A hybrid Approach to model a Bayesian Network of Culture-specific Behavior

(Extended Abstract)

Birgit Endrass, Julian Frommel, Elisabeth André
Human Centered Multimedia, Augsburg University
endrass, andre@hcm-lab.de

ABSTRACT

Cultural background can crucially influence human behavior. For virtual characters, that are supposed to imitate human behavior in a natural manner, integrating cultural background came into focus the recent years. Unlike earlier approaches, this paper presents a hybrid approach of modeling a Bayesian network that generates culture-specific behaviors for virtual characters. For the hybrid approach, the structure of the network is designed based on social science theories, while its parameters are derived from a video corpus.

Categories and Subject Descriptors

I.6.7 [**Simulation and Modeling**]: Model Development

General Terms

Human Factors, Design

Keywords

Bayesian Networks, Culture, Virtual Agents, Nonverbal Behavior

1. MOTIVATION

How non-verbal behaviors are conducted and perceived is, amongst others, dependent on cultural background [14]. Building models that determine culture-related differences in behavior is a challenging task as the causal relation of culture and corresponding behavior needs to be simulated in a convincing and consistent manner.

The majority of approaches is theory-driven, with theories from the social sciences being taken as a basis for computational models. A common way of implementing the theory-driven approach is to start from existing multiagent architectures and extend them to allow for culture-specific adaption of goals, beliefs and plans, e.g. [10], [1].

Data-driven approaches, use data of existing cultures as a basis for computational models. Such a corpus of multimodal behavior was recorded for the Cube-G project in the German and Japanese cultures [12]. In our earlier work, a statistical analysis was performed highlighting differences between the recorded cultures and perception studies were performed with virtual characters [6].

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

2. APPROACH

For the hybrid approach described in this paper, we design the structure of our model based on theoretical knowledge from the social sciences, while the parameters are learned from empirical data. The network is modeled using the GeNie and SMILE modeling environment[1].

Structure of the Network.

To model culture in our network, we use Hofstede's dimensional model [8] that provides mappings for national cultures to five cultural dimensions.

To model non-verbal behavior, nodes are added for gesture types, postures and gestural expressivity. Gesture types are classified according to McNeill's categorization [11]: deictic, beat, emblem, iconic, metaphoric, and adaptors. To distinguish arm postures, we employ Bull's posture categorization [2], holding 32 different arm positions such as *PHEw* (put hands on elbow) or *PHWr* (put hands on wrist). The nodes describing gesture types and postures are directly linked to the culture node in our network.

To model the connection from culture to non-verbal expressivity, we rely on the idea of synthetic cultures [9] who find themselves on one extreme end of each cultural dimension, and for whom prototypical behavioral traits are described. For example, masculine cultures are described as using animated gestures. Expressivity can be broken down into parameters explaining differences in the dynamic variation along different dimensions [7]. For our model, we employ the expressivity parameters spatial extent, power, speed, fluidity and repetition.

Regarding verbal behavior, two nodes are included to our network: one focusing on speech acts, and one on conversational topics, both directly connected to gesture and posture type. Speech acts can be categorized along the DAMSL (Dialog Act Markup in Several Layers) coding scheme [3]. We use the following subset of communicative functions: *statement*, *answer*, *info request*, *agreement / disagreement* (indicating the speaker's point of view), *understanding / misunderstanding* (without stating a point of view), *hold*, *laugh* and *other*. Conversational topics prototypically occurring in first-time meetings can be classified as follows: The *immediate situation*, e.g. the surrounding or atmosphere of the conversation, the *external situation*, describing the larger

[1]http://genie.sis.pitt.edu/

Figure 1: Virtual characters showing prototypical non-verbal behavior.

context of the immediate situation, and the *communication situation*, with topics focusing on the conversation partners [13].

Parameters of the Network.

The first-time meeting scenario of our corpus [12] was annotated regarding the theories and classifications from the social sciences described above. To prepare the data for parameter learning, each annotated conversation is divided into datasets determined by a speech utterance. Each dataset may or may not be accompanied by a gesture and/or posture. For each dataset, the social background (such as gender and culture) is specified using the metadata of the annotations.

