Deception-aware pragmatic inference

Katerina Papantoniou[1,2]

[1] Institute of Computer Science, FORTH, Heraklion, Greece
[2] Department of Computer Science, University of Crete, Heraklion, Greece
papanton@ics.forth.gr

Abstract. Pragmatic competence i.e., the ability to understand intended meaning is a long standing challenge in communication. This work proposes a computational framework for pragmatic inference that is based on reasoning by taking into account the realistic assumption that any communication could be deceptive and that intentions can be reflected in language use. Pragmatic inference is examined under the light of culture since as literature manifests, cultural differences is a crucial parameter in the communication process and often a cause of misinterpretations (e.g., false alarms about the deceptiveness of a message). In contrast to the current situation, the proposed approach takes a holistic stance, aiming to infer knowledge both from surface patterns in raw text and more formal structures beyond the text level.

Keywords: Pragmatics, Deception, Culture, Reasoning, NLP

1 Introduction

Every day, people through their interpersonal interactions run into deceptive-prone situations struggling to decipher deceptive signals in messages. In high-stake situations and in situations, where justice and equal treatment must be ensured, the ability to detect deception is of utmost importance. An example that falls into the above cases is the examination of asylum applications from trained personnel, a task neither easy nor rare.[1]

Let's consider Nizar, a 29 years old Syrian forced to leave his country and seeks asylum in an European country. Some of the questions that are raised during his interactions with the Asylum Office personnel:

- In case the communication is not totally honest, which strategies of deception are employed maybe from both parties (e.g., bluff, information hiding, outright lie, exaggeration etc.)?
- What both parties hypothesize about the portion of information the other party possess or ignore? What are their hypotheses about the beliefs of each side (e.g., Nizar believes that the examiner of his application believes that he didn't lie about his marital status or his political beliefs)?

[1] By the end of 2015, asylum-seekers were 3.2 million (UNHCR).

- How the train of thought will change/revise when a deceptive effort come to light? How the beliefs of both parties revised?
- Does the different culture affect deceptive signals and may this condition cause misconceptions?

In the above context, the participants whoever role they hold must be extremely skilled at mind-reading in order to interpret implicit or seemingly uncorrelated information and to reason about unexpressed beliefs and intentions. The problem is further exacerbated by the different cultural and cognitive background of the involved actors (Nizar and the Asylum office personnel).

The proposed work is motivated by such situations and aims to research deception, by concentrating on a more holistic angle compared to existing literature. We aim at providing a computational framework that will gain advantage from deception detection techniques over a text message in order to enhance and inform complex reasoning tasks (e.g., given that a message of A is deceptive, A hides something else, what points an agent chooses to undermine or stress). We believe that by combining deception detection techniques with reasoning approaches over the communicated content we can achieve a deeper understanding of a message and to infer information that is implied or is hidden. Next, we briefly discuss the three basic axes we decided to rely on namely deception, text and culture.

Deception is omnipresent in every facet of human life [6], for example in fake news, forged reviews, cheaters in exams, white lies, self-deception, concealment of truth for our beloved ones or in order to save face. Despite this pervasiveness of deception, humans are notoriously bad at distinguishing between lies and truth [3]. In experimental studies of detecting deception, accuracy is typically only slightly better than chance [3], even among trained people such as investigators or customs inspectors. So, it is evident the necessity for the development of reliable and ideally proactive and real-time deception detection approaches that can protect individuals and the common weal.

The input to our deception detection algorithms will be simple textual data. This approach has recently gained momentum in the field of deception detection. A combination of factors seem to lead in this turn namely the advances in fields of Natural Language Processing (NLP) and Computational Linguistics, the enormous production of textual data, the seminal work of Vrij [20] and pure need as in many cases text is the only available source or more affordable and less intrusive (e.g., MRI). Last but not least, we prioritize the use of text since it allows us to gain insight over realistic situations.

We place great emphasis on examining deception taking into account the cultural characteristic of the potential liar. As studies show, people of other ethnic group when try to detect deception perform even worse than judging people of their own ethnic group [7]. For example, in a cross-cultural interrogation setting, that the norms of the person under investigation is different from the norms of the investigator, false signals may arise that impede the interrogation process and reduce the investigator's confidence. The importance of culture is recognized by many law enforcement authorities such as the U.K Home Office that list cul-

tural differences and cultural awareness as one of the key issues in investigator training and development [7]. It is indicative that often officers are specialized in different world regions or countries and cases assigned to them accordingly [9]. From a different perspective, the consideration of culture in statistical NLP models can contribute as a form of debiasing, which is currently a vivid research thread [13].

