Parametric design of a family of centrifugal pumps: dealing with unbalancedness of geometries dataset.

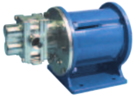
Elisa Ricciettti
Università degli Studi di Firenze
Dipartimento di Matematica e Informatica 'Ulisse Dini'

Joint work with J.Bellucci, M.Checcucci, M.Marconcini, A.Arnone, Dipartimento di Ingegneria Industriale (DIEF)

OBA Summer School 02/07/2017-08/07/2017.

Industrial application:
Enlargement of the pumps catalogue for a company.

H-Series Specifications

SEALED Long-Coupled							
MAG-DRIVE or SEALED Close-Coupled							
SEALED MAG-DRIVE	**H1F H1F-MC**	**H3F H3F-MC**	**H5R H5R-MC**	**H5F H5F-MC**	**H7N H7N-MC**	**H7R H7R-MC**	**H7F H7F-MC**
Max Flow Rate	**0.5** GPM (**1.9** LPM)	**1.4** GPM (**5.3** LPM)	**2.4** GPM (**9.1** LPM)	**3.4** GPM (**13** LPM)	**5.4** GPM (**20** LPM)	**8.6** GPM (**33** LPM)	**10.7** GPM (**40.5** LPM)
Max Diff. Press.	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)
Max Discharge	300 PSI (20.7 BAR)	300 PSI (20.7 BAR)	300 PSI (20.7 BAR)	300 PSI (20.7 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)	225 PSI (15.5 BAR)
Max Temp.	500 °F (260 °C)	500 °F (260 °C)	500 °F (260 °C)	500 °F (260 °C)	500 °F (260 °C)	500 °F (260 °C)	500 °F (260 °C)
Max Viscosity	100,000* CPS	100,000* CPS	100,000* CPS	100,000* CPS	100,000* CPS	100,000* CPS	100,000* CPS
Max Speed	1750 RPM	1750 RPM	1750 RPM	1750 RPM	1750 RPM	1750 RPM	1750 RPM
NPSHR @ Max Speed	3 FT (0.9M)	2 FT (0.6M)	2 FT (0.6M)	2 FT (0.6M)	5.2 FT (1.6M)	5.2 FT (1.6M)	5.2 FT (1.6M)
Weight Sealed, LC Sealed, CC Mag-Drive, CC	2.5 LBS (1.1 KGS) 23 LBS (10 KGS) 31 LBS (14 KGS)	2.5 LBS (1.1 KGS) 23 LBS (10 KGS) 31 LBS (14 KGS)	3.5 LBS (1.6 KGS) 24 LBS (11 KGS) 32 LBS (15 KGS)	3.5 LBS (1.6 KGS) 24 LBS (11 KGS) 32 LBS (15 KGS)	6.5 LBS (2.9 KGS) 29 LBS (13 KGS) 36 LBS (16 KGS)	6.5 LBS (2.9 KGS) 29 LBS (13 KGS) 36 LBS (16 KGS)	6.5 LBS (2.9 KGS) 29 LBS (13 KGS) 36 LBS (16 KGS)

Parametric Design of a new pump: key ingredients

- **Geometry parameterization**: choice of n independent parameters (*degrees of freedom*) that describe the pump geometry: p_1, p_2, ..., p_n.

  A pump: $P = (p_1, p_2, \dots, p_n)$.

- **Design space**: individuate range of variation of parameters: $p_{i\,\mathrm{min}} \le p_i \le p_{i\,\mathrm{max}}$, $i = 1, \dots, n$.

  Design space: $\mathcal{S} = [p_{1\,\mathrm{min}}, p_{1\,\mathrm{max}}] \times \cdots \times [p_{n\,\mathrm{min}}, p_{n\,\mathrm{max}}]$,

- **Performance functions:** Individuation of functions to measure a pump performance (efficiency, flow rate...)

  Performance function $F = (f_1, \dots, f_h)$.

Parametric Design of a new pump: key ingredients

- **Geometry parameterization**: choice of n independent parameters (*degrees of freedom*) that describe the pump geometry: p_1, p_2, ..., p_n.

  A pump: $P = (p_1, p_2, \dots, p_n)$.

- **Design space**: individuate range of variation of parameters: $p_{i\,\min} \leq p_i \leq p_{i\,\max}$, $i = 1, \dots, n$.

