Social Media Marketing: Unleash the Rising Stars

Chin Hui Chew, Ernest Chiew, Leonard Loo, Xinyi Ma
Partners: Ryan Lee, Christian Junge (Tribe Dynamics)
TF: Patrick Ohiomoba

Finding the right influencers has become an important and lucrative part of social media marketing for cosmetic brands. Rising influencers are of interest here because their impact on consumers will increase over time but their engagement cost is significantly lower. By probabilistically modeling user influence on social media graphs and successfully replicating real-life brand mention trends, we have been able to find the rising influencers. As a result we help marketers achieve similar reach as today’s highly paid top social media influencers, at a lower cost.

Introduction

Tribe Dynamics is a San Francisco based startup that measures social media engagement for cosmetics brands. They approached us with an interest in investigating how brand mentions spread across a social media network. Project goals include:

● Creating accurate models of social media influence reach
● Finding the top rising influencers to seed in a brand marketing campaign

Data

Data available include:

1. @-tag Mentions Data: Users, who tags whom, # times
2. Brand Hashtag Data: Users, brands, post count, EMV

Preprocessing

Our model is designed to run on individual brands. To extend our model to encompass all the brands, we perform brand segmentation: group brands based on user behavior and select representative brands from each group. Doing so decreases computational time and reduces overfitting.

Abstract

Finding the right influencers has become an important and lucrative part of social media marketing for cosmetic brands. Rising influencers are of interest here because their impact on consumers will increase over time but their engagement cost is significantly lower. By probabilistically modeling user influence on social media graphs and successfully replicating real-life brand mention trends, we have been able to find the rising influencers. As a result we help marketers achieve similar reach as today’s highly paid top social media influencers, at a lower cost.

Ensuring Our Estimated Reach Matches Actual Data

Reach: Number of activated nodes in a month

1) Constructing network graph with transition matrix, P for each month
   ● Each element in P denotes the probability of influence from one user to another user
   ● The closed-form solution is used here because the predicted reach most closely resembles the actual reach
   ● Note that the difference in the number of posts is used here such that we could accurately model probabilistically the influence from one user to another.

\[ P = \left( \begin{array}{ccc} 0.05 & \cdots & 0.3 \\ \vdots & \ddots & \vdots \\ 0.2 & \cdots & 0 \end{array} \right) \]

2) Cascading with Dynamic Threshold Model (DTM)

LTM Model from Literature | Our improvements: DTM
--- | ---
Initialization: \( \Sigma \alpha_{ui} \leq 1 \) where \( u \) is neighbor of \( v \)
Activation: \( \Sigma \alpha_{ui} \geq \beta(0,1) \)
Cleanup: No new activations between iterations

Dynamic Activation: \( \Sigma \alpha_{ui} \geq \beta(0,1) \)
Speed: Used matrix ops instead of for loops

3) Running DTM with P

Running DTM with our computed transition matrices P closely captures the observed trends for brands in different clusters:

Predicting on Future Data with Past Transition Matrices

1) With access to Ps from the previous months, we can use that to predict future reach. Our algorithm seeds the activated nodes in month t-1, and apply the P calculated from:

\[ P_{t-2|t-1} = (B_{t-2} - B_{t-2} \cdot B_{t-2})^{-1} B_{t-2} \cdot B_{t-1} \]

and DTM to predict reach in month t.
2) Multiply factors on predictions to better match the real world: For each month, the factor is calculated as

\[ \frac{\text{actual reach}}{\text{predicted reach}} \]

The graph shows our predicted reach being close to the actual reach.

Finding Rising Stars

Quality Reach: Summed PageRank scores of activated nodes in a month

1) DTM over all months over all users
   ● Use month t to predict month t+1 for each user
   ● Model this over the first 6 months (training set)

2) Model gradient of Reach and Quality Reach (Sum PageRank)

Gradient fitted with Linear Regression

PageRank scores measure the quality of reach, and are used to find rising stars.

3) Top K Rising Stars: Find one in each cluster

<table>
<thead>
<tr>
<th>Clustering algorithm</th>
<th>Counts per cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbor with k=5</td>
<td>787, 45, 7</td>
</tr>
<tr>
<td>Gaussian Mixture Models with 3 components</td>
<td>751, 71, 17</td>
</tr>
<tr>
<td>Affinity Propagation Clustering</td>
<td>839</td>
</tr>
</tbody>
</table>

No meaningful clusters on user position in network - most belong to one cluster.

4) Top K Rising Stars: Greedy Hill Climbing

● Bottom up approach
● Pick top 1 rising star, then iteratively pick rising stars till K chosen
● Took 8 minutes to run for each iteration (i.e. K=5 takes 40 minutes)

Our top 5 rising stars perform as well as top 5 influencers in after 3 months:

Conclusion

● Close matching of reality: Our closed form P and DTM matches the actual data very closely
● Accurate prediction: We can use historic P to accurately predict future reach given today’s seeds
● Rising stars: Computationally expensive algorithm but we found rising stars that caught up in quality reach with top influencers after 3 months

Reference

2. D. Kempe, J. M. Kleinberg, and ´E. Tardos. Maximizing the spread of influence through a social network. In Proc. of the Ninth ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD'03).