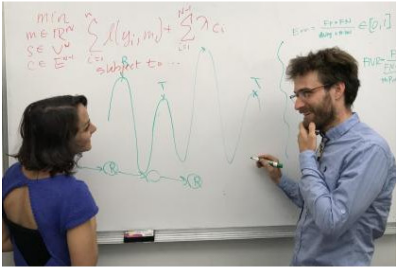
Recent advances in supervised optimal changepoint detection

Toby Dylan Hocking — toby.hocking@nau.edu
Northern Arizona University
School of Informatics, Computing, and Cyber Systems
Machine Learning Research Lab — `http://ml.nau.edu`



Come to Flagstaff!

New algorithms with constraints between adjacent segments

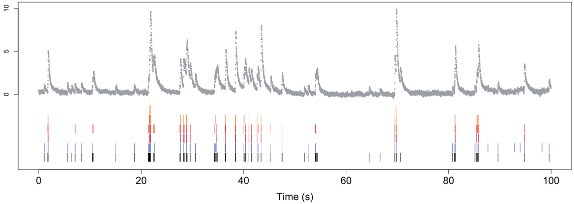
Computing optimal changepoints subject to label constraints

Learning to predict the number of changepoints

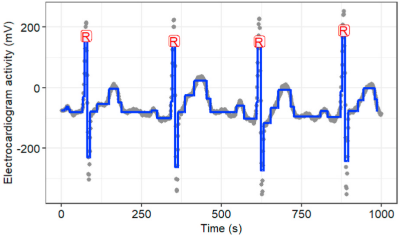
Conclusions

Changepoint detection algorithms for data over time

Neuron spikes, Jewell *et al.*, Biostatistics 2019.

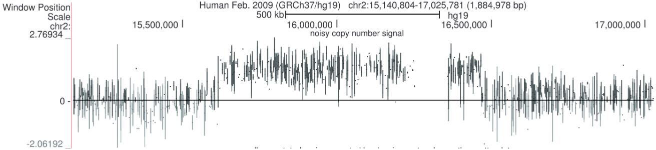


Electrocardiograms (heart monitoring), Fotoohinasab *et al.*, Asilomar 2020.

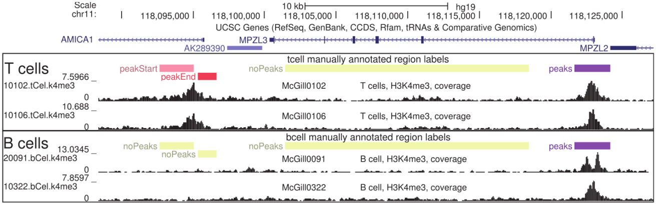


Changepoint detection algorithms for data over space

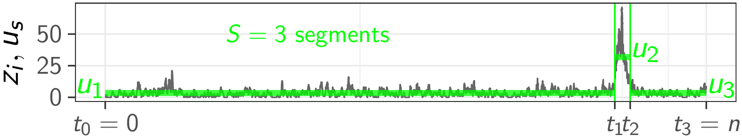
DNA copy number data for cancer diagnosis, Hocking *et al.*, Bioinformatics 2014.



Epigenomic data for understanding the human genome, Hocking *et al.*, Bioinformatics 2017.



Optimal changepoint detection problem and algorithms



$$\min_{\substack{\mathbf{u}\in\mathbb{R}^S \\ 0=t_0<t_1<\cdots<t_{S-1}<t_S=n}} \sum_{s=1}^{S} \sum_{i=t_{s-1}+1}^{t_s} \ell(u_s, z_i)$$

- Algorithm inputs n data $z_1, \dots, z_n$ and $\#$ of segments S.
- Goal is to compute best $S-1$ changepoints $t_1 < \cdots < t_{S-1}$ and S segment parameters $u_1, \dots, u_S$.
- Hard non-convex optimization problem, naïvely $O(n^S)$ time.
- Auger and Lawrence (1989): $O(Sn^2)$ time algorithm.
- Rigaill (2015): $O(n \log n)$ time, unconstrained.
- Hocking *et al.* (2020): $O(n \log n)$, directional constraints.

