Next Generation Learner Modeling by Theory of Mind Model Induction

Klaus P. Jantke\textsuperscript{1}, Bernd Schmidt\textsuperscript{2} and Rosalie Schnappauf\textsuperscript{3}

\textsuperscript{1}Fraunhofer Institute for Digital Media Technology, Erfurt, Germany
\textsuperscript{2}Fachhochschule Erfurt, Erfurt, Germany
\textsuperscript{3}University of Rostock, Rostock, Germany

Keywords: Technology-enhanced Learning, User Modeling, Learner Modeling, Player Modeling, Theory of Mind, Learner Model Induction, Inductive Inference, ToMMI Technology.

Abstract: Learning is a spectrum of involved processes requiring the learner’s engagement and building upon the learner’s prior knowledge and other prerequisites. Educators know how to adapt to their learners’ needs and desires. User modeling is a key technology to enable digital systems such as e-learning environments and serious games to adapt to their users’ peculiarities. There is a huge corpus of scientific research on user modeling, on implementation of user modeling and related system adaptivity, and on the impact on teaching and learning. The aim of the present contribution is to go even further. The concept of theories of mind is adopted and adapted from animal behavioral research. Theory of mind user models allow for the identification and representation of user/learner/player peculiarities beyond the limits of all other preceding approaches to user modeling. Theory of mind learner models allow for the representation of higher quality profiles describing, for instances, intention, misconceptions, or even fear. The acquisition of suchlike expressive profiles is an inductive learning process of the digital system. The inductive inference of learner profiles requires particular concepts and algorithms. An implementation serves as proof of concept.

1 MOTIVATION

This is a technological paper. Although the authors have some running implementation (Schmidt, 2014), application and evaluation are considered secondary. Emphasis is put on an introduction to the innovative technology. The authors’ intention is to coin the term \textit{theory of mind model induction} (nickname: ToMMI) and to discuss how to utilize this technology for the purpose of learner modeling.

The implementation does not only serve as proof of concept, one may go even further. One of the key results in section 6 demonstrates the system developers’ ability to \textit{proof mathematically} that the modeling algorithms succeed in practice.

The present paper is the authors’ first publication in the field and, thus, the first publication on this novel technology at all. Therefore, technology is in focus.

The developed technology relates in an intriguing way to other areas of research such as, prominently, \textit{theories of mind} (Carruthers and Smith, 1996) and \textit{inductive inference} (Jain et al., 1999).

The key idea is to adopt and adapt theory of mind concepts for user modeling. In doing so, user modeling becomes inductive learning. This is investigated in cases where users are learners and/or players.

Besides theory and technology development, the authors work in the area of technology enhanced-learning, in general, and on game-based learning, in particular ranging from earlier publications such as, e.g., (Jantke et al., 2003), (Jantke et al., 2004), and (Jantke and Knauf, 2005) to recent contributions like, e.g., (Knauf et al., 2010), (Jantke and Schulz, 2011), (Fujima and Jantke, 2012), (Arnold et al., 2013), (Krebs and Jantke, 2014), (Jantke and Hume, 2015).

When studying digital games, there arises a really enormous manifold of exciting questions. Playing a game, usually, is fun (Koster, 2005). Among the great excitements of game play, there is the anticipation of an adversary player’s intentions. Knowing or, at least, hypothesizing what a human player wants, may form the basis for advanced ideas of game play such as, e.g., setting a snare for the adversary. Potentially, this does apply to serious games as well (Egenfeldt-Nielsen, 2007). To go even further, in serious games, this knowledge may be used to assist the learner.

Theories of theories of mind (Carruthers and Smith, 1996) allow for underpinning those studies.

To warm up, so to speak, the authors briefly sketch their preliminary case study in which a prototypical \textit{theory of mind induction} has been implemented for the first time ever (Schmidt, 2014).
2 THE GORGE CASE STUDY

In this work, the theory of mind induction means player modeling.

