

The modeling technology **xtractis**® automatically extracts fuzzy rules to solve complex problems, such as predicting consumer preferences from sensory evaluation and instrumental measurements. The learning task involved may be very complex since the reference database can have a very high dimensionality. In order to save the model intelligibility and keep the possibility of inverting the model (i.e. finding exact optimal input values satisfying a request made on the outputs), we impose to use only original variables, and not combined ones like with PCA. By reducing the problem dimensionality, we increase the model robustness, make it easier to interpret, and reduce its generation time. Although we have already implemented an efficient variables selection technique, this study aims to develop and compare a new one based on Support Vector Machines (SVM).



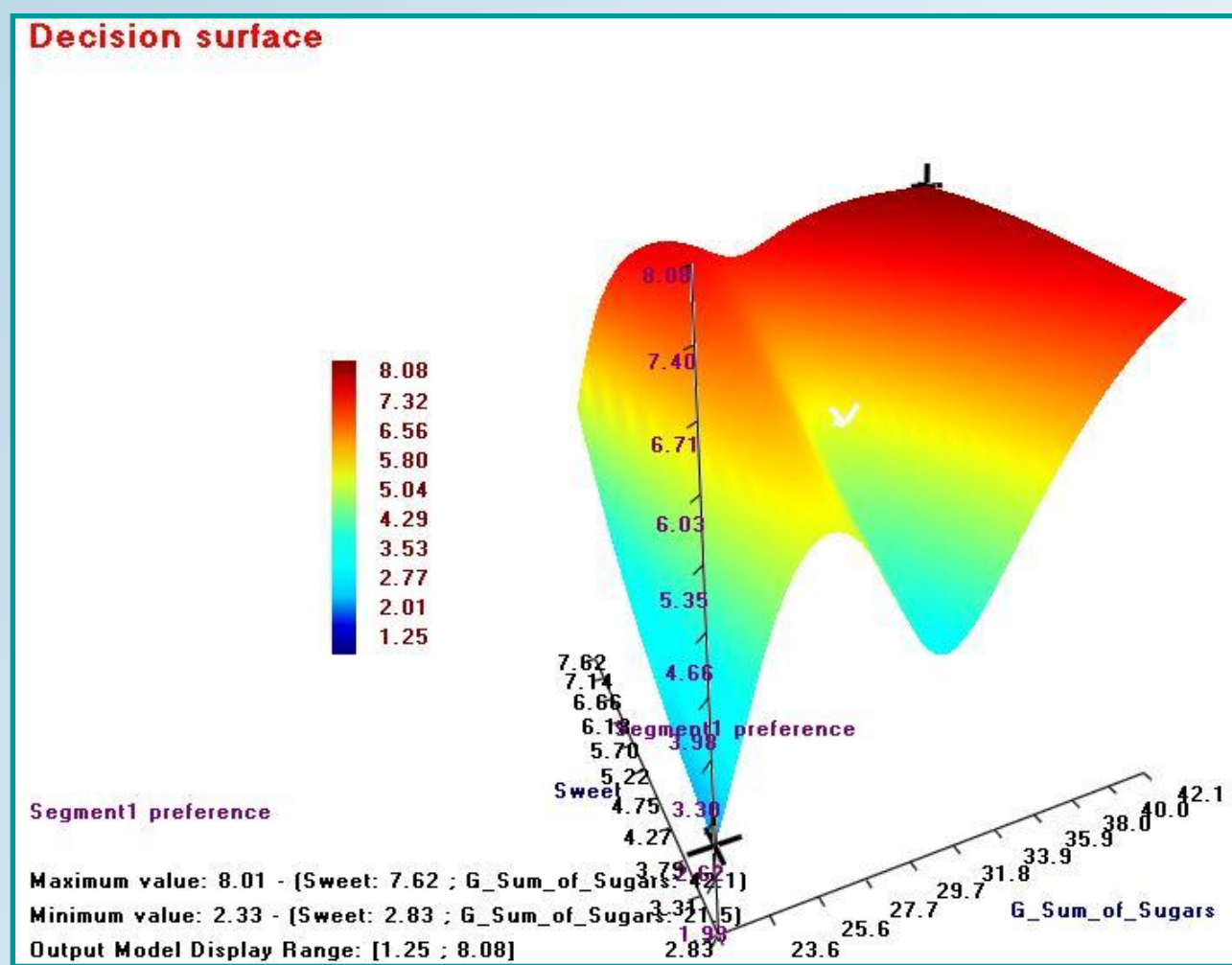
Variables selection by Support Vector Machines for fuzzy non linear modeling



