Health Care in the Era of Machine Learning and Cognitive Computing

Sean Grullon, PhD.
Data Scientist – IBM Watson Health

SLA PHT Division Spring Meeting
April 3rd 2017
Disclaimer

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction, and it should not be relied upon in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remain at our sole discretion.
Key Challenges in Healthcare & Life Sciences Industry

Information deluge

Medical literature doubling every few years

Approximately 700 K new scientific articles per year

Explosion of patient data (EHR, Fitbits, etc)


Demand outstripping supply

American Society of Clinical Oncology suggests that by 2025, overall demand for medical oncology services will grow 42 %

At the same time, supply of hematologists/oncologists is projected to grow only 28 %

2. The State of Cancer Care in America, 2014: A Report by the American Society of Clinical Oncology

Patients not being offered all treatment options

<20% of cancer patients are offered clinical trials as a treatment option

About 20% of cancer patients are eligible for a clinical trial; yet, trial participation is at about 3%

Takes up to 15 years for latest evidence to be fully adopted


Patients expect to participate in decision making around their care

Fewer than half of patients receive clear information on the trade-offs for their treatments and are satisfied with their level of control in medical decisions

New research and advances in medicine increase the complexity of Care and Research

Clinicians are challenged with...

- Understanding the patient condition
  - given disparate sources and varying completeness
- Formulating treatment options
  - based on relevant guidelines and medical literature
- Selecting personalized treatment plans
  - based on co-morbidities, conditions, contraindications, side effects for a patient’s specific clinical attributes

Researchers are challenged with...

- Staying up-to-date on medical literature
  - like rapidly increasing volume of medical literature
- Exploring and uncovering novel connections
  - looking across scientific domains for new relationships between diseases, genes and drugs
- Generating new insights for future research
  - to develop valid hypotheses with the potential to lead to groundbreaking discoveries

---

73
- the number of days it will take medical data to double by 2020

80%
- of the world’s healthcare data is unstructured

<5
- hours or less per month spent reading medical journals by 81% of reporting physicians

Unstructured Medical Data - a huge ocean of unused information

---


© 2017 IBM Corporation
Question: Where was Einstein born?

“One day, from among his city views of Ulm, Otto chose a watercolor to send to Albert Einstein as a remembrance of Einstein’s birthplace.”
What is cognitive computing?

*Cognitive*: of, relating to, or involving conscious mental activities (such as thinking, understanding, learning, and remembering)\(^1\)

Cognitive computing is a new computation paradigm that...

- learns and builds knowledge from various **structured and unstructured** sources of information;
- understands natural language and interacts more naturally with humans;
- Uses machine learning, designing computers to act without being explicitly programmed and developing algorithms to evolve behaviors based on data allowing prediction;
- devises complex models and algorithms that allows computers to find hidden insights without being explicitly programmed where to look;
- enhances the cognitive process of professionals to help improve decision making.

IBM Watson is a technology platform for **cognitive computing** that uses **natural language processing** and **machine learning** to reveal **insights** from large amounts of unstructured data.

---

**Given**
- Rich Questions posed in **Natural Language**
- Over a **Broad Domain of Knowledge**

**Goal is to Deliver**
- Relevant Insights
- Confidence-weighted insights
- Provide Context
- Fast Response Time

---

*Initial Goal of IBM Watson – Augment Human Expertise*
In May 1898, Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

In May, Daniel arrived in India after he celebrated his anniversary in Portugal.
In May 1898 Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

On the 27th of May 1498, Vasco da Gama landed in Kappad Beach.

Question:

Supporting Evidence:

Legend
- Temporal Reasoning
- Statistical Paraphrasing
- GeoSpatial Reasoning
- Reference Text
- Answer

Understand

Reason

Learn
Close to 30 services available today – a few examples provided below

**Sample Use Case:** A company wants to understand the personality characteristics of its customers for fine grained customer segmentation and create highly personalized interaction channels, better customer care

**Solution:** *Personality Insights API* in Watson Developer Cloud can extract and analyze a spectrum of personality attributes to help discover insights about people and entities to fine tune marketing messages

