# 1. Detection of spam reviews: a sentiment analysis approach.

Electronic shopping is highly influenced by online reviews posted by customers against the product quality. Some fraudulent pretenders consider this as an opportunity to write the spam reviews to upgrade or degrade product’s reputation. Hence, detection of those reviews are very essential for preserving the interests of users. To date, number of researches have been proposed in order to detect the spam reviews and to provide the genuine resources for customers and business person. However, we found few limitations in existing supervised approaches. First, most of the supervised approaches have used manual labelling of reviews into spam and non-spam. However, due to identical appearance of reviews manual labelling can not be considered as authentic. Second, the scarcity of spam reviews leads to data imbalance problem. Third, computing similarities among reviews naturally needs expensive computation. In this work, we propose a novel and robust, spam review detection system which efficiently employ following three features: (i) sentiments of review and its comments, (ii) content based factor, and (iii) rating deviation. To address the aforementioned limitations, we investigated all these features for only suspicious review list in which only those reviews have kept which received comments by peer users. The proposed system achieved the F\(\_1\)-score of 91%. The proposed system can be a great asset in spam detection system as it can be used as an stand-alone system to purify the product review datasets.

**2. Spotting fake reviews via collective positive-unlabeled learning**

Online reviews have become an increasingly important resource for decision making and product designing. But reviews systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, ground truth of large scale datasets is still unavailable and most of existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping, the largest Chinese review hosting site, we present the first reported work on fake review detection in Chinese with filtered reviews from Dianping's fake review detection system. Dianping's algorithm has a very high precision, but the recall is hard to know. This means that all fake reviews detected by the system are almost certainly fake but the remaining reviews (unknown set) may not be all genuine. Since the unknown set may contain many fake reviews, it is more appropriate to treat it as an unlabeled set. This calls for the model of learning from positive and unlabeled examples (PU learning). By leveraging the intricate dependencies among reviews, users and IP addresses, we first propose a collective classification algorithm called Multi-typed Heterogeneous Collective Classification (MHCC) and then extend it to Collective Positive and Unlabeled learning (CPU). Our experiments are conducted on real-life reviews of 500 restaurants in Shanghai, China. Results show that our proposed models can markedly improve the F1 scores of strong baselines in both PU and non-PU learning settings. Since our models only use language independent features, they can be easily generalized to other languages.

**3. Prediction of Review Sentiment and Detection of Fake Reviews in Social Media**

For the past dozen years, social media has been widely used and has slowly become major sources of various information. Some social media sites specialize in people's reviews and making recommendations. Some reviews are positive while others negative. Either way, they have a significant impact on users' decisions of whether purchasing the products or services or not. However, fake reviews and fake users can greatly twist the true opinions that are out there. The detection of fake reviews is critical for any progress toward making great usage of the web. In this work, we analyze the statistics of a Yelp Challenge dataset and propose a simple-butpowerful new detection algorithm based on locations of the reviewers and businesses. We also investigate the prediction of the sentiments of the reviews. We compare the accuracy and false alarm rates of several prediction algorithms. While the results are less than optimum, but there are hopes for strong performance with minor revisions.

**4. Fake product review monitoring using opinion mining**

Product reviews play an important role in deciding the sale of a particular product on the ecommerce websites or applications like Flipkart, Amazon, Snapdeal, etc. In this paper, we propose a framework to detect fake product reviews or spam reviews by using Opinion Mining. The Opinion mining is also known as Sentiment Analysis. In sentiment analysis, we try to figure out the opinion of a customer through a piece of text. We first take the review and check if the review is related to the specific product with the help of Decision tree. We use Spam dictionary to identify the spam words in the reviews. In Text Mining we apply several algorithms and on the basis of these algorithms we get the specific results.

**5. A System Using Deep Learning and Fuzzy Logic to Detect Fake Yelp Reviews.**

With the prevalence of online searching, looking up online reviews of businesses, such as restaurants, hotel and other services, is a major factor in people’s decision making. However, fake reviews cause the sentiment analysis of a corpus of reviews to be clouded. This research uses the YELP data set that is publicly available on the internet. I connect both review content, user information and business information to optimize the fake review detection accuracy. In terms of my solution, I use both feature extraction and deep learning-based classification to detect fake reviews. In feature extraction, features are extracted from reviews using term frequency and frequency-inverted document frequency. I extract features from users’ information like number of reviews, review length, number of fake reviews generated, review date, and review helpfulness. I also extract features from businesses like star number, number of reviews, number of associated fake reviews, and type of business. All of these features will be input to different deep learning models (CNN and RNN.) and results compared to determine which machine learning model has the best performance. I then use fuzzy logic to classify the Yelp dataset into 3 groups: fake review, neutral review and true review. This fuzzy classification allows the user to know when a review is real or fake, and to look at those reviews in between to determine their authenticity.

**6. Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training.**

Fake reviews may mislead consumers. A large number of fake reviews will even cause huge property losses and public opinion crises. Therefore, it is necessary to detect and filter fake reviews. However, most existing methods have lower accuracy in detecting fake reviews due to they just use single features and lack of labeled experimental data. To solve this problem, we propose a novelty method to detect fake reviews based on multiple feature fusion and rolling collaborative training. First, the method requires an initial index system with multiple features such as text features, sentiment features of reviews and behavior features of reviewers. Second, the method needs an initial training sample set. Thus, we designed related algorithms to extract all the features of a review. Then the classification of the review is labeled manually. Finally, the method uses the initial sample set to train 7 classifiers, and the most accurate classifier will be selected to classify new reviews. The novelty of the method lies in that the features and the classification labels of the new reviews will be added into the initial sample set as new samples. So the size of the sample set will increase automatically. The experimental results in the reviews of yelp shopping website show that the accuracy of the proposed method for detecting fake reviews is 84.45%, which is 3.5% higher than the baseline methods. And compared with the latest deep learning model, its baseline precision has increased by 5.3%. According to the Friedman test, the support vector machine (SVM) classifier and random forest (RF) classifier has been proven to be the best one by statistical means. It means our method which uses multiple features has higher accuracy than the baseline models. Meanwhile, it also resolves the problem of lacking labeled training samples in fake reviews detection.