Z. Zalila^{1,2}, B. Lorentz^{1,2}, J. Cuquemelle^{1,2}

xtractis®

- Fuzzy modeling technology developed by **intellitech**
- Regression problems solver: from a finite quantity of data (reference database), **xtractis**® generates a fuzzy model with high generalization ability



Complex problems

- High dimensionality (>350 variables)
- High interactions between variables
- Strong non linearity
- Sparse, incomplete or fuzzy reference data

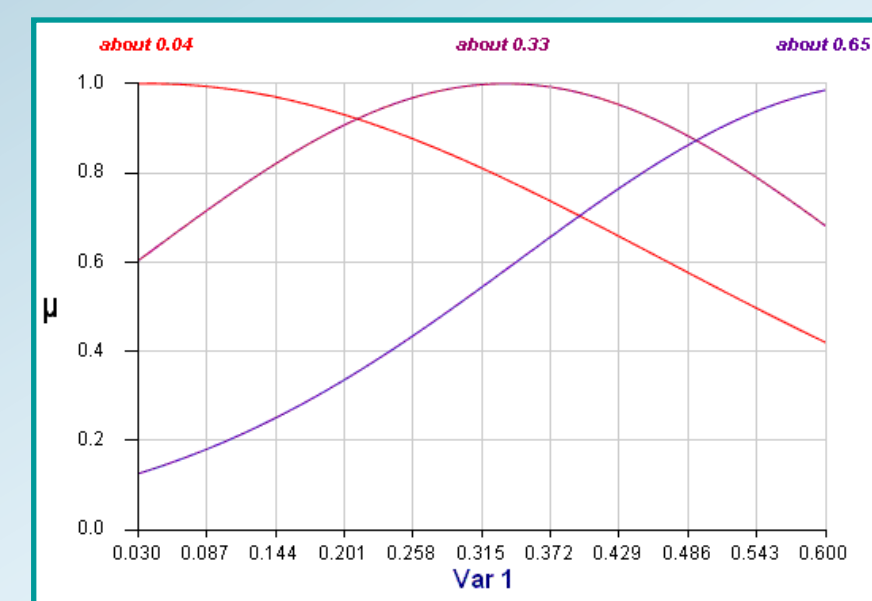
Data type

- Consumer preferences (Subjective data)
- Sensory evaluation (Subjective Objectivized data)
- Instrumental measurements (Objective data)

Fuzzy rules extraction

if (U_1 is A_{1j}) and ... and (U_n is A_{nj}) then (V is B_j)

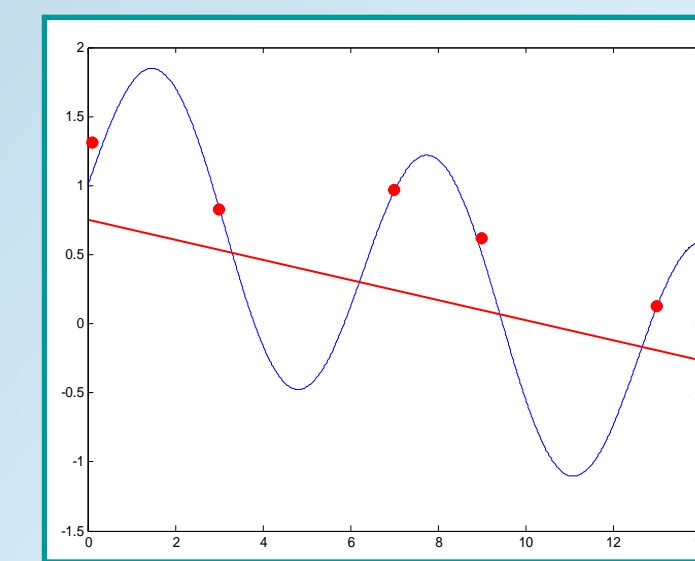
- Intelligibility
- Accuracy
- Robustness



