Predicting Wildfires

Propositional and Relational Spatio-Temporal Pre-processing Approaches

Mariana Oliveira[1,2(✉)], Luís Torgo[1,2], and Vítor Santos Costa[1,2]

[1] DCC – Faculdade de Ciências, Universidade Do Porto, Porto, Portugal
mariana.r.oliveira@inesctec.pt
[2] INESC TEC, Porto, Portugal

Abstract. We present and evaluate two different methods for building spatio-temporal features: a propositional method and a method based on propositionalisation of relational clauses. Our motivating application, a regression problem, requires the prediction of the fraction of each Portuguese parish burnt yearly by wildfires – a problem with a strong socio-economic and environmental impact in the country. We evaluate and compare how these methods perform individually and combined together. We successfully use under-sampling to deal with the high skew in the data set. We find that combining the approaches significantly improves the similar results obtained by each method individually.

1 Introduction

Wildfires are an environmental hazard that affects severely most southern European countries, and Portugal in particular. Although nature relies on fire to rejuvenate the forest, factors such as the introduction of non-indigenous species, the rise of industrial forestry, rural depopulation, and climate changes have compounded the problem [5], severely affecting the country's finances and environment, and sometimes even causing human losses. Given a limited amount of resources to address wildfires, a better understanding of the factors that lead to fire events, and namely to severe fire events, is needed.

Toward this goal, data on wildfires and corresponding geographical context has been continuously collected by several organisations, both at national and at European level. This is an example of the novel environmental and socio-economic databases that store data on entities and how these entities occupy and transform a space while interacting with each other. Besides background data on the entities or the location, such as a site's topology or a country's administrative units, most data will be about the events of interest.

Given the size and complexity of the data, it would be difficult even for a highly qualified expert to fully leverage it. Spatio-temporal data mining techniques offer the promise of finding human-interpretable patterns (e.g., automatically learning association rules), or of models that can be used to successfully predict unknown

We thank Dr. João Torres for providing the data we worked with.

T. Calders et al. (Eds.): DS 2016, LNAI 9956, pp. 183–197, 2016.
DOI: 10.1007/978-3-319-46307-0_12

or future values based on a set of explanatory variables (e.g., regression models to predict risk of fire). We focus on the latter. Dealing with both spatial and temporal dimensions with this goal in mind presents numerous challenges as: (i) the dimensions have different properties, (ii) relationships between spatio-temporal objects are often fuzzy or implicit [1], (iii) multiple levels of granularity and of abstraction of both dimensions impact results differently [27], and (iv) data is often voluminous making scalability a concern.

Propositional data mining methods work on a single table, often assuming that each instance in a data set has been independently sampled from the same underlying distribution. In contrast, multi-relational methods explicitly consider the complex nature of the data, often extending a corresponding propositional approach in order to work on multiple tables from a relational database [12]. In [15], Malerba argues that relational approaches are particularly suited to spatial data mining tasks since they can deal with heterogeneous spatial objects and naturally represent a wide variety of relationships between them. We believe the same argument can be made for spatio-temporal tasks.

Regarding wildfires, the goal will be to estimate the percentage of area burnt (or burn fraction) of a pre-defined unit (in this case, a Portuguese parish) over periods of one year. Moreover, we would like to differentiate major events, that are hard to control, from smallish fire events, that are more frequent but have little impact. To do so, we approach this problem as a regression task.

In order to construct the regression model, we follow the widely used approach of encoding the relevant spatio-temporal information in the form of propositional features through a pre-processing step. These features can be obtained by considering spatio and/or temporal properties in the data [6,18] or even learned through propositional and/or relational techniques. This approach makes it possible to benefit from standard (propositional) prediction models that are both efficient and easy to use.

In this work we present and evaluate two different methods for building spatio-temporal features: a propositional method and a relational method based on Inductive Logic Programming (ILP). We compare how these methods perform individually and combined together. We evaluate performance quantitatively and through the extra knowledge it provides. We also address the high skew in this data set, i.e., the fact that the most important cases of higher burn fraction are under-represented in the data, and demonstrate that under-sampling is quite effective in improving model performance under these conditions.

We proceed to mention some examples of relational and propositional approaches to spatio-temporal prediction that have already been proposed. Purely relational approaches include methods based on ILP [16] and the use of graphical models [4,23]. Propositional approaches include methods based on clustering [3], combinations of spatial and temporal methods [11], extensions of time series forecasting techniques such as ARIMA to account for spatial information [20] and of spatial techniques such as GWR to transfer across time [2]. Further, propositional and relational approaches have been contrasted before in a spatial associative classification setting [10].

In the following section we present the data set we worked with. In Sect. 3 we describe the pre-processing approaches we applied, comparing their results. Section 4 includes concluding remarks and future research directions.

2 Wildfires in Portugal

We proceed to describe the data set motivating our work. We also discuss the pre-computation of spatial relationships below.

