UNIVERSIDAD DE CÓRDOBA

Escuela Politécnica Superior

Departamento de Informática y Análisis Numérico

Nuevos Modelos de Aprendizaje Híbrido para Clasificación y Ordenamiento Multi-Etiqueta

Memoria de Tesis presentada por

Oscar Gabriel Reyes Pupo

como requisito para optar al grado

de Doctor en Informática

Director

Dr. Sebastián Ventura Soto

Córdoba

Octubre de 2016

TITULO: *NUEVOS MODELOS DE APRENDIZAJE HÍBRIDO PARA CLASIFICACIÓN Y ORDENAMIENTO MULTI-ETIQUETA*

AUTOR: *Oscar Gabriel Reyes Pupo*

© Edita: UCOPress. 2016
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

www.uco.es/publicaciones
publicaciones@uco.es

UNIVERSITY OF CÓRDOBA

Polytechnic Superior Institute

Department of Computer Science and Numerical Analysis

New Hybrid Learning Models for Multi-label Classification and Label Ranking

A Thesis presented by

Oscar Gabriel Reyes Pupo

as a requirement to aim for the degree of

Ph.D. in Computer Science

Advisor

Dr. Sebastián Ventura Soto

Córdoba October, 2016

idep
Instituto de Estudios
de Postgrado

TÍTULO DE LA TESIS: Nuevos Modelos de Aprendizaje Híbrido para Clasificación y Ordenamiento Multi-Etiqueta.

DOCTORANDO/A: Oscar Gabriel Reyes Pupo

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

En su tesis, D. Oscar Gabriel Reyes Pupo ha abordado tres temas en el contexto del aprendizaje multi-etiqueta: la estimación de atributos, el aprendizaje basado en instancias y el aprendizaje activo.

En el primer tema, se propusieron un total de cinco métodos de estimación de atributos, dos de ellos basados en la aplicación de algoritmos evolutivos. En el segundo tema, se diseñó un nuevo algoritmo de vecindad inspirado en los principios de clasificación basada en gravitación de datos. En el tercer tema, se desarrollaron dos estrategias de aprendizaje activo, y se construyó una librería de clases que favorece la implementación de métodos de aprendizaje activo y la experimentación en esta área de estudio. Además, se propusieron dos aproximaciones que permiten evaluar de una manera más adecuada el rendimiento de las técnicas de aprendizaje activo.

A partir de los resultados alcanzados en esta tesis, se lograron varias publicaciones en revistas internacionales de impacto y conferencias internacionales, lo que muestra la calidad científica del trabajo realizado. Por otra parte, las líneas de investigación desarrolladas en esta memoria no están aún agotadas, existiendo algunas líneas de trabajo futuro que considero pueden también dar lugar a varias publicaciones científicas de calidad.

En conclusión, considero que la memoria presentada por D. Oscar Gabriel Reyes Pupo reúne, en mi opinión, las condiciones necesarias para su defensa.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 18 de octubre de 2016

Firma del director

Fdo.:____________________

La memoria de Tesis Doctoral titulada “*Nuevos Modelos de Aprendizaje Híbrido para Clasificación y Ordenamiento Multi-Etiqueta*”, que presenta Oscar Gabriel Reyes Pupo para optar al grado de Doctor, ha sido realizada dentro del Programa Oficial de Doctorado “Computación Avanzada, Energía y Plasmas” de la Universidad de Córdoba, España, bajo la dirección del Dr. Sebastián Ventura Soto, cumpliendo, en su opinión, los requisitos exigidos a este tipo de trabajos.

Córdoba, Octubre de 2016

El Doctorando

Oscar

Fdo: Oscar Gabriel Reyes Pupo

El Director

Fdo: Dr. Sebastián Ventura Soto

Tesis Doctoral parcialmente subvencionada por el Ministerio de Economía y Competitividad, proyecto **TIN2014-55252-P**.

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

UNIÓN EUROPEA

Infraestructura/equipo cofinanciada/o por el Fondo Europeo de Desarrollo Regional

“Una manera de hacer Europa”

Agradecimientos

Esta tesis de doctorado no es el resultado del esfuerzo de una sola persona, sino de un conjunto de personas que de una manera u otra han contribuido a la realización de la misma.

Ante todo quiero agradecerle a Dios por la ayuda y fuerza que me ha dado para llevar a cabo esta empresa. A mis amados padres que desde los comienzos de mis estudios me han apoyado incondicionalmente y me han exhortado a seguir superándome. A mi amada esposa por su amor, consejos, aliento y comprensión. Ciertamente, sin el apoyo y sacrificio de mi familia esta tesis no se hubiera podido realizar. Le doy gracias a mi familia por su apoyo, y a la vez le pido perdón por el tiempo robado para realizar esta tesis, tiempo robado que nunca volverá, en especial los momentos que he dejado de ver crecer a mi pequeña y amada hija.

Quiero agradecerle especialmente a mi director Sebastián Ventura, por haber confiado en aquel muchacho desconocido que le solicitó un día que le dirigiera su tesis, y por el apoyo incondicional que me ha prestado a lo largo de todo este tiempo. Le agradezco al Dr. Carlos Morell de la Universidad de Las Villas, Cuba, por sus valiosos comentarios y colaboración en los trabajos realizados.

También quiero agradecerle a mis amigos de la Universidad de Holguín, Cuba, por el tiempo y momentos que compartimos juntos. A los colegas del grupo KDIS de la Universidad de Córdoba, España, por su apoyo en mis estancias realizadas.

Gracias de todo corazón.

Resumen

En la última década, el aprendizaje multi-etiqueta se ha convertido en una importante área de investigación, debido en gran parte al creciente número de problemas reales que contienen datos multi-etiqueta. En esta tesis se estudiaron dos problemas sobre datos multi-etiqueta, la mejora del rendimiento de los algoritmos en datos multi-etiqueta complejos y la mejora del rendimiento de los algoritmos a partir de datos no etiquetados.

El primer problema fue tratado mediante métodos de estimación de atributos. Se evaluó la efectividad de los métodos de estimación de atributos propuestos en la mejora del rendimiento de los algoritmos de vecindad, mediante la parametrización de las funciones de distancias empleadas para recuperar los ejemplos más cercanos. Además, se demostró la efectividad de los métodos de estimación en la tarea de selección de atributos. Por otra parte, se desarrolló un algoritmo de vecindad inspirado en el enfoque de clasificación basada en gravitación de datos. Este algoritmo garantiza un balance adecuado entre eficiencia y efectividad en su solución ante datos multi-etiqueta complejos.

El segundo problema fue resuelto mediante técnicas de aprendizaje activo, lo cual permite reducir los costos del etiquetado de datos y del entrenamiento de un mejor modelo. Se propusieron dos estrategias de aprendizaje activo. La primer estrategia resuelve el problema de aprendizaje activo multi-etiqueta de una manera efectiva y eficiente, para ello se combinaron dos medidas que representan la utilidad de un ejemplo no etiquetado. La segunda estrategia propuesta se enfocó en la resolución del problema de aprendizaje activo multi-etiqueta en modo de lotes, para ello se formuló un problema multi-objetivo donde se optimizan tres medidas, y el problema de optimización planteado se resolvió mediante un algoritmo evolutivo.

Como resultados complementarios derivados de esta tesis, se desarrolló una herramienta computacional que favorece la implementación de métodos de aprendizaje activo y la experimentación en esta área de estudio. Además, se propusieron dos aproximaciones que permiten evaluar el rendimiento de las técnicas de aprendizaje activo de una manera más adecuada y robusta que la empleada comúnmente en la literatura.

Todos los métodos propuestos en esta tesis han sido evaluados en un marco experimental adecuado, se utilizaron numerosos conjuntos de datos y se compararon los rendimientos de los algoritmos frente a otros métodos del estado del arte. Los resultados obtenidos, los cuales fueron verificados mediante la aplicación de test estadísticos no parámetricos, demuestran la efectividad de los métodos propuestos y de esta manera comprueban las hipótesis planteadas en esta tesis.

Abstract

In the last decade, multi-label learning has become an important area of research due to the large number of real-world problems that contain multi-label data. This doctoral thesis is focused on the multi-label learning paradigm. Two problems were studied, firstly, improving the performance of the algorithms on complex multi-label data, and secondly, improving the performance through unlabeled data.

The first problem was solved by means of feature estimation methods. The effectiveness of the feature estimation methods proposed was evaluated by improving the performance of multi-label lazy algorithms. The parametrization of the distance functions with a weight vector allowed to recover examples with relevant label sets for classification. It was also demonstrated the effectiveness of the feature estimation methods in the feature selection task. On the other hand, a lazy algorithm based on a data gravitation model was proposed. This lazy algorithm has a good trade-off between effectiveness and efficiency in the resolution of the multi-label lazy learning.

The second problem was solved by means of active learning techniques. The active learning methods allowed to reduce the costs of the data labeling process and training an accurate model. Two active learning strategies were proposed. The first strategy effectively solves the multi-label active learning problem. In this strategy, two measures that represent the utility of an unlabeled example were defined and combined. On the other hand, the second active learning strategy proposed resolves the batch-mode active learning problem, where the aim is to select a batch of unlabeled examples that are informative and the information redundancy is minimal. The batch-mode active learning was formulated as a multi-objective problem, where three measures were optimized. The multi-objective problem was solved through an evolutionary algorithm.

This thesis also derived in the creation of a computational framework to develop any active learning method and to favor the experimentation process in the active learning area. On the other hand, a methodology based on non-parametric tests that allows a more adequate evaluation of active learning performance was proposed.

All methods proposed were evaluated by means of extensive and adequate experimental studies. Several multi-label datasets from different domains were used, and the methods were compared to the most significant state-of-the-art algorithms. The results were validated using non-parametric statistical tests. The evidence showed the effectiveness of the methods proposed, proving the hypotheses formulated at the beginning of this thesis.

Table of Contents

List of Acronyms

AAM Algorithm Adaptation Methods

AL Active Learning

AUC Area Under the learning Curve

BMAL Batch-Mode Active Learning

BR Binary Relevance

CC Classifier Chain

CMA-ES Covariance Matrix Adaptation Evolution Strategy

CVIRS Category Vector Inconsistency and Ranking of Scores

DGC Data Gravitation Classification

DM Data Mining

ESBMAL Evolutionary Strategy for Batch-Mode Multi-Label Active Learning

FS Feature Selection

FW Feature Weighting

GA Genetic Algorithm

JCLAL Java Class Library for Active Learning

KDD Knowledge Discovery in Databases

KNN K-Nearest Neighbours

LPS Label Power Set

LR Label Ranking

ML Machine Learning

MLAL Multi-label Active Learning

MLC Multi-label Classification

MLL Multi-label Learning

NSGA-II Non-dominated Sorting Genetic Algorithm II

PPT Pruned Problem Transformation

PTM Problem Transformation Methods

RPC Ranking by Pair-wise Comparison

SSL Semi-Supervised Learning

SVM Support Vector Machine

TP True performance of a selection strategy

Part I: Ph.D. Dissertation

1
Introduction

In the last two decades, the volume of data stored in the Internet has exponentially grown. Currently, it is common to find datasets that contain million of examples[1] and thousands (even millions) of features that describe these examples. Nowadays, the making decision process faces new challenges that arise not only of the complexity of the problem to resolve, but also of the complexity of data that must be processed. The Knowledge Discovery in Databases (KDD) is an important tool in the making decision process through large datasets.

KDD is the process of discovering useful, nontrivial, implicit, and previously unknown knowledge from a collection of data [1]. In KDD, the Data Mining (DM) is a crucial step, where the aim is to discover, through advanced data analysis tools, valid patterns and relationships in datasets [2]. DM uses data analysis tools such as statistical models, mathematical methods, and machine learning algorithms. Machine learning (ML) is a branch of artificial intelligence that focus on the construction of computer algorithms that can learn from data [3].

