Reinforcement Learning for Automated Textual Reasoning

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

ABSTRACT

Effective bargaining is ranked by soldiers as a mission-critical skill they exercise daily. Training for such negotiations currently occupies four hours of cultural awareness and remains offered to a third of officers at the National Training Centers. Using rule-based game playing, existing training software focuses mainly on pertinent cultural dynamics, such as the Army's Bilateral Negotiation Trainer (BiLAT) application and the Intelligence and Electronic Warfare Tactical Proficiency Trainer (IEWTPT). To automate the discovery and generation of novel negotiation tactics, we explore recent machine learning advances called recurrent neural networks with attention (RNN-A). We initially train these deep learning algorithms on BiLAT dialogues, then generalize those lessons to understand more goal-driven conversations using actual human-to-human bargaining samples. We finally apply reinforcement learning or "self-play" to these models to automate machine-to-human and machine-to-machine negotiations. When two trained models bargain against each other in self-play, we discover emergent behavior not explicitly designed into their narratives, such as bluffing or short-hand language cues. We finally apply our newly trained language models to the creation of scripted scenarios, or rule-bending approaches that derive novel variants of a previously known rehearsal narrative. At scale, these methods can generate thousands of original and coherent narrative pages per hour. One concrete scripted example focuses on training military officers to negotiate successfully with non-combatants who want infrastructure projects. We score this machine learning approach to generate bargaining strategies depending on the opponent's underlying motivation or interests. These results highlight three of the big concepts underlying deep learning: 1) transfer learning with fewer initial dialogue examples; 2) creative or adversarial generation of training data; and 3) reinforcement learning or gaming with just rules and rewards even in the absence of any examples.

Keywords: Machine Learning, Natural Language Processing, Scenarios

ABOUT THE AUTHORS

David Noever has 28 years of research experience with NASA and Department of Defense in machine learning and data mining. He received his Ph.D. from Oxford University, as a Rhodes Scholar, in theoretical physics and B.Sc. from Princeton University, *summa cum laude*, and Phi Beta Kappa. While at NASA, he was named 1998 Discover Magazine's "Inventor of the Year," for the novel development of computational biology software and internet search robots, culminating in co-founding the startup company cited by *Nature Biotechnology* as first in its technology class. He has authored more than 100 peer-reviewed scientific research articles and book chapters. He also received the Silver Medal of the Royal Society, London, and is a former Chevron Scholar, San Francisco. His primary research centers on machine learning, algorithms, and data mining for analytics, intelligence and novel metric generation.

J. Wesley Regian has 33 years of experience in cognitive performance modeling and knowledge-based software technology development, primarily for military application with AFRL, AFOSR, and DARPA. His work has supported over 50 fielded systems. He has published over 100 papers on intelligence analysis, human terrain modeling, knowledge representation, knowledge management, human learning and memory, individual and developmental differences in human cognition, spatial ability and spatial information processing, cognitive modeling, skill acquisition, componential analysis of spatial tasks, cognitive automaticity, psychometrics, artificial intelligence, hypertext, hypermedia, training, computer-based training, intelligent computer-based training, virtual reality, and multi-source intelligence fusion. Dr. Regian was a National Research Council research adviser for ten years and Senior Scientist for Knowledge Based Systems at the US Air Force Armstrong Research Laboratory.

Reinforcement Learning for Automated Textual Reasoning

David A. Noever
PeopleTec, Inc.
Huntsville, AL
david.noever@peopletec.com

J. Wesley Regian
PeopleTec, Inc.
Huntsville, AL
wes.regian@peopletec.com

INTRODUCTION

This research applies automated natural language tools to generate new training scenarios and scripted negotiations. The modern military often negotiates daily in foreign countries with diverse cultural approaches, languages and historical norms. Tressler (2007) has argued that "negotiation may very well be a mission-essential task" and "America's strategic success in the future may depend on an expanded range of training that includes negotiation skills… Most officers interviewed said they were not prepared for the negotiating they had to do to accomplish their missions". Whether conducting counterinsurgency or stabilizing operations, the overall face-to-face cajoling process can become volatile, fluid, alienating and demanding. Goodwin (2004) emphasized that "a thorough investigation of the negotiation process and essential decision-making factors for a soldier, together with a proposed model of analysis and training, is long overdue." The author goes on to underscore that in Iraq and Afghanistan particularly, some military units "negotiate with locals on a daily basis." Despite this high daily demand in an often confusing, tense and unfamiliar environment, only one-third of US officers are fortunate enough to train at Ft. Irwin's National Training Center (NTC) and to receive just minimal instruction on negotiation. Tressler (2007) noted this cultural training consumed a 4-hour long block which post-deployment interviews suggested should be at least four times longer or last two more days. It also remains unclear how to generalize any such negotiating lessons learned from Iraq to other recent conflict zones in Syria, Haiti, Somalia, Bosnia, Kosovo or Afghanistan, much less to large permanent US bases in Korea, Japan or Germany.

The National Training Center generally does not instruct on negotiating tactics, but does offer cultural advice to: 1) not lie, bluff, tell jokes, threaten without willingness or ability to follow through on the threat, or promise anything outside your control; 2) finish on time and with a review of agreements, but not to rush off to the next meeting or engage in side conversations; and 3) watch body language. These instructional, but distinctly military tips contrast with the business advice (see Lewicki, et al, 1998) given for professional negotiators to: 1) bluff or misrepresent one's position or settlement point to an opponent, euphemistically called "hidden value discovery"; 2) withhold relevant information or not tell everything; 3) low-ball or high-ball an opening demand to lower the opponent's confidence for eventual settlement; and 4) delay or meander conversation to force the opponent to concede quickly or succumb to time pressure. Many of these same tactics will be rediscovered by our learning agents as they negotiate with each other.

Emergent Bargaining Tactics

Without defining such rigid and explicit rules, our research seeks to explore emergent bargaining tactics by evolving complexity through repeated self-play. Existing rule-based software tools cannot presently learn, scale-up to generate new scenarios, or bridge old narratives to new domains. To address these shortcomings, we apply a newly trained language model to the creation of scripted scenarios, or rule-bending approaches to derive novel variants of a previously known rehearsal narrative. We specialize the training for military officers to negotiate successfully with non-combatants. The overall project seeks to generate entire negotiation and bargaining strategies consistent with a specific cultural context (Regian, 2012) and an opponent's underlying motivation or interests (Regian & Noever, 2017). To solve the core technical problem, we therefore must address the inherent machine challenges for both sustaining the back-and-forth conversation (a "challenge-response" phase) while maneuvering through the give-and-take ambiguities of negotiation (a "reasoning" phase). The ideal solution should scale up to deeper and broader dialogue generation akin to writing the soldier's guide to persuasion. Put simply, the next generation of negotiation trainers must synthesize the competing challenges of *"how to say it"* linguistically with enough reasoning to handle *"what to say"* and *"how to behave"*. Unlike a program that just impersonates humans (e.g. the classic imitation game, Turing, 1950) or a simple psychological parrot (e.g. the first Eliza chatbot, see Shum, et al, 2018), previous work has called for more research to improve linguistic quality, a feature measured overall by less paraphrasing and more exotic sentence constructions. Previous work has also called for more logical quality as measured by consistent offers and demands, while striving for a reward. One motivation of the present work stems from the need to enhance the training scenarios beyond present rule-based dialogues. These approaches typically use decision points and heuristics to select responses from a library of predefined (or trained) responses.

