A Supervised Topic Model Approach to Learning Effective Styles within Human-Agent Negotiation*

Extended Abstract

Yuyu Xu
Northeastern University
Boston, Massachusetts
yuyuxu@ccs.neu.edu

David Jeong
University of Southern California
Los Angeles, California
davidjeo@usc.edu

Pedro Sequeira
SRI International
Menlo Park, California
pedro.sequeira@sri.com

Jonathan Gratch
Institute for Creative Technologies
Los Angeles, California
gratch@ict.usc.edu

Javed Aslam
Northeastern University
Boston, Massachusetts
jaa@ccs.neu.edu

Stacy Marsella
Northeastern University
Boston, Massachusetts
marsella@ccs.neu.edu

ABSTRACT

We present a method that analyzes a person's negotiation behavior to automatically detect co-occurrence of tactics and combination of tactics (i.e., negotiation styles). We first identify action features consistent with use of the common negotiation tactics based on prior research in negotiation. Next, we apply regularized linear regression over a negotiation dataset to assess how effective particular tactics are in predicting the negotiation outcome. Finally, we use a supervised variant of a topic model to derive effective negotiation styles. Results from the clusters produced by the topic models provide insights regarding the effectiveness of negotiation styles that people utilize.

KEYWORDS

Explainability in human-agent systems; Socially interactive agents; Agents competing and collaborating with humans

ACM Reference Format:

Yuyu Xu, David Jeong, Pedro Sequeira, Jonathan Gratch, Javed Aslam, and Stacy Marsella. 2020. A Supervised Topic Model Approach to Learning Effective Styles within Human-Agent Negotiation. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

We live in a world of bounded resources, whether it is constraints on our time, actions, material things, or space. Due to these constraints, conflicts between people naturally follow. Negotiation plays a central role to resolve these conflicts, in both our professional and personal lives. Not surprisingly, this has led to a wide range of research on computationally modeling negotiation. Modeling how people negotiate in a negotiation interaction poses multiple challenges. First, negotiation is a complex, highly dynamic social interaction [6]. Second, different people might use different tactics during a negotiation interaction [13, 15], some of them may be more effective than others.

In the work reported here, we look at the question of how negotiators may employ multiple tactics over the course of a negotiation. Specifically, we ask the question of what tactics tend to occur in a single negotiation as part of an overall strategy and what particular combination of tactics were successful. We call such combination of tactics *negotiation styles.* A three step process is proposed to answer this broad question, they are: identifying tactic-related features; assessing tactic-related features and learning effective negotiation styles, as elaborated below.

2 AUTOMATIC DETECTION PROCESS

The dataset used in this work is described briefly, followed by the three step processes.

Negotiation Dataset. We adopt the dataset from [10], which was collected through the IAGO framework [9] and contains 504 negotiation sessions (maximum length being 10 minutes), each between a human participant and an agent. Each human participant was limited to one negotiation with 1 of the 8 agent types, differing across the three dimensions, namely cognitive bias, anchoring, emotion.

The two parties negotiate over 4 items, each with different amounts and assigned payoffs. At the end of each turn, both parties get the payoffs for the items in their hand, which defines the negotiation outcome. To reflect some of the complexity of real world negotiation, Iago framework supports many different combinations of actions. For example, an *offer* is a partial split between four items that may occur simultaneously. *Preference* is communicated with Q&A, such as "do you like x better than y?". *Verbal influence* is conveyed by sending positive messages ("it is important that we are both happy with an agreement") or negative messages ("your offer sucks"). *Emotional* information is communicated via selection of emoticons on the interface. Finally, *agreement* is determined by participants' acceptance of the offer by both (formal agreement) or one party (non-binding agreement).

2.1 Identifying Tactic-related Features

To identify tactic-related features, we propose a compact annotation schema inspired by negotiation literature designed which here we use to quantify the sequences of negotiation action in the IAGO dataset. A single negotiation action is referred to as a *unigram*, while

*This work was largely done while the author studies at Northeastern University and was completed while author works for Google.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

