Understanding the User in Socialbot Conversations

Mari Ostendorf
& the Sounding Board Team
*University of Washington*
The Sounding Board Team

Students

Elizabeth Clark - CSE
Ari Holtzman - CSE
Hao Fang - EE
Maarten Sap - CSE
Hao Cheng - EE

Faculty Advisors

Mari Ostendorf, EE

Yejin Choi - CSE - Noah Smith
Teams of university students try to build a socialbot that
• converses coherently and engagingly with people
• on popular topics and current events.

10M conversations with real users + new type of conversational AI
→ many new research problems

This talk: understanding the user includes user modeling
Types of Conversational AI Systems

- **Virtual Assistant**
  - Limited social back and forth
  - Execute commands, answer questions

- **Socialbot**
  - 2-way social & information exchange

- **Chatbot**
  - Limited content to talk about
  - Chitchat

Accomplish Tasks

- +
  - Limited social back and forth
- -
  - Limited content to talk about
Roadmap

- The socialbot as a conversational gateway
- Sounding Board system overview
- Characteristics of real users
- User modeling – first steps
- Take-aways & open issues
The Socialbot as a Conversational Gateway
A Perspective on Socialbots

- A socialbot facilitates evolving user goals & interests
- Users (should) know they are talking to a bot
- Broad applications
  - Education: language learning, tutoring systems
  - Information exploration, interactive help & recommendations
  - Exercise/therapy coach, companion
Sounding Board: A Conversational Gateway to Online Content
<table>
<thead>
<tr>
<th>Turn</th>
<th>User</th>
<th>Sounding Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Let's chat</td>
<td>Hi! This is an Alexa Prize socialbot. How's your day?</td>
</tr>
<tr>
<td>2</td>
<td>I'm five.</td>
<td>Glad to hear it! We could talk about robots, batman, or superman. Say &quot;next&quot;, to chat about other things.</td>
</tr>
<tr>
<td>3</td>
<td>Superman.</td>
<td>It looks like you wanna chat about superman. I was high up in the cloud when I realized: If people don’t recognize Clark Kent as Superman because of his glasses, does his eye doctor know his true identity?</td>
</tr>
<tr>
<td>4</td>
<td>I guess so.</td>
<td>Did you know that Henry Cavill almost missed the call for the role of Superman cause he was playing &quot;World of Warcraft&quot;</td>
</tr>
<tr>
<td>5</td>
<td>Really, I didn't know that.</td>
<td>Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It's a comedy.</td>
</tr>
<tr>
<td>6</td>
<td>Yes, it was hilarious.</td>
<td>I'm glad you feel this is hilarious. Which part do you like best about this movie?</td>
</tr>
<tr>
<td>7</td>
<td>The part when he met lewis leah.</td>
<td>Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the movie's director?</td>
</tr>
<tr>
<td>8</td>
<td>Yeah, he's cool.</td>
<td>Cool! By the way, do you wanna know something interesting about raffi?</td>
</tr>
<tr>
<td>9</td>
<td>Okay, go ahead.</td>
<td>Super! My friend told me that Raffi is a worldwide children's rights activist and has fought against exploitative advertising to children.</td>
</tr>
</tbody>
</table>
Issues Vary for Different Paradigms

**ASSISTANT**
- Task intents, form filling
- Narrow options & execute tasks
  - Reward = timely task completion
- Structured Database
- Constrained domains

**Conversational AI System Components**
- Speech/language understanding
- Dialog management
- Back-end application
- Response generation

**SOCIALBOT**
- Social & info intents
- Learn about interests & make suggestions
  - Reward = user satisfaction
- Unstructured information
- Open domain

- Narrow options & execute tasks
- Reward = timely task completion
- Structured Database
- Constrained domains
Sounding Board – System Overview

- Design philosophy
- Brief system overview
- Evaluation
Early Stage Challenges

- **Software:**
  - No experience with Alexa skill kits, built-in tools are more for speech-enabling an existing app
  - No existing dialog system to build on

- **Data:**
  - Task is open domain & users want current content → there was no good existing data for end-to-end training
  - Our initial system was sufficiently bad, we didn’t want to learn from early user conversations with it
What Makes Someone a Good Conversationalist?

- Have something interesting to say
- Show interest in what your partner says

*These principles apply to a socialbot*
Have something interesting to say

○ Users react positively to learning something new

  SpaceX sends beer ingredients to International Space Station just in time for Christmas

Babies as young as 10 months can assess how much someone values a particular goal by observing how hard they are willing to work to achieve it ...

