Support Vector Machine classification applied to the parametric design of centrifugal pumps

Elisa Riccietti
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)

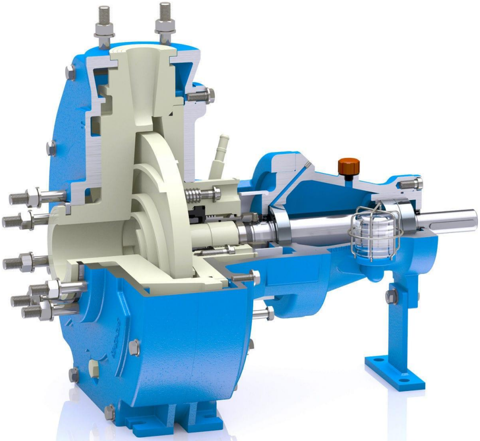
UNIVERSITÀ DEGLI STUDI FIRENZE

SIAM Congress, Politecnico di Milano, 13/09/2016-16/09/2016.

Industrial problem

Motivating application

Design and optimization of a new centrifugal pump.



Design of a centrifugal pump

- Pumps performance is measured by many different functions: efficiency, flow rate...

Design of a centrifugal pump

- Pumps performance is measured by many different functions: efficiency, flow rate...
- To optimize a new centrifugal pump it is necessary to evaluate these functions.

Design of a centrifugal pump

- Pumps performance is measured by many different functions: efficiency, flow rate...
- To optimize a new centrifugal pump it is necessary to evaluate these functions.
- Function evaluations require CFD computations (Computational Fluid Dynamics) $\rightarrow$ Expensive!

Design of a centrifugal pump

- Pumps performance is measured by many different functions: efficiency, flow rate...
- To optimize a new centrifugal pump it is necessary to evaluate these functions.
- Function evaluations require CFD computations (Computational Fluid Dynamics) $\rightarrow$ Expensive!
- The competitiveness of the business requires the design process to be as short as possible.

Design of a centrifugal pump

- Pumps performance is measured by many different functions: efficiency, flow rate...
- To optimize a new centrifugal pump it is necessary to evaluate these functions.
- Function evaluations require CFD computations (Computational Fluid Dynamics) $\rightarrow$ Expensive!
- The competitiveness of the business requires the design process to be as short as possible.
- CFD is coupled with regression models.

Design of a centrifugal pump, standard approach

1. Geometry description.
   - Choice of n independent degrees of freedom, $p_1, p_2, \ldots, p_n$.
   - Setting of the design space:

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

2. Sampling of $\mathcal{S}$.
3. CFD computations to evaluate objective functions values of some samples to form a dataset to build the regression model.
4. Building the regression model.
5. The regression model is used to predict the objective functions values of new samples.
6. Optimization algorithm: selection of an optimal solution.

Standard approach

- If a single pump is considered: the redesign starts from a baseline configuration geometrically close to the final one.
- All the tools are fine tuned for the specific application.
- All the geometrical constraints can be a priori taken into account.

Standard approach

- If a single pump is considered: the redesign starts from a baseline configuration geometrically close to the final one.
- All the tools are fine tuned for the specific application.
- All the geometrical constraints can be a priori taken into account.

$$\Downarrow$$

All the computations performed can be used to form the performance database to build the regression model.

Parametric design of a family of centrifugal pumps

- Parametric design: a family of components has to be considered.
- Tens of parameters are necessary to describe their geometry.
- The parameters vary in a wide range.
- Resulting high dimensional design space.
- It is impossible to take a priori into account all the manufacturing or geometrical constraints.

Parametric design of a family of centrifugal pumps

- Parametric design: a family of components has to be considered.
- Tens of parameters are necessary to describe their geometry.
- The parameters vary in a wide range.
- Resulting high dimensional design space.
- It is impossible to take a priori into account all the manufacturing or geometrical constraints.

$\Downarrow$

Problem: the most part (about 70%) of parameters combinations corresponds to non manufacturable machines or to non-convergent CFD computations!

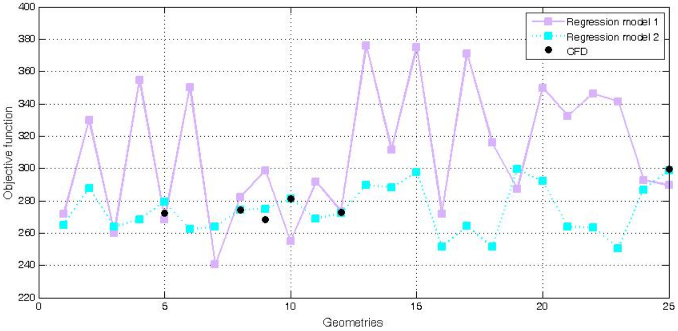
Drawbacks

- The most part of the performed CFD computations are useless, too many CFD computations are necessary to obtain enough data to build the regression model.

