Towards Implementation of Social Interaction

Robert Zubek

Northwestern University Computer Science
1890 Maple Avenue
Evanston, IL 60201
rob@cs.northwestern.edu

Abstract

While it seems desirable to have game characters that interact socially with human players, implementing such interactions in game contexts remains difficult. This paper argues that the implementation of rudimentary social skills can be simplified considerably by explicitly modeling situations in which the agent may find himself and, furthermore, it can be accomplished without having to engage expensive linguistic models. The conclusion is a sketch of a new system that attempts such limited social interactions with human players.

Introduction

It is hardly surprising to notice that people enjoy interacting with other people. Playing, resting, eating, even working and performing chores, are often done in the company of others – sometimes out of necessity, but often out of simple preference. We enjoy the interactions we have with other people, the exchanges of attention, affection, and consideration. Our days are full of dealings with others, and this is often the preferred state of affairs. Even the most mundane endeavors, such as buying groceries or getting a haircut, seem more enjoyable when mediated by sociable people.

This importance of social interactions is clearly reflected in multiplayer computer games. Designers recognize that players enjoy games that connect them together and let them play with and against other human beings. Some genres, such as the persistent worlds exemplified by Ultima Online, EverQuest, and MUDs, make a point of actually grounding much of the gameplay in social interactions with other players. Online lives in these games are filled with engagements with others: conversations about what goes on in the world, building friendships and alliances, haggling, manipulating, congratulating, commiserating. Social engagements make living in an online world more enjoyable, and the world itself more believable. Playing with and against other people in the game is much more fun than playing by yourself.

Since players enjoy and naturally engage in social interactions, it seems desirable to build artificial characters that could interact with players in similar ways. Implementing such skill remains painfully difficult, however, and systems that attempt it usually result in interactions that are visibly scripted and quite brittle. In the current crop of popular multiplayer games in particular, the non-player characters (or NPCs) no longer even attempt to engage in much social interaction at all. They are often used as simple inert background characters, no more active than the trees and rocks around them; or worse yet, they're just vending machines, listening for keywords and dispensing items or bits of story information. In this view, they are no more than appliances, absolutely predictable and dependable. And interacting with these automatic assistants and shopkeepers is indeed about as much fun as talking to an ATM.

The problem is not simply the lack of computational power or autonomy. These characters may be quite complex, but none of their intelligence can show through if they cannot communicate. The point here is that they will be more enjoyable and perceived as more intelligent if they could interact socially and in human-like ways – if they were capable of more than just choice trees and graphical menus. Getting through an interaction efficiently is good, if the player is in a hurry, but equally important is being able to interact in more complex ways, like a real inhabitant of the virtual world, immersing the player further in the game. Such agents could never replace real humans, but it would be vastly more entertaining to sometimes chat up or haggle with an NPC shopkeeper that has a mind of its own, than always being forced to just give money and push the desired button as if the character were a soda machine.

It seems desirable to build more believable autonomous game characters, capable of social interactions with players. As game agents, these systems do not need to be sophisticated, but must be believable and efficient. This paper will argue for an approach to implementing such a believable system that avoids engaging computationally expensive communicative techniques.

Social interaction

As we consider the problem of designing believable, socially interactive characters, we must consider the layers of communication and meaning that constitute even the most common interactions. Their complexity is overwhelming and implementation may seem impossible, as the computer doesn't really understand our ways and our languages. Indeed, in the general case, this is an extraordinarily difficult, "AI-complete" problem.

We can make the problem simpler by allowing ourselves to limit the scope of interaction. Certainly, we cannot build human-level AI – but we can build specialized characters, whose social interactions in specific domains are much more believable. The domains examined in this work will be the mundane interactions that fill the days in computer worlds. These are common enough to be worth examining, and their routine structure will also simplify implementation.

For an example that might put the remainder of this discussion in context, imagine how a traveler might cope in a foreign country, full of people he doesn't understand. He would likely have a basic grasp of how the society works, but not know much language – maybe some of the most useful words, maybe even less.[1] And yet, even without knowledge of the local language, the traveler can nevertheless be quite successful in interacting with others. Even with meager language skills, he will be able to get by: obtain food and the essentials, get from place to place, perhaps even perform some basic jobs. Certainly, he could not participate in all that the society has to offer, but neither will he be completely lost. He will be able to get through interactions with other people, and survive in their society, even though he does not speak their language.

