1st Research Contact Day
Computational Intelligence & Learning
Doctoral School

gembloux
faculté universitaire
des sciences agronomiques

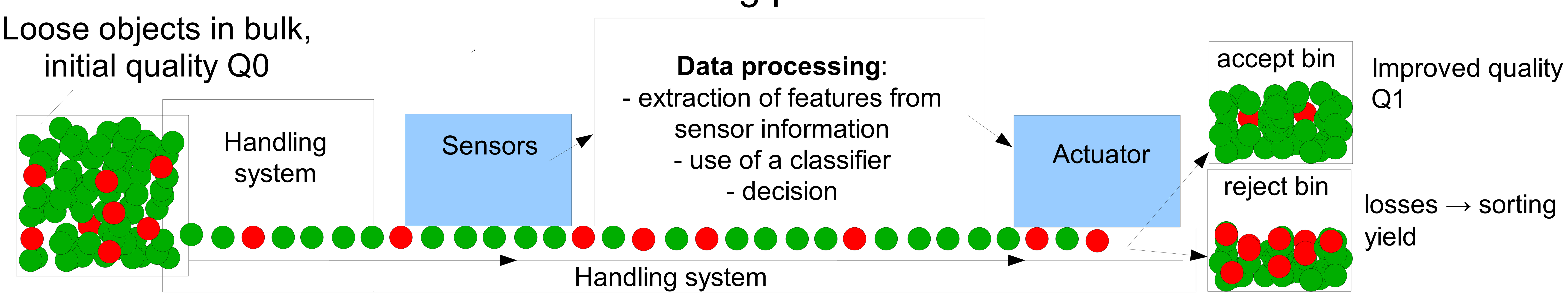
Quality improvement of loose objects using a sorting optimization curve

D. Ooms[1], R. Palm[2], M.-F. Destain[1]
Gembloux Agricultural University, Mechanics and Construction Dpt. (1),
Applied Statistics, Computer Science and Mathematics Dpt. (2)
2 Passage des Déportés, B - 5030 Gembloux, Belgium

Introduction

A sorting optimization curve (SOC) is introduced for the on-line sorting of loose objects. The curve is a keypoint of a global machine learning strategy which simply consists in taking the best advantage of data provided by previous sorting operations and training, with the final choice still dependant on a human operator. This operator then receives complete and adequate information for choosing parameters to be used into the sorting process, by the way of a SOC curve. The introduction of quality and yield for the assessment of the classifier performance is well suited for this kind of product. Quality requirement can be easily taken into account to satisfy to quality norms. The yield factor balances the quality factor as both cannot be maximized simultaneously.

