить используя один раз построенную матрицу-базу достаточно большого числа сочетаний. Меньшие наборы можно получать из матрицы-базы вычеркиванием соответствующих строк и столбцов.

Выводы по работе:

- Разработана стратегия отбрасывания вершин с заранее не подходящими решениями.
- Показано применение самого простого и надежного метода нахождения экстремума – метода прямого перебора решений
- Средствами MathCad реализован метод лексикографического получения сочетаний из n по k.
- Определены возможные направления обобщения процесса отбрасывания «ненужных» вершин для ЗЛП.

Липодат Е.О., магистр гр. ГИС-08-1М

(ГВУЗ "Национальный горный университет", г. Днепропетровск, Украина)

ГИС-АНАЛИЗ НА ОСНОВЕ ПРОСТРАНСТВЕННЫХ ХАРАКТЕРИСТИК ОБЪЕКТОВ

Пространственная характеристика объекта – [1] числовая характеристика элементарной ячейки карты, функционально отражающая пространственно положение этой ячейки относительно области распространения данного фактора.

Фактор – конкретное образование, определяющее строение территории и отображенное на легенде карты в виде самостоятельной единицы.

Цель работы заключается в формализации местоположения объектов для ГИС-анализа пространственной структуры данных.

Входные данные: изображения – слои ГИС. На каждом слое находятся пространственные объекты с известным положением $\{x, y, z\}$.

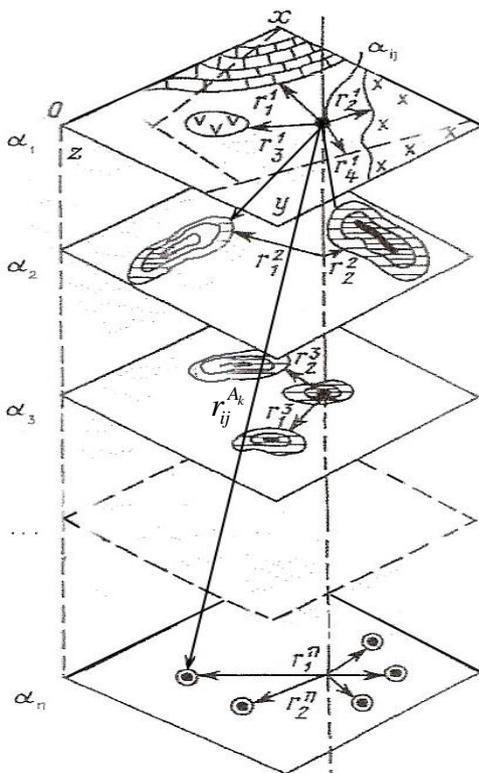


Рисунок 1 – Пространственная структура данных

Разработан следующий алгоритм получения пространственных характеристик:

1. Создание слоев $\alpha_1, \alpha_2, \dots, \alpha_n$, содержащих точечные, линейные или полигональные объекты A_1, A_2, \dots, A_k .
2. Разбиение слоев $\alpha_1, \alpha_2, \dots, \alpha_n$ равномерной сетью на элементарные ячейки a_{ij} , где $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$.
3. Расчет $r_{ij}^{A_k}$ числовых характеристик элементарных ячеек $r_{ij}^{A_k}$, отражающих их пространственное положение (расстояние от центра ячейки до объектов слоя, например, евклидово расстояние).

$$r_{ij}^{A_k} = \sqrt{(x_{a_{ij}} - x_{A_q})^2 + (y_{a_{ij}} - y_{A_q})^2 + (z_{a_{ij}} - z_{A_q})^2},$$

где $\{x_{a_{ij}}, y_{a_{ij}}, z_{a_{ij}}\}$ – положение центра ячейки a_{ij} , $\{x_{A_q}, y_{A_q}, z_{A_q}\}$ – координаты объекта A_q , $q = 1, 2, \dots, k$ слоя.

Выходные данные: набор $M \times N$ характеристик, определяющих пространственное положение элементарных ячеек относительно объектов слоя.

Научная новизна: предложен новый алгоритм для ГИС-анализа пространственной структуры данных на основе расстояний в 2D и 3D пространстве.

