Natural Language Dialogue Systems using AI and ML

Douglas Danforth, Ph.D.
Associate Professor
Department of Obstetrics and Gynecology
Virtual Patient

- Artificial Intelligence - simulate “real” patient
- Conversational – can understand and respond to student questions
- Provide immediate feedback – students can practice multiple times
Virtual Patient

Web-based
chatted conversation

Clinical Skills Center
Spoken conversation
(Dragon)

iPad
Spoken conversation
(Watson)
Natural Language Understanding

- Input type
  - Typed or spoken

- Question identification
  - Rule based vs ML/NLU

- Output
  - TTS vs pre-recorded
**Dialogue Management**

**Input type**

**Chatted (typed)**
- Questions are more succinct
- Conversations shorter
- Typos considerable

**Spoken (STT)**
- Questions are much longer
- Multiple questions in one query
- Conversations longer
- Typos not important
- STT mistranslations important

- “Waht brings you in today?”
- “When did hte pain start?”
- “Anything besieds the back pain?”
- “Tell me more.”

- “So I'd be probing seems to help so that's good and so you started you said this began about 4 to 5 years ago you had another episode.”
- “Are you generally happy with your life now? your job? house?”
Question Identification
Rule based (pattern matching)

- Based on NLP Software ChatScript
- Initial NLP Processing (spell checking, canonization, POS tagging, etc)
- State aware – pronoun resolution
- ~3,000 rules to manage dialogue

ChatScript
Dialogue organized by History of Present Illness, Past Medical History, Family History, Social History

#1 What brings you in today
u: What_brings_you_in_today ("what brings you" "why are you") in today) ^factanswer("I'm hoping you can help me out with a problem I've been having.")
u: ([(when other else] [why what] * ["do for" bring help come] * [today tonight evening afternoon morning day]) ^reuse (What_brings_you_in_today)
u: ([(when other else] "what brings you in") ^reuse (What_brings_you_in_today)
u: ([(when other else] "what brings you in today") ^reuse (What_brings_you_in_today)
u: ([(when other else] "what brings you here") ^reuse (What_brings_you_in_today)

#1 Any other problems
u: Any_other_problems1 (any other problems) ^factanswer("I don't believe so.")
u: (health !medical << [other else] [symptom issue concern problem "bothering you" discuss] >>) ^reuse (Any_other_problems1)
u: (try anything else) ^reuse (Any_other_problems1)

#1 Are you taking any medication for the 1 Schiefcomplaint=1
u: Taking_medication_for_the_1 (are you taking any medication for the Schiefcomplaint?) ^factanswer("I don't take any medications.")
u: ([(how long other] [use take on try] * [alleviate get_rid_of reduce diminish "help with" for] *~2 ["itwords Schiefcomplaint"]) ^reuse (Taking_medication_for_the_1)
u: ([(how long other] [use take on try "no"] * ~medicines * ["itwords Schiefcomplaint"]) ^reuse (Taking_medication_for_the_1)
# Dialogue Accuracy

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dialogue Type</strong></td>
<td>Typed</td>
</tr>
<tr>
<td><strong>Type of Case</strong></td>
<td>Simple</td>
</tr>
<tr>
<td><strong>Type of History</strong></td>
<td>Complete</td>
</tr>
<tr>
<td><strong>Students</strong></td>
<td>First Year (n=21)</td>
</tr>
<tr>
<td><strong>Total questions asked</strong></td>
<td>1396</td>
</tr>
<tr>
<td><strong>Answered correctly</strong></td>
<td>83%</td>
</tr>
<tr>
<td><strong>Answered incorrectly</strong></td>
<td>6.4%</td>
</tr>
<tr>
<td><strong>Not answered</strong></td>
<td>10%</td>
</tr>
</tbody>
</table>
Dialogue accuracy improvements with ML/NLU

- Train Convolutional Neural Network analyzer (CNN)
- Used word - and character-based CNNs to identify and classify input questions
- Binary classifier to choose between CNN and ChatScript

