Business Analytics: The Past, Present and Future of Operations Research

EWO 2013

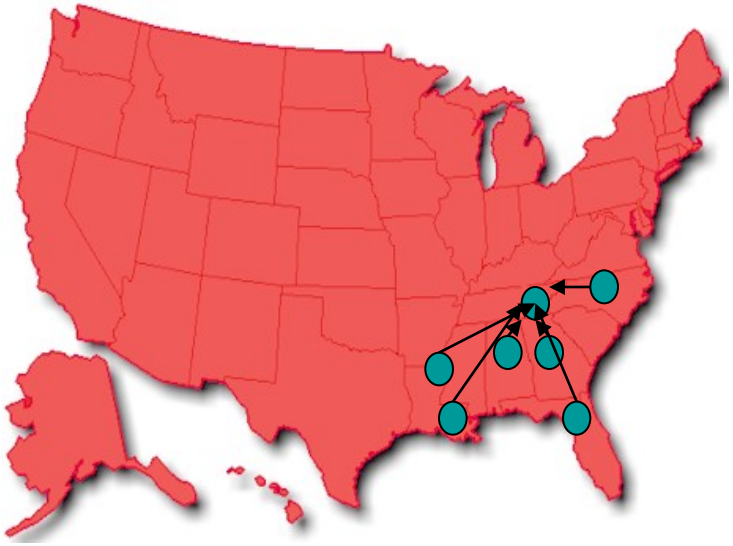
Michael Trick

Tepper School of Business, Carnegie Mellon University

THE PAST

Example 1: A New Model for Package Delivery

- 1973: Fred Smith, avid pilot, had an idea
- Provide “taxi service” for packages
- Fred knew the south-east so began there
- 11 cities were linked in the now-famous model: fly everything to Memphis, re-sort, and fly everything back



Result

- Enormous failure
  - 6 packages on day 1 (including one from Fred)
  - Not much better on days 2 and 3
  - Closed on day 4
- What to do?

Example 2: Yahoo! versus Google

2002. Yahoo! has weathered the crash of the internet bubble. Sees (correctly) that search-based advertising is a *big thing* (it is currently a US$15 billion/year business segment)

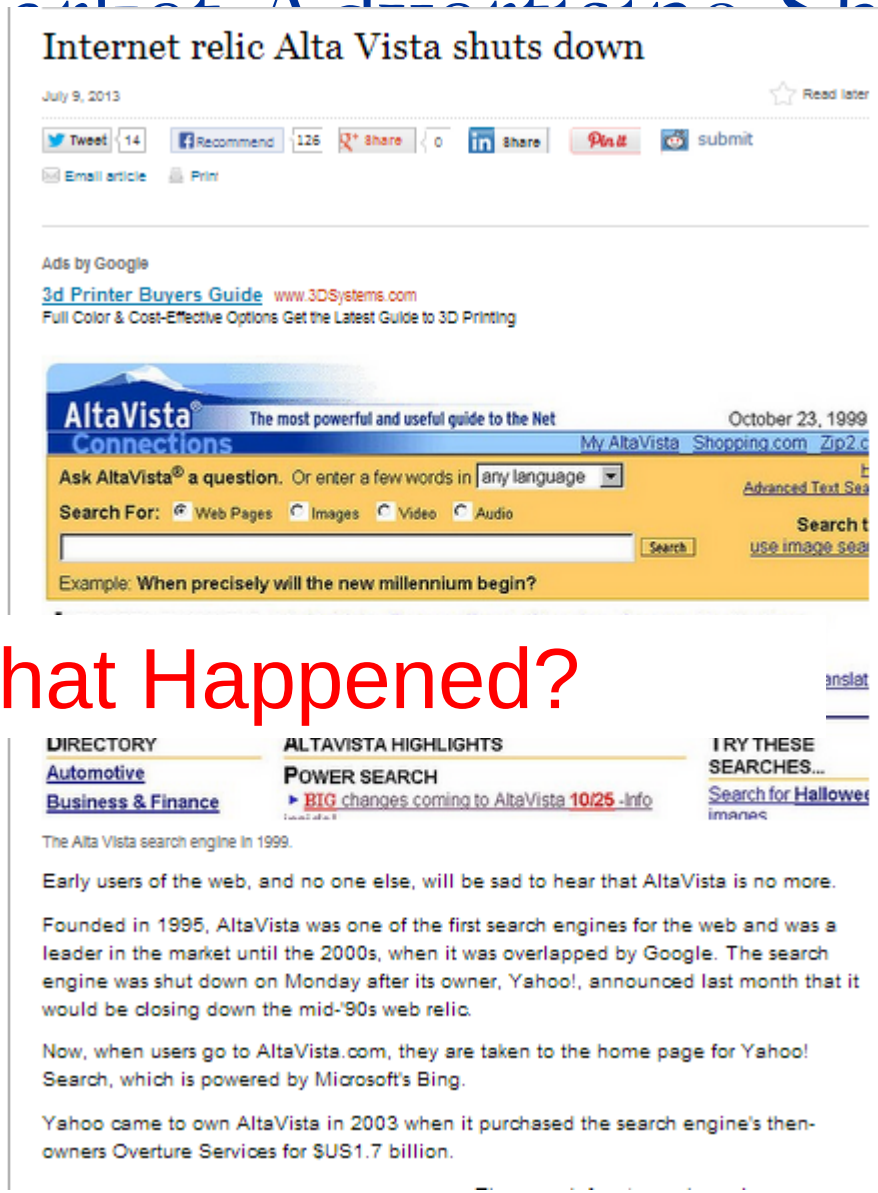
Proposes US$3billion to buy Google (then with yearly revenues of US$240 million) but is rebuffed

Purchases Inktomi and Overture to compete with Google

Search Market Advertising Share

- 2002: Ov ogle
- 2003: Ov
- 2005: Go erture
- 2007: Go market

What Happened?



Internet relic Alta Vista shuts down

July 9, 2013

Read later

Tweet 14 Recommend 126 Share 0 Share Pin it submit

Email article Print

Ads by Google

3d Printer Buyers Guide www.3DSystems.com

Full Color & Cost-Effective Options Get the Latest Guide to 3D Printing

AltaVista Connections The most powerful and useful guide to the Net October 23, 1999

My AltaVista Shopping.com Zip2.c

Ask AltaVista® a question. Or enter a few words in any language

Advanced Text Sea

Search For: Web Pages Images Video Audio

Search

Search t

use image sea

Example: When precisely will the new millennium begin?

anslat

DIRECTORY

Automotive

Business & Finance

ALTAVISTA HIGHLIGHTS

POWER SEARCH

BIG changes coming to AltaVista 10/25 -Info

TRY THESE SEARCHES...

Search for Hallowe

images

The Alta Vista search engine in 1999.

Early users of the web, and no one else, will be sad to hear that AltaVista is no more.

Founded in 1995, AltaVista was one of the first search engines for the web and was a leader in the market until the 2000s, when it was overlapped by Google. The search engine was shut down on Monday after its owner, Yahoo!, announced last month that it would be closing down the mid-'90s web relic.

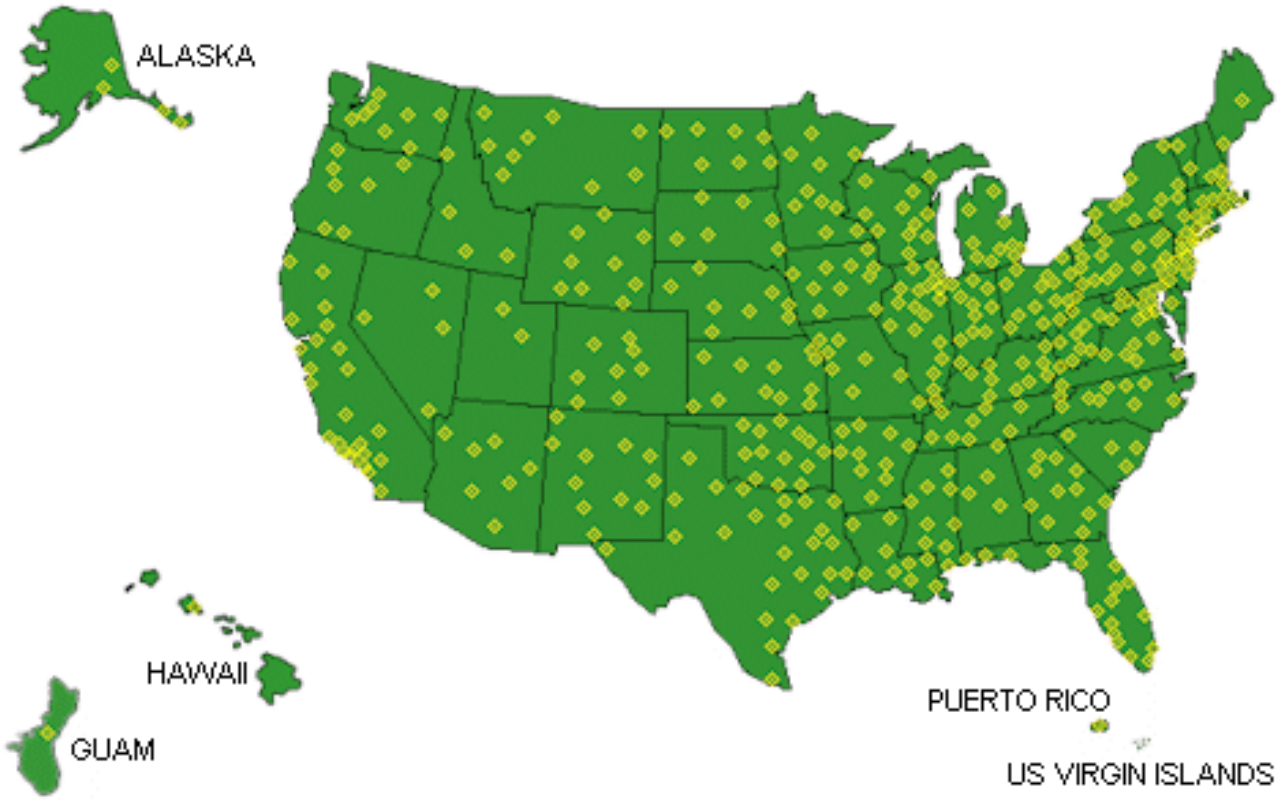
Now, when users go to AltaVista.com, they are taken to the home page for Yahoo! Search, which is powered by Microsoft's Bing.

