MINISTERUL EDUCAŢIEI ŞI CERCETĂRII AL REPUBLICII MOLDOVA
Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică şi Microelectronică
Departamentul Ingineria Software și Automatică

Admis la susţinere
Şef de departament:
Fiodorov I. dr., conf. univ.

„___” _____ 2022

Procesarea limbajului natural utilizând învățarea auto supravegheată

Teza de master

Student: _____ Onescu Alexandru-Vlad, TI-201M

Conducător: _____ Leahu Alexei, prof. univ. dr.

Consultant: _____ Cojocaru Svetlana, lect. univ.

Chişinău, 2022

Abstract

Introducem un nou model de reprezentare a limbajului numit BERT, care înseamnă Reprezentări codificatoare bidirecționale de Transformatoare. Spre deosebire de modelele recente de reprezentare a limbajului (Peters et al., 2018a; Radford et al., 2018), BERT este conceput pentru a antrena prealabil reprezentările bidirecționale profunde din text neetichetat prin condiționarea în comun a ambelor context stânga și dreapta în toate straturile. Ca rezultat, modelul BERT pre-antrenat poate fi ajustat cu doar un singur strat de ieșire suplimentar pentru a crea modele de ultimă generație pentru o largă serie de sarcini, cum ar fi răspunsul la întrebări și inferența limbajului, fără modificări substanțiale ale arhitecturii specifice sarcinii. BERT este conceptual simplu și empiric puternic. Obține rezultate noi de ultimă generație la unsprezece procesare a limbajului natural sarcini, inclusiv împingerea scorului GLUE la 80,5% (7,7% îmbunătățire absolută), Precizie MultiNLI la 86,7% (4,6% absolut de îmbunătățire), Modelele de transducție a secvenței dominante se bazează pe recurente complexe sau rețele neuronale convoluționale care includ un codificator și un decodor. Cel mai bun modelele performante conectează, de asemenea, codificatorul și decodorul printr-o atenție mecanism. Vă propunem o nouă arhitectură simplă de rețea, Transformer, bazată exclusiv pe mecanisme de atenție, dispensând recurența și circumvoluțiile în întregime. Experimentele pe două sarcini de traducere automată arată aceste modele să fie superioară ca calitate, fiind în același timp mai paralelizabilă și necesită în mod semnificativ mai puțin timp pentru antrenament. Modelul nostru atinge 28,4 BLEU la sarcina de traducere din engleză în germană WMT 2014, îmbunătățind cele mai bune rezultate existente, inclusiv ansambluri, de peste 2 BLEU. În sarcina de traducere din engleză în franceză WMT 2014, modelul nostru stabilește un nou scor BLEU de ultimă generație pentru un singur model de 41,8 după antrenament timp de 3,5 zile pe opt GPU-uri, o mică parte din costurile de instruire ale cele mai bune modele din literatură. rătăm că Transformerul generalizează bine la alte sarcini prin aplicarea cu succes la circumscripția engleză analizând ambele cu date mari și limitate de antrenament.

CUPRINS

INTRODUCERE

În ultimii ani, domeniul AI a făcut progrese enorme în dezvoltarea sistemelor de IA care pot învăța din cantități masive de date etichetate cu atenție. Această paradigmă a învățării supravegheate are o experiență dovedită pentru instruirea modelelor de specialitate care se descurcă extrem de bine în sarcina pe care au fost instruiți să o facă. Din păcate, există o limită pentru cât de departe poate merge domeniul AI doar cu învățarea supravegheată. Învățarea supravegheată este un blocaj pentru construirea unor modele generaliste mai inteligente care pot îndeplini sarcini multiple și dobândi noi abilități fără cantități masive de date etichetate.

Un limbaj natural este unul care a evoluat de-a lungul timpului prin utilizare și repetare. Nu implică planificarea și strategia deliberată. Latina, engleza, spaniola și multe alte limbi vorbite sunt toate limbile care au evoluat în mod natural în timp.

Limbile naturale sunt diferite de limbile formale sau construite, care au o origine și o cale de dezvoltare diferite. De exemplu, limbaje de programare, inclusiv C, Java, Python și multe altele au fost create dintr-un motiv specific.

Pentru ca o mașină să fie autonomă, un principiu cheie este să poată comunica printr-unul dintre limbajele naturale cunoscute oamenilor. În lumea largă a inteligenței artificiale, un domeniu se ocupă de activarea mașinilor pentru a interacționa folosind aceste limbaje: Natural Language Processing (NLP).

NLP este un termen umbrelă care cuprinde orice și orice este legat de a face mașinile capabile să proceseze limbajul natural - fie că primește intrarea, înțelege intrarea sau generează un răspuns.

Procesarea limbajului natural (NLP) este un subdomeniu de lingvistică, informatică și inteligență artificială, care se ocupă de interacțiunile dintre calculatoare și limbajul uman, în special modul de programare a computerelor pentru a procesa și analiza cantități mari de date de limbaj natural. Scopul este un computer capabil să „înțeleagă” conținutul documentelor, inclusiv nuanțele contextuale ale limbajului din ele. Tehnologia poate extrage apoi cu precizie informații și informații conținute în documente, precum și clasifica și organiza documentele în sine.

Provocările în procesarea limbajului natural implică frecvent recunoașterea vorbirii, înțelegerea limbajului natural și generarea limbajului natural.

Învățare automată (în engleză, „Machine Learning”) este un subdomeniu al informaticii și o ramură a inteligenței artificiale, al cărui obiectiv este de a dezvolta tehnici care dau calculatoarelor posibilitatea de a învăța. Mai precis, se urmărește să se creeze programe capabile de generalizare pe baza unor exemple.

Este, prin urmare, un proces inductiv. În multe cazuri, domeniul învățării automate se suprapune cu cel al statisticii computaționale, deoarece cele două discipline se bazează pe analiza datelor. Cu toate acestea, învățare automată, se concentrează și pe complexitatea computațională al problemelor. Multe probleme sunt în clasa NP-hard, așa că o mare parte din cercetările efectuate asupra procesul de învățare automată sunt axate pe proiectarea de soluții viabile la aceste probleme. Învățare automată poate fi văzut ca o încercare de a automatiza unele părți din metoda științifică, folosind metode matematice.

O ipoteză de lucru este că cunoașterea generalizată despre lume sau bunul simț formează grosul inteligenței biologice atât la oameni, cât și la animale. Această abilitate de bun simț este luată ca atare la oameni și animale, dar a rămas o provocare deschisă în cercetarea AI de la începuturile sale. Într-un fel, bunul simț este materia întunecată a inteligenței artificiale.

Simțul comun îi ajută pe oameni să învețe noi abilități fără a necesita cantități masive de predare pentru fiecare sarcină. De exemplu, dacă le arătăm doar câteva desene de vaci copiilor mici, ei vor putea în cele din urmă să recunoască orice vacă pe care o văd. Spre deosebire de acestea, sistemele de AI antrenate cu învățare supravegheată necesită multe exemple de imagini de vacă și ar putea să nu reușească încă să clasifice vacile în situații neobișnuite, cum ar fi culcat pe o plajă. Cum se poate ca oamenii să învețe să conducă o mașină în aproximativ 20 de ore de practică cu foarte puțină supraveghere, în timp ce conducerea complet autonomă scapă totuși de cele mai bune sisteme de AI formate cu mii de ore de date de la șoferii umani? Răspunsul scurt este că oamenii se bazează pe cunoștințele de bază dobândite anterior despre modul în care funcționează lumea.

BIBLIOGRAFIE

[1] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. arXiv preprint arXiv:1601.06733, 2016.

[2] Spyros Gidaris, Praveer Singh & Nikos Komodakis. "Unsupervised Representation Learning by Predicting Image Rotations" ICLR 2018.

[3] Carl Doersch, Abhinav Gupta, and Alexei A. Efros. "Unsupervised visual representation learning by context prediction." ICCV. 2015.

[4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[5] Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. Recurrent neural network grammars. In Proc. of NAACL, 2016.

[6] Richard Zhang, Phillip Isola & Alexei A. Efros. "Colorful image colorization." ECCV, 2016.

[7] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attentionbased neural machine translation. arXiv preprint arXiv:1508.04025, 2015.

[8] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. "Adversarial feature learning." ICLR 2017.

[9] Deepak Pathak, et al. "Context encoders: Feature learning by inpainting." CVPR. 2016.

[10] Richard Zhang, Phillip Isola, and Alexei A. Efros. "Split-brain autoencoders: Unsupervised learning by cross-channel prediction." CVPR. 2017.

[11] Xiaolong Wang & Abhinav Gupta. "Unsupervised Learning of Visual Representations using Videos." ICCV. 2015.

[12] Carl Vondrick, et al. "Tracking Emerges by Colorizing Videos" ECCV. 2018.

[13] Ishan Misra, C. Lawrence Zitnick, and Martial Hebert. "Shuffle and learn: unsupervised learning using temporal order verification." ECCV. 2016.

[14] Basura Fernando, et al. "Self-Supervised Video Representation Learning With Odd-One-Out Networks" CVPR. 2017.

[15] Ofir Press and Lior Wolf. Using the output embedding to improve language models. arXiv preprint arXiv:1608.05859, 2016.

[16] Florian Schroff, Dmitry Kalenichenko and James Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering" CVPR. 2015.

[17] Pierre Sermanet, et al. "Time-Contrastive Networks: Self-Supervised Learning from Video" CVPR. 2018.

[18] Alex Graves. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013.

[19] Eric Jang & Coline Devin, et al. "Grasp2Vec: Learning Object Representations from Self-Supervised Grasping" CoRL. 2018.

[20] Oleksii Kuchaiev and Boris Ginsburg. Factorization tricks for LSTM networks. arXiv preprint arXiv:1703.10722, 2017.

[21] Ashvin Nair, et al. "Contextual imagined goals for self-supervised robotic learning" CoRL. 2019.

[22] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. Neural machine translation in linear time. arXiv preprint arXiv:1610.10099v2, 2017.

[23] Olivier J. Henaff, et al. "Data-Efficient Image Recognition with Contrastive Predictive Coding" arXiv preprint arXiv:1905.09272, 2019.

[24] Kaiming He, et al. "Momentum Contrast for Unsupervised Visual Representation Learning." CVPR 2020.

[25] Zhirong Wu, et al. "Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination." CVPR 2018.

[26] Ting Chen, et al. "A Simple Framework for Contrastive Learning of Visual Representations." arXiv preprint arXiv:2002.05709, 2020.

[27] Aravind Srinivas, Michael Laskin & Pieter Abbeel "CURL: Contrastive Unsupervised Representations for Reinforcement Learning." arXiv preprint arXiv:2004.04136, 2020.

[28] Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model. In Empirical Methods in Natural Language Processing, 2016.

[29] Amy Zhang, et al. "Learning Invariant Representations for Reinforcement Learning without Reconstruction" arXiv preprint arXiv:2006.10742, 2020.

[30] Xinlei Chen, et al. "Improved Baselines with Momentum Contrastive Learning" arXiv preprint arXiv:2003.04297, 2020.

[31] Jean-Bastien Grill, et al. "Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning" arXiv preprint arXiv:2006.07733, 2020.

[32] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, Advances in Neural Information Processing Systems 28, pages 2440–2448. Curran Associates, Inc., 2015.

[33] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Disponibil: https://arxiv.org/abs/1810.04805

[34] Self Supervised Representation Learning in NLP. Disponibil: https://amitness.com/2020/05/self-supervised-learning-nlp/

[35] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.

[36] A curated list of awesome self-supervised method. Disponibil: https://github.com/jason718/awesome-self-supervised-learning

[37] Self-Supervised Representation Learning. Disponibil: https://lilianweng.github.io/lil-log/2019/11/10/self-supervised-learning.html

[38] Grokking self-supervised (representation) learning: how it works in computer vision and why | AI Summer. Disponibil: https://theaisummer.com/self-supervised-representation-learning-computer-vision/

[39] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.

[40] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473, 2014.

[41] Denny Britz, Anna Goldie, Minh-Thang Luong, and Quoc V. Le. Massive exploration of neural machine translation architectures. CoRR, abs/1703.03906, 2017.

[42] Alexey Dosovitskiy, et al. "Discriminative unsupervised feature learning with exemplar convolutional neural networks." IEEE transactions on pattern analysis and machine intelligence 38.9 (2015): 1734-1747.

[43] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. CoRR, abs/1406.1078, 2014.

[44] Francois Chollet. Xception: Deep learning with depthwise separable convolutions. arXiv preprint arXiv:1610.02357, 2016.

[45] Junyoung Chung, Çaglar Gülçehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR, abs/1412.3555, 2014.

[46] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. "Representation learning by learning to count." ICCV. 2017.

[47] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122v2, 2017.

[48] Debidatta Dwibedi, et al. "Learning actionable representations from visual observations." IROS. 2018.

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[50] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and Jürgen Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[51] Mehdi Noroozi & Paolo Favaro. "Unsupervised learning of visual representations by solving jigsaw puzzles." ECCV, 2016.

[52] Zhongqiang Huang and Mary Harper. Self-training PCFG grammars with latent annotations across languages. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 832–841. ACL, August 2009.

[53] Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. Exploring the limits of language modeling. arXiv preprint arXiv:1602.02410, 2016.

[54] Łukasz Kaiser and Samy Bengio. Can active memory replace attention? In Advances in Neural Information Processing Systems, (NIPS), 2016.

[55] Łukasz Kaiser and Ilya Sutskever. Neural GPUs learn algorithms. In International Conference on Learning Representations (ICLR), 2016.

[56] Aaron van den Oord, Yazhe Li & Oriol Vinyals. "Representation Learning with Contrastive Predictive Coding" arXiv preprint arXiv:1807.03748, 2018.

[57] Yoon Kim, Carl Denton, Luong Hoang, and Alexander M. Rush. Structured attention networks. In International Conference on Learning Representations, 2017.

[58] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In ICLR, 2015.

[59] Ashvin Nair, et al. "Visual reinforcement learning with imagined goals" NeuriPS. 2018.

[60] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130, 2017. [61] Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser. Multi-task sequence to sequence learning. arXiv preprint arXiv:1511.06114, 2015.

[62] Pascal Vincent, et al. "Extracting and composing robust features with denoising autoencoders." ICML, 2008.

[63] Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of english: The penn treebank. Computational linguistics, 19(2):313–330, 1993.

[64] David McClosky, Eugene Charniak, and Mark Johnson. Effective self-training for parsing. In Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, pages 152–159. ACL, June 2006.

[65] Carles Gelada, et al. "DeepMDP: Learning Continuous Latent Space Models for Representation Learning" ICML 2019.

[66] Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304, 2017.

[67] Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. Learning accurate, compact, and interpretable tree annotation. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, pages 433–440. ACL, July 2006.

[68] Donglai Wei, et al. "Learning and Using the Arrow of Time" CVPR. 2018. [69] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909, 2015.

[70] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538, 2017.

[71] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[72] Abe Fetterman & Josh Albrecht. "Understanding self-supervised and contrastive learning with Bootstrap Your Own Latent (BYOL)" Untitled blog. Aug 24, 2020. [73] Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems, pages 3104–3112, 2014.

[74] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. CoRR, abs/1512.00567, 2015.

[75] Vinyals & Kaiser, Koo, Petrov, Sutskever, and Hinton. Grammar as a foreign language. In Advances in Neural Information Processing Systems, 2015.

[76] The Rise of Self-Supervised Learning. Disponibil: https://jonathanbgn.com/2020/12/31/self-supervised-learning.html

[77] Jie Zhou, Ying Cao, Xuguang Wang, Peng Li, and Wei Xu. Deep recurrent models with fast-forward connections for neural machine translation. CoRR, abs/1606.04199, 2016.

[78] Muhua Zhu, Yue Zhang, Wenliang Chen, Min Zhang, and Jingbo Zhu. Fast and accurate shift-reduce constituent parsing. In Proceedings of the 51st Annual Meeting of the ACL (Volume 1: Long Papers), pages 434–443. ACL, August 2013.