A Trust-based Mixture of Gaussian Processes Model for Robust Participatory Sensing

(Extended Abstract)

Qikun Xiang and Jie Zhang
Nanyang Technological University
Nanyang Avenue, Singapore
qxiang003@e.ntu.edu.sg

Ido Nevat
TUMCREATE
1 CREATE Way, Singapore
ido.nevat@tum-create.edu.sg

Pengfei Zhang
University of Oxford
Walton Well Rd, Oxford, UK
pengfei@robots.ox.ac.uk

ABSTRACT

Data trustworthiness is a crucial issue in real-world participatory sensing applications. Without considering this issue, different types of worker misbehavior, especially the challenging collusion attacks, can result in biased and inaccurate estimation and decision making. In this paper, we propose a novel trust-based mixture of Gaussian processes (GP) model for spatial regression to jointly detect such misbehavior and accurately estimate the spatial field. We develop a Markov chain Monte Carlo (MCMC)-based algorithm to efficiently perform Bayesian inference of the model. Experiments using real-world dataset show the superior robustness of our model compared with existing approaches.

Keywords

Participatory Sensing, Collusion Attack, Gaussian Process Regression

1. INTRODUCTION

Recently, crowdsourcing has become a fast and inexpensive alternative to outsourcing. One of its notable applications is participatory sensing, in which workers collect sensory information of spatial phenomena (e.g. temperature, noise, air pollutant, etc.) via mobile devices [8]. Through collected data, the field of spatial phenomenon can be estimated at any given point in space via regression. However, trustworthiness of collected data is a crucial issue [4]. Faulty sensors, inappropriate methods of measurement, and malicious attacks (especially the challenging collusion attacks), can result in erroneous or malicious data. Without considering this issue, estimations can become biased and inaccurate.

For this type of problem, several robust regression methods have been developed, such as M-estimation [2], and robust parametric methods that employ heavy-tailed distributions such as Student's t-distribution [3]. These methods minimise effects of outliers, but fail to model complex human-like behaviors, such as collusion. In [7], a trust-based heteroscedastic Gaussian process (TrustHGP) model was proposed. Due to overly simplified assumptions, this model lacks the ability to mitigate real-world malicious attacks. In addition, the method is unable to incorporate past trustworthiness of workers in future tasks.

In this paper, we propose a novel trust-based mixture of Gaussian processes (GP) model to yield accurate estimation of spatial fields in the presence of misbehaving workers and untrustworthy data. The mixture model does not assume specific attack strategies, and is robust against various kinds of attacks, including collusion attacks. Our contributions are as follows: (i) We define attacks in spatial regression settings via a mixture of GP model. (ii) We develop a Bayesian trust framework to maintain and update trustworthiness of participatory sensing workers. Updated trustworthiness improves reliability of future tasks. (iii) We design a novel and efficient Markov chain Monte Carlo (MCMC) sampling-based algorithm for Bayesian inference of the proposed model. (iv) We compare our model with state-of-the-art models using real-world dataset and demonstrate its robustness.

2. TRUST-BASED REGRESSION MODEL

Suppose a participatory sensing system has w potentially dishonest workers, and we collected n data points $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ from them in a task to estimate a spatial field $f : \mathbb{R}^d \to \mathbb{R}$. $\mathbf{x}_i \in \mathbb{R}^d$ is the d-dimensional covariate, and $y_i \in \mathbb{R}$ is the response variable dependent on $\mathbf{x}_i$. A data point could either be truthful or untruthful. Truthful data points are observations from the target field f, while untruthful data points are not related to f. Our objective is to reliably estimate $f(\mathbf{x}_*)$ for any covariate $\mathbf{x}_* \in \mathbb{R}^d$ with its probability distribution.

We assume that untruthful data points form K distinct coalitions. We use $c_i \in \{0, 1, \cdots, K\}$ to denote the truthfulness of the i-th data point, where $c_i = 0$ indicates $(\mathbf{x}_i, y_i)$ is truthful and $c_i = j \in \{1, \cdots, K\}$ indicates that it is untruthful and belongs to the j-th coalition. The probability that a data point is truthful depends on the honesty of its contributor $t_m := \Pr(c_i = 0)$, given that the data point is from the m-th worker. When $c_i = j \in \{1, \cdots, K\}$, we define $\epsilon_j := \Pr(c_i = j | c_i \neq 0)$ to be the prior probability of a data point joining the j-th coalition. We place a beta prior on t_m, with parameters $\mathbf{r}_m := (\alpha_m, \beta_m)$, defined as the trustworthiness of the m-th worker. Every new worker is assigned an initial trustworthiness. After completion of each task, the trustworthiness of workers will be updated, and used in future tasks as priors. Workers who showed dishonest behaviors in the past will be distrusted in future tasks. Truthful data points are generated from the target function f. Untruthful data points from the j-th coalition are generated by an attack strategy function $s_j : \mathbb{R}^d \to \mathbb{R}$.

