Chapter 4
Classification and Machine Learning

Damian Alberto
Indian Institute of Technology Bombay, India

ABSTRACT

The manual classification of a large amount of textual materials are very costly in time and personnel. For this reason, a lot of research has been devoted to the problem of automatic classification and work on the subject dates from 1960. A lot of text classification software has appeared. For some tasks, automatic classifiers perform almost as well as humans, but for others, the gap is still large. These systems are directly related to machine learning. It aims to achieve tasks normally affordable only by humans. There are generally two types of learning: learning "by heart," which consists of storing information as is, and learning generalization, where we learn from examples. In this chapter, the authors address the classification concept in detail and how to solve different classification problems using different machine learning techniques.

INTRODUCTION

The manual classification of a large amount of textual materials are very costly in time and personnel for this reason a lot of research has been devoted to the problem of automatic classification and work on the subject dates from 1960.

These days a lot of text classification software have appeared For some tasks, automatic classifiers perform almost as well as humans, but for others, the gap is still large, these systems are directly related to machine learning (machine learning) it aims to achieve tasks normally affordable only by human there are generally two types of learning: learning "by heart" which consists of storing information as is, and learning generalization where we learn from examples a model we will recognize new examples (Ko, 2012).

In this chapter we will address the classification concept in detail how to solve different classification problems were appealed to different machine learning kind. The text classification is a generic task of assigning one or more categories from a predefined list or not a document by finding a functional link

between a set of texts and a set of categories (tags, classes) according to criteria. They apply to many human activities and particularly suited to the problems of automated decision-making.

For example: we are faced with a set of target text and the goal is to make a computer application capable of autonomously determine in which category classified each text, based on statistical data. Habitually the categories refer to the text subjects, but for particular applications, they can take other forms (Nigam, 2000).

Figure 1. Classification of texts (Sriram, 2010)
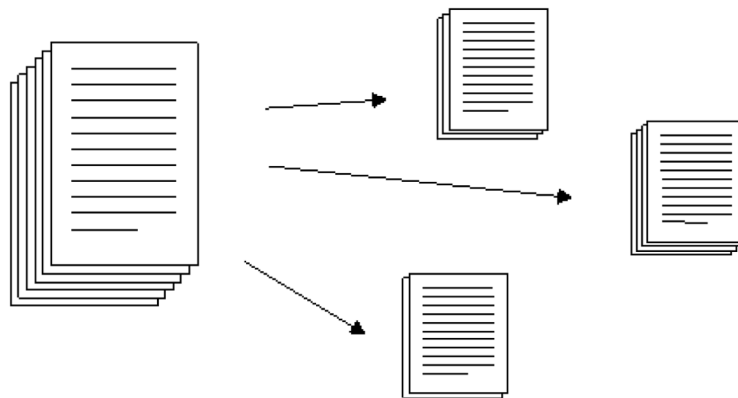
Classification can be found in several application domains as the language identification, recognition of writing, Categorization of multimedia document, spam detection, speech recognition, aid to medical diagnosis.

CLASSIFICATION DUAL CLASS AND MULTI CLASS

The Bi-Classes Classification

The bi-classes classification is a kind of problem examples for which the classification system answers the question: Does the text belongs to the category "C" or to its complementary class (Wu, 2004). some tasks Typical of binary classification are: determining whether a document is it plagiarize or not, Medical tests to determine if a patient has certain disease or not, quality control in the factories decide whether a new product is good enough to be sold, or whether it should be eliminated.

III 3-2-The Multi-Class Classification

The multi-class classification is the problem of the classification of examples in more than two classes or given example can be associated with one or more classes or no class is the most general case of the classification, which can be used in multiple domain: such as handwriting recognition which is part of

Related Content

Eco-Friendly Smart Agricultural Process Using Artificial Intelligence: Economic Benefits
K. Gunasekaran, Sampath Boopathi, S. Suresh, P. Suresh, N. Suthanthira Vanithaand K. Radhika (2024). *Reshaping Environmental Science Through Machine Learning and IoT (pp. 160-175).*
www.irma-international.org/chapter/eco-friendly-smart-agricultural-process-using-artificial-intelligence/346576

ACO_NB-Based Hybrid Prediction Model for Medical Disease Diagnosis
Amit Kumar, Manish Kumarand Nidhya R. (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 526-536).*
www.irma-international.org/chapter/aconb-based-hybrid-prediction-model-for-medical-disease-diagnosis/263336

Development of a Charge Estimator for Piezoelectric Actuators: A Radial Basis Function Approach
Morteza Mohammadzaheri, Mohammadreza Emadi, Mojtaba Ghodsi, Issam M. Bahadur, Musaab Zarogand Ashraf Saleem (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 31-44).*
www.irma-international.org/article/development-of-a-charge-estimator-for-piezoelectric-actuators/249251

Autonomous Last Mile Shuttle ISEAUTO for Education and Research
Raivo Sell, Mairo Leier, Anton Rassõlkinand Juhan-Peep Ernits (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 18-30).*
www.irma-international.org/article/autonomous-last-mile-shuttle-iseauto-for-education-and-research/249250

A CONVblock for Convolutional Neural Networks
Hmidi Alaeddineand Malek Jihene (2021). *Deep Learning Applications in Medical Imaging (pp. 100-113).*
www.irma-international.org/chapter/a-convblock-for-convolutional-neural-networks/260115