Test result on demo-example

This file is used to record the test result of our demo example.. Firstly, we list the event logs and models it need, then do the test and give out the results on it..

In the folder I just created much simple examples to show this. It’s not enough, So change it to the whole data. If we combine them together, it is too much, so we just use the data there and generate the right data and do simple examples. That’s all..

Event log:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data ID | File name | process | cases | events | Event class | Min | Mean | Max |
| D1 | demo-s1-01.xes | Data description:  50 a, b, c1, d, x1,e, f1, g1, g2, h, i ;  50 a, b, c2, d, x2, e, f2, g2, g1, h, i ; |  |  |  |  |  |  |
| D2 | demo-s2-01.xes | Data description:  pos: 50 a, b, c1, d, e,f1, g2, g1, h,i;  50 a, b, c2, d, e,f2, g1, g2, h,i;  Neg: 30 a, c1, d,e, f2, g2, g1, b, h,i;  20 a,c1, b, d, e, f1, g1, g2, h,i; |  |  |  |  |  |  |
| D3 | demo-s3-01.xes | Data description:  Pos:  50 a, b, c1, d, e, f1, g1, g2, h, i ;  50 a, b, c2, d, e, f2, g2, g1, h,i;  Neg:  50 a, b, c1, d, e, f2, g2, g1, h,i;  50 a, b, c2, d, e, f1, g1, g2, h,i; |  |  |  |  |  |  |
| D4 | If we want to create data unsuitable for all of those data, we need have noisy data available at our case.  To design an event log, we have already the cases for the repair method, but for IM, how to change them then??  IM noise data | How to combine all the data together??  50 a, b, c1, d, x1,e,f1, g1, g2, h,i ;  50 a, b, c2, d, x2, e,f2, g2, g1, h,i;  50 a, b, c1, d, x2,e,f1, g2, g1, h,i;  // 50 a, b, c2, d, x1,e,f2, g1, g2, h,i;  20 a, c1, b, e, x2, d, f1, g1, g2, h,i;  Neg: 30 a, c1, d,e, f2, g2, g1, b, h,i;  30 a, c1, b, d, x2, e, f1, g1, g2, h,i;  50 a, c1, d, b,e, f2, g2, g1, h,i;  50 a, c2, d, e, x1,b, f1, g1, g2, h,i; |  |  |  |  |  |  |
| D5 | For IM bad model, change parts of the right sequence ones but others keep the same, then let use check the result… See if it is ok for it…  Not only IM but also the repair model methods...So what to do then?? With noise, we need to repair them, if they are positive or negative deviations… If they are positive, in dfg, it is balanced by negative information… So no need there, | Pos:  50 a, b, c1, d, x1,e,f1, g1, g2, h,i ;  50 a, b, c2, d, x2, e,f2, g2, g1, h,i;  50 a, b, c1, d, x2,e,f1, g2, g1, h,i;  // 50 a, b, c2, d, x1,e,f2, g1, g2, h,i;  20 a, b, c1,d, x2, f1,e, g1, g2, h,i;  5 a, c1, d, b, e, f2, g2, g1, b, h,i;    Neg: 30 a, c1, d, e, f2, g2, g1, b, h,i;  30 a, c1, b, d, x2, e, f1, g1, g2, h,i;  50 a, c1, d, b,e, f2, g2, g1, h,i;  50 a, c2, d, e, x1,b, f1, g1, g2, h,i; |  |  |  |  |  |  |

I have tried so many times but it can not discover the right model, like we expect.. Should we limit the connection of directly-follow relation?? We can not say this..

If we find out the parallel relation of b with another activities, so what to do then?? The existing model has it with others, but we can not see it in the other models.. It is strange but it is true. Only with positive and negative ones…

We can change another parts, and change its order of them.. Just use the data availlable to check the result..

How about we change the structure from sequence to parallel?? I really hope that we can have so many data available.. But if we want to change the structure, it is difficult..

But however, it should work by our definition… We only add those two activities and keep them structure same..

I will use the optimization methods to reject all the data and see what we get..

Models :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Data file | Setting | Model file | Model figures | Description |
| M1 | BPI15\_1\_40\_filter.xes | Inductive Mine:  IM-infrequent : 0.2  concept: name | BPI\_1\_40\_M1\_IM0\_classes.pnml | BPI\_1\_40\_M1\_figure.pdf |  |

Results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Log | Model | Techniques | Result |
| T1.1 | D1 | M1 | Dfg method: | We have a lot of silent transitions in the graph, and it should be there actually.. What causes it behaves like this??  The dfg generated is different from model before, right??  Why do we skip so many activity??  People says that if the same bugs hidden in the programs like the ones you found..  The cardinality goes up, how could it goes up??  It is a strong connection, which we don’t usually have it.. So if we want to assign the cardinality, we should use the event log sizes num??  But it doesn’t change the result, so I need to consider why it is so..  why there is so many silent transitions?? Because it can not distinguish b is parallel to the other situations, then it uses silent transitions to keep it there..  // how to avoid it there??  The reason behind it is clear, but how to avoid it?  If we use the reachability graph to discover the models, we have the directly-follows relation there, it stands a point. So what to do then?? It can result in many models.  If we extract the directly-follows relation from the models directly.. what to do then?? Then the directly-follow relation contains there, like a→b and b→i, the others not change??  If we use the BFS search algorithm, we can have it, but is it complete?  Sometimes it is better just to put them together to have a better result?? I’m not so sure, but the result currently, shows sth wrong.. |
| T1.1.2 | 4.2 | M1 | Fahland’s method: only on positive, test based on train data with all labels | We get models with a lot of complexity to fulfill the needs in the data. But TP=FP=0, don’t know why.. |

We can have a compressed models and data, now let us see the result that if those changes can be reflected by repair methods.. of Farhland or Dees..

Change the introduction part, and finish the first part of this experiments. Need to point out the changes of our data, but what we know is that when the data are here, we can’t see it.. Just combine them all together.. But make sure that you know it

to plot weight changes graphics,

fix one variables, and check the others changes:: the combination are the same, 10\*10 = 100, combinations…

fix the existing weight, and get the values according to others::

if one of them is empty, then we don’t use this value, but should we combine the TP and others together??

Not really, only the valuable values::

1. filter out the values with empty values

2. sort them into existing , pos and neg weight based or we aggregate their sums when the existing is fixed..

The progress on demo-example seems not so well, but I need to overcome it and write in the papers..

With the long-term dependency, and other stuff there. So try it again!!

|  |  |  |  |
| --- | --- | --- | --- |
| Data ID | File name | process | cases |
| D1 | Log-demo--s1-01.xes | Data description:  Pos: 50 a, b, c1, d, x1,e, f1, g1, g2, h, i ;  50 a, c2, b, d, x2, e, f2, g2, g1, h, i ;  Neg: 50 a, c1, b, d, e, f2, g2, g1, h, i ;  50 a, c2, d, e, b, f2, g2, g1, h, i ; |  |
| D2 | demo-s2-01.xes | Data description:  pos: 50 a, b, c1, d, e,f1, g2, g1, h,i;  50 a, b, c2, d, e,f2, g1, g2, h,i;  Neg: 30 a, c1, d,e, f2, g2, g1, b, h,i;  20 a,c1, b, d, e, f1, g1, g2, h,i; |  |
| D3 | demo-s3-01.xes | Data description:  Pos:  50 a, b, c1, d, e, f1, g1, g2, h, i ;  50 a, b, c2, d, e, f2, g2, g1, h,i;  Neg:  50 a, b, c1, d, e, f2, g2, g1, h,i;  50 a, b, c2, d, e, f1, g1, g2, h,i; |  |