Expert Systems with Applications 39 (2012) 4769–4774

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong

Qi Wu [a,b,*], Rob Law [b,*], Xin Xu [c]

[a] Key Laboratory of Measurement and Control of CSE (School of Automation, Southeast University), Ministry of Education, Nanjing, Jiangsu 210096, China
[b] School of Hotel and Tourism Management, Hong Kong Polytechnic University, Hung Horn, Kowloon, Hong Kong
[c] Institute of Automation, College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China

ARTICLE INFO

Keywords:
Sparse Gaussian process
Support vector machine
Tourism demand forecasting
Kernel machines

ABSTRACT

In recent years, Gaussian process (GP) models have been popularly studied to solve hard machine learning problems. The models are important due to their flexible non-parametric modeling abilities using Mercer kernels and the Bayesian framework for probabilistic inference. In this paper, we propose a sparse GP regression (GPR) model for tourism demand forecasting in Hong Kong. The sparsification procedure of the GPR model not only decreases the computational complexity but also improves the generalization ability. We experiment the proposed model with monthly demand data that are relevant to Hong Kong's tourism industry, and compare the performance of the sparse GPR model with those of various kernel-based models to show its effectiveness. The proposed sparse GPR model shows that its forecasting capability outperforms those of the ARMA model and the two state-of-the-art SVM models.

1. Introduction

The tourism industry in Hong Kong has recently undergone drastic changes in terms of international demand for inbound travel, both in tourism receipts and tourist arrivals. Associated with these drastic changes is the restructuring of inbound tourist market segments, and therefore different needs for tourism products and services. As an example, the changes in tourism demand in Hong Kong are more prominent after the introduction of the Individual Visit Scheme to residents in certain cities in Mainland China (Law, To, & Goh, 2008). In view of the new challenges and opportunities that the Hong Kong tourism industry has been, and will be facing, this research is expected to make a major contribution to tourism demand forecasting in Hong Kong.

To a large extent, accurate forecasts of demand for tourism are crucial for effective planning by all providers of services and products in tourism (Perez & Sampol, 2000). In its broadest classification, tourism demand forecasting divides into qualitative and quantitative streams. A qualitative approach, such as a desk reviewing method or an expert-opinion technique, stresses the qualitative insight, intuition, and non-quantifiable knowledge of a particular tourism event. However, these qualitative relationship-identification methods are slow and tedious, and have been criticized as "artistic in nature" and unable to be generalized (Walle, 1997). Quantitative (formal) tourism demand forecasting approaches, employing either causal or time series models, apply mathematical functions to estimate the quantitative relationships of some phenomena in numerical tourism data (Goh, Law, & Mok, 2008; Law, 2000b; Law, Goh, & Pine, 2004; Mok, 1990; Witt & Witt, 1995). On the basis of past performance, these models can then be used to project some future values. However, Goh and Law (2003) commented that given the complex and interrelated nature of today's tourism demand framework, the inclusion of only traditional economic variables in the econometric models could be insufficient. Similarly, Song and Li (2008) stated that AI-based tourism forecasting models can achieve better results than other methods. Due to the limitation of space, this section mainly covers AI-based modeling techniques.

Gaussian processes (GPs) are non-parametric models where *a priori* Gaussian process is directly defined over function values. The direct use of Gaussian processes as priors over functions was motivated by Neal (1996), as he studied priors over weights for ANNs. A model equivalent to GPs, *kriging*, has since long been used for analysis of spatial data (Cressie, 1993). In a more formal way, in a GP, the function outputs $f(x_i)$ are a collection of random variables indexed by the inputs x_i. Any finite subset of outputs has a joint multivariate Gaussian distribution (Rasmussen, 1996).

GP models are routinely used to solve difficult machine learning problems. A recent overview of GP methods is provided by Rasmussen and Williams (2006). Generally speaking, GPs allow a Bayesian use of kernel machines for learning, with the two key

* Corresponding author at: Key Laboratory of Measurement and Control of CSE (School of Automation, Southeast University), Ministry of Education, Nanjing, Jiangsu 210096, China. Tel.: +86 25 51166581; fax: +86 25 511665260.

