Affective model with strategic intentions based on Theory of Mind

Hazaël Jones¹ and Nicolas Sabouret²

¹ LIP6 - UPMC, Paris, FRANCE, hazael.jones@lip6.fr
² LIMSI, Orsay, FRANCE, nicolas.sabouret@limsi.fr

Abstract. This paper presents a computational model for reasoning about affects of the interlocutor, using a Theory of Mind (ToM) paradigm: the system manipulates representation of beliefs about the interlocutor’s affects, preferences and goals. Our affective model is designed for the context of job interview simulation, but it does not depend on a specific set of affects. It relies on simple rules for selecting topics depending on the virtual agent’s personality. We have implemented it using an OCC-based [1] representation of emotions and a PAD (Pleasure, Arousal, Dominance) model for moods.

Keywords: Theory of Mind, Strategic intentions, Affective model, Job interview.

1 Introduction

In order to build a credible interaction between a human and a virtual character, affective computing [2] proposes to simulate human affects in virtual agents, making them more realistic and engaging for interactions. In this context, one main challenge for Artificial Intelligence researchers is to make the virtual character adapt its behaviour to the perceived user’s affective state, which will lead to a more natural and credible interaction for the user.

To this purpose, we claim that virtual characters must not only use reactive behaviour in answer to a wide range of affects (emotions, moods, social attitudes...) such as in [3, 4, 5, 6]. It must also use strategic intentions about the human it interacts with. Strategic intentions can be seen as long term goals [7] for an agent. Indeed, in an interaction, people have intentions about the goal of a conversation, such as obtaining certain information, coming to an agreement, changing the interlocutor’s point of view or having a fun and relaxing conversation. This paper proposes to analyse these strategic intentions and to use them in the reasoning model of an affective agent. To this purpose, we define a general model that can be adapted to different context. In our work, we apply this general model to a specific case of a formal interaction: job interviews in which the goal of the recruiter is to obtain concrete information about the interlocutor’s social and technical skills, so as to select the best candidate.

This work is part of the TARDIS project³ and is integrated in the SEMAINE platform [5]. It proposes new modules for a conversational agent: an affective core

³ TARDIS stands for Training young Adult’s Regulation of emotions and Development of social Interaction Skills. url: http://www.TARDIS-project.eu/
module and a decision module. In this work, we focus on the decision module which use a ToM approach.

Our general model is based on logical rules and follows a theory of mind [8] (or ToM) paradigm. Based on affects perception from social signal interpretation, our virtual agent’s model derives beliefs and intentions of the interlocutor. These informations are confronted to its own goals so as to select the next course of actions in the interaction (in our case, to conduct the job interview).

This paper is organized as follows. Section 2 makes a state of the art on ToM and shows how this can be used in the context of a virtual agent’s reasoner. Section 3 focuses on the job interview context and its specific features. In section 4, we present in details our affective model that integrates the ToM for reasoning about affects in the context of interactions. The rules of this general model are illustrated on examples from the job interview context. The last section concludes on the model and its application to the job interview situation.

2 Related work on Theory of Mind

ToM [8] is the ability to attribute mental states (beliefs, intentions, desires, affects, . . . ) to others. The literature reports numerous ToM studies on the reasoning process of an agent about the reasoning process of another agent [9, 10]. In our work, we want to model the reasoning process of an agent that reasons about the reasoning process of an human (the applicant). This particular configuration raises additional difficulties and leads to a original model for our representation of the ToM. For example, in [11], an agent has beliefs about others in a subjective way. Agent A has belief about agent B following the real structure of agent B beliefs. However, in our work, since agent B is the human applicant, we do not have any information about its belief structure. We must guess them from the outputs of the affect recognition module.

Nevertheless, this model of influence and belief change [11] is based on work in psychology [12]: the authors use influence psychological factors in their simulation framework: consistency, self-interest, speaker’s self-interest and trust (or affinity). We believe that similar high-level reasoning structures must be proposed in reasoning models, to avoid low-level reasoning on perceptions such as what is done in [13, 14]. These papers focus on the perception aspects of ToM such as the desire of engagement, and are tailored for signal interpretation, not for the cognitive model of the virtual agent.

Several other applications have been studied with a ToM approach. For instance, [9] proposes a reasoner for task avoidance, the agent can change its behaviour in order to alter the other agent’s desires, intentions and in fine, actions to occur. This work has been extended in a more generic version [15] that proposes a two-level BDI agent model: the first level is the agent’s reasoner and the second one computes the ToM. Following a different approach, [10] also propose a model based on modal logics that extend the BDI paradigm. Each agent has a set of actions and a set of formulas that represent the agent’s mental state. A formula has a degree of desirability for the agent and a degree of plausibility. The use of modal logic allows researchers to model the recruiter beliefs, desires and intentions, but although it has already been tried in human-agent interaction [16], it seems difficult to represent a real humans’ mental states based only on perceptions.

