THE 10TH ANNUAL MLSP COMPETITION: FIRST PLACE

Arno Solin and Simo Särkkä

Aalto University
Department of Biomedical Engineering and Computational Science
Rakentajanaukio 2 C, FI-02150 Espoo, Finland

ABSTRACT

The goal of the MLSP 2014 Schizophrenia Classification Challenge was to automatically diagnose subjects with schizophrenia based on multimodal features derived from their magnetic resonance imaging (MRI) brain scans. This challenge took place between June 5 and July 20, 2014, and was organized on Kaggle. We present how this classification problem can be solved in terms of a Bayesian machine learning paradigm known as Gaussian process (GP) classification. The proposed solution achieved an AUC score of 0.928, and it ranked first on the Kaggle private leaderboard.

***Index Terms*—** Schizophrenia, magnetic resonance imaging, Gaussian process classification

1. INTRODUCTION

Schizophrenia is a chronic mental disorder that is often characterized by abnormal social behavior and abnormal interpretations of reality. Schizophrenia is associated with small differences in brain structure and activity, even though many details remain largely unknown. Yet, state-of-the-art brain imaging techniques provide data that can be used for assisting the diagnosis of schizophrenia.

The goal of the MLSP 2014 Schizophrenia Classification Challenge [1] was to automatically diagnose subjects with schizophrenia using multimodal features derived from their magnetic resonance imaging (MRI) brain scans. The winning proposition was based on a Gaussian process (GP, see, *e.g.*, [2]) classifier. Gaussian processes enable flexible model specification for Bayesian classification, and their theoretical properties are well suited for this kind of modeling.

In binary GP classification, the observations are considered to be drawn from a Bernoulli distribution. The probability is related to the latent field via a sigmoid function that transforms it to a unit interval. A GP prior with a covariance function defined by a sum of a constant, linear, and Matérn kernel was placed over the latent functions. The model was trained by sampling using the GPSTUFF toolbox [3].

In this paper, we briefly present the data, the classification methods and model setup, and provide details on how replicate the winning solution. Finally, the solution is discussed and some future development ideas are given.

2. MATERIALS AND METHODS

2.1. Data and preprocessing

The data consist of two sets of information collected by different medical imaging modalities: *Functional Network Connectivity* (FNC, [4]) and *Source-Based Morphometry* (SBM, [5]) loadings. The FNC were derived form functional magnetic resonance imaging (fMRI) scans, and can be seen as a functional modality feature describing the subject's overall level of 'synchronicity' between brain areas. SBM loadings are derived from structural MRI scans, and they indicate the concentration of grey matter in different regions of the subject's brain.

Data collection (partially described in [6]) was performed at the Mind Research Network, and funded by a Center of Biomedical Research Excellence (COBRE) grant 5P20RR021938/P20GM103472 from the NIH to Dr. Vince Calhoun. Both the training data and test inputs are available on Kaggle [1].

We denote the training data by $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$. The training data consist of $n = 86$ subjects, where $\mathbf{x}_i \in \mathbb{R}^{410}$ (378 from the FNC and 32 from the SBM, ignoring the constant first terms). The test data $\mathcal{D}_* = \{(\mathbf{x}_{*,i}, y_{*,i})\}_{i=1}^{n_*}$ consists of $n_* = 119{,}748$ subjects (artificially inflated to prevent hand labeling) with unknown labels y_*. As a preprocessing step, we normalize each dimension in the inputs $\mathbf{x}_i$ and $\mathbf{x}_{*,i}$ by dividing them by the standard deviations from training inputs. The labels were transformed to $y_i \in \{-1, 1\}$.

2.2. Gaussian process classification

The winning model was based on Gaussian process classification [2], where the latent functions are assumed to be re-

This work was supported by grants from the Academy of Finland (266940, 273475) and the Finnish Funding Agency for Technology and Innovation (40304/09). We acknowledge the computational resources provided by Aalto Science-IT project.