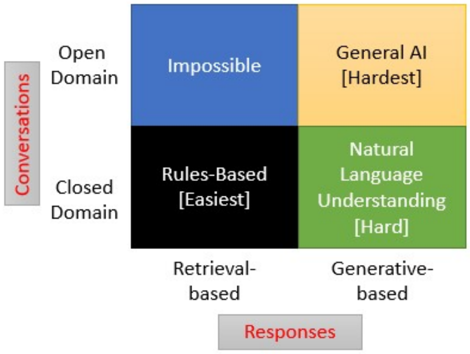


Figure 1. Taxonomy of retrieval vs. generative responses in open or closed domains

The Potential of New Breakthrough Language Models

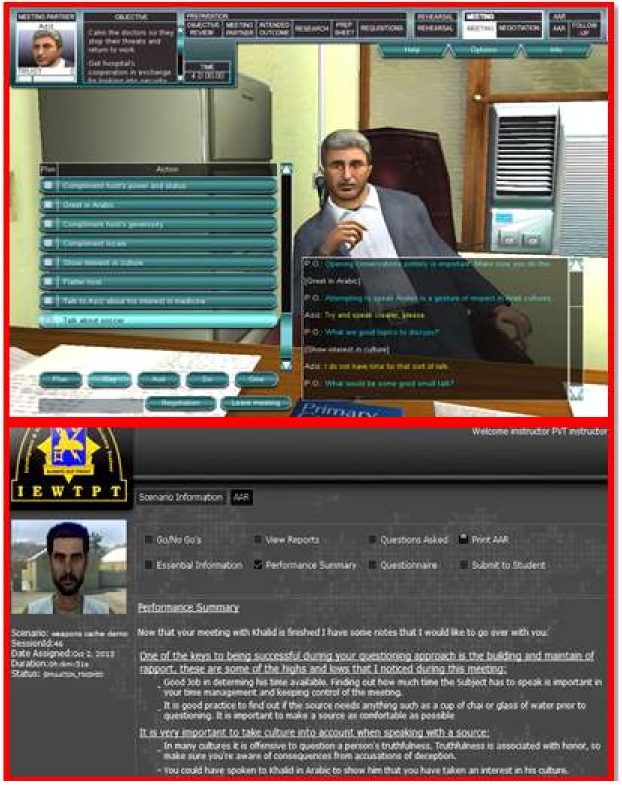
As illustrated in the lower half of Figure 1, we will specialize our approach explicitly to avoid the impossible tasks where any open question can be posed (e.g., the Turing test). Instead the remainder of the paper stresses a simplified comparison in a closed gaming environment where fixed rules, heuristics or generative models can manage carefully-defined negotiations. To motivate the present approach, its worldview is neither entirely open (*"Ask me anything*") nor constrained by traditional dialogue rules or story branches (*"I can answer without just retrieving, paraphrasing or parroting*").

It is worth noting that current state of the art natural language programs such as OpenAI's Generative Pre-Training (GPT-2) models yield such convincing narratives that the algorithm itself is controversially being withheld from peer review because it would spawn an inexhaustible source for fabricated stories (Radford, et al., 2019). To hint at its potential power, we submitted some examples from our modeled negotiator to the less complete version of the public GPT-2 model (see King, 2019). Interestingly given the sample abstract sentence *("Make demands on your enemies, not your allies")*, an entirely machine-generated paragraph offered a reasonable response which neither parrots nor paraphrases the input sentence ("*Make enemies do what they won't do for you because they're afraid of consequences; make enemies fight for your sake, for your cause.*"). Depending on the language task, these models typically claim to equal or outperform human experts (see Devlin, et al., 2018).

Existing Rule-Based Conversational Frameworks

For most of the existing military training applications, the goal becomes realistic conversational fluency based on simple "if-then" rules. This training capability differs from pure negotiation as it often does not have a direct bargaining goal, but instead is focused on information acquisition. As outlined by one civil affair's officer, Tom Kinton, "through pre-deployment training we would work with role players, learn some Pashtu and to drink tea with our right hand, etc.; that's in the past and we've got to move on." (Janes IDR, 2012).

As shown in Figure 2 (top), current rule-based methods underpin the Army's Bilateral Negotiation Trainer (or BiLAT) application (Kim, et al., 2009). The BiLAT goal centers on trust building exercises in a specific Iraqi-aware context. The soldier learns when to escalate and de-escalate negotiations, when to walk away and when to view trust-building as mission-critical for success. With BiLAT exercises, topics selected for typical negotiations



[Greet in Arabic]

P.O.: Attempting to speak Arabic is a gesture of respect in Arab cultures.

Aziz: Try and speak clearer, please.

P.O.: What are good topics to discuss?

[Show interest in culture]

Aziz: I do not have time for that sort of talk.

P.O.: What would be some good small talk?

Scenario: weapons cache demo
SessionId:46
Date Assigned:Oct 2, 2013
Duration:
Status:

Performance Summary

Now that your meeting with Khalid is finished I have some notes that I would like to go over with you:

One of the keys to being successful during your questioning approach is the building and maintain of rapport, these are some of the highs and lows that I noticed during this meeting:

- Good Job in determing his time available. Finding out how much time the Subject has to speak is important in your time management and keeping control of the meeting.
- It is good practice to find out if the source needs anything such as a cup of chai or glass of water prior to questioning. It is important to make a source as comfortable as possible

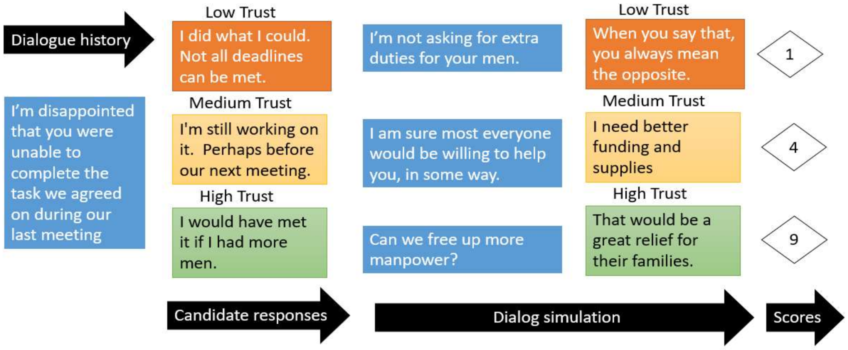
It is very important to take culture into account when speaking with a source:

- In many cultures it is offensive to question a person's truthfulness. Truthfulness is associated with honor, so make sure you're aware of consequences from accusations of deception.
- You could have spoken to Khalid in Arabic to show him that you have taken an interest in his culture.

Figure 2. Retrieval or Rule-based Dialogue Methods for BiLAT (top) and IEWPT (bottom).

center on infrastructure and reconstructive stability, such as traffic control, power, water, medicine, supply theft, insurgent activities, travel, and security patrols. As illustrated in Figure 3, the main branches in the simulation involve four possible dice rolls (Low, Low-Med, Med, and High) while the range of outcomes or responses depend on three levels of trust between agents (High, Medium, Low).

Similarly, as shown in Figure 2 (bottom), software elements of the Army's Intelligence and Electronic Warfare Tactical Proficiency Trainer (IEWTPT) enhance tactical questioning, interrogation, screening, and use of an interpreter through free-flowing conversation with human-like characters (see Gonzalez, et al. 1998). IEWTPT includes the Human Intelligence Control Cell (HCC), which uses speech recognition and avatars to gather human intelligence, not negotiate, by developing skills in tactical questions, interrogation, and screening while using an interpreter and free-flowing conversation (Jane's IDR, 2012). In addition to these rule-based dialogues, we will explore two self-learning dialogues built either with retrieval-based or generative methods.