In the last decade, Multi-label Learning (MLL) has become a popular area of study due to the increasing number of real-world problems that contain multi-label data [4]. The multi-label problems involve examples that belong to a set of

[1] Also known as objects or instances.

labels at the same time. Particular problems involving multi-label data include text categorization [5, 6], semantic annotation of images [7–9], classification of music and videos [10, 11], classification of protein function and gene function [12, 13], chemical data analysis [14] and many more.

Generally speaking, multi-label datasets contain a large number examples and features that describe the examples, e.g. description of texts, images, proteins and genes. Datasets with a large number of examples and features affect in several ways the performance of learning algorithms. For instance, datasets with high dimensionality have a highly negative impact in the efficiency[2], efficacy[3] and effectiveness[4] of the most learning algorithms, know as "The curse of dimensionality" in the literature.

The goal of the MLL paradigm is to learn a model that correctly generalizes unseen multi-label data. On the MLL context two problems have been studied, Multi-label Classification (MLC) and Label Ranking (LR). MLC divides the set of labels into relevant and irrelevant sets, whereas the LR provides an ordering of the labels for a given query example [4, 15].

To date, several MLL algorithms have been proposed. The multi-label algorithms can be divided into two main categories [4, 15]: Problem Transformation Methods (PTM) and Algorithm Adaptation Methods (AAM). The PTM methods transform multi-label datasets into one or more single-label datasets. Then, for each transformed dataset, a single-label classifier is executed, and an aggregation strategy is performed. The Binary Relevance (BR) [15] trains a single-label classifier for each label. The Classifier Chain (CC) [16] is similar to BR, but the dependency between labels is considered. The Ranking by Pair-wise Comparison (RPC) method [17, 18] trains a single-label classifier for each pair of labels. The Label Power Set (LPS) [15] method constructs a new multi-class dataset, where each unique combination of labels is considered as a class of the new multi-class dataset. In studies [19–21], other sophisticated methods based on LPS approach were proposed.

On the other side, the AAM category comprises algorithms that are designed to directly handle multi-label data. In study [22], the Predictive Clustering Trees

[2] The efficiency refers to the amount of computational resources (space and time) used by an algorithm.

[3] The efficacy is related to the probability that has an algorithm to reach an optimal solution.

[4] The effectiveness, or exactness, represents the quality of the solutions found by the algorithms.

method, that has been applied to the MLC task, was proposed. In case [23], an adaptation of the well-known C4.5 algorithm was proposed. Several adaptations of the Artificial Neural Networks have appeared in the literature [24, 25]. In case [26], an extension of the popular AdaBoost algorithm appeared. Several lazy algorithms have been also proposed [27–31].

Despite the large number of studies that exist around the MLL context, there are some open issues to the scientific community. The challenges that arise of learning process from multi-label data inspire the development of new algorithms, mainly focusing in their efficiency, efficacy and effectiveness. Next, the different issues that were faced in the dissertation are introduced, providing their motivation and justification.

Improving performance on complex multi-label data

Generally speaking, multi-label datasets contain a large number of features that describe the examples, e.g. description of texts, images, proteins and genes [5, 6, 32, 33]. The irrelevant, interacting, redundant and noisy features have a highly negative impact in the performance of the learning algorithms. Moreover, the number of features is much bigger than the number of examples in several multi-label applications [5, 32, 33]. On the other hand, in some domains the number of possible labels scales up to hundreds (even thousands) and the distribution of examples per label can be showed in a non-uniform way [11, 12, 32, 34, 35]. Consequently, some multi-label algorithms, specifically lazy algorithms, present a poor performance with regard to time efficiency and effectiveness [36].

Preprocessing techniques have demonstrated to be an important step of KDD process [1, 2]. Feature engineering techniques such as Feature Weighting (FW) and Feature Selection (FS) can significantly improve the performance of learning algorithms [37–39]. FW task assigns a weight to each feature representing the usefulness of the feature to distinguish pattern classes [38]. A weight vector can be used to improve the performance of the lazy algorithms by means of parameterizing the distance function used to retrieve the k-nearest neighbors of a given query example [38]. Furthermore, a weight vector can be used as a ranking of features to guide the search of the best subset of features [40–42]. FS task can be seen as a specific case of the FW process, where the feature weights are binary values representing whether a feature is removed or conserved. FS tries to reduce the dimensionality, which has a positive effect on the efficiency, effectiveness and comprehensibility of machine learning [37, 43, 44].

A large number of studies related to FW and FS tasks on single-label data have been proposed. However, far less studies related to FW and FS tasks on multi-label context have appeared. In studies [45–52], several feature estimation methods were proposed, all of them focused on the FS task. Generally speaking, the feature estimation process in multi-label data is carried out by means of a PTM. However, these approaches have several limitations. First, the performance of a PTM generally depends on the number of labels of the dataset. Consequently, they are very expensive for domains that contain a moderate number of labels. Second, a drawback of some PTM is that they do not consider label correlations. As a result of the above situations, nowadays the designing process of FW and FS methods faces several challenges and it is an open field of research.

In general, the lazy algorithms do not construct a model from the training set, postponing almost all the process until classification. In this family of algorithms, the K-Nearest Neighbors (KNN) [53] algorithms are the simplest and easiest to understand. The KNN algorithms have shown be useful in several domains [54]. However, the main drawback of these algorithms is that they severely deteriorate in data with high dimensionality, imbalanced data, or when the classes are non-separable or they overlap. In case [36], an extensive experimental study was carried out, where the most significant multi-label algorithms were compared. The results showed that the multi-label lazy algorithms obtained the worst performance for almost all the evaluation metrics considered.

In studies [27–31, 55–57], the most significant lazy approaches to multi-label data have appeared. These previous works are important contributions to MLL. However, it is still necessary the development of lazy methods that do not deteriorate their performance on multi-label data with a large number of features, labels, imbalanced data, etc. The multi-label lazy algorithms consider any feature equally important for classifying a query; yet irrelevant, interacting, redundant and noisy features have a highly negative impact in the effectiveness of these algorithms. In this sense, FW methods can help to improve the performance of lazy algorithms by an adequate fitting of the feature weights. Examples with relevant set of labels for the classification of a query example can be retrieved by parameterizing the distance function with a weight vector, leading to a superior performance of the lazy algorithms.

On the other hand, there are other interesting lazy approaches that have been successfully applied on single-label data, and these approaches can be easily adapted to multi-label context. For instance, the Data Gravitation Classification (DGC) approach may be effective in the resolution of multi-label problems. DGC is an approach that applies the principles of the universal law of gravitation to resolve ML problems [58]. One advantage of DGC, compared to other techniques, is that it is based on simple principles with high performance levels [59].

Improving performance through unlabeled data

Generally speaking, the majority of the multi-label problems originate from domains where a huge amount of data is commonly available [6, 7, 20, 60–63]. Data labeling is a very expensive process that requires expert handling. In multi-label settings, experts must label each example several times, as each example belongs to various categories. The situation is further complicated when a multi-label dataset with a large number of examples and label classes is analyzed. Consequently, several real scenarios nowadays contain a small number of labeled data and a large number of unlabeled data simultaneously.

The challenges that arise from problems that contain labeled and unlabeled data at the same time motivate the creation of new computational methods capable of using all these information. Most multi-label algorithms that have been proposed in the literature are designed for working on supervised learning environments, i.e. scenarios where all training examples are labeled. Therefore, the multi-label algorithms can have a poor performance in these scenarios which have a small number of labeled examples.

To date, there are two main areas that are concerned with learning models from labeled and unlabeled data, known as Semi-Supervised Learning (SSL) [64] and Active Learning (AL) [65]. SSL and AL attack the same problem, but from different directions. SSL tries to exploit the latent structure of unlabeled data with the goal of improving label predictions. On the other side, AL is concerned with learning better classifiers by choosing which instances are labeled for training, reducing the costs of data labeling and training an accurate model. AL methods are involved in the acquisition of their own training data. A selection strategy iteratively selects examples from the unlabeled set that seem to be the most informative. Afterwards, an oracle (e.g. a human annotator) annotates the selected examples and they are inserted in the set of labeled data [65].

After more than a decade, an important number of AL methods for single-label data have been proposed (for an interesting survey, see [65]). However, compared to single-label AL, the AL problem within a multi-label context is far less studied. The main challenge in performing AL on multi-label data (MLAL, Multi-label Active Learning) is designing effective strategies that measure the unified informative potential of unlabeled examples across all labels. The most relevant works related to MLAL have appeared in studies [66–70, 70–84].

Most state-of-the-art MLAL strategies employ the Binary Relevance (BR) approach [15] to break down a multi-label problem into several binary classification problems. Consequently, some of these methods are computationally expensive. MLAL strategies are generally tested on the MLC task. However, their performances with regard to the LR task have not been considered. On the other hand, several MLAL strategies have been designed to work with BR-SVM (Binary Relevance with binary Support Vector Machines) as a base classifier. Therefore, the adaptation of these MLAL strategies for working with other type of base classifiers is a hard task to accomplish. In this sense, it would be interesting the development of new MLAL strategies not restricted to a type of base classifier, to directly estimate the utility of the unlabeled examples without using a PTM.

Most state-of-the-art MLAL strategies were designed to select one unlabeled instance at a time. However, in several domains, such as in the speeding up of the process of inducting classifiers with slow training procedures or in systems where a parallel annotation environment is available, the selection of a batch of unlabeled examples is preferred. Batch-mode AL (BMAL) selects a batch of k unlabeled examples in each iteration, in such a way that the selected instances are informative and the overlapping of information between them is minimal [85]. The most significant works related to performing BMAL on multi-label data appeared in [74, 81]. However, it is considered that this research line has not been studied in depth.

2
Objectives

Due to the complexity and importance of the multi-label learning, in this thesis we formulated the following **scientific problem**: How to increase the possibilities to resolve the multi-label classification and label ranking tasks, in order to obtain significantly better solutions than state-of-the-art multi-label algorithms?

The **general objective** of this thesis was to develop new algorithms with a high performance in the resolution of the multi-label classification and label ranking tasks.

The following **specific objectives** were pursued to successfully accomplish this aim:

- $\mathbf{O}_1$: Develop new feature estimation methods that allow to improve the performance of multi-label algorithms.

- $\mathbf{O}_2$: Design a new multi-label lazy algorithm with a good trade-off between efficiency and effectiveness in its solution to learn from complex multi-label data.

- $\mathbf{O}_3$: Develop new MLAL strategies not restricted to a type of base classifier and they directly handle the multi-label data.

After an extensive bibliographic review, the following **hypotheses** were formulated:

- **H**$_1$: If a feature estimation method developed a similarity function more effective in the determination of nearest examples associated to relevant label sets for classification, a significant improvement on the performance of the multi-label lazy algorithms would be achieved.
- **H**$_2$: If a feature estimation method guided the search of relevant subsets of features, the performance of multi-label algorithms would improve.
- **H**$_3$: If a multi-label lazy algorithm based on a data gravitation model was proposed, it would be competitive with the state-of-the-art multi-label lazy algorithms, and it would also provide a good trade-off between efficiency and effectiveness in its solution.
- **H**$_4$: If a MLAL strategy measured the uncertainty on the predictions of the base classifier and the inconsistency of the predicted label sets, it would obtain better solutions than state-of-the-art MLAL strategies.
- **H**$_5$: If the BMAL problem on multi-label data was formulated as a multi-objective problem, and it was resolved by means of an evolutionary algorithm, a significant improvement in the solution of the multi-label BMAL problem would be achieved.

In order to achieve the specific objectives and to test the hypotheses formulated, the following **research tasks** were accomplished:

- Analyze the basis of MLL and review the state-of-the-art multi-label algorithms, identifying open problems in MLC and LR.
- Design and implement new feature estimation methods on multi-label data.
- Validate the effectiveness of the feature estimation methods proposed in the improvement of the performance of multi-label lazy algorithms.
- Validate the effectiveness of the feature estimation methods proposed in the multi-label FS task.