This application is based on a digital game named GORGE (Jantke et al., 2010) (see also (Gaudl et al., 2009) and (Jantke, 2010)). What has been done for game playing is now ready for a transfer to computer supported education. Seen from this perspective, this contribution aims at the transfer of technologies from player modeling to learner modeling.

In GORGE there are different teams of robots (see figure 1) operated by different players who may be humans or computer programs. Originally, GORGE has been designed and implemented as a research tool for studies of the perception of Artificial Intelligence.

![Figure 1: Screenshot of an Earlier Version of GORGE](image)

GORGE is turn-based. A dice is rolled and players may select one of their robots to move it accordingly. When a robot reaches a cell on the game path where another robot is sitting, this one is jostled backwards to the next free cell. Due to this game mechanics, robots tend to form clusters on the path. This leads to opportunities for taking revenge.

Human players may have largely varying intentions about jostling others such as, e.g., some grandparent’s intention not to frustrate the own grandchild. Another aim may be to take immediate revenge whenever possible. Next, consider reciprocal altruism: “If X jostled me never before and if I have the choice to jostle X or to do something else, I will not jostle X.”

Intentions like this may be easily represented by logical formulas. A computer program can monitor human behavior when playing GORGE. Based on the observations, the computer program hypothesizes the player’s goals. The computer learns a theory of mind.

3 CONVENTIONAL MODELING

There is more than 30 years of work on user modeling and adaptation. Consequently, it is not easy to relate the authors’ present efforts.

For the sake of comparison, the authors of the present paper have analyzed (Houben et al., 2009), (De Bra et al., 2010), (Konstan et al., 2011), (Mas-thoff et al., 2012), (Carberry et al., 2013), (Dimitrova et al., 2014), (Ricci et al., 2015), and the papers therein, as well as some of the impressive introductory and survey papers such as (Brusilovsky et al., 1995), (Specht and Weber, 1997) (Brusilovsky, 2001), (Brusilovsky and Millán, 2007).

There is a principle of conventional user modeling. A certain finite number \( n \) of features are selected and human beings are characterized accordingly by points in an \( n \)-dimensional space (Jung, 1921).

The Myers-Briggs Type Indicator (MBTI, for short) (Briggs Myers and Briggs, 1980) relies explicitly on Carl G. Jung’s theory of psychological types. Expressed in formal terms, according to MBTI, profiles of humans are points in a 4-dimensional space.

The Felder-Silverman approach (Felder and Silverman, 1988) is very similar to the MBTI, but puts explicit emphasis on learning such that one may understand it as an attempt to compromise between the MBTI and the David Kolb Learning Style Inventory (LSI, for short).

The LSI (see (Kolb, 1984), (Kolb and Fry, 1975)), analogously, builds a 4-dimensional space to host learner profiles. In contrast to the before-mentioned approaches, the LSI is enriched by a cyclic learning process model that underpins the four dimensions.

The history of learner modeling relies on model spaces of varying dimension (from (Brusilovsky et al., 1995) to (Brusilovsky and Millán, 2007)).

Occasionally, approaches are coming up in which the authors attempt to abandon the conventional limitations and aim at something like learner preferences (Kassak et al., 2015), (Schewe et al., 2007), (Smith et al., 2015). But saying that “preferences are expressed on the level of items” (Kassak et al., 2015) means to fall back to conventional approaches.

However impressive and practically successful, contemporary user modeling has apparent limitations. For illustration, educators who strive hard to treat their students with empathy know about the critical impact of misconceptions and about the importance of conceptual change and want to know their learner’s peculiarities ((Carey, 1985; Carey, 2000), (Thagard, 2012), (Vosniadou, 2013b), and the chapters therein).
4 THEORIES OF MIND

The present short section deals with an inspiration for understanding other individuals’ motivations, goals, desires, preferences, fears, and the like.