978-1-4799-3694-6/14/$31.00

alizations of a Gaussian process. In binary GP classification with observations, $y_i \in \{-1, 1\}, i = 1, \ldots, n$, associated with inputs $\{\mathbf{x}\}_{i=1}^n$, the observations are considered to be drawn from a Bernoulli distribution with a success probability $p(y_i = 1 \mid \mathbf{x}_i)$. The probability is related to the latent function via a sigmoid function that transforms it to a unit interval. We use a probit transformation that defines the likelihood model

$$p(y_i \mid f(\mathbf{x}_i)) = \Phi(y_i f(\mathbf{x}_i)) = \int_{-\infty}^{y_i f(\mathbf{x}_i)} \mathcal{N}(z \mid 0, 1)\,\mathrm{d}z,$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function. We use a Gaussian process to define a prior distribution over the latent functions

$$f \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')).$$

The latent Gaussian process model is characterized by its covariance function (kernel) $k(\cdot, \cdot)$. We want to account for any linear structure plus some additional short-scale nonlinearities in the latent space. Therefore we set up the covariance function as a linear combination of three separate covariance functions:

$$k(\mathbf{x}, \mathbf{x}') = k_{\text{const.}}(\mathbf{x}, \mathbf{x}') + k_{\text{linear}}(\mathbf{x}, \mathbf{x}') + k_{\text{Matérn}}^{\nu=5/2}(\mathbf{x}, \mathbf{x}'),$$

where the individual covariance functions were defined as (see [2] for a similar parametrization):

$$k_{\text{const.}}(\mathbf{x}, \mathbf{x}') = \theta_1,$$

$$k_{\text{linear}}(\mathbf{x}, \mathbf{x}') = \theta_2\, \mathbf{x}^{\mathsf{T}} \mathbf{x}', \quad \text{and}$$

$$k_{\text{Matérn}}^{\nu=5/2}(\mathbf{x}, \mathbf{x}') = \theta_3 \left(1 + \frac{\sqrt{5}r}{\theta_4} + \frac{5r}{3\theta_4^2}\right) \exp\left(-\frac{\sqrt{5}r}{\theta_4}\right),$$

where $r = \|\mathbf{x} - \mathbf{x}'\|$. The first two covariance function components (a constant and linear covariance function) define an affine model, whereas the last covariance function gives the model flexibility to adopt to some non-linearities. This particular Matérn covariance function holds the assumption of the model functions being continuous and rather smooth (twice differentiable). The Matérn class has previously turned out to be suitable for spatio-temporal GP modeling in fMRI applications (see, *e.g.*, [7–9]).

The hyperparameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ were given the following hyper-priors: $\theta_1, \theta_2, \theta_3 \sim$ Log-Uniform, and $\theta_4 \sim t_4(0, 1)$. The hyperparameters were initialized as $\boldsymbol{\theta} = \{1, 1, 1, 0.01\}$.

The training was started by running a Laplace approximation scheme on the model until convergence (see the codes), and then the final training was performed by sampling (1000 samples, 91 after removing burn-in and thinning). We used *Elliptical Slice Sampling* [10] for the latent functions, and the *Surrogate Slice Sampler* [11] for the hyperparameters. These samplers are the defaults in GPSTUFF [3], and they do not require any parameter tuning. The class label probabilities $p(y_{*,i} = 1 \mid \mathcal{D}, \mathbf{x}_{*,i})$ for the test set can now be predicted by the trained model by integrating over the latent functions. For more information and discussion on the methods, see the toolbox manual [12].

2.3. Implementation

To implement GP classification we used the GPSTUFF toolbox [3] for Mathworks Matlab (and Octave):

- http://becs.aalto.fi/en/research/bayes/gpstuff

It is our in-house-developed software package for Gaussian process modeling. All codes were tested in Matlab 8.2.0.701 (R2013b), and GPSTUFF version 4.5 (release 2014-07-22, available online, and distributed under the GNU General Public License) in Ubuntu Linux.

