Paper Number: 71

Output selection for Gaussian Processes classification in iron ore deposits

Silversides, K.L.[1] and Melkumyan, A. [2]

[1]Australian Centre for Field Robotics, University of Sydney, NSW 2006 Australia
katherine.silversides@sydney.edu.au

Modelling stratigraphic ore deposits, such as the banded iron formation hosted iron ore deposits in the Hamersley Ranges of Western Australia, requires detailed knowledge of the location of boundaries. Gaussian Processes (GPs) have been used to identify boundaries in the exploration hole geophysics [1] and chemical assays [2] from these deposits. The boundaries identified include marker shale based ore boundaries and the alluvial to bedded boundary. When using machine learning for feature classification, different techniques may have biases that can affect the results. Gaussian Process predictions are inclined towards the mean value, particularly if there is not sufficient information [3]. To examine the impact that this inclination has on feature identification, a typical Marra Mamba style iron ore deposit was chosen. The alluvial to bedded boundary at the top of the deposit has a signature that can be identified in the TiO_2 and Al_2O_3 chemical assays.

Three GP models were trained using the same starting library, and output ranges of 0 to 1, -0.5 to 0.5 and -1 to 0. In each case the larger number is the output for positive identification of the signature. The GP mean is 0 for all tests. Therefore the outputs were pulled towards the negative, neutral and positive identification respectively. The libraries were progressively updated [1], diverging from each other and producing distinctly different final libraries. The results were compared to a manual geological interpretation (Table 1).

The accuracy and certainty varied between the models, with the 0 to 1 GP having the highest and the -1 to 0 GP having the lowest for both. The 0 to 1 GP also identified the greatest number of correct signatures. The 0 to 1 GP pulled the output towards a negative classification. Therefore the library had to include sufficient positive examples to identify the desired feature, but did not need to include all possible negative examples. This is an advantage due to the very high variability in the negative examples compared to the positive examples, which are variations of a single feature. In comparison, the -0.5 to 0.5 GP pulled all the outputs towards having an uncertain classification. This resulted in a high number of uncertain results as not all possible negative examples could be included in the training library. The -1 to 0 GP pulled the results towards having a positive classification. There were a large number of negative features and including so many dissimilar features in the same category resulted in difficulty training this GP model. Therefore, when using GPs for feature classification the outputs should be arranged so that the pull is against the desired feature. Then the library can be trained to specifically identify that feature as having an output away from the mean.

Table 1: Results for the GP models with different output ranges

-1 to 0
-0.5 to 0.5
0 to 1

No. of holes
Accuracy (%)
No. of holes
Accuracy (%)
No. of holes
Accuracy (%)
Certain holes
All
3649
86.8
4063
89.2
4895
88.7
Signature
2141
80.6
2623
87.2
3098
87.0
No Signature
1508
95.6
1440
92.8
1797
91.7
Uncertain holes
All
2285
66.2
1871
66.2
1039
65.8

Signature
1565
69.8
1082
63.4
591
71.2
No Signature
720
58.5
789
70.1
448
58.7
Overall accuracy
78.9
82.0
84.7
Percent of holes certain
61.5
68.5
82.5

[1] Silversides K et al. (2015) Comput Geosci **77**:118-125
[2] Silversides KL and Melkumyan A (2015) Conf Proc 12th SEGJ
[3] Rasmussen CE and Williams CKI (2006) *Gaussian Processes for Machine Learning*. MIT Press, 248 pp