Exceptional Survival Model Mining

Juliana Barcellos Mattos[1(✉)], Eraylson G. Silva[1], Paulo S. G. de Mattos Neto[1], and Renato Vimieiro[2]

[1] Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil
julianabarcellosmattos@gmail.com, {jbm4,egs,psgmn}@cin.ufpe.br
[2] Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
rvimieiro@dcc.ufmg.br

Abstract. The development of treatments based on the patient's individual characteristics has been an emergent medical approach. The objective is to improve individual responses and overall survival. Thus, there is a need for computational tools able to identify and describe subgroups of patients for which the survival response significantly differs from the overall behaviour. However, there are few algorithms that address this matter. The majority of works of literature aim at building predictive models rather than understanding the characteristics that delineate subgroups with unusual survival. The approaches that provide understanding on factors that interfere in the survival behaviour usually resort to the stratification of the data based on previously known variable's interactions, lacking the ability to shed light into new, possibly unknown, interactions. In contrast to the existent predictive approaches, we propose the use of supervised descriptive pattern mining in order to discover local patterns able to describe subsets of patients that present unusual survival behaviour. In this paper, we present the ESM-AM (Exceptional Survival Model Ant Miner) algorithm, an Exceptional Model Mining approach to the discovery of subgroups with exceptional survival functions that explores the use of ant-colony optimization as search heuristic for the pattern mining task.

Keywords: Exceptional model mining · Survival analysis · Ant-colony optimization · Supervised descriptive pattern mining

1 Introduction

The recent development of large-scale databases and powerful methods for characterising patients is driving medicine towards more personalised (and less aggressive) individual interventions with the ultimate goal of improving the patients' prognosis and survival outcome. Many recent medical works strive for effective methods that can provide a better understanding on factors that interfere in survivability in order to subdivide populations of patients into more specific and uniform subgroups with relation to their survival behaviour. In this sense, many methods for Survival Analysis [10] have been developed over the

R. Cerri and R. C. Prati (Eds.): BRACIS 2020, LNAI 12320, pp. 307–321, 2020.
https://doi.org/10.1007/978-3-030-61380-8_21

years. For most applications, the goal is either to predict the time for the occurrence of a given event or to model the impact of covariates on survival response. Traditional statistical approaches [3] comprise parametric and semi-parametric techniques. The limitations resultant from their distributional and restrictive assumptions have motivated the development of new methods. Machine learning techniques [24] present the advantages of modelling non-linear relationships and deliver high-quality results. These techniques, however, comprise global predictive models usually hard to understand.

Rule induction is a traditional data mining technique that aims at extracting and enumerating patterns by learning sets of rules from a given data. The simplicity of rule models at the representation of understandable patterns gives this technique broad applicability in both predictive and descriptive data mining approaches. We can find in the literature a few rule-based works on survival analysis that strive to extract understandable knowledge from the data. Bazan *et al.* [2] aim at finding groups of patients with different survival estimates by applying rough sets theory to induce a set of decision rules from data. Pattaraintakorn and Cercone [19] employed a rough sets hybrid system to predict the survival time. Liu *et al.* [13] propose the characterisation of high-risk patients by applying the bump hunting method to generate rules. Kronek and Reddy [11] propose a methodology based on Logical Analysis of Data (LAD) for constructing survival patterns used to estimate the survival probability distribution of observations. Wróbel [25] proposes the use of a survival tree to generate a decision list of rules to predict the survival behaviour of new examples. Sikora *et al.* [23] use the survival time to classify patients into *dead* and *alive* classes, and employ a sequential covering strategy for inducing a set of classification rules in order to predict whether the patient would survive longer than a given time. A class of censored observations is used for a post-processing filtration of relevant rules. In [22], the authors handle censoring by proposing a weighting scheme that assigns censored observation to *dead* and *alive* classes with a given probability. Wróbel *et al.* [26] present the LR-Rules, a top-down greedy covering algorithm to induce accurate models for survival time prediction. In order to guide the rule induction process, the authors propose the use of a rule quality measure based on a statistical test that assesses the difference between the survival curve of the subset represented by the rule and its complement. The rules' consequent comprises a survival function fitted to the examples satisfying its premise.

Although the sets of rules deliver explanation over the data, the aforementioned rule-based approaches aim either to predict survival distribution or to classify new observations. Usually, the ones that strive to distinguish subgroups with different survival characteristics resort to the stratification of the time variable or impel the observations to fit predefined classes. This context motivates us to pose the following research question: *is there a more effective approach to characterise subgroups with unusual survival behaviour?*

In this sense, *supervised descriptive rule discovery* [16] is the induction of descriptive rules from labelled data, unifying both descriptive and predictive rule induction approaches. The main goal is, therefore, to understand the underlying

phenomena (according to a target) rather than to arbitrarily explain the data or classify new instances. One great advantage of this approach is the ability to discover local patterns. We call *Supervised Descriptive Pattern Mining* (SDPM) the set of tasks that use local patterns to provide any descriptive knowledge about a property of interest.

