

Algorytmy geoprzetwarzania

Algorytm geoprzetwarzania

Algorytm, który przetwarza **dane geoprzestrzenne**. Oznacza to że wejściem do takiego algorytmu są dane geoprzestrzenne (wektor, raster) lub informacja przestrzenna podana w inny sposób, a wyjściem jest warstwa przestrzenna lub informacja powiązana z warstwą przestrzenną.

Informacja geoprzestrzenna to informacja ściśle powiązana z lokalizacją na Ziemi (lub na innym systemem planetarnym).

Warstwa geoprzestrzenna to zbiór danych rastrowych, wektorowych lub innych, o zdefiniowanym układzie odniesienia przestrzennego oraz zasięgu w ramach tego układu.

Zasięg przestrzenny

Niezależnie od sposobu reprezentacji danych geoprzestrzennych (wektorowy lub rastrowy) proces przetwarzania danych wymaga podania:

- zasięgu przestrzennego operacji;
- układu odniesienia.

W przypadku danych wektorowych, informacja o położeniu obiektu jest zapisana w geometrii obiektu; każdy punkt posiada współrzędne geograficzne (długość i szerokość) lub geodezyjne (x,y). Dane atrybutowe przechowują informację nieprzestrzenną dla każdego obiektu, który posiada unikalny identyfikator.

W przypadku danych rastrowych, informacja nieprzestrzenna o zbiorze danych obejmuje rozmiar zbioru wyrażony w ilości wierszy i komórek, podobnie jak w przypadku obrazu, natomiast informacja geoprzestrzenna dodaje do tego zbioru.

- zasięg przestrzenny krawędzi rastra
- informację o rozdzielczości.

Rozdzielczość jest niezbędna aby określić położenie każdej komórki. Komórki rastra reprezentują pewien obszar na powierzchni Ziemi, a wartości rastra, dla jednego lub więcej pasm są odpowiednikiem atrybutów w modelu wektorowym.

Źródło informacji geoprzestrzennej

Najczęściej dane wejściowe algorytmu georzetwarzania to warstwa geoprzestrzenna, która zawiera informację o zasięgu, układzie odniesienia i, w przypadku rastra o rozdzielczości.

Jeżeli algorytm nie wymaga warstwy geoprzestrzennej jako wejścia - na przykład generator losowej powierzchni, podstawowe informacje, takie jak, jak zasięg, układ odniesienia oraz rozdzielczość (w przypadku rastra) są podawane jako parametry wejściowe.

Wiele systemów geoprzetwarzania (Esri, SAGA, GRASS) posiada zestaw parametrów definiujący środowisko pracy, określające wszystkie wymienione wcześniej parametry.

Ten zestaw paramerów nazywa się odpowiednio:

- środowisko (ESRI)
- grid system (SAGA)
- region (GRASS)

[DODAC ILUSTRACJE]

Relacje pomiędzy rozdzielczością, zasięgiem przestrzennym a wielkością rastra

W przypadku danych rastrowych, zdefiniowanie zasięgu przestrzennego wymaga podania 6 wartości. Mogą to być:

- granice N,S,W,E + rozdzielczość NS, WE. W takiej sytuacji, liczba wierszy i kolumn jest wyliczana z zasięgu i rozdzielczości. jest to podejście najtrudniejsze, gdyż liczba wierszy i kolumn musi być liczbą całkowitą. Oznacza to z reguły konieczność modyfikacji albo zasięgu albo rozdzielczości.
- granice N,S,W,E + liczba wierszy i kolumn. W takiej sytuacji rozdzielczość jest wyliczana poprzez podzielenie zasięgu przez liczbę komórek w wierszach i kolumnach. Rozwiązanie stosowane w QGis, jeżeli wejściem nie jest warstwa geoprzestrzenna. W praktyce oznacza to, że rozdzielczości nie są wartościami okrągłymi.
- Współrzędne N i W + rozdzielczości NS i WE + liczba wierszy i kolumn. Takie rozwiązanie jest najbezpieczniejsze, gdyż nie modyfikuje żadnych parametrów geoprzestrzennych warstwy. Jest stosowane w GDAL

Informacja o współrzędnych N i W oraz rozdzielczości NS i WE określane są terminem geotransformat. Liczba wierszy i kolumn to właściwość obrazu.