E-mail addresses: wuqi7812@163.com, hmwuqi@polyu.edu.hk (Q. Wu).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.09.159

advantages of: (i) GPs provide full probabilistic predictive distributions, including estimations of the uncertainty of the predictions, and (ii) the evidence framework applied to GPs enables the learning of parameters of the kernel.

Treated within a Bayesian framework, very powerful statistical methods can be implemented in GP models which have valid estimations of uncertainties in the predictions, and generic model selection procedures cast as nonlinear optimization problems. However, an important problem with GP methods is that the naive implementation requires computation which grows as $o(n^3)$, where n is the number of training cases. To overcome the computational limitations of GP regression, many authors have recently suggested a variety of sparse approximations (Keerthi & Chu, 2006; Quinonero-Candela & Rasmussen, 2005). These techniques can broadly be divided into two classes of: (1) those that are based on sparse methods, which approximate the full posterior by expressions involving matrices of lower rank $m < n$, and (2) those relying on approximate matrix–vector multiplication conjugate gradient methods. In this paper, class (1) is applied to our tourism demand forecasting tasks.

To our knowledge, sparse GPR models in tourism demand analysis has been entirely unexplored by the academic community. As such, this paper is the first attempt to investigate the issue. Research outcomes are expected to have superb generalization ability compared to other models in tourism demand analysis, and have the advantage of comprehensibility in terms of confidence levels.

The paper is organized as follows. In Section 2 we introduce the AI-based tourism forecasting literature. In Section 3 we describe the proposed method, and in Section 4 we report the results of numerical experiments that demonstrate the effectiveness of our method. The conclusion is arranged in Section 5.

2. Literature review

In recent years, researchers have attempted to develop forecasting models that mimic human nervous systems to tourism demand analyses (Cho, 2003; Law, 2000a; Uysal & Roubi, 1999). This biological resemblance has guided researchers to examine the potential applications of artificial neural networks (ANNs) in pattern classification and modeling. In the context of tourism forecasting, prior studies on single destinations have shown that ANNs are superior to selected forecasting techniques (Cho, 2003; Law, 2000b; Law & Au, 1999). However, the structure of ANNs has to be manually selected and optimized in a trial-and-error way. It is thus difficult to guarantee its generalization performance. Another limitation of ANNs is the problem of over-fitting that frequently occurs after training (Lawrence & Giles, 2000).

Recently, SVMs (Vapnik, 2000) have been the main advance in statistical learning theory for data forecasting and classification (Brahim-Belhouari & Bermak, 2004; Fan, Chen, & Lin, 2005; Guo & Li, 2003; Xu, Law, & Wu, 2007, 2009; Xu & Wang, 2005). The aim of the theory of statistical learning is to study how to select or control the structural complexity of various learning machines (e.g., ANNs and fuzzy sets) so that the generalization ability or precision of forecasting and classification can be guaranteed and optimized. Compared with ANNs, which are based on the traditional empirical risk minimization (ERM) principle, SVMs are a class of statistical learning algorithms derived from the structural risk minimization (SRM) principle, and the generalization ability of SVMs can be guaranteed by selecting the structural complexity automatically. In addition, the training procedure of SVMs is a convex quadratic programming process such that the local minima problem in ANNs can be eliminated (Lin, 2001).

SVM-based models have been shown in many cases to be more accurate than ANNs in regression and prediction problems (Cristianini & Schölkopf, 2002). Tourism researchers should, therefore, refocus their attention to SVM-based modeling techniques as they can better capture the useful information in tourism datasets. Pai and Hong (2005) proposed a preliminary SVM model with genetic algorithms to forecast tourist arrivals. Experimental results showed that the proposed model outperformed the ARIMA approaches. In Chen and Wang's (2007) study, a genetic algorithm was combined with Support Vector Regression (SVR) to model tourist arrivals to China during 1985–2001. Empirical evidence showed that their approach outperformed ANNs and ARIMA models based on the normalized mean square error (NMSE) and mean absolute percentage error (MAPE).

