Multi−Instance Learning with MultiObjective Genetic Programming

Amelia Zafra
University of Cordoba, Spain

Sebastián Ventura
University of Cordoba, Spain

INTRODUCTION

The multiple-instance problem is a difficult machine learning problem that appears in cases where knowledge about training examples is incomplete. In this problem, the teacher labels examples that are sets (also called bags) of instances. The teacher does not label whether an individual instance in a bag is positive or negative. The learning algorithm needs to generate a classifier that will correctly classify unseen examples (i.e., bags of instances).

This learning framework is receiving growing attention in the machine learning community and since it was introduced by Dietterich, Lathrop, Lozano-Perez (1997), a wide range of tasks have been formulated as multi-instance problems. Among these tasks, we can cite content-based image retrieval (Chen, Bi, & Wang, 2006) and annotation (Qi and Han, 2007), text categorization (Andrews, Tsochantaridis, & Hofmann, 2002), web index page recommendation (Zhou, Jiang, & Li, 2005; Xue, Han, Jiang, & Zhou, 2007) and drug activity prediction (Dietterich et al., 1997; Zhou & Zhang, 2007).

In this chapter we introduce MOG3P-MI, a multiobjective grammar guided genetic programming algorithm to handle multi-instance problems. In this algorithm, based on SPEA2, individuals represent classification rules which make it possible to determine if a bag is positive or negative. The quality of each individual is evaluated according to two quality indexes: sensitivity and specificity. Both these measures have been adapted to MIL circumstances. Computational experiments show that the MOG3P-MI is a robust algorithm for classification in different domains where achieves competitive results and obtain classifiers which contain simple rules which add comprehensibility and simplicity in the knowledge discovery process, being suitable method for solving MIL problems (Zafra & Ventura, 2007).

BACKGROUND

In the middle of the 1990's, Dietterich et al. (1997) described three Axis-Parallel Rectangle (abbreviated as APR) algorithms to solve the problem of classifying aromatic molecules according to whether or not they are "musky". These methods attempted to search the appropriate axis-parallel rectangles constructed by their conjunction of features. Their best performing algorithm (iterated-discrim) started with a point in the feature space and grew a box with the goal of finding the smallest box covered at least one instance from each positive bag and no instances from any negative bag. The resulting box was then expanded (via a statistical technique) to get better results.

Following Dietterich et al.'s study, a wide variety of new methods of multi-instance learning has appeared. Auer (1997) tried to avoid some potentially hard computational problems that were required by the heuristics used in the iterated-discrim algorithm and presented a theoretical algorithm, MULTINST. With a new approach, Maron and Lozano-Perez (1998) proposed one of the most famous multi-instance learning algorithms, Diverse Density (DD), where the diverse density of a point, p, in the feature space was defined as a probabilistic measure which considered how many different positive bags had an instance near p, and how far the negative instances were from p. This algorithm was combined with the Expectation Maximization (EM) algorithm, appearing as EM-DD (Zhang & Goldman, 2001). Another study that extended the DD algorithm to maintain multilearning regression data sets was the EM-based multi-instance regression algorithm (Amar, Dooly, Goldman, & Zhang, 2001).

In 1998, Long and Tan (1998) described a polynomial-time theoretical algorithm and showed that if instances in the bags were independently drawn from product distribution, then the APR was PAC-learnable. Following with PAC-learnable research, Kalai and Blum (1998) described a reduction from the problem of PAC-learning under the MIL framework to PAC-learning with one-sided random classification noise, and presented a theoretical algorithm with less complexity than the algorithm described in Auer (1997).

