

ТЕХНОЛОГИЯ СТАТИСТИЧЕСКОГО АНАЛИЗА ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ХАРАКТЕРИСТИК НА ОСНОВЕ БАЗОВЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ

В.В. Парамонов, Р.К. Фёдоров, Г.М. Ружников, П.В. Белых

Институт динамики систем и теории управления Сибирского отделения Российской академии наук
(ИДСТУ СО РАН)

Иркутск, Россия

E-mail: {slv, fedorov, rugnikov, polina}@icc.ru

В работе рассматривается создание информационной технологии, позволяющей проводить статистический анализ для характеристик, имеющих пространственную и временную привязку. Применение технологии дает возможность оценки зависимости различных факторов от пространственного положения.

ВВЕДЕНИЕ

Развитие информационных и телекоммуникационных технологий делает более популярным использование информационных систем, базирующихся на применении Интернет/интранет и имеющих многоуровневую архитектуру. Информационные ресурсы для подобного рода систем целесообразно реализовывать в качестве сервисов. Это позволяет обеспечить межпрограммное взаимодействие на основе веб-стандартов и предоставит возможность более гибкого использования ресурса на различных платформах.

Следует отметить характерную для всего мира тенденцию проникновения геоинформационных технологий в различные сферы деятельности человека. На протяжении двух последних десятилетий происходит активное формирование инфраструктуры пространственных данных, основой которых являются базовые пространственные данные (БПД).

Комплексное использование БПД и различного рода тематической информации, представленной в реляционном виде и имеющей пространственную привязку, дает возможность более эффективно решать широкий спектр задач.

I. ОПИСАНИЕ ЗАДАЧИ И ЕЁ АКТУАЛЬНОСТЬ

Одними из источников получения различного рода тематических данных являются отчеты, представляемые Федеральной Службы государственной статистики РФ, Роспотребнадзором, Министерством Здравоохранения и т.п. Такая информация, как правило, требует нормализации и последующей загрузки в тематические базы данных. После чего возможно проведение исследований.

Оценка связей между различными данными, с учетом их пространственно-временных характеристик, позволяет проводить их качественно-количественные оценки. Например, зависимость содержания в атмосфере муниципальных образований какого-либо загрязняющего вещества (сажа, оксид углерода, диоксид

серы и т.п.) и уровнем заболеваемости населения болезнями органов дыхания (J00-J99 по Международному классификатору болезней X пересмотра). Результат анализа может быть привязан к БПД, а также может быть сформирована тематическая карта, что в итоге это дает возможность наглядно представить результаты исследований.

II. ЭТАПЫ ТЕХНОЛОГИИ СТАТИСТИЧЕСКОГО АНАЛИЗА

Пространственно-статистический анализ реализован в виде WPS-сервиса [1], обеспечивающего проведение статистического анализа (определение коэффициентов регрессии и корреляции) для данных, имеющих пространственно-временные характеристики. Искомый WPS-сервис является компонентой геопортала ИДСТУ СО РАН [2]. Взаимодействие с пользователем осуществляется через стандартный Интернет обозреватель. Это позволяет исключить необходимость установки специализированного программного обеспечения на рабочую станцию пользователя. Архитектура геопортала представлена на рис. 1.

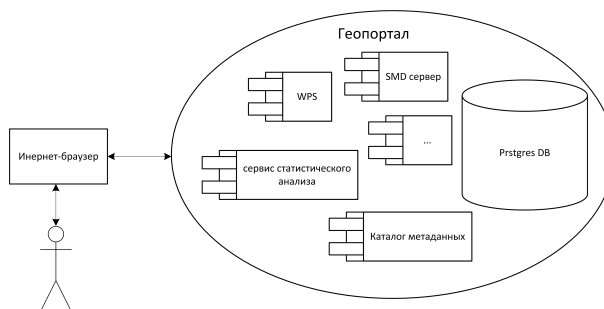


Рис. 1 – Архитектура геопортала ИДСТУ СО РАН

Разработанная технология статистического анализа данных состоит из следующих шагов:

- сбор и загрузка данных;
- нормализация данных;
- пространственно-статистический анализ.

Рассмотрим эти этапы более подробно.

