Third Sprint

Streaming event data compliance checking in Python

Jingjing Huo	376323
Sabya Shaikh	384606
Zheqi Lyu	378653

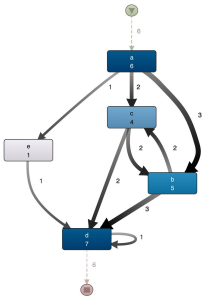
1 Test

1.1 Mutli-client simultaneously logging and compliance checking

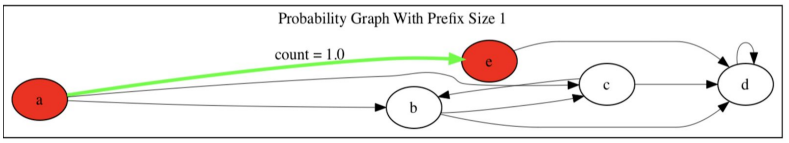
When someone is using uuid ‘cli’ to do compliance checking, then this uuid cannot be used again at the same time.

1.2 Correctness of compliance checking

1. Using the event log “data/Example_EventLogForTraining_Backup.xes” to do training, setting threshold as 0.3. Then using the same event log to do compliance checking.
Here I provide the screenshot using Disco (with 100% Activities and 100% Paths), from this we can see that only the threshold from ‘a’ to ‘e’ is around 0.167 which lower than 0.3. So our system should also only give one alert of type T for the connection of a to e.

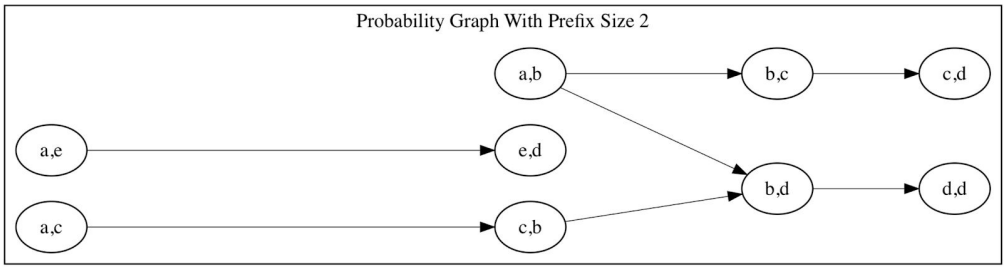


```
[13:13:4:331684][ INFO  ] : ----------------start to do compliance checking, please wait----------------
[13:13:4:388396][Alert T] :  The threshold of the connection in case 'Case3.0' is too low.
   The minimal expected probability from  a --> e :  0.3
              The true probability from  a --> e :  0.166667
[13:13:4:474557][ INFO  ] : ----------------the compliance checking is finishing----------------
```



2. The event log from data/Example_EventLogForTraining_Backup.xes is follows:

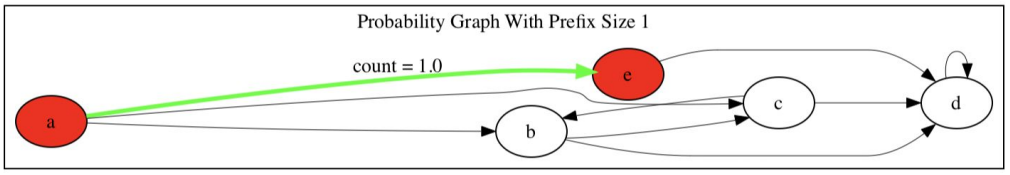
Case3.0: a c b d Case2.0: a c b d Case1.2: a b d d
Case1.1: a b c d Case1.0: a b c d Case2.1: a c b d

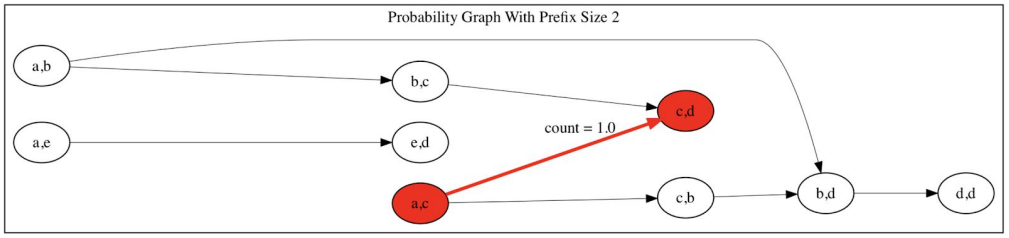


I delete the activity ‘c’ from Case2.0 and activity ‘b’ from Case2.1 to get the file Example_EventLogForTraining.xes. And cases are:

Case3.0: a c b d Case2.0: a b dCase1.2: a b d d
Case1.1: a b c d Case1.0: a b c d Case2.1: a c d

Combining the Probability Graph above we can see that using this event log to do compliance checking there will no other alerts occur in window size ‘1’, except that Alert T described in part 1. But with window size 2, there will be one new alert from ‘ac’ to ‘cd’ and only this one. Our system gives exactly the same result as we analyse.





