Real-Time Adaptation of a Robotic Joke Teller Based on Human Social Signals

Socially Interactive Agents Track

Klaus Weber
Human Centered Multimedia
Augsburg University
Augsburg, Germany
klaus.weber@informatik.uni-augsburg.de

Hannes Ritschel
Human Centered Multimedia
Augsburg University
Augsburg, Germany
hannes.ritschel@informatik.uni-augsburg.de

Florian Lingenfelser
Human Centered Multimedia
Augsburg University
Augsburg, Germany
florian.lingenfelser@informatik.uni-augsburg.de

Elisabeth André
Human Centered Multimedia
Augsburg University
Augsburg, Germany
andre@informatik.uni-augsburg.de

ABSTRACT

Humor is an essential element of human-human communication. Consequently, robots in the role of companions should exploit its potential as well to make interactions more enjoyable. Using a robot as an entertainer requires finding out what kind of humor its audience prefers. However, it is a challenging task for a social robot to learn what users prefer without bothering them by repeatedly asking questions. In this paper, we present an approach based on Reinforcement Learning that enables a robot to continuously adapt to the users' humor preferences without requiring them to explicitly provide feedback. Instead, we designed the robot to analyze the user's ideomotor social cues. We evaluated our approach in a scenario involving a Reeti robot acting as an entertainer. In this role, it is telling different types of jokes, (possibly) underlining its performance with grimaces and sounds. The adaptation process is accomplished only by using the audience's vocal laughs and visual smiles, but no other form of explicit feedback.

KEYWORDS

socially-aware-agents, social adaptation, human-robot-interaction

ACM Reference Format:
Klaus Weber, Hannes Ritschel, Florian Lingenfelser, and Elisabeth André. 2018. Real-Time Adaptation of a Robotic Joke Teller Based on Human Social Signals. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018*, IFAAMAS, 3 pages.

1 INTRODUCTION

Today's social robots can be found in many domains, including assistive scenarios for health and elderly care, education, stores, but also pubs and everyday life. In recent years, robots for entertainment purposes are increasingly becoming popular, recent research sends them into theatres and standup comedy shows with one single goal: let the robot make you laugh. Laughing together may not only be healthy for the human, but also strengthen the relationship to social robots and make interactions more enjoyable over a longer period of time.

In order to entertain the human, a robot should be able to learn which kind of humor the human prefers. Learning, which stimuli should be included in the presentation, as well as which type of jokes and humor is preferred for one single user, is a challenging task for a social robot.

Human feedback from the user is essential to learn these preferences. One potential way is asking the user directly, however, this can be disturbing as it requires additional interaction for both the robot and the user, which disrupts the flow of the conversation. Our work proposes the use of implicit feedback from human social signals, namely facial smiles and vocal laughs, without requiring any explicit feedback to learn humor preferences in a joke telling scenario based on Reinforcement Learning (RL), a machine learning approach based on trial and error.

2 RELATED WORK

Sjöbergh and Araki [7] found that jokes told by a robot are rated significantly funnier than jokes presented by text only. According to the authors, the presentation method is one of many potential factors influencing the perceived funniness of jokes. This is in line with work by Katevas et al. [2] who emphasize that not only the joke itself, but also the joke delivery is of central importance, independent of the audience size. Several factors need to be addressed, including "intonation, posture, gaze, gesture, expression and timing", but also the communication between speaker and audience. Feedback, such as smiles or raised eyebrows, can be very useful for the speaker.

Mastering the art of presenting humor with a robot has already been researched in the context of Japanese Manzai [1] and standup comedy [2, 3] with focus on larger audiences. Research so far investigates how to adapt spoken contents and presentation in a manner that optimizes the show for lots of people, based on their explicit

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

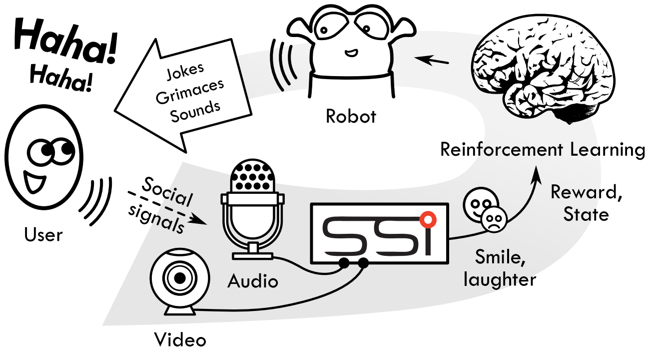

Figure 1: Interaction scenario involving a robot learning how to be funny from human social signals.

