Intrusiveness, Trust and Argumentation: Using Automated Negotiation to Inhibit the Transmission of Disruptive Information

P. Panzarasa, S. D. Ramchurn and N. R. Jennings
Department of Electronics and Computer Science
University of Southampton
{pp, sdr01r, nrj}@ecs.soton.ac.uk

Abstract

The question of how to promote the growth and diffusion of information has been extensively addressed by a wide research community. A common assumption underpinning most studies is that the information to be transmitted is useful and of high quality. In this paper, we endorse a complementary perspective. We investigate how the growth and diffusion of high quality information can be managed and maximized by preventing, dampening and minimizing the diffusion of low quality, unwanted information. To this end, we focus on the conflict between pervasive computing environments and the joint activities undertaken in parallel local social contexts. When technologies for distributed activities (e.g. mobile technology) develop, both artifacts and services that enable people to participate in non-local contexts are likely to intrude on local situations. As a mechanism for minimizing the intrusion of the technology, we develop a computational model of argumentation-based negotiation among autonomous agents. A key component in the model is played by trust: what arguments are used and how they are evaluated depend on how trustworthy the agents judge one another. To gain an insight into the implications of the model, we conduct a number of virtual experiments. Results enable us to explore how intrusiveness is affected by trust, the negotiation network and the agents' abilities of conducting argumentation.

Contact:
Pietro Panzarasa
Department of Electronics and Computer Science
University of Southampton
Highfield
Southampton, SO17 1BJ, U.K.

Tel: +44 (0)23-8059-3089
Fax: +44 (0)23-8059-2865
Email: pp@ecs.soton.ac.uk

Key Words: autonomous agents, persuasion, reputation, network structure, information diffusion..

Intrusiveness, Trust and Argumentation: Using Automated Negotiation to Inhibit the Transmission of Disruptive Information

Pietro Panzarasa, Sarvapali D. Ramchurn and Nicholas R. Jennings

This paper is concerned with the problem of inhibiting the diffusion of low quality, disruptive information within a social context. How can the performance of a joint activity be enhanced by controlling the amount and quality of the information that the participants receive? How should interruptions be managed without losing any potentially important and high quality information? How should a message be evaluated in terms of its potential intrusive impact upon the execution of a joint activity? As a possible mechanism for addressing these problems, we propose a model of argumentation-based negotiation among autonomous agents [Sierra *et al*., 1998]. Messages are received and evaluated by agents on behalf of their users who, in turn, are involved in a parallel joint activity. Upon receiving a message, the agents negotiate with one another in order to reach an agreement as to whether or not their users' joint activity should be interrupted by the transmission of the message. During negotiation, the agents use arguments and counter-arguments in an attempt to impact upon each other's mental states [Panzarasa *et al.*, 2001]. In our model of argumentation-based negotiation, a fundamental role is played by the concept of trust. In fact, the extent to which argumentation is a successful mechanism for managing intrusiveness depends on what type of arguments are used and on how they are evaluated by the agents. In turn, both the choice of arguments and how the agents react to them depend on how trustworthy the agents judge one another. To explore the implications of the model, we conduct a series of virtual experiments. Results show the individual and combined effects of argumentation and trust upon the generation of a final agreement and, ultimately, on how effectively intrusiveness is dealt with.

Intrusiveness and the Diffusion of Information

The question of how to promote the growth and diffusion of information has been extensively addressed by a wide research community. Early work on the diffusion of innovations has typically emphasized the role of the network structure in the promotion and spread of new information [Valente, 1996]. Similarly, recent work on the internet, social networks and the power grid has addressed the resilience of these systems to perturbations of the pattern of interaction between their components [Albert *et al.*, 2000]. A common assumption underpinning all these studies is that the information to be created and propagated is useful and of high quality. However, relatively little research has focussed on the equally pressing question of how to prevent the diffusion of unwanted information. In fact, one of the ways to promote the growth and diffusion of high quality information is to ensure that low quality information can be identified and eradicated. Along these lines, in this paper we address the problem of how to manage and maximize the growth and diffusion of high quality information by focussing upon the complementary issue of how to prevent, dampen and minimize the diffusion of low quality, unwanted information.

