Automatic tools for processing 1H Nuclear Magnetic Resonance data.

Franciszek BINCZYK, *Silesian University of Technology in Gliwice*
(prof. Joanna POLANSKA, *Silesian University of Technology in Gliwice*)

Abstract

Aim of the project was to develop a set of automatic tools for processing 1HNMR human brain data. Nuclear Magnetic Resonance (abr. NMR) is a very popular and widely used medical diagnostic technique. It is based on a phenomenon of magnetic resonance that take place for a nucleons placed in a homogenous magnetic field. As a next step series of electromagnetic pulses (abr. EMP) is applied that cause a change of magnetic spin. When pulse is released nucleons returns to the state of equilibrium. As a natural consequence an electromagnetic wave is emitted and measured by a receiver coil in the NMR scanner. According to the structure of a pulse a result might be observed as a wave consists of harmonics emitted from different types of compounds. Such a type of NMR test is called Magnetic Resonance Spectroscopy (abr. MRS) and is widely used in detection of tumour metabolite profile. Raw signal contains number of unwanted components that must be removed from the signal before metabolite quantification. It is done in a step called pre-processing. A step after, called post-processing, consist of modelling techniques that are used to determine a very accurate value for area under specific peak or group of peaks in frequency spectrum. Author with his supervisor proposed an automatic tool called GNMR for full processing of described data.

Second very popular NMR test is called Diffusion Tensor Imaging (abr. DTI). In this procedure EMP is emitted in such a configuration that it is possible to determine a water nucleons flow in number of directions. Raw data obtained from such are pre-processed in order to eliminate noise and bad measurements. In the final step information about nucleon movement for multiple directions is translated into geometrical estimate named tensor. Such sophisticated information is then used in a process of tractography (estimation of neural tracks) or is used for calculation of anisotropy maps that are very useful in tumour diagnostic process. During visit in DKFZ Heidelberg author proposed an implementation of software solution for tensor estimation. Both mentioned solutions are described in details in the paper.

1. Introduction

Tumours and Cancer are very badly diseases that are main causes of death or serious injure for many people. The most important things are: proper, precise diagnosis and proposition of effective treatment. There exist very popular, non-invasive technique called Nuclear Magnetic Resonance. The technique is based on a phenomenon of magnetic resonance that takes place for nucleons (tissue) placed in homogenous magnetic field. During the first step of test procedure, tissue is put under influence of a series of electromagnetic pulses. It forces a change of magnetic spin of a particle. When the pulse is released nucleons returns to the equilibrium state. As a consequence electromagnetic wave is emitted. There exists plenty of different pulse standards, every used for slightly different purpose. In this work author is mainly focused on two very popular types of NMR techniques.

2. Magnetic resonance spectroscopy

First of mentioned popular techniques is called spectroscopy and is mainly used for detection of tumour metabolite profile. Wave emitted by the particles is a sum of harmonics – each emitted by a group of nucleons in specific chemical compound defined by electron clouds and bond between nucleons. Such an emitted signal is then detected by receiver coil in NMR scanner. It is time domain so to observe harmonics it must be transferred from time into frequency domain. After that it still contains unwanted components that must be removed from the signal. This part of MRS data analysis is called pre-processing and consist of steps such as:

- Water peak subtraction
- Phase correction
- Baseline subtraction
- Noise filtration

There exist plenty of different techniques designed and developed to deal with the problem. In general they might be divided into two groups: manual and automatic. Author briefly studied demands of clinics and decided to implement and improve only automatic ones.

After pre-processing step signal is free of unwanted artefacts but it is still not perfect. At this step it is wise to notice that in analysis of MRS signal one should be interested in very precise measurement of area under specific peak. The problem is not trivial when one takes into account possibility of peak overlap and large discretization of data. Of course it is still possible to simply apply numeric integration but obtained result may contain quite large error. Good and wise decision is application of mathematical modelling in order to extract from the signal interesting information. After literature study authors decided to use GMM that is developed from the mixture of n Gaussian components which are independent Gaussian distributions. Every component may be described by a probability density function. Its parameters might be translated into peak parameters:

- Mean value-peak position;
- Variance-peak width;
- Weight- peak height.

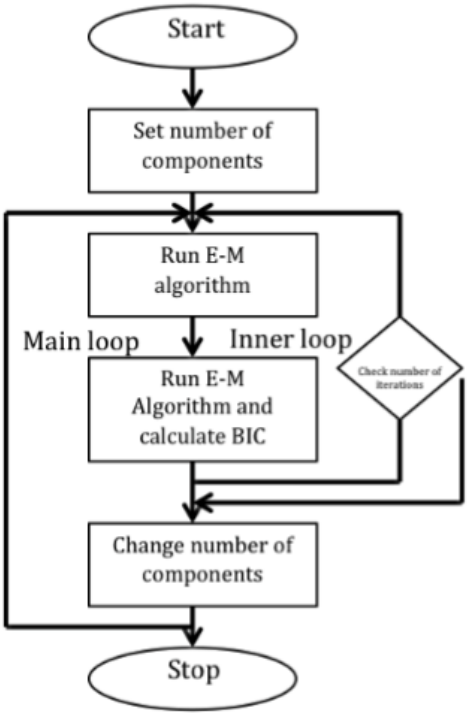

Figure1. Flow chart of proposed modeling routine

To calculate GMM, authors used Expectation-Maximization algorithm (abr. EM). Initial conditions for this method are chosen by random. In a consequence algorithm must be run several times in order to pick the best solution with respect to evaluated Bayesian Information Criterion (abr. BIC) Number of Gaussian components is not strictly define as well. The only limitations are initial and final number of components that are evaluated. The best solution among all calculated is chosen according to the value of calculated BIC. As a last step GMM is scaled to fit the original signal.

In order to fit demands a final result is presented in a form of metabolite map. Such a map is a result of calculation of specific metabolite amount in number of measurement points for which spectra were taken.

3. GNMR-tool for processing NMR spectra

Mentioned methodology was used by author in GNMR, a tool for automatic processing of NMR spectra. Proposed system is able to process spectra obtained from Siemens and Philips scanner. It performs full automatic pre-processing. In order to obtain precise and satisfactory results author has implemented and improved methods:

- For phase correction- adaptive Automics, Shannon entropy minimization, Ernst - dispersion integral minimization, Dispersion vs. absorption.
- For baseline subtraction- authors decided to use median method while it is widely used and obtained results are good. The method is based on so called moving window (a mask of size *n).* Baseline is obtained by derivation of median value of the signal inside the mask per each data point. Such a result is simply subtracted from the original signal
- For noise filtration the Savitzky-Golay filter was chosen. Obtained results were most accurate according to the assumptions that peak amplitude cannot be attenuated and signal to noise-ratio should be approximately at the level of 5%.

After pre-processing, signal is decomposed into mixture model. It is done with use of Dempster's EM algorithm. The flow chart of the whole modelling process is presented on the figure 1. The main window of the GNMR software with result in a form of metabolite map is shown on figure 2.

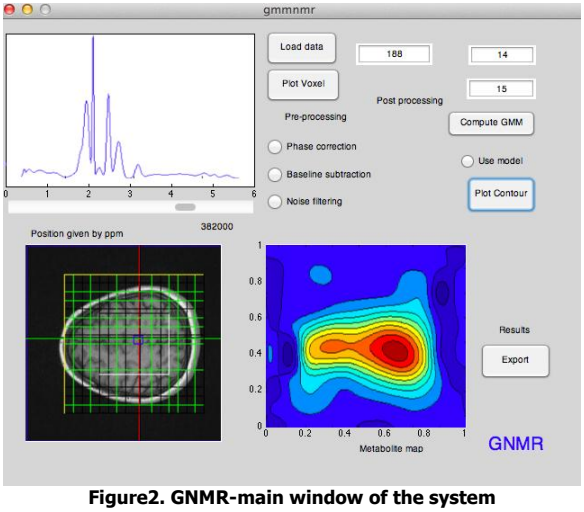

Figure2. GNMR-main window of the system

4. Diffusion tensor imaging

Second very popular NMR technique is called Diffusion tensor imaging. This technique differs from previously mentioned at the point of collection of signal. DTI is a method, which is selective to only one type of nucleons – water. Thanks to the phenomenon of NMR it is possible to change gradient field direction and measure signal after each modification. The process might be repeated for number of directions. In order to properly orient in space several times during test a so-called B0 signal is collected for which there is no gradient direction change. Signal is collected for defined 3-dimensional volume called voxel. Usually data are gathered for number of voxels placed in 2D in "slice" and many slices in 3D volume. Raw signal as previously must be pre- processed in order to remove or correct voxels that does not satisfy defined quality criterion. In case of DTI such a criterion is "grater than 0".
In a first step of pre processing each voxel is examined. If its value is not correct it may be corrected with use of mean value of its closed 26th neighbourhood. In many cases there is no possibility of correcting the voxels because its neighbours has bad value as well.
After pre-processing step it is possible to estimate a tensor from multiple directions measured for a voxel. The first step of tensor estimation is definition of design matrix:

$$D = direction * (-direction)$$

, where
Direction is a matrix consist of n x3 elements (n-number of gradient direction, 3number of dimensions)

Such a result must be then recalculated:

$$D = pseudo\ inverse(D * D') * D'$$

Knowing final form of design matrix one is able to estimate a tensor. In order to do so attenuation signal must be evaluated. It is a result of division of vector containing voxel values for each gradient direction by voxel value obtained for B0- no gradient direction. As a next step tensor structure is estimated:

$$T = D * \log(attenuation\_signal)$$

As a result a six-element tensor is obtained. It is more convenient to use 3x3 symmetric matrix notation of the form:

$$\begin{bmatrix} 1 & 5 & 6 \\ 4 & 2 & 4 \\ 6 & 5 & 3 \end{bmatrix}$$

, where number indicates components of 6 elements tensor.

At this step it is possible to introduce a second correction algorithm to the data. According to the literature eigenvalues of a good tensor should be non-negative. If after tensor calculation it appears that at least one of them is negative it is assumed that for some direction voxel has improper value.
In a consequence voxel value in such a direction is substituted with a mean value of its 26 neighbours. It is done repetitively for each direction.

When all tensors are corrected they are checked again. If number of tensors that had negative eigenvalues in previous procedures are decreased that means – there is a sense of performing correction again. If number of bad tensors remains the same or increased there is no point in performing correction. Processed data are then used in a more sophisticated algorithm of neural track estimation or calculation of anisotropy maps.

5. MITK- DTI tensor reconstruction filter

Knowledge about tensor reconstruction was used by author to implement additional filter into MITK toolkit that is available as open source. Project was done during authors LLP Erasmus stay in Deutsches Krebs forschungs zentrum in Heidelberg. Proposed algorithm was examined with use of large set of DTI data. It was proven that for some cases it works much better than other implemented tensor reconstruction routines. Exemplary tensor map is presented on the figure 3.

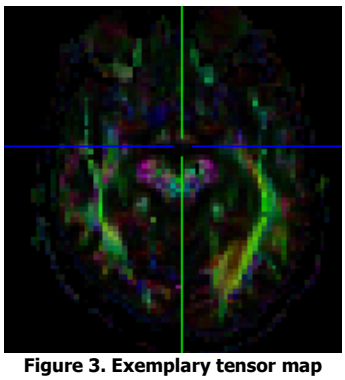

Figure 3. Exemplary tensor map

6. Final remarks

As it was shown as a result of project two tools for automatic analysis of 1H NMR data were developed. Both of them were validated with use of number of clinic data. Authors observed its all pros and cons. All weaknesses will be removed during future research.

7. Acknowledgments

Author is very grateful to the employees of Cancer Center in Gliwice, Prof. Barbara Bobek Bilewicz, Prof. Rafał Tarnawski and Anna Hebda, for providing help and knowledge about medical data.

Author received a grant under the project DoktoRIS - Scholarship program for innovative Silesia co-financed by the European Union under the European Social Fund.

References

[1]. Graff RA: *In vivo NMR spectroscopy. Principles and Techniques*. Wiley (2007)

[2]. P.Hagman et al.: *Understanding Diffusion MR Imaging Tech- niques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond,* RadioGraphics 2006; 26:S205–S223

[3]. Ofer Pasternak, Nir Sochen, Yaniv Gur, Nathan Intrator and Yaniv Assaf , "Free Water Elimination and Mapping from Diffusion MRI". Magentic Resonance in Medicine, vol. 62, pp. 717-730, July 2009.

[4]. D.S.Tuch: *„Q-Ball Imaging",*Magn. Res. Med. 52, 1358-1372, (2004).

[5]. Wang T et al.: *Automics: an integrated platform for NMR-based metabonomics spectra processing and data analysis* BMC Bioinformatics 10:83. (2009)

[6]. Xi Y, Rocke DM: *Baseline Correction for NMR Spectroscopic Metabolomics Data Analysis*, BMC Bioinformatics 9:324 (2008)

[7]. McLachlan G, Peel D: *Finite Mixture Models*, Wiley (2000)

[8]. Pietrowska M et al.: *Optimizing of MALDi-ToF based low-molecular-weight serum proteome pattern analysis In detection of breast cancer* ,Neoplasma 57(6):537-44 (2010)

[9]. Dempster A.P et al.: Maximum likelihood from in- complete data via the EM algorithm. *Journal of the Royal Statistical Society B,* 39(1), (1977), 1-22.

Author:

MSc. Franciszek Binczyk
Silesian University
of Technology
ul. Akademicka 16
44-100 Gliwice
tel. 792727518
email: *franciszek.e.binczyk@polsl.pl*