inSIDE Fair Dialogues: Assessing and Maintaining Fairness in Human-Computer-Interaction

Sabine Janzen[1], Ralf Bleymehl[2], Aftab Alam[2], Sascha Xu[1], Hannah Stein[2]

Deutsches Forschungszentrum für Künstliche Intelligenz[1]
Information and Service Systems, Saarland University[2]

sabine.janzen@dfki.de, s8afalam@stud.uni-saarland.de, ralf.bleymehl@iss.uni-saarland.de, sascha.xu@dfki.de, hannah.stein@iss.uni-saarland.de

Abstract

For simulating human-like intelligence in dialogue systems, individual and partially conflicting motives of interlocutors have to be processed in dialogue planning. Little attention has been given to this topic in dialogue planning in contrast to dialogues that are fully aligned with anticipated user motives. When considering dialogues with congruent and incongruent interlocutor motives like sales dialogues, dialogue systems need to find a balance between competition and cooperation. As a means for balancing such mixed motives in dialogues, we introduce the concept of fairness defined as combination of fairness state and fairness maintenance process. Focusing on a dialogue between human and robot in a retailing scenario, we show the application of the *SatIsficing Dialogue Engine (inSIDE)* - a platform for assessing and maintaining fairness in dialogues with mixed motives.

1 Introduction

For simulating human-like intelligence in dialogue systems, individual and partially conflicting motives of interlocutors have to be processed in dialogue planning. *Motives* are described as desires in the sense of a motivational state according to the belief-desire-intention model, e.g., to find the best price when shopping (Georgeff et al. 1998; Rao and Georgeff 1995a). From a computational linguistics perspective, they are equivalent with the concept of intentions in Levelt (1993). So far, dialogue systems were applied to situations in which interlocutors were highly cooperative and shared congruent motives. Perceived as effective when solving user problems, exceeding satisfaction of user motives and maximizing cooper-

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2018 in
R. Dachselt, G. Weber (Hrsg.):
Mensch und Computer 2018 – Workshopband, 02.–05. September 2018, Dresden.
 https://doi.org/10.18420/muc2018-demo-0482

ativeness (Bunt and Black 2000), these kinds of dialogue systems are well scrutinized, e.g., (Grosz and Kraus 1996; Moore and Paris 1993; Rich and Sidner 1997). Also the counterpart, i.e. dialogue planning with motives that are in pure conflict, was investigated in several works, e.g., (Hadjinikolis et al. 2013; Black and Atkinson 2011; Prakken 2006). In this work, we extend dialog systems to situations with congruent but also incongruent interlocutor motives subsumed by the term mixed motives. When considering dialogues with mixed motives, interlocutors are faced with a conflict between their motives to cooperate and to compete with each other (Schelling 1960). For simulating human cognitive abilities, dialogue systems need to find a balance between competition and cooperation when satisfying mixed motives. Based on research on natural dialogues in conflicting situations, we introduce fairness as a means for balancing mixed motives during the course of a dialogue. Fairness is defined statical as fairness state operationalizing an equal and adequate, i.e. equitable satisfaction of all interlocutor motives at any time (Oxford Dictionaries 2016). Since dialogues are dynamic, fairness state is combined with a fairness maintenance process for handling fairness during the course of dialogue (Doyle 1979).

We propose a model for assessing and maintaining fairness in dialogues that combines a mixed motive model with a game-theoretical equilibrium approach (Nash 1951). One appeal of the model is its holistic consideration and processing of all interlocutor motives in dialogue planning instead of processing individual interlocutor motives selectively (Moore and Paris 1993). Extending existing approaches, e.g., (Grosz and Kraus 1996), motives are represented in an integrated manner on (1) individual level and (2) in aggregation to mixed motives of all interlocutors on collective level. This is due to the fact that interlocutors have more than a single motive when participating in dialogues (Grosz and Sidner 1986). In contrast to approaches exclusively considering positive motivational structures (Georgeff et al. 1998; Rao and Georgeff 1995b), we characterize motives as positive, neutral and negative for capturing their heterogeneity (Schank and Abelson 1977; Konolige and Pollack 1993). We assume that this approach enables a more sophisticated simulation of human behavior in mixed motive interactions as well as a qualitative assessment and maintenance of the dialogue state in terms of fairness.