2.1 Data Set

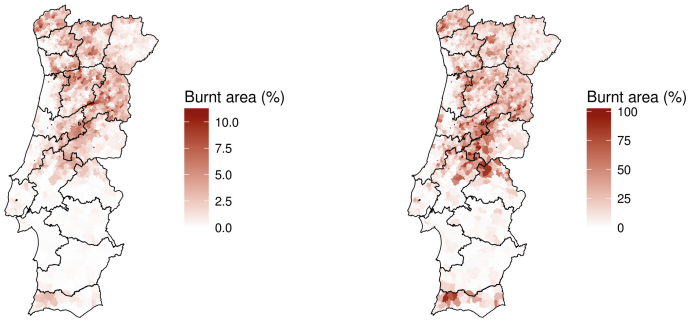
Our motivating application is the evolution of wildfires across mainland Portugal from 1991 to 2010.[1] Spatially, we work at the civil parish level (there are 2882 of them). Our target variable is the percentage of a parish's area burnt yearly by wildfires. The variable is non-cumulative, i.e., an area burning multiple times during the year is considered only once. We used 23 other numerical variables with different temporal granularity as background knowledge (variables measured only once are considered fixed; see Table 1).

Imbalanced Domain. The values taken by our target variable range from 0 % (when no wildfire occurred throughout the year) to 99.8 % (see Fig. 1). The distribution of values is imbalanced in a way that does not correspond to our preference bias, that is, while we are most interested in accurately predicting instances of high burn fraction, these cases are under-represented in the data set. In fact, only about a third of the 57 640 instances have non-zero burn fractions, while less than 9 % of cases present values of 5 % or above and only 0.5 % of cases have a burn fraction of 40 % or more.

Table 1. Explanatory variables used as background knowledge in our data set.

Land cover	Eucalyptus Tall scrubland Small scrubland Broad-leaved forest Pinewood Urban	(%)	Fixed
Terrain	Maximum altitude Mean altitude	(m)	Fixed
	Maximum slope Mean slope		Fixed
Road density	All roads Roads $> 6m$ wide Roads $< 6m$ wide		Fixed
Census data	Irrigable area Meadow area	(%)	Decennial (from 1989)
	Bovine population density Ovine population density Caprine population density	(ha^{-1})	Decennial (from 1989)
	Population density	(ha^{-1})	Decennial (from 1991)
	Population's mean age	(years)	Decennial (from 1991)
	Population of age 65+	(%)	Decennial (from 2001)
	Housing density	(ha^{-1})	Decennial (from 2001)

[1] Most data for this application (with the exception of census data downloaded from ine.pt) provided by Dr. João Torres, researcher at CIBIO. Details regarding data collection can be found in [26].



(a) Mean burn fraction. (b) Maximum burn fraction.

Fig. 1. Mean and maximum percentage of area burnt yearly per parish. Note that they have different scales. Black lines delineate Portuguese districts.

2.2 Computing Spatial Relationships

Both pre-processing approaches we present require the computation of spatial relationships in the data set. We made use of the `PostGIS` spatial extension to a `PostgreSQL` database loaded with the shapefiles of the 2882 Portuguese civil parishes in order to determine spatial neighbourhoods and border parishes.

Defining Neighbourhoods. Neighbourhoods for each parish consist of all intersecting parishes, calculated using the `PostGIS` function `ST_Intersect` (prefix `ST_` identifies `PostGIS` functions).

Neighbour Direction. The relative direction of a neighbour in relation to a reference parish, O, is taken into consideration in an effort to capture effects of dominant winds affecting the spread direction of wildfires. This is not a straightforward problem, given the heterogeneous shapes presented by parishes. Our solution is meant to be fast and easily computable. We first compute cartographic azimuths using the parish of interest as reference and each neighbour's centroids (calculated with `ST_Centroid`). The azimuth is given clockwise relative to the north, resulting in the following definition:

$$\text{neighbour direction} = \begin{cases} \text{east} & \text{if } az \in [45, 135[\,^\circ \\ \text{south} & \text{if } az \in [135, 225[\,^\circ \\ \text{west} & \text{if } az \in [225, 315[\,^\circ \\ \text{north} & \text{if } az \in ([315, 360] \cup [0, 45[)\,^\circ \end{cases} \tag{1}$$

where az is the result of applying the `PostGIS` function `ST_Azimuth` to the centroids of a reference parish and one of its neighbours.

(a) Neighbour direction. (b) Simplified borders.

Fig. 2. On the left, a parish's neighbourhood divided by cardinal directions. Red dots represent the parishes' centroids. The red lines centred at the reference parish, O, divide the neighbourhood in four directions. According to this division, O has no western neighbours. A and B are its northern neighbours; E and D, southern neighbours; and C, an eastern neighbour. On the right, coloured lines define simplified borders between O and each of its neighbours. (Color figure online)

Issue. The example in Fig. 2a illustrates a problem in our proposal. While it is clear O shares borders with neighbours A and E in the western direction, no western neighbour is found since their centroids fall under the northern and southern subspaces. Our propositional approach mitigates this by the way it fills in missing spatio-temporal indicators (Sect. 3.2), while the relational approach relies on the explicit neighbourhood relationship itself (Sect. 3.3).