3. Methodology

Tourism demand analysis can be formulated as a regression problem, where the goal is to estimate an unknown continuous-valued function by collecting a finite number of training samples (x_i,y), $(i = 1,2,\ldots,n)$, where d-dimensional input $x \in R^d$ and output $y \in R$. Assuming the statistical model for data generation has the following form:

$$y = r(x) + \delta \tag{1}$$

where $r(x)$ is the unknown target function, and δ is an additive zero mean noise with noise variance σ^2.

To solve the above regression problem, the proposed sparse GPR-based modeling method for tourism demand forecasting includes the following aspects.

3.1. Basic GPR model

A Gaussian process (GP) is a powerful, non-parametric tool for regression in high dimensional spaces. Key advantages of GPs are their ability to provide uncertainty estimations and to learn the noise and smoothness parameters from training data. More GPR information can be found in Rasmussen and Williams (2006). A GP can be considered as a "Gaussian over functions". More precisely, a GP describes a stochastic process in which the random variables, in this case the outputs of the modeled function, are jointly Gaussian distributed. A GP is fully described by its mean and covariance functions.

A Gaussian process regression (GPR) model is a Bayesian framework which assumes a GP prior over functions, i.e. that *a priori* the function values behave according to the following assumptions:

The training set $D = \{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ is assumed to be drawn from the (noisy) process:

$$y_i = f(x_i) + \varepsilon \tag{2}$$

where x_i is an input vector in R^d and y_i is a scalar output in R (extension to multiple outputs is possible). The noise term ε is drawn from $N(0,\sigma^2)$. For the sake of convenience, the inputs are aggregated into a matrix $X = [x_1,x_2,\ldots,x_n]$. The outputs are likewise aggregated, $y = [y_1,y_2,\ldots,y_n]$. The joint distribution over the (noisy) outputs y given inputs X is a zero-mean Gaussian, and has the form:

$$p(y|X) = N(0, K(X,X) + \sigma^2 I) \tag{3}$$

where the covariance matrix $K(X,X)$ is also called the kernel matrix with elements $K_{ij}(x_i,x_j)$. The kernel function $k(x,x')$ is a measure of the distance between input vectors. The covariance function encodes the assumptions about the function that is wished to learn, by defining a notion of similarity between two function values, as a function of the corresponding two inputs. The term $\sigma^2 I$ introduces the Gaussian noise and plays a similar role to that of ε in (2).

For the kernel function, a common choice is the Gaussian, or squared exponential:

$$K_{ij} = k(x_i, x_j) = v^2 \exp\left(-\frac{(x_i - y_i)}{2\lambda^2}\right) \quad (4)$$

where v^2 controls the prior variance, and λ is an isotropic length-scale parameter that controls the rate of decay of the covariance, i.e., it determines how far away x_i must be from x_j for f_i to be unrelated to f_j.

Given the training samples (x_i, y_i) and a set of test points X^*, the objective of GPR is to find the predictive outputs f^* with probabilistic confidence levels. By making use of the Bayesian inference, the joint posterior distribution is

$$p(f, f^*|y) = \frac{p(y|f)p(f, f^*)}{p(y)} \quad (5)$$

The posterior predictive distribution is produced by marginalizing the unwanted training set latent variables

$$p(f^*|y) = \int p(f, f^*|y)df = \int \frac{p(y|f)p(f, f^*)}{p(y)}df \quad (6)$$

The joint prior distribution and the independent likelihood probability are both Gaussian distribution:

$$p(f, f^*) = N\left(0, \begin{bmatrix} K_{ff} & K_{f^*f} \\ K_{ff^*} & K_{f^*f^*} \end{bmatrix}\right) \quad (7)$$

$$P(y|f) = N(f, \sigma^2 I) \quad (8)$$

where f and f^* is a subscript of the variables between which the covariance is computed and I is the identity matrix. Since both factors in the integral (7) are Gaussian, the integral can be evaluated in a closed form to give the Gaussian predictive distribution:

$$p(f^*|y) = N(\mu, \Sigma) \quad (9)$$

where μ is the predicted output and Σ is the variance, which can be used as the uncertain estimations for confidence levels:

$$\mu = K(X^*, X)[K(X, X) + \sigma^2 I]^{-1} y \quad (10)$$

$$\Sigma = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma^2 I]^{-1} K(X, X^*) \quad (11)$$