**Sample Use Case:** A company wants to translate some patents from Spanish into English for its IP research

**Solution:** *The Watson Language Translation service* provides domain-specific translation utilizing Statistical Machine Translation techniques that have been utilized over the past few decades
To do our part to drive bold and comprehensive transformation, we’ve built best-in-class capabilities in six key areas across the health landscape:

**Oncology and Genomics**
Working to help our clients transform cancer care for patients and providers

**Imaging**
Working to help our clients expand the role of medical imaging for better patient care

**Life Sciences**
Working to help our clients advance more rapid and efficient delivery of targeted and effective therapies

**Value-Based Care**
Working to help our clients manage cost and address quality by managing risk and populations, engaging with consumers, and making more confident decisions

**Government**
Working to help our clients improve the value of health and human services, lower costs, and have a meaningful impact on people’s lives

**Consumer**
Working to help our clients empower individuals to lead healthier lives
Two paths to innovation leveraging the power of Cognitive

Cognitive Solutions
solutions that are configured and trained to address specific use cases in Healthcare and Life Sciences industry

Hardened industry solutions

Cognitive Services
atomic services that address fundamental problems in processing unstructured medical information which can be rapidly combined to create meaningful solutions

Innovation platform for next generation apps

© 2017 IBM Corporation
Watson for Oncology trained by Memorial Sloan Kettering

Ability to surface insights from a vast body of medical literature, publications and guidelines data as physicians consider evidence-based treatment options for patients.

1. Extracts key attributes from a patient’s case.
2. Use those attributes to find candidate treatment options as determined by consulting NCCN Guidelines.
3. Search a corpus of evidence data to find supporting evidence for each option.
4. Use Watson analytics algorithms to rank treatment options based on evidence.

Patient Case
61 y/o woman s/p mastectomy is here to discuss treatment options for a recently diagnosed 4.2 cm grade 2 infiltrating ductal carcinoma...

Prioritized Treatment Options
Evidence Profile

- Inclusion / exclusion criteria
- Co morbidities
- Contraindications
- FDA risk factors
- MSK preferred treatments
- Other guidelines
- Published literature - studies, reports, opinions from Text Books, Journals, Manuals, etc.
- ...

© 2017 IBM Corporation
Life Sciences Cognitive Solution: Watson for Drug Discovery

Watson for Drug Discovery is a cloud-based, end-to-end scalable platform that helps life science researchers discover new disease pathways, new drug targets and additional drug indications.

Three Core WDD Components
- **Visualization**
- **Cognitive**
- **Knowledge**

The Watson Advantage

**Key Capabilities**
- Aggregates diverse content
- Cognitive technology
- Scalability
- Domain understanding
- Agility / speed

**Key Benefits**
- Accelerate insight generation
- Improve researcher productivity
Watson for Drug Discovery looks broadly across public, licensed and private data to unlock hidden information and deliver insights.
Other Cognitive Solutions

- **Watson for Clinical Trial Matching** - enables clinicians to quickly match patients with potential trials for cancer treatment using cognitive computing combined with natural language processing.

- **Watson for Genomics** - can help oncologists deliver precision medicine to their patients by providing information personalized to a patient’s tumor genomics using relevant information extracted from the massive volumes of medical literature.

- **Watson for Patient Safety*** – will support the detection, evaluation, and coding of adverse events contained in thousands of pages of spontaneous reports and medical literature.

* Not available – Solution is currently under development, available as a Proof-of-Concept
Cognitive Services Platform

• Watson cognitive solutions address some of the most difficult data and analytics challenges in healthcare & life sciences industry

• However in order to scale we need a cognitive innovation platform that allows rapid development of cognitive solutions

• Watson Health Cognitive Services Platform is designed to assist application developers to rapidly build apps using available cognitive services in a HIPAA-enabled environment

• A growing portfolio of healthcare services are under development and are expected to be made available in the future

• Some of these are microservices while others are more coarse grained
Conclusion

• Cognitive Computing is able to combine machine learning and natural language to reveal insights from large amounts of unstructured information

• Cognitive Computing can help transform the Healthcare and Life Sciences industry by uncovering data-embedded within unstructured information and generate insights

• IBM Watson has a family of Healthcare and Life Sciences solutions that can help address some of the most difficult problems for our customers