III. СБОР И ЗАГРУЗКА ДАННЫХ

В рамках геопортала ИДСТУ СО РАН разработаны сервисы ввода и редактирования реляционных данных, содержащих пространственные атрибуты. Сервисы обеспечивают:

- многопользовательскую работу через Интернет/интранет — одновременно несколько пользователей могут вводить, редактировать и просматривать данные;
- возможность для пользователя самостоятельно создавать таблицы и определять атрибутный состав таблиц;
- автоматически ввод и отображение на карте пространственных данных;
- возможность применения различных фильтров, в том числе пространственных.

IV. НОРМАЛИЗАЦИЯ ДАННЫХ

Первичным этапом в обработке реляционных данных является их загрузка и нормализация. При этом обеспечивается как идентификация простых типов данных, так и возможная привязка значений к иерархическим справочникам.

V. ПРОСТРАНСТВЕННО-СТАТИСТИЧЕСКИЙ АНАЛИЗ

В работе предлагается технологическое решение позволяющее обрабатывать серию данных, сформированных в соответствии со спецификацией формата GeoTIFF [3]. Предлагаемый подход позволяет обрабатывать как информацию из БД геопортала, так и из пользовательских документов, представленных в этом формате.

Информация в формате GeoTIFF позволяют представить исследуемую территорию в виде ячеек регулярной сетки. В ячейках регулярной сетки можно представить пространственное распределение некоторой характеристики. В геоинформационных системах такое представление называется GRID моделью. Для описания серии файлов используется специализированный файл в формате MTIFF. В этом случае серия GeoTIFF файлов позволяет описать динамику изменения некоторой характеристики в течении времени.

Основой сервиса статистического анализа является библиотека `alglib`[4]. Функциями библиотеки реализовано большое число алгебраических операций.

Структуры данных, получаемые сервисом на вход, представляют собой две трехмерные матрицы. Проводится контроль размерности матриц для обеспечения корректного выполнения операций статистического анализа. В случае работы с данными, представленными в виде файлов, идентичность пространственно-временных характеристик данных должен установить пользователь. Если же данные экстрагируются из БД

Геопортала, то проверка корректности происходит автоматически.

Матричные данные раскладываются в одномерный вектор, над которым выполняется статистический анализ. Результаты анализа могут быть представлены как в виде значения, показывающего связь данных, так и в виде тематической карты, иллюстрирующей взаимосвязь данных на различных территориях. Общая схема работы сервиса статистического анализа данных представлена на рис. 2.

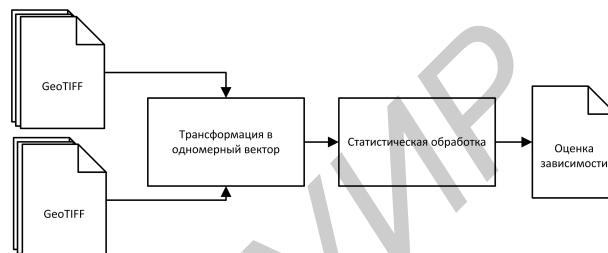


Рис. 2 – Схема работы сервиса статистического анализа данных

Разработанный сервис представляет собой компилированный файл, используемый WPS-сервисом. Реализованный подход позволяет использовать методы статистического анализа и для других ресурсов.

ЗАКЛЮЧЕНИЕ

В результате работы предложена технология проведения статистического анализа пространственно-временных характеристик на основе БПД. Практическое применение технологического решения позволяет проводить анализ зависимости различных тематических данных, имеющих пространственную и временную привязку.

Следующим шагом в развитии технологии является разработка математических моделей, обеспечивающих анализа данных исходя из множества параметров.

Работа частично поддержана РФФИ: грант 14-47-04125, РАН: грант ФНМ-49, Советом по грантам Президента Российской Федерации для государственной поддержки ведущих научных школ (НШ-5007.2014.9)

1. OGC OpenGIS Web Processing Service.[Электронный ресурс] – Режим доступа: <http://www.opengeospatial.org/standards/wps>. – Дата доступа: 08.09.2014.
2. Бычков И. В. Создание инфраструктуры пространственных данных в управлении регионов /, И. В. Бычков, В. М. Плюсин, Г. М. Ружников [и др.] // География и природные ресурсы. – 2013. – № 2. – С. 145 – 150.
3. N. Ritter M. Ruth The GeoTiff data interchange standard for raster geographic images // International Journal of Remote Sensing –Volume 18, Issue 7, 1997, pp. 1637-1647.
4. ALGLIB [Электронный ресурс] –Режим доступа: <http://www.alglib.net/>. – Дата доступа 01.09.2014.