Yahoo came to own AltaVista in 2003 when it purchased the search engine's then-owners Overture Services for $US1.7 billion.

Example 3: The United States Postal Service

- Yearly delivers more than 200 billion pieces of mail to 144 million locations in the US
- Revenue: about US$70 billion, not increasing
- Quasi-governmental agency, required to exist without government subsidies

Mail Processing Facilities



About 650 mail processing facilities!

Reorganization

- All previous attempts to rationalize the structure of USPS were met with congressional intervention and controversy
- No doubt saving possible, but how?
- US$70 billion/year company becomes a “poster child” for inefficiency and bureaucracy
- What to do?

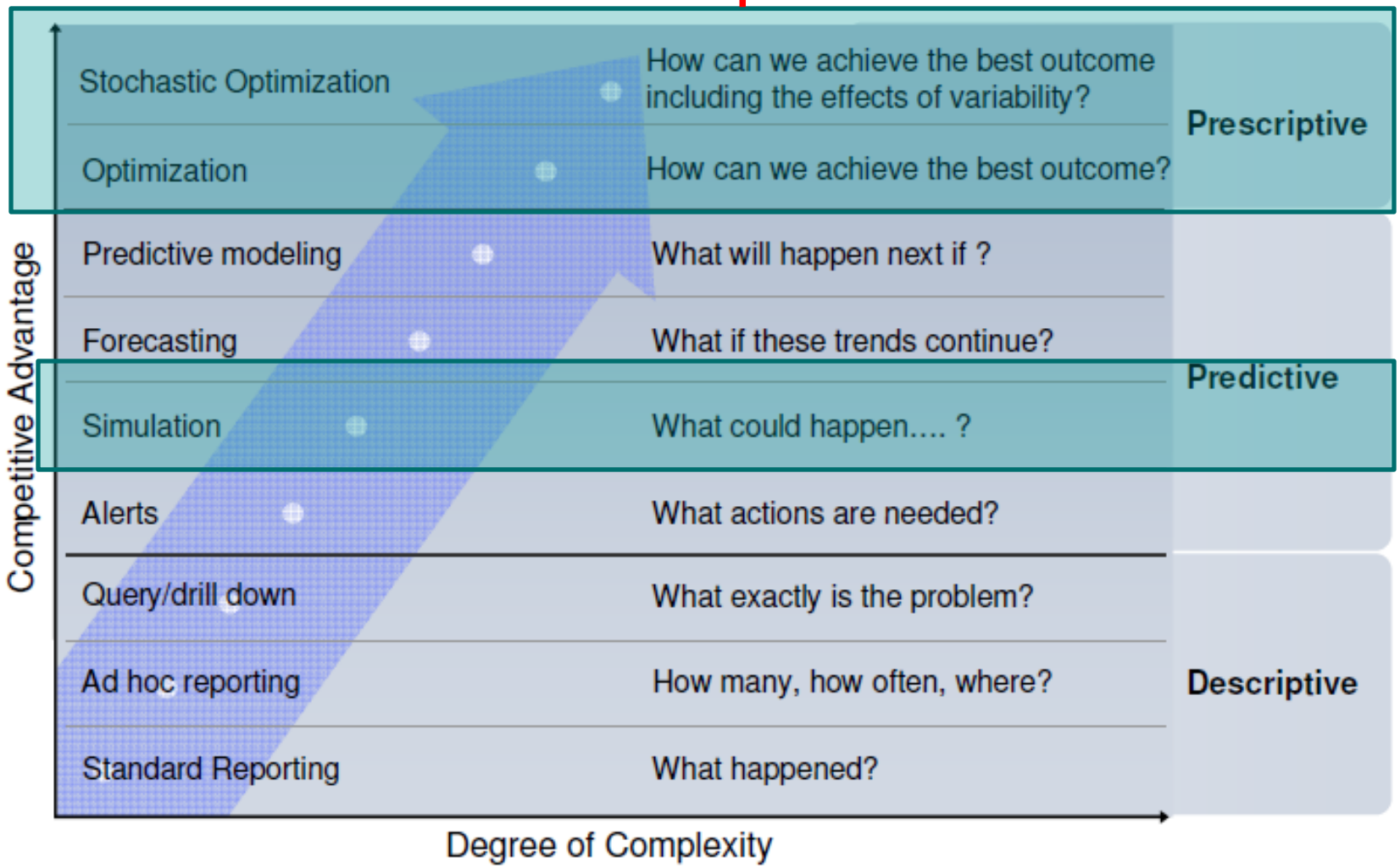
Key Idea

~~Operations Research~~

Business Analytics!

