The Language of Persuasion, Negotiation and Trust

José Lopes and Helen Hastie

SemDial 2021, Postdam

Trust in HRI

- Trust is a hot topic in HRI (Kok and Soh, 2020)
  - Systems need to be able to react and mitigate against over-trust and distrust
- Two types of trust: Conditional and Unconditional (Jones and George, 1998)
  - Conditional trust: the minimum level of trust to facilitate social and economic exchanges toward a common goal
  - Unconditional trust: characterises an experience of trust that starts when individuals abandon the "pretense" of suspending belief
- Trust evolves through interaction (Rempel et al., 1995)
- Trust can fall rapidly (e.g. following an error) (Nesset et al., 2021)
- **How to measure trust in HRI?**

- Bing Cai Kok and Harold Soh. 2020. Trust in robots: Challenges and opportunities. Current Robotics Reports.
- Gareth R. Jones and Jennifer M. George. 1998. The experience and evolution of trust: Implications for cooperation and teamwork. The Academy of Management Review.
- John K Rempel, John G Holmes, and Mark P Zanna.1985. Trust in close relationships. Journal of personality and social psychology.
- Birthe Nesset, David A. Robb, José Lopes, and Helen Hastie. Transparency in hri: Trust and decision making in the face of robot errors. In HRI 21.

Measuring trust

- Problem: designing studies to discover trust signals in language and interaction is difficult
  - Trust requires vulnerability (Rosseau et al, 1998), e.g. In scenarios requiring financial/health decisions
- Current solution: questionnaires to meaure trust in interaction, e.g.
  - Post-interaction (Schaefer, 2013; Jian et al., 2000; Ullman and Malle, 2019)
  - During interaction (Khalid et al, 2019)
- “Success in such cases (financial/health) a reliable approximation of success in terms of persuasion, negotiation and consequently trust and trustworthiness” (Camerer, 2011)
- **Can we use language to measure levels of trust in real-time through proxies?**

- Colin F Camerer. 2011.Behavioral game theory: Experiments in strategic interaction. Princeton University Press.
- Kristin Schaefer. 2013.The perception and measurement of human-robot trust. Ph.D. thesis.
- Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1):53–71.
- Daniel Ullman and Bertram F. Malle. 2018. What does it mean to trust a robot? Steps toward a multidimensional measure of trust. In Companion of HRI ’18.
- Halimahtun Khalid, Wei Shiung Liew, Bin Sheng Voong, and Martin Helander. 2019. Creativity in measuring trust in human-robot interaction using interactive dialogs. In IEA 2018.
- Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. 1998. Not so different after all: Across-discipline view of trust. Academy of management review

Social Signals and Trust

- Predicting trustworthy behaviour is highly connected to the availability of non-verbal cues (DeSteno et al, 2012):
  - Leaning forward and head nods
- Participants are more willing to follow the empathic agent advice (Lisetti et al, 2013)
- Smiling agents have been perceived as more trustworthy, knowledgeable and appealing (Torre et al, 2018)
- Non-verbal immediacy, reinforced with eye gaze, arm gestures and proximity, increases communicative effectiveness, perceived competence and trustworthiness (Chidabaram et al, 2012)

- David DeSteno, Cynthia Breazeal, Robert H. Frank ,David Pizarro, Jolie Baumann, Leah Dickens, and Jin Joo Lee. 2012. Detecting the trustworthiness of novel partners in economic exchange .Psychological Science, 23(12):1549–1556.
- Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! an empathic virtual agent delivers behavior change health interventions. ACM Trans. Manage. Inf. Syst.
- Ilaria Torre, Emma Carrigan, Killian McCabe, Rachel McDonnell, and Naomi Harte. 2018. Survival at the museum: A cooperation experiment with emotionally expressive virtual characters. In ICMI ’18. Chidambaram, V., Chiang, Y. H., & Mutlu, B. (2012).Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In HRI’12.

Language and Trust

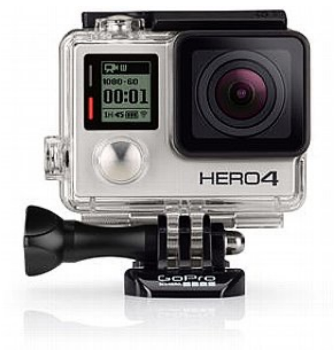
- Deceptive news detection (Rashkin et al, 2017):
  - First-person and second person pronouns are used more in less reliable or deceptive news texts
  - Trusted sources are more likely to use assertive words and less likely to use hedging words
- Providing personal opinions (Newman et al, 2003):
  - Fewer self-references in people telling lies
- Dilemma investment game using instant messaging (Scissors et al., 2008):
  - Higher levels of mimicry were present in high-trusting pairs than low-trusting pairs

- Lauren E. Scissors, Alastair J. Gill, and Darren Gergle. 2008. Linguistic mimicry and trust in text-based CMC. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW'08.
- Matthew L. Newman, James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 2003. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin.
- Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017.Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the EMNLP 2017.

