changTroubleHorizonForecasting2019_Analysis

Definitions

- "conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2 #definition , #conversational-forecasting
- "Antisocial behavior online comes in many forms, including harassment (Vitak et al., 2017), cyberbullying (Singh et al., 2017), and general aggression (Kayany, 1998)." Page 3 #definition
  - antisocial behavior #definition
- "utterance encoder is responsible for generating semantic vector representations of comments" Page 5 #model-architecture
- "This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 #definition , #model-architecture , #to-learn

Applications

Technical Details

- "The former can be pre-trained on large amounts of unsupervised data, similarly to how chatbots are trained. The" Page 2 #application
  - train on dispute data #application
- "latter can piggy-back on the resulting representation after fine-tuning it for classification using relatively small labeled data." Page 2 #conversational-forecasting
  - deploy on dispute data after using supervised learning representation #conversational-forecasting
- " "sequence of N comments C = {c1, . . . , cN }" " Page 5
- "cn = {w1, . . . , wMn}" Page 5 #formula
  - representation of a conversation in a conversation, where each w_i is a "comment" with varying length of n tokens #formula
- "henc m = f RNN(henc m−1, wm)" Page 5 #formula
  - RNN model for utterance decoder. Input is prev hidden state and current token. output is the hidden comment (m<= M tokens) #formula
- "hcon n = f RNN(hcon n−1, henc Mn)" Page 5 #formula , #model-architecture
  - encoder for ALL comments. pass in previous comment encoding. #formula , #model-architecture

- "htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5 #formula , #model-architecture
    - decoder for the (up to) nth comment.

The h_t will be used multiple times as tokens have multiple comment dependency, and each h_t is per comment.

Generates probabilistic output of words (predictions of future words based on context?) #formula , #model-architecture

Conceptual

- "unsupervised representation of conversational dynamics and exploits it to predict future derailment as the conversation develops." Page #research-question , #conversational-forecasting
    - Unsupervised model #research-question , #conversational-forecasting
- "nly identify antisocial content after the fact limits their practicality as tools for assisting pre-emptive moderation in conversational domains." Page #critique , #conversational-forecasting
    - Current state of research has not explored real time moderation as in depth as post-hoc analysis case. #critique , #conversational-forecasting
- "dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2 #challenge , #potential-gap , #conversational-forecasting
    - What are metrics for interaction in conversations?

Could there be a mapping from interaction scores to the emotional labeling that could potentially be beneficial? #challenge , #potential-gap , #conversational-forecasting

- "Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2 #challenge , #potential-gap , #conversational-forecasting
    - first limitation Complexity of knowledge.

Emotional labeling did not need context to predict the dispute outcome in post-hoc, but would context be required in real-time?

We did have in-context learning show better performance-- depends on what we are trying to forecast.

Context will probably be needed to craft the appropriate mediation response. #challenge , #potential-gap , #conversational-forecasting

- "conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2 #challenge , #conversational-forecasting
- "One solution is to assume (unrealistic) prior knowledge of when the to-be-forecasted event takes place and extract features up to that point (Niculae et al., 2015; Liu et al., 2018)." Page 2 #conversational-forecasting , #previous-work
- "Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2 #question , #fairness , #conversational-forecasting , #previous-work
    - Generalizabilitity to dispute theory? We had in-context learning of emotional-labels.

What are fairness considerations in using prior knowledge? #question , #fairness , #conversational-forecasting , #previous-work

- "The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 #challenge , #dispute-theory , #conversational-forecasting
    - What labels do we need for dispute context? #challenge , #dispute-theory , #conversational-forecasting
- "we propose to decouple the objective of learning a neural representation of conversational dynamics from the objective of predicting future events." Page 2 #conversational-forecasting , #research-question
    - Focus on creating the neural representation of conversation dynamics rather than needing to predict events based on labels #conversational-forecasting , #research-question
- "In both datasets, our model outperforms existing fixed-window approaches, as well as simpler sequential baselines that cannot account for inter-comment relations." Page 3 #result
- "utomatically learn neural representations of conversational dynamics through pre-training." Page 3 #goal
- "do not address the issue of aggregating them into a single forecast (i.e., deciding at what point to make 4746 a prediction)" Page 3 #challenge
- "up performing at the level of random guessing.10 This result underscores the need for the pretraining step that can make use of unlabeled data." Page 7 #result
- "both a visibly higher precision-recall curve and larger area under the curve (AUPR) than the baselines" Page 7 #result
    - CRAFT performance #result
- "1) How much early warning does the the model provide? (2) Does the model actually" Page 7 #goal
- "parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8 #challenge , #dispute-theory , #fairness , #potential-gap

    - application of fairness? #challenge , #dispute-theory , #fairness , #potential-gap
- "his model fills a void in the existing literature on conversational forecasting, simultaneously addressing the dual challenges of capturing inter-comment dynamics and dealing with an unknown horizon." Page 9 #contribution , #conversational-forecasting

Personal Insights

- "First, since conversations are dynamic, a forecasting model needs to capture the flow of the discussion, rather than properties of individual comments. Second, real conversations have an unknown horizon: they can end or derail at any time; thus a practical forecasting model needs to assess the risk in an online fashion, as the conversation develops" Page #idea , #robust-optimization
    - Perhaps connect to multi-stage optimization with robust optimization? Also, how to consider "fair" assessment of risk given cross cultural differences? #idea , #robust-optimization
- "window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 #question , #dispute-theory , #conversational-forecasting
    - What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? #question , #dispute-theory , #conversational-forecasting
- "longer windows risk missing the to-be-forecasted event altogether" Page 2 #question , #confusion , #conversational-forecasting
    - why do longer windows miss an event? #question , #confusion , #conversational-forecasting
- "while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2 #question , #conversational-forecasting
    - could something here be used as guide for the mediation response generation? #question , #conversational-forecasting
- " "goal of assisting human moderators of online communities by preemptively signaling at-risk conversations that might deserve their attention." " Page 3
- "This is a useful property for the purposes of model analysis, and hence we focus on this as our primary dataset." Page 4 #question
    - considerations of hand annotated labels? #question
- "attack-containing conversation is paired with a clean conversation from the same talk page, where the talk page serves as a proxy for topic.3" Page 4 #dispute-theory , #Further-exploration-needed , #unsupervised-learning
    - look into techniques for correlation control-- what kinds of structures to consider for the text? #dispute-theory , #Further-exploration-needed , #unsupervised-learning
- "nonlinear gating function (our implementation uses GRU (Cho et al., 2014))." Page 5 #model-architecture , #to-learn

- what is gating function? #model-architecture , #to-learn
- "hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5 #model-architecture , #question
  - Can we vary token length to be more than 1 word #model-architecture , #question
- "order-sensitive representation of conversational context?" Page 8 #conversational-forecasting , #dispute-theory , #question
  - to what extent does order matter for disputes? #conversational-forecasting , #dispute-theory , #question
- "ntuition that comments in a conversation are not independent events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8 #dispute-theory , #idea , #question
  - do we care about patterns for learning dispute structure? Relation to emotional recognition? #dispute-theory , #idea , #question
- "including those for which the outcome is extraneous to the conversation." Page 9 #conversational-forecasting , #question
  - what does this mean? the impact is outside of the actual conversation? #conversational-forecasting , #question
- "A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9 #dispute-theory , #question
  - does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were? #dispute-theory , #question
- "n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9 #conversational-forecasting , #dispute-theory , #potential-gap
  - What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached? #conversational-forecasting , #dispute-theory , #potential-gap

Literary Note To Lookup Later

- "Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , #to-read , #dispute-theory , #conversational-forecasting
  - compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , #to-read , #dispute-theory , #conversational-forecasting

- "Addressing this limitation requires forecasting the future derailment of a conversation based on early warning signs, giving the moderators time to potentially intervene before any harm is done (Liu et al. 2018, Zhang et al. 2018a, see Jurgens et al. 2019 for a discussion)." Page #to-read
  - forecasting real-time antisocial behavior-- described the "warning signs" #to-read
- "similarity between comments (Althoff et al., 2016; Tan et al., 2016) or conversation structure (Zhang et al., 2018b; Hessel and Lee, 2019)—, though neural attention architectures have also recently shown promise (Jo et al., 2018)." Page 2 #to-read , #similarity
- "sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2 #question , #to-read , #conversational-forecasting
  - reference for in-context generation #question , #to-read , #conversational-forecasting
- "any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3 #conversational-forecasting , #fairness , #to-read
  - How do the cautionary warnings in these papers relate to training such a system on dispute literature? #conversational-forecasting , #fairness , #to-read
- "identifying successful negotiations (Curhan and Pentland, 2007; Cadilhac et al., 2013)," Page 3 #negotiation
  - classification of negotiation success post-hoc #negotiation
- "deception (Girlea et al., 2016; P´erez-Rosas et al., 2016; Levitan et al., 2018)" Page 3 #dispute-theory
  - deception classification post-hoc #dispute-theory
- "disagreement (Galley et al., 2004; Abbott et al., 2011; Allen et al., 2014; Wang and Cardie, 2014; Rosenthal and McKeown, 2015)" Page 3 #dispute-theory
  - disagreement classification post-hoc #dispute-theory
- "predicting whether an ongoing conversation will eventually spark disagreement (Hessel and Lee, 2019)" Page 3 #conversational-forecasting
  - download subreddits, do some posyt-hoc analysis of controversial vs. non-controversial labels. the downloaded comment threads may not have currently had a controversial post. #conversational-forecasting
- "statistical measures based on similarity between utterances (Althoff et al., 2016)" Page 3 #conversational-forecasting
- "sentiment imbalance (Niculae et al., 2015)" Page 3 #conversational-forecasting
- "increase in hostility (Liu et al., 2018)" Page 3 #conversational-forecasting
- "graph representations of conversations (Garimella et al., 2017; Zhang et al., 2018b)" Page 3 #conversational-forecasting , #personal-interest

- "hierarchical recurrent encoder-decoder (HRED) architecture (Sordoni et al., 2015a),"
  Page 5 #conversational-forecasting
  - model architechture for generative model in forecatsing; dependency consideration for sequences of tokens #conversational-forecasting

---

🧠 Key Ideas

-

⚙️ Project Relations

-

📄 Related Literature

-

🏷️ Tags

Unique tag Groups

#idea , #robust-optimization
#research-question , #conversational-forecasting
#question , #to-read , #dispute-theory , #conversational-forecasting
#critique , #conversational-forecasting
#to-read
#definition , #conversational-forecasting
#challenge , #potential-gap , #conversational-forecasting
#to-read , #similarity
#challenge , #conversational-forecasting
#conversational-forecasting , #previous-work
#question , #fairness , #conversational-forecasting , #previous-work

#question , #dispute-theory , #conversational-forecasting
#question , #confusion , #conversational-forecasting
#question , #to-read , #conversational-forecasting
#question , #conversational-forecasting
#challenge , #dispute-theory , #conversational-forecasting
#conversational-forecasting , #research-question
#application
#conversational-forecasting
#data-artifacts
#result
#methodology
#conversational-forecasting , #fairness , #to-read
#definition
#negotiation
#dispute-theory
#conversational-forecasting , #personal-interest
#goal
#challenge
#question
#dispute-theory , #Further-exploration-needed , #unsupervised-learning
#data-processing
#model-architecture
#formula
#model-architecture , #to-learn
#formula , #model-architecture
#model-architecture , #question
#data-processing , #question , #to-learn
#definition , #model-architecture , #to-learn
#data-analysis
#conversational-forecasting , #dispute-theory , #question
#dispute-theory , #idea , #question
#challenge , #dispute-theory , #fairness , #potential-gap
#contribution , #conversational-forecasting
#conversational-forecasting , #question
#dispute-theory , #question
#conversational-forecasting , #dispute-theory , #potential-gap

Tags Groups within Current Paper

#idea , #robust-optimization

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"First, since conversations are dynamic, a forecasting model needs to capture the flow of the discussion, rather than properties of individual comments. Second, real conversations have an unknown horizon: they can end or derail at any time; thus a practical forecasting model needs to assess the risk in an online fashion, as the conversation develops" Page
changTroubleHorizonForecasting2019_Analysis	Perhaps connect to multi-stage optimization with robust optimization? Also, how to consider "fair" assessment of risk given cross cultural differences?

#research-question , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"unsupervised representation of conversational dynamics and exploits it to predict future derailment as the conversation develops." Page
changTroubleHorizonForecasting2019_Analysis	Unsupervised model

#question , #to-read , #dispute-theory , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other?

#critique , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"nly identify antisocial content after the fact limits their practicality as tools for assisting pre-emptive moderation in conversational domains." Page
changTroubleHorizonForecasting2019_Analysis	Current state of research has not explored real time moderation as in depth as post-hoc analysis case.