The aim of the authors’ present work is to take the inspiration from the present section to proceed from conventional learner modeling (see section 3) to innovative learner modeling of a higher expressiveness (see section 6) and, therefore, of a higher utility.

The source of inspiration is theories of mind as surveyed in (Carruthers and Smith, 1996).

The gist of the concept is well illustrated in sources such as (Emery, 2004), (Emery et al., 2004), (Goldman, 2006), (Clayton et al., 2006), (Call and Tomasello, 2008), (Emery and Clayton, 2009). Although the treatment in recent publications such as (Call and Tomasello, 2008) and (Emery and Clayton, 2009) is slightly more sceptical than in earlier work ((Carruthers and Smith, 1996) and the chapters therein), the application of the theories of mind perspective to human beings is scientifically justified and practically useful (Mauer, 2012).

Loosely speaking, theories of mind refer to two different individuals, say A and B.

In some of the above-cited sources, both agents A and B are birds, e.g., animals of the food-caching species western scrub jay (Aphelocoma californica). In the authors’ work, A is a computer program and B is a human learner.

The scenario under consideration is as follows. While agent B is acting–chaching food, interacting with an e-learning system, playing a digital game, or whatsoever–the agent A is monitoring B’s behavior. A is pondering the observations made and tries to find explanation for B’s behavior in terms of B’s thoughts.

What A constructs is, in some sense, a model of B.

In cases where B is a human interacting with a digital system, the result is a user model or, more specifically, a learner model or a player model or both at once.

There are particular aspects in theories of mind investigations such as time travel (Suddendorf, 2007) that are relevant to game play.

When game-based learning moves into focus, those particular investigations open completely new opportunities of application.

Notice that the present approach goes beyond the reach of all the above-cited sources. The present work aims at computer programs that are able to monitor human-computer interaction and learn about humans by algorithmic induction of user profiles which have the particular form of a computer’s theory of mind about a human user/learner/player.

5 EDUCATIONAL RELEVANCE OF THEORIES OF MIND

Educators think about the thoughts of their students. This covers a wide spectrum of aspects and enables them to adapt to their students’ peculiarities, needs, and desires (Davis et al., 2000). When computers in educational settings shall become adaptive as well, there arises the need to equip them with digitally, i.e., formally represented knowledge accordingly.

By way of illustration, let us look at preconceptions and misconceptions (Vosniadou, 2013b).

There are wide-spread misconceptions which cause difficulties to learners. A typical one is the belief that motion is caused by a force (Hammer, 1996).

In chemistry, many misconceptions are due to the misinterpretation of molecular equations. This leads learners to the belief that water is just a large amount of H₂O molecules, a theory of mind, so to speak, that makes the electrical conductivity of water completely incomprehensible.

In biology, a quite prominent misconception is the confusion of osmosis and diffusion accompanied by a large number of different misbelieves such as water molecules cease movement at osmotic equilibrium.

Comprehensive publications such as the handbook (Vosniadou, 2013b) or, e.g., (Chi et al., 1994), (diSessa and Sherin, 1998), (MacBeth, 2000), and (Vosniadou, 2013a) illustrate the omnipresence of learners’ thoughts that are likely to hinder learning. There is abundant evidence for the need of knowing a learner’s preoccupation.

Misconceptions are wide-spread in science as demonstrated in the case of biology (Kayoko and Hatano, 2013), in chemistry (Barke et al., 2009), in physics (Brown and Hammer, 2013), and in so-called earth science (Phillips, 1991).

The situation is not different in the humanities; see, e.g., (Leinhardt and Ravi, 2013) for history and (Arabatzis and Kindi, 2013) and (Thagard, 2013) for the history of science.

For curiosity, the rise and fall of phlogiston—the concept and the theory—may be of interest (Wisniak, 2004).

The theories of mind are collections of thoughts ascribed to other individuals. If computers shall be able to construct theories of mind that are intended to characterize human beings, the hypothesized thoughts must be represented formally—inside the computer, constructed and written by the computer and readable by the computer. Theories of mind are sets of logical formulas.