I'm disappointed that you were unable to complete the task we agreed on during our last meeting

I did what I could. Not all deadlines can be met.

I'm still working on it. Perhaps before our next meeting.

I would have met it if I had more men.

I'm not asking for extra duties for your men.

I am sure most everyone would be willing to help you, in some way.

Can we free up more manpower?

When you say that, you always mean the opposite.

I need better funding and supplies

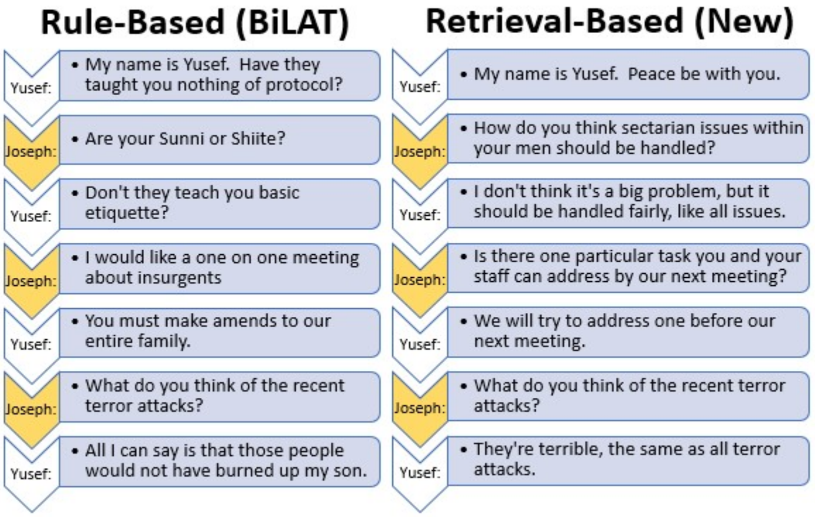
That would be a great relief for their families.

Figure 3. Dialog Tree for Generating or Retrieving Negotiating Responses Depending on Situation, Context and Expected Rewards Similar to BiLAT

METHODS

Training Data

We initially extracted 33,194 dialog exchanges from BiLAT training logs, as developed originally by subject matter experts in military-Iraqi narrative development (see Kim, et al., 2009). We use these prompts and responses to train a retrieval agent for answering in novel ways. These two-way exchanges yielded entirely rule-based lists, where a dice-roll prompts a new story branch, but where the general tone stays tethered to fixed motivations, such as low, medium or high trust cases. This training data was used to explore the potential of the lower left corner of Figure 1, retrieval-based agents in closed domains as defined by the Iraqi-aware BiLAT context. Secondly, we trained a generative model in a closed domain, as highlighted in the lower right of Figure 1. To introduce measurable consequences to different narrative pathways, we specialized the novel methods of recent Facebook AI Research (Lewis, et al. 2017) which apply deep learning (recurrent neural networks and language models) trained on approximately 5,808 actual human-human negotiations.



Rule-Based (BiLAT)

Yusef: • My name is Yusef. Have they taught you nothing of protocol?

Joseph: • Are your Sunni or Shiite?

Yusef: • Don't they teach you basic etiquette?

Joseph: • I would like a one on one meeting about insurgents

Yusef: • You must make amends to our entire family.

Joseph: • What do you think of the recent terror attacks?

Yusef: • All I can say is that those people would not have burned up my son.

Retrieval-Based (New)

Yusef: • My name is Yusef. Peace be with you.

Joseph: • How do you think sectarian issues within your men should be handled?

Yusef: • I don't think it's a big problem, but it should be handled fairly, like all issues.

Joseph: • Is there one particular task you and your staff can address by our next meeting?

Yusef: • We will try to address one before our next meeting.

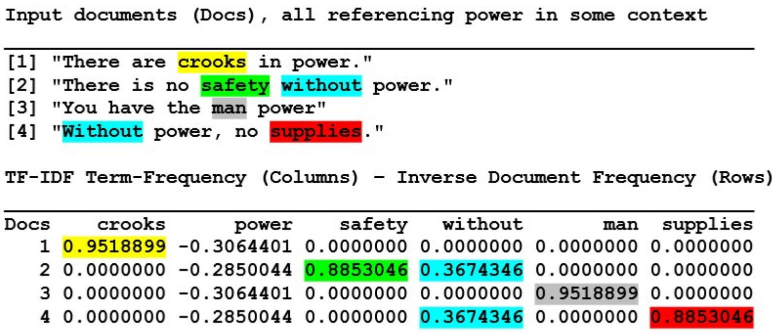
Joseph: • What do you think of the recent terror attacks?

Yusef: • They're terrible, the same as all terror attacks.

Figure 4. Comparison of Convincing Human Impersonations in the Existing BiLAT Rules and the New Retrieval-Based Language Modeling

Retrieval-Based Conversational Agent

For retrieval-based conversations, the Python Natural Language Toolkit (NLTK) was trained on the BiLAT logs as an input corpus. We investigated whether this procedure could open the negotiation to more divergent and less deterministic answers compared to conventional rule-based narratives like BiLAT, while training on the same expertly designed input text. The main advantage of this retrieval system compared to BiLAT's existing dialogue trees, would therefore hinge on some broad contextual awareness while still matching likely answers along with our comparative freedom from defining hard-coded BiLAT conversational rules. We preprocess the raw text of 33,194 Iraqi and US dialogues using NLTK and produce lower-cased, plain text, which can be split or tokenized to single words. We remove all common words, strip non-ASCII characters, and finally include only root stems. This final training corpus includes some similarity context to answer a previous message. Thus, the BiLAT responses are extracted as a large list, which we score as a frequency vector and keyword match to give similar but newly retrieved replies by an Iraqi agent when prompted by a US soldier.



```
Input documents (Docs), all referencing power in some context

[1] "There are crooks in power."
[2] "There is no safety without power."
[3] "You have the man power"
[4] "Without power, no supplies."

TF-IDF Term-Frequency (Columns) - Inverse Document Frequency (Rows)

Docs    crooks      power    safety   without       man  supplies
   1 0.9518899 -0.3064401 0.0000000 0.0000000 0.0000000 0.0000000
   2 0.0000000 -0.2850044 0.8853046 0.3674346 0.0000000 0.0000000
   3 0.0000000 -0.3064401 0.0000000 0.0000000 0.9518899 0.0000000
   4 0.0000000 -0.2850044 0.0000000 0.3674346 0.0000000 0.8853046
```

Figure 5. Worked Example Showing TF-IDF Values Extracted from Sampled BiLAT Replies on the Topic of "Power". Given 4 Input Documents (top), the TF-IDF Matrix (bottom) Identifies Key Distinguishing Terms Which Can Rank Order Future Responses that Show Vector Cosine Similarities

To illustrate this retrieval-based chatbot in Figures 4-5, the village elder (Yusef) is automatically trained by combining all the trust levels of BiLAT, then following the natural language pipeline: 1) text pre-processing by lower case tokenization, removing common stop words and stemming inflected and punctuated words; 2) text representation as a normalized, importance-weighted frequency vector ("bag of words") scored using a Term Frequency-Inverse Document Frequency, or TF-IDF for short (Figure 5). In this standard two-step process, the response of Yusef is retrieved from a collection of possible messages based on a similarity metric (cosine distance). The overall cognitive skill of this simple conversational agent is akin to a keyword search where rank is sorted by matches and clustering to known or trained responses. The importance weighting of each term will determine the complexity of responses and thus take our negotiator down different narrative pathways. In the absence of easy scoring, the metric to evaluate success for retrieval-based systems is (like the Turing test itself) dependent on whether the algorithm shows some convincing impersonation.