Полученные характеристики могут быть использованы для ГИС-анализа пространственной структуры данных и автоматизированной обработки многослойной картографической информации.

Используемые источники:

Б.А. Чумаченко, Е.П. Власов, В.В. Марченко. Системный анализ при геологической оценке перспектив рудоносности территорий – М.: Недра, 1980.

Нікулін С.Л., к.г.н., Підгорна Г.С., магістр гр. ПС-08-1м

(Государственное ВУЗ «Национальный Горный Университет», г.Днепропетровск, Украина)

ПОСТРОЕНИЕ ГЕОЛОГИЧЕСКИ СОДЕРЖАТЕЛЬНЫХ СИСТЕМ ПРИЗНАКОВ ПРИ АВТОМАТИЗИРОВАННОМ РЕШЕНИИ ПРОГНОЗНО-ПОИСКОВЫХ ЗАДАЧ

Основным методом автоматизированного решения прогнозно-поисковых геологических задач является классификация с учителем. Выполнение классификации предполагает наличие а) обучающих выборок, отражающих информацию о локализации известных объектов прогнозируемого типа, а также местоположении заведомо «нецелевых» объектов, и б) признаков, представленных в сеточном виде физических полей, геохимических ореолов, космических снимков и их трансформант, полученных путём математических преобразований первичных материалов.

В результате классификации, проводимой на основе анализа положения объектов в многомерном пространстве признаков, элементарные участки изучаемой территории относятся к одному из взаимоисключающих классов, оцениваются по степени принадлежности объектов к определённому классу.

На качество получаемых результатов решающее влияние оказывает информативность используемой системы признаков. Система считается информативной, если при заданных обучающей выборке и выбранном методе построения решающего правила она обеспечивает классификацию с минимальной величиной потерь. Система признаков, обеспечивающая минимум величины потерь, называется диагностирующей. В настоящий момент оценка величины потерь обычно сводится к расчёту байесовской вероятности ошибки классификации (ВОК), вычисляемой на основе расстояния Фишера, Бхаттачариа, Колмогорова, дивергенции Кульбака или других [1].

Отличительной чертой геологического прогнозирования является малый объём обучающих выборок. Как следствие, система признаков, дающая минимальные значения ВОК, может оказаться в действительности абсолютно бессодержательной с геологической точки зрения. Предлагается рассчитывать величину потерь на основе двух критериев – ВОК и также т.н. геологической содержательности, под которой понимается способность системы достоверно отражать геологическое строение территории.

Предлагается следующая методика построения диагностирующего набора с использованием детерминистских решающих правил. Пусть дан некоторый фиксированный набор признаков X' и цифровая геологическая карта территории. Ячейкам карты приписаны достоверно известные значения некоторого геологического атрибута (например, возраст пород, название литологической разности и т.п.). Рассматривается совокупность ячеек карты с одинаковым значением атрибута, как отдельный кластер C_i , $i=1..M$, где M – количество уникальных значений атрибута. Пусть принадлежность элементарных участков территории определённому кластеру задаётся значениями соответствующих участков геологической карты.

В этом случае геологическую содержательность системы можно оценить как

$$Q = S_{BK} / S_{MK} \quad (1)$$

где S_{BK} – среднее внутрикластерное расстояние для M кластеров;

S_{MK} – среднее межкластерное расстояние.

Q отражает компактность расположения объектов с одинаковым значением атрибута в многомерном пространстве признаков, и по сути, характеризует способность набора признаков отражать наиболее общие черты геологического строения территории. Чем ниже значения Q , тем более геологически содержателен набор признаков.

Построение диагностирующего набора выполняется любым из известных методов (последовательного присоединения, последовательного отбрасывания, случайного поиска и т.п.); на каждой итерации решается задача многокритериальной оптимизации, заключающаяся в поиске такой совокупности признаков, при котором достигается минимум функции она-ла $f(BOK, Q)$. Данная задача может быть решена, например, при помощи выбора решения из множества парето-оптимальных (недоминируемых) альтернатив [2].

Фактически, оптимизация BOK обеспечивает соответствие результатов имеющимся сведениям об объектах прогнозируемого типа, Q – геологическим фактам о геологической среде в целом.

