Discovering Next Generation Product Innovations By Identifying Lead User Preferences Expressed Through Large Scale Social Media Data

Tuesday, August 19th, 2014

Suppawong Tuarob
Computer Science and Engineering

Conrad S. Tucker
Engineering Design and Industrial and Manufacturing Engineering

The Pennsylvania State University
{suppawong,ctucker4}@psu.edu
PRESENTATION OVERVIEW

• Background
• Motivation
• Methodology
  ▪ Extracting Product Features
  ▪ Identifying Latent Features
  ▪ Identifying Product Specific and Global Lead Users
• Case Study
• Results
• Conclusions
• Future Work
A lead user is a consumer of a product who faces needs unknown to the public.

Lead users have two primary characteristics:
- High incentives to solve problems → Innovators
- Ahead of the market

Franke et al. (2006), Schreier et al. (2007)

Everett (2010)
von Hippel (2011)
ACQUIRING LEAD USER INSIGHTS

Existing Approaches: Acquire lead user knowledge through surveys, focus group interviews, or frequent interaction with the customer

CHALLENGES OF EXISTING TECHNIQUES

• Cost and scalability challenges of acquiring lead user insights

• Limited incentive to participate in a lead user study (from the user’s perspective)
**Hypothesis:** topics expressed through social media networks approximate lead user needs with high accuracy
Lead User Need Acquisition

User $X_j$

Unknown Needs $\mathbf{Y}$

Social Network Model

Latent features expressed through social media

Research Hypothesis

Users Post $m_j$

Word Vector $v_j$

Topic $t_k$
METHODOLOGY

**Objective 1: Extract Product Features**

- **Product Specs Documents**
- **Social Media Data**

**Objective 2: Identify Latent Features**

- **Ground-truth Features**
- **User-Discussed Features**

**Objective 3: Identify and Rank Lead Users**

- **Latent Features**
- **Lead Users**

Tuarob, Tucker  http://www.engr.psu.edu/datalab/
Extracting Product Features

Algorithm 1: The feature extraction algorithm from a collection of documents

Input: \( D \): Set of free-text documents to extract product features.
Output: \( E \): Set of extractions. Each \( e \in E \) is a tuple of \( \langle \text{feature, opinion, frequency} \rangle \), for example \( e = \langle \text{'onscreen keyboard'}, \text{'fantastic'}, 5 \rangle \)

1. preprocessing;
2. for \( d \in D \) do
   3. Clean \( d \);
   4. POS tag \( d \);
   5. Extract multi-word features;
3. end
4. initialization;
5. \( E = \emptyset \);
6. \( T = \emptyset \);
7. \( F = \) Seed Features;
8. while \( E \) can still grow do
   9. Learn templates from seed features;
   10. Add new template to \( T \);
   11. foreach \( d \in D \) do
      12. foreach Sentence \( s \in d \) do
         13. \( e \leftarrow \) Extract potential feature-opinion pair using \( T \);
         14. Add \( e \) to \( E \);
      15. end
   16. end
   17. Update \( F \);
18. \( E \leftarrow \) Clustering and normalizing features;
Methodology: Extract Product Features (Example)

**Product Specification Document**

The rechargeable *battery* built with *lithium-ion polymer* with a charge capacity of *1440mAh*[^8] is built in and cannot be replaced by the user; it is rated at \( \leq 225 \text{ hours of standby time} \) and \( \leq 8 \text{ hours of talk time} \).

**Extract**

battery
lithium-ion polymer
1440mAh
225Hr of standby time
8Hr talk time

**Social Media Message**

U know with all the *glass* in the iPhone 4 they really should think about integrating a *solar panel* to recharge the *battery*.

**Extract**

glass
solar panel
battery
Methodology: Identify Latent Features

“A latent feature: a feature that has not yet been implemented in any products within its domain”

• Naïvely, \(\{\text{Latent Features}\} = \{\text{All Features}\} - \{\text{Ground Truth Features}\}\).

