

Name: Kelly Jones

Project: Classification of snoRNA Types Using Classification Relevance Unit Machines

Description

As the number of sequenced snoRNAs becomes larger, it is possible to more accurately determine snoRNAs via computational methods. The Classification Relevance Unit Machine (CRUM) is a recently developed variation on the Support Vector Machine (SVM). While a Gaussian SVM approach can give better accuracy, CRUM offers a decreased model size and increased speed in both training and classification. Since nextgen sequencing methods can yield more sequences, modeling and classification speed could become a useful trait in snoRNA classification.

I would like to do a comparison with existing software as well as add folding features to the classifier. Since much of the CRUM code already exists and the feature selection and training portions of the experiment have been completed so I would like to continue writing a C++ version of CRUM's training as the class continues and I learn more about the workings of classifiers (Particularly kmeans clustering and kernel methods). Also, Since much of the work for generating new models with CRUM comes from feature selection, I'd also like to create a web based tool for generating feature vectors for use with CRUM.

Bibliography

Jana Hertel et al (2007) *SnoReport: computational identification of snoRNAs with unknown targets* *Bioinformatics* Vol. 24 no. 2 2008, pages 158–164

Mark Menor et al (2011) *Relevance Units Machine for Classification* 2011 BMEI

Yask Gupta et al (2014) *ptRNApred: computational identification and classification of post-transcriptional RNA* *Nucleic Acids Research*

<http://nar.oxfordjournals.org/content/early/2014/10/10/nar.gku918.short>

Schedule of work

This week

Crossfold validation of current snoRNA CRUM

Now:Oct

Add folding data as a feature for crum

Now:Nov

Rewrite crum training program

Nov:End

Compare crum-snoRNA classifier with existing methods

As possible

Create web interface for easy DNA/RNA feature selection so users can generate their own models