08/8201 CURIOSITY CLONING - NEURAL MODELLING FOR IMAGE ANALYSIS

Type of activity: Medium Study (4 months)

Background and Motivation

Neurophysiological correlates of human curiosity

Human and animal curiosity can be defined as the natural inquisitive behaviour, which engenders exploration, investigation, and learning. Being these inherent qualities and abilities of intelligent beings, it was not by chance that the Artificial Intelligence community had an attempt to replicate them into artificial agents like robots or specific programmes. Still, what curiosity really is remains a mystery: product of a complex evolutionary chain, there are some details, some particular combination of sensorial data and experiences, which forces attention out from the background noise. These "triggering" details often remain out of our consciousness and they can be hardly reported and fully described.

Consequently, for the Artificial Intelligence community curiosity as such remains a "black box". It is remarkable that until now only rather simple decision-taking modules trying to replicate a similar final behaviour [1] have been elaborated, but still no curiosity engines for AI agents based on a reproduction of curiosity patterns have reached concrete results.

Even if complex, curiosity still remains a mental activity, and as such hinges upon basic chemo-electrical activities of networks of neurons. One can therefore expect that research in the field of neurophysiology could one day answer the question of *what* curiosity is. As by today, neurophysiology could already answer *when* curiosity happens.

"Curious" is often synonymous of inexplicable, highly unusual, odd, or strange. In other terms, curiosity is associated with phenomena, which arouse speculation, interest, or particular attention. These features have been reported to excite the parietal cortex in a very characteristic way. Actually, such a stimulus produces a synchronized peak in the global electrical activity of large groups of neurons in this area, which is perceived from the outside (for example via an EEG) as an electric potential wave. Since this wave has a positive voltage peak and follows the stimulus after 300 ms, it is called P300 [2]. P300 shows interesting features: its magnitude is associated with the level of attention the stimulus is arousing; it can not be fine controlled and it was reported to be, at least partially, independent from consciousness.[2]

Thus, the underlying assumption of the proposed study is that the P300 can serve as a correlate of the level of curiosity of an individual.

Exploiting neuronal correlates for remote data evaluation

Extra-orbital missions can be equipped with several high-definition sensors, allowing to autonomously collecting a potentially enormous amount of data. Typically, the bottleneck in retrieving these data-sets is manifested firstly in the available storing capabilities and secondly in the limited communication bandwidth, which prevents from sending the whole data-set back to Earth. This issue is particularly severe for image data, which is usually quite demanding in terms of dimension (bits) and, since the best possible resolution and quality is normally required by scientist, even hardly compressible in size.

Hence, even if explorative robots could take a vast amount of pictures, these will eventually have to be reduced in number. Separating the scientifically relevant pictures from the less relevant ones is the crucial task. In consequence, the robot has to evaluate in real-time the scientific content of a picture currently by assigning so called Scientific Richness Indexes [3].

These indexes take into account basic features of the picture, like e.g. entropy, contrast, etc. The sorting mechanism is designed a priori, and no learning is involved in the algorithm itself. Classifier systems based on supervised learning could therefore be used as an alternative. Able to learn a possible dependence between those indexes (and maybe other sets of features), with the subjective scientific richness of a picture as evaluated by experts, these systems could give improved performances.

In order to learn how to evaluate pictures, all the supervised learning based systems require a precise and long lasting protocol of actions. First of all several human operators, expert analyzers of scientific pictures (i.e. geologists with an expertise in Martian rocks analysis), will have to evaluate hundreds or thousands of pictures, by rating them with one or several numbers to define their interest(s) value(s). The very same pictures, together with their evaluation, would form the data set required for the training and testing of a classifier, or function approximator.

The learning algorithm will then be run on a subset of the whole data, the so called training set. Parameters will be adjusted by optimizing its performance on that set, or a subset (cross validation), and then the performance of the algorithm will be measured on a separate test set, to assess whether the trained network is finally able to reproduce the same evaluation of the human experts. This approach - which is usually very expensive in terms of human-power and effort required to produce adequate data-sets and systems for images' scientific content evaluation - has not been implemented yet in any space mission.

