

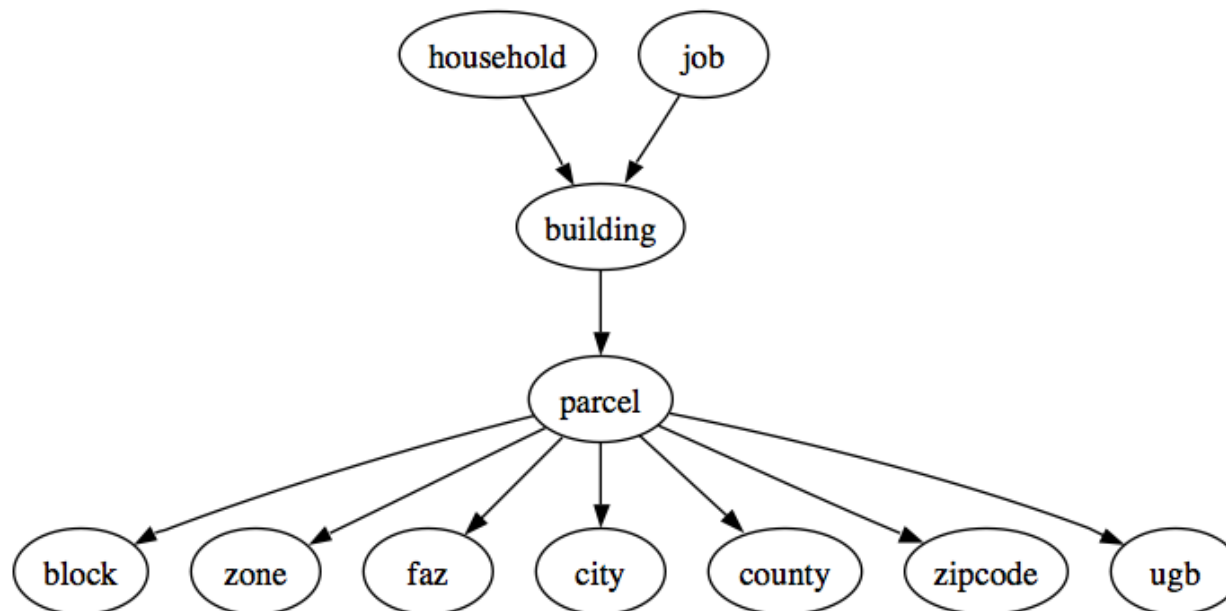
Prepare Data for UrbanSim Application

Liming Wang
05/25/2010

Data Requirement

- ◆ Parcel/Building Data
- ◆ Business Establishments
- ◆ Household Data (Census, Survey)
- ◆ Regional Control Totals
- ◆ Land Use Plan
- ◆ Environmental Constraints
- ◆ Transportation Data

Relationship of datasets



Parcels/Buildings

- Sources: Assessor's, real estate sale transactions, CoStar, etc
- Key variables:
 - Residential/non_residential spaces; “built form”; size/area of geography
 - Price
 - Mapping to various geography
 - Development constraints, zoning
 - Physical and environmental attributes

Households/Persons Data

1. households_for_estimation ← a subset of households of which we have better information, for use in estimating models. Usually from household activity/travel survey.
2. households/persons ← Household/Population Synthesizer
3. Assign synthesized households from synthesized geography to model geography. Take assignment from blockgroup to building as example

Household Assignment (1)

Monte Carlo Assignment

Randomly assign a household to a building with known blockgroup, and possibly, building type:

- alter table households
add household_id int auto_increment key,
add building_id int default -1;
- create table bg_summary
select blockgroup_id, sum(residential_units) as
total_units
from buildings group by blockgroup_id;
- create table fractions

Household Assignment (2)

- Cache households and fractions:
`python opus_core/tools/do_export_sql_to_cache.py -d durham_zone_baseyear -c /data/durham_zone/hh_prep -y 2000 --database_configuration=urbansim2`
- Run Monte Carlo assignment:
`python urbansim/tools/do_monte_carlo_assignment.py -c /data/durham_zone/hh_prep/2000 -i households -f fractions --id-name1=blockgroup_id --id-name2=building_id --fraction-attribute-name=fraction`

Household Assignment (3)

Household Location Choice Model

Run HLCCM to assign households of each synthesis geography:

- estimate HLCCM with data in `households_for_estimation`, possibly include demographic variables of synthesis geography (`blockgroup`)
- cache households, `household_location_choice_model_specification/coefficients`

Household Assignment (4)

- Modify `psrc_parcel/data_preparation/config_households.py`
- Run HLCCM for assignment:
`python opus_core/tools/start_run.py -c psrc_parcel.data_preparation.config_households`

Refer to `psrc_parcel/data_preparation/README.txt` for details.

Employment Data (1)

- Sources: ES202, Dun & Bradstreet, Census Longitudinal Employer Dynamics (LED) etc
- Business/establishment OR individual job
- Unroll from business/establishment to jobs
 - Python script: `psrc_parcel/data_preparation/unroll_jobs_from_establishments.py`
 - sql script (mysql):

Employment Data (2)

```
select max(employees) + 1 into @max_emp from  
businesses;
```

```
# create a temporary tables having @max_emp rows
```

```
create table temp_seq_ids
```

```
select business_id as dummy
```

```
from businesses;
```

```
alter table temp_seq_ids add seq_id int auto_increment  
key;
```

```
delete from temp_seq_ids where seq_id > @max_emp+1;
```

```
alter table businesses add index(employees);
```

Employment Data (2)

```
create table jobs
select seq_id as job_seq_id, #unique indentifer within
    employer
business_id, empid, sector_id, zone_id
from temp_seq_ids j, businesses c
where j.seq_id <= c.employees;
```

```
alter table jobs add job_id int auto_increment primary key
    first;
```

```
drop table temp_seq_ids
```

Employment Data (3)

- Assign business/establishment/jobs to model geography, e.g. building
 - Unrolling process handles most of them if the source data is geo-coded
 - Monte Carlo Assignment
 - Run ELCM

Transportation Data

- OD matrices flattened to a 1d array (column)

Column Name	Data Type	Description
from_zone_id	Integer	origin zone
to_zone_id	Integer	destination zone
am_single_vehicle_to_work_travel_time	Integer	zone-to-zone morning peak period travel time for vehicles traveling in mixed flow lanes

Data Management

- Ways of managing data: DBMS v.s. delimited files
- Get data into and out of opus/UrbanSim: <http://urbansim.org/Documentation/GetDataIntoandOutofOPUS>
- Log data changes: change db chain with scenario_information and change_log table (<http://urbansim.org/downloads/manual/dev-version/opus-userguide/node108.html>)