The first approaches using lazy learning, decision trees and rule learning were researched during the year 2000. In the lazy learning context, Whang and Zucker (2000) proposed two variants of the k nearest-neighbour algorithm (KNN) that they referred to as Citation-KNN and Bayesian-KNN; these algorithms extended the k-nearest neighbor algorithm for MIL adopting Hausdorff distance. With respect to decision trees and learning rules, Zucker and Chevaleyre (2000) implemented ID3-MI and RIPPER-MI, which are multi-instance versions of decision tree algorithm ID3 and rule learning algorithm RIPPER, respectively. At that time, Ruffo (2000) presented a multi-instance version of the C4.5 decision tree, which was known as RELIC. Later, Zhou et al. (2005) presented the Fretcit-KNN algorithm, a variant of Citation-KNN that modified the minimal Hausdorff distance for measuring the distance between text vectors and using multiple instance perspective. There are also many other practical multiple instance (MI) algorithms, such as the extension of standard neural networks to MIL (Zhang & Zhou, 2006). Also there are proposals about adapting Support Vector Machines to multi-instance framework (Andrews et al., 2002; Qi and Han, 2007) and the use of ensembles to learn multiple instance concepts, (Zhou & Zhang, 2007).

We can see that a variety of algorithms have been introduced to learn in multi-instance settings. Many of them are based on well-known supervised learning algorithms following works such as Ray and Craven's (2005) who empirically studied the relationship between supervised and multiple instance learning, or Zhou (2006) who showed that multi-instance learners can be derived from supervised learners by shifting their focuses from the discrimination on the instances to the discrimination on the bags. Although almost all popular machine learning algorithms have been applied to solve multiple instance problems, it is remarkable that the first proposals to adapt Evolutionary Algorithm to this scenario have not appeared until 2007 (Zafra, Ventura, Herrera-Viedma, & Romero 2007; Zafra & Ventura, 2007) even though these algorithms have been applied successfully in many problems in supervised learning.

MAIN FOCUS

Genetic Programming is becoming a paradigm of growing interest both for obtaining classification rules (Lensberg, Eilifsen, & McKee, 2006), and for other tasks related to prediction, such as characteristic selection (Davis, Charlton, Oehlschalager, & Wilson, 2006) and the generation of discriminant functions. The major considerations when applying GP to classification tasks are that a priori knowledge is not needed about the statistical distribution of the data (data distribution free). It can operate directly on the data in their original form, can detect unknown relationships that exist among data, expressing them as a mathematical expression and can discover the most important discriminating features of a class. We can find different proposals that use the GP paradigm to evolve rule sets for different classification problems, both two-class ones and multiple-class ones. Results show that GP is a mature field that can efficiently achieve low error rates in supervised learning, hence making it feasible to adapt to multiple instance learning to check its performance.

We propose, MOG3P-MI, a multiobjective grammar guided genetic programming algorithm. Our main motivations to introduce genetic programming into this field are: (a) grammar guided genetic programming (G3P) is considered a robust tool for classification in noisy and complex domains where it achieves to extract valuable information from data sets and obtain classifiers which contain simple rules which add comprehensibility and simplicity in the knowledge discovery process and (b) genetic programming with multiobjective strategy allows us to obtain a set of optimal solutions that represent a trade-off between different rule quality measurements, where no one can be considered to be better than any other with respect to all objective functions. Then, we could introduce preference information to select the solution which offers the best classification guarantee with respect to new data sets.

In this section we specify different aspects which have been taken into account in the design of the MOG3P-MI algorithm, such as individual representa-

Related Content

Supporting Imprecision in Database Systems
Ullas Nambiar (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1884-1887).*
www.irma-international.org/chapter/supporting-imprecision-database-systems/11076

Using Dempster-Shafer Theory in Data Mining
Malcolm J. Beynon (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2011-2018).*
www.irma-international.org/chapter/using-dempster-shafer-theory-data/11095

Outlier Detection
Sharanjit Kaur (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1476-1482).*
www.irma-international.org/chapter/outlier-detection/11015

Physical Data Warehousing Design
Ladjel Bellatrecheand Mukesh Mohania (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1546-1551).*
www.irma-international.org/chapter/physical-data-warehousing-design/11025

Minimum Description Length Adaptive Bayesian Mining
Diego Liberati (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1231-1235).*
www.irma-international.org/chapter/minimum-description-length-adaptive-bayesian/10979