Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 109 (2015) 46–63
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Knowledge reduction in formal contexts using non-negative matrix factorization ☆

Ch. Aswani Kumar [a,*], Sérgio M. Dias [b,c], Newton J. Vieira [b]

[a] *School of Information Technology and Engineering, VIT University, Vellore - 632014, India*
[b] *Department of Computer Science, Federal University of Minas Gerais (UFMG), Av. Antônio Carlos 6627 - ICEx - 4010, Pampulha, Belo Horizonte, 31.270-901, Minas Gerais, Brazil*
[c] *Federal Service of Data Processing (SERPRO), Av. José Cândido da Silveira, 1.200, Cidade Nova, Belo Horizonte, 31.035-536, MG, Brazil*

Received 3 December 2012; received in revised form 13 July 2014; accepted 16 August 2014
Available online 26 August 2014

Abstract

Formal Concept Analysis (FCA) is a mathematical framework that offers conceptual data analysis and knowledge discovery. One of the main issues of knowledge discovery is knowledge reduction. The objective of this paper is to investigate the knowledge reduction in FCA and propose a method based on Non-Negative Matrix Factorization (NMF) for addressing the issue. Experiments on real world and benchmark datasets offer the evidence for the performance of the proposed method.

Keywords: Concept lattice; Formal concept analysis; Knowledge reduction; Non-negative matrix factorization; Singular value decomposition

1. Introduction

Introduced by Rudolf Wille in the mid 80s, Formal Concept Analysis (FCA) has brought mathematical thinking for knowledge representation and discovery. FCA is formulated based on the two important notions: formal context and formal concept. A formal context consists of a set of objects, set of attributes and a binary relation specifying which objects have what attributes. A formal context is modeled as a cross table, with rows representing the formal objects, columns representing the formal attributes and the crosses representing the relations between them. Concepts derived from the formal context sorted by the inclusion order derived from the formal context sorted by the inclusion order ($\subseteq$) forms a complete lattice known as concept lattice. FCA organizes the information through concept lattices which fundamentally comprises partial order, reflecting the relationship of generalization and specialization among the concepts of the context. As an effective tool for data analysis, knowledge representation and processing, FCA has been widely studied and applied in many diverse scientific fields [1,42].

☆ This paper is an extended version of "Mining association rules using non-negative matrix factorization and formal concept analysis" In Proceedings of ICIP 2012, Springer, CCIS 157, pp. 31–39.

* Corresponding author. Tel.: +91 416 2202771.
E-mail addresses: cherukuri@acm.org, aswanis@gmail.com (Ch. Aswani Kumar), mariano@dcc.ufmg.br, sergio.dias@serpro.gov.br (S.M. Dias), nvieira@dcc.ufmg.br (N.J. Vieira).

http://dx.doi.org/10.1016/j.matcom.2014.08.004

Two central issues in FCA based knowledge discovery are the knowledge representation and knowledge reduction. Recently, there is a growing interest among FCA research community on knowledge reduction. Formal contexts of modest size can produce thousands of formal concepts, further resulting in unreadable and unmanageable concept lattices [28,8]. Another pertinent issue is the density and noise within a context that increase the number of formal concepts. Hence the basic problem in FCA is to find minimal contextual structure which avoids the redundancy while maintaining the structure consistency. In several practical applications the data is of high dimensional in nature and hence often it is required to uncover its low dimensional structure. Matrix decomposition techniques are proved to be successful in this task. However, while revealing the low dimensionality, it is necessary to preserve the non-negative character of the data. NMF has received significant attention from research communities due to its character of preserving the non-negative property of data. NMF is a multivariate data analysis method that presents original data matrix into the product of basis and encoding matrices with non-negative restrictions. Extending upon the analysis of Snasel et al. [49] and Cherukuri [7] the present paper explore the NMF based matrix decomposition for knowledge reduction in formal context. Rest of the paper is presented as follows. To facilitate our discussion and make the paper self-contained, basic concepts in FCA are first introduced in Section 2 and further the knowledge reduction in formal contexts with its related work is described. Non-negative matrix factorization is illustrated in Section 3. To reduce the formal context, we introduce a new method based on NMF in Section 4. We demonstrate the experiments using the proposed method and present the analysis of the results in Section 5.

2. Background

This section recalls the notions and terminology of FCA. Also we summarize the available research on knowledge reduction in FCA. The mathematical principles for this domain were established by Birkhoff in 1960's and the current form of FCA framework was introduced by Wille in 1980's.

2.1. Formal contexts, lattices and FCA

Based upon the partial ordering relations, FCA is one of the effective mathematical framework which represents conceptual hierarchy of data as an algebraic lattice [52,54,59].

Definition 1. A formal context or a dyadic context K is a triple (X, Y, I), where X, called the universe of discourse, is a nonempty and finite set of objects, Y is a nonempty finite set of attributes and $I \subseteq XxY$ is a binary relation between X and Y.

Definition 2. For a formal context K, operators $\uparrow$: $2^X \to 2^Y$ and $\downarrow$: $2^Y \to 2^X$ are defined for every $A \subseteq X$ and $B \subseteq Y$ by $A^\uparrow = \{y \in Y \mid \text{for each } x \in A : (x, y) \in I\}$, $B^\downarrow = \{x \in X \mid \text{for each} y \in B : (x, y) \in I\}$. The operators $\uparrow$ and $\downarrow$ are known as concept forming operators or derivator operators.

Definition 3. A formal concept of the context $K : (X, Y, I)$, is a pair (A, B) of $A \subseteq X$, $B \subseteq Y$, such that $A^\uparrow = B$ and $B^\downarrow = A$. We call A as extent and B as intent of the concept (A, B). Formal concepts are naturally ordered by partial order $\leq$ using sub concept super concept relation, such that for any two formal concepts (A_1, B_1) and (A_2, B_2), $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$.

Extent of a concept represents the objects sharing the same attributes. The objects sharing few or more attributes can be found in the extent of the neighboring concepts that are connected to the current concept. This can be regarded as a form of clustering activity.

Definition 4. The collection of all formal concepts of the context $K : (X, Y, I)$ order by $\leq$ is called a concept lattice.

In contrast to other partial order structures like trees, lattices allow multiple inheritances which facilitate more structured simulation. The lattice structure allows many paths to a particular node while classical hierarchical structures restrict each node to possess only one parent. The sub and super concept relation described by the lattice is a transitive relation. It means that, a concept is the sub concept of any concept that can be reached by traveling upwards from it. The top node of the lattice structure represents generalization capability of FCA by exhibiting all objects in its extent. Similarly, the bottom node of the lattice represents specialization ability of FCA by displaying all attributes of the context in its intent. Based on these generalization and specialization activity when an object-g (attribute-m) attached to a node, all the nodes above (below) it must also contain the object-g (attribute-m).