Expert Systems with Applications 42 (2015) 4805–4817

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Variable neighborhood search heuristics for a test assembly design problem

Jordi Pereira [a,b,*], Mariona Vilà [b]

[a] Departamento de Ingeniería Industrial, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile
[b] Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona, Universitat Politècnica de Catalunya. C. Comte Urgell, 187 1st. Floor, 08036 Barcelona, Spain

ARTICLE INFO

Article history:
Available online 4 February 2015

Keywords:
Bin packing
Test assembly design
Test splitting
Variable neighborhood search

ABSTRACT

Test assembly design problems appear in the areas of psychology and education, among others. The goal of these problems is to construct one or multiple tests to evaluate the test subject. This paper studies a recent formulation of the problem known as the one-dimensional minimax bin-packing problem with bin size constraints (MINIMAX_BSC). In the MINIMAX_BSC, items are initially divided into groups and multiple tests need to be constructed using a single item from each group, while minimizing differences among the tests. We first show that the problem is NP-Hard, which remained an open question. Second, we propose three different local search neighborhoods derived from the exact resolution of special cases of the problem, and combine them into a variable neighborhood search (VNS) metaheuristic. Finally, we test the proposed algorithm using real-life-based instances. The results show that the algorithm is able to obtain optimal or near-optimal solutions for instances with up to 60000-item pools. Consequently, the algorithm is a viable option to design large-scale tests, as well as to provide tests for online small-sized situations such as those found in e-learning platforms.

1. Introduction

Test assembly design studies the problem of selecting the optimal subset of items (questions) among those available to define one or several tests (questionnaires) subject to specific constraints and objectives. This problem appears in multiple areas, mainly in test design (van der Linder, 2005) both for education (Porter, Polikoff, Barghaus, & Yang, 2013; Sun, Chen, Tsai, & Cheng, 2008) and psychology studies (Veldkamp, 2005).

Based on the specific objective of the test, as well as the characteristics of the questions and questionnaires, different problems and formulations may appear. One of these formulations arises when considering the problem as a packing problem, specifically as a one dimensional bin packing problem (BPP) with additional constraints (see Dyckhoff (1990) for a classification of packing problems, and Wäscher, Haußner, & Schumann (2007) for an update of the previous classification). In the BPP formulation, the questions and questionnaires are interpreted as the items and the bins of the BPP respectively, and the weight of the items corresponds to the metric (e.g., the difficulty of the question) that needs to be considered during the test design. The objective of the problem is to find disjoint subsets of all the items so that the sum of weights (i.e., the total difficulty) of each subset is as evenly matched as possible.

In this work we consider the formulation introduced in Brusco, Köhn, and Steinley (2013). This formulation is known as the one-dimensional minimax bin-packing problem with bin size constraints (MINIMAX_BSC) and it departs from the classical BPP by considering that the questions have been previously grouped and questionnaires are constructed selecting a single question from each group. To solve the problem, the authors propose a mixed zero-one integer linear programming model that is then solved using the CPLEX commercial software. For large real-life problems, the authors propose a simulated annealing (SA) algorithm that outperforms the quality of the solutions provided by CPLEX.

While Brusco et al. (2013) do not address the complexity of the problem, the analysis of the results showed that all of the algorithms proposed in the paper were able to optimally solve large problems with 2 questionnaires, leading the authors to hypothesize that the special case of 2 questionnaires may be solvable in polynomial time (Brusco et al., 2013, p. 623). Furthermore, the quality of the solutions provided by the SA deteriorates as the number of questionnaires increases, which may suggest possible improvements if different solution methods were used.

* Corresponding author at: Departamento de Ingeniería Industrial, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile.
E-mail addresses: jorgepereira@ucn.cl, jorge.pereira@upc.edu (J. Pereira), mariona.vila.bonilla@upc.edu (M. Vilà).

http://dx.doi.org/10.1016/j.eswa.2015.01.057
0957-4174/

1.1. Contributions of this work

In this paper, we address the complexity of the MINIMAX_BSC and demonstrate that the MINIMAX_BSC is NP-Hard for two questionnaires, and strongly NP-Hard for more than two questionnaires. We also show that the case with two groups of questions can be optimally solved in polynomial time and that the case with two questionnaires can be efficiently solved as a subset sum knapsack problem (Kellerer, Pferschy, & Pisinger, 2004).

Based on these special solvable cases, we derive a new constructive heuristic and three local search neighborhoods. These methods are then combined in a variable neighborhood search (VNS) metaheuristic (Mladenović & Hansen, 1997) to provide a combined approach to solve the problem. VNS is a popular metaheuristic approach (Hansen, Mladenović, & Moreno Pérez, 2010; Mladenović & Hansen, 1997) based on the use and exploration of different local neighborhoods of an incumbent solution and a mechanism to escape from local optima by restarting the search on random neighbor solutions. Both techniques provide a method to systematically explore the solution space.

