Oleksandr Romanko, Ph.D.
Senior Research Analyst, Risk Analytics, Watson Financial Services, IBM Canada
Adjunct Professor, University of Toronto

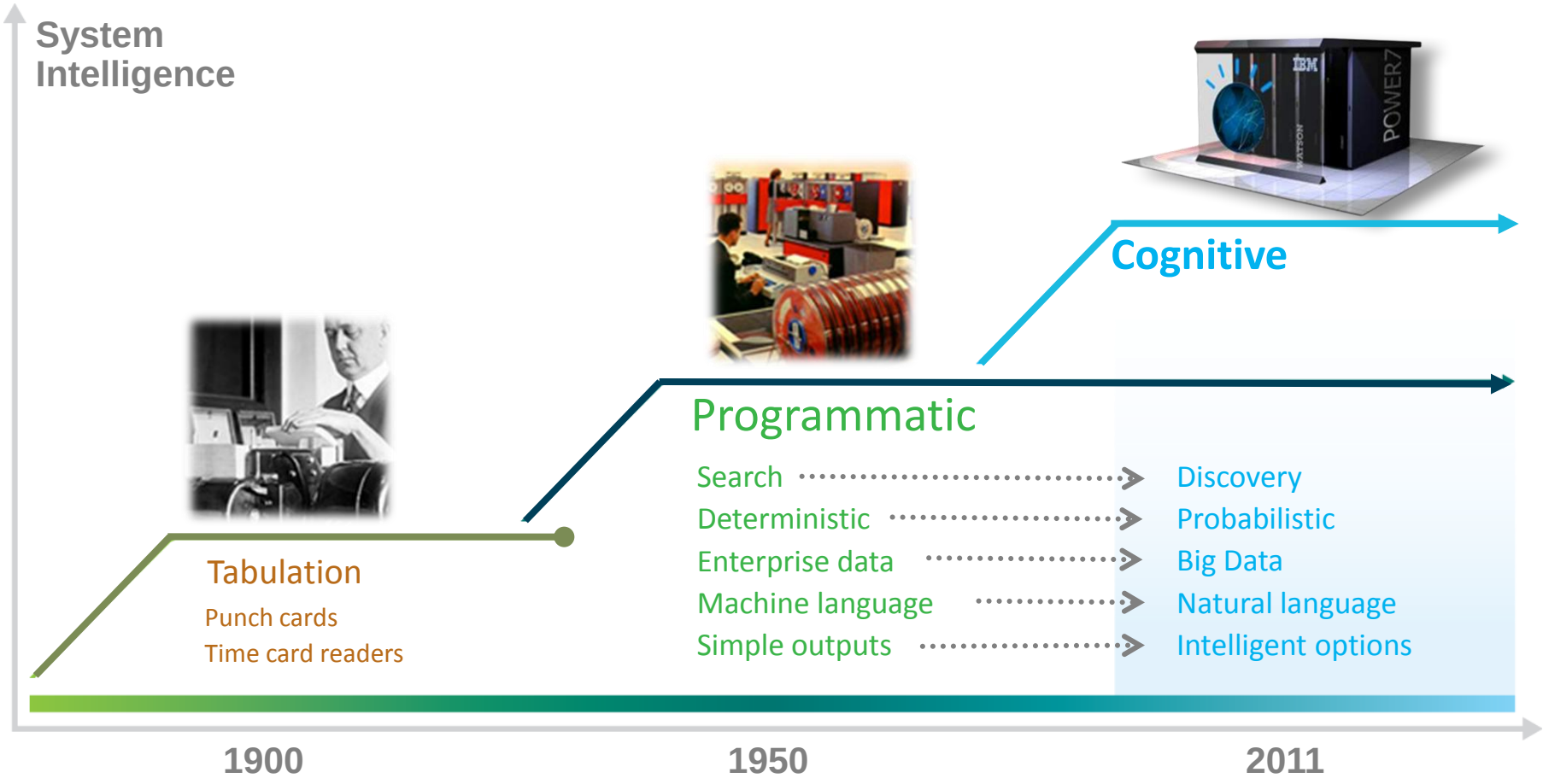
MIE1624H – Introduction to Data Science and Analytics

Lecture 10 – Cognitive Computing and Artificial Intelligence

University of Toronto
December 4, 2018

Era of Cognitive Computing

A new era of computing …



. . .enabling new opportunities on an optimized architecture.

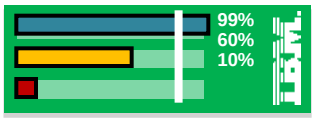
On February 14, 2011, IBM Watson made history

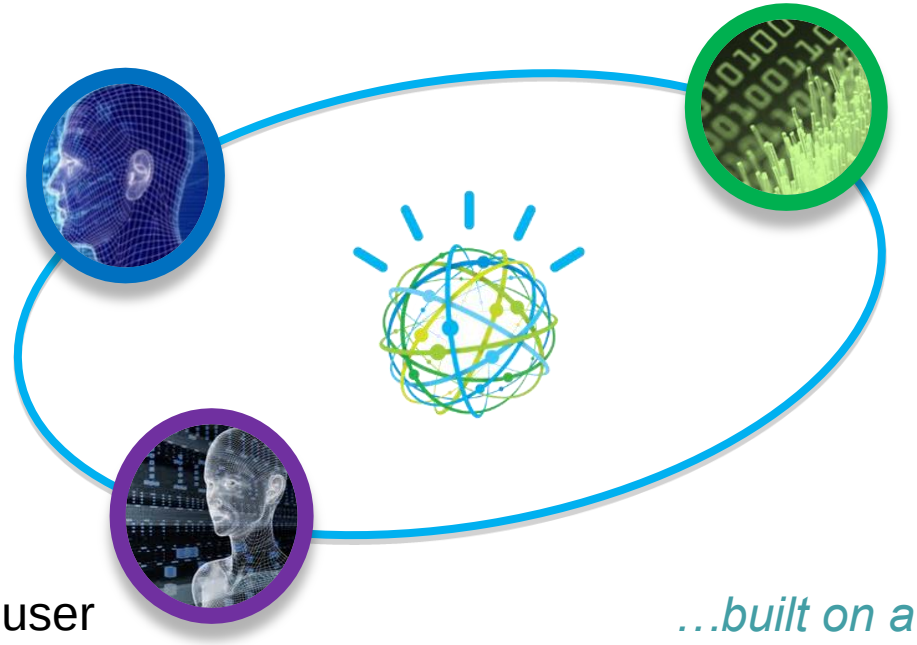
Since then Watson has been put to work transforming how organizations think, act and operate

IBM Watson brings together transformational technologies to drive optimized outcomes

1 Understands natural language and human speech

2 Generates and evaluates hypothesis for better outcomes

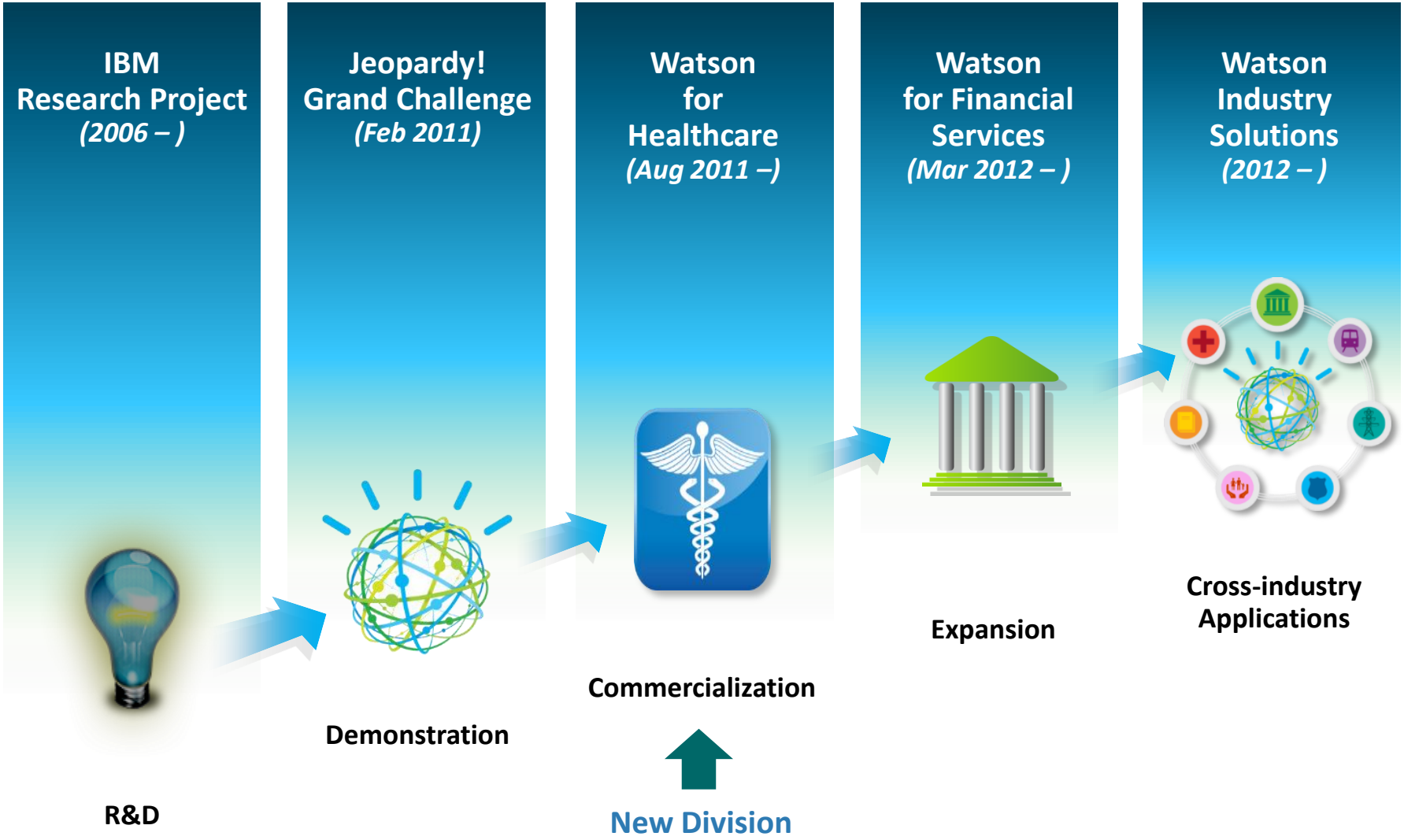




3 Adapts and Learns from user selections and responses

…built on a massively parallel probabilistic evidence-based architecture optimized for POWER

Brief history of IBM Watson



A look behind the scenes at the ultimate in compute intensive architecture

System Specifications

- 2880 Processing Cores
- 90 IBM Power Systems 750 Servers
- 16 Terabytes Memory (RAM) – 20TB Disk
- 80 Teraflops (80 trillion operations per second)
- Workload Optimized Systems

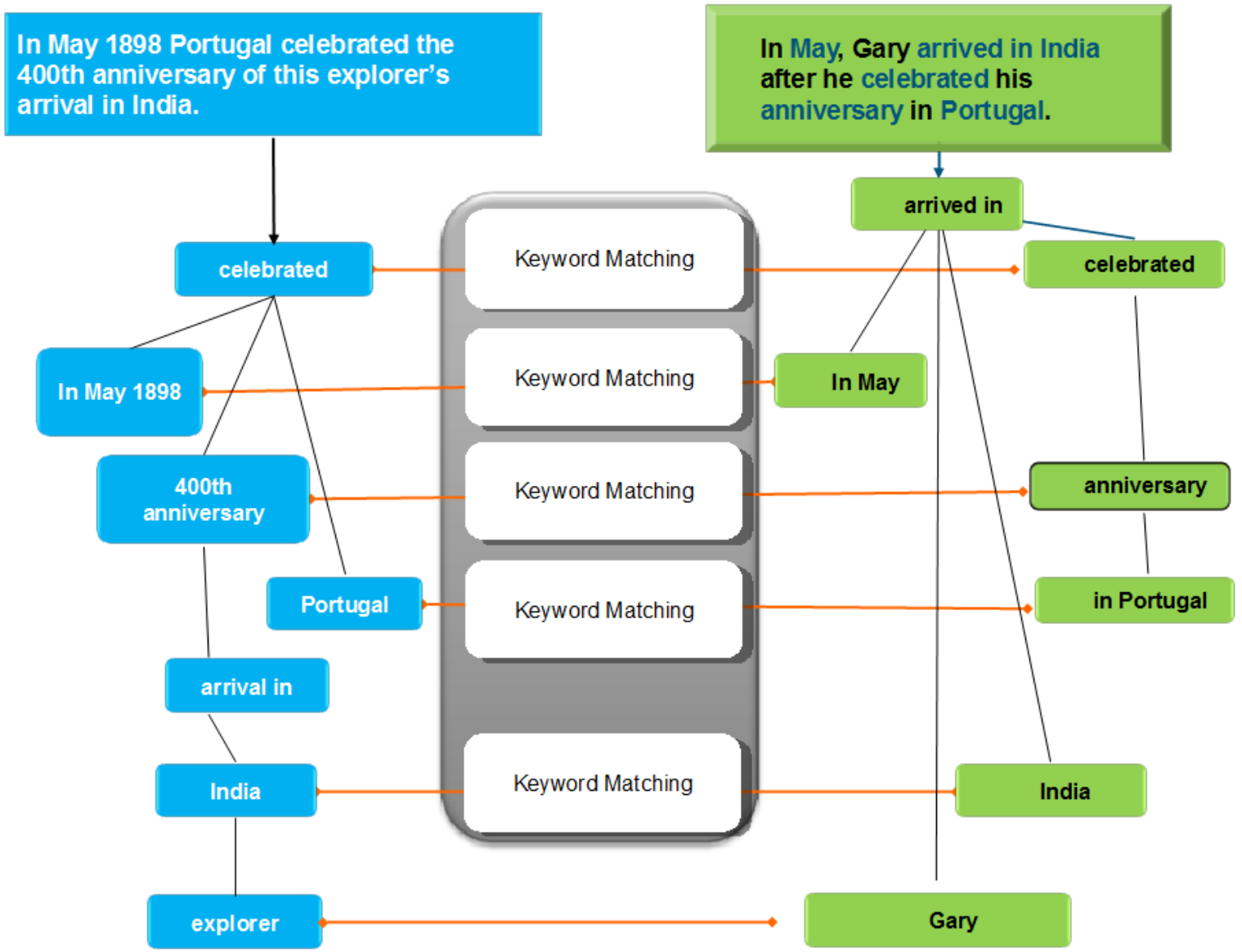
IBM Technology Depth

- Content Analytics
- Business Analytics
- Big Data
- Databases & Data Warehouses

In the past 5 years IBM has spent over $14B in analytical acquisitions and $6B in R&D annually