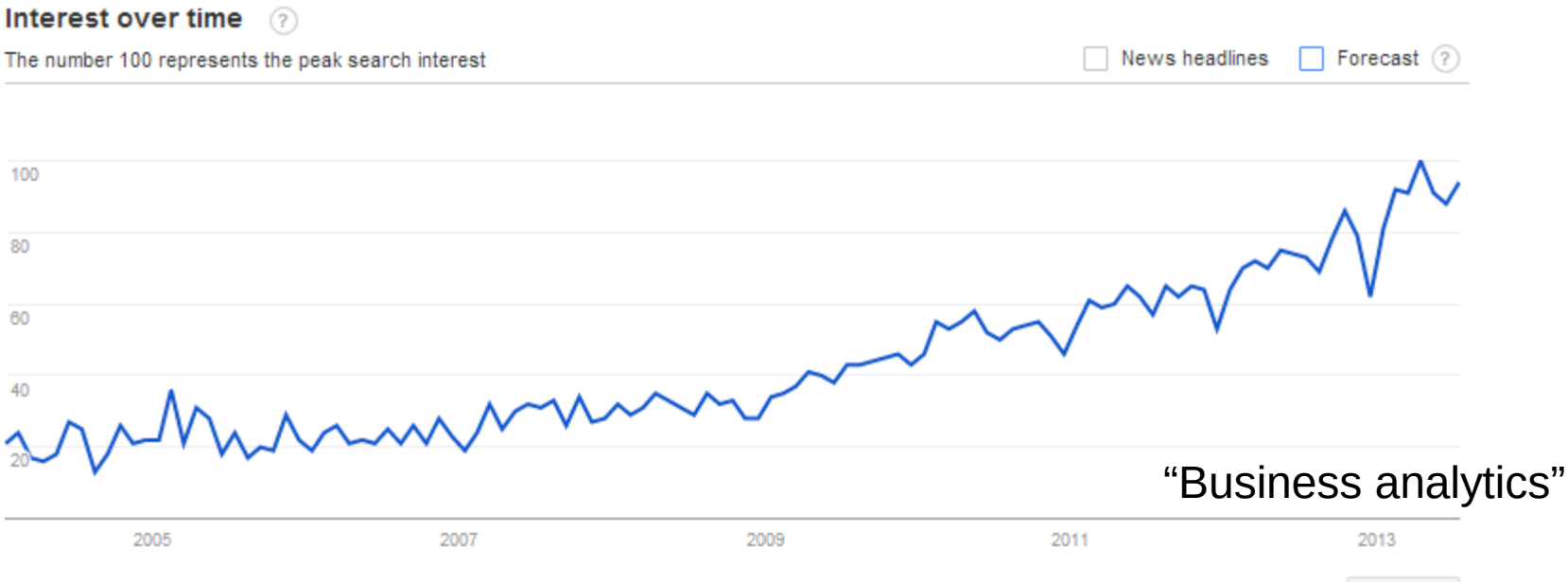
Analytics Landscape

Operations Research



Based on: Competing on Analytics, Davenport and Harris, 2007

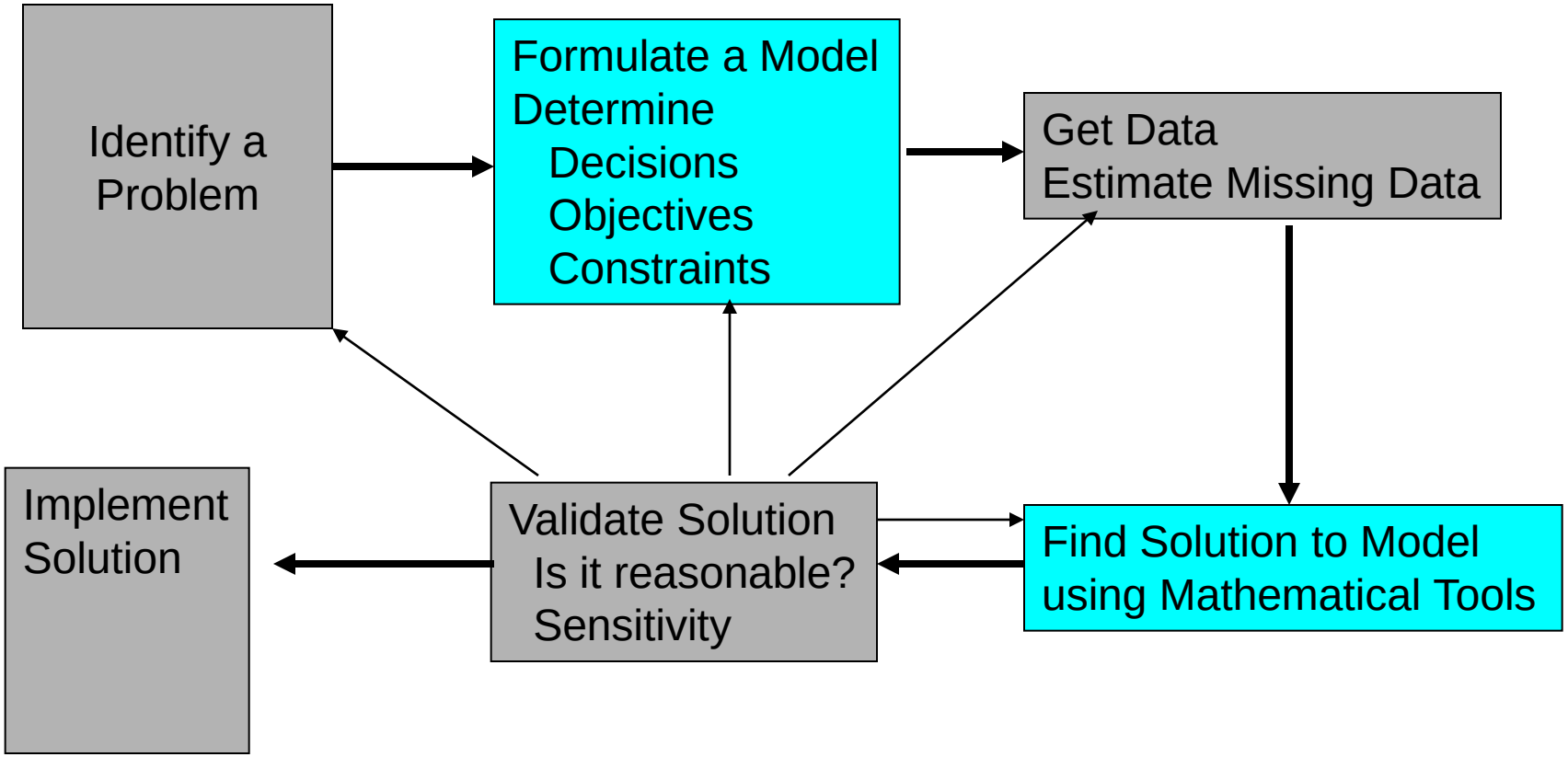
Business Analytics is a Hot Topic



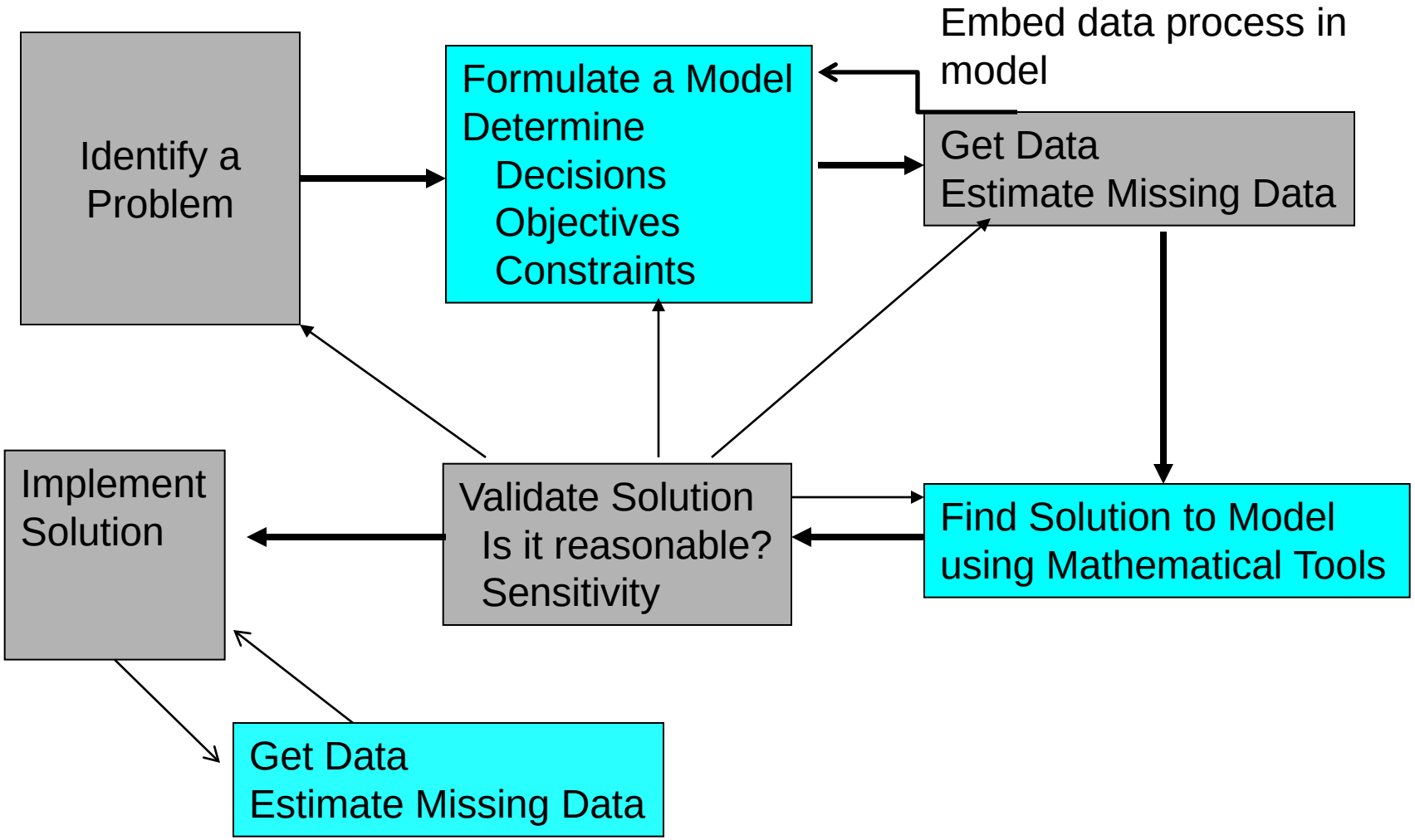
…. Maybe


Interest over time
The number 100 represents the peak search interest
News headlines
Forecast
100
80
60
40
20
Average
2005
2007
2009
2011
Management science
Operations research
Business analytics
Embed

Strawman: The OR Process



The Business Analytics Process



Operations Research/Business Analytics

Using past data to predict the future and make better decisions based on that prediction

- It is really what the best operations research has always been
- Recognition that OR “tool kit” needs to move beyond optimization to include predictive analytics (like data mining)

BUSINESS ANALYTICS: THE PAST OF OPERATIONS RESEARCH

Back to our stories: Fred Smith

- Fortunately, Fred is smart enough to know what he didn't know
- Had a physicist friend analyze the situation
  - Key is route structure
    - Some cities that generate lots of small packages
    - Some cities that require lots of small packages

First model

- Estimate demand for 112 cities (112 by 112 = about 12,000 data points
- Determine competition on routes to get FedEx's share
  - Found some great possibilities: Rochester had tons of small packages (think Kodak) with poor air service
  - New Orleans was poor choice (Delta provided more than enough service)
- Solve problem: what is the most profitable set of cities to include in the network?

- Business Analytics: key is data here!

Result

- 23 city network that was much different than initial one
- Business grew quickly and continues to this day
- Almost no strategic decision is made without OR analysis
- FedEx is an Operations Research company that delivers packages

Example 2: Yahoo! versus Google

- Yahoo! was built on a hand-edited collection of carefully classified links. Pretty reasonable when there were only 30-40 new sites in a day
- My own site was about the 800th site in the world, and joined yahoo as about the fourth operations research site

Google

- Google was built on the automatic identification of the “best page” for any search.
- The question of the “best page” for a search is an operations research question.
- The cleverness of Brin and Page in “solving” this was really in formulating it: the techniques used are (or were: lots has been done since) standard.