Constrained optimization algorithm speed

H, *et al.* Journal of Machine Learning Research 21(87):1–40, 2020.

$$\min_{\substack{\mathbf{u}\in\mathbb{R}^S \\ 0=t_0<t_1<\cdots<t_{S-1}<t_S=n}} \sum_{s=1}^{S} \sum_{i=t_{s-1}+1}^{t_s} \ell(u_s, z_i)$$

$$\text{subject to} \quad u_{s-1} \leq u_s \ \forall s \in \{2, 4, \dots\},$$
$$u_{s-1} \geq u_s \ \forall s \in \{3, 5, \dots\}.$$

Constraints used to force change up to peak state, then change down to background noise state.

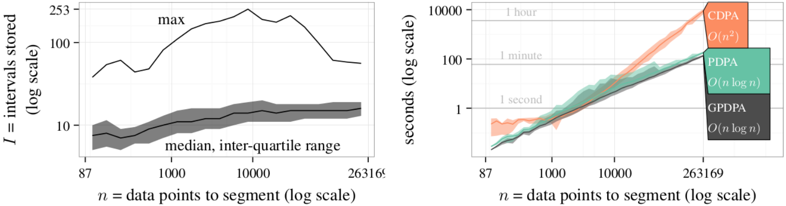
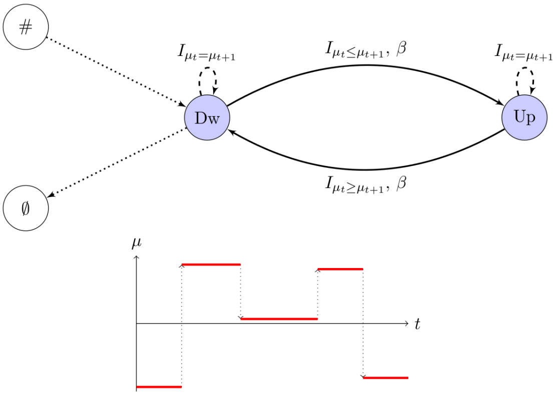


Figure 3: Empirical speed analysis on 2752 count data vectors from the histone mark ChIP-seq benchmark. For each vector we ran the GPDPA with the up-down constraint and a max of $K = 19$ segments. The expected time complexity is $O(KnI)$ where I is the average number of intervals (function pieces; candidate changepoints) stored in the $C_{k,t}$ cost functions. **Left**: number of intervals stored is $I = O(\log n)$ (median, inter-quartile range, and maximum over all data points t and segments k). **Right**: time complexity of the GPDPA is $O(n \log n)$ (median line and min/max band).

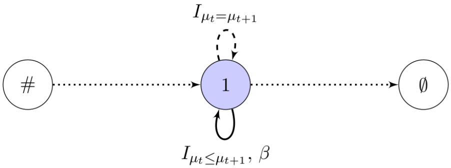
Optimization constraints defined using a graph

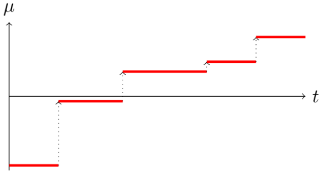
Runge V *et al.* arXiv:2002.03646.



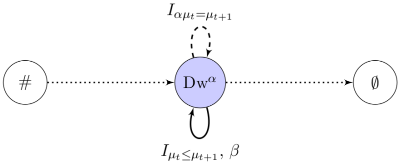
- Purple Dw/Up nodes represent hidden states.
- $\#/\emptyset$ nodes constrain start/end state.
- Edges represent possible state transitions.
- `gfpop` R package with C++ code computes optimal changepoints for user-defined constraint graphs.

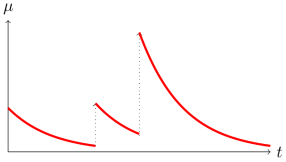
Isotonic regression (all up changes, constant segments)



