Social Media Analytics

Raghu Krishnapuram and Jitendra Ajmera
IBM Research - India

Convergence of Social and Analytic Technologies
## Transform the Way the World Operates

IBM

Socially Synergistic (Enterprise) Solutions
New opportunities, better relationships with citizenry, customers and partners, enhanced talent pool, increased resiliency and efficiency
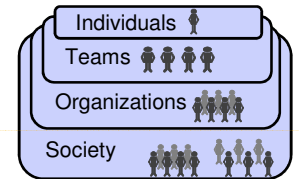
Analytics

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

Process Transformation

Social

Individuals
Teams
Organizations
Society
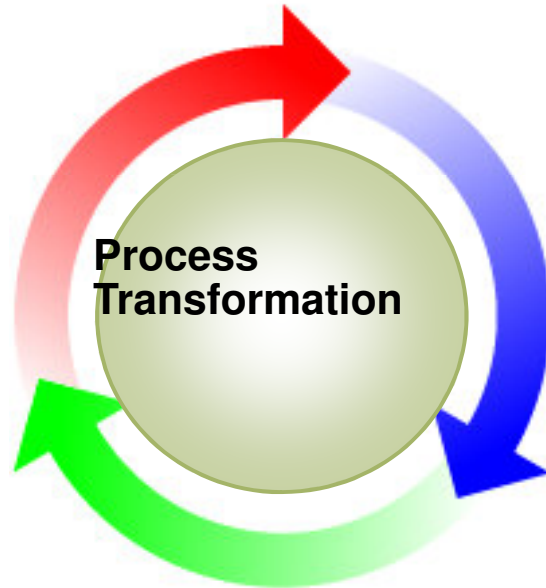
- Customer Sentiment
- Unmet Needs
- Talent Discovery
- Reasoning and Decision Support
- Crowdsensing, Crowdsourcing
- Teaming, Incentives, Motivation

Data

Physical
Sensors & Streams

Enterprise

Social
Data from and about People

Examples of Socially Synergistic Solutions

Customer Care and Insight

Workforce Optimization

Physical Meets Digital

Financial Operations

Smarter Commerce

Advanced Case Management

What is social media analytics? Analytics that helps in forming, understanding, and then leveraging commu-nities for societal activities and business offerings.

IBM

Form

Understand

Leverage
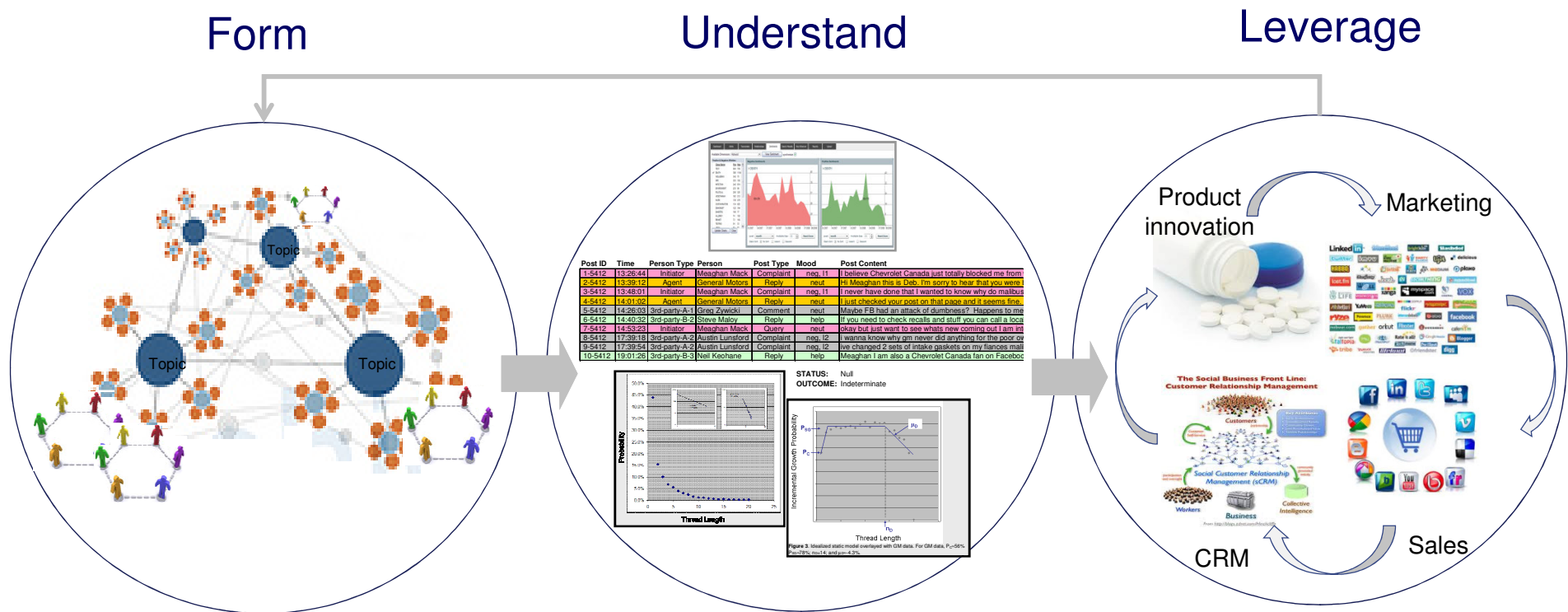
- Analytics technologies to identify/build communities around a given objective.

