Test case development for big data solution evaluation

Maximilian Zocholl[1], Elena Camossi[1] and Anne-Laure Jousselme[1]

[1] NATO STO CMRE, La Spezia 19126, Italy

Exploiting big data for detection tasks requires systematic testing of the big data solution as they are supposed to cope reliably with high dimensional data with large variations. For big data solutions which ingest Automatic Information System (AIS) data [1] typical big data variations are observable in volume, velocity, variety and veracity of the data. The design space, out of which test cases can be selected, is defined by the input factors of the big data solution and the respective big data variations which translate into the dimensions of the design space and the values of the big data variations which are represented the domain of each factor. When combining the typically large number of dimensions with multiple possible values or even continuous value domains for each factor, the number of possible experiments translates seamlessly into combinatorial mayhem and the non-applicability of classical statistical assumptions [2] As the performance of all possible experiments is infeasible, the systematic selection of test cases is crucial. Even though big data dimensions are generally interdependent [3], the reduction of complexity is only feasible by assuming independence between different big data dimensions. To frame the sequential selection process, the described approach serves at narrowing down and defining the design space.

Taking into account domain specific constraints, different big data variations are supposed to be stronger coupled whilst others are only loosely interdependent. In AIS related big data solutions, volume and velocity are functionally stronger interdependent, while veracity and variety are semantically stronger interdependent. For the following, variety and veracity variations are discussed more detailed. While further narrowing down the design space two main challenges arise. Internal and external validity of the test cases need to be guaranteed [4]. Since external and internal validity are necessary conditions for the validity of test cases, they can be used to deduce guidelines.

Firstly, the internal validity of the data needs to be guaranteed. This requires that the difference in the expected results between true and false events is significant. For this purpose, the number of experiments is adapted to the accuracy of the method to be evaluated in the course of the evaluation.

More detailed, measured distances between the information sets that include positive and negative events are reduced stepwise during consecutive tests of a test series. By starting from information sets with a large distance, the initial number of experiments is reduced. After failing one test, there are different ways to proceed.

If the development process was completed before the evaluation started, the test series can either be stopped or the pace of the distance measure can be reduced to refine the evaluation result.

If the development process is not completed, yet, an alternation between evaluation and development phases can lead to an improvement of the evaluation result. For disjoining evaluation and development, a new test has to be created replacing the failed test. This procedure reduces the risk of a development which follows the evaluation criteria but not the concept to be detected. If a purely test-driven development is chosen, the same tests are typically reevaluated.

Secondly, the external validity needs to be guaranteed. This allows for the deduction of two procedures, taking into account dependencies between different big data variations and using different approaches for the test case development.

Firstly, interdependence between different big data variations needs to be taken into account stepwise. For this purpose, the number of test cases evaluating the variations on one big data dimension is adapted to the evaluation result of the test cases evaluating variations of a defining big data dimension. As an example, the variation of the big data veracity requires the existence of a variety of big data to degrade. While unidirectional dependence allows for the complete testing of the independent big data dimension before testing the dependent big data dimension, mutually dependent big data variations require a coordinated extension of the test cases.

For this, big data solution specific constraints on variety and veracity are used successively.
The constraints on the variety are defined by the mapping of the data fields or factors of the big data sources to the input variables of the components of the big data solution. If a data field is not processed, it can be excluded from the following evaluation step.

The constraints on the veracity are defined in dependence to the available big data variety. Starting from a given variety, e.g. variety 1 in Figure 1, the veracity is degraded iteratively until the required performance is testified. If the test is failed, as shown in Figure 1 a), the test is stopped for both veracity and variety variations or continued with similar test cases after the improvement of the big data solution until the required performance is attained, as shown in Figure 1 b). The successful accomplishment of one extent of variation of variety enables the extension of the design space to the next larger test case, Figure 1 c).

