

TSDL: A framework for Tip Spam Detection in Location Based Social Network

Adewole K. S

Department of Computer Science, University of Ilorin, Nigeria
adewole.ks@unilorin.edu.ng

Jimoh R. G

Department of Computer Science, University of Ilorin, Nigeria
jimoh_rasheed@yahoo.com

Isiaka R. M.

Department of Computer, Library and Information Science, Kwara
State University, Malete, Nigeria
abdulrafiu.isiaka@kwasu.edu.ng

Ajiboye A. R.

Department of Computer Science, University of Ilorin, Nigeria
ajibabdulraheem@gmail.com

Abstract—In Web 2.0 systems, Location Based Social Network (LBSN) has become increasingly popular because of its ability for locating places, such as restaurants, hotels and nearby facilities that can render services to the users. Among such LBSN is Apontador, a popular social network introduced in Brazil a couple of years ago. However, despite the numerous opportunities offered by Apontador LBSN, it has become attractive social network for tips spam distribution. In this work, an intelligent machine learning framework is proposed to detect tips spam in Apontador LBSN. The proposed framework explored three classification algorithms: Random Forest, Multilayer Perceptron, and Decorate. Bio-inspired feature identification was studied using Evolutionary algorithm (EA) and Particle Swarm Optimization (PSO) to identify the discriminating features for tip spam detection in LBSN. Three experiments were conducted based on sixty (60), twenty-two (22), and ten (10) features to ascertain the performance of the proposed framework for tip spam detection. Based on the various experiments conducted, Random Forest classification algorithm produced accuracy and F-measure of 90.5% and ROC of 96.7% using 60 features extracted from Apontador LBSN. The algorithm also outperformed the two other classifiers during EA evaluation. However, Decorate classifier produced the best results during the PSO evaluation, achieving F-measure and ROC of 86.3% and 92.9% respectively. The experimental results show that the proposed framework improved in performance for tips spam detection in Apontador LBSN.

Keywords— Location based social network; tip spam; evolutionary computation; machine learning.

I. INTRODUCTION

Location Based Social Network (LBSN), such as Apontador and Yelp, are now becoming attractive social media for the new set of users to obtain information about new locations. LBSN enables users to share information, which may include hotel nearby, restaurants and a host of other places about a specific location to the new users using smartphone with embedded GPS capability. LBSN is a type of online social network (OSN) based on the Web 2.0 technology that is increasingly attracting attention from visitors who are unaware of the layout of the new locations. For instance, the popular Brazil based LBSN, Apontador, includes features that allow users to locate different places for relaxation, add new locations, check-in in locations, and read users tips about a location, such as interesting things to buy or products to avoid when visiting a specific location as recommended by the users on the platform (Costa et al., 2014). This tip allows new users to explore the area and gain more insight on the geography of the new locations. Despite the benefits offer by LBSN to the users of the system they have become attractive platforms for new form of spam message distribution, which propagates

wrong tips about a location. For instance, tip spam in LBSN may appear inform of opinion spam, which targets to mislead readers by posting positive reviews about a specific area in order to promote the location. At the same time, spammer can post negative review to damage the reputation of the area or link users to other external web pages that are completely unrelated to the users need (Guo & Chen, 2014; Molavi Kakhki et al., 2013). This misleading tip can expose LBSN to different attacks from spammers and endanger the trust relationship existing on the platform. Thus, compromising the efforts of promoting LBSN and indeed waste readers' time in filtering spam tips from legitimate information about places (Costa et al., 2013). While existing studies on spam detection in OSN focus more on social networks like Twitter (Cresci et al., 2015) and Facebook (Ahmed & Abulaish, 2012), little efforts have been committed to detecting tip spam in LBSN. For example, Costa et al. (2013) presented machine learning technique to identify tip spam in LBSN using Apontador as test bed. They manually annotate 1,260 spam tips and obtain 1,260 non-spam tips from Apontador. These labeled collections were used to train and test machine learning