# A BioBART-v2

## A.1 Implementation Details

Here, we briefly introduce the implementation details of BioBART-v2. To make the best of the general domain knowledge and save the computational resource, for BioBART-v2, we apply Domain-Adaptive Pre-Training (DAPT) on PubMed abstracts, which contains around 41 GB of biomedical research paper abstracts and is the same corpus used to train BioBART. Compared to BioBART, the main difference is in BioBART-v2 we use a cross-domain vocabulary which comprises of 85,401 tokens, while BioBART inherits the general domain vocabulary of BART which comprises of 50,256 tokens.

We generate the cross-domain vocabulary $\mathcal{V}_{\text{cro}}$ by mixing the origin general domain vocabulary $\mathcal{V}_{\text{gen}}$ and the generated biomedical domain vocabulary $\mathcal{V}_{\text{bio}}$ on the PubMed pre-training corpus. $\mathcal{V}_{\text{bio}}$ generates 60,000 tokens of $\mathcal{V}_{\text{bio}}$ based on the PubMed corpus by Byte Pair Encoding (BPE). The $\mathcal{V}_{\text{cro}}$ is generated by taking the union of $\mathcal{V}_{\text{bio}}$ and $\mathcal{V}_{\text{gen}}$ and results in 85,401 tokens, since many BPE tokens in $\mathcal{V}_{\text{bio}}$ and $\mathcal{V}_{\text{gen}}$ are the same. As we pre-train BioBART-v2 based on BART and new tokens introduce extra parameters to the embedding layers, we initialize the embeddings of the new tokens by first decompose the new tokens into the sub-tokens in $\mathcal{V}_{\text{gen}}$ and then take the average of the sub-token embeddings as the initialized weights.

We use the same experiment setups for DAPT and fine-tuning on downstream tasks as BioBART, except that we extend the pre-training update steps to 200k. The model is pre-trained on DGX with 8 NVIDIA A100 40G with DeepSpeed (Rajbhandari et al., 2020).

## A.2 Results

| Model | ShARe13 F1 | ShARe14 F1 | CADEC F1 | GENIA F1 |
|---|---|---|---|---|
| BART BASE | 76.63 | 77.87 | 68.37 | 78.06 |
| BioBART BASE | 78.78 | 79.17 | 68.39 | 78.43 |
| BioBART-v2 BASE | **79.54** | **79.88** | **70.42** | **79.08** |
| BART LARGE | 79.69 | 80.34 | **70.64** | 78.93 |
| BioBART LARGE | 80.75 | 80.41 | 70.53 | 79.93 |
| BioBART-v2 LARGE | **80.78** | **81.13** | 70.58 | **80.00** |

Table 8: The main result on NER tasks.

| Model | MedMentions Recall@1/@5 | BC5CDR Recall@1/@5 | NCBI Recall@1/@5 | COMETA Recall@1/@5 | AAP Recall@1/@5 |
|---|---|---|---|---|---|
| BART BASE | 69.77/84.59 | 91.56/94.89 | 88.54/95.31 | 78.34/87.40 | 86.37/94.29 |
| BioBART BASE | **71.15/86.22** | **93.01**/95.59 | **89.27/95.31** | **79.63/88.64** | **87.51/94.92** |
| BioBART-v2 BASE | 71.04/85.78 | 92.63/**95.70** | 88.33/95.00 | 79.05/88.37 | 87.36/94.86 |
| BART LARGE | 71.49/84.95 | 92.48/95.26 | 90.21/95.52 | 80.70/88.65 | 88.79/**96.59** |
| BioBART LARGE | **71.78/85.42** | 93.26/**95.74** | 89.90/**95.63** | 81.77/88.87 | **89.40**/95.76 |
| BioBART-v2 LARGE | 70.80/84.33 | **93.33**/95.73 | **90.31**/95.21 | **82.02/89.98** | 89.09/95.74 |

Table 9: The main results on Entity Linking tasks.

| | Covid19-Dialogue | | | | |
|---|---|---|---|---|---|
| **Model** | Rouge-1 | Rouge-2 | Rouge-L | BLEU | BERTscore |
| BART BASE | 27.24 | 12.31 | 25.66 | 10.36 | **0.852** |
| BioBART BASE | 28.14 | **12.77** | 26.32 | **11.40** | 0.849 |
| BioBART-v2 BASE | **29.25** | 12.59 | **27.28** | **11.40** | 0.850 |
| BART LARGE | **29.02** | 12.08 | 26.93 | 10.96 | 0.852 |
| BioBART LARGE | 28.81 | **13.79** | **26.96** | 12.05 | 0.850 |
| BioBART-v2 LARGE | 28.85 | 13.48 | 26.90 | **12.55** | **0.853** |

Table 10: The main results on Dialogue System task.

| | iCliniq | | HealthCareMagic | | MEDIQA-QS | |
|---|---|---|---|---|---|---|
| **Model** | Rouge-1/2/L | BERTscore | Rouge-1/2/L | BERTscore | Rouge-1/2/L | BERTscore |
| BART BASE | **61.43/48.68/59.71** | 0.941 | **46.81/26.19/44.34** | 0.918 | 28.82/10.99/26.99 | 0.896 |
| BioBART BASE | 61.07/48.47/59.42 | 0.941 | 46.67/26.03/44.11 | 0.918 | **30.12/11.28**/27.44 | 0.898 |
| BioBART-v2 BASE | 61.11/48.39/59.26 | **0.945** | 46.43/25.89/43.90 | **0.923** | 29.89/10.98/**27.57** | **0.907** |
| BART LARGE | 59.87/47.01/58.12 | 0.938 | **47.24/26.54/44.68** | 0.919 | 29.97/10.64/28.41 | 0.901 |
| BioBART LARGE | 60.32/47.98/58.69 | 0.940 | 46.54/26.14/44.23 | 0.919 | **31.97/12.39/29.70** | 0.903 |
| BioBART-v2 LARGE | **61.97/48.95/60.05** | **0.946** | 46.43/26.17/44.18 | **0.924** | 30.47/12.33/29.10 | **0.911** |

| | MEDIQA-MAS | | MEDIQA-ANS(Pages) | | MeQSum | |
|---|---|---|---|---|---|---|
| **Model** | Rouge-1/2/L | BERTscore | Rouge-1/2/L | BERTscore | Rouge-1/2/L | BERTscore |
| BART BASE | 31.63/9.98/27.85 | 0.859 | 19.10/6.77/16.90 | 0.851 | 52.93/35.79/50.46 | 0.927 |
| BioBART BASE | 32.90/11.28/29.26 | 0.861 | 18.97/7.46/16.77 | 0.850 | 53.75/36.50/51.27 | 0.929 |
| BioBART-v2 BASE | **33.36/12.26/29.94** | **0.862** | **19.75/7.49/17.46** | **0.858** | **54.31/37.22/51.93** | **0.933** |
| BART LARGE | 29.32/9.00/26.14 | 0.857 | 21.52/9.31/19.15 | 0.853 | 53.68/36.80/51.05 | 0.928 |
| BioBART LARGE | 30.60/**10.37**/27.04 | **0.861** | 21.58/9.34/19.18 | 0.857 | **55.61/38.11/53.15** | 0.933 |
| BioBART-v2 LARGE | **30.75**/9.27/**27.61** | **0.861** | **21.85/9.83/19.54** | **0.863** | 55.17/**38.11**/52.94 | **0.936** |

Table 11: The main results on Summarization tasks.