Применение описанного подхода на ряде нефтегазовых и рудных (рис. 1) месторождений продемонстрирует его преимущества перед традиционным за счёт более полного учёта имеющейся геологической информации [3].

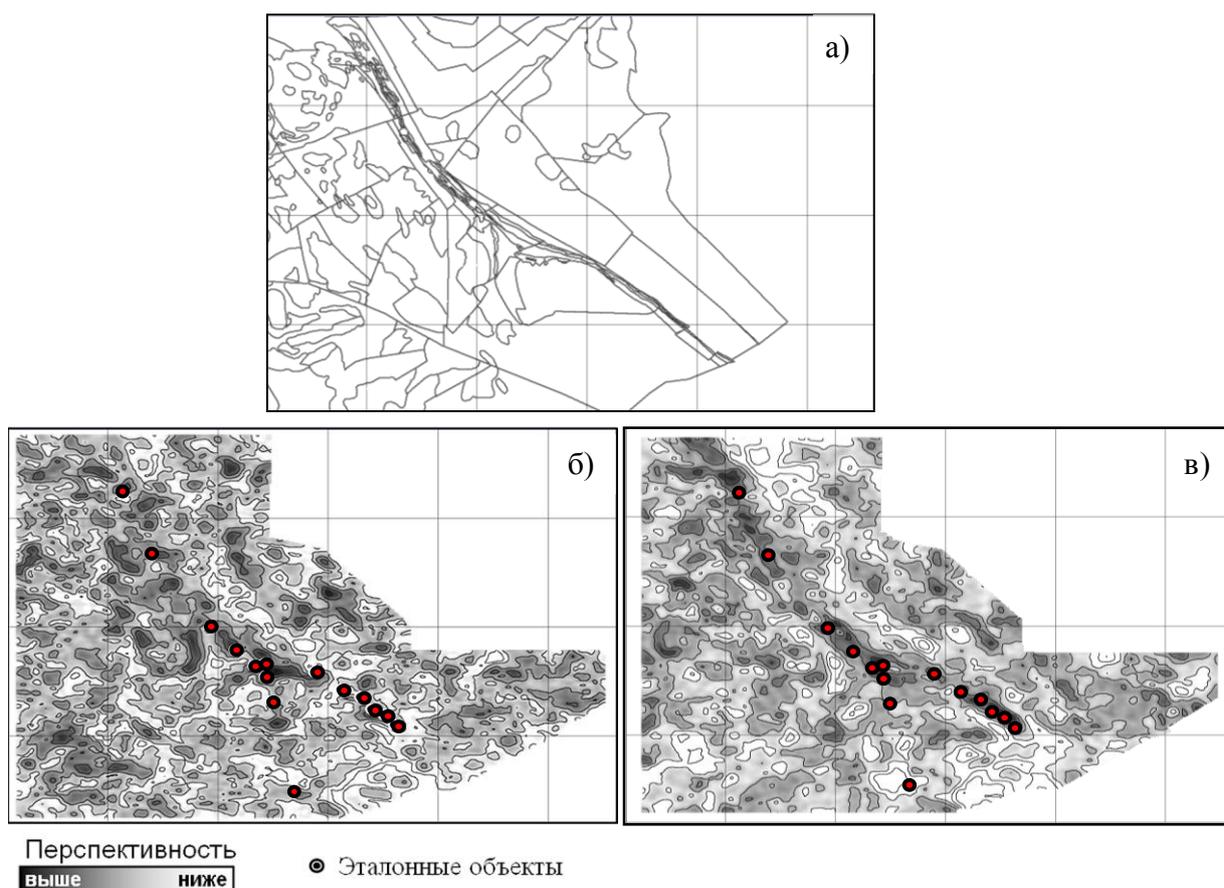


Рисунок 1 – Результаты классификации территории Сорокинской структуры по мере близости к золоторудным объектам: а) контуры объектов исходной геологической карты; б) результаты классификации с использованием традиционного; в) предложенного подходов к формированию информативной системы признаков

Литература

1. Бусыгин Б.С., Мирошниченко Л.В. Распознавание образов при геолого-геофизическом прогнозировании. Днепропетровск: Изд-во ДГУ. – 1991. – 168 с.
2. Лотов В.А., Поспелова И.И. Многокритериальные задачи принятия решений. М: МАКС Пресс. – 2008. – 197 с.
3. Pivnyak G.G., Busygin B.S., Nikulin S.L., Svistun V.K., Garkusha I.N. Geoinformation technology for forecasting methane accumulations within the Donetsk basin mine fields // 22nd World Mining Congress, Istanbul, Turkey, 2011. – P. 641-647.

