Multidimensional dialogue management for tutoring systems

Andrei Malchanau[1], Volha Petukhova[1], Harry Bunt[2] and Dietrich Klakow[1]

[1]Saarland University, Spoken Language Systems, Saarbrücken, Germany
[2]Tilburg Center for Cognition and Communication, Tilburg University, The Netherlands
[1]{andrei.malchanau,v.petukhova,dietrich.klakow}@lsv.uni-saarland.de;
[2]harry.bunt@uvt.nl

Abstract

In this paper we propose an approach to dialogue management for tutoring systems applications. We apply the information state update (ISU) machinery that operates on a multidimensional context model. This approach not only captures the behaviour of dialogue participants more adequately than other approaches but also enables the generation of flexible multimodal behaviour by the system, addressing various task-specific interactive goals and expectations simultaneously. The approach leads to a knowledge-rich representations of the participants' information states and highly flexible dialogue management strategies. Moreover, it offers possibilities for various future extensions, as we illustrate on examples of tutoring scenarios for debating skills.

1. Introduction

The increasing complexity of human-computer systems and interfaces results in an increasing demand for intelligent interaction that is natural to users and that exploits the full potential of spoken and multimodal communication. Much of the research in human-computer system design has been in the area of task-oriented systems especially for information-seeking dialogues concerning well-defined tasks in restricted domains (see e.g. Allen et al., 2000 for the main paradigms used for dialogue modelling for domains of varying complexity).

The research community is currently targeting more flexible, adaptable, open-domain multimodal dialogue systems driven by modelling natural human behaviour. Advances are also made in modelling and managing multi-party interactions, e.g. for meetings or multi-player games.

The genre of tutoring systems lies in between task-oriented and interactive narrative systems created for entertainment (or edutainment). Tutoring systems present narrative scenarios. The explicit goal of teaching generally leads to more straightforward presentation than in the case of entertainment. Several highly visible systems involving tutoring or computer-mediated instruction have been built for military applications, for example to train command behaviour in stressful situations or in different cultural contexts (cf. Riedl &Stern, 2006, Core et al., 2015). Examples involving children include e.g. projects to assess virtual reality augmentation of social skills training for autism (Moore et al, 2005) or for training children to cope with bullying (Paiva et al., 2004).

Most applications are based on well-defined tasks in restricted domains. In some cases, these restrictions are imposed deliberately by the researchers to be able to investigate a limited set of tutoring dialogue phenomena without having to deal with unrelated details. However, this reduces the practical realism of the system. In this paper we present an approach to dialogue management for tutoring systems that has the flexibility to deal with various types of rich multimodal information.

The paper is structured as follows. In Section 2 we discuss the application domain and the tutoring scenario, specifying trainees and tutor roles and tasks. Section 3 presents the framework within which we model interactions by specifying the multidimensional context model and the information state update process in tutoring sessions. Section 4 outlines the process of context-driven generation of the system's tutoring interventions. Section 5 presents the dialogue manager implementation and proposes an evaluation method to validate to what extent the system's tutoring interventions correspond to that of human teachers. Section 6 summarizes our conclusions and indicates perspectives for further research.

2. Tutoring scenario for debates

The targeted scenario is concerned with training debating skills. A debate is a formal interaction that has certain rules, traditions, and even rituals. Participants present their positions by arguments in favor (*Proponent*) or against (*Opponent*) a certain main statement. A trainee of a debating tutoring system may perform in either proponent or opponent role. The performance of a debater is often judged on three main criteria: (1) argument content; (2) argument organization and (3) argument delivery.[1]. Generally, the evaluation of argument content and its quality poses significant challenges requiring a substantial amount of research and development. For instance, to detect logically flawed and/or irrelevant arguments inference machinery and consistency checking need to be implemented, and to the best of our knowledge there is no system that is able to perform this task reliably.

As for organization of arguments, the planning and preparation of an **A**rgument supported by **R**eason and **E**vidence are involved.[2] Argumentative structures have been studied and modelled for argumentative texts and to a certain extent for two-party argumentative dialogues, see (Peldszus and Stede, 2013) for an overview. In order to identify arguments and relations between their constituents, discourse relations are often considered. Dis-

[1]See, for example, the rules in 'How to Debate': http://www.wikihow.com/Debate

[2]See http://www.slideshare.net/Cherye/advanced-debating-techniques

course relations help to identify to which other propositions a proposition serves as evidence and from which other propositions it receives support. For instance in (Teufel, 1999), sentences within one argument and texts as a whole are classified as having one of the discourse relations such as result, purpose, background, solution, and scope achieving an F-score of 0.46.