This logical point of view is taken subsequently.
6 INNOVATIVE TECHNOLOGIES OF LEARNER MODELING

Recall a human player’s reciprocal altruism when playing GORGE (section 2): “If X jostled me never before and if I have the choice to jostle X or to do something else, I will not jostle X.”

If a learner is driven by an intention like this one, how can a computer program learn about the human’s thoughts to represent this particular learner’s highly individual intentions, desires, goals, ideas, fears, and so on in a profile as expressive as a theory of mind?

This section is intended to survey an answer that, hopefully, will inform the readers about the essentials.

6.1 Spaces of ToM Hypotheses (I)

The theories of mind that hypothetically characterize human learners are represented as finite sets of logical formulas.

Modal operators like □ and ◇ are appropriate to make complex expressions readable. The operators, as usual, express the two modalities of necessity and possibility, respectively (Blackburn et al., 2001).

The intention of reciprocal altruism, e.g., may be represented as follows. Let us assume, that predicates such as jostle are ternary. The first of the arguments contains time information. The other two arguments contain player names, where the distinguished constant * names the human player who is currently modeled. The symbols π, perhaps, with upper and lower decorations denote strings of events describing human-computer interactions. The binary relation ⊑ holds for any two strings, exactly if the first one is an initial segment of the second one. All remaining syntax is conventional first order predicate calculus (see, e.g., (Richter, 1978)).

The following formula is a statement about some human-computer interaction encoded by π and some recent player’s action µ. It formalizes a variant of the player’s aim at reciprocal altruism (see section 1).

( 3π′ : π’ ⊑ π ∧ jostle(π’, X, *) → ¬jostle(πµ, X) )

It may be difficult to realize “¬jostle(πµ, X)”, if the action µ is enforced. The following fits better.

( 3π′ : π’ ⊑ π ∧ jostle(π’, X, *) ∧ □¬jostle(π, X, *) → ¬jostle(πµ, X) )

Logical formulas like the two above form the space of hypotheses. A user profile is a finite set of formulas expressing player intentions. In other words, those formulas express “the computer’s thoughts” about the human user currently modeled.

The question is how to learn those formulas from observations.

6.2 Inductive Inference of Human Learner Profiles

… it is not really difficult to construct a series of inferences, each dependent upon its predecessor and each simple in itself. If, after doing so, one simply knocks out all the central inferences and presents one’s audience with the starting-point and the conclusion, one may produce a startling, though possibly a meretricious, effect.

Sherlock Holmes to Dr. Watson in ‘The Adventure of the Dancing Men’ by Arthur Conan Doyle, 1915

Profiles of human learners are inductively inferred from observations of human-computer interactions. Basic steps of inference are simple, but the overall result of learner modeling may appear meretriciously as Arthur Conan Doyle put it (Doyle, 1915).

Logic programming¹ is the authors’ software technology of choice ((Clocksin and Mellish, 1981), (Sterling and Shapiro, 1986)) to implement the steps of logical inference.

Observations are formally represented in the form πµ where π describes the history of interaction and µ denotes the human learner’s current action.

As long as the current user profile is sufficient to explain the human user’s behavior, a property named consistency, there is no need to change the profile.

More formally, assume the necessary background knowledge (domain knowledge, system behavior, . . . ) BK and a current learner profile φ. The consistency is expressed in formal terms² as BK∪{πµ} ⊧ φ.

Note that there is no method, in general, to prove consistency (Jain et al., 1999). Usually, this property is only co-enumerable (Rogers jr., 1967; Sipser, 1997). Therefore, the logical reasoning strategy is refutation. Fortunately, Prolog is refutation-complete.

Every hypothesis φ of the learner profile is kept until its refutation succeeds. Our system in forming learner profiles tries to validate BK∪{πµ} ⊧ φ. Only if this succeeds, the hypothesis is given up and replaced by a refinement. But how to do this . . . ?

