Spam Reviews and Spammer Community Detection Technique using Heterogeneous Information Network

Kajal Sonawane[*1], Prashant Yawalkar[2]

[*1]Computer Department, Savitribai Phule Pune University, Nashik, Maharashtra, India
[2]Computer Department, Savitribai Phule Pune University, Nashik, Maharashtra, India

ABSTRACT

Online reviews and feedback of a product plays a vital role in human tendency to purchase those products. To affect the product sale spammer generates fake reviews on online social media platform. To identify spam reviews and spammer communities is the area of interest of this research work. In literature work, various spam detection techniques are proposed based on Review-Behavioral (RB) Based features, Review-Linguistic (RL) Based Features, User-Behavioral (UB) Based Features but none of the technique provide a simultaneous study of these features and weighting of the features along with finding the relationship among the spam users. The proposed work generates a heterogeneous network having users and reviews are the node type and the spam detection technique is mapped to the heterogeneous network problem. Spam reviews are detected using metapath. A feature weighting method is proposed to detect the relative feature importance. Using generated metapath for similar products spammer community detection is proposed.

Keywords: heterogeneous network, spam detection, fake reviews, social network

I. INTRODUCTION

Customer judge the product based on online reviews. The sale of product depends on these online reviews. To stand in market with positive impression, growing number of businesses tries to receive online praises from their consumers. On the other hand to reach to superior position negative marketing of opponent product is promoted. To float positive or negative reviews based on the business needs artificial reviews are generated and posted on various online social media sites. As per the survey[4] 1/3 customer reviews on social media sites are suspicious. To generate fake reviews and float those reviews on social media site is called as spamming.

In the proposed work analysis of this spam reviews is done using 3 techniques: Review-Behavioral (RB) Based feature, Review-Linguistic (RL) Based Features, User-Behavioral (UB) Based Features. The spam detection probem is converted in to analysis of heterogeneous information network problem. A network is established using users and their reviews as a node and metapaths are identified to detect the network spam reviews. For each type of review weight is assigned. Based on the feature weight importance of feature is calculated in spam detection technique.