Artificial Intelligence in Industry: The good, the bad and the ugly

Gregg Vesonder
AT&T Labs - Research
Shannon Labs
gtv@research.att.com

AT&T

Roadmap

- About me
- Early History
- "Golden Age of AI"
- More Advanced AI
- OO and .com eras
- Current Work
  - Mise en place
  - Preservation
  - Case Studies - Some ESs, KDD Paper
  - Data Mining Gastronomique
- Software Engineering

Vesonder's Relevant Bio

- Software for 30+ years
- PhD in Cognitive Psychology - Computer modeling of learning and memory
- [Bell|AT&T] labs for 25 years
- Dozens of projects, many AI but not all
- OO and .com (Online Platforms)
- Stevens: Software Engineering, Software Architecture and Design

What's AI

- Artificial Intelligence not only tries to understand intelligence but also build intelligent entities (Russell & Norvig)

Cognition	Systems that think like humans	Systems that think rationally
Behavior	Systems that act like humans -> empirical science	Systems that act rationally -> math and engineering

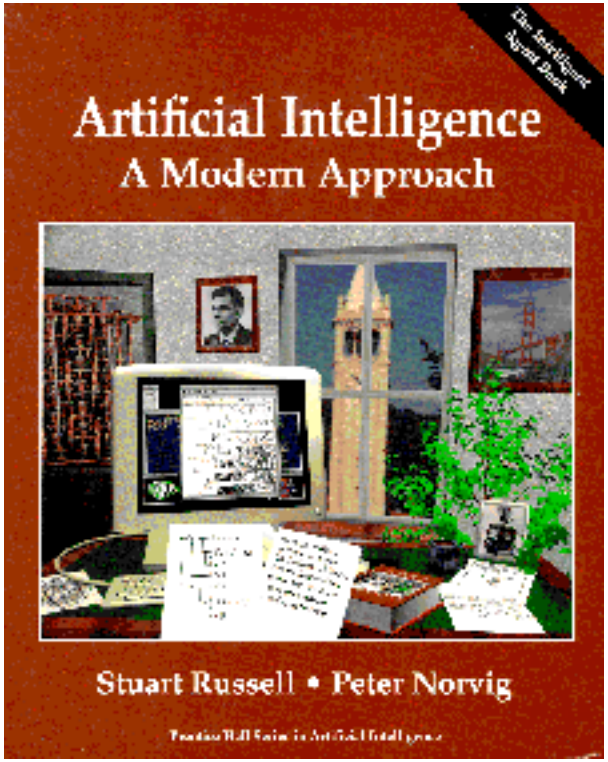
AT&T

AI & Agents

- Agent: something that acts, operates under autonomous control, perceiving the environment, persisting over long time periods, adapting, being able to take on another's goals
- A rational agent achieves the best outcome or, given uncertainty, the best expected outcome
- Perfect rationality - always doing the right thing, is not feasible in complicated environments
- Limited rationality - acting appropriately when there is not enough time (or ability or feasibility) to do all the calculations one might like

Reference

Russell & Norvig Artificial Intelligence: A Modern Approach 2nd Edition, Prentice Hall, ISBN: 0137903952 (now green cover)



