Designing Bots, Virtual Humans, & Other Systems that Hold up their End of the Conversation

Justine Cassell

Carnegie Mellon University & ISIR / La Sorbonne
Chaire Blaise Pascale & Chaire Sorbonne
Mai 2018
The ineffable quality of Rapport in learning

Children who report more rapport are more likely to learn from the virtual peer.
Search Engine vs. Conversation

Justine: “OK Google, I love Manchester United”

Google: Manchester United Football Club is a professional football club
Based in Old Trafford, Greater Manchester, England, that competes
in the Premier League, the top flight of English Football

Justine: “I love Manchester United”

Friend.: “No way! Arsenal wipes the floor with those Red Devils!”

Socially-Aware Robot Asst: “No way! Arsenal wipes the floor with those Red Devils!”
Motivation for Socially-Aware Bots

1. People pursue *multiple conversational goals* in every conversation & expect the same from their interlocutors. To put people at ease, and increase relationship strength, we must understand the *propositional, interactional & interpersonal* functions of conversation.

2. People change interaction styles over time. We must increasingly *manage long-term interactions* with people by changing interaction style in a way that evokes increasing loyalty, rapport and trust.
Rapport improves task performance

Surveys
– Survey respondents gave higher quality answers if they felt rapport with interviewer (Berg (1989)).

Health
– Physicians who build rapport during trial interviews enroll more participants (Albrecht et al., 1999).

Sales
– Rapport with sales staff leads to increased likelihood of purchasing goods/service (Brooks, 1989).
– Customers show increased trust and disclosure when rapport is maintained with sales staff (LaBahn, 1996).
Analyze
Analysis of Rapport

*Thin-slice* (Ambady & Rosenthal, 1992) judgments of every 30 second video segment) presented to 3 annotators in randomized order

**IRR**

- “Consensus” measured by Intra-Class Correlation (single measure): 0.37
- “Consistency” measured by Cronbach alpha: 0.68
- *Inverse-based bias correction* (Kruger et al., 2014) was used to mitigate rater bias & pick single rapport rating for each 30 second video segment.
Implement
Evaluate

Total Interactions:
250+ sessions

Current Dataset:
120 sessions (totally 10+ hours)
- Mean duration = approx. 5 min.
- SD = 69.00 seconds

SARA, 00:13.7, I'm so glad to have the opportunity to meet you in person, 00:16.7
User, 00:18.7, hi how are you [doing, 00:19.2
SARA, 00:19.3, [I'm] in Davos to assist attendees but to be honest I am a little anxious about it it's a big deal to be invited to Davos is it your first time here, 00:28.7
User, 00:29.1, it is and I'm a little anxious too, 00:31.4
SARA, 00:32.7, I found there's a lot to take in I hope that it isn't overwhelming you, 00:37.2
User, 00:37.9, it is overwhelming a little bit, 00:39.8
# Summary of Evaluation Metrics

## Interaction
- Total Time
- System Speaking Time
- User Speaking Time
- System’s Response Time
- User’s Response Time
- System’s Interruptions
- User’s Interruptions

## Verbal
- System’s Intentions
- User’s Intentions
- System’s Conversational Strategies
- User’s Conversational Strategies

## Task
- Session Rec. Acceptance
- Person Rec. Acceptance

## Interpersonal
- Rapport Score
- Mutual Attentiveness
- Coordination
- Positivity
Goal of Socially Aware Systems

Development of a bot that manages *interpersonal rapport* (relationship strength) with users over interactions across time, as well as managing propositional and interactional goals, *in order to improve task performance*.

*Automatically recognize* rapport-managing conversational strategies from *verbal, visual and vocal* modalities of speaker and interlocutor, both within the individual and in the dyad.
Goal of Building Socially Aware Systems

**Theoretical:** Understand the nature of rapport in greater detail, by correlating with associated observable verbal (conversational strategies, vocal (voice quality) and visual (non-verbal) cues

**Methodological:** Leverage this understanding to automatically recognize rapport-building strategies by leveraging and developing statistical machine learning techniques
Ineffective Conversation
(don’t do this with agents)
Intimate Conversation
(don’t do this with agents)
Agent Model of Rapport must be:

1. Dyadic,
2. Multi-level: differentiate between observable signals & underlying psychological states,
3. Sensitive to effect of time
4. Cross-Modal
Data- & Theory-Driven Model of Rapport Management

Strategy
- Initiate mutual self-disclosure
- Refer to shared experience
- Disclose topic-related intimate personal information
- Reciprocal Appreciation
- Embarrassed Laughter
- Praise
- Negative Self-disclosure
- Acknowledge

Goal (Level 1)
- Rapport Enhancement
- Rapport Maintenance

Sub-goal (Level 2)
- Mutual Attentiviness
- Face Management
- Coordination

Sub-goal (Level 3)
- Maintain attention to others
- Support & Appreciate other’s "true-self"
- Enhance other’s positive face
- Act predictably
- *Index Commonality