Drawbacks

- The most part of the performed CFD computations are useless, too many CFD computations are necessary to obtain enough data to build the regression model.
- The regression model cannot be used to predict function values of randomly chosen samples, the prediction for non good samples would yield a meaningless value, and many of them would be part of the optimal solution set.

Need for classification



Proposed approach: classification meta-model

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

Proposed approach: classification meta-model

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

The classifier is used to divide the random samples into two classes:

- 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.

Scheme of the proposed approach

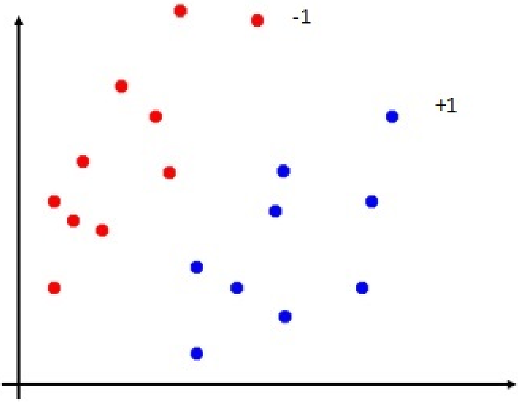
1. Geometry description.
   - Choice of n independent degrees of freedom, $p_1, p_2, \ldots, p_n$.
   - Setting of the design space

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

2. Sampling to $\mathcal{S}$.
3. Classification
4. CFD computations on the samples classified as feasible to obtain values of the objective functions.
5. Regression model is built.
6. Classification
7. Regression model is used to predict function values of new samples classified as feasible.
8. Optimization algorithm: selection of an optimal solution.

Binary classification

- Let consider a binary classification problem.
- Some samples, that are called also features are assumed to belong to two different classes, labelled as $+1$ and -1.



Support Vector Machine

- We used Support Vector Machine as a classifier.
- Machine learning method: the meta-model is trained to do a specific job, in this case it is trained to classify new samples.

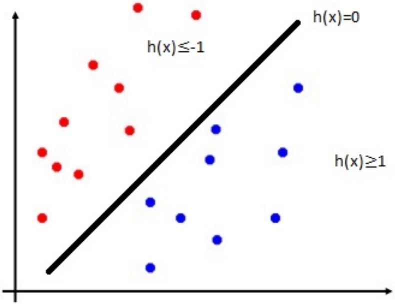
Support Vector Machine

- We used Support Vector Machine as a classifier.
- Machine learning method: the meta-model is trained to do a specific job, in this case it is trained to classify new samples.
- It is assigned a training set $\mathcal{T}$, a set of couples given by a sample $x_i \in \mathbb{R}^n$ and the label of the class it belongs to $y_i \in \{+1, -1\}$, $i = 1, \dots, m_{train}$ from which the machines takes the necessary information to perform the classification process:

$$\mathcal{T} = \{(x_1, y_1), \dots, (x_{m_{train}}, y_{m_{train}})\}.$$

Separating hyperplane

- During SVM training phase a hyperplane that separates samples in the training set belonging to different classes is searched.
- Hyperplane $H = \{x \mid h(x) = w^T x + b, \quad w \in \mathbb{R}^n, b \in \mathbb{R}\}$.



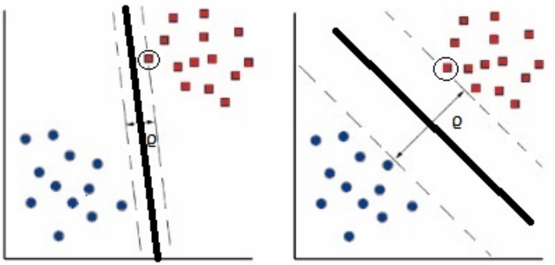
- New samples are assigned to a class according to the sign of function h.

- The separating hyperplane is not unique: for each the margin $\rho(w, b)$ is defined as:

$$\rho(w, b) = \min \frac{|w^T x + b|}{\|w\|}$$

- The optimal hyperplane is the one that maximizes the margin:

$$\max_{w \in \mathbb{R}^n, b \in \mathbb{R}} \rho(w, b)$$



Optimal hyperplane

- If features are linearly separable the optimal hyperplane exists and is unique, it can be found solving

$$\min_{w\in\mathbb{R}^n, b\in\mathbb{R}} \frac{1}{2}\|w\|^2$$
$$\text{s.t. } w^T x_i + b \geq 1, \text{ for all } x_i \in \mathcal{F},$$
$$w^T x_i + b \leq -1, \text{ for all } x_i \in \mathcal{U}.$$

