Chapter 8
A Novel Feature Correlation Approach for Brand Spam Detection

Bharat Tidke
Sardar Vallabhbhai National Institute of Technology, Surat, India

Swati Tidke
College of Engineering, Pune, India

ABSTRACT

In this age of the internet, no person wants to make his decision on his own. Be it for purchasing a product, watching a movie, reading a book, a person looks out for reviews. People are unaware of the fact that these reviews may not always be true. It is the age of paid reviews, where the reviews are not just written to promote one's product but also to demote a competitor's product. But the ones which are turning out to be the most critical are given on brand of a certain product. This chapter proposed a novel approach for brand spam detection using feature correlation to improve state-of-the-art approaches. Correlation-based feature engineering is considered as one of the finest methods for determining the relations among the features. Several features attached with reviews are important, keeping in focus customer and company needs in making strong decisions, user for purchasing, and company for improving sales and services. Due to severe spamming these days, it has become nearly impossible to judge whether the given review is a trusted or a fake review.

INTRODUCTION

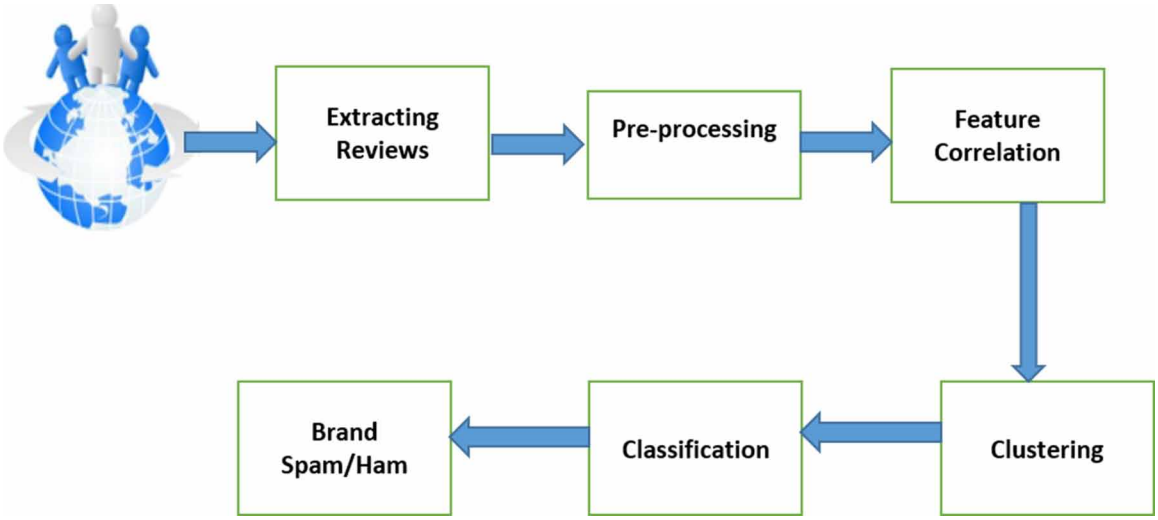
Today, the traditional style of marketing, which is considered as expensive way of promoting, marketing has been taken over by online reviews (Jindal and Liu 2008; Lau et al. 2015). Online reviews are playing a great role in attracting customers and helps in extending communication (Asghar et al. 2020). These reviews are very important part of customer's life, as it helps them to make decisions in purchasing a quality product, Companies harness these reviews to make decisions for improving their businesses. It

DOI: 10.4018/978-1-7998-7371-6.ch008

always looks a great picture when these reviews are true and gets worst if the reviews start getting fake (Fairbanks et al. 2018).

Brand spamming becomes one of the key challenges which increase fake or false reviews. To increase the sales of newly manufactured and launched product, companies generally take support of the existing and famous brands. Though it may not seem wrong, but this biased reviewing describes less about the product and more about supporting product. In case of movie spammer write review about production house. Reviews are not on single attribute of product here. These types of opinions change the direction of reader which may divert to another movie. The proposed work extends effort in detecting these brand spam.

Figure 1. An overview of methodology



Use of clustering has attracted great attention in recent past from researchers and practitioners due to its usefulness in numerous applications (Xu and Wunsch 2008) as shown in Figure 1. Clustering appears as a capable approach to discover group of objects or individual from diverse data. K-means algorithm is an effective technique for clustering and well known in data mining community. However, if cluster is too large then mining the information from users' contents is very expensive and a large share of the content is basically not worthy.

The paper has been divided into five sections. Section one gives a brief introduction. Section two describes motivation. In the third section, detailed literature survey is presented. The fourth section comprises of proposed work for detecting brand spam using feature correlation. In the end, section five gives a conclusion.