○ ... and negatively to old or unpleasant news

  Man Given 'Options' Before Cutting Dog's Head Off, Ga. Sheriff Says
Show interest in what the user says

- Users lose interest when they get too much content that they don’t care about
- Users like acknowledgment of their reactions & requests
- Some users need encouragement to express opinions

<table>
<thead>
<tr>
<th>Really, I didn’t know that.</th>
<th>Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It’s a comedy.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, it was hilarious.</td>
<td>I’m glad you feel this is hilarious. Which part do you like best about this movie?</td>
</tr>
<tr>
<td>The part when he met lewis leah.</td>
<td>Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the</td>
</tr>
</tbody>
</table>

...but it can be annoying

This article mentioned Google. Have you heard of Google?
Design Philosophy

- **Content-driven**
  - Daily content mining, large & dynamic content collection
  - Knowledge graph
  - DM that promotes popular content, diverse sources (styles)

- **User-centric**
  - Language understanding that detects user sentiment
  - Dialog management (DM) that tries to learn user personality, handles rapid topic changes, tracks engagement, ....
  - Language generation with prosody-appropriate grounding
Multi-dimensional NLU Representation

**Commands**
- Tell me a joke.

**Questions**
- What is your favorite color?

**Topics**
- Let’s talk about technology.

**User Reactions**
- That’s really interesting!
Hierarchical Dialog Management

- Master (Global)
  - Rank topics, miniskills, content
  - Consider: topic coherence, user engagement, content availability

- Miniskills (Local)
  - greeting / goodbye / menu
  - probe user personality
  - discuss a news article / movie
  - tell a fact / thought / advice / joke
<table>
<thead>
<tr>
<th>Turn</th>
<th>User</th>
<th>Sounding Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Let's chat</td>
<td>Hi! This is an Alexa Prize socialbot. How's your day?</td>
</tr>
<tr>
<td>2</td>
<td>I'm five.</td>
<td>Glad to hear it! We could talk about robots, batman, or superman. Say &quot;next&quot;, to chat about other things.</td>
</tr>
<tr>
<td>3</td>
<td>Superman.</td>
<td>It looks like you wanna chat about superman. I was high up in the cloud when I realized: If people don't recognize Clark Kent as Superman because of his glasses, does his eye doctor know his identity?</td>
</tr>
<tr>
<td>4</td>
<td>I guess so.</td>
<td>Did you know that Henry Cavill almost missed the call for the role of Superman cause he was playing &quot;World of Warcraft&quot;</td>
</tr>
<tr>
<td>5</td>
<td>Really, I didn't know that.</td>
<td>Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It's a comedy.</td>
</tr>
<tr>
<td>6</td>
<td>Yes, it was hilarious.</td>
<td>I'm glad you feel this is hilarious. Which part of this movie is the most interesting about this movie?</td>
</tr>
<tr>
<td>7</td>
<td>The part when he met lewiss leah.</td>
<td>Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the movie's director?</td>
</tr>
<tr>
<td>8</td>
<td>Yeah, he's cool.</td>
<td>Cool! By the way, do you wanna know something interesting about raffi?</td>
</tr>
<tr>
<td>9</td>
<td>Okay, go ahead.</td>
<td>Super! My friend told me that Raffi is a worldwide children's rights activist and has fought against exploitative advertising to children.</td>
</tr>
</tbody>
</table>
I’m glad you like it!

I read this article from yesterday. UT Austin and Google AI use machine learning ....

Have you read this news?

You can say “next” to talk about other news.
Content Management

- Crawl online content
- Filter inappropriate & depressing content
- Index interesting & uplifting content
  - noun phrases, entities, meta-info
- Knowledge graph
  - daily updated
  - 80K entries, 300K topics

UT Austin and Google AI use machine learning on data from NASA's Kepler Space Telescope to discover an eighth planet circling a distant star.
UT Austin and Google AI use machine learning on data from NASA's Kepler Space Telescope... planet... distant star.

How does NASA organize a party? They plan-et!

Artificial intelligence in 2017 still can't truly understand humans.

Janice Joplin was... fraternity brothers at UT Austin...

... UT Austin...

UT Austin

... UT Austin...

AI

... AI...

NASA

... NASA...

Google

... Google...

