Simple Human-in-the-loop System for Modern Data Science Workflow

An-Dan Nguyen

Outline

1. Motivation
2. Model of Interactions
3. The action model If-this-then-that
4. Implementation
5. Experiments and Results
6. Conclusion and Future Works

Motivation

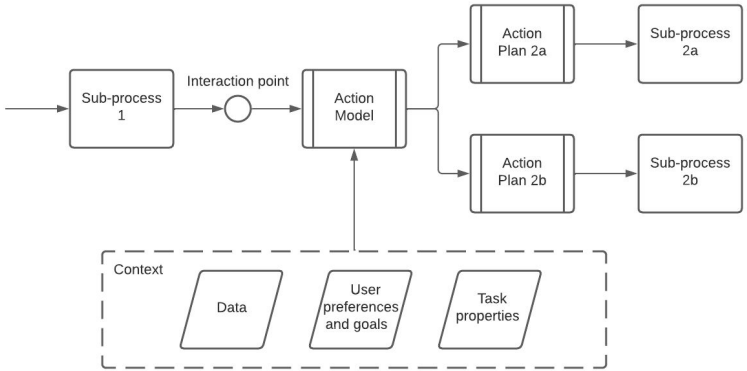
- The advances in computational power helps bring data science techniques into a wide range of fields.
- Each field has its distinctive workflow and usually need to incorporate human expert knowledge.
- Hence, there is a need to for the human to work with the modern data science workflow in a versatile and seamless way.
- Concepts of interactions between human and the system are derived.
- Based on these concepts a prototype has been developed and applied in two data scenarios: streaming data and moving data.

Model of Interactions

- There are three basic concepts in the model of interactions
  - Flow
  - Action Model
  - Communication Method
- Flow are the workflow of a process with **interaction points** where the human can interaction with the process. Flow is not a focus in this research
- Action Model is a predefined (or learned) steps of interactions the user can perform at each interaction points. Action Model is the focus of this research
- Communication method is the mean of which human and the system can communicate. For example, through a UI or APIs. In this work, a simple message passing system is use to facilitate the interaction.

Action Model and Action Plan

- Action Model is an abstract sequence of actions executed when an interaction point is reached in the flow.
- Action Model can be seen as a general template → A class in programming terminology
- Action Plan is a concrete realization of an Action Model where each actions is defined clearly → An object of a class in programming terminology.



Action Model *if-this-then-that*

- Is the first action model developed to demonstrate the presented model of interactions.
- Is versatile and can be used in any interaction points that need decision (if-clause) and the effect after that decision (then-clause)
- Is easy to implemented

```
{
    "if": <if-clause-function>,
    "if-clause-args": {
        <args1>: <args1-value>,
        <args2>: <args2-value>,
        ...
    },
    "then": <then-clause-function>,
    "then-clause-args": {
        <args1>: <args1-value>,
        <args2>: <args2-value>,
        ...
    }
}
```

Implementation

- Python is used.
- Code has been published on GitHub
- The action model is implemented in the module ifttt.py as the class IFTTT
- The action plan is a json file and can be load by the class IFTTT to form an IFTTT object.
- Then, the method *run()* of the class IFTTT execute the action plan.
- Details can be found in the Github repository.

Implementation (cont.)

```
{
  "redis_server": "",
  "redis_port": "",
  "condition_clauses": [
    {
      "condition_variable": "",
      "condition_operator":  "",
      "threshold_value": ""
    }
  ],
  "target_module_file_path": "",
  "target_module_name": "",
  "target_function_name": "",
  "target_function_params": {
    "dataset_path": "",
    "result_file_path": ""
  },
  "num_repeat": "",
  "num_processes": ""
}
```

Action model

```
{
  "redis_server": "localhost",
  "redis_port": 6379,
  "condition_clauses": [
    {
      "condition_variable": "moved",
      "condition_operator":  "eq",
      "threshold_value": 1
    }
  ],
  "target_module_file_path": "./plotter.py",
  "target_module_name": "plotter",
  "target_function_name": "moving_data_consumer",
  "target_function_params": {
    "dataset_path": "../../../data/move_des_data",
    "result_file_path": "../../../result/moving.png"
  },
  "num_repeat": 1,
  "num_processes": 1
}
```

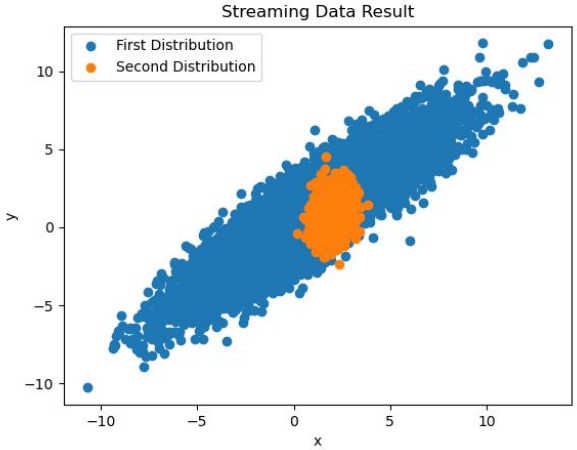
Action plan

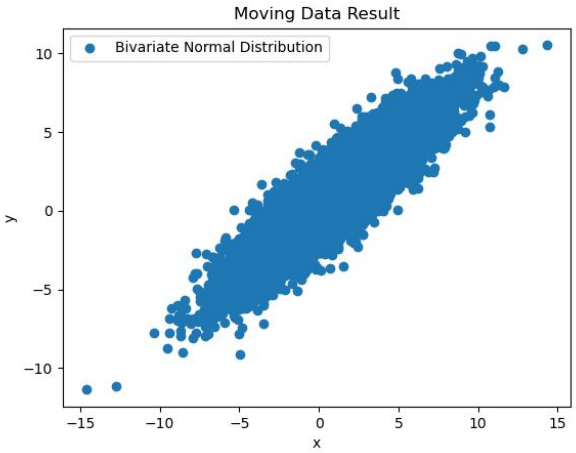
Experiment and Result

- Two data scenarios, streaming data and moving data, are used to test the system.
- Streaming data scenario represents the situations where volume and velocity of data is fast.
- While moving data scenario represents the cases where the completion of a data science task is uncertain. It can be due to unreliable network connection or APIs.

Experiment and Result (cont.)

- In both scenario, the interaction points are placed when the data have been accumulate enough to have a simple plot about the data.
- The action plans (from the action model *if-then-this-that*) are created to perform actions at those interaction plots to do the plotting.
- Steps to reproduce the results are detailed in the Github repository.





Conclusion and Future Works

- This project has derived the concepts of models of interactions and used those concepts to develop a simple system HITL to demonstrate the ideas.
- Major drawbacks of the works are:
  - The experiment is on hypothetical scenario
  - The variety of action models and the flexibility of current *if-this-then-that* action model is limited
- Those will be addressed in the upcoming iterations of the research.