Error Propagation

20th October 2005

Abstract

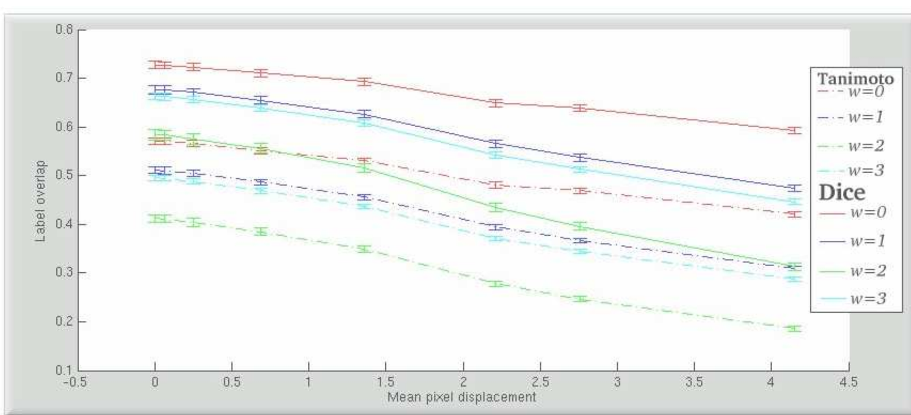
This technical document assembles and presents a collection of notes on how one should calculate sensitivity of measures such as model Specificity and Generalisability, as well as label overlap. These measures are used, in this particular context, to evaluate non-rigid registration. Since they are separable families of methods, they ought to be compared somehow; the eventual comparison must be compact and easily-comprehended too. We have chosen the notion of sensitivity as potentially the most valid way of reasoning about these two methods and comparing them with little bias, if any at all. Derivation of error bars is included throughout and is also the primary point of focus of this document, which is intended to provide better understanding of the current methods.

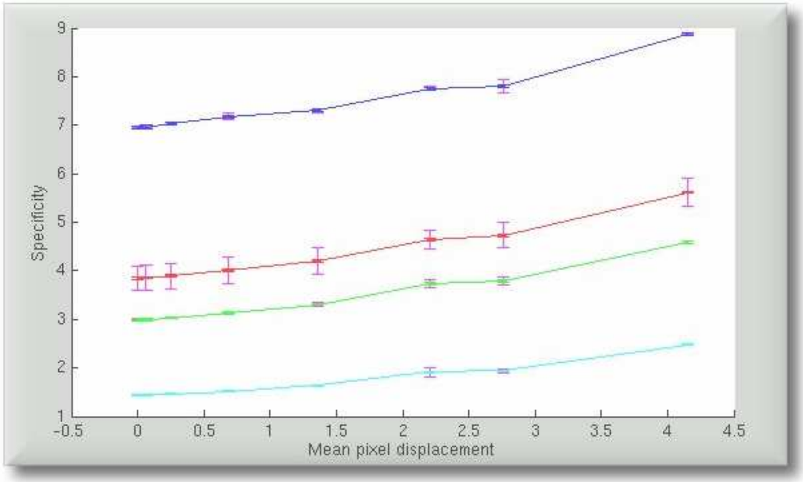
We begin with a brief introduction to the problem in question and then describe the variables that we have available. A derivation of sensitivity is then described in some depth and the corresponding errors too. Lastly, aggregation of figures that we handle as to make them more compact, is described somewhat briefly. Bound to that description is the formulation that we use to propagate the error bars, which altogether shape our final figure of interest.

1 Background

Example plots are included in this section, which attempts to illustrate the general aims and add some context. We are investigating plots which depict a certain measure m versus the extent of mean pixel displacement. These plots typically reflect on some measure of 'goodness' as data is degraded.

Mean pixel displacement is closely-related to the degree of *mis-registration* in the set of images under investigation. In simple terms, as we move rightwards in the plots, the overlap (or *correspondence*) among the images is made worse. We expect our measure m to be able to detect these overlap changes, preferably with a linear response too, which implies robustness. Below are 2 blended figures which practically visualise that. In the former case, overlap between label in the images is used as our measure, whereas model-based measures are used in the latter.





