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

Jhonn Pablo Rodríguez, University of Cauca, Popayán, Colombia

David Camilo Corrales, University of Cauca, Popayán, Colombia

Juan Carlos Corrales, 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.

KEYWORDS

Crops, Dataset, Incidence, Synthetic Data

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 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.

2. BACKGROUND

2.1. Coffee Rust

Coffee rust is caused by the fungus Hemileia vastatrix, among the cultivated species C. Arabica is the most severely attacked. The disease causes defoliation, sometimes this one can lead to death of branches and crop losses (Figure 1a). The first symptoms are small yellowish lesions that appear on the underside of the leaves, where the fungus has penetrated through the stomata (Figure 1b). These lesions grow, coalesce and produce uredospores with their distinctive orange color. Chlorotic spots can be observed on the upper surface of the leaves. During the last stage of the disease, lesions become necrotic (Avelino et al., 2015).

The progression of coffee rust depends on four factors (Figure 2) that appear simultaneously (Rivillas Osorio, 2011):

- **The Host:** there are varieties of coffee plants susceptible and resistant to rust. Varieties such as Típica, Borbón and Caturra suffer severe rust attacks, while Colombia and Castillo varieties are highly resistant to rust.
- **Pathogenic Organism:** Hemileia vastatrix lifecycle begins with the germination of uredospores in 2-4 hours in optimal conditions. Within 24 – 48 hours, infection is completed. Once the infection is completed, the underside of the leaf is colonized and sporulation will occur through the stomata (Nutman, Roberts, & Clarke, 1963).

19 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/article/a-process-for-increasing-the-samples-of-coffee-rust-through-machine-learning-methods/203021

Related Content

A Conceptual Model of Grassland-Based Beef Systems

Guillaume Martin, Roger Martin-Clouaire, Jean-Pierre Rellierand Michel Duru (2011). *International Journal of Agricultural and Environmental Information Systems (pp. 20-39).*

www.irma-international.org/article/conceptual-model-grassland-based-beef/51630

SiRCub, A Novel Approach to Recognize Agricultural Crops Using Supervised Classification

Jordi Creus Tomàs, Fabio Augusto Faria, Júlio César Dalla Mora Esquerdo, Alexandre Camargo Coutinhoand Claudia Bauzer Medeiros (2017). *International Journal of Agricultural and Environmental Information Systems (pp. 20-36).*

www.irma-international.org/article/sircub-a-novel-approach-to-recognize-agricultural-crops-using-supervised-classification/188644

Vegetation Filters: The Potential of Short Rotation Woody Crops for the Treatment of Municipal Wastewater

Mohini Singhand R.K. Srivastava (2015). *Handbook of Research on Uncovering New Methods for Ecosystem Management through Bioremediation (pp. 196-221).*

www.irma-international.org/chapter/vegetation-filters/135096

Review on Plastic Waste Disposal and Role of Microorganisms in Bioremediation of Plastics

Vyshnavi V. Rao, Sonashree R.and Rashmi R. Halbavi (2022). *Research Anthology on Emerging Techniques in Environmental Remediation (pp. 481-492).*

www.irma-international.org/chapter/review-on-plastic-waste-disposal-and-role-of-microorganisms-in-bioremediation-of-plastics/291251

Green Chemistry: A New Frontier of Challenges and Opportunities for Pollution Abatement

Prakash Chandra Mishraand Anil Kumar Giri (2018). *Microbial Biotechnology in Environmental Monitoring and Cleanup (pp. 152-161).*

www.irma-international.org/chapter/green-chemistry/196797