[DODAC ILUSTRACJE]

wejście i wyjście

Dane wejściowe to najczęściej jedna lub więcej warstw geoprzestrzennych. Algorytmy akceptują również dane geoprzestrzenne zapisane w innych formatach, na przykład listy punktów, lub macierzy w takiej sytuacji informacja geoprzestrzenna (projekcja, zasięg, rozdzielczość) musi być dodana jawnie lub jest pobierana ze środowiska.

Wyjściem operacji geoprzestrzennych jest również jedna lub więcej warstw geoprzestrzennych. Warstwy te z reguły są zgodne z układem odniesienia, zasięgiem i rozdzielczością warstw wejściowych, za wyjątkiem wybranych algorytmów, takich jak reprojektacja. W specyficznych sytuacjach informacja wyjściowa zapisywana jest w postaci tabeli, raportu tekstowego lub innej, nieprzestrzennej formie. Wyjściem algorytmu geoprzetwarzania może być również obraz pozbawiony atrybutów geoprzestrzennych lub wykres.

W przypadku, jeżeli dane wyjściowe mają charakter nieprzestrzenny, ich użycie jako wejścia do kolejnych algorytmów w skrypcie geoprzetwarzania może być trudne lub niemożliwe

Parametry algorytmów geoprzetwarzania

Najważniejsze parametry nieprzestrzenne, jakie można przekazać do algorytmów geoprzetwarzania

- wartości lub lista wartości lub zakres (dwie liczby) liczba
- tekst lub lista łańcuchów tekstowych
- opcja - nazwa, ale w praktyce liczba całkowita (ENUM)
- flaga - wartość logiczna tak/nie
- pole lub lista pól (nazwa)
- wyrażenie na temat pól
- informacja geoprzestrzenna lub jej elementy
- kolor, data-czas, czas trwania itp

ENUM(erator) to specjalny typ zmiennej, która kolejnym symbolom, z reguły łańcuchom tekstu, przypisuje kolejne wartości całkowite. ma charakter pomocniczy, w sytuacji gdy algorytm jako opcje przyjmuje kolejne liczby całkowite, co nie jest czytelne.

Status danych wyjściowych

Dane geoprzestrzenne z reguły zajmują dużo przestrzeni dyskowej. Rozmiar indywidualnych zbiorów danych natomiast często przekracza pojemność pamięci operacyjnej. Z tego powodu zarządzanie statusem danych wyjściowych jest ważnym elementem skuteczności podejmowanych działań.

Można wyróżnić 5 głównych statusów danych wyjściowych (nowych zbiorów)

- zbiór danych będący wynikiem docelowym
- zbiór danych będący wynikiem pośrednim, który powinien być zachowany
- zbiór danych będący wynikiem pośrednim, który nie musi być zachowany
- zbiór danych będący kopią danych wejściowych, do którego dodano dodatkową informację (z reguły dane wektorowe)
- zbiór danych będący zmodyfikowaną wersją danych wejściowych (np. usunięcie artefaktów)

Nadpisywanie danych w geoprzetwarzaniu jest ryzykowne. Nie ma możliwości cofnięcia!

Polityki zachowania danych wyjściowych

- pozostawianie każdego wyniku danych wyjściowych:
 - jest rozwiązaniem bezpiecznym ale...
 - prowadzi do szybkiej utraty przestrzeni dyskowej,
 - prowadzi do chaosu w przestrzeni nazw
- modyfikacja istniejącego zbioru
 - brak możliwości cofnięcia
 - lepsze zarządzanie nazwami
- weryfikacja wyników ex-post
 - zakładamy, że wyniki i stany pośrednie są przeglądane i usuwane po zakończeniu pracy
 - dodatkowa prac
- zmienne tymczasowe
 - tylko skrypty geoprzetwarzania
 - czas życia zbioru kontrolowany przez algorytm

Geoprzetwarzanie

Procedura uruchamiania algorytmów geoprzestrzennych (narzędzi), wykonywanych w określonych ramach (framework). Główne zadania geoprzetwarzania to:

- zarządzanie danymi geoprzestrzennymi
- analiza danych geoprzestrzennych
- wydobywanie informacji przestrzennej
- konserwacja danych geoprzestrzennych
- tworzenie i aktualizacja danych przestrzennych
- dystrybucja danych geoprzestrzennych

Najważniejsze operacje geoprocessingu

- łączenie i przycinanie warstw (integracja i ekstrakcja)
- operacje przestrzenne (logiczne) na obiektach (przecięcie, suma, różnica, różnica rozłączna)
- selekcja i ekstrakcja obiektów na podstawie relacji przestrzennych (predykaty przestrzenne - zawiera, równa się, styka z itp.)
- selekcja i ekstrakcja obiektów na podstawie atrybutów i własności danych
- algebra rastrów
- analizy sąsiedztwa i odległości (bufor, analiza gęstości itp.)
- generalizacja i integracja danych z różnych źródeł
- konserwacja, czyszczenie i naprawa danych
- konwersja danych pomiędzy formatami i typami
- zmiana odwzorowania
- algorytmy specjalistyczne