Good debaters are distinguished by concise clear arguments connected by explicitly signalled structure, e.g. by discourse markers and dialogue act announcements. For example, 'I will speak in favour of ... Because ... Since international research shows...'. For our task, we concentrate on detection of *justification* and *evidence* relations, and provide feedback to the trainee whether the way he structured his arguments is in accordance with the tutoring system's expectations.

Finally, it is important in debate not only what arguments are brought up and how they are structured but also how they are presented or delivered. In this respect, five aspects are considered: Audibility, Engagement, Conviction, Authority and Likability (AECAL). Good debaters should give a strong impression that they truly believe what they say. To express authority, confidence, respect and friendliness, the debater needs to use his body properly, control his voice, posture, emotions and maintain eye contact.

2.1. Tutoring system tasks

A tutoring dialogue system has two main tasks: (1) to track and understand the behaviour of dialogue participants, in particular those of the trainee; and (2) to perform certain tutoring interventions. As for the first task, to capture visual information various modern tracking devices can be used, e.g. Kinect[3] or Intel RealSense[4]; and for the speech modality speech recognizers can be deployed using open source toolkits, e.g. HTK[5] or KALDI[6]. Captured data needs to be processed, e.g. tokenized, parsed, etc., in order to enable the system's understanding of the multimodal actions performed by the participants. Given the set of debating skills discussed above and in order for the system to provide feedback on the trainee's performance, the following aspects are analysed: (1) *presentational* aspects such as voice quality, speaking rate and overall posture orientation; (2) *interactional* aspects such as turn, time, contact and own communication management; and (3) aspects related to *argument structure*.[7]

Given the system's understanding of the trainee's behaviour, the second task of the system is to perform tutoring interventions to inform the trainee of a mistake or to propose corrections (or to provide positive feedback). The performance on this task requires immediate real-time feedback, often called 'in-action' feedback (Schön, 1983) on the three kinds of aspects mentioned above: presentational, interactional, and argument structure.

3. Multidimensional context modelling

Dialogue behaviour, when understood by a dialogue participant, evokes certain changes in the participants' context model (or 'information state'). Since we deal with several different tutoring aspects, an articulate context model should contains all information considered relevant for interpreting such rich multimodal dialogue behaviour in order to enable the tutoring system to generate an adequate reaction. Since tutoring interventions are concerned with trainee performance on argument and overall interaction structuring, fluency of spoken contributions, turn and time management, and managing perceptual and physical presentational aspects. Thus, a rather complex but also flexible model is required to deal with such complex communicative scenario. A dialogue model has been proposed by Bunt (1999) and refined by Keizer et al. (2011) and by Petukhova (2011). Complexities of natural human dialogue are handled by analysing dialogue behaviour as having communicative functions in several dimensions. 10 Dimensions are distinguished (see Dynamic Interpretation Theory (DIT), Bunt, 2000), addressing information about the task or activity domain (*Task*), speaker's processing of the previous utterance(s) (*Auto-feedback*) or that of the addressee (*Allo-feedback*), the speaker editing his own contributions (*Own Communication Management*) or those of the addressee (*Partner Communication Management*), the speaker's need for time (*Time Management*), maintaining contact (*Contact Management*), allocation of speaker role (*Turn Management*), future structure of dialogue (*Dialogue Structuring*), and social constraints (*Social Obligations Management*). Activities in these various 'dimensions' are called *dialogue acts* and are formally interpreted as update operations on the participants' information states. Dialogue acts have two main components: a *semantic content*, which specifies what the act is about; and a *communicative function*, which specifies how an addressee updates his information state with the semantic content when he understands the corresponding aspect of the meaning of a dialogue utterance. Bunt (2014) provides a specification of the dialogue acts update semantics.

The proposed context model has five components: (1) **Linguistic Context (LC)** with information about (a) 'dialogue history'; (b) 'latest state'; and (c) 'dialogue future' or 'planned state'; (2) **Semantic Context (SemC)** containing information about the task and domain; (3) **Cognitive Context (CC)** representing information about the current and expected participants' processing states; (4) **Perceptual/Physical Context (PC)** having information about the perceptible aspects of the communication process and the task/domain; (5) **Social Context (SocC)** containing information about current speaker's and partner's social context.

