

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta
Dissertação será disponibilizado
somente a partir de 16/07/23.



Universidade Estadual Paulista "Júlio de Mesquita Filho"
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Uso de Redes Neurais Artificiais para Extração de Dados de Prontuários Médicos

Naila Camila da Rocha

Botucatu
2021

Naila Camila da Rocha

Uso de Redes Neurais Artificiais para Extração de Dados de Prontuários Médicos

Dissertação de Mestrado apresentada ao Curso de Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Mestre em Biometria.

Orientador: Prof(a). Dr(a). Liciania Vaz de Arruda Silveira

Coorientador: Prof. Dr. José Eduardo Corrente

Botucatu
2021

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Rocha, Naila Camila da.

Uso de redes neurais artificiais para extração de dados de prontuários médicos / Naila Camila da Rocha. - Botucatu, 2021

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Liciania Vaz de Arruda Silveira

Coorientador: José Eduardo Corrente

Capes: 90194000

1. Registros médicos.
2. Análise por agrupamento.
3. Redes neurais (Computação.
4. Distância de Gower.

Palavras-chave: Análise de agrupamentos; Distância de Gower; Prontuários médicos; Reconhecimento de entidades nomeadas; Redes neurais.

Resumo

Diversos estudos recentes têm utilizado inteligência artificial na extração e tratamento de dados secundários na área da saúde, obtidos em prontuários eletrônicos hospitalares. No entanto, alguns estudos são inviáveis devido a informações incompletas ou inseridas apenas em campos narrativos. O objetivo deste trabalho é desenvolver uma rede neural que utilize os dados desses campos para obter informações estruturadas referentes aos sintomas, diagnósticos, medicamentos, condições, exames e tratamentos. A rede neural proposta facilitará a descoberta de relações entre doenças e sintomas, prevalências e incidências, a identificação de condições clínicas, a evolução de enfermidades e os efeitos das medicações prescritas. O algoritmo utiliza métodos de processamento de linguagem natural para extração de textos e redes neurais convolucionais para reconhecimento de padrões. Foram simulados diferentes valores e funções para a determinação dos hiperparâmetros e otimizadores mais adequados para o modelo de Reconhecimento de Entidades Nomeadas (NER) desenvolvido através da biblioteca *spaCy* em *Python*. Para uma análise exploratória dos dados extraídos e demonstração da aplicabilidade do modelo foram executadas técnicas da estatística multivariada de análise de agrupamento, obtendo quatro grupos que melhor representam os perfis dos pacientes e os medicamentos por eles utilizados. Os resultados obtidos foram significativos considerando a complexidade do modelo, com um *F-Score* de 63,9% e *Precision* de 72,7%. A classe Condição do Paciente chegou a atingir 90,3% de *Precision*, seguido por Medicação com 87,5%. No desenvolvimento do presente trabalho, foram utilizados dados de 30.000 prontuários de pacientes do Hospital das Clínicas da Faculdade de Medicina de Botucatu/SP - Brasil (HCFMB), gerando um *corpus* com 1.200 textos clínicos. A utilização de NER em dados clínicos se mostrou uma ferramenta capaz de extrair informações que não existem em campos estruturados de prontuários médicos. Além disso, análises de agrupamento utilizando esses dados revelam comportamentos e características até então desconhecidas, relacionadas com as Entidades extraídas.

Palavras-chave: Redes Neurais Convolucionais, Análise de Agrupamento, Reconhecimento de Entidades Nomeadas, Prontuários Médicos.