#to-read

Paper (9)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"Addressing this limitation requires forecasting the future derailment of a conversation based on early warning signs, giving the moderators time to potentially intervene before any harm is done (Liu et al. 2018, Zhang et al. 2018a, see Jurgens et al. 2019 for a discussion)." Page
changTroubleHorizonForecasting2019_Analysis	forecasting real-time antisocial behavior-- described the "warning signs"
changTroubleHorizonForecasting2019_Analysis	"similarity between comments (Althoff et al., 2016; Tan et al., 2016) or conversation structure (Zhang et al., 2018b; Hessel and Lee, 2019)—, though neural attention architectures have also

Paper (9)	Related Annotations
	recently shown promise (Jo et al., 2018)." Page 2 , #similarity
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2 #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3 #conversational-forecasting , #fairness ,
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature? #conversational-forecasting , #fairness ,

#definition , #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2

#challenge , #potential-gap , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each

Paper (2)	Related Annotations
	individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2

#to-read, #similarity

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"similarity between comments (Althoff et al., 2016; Tan et al., 2016) or conversation structure (Zhang et al., 2018b; Hessel and Lee, 2019)—, though neural attention architectures have also recently shown promise (Jo et al., 2018)." Page 2

#challenge, #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2

#conversational-forecasting, #previous-work

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"One solution is to assume (unrealistic) prior knowledge of when the to-be-forecasted event takes place and extract features up to that point (Niculae et al., 2015; Liu et al., 2018)." Page 2
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the

Paper (2)	Related Annotations
	conversation" Page 2 #question , #fairness ,

#question , #fairness , #conversational-forecasting , #previous-work

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2

#question , #dispute-theory , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window?

#question , #confusion , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event?

#question , #to-read , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation

#question , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation?

#challenge , #dispute-theory , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context?

#conversational-forecasting , #research-question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"we propose to decouple the objective of learning a neural representation of

Paper (2)	Related Annotations
	conversational dynamics from the objective of predicting future events." Page 2
changTroubleHorizonForecasting2019_Analysis	Focus on creating the neural representation of conversation dynamics rather than needing to predict events based on labels

#application

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"The former can be pre-trained on large amounts of unsupervised data, similarly to how chatbots are trained. The" Page 2
changTroubleHorizonForecasting2019_Analysis	train on dispute data

#conversational-forecasting

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2 #definition ,
changTroubleHorizonForecasting2019_Analysis	"latter can piggy-back on the resulting representation after fine-tuning it for classification using relatively small labeled data." Page 2
changTroubleHorizonForecasting2019_Analysis	deploy on dispute data after using supervised learning representation
changTroubleHorizonForecasting2019_Analysis	"unsupervised representation of conversational dynamics and exploits it to predict future derailment as the conversation develops." Page #research-question ,
changTroubleHorizonForecasting2019_Analysis	Unsupervised model #research-question ,

Paper (43)	**Related Annotations**
changTroubleHorizonForecasting2019_Analysis	"nly identify antisocial content after the fact limits their practicality as tools for assisting pre-emptive moderation in conversational domains." Page #critique ,
changTroubleHorizonForecasting2019_Analysis	Current state of research has not explored real time moderation as in depth as post-hoc analysis case. #critique ,
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2 #challenge , #potential-gap ,
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2 #challenge , #potential-gap ,
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2 #challenge ,
changTroubleHorizonForecasting2019_Analysis	"One solution is to assume (unrealistic) prior knowledge of when the to-be-forecasted event takes place and extract features up to that point (Niculae et al., 2015; Liu et al., 2018)." Page 2 , #previous-work
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2 #question , #fairness , , #previous-work
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 #challenge , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? #challenge , #dispute-theory ,

Paper (43)	**Related Annotations**
changTroubleHorizonForecasting2019_Analysis	"we propose to decouple the objective of learning a neural representation of conversational dynamics from the objective of predicting future events." Page 2 , #research-question
changTroubleHorizonForecasting2019_Analysis	Focus on creating the neural representation of conversation dynamics rather than needing to predict events based on labels , #research-question
changTroubleHorizonForecasting2019_Analysis	"his model fills a void in the existing literature on conversational forecasting, simultaneously addressing the dual challenges of capturing inter-comment dynamics and dealing with an unknown horizon." Page 9 #contribution ,
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 #question , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? #question , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2 #question , #confusion ,
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event? #question , #confusion ,
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2 #question ,
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation? #question ,
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 , #dispute-theory , #question
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? , #dispute-theory , #question

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9 , #question
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation? , #question
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9 , #dispute-theory , #potential-gap
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached? , #dispute-theory , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , #to-read , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , #to-read , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2 #question , #to-read ,
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation #question , #to-read ,
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3 , #fairness , #to-read

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature? , #fairness , #to-read
changTroubleHorizonForecasting2019_Analysis	"predicting whether an ongoing conversation will eventually spark disagreement (Hessel and Lee, 2019)" Page 3
changTroubleHorizonForecasting2019_Analysis	download subreddits, do some posyt-hoc analysis of controversial vs. non-controversial labels. the downloaded comment threads may not have currently had a controversial post.
changTroubleHorizonForecasting2019_Analysis	"statistical measures based on similarity between utterances (Althoff et al., 2016)" Page 3
changTroubleHorizonForecasting2019_Analysis	"sentiment imbalance (Niculae et al., 2015)" Page 3
changTroubleHorizonForecasting2019_Analysis	"increase in hostility (Liu et al., 2018)" Page 3
changTroubleHorizonForecasting2019_Analysis	"graph representations of conversations (Garimella et al., 2017; Zhang et al., 2018b)" Page 3 , #personal-interest
changTroubleHorizonForecasting2019_Analysis	"hierarchical recurrent encoder-decoder (HRED) architecture (Sordoni et al., 2015a)," Page 5
changTroubleHorizonForecasting2019_Analysis	model architechture for generative model in forecatsing; dependency consideration for sequences of tokens

#data-artifacts

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#result

Paper (4)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"In both datasets, our model outperforms existing fixed-window approaches, as well as simpler sequential baselines that cannot account for inter-comment relations." Page 3
changTroubleHorizonForecasting2019_Analysis	"up performing at the level of random guessing.10 This result underscores the need for the pretraining step that can make use of unlabeled data." Page 7
changTroubleHorizonForecasting2019_Analysis	"both a visibly higher precision-recall curve and larger area under the curve (AUPR) than the baselines" Page 7
changTroubleHorizonForecasting2019_Analysis	CRAFT performance

#methodology

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#conversational-forecasting, #fairness, #to-read

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature?

#definition

Paper (4)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior online comes in many forms, including harassment (Vitak et al., 2017), cyberbullying (Singh et al., 2017), and general aggression (Kayany, 1998)." Page 3
changTroubleHorizonForecasting2019_Analysis	antisocial behavior
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 , #model-architecture , #to-learn

#negotiation

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"identifying successful negotiations (Curhan and Pentland, 2007; Cadilhac et al., 2013)," Page 3
changTroubleHorizonForecasting2019_Analysis	classification of negotiation success post-hoc

#dispute-theory

Paper (22)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 #challenge , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? #challenge , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8 #challenge , , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	application of fairness? #challenge , , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"attack-containing conversation is paired with a clean conversation from the same talk page, where the talk page serves as a proxy for topic.3" Page 4 , #Further-exploration-needed , #unsupervised-learning
changTroubleHorizonForecasting2019_Analysis	look into techniques for correlation control- - what kinds of structures to consider for the text? , #Further-exploration-needed , #unsupervised-learning
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting , , #question
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting , , #question
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent events; rather, the

Paper (22)	Related Annotations
	order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8 , #idea , #question
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition? , #idea , #question
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9 , #question
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were? , #question
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9 #conversational-forecasting , , #potential-gap
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached? #conversational-forecasting , , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , #to-read , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , #to-read , , #conversational-forecasting

Paper (22)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"deception (Girlea et al., 2016; P´erez-Rosas et al., 2016; Levitan et al., 2018)" Page 3
changTroubleHorizonForecasting2019_Analysis	deception classification post-hoc
changTroubleHorizonForecasting2019_Analysis	"disagreement (Galley et al., 2004; Abbott et al., 2011; Allen et al., 2014; Wang and Cardie, 2014; Rosenthal and McKeown, 2015)" Page 3
changTroubleHorizonForecasting2019_Analysis	disagreement classification post-hoc

#conversational-forecasting , #personal-interest

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"graph representations of conversations (Garimella et al., 2017; Zhang et al., 2018b)" Page 3

#goal

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"utomatically learn neural representations of conversational dynamics through pre-training." Page 3
changTroubleHorizonForecasting2019_Analysis	"1) How much early warning does the the model provide? (2) Does the model actually" Page 7

#challenge

Paper (8)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2 , #potential-gap , #conversational-forecasting

Paper (8)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2 , #potential-gap , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"do not address the issue of aggregating them into a single forecast (i.e., deciding at what point to make 4746 a prediction)" Page 3
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8 , #dispute-theory , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	application of fairness? , #dispute-theory , #fairness , #potential-gap

#question

Paper (23)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the

Paper (23)	**Related Annotations**
	conversation" Page 2 , #fairness , #conversational-forecasting , #previous-work
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2 , #confusion , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event? , #confusion , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation? , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"This is a useful property for the purposes of model analysis, and hence we focus on this as our primary dataset." Page 4
changTroubleHorizonForecasting2019_Analysis	considerations of hand annotated labels?
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5 #model-architecture ,
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word #model-architecture ,
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting , #dispute-theory ,

Paper (23)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8 #dispute-theory , #idea ,
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition? #dispute-theory , #idea ,
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9 #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation? #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9 #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were? #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page , #to-read , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? , #to-read , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban

Paper (23)	Related Annotations
	et al., 2016, 2017)," Page 2 , #to-read , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation , #to-read , #conversational-forecasting

#dispute-theory , #Further-exploration-needed , #unsupervised-learning

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"attack-containing conversation is paired with a clean conversation from the same talk page, where the talk page serves as a proxy for topic.3" Page 4
changTroubleHorizonForecasting2019_Analysis	look into techniques for correlation control- - what kinds of structures to consider for the text?

#data-processing

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#model-architecture

Paper (9)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"utterance encoder is responsible for generating semantic vector representations of comments" Page 5

Paper (9)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 #definition , , #to-learn
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5 #formula ,
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding. #formula ,
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5 #formula ,
changTroubleHorizonForecasting2019_Analysis	"nonlinear gating function (our implementation uses GRU (Cho et al., 2014))." Page 5 , #to-learn
changTroubleHorizonForecasting2019_Analysis	what is gating function? , #to-learn
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5 , #question
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word , #question

#formula

Paper (7)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"cn = {w1, . . . , wMn}" Page 5
changTroubleHorizonForecasting2019_Analysis	representation of a conversation in a conversation, where each w_i is a "comment" with varying length of n tokens
changTroubleHorizonForecasting2019_Analysis	"henc m = f RNN(henc m−1, wm)" Page 5
changTroubleHorizonForecasting2019_Analysis	RNN model for utterance decoder. Input is prev hidden state and current token. output is the hidden comment (m<= M tokens)
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5 , #model-architecture

Paper (7)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding. , #model-architecture
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5 , #model-architecture

#model-architecture , #to-learn

Paper (3)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 #definition ,
changTroubleHorizonForecasting2019_Analysis	"nonlinear gating function (our implementation uses GRU (Cho et al., 2014))." Page 5
changTroubleHorizonForecasting2019_Analysis	what is gating function?

#formula , #model-architecture

Paper (3)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding.
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5

#model-architecture , #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word

#data-processing, #question, #to-learn

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#definition, #model-architecture, #to-learn

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6

#data-analysis

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#conversational-forecasting , #dispute-theory , #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes?

#dispute-theory , #idea , #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition?

#challenge , #dispute-theory , #fairness , #potential-gap

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8
changTroubleHorizonForecasting2019_Analysis	application of fairness?

#contribution, #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"his model fills a void in the existing literature on conversational forecasting, simultaneously addressing the dual challenges of capturing inter-comment dynamics and dealing with an unknown horizon." Page 9

#conversational-forecasting, #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation?

#dispute-theory, #question

Paper (4)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were?

#conversational-forecasting, #dispute-theory, #potential-gap

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached?

Tags Groups within All Papers

#idea, #robust-optimization

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"First, since conversations are dynamic, a forecasting model needs to capture the flow of the discussion, rather than properties of individual comments. Second, real conversations have an unknown horizon: they can end or derail at any time; thus a practical forecasting model needs to assess the risk in an online fashion, as the conversation develops" Page
changTroubleHorizonForecasting2019_Analysis	Perhaps connect to multi-stage optimization with robust optimization? Also, how to consider "fair" assessment of risk given cross cultural differences?

