Predição de Estruturas Secundárias de Proteínas usando Aprendizado de Máquina e BLAST

Gabriel Bianchin de Oliveira

Orientador: Prof. Dr. Zanoni Dias
Coorientador: Prof. Dr. Hélio Pedrini

11 de Março de 2021

Instituto de Computação
Universidade Estadual de Campinas

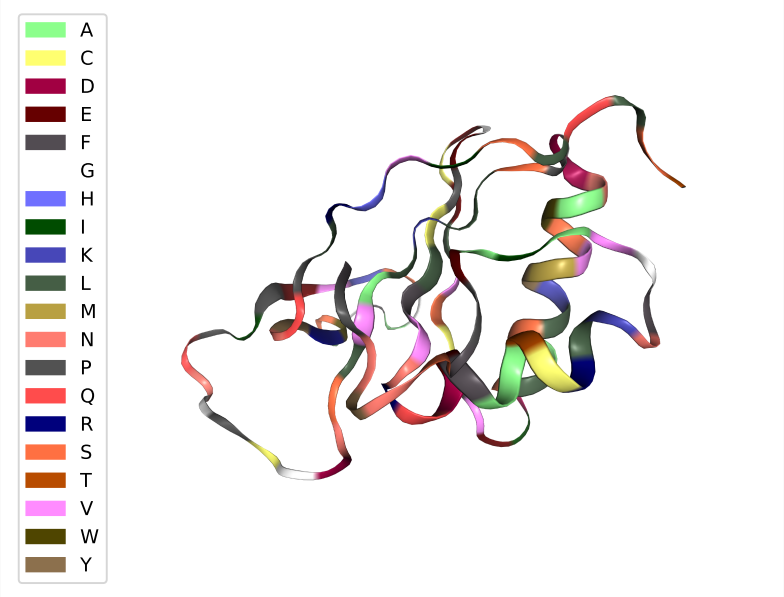
Sumário

Sumário

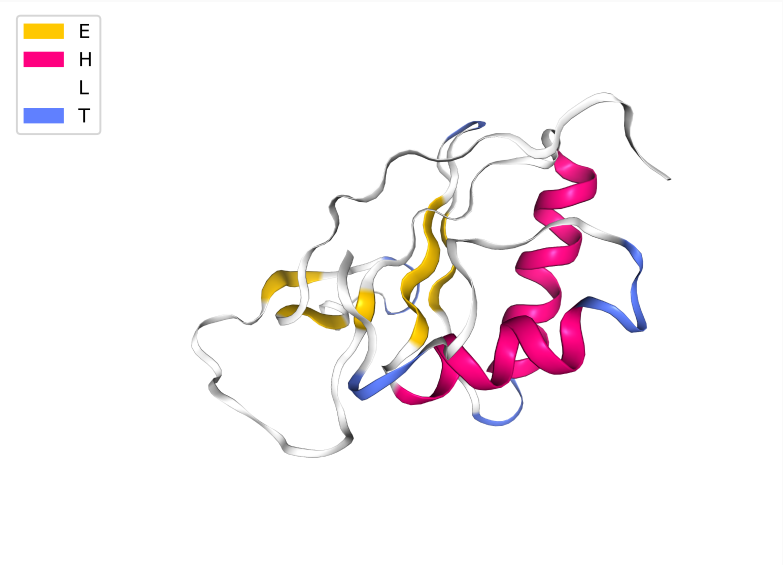
Introdução

- Importância de proteínas em processos biológicos
- Aminoácidos
- Estruturas tridimensionais das proteínas

Introdução



(a) Aminoácidos



(b) Estruturas Secundárias

Figura 1: Aminoácidos e estruturas secundárias da proteína PDB ID: 6BI6

Introdução

- Impacto das estruturas tridimensionais
- Métodos laboratoriais para a determinação de estruturas secundárias
- Diferença do número de proteínas sequenciadas e proteínas com estruturas definidas

Introdução

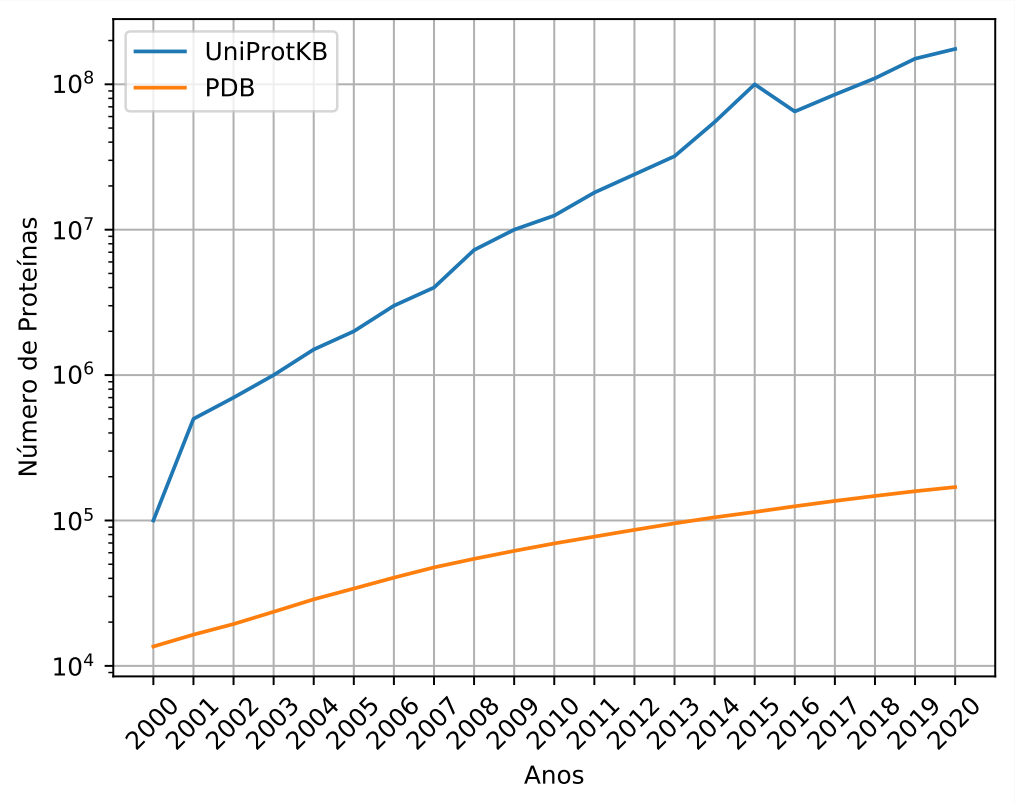


Figura 2: Diferença entre o número de proteínas sequenciadas e proteínas com estruturas definidas nos últimos 20 anos

Introdução

- Predição de estruturas secundárias com métodos computacionais
- Métodos livres de modelo
  - Sequência de aminoácidos no formato *one-hot encoding* e informações evolutivas
  - Classificadores locais e globais
  - Poucos métodos podem classificar estruturas usando apenas as sequências de aminoácidos
- Métodos baseados em modelo
  - Proteínas homólogas
  - BLAST [1]

Objetivos

Propor, implementar e avaliar métodos usando classificadores de aprendizado de máquina e BLAST para predizer estruturas secundárias de proteínas

- Definição de métodos livres de modelos usando diferentes algoritmos com classificação local, global e local-global
- Avaliação da fusão de diferentes classificadores livres de modelo
- Investigação do BLAST como um método baseado em modelo
- Avaliação da fusão entre métodos livres de modelo e baseados em modelo

Questões de Pesquisa

1. Utilizar apenas sequência de aminoácidos pode produzir resultados próximos aos resultados utilizando sequência de aminoácidos e informações evolutivas?
2. A transformação do vetor *one-hot encoding* esparso em um vetor denso pode ajudar na classificação de estruturas secundárias das proteínas?
3. Qual é o impacto da fusão de classificadores locais e globais na classificação?
4. A fusão entre métodos baseados em modelo e livres de modelo pode melhorar os resultados?

