

Exploring Context-Free Languages via Planning: The Case for Automating Machine Learning



M. Katz, P. Ram, S. Sohrabi, O. Udrea

Current automated data preparation & model selection

Bayesian opt, Genetic algo, HTN Planning, MCTS

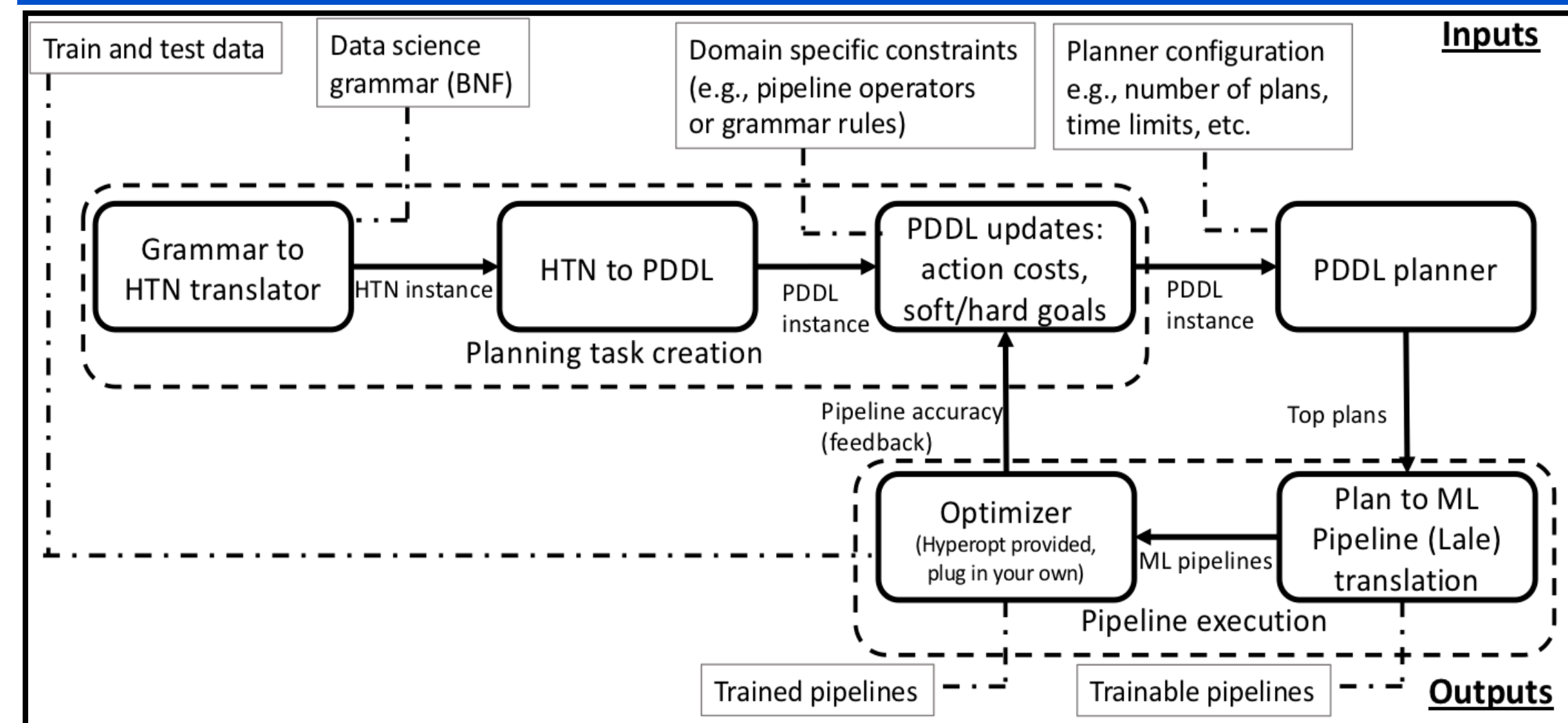
Limited to hand-crafted, chain-shaped pipelines; no user constraints