#research-question, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"unsupervised representation of conversational dynamics and exploits it to predict future derailment as the conversation develops." Page
changTroubleHorizonForecasting2019_Analysis	Unsupervised model

#question, #to-read, #dispute-theory, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other?

#critique, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"nly identify antisocial content after the fact limits their practicality as tools for assisting pre-emptive moderation in conversational domains." Page
changTroubleHorizonForecasting2019_Analysis	Current state of research has not explored real time moderation as in depth as post-hoc analysis case.

#to-read

Paper (49)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"Most work uses simple dictionary-based approaches [40], though more recent work takes advantage of powerful transformer-based models [12]." Page 2 #emotion-recognition , , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	"methodological tools for studying human emotion [37]." Page
EmotionallyAwareAgentsDispute_Analysis	emotionally aware agent
EmotionallyAwareAgentsDispute_Analysis	"ability to incorporate contextual information [10, 54, 58, 60, 61]." Page #emotion-recognition ,
EmotionallyAwareAgentsDispute_Analysis	LLM's ability to reason about emotion with provided context. #emotion-recognition ,
EmotionallyAwareAgentsDispute_Analysis	"multi-issue bargaining [24]: parties are concerned with multiple issues that combine to form an overarching (private) objective function they seek to optimize." Page 2 , #negotiation
EmotionallyAwareAgentsDispute_Analysis	"As subjective perceptions of dispute are a better predictor of future negotiation decisions than the objective result [11]," Page 4 #interesting , , #negotiation
EmotionallyAwareAgentsDispute_Analysis	"better" why? how are predictions currently done and what are the shortcomings of objective measures? #interesting , , #negotiation
EmotionallyAwareAgentsDispute_Analysis	"[2, 62]. This suggests that an alternative labeling scheme might improve results." Page 4 , #emotion-labeling
EmotionallyAwareAgentsDispute_Analysis	"In-context Learning. Recent research suggests accuracy can be improved via “in-context

Paper (49)	Related Annotations
	learning" (including training examples within the prompt [19])" Page 5 #Further-exploration-needed ,
EmotionallyAwareAgentsDispute_Analysis	does this bias the model result at all? #Further-exploration-needed ,
EmotionallyAwareAgentsDispute_Analysis	"Prior research on disputes suggests that disputants often reach an impasse because anger provokes anger, leading to escalatory spirals and hardened positions [48]" Page 7 , #escalation
EmotionallyAwareAgentsDispute_Analysis	For what kinds of disputes? , #escalation
EmotionallyAwareAgentsDispute_Analysis	"To this end, other recent work demonstrates that LLMs show promise for annotating the semantic content of conflict dialogs [27, 39]." Page 8
CanLanguageModels_Analysis	"come in the form of endowing the LLM's prompt with psychology-based negotiation strategies, as similar prior work exists with rule-based agents [19, 25]." Page 8 #potential-gap , #question ,
CanLanguageModels_Analysis	is this for post dispute or during? #potential-gap , #question ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"County's AFST website [38]" Page #AFST ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"ifferential compliance has been shown to be a factor driving increased poor-rich [42] and black-white [46, 2] disparities in the postdeployment period." Page 2 #differential-compliance ,
hardtEqualityOpportunitySupervised2016_Analysis	"Barocas and Selbst [BS16]" Page 5

Paper (49)	Related Annotations
hardtEqualityOpportunitySupervised2016_Analysis	"Romei and Ruggieri [RR14]" Page 5
hardtEqualityOpportunitySupervised2016_Analysis	"Dwork et al. [DHP+12]" Page 5 , #fairness
hardtEqualityOpportunitySupervised2016_Analysis	limitations of demographic parity , #fairness
haleKODIS_NAACL_Annotatedpdf_Analysis	"Brett 090 et al., 1998; Halperin, 2008; Pruitt, 2007" Page 2 #conflict ,
haleKODIS_NAACL_Annotatedpdf_Analysis	mention the elicitation of heightened emotions and how they manifest in conflict dialogues #conflict ,
haleKODIS_NAACL_Annotatedpdf_Analysis	"(Leung 102 and Cohen, 2011; Yao et al., 2017)." Page 2 , #theoretical-framework , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	Literature used to guide multicultural measurement variables in study , #theoretical-framework , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	"(Abdurahman et al., 2024; 168 Messeri and Crockett, 2024)" Page 2 , #ethical-considerations
haleKODIS_NAACL_Annotatedpdf_Analysis	ethical risk of deployment in considering ethical implications on cultural inequalities , #ethical-considerations
haleKODIS_NAACL_Annotatedpdf_Analysis	"CaSiNo framework of Chawla et al. 189 (2021)," Page 3 #Further-exploration-needed ,
haleKODIS_NAACL_Annotatedpdf_Analysis	look at dataset +methodology #Further-exploration-needed ,
haleKODIS_NAACL_Annotatedpdf_Analysis	"(Chawla et al., 2023a)" Page 3 #Further-exploration-needed ,
haleKODIS_NAACL_Annotatedpdf_Analysis	methodology based on this paper but from deal-making application #Further-exploration-needed ,

Paper (49)	Related Annotations
haleKODIS_NAACL_Annotatedpdf_Analysis	"Giamattei et al., 220 2020)" Page 3
haleKODIS_NAACL_Annotatedpdf_Analysis	describes software used for dispute task scenarios
haleKODIS_NAACL_Annotatedpdf_Analysis	"multicultural 314 deal-making (Aslani et al., 2016)" Page 4 , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	discusses the variables to be measured. , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	"(Curhan et al., 328 2006; Brown and Curhan, 2012)" Page 4 , #theoretical-framework , #dispute-theory
haleKODIS_NAACL_Annotatedpdf_Analysis	relates to agreeability and behavior theory to guide measurement variables in stidy , #theoretical-framework , #dispute-theory
haleKODIS_NAACL_Annotatedpdf_Analysis	"but disputes involve unique social processes (Brett, 2007)" Page
haleKODIS_NAACL_Annotatedpdf_Analysis	Look at how disputes differ from deal-making from Brett paper in the social interactions
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"Addressing this limitation requires forecasting the future derailment of a conversation

Paper (49)	Related Annotations
	based on early warning signs, giving the moderators time to potentially intervene before any harm is done (Liu et al. 2018, Zhang et al. 2018a, see Jurgens et al. 2019 for a discussion)." Page
changTroubleHorizonForecasting2019_Analysis	forecasting real-time antisocial behavior-- described the "warning signs"
changTroubleHorizonForecasting2019_Analysis	"similarity between comments (Althoff et al., 2016; Tan et al., 2016) or conversation structure (Zhang et al., 2018b; Hessel and Lee, 2019)—, though neural attention architectures have also recently shown promise (Jo et al., 2018)." Page 2 , #similarity
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2 #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3 #conversational-forecasting , #fairness ,
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature? #conversational-forecasting , #fairness ,

#definition, #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2

#challenge, #potential-gap, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2

#to-read, #similarity

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"similarity between comments (Althoff et al., 2016; Tan et al., 2016) or conversation structure (Zhang et al., 2018b; Hessel and Lee, 2019)—, though neural attention architectures have also recently shown promise (Jo et al., 2018)." Page 2

#challenge, #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2

#conversational-forecasting, #previous-work

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"One solution is to assume (unrealistic) prior knowledge of when the to-be-forecasted event takes place and extract features up to that point (Niculae et al., 2015; Liu et al., 2018)." Page 2
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2 #question, #fairness,

#question, #fairness, #conversational-forecasting, #previous-work

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2

#question, #dispute-theory, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in

Paper (2)	Related Annotations
	comments 3 through 6 of Figure 1)" Page 2
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window?

#question, #confusion, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event?

#question, #to-read, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation

#question, #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation?

#challenge , #dispute-theory , #conversational-forecasting

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context?

#conversational-forecasting , #research-question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"we propose to decouple the objective of learning a neural representation of conversational dynamics from the objective of predicting future events." Page 2
changTroubleHorizonForecasting2019_Analysis	Focus on creating the neural representation of conversation dynamics rather than needing to predict events based on labels

#application

Paper (8)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"ts that expressions of com261 passion are in important predictor of negotiated 262 outcomes (Allre" Page 3 , #agent-intervention
EmotionallyAwareAgentsDispute_Analysis	applications of agent intervention , #agent-intervention
haleKODIS_NAACL_Annotatedpdf_Analysis	"social 138 science perspective, disputes offer a unique test- 139 bed to study

Paper (8)	Related Annotations
	critical social processes such as the 140 social function of emotional expressions" Page 2 #conflict , , #social-science
haleKODIS_NAACL_Annotatedpdf_Analysis	"perspective- 142 taking" Page 2 #conflict , , #social-science
haleKODIS_NAACL_Annotatedpdf_Analysis	"role of cultural and cultural misun- 145 derstandings" Page 2 , #social-science
haleKODIS_NAACL_Annotatedpdf_Analysis	"growing interest in creating artificial role-players 154 where students can practice and receive feedback 155 on their dispute-resolution skills (Shaikh et al., 156 2024; Murawski et al., 2024), in creating dialogue 157 agents that monitor interactions between people 158 and intervene to mitigate conflict (Cho et al., 2024), 159 or even replace people in customer service disputes 160 with AI (Ebers, 2022)." Page 2
changTroubleHorizonForecasting2019_Analysis	"The former can be pre-trained on large amounts of unsupervised data, similarly to how chatbots are trained. The" Page 2
changTroubleHorizonForecasting2019_Analysis	train on dispute data

#conversational-forecasting

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2 #definition ,
changTroubleHorizonForecasting2019_Analysis	"latter can piggy-back on the resulting representation after fine-tuning it for classification using relatively small labeled data." Page 2
changTroubleHorizonForecasting2019_Analysis	deploy on dispute data after using supervised learning representation

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"unsupervised representation of conversational dynamics and exploits it to predict future derailment as the conversation develops." Page #research-question ,
changTroubleHorizonForecasting2019_Analysis	Unsupervised model #research-question ,
changTroubleHorizonForecasting2019_Analysis	"nly identify antisocial content after the fact limits their practicality as tools for assisting pre-emptive moderation in conversational domains." Page #critique ,
changTroubleHorizonForecasting2019_Analysis	Current state of research has not explored real time moderation as in depth as post-hoc analysis case. #critique ,
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2 #challenge , #potential-gap ,
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2 #challenge , #potential-gap ,
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur at any time." Page 2 #challenge ,
changTroubleHorizonForecasting2019_Analysis	"One solution is to assume (unrealistic) prior knowledge of when the to-be-forecasted event takes place and extract features up to that point (Niculae et al., 2015; Liu et al., 2018)." Page 2 , #previous-work
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2 #question , #fairness , , #previous-work

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 #challenge , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? #challenge , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"we propose to decouple the objective of learning a neural representation of conversational dynamics from the objective of predicting future events." Page 2 , #research-question
changTroubleHorizonForecasting2019_Analysis	Focus on creating the neural representation of conversation dynamics rather than needing to predict events based on labels , #research-question
changTroubleHorizonForecasting2019_Analysis	"his model fills a void in the existing literature on conversational forecasting, simultaneously addressing the dual challenges of capturing inter-comment dynamics and dealing with an unknown horizon." Page 9 #contribution ,
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 #question , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? #question , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2 #question , #confusion ,
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event? #question , #confusion ,
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2 #question ,

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation? #question ,
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 , #dispute-theory , #question
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? , #dispute-theory , #question
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9 , #question
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation? , #question
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9 , #dispute-theory , #potential-gap
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached? , #dispute-theory , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page #question , #to-read , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , #to-read , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et al., 2015b; Serban et al., 2016, 2017)," Page 2 #question , #to-read ,

Paper (43)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation #question , #to-read ,
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3 , #fairness , #to-read
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature? , #fairness , #to-read
changTroubleHorizonForecasting2019_Analysis	"predicting whether an ongoing conversation will eventually spark disagreement (Hessel and Lee, 2019)" Page 3
changTroubleHorizonForecasting2019_Analysis	download subreddits, do some posyt-hoc analysis of controversial vs. non-controversial labels. the downloaded comment threads may not have currently had a controversial post.
changTroubleHorizonForecasting2019_Analysis	"statistical measures based on similarity between utterances (Althoff et al., 2016)" Page 3
changTroubleHorizonForecasting2019_Analysis	"sentiment imbalance (Niculae et al., 2015)" Page 3
changTroubleHorizonForecasting2019_Analysis	"increase in hostility (Liu et al., 2018)" Page 3
changTroubleHorizonForecasting2019_Analysis	"graph representations of conversations (Garimella et al., 2017; Zhang et al., 2018b)" Page 3 , #personal-interest
changTroubleHorizonForecasting2019_Analysis	"hierarchical recurrent encoder-decoder (HRED) architecture (Sordoni et al., 2015a)," Page 5
changTroubleHorizonForecasting2019_Analysis	model architechture for generative model in forecatsing; dependency consideration for sequences of tokens

