

Accepted Manuscript

Ensemble based spam detection in social IoT using probabilistic data structures

Amritpal Singh, Shalini Batra

PII: S0167-739X(17)31313-4
DOI: <https://doi.org/10.1016/j.future.2017.09.072>
Reference: FUTURE 3728

To appear in: *Future Generation Computer Systems*

Received date : 21 June 2017
Revised date : 19 August 2017
Accepted date : 27 September 2017

Please cite this article as: A. Singh, S. Batra, Ensemble based spam detection in social IoT using probabilistic data structures, *Future Generation Computer Systems* (2017), <https://doi.org/10.1016/j.future.2017.09.072>

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.



Ensemble based Spam Detection in Social IoT using Probabilistic Data Structures

Amritpal Singh^{a,1,*}, Shalini Batra^{a,2}

^aThapar University, Punjab, India 147001

Abstract

A social approach can be used for the Internet of Things (IoT) to connect large number of objects in social networks like Twitter, Facebook, Instagram, etc. Social networks within the IoT domain have simplified the task of dynamic discovery of services and information. Detecting spam in social media, especially when massive data flows continuously and large number of attributes are associated with it, is a daunting task which requires lot of technical insight. This paper proposes a semi-supervised technique for spam detection in Twitter by employing ensemble based framework comprising of four classifiers. The framework is based on usage of Probabilistic Data Structures (PDS) like Quotient Filter (QF) to query the URL database, spam users, spam words databases and Locality Sensitive Hashing (LSH) for similarity search, as classifiers in various stages which provide fast results with less computational effort. Performance of the framework has been evaluated by comparative analysis of PDS with the similar data structures and through the standard evaluation parameters which include precision, recall and F-score.

*Corresponding author

Email addresses: amritpal.singh203@gmail.com (Amritpal Singh), sbatra@thapar.edu (Shalini Batra)

¹A. Singh is Research Associate in the Department of Computer Science and Engineering, Thapar university, Patiala, Punjab, INDIA.

²S. Batra is Associate Professor in Department of Computer Science and Engineering, Thapar university, Patiala, Punjab, INDIA, since 2002.

Download English Version:

<https://daneshyari.com/en/article/6873281>

Download Persian Version:

<https://daneshyari.com/article/6873281>

[Daneshyari.com](https://daneshyari.com)