Each of these five components contains the representation of three parts: (1) the speaker's beliefs about the task, about the processing of previous utterances, or about certain aspects of the interactive situation; (2) the addressee's beliefs of the same kind, according to the speaker; and (3) the beliefs of the same kind which the speaker assumes

[3] https://dev.windows.com/en-us/kinect

[4] https://software.intel.com/en-us/realsense/home

[5] http://htk.eng.cam.ac.uk/

[6] http://kaldi.sourceforge.net/

[7] Details on each processing step leading to multimodal behaviour understanding are out of scope of this paper. For this we refer to the following papers (van Rosmalen et al., 2015; Petukhova et al.,)

$$
\left[\begin{array}{ll}
LingContext: & \left[\begin{array}{ll}
speaker: & \left[\begin{array}{ll}
dialogue\_history: & \left\{\ \langle previous\_segments\ \boxed{\text{Fig2}}\ \rangle\ \right\} \\
latest\_segment: & \left[\begin{array}{l} FS:\ \boxed{\text{Fig2}} \\ state = opening|body|closing \end{array}\right] \\
dialogue\_future: plan: & \left[\begin{array}{l} candidates: \langle list\_DAs\rangle \\ order: \langle ordered\_list\_DAs\rangle \end{array}\right]
\end{array}\right] \\
\multicolumn{2}{l}{partner: \langle partner\_linguistic\_context\rangle\ (according\ to\ speaker)} \\
\multicolumn{2}{l}{shared: \langle shared\_linguistic\_context\rangle}
\end{array}\right] \\
SemContext: & \left[\begin{array}{l}
speaker\_task\_model: \langle beliefs\rangle \\
partner\_task\_model: \langle beliefs\rangle\ (according\ to\ speaker) \\
shared\_task\_model: \langle mutual\_beliefs\rangle
\end{array}\right] \\
CogContext: & \left[\begin{array}{ll}
speaker\_own\_proc\_state: & \left[\begin{array}{l} proc\_problem: yes|no \\ problem\_input: FS \\ time\_need: negligible|small|substantial \end{array}\right] \\
\multicolumn{2}{l}{partner\_proc\_state: \langle partner\_cognitive\_context\rangle(according\ to\ speaker)} \\
\multicolumn{2}{l}{shared: \langle shared\_cognitive\_context\rangle}
\end{array}\right] \\
PercContext: & \left[\begin{array}{ll}
speaker: & \left[\begin{array}{l} own\_presence: positive|negative \\ own\_readiness: positive|negative \end{array}\right] \\
\multicolumn{2}{l}{partner: \langle partner\_perceptual\_context\rangle(according\ to\ speaker)} \\
\multicolumn{2}{l}{shared: \langle shared\_perceptual\_context\rangle}
\end{array}\right] \\
SocContext: & \left[\begin{array}{ll}
speaker: & \left[\begin{array}{l} interactive\_pressure: none|greet|apology|thanking|\ldots \\ reactive\_pressure: \{\langle dialogue\_acts\rangle\} \end{array}\right] \\
\multicolumn{2}{l}{partner: \langle partner\_social\_context\rangle(according\ to\ speaker)} \\
\multicolumn{2}{l}{shared: \langle shared\_social\_context\rangle}
\end{array}\right]
\end{array}\right]
$$

Figure 1: Feature structure representation of the context model.

to be shared (or 'grounded') with the addressee. A context model for multi-party dialogues is more complex containing representation of speaker's beliefs about contexts of more than one addressee and possibly also of other participants, e.g. the audience in a debate. Figure 1 shows the proposed context model with its component structure.

Each of the parts of the model can be updated independently while other parts remain unaffected. For instance, Linguistic Context is updated when dealing with presentational aspects and some interactional aspects, such as turn management; in the Cognitive Context participant's processing states are modelled, as well as aspects related to time and own communication management (e.g. error in speech production). Semantic context contains representations of task-related actions, in our scenario participant's arguments and their structures, and system's tutoring goals and expectations on trainee's learning progress.

