

# Selection, Schedules and Configuration on *ASlib*

Albert-Ludwigs-Universität Freiburg



**UNI  
FREIBURG**

Marius Lindauer

Research Group on Learning, Optimization, and Automated Algorithm Design



- *ASlib* for benchmarking algorithm selectors



- *ASlib* for benchmarking algorithm selectors
- First results of *LLAMA* + *mlr* on *ASlib*

- *ASlib* for benchmarking algorithm selectors
- First results of *LLAMA* + *mlr* on *ASlib*
- Open Questions:
  - Performance of other selectors on *ASlib*?

- *ASlib* for benchmarking algorithm selectors
- First results of *LLAMA* + *mlr* on *ASlib*
- Open Questions:
  - Performance of other selectors on *ASlib*?
  - How important are pre-solving schedules on *ASlib*?

- *ASlib* for benchmarking algorithm selectors
- First results of *LLAMA* + *mlr* on *ASlib*
- Open Questions:
  - Performance of other selectors on *ASlib*?
  - How important are pre-solving schedules on *ASlib*?
  - How to use algorithm configuration to improve performance?



- *LLAMA*[Kotthoff 2013]+*mlr*[Bischl]: random forest regression, default features



- *LLAMA*[Kotthoff 2013]+*mlr*[Bischl]: random forest regression, default features
- *SATzilla'11*[Xu et al. 2011]: cost-sensitive random forest classification with 3 pre-solvers, selected features



- *LLAMA*[Kotthoff 2013]+*mlr*[Bischl]: random forest regression, default features
- *SATzilla'11*[Xu et al. 2011]: cost-sensitive random forest classification with 3 pre-solvers, selected features
- *SNNAP*[Collautti et al. 2013]: random forest regression + *k*-NN, all features

- *LLAMA*[Kotthoff 2013]+*mlr*[Bischl]: random forest regression, default features
- *SATzilla'11*[Xu et al. 2011]: cost-sensitive random forest classification with 3 pre-solvers, selected features
- *SNNAP*[Collautti et al. 2013]: random forest regression + *k*-NN, all features
- *ISAC*[Kadioglu et al. 2010]: clustering with *g*-means, all features

- *LLAMA*[Kotthoff 2013]+*mlr*[Bischl]: random forest regression, default features
- *SATzilla'11*[Xu et al. 2011]: cost-sensitive random forest classification with 3 pre-solvers, selected features
- *SNNAP*[Collautti et al. 2013]: random forest regression + *k*-NN, all features
- *ISAC*[Kadioglu et al. 2010]: clustering with *g*-means, all features
- *claspfolio 2*[Hoos et al. 2014]: cost-sensitive random forest classification with 3 pre-solvers default features

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	<b>285.3</b>

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	<b>285.3</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	<b>2185.1</b>

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	<b>285.3</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	<b>2185.1</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	<b>937.1</b>



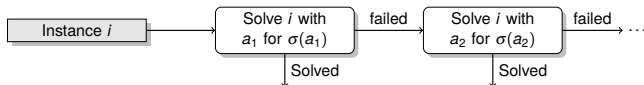
# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	<b>285.3</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	<b>2185.1</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	<b>937.1</b>
<i>SAT11-HAND</i>	9637.1	<b>6138.1</b>	9774	15650.2	7561.1
<i>SAT11-INDU</i>	7465.8	<b>5889.3</b>	6731	8078	7981.5
<i>SAT11-RAND</i>	4856.9	<b>990.2</b>	4658.3	5274.1	3593.5

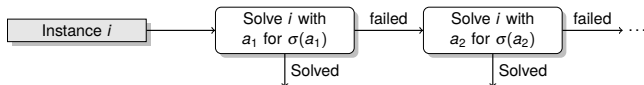
# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>
<i>ASP-POTASSCO</i>	<b>124.8</b>	NA	208.4	464.9	132.5
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	<b>285.3</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	<b>2185.1</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	<b>937.1</b>
<i>SAT11-HAND</i>	9637.1	<b>6138.1</b>	9774	15650.2	7561.1
<i>SAT11-INDU</i>	7465.8	<b>5889.3</b>	6731	8078	7981.5
<i>SAT11-RAND</i>	4856.9	<b>990.2</b>	4658.3	5274.1	3593.5
<i>SAT12-ALL</i>	1843.3	NA	<b>1344.1</b>	3015.6	1615.0
<i>SAT12-HAND</i>	2556.1	NA	2342.3	3814.6	<b>1859.9</b>
<i>SAT12-INDU</i>	1058.3	NA	<b>838.5</b>	1608.5	1182.8
<i>SAT12-RAND</i>	618.3	NA	700.4	<b>410.7</b>	695.8

