ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 6 Issue 6 November 2018

International Manuscript ID : 23482001V6I6112018-5

(Approved and Registered with Govt. of India)

Programming Patterns with Bioinformatics and Molecular Biology

Shahnawaz Ayoub
Research Scholar
Shri Venkateshwara University
Gajraula, U.P., India

Dr. Rakesh Kumar
Associate Professor
Shri Venkateshwara University
Gajraula, U.P., India

Abstract

Bioinformatics is the interdisciplinary area which integrates biology, computer science, mathematics, engineering, chemistry and statistics for advanced predictions and analytics. The field of molecular biology is also closely associated with Bioinformatics for accurate analysis of biological structures. Molecular biology deals with the deep analysis of the bimolecular movements in the cell of body along with the details of Proteins, DNA, RNA and biosynthesis. In addition to the enormous diagnosis machines in medical sciences, the software tools and libraries are also used. These software tools and applications evaluate the biological data which are fetched from the computerized diagnosis machines. Here, the concept of Bioinformatics comes to the scenario in which the software tools and applications are used to understand the biological and medical data. These software suites make use of high performance programming languages at the back-end to process and evaluate the biological dataset with the objectives to find out the human body parameters for effective treatment.

Keywords: Bioinformatics, Data Analytics, Data Science

Introduction

Now days, the applications of Information and Communications Technology (ICT) are not limited to data transmission, cloud deployments, social media, web servers and mobile applications. From last decade, Information Technology is touching every area of social and corporate world including health and medical sciences. Most of the medical diagnosis laboratories are now equipped with the advance computerized machines to accurately diagnose and fetch the parameters of human body including Magnetic Resonance Imaging (MRI), Computed tomography (CT), Electroencephalography (EEG), Electrocardiography (ECG), Ultrasound, Mammography, Laparoscopy, Blood Examination, X-Ray and many others. These systems provide higher degree of accuracy in the analysis of human body which assists the medical practitioner or doctor to predict the disease. By this process, the medical practitioners are able to recommend the suitable treatment to the patients.

Datasets for Research in Medical and Biological Areas

With the deployments of computerized machines, the researchers in diagnosis and medical sciences are taking assistance from software professionals in their field so that the programming modules can be processed by the software developers. Even, the computer scientists are now taking the interdisciplinary field of bioinformatics for their research so that their programming knowledge can be utilized for health sciences.

There are enormous medical datasets available for researches which are released by the diagnosis laboratories so that the overall architecture and structure of medical-biological data can be analyzed by the software experts. The programmers working in bioinformatics can download these medical datasets and they can perform the analysis using their effective algorithms.

Following are few links from where the medical data on EEG, ECG, MRI and X-Ray can be fetched for analysis using programming languages and tools

Dataset Library	**Link**
UCI Machine Learning Repository	https://archive.ics.uci.edu/ml/datasets.html
Health Data	https://www.healthdata.gov
Physionet	https://www.physionet.org/pn6/chbmit/
BrainSignals	http://www.brainsignals.de/
EEG Dataset	http://www.bsp.brain.riken.jp/~qibin/homepage/Datasets.html
OpenfMRI	https://openfmri.org/dataset/
Alyward	http://www.aylward.org/notes/open-access-medical-image-repositories
ECG Dataset	https://www.physionet.org/physiobank/database/ptbdb/
ECG Library	https://ecglibrary.com/ecghome.php
Ultrasound	http://splab.cz/en/download/databaze/ultrasound

Besides the abovementioned links, there are many resources available from where the medical and microbiological data can be downloaded for research and predictions.

Free and Open Source Tools for analysis of Medical Data

Following are the software tools which can be used for the analysis and evaluation of medical data for specific type of dataset

OpenEEG (URL : http://openeeg.sourceforge.net/doc/)

OpenEEG is free and open source software which can be used for EEG Signal analysis with enormous libraries as add-on including Neuroserver, BioEra, BrainBay, Brainathlon, BrainWave Viewer and EEGMIR.

EEGNET (URL : https://sites.google.com/site/eegnetworks/)

It is Free and Open Source Tool for the analysis and visualization of EEG Brain Signals. It is having features to visualize the brain network.

BioSig (URL : http://biosig.sourceforge.net/)

BioSig is a software library under free and open source distribution with the enormous features of biomedical signal processing. This library is having excellent features to process the biosignals including electrocorticogram (ECoG), electromyogram (EMG), electrocardiogram (ECG), electrooculogram (EOG), electroencephalogram (EEG), respiration and many others. In addition, the interfacing toolboxes and drivers for Octave, MATLAB, Python, PHP, Perl, Ruby, Tcl, C and C++ are also available. The key areas of brain-computer interfaces, psychology, Neuroinformatics, Cardiovascular Systems, Neurophysiology and sleep research are effectively processed in BioSig.

GenomeTools (URL : http://genometools.org/)

GenomeTools is the open source software for the analysis of genome and biological parameters. It is having a free library of tools for bioinformatics. The APIs in C are available with the detailed manual of usage. In addition, the deep analysis of biological structures are integrated in GenomeTools.

Working with BioPython for Molecular Biology (URL : http://biopython.org/)

Biopython provides the set of tools and libraries for the analysis and computations of biological structures. Biopython is available in free and open source distribution and member of Open Bioinformatics Foundation (OBF). Biopython can parse the files of bioinformatics into the data structures which can be processed by Python code.

Following international formats are supported in Biopython

- UniGene
- PubMed

- GenBank
- Medline
- GenBank
- FASTA
- Clustalw
- Blast

Installation of Biopython on Ubuntu

```
$ sudo apt-get install python-biopython
```

Installation of Biopython with Documentation

```
$ sudo apt-get install python-biopython-doc
```

BioSQL (http://biosql.org) can be used with Biopython to store the biological database. To integrate BioSQL, following instruction is executed

```
$ sudo apt-get install python-biopython-sql
```

Sequence is the key object in bioinformatics. The sequences can be processed in Biopython with following instructions

```
>>> from Bio.Seq import Seq
>>> my_seq = Seq("MyDefinedSequence")
>>> my_seq
Seq(' MyDefinedSequence ', Alphabet())
>>> print(my_seq)
MyDefinedSequence
>>> my_seq.alphabet
Alphabet()
```

Transcription Functions on DNA and RNA

If you have a DNA sequence, you may want to turn it into RNA. In bioinformatics we normally assume the DNA is the coding strand (not the template strand) so this is a simple matter of replacing all the thymines with uracil:

Complement and reverse complement

These are very simple - the methods return a new Seq object with the appropriate sequence and the same alphabet:

```
>>> from Bio.Seq import Seq
>>> from Bio.Alphabet import generic_dna

>>> my_values_dna = Seq("MY_VALUES_DNA", generic_dna)
>>> my_values_dna
Seq('MY_VALUES_DNA', DNAAlphabet())

>>> my_values_dna.complement()
Seq('ATCATGTGACCA', DNAAlphabet())

>>> my_values_dna.reverse_complement()
Seq('ACCAGTGTACTA', DNAAlphabet())
```

If you have a DNA sequence, you may want to turn it into RNA. In bioinformatics we normally assume the DNA is the coding strand (not the template strand) so this is a simple matter of replacing all the thymines with uracil:

```
>>> my_values_dna
Seq('MY_VALUES_DNA', DNAAlphabet())
```

```
>>> my_values_dna.transcribe()
Seq('AGUACACUGGU', RNAAlphabet())
```

With the specification of RNA, the associated DNA can be fetched

```
>>> my_values_rna = my_values_dna.transcribe()

>>> my_values_rna
Seq('AGUACACUGGU', RNAAlphabet())

>>> my_values_rna.back_transcribe()
Seq('MY_VALUES_DNA', DNAAlphabet())

>>> my_values_rna
Seq('AGUACACUGGU', RNAAlphabet())

>>> my_values_rna.back_transcribe().reverse_complement()
Seq('ACCAGTGTACT', DNAAlphabet())
```

Sleep EEG Analysis in GNU Octave

Assorted signals are delivered to all parts of the body so that the other organs can communicate each other for specific or general purposes. One of the key signals in the human brain is Electroencephalography (EEG) which is generated from the brain including during the state of sleep and unconscious. Electroencephalography (EEG) signals comprise the brain waves which can be evaluated using GNU Octave. The analysis on sleeping disorders and various diseases can be done with EEG evaluation.

