Store Review Spammer Detection Based on Review Relationship

Qingxi Peng

State Key Laboratory of Software Engineering, Wuhan University, Luojiashan Road, 430072 Wuhan, China
pengqingxi@google.com

Abstract. Reviews and comments play important role in online shopping. It can help people getting more information about stores and products. The potential customers tend to make decision according to it. However, driven by profit and fame, spammers post spurious reviews to mislead the customers by promoting or demoting target store. Previous studies mainly focused on the text features and ratings to identify fake reviews. However, these studies ignore the importance of relationship between store and reviewer. This paper first proposes sentiment analysis techniques to calculate the sentiment score of reviews. Then a relationship-based method has been proposed to identify the spammers. We also present an algorithm which can detect both single-mode spammers and multi-mode spammers. A subset of highly suspicious reviewers is selected for evaluation by human judges. Experimental results show that the proposed method can find out the review spammer efficiently.

Keywords: review spam, review relationship, sentiment analysis.

1 Introduction

With the development of Web 2.0, online shopping has become a popular activity of many peoples life. Previously, people usually transfer experience within small scope such as friends and relatives. Nowadays, customers prefer to share their experiences about store and product by posting reviews, while potential customers tend to gain indirect experience by watching user generating reviews before online shopping. The online comments have also been used by stores to identify their product and service problems to attract more customers. However, Spammers publish spurious reviews to promote or demote target online store, inducing users buy or not to buy something from particular store. More and more spam reviews emerge in major review websites such as Epinion.com, Resellerrating.com, and Shopzilla.com. Therefore, detecting such spurious reviews and spammers become a pressing issue.

Previous works have paid more emphasis on text features of reviews. Similarity of text has been regarded as main evidence to identify the spam reviews. While cunning spammers avoid writing near duplicated text, the text-based methods dont work now. Other works use rating score and feedback score as indicator to
detect the spam reviews. In this paper, we study the relationship between store, reviewer and review. We also incorporate the sentiment analysis techniques into spam review detection.

There are two typical spammers: single-mode spammers and multi-mode spammers. Single-mode spammer is spammer who only uses one userid to post reviews. In [15], time series analysis has been employed to detect the single-mode spammers. Multi-mode spammers usually post several reviews using one userid [4]. Previous works either capture single-mode spammers by time-series analysis of reviews, or capture multi-mode spammers by means of review graph. Due to the analysis of relationship between stores and reviewers, our method can capture both single-mode spammers and multi-mode spammers, which is different from previous works.

In this paper, our researches mainly focus on the relationship between store and reviewer. Firstly we incorporate the sentiment analysis techniques to calculate the sentiment score. Then deviation of sentiment and rating has been analyzed by a relationship-based method. We also propose an algorithm to capture the spammers mentioned above. It is important to note that our idea is different from [9] even if we all use relationships. We map the reviewers and stores into a heterogeneous graph, and consider the entities and the relationship between them. To our knowledge, this is first study detecting both single-mode and multi-mode spammers.

Our contributions in this paper are then as follows:
1. We incorporate the sentiment analysis techniques into spam review detection, and propose a method to calculate the sentiment score of reviews.
2. We map the reviews and stores into a heterogeneous graph, and propose a relationship-based approach to calculate the deviation of sentiment and rating scores.
3. We propose an algorithm to detect both single-mode spammers and multi-mode spammers. Experimental results demonstrate the effectiveness of the proposed method.

The remainder of the paper is organized as follows. Section 2 discusses related work in the area of spam detection and their limitations. In Section 3, we first propose sentiment analysis method to calculate sentiment score of reviews. Then the deviation of review has been calculated. Finally, an algorithm has been proposed to capture both single-mode and multi-mode spammers by means of store-reviewer relationship. In section 4 we present an evaluation of the spammer detecting on the real world dataset. Finally, conclusions and directions for future work are given in section5.

2 Related Work

In this section we summarize the previous approaches, and how our method differs from them.