Watson Beyond Jeopardy!: Adaptation to the Medical Domain

Jennifer Chu-Carroll
IBM T.J. Watson Research Center

Research Work by the Watson Technologies Team
Automatic Open-Domain Question Answering
A Long-Standing Challenge in Artificial Intelligence to emulate human expertise

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

- **Deliver**
  - **Precise Answers:** Determine what is being asked & give precise response
  - **Accurate Confidences:** Determine likelihood answer is correct
  - **Consumable Justifications:** Explain why the answer is right
  - **Fast Response Time:** Precision & Confidence in <3 seconds
The Jeopardy! Challenge
A palpable, compelling and notable way to drive the technology of Question Answering along Key Dimensions

- **Broad/Open Domain**
- **Complex Language**
- **High Precision**
- **Accurate Confidence**
- **High Speed**

**$200**
The juice of this bog fruit is sometimes used to treat urinary tract infections

**$400**
This Bavarian city traces its origins to a Benedictine monastery at Tegernsee

**$600**
In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus

**$800**
Grace Murray Hopper is credited with applying this 3-letter term to a mysterious computer problem

**$1000**
Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

© 2013 IBM Corporation
DeepQA: The architecture underlying 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.
Watson answers by finding, reading, scoring and combining evidence.

**Content (Structured & Unstructured)**

Watson answers by finding, reading, scoring and combining evidence.

**Candidate Answer/Hypothesis Generation**

**High-Speed Evidence Retrieval**

- Isaac Newton
- Wilhelm Tempel
- HMS Paramour
- Christiaan Huygens
- Halley’s Comet
- Edmond Halley
- Pink Panther
- Peter Sellers

**Important Terms:** 1698, comet, paramour, pink, ...

**AnswerTypes:** comet discoverer

- Date(1698),
- Took(discoverer, ship)
- Called(ship, Paramour Pink)

**Diverse and Extensible Evidence Scoring**

- [0.58 0.5 -1.3 ... 0.97]
- [0.71 1 13.4 ... 0.60]
- [0.42 0 2.0 ... 0.90]
- [0.84 0.5 10.6 ... 0.88]
- [0.33 0 6.3 ... 0.83]
- [0.21 1 11.1 ... 0.92]
- [0.91 0 -8.2 ... 0.31]
- [0.91 0 -1.7 ... -.20]

**100’s of NLP Scoring Algorithms**

- 1) Edmond Halley (0.85)
- 2) Christiaan Huygens (0.20)
- 3) Peter Sellers (0.05)
- 4) ...

**Merging & Ranking Based on Statistical Machine Learning**

IBM Watson
Playing in the Winners Cloud

Baseline 12/06

v0.1 12/07
v0.2 05/08
v0.3 08/08
v0.4 12/08
v0.5 05/09
v0.6 10/09
v0.7 04/10
v0.8 11/10
With Precision, Accurate Confidence and Speed, the rest was History
Watson: A New Era of Computing

System Intelligence

1900
1950
2011

Tabulation
Punch cards
Time card readers

Programmatic
Search
Deterministic
Enterprise data
Machine language
Simple outputs

Cognitive
Discovery
Probabilistic
Big Data
Natural language
Intelligent options

Watson: A New Era of Computing
Watson Beyond Factoid Question Answering

1. Understands natural language input

2. Generates and evaluates evidence-based hypothesis

3. Adapts and learns from user selections and responses
Use of Question Answering in Medical Diagnosis

After Watson’s win on Jeopardy!, people assumed that anything that could be phrased as a question could be correctly answered by Watson:

Watson, “Given my medical record <insert hundreds of pages of structured and unstructured data here>, what’s wrong with me?”

But that isn’t what Watson was designed for:

- Watson wants a single sentence question
- Watson wants to find passages based on concepts in the question
- Watson wants to explore candidates found in relevant passages
- Watson wants to align answer-bearing passages with questions
The New Watson Challenge

We accepted the implied challenge to facilitate the reasoning process over a complex scenario:

**Input**
- Complex natural language description of a problem

**Output**
- Evidence-based inference chains leading to hypotheses

Our first domain of exploration is medical diagnosis because of its mature, complex and meaningful problem solving nature
Taking Watson beyond Jeopardy!