The results of the method clearly outperform those provided by previous algorithms, such as the SA method proposed in Brusco et al. (2013). More specifically, the proposed method is able to routinely reach a BPP-based lower bound in instances with over ten questions per questionnaire featuring real-life characteristics, and still manages to obtain small deviations from the theoretical bound in other cases.

1.2. Paper outline

The remainder of the paper is structured as follows. In Section 2, we study the problem, describe its mathematical formulation, address the issue of complexity, and identify special cases that can be efficiently solved. We also outline some of the work in the literature devoted to test design and other related problems, such as the BPP. In Section 3, we describe the proposed local searches and a constructive heuristic. We also propose their combination into a VNS. In Section 4, we describe the results of various computational experiments to assess the quality of the proposed algorithm. Finally, in Section 5, we provide conclusions of the present work and identify some future areas of research.

2. Problem description

2.1. Literature review

Test assembly design was early identified as a combinatorial optimization problem. According to van der Linden (1998), an initial work in the field (Birnbaum, 1968) proposed a three step process to design and construct tests: (1) identification of the goal of the test under construction; (2) identification of a target information function in accordance with the goal; and (3) selection of items based on the information function and the fulfillment of some constraints. These steps clearly mimic the formulation and resolution of any other combinatorial optimization problem.

Note that these three steps are preceded by a preliminary one in which the item pool is designed. Item-pool design requires the estimation of the information function of the items. The step is performed using item response theory (IRT), see Veldkamp (2013) and Lu (2014) for a description and a critique of IRT. The information function is then discretized into specific ability levels of interest (such as the pass/fail ability level for the test) in order to provide coefficients for the objective and/or the constraints of the combinatorial optimization model.

While IRT is the predominant method to evaluate the information provided by the items, alternatives exist, see Smits and Finkelman (2014) for a case study using ordinal regression on the total expected test score.

The present paper only considers the item selection process. Even when we limit our attention to the selection step, the literature is extensive. We can classify previous work into three groups: (1) single test construction before test use (offline single test); (2) single test construction using the information provided by answered questions (online single test); and (3) joint test construction in which the tests are constructed in order to evaluate examinees with different, but equivalent, test forms (multiple test construction).

Note that any of these methods can be used to solve other problems after some modifications. Therefore, we offer a review of recent procedures for each of these three methods.

Offline single test construction is known in the literature as static test generation (Nguyen & Fong, 2013), or automated test assembly (Veldkamp, 2013). van der Linder (2005) provides a general integer programming formulation for the problem. The proposed objective is to maximize the minimum amount of discrimination index on any ability level of interest, while satisfying some constraints on the different requirements of the test. Veldkamp (2013) considers the model in van der Linder (2005) under robustness considerations on the coefficients provided during item-pool design. The effect of using a robust formulation is then verified by comparing the solutions provided by both models.

Nguyen and Fong (2013) identifies the problem as a multidimensional knapsack problem (Kellerer et al., 2004), in which the test tries to maximize the discrimination degree (i.e., how good the test is at recognizing user proficiency) under the following constraints: number of questions, time to perform the test, average difficulty and number of questions per topic. The model is then solved using a branch-and-cut based procedure and the solutions are compared to previous methods found in the literature on instances with up to 50 000 items.

Online single test construction corresponds to the construction of the test along with its resolution, and it is usually associated to CAT (computerized adaptive test). Due to the variable nature of the problem, the methods are usually constructive and item selection makes use of greedy rules. Furthermore, CATs try to take into account that multiple tests are to be generated and thus, some randomization is introduced to avoid excessive use of some items.

In He, Diao, and Hauser (2014), four different item-selection methods are evaluated. These methods take into account multiple particularities, such as side constraints, content specifications, time requirements, or item formats among others. All of the methods offer similar results in terms of their use of items and their accuracy to measure the information function. Among them, we highlight the shadow test approach, see also van der Linden and Veldkamp (2004). This method constructs a complete test in each selection step and then randomly selects the item from the constructed test.

Edmonds and Armstrong (2009) considers a hybrid between offline and online single test construction, denoted as multiple stage adaptive test (MST) design. The hybrid divides items into groups, each belonging to a different stage of the test. When an item is to be selected, the method randomly chooses the item from the corresponding group by taking into account the current level of the examinee. The definition of each group is modeled, and then solved using a commercial solver (CPLEX) for a 1336-item pool.

Multiple test construction, or parallel test design, is used when interchangeable tests need to be constructed. Examples are the evaluation of candidates at different time frames, or the assignment of different tests to students in order to avoid cheating.