It is worth noticing that all curves are sampled at 8 separate points. Each such point corresponds to one particular value of mean pixel displacement. The value measured, namely m, is calculated over 10 instantiations of an image set, from which the average has been derived. This repeatability factor will later enable us to argue that the methods work consistently, so no handpicking or fluke are involved.

We are interested in two separate 'sources' of error (uncertainty):

1. **Error that is associated with the instantiation process.** As the number of instantiation is finite, our measures are susceptible to some fluctuation. We absolutely must account for bias, which is due to the random instantiation process. In our circumstance, we have 10 instantiations to consider.

2. **Error that is associated with the calculation of the value** m. To obtain the values which we seek, a set of synthetic images is generated. Since that set is limited in term of its size, corresponding error bars need to be bound.

2 Combining the Errors

To get meaningful plots, where error bars faithfully reflect on truth, we ought to better understand our error sources. They appear to be independent, but this observation does not simplify matters.

Let us look at the plots vertically, taking one value of mean pixel displacement at a time. Each value of displacement, denoted by d, is produced by warping a set of aligned images, i.e. images where there is no inherent displacement.

Let us define σ_{mi} to be the predicted error in the estimate of m for a warp instance $i : (1..N)$. We can then obtain the mean

$$\overline{\sigma_m} = \sum_i \frac{\sigma_{mi}}{N} \tag{1}$$

and the standard error is thus

$$SE_{\sigma_m} = \frac{SD(\sigma_{mi})}{\sqrt{N-1}}. \tag{2}$$

The above error simply tells us the certainty we have in our values of Specificity, Generalisability, or overlap, averaged over 10 instantiations.

As for the mean of the measurement m for the given displacement value m,

$$\overline{m_d} = \sum_i m_i/N \tag{3}$$

and the corresponding standard error

$$SE_{\overline{m}} = \frac{SD(m_i)}{\sqrt{N-1}}. \tag{4}$$

This error corresponds to the effect of having 10 different values measured for a given extent of displacement. The values tend to fluctuate slightly, thereby leading to uncertainty. This uncertainty is separate, however, from that which is associated with the computation of the measure itself.

3 Sensitivity

Sensitivity allows us to reason about the ability of a given measure to discern, by means of simple calculation, one magnitude of displacement from another. The value of sensitivity is affected by uncertainty as well – uncertainty which must be propagated from formulations in §2.

We introduce another index j to avoid confusion with i, which re-appears once the formulae are expanded. The sensitivity at a point j in our new sensitivity plot should be computed as follows:

$$Sens_j = \frac{(\overline{m_j} - \overline{m_0})}{d_j} / (\overline{\sigma_m})_j = \frac{TOP}{BOTT} \tag{5}$$

TOP and $BOTT$ are used merely to serve as aliases, which later reduce the complexity (scale) of the equations, making them more digestible.

4 Error in Sensitivity

From the above, the corresponding errors can now be derived:

$$\sigma^2_{TOP} = (SE_{\overline{m}})^2_j + (SE_{\overline{m}})^2_0 \tag{6}$$

$$\sigma^2_{BOTT} = (SE_{\sigma_m})^2_j \tag{7}$$

$$\sigma^2_{Sens_j} = F(\sigma^2_{TOP}, \sigma^2_{BOTT}). \tag{8}$$

The equation should be expanded to become

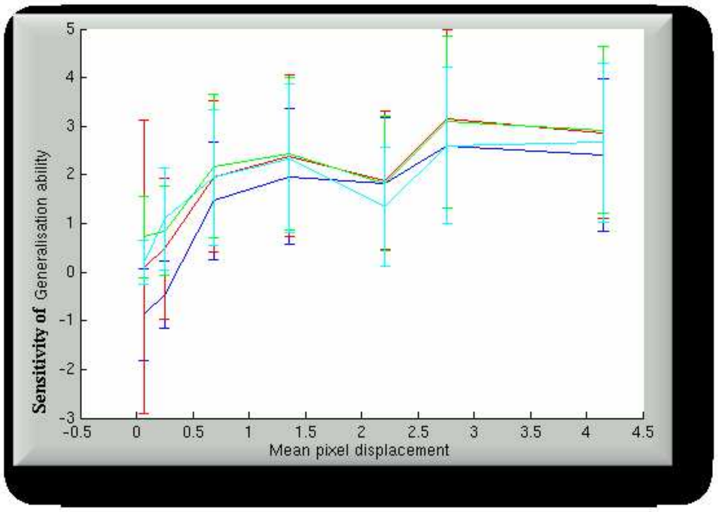
$$(\frac{\sigma_{Sens_j}}{Sens_j})^2 = (\frac{\sigma_{TOP}}{TOP})^2 + (\frac{\sigma_{BOTT}}{BOTT})^2 - 2(\frac{\sigma_{TOP}\sigma_{BOTT}}{TOP})(\frac{\sigma_{TOP}\sigma_{BOTT}}{BOTT}) \tag{9}$$

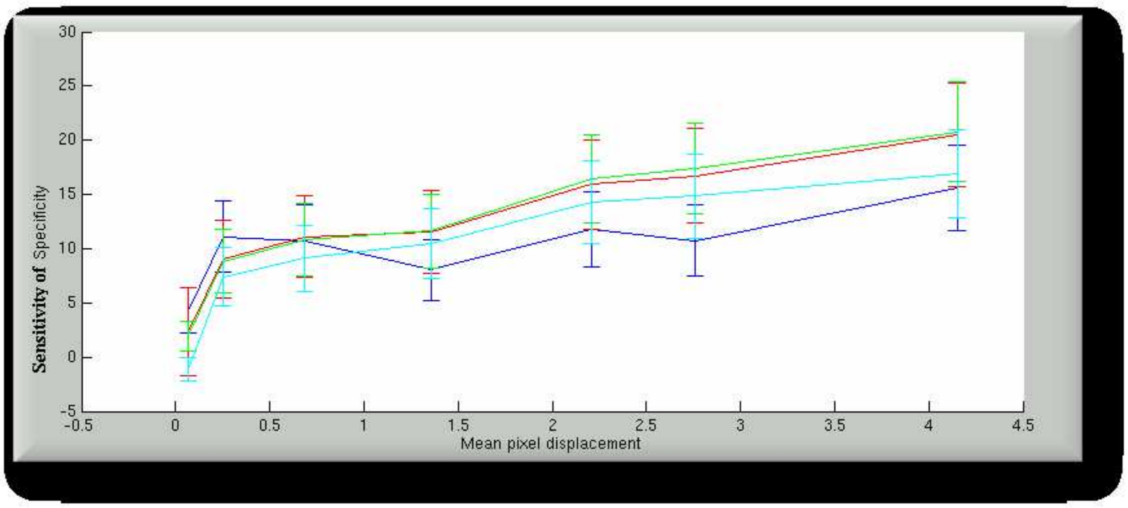
and then reduced to

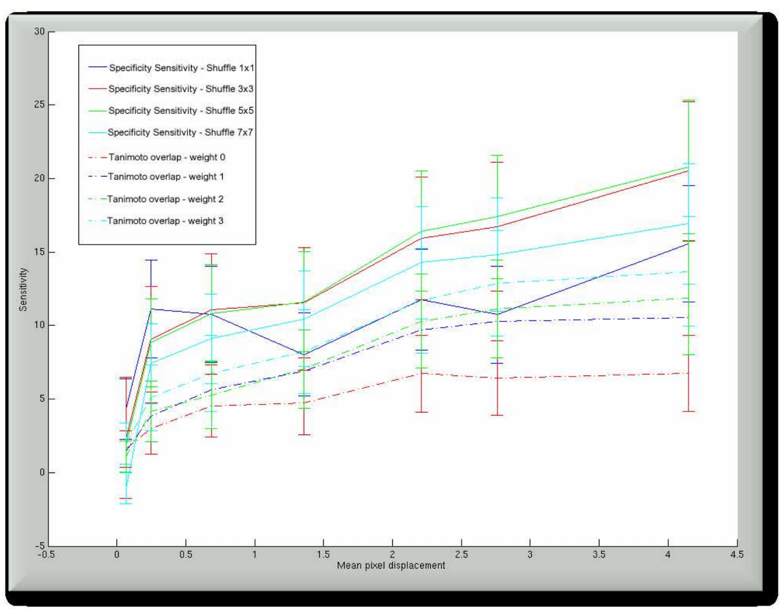
$$\sigma_{Sens_j} = Sens_j\sqrt{(\frac{\sigma_{TOP}}{TOP})^2 + (\frac{\sigma_{BOTT}}{BOTT})^2 - 2(\frac{\sigma_{TOP}\sigma_{BOTT}}{TOP})(\frac{\sigma_{TOP}\sigma_{BOTT}}{BOTT})} \tag{10}$$