Subgroup Discovery (SD) [1,8] is one of the earliest SDPM tasks that aims at the discovery of local patterns that describe *interesting* subsets of the data. *Interestingness* is defined as distributional unusualness of a specific target attribute over the subset when compared to its distribution on the whole data (or on the subset's complement). However, one can understand that a deviating distribution of one target attribute does not encompass all forms of interestingness [5] and, thus, SD becomes impracticable in cases where a single target variable cannot express the property of interest. In this context, Exceptional Model Mining (EMM) [5,12] can be considered a multi-target generalisation of SD task, where the concept of the property of interest is extended to a target model. Given a model that best represents the data and thus constitutes the property of interest, the EMM task searches for subgroups of the data for which the model fitted to the subgroup differs substantially from the respective model fitted to the whole data (or to the subgroup's complement).

A perspective from EMM on survival analysis comprehends the data as a potential composition of different data subsets that present distinct survival behaviour. Hence, the EMM rule-based model comprises individual local patterns related to a target function instead of predictive models. Park *et al.* [17] propose an SD approach to analyse survival in breast cancer by presenting a tree-based rule induction approach that uses mean survival time as the target variable. High/low survival groups are derived from subgroups with average survival time significantly higher/smaller than the average for the remaining samples. However, the distributional unusualness of the mean survival time does not encompass all useful insights about the cohort's survival rate provided by a survival function. Although EMM poses a robust computational method for providing comprehensible identification of subgroups with exceptional survival response, there is still little research on this topic.

Finally, the combinatorial nature of rule induction processes poses a great challenge concerning computational cost. When focusing on EMM, these challenges can be even more significant, once the task usually deals with large sets of numerical data and induction of numerical models. Therefore, the strategy employed for traversing large search spaces is an essential issue for a good performance of the method. Apart from exhaustive algorithms, existing approaches on SD and EMM tasks mainly explore the use of greedy heuristic *Beam Search* [7] and of *Evolutionary Computing* [4,14,15,21]. However, to the best of our knowledge, there is no work on EMM exploring the use of bio-inspired meta-heuristic as a search strategy to the pattern discovery process.

In this paper, we address the problem of discovering subgroups of patients with unusual survival behaviour through the perspective of EMM in contrast to the majority of existent predictive approaches. Rather than strive to build

accurate models, we aim at building models capable of describing the local exceptionalities existent in the data. Hence, the main goal of this paper is to present the Exceptional Survival Model Ant Miner (ESM-AM) algorithm, an EMM framework designed for discovering subgroups with exceptional survival functions. ESM-AM relies on a measure of exceptionality between survival curves to guide the search for subgroups and returns a ruleset in which each rule describes a discovered subgroup. Differently from most EMM frameworks that employ greedy heuristics, our algorithm employs Ant-Colony Optimisation (ACO) as a search strategy for the rule induction process. We assess the performance of ESM-AM by comparing it to the covering greedy algorithm LR-Rules that, to the best of our knowledge, is the only work in literature for inducing rules based on the difference between survival functions. We evaluate the performance of our ACO-based approach in terms of the characteristics of the resultant rulesets, by comparing ESM-AM and LR-Rules results on 14 survival data sets. We also analyse some individual discovered rules in order to evaluate whether our EMM approach is capable of discovering exceptional subgroups and retrieving essential characteristics from the data. The remainder of this paper is organised as follows. Section 2 gives a brief review of Survival Analysis and introduces the main concepts of EMM framework. Section 3 describes the algorithm proposed in this paper, followed by Sect. 4 that presents the experiments and achieved results. Finally, in Sect. 5, we draw some conclusions and present directions to extend this proposal.

2 Background

2.1 Survival Analysis (SA)

Survival Analysis [10] is the collection of methods and techniques designed to analyse data in which the target variable is the *time* until a given *event* occurs. For *event* one can understand any designated experience of interest that may happen to a subject under study, e.g. a patients' death, relapse from remission, or any other experience. The *time* variable – also referred to as *survival time* – represents the time since the beginning of the study until the occurrence of the event of interest. The main characteristic of survival data that differentiate it from regression problems is the existence of observations for which the survival times are unknown; this phenomenon is called *censoring*. The most common type of censoring happens when an individual's survival time is greater than the time observed during the study – usually because the subject did not suffer the event yet when the study ended or because the subject left the study due to any reason other than the event occurrence. In these cases, we say that the observation is *right-censored*.

Let $(A, T, δ) be a survival data set with |$| observations (instances). Each observation in $ is characterized by a set of descriptive attributes $A = \{A_1, A_2, \ldots, A_{|A|}\}$, a survival time T and a censoring status δ, that indicates whether the subject is censored ($\delta = 0$) or has experienced the event ($\delta = 1$).

Therefore, the ith observation in the data set can be represented as a vector $o_i = (A_{1_i}, A_{2_i}, \ldots, A_{|A|_i}, T_i, \delta_i)$.

In general, survival data is modelled in terms of two functions, namely survival and hazard. The *survival function* $S(t)$ indicates the probability that an individual o_i survives up to a specified future time t, i.e., $S(t) = P(T_i > t)$. The initial value, $S(t = 0) = 1$, represents the fact that no observation has yet suffered the event at the beginning of the study, and thus, the probability of surviving past the initial time is one. The *hazard function* – also referred as conditional failure rate – is usually denoted by $h(t)$ and provides the probability that an individual o_i experiences the event at a certain time t given that it has survived up to that time. In other words, $h(t)$ is a measure of instantaneous failure potential and is the mechanism used at the mathematical modelling of survival data in a large number of survival analysis approaches.

