Time Series Analysis of Twitter Data

Matthew Tiger

December 2, 2015

1. Introduction

With the advent of social media, the way in which society communicates has evolved. Popular social media networks document these new forms of communications and as a result a rich set of data surrounding these interactions emerges. In particular, Twitter remains one of the most popular social media platforms to date. Twitter was first launched in July 2006 as a social networking service designed to allow its users to communicate via short 140-character messages called "tweets". These tweets are sometimes affixed by the sender with a meta-label called a "hashtag", denoted by a string leading with the # symbol, that is meant to categorize the information contained in the message. Twitter currently has an active user base of 302 million users sending these categorized messages every second.

In this report, we will analyze data pertaining to a popular television show collected from Twitter's streaming API over the course of three weeks. We will then fit a time series model to this data and provide a forecasting model of the next week's projected data.

2. Data Description

3. Model Fitting

4. Forecasting

5. Conclusion