参考文献

第 1 章

[1] Turing A M. On computable numbers with an application to the Entscheidungsproblem[J]. Journal Of Mathematics, 1936, 58(5): 345-363.

[2] https://zh.wikipedia.org/wiki/图灵机

[3] https://en.wikipedia.org/wiki/Warren_Sturgis_McCulloch

[4] https://zh.wikipedia.org/wiki/动作电位

[5] Schmidhuber J. Annotated history of modern AI and Deep learning[J]. arXiv preprint, 2022, arXiv:2212.11279.

[6] https://zh.wikipedia.org/wiki/电子数值积分计算机

[7] Shannon C E. A mathematical theory of communication[J]. The Bell system technical journal, 1948, 27(3): 379-423.

[8] Turing A M. Computing machinery and intelligence[M]. Springer Netherlands, 2009.

[9] https://en.wikipedia.org/wiki/Dartmouth_workshop

[10] https://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)

[11] https://zh.m.wikipedia.org/wiki/艾伦·纽厄尔

[12] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.

[13] Minsky M, Papert S. An introduction to computational geometry[J]. Cambridge tiass, HIT, 1969, 479(480): 104.

[14] https://en.wikipedia.org/wiki/Lighthill_report

[15] https://en.wikipedia.org/wiki/Expert_system

[16] Ledley RS, and Lusted LB. Reasoning foundations of medical diagnosis. Science, 1959, 130(3366): 9–21.

[17] Weiss SM, Kulikowski CA, Amarel S, Safir A. A model-based method for computer-aided medical decision-making. Artificial Intelligence, 1978, 11(1–2): 145–172.

[18] https://en.wikipedia.org/wiki/Dendral

[19] Lemarechal C. Cauchy and the Gradient Method[J]. Doc Math Extra, 2012, 2:251-254.

[20] Hadamard J. Mémoire sur le problèmed'analyserelatif à l'équilibre des plaques élastiquesencastrées[M]. Imprimerienationale, 1908.

[21] Robbins H, Monro S. A Stochastic Approximation Method. The Annals of Mathematical Statistics. 1951, 22(3):400.

[22] Linnainmaa S. Taylor expansion of the accumulated rounding error[J]. BIT Numerical Mathematics, 1976, 16(2): 146-160.

[23] Werbos P J. Applications of advances in nonlinear sensitivity analysis[C] Berlin, Heidelberg: Springer Berlin Heidelberg, 2005(1): 762-770.

[24] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.

[25] Fukushima K. Neural network model for a mechanism of pattern recognition unaffected by shift in position—Neocognitron[J]. Trans. IECE, 1979, (62)10:658-665.

[26] Fukushima K. Visual feature extraction by a multilayered network of analog threshold elements[J]. IEEE Transactions on Systems Science and Cybernetics. 1969, 5(4): 322-333.

[27] LeCun Y, Boser B, Denker J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J], Neural Computation, 1989, 1(4): 541-551.

[28] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[29] Ciresan D C, Meier U, Masci J, et al. Flexible, High Performance Convolutional Neural Networks for Image Classification[C]. International Joint Conference on Artificial Intelligence, 22th International Joint Conference on Artificial Intelligence, 2011.

[30] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks. 25th Neural Information Processing Systems, 2012.

[31] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778.

[32] Ising E. BeitragzurTheorie des Ferromagnetismus[J]. Z. Phys. 1925, 1(1): 253-258.

[33] Amari S I. Learning patterns and pattern sequences by self-organizing nets of threshold elements[J]. IEEE Transactions, 1972, 1197-1206.

[34] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the national academy of sciences, 1982, 79(8): 2554-2558.

[35] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.

[36] Jordan M I. Serial order: A parallel distributed processing approach[M] Advances in psychology. North-Holland, 1997, 121: 471-495.

[37] Werbos P J. Generalization of backpropagation with application to a recurrent gas

market model[J]. Neural networks, 1988, 1(4): 339-356.

[38] Werbos P J. Backpropagation through time: what it does and how to do it[J]. Proceedings of the IEEE, 1990, (78)10:1550-1560.

[39] Rumelhart D E, Hinton G E, Williams R J. Learning Internal Representations by Error Propagation[J], Parallel Distributed Processing, 1986, (1)71: 599-607.

[40] Hochreiter S. UntersuchungenzudynamischenneuronalenNetzen[J]. Diploma thesis, TUM, 1991.

[41] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[42] https://en.wikipedia.org/wiki/Long_short-term_memory

[43] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint, 2014, arXiv:1406.1078.

[44] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.

[45] Thireou T, Reczko M. Bidirectional long short-term memory networks for predicting the subcellular localization of eukaryotic proteins[J]. IEEE/ACM transactions on computational biology and bioinformatics, 2007, 4(3): 441-446.

[46] Zhou C, Sun C, Liu Z, et al. A C-LSTM neural network for text classification[J]. arXiv preprint, 2015, arXiv:1511.08630.

[47] Li Y, Wang X, Xu P. Chinese Text Classification Model Based on Deep Learning[J]. Future Internet, 2018, 10(11):113.

[48] Kunfu Wang, Pengyi Zhang, et al. A Text Classification Method Based on the Merge-LSTM-CNN Model[C] Journal of Physics: Conference Series, 2020, 1646(1): 012110.

[49] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J].arXiv preprint, 2013, arXiv:1301.3781.

[50] Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation[C]. In Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing, 2014, 1532–1543.

[51] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, arXiv:1409.0473.

[52] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint, 2015, arXiv:1508.04025.

[53] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[54] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. arXiv preprint, 2016, arXiv:1609.08144.

[55] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2018, arXiv:1810.04805.

[56] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.

[57] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask

learners[J]. OpenAI blog, 2019, 1(8): 9.

[58] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[59] Matthew E Peters, Neumann M, et al. Deep Contextualized Word Representations[C]. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, 1: 2227–2237.

[60] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[J]. Advances in neural information processing systems, 27, 2014.

[61] Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report[J]. arXiv preprint, 2023, arXiv:2303.08774.

[62] https://en.wikipedia.org/wiki/LaMDA

[63] https://en.wikipedia.org/wiki/PaLM

[64] https://zh.wikipedia.org/wiki/Bard

[65] https://en.wikipedia.org/wiki/Gemini_(language_model)

[66] Bard 体验地址：https://bard.google.com/chat/

[67] 文心一言体验地址： https://yiyan.baidu.com/

[68] 通义千问体验地址： https://tongyi.aliyun.com/qianwen/

[69] 百川大模型体验地址：https://www.baichuan-ai.com/chat

[70] https://zh.wikipedia.org/wiki/艾伦·图灵

[71] https://zh.wikipedia.org/zh-cn/约翰·冯·诺伊曼

[72] https://zh.wikipedia.org/zh-cn/约翰·麦卡锡

[73] https://zh.wikipedia.org/zh-cn/克劳德·香农

[74] https://zh.wikipedia.org/wiki/沃伦·麦卡洛克

[75] https://en.wikipedia.org/wiki/Walter_Pitts

[76] https://en.wikipedia.org/wiki/John_Mauchly

[77] https://en.wikipedia.org/wiki/J._Presper_Eckert

[78] https://en.wikipedia.org/wiki/Allen_Newell

[79] https://en.wikipedia.org/wiki/Herbert_A._Simon

[80] https://en.wikipedia.org/wiki/Frank_Rosenblatt

[81] https://en.wikipedia.org/wiki/Marvin_Minsky

[82] https://en.wikipedia.org/wiki/Seymour_Papert

[83] https://en.wikipedia.org/wiki/Seppo_Linnainmaa

[84] https://en.wikipedia.org/wiki/Paul_Werbos

[85] https://en.wikipedia.org/wiki/David_Rumelhart

[86] https://en.wikipedia.org/wiki/Geoffrey_Hinton

[87] https://en.wikipedia.org/wiki/Yann_LeCun

[88] https://en.wikipedia.org/wiki/Yoshua_Bengio

[89] https://en.wikipedia.org/wiki/Alex_Krizhevsky

[90] https://en.wikipedia.org/wiki/Ilya_Sutskever

[91] https://en.wikipedia.org/wiki/Andrew_Ng

[92] https://en.wikipedia.org/wiki/Ronald_J._Williams

[93] https://en.wikipedia.org/wiki/Kunihiko_Fukushima

[94] https://www.linkedin.com/in/dan-c-49317a8b/

[95] https://kaiminghe.github.io/

[96] https://en.wikipedia.org/wiki/Ernst_Ising

[97] https://en.wikipedia.org/wiki/Wilhelm_Lenz

[98] https://en.wikipedia.org/wiki/John_Hopfield

[99] https://en.wikipedia.org/wiki/Jeffrey_Elman

[100] https://en.wikipedia.org/wiki/Michael_I._Jordan

[101] https://en.wikipedia.org/wiki/Sepp_Hochreiter

[102] https://en.wikipedia.org/wiki/J%C3%BCrgen_Schmidhuber

[103] https://en.wikipedia.org/wiki/Ashish_Vaswani

[104] https://en.wikipedia.org/wiki/Sam_Altman

[105] https://en.wikipedia.org/wiki/Demis_Hassabis

[106] https://zh.wikipedia.org/zh-cn/李飞飞

[107] https://en.wikipedia.org/wiki/Ian_Goodfellow

[108] https://en.wikipedia.org/wiki/Oriol_Vinyals

[109] https://en.wikipedia.org/wiki/Quoc_V._Le

[110] https://en.wikipedia.org/wiki/Jeff_Dean

[111] https://en.wikipedia.org/wiki/Greg_Brockman

[112] https://en.wikipedia.org/wiki/Torch_(machine_learning)

[113] https://en.wikipedia.org/wiki/Theano_(software)

[114] https://github.com/Theano/Theano

[115] https://mxnet.apache.org/

[116] https://en.wikipedia.org/wiki/Apache_MXNet

[117] https://caffe.berkeleyvision.org/

[118] https://zh.wikipedia.org/wiki/Caffe

[119] https://en.wikipedia.org/wiki/Keras

[120] https://keras.io/

[121] https://chainer.org/

[122] https://en.wikipedia.org/wiki/Chainer

[123] https://zh.wikipedia.org/zh-cn/TensorFlow

[124] https://www.tensorflow.org/

[125] https://www.paddlepaddle.org.cn/

[126] https://github.com/PaddlePaddle/Paddle

[127] https://zh.wikipedia.org/zh-cn/PyTorch

[128] https://pytorch.org/

第 2 章

[1] https://www.pytorch.org

[2] https://www.tensorflow.org

[3] https://mxnet.apache.org

[4] https://www.paddlepaddle.org.cn

[5] https://docs.nvidia.com/deeplearning/cudnn/pdf/cuDNN-Support-Matrix.pdf

[6] https://github.com/pytorch/pytorch#from-source

[7] https://docs.conda.io

[8] https://docs.conda.io/projects/miniconda/en/latest

[9] https://www.anaconda.com

[10]https://developer.nvidia.com/cuda-toolkit-archive

[11]https://pytorch.org/get-started/previous-versions

[12]https://download.pytorch.org/whl/torch_stable.html

[13]https://github.com/moon-hotel/GPT2-Chinese/tree/master

[14]https://www.jetbrains.com.cn/pycharm/download/?section=windows

[15]https://www.netsarang.com/en/xshell/

[16]https://mobaxterm.mobatek.net/

[17]https://termius.com/

[18] https://winscp.net/eng/index.php

第 3 章

[1] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[2] Yann LeCun, Deep Learning, Alfredo Canziani, NYU, Spring 2020.

[3] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. the Journal of machine Learning research, 2011, 12: 2825-2830.

[4] Hungyi L, MachineLearning, National Taiwan University, Spring, 2020.

[5] Hastie T, Tibshirani R, Friedman J, The Elements of Statistical Learning, Second Edition, Springer Series in Statistics, Jan 13, 2017.

[6] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.

[7] 赵申剑，黎彧君，符天凡，李凯 译，深度学习 [M]. 北京：人民邮电出版社, 2017.

[8] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. the Journal of machine Learning research, 2011, 12: 2825-2830.

[9] Sorower M S. A literature survey on algorithms for multi-label learning[J]. Oregon State University, Corvallis, 2010, 18(1): 25.

[10] Godbole S, Sarawagi S. Discriminative Methods for Multi-labeled Classification. 2004, Lecture Notes in Computer Science, 22–30.

[11] https://mmuratarat.github.io/2020-01-25/multilabel_classification_metrics

第 4 章

[1] 赵申剑, 黎彧君, 符天凡, 李凯 译, 深度学习 [M]. 北京：人民邮电出版社, 2017.

[2] 卷积定义：https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF

[3] 图 片 来 源 : https://stats.stackexchange.com/questions/208936/what-is-translation-invariance-in-computer-vision-and-convolutional-neural-netwo/208949

[4] 于建国, 卷积神经网络介绍. https://zhuanlan.zhihu.com/p/27642620

[5] Hungyi L, MachineLearning, National Taiwan University, Spring, 2020.

[6] https://cs231n.github.io/convolutional-networks/#norm

[7] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

[8] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint, 2014, arXiv:1408.5882.

[9] LV Z, XU J, ZHENG K, et al. LC-RNN: A Deep Learning Model for Traffic Speed Prediction [C] Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden. 2018.

[10] https://cs231n.github.io/convolutional-networks/#pool

[11] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11), 2278-2324.

[12] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] Advances in neural information processing systems, 2012, 1097-1105.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. International Conference on Learning Representations, 2015.

[15] Sermanet P, Eigen D, Zhang X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[C]. International Conference on Learning Representations, 2014.

[16]Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint, 2013, arXiv:1312.4400.

[17] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C] Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, 1-9.

[18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C] Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 770-778.

[19] Ioffe S, Szegedy C, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML, PMLR, 2015 37:448-456.

[20] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C] Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, 4700-4708.

[21] https://www.tensorflow.org/api_docs/python/tf/keras/applications/densenet/DenseNet121

[22] https://github.com/liuzhuang13/DenseNet

第 5 章

[1] https://pytorch.org/docs/master/tensorboard.html#torch-utils-tensorboard

[2] Saving And Laoding Models https://pytorch.org/tutorials/beginner/saving_loading_models.html

[3] https://www.tensorflow.org/tutorials/images/transfer_learning

[4] https://www.image-net.org/index.php

[5] https://zh.wikipedia.org/wiki/ImageNet

[6] CS231n Convolutional Neural Networks for Visual Recognition. https://cs231n.github.io/transfer-learning/

[7] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[8] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance

deep learning library[J]. Advances in neural information processing systems, 32, 2019.

第 6 章

[1] https://huggingface.co/docs/transformers/index

[2] https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)

[3] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.

[4] Ioffe S, Szegedy C, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML, PMLR, 2015, 37:448-456.

[5] Ba J L, Kiros J R, Hinton G E. Layer normalization[J]. arXiv preprint, 2016, arXiv:1607.06450.

[6] Wu Y, He K. Group normalization[C]//Proceedings of the European conference on computer vision (ECCV). 3-19, 2018.

[7] https://cocodataset.org/#home

[8] Polyak B T. Some methods of speeding up the convergence of iteration methods. USSR Computational Mathematics and Mathematical Physics, 1964, 4(5):1–17.

[9] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of machine learning research, 2011, 12(7).

[10] Zeiler M D. Adadelta: an adaptive learning rate method[J]. arXiv preprint, 2012, arXiv:1212.5701.

[11] Tieleman T, Hinton G E. Lecture 6.5 - RMSProp, COURSERA: Neural Networks for Machine Learning. Technical report, 2012.

[12] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint, 2014, arXiv:1412.6980.

[13] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.

[14] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]. Proceedings of the IEEE international conference on computer vision. 2015, 1026-1034.

第 7 章

[1] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.

[2] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[3] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, arXiv:1409.0473.

[4] Werbos P J. Generalization of backpropagation with application to a recurrent gas market model [J]. Neural networks, 1988, 1(4): 339-356.

[5] Werbos P J. Backpropagation through time: what it does and how to do it [J]. Proceedings of the IEEE, 1990, (78)10:1550-1560.

[6] Rumelhart D E, Hinton G E, Williams R J. Learning Internal Representations by Error Propagation [J], Parallel Distributed Processing, 1986, (1)71: 599-607.

[7] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997,

9(8): 1735-1780.

[8] 诸葛越, 葫芦娃. 百面机器学习[M]. 北京: 人民邮电出版社, 2018.

[9] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint, 2014, arXiv:1406.1078.

[10] Gers F A, Schmidhuber J. Recurrent nets that time and count[C]. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing, 2000, 3: 189-194.

[11] Greff K, Srivastava R K, Koutník J, Steunebrink B R, Schmidhuber J. LSTM: A Search Space Odyssey, in IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10):2222-2232.

[12] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.

[13] Blog A K. The Unreasonable Effectiveness of Recurrent Neural Networks[J]. May, 2015.

[14] https://github.com/chinese-poetry/chinese-poetry

第 8 章

[1] Kim Y. 2014. Convolutional Neural Networks for Sentence Classification [C]. In Proceedings of the 2014 Conference on EMNLP, pages 1746–1751.

[2] https://github.com/fxsjy/jieba

[3] https://www.tensorflow.org/text/tutorials/text_classification_rnn

[4] Zhou C, Sun C, Liu Z, et al. A C-LSTM neural network for text classification[J]. arXiv preprint, 2015, arXiv:1511.08630.

[5] Zhang Y, Yuan H, Wang J, et al. YNU-HPCC at EmoInt-2017: Using a CNN-LSTM model for sentiment intensity prediction[C]. Proceedings of the 8th workshop on computational approaches to subjectivity, sentiment and social media analysis. 2017, 200-204.

[6] Lin S, Xie H, Yu L C, et al. SentiNLP at IJCNLP-2017 Task 4: Customer feedback analysis using a Bi-LSTM-CNN model[C]. Proceedings of the IJCNLP 2017, Shared Tasks, 2017, 149-154.

[7] Sosa P M. Twitter sentiment analysis using combined LSTM-CNN models[J]. EprintArxiv, 2017, 2017: 1-9.

[8] Li Y, Wang X, Xu P. Chinese Text Classification Model Based on Deep Learning. Future Internet, 2018, 113(10).

[9] Kunfu Wang, Pengyi Zhang, et al. A Text Classification Method Based on the Merge-LSTM-CNN Model[J] . Journal of Physics: Conference Series, 2020, 1646(1): 012110.

[10] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [J].

Advances in neural information processing systems, 2015, 28.

[11] https://www.csc.kth.se/cvap/actions/

[12] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action

recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(1): 221-231.

[13] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3d convolutional networks[C]. Proceedings of the IEEE international conference on computer vision. 2015, 4489-4497.

[14] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]. Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

第 9 章

[1] https://zh.wikipedia.org/zh-cn/图灵测试

[2] 吴军，《数学之美》，北京: 人民邮电出版社, 2014.

[3] https://en.wikipedia.org/wiki/Natural_language_processing

[4] https://en.wikipedia.org/wiki/Language_model

[5] Natural Language Processing with Deep Learning, Stanford CS224n, Winter 2022.

[6] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[7] Dan J, James M H. N-gram Language Models. Speech and Language Processing, 2013.

[8] https://en.wikipedia.org/wiki/N-gram_language_model

[9] Brants T, et al. Large Language Models in Machine Translation[J]. EMNLP-CoNLL 2007, 2017, 858–867.

[10] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint, 2013, arXiv:1301.3781.

[11] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural information processing systems, 2013, 26.

[12] Řehůřek R, Sojka P. Software framework for topic modelling with large corpora[J]. 2010.

[13] https://drive.google.com/open?id=0B7XkCwpI5KDYNlNUTTlSS21pQmM&authuser=0

[14] https://github.com/moon-hotel/Chinese-Word-Vectors

[15] https://distill.pub/2016/misread-tsne/

[16] Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation[C]. In Proceedings of the 2014 Conference on EMNLP, 2014, 1532–1543.

[17] https://nlp.stanford.edu/projects/glove/

[18] https://www.cs.cornell.edu/people/pabo/movie-review-data/

[19] Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information[J]. Transactions of the association for computational linguistics, 2017, 5: 135-146.

[20] Grave E, Bojanowski P, Gupta P,et al. Learning Word Vectors for 157 Languages[C]. In Proceedings of the Eleventh International Conference on LREC, Miyazaki, Japan, 2018.

[21] https://nlp.stanford.edu/software/segmenter.html

[22] https://radimrehurek.com/gensim/

[23] https://fasttext.cc

[24] Joulin A, Grave E, Bojanowski P, et al. Bag of Tricks for Efficient Text Classification.[C] In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, Spain, 2017, 2: 427–431.

[25] Sutskeveri, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[J]. Advances in neural information processing systems, 2014, 27.

[26] Chen C, Kahn J. Fast Beam Search Decoding in PyTorch with TorchAudio and Flashlight Text, PyTorch.org.

[27] Neubig G. Neural machine translation and sequence-to-sequence models: A tutorial[J]. arXiv preprint, 2017, arXiv:1703.01619.

[28] https://en.wikipedia.org/wiki/Entropy_(information_theory)

[29] https://en.wikipedia.org/wiki/Perplexity

[30] Papineni K, Roukos S, Ward T, et al. Bleu: a method for automatic evaluation of machine translation[C] Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002, 311-318.

[31] https://en.wikipedia.org/wiki/Google_Translate

[32] https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

[33] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. arXiv preprint, 2016, arXiv:1609.08144.

[34]https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html

[35] Luong M T, Brevdo E, Zhao R, Neural Machine Translation (seq2seq) Tutorial, https://github.com/tensorflow/nmt, 2017.

[36] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, arXiv:1409.0473.

[37] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv preprint, 2014, arXiv:1409.1259.

[38] Attention | Definition, Theories, Aspects, & Facts | Britannica". www.britannica.com.

[39] https://en.wikipedia.org/wiki/Attention

[40] https://en.wikipedia.org/wiki/Visual_spatial_attention

[41] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV). 3-19, 2018.

[42] Mnih V, Heess N, Graves A. Recurrent models of visual attention[J]. Advances in neural information processing systems, 27, 2014.

[43] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[44] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint, 2015, arXiv:1508.04025.

第 10 章

[1] Han X, Zhang Z, Ding N, et al. Pre-trained models: Past, present and future[J]. AI Open, 2021, 2: 225-250.

[2] Matthew E. Peters, Neumann M, et al. Deep Contextualized Word Representations[C]. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, 1:2227–2237.

[3] Melamud O, Goldberger J, Dagan I. context2vec: Learning generic context embedding with bidirectional lstm[C] Proceedings of the 20th SIGNLL conference on computational natural language learning. 2016: 51-61.

[4] McCann B, Bradbury J, Xiong C, et al. Learned in translation: Contextualized word vectors[J]. Advances in neural information processing systems, 2017, 30.

[5] Wieting J, Bansal M, Gimpel K, et al. Charagram: Embedding words and sentences via character n-grams[J]. arXiv preprint, 2016, arXiv:1607.02789.

[6] Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information[J]. Transactions of the association for computational linguistics, 2017, 5: 135-146.

[7] Neelakantan A, Shankar J, Passos A, et al. Efficient non-parametric estimation of multiple embeddings per word in vector space[J]. arXiv preprint, 2015, arXiv:1504.06654.

[8] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks[J]. Advances in neural information processing systems, 2015, 28.

[9] https://allenai.org/allennlp

[10] 阿斯顿·张、李沐、扎卡里 C. 立顿等，动手学深度学习[M]，2 版. 北京:人民邮电出版社, 2019.

[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 30, 2017.

[12]

https://github.com/BAI-Yeqi/Statistical-Properties-of-Dot-Product/blob/master/proof.pdf

[13] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2018, arXiv:1810.04805.

[14] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.

[15] https://github.com/aceimnorstuvwxz/toutiao-text-classfication-dataset

[16] https://rowanzellers.com/swag/

[17] https://github.com/rowanz/swagaf/tree/master/data

[18] https://rajpurkar.github.io/SQuAD-explorer/

[19] https://github.com/zjy-ucas/ChineseNER

[20] https://s3.amazonaws.com/research.metamind.io/wikitext/wikitext-2-v1.zip

[21] https://github.com/google-research/bert/

[22] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.

[23] https://github.com/openai/gpt-2

[24] https://github.com/moon-hotel/gpt-2

[25]

https://openaipublic.blob.core.windows.net/gpt-2/models/1558M/model.ckpt.data-00000-of-00001

[26] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[27] Rosset C, Turing-NLG: A 17-billion-parameter language model by Microsoft, Published February 13, 2022.

[28] Child R, Gray S, Radford A, et al. Generating long sequences with sparse transformers[J]. arXiv preprint, 2019, arXiv:1904.10509.

[29] https://commoncrawl.org

[30] https://platform.openai.com/playground

[31] https://github.com/Morizeyao/GPT2-Chinese

[32] https://github.com/moon-hotel/GPT2-Chinese

[33] https://en.wikipedia.org/wiki/PyTorch_Lightning

[34] https://lightning.ai/docs/app/stable/

[35] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[J]. Advances in Neural Information Processing Systems, 2022, 35: 27730-27744.

[36] https://zh.wikipedia.org/wiki/强化学习

[37] https://en.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback

[38] https://openai.com/research/instruction-following

[39] https://openai.com/blog/chatgpt

[40] https://platform.openai.com/docs/models/gpt-3-5

[41] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.

[42] 张天熙.论运动减肥的基本内容与方法[J]. 宝鸡文理学院学报(社会科学版), 2013,

33(06):127-129

[43] https://www.chainbrainai.com.

[44] https://www.mindshow.fun/

[45] https://markmap.js.org/

[46] https://www.bing.com/images/create

[47] https://github.com/moon-hotel/DeepLearningWithMe

[48] https://github.com/moon-hotel/GPTs

[49] Jiang F, Xu Z, Niu L, et al. ArtPrompt: ASCII Art-based Jailbreak Attacks against Aligned LLMs[J]. arXiv preprint, 2024, arXiv:2402.11753.

[50] https://github.com/baichuan-inc/Baichuan2

[51] https://huggingface.co/baichuan-inc/Baichuan2-7B-Chat/tree/main

[52] https://www.baichuan-ai.com/chat

[53] Su J, Ahmed M, Lu Y, et al. Roformer: Enhanced transformer with rotary position embedding[J]. Neurocomputing, 2024, 568: 127063.

[54] GPT-4 Technical Report https://arxiv.org/abs/2303.08774

[55] gpt-4 turbo https://openai.com/blog/new-models-and-developer-products-announced-at-devday

[56]OpenAI's first developer conference November 6, 2023 · San Francisco, CA

[57] https://openai.com/research/gpt-4

[58] GPT 系 列 模 型 版 本 对 应 https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

[59] https://platform.openai.com/docs/api-reference/chat/create

[60] https://openai.com/blog/introducing-gpts