3 Predicting Wildfires

In the following sections, we define our problem clearly and detail the different steps involved in each pre-processing methodology, as well as steps common to both approaches.

3.1 Problem Definition

Predictive data analysis tasks face the problem of approximating an unknown function $Y = f(X_1, X_2, \ldots, X_p)$ mapping values of a set of predictors or *explanatory* variables, $\mathbf{X}$, into the values of a *target* variable, Y, where the approximation is called the *model*.

In a spatio-temporal setting, the aim is to predict values at different times and locations. In this work, we aim at forecasting future values given past information from neighbouring locations in the past. Consider a data set $D = \{\{y_1^1, x_{a_1}^1, \ldots, x_{p_1}^1\}, \ldots, \{y_n^m, x_{a_n}^m, \ldots, x_{p_n}^m\}\}$ where y_t^l and x_{it}^l correspond, respectively, to the values of the target variable Y and explanatory variables X_i at geographical location l and time t. The goal is to predict the value of Y at a

location of interest, s, at a future time, k, given the observed values y_t^l and $\mathbf{x_t^l}$, such that $t < k$.

3.2 Propositional Pre-processing Approach

The pre-processing stage of our propositional approach can be divided into two steps: calculation of spatio-temporal indicators and imputation of missing data.

Building Spatio-Temporal Indicators. We build two types of indicators: purely temporal, and spatio-temporal. A purely temporal indicator (or, self-indicator) is obtained by calculating the Exponential Moving Average (EMA) of n past values of the target variable for the reference parish in the previous 9 years. We also build spatio-temporal indicators, inspired by the work of [18], considering historical values of the target variable for direct neighbours located at each cardinal direction in the previous 5 years. We compute the indicator for a particular direction in two steps. First, we calculate the EMA (with ratio $\frac{2}{n+1}$) of the target variable for each neighbour whose centroid falls in that direction. Then, if there is more than one, a weighted mean of these values is calculated.

Weighing Neighbours. The weights above are designed to roughly approximate the risk of exposure of a parish to wildfire spread from each neighbour. The strength of connection between neighbours could be directly measured by the fraction of the border shared with them. However, meandering borders can easily increase in length without proportionately increasing the degree of exposure of the reference parish to wildfires originating in that particular neighbour. Therefore, we define a simplified border as the maximum distance between any two points of the intersection using `ST_MaxDistance` (see Fig. 2b). The weight of a neighbour is the length of its simplified border divided by the sum of the lengths of the simplified borders of all neighbours in that direction.

Issues. Weighting a neighbour's EMA in this fashion raises a problem: the centroid of a neighbour may fall in a subspace of a certain direction while most of its border with the reference parish belongs to another. However, since we do not have information regarding which portion of a neighbour was burnt (whether it was close to the border or not), and the level of temporal and spatial granularity of our data is low, these approximations are still reasonable.

Filling in Missing Data. In order to use standard learning algorithms, we pre-selected reasonable procedures to fill in missing data. First, independent spatial-only Inverse Distance Weighting (IDW) (as implemented in [19]) is used to fill in values missing due to unavailability of spatial data (2.8 % of all cells). Next, missing values due to heterogeneous temporal granularity (20.6 % of all cells) are filled in with the latest measurement as there are not enough points to meaningfully smooth over values. Finally, spatio-temporal indicators missing due to no neighbour centroids falling within a certain direction (6.8 % of all cells)

are filled in with zero if the parish borders with the sea/ocean or the average of the two contiguous directions, otherwise.

3.3 Relational Pre-processing Approach

The relational approach we propose follows three steps: first, we rely on the clause search mechanism implemented in the Aleph ILP system [22]; we then propositionalise by associating each clause to a different attribute; last, we construct the attribute examples table. In this table, given an example e and a clause (attribute) i, $e[i] = 1$ if the clause is true for the instance, and $e[i] = 0$ otherwise. This approach has been used before in diverse contexts, including spatial classification [10]. Although Aleph searches for clauses that are optimized to perform well for binary classification, we hypothesise that standard regression algorithms such as Support Vector Regression machines (SVRs) can successfully use the binary features representing interesting clauses to accurately approximate the numerical values of the target variable.

Background Knowledge and Examples

Explanatory Attributes. In order to use Aleph, each explanatory attribute was converted into a binary (if fixed) or ternary (if time-varying) Prolog predicate.