This is the reason why we propose a model based on general rules that takes as inputs recognised affects from the interlocutor and strategic intentions for the virtual
agents, and combines them in the ToM-based affective model. The goal of our model is to represent the reasoning process of an agent that reasons about the reasoning process of a human. Our model will be applied and illustrated in the context of the TARDIS project that considers a job interview simulation as an interaction.

3 Job interview context

In job interviews, the recruiter has to reason about the actual and potential behaviour of the applicant in front of him [17]. In our simulation the recruiter is an agent and we want this agent to be able of such a reasoning about the applicant’s social skills. For this purpose, it is necessary to have the capability to predict in which situation the applicant will show certain behaviour.

Furthermore, in job interviews, recruiters want to ask questions in order to influence or to provoke particular reactions on the interviewee [17]. For example, to test the applicant’s capability to manage his or her stress, the recruiter can be voluntarily aggressive during the interview. Our goal in the TARDIS project is to model that kind of recruiter strategic intentions. To this purpose, we need a formalism to represent these strategies and to reason about the applicant’s state of mind, using a ToM approach that relies on perceptions from a social signal interpretation module [18].

Job interviews prove to be an interesting context for our study for three reasons. First, it consists in a one-to-one interaction with limited external perturbances; second, both participants strongly rely on ToM to adapt to the situation; third, job interview situations are stressful for the applicants and the self-control of affects is a key to make a good impression. In [19], a study shows that people who tried to suppress or hide negative emotions during a job interview are considered more competent by evaluators. Anxiety attitude is inherent to a job interview [20] and it has been shown that anxious individuals have less success in job interviews [21]. Similarly, Tiedens [22] shows that anger and sadness play an important role in the job interview. A person that express sadness is often considered incompetent and anger is not good for job access. Thus, emotion regulation is a key element to obtain a job.

According to Ekman, emotion expressions are regulated by situative norms according to social display rules [23]. One difficulty in our model is to allow programmers to define these social norms. In addition to that, we must also consider the recruiter intentions for the recruitment. In the selection process, recruiters want to collect data about the future applicant in order to make the most accurate choice. In [24], the main data that interest the recruiters are general work performance, individual qualities and specific competencies regarding the job (the match between the candidate and the job). In [25], recruitment objectives are analysed, they can differ between companies. The most important aspect about a recruitment are the retention rate and job performance of an employee. In order to fulfil these 2 financial objectives for a company, the job interview considers applicant attention, comprehension, credibility, interest in the job, accuracy of applicant’s expectations and self insight (knowledge, skills, abilities and needs). Based on this literature, we will describe in section 4.2, the topics of interest for questions during a job interview.

Reciprocally, recruiter behaviour is important during a job interview. Rynes [26] shows that it influences applicant attraction about the job. In [24], recruiters methods in order to sell their organisation to one applicant are stressed out. Recruiters that
want to sell their organization focus on these subjects: salary, co-workers, physical facilities, flexible hours, child care, advancement, benefits, job challenge/interest, organisation characteristics because they are the main intentions for an applicant when applying for a job. Furthermore, the ability for the applicant to adapt to the behaviour and intrinsic strategy of the recruiter is a key element in order to get a job [19]. For this reason, we propose several strategies for a recruiter (in section 4.2): provocative, pugnacious, friendly and helpful.

The following section presents our model of strategic intention for a general purpose one-to-one interacting virtual agent, based on a Theory of Mind about affects and the ongoing interaction. We show how this model can be used in the specific context of a job interview scenario.

4 A ToM-based model for a cognitive virtual agent

In this section, we will present our generic model, and then show its application in the TARDIS project. We propose to build a representation of the other’s attitude toward a set of topics. This is not properly speaking a ToM, but rather a model of the other. However, some dimensions such as “importance of the salary for the interlocutor” clearly refer to a representation of what the interlocutor thinks about a subject. Similarly, the fact that these dimensions are built based on the feeling expressed during the interaction makes this a representation of what the interlocutor feels about the current interaction/topics. This is the reason why we speak of ToM.

