

Автоматическая локализация пространственных некорректностей при массовом импорте геоинформационных данных

М.Ю. Трухачев

Южный федеральный университет, Ростов-на-Дону

Аннотация: При подготовке тематических карт средствами различных геоинформационных систем вопрос интеграции данных решается обменом файлами данных открытых форматов. Методика массового импорта таких файлов с полуавтоматическим отделением и ручным исправлением дефектных площадных пространственных объектов предполагает значительные трудозатраты на поиск всех некорректностей в каждом таком объекте. Предложено программное решение для автоматической локализации пространственных некорректностей, рассмотрена его новизна и перспектива качественного совершенствования.

Ключевые слова: градостроительная документация, тематическая карта, геоинформационная система, ObjectLand, интеграция данных, обменный формат, источник ошибок, некорректность полигонального объекта, кластеризация точек на плоскости, K-медоид.

Введение

В практике выпуска градостроительной документации (планы землепользования и застройки, генеральные планы, схемы территориального планирования) значительное место занимает подготовка тематических карт. Тематическое картографирование подразумевает применение операций пространственного анализа, направленное на создание информации, производной от географически неоднородных статистических данных. При создании тематических карт применяются геоинформационные системы (далее ГИС), представляющие собой «удобный механизм анализа и визуализации географических данных» [1]. В статье описана сложившаяся практика [2-4] полуавтоматического импорта пространственных и атрибутивных данных без потери дефектных объектов в отечественную ГИС ObjectLand [5], уделено внимание принципиальному недостатку существующей методики – необходимости вручную искать участки, содержащие некорректность. Предлагается автоматизирование процесса локализации некорректностей пространственных метрик замкнутых

полигональных (далее – площадных) объектов при импорте данных. Новизна предлагаемого подхода заключается в его большей универсальности по сравнению с решениями, привязанными к определённой ГИС [6-8]. Рассмотрена возможность дальнейшего усовершенствования качества выходных данных алгоритма локализации.

Существующая методика импорта данных

Далее описана существующая ныне методика массового импорта геоинформационных данных из файлов открытых обменных форматов MIF (Map Interchange Format) и SHP (Shapefile) с полуавтоматической локализацией и ручным исправлением ошибок, допущенных в площадных пространственных объектах. Возможности импорта данных из обменных форматов в ГИС ObjectLand [9] позволяют организовать последовательный ввод «файл за файлом» в интерактивном режиме, с ручным заполнением либо правкой состава атрибутивных полей для каждого файла. Чтобы избавить пользователя от рутинной работы и избежать появления ошибок при заполнении полей диалоговых окон, была написана программа BatchImporter [2], автоматизирующая импорт множества файлов обменных форматов MIF и SHP. Для организации массового импорта данных имеет значение факт наличия структурной обработки исключений (SEH, Structured Exception Handling), реализованной в BatchImporter средствами языка VB.Net, входящего в пакет MS Visual Studio 2005 [10]. Большинство нештатных ситуаций, происходящих в процессе импорта данных и при подготовительных и/или завершающих ситуациях, обрабатывается без аварийного завершения всей программы и сопровождается соответствующей записью в файле журнала, имеющего формат «плоский текст» для удобства чтения.

Перечислим некоторые источники ошибок в пространственных данных площадного типа:

- аппроксимация отрезками дуг и сплайнов порождает множество слишком близко расположенных точек в пределах одного контура, которые могут восприниматься алгоритмами пространственного анализа как одна вершина, что вызовет сбой при дальнейшей обработке;

- в результате операций булевой алгебры над пространственными объектами могут появляться самопересечения контуров, а также участки площадных фигур и даже целые объекты с нулевой площадью;

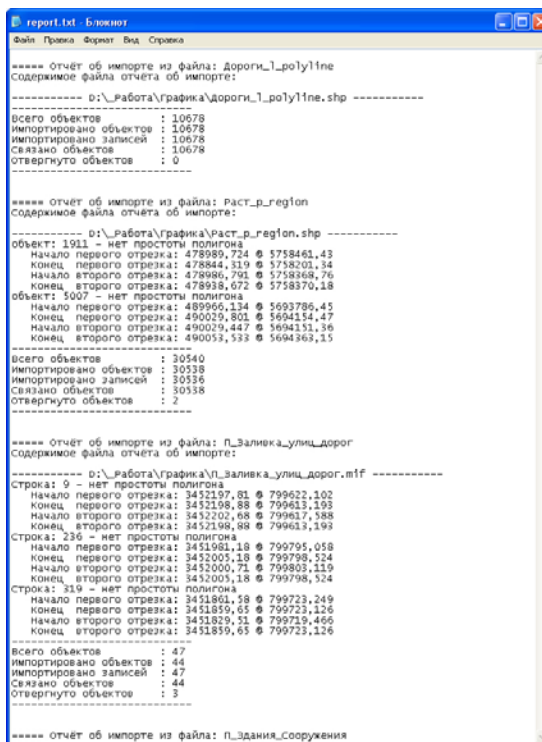
- сбои при автоматической и полуавтоматической векторизации растровых данных, вызванные несовершенной работой алгоритма в сомнительных участках растра, проявляются как хаотичное переплетение на небольшой площади множества отрезков;

- небрежность оператора, например, самопересечения вследствие повторных щелчков мышью при построении некоторых вершин, либо случаи пренебрежения, когда по построению внутренние подконтурные соединены участком нулевой площади между собой и/или с внешним подконтуром.

Площадные объекты, хранящиеся в файлах обменных форматов, могут иметь следующие некорректности: совпадение двух и более вершин, касание вершиной одного подконтурного отрезка другого, самопересечения либо взаимные пересечения подконтуров, выход внутренних подконтуров за пределы внешнего. В ГИС ObjectLand реализована проверка на этапе импорта данных площадных объектов на наличие вышперечисленных некорректностей. Подсистема импорта не импортирует такие площадные объекты, выводя в файл отчёта об импорте указание номера строки MIF-файла либо номера объекта, в случае SHP-файла. На рис. 1 представлен фрагмент содержимого текстового файла отчёта.

В настоящее время, при помощи вспомогательных программ MifConvert [3] и ShpConvert [4] оператор импорта данных отделяет от

обменного файла только те объекты, в которых при импорте была обнаружена хотя бы одна некорректность.



```
geronf.txt - Блокнот
Файл Правка Формат Вид Справка

===== Отчет об импорте из файла: дороги_1_polyline
Содержимое файла отчета об импорте:
-----
D:\Работа\Графика\дороги_1_polyline.shp
-----
Всего объектов : 10678
Импортировано объектов : 10678
Импортировано записей : 10678
Связано объектов : 10678
Отвергнуто объектов : 0

===== Отчет об импорте из файла: Раст_p_region
Содержимое файла отчета об импорте:
-----
D:\Работа\Графика\Раст_p_region.shp
-----
Объект: 1911 - нет простоты полигона
Начало первого отрезка: 478989,724 @ 5758461,43
Конец первого отрезка: 478984,319 @ 5758201,34
Начало второго отрезка: 478986,791 @ 5758368,76
Конец второго отрезка: 478938,672 @ 5758370,18
Объект: 1007 - нет простоты полигона
Начало первого отрезка: 489956,134 @ 5693786,45
Конец первого отрезка: 490029,801 @ 5694154,47
Начало второго отрезка: 490029,447 @ 5694151,36
Конец второго отрезка: 490053,533 @ 5694363,15
-----
Всего объектов : 30540
Импортировано объектов : 30538
Импортировано записей : 30538
Связано объектов : 30538
Отвергнуто объектов : 2

===== Отчет об импорте из файла: П_Заливка_улиц_дорог
Содержимое файла отчета об импорте:
-----
D:\Работа\Графика\П_Заливка_улиц_дорог.tif
-----
Строка: 0 - нет простоты полигона
Начало первого отрезка: 3452197,81 @ 799622,102
Конец первого отрезка: 3452198,88 @ 799613,193
Начало второго отрезка: 3452202,68 @ 799617,388
Конец второго отрезка: 3452198,88 @ 799613,193
Строка: 136 - нет простоты полигона
Начало первого отрезка: 3451981,18 @ 799795,058
Конец первого отрезка: 3452003,18 @ 799796,324
Начало второго отрезка: 3452000,71 @ 799803,119
Конец второго отрезка: 3452003,18 @ 799798,524
Строка: 319 - нет простоты полигона
Начало первого отрезка: 3451861,58 @ 799723,249
Конец первого отрезка: 3451859,65 @ 799723,126
Начало второго отрезка: 3451859,51 @ 799719,466
Конец второго отрезка: 3451859,65 @ 799723,126
-----
Всего объектов : 47
Импортировано объектов : 44
Импортировано записей : 47
Связано объектов : 44
Отвергнуто объектов : 3

===== Отчет об импорте из файла: П_Здания_Сооружения
```