By systematically detecting P300 events and analyzing their features, it is possible to infer curiosity in terms of timing, magnitude, and persistence [4]; by correlating it with the corresponding sensorial stimulus, it becomes possible to assign an interest level to each single stimulus presented. A curiosity-level derived set of data can then be used to train and test a classifier. Ideally such classifier would react to stimuli showing the same level of curiosity that had been monitored from the person who was "curiosity-recorded" in the first place. In short, the person's curiosity would somehow be replicated – or "cloned" – into an artificial system.

Moreover, since the P300 shows curiosity-raising at its very beginning, it should be possible to classify the interest-level of an image faster than by the commonly performed interview technique, and without a possible bias operated by the subject's conscious filtering. For a large set of images, as it is required for the training-testing data-set, reducing the time dedicated to the analysis of an image can have drastic effects on the total time required to the subject to spend "looking at images".

How far this approach can simulate the human curiosity and actually improve the performance of an artificial network is the open question at the core of this study. To focus on a well-defined problem we propose an experiment on image evaluation, which is a serious issue for most of the extra-orbital space missions.

The idea is to perform an experiment, where a group of test subjects – instructed to look for ‘scientifically interesting’ images - is confronted with a set of images, while EEG recorded. Data coming from EEG recordings are analysed to find correlations between the level of P300 expressed by the subjects and the picture displayed. Once the EEG recording and analyzing experimental sessions has provided enough information, this data-set can be used to train a classifier.

Eventually, the system should be able to work as a trained artificial image sorter, sorting out a test-set of pictures for their ability of arousing the scientific curiosity of the subject

Research and Study Objectives

The objective of this study is to prove that the scientific curiosity about images can be detected by reading EEG measurements, including the P300 signal. This should be achieved by performing a set of experiments at the universities’ or research centres’ premises.

Assuming success and following this study, these data will be used to develop an artificial image sorter trained with these curiosity measurements. The research centres and university/ies participating in the first phase will be invited to participate also to this follow-up activity.

As part of its contribution, the ACT will provide at the study kick-off sets of appropriately pictured, representative images. These will consist of approximately 3 times 2000 pictures.

In their proposal, university(ies) and research centres are asked to provide an experimental protocol allowing to answer the question whether the P300 signal can effectively measure the curiosity level of humans on these picture sets. Universities and research centres are encouraged to propose the use of additional signals to measure curiosity.

The proposed system should be aimed to monitor the level of curiosity aroused by each individual picture presented. University(ies) and research centres may also propose an additional set of images, pictured in order to maximise the responsiveness and the effectiveness of the proposed system.

As part of the proposed protocol, the proposals should include measurements allowing understanding the maximum reliable picture display rate (pictures displayed per minute) achievable, by correlating error rates with picture display rates.

In summary, proposals for this study to be performed together with the ACT should contain information on:

- the proposed experimental procedure in order to continuously record the level of P300 of a subject while he/she is analyzing the scientific interest of a set of randomly sorted pictures.
- the proposed procedure to correlate the P300 peaks (or other proposed signals) with the exact picture-stimulus, which triggers them.
- the proposed protocol to find the fastest picture display-rate which will still permit the arousing and correlation of P300 peaks and to define the relationship between performance of the correlating system and picture display-rate.

- the proposed facilities to perform the experiments.

Universities and research centres are required to possess or have access to all the facilities required to perform the proposed study within the specified time frame.

The processed data sets should be made available in ascii format.

References

1. Phillips-Wren, G., Ichalkaranje, N., Jain, L., Intelligent Decision Making: An AI-Based Approach, Springer, Series: Studies in Computational Intelligence, Vol. 97, 2008
2. Picton, T.W., The P300 wave of the human event-related potential, *J. Clin Neurophysiol*., vol. 9, no. 4, pp. 456-79, October 1992
3. Barnes, D., Shawa, A., Summersa, P., Wardb, R., Woodsb, M., Evansb, M., Paarc, G., and Simsd, M., Imaging and localisation software demonstrator for planetary aerobots, *Acta Astronautica, 59*, pp. 1062-1070. Elsevier. doi:10.1016/j.actaastro.2005.07.050, 2006
4. Gerson, A.D, Parra, L. C., and Sadja, P., Cortically Coupled Computer Vision for Rapid Image Search, *IEEE Transactions on neural systems and rehabilitation engineering*, pp. 174-179, Vol 14, No 2, June 2006.