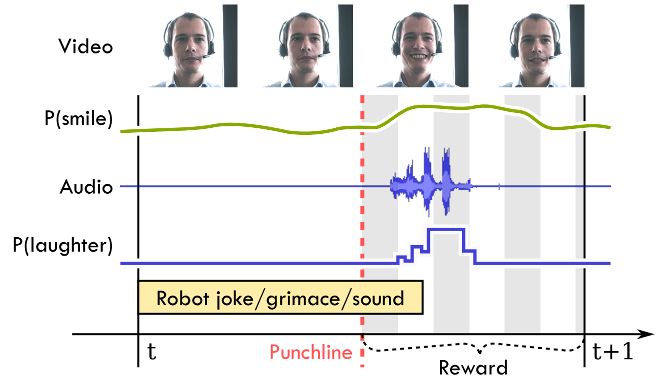

Figure 2: Reward calculation based on social signals.

and implicit feedback. However, for joke-telling scenarios, adaptation is also important for smaller audiences like single humans, when you want to equip a domestic companion with joke-telling abilities, which are optimized for this particular person. As an aggravating factor, joke preferences are individual: people enjoy different types, whether it be black, gross-out, slapstick, academic jokes or nonsense.

Our contribution is a fully autonomous, real-time adaptation approach to learn humor preferences based on Reinforcement Learning with linear function approximation and social signals derived from voiced laughter and smile.

3 SOCIAL ADAPTATION PROCESS

The main requirement of our adaption process is not forcing the user to provide any explicit feedback in order to enable the robot to adapt. To meet this requirement, the learning process uses social signals for learning [6]. This is a useful approach for the simple reason that social signals are produced by the user anyway without requiring specific effort. Those signals, the probabilities of vocal laughs and the facial smiles, are used for defining the state space of our robot in order to give the robot the ability to learn the relationship between the current level of smile/laughter and the best action to be chosen. This is because the automated recognition of human laughs has been applied to a number of conversational interfaces in order to generate an engaging and pleasant user experience [5, 10].

We developed a robot serving as an entertainer whose goal is to elicit laughter in users by making funny grimaces, producing sounds or telling jokes. We provided the robot with several categories as action set, such as three different joke categories, grimaces, sounds and the combination of grimaces+sounds and grimaces+jokes. Overall, we defined 108 jokes split into the three joke categories, 19 grimaces and 23 sounds. Consequently, there exist 437 grimace-sound and 2052 grimace-joke combinations.

The robot selects an action depending on the learned preference and presents it to the user who reacts by smiling and laughing as illustrated in Figure 1. These signals are sensed and processed to compute the probabilities of laughter and smile. These are then used to update the user's preferences, i.e. the robot's knowledge, based on RL and linear function approximation (Sutton et al. [8]). Depending on these preferences, the user gets presented the jokes that he or she finds the funniest in the way that is liked the most.

We employ the Social-Signal-Interpretation framework (SSI) [11] to capture and analyze the required audio-visual signals: Audio signals are used for estimating the probability of the user's vocal laughter, whereas video signals are used to compute the estimated probability of the user's smile. Though there are of course more sophisticated approaches to quantify the intensity of vocal laughs [9] and facial smiles [4], the probabilistic output of the implemented classification system turns out to be a good intensity measure for our application.

The robot's goal is to keep the amount of the user's vocal laughs and facial smiles as high as possible. Therefore, both laughs and smiles are used to compute a reward $\mathcal{R}_t \in [0, 1]$ at time t. The closer to 1, the better the performance. Both the laughs and smiles are defined as value $\rho \in [0, 1]$. According to Katevas et al. [2] people's response to a joke usualy peaks out just after the punchline. Following their findings, our robot measures the smiles and laughter at rate 2.5Hz from a predefined punchline, takes the average and sums up both values weighted by 0.5 to get the final reward $\mathcal{R}_t$ used for the learning process as visualized in figure 2. Smiles are sensed as continuous stream, whereas jokes are only detected as they occur and sent as event. For actions not having a punchline, such as grimaces and sounds, Katevas et al. also showed that the best time for measurement is during and right after the action, but not before either.