Codes for replicating the winning submission are available online:

- http://github.com/asolin/MLSP2014-kaggle-challenge

3. DISCUSSION AND CONCLUSIONS

The GP classifier trained following the previously described steps received a final private leaderboard AUC score of 0.92821 on Kaggle and hence winning the competition. The solutions scoring second and third were based on an SVM classifier (scoring 0.923/0.647 on the private/public leaderboards) and Distance Weighted Discrimination (scoring 0.913/0.844), respectively.

This particular GP classifier model was chosen by trying out a couple of models and comparing their performance by leave-one-out cross-validation (LOOCV). This model did show promising performance using LOOCV, but the score (AUC) on the public leaderboard (calculated on approximately 52% of the data) on Kaggle was only 0.70536. This sort of discrepancy is not uncommon in fields of study, where data is scarce, thus the limited size of the test data set did clearly affect the coherence of the public and private leaderboard scores, making it difficult to predict the true performance of the method based on the public score.

The present classifier could be improved in several ways. One evident choice of improvement would be to consider two separate length-scale hyperparameters for the FNC and SBM loadings, rather than normalizing the data. It is also generally well-known [12] that in GP classification MCMC is more accurate than approximative inference methods such as Expectation propagation (EP) or the Laplace approximation. However, the inference times line up in the opposite order. Therefore, for example EP could be a viable option to speed up the inference in this case.

4. REFERENCES

[1] "MLSP 2014 Schizophrenia classification challenge," Online: https://www.kaggle.com/c/mlsp-2014-mri, 2014, Competition chairs Rogers F. Silva and Vince D. Calhoun.

[2] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[3] Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, and Aki Vehtari, "GPstuff: Bayesian modeling with Gaussian processes," *Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1175–1179, 2013.

[4] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, pp. 663–676, 2012.

[5] Judith M. Segall, Elena A. Allen, Rex E. Jung, Erik B. Erhardt, Sunil Kumar Arja, Kent A. Kiehl, and Vince D. Calhoun, "Correspondence between structure and function in the human brain at rest," *Frontiers in Neuroinformatics*, vol. 6, no. 10, 2012.

[6] Mustafa S. Çetin, Fletcher Christensen, Christopher C. Abbott, Julia M. Stephen, Andrew R. Mayer, José M. Cañive, Juan R. Bustillo, Godfrey D. Pearlson, and Vince D. Calhoun, "Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across varying sensory loads in schizophrenia," *NeuroImage*, vol. 97, pp. 117–126, 2014.

[7] Simo Särkkä, Arno Solin, Aapo Nummenmaa, Aki Vehtari, Toni Auranen, Simo Vanni, and Fa-Hsuan Lin, "Dynamical retrospective filtering of physiological noise in BOLD fMRI: DRIFTER," *NeuroImage*, vol. 60, no. 2, pp. 1517–1527, 2012.

[8] Simo Särkkä, Arno Solin, and Jouni Hartikainen, "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, 2013.

[9] Arno Solin and Simo Särkkä, "Infinite-dimensional Bayesian filtering for detection of quasiperiodic phenomena in spatiotemporal data," *Physical Review E*, vol. 88, pp. 052909, 2013.

[10] Iain Murray, Ryan P. Adams, and David Mackay, "Elliptical slice sampling," in *International Conference on Artificial Intelligence and Statistics*, 2010, pp. 541–548.

[11] Iain Murray and Ryan P. Adams, "Slice sampling covariance hyperparameters of latent Gaussian models," in *Advances in Neural Information Processing Systems*, 2010, pp. 1732–1740.

[12] Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, and Aki Vehtari, *Bayesian Modeling with Gaussian Processes Using the GPstuff Toolbox*, 2012, arXiv preprint arXiv:1206.5754.