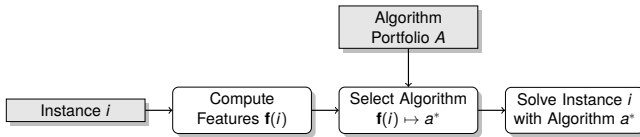
- algorithms schedules via *aspeed* [Hoos et al. 2014]:  
run algorithm  $a$  for at most  $\sigma(a)$  seconds



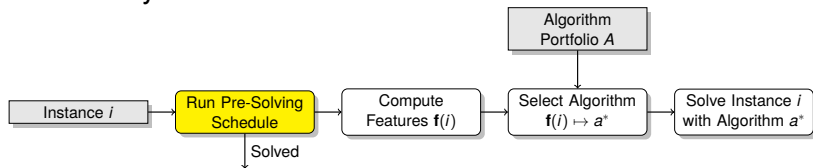
- algorithms schedules via *aspeed* [Hoos et al. 2014]:  
run algorithm  $a$  for at most  $\sigma(a)$  seconds



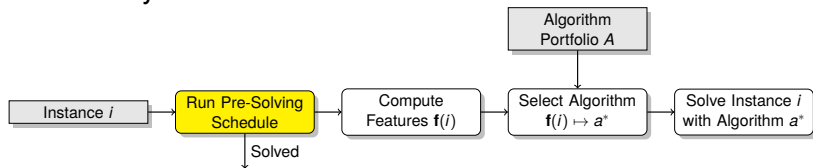
- How to integrate algorithm pre-solving schedules into algorithm selection frameworks?



*SATzilla* Style:



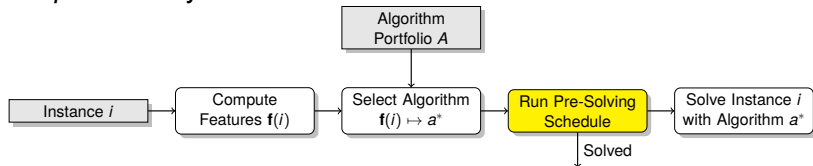
## SATzilla Style:



## Advantage:

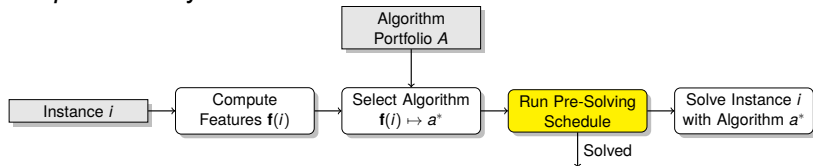
- No overhead due to feature computation if solved during pre-solving

*claspfolio 2* Style:





*clasptfolio 2* Style:



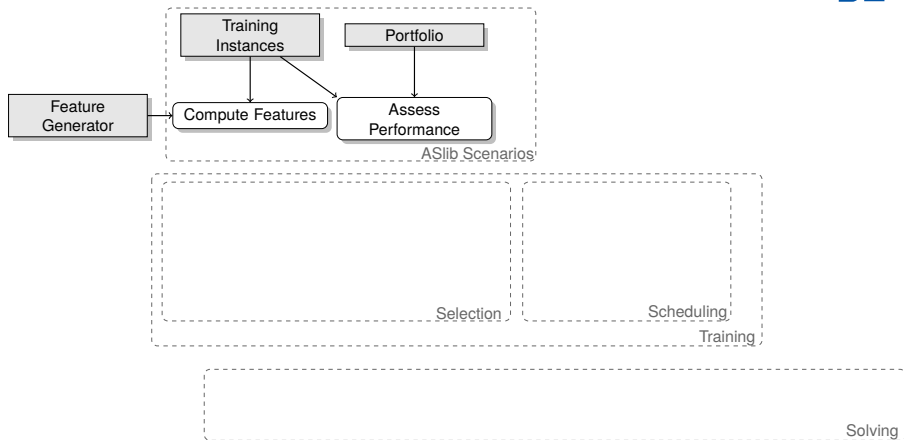
Advantage:

