Boosting in Deep Semi-supervised Learning with Multi-view Redundant Information

Juan Castorena CCS-3

Student: Steven Wang UCSD

10/13/2021

LA-UR-21-289
98

Problem

- Data analysis with machine learning.

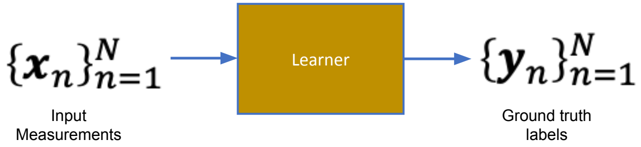

Figure 1: Supervised Learning Schematic

- Label scarcity problem in many scientific experimental settings.
  - Data labeling can be time consuming.
  - Can require human expertise.
  - Can be expensive.

Focus

- Semi-supervised learning (SSL).

$$\{x_n, y_n\}_{n=1}^{N_1} \qquad \{x_n\}_{n=N_1+1}^{N}$$

Labeled data Unlabeled data

exploits both labeled and unlabeled data available.

- Explain how semi-supervised learning works in general (overview).

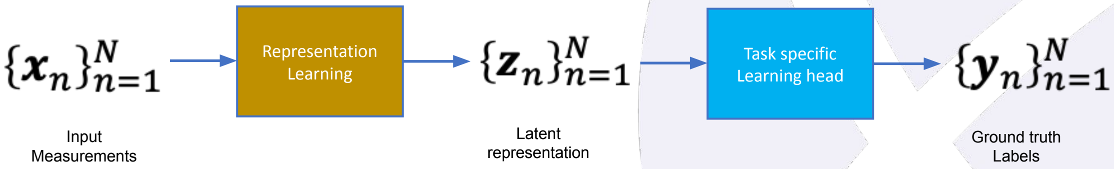

Figure 2: Semi-supervised learning schematic

  - Imposing constraints in latent representation space.

[1] Tarvainen, Antti, and Harri Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results." *arXiv preprint arXiv:1703.01780* (2017).

Problem: Noisy unlabeled data

- Problem of noisy unlabeled examples in SSL
  - Example: A measurement that is negatively correlated with background of a different class.
  - Problematic because we do not have ground truth for unlabeled data points

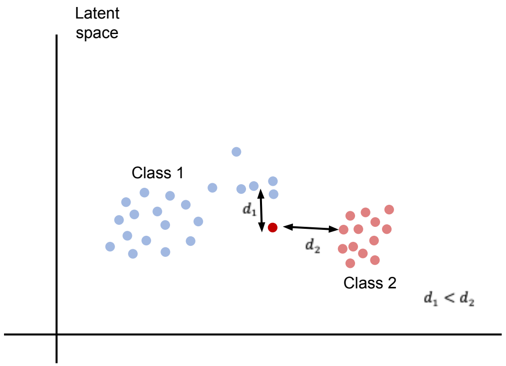

Figure 3:Noisy unlabeled data

Idea: Leverage Multiview information

- Exploit multi-view information typically available in many scientific applications (with aims to reduce uncertainty).
- Most approaches would use synthetic augmentations.

Figure 4: Axial perspective of COVID-19 infected patient [1].

- Multiview information available in many applications:
  - Biomedical
  - Planetary Science
  - Materials
  - Robotics applications.

[1] Afshar, P., Heidarian, S., Enshaei, N. *et al.* COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. *Sci Data* **8,** 121 (2021). https://doi.org/10.1038/s41597-021-00900-3

Idea: Boosting

- Boosting: Idea is to either weight or select (sparsity imposed through ℓ_1) training examples adaptively to remove noisy unlabeled examples.

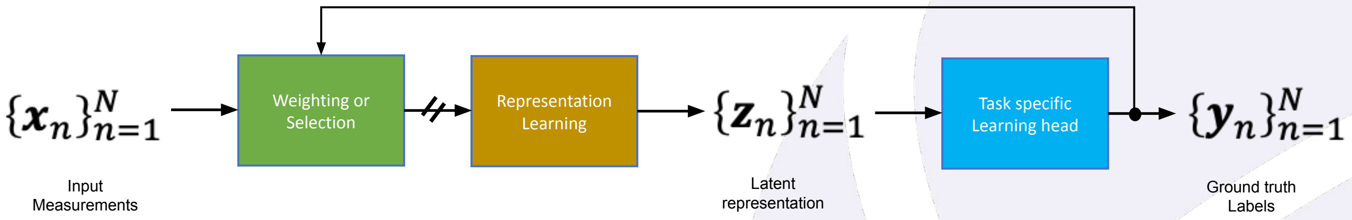

Figure 5: Semi-supervised learning schematic with Boosting

- Problem: minimize regret

$$\sum_{t=1}^{T} \mathbf{w}^t \boldsymbol{\ell}^t - \min_i \sum_{t=1}^{T} \ell_i^t \qquad (1)$$