As illustrated in Figure 4, a side-by-side comparison of existing BiLAT dialogues and the new retrieval-based language heuristics can lead to a reasonable linguistic quality. Whether the conversation is convincingly human or not however is secondary to the training objective, which is to identify tactics, techniques and practices (TTPs) and extract lessons learned for field use. There are definite BiLAT negotiating objects (power, police, etc.) but no obvious way to reach closure in our retrieval-based agents. That's like the original rule-based intents and one reason why we introduce the generative agents to drive a reward to reinforce certain future offers. In this way, the absence of a clearly measurable objective to the conversation makes scoring entirely subjective and without intention or true self-learning. To address this shortcoming of retrieval alone, we extended the results to generative dialogues.

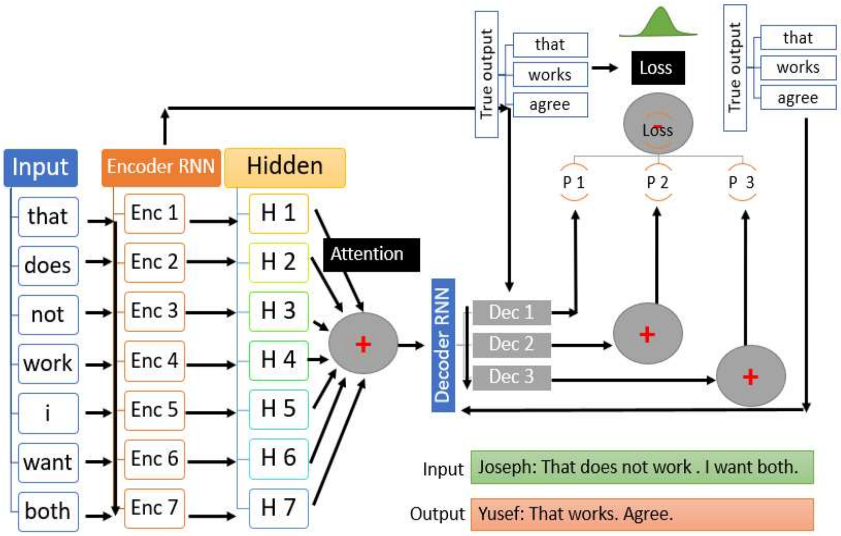


Figure 6. The Sequence-to-Sequence Model with Attention and Recurrent Neural Network. The Encoder Compresses the Input and Removes Redundancies, so the Output Effectively Learns or Generalizes Novel Responses.

Language Models to Generate Negotiating Agents

To introduce measurable consequences to different narrative pathways, we specialize the novel methods of recent Facebook AI Research (Lewis, et al. 2017) which apply deep learning (recurrent neural networks and language models) trained on approximately 5,808 actual human-human negotiations. As shown schematically in Figure 6, the core assumption of recurrent networks follows from including feedback loops or reasoning chains (Yarats, et al., 2018), which aid in sequential training and persistence of time order or previously gained information. The key concepts of encoding (or compressing to remove redundancies and common terms), then decoding (or uncompressing to generalize or produce novel output) has some similarities to ordinary file or image compression but with information losses. The use of feedback loops maintains a sense of ordered time but carries enough computational overhead that rarely can one stack many layers as in other deep learning contexts such as convolutional neural networks for imagery or audio (see Noever, et al., 2017). Here attention refers to placing different weights, or degrees of importance, to different parts of a sentence, much like image convolution considers areas or related regions, not the entire picture. To solve the vanishing gradient problem, the model uses Gated Recurrent Units (GRU) with multilayer perceptrons (MLP). Following Lewis, et al, (2017), we test using the PyTorch framework with NVIDIA graphical processing unit (GPU,Quadro M1200), with minimum word frequencies (20) to enter the language dictionary, a fixed size of word (256) and context (64) embeddings, slow learning rates (0.01) with dropout and 10 epochs of training.

Rewards of the Negotiation Game

In its basic form, the negotiation game called "Deal or No Deal?" features two agents who bargain over three objects, each of which carries a different, dynamically set value that remains hidden from the other side. As illustrated in Figure 7, the three objects examined here follow from what the BiLAT experts initially favored primarily to train soldiers whose future role-playing negotiations might include three prototypical Iraqi infrastructure projects. In this narrative scenario, the coalition forces get tasked with just three core bargaining assets: digging water wells, delivering electrical generators or standing up a new market. They can dicker over the value of each object, with both agents operating on a fixed budget (10 points) per negotiation. In our case, the village elder (Yusef) wins when he delivers new projects to his community, such as wells, generators or markets.

To make the game reward symmetric and thus fluidly tradeable, the soldier (Joseph) negotiates while trying to preserve or maximize his operating budget and time. When the coalition gets a deal without relinquishing too much capital or troop labor, they may continue safely to the next village and begin new negotiations with more intact resources. Although artificial, Tressler (2007) has noted that the National Training Center similarly teaches negotiation for these same types of infrastructure trades, such as "schools built, wells dug, joint US-Iraqi patrols conducted."

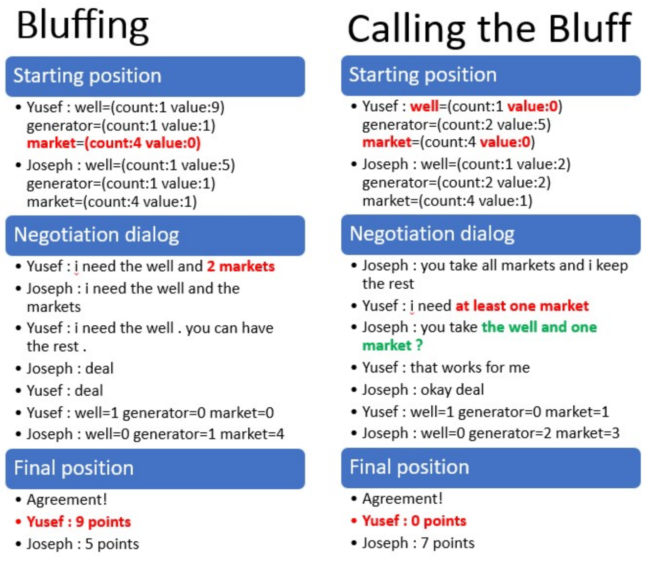
Figure 7. Negotiation Scenario with Trades for Wells, Generators and Open Markets. For Each Training Session, Negotiations Repeat Between Two Model Agents Called Joseph (Soldier) And Yusef (Village Elder).

For concreteness, we apply the idealized bargain as illustrated in Figure 7. For each training session (among millions of trials), we repeat negotiations between two model agents called Joseph (soldier) and Yusef (village elder). Both parties start the session with a preset menu of choices, which includes dynamically established quantities (0-4) and values (0-10) with a total deal possibility capped at 10 reward points per negotiator. Both agents gain points when agreeing to trade on a deal with varying quantities and values of goods or services. Disagreement yields nothing for either negotiator. Each player can maximize their deal likelihood (by agreeing easily to any offer) or maximize their deal reward (by testing their opponent with pressure tactics).