The parameters of the kernel function, $\theta = [\lambda, v, \sigma_n]$, are called the hyper-parameters of the Gaussian process. These hyper-parameters can be learned by maximizing the log likelihood of the training outputs given the inputs:

$$\theta_{\max} = \arg\max_{\theta}\{\log(p(y|X, \theta))\} \quad (12)$$

where

$$\log(p(y|x)) = -\frac{1}{2} y^T (K(X, X) + \sigma^2 I)^{-1} y - \frac{1}{2}\log\left|K(X, X) + \sigma^2 I\right| - \frac{n}{2}\log 2\pi \quad (13)$$

3.2. Sparse approximation to GPR model for tourism demand forecasting

The problem with the above GPR model is that it requires inversion of a matrix of size $n \times n$ which requires $o(n^3)$ operations, where n is the number of training cases. To reduce the computational complexity of the basic GPR model, many computationally efficient approximation methods for GPR have been proposed. Common to all these approximation schemes is that only a subset of the latent variables is treated exactly, and the remaining variables are given some approximate, but computationally cheaper treatments. In the study of Quinonero-Candela and Rasmussen (2005), a unifying description of sparse approximations for GP regression was provided by modifying the joint prior $p(f^*, f)$ from (6) in ways which will reduce the computational requirements from (8). This general sparse approximation procedure is briefly presented as follows.

Let us first rewrite that prior by introducing an additional set of m latent variables $\mu = [\mu_1, \ldots, \mu_m]$ which we call the inducing variables. These latent variables are values of the Gaussian process (as also f and f^*), corresponding to a set of input locations X_μ, which we call the inducing inputs.

Due to the consistency of Gaussian processes, we know that we can recover $p(f^*, f)$ by simply integrating (marginalizing) out μ from the joint GP prior $p(f^*, f, \mu)$

$$p(f^*, f) = \int p(f, f^*, \mu) = \int p(f, f^*|\mu)p(\mu)d\mu \quad (14)$$

where $p(\mu) = N(0, K_{\mu,\mu})$.

As we shall detail in the following sections, the different computationally efficient algorithms proposed in the literature correspond to different additional assumptions about the two approximate inducing conditionals $q(f|\mu), q(f^*|\mu)$ of the integral in (9).

We approximate the joint prior by assuming that f^* and f are conditionally independent given μ, such that:

$$p(f, f^*) \approx q(f, f^*) = \int q(f * |\mu)q(f|\mu)p(\mu)d\mu \quad (15)$$

$$p(f|\mu) = N(K_{f,\mu}K_{\mu,\mu}^{-1}\mu, K_{ff} - Q_{ff}) \quad (16)$$

$$p(f^*|\mu) = N(K_{f^*,\mu}K_{\mu,\mu}^{-1}\mu, K_{f^*f^*} - Q_{f^*f^*}) \quad (17)$$

where

$$Q_{a,b} \triangleq K_{a,\mu}K_{\mu,\mu}^{-1}\mu K_{\mu,b} \quad (18)$$

Most of the sparse GPR methods, such as those studied in Keerthi and Chu (2006), correspond simply to different approximations to the conditionals in (10) with exact likelihood and inducing prior.

4. Performance evaluation and comparisons

To illustrate the sparse GPR method, tourism demand forecast with multi-factors is studied. In general, demands for international travel can be estimated by the following function f (Goh & Law, 2003; Kim & Uysal, 1998; Lathiras & Siriopoulos, 1998; Lee, Var, & Blaine, 1996; Lim, 1997; Song & Witt, 2000)

$$Q_{ijt} = f(\mathrm{Inc}_{it}, \mathrm{RP}_{ijt}, \mathrm{TC}_{ijt}, \mathrm{FER}_{ijt}, \mathrm{Pop}_{it}, \mathrm{Mkt}_{jt}, \mathrm{QF}_{jt}, \xi_t) \quad (19)$$