GNU Octave (https://www.gnu.org/software/octave/) is one the powerful and multifunctional tool for engineering and scientific applications of research. The simulations related to engineering as well as medical can be implemented with the assorted toolboxes and functions in Octave. It is used as an effective alternate to MATLAB under open source distribution. A number of toolboxes for different applications are available in GNU Octave which can be used for optimization and predictive analysis.

The Wave Form Database (WFDB) Package can be integrated with GNU Octave. This package is equipped with the functions and modules for EEG and Brain Signal evaluations. Similar process is followed in case of Brain Mapping or Brain Fingerprinting for criminal investigation in their unconscious state. There are assorted stages of sleep or unconscious states which can be analyzed from EEG signals after recording from the electrodes. This process assists in the forensic analysis of the person while in unconscious state. By this evaluation, the medical disorders can also be detected using WFDB package in Octave. Following are the excerpts of Benchmark Sleep Stages which can be evaluated using WFDB package in GNU Octave so that the overall nervous system can be predicted along with the brain disorders.

Conclusion

Now days, Bioinformatics and Biomedical Predictive Analytics is one of the key domains of research for assorted applications. The extraction, processing and predictive mining from brain, heart and other human body generated signals are evaluated with the use of information technology. The datasets from Physionet, UCSD, FPMS and others can be used for the research work in bioinformatics with the integration of data mining and machine learning tools.

References

[1] Nabian, Mohammad Amin; Meidani, Hadi (2017-08-28). "Deep Learning for Accelerated Reliability Analysis of Infrastructure Networks". Computer-Aided Civil and Infrastructure Engineering. 33 (6): 443–458. arXiv:1708.08551. Bibcode:2017arXiv170808551N. doi:10.1111/mice.12359. S2CID 36661983.

[2] Nabian, Mohammad Amin; Meidani, Hadi (2018). "Accelerating Stochastic Assessment of Post-Earthquake Transportation Network Connectivity via Machine-Learning-Based Surrogates". Transportation Research Board 97th Annual Meeting.

[3] Nabian, Mohammad Amin; Meidani, Hadi (2017). "Uncertainty Quantification and PCA-Based Model Reduction for Parallel Monte Carlo Analysis of Infrastructure System Reliability". Transportation Research Board 96th Annual Meeting.

[4] Climate Change 2013 The Physical Science Basis (PDF). Cambridge University Press. 2013. p. 697. ISBN 978-1-107-66182-0. Retrieved 2 March 2016.

[5] Cassey; Smith (2014). "Simulating confidence for the Ellison-Glaeser Index". Journal of Urban Economics. 81: 93. doi:10.1016/j.jue.2014.02.005.

[6] Sawilowsky & Fahoome 2003

[7] Spall, James C. (2005). "Monte Carlo Computation of the Fisher Information Matrix in Nonstandard Settings". Journal of Computational and Graphical Statistics. 14 (4): 889–909. CiteSeerX 10.1.1.142.738. doi:10.1198/106186005X78800. S2CID 16090098.

[8] Das, Sonjoy; Spall, James C.; Ghanem, Roger (2010). "Efficient Monte Carlo computation of Fisher information matrix using prior information". Computational Statistics & Data Analysis. 54 (2): 272–289. doi:10.1016/j.csda.2009.09.018.

[9] Guillaume Chaslot; Sander Bakkes; Istvan Szita; Pieter Spronck. "Monte-Carlo Tree Search: A New Framework for Game AI" (PDF). Sander.landofsand.com. Retrieved 28 October 2017.

[10] "Monte Carlo Tree Search - About". Archived from the original on 2015-11-29. Retrieved 2013-05-15.

[11] Chaslot, Guillaume M. J. -B; Winands, Mark H. M; Van Den Herik, H. Jaap (2008). Parallel Monte-Carlo Tree Search. Lecture Notes in Computer Science. 5131. pp. 60–71. CiteSeerX 10.1.1.159.4373. doi:10.1007/978-3-540-87608-3_6. ISBN 978-3-540-87607-6.