ActivitySim Memory Memo Update

Sijia Wang, WSP

August 18th, 2022
ActivitySim Phase 7 Meeting

1. Introduction

Memory Problems ActivitySim Users Are Experiencing

Anything We Missed?

- ActivitySim models cannot run on machines even with significant memory resources (e.g. 128 GB).
- Chunking is hard to set up or not working.

Memory Problems

Outline Of The Memo

1. Explore possible causes to the high memory footprint;
2. Compare to other ABMs' software engineering approaches; and
3. Offer suggestions on fixes

2. Possible Causes

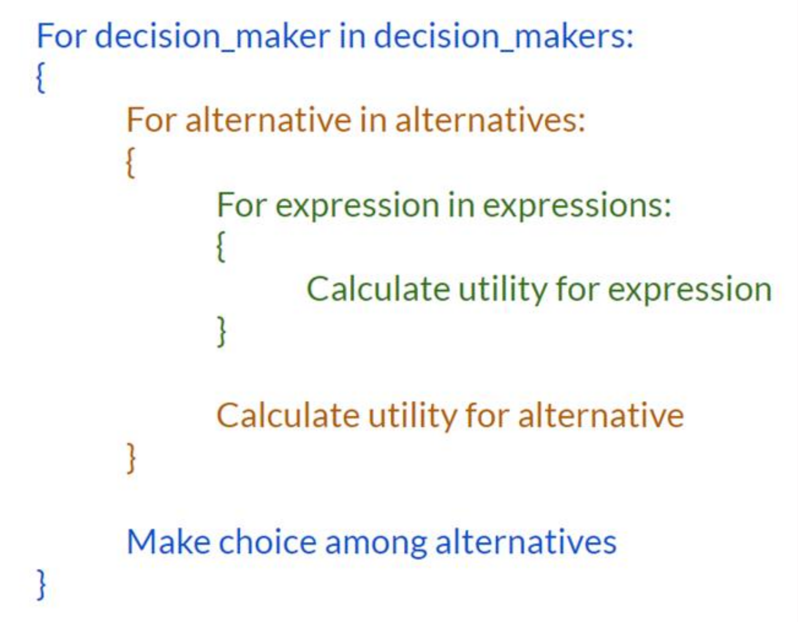
We explored the following feature designs of ActivitySim, and compared them with other ABMs

1. Choice model simulation
2. Data model
3. Data type

1. Choice Model Simulation

- CT-RAMP for-loop
- ActivitySim vector math

CT-RAMP For-Loop

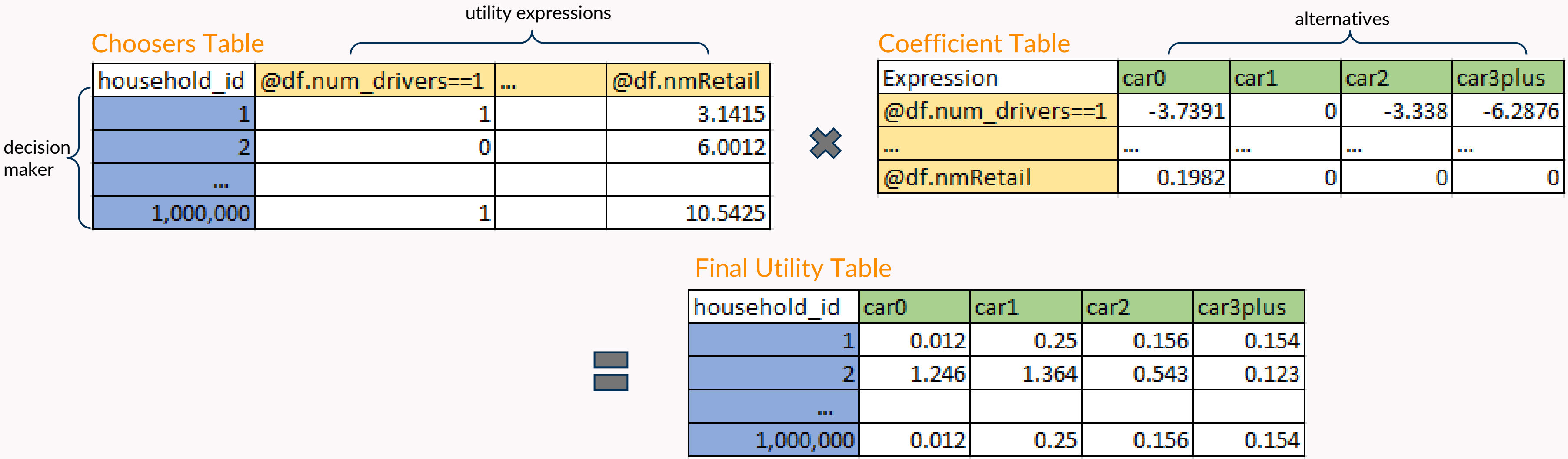


```
For decision_maker in decision_makers:
{
        For alternative in alternatives:
        {
                For expression in expressions:
                {
                        Calculate utility for expression
                }

                Calculate utility for alternative
        }

        Make choice among alternatives
}
```