Notes:
- Adhere to sociocultural or interpersonal norm (including relational definition)
- Reciprocate previous action (ex. Respond to other's self-disclosure)
- *Violate sociocultural norm to match interpersonal norm
## Conversational Strategies

<table>
<thead>
<tr>
<th>Conversational Strategies</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN (Violation of Social Norm)</td>
<td>“man you take forever to write”</td>
</tr>
<tr>
<td>SD (Self Disclosure)</td>
<td>“I hate math”</td>
</tr>
<tr>
<td>PR (Praise)</td>
<td>“well done”</td>
</tr>
<tr>
<td>SE (Reference to Shared Experience)</td>
<td>“I shared m&amp;m's with you last time”</td>
</tr>
<tr>
<td>BC (Back Channel)</td>
<td>“yup”</td>
</tr>
<tr>
<td>QE (Question Eliciting SD)</td>
<td>“are you an atheist”</td>
</tr>
</tbody>
</table>
Data- & Theory-Driven process model

**Rapport building**
- **Rapport**
  - **Friends**
    - Violate sociocultural norms to fit interlocutor’s behavior expectation
    - Mark in- and out-group
    - Learn behavior expectations
  - **Strangers**
    - Politeness according to sociocultural norms
    - Learn behavior expectations

**Reasoning**
- **Shared personal knowledge**
- **Self-disclosure**
- **Reciprocity**
- **Attributed Process**

**Strategies**
- **Discloser**
  - Trust
  - Social Validation
  - Social Control
  - Reciprocity of Self-disclosure

**Reciprocity condition**
- **Goal**
- **Norm Recip.**
- **Conv. Norm**
- **Relationship Stage (T1/T2)**

**Estimate**
- Appropriate social distance at right time

**Generate**
- Imagined interaction:
  - Min. vulnerability
  - Influence subs. attr. process

**Match “likeness”** and rel. stage (T1/T2)

**Match “likeness”** and rel. stage (T1/T2)

**Justine Cassell**

**Carnegie Mellon University**
Data-Driven: Temporal association rules

Form of temporal rules

“If event A happens at time t, there is 50% chance of event B happening between time t+3 to t+5”
Example: Friend in high rapport

**Tutor:** Sweeney you can't do that, that's the whole point {smile} [Violation of Social Norm]

**Tutee:** I hate you. I'll probably never never do that [Reciprocate Social Norm Violation]

**Tutor:** Sweeney that's why I'm tutoring you {smile}

**Tutee:** You're so oh my gosh {smile}. We never did that ever [Violation of Social Norm]

**Tutor:** {smile} What'd you say?
Temporal association rules: Strangers

Example: Stranger in low rapport

Tutee: divide oh this is so hard let me guess: 11

[Violation of Social Norm]

Tutor: you know

Tutee: 6

Tutor: next problem is is exactly the same {smile}: over 11 equals, 11 x over 11

Tutee: I don't need your help; [Violation of Social Norm]

Tutor: {Overlap} That is seriously like exactly the same.
Socially-Aware Agent Architecture
Socially-Aware Agent Architecture
Some Applications:
Rapport-Aware Peer Tutor (RAPT)
RAPT
WoZ System Architecture

**Learner Model**
- Problem Selection

**Problem-Solving Interface**
- Step-level problem correctness

**Tutoring Move**
- Finite State Machine
- Agent’s next tutoring move

**Tutoring and Social utterance database**

**Mic**
- Speech Signal

**ASR**
- Text
- Acoustic Features

**OpenSmile**
- Social Conversational Strategy Classifier
- Rapport Score Estimator

**Camera**
- Images (640x480, 30fps)

**OpenFace**
- FACS
- Recorded Video/Audio of User

**Social Reasoner**
- WoZ: Task NLU and Task Reasoner
- User’s social utterance label

**Condition 1:**
- Social behaviors gradually increasing over time

**Condition 2:**
- Social behavior chosen based on current rapport level, previous social moves, and users’ nonverbal behaviors

**Socially-inflected task utterance**

**Agent’s selected task utterance**

**BEAT**
- Behavioral Markup Language (BML)

**Smart Body**
- Unity 3D (agent)

**TTS**
- Synthetic Speech Markup Language (SSML)

**Screen Recorder**

**Pre- and Post-conditions for Algebra task**

**RAPT WoZ System Architecture**

**Justine Cassell**

**Carnegie Mellon University**
Evaluation: rule-based vs. adaptive

control condition: (fixed heuristics for social dialogue usage)

