Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 94 (2016) 199 – 206

Procedia Computer Science

The 13th International Conference on Mobile Systems and Pervasive Computing (MobiSPC2016)

A new categorization numerical scheme for mobile robotic computing using odor Data-set recognition as a case

Choukri Djellali, Mehdi Adda

Mathematics, Computer Science and Engineering Dep.
University of Quebec At Rimouski
300, Allée des Ursulines, Rimouski , QC G5L 3A1, Canada

Abstract

Categorization is one of the most active research and application areas of Data Mining. In this paper, we address the problem of pattern categorization in mobile robotic computing. It is the task of automatically sorting a set of patterns into categories from a predefined set. Most categorization algorithms are sensitive to noise, architecture configuration, Bellman's curse of dimensionality, instability, and complex shapes. Hence, in the present study, a novel numerical scheme (RC) for pattern categorization which provides a good generalization ability with a small empirical error, is described. The experimental study with E-nose of six different MOX gas sensors is presented. Our evaluation method demonstrates the effectiveness and multidisciplinary applications of our approach.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Machine Learning, Information Retrieval, Pattern Recognition, Categorization, Features Selection, Support Vector Machines.

1. Introduction

In Data Mining and pattern recognition, categorization is the problem of assigning each input pattern to one of a given set of categories. *Categorization can be defined as the placement of entities in groups whose members bear some similarity to each other*[1]. It is a powerful broadly applicable Data Mining technique that uses supervised learning in order to infer a complex computing function from labeled training patterns. Supervised learning entails approximating the underlying mapping between an input pattern and a desired output value (also known as label).
Nowadays, the categorization approach has been extensively used to study Mobile Communication[15], Wireless Sensor Networks[18], Mobile Web[19], Database[3], Web Services[2], robot recognition[4], video Mining[5], Information Retrieval[6], System security[8], image Mining[10] and Networking[7].

* Corresponding author. Tel.: +1-514-987-3000 ; fax: +1-514-987-8477.
E-mail address: Choukri.Djellali@uqar.ca

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.08.031

Most categorization algorithms are sensitive to outliers, noise, presentation order, architecture configuration and complex shapes. On one hand, the machine learning schemes deal with input patterns that are not linearly separable and the decision boundaries learned by the categorization algorithms can be complex and irregular. On the other hand, the categorization algorithms aim to converge to an optimized configuration. This state can be a local minimum of the function to be optimized (also known as loss function). This locally learning state ensures low training error and provides tight control on over-fitting but can not approximate the complex decision boundaries.
In order to avoid these limitations, we used a new categorization scheme based on kernel learning machine theory and Bootstrap aggregating scheme.
This paper is structured as follows: In Section 2, we present the current state of the art, our research questions and the problematic of categorization. The conceptual architecture of our categorization model is given in Section 3. We present in Section 4 a short evaluation with a benchmarking model for pattern categorization. Finally, a conclusion (Section 5) ends the paper with future works.

2. State of the Art, Problem and Research Questions

Categorization is one of the most important methodologies in Data Mining and it also has a central importance in pattern recognition tasks. It is considered as a separate class of supervised learning that analyzes the training patterns and produces the relevant model, i.e. representing the relationships, correlations, distribution, etc., which can be used for prediction. Supervised learning is used to infer a target function from labeled learning patterns [22].
Formally, categorization is an approximation of a target function ψ by a classifier $\tilde{\tilde{\psi}}$ which is defined as follows:

$$\begin{cases} \psi : P \times C \mapsto \{T, F\} \approx \tilde{\tilde{\psi}} : P \times C \mapsto \{T, F\} \\ \text{if } \psi(p_i, c_j) = T \rightarrow p_i \in c_j \;\; else \; p_i \notin c_j \end{cases} \tag{1}$$

The machine learning task is to select a function $\tilde{\tilde{\psi}}$ that closely approximates a target function ψ by minimizing the generalization error defined by the following formula:

$$E = \underset{\psi}{\text{Argmin}} \, (\frac{1}{n} \sum_{i=1}^{i=n} f_L(\psi(p_i), c_i)), \;\; \forall (p_i, c_i) \in S_n \tag{2}$$

Where, $P \subset \mathbb{R}^n$, $C \subset \mathbb{R}^d$ and $S_n = \{(p_1, c_1), (p_2, c_2), ...(p_n, c_d)\}$, $d \leq n.$, f_L: loss function.

Several categorization models have been suggested using machine learning as a basis for pattern recognition. Luiz M, G. Gonplves' work [9] (2000) was among the earliest efforts in which multi-feature maps are used as input to an associative memory to categorize a set of sensory patterns. In order to build the categorization model, they used a Multi-Layer Perceptron trained with a back-propagation algorithm (BPNN) and a neural network based on the Self-Organizing Map or (SOM).
In the study (Tapomayukh Bhattacharjee et al., 2013) [13] the Hidden Markov Models are used to capture the dynamic robot-environment interactions and to categorize objects. Two HMM models for categorizing trunk vs. leaf was trained. The evaluation based on cross-validation showed that the proposed algorithms yield good results.
The study of (Gonzalez-Aguirre et al., 2013) [12] presented a system to categorize small, rigid and graspable objects with limited visual sensing capabilities in a human household environment. In order to improve the categorization performance, the system used a bagging scheme based on Radial Basis Function or (RBF) kernels, MultiLayer Perceptrons or (MLP) with one hidden layer and K-nearest neighbor classifiers or (K-nn). Visual sensing from different vantage points is used to reconstructe the objects 3D mesh models. This 3D reconstruction is used for shape feature extraction.
J.R. Ruiz-Sarmiento, C. Galindo and J. Gonzalez-Jimenez (2015) [11] employed a Conditional Random Field or (CRF) model to categorize objects and rooms into robot workspace. The evaluation based on home scenes from the NYU2 Data set showed that the proposed model yields good results.
The revolution, that the mobile computing and robotic are witnessing, has led to the appearance of several categorization models. We studied a dozen or so of categorization models that originated from a variety of scientific applications,