- Design and implement a multi-label lazy algorithm based on DGC principles.
- Validate the effectiveness of the multi-label lazy algorithm proposed by means of comparing with the most relevant state-of-the-art lazy algorithms.
- Design and implement a MLAL strategy not restricted to a type of base classifier and that directly handles the multi-label data.
- Design and implement an evolutionary strategy to resolve the multi-label BMAL problem.
- Validate the effectiveness of the MLAL strategies proposed by means of comparing with the most relevant state-of-the-art MLAL strategies.

In the execution of these tasks, the following **scientific methods** were used:

- General methods: the hypothetico-deductive method was used to elaborate the hypotheses and to propose research lines from partial results. The systematic method for the development of computational tools. The bibliographic revision method for the analysis of previous works.
- Logic methods: the method of analysis and synthesis to decompose the information in logical and related parts, simplifying the information to process. The modeling method in the designing of algorithms and computational tools.
- Empirical methods: the experimentation to assess the methods proposed.
- Mathematical methods: statistical tests to validate the quality of the results. The statistical comparisons between algorithms were carried out by means of non-parametric statistical tests as proposed in [86–88].

3
Methodology

This chapter summarizes the methods, tools and dataset used for the development and evaluation of the algorithms proposed in this thesis. Detailed information about the methodology employed in each of the experimental studies is provided in their respective article's documentation.

Multi-label datasets

The multi-label datasets used in all experiments were obtained from the repository of real-world multi-label problems of MULAN library[1] [89]. Multi-label datasets with different scale and from different application domains were included to analyze the behavior of the methods proposed in this thesis.

Table 3.1 shows some statistics of the multi-label datasets. The values of the properties of the Corel16k dataset were averaged over all ten samples used. The label cardinality is the average number of labels per example. The label density is the label cardinality divided by the total number of labels. The label cardinality, label density and different subsets of labels are measures that represent the complexity of a multi-label dataset. The datasets vary in size: from 194 up to 43,907 examples

[1] http://mulan.sourceforge.net/datasets-mlc.html

(n), from 19 up to 52,350 features (d), from 6 up to 374 labels (q), from 15 up to 6555 different subset of labels (d_s), from 1.014 up to 26.044 label cardinality (l_c), and from 0.009 up to 0.485 label density (l_d).

Dataset	**Domain**	**Source**	n	d	q	d_s	l_c	l_d
Arts	Text	[32]	7484	23146	26	599	1,654	0,064
Bibtex	Text	[6]	7395	1836	159	2856	2,402	0,015
Birds	Audio	[90]	645	260	19	133	1,014	0,053
Business	Text	[32]	11214	21924	30	233	1,599	0,053
Cal500	Music	[11]	502	68	174	502	26,044	0,150
Computers	Text	[32]	12444	34096	33	428	1,507	0,046
Corel16k (10 samples)	Image	[7]	13811	500	161	4937	2,867	0,018
Corel5k	Image	[91]	5000	499	374	3175	3,522	0,009
Education	Text	[32]	12030	27534	33	511	1,463	0,044
Emotions	Music	[92]	593	72	6	27	1,869	0,311
Enron	Text	[93]	1702	1001	53	753	3,378	0,064
Entertainment	Text	[32]	12730	32001	21	337	1,414	0,067
Flags	Image	[9]	194	19	7	54	3,392	0,485
Genbase	Biology	[33]	662	1186	27	32	1,252	0,046
Health	Text	[32]	9250	30605	32	335	1,644	0,051
Mediamill	Video	[61]	43907	120	101	6555	4,376	0,043
Medical	Text	[5]	978	1449	45	94	1,245	0,028
Recreation	Text	[32]	12828	30324	22	530	1,429	0,065
Reference	Text	[32]	8027	39679	33	275	1,174	0,035
Scene	Image	[10]	2407	294	6	15	1,074	0,179
Science	Text	[32]	6428	37187	40	457	1,450	0,036
Social	Text	[32]	12111	52350	39	361	1,279	0,033
Society	Text	[32]	14512	31802	27	1054	1,670	0,062
TMC2007-500	Text	[60]	28596	500	22	1341	2,16	0,098
Yeast	Biology	[17]	2417	103	14	198	4,237	0,303

Table 3.1: Some statistics of the benchmark datasets.

Software

The MULAN library [89] was used for the implementation of the algorithms proposed and the existing methods in the literature. MULAN is a Java library which contains several methods for MLL, and it is constructed over the popular data mining tool WEKA [94]. On the other hand, the JCLEC library [95], which is a framework for evolutionary computation, was used to implement those methods that use evolutionary techniques.

Performance evaluation

In all experiments, a stratified 10-fold cross validation method [96] was carried out. To stratify the multi-label data, the methods proposed in [97] were used. Owing to the random nature of the evolutionary techniques, for each experiment, several runs were executed and the average value was calculated. In the experiments that involved lazy algorithms, the best number of neighbors was determined for each classifier on each dataset. In the experiments related to AL, a pool-based scenario [98] was employed.

Several evaluation measures proposed in [4, 15, 36] were used to assess the effectiveness of the multi-label algorithms. The formulation of these measures, as also their interpretations, can be consulted in the articles derived from this thesis.

In all experiments, the results were statistically validated to analyze if there were significant differences between the algorithms compared. The comparisons between algorithms were carried out using non-parametric statistical tests as proposed in [86–88]. The Wilcoxon's test [99] was conducted to compare a pair of algorithms. The Friedman's test [100] was used to perform multiple comparisons. In case that Friedman's test detected significant differences, the Bergmann-Hommel [101] and Shaffer [102] tests were used to perform all pairwise comparisons, and the Hommel procedure [103] was employed to conduct multiple comparisons with a control method.

4 Results

This chapter summarizes the different methods proposed and briefly presents the results achieved in regard to the objectives aimed in this thesis.

Improving performance on complex multi-label data

The performance of learning algorithms, specifically the lazy algorithms, is affected in datasets which have a high dimensionality. Generally speaking, lazy algorithms consider any feature equally important for classifying a query; yet irrelevant, interacting, redundant and noisy features have a highly negative impact in the performance of these algorithms.

In study [104], a feature estimation method for multi-label data was proposed. In this work, a heuristic based on a similarity measure to compute an adequate weight vector was designed. The proposal takes as premise that the similarity between label sets is a good heuristic to estimate the similarity between examples in the feature space. Given a subset of training examples (validation set), a weight vector is estimated. For each validation example, two rankings of examples are calculated, a ranking of examples based on feature space and another ranking of examples based on label space. The aim is to learn a weight vector that minimizes the distance between the two rankings associated to each validation example. For solving the optimization problem, a Genetic Algorithm (GA) with a real codification was designed.

The best weight vector found by the GA is used to improve the effectiveness of the multi-label lazy algorithms. The weight vector allows the distance function to have a greater probability to recover examples with labels sets more relevant for classification. The effectiveness of the method proposed was tested with the MLkNN [27], BRkNN [28] and IBLRML [30] lazy algorithms. Several evaluation metrics related to MLC and LR tasks were used to assess the effectiveness of the feature estimation method proposed. A statistical test validated that the weighted lazy algorithms, which parameterize their distance functions using the weight vectors learned, obtained significantly better results than the original versions (non-weighted) of the lazy algorithms.

In study [105], a more sophisticated feature estimation method was presented. A new way of computing the rankings of examples is proposed, where only the k nearest neighbors of each validation example are considered. On the other hand, a new metric to compute the distance between the rankings of examples is formulated. In this work, the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [106, 107] algorithm was employed for the resolution of the optimization problem. CMA-ES optimizes the metric defined as a heuristic to estimate an adequate weight vector. As in case [104], the main goal was to examine the benefit of feature estimation methods to improve the performance of multi-label lazy algorithms. The effectiveness of the method proposed was tested with the MLkNN [27], BRkNN [28], DMLkNN [29], IBLRML [30] and MLCWkNN [57] algorithms.

In study [108], an extension of the well-known ReliefF algorithm [40] to multi-label data was proposed. The method proposed, named ReliefF-ML, directly estimates the utility of the features, i.e. it does not use any PTM. The concepts *Hits* and *Misses* used by the classic ReliefF algorithm were redefined. ReliefF-ML can be considered a generalization of the classic ReliefF, where the equation used to update the weights was modified. The effectiveness of the method proposed was validated on the improvement of the performance of three multi-label lazy algorithms, MLkNN [27], DMLkNN [29] and MLCWkNN [57].

In study [109], two another extensions of the ReliefF algorithm for multi-label data were proposed. The first extension proposed, named PPT-ReliefF, uses the Pruned Problem Transformation (PPT) method [19] to convert the original multi-label dataset into a new multi-class dataset. PPT has the power of LPS approach, where the correlation among labels is implicitly taken into account, but PPT only considers the most important label relationships. PPT approach reduces the scarcity of labels and the over-fitting of data. The second extension proposed, named RReliefF-ML, is based on the well known ReliefF adaptation to regression problems [110]. RReliefF-ML does not use a PTM for the estimation of the feature weights, it retrieves only k-nearest neighbors for each sampling example. In this work, the two extensions proposed PPT-ReliefF and RReliefF-ML, and the extension ReliefF-ML that was proposed in [108], were compared to other existing state-of-the-art ReliefF extensions. The experimental study showed that the three methods significantly improved the performance of the multi-label lazy algorithms.

On the other hand, the effectiveness of the three methods (PPT-ReliefF, RReliefF-ML and ReliefF-ML) was evaluated in FS task. The weight vectors were converted into feature rankings, the features are ordered according their relevance, and these rankings guided the search of the best subset of features. The evidence suggested that the distributions of the relevant features on the top of the rankings determined by PPTReliefF, ReliefF-ML and RReliefF-ML were better than the distributions of the relevant features determined by the other ReliefF extensions considered in the comparison. The study showed that the baseline classifiers can obtain formidable results on complex multi-label datasets considering a small number of features.

In study [111], a multi-label lazy algorithm based on the principles of DGC approach was proposed. The method proposed, named MLDGC, directly handles multi-label data, and considers each example as an atomic data particle. Considering each example as an atomic data particle, the problems that arise in the creation of artificial particles from several examples are avoided. In this work, the concept of *Neighborhood-based Gravitation Coefficient* was introduced, which is used in the calculation of the gravitation forces. MLDGC has an acceptable computational complexity. MLDGC was compared to 12 multi-label lazy algorithms, confirming the effectiveness of this data gravitation model for better multi-label lazy learning.

The publications associated to this part of the dissertation are:

O. Reyes, C. Morell and S. Ventura. ***Learning Similarity Metric to improve the performance of Lazy Multi-label Ranking Algorithms***. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA'2012). IEEE, pp. 246-251, 2012.

O. Reyes, C. Morell and S. Ventura. ***ReliefF-ML: an extension of ReliefF algorithm to multi-label learning***. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. LNCS, Springer, vol. 8259, pp. 528-535, 2013.

O. Reyes, C. Morell and S. Ventura. ***Evolutionary feature weighting to improve the performance of multi-label lazy algorithms***. Integrated Computer-Aided Engineering, vol. 21, no. 4, pp. 339-354, 2014.

O. Reyes, C. Morell and S. Ventura. ***Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context***. Neurocomputing, vol. 161, pp. 168-182, 2015.

O. Reyes, C. Morell and S. Ventura. ***Effective lazy learning algorithm based on a data gravitation model for multi-label learning***. Information Sciences, vol. 340-341, pp. 159-174, 2016.

Improving performance through unlabeled data

The main challenge in performing AL on multi-label data is designing effective strategies that measure the unified informative potential of unlabeled examples across all labels. On the other hand, developing efficient strategies is a crucial point in scenarios where a base classifier which has a costly training process is used, or the time that the expert can wait to label the unlabeled examples is limited.

In study [112], a MLAL strategy, named Category Vector Inconsistency and Ranking of Scores (CVIRS), was proposed. Two uncertainty measures based on the predictions of the base classifier and the inconsistency of a predicted label set regarding to the label dimension of the labeled dataset, respectively, were defined to select the most uncertain examples. Given an unlabeled example, the difference margins[1] in predictions of classifier with respect to whether the example belongs or does not belong to each label is computed. An example with large margin value on a label means that the classifier has small error in differentiating whether the example belongs or does not belong to this label. On the other hand, an example with small margin value on a label means that it is more ambiguous for the current classifier to predict whether the example belongs or does not belong to this label. The calculus of the unified uncertainty that has the classifier with respect to an unlabeled example was formulated as an rank aggregation problem. A simple and efficient positional method was used to resolve the rank aggregation problem formulated.

