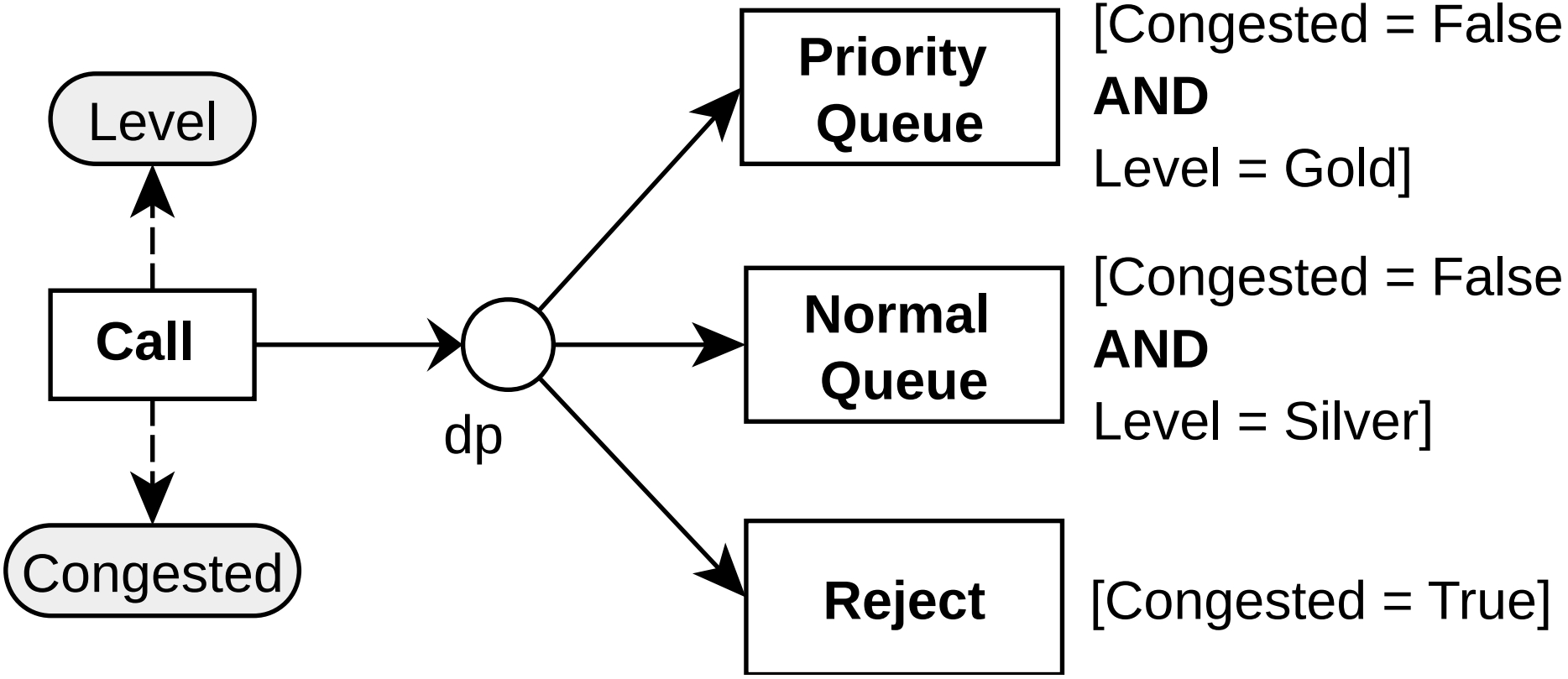
Decision Mining
Discovering Overlapping Rules

Felix Mannhardt
joint work with
Massimiliano de Leoni, Hajo A. Reijers,
Wil M.P. van der Aalst

TU/e Technische Universiteit Eindhoven University of Technology

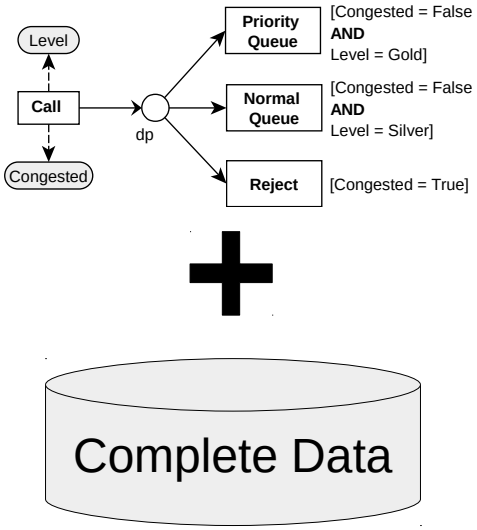
Where innovation starts

Mutual-exclusive Decision Rules



Why discover “Overlapping Decision Rules”?

Mutually-exclusive Rules



Deterministic System

#	Level	Congested	Outcome
1	Gold	False	Priority Queue
2	Silver	False	Normal Queue
3	Silver	True	Reject
4	Gold	False	Priority Queue
5	Silver	False	Normal Queue

Why discover decision rules that overlap?

Reason for overlap – Incompleteness

Incomplete Information

- Not recorded
- Process context
- Confidential
- ...

- Expert approval
- Deferred choice
- Randomized check
- ...

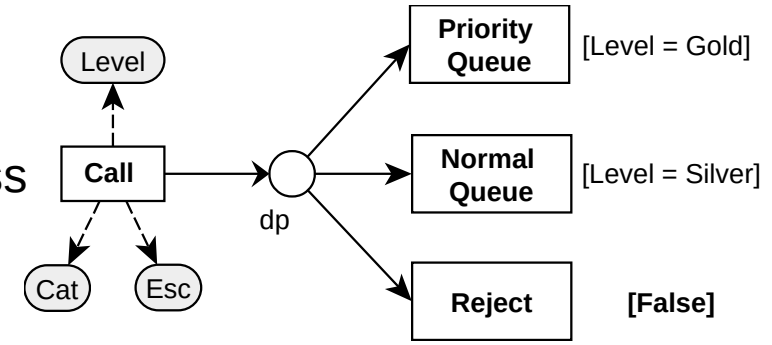
Example

#	Level	Esc	Cat	Outcome
1	Gold	No	A	**Priority**
2	Silver	No	A	**Normal**
3	Silver	No	A	**Reject**
4	Gold	No	A	**Priority**
5	Silver	No	A	**Normal**

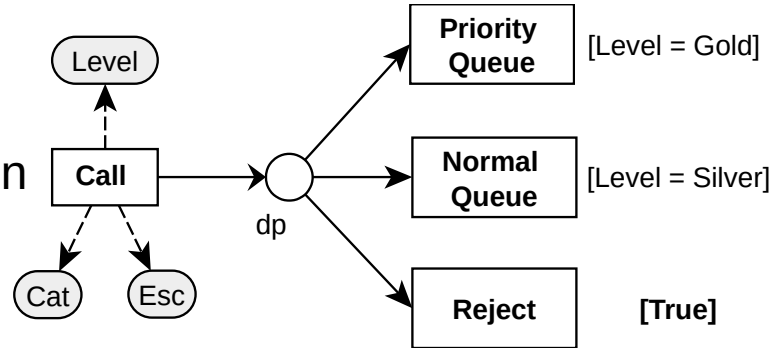
Trade-off between fitness and precision

#	Level	Esc	Cat	Outcome
1	Gold	No	A	**Priority**
2	Silver	No	A	**Normal**
3	Silver	No	A	**Reject**
4	Gold	No	A	**Priority**
5	Silver	No	A	**Normal**