Contribuições

- Investigação da importância da Matriz de Pontuação de Posição Específica (PSSM) na predição de estruturas secundárias de proteínas
- Desenvolvimento e avaliação de um método de fusão, que chamamos de Sacola de Otimizadores
- Avaliação de *Transformers* no problema de predição de estruturas secundárias de proteínas
- Avaliação de fusões entre métodos livres de modelo e baseados em modelo

Sumário

Trabalhos Relacionados - Primeira Fase

- Em 1974, Chou e Fasman [2] propuseram o primeiro trabalho sobre predição de estruturas de proteínas
- Métodos de regras quantitativas e qualitativas e medidas estatísticos
- Comparação de proteínas homólogas, como Levin *et al.* [3] e Nishikawa e Ooi [4]
- Baixo poder computacional e bases de dados pequenas

Trabalhos Relacionados - Segunda Fase

- Aumento do poder computacional e criação de grandes bases de proteínas
- Métodos utilizando classificadores baseados em janela deslizante
- Holley e Karplus [5] propuseram o primeiro algoritmo baseado em redes neurais
- Desenvolvimento do BLAST e PSI-BLAST nos anos 1990
- Jones [6] verificou a utilização da Matriz de Pontuação de Posição Específica (PSSM)
- Przybylski e Rost [7] avaliaram o BLAST e PSI-BLAST para a predição de estruturas secundárias
- Aumento da acurácia Q3 e utilização da acurácia Q8
- Limitação de análise por janelas

Trabalhos Relacionados - Terceira Fase

- Grande evolução do poder computacional nos anos 2010
- Redes neurais convolucionais e recorrentes
- Fusão de métodos locais e globais
- Transformação da sequência de aminoácidos em um vetor denso, como utilizado em Guo *et al.* [10] e Ratul *et al.* [8]
- Método baseado em modelo pouco explorado

Sumário

Base CB6133

- Base com 6.133 proteínas com até 700 aminoácidos
- Proteínas com menos de 30% de similaridade
- 5.600 proteínas para treinamento, 256 proteínas para validação e 272 proteínas para teste

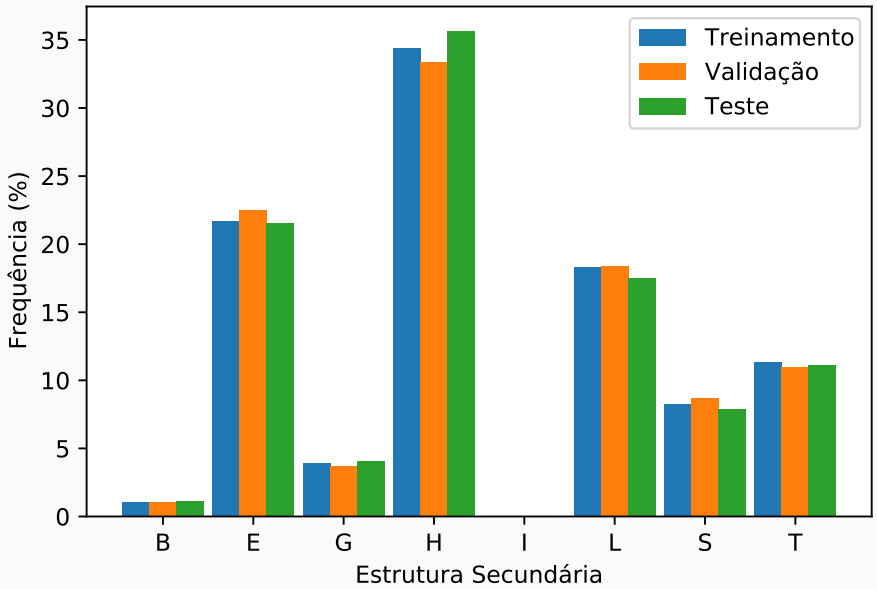


Figura 3: Distribuição das classes nos conjuntos de treinamento, validação e teste na base de dados CB6133

Base CB513

- Base com 513 proteínas usada para teste
- Base CB6133 filtrada utilizada para treinamento e validação
- 5.278 proteínas para treinamento, 256 proteínas para validação e 513 proteínas para teste

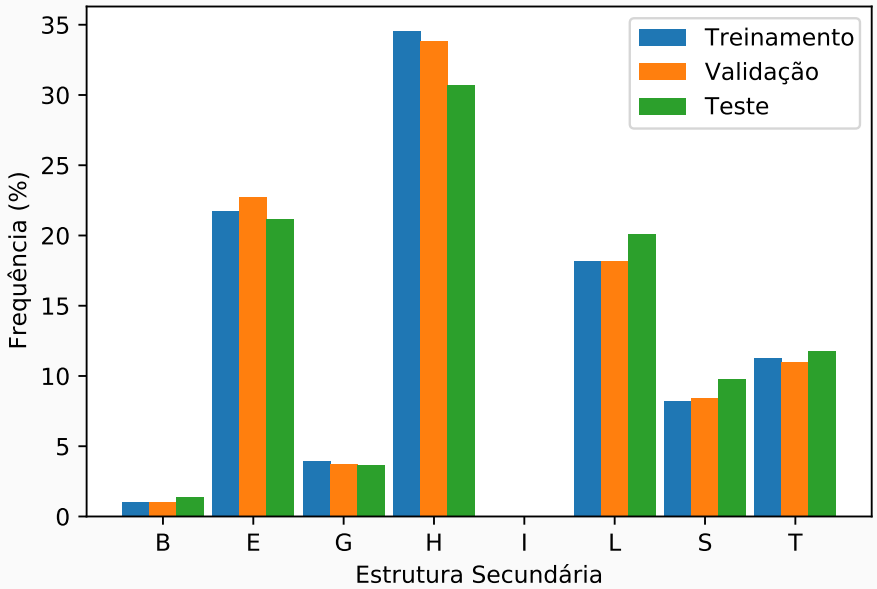


Figura 4: Distribuição das classes nos conjuntos de treinamento, validação e teste na base de dados CB513

Base PDB - Classificação Q8

- Proteínas depositadas no PDB em 2018 com até 700 aminoácidos
- Pré-processamento nos aminoácidos
- 6.979 proteínas para treinamento, 500 proteínas para validação e 500 proteínas para teste