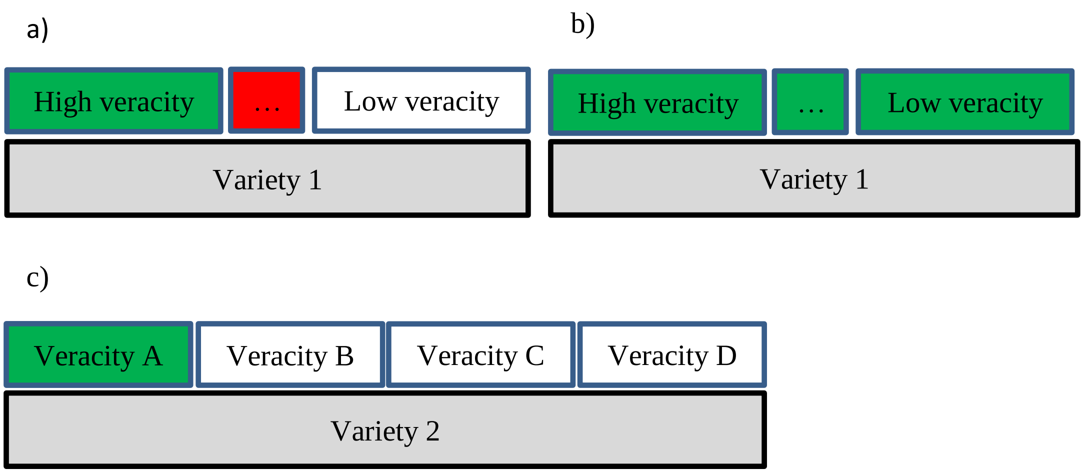

Figure 1: Testing of veracity variation depending on the progress in testing variety variation

For each class of detection task, e.g. tasks relating to “area”, “speed”, “course” and “heading”, “connectivity” and maritime status and each detection task, e.g. “null speed” or “mismatch speed vessel type” can be defined separately by the combination of input data value levels of the data fields, identified to contribute to the detection result. E.g. if a component takes into account longitude and latitude for detecting a speed event, only the degradation of those input variables can be tested.
In case the variables being taken into account for a detection task are unknown, the impact of a veracity variation yields information on it. E.g. varying the “gps_speed” signal and observing the detection of “null speed” events unveils whether the respective AIS field is used for the detection task “null speed”.

With an increasing variety of the information taken into account, detections and testing of these detections becomes feasible for data with very low veracity but only marginal differences to data with high veracity, e.g. as current eddies can be estimated from AIS signals [5] weather and ocean conditions can yield information about the veracity of AIS signals.

Secondly, for guaranteeing the representative character of the developed test cases the addressed variations need to be set into context by the actual big data variations. By mapping the performed test cases to the design space, gaps can be detected and addressed by future test case developments. The definition and narrowing down of the design space can be supported by the combination of theoretical event descriptions and sample data.

With respect to theoretical concept descriptions, the definition of the design space is improved by two aspects. Firstly, events which are known to exist but which are not included in the sample dataset can be added artificially, e.g. by simulation. Secondly, human interpretable event definitions are specified and instantiated, e.g. collision is specified into collisions between vessels of different types and collisions between one vessel and the mainland.

With respect to the sample data, the definition of the design space benefits from three aspects. Firstly, theoretically impossible factor combinations that are observed in the sample data (e.g. “engaged in fishing” and “at anchor” while changes in position data indicates a moving vessel). Secondly, unknown variations, i.e. variations which are not described, yet, increase the array of testable variations, e.g. the fast and consecutive sending of two different AIS positions from two transceivers from bow and stern of a vessel. Thirdly, Depending on the frequency of factor combination observations, the design space can be characterized in a probabilistic way.

References

1. ITU-R: Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band, International Telecommunications Union, itu-r m.1371-4, (2010).
2. Fan, Jianqing and Han, Fang and Liu, Han: Challenges of Big Data analysis, National Science Review, vol. 1, no. 2, pp. 293-314, (2014).
3. Gandomi, Amir and Haider, Murtaza: Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, vol. 35, no. 2, pp. 137-144, Elsevier, (2015).
4. Cook, Thomas D and Campbell, Donald Thomas and Shadish, William: Experimental and quasi-experimental designs for generalized causal inference, Houghton Mifflin Boston, (2002).
5. Jakub, Thomas D: Estimating Ocean Currents from Automatic Identification System Based Ship Drift Measurements, University of Colorado at Boulder, (2013).