Dear Gaël,

thanks for the answer your intention to participate into the contest.

The following is the accuracy wrt. your February log:

	model_1	model_2	model_3	model_4	model_5	model_6	model_7	model_8	model_9	model_10
True Positive:	15	15	16	13	20	17	20	16	13	17
False Positive:	8	0	9	13	11	5	10	7	7	7
True Negative:	12	20	11	7	9	15	10	13	13	13
False Negative:	5	5	4	7	0	3	0	4	7	3
Number of traces correctly classified	**27**	**35**	**27**	**20**	**29**	**32**	**30**	**29**	**26**	**30**
Final result	**285**									

We had to change the "intermediate" test event logs of March because they seemed problematic. We uploaded the new "intermediate" test event logs of March on 5. April 2019 at 16:11.
If you did not download the set of event log after that date, you should repeat the analysis. I really apologize for this.

Regarding the other questions of yours, see below the answers

> - according the "Behavioral characteristics" the log7 is not associated with a DMN table. This is indeed the case in the training dataset. However, in the test dataset (February and March) I see additional data. Can you please explain why?

I confirm that log 7 does not have an associated DMN table. You can safely ignore the additional data. I will remove the additional data from the test logs.

> - Is it possible to have uncomplete traces in the test logs? If yes, should we consider them as fitting or not?

If incomplete traces are in the test logs, they should be considered as not fitting.

> - Is it possible that some "Behavioral characteristics" (BC) mentioned are not true? E.g., Can the BC indicates a loop where, in fact, there is no loop in the "ground truth" process model that we have to discover?

It seems that, indeed, there has been some issues with loops in the generated models. It seems that the model generator had some problems with generating loops. **In fact, no model indeed contains a loop.**
Basically, it is a feature that is not tested here.

Please let me iterate my apology for the issues.

Regards,
Massimiliano

PS: Because of all the issues above, **we have decided that the deadline will be extended of 15 days**. We will soon update the website accordingly. This is to account for the extra time that other contestants and you will have to spend because of these issues.
Also, because of the issues, **we have decided that this first classification attempt of your is considered as the first of the two that you are allowed. Basically, your counter resets.**

Gaël Bernard wrote on 05/04/2019 at 19:03:
> Dear organizers of the process discovery challenge,
>
> Here is my first attempt to classify the training logs. Can you pls provide the accuracy of my evaluation?
>
> If I may, I have few additional questions concerning the competition:
> - according the "Behavioral characteristics" the log7 is not associated with a DMN table. This is indeed the case in the training dataset. However, in the test dataset (February and March) I see additional data. Can you please explain why?
> - Is it possible to have uncomplete traces in the test logs? If yes, should we consider them as fitting or not?
> - Is it possible that some "Behavioral characteristics" (BC) mentioned are not true? E.g., Can the BC indicates a loop where, in fact, there is no loop in the "ground truth" process model that we have to discover.
>
> Thanks a lot.

Have a great weekend.

Best regards
Gaël

_____

Gaël Bernard
PhD Student
IS - Big Data & Analytics