The survival function is usually estimated by the non-parametric Kaplan-Meier (KM) survival estimate [9]. Considering $\mathbb{S}$, we define $\mathcal{T} = \{t_1, t_2, \ldots, t_k | t \in T, k \leq |\mathbb{S}|\}$ the set of unique ordered survival times of $\mathbb{S}$. The estimated probability $\hat{S}(t_j)$ of surviving past a time $t_j \in \mathcal{T}$ is given by Eq. 1:

$$\hat{S}(t_j) = \left(\prod_{\forall t_i \in \mathcal{T}}^{t_{j-1}} \hat{P}(\mathcal{T} > t_i | \mathcal{T} \geq t_i) \right) \cdot \hat{P}(\mathcal{T} > t_j | \mathcal{T} \geq t_j) \equiv \hat{S}(t_{j-1}) \left(1 - \frac{d_j}{r_j}\right) \quad (1)$$

where $\hat{S}(t_{j-1})$ is the probability of being alive at the time interval $[t_{j-1}, t_j)$, r_j is the number of patients alive (at risk) just before t_j, and d_j the number of events that happened at the time interval $[t_j, t_{j+1})$; for $t_0 = 0$, $S(t_0) = 1$. The KM survival curve is the plot of the KM survival probabilities against time.

For comparing the survival experience of different groups, the *logrank* [20] statistical test is the most widely used method. It tests the null hypothesis that there is no overall difference between the KM curves of the groups, making use of the events observed within each group versus the number of events that are expected to happen.

Let $G \subseteq \mathbb{S}$ be a group of patients. We define r_j^G the number of patients in G that are at risk just before $t_j \in \mathcal{T}$. The logrank test assumes that the number of events that are expected to happen within a group is proportional to the extent of its risk, i.e. to the proportion r_j^G / r_j. Hence, the number E^G of expected events suffered by G over $\mathcal{T}$ is given by Eq. 2.

$$E^G = \sum_{\forall t_j \in \mathcal{T}} \frac{r_j^G}{r_j} \times d_j \quad (2)$$

For comparing G to its complement $\overline{G} = \mathbb{S} \setminus G$, the logrank test $X^2 \sim \chi_1^2$ is given by Eq. 3, where $O^G (O^{\overline{G}})$ is the number of observed events in $G(\overline{G})$.

$$X^2 = \frac{(O^G - E^G)^2}{E^G} + \frac{(O^{\overline{G}} - E^{\overline{G}})^2}{E^{\overline{G}}} \quad (3)$$

2.2 Exceptional Model Mining (EMM)

The EMM task aims at discovering subgroups that are exceptional in relation to a property of interest. Given a survival data set $\mathbb{S}(\mathrm{A}, \mathrm{T}, \delta)$, let A be the set of descriptive attributes used to define subgroups and let $Y = \{T, \delta\}$ be the set of target attributes used to evaluate the described subgroups. Our property of interest is, therefore, defined as the target model $f(Y) = \hat{S}(t)$. The *patterns*, or *descriptions*, are usually taken from a *description language* $\mathcal{D}$ of free choice within EMM design. Here we define a pattern $P \in \mathcal{D}$ as a rule representation in the form

$$P : \quad IF \quad c^1 \ AND \ c^2 \ AND \ \ldots \ AND \ c^L \quad THEN \ \hat{S}_P(t)$$

where the *conditions* c^l are of the form $a_i = v_{ij}$ for $a_i \in A$ and $v_{ij} \in Domain(a_i)$. We call $L \leq |A|$ the rule's *length*. The consequent $\hat{S}_P(t)$ of the rule comprises the KM model fitted for the observations that satisfy its antecedent.

A *subgroup* corresponding to a pattern P is the set of observations $G_P \subseteq \Omega$ of size $|G_P|$ that are covered by P, i.e. that satisfy its antecedent. The *coverage* of a subgroup is given by the number of observations that it covers. One can understand that *rules* are representations of *patterns*/*descriptions*, being those three terms equivalent and, thus, implying *subgroups*. The *complement* of a subgroup G_P is the set of observations $\overline{G}_P = \Omega \setminus G_P$ of size $|\Omega| - |G_P|$.

For each subgroup G_P under evaluation, the target model $\hat{S}(t)$ is induced on the set of observations $o_i \in G_P$. Then, the subgroup is evaluated with a *quality measure* $\varphi : \mathcal{D} \rightarrow \mathbb{R}$ that quantifies the difference between the target model fitted on G_P and the target model fitted on $\overline{G}_P$. Here we define the quality measure $\varphi = (1 - p\text{-}value)$ based on the logrank statistical test, assuming values in $(0, 1)$ interval. The smaller the logrank's p-value, the more exceptional is the KM model fitted to the subgroup, and thus the higher is its quality.

3 EMS-AM: Exceptional Survival Model Ant Miner

EMS-AM algorithm is an adaptation of the well-known classification rule induction algorithm Ant-Miner [18]. We adapted the Ant Colony Optimization heuristic to discover subgroups with exceptional KM curves. ESM-AM returns a list of discovered rules of the aforementioned form, and is presented in Algorithm 1.