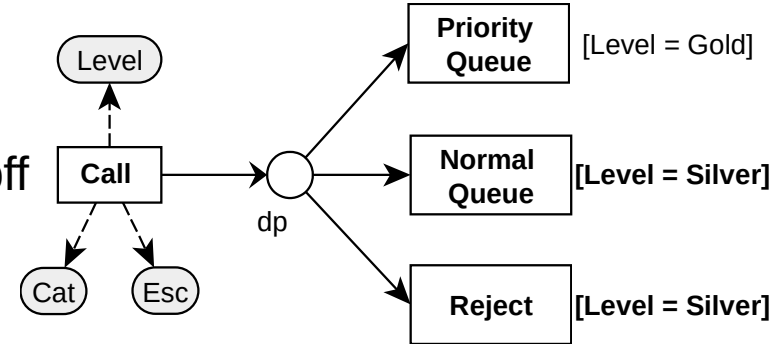
Poor Fitness



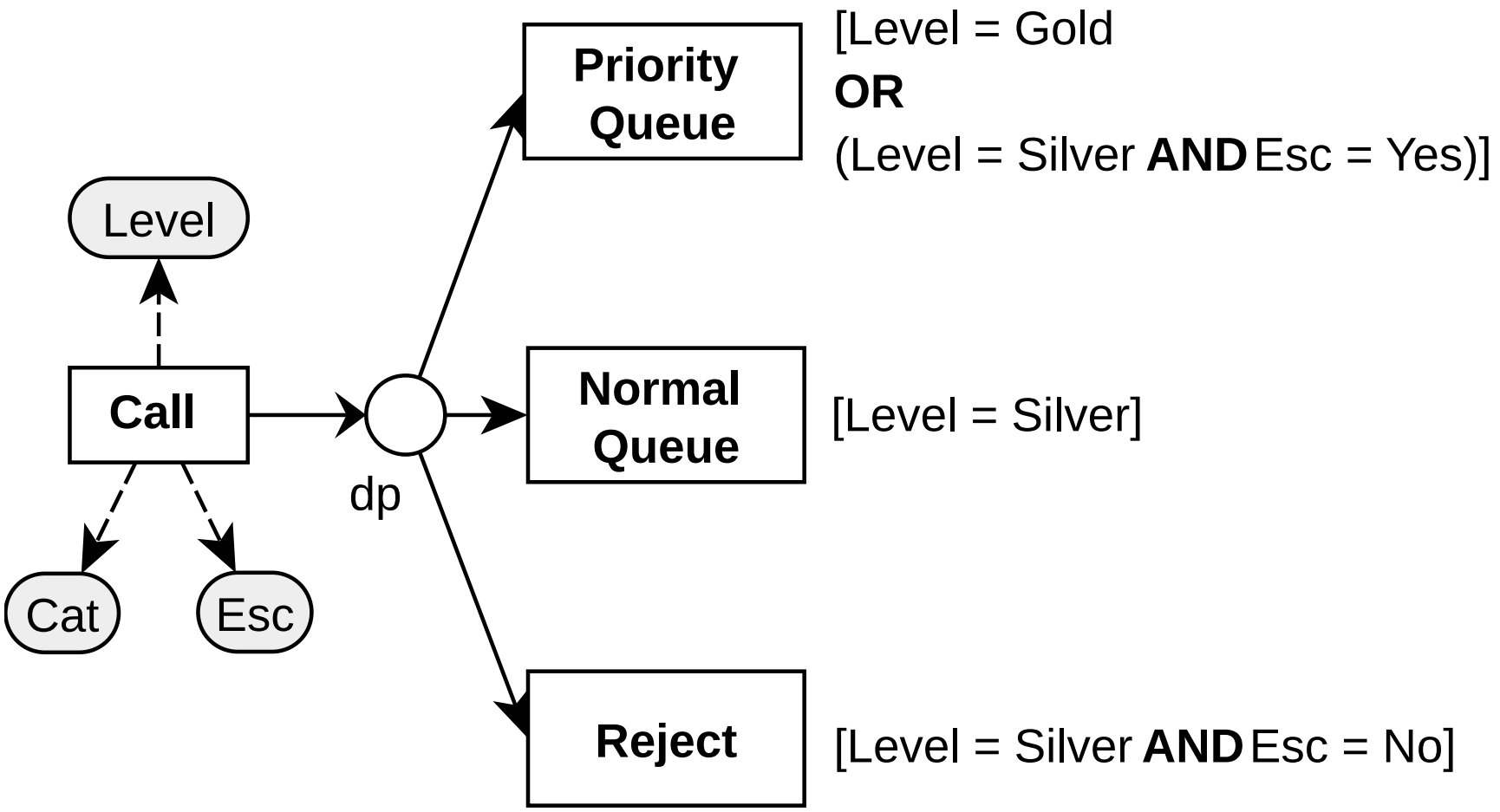
Poor Precision



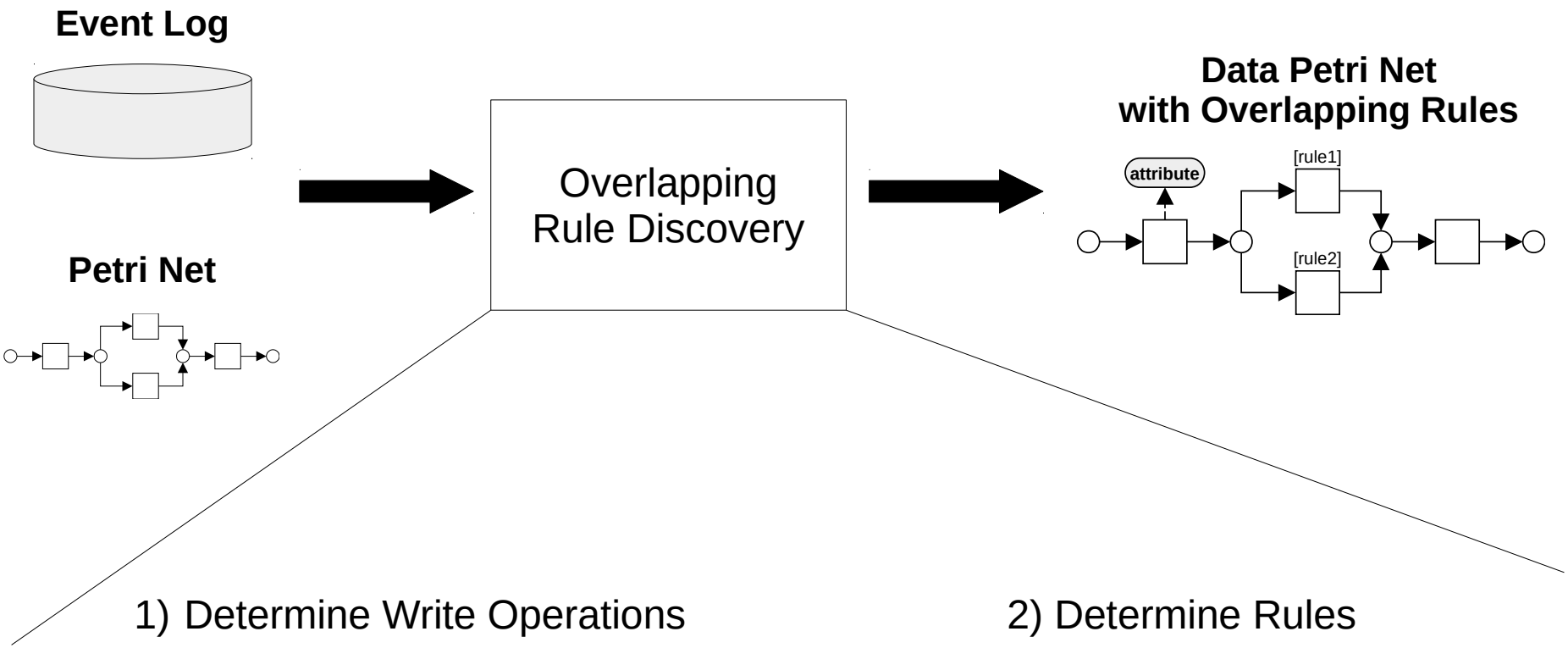
Good Trade-off



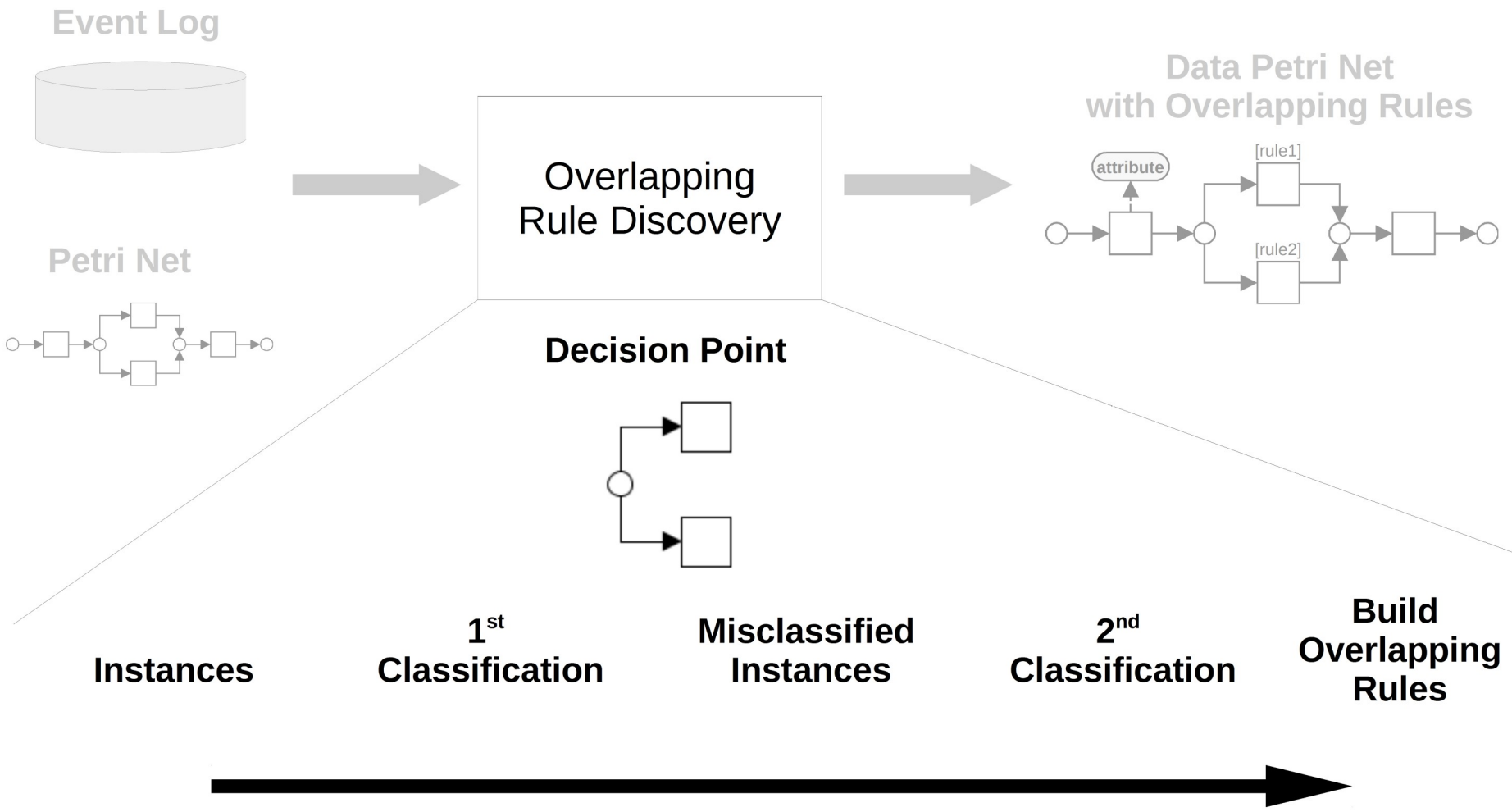
Our goal: Partially overlapping decision rules



Overview of the discovery method



Overview of the discovery method



1) Collect observation instances

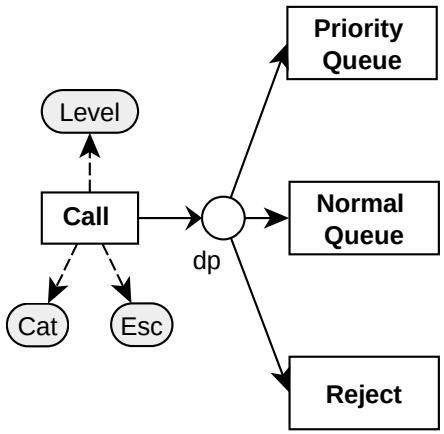
Event Log

#	Event	Level	Esc	Cat
1	Call	Gold	No	A
1	Priority Queue			
2	Call	Silver	Yes	B
2	Priority Queue			
3	Call	Silver	No	A
3	Reject			
..	..	..	..	

Observation Instances

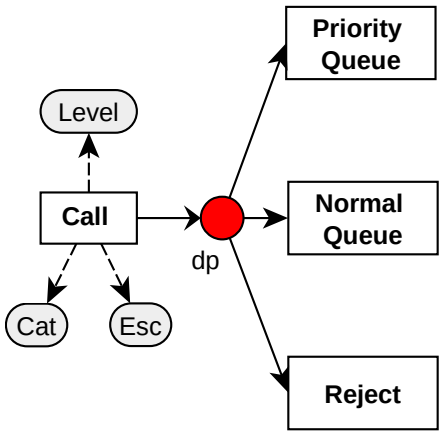