![Combined System Accuracy Chart]

- ChatScript: 75.0%
- CNN: 80.0%
- full: 90.0%
- oracle: 95.0%
Fall 2018 – Mr. Carlos Martinez

- Practice history taking by first year medical students
- HOPI, PMH, Family History, Social History
- ~200 Conversations
- > 12,000 Questions

<table>
<thead>
<tr>
<th></th>
<th>All inputs</th>
<th>Watson Exact (71.4%)</th>
<th>Watson Acceptable (89.5%)</th>
<th>Watson Gibberish (10.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatScript Correct</td>
<td>73.5%</td>
<td>78.3%</td>
<td>76.0%</td>
<td>50.2%</td>
</tr>
<tr>
<td>CNN Correct</td>
<td>72.7%</td>
<td>74%</td>
<td>73.6%</td>
<td>59.6%</td>
</tr>
<tr>
<td>Acceptable Response</td>
<td>76.2%</td>
<td>80.8%</td>
<td>78.7%</td>
<td>53.8%</td>
</tr>
<tr>
<td>Question Not Understood</td>
<td>16.5%</td>
<td>11%</td>
<td>11.7%</td>
<td>35.4%</td>
</tr>
</tbody>
</table>
Fall 2018 – Mr. Carlos Martinez

- Practice history taking by first year medical students
- HOPI, PMH, Family History, Social History
- ~200 Conversations
- > 12,000 Questions

<table>
<thead>
<tr>
<th>All inputs</th>
<th>Watson Exact (71.4%)</th>
<th>Watson Acceptable (89.5%)</th>
<th>Watson Gibberish (10.5%)</th>
</tr>
</thead>
</table>

Wexner Medical Center
Fall 2018 – Mr. Carlos Martinez

- Practice history taking by first year medical students
- HOPI, PMH, Family History, Social History
- ~200 Conversations
- > 12,000 Questions

<table>
<thead>
<tr>
<th></th>
<th>All inputs</th>
<th>Watson Exact (71.4%)</th>
<th>Watson Acceptable (89.5%)</th>
<th>Watson Gibberish (10.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Error Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsoft</td>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson untrained</td>
<td>0.079</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson Acoustic</td>
<td>0.069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson Language</td>
<td>0.058</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson Both</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Natural Language Understanding

Output

Text-to-Speech

- “Voices” limited
- Quality varies
- Limited inflection
- Somewhat slow

Pre-recorded

- Time-consuming to create
- More difficult to deliver
- Difficult to update or modify
- More realistic
Summary

• Using Natural Language Understanding, Virtual Patients can understand, respond, categorize, and “assess” student performance, enabling students to practice their history taking skills and receive immediate feedback.

• Dialogue management can use fairly simply rule based approaches or more sophisticated NLU systems.
  • Rule based systems require hand crafting rules but can work with limited data.
  • NLU systems can be more robust but require significant training data.
  • Hybrid systems can maximize accuracy.

• Depending on the application, modern cloud-based systems may be easier to implement.
Support

Artificial Intelligence based Virtual Reality (VR) Simulation of Provider-patient interaction to enhance cultural competency using a simulated patient with limited English proficiency. Medicaid Equity Simulation Project $455,103 2018 – 2019

Virtual Patients for Medication Assisted Treatment of Opioid Use Disorder. DHHS Health Resources and Services Administration (Supplement) $100,000 2018 – 2019


Virtual Patient Simulations to Assess Data Gathering and Clinical Reasoning National Board of Medical Examiners Edward J. Stemmler Education Research Fund, $149,862, 2012 – 2014

Virtual Patients in the 2012 Curriculum The Ohio State University College of Medicine Innovation Fund, $45,000 2011- 2013

Virtual Reality: A Unique Means to Teach the Reality of the Patient-Centered Medical Home. DHHS Health Resources and Services Administration, $ 1,465,655 2010 – 2015

Development of Virtual Patients using the Second Life platform. Perinatal Resources Inc, $30,000 2008-2011