Transactions on Machine Learning
and Data Mining
Vol. 5, No.1 (2012) 1-2

ISSN: 1865-6781,
ISBN: 978-3-942952-11-8

ibai Publishing
www.ibai-publishing.org

Editorial

Adam Krzyżak

Department of Computer Science and Software Engineering, Concordia University, Montreal, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8

Abstract. Special issue of "Transactions on Machine Learning and Data Mining" includes three papers spanning a wide spectrum of data mining and machine learning fields. The first paper applies contrasting correlations between databases to detect changes of interest. The second paper uses frequent sequential patterns to extract sets of connected pixels to reveal meaningful patterns in satellite radar images of agriculture cultivated areas. The third paper is concerned with recognition of wood species using invariant features based on pore distribution in wood microscopic images. An expert system based on the rules extracted from C4.5 classification tree is proposed and tested with real data.

Keywords: Contrasting correlations, clique, computational complexity, satellite imagery, spatio-temporal mining, recognition of wood porosity, C4.5 tree classifiers

Summary

Special issue of "Transactions on Machine Learning and Data Mining" presents three carefully selected papers from World Congress "The Frontiers in Intelligent Data and Signal Analysis" DSA 2011 - one paper form MLDM 2011and two papers from ICDM 2011. The papers make significant contributions to the fields of data mining and machine learning.

The first paper presents [1] a new algorithms for discovering changes in the correlations of an itemset among contrasted databases to detect potential changes among them. This is accomplished by a simple double-clique algorithm that enumerates itemsets that show higher, but not too high, correlation in one database and lower correlation in another one for contrasting. The itemset correlation is measured by k-way mutual information. The authors are looking for cases when weak

correlation is present in the first database and moderate correlation in the second database, while there is a minimum correlation increase from the first to the second database. The main contribution of the paper lies in proposing an efficient algorithm for identifying itemsets satisfying correlation constraints. This is accomplished by enumerating double-cliques in anti-correlation graphs. In simulation experiments the proposed algorithm performs almost two orders of magnitudes faster than the naive approach.

The second paper [2] presents a novel technique for agricultural monitoring by mining Satellite Image Time Series over agriculture cultivated areas. It uses Grouped Frequent Sequential patterns (GFS-patterns) extended to spatio-temporal context in order to extract sets of connected pixels sharing a similar temporal evolution. It allows to uncover sets of pixels satisfying two properties of cultivated areas, namely being spatially connected and sharing similar temporal evolutions. No prior knowledge of the identified regions is assumed and no distance measures are needed. The general framework of GFS-patterns is extended in two directions. Firstly, the connectivity constraint is used in the search space exploration leading to significant reduction of execution times on real Satellite Image Time Series of cultivated areas. Secondly, simple post-processing with a maximality constraint over the patterns significantly improves efficiency. The experiments were performed on database of real images ADAM (Data Assimilation by Agro-Modeling) of SITS (Centre National d'Etudes Spatiales (2010)) demonstrate that pushing the average connectivity measure constraint, during GFS-pattern extraction is effective to reduce the search space. It is also demonstrated that together with a maximality constraint, the proposed approach is useful to find meaningful patterns in real data.

The third paper [3] proposes a classification system for recognition of wood species based on microscopic images of wood pores. The system introduces two new sets of features invariant to rotations, scale and translations. The features are based on the nearest pore pairs and on pore diameter change distributions. The machine learning algorithm C4.5 was used to generate decision trees and decision rules. In simulation experiments with real wood data 83.7% classification rates were obtained which compare favorably with the performance of human experts.

All three papers provide novel and significant contributions to the field of data mining and machine learning. All proposed algorithms are tested on real data and they perform well. The methodology and algorithms presented in the papers may be applied in other application domains and thus have lasting value. The editors ought to be commanded for choosing valuable and important contributions for the special issue.

References

1. Li, A., Haraguchi, M., Okubo, Y.: Contrasting Correlations by an Efficient Double-Clique Condition. Trans.on MLDM 5(1), 3-22 (2012)
2. Julea, A., Meger, N., Rigotti, C., Trouve, E., Jolivet, R., Bolon, P.: Efficient Spatio-temporal Mining of Satellite Image Time Series for Agricultural Monitoring. Trans.on MLDM 5(1), 23-44 (2012)
3. Pan, S., Kudo, M.: Recognition of Wood Porosity Based on Direction Insensitive Feature Sets. Trans.on MLDM 5(1), 23-44 (2012)