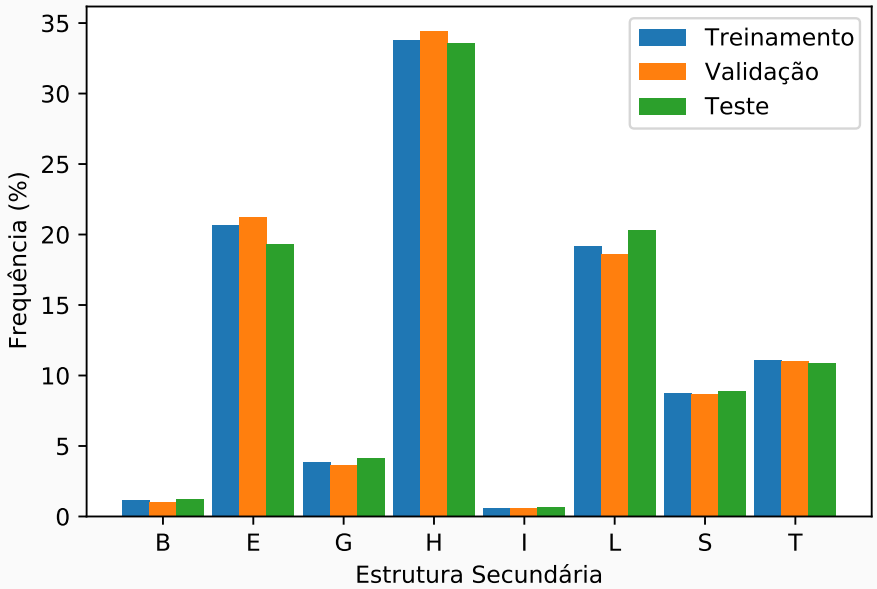


Figura 5: Distribuição das classes nos conjuntos de treinamento, validação e teste na base de dados PDB para a classificação Q8

Base PDB - Classificação Q3

- Transformação da classificação Q8 em classificação Q3
- 6.979 proteínas para treinamento, 500 proteínas para validação e 500 proteínas para teste
- 20 características por aminoácido

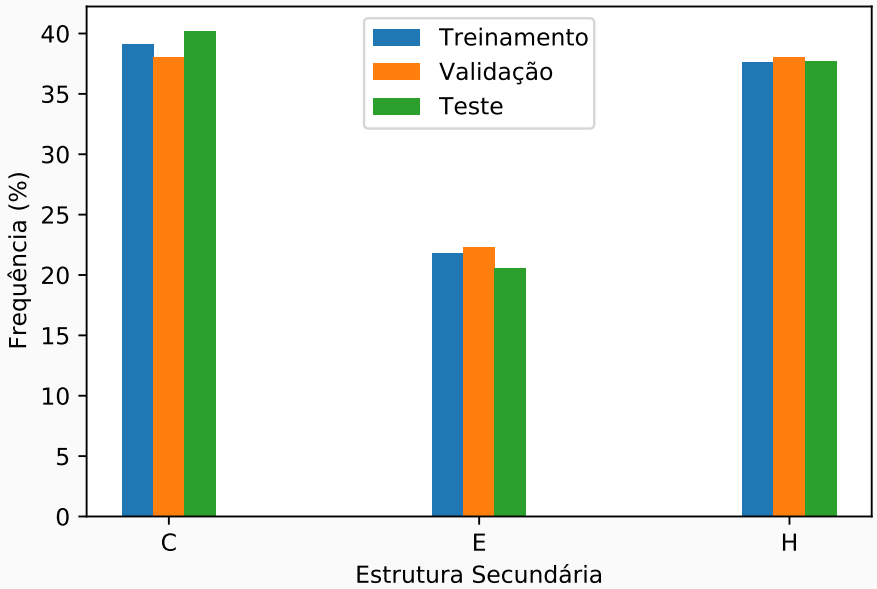


Figura 6: Distribuição das classes nos conjuntos de treinamento, validação e teste na base de dados PDB para a classificação Q3

Métricas de Avaliação

- Precisão
- Revocação
- Acurácia Q3
- Acurácia Q8

Sumário

Redes Neurais Recorrentes

- RNNs como classificadores globais
- RNNs com camadas de *embedding*, 600 neurônios por camada e entre 2 e 6 camadas