Clearly, whether or not a piece of information is of high quality and is, therefore, to be transmitted, depends not only on how the information was generated and by whom, but also on the target of the transmission. This involves the recipient's mental state (including goals, beliefs, preferences, etc.), what task is being carried out and the environment in which the recipient is located. Typically, low quality information includes unverified, inaccurate, erroneous messages. However, there are cases in which even severely flawed messages might convey high quality meta-information. This might happen if, for example, the main task of the recipient is precisely to evaluate the trustworthiness of the sender. This suggests that an appropriate criterion for determining whether a piece of information is of high or low quality is to evaluate whether or not that information helps the recipient to sustain or improve the performance of its task. The diffusion of low quality information typically brings about interference with the realization of the main task. Here, we use the notion of intrusiveness to refer to such interference caused by the spread of useless, disruptive and damaging information.

Recent work in human-computer interaction has suggested a number of ways for dealing with intrusiveness [McFarlane, 1999]. The idea is to promote the diffusion of constructive task-related information by efficiently handling the interruptions caused by incoming messages. On the one hand, any interruption should be considered in order to ensure that any important incoming information will be received. On the other, however, being constantly interrupted by any incoming message might jeopardize the efficient performance of the main task should the received message convey unimportant, disruptive information. Thus, the recipient of a message of unknown informational value is confronted with a trade-off. Being interrupted by any incoming message ensures that no potentially important information is lost, but this brings about the risk of being interrupted by intrusions detrimental to the performance of the recipient's main task. Overcoming this trade-off requires mechanisms for dealing with

incoming messages that, while maximizing the access to interruptions of informational value, at the same time minimize the effects that possible intrusions might have upon the recipient's task. For example, repeatedly checking an interface can provide information updates, while at the same time allowing the recipient to choose when to deal with the interruptions and possible intrusions. Similarly, interruptions can be managed through an alerting interface or through peripheral broadcast of the messages.

All these mechanisms are concerned with the problem of how to manage interactions by determining how and when to have access to them. However, the core problem of elucidating efficient mechanisms for filtering out intrusiveness remains relatively unexplored. Furthermore, most research has concentrated on the effects of interruptions on isolated recipients. Conversely, little work has focussed on the problem of how to manage the potentially disruptive effects caused by unexpected interruptions upon the execution of a joint task (e.g. within a group, an organization). This paper presents one possible approach to these problems by proposing argumentation-based negotiation among software agents as a mechanism for dealing with intrusiveness within a social context. Agents receive messages on behalf of their users who, in turn, are engaged in the execution of a joint task. Upon receiving a message, the agents assess the value of the information conveyed. They then negotiate with one another in order to get to a joint decision as to whether or not the execution of the users' joint task should be interrupted by the transmission of the message. This joint decision will depend on the value that each agent assigns to the message received, on the arguments used during negotiation and on how trustworthy the agents judge one another during their interactions.

Persuasive Negotiation among Autonomous Agents

A fundamental component of our model of negotiation is the agents' abilities to persuade each other by using arguments. Argumentation-based negotiation represents one of the ways in which autonomous interacting agents manage their interdependencies and coordinate their activities [Sierra *et al.*, 1998]. In this form of negotiation, agents generate and exchange arguments that indicate the reasons why their proposals should be accepted. In this sense, argumentation-based negotiation is persuasive because the arguments exchanged are aimed at impacting upon, and thereby altering, the mental states (e.g. preferences, acceptability thresholds, goals) of the agents involved [Panzarasa *et al*., 2001]. Arguments may take the form of threats, rewards, appeals or they may merely convey information reflecting the agents' beliefs. Their key function is to speed up the generation of a final agreement by making the agents' proposals more attractive to the counterparts. However, how effective and appropriate an argument is depends on the agents involved in negotiation and on the social structure in which argumentation takes place. In our model, agents are endowed with a reasoning mechanism through which they can autonomously choose what argument to use for justifying or supporting their proposals. The value that a proponent assigns to an argument is here formalized as a function of a number of attributes including the number of times the argument turned out to be successful in a set of previous negotiations. In addition, once an argument has been received, it needs to be evaluated. Evaluation occurs by checking the validity of the argument against the recipient's mental state. An argument is rejected when, once evaluated against the recipient's goals and beliefs, it does not provide sufficient support to a proposal. In this case, the argument fails to impact on the recipient's mental state. Conversely, the acceptance of an argument alters the recipient's mental state by modifying its range of acceptability and/or its rating function over this range.