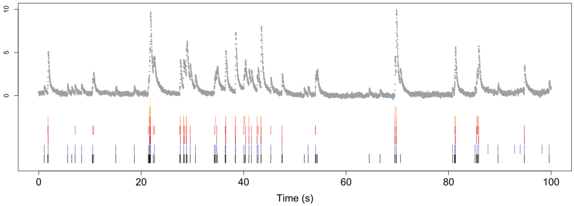


All up changes, exponential decaying segments

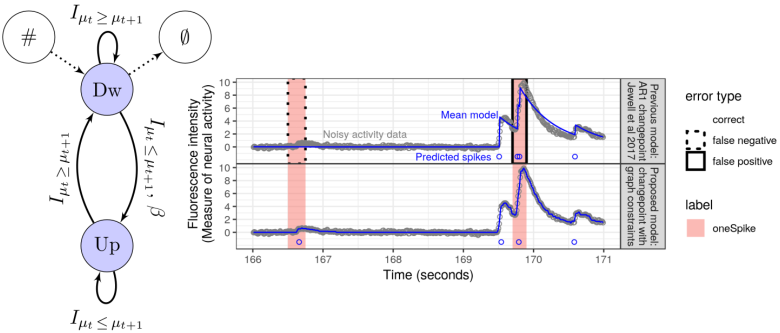




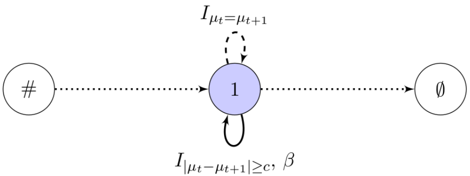
Jewell S, Hocking TD, Fearnhead P, Witten D. Fast Nonconvex Deconvolution of Calcium Imaging Data. Biostatistics (2019).

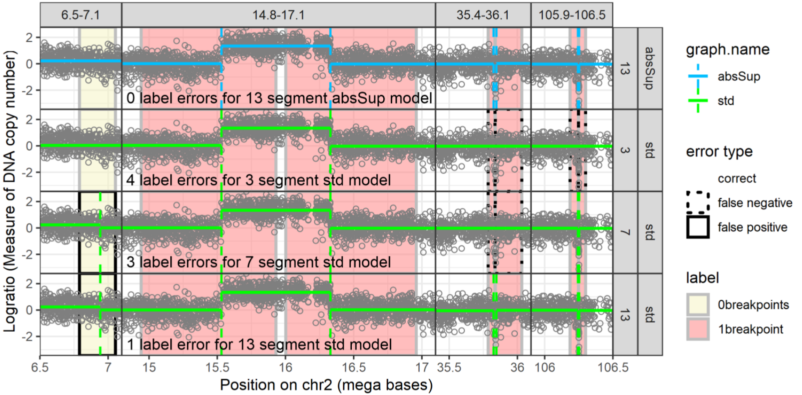


Many changes up to and down from each spike

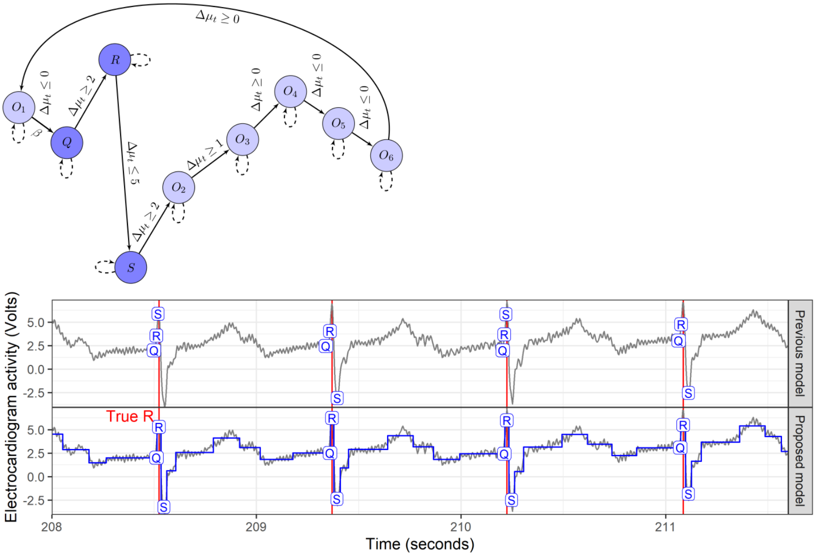


Relevant changes (any direction, large in absolute value)





Complex graph for electrocardiogram data



Fotoohinasab *et al.*, Asilomar conference 2020.

