Table 6: Comparison of training time based on GPU and CPU. (Test evaluations for all the models were performed by CPU.)

Types	Model	Datasets	1-Batch Time (sec.)	Train Time (hr.)	Test Time (hr.)	Macro-F_1	Micro-F_1
CPU	DGCNN-3	RCV1	1.416	12	2.5	0.4320	0.7611
GPU	DGCNN-3	RCV1	0.767	6	2.5	0.4322	0.7611
CPU	DGCNN-6	NYTimes	1.437	168	0.6	0.2985	0.6566
GPU	DGCNN-6	NYTimes	0.401	48	0.6	0.2991	0.6566

Table 7: Comparison of results on NYtimes dataset.

Models	Classes	Macro-F_1	Micro-F_1
SVM	2,318	0.2158	0.5217
HSVM	2,318	0.2187	0.5213
TD-SVM	2,318	0.2249	0.5404
HR-SVM	2,318	0.2571	0.6123
HLSTM	2,318	0.2141	0.5271
HAN	2,318	0.2217	0.5395
RCNN	2,318	0.2019	0.5311
XML-CNN	2,318	0.2001	0.5292
DCNN-3	2,318	0.2471	0.5793
DCNN-6	2,318	0.2669	0.6055
DGCNN-1	2,318	0.2147	0.5195
DGCNN-3	2,318	0.2791	0.6030
DGCNN-6	2,318	0.2991	0.6566
HR-DGCNN-1	2,318	0.2209	0.5293
HR-DGCNN-3	2,318	0.2807	0.6146
HR-DGCNN-6	2,318	**0.2995**	**0.6612**

Table 8: Comparsion of training time and results on NY-Times dataset. The evaluations for stand-alone (Native) and recursive hierarchical segmentation (RHS) programs were performed by DGCNN-6 and GPU.

Model	Macro-F_1	Micro-F_1	Training Time (days)
Native	0.2993	0.6481	15
RHS	0.2991	0.6566	2

Table 9: Number of parameters (in millions).

Dataset	1 Layer	3 Layers	6 Layers
RCV1	3.73	4.99	9.57
NYTimes	3.92	5.21	12.32

KDD. 257–265.

[14] Siddharth Gopal and Yiming Yang. 2015. Hierarchical Bayesian inference and recursive regularization for large-scale classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 3 (2015), 18.

[15] Siddharth Gopal, Yiming Yang, Bing Bai, and Alexandru Niculescu-Mizil. 2012. Bayesian models for large-scale hierarchical classification. In *NIPS*. 2411–2419.

[16] Mikael Henaff, Joan Bruna, and Yann LeCun. 2015. Deep Convolutional Networks on Graph-Structured Data. *CoRR* abs/1506.05163 (2015). http://arxiv.org/abs/1506.05163

[17] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[18] Thorsten Joachims. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In *ECML*. 137–142.

[19] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In *ACL*. 655–665.

[20] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *EMNLP*. 1746–1751.

[21] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*. 1097–1105.

[23] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent Convolutional Neural Networks for Text Classification. In *AAAI*. 2267–2273.

[24] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep Learning. *Nature* 521 (2015), 436–444.

[25] Yann Lecun, LÃŢon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*. 2278–2324.

[26] Omer Levy and Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. In *NIPS*. 2177–2185.

[27] David D Lewis, Yiming Yang, Tony G Rose, and Fan Li. 2004. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research* 5, Apr (2004), 361–397.

[28] Xin Li and Dan Roth. 2002. Learning question classifiers. In *ACL*. 1–7.

[29] Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. 2017. Deep Learning for Extreme Multi-label Text Classification. In *SIGIR*. 115–124.

[30] Tie-Yan Liu, Yiming Yang, Hao Wan, Hua-Jun Zeng, Zheng Chen, and Wei-Ying Ma. 2005. Support vector machines classification with a very large-scale taxonomy. *ACM SIGKDD Explorations Newsletter* 7, 1 (2005), 36–43.

[31] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *Computer Science* (2013).

[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*. 3111–3119.

[33] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. 2016. Learning Convolutional Neural Networks for Graphs. In *ICML*. 2014–2023.

[34] François Rousseau, Emmanouil Kiagias, and Michalis Vazirgiannis. 2015. Text categorization as a graph classification problem. In *ACL*, Vol. 15. 107.

[35] Evan Sandhaus. 2008. The New York Times Annotated Corpus LDC2008T19. In *Linguistic Data Consortium*.

[36] Sam Scott and Stan Matwin. 1999. Feature Engineering for Text Classification. In *ICML*. 379–388.

[37] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *ICLR*.

[38] Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011. Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection. In *NIPS*. 801–809.

[39] Aixin Sun and Ee-Peng Lim. 2001. Hierarchical Text Classification and Evaluation. In *ICDM*. 521–528.

[40] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. In *EMNLP*. 1422–1432.

[41] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research* 6, Sep (2005), 1453–1484.

[42] Wei Wang, Diep Bich Do, and Xuemin Lin. 2005. Term graph model for text classification. In *International Conference on Advanced Data Mining and Applications*. Springer, 19–30.

[43] Lin Xiao, Dengyong Zhou, and Mingrui Wu. 2011. Hierarchical classification via orthogonal transfer. In *ICML*. 801–808.

[44] Gui-Rong Xue, Dikan Xing, Qiang Yang, and Yong Yu. 2008. Deep classification in large-scale text hierarchies. In *SIGIR*. 619–626.

[45] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J. Smola, and Eduard H. Hovy. 2016. Hierarchical Attention Networks for Document Classification. In *NAACL–HLT*. 1480–1489.

[46] Min-Ling Zhang and Zhi-Hua Zhou. 2014. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* 26, 8 (2014), 1819–1837.

[47] Wenjie Zhang, Liwei Wang, Junchi Yan, Xiangfeng Wang, and Hongyuan Zha. 2017. Deep Extreme Multi-label Learning. *CoRR* abs/1704.03718 (2017).

[48] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In *NIPS*. 649–657.