Understanding

Specific Questions

The type of murmur associated with this condition is harsh, systolic, and increases in intensity with Valsalva

Interacting

Question-In/Answer-Out

From specific questions to rich, incomplete problem scenarios (e.g. EHR)

Explaining

Precise Answers & Accurate Confidences

Move from quality answers to quality answers and evidence

Learning

Batch Training Process

Scale domain learning and adaptation rate and efficiency

Entire Medical Record

Rich Problem Scenarios

Interactive Dialog

Teach Watson

Comparative Evidence Profiles

Continuous Training & Learning Process

Input, Responses

Dialog

Refined Answers, Follow-up Questions

Answers, Corrections, Judgements

Responses, Learning Questions

Evidence analysis and look-ahead, drive interactive dialog to refine answers and evidence

Comparative Evidence Profiles
Sample Patient Scenario from US Medical Licensing Exam

A 70-year-old man comes for a follow up with his cardiologist. There are no specific complaints. Findings at the physical exam are BP- 130/80 mmHg, HR- 80 beats/min, and appearance of pale mucous membranes. Lungs are clear to auscultation, and there is no edema of lower extremities. Fecal occult blood test (FOBT) was negative. Blood test shows hypochromic microcytic RBCs. Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin. What is the most probable diagnosis in this patient?

(A) Anemia of chronic disease
(B) Anemia secondary to iron deficiency
(C) Beta thalassemia
(D) Megaloblastic anemia
(E) Sideroblastic anemia

A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta-hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?

(A) Streptococcus pneumoniae
(B) Staphylococcus aureus
(C) Peptostreptococcus
(D) Streptococcus pyogenes
(E) Staphylococcus epidermidis

➢ The answers are not one step away
➢ Finding them requires connecting the dots
➢ Shallow language understanding is not enough
➢ Discovering rationalized paths through the content becomes a key value
WatsonPaths: Beyond Factoid QA
A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?

- Objectives:
  1. Figure out what’s wrong with the patient
  2. Identifying effective treatment/next steps

- Identify critical information from scenario for diagnosis
  - Parsing
  - Co-reference resolution
  - Negation detection
  - Clinical factor identification
Scenario Analysis Results

A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?
“She has pain in the epigastric region and sometimes on the right side of her abdomen.”
A 70-year-old man comes for a follow up with his cardiologist. There are no specific complaints. Findings at the physical exam are BP- 130/80 mmHg, HR- 80 beats/min, and appearance of pale mucous membranes. Lungs are clear to auscultation, and there is no edema of lower extremities. Fecal occult blood test (FOBT) was negative. Blood test shows hypochromic microcytic RBCs. Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin. What is the most probable diagnosis in this patient?
WatsonPaths for Medical Diagnosis (Cont’d)

Blood test shows hypochromic microcytic RBCs.

Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin.

- low total iron-binding capacity (TIBC)
- increased ferritin
- low serum iron

Ask Watson: What disease causes low serum iron?

- Anemia of chronic disease
- Anemia secondary to iron deficiency
- Beta thalassemia
- Megaloblastic anemia
- Sideroblastic anemia

Ask Watson: What are findings of anemia of chronic disease?
WatsonPaths for Medical Diagnosis (Cont’d)

2000057 What disease causes low serum iron?

Evidence Profile

<table>
<thead>
<tr>
<th>Disease</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anemia of chronic disease</td>
<td>59%</td>
</tr>
<tr>
<td>Sideroblastic anemia</td>
<td>8%</td>
</tr>
</tbody>
</table>

Details

100% "Rheumatoid arthritis" Corpus: Web Corpus Expansion

In most cases, the reduced red cell mass is caused by the **anemia of chronic disease**, a normocytic-normochromic process characterized by a low concentration of serum iron, a low serum iron-binding capacity, and a normal or increased serum ferritin concentration.

65% "Iron deficiency anemia" Corpus: Web Corpus Expansion

Increased stainable iron in macrophages. **Sideroblastic anemia** is suspected in patients with microcytic anemia or a high RDW anemia, particularly with increased serum iron, serum ferritin, and transferrin saturation (see Anemias Caused by Deficient Erythropoiesis: Iron Deficiency Anemia).
A young woman presented with pneumonia, and a course of cephalosporins were given to treat this condition. She developed a non-bloody, watery diarrhea with abdominal cramps in a few days after starting the treatment and came back to the doctor’s office. Her stool was sent to the laboratory and a colonoscopy performed to find out the cause of this diarrhea. Colonoscopy showed yellow-white plaques and membranes on colonic mucosa. Which of the following is the most likely toxin, secreted from this microorganism, which is causing these colonic changes?

- Neurotoxin
- Endotoxin
- Tetanospasmin
- Alpha toxin
- Cytotoxin B
Ask Watson: What causes colonoscopy to show yellow-white plaques and membranes on colonic mucosa?
Finding More Missing Links And Answering the Question

Ask Watson: What is a finding of pseudomembranous colitis?