• Retrieve *meaningful* latent features. So \(\{\text{Latent Features}\} = \{\text{Social Discussed Features}\} - \{\text{Ground Truth Features}\}\).
Methodology: Identify Lead Users (Product Specific vs. Global)

- Lead Users
  - Product Specific
    - Products familiar with
  - General
    - Products used
    - Products familiar with
    - All products within a domain
Methodology: Identify Latent Features

- \( \mathcal{S} = \{s_1, s_2, \ldots, s_n\} \): a product domain
- \( F^*(\mathcal{S}) \) set of \textbf{global} latent features
- \( F^*(s) \): the set of \textbf{product specific} latent features of product \( s_i \)

\[
FF(f, F^*) = 0.5 + 0.5 \cdot \frac{|\text{Frequency}(f)|}{\sum_{f' \in F^*} |\text{Frequency}(f')|}
\]

\[
IPF(f, \mathcal{S}) = \log \frac{|\mathcal{S}|}{|\{s \in \mathcal{S} : f \in s\}|}
\]

\[
FF - IPF(f, F^*, \mathcal{S}) = FF(f, F^*) \cdot IPF(f, \mathcal{S})
\]
Methodology: Identify Product Specific Lead Users

“Product specific lead users emit innovative ideas about the products that they use or are familiar with”

Algorithm 2: Algorithm for identifying and ranking product specific lead users of a particular product s

**Input:** $s \in S$: The product. U: The set of all users. $F(G_s)$: Ground-truth features. $F(M_s)$: User discussed features. $F^*(s)$: Latent features.

**Output:** Ranked list of users with respect to $P(u|s)$

1. initialization;
2. $I = \emptyset$;
3. **foreach** user $u \in U$ **do**
4. $M_u \leftarrow$ The messages posted by $u$;
5. Compute $F(M_u)$ using Algorithm 1;
6. iScore $\leftarrow$ Compute $P(u,s)$;
7. Add $\langle u, \text{iScore} \rangle$ to I;
8. **end**
9. $I \leftarrow$ Rank users in I by iScores;
10. **return** I

$$P(u|s) = \sum_{f \in F(M_u)} P(u|f,s) \cdot P(f|s)$$

$$P(u|f,s) = \begin{cases} 1 & ; f \in F^*(S) \\ 0 ; Otherwise \end{cases}$$

$$P(f|s) = \frac{1}{|F(G_s) \cup F(M_s)|}$$
Methodology: Identify Global Users

“Global lead users have critical, innovative ideas about all the products within the product domain.”

\[ P(u) = \sum_{s \in S} P(u|s) \cdot P(s) \]

\[ P(s) = \frac{|Positive(s)|}{\sum_{s' \in S} |Positive(s')|} \]

Positive(s) is the set of positive messages associated with the product s.
Case Study: Smart phones and Twitter Users

• 27 smart phone products
  – Smartphone Specification Product Specification Manuals

  – Product Related Twitter Data:
    • 2.1 billions tweets in the United States during the period of 31 months, from March 2011 to September 2013.
    • Preprocess by cleaning and mapping sentiment level (positive, neutral, negative).
Monthly distribution of Twitter discussion of each smart phone model across the 31 month period of data collection.
## Results: Extract Product Features