which corresponds to the method of propagating errors in the case of division (see Web link below).

Finally, some figures of sensitivity can be generated. They can either be plotted apart or blended for comparison purposes.


Sensitivity of Generalisation ability
Mean pixel displacement


Sensitivity of Specificity
Mean pixel displacement

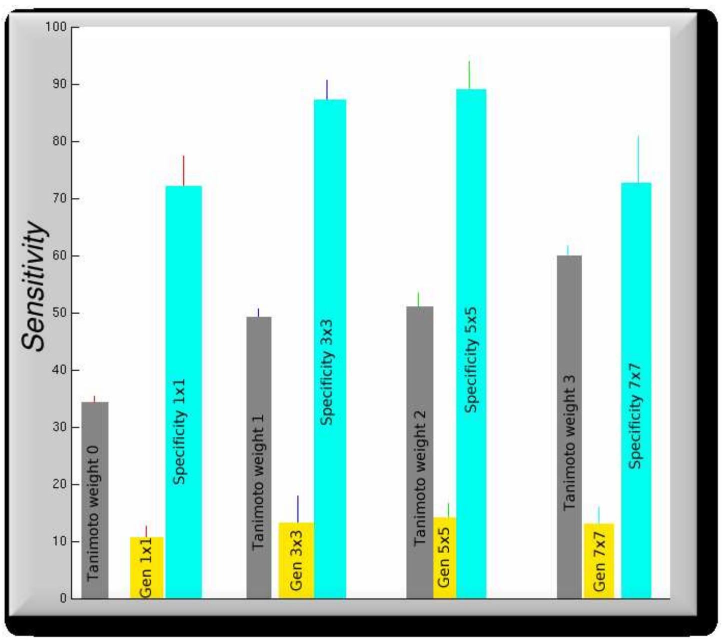

Specificity Sensitivity - Shuffle 1x1
Specificity Sensitivity - Shuffle 3x3
Specificity Sensitivity - Shuffle 5x5
Specificity Sensitivity - Shuffle 7x7
Tanimoto overlap - weight 0
Tanimoto overlap - weight 1
Tanimoto overlap - weight 2
Tanimoto overlap - weight 3
Sensitivity
Mean pixel displacement

At the last stage, we seek to aggregate the sensitivities by accumulating their values. The corresponding error bars must be propagated properly in accordance with rules of error propagation for addition.

Since in our case particular case, 8 values[1] are being aggregated, we retain a running sum and at each point re-calculate the error bars. The process can be expressed mathematically as follows:

$$\sigma_{Aggregated\, Sens} = \sqrt{\sum_{i} \sigma^2_{Sens_j} + \sigma^2_{Sens_{j+1}} - 2\sigma_{Sens_j}\sigma_{Sens_{j+1}}}. \tag{11}$$

This given an initial (very coarse) figure that serves as somewhat of a benchmark for our methods. This benchmark, however, is tightly-dependent on the hypothesis that sensitivity, as we decided to formulate its definition, is the correct criterion for appraisal.



For more details on error propagation, confer:

http://mathworld.wolfram.com/ErrorPropagation.html

[1] There are 8 values of the mean pixel displacement, although in principle, data for 16 values of the mean pixel displacement is available, but has not been processed, yet.