SVIT Research Group

Traceability Links Recovery in BPMN Models through Evolutionary Learning to Rank

Raúl Lapeña, Ana Marcén, Jaime Font, and Carlos Cetina

Agenda

- Problem description & motivation
- Approach
- Preliminary results
- Conclusions
- Q&A

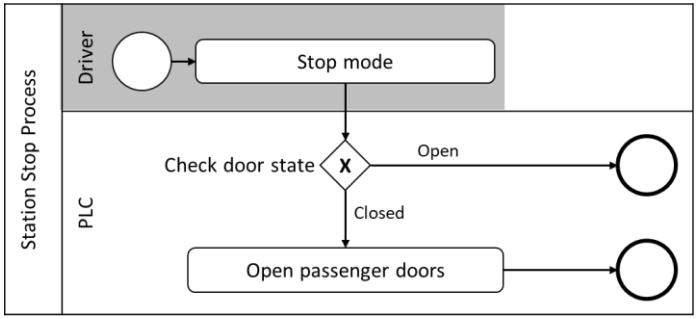
Problem description & motivation

- Traceability Links Recovery (TLR) is key to success in industrial software projects
- Industry partner (railway domain) → BPMN models
- BPMN models capture interaction, and are also used to design and derive other artifacts

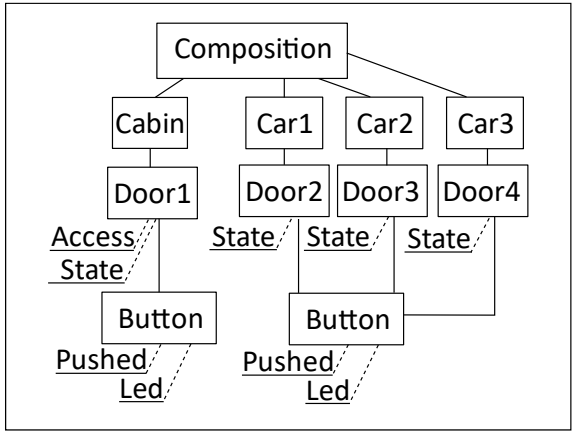
Problem description & motivation

- TLR techniques depend greatly on the language of the software artifacts under study