4. Context-driven generation of tutoring interventions

As specified in Section 2, the Dialogue Manager (DM) has to generate in-action feedback. We illustrate how the DM functions in the training of presentational skills, i.e. on auditive and visual performance. This mainly triggers updates in the cognitive and semantic contexts of the model. Additionally, the linguistic context is regularly updated with the system's recognition of the trainee's multimodal behaviour, specifically posture shifts and speaking volume. The system's understanding of the trainee's behaviour is illustrated in Figure 2.

When the system (S) recognizes that participant P1 occupies the speaker role (i.e. has a turn) and interprets his/her behaviour as P1 speaking too loud and/or performing an inappropriate body movement, it should react by either informing the addressee of his infelicitous use of voice and body, or propose how this behaviour can be corrected. At the same time the system does not want to take the turn over, but rather communicate its messages in a non-intrusive manner. Thus, system responses are generated visually using colors (red meaning something wrong

$$
\left[\begin{array}{ll}
FunctionalSegment(FS_1): & \left[\begin{array}{ll}
\multicolumn{2}{l}{start: \langle token_{15}|1260\rangle} \\
\multicolumn{2}{l}{end: \langle token_{24}|5640\rangle} \\
\multicolumn{2}{l}{verbatim: \langle smoking, should, not, be, banned, in, all, public, places\rangle} \\
\multicolumn{2}{l}{prosody: \langle intensity = 73dB\rangle} \\
\multicolumn{2}{l}{gesture: \langle ARMSCROSSED\rangle} \\
\multicolumn{2}{l}{sender: \langle p1\rangle} \\
dial\_acts(DA_2): \{\langle & \left[\begin{array}{ll}
\multicolumn{2}{l}{dimension(D): \langle task\rangle} \\
\multicolumn{2}{l}{comm\_function(CF): \langle inform\rangle} \\
\multicolumn{2}{l}{sem\_content(SC): \langle content\rangle} \\
\multicolumn{2}{l}{sender/speaker: \langle p1\rangle} \\
\multicolumn{2}{l}{addressee: \langle p2\rangle} \\
rhetorical\_relation: & \left[\begin{array}{l} antecedent: \{\langle DA_1\rangle\} \\ type: \langle elaborate\rangle \end{array}\right]
\end{array}\right] \rangle\}
\end{array}\right]
\end{array}\right]
$$

Figure 2: Example of feature structure representation of a functional segment.

happened, green - participant's performance is according to expectations, thus providing positive feedback) and pictures depicting correct body position, plus a verbal message, e.g. *'Reset your posture'*.

The context model is updated as shown in Table 1.[8] The system believes to have interpreted the contribution by participant P1 having certain semantic content (representation of verbal component, speakingVolume:HIGH and gesture:ARMSCROSSED, which P1 believes are correct) and communicative function (Inform). Using the knowledge available to the system, e.g. a database with prosodic properties and body postures that are inappropriate in certain types of debate situation, the system's task is to inform the addressee about presentational errors. Thus, the system has a choice to generate either an Inform act with content $\neg appropriate(volume(fs_1) = high)$ and $\neg appropriate(gesture(fs_1) = ARMSCROSSED)$, or a Correction act, as illustrated in Table 1. The system expects that its dialogue acts are successfully interpreted (s2, s3) and their contents adopted (s07, s08) by the participant P1, and when continuing the dialogue he will lower his volume and uncross his arms (adoption in u08 and u09 leading to dialogue acts da_4 and da_5 expressed in one multifunctional functional segment fs_4).

If, by contrast, the trainee does not recognize the system's dialogue act, misinterprets, ignores it, or is not able to perform corrected actions, this will lead to *cancellation* of the expected adoption effects. A belief or goal is further canceled if it does not apply any more. Cancellation of a goal will also occur when the goal has been achieved or has been understood to be unachievable. Weak beliefs can be *strengthened* later as supporting evidence becomes available (see Bunt et al., 2007).

5. Evaluation

The Dialogue Manager (DM) prototype has been evaluated together with the Fission module using simulated input. The DM prototype is designed as a set of processes (threads) that receive data, update the information state and generate output. Figure 3 presents the overall DM architecture. Firstly, receives data produced by the Fusion/Interpretation module. Next, an update of information state is performed based on the received input. What part of context model to update is decided by Process Manager.

[8]NOTE: For the sake of simplicity we do not spell out the updates on other debate participant's state - opponent, since he is not in the trainee role in this situation and whose behaviour interpretation is out of scope of this paper. Example serves to illustrate of the main underlying principle.