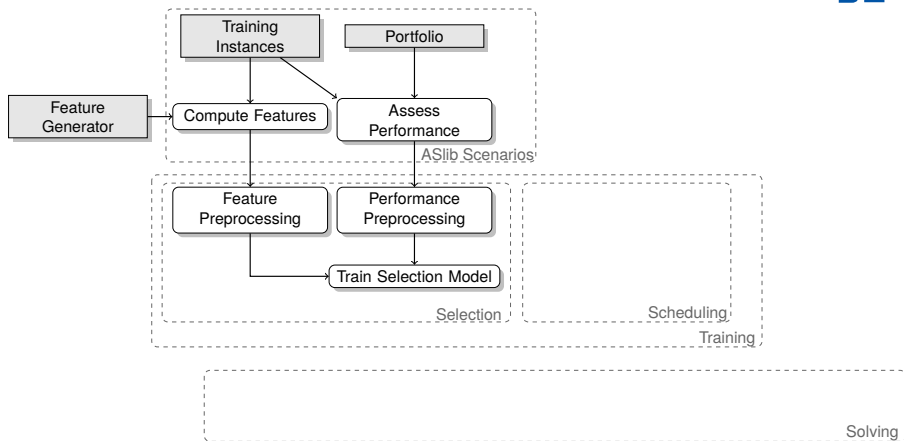
- Pre-solving schedule can be optimized, i.e., remove  $a^*$  from schedule

