Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 175 (2006) 649–671
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Discrete Optimization

Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering

Gareth R. Beddoe *, Sanja Petrovic

Automated Scheduling Optimisation and Planning Research Group, Department of Computer Science, University of Nottingham, Nottingham NG8 1BB, United Kingdom

Received 21 May 2004; accepted 1 December 2004
Available online 16 August 2005

Abstract

Personnel rostering problems are highly constrained resource allocation problems. Human rostering experts have many years of experience in making rostering decisions which reflect their individual goals and objectives. We present a novel method for capturing nurse rostering decisions and adapting them to solve new problems using the Case-Based Reasoning (CBR) paradigm. This method stores examples of previously encountered constraint violations and the operations that were used to repair them. The violations are represented as vectors of feature values. We investigate the problem of selecting and weighting features so as to improve the performance of the case-based reasoning approach. A genetic algorithm is developed for off-line feature selection and weighting using the complex data types needed to represent real-world nurse rostering problems. This approach significantly improves the accuracy of the CBR method and reduces the number of features that need to be stored for each problem. The relative importance of different features is also determined, providing an insight into the nature of expert decision making in personnel rostering.

Keywords: Scheduling; Knowledge based systems; Case-based reasoning; Meta-heuristics; Feature selection and weighting

1. Introduction

Nurse rostering can be defined to be the problem of placing resources (nurses), subject to constraints, into slots in a pattern, where the pattern denotes a set of legal shifts defined in terms of work that needs to be done [30]. A wide variety of constraints can be imposed on rosters depending on the legal,

* Corresponding author. Address: Medicsight PLC, 46 Berkeley Square, London W1J 5AT, United Kingdom. Tel.: +44 796 115 7976.
E-mail addresses: grb@cs.nott.ac.uk, gareth.beddoe@medicsight.com (G.R. Beddoe), sxp@cs.nott.ac.uk (S. Petrovic).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2004.12.028

management, and staffing requirements of individual organisations. Definitions of roster quality and optimality are highly subjective and therefore difficult to represent systematically using utility functions or rule bases. Human rostering experts have many years of experience in making rostering decisions which reflect their individual goals and objectives.

Nurse rostering problems have been solved using a variety of different mathematical and artificial intelligence methods. They are usually modelled as optimisation problems but the objective functions used vary considerably between problems. Bailey [3], Beaumont [6], and Warner [28] use mathematical programming techniques to generate nurse rosters optimised with respect to staffing costs, under-staffing costs, and shift pattern penalties. Constraint satisfaction techniques have been developed by Abdennadher and Schlenker [1], Cheng et al. [11], and Meyer auf'm Hofe [20] which allow the definition of many different types of constraint. A number of meta-heuristic approaches have been explored including genetic algorithms [13], simulated annealing [4], tabu search [8,12], and hyper-heuristics [10]. A CBR approach by Scott and Simpson [26] combined case-based reasoning with constraint logic programming by storing shift patterns used for the construction of nurse rosters.

Case-based repair generation (CBRG) is a technique developed by the authors to solve nurse rostering problems [7] which uses *case-based reasoning* (CBR). CBR is a reasoning paradigm in which new problems are solved using the solutions to similar problems that have previously been encountered [18]. Previous problems and their corresponding solutions are stored as *cases* in a database called a *case-base*. New problems are compared to the cases in the case-base and the most *similar* is retrieved. The solution to the problem from the retrieved case is then adapted to the context of the new problem. If the new solution could be useful for future problem solving then it is stored in the case-base, thus increasing the total knowledge held.

The CBRG method considers each constraint violation in a roster as a separate problem. The case-base contains a history of previous constraint violations and the operations that were used to repair them. Cases are retrieved from the case-base using a two stage retrieval process [23]. The first stage retrieves those cases containing violations of the same type as the current problem. The second stage calculates the similarity of these cases to the current problem using the *weighted nearest neighbour* method. The violations are represented by a set of characteristic *features* and can be interpreted as points in a *feature space*. Weights are assigned to the features representing their relative importance. The most similar case is then defined as the one with the smallest weighted distance from the feature vector representing the current problem. It is vital for the retrieval process that appropriate features are selected to represent the violations and that these features are carefully weighted.

One of the most common ways to determine the accuracy of a case-base is to measure its *classification accuracy*. The CBRG method can be seen as a classifier which determines the type and parameters of a repair for a given violation. Its classification accuracy can be measured by repeatedly removing a case from the case-base, performing a retrieval to determine the nearest case to the removed case, and then comparing the repairs. In the literature, nearest neighbour classification algorithms [14] have been used successfully to solve a number of different classification problems. They allow complex relationships between input parameters to be captured without the need to model them explicitly. However, they can be sensitive to noise in the data sets and erroneous or irrelevant features [2]. These effects can be reduced by selecting only relevant features from the feature set and assigning a weight to each feature representing its relative importance. A number of different feature weighting and selection methods have been developed including Salzberg's [25] feature weighting algorithm based on a heuristic approach for his EACH classification method, a random mutation hill climbing approach for feature selection by Skalak [27], and a genetic algorithm by Kuncheva and Jain [19]. Many more algorithms are described in a review by Wettschereck et al. [29]. We investigate an approach to automated weighting and feature selection based on the genetic algorithm based GA-WKNN developed by Kelly and Davis [17] and a dimensionality reduction algorithm developed by Raymer et al. [24]. These approaches are adapted so that they can handle the types of data used in the CBRG method to model the nurse rostering problem.