**Praise**
Decreasing in frequency [Kumar et al., 2010]

**Self-Disclosure**
Gradually increasing in frequency and intimacy [Ogan, 2011; Bickmore and Schulman, 2010]

**Questions eliciting self-disclosure**
Gradually increasing in breadth and depth of topics [Altman and Taylor, 1973]

**References to shared experiences**
Gradually increasing frequency [Kumar et al., 2010; Cassell and Bickmore, 2003]

**Violation of social norms**
Use only after a given threshold in number of turns or elapsed time [Ogan et al., 2012]

**Indirectness**
Decreasing in frequency [Madaio et al., 2017]
Evaluation: rule-based vs. adaptive

**Experimental condition:** (adaptive usage of social dialogue)

**Based on:**
- Current rapport state
- Changes in rapport state (increasing, decreasing, maintaining)
- User’s social behavior (self-disclosure, violation of social norms, etc)
- Agent’s previous social behavior
- User’s nonverbal behavior (smiling, nodding, gaze patterns)
- Agent’s previous tutoring behaviors (feedback, questions, explanations, etc)
Application: Mobile front-end to apps

Hi, I'm your InMind agent.
Justine Cassell

with Yahoo InMind Team

Carnegie Mellon University
Integrated InMind Dialog Architecture

General architecture

Multi User Framework

NLU

User intent + entities

User conv. strategy

Conversational Classifier

Facial Features

Non-verbal analysis

Non-verbal behavior

Open Face

Rapport Estimator

Rapport score

Dialogue Manager

Movies

System intention +
User frame +
System conv. strategy

Utterance realization
+ non-verbal realization

BEAT

System intention +
User frame +
System conv. strategy

Utterance realization

Google TTS

Capture stream

Google ASR

Video Streaming

Dialogue management

Signal understanding

Front-end application

Multi-user management

Justine Cassell

Carnegie Mellon University

Long-Term Memory

NLG

Utterance realization
SARA: Socially Aware Robot Assistant
WHAT SARA UNDERSTANDS

OpenFace

User-Sara Rapport 6 / 7

CAMERA FEED LIVE

HEAD TRACKING...

User Conversation Strategy

SELF DISCLOSURE (SD) 88%

SHARED EXPERIENCE (SE) 72%

PRAISE (PR) 78%

VIOLATE SOCIAL NORM (VSN) 50%

FOLLOW SOCIAL NORM (FSN) 50%

Smile NO

Eye Gaze NO

Head Nod NO
Tracking Facial Movements

OpenFace: L.P. Morency

Justine Cassell

Carnegie Mellon University
this is my favorite part let me look this up one minute
SARA: Socially-Aware Robot Assistant at Davos
Methodology

Start here

Study

Theorize & Model

Build

Test

Collect Natural data

Realize gaps in understanding

Build formal models

Implement system on the basis of model

Design evaluation of use

Start here

Collect Natural data

Realize gaps in understanding

Build formal models

Implement system on the basis of model

Design evaluation of use
New SARA Framework

Client Side

- Kinect
- Touch Sensor
- Speech Recognition
- OpenFace
- OpenSmile
- Unity
- TTS

Server Side

- Language Understanding
- CSC
- Rapport Estimator
- User Model
- Social BEAT
- Incremental NLG
- Content Planner

Surface

- MSG_ASR
- MSG_NBG
- MSG_CSC
- MSG_RPT
- MSG_UM
- MSG_SR
- MSG_REC
- MSG_CP
- MSG_NLG
- MSG_BEAT
- MSG_INC

MUF

Reasoner (Task + Social Reasoner)

Justine Cassell

Carnegie Mellon University
Indirectness Strategy Classifier

- Corpus
  - RAPT 2013: indirectness annotation of peer-tutoring corpus
  - ConLL 2010 shared task on uncertainty detection
    - Wikipedia dataset (Wikipedia articles)
    - BioScope dataset (abstracts and articles from biomedical literature)

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apology</td>
<td>Apologies used to soften direct speech acts</td>
<td>Sorry, its negative 2.</td>
<td>7.7%</td>
</tr>
<tr>
<td>Qualifiers</td>
<td>Qualifying words for reducing intensity or certainty</td>
<td>You just add 5 to both sides.</td>
<td>66.1%</td>
</tr>
<tr>
<td>Extenders</td>
<td>Indicating uncertainty by referring to vague categories</td>
<td>You have to multiply and stuff.</td>
<td>3.6%</td>
</tr>
<tr>
<td>Subjectivizer</td>
<td>Making an utterance seem more subjective to reduce intensity</td>
<td>I think you divide by 3 here.</td>
<td>22.6%</td>
</tr>
</tbody>
</table>

IN (Indirect Delivery) “so I think what I'm gonna do is make that 15 minus 3 a 12”
Architecture: Indirectness Classifier

Justine Cassell

SARA Receptionist
With thanks for generous funding and support to the following organizations, and to the students and staff of the ArticuLab