On the other hand, a measure that represents the inconsistency of a predicted label set was defined. This measure is based on the premise that as the labeled set and unlabeled set are drawn from the same underlying distribution, is expected that predicted label sets and the label sets of labeled examples share common properties. The inconsistency of a predicted label set is calculated by means of the Hamming and entropic distances between two binary vectors. Based on the two measures defined (uncertainty and inconsistency), CVIRS iteratively selects the unlabeled examples that have high uncertainty levels and, at the same time, high inconsistency in their predicted label sets. This approach can be used with any base classifier which can obtain proper probability estimates from its outputs. The proposal is not restricted to base classifiers that use PTM, it can also be used with multi-label algorithms that belong to AAM category. CVIRS was compared to seven state-of-the-art MLAL strategies, confirming the effectiveness of the proposal for better MLAL.

Most state-of-the-art MLAL strategies were designed to select one unlabeled example at a time. This type of AL strategy can be easily used to select a batch of unlabeled examples, e.g. by selecting the k best instances in a greedy manner, but the information overlapping between the selected instances is not considered. The most significant works related to performing batch-mode AL on multi-label

[1] The difference margin is defined as the difference between the probabilities that an example belongs or does not belong to a particular label.

data appeared in [74, 81]. In these previous works, the batch selection task is commonly formulated as a NP-hard integer programming problem. However, the use of this type of methods is difficult, practically speaking, for their application to large-scale multi-label datasets. On the other hand, most MLAL strategies only use informativeness-based[2] criteria to select the most useful unlabeled examples, leading to a sub-optimal performance [82]. Other types of selection criteria, such as representativeness[3] and diversity[4], have been rarely considered in the MLAL context. Few works have combined two selection criteria to select the best unlabeled examples [74, 78, 79, 81, 82], notably informativeness and representativeness, or informativeness and diversity. To date, to the best of our knowledge, a MLAL strategy that combines the three criteria (informativeness, representativeness and diversity) had not been proposed.

In study [113], a MLAL strategy, named Evolutionary Strategy for Batch-Mode Multi-Label Active Learning (ESBMAL), was proposed. ESBMAL formulates the BMAL problem as a multi-objective optimization problem, and the optimization problem is solved by the well-known NSGA-II algorithm [114]. The evolutionary algorithm tries to optimize three measures based on informativeness, diversity and representativeness, respectively. An individual of the population represents a candidate batch of examples. In each AL iteration, ESBMAL aims to select a set of unlabeled examples which are usually informative across all labels, diverse between each other, and representative of the underlying distribution. ESBMAL can be used with any base multi-label classifier which can obtain proper probability estimates from its outputs. ESBMAL is more efficient, in computational terms, than state-of-the-art multi-label BMAL strategies. The experimental study showed the effectiveness of the proposal for better multi-label BMAL.

Next, complementary results derived of this thesis are briefly exposed:

- Currently, there are several software tools which assist the experimentation process and development of new algorithms in DM and ML areas, such as Rapid Miner, WEKA, Scikit-learn, Orange and KEEL. However, these tools are focused to supervised and unsupervised learning problems. To date,

[2] Informativeness measures the effectiveness of an example by reducing the uncertainty of a model.

[3] Representativeness measures whether an unlabeled example is representative of the underlying distribution.

[4] Diversity measures the information redundancy that exist among a set of examples.

there has been insufficient effort towards the creation of a computational tool mainly focused to AL.

The above situation motivated the development of the JCLAL[5] (Java Class Library for Active Learning) framework [115]. JCLAL is an open source software for researchers and end-users to develop AL methods. JCLAL aims to bring the benefits of open source software to people working in the area of AL. It includes the most relevant strategies that have been proposed in single-label and multi-label learning paradigms. It provides the necessary interfaces, classes and methods to develop any AL method.

JCLAL is an open source project under the GNU General Public License (GPL). It has an architecture that follows strong principles of object-oriented programming, where it is common and easy to reuse code. Through a flexible class structure, the library provides the possibility of including new AL methods, as well as the ability to adapt, modify or extend the framework according to developer's needs.

- Despite the call made by the ML community for a rigorous and correct statistical analysis of published results, the use of statistical tests for analyzing the performance of AL methods has not been rigorous. Through an extensive bibliographic review of works published in the AL area, we observed that many excellent and innovative AL papers end by drawing conclusions by means of visually comparing learning curves.

  The visual comparison of learning curves is effective when a small number of AL strategies are compared, and their performances differ sufficiently so that the learning curves do not overlap greatly. Conversely, the visual comparison of several learning curves can be very confusing, as the learning curves may intersect at many points. If several active learning strategies are compared over multiple datasets, and their performances are similar over multiple datasets, the resulting graphs may be very difficult to interpret, and the visual analysis of the AL performance may be a very difficult task to accomplish. Consequently, conclusions from questions such as, which is the AL strategy that delivers the best performance?, are not possible, or very difficult to draw.

[5] http://jclal.sourceforge.net

In study [116], two approaches, based on the use of non-parametric statistical tests, to statistically compare AL strategies over multiple datasets were proposed. The first approach is based on the analysis of the Area Under learning Curve (AUC) and the rate of performance change. The concept *True Performance of a selection strategy* (TP) is defined. A TP score can be interpreted as a general view of the performance of an AL strategy. After computing the TP scores of the AL strategies on each dataset, a statistical analysis can be carried out, and then final considerations could be given with a statistical support.

The second approach, instead of only considering the final results (TP scores), analyzes the intermediate results generated in each iteration of the AL process. This can reveal very significant information when AL strategies are compared, especially in cases where TP scores are statistically similar. The application of both approaches was illustrated by means of an experimental study, demonstrating the usefulness of the proposal for improving analysis of AL performance.

The publications associated to this part of the dissertation are:

O. Reyes, C. Morell and S. Ventura. ***Effective active learning strategy for multi-label learning***. Neurocomputing, submitted, 2015.

O. Reyes and S. Ventura. ***Evolutionary Strategy to perform Batch-Mode Active Learning on Multi-label Data***. ACM Transactions on Intelligent Systems and Technology, submitted, 2016.

O. Reyes, E. Pérez, M. C. Rodríguez Hernández, H. M. Fardoun and S. Ventura. ***JCLAL: A Java Framework for Active Learning***. Journal of Machine Learning Research, vol. 17 (95), pp. 1-5, 2016.

O. Reyes, A. H. Altahi and S. Ventura. ***Statistical Comparisons of Active Learning Strategies over Multiple Datasets***. Information Sciences, submitted, 2016.

5

Conclusions and future work

This chapter briefly summarizes the concluding remarks obtained from the research of the thesis and provides research lines for future work.

Conclusions

This Ph.D. thesis focused on MLL paradigm. The bibliographic study allowed to detect open problems, and formulate the hypotheses that guided this research. The work developed cannot be considered to be concluded, due to the extent of the topics treated and their possible application to other areas.

Improving performance on complex multi-label data

The first part of the thesis focused on the improvement of the performance of multi-label algorithms, specifically the lazy algorithms, on complex multi-label data.

In studies [104, 105, 108, 109], five feature estimation methods were proposed. Two methods are based on the application of evolutionary algorithms to estimate an adequate weight vector. The three other methods proposed are extensions of the well-known ReliefF algorithm. The methods based on ReliefF approach are more computationally efficient than the two other based on evolutionary techniques.

The results showed that it is possible to obtain a good estimation of the feature weights by directly handling the multi-label data, i.e. without using a PTM. The parameterization of the distance functions with a weight vector allowed to recover examples with relevant label sets for classification. The evidence showed that the methods proposed significantly improve the performance of the multi-label lazy algorithms on complex multi-label data, proving the hypothesis $\mathbf{H}_1$ formulated in this thesis.

On the other hand, a weight vector can be useful to guide the search process of the best subsets of features. In case [109], it is showed how by converting the weight vectors into feature rankings, it is possible to select small subsets of features efficiently, leading to a significant improvement of the performance of multi-label algorithms. The evidence showed that the methods perform well on the FS task, proving the hypothesis $\mathbf{H}_2$ formulated in this thesis.

In study [111], a multi-label lazy algorithm based on DGC approach was proposed. Considering each example as an atomic data particle avoided the problems that arise in the creation of artificial particles from several examples. The introduction of the new concept *Neighborhood-based Gravitation Coefficient* achieved better levels of classification. This coefficient strengthens or weakens the effect that a particle has over a test example. Two particles at the same distance of a test example, but with different levels of purity in their neighborhood, will exert different gravitational forces. The evidence showed that the method proposed significantly outperformed the state-of-the-art multi-label lazy algorithms. The method provides a good trade-off between efficiency and effectiveness in its solution, proving the hypothesis $\mathbf{H}_3$ formulated in this thesis.

In general, the specific objectives $\mathbf{O}_1$ and $\mathbf{O}_2$, declared at the beginning of this thesis, were fulfilled through the results obtained in the works [104, 105, 108, 109, 111].

Improving performance through unlabeled data

The second part of the thesis focused on the development of MLAL strategies not restricted to a type of base classifier, and they directly handle the multi-label data. The multi-label BMAL problem was also analyzed.

In study [112], an efficient MLAL strategy was proposed. Two measures to select the unlabeled examples were defined. The first measure is related to the uncertainty in the predictions of the base classifier. A rank aggregation problem was formulated to compute the unified uncertainty of an unlabeled example, and this problem was solved by an efficient positional method. The second measure is related to the inconsistency of the predicted labels sets. The combination of these two measures allowed to select examples that not only are informative for the current model, but they are also not representative of the data distribution of the labeled set. The method proposed can be used with any base classifier which can obtain proper probability estimates from its outputs. The evidence showed that the method significantly outperformed several state-of-the-art MLAL strategies, proving the hypothesis $\mathbf{H}_4$ formulated in this thesis.

In study [113], a multi-label BMAL strategy was proposed. Three measures based on informativeness, representativeness and diversity were defined, respectively. The multi-label BMAL problem was formulated as a multi-objective problem, and the optimization problem was solved by an evolutionary algorithm. The solutions reached by the evolutionary algorithm represent sets of examples which are usually informative across all labels, diverse between each other, and representative of the underlying distribution. The results showed that the multi-objective problem formulated is a good heuristic to resolve the multi-label BMAL problem, as also that the evolutionary algorithms are effective in the resolution of this type of problem. The method is more computationally efficient than other existing approaches, which commonly formulate the multi-label BMAL problem as a complex integer programming problem. The method can be used with any base multi-label classifier which can obtain proper probability estimates from its outputs. The evidence showed that the method significantly outperformed the state-of-the-art multi-label BMAL strategies, proving the hypothesis $\mathbf{H}_5$ formulated in this thesis.

As complementary results derived of this thesis, a computational tool that allows the implementation of AL methods in a simple manner and favors the experimentation on this area was developed [115]. This framework was announced to AL community in November 2014 and has had a good acceptation.

Finally, in study [116], two approaches to assess the performance of AL methods were proposed. The first approach is based on the analysis of the AUC and the rate of performance change. The second approach analyses the intermediate results derived from AL iterations. The second approach is more robust and powerful than the first one, it is able to detect less significant differences. The evidence showed the usefulness of non-parametric tests in the evaluation of AL performance.

In general, the specific objective $\mathbf{O}_3$, declared at the beginning of this thesis, was fulfilled through the results obtained in the works [112, 113].

Future work

In this section, some remarks for future lines of research that arise from the studies developed in this thesis are provided.

