CARNEGIE MELLON UNIVERSITY - COMPUTATIONAL BIOLOGY DEPARTMENT

Active Learning for Drug Selection on Identified Target Protein

Christine Baek
christib@andrew.cmu.edu

Qi Chu
qchu@andrew.cmu.edu

1 INTRODUCTION

In this project, we explore three datasets of different noise level, for identification compounds, or drugs that bind to specific target protein associated with disease. We use DHM as our active learning strategy to determine when to query the Oracle, and SVM to train our model on the oracle-obtained as well as inferred labels.

2 METHODS

2.1 ACTIVE LEARNING STRATEGY

3 METHODS

3.1 BASE LEARNER STRATEGY

TODO: DHM

- why chosen
- any modifications/source?

3.2 CLASSIFIER STRATEGY

TODO: SVM

- why chosen
- any modifications/source?

4 Results

4.1 Easy

TODO: Test error curve, f1 score curve, test error/calls to oracle curve

4.2 Moderate

TODO: Test error curve, f1 score curve, test error/calls to oracle curve

4.3 Difficult

TODO: Test error curve, f1 score curve, test error/calls to oracle curve
TODO: error curve (Traint,test)
TODO: f1 score curve (train, test)
TODO: error / num calls to oracle (train, test)

4.4 Moderate

TODO: error curve (Traint,test)
TODO: f1 score curve (train, test)
TODO: error / num calls to oracle (train, test)

4.5 Difficult

TODO: error curve (Traint,test)
TODO: f1 score curve (train, test)
TODO: error / num calls to oracle (train, test)

5 Conclusion

TODO: briefly summarize

References

[1] S. Dasgupta, D. Hsu, C. Monteleoni, *A general agnostic active learning algorithm*, NIPS, 2008

[2] S. Dasgupta, *Two faces of active learning*, `http://cseweb.ucsd.edu/~dasgupta/papers/`, 2010