4.1 Generic model for theory of mind

Our main objective is to deduce real user beliefs from the real user/virtual agent interaction. In this interaction, inputs (affective states of the user) are given by non verbal signals deduced from social signal interpretation. It can be used on different simulations involving the interaction of a human with a virtual avatar: teaching, formation, training, . . . The common aspect of these simulations is the use of questions by the avatar. Our model considers agent’s question in order to manage the context of the answers of the person in interaction with the simulation.

Context management. With a view to manage the context, labels are given to the questions/sentences of the virtual character in order to interpret the answer/reaction of the human in term of beliefs on some topics. A list of topics can be done for each specific application. The set of topics $\text{set}_{\text{topic}}$ contains N topics: \{topic$$_1$$, topic$$_2$$, . . . , topic$$_N$$\}. Each subject is application dependent and based on the domain of the simulation. A question is concerned by 0 to n topics.

Beliefs build. In order to build beliefs about the human who interacts with the system, we consider the questions/sentences that were just expressed by the virtual agent (identified by labels about topics) and the quality of the answer of the human from an affective point of view (which is obtained by Social Signal Interpretation, or SSI). Based on that, the agent will update its beliefs about the human on a particular subject. We denote the beliefs of the agent about the human $B_{\text{Human}(\text{topic}_i)}$ for $i$ in \{1, . . . , N\}. 
According to the topic(s) raised by the question/remark of the agent, beliefs will be updated. In pursuance of building the beliefs of the human, we consider its answer (perceived via SSI) and decide if the answer is rather positive, negative or neutral. The SSI module gives us affects with an associated degree (between 0 and 1). In order to determine if the global answer is positive or not, an average of positive and negative perceptions is done. This average considers the degree of the detected affects because we don’t want a big degree of joy to be compensated by a small degree of distress. The degree of the detected affects is valued in the interval [0,1].

Let’s denote $W_{\text{Weight}}(Aff_+) = \text{Number}(Aff_+) \times \sum_i \text{Degree}(Aff_{i+})$ and $W_{\text{Weight}}(Aff_-) = \text{Number}(Aff_-) \times \sum_i \text{Degree}(Aff_{i-})$ for $i$ in the set of detected affects (positive for $Aff_+$ and negative for $Aff_-)$.

The following rule determines how the average answer is computed:

**if** $(W_{\text{Weight}}(Aff_+)) > W_{\text{Weight}}(Aff_-)$ **then**

\[
\text{AverageAns} \leftarrow \frac{W_{\text{Weight}}(Aff_+)}{\text{Number}(Aff)}
\]

**else**

\[
\text{AverageAns} \leftarrow -\frac{W_{\text{Weight}}(Aff_-)}{\text{Number}(Aff)}
\]

with $Aff$ all the affects expressed by the human. The average answer $\text{AverageAns}$ is between $-1$ and $1$ and an $\text{AverageAns}$ around $0$ means a neutral average answer.

Based on the human’s average answer and the topic tags of the question/remarks just done by the agent, the beliefs can be computed. Updates of each belief is given by algorithm 1. $\alpha$ is a weight between 0 and 1 that can be altered if we want the recruiter beliefs about the human to evolve quickly ($\alpha = 1$) or not ($\alpha$ near of 0). It can rely on the personality of the agent. An impulsive agent has an $\alpha$ near of 1 and a moderate one near of 0.

**Algorithm 1 Beliefs computation**

\[
\text{for } topic_i \in \text{set_topic} \ \text{do}
\]

\[
B_{\text{Human}}(topic_i) \leftarrow B_{\text{Human}}(topic_i) + \alpha \times \text{AverageAns}
\]

**Desires and goals.** The desires are used to define the strategic intentions of the agent. We organize our desires in two categories: the high-level ones and the more specific ones. The high level intentions are directly linked to social attitudes. Attitudes can be initialized with personality and can evolve during the simulation but with a dynamic slower than the emotional one which is quite reactive. For more detail about the computation of social attitudes, refer to [27]. The high-level intentions are about the general intentions of the agent for the interaction, the specific ones are about specific beliefs about the human that interest the agent during the interaction.

The high-level desires are denoted: $D(Attitude)$. The specific desires are denoted: $D(B_{\text{Human}}(topic_i))$ because specific desires in an interaction are about beliefs of the
human on a particular topic. For instance $D(B_{John}(football))$ is the desire of the agent to know if John has knowledge in the football topic.

**Dynamics of goals.** The high-level desires evolve in function of the social attitude of the agent. Social attitudes are based on Leary circumplex [28]. Depending on the application, some attitudes will be relevant and some will not. As shown by Leary, attitudes can be separated in two categories, the positive ones (friendly, cooperative, extroverted, . . .) and the negative ones (hostile, critical, . . .). Based on these two kind of attitudes, let’s define algorithm 2 in order to update the desires of the agent. This algorithm works as follows: if the agent has a negative attitude, he intends to select topics with a negative answer for the human. On the contrary, if the agent has a positive attitude, its desires are about topics with a positive answer from the human.