Dimensionality reduction

Objectives

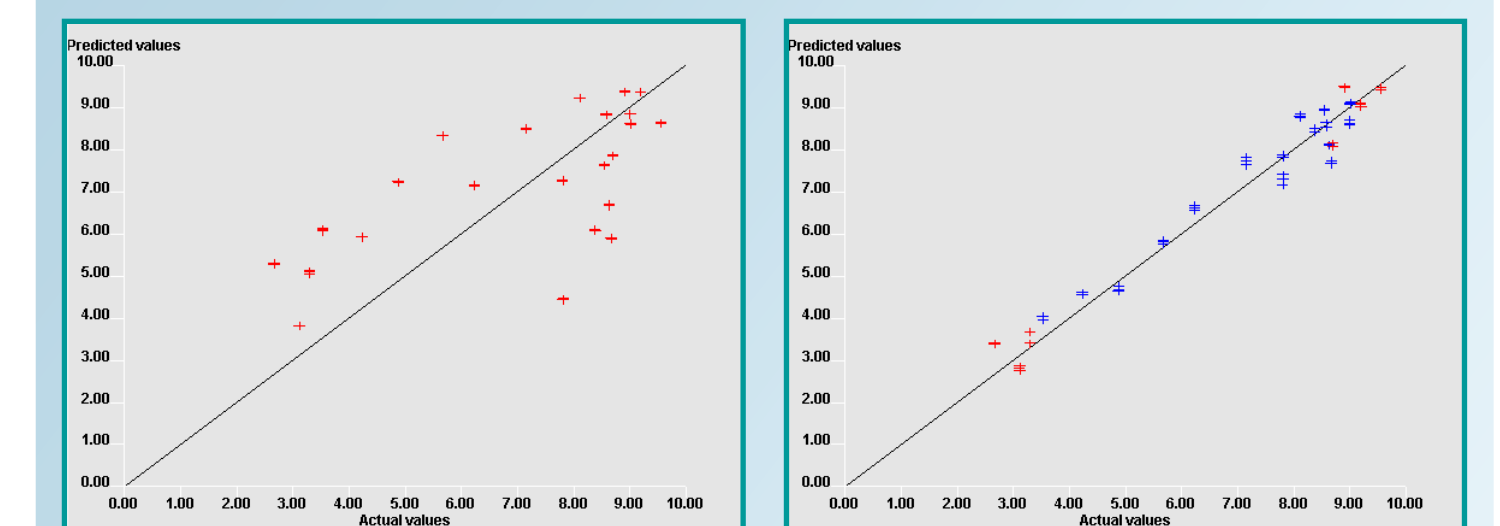
- ↘ risk of overfitting + ↗ model robustness:
- ↗ number of variables
 - ↗ accuracy
 - ↘ generalization ability
 - ↘ interpretability



- Keep the possibility of inverting the model (optimal inputs) and save its intelligibility: only original variables (no PCA)
- Reduce model generation time

Issues

- Find a set of variables that ↗ accuracy and robustness of the fuzzy model



3 x Leave One Out with 111 variables (left) vs 6 variables (right)

- Detect redundant information or irrelevant variables, and only keep the necessary ones
- Search through all the possible variables subsets: ~~exhaustive search~~