Optimal hyperplane

- If features are linearly separable the optimal hyperplane exists and is unique, it can be found solving

$$\min_{w\in\mathbb{R}^n, b\in\mathbb{R}} \frac{1}{2}\|w\|^2$$

$$\text{s.t. } w^T x_i + b \geq 1, \text{ for all } x_i \in \mathcal{F},$$

$$w^T x_i + b \leq -1, \text{ for all } x_i \in \mathcal{U}.$$

- If the features are not linearly separable it is necessary to allow the presence of some outliers inserting some slack variables ζ_i $i = 1, \dots, m_{train}$ in the model:

$$w^T x_i + b \geq 1 - \zeta_i \text{ for all } x_i \in \mathcal{F},$$

$$w^T x_i + b \leq -1 + \zeta_i \text{ for all } x_i \in \mathcal{U},$$

$$\zeta_i \geq 0, i = 1, \dots, m_{train}.$$

Optimal hyperplane

- If x_i is incorrectly classified $\zeta_i > 1$, so $\sum_{i=1}^{m_{train}} \zeta_i$ is an upper bound of the number of training features misinterpreted:

$$\min_{\omega,b,\zeta} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m_{train}} \zeta_i$$

$$\text{s.t. } y_i(w^T x_i + b) \leq 1 - \zeta_i,$$

$$\zeta_i \geq 0, \; i = 1, \ldots, m_{train}.$$

Unbalanced Dataset

- Parametric design: a training set has to be formed to train SVM meta-model sampling randomly the design space.
- Problem: the unfeasible samples are many more than the feasible ones: SVM has to be trained on a strongly unbalanced training set

Unbalanced Dataset

- Parametric design: a training set has to be formed to train SVM meta-model sampling randomly the design space.
- Problem: the unfeasible samples are many more than the feasible ones: SVM has to be trained on a strongly unbalanced training set

$$\downarrow$$

  - SVM has few information about the minority class to make an accurate prediction
  - It is easy to have many feasible features misclassified.

Unbalanced Dataset

- We are interested in detecting feasible samples: it is necessary to force the classifier to take features belonging to the different classes into different consideration.

Unbalanced Dataset

- We are interested in detecting feasible samples: it is necessary to force the classifier to take features belonging to the different classes into different consideration.
- Two different weights are used for the positive and the negative features:

$$\min_{\omega,b,\zeta} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m_{train}} \zeta_i \rightarrow \min_{\omega,b,\zeta} \frac{1}{2}\|w\|^2 + C_+ \sum_{x_i\in\mathcal{F}} \zeta_i + C_- \sum_{x_i\in\mathcal{U}} \zeta_i .$$

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

Numerical Tests

- We used the SVM implemented by Chih-Chung Chang and Chih-Jen Lin in LIBSVM - A Library for Support Vector Machines.

Numerical Tests

- We used the SVM implemented by Chih-Chung Chang and Chih-Jen Lin in LIBSVM - A Library for Support Vector Machines.
- The results of the classification procedure determines the savings in terms of CFD computations.
- Compromise between
  - finding as much feasible features as possible,
  - allowing in their set as few false positives as possible.

Performance evaluation

Performance is evaluated by the confusion matrix, in which are reported:

- *TPR* true positive rate $TPR = \frac{TP}{TP+FN}$,
- *FPR* false positive rate $FPR = \frac{FP}{TN+FP}$,
- *TNR* true negative rate $TNR = \frac{TN}{TN+FP}$,
- *FNR* false negative rate $FNR = \frac{FN}{TP+FN}$.

Numerical Tests

- The choice of the free parameters deeply affects the classifier performance.
- We investigated the best parameter choice, fixing $C_- = 1$ and varying C_+.

Numerical Tests

- The choice of the free parameters deeply affects the classifier performance.
- We investigated the best parameter choice, fixing $C_{-} = 1$ and varying C_{+}.
- Three different databases are considered, with n=40, 44, 42 degrees of freedom, and ratio between unfeasible and feasible features 3:1, 7:1, 6:1.
- SVM was trained over a set of $m_{train} = 30000$ geometries.

