Communication Algorithms via Deep Learning

Conference Paper

Author(s):
Kim, Hyeji; Jiang, Yihan; Rana, Ranvir; Kannan, Sreeram; Oh, Sewoong; Viswanath, Pramod

Publication date:
2018-02

Permanent link:
https://doi.org/10.3929/ethz-b-000245054

Communication Algorithms via Deep Learning

Hyeji Kim*, Yihan Jiang†, Ranvir Rana*, Sreeram Kannan†, Sewoong Oh*, Pramod Viswanath*
University of Illinois at Urbana Champaign*, University of Washington†

Reliable digital communication, both wireline (ethernet, cable and DSL modems) and wireless (cellular, satellite, deep space), is a primary workhorse of the modern information age. A critical aspect of reliable communication involves the design of codes that allow transmissions to be robustly (and computationally efficiently) decoded under noisy conditions. This is the discipline of coding theory; over the past century and especially the past 70 years (since the birth of information theory [1]) much progress has been made in the design of near optimal codes. Landmark codes include convolutional codes, turbo codes, low density parity check (LDPC) codes and, recently, polar codes. The impact on humanity is enormous – every cellular phone designed uses one of these codes, which feature in global cellular standards ranging from the 2nd generation to the 5th generation respectively, and are text book material [2].

The canonical setting is one of point-to-point reliable communication over the additive white Gaussian noise (AWGN) channel and performance of a code in this setting is its gold standard. The AWGN channel fits much of wireline and wireless communications although the front end of the receiver may have to be specifically designed before being processed by the decoder (example: intersymbol equalization in cable modems, beamforming and sphere decoding in multiple antenna wireless systems); again this is text book material [3]. There are two long term goals in coding theory: (a) design of new, computationally efficient, codes that improve the state of the art (probability of correct reception) over the AWGN setting. Since the current codes already operate close to the information theoretic "Shannon limit", the emphasis is on *robustness* and *adaptability* to deviations from the AWGN settings (a list of channel models motivated by practical settings, (such as urban, pedestrian, vehicular) in the recent 5th generation cellular standard is available in Annex B of 3GPP TS 36.101. (b) design of new codes for multi-terminal (i.e., beyond point-to-point) settings – examples include the feedback channel, the relay channel and the interference channel.

Progress over these long term goals has generally been driven by individual human ingenuity and, befittingly, is sporadic. For instance, the time duration between convolutional codes (2nd generation cellular standards) to polar codes (5th generation cellular standards) is over 4 decades. Deep learning is fast emerging as capable of learning sophisticated algorithms from observed data (input, action, output) alone and has been remarkably successful in a large variety of human endeavors (ranging from language [4] to vision [5] to playing Go [6]). Motivated by these successes, we posit that deep learning methods can play a crucial role in solving both the aforementioned goals of coding theory and show that we can make significant progress on both these goals in this work.

While the learning framework is clear and there is virtually unlimited training data available, there are two main challenges: (a) The space of codes is very vast and the sizes astronomical; for instance a rate 1/2 code over 100 information bits involves designing 2^{100} codewords in a 200 dimensional space. Computationally efficient encoding and decoding procedures are a must, apart from high reliability over the AWGN channel. (b) Generalization is highly desirable across block lengths and data rate that each work very well over a wide range of channel signal to noise ratios (SNR). In other words, one is looking to design a family of codes (parametrized by data rate and number of information bits) and their performance is evaluated over a range of channel SNRs.

In part due to these challenges, recent deep learing works on coding theory focus on decoding known codes using data-driven neural decoders for short block lengths [7, 8, 9]. The main challenge is to restrict oneself to a class of codes that neural networks can naturally encode and decode. In this work, we restrict ourselves to a class of *sequential* encoding and decoding schemes, of which convolutional and turbo codes are part of. These sequential coding schemes naturally meld with the family of recurrent neural network (RNN) architectures, which have recently seen large success

*H. Kim, R. Rana and P. Viswanath are with Coordinated Science Lab and Department of Electrical Engineering at University of Illinois at Urbana Champaign. S. Oh is with Coordinated Science Lab and Department of Industrial and Enterprise Systems Engineering at University of Illinois at Urbana Champaign. Email: {hyejikim,rbrana2,swoh,pramodv}@illinois.edu (H. Kim, R. Rana, S.Oh, and P.Viswanath)

†Y. Jiang and S. Kannan are with the Department of Electrical Engineering at University of Washington. Email: yihanrogerjiang@gmail.com (Y. J), ksreeram@uw.edu (S. K.)

The full manuscript of this work is under review at the 6th International Conference on Learning Representations.

in a wide variety of time-series tasks. The ancillary advantage of sequential schemes is that arbitrarily long information bits can be encoded and also at a large variety of coding rates. Working within sequential codes parametrized by RNN architectures, we make the following contributions.

(1) Focusing on *convolutional codes* we aim to decode them on the AWGN channel using RNN architectures. Efficient optimal decoding of convolutional codes has represented historically fundamental progress in the broad arena of algorithms; optimal bit error decoding is achieved by the 'Viterbi decoder' [10] which is simply dynamic programming or Dijkstra's algorithm on a specific graph (the 'trellis') induced by the convolutional code. Optimal block error decoding is the BCJR decoder [11] which is part of a family of forward-backward algorithms. While early work had shown that vanilla-RNNs are capable in *principle* of emulating both Viterbi and BCJR decoders [12, 13] we show empirically, through a careful construction of RNN architectures and training methodology, that neural network decoding is possible at very near optimal performances (both bit error rate (BER) and block error rate (BLER)). The key point is that we train a RNN decoder at a *specific* SNR and over *short information bit* lengths (100 bits) and show *strong generalization* capabilities by testing over a wide range of SNR and block lengths (up to 10,000 bits). The specific training SNR is closely related to the Shannon limit of the AWGN channel at the rate of the code and provides strong information theoretic collateral to our empirical results.

(2) *Turbo codes* are naturally built on top of convolutional codes, both in terms of encoding and decoding. A natural generalization of our RNN convolutional decoders allow us to decode turbo codes at BER comparable to, and at certain regimes, even *better* than state of the art turbo decoders on the AWGN channel. That data driven, SGD-learnt, RNN architectures can decode comparably is fairly remarkable since turbo codes already operate near the Shannon limit of reliable communication over the AWGN channel.

(3) We show the afore-described neural network decoders for both convolutional and turbo codes are *robust* to variations to the AWGN channel model. We consider a problem of contemporary interest: communication over a "bursty" AWGN channel (where a small fraction of noise has much higher variance than usual) which models inter-cell interference in OFDM cellular systems (used in 4G and 5G cellular standards) or co-channel radar interference. We demonstrate empirically the neural network architectures can adapt to such variations and beat state of the art heuristics comfortably (despite evidence elsewhere that neural network are sensitive to models they are trained on [14]). Via an innovative local perturbation analysis (akin to [15]), we demonstrate the neural network to have learnt sophisticated preprocessing heuristics in engineering of real world systems [16].

(4) We demonstrate new RNN-driven encoders (with matching decoders) that operate significantly better than state of the art on the AWGN channel with (noisy) output feedback. While feedback does not improve the Shannon capacity of the AWGN channel [17], it is known to provide better reliability at finite block lengths [18], although very sensitive to even tiny amounts of noise in the output feedback; more generally any linear code incorporating the noisy output feedback cannot achieve a non-zero reliable rate of communication [19] – this is very troubling since all practical codes are linear and linear codes are known to achieve capacity (without feedback) [20]. Our RNN parameterized encoders are inherently *nonlinear* and map information bits *directly* to real-valued transmissions. Their performance vastly improves the state of the art on the long standing open problem in information theory on communicating over the AWGN channel with noisy output feedback.

References

[1] C. E. Shannon, "A mathematical theory of communication, part i, part ii," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.

[2] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[3] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[6] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[7] T. J. O'Shea and J. Hoydis, "An introduction to machine learning communications systems," *CoRR*, vol. abs/1702.00832, 2017. [Online]. Available: http://arxiv.org/abs/1702.00832

[8] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, “Deep learning-based communication over the air,” *arXiv preprint arXiv:1707.03384*, 2017.

[9] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink, “On deep learning-based channel decoding,” in *Information Sciences and Systems (CISS), 2017 51st Annual Conference on*. IEEE, 2017, pp. 1–6.

[10] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[11] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate (corresp.),” *IEEE Transactions on information theory*, vol. 20, no. 2, pp. 284–287, 1974.

[12] X.-A. Wang and S. B. Wicker, “An artificial neural net viterbi decoder,” *IEEE Transactions on Communications*, vol. 44, no. 2, pp. 165–171, Feb 1996.

[13] M. H. Sazlı and C. Icsık, “Neural network implementation of the bcjr algorithm,” *Digital Signal Processing*, vol. 17, no. 1, pp. 353 – 359, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200406000029

[14] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, “Intriguing properties of neural networks,” *arXiv preprint arXiv:1312.6199*, 2013.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, “"why should i trust you?": Explaining the predictions of any classifier,” in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. New York, NY, USA: ACM, 2016, pp. 1135–1144. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939778

[16] J. Li, X. Wu, and R. Laroia, *OFDMA mobile broadband communications: A systems approach*. Cambridge University Press, 2013.

[17] C. Shannon, “The zero error capacity of a noisy channel,” *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[18] J. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback–i: No bandwidth constraint,” *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[19] Y.-H. Kim, A. Lapidoth, and T. Weissman, “The gaussian channel with noisy feedback,” in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1416–1420.

[20] P. Elias, “Coding for noisy channels,” in *IRE Convention record*, vol. 4, 1955, pp. 37–46.