#	Event	Level	Esc	Cat
1	Priority Queue	Gold	No	A
2	Priority Queue	Silver	Yes	B
3	Reject	Silver	No	A
..	..	..	..	

Petri Net



Method handles:
- Loops
- Noise
- Unassigned values

Decision Point

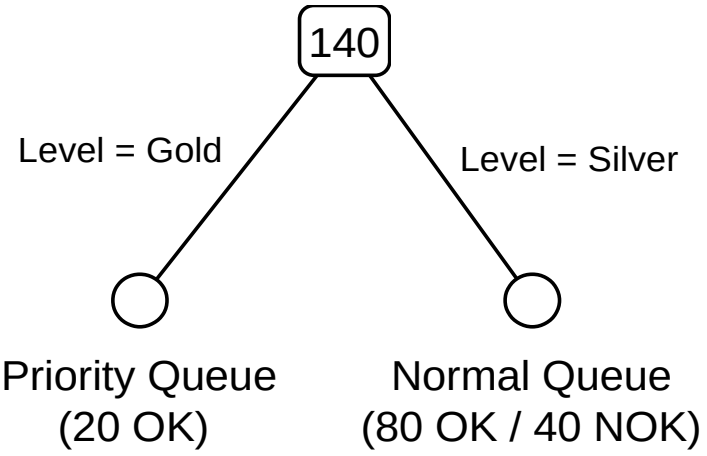


2) Build 1st Classification Using C4.5

Observation Instances

Count	Level	Esc	Cat	Outcome
x10	Gold	No	A	**Priority Queue**
x10	Gold	No	B	**Priority Queue**
x10	Silver	No	A	**Priority Queue**
x10	Silver	Yes	B	**Priority Queue**
x30	Silver	No	A	**Normal Queue**
x30	Silver	No	B	**Normal Queue**
x20	Silver	Yes	B	**Normal Queue**
x20	Silver	No	A	**Reject**