Sorting process

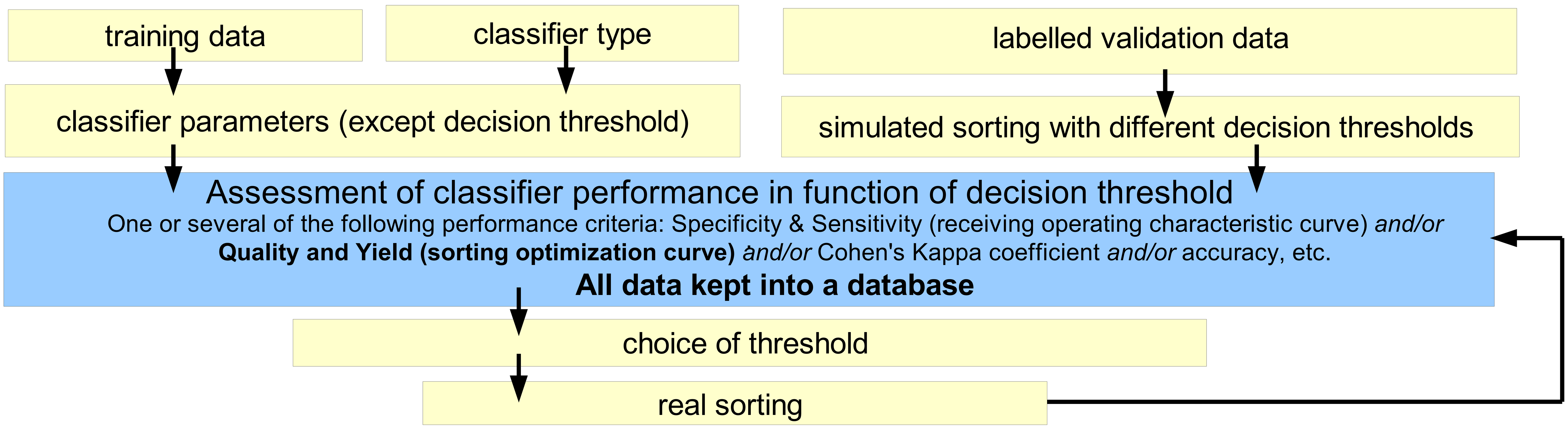


Building and improvement of a sorting database

The data gathered during the sorting process set-up and use can be recorded and exploited when a new situation arises. By searching similar past operations within a searchable database, the user can choose the best set classifier + parameters with a minimum of training data, or none at all. It is thus important to keep a record of data under a simple and explicit format, associated with the prediciton of output in terms of quality and quantity. The sorting optimization curves (SOC) are an important element of such a database. This is most relevant with agricultural products such as grain and seeds which have a great variability due to growing, harvest and stockage conditions. SOC curves may replace or may be recorded together with receiver operating characteristic curves (ROC).

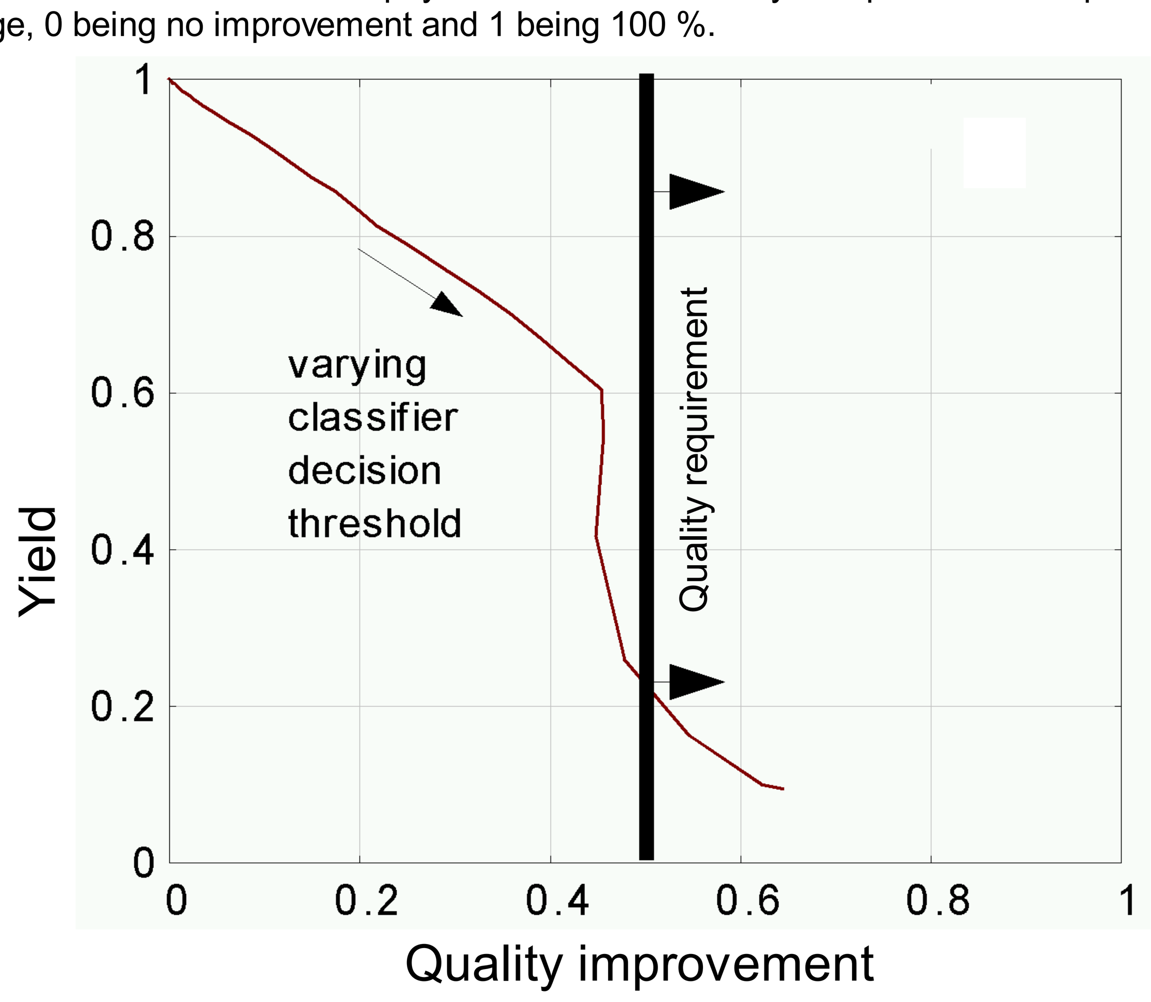
The **ROC curve** is widely used for binary classification. A typical example is diagnosis from medical imaging. True positive rate is plotted against false positive rate into a 2D graph in function of the decision threshold. The optimal threshold may be deduced from ROC curves when misclassification can be predicted. However, individual misclassification costs are not easily predictable with loose objects in bulk.