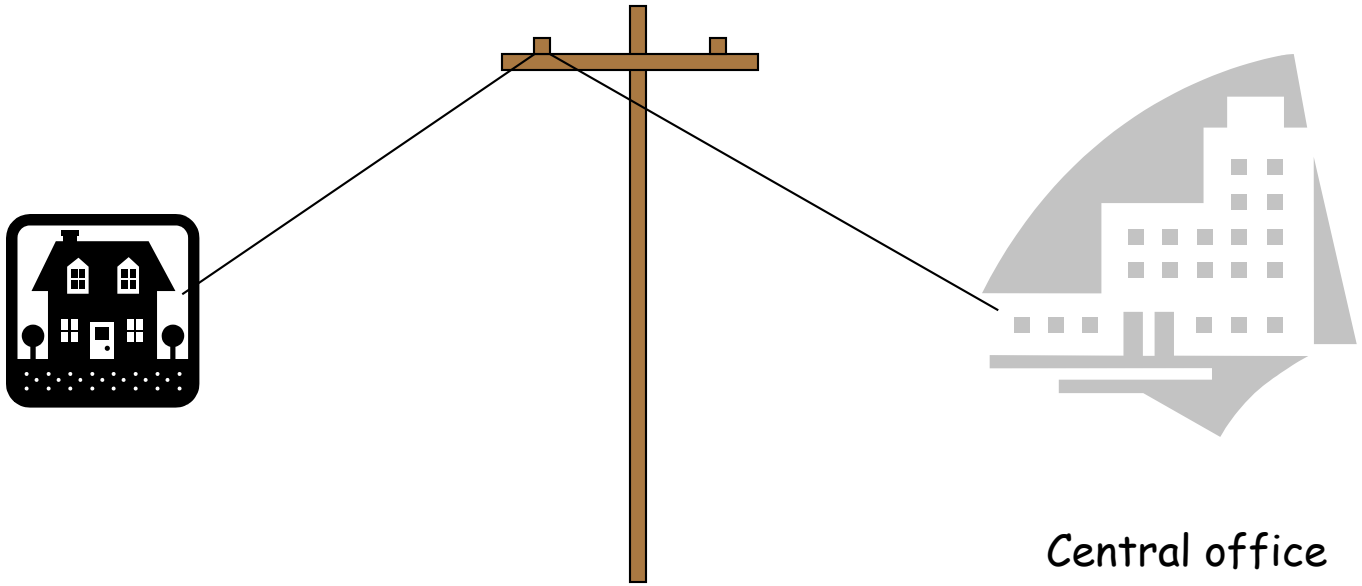
AT&T

My First AI Project

(in industry)

- Serendipity - in the field with another system (CRAS)
- Noticed analysts struggling with reports -- doing correlations
- Met after work
- Generated a proposal -- this was in 1979
  - DEC had R1 - rule based system for configuring vaxes
  - Schlumberger had the Dipmeter Advisor
  - Stanford had MYCIN

ACE



Knowledge Engineering Technique

- Select a specific topic
- Gather "book knowledge" of topic and construct rules
- Incorporate rules in a computer program
- Run program on real situations
- Discover where program differs from what the expert suggests
- Repeat last three steps until program agrees with expert (satisficing)

ACE Rule

IF a range of pairs within a cable has generated a large number of customer trouble reports

ANDIF a majority of work on these pairs was done in the terminal block

THEN look for a common address for these pairs

AT&T

Some other rules

MYCIN:

IF the infection type is primary-bacterima
ANDIF the suspected entry point is the gastrointestinal tract
ANDIF the site of the culture is one of the sterile sites
THEN there is evidence that the organism is bacteriodes

DIPMETER ADVISOR:

IF there is a red pattern over a fault
ANDIF the direction is perpendicular to the fault
ANDIF the length is > 200 feet
THEN the fault is a growth fault

AT&T

Lessons Learned

- Autocratic vs. Democratic
- Flavors of expertise: medical diagnosis versus cable analysis
- Comfortable user interface
- Software interface:
  - UNIX + AI
  - ~30% of work is AI related
  - Quick code development
- Rules and procedures
- Skilled Knowledge engineers - know the domain
- Added value, part of the work flow
- Personal - role of luck and the thrill of seeing your efforts used.

Digression

- A2b music: November 1997
  - Really seeing folks use stuff
- iTunes today
- Back to AI

Led to Major Initiatives In AI: The "golden" age

- ALVEY - Great Britain
- ESPIRIT - European Common Market
- FIFTH Generation - JAPAN
- Strategic Computing - United States
  - Autonomous land vehicle
  - Pilot's associate
  - Fleet manager
- And most corporations

AT&T

More Expert Systems

- OKIES
- MIDS
- SARTS AUTOTEST
- Other AI LODE - kernel knowledge bases from databases
- Led to Environment tuned to our needs - C5 and ES technology embedded in larger systems

AT&T

In Telecom

- Maintenance
- Provisioning (forecasting, planning & designing)
- Network Administration (routing, billing, facility assignment)

AT&T

Expert Systems in Telecom

	'83	'84	'85	'86	'87	'88	'89
Maintenance	1	1	2	5	3	9	4
Provisioning	-	-	-	1	1	3	1
Network Administration	-	-	2	1	-	3	1
TOTAL	1	1	4	7	4	15	6

AT&T

More Advanced AI: PROSE

- Knowledge based configurator for switch adjuncts
- Used C-Classic - pure declarative form, provided quick and consistent knowledge editing
  - Forces consistency by:
    - Straightforward representation of product's compatibility rules
    - Enforcement of consistency and integrity of the compatibility rules and the configurations they generate
    - Generates configurations from minimal number of inputs

AT&T

C-Classic Inference

- Classification - find all descriptions applicable to an object: find all descriptions more general and specific than a proposed object description
- Completion or propagation of logical consequences, including, but not limited to inheritance
- Contradiction detection
- Simple forward chaining rules or triggers
- Dependency maintenance for retraction and recovery

AT&T

AI Winter - OO and .com

- Many of the initiatives did not meet expectations - drop in attendance
- Most went to OO (cfront) and then on to web (a2b, cospace-community in general)
- Signs of spring

Web- Cospace


Netscape: Cospace
SHARED-CORE research.att.com
construct.net
giger.com
mojo
tk@tess
cospace.research.att.com 7 users 10 portals [-10.79 5.03 7.29 5.39] <1/s>
;; mojo has arrived
;; solstice has arrived
;; solstice has left
;; solstice has arrived
;; solstice has left
mojo: so long, solstice, and may all your chi-squared's be unity
mojo: i'm outta here...
tk: likewise
Who
Admin
Unlink
Core
Help
Quit
Type:

AT&T

Datamining

- An information extraction activity whose goal is to discover hidden facts contained in databases. Using a combination of machine learning, statistical analysis, modeling techniques and database technology, data mining finds patterns and subtle relationships in data and infers rules that allow the prediction of future results. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis.
- http://www.twocrows.com/glossary - datamining company

AT&T

Newest Stuff: AI and Datamining

- Systems Approach (work with Wright and Dasu)
- Unique issues and combinations of issues
  - Mise en place
  - [most|all] runs are unique
  - Data Quality is crucial
  - Granularity
  - Downstream systems
- Process issues
  - Knowledge engineering throughout
  - Verification and validation issues

AT&T

"Bouillabaisse" Data Mining

- Data exists in some repository/corpus
- Know the fields and relationships
- At least familiar with some domain
- Others have mined the data - community
- Reference efforts -- helps Verification (built system right) and Validation (built right system)
- …
- World Wide Telescope - Jim Gray

AT&T

Stone Soup Data Mining

- A Fable in many parts
- The data is not in one place, in fact it is in many places
  - Don't know the quality
  - Don't know what it means and there is no one source to discover it (multiple, conflicting experts - Brooks "never go to sea with two chronometers, go with one or three")
- Data does not remain there - have to capture it -- usually on arcane systems

AT&T

Stone Soup -2

- Once you get it - more experts, pilot runs (very much like Knowledge Engineering technique)
  - BTW it is in EBCDIC, described by COBOL copybooks, you're running UNIX…
- Discover you need other data to interpret it - back to previous page
- At this point it has been months - if lucky
- Time to formalize the collection process
- Did I mention the data is huge! Terabytes to Petabytes
- Time to do some "data mining" - knowledge and quality
- Archiving issues - reproduction (depends on what is available and who contributes)

AT&T

Byte Digression

- 1 Megabyte ~ 1 Minute MP3 music
- 1,000 Meg = 1 Gigabyte = 250 4 min songs (1.6 CDs)
- 1,000 Gig = 1 Terabyte = 250,000 songs (1600 CDs)
- 1,000 Terabytes = 1 Petabyte = 250,000,000 songs, 150,000,000 hours of music - 6,250,000 hours ~ 713 years of continuous music listening! 1.6 Million CDs

AT&T

More Bytes

Information	Bytes
Typewritten page	2 Kilobytes
Shakespeare	5 Megabytes
Pickup truck filled with books	1 Gigabyte
50,000 trees made into paper and printed	1 Terabyte
All US academic research libraries	2 Petabytes
All hard disk capacity developed in 1995	20 Petabytes