2 Code Comments Format

We use following templete to comment our code:

For parameter type - *class* we write as below. The class name is enclosed within " ` "
class `streaming_event_compliance.object.automata.Automata`: <param_description>
For other parameter types we write as below
int : <param_description>

2.1 Comments for functions

Syntax:

```
"""
Description:
        <function description is written here>

:param param_name: :param_type: param_description
:return param_name: :param_type: param_description
"""
```

Example:

```
"""
Description:
        This function takes sink_node, source_node and checks if the automata of
        'windowsize' has any source_node and sink_node that matches tbe source_node,
        sink_node  passed.

:param event: :`dict`={'case_id': `string`, 'activity': `string`}
:param autos: :`dict`={int: class `streaming_event_compliance.object.automata.Automata`}:
                        {window size: the automata used to do compliance checking}
:param alogs: :`dict`={int: class `streaming_event_compliance.object.automata.Alertlog`}:
                        {window size: the alertlog used to stored the alert-information}
:param uuid: :`string`: client-id
:return: :int: {0: not created, 1: created successfully}
"""
```

2.2 Comments for classes

Syntax:

```
"""
Description:
        <Class description is written here>

Instance Variables:
```

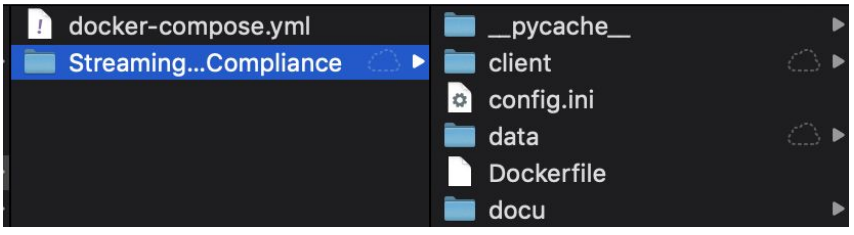
```
:param param_name: :param_type: param_description
"""
Example:
"""
Description:
   This class describes the basic elements of an automata.

Private Instance Variables:
_node: :`dict`={`string`: int}: {nodename: the number of this nodes appear}.
_connections: :`dict`={`string`: class
`streaming_event_compliance.object.automata.automata.Connection`}:
            {hash value: one connection object}.
_total_number: :int: the number of cases, calculated by taking count of node named 'None'
            (using 'None' to indicate the beginning of one case in our setting).
"""
```

3 Deployment

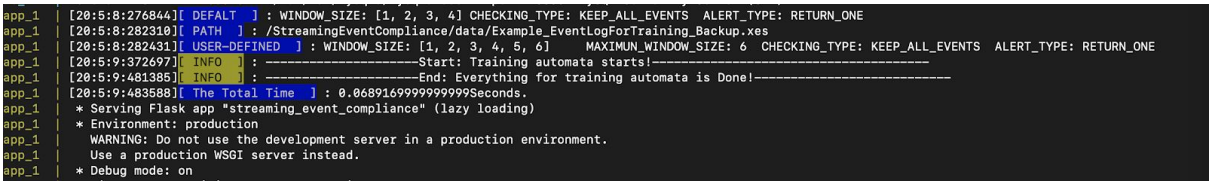
1. Using *docker-compose.yml* to create a container, which links the **mysql in local** and project together.
   Setup:
   a. Change **deploy = True** in File `/StreamingEventCompliance/streaming_event_compliance/__init__.py ` (Default = False)
   b. We use `docker.for.mac.host.internal` as hostname for Mac. And need to change if works on Windows or Linux.
   c. We are now just using user root to connect to our database. Mysql root password should be consistent with your local root password. Please set the password in file `/StreamingEventCompliance/streaming_event_compliance/__init__.py ` and `/StreamingEventCompliance/Dockerfile`. (Default = root)

```
deploy = True
if deploy:
    DATABASE_PATH = 'mysql+pymysql://root:root@docker.for.mac.host.internal/compliancechecker'
    app.config['BASE_DIR'] = '/StreamingEventCompliance/'
```

   d. Copy docker-compose.yml to to the directory containing StreamingEventCompliance project.



   e. Run docker-compose up in this directory in the terminal



2. Build mysql container and project separately.

4 Phase Review

Zheqi Lyu:
We are try to deploy our project using docker, but we don't have so much production. We tried several times with different ways, but it still works fine. Sometimes it occurs different problems, when we use different version of mysql and different version of python. I think, I need to improve the ability to solve problem and retrieve in the Internet. We hope it can run in the next phase.

Sabya Shaikh:
We are facing multiple issues with dockerizing our code. We tried different options and still fail to do so(with windows). It was the most time consuming part. Project is working fine. All the functionalities are working as expected. To make it easily understandable we have commented the code.

Jingjing Huo:
We don't know how to use docker and try again and again, it's just like wasting time, it's the way to learn something. But maybe it's not the best way, because the mechanism behind should be the most important thing, we should first understand that. Maybe because the duration of this phase is not much long or others, I cannot sit down and learn it carefully. I hope we can improve in the next week.