R. Belohlavek, J. Outrata, M. Trnecka: Impact of Boolean factorization on classification

Department of Computer Science, Palacky University, Olomouc (17. listopadu 12, CZ–77146 Olomouc, Czech Republic)

Introduction

:: classification of Boolean data: objects described by Boolean (binary, yes-no) input attributes and assigned a (many-valued) class label

:: preprocessing of data to improve the quality of classification – wanted for Boolean data with a clear semantics

:: in our previous papers: preprocessing based on Boolean matrix factorization (BMF) – replacing input attributes by factors computed from the attributes

:: factors = (some) formal concepts (from Formal Concept Analysis) of input data – clear meaning, easy interpretation

:: classification using a smaller number of factors as new attributes, yet with improvement of the quality

:: how various BMF algorithms impact the quality of classification?

Boolean Matrix Factorization (BMF)

= decomposition of Boolean matrix I to Boolean matrices A and B s.t. I is (approximately) equal to $A \circ B$, $(A \circ B)_{ij} = \vee_{l=1}^{k} A_{il} \cdot B_{lj}$.

$$\begin{pmatrix} 1&1&0&0&0\\ 1&1&0&0&1\\ 1&1&1&1&0\\ 1&0&0&0&1 \end{pmatrix} = \begin{pmatrix} 1&0&0&1\\ 1&0&1&0\\ 1&1&0&0\\ 0&0&1&0 \end{pmatrix} \circ \begin{pmatrix} 1&1&0&0&0\\ 0&0&1&1&0\\ 1&0&0&0&1\\ 0&1&0&0&0 \end{pmatrix}$$

:: I ... object-attribute Boolean data, A ... object-factor (usage) matrix, B ... factor-attribute (basis vector) matrix

~ discovery of k factors (as new Boolean attributes) that (approximately) explain the data – factor l represented by lth column of A and lth row of B:

$A_{il} = 1$ ... factor l applies to object i
$B_{lj} = 1$ ... attribute j is a manifestation of factor l

:: $(A \circ B)_{ij}$ ... "object i has attribute j iff there is a factor l such that l applies to i and j is a manifestation of l"

Problem 1, discrete basis problem (Miettinen et al., 2008):
For I and positive k, find A and B minimizing

$$||I - A \circ B|| = \Sigma_{i,j} |I_{ij} - (A \circ B)_{ij}|.$$

... the importance of the first k (most important) factors

Problem 2 (Belohlavek, Vychodil, 2010):
For I and positive integer ε, find A and B with k as small as possible such that $||I - A \circ B|| \leq \varepsilon$.
... the need to account for (and thus to explain) a prescribed portion of data

:: NP-hard optimization problems ⇒ approximation algorithms

BMF in preprocessing

Boolean data for classification $\langle X, Y, I, c \rangle$:
$X = \{1, \ldots, n\}$ ... objects
$Y = \{1, \ldots, m\}$ ... input Boolean attributes
$I \subseteq X \times Y$ ... object-attribute relational Boolean data
$c : X \to C$ ... class attribute, mapping assigning to each object $i \in X$ its class label $c(i) \in C$

Preprocessing

:: (approximately) decompose I to A and B, either as Problem 1 or Problem 2

:: $\langle X, F, A, c \rangle$ ... new Boolean data for classification, $F = \{1, \ldots, k\}$ ... factors = new Boolean attributes

Classification

:: use any classification model (e.g. a decision tree) developed for $\langle X, F, A, c \rangle$ to classify objects described by the original Boolean attributes from Y:

for object represented by vector P, apply the model to $\mathbf{g(P)}$ using transformations $g : \{0,1\}^m \to \{0,1\}^k$ and $h : \{0,1\}^k \to \{0,1\}^m$ between the space of the original attributes and the space of factors

$$(\mathbf{g(P)})_\mathbf{l} = \wedge_{\mathbf{j=1}}^{\mathbf{m}} (\mathbf{B_{lj}} \to \mathbf{P_j}) \quad \text{and} \quad (h(Q))_j = \vee_{l=1}^{k} (Q_l \cdot B_{lj})$$