<table>
<thead>
<tr>
<th>Model</th>
<th># Base Features</th>
<th># User Features</th>
<th># Latent Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlackBerry Bold 9900</td>
<td>1126</td>
<td>126</td>
<td>101</td>
</tr>
<tr>
<td>Dell Venue Pro</td>
<td>497</td>
<td>50</td>
<td>36</td>
</tr>
<tr>
<td>HP Veer</td>
<td>1206</td>
<td>76</td>
<td>56</td>
</tr>
<tr>
<td>HTC ThunderBolt</td>
<td>627</td>
<td>335</td>
<td>281</td>
</tr>
<tr>
<td>iPhone 3G</td>
<td>1330</td>
<td>532</td>
<td>420</td>
</tr>
<tr>
<td>iPhone 3GS</td>
<td>891</td>
<td>775</td>
<td>652</td>
</tr>
<tr>
<td>iPhone 4</td>
<td>995</td>
<td>6057</td>
<td>5720</td>
</tr>
<tr>
<td>iPhone 4S</td>
<td>963</td>
<td>5922</td>
<td>5582</td>
</tr>
<tr>
<td>iPhone 5</td>
<td>1020</td>
<td>13493</td>
<td>13050</td>
</tr>
<tr>
<td>iPhone 5C</td>
<td>895</td>
<td>833</td>
<td>717</td>
</tr>
<tr>
<td>iPhone 5S</td>
<td>973</td>
<td>1962</td>
<td>1740</td>
</tr>
<tr>
<td>Kyocera Echo</td>
<td>895</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>LG Cosmos Touch</td>
<td>769</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>LG Enlighten</td>
<td>1084</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Motorola Droid RAZR</td>
<td>582</td>
<td>593</td>
<td>496</td>
</tr>
<tr>
<td>Motorola DROID X2</td>
<td>504</td>
<td>162</td>
<td>138</td>
</tr>
<tr>
<td>Nokia E7</td>
<td>749</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Nokia N9</td>
<td>745</td>
<td>83</td>
<td>62</td>
</tr>
<tr>
<td>Samsung Dart</td>
<td>1178</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Samsung Exhibit 4G</td>
<td>1331</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Samsung Galaxy Nexus</td>
<td>456</td>
<td>1147</td>
<td>1017</td>
</tr>
<tr>
<td>Samsung Galaxy S 4G</td>
<td>1322</td>
<td>62</td>
<td>37</td>
</tr>
<tr>
<td>Samsung Galaxy S II</td>
<td>1319</td>
<td>801</td>
<td>662</td>
</tr>
<tr>
<td>Samsung Galaxy Tab</td>
<td>771</td>
<td>884</td>
<td>762</td>
</tr>
<tr>
<td>Samsung Infuse 4G</td>
<td>1121</td>
<td>85</td>
<td>60</td>
</tr>
<tr>
<td>Sony Ericsson Xperia Play</td>
<td>726</td>
<td>132</td>
<td>102</td>
</tr>
<tr>
<td>T-Mobile G2x</td>
<td>945</td>
<td>39</td>
<td>23</td>
</tr>
</tbody>
</table>

Tuarob, Tucker  
http://www.engr.psu.edu/datalab/
Results: Identify Latent Features

- A set of 25,816 global latent features are extracted from the smart phone related social media data.

- Latent features with FF-IPF scores lower than 1.1 are treated as noise and eliminated.
## Results: Identify Latent Features

<table>
<thead>
<tr>
<th>Latent Feature</th>
<th>FF-IPF</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterproof</td>
<td>1.3087</td>
<td>I hope Apple incorporates some of that new <strong>waterproof</strong> technology in the iPhone 5. iPhone 5 better be <strong>waterproof</strong>, shockproof, scratchproof, thisproof, thatproof, and all the rest of the proofs for $800.</td>
</tr>
<tr>
<td>Solar Panel</td>
<td>1.3061</td>
<td>... and what else would make the iPhone 5 even better, built in <strong>solar power charging</strong>! U know with all the glass in the iPhone 4 they really should think about integrating a <strong>solar panel</strong> to recharge the battery.</td>
</tr>
<tr>
<td>Hybrid</td>
<td>1.3027</td>
<td>I wish there was an #android phone out there that was a <strong>hybrid of the best features on the droid razr maxx and the galaxy nexus</strong>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I need a <strong>hybrid-iPhone 4s</strong> so the battery can hold on all day when I'm at #vmworld. Steve, are you listening? :)</td>
</tr>
<tr>
<td>Tooth Pick</td>
<td>1.3023</td>
<td>I hope iPhone 5 borrows from Swiss Army and finally adds a <strong>removable tooth pick</strong>.</td>
</tr>
<tr>
<td>iHome</td>
<td>1.3021</td>
<td>My life would be 827492916 times better if my <strong>iHome took my iPhone 5</strong> First world problem: mad because my <strong>iPhone 5 is not compatible with this iHome</strong> dock in the hotel room.</td>
</tr>
</tbody>
</table>