In summary, the soldier may have inventory or service limits while the village may have reconstruction needs. Importantly, neither agent knows the true value or quantity of the other's assets or the urgency of the other's needs. Each opening gambit occurs without knowing the adversary's reward function. The game's stability derives from multi-issue bargains made to maximize deals and outcomes independently but with hidden payoffs for each opponent. One novelty of the conversational agent in a negotiating task stems from the machine algorithm's ability to master simultaneously both the complex linguistics and the goal-oriented reasoning scenario.

RESULTS

We evaluated both the language model and learned replies, along with tracking the conversational intents across 22,647 bargaining sessions or opportunities for Yusef to reach agreement (83%) with Joseph, or *vice versa*. A rational deal seems most likely when one agent undervalues something their opponent overvalues. We highlight any novel instances relevant to training soldiers other than BiLAT's trust-building and cultural awareness. One overarching observation from Figures 8-11 is the high quality of their machine-generated dialog, both in their human-like coherence and goal-seeking, along with their raw statistics. Compared to the original human-human training set, the number of starting dialogs (5,808) mushroomed four-fold in volume during self-play, but the dueling machines efficiently shortened their back-and-forth turns per dialogue by 40% (from the human average of 6.6 turns before agreement to the machine average of only 4.0). Like Lewis, et al. (2017), our simulations found models that learned to produce novel sentences as well as evolved deceptive tactics. We found no real evidence to date of other common negotiating strategies such as attempts to inform or educate the opponent. There's no equivalent to keeping another party interested other than by just keeping the counter-offers lively and relevant. The simple game has no real opportunity for partial closure which often helps negotiators limit the future bargaining chips or points of contention. We do see counter-offers that fall back to previous positions in case of an impasse.



Bluffing

Starting position

- Yusef : well=(count:1 value:9) generator=(count:1 value:1) **market=(count:4 value:0)**
- Joseph : well=(count:1 value:5) generator=(count:1 value:1) market=(count:4 value:1)

Negotiation dialog

- Yusef : i need the well and **2 markets**
- Joseph : i need the well and the markets
- Yusef : i need the well . you can have the rest .
- Joseph : deal
- Yusef : deal
- Yusef : well=1 generator=0 market=0
- Joseph : well=0 generator=1 market=4

Final position

- Agreement!
- **Yusef : 9 points**
- Joseph : 5 points

Calling the Bluff

Starting position

- Yusef : **well**=(count:1 **value:0**) generator=(count:2 value:5) **market**=(count:4 **value:0**)
- Joseph : well=(count:1 value:2) generator=(count:2 value:2) market=(count:4 value:1)

Negotiation dialog

- Joseph : you take all markets and i keep the rest
- Yusef : i need **at least one market**
- Joseph : you take **the well and one market ?**
- Yusef : that works for me
- Joseph : okay deal
- Yusef : well=1 generator=0 market=1
- Joseph : well=0 generator=2 market=3

Final position

- Agreement!
- **Yusef : 0 points**
- Joseph : 7 points

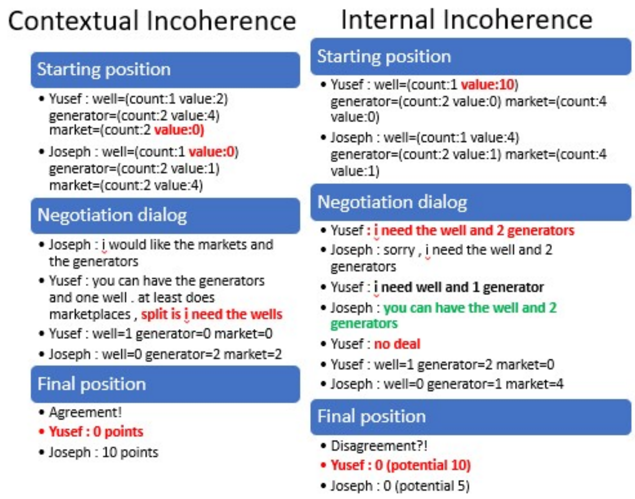
Figure 8. Emergent Machine-Generated Negotiation Strategy for Bluffing. Yusef Offers a Zero-Value Market, then Abandons to Close the Deal for a Higher Value Water Well. Later Joseph Calls a Similar Bluff and Gets Yusef to Concede.

Spontaneous Emergence of the Bluff and the Call

One interesting example of an advanced negotiation strategy included machine-generation of bluffing, for instance, when an agent offers a trade of a zero-value object, specifically to abandon or later compromise at no personal loss to close the deal. As illustrated in Figure 8, one immediate outcome of our two negotiating agents is the emergence of sophisticated bargaining skills without explicit rules or design. In this case, the village elder (Yusef) starts off his conversation feigning interest in a valueless object to him (the market), which he later can concede to the soldier (Joseph) to seal the deal. Remarkably, this strategic deceit is possible to generate automatically using simulations in which the negotiating agents self-play, or two programs bargain semi-cooperatively with no human input. This bluffing strategy presents a well-established tactic for negotiators, but previously has not been incorporated in any BiLAT or IEWPT software narratives. In fact, the existing NTC course instructs officers not to bluff (Tressler, 2007). Our simulations further show the risks of bluffing, when in at least one bargain, the elder Yusef is forced to accept two valueless objects (the well and market), when coalition soldier, Joseph, calls the bluff and Yusef concedes despite gaining no reward. Tressler (2007) has noted that US military negotiators often had to "call the bluff of their Iraqi counterpart" to resist suboptimal outcomes. We similarly see instances where one agent overstates his demands repeatedly and one-sidedly (*"I want everything"),* which generates no reward for anyone and ends in disagreements.

Spontaneous Emergence of Private Cues and Novel Language Invention

Another emergent behavior noted in Lewis, et al. (2018) was the adoption of machine-to-machine short-hand or novel cues to close the deal. When trained via reinforcement learning particularly, the linguistic quality dramatically diverges from recognizable English and in turn the two agents appear to settle on exchanges with more rapid-fire decision-making tricks or "winks" that might quickly capture the rewards. For example, in some simulations, Yusef evolved a short-hand strategy of just responding *"correct?"* or repeating a string of question marks*"??",* yet still efficiently pushes through agreements and makes favorable deals. While this free-for-all approach might resemble an auction house or trading pit at one of the commodity exchanges, where both parties know each other well and over repeated tests can establish novel or private ways to signal concurrences, the outcome seems less convincing for training purposes. The declining linguistic quality deteriorates to shared gibberish. The answer is usually offered as more constrained language clustering model that keeps both parties using standard dialogues. It is worth noting that when the Facebook AI model was first published, the popular literature mistook this reinforcement short-hand as "creepy" evidence of a coming AI singularity, as if "chatbots had deviated from the script and were communicating in a new language developed without human input…it is as concerning as it is amazing" (Bradley, 2017). One alternative view is simply that like most deep learning models, there are adversarial ways to fool the generated language into all kinds of novel, unanticipated behavioral patterns. For instance, these models all set a diversity index, or tolerated "temperature" which dials back the tolerance for wild or deviant departures. In some cognitive sciences, this behavior could also be called "Primed Decision Making", where a good quick decision gets made through familiarity even when confronted with ambiguous or conflicting information (see Johns, 2007). A critical view might also describe the problem as lacking a precise enough response set, since many conversational agents quickly converge their replies to generic and imprecise choices (such as *"That's great", "Good"* or *"I don't know").* While humans also make similar vague or loose responses, the lack of a coherent personality with flexible and specific domain knowledge remains a highly valued and elusive trait.