Ask Watson: Cytotoxin B is the most likely toxin secreted from what microorganism?
WatsonPaths builds complex inference graphs by relying on various systems (including Watson) to generate relations and confidence between nodes. With this capability WatsonPaths can answer questions that the base Watson system cannot. It provides a powerful and interactive decision support paradigm over large volumes of unstructured content.
Leveraging Existing Medical Resources

- UMLS (Unified Medical Language System) from NLM
  - ~100 sources, sort of merged
  - ~3M unique concept identifiers (not unique concepts), organized in a type hierarchy
    - activities, anatomy, chemicals/drugs, devices, disorders, genetics, organisms, physiology, procedures, ...
  - ~350 relation types; ~30M unique relation instances
    - diagnoses, treats, finding_site_of, has_causative_agent, contraindicates, ...

- Sample Uses of UMLS
  - In Watson QA system
    - Type Coercion: does a candidate answer match the type the question is seeking
    - Candidate generation
    - Term matching
  - In WatsonPaths
    - Clinical factor identification
    - Relation generation in inference graph
    - Term matching
As with other NSAIDs, ibuprofen may be useful in the treatment of severe orthostatic hypotension. Lasix (furosemide), a diuretic, and ibuprofen, an NSAID, can be taken together.

Rule-based relation detector identifies hyponymy relations in text:

Frame 01
- subj: Ibuprofen
- type: NSAID

Frame 02
- subj: Lasix
- type: diuretic

Frame 03
- subj: NSAID
- type: drug

Ibuprofen is a NSAID
Lasix is a diuretic
NSAID is a drug

200M
Urinary Tract Infection
The most common symptoms of a bladder infection are burning with urination, frequency of urination, an urge to urinate, without vaginal discharge or significant pain.[4] An upper urinary tract infection or pyelonephritis may additionally present with flank pain and a fever. Healthy women have an average of 5 days of symptoms.[4]

The symptoms of urinary tract infections may vary with age and the specific part of the urinary system that was affected. In young children, urinary tract infections may present with diarrhea, loss of appetite, nausea and vomiting, fever and excessive crying that cannot be resolved by typical measures. [5] [6]

Cystitis Symptoms:
- Sudden onset
- Dysuria (painful urination)
- Nocturia
- Low back pain
- Pneumaturia

### Disease Symptom KB

<table>
<thead>
<tr>
<th>Disease</th>
<th>Symptom</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTI</td>
<td>burning with urination</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>UTI</td>
<td>fever</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Cystitis</td>
<td>dysuria</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Unsupervised Learning

Topics Clusters: term co-occurrence in documents - LSA

Type Clusters: terms share similar syntactic roles (Prismatic)
Complexity of Language in the Medical Domain

- **Terminology**: 
  - Food would “get stuck” when she was swallowing.
  - Swallowing difficulty...
  - Food gets held-up...

- **Causation**: 
  - ...can cause food to move slowly in the esophagus.

- **Chronology**: 
  - Sudden onset of chills
  - Cold chills
  - Delta-Wave
  - Abnormal QRS Complex
  - Normal QRS Pattern
  - PR Interval
  - Domain

- **Magnitude**: 
  - High Temperature
  - Temperature
  - Fever
  - Fever after acute symptoms subside...

- **Negation**: 
  - Nonproductive cough
  - Productive cough

- **Causation**: 
  - Abdomen Pain
  - Urination Pain
  - Dysuria

- **Location**: 
  - Flank Pain
  - Kidney Pain
  - Lower Back Pain

- **Terminology**: 
  - Pneumaturia
  - Bubbles in the urine
  - Abdomen Pain exacerbated by exercise

- **Domain**: 
  - Abdomen Pain
  - Flank Pain
  - Lower Back Pain
  - Kidney Pain
### Sign/Symptom Matching

| Syndrome of sore throat, fever, **sepsis** and **unilateral neck swelling** | ...cause **septic thrombophlebitis** of the internal jugular vein (Lemierre syndrome). Most patients have fever, sore throat, odynophagia, and **swelling in the neck down to the hyoid bone** |
| Nasal mucosal atrophy and **foul-smelling crusts in the nasal passages** | Atrophic rhinitis is characterized by progressive nasal atrophy, mucosal colonization with Klebsiella ozaenae or other organisms and **foul smelling nasal discharge** |
| Syndrome characterized by hypokalemic metabolic alkalosis, mild hypotension, **calluses on the knuckles** and **enamel erosion** | Many individuals with bulimia have **skin abrasions on their knuckles** from inducing vomiting. The most common effect of anorexia and bulimia is **tooth enamel erosion**. |
| **Flexing patient's right hip and knee to elicit pain** is used to diagnose this condition | For example, the obturator sign is present when the **internal rotation of the thigh elicits pain** (i.e., pelvic appendicitis), and the psoas sign is present when the **extension of the right thigh elicits pain** (i.e., retroperitoneal or retrocecal appendicitis) |
### Paraphrases/Entailment