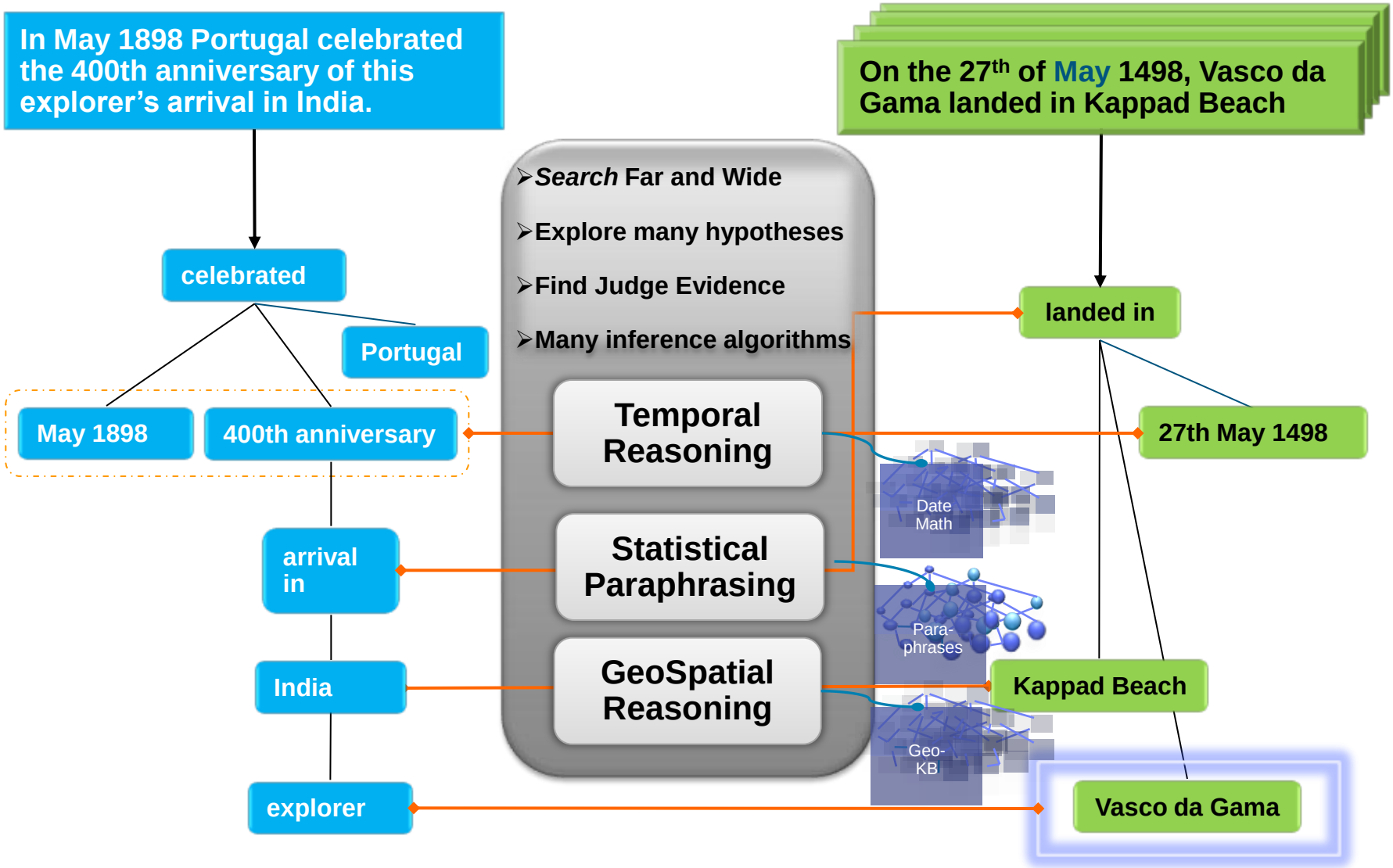
Artificial Intelligence

Artificial intelligence



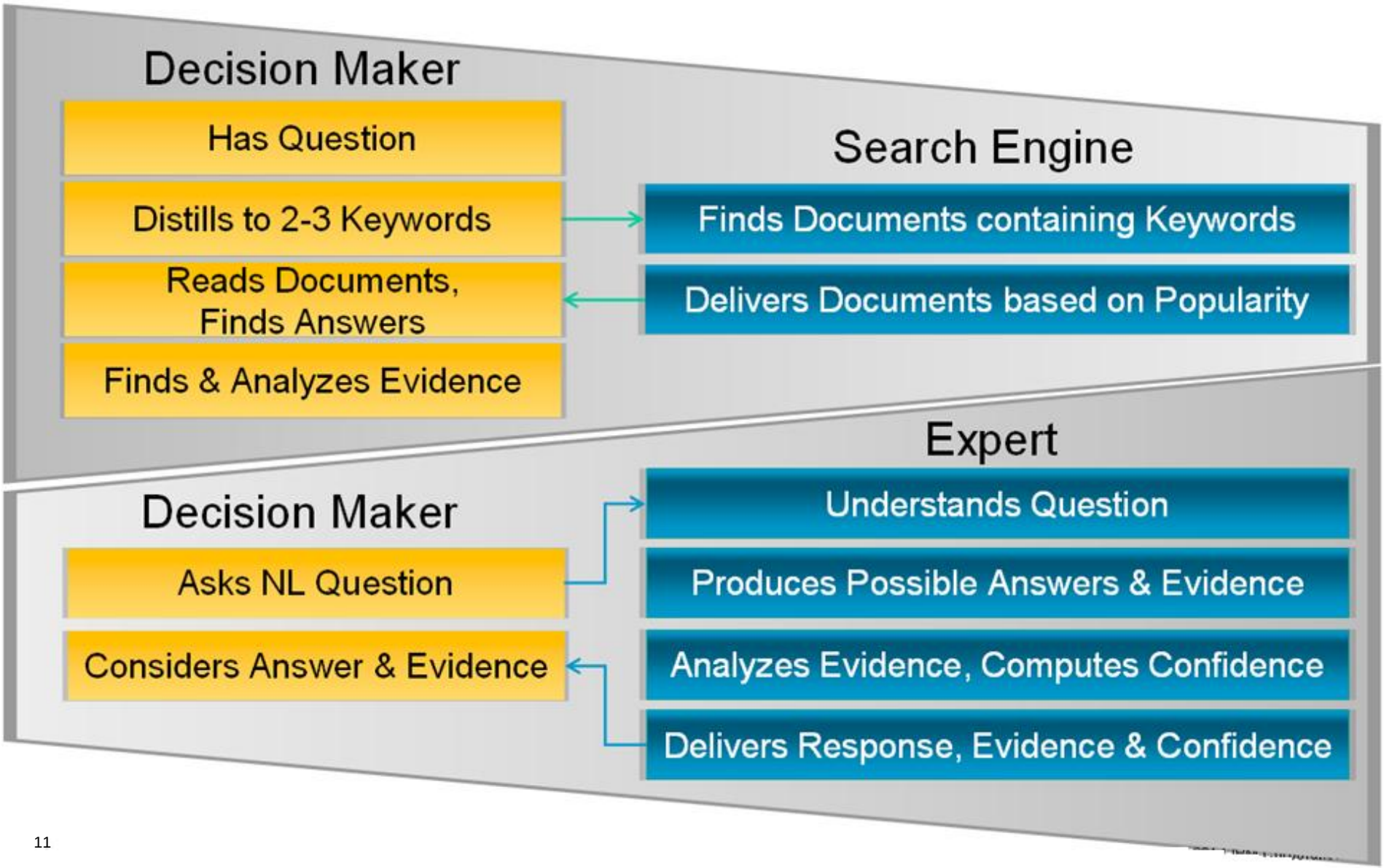
Source: A Brief Overview and Thoughts for Healthcare Education and Performance Improvement by Watson Team

Artificial intelligence



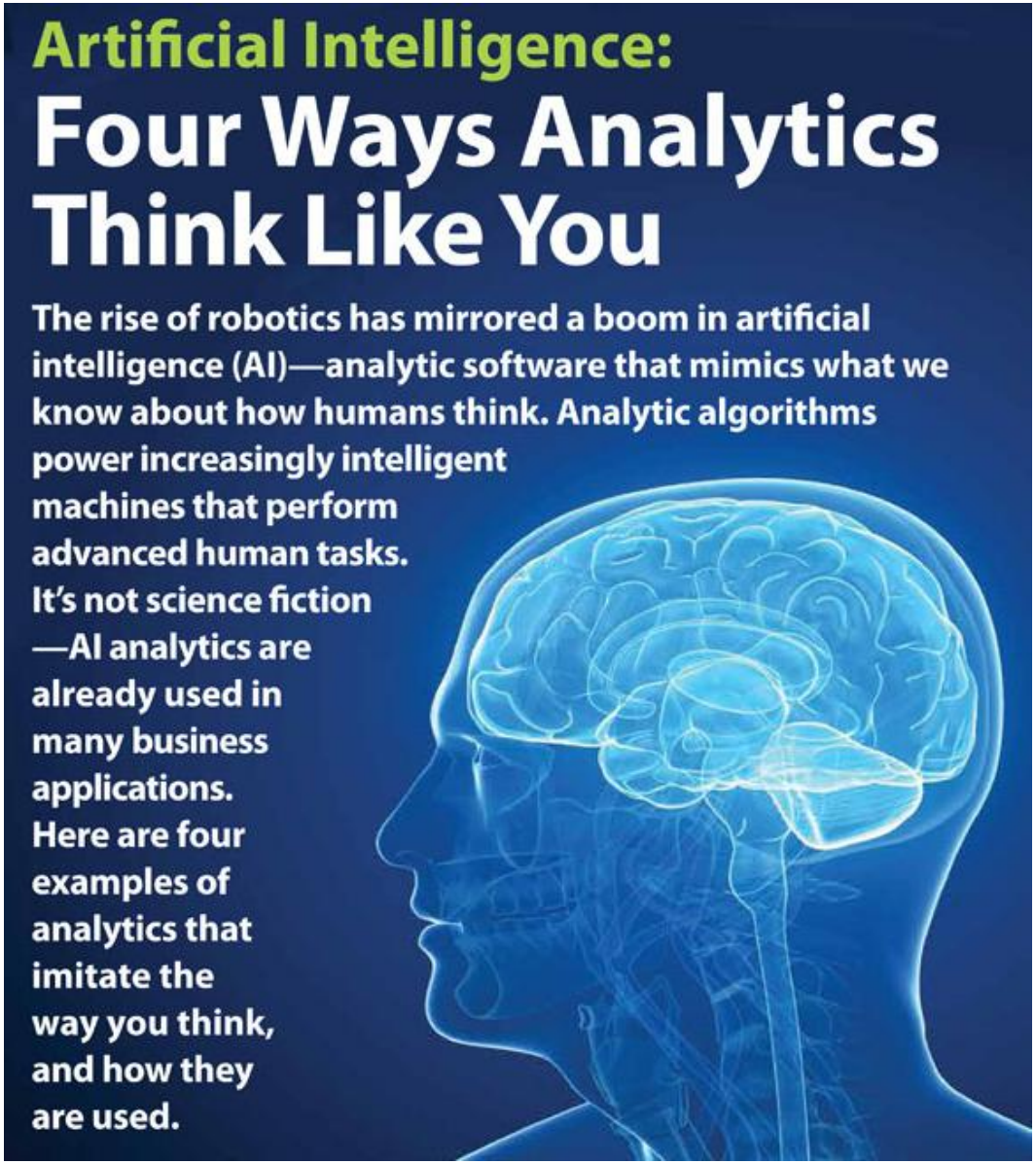
Source: A Brief Overview and Thoughts for Healthcare Education and Performance Improvement by Watson Team