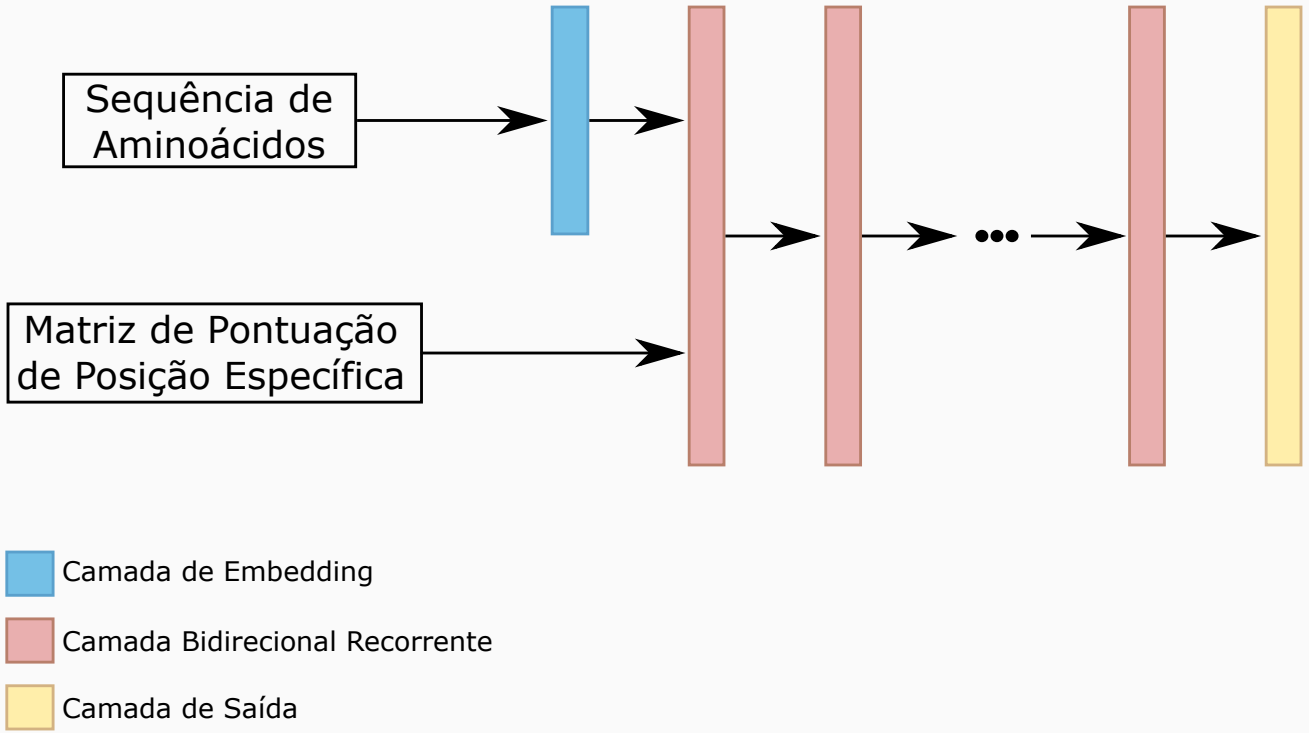


Figura 7: Arquitetura geral da RNN

Redes Neurais Recorrentes

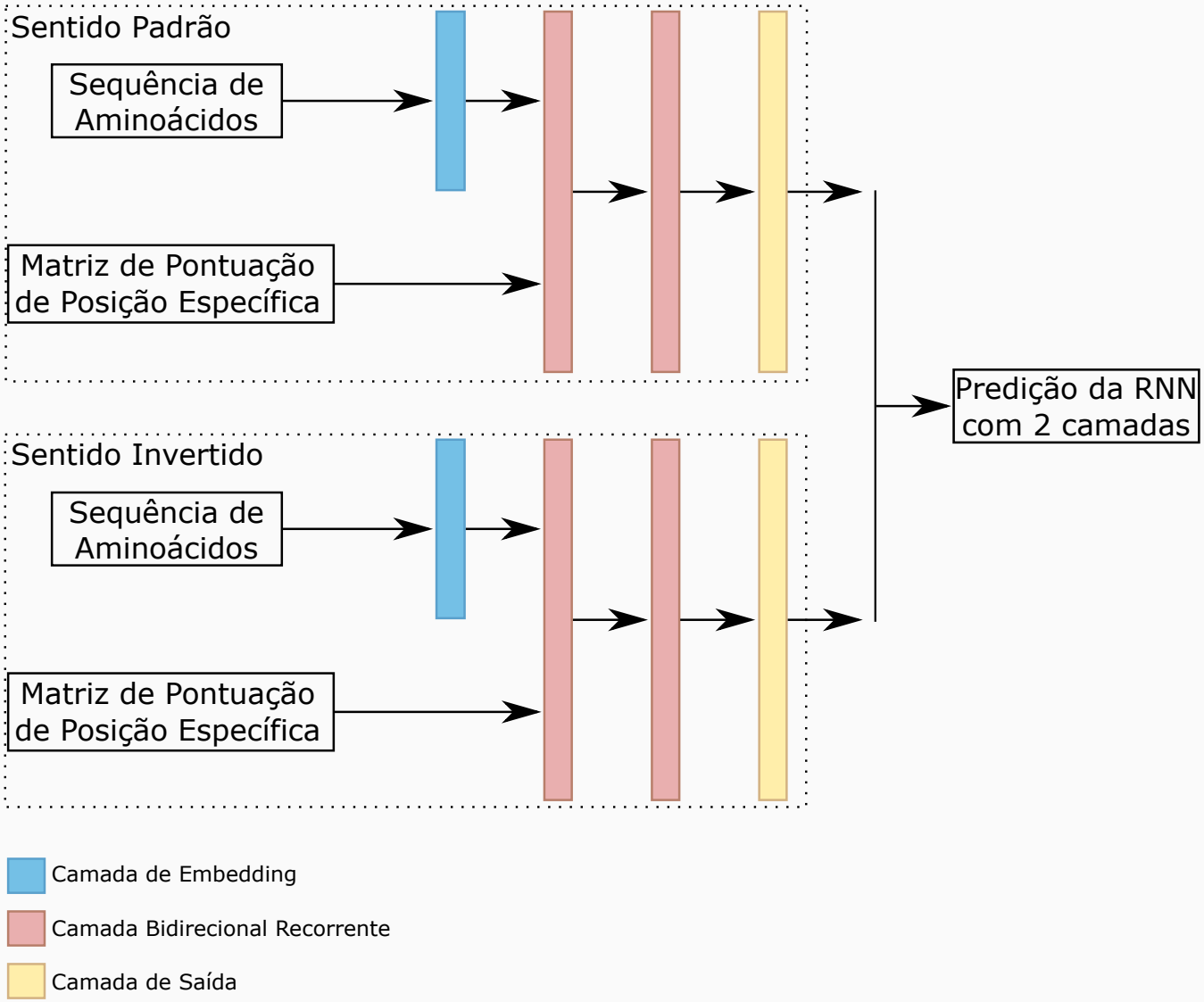


Figura 8: Fusão de RNNs com duas camadas bidirecionais recorrentes

Florestas Aleatórias

- RFs como classificadores locais
- RFs com janelas de tamanho 9 até 17 (9, 11, 13, 15 e 17)
- Preenchimento com 0

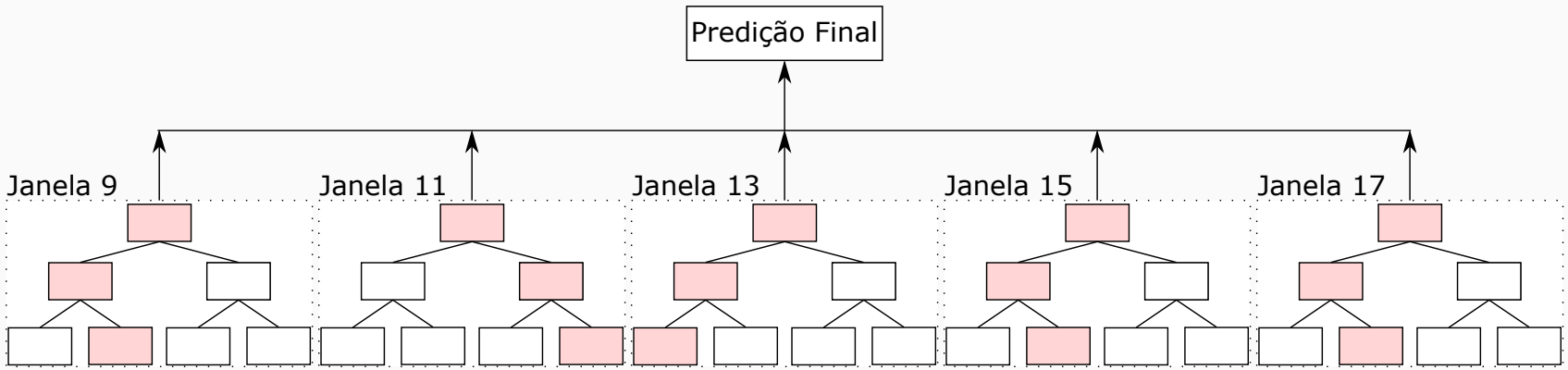


Figura 9: Fusão entre florestas aleatórias

Blocos Inception-v4

- BIv4s como classificadores locais e globais
- Convolução unidimensional e sem camadas de *pooling* entre blocos *inception*
- BIv4s com camadas de *embedding* e com 3 até 7 blocos do tipo B