We conducted a preliminary user study with 24 participants to compare the learning agent with a random agent (as a baseline). The study revealed that the learning agent acquired a slightly more positve reward than the random agent.

4 SUMMARY AND FUTURE WORK

In this paper, we presented a real-time RL approach solely based on social signals in order to have a robot adapt to a single user's humor preferences without requiring him or her to give explicit feedback. As a proof-of-concept we implemented and tested a demonstrator where a robot has been serving as an entertainer telling the user's different kind of jokes. In our future work, we would like to explore a richer set of nonverbal behaviors for presenting jokes to users in a personalized manner.

REFERENCES

[1] Kotaro Hayashi, Takayuki Kanda, Takahiro Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2008. Robot *Manzai*: Robot Conversation as a Passive-Social Medium. *I. J. Humanoid Robotics* 5, 1 (2008), 67–86. https://doi.org/10.1142/S0219843608001315

[2] Kleomenis Katevas, Patrick GT Healey, and Matthew Tobias Harris. 2015. Robot Comedy Lab: experimenting with the social dynamics of live performance. *Frontiers in psychology* 6 (2015).

[3] Heather Knight. 2011. *Eight Lessons Learned about Non-verbal Interactions through Robot Theater.* Springer Berlin Heidelberg, Berlin, Heidelberg, 42–51. https://doi.org/10.1007/978-3-642-25504-5_5

[4] Robert Lynch. 2010. It's funny because we think it's true: laughter is augmented by implicit preferences. *Evolution and Human Behavior* 31, 2 (2010), 141–148.

[5] Radosław Niewiadomski, Jennifer Hofmann, Jérôme Urbain, Tracey Platt, Johannes Wagner, Bilal Piot, Huseyin Cakmak, Sathish Pammi, Tobias Baur, Stephane Dupont, et al. 2013. Laugh-aware virtual agent and its impact on user amusement. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems.* International Foundation for Autonomous Agents and Multiagent Systems, 619–626.

[6] Hannes Ritschel, Tobias Baur, and Elisabeth André. 2017. Adapting a Robot's Linguistic Style Based on Socially-Aware Reinforcement Learning. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 378–384.

[7] Jonas Sjöbergh and Kenji Araki. 2009. *Robots Make Things Funnier.* Springer Berlin Heidelberg, Berlin, Heidelberg, 306–313. https://doi.org/10.1007/978-3-642-00609-8_27

[8] Richard S. Sutton, Hamid Reza Maei, Doina Precup, Shalabh Bhatnagar, David Silver, Csaba Szepesvári, and Eric Wiewiora. 2009. Fast Gradient-descent Methods for Temporal-difference Learning with Linear Function Approximation. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09).* ACM, New York, NY, USA, 993–1000. https://doi.org/10.1145/1553374.1553501

[9] Jérôme Urbain, Hüseyin Çakmak, Aurélie Charlier, Maxime Denti, Thierry Dutoit, and Stéphane Dupont. 2014. Arousal-driven synthesis of laughter. *IEEE Journal of Selected Topics in Signal Processing* 8, 2 (2014), 273–284.

[10] Jérôme Urbain, Radoslaw Niewiadomski, Elisabetta Bevacqua, Thierry Dutoit, Alexis Moinet, Catherine Pelachaud, Benjamin Picart, Joëlle Tilmanne, and Johannes Wagner. 2010. Avlaughtercycle. *Journal on Multimodal User Interfaces* 4, 1 (2010), 47–58.

[11] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The Social Signal Interpretation (SSI) Framework: Multimodal Signal Processing and Recognition in Real-time. In *Proceedings of the 21st ACM International Conference on Multimedia (MM '13).* ACM, New York, NY, USA, 831–834. https://doi.org/10.1145/2502081.2502223