Algorithm 2 Desires computation

```
if (Attitude ∈ set(attitude_−)) then
    for B_{Human}(topic) ∈ set-topic do
        if (AverageAns < 0) then
            D(topic) ← D(topic) + α × |AverageAns|
        else
            D(topic) ← D(topic) − α × |AverageAns|
    if (Attitude ∈ set(attitude_+)) then
        for B_{Human}(topic) ∈ set-topic do
            if (AverageAns < 0) then
                D(topic) ← D(topic) − α × |AverageAns|
            else
                D(topic) ← D(topic) + α × |AverageAns|
```

**Goal selection.** Several strategies can be defined for the selection of one desire in the list of possible desires. The most natural one is to select the desire with the maximum value in the available desires. At one moment of the dialogue, every possibilities (topics) can not be approached in order to conserve the logical sequence of the conversation (a scenario for instance).

4.2 Model in the job-interview context

Figure 1 shows our global architecture. The TARDIS architecture considers four main components:

- The Social Signal Interpretation (SSI) component provides the affective model with information about the youngster’s affects and social attitudes that are detected by the system.
- The Interview Scenario component tells the virtual recruiter the expectation in terms of emotions and attitudes, depending on the interview progress. In TARDIS, the agent has no understanding of the youngster’s actual answers to the questions. It follows a scenario, that can be influenced by the recruiter perception and internal states and focus on the affective recognition and adaptation.
The Animation component is responsible for expressing the virtual recruiter’s affective state through its behaviour and expressions.

The virtual recruiter component which is composed of two modules:

- The Affective module. It provides a reactive model based on expectations from the recruiter and SSI of the youngster’s affective expression [27] and allows the computation of recruiter emotions, moods and social attitudes.

- The Decision module, that is the focus of this document. The goal of this module is to build a ToM for a cognitive agent in the context of a job interview. Our agent (the recruiter) will deduce intentions of the youngster considering its answers (based on SSI) in a particular context (the question that has just been asked by the recruiter). This model will also influence new questions, because if we know what question are difficult for the applicant, we should try to avoid them or not according to the agent personality (pugnacious or peaceful recruiter).

The decision module is based on a BDI architecture [29]. The BDI architecture is often used for its intuitive representation of the agent’s reasoning. The reasoning is organised by modules (beliefs, desires, intuitions) for a clear structuring.

**Context management.** In order to manage the context, some labels are given to the questions of the recruiter in order to interpret the answer of the youngster in term of beliefs on certain subject. Here is the set of topics set_topic that can be tagged for a job interview:

- **topic_youth**: questions about the youngster (general questions),
– topic\textsubscript{job}: questions about the job,
– topic\textsubscript{salary}: questions about the salary,
– topic\textsubscript{hours}: questions about the working hours,
– topic\textsubscript{skill}: questions about the competencies, the skills of the youngster regarding the job,
– topic\textsubscript{socialSkill}: questions about the general social skills of the youngster,

A question is concerned by 0 to \( n \) topics. For example, the question "In what position will you like to work in our enterprise?" can be tagged by two different topics: topic\textsubscript{skill} and topic\textsubscript{job} because it tells about the youngster skills (the position he thinks he can apply for in this job) and its knowledge of the job (organisation of the enterprise).

\textbf{Beliefs build.} Our goal is to build and update beliefs about the youngster. Algorithm 1 allows us to compute the following beliefs: importance of the salary for the youngster \( B_{Young}(salary) \), importance of scheduling and working hours for the youngster \( B_{Young}(hours) \), qualities of job skills \( B_{Young}(skills) \) and qualities of social skills \( B_{Young}(socialSkills) \),

– self-confidence \( B_{Young}(young) \),
– knowledge about the job \( B_{Young}(job) \),
– importance of the salary for the youngster \( B_{Young}(salary) \),
– importance of scheduling and working hours for the youngster \( B_{Young}(hours) \),
– qualities of job skills \( B_{Young}(skills) \),
– qualities of social skills \( B_{Young}(socialSkills) \),
– other beliefs can be added to this list.

According to the topic raised by the question of the recruiter, beliefs can mean different things. For instance \( B(young) \) is the belief of the youngster in himself. For a belief equal to 1, the youngster is very confident, and for \(-1\), he has an important lack of self-esteem. Here, the value quantifies the confidence of the youngster. If we look at \( B(salary) \), the signification is different, it is about the importance that the youngster put in the salary when applying for this job.