- Analytics & modeling provide deep understanding of structure & dynamics of community interactions

- 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

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

SCION | osts | Respond | Cust | Enter Keywords

Categories

Category ∧	# Posts	# Unresponded	Sentiment	Health
Phones and Devices	219 posts	97 unresponded	15% negative - 28% positive	60%
Services and Features	167 posts	69 unresponded	15% negative - 21% positive	59%
Mobile Applications	217 posts	102 unresponded	28% negative - 19% positive	47%
Service Plans	110 posts	51 unresponded	25% negative - 10% positive	41%
Billing / Accounts	191 posts	81 unresponded		
Network & Coverage	517 posts	253 unresponded		
Trending Technology (4G)	606 posts	144 unresponded	16% negative - 13% positive	62%
Technical Support	60 posts	20 unresponded	25% negative - 12% positive	50%
Urgent Help	71 posts	11 unresponded	32% negative - 10% positive	55%

Sources

Source ∧	# Posts	# Unresponded	Sentiment	Health
Facebook	721 posts	411 unresponded	25% negative - 13% positive	42%
Twitter	1198 posts	432 unresponded	13% negative - 2% positive	36%
Forums	2165 posts	828 unresponded	23% negative - 17% positive	52%

Concepts

account-info android-update battery cell-sites coverage htc-arrive error-msg network-coverage power-cycle retail-store service-center silver-status system-updates

Top Authors

carleebb	28
GamerBlogTv	19
K9NSP	17
lexy_229	14
andykaluzny	14
Morenicano	13
camarojones	12
robs8249	10

United States Mainland

Chicago 57% negative
Nashville 100% negative
San Diego 60% negative
Houston 100% negative

Callout boxes:
- 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

IBM

Representative words

Facet topic words

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

Sentiment Topic	Words from the topic
Positive	good, nice, helpful, polite, courteous, genuinely, fantastic knowledgeable, professional
Negative	bad, poor, unhappy, fault, faulty, ignored, insensitive, untrained, disappointing

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

On click

The agent was very helpful....

Manager spoke politely and helped me...

Rep was very pleasant to speak to...
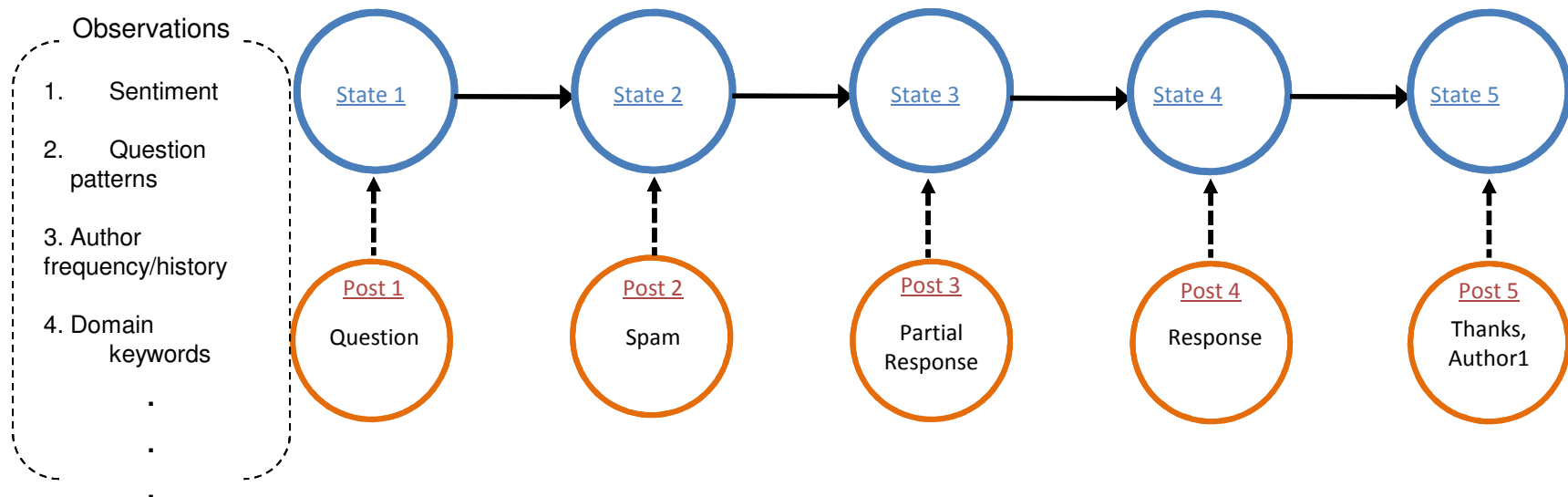
Staff was very professional and courteous...
....

- **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**
  - Indentify 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

.
.
.

State 1 → State 2 → State 3 → State 4 → State 5

Post 1 — Question
Post 2 — Spam
Post 3 — Partial Response
Post 4 — Response
Post 5 — Thanks, Author1

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