Informed decision making: search vs. expert Q&A



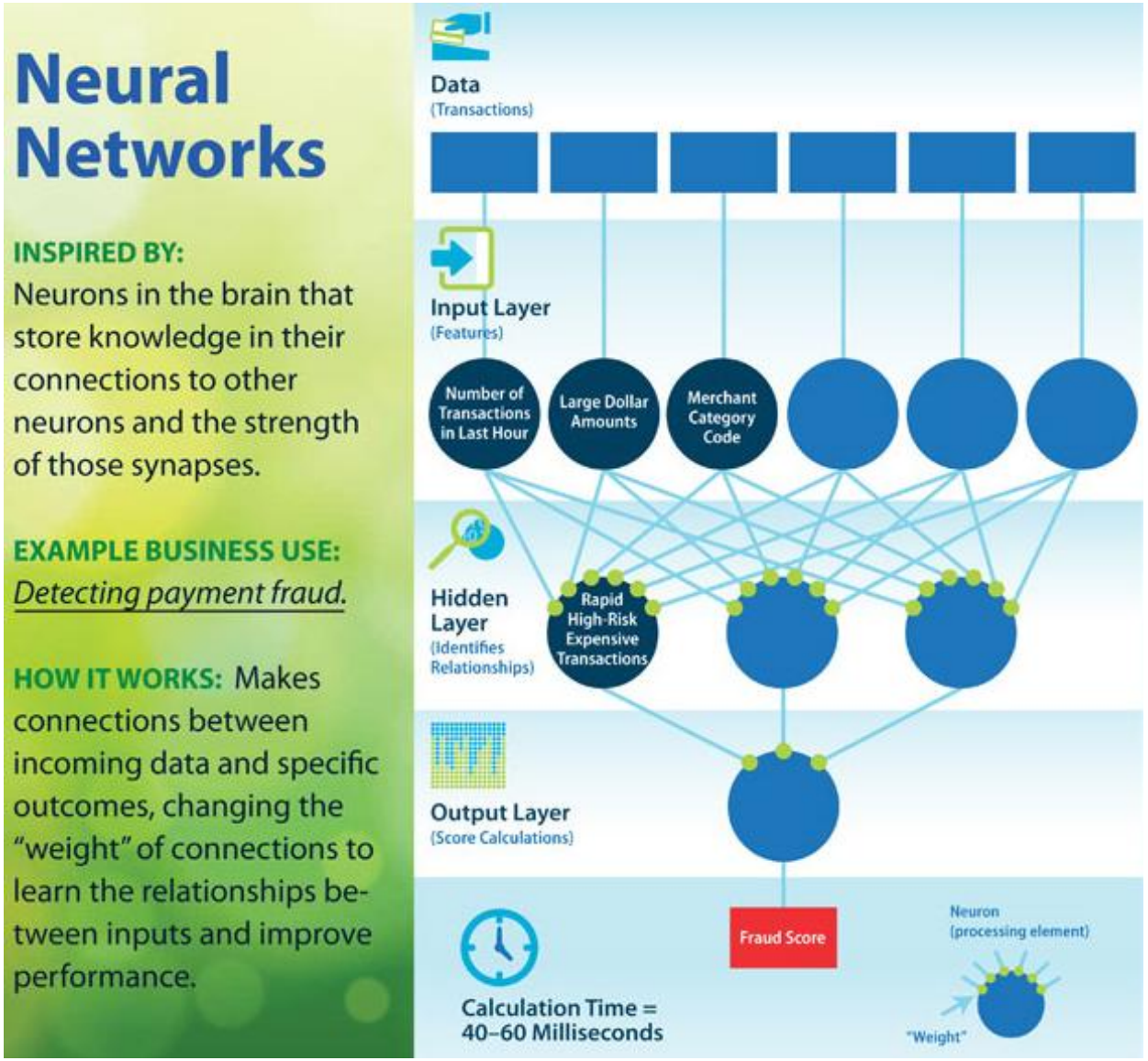
AI Algorithms

Artificial intelligence and analytics



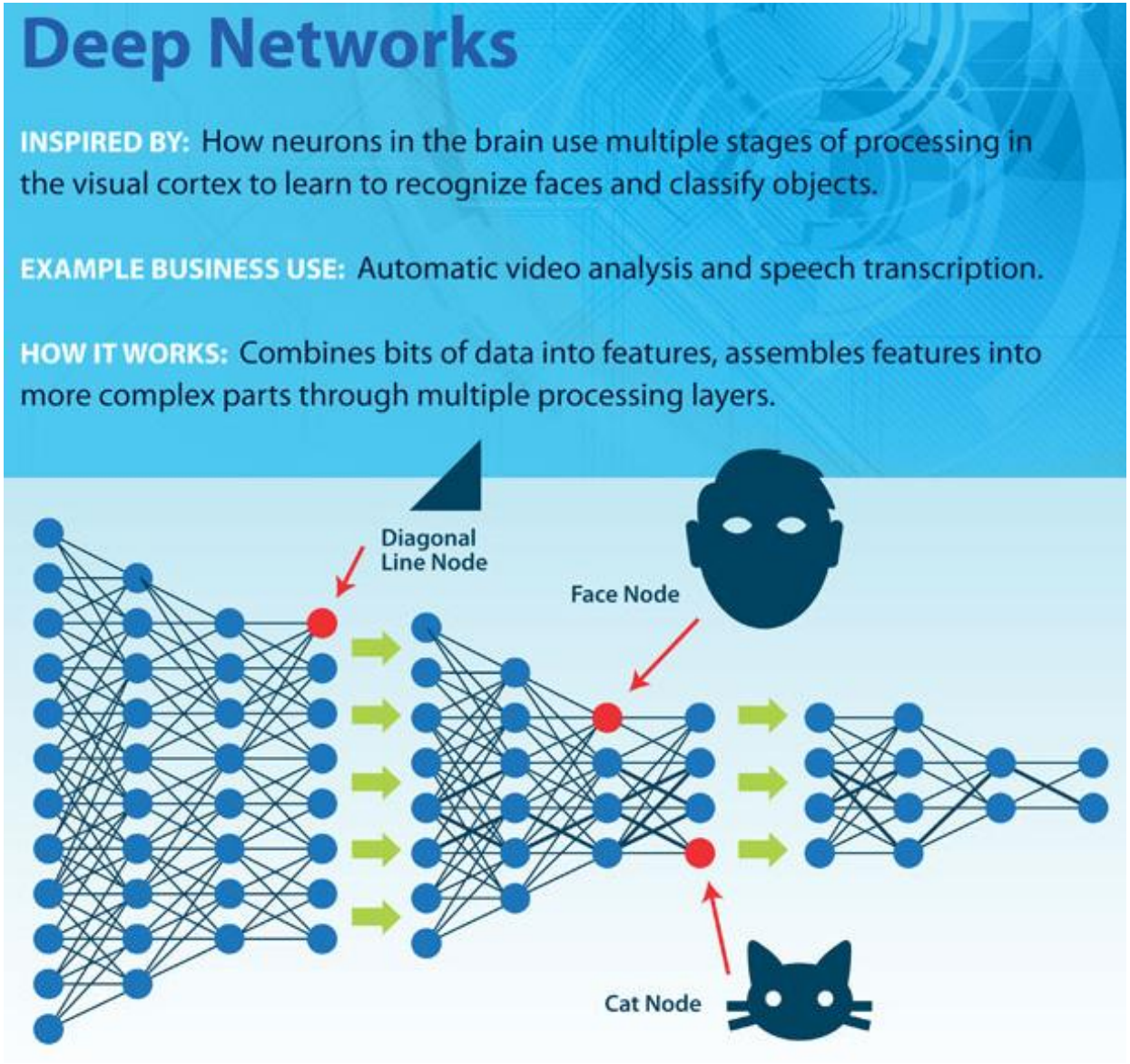
Source: Fair Isaac Corporation http://www.fico.com/infographics

Neural networks



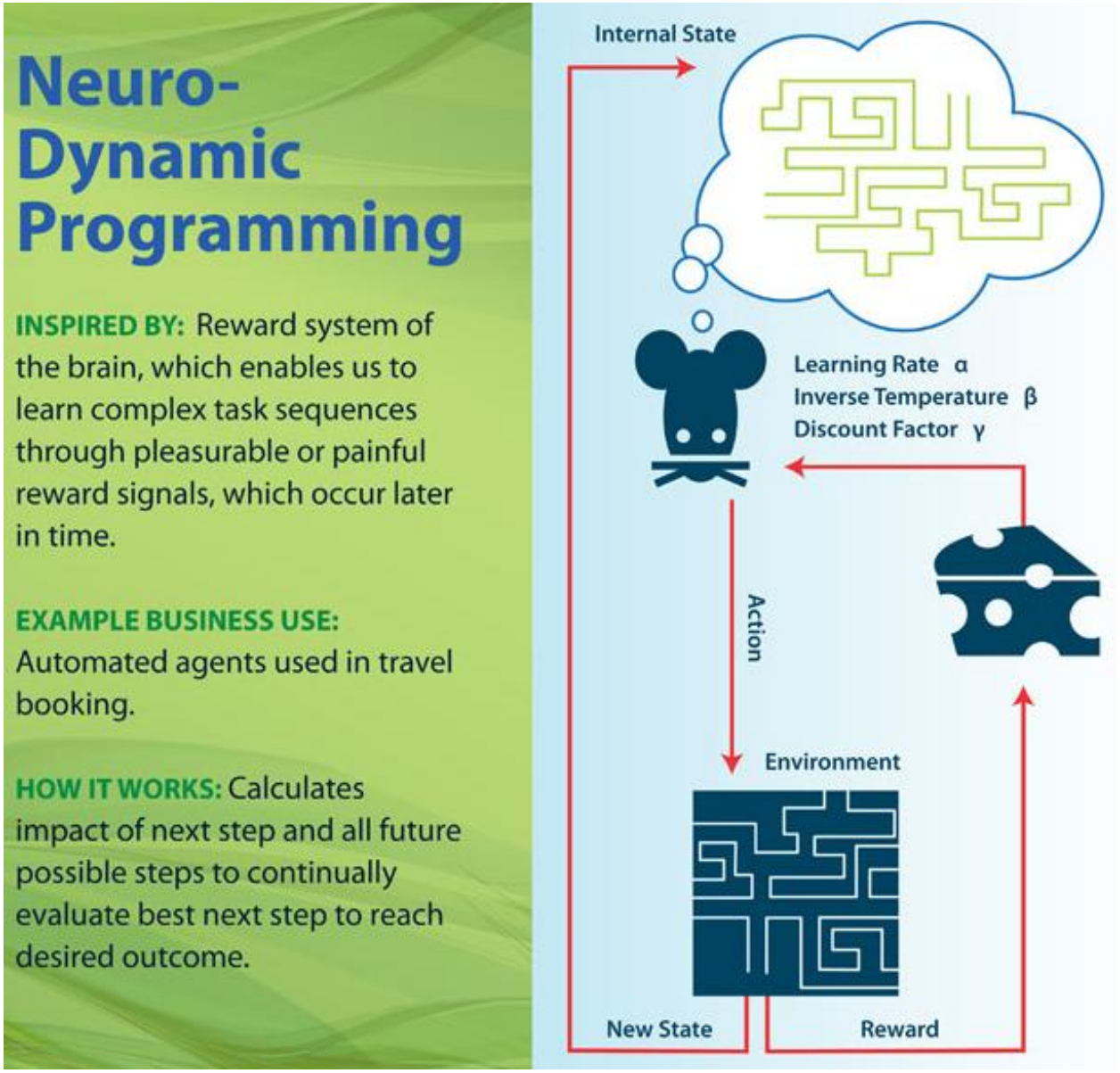
Source: Fair Isaac Corporation http://www.fico.com/infographics

Deep networks



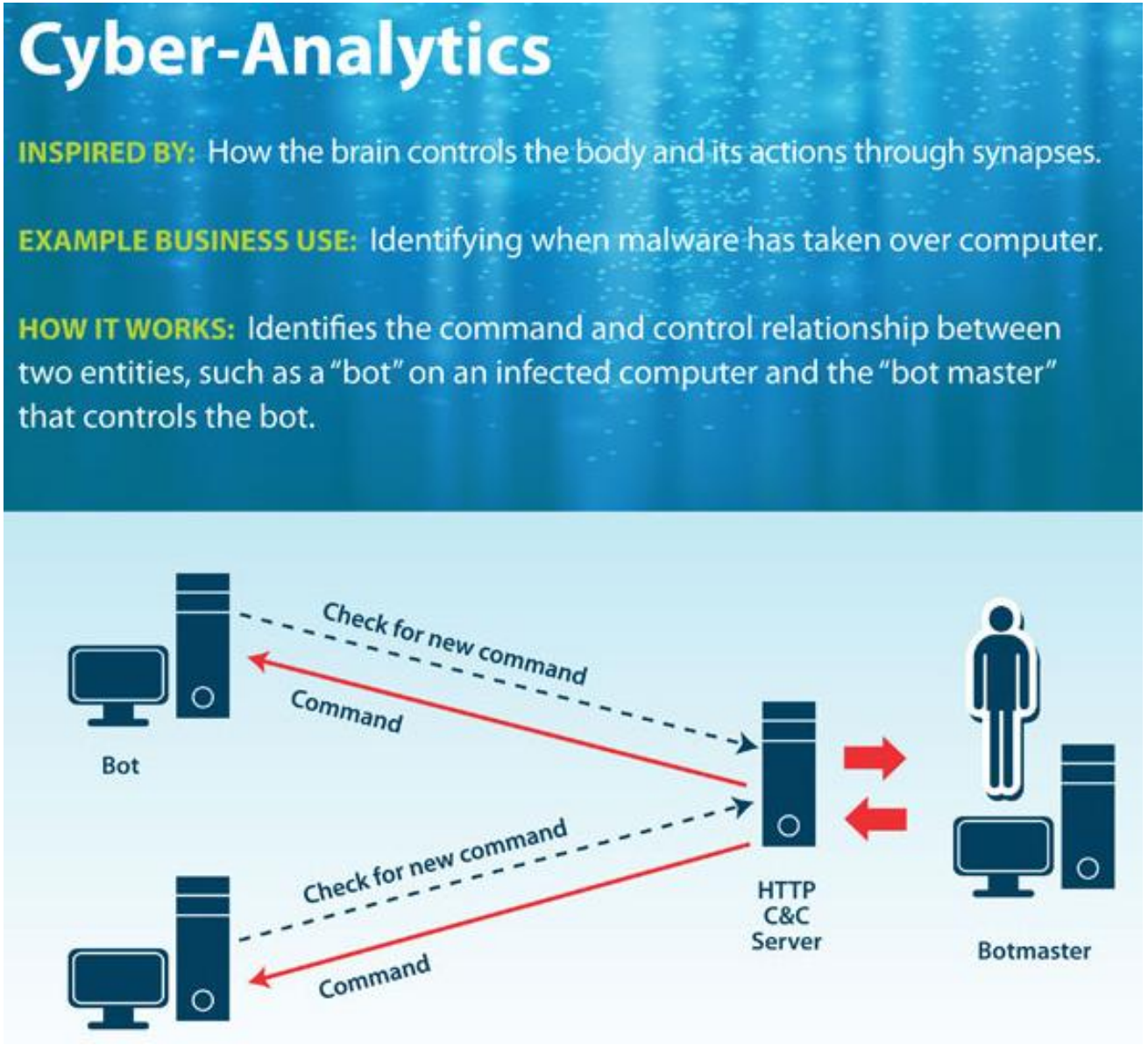
Source: Fair Isaac Corporation http://www.fico.com/infographics

Neuro-dynamic programming



Source: Fair Isaac Corporation http://www.fico.com/infographics

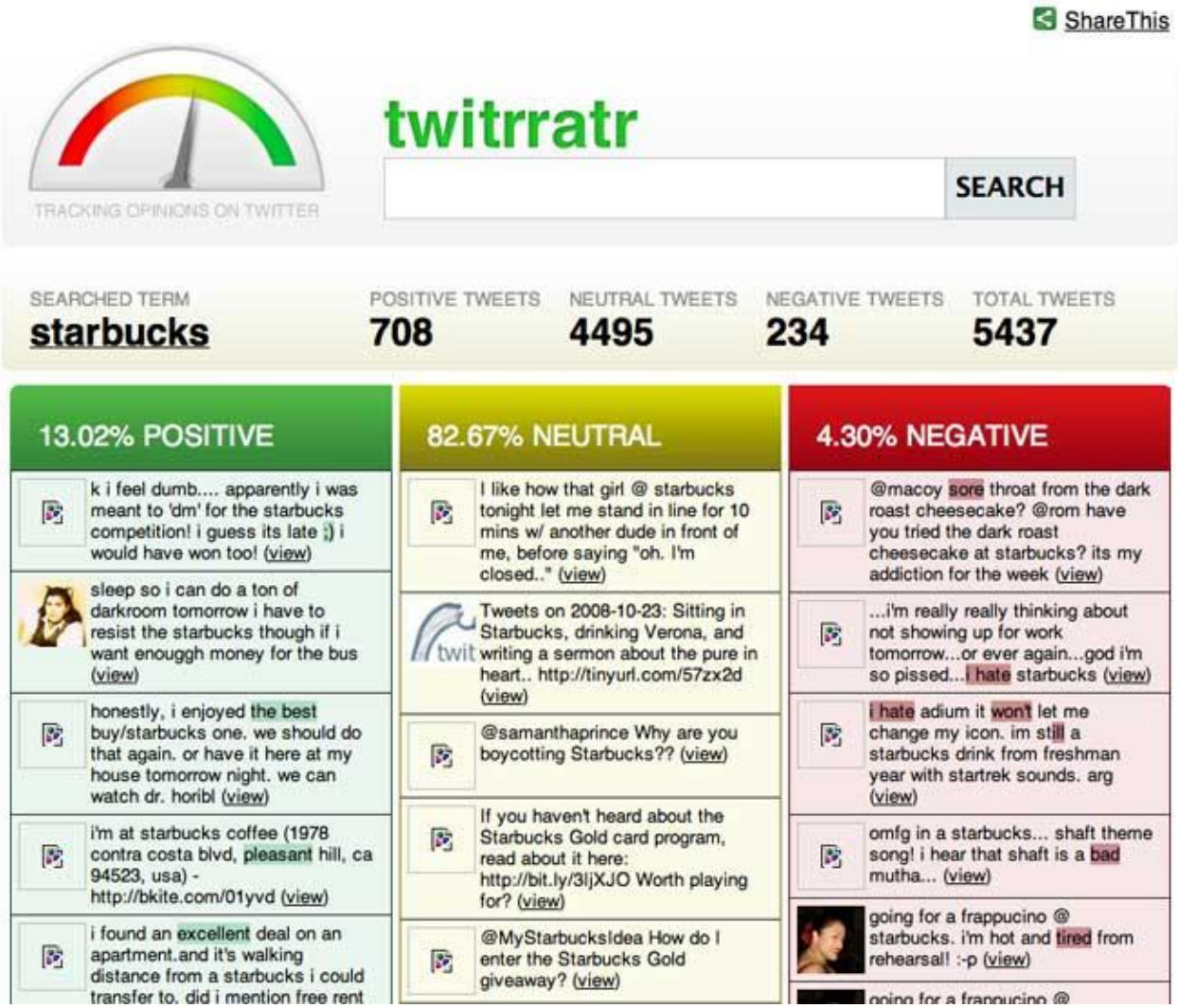
Cyber-analytics



Source: Fair Isaac Corporation http://www.fico.com/infographics

📖 Text Analytics

Text analytics and sentiment analysis



Sentiment analysis of tweets

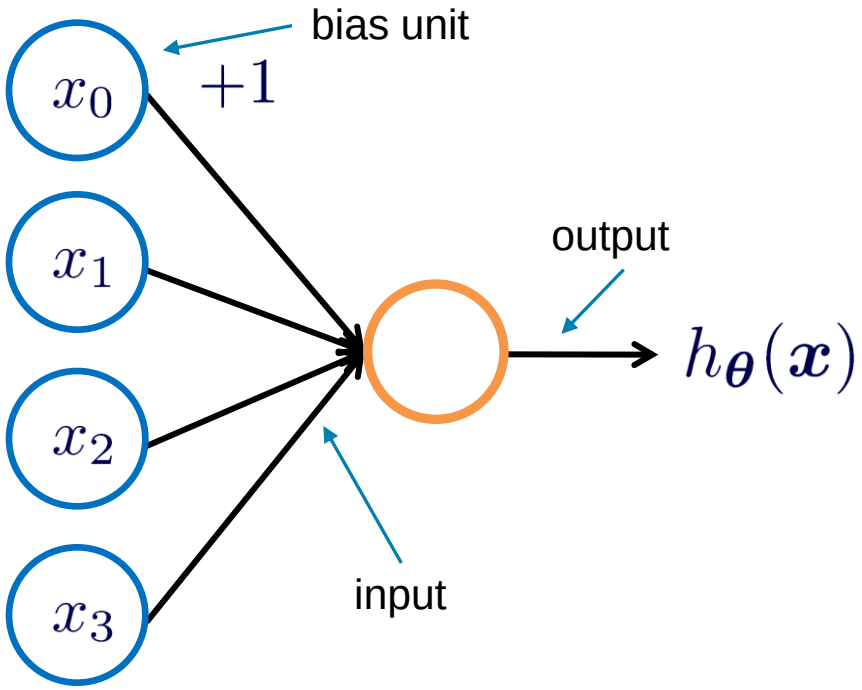
Artificial Neural Networks

Neuron model – logistic unit

- **Logistic regression** learns **parameter** vector $\boldsymbol{\theta}$:
- On **input** $\boldsymbol{x}$, it **outputs**:

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\boldsymbol{\theta}^T \boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}}}$$

- **Sigmoid (logistic) activation function** $g(z) = \frac{1}{1 + e^{-z}}$



$$x_0 = 1$$

$$\boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

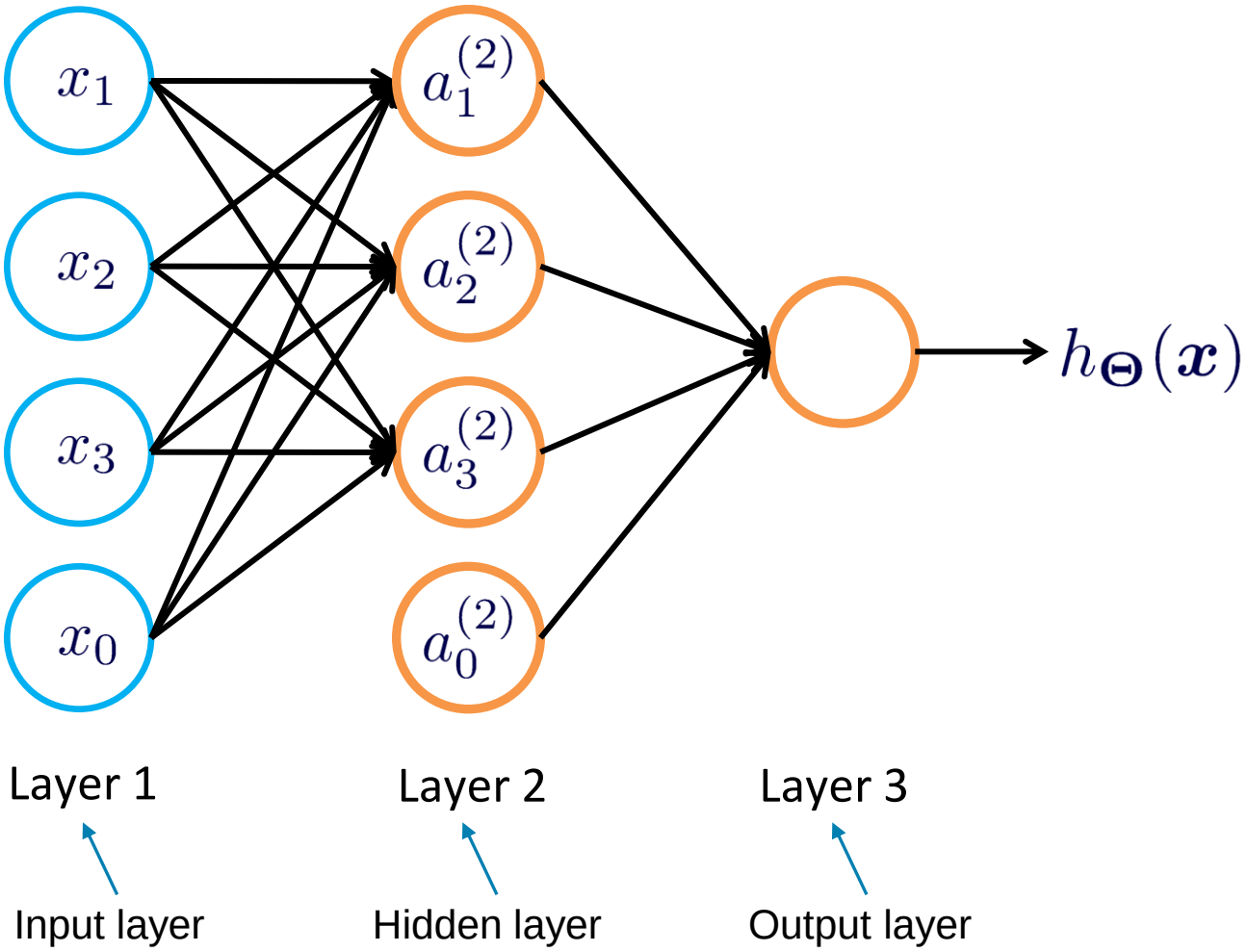
features

weights (parameters)

Source: Andrew Ng “Machine Learning”

Neural network

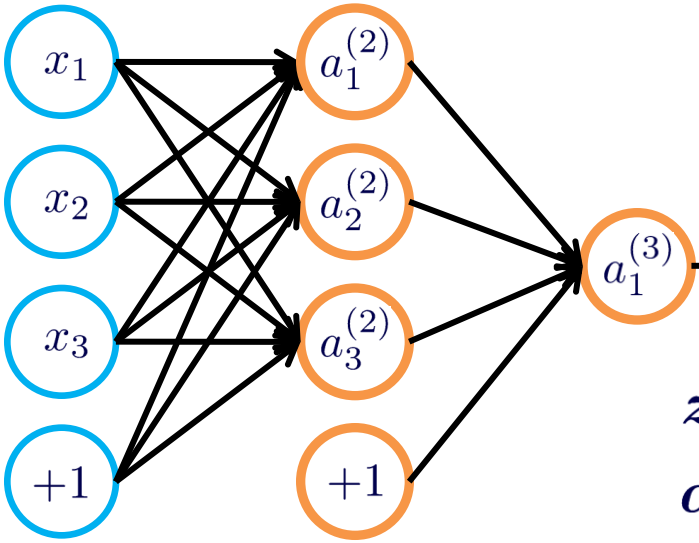
- Put **a number** of **logistic units** together to get a 3-layer network:



Source: Andrew Ng “Machine Learning”

Neural network – forward propagation

- Put **a number** of **logistic units** together to get a 3-layer network:



$$\boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \boldsymbol{z}^{(2)} = \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{pmatrix}$$

$$\boldsymbol{z}^{(2)} = \Theta^{(1)}\boldsymbol{a}^{(1)} \qquad \boldsymbol{z}^{(3)} = \Theta^{(2)}\boldsymbol{a}^{(2)}$$

$$\boldsymbol{a}^{(2)} = g(\boldsymbol{z}^{(2)}) \qquad h_\Theta(\boldsymbol{x}) = \boldsymbol{a}^{(3)} = g(\boldsymbol{z}^{(3)})$$

$\boldsymbol{a}^{(1)} = \boldsymbol{x}$ $\quad a_0^{(2)} = 1$

$$a_1^{(2)} = g(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3)$$

$$a_2^{(2)} = g(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3)$$

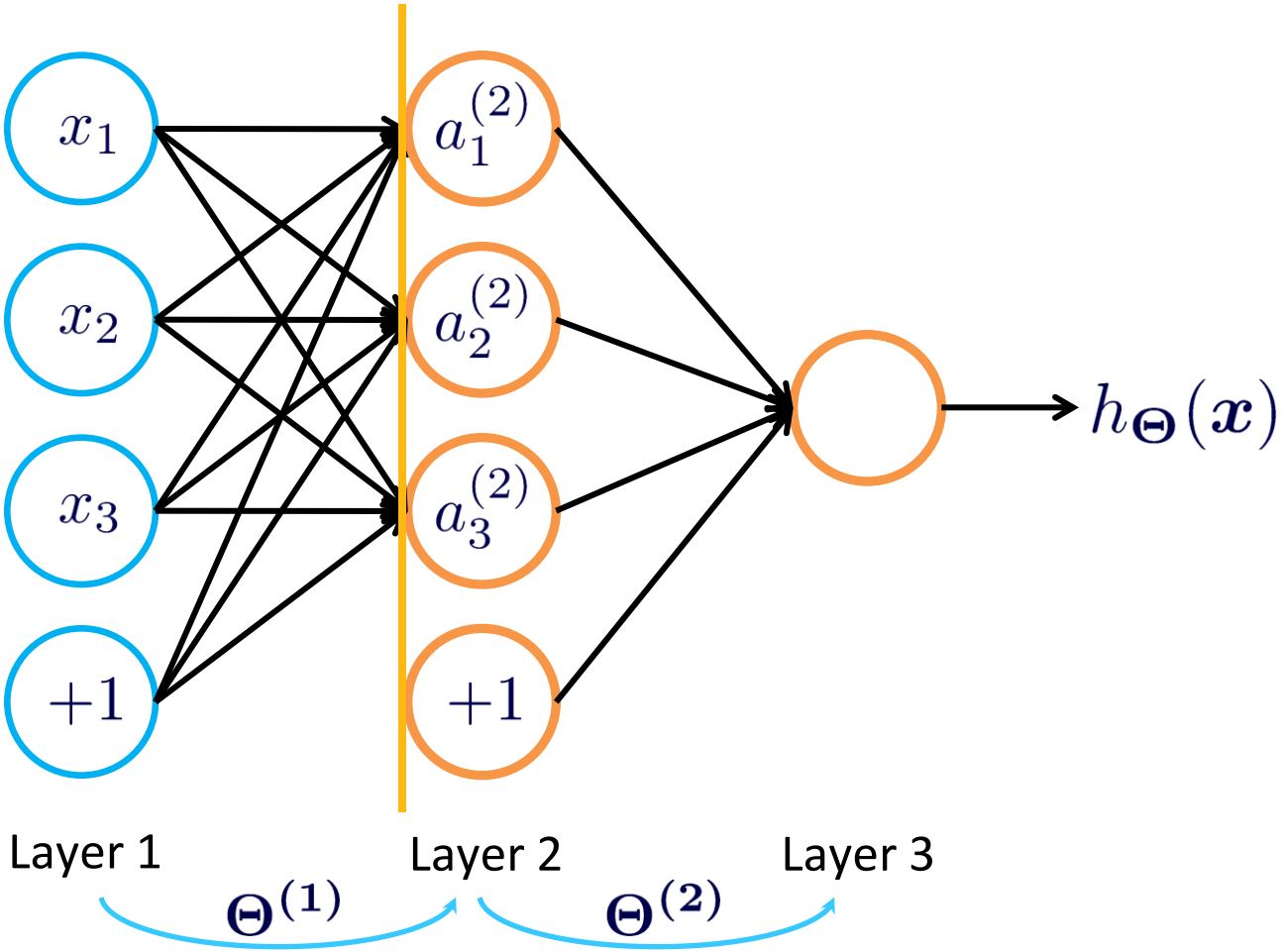
$$a_3^{(2)} = g(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3)$$

$$h_\theta(\boldsymbol{x}) = g(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)})$$

Source: Andrew Ng “Machine Learning”

Neural network learning its features

- Put **a number** of **logistic units** together to get a 3-layer network:

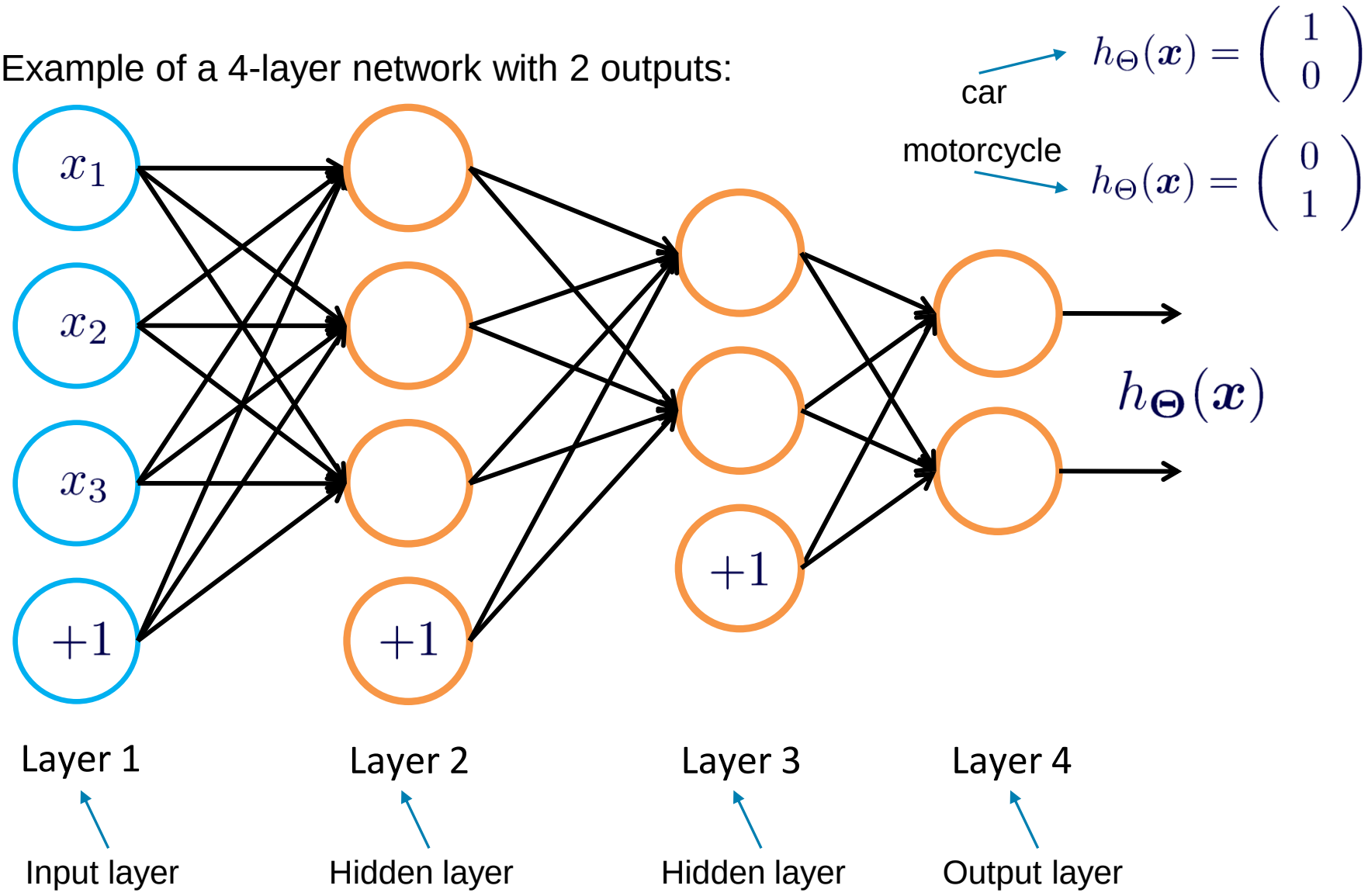


$$h_\theta(x) = g(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)})$$

Source: Andrew Ng “Machine Learning”