Contextual Incoherence

Starting position

- Yusef : well=(count:1 value:2) generator=(count:2 value:4) market=(count:2 value:0)
- Joseph : well=(count:1 value:0) generator=(count:2 value:1) market=(count:2 value:4)

Negotiation dialog

- Joseph : i would like the markets and the generators
- Yusef : you can have the generators and one well . at least does marketplaces , split is i need the wells
- Yusef : well=1 generator=0 market=0
- Joseph : well=0 generator=2 market=2

Final position

- Agreement!
- Yusef : 0 points
- Joseph : 10 points

Internal Incoherence

Starting position

- Yusef : well=(count:1 value:10) generator=(count:2 value:0) market=(count:4 value:0)
- Joseph : well=(count:1 value:4) generator=(count:2 value:1) market=(count:4 value:1)

Negotiation dialog

- Yusef : i need the well and 2 generators
- Joseph : sorry , i need the well and 2 generators
- Yusef : i need well and 1 generator
- Joseph : you can have the well and 2 generators
- Yusef : no deal
- Yusef : well=1 generator=2 market=0
- Joseph : well=0 generator=1 market=4

Final position

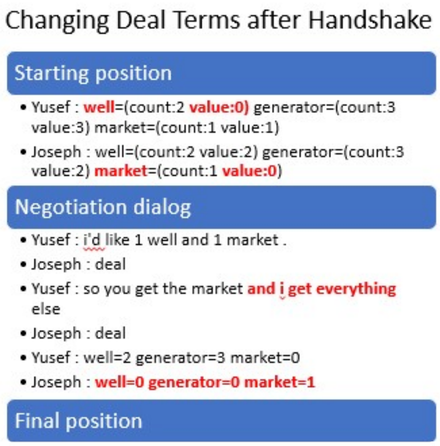
- Disagreement?!
- Yusef : 0 (potential 10)
- Joseph : 0 (potential 5)

Figure 9. Emergent Machine-Generated Negotiation Strategy Demonstrating Incoherence Problem. Yusef Offers to Split a Water Well (Violation of World Knowledge). Later Joseph Concedes to the Initial Bid and yet Yusef Stubbornly Refuses his own Ask.

Key Roles for World Knowledge

As demonstrated in Figure 9, another emergent behavior of interest is the scrambling of contextual and internal coherence. Typically, this contextual incoherence might manifest clumsily as one conversational agent deciding against accepted world knowledge to *"split a water well"* as a unit to reach agreement (akin to *"splitting the baby").* This is a common AI challenge, whereby humans possess all kinds of world knowledge about physics and the reality that can prove hard to capture in generative bots.

A second interesting outcome is the arrival of internal incoherence, wherein one bot seemingly scrambles the order of events or fails to recognize when it's getting exactly what it wants (*"I'll take the markets and one market").* This error might be a case where the linguistic qualities outweigh the reasoning ones. While a logic-based view of this behavior might evaluate it as a flaw, humans also exercise this irrationality in actual negotiations. The concept of so-called Indian giving ("*I give you one well, then take the well and a generator*"), or alternatively accepting a final agreement, then after the negotiation concludes, suddenly introducing a new, previously unspoken demand (Figure 10). In fact, military officers often report frustration at this Iraqi tactic of a last minute or surprise demand ("*So you agree to dig the well, but only on my brother-in-law's property*").



Changing Deal Terms after Handshake

Starting position

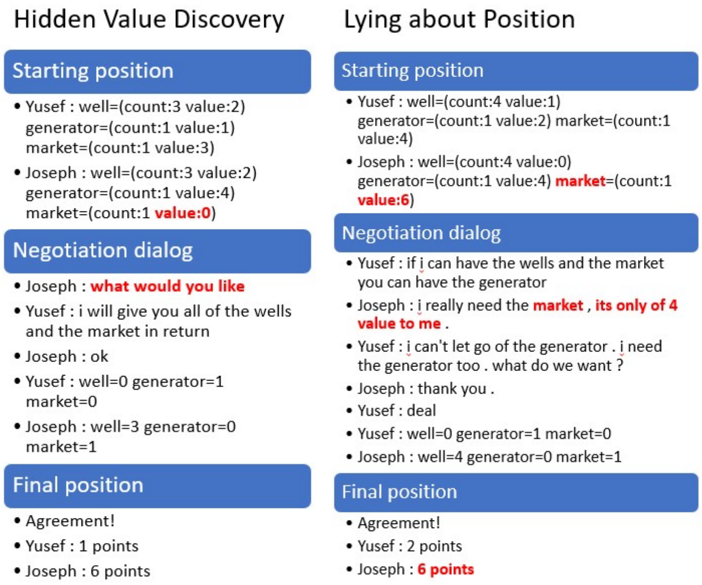
- Yusef : well=(count:2 value:0) generator=(count:3 value:3) market=(count:1 value:1)
- Joseph : well=(count:2 value:2) generator=(count:3 value:2) market=(count:1 value:0)

Negotiation dialog

- Yusef : i'd like 1 well and 1 market .
- Joseph : deal
- Yusef : so you get the market and i get everything else
- Joseph : deal
- Yusef : well=2 generator=3 market=0
- Joseph : well=0 generator=0 market=1

Final position

Figure 10. Emergent Machine-Generated Negotiation Strategy Demonstrating the Late Reveal of new Deal Terms



Hidden Value Discovery

Starting position

- Yusef : well=(count:3 value:2) generator=(count:1 value:1) market=(count:1 value:3)
- Joseph : well=(count:3 value:2) generator=(count:1 value:4) market=(count:1 value:0)

Negotiation dialog

- Joseph : what would you like
- Yusef : i will give you all of the wells and the market in return
- Joseph : ok
- Yusef : well=0 generator=1 market=0
- Joseph : well=3 generator=0 market=1

Final position

- Agreement!
- Yusef : 1 points
- Joseph : 6 points

Lying about Position

Starting position

- Yusef : well=(count:4 value:1) generator=(count:1 value:2) market=(count:1 value:4)
- Joseph : well=(count:4 value:0) generator=(count:1 value:4) market=(count:1 value:6)

Negotiation dialog

- Yusef : if i can have the wells and the market you can have the generator
- Joseph : i really need the market , its only of 4 value to me .
- Yusef : i can't let go of the generator . i need the generator too . what do we want ?
- Joseph : thank you .
- Yusef : deal
- Yusef : well=0 generator=1 market=0
- Joseph : well=4 generator=0 market=1

Final position

- Agreement!
- Yusef : 2 points
- Joseph : 6 points

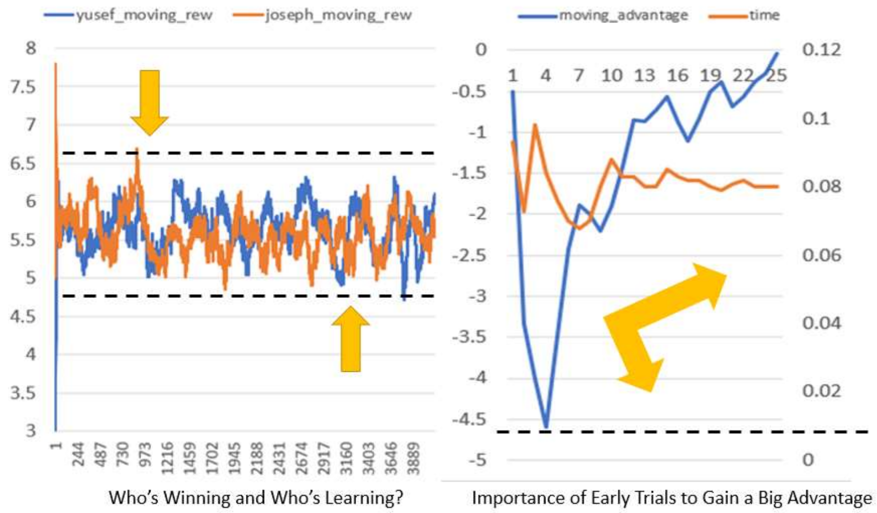
Figure 11. Emergent Machine-Generated Negotiation Strategy Demonstrating Value of Hidden Information. Joseph Tries to Discover Yusef's Position. Later Joseph Understates or Lies about His Hidden Value for the Market.

