



A fuzzy genetic programming-based algorithm for subgroup discovery and the application to one problem of pathogenesis of acute sore throat conditions in humans



C.J. Carmona^{a,*}, V. Ruiz-Rodado^b, M.J. del Jesus^c, A. Weber^d, M. Grootveld^b, P. González^c, D. Elizondo^e

^a Department of Civil Engineering, Languages and Systems Area, University of Burgos, 09001 Burgos, Spain

^b Leicester School of Pharmacy, De Montfort University, LE1 9BH Leicester, United Kingdom

^c Department of Computer Science, University of Jaén, 23071 Jaén, Spain

^d Institute for Materials Research and Innovation, University of Bolton, BL3 5AB Bolton, United Kingdom

^e School of Computer Science and Informatics, De Montfort University, LE1 9BH Leicester, United Kingdom

ARTICLE INFO

Article history:

Received 24 May 2014

Received in revised form 19 November 2014

Accepted 21 November 2014

Available online 5 December 2014

Keywords:

Genetic programming
Subgroup discovery
Evolutionary fuzzy system
Bioinformatics

ABSTRACT

This paper proposes a novel algorithm for subgroup discovery task based on genetic programming and fuzzy logic called Fuzzy Genetic Programming-based for Subgroup Discovery (FuGePSD). The genetic programming allows to learn compact expressions with the main objective to obtain rules for describing simple, interesting and interpretable subgroups. This algorithm incorporates specific operators in the search process to promote the diversity between the individuals. The evolutionary scheme of FuGePSD is codified through the genetic cooperative-competitive approach promoting the competition and cooperation between the individuals of the population in order to find out the optimal solutions for the SD task.

FuGePSD displays its potential with high-quality results in a wide experimental study performed with respect to others evolutionary algorithms for subgroup discovery. Moreover, the quality of this proposal is applied to a case study related to acute sore throat problems.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Subgroup discovery (SD) is a descriptive data mining technique for describing unusual features with monitored properties of interest [40,66]. This task contributes interesting knowledge to the scientific community from two view-points, specifically both features those including the provision of interest and precision. SD has been included within the concept of Supervised Descriptive Rule Discovery [42], together with further descriptive techniques such as emerging patterns [18] and contrast set mining [5].

Differing SD algorithms have been implemented throughout the literature in order to solve SD tasks based on beam search such as CN2-SD [45] or SD [27], exhaustive such as SD-Map [3], or genetic algorithms such as SDIGA [15] and NMEEF-SD [8], among others.

* Corresponding author. Fax: +34 259478.

E-mail address: cjcarmona@ubu.es (C.J. Carmona).

Genetic programming [41] is a methodology based on evolutionary algorithms (EAs) and it has been used for classification purposes [20], rule learning [43,44] and genetic-based machine learning [21]. Amongst its advantages can be highlighted the following:

- flexibility in the learning process due to the use of populations with dynamic size and individuals with structure and size variable. This property facilitates the obtaining of descriptive rules for the search space,
- simplicity, since it allows to learn rules in a flexible way without the necessity to include all variables in the individuals,
- diversity amongst the rules, which is acquired through specific operators promoting the diversity at phenotype or genotype level.

This paper presents a new approach named Fuzzy Genetic Programming-based for Subgroup Discovery: FuGePSD. This algorithm represents an evolutionary fuzzy system (EFS) [34] based on genetic programming [41] which employs a tree structure with a variable-length to represent the individuals of the population. FuGePSD employs several genetic operators in order to obtain rules to which are as general and precise as possible describing new information of the search space. In this way, FuGePSD includes an operator to promote the diversity at genotype level where rules describing the same examples are penalised. Moreover, drop and the insertion of genetic operators enhances the increase in precision and generality of the rules.

Benefits offered by the FuGePSD technique are delivered in a complete experimental study supported by appropriate statistical tests. The study is focused on datasets with continuous variables and the validity of FuGePSD is analysed with respect to alternative EAs for SD. Statistical tests confirm the highly effective performance and suitability for this new approach. Moreover, the behaviour of FuGePSD in real problems is applied to a study related to sore throat. This problem is an acute upper respiratory tract infection that impinges on the throat's respiratory mucosa, and can be linked with fever, headache and general malaise. The dataset analysed distinguishes for the high dimensionality with a wide number of features. Results acquired show the quality of the new proposal presented in this paper which are highlighted by experts in this field.

The paper is organised as follows. Firstly, preliminary concepts are described in Section 2. Next, Section 3 presents the new approach in which a description of the algorithm, operation scheme, fitness functions and genetic operators required in order to facilitate its analysis can be observed. Sections 4 and 5 present all information related to the experimental framework and the study, respectively. In Section 6, a case study is presented, and results arising there from are discussed by researchers with expertise in this field. Finally, the major salient conclusions are outlined.

2. Preliminaries

This section introduces the main concepts used for the algorithm presented. Firstly, a brief introduction to EFSs, and a short review of the SD proposals based on EAs in the specialised literature are presented in Section 2.1. Secondly, the definition, main properties and elements of the SD technique are outlined in Section 2.2. Thirdly, major properties and quality measures for fuzzy rules in SD are summarised in 2.3. Finally, the use of EFSs in SD throughout the literature is presented in Section 2.4.

2.1. Evolutionary fuzzy systems

An EFS [34] can be described as a fuzzy system [68] augmented with a learning process based on evolutionary computation [19], such as those involving genetic algorithms [30,36], genetic programming [41], evolutionary programming [23] or evolution strategies [55], amongst others.

Fuzzy systems are usually considered in the form of fuzzy-rule based systems (FRBSs), which are composed of “IF–THEN” rules where both the antecedent and consequent can contain fuzzy logic statements. This simple and interpretable representation facilitates their application in a wide range of real-world problems such as the pioneering problems in control [52], modelling [53], classification [39]. On the other hand, the EAs are well known and widely used global search technique with the ability to explore a large search space such as regression [26], association rule mining [50] or instance selection [17], for example. Therefore, EAs can be used in the development of FRBSs offering much potential as a search tool, allowing the inclusion of domain knowledge and the obtaining of better rules.

Different schemes of representation for the EAs are considered within EFS:

1. “Chromosome = rule” approach, in which each individual codifies a single rule, and the whole rule set is provided by combining several individuals within the population. Three categories are considered: the Michigan approach usually known as the as learning classifier system [37], the Iterative Rule-Learning approach [60] and the genetic cooperative-competitive learning approach [31].
2. “Chromosome = set of rules” approach, also known as the Pittsburgh approach, in which each individual represents a set of rules [59]. In this case, a chromosome evolves a complete set of rules that compete amongst them along the evolutionary process.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات