THE EFFICIENCY OF DATA VISUALIZATION SOFTWARE FOR PHYSIOBANK

Sándor Kaczur

1. ABSTRACT

This project started in 2010. The ECG curves can be characterized as certain waves (P, Q) length of the sign, the symmetry of the distance (PR, QT interval, RR, ST), in the QRS complex. The PhysioToolkit is an open-source, multi-language, programming available software collection used for biological, physiological researches. Those software products include data acquisition, signal processing, modeling, time series analysis software modules, components, packages. The author's research theme is the analysis of the efficiency of ECG algorithms. This article contains some related and complementary ranges of analytical examination on the efficiency of algorithms.

2. INTRODUCTION

In an ECG analysis the detection of the too rapidly extending (tachycardia), or hardly stimulated (bradycardia), or behave chaotically (fibrillation) is easy to find in which part of the QRS complex contains a problem. The fact that there is a problem with waves determines which part of the pacemaker and conduction system is malfunctioning. The various recordings (of 12) may be determined by comparing the exact location of the problem, which can be very important in case of infarction (heart attack).

The ECG is analyzed by grouping, its characteristics can be separated, which is usually called, interpreting. After separation and grouping you may draw conclusions in order to have a diagnosis that supports medical expert activities in the above data. In by doing so – after preprocessing, various learning and statistical methods – you may improve the automated diagnosis followed by the interpretation of the ECG wave characteristics.

The aim is to upgrade the algorithms that analyse efficiency with statistical analyses in order to facilitate the investigation of ECG waves.

3. SCHEDULE OF RESEARCH

The PhysioToolkit biological, physiological research used open-source, multi-lingual software collection [1] is recognized. The main topics of the available materials: presentation of data, import/export possibilities of signal processing and time series analysis, modeling, developing software. Studying the PhysioBank data visualization software from the wave display and PLT wave function plotting software, source code and algorithms. The available samples, data sets for processing, the necessary file format conversion, and the specifications of the above in rdann, rdsamp, wrann and wrsamp packages are learned. The software modules are also studied for processing physiological signals from the wqrs and qcgpuwave QRS detectors, the EDR respiratory signal analyzer, and the comp pr-wave analyzer.

The studied libRASCH system [2] provides a common interface between different modes of signals storage. The availability of source algorithms significantly supports the selected subfields of science. The detect-simple, Fiducial point and freq analysis plug-ins, including those with a QRS detection, the sliding benchmarks mapping and ECG analysis frequency are learned. HRV plug-in shows different wavelength intervals that are used to show such a plot view, ECG-view plug-ins.

The detailed documentation, source code and algorithms of the eplimited open source ECG analysis software [3] are analyzed.

The Weka 2 data analysis, expert system toolsedge [4] has been being studying, which used to function as an interface anthropometric database.

First, the selection will perform the necessary algorithms, analysis and efficiency analysis. For those, it is essential to know the available warning, corrugated drawing software in source code, and the known file formats of PhysioBank and libRASCH. The possibilities of more covering smaller subfields are also analyzed and how they connect to each other creating a large software where the algorithms are built on each other, that is, one output creating the other input (e.g., the detect-simple QRS wave detection method to complement the Fiducial-point algo-

rithm that gives the ECG sliding benchmark). For software development the necessary algorithms in Java was integrated. My results in an area overview that summarizes the publications and conference presentations will be presented. The parts that need improvement in algorithms will be examined. What paradigm can be built on the existing algorithms to further develop or find new ones – is being mapped out. The efficiency of specific algorithms is analyzed, optimized them, using learning algorithms. Finally, my achieved results are published.

4. SOME ASPECTS OF THE EFFICIENCY OF ALGORITHMS [5-8]

The goal of the analysis is to map system requirements.

Execution time:

- Can be measured by the number of instructions executed and it depends on the hardware.
- Does not necessarily depend on the number of input data.
- Can be the minimum, average and maximum.

Disk space/Memory space:

- To measure the variables and data structures associated with the demand.
- Depends on the data type and operation of the program, including language-specific and depends on the implementation, as well.
- Can be expressed in code size and memory storage costs occupied space requirement.
- Reducing the reservation generally increases the execution time.

Complexity:

- Global, when the algorithm as a whole has to be understood in order to reduce.
- Local, if the source code optimization (at least partially) is to be solved.
- It also depends on the data type and operation of the program, including language-specific and depends on the implementation
- Less objectively defined.
- Grouped as a logical and structural complexity.

We should always ask what matters and what does not; and what the reasonable compromises are.

