

# Boxes and dual clusters for formal concept analysis and biclustering

Boris Mirkin

Birkbeck University of London, UK, [mirkin@dcs.bbk.ac.uk](mailto:mirkin@dcs.bbk.ac.uk)

We apply the data recovery approach to two areas of research: formal concept analysis and bi-clustering. The former involves lattice-theoretic analysis of one-only rectangles on a zero-one data matrix, the other is oriented towards finding high-density rectangles on the data. Box model is of an SVD-type approximation of the data with a binary matrix in which all entries within a box (rectangle) are equal to one positive value, alpha, whereas all the other entries are equal to another real, beta. The corresponding equation involves  $N \times M$  matrices of the size of the data. Relations between optimal boxes and formal concepts are analyzed. The equation leads to two other equations involving  $M \times M$  and  $N \times N$  matrices, respectively. The corresponding binary solutions are referred to as dual clusters. The matrices generate a reasonable similarity measure between binary data. A three-sided alternating minimization method for finding (a set of) dual clusters is proposed and tried in a number of experiments and real world data sets.