Reputation and the Social Structure of Trust

Another basic component of our model is trust. It is often argued that trust is a fundamental prerequisite for good organizational performance [Nicholas, 1993]. Without trust, information needs to be rechecked, decisions need to be reevaluated, agreements take longer to be reached and collaboration diminishes. Recent work on communication in a simulated organizational task suggests that the extent to which truth telling affects organizational performance depends on the environmental conditions and the size of the organization [Prietula and Carley, 1999]. Trust is typically facilitated by face-to-face interaction and affects people's judgements by triggering emotional and cognitive responses. Thus, as we move to distributed groups, virtual organizations and web-based teams, the sources of trust and its effects on the agents' behavior are likely to become more critical and controversial. In such distributed environments, is trust always necessary for good performance? To what extent does trust impact on the choice and evaluation of the arguments that agents exchange among one another to come to an agreement? Under what conditions is trust likely to emerge within a network of negotiating agents?

We address these questions by modeling trust as a cognitive and structural construct. In our model, a critical component of trust is reputation. An agent's reputation is a characteristic or attribute ascribed to it by its partners [Raub and Weesie, 1990]. An important source of an agent's reputation is its observed behavior. Information about

an agent's interaction with a current partner can be used in subsequent interactions by this partner or by other partners when contemplating their own interactions with that agent. Thus, a fundamental empirical basis of an agent's reputation is the direct observation of its current and past social behavior. However, another important source of an agent's reputation is the information about its social behavior that can be obtained by communicating with other agents. In turn, the extent to which this information is judged accurate and trustworthy depends upon the established reputation of the agents from whom the information is obtained. Therefore, cognition and social structure work together to produce a combined effect upon the generation of trust. How trustworthy an agent is judged by another agent depends on what the latter knows about the former. And this knowledge results from the agents' positions in a network of social relations [Carley, 2001]. "Who interacts with whom" affects "who knows what", and "who knows what" represents the cognitive source of reputation. Thus, trust can be regarded as dependent upon the "embeddedness" of the agents' interactions in structures of social relations [Buskens, 1998]. In turn, since "who knows what" impacts upon "who interacts with whom", trust motivates the agents' social behavior and, ultimately, drives the establishment and evolution of social relations.

Trust and Arguments

Trust and argumentation are inherently intertwined and dynamically affect each other during the course of negotiation. Since an agent's reputation depends on its observed behavior, it is affected by information about the type of arguments used by the agent in its past negotiations. To capture how the cognitive and structural components of trust are shaped by a process of persuasive negotiation, we formalize trust as a function of: (i) the structural properties of the negotiation network; (ii) the frequency of negotiation over time; (iii) the outcomes of past negotiations; and (iv) the arguments used to impact on the partners' mental states. Clearly, how an agent is structurally connected to a potential partner determines how accurate the information it has about the partner's past behavior is. This, in turn, will determine the extent to which that information will affect the agent's assessment of its partner's reputation. Furthermore, trust is affected by the frequency of negotiation: interacting repeatedly with the same agent over a long period of time is likely to generate stronger effects on that agent's reputation than only one single interaction in the past. How trustworthy an agent is regarded also depends on the success or failure of previous negotiations with that agent. A broken commitment, for example, is likely to be detrimental to an agent's reputation, thus making future negotiation with that agent less attractive. Finally, trust is affected by how an agreement is reached, namely by the type of arguments that the agents use to persuade one another. For example, an agent's decision to speed up negotiation by threatening its partners will make these and other potential partners less inclined to interact with the same agent in future negotiations. Or, if an agent does not fulfil the promises made to persuade its previous partners, then making promises will become a less effective strategy for impacting upon future partners' mental states.

Even though trust is affected by negotiation, nonetheless it affects negotiation in a two-fold way. Firstly, judgments about an agent's trustworthiness impact upon the type of arguments used by the agent's partners in their attempts to support their proposals. For example, using a threat during negotiation with an agent of low reputation is more likely than simply providing meta-level information in the form of logical statements. Secondly, an agent's reputation determines how its partners evaluate and react to the arguments they receive from that agent. An agent's high reputation is likely to emphasize the effects that its arguments have on its partners' mental states. Conversely, low reputation mitigates the strength of arguments in terms of their potential for making proposals more attractive and acceptable. For example, the extent to which a statement indicating the reasons why a proposal should be accepted is convincing, depends on how trustworthy the sender is. Similarly, the effectiveness of a promise depends on the reputation of the agent who made it, and particularly on whether or not its past promises were honored. Thus, trust impacts on the recipient's reaction to the received proposal, and in particular on whether the proposal will be rejected, accepted or modified. Finally, trust will affect the generation of further arguments for supporting counterproposals.