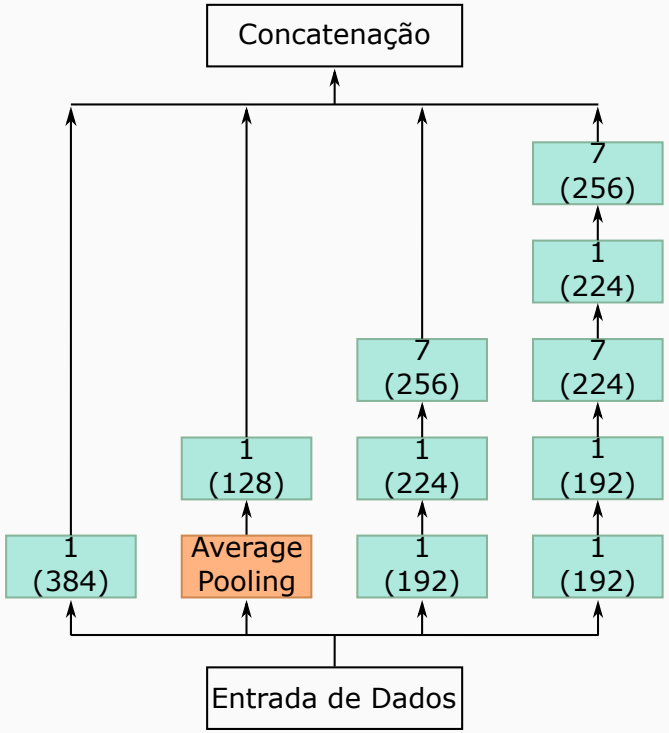


Figura 10: Bloco B com convoluções unidimensionais

Redes Inception Recorrentes

- RIRs como classificadores locais seguidos por classificadores puramente globais
- RIRs com camadas de *embedding*, 3 até 7 blocos do tipo B e 3 camadas bidirecionais recorrentes com 100 neurônios por camada

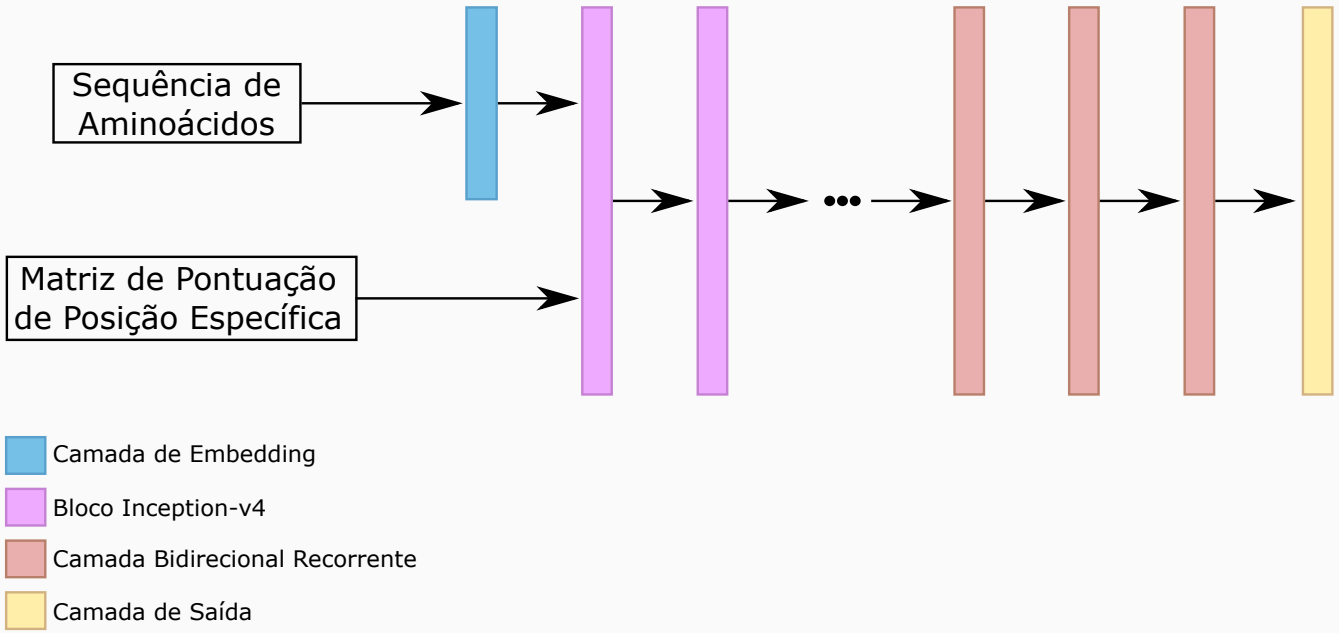


Figura 11: Arquitetura geral da RIR

Transformers

- *Transformers* em tarefas biológicas
- Transformação dos aminoácidos em *tokens*
- BERT com janelas deslizantes de tamanho igual a 101, 121, 141, 161 e 181

