

An Approach for Pattern Recognition and Prediction of Information Diffusion Model on Twitter

Authors : Amartya Hatua, Trung Nguyen, Andrew Sung

Abstract : In this paper, we study the information diffusion process on Twitter as a multivariate time series problem. Our model concerns three measures (volume, network influence, and sentiment of tweets) based on 10 features, and we collected 27 million tweets to build our information diffusion time series dataset for analysis. Then, different time series clustering techniques with Dynamic Time Warping (DTW) distance were used to identify different patterns of information diffusion. Finally, we built the information diffusion prediction models for new hashtags which comprise two phrases: The first phrase is recognizing the pattern using k-NN with DTW distance; the second phrase is building the forecasting model using the traditional Autoregressive Integrated Moving Average (ARIMA) model and the non-linear recurrent neural network of Long Short-Term Memory (LSTM). Preliminary results of performance evaluation between different forecasting models show that LSTM with clustering information notably outperforms other models. Therefore, our approach can be applied in real-world applications to analyze and predict the information diffusion characteristics of selected topics or memes (hashtags) in Twitter.

Keywords : ARIMA, DTW, information diffusion, LSTM, RNN, time series clustering, time series forecasting, Twitter

Conference Title : ICDMKM 2018 : International Conference on Data Mining and Knowledge Management

Conference Location : Madrid, Spain

Conference Dates : March 26-27, 2018