To date, there are still some issues that remain far less studied in the MLL paradigm. Through the bibliographic study carried out in the development of this work, the following promising research lines were detected:

- Instance selection algorithms for multi-label data. To date, very few works have been proposed in this sense.
- Imbalanced learning techniques for multi-label data. In last years, this line of research has gained the attention of the scientific community due to multi-label datasets commonly show a non-uniform distribution of examples per label.
- Algorithms for multi-label data streams. In this sense, it is important the development of incremental multi-label algorithms.
- Dimensionality reduction in the label space. To date, very few proposals have been presented in this sense.
- SSL algorithms for multi-label data. SSL algorithms for multi-label data, in comparison to single-label data, have been far less studied. On the other hand, the combination of AL and SSL approaches is an interesting research line.

On the other hand, the following research lines are also proposed from the results obtained in this work:

- Define new heuristics to learn similarity metrics in order to improve the performance of multi-label lazy algorithms.
- Propose feature estimation methods based on evolutionary techniques that allow to effectively select subsets of relevant features on multi-label data.
- Propose other models based on DGC approach for better multi-label lazy learning.

- Design new AL strategies based on evolutionary techniques to effectively resolve the multi-label BMAL problem.

- Extend the AL methods proposed in this thesis to select example-label pairs, instead of consulting all possible labels of the selected examples. The selection of example-label pairs, taking into account the dependence between labels, can lead to a considerable reduction on the data labeling cost.

- Adapt the methods proposed in this thesis to the multi-instance multi-label problem. In multi-instance multi-label problems, examples are described by multiple instances and they are associated with multiple class labels.

- Extend JCLAL library by including other AL strategies, for instance AL strategies for multi-instance learning and multi-instance multi-label learning. On the other hand, it would be interesting the development of a module that allows the distributed computation of AL strategies, thus enabling the use of the library in Big Data area.

Bibliography

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

[2] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed. New Jersey, United States of America: John Wiley & Sons, 2014.

[3] T. M. Mitchell, *Machine Learning.* McGraw-Hill, 1997.

[4] E. Gibaja and S. Ventura, “Multi-label learning: a review of the state of the art and ongoing research,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 6, pp. 411–444, 2014.

[5] J. P. Pestian, C. Brew, P. Matykiewicz, D. J. Hovermale, N. Johnson, K. B. Cohen, and W. Duch, “A shared task involving multi-label classification of clinical free text,” in *Proceedings of the Workshop on Biological, Translational, and Clinical Language Processing (BioNLP'2007).* Stroudsburg, PA, United States of America: Association for Computational Linguistics, 2007, pp. 97–104.

[6] I. Katakis, G. Tsoumakas, and I. Vlahavas, “Multilabel text classification for automated tag suggestion,” in *Proceedings of the ECML/PKDD Discovery Challenge*, vol. 75, 2008.

[7] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan, “Matching words and pictures,” *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, 2003.

[8] S. Yang, S. Kim, and Y. Ro, “Semantic home photo categorization,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp. 324–335, 2007.

[9] E. Correa, A. Plastino, and A. Freitas, "A Genetic Algorithm for Optimizing the Label Ordering in Multi-Label Classifier Chains," in *Proceedings of the 25th International Conference on Tools with Artificial Intelligence (ICTAI'2013).* IEEE, 2013, pp. 469–476.

[10] M. Boutell, J. Luo, X. Shen, and C. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[11] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 467–476, 2008.

[12] F. Otero, A. Freitas, and C. Johnson, "A hierarchical multi-label classification ant colony algorithm for protein function prediction," *Memetic Computing*, vol. 2, no. 3, pp. 165–181, 2010.

[13] M. G. Larese, P. Granitto, and J. Gómez, "Spot defects detection in cDNA microarray images," *Pattern Analysis and Applications*, vol. 16, no. 3, pp. 307–319, 2013.

[14] E. Ukwatta and J. Samarabandu, "Vision based metal spectral analysis using multi-label classification," in *Canadian Conference on Computer and Robot Vision (CRV'2009).* IEEE, 2009, pp. 132–139.

[15] G. Tsoumakas, I. Katakis, and I. Vlahavas, *Data Mining and Knowledge Discovery Handbook*, 2nd ed. New York, United States of America: Springer-Verlag, 2010, ch. Mining Multi-label Data, pp. 667–686.

[16] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333–359, 2011.

[17] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2001, pp. 681–687.

[18] J. Furnkranz, E. Hullermeier, E. Mencía, and K. Brinker, "Multilabel classification via calibrated label ranking," *Machine Learning*, vol. 73, no. 2, pp. 133–153, 2008.

[19] J. Read, "A pruned problem transformation method for multi-label classification," in *Proceedings of the New Zealand Computer Science Research Student Conference (NZCSRS'2008)*, 2008, pp. 143–150.

[20] G. Tsoumakasa, I. Katakis, and I. Vlahavas, "Effective and efficient multilabel classification in domains with large number of labels," in *Proceedings of the ECML/PKDD Workshop on Mining Multidimensional Data (MMD'2008)*, 2008, pp. 30–44.

[21] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 1079–1089, 2011.

[22] H. Blockeel, L. Raedt, and J. Ramon, "Top-down induction of clustering trees," in *Proceedings of the 15th International Conference on Machine Learning*, 1998, pp. 55–63.

[23] A. Clare and R. King, "Knowledge discovery in multi-label phenotype data," in *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'2001).* Springer, 2001, pp. 42–53.

[24] K. Crammer and Y. Singer, "A family of additive online algorithms for category ranking," *Journal of Machine Learning Research*, vol. 3, pp. 1025–1058, 2003.

[25] M. L. Zhang and Z. H. Zhou, "Multi-label neural networks with applications to functional genomics and text categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, pp. 1338–1351, 2006.

[26] R. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization," *Machine Learning*, vol. 39, no. 2, pp. 135–168, 2000.

[27] M. L. Zhang and Z. H. Zhou, "ML-kNN: A lazy learning approach to multilabel learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[28] E. Spyromitros, G. Tsoumakas, and I. Vlahavas, "An empirical study of lazy multi-label classification algorithms," in *Artificial Intelligence: Theories, Models and Applications.* Springer, 2008, pp. 401–406.

[29] Z. Younes, F. Abdallah, and T. Denceux, “Multi-label classification algorithm derived from k-nearest neighbor rule with label dependencies,” in *Proceedings of the 16th Eropean Signal Processing Conference (EUSIPCO’2008).* IEEE, 2008, pp. 1–5.

[30] W. Cheng and E. Hullermeier, “Combining instance-based learning and logistic regression for multilabel classification,” *Machine Learning*, vol. 76, no. 2-3, pp. 211–225, 2009.

[31] Z. Younes, F. Abdallah, and T. Denceux, “An Evidence-Theoretic K-Nearest Neighbor Rule for Multi-label Classification,” in *Scalable Uncertainty Management*, ser. LNAI, I. Godo and A. Pugliese, Eds., vol. 5785. Springer, 2009, pp. 297–308.

[32] N. Ueda and K. Saito, “Parametric mixture models for multi-labeled text,” in *Proceedings of Advances in Neural Information Processing Systems (NIPS’2015).* MIT Press, 2002, pp. 737–744.

[33] S. Diplarisa, G. Tsoumakas, P. Mitkas, and I. Vlahavas, “Protein classification with multiple algorithms,” in *Proceedings of the 10th Panhellenic Conference on Informatics (PCI’2005)*, ser. LNCS, vol. 3746. Springer, 2005, pp. 448–456.

[34] S. Dendamrongvit, P. Vateekul, and M. Kubat, “Irrelevant attributes and imbalanced classes in multi-label text-categorization domains,” *Intelligent Data Analysis*, vol. 15, no. 6, pp. 843–859., 2011.

[35] F. Charte, A. J. Rivera, M. J. del Jesus, and F. Herrera, “Addressing imbalance in multilabel classification: Measures and random resampling algorithms,” *Neurocomputing*, vol. 163, pp. 3–16, 2015.

[36] G. Madjarov, D. Kocev, and D. Gjorgjevikj, “An extensive experimental comparison of methods for multi-label learning,” *Pattern Recognition*, vol. 45, pp. 3084–3104, 2012.

[37] K. Kira and L. Rendell, “A practical approach to feature selection,” in *Proceedings of the ninth International Workshop on Machine learning.* Morgan Kaufmann, 1992, pp. 249–256.

[38] D. Wettschereck, D. W. Aha, and T. Mohri, “A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms,” *Artificial Intelligence Review*, vol. 11, pp. 273–314, 1997.

[39] A. Abraham, E. Corchado, and J. Corchado, “Hybrid learning machines,” *Neurocomputing*, vol. 72, pp. 2729–2730, 2009.

[40] I. Kononenko, “Estimating attributes: analysis and extensions of ReliefF,” in *Proceedings of the European Conference on Machine Learning (ECML’1994)*. Catania, Italy: Springer, 1994, pp. 171–182.

[41] M. Robnik-Sikonja and I. Kononenko, “Theoretical and empirical analysis of ReliefF and RReliefF,” *Machine Learning*, vol. 53 (1-2), pp. 23–69, 2003.

[42] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, “Heuristic search over a ranking for feature selection,” in *Proceedings of IWANN’2005, Computational intelligence and bioinspired systems*, ser. LNCS, vol. 3512. Springer, 2005, pp. 742–749.

[43] L. Yu and H. Liu, “Feature selection for high-dimensional data: a fast correlation-based filter solution,” in *Proceedings of the 20th International Conference on Machine Learning (ICML’2000)*, Washington DC, 2003, pp. 856–863.

[44] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[45] M. L. Zhanga, J. M. Peña, and V. Robles, “Feature selection for multi-label naive bayes classification,” *Information Sciences*, vol. 179, pp. 3218–3229, 2009.

[46] Q. Gu, Z. Li, and J. Han, “Correlated multi-label feature selection,” in *Proceedings of the 20th ACM international Conference on Information and knowledge Management (CIKM’2011)*. Scotland, United Kingdom: ACM, 2011.

[47] N. Spolaôr, E. Cherman, and M. Monard, “Using ReliefF for multi-label feature selection,” in *Proceedings of the Conferencia Latinoamericana de Informática*, Brazil, 2011, pp. 960–975.

[48] N. Spolaôr, E. Cherman, M. Monard, and H. Lee, "Filter approach feature selection methods to support multi-label learning based on ReliefF and Information Gain," in *Proceedings of the Advances in Artificial Intelligence (SBIA'2012)*, ser. LNCS, vol. 7589. Springer, 2012, pp. 72–81.

[49] D. Kong, C. Ding, H. Huang, and H. Zhao, "Multi-label ReliefF and F-statistic feature selections for image annotation," in *Proceedings of Computer Vision and Pattern Recognition (CVPR'2012).* IEEE, 2012, pp. 2352–2359.

[50] J. Lee and D. W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, vol. 34, pp. 349–357, 2013.

[51] N. Spolaôr, E. Alvares, M. Carolina, and H. Diana, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electronic Notes in Theoretical Computer Science*, vol. 292, pp. 135–151, 2013.

[52] G. Doquire and M. Verleysen, "Mutual information-based feature selection for multilabel classification," *Neurocomputing*, 2013.

[53] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[54] I. Kononenko and M. Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms.* Cambridge, United Kingdom: Horwood Publishing, 2007.

[55] K. Brinker and E. Hüllermeier, "Case-based multilabel ranking," in *Proceedings of the 20th International Conference on Artificial Intelligence (IJCAI'2007)*, 2007, pp. 702–707.

[56] Z. Younes, F. Abdallah, and T. Denoux, "Fuzzy multi-label learning under veristic variables," in *Proceedings of International Conference on Fuzzy Systems.* IEEE, 2010, pp. 1–8.

[57] J. Xu, "Multi-label weighted k-nearest neighbor classifier with adaptive weight estimation," in *Proceedings of the ICONIP'2011, Neural Information Processing*, ser. LNCS, vol. 7073. Springer, 2011, pp. 79–88.

[58] L. Peng, B. Peng, Y. Chen, and A. Abraham, "Data gravitation based classification," *Information Sciences*, vol. 179, no. 6, pp. 809–819, 2009.

