Motivation
Impact of the n-gram extension
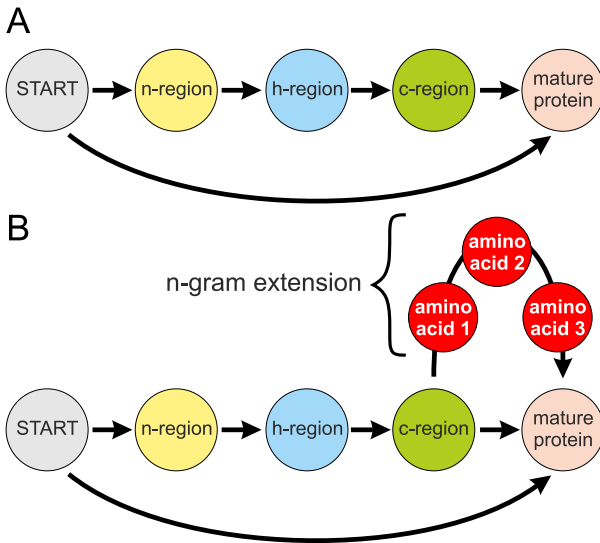Greedy n-gram choosing algorithm

Expanding signalHsmm using n-grams

Michał Burdukiewicz, Piotr Sobczyk

University of Wrocław, Department of Genomics, Poland

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

Outline

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

To improve our model, we added few supplementary states representing specific motives that might occur in the proximity of cleavage site. The structure of cleavage sites, more conserved than other parts of signal peptide, may be reflected by n-grams (k-mers), short vectors of n characters derived from input sequences.

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

The diagram of simple (A) and extended version of signalHsmm with the n-gram cleavage site model (B).

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

Outline

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm
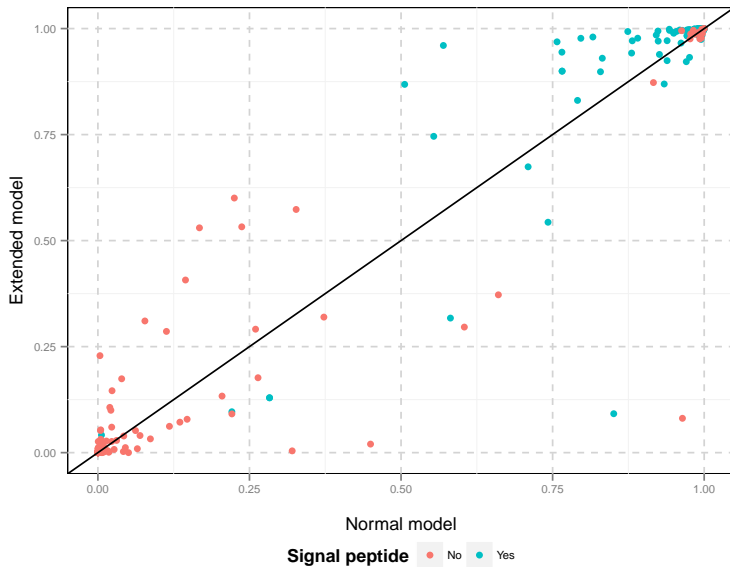
Performance of different classifiers.

	AUC	TP	TN	FP	FN
signalPnotm	0.9416	208	195	19	6
signalPtm	0.9673	205	209	5	9
predsi	0.8949	194	189	25	20
phobius	0.9509	207	200	14	7
philius	0.9369	204	197	17	10
signalHsmm2010	0.9526	198	191	23	16
signalHsmm1989	0.9562	202	194	20	12
signalKmer	0.9695	206	194	20	8

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

Outline

Motivation
Impact of the n-gram extension
Greedy n-gram choosing algorithm

bla