#data-artifacts

Paper (6)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"or anger while GPT assigns more diverse labels and utilizes neutral as a dampener. GPT suggests the dialogues contain far more anger than joy, especially for buyers." Page 5 #question ,
EmotionallyAwareAgentsDispute_Analysis	What is the actual distribution of emotions for the data set? #question ,
EmotionallyAwareAgentsDispute_Analysis	"Additionally, the high amount of training data in the agreement case and the low amount of data in the impasse case caused a class imbalance, which we had to account for so our model could pick up on meaningful information" Page 7 , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	Class imbalance for objective outcomes of disputes , #potential-gap
haleKODIS_NAACL_Annotatedpdf_Analysis	"KObe DISpute corpus (KODIS)" Page 3 #todo ,
haleKODIS_NAACL_Annotatedpdf_Analysis	look at this dataset #todo ,

#result

Paper (33)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"To foreshadow our findings, we find that automatically recognized emotional expressions explain up to 45% of the variance in dispute outcomes (compared with 5% in prior negotiation research)" Page #dispute-theory ,
EmotionallyAwareAgentsDispute_Analysis	"T5-Twitter failed to perceive anger in conversations that were seen as obviously angry to participants." Page 4 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"across all four SVI subscales. This shows predictions that

Paper (33)	Related Annotations
	simply replacing T5-Twitter with GPT-4o yields" Page 5 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"GPT explains almost half the variance in feelings about buyers' feelings about the process and relationship." Page 6 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	By , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"In contrast, sellers who reach an agreement begin the dispute by expressing far more compassion for the situation, and buyers reciprocate those expressions. Thus, our automatic analysis highlights a pathway to dispute resolution that has been under-emphasized in the dispute literature. Together, these pat" Page 7 #Further-exploration-needed , , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	For other kinds of disputes scenarios, could we classify different kinds of spirals? #Further-exploration-needed , , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	"find that the subjective outcome of a dispute can be predicted from emotional expressions alone, ignoring the actual content of the dialog." Page 8 #critique , , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	"we also find evidence that “spirals of compassion” might reverse these effects (see Fig. 6(b)). Together, our findings suggest agents could intervene early to encourage participants to avoid costly escalation (see [44, 52])." Page 8 #critique , , #potential-gap

Paper (33)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	What do we mean by "reverse the effect?" #critique , , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	"figure climbs to 33% in the dialogues in the top quartile of self-reported frustration and drops to only 6% in the dialogues with the lowest." Page 4 #dispute-theory , , #potential-gap
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"marked change in workers' screening decisions in the post-deployment period." Page 2
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"deployment of the tool neither significantly mitigates nor exacerbates disparities observed at the given level of analysis." Page 2
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"2.5% of referrals had an associated score shown" Page 4 s
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"verall screen-in rate did not vary before and after deployment, remaining around 45%" Page 5 #behavior ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"bserve a change on which cases were investigated" Page 5 #behavior ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"ronounced increase in the screen-in rate of highest risk cases, and a pronounced decrease in the screen in rates for low and moderate risk cases." Page 5 #behavior ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Overall, in the period before deployment, cases that would have fallen in the category M = 1 had a screen-in rate of 58%, while after deployment this rate increased to 71%" Page 5 #behavior ,

Paper (33)	Related Annotations
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Even though there is a cost-barrier to override the machine (required supervisory approval), only 66% of cases shown as mandatory screen-in, $\tilde{M} = 1$, are screened in." Page 6 #behavior ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"all workers are effectively using information at their disposal to avoid many errors of omission, and choosing to screen in around 30% of cases that are shown to have the lowest risk scores" Page 7 #behavior,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	good vigilance #behavior ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"his indicates that humans are more likely to override a shown mandatory screen in when it is not an assessed mandatory." Page 7 #behavior,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"core was an underestimation of the assessed score were screened in at a rate of almost 60%, while the other cases with a shown score between 11 and 15 were screened in at rates around 30%." Page 7 #behavior , #data ,

Paper (33)	Related Annotations
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Before deployment of the tool, 18% of investigated referrals were accepted for services, 19% were connected to an existing open case involving the family, and 63% were not accepted for service. Post-deployment, these rates were 21%, 23%, and 56%, respectively. This indicates a higher precision in the post-deployment period: more of the screened-in referrals were being provided with services." Page 8 #impact ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	case worker was able to direct resources and have casese opened #impact ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"or the cases with a shown score between 11 and 15, more than half of those for which this score was an underestimation of the assessed score were accepted for services upon screen-in." Page 8 #impact ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	similairty of resources assisgned across buckets #impact ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"hese results indicate that, unlike what has been observed in other domains, there does not appear to be a difference in willingness to adhere to the recommendation that would compound previous racial injustices." Page 8 #impact , #racial ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"screen in rates do not appear to correlate with neighbourhood poverty levels." Page 9 #impact , #income ,
changTroubleHorizonForecasting2019_Analysis	"In both datasets, our model outperforms existing fixed-window approaches, as well as simpler sequential baselines that

Paper (33)	Related Annotations
	cannot account for inter-comment relations." Page 3
changTroubleHorizonForecasting2019_Analysis	"up performing at the level of random guessing.10 This result underscores the need for the pretraining step that can make use of unlabeled data." Page 7
changTroubleHorizonForecasting2019_Analysis	"both a visibly higher precision-recall curve and larger area under the curve (AUPR) than the baselines" Page 7
changTroubleHorizonForecasting2019_Analysis	CRAFT performance

#methodology

Paper (39)	Related Annotations
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"information communicated in the referral call, along with multi-system administrative data on demographics, child welfare involvement, criminal history, and other information related to the children and adults associated to a referral" Page 3
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"single risk score reflecting the likelihood that the children on the referral will experience adverse child welfare related outcomes in the months" Page 3
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"reflecting the immediate or long-arc risk of the childre" Page 4
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"“mandatory screen-in” to certain referrals, which means that the supervisor’s approval is required in order to screen out the referral." Page 4
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"litch in the system led to certain model inputs not being

Paper (39)	Related Annotations
	calculated correctly in realtime." Page 4 , #model-error
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"analyze the behavior of call workers with respect to the assessed score, S, before and after the deployment of the tool." Page 5
haleKODIS_NAACL_Annotatedpdf_Analysis	"For culture, we focus on differences 101 between Dignity, Face, and Honor cultures (Leung 102 and Cohen, 2011; Yao et al., 2017)." Page 2 #conflict , , #theoretical-framework
haleKODIS_NAACL_Annotatedpdf_Analysis	"We collected a diverse sample 096 of participants from over ninety countries, and par097 ticipants were matched within and across countries. 098 We measure cultural and individual differences that 099 have previously been shown to shape negotiated 100 outcomes." Page 2 , #sampling
haleKODIS_NAACL_Annotatedpdf_Analysis	diverse sampling , #sampling
haleKODIS_NAACL_Annotatedpdf_Analysis	"risk-propensity (Meertens and Lion, 2008) and the 116 specific goals parties bring to a dispute." Page 2 , #variables , #individual
haleKODIS_NAACL_Annotatedpdf_Analysis	"Finally, 117 we assess several theoretical mechanisms surround118 ing the dispute, including process variables (e.g., 119 what tactics did parties use, what emotions were 120 expressed, and did parties understand their part121 ner's interests?) and outcome variables, including 122 objective and subjective measures concerning the 123 outcome of the dispute." Page 2 , #data-analysis , #variables

Paper (39)	Related Annotations
haleKODIS_NAACL_Annotatedpdf_Analysis	"mix of human-human and human-AI 126 disputes." Page 2 #data-type ,
haleKODIS_NAACL_Annotatedpdf_Analysis	"Post130 conflict measures include beliefs about whether 131 their partner was human or AI and attitudes towards 132 using AI technology for such applications." Page 2 , #data-analysis , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants are re209 cruited through an online service such as Prolific or 210 Amazon Mechanical Turk." Page 3 , #convenience-sampling
haleKODIS_NAACL_Annotatedpdf_Analysis	"buyer who purchased a basket-256 ball jersey for their sick nephew and claims they 257 received the wrong item. The seller disputes this 258 claim, arguing the correct item was sent and no 259 refunds are allowed." Page 3
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants are told they can dis267 cuss the refund, drop their review, request their 268 opponent drop their review, and discuss who, if 269 anyone, should apologize." Page 4
haleKODIS_NAACL_Annotatedpdf_Analysis	are these the resolution tactic constraints? is there a measure for deviation from possibilities?
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants receive a base pay for 277 attempting the task, but this can be nearly doubled 278 if they achieve all their objectives in the dispute" Page 4 , #incentive
haleKODIS_NAACL_Annotatedpdf_Analysis	"cultural variability, 284 we use “elicited preferences,” meaning that partic285 ipants are provided a fixed number of points

Paper (39)	Related Annotations
	they 286 can allocate across the four issues" Page 4 , #data-analysis
haleKODIS_NAACL_Annotatedpdf_Analysis	choice modeling , #data-analysis
haleKODIS_NAACL_Annotatedpdf_Analysis	"Dignity, Face, and Honor) shape negotiation tactics 317 (e.g., tendency to express emotion, willingness to 318 exchange information)" Page 4 , #theoretical-framework
haleKODIS_NAACL_Annotatedpdf_Analysis	Theoretical model used to measure multicultural effects on deal-making , #theoretical-framework
haleKODIS_NAACL_Annotatedpdf_Analysis	"Agreement quality can be mea- 323 sured in objective terms, but subjective feelings 324 about the agreement and partner are more predic- 325 tive of subsequent behavior, such as willingness 326 to follow through on agreements and maintain a 327 future relationship with the partner (Curhan et al., 328 2006; Brown and Curhan, 2012)" Page 4 , #theoretical-framework , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	is there a relation or work to the cultural aspect on individual feelings about negotiation outcomes and what kind of emotional frameworks may be used to assess this? , #theoretical-framework , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"They are first asked a commitment request 334 asking if they could commit to providing thoughtful 335 responses." Page 5 , #self-report
haleKODIS_NAACL_Annotatedpdf_Analysis	Is this related to emotional elicitation/ manipulation in a sense , #self-report

Paper (39)	Related Annotations
haleKODIS_NAACL_Annotatedpdf_Analysis	"demo342 graphic survey including gender, education level, 343 country" Page 5 , #data-collection
haleKODIS_NAACL_Annotatedpdf_Analysis	"18-item Dignity, Face, 346 Honor measure of cultural attitudes" Page 5 , #multicultural , #data-collection
haleKODIS_NAACL_Annotatedpdf_Analysis	related to understanding identity and influence of culture , #multicultural , #data-collection
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants complete a short measure 364 of risk propensity (Meertens and Lion, 2008)" Page 5 , #data-collection , #individual
haleKODIS_NAACL_Annotatedpdf_Analysis	"participants self-report their goals for the up367 coming dispute by allocating 100 points over four 368 issues to be discussed" Page 5
haleKODIS_NAACL_Annotatedpdf_Analysis	"The concept of “integrative 392 potential” measures the potential for joint gains 393 (which may or may not be realized)." Page 5 , #data-analysis , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	this is related to the monetary bonus of good performance and the participant's self-reported goals. Utility has multiple possible solutions-- what goal they expect may change in weight , #data-analysis , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"They are asked to do the same for 411 their partner. The distance between these estimates 412 and their partner’s actual preferences can serve as 413 a measure of perspective-taking accuracy." Page 5 , #data-analysis , #self-report , #perspective-taking

Paper (39)	Related Annotations
haleKODIS_NAACL_Annotatedpdf_Analysis	This is compared to the initial assignment, but is the scale different? , #data-analysis , #self-report , #perspective-taking
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants are asked a 10-item ques- 415 tionnaire (Aslani et al., 2016) about the tactics they 416 and their partner used during the dispute." Page 5 , #self-report , #tactics
haleKODIS_NAACL_Annotatedpdf_Analysis	"dignity, face, honor, and risk." Page 6 , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"GPT4o (run on 06/28/2024) (Achiam et al., 548 2023) is prompted to annotate each dialogue 549 turn." Page 7 #emotion , , #data-analysis
haleKODIS_NAACL_Annotatedpdf_Analysis	annotate text in turn-based manner #emotion , , #data-analysis

#conversational-forecasting , #fairness , #to-read

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"any automated systems might encode or even amplify the biases existing in the training data (Park et al., 2018; Sap et al., 2019; Wiegand et al., 2019)," Page 3
changTroubleHorizonForecasting2019_Analysis	How do the cautionary warnings in these papers relate to training such a system on dispute literature?

