Analyzing the Quality of Information Solicited from Targeted Strangers on Social Media

Jeffrey Nichols, Michelle X. Zhou, Huahai Yang, Jeon Hyung Kang, Xiaohua Sun
IBM Research, USC ISI, Tongji University
jwnichols@us.ibm.com
@jwnichls
Asking questions on social media...

What’s a good digital camera?
Can we leverage the large amount of data created by public social networks?

Results for my nikon

CBxBacquel Christian Bacquel
I won't even lie, I haven't touched the Nikon for a shoot of my own in months. But there's reasons... GOOD reasons.
18 minutes ago

vinaykashy Vinay Kashyap
Inspired frm @Flyfiddlesticks "Might" use my dad's Nikon SLR during @ashok1282 Engagement this Sunday!
1 hour ago

lynnhill99 Lynn Hill
Nikon's New Cool DSLR Camera: Trouble is, since I am taking my sweet time deciding, I find that the advances in ... bit.ly/woh6t2
1 hour ago

treyhyde Richard "Trey" Hyde
My Nikon D90, didn't work in GPS, used LEVER event built-in.
Where might this be helpful?

• Questions that have spatial and/or temporal specificity (e.g., about an event)

• Questions for which there might be a diversity of opinion

• More?
How feasible is this approach?

• Will people answer questions from strangers?
• Will use of an incentive increase responses?
• What is the quality of the answers?
How feasible is this approach?

- Will people answer questions from strangers?
- Will use of an incentive increase responses?
- What is the quality of the answers?
How feasible is this approach?

- Will people answer questions from strangers?
- Will use of an incentive increase responses?
- What is the quality of the answers?
Scenario: Crowdsourcing Product Reviews

Step 0. Develop list of questions about a product or service

Goal: Collected answers to all questions should give a good product review

- How is the image quality?
- Does it take good low light pictures?
- How durable is it?
- What accessories are must haves?
- Etc...

Step 1. Identify owners or users of a product or service on Twitter

Step 2. Ask a question from list

Step 3-4. Ask more questions if user responds

Step 5. Visualize results as structured product review (future work)
Why Product Reviews?

Study:
Can utilize/compare with existing review data sets
  • Amazon.com, Yelp, etc.
  • Determine questions to ask, compare resulting reviews

Concept:
Different Properties Than Other Review Methods
  • More control
  • Structured by questions
  • Question list can be easily augmented
Product Review Scenarios

Samsung Galaxy Tab 10.1

• Popular consumer electronics product at the time of the study (didn’t want to use iPad)
• Compared to reviews from Amazon.com

L.A.-area Food Trucks

• Vibrant scene and Twitter is a primary means of communication
• Food trucks usually identified in tweet by @handle
• Compared to reviews from Yelp.com
Samsung Galaxy Tab 10.1

Questions

• 2 iterations
• First round Qs based on CNET and Engadget editor reviews
• Second round modified based on top 10 user reviews of tablets on Amazon.com

Procedure

• Identified users from real-time twitter stream
• Keywords and then manual human inspection
• Questions chosen semi-randomly based on content of tweet, answers received so far

Round #2 Questions

<table>
<thead>
<tr>
<th>Q#</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How fast is it?</td>
</tr>
<tr>
<td>2</td>
<td>How does it feel?</td>
</tr>
<tr>
<td>3</td>
<td>How is the display?</td>
</tr>
<tr>
<td>4</td>
<td>How is the camera quality?</td>
</tr>
<tr>
<td>5</td>
<td>How are the speakers?</td>
</tr>
<tr>
<td>6</td>
<td>How is the battery life?</td>
</tr>
<tr>
<td>7</td>
<td>Where do you use yours most?</td>
</tr>
<tr>
<td>8</td>
<td>Who would you recommend it for?</td>
</tr>
<tr>
<td>9</td>
<td>How easy is it to carry around?</td>
</tr>
<tr>
<td>10</td>
<td>Are you finding all of the apps that you want?</td>
</tr>
<tr>
<td>11</td>
<td>Is it easy to personalize the device for your use?</td>
</tr>
<tr>
<td>12</td>
<td>What do you use it for the most?</td>
</tr>
<tr>
<td>13</td>
<td>How easy is it to connect to other devices?</td>
</tr>
</tbody>
</table>
Samsung Galaxy Tab 10.1 Example

Dr. Sarmad Qureshi @SarmadQureshi  

Dec 20, 2011

Enjoying the bluetooth keyboard for my Galaxy Tab 10.1!!!!!!

Tabletsqa @tabletsqa

Dec 20, 2011

@SarmadQureshi Trying to learn about tablets... sounds like you have Galaxy Tab 10.1. How easy is it to connect to other devices?