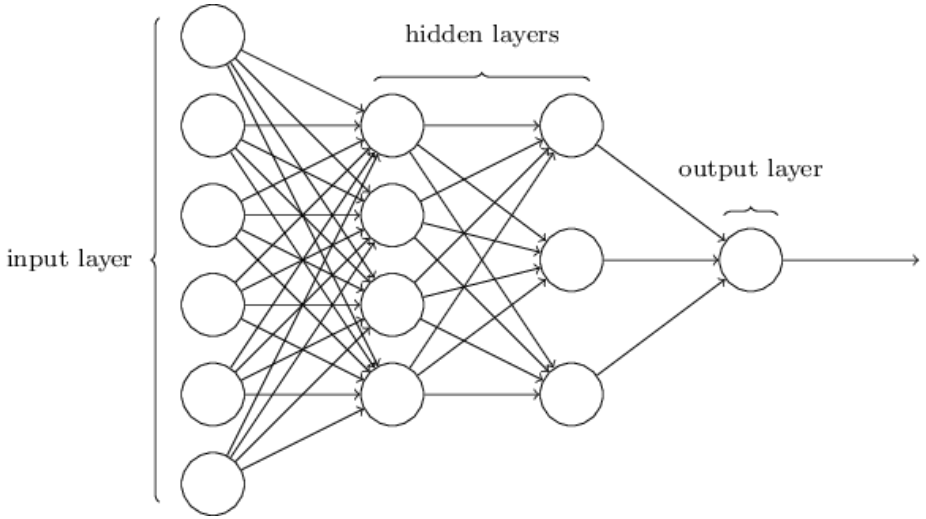
Neural network architectures

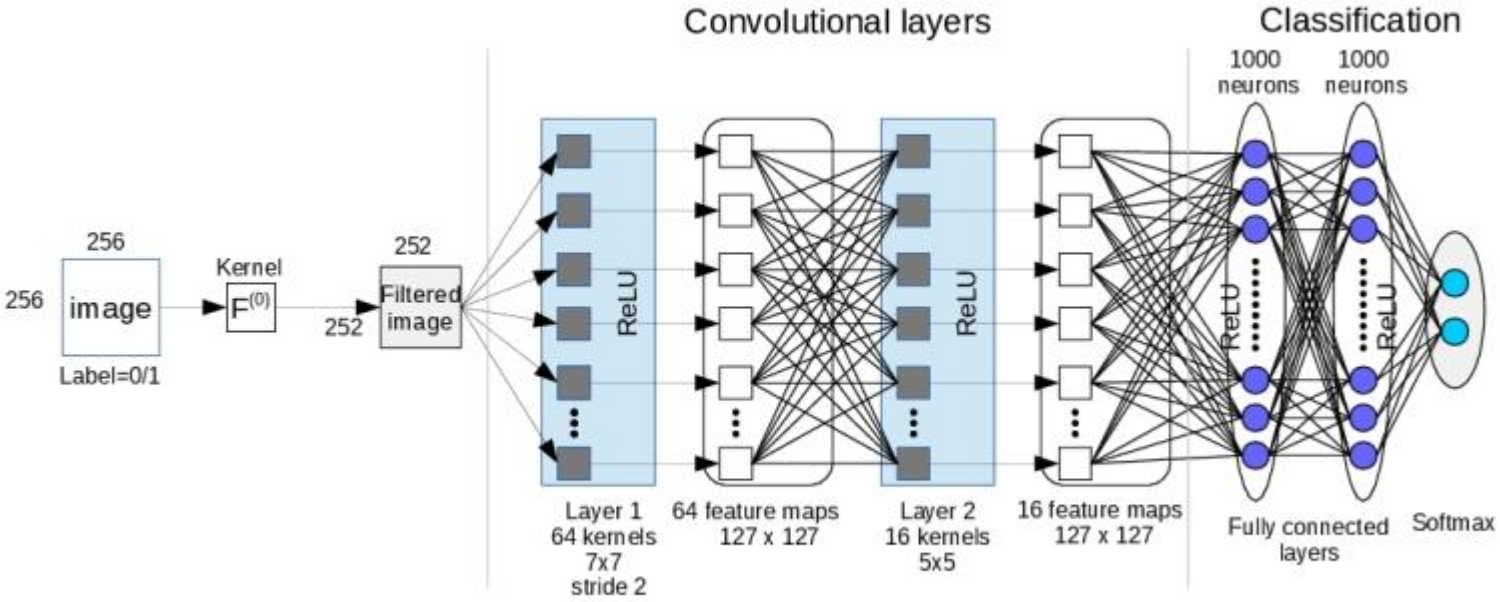
- Example of a 4-layer network with 2 outputs:



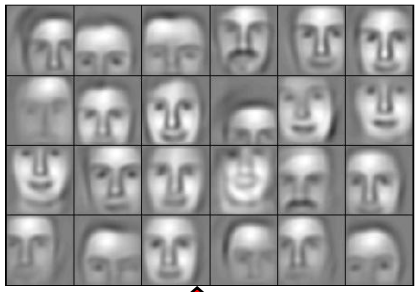
Source: Andrew Ng “Machine Learning”

Simple neural network vs. deep neural network

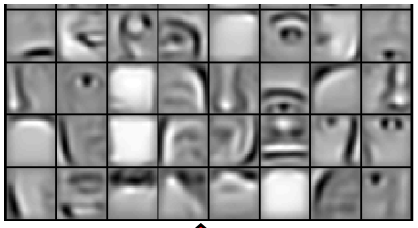




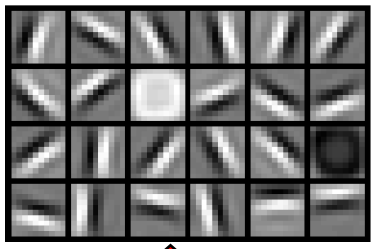
Convolutional neural net on face images



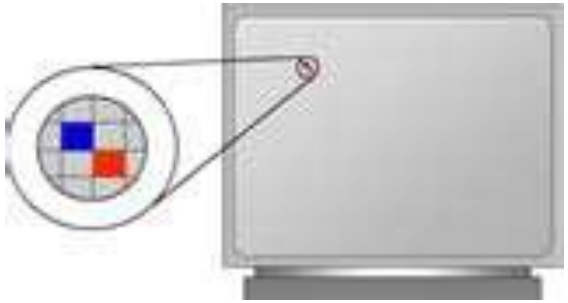
3rd layer
object models



2nd layer
object parts
(combination
of edges)



1st layer
edges



pixels

In vision, each layer of the neural network is mimicking how humans process images

Source: Andrew Ng “Machine Learning”

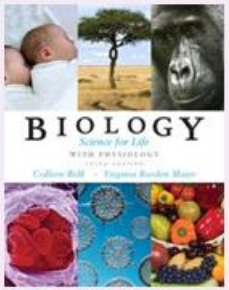
Applications

Opportunities for Watson in Adjacent Spaces

IBM WATSON

Healthcare and Life sciences

Drug Discovery



New Biology Science

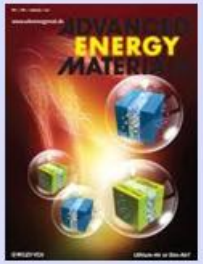
New Bio-medical Research

- ***Drug discovery: ~12-15 yrs, $B per drug, 90+% fallout rate***

Chemical and Petroleum

Oil Reservoir Discovery

Crop Sciences



New Energy Materials

- ***Lithium ion Battery: ~20 years development time***

Existing Discovery is Slow, Expensive, Ad hoc and Manual

Consumer Goods and Products

Product Innovation

New Market Identification

New Partnerships

- ***Product formation: based on Ad hoc manual trial & error***

Semi-Conductor and Materials

Nano Materials

Energy Storage

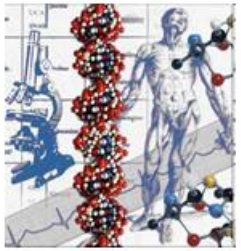
Water Filtration

- ***Water filtration: Billions still do not have clean water today***

Drug Development Processes

IBM WATSON

What are a set of targets for a given disease area	*Which compounds are good for which targets?*	*What's the safety and toxicity profile for a given compound?*	*What kinds of patient profiles are good for a clinical trial?*
Which type of pathways should we consider	*What similar compounds can we find to have activities against a given target*	*What other compounds might have similar tox and safety profiles*	*How do we design clinical trials for the patient population?*



Key Analytics Capabilities: BI, Text analytics, Network Analysis, Relationship Discovery, ML, Modeling, …

BigData / BigInsight Infrastructure

Scientific Literature and Patents | Drug and disease related ontologies | Pre-clinical Clinical Trials | Electronic Medical Records | Lab and Imaging data | Genomic Data | Claims Data | Web Social Media

📖 Spatio-Temporal Analytics

Spatio-temporal analytics

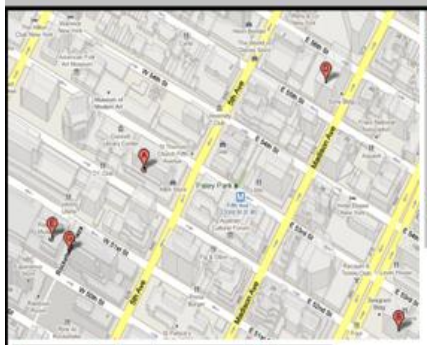
Spatio-Temporal Analysis Helps Water Management

User Configured Alert
- If there are more than 3 brown water complaints within a 1 mile radius of the capitol within a 3 hour time frame – alert – potential security situation

Automatic Spatial Alert
- Defects seem to be spatially clustered and occurring within short time window in this part of the pipeline

Spatio-Temporal User Query
- Select all temperatures changes lager than 30 degrees within 24 hours – that occurred "near defect location –both spatially / temporally"

Spatio-Temporal Analysis Helps Building Energy Management



Spatio-temporal visualization
Display multiple map layers:
- District Bureau map layer
- Building map layer
  - Marker size reflects bldg size range
  - Marker colors reflects energy consumption density
- Pop-up window to show other properties of selected bldg

Automatic Spatio-temporal Alert
- Alert when a cluster of high energy density building within one mile in a very hot summer day – may cause power supply problem.

Spatio-Temporal User Query
- Select the top 10 high energy density buildings within one mile of the UN building.

Predictive insights

Spatial pattern recognition

Automatic Spatio-Temporal Alerts

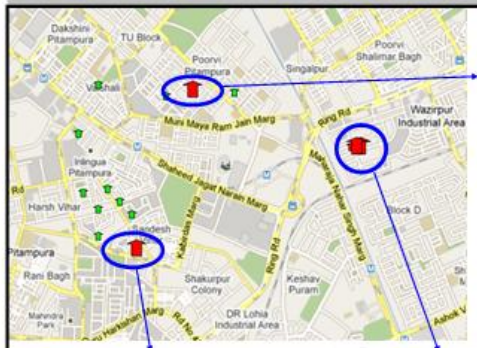
Use Cases in various IBM projects

User-defined Spatio-temporal Alert

Scheduling Decision Support

Operational dashboard

Spatial Temporal Analysis Helps Liquefied Petroleum Gas Cylinder Tracking



Automatic Spatial Alert
- This LPC cylinder is delivered to a commercial area, which might indicate an illegal diversion of subsidized cylinders meant for domestics use for commercial use

Spatio-temporal query
- Select all cylinders that have not changed location or been read in the last N days – possible container lost.

User Configured Alert
- If more than one LPG cylinder delivered to the same address within 2 days - provide alert – it might indicate multiple accounts using the same address

Spatio-Temporal Analysis Helps Railroad Safety

User Configured Alert
- If there are more than 3 incidents occurring when two or more trains are within a mile of the level crossing – provide alert

Automatic Spatial Alert
- Multiple near miss incidents recorded near crossings with boom gates

Spatio-Temporal User Query
- Select all traffic jams reported within 3 hours of a rain storm near these crossings

Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

Spatio-temporal analytics applications in city space

- Spatio-temporal hotspot analysis to identify where are the hot areas for an event (e.g., traffic crash accidents, 311 calls, 911 calls, and etc)
- GPS analytics to help with patrol activity planning

Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

Spatio-temporal hotspot analysis

Problem: Detect the “hot” area where an event happened more frequently

DBSCAN: Density Based Spatial Clustering of Applications with Noise (Ester et al. 1996)

Extend DBSCAN with time dimension

Basic Idea: density reachability

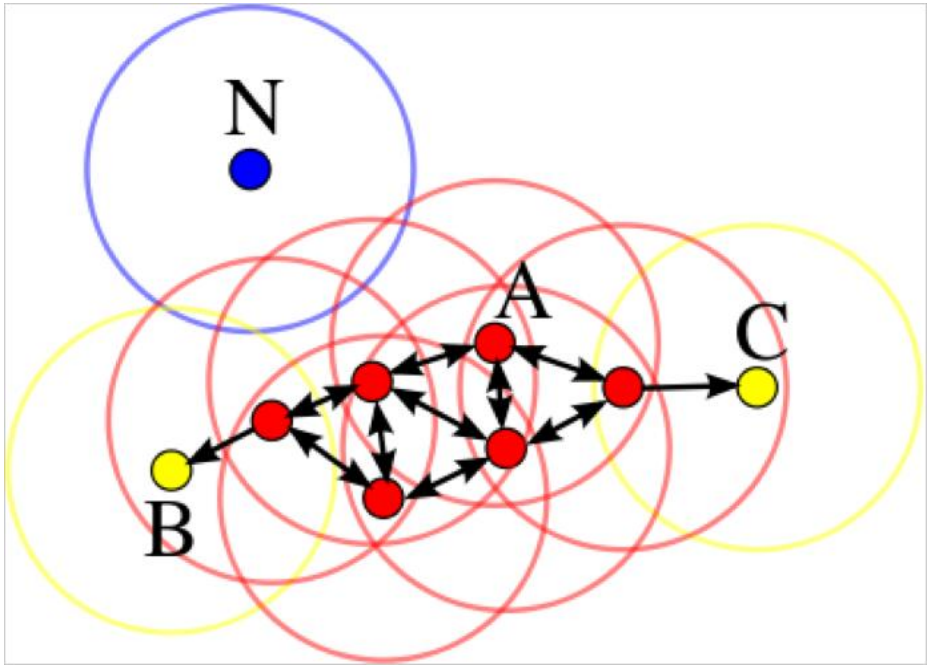
Point B is directly density-reachable from point A

If:

(1) distance between A and B
is smaller than fixed value Ɛ

(2) A is surrounded by
sufficiently many points (K)

(3) Max time between A and B is less
than a fixed value t



Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

About the parameters

Three parameters K , Ɛ & t:

(1) Parameter k: minimum number of points to form a cluster

<1> A smaller k would produce more small clusters – more small regions on the map

<2> A lager k would produce less large clusters – less larger regions on the map

<3> The selection of k is supposed to be guided by the total number of events

(2) Parameter Ɛ: maximum neighborhood radius

<1> The value of Ɛ is “proportional” to the average size of the clusters

<2> A large Ɛ encourages to cover a larger portion of events on map

<3> The selection of Ɛ is guided by the average distance between the locations

(3) Parameter t: maximum time between events

Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

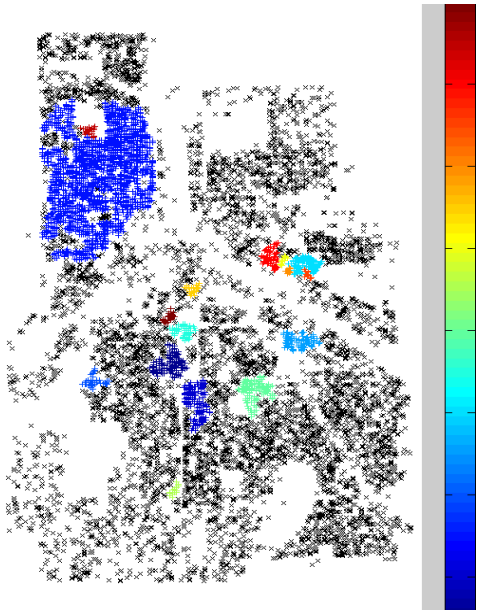
Sample hotspots for Burglary incidents

Case 1:

Few points required to form hotspot
(k=15 , Ɛ=0.001, t=1 day)

# of samples: 11820
Runtime: 3.4319s

Covered points: 31.0359%

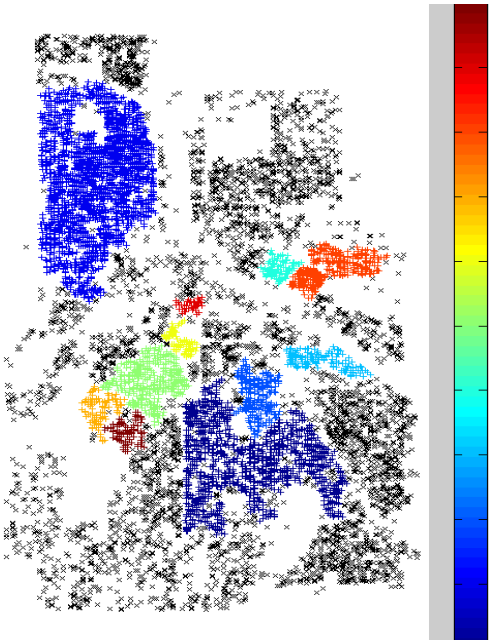


Case 2:

Many points required to form hotspot
(k=20, Ɛ=0.0013, t = 1day)

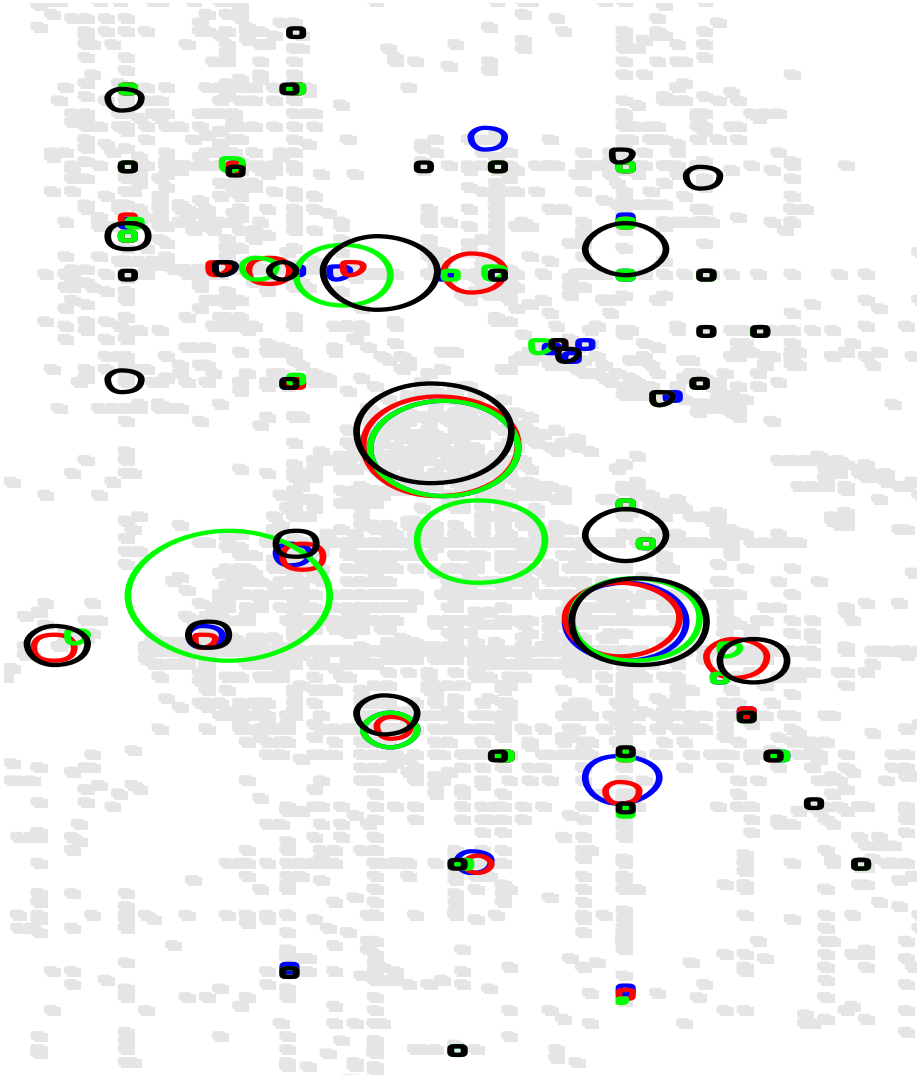
# of samples: 11820
Runtime: 3.5182s

Covered points: 54.3235%



Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

How hotspots change over the years?

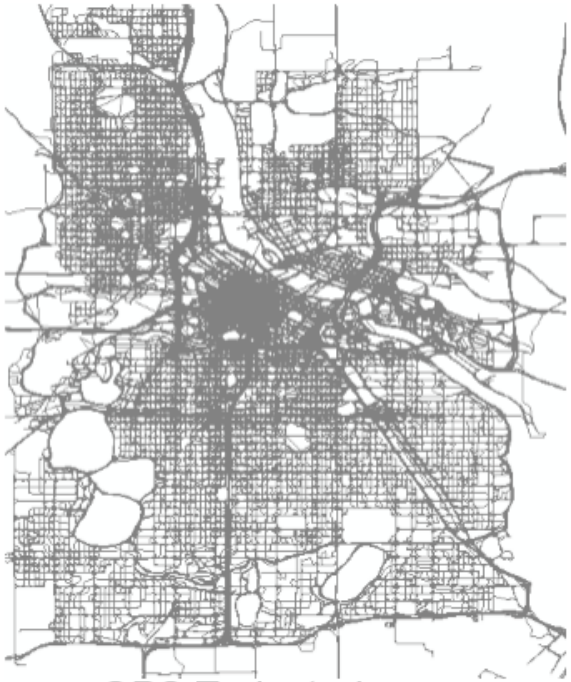


Blue = 2007
Red = 2008
Green = 2009
Black = 2010

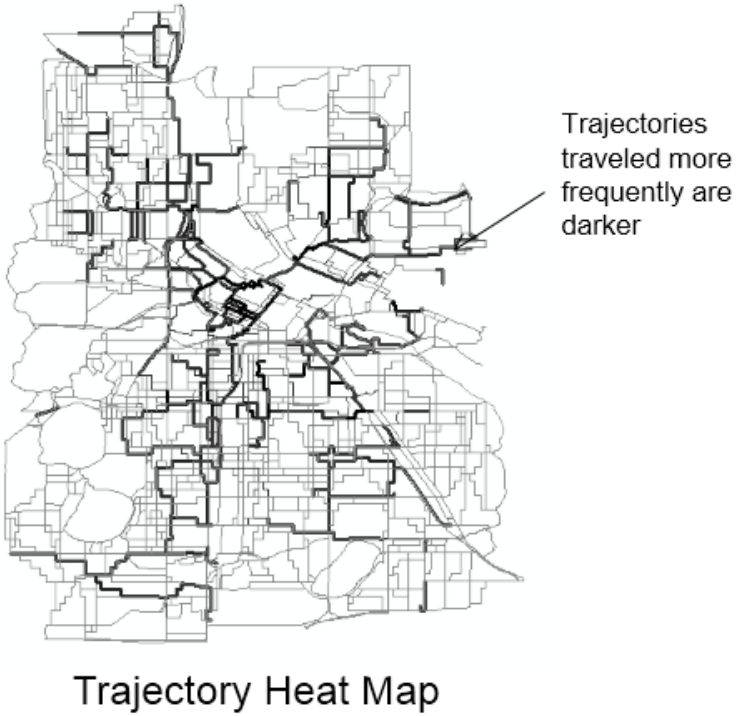
Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

GPS trajectory analysis

- Background: The increasing availability of GPS equipped devices and the ubiquitousness of location sensing devices has resulted in a tremendous amount of data about moving objects.
- Problem: Given GPS logs of participating vehicles, generate a **trajectory heat map** with temporal dimension, which captures the travel patterns, to enable online spatial temporal analytics



GPS Trajectories

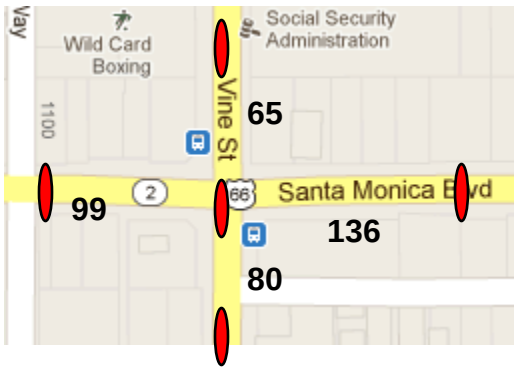


Trajectory Heat Map

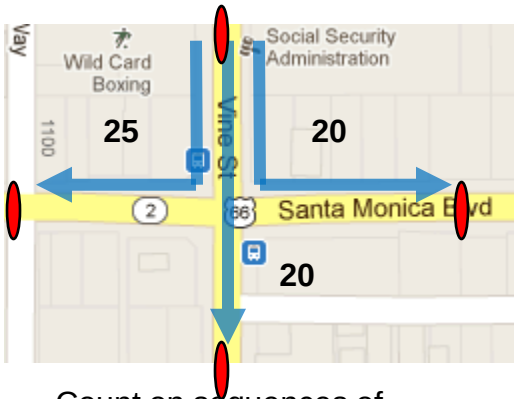
Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

GPS trajectory analysis

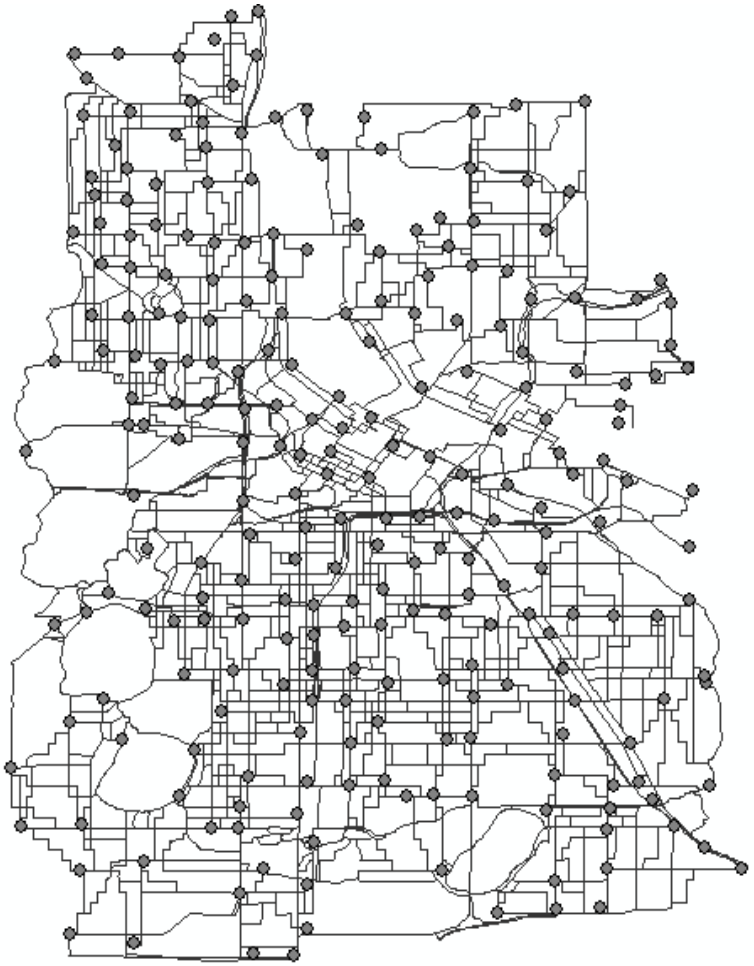
- Main Challenges
  - Problem 1: The total number of segments is too large thus difficult to visualize travel pattern.
    - Goal 1: Aggregated level heat map is necessary
  - Problem 2: The segment is too short, difficult to identify travel pattern.
    - Goal 2: Each “edge” on the trajectory heat map should be long enough to represent one travel pattern