#definition

Paper (76)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"Disputes arise when one party in a relationship (an individual, group, or nation) claims that another party refuses to accept, thus threatening the future of the relationship [25]" Page , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	"egotiation (or Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), A. El Fallah Seghrouchni, Y. Vorobeychik, S. Das, A. Nowe (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (**www.ifaamas.org**). This work is licenced under the Creative Commons Attribution 4.0 International (CC-BY 4.0) licence. deal-making) involves coming together to create a new relationship (i.e., focus on opportunities for gains), and when parties can't reach a deal, they can seek other partners" Page , #negotiation
EmotionallyAwareAgentsDispute_Analysis	dispute resolution is NOT deal-making. I think deal-making is the same as negotiation (in the context of this paper). Disputes have invariant partners with the relationship quality as the outcome, whereas negotiation is focused on the mutual objective, with relationships possibly variant. , #negotiation
CanLanguageModels_Analysis	"parties are focused on a perceived injustice by the other party and the potential costs of

Paper (76)	Related Annotations
	ending an existing relationship" Page
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Algorithm aversion–the tendency to ignore tool recommendations after seeing that they can be erroneous–originates from a lack of agency [29, 12] and lack of transparency of the algorithm [49]" Page 2 #algorithm-aversion ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	define lack of agency of the tool #algorithm-aversion ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"automation bias, on the other hand, will follow tool recommendations despite available (but unnoticed or unconsidered) information that would indicate that the recommendation is wrong" Page 2
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Commission errors refer to instances where humans take action on the basis of an erroneous algorithmic recommendation, failing to incorporate contradictory external information into the decision process" Page 2
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	algorithm is wrong, human doesn't notice
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"adherence is taken to be synonymous with trust. Indeed, while there is no single commonly adopted definition of trust in the HCI literature, the term trust typically refers to a measure of, or the factor influencing, the degree to which the human is willing to delegate decision-making to the machine in absence of complete knowledge of the algorithmic

Paper (76)	Related Annotations
	pipeline [28, 51, 3, 37, 50]" Page 3 #adherence ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"Out-of-home placement refers to whether the child is placed out of the home following an investigation, and a future referral refers to a future call involving the child coming in to the hotline." Page 4 #ASFT ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	model predictions #ASFT ,
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"commit errors of omission—screening out shown low-risk cases that are assessed as high(er) risk" Page 6
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	failed to detect by human judgement where (model error) but needed higher vigilance
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	"hown high-risk cases that are assessed as lower risk" Page 7
de-arteagaCaseHumansintheLoopDecisions2020_Analysis	agree with model error, but actually not bad-> more consequences on child
gratchFieldAffectiveComputing_Analysis	"emotions" Page 2 #supportingevidence ,
gratchFieldAffectiveComputing_Analysis	"moods" Page 2
gratchFieldAffectiveComputing_Analysis	"interpersonal stances" Page 2 , #main-idea
gratchFieldAffectiveComputing_Analysis	"affective dispositions" Page 2
hardtEqualityOpportunitySupervised2016_Analysis	"oblivious: it depends only on the joint statistics of the predictor, the target and the protected attribute, but not on interpretation of individual features." Page
hardtEqualityOpportunitySupervised2016_Analysis	"protected attributes" Page
hardtEqualityOpportunitySupervised2016_Analysis	"Demographic parity requires that a decision—such as accepting or denying a loan

Paper (76)	Related Annotations
	application—be independent of the protected attribute." Page
hardtEqualityOpportunitySupervised2016_Analysis	"“oblivious”, in that it is based only on the joint distribution, or joint statistics, of the true target Y , the predictions Ŷ , and the protected attribute A." Page 2 , #theoretical-framework , #fairness
hardtEqualityOpportunitySupervised2016_Analysis	"Y satisfies equalized odds with respect to protected attribute A and outcome Y , if Ŷ and A are independent conditional on Y ." Page 3
hardtEqualityOpportunitySupervised2016_Analysis	Equalized Odds
hardtEqualityOpportunitySupervised2016_Analysis	"property of a predictor Ŷ or score R is said to be oblivious if it only depends on the joint distribution of (Y , A, Ŷ) or (Y , A, R)" Page 4 , #fairness
hardtEqualityOpportunitySupervised2016_Analysis	"A predictor Ỹ is derived from a random variable R and the protected attribute A if it is a possibly randomized function of the random variables (R, A) alone. In particular, Ỹ is independent of X conditional on (R, A)." Page 5 , #pre
hardtEqualityOpportunitySupervised2016_Analysis	derived predictor , #pre
hardtEqualityOpportunitySupervised2016_Analysis	"predictor can be obtained as an derived threshold predictor" Page 10
hardtEqualityOpportunitySupervised2016_Analysis	a predictor is a classifier based on the thresholded R
hardtEqualityOpportunitySupervised2016_Analysis	"(Identical ROC Curves). We say that a score R has identical conditional ROC curves if Ca(t) = Ca′ (t) for all groups of a, a′ and all t ∈ R." Page 14
rudinInterpretableMachineLearning2022	"case-based reasoning" Page #Further-exploration-needed ,

Paper (76)	Related Annotations
rudinInterpretableMachineLearning2022	define and find examples #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"disentanglement of neural networks" Page #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	define disentanglement #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"generative or causal constraints" Page #actionitem ,
rudinInterpretableMachineLearning2022	find examples and define #actionitem ,
rudinInterpretableMachineLearning2022	"“Rashomon set”" Page #actionitem , #important ,
rudinInterpretableMachineLearning2022	Define #actionitem , #important ,
rudinInterpretableMachineLearning2022	"Interpretable predictive models, which are constrained so that their reasoning processes are more understandable to humans" Page 2 #Interpretable-machine-learning , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	high level defintion of interpretability the paper uses as basis for defintion. Could refine further #Interpretable-machine-learning , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"“Clever Hans” phenomenon" Page 2 #actionitem , #important , #blackbox ,
rudinInterpretableMachineLearning2022	Need to define in relation to impact of using black box models in practice #actionitem , #important , #blackbox ,
rudinInterpretableMachineLearning2022	"underlying distribution of data changes (called domain shift" Page 2 #Interpretable-machine-learning , #actionitem ,
rudinInterpretableMachineLearning2022	define domain shift, and whether thee are generalize frameworks

Paper (76)	Related Annotations
	for how this arises in 2-3 key domains were interperetability has a huge effect. #Interpretable-machine-learning , #actionitem ,
rudinInterpretableMachineLearning2022	"Clean" means that the data do not have too much noise or systematic bias" Page 4
rudinInterpretableMachineLearning2022	"Tabular" means that the features are categorical or real," Page 4
rudinInterpretableMachineLearning2022	"Raw" data is unprocessed and has a complex data type" Page 4
rudinInterpretableMachineLearning2022	"soft and hard interpretability constraints" Page 4 #Interpretable-machine-learning , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	define differences #Interpretable-machine-learning , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"This definition might require refinement, sometimes over multiple iterations with domain experts. There are many papers detailing these issues, the earliest dating from the mid-1990s [e.g., 157]." Page 5 #Interpretable-machine-learning , #important , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	What drives the definition of interpertabilitiy from a domain level and from the CS community? #Interpretable-machine-learning , #important , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"Interpretability penalties or constraints can include sparsity of the model, monotonicity with respect to a variable, decomposibility into sub-models, an ability to perform case-based reasoning or other types of visual

Paper (76)	Related Annotations
	comparisons, disentanglement of certain types of information within the model's reasoning process, generative constraints (e.g., laws of physics), preferences among the choice of variables, or any other type of constraint that is relevant to the domain" Page 5 #Interpretable-machine-learning , #important , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	Need to understand why knowledge of these penalizes the outcome from (*) #Interpretable-machine-learning , #important , #Further-exploration-needed ,
rudinInterpretableMachineLearning2022	"coring systems are linear classification models that require users to add, subtract, and multiply only a few small numbers in order to make a prediction." Page 16 , #logical-model
rudinInterpretableMachineLearning2022	"Case-based reasoning is a paradigm that involves solving a new problem using known solutions to similar past problems [1]. It is a problem-solving strategy that we humans use naturally in our decision-making processes [219]" Page 25 , #case-based-reasoning
rudinInterpretableMachineLearning2022	"Nearest neighbor-based techniques. These techniques make a decision for a previously unseen test instance, by finding training instances that most closely resemble the particular test instance" Page 26 , #case-based-reasoning
rudinInterpretableMachineLearning2022	"Given a previously unseen test instance, they make a decision by finding prototypical cases

Paper (76)	Related Annotations
	(instead of training instances from the entire training set) that most closely resemble the particular test instance" Page 27 , #case-based-reasoning
rudinInterpretableMachineLearning2022	"Disentanglement" here refers to the way information travels through the network: we would perhaps prefer that all information about a specific concept (say "lamps") traverse through one part of the network while information about another concept (e.g., "airplane") traverse through a separate part." Page 31 , #disentanglement
rudinInterpretableMachineLearning2022	"n general, a PINN is a neural network that approximates the solution of a set of PDEs with initial and boundary conditions. The training of a PINN minimizes the residuals from the PDEs as well as the residuals from the initial and boundary conditions." Page 46 , #physics-machine-learning
rudinInterpretableMachineLearning2022	constraints based on laws of nature, PDE/ODES for material science, etc. , #physics-machine-learning
rudinInterpretableMachineLearning2022	"The Rashomon effect occurs when there are multiple descriptions of the same event [41] with possibly no ground truth" Page 48 , #Rashomon-Set
rudinInterpretableMachineLearning2022	"computing statistics of the Rashomon set in parameter space, such as the volume in parameter space, which is called the Rashomon volume" Page 51 , #Rashomon-Set

Paper (76)	Related Annotations
rudinInterpretableMachineLearning2022	"Unfortunately, these topics are much too often lumped together within the misleading term “explainable artificial intelligence” or “XAI” despite a chasm separating these two concepts [250]" Page 9 #question , #Further-exploration-needed , , #To-read
haleKODIS_NAACL_Annotatedpdf_Analysis	"Conflict dialogues are task036 oriented non-collaborative conversations" Page , #conflict
haleKODIS_NAACL_Annotatedpdf_Analysis	Definition of a conflict dialogue. there is a set of goals for both parties as well as competing interests. , #conflict
haleKODIS_NAACL_Annotatedpdf_Analysis	"disputes are 071 backward-looking, typically involving an existing 072 relationship that has gone badly." Page , #conflict
haleKODIS_NAACL_Annotatedpdf_Analysis	distinction between deal-making and disputes. Deal making is opportunistic benefits of the relationship while disputes are focused on the cost of ending the relationship. More emotionally driven, increases the effect of influence methods. , #conflict
haleKODIS_NAACL_Annotatedpdf_Analysis	"Dignity relates to the 350 Western ideal that each individual has intrinsic self351 worth and thus tends to be impervious to threats 352 from others." Page 5 , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	"Honor cultures tend to view self-worth 353 as something that must be claimed and defended 354 from external threats." Page 5 , #multicultural
haleKODIS_NAACL_Annotatedpdf_Analysis	"Face cultures also see self355 worth as conferred by others but see retaliation as 356 further

Paper (76)	Related Annotations
	eroding self-worth." Page 5 , #multicultural
changTroubleHorizonForecasting2019_Analysis	"conversational forecasting, which includes future-prediction tasks such as predicting the eventual length of a conversation" Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior online comes in many forms, including harassment (Vitak et al., 2017), cyberbullying (Singh et al., 2017), and general aggression (Kayany, 1998)." Page 3
changTroubleHorizonForecasting2019_Analysis	antisocial behavior
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 , #model-architecture , #to-learn

#negotiation

Paper (10)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"egotiation (or Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), A. El Fallah Seghrouchni, Y. Vorobeychik, S. Das, A. Nowe (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (**www.ifaamas.org**). This work is licenced under the Creative Commons Attribution 4.0 International (CC-BY 4.0) licence. deal-making) involves coming together to create a new relationship (i.e., focus on opportunities for gains),

