Finding Deceptive Opinion Spam by Any Stretch of the Imagination
Myle Ott – Yejin Choi – Claire Cardie – Jefferey T Hancock

Jayaram Gokulan – jg929 - Critique #1

Among the several interesting aspects introduced and elaborated in the study, perhaps the more compelling discussion would be how the experimental results from investigations helped corroborate and reinforce formerly established empirical beliefs and literature findings on human's inefficient authentication of review opinions. Spatial configurational details appearing as some of the top-weighted 'truthful' features captured by the automated LIWC-BIGRAMS combinational method is a classic example demonstrating how deceptive opinions falling short on including such particulars can be erroneously classified acceptable by human judges. Yet the more important development stemming from this detail and deeming better attention could be the witnessing of counter-intuitive features as outcomes as well, hinting at the limitations and strengths of the LIWC technique and BIGRAMS method respectively, all the while allowing readers to ponder over the larger issue of how reverse psychology can be a possible future challenge. For instance, the generally conceived notion of decreased usage of the first-person singular in deception reviews saw a reversal during the course of this study. While the report goes on to suggest considering motivation and context in feature selection as possible mitigatory strategies, the more pressing issue at hand would be comprehending the cognizance and adaptation of players in this field, to newly introduced detection techniques, via information dissemination (in the area of social psychology) i.e. if the review creators themselves can foretell the outcomes of even the most advanced deception detection algorithm and acclimatize accordingly, then it renders any, if not all, such detection implementations moot. Understandably, the effectiveness of the strategies we employ can be viewed varying in inverse proportions to the informational efficiency of deceptive opinion review writers, thus prompting us to take a more fundamental approach in figuring how reverse psychology tactics are formulated and executed under different settings of deception detection methods that have been introduced in this investigation.

Considering the general notion of how ML classifiers including Naïve Bayes and SVM are susceptible to feature correlation, perhaps a topic needing further research would be the data collection technique presented in this literary work. Data in this context correspond to attributes picked up by POS, LIWC, and N-gram methods from the gold standard deceptive opinion datasets. AMT-provided Turkers helped fabricate deceptive reviews, though little has been studied about the Turkers themselves, in the context of how the reviews they generate (and hence the features) possibly contribute to the phenomenon of feature correlation (if at all any). Some interesting deliberations to ruminate on, would be the details of the selection process for these Turkers – What are the criteria for choosing Turkers? Are there any processing stages to rectify correlation and select limited correlated features? Is there a random sampling procedure involved in their selection and how effective are they? And so on. Reviewing such information would provide improved clarity around the model efficacy and eventually help us better appreciate the experimental results provided in this study.

On a smaller note, the research study reflects on the difficulty of charting human performance using Mechanical Turks, noting their tendency to be manipulated by the very same monetary benefits, introduced to stimulate a candid response while verifying review opinions. University students are considered more reliable in that respect, observing how they stand to gain no monetary perks for the same task. Accepting this rationale, with limited information on parameters compromising student's performance during the task, cannot be fully justified and thus asks for better investigation into the issue. For instance, one can consider a contrasting example wherein students, amidst their university schedule and generally without an income source, are inclined to work more honestly given monetary incentives. Understanding such determining factors at play could help provide a more accurate picture of human performance in deceptive opinion detection and avoid misleading results.