Hierarchical Nested CRFs for Segmentation and Labeling of Physiological Time Series

Roy Adams\textsuperscript{1}, Edison Thomaz\textsuperscript{2}, and Benjamin Marlin\textsuperscript{1}
\textsuperscript{1}: UMass Amherst CICS. \textsuperscript{2}: Georgia Institute of Technology

INTRODUCTION

Given a densely observed time series, our goal is to infer a nested hierarchical segmentation and labeling at multiple temporal scales.

This work is motivated by problems in mHealth including the detection and delineation of health related activities like eating, drinking, smoking, and drug use based on continuously recorded on-body sensor data (e.g. respiration wave-forms, electrocardiogram waveforms, actigraphy data, etc.).

CONTRIBUTIONS

1. A flexible Hierarchical Nested Segmentation (HNS) model framework for jointly labeling and segmenting discrete time series that can be extended to incorporate a number of different hard constraints and high-order factors.

2. A quadratic time inference method based on the inside-outside algorithm.

3. Positive results on an eating detection task using actigraphy data.

RELATED WORK

Computer Vision: Models for hierarchical segmentation of images exist, but typically assert a Pott’s model for labels \cite{7} or require a predefined hierarchy \cite{2}.

Natural Language Processing: Our work is closely related to probabilistic parsing models, such as \cite{3}, and can be formulated as a probabilistic context-free grammar.

mHealth: Our applications build on past methods for activity detection using actigraphy data including \cite{4}.

KEY NOTATION

We use the following variables to represent cycle and segment level features and labels.

\begin{itemize}
  \item \(Y_{i0}\): Event level label for position \(i\).
  \item \(X_{i0}\): Event level features for position \(i\).
  \item \(Y_{jk}\): Label for the segment from \(j\) to \(k\) in level \(l\).
  \item \(X_{jk}\): Features for the segment from \(j\) to \(k\) in level \(l\).
\end{itemize}

HIERARCHICAL SEGMENTATION CONDITIONAL RANDOM FIELD

The Hierarchical Nested Segmentation (HNS) model induces a probability distribution over labeled nested segmentations. The model is based on a high-order Conditional Random Field (CRF). The probability of the segmentation and labeling \(y\) given the features \(x\) is:

\[ P(y = y|x = x, w) = \frac{1}{Z(x)} \prod_{j} \prod_{k} \psi_{y}(y_{jk}, x, w) \]

We define two fixed collections of binary factors to enforce the consistency of the segmentation at each level \(l\) and the proper nesting of segments between levels \(l'\):

\begin{itemize}
  \item \(\psi_{y}(1) = 1\) if \(\forall i \exists_{1}(j, k)\text{ s.t. } j < i \leq k \text{ and } y_{jk} \neq 0\)
  \item \(\psi_{y}(0) = 0\) otherwise
\end{itemize}

These two sets of factors induce a uniform distribution over valid nested segmentations. We can flexibly model different properties of segmentations using additional factors.

RESULTS ON EATING DETECTION

We use the following variables to represent cycle and segment level features and labels.

\begin{itemize}
  \item \(y_{0}\): Event level label for position \(i\).
  \item \(x_{0}\): Event level features for position \(i\).
\end{itemize}

Figure (a) shows a Linear Chain CRF (LC-CRF), (b) shows a two-level tree structured CRF (T-CRF), and (c) shows a two-level HNS model.

HNS FOR EVENTS AND ACTIVITIES

We construct a three-layer HNS model on top of an unsupervised base segmentation (e.g. into individual hand gestures) where the bottom layer represents events, the mid layer represents inter-event intervals, and the top layer represents complete activity sessions.

- **Event Level**: The first level represents event type for individual cycles. For example, an individual gesture may be either a food-to-mouth gesture or not.
- **Inter-event Level**: The first segmentation level represents inter-event durations. We constrain inter-event segments to start and end on a positive event cycle and incorporate duration information (e.g. length in cycles) again using log-linear factors.
- **Session Level**: Sessions represent complete activities from start to finish, such as a meal. Negative sessions are restricted to only cover cycle durations. We include a session level cardinality factor that models the number of positive events beneath a positive session.

INFERENCE AND LEARNING

We perform inference using a modified version of the inside-outside algorithm, an exact MAP inference method based on dynamic programming. Specifically, our algorithm closely resembles that of Semi-Markov CRFs [5] and runs in O(Cn'^2) time, where C is the maximum number of positive events allowed in a single positive session.

To learn our model, we use loss-augmented maximum margin learning, as presented in [6] and implemented for python in [7].

EXPERIMENTAL DETAILS

We evaluated the performance of a three-layer HNS model on the problem of eating detection from wrist-worn actigraphy data [4]. We compared performance on two tasks:

- **Event Labeling** involves labeling each discrete cycle as containing an eating gesture or not. On this task we compared against Logistic Regression [LR] and the T-CRF model shown above.
- **Session Labeling** involves segmenting the discrete sequence into periods of eating and other activities. On this task we compared against the T-CRF model. For both tasks we used a 10-fold cross-validation set up with further cross-validation on the train set to tune the hyperparameters of each model.

CONCLUSIONS AND FUTURE WORK

Our method garners moderate performance improvements over baseline methods for the problem of eating detection. We plan to apply this method to other mHealth problems, such as smoking detection from respiration monitors. We believe that, given the size of the training data available in this problem, model averaging via ensemble methods or full Bayesian inference will boost and stabilize performance. Finally, given the personal nature of the activities in question, we plan to extend this framework by learning personalized models for individual subjects.

REFERENCES

\begin{itemize}
  \item [1] Jakob Verbeek and William Triggs. Some segmentation with CRFs learned from partially labeled images. 2007.
\end{itemize}