

Combining Word Embeddings of Protein
Sequences with Evolutionary Information for
Secondary Structure Prediction Using Deep
Neural networks
Bachelor's thesis in Informatics WS18/19

Anna Reithmeir

December 4, 2018

Contents

1	Introduction	4
1.1	Problem Statement	5
1.2	Motivation and Objectives	5
1.3	Outline	5
2	Background	5
2.1	Neural Networks	5
2.1.1	CNNs	5
2.1.2	LSTMs	5
2.2	Structural Feature Prediction of Proteins with Neural Networks .	5
2.3	ProtVec representation	5
3	Related Work	5
4	Implementation	5
4.1	Overview	5
4.2	Data preprocessing	5
4.3	Training Parameters and Procedure	5
5	Evaluation	5
5.1	Results	5
5.1.1	Performance of CNN	5
5.1.2	Performance of LSTM	5
5.2	Runtime and Memory	5
5.3	Discussion	5
5.4	Limitations	5
6	Summary	5
6.1	Conclusion	5
6.2	Outlook	5
	Appendices	5
A	Appendix A	5
B	Appendix B	6

Abstract

Abstract goes here

1 Introduction

Intro goes here

1.1	Problem Statement
1.2	Motivation and Objectives
1.3	Outline
2	Background
2.1	Neural Networks
2.1.1	CNNs
2.1.2	LSTMs
2.2	Structural Feature Prediction of Proteins with Neural Networks
2.3	ProtVec representation
3	Related Work
4	Implementation
4.1	Overview
4.2	Data preprocessing
4.3	Training Parameters and Procedure
5	Evaluation
5.1	Results
5.1.1	Performance of CNN
5.1.2	Performance of LSTM
5.2	Runtime and Memory
5.3	Discussion
5.4	Limitations
6	Summary
6.1	Conclusion
6.2	Outlook

Appendices

5

A Appendix A

Appendix A

B Appendix B

Appendix B