xtractis® : dimensionality reduction

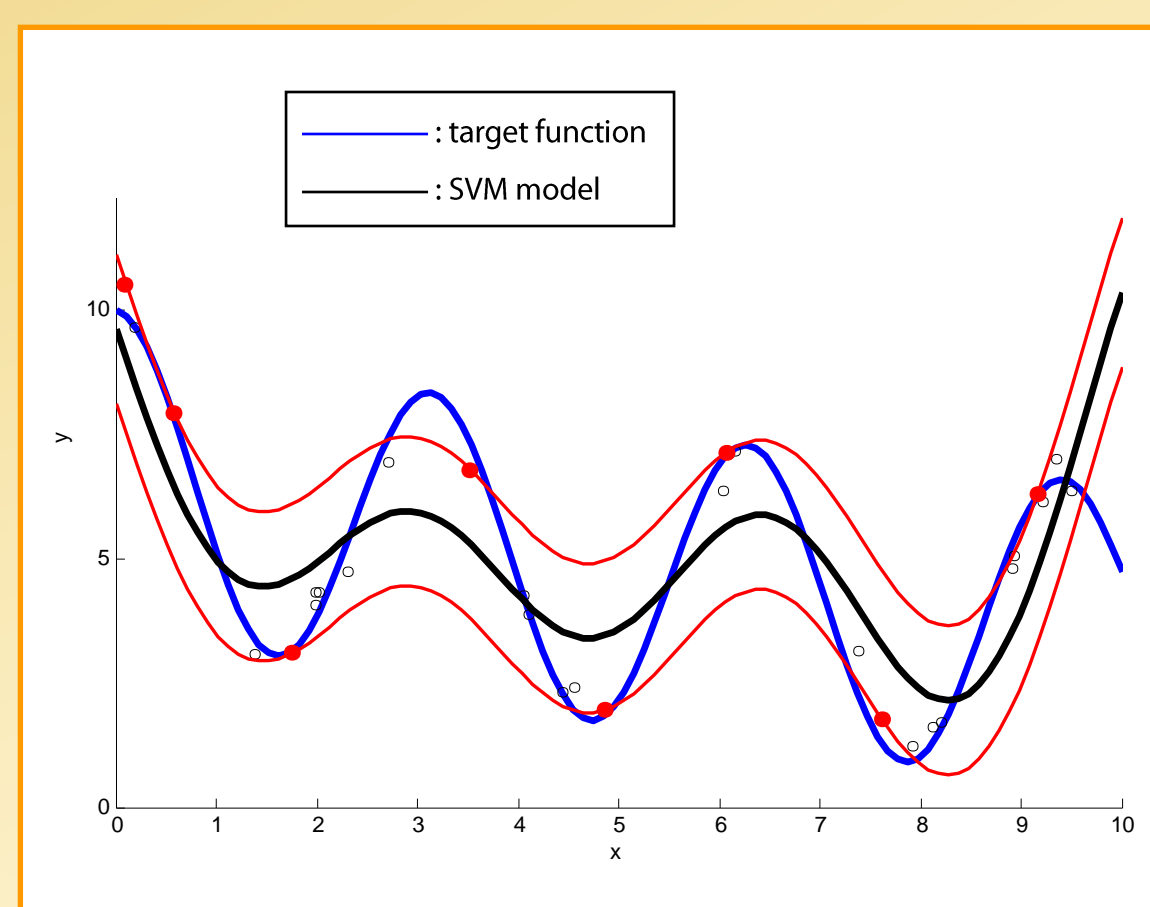
- Efficient, directly based on fuzzy models performance
- Lots of models generations: slow variables selection

SVM: Support Vector Machine

- Fast solver for non linear regression problems [Vapnik, 1995]
- Accurate models
- Control of the model complexity
- Not interpretable models:

$$f(x) = \sum_{i=1}^m (\beta_i K(x, x_i)) + \omega_0$$

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right)$$



SVM: dimensionality reduction

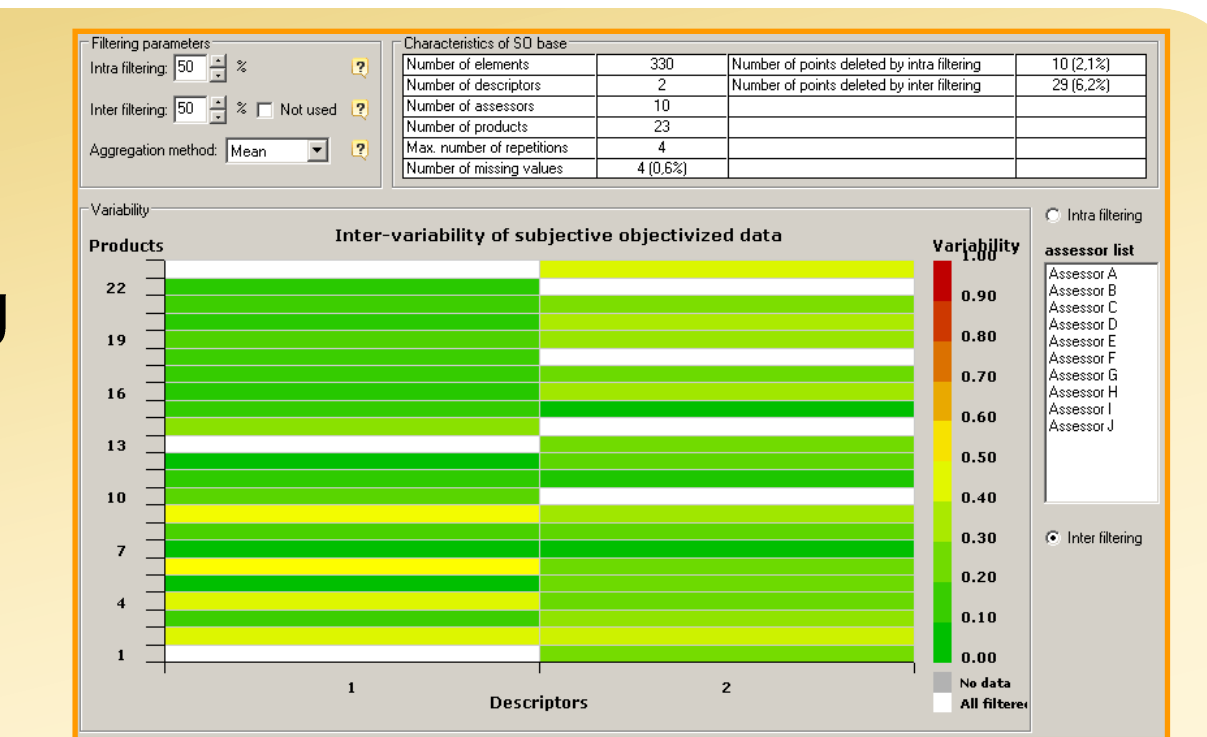
- Use of SVM models instead of fuzzy models
- Most irrelevant variable = variable with the less influence on SVM model (brings little information)
- Descending search: starting from the whole initial set, most irrelevant variable removed at each step
- Estimation of the influence of a variable on the model: ↗ accuracy + ↘ complexity

xtractis® approach

Learning base

- 111 input variables } more variables than learning points: risk of overfitting
- 23 learning points

Hybrid approach



SVM dimensionality reduction tool

Preselects 12 variables **in less than 1 minute**: var 1, 5, 15, 20, 24, 64, 75, 77, 82, 94, 96, 104

xtractis® dimensionality reduction tool

- Selects the best subset among the initial 111 variables set: var 1, 5, 23, 24, 75, 82
- Search time: 4.75 min**
- Selects the best subset among the 12 preselected variables set: var 1, 24, 75, 82, 104
- Search time: 0.25 min**

Generation of top-models

- | | |
|--|--|
| <ul style="list-style-type: none"> Structure: 6 variables, 4 fuzzy rules Accuracy: 0.994 Robustness (Monte Carlo cross validation): <ul style="list-style-type: none"> 0.949 (MC 200 x 10%) 0.924 (MC 200 x 20%) 0.898 (MC 200 x 30%) | <ul style="list-style-type: none"> Structure: 5 variables, 5 fuzzy rules Accuracy: 0.999 Robustness (Monte Carlo cross validation): <ul style="list-style-type: none"> 0.951 (MC 200 x 10%) 0.940 (MC 200 x 20%) 0.928 (MC 200 x 30%) |
|--|--|

New dimensionality reduction / Models extraction with others learning parameters

Results

- Fuzzy models generated using the SVM preselection have good performance (accuracy and robustness)
- Among the initial 111 variables set, **xtractis**® selects 5 of the 12 variables considered as the most important by the SVM tool
- For huge dimensionality (>100 variables), possibility to preselect the most relevant variables for a fuzzy model with a fast modeling technology (SVM)
- Transformation of 111 variables complex regression problem into a 12 variables problem easier to solve
- Time saved can be used for selecting the best **xtractis**® learning parameters and the best fuzzy models structures **more top models generated**