Supplementary Materials of the Paper "Parameter Selection: Why We Should Pay More Attention to It"

Jie-Jyun Liu[1], Tsung-Han Yang[1], Si-An Chen[1,2], and Chih-Jen Lin[2]

[1]ASUS Intelligent Cloud Services
[2]National Taiwan University

July 15, 2021

1 Additional Experimental Results

In Table I and Table II, we provide details of configurations that achieve top validation results. Clearly, quite a few parameter settings can lead to optimal results. Thus, even if only a casual search is conducted, better results than those in Mullenbach et al. (2018) may be obtained.

Table I: CNN parameter-selection results on MIMIC-III-50, ordered according to validation P@5

	Validation performance			Test performance			Parameters selected			
Seed	Macro-F1	Micro-F1	P@5	Macro-F1	Micro-F1	P@5	d_c	k	q	η
1337	0.6265	0.6714	0.6378	0.6185	0.6641	0.6370	550	6	0.6000	0.0001
	0.6201	0.6656	0.6376	0.6096	0.6585	0.6346	550	8	0.6000	0.0001
	0.6106	0.6674	0.6362	0.6019	0.6617	0.6345	550	4	0.6000	0.0001
	0.6068	0.6521	0.6361	0.6053	0.6492	0.6325	450	4	0.6000	0.0003
	0.6094	0.6669	0.6356	0.6049	0.6616	0.6312	450	4	0.6000	0.0001
1331	0.6078	0.6666	0.6412	0.6002	0.6621	0.6378	550	4	0.4000	0.0001
	0.6163	0.6686	0.6404	0.6089	0.6647	0.6337	450	4	0.6000	0.0001
	0.6186	0.6726	0.6379	0.6161	0.6671	0.6378	550	4	0.6000	0.0001
	0.5810	0.6543	0.6376	0.5780	0.6484	0.6304	550	4	0.2000	0.0001
	0.6092	0.6643	0.6374	0.6000	0.6554	0.6319	450	10	0.6000	0.0001
42	0.6232	0.6710	0.6426	0.6173	0.6656	0.6376	550	6	0.6000	0.0001
	0.6069	0.6630	0.6356	0.5963	0.6559	0.6305	550	6	0.4000	0.0001
	0.6156	0.6527	0.6353	0.6096	0.6467	0.6335	550	6	0.8000	0.0001
	0.6207	0.6667	0.6348	0.6133	0.6616	0.6320	550	8	0.6000	0.0001
	0.6153	0.6652	0.6347	0.6056	0.6578	0.6304	450	6	0.6000	0.0001

Table II: CAML parameter-selection results on MIMIC-III-50, ordered according to validation P@5

	Validation performance			Test performance			Parameters selected			
Seed	Macro-F1	Micro-F1	P@5	Macro-F1	Micro-F1	P@5	d_c	k	q	η
1337	0.6329	0.6794	0.6506	0.6352	0.6797	0.6514	250	10	0.8000	0.0001
	0.6435	0.6884	0.6497	0.6405	0.6874	0.6471	250	10	0.8000	0.0003
	0.6385	0.6863	0.6487	0.6393	0.6829	0.6501	350	8	0.8000	0.0001
	0.6396	0.6880	0.6478	0.6407	0.6871	0.6507	450	10	0.8000	0.0003
	0.6450	0.6912	0.6474	0.6467	0.6911	0.6527	550	10	0.8000	0.0003
1331	0.6322	0.6790	0.6463	0.6229	0.6791	0.6502	350	10	0.8000	0.0001
	0.6233	0.6801	0.6463	0.6335	0.6803	0.6519	250	10	0.8000	0.0001
	0.6210	0.6748	0.6462	0.6298	0.6795	0.6510	550	10	0.8000	0.0001
	0.6466	0.6884	0.6462	0.6349	0.6864	0.6492	250	6	0.8000	0.0001
	0.6247	0.6807	0.6460	0.6335	0.6829	0.6514	450	10	0.8000	0.0001
42	0.6263	0.6828	0.6486	0.6278	0.6824	0.6522	150	8	0.8000	0.0001
	0.6330	0.6841	0.6478	0.6269	0.6804	0.6531	450	10	0.8000	0.0001
	0.6350	0.6816	0.6474	0.6249	0.6803	0.6489	250	10	0.8000	0.0001
	0.6463	0.6922	0.6472	0.6435	0.6891	0.6492	550	8	0.8000	0.0003
	0.6482	0.6935	0.6470	0.6450	0.6894	0.6515	550	10	0.8000	0.0003

2 Parameter Selection in Past Works

The current work concerns about the need of parameter selection. It is important to check if the later works mentioned in the paper have performed parameter selection. In Table III we show descriptions given in these works. Interestingly, top-performing approaches in Table 2 of the paper tend to be those that have paid more attention to parameter selection; see, for example, LAAT (Vu et al., 2020), MSATT-KG (Xie et al., 2019), and HyperCore (Cao et al., 2020).

Table III: Description about parameter selection in works that have considered the date MIMIC-III-50

	Hyperparameter search algorithms
MVC-LDA, MVC-RLDA (Sadoughi et al., 2018)	HyperBand
MultiResCNN (Li and Yu, 2020)	hyperparameters values are chosen empirically or following CAML
DCAN (Ji et al., 2020)	random search
LAAT (Vu et al., 2020)	grid search over LSTM hidden size u: 128, 256, 384, 512 and the projection size: 128, 256, 384, 512
LEAM (Wang et al., 2018)	study the impact of window size
MSATT-KG (Xie et al., 2019)	hyper-parameters are fine-tuned on the validation set
HyperCore (Cao et al., 2020)	grid search

References

P. Cao, Y. Chen, K. Liu, J. Zhao, S. Liu, and W. Chong. HyperCore: Hyperbolic and co-graph representation for automatic ICD coding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3105–3114, 2020. doi: 10.18653/v1/2020.acl-main.282.

S. Ji, E. Cambria, and P. Marttinen. Dilated convolutional attention network for medical code assignment from clinical text. In *Proceedings of the Third Clinical Natural Language Processing Workshop*, pages 73–78, 2020. doi: 10.18653/v1/2020.clinicalnlp-1.8.

F. Li and H. Yu. ICD coding from clinical text using multi-filter residual convolutional neural network. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI)*, pages 8180–8187, 2020. doi: 10.1609/aaai.v34i05.6331.

J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein. Explainable prediction of medical codes from clinical text. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1101–1111, 2018. doi: 10.18653/v1/N18-1100.

N. Sadoughi, G. P. Finley, J. Fone, V. Murali, M. Korenevski, S. Baryshnikov, N. Axtmann, M. Miller, and D. Suendermann-Oeft. Medical code prediction with multi-view convolution and description-regularized label-dependent attention, 2018. URL `http://arxiv.org/abs/1811.01468`. arXiv preprint arXiv:1811.01468.

T. Vu, D. Q. Nguyen, and A. Nguyen. A label attention model for ICD coding from clinical text. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3335–3341, 2020. doi: 10.24963/ijcai.2020/461.

G. Wang, C. Li, W. Wang, Y. Zhang, D. Shen, X. Zhang, R. Henao, and L. Carin. Joint embedding of words and labels for text classification. In *Proceedings of the 56th Annual Meeting of the Association of Computational Linguistics (ACL)*, 2018. doi: 10.18653/v1/P18-1216.

X. Xie, Y. Xiong, P. S. Yu, and Y. Zhu. EHR coding with multi-scale feature attention and structured knowledge graph propagation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 649–658, 2019. doi: 10.1145/3357384.3357897.