# Mining IPC-2011 Results

Isabel Cenamor
Tomás de la Rosa
Fernando Fernández

June 25, 2012

Universidad Carlos III de Madrid

# Motivation

- Each competition (IPC) produces a big amount of data

# Motivation

- Each competition (IPC) produces a big amount of data
- This data opens a wide variety of analysis from a Data Mining perspective

# Motivation

- Each competition (IPC) produces a big amount of data
- This data opens a wide variety of analysis from a Data Mining perspective
- The results of the analysis can help us to find some insights about the performance of the planners

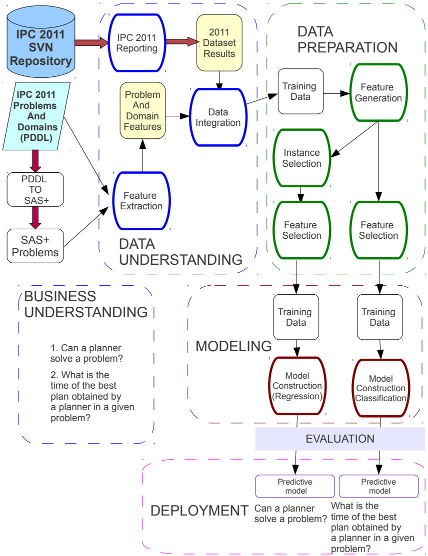
# Motivation

- Each competition (IPC) produces a big amount of data
- This data opens a wide variety of analysis from a Data Mining perspective
- The results of the analysis can help us to find some insights about the performance of the planners
- And can be used to configure a portfolio of planners that takes into account the particular features of a planning problem

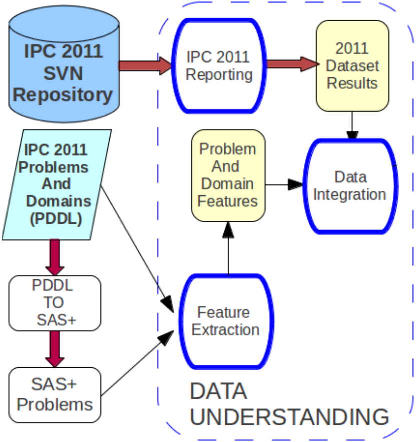
# Target

- It is posible to generate a model that predicts:
  - If a planner will be able to find a solution
  - How long it will take

# General Description



# Data Understanding




We have used all the problems from the Sequential Satisficing and Sequential Optimization tracks:

1. Processing PDDL to SAS+
2. Extraction of features from the problems
3. Extraction of the results of the last competition
4. Data Integration.

# Features

The features have two different sources:

1. The IPC 2011 Results
2. The IPC 2011 Domains and Problems

## Total Instances

- Seq-sat has 7560 instances: 27 planners with 20 problems in 14 domains (3837 solved / 3723 unsolved)
- Seq-opt has 3360 instances: 12 planners with 20 problems in 14 domains (1831 solved / 1529 unsolved)

# The IPC 2011 Results

These features are a subset of the elementary variables offered by the software of the IPC:

1. Planner
2. Domain
3. Problem
4. Time vector (CPU time of each solution found)
5. Quality vector (Plan quality of each solution found)

# The IPC 2011 Domains and Problems

The objective of this process is the characterization of the problem. These features are divided in:

1. Basic: based on PDDL
2. Elaborated: based on SAS+

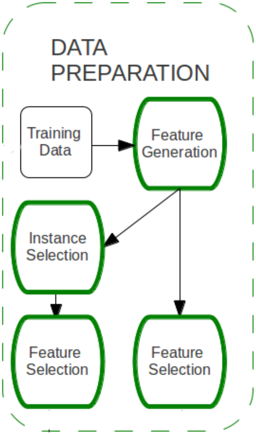
The size of the set of features extracted is 47.

