Applied Soft Computing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

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

A spam filtering multi-objective optimization study covering parsimony maximization and three-way classification

Q1 Vitor Basto-Fernandes[a], Iryna Yevseyeva[b], José R. Méndez[c], Jiaqi Zhao[d], Florentino Fdez-Riverola[c,*], Michael T.M. Emmerich[e]

[a] School of Technology and Management, Computer Science and Communications Research Centre, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal
[b] School of Computer Science and Informatics, Faculty of Technology, De Montfort University, LE1 9BH Leicester, United Kingdom
[c] Informatics Engineering School, University of Vigo, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain
[d] Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xian Shaanxi Province 710071, China
[e] Leiden Institute of Advanced Computer Science, Faculty of Science. Leiden University, 2333-CA Leiden, the Netherlands, the Netherlands

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 17 June 2016
Accepted 27 June 2016
Available online xxx

Keywords:
Spam filtering
Multi-objective optimization
Parsimony
Three-way classification
Rule-based classifiers
SpamAssassin

ABSTRACT

Classifier performance optimization in machine learning can be stated as a multi-objective optimization problem. In this context, recent works have shown the utility of simple evolutionary multi-objective algorithms (NSGA-II, SPEA2) to conveniently optimize the global performance of different anti-spam filters. The present work extends existing contributions in the spam filtering domain by using three novel indicator-based (SMS-EMOA, CH-EMOA) and decomposition-based (MOEA/D) evolutionary multi-objective algorithms. The proposed approaches are used to optimize the performance of a heterogeneous ensemble of classifiers into two different but complementary scenarios: parsimony maximization and e-mail classification under low confidence level. Experimental results using a publicly available standard corpus allowed us to identify interesting conclusions regarding both the utility of rule-based classification filters and the appropriateness of a three-way classification system in the spam filtering domain.

1. Introduction

Nowadays, the use of Internet mailing services has become indispensable in the daily life of millions of users worldwide. Additionally, the combination of e-mail with the latest mobile always-connected smart-phones provides a simple but powerful method to stay in touch with other people and efficiently exchange documents at any time. As a result, both instant messaging (IM) applications and e-mail are commonly used for this purpose. However, a fundamental difference between popular IM applications (including Whatsapp or GTalk) and Internet mailing services is the existence of consent management methods, which can be found only among the former (e.g. blocking users, etc.). This situation has greatly facilitated the use of e-mails as an aggressive/massive advertisement method and virus distribution platform, originating the spam phenomenon.

Since the inception of spam, many companies and research teams have combined their efforts to fight against spam deliveries using different approaches and methods [1]. In this context, and from a scientific perspective, several machine learning (ML) algorithms have been successfully adapted and applied to filter spam messages, mainly including Naïve Bayes (NB) [2], ensemble techniques [3], Support Vector Machines (SVM) [4] and other memory-based systems [5]. Additionally, the computer security industry and the open source community also contributed with effective techniques such as DNS black and white lists [6,7], hashing schemes [8] and the development of SpamAssassin [9], the most popular filtering framework used to combine heterogeneous and complementary anti-spam techniques.

Q3 Since its creation, SpamAssassin has been widely used as the base of commercial products and filtering services including McAfee SpamKiller and Symantec Brightmail [10]. It allows system administrators to define specific filters using *ad hoc* rules. Each rule contains a logical expression (used as a trigger) and defines its associated score. Every time an e-mail is received for evaluation, SpamAssassin finds all the rules matching the target message

Q2 * Corresponding author at: ESEI: Escuela Superior de Ingeniería Informática Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004, Ourense, Spain.
E-mail addresses: vitor.basto.fernandes@ipleiria.pt (V. Basto-Fernandes), iryna.yevseyeva@dmu.ac.uk (I. Yevseyeva), moncho.mendez@uvigo.es (J.R. Méndez), jiaqizhao@stu.xidian.edu.cn (J. Zhao), riverola@uvigo.es (F. Fdez-Riverola), emmerich@liacs.nl (M.T.M. Emmerich).

http://dx.doi.org/10.1016/j.asoc.2016.06.043
1568-4946/

and computes the sum of their scores. This accumulative value is then compared with a configurable threshold (*required score*) to finally classify the new incoming message as spam or legitimate (also known as *ham*).

In order to define accurate anti-spam filters, the SpamAssassin framework provides implementations of several techniques including regular expressions, DNS black and white lists, Distributed Checksum Clearinghouses [11], Naïve Bayes [12,13], Sender Policy Framework [14], Hashcash [15], DomainKeys Identified Mail [16], language guessing [17], as well as several extra protocol error checks. Additionally, SpamAssassin allows the use of user-defined plugins to further extend the number of available techniques that compose a given filter.

