ADVANCED CONTENT ANALYSIS TECHNIQUES

Content Analysis PDW
Academy of Management – 2014
Philadelphia, PA

Tim Hannigan (University of Oxford)
Robert Vesco (Yale University)
GOALS

• Share examples of cutting edge text analysis techniques by running through a sample case

• Discuss challenges and future promise

• Provide tips and resources for implementation
ADVANCED TECHNIQUES

- Topic Models
- Name Entity Recognition (NER)
- High Accuracy Sentiment Analysis
- Concept Networks
TOPIC MODELING OVERVIEW (BLEI, 2011)
**TOPIC ANALYSIS**

- **What is it?**
  - Generates probabilistic models of topic/categories within text
  - Most commonly Latent Dirichlet Allocation (LDA)

- **When is it useful? Examples?**
  - Looking at the change in ideas over time (Kaplan & Vakili, 2014; Mohr et al, 2013; DiMaggio et al, 2013)
  - Identifying relationships between entities based on shared meanings
Example: AOM PDWs Abstracts

- We web-scraped 2014 AOM PDWs
- Texts of Abstracts
- Sponsors (BPS, OMT, COG, TIM ...)
- Extracted “topics” using LDA Topic Modeling
- Analysed relationship between topics and sponsors
AOM PDWs

(AAA) All Academy Activities
(AAT) All Academy Theme
(AAC) Affiliate Activities & Committees
(AAM) Asia Academy of Management
(BPS) Business Policy & Strategy
(CAR) Careers
(CAU) Caucuses
(CM) Conflict Management
(CMS) Critical Management Studies
(D&IT) Diversity & Inclusion Theme Committee
(ENT) Entrepreneurship
(EXH) Exhibits
(GDO) Gender & Diversity in Organizations
(HCM) Health Care Management
(HR) Human Resources
(HAM) Iberoamerican Academy of Management
(ICW) In Conjunction With Activities
(INDAM) Indian Academy of Management
(IM) International Management
(ITC) International Theme Committee
(MC) Management Consulting
(MED) Management Education & Development
(MH) Management History
(MSR) Management Spirituality & Religion
(MOC) Managerial & Organizational Cognition
(NDSC) New Doctoral Student Consortium
(OM) Operations Management
(OMT) Organization & Management Theory
(ODC) Organization Development & Change
(OB) Organizational Behavior
(OCIS) Organizational Communication & Information Systems
(ONE) Organizations & the Natural Environment
(PTC) Practice Theme Committee
(PNP) Public & Nonprofit
(RM) Research Methods
(SIM) Social Issues in Management
(SAP) Strategizing Activities and Practices
(TLC) Teaching & Learning Conference
(TTC) Teaching Theme Committee
(TIM) Technology & Innovation Management
### Need to Interpret Topics

<table>
<thead>
<tr>
<th>Topic 11</th>
<th>Topic 15</th>
<th>Topic 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;social&quot;</td>
<td>&quot;entrepreneur&quot;</td>
<td>&quot;gender&quot;</td>
</tr>
<tr>
<td>&quot;ethic&quot;</td>
<td>&quot;ventur&quot;</td>
<td>&quot;women&quot;</td>
</tr>
<tr>
<td>&quot;csr&quot;</td>
<td>&quot;entrepreneursh&quot;</td>
<td>&quot;career&quot;</td>
</tr>
<tr>
<td>&quot;stakehold&quot;</td>
<td>&quot;entrepreneur&quot;</td>
<td>&quot;diers&quot;</td>
</tr>
<tr>
<td>&quot;respons&quot;</td>
<td>&quot;new&quot;</td>
<td>&quot;femal&quot;</td>
</tr>
<tr>
<td>&quot;corpor&quot;</td>
<td>&quot;busi&quot;</td>
<td>&quot;negoti&quot;</td>
</tr>
<tr>
<td>&quot;moral&quot;</td>
<td>&quot;opportun&quot;</td>
<td>&quot;studi&quot;</td>
</tr>
<tr>
<td>&quot;studi&quot;</td>
<td>&quot;startup&quot;</td>
<td>&quot;work&quot;</td>
</tr>
<tr>
<td>&quot;manag&quot;</td>
<td>&quot;studi&quot;</td>
<td>&quot;find&quot;</td>
</tr>
<tr>
<td>&quot;busi&quot;</td>
<td>&quot;find&quot;</td>
<td>&quot;effect&quot;</td>
</tr>
<tr>
<td>&quot;paper&quot;</td>
<td>&quot;model&quot;</td>
<td>&quot;men&quot;</td>
</tr>
<tr>
<td>&quot;valu&quot;</td>
<td>&quot;effect&quot;</td>
<td>&quot;research&quot;</td>
</tr>
<tr>
<td>&quot;theori&quot;</td>
<td>&quot;research&quot;</td>
<td>&quot;differ&quot;</td>
</tr>
<tr>
<td>&quot;practic&quot;</td>
<td>&quot;theori&quot;</td>
<td>&quot;manag&quot;</td>
</tr>
<tr>
<td>&quot;find&quot;</td>
<td>&quot;firm&quot;</td>
<td>&quot;posit&quot;</td>
</tr>
<tr>
<td>&quot;engag&quot;</td>
<td>&quot;uncertainti&quot;</td>
<td>&quot;experi&quot;</td>
</tr>
<tr>
<td>&quot;uneth&quot;</td>
<td>&quot;use&quot;</td>
<td>&quot;use&quot;</td>
</tr>
<tr>
<td>&quot;use&quot;</td>
<td>&quot;paper&quot;</td>
<td>&quot;male&quot;</td>
</tr>
<tr>
<td>&quot;firm&quot;</td>
<td>&quot;capit&quot;</td>
<td>&quot;age&quot;</td>
</tr>
<tr>
<td>&quot;public&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## TOP TOPICS BY DIVISION