http://searchstorage.techtarget.com

Knowledge Engineering Technique

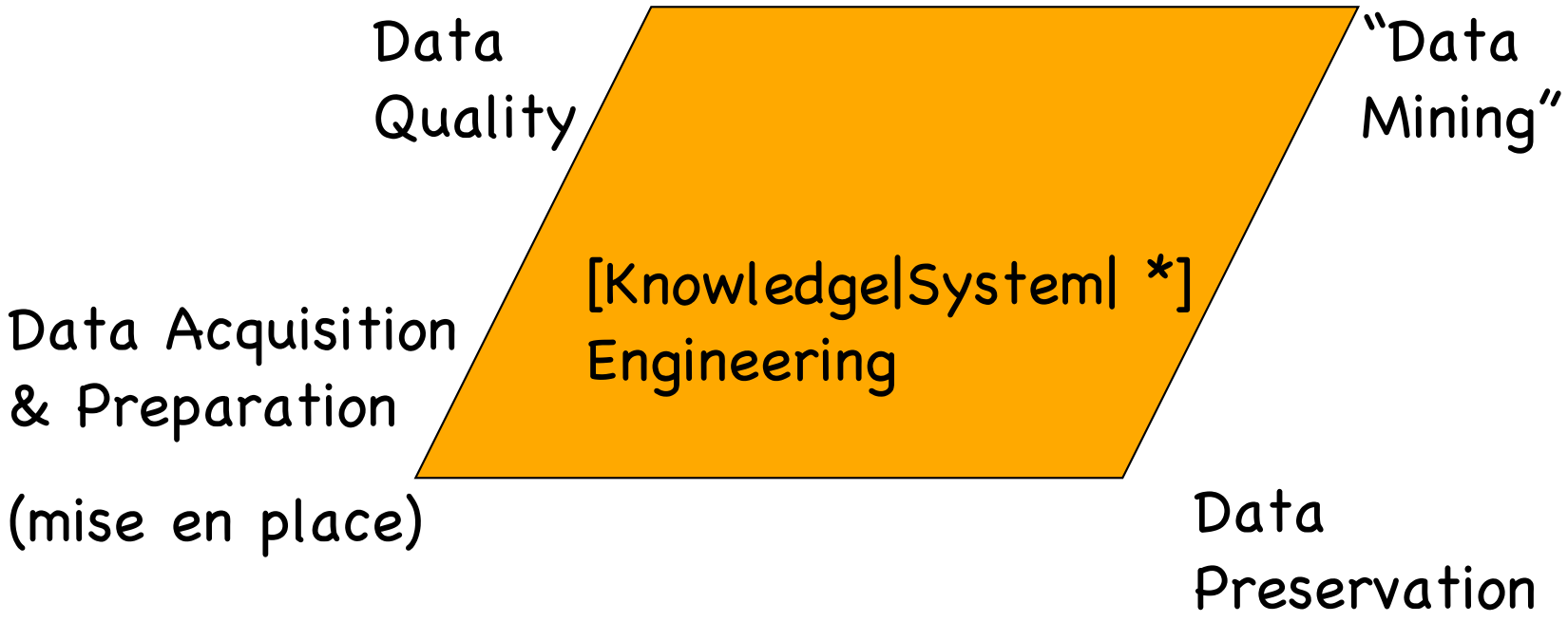
- Knowledge Engineer becomes familiar with domain, architecture and operation
- KE meets with experts to understand operations and issues
- Team uses knowledge to create first (and subsequent) passes at working system
- Experts critique results, provide new knowledge and iterate on previous step until a satisfactory (or best possible) conclusion is achieved

AT&T

Stone Soup-3

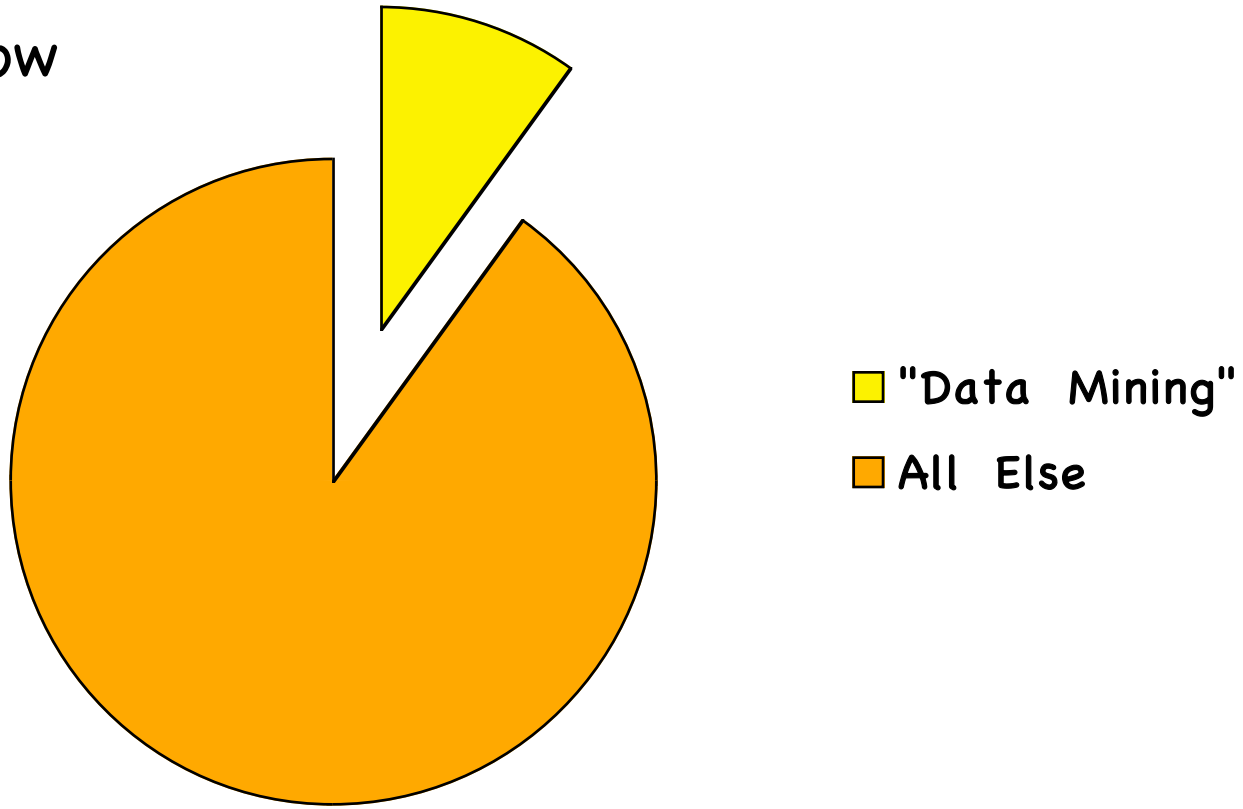
- About this time one of your feeds changes - actually it was several months ago
- Verification and validation throughout
- Preservation of data, summarized data, interim reports and techniques - really time "capsules"