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

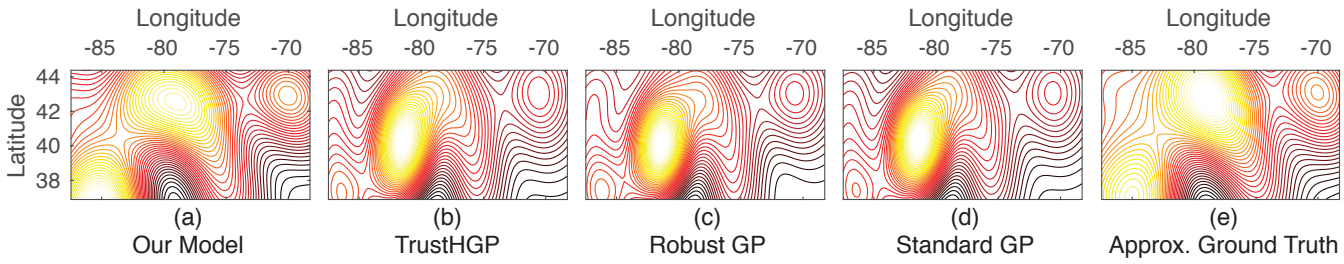

Figure 1: Contour plots of predictive mean resulted from contaminated AQI dataset with (a) the proposed model, (b) TrustHGP, (c) robust GP, (d) standard GP, (e) approximate ground truth of the dataset [**Best Viewed in Color**]

We assume that all observations contain Gaussian noises.

$$(y_i|\mathbf{x}_i, c_i = 0) \sim N(f(\mathbf{x}_i), \sigma_{n0}^2), \quad (1)$$

$$(y_i|\mathbf{x}_i, c_i = j) \sim N(s_j(\mathbf{x}_i), \sigma_{nj}^2),\ j \in \{1, \cdots, K\}. \quad (2)$$

We model f and $\{s_j\}_{j=1}^K$ as Gaussian processes (GP) [5] realizations. Their priors are specified by GP with zero mean,

$$f \sim \mathcal{GP}(0, \mathcal{C}_f; \boldsymbol{\theta}_f), \quad (3)$$

$$s_j \sim \mathcal{GP}(0, \mathcal{C}_{s_j}; \boldsymbol{\theta}_{s_j}),\ \text{for } j \in \{1, \cdots, K\}, \quad (4)$$

where $\mathcal{C}_f$ [1] and $\{\mathcal{C}_{s_j}\}_{j=1}^K$ are covariance functions parameterized by GP hyperparameters $\boldsymbol{\theta}_f$ and $\{\boldsymbol{\theta}_{s_j}\}_{j=1}^K$ accordingly. This creates a mixture of $(K+1)$ GP. In reality, the number of coalitions K is unknown. We estimate the model marginal likelihood and choose the value of K via a Bayesian model determination procedure similar to [1].

Let $\mathbf{L} := (\{c_i\}_{i=1}^n, \boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_j}\}_{j=1}^K, \boldsymbol{\epsilon})$ denote latent variables [2]. Let $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n$, $\mathbf{y} = \{y_i\}_{i=1}^n$. To obtain the posterior predictive distribution of $f_* := f(\mathbf{x}_*)$, we need to marginalize over $\mathbf{L}$, i.e.

$$p(f_*|\mathbf{x}, \mathbf{y}) = \int_{\mathbf{L}} p(f_*|\mathbf{x}, \mathbf{y}, \mathbf{L}) p(\mathbf{L}|\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{L}. \quad (5)$$

The first term within the integral in (5) is a Gaussian density, and the second term is the posterior distribution of $\mathbf{L}$, which is analytically intractable. Hence, we apply Markov chain Monte Carlo (MCMC) to generate samples from $p(\mathbf{L}|\mathbf{x}, \mathbf{y})$ to approximate the integral in (5). Each component of $\mathbf{L}$ can be sampled from its conditional posterior distribution via Gibbs sampling. In particular, $\{c_i\}_{i=1}^n$ is sampled component-wise by Gibbs sampling, and all GP hyperparameters $(\boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_j}\}_{j=1}^K)$ are sampled via Hybrid Monte Carlo (HMC). $\boldsymbol{\epsilon}$ is directly sampled from its posterior, which is a Dirichlet distribution. Additionally, we apply parallel tempering [6] to facilitate convergence of the Markov chain.

