ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

0-1 multilinear programming as a unifying theory for LAD pattern generation[☆]

Kedong Yan [a], Hong Seo Ryoo [b,*]

[a] Department of Information Management Engineering, Graduate School of Information Management & Security, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul, 02841, Republic of Korea

[b] School of Industrial Management Engineering, College of Engineering, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul, 02841, Republic of Korea

ARTICLE INFO

Article history:
Received 9 July 2015
Received in revised form 30 November 2015
Accepted 9 August 2016
Available online xxxx

Keywords:
Boolean logic
Logical analysis of data
Pattern generation
Multilinear programming
0-1 linearization

ABSTRACT

This paper revisits the Boolean logical requirement of a pattern and develops 0-1 multilinear programming (MP) models for (Pareto-)optimal patterns for logical analysis of data (LAD). We show that all existing and also new pattern generation models can naturally be obtained from the MP models via linearization techniques for 0-1 multilinear functions. Furthermore, 0-1 MP provides an insight for understanding how different and independently developed models for a particular type of pattern are inter-related. These show that 0-1 MP presents a unifying theory for pattern generation in LAD.

1. Introduction

LAD is a Boolean logic-based data analytics methodology [14,22,35] that has aroused much interest in optimization and data mining communities during the past two decades and has proven useful in practical decisions [1,2,4,5,24,29,31]. For separation of a finite set of two different types of observations/data, without loss of generality, a critical step in LAD successively discovers a piece of structural information that distinguishes one or more of one type of data from all of the other type. Such knowledge is called a pattern, which corresponds to a conjunction of literals in Boolean logic, where a literal refers to a 0-1 feature or its negation. Patterns are the building blocks of a LAD classification theory, and the construction of a pattern with respect to a desired pattern selection criterion forms the key stage in data analytics via LAD. The difficulty is that pattern generation is a combinatorial optimization problem that forms the bottleneck stage in the application of LAD.

Deferring the presentation of the background material until Section 2, pattern generation approaches in the literature can be classified as either enumeration-based [12,23,26] or optimization-based [10,20,27,35]. In note of numerical difficulties associated with pattern generation, term-enumeration methods are first developed, and they rely on simple rules for constructing terms that merely satisfy certain requirements imposed on patterns. As heuristic approaches, these have limitations in identifying a pattern that is optimal or Pareto-optimal with respect to one or more pattern preference, however. As a result, these methods ironically suffer from a poor average run-time complexity, for the number of LAD patterns made

☆ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant Number: NRF-2013R1A1A2011784).

* Corresponding author.
E-mail addresses: yankedong@korea.ac.kr (K. Yan), hsryoo@korea.ac.kr (H.S. Ryoo).

http://dx.doi.org/10.1016/j.dam.2016.08.007

up of d of the total of n literals can be as many as $\binom{n}{d}2^d$. To alleviate limitations of the term-enumeration methods in finding useful patterns, tools have also been developed for transforming an existing pattern into a Pareto-optimal one in polynomial run time in the number of terms of d literals for $d \in \{1, \dots, n\}$ [3,6,26], which is exponentially many, however.

Unlike enumeration methods, optimization-based approaches can generate optimal and Pareto-optimal patterns and also patterns of different degrees with an equal amount of efforts and suffer from the numerical complexity of the associated optimization problem only in the worst case. In brief, for identifying a pattern with a maximum coverage among those that distinguish a reference data A_ℓ from all data of the other type, [23] first presented a polynomial set covering formulation and used the best linear approximation of the pseudo-Boolean functions from [25] to identify a pattern that is almost an 'A_ℓ-maximum' pattern. Subsequently, the authors linearized the same polynomial set covering model to obtain an optimization-based approach for generating A_ℓ-maximum patterns in [10]. On a related note, [9] considers a combined pattern generation and theory formation problem as a way to give rise to a 'large margin' theory that supposedly performs well in testing. The same formulation appears in [27] within a column generation scheme for large-scale integer programming in search of a way to find 'near/fuzzy patterns' that perform well in testing. Although optimization-flavored, these two methods are neither designed nor guaranteed to generate patterns, hence can be classified as heuristic methods.

