BIRZEIT UNIVERSITY

Electrical and Computer Engineering Department

ENCS434 Artificial Intelligence, First Semester, 2020-2021

Programming Project 2 Instructor: Aziz Qaroush

Due: January 20, 2021

Automatic Spam Review Detection

This assignment is for groups of 3 students each (at most). If you want to do it alone you must get the permission of the instructor.

1. Motivation and Introduction

The Internet is now considered to be the most demanded service all around the world. People now can communicate and share their knowledge very easily through different ways such as writing comments in the social sites or review the products they have used describing their experience using that product. As a result, the reviews became very important for the customer and the business as many online shopping paradigms have spread out which help users to review products and share their reviews on social media. Such reviews impose a huge impact on product sales since people tend to check the reviews before deciding to purchase products. Since reviews can be provided by anyone on the public domain, users misuse the reviewing services by providing fake reviews to promote or demote a particular product. Later in this research, fake reviews will be referred as 'spam' and individuals who provide spam reviews will be referred as spammers.

The spam reviews (i.e. Fake reviews provided by spammers to promote or demote particular product) is the type to be covered in this research. Spam reviews are classified by Jindal et. al. into mainly three types:

1. **Type 1 Fake Reviews:** Providing fake opinions about products to promote or demote products by undeserving positive and negative opinions respectively which are written with hidden motives. This type of reviews is also called fake reviews
2. **Type 2 On Brands Reviews:** this type of spam doesn't sufficient to concentrate on the products or services, since it can be realted the cost of products, the sellers, the manufacturers of the products and other aspects. Although it supposed to be on true, but they are considered as spam as they cannot provide any helpful about the concerned product. (e.g. "I hate Dell products and I'll never buy their products").
3. **Type 3 Non-Related Reviews:** it also called Non-related reviews which are considered as spam since it can be an advertisement or unhelpful opinions such as

answers, questions and any random texts. (e.g. “<3 <3 <3 <3 <3” and “how to make a deliver request?”).

Due to the big challenge in the business today some companies illegally tend to hire employees to write fake reviews to promote their products or demote other competitors’ products. So, there is a big concern from the customer and the business to have real reviews, for these reasons many sites tend to filter spams and display only the real reviews to maintain credibility. Handling the problem of review spam can be done in two ways which is manually filtering which is time consuming and will not be effective enough because the spammers now use new methods to write spam reviews, also they can use an automatic ways of inserting spam reviews which make it hard to detect by human since the ordinary people cannot measure the semantic similarity, also, reliance of multiple features of identifying spams is limited in case of manually filtering. The second way is automated filtering based on a set of features extracted from review text and reviewer account to identify the spam reviews which is fast and more efficient than manually filtering, but the main challenge is developing new algorithms for detecting spams efficiently taking into consideration the new methods spammers use to overcome these algorithms, this approach has an important space in recent researches.
One of the examples from websites that use automated filtering is YELP[1] since they use filters to detect the spam reviews on the website and just showing the user’s helpful reviews.

For more information about Spam reviews please read the following papers:

- What Yelp Fake Review Filter Might Be Doing
- On the Temporal Dynamics of Opinion Spamming: Case Studies on Yelp
- S. Dixit and A. Agrawal, "Survey on review spam detection," in *Int J Comput Commun Technol ISSN (PRINT)*, 2013.
- N. Jindal, B. Liu and E.-P. Lim, "Finding Unusual Review Patterns Using Unexpected Rules," in *19th International Conference on Information and Knowledge Management and Co-located Workshops*, Canada, 2010.
- Survey of review spam detection using machine learning techniques
- Spam Review Detection Techniques: A Systematic Literature Review

2. Problem Statement

Spam reviews problem is a challenging problem; as most people nowadays use online shopping to make purchase decisions based on the other reviews, helpful and non-fake reviews became very important business owner to provide the honest reviews to the customers, also it is crucial for customers to find helpful reviews in order to make the right decision. The goal is to develop an automated solution that filters the reviews out based on extracted features from the review content, reviewer behaviors and information. This application will help the customers to trust the websites that use spam filters to display the honest reviews only.

[1] Yelp: https://www.yelp.com/dataset

3. Data set:
You will train and test your system using the following Yelp dataset:
https://drive.google.com/file/d/1uo1WjnBIcKPaKTvl0NaGFRYR6Di1ovwu/view?fbclid=IwAR0UjrJAqhiEp6hxxK7rSjN0V-5Jw0uDq6gYce9cNK3KHtKeehTmeFZqBDo

The dataset contains the following fields:

reviewerID	reviewContent	rating	usefulCount	date	reviewerID	firstCount	reviewCount	Filtered

Where:

- reviewerID: is the ID of the reviewer
- reviewCountent: text of the review
- rating: is the rate provided by the reviewer
- usfulCount: number of times this review marked as useful
- firstCount: the number of times this reviewer is the first one that make review
- reviewCount: total number of reviews for the reviewer
- Filtered: is the label [0 mean non-spam, 1 mean spam]

For more information visit the following link:

5. Submissions: Please submit the following:
1. **Report**:

- Describe in details your formalization of the problem including the stages of your solution, selected features, and results.
- describe in details how you designed each feature.
- The results you obtained including evaluation method, measures, and comparing t other related work.

2. **Source Code** : Include all the source code you developed or extended from the program. These need to be submitted only electronically (no hardcopies of the code). The running program needs also to be submitted electronically.

3. **Demo:** You will be asked to demo your work to your instructor. For that you need to be able to work with your program, introduce minor modifications and defend your choices.

Honor Policy: All are required to adhere to the University honor policy and violations will be dealt with according to University regulations.

Good Luck