Data

- Investigated linguistic cues in:
  - Negotiation (He et al, 2018): Craigslist Bargain dataset
    - 6555 negotiation dialogues

Craigslist Bargain Dataset



Price: $265
Seller target: $265
Buyer target: $243

BUYER: hi there
SELLER: Good Day!
BUYER: how are you today?
SELLER: I'm well. Broke my arm and can't use my go pro for a while.
BUYER: oh geeze, im sorry to here that.
SELLER: I can't use it but maybe you're interested?
BUYER: Yes, A go pro is something I have been interested in for a while, how does $243 for it sound?
SELLER: That will work for me. I can't use it anyway!

Data

- Investigated linguistic cues in:
  - Negotiation (He et al, 2018): Craigslist Bargain dataset
    - 6555 negotiation dialogues
  - Persuasion (Wang et al, 2019): Persuasion for Good dataset
    - 1017 persuasion dialogues

Persuasion for Good Dataset

Donation Persuader: 0.0
Donation Persuadee: 0.0
Intended donation Persuadee: 0.2

PERSUADER: Good morning. How are you doing today?
PERSUADEE: Hi. I am doing good. How about you?
PERSUADER: I'm doing pretty good for a Tuesday morning.
PERSUADEE: Haha. Same here, but it really feels like a Monday.
PERSUADER: Ugh yes it does!
PERSUADEE: I can not believe how warm it is already.
(...)
PERSUADER: We do. I guess I should get into what this chat is supposed to be about. Have you heard of the Charity Save The Children?
PERSUADEE: I have heard about them. What do you like about them?
PERSUADER: I like that they're committed to helping children in need. They're very transparent in their work and do great things to help children in underprivileged countries.
PERSUADEE: Yes, I also like what they do. They are a great organization.
PERSUADER: I'm planning on donating most of my earnings today. Would you like to donate as well?
PERSUADEE: I would like to dotate $0.20. Would that help?
PERSUADER: Yes it would. Any little bit helps. Thank you for your donation!

Data and Research Goals

- Investigated linguistic cues in:
  - Negotiation (He et al, 2018): Craigslist Bargain dataset
    - 6555 negotiation dialogues
  - Persuasion (Wang et al, 2019): Persuasion for Good dataset
    - 1017 persuasion dialogues
- Both cases require a trustful relationship between interlocutors to be successful
- Research goals:
  - Identify linguistic indicators of trustworthiness in successful interactions
  - Identify role-specific linguistic indicators
  - Use data-driven methods to identify the outcome of the dialogue

- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In EMNLP 2018.
- Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In ACL 2019.

Method

- Features (following De Kock and Vlachos, 2021):
  - Politeness (Zhang et al, 2018) [POLI]
  - Collaboration (Niculae and Danescu-Niculescu-Mizil, 2016) [COLL]
  - LIWC (Pennebaker, 2001)
- For each dialogue features, we took include the:
  - Average
  - Gradient of a straight line fit of the feature value throughout the conversation (fit)
- Using Linear Regression (LR), we
  - Predict the outcome of the dialogue
  - Identify most relevant features

Christine De Kock and Andreas Vlachos. 2021. I beg to differ: A study of constructive disagreement in on-line conversations. In EACL 2021.
Justine Zhang, Jonathan Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Dario Taraborelli, and Nithum Thain. 2018. Conversations gone awry: Detecting early signs of conversational failure. In ACL 2018.
Vlad Niculae and Cristian Danescu-Niculescu-Mizil. 2016. Conversational markers of constructive discussions. In NAACL 2016.
James W Pennebaker. 2001. Linguistic inquiry and word count: LIWC 2001.

List of (selected) Features

Feature	Definition	Type
n_repeated_content	number of repeated content words in consecutive turns	COLL
agree	whether there is an agreement expression	COLL
disagree	whether there is a disagreement expression	COLL
n_repeated_stop	number of repeated stop words in consecutive turns	COLL
n_adopted_w_hedge	number of words re-used from hedges lexicon	LIWC
n_words	number of words per utterance	LIWC
geo	number of usages of words from the geographic terms lexicon	LIWC
hedge	number of words from the hedges lexicon	LIWC
please_start	if utterance starts with please	Politeness
apologising	if utterance contains apologetic words	Politeness
2nd_person	if utterance contains second person words	Politeness
direct_question	if utterance starts with what, why, who or how	Politeness
gratitude	if utterance contains gratitude words	Politeness
has_positive	if utterance as positive words	Politeness
1st_person	if utterance has first person pronouns	Politeness
has_negative	if utterance has negative words	Politeness
2nd_person_start	if uttreance starts with a second person pronoun	Politeness

Table 7: Linguistic indicators reference.

Results

- Craigslist Bargain

Features	**Accuracy**	**F1-score**	R^2	**Top-5 features**
Baseline Majority	0.769	0.869	-	-
COLL + LIWC + Politeness	**0.847**	**0.904**	**0.489**	-fit_n_words -avg_has_negative +avg_has_positive +avg_gratitude -avg_apologising
Buyer+Seller Features	**0.857**	**0.910**	-0.519	-avg_seller_1st_person -avg_buyer_2nd_person_start +fit_seller_apologizing +fit_buyer_please_start +fit_seller_n_adopted_w_hedge

Table 1: Accuracy, F1-score and McFadden's R^2 for predicting negotiation success in the Craigslist Bargain dataset. The speaker-independent features are in the top part of the table. Speaker-dependent features are in the bottom part of the table where the buyer and seller features include LIWC+Politeness separated out and calculated per role. The top-5 features are sorted according the absolute coefficient value.

Results

Craigslist Bargain, role-dependent

Buyer Features	0.832	0.896	**0.380**	-fit_pron_me +fit_pron_we +fit_1st_person +fit_indicative +avg_subjunctive
Seller Features	0.834	0.898	-0.222	+fit_n_introduced -avg_direct_start -fit_pron_you -fit_hedges +fit_indicative
Buyer+Seller Features	**0.857**	**0.910**	-0.519	-avg_seller_1st_person -avg_buyer_2nd_person_start +fit_seller_apologising +fit_buyer_please_start +fit_seller_n_adopted_w_hedge

Table 1: Accuracy, F1-score and McFadden's R^2 for predicting negotiation success in the Craigslist Bargain dataset. The speaker-independent features are in the top part of the table. Speaker-dependent features are in the bottom part of the table where the buyer and seller features include LIWC+Politeness separated out and calculated per role. The top-5 features are sorted according to the absolute coefficient value.

Results

Persuasion For Good

Features	**Accuracy**	**F1-score**	R^2	**Top-5 features**
Baseline Majority	0.536 (0.001)	**0.698 (0.001)**	-	-
COLL	**0.571 (0.029)**	0.653 (0.022)	-0.088 (0.063)	+avg_agree +avg_n_repeated_content +avg_n_repeated_stop -fit_disagree +fit_repeated_stop
COLL + LIWC + Politeness	0.556 (0.039)	0.591 (0.038)	**0.025 (0.058)**	-avg_geo -avg_has_negative +avg_has_positive +avg_agree -avg_direct_question

Table 2: Mean Accuracy, F1-score and McFadden's R^2 for predicting persuasion in the Persuasion for Good Dataset in the 5-folds. The figure between brackets represent the standard deviation across the different folds. The top-5 features are sorted according the mean of absolute coefficient values.

Opaque Methods

- Sentence Representation
  - RoBERTa-SE (Reimers and Gurevych, 2019): average sentence embeddings for all turns in the dialogue
  - ConvERT (Henderson et al, 2019): dialogue embedding
- Methods:
  - Linear-NN: Linear layer followed by a softmax layer
  - Linear regression

- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In Proceedings of EMNLP 2019.
- Matthew Henderson et al. .2019. Convert: Efficient and accurate conversational representations from transformers. arXiv.

Results: Opaque Methods

Features	Model	Accuracy	F1-score	R^2
RoBERTa-SE	LR	0.854	0.906	**0.560**
ConvERT	LR	**0.895**	**0.932**	0.533
RoBERTa-SE	Linear-NN	0.843	0.904	-
ConvERT	Linear-NN	0.859	0.913	-

Craigslist Bargain

Features	Model	Accuracy	F1-score	R^2
RoBERTa-SE	LR	0.611 (0.038)	0.638 (0.052)	0.050 (0.331)
ConvERT	LR	0.602 (0.022)	0.665 (0.027)	0.120 (0.003)
RoBERTa-SE	Linear-NN	0.607 (0.010)	**0.724 (0.013)**	-
ConvERT	Linear-NN	**0.622 (0.018)**	0.715 (0.004)	-

Persuasion for Good

Discussion

- It can be useful to look at linguistic features from a speaker dependent perspective. In no deal dialogues:
  - Sellers use more 1st person pronouns
  - Buyers use more 2nd person pronouns
- In dialogues where there was a deal achieved, length of the utterance tends to decrease over time
  - The challenge is when systems need to be transparent
- Collaborative features are more relevant in predicting persuasion
  - Language style is context dependent (competitive vs collaborative)
- Neural methods improve dialogue outcome detection

Conclusion

- We have investigated linguistic indicators that reflect two tasks:
  - when a deal has been reached in negotiation dialogues, and
  - persuasion for a donation
- These two interaction outcomes can be seen as examples of *conditional trust* (Jones and George, 1998)
- Various lexicon-based features were identified as being indicators of success through our transparent method of training regressors
- A role-based analysis showed differences in the relevant features in negotiation
- Methods based on dialogue embeddings achieved the best performance in both problems, but are not transparent

Future work

- Trustworthiness data collection
  - In this work, success in negotiation and persuasion were used as proxies for trust
  - Collect trustworthiness scores and propensity to trust
  - Fine-grained trustworthiness scores (turn level)
- Condition language generation to instill trust
  - Using neural models

jobs.ac.uk

Find a Job Find PhDs Career Advice Jobs by Email

< Return to the home page

Research Fellow - Conversational Systems and Human-Robot Interaction

Heriot-Watt University - School of Mathematical and Computer Sciences

h.hastie@hw.ac.uk