Hierarchical Nested CRFs for Segmentation and Labeling of Physiological Time Series

Roy J. Adams
College of Information and Computer Science
UMass Amherst
rjadams@cs.umass.edu

Edison Thomaz
School of Interactive Computing
Georgia Institute of Technology

Gregory D. Abowd
School of Interactive Computing
Georgia Institute of Technology

Benjamin M. Marlin
College of Information and Computer Science
UMass Amherst

In this paper we address a key problem in the emerging field of mobile health (mHealth) research [4] that we will refer to as the *hierarchical event detection and activity segmentation* problem. Given dense time series of biosignals continuously recorded using on-body sensors (e.g: respiration waveforms, electrocardiogram waveforms, actigraphy data, etc.), the goal is to infer a hierarchy of event labels at multiple temporal scales. For general time series, the goal in this problem is to both infer a hierarchical segmentation of the input time series consisting of a collection of non-overlapping spans at each level that are nested across levels, and to infer a label for each span at each level. Some examples of such labeled, nested segmentations are shown in Figure 1 (e)-(g). We present a hierarchical span-based conditional random field (CRF) framework for this problem that leverages higher-order factors to enforce the nesting constraint.

We focus in this paper on the specific problem of eating detection from actigraphy signals which has two levels: the detection of individual eating gestures (the events), and the delineation of eating sessions (activity segmentation) [5, 9]. This two-level structure appears in a variety of other mHealth problems including smoking detection, drinking detection, and conversation detection [5, 1, 6]. While the first-level detection task can be addressed using standard classification methods, the second-level segmentation task is significantly more complex due to the fact that the segments can have arbitrary lengths, and the event labels that sit below a particular activity segment can be heterogeneous (e.g: the gestures that occur during an eating segment are a mixture of both eating and non-eating gestures) [9]. As a result, prior work within the mHealth research community has either ignored the session delineation problem completely, used methods based on ad-hoc post-processing of detections, or stacked simple segmentation models like linear chain CRFs on top of the detection outputs [1, 6, 5, 9].

Our proposed framework, which we call the Hierarchical Nested Segmentation (HNS) model, instead solves the detection and segmentation problems jointly in a single unified model with a single consistent *maximum a posteriori* (MAP) inference algorithm. We present our model as a conditional random field and compare against other, similar CRF based structured prediction models (see Figure 1 (a)-(c)). Our framework allows for a variety of factors in addition to those enforcing a valid nested segmentation including high-order cardinality factors and constraints on label positions across layers. These factors can be used to model regularities in the structure of labels within and across levels. These factors make the model more expressive than standard pairwise Potts models used in computer vision and other areas [8].

For the specific case of the event detection and activity segmentation model described above, we show that it is possible to perform exact MAP inference in time quadratic in the length of the input sequence using dynamic programming. The inference algorithm is closely related to both inference in semi-Markov CRFs [7] and the inside-outside algorithm used in CRF-based parsing [2]. We leverage the MAP inference algorithm to learn the model parameters within the structured support

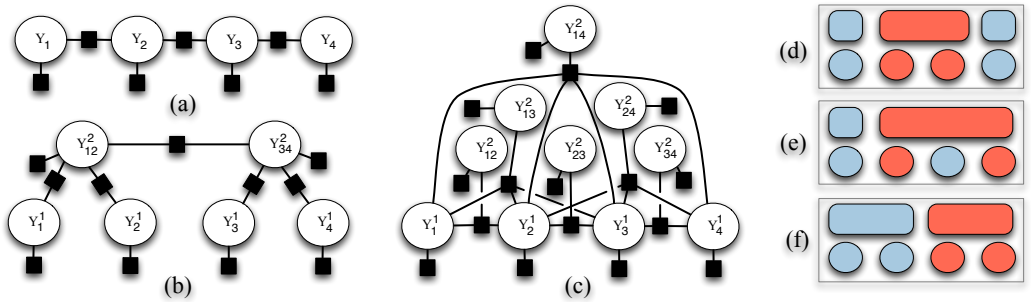

Figure 1: Figure (a) shows a factor graph model for a standard linear chain CRF over a length-four sequence. Figure (b) shows a two-level hierarchical CRF where the first level labels are grouped into fixed size blocks at the second level, which has linear chain structure. Figure (c) shows a two-level version of the proposed model, which includes a quadratic number of second level span variables, one for each possible span. The global coordinating factor that ensures a valid segmentation connects to all second-level span variables and is not pictured. Figures (d)-(f) show example segmentations and labelings for a length four sequence. Our model (c) represents a distribution over all such structures conditioned on the input features.

