

TargetVue

Analysis of Online Anomalous User

Hung-li Chen (Henry)

1

Target Vue: Visual Analysis of Anomalous User Behaviours in Online Communication Systems (TVCG, 2016)

Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, Ching-Yung Lin

First five authors are from IBM T.J. Watson Research Centre
The last author is from University of Pittsburgh not Pissburg...

2

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

3

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

4

Context

- Anomaly Detection is important.
- Challenging to find completely automated solutions

Contribution

- TargetVue: a system that **detects** and supports **interactive exploration** of **anomalous users**
- New **glyph design** and the grid layout
- Evaluation through a bot detection challenge and case a case study

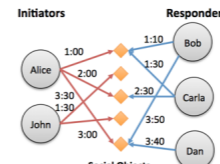
5

Agenda

- Context and Contribution
- Requirements, Data and Tasks**
- Design of TargetVue
- Evaluation and Comments

6

Data Model



- Initiator - Social Objects - Responders
- High-level features: Behaviour, Content, Interaction, Temporal, Network, User Profile
- Data: **time series** of **feature vectors** (for each user)

7

Requirements

- Feature Selection
- Anomaly Detection in Context
- Ranking Threats
- Learn from User Feedback

8

Tasks

- Showing the data **overview** and detection results
- Interpreting user behaviours from **different perspectives**
- Facilitate visual data **comparisons**
- Revealing users' impacts in social communication
- Easy **browsing** of raw data
- Flexible data labeling

9

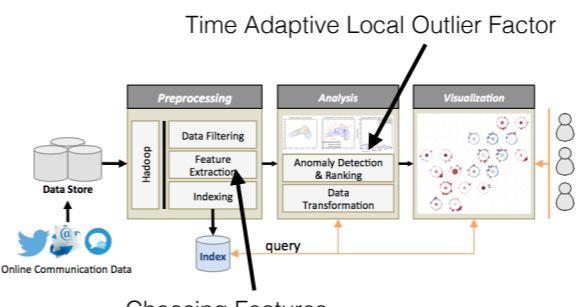
Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue**
- Evaluation and Comments

10

TargetVue: System Design

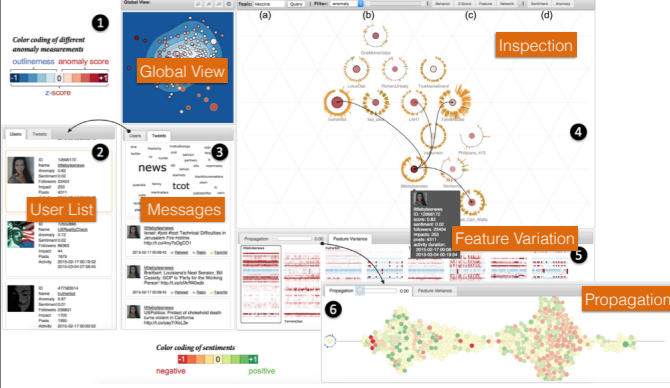
Time Adaptive Local Outlier Factor



Choosing Features

11


TargetVue: Interface



12

Global Encodings

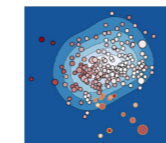
- Users** as circular nodes
- Importance** as the size of the nodes
- Sentiments** or **anomaly scores** as color



13

Global View

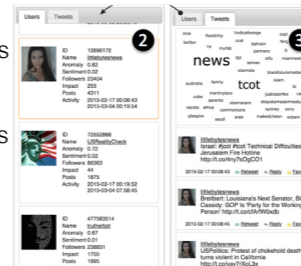
- Data: dimensionality reduced mean feature vector, kernel density estimation
- Encoding: location, contour map (white to blue)
- Task: overview
- Outliers are in the low density areas



14

User list and Messages

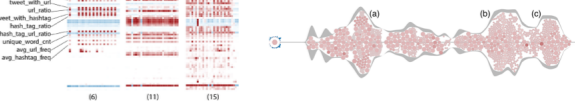
- Data: user profile information, raw messages
- Encoding: high frequency tag cloud, list of messages and user profiles
- Task: browsing raw data and overview



15

Feature Variation and Propagation View

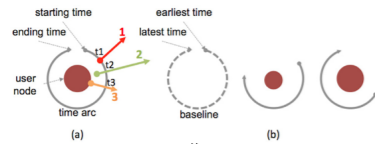
- Data: derived feature z-score (difference from baseline), users in communication threads
- Encoding: temporal heatmap, propagation view (introduced in FluxFlow)
- High impact users have many other users in the thread



16

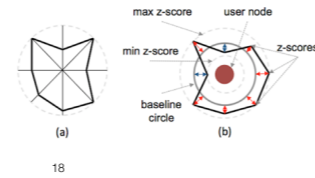
Inspection: Behaviour Glyph

- Data: posting and responding activity timeline, duration, number of users involved, sentiment of the threads
- Encoding: circular timeline, line mark (see below)
- line mark: thickness (number of users), length (duration), color (sentiment), intersection (time when the user join).



Inspection: Z-glyph

- Data: derived z-score of different features, (based on mean and standard deviation of features)
- Encoding: baseline circle, color coded area mark



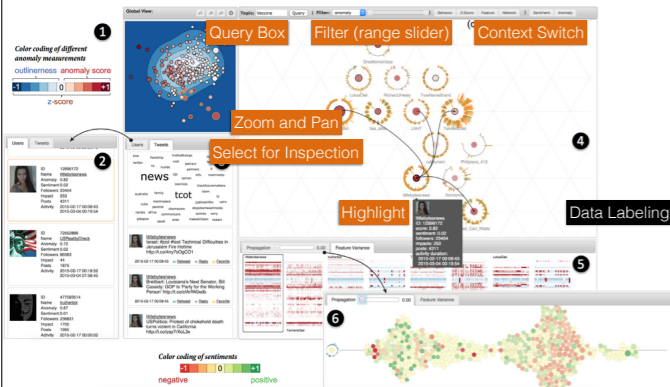
Inspection: Relation glyph

- Data: interaction relationships between users
- Encoding: directed links

Inspection: Layout

- Triangle mesh for placement
- Fast linear layout
- Preserve topology
- Maximize average similarity (clustering) between neighbouring glyphs

Interaction



Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

Evaluation

- The investigators used the system in social bot detection challenge.
 - Use global view to pick out outliers and anomalies
 - Inspect the users, and study their behaviour
 - Inspect specific features of users
 - Tune the model
- Example usage on Email data.
- Domain expert interview (2 experts): "Comprehensive", "very powerful"

Comments

- Delivers what are promised (Explicit reference to the requirements and tasks).
- Glyph design is information dense, effective for identifying anomalies, encoding may not be the most visually effective.
- Scaling limit is unclear (mentioned that the pipeline is built on hadoop, used the system for twitter data of 8000 users and 4M tweets)
- Evaluation in the future would be helpful.

Questions?