# Elaborated Features

Based on SAS+:

- Based on Causal Graph (CG)
  - General(4)
  - General Ratios (4)
  - High Level Statistics Information (6)
  - Topology Statistics Information(12)
- Based on Transition Graph (DTG)
  - General (3)
  - Topology Statistics Information (12)

# Data Preparation




With the data set created in the previous step:

1. We estimate output attributes:
   - Solution
   - Time of first solution
   - Quality of first solution
   - Time of median solution
   - Quality of median solution
   - Time of best solution
   - Quality of best solution
2. Automatic Selection of Features

# Data Modeling

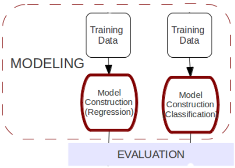



Figure: Data Modeling

Different sets based on the prediction variable:

1. Classification $\longrightarrow$ Solution?
2. Regression:
   - Time of the first solution
   - Median time of the solutions
   - Execution time of the best solution

# Algorithms

We used Weka Software in the modeling process:

- Classification
  - Decision Tree (J48)
  - Support Vector Machine (SMO)
  - Instance Based Learning Algorithm (IBK)
- Regression
  - Regression Rules (M5Rules)
  - Support Vector Machine (SMO)
  - Instance Based Learning Algorithm (IBK)

# Metric Used

- $Accuracy \;=\; \left(\frac{number\ \ TP + number\ \ TN}{Total}\right)$
- $RelativeError \;=\; \frac{Absolute\ \ Error}{Real\ \ Value}$

# Model evaluation

Using data from the competition we have taken the classes for the models

- Is the estimation valid for new problems in the same domains seen in the IPC 2011?

# Model evaluation

Using data from the competition we have taken the classes for the models

- Is the estimation valid for new problems in the same domains seen in the IPC 2011?
- Yes , with Cross Validation

# Model evaluation

Using data from the competition we have taken the classes for the models

- Is the estimation valid for new problems in the same domains seen in the IPC 2011?
- Yes , with Cross Validation
- Is the estimation valid for new problems in domains differents to the IPC 2011 ones?

# Model evaluation

Using data from the competition we have taken the classes for the models

- Is the estimation valid for new problems in the same domains seen in the IPC 2011?
- Yes , with Cross Validation
- Is the estimation valid for new problems in domains differents to the IPC 2011 ones?
- Yes, with Leave - one - domain - out

# Cross Validation

Cross Validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set.

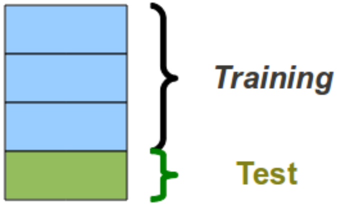



Figure: Cross Validation I

# Cross Validation

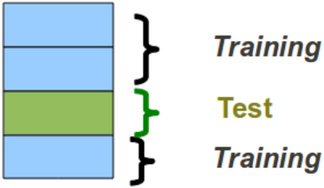



Figure: Cross Validation II

# Cross Validation

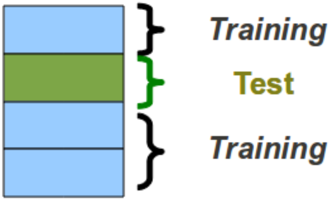



Figure: Cross Validation III

# Cross Validation

The error is the mean of the evaluations

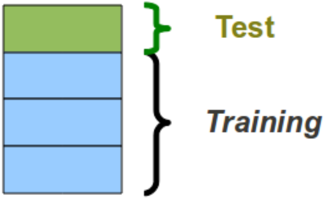



Figure: Cross Validation IV

# Leave - one - domain - out

This is the same as a K-fold cross-validation with K being equal to the number of observations in the original sample. (Domains)

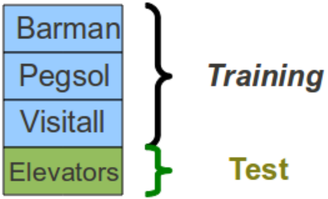



Figure: Leave - one - domain - out I

# Leave - one - domain - out

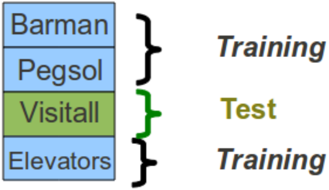



Figure: Leave - one - domain - out II

# Leave - one - domain - out

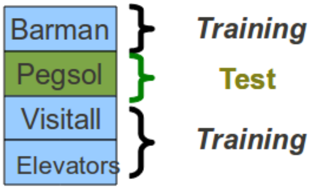



Figure: Leave - one - domain - out III

# Leave - one - domain - out

The error is the mean of the evaluations

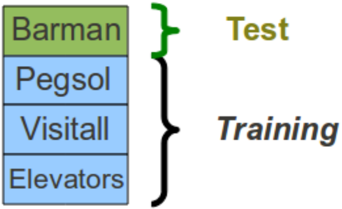



Figure: Leave - one - domain - out IV

# Seq-sat Classification

| Dataset | Cross Validation | Leave Domain Out |
|---|---|---|
| J48 | **88.75(1.05)** | 59.14(12.13) |
| IBk -K 1 | 88.67(1.29) | 60.83(10.13) |
| IBk -K 3 | 87.63(1.07) | 60.58(11.76) |
| IBk -K 5 | 88.58(1.07) | **61.95(11.10)** |
| SMO | 72.48(1.58) | 61.34(10.10) |