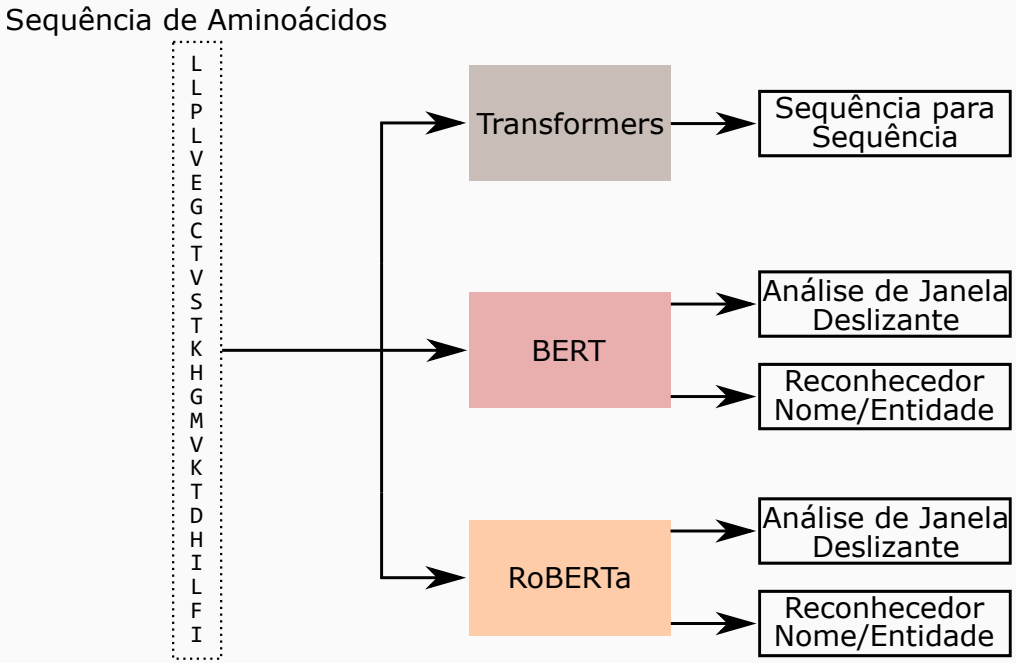


Figura 12: Arquiteturas baseadas em *Transformers* avaliadas

Fusão dos Métodos Livres de Modelo

- Fusão entre RNN, RF, RIR, BIv4 e BERT

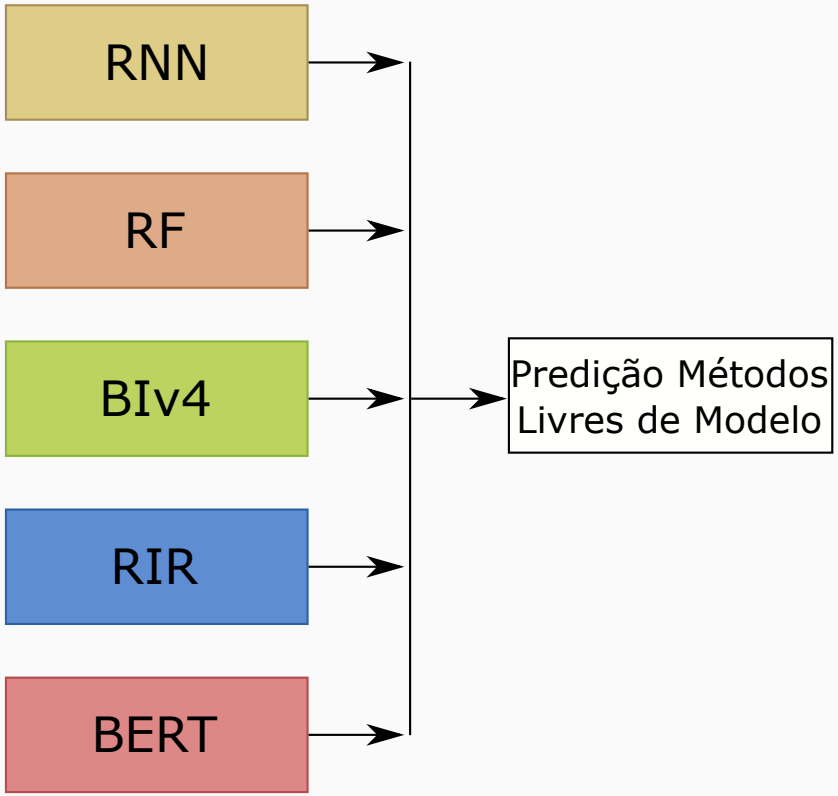


Figura 13: Fusão dos métodos livres de modelo

BLAST

- BLAST como busca entre proteínas homólogas
- Todas as proteínas do PDB como base de busca
- Bons Alinhamentos e Alinhamentos Gerais

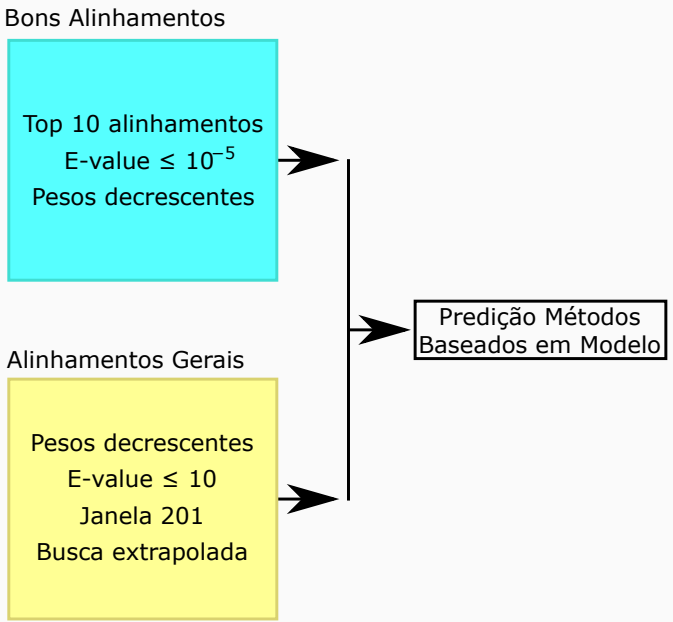


Figura 14: Fusão entre Bons Alinhamentos e Alinhamentos Gerais

Sacola de Otimizadores

- Algoritmo Genético
- Busca Cuco
- Otimização por Enxame de Partículas

Sumário

Fusão dos Métodos Livres de Modelo - CB6133

Método	Acurácia Q8 (%)
Fusão	**62,69**
RNN	61,63
RIR	60,99
BIv4	60,40
BERT	59,60
RF	54,20

Tabela 1: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base CB6133

Fusão dos Métodos Livres de Modelo - CB513

Método	Acurácia Q8 (%)
Fusão	**57,81**
RIR	57,68
RNN	57,58
BIv4	56,73
BERT	55,86
RF	49,64

Tabela 2: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base CB513

Fusão dos Métodos Livres de Modelo - CB513

Método	Acurácia Q8 (%)
Ratul *et al.* [8]	61,6
Nosso método	**57,8**
Guo *et al.* [11]	57,1

Tabela 3: Comparação do resultado obtido com outros trabalhos da literatura no conjunto de teste da base CB513

Fusão dos Métodos Livres de Modelo - PDB Q8

Método	Acurácia Q8 (%)
Fusão	**74,76**
RIR	71,98
BIv4	71,64
RF	69,75
RNN	68,48
BERT	63,04

Tabela 4: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base PDB para a classificação Q8

Fusão dos Métodos Livres de Modelo - PDB Q3

Método	Acurácia Q3 (%)
Fusão	**83,03**
RIR	81,91
RNN	79,22
RF	78,65
BIv4	78,26
BERT	75,68

Tabela 5: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base PDB para a classificação Q3

Fusão dos Métodos Baseados em Modelo - CB6133

Método	Acurácia Q8 (%)
Fusão	**78,73**
Alinhamentos Gerais	75,96
Bons Alinhamentos	69,15

Tabela 6: Resultados dos métodos individuais e da fusão dos métodos baseados em modelo no conjunto de teste da base CB6133

Fusão dos Métodos Baseados em Modelo - CB513

Método	Acurácia Q8 (%)
Fusão	**89,16**
Alinhamentos Gerais	86,43
Bons Alinhamentos	76,34

Tabela 7: Resultados dos métodos individuais e da fusão dos métodos baseados em modelo no conjunto de teste da base CB513

Fusão dos Métodos Baseados em Modelo - CB513

Método	Acurácia Q8 (%)
Nosso método	**89,1**
Ratul *et al.* [8]	61,6
Guo *et al.* [11]	57,1

Tabela 8: Comparação do resultado com outros trabalhos da literatura no conjunto de teste da base CB513

Fusão dos Métodos Baseados em Modelo - PDB Q8

Método	Acurácia Q8 (%)
Fusão	**88,55**
Alinhamentos Gerais	86,36
Bons Alinhamentos	68,63

Tabela 9: Resultados dos métodos individuais e da fusão dos métodos baseados em modelo no conjunto de teste da base PDB para a classificação Q8

Fusão dos Métodos Baseados em Modelo - PDB Q3

Método	Acurácia Q3 (%)
Fusão	**93,21**
Alinhamentos Gerais	91,74
Bons Alinhamentos	82,11

Tabela 10: Resultados dos métodos individuais e da fusão dos métodos baseados em modelo no conjunto de teste da base PDB para a classificação Q3