A View of the Space



A Rough Estimate of the Effort

Of course the 10% can grow over time, but…



AT&T

The Life Cycle

- Discover data needed - KE
- Get data/Establish Feed
  - Discover and perhaps get additional data to interpret data - KE
  - Verify & Validate feed
  - Assess data quality
- Discover Reference results for V & V (may be earlier)
- Prepare environment and Run Data
- V &V - KE (iterate - may take you to top again)
- Preserve environment and archive
- Continuously check "upstream" issues - improve data quality
- Usually there is increased level of understanding

AT&T

Preservation

- No ready made archives - recall the size, tape is daunting. 1+ Petabyte year
- Preserve data, software and comparisons
  - Data and meta data synchronized (e.g. time dependent)
  - Redundancy, security, ..
  - Recoverability
- Testing

AT&T

Mise en place

- "put in place" chopping, mincing, measurement, peeling, washing
- Significant planning activity to start a run
  - Data ready - off tape and accessible - could be N different feeds
  - Data verified
  - Sufficient system resources (disk, memory, …)
  - Consistent software builds
- Candidate for AI planning techniques, ES for monitoring run (insuring available disk resources, trapping failures, …)

AT&T

Data Mining Gastronomique

- Data Quality - see Parni & Ted book reference
- AI Techniques:
  - Planning - especially for Mise en place
  - Expert Systems - Rule base/Agent systems for monitoring/quality
    - Also use Ganglia and other tools
  - KE at most points

AT&T

Datamining Conclusions

- Provider a broader view of what constitutes data mining
- Process orientation - addresses complete system development
  - Sometimes the data isn't on the web, in a corpus or on a CD
  - Quality issues
- Mise en place a big issue, since each run is special
- AI everywhere

AT&T

General Conclusions

- The Good: new technology, high productivity, satisfied users, sense of accomplishment, span of effect, resources as a function of success. Not a paper chase, chance to work on significant problems
- The Bad: many jump on bandwagon, signal to noise ratio is high, many misconceptions of AI depending on experience, management buy in
- And The Ugly - the real world is, it is not all AI, really need to have a systems and software engineering approach and there is beauty in the ugly

AT&T

Shameless Plug

- Check out my blog, the Software Universe, http://vesonder.typepad.com/universe/
- Or go through my very uncreative homepage: http://homepage.mac.com/vesonder

AT&T

References

- Tamraparni Dasu, Gregg Vesonder, Jon R. Wright: Data quality through knowledge engineering. KDD-2003: 705-710
- Jon R. Wright, Elia Weixelbaum, Gregg T. Vesonder, Karen Brown, Stephen Palmer, Jay Berman, Harry Moore: A Knowledge-Based Configurator that Supports Sales, Engineering, and Manufacturing at AT&T Network Systems. AI Magazine 14(3): 69-80 (1993)
- Tamraparni Dasu, Theodore Johnson: Exploratory Data Mining and Data Cleaning. John Wiley 2003.