Regression analysis for paths inference in a novel Proton CT system

Liyun Gong, Miao Yu, Xujiong Ye and Nigel Allinson
School of Computer Science
University of Lincoln, UK
Email: {lgong, myu, xye, nallinson}@lincoln.ac.uk

***Abstract*—In this work, we analyse the proton paths inference for the construction of CT imagery based on a new proton CT proton system, which can record multiple proton paths/residual energies. Based on the recorded paths of multiple protons, every proton path is inferred. The inferred proton paths can then be used for the residual energies detection and CT imagery construction for analyzing a specific tissue. Different regression methods (linear regression and Gaussian process regression models) are exploited for the path inference of every proton in this work. The studies on a recorded proton trajectories dataset show that the Gaussian process regression method achieves better accuracies for the path inference, from both path assignment accuracy and root mean square errors (RMSEs) studies.**

I. Introduction

Proton therapy is becoming more and more important in the field of radiotherapy, i.e., CT imagery for treatment planning of cancer, as it can deliver the planned dose of radiations over a specific region of the tumour,minimising the damage to healthy tissue [1]. For a typical proton CT instrument, protons are emitted and individual incident and exit paths before/after passing through the tissues are recorded by the two pairs of proton trackers (position sensitive detectors); while the residual energy of protons are recorded using a scintillator calorimeter. Based on the recoded paths and residual energy, corresponding CT imagery can be constructed. A limitation of such a CT instrument is that only *one event* (i.e., incident/exit paths of one proton and related residual energy) can be recorded per instrument cycle, which is time-consuming.

Recently, a new proton CT system is developed [2]. Compared with the traditional one, multiple silicon strip sensors (SSDs) operating at a multiple of the cyclotron (proton generator) frequency are used as proton trackers; while the residual energy can be recorded using a 'range telescope' - an assemble of closely-packed large area CMOS imagers. These CMOS imagers are 2-dimensional sensors which allow multiple protons to pass through during a certain time interval and store their residual energies. Instead of recording only one event per instrument cycle, *multiple events* (i.e., paths of multiple emitted protons passing through the range telescope and related residual energies) can be recorded by the new proton CT system, which can potentially increase the efficiency of the CT imagery generation.

From the recorded multiple events, the path of every individual proton going through the range telescope needs to be inferred. The corresponding residual energy can then be derived by finding the recorded one whose position is closest to the inferred path for the CT imagery construction. One straightforward idea for the path inference is to apply multiple target tracking algorithms [3] to track the paths of multiple protons by data association (i.e., associating each proton with its corresponding measurement) [4]; however, they are usually time-consuming which is not suitable for the real-time application.

In this work, we refer to the regression method for solving the path inference problem. By exploiting the regression method [5], we estimate the relationship between the entry position and exit position of protons for every range telescope layer, which derives the inferred individual path. Two types of regression methods have been tested in this work:

i).Linear regression, which assumes linear relationship between entry/exit positions of every layer

ii).Gaussian process regression methods, which assumes non-linear position relationship considering both the reaction and reflection effects as protons pass through different layers

The rest of the paper is organized as follows: Section II gives the overview of the proposed proton CT system as in [2]. Regression methods based proton trajectory inferring are presented in Section III. Experimental results and comparative analyses are provided in Section IV and the final conclusion is given in Section IV.

II. Overview of the proton CT system

The overview of the new developed proton CT system is shown in Fig. 1. The proton trackers on both sides of the patient can detect each proton and its position. They are composed of multiple silicon strip sensors (SSDs) operating at a multiple of the proton generator frequency. The paths and residual energies of multiple protons can be recorded using a range telescope - a sandwich of closely-packed large area CMOS imagers, which have passive absorbers between each other as shown in Fig. 2. For

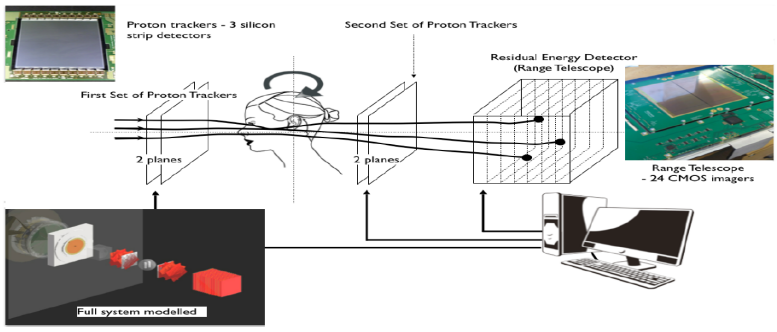

Fig. 1. The new CT system which can record multiple proton paths and residual energies.

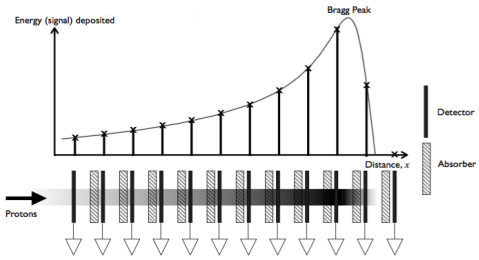

Fig. 2. The structure of stacked CMOS imagers.

a particular proton, its residual energy is given by the location of the Bragg Peak, which is recorded as the maximum absorbed energy on the CMOS imagers for this proton.

During a certain time interval, multiple protons could pass through the CMOS imagers while the residual energies of multiple protons are recorded. By applying a particular regression model, we can infer the path of each proton, from which its related recorded residual energy can be obtained for the CT image construction.

III. Regression for the Path Inference

The positions relationship between different layers can be constructed through regression models, from which the path of each proton through the CMOS imager can be inferred.

A. Linear regression

For the linear regression, it is assumed that there exists linear relationship between input $\mathbf{x}$ and output y, with:

$$y = \mathbf{w}^T \mathbf{x} \tag{1}$$

where $\mathbf{w}$ represents a weight vector. Based on a training dataset $\{(\mathbf{x}_1, y_1)..., (\mathbf{x}_N, y_N)\}$, we can find the optimal $\mathbf{w}$ by minimizing the following cost function F representing errors between the actual values and predicted results as:

$$F = \|X\mathbf{w}^T - \mathbf{y}\|^2 \tag{2}$$

where $X = [\mathbf{x}_1, ..., \mathbf{x}_N]^T$ and $\mathbf{y} = [y_1, ..., y_N]^T$.

By setting the derivation of (1) to be zero, we can find the optimal $\mathbf{w}^*$ as:

$$\mathbf{w}^* = (X^T X)^{-1} X^T \mathbf{y} \tag{3}$$

which is used for predicting the output y^* when new input $\mathbf{x}^*$ comes.

B. Gaussian process regression

Different from traditional regression methods (i.e., linear regression), it estimates both the predicted output value and uncertainty based on an input vector. Gaussian process regression is based on the following model:

$$y = \mathbf{w}^T \phi(\mathbf{x}) + \varepsilon, \tag{4}$$

where $\mathbf{x}$, y represent the input and output, $\phi(\cdot)$ is a mapping function which maps the input $\mathbf{x}$ to a higher dimensional space for modelling the non-linear relationship between $\mathbf{x}$ and y, ε represents a noise term with the following distribution:

$$\varepsilon \sim N(0, \sigma^2). \tag{5}$$

where $N(0, \sigma^2)$ represents a Gaussian distribution with the mean 0 and standard deviation σ.

The model (4) can be trained based on a dataset containing N input-output pairs–$\{(\mathbf{x}_1, y_1)..., (\mathbf{x}_N, y_1)\}$ associated with that model. According to equations (4) and (5), we have:

$$
\begin{aligned}
p(\mathbf{y}|X,\mathbf{w}) &= \prod_{i=1}^{N} p(y_i|\phi(\mathbf{x}),w) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y_i-\phi(\mathbf{x})^T\mathbf{w})^2}{2\sigma^2}) \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}} exp(-\frac{1}{2\sigma^2}\|\mathbf{y}-\Phi^T(X)\mathbf{w}\|^2) \\
&= N(\Phi^T(X)\mathbf{w},\sigma^2\mathbf{I})
\end{aligned} \tag{6}
$$

where the input vector ensemble $X = \{\mathbf{x}_1,...,\mathbf{x}_N\}$, $\Phi(X) = \{\phi(\mathbf{x}_1),...,\phi(\mathbf{x}_N)\}$, $\mathbf{y} = [y_1,...,y_N]$ represents the corresponding outputs and $\mathbf{I}$ represents an identity matrix.

Meanwhile, we assume that weights $\mathbf{w}$ follow a Gaussian distribution:

$$
\mathbf{w} \sim N(\mathbf{0},\Sigma_{\mathbf{w}}) \tag{7}
$$

According to the Bayesian rule, we have:

$$
p(\mathbf{w}|\mathbf{y},X) = \frac{p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)} \propto p(\mathbf{y}|X,\mathbf{w})p(\mathbf{w}) \tag{8}
$$

By substituting (6) and (7) into (8), we have:

$$
\begin{aligned}
&p(\mathbf{w}|\mathbf{y},X) \propto exp(-\frac{1}{2\sigma^2}(\mathbf{y}-\Phi(X)^T\mathbf{w})^T(\mathbf{y}-\Phi(X)^T\mathbf{w})) \\
&exp(-\frac{1}{2}\mathbf{w}^T\Sigma_{\mathbf{w}}\mathbf{w}) \\
&\propto exp(-\frac{1}{2}(\mathbf{w}-\bar{\mathbf{w}})^T(\frac{1}{\sigma^2}(\Phi(X)\Phi(X)^T+\Sigma_{\mathbf{w}}^{-1})(\mathbf{w}-\bar{\mathbf{w}}))
\end{aligned} \tag{9}
$$

where $\bar{\mathbf{w}} = \sigma^{-2}(\sigma^{-2}\Phi(X)\Phi(X)^T+\Sigma_{\mathbf{w}}^{-1})^{-1}\Phi(X)\mathbf{y}$. From (9), we can see that the posterior of $\mathbf{w}$ based on the training data follows a Gaussian distribution, with:

$$
p(\mathbf{w}|\mathbf{y},X) \sim N(\bar{\mathbf{w}} = \sigma^{-2}A^{-1}\Phi(X)\mathbf{y}, A^{-1}) \tag{10}
$$

where $A = \sigma^{-2}\Phi(X)\Phi(X)^T + \Sigma_{\mathbf{w}}^{-1}$.

For a new input $\mathbf{x}^*$, the distribution of the predicted value y^* is also a Gaussian based one, which can be estimated from (4) and (10) as:

$$
p(y^*|\mathbf{x}^*,\mathbf{y},X) = N(\frac{1}{\sigma^2}\phi^T(\mathbf{x}^*)A^{-1}\Phi(X)\mathbf{y}, \phi^T(\mathbf{x}^*)A^{-1}\phi(\mathbf{x}^*)) \tag{11}
$$

(11) forms the basis of the Gaussian process regression. From (11), we obtain a whole distribution related to the predicted y^* instead of a single value, from which more information (i.e., the predicted value and uncertainty) can be obtained.