where $i = 1, 2, \ldots, I$ (for I origins); I is set for 13 in this research; $j = 1, 2, \ldots, J$ (for J destinations); in this research, J is set for 1, which represents Hong Kong; $t = 1, 2, \ldots, K$ (for K time periods); K is set for 38 for annual data and 456 for monthly data ; Q_{ijt} = origin i's demand for travel to destination j at time t; Inc_{it} = income of origin i at time t; RP_{ijt} = prices in destination j relative to origin i at time t; TC_{ijt} = transportation costs between origin i and destination j at time t; FER_{ijt} = foreign exchange rate, measured as units of destination j's currency per unit of origin i's currency at time t; Pop_{it} = population in origin i at time t; Mkt_{jt} = marketing expenses to promote destination j's tourism industry at time t; QF_{jt} = qualitative factors (including special events, changes in tourism policies, climate, and leisure time) in destination j that are relative to origin at time t; ξ_t = random error at time t.

Depending on data availability, tourism researchers use real or proxy values for the dependent variable and the explanatory variables described in function (19) will be used for model calibration and data analysis.

The study covered the time period 1985–2008 for monthly analysis (due to data availability). Since it is not possible to include all countries/origins in this research, we conducted an initial search and thus determined 13 countries/origins that had generated the most number of inbound tourists to Hong Kong with the most number of tourist arrivals during the time period. During the research process, we have collected secondary data from university libraries, library of the Hong Kong Tourism Board, and other research institutes. Both online and printed versions of data source were used. The data set contains 276 samples representing the recording period of the number of tourist arrivals. We obtain 142 patterns for training and 134 for testing candidate models. Predictive patterns are generated by windowing 10 inputs and one output.

In addition, distinct numerical variables have different dimensions and should be normalized firstly. The following normalization is adopted:

$$\bar{x}_i^d = \frac{x_i^d - \min(x_i^d|_{i=1}^l)}{\max(x_i^d|_{i=1}^l) - \min(x_i^d|_{i=1}^l)}, \quad d = 1, 2, \ldots, n_2 \tag{20}$$

where l is the number of samples, x_i^d and $\bar{x}_i^d$ denote the original value and the normalized value respectively. In fact, all the numerical variables from (1)–(19) are the normalized values although they are not marked by bars.

Obviously, the tourism demand is a multivariable series. The normalized data set with an additive zero mean noise is presented in Fig. 1.

The experiments are conducted on a 1.80 GHz Core(TM)2 CPU personal computer (PC) with 1.0G memory under Microsoft Windows XP professional. Some criteria, such as mean absolute error (**MAE**), mean square error (**MSE**) and mean absolute percentage error (**MAPE**), are adopted to evaluate the forecasting performance of the sparse GPR method. The predictive performance given by sparse GPR method is illuminated in Figs. 2–5. It is obvious that the predictive performance of the sparse GPR has good generalization capability for the tourism data.

Apparently, tourism demand forecasting as an application of time series forecasting is a complex dynamic system, and the demand behavior is affected by many factors. Many of these factors have the random, nonlinear, seasonal, and uncertain characteristics. There is a kind of nonlinear mapping relationship between the influencing factors and demand series. As such, it is difficult to present the relationship by traditional models.

To analyze the forecasting performance of the proposed sparse GPR, some state-of-the-art models such as ARMA, v-SVM (Wu, 2009) and g-SVM (Wu, 2010) are selected to handle the above multidimensional tourism demand series. The indexes **MAE**, **MAPE** and **MSE** are used to evaluate the forecasting capability of the above four models. The results of the experiments indicate the running cost of the model ARMA is the highest among all selected models. The running cost of v-SVM is close to that of g-SVM. Although the forecasting precision of the proposed sparse GPR is better than that of the models v-SVM and g-SVM, the running cost of the proposed sparse GPR is higher than that of the models v-SVM and g-SVM. Table 1 shows the forecasting accuracy of the four models.

Then we can find that ARMA model is not appropriate for the tourism demand series. The comparison results indicate the accuracy indexes **MAE**, **MAPE** and **MSE** as provided by sparse GPR are better than ones of ARMA. Moreover, the forecasting accuracy given by sparse GPR excels the ones by other kernel-based models such as v-SVM and g-SVM. The proposed sparse GPR thus has a good generalization performance and is appropriate to those cases with tourism demand forecasting.

