Social Media Analytics

Raghu Krishnapuram and Jitendra Ajmera
IBM Research - India
Convergence of Social and Analytic Technologies Transform the Way the World Operates

**Socially Synergistic (Enterprise) Solutions**

- New opportunities, better relationships with citizenry, customers and partners, enhanced talent pool, increased resiliency and efficiency

**Analytics**
- Data aggregation
- Smart filtering
- Meaning extraction
- Consumable analytics
- Process orchestration
- Stream processing

**Social**
- Customer Sentiment
- Unmet Needs
- Talent Discovery
- Reasoning and Decision Support
- Crowdsensing, Crowdsourcing
- Teaming, Incentives, Motivation

**Data**
- Physical: Sensors & Streams
- Enterprise
- Social: Data from and about People

© 2011 IBM Corporation
Examples of Socially Synergistic Solutions

Customer Care and Insight

Workforce Optimization

Physical Meets Digital

Financial Operations

Smarter Commerce

Advanced Case Management

© 2011 IBM Corporation
What is social media analytics? Analytics that helps in **forming, understanding, and then leveraging** communities for societal activities and business offerings.

- **Form**: Analytics technologies to identify/build communities around a given objective.
- **Understand**: Analytics & modeling provide deep understanding of structure & dynamics of community interactions
- **Leverage**: Manage communities to achieve specific goals or business objectives.
## Examples of Technical Successes

### Natural Language Processing
- Noisy text analytics
- Brand and reputation management
- Sentiment analysis

### Data Mining and Knowledge Discovery
- Profiling
- Personalization
- Customization

### Community analytics
- Influencers
- Topic detection
- Churn prediction
- Event detection and tracking

### Big Data
- Massive scale analytics platforms
- Stream computing

© 2011 IBM Corporation
Examples of Successful Applications

- Collaborative (crowd-sourced) knowledge creation – wikipedia
- Public sector/government - Grievance redress, opinion gathering, political campaign (propaganda)
- Telco – promotions, churn prediction, personalization
- Consumer products/Retail - customer service, customer engagement, sales growth, labor pooling,
- Travel - Cost-effective, direct customer feedback; linked community offerings; new customer acquisition, etc.
- Contact center – Agent/community help. FAQ, generate agent engagement rules, relevant KPIs.
Examples of Projects at
IBM Research - India
Top concepts and authors, thread categorization, forum monitoring

Threads filtered & organized into categories

Analytics look for other topics and concepts

Analytics track sentiment and health of the categories

Threads from many sources are integrated into one tool
## Automated Concept-Sentiment Mining

### Representative words

<table>
<thead>
<tr>
<th>Facet Topic (representative words)</th>
<th>More Words from the topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>question, answers, problems</td>
<td>question, questions, questioning, answers, problems, understanding, issues</td>
</tr>
<tr>
<td>agent, accent, conversation</td>
<td>accent, agent, conversation, english, foreign, voice, spanish, record, problem</td>
</tr>
<tr>
<td>information, assistance, knowledge</td>
<td>information, assistance, specialist, knowledge, resolve, decisions, ability</td>
</tr>
<tr>
<td>dealership, warranty, contract</td>
<td>dealership, warranty, coupon, dealer, contract, recall, repair, claim, care</td>
</tr>
<tr>
<td>vehicle, car, gm</td>
<td>vehicle, gm, car, toyota, chevrolet, toyota, saturn, pontiac, chevy</td>
</tr>
<tr>
<td>agent, service, rep</td>
<td>assistant, manager, supervisor, agent, rep, staff, customer, service</td>
</tr>
</tbody>
</table>

### Sentiment Topic

<table>
<thead>
<tr>
<th>Facet Topic</th>
<th>Words from the topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>good, nice, helpful, polite, courteous, genuinely, fantastic, knowledgeable, professional</td>
</tr>
<tr>
<td>Negative</td>
<td>bad, poor, unhappy, fault, faulty, ignored, insensitive, untrained, disappointing</td>
</tr>
</tbody>
</table>

