Neurocomputing 207 (2016) 141–149

Contents lists available at ScienceDirect

Neurocomputing

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

Discriminative directional classifiers

Kelwin Fernandes [a,b,*], Jaime S. Cardoso [a,b,*]

[a] INESC TEC, Porto, Portugal
[b] Universidade do Porto, Porto, Portugal

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form
4 February 2016
Accepted 23 March 2016
Communicated by Ding Wang
Available online 13 May 2016

Keywords:
Directional statistics
Supervised classification
Discriminative models
Logistic regression

ABSTRACT

In different areas of knowledge, phenomena are represented by directional-angular or periodic-data; from wind direction and geographical coordinates to time references like days of the week or months of the calendar. These values are usually represented in a linear scale, and restricted to a given range (e.g. $[0, 2\pi)$), hiding the real nature of this information. Therefore, dealing with directional data requires special methods. So far, the design of classifiers for periodic variables adopts a generative approach based on the usage of the von Mises distribution or variants. Since for non-periodic variables state of the art approaches are based on non-generative methods, it is pertinent to investigate the suitability of other approaches for periodic variables. We propose a discriminative Directional Logistic Regression model able to deal with angular data, which does not make any assumption on the data distribution. Also, we study the expressiveness of this model for any number of features. Finally, we validate our model against the previously proposed directional naïve Bayes approach and against a Support Vector Machine with a directional Radial Basis Function kernel with synthetic and real data obtaining competitive results.

1. Introduction

Several phenomena and concepts in real life applications are represented by angular data or, as is referred in the literature, directional data. Some examples of directional information are the wind direction as analyzed by meteorologists, magnetic fields in rocks studied by geologists, geographic coordinates, among others [1]. Also, some entities are usually referenced in an angular manner; gynecologists denote the location to perform a biopsy, when performing a colposcopic screening, using the angle formed by the vertical axis of the cervix. Another example can be found in the area of computer vision, where color is often defined in cylindrical spaces like the Hue-Saturation-Value (HSV) color space. However, directional information is not constrained to scientific contexts; on a daily basis we naturally use angular variables. For example, time is usually represented by hours, days of the week, day of the month, season, etc. This reference system is cyclic by nature.

Directional variables are usually encoded as a periodic value in a given range (e.g. $[0, 2\pi)$, $[0°, 360°)$). This work focuses merely in this representation of directionality, where an angular variable is a real-value number with periodicity defined by a range. However, directional data can also be found in other representations, such as discrete categorical values ordered by a circular relation [2]. Also, some literature makes use of histograms which lie in a circular space instead of the linear one.

Working effectively with directional data requires dealing with techniques that are aware of the angular nature of the information [1]. For example, 0 and 2π are indeed the same angle and their average is not π but 0. In this sense, directional statistics concerns the problems derived from using traditional linear statistics with this type of data [1]. Even visualization of this type of data requires different representations to illustrate its periodic behavior (e.g. rose diagrams and circular histograms). In order to formalize the definition of a directional function, consider the predicate *dir* defined in Eq. (1), where $\mathbb{N}$ is the set of integers and $\mathbb{B} = \{\text{true}, \text{false}\}$:

$$dir : \mathbb{N} \longrightarrow \mathbb{B}$$
$$dir(i) = \text{true}, \quad \text{iff the } i\text{th feature is directional} \qquad (1)$$

We will say that the function f, with domain in $\mathbb{R}^n$, is directional with period $\vec{P}$ (i.e. the feature in the position i has period $\vec{P}_i$), if and only if Eq. (2) holds, where non-directional features are assumed to have infinite period (i.e. $\neg dir(i) \Rightarrow \vec{P}_i = \infty^+$):

$$f(\vec{\theta}) = f(\vec{\theta} + \vec{k} \circ \vec{P}), \quad \vec{k} \in \mathbb{Z}^n \qquad (2)$$

Here on, we will restrict the periodicity of the directional values to $P_i = 1$, without loss of generality.

* Corresponding authors at: INESC TEC, Porto, Portugal.
E-mail addresses: kafc@inesctec.pt (K. Fernandes), jaime.cardoso@inesctec.pt (J.S. Cardoso).

http://dx.doi.org/10.1016/j.neucom.2016.03.076
0925-2312/

Supervised learning can be understood as the process of learning a function f based on the so-called training data that comprises examples of the input vectors and their corresponding target values [3]. In this work, we are interested in the learning task known as classification, where the target can take a finite number of values. These values are usually denoted as classes or labels and the input vector defines a set of features that describe objects in the domain of the function. As the result of a supervised classification task, we obtain a classifier, which is used to assign a class to an object that has not been seen at the training stage. The ability to correctly label new instances is known as generalization [3]. Traditional models that do not take into account directionality may suffer drop of generalization in areas near to the period of the function. Furthermore, the function may return different decisions for different $\Delta + \vec{k} \circ \vec{P}$, $\vec{k} \in \mathbb{Z}^n$, and a fixed $\Delta \in \mathbb{R}^n$, despite all of them semantically represent the same angle.

