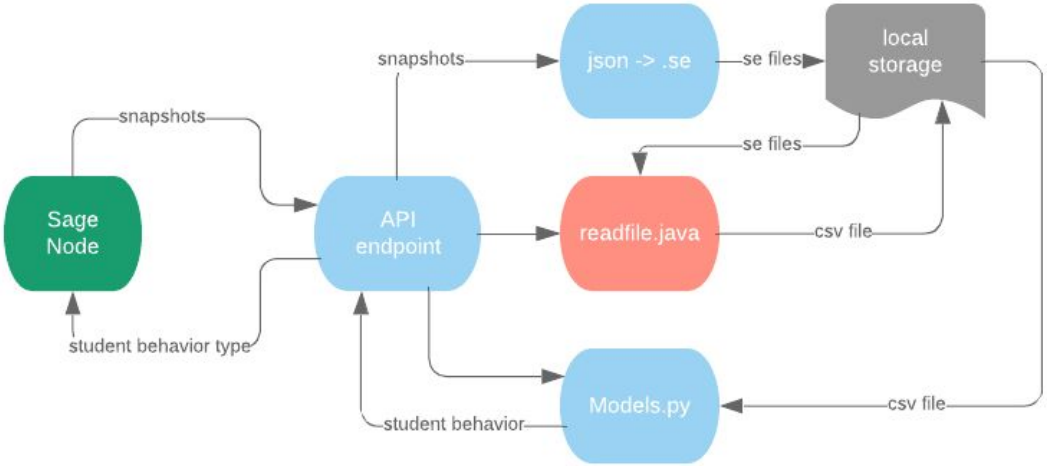
SAGE Behavior Detection

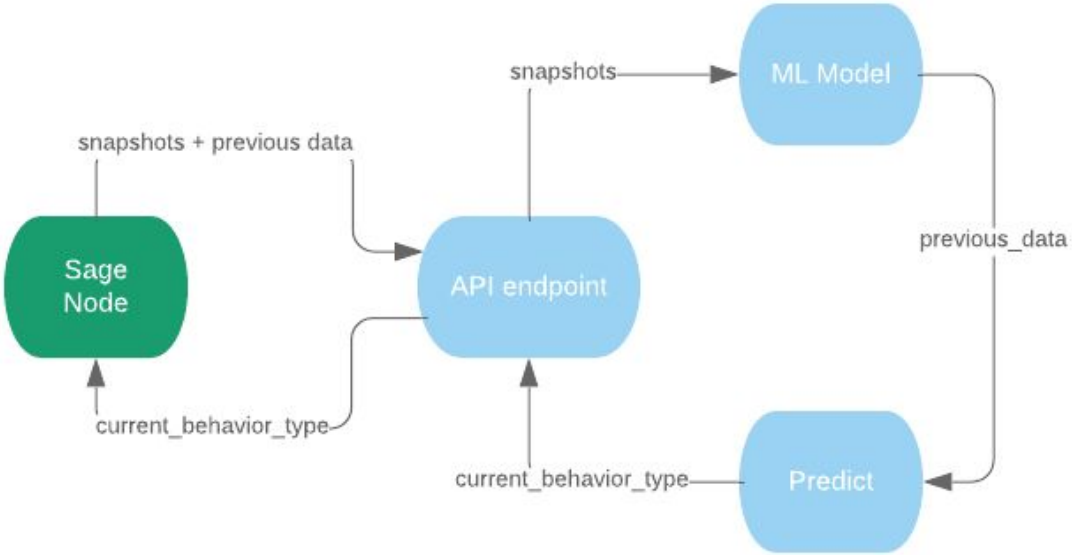
Vivian Han, Joe Huang

Previous Data Flow

- Combination of multiple modules
- Redundant data transformation
- Local storage



Simplified Data Flow



Changes in data format

```
{
   'prev_data':{
   ...
   },
   'cur_game': 0,
   'snapshots':{
   ...
   }
}
```