By exemplifying our model in a dialogue system as natural language sales assistant for conducting sales dialogues, we were able to evaluate the proposed model in an empirical study combining results of a run-time experiment with feedback by users (N=107) in terms of perceived achievement of motives as well as fairness of created dialogues with promising results (Janzen et al., 2016).

2 Mixed Motives in Dialogues

For giving a practical example, we describe a dialogue with mixed motives between a customer and a retailer in a shopping scenario:

Customer: *"I am searching for a low cost router. Is the range of this wifi router appropriate for a house with 3 floors?"*

Retailer: *"In case of larger distances or several obstacles as given by 3 floors, this router will come to an end. I would recommend an additional wifi repeater that got very good feedback by other customers. You can buy both router and repeater as a bundle with 15% discount."*

The example shows a dialogue snippet between two actors with different motives that are congruent and incongruent. Motives by the customer and the retailer for gaining respectively providing comprehensive product information are congruent. But, furthermore, the customer intends to find the best price whereas the retailer wants to increase revenue and to improve customer relationship. These motives are incongruent and partially conflicting. Nonetheless, actors are able to find a balance between selfishness, i.e. pursuing individual motives, and fair play, i.e. responding to anticipated motives of counterparts for creating fair dialogues. Here, a balance between mixed motives is found by giving information regarding the wifi router as well as preferences of other customers followed by a discounted bundle offer.

3 Demo Satisficing Dialogue Engine (inSIDE)

In order to explain this kind of behavior in computational terms, we specified our model for assessing and maintaining fairness under mixed motives in dialogues. The model separates linguistics from conceptual non-linguistic aspects (Traum and Larsson 2003) and consists of three main components: *mixed motive model, mapper* and *linguistic intention model*. The mixed motive model combines the explicit representation and situated processing of mixed motives with a game-theoretical equilibrium approach operated by the equilibrium identifier. The model operates by assuming that interlocutors are rational. That means they act strategically in pursuit of their own motives that they try to maximally satisfy. Therefore, we assume that game theory is an adequate prospect to deliver the analytical tools for assessing fairness in dialogues. In game theory literature, equilibrium concepts are widely applied, e.g., Nash equilibrium (Nash 1951). A Nash equilibrium is an outcome that holds because no involved actor has a rational incentive to deviate from it, i.e., the final result is "good enough" for all actors in the sense of a happy medium. Adapted to this work, this refers to a combination of motives at a particular time in the dialogue, that is good enough for planning dialogues that support equitable, i.e. fair satisfaction of mixed motives. By means of the second component - the mapper - motives are mapped onto linguistic intentions and vice versa. Therefore domain-specific knowledge about correlations between mixed motives and linguistic intentions is required that is induced by a domain configurator and has to be derived empirically. Last, the linguistic intention model covers linguistic intentions that capture intended effects of single text segments (Grosz and Sidner 1986). This reflects the fact that text segments fulfill specific functions regarding the whole text (Moore and Paris 1993; Hovy 1988). Linguistic intentions can be used as triggers for generating text and therefore, they contribute to the achievement of motives in dialogues.

Based on the proposed model, we implemented the dialogue platform *SatIsficing Dialogue Engine (inSIDE)* that offers a REST API for the development of light-weight dialogue interfaces.

