Tina Memo No. 2012-004

To Appear in Proc. Remote Sensing and Photogrammetry Soc. Conf. 2012.

A Quantitative Approach to the Analysis of Planetary Terrains.

P.D.Tar, N.A.Thacker, M.A.Jones and J.D.Gilmour.

Last updated
21 /8 / 2012

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

A Quantitative Approach to the Analysis of Planetary Terrains

Tar P.D., Thacker N.A., Jones M.A. Gilmour J.D.

School of Earth Atmospheric and Environmental Sciences, University of Manchester UK
Contact details: 0044798 2428522; paul.d.tar@gmail.com

Abstract

The aim of this paper is to introduce a quantitative approach to the use of pattern recognition in image analysis tasks. We investigate the use of Bayes Theorem in regional quantitation for a texture (terrain) based interpretation of satellite images of planetary surfaces. We show how the conventional approaches to density based pattern recognition can be interpreted as a likelihood estimation of quantities. We then use the methodologies of error estimation and propagation to develop a general theory for the prediction of the statistical and systematic errors required to interpret the level of agreement between theory and measured data. We argue that such techniques are necessary when applying these methods to scientific investigations and demonstrate the limits of validity for computed uncertainties in real satellite image data. In so doing we illustrate how conventional ROC style evaluation of algorithms are inappropriate, both for calibration of uncertainty and also as the sole mechanism for selecting between analysis techniques in scientific tasks. We conclude that the main challenge in scientific applications relates to our inability to assess the suitability of a trained pattern recognition system for application to incoming data.

Keywords: *imaging, quantitative, errors, terrain, classification*

1. Introduction

Imaging of planetary surfaces is a high profile element of missions to explore the solar system. Over the last 10-15 years orbiting missions have returned images of the Martian surface with unprecedented resolution. Current missions Messenger, Dawn and Lunar Reconnaissance Orbiter are providing high resolution coverage of the surfaces of Mercury, Vesta and the Moon, respectively. Within the next few years Juno and New Horizons will provide new images of Jupiter and its moons, and for the first time in the history of planetary science close-up images of Pluto. However, the quantity of available surface imagery has surpassed the capacity of individuals to make first pass assessments seeking the presence or absence of features diagnostic of terrain modifying processes which may be of interest to planetary researchers.

There is a clear need for a reliable automated solution for applications including identifying evidence of drainage networks or desiccation, whose distribution across a planetary surface is diagnostic of past or present climate patterns. There is interest in the mapping of dunes and their relationship with climate, wind direction and grain availability. The size, quantity and orientation of surface fissures are of interest, as they can be used to infer the presence of stress fields. There is also significant interest in the determination of relative surface ages from size frequency distributions (SFDs) of impact craters. All of these applications involve, as a first stage, the classification of terrain types and the estimation of quantities of terrain types present within regions of interest. This is a machine learning and pattern recognition problem which has not, thus far, been solved in a quantitatively (or scientifically) meaningful way.

A quantitatively valid solution requires not only estimated quantities of terrain types but also uncertainties, i.e. confidence intervals (error bars) on measurements, as measurements can only be meaningfully compared with scientific predictions (theories / models) if the measurement accuracy is known. Basic error analysis of classification and pattern recognition systems involves counting misclassification rates (error rates) on training or calibration sets. A more comprehensive analysis involves the construction of Reciever Operating Characteristics (ROC) curves showing error rates over a range of parameters. Both methods are empirical in nature, giving a set of singular statements about a classifier's performance on particular sets of data. However, reliance upon the resulting errors as estimates of future errors is problematic as it assumes training data is strongly representative of incoming data. Upon manual visual inspection incoming images may subjectively be considered representative, but even small perturbations in data distributions can invalidate the use of prior error estimates. We believe that the lack of a rigorous theoretical analysis of uncertainties in related work has prevented researchers from trusting automated systems, and hence forced their adoption of semi-automated and crowd-sourcing alternatives such as Moon Zoo (Joy K et al. 2011). These enlist members of the public to assist with the quantity estimation of basic features such as craters and boulders. Our work towards the creation of a trustworthy automated system emphasises the needs of scientists with a primary criterion of success being the ability to reliably predict errors on estimated quantities derived from, and applicable to, previously unseen incoming data.