Google's page rank

Here's how the PageRank is determined. Suppose that page P_j has l_j links. If one of those links is to page P_i, then P_j will pass on $1/l_j$ of its importance to P_i. The importance ranking of P_i is then the sum of all the contributions made by pages linking to it. That is, if we denote the set of pages linking to P_i by B_i, then

$$I(P_i) = \sum_{P_j \in B_i} \frac{I(P_j)}{l_j}$$

This may
importanc
familiar.

Let's first

Notice th
column is
nonnegati
our story.

We will al
pages. Th

Details unimportant: based on idea of assigning importance to each web page in a consistent way (where importance flows through links on a page)

Easily understood by OR undergrads (you, if you understand eigenvectors!)

In other words, the vector I is an eigenvector of the matrix **H** with eigenvalue 1. We also call this a *stationary vector* of **H**.

Business Analytics

- Real key is to recognize the network of links as data
- Constantly updated as network updates
- Alternative of a static analysis unappealing in this environment

Why did Yahoo! “fail”?

From a Wired article on Terry Semel (CEO 2001-2007)

Many who have met with him at Yahoo say he still doesn't know the right questions to ask about technology. "Terry could never pound the table and say, 'This is where we need to go, guys,'" one former Yahoo executive says. "On those subjects, he always had to have someone next to him explaining why it was important." One could have made a convincing argument two years ago that such deep technical knowledge didn't matter much. But now we have empirical evidence: At Yahoo, the marketers rule, and at Google the engineers rule. And for that, Yahoo is finally paying the price.

By Fred Vogelstein

« back Page 3 of 3

US Postal Service

- To break the political logjam, create an independent, politics-free model and determine an “optimal solution”
- Test the solution against the real-world.
- Use the model to determine cost of political interference

Result

- Results suggest that the correct size of the network is approximately 350 facilities
- Network “looks different”: many more facilities specializing in types of mail, rather than handling all mail from a region

- Initial proposal to close more than 130 facilities. Expected savings: more than US$2 billion/year with improved service

US Postal Service: Results

Sample Conversation:

Congressman: why are you closing Wheeling, WV and moving the work to Pittsburgh? Why don't you close Pittsburgh and move it to Wheeling?

Reponse: That change would cost an additional $12 million in capital costs and $7.8 million in yearly costs. If you can convince your colleagues (including the Pittsburgh congressman) that it is worth it, we'd be happy to make the change.

Congressman: Oh..

Too slow to implement?

Washington Post, March 29, 2013

Postal Service to close more than 50 mail-processing centers

By Josh Hicks, Published: March 29 at 6:00 am E-mail the writer

The U.S. Postal Service announced on Wednesday that it would close more than 50 mail-processing centers this year, accelerating a consolidation plan that the agency developed to trim more than $2 billion in annual costs.

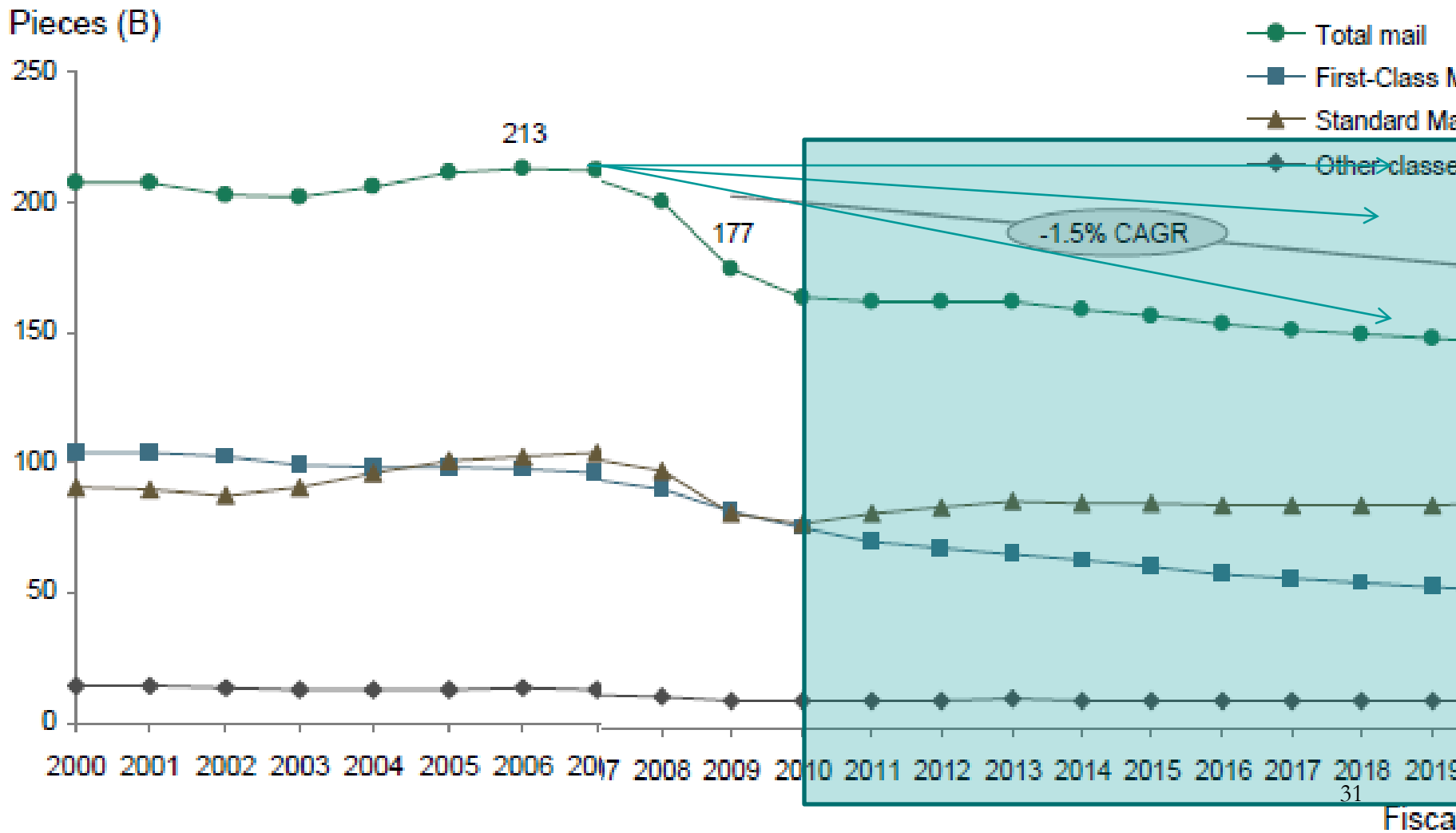
"As first-class mail declines, we must continue to improve operational efficiencies and reduce costs by making better use of space, staffing, equipment and transportation in processing the nation's mail," the agency said in a statement Thursday.

The Postal Service's original consolidation plan called for closing 140 mail-processing centers during the last year and another 89 facilities starting in early 2014.

Business Analytics?

- Given the data environment, difficult to do much prediction
- Where is mail volume going?

Mail volume projections



BUSINESS ANALYTICS: THE PRESENT OF OPERATIONS RESEARCH

Why now?

- Trends are moving towards more OR
  - Increased Data
  - Faster Computers
  - Better Algorithms
  - Lower Fixed Cost for OR

Increased Da

ScienceDaily®

Your source for the latest research news

News | Articles | Videos | Images | Books

Health & Medicine | Mind & Brain | Plants & Animals | Earth & Climate | Space & Time | Matter &

Science News

... from universities, journals, and other research organizations

Big Data, for Better or Worse: 90% of World's Data Generated Over Last Two Years

May 22, 2013 — A full 90% of all the data in the world has been generated over the last two years. The internet companies are awash with data that can be grouped and utilised. Is this a good thing?

An increasing amount of data is becoming available on the internet. Each and every one of us is constantly producing and releasing data about ourselves. We do this either by moving around passively -- our behaviour being registered by cameras or card usage -- or by logging onto our PCs and surfing the net.

The volumes of data make up what has been designated 'Big Data' -- where data about individuals, groups and periods of time are combined into bigger groups or longer periods of time.