<table>
<thead>
<tr>
<th>Question Text</th>
<th>Passage Text</th>
<th>Learned Axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murmur associated with this condition is harsh, systolic, diamond-shaped, and increases in intensity with Valsalva</td>
<td>A systolic murmur that increases with the valsalva maneuver and disappears with squatting suggests hypertrophic cardiomyopathy</td>
<td>X suggests Y =&gt; X associated with Y</td>
</tr>
<tr>
<td>Class of drugs causing regression of polyposis in familial adenomatous polyposis</td>
<td>NSAIDs have been shown to induce adenoma regression in patients with familial adenomatous polyposis</td>
<td>X has been shown to induce Y =&gt; X causes Y</td>
</tr>
<tr>
<td>Intravenous treatment for cyanide poisoning</td>
<td>Antidotes for cyanide poisoning include amyl nitrate, sodium nitrate, and intravenous sodium thiosulfate.</td>
<td>Antidotes for X include Y =&gt; Y is treatment for X</td>
</tr>
<tr>
<td>Syndrome characterized by narrowing of the extra-hepatic bile duct from mechanical compression by a gallstone impacted in the cystic duct</td>
<td>Mirizzi’s syndrome, a rare condition in which a gallstone impacting the cystic duct obstructs the common bile duct by edema and extrinsic compression</td>
<td>X obstructs Y =&gt; narrowing of Y by X</td>
</tr>
<tr>
<td>Preferred corrective treatment for acute episodes of angioedema in patients with hereditary angioedema</td>
<td>For acute episodes of angioedema in hereditary angioedema, administer intravenous, purified, nanofiltered C1-INH concentrate as first-line therapy</td>
<td>For X, administer Y as first-line therapy =&gt; Y is preferred treatment for X</td>
</tr>
</tbody>
</table>
Explore the Evidence and Teach Watson

See Watson’s evidence for making a connection. Judge it to improve the inference and Teach Watson in the process.
What would have to be true for seizure disorder to be correct?

What neurological condition contraindicates the use of bupropion?

Patients with preexisting seizure disorder should not use bupropion due to a higher-than-proportional increase in the possibility of seizure as the dose is increased.

Does *should not use* bupropion mean *contraindicates the use of* bupropion?

Does “*should not use*” mean “*contraindicates the use of*” in general?
Learning Entailments

Watson considers...

Syndrome characterized by **narrowing** of the extra-hepatic bile duct from mechanical compression by a gallstone impacted in the cystic duct

**Mirizzi’s syndrome**, a rare condition in which a gallstone impacting the cystic duct **obstructs** the common bile duct by edema and extrinsic compression

Automatically generates learning questions...

Does a **gallstone** **obstructs** the common bile duct mean **narrowing** of bile duct by a gallstone

Does **“obstructs”** entail **“narrowing”** in general?
Learning Types

Watson considers...

Thienopyridine derivative that reduces the incidence of MI in acute coronary syndrome

Automatically generates learning questions...

Is clopidogrel a thienopyridine derivative?

What would have to be true for clopidogrel to be correct?

Large clinical trials have demonstrated that combination antiplatelet therapy with clopidogrel and aspirin significantly reduces the risk of adverse cardiac events in patients with acute coronary syndromes.
Some Language Challenges Require a Longer-Term Investment

Tackling the language/logic problem in a general and comprehensive way will take a longer-term sustained research investment. We need to provide value to the user today while the system actively learns through collaboration.

Attacks of Meniere's disease are precipitated by this dietary indiscretion

For treating Meniere's disease, treatment includes use of antibiotics and dietary changes. A low salt diet might also help in alleviating the symptoms of tinnitus and Meniere's disease to some extent.

Both are referring to eating too much salt. But a low-salt diet is not a dietary indiscretion. However, if it alleviates a problem then perhaps the opposite (eating too much salt) is?

Drug treatment for nerve gas exposure that reactivates acetylcholinesterase

Some of the nerve agents attack and destroy acetylcholinesterase by phosphorylation, so the action of acetylcholine becomes prolonged, pralidoxime (2-PAM) is the cure for organophosphate poisoning because it can cleave this phosphorylation.

Reactivation may be described by cleaving an action that may have caused an attack and destruction? If we worked hard to identify and manually model and generate such a complex and highly contextual rule – how many times would it be used? There are potentially millions of these. But where? Will they apply generally? Are they worth the investment?
Taking Watson beyond Jeopardy!: Recap

- Understanding
- Interacting
- Explaining
- Learning

Rich Problem Scenarios -> Interactive Dialog Teach Watson -> Comparative Evidence Profiles -> Continuous Training & Learning Process

- Entire Medical Record
- Input, Responses, Dialog
- Refined Answers, Follow-up Questions
- Answers, Corrections, Judgements, Responses, Learning Questions
Additional Business Applications

Healthcare / Life Sciences: Diagnostic Assistance, Evidence-Based, Collaborative Medicine

Tech Support: Help-desk, Contact Centers

Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing and Education
THANK YOU