Fusão Final - CB6133

Método	Acurácia Q8 (%)
Fusão Horizontal	**82,83**
Fusão Hierárquica	82,61
Fusão dos Métodos Baseados em Modelo	78,73
Alinhamentos Gerais	75,96
Bons Alinhamentos	69,15
Fusão dos Métodos Livres de Modelo	62,69
RNN	61,63
RIR	60,99
BIv4	60,40
BERT	59,60
RF	54,20

Tabela 11: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base CB6133

Fusão Final - CB513

Método	Acurácia Q8 (%)
Fusão dos Métodos Baseados em Modelo	**89,16**
Fusão Horizontal	88,79
Fusão Hierárquica	88,69
Alinhamentos Gerais	86,43
Bons Alinhamentos	76,34
Fusão dos Métodos Livres de Modelo	57,81
RIR	57,68
RNN	57,58
BIv4	56,73
BERT	55,86
RF	49,64

Tabela 12: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base CB513

Fusão Final - PDB Q8

Método	Acurácia Q8 (%)
Fusão Horizontal	**89,66**
Fusão Hierárquica	89,60
Fusão dos Métodos Baseados em Modelo	88,55
Alinhamentos Gerais	86,36
Fusão dos Métodos Livres de Modelo	74,76
RIR	71,98
BIv4	71,64
RF	69,75
Bons Alinhamentos	68,63
RNN	68,48
BERT	63,04

Tabela 13: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base PDB para a classificação Q8

Fusão Final - PDB Q3

Método	Acurácia Q3 (%)
Fusão Hierárquica	**93,74**
Fusão Horizontal	93,72
Fusão dos Métodos Baseados em Modelo	93,21
Alinhamentos Gerais	91,74
Fusão dos Métodos Livres de Modelo	83,03
Bons Alinhamentos	82,11
RIR	81,91
RNN	79,22
RF	78,65
BIv4	78,26
BERT	75,68

Tabela 14: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base PDB para a classificação Q3

Sumário

Fusão dos Métodos Livres de Modelo - CB6133

Método	Acurácia Q8 (%)
Fusão	**76,67**
RNN	76,09
BIv4	75,42
RIR	75,29
RF	69,30
BERT	59,60

Tabela 15: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base CB6133

Fusão dos Métodos Livres de Modelo - CB6133

Método	Acurácia Q8 (%)
Ratul *et al.* [8]	76,9
Nosso método	**76,7**
Drori *et al.* [12]	76,3
Gou *et al.* [10]	75,7
Johansen *et al.* [13]	74,8
Guo *et al.* [11]	74,2
Zhou *et al.* [14]	74,0
Zhou e Troyanskaya [15]	72,1

Tabela 16: Comparação do resultado obtido com outros trabalhos da literatura no conjunto de teste da base CB6133

Fusão dos Métodos Livres de Modelo - CB513

Método	Acurácia Q8 (%)
Fusão	**71,48**
RIR	71,16
BIv4	70,93
RNN	69,88
RF	64,60
BERT	55,86

Tabela 17: Resultados dos métodos individuais e da fusão dos métodos livres de modelo no conjunto de teste da base CB513

Fusão dos Métodos Livres de Modelo - CB513

Método	Acurácia Q8 (%)
Ratul *et al.* [8]	71,9
Nosso método	**71,5**
Busia *et al.* [16]	71,4
Uddin *et al.* [17]	70,9
Johansen *et al.* [13]	70,9
Drori *et al.* [12]	70,7
Fang *et al.* [9]	70,6
Zhou *et al.* [14]	70,3
Gou *et al.* [10]	70,2
Li e Yu [18]	69,4
Lin *et al.* [19]	68,4
Wang *et al.* [20]	68,2
Zhou e Troyanskaya [15]	66,4

Tabela 18: Comparação do resultado obtido com outros trabalhos da literatura no conjunto de teste da base CB513

Fusão Final - CB6133

Método	Acurácia Q8 (%)
Fusão Horizontal	**85,08**
Fusão Hierárquica	84,86
Fusão dos Métodos Baseados em Modelo	78,73
Fusão dos Métodos Livres de Modelo	76,67
RNN	76,09
Alinhamentos Gerais	75,96
BIv4	75,42
RIR	75,29
RF	69,30
Bons Alinhamentos	69,15
BERT	59,60

Tabela 19: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base CB6133

Fusão Final - CB6133

Método	Acurácia Q8 (%)
Nosso método	**85,1**
Ratul *et al.* [8]	76,9
Drori *et al.* [12]	76,3
Gou *et al.* [10]	75,7
Johansen *et al.* [13]	74,8
Guo *et al.* [11]	74,2
Zhou *et al.* [14]	74,0
Zhou e Troyanskaya [15]	72,1

Tabela 20: Comparação do resultado obtido com outros trabalhos da literatura no conjunto de teste da base CB6133

Fusão Final - CB513

Método	Acurácia Q8 (%)
Fusão Horizontal	**89,55**
Fusão dos Métodos Baseados em Modelo	89,16
Fusão Hierárquica	89,14
Alinhamentos Gerais	86,43
Bons Alinhamentos	76,34
Fusão dos Métodos Livres de Modelo	71,48
RIR	71,16
BIv4	70,93
RNN	69,88
RF	64,60
BERT	55,86

Tabela 21: Resultados dos métodos individuais e das fusões dos métodos livres de modelo e baseados em modelo no conjunto de teste da base CB513

Fusão Final - CB513

Método	Acurácia Q8 (%)
Nosso método	**89,5**
Ratul *et al.* [8]	71,9
Busia *et al.* [16]	71,4
Uddin *et al.* [17]	70,9
Johansen *et al.* [13]	70,9
Drori *et al.* [12]	70,7
Fang *et al.* [9]	70,6
Zhou *et al.* [14]	70,3
Gou *et al.* [10]	70,2
Li e Yu [18]	69,4
Lin *et al.* [19]	68,4
Wang *et al.* [20]	68,2
Zhou e Troyanskaya [15]	66,4

Tabela 22: Comparação do resultado obtido com outros trabalhos da literatura no conjunto de teste da base CB513

Sumário

Conclusões

- Métodos livres de modelo
  - Métodos baseados em redes neurais
    - Rede Neural Recorrente
    - Rede Neural Convolucional
    - *Transformers*
  - Métodos de aprendizado de máquina
    - Florestas Aleatórias
- Métodos baseados em modelo
  - BLAST
- Métricas de avaliação

Questões de Pesquisa

1. Utilizar apenas sequência de aminoácidos pode produzir resultados próximos aos resultados utilizando sequência de aminoácidos e informações evolutivas?
2. A transformação do vetor *one-hot encoding* esparso em um vetor denso pode ajudar na classificação de estruturas secundárias das proteínas?
3. Qual é o impacto da fusão de classificadores locais e globais na classificação?
4. A fusão entre métodos baseados em modelo e livres de modelo pode melhorar os resultados?