  - Examples $\{\boldsymbol{x}_n, y_n\}_{n=1}^N$
  - $\boldsymbol{z}_n = h_1^t(\boldsymbol{x}_n), \quad \boldsymbol{y}_n = h_2^t(\boldsymbol{z}_n)$
  - $\boldsymbol{\ell}^t = \sum_{n=1}^{N} d(h_2^t(\boldsymbol{z}_n), \boldsymbol{y}_n)$. where $d: \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ is a distance
  - Weighting or selection of examples $\mathbf{w}^t$
  - Solution Adaboost [1]

[1] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55.1 (1997): 119-139.

Approach: Latent representation

- Idea is to encourage a latent representation that promotes consistency in the representation between views of a CT scan (i.e., slices with same label in a patient) while also promoting inconsistency with representations from slices with other labels.
- Formulated as:

$$\ell = -\sum_{p \in \mathcal{P}} \frac{1}{|\mathcal{P}| - 1} \sum_{p' \in \mathcal{P} \setminus \{p\}} \log \frac{\exp(\langle \mathbf{z}_p, \mathbf{z}_{p'} \rangle)}{\sum_{n \in \mathcal{N}} \exp(\langle \mathbf{z}_p, \mathbf{z}_n \rangle^2)} \qquad (2)$$

  - Numerator is a proxy for MI between views that should be consistent
  - Denominator is sort of a proxy for MI between views that should be inconsistent.

- Impose structure on latent representation such that representation makes sense to humans (e.g., return a representation that yields the region(s) of anomalies as hot in image space). Looked for neural net architecture, developed a new loss.

[1] Oord, Aaron van den, Yazhe Li, and Oriol Vinyals. "Representation learning with contrastive predictive coding." *arXiv preprint arXiv:1807.03748* (2018).

Approach: SSL Architecture

- Architecture: Semi-supervised learning

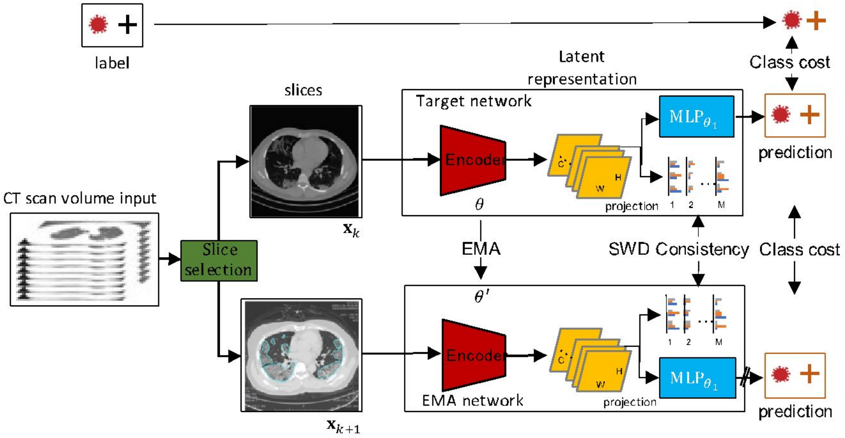

Figure 7: Schematic of semi-supervised approach.

- Deep neural-nets for latent representations
- Dual frameworks that allow pairwise comparisons between representations of examples.
- Induces some form of competition between dual networks.
- To discriminates anomaly from normal tissue.
- One of the benefits is that it gives some control about what should be similar and what shouldn't.
- Properties of exponential averaging

Experiments

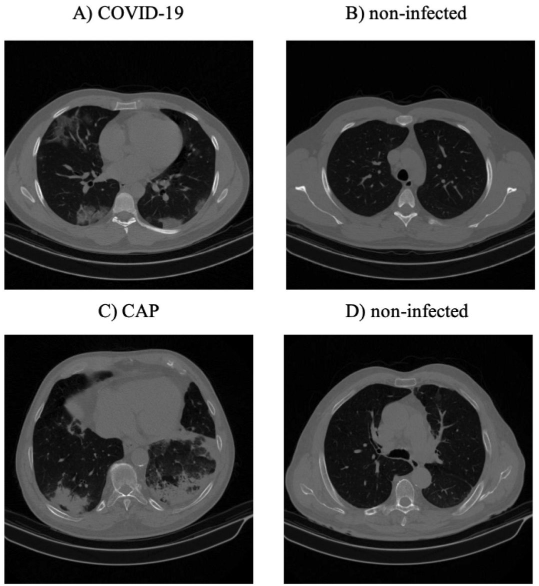

Figure 6: Examples of COVID-19, CAP and non-infected CT slices.

