Forecasting Innovation Pathways: Using Data on ‘Big Data’

Alan Porter
Georgia Tech
alan.porter@isye.gatech.edu
And
Search Technology, Inc.

Thanks to Ying Huang, Jannik Schuehle, Jan Youtie, Jianhua Liu, Ying Quo & Doug Robinson for key ideas and research. We acknowledge support from the US National Science Foundation (Award #1527370 – “Forecasting Innovation Pathways of Big Data & Analytics”). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Agenda

1. My slanted view on Management of Technology (MOT) “intel”
2. Tech Mining: A classic case – advanced ceramics
3. Forecasting Innovation Pathways (FIP): The case of data on “Big Data” (BD)
4. Message: Think empirically – use those ‘big’ data (+ more)
PICMET Mission

- Portland International Conference on Management of Engineering & Technology
- Advance analyses of changing technologies to
- Inform Management of Technology (MOT)

Managing (other than ‘technology’):
On the Shop Floor

1. The font of knowledge: The machine operator's intuition (experience)
2. Role of data: Deming’s Quality Control? [nonsense; see #1]
3. Action: Japan – empirical data provides enhanced knowledge to manage production processes far better
4. Result: Six Sigma manufacturing world (contrast to hospital world)
Managing (other than ‘technology’):
Athletics
1. The font of knowledge: The old veteran coach/manager ~ intuition (experience)
2. Role of data: [nonsense; see #1]
3. Action: **Moneyball** – empirical data provide enhanced knowledge to get superior major league baseball players on a tight budget [Oakland A’s quite successful]
4. One result -- NBA-2015/16 – Golden State Warriors
   – Stats inform personnel decisions
   – Stats inform player combinations & matchups [~Big Data = “Advanced Scout reveals hidden patterns from play-by-play data]
   – Andre Iguodala – first start after 82 regular season + 3 playoff series – in the 2015 finals
   – **NBA champs** (2015)
   [but, hey, you can’t win ‘em all!]

Managing: Science, Technology & Innovation (ST&I)
1. The font of knowledge in the US = personal judgment
   – scientists (peer review), or
   – engineers ~ collective experience, or
   – MOT managers ~ tacit knowledge
2. Role of data: [nonsense; see #1]
   – Scientometrics to inform R&D funding decisions? (in the US, request ear plugs)
3. Action: “Tech Mining” for **empirical intelligence** to complement expertise
4. Result → Better decisions!? 
Tech Mining

• To generate effective intelligence
• From ST&I information resources
• on topics of concern

Questions to Answer from field-structured data:
e.g., R&D publication or patent abstract records
gathered via global database search

Who?  Where?
What?  When?
How? & Why? – Need human analyst to interpret the data
TechMining success story: Ceramics in Engines by Bob Watts (US Army)  
**PICMET Best Student Paper**

- Overcoming Management Resistance
- Jumping Domains
- “Discovering” new technology

Informing a tough decision

- US Army Tank-Automotive Research, Development & Engineering Center
- Task in 1996: Reassess a “loser technology” – could thin-film ceramics be used in tank engines?
- TechMining: R&D Profile -- Amount of activity up a little -- but clues of significant maturation (next slide)
The data speak:

Ceramic Engine Publications (85-96)
Technology Maturity & Keyword Diffusion

The rest of the story

• Experts support the empirical findings
• Management buys in – search out potential in “coating engine parts”
• Who to go to? search finds ~95% of the research is NOT in their mechanical engineering domain -- so who would lead in ceramic coatings R&D?
• Identify R&D leaders – in semiconductor ceramics!
• $million projects funded with Sandia National Lab and a company to adapt “vapor deposition” to turbine blades
• Production plant coats used (Gulf War) Abrams tank turbine blades back to spec begins successful operation (2004)
## Forecasting Innovation Pathways (FIP)

### 10 Steps (non-linear!) to Forecast Newly Emerging Science/Technology (NEST) Innovation Pathways

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Understand the NEST and its TDS (Technology Delivery System)</td>
</tr>
<tr>
<td>Step A:</td>
<td>Characterize the technology’s nature</td>
</tr>
<tr>
<td>Step B:</td>
<td>Model the TDS</td>
</tr>
<tr>
<td>2.</td>
<td>Tech Mine</td>
</tr>
<tr>
<td>Step C:</td>
<td>Profile R&amp;D</td>
</tr>
<tr>
<td>Step D:</td>
<td>Profile innovation actors &amp; activities</td>
</tr>
<tr>
<td>Step E:</td>
<td>Determine potential applications</td>
</tr>
<tr>
<td>Step J:</td>
<td>Engage experts</td>
</tr>
<tr>
<td>3.</td>
<td>Forecast likely innovation paths</td>
</tr>
<tr>
<td>Step F:</td>
<td>Lay out alternative innovation pathways</td>
</tr>
<tr>
<td>Step G:</td>
<td>Explore innovation components</td>
</tr>
<tr>
<td>Step H:</td>
<td>Perform Technology Assessment</td>
</tr>
<tr>
<td>4.</td>
<td>Synthesize &amp; report</td>
</tr>
<tr>
<td>Step I:</td>
<td>Synthesize and Report</td>
</tr>
</tbody>
</table>