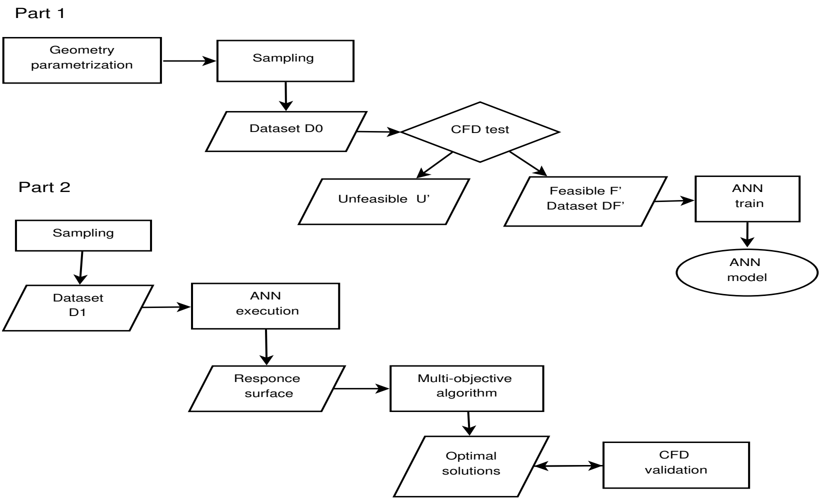
  Design space: $\mathcal{S} = [p_{1\,\min}, p_{1\,\max}] \times \cdots \times [p_{n\,\min}, p_{n\,\max}]$,

- **Performance functions:** Individuation of functions to measure a pump performance (efficiency, flow rate...)

  Performance function $F = (f_1, \dots, f_h)$.

Computation through CFD (Computational Fluid Dynamics) expensive! $\rightarrow$ regression meta-model (ANN)

Single pump design scheme



Single pump design vs family design

Single pump design:

- The redesign starts from a baseline configuration geometrically close to the final one.
- The number of parametres is small (~ 10)
- All the tools are finely tuned for the specific application. (e.g. geometry parameterization, mesh generation, CFD solver etc.)
- All the geometrical constraints can be a-priori taken into account.

Family design:

- The design starts from scratch.
- Tens of parameters are necessary (~ 40), to model the different pumps geometries in a family, resulting in a high dimensional design space.
- The parameters vary in a wide range.
- It is impossible to take a-priori into account all the geometrical constraints.

Single pump design:

$\Downarrow$

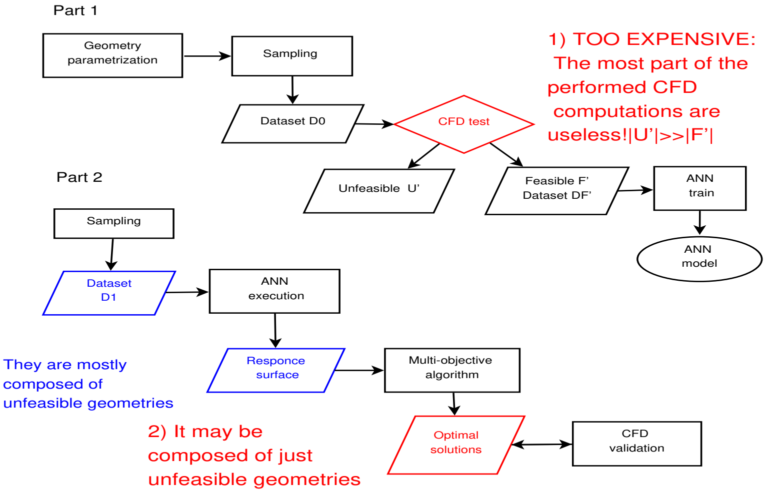
When the design space is sampled, all the geometries results to be **FEASIBLE**: they correspond to manufacturable machines and to convergent CFD computations

Family design:

$\Downarrow$

The most part (about 70%) of parameters combinations are **UNFEASIBLE**: they correspond to non manufacturable machines or to non-convergent CFD computations

Drawbacks of the application of single pump design scheme to the design of a family of pumps



Proposed approach: classification meta-model

IDEA: We need a cheap tool to identify the feasible parameters combinations.

$\Downarrow$

We propose an approach based on coupling CFD computations and the regression model with a classification meta-model.

Binary classification problem:

- class $\mathcal{F}$ of feasible geometries: manufacturable machines and convergent CFD calculations,
- class $\mathcal{U}$ of unfeasible geometries: non manufacturable machines or non-convergent CFD calculations.