The starting point of this work is that deception indicators are critical to be extracted not only from surface patterns in textual data (e.g., news articles, dialogues, reports etc) but also by taking advantage of the overall context that can reveal the intention behind a deception effort and implied information. This requires the formal (abstract) representation of the communication content and subsequently the application of the appropriate forms of reasoning.

2 State-of-the-art

The passage from unrestricted free form text to a formalization that will enable complex reasoning process is a genuine arduous project. A lot of endeavours for the representation of semantics and pragmatics to logical forms have been presented in the literature. A comprehensive overview of this efforts such as LFG (Lexical functional Grammar), HPSG (Head-driven phrase structure grammar) and DRT (Discourse Representation Theory) is provided in [5] and [17]. The current state-of-the art in transforming dependency structures to logical forms succeeds in representing underlying predicate-argument structures, in an almost language-independent manner [17]. This work will be the basis for our transition to logical forms.

As far as the deception representation and reasoning is concerned a recent and very close to our goals work is that of Licato [14] that attempts to model the complex reasoning and deceptive planning used in an episode of the popular TV series "Breaking Bad". The author extend Cognitive Event Calculus to represent knowledge that involves nested beliefs, desires and intentions. For the representation of plans actions schemas were used, while for nested beliefs and all the alternative possibilities in a plan, he used non-monotonic reasoning and specifically *default reasoning*. In [4] the authors propose a model of belief and intention change over the course of a dialogue in which the decisions taken during the dialogue affect the possibly conflicting goals of the agents involved. They used Situation Calculus to model the evolution of the world and an observation model to analyze the evolution of intentions and beliefs. Their formalization is illustrated within the game of Werewolf, a party game that is frequently is used as use case in deceptive studies. A complete but mainly theoretical formal account of dishonesty is presented in [18]. The authors introduce a propositional multi-modal logic that can represent an agent's belief and intention as well as communication between agents. They handle different categories of dishonesty namely lies, bullshit, withholding information and half-truths.

In the context of culture modelling an important contribution is offered in [19]. The CARA architecture (Cognitive Architecture for Reasoning about Ad-

versaries) supports methods to gather data about different cultural groups and learn the intensity of those groups' opinions on various topics. The aspect of culture is modelled through rules taking advantage of knowledge extracted from the Web [1] and from prior theoretical studies. Rules are also used in [12] to model culture for a trade agents scenario. They model culture based on one of the five dimensions of culture according to Hofstede: individualism versus collectivism.

3 Proposed Approach

In an abstract level the the proposed approach constitutes of two processes (Figure 1). The first one is responsible to decide about the overall deceptiveness of a textual message and the extraction of fine-grained information such as type of deception based mainly on NLP approaches. This task feeds the pragmatics inference task that realizes the modelling of complex situations where reasoning beyond the text level is needed to understand hidden intentions and beliefs as those reflected in the introductory example. As Figure 1 depicts, culture is a modular parameter for both tasks, thus our approach can be applied also in contexts where culture is not so critical.

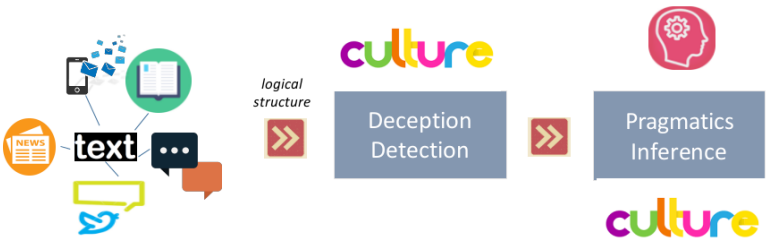

Fig. 1. A high level overview of the proposed approach

3.1 Deception Detection

Deception detection from a computational linguistics viewpoint focuses on linguistics differences between deception and truth-telling. A lot of theories back up this approach (psychoanalytic approach of Freud, Lexical Hypothesis) and in this respect a long list of lexical features almost for any level level of lexical analysis (e.g., morphology, syntax, discourse, psycholinguistics) has been examined. Since this research goal is out of the focus of this paper we briefly mention the directions that we will base our efforts upon:

Cross-linguistic and cross-culture deception detection
A reasonable argument that has already started to be explored [15] is if the

world's languages differences affect deception linguistic cues.[2] Drawing on prior research on psychology and sociology we want to examine in a larger scale indications about the existence of discriminating cues that are universally applicable across cultures. Equally important and under investigation is to understand the differences between cultures that are reflected in language use and maybe lead to misconceptions. For example, anxiety and awkwardness because of communication obstacles or politeness as an inherent characteristic in some cultures.