¹Logic programming is an appropriate approach to the authors’ ambitious task of modeling learners by theories of mind. For comparison, look at IBM’s Watson which is powered by 10 racks of IBM Power 750 servers running Linux, and uses 15 terabytes of RAM, 2,880 processor cores and is capable of operating at 80 teralops. Watson was written in mostly Java but also significant chunks of code are in C++.

²A core part to perform reasoning is implemented in Prolog.

© 2016 CSEDU - 8th International Conference on Computer Supported Education
6.3 Spaces of ToM Hypotheses (II)

To learn learner profiles, one needs to agree about “what to say about a human learner”. After such a decision has been made, one has a potentially infinite set of formulas which may describe characteristics of human learners, i.e., theories of mind.

As described before, the refutation-completeness of Prolog allows for a fully computerized refutation of logical expressed hypotheses. But how to refine refuted hypotheses within a theory of mind? How to step forward from one hypothetical learner model to the next one, if necessary?

To arrive at a completely algorithmic approach to learning of theories of mind forming learner models, we adopt and adapt a fundamental concept introduced by Dana Angluin (Angluin, 1980). Her ingenious concept is called indexed family of formal languages in which every particular language has a decidable word problem. In contrast to Dana Angluin’s approach, we can not assume decidability. This means that our approach is more expressive, but somehow less comfortable. We have to invest more algorithmic effort. These thoughts lead to the concept below that our approach is more expressive, but somehow less comfortable. We have to invest more algorithmic effort. These thoughts lead to the concept below that describes a learner’s intention, goal, motivation, or so, that our approach is more expressive, but somehow less comfortable. We have to invest more algorithmic effort. These thoughts lead to the concept below that describes a learner’s intention, goal, motivation, or so, that our approach is more expressive, but somehow less comfortable. We have to invest more algorithmic effort.

A so-called indexed family of logical formulas is a sequence of formulas \( \Psi = \{ \psi_n \}_{n=0, 1, 2, \ldots} \) as follows.

(i) \( \Psi = \{ \psi_n \}_{n=0, 1, 2, \ldots} \) is recursively enumerable.

(ii) For any \( \pi \in \Pi \) and for any index \( n \), \( BK \cup \{ \pi \} \neq \psi_n \) is recursively enumerable.

(iii) For any two indices \( i \) and \( j \) with \( i < j \), \( \psi_j \) does not logically imply \( \psi_i \).

Given an indexed family of logical formulas \( \Psi \), one can easily implement a computer program able to learn whatever formula characterizes a human learner’s intentions.

According to fundamental results of inductive learning based on usually incomplete information (Jain et al., 1999), there is a computer program able to learn whatever formula in \( \Psi = \{ \psi_n \}_{n=0, 1, 2, \ldots} \) might describe a human learner’s peculiarities. The inference principle is called identification by enumeration. For every set of observed learner behavior, it returns the first formula which is not (or not yet) refuted.

Note that there may be any finite number of spaces of hypotheses in use, i.e., any collection of indexed families of logical formulas \( \Psi^1, \Psi^2, \Psi^3, \ldots, \Psi^{(k)} \). Similar inference procedures may run on the different enumerations in parallel returning one statement from each of the (sub-)spaces of hypotheses.

6.4 Features of Learner Modeling by Theory of Mind Induction

To sum up sections 6.1, 6.2, and 6.3, if we are able (i) to describe potential human learner peculiarities of interest by means of logical formulas and if we can (ii) enumerate these formulas appropriately, there is a universal computational method that is able to learn whatever human peculiarity might be on hand.

Recall that a formula \( \psi \) occurs as a component of a profile of some human learner, if it is listed within an indexed family of logical formulas, which is in use for theory of mind induction, where this formula has some index \( n \), i.e., \( \psi = \psi_n \). Furthermore, for every preceding index \( i \), the formula \( \psi_i \) has been refuted. Finally, there is no refutation of \( \psi_n \), at least, not yet.

