

Exchange Rate Prediction from Twitter's Trending Topics

Fulya Ozcan

University of California, Irvine

September 30, 2016

Abstract

This paper investigates whether incorporating sentiment extracted from Twitter's trending topics would improve the intra-day exchange rate predictions. What makes this paper unique is that unlike previous similar studies which only consider tweets that contain the symbol or the name of the currency or stock, it looks at all trending topics irrespective of whether they contain the name or the symbol of the currency. This allows to capture the general sentiment among the users, which implicitly affects the exchange rates.

This paper contributes to two different strands of literature in two disciplines: exchange rate prediction and machine learning. This is the first paper that makes use of data from social media to predict exchange rates with an unsupervised topic clustering model. Studies available in this literature mostly make use of news headlines as their data, and use neural networks or ensembles as their models. The novelty of this paper in exchange rate prediction literature is using Twitter data and also employing a Dirichlet process mixture algorithm to cluster the sentiment-based topics to use them in a time series model to predict the short term exchange rates. This paper is also the first paper that makes use of unsupervised learning in topic extraction, in exchange rate prediction using natural language processing as sentiment from all trending topics are extracted regardless of whether they contain the symbol of the currency in question.

Intra-day foreign exchange rates are predicted by making use of the trending topics from Twitter, using a sentiment based topic clustering algorithm. Twitter trending topics data provide a good source of high frequency information, which would improve the short-term or intra-day exchange rate predictions. This project uses an online dataset, where trending topics in the world are fetched from Twitter every ten minutes since July 2013. First, using a sentiment lexicon, the trending topics are assigned

a sentiment (negative, positive, or uncertain), and then using a continuous Dirichlet process mixture model, the trending topics are clustered regardless of whether they are explicitly related to the currency under consideration. This unique approach enables to capture the general sentiment among users, which implicitly affects the currencies. Finally, the exchange rates are estimated using a linear model which includes the topic based sentiment series and the lagged values of the currencies, and a VAR model on the topic based sentiment time series. The main variables of interest are Euro/USD, GBP/USD, Swiss Franc/USD and Japanese Yen/USD exchange rates. The linear model with the sentiments from the topics and the lagged values of the currencies is found to perform better than the benchmark AR(1) model. Incorporating sentiments from tweets also resulted in a better prediction of currency values after unexpected events.

One of the important results from this study is that although this approach cannot predict surprise events, since events need to be heard by the users before a user sentiment can be obtained, it performs better than the benchmark AR(1) model when predicting what will happen after such unexpected events. Moreover, when the topics are added as dummies to the lagged values of the currencies, the prediction errors are found to be lower than the ones from using only the lagged values AR(1). This shows the implicit effect of the sentiment among users on the currency values.

Keywords: Natural Language Processing, Dirichlet Mixture Models, Exchange Rate Prediction, Trending Topics, Twitter.