5. HOW PR-COMP SOFTWARE WORKS

The pr-comp software is included in the pr-comp bundled software [1, 9]; it is one of the several physiologic signals processing software. This is an open-source product in C++, consisting of five parts: one for the definition and four as the controller.

The defs program defines the data structures used for storing data and parameters for the other programs of this module. The makepv reads WFDB record as vectors then the makecv converts covariance matrices from vector pattern processing. The makeev is able to process covariance matrices to make the ordered list of eigenvectors. After that the makefv program does the final joint processing. The mathematical background of algorithms is known.

6. ANALYSING OF THE PARTS OF THE PR-COMP

This analysing of the parts of the pr-comp was carried out in order to investigate the efficiency point of view. The upper limits of the execution time were reviewed: a linear number of steps was not founded, but there are quadratic and cubic one instead. The subsistence needs of space were tested: there were no higher than quadratic demand. When the complexity was reviewed: how many programming theorems can be used; are there methods or inner-nested methods; are there many levels of nested loops were distinguished. Details and specifications have already been published [10]. The following table summarizes the study results (Fig. 1).

	Execution time	Disk space/ Memory space	Complexity
makepv	N^2+k	$3N+k$	4 programming theorems methods 2-level nested loops
makecv	N^3+N^2+3N+k	$2N^2+N+k$	7 programming theorems 3-level nested loops
makeev	$2N^3+6N^2+N+k$	$3N^2+2N+k$	8 programming theorems inner/nested methods 3-level nested loops
makefv	N^3+2N^2+4N+k	$2N^2+2N+k$	5 programming theorems methods 3-level nested loops

Figure 1.
The 4 controller parts of the pr-comp software (efficiency)

7. RESULTS AND RECOMMENDATIONS

The results and recommendations are summarized here for the pr-comp software [11], giving a guideline for further development of the program:

- reduce the covariance matrix,
- rationalize data conversion,
- migrate to smarter data structure,
- teach schema, use pattern matching.

The pr-comp software is part of the PhysioBank ATM software [12]. In relation to the hierarchical structure, some conceptual proposals are given:

- use uniform data formats (at least avoid using such a huge variety),
- use OO design patterns to reduce logical and structural complexity,
- there are more efficient internal data structure and algorithm processing,
- use adaptable visualisation,
- global (rather than local) efficiency considerations are more appropriate in reaching our final goals.

8. SUMMARY

This research topic has always been and still is a current and modern one. After studying and understanding several ECG analyzer algorithms, the challenge will be to propose a paradigm for selecting and developing an integrated and comprehensive methodology. The development can progress further [13]. Theoretical paradigm must be found or produced, which will be useable regardless of the language.

9. REFERENCES

[1] PhysioToolkit, http://physionet.org/physiotools/software-index.shtml, 2013.04.05.
[2] libRASCH, http://www.librasch.org/librasch, 2013.04.05.
[3] eplimited, http://www.eplimited.com, 2013.04.05.
[4] Weka 2, http://www.cs.waikato.ac.nz/ml/weka, 2013.04.05.
[5] Zsakó L.: Módszeres programozás: Hatékonyság, Mikrológia 6., ELTE, 1999
[6] Szlávi P., Zsakó L., Temesvári T.: Módszeres programozás: A programkészítés technológiája, Mikrológia 21., ELTE, 2000
[7] Iványi A.: Informatikai algoritmusok I., II. ELTE, 2004., 2005., ISBN 9634637752
[8] Kaczur S.: A hozzárendelési feladat két megoldási módszerének összehasonlító elemzése, Informatika, A GDF Közleményei, Budapest, XI. évfolyam 2. szám, 2009, ISSN 1419-2527, p. 24-27
[9] pr-comp, http://www.physionet.org/tutorials/pr-comp, 2013.04.05.
[10] Kaczur S.: Néhány nyílt forráskódú EKG analizáló algoritmus hatékonysága, A Tudomány Hete a Dunaújvárosi Főiskolán, Dunaújváros, Dunaújvárosi Főiskola, 2011. november 7-11. (folyamatban)
[11] Bánhalmi A.: Adatbányászat 2. előadás, http://www.inf.u-szeged.hu/~banhalmi/adatb/2_eloadas.ppt, 2013.04.05.
[12] PhysioBank ATM, http://www.physionet.org/cgi-bin/atm/ATM, 2013.04.05.
[13] Kaczur S.: Részletes kutatási terv, ELTE Informatikai Doktori Iskola, 2011