:: usually $k < m$ (⇒ reduction of dimensionality of data) and g is not an injective mapping

→ assigning to objects in $\langle X, F, A, c \rangle$ with equal $g(P)$ representations transformed from the objects in $\langle X, Y, I, c \rangle$ with different P representations and different assigned class labels the majority class label of those class labels

BMF Algorithms

:: Tiling/GreCon (Geerts et al./Belohlavek, Vychodil) – columns of A and rows of B are objects and attributes of tiles/formal concepts of I greedily selected to maximize the drop of $||I - A \circ B||$

:: GreConD (Belohlavek, Vychodil) – as GreCon, formal concepts computed on demand (instead of in advance)

:: Asso (Miettinen et al.) – for k required factors, greedy selection of rows of B from rows of C and computation of columns of A so that $A \circ B$ approximates I, C ... $m \times m$ matrix, $C_{ij} = 1$ if confidence of association rule $\{i\} \Rightarrow \{j\}$ in I is at least parameter τ

:: GreEss (Belohlavek, Trnecka) – as GreConD, formal concepts selected from intervals in concept lattice of I defined by factors/formal concepts from first factorization of Boolean matrix of so-called essential 1s of I

:: GreEssQ (Belohlavek, Trnecka) – as GreEss, formal concepts of the Boolean matrix of essential 1s of I, selected with additional information regarding a kind of quality of 1s

Experimental Evaluation

:: evaluating and comparing classification accuracy of models created by machine learning methods: decision tree C4.5, Nearest Neighbor (NN), Naive Bayes (NB), neural network trained by back propagation (MLP)

:: datasets selected from UCI ML Repository:

dataset	attrs (binary)	objects	class distribution
breast-cancer	9 (51)	277	196/81
car	6 (21)	1728	1210/384/69/65
kr-vs-kp	36 (74)	1598	835/763
mushroom	22 (125)	564	362/202
solar-flare 2	12 (45)	1066	147/211/239/95/43/331
zoo	15 (30)	101	41/20/5/13/4/8/10