However, sometimes the dimensionality of a mapped feature $\phi(\mathbf{x})$ may be high for highly non-linear regression problems or even without an explicit representation [5], thus (11) can not be derived directly. To solve this issue, we can re-write (11) in the following form by applying the matrix operations and the matrix inverse lemma [6] as:

$$
\begin{aligned}
&p(y^*|\mathbf{x}^*,\mathbf{y},X) = N(\phi^T(\mathbf{x}^*)\Sigma_{\mathbf{w}}\Phi(X)(K+\sigma^2\mathbf{I})^{-1}\mathbf{y}, \\
&\phi^T(\mathbf{x}^*)\Sigma_w\phi(\mathbf{x}^*) - \phi^T(\mathbf{x}^*)\Sigma_w\Phi(X)(K+\sigma^2\mathbf{I})^{-1} \\
&\Phi^T(X)\Sigma_w\Sigma_w\phi(\mathbf{x}^*)
\end{aligned} \tag{12}
$$

where $K = \Phi^T(X)\Sigma_{\mathbf{w}}\Phi(X)$.

According to [5], the concept of the kernel function is introduced and defined in (13) as:

$$
k(\mathbf{x}_i,\mathbf{x}_j) = \phi^T(\mathbf{x}_i)\Sigma_{\mathbf{w}}\phi(\mathbf{x}_j) \tag{13}
$$

From (13), (12) can be represented by the kernel function as:

$$
\begin{aligned}
&p(y^*|\mathbf{x}^*,\mathbf{y},X) = N(\mathbf{k}(\mathbf{x}^*,X)(K+\sigma^2\mathbf{I})^{-1}\mathbf{y}, \\
&k(\mathbf{x}^*,\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*,X)(K+\sigma^2\mathbf{I})^{-1}\mathbf{k}^T(\mathbf{x}^*,X)
\end{aligned} \tag{14}
$$

where $\mathbf{k}(\mathbf{x}^*,X) = [k(\mathbf{x}^*,\mathbf{x}_1),...,k(\mathbf{x}^*,\mathbf{x}_N)]$. Different types of kernel functions can be chosen in (14), in this work a Gaussian kernel–$k(x,x') = exp(-\gamma\|x-x'\|)$ is chosen, with γ being a parameter controlling the radius of this kernel.

From a particular regression model, a proton path will be inferred by mapping the positions between every pair of consecutive layers for every proton by the model. The corresponding proton residual energy can then be further detected as the largest recorded energy value among the ones whose corresponding position is closest to the inferred path for every layer, from which the CT imagery can be finally obtained.

IV. Experimental results

In the experimental studies, we evaluate the paths inference by different regression models. The dataset we use come from the University of Birmingham [7], containing both the positions and recorded energy values of protons from thirteen CMOS layers in the telescope range. For a preliminary study, it is assumed that the recorded energies single point measurements (spread clutter measuremnts will be investigated for future studies) and all protons have gone through the thirteen CMOS layer. 100 proton trajectories are chosen which are shown in Fig. 3, with 75 of which being chosen as the training data and 25 of which being chosen as testing. Both linear and Gaussian models have been trained for deriving position relationships of protons between consecutive layers according to the training dataset.

For every single particular proton, its position on every single layer is predicted by the trained regression model. The measurements whose positions are closet to the predicted ones for every layer are connected as the inferred path. Table I shows the percentage of the 'correctly inferred path' (if the inferred path for a proton is equivalent to the groundtruth one) for the test dataset. Results show that the Gaussian process regression achieves a better performance with more paths being correctly inferred.

TABLE I
THE MEAN AND STANDARD DEVIATION OF RMSES FOR ALL LAYERS.

	Linear regression	Gaussian process regression
Mean (1st dimensionality)	2.53	1.48
Std (1st dimensionality)	1.11	0.71
Mean (2nd dimensionality)	2.57	1.18
Std (2nd dimensionality)	1.22	0.56
Mean (total)	3.61	1.91
Std (total)	1.63	0.88

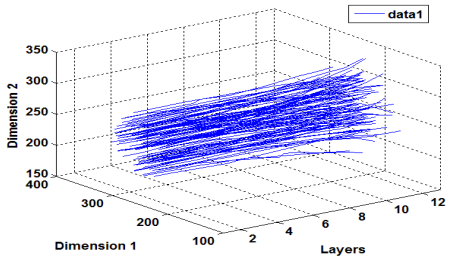

Fig. 3. The proton trajectories used in the experiment.