Deception types

A large body of the deception literature has been devoted to the typology of deception and to the identification of subtle differences between types of deception [16] [10] (e.g., distraction, concealments, white lies etc.). As very little work has been done [2] towards the discrimination between deception strategies from text data, we concentrate our efforts in this direction.

As far as the input data availability is concerned the vivid interest for deception detection has lead to the creation of a considerable number of publicly available textual datasets from diverse domains (e.g., news, reviews, court data). We plan to base our work is such datasets and perhaps to expand this pool of data with datasets for the Greek language.

3.2 Pragmatics Inference

The first step in this task is the structured representation of the communicated content. We anticipate to formalize natural language just to the extent that it allows us to transfer certain information to a formal context taking advantage of the recent advances in this direction [17]. Since none of the available modal logics we reviewed is able to fully cover our requirements [18], we plan to introduce a new modal logic that builds upon the first-order Event Calculus (EC). EC has become almost the natural choice for the modelling of natural language narratives since to some extent it can capture natural language semantics. In addition, the reformulation of EC in terms stable model semantics that can be computed by Answer Set Programming (ASP) solvers make EC a robust choice.

In our case, a critical requirement is nonmonotonicity due to dynamic changes in belief, intentions and knowledge. For instance, the presence of a deceptive message maybe lead to the revision of the existing knowledge or the creation of alternative paths that due to explosion in the quantity of knowledge must be handled. The parameter of time is another important requirement since the sequence of facts cannot be ignored.

As we have already argued, in an inter-cultural setting, the cultural gap may play a decisive role in communication since often it could be the cause of misconceptions and misunderstandings. In this respect, we must be able to incorporate knowledge about culture. This culture-driven knowledge can take two forms: a. the form of context-dependent knowledge about values, norms, relations, opinions, stance towards life (e.g., perceptions about family bonds) and b. linguistic

[2] The relationship between language and culture is supported by several theories among them the Sapir-Wholf hypothesis that language influences cognition.

expression of cultural differences (e.g., some cultures are engaged to small talk while others are more tolerant to persuasion). The theoretical support is provided by the research for cross-cultural deception detection [8], the differences between high and low context cultures [8] and the cultural dimensions as reflected in the work of Hofstede [11]. From an implementation perspective the obvious choice is the modelling through rules however the challenge is the validation of these rules perhaps by using external sources as validation mechanism in order to avoid the modelling of stereotypical and prejudged knowledge.

The nature of the examined problem guide us to forms of reasoning that make inferences beyond the scope of the premises as well as the ability to backtrack. For that reason, we plan to define a new form of reasoning that can be placed under the umbrella of *Ampliative and Defeasible Reasoning*. In addition, we must also take into account the differences in reasoning that emerge from culture (e.g., since a culture expressed in a particular way what conclusion could be inferred). Lastly, returning to the initial discussion about mind-reading a challenge is to incorporate aspects of *Counterfactual Reasoning* as a way to examine the viewpoint of the "Other" and avoid problems like confirmation bias. Table 1 provides a sketch of the key requirements that must be fulfilled in respect of representation and reasoning.

Table 1: Requirements analysis

Representation		
Type	**Nested**	**Example**
Time		Events in a course of a dialogue, time sequence in a narrative
Extensional		All Europeans citizens are individualists
Intensional	✓	Nizar believes that the examiner of his application believes that the well-being of his family is a priority for him
Ignorance	✓	Investigator ignores that...
Deception Strategies	✓	Investigator **hides** from Nizar that he holds information about his past, Nizar **lied** about his involvement in the civil war of this country, Officer **bluffs** about...

Reasoning	
Type	**Example**
Defeasible	If Nizar lied for his economical status, he previously also lied...
Ampliative	Nizar does not disclose information about his war experiences. I can conclude that maybe suffers for post-traumatic stress disorder
Counterfactual	Examine the case when a lie was not a lie but the result of confirmation bias
Cultural	Nizar as collectivist, values high the institution of family, so the white lie in relation to some family members was indeed an effort to protect them

4 Conclusions & Future Steps

In this paper, we present our proposal for a deceptive-aware computational framework for pragmatics inference. We aim at a closer collaboration between Knowledge Representation and Reasoning with NLP in order to offer a deeper understanding of the communicated content. We prioritize the influence of culture since, as literature emphatically manifests, it is a crucial parameter and a constant source of misconceptions. Our prospective goal is two-fold, from one side to offer more realistic deceptive aware multi-agent environments that require complex forms of reasoning and from the other side to contribute to Computational Pragmatics by based on formal logic for the pragmatic interpretation.