Proposed approach: classification meta-model

IDEA: We need a cheap tool to identify the feasible parameters combinations.

$\Downarrow$

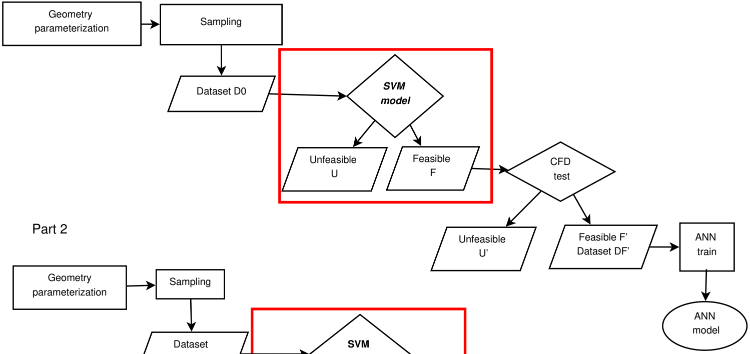
We propose an approach based on coupling CFD computations and the regression model with a classification meta-model.

Binary classification problem:

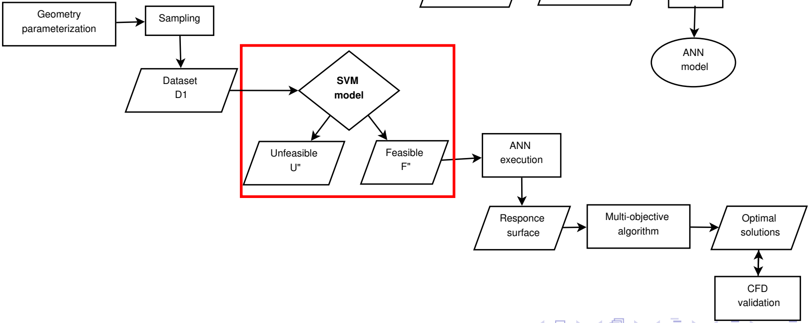
- class $\mathcal{F}$ of feasible geometries: manufacturable machines and convergent CFD calculations,
- class $\mathcal{U}$ of unfeasible geometries: non manufacturable machines or non-convergent CFD calculations.

We used Support Vector Machine as a classifier.

Part 1


Geometry parameterization
Sampling
Dataset D0
SVM model
Unfeasible U
Feasible F
CFD test
Unfeasible U'
Feasible F' Dataset DF'
ANN train
ANN model

Part 2


Geometry parameterization
Sampling
Dataset D1
SVM model
Unfeasible U"
Feasible F"
ANN execution
Responce surface
Multi-objective algorithm
Optimal solutions
CFD validation

Main feature of the classification process

We chose *Support Vector Machine* as a classifier, and we used the LIBSVM - A Library for Support Vector Machines, implemented by Chih-Chung Chang and Chih-Jen Lin.

Main feature of the classification process

We chose *Support Vector Machine* as a classifier, and we used the LIBSVM - A Library for Support Vector Machines, implemented by Chih-Chung Chang and Chih-Jen Lin.

MAIN DIFFICULTY:

SVM has to be trained on an unbalanced dataset!

$$|\mathcal{U}| >> |\mathcal{F}|$$

Unbalanced Dataset

- Two different weights are used for the positive and the negative features:

$$\min_{\omega,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i$$

$$\Downarrow$$

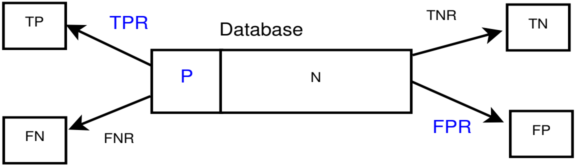
$$\min_{\omega,b} \frac{1}{2}\|w\|^2 + C_+ \sum_{x_i \in \mathcal{F}} \xi_i + C_- \sum_{x_i \in \mathcal{U}} \xi_i$$

$$\xi_i = \xi(\omega, b; x_i, y_i) = \max(1 - y_i(\omega^T \Phi(x_i) + b), 0).$$

- $C_+ > C_-$: the misinterpretation of feasible features is penalized with more severity.

Performance evaluation

- *TPR* true positive rate $TPR = \frac{TP}{TP+FN}$, fraction of positive samples correctly classified over all positive samples available in the test,
- *FPR* false positive rate $FPR = \frac{FP}{TN+FP}$, fraction of unfeasible features misinterpreted over all negative samples available in the test.