Spatial Relationships. Spatial relationships are expressed by the following predicates: `neighbour(Parish, Neighbour)` where recursion on `Neighbour` is avoided, `neighbourDirection(Parish, Neighbour, Direction)` where `Direction` is defined in Sect. 2.2, and `border(Parish, Object)` where `Object` can take the values `sea` or `spain`.

Temporal Relationships. We use the number of years past since a wildfire last affected a certain parish. This is represented by a pair of predicates: `yearsSinceFireLE(Parish, Year, TimeDist)` and `yearsSinceFireGE(Parish, Year, TimeDist)`, which are true if by `Year` the `Parish` has suffered a wildfire `TimeDist` or less years ago, or `TimeDist` or more years ago, respectively. We use two definitions, conditioned on `TimeDist` being a variable: the first holds when `TimeDist` in unbound, and is used in Aleph's saturation step; the second, is always called with the argument bound to one of the values found during saturation (`TimeDist` is a constant).

Additional Predicates. Since Aleph does not deal with numerical attributes directly, auxiliary predicates were designed. That is, each time-varying attribute had a corresponding `attributeLE(Parish, Year, Value)` and `attributeGE(Parish, Year, Value)` meaning that the value of attribute in `Parish` measured in or before `Year` is lesser or equal (or greater or equal) to `Value`. Similar predicates `attributeLE(Parish, Value)` and `attributeGE(Parish, Value)` were created for fixed attributes.

Examples. The predicate at the head of clauses is `burnt(Parish, Year)`, where a positive example is a `Parish` burnt more than 5 % in a given `Year`.

Clause Search and Selection. The standard Aleph command `induce/0` is not appropriate in this context. Instead, we devise our own method of clause search and selection. We set clause cost (used for generalisation on the reduction step) to be the F_β-measure defined as

$$F_\beta\text{-measure} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}. \tag{2}$$

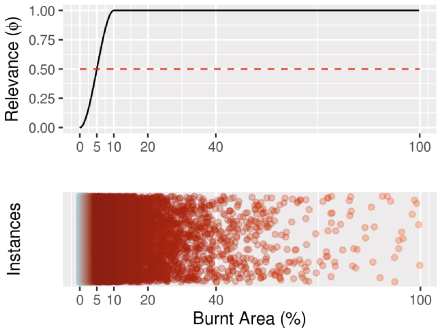
First, we randomly chose an example as the seed for search. We then generate clauses until a certain threshold is reached. Instead of trying to find a theory covering all examples, we store each and every clause that has been the best so far for each saturated example according to our chosen metric. In the style of Gleaner [13], we experimented with different values of β for the F_β-measure, trying 60 random seed examples for each $\beta \in \{0.75, 0.9, 1.0, 1.1, 1.25\}$. Note that this requires that the clause found to be the best so far be reset every time we change the value of β. By varying β, we hope to add some diversity to our discovered clauses, while keeping it around 1.0 assigns similar importance to their precision and recall. We set the Aleph parameters controlling the maximum number of layers of new variables and nodes to 3 and 7500, respectively.

Propositionalisation. Having found the clauses, a Prolog program converts the stored clauses into a CSV file with rows corresponding to instances and columns to clauses. This program is capable of filtering out clauses that are exact repetitions of others, but cannot filter clauses that are even extremely similar except for some minor change in a constant numeric literal, for example.

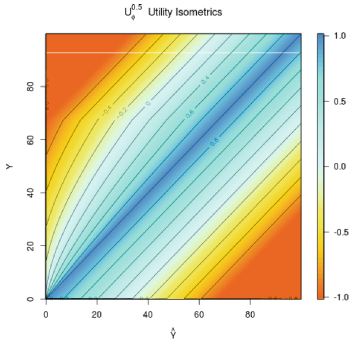
3.4 Common Steps

After the pre-processing methods described in Sects. 3.2 and 3.3, we apply a re-sampling technique that aims to deal with the imbalanced target domain as an extra pre-processing step. The learning algorithms used and post-processing steps are also shared by both approaches.

Re-sampling. Several pre-processing techniques exist to tackle the problem of imbalanced domains that do not correspond to the user's preference bias. Re-sampling techniques are quite effective and have the advantage of having already been proposed for both classification and regression [9], working equally well with numerical and categorical attributes. Besides balancing the domain, under-sampling reduces the dimensionality of the data set by removing instances of the most common class (or range of values), mitigating scalability issues in the process. We pre-selected the re-sampling technique for regression proposed by [25] as implemented in [8]. This method automatically calculates the amount



(a) Relevance function (full black line) used for re-sampling technique and performance metrics adapted to regression under imbalanced domains. Note that the threshold of relevance (dashed red line) coincides with 5% of burnt area. Below, a visual representation of instances across the domain (most are concentrated at 0%).



(b) Contour map of regression utility. The x-axis shows predicted values ($\hat{Y}$) for true values (Y) in the y-axis. Colouring and contour lines map the utility of each prediction as defined by the relevance function ϕ pictured on the left.

