Project Ideas - General Approaches and possible Problems

October 18, 2017

1 Generalized Conformance Checking by incremental Computation

1.1 Goal

Find Log $L^* \subseteq L$ and Model M^* that is similar to M
Does this approach yield similar results?

1.2 Related Work

$https://www.matthiasweidlich.com/paper/log_model_trust_BPM_2016.pdf$

1.3 Approach

-choose trace by trace from Log L
-mine model from new Sublog and check for tree-edit distance in newly mined Model
-if tree-edit distance is in trust-based proximity to model stop
-this gives L^* and M^*

1.4 Problems

-does this approach guarantee, that L^* and M^* both are in trust-based proximity?

2 Offline, alignemnt-based Conformance Checking by incremental Computation

2.1 Goal

Try to approximate the correct fitness value for M and L by incremental approximation.

2.2 Approach

-choose trace by trace
-calculate alignment for trace and throug this the cost of trace
-save the alignment and the cost of it
-use already seen alignments to deduct costs of concatenations of already seen traces
-take average of costs and check how much it changes - maybe stop when change in average is below an ϵ

2.3 Problems

-should alignment costs just be added upon each other -can one approximate how big the difference in alignment costs changes, based on the relative position of the difference? i.e. earlier differentiations cost more than later ones -how to stitch together two already seen alignments, just add up? check for path in alignment from one to the other -what is new information?

3 Stratified Sampling for incremental Computation (Discovery + Conformance Checking)

3.1 Goal

Deduce more efficient sampling of traces

3.2 Approach

-partition Log into buckets and sample from buckets that have higher entropy

3.3 Approach

How can one efficiently deduct a good Bucket size?
How should entropy be measured?