1st Decision Tree

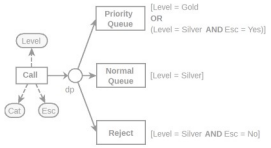


Overview of the discovery method

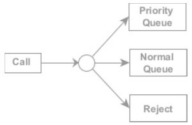
Event Log

Overlapping
Rule Discovery

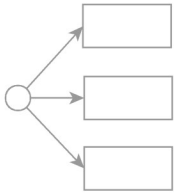
Data Petri Net
with Overlapping Rules



Petri Net



Decision Point



Instances

1st Classification

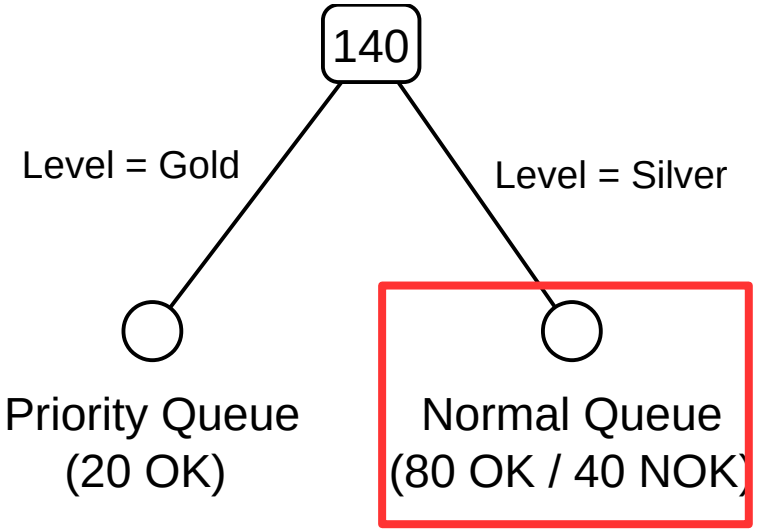
Misclassified Instances

2nd Classification

Build Overlapping Rules

3) Collect Misclassified Instances for Leafs

1st Decision Tree



Misclassified Instances

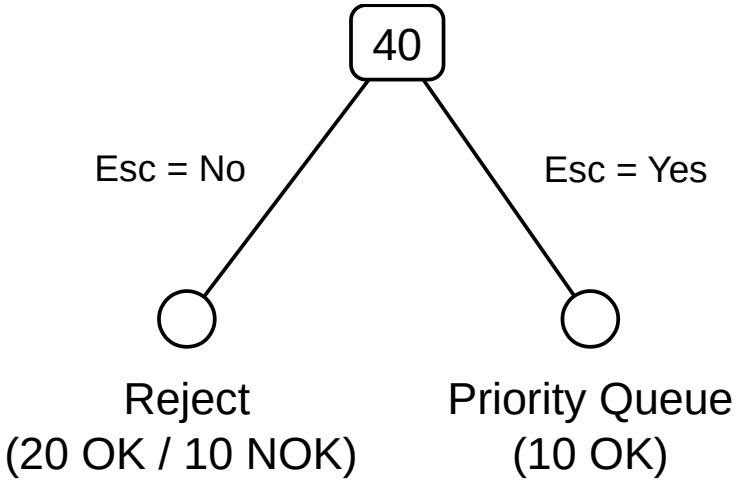
Count	Level	Esc	Cat	Outcome
x10	Silver	No	A	**Priority Queue**
x10	Silver	Yes	B	**Priority Queue**
x20	Silver	No	A	**Reject**

4) Build 2nd Classification Using C4.5

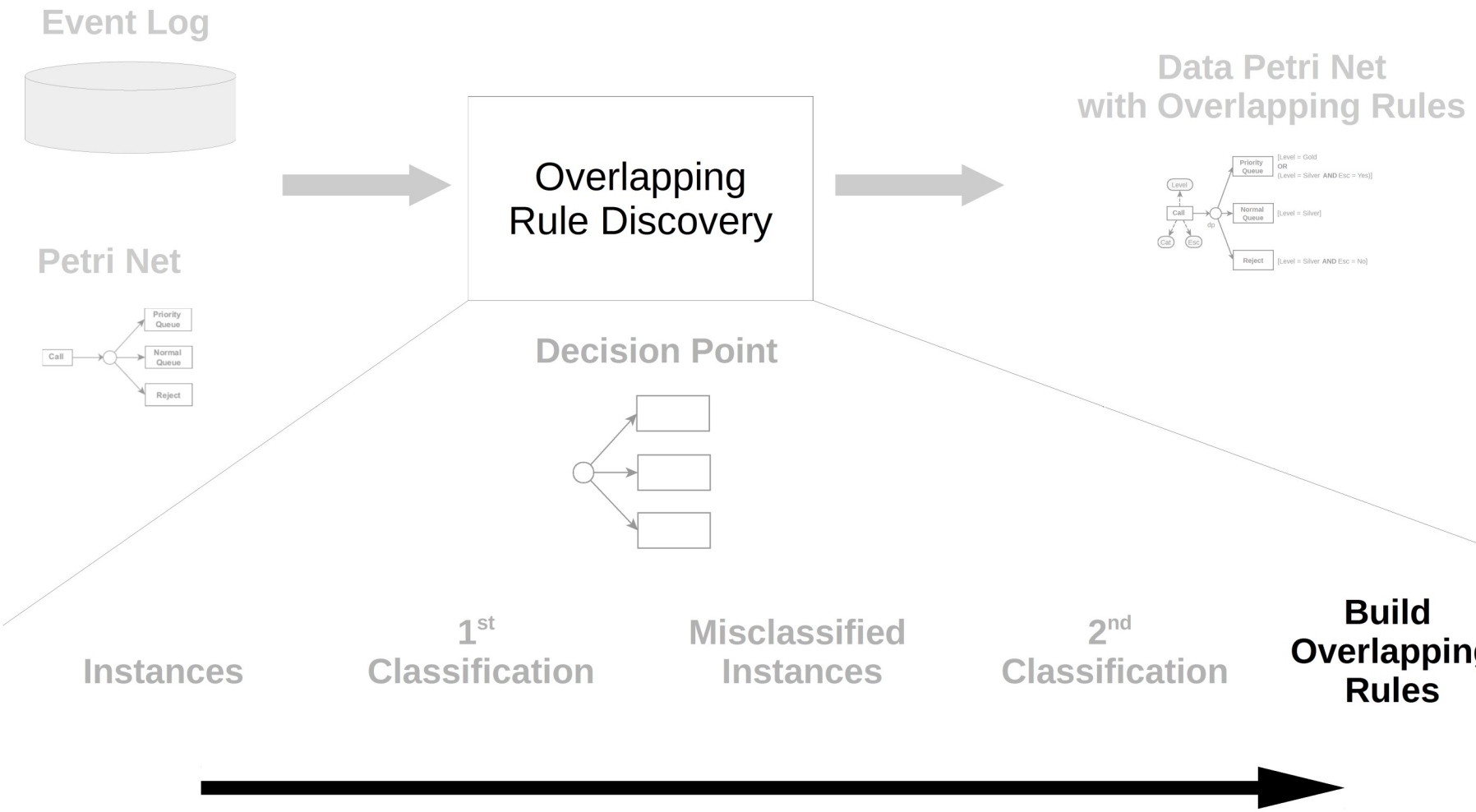
Misclassified Instances

Count	Level	Esc	Cat	Outcome
x10	Silver	No	A	**Priority Queue**
x20	Silver	No	A	**Reject**
x10	Silver	Yes	B	**Priority Queue**