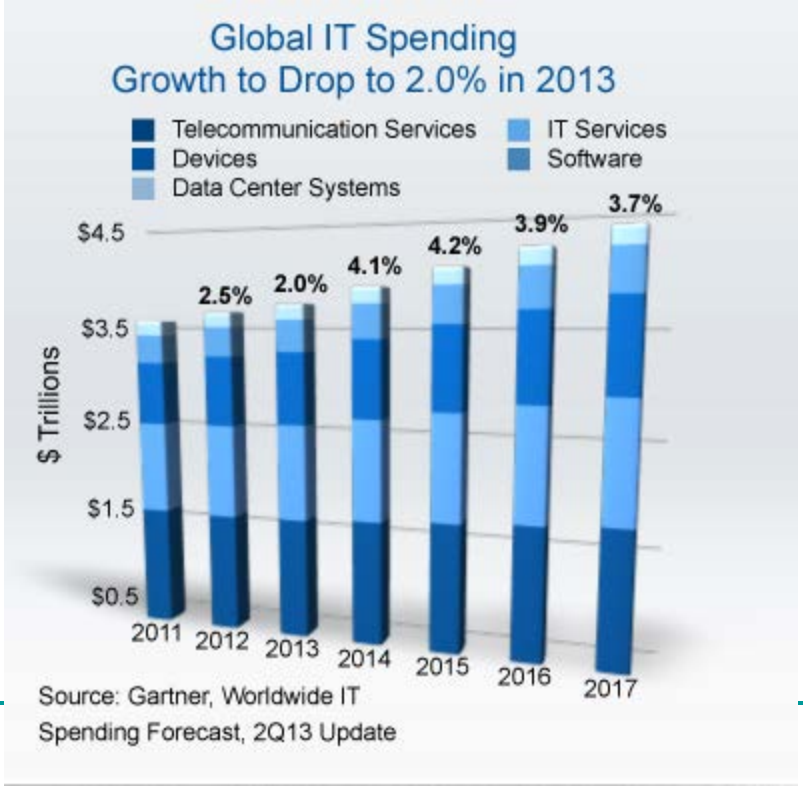
Big Data makes it possible to achieve research results that cover a wide range of issues, and can tell us a great deal about developments in the world in many different areas. (Credit: © zothen / Fotolia)

Research advantages

Petter Bae Brandtzæg of SINTEF ICT points to the huge research centres now developed at internet companies such as Facebook and Google

Big Spending

- Yearly global spending on information technology: $3.5 trillion.



OR Role

- Data is not information; information is not improved decision making

- Operations research/business analytics allows companies to make better decisions based on the data they are collecting

Faster Computers: Supercomputer and otherwise

Intel i7 (6 core)
$569 from
Best Buy

www.top500.org

Faster computers increases the relevance and applicability of OR

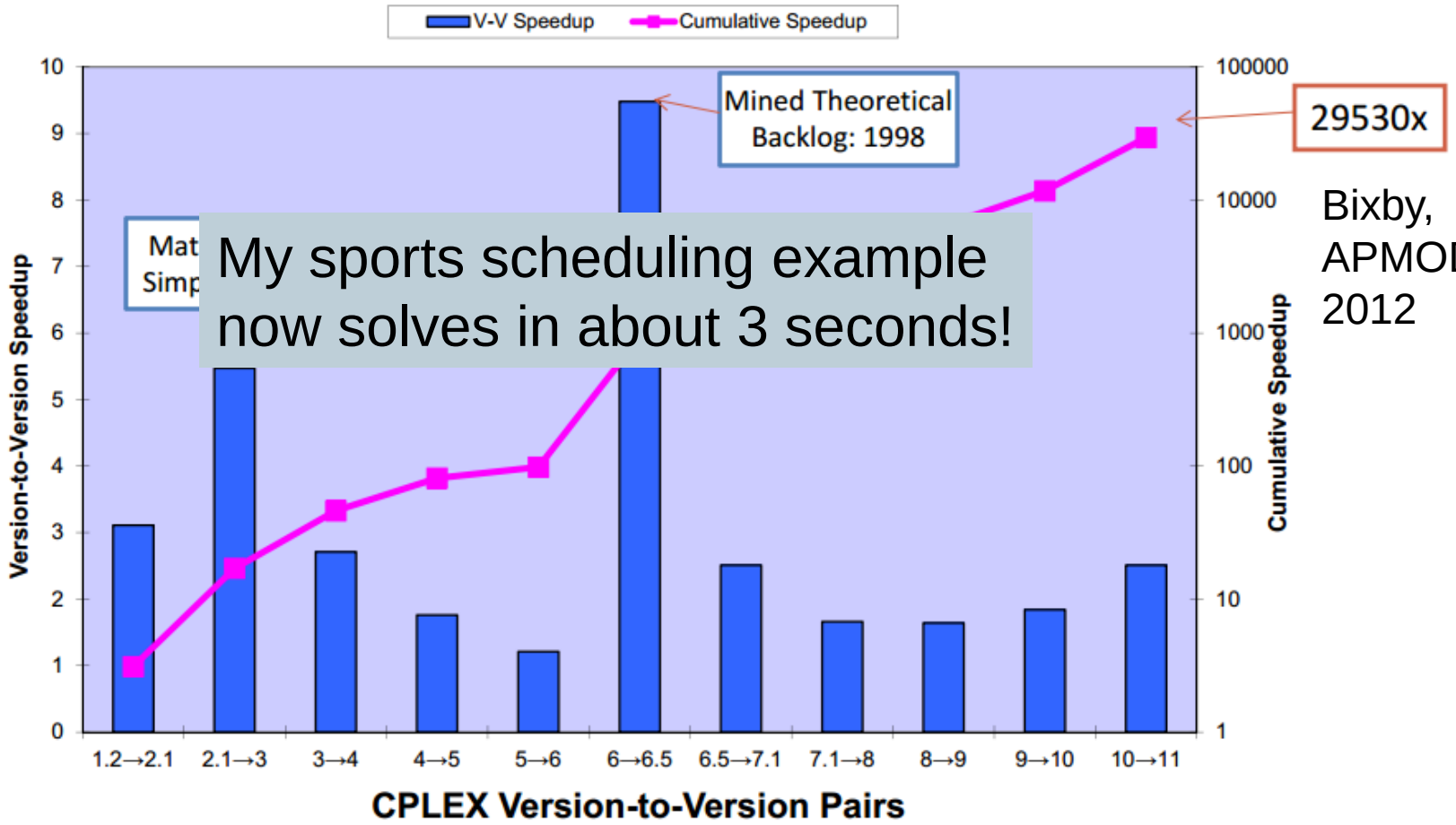
- Example. Sports scheduling example
  - Time (1996): 18 days to solve (impractical)
  - Time (2007): 30 minutes to solve (practical)

Algorithms

- Key to OR is algorithms to find “best decisions” relative to a mathematical model
- Algorithms often very sophisticated: goal is to quickly find the best decision among a huge (e.g. more than the number of atoms in the universe) number of choices
- Extremely active area for both academia and business: there are many firms whose business model revolves around a new algorithm

Algorithms are getting better also!

Speedups 1991-2007



Bixby, APMOD 2012

Speed

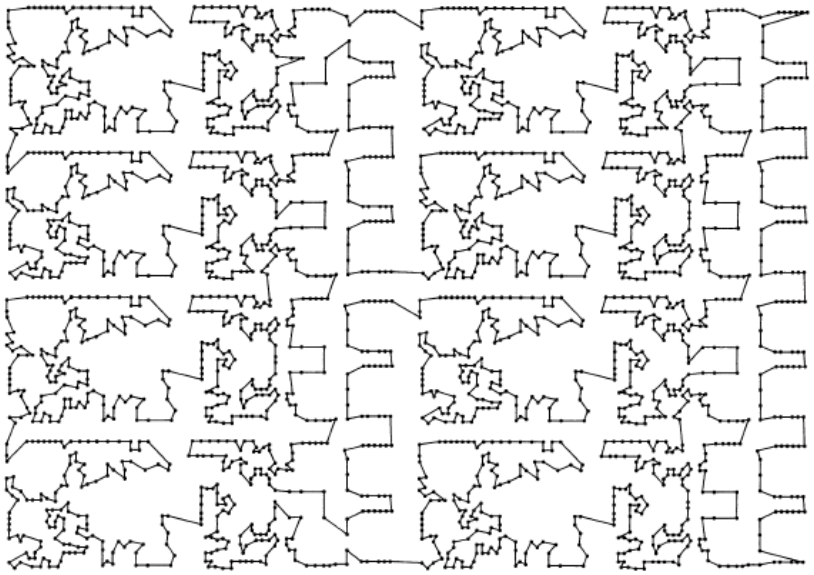
- The combination of improved algorithms and faster computers mean that many problems can now be solved 1 billion times faster than they could 10 years ago.

- Huge increase in applicability of OR methods

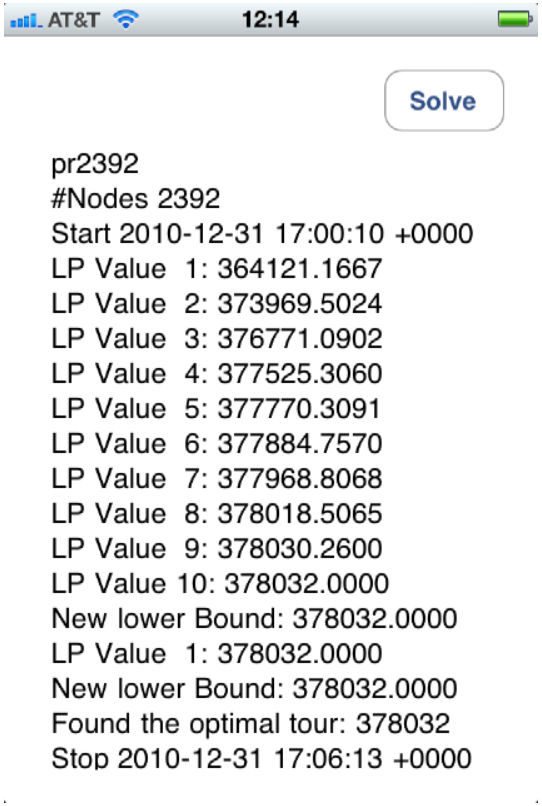
Illustration: TSP with 2392 nodes

1987: Padberg and Rinaldi

when using OSL as the LP solver: While the CYBER 205 runs of the two largest problems TK1002 and TK2392 of our study required about 7 hrs., 18 mins. and 27 hrs., 20 mins., respectively, the IBM 3090/600 runs required about 3 hrs., 10 mins.



