Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M15-222P

Published: 6/1/2016

Inventors

Paulo Shakarian

Ruocheng Guo

Elham Shaabani

Abhinav Bhatnagar

Contact

Shen Yan
shen.yan@skysonginnovations.com

Cascade Prediction

In social media, the term “viral” refers to when a piece of information (microblog, photograph, video, link, etc.) starts to spread in a social network and increases in popularity by an order-of-magnitude. Much work is currently being done to better understand the patterns and implications of this new phenomenon. Previous studies have devised a regression model for viral media and noted a severe imbalance of the data. This is due to the power-law relationship between cascade size and frequency, which hinders the creation of a useful model. Additional studies have predicted viral cascades with high precision and recall, but unrealistically balance the data sets and define “viral” as cascades that only double in size. Therefore, there is a need for a tool that can accurately predict viral phenomena while realistically representing the data.

Researchers at Arizona State University have devised a suite of measurements based on “structural diversity” – the variety of social contexts or communities in which individuals partaking in a given cascade engage. This work presents a set of measurements that are not based on the specific content itself, but rather based on data about the post and the network topology that it spreads through. This tool demonstrates that these measurements are able to distinguish viral from non-viral cascades, despite the imbalance of the data. This approach also identifies if cascades observed for 60 minutes will grow to 500 reposts, and demonstrates the tradeoff between precision and recall. Overall, this method significantly outperforms current state-of-the-art techniques.

Potential Applications

- Social media
- Social networks
- Spread of information

Benefits and Advantages

- Improved Accuracy - Better accuracy in prediction of viral occurrences.
- Structural Diversity-based Prediction – Measurements are based on data about the post and the network topology that it spreads through rather than the specific content itself.
- Increased Precision and Recall –
  - Able to identify cascades of size 50 reposts that grow to 500 reposts with a precision of 0.69 and recall of 0.52 for the viral class.
  - Able to identify cascades that have advanced for 60 minutes that will reach 500 reposts with a precision of 0.65 and recall of 0.53 for the viral class.
  - Identifies tradeoffs, such that precision of 0.78 or recall of 0.71 were obtained at the expense of the other.
- Realistic Assessments - Results achieved while maintaining the imbalances of

the dataset - which better mimics reality.

For more information about the inventor(s) and their research, please see

Dr. Paulo Shakarian's directory webpage