Fig. 3. Relevance function and resulting regression utility map. (Color figure online)

of under-sampling needed to balance the domain, given a user-specified relevance function for the target's domain and a threshold of relevance below which instances can be removed. We have settled on the function shown in Fig. 3a with a relevance threshold of 0.5, corresponding to 5 % of burnt area. This relevance function is also used for performance evaluation in Sect. 3.5.

Modelling and Post-processing We applied the Random Forest (RF) and SVR algorithms, as implemented in [14] and [17], respectively, to our transformed data sets. The predictions were then forced into the range of our target variable.

3.5 Experimental Analysis

The main goal of our experimental analysis is to compare the results obtained by standard regression models on our data set after a propositional versus a relational pre-processing approach. We also test if a combination of the two pre-processing approaches results in improved predictions of burn fraction. We try to provide insight into the effect of different variables in the results. Further, we assess the impact of the re-sampling step mentioned in Sect. 3.4.

Experimental Setup Standard metrics such as Mean Squared Error (MSE) are not well equipped to deal with domains where the user preference bias does not correspond to the target domain distribution as in our case [9]. Therefore, we

evaluate the quality of our numeric predictions using F_1-measure$_R$, which is the standard F_1-measure (2) calculated using the following definitions of precision and recall[2], adapted to utility-based regression [7],

$$\text{precision}_R = \frac{\sum_{\phi(\hat{y}_i)>t_R} (1+u_i)}{\sum_{\phi(\hat{y}_i)>t_R} (1+\phi(\hat{y}_i))}, \qquad \text{recall}_R = \frac{\sum_{\phi(y_i)>t_R} (1+u_i)}{\sum_{\phi(y_i)>t_R} (1+\phi(y_i))} \tag{3}$$

where ϕ is the relevance function (depicted in Fig. 3a), t_R is a relevance threshold (set to 0.5), and u is a function of utility of a prediction (defined in [21]) depending on the numeric error of the prediction and the importance of both the predicted $\hat{y}$ and true y values (see Fig. 3b). Moreover, we measure the time spent pre-processing data, training and testing models.

In order to obtain reliable estimates of these metrics, we divide the data set into 10 pairs of sliding training and test sets, and calculate statistics of the results over them. We train our methodologies on data for a stretch of 8 years (23056 instances), and test them in the following 3 years (8646 instances). The first training set starts in 1991 and the last in 2000.

We repeat the experiments for each pre-processing approach (and their combination) with and without the under-sampling step[3]. We test for statistical significance in difference of performance using the Wilcoxon signed-rank test.

Results and Discussion. Tables 2 and 3 summarise the results obtained. For each setup described above, we only show the results achieving the highest F_1-measure$_R$ after grid search parameter optimisation. The propositional and the relational pre-processing approach both obtain very similar results, behaving well when the under-sampling step is performed. However, the best results are obtained by the combination of the two methods, with statistical significance in terms of F_1-measure$_R$ (as determined by pairwise comparisons with this approach as baseline). This seems to validate the incorporation of both approaches when dealing with this kind of data sets, although there is an obvious trade-off on pre-processing and training time when adopting a relational approach. Note also, that in terms of recall, this combination of approaches does not work significantly better than simply applying either approach individually; and in terms of precision, it is also not significantly better than using a relational approach only. The fact that the relational approach works so well, with results competitive with or even better than the propositional approach, confirms our hypothesis that it is possible to build a good regression model by applying a standard algorithm to a table with a numerical target variable and Boolean features optimised for classification (of the categorised target variable).

Moreover, it is clear that under-sampling not only greatly improves the predictive ability of the models, but also decreases the training time needed to build the model. On average, the under-sampling methodology used reduces the train-

[2] Implemented in `R` package `uba` (http://www.dcc.fc.up.pt/~rpribeiro/uba/).
[3] Using `R` package `performanceEstimation` [24].

Table 2. Median (med) and interquartile range (IQR) of results obtained with each methodology. The best results for each pre-processing method are in italic, while the best results overall are in bold. The Wilcoxon signed rank test was used to obtain p-values of the differences in performance with the best approach as baseline (bold p-values mean that the difference is statistically significant).

