The benefits of hybrid methodologies in Data Mining and Knowledge Discovery. Applications to Human Behaviour

K. Gibert

Knowledge Engineering and Machine Learning group, Universitat Politècnica de Catalunya
Dep. Statistics and Operations Research, Universitat Politècnica de Catalunya
Ed. C5. Pta 2. Campus Nord. C. Jordi Girona 1-3. Barcelona 08028. (karina.gibert@upc.edu)

Data Mining, also known quite often as Knowledge Discovery from Databases (KDD) is a quite recent discipline oriented to extract useful knowledge from data. The results of KDD analyses fulfill the main goal of providing non-trivial new knowledge useful for:

a) providing decision support
b) contribute to better manage or control of highly complex organizations or processes

even when massive data sets have to be faced.

KDD is a global umbrella providing a methodological framework in which

A)pre and post-processing are also considered as important parts of the process as data exploitation itself.

B) Both qualitative and quantitative data are rigourously analyzed

C) Prior domain knowledge should be used for improving quality of discovered knowledge

The data exploitation itself is in the kernel of the KDD process and many different data mining techniques, can be considered, depending on the target problem and data structure, among others. Many data mining techniques come from Statistical field (particularly from multivariate analysis) and work well with numerical data. Another important source of techniques is Artificial Intelligence (AI) , which use to perform better with qualitative data.

However, in real applications, frequently numerical and qualitative variables are relevant for later decision making and it is better to analyze jointly, respecting their original forms. In many occasions, hybrid AI&Stats methodologies for Data Mining and KDD constitute very powerfull tools to cope the structure of very complex domains and provide useful knowlege.

An introduction to the discipline and the main characteristics of the KDD methodological approach will be presented, together with some AI&Stats methodologies and some specific applications related with human behaviour.

References:

[1] Fayyad U, et al 1996: From Data Mining to Knowledge Discovery: An overview. In Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press.

[2] K. Gibert, M. Sànchez-Marrè (2011): Outcomes from the iEMSs Data Mining in the Environmental Sciences Workshop Series. Environmental Modelling and Software doi: 10.1016/j.envsoft.2011.01.009

[3] K. Gibert, C. García-Alonso, L. Salvador Carulla (2010) Integrating clinicians, knowledge and data: expert-based cooperative analysis in healthcare decision support. Health Research Policy and Systems 8(28):1-16