Virtual Experiments

Using our model of argumentation-based negotiation, a series of experiments are conducted. The domain under consideration is a meeting room scenario in which each participant has a number of personal devices (e.g. cellular phones, laptops, PDAs). Each message received through these devices might have different informational value. For example, it might be important for both the recipient and the other individuals attending the meeting. In this case, the meeting should be interrupted and the message displayed within the meeting room. However, messages can also have low informational value for both the recipient and the meeting, or they can be important to the recipient, but intrude on the joint task undertaken within the meeting. Thus, whenever a message is received by any of the

participants' devices, a decision is to be made as to whether or not the meeting should be interrupted and the message displayed within the meeting room. The idea is to explore the potential conflict between activities in a local physical and social context (the meeting room) and parallel activities in distributed contexts (the software environment). When technologies for distributed activities (e.g. mobile technology) develop, both artifacts and services that enable people to participate in non-local contexts are likely to intrude more and more on local situations. Potentially this phenomenon is a threat to the effectiveness and efficiency of the activities undertaken in such local situations. Thus, to minimize the intrusion of the technology, pervasive computing environments necessitate intelligent management.

Software agents can be used to handle such heterogeneous computing environments. Each individual (user) in the meeting room is represented by a personal agent located in a parallel software environment. Each agent controls the devices possessed by its user and evaluates the messages received by these devices against the user's preferences. To minimize the intrusive impact of the devices they control, agents then negotiate with one another and eventually get to a final joint decision as to whether or not to display the messages received. Agents use argumentation to support negotiation, and argumentation, in turn, affects and is affected by how trustworthy the agents judge one another. We explore the individual and combined effects of argumentation and trust on intrusiveness from a three-fold perspective. Firstly, we look at different degree of sophistication of the agents' cognition, and we examine to what extent different capabilities of conducting argumentation affect the rate of information diffusion and the level of intrusiveness caused by the messages displayed. Secondly, we investigate the impact that the network structure in which negotiation takes place has upon the agents' reputation and, ultimately, upon agreement generation and intrusiveness. Along these lines, we ask whether the transmission of high quality information can be enhanced by destabilizing the negotiation network and by isolating agents according to their structural positions and/or their cognitive skills. Finally, we analyze the combined effects that the network structure and the agents' abilities to negotiate have upon intrusiveness. This will shed light on a number of issues concerned with the relation between network and cognitive complexity.

References

[Albert *et al*., 2000] Albert, R, H. Jeong & A.-L. Barabási, 2000, "Attack and Error Tolerance of Complex Networks." *Nature*, 406: 378-382.

[Buskens, 1998] Buskens, V., 1998, "The Social Structure of Trust." *Social Networks*, 20: 265-289.

[Carley & Hill, 2001] Carley, K. M. &V. Hill, 2001, "Structural Change and Learning within Organizations." In Lomi A. & E. R. Larsen (eds.), *Dynamics of Organizations. Computational Modeling and Organization Theories*, AAAI Press/ The MIT Press, Menlo Park, CA, 63-92.

[McFarlane, 1999] McFarlane, D, 1999, "Coordinating the Interruption of People in Human-Computer Interaction." In *Human-Computer Interaction - INTERACT99*, 295-303.

[Panzarasa *et al*., 2001] Panzarasa, P., T. J. Norman & N. R. Jennings, 2001, "Social Mental Shaping: Modelling the Impact of Sociality on Autonomous Agents' Mental States." *Computational Intelligence*, 17(4): 738-782.

[Prietula & Carley, 1999] Prietula, M. & K. M. Carley, 1999, "Exploring the Effects of Agent Trust and Benevolence in a Simulated Organizational Task." *Applied Artificial Intelligence*, 13(3): 321-338.

[Raub & Weesie, 1990] Raub, W. & J. Weesie, 1990, "Reputation and Efficiency in Social Interactions: An Example of Network Effects." *American Journal of Sociology*, 96(3): 626-654.

[Sierra *et al*., 1998] Sierra, C., N. R. Jennings, P. Noriega & S. Parsons, 1998, "A Framework for Argumentation-based Negotiation." In *Intelligent Agents IV*, Springer Verlag, 177-192.

[Valente, 1996} Valente, T. W, 1996, "Social Network Thresholds in the Diffusion of Innovations." *Social Networks*, 18(1): 69-89.