Count on each segment on the intersection



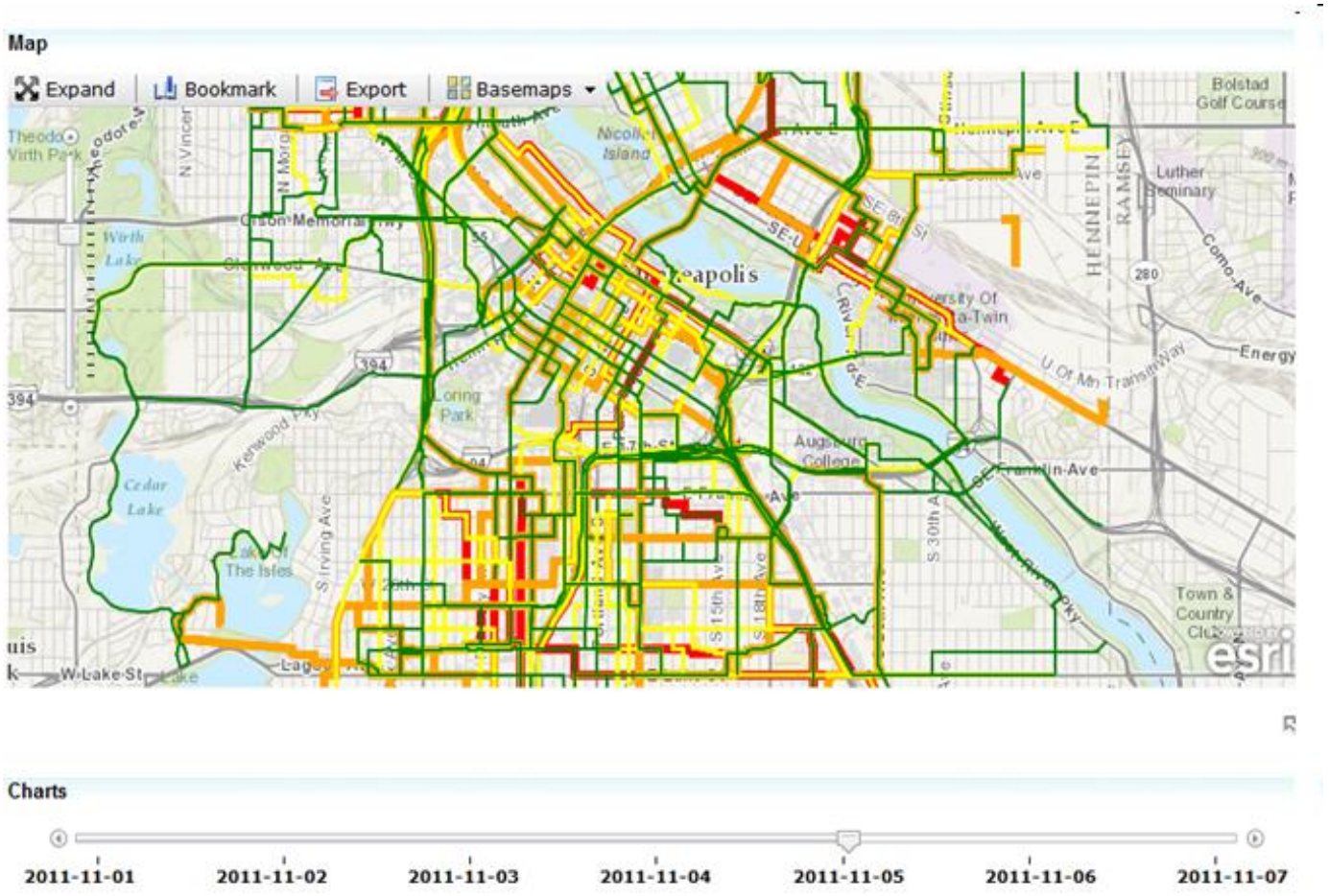
Count on sequences of segments



14k segments

Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

Trajectory heat map

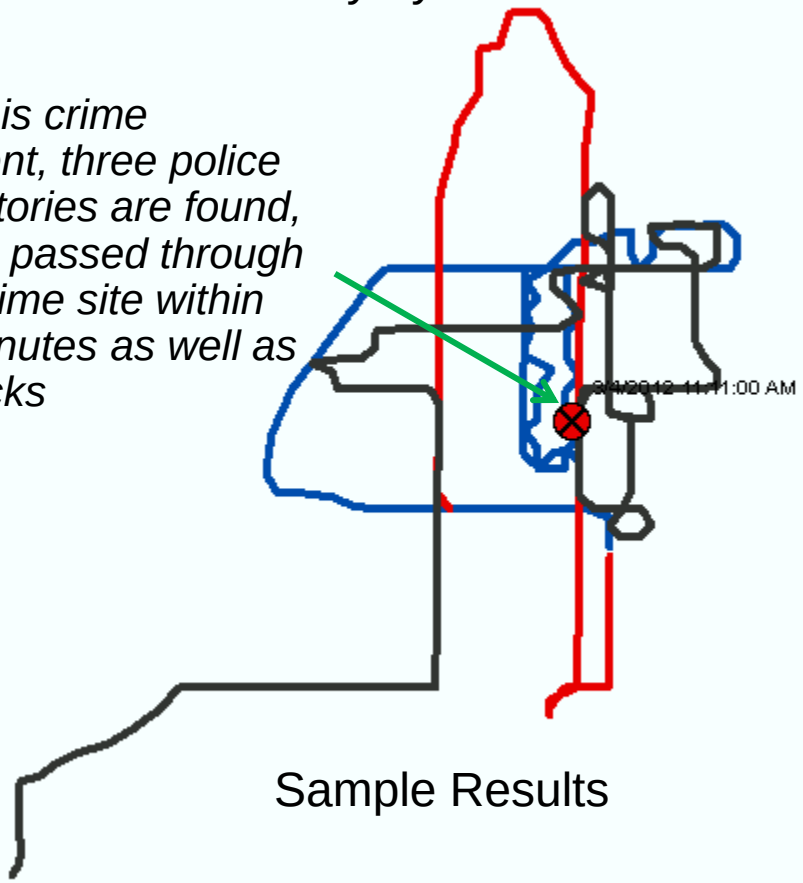


Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

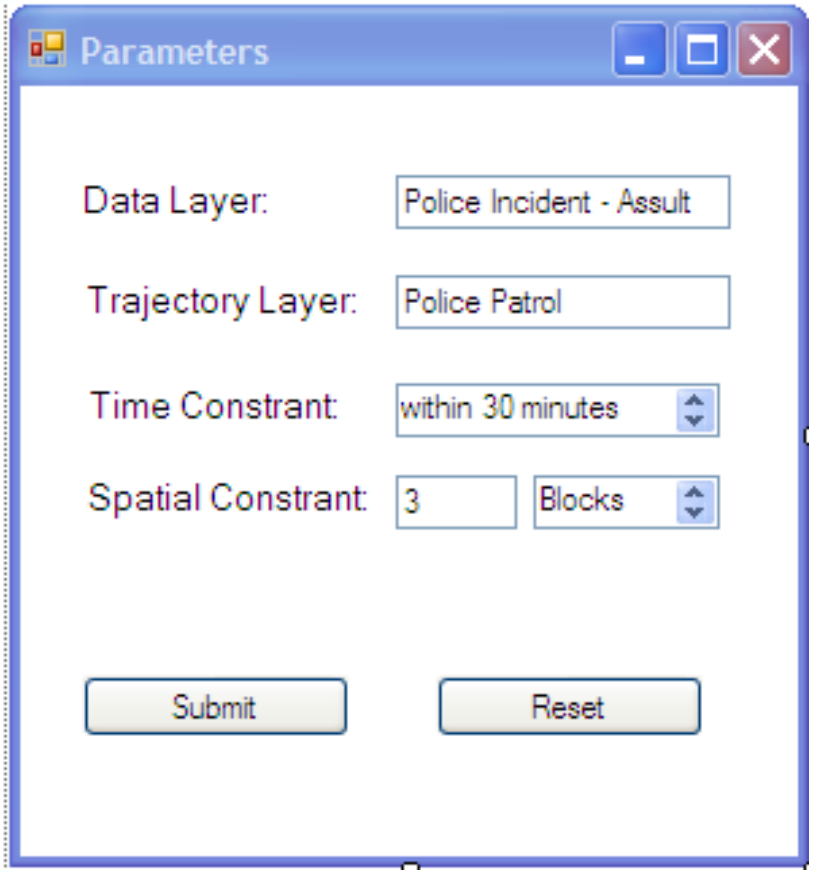
Cross layers trajectory analysis – individual trajectory study

- Individual trajectory study
  - Given a temporal point (e.g., incident, event), locate the trajectories, which are within both the spatial and temporal bounds
  - Note that the spatial and temporal constraint must be meet simultaneously by at least one GPS point on the trajectory

For this crime incident, three police trajectories are found, which passed through the crime site within 30 minutes as well as 3 blocks



Sample Results

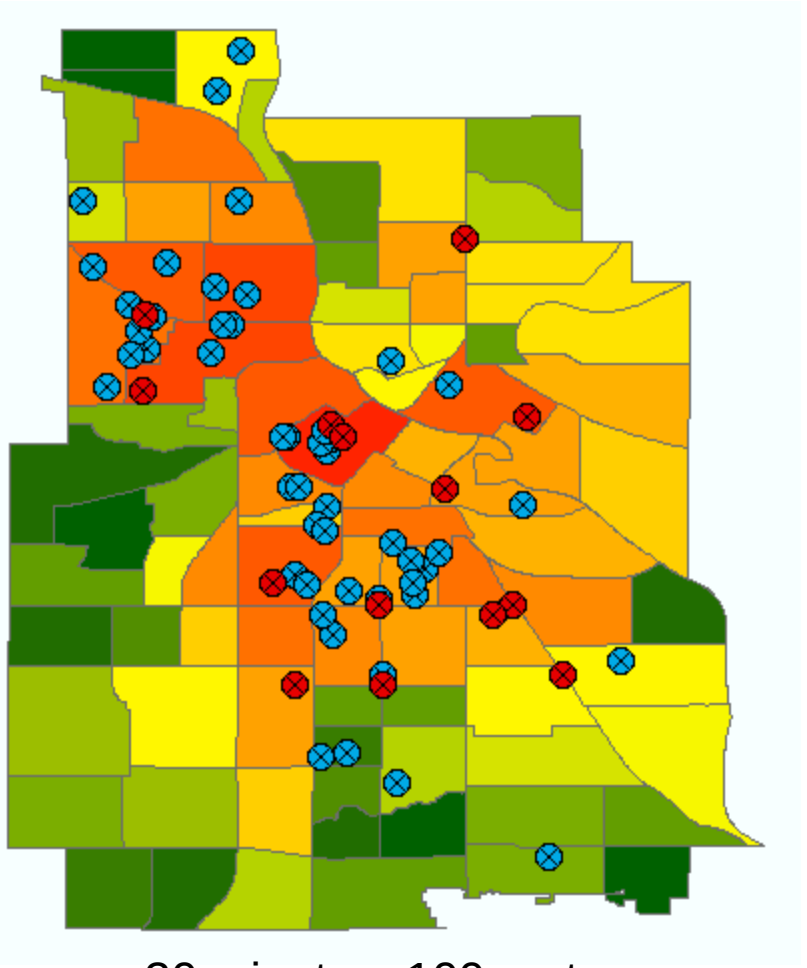


Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

Spatio-temporal correlation study

- Spatio-Temporal Analytics – ASLT+ROB 66 incidents

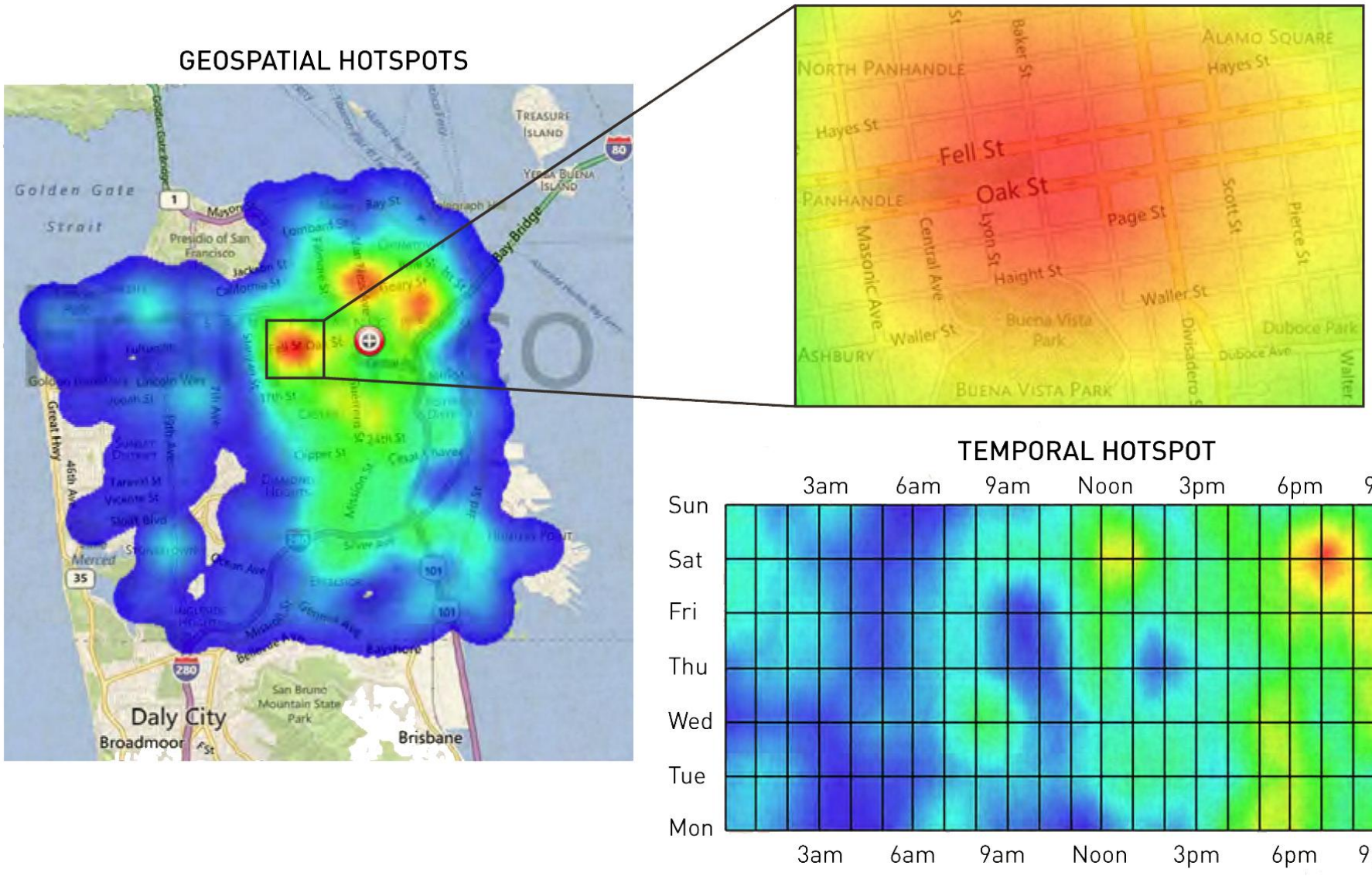
	50 Meters	100 Meters	200 Meters
Within 15 minutes	6(9%)	12(18%)	28(42%)
Within 30 minutes	7(11%)	**14(21%)**	37(56%)
Within 1 hour	13(20%)	23(35%)	46(70%)
Within 2 hours	27(41%)	37(56%)	57(86%)



30 minutes, 100 meters

Source: Hongfei Li, Xuan Liu, Spatio-Temporal Analytics

Geospatial and temporal car theft hotspots



Source: Booz Allen Hamilton, Field Guide to Data Science

📖 Decision-Making Based on Analytics

How good is analytics?

The Economist

National security

Intelligent intelligence

Just how good are government analysts?

Jul 19th 2014 | From the print edition

PREDICTION is difficult, as Niels Bohr once observed, especially about the future. But that has not stopped a huge industry that promises to do so from springing up. Commuters pore over newspaper columns; politicians spend fortunes on deciphering polling data; bosses sit in thrall to the pronouncements of coiffed young flip-chart merchants from management-consulting firms.

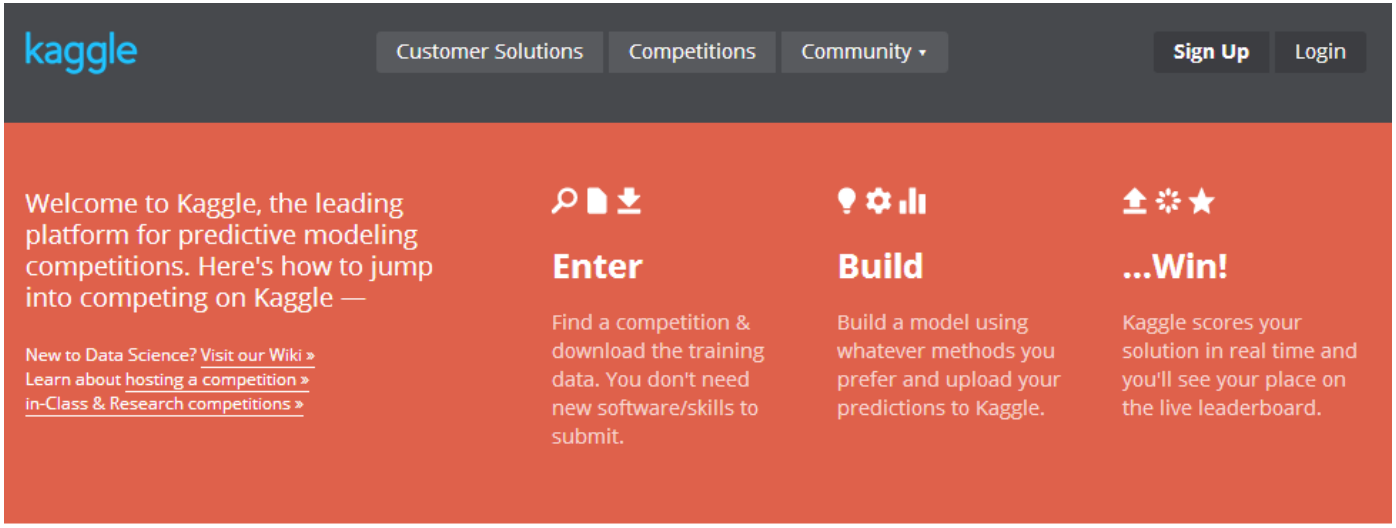
Governments are as keen as anyone to know the future, and one of the main jobs of intelligence agencies is to arm them with forecasts about how the world will look months or years from now. How good these are is an open question: spies often complain the public hears about their failures (to predict the September 11th attacks, for instance) but never of their successes.

