Semi-Supervised Learning for Sparsely-Labeled Sequential Data: Application to Healthcare Video Processing
Supplementary Materials

Florian Dubost*, Erin Hong*, Siyi Tang
Nandita Bhaskhar, Christopher Lee-Messer, and Daniel Rubin

*Stanford University

{floriandubost1, erin.hong17}@gmail.com, {cleemess, rubin}@stanford.edu

Table 1. Statistics of the full hospital video dataset.
| Label   | Nbr of windows | Average length (sec) | Median length (sec) | Standard Deviation (sec) | Total length (sec) | Number of patients |
|---------|----------------|----------------------|---------------------|--------------------------|-------------------|-------------------|
| Suctioning | 45             | 14.36                | 10                  | 14.46                    | 646               | 20                |
| Chewing  | 15             | 89.07                | 45                  | 91.23                    | 1336              | 12                |
| Rocking  | 21             | 54.67                | 24                  | 71.08                    | 1148              | 10                |
| Cares    | 44             | 88.66                | 46                  | 170.59                   | 3901              | 23                |
| Patting  | 33             | 38.64                | 21                  | 40.49                    | 1275              | 9                 |
| All      | 158            | 52.57                | 25                  | 104.47                   | 8306              | 59                |

Table 2. Statistics of the training split of the hospital video dataset.
| Label   | Nbr of windows | Average length (sec) | Median length (sec) | Standard Deviation (sec) | Total length (sec) | Number of patients |
|---------|----------------|----------------------|---------------------|--------------------------|-------------------|-------------------|
| Suctioning | 12             | 20.92                | 16                  | 14.92                    | 251               | 5                 |
| Chewing  | 5              | 147.4                | 130                 | 98.04                    | 737               | 4                 |
| Rocking  | 6              | 92.33                | 56                  | 97.53                    | 554               | 2                 |
| Cares    | 18             | 122.56               | 41                  | 255.59                   | 2206              | 8                 |
| Patting  | 10             | 42.2                 | 18                  | 56                       | 422               | 3                 |

Table 3. Statistics of the validation split of the hospital video dataset.
| Label   | Nbr of windows | Average length (sec) | Median length (sec) | Standard Deviation (sec) | Total length (sec) | Number of patients |
|---------|----------------|----------------------|---------------------|--------------------------|-------------------|-------------------|
| Suctioning | 18             | 13.5                 | 11                  | 10.96                    | 243               | 7                 |
| Chewing  | 6              | 44.33                | 20                  | 42.13                    | 266               | 5                 |
| Rocking  | 8              | 45.75                | 24                  | 53.7                     | 366               | 4                 |
| Cares    | 15             | 75.33                | 65                  | 50.24                    | 1130              | 10                |
| Patting  | 17             | 31.41                | 17                  | 28.8                     | 534               | 4                 |

Table 4. Statistics of the testing split of the hospital video dataset.
| Label   | Nbr of windows | Average length (sec) | Median length (sec) | Standard Deviation (sec) | Total length (sec) | Number of patients |
|---------|----------------|----------------------|---------------------|--------------------------|-------------------|-------------------|
| Suctioning | 15             | 10.13                | 7                   | 15.88                    | 152               | 8                 |
| Chewing  | 4              | 83.25                | 44                  | 96.14                    | 333               | 3                 |
| Rocking  | 7              | 32.57                | 17                  | 43.74                    | 228               | 4                 |
| Cares    | 11             | 51.36                | 34                  | 50.54                    | 565               | 5                 |
| Patting  | 6              | 53.17                | 52                  | 32.53                    | 319               | 2                 |
Table 5. Results on the HMDB51 dataset. Brushing hair vs. draw sword. We observe an early positive contribution of N with a late and progressively increasing negative contribution (same as CIFAR) with $\alpha_1 = 3$, $\beta_1 = 1$, $\alpha_2 = 0.5$, and $\beta_2 = 9$.

Table 6. Recall for the CIFAR-10 dataset. Means (standard deviations) are computed for 10 repetitions of the experiments with different random initialization of the weights. Best statistically significant results (p-value < 0.05) are highlighted in bold.

Table 7. F1-scores for the CIFAR-10 dataset. Means (standard deviations) are computed for 10 repetitions of the experiments with different random initialization of the weights. Best statistically significant results (p-value < 0.05) are highlighted in bold.

Table 8. AUC for the CIFAR-10 dataset. Means (standard deviations) are computed for 10 repetitions of the experiments with different random initialization of the weights. Best statistically significant (p-value < 0.05) results are highlighted in bold.
| Task:   | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| car vs. |       |       |       |       |       |       |       |       |       |
| airplane | 75.6 (3.6) | 81.2 (4.0) | **84.6 (1.4)** | 79.7 (2.5) | 77.2 (4.1) | 78.1 (2.1) | 82.7 (2.2) | 75.5 (3.7) | 68.1 (4.9) |
| bird    | **72.7 (2.6)** | 63.0 (6.8) | 68.4 (2.5) | 69.0 (2.3) | 70.5 (3.1) | 70.9 (3.3) | 71.3 (2.7) | 67.7 (1.5) | 70.4 (2.2) |
| cat     | 57.1 (1.1) | 56.0 (1.1) | 63.3 (2.4) | 65.0 (2.6) | 62.5 (1.5) | **65.1 (1.7)** | 56.3 (1.1) | 57.2 (2.0) | 59.2 (2.8) |
| deer    | 62.1 (3.8) | 64.4 (3.7) | 61.6 (2.8) | **66.6 (2.3)** | 63.9 (1.9) | 61.7 (1.1) | 60.6 (1.7) | 64.2 (1.5) | 64.2 (2.0) |
| dog     | 60.6 (1.5) | 57.8 (1.7) | **62.8 (2.0)** | 59.5 (1.1) | 59.4 (1.9) | 60.7 (3.2) | 57.4 (1.8) | 61.5 (1.2) | 61.5 (2.5) |
| frog    | 49.9 (1.6) | 52.1 (0.7) | 57.4 (2.9) | 53.8 (2.3) | 53.6 (1.9) | **61.3 (4.1)** | 53.5 (0.8) | 51.5 (1.7) | 48.9 (1.1) |
| horse   | 56.6 (2.5) | 52.7 (1.6) | 54.6 (0.9) | 56.2 (3.0) | 53.4 (1.8) | 63.0 (5.7) | **70.6 (3.1)** | 68.3 (4.2) | 55.9 (2.0) |
| ship    | 73.9 (0.6) | 79.5 (2.3) | **82.2 (2.4)** | 75.4 (1.1) | 77.8 (1.0) | 76.5 (4.1) | 76.1 (2.3) | 74.2 (2.0) | 76.0 (2.4) |
| truck   | 58.6 (1.2) | 68.5 (7.1) | 58.5 (3.5) | 60.8 (14.1) | 52.0 (2.6) | 65.2 (4.4) | 47.1 (0.8) | 65.2 (4.9) | **75.5 (3.9)** |

Table 9. **Precision for the CIFAR-10 dataset.** Means (standard deviations) are computed for 10 repetitions of the experiments with different random initialization of the weights. Best statistically significant (p-value < 0.05) results are highlighted in bold.