Part G

Evaluation and Benchmarking

The final part of this handbook deals with the problem of benchmarking. Benchmarking provides a fundamental way of evaluating various proposed solutions against the state of the art. This is typically done for one of three reasons. First to see if a given system or process meets specific performance criteria in terms of accuracy, speed, or, more generally, costs. Performance criteria are set up and used to check whether the systems pass the performance test. Second, systems may be evaluated to determine the best at a particular task... to evaluate the state of the art. In this case, datasets and metrics are set up and solutions compete against each other to determine a winner. In the academic community, competitions such as these are often set up and are very successful at driving research, and motivating the development of new algorithms. Finally benchmarking can be used to simply evaluate current approaches against the challenges of the problem itself to know where effort should be put to improve performance. Such benchmarks are common internally during the research process where development datasets provide both regression data to make sure systems are still performing adequately and challenge data, representing yet unsolved problems.

In ▶Chap. 29 (Datasets and Annotations for Document Analysis and Recognition), Ernest Valveny focuses on data set problems, describing various evaluations and datasets that are common in the community, what the needs are for new datasets and the challenges of building them. In ▶Chap. 30 (Tools and Metrics for Document Analysis Systems Evaluation), Volker Märgner and Haikal El Abed wrap up the handbook by providing a survey to tools and performance metrics used in the field. Over the past decade we have seen a consolidation of such tools and definite sign that the community is maturing.

There is also a lot to be learned in this area, especially from related communities. While data is not cheap to produce, providing quality datasets for researchers to work on has proven time and again to be one of the most valuable resources to motivate the community to address and solve its most challenging problems.