2. Methodology

Our methodology for summarising the content of images produces two important outputs:

- a vector, **Q**, of quantities where each element, Q_K, is an estimate of the amount of terrain of type K (e.g. dunes, fissures, etc. as defined by the user) present within the data;
- and a covariance matrix, **C**, giving a theoretical estimate of the certainty with which quantities of terrain have been measured, i.e. the basis of error bars required for meaningful interpretation.

Raw images are complex and difficult to work with directly. To simplify the problem images are translated into a series of binary patterns describing locally repeating structures. Bits within these patterns correspond to pixel pair brightness comparisons similar to those used in the BRIEF representation (Calonder M et al. 2010). Each binary pattern, X, is assigned to a histogram bin and a histogram, H(X), is constructed for each image to be analysed providing statistical distributions of binary pattern repetitions within the data. Any single terrain type, K, is composed of numerous subcomponent elements, k, which generate correlated groups of pixel patterns. These correlated patterns are extracted using an Independent Component Analysis (ICA) training algorithm which we have developed especially for histogram data. This training algorithm estimates a set of probability mass functions (PMFs) giving the probability of observing each pattern within each terrain type. During analysis these PMFs are linearly combined, weighted by an estimate of the quantity of each terrain element found within an incoming histogram. This forms a statistical model of the terrains believed to exist within an image under analysis in terms of previously learned terrain types.

The ICA algorithm (to extract PMFs) and quantity estimation technique (to make measurements) are both based upon Expectation Maximisation (EM) which iteratively updates terrain model parameters to converge upon the maximum likelihood solutions. Each EM iteration weights terrain model parameters by the probability of the corresponding terrain element being the source of the observed data:

$$P_t(k|X) = \frac{P_t(X|k)Q_{t-1}(k)}{\sum_l P_t(X|l)Q_{t-1}(l)} \qquad P_t(X|k) = \frac{P_t(k|X)H(X)}{Q_t(k)} \qquad Q_t(k) = \sum_X P_t(k|X)H(X)$$

where subscript t denotes the current iteration of the algorithm; P(k|X) is the current estimated probability that terrain component k was the source of the observed pattern X; P(X|k) is the PMF of terrain type k, i.e. the probability of X occurring within k; Q(k) is the current estimate of the quantity of k within the data; and H(X) is the histogram describing the distribution of patterns within the incoming image. These formulas are recursively computed from initial random starting points until stable estimates of the parameters are reached. During training the PMFs are computed and averaged over many example histograms taken from user defined terrains. During analysis of new incoming data only quantities are estimated. The converged final quantity estimates give the measurement vector **Q** with an element for each terrain component k.

Stabilities of quantity measurements are estimated using error propagation (Barlow R, 1989). Sources of uncertainty are noise in incoming data, H(X), and noise in trained model components, P(X|k). These terms are computed for one EM iteration at the point of convergence and summed as independent contributions:

$$\mathbf{C}_{ij(data)} = \sum_X \left[\left(\frac{\partial Q(i)}{\partial H(X)} \right) \left(\frac{\partial Q(j)}{\partial H(X)} \right) \sigma^2_{H(X)} \right]$$

$$\mathbf{C}_{ij(model)} = \sum_X \left[\sum_k \left(\frac{\partial Q(i)}{\partial P(X|k)} \right) \left(\frac{\partial Q(j)}{\partial P(X|k)} \right) \sigma^2_{P(X|k)} \right]$$

$$\mathbf{C}_{EM} = \mathbf{C}_{data} + \mathbf{C}_{model}$$

where subscripts i and j indicate covariance elements between quantities Q_i and Q_j; sigma squared with subscript H(X) is the expected variance on incoming histogram bins, which is assumed to be Poisson; and sigma squared with subscript P(X|k) is the propagated error from training histograms used to estimate the model PMFs. The iterative nature of EM produces an error feedback loop from one step to the next so an amplification term is required to scale upwards the EM covariance to produce a final covariance matrix:

$$\mathbf{C} = \mathbf{A}^T\mathbf{C}_{EM}\mathbf{A} \qquad \mathbf{A} = [\mathbf{I} - \nabla\mathbf{Q}]^{-1} \qquad \nabla\mathbf{Q}_{ij} = \frac{\partial\mathbf{Q}_i}{\partial\mathbf{\Delta}_j}$$

where **C** is the covariance estimate for measurement vector **Q**; **A** is an amplification matrix which scales the single EM step covariance; **I** is the identity matrix; and **grad Q** is the Jacobian matrix (first order partial derivatives) of measurement vector **Q** with respect to feedback errors (deltas) on each estimated quantity. This amplification models the iterative feedback as a single linear scaling which is derived from interpreting the error feedback loop as a geometric series. Together, the quantity vector and covariance matrix provides researchers with sufficient information to make quantitative assessments of image content within known confidence intervals, assuming the image formation process is in accordance with the statistical assumptions made in the theory. Additionally, a chi-squared per degree of freedom test can be used to assess the quality of a fitted model, M(X), to data, H(X), to identify those unrepresentative datasets which do not meet statistical assumptions or are poorly approximated by the known modelled terrains:

$$\chi^2_{(N-c)} = \frac{1}{N-c}\sum_X \frac{(H(X)^{0.5} - M(X)^{0.5})^2}{\sigma^2_{H(X)^{0.5}} + \sigma^2_{M(X)^{0.5}}} \qquad M(X) = \sum_k P(X|k)Q(k)$$

where N is the total number of histogram bins; c is the number of model parameters (estimated quantities); the sigma squared terms are the expected variances on data and model respectively; and a square-root transform has been applied to stabilise the assumed Poisson histogram noise (Thacker N et al., 1997).

3. Data

All aspects of our quantity estimation and error theory have been corroborated using Monte-Carlo simulated data. The method has also been applied to 30 Martian HiRise images containing numerous terrain types to assess the applicability of the method to real data. Over repeated randomly sampled trials theoretical errors (via our error theory) and empirically observed errors (compared to known ground truths) were recorded.

4. Results

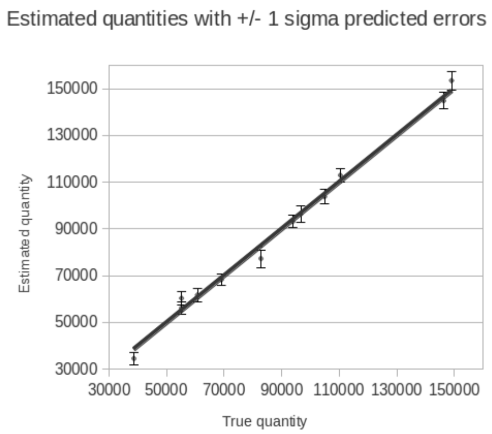

Figure 1a: *Example of Monte-Carlo data quantity estimations. The ground truth quantities are along the x-axis with estimated quantities on the y-axis. 1 standard deviation error bars are overlaid using our error theory to predict accuracy of estimates.*

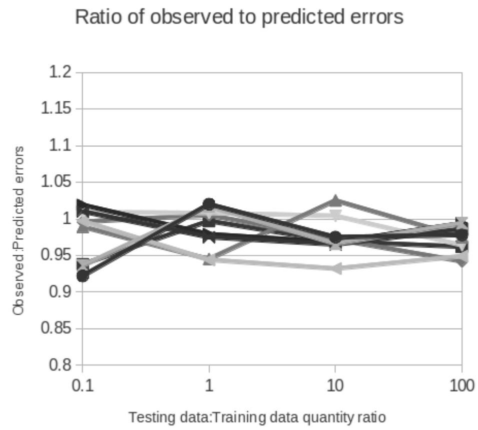

Figure 1b: *Agreement between predicted and observed errors over a range of test data quantities for different Monte-Carlo simulated distributions. Agreement is measured as the ratio of empirically observed errors to theoretically predicted errors.*

Monte-Carlo results are presented in Figure 1 showing quantities of simulated data estimated correctly within the statistically predicted errors provided by our error theory showing the validity of the method. Figure 2a shows the limits of the method when applied to real Martian data confirming the theory is capable of providing predictable quantity and error estimates on some data sets, but not others, but a chi-squared test (Figure 2b) is capable of identifying problematic datasets allowing them to be highlighted to the user.

