Predicting Single Observer's Votes from Objective Measures using Neural Networks

Lohic Fotio Tiotsop [1] Tomas Mizdos [2] Miroslav Uhrina [2]
Peter Pocta [2] Marcus Barkowsky [3] Enrico Masala [1]

[1] Control and Computer Engineering Department, Politecnico di Torino, Torino, Italy

[2] University of Zilina, Zilina, Slovakia

[3] Deggendorf Institute of Technology (DIT), Deggendorf, Germany

VQEG JEG-Hybrid Group

Media Quality Assessment using Neural Networks (NN)

NNs have been extensively used for MOS prediction of processed video sequences (PVS)

- Shallow NN approach

$$NN(PVS_{features}) \rightarrow MOS$$

- Deep NN approach

$$NN(PVS) \rightarrow MOS$$

Modelling Subjects Opinions: The Approach

- Design NNs that attempt to mimic single observer behaviour to predict observers' opinion scores (OS) rather than the MOS
- Train a NN_{obs} for each observer such that

$$NN_{obs}(PVS_{features}) \rightarrow [p_1, p_2, p_3, p_4, p_5]$$

$$p_i = Probability(OS = i)$$

- Predicted OS is determined as $i = \arg\max_i(p_i)$
- The uncertainty of the observer when assessing the quality of a PVS is measured by

$$\sigma_{obs}^{pvs} = \sum_{i=1}^{5} i^2 \cdot p_i - \left(\sum_{i=1}^{5} i \cdot p_i\right)^2$$

Modelling Subjects Opinions: Motivations

- To be able to predict not only the MOS, but also the SOS and confidence intervals.
- To be able to perform statistical tests regarding the quality of a PVS compared to another.
- Evaluate how much confident is an observer when assigning a score to each PVS
- To be able to estimate the distribution of observers' opinions
- Enabling the simulation of subjective experiments

Dataset Description

- Subjective data: VQEG-HDTV experiments set 1, 3 and 5
- Video Quality metrics (VQMs) for each PVS: PSNR, SSIM, MSSSIM, VIF and VMAF
- Features for each PVS: Blockiness, Blockloss, Blur, Noise, Contrast, Flickering, Spatial activity index (SI) and Temporal activity index (TI)
- VQEG-HDTV subjective data comes from experiments in different labs; data has been aligned by means of a common set: 24 PVSs evaluated in all the 3 experiments

Numerical Experiments Setup

- Sets are considered together to have more data for training
    - Implies forming triplets of observers, one from each set
    - Input: 72 observers, 24 for each set
    - Output: 24 "virtual" observers that "virtually" participated to all the 3 experiments, e.g.,

$$Obs_{1'} \rightarrow [Obs_1^{set1}, Obs_{17}^{set3}, Obs_{11}^{set5}]$$

    - Triplets are formed in such a way that the total mutual RMSE on the score of the observers is minimized on the common set of PVS.
    - Subjects that voted similarly on the common set are likely to be connected to form the triplets

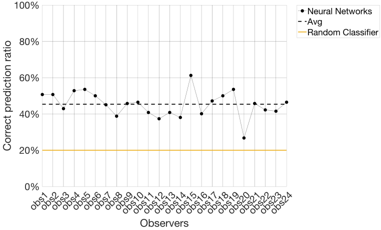
Numerical Experiments Setup

- Train 24 NN, one for each observer, using 2 sets and test on the one kept out.
- Input: VQMs, VQMs+features or VQMs+features distribution
- Labels: OSs of the corresponding observer
- We define the following indexes for each observer

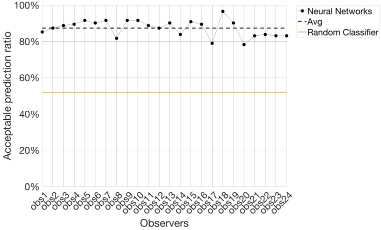
$$\textit{Correct prediction ratio} = \frac{\#(\textit{predicted OS=actual OS})}{\#(\textit{PVS in test set})}$$

$$\textit{Acceptable prediction ratio} = \frac{\#(|\textit{predicted OS-actual OS}| \leq 1)}{\#(\textit{PVS in test set})}$$