Носова Д.В., магистр гр.КМ08-М,

Научный руководитель Гаркуша И.Н., к.т.н., доцент

(ГБУЗ «Национальный Горный Университет», г. Днепропетровск, Украина)

РАЗРАБОТКА ТЕХНОЛОГИИ МОНИТОРИНГА ПЛОЩАДИ РАСТИТЕЛЬНОСТИ ДНЕПРОПЕТРОВСКА ПО КОСМИЧЕСКИМ ИЗОБРАЖЕНИЯМ

Специфика современной экологической обстановки определяет особую значимость проблемы зеленых насаждений, которую невозможно решить без оценки их состояния.

В основу разработки технологий мониторинга зеленых насаждений положена классификация объектов растительного мира расположенных на землях населенного пункта, по функциональному назначению с учетом особенностей экологической обстановки урбанизированной территории. Мониторинг охватывает: насаждения на улицах и дорогах, общегородские парки, скверы, бульвары.

Цель работы – создание технологии мониторинга зеленой зоны по данным космосъемки. Объект исследования – космоснимки высокого и низкого разрешений г. Днепропетровска. Предмет исследования – алгоритмы, технологии обработки космических снимков и методы расчета определения зеленой массы.

В качестве исходных данных выступают снимки г. Днепропетровска со сканеров LANDSAT-5 TM и Terra MODIS. При выполнении исследований используются программные комплексы: ERDAS Imagine, ENVI, MatLab.

Для достижения поставленной цели решаются следующие задачи:

- 1) Поиск площадей растительности и последующая обработка изображений на основе многоуровневой сегментации;
- 2) Разработка критериев определения границ площадей растительности;
- 3) Описание количественной оценки при обработке изображений различными алгоритмами.

Поиск площадей зеленых насаждений проводится несколькими способами. При анализе алгоритмов и технологий обработки космических изображений, а также методах расчета определения зеленой массы используют два подхода к решению этой задачи:

- использование вегетационных индексов;
- преобразование «колпак с кисточкой»;
- метод главных компонент (РСА).

Отражение растительного покрова в красной и ближней инфракрасной (ИК) областях электромагнитного спектра связано с его зеленой фитомассой. Чтобы оценить состояние растительности, применяется количественный показатель - нормализованный разностный вегетационный индекс NDVI (Normalized Difference Vegetation Index):

$$NDVI = \frac{B_{ИК} - B_K}{B_{ИК} + B_K} \quad (1),$$

где $B_{ИК}$ - отражение в ближней ИК-области спектра; B_K - отражение в красной области спектра.

В красной области спектра (0,6-0,7 мкм) лежит максимум поглощения солнечной радиации хлорофиллом высших сосудистых растений, а в ИК-области (0,7-1,0 мкм) находится область максимального отражения клеточных структур листа. То есть высокая фотосинтетическая активность (связанная, как правило, с густой растительностью) ведет к меньшему отражению в красной области спектра и большему в ИК. От-

ношение этих показателей друг к другу позволяет четко отделять и анализировать растительные объекты от прочих природных. Использование нормализованной разности между минимумом и максимумом отражений увеличивает точность измерения, позволяет уменьшить влияние таких явлений как различия в освещенности снимка, облачности, дымки, поглощение радиации атмосферой и пр.

