FACIAL EXPRESSIONS OF MOVIE-GOERS

Entry to Challenge: *How would you teach AI to be kind?*

Nandi Schoots

nandischoots@yahoo.com

Introduction

We would like a data-set that contains moral or social situations with normative labels.

Storytelling emerged in almost all cultures around the world. It evolved to entertain, educate and instill values. As a result the density of socially didactic situations in movies (and literature) is very high. The proposal is to collect stories (or situations with moral weight from stories) together with human emotional responses to those stories. More concretely, the proposal is to abstract a script and label the narrative with human emotional responses by observing facial expressions of people watching the movie.

The idea of using narrative extraction to teach AI human values has been preliminarily explored before, see [1]. We propose to extract moral information from stories by combining existing machine learning approaches for narrative extraction and facial expression recognition.

Encoding scripts into narrative event chains

To collect and encode the stories, we can use a machine learning algorithm that transforms a movie-script into a more abstract representation of the narrative. Unsupervised learning can be applied to scripts to extract narrative event chains (partially ordered sets of events centered around a common protagonist) as in [2]. Alternatively, instead of applying machine learning to a script, more advanced machinery could be used to interpret the narrative portrayed through a sequence of images. Related work includes [3], [4] and [5]

Translating facial expressions into emotional responses

To collect human emotional responses to situations we propose that people's facial expression upon watching a movie are recorded and classified.

Facial expressions could be classified into a neutral expression or an expression of one of the basic emotions. Important factors when choosing which emotions to include in the mapping of facial expressions to emotions are how basic the emotion is (to what extent other emotions are composed of it) and how universal the expression of the emotion is. An option would be to use the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) identified by Paul Ekman, which have corresponding (cross-culturally) universal facial expressions.

Emotional responses can be classified by how desirable they are, e.g., happiness is considered positive and anger negative. One option is to translate the facial expressions into a normative label. This makes the dataset readily usable by AI to start predicting which of its possible actions are good. However, a small mistake in the translation of emotion to normative value may result in a large accumulation of error if the dataset is used widely. In case we conclude after investigation that the risk of an error in the mapping from emotion to normative label is too large, we may choose to leave the data rawer.

To get a more nuanced view of human emotional responses, we suggest showing the same movie to multiple people. For example, a movie could be shown at the cinema to a whole audience. We could then use the most common emotional response as label. Alternatively, instead of just labeling an event or situation with just one emotion, we could choose to label it with a set of probability distributions, one for each emotion, that show how likely the event is to elicit the emotion.

It is important that people are not aware of being recorded, because this may influence how they express themselves. An option would be to ask movie-goers if they want to be part of an experiment (in exchange for a small reward) and provide a small questionnaire before the start of the movie, so that people think the experiment is already over when they start watching the movie.

Work into using machine learning to classify facial expression includes [6], [7], [8] and [9].

Combining the two: Labeling narrative event chains with emotional responses

Each event of the narrative event chain (based on the script or movie) should be labeled with (the distribution of) emotional responses. This would require a machine learning project to build narrative event chains based on scripts and a machine learning project on recognizing facial expressions. Lastly, building this dataset requires the installation of cameras in cinemas and willing participants.

The resulting data consists of a large set of social situations and common emotional reactions to those situations, which allows the AI to interpret them and understand what they mean to people. From this dataset the AI can learn to predict what events or actions will elicit positive or negative emotions.

References

[1] Mark O. Riedl and Brent Harrison *Using Stories to Teach Human Values to Artificial Agents.* AAAI Workshop: AI, Ethics, and Society, 2016.

[2] Chambers, N., and Jurafsky, D. *Unsupervised learning of narrative event chains.* In Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2008.

[3] Finlayson, M. *Learning Narrative Structure from Annotated Folktales.* Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2011.

[4] Mueller, E. *Understanding script-based stories using commonsense reasoning.* Cognitive Systems Research 5(4):307–340, 2004.

[5] Schank, R., and Abelson, R. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.* Lawrence Erlbaum Associates, 1977.

[6] Giorgana, G., and Ploeger, P. G. *Facial expression recognition for domestic service robots.* Robo Cup 2011: Robot Soccer World Cup XV, pp. 353-364, 2012.

[7] Sarawagi, V., and Arya, K. V. *Automatic facial expression recognition for image sequences.* Contemporary Computing (IC3), Sixth International Conference on, 278-282, 2013.

[8] Yu, H., and Liu, H. *Combining appearance and geometric features for facial expression recognition.* Sixth International Conference on Graphic and Image Processing , 944308-944308, 2015.

[9] Zhang, L., and Tjondronegoro, D. *Facial expression recognition using facial movement features.* IEEE Transactions on Affective Computing, 2(4), 219-229, 2011.