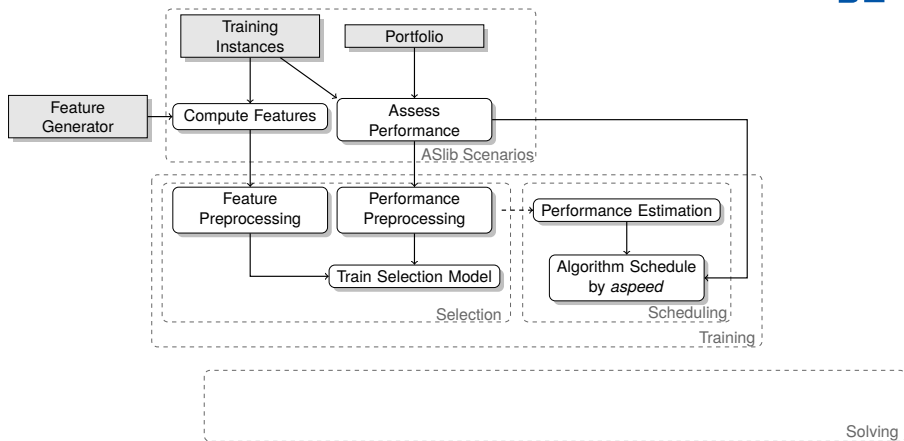
# *clasptfolio 2* Framework

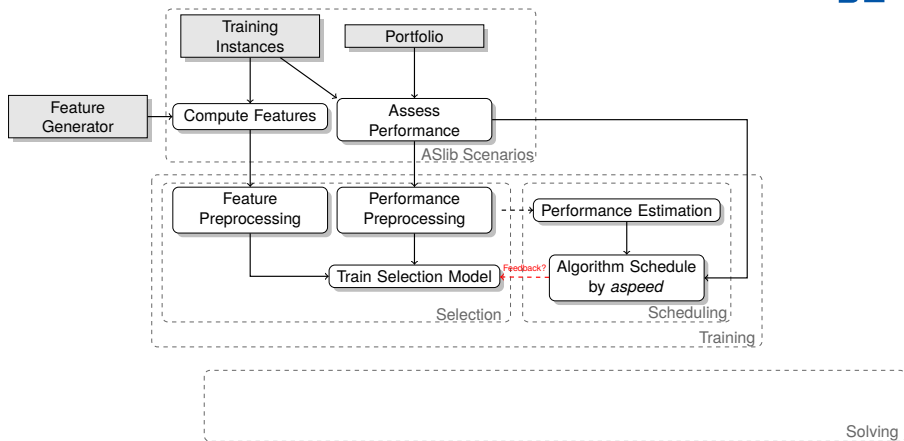


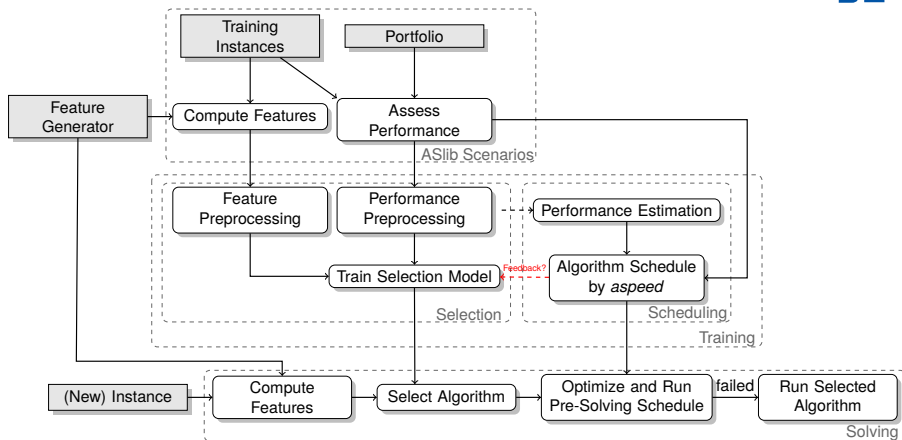
# claspfolio 2 Framework













- Pure Selection: no pre-solving



- Pure Selection: no pre-solving
- Early vs Late: First pre-solving or first selection?

- Pure Selection: no pre-solving
- Early vs Late: First pre-solving or first selection?
- Feedback vs. No Feedback from pre-solving schedule to the training of the selection model?

- Pure Selection: no pre-solving
- Early vs Late: First pre-solving or first selection?
- Feedback vs. No Feedback from pre-solving schedule to the training of the selection model?

## Benchmark:

- Cost-sensitive random forests
- At most 256 second for at most 3 pre-solvers

# Benchmarks of Pre-Solving Integration

	Pure Sel.	Early + No F.	Late + No F.	Early + F.	Late + F.
<i>ASP-POTASSCO</i>	<b>114.1</b>	132.3	132.2	132.3	132.2
<i>CSP-2010</i>	<b>349.9</b>	352.9	352.8	352.9	352.8
<i>MAXSAT12-PMS</i>	295.1	299.6	<b>281.0</b>	299.6	<b>281.0</b>
<i>PREMARSHALLING</i>	5083.0	2185.0	<b>2183.9</b>	2185.0	<b>2183.9</b>
<i>QBF-2011</i>	1279.1	1013.9	<b>924.6</b>	1013.9	<b>924.6</b>

# Benchmarks of Pre-Solving Integration

	Pure Sel.	Early + No F.	Late + No F.	Early + F.	Late + F.
<i>ASP-POTASSCO</i>	<b>114.1</b>	132.3	132.2	132.3	132.2
<i>CSP-2010</i>	<b>349.9</b>	352.9	352.8	352.9	352.8
<i>MAXSAT12-PMS</i>	295.1	299.6	<b>281.0</b>	299.6	<b>281.0</b>
<i>PREMARSHALLING</i>	5083.0	2185.0	<b>2183.9</b>	2185.0	<b>2183.9</b>
<i>QBF-2011</i>	1279.1	1013.9	<b>924.6</b>	1013.9	<b>924.6</b>
<i>SAT11-HAND</i>	14546.2	<b>7332.0</b>	7349.3	<b>7332.0</b>	7349.3
<i>SAT11-INDU</i>	<b>7824.4</b>	7846.6	7860.5	7846.6	7860.5
<i>SAT11-RAND</i>	4959.0	<b>3473.7</b>	3482.5	<b>3473.7</b>	3482.5