Throughout different tasks, we maintain the trustworthiness of workers via $\{\mathbf{r}_m\}_{m=1}^w$. After a task, the updated trustworthiness of the m-th worker is $\mathbf{r}'_m = (\alpha'_m, \beta'_m)$. We first formulate the posterior of t_m from samples obtained from MCMC, denoted by P, and then approximate P by a beta distribution $\mathrm{Beta}(\alpha'_m, \beta'_m)$, denoted by $\widehat{P}$, such that the Kullback-Leibler (KL) divergence is minimised,

$$(\alpha'_m, \beta'_m) = \arg\min_{\alpha'_m, \beta'_m} D_{\mathrm{KL}}(P||\widehat{P}). \quad (6)$$

This is a convex optimization problem and can be computed efficiently via Newton's method.

[1] $\mathcal{C}_f$ may be defined to be stationary or non-stationary.
[2] Note that $(t_1, \cdots, t_w)$ has been integrated out.

3. EXPERIMENTATION

We demonstrate the robustness of our model by applying it to a real-world dataset and comparing with several state-of-the-art models. The dataset we use is an air quality index (AQI) dataset retrieved from the World Air Quality Index project (http://aqicn.org) on July 13, 2016, which consists of 213 AQI readings from North America. In this dataset, $\{\mathbf{x}_i\}_{i=1}^n$ corresponds to longitudes and latitudes where data were observed, and $\{y_i\}_{i=1}^n$ corresponds to measured PM2.5 AQIs. Since the dataset came from an official source, we assume all readings are truthful. We randomly assign data points to 20 imaginary honest workers, and then contaminate the original dataset by adding 50 untruthful data points from a coalition of 5 imaginary dishonest workers. These untruthful data points contain false AQI readings that are higher than the actual value. The contaminated dataset simulates a collusion attack scenario, and is used to evaluate the robustness of our model against three baseline models.

The three baseline models we compare our model against are the standard GP, robust GP with Student's t-distributed noises [3], and TrustHGP [7]. We give weakly-informative priors to GP hyperparameters of all four models, and use them to perform Bayesian inference on the contaminated dataset. Figure 1(a), 1(b), 1(c), 1(d) show the contour plots of the posterior mean resulted from our proposed model, TrustHGP, robust GP, and standard GP, respectively. Figure 1(e) shows the approximate ground truth of this dataset obtained by applying standard GP regression to the original dataset. Our proposed model shows accurate reconstruction of the spatial field that is close to the approximate ground truth, despite the presence of collusion attacks, while three baseline models have produced distorted results due to the attacks. In addition, the worker trustworthiness is efficiently updated after the inference, which is able to further improve the predictive accuracy of future tasks.

4. CONCLUSIONS

This paper introduced a novel trust-based mixture of Gaussian processes model for spatial field regression in the presence of untruthful data in participatory sensing. Bayesian inference of the model is done via a MCMC-based sampling algorithm. We demonstrated the robustness of the proposed model using real-world dataset, comparing with three baseline models. Our model was shown to be significantly more effective than baseline models against attacks. The trustworthiness of workers is efficiently updated after each task, and is used to improve accuracy of future tasks. For future work, we will investigate the performance of the model when dealing with non-stationary spatial processes.

REFERENCES

[1] N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607, 2008.

[2] P. J. Huber. *Robust statistics*. Springer, 2011.

[3] P. Jylänki, J. Vanhatalo, and A. Vehtari. Robust gaussian process regression with a student-t likelihood. *Journal of Machine Learning Research*, 12:3227–3257, 2011.

[4] H. Mousa, S. B. Mokhtar, O. Hasan, O. Younes, M. Hadhoud, and L. Brunie. Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, 90:49–73, 2015.

[5] C. E. Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[6] R. H. Swendsen and J.-S. Wang. Replica monte carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607, 1986.

[7] M. Venanzi, A. Rogers, and N. R. Jennings. Crowdsourcing spatial phenomena using trust-based heteroskedastic gaussian processes. In *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, 2013.

[8] A. Zenonos, S. Stein, and N. R. Jennings. Coordinating measurements for air pollution monitoring in participatory sensing settings. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 493–501, 2015.