:: binary attributes were obtained by plain nominal scaling

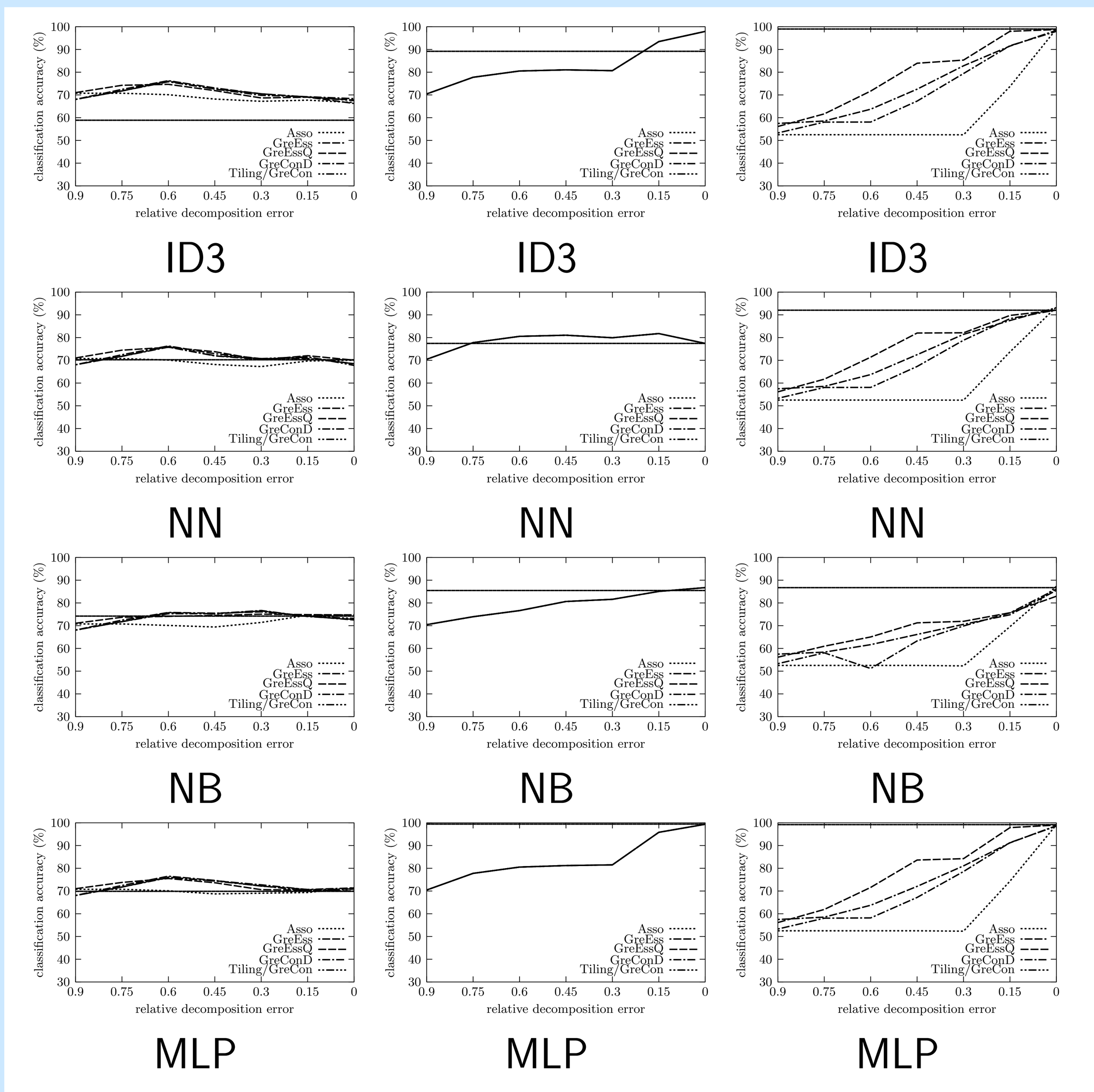
Results

:: for testing parts of the datasets

:: Tiling/GreCon, GreConD, Asso, GreEss, GreEssQ, original input data

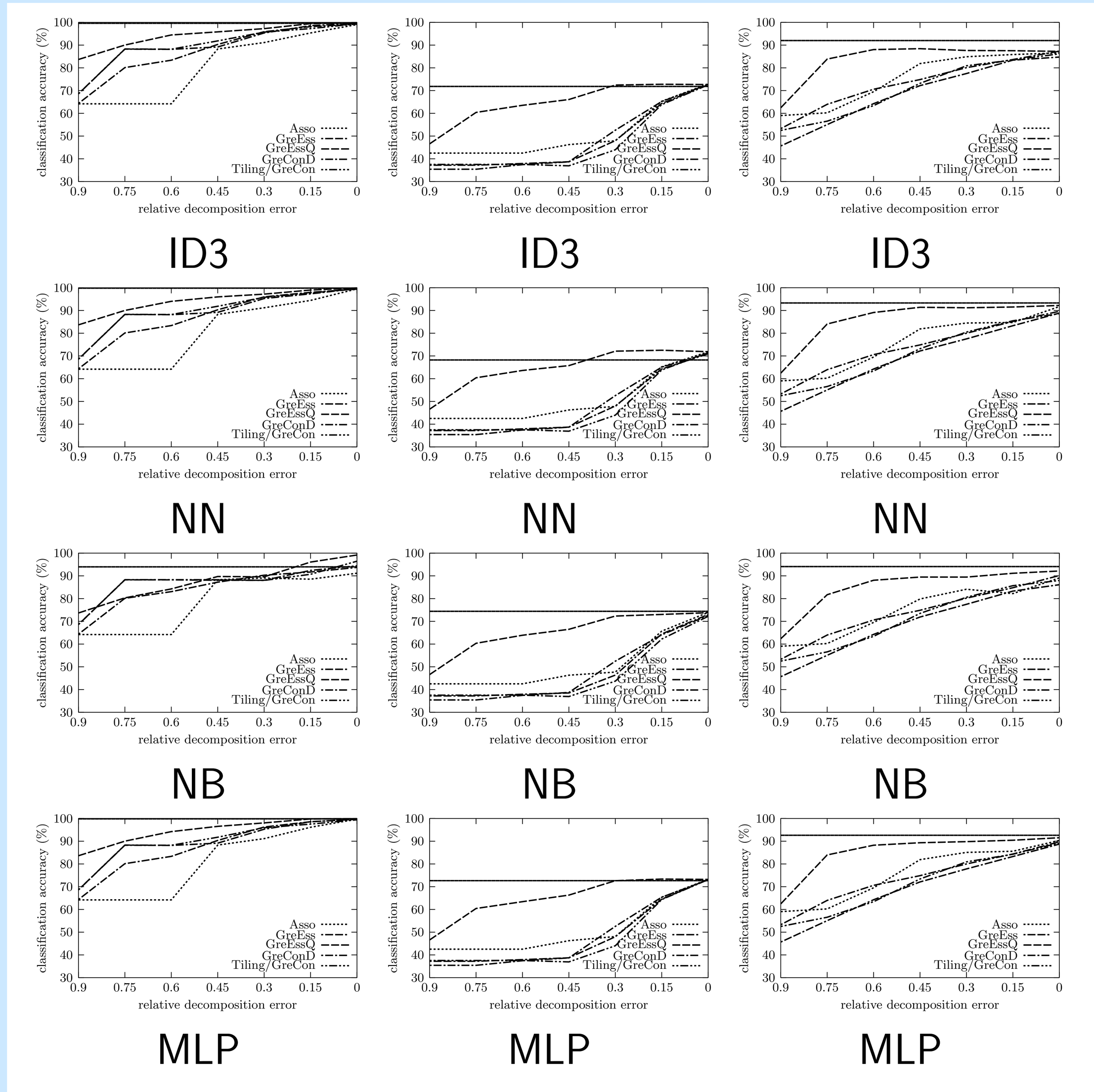
:: x-axis ... relative decomposition error = ratio of 1s of input data left uncovered in factor decomposition

breast-cancer car kr-vs-kp



Experimental Evaluation

mushroom solar-flare 2 zoo



Observations

:: GreEssQ outperforms all other algorithms – so-called attribute concepts utilized as factors, covering areas of 1s of input data matrix covered by small number of factors

:: poor classification performance of Asso– factors need not be rectangular areas full of 1s in input data matrix

:: similar performance of Tiling/GreCon, GreConD and GreEss – similar factorizations or strategy

:: sometimes preprocessed data lead to better classification accuracy than original data, even with a few factors covering less than 100 % of input data, e.g. for car, kr-vs-kp, mushroom or solar-flare 2 datasets

:: car dataset: same performance of all BMF algorithms – very natural found factors

:: breast-cancer dataset: preprocessed data are (much) better classified than original data (best with 40% coverage of input data by factors!)– many superfluous attributes or large noise in input data overcome by factors

Acknowledgment

Supported by the ESF project No. CZ.1.07/2.3.00/20.0059 (R. Belohlavek) and grants No. 202/10/P360 of the Czech Science Foundation (J. Outrata) and by the IGA of Palacky University, No. PrF 20124029 (M. Trnecka).

International Center for Information and Uncertainty
SSIU 2013: International Summer School "Information and Uncertainty"
http://mcin.upol.cz/SSWIU-2013/

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT