



Event detection on large social media using temporal analysis

Abdulrahman Aldhaheri

Advisor: Professor Jeongkyu Lee

Computer Science & Engineering Department
University of Bridgeport, Bridgeport, CT

Abstract

The problem of event detection based on social media has attracted researchers' attention recently because of the enormous popularity of social media. Existing approaches focus on features that don't reflect full characteristics of the social network. For the purpose of this research, we define an event as an occurrence that has enough force and momentum that could create an observable change of the context of a social network. Such a definition provides us with a wider perspective through which we can view the big picture of the social network. In this research, we propose a novel framework for detecting events on social media. We introduce a temporal approach to detect structural change of the social network that reflects an occurrence of an event using machine learning algorithms. In this study, we show that processing temporal social networks captures the complete complexity of the social network, which results in a higher accuracy of event detection.

Introduction

In this study, we show that processing temporal social networks as a graph captures the complete complexity of the social network, which results in a higher accuracy of event detection model. We propose a temporal social network graphs event detection framework based on which we propose a novel social network transformation approach that transforms social media streams into temporal images. This allows for building a better event detection predictive model. We validate the proposed approach by performing experiments on streamed social media data collected for the purpose of this research. We achieve an accuracy rate in detecting events that surpasses existing approaches. We evaluate our proposed approach by using commonly used model evaluation metrics.. We calculated and compared accuracy, precision, recall, and F1-score. We also used precision-recall and ROC curves to evaluate the performance of our proposed approach.

Approach

Modeling social networks as a graph enables researchers to perform a wide range of social networks analysis. However, processing huge social networks imposes many challenges. We are proposing a novel approach to process social networks. By taking advantage of the adjacency matrix representation of the social network graph. We convert the social network graph into an $n \times n$ adjacency matrix, where n is the number of nodes in the social network. Then, we convert the adjacency matrix to a gray-scale image with an $n \times n$ pixels as shown in the following figure.

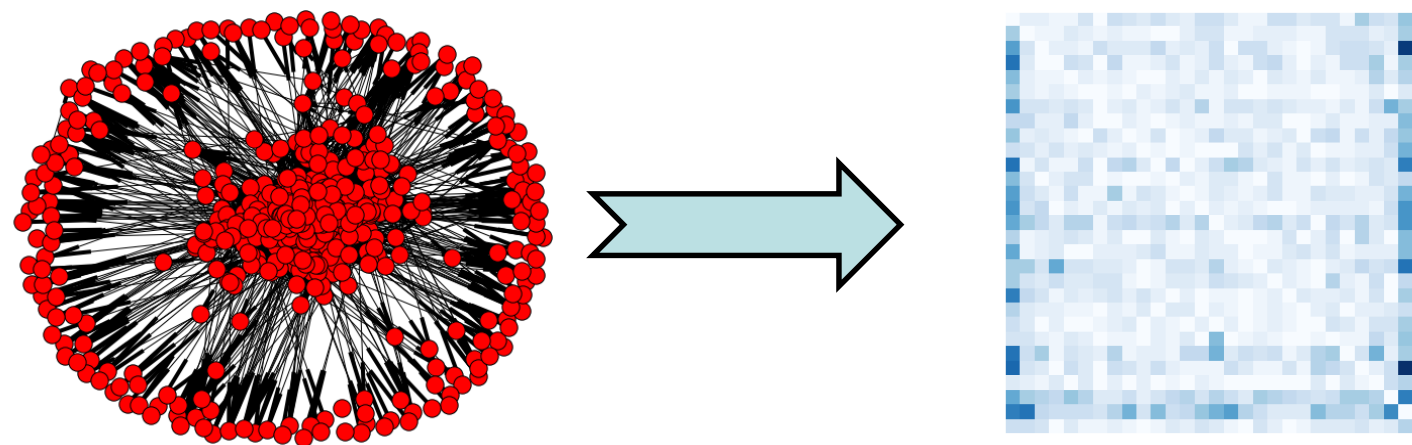


Figure 1: Converting a social network to a gray-scale image.

We scaled down the images using a well-known down sampling technique, i.e. max pooling. Max pooling is a down sampling technique that results in reducing the size of the image drastically. In the max pooling, a defined region with a dimension $p \times q$ returns a single value, which is the maximum value in the region. The following figure shows an image after applying max-pooling using different values of p and q .

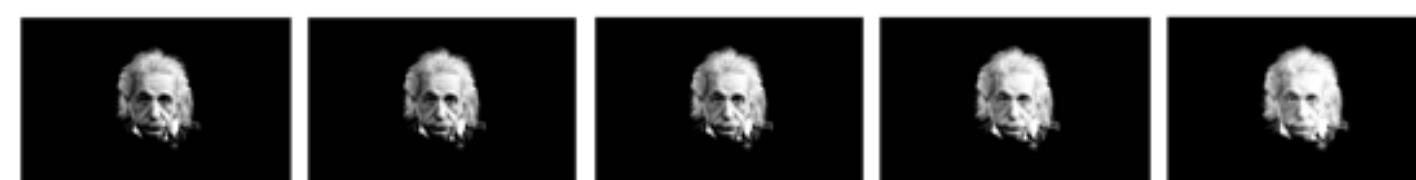


Figure 2: Down sampling an image. The original image on the left is being down sampled using values 2, 4, 6, and 8.

Figure 3 shows how the number of features is reduced drastically as the dimension of the region increase. Lowering the number of features without compromising the quality of the data allows the proposed approach to be scalable.

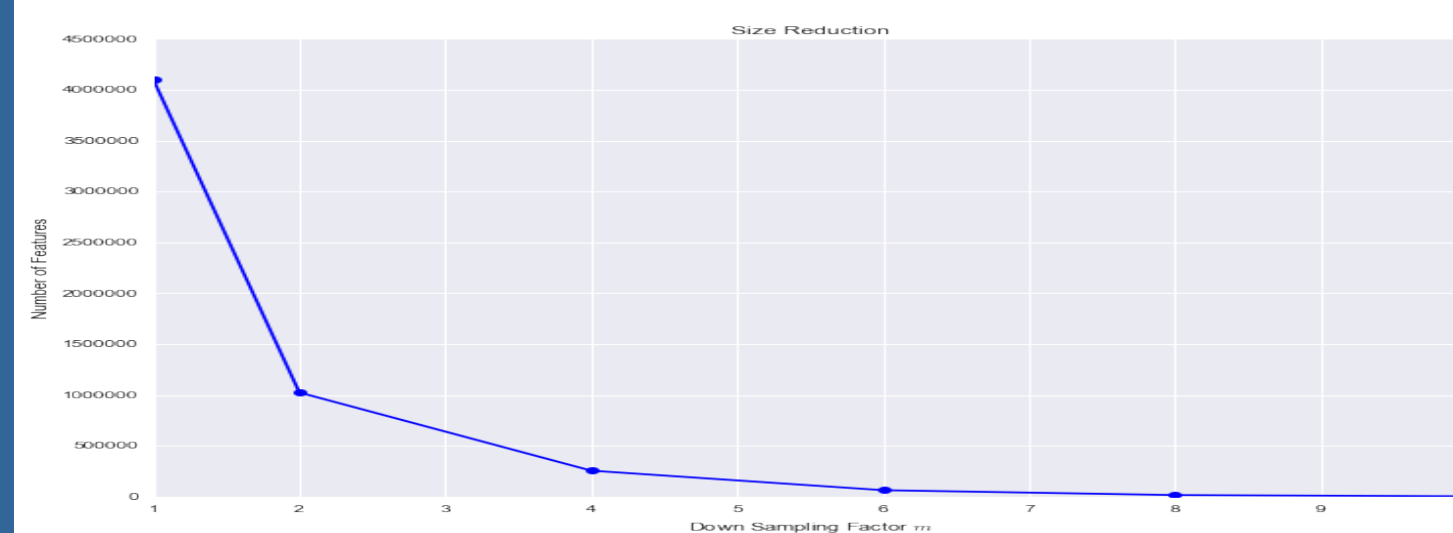


Figure 3: A drastic reduction in the number of features.

Figure 4 shows the overall frame work of the proposed model.

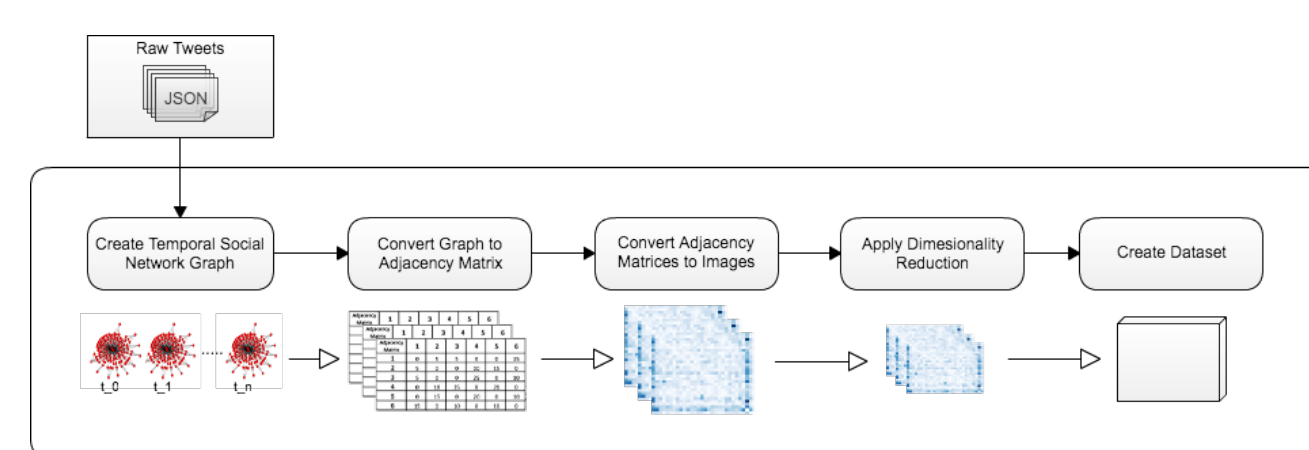


Figure 4: Social network transformation and compression

Results

We implemented the proposed framework on an actual dataset that we collected using Twitter public API. The dataset contains more than 4.4 million raw tweets with a size that exceeds 17Gb. Using the proposed framework, we extracted temporal social network graphs from the dataset and then transform them into temporal snapshots. We examine the impact of the applied dimensionality reduction technique on the performance of the event detection model. We applied a multilayers neural network on every dataset and recorded the model performance's results in the following table.

Region	Accuracy	Precision	Recall	F1-Score
100 × 100	0.970	0.980	0.980	0.980
110 × 110	0.985	1	0.981	0.990
120 × 120	0.978	0.991	0.981	0.986
130 × 130	0.949	0.981	0.953	0.967
140 × 140	0.963	0.982	0.973	0.978
150 × 150	0.993	0.991	1	0.995
160 × 160	0.956	0.990	0.953	0.971
170 × 170	0.978	0.973	1	0.986
180 × 180	0.949	0.991	0.948	0.969
190 × 190	0.978	0.991	0.983	0.987
200 × 200	0.971	0.982	0.982	0.982

Table 1: Impact of image down sampling on model performance

Conclusion

We introduce a temporal approach to detect structural changes in the social network that reflects an occurrence of an event using machine learning algorithms. In this study, we show that processing temporal social networks graphs captures the complete complexity of the social network, which results in a higher event detection model accuracy. We proposed a Temporal Social Network Graphs Event Detection framework based on a novel social network transformation approach that transforms social media streams into temporal images, which allows for building a better event detection predictive model. We evaluated the proposed model by applying well-known metrics used to evaluate data mining, document retrieval, and machine learning algorithms. We have also shown that the proposed framework detects events with very high accuracy.

References

A. Aldhaheri, and J. Lee. "Event detection on large social media using temporal analysis". In 2016 IEEE 7th Annual Computing and Communication Workshop and Conference, Las Vegas, USA, Jan. 2017.