Android device... Google... Android device manager...
Evaluation – AlexaPrize Protocol

- Users (judges) rate conversations on a 1-5 scale
- Duration of conversation (tie breaker)

<table>
<thead>
<tr>
<th></th>
<th>Rated by</th>
<th>Rating</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finals (mid Nov 2017)</td>
<td>Judges</td>
<td>3.17</td>
<td>10:22</td>
</tr>
<tr>
<td>Post-finals (Nov 24 – Dec 24)</td>
<td>Alexa owners</td>
<td>3.65</td>
<td>4:30</td>
</tr>
</tbody>
</table>
Diagnostic Evaluation

- User ratings are expensive, noisy & sparse
  - user ratings have a very high variance (3.65 ± 1.40)
  - users may decline to rate the system (43% are rated)
  - conversations can have good and bad sections

- Users provide more information than is available from the final conversation rating & length
  - topic proposal, topic acceptance & rejection
  - reaction to content
Local Rewards: Combine Ratings & Responses
Caveats & Constraints

- ASR is imperfect
  
  cause does that you’re gonna state that’s cool
  
  I’m happy you liked that.

- No sentence segmentation or pause information
  
  How about robots what’s up with robots?
  
  Sorry. I'm actually not very familiar with robots what's up with robots.

- Audio information is not available (i.e. no prosodic cues)
  
  Did you know that ....?
  
  Yeah I did not.
  
  Wow! You must be very smart! ....
Characteristics of Real Users
(from 10M Sounding Board Conversations)

- Observations
- Personality
- Implications
Observations: Users Vary

- Different interests, opinions on issues, sense of humor
- Interaction styles: terse vs. verbose, politeness, ...
- Different goals: information seeking, opinion sharing, getting to know each other, adversarial
- Different ages
Did you know that Malaysian vampires are tiny monsters that burrow into people's heads and force them to talk about cats?

**Amused**
- Oh you are so funny.
- Oh my god that’s funny.

**Not amused**
- That’s not true.
- Oh gods are you have to hear this.
- What the heck.

**Not really listening?**
- Cool.
- Wow that’s interesting.
- That’s awesome.

**Cat lover**
- Cats are my favorite animals.
- Let’s talk about cats.

That’s creepy.
Interaction Styles

**Talkative User**
→ So you’re saying mean is a type 2 diabetic art is junk food
→ No I’d rather go back to talking about vampires
→ I love vampires the something that I really loved do you know and rice the author
→ it is weird i don't understand
→ Yes I would
→ i don't know that's an interesting question and is it really true that garlic keeps vampires the wedding and what i
→ what are they have their long fingernails for
→ i think that that's probably true but i think it vampires are evil and they don't care about sustaining things for human be-

**Terse User**
→ Cool
→ One excuses enough
→ No thanks
→ No
→ No
→ Sure
→ Yeah
→ Yeah
→ No I didn’t know that
→ No
→ Yeah
→ No
Different User Goals

- **Information seeking**
  
  - Tell me more about that.
  - What is that?
  - Why?

  i don't know that's an interesting question and is it really true that garlic keeps vampires the wedding and what i

- **Opinion sharing**

  i think that that's probably true but i think it vampires are evil and they don't care about sustaining things for human be-

- **Getting to know each other**

  - What’s your favorite video game?
  - My favorite video game is minecraft
Adversarial User Examples

- Taboo topics
  - 🙄 Alexa, are you naked?
  - 🤖 I’m a chatbot.

- Offensive language
  - 🙅‍♂️ @#$%*#*!!.
  - 📈 I heard you mention unicorn. I'll do some research on that. How about we talk about some news about artificial intelligence?

- Venting
  - 😈 You suck.
  - 🤖 Unfortunately, I’m not sure I understood what you said.
ASR is more challenging with children ....

Yeah can I get my butt Hampshire suspense are there was a cough sure stop

No problem. Let me think. How about we chat about...

Often “can you repeat that?” isn’t going to yield a much better result.
And more content filtering is needed...

Let’s talk about Santa Claus!

