Proposal: Transformer-CRF Integration for Sequence Labeling on EEG Data

Xiaohang Ma[1], Shiying Xiao[2], and Xiaohui Yin[2]

[1]Department of Mathematics, University of Connecticut

[2]Department of Statistics, University of Connecticut

October 11, 2024

1 Project Overview

Sequence labeling tasks are pivotal in diverse domains, especially healthcare, where precise and timely data classification can significantly impact patient outcomes. Medical datasets, such as physiological signals, often exhibit complexity and heterogeneity, posing challenges for traditional modeling techniques. Therefore, developing models capable of extracting meaningful patterns from these datasets is crucial for accurate diagnostic and prognostic decisions.

In recent years, deep learning has emerged as a powerful tool for sequence labeling. Neural network architectures like long short-term memory (LSTM) and Transformer models have revolutionized sequential data processing. Additionally, conditional random fields (CRFs) have been employed for modeling label dependencies, demonstrating promising results. However, while these methods have shown individual strengths, a comprehensive approach that leverages their complementary capabilities is essential for tackling complex, multi-modal datasets.

This project introduces a novel framework that combines deep neural networks and probabilistic graphical models (PGMs) to enhance sequence labeling performance. By employing Transformer (Vaswani et al., 2017) for feature extraction and integrating them with CRF models, we aim to improve sequence tagging accuracy. Furthermore, mean-field approximation techniques are utilized for efficient inference, ensuring a computationally efficient yet effective approach. This integration of modern neural network architectures with CRFs offers a robust and promising solution for complex sequence labeling tasks.

2 Literature Review

Deep neural networks, particularly recurrent neural networks (RNNs) like LSTM, are widely used for sequence labeling due to their ability to capture long-range dependencies in sequences. Bidirectional LSTM (Bi-LSTM) further enhances this by processing sequences in both directions, improving context understanding. CRFs, on the other hand, are commonly used for structured predictions, modeling dependencies between neighboring labels to ensure valid output sequences. Huang et al. (2015) pioneered the use of Bi-LSTM integrated with CRFs for sequence labeling tasks, demonstrating their effectiveness in extracting features and ensuring label consistency.

However, Bi-LSTMs can still struggle to fully account for structural dependencies between predicted labels, leading to potential inconsistencies in the output sequence. Additionally, they typically require large amounts of labeled data and are computationally expensive, making them challenging to deploy in resource-constrained environments, such as hospitals.

Transformer-based models, such as BERT (Devlin et al., 2019), have emerged as powerful alternatives to traditional RNN-based approaches. Trained on massive datasets in an unsupervised manner, Transformer can be fine-tuned for various natural language processing tasks, including sequence labeling. Devlin et al. (2019) directly compared BERT to traditional models, including Bi-LSTM, and demonstrated its superior performance on tasks like named entity recognition, highlighting its ability to capture more feature information than LSTM models.

While CRFs are effective for modeling label dependencies, they often lack efficient closed-form solutions, especially for complex models beyond linear-chain CRFs. To address this, Krähenbühl and Koltun (2011) introduced the use of mean-field approximation to solve more general CRFs in the context of complex prediction tasks. By minimizing the Kullback-Leibler (KL) divergence, mean-field approximation can be reduced to a fixed-point iteration, providing a faster and more efficient approach for training and inference while maintaining suitable accuracy. Zheng et al. (2015) further extended this technique by reformulating mean-field approximation for fully connected CRFs, integrating an RNN structure to capture temporal dependencies.

3 Data

The harmful brain activity classification dataset (Jing et al., 2023) offers a valuable resource for studying brain health. This dataset, originally used in a Kaggle competition that concluded on April 8th, 2024 (available at `https://www.kaggle.com/competitions/`

hms-harmful-brain-activity-classification), contains electroencephalography (EEG) signals from hospital patients, collected by Harvard Medical School. The data comes in two primary formats: (i) EEG time series data, which records 50-second windows of brain activity measured by 19 electrodes placed on the scalp; (ii) Spectrogram data, which provides frequency-based representations of the EEG signals.

4 Activities Plan

1. **Probability graphical model** (Shiying Xiao and Xiaohang Ma)
   - Build the CRF layer for label predition
   - Mean field approximation inference and training of CRF model
2. **Deep learning methods** (Shiying Xiao, Xiaohui Yin and Xiaohang Ma)
   - CNN and RNN layers for feature extraction
   - RNN-CRF layers for label discrimination
3. **Manuscript writing** (Shiying Xiao, Xiaohui Yin and Xiaohang Ma)
4. **Project presentation** (Shiying Xiao, Xiaohui Yin and Xiaohang Ma)

References

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics.

Fan, J., C. Fang, Y. Gu, and T. Zhang (2023). Environment invariant linear least squares.

Huang, Z., W. Xu, and K. Yu (2015). Bidirectional LSTM-CRF models for sequence tagging.

Jing, J., W. Ge, S. Hong, M. B. Fernandes, Z. Lin, C. Yang, S. An, A. F. Struck, A. Herlopian, I. Karakis, J. J. Halford, M. C. Ng, E. L. Johnson, B. L. Appavu, R. A. Sarkis, G. Osman, P. W. Kaplan, M. B. Dhakar, L. A. Jayagopal, Z. Sheikh, O. Taraschenko, S. Schmitt, H. A. Haider, J. A. Kim, C. B. Swisher, N. Gaspard, M. C. Cervenka, A. A. Rodriguez Ruiz, J. W. Lee, M. Tabaeizadeh, E. J. Gilmore, K. Nordstrom, J. Y. Yoo, M. G. Holmes, S. T.

Herman, J. A. Williams, J. Pathmanathan, F. A. Nascimento, Z. Fan, S. Nasiri, M. M. Shafi, S. S. Cash, D. B. Hoch, A. J. Cole, E. S. Rosenthal, S. F. Zafar, J. Sun, and M. B. Westover (2023). Development of expert-level classification of seizures and rhythmic and periodic patterns during EEG interpretation. *Neurology 100*(17), e1750–e1762.

Krähenbühl, P. and V. Koltun (2011). Efficient inference in fully connected CRFs with gaussian edge potentials. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, Red Hook, NY, USA. Curran Associates, Inc.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017).* Curran Associates, Inc.

Zheng, S., S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr (2015). Conditional random fields as recurrent neural networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1529–1537. IEEE.