Cross Validation



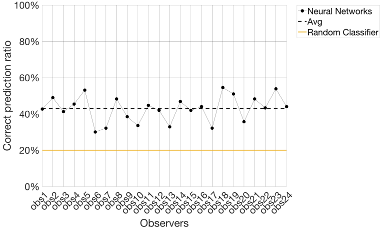
(a) Percentage of correct prediction



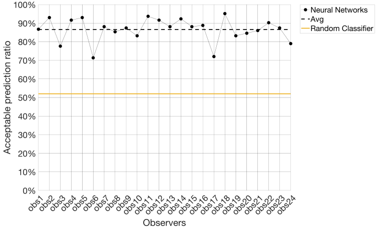
(b) Percentage of acceptable predictions

Figure: The models are trained on the VQEG-HDTV set 1 and 5 then tested on set 3 using only the VQMs as features. For each observers, the related NN performs better than a random classifier

Cross Validation



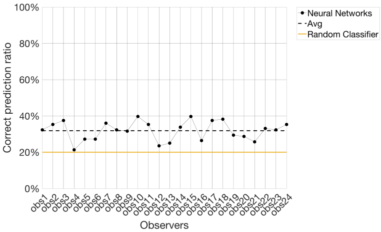
(a) Percentage of correct prediction



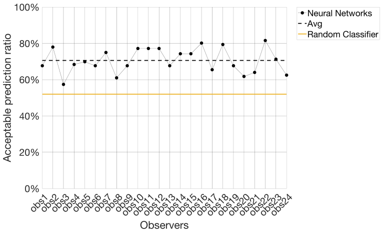
(b) Percentage of acceptable predictions

Figure: The models are trained on the VQEG-HDTV set 3 and 5 then tested on set 1 using only the VQMs as features. For each observers, the related NN performs better than a random classifier

Cross Validation



(a) Percentage of correct prediction



(b) Percentage of acceptable predictions

Figure: The models are trained on the VQEG-HDTV set 1 and 3 then tested on set 5 using only the VQMs as features. For each observers, the related NN performs better than a random classifier

Contribution of No Reference Measures

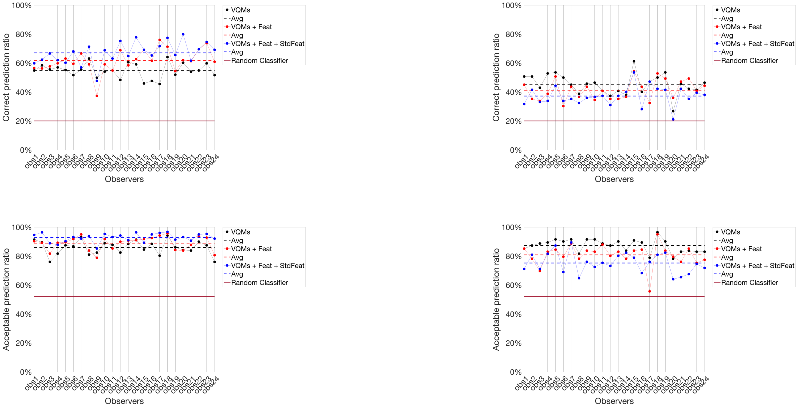


Figure: Contribution of no reference measures on the training (left) and validation (right) set, when predicting the OS.