Таблица 1. Значение NDVI для различных типов земного покрытия [1]

Тип объекта	Отражение в красной области спектра	Отражение в ИК-области спектра	Значение NDVI
Густая растительность	0.1	0.5	0.7
Разряженная растительность	0.1	0.3	0.5
Открытая почва	0.25	0.3	0.025
Облака	0.25	0.25	0
Снег и лед	0.375	0.35	-0.05
Вода	0.02	0.01	-0.25
Искусственные материалы (бетон, асфальт)	0.3	0.1	-0.5

Преобразование «колпак с кисточкой». Если необходимо определить «бедную» растительность, - используется перпендикулярный индекс (Perpendicular Vegetation Index) PVI. Имеет узкий динамический диапазон, чувствителен к изменению атмосферы. Прост в использовании и нахождении почвенной линии.

Почвенная линия - это гипотетическая линия в спектральном пространстве, которая описывает вариацию спектра открытой почвы на снимке. Линия определяется через определение двух или более эталонов голой почвы на снимке, имеющих разное отражение и определение наиболее подходящей линии в спектральном пространстве. [2]. Перпендикулярный индекс вычисляется:

$$PVI = \sin(a) * B_{ИК} - \cos(a) * B_K \quad (2),$$

где a - угол между почвенной линией и осью $B_{ИК}$.

Параметры индекса:

- Перпендикулярный индекс
- Изовегетационные линии параллельны почвенной линии;
- Почвенная линия: может иметь разный наклон, проходит через начало координат;
- Возможные значения: -1..1

Метод Главных Компонент (РСА) – один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных или к сингулярному разложению матрицы данных. Другие способы уменьшения размерности данных — это метод независимых компонент, многомерное шкалирование, а также многочисленные нелинейные обобщения: метод главных кривых и многообразий, поиск наилучшей проекции (англ. Projection Pursuit), нейросетевые методы «узкого горла», самоорганизующиеся карты Кохонена и др.

Литература

1. http://mapexpert.com.ua/index_ru.php?id=20&table=news
2. R.J. Kauth and G.S. Thomas, "The tasseled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT." Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University of West Lafayette, Indiana, 1976, pp. 4B-41 to 4B-51.

**Чуб А.А., магистр, группа КМ-08-1м,
Научный руководитель Сарычева Л.В., канд. ф.-м. н., проф. каф. ГИС
(ГВУЗ «Национальный горный университет», г. Днепропетровск, Украина)**

РАЗРАБОТКА ПРОГРАММНОГО ПАКЕТА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ПРОСТРАНСТВЕННЫХ ДАННЫХ МОНИТОРИНГА РЕГИОНОВ

В настоящее время интеллектуальный анализ данных (ИАД) мониторинга регионов – актуальная задача. ИАД-программы – коммерческие. Данные мониторинга регионов характеризуются пространственной привязкой. Для их исследования требуется создание новых алгоритмов интеллектуального анализа пространственных данных (ИАПД).

Цель работы – разработка пакета программ ИАПД с открытым кодом.

Структура пакета программ ИАПД следующая:

- 1) Блок управления данными и картами
 - а) масштабирование
 - б) перемещение карты
 - в) открытие нового файла
 - г) просмотр базы данных
 - д) редактор базы данных
 - е) установление подписей на карте
- 2) Блок классификации и кластеризации
 - а) классификация
 - б) кластеризация
 - в) выбор атрибутов для кластеризации
 - г) установление количество классов (кластеров)
 - д) выбор поля для классификации
- 3) Блок выделения аномалий
 - а) выделение пространственно-временных аномалий

В блоке (2) реализованы такие методы анализа данных:

- классификация («равные интервалы», «стандартное отклонение»)
- кластеризация «K-средних»

Выделение пространственно-временных аномалий (блок 3) основано на классификации «стандартное отклонение».

Особенности применения методов ИАПД для анализа показателей эколого-социально экономического (ЭСЭ) мониторинга регионов обусловлены характером исходных данных и содержательной постановкой решаемых задач. Одни и те же методы ИАПД используются для различных по содержанию задач. Например, кластеризация применяется и для анализа сети мониторинга, и для районирования территории (по степени загрязнения, по совокупности показателей тематического сюжета «инвестиционная привлекательность» регионов и т.п.), и для выделения аномальных регионов. Кластеризация используется как на этапе разведочного анализа, так и на этапе извлечения знаний из ЭСЭМ-данных, проверки гипотез о сходстве протекающих в различных регионах ЭСЭ-процессов.