Table 1: Example of context update. (LC = Linguistic Context; CC = Cognitive Context; SC = Semantic context; prec = preconditions; du = dialogue utterance; da = dialogue act; fs = functional segment; D = dimension; CF = communicative function; exp.und = expected understanding; und = understanding; exp.ad = expected adoption; ad = adoption; Bel = believes; MBel = mutually believed; WBel = weakly believes)

Context	num	source	S's context	num	source	P_1's context
LC				u001	*prec*	$Bel(P_1, Next\_Speaker(P_1))$
LC	s1		$Bel(S, Current\_Speaker(P_1))$	u1		$Bel(P_1, Current\_Speaker(P_1))$
	$fs_1 : du1$	*latest*	$\langle verbatim\rangle$; $volume(fs_1) = high$; $gesture(fs_1) = armscrossed$	$fs_1 : du1$	*latest*	$\langle verbatim\rangle$; $volume(fs_1) = high$; $gesture(fs_1) = armscrossed$
	$fs_1 : da_1$	*D;CF* *sem_content*	Task; Inform $\langle p\rangle$ Speaker:P_1; Addressee: P_2	$fs_1 : da_1$	*D;CF* *sem_content*	Task; Inform $\langle p\rangle$ Speaker:P_1;Addressee: P_2
SC	s01a	exp.und:$fs_1 : da_1$	$Bel(S, MBel(\{S,P_1\}, WBel(P_1,$ $Bel(S, Bel(P_1, appropriate(volume(fs_1) = high)))$	u01a	exp.und:$fs_1 : da_1$	$Bel(P_1, MBel(\{S,P_1\}, WBel(P_1,$ $Bel(S, Bel(P_1, appropriate(volume(fs_1) = high)))$
	s01b	exp.und:$fs_1 : da_1$	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Bel(S,$ $Bel(P_1, appropriate(gesture(fs_1) = armscrossed)))$	u01b	exp.und:$fs_1 : da_1$	$Bel(P_1, MBel(\{S,P_1\}, WBel(P_1, Bel(S,$ $Bel(P_1, appropriate(gesture(fs_1) = armscrossed)))$
SC	s2a	*prec*	$Bel(S, \neg appropriate(volume(fs_1) = high))$; $Bel(S, appropriate(volume(fs_1) = medium))$			
	s3a		$Want(S, Bel(P_1, appropriate(volume(fs_1) = medium)))$			
	s2a	*prec*	$Bel(S, \neg appropriate(gesture(fs_1) = armscrossed))$; $Bel(S, appropriate(gesture(fs_1) = armsUncrossed))$			
	s3b		$Want(S, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed)))$			
LC	da_2	*plan:s03a* *sem_content*	Task; Correct $appropriate(volume(fs_1) = medium)$			
	da_3	*plan:s03b* *sem_content*	Task; Correct $appropriate(gesture(fs_1) = armsUncrossed)$			
LC	s04	*prec*	$Bel(S, Current\_Speaker(P_1))$; $Want(S, Next\_Speaker(P_1))$			
LC	$fs_2 : du2$	*latest*	$\langle VOLUME\_MEDIUM\rangle$			
	$fs_2 : da_2$	*D;CF*	Task; Correct			
	$fs_3 : du3$	*latest*	$\langle UNCROSS\_ARMS\rangle$			
	$fs_3 : da_3$	*D;CF*	Task;Correct			
CC	s2	exp.und:da_2	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Interpreted(P_1, du_2)))$	u2	exp.und:da_2	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Interpreted(P_1, du2)))$
	s3	exp.und:da_3	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Interpreted(P_1, du_3)))$	u3	exp.und:da_3	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Interpreted(P_1, du3)))$
SC	s05	exp.und:da_2	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Want(S, Bel(P_1,$ $appropriate(volume(fs_1) = medium))))))$	u02	exp.und:da_2	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Want(S, Bel(P_1,$ $appropriate(volume(fs_1) = medium))))))$
	s06	exp.und:da_3	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Want(S, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed))))))$	u03	exp.und:da_3	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Want(S, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed))))))$
	s07	exp.ad:da_2	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Bel(P_1,$ $appropriate(volume(fs_1) = medium)))))$	u04	exp.ad:da_2	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Bel(P_1,$ $appropriate(volume(fs_1) = medium)))))$
	s08	exp.ad:da_3	$Bel(S, MBel(\{S,P_1\}, WBel(P_1, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed)))))$	u05	exp.ad:da_3	$Bel(P_1, MBel(\{S,P_1\}), WBel(P_1, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed))))$
SC				u06	und:da_2	$Bel(P_1, Want(S, Bel(P_1,$ $appropriate(volume(fs_1) = medium))))$
				u07	und:da_3	$Bel(P_1, Want(S, Bel(P_1,$ $appropriate(gesture(fs_1) = armsUncrossed)))$
				u08	ad:da_2	$Bel(P_1, appropriate(volume(fs_1) = medium))$
				u09	ad:da_3	$Bel(P_1, appropriate(gesture(fs_1) = armsUncrossed))$
LC				da_4	*plan:u08* *sem_content*	Task; Inform $appropriate(volume(fs_1) = medium)$
				da_5	*plan:u09* *sem_content*	Task; Inform $appropriate(gesture(fs_1) = armsUncrossed)$
LC				u002	*prec*	$Bel(P_1, Next\_Speaker(P_1))$
LC				$fs_4 : du4$	*latest*	$\langle verbatim\rangle$; $volume(fs_4) = medium$; $gesture(fs_4) = armsUncrossed$

