Parsing Wireless Electrocardiogram Signals with the CRF-CFG Model

Thai Nguyen, Roy J. Adams, Annamalai Natarajan, and Benjamin M. Marlin
University of Massachusetts Amherst

Introduction: Recent advances in wearable sensor technology have made it possible to simultaneously collect multiple streams of physiological and context data from individuals as they go about their daily activities in natural environments. However, data collected by such wireless on-body sensors is often fraught with sensor drop out, drift, and noise. In this work, we focus specifically on the case of analyzing data from low-cost wireless electrocardiogram (ECG) sensors where the limited number of electrodes, variable electrode placement, and poor electrode contact with the skin can lead to substantial electrical noise and thus uncertainty in predictions derived from the data.

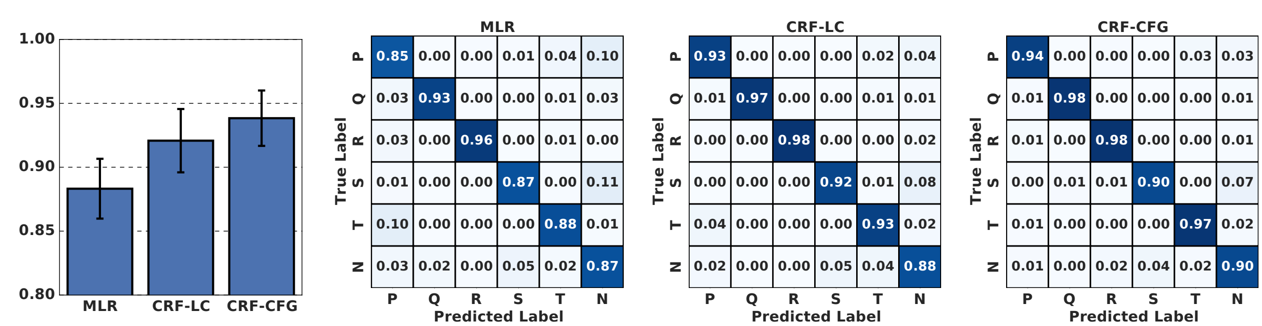
Background: An ECG sensor produces a data stream corresponding to the electrical activity of the muscles of the heart. A normal cardiac cycle (corresponding to a single heart beat) produces a characteristic sequence of five waves (the P, Q, R, S, and T waves). Applications of ECG including arrhythmia detection [3], stress detection [1], and the detection of drug use [4] require analysis of the shapes of these waves and the intervals between them. Prior work has shown that probabilistic approaches to ECG morphological analysis, specifically linear chain conditional random field (CRF) models, can significantly out-perform deterministic rule-based methods when applied to wireless ECG data [4].

Proposed Approach: In this work, we present a new approach to the ECG morphology extraction problem based on the use of Context Free Grammar Conditional Random Field (CRF-CFG) models [2, 5]. The CRF-CFG model is a log-linear model in the CRF family that can capture higher-order structure in data through the CFG, while allowing flexible feature functions to be defined for each production. We propose a novel robust grammar for ECG data along with feature functions that capture the shapes of ECG waves as well as the time between them. As in [4], we use an augmented set of wave types that explicitly includes a noise label (n) to deal with noise artifacts in the raw data. The proposed grammar is shown below with the terminal symbols p, q, r, s, t, n corresponding to the wave labels. The internal symbols and productions allow for skipping missing wave types.

$$
\begin{aligned}
\alpha &\to SP \mid SQ \mid SR \mid SS \mid ST \mid SN \\
SP &\to P\ SQ \mid P\ SP \mid P\ SR \mid P\ SS \mid P\ ST \\
SQ &\to Q\ SQ \mid Q\ SP \mid Q\ SR \mid Q\ SS \mid Q\ ST \\
SR &\to R\ SQ \mid R\ SP \mid R\ SR \mid R\ SS \mid R\ ST \\
SS &\to S\ SQ \mid S\ SP \mid S\ SR \mid S\ SS \mid S\ ST \\
ST &\to T\ SQ \mid T\ SP \mid T\ SR \mid T\ SS \mid T\ ST \\
SN &\to N\ SQ \mid N\ SP \mid N\ SR \mid N\ SS \mid N\ ST
\end{aligned}
$$

$$
\begin{aligned}
P &\to p \mid p\ N & \qquad S &\to s \mid s\ N \\
Q &\to q \mid q\ N & \qquad T &\to t \mid t\ N \\
R &\to r \mid r\ N & \qquad N &\to n \mid n\ N
\end{aligned}
$$

Experiments and Results: We evaluate the proposed CRF-CFG model on a wireless ECG data set collected from six subjects. The data set includes approximately $3,700$ labeled waves per subject. We apply a leave-one-subject-out (LOSO) train-test protocol with nested LOSO cross validation to set hyperparameters. We compare the performance of the CRF-CFG model with two baseline methods: the linear chain CRF (CRF-LC) and multinomial logistic regression (MLR). In the figure below, we report the mean ECG wave labeling accuracy and confusion matrices for each model. The CRF-CFG model achieves a mean ECG peak labeling accuracy of 94%, which is a 22.2% reduction in error when compared to using the previously proposed linear chain CRF model [4].



References

[1] F Alamudun, J Choi, R Gutierrez-Osuna, H Khan, and B Ahmed. Removal of subject-dependent and activity-dependent variation in physiological measures of stress. In *Pervasive Computing Technologies for Healthcare, Proceedings of the 6th International Conference on*, pages 115–122, 2012.

[2] Jenny Rose Finkel, Alex Kleeman, and Christopher D Manning. Efficient, feature-based, conditional random field parsing. In *ACL*, volume 46, pages 959–967, 2008.

[3] Sheng Hu, Zhenzhou Shao, and Jindong Tan. A real-time cardiac arrhythmia classification system with wearable electrocardiogram. In *Body Sensor Networks, Proceedings of the 2011 International Conference on*, pages 119–124, 2011.

[4] Annamalai Natarajan, Edward Gaiser, Gustavo Angarita, Robert Malison, Deepak Ganesan, and Benjamin Marlin. Conditional random fields for morphological analysis of wireless ecg signals. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 370–379. ACM, 2014.

[5] Charles Sutton and Andrew McCallum. *Conditional probabilistic context-free grammars*. PhD thesis, Citeseer, 2004.