На основе результатов ИАПД мониторинга создаются геоинформационные модели – принципиально новый класс ЭСЭ-моделей, основанный на иконическом способе отображения ЭСЭ-информации, упорядоченной с помощью географической карты. В соответствии с тематическим сюжетом по совокупности изучаемых ЭСЭ-показателей создаются карты. Информация представляется для дальнейшего анализа в сжатом, синтетическом виде, а карта рассматривается как аналитическое средство.

При написании пакета программ используется язык Java. Его преимущество – кроссплатформенность (работает одинаково на разных операционных системах), производительность.

Пакет программ создается с использованием бесплатной библиотеки GeoTools проекта OSGeo. Проект предоставляет различные инструменты и библиотеки для создания и расширения возможностей ГИС-программ.

В качестве входных данных используется шейп-файл. Созданный компанией Esri с целью интероперабельности между различными ГИС-программами. Каждый объект (геометрический объект) позволяет хранить несколько атрибутов («Валовый внутренний продукт», «Выбросы вредных веществ в атмосферу» и т.п.), используемые для ИАПД. При использовании требуется выбрать количество классов и признаки, по которым будет производиться анализ. Результатом выполнения программы является раскрашенная карта с легендой – геоиконическая модель, отражающая ИАПД мониторинга ЭСЭ-показателей регионов. На рисунке 1 пример геоиконической модели с тематическим сюжетом «Количество выбросов вредных веществ в атмосферу»

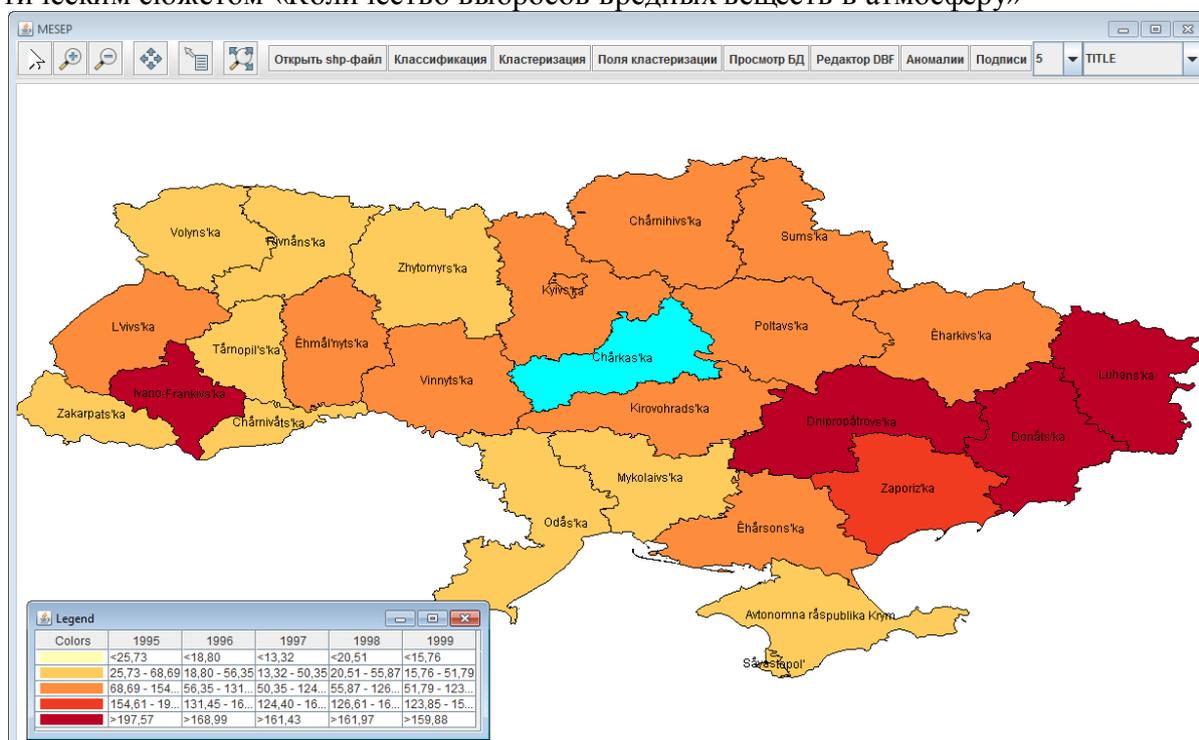


Рисунок 1. Выделение пространственно-временных аномалий по тематическому сюжету «Количество выбросов вредных веществ в атмосферу» за 5 лет

Разработанная программа осуществляет интеллектуальный анализ пространственных данных мониторинга регионов. Новизна: созданный ПП ИАПД внедряет функции ИАД в ГИС(в существующих ГИС нет блока анализа аномалий, кластеризации).