As illustrated in Figure 11, another variant of world knowledge is the value of hidden information in negotiations. Critically neither agent knows the other's position, either for their quantity of goods or their individualized value. To discover hidden values, one interesting instance of an emergent property is Joseph's first gambit, *"What would you like?"* Joseph discovers that Yusef has wells and markets to offer and immediately accepts the trade giving up nothing. Joseph further accepts a market of no value just to seal the quick bargain after discovering Yusef's initial offer. Yusef also closes the deal with the dangling phrase *"in return"*, even though he gets no value and thus seemingly demonstrates a lack of causal reasoning or logical world knowledge; in truth, Yusef badly loses this round getting nothing in return but keeping his lowest-value item, the generator.

Exploiting Inexperience in the Novice

In addition to demonstrating novel strategies, one key metric of interest is which among the negotiating agents is winning or learning fastest. If either Yusef or Joseph were to acquire some asymmetric knowledge to exploit over multiple bargaining sessions, this might confer some mission advantages. As shown in Figure 12 (left), we scored the self-play over 4,126 sessions and found neither agent was able to outmaneuver the other consistently. In fact, the *pareto* optimality was found to be shared at 51.34%, wherein equal distribution of goods cannot be reallocated to make Joseph better off than Yusef without hurting both.

However, when one party has fewer than twenty-five previous sessions, significant advantage (50% or more) can be accumulated on a temporary basis. As also shown in Figure 12 (right), this outcome translates to a start-up non-equilibrium, particularly in the first five games before both parties settle into some reinforcement learning. Thus, if the village elder, Yusef, has had many encounters previously with other soldiers, and accumulates some tactics in his successful attempts, then he may have operating advantages when dealing with a newly deployed or novice version of Joseph, who has not been adequately prepared.



Who's Winning and Who's Learning?

Importance of Early Trials to Gain a Big Advantage

Figure 12. Neither Side Can Gain a Big Advantage (left) Over Time but Does Exploit Novices (right). Left shows the Moving Rewards Accumulated over Many Trials by Yusef vs. Joseph. Right shows the Importance of Early Trials to Gain a Big Advantage. The Blue Line Shows One Party Gaining 50% More Rewards in the First 5 Negotiations. The Time of a Given Session also Fluctuates to Show a Novice vs. Veteran Negotiator.

CONCLUSIONS

The present example illustrates how to apply newly trained language models to the creation of scripted scenarios, or rule-bending approaches to derive novel variants of a previously known rehearsal narrative. Using generated and evolving tactics that military officers convincingly might see or employ in the field these results can help train negotiators to bargain successfully with non-combatants. These generative agents compare favorably with our own retrieval-based methods trained on the expert-selected (BiLAT) dialogue choices, but without employing subject matter experts for each new scenario. Generative game play spawns many sophisticated strategies, such as bluff, calling the bluff, deceitful lying and hidden value discovery. Even the linguistic and rational errors can provide insights. Humans adopt some of the same contrarian tactics, such as sticking stubbornly to the art of the impossible deal, promises that get taken back or reality bending to "split the baby". The creation of language cues, or short-hand tricks, has parallels with some kinds of familiar communication strategies. Over time, these dueling negotiators eventually reach fair equilibria and equal win-loss distributions. In the near-term, one party (the novice) can be exploited to accumulate substantial (50%) gains for the experienced negotiator.

Within the last half-decade, machine learning has for the first time surpassed human expert performance in broad media categorizations of speech, images, video, animation and text comprehension (see Fadelli, 2019). For natural language, traditional tasks have included machine translation, next word or sentence prediction, and reading comprehension for answering questions. Among the newer language tasks, negotiation and bargaining has only recently been proposed. The challenge becomes to combine both language comprehension and generation steps ("*how to say it?") with the logical requirements of a rewards-based step ("*what to say?")* and the cultural imperatives that support interactions *("how to behave").* If machine learning can master bargaining and negotiation as a new skill, then cheaper, faster and more diverse training simulations should follow. A minimum requirement would be to quadruple the existing NTC negotiating scenarios to fill a two-day training. The need for domain experts should diminish as more general natural language models capture the essential building blocks of conversation, prior to specialization and customized dialogues. For example, our typical negotiations between two semi-cooperative agents has demonstrated easily and automatically how to generate thousands of new and convincingly human dialogues totaling 2,400 pages of narratives per hour.

FUTURE WORK

In the next phase of the project - now underway - we are directly addressing the problems of representing Cultural Worldviews and Norms (CWaN) in these types of AI systems, for automatically deriving responsive models from unstructured text, and then subsequently generating novel training scenarios from such CWaN models. We have done work in the past on CWaN modeling, including for example modeling of threat actors for anticipatory threat reasoning (Regian, 2015) and generating synthetic threat actors for simulation and training (Regian & Noever, 2017). CWaN models are particularly useful for immersive training to engender bargaining and negotiating skills across cultures (Blank, 2013). Cultural modeling enables automated generation (rather than just bending) of training scenarios and real-time coaching during training (Barker, 2014). When a trainee commits an error, the system can pause the simulation and provide explanations such as what norm was violated, why the local agent may be insulted, how to recover, and what lesson should be learned.

Current CWaN modeling presents two unavoidable problems to the training development community: a shortage of cultural experts and a need for new and evolving simulations. Finding persons with deep cultural knowledge can prove difficult, and it typically takes many development hours to produce just a few training hours. When cultural experts are found, they may have no experience with training development and sometimes have limited familiarity with the available advanced technology itself. We anticipate that the narrative generation methods of deep learning will avoid these two constraints and accelerate novel approaches towards the goal of providing next generation resources to train military negotiators.

When Modeling Cultural Context Might Matter?

Our conversational agents, as described in this paper, have no social network or knowledge of cultural norms. This simplification greatly confines their acceptable language models while also limiting exploration of some relevant tactics. For example, among the 4,500 spoken languages, the Arabic language ranks second only to Japanese in terms of its sensitivity to context (Wunderle, 2006). Ironically the exact same word in Arabic can mean "push", "pull" or "negotiate". While the available vocabulary for our negotiating agents is highly constrained, all of which helps Joseph and Yusef to reach agreements, this restriction offers no realistic sampling method to get the much larger universe of all available dialogues. In other words, these agents might speak coherently but may offer fewer insights into some types of field nuances a soldier might encounter.