Skrypty geoprzetwarzania

Skrypt geoprzetwarzania to skrypt danego środowiska (systemu operacyjnego, aplikacji), który uruchamia narzędzia (programy, algorytmy) geoprzetwarzania i używa jako wyjścia oraz i/lub wejścia danych geoprzestrzennych.

Skrypt uruchamiany w systemie operacyjnym, może być napisany w dowolnym języku programowania. Skrypty tworzone wewnątrz aplikacji GIS tworzone są w dedykowanym języku i korzystają z API (Application Programming Interface) danej aplikacji

Celem skryptów jest automatyzacja procesu geoprzetwarzania

Do czego używamy skryptów? [ZMODYFIKOWAC]

automatyzacja sekwencji działań

Sekwencja kolejnych poleceń. Wyjście wcześniejszych algorytmów staje się wejściem dla następnych.

iteracyjne przetwarzania dużych zbiorów danych

Zbiorem danych może być baza danych geoprzestrzennych, katalog, lub pojedynczy plik. Procedury iteracyjne zakładają również filtrowanie danych (na przykład pomijanie małych obiektów) lub odmienne ścieżki przetwarzania w zależności od analizowanych parametrów

dobór parametrów algorytmów geoprzetwarzania na podstawie danych

Stosowane w sytuacjach, gdy poszczególne obiekty są przetwarzane z innymi parametrami (na przykład dobór rozdzielczości do wielkości obiektu, dobór parametrów interpolacji itp.)

symulacje i optymalizacje

Złożone procedury geoprzetwarzania, których parametry kolejnych kroków zmieniają się na podstawie wyników poprzednich kroków

zarządzanie danymi geoprzestrzennymi

Reprojekcja

Wizualizacja i publikowanie danych

Polecenia wybranych aplikacji

Uruchamianie algorytmów jako indywidualnych programów (poziom systemu)

GRASS GIS

Grass wymaga importu i eksportu danych do bazy danych

```
r.import dem.tif output=dem
r.slope.aspect -n elevation=dem slope=slope
r.mapcalc 'inclined=slope>2?1:0'
r.export inclined output=inclined.tif
```

GDAL

GDAL pracuje na dowolnych plikach danych geoprzestrzennych

```
gdaldem slope dem.tif slope.tif -compute_edges
gdal_calc.py --calc='A>21:0' -A slope.tif --outfile=inclined.tif
```

Uruchamianie algorytmów przy pomocy wykonawców (executors) (poziom aplikacji)

SAGA GIS

Wykonawca sagi `saga_cmd` wymaga podania modułu, z którego zostanie pobrany algorytm (numer).

```
saga_cmd ta_morphometry 23 -DEM dem.tif -SLOPE slope -CONSTRAIN 1
saga_cmd grid_calculus 1 -FORMULA "ifelse(g1>100,1,0)" -GRIDS slope.dat -TYPE 0 -RESULT
inclined
saga_cmd io_gdal 1 -GRIDS inclined.dat -FILE inclined -EXTENSION tif -FORMAT 9
```

wykonawca QGis `qgis_process` wymaga nazwy algorytmu. Parametry są podwane w sposób nietypowy

Qgis

```
qgis_process run qgis:slope -- INPUT=dem.tif -- OUTPUT=slope.tif
qgis_process run qgis:rastercalculator -- FORMULA '"slope@1">1' -- LAYERS=slope.tif --
OUTPUT=inclined.tif
```

QGIS uruchamianie algorytmów z parametrami w formacie JSON

```
echo '{"inputs': {'INPUT': 'my_shape.shp', 'DISTANCE': 5}}' | qgis_process run  
native:buffer -
```

JSON - Java Script Object Notation: alternatywny względem XML sposób obsługi danych semistrukuralnych. Struktura JSON przypomina znany ze słownika Pythona model *key:value*, gdzie *value* może być kolejnym zagnieżdżonym słownikiem zawierającym własne klucze i wartości

`echo` komenda wysyłająca tekst na standardowe wyjście; `|` operator potokowania, przyjmujący strumień danych i kierujący go do standardowego wejścia

uruchamianie algorytmów wewnątrz aplikacji - GRASS

Tworzenie skryptów geoprzetwarzania w środowisku aplikacji wymaga udostępnienia przez aplikację środowiska programowania oraz sposobu wywoływania algorytmów.

