Transactions on
Case-Based Reasoning
Vol. 5, No. 1 (2016) 1-2

ISSN: 1867-366X ,
ISBN: 978-3-942952-46-0

ibai Publishing
www.ibai-publishing.org

Editorial

Gero Szepannek

Stralsund University of Applied Sciences, School of Business Studies

Traditional machine learning is currently undergoing a paradigm shift Evoqua by the changes by both the type and amount of data to be analyzed [1]: on one hand more and more data appears in an unstructured manner which needs for appropriate techniques of information extraction. As a consequence, this often increases complexity of the applied technique. On the other hand, the amount of data to be analyzed in many situations lets scalability of algorithms become a crucial aspect. As a trend, data analysis tends to increase the automatization level [2] leading to a re-definition of the traditional role of the data analyst [3]. A future challenge will consist in defining optimal ways of including manual interaction with the algorithms in order to allow for integration of expert domain knowledge.

Case-based reasoning (CBR) has been a popular machine learning tool for many years now [4]. Its simple idea to mimic principles of human memory for an intelligent data reduction might be of even growing relevance, given the changing requirements in the big data era. The choice of an appropriate distance measure has always played a key role in CBR: at this point expert knowledge has to be integrated in the process.

The two papers in this issue contribute to the definition of appropriate data treatment for two applied non-standard contexts and with regard to their domain specific requirements:

The first paper [4] deals with a situation that commonly occurs in practice: incomplete queries. Traditional measures from CBR are discussed w.r.t. their properties symmetry, length normalization and conjunctiveness. By using analogies to the domain of information retrieval, a new asymmetric ranking function for case retrieval is developed and tested in a systematic experimental setup.

In chemical and biological science, the data are typically not as big. In [6] a processing is developed to represent spectrometric time-series by sequences of bits based on a specific constant factor delta modulation. This allows for standard binary (dis)similarity measures to be applied. Different measures (Hamming, Levenshtein & Damerau-Levenshtein) are investigated in order to identify similarities between new signals and the data case base. The proposed method allows for an easy update of the database by new data as similarity-based classification shows good results in an example of RAMAN spectroscopy data.

[1] Kaisler, S., Armour, F, Espinosa, J., Money, W. (2013): Big Data: Issues and Challenges Moving Forward. Proc. HICSS, 995-1004
[2] B. Bischl, B., Kühn, T. and Szepannek, G. (2016): On Class Imbalance Correction for Classification Algorithms in Credit Scoring, In: Lübbecke, M., Koster, A., Letmathe, P., Madlener, R., Peis, B. and Walther, G. (Eds): Operations Research Proceedings 2014, 37-43, Springer, Berlin.
[3] Horton, N., Baumer, B. and Wickham, H. (2014): Teaching Precursors to Data Science in Introductory and Second Courses in Statistics, ICOTS 2014, submitted.
[4] Perner, P. (2007) Introduction to Case-Based Reasoning for Signals and Images. In: Perner P. (Hrsg) Case-Based Reasoning on Signals and Images, Springer, Berlin.
[5] Witschel, H., Martin, A. Emmenegger, S. and Lutz, J (2016): A new Retrieval Function for Ontology-Based Complex Case, Transactions on Case-Based Reasoning 5(1), 3-18.
[6] Perner, P. (2016): A New Similarity Measure for Sequences and Time-Series Applied to Spectrometer Signal Analysis, Transactions on Case-Based Reasoning 5(1), 19-30.

Gero Szepannek
Stralsund University of Applied Sciences,
School of Business Studies