

Automatic emotion analysis on human-robot interaction

Student: Daniel Llatas Spiers

Supervisor: Maike Paetzel

January, 2016.

Background

The use of robots in society has increased widely in the last decades. Nowadays, robots are used in many different industries. Even their presence on daily activities are starting to become common for people. As the exposure between human and robots is increasing, the interaction also has to become smoother and more natural. In order to achieve this, robots have to be provided with a capability that let them understand the surrounding environment. Specially, the intentions of a human being.

In the last years, the use of Deep Learning algorithms has been proven to be very successful in different areas. For instance, Jeremy Howard showed on his Brussels 2014 TEDx's talk how computers trained using Deep Learning were able to achieve some amazing tasks. These tasks include the ability to learn Chinese language, to recognize objects in images, and to help on medical diagnosis [1].

This project attempts to improve human-robot interaction. The focus is to explore on the possibilities that face-to-face communication data can provide us. As stated by Alex Pentland, face-to-face communication provided us not only with the conversation itself (words) but with a set of features that reveals the speaker's inner state [2]. This field is called Social Signal Processing.

Finally, this project plans to use Deep Learning for Social Signal Processing to provide a robot with a capability to react on human emotions in a proper way. Existing studies training artificial neural networks have been able to classify emotions with good results [3] [4]. By succeeding in this task, it is one step closer to extend the use of robots to other fields where the interaction with humans is critical.

Description of the task

This project has been divided into two phases. The first phase intends to evaluate available datasets, to find relevant features, and to train and to test the artificial neural network. It is worthy to mention the dependence between the data captured by the robot's camera (Kinect) and the datasets available. Moreover, it is important to select a dataset for which it is easy to extract the same features as the Microsoft Kinect framework extracts, because feature extraction is a very complex field in itself and it would be beyond the scope of a Master thesis to come up with an own framework for that. Thus, it is preferable to look for a labeled Kinect dataset. The second phase will focus on designing a user study to collect video data. A set of participants will be recorded using a Kinect

device when interacting with the robot. This dataset will be very specific for this project, so it will be very useful to test the performance of the neural network trained on the previous phase.

The following list presents the two phases, as well as a brief description for each task:

Phase 1

- Evaluate datasets: Search for existing labeled Kinect datasets available for research on this topic. It might be preferable to look for datasets providing features extracted from videos using Kinect framework.
- Feature extraction: So far only one relevant Kinect dataset has been found [9]. Extensive research has to be done in order to choose the features that will help us achieve the best performance on the experiment.
- Data preprocessing: Involves cleaning the data or changing its format to fit a particular one required by the deep learning framework.
- Set up experiment environment: Installing tools and software that will be used to perform the deep learning experiment. Also, implies deciding on the architecture of the artificial neural network.
- Train artificial neural network: Executing a piece of software that will teach the network to classify on the human emotions based on features previously selected. The original dataset will be divided in training set and test set. The training set is going to be used on this task.
- Test artificial neural network: Checking the accuracy of the trained network on new observations. As it was previously stated, the testing will use test set, a subset of the original dataset.

Phase 2

- Design user study: Decide on patterns to conduct a study on people interacting with the robot.
- Set up user study: Get a place, and the appropriate equipment to record the experiment. Also, gather around volunteers to interact with the robot.
- Perform user study: Conduct the study with participants. Store the videos collected on the previous task on a persistent way. Also, classify them accordingly to the emotions expressed.
- Test artificial neural network with new dataset: Checking the accuracy of the trained network on Phase 1 with the new observations collected on this phase. Another option is to re train the artificial neural network using Phase 1 dataset as the training set. And, then use the Phase 2 dataset to perform the testing.

Methods

On the Phase 1, a set of diverse tools will be needed to complete all the tasks. Evaluating datasets is dependent on the position and type of robot's camera. Some datasets seems promising for the experiment. One of them has data related to colorful conversations. It is available for academic purposes and it is provided by SSPNET (Social signal processing network) [6]. Affectiva (a company that researches on emotion) in collaboration with MIT have released a dataset of spontaneous facial responses to three Super Bowl Ads [5].

Another critical task in this project is to find relevant features. The use of convolutional neural networks on videos have showed some interesting results on this task [8]. However, further research has to be done before deciding on the use of a particular algorithm.