2011: Bill Cook and Concorde with iPhone



Lower Fixed Cost for OR

- Past. Big projects for big companies
  - Fighting World War II
  - Airline Crew Scheduling
  - Material Planning at Ford
- Current.
  - Much more accessible
  - OR on your computer (Solver in Excel)

Microsoft Excel - mlb.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

B38 =SUMPRODUCT(B4:M17,B20:M33)

	A	B	C	D	E	F	G	H	I	J	K
8	PIT	480	200	865	226	115	213	740	781	2249	1063
9	CIN	739	428	782	409	211	215	592	535	2021	807
10	STL	1034	731	862	660	484	435	460	237	1728	541
11	HOU	1617	1261	807	1317	1121	1102	1067	654	1623	242
12	MIL	856	646	1078	431	333	239	296	442	1827	846
13	ARZ	2295	2001	1801	1888	1741	1670	1278	1047	635	885
14	LA	2600	2325	2165	2181	2050	1973	1531	1364	328	1246
15	SF	2695	2454	2398	2267	2160	2074	1586	1504	13	1474
16	SD	2581								452	1180
17	COL	1753								949	643
18											
19	***Assignments***	BOS								OAK	TEX
20	WAS	0								0	0
21	PHI	0								0	0
22	ATL	0								0	0
23	FLA	0								0	0
24	PIT	0								0	0
25	CIN	0								0	0
26	STL	0								0	0
27	HOU	0								0	0
28	MIL	0								0	0
29	ARZ	0								0	0
30	LA	0								0	0
31	SF	0	0	0	0	0	0	0	0	0	0
32	SD	0	0	0	0	0	0	0	0	0	0
33	COL	0	0	0	0	0	0	0	0	0	0
34		0	0	0	0	0	0	0	0	0	0
35		=	=	=	=	=	=	=	=	=	=
36		1	1	1	1	1	1	1	1	1	1
37											
38	***Total Distance***	0									

Solver Parameters

Set Target Cell: $B$38

Equal To: Max Min Value of: 0

By Changing Cells:

$B$20:$M$33

Subject to the Constraints:

$B$34:$M$34 = $B$36:$M$36
$N$20:$N$33 <= $P$20:$P$33

Solve | Close | Guess | Options | Add | Change | Delete | Reset All | Help

DIST

BUSINESS ANALYTICS: THE FUTURE OF OPERATIONS RESEARCH

Future is bright!

1010101010
0101010101

Business Analytics

$16 billion
in revenue by 2015

Opportunity:
Global data volumes are predicted to increase by 29 times over the next 10 years to 35 zettabytes.* (A zettabyte is a 1 followed by 21 zeros.)

2015 Road Map Objective:
Business analytics revenue is expected to be $16 billion by 2015.

Enterprises need a way to manage and mine the deluge of potentially valuable information, and the key is advanced data analytics. IBM spotted this emerging need early, building the world's leading analytics practice—with 7,800 expert consultants, the world's premier nonacademic mathematics function and the acquisition of 25 companies, for $14 billion in gross spending, to deepen our capabilities.

Our scientists have received more than 500 analytics patents. They are expanding technology frontiers through breakthroughs like the powerful new computer named Watson, which competed and won on the television quiz show *Jeopardy!* Applying Watson's use of advanced analytics to decipher natural language, IBM is working to identify better healthcare diagnoses, potential drug interactions and "what if" scenarios in finance and compliance.

IBM 2010 Annual report; contrast "cloud" at $7 b and smarter planet $10 b

Themes for the Future

- Closer link between predictive analytics and prescriptive analytics
- Handling uncertainty to create robust solutions
- Exploiting parallelism, or lose Moore's Law
- New areas of application
- Optimization with adversaries

Example

- Large automobile parts supplier
- Receives estimates of order quantities six months in advance (which are then regularly updated)
- Uses estimates to do capacity planning, shift scheduling, etc.

Capacity Planning

- Well solved, classical operations research problem
- Issues of level of detail, scheduling, etc. etc.
- But what about data…?

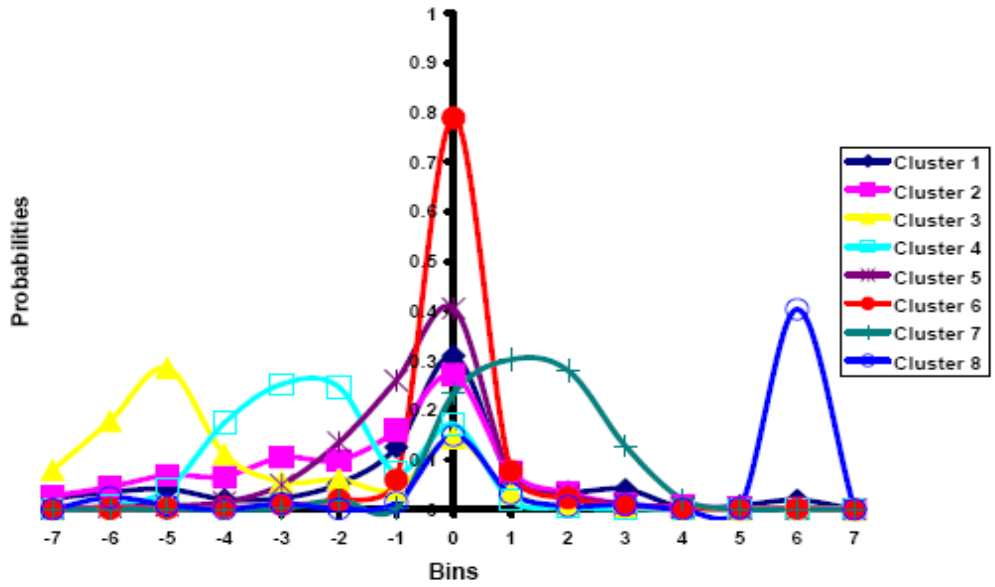
Problem

- Estimates were (sometimes) terrible:
- January: “I’ll need 1000 in June”
- February: “I’ll need 2500 in June”
- March: “I won’t need any in June”
- April: “I’ll need 10,000 in June”
- May: “I’ll need 500 in June”
- June: “Where are the 5,000 I need?”

Business Analytics Approach to Capacity planning

- Capacity planning needs to adapt to the messy data environment.

One step: cluster customers based on accuracy of their forecasts



Other Examples

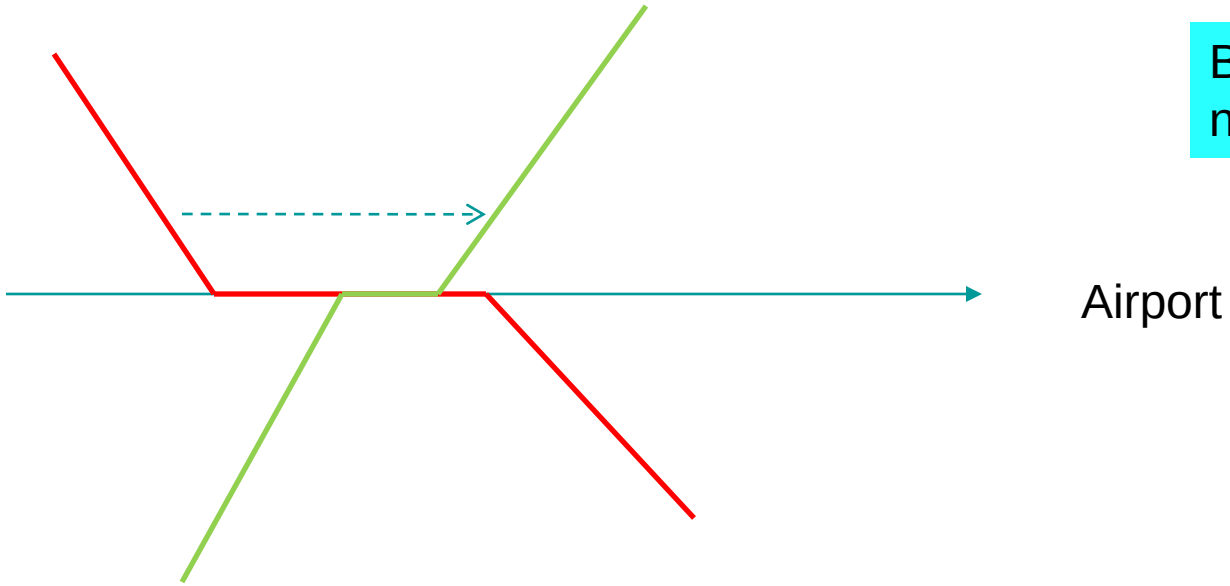
- Revenue management and prediction of “no-shows”
- Text mining to estimate stock variability and optimal portfolios

Handling Uncertainty and Robustness

- Many approaches for handling uncertainty in models
  - Replacing with mean, or some %tile (USPS capacity planning)
  - Scenario optimization
  - Stochastic programming
  - “Robust” optimization (assumptions on ranges, number of changes of data)

True measures of robustness?