In this work we propose a binary classifier aware of the directional constraint. The rest of this paper is organized as follows. Section 2 describes related work in the area of directional statistics and learning. Sections 3–5 detail the proposed model, its expressiveness and the optimization strategy, respectively. Section 6 summarizes the performed experiments to assess the relevance of the proposed model and, finally, Section 7 summarizes some conclusions and future work.

2. Related work

Most different types of problems and approaches in Machine Learning can be broadly defined as a classification, regression or clustering tasks. Classification and Regression are the most common supervised learning tasks. On the other hand, clustering is probably the best known unsupervised learning task, where the objective is to group data into non predefined categories based on some similarity criterion.

Previous attempts to address learning tasks with directional data have been carried out in each of the aforementioned areas. Most of them take advantage of circular distributions (such as von Mises and von Mises–Fisher). For instance, Banarjee et al. [4] proposed a generative mixture-model approach for clustering directional data using the von Mises–Fisher distribution. Moreover, they conclude that the spherical k-means is a special case of the mixture of von Mises–Fisher model. Fitting mixtures of angular distributions have been separately studied by Mooney et al. [5] and Mardia et al. [6].

Regression scenarios with directional data have been studied in several contexts [7–9]. Xu and Schoenberg [9] proposed a kernel regression method based on the von Mises distribution. Their method was used to discover the relationship between a single directional explanatory variable (wind direction) and a real-valued linear response variable (total area burned per day in wildfires). Fisher and Lee [7] studied the regression problem where the predictive variables are linear and the model outcome is directional. Their work also assumes that angular observations follow von Mises distributions and focuses on the estimation of the distribution parameters. Finally, Kato et al. [8] addressed the circular–circular problem, wherein both, predictive and target observations, have a circular nature.

Circular ordinal regression is an intermediate problem in this area, which lies between regression and classification. It considers a discrete number of labels which preserve a certain circular order. Devlaminck et al. [2] proposed two methods to solve this problem. The first one is an SVM variation, and the second method transforms the circular ordinal regression problem into multiclass classification. However, the directionality concerns in [2] are focused on the model outcome rather than on the feature space.

In the area of directional classification, different approaches have been considered: from Discriminant Analysis [10,11] to generative models [1,12,13]. SenGupta and Roy [14] proposed a distance-based classification rule using the chord-length between two points on the circle to classify unidimensional data. In more recent work, SenGupta and Ugwuowo [15] developed a multidimensional method for binary classification using directional data; they studied data on torus (two directional variables) and cylinder (one linear variable and one directional variable). Their approach has the limitation that it assumes as known the probabilities of misclassification [15].

Kirby and Miranda [16] proposed a variation on the classic feed-forward neural network by including the notion of a circular node, able to store and transmit angular information. In fact, their node is an abstraction for the combination of a pair of coupled nodes, whose combined values are constrained to lie on the unit circle. However, their solution is not invariant to the same inputs at different periods, namely, a pair of coupled nodes may return different responses to the same angular input. Furthermore, their model requires to manually define the hybrid architecture.

Finally, adaptions to generative models were studied in the past. First, Zemel et al. [13] extended the Boltzmann machine to consider cyclic units. On the other hand, López et al. proposed a directional naïve Bayes formulation [1,12]. Their contribution involves using the von Mises and von Mises–Fisher distributions for the directional variables instead of the classic Gaussian distribution. The effectiveness of this method relies on the independence assumption of the features and the adequacy of the von Mises distribution to model the behavior of the directional features.

In this work, we propose a Directional Logistic Regression, the discriminative counterpart to the Naïve Bayes model, which does not make assumptions on the distribution of the input data.

3. Directional logistic regression

Generative classifiers aim to model the joint probability $p(x, y)$, where x and y respectively denote the input and output variables. Traditional generative models would then make their predictions by choosing the label y that maximizes $p(x, y)$, computed using Bayes rules [17]. Instead, discriminative classifiers model the posterior probability $p(y|x)$. This computation is done in a direct manner or by learning a map from inputs x to the class labels [17].

As we have shown in Section 2, previous attempts to design classifiers for periodic data adopted a generative approach based on the von Mises distribution or variants [1]. Since state of the art approaches are based on non-generative methods for non-periodic variables [18], in this work we propose a discriminant approach to classify directional data. Our contribution stands as a directional-aware version of the Logistic Regression [19], which is the discriminant counterpart of the naïve Bayes classifier, previously used to address this problem. This relation is known as a Generative-Discriminative pair [17].

Eq. (3) defines the Directional Logistic Regression (dLR) model. This model can be understood as a Logistic Regression with a mapping from the original angular space to a linear one. As we show in Section 5, this mapping is learned simultaneously with the feature coefficients. Hereinafter, the two possible labels belong to $\{0, 1\}$, and n is the number of features:

$$f(\theta) = \frac{1}{1 + e^{-k \cdot h(\theta)}}$$