Paper (10)	Related Annotations
	and when parties can't reach a deal, they can seek other partners" Page #definition ,
EmotionallyAwareAgentsDispute_Analysis	dispute resolution is NOT deal-making. I think deal-making is the same as negotiation (in the context of this paper). Disputes have invariant partners with the relationship quality as the outcome, whereas negotiation is focused on the mutual objective, with relationships possibly variant. #definition ,
EmotionallyAwareAgentsDispute_Analysis	"In negotiations, emotion recognition has shown only weak promise to predict outcomes" Page #emotion-recognition ,
EmotionallyAwareAgentsDispute_Analysis	"GPT-4 yielded the best performance compared with a wide range of other LLMs on making inferences from negotiation dialogues [39]." Page 4 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	Look into what are performance metrics typically used. , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"multi-issue bargaining [24]: parties are concerned with multiple issues that combine to form an overarching (private) objective function they seek to optimize." Page 2 #to-read ,
EmotionallyAwareAgentsDispute_Analysis	"As subjective perceptions of dispute are a better predictor of future negotiation decisions than the objective result [11]," Page 4 #interesting , #to-read ,
EmotionallyAwareAgentsDispute_Analysis	"better" why? how are predictions currently done and what are the shortcomings of objective measures? #interesting , #to-read ,
changTroubleHorizonForecasting2019_Analysis	"identifying successful negotiations (Curhan and Pentland, 2007; Cadilhac et al., 2013)," Page 3
changTroubleHorizonForecasting2019_Analysis	classification of negotiation success post-hoc

#dispute-theory

Paper (51)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"Disputes arise when one party in a relationship (an individual, group, or nation) claims that another party refuses to accept, thus threatening the future of the relationship [25]" Page #definition ,
EmotionallyAwareAgentsDispute_Analysis	"disputes evoke much stronger emotions than negotiations, particularly anger." Page
EmotionallyAwareAgentsDispute_Analysis	"anger provokes retaliation [48]" Page
EmotionallyAwareAgentsDispute_Analysis	"dispute literature shows how specific emotions convey specific intentions and examines how these conveyed intentions change within a text. This is challenging as interpreting each utterance depends on the context of prior utterances [47" Page #emotion-recognition , #contribution ,
EmotionallyAwareAgentsDispute_Analysis	Lead with emotional mapping to dispute intentions based on theory to analyze context in unseen disputes to forecast intention meaning. #emotion-recognition , #contribution ,
EmotionallyAwareAgentsDispute_Analysis	"To foreshadow our findings, we find that automatically recognized emotional expressions explain up to 45% of the variance in dispute outcomes (compared with 5% in prior negotiation research)" Page , #result
EmotionallyAwareAgentsDispute_Analysis	"Whereas social science findings have emphasized the role of anger in disputes, our findings highlight the importance of other emotional expressions, such as compassion." Page 2 #emotion-recognition , #contribution ,

Paper (51)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"LMs perform substantially better in predicting dispute outcomes than previous text emotion recognition methods (highlighting the impact of several prompting strategies)" Page 2 #emotion-recognition , #contribution ,
EmotionallyAwareAgentsDispute_Analysis	"results replicate prior social science findings on the role of anger in escalation and lay a foundation for autonomous agents that could detect escalation and intervene before a dispute ends in an impasse." Page 2 #emotion-recognition , #contribution ,
EmotionallyAwareAgentsDispute_Analysis	"parties can often find better solutions if they exchange information about each other's interests and find solutions that maximize joint gains [5, 56]." Page 2
EmotionallyAwareAgentsDispute_Analysis	"Neither line of work considers direct interactions between people nor do they consider the role of emotional expressions." Page 2 #critique ,
EmotionallyAwareAgentsDispute_Analysis	in relation to mediation and argument-based agent-agent negotiation #critique ,
EmotionallyAwareAgentsDispute_Analysis	"Most work uses simple dictionary-based approaches [40], though more recent work takes advantage of powerful transformer-based models [12]." Page 2 #emotion-recognition , #to-read ,
EmotionallyAwareAgentsDispute_Analysis	"hough we focused on dispute resolution, our findings could inform techniques that intentionally escalate conflicts." Page 8 , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	Can we apply our results to other dispute types? , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	"As parties are already linked," Page #question ,
EmotionallyAwareAgentsDispute_Analysis	What defines a link--furthermore, the type of the relationship? Historic agreement? Self-interests? #question ,

Paper (51)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"Prior research on emotions in text focuses on a document as the unit of analysis and outputs coarse-grained “sentiment” rather than specific emotions like anger [59]." Page #question ,
EmotionallyAwareAgentsDispute_Analysis	Difference between intention and sentiment? #question ,
EmotionallyAwareAgentsDispute_Analysis	"figure climbs to 33% in the dialogues in the top quartile of self-reported frustration and drops to only 6% in the dialogues with the lowest." Page 4 , #result , #potential-gap
EmotionallyAwareAgentsDispute_Analysis	"This reinforces findings in the social sciences on how anger shapes conflict [30, 48]," Page 8 #question , #Further-exploration-needed ,
EmotionallyAwareAgentsDispute_Analysis	What is "new" result versus not something currently backed by the dispute literature? #question , #Further-exploration-needed ,
CanLanguageModels_Analysis	"Real-world corpora involve rich and authentic conversations, but one rarely has access to each party’s true underlying goals and emotions, making it difficult to quantify success or failure objectively" Page
CanLanguageModels_Analysis	Benefits/drawbacks of observed disputes
CanLanguageModels_Analysis	"anger in disputes provokes retaliation [21]" Page , #evidence
haleKODIS_NAACL_Annotatedpdf_Analysis	"Subjective 423 perceptions of the outcome of a dispute are a bet- 424 ter predictor of future negotiation decisions than 425 the actual economic result (Brown and Curhan, 426 2012)." Page 5 , #subjective
haleKODIS_NAACL_Annotatedpdf_Analysis	"do ex534 pressions of anger provoke escalation and impasses 535 in disputes (as previously claimed), can negotiation 536 satisfaction be predicted by emotional expression 537 alone, and

Paper (51)	Related Annotations
	how does culture shape these findings?" Page 7 , #research-question
haleKODIS_NAACL_Annotatedpdf_Analysis	"(Curhan et al., 328 2006; Brown and Curhan, 2012)" Page 4 #to-read , #theoretical-framework ,
haleKODIS_NAACL_Annotatedpdf_Analysis	relates to agreeability and behavior theory to guide measurement variables in stidy #to-read , #theoretical-framework ,
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 #challenge , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? #challenge , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8 #challenge , , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	application of fairness? #challenge , , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? #question , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"attack-containing conversation is paired with a clean conversation from the same talk page, where the talk page serves as a proxy for topic.3" Page 4 , #Further-exploration-needed , #unsupervised-learning
changTroubleHorizonForecasting2019_Analysis	look into techniques for correlation control-- what kinds of structures to consider for the text? , #Further-exploration-needed , #unsupervised-learning

Paper (51)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting , , #question
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting , , #question
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8 , #idea , #question
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition? , #idea , #question
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9 , #question
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were? , #question
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9 #conversational-forecasting , , #potential-gap
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached? #conversational-forecasting , , #potential-gap
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan,

Paper (51)	Related Annotations
	2014)" Page #question , #to-read , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? #question , #to-read , , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"deception (Girlea et al., 2016; P´erez-Rosas et al., 2016; Levitan et al., 2018)" Page 3
changTroubleHorizonForecasting2019_Analysis	deception classification post-hoc
changTroubleHorizonForecasting2019_Analysis	"disagreement (Galley et al., 2004; Abbott et al., 2011; Allen et al., 2014; Wang and Cardie, 2014; Rosenthal and McKeown, 2015)" Page 3
changTroubleHorizonForecasting2019_Analysis	disagreement classification post-hoc

#conversational-forecasting , #personal-interest

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"graph representations of conversations (Garimella et al., 2017; Zhang et al., 2018b)" Page 3

#goal

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"utomatically learn neural representations of conversational dynamics through pre-training." Page 3
changTroubleHorizonForecasting2019_Analysis	"1) How much early warning does the the model provide? (2) Does the model actually" Page 7

#challenge

Paper (11)	Related Annotations
rudinInterpretableMachineLearning2022	"Improve the scalability of optimal sparse scoring systems: As discussed, for scoring systems, the only practical approaches that produce optimal scoring systems require a MIP solver, and these approaches may not be able to scale to large problems, or optimally handle continuous variables." Page 20 #logical-model ,
rudinInterpretableMachineLearning2022	"But in the case where we do not know the concepts, or in the case where the concepts are numerous and we do not know how to parameterize them, we cannot use the techniques from Challenge 5. In other words, the concept c in Constraint (5.1) is no longer a concept we predefine, but it must still be an actual concept in the existing universe of concepts" Page 35 #unsupervised-learning , , #disentanglement
rudinInterpretableMachineLearning2022	we do not know the classes ahead of time unlike other case. Makes quantitavely evaluating harder if we know nothing about the classes. we may also ignore key iunformaiton if we have human annotators. #unsupervised-learning , , #disentanglement
changTroubleHorizonForecasting2019_Analysis	"dynamic and their outcome might depend on how subsequent comments interact with each other" Page 2 , #potential-gap , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"Thus a forecasting model needs to capture not only the content of each individual comment, but also the relations between comments. Previous work has largely relied on hand-crafted features to capture such relations" Page 2 , #potential-gap , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"conversations have an unknown horizon: they can be of varying lengths, and the to-beforecasted event can occur

Paper (11)	Related Annotations
	at any time." Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"The main difficulty in directly adapting these models to the supervised domain of conversational forecasting is the relative scarcity of labeled data:" Page 2 , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What labels do we need for dispute context? , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"do not address the issue of aggregating them into a single forecast (i.e., deciding at what point to make 4746 a prediction)" Page 3
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8 , #dispute-theory , #fairness , #potential-gap
changTroubleHorizonForecasting2019_Analysis	application of fairness? , #dispute-theory , #fairness , #potential-gap

#question

Paper (103)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"As parties are already linked," Page , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	What defines a link--furthermore, the type of the relationship? Historic agreement? Self-interests? , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	"Prior research on emotions in text focuses on a document as the unit of analysis and outputs coarse-grained “sentiment” rather than specific emotions like anger [59]." Page , #dispute-theory

Paper (103)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	Difference between intention and sentiment? , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	"e model classified each utterance of the dispute in isolation, and we created an overall score for the dialog by summing across these labels afterward." Page 4
EmotionallyAwareAgentsDispute_Analysis	what does "in isolation" mean here?
EmotionallyAwareAgentsDispute_Analysis	"“No I do not because you clearly did not read the description” might be seen as neutral in isolation but more negative in the context of the preceding line “So you do not see a need to apologize to me for sending me the wrong jersey.”" Page 4 , #data-analysis
EmotionallyAwareAgentsDispute_Analysis	how many lines in advance is enough context? , #data-analysis
EmotionallyAwareAgentsDispute_Analysis	"conflict literature suggests that expressions of compassion are an important predictor of negotiated outcomes" Page 4 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	Why does the scale not initially differentiate positive emotions as much? If a scale is adjusted for a specific use case, can we compare it well with prior research? , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"o address concerns that the emotions of each dialogue turn can depend on prior turns [47]" Page 5 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	Can we further refine how far back the context needs to be to get better accuracy? Further if we assume in isolation, how are we sure the model "forgets" previous input to make a fresh decision? , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"address concerns with T5-Twitter’s labels, we substituted T5-Twitter’s

Paper (103)	**Related Annotations**
	love with compassion and added a neutral label to not force GPT to pick an emotion where none is apparent." Page 5 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	is this too mach variation to compare performance with T5? I feel like this would have a lot of confounding variables in a better outcome. But the goal is to increase the accuracy to align better with the subjective value , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"we assess how each emotion label correlates with self-reported frustration for each annotation method." Page 5 , #data-analysis
EmotionallyAwareAgentsDispute_Analysis	"or anger while GPT assigns more diverse labels and utilizes neutral as a dampener. GPT suggests the dialogues contain far more anger than joy, especially for buyers." Page 5 , #data-artifacts
EmotionallyAwareAgentsDispute_Analysis	What is the actual distribution of emotions for the data set? , #data-artifacts
EmotionallyAwareAgentsDispute_Analysis	"Backwards Regression. A subsequent backward regression clarifies which expressions significantly contribute to predicting SVI (* p < .001, p < .01, and * p < .05)" Page 6 , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	for backwards comparison, how do we choose the p-level for this use case? Difference between the magnitude/ sign and the p-level? What is the scale based on? , #model-comparison
EmotionallyAwareAgentsDispute_Analysis	"This reinforces findings in the social sciences on how anger shapes conflict [30, 48]," Page 8 , #Further-exploration-needed , #dispute-theory
EmotionallyAwareAgentsDispute_Analysis	What is "new" result versus not something currently backed by the

