Chapter 5
Ant Programming Algorithms for Classification

Juan Luis Olmo
University of Córdoba, Spain

José Raúl Romero
University of Córdoba, Spain

Sebastián Ventura
University of Córdoba, Spain

ABSTRACT

Ant programming is a kind of automatic programming that generates computer programs by using the ant colony metaheuristic as the search technique. It has demonstrated good generalization ability for the extraction of comprehensible classifiers. To date, three ant programming algorithms for classification rule mining have been proposed in the literature: two of them are devoted to regular classification, differing mainly in the optimization approach, single-objective or multi-objective, while the third one is focused on imbalanced domains. This chapter collects these algorithms, presenting different experimental studies that confirm the aptitude of this metaheuristic to address this data-mining task.

INTRODUCTION

Data mining (DM) tasks and some parts of the knowledge discovery process can be addressed as optimization and search problems, in account of their difficulty to be modeled and the high size of the space of solutions. To this end, biologically-inspired techniques appear as a good technique, since they are techniques tolerant to certain imprecision and uncertainty that are able to model in an approximate way natural phenomena.

The DM classification task focuses on predicting the value of the class given the values of certain other attributes (referred to as the predicting attributes). A model or classifier is inferred in a training stage by analyzing the values of the

DOI: 10.4018/978-1-4666-6078-6.ch005

predicting attributes that describe each instance, as well as the class to which each instance belongs. Thus, classification is considered to be supervised learning, in contrast to unsupervised learning, where instances are unlabeled. Once the classifier is built, it can be used later to classify other new and uncategorized instances into one of the existing classes.

Genetic programming (GP) (Koza, 1992) was the first biologically-inspired automatic programming technique used for addressing the classification task of DM. A survey focused on the application of GP to classification can be found in (Espejo, Ventura, & Herrera, 2010). Another automatic programming technique, less widespread but more recent than GP, is ant programming (AP) (Roux & Fonlupt, 2000) which uses ant colony optimization as the search technique to look for computer programs. Actually, individuals in AP are known as artificial ants and they encode a solution that is represented by a path over a graph or a tree. Recent research has put the spotlight on the application of AP to DM, specifically to classification (Olmo, Romero, & Ventura, 2011) and association rule mining (Olmo, Luna, Romero, & Ventura, 2013), demonstrating the suitability of this metaheuristic to find good and comprehensible solutions to these tasks.

In this chapter we present the AP algorithms for inducing rule-based classifiers that have been presented in literature. Two of them are devoted to regular classification, and they mainly differ in the optimization approach, while the third proposal is specific for imbalanced classification (Olmo et al., 2012). All the algorithms can cope both with binary and multiclass data sets.

The first section of this chapter presents the original single-objective AP algorithm for classification, called GBAP (Grammar-Based Ant Programming). The second section describes the multi-objective AP proposal, called MOGBAP (Multi-Objective GBAP). The third section explains the main workings of the imbalanced APIC (Ant Programming for Imbalanced Classification) algorithm. The fourth section presents the experimental studies carried out to show the performance of these algorithms. Finally, the last section gives some concluding remarks and ideas for future work.

THE GBAP ALGORITHM: GRAMMAR-BASED ANT PROGRAMMING

This section introduces the first AP algorithm for classification, called GBAP, which is based on the use of a context-free grammar (CFG) for ensuring the generation of individuals syntactically valid, as well as the other AP algorithms presented in this work. The algorithm evolves a population of rules from the training set that are combined at the end of the last generation into a decision-list like classifier. Then, the model induced is test over the test set and the results obtained are reported. The flowchart of GBAP is shown in Figure 1, and its characteristics are described in the following subsections.

Environment and Rule Encoding

The AP models presented here are founded on the use of a context-free grammar (CFG) that defines all the possible states that individuals can visit. Actually, the environment that permits ants to communicate indirectly with each other is the derivation tree that can be generated from the grammar, as shown in Figure 1. This grammar is expressed in Backus-Naur form, and its definition is given by $G = (\Sigma_N, \Sigma_T, P, S)$:

$G = (\Sigma_N, \Sigma_T, P, S)$

Σ_N = {<Rule>, <Antecedent>, <Consequent>, <Condition>}

Σ_T = {-->, AND, =, !=, $attr_1$, $attr_2$, ..., $attr_n$, $value_{1,1}$, $value_{1,2}$, ..., $value_{1,m}$, $value_{2,1}$, $value_{2,2}$, ..., $value_{2,m}$, ..., $value_{n,1}$, $value_{n,2}$, ..., $value_{n,m}$}

Related Content

Improving Rumor Detection by Image Captioning and Multi-Cell Bi-RNN With Self-Attention in Social Networks
Jenq-Haur Wang, Chin-Wei Huangand Mehdi Norouzi (2022). *International Journal of Data Warehousing and Mining (pp. 1-17).*
www.irma-international.org/article/improving-rumor-detection-by-image-captioning-and-multi-cell-bi-rnn-with-self-attention-in-social-networks/313189

Dynamic View Selection for OLAP
Michael Lawrenceand Andrew Rau-Chaplin (2008). *International Journal of Data Warehousing and Mining (pp. 47-61).*
www.irma-international.org/article/dynamic-view-selection-olap/1799

Data Mining Association Rules for Making Knowledgeable Decisions
A.V. Senthil Kumarand R. S.D. Wahidabanu (2009). *Data Mining Applications for Empowering Knowledge Societies (pp. 43-53).*
www.irma-international.org/chapter/data-mining-association-rules-making/7545

DMA: Matrix Based Dynamic Itemset Mining Algorithm
Damla Oguz, Baris Yildizand Belgin Ergenc (2013). *International Journal of Data Warehousing and Mining (pp. 62-75).*
www.irma-international.org/article/dma/105120

Advances in Classification of Sequence Data
Pradeep Kumar, P. Radha Krishna, Raju S. Bapiand T. M. Padmaja (2008). *Data Mining and Knowledge Discovery Technologies (pp. 143-174).*
www.irma-international.org/chapter/advances-classification-sequence-data/7517