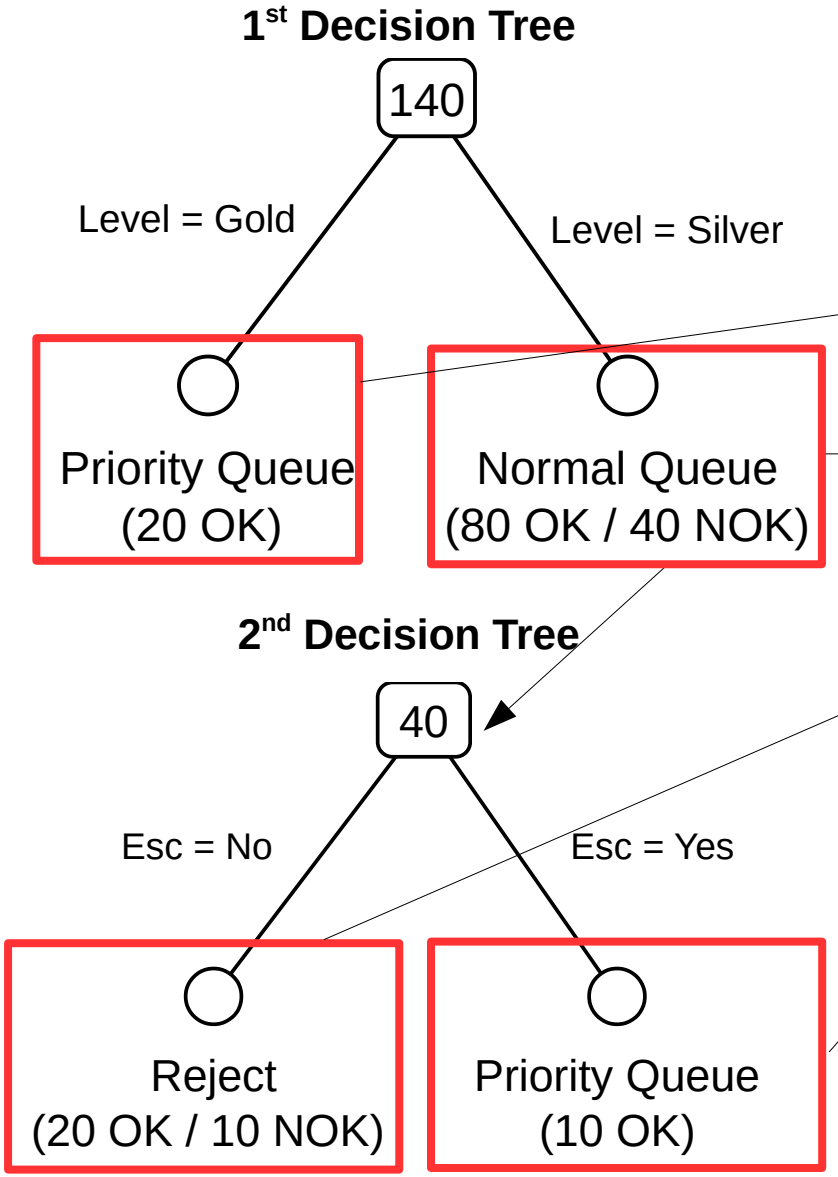
2nd Decision Tree



Overview of the discovery method



5) Build Overlapping Decision Rule



Overlapping Rules

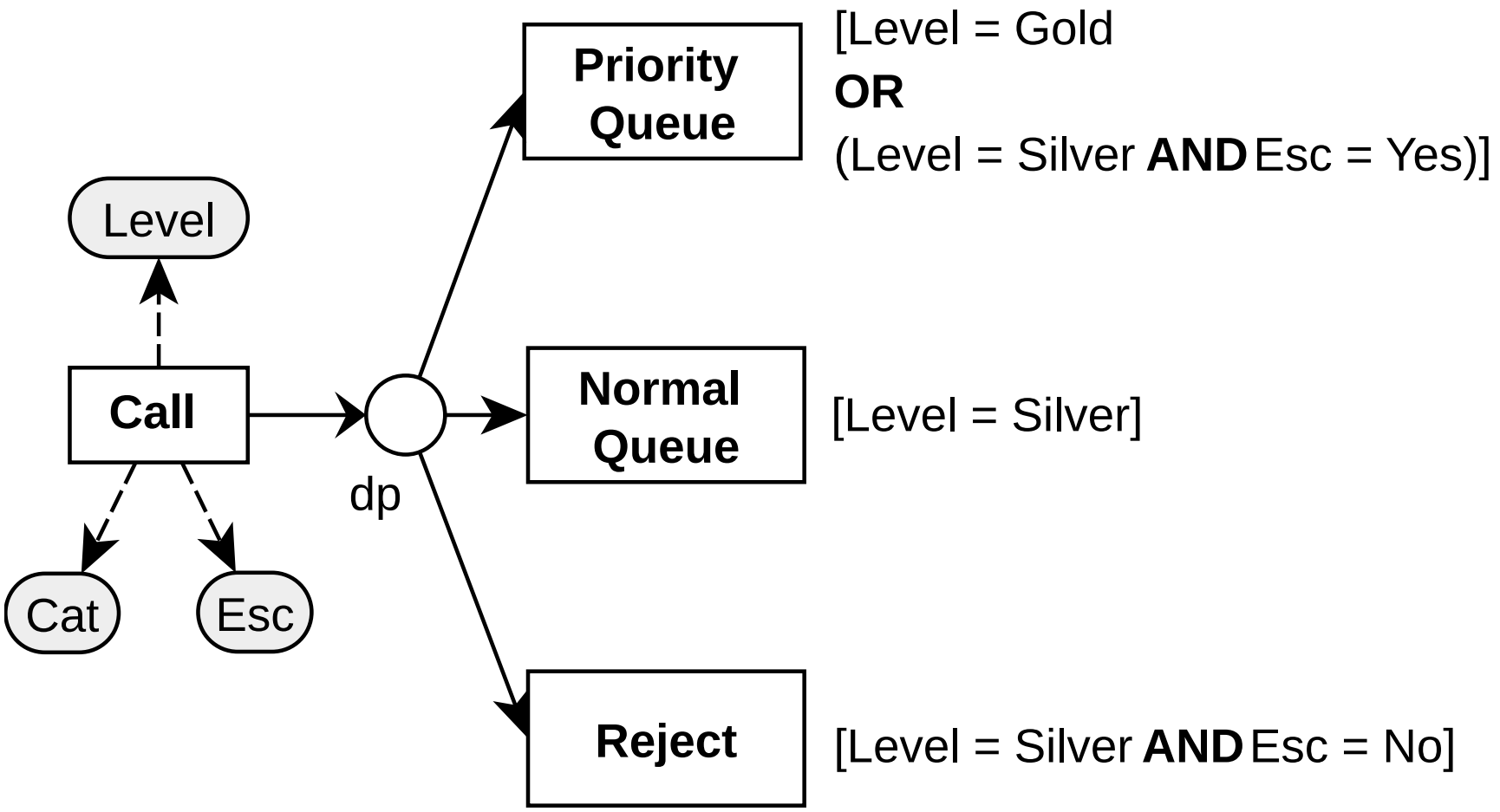
If Level = Gold
then Priority Queue

If Level = Silver
then Normal Queue

If Level = Silver **AND** Esc = No
then Reject

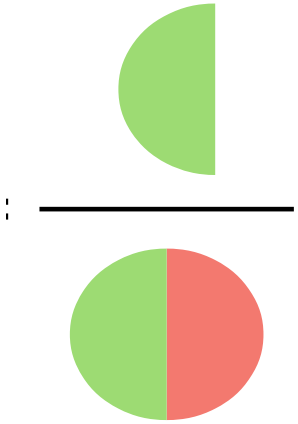
If Level = Silver **AND** Esc = Yes
then Priority Queue

Result: Data Petri Net with Overlapping Rules



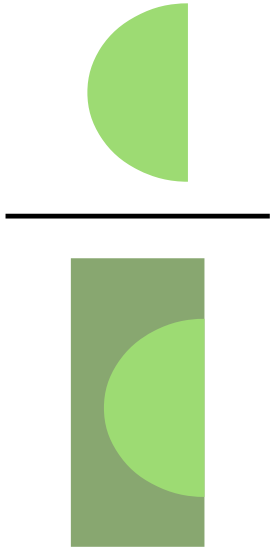
Evaluation – Measures

Place Precision



How much
unobserved behavior
is allowed?

Place Fitness



How much
observed behavior
is allowed?

Image source: https://commons.wikimedia.org/wiki/File:Precisionrecall.svg

Evaluation – Datasets

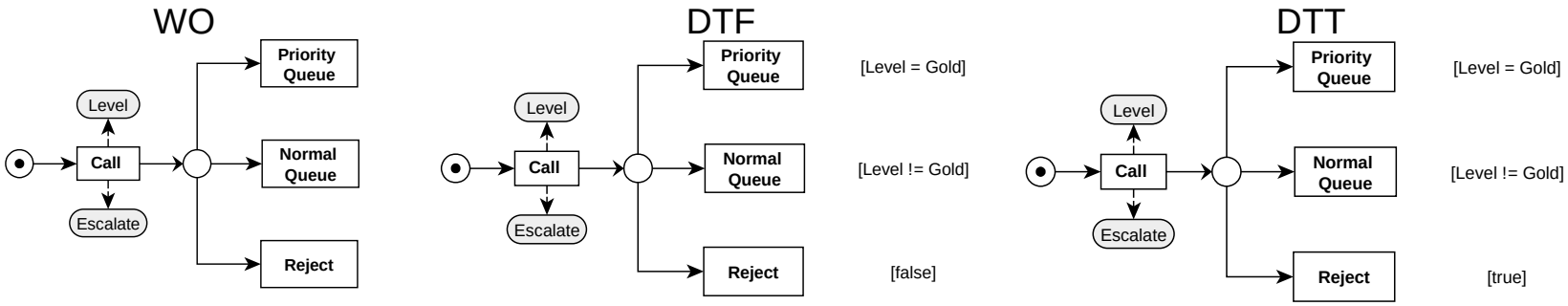
Datasets

Dataset	# Traces	# Events	# Attributes	# Decision Points
Road Fines Management	150,000	500,000	9	5
SEPSIS Patients	1,000	15,000	39	11