The key insight is that a component such as \( \psi_n \) within a learner profile is unavoidably hypothetical. This has to be stressed, though it is not a big surprise. When humans think about other human beings and about their thoughts, they usually are aware of uncertainty. This applies to computer “thoughts” as well.

Consequently, it is highly advisable to deal with hypothetical learner profiles carefully. A particularly interesting option is to refine adaptivity by reflection (Jantke et al., 2013). This approach has deep roots in research about reflective inductive inference (see Grieser, 2008) which expands upon (Jantke, 1994), (Grieser and Jantke, 1995), and (Jantke, 1995)).

The crux is that there is no way to definitely say whether or not a hypothesis is consistent with given observations. The only way of automated reasoning is to try refutation. If a hypothesis is not consistent, Prolog will find this out after some time.

A closer look at the methodology described in the previous subsections reveals a further peculiarity. First, naturally, if a formula has been found that truly describes a learner’s intention, goal, motivation, or so, this is consistent and, thus, will never be abandoned. But, second, what might be the reason not to arrive at such a correct hypothesis? The only reason may be another formula, say \( \psi_k \), which occurs earlier in the enumeration, but which has not yet been refuted.

In such a case, this earlier hypothesis the learning algorithm gets stuck with is equivalent—within the limited information available—to the “true” one.

---

3The word problem is the question whether or not any word belongs to a language (Hopcroft and Ullman, 1979).

4By means of logic programming and due to the known refutation completeness of Prolog, this means that taking (a) the basic knowledge, (b) the observation of the learner’s behavior, and (c) the candidate formula \( \psi \), there has been derived the empty clause. This means the discovery of a contradiction and, thus, it establishes a refutation of \( \psi \).

5See (Bläsius und Bürckert, 1978) and (Richter, 1978) for background and (Popper, 1934) for the bigger picture.
7 SUMMARY & CONCLUSIONS

To the authors’ very best knowledge, the present approach of modeling users, especially learners, by theories of mind—more precisely by the induction of user profiles that form theories of mind—is a novelty. By theories of mind one may express peculiarities of human users far beyond the limits of conventional user modeling (see the rich corpus of work (Houben et al., 2009), (De Bra et al., 2010), (Konstan et al., 2011), (Masthoff et al., 2012), (Carberry et al., 2013), (Dimitrova et al., 2014), (Ricci et al., 2015), and the papers therein). This is of a particular relevance to technology-enhanced learning. Teachers and digital systems to assist them for purposes of teaching and learning need to “know” about their human learners.

There may be individual goals, misconceptions, and even fears of human learners. Theories of mind are appropriate to express such individual conditions and the so-called theory of mind model induction is appropriate for the computerized learning of human peculiarities. This is a bio-inspired technology that has a firm foundation in behavioral research.

But do we really need so much technology on a CSEDU conference?

Yes, we do. First of all, educators need to be informed about the next opportunities available to them. Second and even more importantly, they are needed to carve out the future of education. Only educators can name the human learner conditions that are not yet covered sufficiently well by conventional learner modeling. Only educators can help the technologists to determine the formalism needed to express learner peculiarities of relevance.

Seen from this perspective, the present paper is intended to be understood as a call for co-operation. Let us go together for the design, implementation, application, and evaluation of the next generation of e-learning systems that are able to “understand” their human learners by the induction of learner needs from observed behavior.

The observation of learners leads to hypothetical learner profiles (see section 6 above). Based on these learner profiles, personalization takes place. But how to treat learners appropriately? This is not a question for technologies. Think of a trainee who strives hard to avoid repeated conversation with the same staff member. How to react appropriately?

It needs psychologists, educational psychologists, educators, and domain experts to determine suitable adaptive behavior in response to hypothesized learner profiles. As a firm basis, the specialists need to understand learner profiles and the way they are created.

REFERENCES


