Machine Learning (XAI501) Term Project Bi-weekly report

*Created on* December 9, 2020

Bayesian Uncertainty Estimation for Ultrasound Medical Image Segmentation

Sunwoo Kim1, Kwanseok Oh2, Jinhyo Shin2, Sangjin Kim2

1Department of Computer Science and Engineering, 2Department of Artificial Intelligence

**Milestones**

|  |  |  |
| --- | --- | --- |
| Date | Milestones | Remark |
| 10/28 (Wed) | Proposal | Done |
| 11/04 (Wed) | Data preprocessing and understanding | Done |
| 11/11 (Wed) | Proposed architecture design and implementation | Done |
| 11/18 (Wed) | Experiments conducting (1) | Done |
| 11/25 (Wed) | Experiments conducting (2) | Done |
| 12/02 (Wed) | Analysis | Proceeding |
| 12/09 (Wed) | Final documentation and presentation preparation | Proceeding |
| 12/13 (Sun) | Final documentation and presentation preparation |  |

**Group progress**

* Writing report / presentation
* Finding Reference
* Conducting experiment
* Analysis of the result

**Individual progress**

* Sangjin : Finding analytical methods to evaluate experimental outcomes
* Jinhyo : Finding matrix functions and visual analysis methods to estimate uncertainty
* Sunwoo : Search for modules and ideas to improve evaluation performance
* Kwanseok : Using various regularizers and finding parameters for learning bayesian neural networks

**New findings**

* Experiment conducting
  + Attempt to improve segmentation performance
    - Use Dice Loss
      * Dice coefficient : 2TP / (2TP + FP +FN)
      * Dice loss : 1-DC
    - Apply dropout to network
      * Goal : To do representation learning and turn basic network into bayesian network
      * Ablation study (Which layer to apply dropout)
        + Encoder and Decoder both
        + High level of Encoder
        + High level of Decoder
        + Low level of Decoder
      * Ablation result
        + When dropout applied encoder and decoder at the same time, the learning did not work.
        + When dropout applied encoder or decoder only one side, work well and It does not matter which side it is.(very little difference in performance)
      * Ablation analysis
        + A well-trained network has well-defined neurons enabled by each layer of the network, so there is no difference in performance by applying dropout to any layer.
        + Applying dropout means that a basic network is able to interpreted as a bayesian model,
    - Hyper parameter setting
      * Learning rate : 0.001
      * Learning decay (exponential) : 0.99
      * Epoch : 150
      * Batch size : 8
      * Dropout ratio : 0.2
  + Attempt to predict uncertainty map
    - Using MCdropout as dropout method
    - Soft mask
      * Goal : We tried to predict uncertainty by using softmask segmentation, regardless of dropout method
      * Goal 2 : We thought that the use of softmask will be more useful as a medical image analysis aid because it will express more highlight at certain point where model predict strongly sure
  + Metrics for evaluating Performance and Uncertainty
    - Existing metrics from references
      * P(accurate|certain)
      * P(uncertain|inaccurate)
      * These evaluations are not accurate for our task.
    - Imbalanced label problem
      * Our task is a binary segmentation and labels in each image are highly imbalanced. (almost background)
      * The imbalance interferes with accurate evaluation.
      * To overcome we used other well known metrics and custom ver.
    - Other metrics
      * Cohen’s kappa coefficient : A statistic which is used to measure inter-rater reliability. In this, the raters are rater for accuracy and rater for certainty.
      * Canonical IoU (Intersection of Union): One of the popular metrics in computer vision. It is calculated as the ratio of intersection of prediction and Ground Truth and union of them. Union constraints area of interest in evaluation, excluding not meaningful background.
    - Custom Metrics
      * Certain IoU : only predicted pixels where the model is confident are used.
      * Uncertain IoU : predicted pixels where model is not confident are also used.
    - Though our various trials to evaluate the uncertainty properly, these Metrics were not enough to evaluate the uncertainty. How to evaluate the uncertainty still remains unsolved.
  + Result :
    - p(accurate|certain) : 0.98
    - p(uncertain|inaccurate) : 0.39
    - Cohen’s kappa coefficient : 0.31
    - Variance of IoU series for segmentation in the test set is too large : 0.56 (Each IoU score between 0.3 and 0.8)

**Changes to the proposal**

* None