Trabalhos Futuros

- Pré-processamento
  - Aumentação de dados
  - Outras características evolutivas
  - *Transformers*
- Pós-processamento
  - Explicabilidade dos resultados
  - Técnicas de fusão

Publicações

- "Ensemble of Bidirectional Recurrent Networks and Random Forests for Protein Secondary Structure Prediction" (Gabriel Bianchin de Oliveira, Hélio Pedrini e Zanoni Dias) - 27th International Conference on Systems, Signals and Image Processing (IWSSIP 2020) [21]
- "Fusion of BLAST and Ensemble of Classifiers for Protein Secondary Structure Prediction" (Gabriel Bianchin de Oliveira, Hélio Pedrini e Zanoni Dias) - 33rd Conference on Graphics, Patterns and Images (SIBGRAPI 2020) [22]
- "Protein Secondary Structure Prediction Based on Fusion of Machine Learning Classifiers" (Gabriel Bianchin de Oliveira, Hélio Pedrini e Zanoni Dias) - 36th ACM/SIGAPP Symposium On Applied Computing - Bioinformatics Track (ACM SAC BIO 2021) [23]

Referências

[1] Stephen F Altschul, Thomas L Madden, Alejandro A Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J Lipman.
Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
Nucleic Acids Research, 25(17):3389–3402, 1997.

[2] Peter Y Chou and Gerald D Fasman.
Prediction of protein conformation.
Biochemistry, 13(2):222–245, 1974.

[3] Jonathan M Levin, Barry Robson, and Jean Garnier.
An algorithm for secondary structure determination in proteins based on sequence similarity.
FEBS Letters, 205(2):303–308, 1986.

Referências

[4] Ken Nishikawa and Tatsuo Ooi.
Amino acid sequence homology applied to the prediction of protein secondary structures, and joint prediction with existing methods.
Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 871(1):45–54, 1986.

[5] Howard Holley and Martin Karplus.
Protein secondary structure prediction with a neural network.
National Academy of Sciences, 86(1):152–156, 1989.

[6] David T Jones.
Protein secondary structure prediction based on position-specific scoring matrices.
Journal of Molecular Biology, 292(2):195–202, 1999.

Referências

[7] Dariusz Przybylski and Burkhard Rost.
Alignments grow, secondary structure prediction improves.
Proteins: Structure, Function, and Bioinformatics, 46(2):197–205, 2002.

[8] Md Aminur Rab Ratul, Maryam Tavakol Elahi, M Hamed Mozaffari, and WonSook Lee.
PS8-Net: A deep convolutional neural network to predict the eight-state protein secondary structure.
arXiv preprint arXiv:2009.10380, 2020.

[9] Chao Fang, Yi Shang, and Dong Xu.
MUFOLD-SS: New deep inception-inside-inception networks for protein secondary structure prediction.
Proteins: Structure, Function, and Bioinformatics, 86(5):592–598, 2018.

Referências

[10] Yanbu Guo, Bingyi Wang, Weihua Li, and Bei Yang.
Protein secondary structure prediction improved by recurrent neural networks integrated with two-dimensional convolutional neural networks.
Journal of Bioinformatics and Computational Biology, 16(05):1850021, 2018.

[11] Yanbu Guo, Weihua Li, Bingyi Wang, Huiqing Liu, and Dongming Zhou.
DeepACLSTM: deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction.
BMC Bioinformatics, 20(1):341, 2019.

Referências

[12] Iddo Drori, Isht Dwivedi, Pranav Shrestha, Jeffrey Wan, Yueqi Wang, Yunchu He, Anthony Mazza, Hugh Krogh-Freeman, Dimitri Leggas, Kendal Sandridge, Linyong Nan, Kaveri Thakoor, Chinmay Joshi, Sonam Goenka, Chen Keasar, and Itsik Peer.
High quality prediction of protein Q8 secondary structure by diverse neural network architectures.
arXiv preprint arXiv:1811.07143, 2018.

[13] Alexander R Johansen, Casper K Sønderby, Søren K Sønderby, and Ole Winther.
Deep recurrent conditional random field network for protein secondary prediction.
In *8th International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, pages 73–78. ACM, 2017.

Referências

[14] Jiyun Zhou, Hongpeng Wang, Zhishan Zhao, Ruifeng Xu, and Qin Lu.
CNNH_PSS: protein 8-class secondary structure prediction by convolutional neural network with highway.
BMC Bioinformatics, 19(4):60, 2018.

[15] Jian Zhou and Olga Troyanskaya.
Deep Supervised and Convolutional Generative Stochastic Network for Protein Secondary Structure Prediction.
In *31st International Conference on Machine Learning (ICML)*, pages 745–753, 2014.

[16] Akosua Busia and Navdeep Jaitly.
Next-step conditioned deep convolutional neural networks improve protein secondary structure prediction.
arXiv preprint arXiv:1702.03865, 2017.

Referências

[17] Mostofa Rafid Uddin, Sazan Mahbub, Md Saifur Rahman, and Md Shamsuzzoha Bayzid.
SAINT: self-attention augmented inception-inside-inception network improves protein secondary structure prediction.
bioRxiv, page 786921, 2019.

[18] Zhen Li and Yizhou Yu.
Protein secondary structure prediction using cascaded convolutional and recurrent neural networks.
In *25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2560–2567. AAAI Press, 2016.

Referências

[19] Zeming Lin, Jack Lanchantin, and Yanjun Qi.
MUST-CNN: a multilayer shift-and-stitch deep convolutional architecture for sequence-based protein structure prediction.
In *30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 27–34, 2016.

[20] Yangxu Wang, Hua Mao, and Zhang Yi.
Protein secondary structure prediction by using deep learning method.
Knowledge-Based Systems, 118:115–123, 2017.

Referências

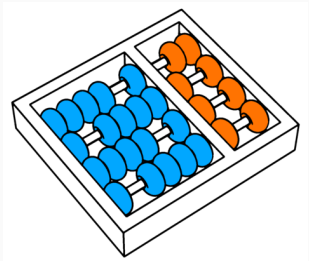
[21] Gabriel Bianchin de Oliveira, Helio Pedrini, and Zanoni Dias. **Ensemble of bidirectional recurrent networks and random forests for protein secondary structure prediction.** In *27th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 311–316, Rio de Janeiro, RJ, Brazil, 2020. IEEE.

[22] Gabriel Bianchin de Oliveira, Helio Pedrini, and Zanoni Dias. **Fusion of BLAST and Ensemble of Classifiers for Protein Secondary Structure Prediction.** In *33rd Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 308–315. IEEE, 2020.

Referências

[23] Gabriel Bianchin de Oliveira, Helio Pedrini, and Zanoni Dias. **Protein Secondary Structure Prediction Based on Fusion of Machine Learning Classifiers.** In *36th ACM/SIGAPP Symposium On Applied Computing - Bioinformatics Track (ACM SAC BIO)*, pages 26–29, Gwangju, South Korea, 2021. ACM.

Agradecimentos



LIV

Laboratório de Informática Visual

Predição de Estruturas Secundárias de Proteínas usando Aprendizado de Máquina e BLAST

Gabriel Bianchin de Oliveira

Orientador: Prof. Dr. Zanoni Dias
Coorientador: Prof. Dr. Hélio Pedrini

11 de Março de 2021

Instituto de Computação
Universidade Estadual de Campinas