In parallel to receiving and updating, the output based on the analysis of the information state is generated. The DM keeps track of its own dialogue history in the Linguistic Context of the context model.

The multimodal behaviour of the Dialogue Manager can be illustrated by an example from the training of presentational skills, where feedback on a trainee's postures and prosody are generated. When certain postures or prosody types are detected for more than a pre-set time span (1000ms) inform or correction acts are sent to the Fission module. Subsequently, when the trainee corrected his posture for more than a pre-set time span (500ms) a positive feedback act is generated.

Generally, it is hard to evaluate the performance of a dialogue manager as a single module since it depends heavily on the quality of its inputs. Thus, performance of the DM is best evaluated as a part of evaluating the integrated tutoring system. One way to do this is to compare the system's tutoring interventions to those of human tutors. We conducted a series of WoZ experiments whose main goal was to study the effects of tutoring interventions. The output from this experiments has been used as simulated input for our tutoring system. Three human Wizard tutors provided feedback on the presentational, interactive and

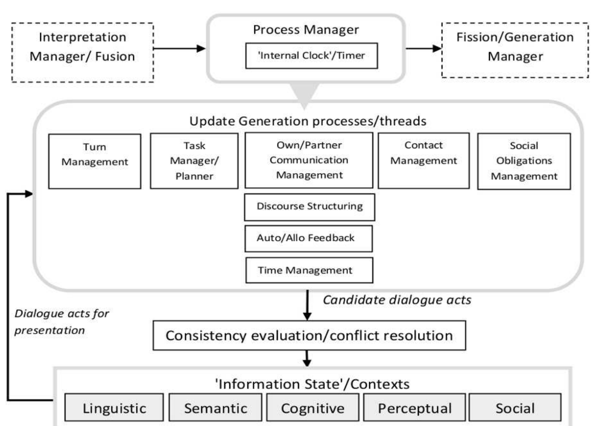

Figure 3: Overall Dialogue Manager architecture.

structural aspects of a trainee's performance in real time by pressing a red button for negative feedback, e.g. 'talks too loud', 'talks too much', 'rude interruption', 'no evidence provided', 'not clear arguments', etc., and a green button for positive feedback. Two debate sessions were evaluated with the total duration of 22 minutes consisting of 57 turns of 4 different speakers, and comprising 426 utterances. Time-stamped human Wizard's and automatically generated tutoring interventions events were logged and compared. As can be observed in Table 2, human and system interventions differ a lot both quantitatively and qualitatively. The system generated about 50% more feedback messages, with a significantly higher portion of negative feedback than human tutors do. This does not mean that the system actions were wrong, however. Upon close inspection, the majority of them make perfect sense. Errors were attributed mostly due to imperfect interpretation of spoken trainee's behaviour. Clearly, automatic natural language recognition and understanding are not ideal for many tasks. We found that important issues in dialogue management which are still largely open concern the amount, type and complexity of feedback which is appreciated most and considered useful. Thus, user-based evaluation and usability testing are very important and should be performed on a large scale involving both trainees and tutors. From the evaluation with trainees insights can be gained on what skills and what aspects are most important for them to master, and from the evaluation with tutors what type, amount and timing of interventions they provide lead to the best learning outcome. The dialogue management strategies can be specified in more detail by defining additional constrains and control strategies. The model underlying the DM architecture and multi-thread design enables such modifications and extensions.