In summary, based on the forecasting method using sparse GPR models, the experiment has shown that reasonable prediction and

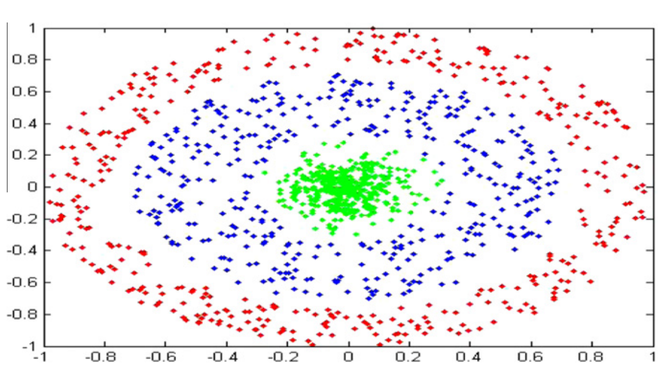

Fig. 1. Training and testing set with an additive zero mean noise.

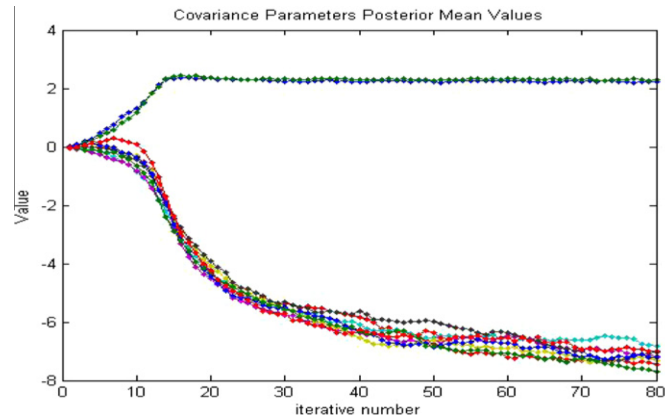

Fig. 2. Covariance parameters posterior mean values.

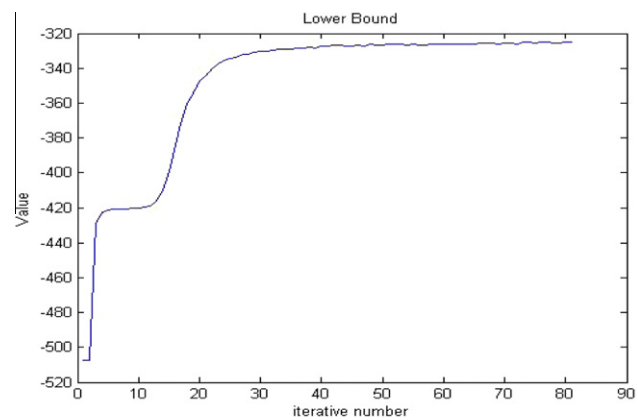

Fig. 3. Lower bound.

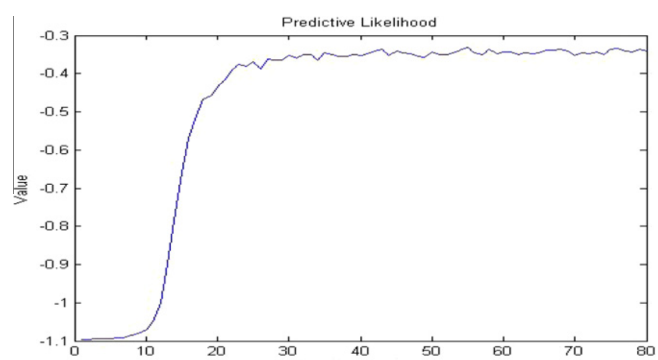

Fig. 4. Predictive likelihood.