---

### Tech Mining for “Big Data” (BD)

**Why study BD? To inform a potent Social Innovation**

- US Government Accountability Office (GAO) doing a Technology Assessment of 21<sup>st</sup> Century Data
- To inform the US Congress and stakeholders re:
  - Issues and opportunities
  - Potential legislation – e.g., to protect privacy & security
  - Potential incentives to promote commercial development
- NSF support for our Georgia Tech group to improve our “Forecasting Innovation Pathways” methodology, using BD case
  - Interacted informally with GAO
Various Data on Big Data

Our initial focus is on R&D data:
- **INSPEC, EI Compendex** – treat the strong computer science core
- **Derwent Innovation Index** – patents
- **Research awards** – NSF, NSFC
- **Web of Science** – fundamental research; include conference proceedings [focus here]

+ Social Innovation facets:
  - **Commercial activity** (databases like ABI/Inform)
    - Databases like Lexis Nexis ~10,000
  - **Popular attention**
    - Google hits ~> 274 million (as of 7/31/2016)

Locating & Retrieving Data

<table>
<thead>
<tr>
<th>No</th>
<th>Search Strategy</th>
<th>Search Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Core Lexical Query</td>
<td>TS= ('Big Data' or Bigdata or &quot;Map Reduce&quot; or MapReduce or Hadoop or Hbase or Nosql or Newsql)</td>
</tr>
<tr>
<td>2</td>
<td>Expanded Lexical Query</td>
<td>TS=(Big Near/1 Data or Huge Near/1 Data) or &quot;Massive Data&quot; or &quot;Data Lake&quot; or &quot;Massive Information&quot; or &quot;Huge Information&quot; or &quot;Big Information&quot; or &quot;Large-scale Data&quot; or Petabyte or Exabyte or Zettabyte or &quot;Semi-Structured Data&quot; or &quot;Semistructured Data&quot; or &quot;Unstructured Data&quot;) AND TS=&quot;(Cloud Comput&quot; or &quot;Data Min&quot; or &quot;Analytic&quot; or &quot;Privacy&quot; or &quot;Data Manag&quot; or &quot;Social Media&quot; or &quot;Machine Learning&quot; or &quot;Social Network&quot; or &quot;Security&quot; or &quot;Twitter&quot; or &quot;Predict&quot; or &quot;Stream&quot; or &quot;Architect&quot; or &quot;Distributed Comput&quot; or &quot;Business Intelligence&quot; or &quot;GPU&quot; or &quot;Innovat&quot; or &quot;GIS&quot; or &quot;Real-Time&quot; or &quot;Sensor Network&quot; or &quot;Smart Grid&quot; or &quot;Complex Network&quot; or &quot;Genomics&quot; or &quot;Parallel Comput&quot; or &quot;Support Vector Machine&quot; or &quot;SVM&quot; or &quot;Distributed&quot; or &quot;Scalab&quot; or &quot;Time Serie&quot; or &quot;Data Science&quot; or &quot;Informatics&quot; or &quot;OLAP&quot;)</td>
</tr>
<tr>
<td>3</td>
<td>Specialized Journals</td>
<td>The papers published in these specialized journals are not indexed by WOS</td>
</tr>
<tr>
<td>4</td>
<td>Cited References</td>
<td>The publications, which were cited more than 20 times did not fulfill the criteria for inclusion (see paragraph &quot;Cited Reference Analysis&quot;)</td>
</tr>
</tbody>
</table>
Big Data Research Publication Profile

- Web of Science
- 2008 thru 2015 (as of early 2016)
- 11684 records, of which 53% are conference papers, with 39% journal articles
- Leading sources are IEEE International Conference on Big Data and IEEE International Congress on Big Data
- Lots of computer science attention
- Notable health Big Data research activity

Research Trend: “Big Data” [Web of Science]{“when??}
“Who Is doing Big Data Research? Top 10 Organizations publishing

<table>
<thead>
<tr>
<th>Author Organization</th>
<th>Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese Acad Sci</td>
<td>293</td>
</tr>
<tr>
<td>Tsinghua Univ</td>
<td>151</td>
</tr>
<tr>
<td>IBM</td>
<td>101</td>
</tr>
<tr>
<td>Harvard Univ</td>
<td>95</td>
</tr>
<tr>
<td>MIT</td>
<td>93</td>
</tr>
<tr>
<td>Beijing Univ Posts &amp; Telecommun</td>
<td>90</td>
</tr>
<tr>
<td>Univ Calif Berkeley</td>
<td>87</td>
</tr>
<tr>
<td>Stanford Univ</td>
<td>86</td>
</tr>
<tr>
<td>Univ Illinois</td>
<td>86</td>
</tr>
<tr>
<td>Huazhong Univ Sci &amp; Technol</td>
<td>85</td>
</tr>
</tbody>
</table>