Paper (103)	Related Annotations
	dispute literature? , #Further-exploration-needed , #dispute-theory
CanLanguageModels_Analysis	"However, many of the disputes escalated and ended without agreement, thus forgoing their bonus. Even when agreements were reached, disputants often reported high frustration with their partner." Page 2
CanLanguageModels_Analysis	Why did many disputes lead to conflict?
CanLanguageModels_Analysis	"(1) whether or not the dispute ended in success or failure, and (2) whether or not the participants reported frustration with each other." Page 2
CanLanguageModels_Analysis	How does frustration get mapped to anger? In the other paper, we have a 10-item tactics survey, but no label on GPT model for frustration
CanLanguageModels_Analysis	"We find that the LLMs were rated significantly better in predicting when to intervene, rated as providing a better rationale for intervening, and rated as providing a more effective mediation message to the disputants" Page 2 , #critique , #potential-gap
CanLanguageModels_Analysis	"importance to these issues via a payoff matrix, participants are free to assign their own importance to each issue." Page 3
CanLanguageModels_Analysis	why move away from payoff matrix as in initial KODIS paper on the cross-cultural differences
CanLanguageModels_Analysis	"ing LLMs and explore several prompt" Page 3 , #Further-exploration-needed
CanLanguageModels_Analysis	for what kinds of disputes? , #Further-exploration-needed
CanLanguageModels_Analysis	"Given the human-mediated dialogs alongside the GPT-mediated ones1,

Paper (103)	Related Annotations
	we move to subjective evaluations of each." Page 6 , #critique , #data-collection
CanLanguageModels_Analysis	But how do we define how a human interperets "exchange level"? , #critique , #data-collection
CanLanguageModels_Analysis	"he humanmediator followed instructions (i.e., they role-played as a mediator);" Page 6 , #data-analysis
CanLanguageModels_Analysis	in what cases did they not? , #data-analysis
CanLanguageModels_Analysis	"In this new task, we recruit N = 106 participants so that each mediation receives at least five annotations." Page 7 , #sampling
CanLanguageModels_Analysis	Any issues with sampling again and outcome qualioty/statistical effects? Is there more subjectivity now in the 20-to evaluate the justification? , #sampling
CanLanguageModels_Analysis	"Concretely, we ask to what extent the participant disagrees or agrees (1-10) with three statements depicted in Table 2." Page 7 , #experiment
CanLanguageModels_Analysis	Why a 10 point scale for agreeability? Also, what were the comstrains for reason to intervene on giving the reaso? DO we need to include any ethical considerdations in the kinds of repsonses? , #experiment
CanLanguageModels_Analysis	"come in the form of endowing the LLM's prompt with psychology-based negotiation strategies, as similar prior work exists with rule-based agents [19, 25]." Page 8 #potential-gap , , #to-read
CanLanguageModels_Analysis	is this for post dispute or during? #potential-gap , , #to-read

Paper (103)	Related Annotations
gratchFieldAffectiveComputing_Analysis	"d) action tendencies (such as preparation for fight versus flight)" Page 2
gratchFieldAffectiveComputing_Analysis	How are these different from psychological changes, and do these vary across individuals or are there generalizations?
hardtEqualityOpportunitySupervised2016_Analysis	"‖p − q‖2 6 √ 2 · dK (R, R′)." Page 11 #Bayesian-methods , #formula ,
hardtEqualityOpportunitySupervised2016_Analysis	"The feasible set of false/true positive rates of possible equalized odds predictors is thus the intersection of the areas under the A-conditional ROC curves, and above the main diagonal (see Figure 2). Since for any loss function the optimal false/true-positive rate will always be on the upper-left boundary of this feasible set, this is effectively the ROC curve of the equalized odds predictors." Page 9 #explanation , , #ROC
hardtEqualityOpportunitySupervised2016_Analysis	"y = 1, the constraint requires that Ὑ has equal true positive rates across the two demographics A = 0 and A = 1. For y = 0, the constraint equalizes false positive rates." Page 3 , #Further-exploration-needed
hardtEqualityOpportunitySupervised2016_Analysis	"R ∈ [0, 1]." Page 8 #todo ,
hardtEqualityOpportunitySupervised2016_Analysis	see if this is same R as mentioned to be random variable? #todo ,
hardtEqualityOpportunitySupervised2016_Analysis	"In this case, any thresholding of R yields an equalized odds predictor (all protected groups are at the same point on the curve, and the same point in false/true-positive plane)." Page 8 , #analysis
hardtEqualityOpportunitySupervised2016_Analysis	ROC curve-- since we condition on the distribution of A, picking t selects a point on the ROC curve. we generate the same graph for all possible (A,Y) distributions , #analysis

Paper (103)	Related Annotations
rudinInterpretableMachineLearning2022	"uch issues with explanations have arisen with assessment of fairness and variable importance [258, 82] as well as uncertainty bands for variable importance [113, 97]" Page 10 , #Interpretable-machine-learning , #critique , #Explainability
rudinInterpretableMachineLearning2022	What are uncertainty bands? , #Interpretable-machine-learning , #critique , #Explainability
rudinInterpretableMachineLearning2022	"inflated by including many “obvious” cases" Page 7 • Why are "obvious' cases misleading? ROC- Distinguishability between classes
rudinInterpretableMachineLearning2022	"n interpretable robotic surgeon would be worse than its black box counterpart. The question ultimately becomes whether the Rashomon set should permit such an interpretable robotic surgeon–and all scientific evidence so far (including a large-and-growing number of experimental papers on interpretable deep learning) suggests it would." Page 9 , #blackbox , #Further-exploration-needed
rudinInterpretableMachineLearning2022	What do we mean by "worse"? Are we defining this in terms of accuracy? , #blackbox , #Further-exploration-needed
rudinInterpretableMachineLearning2022	"Unfortunately, these topics are much too often lumped together within the misleading term “explainable artificial intelligence” or “XAI” despite a chasm separating these two concepts [250]" Page 9 , #Further-exploration-needed , #definition , #To-read
rudinInterpretableMachineLearning2022	"But function approximators are not used in interpretable ML; instead of

Paper (103)	Related Annotations
	approximating a known function (a black box ML model), interpretable ML can choose from a potential myriad of approximatelyequally-good models, which, as we noted earlier, is called “the Rashomon set”" Page 11 , #Interpretable-machine-learning , #Explainability
rudinInterpretableMachineLearning2022	difference between model and function in this context , #Interpretable-machine-learning , #Explainability
rudinInterpretableMachineLearning2022	"linear models, which includes scoring systems (and risk scores)" Page 21 , #logical-models
rudinInterpretableMachineLearning2022	linear models are logical models? , #logical-models
rudinInterpretableMachineLearning2022	"boosted models are not naturally sparse, and issues with bias arise under 1 regularization, as discussed in the scoring systems section." Page 23 , #Further-exploration-needed , #Generalized-Additive-Model
rudinInterpretableMachineLearning2022	What are proxies? Connects to how regularizaiton issues (stripping away too many small weights) if tree is not sparse can be an issue. , #Further-exploration-needed , #Generalized-Additive-Model
rudinInterpretableMachineLearning2022	"using a deep neural network that transforms the input space into a feature space where a kNN classifier will perform well (i.e., deep kNN). Papernot and McDaniel [229]" Page 26
rudinInterpretableMachineLearning2022	what is the relation to latent spaces here?
haleKODIS_NAACL_Annotatedpdf_Analysis	"This theoretical 103 framework distinguishes cultures by the degree to 104 which people’s social identity is independent ver105 sus interdependent and thus shapes the importance 106 given to norms of

Paper (103)	Related Annotations
	reciprocity and honesty." Page 2 , #important , #theoretical-framework
haleKODIS_NAACL_Annotatedpdf_Analysis	what are norms? Why specifically recipricocity and honesty in the context of cultural analysis here. It seems the cultures are categorized according to values and how they relate to these norms? , #important , #theoretical-framework
haleKODIS_NAACL_Annotatedpdf_Analysis	"information 195 expressed in the dialog revealed participant's pri196 vate goals for the negotiation" Page 3 , #Further-exploration-needed
haleKODIS_NAACL_Annotatedpdf_Analysis	"Participants are next directed 217 to the dispute task implemented in Lioness Labs, 218 a software framework used for multi-participant 219 behavioral economic experiments" Page 3
haleKODIS_NAACL_Annotatedpdf_Analysis	in this general section consider the effects of randomization and some of those concepts from casual inference (?) from ISE 625. Do we care about individual characteristics in the survey that may be confounding variables in how "important" a dispute may be?
haleKODIS_NAACL_Annotatedpdf_Analysis	"otherwise, at one minute left, the 226 participant moves on and converses with an AI 227 counterpart." Page 3
haleKODIS_NAACL_Annotatedpdf_Analysis	is one minute enough time to resolve with an AI chatbot? 8 minutes total possible, but only 1 needed for AI?
haleKODIS_NAACL_Annotatedpdf_Analysis	"defines their goals in the negotiation" Page 4
haleKODIS_NAACL_Annotatedpdf_Analysis	do we set what the goals are for them in the negotiation?
haleKODIS_NAACL_Annotatedpdf_Analysis	"These points are 444 also used to determine the bonus." Page 6 , #confusion
haleKODIS_NAACL_Annotatedpdf_Analysis	I am unsure on what outcomes lead to "doubling". because if they have fixed

Paper (103)	Related Annotations
	weight and this formula holds the points + binary weights (I), then there is no doubling indicated with this linear additive utility. , #confusion
haleKODIS_NAACL_Annotatedpdf_Analysis	"We only exam- 565 ine human-human dyads; the final agreement (or 566 non-agreement) is excluded from the dialogues." Page 7
haleKODIS_NAACL_Annotatedpdf_Analysis	why is the outcome excluded?
haleKODIS_NAACL_Annotatedpdf_Analysis	"This is remarkable as participants forfeit 578 a cash bonus if they fail to achieve an agreement, 579 even though this was merely a simulated dispute." Page 8
haleKODIS_NAACL_Annotatedpdf_Analysis	Why was this the case? Was the reasoning for the high turnover of impasse accurate, or is there some bias/ influence in the study that may have lead to low resolution outcome
haleKODIS_NAACL_Annotatedpdf_Analysis	"Fig609 ure 3 illustrates the differences in occurrences of 610 emotions for the five countries we consider when 611 disputing against the same or a different country." Page 8
haleKODIS_NAACL_Annotatedpdf_Analysis	Is this based on the GPT turn-based rating?
changTroubleHorizonForecasting2019_Analysis	"Another compromising solution is to extract features from a fixed-length window, often at the start of the conversation" Page 2 , #fairness , #conversational-forecasting , #previous-work
changTroubleHorizonForecasting2019_Analysis	"window of only two comments would miss the chain of repeated questioning in comments 3 through 6 of Figure 1)" Page 2 , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	What kinds of patterns can we consider for "spiraling"? How do we define structure of a window? ,

Paper (103)	Related Annotations
	#dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"longer windows risk missing the to-be-forecasted event altogether" Page 2 , #confusion , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	why do longer windows miss an event? , #confusion , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"while concomitantly being able to process the conversation as it develops (see Gao et al. (2018) for a survey)." Page 2 , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	could something here be used as guide for the mediation response generation? , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"This is a useful property for the purposes of model analysis, and hence we focus on this as our primary dataset." Page 4
changTroubleHorizonForecasting2019_Analysis	considerations of hand annotated labels?
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5 #model-architecture ,
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word #model-architecture ,
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting , #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent

Paper (103)	Related Annotations
	events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8 #dispute-theory , #idea ,
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition? #dispute-theory , #idea ,
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9 #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation? #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9 #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were? #dispute-theory ,
changTroubleHorizonForecasting2019_Analysis	"Antisocial behavior is a persistent problem plaguing online conversation platforms; it is both widespread (Duggan, 2014)" Page , #to-read , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	compare antisocial behavior to dispute characteristics-- are they distinct concepts, or is one a subclass of the other? , #to-read , #dispute-theory , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	"sequential neural models that make effective use of the intra-conversational dynamics (Sordoni et

Paper (103)	Related Annotations
	al., 2015b; Serban et al., 2016, 2017)," Page 2 , #to-read , #conversational-forecasting
changTroubleHorizonForecasting2019_Analysis	reference for in-context generation , #to-read , #conversational-forecasting

#dispute-theory , #Further-exploration-needed , #unsupervised-learning

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"attack-containing conversation is paired with a clean conversation from the same talk page, where the talk page serves as a proxy for topic.3" Page 4
changTroubleHorizonForecasting2019_Analysis	look into techniques for correlation control- - what kinds of structures to consider for the text?