• At IBM Watson Health we are building a Cognitive Innovation Platform for Healthcare that is designed to allow rapid development of healthcare apps leveraging a set of cognitive services that address some of the fundamental challenges in processing unstructured medical content
Backup
Learn

From Past Experience to Predict Better Outcomes

Identifies many answers to questions and gathers supporting evidence

Evaluates hypotheses and quantifies confidence in answers

Adapts and learns to improve results over time

Improved accuracy based on learning from additional evidence, additional questions and feedback

Structured and Unstructured Data

Consumes large amounts of data and interprets and understands natural language

Deep semantic and contextual understanding of the data

For Possible Answers

Identifies many answers to questions and gathers supporting evidence

Candidate answers with supporting evidence

Supporting Evidence to Generate Confidence

Evaluates hypotheses and quantifies confidence in answers

Candidate answers ranked by confidence score
IBM Watson Health

How Watson answers questions

First Watson learns a new subject..

- All related materials are **curated and loaded** into Watson, such as Word documents, PDFs and web pages
- Watson taught on **relevant concepts, relationships** in that subject
- Sample **questions and answers pairs** used to **train Watson** on the subject by SMEs
- Watson is **updated** as new information is published

Then Watson answers a question..

- Watson can **search** millions of documents to find thousands of possible answers
- **Collects evidence** and uses **scoring algorithms** to rate the quality of this evidence
- **Ranks** all possible answers based on the score of its supporting evidence and as per training from SMEs
Diagram the sentence

- Look for verbs, nouns, subjects, etc.
- Look for punctuation, clauses, conjunctions, etc.
- Sentence structure and syntax provides insight into the meaning

Standardize the sentence

- Lemma words
- Number standardization
- Many more
Understand: Named Entity Detection

- The man is sixteen years old.

- What do the words represent?
  - “Louvre” probably refers to a museum
  - “Porsche 911” probably refers to a car

- Man
  - may refer to a human
  - may refer to all people
  - may be a specific person in the sentence

- How rare are these words? Rare words are often important words in the detection of concepts
Understand: Inference

- Relies on results of grammar, named entity detection, etc.
- Determines when some text refers back to some previously mentioned entity
- Watson knows that “He” refers to “Vasco da Gama” because this relationship is expressed in the text. **Watson only knows what he can read.**

Vasco da Gama commanded a Portuguese ship. **He** sailed more than 6000 miles over open ocean.
The DeepQA architecture inside Watson

Generates many hypotheses, collects a wide range of evidence and balances the combined confidences of over 100 different analytics that analyze the evidence from different dimensions.
DeepQA Architecture

Multiple interpretations

Hundreds of answer sources
• Primary search
• Candidate answer generation

Tens of thousands of evidence sources and scores
• Answer scoring
• Evidence retrieval
• Deep evidence scoring

Learned models
• Combine and weigh evidence

Natural language processing

Information retrieval

Knowledge representation and reasoning

Deep analytics

Parallel and distributed computing

Machine learning

Inquiry analysis
Decomposition
Hypothesis generation
Hypothesis and evidence scoring
Synthesis
Final confidence merging and ranking

© 2017 IBM Corporation
Watson for Clinical Trial Matching

Business challenge:
- Clinicians have no easy way to search across eligibility criteria of relevant clinical trials for their patient.

Watson solution:
- Use patient data to identify candidate trials for which the patient might be eligible by providing criteria level evaluation based on a patient’s attributes.

Two use cases:
- Point of care: assess patient eligibility against all available trials.
- For the Sponsor: provide the sponsor insight into the recruitment pipeline for their trial at a specific site.

Initial focus is oncology, Watson is trained in breast, lung, colon and rectal cancer, with other tumor types to follow.

---

15%  
Cancer patients are aware of clinical trials as a treatment option.  

20%  
.. of oncology clinical studies are completed on time.  

<5%  
Of cancer patients enroll in a clinical trial.  

>12,000  
Active clinical trials worldwide.