New algorithms with constraints between adjacent segments

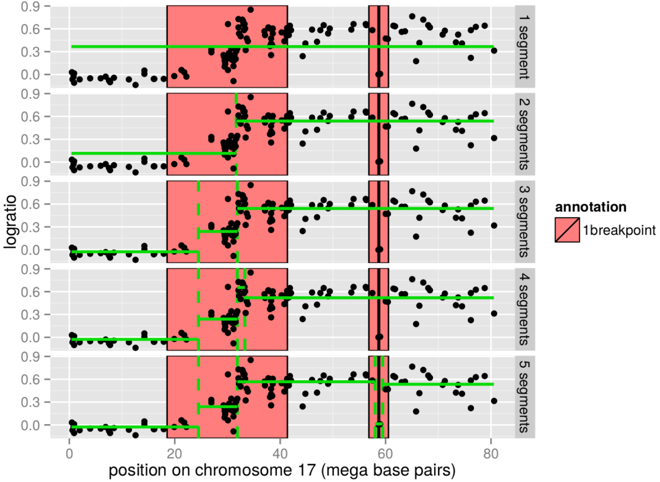
Computing optimal changepoints subject to label constraints

Learning to predict the number of changepoints

Conclusions

What if no models agree with expert labels?

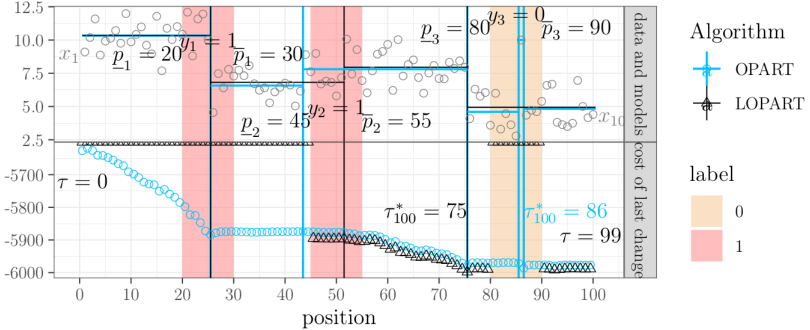
Hocking and Rigaill, Pre-print hal-00759129.



- Want: one changepoint in each label (red rectangle).
- No model is consistent with all three labels.

Using expert labels as optimization constraints

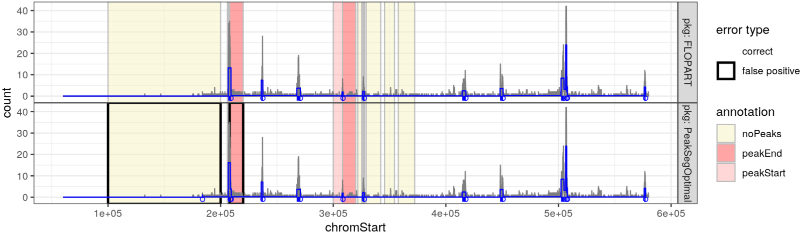
Hocking and Srivastava, Pre-print arXiv:2006.13967.



- Previous OPART model (blue) ignores labels (two errors).
- Main idea: add optimization constraints to ensure that there is the right number of changepoints predicted in each label.
- Proposed LOPART model (black) consistent with labels.

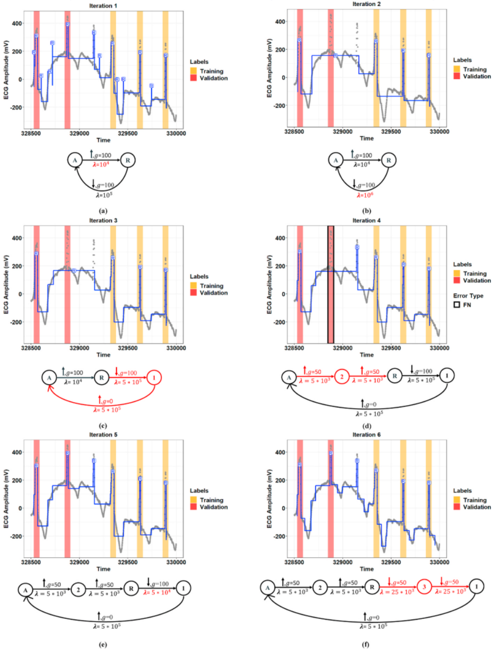
Label constraints and up-down constraints

Stenberg and Hocking, in progress.



- Previous PeakSegOptimal algorithm (bottom) ignores labels (two errors).
- Proposed FLOPART model (top) consistent with labels, and interpretable in terms of peaks and background.

Learning a complex graph using labels



- Fotoohinasab *et al.*, 2021.
- Simple initial graph is iteratively edited (red) to agree with expert labeled regions (orange rectangles).
- Easier for expert to provide labels than graph.