For instance, after a question about the job, considering algorithm 1, with \( \alpha = 1 \) (impulsive recruiter) and an actual belief value \( B_{Young}(job) = 0.5 \), \( B_{Young}(job) \) will become \(-0.3\) for an Average\textsubscript{Ans} of \(-0.8\) making the recruiter beliefs about the youngster change sign in one answer. A moderate recruiter (\( \alpha = 0.2 \)) will obtain a belief \( B_{Young}(job) = 0.34 \) which will change the dynamic of our simulation. An impulsive recruiter will cause strong dynamics and a moderate one smoother ones, which is the expected behaviour.

\textbf{Desires and goals.} For a job-interview simulation, the recruiter will have a limited set of high-level intentions (provocative, pugnacious, friendly and helpful) and only one of them will be triggered in the same time.

The specific intentions are about subjects that the recruiter want to favour during the interview. The level of each subject will be adapted in function of the high-level intentions and will also consider the beliefs about the youngster. Actually, these specific goals for the recruiter are about the knowledge of youngster’s beliefs on certain subjects. For instance, a question about the job will be associated to the goal \( G(B_{Young}(job)) \) because this question will give more information to the recruiter about the belief \( B_{Young}(job) \).
**Goal selection based on recruiter’s high level intentions.** The high level goals can be defined directly in the scenario or be computed on the personality of the recruiter. We define 4 main strategies (2 for the positive attitudes and 2 for the negative attitudes). Let’s remind that a recruiter with positive attitudes will have positive desires on topics where the youngster has positive average answers. On the contrary, a recruiter with negative attitude will have positive desires on topics where the youngster has negative average answers. Here are some strategies that we use for the virtual recruiter:

- **Provocative recruiter:** the recruiter will have a negative attitude and will always select the worst topic for the user (the one with the maximum Desire for the negative agent). \( \text{Intention} = \max(\text{Byoung}(\text{subject})) \).

- **Pugnacious recruiter:** the recruiter will have a negative attitude but will randomly select one of the worst topics but not always the same.

  \[ \text{Intention} = \text{random}(\max_n(\text{Byoung}(\text{subject}))) \]

  with \( \max_n \), the \( n \) worst subjects.

- **Friendly recruiter:** the recruiter will have a positive attitude and will randomly select one of the best topics for the user but not always the same. \( \text{Intention} = \text{random}(\max_n(\text{Byoung}(\text{subject}))) \).

- **Helpful recruiter:** the recruiter will have a positive attitude and will always select the best topic (the one with the maximum Desire for the positive agent). \( \text{Intention} = \max(\text{Byoung}(\text{subject})) \).

These different strategies lead to different goals. At one moment of the interaction, only some subjects can be approached according to the possibilities of the scenario.

### 5 Conclusion

In this article, we propose a ToM model for an affective virtual agent. The ToM is about a real person in interaction with the system. It is centred on the interpretation of the affective states perceived through Social Signal Interpretation. By building beliefs about the person in interaction with the simulation, we allow an interaction by understanding the subjects where the person is confident or not. Then, according to the virtual agent high level intentions, new questions will be selected in a coherent and credible way regarding the personality of the agent.

This work is actually integrated in the TARDIS platform. It has been used with 18 youngsters and results will allow us to validate the global behaviour of our ToM model in the context of job interview. With this evaluation, we want to confirm that our model proposes a credible virtual recruiter for a job interview scenario. The ToM should provide coherent actions of the recruiter according to the reactions of the youngster and the personality of the recruiter. This aspect can be evaluated through literature and thanks to the youngsters that will interact with the system.

Our immediate next step is to offer different levels of sympathy or politeness in each possible step of the job interview scenario. Each utterance will be annotated and several possibilities will allow to modulate the verbal behaviour of the recruiter, depending on its attitude and an expected level of difficulty. One current limit of our model is that it requires manual annotation of the scenario. The definition of an automated annotation process, based on the utterance’s semantic and contextual information, would greatly increase the scalability of the model.
Acknowledgment

This research has received funding from the European Union Information Society and Media Seventh Framework Programme FP7-ICT-2011-7 under grant agreement 288578.

References

[27] Jones, H., Sabouret, N.: An affective model for a virtual recruiter in a job interview context. In: 4th International Conference on Games and Virtual Worlds for Serious Applications, Genoa, Italy, 29/10/12-31/10/12, VS-GAMES’12 (October 2012) in press