Dr. Sarmad Qureshi @SarmadQureshi

Dec 24, 2011

@tabletsqa very easy. Since it got a software update recently

Tabletsqa @tabletsqa

Jan 1, 2012

@SarmadQureshi Thanks! What do you use it for the most?

Dr. Sarmad Qureshi @SarmadQureshi

Jan 2, 2012

@tabletsqa emails, facebook, twitter, youtube and sometimes for watching movies while in bed :-(

Tabletsqa @tabletsqa

Jan 2, 2012

@SarmadQureshi Thanks!
Los Angeles Food Trucks

Questions

- Based on our own intuitions of what information would be interesting

Procedure

- Identified users from real-time twitter stream
- @handles for food trucks and then manual human inspection
- Asked questions for 90 active LA food trucks at time of study
- Most traffic was concentrated for just three (Kogi Taco, Grilled Cheese, and GrillEmAll), and we report results only for those

<table>
<thead>
<tr>
<th>Q#</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What do you prefer to order?</td>
</tr>
<tr>
<td>2</td>
<td>Are there vegetarian options?</td>
</tr>
<tr>
<td>3</td>
<td>Is the menu large or small (for a food truck)?</td>
</tr>
<tr>
<td>4</td>
<td>How is the service?</td>
</tr>
<tr>
<td>5</td>
<td>Does the price match the amount of food you get?</td>
</tr>
<tr>
<td>6</td>
<td>Is it clean?</td>
</tr>
<tr>
<td>7</td>
<td>Does it move a lot, or is it often in the same places?</td>
</tr>
<tr>
<td>8</td>
<td>How far would you travel to get food from this truck?</td>
</tr>
</tbody>
</table>
Los Angeles Food Trucks Example

Gloria Vega @zooey02
Heading to @irvinelanes after work. Can’t wait to see @chunknchip and @FreshFriesLA! It’s been too long. :)

Food Truck Q&A @foodtruckqa
@zooey02 Interested in FreshFriesLA ...sounds like you’ve eaten there. Is it clean?

Gloria Vega @zooey02
@foodtruckqa Yes! @FreshFriesLA is one of the best trucks I’ve eaten ate. U should try them..and that’s an order... :)

Food Truck Q&A @foodtruckqa
@zooey02 Thanks! How far would you travel to get food from this truck?

Gloria Vega @zooey02
@foodtruckqa Usually don’t travel far. The farthest I’ve gone is Cerritos...I’ve known to hit up 2-3 lots in a dinner service. :)

Food Truck Q&A @foodtruckqa
@zooey02 Thanks!
Question Asking Dashboard*

Keyword-filtered stream

Responses

User’s recent tweets

* This is actually the TSA Tracker dashboard [Nichols and Kang, CSCW 2012], but the product reviews version was very similar
Quality Evaluation Methods

Human Coding
- Twitter responses & Traditional Reviews
- Relevance of response
- Information Types

Information Entropy
- Comparison between Twitter/Amazon, Twitter/Yelp

Mechanical Turk Questionnaire
- Usefulness, Objectiveness, Trustworthiness, Balance, Readability
Results...
### Response Rates/Times

<table>
<thead>
<tr>
<th></th>
<th>Tablet</th>
<th>Food Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Questions</strong></td>
<td>633</td>
<td>171</td>
</tr>
<tr>
<td><strong>Responses</strong></td>
<td>183</td>
<td>70</td>
</tr>
<tr>
<td><strong>Response Rate (%)</strong></td>
<td>29%</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Median Response Time (min)</strong></td>
<td>180.5</td>
<td>34.7</td>
</tr>
<tr>
<td><strong>% Responses in 30 Min</strong></td>
<td>36%</td>
<td>46%</td>
</tr>
<tr>
<td><strong>95% Responses Received (hrs)</strong></td>
<td>128</td>
<td>128</td>
</tr>
</tbody>
</table>

**Graphs:**
- Tablet: First Question vs Follow-up Question
- Food Truck: First Question vs Follow-up Question
# Response Quality (Coding)

## Overall Breakdown

<table>
<thead>
<tr>
<th></th>
<th>Response Count</th>
<th>Relevant Answer</th>
<th>Wrong Answer But Useful Info</th>
<th>Multi-Message Response</th>
<th>Average Info per Response</th>
<th>Off-topic Info per Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tablet</td>
<td>258</td>
<td>71%</td>
<td>19%</td>
<td>3%</td>
<td>1.82</td>
<td>0.48</td>
</tr>
<tr>
<td>Food Truck</td>
<td>111</td>
<td>82%</td>
<td>6%</td>
<td>6%</td>
<td>1.69</td>
<td>0.46</td>
</tr>
</tbody>
</table>

