Name : Ajay Rastogi

Supervisor : Prof. Monica Mehrotra

Title : Detection of Opinion Spam in Online Social Media

Department : Department of Computer Science, Jamia Milla Islamia

Abstract

Online social media has provided an excellent way for people to share opinions or experiences, exchange ideas, provide suggestions etc. by using different means of online communication. With the advancements in web 2.0 technology, a variety of opinion sharing platforms such as *Amazon*, *Flipkart*, *Yelp*, *TripAdvisor* etc. have emerged as successful business organizations in terms of online marketing. These platforms offer their users a way to write online reviews about various products or services, which play a central role in people's decision-making process. Nowadays, people use online reviews more often to know about certain quality characteristics of online products or services, before deciding to make a purchase. These reviews are highly important as these are a product's quality indicators and have the potential to influence the decision making of a consumer. Moreover, these reviews are not only used by the consumers in decision making, but also utilized by product manufacturers or businesses for the betterment of their products or services.

In this thesis, initially we perform an in-depth and systematic review of the problem of opinion spam detection with a brief summary of the research articles published between 2007 and 2015. To gain a better understanding of the different aspects of opinion spam detection, the literature has been categorized based on three factors: detection targets, spamming features, and opinion-spam detection methods. The first category discusses the different detection targets of opinion spam, which include detecting opinion spam, opinion spammers, and

collusive opinion spammers' groups. Thereafter, the different spamming features referred to in previous works have been grouped into three major categories based on textual and linguistic features, behavioral features, and relational features. The third categorization includes a variety of opinion-spam detection methods with a focus on their results. Finally, a brief analysis has been presented summarizing the different views on opinion spamming. Furthermore, some of the recent advancements in the field of opinion spam detection are highlighted to show the current progress till date. After the rigorous review of the problem of opinion spam detection, we found some important research gaps for further improvements in this domain.

After getting insights from the literature review, an analysis of behavioral and textual characteristics of opinion spamming to distinguish between fake and genuine reviews has been performed in this thesis. The primary motive for this kind of analysis is to examine the impact of behavioral features - describing user's complex behavior on review platforms - and textual features - mainly focusing on writing patterns of users' text - on the performance of classifying spam reviews and review-spammers. This analysis is of great importance as it can help researchers in understanding complex spamming behavior of reviewers as well as writing clues of review spamming.

Thereafter, we propose several novel metadata and content-based features to effectively identify opinion spam. This research effectively utilizes all perspectives of opinion spamming by introducing one novel angle i.e., product-centric view, along with two common perspectives, review-centric and reviewer-centric, for opinion spam detection. Moreover, this work also deals with providing a novel solution to one of the major problems, i.e., class imbalance problem, in the field of opinion spam detection. Furthermore, an effective opinion spam detection model is introduced which utilizes some previously well-known opinion spamming features along with the novel set of proposed features in improving the performance over existing state-of-the-art approaches. This research can be used by the product manufacturers or online businesses in

designing effective opinion spam filters, and therefore, maintaining the trustworthiness of online opinions.

Furthermore, we examine the impact of feature selection by considering various important factors which can improve the performance in opinion spam detection domain. As feature engineering by using text mining and NLP techniques can generate hundreds or even thousands of content-based features, this leads to a high-dimensional feature space where many of the features might be irrelevant. This gives us a strong motivation to use proper feature selection techniques to significantly reduce the cost and improve performance of opinion spam filters. Although, machine learning has attracted significant research attention in the field of opinion spam detection, very limited attention is paid on feature selection and ranking. This research has focused on this neglected part of opinion spam detection by using several feature selection techniques for performance evaluation of several machine learning models. By providing useful insights on feature selection to further improve the detection performance in opinion spam domain, we believe this study is of immense importance to researchers and practitioners.

In this thesis, the proposed research work mainly considers fake reviews as most common type of opinion spam. However, this research can also be extended to identify different types of misinformation on web such as *fake news detection*, *rumor detection*, etc. . It will be of great interest to know how good the proposed solutions would work for different types of spamming content spreading on the internet.