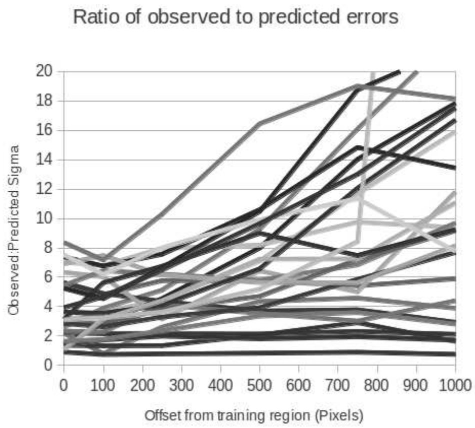

Figure 2a: *Agreement between predicted and observed errors for different Martian terrains. The x-axis indicates how spatially offset the randomly sub-sampled testing regions were from the training regions, i.e. how "unseen" (100% at x=1000).*

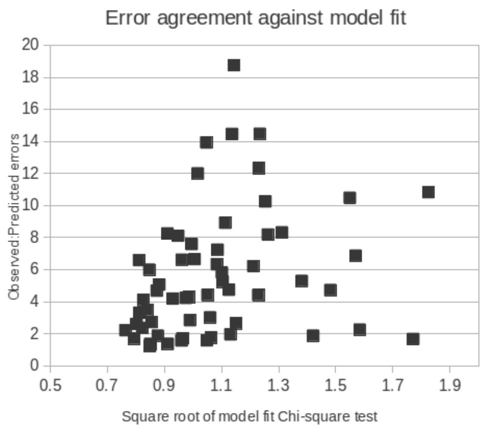

Figure 2b: *A clear trend can be identified between a Chi-squared per degree of freedom and poorly predicted errors. When errors are significantly underestimated the model fit is correspondingly poor. This is a useful data "representativeness" indicator.*

5. Discussion / Analysis

Supervised classification methods including Support Vector Machines and Random Forrest classifiers have been applied by others to terrain analysis tasks with performance measured empirically against known ground truths. Users of such techniques have no method of predicting the performance of their classifiers on unknown data and must make the extremely optimistic assumption that once trained or calibrated systems will only ever encounter data which is strongly representative of the exemplars presented during training. Furthermore, no clear technique to assess the representativeness of incoming data exists when using such methods making it difficult to interpret results in a quantitatively/scientifically meaningful way. In contrast our method provides theoretical predictions of accuracy on a per-dataset basis and chi-squared testing can filter unrepresentative data. Martian results are encouraging, with many errors predicted to within a factor of 10 or better, but results also highlight the difficulties in building representative models for complex data.

6. Conclusions

We have presented a new statistical technique for analysing histogram based data for estimating the composition of histograms, assuming incoming data conforms to a linear combination of Poisson data generating processes. The method's potential has been demonstrated in Monte-Carlo simulation and Martian terrain images. The most important contribution has been to provide an original approach to error analysis allowing pattern recognition to be used, under the correct conditions, as a scientific measuring tool. We believe an improved input representation (definition of patterns X) may improve performance on real data.

7. References

BARLOW R, 1989, Statistics: a guide to the use of statistical methods in the physical sciences (John Wiley and Sons), ISBN 0 471 92294 3, pp. 48-67

CALONDER M ET AL. 2010, BRIEF: Binary Robust Independent Elementary Features, ECCV 2010 conference paper

JOY K ET AL. 2011, Moon Zoo: citizen science in lunar exploration, Astronomy & Geophysics, 52, pp. 2.10-2.12

THACKER N ET AL., 1997, The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data, Kybernetika, 34, 4, pp. 363-368