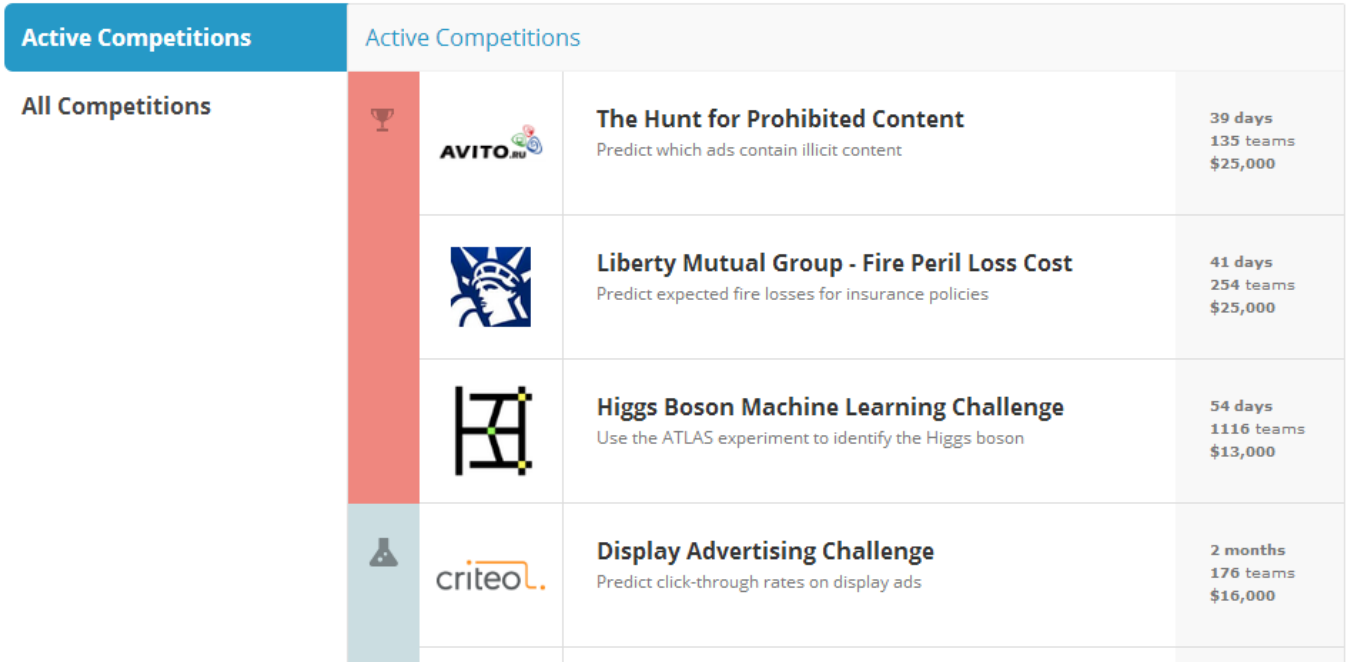
In a paper in the *Proceedings of the National Academy of Sciences*, David Mandel, of Defence Research and Development Canada, and Alan Barnes, a former intelligence analyst for the same country, take a stab at an answer. They analysed more than 1,500 intelligence forecasts produced by a nameless (but presumably Canadian) agency, covering the period from March 2005 to December 2011.

Their results suggest that the old joke about "military intelligence" being an oxymoron is unfair. When they compared what the analysts had said with what actually came to pass, they found that the predictions were right about three-quarters of the time. Cynics might wonder if the analysts merely restricted themselves to easy cases, but Dr Mandel and Dr Barnes also found they were good at calibrating their judgments. Events they deemed unlikely did not happen often, whereas those they thought likely occurred frequently. Indeed, if anything they were underselling themselves, tending to err more than necessary on the side of uncertainty. And there was evidence that their skills could be learnt—for more-experienced analysts tended to do better than their junior counterparts.

The result is even more striking because of its contrast with a famous earlier finding. In 2005 Philip Tetlock, a management theorist at the University of Pennsylvania, announced the results of a 20-year study in which 284 experts—professors, journalists, civil servants and so forth—were invited to make more than 28,000 predictions. Their performance was abysmal: barely better than chance, and inferior even to simple computer algorithms.

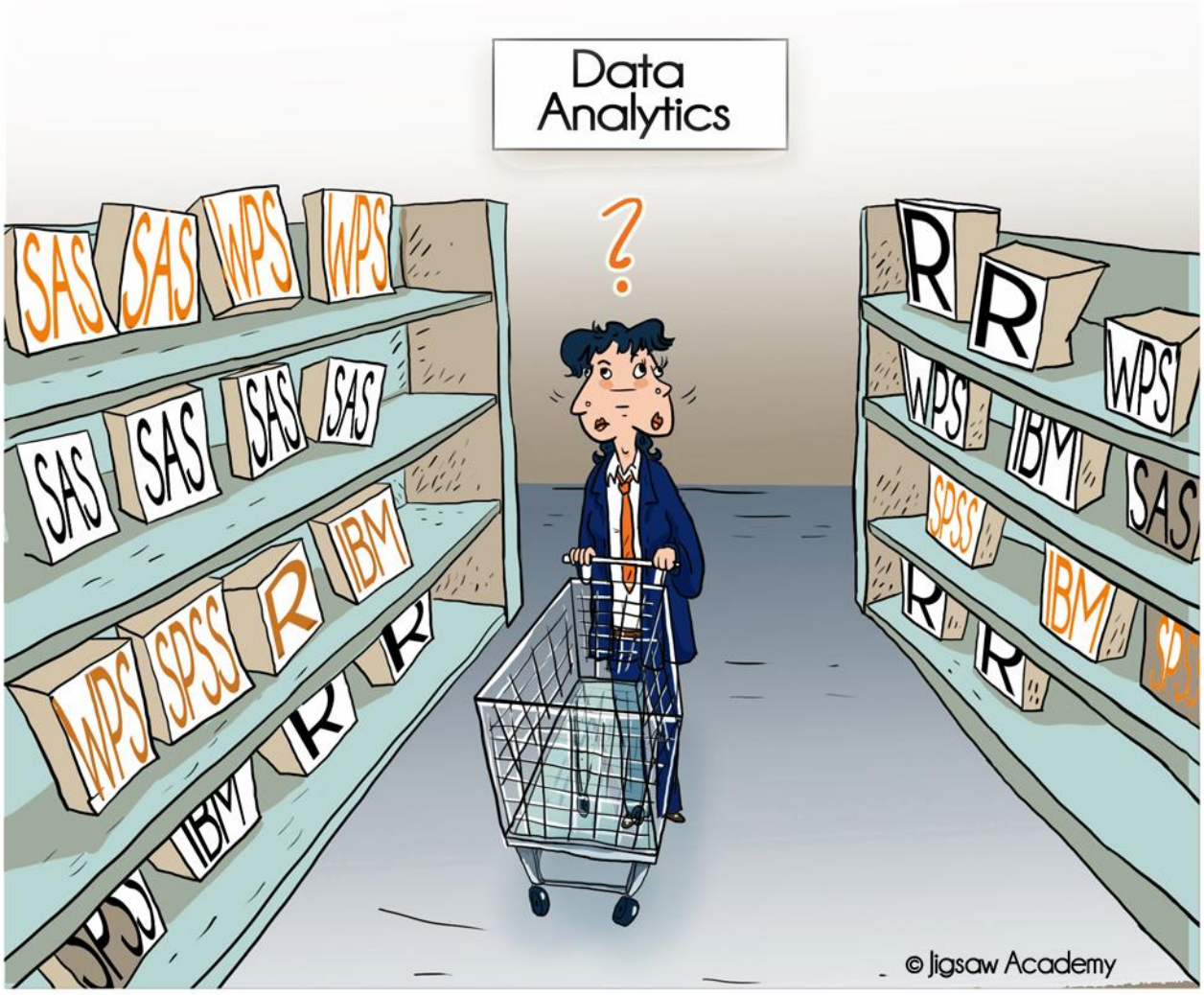
Crowd-sourcing analytics



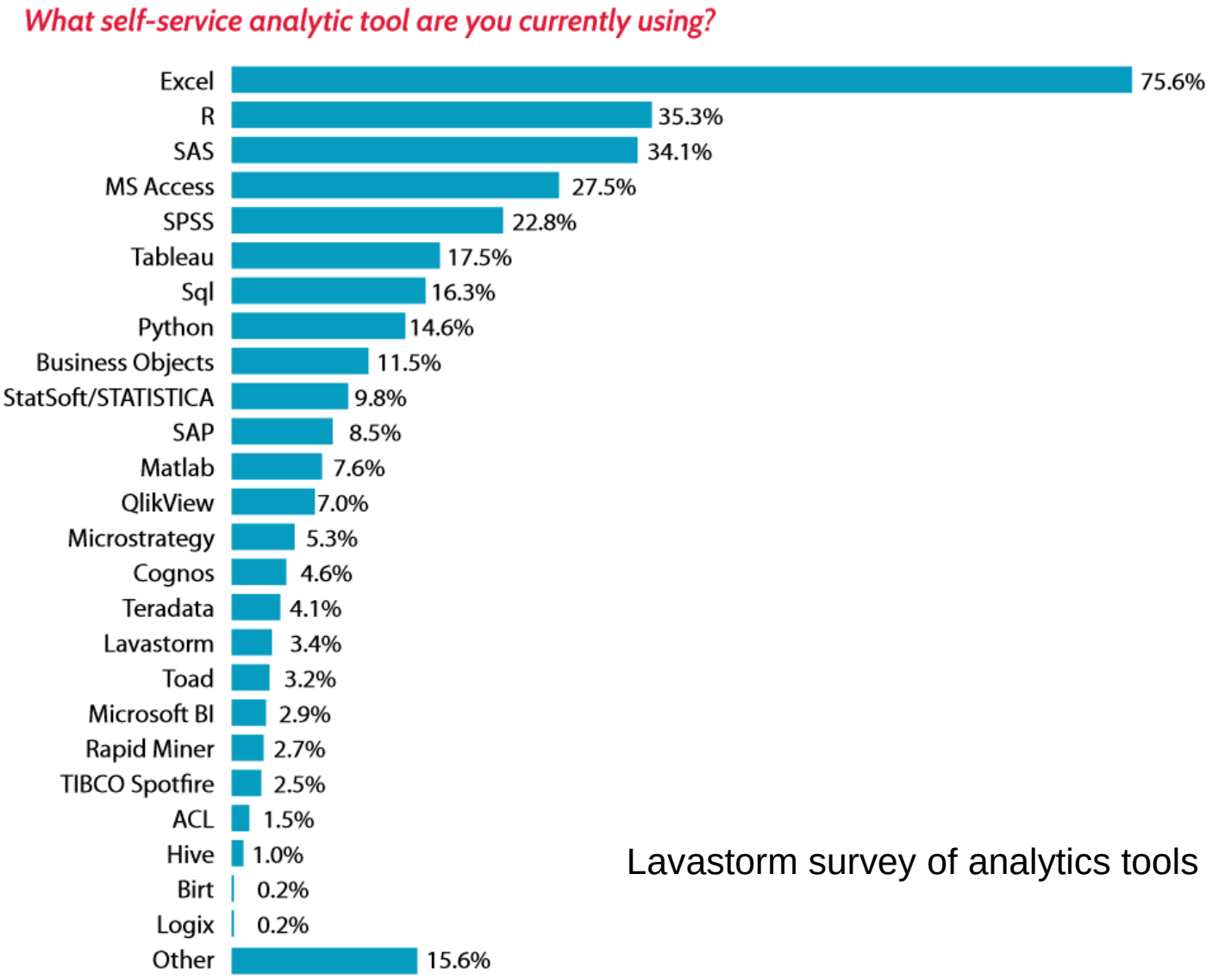


📖 Analytics Software

Software for analytics



Software for analytics



Lavastorm survey of analytics tools

Source: R. Muenchen "The Popularity of Data Analysis Software", http://r4stats.com/articles/popularity/

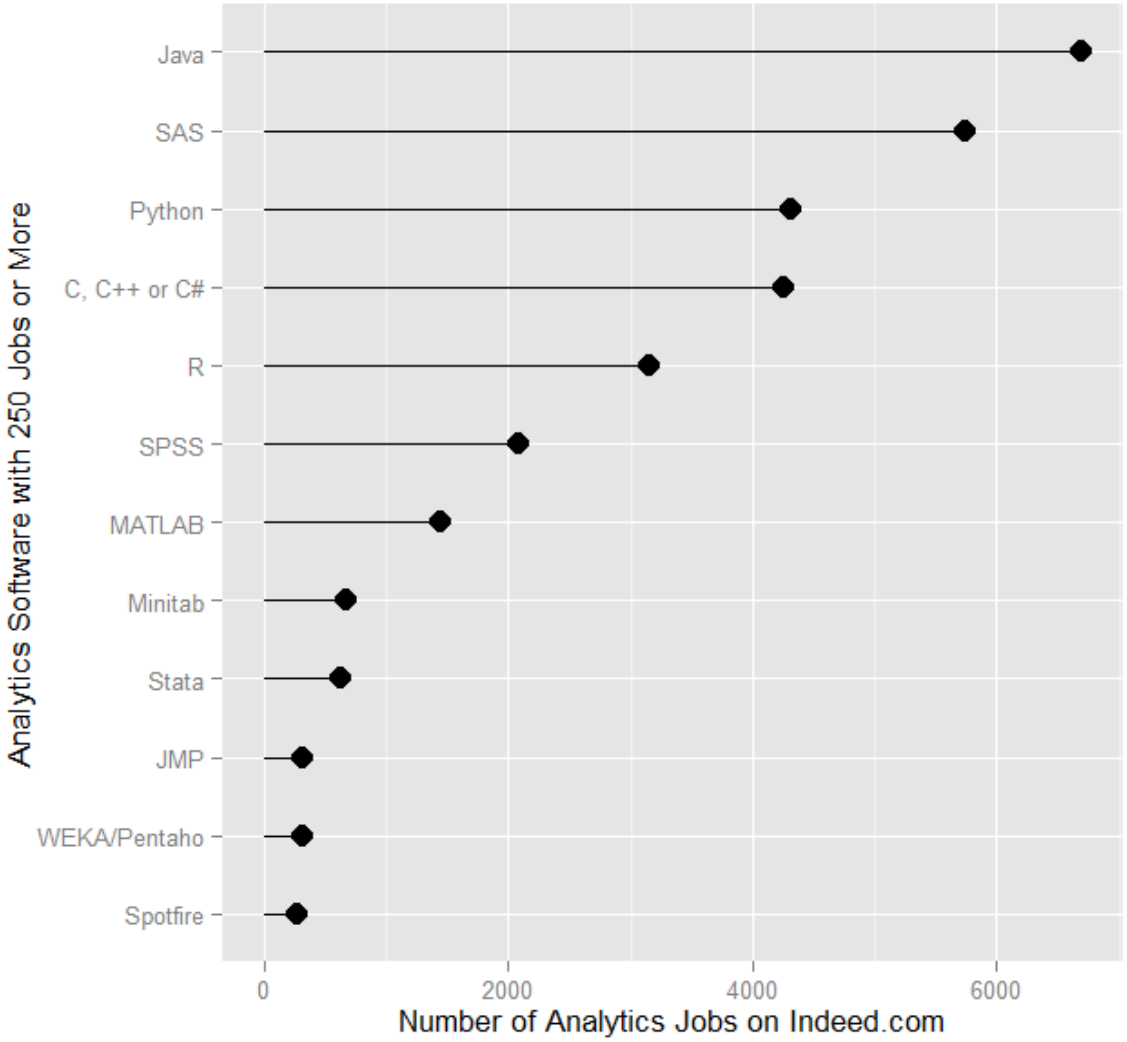
Software for analytics

Gartner “Magic Quadrant” plot of companies that sell advanced analtyics software (2014)

Source: R. Muenchen "The Popularity of Data Analysis Software", http://r4stats.com/articles/popularity/

Software for analytics

The number of analytics jobs for the more popular software (250 jobs or more, 2/2014)



Source: R. Muenchen "The Popularity of Data Analysis Software", http://r4stats.com/articles/popularity/