Our demo shows the application of inSIDE platform for realizing sales dialogues between customers and a service robot in a retail store (cf. Fig. 1). Both, customer and robot, have different motives for participating in the sales dialogue, e.g., searching for the best price or increasing revenue. Nonetheless, empowered by inSIDE the robot has the ability to find a balance between selfishness, i.e. pursuing individual motives, and fair play, i.e. responding to anticipated motives of the customer for creating a dialogue perceived as fair. Besides the ability to pro-actively initiating and conducting dialogues, the service robot is able to guide customers within the retail store for showing products and to present appropriate information on its display when required by the ongoing dialogue. In the demo itself, the first step is that the robot welcomes the customer. Then, the customer may ask questions about products. The robot reacts accordingly by showing details on its built-in table, listing potential products for the customer or giving spoken information. Furthermore, it is able to guide the customer to a specific product in the environment, i.e. if the customer asks for testing the product, the robot will offer to guide the customer to the product, so that he/she can test it. Another aspect of the demo is the proactive behavior when the robot detects that the customer remains in front of a product for a longer period of time. As a reaction, the robot approaches the customer again and asks if there is any more help needed. Main issues of the demo are: (1) assessing and maintaining fairness in dialogues with mixed motives in human-robot-interaction by means of SatIsficing Dialogue Engine (inSIDE); (2) proactive behavior by the robot when required in interaction; and (3) spatial guiding of customer when required in interaction.

Fig. 1: Sales dialogue between customer and robot in demo

References

Black, E., & Atkinson, K. (2011). Choosing persuasive arguments for action. In *10th Int. Conf. on Autonomous Agents and Multiagent Systems*, 905–912.

Bunt, H., & Black, W. (2000). The abc of computational pragmatics. *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics* 1–46.

Doyle, J. (1979). A truth maintenance system. *Artificial intelligence* 12(3):231–272.

Georgeff, M. P.; Pell, B.; Pollack, M. E.; Tambe, M.; & Wooldridge, M. (1998). The belief-desire-intention model of agency. In *Proc. of the 5th Int. Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages*, 1– 10.

Grosz, B. J., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence* 86(2):269 – 357.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Comput. Linguist.* 12(3):175– 204.

Hadjinikolis, C.; Siantos, Y.; Modgil, S.; Black, E.; & McBurney, P. (2013). Opponent modelling in persuasion dialogues. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence*, 164–170.

Hovy, E. H. (1988). Generating natural language under pragmatic constraints. *Journal of Pragmatics* 11(6):689–719.

Janzen, S., Maass, W. & Kowatsch, T. (2016). Finding the Middle Ground - A Model for Planning Satisficing Answers,.*Proc. of 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*.

Konolige, K., & Pollack, M. E. (1993). A representationalist theory of intention. In *Proc. of the 13th Int. Joint Conf. on Art. Int. (IJCAI 1993)*, 390–395.

Levelt, W. J. (1993). *Speaking: From intention to articulation*, volume 1. MIT press.

Moore, J. D., & Paris, C. L. (1993). Planning text for advisory dialogues: capturing intentional and rhetorical information. *Comput. Linguist.* 19(4):651–694.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics* 54(2):286–295.

Oxford Dictionaries. (2016). ”fair”. Oxford University Press (http://oxforddictionaries.com).

Prakken, H. (2006). Formal systems for persuasion dialogue. *The Knowledge Engineering Review* 21(02):163–188.

Rao, A. S., & Georgeff, M. P. (1995a). Bdi agents: From theory to practice. In *Proc. of the 1st Intl. Conf. on Multiagent Systems (ICMAS)*, 312–319. The MIT Press.

Rao, A. S., & Georgeff, M. P. (1995b). *Formal models and decision procedures for multi-agent systems.* Technical Note 61, Australian Artificial Intelligence Institute.

Rich, C., & Sidner, C. L. (1997). Collagen: when agents collaborate with people. In *Proceedings of the first international conference on Autonomous agents*, AGENTS ’97, 284–291. New York, NY, USA: ACM.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard university press.

Traum, D., & Larsson, S. (2003). *The information state approach to dialogue management*. Springer. 325–353.