<table>
<thead>
<tr>
<th></th>
<th>BPS</th>
<th>OB</th>
<th>MOC</th>
<th>OMT</th>
<th>TIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EmployeeHR</td>
<td>2.00</td>
<td>28.00</td>
<td>11.00</td>
<td>6.00</td>
<td>0.00</td>
</tr>
<tr>
<td>FirmPerformance</td>
<td>46.00</td>
<td>1.00</td>
<td>0.00</td>
<td>5.00</td>
<td>41.00</td>
</tr>
<tr>
<td>Innovation</td>
<td>9.00</td>
<td>5.00</td>
<td>1.00</td>
<td>9.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Institutions</td>
<td>11.00</td>
<td>1.00</td>
<td>6.00</td>
<td>37.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Theory</td>
<td>50.00</td>
<td>62.00</td>
<td>20.00</td>
<td>48.00</td>
<td>28.00</td>
</tr>
</tbody>
</table>
NAMED ENTITY RECOGNITION

What is it?
Find names (people and places) within text

When is it useful? Examples?
You want to identify individuals or geographic locations mentioned
Allows you to do social network analysis and spatial econometrics
Currently using to identify government agencies in patent data
Advanced Sentiment Analysis

• **What is it?**
  • Is the text negative, positive or neutral; can bring Natural Language Processing, so negation of valence can be captured.. perhaps even sarcasm (yeah, right!)
  • Machine learning

• **Advantages**
  • Same as always – but many high-end services provide high accuracy (>90%)
  • Going beyond word dictionaries and simple word counts, moving beyond LIWC (Kaplan, 2011)

• **Downsides:**
  • Methodological blackbox
  • Complexity in process / cost

• **Services will do it for you**
  ▪ Have access to large corpora (ie. Google Books, NGrams), language is evolving (ie. Google, big data)
  ▪ Relatively cheap … to not so much
Can use collocations of concepts to form a network; then can use network tools such as centrality to measure salience

To be covered more in **PDW “Revisiting the Product Ontology”** (Sat Aug 2, 10:14-12:45, Pennsylvania Convention Center Room 203B) and **Symposium “The Power of Words in Big Data”** (Sun Aug 3 11:15-12:45, Pennsylvania Convention Center, Room 122 A)
NEXT CHALLENGES

• Promise
  • there’s no out of the box tool to do this for you
  • there are opportunities to collaborate with computer science researchers
  • opportunities to integrate this with Network Analysis Tools
  • Ethnographers and computers scientists can work together using topic modelling and complement one another (eg. Levy & Franklin, 2013)
  • emergent properties (large amounts of data)

• Challenges:
  • tools can be a black box; may be sensitive to certain assumptions
  • despite their scale and speed, there still remains a lot of researcher degrees of freedom
RESOURCES


• Topic Modeling in R using LDA

• Topic Modeling Tutorial in R and Python

• http://java.dzone.com/articles/topic-modeling-python-and-r

• Python programming for the Humanities - http://fbkarsdorp.github.io/python-course/

• Text analysis with topic models for the Humanities and social sciences

• https://de.dariah.eu/tatom/index.html

• https://github.com/rlvesco7/aom2014-content-analysis-pdw