Paradigm for selecting the optimum classifier in semiconductor automatic defect classification applications

Martin A. Hunt, James S. Goddard, James A. Mullens, Regina K. Ferrell, Bobby R. Whitus, Ariel Ben-Porath (Oak Ridge National Laboratory, Oak Ridge, TN, 37831) (ABP, Applied Materials, Réhovot, Israel)

The automatic classification of defects found on semiconductor wafers using a scanning electron microscope (SEM) is a complex task that involves many steps. The process includes detecting the defect, measuring attributes of the defect, and automatically assigning a classification. In many cases, especially during product ramp-up, there are few training examples for an automatic defect classification (ADC) system. This condition presents a problem for traditional supervised parametric and nonparametric learning techniques.

In this paper we investigate the attributes of several approaches to ADC and compare their performance under a variety of available training data scenarios. Supervised parametric classifiers such as Bayes or Maximum likelihood classifiers rely on assumptions of the form of the underlying probability distributions of the data that often are invalid. A typical parameterization approach is to assume the class-conditional probabilities are normal and unimodal. Supervised nonparametric techniques remove the probability distribution assumptions and converge to the optimum Bayes classifier when a large number training of examples are available. In general the supervised nonparametric techniques still need a moderate amount of training data if they are to be considered applicable in a broad range of input conditions (not just accurate for the given training set), but they can function adequately with a limited training set. A final classification option available is a rule-based approach that is either supervised or unsupervised. A rule-based approach can function with no training examples (unsupervised) or with training examples (supervised). In the unsupervised case expert knowledge of the process and historical information from similar processes can be formulated into decision rules. In the supervised case training data can be used to automatically generate a decision tree that optimizes the hard decision boundaries to achieve the best performance on the given, labeled training data.

We have selected to characterize the attributes and performance of a traditional K-nearest neighbor classifier, probabilistic neural network (PNN), and rule-based classifier in the context of SEM ADC. The PNN classifier is a nonparametric supervised classifier that is built around a radial basis function (RBF) neural network architecture. A basic summary of the PNN will be presented along with the generic strengths and weakness described in the literature and observed with actual semiconductor defect data. The PNN classifier is able to manage conditions such as non-convex class distributions and single class multiple clusters in feature space. A rule-based classifier producing built-in core classes provided by the Applied Materials SEMVision tool will be characterized in the context of both few examples and no examples.

An extensive set of fab generated data is used to characterize the performance of these ADC approaches. Typical data sets contain from 30 to greater than 200 defects and have many feature descriptors for each defect. The number of classes in the data set range from 4 to more than 12. The conclusions reached from this analysis indicate that the strengths of each method are evident under specific conditions that are related to different stages within the VLSI yield curve.