Recommending Investors for Crowdfunding Projects

Jisun An
University of Cambridge

with Daniele Quercia (Yahoo Labs Barcelona) and Jon Crowcroft (Univ. of Cambridge)
CROWDFUNDING

A founder proposes a project (e.g., smart watch, documentary, video game) and asks the Internet crowd for money.

More than 450 sites
Raised $2.8 billion in 2012
Bring creativity to life

Curious how Kickstarter works?

Learn more

With the help of his 961 backers, Jack published his first novel.

Staff Picks: Publishing

DEEP: The Story of Skiing and the Future of Snow
by DEEP: The future of skiing and snow in Jackson, WY

Help publish a book about the culture of skiing, the miracle of snow and how climate change could wipe out both in 75 years.

77% funded $19,294 pledged 7 days to go
On March 3, 2014, Kickstarter passed $1 BILLION in pledges. That’s $1,000,000,000 pledged by 5.7 million people to creative projects. More than half was pledged in the last 12 months alone.
PEBBLE WATCH

Pebble: E-Paper Watch for iPhone and Android
by Pebble Technology

Funded: This project successfully raised its funding goal on May 19, 2012.

68,929 backers
$10,266,845 pledged of $100,000 goal
0 seconds to go

Funding period
Apr 11, 2012 - May 18, 2012 (37 days)

Project by
Pebble Technology
Palo Alto, CA

Contact me
Check out the video

Our Kickstarter campaign is over, but you can still get a Pebble. Head over to www.getpebble.com for more info and to place an order.

May 8 - Pebble now supports Bluetooth 4.0!

If you're an app developer, big or small, please keep Pebble in mind! Sign up for our Developer's mailing list here.

High resolution photos for PRESS download here. For press inquiries, please contact media@getinpulse.com. Follow Pebble on Twitter @pebble and on Facebook.

April 12 pt 2 - Pebble is now even more water resistant! You can go swimming, run in the rain with Pebble.

April 12 update - We're absolutely blown away by your support, Kickstarter. $1M in 28 hours!

Darling Fireball - *The watch itself is a very cool idea; I'm in as a backer*

Forbes - *Proven track record...incredibly useful product*

Engadget - *Alerta intros Pebble smartwatch, inPulse's attractive younger sibling*

Wired Gadget Lab - *Smartwatches haven't really caught on with mainstream buyers -
KICKSTARTER

Not all projects are successfully financed.
Success rate: **43.85%** (by Kickstarter)

A recent study has found that “the majority of failed project creators cited the inability to *successfully leverage an online audience* as a main reason for failing.”
OUR GOAL

To propose automatic ways of matching Kickstarter founders with online investors
PROBLEM

We need to understand why people donate to which projects.

Founder cannot advertise through Kickstarter sites.
OUR GOAL

To propose automatic ways of matching Kickstarter founders with Twitter investors
METHODOLOGY

Step 1. **Crawling** Kickstarter sites and Tweets
   Project details, investors’ profiles, tweets

Step 2. **Characterizing** pledging behavior

Step 3. **Recommending** Twitter users given a project
   Predicting pledging behavior
   Ranking investors
DATASET

PLEDGING BEHAVIOR

RECOMMENDING INVESTORS
| DATASET | PLEDGING BEHAVIOR | RECOMMENDING INVESTORS |
DATASET

- **Data collection**
  - Scraped all projects featured on ‘Recently Launched Kickstarter page’ between July - October 2013
  - Regularly checked each project for any changes in pledged money and investors
  - Collected all tweets containing “kickstarter” or project title/URL

- **Focus on 1,149 USA based Kickstarter projects**
  - A total of 78,460 investors and their pledges (177,882) raised a total of $12.3M
  - 71,315 tweets relating to those projects
similar to success rate published by Kickstarter itself: 43.85%