#data-processing

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#model-architecture

Paper (9)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"utterance encoder is responsible for generating semantic vector representations of comments" Page 5

Paper (9)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 #definition , , #to-learn
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5 #formula ,
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding. #formula ,
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5 #formula ,
changTroubleHorizonForecasting2019_Analysis	"nonlinear gating function (our implementation uses GRU (Cho et al., 2014))." Page 5 , #to-learn
changTroubleHorizonForecasting2019_Analysis	what is gating function? , #to-learn
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5 , #question
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word , #question

#formula

Paper (40)	Related Annotations
hardtEqualityOpportunitySupervised2016_Analysis	"Pr{Y = 1 | A = 0} = Pr{Y = 1 arXiv:1610.02413v1 [cs.LG] 7 Oct 2016 1 | A = 1}" Page
hardtEqualityOpportunitySupervised2016_Analysis	demographic parity formula
hardtEqualityOpportunitySupervised2016_Analysis	"Y satisfies equal opportunity with respect to A and Y if Pr { Ύ = 1 | A = 0, Y = 1} = Pr { Ύ = 1 | A = 1, Y = 1} ." Page 4
hardtEqualityOpportunitySupervised2016_Analysis	"γa(Ύ) d=ef (Pr { Ύ = 1 | A = a, Y = 0} , Pr { Ύ = 1 | A = a, Y = 1})" Page 6 , #fairness

Paper (40)	Related Annotations
hardtEqualityOpportunitySupervised2016_Analysis	equal opportunity in terms of FPR and TPR from distribution , #fairness
hardtEqualityOpportunitySupervised2016_Analysis	"Lemma 4.3. A predictor Ỹ is derived if and only if for all a ∈ {0, 1}, we have γa(Ỹ) ∈ Pa(Y)." Page 6
hardtEqualityOpportunitySupervised2016_Analysis	derived predictor condition. The Ytilde must come from the convex hull of probabilities for TPR/FPR
hardtEqualityOpportunitySupervised2016_Analysis	"min Ỹ E`(Ỹ , Y)" Page 7
hardtEqualityOpportunitySupervised2016_Analysis	optimization problem for expected loss from equalized odds and ROC graphs
hardtEqualityOpportunitySupervised2016_Analysis	"Ca(t) d=ef (Pr { R > t | A = a, Y = 0} , Pr { R > t | A = a, Y = 1}) ." Page 8
hardtEqualityOpportunitySupervised2016_Analysis	Application of non-discrimination for score-based continuous models using ROC score.
hardtEqualityOpportunitySupervised2016_Analysis	"Figure 2: Finding the optimal equalized odds threshold predictor (middle), and equal opportunity threshold predictor (right). For the equal opportunity predictor, within each group the cost for a given true positive rate is proportional to the horizontal gap between the ROC curve and the profit-maximizing tangent line" Page 9
hardtEqualityOpportunitySupervised2016_Analysis	"Ta is a randomized threshold assuming the value ta with probability pa and the value ta with probability pa." Page 9
hardtEqualityOpportunitySupervised2016_Analysis	Randomized interpretation for mixed ROC. We have another convex hull when the two curves do not intersect.
hardtEqualityOpportunitySupervised2016_Analysis	"∀a : γ∈Da γ0 `(1, 0) + (1 − γ1)` (0, 1)" Page 10 , #ROC
hardtEqualityOpportunitySupervised2016_Analysis	Equalized Odds Loss Function randomization curve , #ROC

Paper (40)	Related Annotations
hardtEqualityOpportunitySupervised2016_Analysis	"R = arg minr(x,a) E [(Y − r(X, A))2] = r∗(X, A) with r∗(x, a) = E[Y | X = x, A = a]." Page 10 #Bayesian-methods ,
hardtEqualityOpportunitySupervised2016_Analysis	"K(R, R′) d=ef max a,y∈{0,1} sup t∈[0,1] |||Pr {R > t | A = a, Y = y} − Pr {R′ > t | A = a, Y = y}||| ." Page 11 #Bayesian-methods ,
hardtEqualityOpportunitySupervised2016_Analysis	Maximum least upper bound (R) conditioned on all possible A, Y values. #Bayesian-methods ,
hardtEqualityOpportunitySupervised2016_Analysis	"‖p − q‖2 6 √ 2 · dK (R, R′)." Page 11 #Bayesian-methods , , #question
hardtEqualityOpportunitySupervised2016_Analysis	"E (Ŷ , Y) 6 E (Y ∗, Y) + 2 √ 2 · dK(R, R∗) ," Page 11 #Bayesian-methods ,
hardtEqualityOpportunitySupervised2016_Analysis	for an equalized odds predictor, the Kolmogrov distance for restricted ROC curves #Bayesian-methods ,
hardtEqualityOpportunitySupervised2016_Analysis	"Pr {Y = 1 | R = t, A = a} = Pr {Y = 1 | R = t, A = a′}" Page 15 #discrimination-measure ,
hardtEqualityOpportunitySupervised2016_Analysis	matching frequencies #discrimination-measure ,
rudinInterpretableMachineLearning2022	"min f ∈F 1 n ∑ i Loss(f, zi) + C · InterpretabilityPenalty(f), subject to (*) InterpretabilityConstraint(f)" Page 4 #Interpretable-machine-learning ,
rudinInterpretableMachineLearning2022	"sparsity is measured by the number of leaves in the tree" Page 14 , #sparsity
rudinInterpretableMachineLearning2022	"g(E[y]) = β0 + f1(x·1) + ... + fp(x·p)," Page 21
rudinInterpretableMachineLearning2022	generative additive model
rudinInterpretableMachineLearning2022	"R(F , f ∗,) = {f ∈ F such that Loss(f) ≤ Loss(f ∗) + }," Page 49 , #Rashomon-Set
haleKODIS_NAACL_Annotatedpdf_Analysis	"IP = 1 − □Xbuyer · □Xseller □ →Xbuyer□□ □Xseller□" Page 5 , #preference-elicitation

Paper (40)	Related Annotations
haleKODIS_NAACL_Annotatedpdf_Analysis	Integrative potential formula related to payoff of 4 choices , #preference-elicitation
haleKODIS_NAACL_Annotatedpdf_Analysis	"Ua = X i∈I 447 wi,a ∗ li,a" Page 6 , #objective
haleKODIS_NAACL_Annotatedpdf_Analysis	measures the outcome value of the dispute based on the preferences the users assigned to the 4 issues prior to the dispute. , #objective
changTroubleHorizonForecasting2019_Analysis	"cn = {w1, . . . , wMn}" Page 5
changTroubleHorizonForecasting2019_Analysis	representation of a conversation in a conversation, where each w_i is a "comment" with varying length of n tokens
changTroubleHorizonForecasting2019_Analysis	"henc m = f RNN(henc m−1, wm)" Page 5
changTroubleHorizonForecasting2019_Analysis	RNN model for utterance decoder. Input is prev hidden state and current token. output is the hidden comment (m<= M tokens)
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5 , #model-architecture
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding. , #model-architecture
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5 , #model-architecture

#model-architecture , #to-learn

Paper (3)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6 #definition ,

Paper (3)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"nonlinear gating function (our implementation uses GRU (Cho et al., 2014))." Page 5
changTroubleHorizonForecasting2019_Analysis	what is gating function?

#formula, #model-architecture

Paper (3)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"hcon n = f RNN(hcon n−1, henc Mn)" Page 5
changTroubleHorizonForecasting2019_Analysis	encoder for ALL comments. pass in previous comment encoding.
changTroubleHorizonForecasting2019_Analysis	"htdec = f RNN(htde−c1, wt−1) wt = f out(htdec) (3)" Page 5

#model-architecture, #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"hidden state hcnon can then be viewed as an encoding of the full conversational context up to and including the n-th comment." Page 5
changTroubleHorizonForecasting2019_Analysis	Can we vary token length to be more than 1 word

#data-processing, #question, #to-learn

Paper (0)	Related Annotations

Dataview: No results to show for table query.

#definition, #model-architecture, #to-learn

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"This process, known as fine-tuning, reshapes the representation learned during pre-training to be more directly useful to prediction (Howard and Ruder, 2018)." Page 6

#data-analysis

Paper (16)	Related Annotations
EmotionallyAwareAgentsDispute_Analysis	"They used a dictionary-based approach and several pre-trained models, finding the best results for a T5 model fine-tuned on Twitter corpus [49]" Page 4
EmotionallyAwareAgentsDispute_Analysis	"“No I do not because you clearly did not read the description” might be seen as neutral in isolation but more negative in the context of the preceding line “So you do not see a need to apologize to me for sending me the wrong jersey.”" Page 4 #question,
EmotionallyAwareAgentsDispute_Analysis	how many lines in advance is enough context? #question,
EmotionallyAwareAgentsDispute_Analysis	"we assess how each emotion label correlates with self-reported frustration for each annotation method." Page 5 #question,
CanLanguageModels_Analysis	"he humanmediator followed instructions (i.e., they role-played as a mediator);" Page 6 #question,
CanLanguageModels_Analysis	in what cases did they not? #question,
haleKODIS_NAACL_Annotatedpdf_Analysis	"Finally, 117 we assess several theoretical mechanisms surround118 ing the dispute, including process variables (e.g., 119 what tactics did parties use, what emotions were 120 expressed, and did parties understand

Paper (16)	Related Annotations
	their part121 ner’s interests?) and outcome variables, including 122 objective and subjective measures concerning the 123 outcome of the dispute." Page 2 #methodology , , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"Post130 conflict measures include beliefs about whether 131 their partner was human or AI and attitudes towards 132 using AI technology for such applications." Page 2 #methodology , , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"cultural variability, 284 we use “elicited preferences,” meaning that partic285 ipants are provided a fixed number of points they 286 can allocate across the four issues" Page 4 #methodology ,
haleKODIS_NAACL_Annotatedpdf_Analysis	choice modeling #methodology ,
haleKODIS_NAACL_Annotatedpdf_Analysis	"The concept of “integrative 392 potential” measures the potential for joint gains 393 (which may or may not be realized)." Page 5 #methodology , , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	this is related to the monetary bonus of good performance and the participant's self-reported goals. Utility has multiple possible solutions-- what goal they expect may change in weight #methodology , , #variables
haleKODIS_NAACL_Annotatedpdf_Analysis	"They are asked to do the same for 411 their partner. The distance between these estimates 412 and their partner’s actual preferences can serve as 413 a measure of perspective-taking accuracy." Page 5 #methodology , , #self-report , #perspective-taking
haleKODIS_NAACL_Annotatedpdf_Analysis	This is compared to the initial assignment, but is the scale different? #methodology , , #self-report , #perspective-taking
haleKODIS_NAACL_Annotatedpdf_Analysis	"GPT4o (run on 06/28/2024) (Achiam et al., 548 2023) is prompted to annotate each dialogue 549 turn." Page 7 #emotion , #methodology ,
haleKODIS_NAACL_Annotatedpdf_Analysis	annotate text in turn-based manner #emotion , #methodology ,

#conversational-forecasting, #dispute-theory, #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes?

#dispute-theory, #idea, #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"ntuition that comments in a conversation are not independent events; rather, the order in which they appear matters (e.g., a blunt comment followed by a polite one feels intuitively different from a polite comment followed by a blunt one)." Page 8
changTroubleHorizonForecasting2019_Analysis	do we care about patterns for learning dispute structure? Relation to emotional recognition?

#challenge, #dispute-theory, #fairness, #potential-gap

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"parency, precluding an analysis of how exactly CRAFT models conversational context." Page 8
changTroubleHorizonForecasting2019_Analysis	application of fairness?

#contribution, #conversational-forecasting

Paper (1)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"his model fills a void in the existing literature on conversational forecasting, simultaneously addressing the dual challenges of capturing inter-comment dynamics and dealing with an unknown horizon." Page 9

#conversational-forecasting , #question

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"including those for which the outcome is extraneous to the conversation." Page 9
changTroubleHorizonForecasting2019_Analysis	what does this mean? the impact is outside of the actual conversation?

#dispute-theory , #question

Paper (4)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"order-sensitive representation of conversational context?" Page 8 #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	to what extent does order matter for disputes? #conversational-forecasting ,
changTroubleHorizonForecasting2019_Analysis	"A practical limitation of the current analysis is that it relies on balanced datasets, while derailment is a relatively rare event for which a more restrictive trigger threshold would be appropriate." Page 9
changTroubleHorizonForecasting2019_Analysis	does balanced matter for disputes? do we need to quantify how detailed impasses or spirals were?

#conversational-forecasting , #dispute-theory , #potential-gap

Paper (2)	Related Annotations
changTroubleHorizonForecasting2019_Analysis	"n reality, derailment need not spell the end of a conversation; it is possible that a conversation could get back on track, suffer a repeat occurrence of antisocial behavior, or any number of other trajectories." Page 9
changTroubleHorizonForecasting2019_Analysis	What can we say about spiral shapes and impasses? Can disputes esalate and descalate in such a way an impasse would not be reached?