None of this should be surprising – as mature individuals, we can maneuver our way through typical social interactions even without knowing the language, so long as we know how to behave in each situation. We know what is involved in buying groceries, getting on a bus, and other common interactions. These stereotyped situations only require us to get a general grasp of the interaction, coordinate with others on just the few missing bits of information, and generally behave in the way one generally behaves in these situations. And even though each interaction will be different, we all have ways of monitoring and handling its flow, detecting potential problems, and repairing it when it breaks.

[1] The inverse is also interesting – when a traveler understands the language, but not the society. This leads to much more subtle problems that can be quite embarrassing for the participants, rather fun for the onlookers, but unfortunately completely outside the scope of this paper.

Equipped with the knowledge of stereotyped interactions, and knowledge of how to maintain and repair an ongoing interaction, the traveler is thus prepared to cope in a foreign society. When faced with a need to communicate, he will try his best to navigate the situation he is in, and come to an understanding with his conversational partner. Expectably, the conversation will never be smooth. The traveler will often make mistakes, find himself at a loss for words, or need to convey what he means using gestures and facial expressions. When he fails to understand what had been said to him, he might have to simply try and guess, from the emotional tone and body language, what the other may be trying to communicate given the interactions thus far. But he will try to understand, try to communicate, and above all, try to reach some end-goal of interaction.

These mundane stereotyped interactions are exactly the kind of a resource people use when faced with the need to participate in a society whose language they do not understand. We can also use them as a starting point in implementation of social participation in a computer – considering them as a base mechanism of social interaction.

Stereotyped social interaction

Let us examine in greater detail what we mean by *mundane, stereotyped social interactions*.

In the process of socialization, human beings appear to acquire a large repertoire of constructs for referring to routine social activities. They can be regarded as representations of typical actions performed by typical actors under typical circumstances (Heritage, 1984, pp. 57-58), and they let us get through everyday routines without paying attention to them. For example, our knowledge about how to make purchases in a store, take public transportation, or navigate through an intersection, are exactly those kinds of typified constructs – all participants have an idea of the information the interaction requires, a standard progression of steps to get through, and we know how to go about getting through them. Thanks to such stereotyped constructs, we do not need to analyze the situation to come up with the right set of actions to perform. Rather, we know very well the standard way of, say, going about making a purchase – give the items to the cashier, who will tell you the total price, give them the money, and so on. The routine works so well that it disappears from our attention – indeed, we do not notice it until it breaks.

Furthermore, the nature of these stereotyped routines is that they are 'anonymous', in that they do not make reference to the participants' individual feelings, goals, intentions, or even a more general organizational rationale for their actions. Indeed, if the interaction is simple and not collaborative or combative (that is, one is not deliberately trying to help or harm the other), there is no need to assess the individual characteristics of the participants. Thanks to this anonymity, a large number of very different individu-

als with very different goals and abilities can effectively interact with each other in a number of diverse routines.

The pervasiveness of such stereotyped interactions is fortunate and quite useful. This view of stereotyped interactions also suggests a possible finite-state implementation – but before we discuss this in greater detail, let us briefly examine the issues of participating in such an interaction.

Interaction as grounding for communication

Getting through a social interaction seems to require communication, and the common mode of communication is, expectably, conversation. This suggests that in order to engage in social interactions, we need to have at least rudimentary conversational skills, including some language understanding.

Approaching language processing from a general linguistic perspective, however, would immediately lead us to extraordinarily difficult problems. Conversational language is full of contingencies that complicate traditional, decontextualized approaches to understanding. Syntax can certainly be contingent on the context – utterances in a conversation are notoriously ungrammatical, often missing important elements, repeating them, mangling their order, and so on. Similar contingencies are present on the semantic level – utterances are full of ambiguities that require the history of interaction to resolve, or expressions such as indexicals that require a grounding in the situation in order to have any discernible meaning at all. Finally, on the pragmatic level, expressions rarely mean simply what they mean – to really grasp the import of an expression we need to understand its place in the interaction at hand. And yet, bad grammar and vague meanings usually pass unnoticed by participants. Only when we read transcripts of casual conversations do we begin to realize how hopelessly messy spoken language really is.

These problems are characteristic of everyday speech, and very difficult to approach computationally. We will inevitably run into them, should we attempt understanding of informal communication based only on the *text* of each utterance. As Garfinkel eloquently demonstrates in many of his essays (1967), human speakers tend to communicate very little explicitly, instead relying on the mutual understanding of the evolving situation to provide the import of their words. In a very concrete sense, the situation grounds much of informal communication. This is also the kind of communication that accomplishes mundane interactions.