### Sentiment Analysis

<table>
<thead>
<tr>
<th>Facet Topic</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
<th>Correlated CSAI score</th>
</tr>
</thead>
<tbody>
<tr>
<td>question, answers, problems</td>
<td>22.23</td>
<td>47.18</td>
<td>30.59</td>
<td>3</td>
</tr>
<tr>
<td>agent, accent, conversation</td>
<td>38.16</td>
<td>51.82</td>
<td>10.02</td>
<td>2</td>
</tr>
<tr>
<td>information, assistance, knowledge</td>
<td>29.48</td>
<td>14.52</td>
<td>56</td>
<td>3</td>
</tr>
<tr>
<td>dealership, warranty, contract</td>
<td>19.22</td>
<td>44.18</td>
<td>36.6</td>
<td>1</td>
</tr>
<tr>
<td>vehicle, car, gm</td>
<td>18.78</td>
<td>37.26</td>
<td>43.96</td>
<td>3</td>
</tr>
<tr>
<td>agent, service, rep</td>
<td>40.76</td>
<td>55.87</td>
<td>3.37</td>
<td>2</td>
</tr>
</tbody>
</table>

The agent was very helpful....
Manager spoke politely and helped me...
Rep was very pleasant to speak to...
Staff was very professional and courteous...

On click
Influencer finding

- Some users on social media are more influential than others
  - They get heard by a large number of users (followers)
  - They are subject matter experts
  - They are super active, specially on a particular topic

- Issues
  - All the followers are not listening
  - Active users can also be spammers
  - People may not like to make formal connections

- Our Approach
  - Identify effective number of followers
  - Consider relevant (topical) activity (forwards, shares)
  - Analyze this implicit network using methods such as pagerank
  - Initial results show 85% accuracy

- How to leverage?
  - Prioritize for customer care
  - Influence by offers/service
  - Prioritize posts to respond
  - Encourage people to write meaningful content on a forum
  - Can be used to define a network of experts
  - Can be used to recommend to people “who to subscribe”
Thread Analytics

Social media communications are threaded:
- People respond to previous posts to express opinions, spread knowledge, etc.
- Each post itself is not complete
- The structure of the thread varies depending on the type of discussion
- This structure can be exploited to answer questions like:
  - Who is the original poster (company, individual user, spammer, ...)?
  - What's the intent of the original poster (marketing, query, complaint, praise, ...)?
  - What's the intent of each comment (agree, disagree, ask question, spam, flame, ...)?
  - Is the conversation "resolved" in some way or still unanswered?
  - Is the author of a comment an expert?
  - Is the discussion event oriented?

Observations
1. Sentiment
2. Question patterns
3. Author frequency/history
4. Domain keywords
Characterizing Micro-text

Problem: Traditional text analysis techniques cannot be applied because:
- Few words, missing context
- Missing syntax makes POS tagging etc less accurate,
- Rule-based approaches also fail.

Hypothesis: Content generation in micro-text is governed by:
- User preferences (theme, user specific)
- Events (temporal, social phenomenon)

Approach:
- A non-parametric model that captures ‘themes’ and ‘events’ and explains their influence on the content generation in micro-text
- A parallel Gibbs sampling based online inference algorithm for the non-parametric model

Experimental Results
- Accuracy of 76.3% on the user-theme identification task
- For the user authorship prediction task, an accuracy of over 70% as compared to 50% accuracy obtained by baseline LDA model
Attention Prediction on Brand Pages

**Problem:** Brand pages are increasingly being set-up by enterprises to have social media presence. Given a new ‘post’ on a brand-page, predict how much ‘attention’ (comments, likes, shares etc.) it will attract.

**Application:** Social media monitoring for critical topics, prioritization of posts on forums for manual viewing

**Hypothesis:** The factors responsible for drawing attention include factors from all three dimensions: content, author and the network ad placing/pricing

**Approach:** Extract features from all the three dimensions and then classify in terms of discreet categories or a regression function that predicts the exact number of comments

**Experimental results:** (SVM classification and regression) For the classification task, obtained a precision (recall) of 77 (68%) as compared to a baseline performance of 63% (56%). For the regression task, obtained a predictive $R^2$ of .54 as compared to a baseline $R^2$ of .24. The baseline system consisted of a topic model based on users’ previous commenting history
Some Research Challenges

- Noisy data, spam, rumors
- Data fusion/correlation (conflicting data, missing data)
- Heterogeneous data sources
- Models for user generated content
- Models for social media behavior including incentives
- Pattern evolution in social networks
- User intent mining
- Prediction under uncertainty
- Scalable analytics
- Socially synergistic systems (including within enterprise)
- Spatio-temporal analytics on social data