For optimization-based pattern generation, [35] gave the first full treatment and presented a set of mixed integer and linear programming (MILP) models for constructing various optimal and Pareto-optimal LAD patterns. Through extensive experiments, [35] also demonstrated advantages of optimization-based approaches over enumeration-based methods. In [20], a set of MILP models in a much smaller number of 0-1 decision variables was presented for a more efficient generation of useful LAD patterns via optimization principles. Equipped with new models that are much easily solved, [20] also demonstrated different utilities of strong prime and strong spanned patterns in their use in classifying new data.

Motivated by a surge of interest in more efficient optimization-based LAD pattern generation, this paper revisits the Boolean logical definition of a LAD pattern and presents a unifying theory of LAD pattern generation in 0-1 MP. Specifically, we use the Boolean logical definition of a LAD pattern to first develop 0-1 MP models for well-studied (Pareto-)optimal LAD patterns. Next, we demonstrate that all existing pattern generation models from the literature and also new ones can be obtained from these MP models via 0-1 linearization techniques for multilinear functions. Furthermore, we demonstrate that 0-1 MP provides an insight for understanding how all different and independently developed pattern generation models for a particular type of LAD pattern are inter-related. In summary, these show that 0-1 MP holds a unifying theory for pattern generation in LAD (refer Fig. 1.)

As for the organization of this paper, Section 2 provides a brief background on LAD pattern, and Section 3 develops 0-1 MP models for optimal and Pareto-optimal LAD patterns that are well-studied in the literature. In Section 4, we use McCormick's envelopes for multilinear functions, standard probing techniques in integer programming and also new valid inequalities for 0-1 multilinear functions to demonstrate how the 0-1 MILP/IP pattern generation models from the literature and new models can be obtained from the MP models of Section 3. As a bonus treatment, Section 5 illustrates the construction of different pattern generation models on a small dataset and demonstrates their relative efficiency in pattern generation on six machine learning benchmark datasets. Results in this section seem to indicate that the efficiency in pattern generation by the 0-1 linear models is proportionally related to the advancedness of linearization techniques used for their derivation. This invites more attention and efforts to be directed toward the development of more advanced 0-1 linearization techniques and stronger valid inequalities for 0-1 multilinear functions. Finally, concluding remarks are provided in Section 6.

2. Background on LAD pattern

Without loss of generality, let us consider the separation of a finite number of $+$ and $-$ observations/data. For $\bullet \in \{+, -\}$, let us denote by $S^\bullet$ the index set of $m^\bullet$ observations of $\bullet$ type. We assume that the dataset is contradiction and duplicate free and let $S^+ \cap S^- = \emptyset$ and $S = S^+ \cup S^-$. Denote by $a_1, \dots, a_n$ the n binary features that uniquely describe each data A_i, $i \in S$. Let a_{ij} denote the binary value of the jth feature of A_i for $i \in S$.

In Boolean algebra, a *term* refers to a conjunction of one or more literals, where a literal is either a feature a_j or its complement $\neg a_j$ (or $\bar{a}_j$) for $j \in N := \{1, \dots, n\}$. Therefore, a term takes the form

$$T := \bigwedge_{j \in N_1} a_j \bigwedge_{j \in N_2} (\neg a_j),$$

where $N_1, N_2 \subseteq N$ such that $N_1 \cap N_2 = \emptyset$. We say that a term T *covers* A_i if

$$T(A_i) := \bigwedge_{j \in N_1} a_{ij} \bigwedge_{j \in N_2} (\neg a_{ij}) = 1.$$

In LAD, a term is called a $+$ $(-)$ *pattern* if it distinguishes at least one $+$ $(-)$ observation from all $-$ $(+)$ observations. That is, if it covers one or more $+$ $(-)$ data and none of the $-$ $(+)$ data. (In note of the symmetry in the definitions of $+$ and $-$ patterns, we deal with the generation of $+$ patterns in this paper for convenience in presentation.) Thus, a term T is a pattern if and only if $T(A_i) = 1$ for some $i \in S^+$ and $T(A_i) = 0$, $\forall i \in S^-$.

Let us introduce n additional 0-1 features a_{n+j} to negate a_j as $a_{n+j} := \neg a_j = 1 - a_j$ for $j \in N$. Let $N' := \{n+1, \dots, 2n\}$ and $\mathcal{N} := N \cup N'$. Now, each A_i, $i \in S$, is uniquely described by $2n$ 0-1 features/attributes a_j, $j \in \mathcal{N}$. For constructing patterns,