Human Outcome					
Effect of cluster to outcome	$c_7(w = 0.39)$	$c_5(w = -0.38)$	$c_0(w = 0.26)$	$c_8(w = -0.25)$	$c_6(w = -0.12)$
Tactics	Preference_ask:0.32 Preference_tell:0.15 Emotion_happy:0.15 Verbal_positive:0.11	Logrolling_h_lt_a:0.18 Preference_tell:0.16 Logrolling_h_eq_a:0.12 Verbal_positive:0.11	Logrolling_h_gt_a:0.28 Preference_ask:0.13 NonbindingAgreement_reject:0.11 Emotion_angry:0.09	NonbindingAgreement_reject:0.23 MultiIssue_true:0.16 Verbal_positive:0.13 Verbal_negative:0.12	Emotion_neutral:0.27 Preference_misclick:0.25 Emotion_angry:0.10 Verbal_positive:0.07
Negotiation styles	"PreferenceExchange_Positive"	"Positive_NonAggressive"	"Aggressive"	"Rejection"	"Manipulative_Careless"
Joint Outcome					
Effect of cluster to joint outcome	$c_8(w = 0.33)$	$c_5(w = 0.31)$	$c_1(w = 0.17)$	$c_6(w = 0.11)$	$c_9(w = -0.07)$
Tactics	Preference_ask:0.53 Preference_tell:0.19 Emotion_happy:0.18 Verbal_positive:0.05	Preference_ask:0.17 Preference_tell:0.17 Verbal_positive:0.11 Logrolling_h_gt_a:0.10	Verbal_positive:0.16 Emotion_surprised:0.15 Concession_above_avg:0.15 Emotion_neutral:0.12	MultiIssue_true:0.35 Preference_tell:0.13 Preference_ask:0.10 Verbal_positive:0.10	Emotion_happy:0.21 Verbal_positive:0.15 Emotion_neutral:0.10 Concession_below_avg:0.09
Negotiation styles	"PreferenceExchange_Positive"	"PreferenceExchange_Positive_Strong"	"Positive_Concession_Above"	"PreferenceExchange_MultiIssue"	"Positive"

Table 1: Supervised LDA produced tactics cluster based on human side tactic-related features.

three consecutive actions are referred to as a *trigram*, *e.g.*, consisting of an offer, a counter-offer, and another offer of the first negotiator. We assume that each tactic may potentially be realized by behavior patterns of unigrams or trigrams, *i.e.*, *tactic-related features*. Table 2 provides detailed descriptions of the data schema and relates tactics based on relevant theories to classes of discretized actions.

Our annotation schema connects tactics with particular actions and interpret the results. Specifically, each tactic is realized by a set of *tactic-related features*. *E.g.*, tactic *Emotion Expressivity* comprises a set of features {Happy, Sad, Neutral, Surprise, Angry}, each of which is realized by the behavior patterns of unigrams communicated across different channels. *E.g.*, feature *Happy* is realized by counting the occurrences of unigram "sending an happy emoticon", conveyed during a negotiation through the emotion channel.

2.2 Assessing Tactic-related Features

We adopt a supervised learning model, where the *features* for this model are the occurrences of each distinct sub-sequence (unigrams or trigrams, from human side or both sides), and the *labels* are either human outcome or joint outcome. Specifically, we used Linear Regression with ElasticNet [16] regularizer, as an effective solution to tackle the challenge that the feature matrix might be sparse. The feature matrix is composed of the occurrences of distinct sub-sequences within a single negotiation sequence and some of them might be non-existent given how rich this dataset is. To train model, we use K-folds cross-validation ($K = 10$) and model selection.

Tactics based on Relevant Theories	Tactic-related Features	Actions
Concession in Offer [1]	Concession_below_avg Concession_avg Concession_above_avg	offer_one_sided:below_avg offer_one_sided:avg offer_one_sided:above_avg
Logrolling [7]	Logrolling_h_eq_a Logrolling_h_lt_a Logrolling_h_gt_a	offer_two_sided:h_below_avg_a_below_avg offer_two_sided:h_avg_a_avg offer_two_sided:h_above_avg_a_above_avg offer_two_sided:h_below_avg_a_avg offer_two_sided:h_below_avg_a_above_avg offer_two_sided:h_avg_a_above_avg offer_two_sided:h_avg_a_below_avg offer_two_sided:h_above_avg_a_below_avg offer_two_sided:h_above_avg_a_avg
Preference Exchange [14]	Preference_tell Preference_tell_untruthful Preference_ask Preference_misclick	preference_tell:critical preference_tell:comparative preference_tell:untruthful preference_ask:critical preference_ask:comparative preference_tell:misclick preference_ask:misclick
Verbal Influence [12]	Verbal_positive Verbal_negative	verbal_influence:positive verbal_influence:negative
Emotion Expressivity [11]	Emotion_angry Emotion_neutral Emotion_happy Emotion_sad Emotion_surprised	emoticon:angry emoticon:neutral emoticon:happy emoticon:sad emoticon:surprised
MultiIssue Offer [5]	MultiIssue_false MultiIssue_true	offer_two_sided:zero_item offer_two_sided:one_item offer_two_sided:two_items offer_two_sided:three_items offer_two_sided:four_items
Anchoring [2]	Anchoring_size	offer_one_sided:initial_payoff
Nonbinding Agreement	Nonbinding_accept Nonbinding_reject	agreement:nonbinding_accept agreement:nonbinding_reject

Table 2: Data schema. Nonbinding Agreement (last row) is based on observational heuristics of the IAGO dataset. ("h": human, "a": agent, "avg": average, "lt": less than, "gt": greater than, "eq": equal).

