A Study on Text-Score Disagreement in Online Reviews

In this paper, we focus on online reviews and employ artificial intelligence tools, taken from the cognitive computing field, to help understand the relationships between the textual part of the review and the assigned numerical score. We move from the intuitions that (1) a set of textual reviews expressing different sentiments may feature the same score (and vice-versa), and (2) detecting and analyzing the mismatches between the review content and the actual score may benefit both service providers and consumers, by highlighting specific factors of satisfaction (and dissatisfaction) in texts. To prove the intuitions, we adopt sentiment analysis techniques and we concentrate on hotel reviews, to find polarity mismatches therein. In particular, we first train a text classifier with a set of annotated hotel reviews, taken from the Booking website. Then, we analyze a large dataset, with around 160k hotel reviews collected from TripAdvisor, with the aim of detecting a polarity mismatch, indicating if the textual content of the review is in line, or not, with the associated score. Using well-established artificial intelligence techniques and analyzing in depth the reviews featuring a mismatch between the text polarity and the score, we find that-on a scale of five stars-those reviews ranked with middle scores include a mixture of positive and negative aspects. The approach proposed here, beside acting as a polarity detector, provides an effective selection of reviews-on an initial very large dataset-that may allow both consumers and providers to focus directly on the review subset featuring a text/score disagreement, which conveniently convey to the user a summary of positive and negative features of the review target.