The **decision threshold** is the value used to determine if each object is intended for the accept or reject bin. The binary classifier issues an output value and the threshold is used for making a decision between the two classes. A high threshold value would reduce the amount of objects into the accept bin, but gives a purer product. It would give a high quality with a poor yield. A low threshold value would give a higher amount of objects into the accept bin, with more unwanted objects in it. It would give a higher yield with a poorer quality.



The Sorting Optimization Curve (SOC)

The sorting optimization curve is plotted by predicting output yield against quality in function of the decision threshold. The sorting can be simulated using labelled data or made effectively. The following curve was built from a dataset of labelled samples of chicory seeds. The sorting is intended to remove immature seeds on basis of chlorophyll fluorescence. Quality is expressed as improvement of germination percentage, 0 being no improvement and 1 being 100 %.



Requirement about quality and/or yield can be easily represented in SOC curves using straight lines: horizontal for yield requirement and vertical for quality requirement. In the above example, the quality requirement can only be met at the cost of heavy material losses.

Different expressions for yield and quality

The producer must quantify **quality** as a function of the proportion of unwanted objects in order to make SOC curves useful. In such curves, quality may be expressed as its simplest expression Q1, as the quality improvement Q1-Q0 or as the relative quality improvement (Q1-Q0)/(1-Q0). The latter has the advantage of uniform scaling.

Yield can refer to the total amount of output product into the accept bin, or to the total amount of wanted objects (green in first scheme above). Some caution must be observed to maintain coherent units into the database

Perspective: using the database to set up the parameters and decision threshold for new objects

The sorting optimization curve is dependant on the following:
- the objects: their origin, nature and initial quality, expressed as various factors such as variety (for grain and seeds: harvest time, growing temperature, germination percentage, etc.);
- the kind of classifier;
- the classifier internal parameters (examples: number of nodes and training pattern of a neural network, smoothing factor of a kernel).

Therefore, it is necessary to provide a search engine for retrieving the SOC curve, within the database, which corresponds best to the present objects and classifier. It is a classification task in itself where features are the above listed topics: objects, classifier and classifier parameters. The search engine must provide a confidence level on the SOC matching the actual sorting conditions and material. Interpolation methods could be used to provide an adapted SOC curve without the need for re-testing the classifier(s).

Conclusion

The sorting optimization curve has a great potential for sorting applications of bulk materials where output quality is a key factor. Variability of material and sorting methods can be incorporated into a machine learning process in which past experiences are continuously recorded and used for predicting and controlling the future behavior of a sorting machine