The algorithm is initialized with an empty list of rules and with a set of uncovered cases comprising all cases in the data set and then it follows a covering-based approach. In each iteration (lines 4–26), a colony of ants constructs a number of rules R_t. Then the best rule R_{best} – according to φ – is selected to be added to the list of discovered rules, and the examples covered by R_{best} are removed from the set of uncovered cases. This process is repeated while the number of remaining uncovered observations do not achieve a maximum threshold or until a maximum number of iterations is reached. Non-significant rules are discarded at a level of significance of α. In case the ant colony is no longer able to discover significant rules, the algorithm is finalized, and the list of rules is returned.

Algorithm 1: High-Level description of ESM-AM algorithm

```
1: uncoveredCases = {all data set cases}
2: DiscoveredRuleList = [ ]
3: it = 0
4: while (uncoveredCases > max_uncovered_cases) or (it < no_of_ants) do
5:    t , j = 1 # ant, convergence indexes
6:    Initialize all trails with the same amount of pheromone
7:    repeat
8:       R_t = ConstructRule(dataset, min_cases_per_rule)
9:       PruneRule(R_t);
10:      PheromoneUpdating(R_t);
11:      if (R_t = R_{t-1}) then
12:         j += 1
13:      else
14:         j = 1;
15:      end if
16:      t += 1;
17:   until (t ≥ no_of_ants) or (j ≥ no_rules_converg)
18:   Choose the best rule R_best;
19:   if quality(R_best) ≥ (1 − α) then
20:      DiscoveredRuleList = DiscoveredRuleList ∪ R_best
21:      uncoveredCases = uncoveredCases \ {examples covered by R_best}
22:   else
23:      break
24:   end if
25:   it += 1
26: end while
27: return: DiscoveredRuleList
```

For the rule induction process (lines 7–17), the ants in the colony generate complete rules following two procedures: rule construction (line 8) and pruning (line 9). Each ant t constructs a rule R_t in a general-to-specific approach by iteratively adding conditions to a initially empty rule until there is no more conditions to be added or until the addition of a condition results in a rule coverage bellow a threshold. The probabilistic choice of a condition $c_{ij} : a_i = v_{ij}$ to be added to the current partial rule depends on both the heuristic function (η) and the pheromone (τ) associated with each condition. It is determined by the probability P_{ij} given in Eq. 4, where $|A|$ is the size of the attribute set, $|D_i|$ is the size of the attribute $a_i \in A$ domain, $x_i = 1$ if the a_i was not yet added to the current rule, or $x_i = 0$ otherwise.

$$P_{ij} = \frac{\eta_{ij} \cdot \tau_{ij}(t)}{\sum_{i=1}^{|A|} x_i \cdot \sum_{j=1}^{|D_i|} (\eta_{ij} \cdot \tau_{ij}(t))} \tag{4}$$

The heuristic function η_{ij} associated with c_{ij} is based on Shannon's entropy (Eq. 5). We considered an initial partition of the observations as those with

survival time at least as long as the cohort's average survival time, and those with shorter survival time. The quality of a condition is then the normalized information gain, obtained by further partitioning observations based on it. The class entropy was computed inducing a partition on the observations according to a condition. The heuristic function is given in Eq. 6.

$$H(W|a_i = v_{ij}) = -\sum_{w=1}^{k} P(w|a_i = v_{ij}) \cdot \log_2 P(w|a_i = v_{ij}) \tag{5}$$

$$\eta_{ij} = \frac{\log_2 k - H(W|a_i = v_{ij})}{\sum_{i=1}^{|A|} x_i \cdot \sum_{j=1}^{|D_i|} \log_2 k - H(W|a_i = v_{ij})} \tag{6}$$

After the ant constructs a complete rule R_t, the quality of the discovered subgroup (rule) is computed on the basis of the logrank test as $\varphi = (1 - p\text{-}value)$. Then, a rule pruning procedure iteratively removes conditions from the rule's antecedent, each time eliminating the condition c_{ij} that leads to the largest improvement in the rule quality φ. The pruning stops when no conditions can be removed without decreasing the rule's quality, or when the rule contains only one condition.

Lastly, the pheromone updating process increments the amount of pheromone associated to the conditions used on the pruned R_t according to Eq. 7, where $\tau_{ij}(t+1)$ is the amount of pheromone associated to c_{ij} for the next ant iteration. For the conditions not used on R_t, evaporation process is simulated by the normalization of τ values in $(t + 1)$.

$$\tau_{ij}(t + 1) = \tau_{ij}(t) + \varphi \cdot \tau_{ij}(t) \tag{7}$$

This ant-based rule induction process is repeated until all ants in the colony have constructed their rules, or until the ants converge to a single rule – according to a threshold of identical sequential rules. For each new algorithm iteration (line 4), a new colony is created and all available conditions receive the same initial amount of pheromone.

Finally, the ESM-AM has five user-defined parameters: (1) *no_of_ants* defines the size of the ant colony; (2) *max_uncovered_cases* defines the maximum number of remaining uncovered cases; (3) *no_rules_converg* defines the number of sequential identical constructed rules R_t in order to consider that the ants have converged to a solution (rule); (4) *min_cases_per_rule* is the minimum required rule coverage; and (5) α defines the level of significance of the logrank test.