Our immediate steps is to complete our requirements analysis for the deceptive-aware modal logic while in parallel we work on towards the deception detection from text task.

Acknowledgements

This work is funded by the Institute of Computer Science (ICS) of the Foundation for Research and Technology - Hellas (FORTH) and is conducted under the supervision of Prof. Dimitrios Plexousakis and in cooperation with Dr. Giorgos Flouris, Dr. Theodoros Patkos and Prof. Ion Androutsopoulos.

References

1. Massimiliano Albanese and VS Subrahmanian. T-rex: A domain-independent system for automated cultural information extraction. In *Proceedings of the First International Conference on Computational Cultural Dynamics (ICCCD 2007)*, 2007.
2. Darren Scott Appling, Erica J. Briscoe, and Clayton J. Hutto. Discriminative models for predicting deception strategies. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 947–952, New York, NY, USA, 2015. ACM.
3. Bella M DePaulo, Julie I Stone, and G Daniel Lassiter. Deceiving and detecting deceit. *The self and social life*, 323, 1985.
4. Codruta Liliana Gîrlea, Eyal Amir, and Roxana Girju. Tracking beliefs and intentions in the werewolf game. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*, 2014.
5. Naveen Sundar Govindarajulu, Selmer Bringsjord, and John Licato. On deep computational formalization of natural language. In *Proceedings of the Workshop:Formalizing Mechanisms for Artificial General Intelligence and Cognition(Formal MAGiC) at Artificial General Intelligence*, 2013.
6. PA Granhag, LA Strömwall, and (Eds). *The detection of deception in forensic contexts*. Cambridge University Press, 2004.
7. Pär A. Granhag, Aldert Vrij, and Bruno Verschuere. *Detecting Deception: Current Challenges and Cognitive Approaches (Wiley Series in Psychology of Crime, Policing and Law)*. Wiley-Blackwell, 1 edition, October 2014.

8. Par Granhag, A. Vrij, and B. Verschuere. *Detecting deception : current challenges and cognitive approaches*. Wiley, Hoboken, 2015.
9. Pr Anders Granhag, Leif A. Strmwall, and Maria Hartwig. Granting asylum or not? migration board personnel's beliefs about deception. *Journal of Ethnic and Migration Studies*, 31(1):29–50, 2005.
10. Maria Hartwig. *Interrogating to detect deception and truth: Effects of strategic use of evidence*. 2005.
11. Geert H. Hofstede. *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Sage, Thousand Oaks, CA, 2nd and enlarged edition, 2001.
12. Gert Jan Hofstede, Catholijn M. Jonker, and Tim Verwaart. *Individualism and Collectivism in Trade Agents*, pages 492–501. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
13. Dirk Hovy. Demographic factors improve classification performance. 2015.
14. Licato John. Formalizing deceptive reasoning in breaking bad: Default reasoning in a doxastic logic. In *Proceedings from the 2nd AAAI Symposium on Deceptive and Counter-Deceptive Machines (DCDM 2015), Arlington*, 2015.
15. Verónica Pérez-Rosas and Rada Mihalcea. Cross-cultural deception detection. In *ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, pages 440–445, 2014.
16. Tiantian Qin and Judee K. Burgoon. *An Empirical Study on Dynamic Effects on Deception Detection*, pages 597–599. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
17. Siva Reddy, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman, and Mirella Lapata. Transforming Dependency Structures to Logical Forms for Semantic Parsing. *Transactions of the Association for Computational Linguistics*, 4:127–140, 2016.
18. Chiaki Sakama, Martin Caminada, and Andreas Herzig. A formal account of dishonesty. *Logic Journal of the IGPL*, 23(2):259, 2014.
19. VS Subrahmanian, Massimiliano Albanese, Maria Vanina Martinez, Dana Nau, Diego Reforgiato, Gerardo I Simari, Amy Sliva, Octavian Udrea, and Jonathan Wilkenfeld. Cara: A cultural-reasoning architecture. *IEEE Intelligent Systems*, pages 12–16, 2007.
20. Aldert Vrij, P. Granhag, and S. Porter. Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science In The Public Interest*, 11(3):89–121, 12 2010.