Given the configurable structure of the SpamAssassin framework, and taking into consideration that the final accuracy of each user-defined filter strongly depends on the diversity of the underlying classifiers, the optimization of rule weights and other parameters governing the primary rule-based filtering process is still a challenge. In such a situation, initial approaches for the optimization of rule-based filters have been formulated as a single objective problem, where a general performance index (e.g., number of errors, kappa index or f-score, or Total Cost Ratio) is commonly used [18]. However, a more intuitive formulation of this problem involves several objectives. In fact, at least two complementary indexes should be simultaneously considered for minimization in the development of novel accurate anti-spam filters: (*i*) number of false negative (FN) errors (i.e., spam messages classified as legitimate) and (*ii*) number of false positives (FP) errors (i.e., legitimate messages classified as spam). Nevertheless, these objectives are in conflict, since minimizing the number of FP errors can be done only at the expense of increasing the number of spam messages going into e-mail boxes, and vice versa.

Single objective optimization approaches (also known as 'a priori' methods) require that sufficient preference information is expressed (blindly) before solution set is computed (i.e. assigning weights for the target objectives to aggregate objectives within a single objective which is optimized subsequently). In contrast, multi-objective optimization ('a posteriori') methods provide insights of conflicts between the objectives, i.e. at which extend one objective can be improved at the cost of other(s). Thus, user can select the resulting solution that best fits his/her preferences. In this context, some initial approaches [19,20,10] have evaluated the suitability of applying different multi-objective evolutionary algorithms (MOEA) in the spam filtering domain for optimizing both FN and FP errors at the same time. However, in the aforementioned studies only classical MOEA techniques (NSGA-II and SPEA2) were applied, and questions such as how to better adapt these algorithms using domain specific knowledge and how to consider other objective functions remained unanswered.

In this line, we carried out a preliminary study about the performance of several MOEA approaches to solve different optimization questions [21]. An extended version of our preliminary work, including only the study on parsimony but additional benchmarks, has very recently been published in [22]. The problem of three-way classification and the postprocessing of results on SPAM filters were however not addressed. In detail, this study included the spam filtering problem as a part of the MOEA benchmarking protocol with the goal of showing the insights of conflicts between those objectives to be minimized. However, our past work did not contribute a method to accurately evaluate the structure of the decision space (i.e., a detailed analysis of the relevance of each rule), which is essential for administrators to maintain (and continuously improve) filtering services.

Moreover, in order to fight against spam in environments where the cost associated to misclassification errors is high, the three-way classification scheme [23–25] emerged as a reliable way of mitigating information loss and security risks. Under this scheme, classifiers can avoid providing a solution in case there are not enough evidences to assign target instances to one of the two available classes (i.e., spam or legitimate). In such a situation, these messages are labelled as 'suspicious', 'doubtful' or 'borderline', being the subject of a further examination manually carried out by the final user. In this context, to increase security while revising suspicious e-mails, images, links and dangerous attachments should not be automatically loaded. As long as suspicious e-mails do not count as errors but are classified at the expense of increasing user efforts, the amount of messages labelled in this way should be also minimized (i.e., if all the messages belonging to a given corpus are labelled as 'suspicious', the number of misclassifications will be zero). Complementarily, the appropriateness of using a three-way classification scheme was also suggested as future work in our preliminary study [21].

In the present work, we complement previous findings by using three modern plus two classic MOEA approaches in two different ways: (*i*) by studying the structure of the decision space in the optimization of traditional binary classification processes (i.e., minimizing the amount of necessary rules and the number of FP and FN errors) and (*ii*) evaluating the suitability of three-way classification schemes to accurately filter spam contents. In the former case, we take advantage of the first optimization objective (parsimony) to specifically assess the contribution of each rule when generating a correct classification. In the second case, we carry out a performance study about the minimization of FP and FN errors when working with a three-way classification filter. These analyses have been implemented as two different optimization scenarios, making use of a well-known publicly available corpus.

While this section has introduced the motivation for this work, the rest of the paper is organized as follows: Section 2 presents the problem formulation, explains how to optimize ML classifiers with evolutionary algorithms (EAs) and summarizes previous works in anti-spam filter optimization using MOEA. Section 3 introduces the two case studies, defines the benchmarking protocol, establishes the performance metrics to be used and presents and discusses relevant issues regarding each case study. Finally, Section 4 provides conclusions and identifies future research work.

2. Materials and methods

In spite of the fast progress in computer technology and the constant increase of computational power, performing exhaustive searches in large continuous and combinatorial spaces is still challenging. In this context, the remarkable popularity of EAs over other optimization techniques is mainly motivated by their ability to search these spaces and find approximate (near) optimal solutions [26]. In the particular domain of multi-objective optimization, EAs stand for well-established computational methods where the population-based approach makes them suitable to search for approximation sets to the efficient set.

In this way, EAs were found to be particularly useful for dealing with multi-objective problems characterized by several conflicting goals, for which not simply a single optimum solution, but a set of Pareto optimal or non-dominated solutions need to be obtained. Together, these solutions represent the trade-offs between the existing objectives, being optimal in the sense of Pareto dominance. In such a situation, a Pareto optimal solution can only be improved in one objective at the expense of loss in other(s). As long as a population of possible solutions is used in parallel to solve these problems, the search is directed not a single optimum but towards multiple Pareto optimal solutions, which is the case of MOEAs (also known as Evolutionary Multi-objective Optimization Algorithms, EMOAs).

Please cite this article in press as: V. Basto-Fernandes, et al., A spam filtering multi-objective optimization study covering parsimony maximization and three-way classification, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2016.06.043