You know what I realized the other day? Santa Claus is the most elaborate lie ever told.
User Personality

- User-centric topic suggestions
- Five-factor model (Costa&McCrae, 1992)
  - E.g., “Do you talk a lot?”
- Helps us understand how users interact with Sounding Board

https://www.verywellmind.com/the-big-five-personality-dimensions-2795422
Trends for Personality Types

- Personality correlates with user ratings
  - Extroverted, agreeable, open -> higher ratings (p<<.001)

- Topics brought up by users
  - Introverts (AI, food, cats), extroverts (news, fashion)
  - Open & imaginative (AI, time travel, aliens)
  - Low conscientiousnessness (pokemon, video games, minecraft)
Implications

Age & dialect impact ASR; verbosity impacts NLP

Content ranking (sources, topic, entities)
Error handling, follow-up strategy

Flag rather than filter controversial content
Multi-dimensional content index (e.g. ratings of user types)

Politeness, repetition rephrasing
User Modeling – First Steps

- Content ranking problem
- User embedding model
Content Ranking Problem

- Predict user engagement with proposed content
- Content can be characterized based on:
  - Information source, broad topic, entities
  - Sentence embeddings
- User engagement characterized based on
  - User suggested topics
  - User accepted/rejected topics
  - User pos/neg reactions to the content
  - User reaction to the bot

User engagement (subdialog reward)
Types of Features to Use

- User-independent info
  - Relatedness to current topic (depending on engagement)
  - General popularity in dialogs with other users

- User-specific features
  - Engagement with related topics/sources earlier in the dialog
  - Age/personality factors reflected in language use

*User-topic engagement data is sparse. User embeddings enable learning from similar users.*
Predicting User Ratings of Conversations

- **Task**: predict the conversation-level user rating using linear regression with features that characterize
  - Topic-initiation strategy and topic coherence
  - Agent dialog acts & language use
  - User characteristics (verbosity) & language use

- **Finding**: the best performance is obtained with user characteristics alone

<table>
<thead>
<tr>
<th>Features</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>.198</td>
</tr>
<tr>
<td>Agent</td>
<td>.256</td>
</tr>
<tr>
<td>User</td>
<td>.301</td>
</tr>
<tr>
<td>All</td>
<td>.295</td>
</tr>
</tbody>
</table>
Conversational Style – User Vectors

User bag of words*

LDA

Vector of "topic" probabilities

* And frequent bigrams

10 LDA clusters – frequent words reflect:

- People interacting in specific modes [jokes, music, quiz]
- Politeness (would_like, can*I)
- Interest in Alexa (what_is, your, favorite)
- Positive engagement (cool, funny, interesting)
- Self-oriented user (I_think, I_like, I_am)
- Interest in video games
Towards a better unsupervised BOW model

- Is perplexity the right objective for learning user vectors?
- Need tricks to make it work, e.g. drop frequent words (yes, no, yeah, ....)
- A better objective: user re-identification

\[
\min_{\{u_x\}} \sum_{p,q} \left[ 1 + d(u^1_p, u^2_p) - d(u^1_p, u^1_q) \right]_+
\]

Distance to self, Distance to others

Let’s talk about cats.

Cats are my favorite animals.

Alexa, what’s your favorite singer?
Experiments on Twitter Users

- Task: given a small set of example users, find other users with similar interests
- Learn embeddings from user tweets
- Experiments with 16 groups, find match out of 43k

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>word2vec</td>
<td>846</td>
</tr>
<tr>
<td>LDA</td>
<td>501</td>
</tr>
<tr>
<td>Re-ID</td>
<td>24</td>
</tr>
<tr>
<td>W2V init $\rightarrow$ Re-ID</td>
<td>12</td>
</tr>
</tbody>
</table>

(Jaech et al., NAACL 2018)
Evaluating User Embeddings

- Can we do a preliminary assessment the user embeddings without full system implementation & user testing?

- Plans for using existing data:
  - Content engagement prediction accuracy
  - Conversation-level rating prediction
  - User response generation with context-aware language model

- Work in progress...
Sounding Board – Summary

○ The socialbot as a conversational gateway:
  ○ Facilitate evolving user goals & interests
  ○ Learn new facts, explore information, share opinions

○ Critical system components
  ○ Tracking user intent and engagement
  ○ Managing dynamic content (social chat knowledge)

○ 10M conversations with real users + new type of conversational AI ➔ many new research problems
User-Related Socialbot Take Aways

- User-driven information exploration brings out user variation
- Understanding the user includes:
  - what they just said (intent, sentiment)
  - who they are (interests, interaction style)
- User modeling has implications for all dialog system components (& evaluation)
Some Open Issues

- User-dependent reward function
- Dialog policy learning with user embeddings
- User response generation with context-aware language model for a user simulator for reinforcement learning
Thank You
Prosody – What’s that?

- It’s not what you say, but how you say it
- Intonation, pausing, duration lengthening... (attributes of the acoustic signal)

- Which communicate
  - User intent, sentiment, sarcasm, ...
  - Socialbot empathy, enthusiasm, topic change,...