Sumário

	1 INTRODUÇÃO	1
1.1	Objetivos	2
	2 FUNDAMENTAÇÃO TEÓRICA	4
2.0.1	Mineração de Texto e Linguagem Natural	4
2.0.2	Redes Neurais e Aprendizagem Profunda	5
2.0.3	Reconhecimento de Entidades Nomeadas (REN)	7
2.0.3.1	Modelos Ocultos de Markov	9
2.0.3.2	Máxima Entropia	9
2.0.3.3	Campos Aleatórios Condicionais	10
2.0.3.4	Redes Neurais Convolucionais	11
2.0.3.4.1	Redes Neurais Convolucionais em Processamento de Linguagem Natural (PLN)	11
2.0.3.4.2	Camadas do tipo <i>Embedding</i>	12
2.0.3.4.3	Arquitetura e Funcionamento de uma Rede Neural Convolucional para a Classificação de Texto	12
2.0.3.4.4	Camada do tipo <i>Pooling</i>	13
2.0.3.4.5	Funções de Ativação	13
2.0.3.4.6	Funções de Custo, Funções de Perdas ou Funções de Erro	17
2.0.3.4.7	Taxa de Aprendizado - <i>Learning Rate</i>	19
2.0.3.4.8	Otimizadores de Parâmetros	19
2.0.3.4.9	<i>Dropout</i> de unidades (ocultas e visíveis) em uma Rede Neural	22
2.0.3.4.10	Camada Final <i>Fully Connected</i>	22
2.0.3.4.11	Redes Neurais Convolucionais Residuais (ResNet)	22
2.0.4	Análise da Qualidade dos Dados e Validação do Modelo	23
2.0.5	Métodos Estatísticos Multivariados - Análise de Agrupamentos	26
	3 METODOLOGIA	29
3.0.1	Apresentação dos dados	29
3.0.2	Determinação dos Métodos	29
3.0.3	Tamanho do conjunto de dados	30
3.0.4	Ferramentas para o Reconhecimento de Entidades Nomeadas	30
3.0.5	Pré-processamento dos dados	30
3.0.6	<i>Corpus</i> Clínico	31
3.0.7	Descrição do Modelo e suas Configurações	32
3.0.8	Análise de Desempenho do Modelo para o Reconhecimento de Entidades Nomeadas	34

3.0.9	Pós-processamento dos dados	35
3.0.10	Métodos Estatísticos Multivariados - Análise de Agrupamentos	37
	4 RESULTADOS E AVALIAÇÃO DO DESEMPENHO DO MODELO	39
4.0.1	Resultados e Desempenho do Modelo	39
4.0.2	Hiperparâmetros	41
4.0.3	Entidades Extraídas	42
4.0.4	Análise da Qualidade dos Dados e Validação do Modelo	44
4.0.5	Métodos Estatísticos Multivariados - Análise de Agrupamentos	45
	5 CONCLUSÃO	49
	Referências	51

Referências

- AIHUB, T. M. *Named Entity Recognition using Spacy and Tensorflow*. 2020. Disponível em: <<https://aihub.cloud.google.com>>. 7, 8
- ANANIADOU, S.; KELL, D. B.; TSUJII, J.-i. Text mining and its potential applications in systems biology. *Trends in biotechnology*, Elsevier, v. 24, n. 12, p. 571–579, 2006. 1
- ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, v. 5, n. 2, 2006. 2
- BAKER, R. S. J. D. Data mining for education. *International encyclopedia of education*, Elsevier Oxford, UK, v. 7, n. 3, p. 112–118, 2010. 26
- BALDI, P. Autoencoders, unsupervised learning, and deep architectures. In: *Proceedings of ICML workshop on unsupervised and transfer learning*. [S.l.: s.n.], 2012. p. 37–49. 6
- BERKHIN, P.; BECHER, J. D. Learning simple relations: Theory and applications. In: *SIAM. Proceedings of the 2002 SIAM International Conference on Data Mining*. [S.l.], 2002. p. 420–436. 26
- BISHOP, M.; LOONEY, C. 006.4—computer systems. pattern recognition. *The British National Bibliography*, British Library, Bibliographic Services Division., v. 1, p. 30, 1998. 6
- BONNIN, R. *Building Machine Learning Projects with TensorFlow*. [S.l.]: Packt Publishing Ltd, 2016. 12, 17
- BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 111–118. 13, 22
- BRAGA, L. P. V. B. *Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada*. [S.l.]: Editora E-papers, 2005. 5
- BURKE, D. S. et al. Measurement of the false positive rate in a screening program for human immunodeficiency virus infections. *New England Journal of Medicine*, Mass Medical Soc, v. 319, n. 15, p. 961–964, 1988. 23, 24
- CASTRO, L. N. d.; FERRARI, D. G. *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. São Paulo: Saraiva, 2016. 5
- CHEN, D.; MANNING, C. D. A fast and accurate dependency parser using neural networks. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. [S.l.: s.n.], 2014. p. 740–750. 13, 14
- CHEN, H. et al. A textual database/knowledge-base coupling approach to creating computer-supported organizational memory. *MIS Department, University of Arizona*, v. 5, 1994. 4
- CHIEU, H. L.; NG, H. T. Named entity recognition with a maximum entropy approach. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. [s.n.], 2003. p. 160–163. Disponível em: <<https://www.aclweb.org/anthology/W03-0423>>. 9, 10