# Benchmarks of Pre-Solving Integration

	Pure Sel.	Early + No F.	Late + No F.	Early + F.	Late + F.
<i>ASP-POTASSCO</i>	<b>114.1</b>	132.3	132.2	132.3	132.2
<i>CSP-2010</i>	<b>349.9</b>	352.9	352.8	352.9	352.8
<i>MAXSAT12-PMS</i>	295.1	299.6	<b>281.0</b>	299.6	<b>281.0</b>
<i>PREMARSHALLING</i>	5083.0	2185.0	<b>2183.9</b>	2185.0	<b>2183.9</b>
<i>QBF-2011</i>	1279.1	1013.9	<b>924.6</b>	1013.9	<b>924.6</b>
<i>SAT11-HAND</i>	14546.2	<b>7332.0</b>	7349.3	<b>7332.0</b>	7349.3
<i>SAT11-INDU</i>	<b>7824.4</b>	7846.6	7860.5	7846.6	7860.5
<i>SAT11-RAND</i>	4959.0	<b>3473.7</b>	3482.5	<b>3473.7</b>	3482.5
<i>SAT12-ALL</i>	2055.4	<b>1657.9</b>	1691.4	<b>1657.9</b>	1662.8
<i>SAT12-HAND</i>	2923.1	<b>1841.7</b>	1850.8	<b>1841.7</b>	1850.8
<i>SAT12-INDU</i>	<b>1042.0</b>	1050.0	1049.4	1051.1	1048.5
<i>SAT12-RAND</i>	<b>619.5</b>	666.1	672.1	666.1	672.1

# Benchmarks of Pre-Solving Integration

	Pure Sel.	Early + No F.	Late + No F.	Early + F.	Late + F.
<i>ASP-POTASSCO</i>	<b>114.1</b>	132.3	132.2	132.3	132.2
<i>CSP-2010</i>	<b>349.9</b>	352.9	352.8	352.9	352.8
<i>MAXSAT12-PMS</i>	295.1	299.6	<b>281.0</b>	299.6	<b>281.0</b>
<i>PREMARSHALLING</i>	5083.0	2185.0	<b>2183.9</b>	2185.0	<b>2183.9</b>
<i>QBF-2011</i>	1279.1	1013.9	<b>924.6</b>	1013.9	<b>924.6</b>
<i>SAT11-HAND</i>	14546.2	<b>7332.0</b>	7349.3	<b>7332.0</b>	7349.3
<i>SAT11-INDU</i>	<b>7824.4</b>	7846.6	7860.5	7846.6	7860.5
<i>SAT11-RAND</i>	4959.0	<b>3473.7</b>	3482.5	<b>3473.7</b>	3482.5
<i>SAT12-ALL</i>	2055.4	<b>1657.9</b>	1691.4	<b>1657.9</b>	1662.8
<i>SAT12-HAND</i>	2923.1	<b>1841.7</b>	1850.8	<b>1841.7</b>	1850.8
<i>SAT12-INDU</i>	<b>1042.0</b>	1050.0	1049.4	1051.1	1048.5
<i>SAT12-RAND</i>	<b>619.5</b>	666.1	672.1	666.1	672.1

- Pre-Solving is important on 6 scenarios

# Benchmarks of Pre-Solving Integration

	Pure Sel.	Early + No F.	Late + No F.	Early + F.	Late + F.
<i>ASP-POTASSCO</i>	<b>114.1</b>	132.3	132.2	132.3	132.2
<i>CSP-2010</i>	<b>349.9</b>	352.9	352.8	352.9	352.8
<i>MAXSAT12-PMS</i>	295.1	299.6	<b>281.0</b>	299.6	<b>281.0</b>
<i>PREMARSHALLING</i>	5083.0	2185.0	<b>2183.9</b>	2185.0	<b>2183.9</b>
<i>QBF-2011</i>	1279.1	1013.9	<b>924.6</b>	1013.9	<b>924.6</b>
<i>SAT11-HAND</i>	14546.2	<b>7332.0</b>	7349.3	<b>7332.0</b>	7349.3
<i>SAT11-INDU</i>	<b>7824.4</b>	7846.6	7860.5	7846.6	7860.5
<i>SAT11-RAND</i>	4959.0	<b>3473.7</b>	3482.5	<b>3473.7</b>	3482.5
<i>SAT12-ALL</i>	2055.4	<b>1657.9</b>	1691.4	<b>1657.9</b>	1662.8
<i>SAT12-HAND</i>	2923.1	<b>1841.7</b>	1850.8	<b>1841.7</b>	1850.8
<i>SAT12-INDU</i>	<b>1042.0</b>	1050.0	1049.4	1051.1	1048.5
<i>SAT12-RAND</i>	<b>619.5</b>	666.1	672.1	666.1	672.1

- Pre-Solving is important on 6 scenarios
- Difference between pre-solving approaches is small



# How to improve further?

- *claspfolio 2* implements many different strategies.