<table>
<thead>
<tr>
<th></th>
<th>Successful</th>
<th>Failed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects</td>
<td>520</td>
<td>629</td>
<td>1,149</td>
</tr>
<tr>
<td>Proportion</td>
<td><strong>45.3%</strong></td>
<td>54.7%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Successful</th>
<th>Failed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal ($)</td>
<td>11,033.90</td>
<td>&lt; 30,716.86</td>
<td>20,875.38</td>
</tr>
<tr>
<td>Duration (days)</td>
<td>28.56</td>
<td>29.25</td>
<td>28.91</td>
</tr>
<tr>
<td>Number of investors</td>
<td>285.11</td>
<td>&gt; 47.09</td>
<td>166.10</td>
</tr>
<tr>
<td>Pledge ($)</td>
<td>79.71</td>
<td>60.13</td>
<td>68.99</td>
</tr>
<tr>
<td>Final amount</td>
<td>168.93%</td>
<td>&gt; 19.51%</td>
<td>94.22%</td>
</tr>
<tr>
<td>Number of tweets</td>
<td>101.93</td>
<td>&gt; 44.43</td>
<td>73.18</td>
</tr>
</tbody>
</table>
DATASET

78,460 investors

On average, investors supported three projects
DATASET  PLEDGING BEHAVIOR  RECOMMENDING INVESTORS
INVESTORS VS. DONORS

“20-40% of initial fundings in Kickstarter come from family and friends.”

Frequent Investors

Occasional Investors
Frequent investors tend to behave as an investor rather than a donor.

- **Pay Attention to Founder Skills**: 
  - # of updates,
  - # of comments,
  - # of reward levels,
  - and dedicated website

- **Invest in “High Capital” Projects**: project’s monetary goal

- **Invest in Geographically Global Projects**: Harvesine distance between the founder’s location and each investor’s location

- **Pay Attention to Fast Growing Projects**: growth rate per hour

- **Invest in Project Categories of Interest**: topical interests extracted from one’s tweets
## PLEDGING BEHAVIOR

**Hypotheses**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[H1]</td>
<td>A project is likely to be financed by frequent investors if its founder: [H1.1] frequently updates the project after launching it. [H1.2] answers the potential investors’ requests. [H1.3] allows for fine-grained funding levels. [H1.4] sets a dedicated web site.</td>
</tr>
<tr>
<td>[H2]</td>
<td>A project with a high goal is likely to be financed by frequent investors.</td>
</tr>
<tr>
<td>[H3]</td>
<td>A local project is likely to be supported by occasional investors.</td>
</tr>
<tr>
<td>[H4]</td>
<td>A fast-growing project is likely to be financed by frequent investors.</td>
</tr>
<tr>
<td>[H5]</td>
<td>Active investors tend to fund projects that match their own interests.</td>
</tr>
</tbody>
</table>
PLEDGING BEHAVIOR

Probability that an investor of type $B$ will fund a project of type $P$:

$$p(B|P) = \frac{p(B \cap P)}{p(P)}$$

- Investors of type $B$ who funded projects of type $P$
- All investors who backed projects of type $P$
PLEDGING BEHAVIOR

[H1] A project is likely to be financed by frequent investors if its founder:

[H1.1] frequently updates the project after launching it.
PLEDGING BEHAVIOR

[H1] A project is likely to be financed by frequent investors if its founder:

[H1.1] frequently updates the project after launching it. \( r=0.26 \)

[H1.2] answers the potential investors’ requests. \( r=0.19 \)

[H1.3] allows for fine-grained funding levels. \( r=0.05 \)

[H1.4] sets a dedicated web site. \( r=0.05 \)
PLEDGING BEHAVIOR

[H2] A project with a high goal is likely to be financed by frequent investors. $r=0.21$

[H3] A local project is likely to be supported by occasional investors. $r=0.32$

[H4] A fast-growing project is likely to be financed by frequent investors. $r=0.17$

[H5] Active investors tend to fund projects that match their own interests. $r=0.20$
PLEDGING BEHAVIOR

Those who have supported a considerable number of projects act in ways similar to how investors would do, while occasional supporters appear to be behaving as charitable donors.