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]:

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

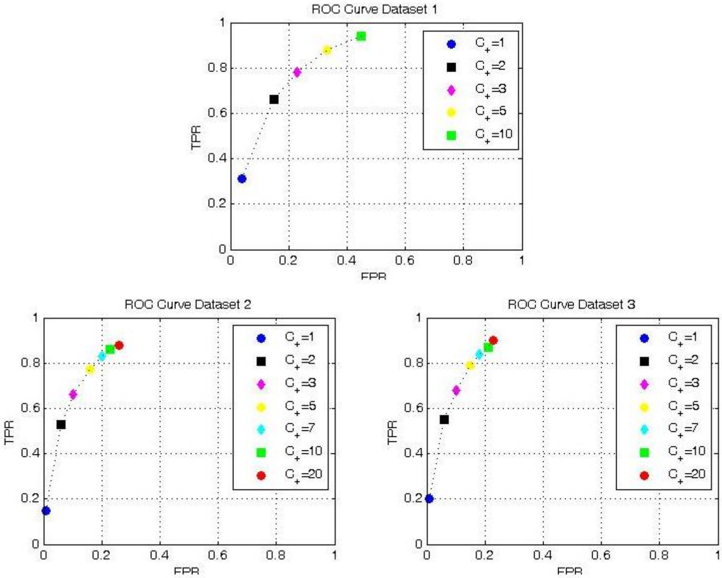
Best parameter choice

- 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 curve.

ROC curve


ROC Curve Dataset 1
TPR
FPR
0
0.2
0.4
0.6
0.8
1
C+=1
C+=2
C+=3
C+=5
C+=10
ROC Curve Dataset 2
TPR
FPR
C+=1
C+=2
C+=3
C+=5
C+=7
C+=10
C+=20
ROC Curve Dataset 3
TPR
FPR
C+=1
C+=2
C+=3
C+=5
C+=7
C+=10
C+=20

Confusion matrix

3 : 1	$C_+ = 1$	$C_+ = 2$	$C_+ = 3$	$C_+ = 5$	$C_+ = 10$
TPR	31%	66%	78%	88%	94%
FPR	4%	15%	23%	33%	45%
TNR	96%	85%	77%	67%	55%
FNR	68%	34%	22%	11%	5%

7 : 1	$C_+ = 1$	$C_+ = 2$	$C_+ = 3$	$C_+ = 5$	$C_+ = 7$	$C_+ = 10$	$C_+ = 20$
TPR	15%	53%	66%	77%	83%	86%	88%
FPR	0.7%	6%	10%	16%	20%	23%	26%
TNR	99%	94%	90%	84%	80%	77%	74%
FNR	85%	46%	34%	22%	17%	13%	11%

6 : 1	$C_+ = 1$	$C_+ = 2$	$C_+ = 3$	$C_+ = 5$	$C_+ = 7$	$C_+ = 10$	$C_+ = 20$
TPR	20%	55%	68%	79%	84%	87%	90%
FPR	1%	6%	10%	15%	18%	21%	23%
TNR	99%	94%	90%	85%	82%	79%	77%
FNR	79%	44%	32%	21%	16%	13%	10%

Conclusions

Benefits of the proposed approach:

- Strategy to handle the unbalancedness of the dataset: good classification results on datasets with different number of degrees of freedom and different ratio between feasible and unfeasible features.

Conclusions

Benefits of the proposed approach:

- Strategy to handle the unbalancedness of the dataset: good classification results on datasets with different number of degrees of freedom and different ratio between feasible and unfeasible features.
- First classification process: form the same dataset to build the regression model as in the standard approach but saving about 40% of the CFD computations.

Conclusions

Benefits of the proposed approach:

- Strategy to handle the unbalancedness of the dataset: good classification results on datasets with different number of degrees of freedom and different ratio between feasible and unfeasible features.
- First classification process: form the same dataset to build the regression model as in the standard approach but saving about 40% of the CFD computations.
- Second classification process: restrict the regression phase to a set of samples that is far more balanced than the starting dataset (from $7:1 \rightarrow 2:1$), reducing the undesirable presence of unfeasible samples in the optimal solution set.

Conclusions

Benefits of the proposed approach:

- Strategy to handle the unbalancedness of the dataset: good classification results on datasets with different number of degrees of freedom and different ratio between feasible and unfeasible features.
- First classification process: form the same dataset to build the regression model as in the standard approach but saving about 40% of the CFD computations.
- Second classification process: restrict the regression phase to a set of samples that is far more balanced than the starting dataset (from $7:1 \rightarrow 2:1$), reducing the undesirable presence of unfeasible samples in the optimal solution set.

THANK YOU FOR YOUR ATTENTION!

Performance evaluation

Performance is evaluated by the confusion matrix, in which are reported:

- TPR true positive rate or sensitivity or recall $TPR = \frac{TP}{TP+FN}$, fraction of positive samples correctly classified over all positive samples available during the test,
- FNR false negative rate $FNR = \frac{FN}{TP+FN}$ fraction of feasible features misinterpreted over all positive samples available during the test,
- TNR true negative rate or specificity $TNR = \frac{TN}{TN+FP}$, fraction of negative samples correctly classified over all negative samples available during the test,
- FPR false positive rate $FPR = \frac{FP}{TN+FP}$, fraction of unfeasible features misinterpreted over all negative samples available during the test.

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. Information Sciences 2014]