# How to improve further?

- *claspfolio 2* implements many different strategies.
- *claspfolio 2* can be a lot better!

# How to improve further?

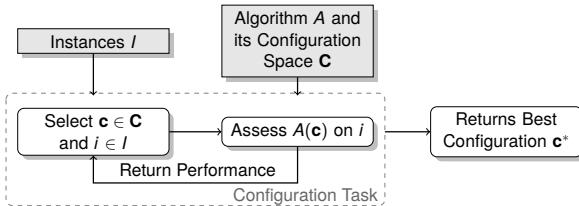
- *claspfolio 2* implements many different strategies.
- *claspfolio 2* can be a lot better!
- How to choose a well-performing configuration of *claspfolio 2*?

# How to improve further?

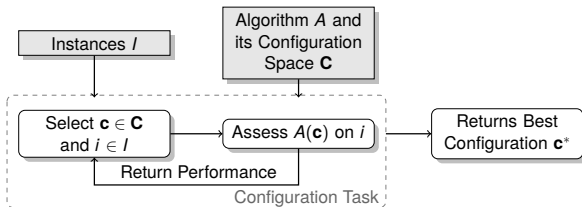
- *claspfolio 2* implements many different strategies.
- *claspfolio 2* can be a lot better!
- How to choose a well-performing configuration of *claspfolio 2*?

→ Algorithm Configuration!

# How to configure an algorithm selector?

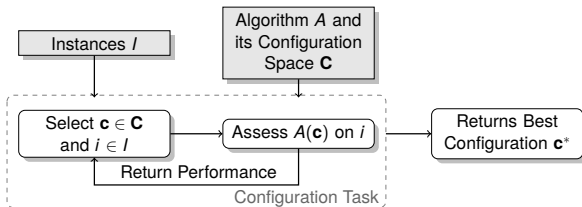


# How to configure an algorithm selector?



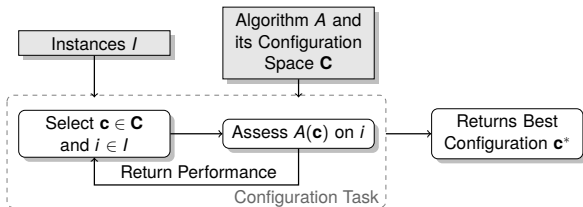
## 1 Parametrize your selector

# How to configure an algorithm selector?



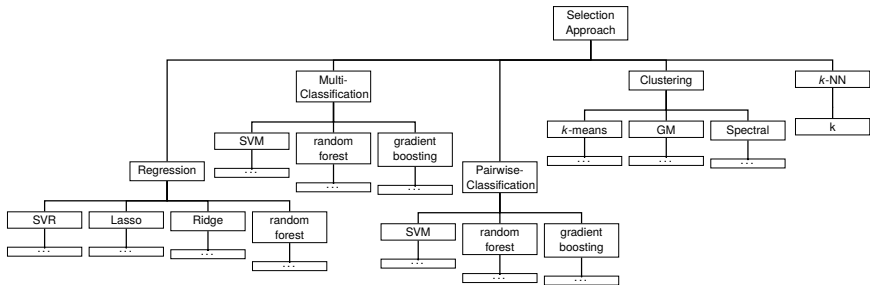
- 1 Parametrize your selector
- 2 Define the configuration space

# How to configure an algorithm selector?



- 1 Parametrize your selector
- 2 Define the configuration space
- 3 Define meta instances for configuration process







## Feature Preprocessing:

- Normalization
- PCA
- Imputation
- Maximal runtime to compute features
- Subset of feature groups

## Feature Preprocessing:

- Normalization
- PCA
- Imputation
- Maximal runtime to compute features
- Subset of feature groups