Шевченко А.А., магистр, гр.КМ-08-1м
Научный руководитель – проф. каф. ГИС Коротенко Г.М.
(ГВУЗ «Национальный горный Университет»)

РАЗРАБОТКА ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ СОЗДАНИЯ ПРОСТРАНСТВА ДАННЫХ УЧЁТА АБИТУРИЕНТОВ ДЛЯ КАФЕДРЫ ГИС

Пространство данных – организованная в соответствии с определёнными правилами и поддерживаемая в памяти компьютера совокупность данных, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователя.

Множество разнородных наборов данных и файлов предметной области, именуемое пространством данных можно представить следующим образом:

$$DS = \langle DB, DW, Wb, Nd, Gr \rangle,$$

где DB, DW, Wb, Nd, Gr разнородные наборы данных, которые представляют содержимое множества баз данных (БД), хранилищ данных, Веб-страниц, текстовых файлов, электронных таблиц, графических данных соответственно.

Цель работы заключается в создании пространства данных «Приём абитуриентов», для автоматизации обработки информации, получаемой в процессе профориентационной работы кафедры ГИС с учебными заведениями Днепропетровской области, а также мониторинга рейтингов абитуриентов кафедры ГИС в структуре данных МОНУ на основании разработанной информационной технологии.

Актуальность решаемых задач заключается в создании информационной технологии оперативного мониторинга рейтингов абитуриентов кафедры ГИС в структуре учебных заведений Украины.

Структура разрабатываемой информационной технологии состоит из трёх частей:

- базы данных, в которой накапливаются данные о студентах 1-5 курсов, обучающихся на кафедре ГИС;
- базы данных учебных заведений (школ, техникумов, училищ, ВУЗов) Днепропетровской области;
- базы данных информации об абитуриентах, постоянно обновляемой на сайте МОНУ.

В работе решаются следующие задачи:

- разрабатывается комплекс форм, реализующих пользовательский интерфейс пространства данных «Прием абитуриентов»;
- создаётся инструментарий отбора и сортировки необходимой информации из наборов данных большого объёма и разной структуры;
- автоматизируется процесс вывода отчетных документов;
- формируется структура пространства данных и комплекс средств автоматизации их обработки. Данный комплекс позволяет производить анализ различных взаимосвязей в структуре ВУЗов Украины между рейтингами абитуриентов для продолжения профориентационной работы.

Пространство данных разворачивается в среде СУБД MS Access. Элементами пространства являются базы данных «Студенты», «Учебные заведения», БД «Абитуриенты», на основе мониторинга Веб-сайта МОНУ. В основу информационной технологии положены процедурные решения, реализованные на скриптовом языке Visual Basic for

Application (VBA). Модели данных пространства данных разрабатываются с использованием языка визуализации UML, входящего в состав пакета VoUML.

В результате получены таблицы исходных данных, схемы целостности данных, реляционные диаграммы, диаграммы классов (UML-диаграммы), а также ряд форм, отчётов и запросов для отбора необходимой информации.

Впервые разрабатывается пространство данных «Прием абитуриентов» кафедры ГИС, позволяющее накапливать информацию о результатах профориентационной работы и выполнять анализ этих данных с целью совершенствования деятельности по привлечению абитуриентов для обучения на кафедре. Созданное пространство наполняется реальными данными об учебных заведениях Днепропетровской области, в том числе и о тех, с которыми работают сотрудники кафедры, данными о студентах, а также данными об абитуриентах, привлеченных на дни открытых дверей Государственного ВУЗ «НГУ».