- Dataset [1]:
  - 2021 ICASSP COVID19 Challenge.
    - Task: Classify CT scans into COVID19, Community acquired pneumonia (CAP) and normal.
    - Experienced thoracic radiologists provided slice level labels for a small subset of the training set (i.e., 4990/46000) and also rRT-PCR test labels where provided for each patient (307 patients).
    - Training 307, validation 98, test1, test2, test3, test4 disjoint CT scans.
    - CT scans acquired by different scanners in multiple medical centers, using different settings for slice thickness, and exposure time.

[1] Afshar, P., Heidarian, S., Enshaei, N. *et al.* COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. *Sci Data* **8,** 121 (2021). https://doi.org/10.1038/s41597-021-00900-3

Experiments

- Trained the network module to learn latent representation only with labeled CT slices data
- Trained the SSL dual neural net to refine representation now also including labeled/unlabeled CT slices batches.
- After several epochs we trained linear head independently (freezing latent representation networks) using only labeled examples.
- Preliminary performance was evaluated in the 4 test sets.
- All experiments were performed on Darwin using PyTorch.

Preliminary Results

Table 1: Performance comparison

	Baseline [1]	Labeled only (ours)	Labeled/Unlabeled (ours)
Partition	Acc	Acc	Acc
test1	0.867	0.919	0.945
test2	0.911	0.946	0.968
test3	0.867	0.930	0.971
test4	0.833	0.902	0.913

$$\mathrm{Acc} = \frac{\sum_i c_{i,i}}{\sum_{i,j} c_{i,j}}$$

[1] S. Xue and C. Abhayaratne, "Covid-19 Diagnostic Using 3d Deep Transfer Learning for Classification of Volumetric Computerised Tomography Chest Scans," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8573-8577, doi: 10.1109/ICASSP39728.2021.9414947.

Conclusion

- What we completed…
  - Learning latent representations labeled data.
    - Developed an objective function for learning.
    - Representation is interpretable
  - Dual framework for labeled/unlabeled

- Working on…
  - Boosting

END

Boosting in Deep Semi-supervised Learning with Multi-view Redudant Information

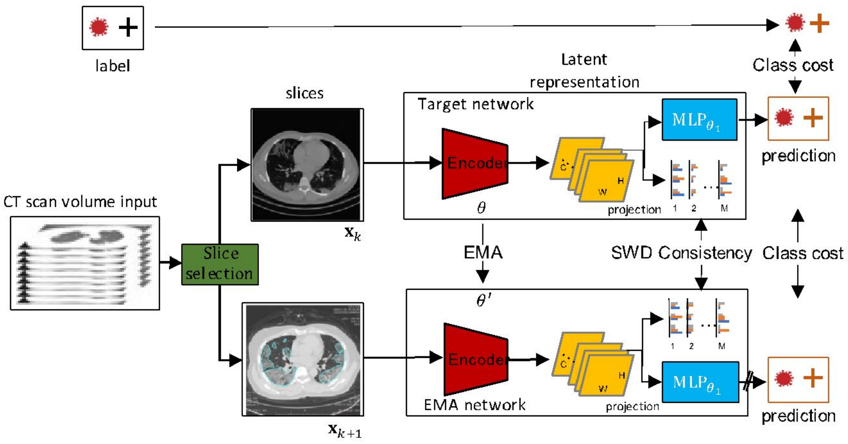

Figure 1: Schematic of semi-supervised approach.

Project Description:

- Our work addresses the problems of label scarcity by leveraging labeled/unlabeled multi-view observations.

Project Outcomes:

- ***A framework leveraging labeled/unlabeled and multi-view data to extract latent representations.***
- ***Implementation of the proposed approach***

- Working on summarizing our work in a manuscript for IEEE ICASSP due Oct 2, 2021.

PI: Juan Castorena
Total Project Budget: 60K
ISTI Focus Area: Computational and Data Integrity

Los Alamos NATIONAL LABORATORY