Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 292 (2016) 85–97
FUZZY sets and systems
www.elsevier.com/locate/fss

Factorization of matrices with grades

Radim Belohlavek *, Vilem Vychodil

Dept. Computer Science, Palacký University, Czech Republic

Received 11 February 2014; received in revised form 22 January 2015; accepted 24 March 2015

Available online 1 April 2015

Dedicated to Francesc Esteva on the occasion of his 70th birthday

Abstract

We present an approach to decomposition and factor analysis of matrices with ordinal data. The matrix entries are grades to which objects represented by rows satisfy attributes represented by columns, e.g. grades to which an image is red, a product has a given feature, or a person performs well in a test. We assume that the grades are taken from bounded scales equipped with certain aggregation operators that are involved in the decompositions. Particular cases of the decompositions include the well-known Boolean matrix decomposition, and the sup-t-norm and inf-residuum decompositions. We consider the problem of decomposition of a given matrix into a product of two matrices with grades such that the number of factors, i.e. the inner dimension, be as small as possible. We observe that computing such decompositions is NP-hard and present a greedy approximation algorithm. Our algorithm is based on a geometric insight provided by a theorem identifying particular rectangular-shaped submatrices as optimal factors for the decompositions. These factors correspond to fixpoints of certain Galois connections associated with the input matrix, which are called formal concepts, and allow an easy interpretation of the decomposition. We present illustrative examples and experimental evaluation of the algorithm.

Keywords: Matrix decomposition; Ordinal data; Residuation; Galois connection; Algorithm

1. Introduction

1.1. Problem description

In traditional approaches to dimensionality reduction, such as factor analysis, a decomposition (factorization) of an object–variable matrix is sought into an object–factor matrix and a factor-variable matrix with the number of factors reasonably small. The factors are considered as new variables, hidden in the data and likely more fundamental than the original variables. Computing the factors and interpreting them is the central topic of this paper.

We consider decompositions of matrices I with a particular type of ordinal data. Namely, each entry I_{ij} of I represents a grade which the object corresponding to the ith row has, or is incident with, the attribute corresponding to

* Corresponding author.
E-mail addresses: radim.belohlavek@acm.org (R. Belohlavek), vychodil@acm.org (V. Vychodil).

http://dx.doi.org/10.1016/j.fss.2015.03.020
0165-0114/

the jth column. Examples of such data are results of questionnaires where respondents (rows) rate services, products, etc., according to various criteria (columns); results of performance evaluation of people (rows) by various tests (columns); or binary data in which case there are only two grades, 0 (no, failure) and 1 (yes, success). Our goal is to decompose an $n \times m$ object–attribute matrix I into a product

$$I = A \circ B \tag{1}$$

of an $n \times k$ object–factor matrix A and a $k \times m$ factor–attribute matrix B with the number k of factors as small as possible.

The scenario is thus similar to ordinary matrix decomposition problems but there are important differences. First, we assume that the entries of I, i.e. the grades, as well as the entries of A and B are taken from bounded scales L of grades, such as the real unit interval $L = [0, 1]$ or the Likert scale $L = \{1, \ldots, 5\}$ of degrees of satisfaction. Second, the matrix composition operation $\circ$ used in our decompositions is not the usual matrix product. Instead, we use a general product based on supremum-preserving aggregation operators introduced in [5,6], see also [12]. Two important, well-known [2,11] cases of this product are the sup-t-norm-product defined by

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj}, \tag{2}$$

and the inf-residuum-product (denoted also by $\triangleleft$) defined by

$$(A \circ B)_{ij} = \bigwedge_{l=1}^{k} A_{il} \rightarrow B_{lj}, \tag{3}$$

where $\otimes$ and $\rightarrow$ denote a (left-)continuous t-norm and its residuum [11,15], and $\bigvee$ and $\bigwedge$ denote the supremum and infimum. The ordinary Boolean matrix product is a particular case of the sup-t-norm product in which the scale L has 0 and 1 as the only grades and $a \otimes b = \min(a, b)$. It is to be emphasized that we attempt to treat graded incidence data in a way which is compatible with its semantics. This need has been recognized long ago in mathematical psychology, in particular in measurement theory [16]. For example, even if we represent the grades by numbers such as $0 \sim$ strongly disagree, $\frac{1}{4} \sim$ disagree, $\ldots$, $1 \sim$ strongly agree, addition, multiplication by real numbers, and linear combination of graded incidence data may not have natural meaning. Consequently, decomposition of a matrix I with grades into the ordinary matrix product of arbitrary real-valued matrices A and B may suffer from a difficulty to interpret A and B, see [20,28]. In this paper, we present an algorithm which is based on a theorem from [6] regarding the role of fixpoints of certain Galois connections associated with I as factors for decomposition of I. This is important both from the technical viewpoint, since due to [6] optimal decompositions may be obtained this way, and the knowledge discovery viewpoint, since the fixpoints, called formal concepts may naturally be interpreted. The algorithm runs in polynomial time and delivers suboptimal decompositions. This is a necessity because, as we show, computing optimal decompositions is an NP-hard optimization problem. In addition, we present an illustrative example demonstrating the usefulness of such decompositions, and an experimental evaluation of the algorithm.

1.2. Related work

Recently, new methods of matrix decomposition and dimensionality reduction have been developed. One aim is to have methods which are capable of discovering possibly non-linear relationships between the original space and the lower dimensional space [23,29]. Another is driven by the need to take into account constraints imposed by the semantics of the data. Examples include nonnegative matrix factorization, in which the matrices are constrained to those with nonnegative entries and which leads to additive parts-based discovery of features in data [17]. Another example, relevant to this paper, is Boolean matrix decomposition. Early work on this problem was done in [22,26]. Recent work on this topic includes [7,9,19,20,22]. As was mentioned above, Boolean matrix decomposition is a particular case of the problem considered in this paper. Note also that partly related to this paper are methods for decomposition of binary matrices into non-binary ones such as [18,24,25,27,31], see also [28] for further references.

2. Decomposition and factors

2.1. Decomposition and the factor model

As was mentioned above, we assume that the matrix entries contain elements from scales (grades) equipped with certain aggregation operators. In particular, we assume a general model of (1) in which the entries of A, B, and I