Рис. 1. – Записи в файле журнала импорта при наличии некорректных площадных объектов

Затем эти данные доимпортируются в качестве линейных, а не площадных, объектов в ГИС ObjectLand, возможно, при помощи уже упоминавшейся программы BatchImporter, поддерживающей доимпорт в ранее созданную графическую базу данных. В среде ObjectLand проблемные данные редактируются вручную для устранения всех некорректностей, далее каждому объекту пользователь вручную меняет тип геометрии с линейного обратно на площадной в соответствии с порядком, описанным в [11]. Внимание при ручной обработке следует уделить связанной с объектом табличной (атрибутивной) записи, эта информация не должна игнорироваться. На рис. 2 показана блок-схема описанного процесса импорта данных.

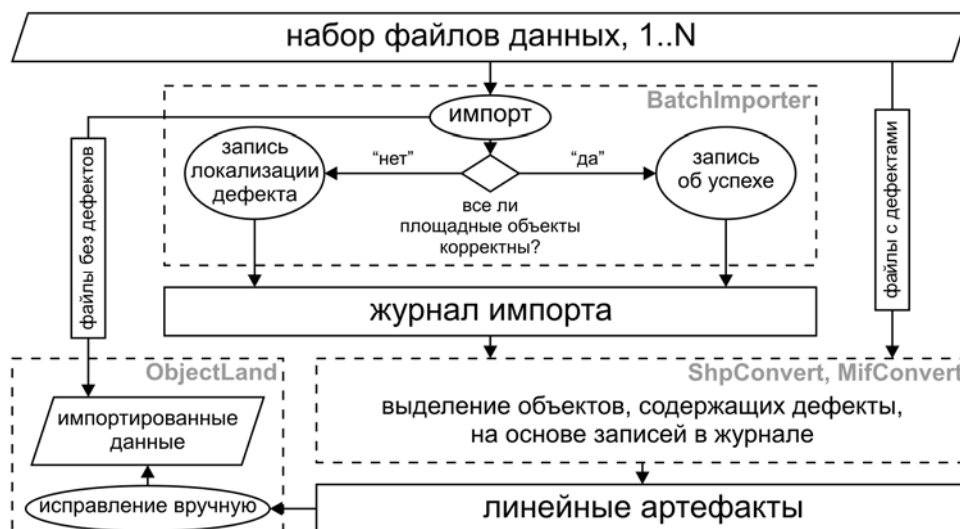


Рис. 2. – Блок-схема процесса импорта данных из обменных форматов с доимпортом исправленных площадных объектов

Принципиальным недостатком описанной методики импорта данных является необходимость просмотра оператором всего контура каждого проблемного объекта путём длительного панорамирования при достаточно крупном масштабе отображения в целях последовательной локализации и исправления всех имеющихся некорректностей, что занимает неоправданно долгое время и вызывает утомление при большом количестве однотипных векторных данных. Так происходит потому, что в общем случае для дефектного площадного объекта в журнал отчёта об импорте заносятся не все точки локализации некорректностей.