- Has small memory footprint
- Needs additional software engineering to be computationally efficient
  - Multi-threading or distributed computing
  - Stand-alone data servers


utility expressions
alternatives
decision maker

Choosers Table

household_id	@df.num_drivers==1	...	@df.nmRetail
1	1		3.1415
2	0		6.0012
...			
1,000,000	1		10.5425

×

Coefficient Table

Expression	car0	car1	car2	car3plus
@df.num_drivers==1	-3.7391	0	-3.338	-6.2876
...	...	...	...	...
@df.nmRetail	0.1982	0	0	0

=

Final Utility Table

household_id	car0	car1	car2	car3plus
1	0.012	0.25	0.156	0.154
2	1.246	1.364	0.543	0.123
...				
1,000,000	0.012	0.25	0.156	0.154

- Takes advantage of modern Python libraries like Pandas
- Synchronous computing, but has large memory footprint
- Requires chunking for batch-processing

ActivitySim Vector Math

2. Data Model

- ActivitySim pipeline
- CT-RAMP data servers

- A data pipeline that
  - Runs a series of model steps
  - Manages data tables throughout the model run
  - Allows for restarting at any model step via checkpoints
- Checkpoints are saved at every pipeline step
  - Users can turn checkpoints off to save run time
- Checkpoints copy and save out complete data tables - lots of duplication!
  - Update: This contributes more to a storage problem than a memory problem
- Synchronous computing

ActivitySim Pipeline

- Household server
  - Stand-alone central hub for households and persons
  - CT-RAMP 2S - Every model result is an “Attribute” object added to the existing data
    - Vehicles are set as an “Attribute” to Households
    - Tours are set as an “Attribute” to Persons
    - “Attribute” will only be created for relevant decision-makers - saves memory
  - Checkpoints are saved at every model step without duplication
- Matrix server
  - Stand-alone central hub for matrices
  - CT-RAMP 1, 2 - Compressed matrix collection
  - CT-RAMP 2S - Does not preload skims. Loads skims on demand.
- Asynchronous computing
- Cloud computing

CT-RAMP Data Servers - CT-RAMP 2S as Example

3. Data Type

- ActivitySim data type
  - Data tables have lots of int64 - is it necessary?
  - Skims are float 64
- DaySim data type
  - Skims are 2-byte unsigned integers

3. Solutions to Consider

Ideal Solutions

- Convert ActivitySim to For-Loop
- Separate data tables and skims from pipeline - have a stand-alone household server and a matrix server
- Orchestrated cloud computing

This requires more fundamental code refactoring.

Practical Next Steps

- Improve chunking performance
- Audit ActivitySim data types
    - How much memory can we save by just changing data types?
- Monitor memory peaks in model runs
- Re-assess with Sharrow implementation

- TourCast
  - Data Model - Are there household and matrix servers?
  - Data Type - What types are used?
  - How is memory managed?
- DaySim
  - Data Model - Are there household and matrix servers?
- Jeff Doyle’s comment about changing data type in ActivitySim
- How many of the issues will be solved by Sharrow?

Items to follow up on