The SMILE-Framework[2] provides, amongst others, an implementation of the EM-algorithm [5], that requires a decision network and a list of datasets. With it, the structure of our network that was modeled based on theoretical knowledge, is filled with parameters learned from the multimodal corpus.

3. APPLICATION

The Bayesian network is integrated into a virtual environment holding culture-specific characters [4]. First-time meeting dialogues, similar to the ones recorded in the corpus, were scripted and tagged with the same categorizations of speech acts and topics as in our network. Animations are labeled according to the gesture and posture types. Each animation can be customized to match different levels of expressivity. The probabilities for culture-dependent non-verbal behaviors are generated by the network based on verbal behavior and cultural background, and then simulated with the characters.

Figure 1 shows a screenshot of a male German character in a prototypical German posture (FA - fold arms) in conversation with a female Japanese character performing an iconic gesture with a small spatial extent.

4. CONCLUSION AND FUTURE WORK

The hybrid approach described in this paper combines the advantages of the commonly used theory-driven and data-driven approaches, as it explains the causal relations of cultural background and resulting behavior, and augments them by findings from empirical data.

In the resulting network, on the one hand, the most probable culture-related level of expressiveness in non-verbal behavior is calculated, and on the other hand the most likely culture-specific non-verbal behavioral type (gesture and posture) is determined based on the verbal behavior (speech act and topic) that it should accompany.

For our future work we aim on evaluating the network and have a deeper look into further aspects such as gender-specific differences. In addition, studies will be conducted that test how virtual characters, performing the behavior calculated by our model, are perceived by human observers.

[2]http://genie.sis.pitt.edu/

5. REFERENCES

[1] R. Aylett, A. Paiva, N. Vannini, S. Enz, E. André, and L. Hall. But that was in another country: agents and intercultural empathy. In *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems*, 2009.

[2] P. Bull. *Posture and Gesture*. Pergamon Press, Oxford, 1987.

[3] M. Core and J. Allen. Coding Dialogs with the DAMSL Annotation Scheme. In *Working Notes of AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, 1997.

[4] I. Damian, B. Endrass, P. Huber, N. Bee, and E. André. Individualized Agent Interactions. In *Proc. of 4th Int. Conf. on Motion in Games*, 2011.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[6] B. Endrass, E. André, M. Rehm, and Y. Nakano. Investigating culture-related aspects of behavior for virtual characters. *Autonomous Agents and Multi-Agent Systems*, 2013.

[7] P. E. Gallaher. Individual Differences in Nonverbal Behavior: Dimensions of Style. *Journal of Personality and Social Psychology*, 63(1):133–145, 1992.

[8] G. Hofstede, G.-J. Hofstede, and M. Minkov. *Cultures and Organisations. SOFTWARE OF THE MIND. Intercultural Cooperation and its Importance for Survival.* McGraw Hill, 2010.

[9] G. J. Hofstede, P. B. Pedersen, and G. Hofstede. *Exploring Culture - Exercises, Stories and Synthetic Cultures.* Intercultural Press, 2002.

[10] W.-J. Johnson, S. Marsella, and H. Vilhjálmsson. The DARWARS Tactical Language Training System. In *Interservice / Industry Training, Simulation, and Education Conference*, 2004.

[11] D. McNeill. *Hand and Mind - What Gestures Reveal about Thought.* University of Chicago Press, 1992.

[12] M. Rehm, E. André, Y. Nakano, T. Nishida, N. Bee, B. Endrass, H.-H. Huan, and M. Wissner. The CUBE-G approach - Coaching culture-specific nonverbal behavior by virtual agents. In *ISAGA 2007: Organizing and Learning through Gaming and Simulation*, 2007.

[13] K. P. Schneider. *Small Talk: Analysing Phatic Discourse.* Hitzeroth, Marburg, 1988.

[14] S. Ting-Toomey. *Communicating across cultures.* The Guilford Press, New York, 1999.