Of Top 30 -- **17 American, 13 Chinese**;
Of 11684 WoS papers, 3656 with US author; 3022 with Chinese author
(trailed by UK (672) & Germany (594))
Disciplinary Span of Big Data Research

Method from Rafols, Porter and Leydesdorff (2009)

Psychological Sci.
Agri Sci
Biomed Sci
Chemistry
Physics
Mech Eng
Environ Sci & Tech
Materials Sci
Infectious Diseases
Clinical Psychology
Social Studies
Clinical Med
Computer Sci
Business & MGT
Geosciences
Ecol Sci
Econ Polit & Geography
Health & Social Issues

Topical Big Data Factors (data thru 2014) {what??}
BD R&D Policy Analyses {what??}

- Amazing increase in research on BD 2011-2014
- BD research is dominated by two countries – implications?
- Multidisciplinary attention, centered on computer science, but involving most R&D areas
- **Social Innovation** is a priority: sharp increases in social science analyses, patenting, commercial interest, and popular coverage [4,300,000 hits in Google Scholar; 274,000,000 in Google]
- We are pursuing Social Innovation aspects:
  - technology delivery system (players & influences)
  - technology assessment

Exploring Big Data Social Innovation

- **Technology Assessment**
  - Likelihood X Magnitude impact screening
  - Mitigation options
  - Survey + open web opinion seeking
    [Jianhua Liu, Ying Guo, et al. PICMET paper]

- **GAO “21st Century Data” TA study ongoing**

Big Data Social Innovation Scenario Analyses

**Key on application variations**

- **Morphological strategy:**
  - List factors
  - Identify alternatives for each
  - Package – driven by applications (real + potential)

- **Construct ~ 15-20 diverse application scenarios**
  - Present via blog or such to elicit feedback
    (elaboration, correction, addition)
  - Of various stakeholders

- **Scan for**
  - Sensitivities – factors (states) affecting multiple development paths
  - Opportunities
  - Emergent issues
Morphological Structuring for Scenario Analyses

<table>
<thead>
<tr>
<th>Factor</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market segments</td>
<td>A. Health</td>
</tr>
<tr>
<td>Notable Apps</td>
<td>F. Google Flu</td>
</tr>
<tr>
<td>Architecture</td>
<td>Distributed data</td>
</tr>
<tr>
<td>Regions</td>
<td>US</td>
</tr>
<tr>
<td>Standards</td>
<td>Bodies</td>
</tr>
<tr>
<td>Regulations</td>
<td>International collaboration</td>
</tr>
<tr>
<td>BD Development guiding policies</td>
<td>Open source emphasis</td>
</tr>
<tr>
<td>Boosting policies</td>
<td>R&amp;D funding</td>
</tr>
<tr>
<td>Industry Structures</td>
<td>Start-ups</td>
</tr>
<tr>
<td>External forces</td>
<td>Economic health</td>
</tr>
<tr>
<td>Infrastructures</td>
<td>Knowledge reservoirs</td>
</tr>
<tr>
<td>Key Actors</td>
<td>Google</td>
</tr>
<tr>
<td>Societal concerns</td>
<td>Privacy</td>
</tr>
<tr>
<td>Privacy Issues</td>
<td>Awareness (of threats to us)</td>
</tr>
<tr>
<td>Security</td>
<td>Hacking – private</td>
</tr>
<tr>
<td>Technologies</td>
<td>Distributed storage</td>
</tr>
<tr>
<td>Key Tech Capabilities</td>
<td>Sensors development</td>
</tr>
<tr>
<td>Enabling IT</td>
<td>Memory</td>
</tr>
<tr>
<td>R&amp;D Thrusts</td>
<td>Data science</td>
</tr>
<tr>
<td>(Data resources)</td>
<td>R&amp;D funding</td>
</tr>
<tr>
<td>(comparables – to compare trends)</td>
<td>Electrification</td>
</tr>
</tbody>
</table>

Our Big Data Papers


**Selected FIP References**

- Porter, A.L., and Huang, Y. (2016), Forecasting future innovation pathways with big data analytics, *CIMS Innovation Management Report*, 8-13 (July/August), Poole College of Management, NC State University, Raleigh.
- Global Tech Mining Conference **Special Issues**: *Technology Analysis & Strategic Management; Technological Forecasting & Social Change; Scientometrics; International Journal of Technology Management.*

**Resources**

- The text mining software used: *VantagePoint*
  [www.theVantagePoint.com](http://www.theVantagePoint.com)
  [or Thomson Data Analyzer]

- Global Tech Mining Conference, in conjunction with Atlanta Conference on Science & Innovation Policy, September, 2017
One more data-based example:
**Indicators of technical “emergence”**

Emergence Indicators:
Topical Terms
Secondary Emergence Indicators: Frontier Players

Managing ST&I
Combine multiple knowledge sources
(Hal Linstone’s multiple perspectives):

** My charge: ST&I management is too intuitive
[exceptions]
⇒ Competitive Technical Intelligence (CTI)

** Direct experience + expert opinion + empirical intelligence
⇒ Better decisions!