A Virtual Subjective Experiment

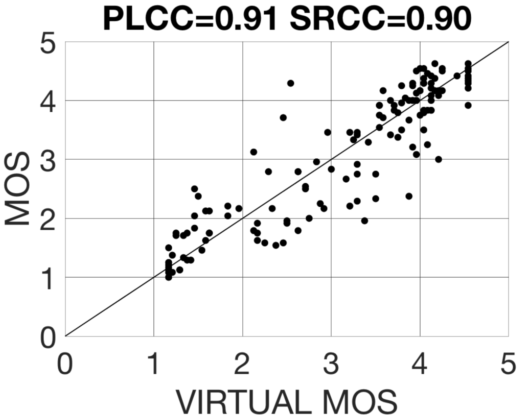


Figure: The NNs (trained on the set 1 and 5) are considered as "Virtual Observer" and used to run a "Virtual subjective experiments" on set 3. The obtained MOS is compared to the actual one

A Virtual Subjective Experiment

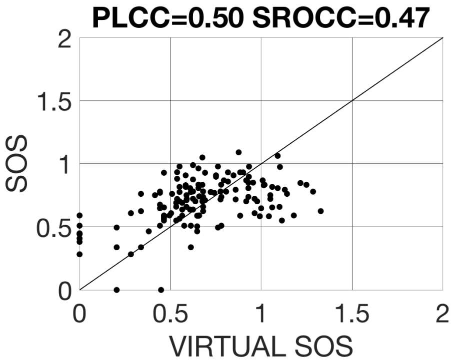


Figure: The NNs are considered as "Virtual Observer" and used to run a "Virtual subjective experiments". The obtained SOS is compared to the actual one. The correlation coefficient is significantly different from 0 ($p_{value} = 2.1 \times 10^{-10}$)

Possible Applications

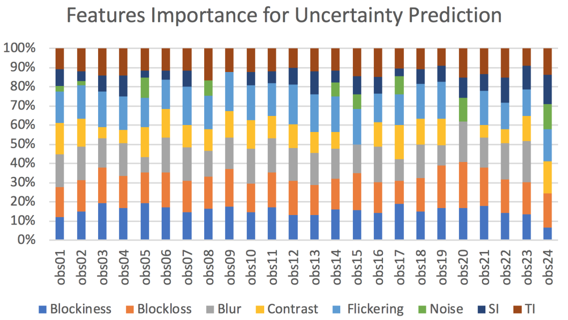
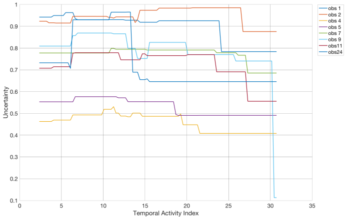


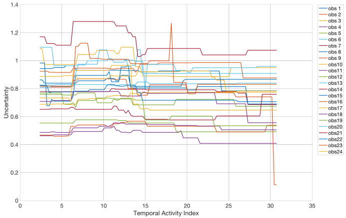
Figure: Importance of (no reference) features for uncertainty prediction using a neighbourhood component analysis (NCA[1]) approach. For instance, it seems that the noise feature is less influential on uncertainty rather than blockiness.

[1] W. Yang, K. Wang, W. Zuo, Neighbourhood Component Feature Selection for High-Dimensional Data, Journal of Computers, vol. 7, n. 1, 2012, pp. 161–168

Possible Applications



(a) Some observers



(b) All observers

Figure: Lower uncertainty values are observed for sequences with high temporal activity. Curves are obtained fitting perceptual features to the uncertainty through regression tree models

Conclusions

- We predict single OSs rather than the MOS by designing a NN for each observer, hence more statistical indicators regarding the perceived quality can be computed.
- This approach allows to quantify how uncertain is an observer when assigning a vote. We study the importance of some perceptual features in predicting such uncertainty.
- Preliminary numerical results showed that NNs can actually learn some information about the process guiding the choices of a single observer.

Future Work

- Designing subjective experiments specifically aimed at subjects modelling
- Employ DL: the perceptual features that affect the judgement of an observer are learned by the NN while training it directly on content.

Thanks for your attention