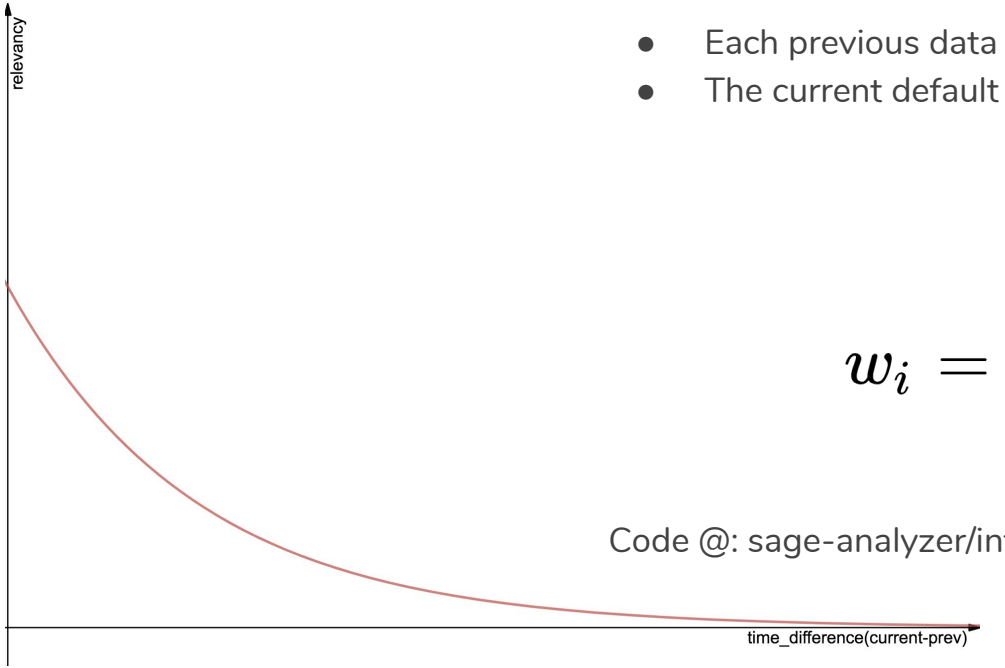
```
"snapshots": [
    {
        "timestamp": "18-40-28-GMT-1204-2018",
        "content": "<<Object Stage>>\n"
    },
    {
        "timestamp": "18-40-30-GMT-1204-2018",
        "content": "<<Object Stage>>\n\t\twhenGreenFlag\n"
    },
    {
        "timestamp": "18-40-31-GMT-1204-2018",
        "content": "<<Object Stage>>\n\t\twhenGreenFlag\n\t\tstartScene\n"
    },
```

Taking Previous Data into Consideration

```
"19-55-11-GMT-1207-2018": [
    0,
    1
],
"19-55-11-GMT-1205-2018": [
    3,
    6
],
```

```
timestamp: [
    behavior_type,
    game_type
]
```

Previous Behavior Type



- Each previous data weighted by time difference
- The current default cutoff time is set to 20 days

$$w_i = e^{\frac{ln(0.01) \times \text{time\_diff[i]}}{\text{cutoff}}}$$

Code @: sage-analyzer/intelligentHint/behaviordetection

Previous Game Type

- Basic Probability Model
- Smoothing

$$P(b|g) = \frac{P(b,g)+k}{P(g)+4k}$$

Behavior Type	Game Type	P(behavior|game)
0	3	0.05
1	3	0.75
...	...	...

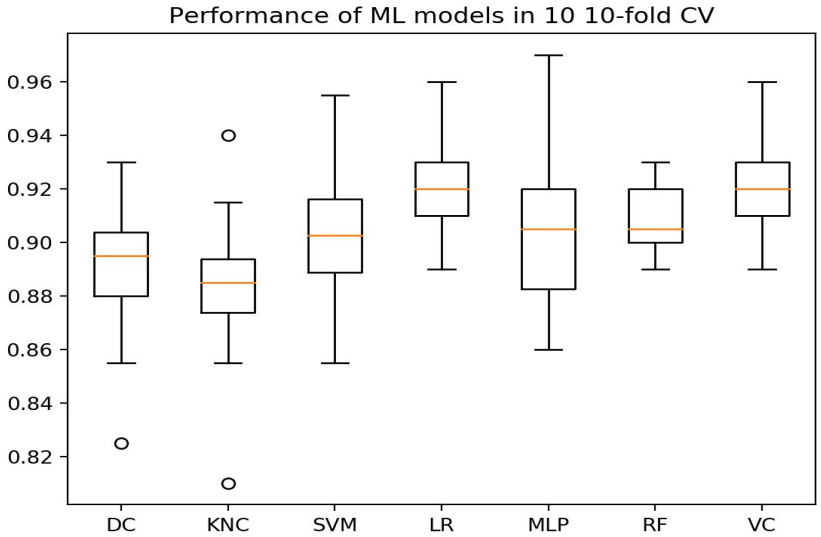
Code @: sage-analyzer/intelligentHint/behaviordetection