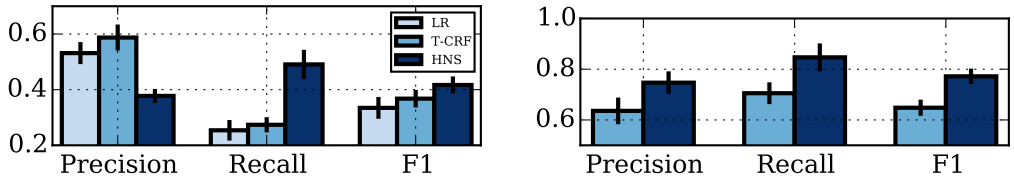

Figure 2: Average precision, recall, and F_1 results for event level (Left) and segment level (Right) prediction experiments along with one standard error bars.

vector machine (SSVM) framework [10]. We note that the model we propose may thus be equivalently viewed as a higher-order CRF or an SSVM. We choose to describe the model as a CRF and represent it graphically using a factor graph [3].

We evaluate the HNS model on the problem of eating detection in data gathered using wrist worn accelerometers. The data was originally published in [9] and contains hand labeled time series from 20 subjects performing a variety of tasks in a laboratory setting. For each subject, features were calculated on a 6 second sliding window. We test the HNS model on two detection tasks. Event detection consists of labeling each bottom level position in the sequence as an eating gesture or not. On the event labeling task, we compare the HNS model against an independent logistic regression (LR) model and the tree-structured CRF model shown in Figure 1 (b) (T-CRF). Segment level detection consists of segmenting the sequence and labeling each sequence as eating or not. On this task we compare against the T-CRF model. We calculate precision, recall, and F_1 score for each task across a 10-fold cross validation split. The average results along with one standard error bars are shown in Figure 2. The HNS outperforms all models on both tasks with the biggest improvement coming on the segmentation task. The improvement over LR on the event detection task and T-CRF on the segmentation task are significant at the $p = 0.05$ level using a paired t-test.

References

[1] Amin Ahsan Ali, Syed Monowar Hossain, Karen Hovsepian, Md Mahbubur Rahman, Kurt Plarre, and Santosh Kumar. mpuff: automated detection of cigarette smoking puffs from respiration measurements. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, pages 269–280. ACM, 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] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 47(2):498–519, 2001.

[4] Santosh Kumar, Wendy Nilsen, Misha Pavel, and Mani Srivastava. Mobile health: Revolutionizing healthcare through transdisciplinary research. *Computer*, (1):28–35, 2013.

[5] Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, and Evangelos Kalogerakis. Risq: recognizing smoking gestures with inertial sensors on a wristband. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pages 149–161. ACM, 2014.

[6] Nazir Saleheen, Amin Ahsan Ali, Syed Monowar Hossain, Hillol Sarker, Soujanya Chatterjee, Benjamin Marlin, Emre Ertin, Mustafa al'Absi, and Santosh Kumar. puffmarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 999–1010, 2015.

[7] Sunita Sarawagi and William W Cohen. Semi-markov conditional random fields for information extraction. In *Advances in Neural Information Processing Systems*, pages 1185–1192, 2004.

[8] Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Computer Vision–ECCV 2006*, pages 1–15. Springer, 2006.

[9] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 1029–1040. ACM, 2015.

[10] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. Large margin methods for structured and interdependent output variables. In *Journal of Machine Learning Research*, pages 1453–1484, 2005.