

CS Student Conference

Amanda Minnich
aminnich@cs.unm.edu

February 25, 2014

Abstract

In an age when many decisions about where to stay or what to buy are based on online reviews, businesses have much to gain by hiring spammers to write fake reviews that boost their overall product ratings. These fake reviews can have strong negative effects on consumers' purchasing experiences, and unfortunately they can be very difficult to detect. As with spam email and social spamming, there is an arms race occurring between spam review writers and spam review detectors.

Our research is focused specifically on hotel review spam. Our data consists of 16 million reviews for over 160,000 hotels from the three largest hotel review websites: TripAdvisor.com, Booking.com, and Hotels.com. By combining heterogeneous review sources for the same product, we can examine discrepancies in rating distributions, as well as abnormal patterns in spatial and temporal distributions of both reviews and users. Booking.com and Hotels.com both require the reviewer to book through their website in order to write a review. Conversely, TripAdvisor.com allows anyone to create an account and leave a review. We therefore hypothesize that this website contains the majority of the review spam. The other two websites will act as a baseline so we know what "normal" review patterns look like.

Our goal in this research is to identify suspicious users, hotels, and reviews. We hope that our results can contribute to both the creation of more rigorous review filters as well as the analysis of opinion spammers' general behaviors.