

Project Review

Rating Extraction through Sentiment and Sarcasm Analysis

Joe Abley, jabley@uwo.ca

December 8, 2018

This is a review of [6], as presented in ECE 9603; the full project description and detailed results have not been made available. This review is submitted as part of the coursework for ECE 9603.

1 Project Summary

Many web pages provide facilities to collect feedback, e.g. from customers or users. In many cases this feedback is collected in the form of both a rating and as free-text comments. Ratings are observed frequently to have a J-shaped distribution, with extremely high and low ratings appearing inflated compared to the normal distribution that might be expected. In the case where both ratings and comments are available, it is observed anecdotally that the distribution of comments is often different from the distribution of ratings. This project aims to address the incoherence between the user star rating and user comments using machine learning techniques to construct a model trained using public datasets. The resulting model was tested on a public dataset of reviews obtained from Yelp [2].

The Yelp dataset was preprocessed to remove various items such as URLs and hashtags from individual reviews. The result was subsequently tokenised, tagged according to Part of Speech (POS) and subject to lemmatization, a technique from linguistics which involves grouping together various related forms of a single word such that they can be identified by a single item, known as the word's lemma.

Features were derived from each preprocessed review such as the number of particular parts of speech observed, the number as a proportion of sentence length and the SentiWordNet [3] score.

Binary classifiers were constructed based on the Sentiment140 [1] dataset, originally from Twitter, and the SARC [5] dataset, originally from Reddit. 21 hyper-optimised classification models and one Long Short Term Memory (LSTM) model were tested using 80:20 hold-out validation on a test set of 100,000 observations; LSTM was found to be the best approach for both applications. The LSTM model was further tuned using a grid search of 648 combinations of parameters, and achieved an F-Score of 0.737 (sentiment) and 0.702 (sarcasm).

When applied to the Yelp dataset, the model predicted ratings that formed a balanced, normal distribution based on the natural language comments in reviews, while the star ratings applied by the same reviewers formed the J-shaped distribution that was noted earlier to have been observed elsewhere.

2 Constructive Feedback

The project was eloquently presented, and paced to deliver a high volume of information effectively in a relatively short time. The examples given in particular were engaging and familiar, and allowed the normalisation and feature selection processes to be easily understood by an audience without a thorough grounding in natural language processing.

This study clearly represents an impressive amount of work. The grid searches used for model and neural network topology selection were comprehensive.

The problem statement framed the observed incoherence between summary ratings and comments well, but the objective to *address* them did not clearly identify an intended result. The work described in this project could be used as part of a user interface to influence the way that ratings are collected; it could be used to obviate the need for reviewer-supplied ratings; it could be used to solve business problems such as recommendations or promotions in place of, or in combination with, the ratings provided by reviewers. The degree of dissonance between the ratings and comments left by particular reviewers might even be used to help classify the reviewer. The possible applications of this work seem varied, and it would have been useful to identify some of them in the presentation.

References

- [1] About sentiment140. <http://help.sentiment140.com/>. Accessed: 2018-12-08.
- [2] Yelp dataset challenge. <https://www.yelp.com/dataset/challenge>. Accessed: 2018-12-08.
- [3] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *in Proc. of LREC*, 2010.
- [4] M. Khodak, N. Saunshi, and K. Vodrahalli. A large self-annotated corpus for sarcasm. Technical report, Princeton University, 2018.
- [5] D. Manias and I. Shaer. Rating extraction through sentiment and sarcasm analysis. ECE-9603 Presentation 2018-12-06, December 2018.