Top 5 latent features across the chosen smart phone models, FF-IPF scores, and example tweets that related to the latent features
Results: Identify Lead Users

Average product specific iScore (i.e. $P(u|s)$) and global iScore (i.e. $P(u)$) of top 100 lead users across all the selected smartphone models, along with average numbers of tweets both related to each smartphone model (Avg Num Msg @TOP100) and average number of tweets related to smartphone in general (Avg Num All Msg @TOP100)
Results: Identify Lead Users

<table>
<thead>
<tr>
<th>Model</th>
<th>Product iScore</th>
<th>Sample Twitter Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samsung Galaxy Nexus</td>
<td>0.0496</td>
<td>I wish there was an #android phone out there that was a <em>hybrid</em> of the best features on the droid razr maxx and the galaxy nexus.</td>
</tr>
<tr>
<td>HTC ThunderBolt</td>
<td>0.0308</td>
<td>HTC Thunderbolt <em>fail</em>: Connect phone to PC to access drivers on included <em>SD card</em> ... but need drivers installed to access SD card from PC</td>
</tr>
<tr>
<td>iPhone 5</td>
<td>0.0174</td>
<td>but unless <em>Siri</em> can do more that just talk ...I'm not sold! #iPhone5</td>
</tr>
<tr>
<td>Sony Ericsson Xperia Play</td>
<td>0.0085</td>
<td>Hmm.. Playing games supporting Xperia Play controls. Wish I could use <em>PS3 controller</em> .. Makes me want an LTE Xperia Play with Tegra3..</td>
</tr>
<tr>
<td>Kyocera Echo</td>
<td>0.0077</td>
<td>Kyocera Echo needs to develop its own <em>apps</em></td>
</tr>
</tbody>
</table>

Sample tweets from the top lead user of each sample five smart phone models. These tweets suggest product innovative improvement for each corresponding product.
### Results: Identify Lead Users

<table>
<thead>
<tr>
<th>Global iScore</th>
<th>Sample Twitter Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0127</td>
<td>I wish there were a tweak for the iPhone 4S that would <em>indicate &quot;4G&quot; instead of just 3G</em> when I'm connected with a HSDPA+ connection.</td>
</tr>
<tr>
<td>0.0126</td>
<td>If you trust my instinct, the iPhone 5S will come in <em>multiple colors and two display sizes</em></td>
</tr>
<tr>
<td>0.0113</td>
<td>Very exciting Siri on the iPhone 4S <em>activates when you &quot;raise it to your ear&quot;</em> that'd b awesome.</td>
</tr>
<tr>
<td>0.0107</td>
<td>I wish i could use my iPhone as a <em>universal remote</em> control.</td>
</tr>
<tr>
<td>0.0105</td>
<td>Since iPhone already does fingerprint, Sumsung should <em>scan eyes</em> .</td>
</tr>
</tbody>
</table>

Sample tweets from the top 5 global lead users of the smartphone domain. These tweets suggest product innovation.
Conclusion and Path Forward

Social Network Model

- Unknown Needs $\mathbf{Y}$
- User $X_j$

Research Hypothesis

- Latent features expressed through social media
- Users' Posts $m_j$
- Word Vector $v_j$
- Topic $t_k$
Acknowledgement & References

Contributors:
• D.A.T.A. Lab: Suppawong Tuarob, Conrad Tucker

References