			Precision$_R$		**Recall**$_R$		**F1-measure**$_R$	
Method	Re-sample	Model	med±IQR	p-val.	med±IQR	p-val.	med±IQR	p-val.
Propositional	None	RF	0.70 ± 0.13	(**0.002**)	0.22 ± 0.13	(**0.002**)	0.33 ± 0.13	(**0.002**)
		SVR	0.68 ± 0.10	(**0.002**)	0.49 ± 0.10	(**0.002**)	0.56 ± 0.10	(**0.002**)
	Under	RF	0.81 ± 0.13	(**0.002**)	0.67 ± 0.13	(**0.002**)	0.72 ± 0.13	(**0.002**)
		SVR	*0.84 ± 0.07*	(**0.002**)	*0.76 ± 0.07*	(0.01)	*0.80 ± 0.07*	(**0.002**)
Relational	None	RF	0.71 ± 0.12	(**0.002**)	0.18 ± 0.12	(**0.002**)	0.29 ± 0.12	(**0.002**)
		SVR	0.68 ± 0.09	(**0.002**)	0.50 ± 0.09	(**0.002**)	0.57 ± 0.09	(**0.002**)
	Under	RF	0.80 ± 0.09	(**0.002**)	0.58 ± 0.09	(**0.002**)	0.66 ± 0.09	(**0.002**)
		SVR	*0.85 ± 0.06*	(0.02)	*0.76 ± 0.06*	(0.04)	*0.80 ± 0.06*	(**0.002**)
Propositional + Relational	None	RF	0.72 ± 0.11	(**0.002**)	0.22 ± 0.11	(**0.002**)	0.33 ± 0.11	(**0.002**)
		SVR	0.70 ± 0.10	(**0.002**)	0.52 ± 0.10	(**0.002**)	0.59 ± 0.10	(**0.002**)
	Under	RF	0.80 ± 0.12	(**0.002**)	0.65 ± 0.12	(**0.002**)	0.70 ± 0.12	(**0.002**)
		SVR	**0.85 ± 0.06**	–	**0.77 ± 0.06**	–	**0.81 ± 0.06**	–

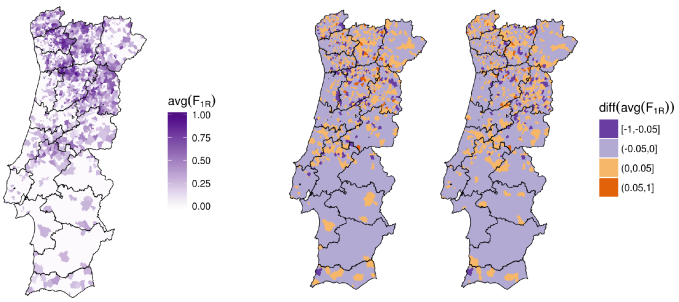
Table 3. Median and IQR of time taken per observation, in seconds, by various methodologies. The pre-processing time shown for propositional approaches includes time spent calculating spatio-temporal indicators and imputing missing data; for relational approaches, it includes time spent finding clauses using the Aleph system and converting them to propositional form (but not time spent encoding auxiliary predicates). Both exclude time spent computing spatial relationships, but include re-sampling time when appropriate. Since part of the pre-processing was performed on the data set as a whole, we omit its IQR.

Method	Re-sample	Model	**Pre-proc.**	**Training**	**Testing**	**Total time**
Propositional	None	RF	2.2e-3	5.8e-1 ± 6e-2	8.0e-4 ± 4e-4	5.8e-1
		SVR	1.7e-3	8.5e-3 ± 5e-4	6.7e-4 ± 7e-5	1.1e-2
	Under	RF	6.8e-3	2.6e-2 ± 6e-3	3.3e-4 ± 6e-5	3.3e-2
		SVR	3.1e-3	1.8e-4 ± 6e-5	2.1e-4 ± 4e-5	**3.5e-3**
Relational	None	RF	1.7	2.1e-1 ± 7e-2	3.6e-4 ± 7e-5	1.9
		SVR	1.7	2.0e-2 ± 1e-2	2.7e-3 ± 6e-4	1.7
	Under	RF	1.7	2.2e-2 ± 6e-3	5.0e-4 ± 4e-4	1.7
		SVR	1.7	6.0e-4 ± 1e-4	7.0e-4 ± 2e-4	*1.7*
Propositional + Relational	None	RF	1.7	1.5e-1 ± 2e-2	2.8e-4 ± 5e-5	1.9
		SVR	1.7	7.0e-2 ± 1e-2	6.0e-3 ± 2e-3	1.8
	Under	RF	1.7	1.9e-2 ± 7e-3	3.2e-4 ± 8e-5	1.7
		SVR	1.7	1.0e-3 ± 3e-4	1.0e-3 ± 1e-3	*1.7*

ing sets to 20 % of their original size, but increases the F_1-measure$_R$ obtained by RFs and SVRs by 118 % and 42 %, respectively.

Furthermore, SVRs consistently (and significantly) outperformed RFs. SVR presented higher susceptibility to parameter tuning, as evidenced by the fact that, when using default parameters, RF routinely outperformed SVR (these results are omitted in favour of those obtained after parameter optimisation).