2.3 Learning Effective Negotiation Styles

After identifying the most predictive features, we can determine what negotiation styles exist, as well as their effectiveness. To achieve that, we resorted to a topic model, using a well-established solution: Latent Dirichlet Allocation (LDA) [3],where the *features* are the occurrences of each distinct sub-sequence, and the *labels* correspond to the outcomes (human, joint). LDA is a natural language processing-based probabilistic unsupervised learning method for modeling topics. This model assumes a generative process, where each document (in our case a negotiation sequence) in a corpus (*i.e.*, all negotiation sequences) is composed of collections of words or phrases (sub-sequences), each generated from multiple underlying topics (clusters of the distinct sub-sequences). The underlying assumption of LDA fits our objective of automatically clustering features, where the clusters characterize *styles* of the negotiation. To address the limited effectiveness of capturing relations between clusters and outcomes using (unsupervised) LDA, we also adopted its supervised variant (sLDA) [4, 8].

3 DEMONSTRATION

Our results show that the identified tactic-related features are able to achieve a much higher performance in outcome prediction as compared to the random guess baseline (up to 30% boost). This demonstrates that we could systematically acquire a predictive set of tactic-related features. Furthermore, our models can characterize the distinguishing negotiation styles through co-occurrences of the tactics, as well as show how effective these clusters are in terms of predicting outcome (see Table 1). For example, one produced cluster reveals the negotiation style called "Aggressive" because it contains tactics: *log-rolls where human gets more payoff as compared to agent*; *human asks issue preference from the agent*; *human reject agent's offer* and *human showing angry emoticon*. These tactics combined suggest an aggressive strategy. This "Aggressive" style has a positive effect on human's negotiation outcome (*i.e.*, human adopting aggressive strategy could end up having more payoff). Although our techniques were applied to a particular human-agent dataset, we see the approach to be more broadly applicable to human-human interactions.

REFERENCES

[1] Alan A. Benton, Harold H. Kelley, and Barry Liebling. 1972. Effects of Extremity of Offers and Concession Rate on the Outcomes of Bargaining. *Journal of Personality and Social Psychology* 24 (10 1972), 73–83.

[2] Thomas Mussweiler Adam D. Galinsky. 2001. First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology* (2001).

[3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022. http://dl.acm.org/citation.cfm?id=944919.944937

[4] Angelos Katharopoulos, Despoina Paschalidou, Christos Diou, and Anastasios Delopoulos. 2016. Fast Supervised LDA for Discovering Micro-Events in Large-Scale Video Datasets. In *Proceedings of the 2016 ACM on Multimedia Conference (MM '16)*. ACM, New York, NY, USA, 332–336. https://doi.org/10.1145/2964284.2967237

[5] H.H. Kelley. 1996. *A Classroom Study of the Dilemmas in Interpersonal Negotiations*. Berkeley Institute of International Studies. https://books.google.com/books?id=ofBjAQAACAAJ

[6] Min Li, Leigh Plunkett Tost, and Kimberly Wade-Benzoni. 2007. The dynamic interaction of context and negotiator effects: A review and commentary on current and emerging areas in negotiation. *International Journal of Conflict Management* 18, 3 (2007), 222–259. https://doi.org/10.1108/10444060710825981 arXiv:https://doi.org/10.1108/10444060710825981

[7] Jeffrey Loewenstein and Leigh Thompson. 2006. *Learning to Negotiate: Novice and Experienced Negotiators*. Psychology Press, 77–97.

[8] Jon D Mcauliffe and David M Blei. 2008. Supervised topic models. In *Advances in neural information processing systems*. 121–128.

[9] Johnathan Mell and Jonathan Gratch. 2017. Grumpy & Pinocchio: Answering Human-Agent Negotiation Questions Through Realistic Agent Design. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS '17)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 401–409. http://dl.acm.org/citation.cfm?id=3091125.3091186

[10] Sarah Roediger. 2018. The Effect of Suspicion on Emotional Influence Tactics in Virtual Human Negotiation.

[11] Joanna Schug, David Matsumoto, Yutaka Horita, Toshio Yamagishi, and Kemberlee Bonnet. 2018. Emotional expressivity as a signal of cooperation.

[12] Marwan Sinaceur, Gerben A van Kleef, Margaret A. Neale, Hajo Adam, and Christophe Haag. 2011. Hot or cold: is communicating anger or threats more effective in negotiation? *The Journal of applied psychology* 96 5 (2011), 1018–32.

[13] Leigh Thompson. 1990. Negotiation Behavior and Outcomes: Empirical Evidence and Theoretical Issues. *Psychological Bulletin* 108, 3 (1 1 1990), 515–532. https://doi.org/10.1037/0033-2909.108.3.515

[14] Leigh L Thompson. 1991. Information exchange in negotiation. *Journal of Experimental Social Psychology* 27, 2 (1991), 161 – 179. https://doi.org/10.1016/0022-1031(91)90020-7

[15] Laurie R Weingart, Leigh L Thompson, Max H Bazerman, and John S Carroll. 1990. Tactical behavior and negotiation outcomes. *International Journal of Conflict Management* 1, 1 (1990), 7–31.

[16] Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67 (2005), 301–320.