Paper Reading & Discussion

Paper Title	Improving Mobile Interactive Video QoE via Two-Level Online Cooperative Learning
DOI	10.1109/TMC.2022.3179782
Name	Sahil Pattni
Concordia ID	40216177
Course	COMP 691
Lecture	7

Questions

- The authors state that as neural processing units become available on mobile devices, training such a model on devices would become feasible. With developments in such NPUs, will mobile devices be able to handle not just inference, but continuous training?
  - Additionally, how would it affect both battery life and device performance, assuming that it is likely to be competing with other applications/model for NPU access?
- If Legato was deployed at scale over a significant period of time, would it be possible that the model would stagnate after a given period of time? (i.e. would the model converge to a point where new aggregations of individualized models provide little to no changes in the model?).

Comments

- I like the balance Legato seeks to strike between generalized QoE and individual experience, by iteratively providing an individualized model to each user based on the last aggregated model.
- The real-world data was collected from only 5 users, a small sample size. It might be the case that a larger sample size would provide more varied data, and therefore change the evaluation performance of Legato.
- Using the legacy rule-based system as a base is valuable at the initial learning stages for Legato, where the RL system is in its infancy and may provide inconsistent and incorrect output often.

Paper Title	GRACE: Loss-Resilient Real-Time Video through Neural Codecs
DOI	10.48550/arXiv.2305.12333
Name	Sahil Pattni
Concordia ID	40216177
Course	COMP 691
Lecture	7

Questions

- The authors proposed the adaptation of GRACE towards multi-party conferencing as a potential future work, but wouldn't the limitations of GRACE be more pronounced? For example, the synchronization of the encoder with multiple decoders [1] would have a greater computational cost than currently exists.
- Given that the focus of GRACE is loss resilience, what real-time video applications exist for GRACE where loss resilience is more important than compression efficiency, due to strict latency requirements?

Comments

- The joint training employed by GRACE ensures that even the encoder is aware of the packet losses, thereby allowing it to modify its redundancy rate accordingly.
- Given the complexity of this publication, making the code-base publicly available - as the authors have - will allow for democratized improvements and optimizations of the system.
- Compared to the previous paper, this one has been evaluated on a much larger sample size: 240 participants & nearly a thousand subjective ratings. This allows for a more generalized model as the (sampling) variance reduces as the sample size grows.
- This is (probably) the first effort to jointly train both the encoder and decoder, with the aim of improving loss resilience in real-time video applications. To that end, the authors have developed an end-to-end system, with optimization techniques related to packetization, encoder/decoder state synchronization, and runtime efficiency. Such an end-to-end system can be re-used by future researchers, removing the need to *reinvent the wheel*.

[1] assuming a simple one-to-many stream, a many-to-many stream would be even heavier