Figure 4 allows us to examine the spatial distribution of results. We should remark that some parishes have zero or very low numbers of wildfires that exceed the minimum threshold in the considered time period. In this case, one error may have a disproportionate impact on our measure, as it can be observed in areas such as the center-south Alentejo region and highly urbanised areas such as Lisbon metro area. The higher average F_1-measure$_R$ per parish, depicted in Fig. 4a, is strongly (and positively) correlated with the average historic and neighbourhood values of the target variable itself, i.e., with our spatio-temporal indicators. This is not too surprising considering their higher level of temporal granularity but, coupled with the fact that they also achieve the highest RF importance (computed from permuting OOB data), still validates our propositional approach. This strong correlation is closely followed by positive correlations with mean altitude and slope, percentage of area covered by scrubland and caprine population density. Negative correlations were topped by mean percentage of urbanised area, housing, population and road density. We believe the negative correlations are explained by the very few cases of wildfires in urban regions (our data set does not include house fires) as discussed above. By examining Fig. 4b, we can see that although parishes benefiting from the use of each pre-processing approach on its own differ, they are similarly distributed across regions.



(a) Best F_1-measure$_R$. (b) Differences of performance.

Fig. 4. On the left, best average F_1-measure$_R$ overall per parish (obtained by the combination of propositional and relational pre-processing approaches). To the right, the categorised difference of average F_1-measure$_R$ per parish obtained by the best propositional only (in the middle) and relational only (on the right) approaches in relation to the combination of approaches on the left. A negative difference means that the combination of approaches, on average, performs better in that parish. Note that these were obtained using under-sampling and SVR.

Note that the propositional feature space is the same for every training set (29 explanatory variables), while the number of features used by the relational approach depends on the number of clauses found, ranging from 39 to 89 binary features, with a median of 67 features – more than double the amount for the propositional approach. Below, we present two interesting examples of clauses found on the first and last training sets, respectively:

```
burnt(ParishA,Year) :-
    yearsSinceFireGE(ParishA,Year,2),
    neighbourDirection(ParishA,ParishB,west),
    yearsSinceFireGE(ParishB,2).

burnt(ParishA, Year) :-
    maxAltitudeGE(ParishA, 507),
    neihgbourdirection(ParishA, ParishB, south),
    yearsSinceFireLE(ParishB, Year, 5).
```

They can be written as *"A parish burned if it has a western neighbour and both of them hadn't burnt for at least a year"* and *"A parish burned if its maximum altitude was higher than (or equal to) 507 m and it has a southern neighbour that burnt at least once in the last five years"*. Both clauses include temporal predicates as well as information on a parish's neighbours. For a notion of the coverage of these clauses, see Fig. 5. The clauses appear, respectively, on the top 30 and top 50 most important features (according to RF importance) when using under-sampling and a combination of propositional and relational approaches (top 15 and top 10 if using relational only).

Overall, the propositional and relational approaches obtain very good (and similar) results. Although the relational approach performs slightly better in some cases, it requires longer processing times. The results are significantly improved by combining both strategies, which is interesting.

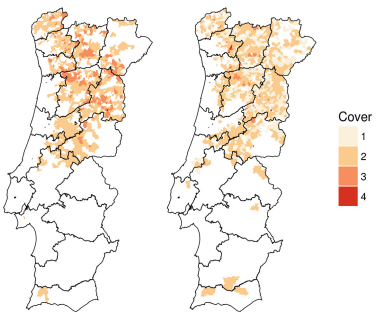


Fig. 5. Spatial coverage of example relational clauses ordered from left to right. Coverage is defined as the number of years for which only the body of the clause is true subtracted from the total number of years that the whole clause is true for each parish. If the body of the clause is always false, the parish is left white.

4 Conclusion

In order to predict the annual burn fraction of Portuguese parishes, we compared two approaches that encode spatio-temporal information in propositional form, each using different pre-processing methods, so that standard regression algorithms can be used. The fully propositional approach builds spatio-temporal indicators considering simplified borders. The relational one, uses an ILP system to find relational clauses that can be transformed into binary features. Both used a notion of spatio-temporal neighbourhood including spatial direction and an utility-based re-sampling technique to deal with this imbalanced domain. Further, we compared each method with a combination of both.

In spite of the features produced by the relational approach having been optimised for classification, the results obtained by the former method are still competitive with (and sometimes slightly better than) the propositional approach in this regression task. Propositional features are, however, much faster to compute. Despite both strategies behaving well after under-sampling, they still perform significantly worse than their combination.

Future work includes exploration of other propositional clustering-based approaches (such as [3]) and graphical modelling techniques (such as Markov Logic Networks), and their application to different data sets. We also plan on investigating whether our results transfer between different countries.

Acknowledgements. We would like to thank Dr. João Torres for providing most of the data we worked with. This work is financed by the ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013.