Models - stacking

Voting Model

- Decision Tree
- SVM
- K Nearest Classifier
- Logistic Regression
- Multilayer Perceptron
- Random forest

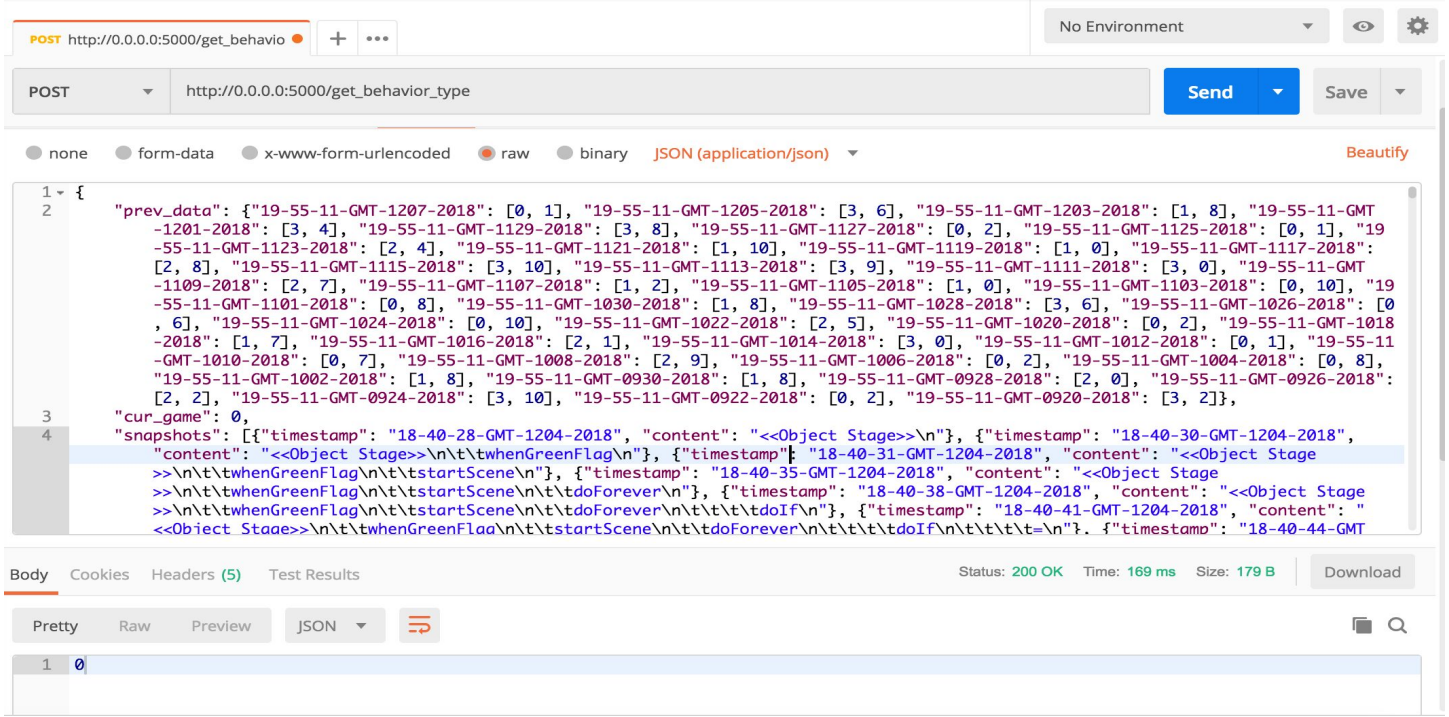
The model is saved locally as saved_model.pkl



Output the current behavior type

$$P_{model}(b)P(b|g)\sum_{b_i=b} w_i$$

API Demo (Postman)



Limitation & Future Work

- Improving ML Model Accuracy (currently ~90% w/ mock data)
  - Improving Probability Model
  - Adjusting Function Parameters
  - Unique ID for each block rather than just name
  - More variation in mock data