## Performance Preprocessing:

- Transformation (e.g., log)
- Instance weighting
- Solver subset selection (based on marginal contribution)

## Feature Preprocessing:

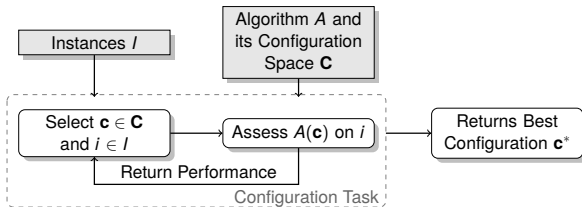
- Normalization
- PCA
- Imputation
- Maximal runtime to compute features
- Subset of feature groups

## Performance Preprocessing:

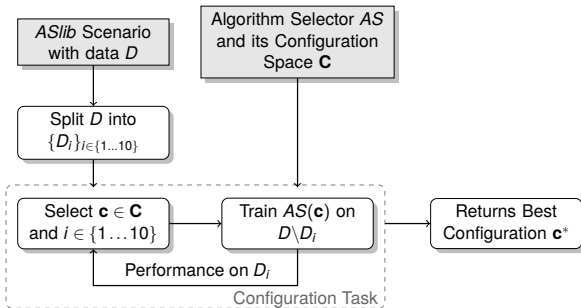
- Transformation (e.g., log)
- Instance weighting
- Solver subset selection (based on marginal contribution)

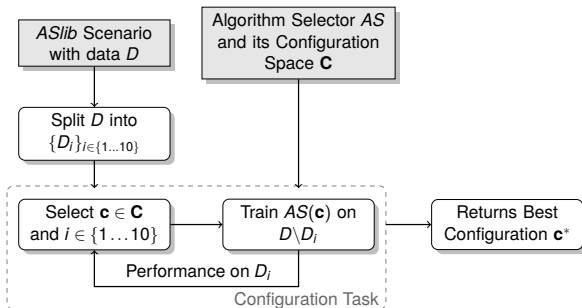
## Pre-Solving:

- Maximal number of solvers
- Maximal pre-solving runtime
- Approach



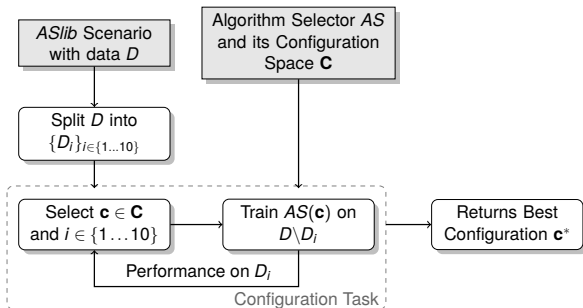
# Configuration of Selectors





## Remarks:

- If we assess  $AS(\mathbf{c})$  on all  $i \in \{1 \dots 10\}$ , we have performed a 10-fold cross validation



## Remarks:

- If we assess  $AS(\mathbf{c})$  on all  $i \in \{1 \dots 10\}$ , we have performed a 10-fold cross validation
- Some configurators can discard configurations when they perform poorly on a subset of instances



*AutoFolio* = Configurator + Selector

*AutoFolio* = Configurator + Selector

First Prototype:

*AutoFolio* = [SMAC | ParamILS] + *claspfolio 2*

*AutoFolio* = Configurator + Selector

First Prototype:

*AutoFolio* = [SMAC | ParamILS] + *claspfolio 2*

Experimental setup:

- 12 *ASlib* scenarios
  - 2 configurators
  - 16 independent runs
  - 2 days as configuration budget
  - 10-fold **outer** cross validation
- 21.8 CPU years

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>	<i>AutoFolio</i>
<i>ASP-POTASSCO</i>	124.8	NA	208.4	464.9	132.5	<b>123</b>
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>	413.3
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	285.3	<b>169.9</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	2185.1	<b>1663.7</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	937.1	<b>924.9</b>