References

1. Andrienko, G., Malerba, D., May, M., Teisseire, M.: Mining spatio-temporal data. J. Intell. Inf. Syst. **27**(3), 187–190 (2006). doi:10.1007/s10844-006-9949-3
2. Appice, A., Ceci, M., Malerba, D., Lanza, A.: Learning and transferring geographically weighted regression trees across time. In: Atzmueller, M., Chin, A., Helic, D., Hotho, A. (eds.) MSM/MUSE 2011. LNCS, vol. 7472, pp. 97–117. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33684-3_6
3. Appice, A., Pravilovic, S., Malerba, D., Lanza, A.: Enhancing regression models with spatio-temporal indicator additions. In: Baldoni, M., Baroglio, C., Boella, G., Micalizio, R. (eds.) AI*IA 2013. LNCS, pp. 433–444. Springer, Heidelberg (2013). doi:10.1007/978-3-319-03524-6_37
4. Barber, C., Bockhorst, J., Roebber, P.: Auto-regressive HMM inference with incomplete data for short-horizon wind forecasting. In: NIPS, pp. 136–144 (2010)
5. Bassi, S., Kettunen, M., Kampa, E., Cavalieri, S.: Forest Fires: Causes and Contributing Factors in Europe. European Parliament, Brussels (2008)

6. Bilgili, M., Sahin, B., Yasar, A.: Application of artificial neural networks for the wind speed prediction of target station using reference stations data. Renew. Energ. **32**(14), 2350–2360 (2007). doi:10.1016/j.renene.2006.12.001
7. Branco, P.: Re-sampling approaches for regression tasks under imbalanced domains. Master's thesis, University of Porto (2014)
8. Branco, P., Ribeiro, R.P., Torgo, L.: UBL: utility-based learning (2014). R package version 0.0.1
9. Branco, P., Torgo, L., Ribeiro, R.P.: A survey of predictive modelling under imbalanced distributions. CoRR abs/1505.01658 (2015)
10. Ceci, M., Appice, A.: Spatial associative classification: propositional vs structural approach. J. Intell. Inf. Syst. **27**(3), 191–213 (2006). doi:10.1007/s10844-006-9950-x
11. Cheng, T., Wang, J.: Integrated spatio-temporal data mining for forest fire prediction. Trans. GIS **12**(5), 591–611 (2008). doi:10.1111/j.1467-9671.2008.01117.x
12. Dzeroski, S.: Multi-relational data mining: an introduction. SIGKDD Explor. **5**(1), 1–16 (2003). doi:10.1145/959242.959245
13. Goadrich, M., Oliphant, L., Shavlik, J.W.: Gleaner: creating ensembles of first-order clauses to improve recall-precision curves. Mach. Learn. **64**(1–3), 231–261 (2006). doi:10.1007/s10994-006-8958-3
14. Liaw, A., Wiener, M.: Classification and regression by randomforest. R News **2**(3), 18–22 (2002)
15. Malerba, D.: A relational perspective on spatial data mining. IJDMMM **1**(1), 103–118 (2008). doi:10.1504/IJDMMM.2008.022540
16. McGovern, A., Gagne, D.J., Williams, J.K., Brown, R.A., Basara, J.B.: Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. Mach. Learn. **95**(1), 27–50 (2014). doi:10.1007/s10994-013-5343-x
17. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien (2014). R package version 1.6-4
18. Ohashi, O., Torgo, L.: Wind speed forecasting using spatio-temporal indicators. In: ECAI, pp. 975–980 (2012). doi:10.3233/978-1-61499-098-7-975
19. Pebesma, E.J.: Multivariable geostatistics in S: the gstat package. Comput. Geosci. **30**(7), 683–691 (2004). doi:10.1016/j.cageo.2004.03.012
20. Pravilovic, S., Appice, A., Malerba, D.: An intelligent technique for forecasting spatially correlated time series. In: Baldoni, M., Baroglio, C., Boella, G., Micalizio, R. (eds.) AI*IA 2013. LNCS, pp. 457–468. Springer, Heidelberg (2013). doi:10.1007/978-3-319-03524-6_39
21. Ribeiro, R.P.: Utility-based regression. Ph.D. thesis, University of Porto (2011)
22. Srinivasan, A.: The Aleph Manual. University of Oxford, Oxford (2007)
23. Thompson, C.S., Thomson, P.J., Zheng, X.: Fitting a multisite daily rainfall model to New Zealand data. J. Hydrol. **340**(1), 25–39 (2007). doi:10.1016/j.jhydrol.2007.03.020
24. Torgo, L.: An infra-structure for performance estimation and experimental comparison of predictive models in R. CoRR abs/1505.01658 (2014)
25. Torgo, L., Ribeiro, R.P., Pfahringer, B., Branco, P.: SMOTE for regression. In: Pereira, F., Machado, P., Costa, E., Cardoso, A. (eds.) EPIA 2015. LNCS, vol. 9273, pp. 378–389. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40669-0_33
26. Torres, J.: Patterns and drivers of wildfire occurrence and post-fire vegetation resilience across scales in Portugal. Ph.D. thesis, University of Porto (2014)
27. Yao, X.: Research issues in spatio-temporal data mining. In: Workshop on Geospatial Visualization and Knowledge Discovery, UCGIS, pp. 18–20 (2003)