Aspect	Human Tutor		System	
	positive	negative	positive	negative
Presentation	0	26	14	58
Interaction	40	18	27	46
Structure	8	3	2	1
Totally	103		148	
Completely matched	40			

Table 2: Tutoring interventions generated by human teachers and by evaluated tutoring system.

6. Conclusions and future work

In this paper we proposed the application of multidimensional dialogue management design for tutoring systems. Most state-of-the-art dialogue systems contain a dialogue manager, a module which takes care of deciding which action to take next in the dialogue, given some form of information state or context model that is monitored and updated during the dialogue. The dialogue manager is mostly designed to handle one particular dialogue task at a time and have only one set (typically very rigid) of possible dialogue state transitions for this task. As a result, the dialogue system is unable to handle complex real life communicative situations in complex domains and to exploit the full potential of spoken and multimodal interaction, for example in multi-party debates or tutoring sessions. Since dialogue models provide the basis for interpreting the speaker's behaviour and for decisions about future actions, a multidimensional approach to dialogue modelling opens the perspective for more adequate and rich human-system interaction. It supports more accurate understanding and multimodal and multi-tasking behaviour which is tuned to the situation. Thanks to its domain- and task-independent nature, the model offers possibilities for sophisticated refinements and structured extensions, but also for specific constraints, if required.

This paper also raised new issues and directions for future work. As a first step we plan to perform large-scale user-based evaluation with tutors and trainees as indicated above. A particular challenge for future work will be to define and incorporate control strategies and constraints necessary for the system's behaviour generation, especially in setting intervention goals.

7. References

Allen, J., D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, 2000. An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3-4):213–274.

Bunt, H., 1999. Dynamic interpretation and dialogue theory. In M. Taylor, F. Neel, and Bouwhuis D. (eds.), *The structure of multimodal dialogue II*. Amsterdam: John Benjamins.

Bunt, H., 2000. Dialogue pragmatics and context specification. In H. Bunt and W. Black (eds.), *Abduction, Belief and Context in Dialogue*. Amsterdam: John Benjamins, pages 81–105.

Bunt, H., 2014. A context-change semantics for dialogue acts. In Bos J. Bunt, H. and S. Pulman (eds.), *Computing Meaning*, volume 4. Dordrecht: Springer.

Bunt, H., S. Keizer, and R. Morante, 2007. A computational model of grounding in dialogue. In *Proceedings of the Workshop in Discourse and Dialogue*. Antwerp, Belgium.

Keizer, S., H. Bunt, and V. Petukhova, 2011. Multidimensional dialogue management. In A. van den Bosch and G. Bouma (eds.), *IMIX book*. Springer.

Moore, D., Y. Cheng, P. McGrath, and N. Powell, 2005. Collaborative virtual environment technology for people with autism. *Journal of the Hammill Institute on Disabilities*, 20(4):231243.

Paiva, A., J. Dias, D. Sobral, R. Aylett, R. Sobreperez, S. Woods, C. Zoll, and L. Hall, 2004. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *AAMAS '04 Proceedings, IEEE Computer Society*. New York.

Peldszus, A. and M. Stede, 2013. From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1):1–31.

Petukhova, V., 2011. *Multidimensional Dialogue Modelling. PhD dissertation*. The Netherlands: Tilburg University.

Petukhova, V., A. Malchanau, and H. Bunt. Modelling argumentation in parliamentary debates. In *Proceedings of the 15th CMNA Workshop*.

Riedl, M. and A. Stern, 2006. Believable agents and intelligent story adaptation for interactive storytelling. In *Proceedings of the TIDSE'06 Conference*.

Schön, D. A., 1983. The reflective practitioner: How professionals think in action. In T. Smith (ed.), *Basic Books*. London: Temple Smith.

Teufel, S., 1999. *Argumentative Zoning: Information Extraction from Scientific Text, Ph.D. thesis*. University of Edinburgh.

van Rosmalen, P., D. Börner, J. Schneider, V. Petukhova, and J. van Helvert, 2015. Feedback design in multimodal dialogue systems. In *Proceedings of the CSEDU Conference*. Portugal.