[59] A. Cano, A. Zafra, and S. Ventura, "Weighted Data Gravitation Classification for Standard and Imbalanced Data," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1672–1687, 2013.

[60] A. Srivastava and B. Zane-Ulman, "Discovering recurring anomalies in text reports regarding complex space systems," in *Proceedings of the Aerospace Conference*. IEEE, 2005, pp. 55–63.

[61] C. Snoek, M. Worring, J. van Gemert, J. Geusebroek, and A. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," in *Proceedings of the 14th annual ACM International Conference on Multimedia*. Santa Barbara, United States of America: ACM, 2006, pp. 421–430.

[62] E. L. Mencía and J. Furnkranz, "Efficient pairwise multi-label classification for large-scale problems in the legal domain," in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'2008)*. Antwerp, Belgium: Springer-Verlag, 2008, pp. 50–65.

[63] T. S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. T. Zheng, "NUS-WIDE: A Real-World Web Image Database from National University of Singapore," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. Greece: ACM, 2009.

[64] X. Zhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publishers, 2009.

[65] B. Settles, *Active Learning*, 1st ed., ser. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2012.

[66] X. Li, L. Wang, and E. Sung, "Multi-label SVM active learning for image classification," in *Proceedings of the International Conference on Image Processing (ICIP'2004)*, vol. 4. IEEE, 2004, pp. 2207–2210.

[67] K. Brinker, *From Data and Information Analysis to Knowledge Engineering.* Springer, 2006, ch. On Active Learning in Multi-label Classification, pp. 206–213.

[68] G. Qi, X. Hua, Y. Rui, J. Tang, and H. Zhang, "Two-dimensional active learning for image classification," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'2008).* IEEE, 2008, pp. 1–8.

[69] ——, "Two-dimensional multi-label active learning with an efficient online adaptation model for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. 1, 2009.

[70] B. Yang, J. Sun, T. Wang, and Z. Chen, "Effective multi-label active learning for text classification," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Paris, France: ACM, 2009, pp. 917–926.

[71] X. Zhang, J. Cheng, C. Xu, H. Lu, and S. Ma, "Multi-view multi-label active learning for image classification," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'2009).* IEEE, 2009, pp. 258–261.

[72] A. Esuli and F. Sebastiani, "Active Learning Strategies for Multi-Label Text Classification," in *Advances in Information Retrieval.* Springer, 2009, pp. 102–113.

[73] M. Singh, E. Curran, and P. Cunningham, "Active learning for multi-label image annotation," in *Proceedings of the 19th Irish Conference on Artificial Intelligence and Cognitive Science*, 2009, pp. 173–182.

[74] S. Chakraborty, V. Balasubramanian, and S. Panchanathan, "Optimal Batch Selection for Active Learning in Multi-label Classification," in *Proceedings of the 19th ACM international conference on Multimedia (MM's2011).* Scottsdale, Arizona, United States of America: ACM, 2011, pp. 1413–1416.

[75] C. W. Hung and H. T. Lin, "Multi-label active learning with auxiliary learner," in *Proceedings of the Asian Conference on Machine Learning.* JMLR, 2011, pp. 315–330.

[76] P. Wang, P. Zhang, and L. Guo, “Mining multi-label data streams using ensemble-based active learning,” in *Proceedings of the 12th SIAM International Conference on Data Mining*, 2012, pp. 1131–1140.

[77] J. Tang, Z.-J. Zha, D. Tao, and T.-S. Chua, “Semantic-gap-oriented active learning for multilabel image annotation,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2354–2360, 2012.

[78] X. Li and Y. Guo, “Active Learning with Multi-Label SVM Classification,” in *Proceedings of the 23th International joint Conference on Artificial Intelligence.* AAAI Press, 2013, pp. 1479–1485.

[79] S. Huang and Z. Zhou, “Active query driven by uncertainty and diversity for incremental multi-label learning,” in *Proceedings of 13th International Conference on Data Mining.* IEEE, 2013, pp. 1079–1084.

[80] J. Wu, V. Sheng, J. Zhang, P. Zhao, and Z. Cui, “Multi-label active learning for image classification,” in *Proceedings of the International Conference on Image Processing.* IEEE, 2014, pp. 5227–5231.

[81] B. Zhang, Y. Wang, and F. Chen, “Multilabel image classification via high-order label correlation driven active learning,” *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1430–144, 2014.

[82] S. Huang, R. Jin, and Z. Zhou, “Active learning by querying informative and representative examples,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 1936–1949, 2014.

[83] D. Vasisht and A. Damianou, “Active learning for sparse bayesian multilabel classification,” in *Proceedings of the 20th SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2014, pp. 472–481.

[84] S. Huang, S. Chen, and Z. Zhou, “Multi-label active learning: Query type matters,” in *Proceedings of the 24th International Conference on Artificial Intelligence.* AAI Press, 2015, pp. 946–952.

[85] Y. Fu, X. Zhu, and A. K. Elmagarmid, “Active learning with optimal instance subset selection,” *IEEE Transactions on Cybernetics*, vol. 43, no. 2, 2013.

[86] J. Demšar, “Statistical Comparisons of Classifiers over Multiple Data Sets,” *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[87] S. García and F. Herrera, “An extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all pairwise comparisons,” *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.

[88] S. García, A. Fernández, J. Luengo, and F. Herrera, “Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power,” *Information Sciences*, vol. 180, pp. 2044–2064, 2010.

[89] G. Tsoumakas, E. Spyromitros-Xioufi, J. Vilcek, and I. Vlahavas, “MULAN: A java library for multi-label learning,” *Journal of Machine Learning Research*, vol. 12, pp. 2411–2414, 2011.

[90] F. Briggs and et. al., “The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment,” in *Proceedings of the International Workshop on Machine Learning for Signal Processing (MLSP’2013)*. IEEE, 2013.

[91] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth, “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,” in *Proceedings of the 7th European Conference on Computer Vision*, ser. LNCS, vol. 2353. Springer, 2002, pp. 97–112.

[92] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, “Multilabel classification of music into emotions,” in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR’2008)*, 2008, pp. 325–330.

[93] B. Klimt and Y. Yang, “The Enron corpus: a new dataset for email classification research,” in *Proceedings of the 15th European Conference on Machine Learning (ECML’2004)*. Springer, 2004, pp. 217–226.

[94] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA Data Mining Software: An Update,” in *SIGKDD Explorations*, vol. 11, no. 1. ACM, 2009, pp. 10–18.

[95] S. Ventura, C. Romero, A. Zafra, J. A. Delgado, and C. Hervás, "JCLEC: A java framework for evolutionary computation," *Soft Computing*, vol. 12, pp. 381–392, 2008.

[96] R. Kohavi, "A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, vol. 2, 1995, pp. 1137–1143.

[97] K. Sechidis, G. Tsoumakas, and I. Vlahavas, "On the stratification of multi-label data," in *Machine Learning and Knowledge Discovery in Databases.* Springer, 2011, pp. 145–158.

[98] D. Lewis and W. Gale, "A sequential algorithm for training text classifier," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* Dublin, Ireland: Springer, 1994, pp. 3–12.

[99] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

[100] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, pp. 86–92, 1940.

[101] G. Bergmann and G. Hommel, "Improvements of general multiple test procedures for redundant systems of hypotheses," *Multiple Hypotheses Testing*, pp. 100–115, 1988.

[102] J. Shaffer, "Modified sequentially rejective multiple test procedures," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 826–831, 1986.

[103] G. Hommel, "A stagewise rejective multiple test procedure based on a modified bonferroni test," *Biometrika*, vol. 75, no. 2, pp. 383–386, 1988.

[104] O. Reyes, C. Morell, and S. Ventura, "Learning similarity metric to improve the performance of lazy multi-label ranking algorithms," in *Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA'2012).* IEEE, 2012, pp. 246–251.

[105] ——, “Evolutionary feature weighting to improve the performance of multi-label lazy algorithms,” *Integrated Computer-Aided Engineering*, vol. 21, no. 4, pp. 339–354, 2014.

[106] N. Hansen and A.Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[107] A. Auger and N. Hansen, “A restart CMA evolution strategy with increasing population size,” in *Proceedings of the IEEE Congress on Evolutionary Computation*, vol. 2. IEEE, 2005, pp. 1769–1776.

[108] O. Reyes, C. Morell, and S. Ventura, “ReliefF-ML: an extension of ReliefF algorithm to multi-label learning,” in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. LNCS. Springer, 2013, vol. 8259, pp. 528–535.

[109] ——, “Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context,” *Neurocomputing*, vol. 161, pp. 168–182, 2015.

[110] M. Robnik-Šikonja and I. Kononenko, “An adaptation of Relief for attribute estimation in regression,” in *Proceedings of the Fourteenth International Conference on Machine Learning (ICML’97)*, 1997, pp. 296–304.

[111] O. Reyes, C. Morell, and S. Ventura, “Effective lazy learning algorithm based on a data gravitation model for multi-label learning,” *Information Sciences*, vol. 340-341, pp. 159–174, 2016.

[112] ——, “Effective active learning strategy for multi-label learning,” *Neurocomputing, submitted*, 2015.

[113] O. Reyes and S. Ventura, “Evolutionary strategy to perform batch-mode active learning on multi-label data,” *ACM Transactions on Intelligent Systems and Technology, submitted*, 2016.

[114] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multi-objective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[115] O. Reyes, E. Pérez, M. C. Rodríguez-Hernández, H. M. Fardoun, and S. Ventura, "JCLAL: A Java Framework for Active Learning," *Journal of Machine Learning Research*, vol. 17, pp. 1–5, 2016.

[116] O. Reyes, A. H. Altahi, and S. Ventura, "Statistical comparisons of active learning strategies over multiple datasets," *Information Sciences, submitted*, 2016.

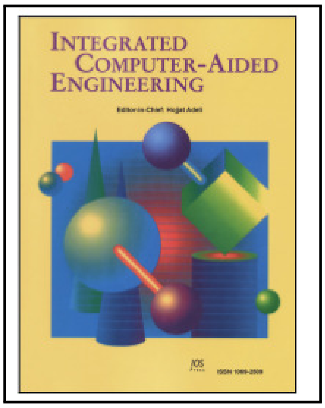
Part II: Journal Publications

Title:

Evolutionary feature weighting to improve the performance of multi-label lazy algorithms

Authors:

O. Reyes, C. Morell, and S. Ventura



Integrated Computer-Aided Engineering, *Volume 21, pp. 339-354, 2014*

Ranking:

Impact factor (JCR 2014): 4.698

Knowledge area:

Computer Science, Interdisciplinary Applications: 2/102

Computer Science, Artificial Intelligence: 5/123

DOI: 10.3233/ICA-140468

Integrated Computer-Aided Engineering 21 (2014) 339–354
DOI 10.3233/ICA-140468
IOS Press

Evolutionary feature weighting to improve the performance of multi-label lazy algorithms

Oscar Reyes[a], Carlos Morell[b] and Sebastián Ventura[c,d,*]
[a]*Computer Science Department, University of Holguín, Holguín, Cuba*
[b]*Computer Science Department, Universidad Central de Las Villas, Santa Clara, Cuba*
[c]*Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain*
[d]*Information Systems Department, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract. In the last decade several modern applications where the examples belong to more than one label at a time have attracted the attention of research into machine learning. Several derivatives of the k-nearest neighbours classifier to deal with multi-label data have been proposed. A k-nearest neighbours classifier has a high dependency with respect to the definition of a distance function, which is used to retrieve the k-nearest neighbours in feature space. The distance function is sensitive to irrelevant, redundant, and interacting or noise features that have a negative impact on the precision of the lazy algorithms. The performance of lazy algorithms can be significantly improved with the use of an appropriate weight vector, where a feature weight represents the ability of the feature to distinguish pattern classes. In this paper a filter-based feature weighting method to improve the performance of multi-label lazy algorithms is proposed. To learn the weights, an optimisation process of a metric is carried out as heuristic to estimate the feature weights. The experimental results on 21 multi-label datasets and 5 multi-label lazy algorithms confirm the effectiveness of the feature weighting method proposed for a better multi-label lazy learning.