- COSTA, H. G. Modelo para webibliomining: proposta e caso de aplicação. *Revista da FAE*, v. 13, n. 1, p. 115–126, 2010. 5
- DAWAR, K.; SAMUEL, A. J.; ALVARADO, R. Comparing topic modeling and named entity recognition techniques for the semantic indexing of a landscape architecture textbook. In: IEEE. *2019 Systems and Information Engineering Design Symposium (SIEDS)*. [S.l.], 2019. p. 1–6. 9
- DENG, L.; YU, D. Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014. 6, 7
- EFRON, B. *The jackknife, the bootstrap and other resampling plans*. [S.l.]: SIAM, 1982. 25, 26
- FALCÃO, A. E. J. et al. Indecs: método automatizado de classificação de páginas web de saúde usando mineração de texto e descritores em ciências da saúde (decs). *Journal of Health Informatics*, v. 1, n. 1, 2009. 1
- GOODFELLOW, I. et al. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1. 5, 7, 14, 15, 16, 17
- GOTH, G. Analyzing medical data. *Communications of the ACM*, ACM New York, NY, USA, v. 55, n. 6, p. 13–15, 2012. 1
- GUPTA, D. Fundamentals of deep learning—introduction to recurrent neural networks. *Analytics Vidhya (2107)*, 2017. 14, 15, 17
- HAYKIN, S. S. *Neural Networks and learning machines/Simon Haykin*. [S.l.]: New York: Prentice Hall,, 2009. 5, 6
- HCFMB. *HOSPITAL DAS CLINICAS DA FACULDADE DE MEDICINA DE BOTUCATU (HCFMB)*. 2019. Disponível em: <<http://www.hcfmb.unesp.br/>>. 29
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. 22, 23
- HEARST, M. A. Untangling text data mining. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. [S.l.], 1999. p. 3–10. 4
- HINTON, G.; OSINDERO, S. The, y. 2006, a fast learning algorithm for deep belief nets. *Neural computation*, v. 18, n. 7, 2006. 6, 33
- JONES, M. T. *Arquiteturas de aprendizado profundo: O surgimento da inteligência artificial*. 2017. Disponível em: <<http://www.ibm.com/developerworks/br/library/cc-machine-learning-deeplearning-architectures/cc-machine-learning-deep-learning-architectures-pdf.pdf/>>. 7
- JURAFSKY, D.; MARTIN, J. H. Speech and language processing (draft). *Chapter A: Hidden Markov Models (Draft of September 11, 2018)*. Retrieved March, v. 19, p. 2019, 2018. 9, 10
- JUSOH, S.; ALFAWAREH, H. M. Techniques, applications and challenging issue in text mining. *International Journal of Computer Science Issues (IJCSI)*, International Journal of Computer Science Issues (IJCSI), v. 9, n. 6, p. 431, 2012. 4
- KASSAMBARA, A. *Practical guide to cluster analysis in R: Unsupervised machine learning*. [S.l.]: Sthda, 2017. v. 1. 26, 27, 28

- KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. [S.l.]: John Wiley & Sons, 2009. v. 344. 26, 27, 28
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 21
- KOHANE, I. S. Using electronic health records to drive discovery in disease genomics. *Nature Reviews Genetics*, Nature Publishing Group, v. 12, n. 6, p. 417–428, 2011. 1
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. 13, 14
- KULKARNI, T. D. et al. Deep convolutional inverse graphics network. *Advances in neural information processing systems*, v. 28, p. 2539–2547, 2015. 11
- LAFFERTY, J.; MCCALLUM, A.; PEREIRA, F. C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001. 10
- LEOPOLD, H. et al. Using hidden markov models for the accurate linguistic analysis of process model activity labels. *Information Systems*, Elsevier, v. 83, p. 30–39, 2019. 9
- LI, J. et al. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 2020. 8, 23
- LOH, S.; GARIN, R. S. Web intelligence–inteligência artificial para descoberta de conhecimento na web. *Oficina de Inteligência Artificial*, v. 5, p. 11–34, 2001. 5
- LOPES, F.; TEIXEIRA, C.; OLIVEIRA, H. G. Contributions to clinical named entity recognition in Portuguese. In: *Proceedings of the 18th BioNLP Workshop and Shared Task*. Florence, Italy: Association for Computational Linguistics, 2019. p. 223–233. Disponível em: <<https://www.aclweb.org/anthology/W19-5024>>. 8
- MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In: *Proc. icml*. [S.l.: s.n.], 2013. v. 30, n. 1, p. 3. 13, 14
- MAIA, L. B. et al. Evaluation of melanoma diagnosis using deep features. In: IEEE. *2018 25th international conference on systems, signals and image processing (IWSSIP)*. [S.l.], 2018. p. 1–4. 11
- MALI, M.; ATIQUE, M. Applications of text classification using text mining. *International Journal of Engineering Trends and Technology*, v. 13, n. 5, p. 209, 2014. 5
- MCCALLUM, A.; LI, W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. 2003. 11
- MONTAVON, G.; SAMEK, W.; MÜLLER, K.-R. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, Elsevier, v. 73, p. 1–15, 2018. 2
- MURUGAVEL, M. *Spacy Annotation Tool*. 2020. Disponível em: <<https://manivannanmurugavel.github.io/annotating-tool/spacy-ner-annotator/>>. 31, 32
- NETO, J. F. d. S. Reconhecimento de entidades nomeadas para o português usando redes neurais. Pontifícia Universidade Católica do Rio Grande do Sul, 2019. 25

- NIELSEN, M. A. *Neural Networks and deep learning*. [S.l.]: Determination press San Francisco, CA, USA:, 2015. v. 2018. 6
- OLSON, D. L.; DELEN, D. Performance evaluation for predictive modeling. In: *Advanced data mining techniques*. [S.l.]: Springer, 2008. p. 137–147. 23, 24, 25, 26
- PAKHOMOV, S. et al. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care*, v. 13, n. 6 Part 1, p. 281–288, 2007. 2
- PEI, W.; GE, T.; CHANG, B. An effective neural network model for graph-based dependency parsing. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. [S.l.: s.n.], 2015. p. 313–322. 13, 14
- PEISSIG, P. L. et al. Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *Journal of the American Medical Informatics Association*, BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, v. 19, n. 2, p. 225–234, 2012. 2
- PERES, R. Algoritmo back-propagation. *Revista Programar*, 2017. 5
- PETERS, A. C. et al. Semclinbr—a multi institutional and multi specialty semantically annotated corpus for portuguese clinical nlp tasks. *arXiv preprint arXiv:2001.10071*, 2020. 8
- PINTO, V. B. Prontuário eletrônico do paciente: documento técnico de informação e comunicação do domínio da saúde 10.5007/1518-2924.2006 v11n21p34. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 11, n. 21, p. 34–48, 2006. 1
- PIRES, A.; DEVEZAS, J.; NUNES, S. Benchmarking named entity recognition tools for portuguese. *Proceedings of the Ninth INForum: Simpósio de Informática*, p. 111–121, 2017. 24
- PODANI, J.; SCHMERA, D. On dendrogram-based measures of functional diversity. *Oikos*, Wiley Online Library, v. 115, n. 1, p. 179–185, 2006. 26
- PONTI, M. A.; COSTA, G. B. P. da. Como funciona o deep learning. *arXiv preprint arXiv:1806.07908*, 2018. 5
- PYPI. *Python Package Index (PyPI)*. 2021. Disponível em: <<https://pypi.org/>>. 31
- RABINER, L.; JUANG, B. An introduction to hidden markov models. *ieee assp magazine*, IEEE, v. 3, n. 1, p. 4–16, 1986. 9
- RAMACHANDRAN, P.; ZOPH, B.; LE, Q. V. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017. 14, 15, 16, 17
- RASCHKA, S.; MIRJALILI, V. *Python machine learning*. [S.l.]: Packt Publishing Ltd, 2017. 17
- RATNAPARKHI, A. Maximum entropy models for natural language ambiguity resolution. 1998. 9, 10
- REYNOLDS, A. P.; RICHARDS, G.; RAYWARD-SMITH, V. J. The application of k-medoids and pam to the clustering of rules. In: SPRINGER. *International Conference on Intelligent Data Engineering and Automated Learning*. [S.l.], 2004. p. 173–178. 27, 28
- ROQUE, F. S. et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS computational biology*, Public Library of Science, v. 7, n. 8, 2011. 2

