Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network

1 INTRODUCTION

In this paper, we attempt to analyze the customer reviews on women clothing e-commerce by employing statistical analysis and sentiment classification. We first analyze the non-text review features (e.g. age, class of dress purchased, etc.) found in the dataset, as an attempt to unravel any connection between them and customer recommendation on the product. Then, we implement a bidirectional recurrent neural network (RNN) with long-short term memory (LSTM) for classifying whether a review text recommends the purchased product or not, and for classifying the user review sentiment towards the product.

2 METHODOLOGY

2.1 Machine Intelligence Library

2.2 The Dataset

2.3 Data Analysis

2.3.1 Analysis on Univariate Distributions

2.3.2 Analysis on Multivariate Distributions

2.3.3 Multivariate Analysis and Descriptive Statistics

2.3.4 Word Frequency Distributions

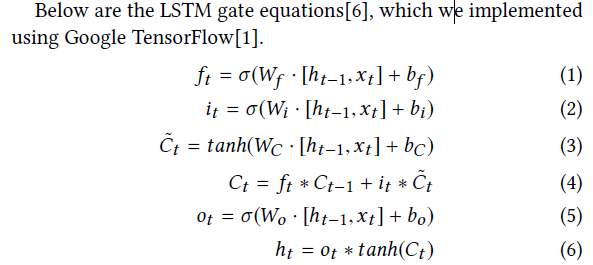
2.4 Dataset Preprocessing

2.4.1 Text Cleaning

2.4.2 Sentiment Analysis

2.4.3 Word Embeddings

2.5 Machine Learning



We employed this machine learning model on two text classification problems on the dataset: (1) recommendation classification which determines whether a review text recommends the reviewed product, and (2) sentiment classification, which determines the tone of the review text towards the purchased product.

2.5.1 Recommendation Classification

2.5.2 Sentiment Classification

3 RESULTS AND DISCUSSION

4 CONCLUSION AND RECOMMENDATION