Obecnie takie możliwości są udostępnione przez środowisko ArcGIS Pro w postaci biblioteki ArcPy, pozwalającej na wywoływanie algorytmów geoprzetwarzania jako funkcji lub metod języka Python.

W środowisku GRASS GIS mechanizm tworzenia skryptów z użyciem gotowych algorytmów geoprzetwarzania wykorzystuje procedury uruchamiania algorytmów w środowisku systemowym.

```
grass.run_command('r.slope.aspect',elevation='dem',slope='slope')
params=['r.slope.aspect','elevation=dem','slope=slope']
params.join(" ") > ...
```

Więcej na osobnym wykładzie

Uruchamianie algorytmów wewnątrz aplikacji - GRASS

Wbudowane algorytmy QGIS, oraz te, które są dostarczane przez zarejestrowane providery uruchamiane są metodą `processing.run()`:

```
params = {'INPUT': 'dem.tif',  
          'OUTPUT': 'slope.tif'}  
  
processing.run('qgis:slope', params)
```

algorytm i słownik `params` zamieniany jest na obiekt JSON i przekazywane do egzekutora w środowisku systemowym

Przechowywanie wyników

W procesie geoprzetwarzania wyniki mogą być przechowywane jako zbiory danych na dysku lub zbiory danych w pamięci operacyjnej. Zbiory dyskowe mogą mieć status tymczasowych, czyli, albo przechowywanych w katalogach tymczasowych, albo utrzymywanych tylko w czasie sesji aplikacji.

Polityka przechowywania warstw tymczasowych zależy od ich potencjalnej wielkości. W przypadku zbiorów rastrowych, rozmiar realnie przetwarzanych danych - na przykład scena satelity Sentinel 2, przekracza dostępną pamięć większości komputerów osobistych. Z tego powodu tymczasowe warstwy rastrowe są przechowywane jako tymczasowe zbiory dyskowe

Warstwy tymczasowe w pamięci są przechowywane w specjalnym formacie `OGR:memory` lub jego rastrowym odpowiedniku `GDAL:MEM`.

W skryptach geoprzetwarzania dane pośrednie często mają charakter tymczasowy (w pamięci lub jako zbiory tymczasowe)

Formaty uniwersalne a formaty natywne

Pojęcie formatu uniwersalnego w GIS jest nieostre. Formatami uniwersalnymi są `.gpkg` oraz `.tif`, ze względu na to że nie są powiązane z żadnym środowiskiem GIS. Podobną funkcję pełnią formaty tekstowe: `.csv` czy ASCII GRID (`.asc`) oraz formaty wymiany danych, jak GML czy GeoJSON.

Formaty natywne to takie, które do których zapisują wyniki swoich prac konkretne systemy geoinformacyjne.

GDAL daje dostęp w trybie odczytu do właściwie wszystkich natywnych typów danych geoprzestrzennych oraz do większości w trybie zapisu.

- **QGIS i GDAL** - nie posiadają własnych formatów przechowywania danych, korzysta domyślnie z `.gpkg` oraz `.tif`.
- **SAGA** - posiada własny format plików, obejmujących kilka osobnych zbiorów (`*dat`, `*sgrd`). Może odczytywać pliki obsługiwane przez GDAL, ale zapis do plików wymaga eksportu. Formaty SAGA są odczytywane przez GDAL (czyli również przez QGIS)
- **GRASS** - posiada własną rozbudowaną bazę danych do której dostęp spoza systemu jest dość trudny (głównie z powodu archaiczności środowiska)

Co jeśli OUTPUT nie jest geoprzestrzenny

Niektóre algorytmy i skrypty mogą w wyniku zwracać obiekt nieprzestrzenny mogą to być:

- zmienna lub kilka zmiennych: powinny być zwracane w formie, wartości (jedna), krotki (wiele, tego samego typu) lub słownika (wiele różnego typu)
- tabela lub tablica: zwracana jako DataFrame albo macierz, docelowo w formie pliku tekstowego lub arkusza kalkulacyjnego
- grafika (plik pozbawiony informacji przestrzennej, wykres), w formie obrazu jpg, lub png
- raport, najczęściej w formie pliku tekstowego, html lub pdf

zmienne oraz tabele mogą stanowić INPUT do kolejnych składowych algorytmu geoprzetwarzania