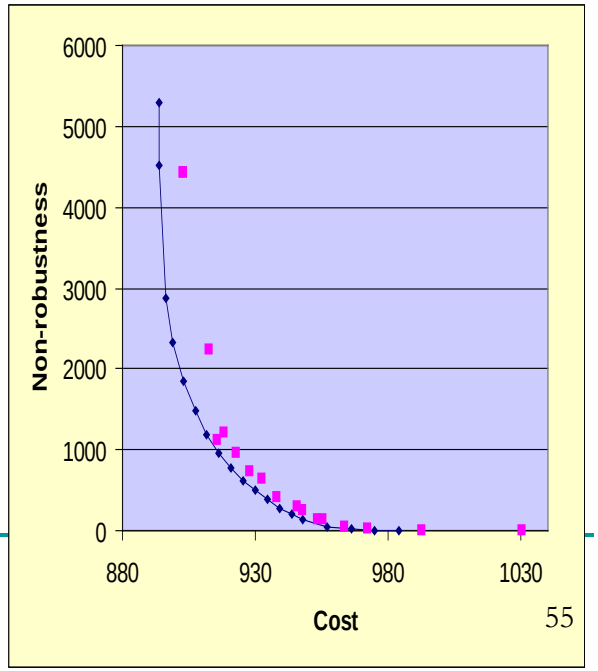
- What does a robust solution really mean?
- David Ryan and others on robust airline crew schedules



Both planes now late

Measure of Robustness

- Unclear where uncertainty is. Model storms? Politics? Baggage handling?
- Use “plane changes” as surrogate. Schedules with few plane changes are naturally more robust.
- Get robustness at little cost

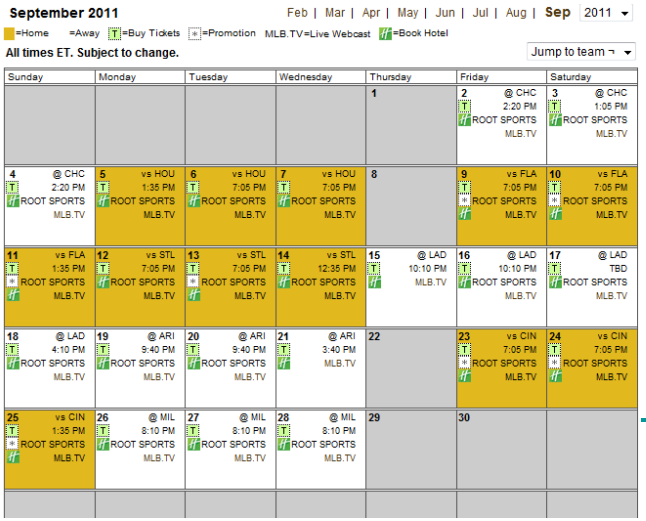


Data-free Robustness

- Bartholdi and Eisenstein: scheduling shuttle buses. Don't! At checkpoint, use rule "Wait 60% of the time until the bus behind will arrive at the checkpoint"
- Robust to changes in traffic congestion, passenger load times, and even number of buses!

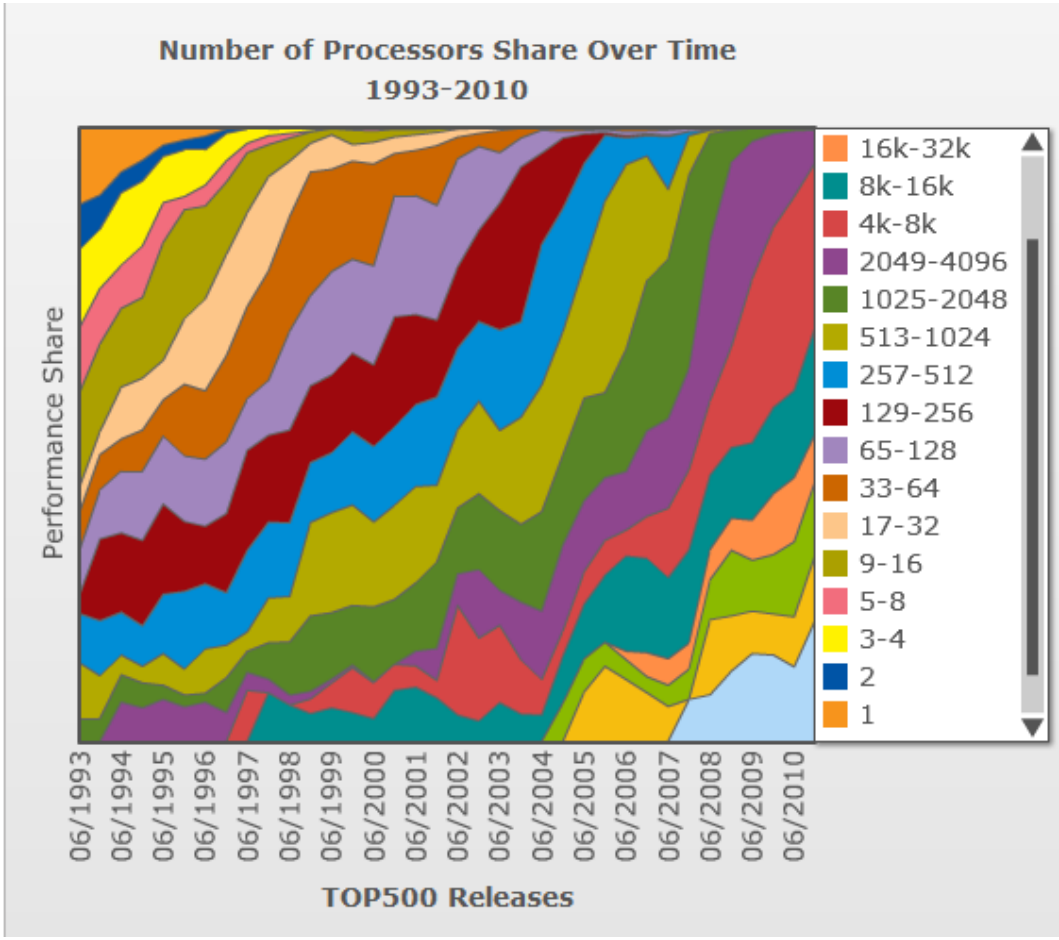
Other applications?

- Badly need some measure of robustness for financial markets: historical variance doesn't seem to work well.
- Scheduling: include simple robustness in Major League Baseball scheduling



Parallelism

- Back to the “top 500”:



Risk for some types of operations research

- Some of what we do seems hard to parallelize:
  - Simplex-based linear programming
  - Many network algorithms
  - Dynamic Programming

Effect on Mixed Integer Programming

- Hans Mittelmann's benchmarks:

s	problem	CPLEX6	GUROBI6	CBC6	CPLEX12	GUROBI12	CBC12
1	air04	8	7	31	9	7	31
	cap6000	1	1	1	1	1	2
	mas74	103	56	166	100	47	95
	mod011	15	12	13	15	12	14
	mzzv11	25	12	97	24	12	100
	mzzv42z	13	3	67	13	3	78
	pk1	12	6	13	8	5	9
	qiu	6	7	55	4	5	35
2	ran12x21	6	10	22	6	8	12
	ran13x13	3	5	12	3	3	9
3	binkar10_1	4	2	95	6	1	65
	lrn	24	65	475	25	24	355
	prod2	18	8	17	14	8	17
4	bc1	25	30	228	14	30	178
	bienst2	13	7	390	9	5	162
	dano3_5	140	81	169	173	79	162
	mark._4_0	20	11	42	10	10	34
	mark._5_0	1053	1837	f	575	691	f
	qap10	3	56	156	3	56	157
	seymour1	37	32	693	28	32	210
	swath2	7	4	11	7	3	9
	swath3	27	9	181	27	13	60
	30_05_100	9	9	153	9	9	141

Not a lot of halving of time between the 6 and the 12 columns!

