

Chapter 23

A Process for Increasing the Samples of Coffee Rust Through Machine Learning Methods

Jhonn Pablo Rodríguez

University of Cauca, Popayán, Colombia

David Camilo Corrales

Telematic Engineering Group, University of Cauca, Popayán, Colombia and Department of Computer Science and Engineering, Carlos III University of Madrid, Madrid, Spain

Juan Carlos Corrales

Telematic Engineering Group, University of Cauca, Popayán, Colombia

ABSTRACT

This article describes how coffee rust has become a serious concern for many coffee farmers and manufacturers. The American Phytopathological Society discusses its importance saying this: "...the most economically important coffee disease in the world..." while "...in monetary value, coffee is the most important agricultural product in international trade..." The early detection has inspired researchers to apply supervised learning algorithms on predicting the disease appearance. However, the main issue of the related works is the small number of samples of the dependent variable: Incidence Percentage of Rust, since the datasets do not have a reliable representation of the disease, which will generate inaccurate predictions in the models. This article provides a process about coffee rust to select appropriate machine learning methods to increase rust samples.

1. INTRODUCTION

Coffee rust has become a serious concern for many coffee farmers and manufacturers. The American Phytopathological Society discusses its importance saying this: "the most economically important coffee disease in the world," while "in monetary value, coffee is the most important agricultural product

DOI: 10.4018/978-1-6684-6291-1.ch023

A Process for Increasing the Samples of Coffee Rust Through Machine Learning Methods

in international trade”. Without a solution, the effects on the coffee industry may soon be reflected in price and availability (Arneson, 2000).

For several years, the disease was managed through the combination of various techniques such as quarantine, cultural management, fungicides and resistant crops. Due to the effectiveness of chemical control and the relatively limited damage caused by the disease, particularly at high altitudes, Mesoamerican coffee farmers and technical authorities considered it manageable. This view prevailed until the epidemic between 2008 and 2013 along Mesoamerica, from Colombia to Mexico, including Peru, Ecuador and some Caribbean countries (Avelino et al., 2015). Coffee farmers were desperate to obtain an answer to this terrible situation since the intensity was higher than anything previously observed, affecting a large number of countries including: Colombia, from 2008 to 2011, affecting an average of 31% of coffee production compared with the production in 2007; Central America and Mexico, in 2012–13, affecting an average of 16% of the production in 2013 compared with 2011-12 and an average of 10% in 2013-14 compared with 2012-13; and Peru and Ecuador in 2013 (Avelino et al., 2015). More specifically, in 2013, the Guatemalan government and the Guatemalan Nation Coffee agency declared a national state of emergency after a projection of nearly 15% crop loss in their region. The devastation has continued to spread due to higher temperatures in this region, which are making fungus growth at higher altitudes possible (“A Solution to the Coffee Rust Epidemic,” 2015). Higher temperatures may be linked to climate change. And several/many experts are worried about the persistence of these conditions (high temperatures) will not change in the near future. In this regard, several reports and experts proposed solutions related with early detection of the disease and the eradication of infected plants.

The early detection has inspired researchers to apply supervised learning algorithms on predicting the disease appearance. The data collected about conditions and soil fertility properties, physical properties and management of a coffee crop, can be used to forecast the rust infection rate. In the same way, weather conditions such as the minimum and maximum levels of temperature, humidity and rainy days can help to estimate the behavior of the disease. Several Colombian and Brazilian researches in supervised learning attempt to detect the incidence percentage of rust (IPR) in coffee crops using Neural Networks, Decision Trees, Support Vector Machines, Bayesian Networks, K Nearest Neighbor, and Ensemble Methods (Cesare di Girolamo, 2013b; Cintra, Meira, Monard, Camargo, & Rodrigues, 2011; Corrales, Corrales, & Figueroa-Casas, 2015; Corrales, Figueroa, Ledezma, & Corrales, 2015; Thamada, Rodrigues, & Meira, 2015). However, the main drawback of the related works is the few data samples of the dependent variable: Incidence Percentage of Rust, since the datasets do not have a reliable representation of the disease, which will generate inaccurate classifiers (Corrales, Figueroa, et al., 2015).

This paper provides a process to increase coffee rust samples applying machine learning methods through a systematic review about coffee rust in order to select appropriate algorithms to increase rust samples. The paper is structured as follows: in the next section, we describe the coffee rust disease and supervised learning concepts. Section 3 exposes the supervised learning approaches applied to coffee rust detection and the main challenges due to low accuracy of rust detecting models; Section 4 shows a systematic review of the approaches to generate synthetic data. Section 5, proposes a process for building large dataset of coffee rust based on the Section 4. Finally, the section 6 presents the conclusions.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/a-process-for-increasing-the-samples-of-coffee-rust-through-machine-learning-methods/307464

Related Content

Arrhythmia Classification Using Deep Learning Architecture

Kuldeep Singh Chouhan, Jyoti Gajrani, Bhavna Sharma and Satya Narayan Tazi (2022). *Real-Time Applications of Machine Learning in Cyber-Physical Systems* (pp. 148-172).

www.irma-international.org/chapter/arrhythmia-classification-using-deep-learning-architecture/299160

Using Open-Source Software for Business, Urban, and Other Applications of Deep Neural Networks, Machine Learning, and Data Analytics Tools

Richard S. Segall and Vidhya Sankarasubbu (2022). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-28).

www.irma-international.org/article/using-open-source-software-for-business-urban-and-other-applications-of-deep-neural-networks-machine-learning-and-data-analytics-tools/307905

Multi-Criteria Decision Making in Healthcare: A Bibliometric Review

Beena John Jiby, Sachin Sakhare, Mandeep Kaur and Gaurav Dhiman (2022). *Demystifying Federated Learning for Blockchain and Industrial Internet of Things* (pp. 186-213).

www.irma-international.org/chapter/multi-criteria-decision-making-in-healthcare/308119

Application of Machine Learning to Analyse Biomedical Signals for Medical Diagnosis

Upendra Kumar and Shashank Yadav (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning* (pp. 205-236).

www.irma-international.org/chapter/application-of-machine-learning-to-analyse-biomedical-signals-for-medical-diagnosis/263321

Applications of Feature Engineering Techniques for Text Data

Shashwati Mishra and Mrutyunjaya Panda (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science* (pp. 182-194).

www.irma-international.org/chapter/applications-of-feature-engineering-techniques-for-text-data/268755