Propositions (Stellingen)

by Fabrice Pierre Robert Colas, author of

Data Mining Scenarios
for the Discovery of Subtypes and
the Comparison of Algorithms

1. In the discovery of valid subtypes in data, there are many more steps involved than only the selection and application of the clustering algorithm. [This thesis]

2. Identifying subtypes can be seen as a data mining scenario that can be applied to different application areas, yet, the step of validation of the subtypes remains specific to the application area. [This thesis]

3. In binary text classification problems, the support vector machines are subject to a performance deterioration for particular combinations of number of features and documents. [This thesis]

4. In large bag of words feature spaces, tightly constrained support vector machines with small C tend to be high performing. However, in this setting, the support vector machines solution is equivalent to a nearest mean classifier. [This thesis]

5. Data mining focuses on generating hypothesis, statistics on validating them. Both research disciplines are complementary and answer different research questions.

6. The R platform for statistical computing is a convenient means to implement data mining scenarios, to extend them to parallel computing environments and to repeat previous research via the use of software packages.

7. In bioinformatics, considerable scientific interaction is necessary to develop a data mining scenario.

8. The most important phase in data analysis is the preparation of the data: known factors of variability must be identified and eventually accounted for.

9. As the language is an instrument of symbolic power, having difficulties with it in work or life leads to professional or social weaknesses. It is the key to social integration.

10. We may feel reassured to try to name and know every single thing. However, the task is both boundless and useless. Focusing will only lead to greater tensions whereas greater attention could enable us to experience life more fully.