4 Experiments and Results

4.1 Experimental Setup

The performance of the ESM-AM algorithm was evaluated on 14 real-world data sets listed in Table 1. All data sets were processed by removing observations containing missing values and by filtering the features. Table 2 presents the

Table 1. Characteristics of 14 datasets used in the experimental study: the number of observations (#obs), the number of descriptive attributes (#att), the number of discretized descriptors (#disc), the percentage of censored observations (%cens), and the survival event description (Event)

Dataset	#obs	#attr	#disc	%cens	Subject of research	Event
actg320	1151	11	3	91.66	HIV-infected patients	AIDS diagnosis/death
breast-cancer	196	80	78	73.98	Node-Negative breast cancer	distant metastasis
cancer	168	7	5	27.98	Advanced lung cancer	death
carcinoma	193	8	1	27.46	Carcinoma of the oropharynx	death
gbsg2	686	8	5	56.41	Breast cancer	recurrence
lung	901	8	0	37.40	Early lung cancer	death
melanoma	205	5	3	72.20	Malignant melanoma	death
mgus	176	8	6	6.25	Monoclonal gammopathy	death
mgus2	1338	7	5	29.90	Monoclonal gammopathy	death
pbc	276	17	10	59.78	Primary biliary cirrhosis	death
ptc	309	18	1	93.53	Papillary thyroid carcinoma	recurrence/progression
uis	575	9	4	19.30	Drug addiction treatment	return to drug use
veteran	137	6	3	6.57	Lung cancer	death
whas500	500	14	6	57.00	Worcester Heart Attack	death

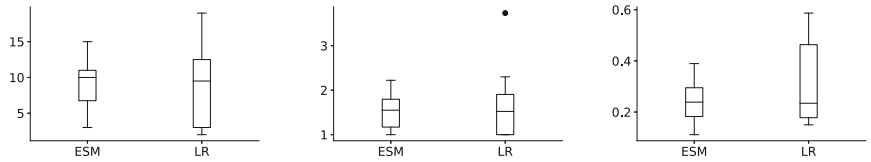
Table 2. Selected descriptive attributes of the data sets with feature selection

Dataset	Descriptive attributes
actg320	tx, txgrp, strat2, sex, raceth, ivdrug, hemophil, karnof, cd4, priorzdv, age
mgus	age, sex, dxyr, pcdx, alb, creat, hgb, mspike
ptc	risk_group, histological_type, age, sex, path_t_stage, path_n_stage, path_m_stage, tumor_status, exome, extrathyroidal_extension, mrna_cluster, mirna_cluster, arm_scna_cluster, methylation_cluster, disease_stage, primary_exome, lowpass, wgs_status
whas500	age, gender, hr, sysbp, diasbp, bmi, cvd, afb, sho, chf, av3, miord, mitype, los

descriptive attributes for the sets that were filtered. As the algorithm only copes with categorical attributes, all numerical variables were discretized with K-Means into five interval categories. The five ESM-AM user-defined parameters were set as follows: $no\_of\_ants = 3000$; $max\_uncovered\_cases = 0$; $no\_rules\_converg = 10$; $min\_cases\_per\_rule = 10$; and $\alpha = 0.05$.

In order to evaluate the models generated by our proposed ACO heuristic approach, ESM-AM algorithm was compared with the LR-Rules [26], a greedy covering algorithm. We used the LR-Rules default parameters defined in the available[1] implementation. The rule models resultant from both algorithms were evaluated according to the following characteristics: number of discovered rules;

[1] LR-Rules algorithm: https://github.com/adaa-polsl/LR-Rules/releases.



(a) #rules (b) length (c) ruleCov

Fig. 1. Boxplots of the ESM-AM (ESM) and LR-Rules (LR) algorithms results for: (a) the number of discovered rules; (b) the average rule length; and (c) the average relative rule coverage

average rule length; average relative rule coverage; rule set coverage; and integrated Brier score (IBS). The Brier score (BS) [6] measures the square difference between an observation survival status δ_i and its estimated survival probability $\hat{S}(t)$, in a given time T^*. The BS value for an observation o_i (incorporating censoring) is given by Eq. 8. The IBS is the Brier score integrated over all survival times T for all n observations, and is given by Eq. 9, where $\hat{G}(t)$ is the KM estimate of the censoring distribution, obtained from estimating the survival function for $\delta = (1 - \delta)$. The IBS was calculated for each discovered rule and summed over the entire ruleset.

$$BS_i(T^*) = \begin{cases} \frac{1}{\hat{G}(T_i)}[0 - S(T^*)]^2 & \text{if } T_i \leq T^*, \delta_i = 1 \\ \frac{1}{\hat{G}(T^*)}[1 - S(T^*)]^2 & \text{if } T_i > T^* \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$IBS = \frac{1}{max(T)} \int_0^{max(T)} \left(\frac{1}{n} \sum_{i=1}^{n} BS_i(T^*) \right) dT^* \tag{9}$$

4.2 Results Analysis

The results for both ESM-AM and LR-Rules algorithms on each data set are presented in Table 3. All figures and additional results, as well as the data sets used in the experiments, are available on ESM-AM website[2].