- BPMN models present less text than other artifacts



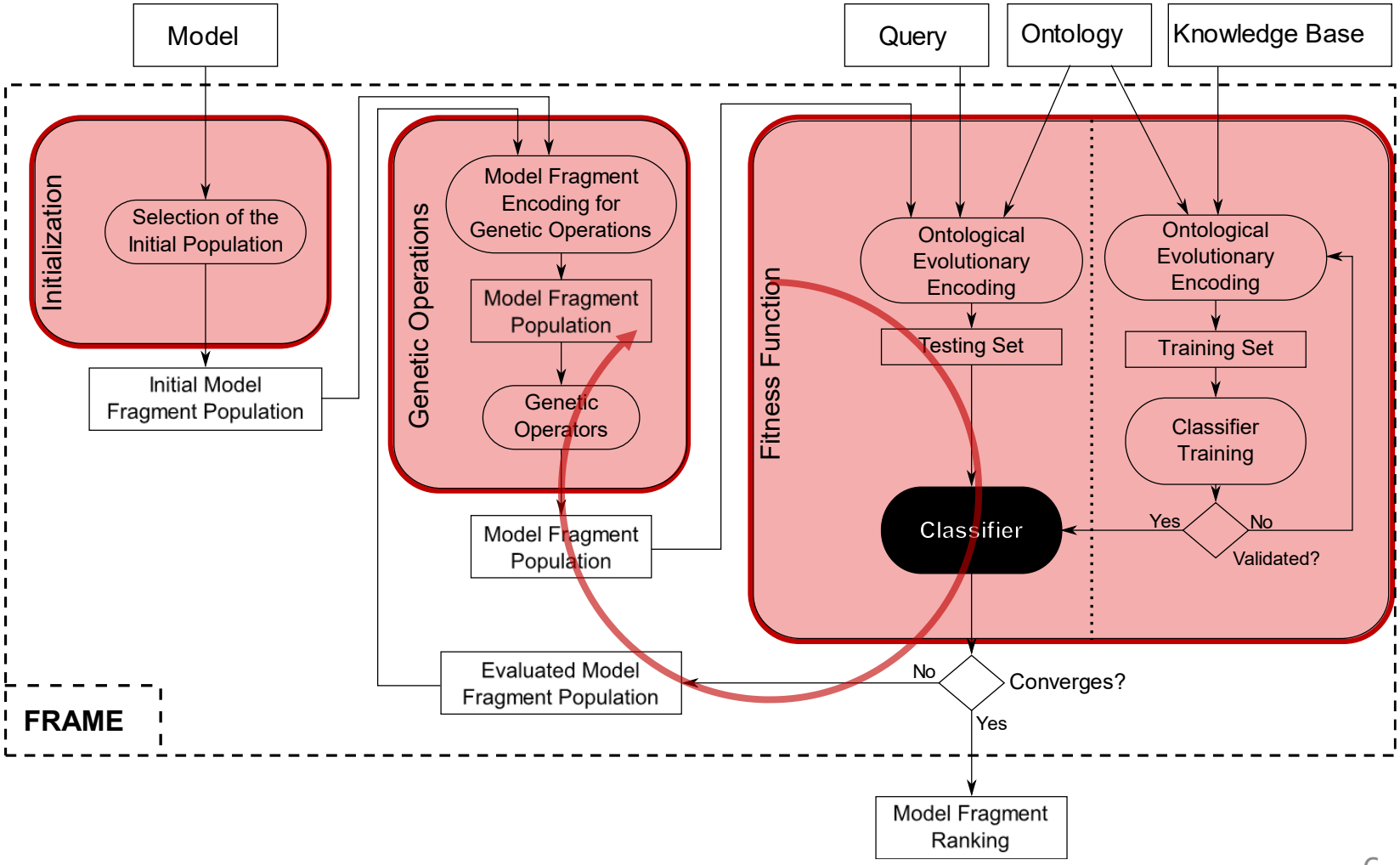
VS



Problem description & motivation

- Evolutionary Learning to Rank (ELtoR)
  - Not so dependent on linguistics
  - Better results than traditional TLR techniques when artifacts have less textual content
- Idea: adapt ELtoR for TLR between requirements and BPMN models