# Seq-opt Classification

| Dataset | Cross Validation | Leave Domain Out |
|---|---|---|
| J48 | **90.14(1.58)** | 60.36 (23.69) |
| IBk -K 1 | 86.96(1.57) | 60.36 (21.26) |
| IBk -K 3 | 87.81(1.81) | 58.78 (21.66) |
| IBk -K 5 | 83.91(1.90) | 60.86 (20.53) |
| SMO | 79.96(2.30) | **67.41 (16.55)** |

# Seq-sat Regression(I)

| Dataset | Cross Validation | | |
|---|---|---|---|
| | First Time | Median Time | Best Time |
| M5Rules | 73.81(4.78) | 74.02(3.90) | 73.66(3.61) |
| IBk -K 1 | 59.84(5.15) | 65.25(5.28) | 67.57(4.07) |
| IBk -K 3 | **55.05(3.72)** | **60.02(4.00)** | **62.98(3.12)** |
| IBk -K 5 | 56.61(3.66) | 60.93(3.51) | 64.39(3.00) |
| SMOreg | 60.18(4.06) | 64.08(3.65) | 69.50(2.87) |

# Seq-sat Regression(II)

| Dataset | Leave Domain Out | | |
|---|---|---|---|
| | First Time | Median Time | Best Time |
| M5Rules | 17204.81(60518.16) | 1492.24(2798.89) | 985.64(2200.93) |
| IBk -K 1 | 87.94(30.76) | 91.12(29.39) | 93.66(23.38) |
| IBk -K 3 | **79.31(28.27)** | 89.87(31.70) | 85.96(22.26) |
| IBk -K 5 | 92.12(29.73) | **89.70(26.57)** | **85.57(19.21)** |
| SMOreg | 835.17(2264.22) | 184.10(165.75) | 907.32(2620.74) |

# Seq-opt Regression

| Dataset | Cross Validation | Leave Domain Out |
|---|---|---|
| M5Rules | 67.08(7.63) | 213.87 (231.95) |
| IBk -K 1 | **59.74(8.37)** | 141.54 (47.40) |
| IBk -K 3 | 59.99(6.32) | **123.37 (11.26)** |
| IBk -K 5 | 63.59(6.38) | 127.21 (10.96) |
| SMOreg | 66.84(5.71) | 15151.04 (54178.83) |

# Different classification accuracies achieved with individual models

| Planners | | Accuracy |
|---|---|---|
| Lama-2008 | | 81,43 ±6,35 |
| Lamar | | 81,43 ±5,71 |
| Satplanlm-c | | 86,79±5,99 |
| Forkuniform | | 88,93±3,73 |
| Cpt4 | | 92,5±4,36 |
| Minimum | Fd-autotune2 | 78,2 |
| Maximum | Acoplan, Acoplan2 | 97,5 |
| Average | – | 88,5 ± 5,3 |
| Track Winner | Lama-2011 | 81,4 |

# Conclusions

- In this analysis we have given some insights about the performance of planners

# Conclusions

- In this analysis we have given some insights about the performance of planners
- We have created classification models for predicting whether a planner will succeed or not in a given problem

# Conclusions

- In this analysis we have given some insights about the performance of planners
- We have created classification models for predicting whether a planner will succeed or not in a given problem
- And we have created regression models for predicting the time a planner will need to solve the problem

# Conclusions

- The leave one doamin out evaluation is an alternative to estimate how good the learned models in unknown domain

# Conclusions

- The leave one doamin out evaluation is an alternative to estimate how good the learned models in unknown domain
- The results on known domains have a good accuracy

# Conclusions

- The leave one doamin out evaluation is an alternative to estimate how good the learned models in unknown domain
- The results on known domains have a good accuracy
- But it seems that this does not hold in unknown domains

# Conclusions

- The leave one doamin out evaluation is an alternative to estimate how good the learned models in unknown domain
- The results on known domains have a good accuracy
- But it seems that this does not hold in unknown domains
- The results have shown that the elaborated features are relevant for partially characterizing the complexity of planning problems

# Future Work

- Creating new feature to improve the results in regression models

# Future Work

- Creating new feature to improve the results in regression models
- Developing a portfolio of planner with the created models