The ESM-AM algorithm returned rule models with, on average, 9.43 rules of size 1.52 (condition), compared to the LR-Rules' average of 8.93 discovered rules of size 1.63. We notice then that ESM-AM was able of generating compact models concerning both the size of the ruleset and the length of the rules. The coverage of ESM-AM rules was, on average, 25% of the total cases in the data sets, comprising rules that neither cover the majority of the cases nor very small groups. Figure 1 shows the boxplots of the performance of both algorithms with relation to the number of discovered rules, rule length and rule coverage. We

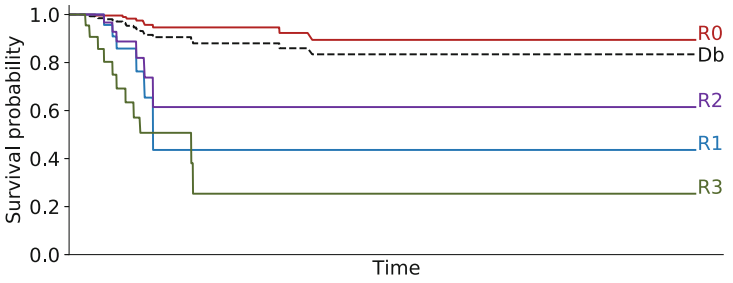
[2] ESM-AM algorithm website: https://github.com/jbmattos/ESM-AM_bracis2020.

also notice that comparing to LR-Rules, ESM-AM results presented smaller variability. When evaluating the coverage of the final rulesets (*setCov*), ESM-AM showed greater variability, presenting in some cases, a higher percentage of observations that remained not covered by any rule. For the IBS results, ESM-AM algorithm presented an average of 0.15 comparing to 0.18 presented by LR-Rules. One could understand the IBS as a measure of the quadratic error between the survival estimates of the observations covered by a rule and their true survival status. Therefore, we notice that the ESM-AM algorithm was able to discover more homogeneous subgroups concerning survival response. Finally, for a level of significance of 5%, the Wilcoxon test showed statistically significant difference between ESM-AM and LR-Rules performances only in terms of the *setCov* criterion (p-value = 0.036).

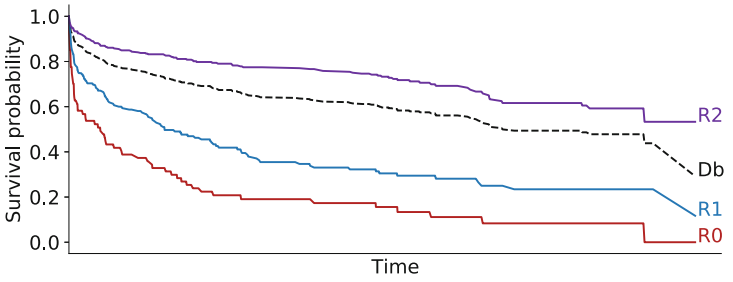
Besides, to evaluate ESM-AM final models in terms of the subgroups discovered by our EMM framework, we assess whether the induced rules present statistically significant survival functions. Figure 2 presents the KM curves for the rulesets induced on *ptc* and *whas500* data sets. The plots additionally include the cohort's KM curve, given by the *Db* curve. It is possible to observe the significant difference between the survival curve of the study cohort in comparison to the curves induced over the subgroups discovered by ESM-AM, indicating that the algorithm is able to identify local patterns with significant distinct survival response. In a more detailed analysis of the individual discovered rules, we found that the algorithm was able to retrieve information on attributes that stratify the data into different survival experiences.

Table 3. Characteristics on the resultant rule models for ESM-AM (ESM) and LR-Rules (LR) algorithms: the number of discovered rules (#rules), the average rule length (length), the average relative rule coverage (cov±std, ruleCov), the rule set coverage(setCov), and integrated Brier score on the rule set (IBS). Bold values represent the best results.

Metrics	#rules		length		ruleCov		setCov		IBS	
Algorithms	ESM	LR	ESM	LR	ESM	LR	ESM	LR	ESM	LR
actg320	**9**	15	**2.22**	3.73	0.26 ± 0.16	0.15 ± 0.16	1.00	1.00	**0.43**	0.45
breast-cancer	**11**	19	**1.18**	1.95	0.24 ± 0.15	0.17 ± 0.10	0.94	**0.98**	0.01	0.01
cancer	11	**9**	2.00	**1.78**	0.17 ± 0.16	0.25 ± 0.16	0.74	**1.00**	0.08	**0.07**
carcinoma	10	**3**	1.80	**1.00**	0.27 ± 0.20	0.33 ± 0.30	0.99	0.99	0.06	**0.02**
gbsg2	14	**10**	**1.79**	2.30	0.19 ± 0.24	0.22 ± 0.23	1.00	1.00	0.13	**0.11**
lung	9	**7**	**1.00**	1.14	0.35 ± 0.13	0.35 ± 0.14	1.00	1.00	0.12	**0.09**
melanoma	6	**2**	1.00	1.00	0.39 ± 0.17	0.50 ± 0.06	1.00	1.00	0.02	**0.01**
mgus	13	**11**	**1.62**	1.73	0.11 ± 0.08	0.18 ± 0.11	0.71	**1.00**	0.01	0.01
mgus2	**6**	18	**1.17**	1.50	0.18 ± 0.09	0.17 ± 0.09	0.70	**1.00**	**0.38**	1.25
pbc	11	**3**	1.36	**1.00**	0.20 ± 0.28	0.59 ± 0.37	1.00	1.00	0.03	**0.02**
ptc	4	**2**	1.50	**1.00**	0.30 ± 0.36	0.50 ± 0.42	1.00	1.00	0.33	**0.08**
uis	15	**13**	**2.00**	2.08	0.23 ± 0.18	0.22 ± 0.17	0.99	**1.00**	0.36	**0.27**
veteran	**10**	11	1.60	**1.55**	0.18 ± 0.08	0.18 ± 0.08	0.84	**1.00**	0.16	**0.09**
whas500	3	**2**	1.00	1.00	0.38 ± 0.23	0.50 ± 0.19	1.00	1.00	0.04	**0.03**