New algorithms with constraints between adjacent segments

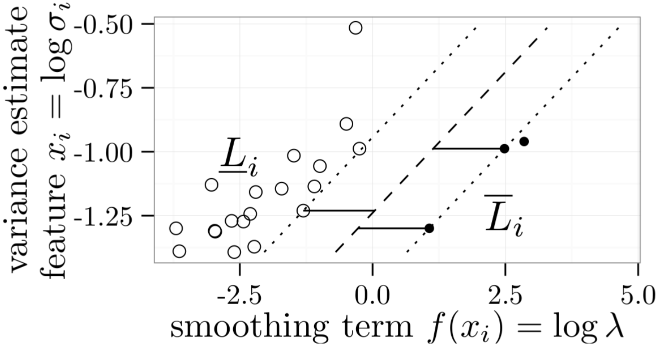
Computing optimal changepoints subject to label constraints

Learning to predict the number of changepoints

Conclusions

Max margin interval regression problem similar to SVM

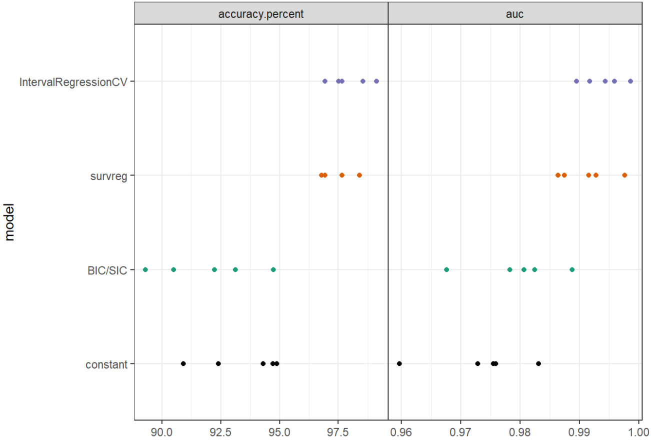
Hocking *et al.*, 2013.



- Train on several data sequences with labels (dots).
- Want to compute function between white and black dots.
- SVM margin is multi-dimensional (diagonal).
- Here margin to maximize is one-dimensional (horizontal).
- Learned function predicts number of changepoints/segments.

Test accuracy/AUC in five-fold cross-validation

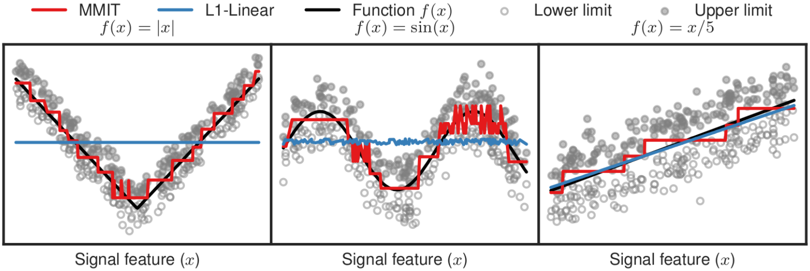
Hocking and Killick, useR2017 conference tutorial.



Learned linear functions for predicting the number of changepoints (IntervalRegressionCV, survreg) are much more accurate than constant baseline and unsupervised BIC/SIC.

How to predict

Drouin *et al.*, Neural Information Processing Systems 2017.



- Generalization of classical CART regression tree learning algorithm.
- Can learning non-linear functions of inputs.
- More recently we implemented a similar idea in xgboost, Barnwal *et al.*, Pre-print arXiv:2006.04920.

New algorithms with constraints between adjacent segments

Computing optimal changepoints subject to label constraints

Learning to predict the number of changepoints

Conclusions

Conclusions

- Optimal detection of $S-1$ changepoints in n data is naively a $O(n^S)$ computation.
- Functional pruning method yields algorithms with worst case time complexity of $O(n^2)$ (same as classical dynamic programming).
- Empirically the functional pruning algorithms are much faster, $O(n \log n)$.
- Only one proof of average time complexity for 1 changepoint and the uniform loss function (never used in practice).
- Would be interesting to prove $O(n \log n)$ average time complexity in other more realistic situations. (square/Poisson loss, λ) How?
- Let's collaborate! toby.hocking@nau.edu