## Irrelevant Response Breakdown

<table>
<thead>
<tr>
<th></th>
<th># Irrelevant Responses</th>
<th>No Experience</th>
<th>Didn't know or understand</th>
<th>Thinks we're a bot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tablet</td>
<td>75</td>
<td>63%</td>
<td>11%</td>
<td>7%</td>
</tr>
<tr>
<td>Food Truck</td>
<td>20</td>
<td>25%</td>
<td>30%</td>
<td>0%</td>
</tr>
</tbody>
</table>
### Information Entropy

#### All Information

<table>
<thead>
<tr>
<th>Information (bits)</th>
<th>Tablet</th>
<th>Food Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amazon</td>
<td>Twitter</td>
</tr>
<tr>
<td></td>
<td>4.25</td>
<td>3.76</td>
</tr>
</tbody>
</table>

#### Information In Both Sets

<table>
<thead>
<tr>
<th>Information (bits)</th>
<th>Tablet</th>
<th>Food Truck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amazon</td>
<td>Yelp</td>
</tr>
<tr>
<td></td>
<td>4.09</td>
<td>3.27</td>
</tr>
<tr>
<td></td>
<td>Twitter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.73</td>
<td></td>
</tr>
</tbody>
</table>

The Twitter method is dependent on the questions
- Despite trying to base our questions on the contents of Amazon reviews, those reviews still contained more information.
- Our food truck questions went beyond Yelp reviews

* Calculated using a shrinkage entropy estimator
Mechanical Turk Evaluation

Completion Times

- Tablet
  26.5 minutes for Amazon
  25.8 minutes for Twitter
- Food Truck
  19.9 minutes for Yelp
  16.8 minutes for Twitter

Explanation of Results

- Few concrete examples of experiences in Twitter answers
- Limited information about Twitter reviewers

<table>
<thead>
<tr>
<th>Tablet</th>
<th>Amazon</th>
<th>Twitter</th>
<th>Mann-Whitney</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>3.19</td>
<td>2.64</td>
<td>868.5</td>
<td>0.006</td>
</tr>
<tr>
<td>Objectiveness</td>
<td>2.94</td>
<td>2.53</td>
<td>814.5</td>
<td>0.042</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>2.94</td>
<td>2.39</td>
<td>861.0</td>
<td>0.008</td>
</tr>
<tr>
<td>Balance</td>
<td>3.00</td>
<td>2.11</td>
<td>936.0</td>
<td>0.001</td>
</tr>
<tr>
<td>Readability</td>
<td>2.92</td>
<td>2.61</td>
<td>741.5</td>
<td>0.270</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food Truck</th>
<th>Yelp</th>
<th>Twitter</th>
<th>Mann-Whitney</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>2.86</td>
<td>2.56</td>
<td>734.0</td>
<td>0.309</td>
</tr>
<tr>
<td>Objectiveness</td>
<td>2.17</td>
<td>2.08</td>
<td>672.0</td>
<td>0.783</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>2.58</td>
<td>2.14</td>
<td>800.5</td>
<td>0.071</td>
</tr>
<tr>
<td>Balance</td>
<td>2.47</td>
<td>1.72</td>
<td>921.0</td>
<td>0.002</td>
</tr>
<tr>
<td>Readability</td>
<td>2.89</td>
<td>2.11</td>
<td>896.0</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Conclusions

• Answer quality is fairly high for both domains: 70-80%

• Quality seems to be tied to targeting accuracy
  • Most “bad” answers come from people who didn’t know the answer to our question

• For Product Reviews: Not yet comparable to other methods
  • Improvements to question method and visualization may help
  • Structure questions to elicit more context around existing answers
What’s next?

- Can we infer who will answer and who will not? Coming soon at IUI 2013!
- Can we build a tool to support this process?
- Or a crowdsourced product review tool that let’s people enter questions and helps them find answers?
- What will be the impact of this approach on social networks?
Thanks!

For more information, contact:
Jeffrey Nichols
jwnichols@us.ibm.com
@jwnichls
Engagement Continuum

qCrowd

Manual

Assisted

Automatic

Humans do all the work

• Keyword filtering
• Unstructured engagement
• Domain-independent analytics

Analytics streamline decisions: “press button to engage”

• Scenario-based filtering
• Smart engagement recommendations (e.g., based on location inference)
• Customizable engagement scenarios
• Domain-specific analytics

System-driven engagement

• Rule-based engagement
• Exception identification and notification
• Intelligent transition to human-driven engagement as desired