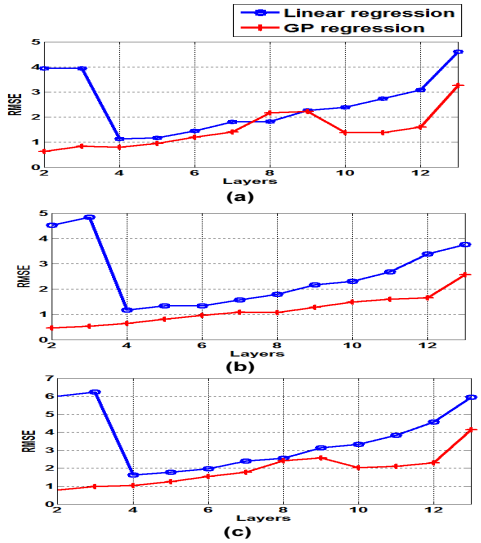

Fig. 4. The RMSEs comparisons between the inferred and groundtruth proton positions for every layer by different regression models. (a). RMSEs comparison for the Dimension 1 of the CMOS imager (b). RMSEs comparison for the Dimension 2 of the CMOS imager (c). 2-D RMSEs comparison of the CMOS imager

Next, we calculate the RMSEs between the groundtruth and predicted positions for protons in the test dataset. The averaged RMSEs of all tested protons for every layer are plotted in the Fig. 4. Furthermore,both the mean and standard deviation of the averaged RMSEs for all layers are calculated, which are shown in Table II. Again, we can see that the Gaussian process method achieves a better performance, with both smaller mean and standard deviation of errors. The reason why the Gaussian process method achieves the best performance is that: there exist reaction and reaction for the proton paths which makes paths non-linear while Gaussian process regression can better deal with these non-linear properties.

TABLE II
THE ASSIGNMENT ACCURACIES BY DIFFERENT REGRESSION METHODS.

	Linear regression	GP regression
Accuracy	68%	72%

V. CONCLUSION

In this work, we have evaluated the performance of different regression models, for the inferring the proton paths in the range telescope of a new proton CT system, which can be applied for the CT imagery construction. The results show that the Gaussian regression method achieves better performance than the linear regression method.

Our future works could be expanded from the following two directions: firstly, we will increase the size of the dataset (both training and testing) for evaluation; besides, other machine learning methods which can better deal with the large dataset, such as deep neural networks, will be applied for constructing the corresponding regression models.

REFERENCES

[1] J. Taylor, P. Allport, G. Casse, N.A. Smith, I. Tsurin, N. Allinson, M. Esposito, A. Kacperek, and J. Nieto-Camero, "Proton tracking for medical imaging and dosimetry," *Journal of Instrumentation*, vol. 10, 2015.

[2] T. Price, M. Esposito, G. Poludniowski, J. Taylor, C. Waltham, D. Parker, S. Green, S. Manolopoulos, N. Allinson, and T. Anaxagoras, "Expected proton signal sizes in the pravda range telescope for proton computed tomography," *Journal of Instrumentation*, vol. 10, 2015.

[3] S. Blackman and R. Popoli, "Design and analysis of modern tracking systems," in *Artech House, Norwood, MA*, 1999.

[4] R. Collins, "Multitarget data association with higher-order motion models," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA*, 2012.

[5] C. Bishop, "Pattern recognition and machine learning," in *Springer*, 2007.

[6] G. Strang, "Introduction to linear algebra, 5th edition," in *Wellesley-Cambridge Press*, 2016.

[7] "The PRaVDA consortium, http://www.pravda.co.uk, 2014, accessed October 8," .