Challenges

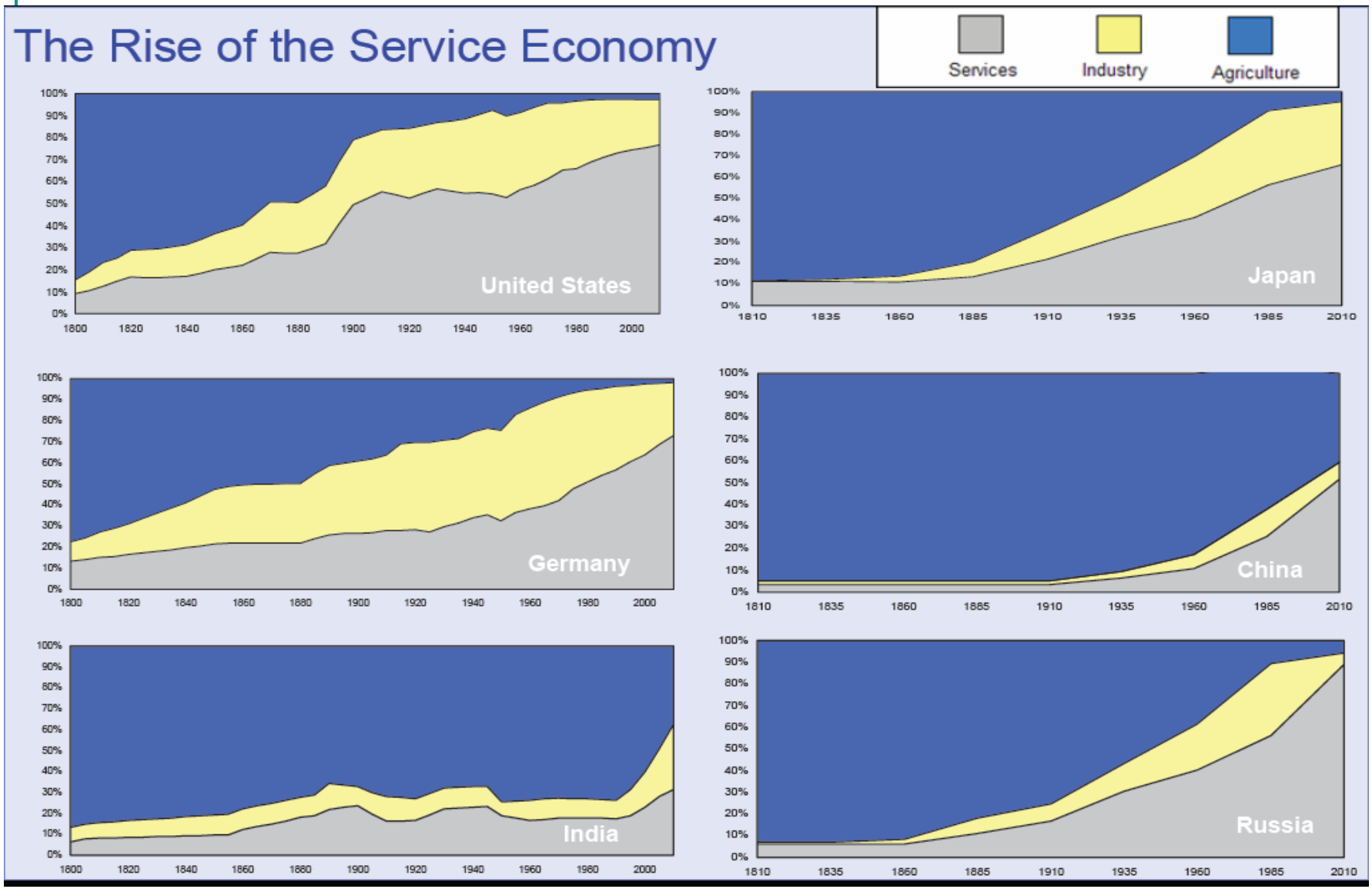
- In order to keep up with computer speed improvements, we need to improve parallel performance
- Alternative algorithms? Perhaps formerly “failed” algorithms now seem more appealing?

New Areas of Application

- Traditional view of OR:
  - Manufacturing
  - Services limited to transportation and logistics

- New view
  - OR everywhere
  - Services are a great opportunity

The Rise of the Service Economy



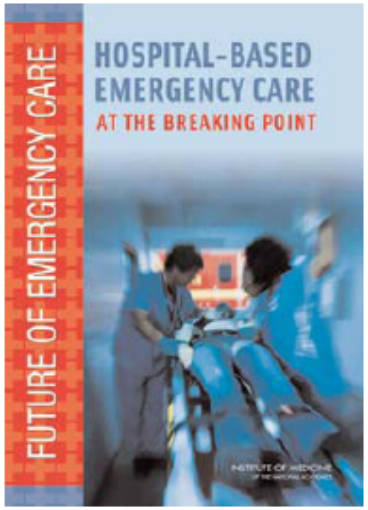
REPORT BRIEF • JUNE 2006

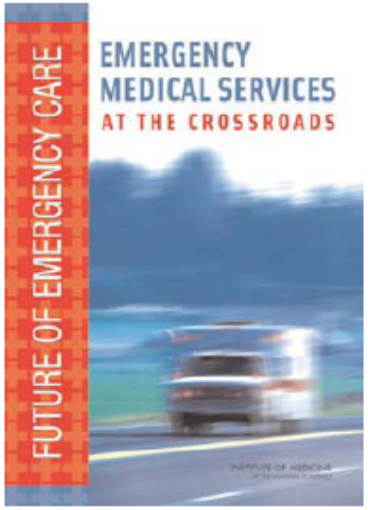
The Future of Emergency Care in the United States Health System

If one were to judge solely from popular TV shows, the nation's emergency care system is in fine shape. Its doctors, nurses, and ambulance personnel are dedicated and competent professionals who save lives with their expertise and state-of-the-art equipment and can always be trusted to come through in a crunch. And, indeed, there is a great deal of truth to this picture: Our emergency and trauma care system has made tremendous strides over the past few decades, and today it manages to save many lives that just ten or twenty years ago would have been inevitably lost.

But underneath the surface, a national crisis in emergency care has been brewing and is now beginning to come into full view. Emergency departments (EDs) across the country are overcrowded. Ambulances are turned away, and patients, once they are admitted, may wait in hallways for hours or even days before inpatient beds open up for them. Often the specialists that patients need to see are not available. And the system that transports patients to the hospitals is fragmented and inconsistent in the level of quality it provides.

It was against this backdrop that the Institute of Medicine's





RECOMMENDATIONS

To improve the nation's emergency care system and deal with the growing demands placed on it, the committee recommends a multi-pronged strategy. Together the three reports contain a number of recommendations, but the main thrusts of the recommendations can be summarized by four basic themes:

Improving hospital efficiency and patient flow

Tools developed from engineering and operations research have been successfully applied to a variety of businesses, from banking and airlines to manufacturing companies. These same tools have been shown to improve the flow of patients through hospitals, increasing the number of patients that can be treated while minimizing delays in their treatment and improving the quality of their care. One such tool is queuing theory, which by smoothing the peaks and valleys of patient admissions has the potential to eliminate bottlenecks, reduce crowding, improve patient care, and reduce cost. Another promising tool is the clinical decision unit, or 23-hour observation unit, which helps ED staff determine whether certain ED patients require admission. Hospitals should use these tools as a way of improving hospital efficiency and, in particular, reducing ED crowding.

Lots of other areas of application

- Medical procedures
  - Kidney transplant chains
- Education
- Urban development
- Call center optimization
- Customer optimization

Adversaries

- As OR combines with predictive analytics, more likely to face adversaries working against model assumptions
  - Best Buy searching for “bad customers”
  - Google in never-ending fight against spammers and those who game rankings
- But we are pretty good at that.

NATO-UNCLASSIFIED

MXP-1(D)(N

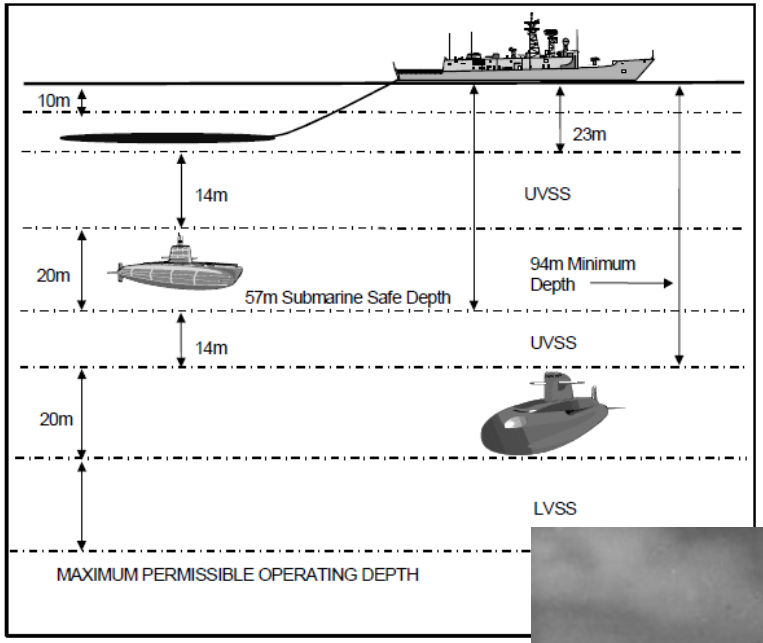


Figure 2-2 Submarine Safe Depth Ca

Which completes the loop to OR Past

Summarize: Business Analytics and…

- OR Past
  - Fedex
  - Google
  - USPS

  All examples of key role of data
- OR Present
  - More Data
  - Faster computers
  - Better algorithms
  - Lower fixed Costs
- OR Future
  - Predictive Analytics
  - Robustness
  - Parallelism
  - New applications
  - Adversaries

More at

http://mat.tepper.cmu.edu/blog