Lured into Kickstarter by their own friends and family members who might happen to be on Facebook.

Probability that an investor supports a project as a function of the number of the project founder’s Facebook friends.
DATASET

PLEDGING BEHAVIOR

RECOMMENDING INVESTORS
| DATASET | PLEDGING BEHAVIOR | RECOMMENDING INVESTORS |
RECOMMENDING INVESTORS

1. **Linking** Kickstarter users to Twitter accounts
   Matching the names of Kickstarter users interested in a project with Twitter users mentioning the project.
   7,429 investors who are on Twitter with 891 projects they funded

2. **Predicting** pledging behavior *(who funds what)*
   Using Logistic Regression *(LR)* and Support Vector Machine *(SVM)*
   with three kernels: linear, polynomial, and RBF *(Radial Basis Function)*

3. **Ranking** investors
Dependent variable
whether
the investor supports
the project
(prediction is 1)
or not
(prediction is 0)

Independent variable

- Static
  - Project feature: project’s pledging goal, reward level, category.
  - Investor feature: Past supported project categories and his/her interests expressed on Twitter

- Dynamic
  - Project feature: growth rate, number of project updates, geographic dispersion of investors, and the number of comments exchanged
Problem
Our data only include positive cases—that is, the set of pledges that actually happened.

Solution

Augment our dataset with negative cases:
adding an equal number of negative cases (50-50 split)
(a set of random project-investor pairs)
PREDICTION WITH BALANCED DATASET

5-fold cross validation

SVM with polynomial and RBF kernels work best

82% of accuracy in predicting an unordered list of investors only by static features and 73% of accuracy by dynamic features.

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>ACC</th>
<th>P</th>
<th>R</th>
<th>$F_1$</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Static</td>
<td>0.57</td>
<td>0.57</td>
<td>0.55</td>
<td>0.56</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.57</td>
<td>0.58</td>
<td>0.55</td>
<td>0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>SVM-linear</td>
<td>Static</td>
<td>0.58</td>
<td>0.60</td>
<td>0.51</td>
<td>0.55</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.58</td>
<td>0.60</td>
<td>0.50</td>
<td>0.55</td>
<td>0.58</td>
</tr>
<tr>
<td>SVM-poly</td>
<td>Static</td>
<td>0.80</td>
<td>0.81</td>
<td>0.75</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.68</td>
<td>0.76</td>
<td>0.54</td>
<td>0.63</td>
<td>0.68</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>Static</td>
<td><strong>0.82</strong></td>
<td><strong>0.79</strong></td>
<td><strong>0.83</strong></td>
<td><strong>0.82</strong></td>
<td><strong>0.81</strong></td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td><strong>0.73</strong></td>
<td><strong>0.75</strong></td>
<td><strong>0.68</strong></td>
<td><strong>0.71</strong></td>
<td><strong>0.73</strong></td>
</tr>
</tbody>
</table>
Run classifications on input of different combinations of features:

- Number of comments (C),
- Reward levels (R),
- Geographic span (S),
- Growth rate (G),
- Category matching (E),
- Topic similarity (TS)

Adding category matching and topical similarity results in considerable performance improvements.
Frequent investors: projects on technology, games, and comics
<-> Occasional investors: art projects
PREDICTION WITH IMBALANCED DATASET

Creating an alternative test set:
20% positive cases and 80% negative cases (20/80 split)

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>ACC</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Static</td>
<td>0.56</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>SVM-linear</td>
<td>Static</td>
<td>0.60</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.61</td>
<td>0.59</td>
</tr>
<tr>
<td>SVM-poly</td>
<td>Static</td>
<td>0.81</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.77</td>
<td>0.70</td>
</tr>
<tr>
<td><strong>SVM-RBF</strong></td>
<td><strong>Static</strong></td>
<td><strong>0.82</strong></td>
<td><strong>0.81</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Dynamic</strong></td>
<td><strong>0.74</strong></td>
<td><strong>0.73</strong></td>
</tr>
</tbody>
</table>
RANKING INVESTORS