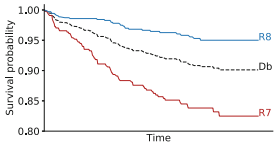


(a) ptc

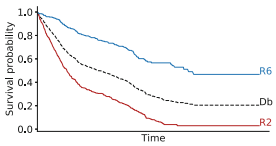


(b) whas500

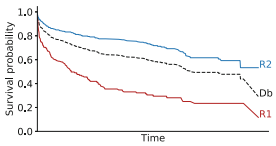
Fig. 2. Analysis of the discovered rules for *ptc* (a) and *whas500* (b) data sets; the *Db* curves represent the KM estimates on the entire cohort



(a) actg320



(b) lung



(c) whas500

Fig. 3. Analysis of individual rules induced for *actg320* (a), *lung* (b) and *whas500* (c) datasets; the *Db* curves represent the KM estimates of the study cohort

In the *actg320* data set, the *strat2* variable represents the counting of cells with expression of the CD4 protein, dividing the observations into low/high ($strat2 = 0$ /$strat2 = 1$) counting – where a low counting imply a higher risk for the patient. Among the nine resultant rules induced on this data set, the algorithm recovered such information presenting the following two rules: **R7**: $\{strat2 = 0\}$ and **R8**: $\{strat2 = 1\}$. Figure 3a presents the KM plot of both rules reflecting the expected survival behaviour.

In the *lung* data set, the $stage1 = \{1, 2, 3\}$ variable reflects the overall stage of lung cancer, for $stage1 = 1$ earlier than $stage1 = 3$. For the rule set induced on this data set, the ESM-AM algorithm returned also nine rules, two of them:

R2: $\{stage1 = 3\}$ and **R6**: $\{stage1 = 1\}$. Figure 3b presents the KM curves for both rules, showing that the survivability is better for early lung cancer stage.

In the *whas500* data set, the *chf* variable stands for *congestive heart complications*, dividing the observations into a group of patients that present complications and the ones that do not. The ESM-AM algorithm returned a rule set comprising three rules, two of them: **R1**: $\{chf = True\}$ and **R2**: $\{chf = False\}$. Figure 3c present the plot of both rules, showing that the presence of heart complications decreases the chances of survival.

5 Conclusions

In this paper, we introduce a novel approach to the discovery of subgroups with unusual survival behaviour based on supervised descriptive pattern mining, in contrast to the predictive approaches existent in literature. We presented the ESM-AM (Exceptional Survival Model Ant Miner) algorithm, an EMM framework that uses ACO meta-heuristic for the rule induction process. The algorithm returns a rule set where each rule can be understood as the description of a subgroup that is exceptional with relation to its survival function. Our proposed algorithm is the first approach for EMM task that explores a bio-inspired meta-heuristic as search heuristic.

We evaluated our proposal assessing its capability of returning accurate rule models and of discovering interesting subgroups. Therefore, we tested our ACO-based heuristic approach to the discovery of local survival exceptionalities on 14 data sets. The performance of ESM-AM was evaluated in comparison to the LR-Rules algorithm – a greedy covering rule induction algorithm for survival data analysis. Our approach achieved competitive results concerning characteristics of the rulesets, performing similarly to LR-Rules with relation to model size, rule length, rule coverage and IBS. When comparing to LR-Rules, ESM-AM algorithm returned compact rule models and lower IBS, i.e. a smaller difference between a subgroup's survival function and the true survival experience of the observations that it comprises. The low IBS also indicates that the rules discovered by the ESM-AM comprise homogeneous subgroups with respect to survival behaviour. When assessing the algorithm's capability of discovering local survival behaviour exceptionalities, we notice that the ESM-AM was able to discover significant subgroups and to identify data characteristics that interfere in survival experience.

Finally, there are different directions to expand this study: (1) cope with numerical attributes; (2) investigate other quality measures for the subgroups, in other to consider not only its exceptionality but also its coverage, according to the definitions of EMM framework; (3) investigate new heuristic functions and new pheromone updating procedures for the ACO meta-heuristic in order to better capture survival relations on the induction process; (4) tackle problems such as pattern's redundancy, high-dimensionality and false statistical discoveries; and (5) expand the results' analysis with further experimental statistical procedures and more detailed exploratory data analysis.