# Comparison on *ASlib* (PAR10)

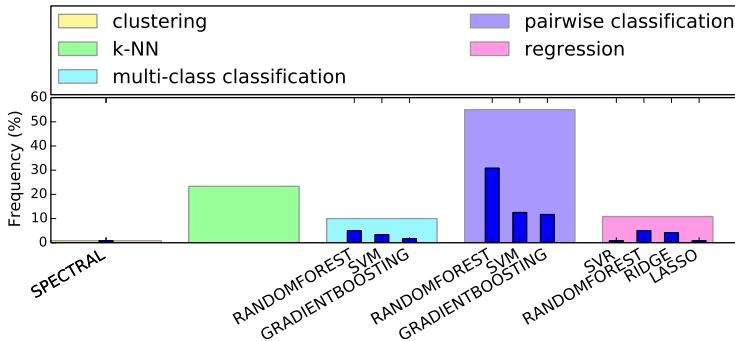


	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>	<i>AutoFolio</i>
<i>ASP-POTASSCO</i>	124.8	NA	208.4	464.9	132.5	<b>123</b>
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>	413.3
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	285.3	<b>169.9</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	2185.1	<b>1663.7</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	937.1	<b>924.9</b>
<i>SAT11-HAND</i>	9637.1	6138.1	9774	15650.2	7561.1	<b>5935.8</b>
<i>SAT11-INDU</i>	7465.8	<b>5889.3</b>	6731	8078	7981.5	7536.8
<i>SAT11-RAND</i>	4856.9	<b>990.2</b>	4658.3	5274.1	3593.5	1294.8

# Comparison on *ASlib* (PAR10)

	<i>LLAMA</i> (RF)	<i>SATzilla'11</i>	<i>SNNAP</i>	<i>ISAC</i>	<i>claspfolio 2</i>	<i>AutoFolio</i>
<i>ASP-POTASSCO</i>	124.8	NA	208.4	464.9	132.5	<b>123</b>
<i>CSP-2010</i>	378	NA	1087.5	1098.4	<b>356.9</b>	413.3
<i>MAXSAT12-PMS</i>	294.5	NA	867.4	1004.1	285.3	<b>169.9</b>
<i>PREMARSHALLING</i>	3921.9	NA	9657.8	5073.3	2185.1	<b>1663.7</b>
<i>QBF-2011</i>	1038.9	NA	7241	3951.2	937.1	<b>924.9</b>
<i>SAT11-HAND</i>	9637.1	6138.1	9774	15650.2	7561.1	<b>5935.8</b>
<i>SAT11-INDU</i>	7465.8	<b>5889.3</b>	6731	8078	7981.5	7536.8
<i>SAT11-RAND</i>	4856.9	<b>990.2</b>	4658.3	5274.1	3593.5	1294.8
<i>SAT12-ALL</i>	1843.3	NA	1344.1	3015.6	1615.0	<b>925.8</b>
<i>SAT12-HAND</i>	2556.1	NA	2342.3	3814.6	1859.9	<b>968.2</b>
<i>SAT12-INDU</i>	1058.3	NA	838.5	1608.5	1182.8	<b>638.3</b>
<i>SAT12-RAND</i>	618.3	NA	700.4	410.7	695.8	<b>382.9</b>

# Selected Configurations



- Over-tuning on some scenarios (e.g., *CSP-2010*)



- Over-tuning on some scenarios (e.g., *CSP-2010*)
- *claspfolio 2* can be better than *SATzilla'11* also on SAT11 scenarios
- Restricted configuration space of *claspfolio 2* can improve the performance of *AutoFolio* on SAT11 scenarios

- Over-tuning on some scenarios (e.g., *CSP-2010*)
- *claspfolio 2* can be better than *SATzilla'11* also on SAT11 scenarios
- Restricted configuration space of *claspfolio 2* can improve the performance of *AutoFolio* on SAT11 scenarios
- Reduction of configuration budget
- We do not yet know how far we can decrease the budget without losing performance



- There is no silver bullet for *ASlib*

- There is no silver bullet for *ASlib*
- Pre-solving is important on some of the scenarios
- Implementation details are not so important

- There is no silver bullet for *ASlib*
- Pre-solving is important on some of the scenarios
- Implementation details are not so important
- Configurator automatically finds well performing configurations of selectors

- There is no silver bullet for *ASlib*
- Pre-solving is important on some of the scenarios
- Implementation details are not so important
- Configurator automatically finds well performing configurations of selectors

Thank you!