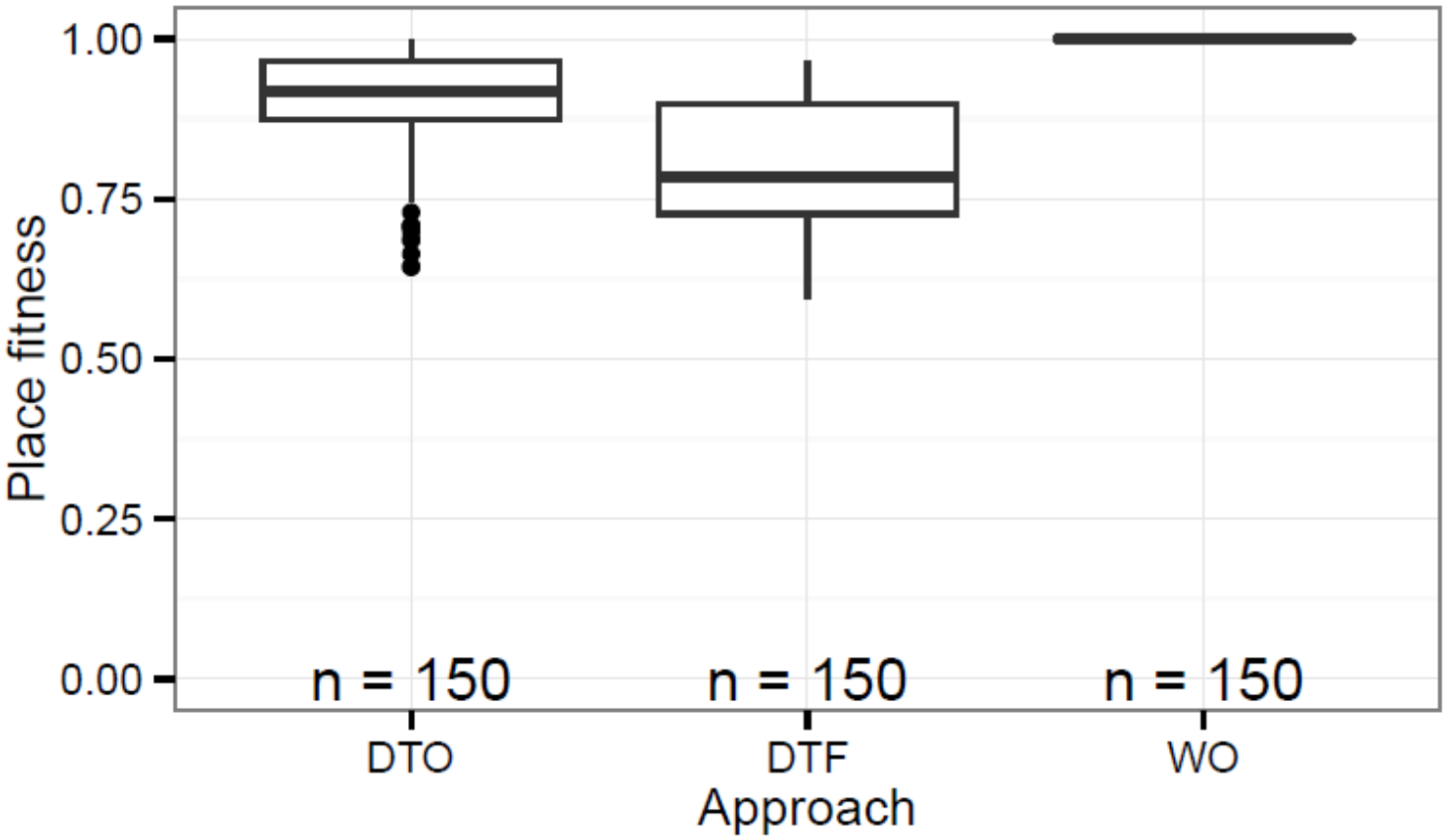
Evaluation – Setup

Compared Approaches

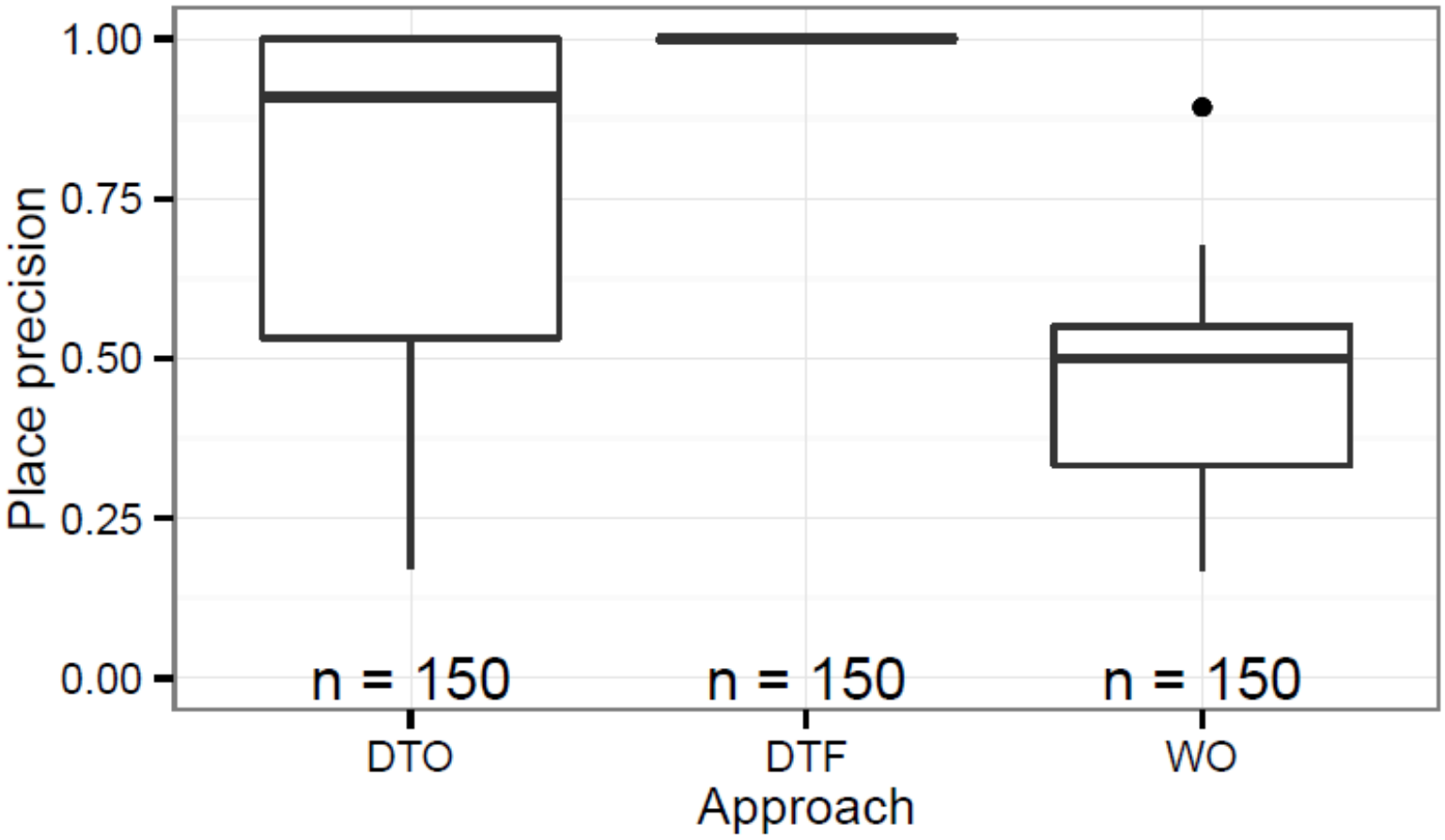
Approach	Description	Expected Precision	Expected Fitness
WO	Without rules	poor	good
DTF	Using false as default	good	poor
DTT	Using true as default	poor	good
DTO	The presented approach	balanced	balanced