Finally, in order to execute the training, it would be better to use some existing solution. For example, Google released Tensor Flow [7] some months ago. Tensor Flow is an open source library that allows to set up and to execute machine learning experiments representing them through data flow graphs. This tool seems particular interesting to use during this project since it was designed and implemented mainly by Google Brain Team, who were conducting deep neural networks research.

On Phase 2, a qualitative research is necessary. A focus group might be arranged with a group of volunteers. The volunteers will interact with the robot in different ways. Then, Kinect framework has to be accessed in order to get the collected information. Then, it might be worthy to manipulate the data in order to clean it up. A programming language as C++ could be used on both tasks because it has shown to have good support for processing Kinect data. Finally, methods on Phase 1 will be reuse to test the trained neural network's performance with the new dataset.

Relevant courses

1. Machine Learning
2. Artificial Intelligence
3. Data Mining
4. Image Analysis

Delimitations

This project is just a puzzle piece on the main goal of Social Robotics Lab at Uppsala University. The main goal is to develop a robot capable of multimodal interaction. On that sense, this project is aimed to research on the possibilities that the Kinect device could offer in order to detect emotions. Originally, an extra phase was discussed. The focus on that phase was to implement an engine to classify emotions on real time. However, it has been decided later that this would not be feasible to achieve on the project's time span.

Time plan

Week	Date	Phase	Task 1	Task 2	Task 3
1	25/01/2016	1	Evaluate datasets		
2	01/02/2016	1	Feature extraction	Meeting with reviewer	Write report
3	08/02/2016	1	Feature extraction	Write report	
4	15/02/2016	1	Feature extraction	Meeting with reviewer	
5	22/02/2016	1	Feature extraction	Data preprocessing	Write report
6	29/02/2016	1	Data preprocessing	Meeting with reviewer	
7	07/03/2016	1	Set up experiment environment	Write report	
8	14/03/2016	1	Set up experiment environment	Meeting with reviewer	
9	21/03/2016	1 - 2	Train ANN	Design user study	Write report
10	28/03/2016	1	Train ANN	Meeting with reviewer	Write report
11	04/04/2016	1 - 2	Test ANN	Design user study	Write report
12	11/04/2016	1 - 2	Test ANN	Set up user study	Meeting with reviewer
13	18/04/2016	2	Set up user study	Write report	
14	25/04/2016	2	Set up user study	Meeting with reviewer	Write report
15	02/05/2016	2	Set up user study		
16	09/05/2016	2	Perform user study	Meeting with reviewer	
17	16/05/2016	2	Perform user study		
18	23/05/2016	2	Test ANN with new dataset	Meeting with reviewer	
19	30/05/2016	2	Write report	Project presentation at vi2 seminar	
20	06/06/2016	2	Perform project presentation	Meeting with reviewer	

References

- [1] Howard, J. (2014, December). Jeremy Howard: The wonderful and terrifying implications of computers that can learn. [Video file]. Retrieved from https://www.ted.com/talks/jeremy_howard_the_wonderful_and_terrifying_implications_of_computers_that_can_learn
- [2] Pentland, A. S. (2007). Social signal processing [exploratory DSP]. *Signal Processing Magazine, IEEE*, 24(4), 108-111.
- [3] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in *Proc. Interspeech*, 2014.
- [4] Yelin Kim, Honglak Lee, and Emily Mower Provost. "Deep learning for robust feature generation in audiovisual emotion recognition". University of Michigan, Electrical Engineering and Computer Science, Ann Arbor, Michigan, USA. 2013.
- [5] Daniel McDuff, Rana El Kaliouby, Thibaud Senechal, May Amr, Jeffrey Cohn, Rosalind Picard, Affectiva MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In the Wild", *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013.
- [6] Gary McKeown, Michel F. Valstar, Roderick Cowie, and Maja Pantic, 'The SEMAINE Corpus of Emotionally Coloured Character Interactions', *Proc. IEEE Int'l Conf. Multimedia & Expo (ICME'10)*, pp. 1079-1084, Singapore, July 2010.
- [7] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from tensorflow.org.
- [8] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar and Li Fei-Fei. 2014. "Large-scale Video Classification with Convolutional Neural Networks". *Proceedings of International Computer Vision and Pattern Recognition (CVPR 2014)*
- [9] Rui Min, Neslihan Kose, Jean-Luc Dugelay, "KinectFaceDB: A Kinect Database for Face Recognition," *Systems, Man, and Cybernetics: Systems*, *IEEE Transactions on*, vol.44, no.11, pp.1534,1548, Nov. 2014, doi: 10.1109/TSMC.2014.2331215