---

IBM launched an initiative to accelerate a new era of genomic medicine

**Challenges**

- As the cost of Next Generation Sequencing decreases, we anticipate that there will be an increase in tumor genome sequencing resulting in massive quantities of genetic data to analyze.
- Clients stated that it can take from days to weeks to perform a comprehensive manual analysis of the genetic alterations for one patient.
- Identifying alterations driving the patient cancer and matching them with molecular targeted therapies using multiple data sources is extremely complex and labor-intensive.

**Value**

“Using Watson’s cognitive computing capabilities, I hope it will be possible for oncologists like me to quickly mine insights from the immense amount of genomic data that’s becoming available about individual patients by using Watson to identify potential drugs that target our patients’ specific genetic profiles.”

– Dr. Lukas Wartman, Cancer Survivor and Assistant Director, Cancer Genomics, The Genome Institute at Washington University.

The more exciting part is that in 30 percent of patients, Watson found something new—so that’s 300-plus people where Watson identified a potential treatment for consideration that a well-meaning, hard-working group of physicians hadn’t found – Dr. Ned Sharpless, head of the University of North Carolina’s Lineberger Comprehensive Cancer Center.

**Success**

Launched and available in two offerings:
- WfG (standalone)
- Watson Genomics from Quest Diagnostics® (End-to-End)

© 2017 IBM Corporation
Watson for Genomics Overview

Service Analysis, Reports, & Visualizations

Molecular Profile Analysis

Pathway Analysis

Drug Analysis

Content

- 20+ Content Sources Including:
  - Medical Articles
  - Drug Information
  - Clinical Trial Information
  - Genomic Information
  - OncoKB by MSKCC
Functionality Highlights

- “Born in the cloud” multi-user and multi-tenant solution with a single code base
- No customization, configuration or integration required for initial use
- Security rich environment managed by IBM and industry standards
- Patient data uploaded to WfG is de-identified (de-identified mutated DNA)
  - Accepted input data includes somatic mutations, copy number variations, gene expression and fusion
  - Supports gene panels, whole exome and whole genome sequenced files
  - Natural Language Processing (NLP) used to extract information from extensive medical literature (over 23 millions articles)
  - 20+ structures and unstructured data sources ingested
- Analytics engine to identify relevant alterations, drugs and clinical trials for specific types of cancer
- Report and interactive visualizations of the molecular profile, drugs and pathways
- Summary report shows target therapeutic options categorized by FDA approved for the patient cancer type, Investigational and FDA approved for other cancer types
- Evidences presented via hyperlinks to sources for easy drill down
What if case processing for life sciences companies could become a largely automated process?

Watson creates E2B File
- Individual Case Safety Reports (ICSRs) Detection
- Seriousness
- Expectedness
- MedDRA Coded
Priority 15 days

Causality assessed & narrative generated

Watson identifies the case as an addendum E2B File amended with addendum
New MedDRA Coded
Priority 7 days

Medical review Final Case checked by a reviewer and sent to reporting & agency

With Watson
Watson for Patient Safety uses cognitive computing to help increase speed and efficiency of case processing

With Watson, we aim to transform these processes by reducing the time and manual effort required, while maintaining a high level of quality, consistency and accuracy.
Developers can rapidly build healthcare apps leveraging the healthcare cognitive services platform

Sample Use Case: Build a cognitive pipeline that will take in patient EMR data (unstructured) and extract Oncology specific attributes (structured)
Watson Health Core Services and the Watson Health Portfolio

Watson Health Core offers a landing zone where multi-dimensional data can be captured and normalized for other health and research applications to use.

- Life Sciences Solutions
- Genomics and Oncology Solutions
- Imaging Solutions
- Value-based Care

Enriched Data Sets: Explorys, Phytel, Merge, etc.

Informatics Studio (coming soon)
Improve patient care through aggregated views of clinical, research, and social health data.

Watson Health Core
Innovate with HIPAA-enabled health-data platform as a service.

Developer tools and ecosystem
(more to come)

Watson Health for ResearchKit
Enables the use of iOS devices for medical research, integrated into the HIPAA-enabled cloud platform.

IBM SoftLayer Infrastructure as a Service

This information is intended to outline our general product direction, and it should not be relied on in making a purchasing decision. Information on roadmaps is not a commitment and may be subject to change.