Using our SVM-RBF,
Rank all Twitter users for each project

Probability that B will fund P
= SVM-RBF (investor B features, project P features)

For a given project,

<table>
<thead>
<tr>
<th>All Twitter users</th>
<th>Probability</th>
<th>Funded</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9</td>
<td>Yes</td>
</tr>
<tr>
<td>C</td>
<td>0.7</td>
<td>Yes</td>
</tr>
<tr>
<td>D</td>
<td>0.6</td>
<td>No</td>
</tr>
<tr>
<td>E</td>
<td>0.3</td>
<td>No</td>
</tr>
<tr>
<td>B</td>
<td>0.2</td>
<td>No</td>
</tr>
</tbody>
</table>
RANKING INVESTORS

Using our SVM-RBF, Rank all Twitter users for each project

Then measure: MeanRR (Mean Reciprocal Rank) and MaxRR (Maximum Reciprocal Rank)

\[
\text{rank} = \frac{\sum_{i,P} \text{funded}_{i,P} \times \text{rank}_{i,P}}{\sum_{i,P} \text{rank}_{i,P}}
\]

- a flag that reflects whether investor \( i \) has supported project \( P \)
- the percentile-ranking of investor \( i \) within the ordered list of investors predicted for project \( P \)
### RANKING INVESTORS

<table>
<thead>
<tr>
<th>All Twitter users</th>
<th>Probability</th>
<th>Funded</th>
<th>rank</th>
<th>funded</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9</td>
<td>Yes</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0.7</td>
<td>Yes</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0.6</td>
<td>No</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0.3</td>
<td>No</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0.2</td>
<td>No</td>
<td>0.9</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
rank = \frac{\sum_{i,P} funded_{i,P} \cdot \text{rank}_{i,P}}{\sum_{i,P} \text{rank}_{i,P}} = \frac{0.1 + 0.3}{0.25} = 0.16
\]
RANKING INVESTORS

33% gain over the random baseline in predicting an ordered list

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>MeanRR</th>
<th>MaxRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>-</td>
<td>0.50</td>
<td>0.87</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>Static</td>
<td>0.34</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.37</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0.32</td>
<td>0.38</td>
</tr>
</tbody>
</table>
COLD-START PROBLEM

Extend investor pool

The Twitter-derived features

1. **Activity**: the logarithm of the total number of tweets
2. **Status**: the logarithm of the total number of followers divided by the number of followees
3. **Influence**: the sum of the average number of retweets, favorites, and mentions of the account’s tweets

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>ACC</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM-RBF</td>
<td>Static</td>
<td>0.68</td>
<td>0.71</td>
<td>0.61</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.67</td>
<td>0.72</td>
<td>0.58</td>
<td>0.64</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>model</th>
<th>Features</th>
<th>MeanRR</th>
<th>MaxRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>-</td>
<td>0.50</td>
<td>0.87</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>Static</td>
<td>0.44</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>0.44</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0.40</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Prediction accuracy (69%)  
Ranking performance (20% gain)
Artistic projects should rely on the traditional 3Fs (friends, family, and fools), employing social media sites to efficiently reach them.

Technology projects should broaden their search and look for active and frequent investors.
PLEDGING BEHAVIOR

**Occasional vs. Frequent**
Number of supported projects

- Matching interests

**Investors**

**Founder**
- Management skills
  - Updates
  - Comments
  - Fine-grained reward
  - Web sites

**Project**
- Pledging goal
  - Monetary goal
- Local vs. Global
  - Geographic dispersion
- Growth
  - Growth rate per hour
Related work on predicting success of crowdfunding projects

Not all projects are successfully financed. Success rate: 43.85% (by Kickstarter)