Pipeline Grammar

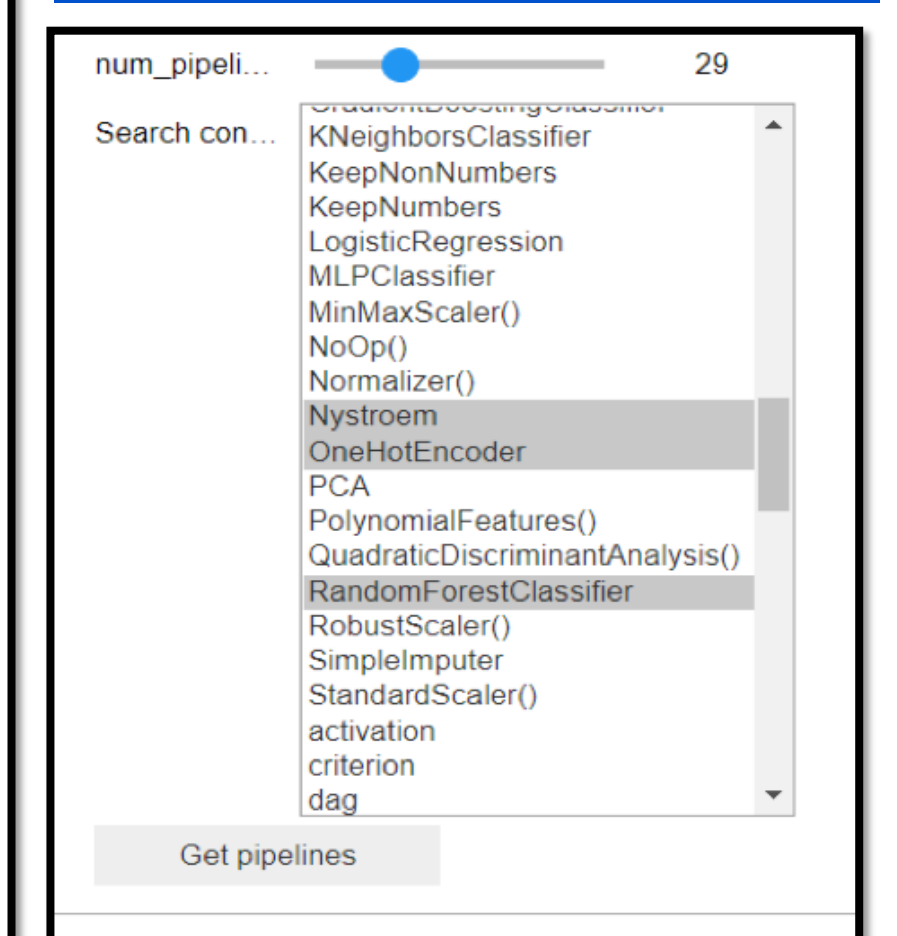
```

<start> → <dag>>><est>
<dag> → NoOp()|<est>|<tfm>|<dag>>><dag>
<dag> → (((<dag>))&((<dag>)))>>Concat()
<est> → <glm>|<mlpc>|<dtc>|<ebm>|<gnb>|<knc>|<qda>
<glm> → LogisticRegression(solver=<glmslv>, penalty=<glmpen>)
<glmpen> → 'l1'|'l2'
<glmslv> → 'newton-cg'|'sag'|'saga'|'lbfgs'|'liblinear'
<mlpc> → MLPClassifier(activation=<mlpca>, solver=<mlpcs>,
learning_rate=<mlpcl>)
<mlpca> → 'identity'|'relu'|'tanh'|'logistic'
<mlpcs> → 'lbfgs'|'sgd'|'adam'
<mlpcl> → 'constant'|'invscaling'|'adaptive'
<dtc> → DecisionTreeClassifier(criterion=<dtcc>, splitter=<dtcs>)
<dtcc> → 'gini'|'entropy'
<dtcs> → 'best'|'random'
<ebm> → <rfc>|<gbc>|<etc>
<rfc> → RandomForestClassifier(criterion=<rfcc>)
<rfcc> → 'gini'|'entropy'
<abc> → GradientBoostingClassifier(loss=<abcl>)
    
```

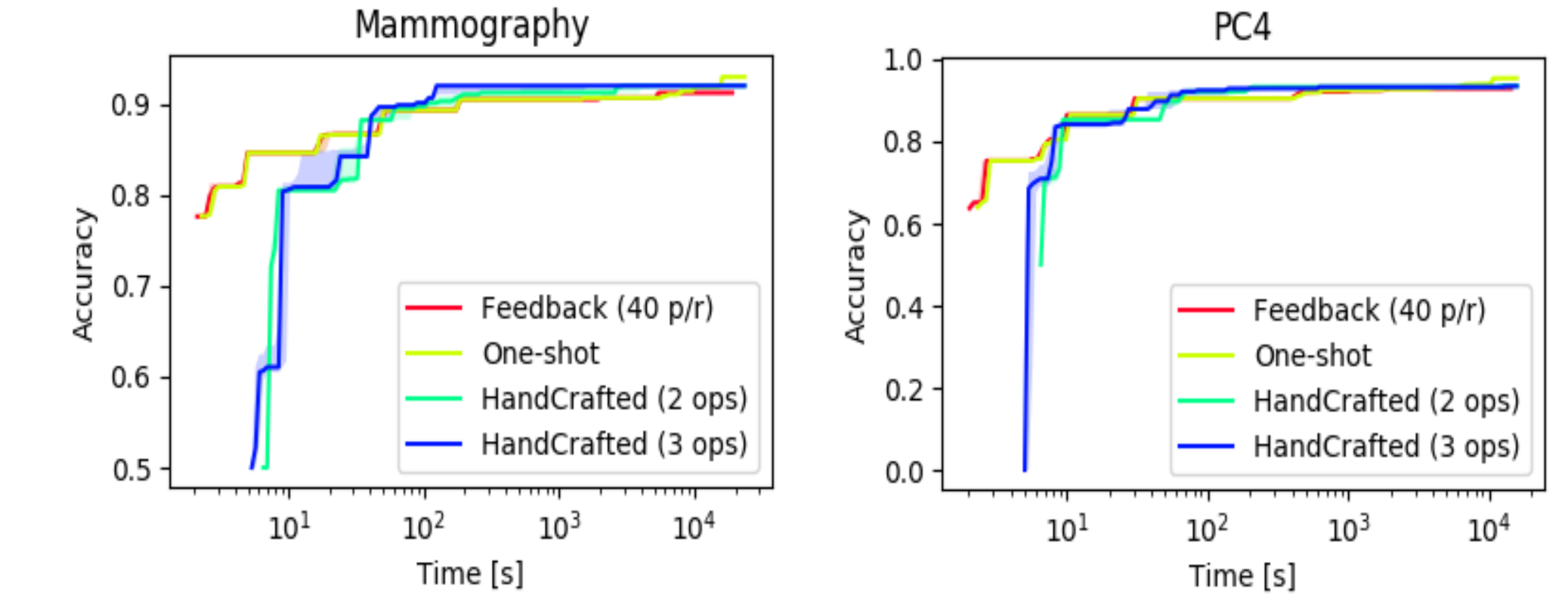
AI Planning based AutoML



User specified constraints



Competitive accuracy



Efficient scaling w. constraints