We expect that these interactions, because they are so predictable, will provide a good grounding for communicative understanding. The situation at hand can be used to set up communicative expectations – we know more or less what to expect at each point in a stereotyped interaction – and these expectations will help us interpret the other person's utterances. For example, when walking into a bakery, we place ourselves in the context of buying what is sold there, and our attempts to communicate will be interpreted in that light by others. Thus we do not need to order using full declarative sentences such as "I would like a poppy seed bagel, please." By the virtue of the situation, all one needs to say is "poppy seed bagel, please," or perhaps just "poppy bagel," or even just grunt while pointing at the desired item. Issues of politeness aside, this would have communicated enough for the interaction to proceed to the next stage. This reliance on context is especially useful if we do not fully understand the language, in which case we will succeed by force-fitting what we do understand into what we expect to hear, and respond according to how well we think we understand what is going on.

The point here is that the participants only need to understand each other well enough to get through the interaction. This will be made simpler by the situation, as there is always an interpretation bias at play: once a typical interaction is engaged, participants are strongly biased to interpret the communicative attempts of the other as relevant to advancing the interaction. Stereotyped situations will provide better grounding than other contexts, because of their predictability.

Implementation outline

As we have seen, this project approaches the problem of participation in social interactions by restricting it in two ways:

1. Focus on mundane, stereotyped social interactions.
2. Perform understanding of language-in-context only.

This limited form of social interaction – using context-sensitive communication in well-defined contexts – is vastly simpler to engage in than general social interaction. Even humans make use of its simplicity, when they lack sophisticated linguistic abilities. It does not mean, however, that these stereotyped interaction are all that there is to interacting socially with others. Quite to the contrary, they are only the beginning; we will treat them as a base competence level, on which we could build more sophisticated characters.

The working hypothesis is that we could actually implement participation in these stereotyped interactions by explicitly modeling their progression. We can think of a single stereotyped situation as a set of state spaces, representing the anonymous roles that participants take on, the standard progression of events, and the information that needs to be exchanged in order for the interaction to progress. The complete system will include state models of many such interactions, often stacking them in layers of competence; layers of models from very general to highly specialized. They all represent some partial understanding of what goes on at the moment, all try to process what they observe unfolding around them, and all recommend what to do to further advance them.

Explicating the communicative expectations hiding in each interaction will in turn help us communicate with the player. Our movement through the interaction state space will tell us what information we expect to be presented at each point to advance the interaction, and the state of the interaction will in turn tell us what we need to communicate to our interlocutor. In this manner, communication only serves to drive the interaction through expectable states into a desirable end, and is never processed outside of the context of some interaction.

This model means that we could treat conversation as navigation through a complex but explicitly modeled state space. Our position in this state space will bias the understanding of incoming utterances, and this understanding will in turn transition us between states. Positioning ourselves and navigating through the ongoing interaction can then be accomplished using probabilistic state estimation techniques, including mechanisms for detecting and repairing erroneous estimation. Language processing can also be initially limited to very simple parsing and state-sensitive understanding, although it would benefit from later additions of a more complex linguistic apparatus. The initial implementation can therefore be quite efficient, and it should be rather inexpensive compared to more general conversational models.

Work on the system has begun only recently – thus the details are still rather vague. The benefits and difficulties of the approach discussed above will have to remain unaddressed in this paper – at the time of writing the system is not yet completely implemented, and it would be a disservice to the reader to present a premature discussion at such an early stage of development. I hope, however, to be able to present further details and, with some luck, some preliminary results by the time of the symposium.

Related work

One standard approach towards implementing NPCs was already mentioned in passing – that of completely scripting the dialogue, turning the conversation into a choice tree. That is, at every point in the conversation the player is presented with a small set of possible pre-written responses, and they navigate it by choosing the answer they like best, which leads to the next bit of conversation, next choice point, and so on. This well-known mechanism is popular in adventure and role-playing games, but its limitations are equally well known. Most significantly, it can be difficult to script them in such a way as to provide the player with maximum conversational choices while retaining simplicity of the overall choice tree. Furthermore, even in the most extensive choice tree the artificiality of the interaction is painfully transparent. Another approach, used not so much in games as in interactive avatars, is simple text pattern-matching descended from Eliza. Online avatars engage these techniques for purposes ranging from novelty and marketing (such as Alicebot[2] used on the web site for the movie *A.I.*), to simple amusement (such as common chatterbots). All Elizas, however, share the same classic weakness of lacking situation sensitivity and state retention, which makes them incapable of participating in any sort of an extended interaction. Neither of these approaches is capable of supporting a conversation.