Keywords: Feature weighting, lazy learning algorithms, multi-label classification, label ranking, learning metric, evolutionary algorithms

1. Introduction

In the last few decades, studies in the field of supervised learning have dealt with the analysis of data where the examples were associated with a single label [47,49,66]. However, there are several real problems where the examples belong to a set of labels at the same time, known as multi-label problems [63]. In the last few years an increasing number of modern applications that contain multi-label data have appeared, such as text categorisation [38], emotions evoked by music [35], semantic annotation of images [73] and videos [8], classification of protein function and gene [76].

Several multi-label lazy algorithms derivate of the k-nearest neighbours (k-NN) classifier scheme have been proposed on the multi-label learning context [12,56,72, 74,77]. In general, these algorithms do not construct a model from the training set, postponing almost all the process until classification. They classify a query by retrieving its k-nearest neighbours in feature space and after that, an aggregation strategy is performed to predict the set of labels of a query instance [63]. In the same way of single-label k-NN classifier, the multi-label lazy algorithms have a high dependency with respect to the definition of a distance function that is used to determine the k-nearest neighbours of a query instance. The main disadvantage of the multi-label lazy algorithms is that they consider any feature equally important for classifying a query; yet irrelevant, interacting, redundant and noisy features have a highly negative impact in the precision of these algorithms [67].

[*] Corresponding author: Sebastián Ventura, Department of Computer Science and Numerical Analysis, University of Córdoba, Albert Einstein Building, Rabanales Campus, Córdoba, Spain. Tel.: +34 957 212 218; Fax: +34 957 218 630; E-mail: sventura@uco.es.

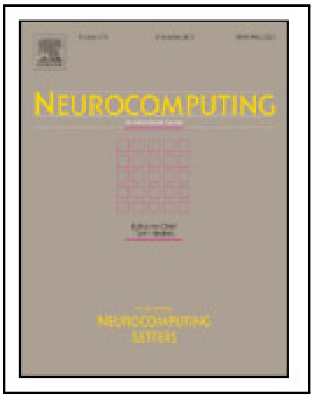
ISSN 1069-2509/14/$27.50 © 2014 – IOS Press and the author(s). All rights reserved

Title:

Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context

Authors:

O. Reyes, C. Morell, and S. Ventura



Neurocomputing, *Volume 161, pp. 168-182, 2015*

Ranking:

Impact factor (JCR 2015): 2.392

Knowledge area:

Computer Science, Artificial Intelligence: 31/130

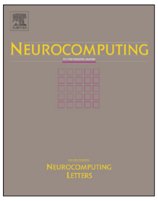
DOI: 10.1016/j.neucom.2015.02.045

Neurocomputing 161 (2015) 168–182

Contents lists available at ScienceDirect

Neurocomputing

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



Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context

Oscar Reyes [a], Carlos Morell [b], Sebastián Ventura [c,d,*]

[a] Department of Computer Science, University of Holguín, Cuba
[b] Department of Computer Science, Universidad Central de Las Villas, Cuba
[c] Department of Computer Science and Numerical Analysis, University of Córdoba, Spain
[d] Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form
12 February 2015
Accepted 13 February 2015
Communicated by Jiayu Zhou
Available online 21 February 2015

Keywords:
Multi-label learning
ReliefF algorithm
Feature weighting
Feature selection
Multi-label classification
Label ranking

ABSTRACT

Multi-label learning has become an important area of research due to the increasing number of modern applications that contain multi-label data. The multi-label data are structured in a more complex way than single-label data. Consequently the development of techniques that allow the improvement in the performance of machine learning algorithms over multi-label data is desired. The feature weighting and feature selection algorithms are important feature engineering techniques which have a beneficial impact on the machine learning. The ReliefF algorithm is one of the most popular algorithms to feature estimation and it has proved its usefulness in several domains. This paper presents three extensions of the ReliefF algorithm for working in the multi-label learning context, namely ReliefF-ML, PPT-ReliefF and RReliefF-ML. PPT-ReliefF uses a problem transformation method to convert the multi-label problem into a single-label problem. ReliefF-ML and RReliefF-ML adapt the classic ReliefF algorithm in order to handle directly the multi-label data. The proposed ReliefF extensions are evaluated and compared with previous ReliefF extensions on 34 multi-label datasets. The results show that the proposed ReliefF extensions improve preceding extensions and overcome some of their drawbacks. The experimental results are validated using several nonparametric statistical tests and confirm the effectiveness of the proposal for a better multi-label learning.

1. Introduction

Traditional machine learning applications have been derived from the analysis of data where the examples are associated with a single label [1]. However, recently studies over data that are structured in a more complex way than single-label data have received especial attention. Multi-label problems are concerned to those problems where the examples belong to a set of labels at the same time [2,3]. The goal of the Multi-Label Learning (MLL) paradigm is to learn a model that correctly generalises unseen multi-label data [2,3]. On the MLL context two problems are studied, multi-label classification (MLC) and label ranking (LR). MLC divides the set of labels into relevant and irrelevant sets, whereas the LR provides an ordering of the labels for a given query instance [3,4].

In the last few years, an increasing number of modern applications that contain multi-label data have appeared, such as text categorisation [5], emotions evoked by music [6], semantic annotation of images [7] and videos [8], classification of protein function [9] and gene [10,11].

Generally speaking, the multi-label datasets contain a large number of features that describe the instances, e.g. description of texts, images, proteins and genes [5,7–16]. The irrelevant, interacting, redundant and noisy features have a highly negative impact in the performance of the learning algorithms [17]. Moreover, the number of features is much bigger than the number of instances in several multi-label applications [13]. On the other hand, in some domains the number of possible labels can be in the region of hundreds (even thousands) and the distribution of instances per label can be showed in a non-uniform way [8,12,14–16]. Consequently, some multi-label learning algorithms present a poor performance with regard to time complexity and efficiency [4]. As a result of the above situations, nowadays the designing process of MLL algorithms faces several challenges and it is an open field of research.

The preprocessing techniques have demonstrated to be an important step of the knowledge discovery in databases [1,18]. Feature engineering techniques such as feature weighting (FW) and feature selection (FS) improve the performance of machine

[*] Corresponding author at: Department of Computer Science and Numerical Analysis, University of Córdoba, Spain, Tel.: +34 957212218; fax: +34 957218630.
E-mail addresses: oreyesp@facinf.uho.edu.cu (O. Reyes), cmorellp@uclv.edu.cu (C. Morell), sventura@uco.es (S. Ventura).

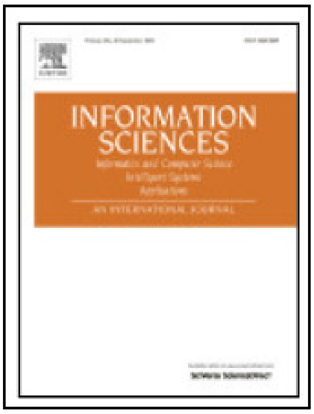
http://dx.doi.org/10.1016/j.neucom.2015.02.045
0925-2312/© 2015 Elsevier B.V. All rights reserved.

Title:

Effective lazy learning algorithm based on a data gravitation model for multi-label learning

Authors:

O. Reyes, C. Morell, and S. Ventura



Information Sciences, *Volume 340-341, pp. 159-174, 2016*

Ranking:

Impact factor (JCR 2015): 3.364

Knowledge area:

Computer Science, Information Systems: 8/143

DOI: 10.1016/j.ins.2016.01.006

Information Sciences 340–341 (2016) 159–174

Contents lists available at ScienceDirect

Information Sciences

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

Effective lazy learning algorithm based on a data gravitation model for multi-label learning

Oscar Reyes[a], Carlos Morell[b], Sebastián Ventura[c,d,*]

[a] Department of Computer Science, University of Holguín, Holguín, Cuba
[b] Department of Computer Science, Universidad Central de Las Villas, Villa Clara, Cuba
[c] Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain
[d] Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 14 April 2015
Revised 22 December 2015
Accepted 1 January 2016
Available online 8 January 2016

Keywords:
Multi-label learning
Data gravitation model
Lazy learning
Multi-label classification
Label ranking

ABSTRACT

In the last decade, an increasing number of real-world problems surrounding multi-label data have appeared, and multi-label learning has become an important area of research. The data gravitation model is an approach that applies the principles of the universal law of gravitation to resolve machine learning problems. One advantage of the data gravitation model, compared with other techniques, is that it is based on simple principles with high performance levels. This paper presents a multi-label lazy algorithm based on a data gravitation model, named MLDGC. MLDGC directly handles multi-label data, and considers each instance as an atomic data particle. The proposed multi-label lazy algorithm was evaluated and compared to several state-of-the-art multi-label lazy methods on 34 datasets. The results showed that our proposal outperformed state-of-the-art lazy methods. The experimental results were validated using non-parametric statistical tests, confirming the effectiveness of this data gravitation model for multi-label lazy learning.

1. Introduction

The study of problems where examples are simultaneously associated with a set of labels has received special attention. Problems where this type of data appears are known as multi-label problems. Particular problems involving multi-label data include text categorization [27,35,40,48], emotions evoked by music [32], semantic annotation of images [1,11,65], classification of music [54] and videos [3,26], classification of protein function [14,37] and gene function [8,31,69], acoustic classification [4], chemical data analysis [56] and many more.

Multi-label learning is with a form of learning method that deals with a model which correctly generalizes unseen multi-label data [21,51]. Two tasks have been studied concerning the question of multi-labels: multi-label classification and label ranking. Multi-label classification divides a set of labels into relevant and irrelevant sets, whereas the label ranking task establishes an order of the labels for a given test instance [34,51].

For more than a decade, a considerable number of multi-label learning algorithms have been proposed. These multi-label algorithms can be divided into problem transformation methods and algorithm adaptation methods [21]. The former transform multi-label problems into one or more single-label problems, in order for classic learning algorithms to be used.

[*] Corresponding author at: Department of Computer Science and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain. Tel.:+34 957212218; fax:+34 957218630.

E-mail addresses: ogreyesp@gmail.com (O. Reyes), cmorellp@uclv.edu.cu (C. Morell), sventura@uco.es (S. Ventura).

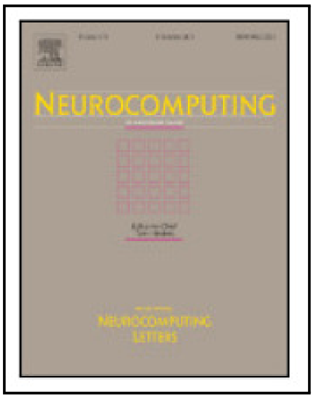
http://dx.doi.org/10.1016/j.ins.2016.01.006

Title:

Effective active learning strategy for multi-label learning

Authors:

O. Reyes, C. Morell, and S. Ventura



Neurocomputing, *submitted, 2015*

Ranking:

Impact factor (JCR 2015): 2.392

Knowledge area:

Computer Science, Artificial Intelligence: 31/130

Effective active learning strategy for multi-label learning

Oscar Reyes[a], Carlos Morell[b], Sebastián Ventura[c,d,*]

[a]Department of Computer Science, University of Holguín, Cuba
[b]Department of Computer Science, Universidad Central de Las Villas, Cuba
[c]Department of Computer Science and Numerical Analysis, University of Córdoba, Spain
[d]Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Data labelling is an expensive process that requires expert handling. In multi-label data, data labelling is further complicated owing to experts must label each example several times, as each example belongs to various categories. Active learning is concerned with learning accurate classifiers by choosing which examples will be labelled, reducing the labelling effort and the cost of training an accurate model. This paper presents a new active learning strategy for working on multi-label data. Two uncertainty measures based on the predictions of base classifier and the inconsistency of a predicted label set regarding the label dimension of the labelled dataset, respectively, are defined to select the most uncertain examples. The proposed strategy was evaluated and compared to several state-of-the-art strategies on 18 datasets. The experimental results were validated using non-parametric statistical tests and confirmed the effectiveness of the proposal for better multi-label active learning.

