View Reviews

Paper ID
7005

Paper Title
Confidence-based federated distillation for vision-based lane-centering

Track Name
Satellite Workshop: Timely and Private Machine Learning over Networks

Reviewer #1

Questions

2. Importance/Relevance
3. Of sufficient interest

4. Novelty/Originality
2. Minor originality

5. Justification of Novelty/Originality Score (required if score is 1 or 2)
The novelty is limited, since the authors propose the usage of entropy as indicator of confidence, this has been done extensively in the literature, but not necessarily for the FL. Thus, I proposed the minor originality.

6. Technical Correctness
3. Probably correct

8. Experimental Validation
3. Limited but convincing

10. Clarity of Presentation
3. Clear enough

12. Reference to Prior Work
3. References adequate

14. Overall evaluation of this paper
2. Marginal reject

15. Justification of Overall evaluation of this paper (required if score is 1 or 2)
The paper is interesting and has some interesting aspects. There are some aspects that can be improved, for example a more thorough literature review on the usage of entropy for confidence, a more thorough discussion on the computation cost of such methods. Furthermore, the authors used the Shannon entropy, what about other types of entropy? Also, a more thorough discussion on convergence is required.

19. Detailed assessment of the paper (seen by the authors):
Please see the comments on 15.

Reviewer #2

Questions

2. Importance/Relevance
3. Of sufficient interest

4. Novelty/Originality

2. Minor originality

5. Justification of Novelty/Originality Score (required if score is 1 or 2)
The authors claim that the idea of selecting one with the lowest entropy by sorting local clients' hypotheses is original in the literature on knowledge distillation. However, there exist several prior works that have proposed similar methods for identifying the best teacher model, such as Bayesian model distillation, AIC-based knowledge distillation, and MDL-based knowledge distillation. These approaches employ analogous mathematical frameworks to select the most appropriate teacher model from a set of candidates.

6. Technical Correctness
3. Probably correct

8. Experimental Validation
2. Lacking in some respect

9. Justification of Experimental Validation Score (required if score is 1 or 2)
In Fig. 2 and Fig. 3, the experiments in case of non-i.i.d. data distribution do not seem to be implemented until convergence. The decreasing trend of losses is clearly visible, but it could have been implemented with more training rounds to reach a saturation point.

10. Clarity of Presentation
2. Difficult to read

11. Justification of Clarity of Presentation Score (required if score is 1 or 2)
The paper has room for improvement in readability. One of the striking examples is that the term "confidence" is not clearly explained in the introduction, although it is an important concept that even appears in the title. It is unclear whether the same data sample must be fed to acquire penultimate layer outputs from each local model. Overall, the general description takes a high proportion, which leads to a longer time for readers to figure out the main contribution of this work.

12. Reference to Prior Work
4. Excellent references

14. Overall evaluation of this paper
2. Marginal reject

15. Justification of Overall evaluation of this paper (required if score is 1 or 2)
This study proposes a confidence-based distillation approach for improving steering angle prediction performance in federated networks. In this approach, the central server gathers penultimate layer outputs of local models and selects the one with the lowest entropy to serve as a teacher model for a global model at the epoch.

Although the paper includes all the essential elements to demonstrate the effectiveness of the proposed approach, the introduction and experiment sections could be improved in terms of organization and language to better persuade readers and enhance readability.

19. Detailed assessment of the paper (seen by the authors):
The proposed method might take a lot of computation cost. The center seems to possess multiple local model copies at once in order to find entropy of each in parallel. Then only one with the lowest entropy is chosen to supervise the global model, which implies that all other copies than the chosen one are discarded, according to what the reviewer has understood.

For the overall assessment, please refer to the review above.

Reviewer #3

Questions

2. Importance/Relevance

2. Of limited interest

4. Novelty/Originality

3. Moderately original

5. Justification of Novelty/Originality Score (required if score is 1 or 2)

The authors leverage ideas from the active learning to solve the non-i.i.d. data problem in FL, which is not seen before.

6. Technical Correctness

3. Probably correct

7. Justification of Technical Correctness Score (required if score is 1 or 2)

The authors have conducted experiments to validate their hypothesis.

8. Experimental Validation

3. Limited but convincing

9. Justification of Experimental Validation Score (required if score is 1 or 2)

The authors have conducted experiments based on widely used dataset and compared their algorithm with other well-known ones.

10. Clarity of Presentation

2. Difficult to read

12. Reference to Prior Work

3. References adequate

13. Justification of Reference to Prior Work Score (required if score is 1 or 2)

The authors have surveyed various existing works in the introduction part.

14. Overall evaluation of this paper

3. Marginal accept

15. Justification of Overall evaluation of this paper (required if score is 1 or 2)

The authors proposed a novel algorithm to solve the non-i.i.d. data problem in FL. Although the algorithm lacks solid theoretical basis, the authors have conducted sufficient experiments to validate the correctness and effectiveness of their algorithm.

19. Detailed assessment of the paper (seen by the authors):

1. Please explain in detail how to obtain p(x_i) in equation (1).
2. y_t and y_s in equation (3) should have subscript i.
3. What is the relationship between y_t, y_s and x_i ?
4. The first item in the right-hand-side of equation (3) should be w^s_{r,j-1}.
5. In line 6 of Alg 1, the symbol n is used without definition.
6. In line 7 of Alg 1, the symbol N is used as the number of steps in each round. However, N is also defined as the number of elements in the output of the penultimate layer.
7. In Fig. 2(b), the proposed algorithm performs worse than the other two in the first 30 rounds. Could you please explain the reason?