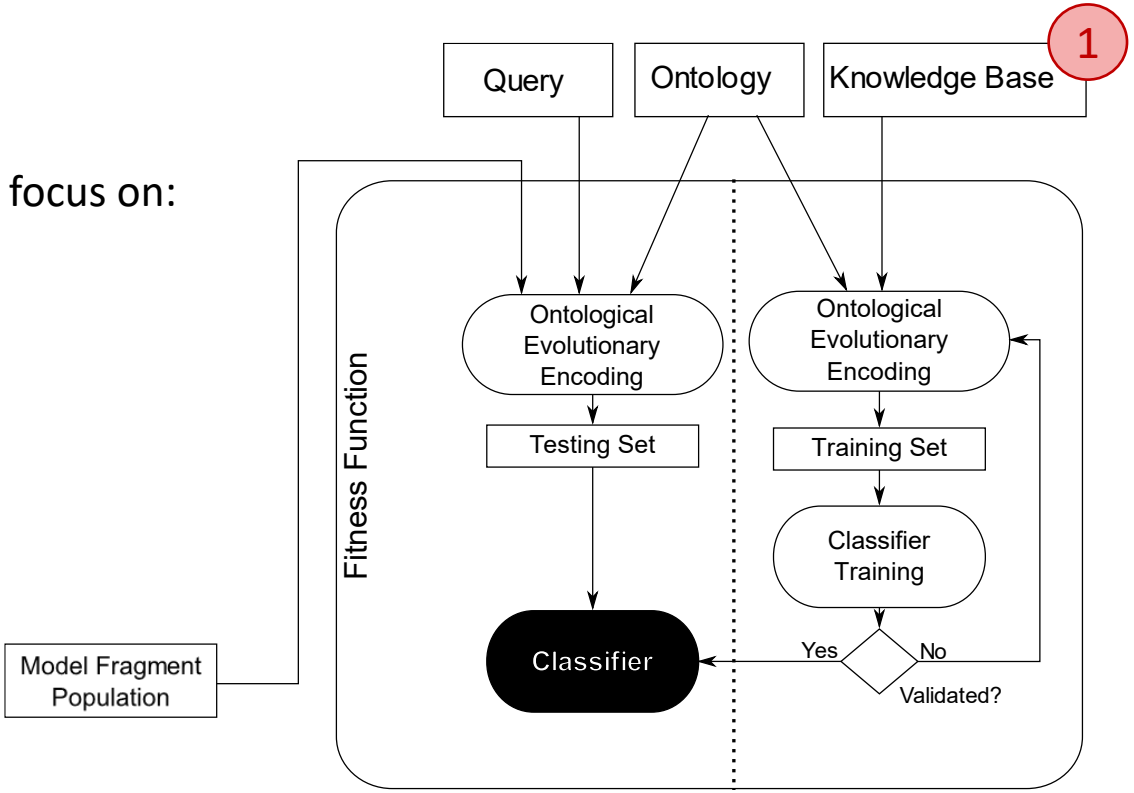
Approach


Model
Query
Ontology
Knowledge Base
Initialization
Selection of the Initial Population
Initial Model Fragment Population
Genetic Operations
Model Fragment Encoding for Genetic Operations
Model Fragment Population
Genetic Operators
Model Fragment Population
Fitness Function
Ontological Evolutionary Encoding
Testing Set
Classifier
Ontological Evolutionary Encoding
Training Set
Classifier Training
Yes
No
Validated?
Evaluated Model Fragment Population
No
Converges?
Yes
FRAME
Model Fragment Ranking

Preliminary results

To properly train the classifier, we focus on:

- The knowledge base

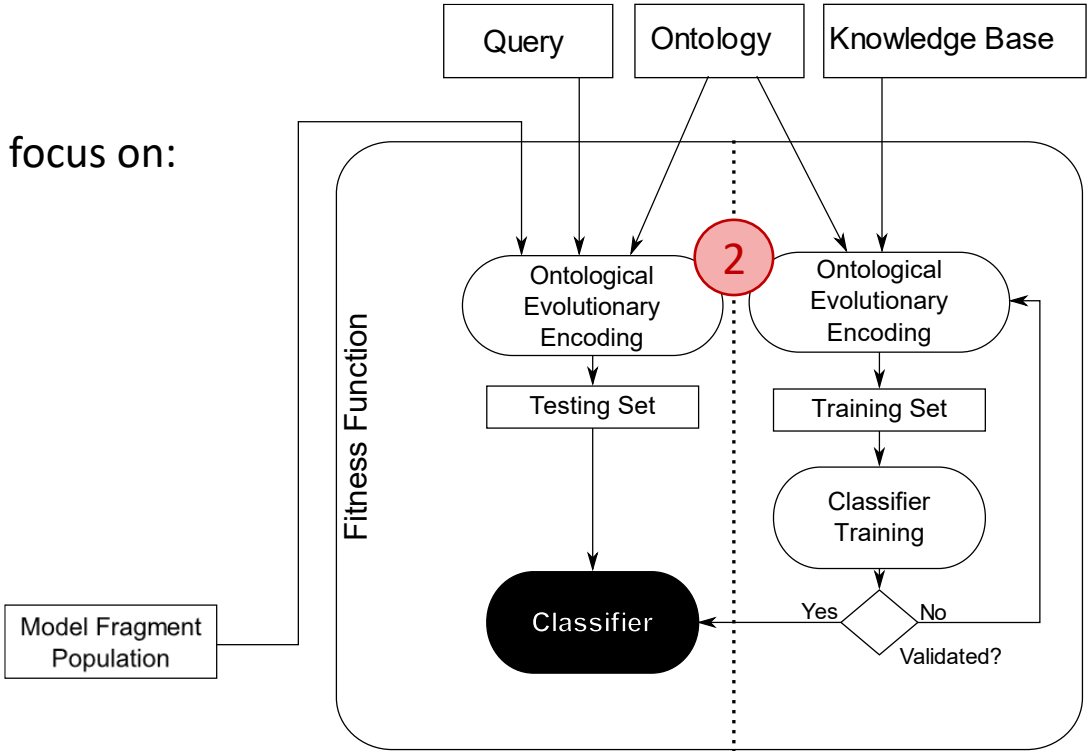


Idea: BPMN has little text, but there are language patterns that can be leveraged to link requirements and models → knowledge base should also have examples of all possible patterns.