- RUDER, S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016. 17, 19, 20, 21, 22
- SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, Elsevier, v. 61, p. 85–117, 2015. 6
- SCHNEIDER, E. T. R. et al. BioBERTpt - a Portuguese neural language model for clinical named entity recognition. In: *Proceedings of the 3rd Clinical Natural Language Processing Workshop*. Online: Association for Computational Linguistics, 2020. p. 65–72. Disponível em: <<https://www.aclweb.org/anthology/2020.clinicalnlp-1.7>>. 8
- SETTLES, B. Biomedical named entity recognition using conditional random fields and rich feature sets. In: *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP)*. [S.l.: s.n.], 2004. p. 107–110. 10, 11
- SHELAR, H. et al. Named entity recognition approaches and their comparison for custom ner model. *Science & Technology Libraries*, Taylor & Francis, v. 39, n. 3, p. 324–337, 2020. 8
- SILVA, L. A. da; PERES, S. M.; BOSCARIOLI, C. *Introdução à mineração de dados: com aplicações em R*. [S.l.]: Elsevier Brasil, 2017. 5
- SLATTON, T. G. A comparison of dropout and weight decay for regularizing deep neural networks. 2014. 33
- SMITH, L. N. Cyclical learning rates for training neural networks. In: IEEE. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.], 2017. p. 464–472. 19
- SONG, M. Opinion: Text mining in the clinic. *The Scientist*, 2013. 1
- SPACY. *Language Processing Pipelines*. 2021. Disponível em: <<https://spacy.io/usage/processing-pipelines>>. 33
- SPASIC, I. et al. Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in bioinformatics*, Henry Stewart Publications, v. 6, n. 3, p. 239–251, 2005. 1
- SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014. 22, 33
- SULLIVAN, D. The need for text mining in business intelligence. *DM REVIEW*, POWELL PUBLISHING INC, v. 10, p. 12–16, 2000. 4
- TAN, A.-H. Text mining: The state of the art and the challenges. In: SN. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*. [S.l.], 1999. v. 8, p. 65–70. 4
- THINC. *Loss Calculators*. 2021. Disponível em: <<https://thinc.ai/docs/api-loss>>. 33
- THOMAS, J.; MCNAUGHT, J.; ANANIADOU, S. Applications of text mining within systematic reviews. *Research Synthesis Methods*, Wiley Online Library, v. 2, n. 1, p. 1–14, 2011. 5
- THURASINGHAM, B. *Data mining: technologies, techniques, tools, and trends*. [S.l.]: CRC press, 2014. 4

- VYCHEGZHANIN, S.; KOTELNIKOV, E. Comparison of named entity recognition tools applied to news articles. In: IEEE. *2019 Ivannikov Ispras Open Conference (ISPRAS)*. [S.l.], 2019. p. 72–77. 8
- WANG, Y.-Y.; DENG, L.; ACERO, A. Spoken language understanding. *IEEE Signal Processing Magazine*, IEEE, v. 22, n. 5, p. 16–31, 2005. 9
- XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on neural networks*, Ieee, v. 16, n. 3, p. 645–678, 2005. 26, 27
- ZEILER, M. D. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012. 19
- ZHANG, Y.; WALLACE, B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015. 11, 12, 13, 14, 22
- ZHOU, S. K.; GREENSPAN, H.; SHEN, D. *Deep learning for medical image analysis*. [S.l.]: Academic Press, 2017. 6, 7
- ZWEIGENBAUM, P. et al. Frontiers of biomedical text mining: current progress. *Briefings in bioinformatics*, Oxford University Press, v. 8, n. 5, p. 358–375, 2007. 1