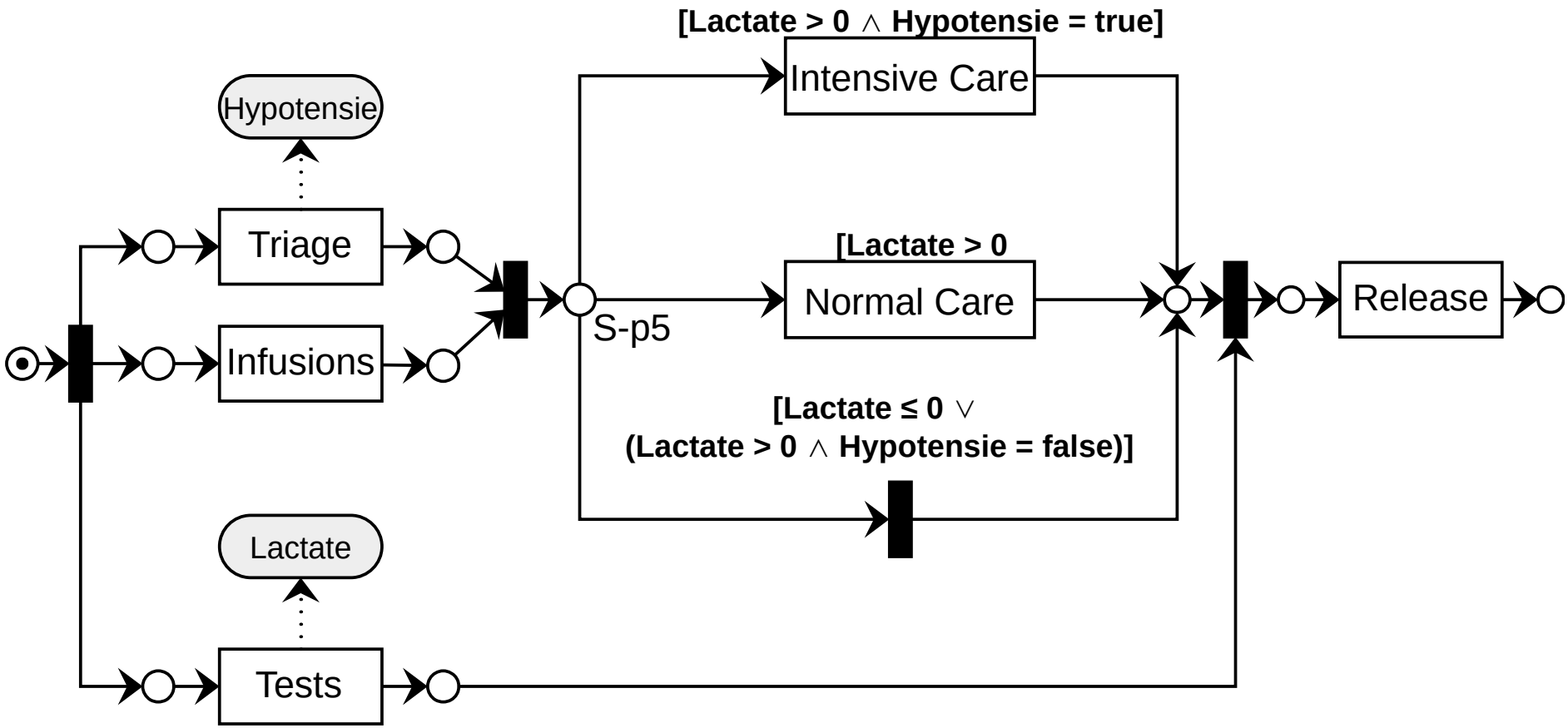
Evaluation – Fitness



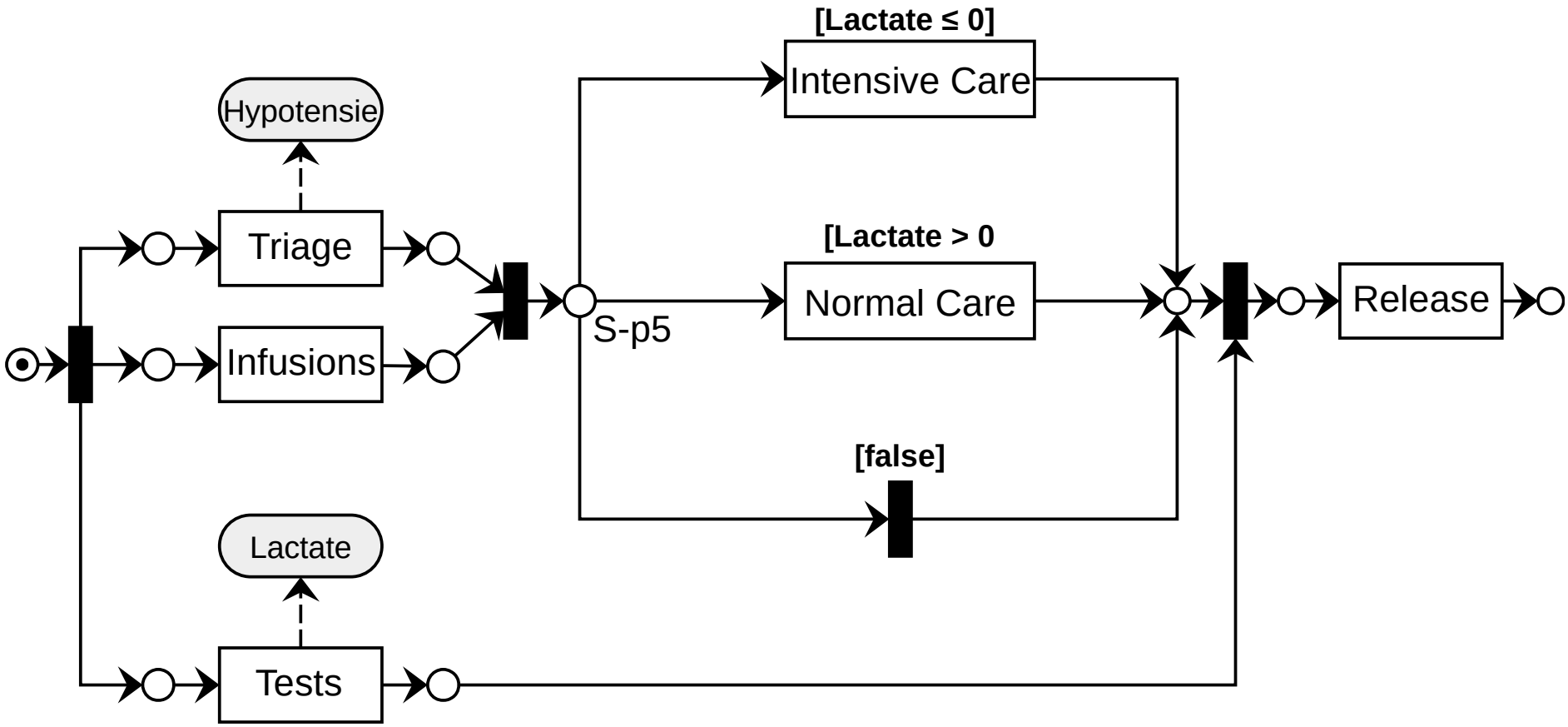
Evaluation – Precision



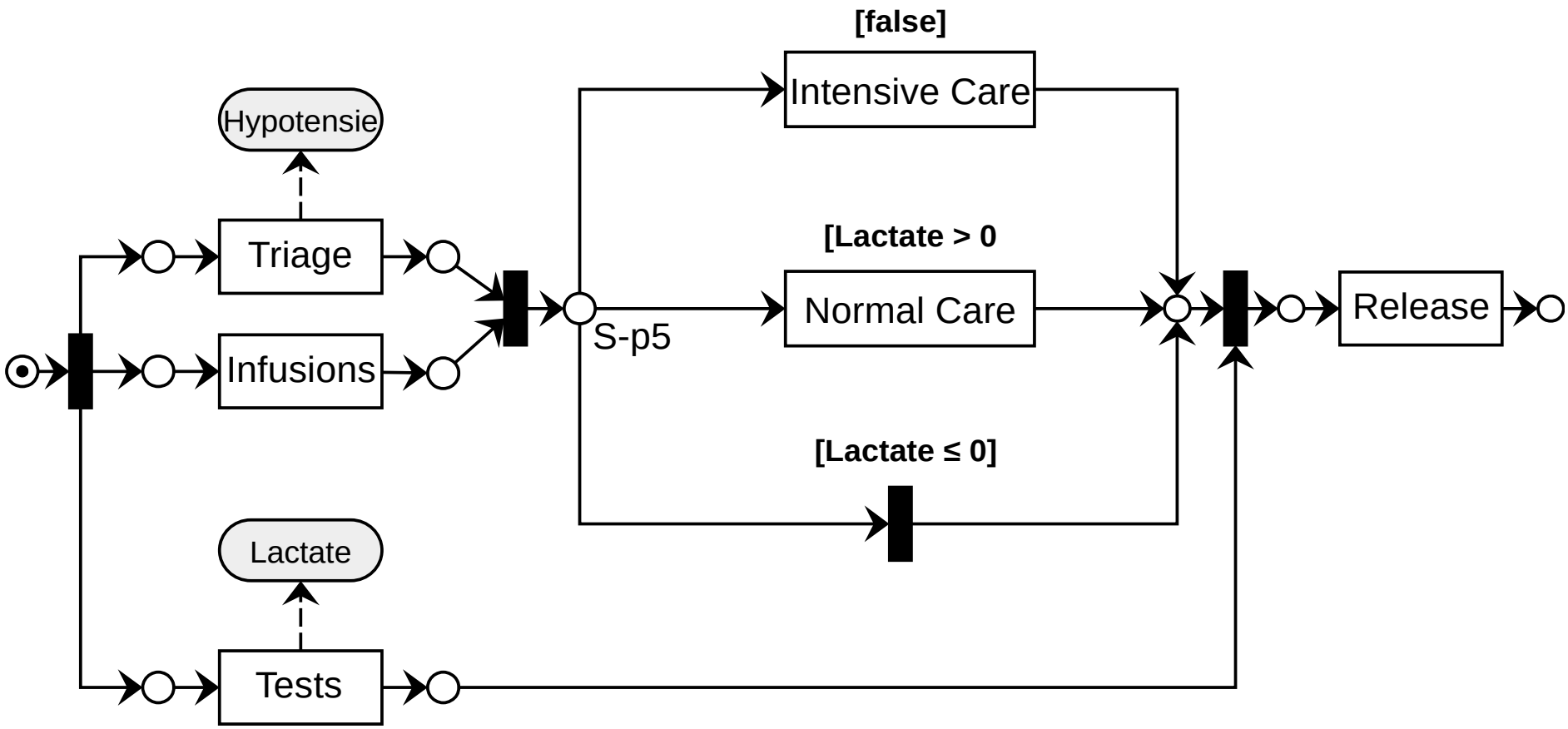
Example Rule SEPSIS - Overlapping



Example Rule SEPSIS - DTF

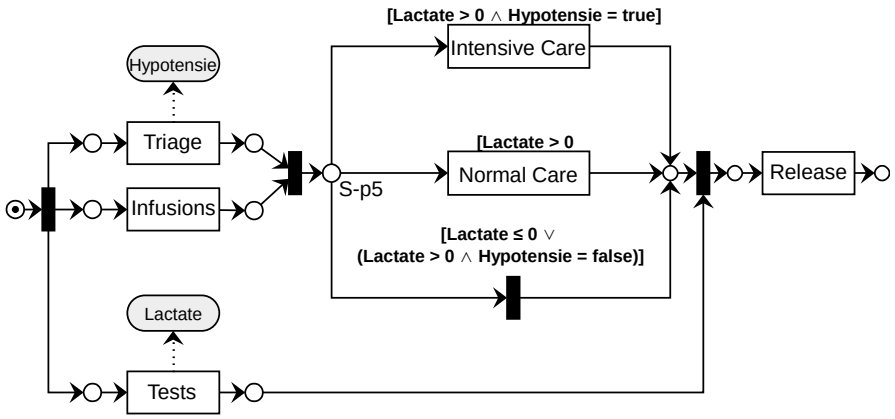


Example Rule SEPSIS - DTT

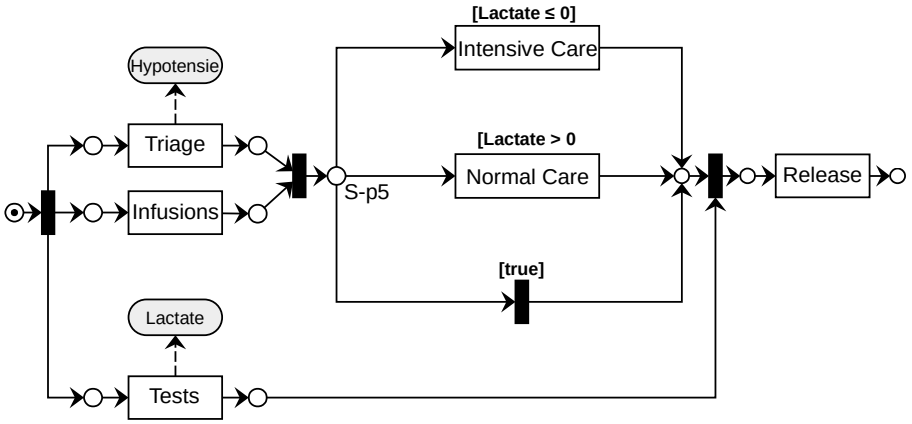


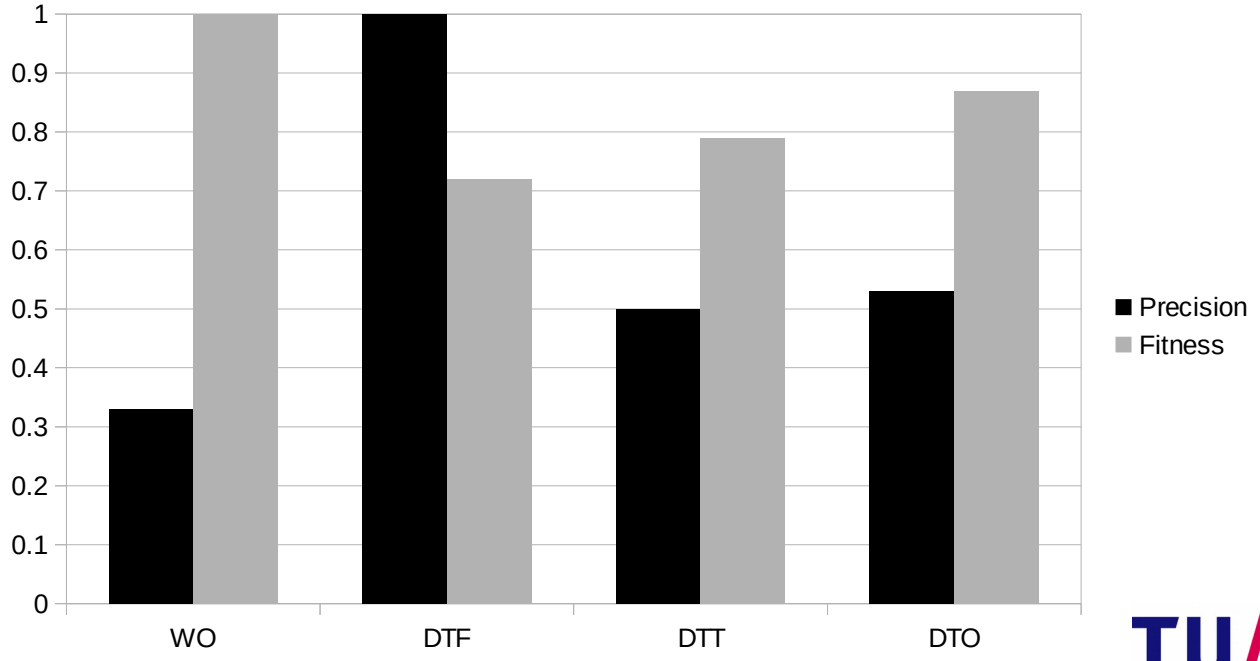
Example Rule SEPSIS – Precision and Fitness

Presented Approach (DTO)



Naive Approach (DTT)





Conclusion

- Discovery of Overlapping Rules in Event Data
  - Based on Decision Tree Induction
  - Implemented in *MultiPerspectiveExplorer (ProM)*
- Results
  - Improves model fitness over standard decision trees (DTF)
  - Improves model precision over the naive approach (DTT)
  - Limitations: Experimental validation & Complex rules
- Related work
  - Multi-label classification assumes instances to have multiple-labels
  - Association rule mining returns large set of rules
- Future work
  - Simplicity of Data-aware Process Models
  - Imbalanced Distribution of Classes

Questions? Remarks? Ideas?

Image source: http://commons.wikimedia.org/wiki/File:Pictofigo_-_Idea.png