Compromise between

- finding as much feasible features as possible (TP),
- allowing in their set as few false positives as possible (FP).

Numerical Tests

- Performance depends on the free parameters C_-, C_+.
- Different databases were considered with different unbalancedness levels.
- We consider one such that:
  - $n = 44$ features,
  - $m = 80000$ samples
  - $|\mathcal{U}|/|\mathcal{F}| = 7/1$.
- We investigated the best parameter choice, fixing $C_- = 1$ and varying C_+.

Numerical Tests

- Performance depends on the free parameters C_-, C_+.
- Different databases were considered with different unbalancedness levels.
- We consider one such that:
  - $n = 44$ features,
  - $m = 80000$ samples
  - $|\mathcal{U}|/|\mathcal{F}| = 7/1$.
- We investigated the best parameter choice, fixing $C_- = 1$ and varying C_+.

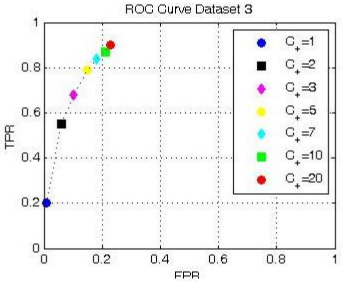
Best parameter choice

- Literature: the coefficients corresponding to feasible and unfeasible features should be inversely proportional to the ratio of the corresponding features set sizes, [Shin, Cho, 2003]:

$$\frac{C_+}{C_-} \backsimeq \frac{|\mathcal{U}|}{|\mathcal{F}|}.$$

- Best parameter choice: the ROC (Receiver Operating Characteristic) curve.

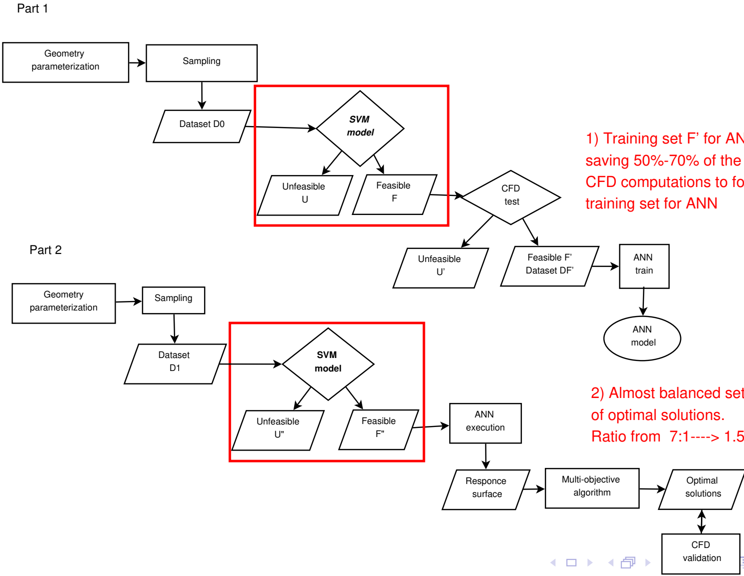
ROC curve



$$\frac{|\mathcal{U}|}{|\mathcal{F}|} = \frac{7}{1} \rightarrow \frac{C_+}{C_-} = 7$$

7 : 1	$C_+ = 1$	$C_+ = 2$	$C_+ = 3$	$C_+ = 5$	$C_+ = 7$	$C_+ = 10$	$C_+ = 20$
TPR	14.7%	53.5%	66.0%	77.5%	83.0%	86.5%	88.4%
FPR	0.7%	5.9%	9.9%	16.1%	19.8%	23.1%	25.7%

Conclusions



1) Training set F' for ANN saving 50%-70% of the CFD computations to form training set for ANN

2) Almost balanced set of optimal solutions. Ratio from 7:1----> 1.5:1.

THANK YOU FOR YOUR ATTENTION!

Kernel choice

The classification performance is influenced by the choice of the free parameters, for example the kernel choice.

- The radial basis function kernel (RBF) was chosen:

$$K(x, y) = e^{-\gamma \|x-y\|^2}.$$

- γ was set to the average squared distance among training patterns, [Nanculef R, Frandi E, Sartori C, Allende H. *A novel frank wolfe algorithm, analysis and applications to large-scale svm training*, 2014].