Temporal Opinion Spam Detection by Multivariate Indicative Signals

Problem

**Background**

Important to businesses’ revenue!

Pay spammers to write fake reviews

**Problem Statement**

Input: products’ review streams

Output: targeted products at time $t$

Approach

**Indicative Signals**

<table>
<thead>
<tr>
<th>Name</th>
<th>Range</th>
<th>Suspicous if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Rating</td>
<td>[1, 5]</td>
<td>Change</td>
</tr>
<tr>
<td>Number of (+/-) Reviews</td>
<td>[0, $\infty$]</td>
<td>Increase</td>
</tr>
<tr>
<td>Rating Entropy</td>
<td>[0, $\log_5 5$]</td>
<td>Decrease</td>
</tr>
<tr>
<td>Ratio of Singletons</td>
<td>[0, 1]</td>
<td>Increase</td>
</tr>
<tr>
<td>Ratio of First-timers</td>
<td>[0, 1]</td>
<td>Increase</td>
</tr>
<tr>
<td>Youth Score</td>
<td>[0, 1]</td>
<td>Increase</td>
</tr>
<tr>
<td>Temporal Gap Entropy</td>
<td>$[0, \max] + 1$</td>
<td>Decrease</td>
</tr>
</tbody>
</table>

With windows size $\Delta T$ and logarithmic binning, number of bins is $[\log_2 \Delta T] + 1$ and $\max = \log_2 ([\log_2 \Delta T] + 1)$.

Overview

1. Temporal Signal Extraction;
2. Anomaly Detection in Lead Signal:
   i. CUSUM for average rating;
   ii. Autoregressive model (AR) for others;
3. Anomaly Detection in Supporting Signals;
   i. Analyze local values only when “alarms” triggered by lead signal;
   ii. Use AR to detect anomalies;
4. Suspicousness Quantification;
   i. 4 features to characterize anomalies;
   ii. Integrate features into single value;

Contributions

- Problem formulation: Descriptive and general to detect spam reviews by monitoring indicative signals;
- New methodology: Online and efficient algorithm;
- Validated the method on real-world datasets.

- Code and Data available: http://www3.cs.stonybrook.edu/~juyye/

Experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Reviewer#</th>
<th>Product#</th>
<th>Review#</th>
<th>Start Date</th>
<th>End Date</th>
<th>Week#</th>
</tr>
</thead>
</table>

Case 1

- Product from iTunes
  - Burst in # of positive reviews: every 7 weeks;
  - Duplicate review texts, e.g.
    - “Great app for gamers”
    - “Great App For Gaming news”
    - “Must have app for gamers”
    - “One of my favorite apps”
  - Synchronized extreme ratings (5-star)

Case 2

- Product from Flipcart
  - Week 35 is detected suspicious;
  - Spammer group is detected;

Case 3

- Product from Flipcart
  - 125 5-star reviews in 2 days, while less than 10 reviews during other weeks;
  - These reviewers also reviewed another book, same author, same order during the same days!