Preliminary results

To properly train the classifier, we focus on:

- The knowledge base
- The encoding

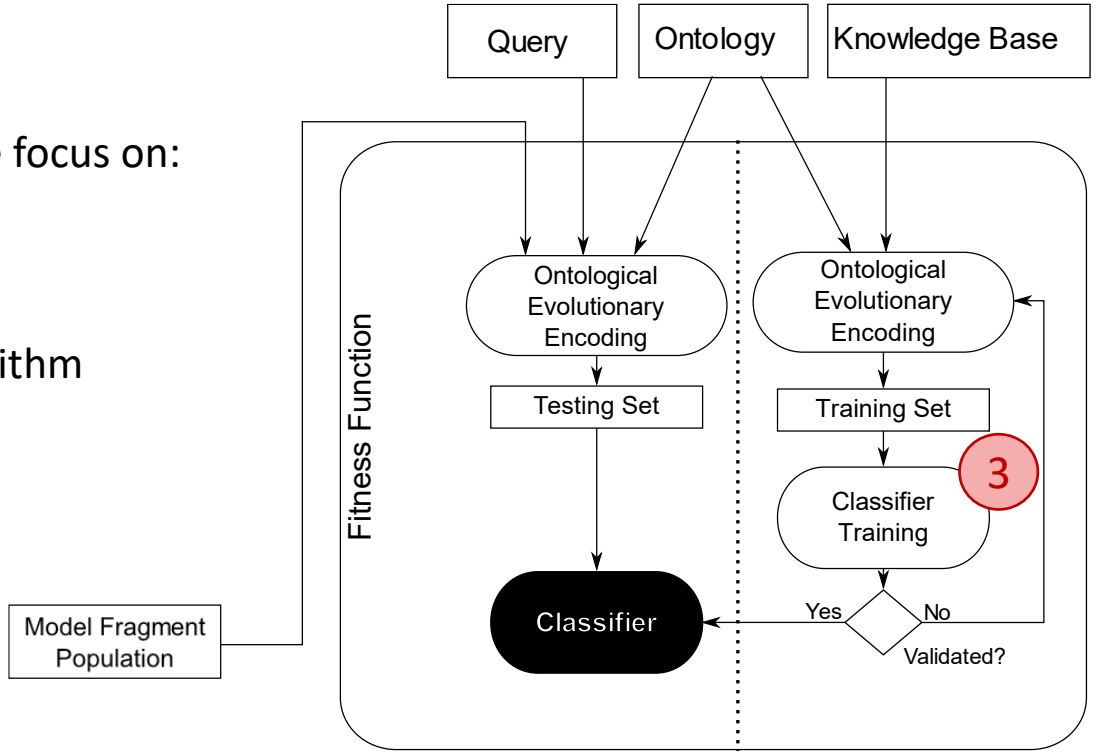


Idea: Adapt existing encodings of MDD models for BPMN models → Count presence/absence of model element **and** occurrences of element types and BPMN language patterns

Preliminary results

To properly train the classifier, we focus on:

- The knowledge base
- The encoding
- The machine learning algorithm



Idea: So far, we used Rankboost for ELtoR. However, due to the lack of text in BPMN, we expect to need a larger knowledge base → we might need other ML techniques with different capabilities

Conclusions

- TLR is key to success for industrial software
- Additional challenge with BPMN Models
- ELtoR can improve the state of the art

✓ Encoding transported to BPMN models

➢ Adapt the training process
➢ Adapt the knowledge base
➢ Explore best ML algorithms

Thanks for your attention

Q&A