Acknowledgment. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by the National Council for Scientific and Technological Development – CNPq.

References

1. Atzmueller, M.: Subgroup discovery. Wiley Interdiscip. Rev.: Data Mining Knowl. Discov. **5**, 35–49 (2015)
2. Bazan, J., Osmólski, A., Skowron, A., Ślçezak, D., Szczuka, M., Wróblewski, J.: Rough set approach to the survival analysis. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 522–529. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45813-1_69
3. Bradburn, M.J., Clark, T.G., Love, S., Altman, D.: Survival analysis part ii: multivariate data analysis-an introduction to concepts and methods. Br. J. Cancer **89**(3), 431 (2003)
4. Carmona, C.J., González, P., del Jesus, M.J., Herrera, F.: Overview on evolutionary subgroup discovery: analysis of the suitability and potential of the search performed by evolutionary algorithms. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. **4**(2), 87–103 (2014)
5. Duivesteijn, W., Feelders, A.J., Knobbe, A.: Exceptional model mining. Data Min. Knowl. Discov. **30**(1), 47–98 (2016). https://doi.org/10.1007/s10618-015-0403-4
6. Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M.: Assessment and comparison of prognostic classification schemes for survival data. Stat. Med. **18**(17–18), 2529–2545 (1999)
7. Helal, S.: Subgroup discovery algorithms: a survey and empirical evaluation. J. Comput. Sci. Technol. **31**(3), 561–576 (2016). https://doi.org/10.1007/s11390-016-1647-1
8. Herrera, F., Carmona, C.J., González, P., Del Jesus, M.J.: An overview on subgroup discovery: foundations and applications. Knowl. Inf. Syst. **29**(3), 495–525 (2011). https://doi.org/10.1007/s10115-010-0356-2
9. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**(282), 457–481 (1958)
10. Kleinbaum, D.G.: Survival analysis, a self-learning text. Biomet. J.: J. Math. Methods Biosci. **40**(1), 107–108 (1998)
11. Kronek, L.P., Reddy, A.: Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data. Bioinformatics **24**(16), i248–i253 (2008)
12. Leman, D., Feelders, A., Knobbe, A.: Exceptional model mining. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008. LNCS (LNAI), vol. 5212, pp. 1–16. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87481-2_1
13. Liu, X., Minin, V., Huang, Y., Seligson, D.B., Horvath, S.: Statistical methods for analyzing tissue microarray data. J. Biopharm. Stat. **14**(3), 671–685 (2004)
14. Lucas, T., Silva, T.C., Vimieiro, R., Ludermir, T.B.: A new evolutionary algorithm for mining top-k discriminative patterns in high dimensional data. Appl. Soft Comput. **59**, 487–499 (2017)
15. Lucas, T., Vimieiro, R., Ludermir, T.: SSDP+: a diverse and more informative subgroup discovery approach for high dimensional data. In: 2018 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. IEEE (2018)

16. Novak, P.K., Lavrač, N., Webb, G.I.: Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining. J. Mach. Learn. Res. **10**(Feb), 377–403 (2009)
17. Park, J.V., Park, S.J., Yoo, J.S.: Finding characteristics of exceptional breast cancer subpopulations using subgroup mining and statistical test. Expert Syst. Appl. **118**, 553–562 (2019)
18. Parpinelli, R.S., Lopes, H.S., Freitas, A.A.: Data mining with an ant colony optimization algorithm. IEEE Trans. Evol. Comput. **6**(4), 321–332 (2002)
19. Pattaraintakorn, P., Cercone, N.: A foundation of rough sets theoretical and computational hybrid intelligent system for survival analysis. Comput. Math. Appl. **56**(7), 1699–1708 (2008)
20. Peto, R., et al.: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. ii. Analysis and examples. Br. J. Cancer **35**(1), 1 (1977)
21. Pontes, T., Vimieiro, R., Ludermir, T.B.: SSDP: a simple evolutionary approach for top-k discriminative patterns in high dimensional databases. In: 2016 5th Brazilian Conference on Intelligent Systems (BRACIS), pp. 361–366. IEEE (2016)
22. Sikora, M., et al.: Censoring weighted separate-and-conquer rule induction from survival data. Methods Inf. Med. **53**(02), 137–148 (2014)
23. Sikora, M., Mielcarek, M., Kałwak, K., et al.: Application of rule induction to discover survival factors of patients after bone marrow transplantation. J. Med. Inform. Technol. **22**, 35–53 (2013)
24. Wang, P., Li, Y., Reddy, C.K.: Machine learning for survival analysis: a survey. ACM Comput. Surv. (CSUR) **51**(6), 110 (2019)
25. Wróbel, Ł.: Tree-based induction of decision list from survival data. J. Med. Inform. Technol. **20**, 73–78 (2012). http://jmit.us.edu.pl/cms/index.php?page=vol-20-2012
26. Wróbel, Ł., Gudyś, A., Sikora, M.: Learning rule sets from survival data. BMC Bioinform. **18**(1), 285 (2017). https://doi.org/10.1186/s12859-017-1693-x