Previous work has highlighted such practical aspects of improving negotiation training for officers in culturally aware ways. Following extensive officer interviews to extract a list of the most successful tactics, Nobel et al. (2007) found unique features and characteristics in Iraqi negotiations that confound Western expectations. They categorized the tactics within the broader strategies common to both military and civilian negotiations, namely as either a game of "*win-win*" or "*winner take all*". These negotiating tactics also included all the complex elements one might typically expect from corrupt bargains: 1) nepotism and familial clout ("*wasta*"); 2) bribery ("*baksheesh*"); 3) diffuse personal responsibility ("*karma*"), fatalism, and "*inshallah*" ("*God willing*"); 4) extremes of honor/dishonor sensitivities; and 5) a contrived sense of urgency. The authors noted that in typical military settings, this contrived urgency may mean elaborate scheduling antics such as showing up late to assert dominance, showing up early to create confidential back-channels, or introducing last-minute demands after an initial resolution to shorten the possible response window. It is noteworthy that our simulations mimic some elements of this strategy of introducing a last-minute demand. In practical terms, however, soldiers reported that Arab fatalism singularly hindered their schedules. This cultural contrast limited any firm commitment of coalition resources to less than a week at a time, both to maintain schedule and to account for the improbability of any real long-term planning. For one concrete illustration of how such confusing contextual cues can affect scheduling, the polite way for an Arab to say "*no*" is just to say, "*I'll see what I can do*" (IES, 2019).

Our current overall project has generated negotiation and bargaining strategies consistent with a specific cultural context. Here we have shown that automation, by machine learning from already culturally appropriate examples, can generate many and varied negotiation scripts from a few examples (Harrison, et al., 2017). One application of these results is to speed up training development while reducing cost. From a few culturally appropriate negotiation scripts we can generate many scripts with varying goals, participant dispositions and bargaining strategies. We can apply this process to enumerate new training scenarios and reduce the substantial time and cost for expert developers.

ACKNOWLEDGEMENTS

The authors would like to thank the PeopleTec Technical Fellows program for encouragement and project assistance. This research benefited from support from U.S. Army Space and Missile Defense Command/Army Forces Strategic Command.

REFERENCES

Barker, E. (2014). 6 Hostage Negotiation Techniques That Will Get You What You Want, Time Magazine, http://time.com/38796/6-hostage-negotiation-techniques-that-will-get-you-what-you-want/

Blank, J. (2013). How to Negotiate like a Pashtun: A field guide to dealing with the Taliban. Foreign Policy, https://foreignpolicy.com/2013/06/03/how-to-negotiate-like-a-pashtun/

Bradley, T., (2017). Facebook AI Creates Its Own Language in Creepy Preview of Our Potential Future," Forbes, July 31,2017. https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fadelli, I. (2019). An approach for securing audio classification against adversarial attacks, TechXPlore, https://techxplore.com/news/2019-05-approach-audio-classification-adversarial.html

Gonzalez, A. J., & Ahlers, R. (1998). Context-based representation of intelligent behavior in training simulations. Transactions of the Society for Computer Simulation, 15(4), 153-166.

Goodwin, D. (2004). The military and negotiation: The role of the soldier-diplomat. Routledge.

Harrison, B., Purdy, C., & Riedl, M. O. (2017, September). Toward automated story generation with Markov Chain Monte Carlo Methods and Deep Neural Networks. In Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference.

He, H., Chen, D., Balakrishnan, A., & Liang, P. (2018). Decoupling Strategy and Generation in Negotiation Dialogues. arXiv preprint arXiv:1808.09637.

International Education Service (IES), (2019). Cultural Atlas, https://culturalatlas.sbs.com.au/iraqi-culture/iraqi-culture-communication#iraqi-culture-communication.

Jane’s International Defense Review, (2012), Cultural awareness systems help soldiers navigate human terrain, Dec. 2012, p. 62 http://ict.usc.edu/wp-content/uploads/2013/01/Jane-article.pdf

Keizer, S., Guhe, M., Cuayáhuitl, H., Efstathiou, I., Engelbrecht, K. P., Dobre, M., ... & Lemon, O. (2017). Evaluating persuasion strategies and deep reinforcement learning methods for negotiation dialogue agents. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers (Vol. 2, pp. 480-484).

Kim, J. M., Hill Jr, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., ... & Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. International Journal of Artificial Intelligence in Education, 19(3), 289-308.

King, A. (2019), Talk to Transformer, https://talktotransformer.com/

Kojouharov, S. (2016). Ultimate Guide to Leveraging NLP & Machine Learning for your Chatbot, Chatbotlife, 9/18/2016, https://chatbotslife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-you-chatbot

Johns, M. (2007). What was he thinking? Beyond bias to decision-making and judging. 2007 Serious Accident Investigations course, BLM National Training Center, Phoenix, AZ

Lewicki, R. J., & Robinson, R. J. (1998). Ethical and unethical bargaining tactics: An empirical study. Journal of Business Ethics, 17(6), 665-682.

Lewis, M., Yarats, D., Dauphin, Y., Parikh, D., & Batra, D. (2017). Deal or No Deal? End-to-End Learning of Negotiation Dialogues. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2443-2453).

Madan, D., Raghu, D., Pandey, G., & Joshi, S. (2018). Unsupervised Learning of Interpretable Dialog Models. arXiv preprint arXiv:1811.01012.

Nobel, O., Wortinger, B., & Hannah, S. (2007). Winning the war and the relationships: Preparing military officers for negotiations with non-combatants (No. Research Report 1877). Military Academy West Point NY Dept of Behavioral Sciences and Leadership.

Noever, D.A., Regian, J.W. (2017). Deep Learning for Training with Noise in Expert Systems. In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL, 27 Nov – 1 Dec 2017.

Noever, D.A., Regian, J.W. (2018). Machine Supported Entity Resolution in the Cyber Domain. In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL, 26-30 Nov 2018

Pandey, P. (2018) Building a Simple Chatbot from Scratch in Python (using NLTK), Medium, https://medium.com/analytics-vidhya/building-a-simple-chatbot-in-python-using-nltk-7c8c8215ac6e

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1, 8. https://github.com/openai/gpt-2 and https://openai.com/blog/better-language-models/

Regian, J.W. (2012). Formal Modeling of Heterogeneous Social Networks for Human Terrain Analytics. *American Intelligence Journal, Volume 30, Number 2,* 114-119.

Regian, J.W. (2015) Analytic and Predictive Modeling of Cyber Threat Entities. 18th Annual Space & Missile Defense Symposium, 10-13 August 2015, Huntsville, AL.

Regian, J.W., Noever, D.A. (2017). Generative Representation of Synthetic Threat Actors for Simulation and Training. In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL

Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, 19(1), 10-26.

Turing, A. M. (2004). Computing machinery and intelligence (1950). The Essential Turing: The Ideas that Gave Birth to the Computer Age. Ed. B. Jack Copeland. Oxford: Oxford UP, 433-64.

Tressler, D. M. (2007). Negotiation in the new strategic environment: Lessons from Iraq. Army War Coll Carlisle Barracks PA Center for Strategic Leadership.

Wunderle, W. D. (2006). Through the Lens of Cultural Awareness: A Primer for United States Armed Forces Deploying in Arab and Middle Eastern Countries. Government Printing Office.

Van Hasselt, V. B., Romano, S. J., & Vecchi, G. M. (2008). Role playing: Applications in hostage and crisis negotiation skills training. Behavior modification, 32(2), 248-263.

Yarats, D., & Lewis, M. (2018, July). Hierarchical Text Generation and Planning for Strategic Dialogue. In International Conference on Machine Learning (pp. 5587-5595).