Keywords: Multi-label classification, label ranking, multi-label active learning, active learning strategy, pool-based scenario, rank aggregation problem

1. Introduction

In recent years, the study of problems that involve data associated with more than one label at the same time has attracted a great deal of attention. Particular multi-label problems include text categorization [1–3], classification of emotions evoked by music [4], semantic annotation of images [5–7], classification of music and videos [8–10], classification of protein and gene function [11–16], acoustic classification [17], chemical data analysis [18] and many more.

Multi-label learning is concerned with learning a model that correctly generalizes unseen multi-label data. In multi-label learning, two tasks have been studied [19–21]: Multi-label Classification and Label Ranking. Multi-label Classification task aims to find a model where, for a given test instance, the label space is divided into relevant and irrelevant label sets. On the other hand, Label Ranking task aims to provide, for a given test instance, a ranking of labels according to their relevance values. In the literature, is named Multi-label Ranking task [22] the generalization of Multi-label Classification and Label Ranking tasks. Multi-label Ranking aims to produce, at the same time, both a bipartition of label space and a consistent ranking of labels.

Most multi-label learning algorithms that have been proposed in the literature are designed for working on supervised learning environments, i.e. scenarios where all training instances are labelled. However, data labelling is a very expensive process that requires expert handling. In multi-label data, experts must label each example several times, as each example belongs to various categories. The situation is further complicated when a multi-label problem with a large number of examples and label classes is analyzed. Consequently, several real scenarios nowadays contain a small number of labelled data and a large number of unlabelled data simultaneously.

To date, there are two main areas that are concerned with learning models from labelled and unlabelled data, known as Semi-Supervised Learning [23] and Active Learning [24]. Active Learning is concerned with learning better classifiers by choosing which instances are labelled for training. Consequently, the labelling effort

*Corresponding author. Tel:+34957212218; fax:+34957218630.
Email addresses: oreyesp@facinf.uho.edu.cu (Oscar Reyes), cmorellp@uclv.edu.cu (Carlos Morell), sventura@uco.es (Sebastián Ventura)

Title:

Evolutionary Strategy to perform Batch-Mode Active Learning on Multi-Label Data

Authors:

O. Reyes and S. Ventura

ACM Transactions on

Intelligent Systems and Technology

Association for Computing Machinery

ACM Transactions on Intelligent Systems and Technology, *submitted, 2016*

Ranking:

Impact factor (JCR 2015): 2.414

Knowledge area:

Computer Science, Artificial Intelligence: 30/130

Computer Science, Information Systems: 20/143

Evolutionary Strategy to perform Batch-Mode Active Learning on Multi-Label Data

OSCAR REYES, University of Córdoba, Spain
SEBASTIÁN VENTURA, University of Córdoba, Spain, and King Abdulaziz University, Jeddah, Saudi Arabia

Multi-label learning has become an important area of research, owing to the increasing number of real-world problems that contain multi-label data. Data labeling is a very expensive process that requires expert handling. Consequently, numerous modern problems involve a small number of labeled examples and a large number of unlabeled examples simultaneously. Batch-mode active learning focusses on constructing accurate classifiers by means of choosing which instances will be labeled, in such a way that the selected instances are informative and the overlapping of information between them is minimal, reducing the labeling effort and the cost of training an accurate model. This paper presents a new strategy, named ESBMAL, to perform batch-mode active learning on multi-label data. ESBMAL formulates batch-mode active learning as a multi-objective problem and solves it by means of an evolutionary algorithm. Extensive experiments were conducted to validate the effectiveness of the proposal. The experimental results were validated using non-parametric statistical tests and confirmed the effectiveness of the proposal for better batch-mode multi-label active learning.

CCS Concepts: •**Theory of computation** → **Active learning;**

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Batch-mode multi-label active learning, multi-objective problem, genetic algorithm, multi-label classification, label ranking

ACM Reference Format:
Oscar Reyes and Sebastián Ventura. Evolutionary Strategy to perform Batch-Mode Active Learning on Multi-Label Data. *ACM Trans. Intell. Syst. Technol.* V, N, Article A (January YYYY), 20 pages.
DOI: -

1. INTRODUCTION

Multi-label problems concern problems where examples belong to multiple labels at the same time. The goal of the Multi-Label Learning paradigm is to develop a model that correctly generalizes unseen multi-label data. Multi-Label Classification and Label Ranking are two tasks that have been studied in the context of multi-label learning. The Multi-Label Classification (MLC) task aims to find a model where, for a given test instance, the labels are divided into relevant and irrelevant label sets. On the other hand, the Label Ranking (LR) task provides, for a given test instance, a permutation of the labels; the labels are ordered according to their relevance values. [Gibaja and Ventura 2014; Tsoumakas et al. 2010]

Particular real-world problems that involve multi-label data include text categorization [Katakis et al. 2008; Pestian et al. 2007], classification of emotions evoked by music [Li and Ogihara 2003], semantic annotation of images [Barnard et al. 2003], classi-

Author's addresses: O. Reyes, Department of Computer Science and Numerical Analysis, University of Córdoba, Spain; S. Ventura, Department of Computer Science and Numerical Analysis, University of Córdoba, Spain, and Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia.

DOI: -

Title:

JCLAL: A Java Framework for Active Learning

Authors:

O. Reyes, E. Pérez, M. C. Rodríguez Hernández, H. M. Fardoun and S. Ventura

Journal of Machine Learning Research, *Volume 17 (95), pp.1-5, 2016*

Ranking:

Impact factor (JCR 2015): 2.450

Knowledge area:

Computer Science, Artificial Intelligence: 29/130

JCLAL: A Java Framework for Active Learning

Oscar Reyes OGREYESP@GMAIL.COM
Eduardo Pérez EPEREZP@FACINF.UHO.EDU.CU
Department of Computer Science
University of Holguín
Holguín, Cuba

María del Carmen Rodríguez-Hernández 692383@UNIZAR.ES
Department of Computer Science and Systems Engineering
University of Zaragoza
Zaragoza, Spain

Habib M. Fardoun HFARDOUN@KAU.EDU.SA
Department of Information Systems
King Abdulaziz University
Jeddah, Saudi Arabia

Sebastián Ventura SVENTURA@UCO.ES
Department of Computer Science and Numerical Analysis
University of Córdoba
Córdoba, Spain
Department of Information Systems
King Abdulaziz University
Jeddah, Saudi Arabia

Editor: Geoff Holmes

Abstract

Active Learning has become an important area of research owing to the increasing number of real-world problems which contain labelled and unlabelled examples at the same time. JCLAL is a Java Class Library for Active Learning which has an architecture that follows strong principles of object-oriented design. It is easy to use, and it allows the developers to adapt, modify and extend the framework according to their needs. The library offers a variety of active learning methods that have been proposed in the literature. The software is available under the GPL license.

Keywords: active learning, framework, java language, object-oriented design

1. Introduction

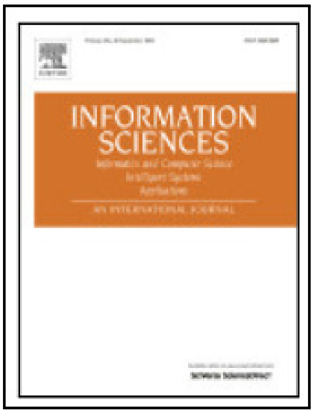
In the last decade, the study of problems which contain a small number of labelled examples and a large number of unlabelled examples at the same time have received special attention. Currently, there are two main areas that research the learning of models from labelled and unlabelled data, namely Semi-Supervised Learning and Active Learning (AL). AL is

Title:

Statistical Comparisons of Active Learning Strategies over Multiple Datasets

Authors:

O. Reyes, A. H. Altahi, and S. Ventura



INFORMATION SCIENCES

Informatics and Computer Science
Intelligent Systems
Applications

AN INTERNATIONAL JOURNAL

Information Sciences, *submitted, 2016*

Ranking:

Impact factor (JCR 2015): 3.364

Knowledge area:

Computer Science, Information Systems: 8/143

Statistical Comparisons of Active Learning Strategies over Multiple Datasets

Oscar Reyes[a], Abdulrahman H. Altahi[b], Sebastián Ventura[a,b,*]

[a] *Department of Computer Science and Numerical Analysis, University of Córdoba, Spain*
[b] *Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Active learning has become an important area of research owing to the increasing number of real-world problems where a huge amount of unlabeled data is available. Active learning strategies are commonly compared by means of visually comparing learning curves. However, in cases where several active learning strategies are tested on multiple datasets, the visual comparison of learning curves may not be the best choice to decide whether a strategy is significantly better than another one. In this paper, two approaches are proposed, based on the use of non-parametric statistical tests, to statistically compare active learning strategies over multiple datasets. The application of the two approaches is illustrated by means of an experimental study, demonstrating the usefulness of the proposal for improving analysis of active learning performance.

Keywords: Active learning, Non-parametric statistical test, Area under learning curve, Rate of performance change, Active learning iterations

1. Introduction

Machine learning aims to construct computational algorithms able to determine general patterns from available data. In the learning process, not all data are useful, because noisy, redundant and incomplete data can affect in many ways the performance of learning algorithms. Consequently, the acquisition of a high-quality and compact dataset (a.k.a. training set) from which a learning algorithm can determine useful patterns is very important [1].

Sample selection is an important preprocessing step in data mining. Sample selection aims to select a representative subset from the original dataset, in such a manner that the performance of the learner generated from the selected subset will be the same (even higher) as if the original dataset is used [2]. The main advantages in applying sample selection methods are as follows [1–4]: reduce storage requirements by means of removing redundant information present in datasets, reduce computation effort in the classification of new patterns, increment the performance of learning algorithms by means of removing noisy points and outliers, enable learning algorithms to work effectively with large-scale datasets, and reduce the labeling cost.

Sample selection methods can be roughly classified into two categories [1]: instance selection and active learning. Instance selection aims to condense a dataset by filtering noisy and redundant data. Instance selection methods can be categorized into two groups [5]: wrapper methods where the selection criterion is based on the accuracy obtained by a learner, and filter methods where the selection criterion is not based on the results of a learner.

On the other side, active learning aims to process incomplete data, referring to data with missing labels, by means of selecting instances from unlabeled datasets, reducing the labeling effort and cost of training an accurate learner [6, 7]. Nowadays, we find many modern problems where a huge amount of unlabeled data is available. Sometimes, the labeling process may be subject to little or no cost. However, for many supervised learning tasks, data labeling is a time-consuming process that requires expert handling [6]. Successful applications of active learning include text

[*] Corresponding author. Tel:+34957212218; fax:+34957218630.
Email addresses: ogreyesp@gmail.com (Oscar Reyes), ahaltalhi@kau.edu.sa (Abdulrahman H. Altahi), sventura@uco.es (Sebastián Ventura)

Conference publications

- O. Reyes, C. Morell and S. Ventura. ***Learning Similarity Metric to improve the performance of Lazy Multi-label Ranking Algorithms***. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA'2012). IEEE, pp. 246-251, 2012.

- O. Reyes, C. Morell and S. Ventura. ***ReliefF-ML: an extension of ReliefF algorithm to multi-label learning***. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. LNCS. Springer, vol. 8259, pp. 528-535, 2013.

- O. Reyes, C. Morell, and S. Ventura. ***Feature weighting on multi-label data through quadratic loss minimization***. In Congreso Internacional de Matemática y Computación, COMPUMAT-2013, Habana, Cuba, 2013.

- E. Pérez, O. Reyes and S. Ventura. ***Application of active learning in medical diagnosis***. In IV Encontro Regional de Computacão e Sistemas de Informacão, ENCOSIS-2015, Manaus/Amazonas, Brasil, 2015.

- O. Reyes and S. Ventura. ***Estrategia efectiva para el aprendizaje activo multi-etiqueta***. In XVII Conferencia de la Asociación Española para la Inteligencia Artificial, pp. 835-844, Salamanca, Spain, 2016.