Предлагаемое программное решение

Представляется целесообразным проведение автоматической локализации некорректностей площадных объектов до импорта в ГИС. На рис. 3 приведена блок-схема верхнего уровня функциональной декомпозиции программного решения поставленной задачи. По сравнению с существующими решениями [6-8], привязанными по реализации к определённой ГИС, предлагаемое в данной статье дополнение к существующей методике более универсально, так как для получения

пространственных данных оперирует непосредственно с обменными файлами.



Рис. 3. – Блок-схема программной локализации некорректностей

Алгоритм основан на исследовании взаимного расположения точки и отрезка и выявляет следующие ситуации:

- расстояние между точками менее ε (заданный допуск);
- расстояние между точкой и отрезком менее ε ;
- отрезки пересекаются.

Возможны также случаи, когда одной некорректности соответствуют два и более срабатывания алгоритма локализации, что приводит к появлению в выходных данных серии близко расположенных, либо одинаковых, точек. Таким образом, для дальнейшего применения необходимо агрегировать выходные данные, что значительно сократит их объём и, следовательно, количество локализаций некорректностей, которые предстоит обработать пользователю средствами ГИС, как это было описано выше. Агрегирование пространственных точек предполагается целесообразным провести на основе алгоритма кластеризации «К-медоид», описанного в [1], как наиболее подходящего для совершенствования выходных данных предложенного дополнения к существующей методике по следующим причинам [12-14]:

- по построению, алгоритм пытается отыскать «сферические» кластеры, что соответствует имеющемуся характеру распределения данных;

- меаноид менее чувствителен к посторонним значениям, то есть выбросам, что также удобно для существующих данных;

- центральные точки кластеров, называемые меаноидами, по определению принадлежат множеству исходных точек.

С учётом предлагаемого программного решения, не зависящего от выбора определённой ГИС, на этапе массового импорта данных из файлов открытых обменных форматов удастся избежать длительной процедуры ручного поиска некорректностей контуров площадных объектов.

Литература

1. Shashi Shekhar, Sanjay Chawla Spatial databases - a tour. Upper Saddle River, NJ : Prentice Hall, 2003. 262 p.

2. Трухачев М.Ю. Массовый импорт данных из форматов SHP и MIF в ГИС ObjectLand // Южный градостроительный центр URL: urgc.info/?p=459 (дата обращения: 01.02.2020).

3. Трухачев М.Ю. Исправление ошибок геометрии в площадных MIF-файлах // Южный градостроительный центр URL: urgc.info/?p=437 (дата обращения: 01.02.2020).

4. Трухачев М.Ю. Исправление ошибок геометрии в площадных SHP-файлах // Южный градостроительный центр URL: urgc.info/?p=440 (дата обращения: 01.02.2020).

5. Розенберг И.Н., Святков Д.С., Гитис С.А. Геоинформационная система Objectland // Известия Южного федерального университета. Технические науки. 2000. Т. 16. № 2. С. 361–362.

6. Проверка топологии с помощью правил топологии // ArcGIS Desktop URL: desktop.arcgis.com/ru/arcmap/latest/extensions/data-reviewer/validating-a-topology-against-topology-rules.htm (дата обращения: 01.02.2020).



7. Федорова Н. Модуль проверки пространственных данных на наличие топологических ошибок // ИТП Град URL: itpgrad.ru/node/1575 (дата обращения: 01.02.2020).
 8. Simashkov Как в MapInfo инспектировать объекты на топологическую корректность // Лаборатория АгроГИС-технологий г. Калуга URL: nrk-kaluga.ru/TopologyMI.htm (дата обращения: 01.02.2020).
 9. Руководство пользователя. Импорт и экспорт данных // ГИС ObjectLand URL: objectland.ru/support/doc/usermanual/MIMX.html (дата обращения: 01.02.2020).
 10. Использование структурированных исключений обработки в Visual Basic .NET или Visual Basic 2005 // Служба поддержки Майкрософт URL: support.microsoft.com/ru-ru/help/315965/how-to-use-structured-exception-handling-in-visual-basic-net-or-in-vis (дата обращения: 01.02.2020).
 11. Руководство пользователя. Преобразование площадного объекта в линейный и линейного в площадной // ГИС ObjectLand URL: objectland.ru/support/doc/usermanual/MEDC_ObjLine2Area.html (дата обращения: 01.02.2020).
 12. Kaufman L., Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, 2009. 342 p.
 13. Куринных Д.Ю., Айдинян А.Р., Цветкова О.Л. Подход к кластеризации угроз информационной безопасности предприятий // Инженерный вестник Дона. 2018. №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4803 (дата обращения: 01.02.2020).
 14. Гинис Л.А., Давыденко О.В. Применение когнитивного теоретико-множественного подхода к задаче определения кадастровой стоимости земель // Инженерный вестник Дона. 2019. №7. URL: ivdon.ru/ru/magazine/archive/N7y2019/6110 (дата обращения: 01.02.2020).
-

