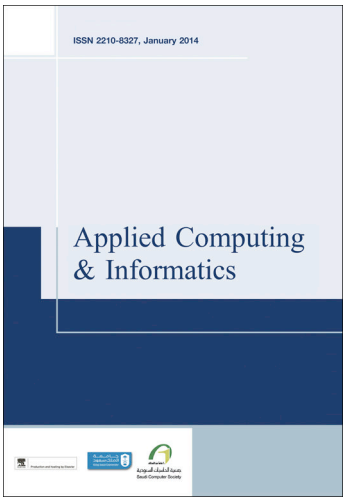
Accepted Manuscript



Original Article

Research on social data by means of cluster analysis

Camila Maione, Donald R. Nelson, Rommel Melgaço Barbosa

PII: S2210-8327(17)30352-6
DOI: https://doi.org/10.1016/j.aci.2018.02.003
Reference: ACI 104

To appear in: *Applied Computing and Informatics*

Received Date: 13 December 2017
Revised Date: 20 February 2018
Accepted Date: 20 February 2018

Please cite this article as: Maione, C., Nelson, D.R., Melgaço Barbosa, R., Research on social data by means of cluster analysis, *Applied Computing and Informatics* (2018), doi: https://doi.org/10.1016/j.aci.2018.02.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Research on social data by means of cluster analysis

Camila Maione[a], Donald R. Nelson[b], Rommel Melgaço Barbosa[a,*]

[a] Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[b] Department of Anthropology, University of Georgia, Athens, USA

*Corresponding author. Email: rommel@inf.ufg.br. Postal address: Alameda Palmeiras, Quadra D, Câmpus Samambaia, Instituto de Informática, CEP 74690-900, Goiânia, Goiás, Brazil.

Declarations of interest: none

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ABSTRACT

This paper presents a data mining study and cluster analysis of social data obtained on small producers and family farmers from six country cities in Ceará state, northeast Brazil. The analyzed data involve demographic, economic, agriculture and food insecurity information. The goal of the study is to establish profiles for the small producer families that reside in the region and to identify relevant features which differentiate these profiles. Moreover, we provide an efficient data mining methodology for analysis of social data sets which is capable of handling its natural challenges, such as mixed variables and abundance of null values. We use the Silhouette method for the estimation of the best number of natural groups within the data, along with the Partitioning Around Medoids clustering algorithm in order to compute the profiles. The Correlation-Based Feature Selection method is used to identify which social criteria are the most important to differentiate the families from each profile. Classification models based on support vector machines, multilayer perceptron and decision trees were developed aiming to predict in which of the identified clusters an arbitrary family would be best fit. We obtained a good separation of the families into two clusters, and a multilayer perceptron model with approximately 93.5% prediction accuracy.

Keywords *- clustering, social data, classification, pam, data mining*

1. INTRODUCTION

The Brazilian law (Lei 11.326, 2006) defines the family farmer as the family which practices agricultural activities in a rural area that is not larger than four fiscal modules and uses predominantly the family's own labor in their business activities. Moreover, according to the law, the income of a family farmer must be originated from activities related to their own business, which must be ran by the farmer and his family. According to Guanziroli (2001), in 2001, family-based agriculture in Brazil was so far responsible for almost 40% of the total agricultural production, and accounts for 76.8% of agricultural employment.

Most of the family-based agriculture in Brazil is consolidated in the Northeast region. However, the family farmers which reside in this region face various challenges related to the soil and climate. Approximately 70% of the semiarid region is

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات