Emily Mulhall and Elias Stengel-Eskin
COMP 550
Final Project Proposal
October 20, 2017

1 Proposal

Hidden Markov models (HMMs) achieve high accuracies on text decoding tasks when the encoding is a simple cipher. We obtain 98.61% accuracy using an HMM with Laplace smoothing and a language model for a simple substitution cipher (a Caesar cipher). The same HMM obtained 84.50% accuracy on text encoded using a two letter substitution cipher, where one substitution is randomly chosen for each input character. However, in ciphers where the encoded text results from adding integer values of a key (produced by permuting the characters of the plain text), the same HMM architecture fails. Using sequence-to-sequence modeling, specifically long short-term memory networks (LSTMs) and recurrent neural networks (RNNs) encoder-decoder models, we hope to more accurately decode this cipher text.

Sequence-to-sequence modeling involves learning the function between one sequence of tokens and another. This is most applicable in machine translation, where a sequence of words in one language is mapped to words in another language, but also has applications in summarization and cryptography. Many of the most successful models fall under encoder-decoder models.

Encoder-decoder models are a special case of RNNs which can be used to generate the next element in a sequence using the probability of that element given the preceding elements and some vector c which is a representation of the input sequence built using an "encoding" network. Thus, each token in the output sequence is conditioned on the representation of the entire input sequence as well as the previously predicted tokens (Goldberg, 2017). In our case, the encoder-decoder network will use an RNN or an LSTM to obtain a vector representation of the coded sequence and then condition on that representation to decode the sequence into plain text.

To test whether encoder-decoder models better decode complex cipher text, we will create both an RNN and LSTM encoder-decoders. We will also create simpler RNN and LSTM models which do not condition on the entire sequence, but merely on the current input character. Greydanus (2017) showed that this simple LSTM architecture could decode the famously complex enigma cipher, which leads us to believe that the simpler models will be able to decode the ciphers as well as the more complex encoder-decoder model.

We would like to test both RNN and LSTM models because LSTMs are less susceptible to vanishing or exploding gradients. Given the extremely small amount of data (13 exemplars) needed to obtain high results on the simpler two ciphers with the HMM, we expect that the HMM will outperform the neural network models on small datasets. However, with more training data, we hypothesize that the neural network models will obtain higher accuracies on all three ciphers; in particular, we are interested in seeing if the neural network models are able to learn the third cipher, which the HMM was unable to learn. We also hypothesize that even with additional training data the HMM will be unable to capture the structure of the third cipher, and the accuracy will remain low. Additionally, we expect the LSTM models to be more accurate than the simple RNN models on all three cipher texts.

References

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, 10(1):1–309.

Greydanus, S. (2017). Learning the enigma with recurrent neural networks. *arXiv preprint arXiv:1708.07576*.