References

1. Shashi Shekhar, Sanjay Chawla Spatial databases - a tour. Upper Saddle River, NJ: Prentice Hall, 2003. 262 p.
 2. Trukhachev M.Yu. Massovyy import dannykh iz formatov SHP i MIF v GIS ObjectLand [Batch import of SHP and MIF data files into GIS ObjectLand] Yuzhnyy gradostroitel'nyy tsentr URL: urgc.info/?p=459.
 3. Trukhachev M.Yu. Ispravlenie oshibok geometrii v ploshchadnykh MIF-faylakh [A way to get rid of some errors in area topologies in the MIF data files] Yuzhnyy gradostroitel'nyy tsentr URL: urgc.info/?p=437.
 4. Trukhachev M.Yu. Ispravlenie oshibok geometrii v ploshchadnykh SHP-faylakh [A way to get rid of some errors in area topologies in the SHP data files] Yuzhnyy gradostroitel'nyy tsentr URL: urgc.info/?p=440.
 5. Rozenberg I.N., Svyatov D.S., Gitis S.A. Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki. 2000. V. 16. № 2. pp. 361–362.
 6. Proverka topologii s pomoshch'yu pravil topologii [Topology checking with the assistance of topology rules]. ArcGIS Desktop. URL: desktop.arcgis.com/ru/arcmap/latest/extensions/data-reviewer/validating-a-topology-against-topology-rules.htm.
 7. Fedorova N. Modul' proverki prostranstvennykh dannykh na nalichie topologicheskikh oshibok [Program solution for checking the spatial data with topological errors]. ITP Grad URL: itpgrad.ru/node/1575.
 8. Simashkov Kak v MapInfo inspektirovat' ob"ekty na topologicheskuyu korrektnost' [On inspecting the objects for topological correctness in MapInfo]. Laboratoriya AgroGIS-tekhnologiy g. Kaluga. URL: npk-kaluga.ru/TopologyMI.htm.
 9. Rukovodstvo pol'zovatelya. Import i eksport dannykh [User manual. Data import and export] GIS ObjectLand URL: objectland.ru/support/doc/usermanual/MIMX.html.
-



10. Ispol'zovanie strukturirovannykh isklyucheniy obrabotki v Visual Basic .NET ili Visual Basic 2005. [How to use structured exception handling in Visual Basic .NET or in Visual Basic 2005] Sluzhba podderzhki Microsoft URL: support.microsoft.com/ru-ru/help/315965/how-to-use-structured-exception-handling-in-visual-basic-net-or-in-vis.
11. Rukovodstvo pol'zatelya. Preobrazovanie ploshchadnogo ob"ekta v lineynyy i lineynogo v ploshchadnoy [User manual. Transformation of areal object into linear one and vice-versa] GIS ObjectLand URL: objectland.ru/support/doc/usermanual/MEDC_ObjLine2Area.html.
12. Kaufman L., Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, 2009. 342 p.
13. Kurinnykh D.Yu., Aydiyanyan A.R., Tsvetkova O.L. Inzhenernyj vestnik Dona, 2018, №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4803.
14. Ginis L.A., Davydenko O.V. Inzhenernyj vestnik Dona, 2019, №7. URL: ivdon.ru/ru/magazine/archive/N7y2019/6110.