In terms of language-based interaction, several projects at Rochester and elsewhere attempt sophisticated, goal-directed cooperative conversation with a human user. For example, in the TRAINS system (Allen et al., 1994), the human and the computer work together to find good train schedules for a railroad system; in TRIPS (Allen et al., 2001), they cooperate to set up complicated and sensitive transportation schedules, such as for evacuation of an imperiled island. The insight of these systems was that, even though the system was not able to understand general language, the developing situation tended to provide a lot of constraint to help interpret the human's speech, and further inference would allow the machine to guess their intentions and plans, providing even more interpretative help. The systems also attempted to extract linguistic speech acts from the ongoing interaction, as they should help to model the human participant as well as clarify what is going on in the conversation. These valuable insights find reflection in this project.

In terms of agents that engage in believable interactions, important work was pioneered by the OZ group projects (for example, Bates et al., 1991, and Reilly, 1996). Their systems were some of the first to present autonomous, naturalistic performance of scripted interactions, as well as techniques for modeling emotions and communicative displays. Related work was also done on multimodal conversational agents that communicate through more than just text. Projects such as REA and SAM (Cassell 2001) examine in great detail non-verbal elements of human conversation, such as the use of posture, gesture, facial expression, or emotive display. This project shares the conviction of those works – that context-sensitive non-verbal communication is as important for getting through an interaction as the words actually uttered by the participants. However, because we concentrate on the game domain, our examination of non-verbal communication will have to remain limited to what games can actually support.

Finally, in the explicit modeling of interaction, the project is also heavily influenced by situation scripts of Schank and Abelson (1977). Unlike scripts, however, this project concentrates on prescriptive descriptions of situations, to be used for actual engagement in the interaction.

[2] See www.alicebot.org for details

Conclusions

In order to build believable social agents, we seek to find a level of base social competence, which we hope is implementable and would provide a good foundation for more complex mechanisms. So long as we allow ourselves to work within the limitations of this limited model – understanding of communication-in-context only, for well-defined, stereotyped contexts – implementing such a grounding should be feasible using current techniques. This is the kind of a basic agent we hope to model: a communicatively eager non-player character, capable of participating believably in a number of interactions with human players, making up for the lack of language skills with knowledge of how to engage in basic social interactions. On top of this foundation we hope to be able to build agents with better language understanding and more knowledge about the world.

The general approach is this. Robust models of social interaction form the basis for understanding of communicative attempts. These models present stereotyped interactions common in online worlds – including trade, coordinating resources, building rapport, and so on – and are accomplished using state estimation mechanisms and corrective techniques that warn about and try to correct misunderstandings. Communication is treated not as a separate cognitive system, but rather only as a means of propelling an ongoing interaction – in a sense, communicative attempts transition the participants between different stages of the interaction. This means that the communication need not be grammatical, or even verbal, so long as it conveys the expected bit of information given the particular stage of the particular interaction.

The hypothesis is that the models of social interaction and the communication necessary to advance it should be reasonable to express computationally, and implementable using fast and efficient mechanisms. A proof-of-concept implementation is currently being built.

References

Allen, J. F. Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N. G., Miller, B. W., Poesio, M., Traum, D. R. 1994. The TRAINS Project: A case study in building a conversational planning agent. TRAINS Technical Note 94-3. Computer Science Department, University of Rochester.

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A. 2001. Towards Conversational Human-Computer Interaction. *AI Magazine.* Vol. 22 No. 4. Winter 2001.

Bates, J., Loyall, B., and Reilly, W. S. 1991. Broad Agents. In Proceedings of the AAAI Spring Symposium on Integrated Intelligent Architectures, Stanford University, March 1991. *SIGART Bulletin.* Vol. 2. No. 4. August 1992.

Cassell, J. 2001. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine.* Vol. 22 No. 4. Winter 2001.

Garfinkel, H. 1967. *Studies in Ethnomethodology.* Malden, MA: Blackwell Publishers.

Heritage, J. 1984. *Garfinkel and Ethnomethodology.* Cambridge, MA: Blackwell Publishers.

Reilly, W. S. 1996. Believable Social and Emotional Agents. Ph.D. diss., School of Computer Science, Carnegie Mellon University.

Schank, R. C., and Abelson, R. P. 1977. *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.