Using Star Ratings as Trend-Aware Performance Metrics

Kenta Takatsu, Caroline Chang

November 9, 2017

Abstract

In this paper, we devise a method to predict Yelp review star rating in a given time period for a particular restaurant category: Chinese restaurants. To create the best star rating system possible, we investigate which types of text embedding techniques give us the most useful feature of latent rating distributions, which types of classifiers (multi-feature linear regression and support vector machines) perform the best when predicting star ratings, and in what ways certain types of topics affect business performance. We present our findings of how multi-feature linear regression and support vector machines compare in their in conjunction with multiple different embedding generating methods. We also discuss the most effective time interval with which we can get the most successfully predicted star ratings and provide an analysis of the types of topics that act as indicators of business performance.

1 Introduction

Our main objective is to develop a streamlined star rating prediction system that can ultimately give users a long-term idea whether particular businesses are undergoing positive and negative trends in their own performance. Past research has successfully generated predicted star ratings after generating latent subtopics from review text using Latent Dirichlet Allocation (or LDA), but has not taken the next step as to use these predicted star ratings to make an overarching conclusion about performance in businesses in different categories (Huang et al. 2014). A key element to developing such a system lies in accurate prediction, which can only be achieved through developing a fine-tuned classifier that will be able to predict star ratings over consistent time intervals. Accumulation of such star ratings over specified time intervals will give us the means to properly identify performance trends for any given business in a pre-chosen category. Every individual assessment of a business consists of review text and a star rating. There are different topics that arise within each review text, and for similar businesses in the same category, past research has shown that certain types of topics tend to occur more given the type of category. We extract such topics from review text using methods such as LDA. To translate the linguistic context that we have for each review in our analysis, we need to come up with embeddings that essentially translate a corpus of text into a vector space. In this project, we try to three different methods for generating a topic vector: gensim, tfi-df, and word2vec. In order to have an effective classifier, the quality of the embeddings is significant. Though our goal is to ultimately determine business performance using our developed classifiers, we also want to investigate whether types of topics themselves could also act as litmus tests of business performance, due to prevalence, changing popularity over time, etc.

2 Research Design

The objectives that we focus on in this project are determining the best topic embedding methods and the most effective classifiers for star rating prediction. We also define the best time periods by comparing the best performance among the different bin sizes. Lastly, we infer time-sensitive business performance from the topics themselves and explore in what way the topics themselves can be used as indicators of business performance.

2.1 Time Interval Bins

In this section, we will briefly discuss the restraints on our data set and how we need to consider the idea of data-richness (time variance among the reviews of a particular business as well as the

number of reviews per business). We will talk about how we come up with a subset of reviews within a category to run our classifiers on and what all these reviews in our subset have in common. Next, we will detail the method of training and testing our classifiers over a certain time interval and then subsequently predicting the next time interval bin's star rating. We will then check the accuracy of that rating compared to original value. We will also discuss our method for selecting the consistent and relevant topics we need as features for our classifiers.

2.2 Building Embedding Vectors

In this section, we will discuss the specifics of LDA as well as our three constructed different techniques of generating embedding vectors that will ultimately be used as features in our classifiers.

2.3 Multi-feature Linear Regression, SVM Classifiers

In this section, we will discuss how we construct our classifier models, and how the embeddings are implemented in our classifiers where topics are the features and star ratings are the response variable. We will briefly list the factors that we could use as points of comparison: i.e. hyperparameters, time windows, and ultimately how the different embeddings affect the different classifiers.

3 Results

3.1 Initial Embedding Vector & Prediction Baseline (along with N+1 Bin Prediction Results)

In this section, we will discuss our preliminary attempt at star rating prediction using a 127-topic embedding vector with the two classifiers. We will show the frequency count of the topics across all the reviews in the Chinese restaurant category and subsequently show which topics we selected as prominent features. We will discuss our comparison of how both performed, and discuss why this method might be a bit simplistic (i.e. it does not account for radical shifts in business performance i.e. drastic downturn in business performance) On average, rating predictions from both multi-feature linear regression and support vector machines were roughly 2 stars off, which is very significant.

3.2 Embedding & Classifier Comparison

In this section, we will discuss which embedding in which classifier performed the best. We will specifically show a plot where we show how we predict the offsets of given time intervals from the overall average business rating. We will also compare time windows and hyper parameters from the classifiers.

4 Conclusion & Future Work (to be determined)