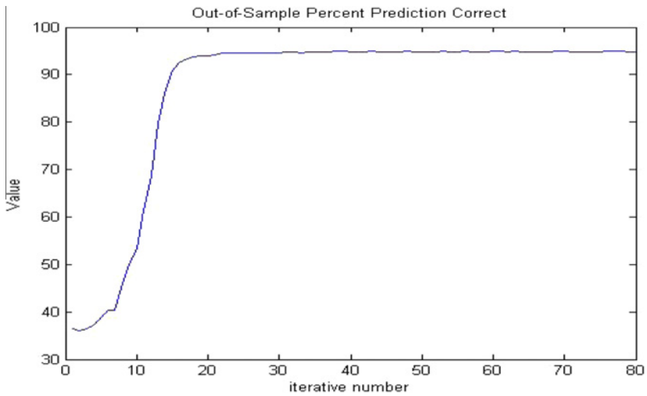

Fig. 5. Out-of-sample percent prediction correct.

Table 1
Forecasting accuracy of the four models.

Model	MAE	MAPE	MSE
ARMA	153.8	43.6	13,657
v-SVM	17.2	5.7	448
g-SVM	11.38	3.78	188
Sparse GPR	1.834	0.64	75.3

tracking performance can be achieved in the case of nonstationary tourism series. Therefore, sparse GPR models are a practical and powerful Bayesian tool for data analysis in the context of tourism demand forecasting.

5. Conclusions

In this paper we proposed the use of the sparse GPR models to keep us from avoiding over-fitting problems and to provide a predictive distribution of tourism demand. Although sparse GPRs are very flexible regression models, they are still limited by the form of the covariance function. For example it is difficult to model non-stationary processes with a sparse GPR model because it is hard to construct sensible non-stationary covariance functions. Although the sparse GPR is not specifically designed to model non-stationarity, the extra flexibility associated with moving active inputs around can actually achieve this to a certain extent. The experiment shows the sparse GPR model fits to some data with an input dependent noise variance. The sparse GPR achieves a much better fit to the data than the standard GP by moving almost all the active input points outside the region of data. It will be interesting to test these capabilities further in the future.

To show the effectiveness of the proposed method compared with other tradition model and kernel-based models, we have conducted simulation with the tourism demand data of Hong Kong. Experimental results show that the forecasting capability of the sparse GPR model outperforms those of ARMA and other kernel-based models in terms of error index, thereby generating consistent results. The extension to classification is also a natural avenue to explore.

Acknowledgements

This research was partly supported by a research grant funded by the Hong Kong Polytechnic University. The work described in this paper was supported by a grant funded by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Number: PolyU 5461/11H, 546111). This research was also partly supported by the National Natural Science Foundation of China under Grant 60904043 and 70761002; the Hong Kong Polytechnic University (G-YX5J) China Postdoctoral Science Foundation (20090451152), Jiangsu Planned Projects for Postdoctoral Research Funds (0901023C) and Southeast University Planned Projects for Postdoctoral Research Funds.

References

Brahim-Belhouari, S., & Bermak, A. (2004). Gaussian process for nonstationary time series prediction. *Computational Statistics and Data Analysis, 47*(4), 705–712.
Chen, K. Y., & Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management, 28*(1), 215–226.
Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management, 24*(3), 323–330.
Cressie, N. (1993). *Statistics for spatial data*. New York: John Wiley & Sons.
Cristianini, N., & Schölkopf, B. (2002). Support vector machines and kernel methods: The new generation of learning machines. *AI Magazine, 23*(3), 31–42.
Fan, R.-E., Chen, P.-H., & Lin, C.-J (2005). *Working set selection using the second order information for training SVM*. Technical report, Department of Computer Science, National Taiwan University.
Goh, C., & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management, 24*(5), 511–517.
Goh, C., Law, R., & Mok, H. M. K. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research, 46*(3), 327–338.
Guo, G. D., & Li, S. Z. (2003). Content-based audio classification and retrieval by support vector machines. *IEEE Transactions on Neural Networks, 14*(1), 209–215.
Keerthi, S. S., & Chu, W. (2006). A matching pursuit approach to sparse GP regression. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.). *Advances in neural information processing systems* (Vol. 18, pp. 643–650). Cambridge, MA: MIT Press. NIPS18.
Kim, Y., & Uysal, M. (1998). Time-dependent analysis for international hotel demand in Seoul. *Tourism Economics, 4*(3), 253–263.
Lathiras, P., & Siriopoulos, C. (1998). The demand for tourism to Greece: A cointegration approach. *Tourism Economics, 4*(2), 171–185.
Law, R. (2000a). Demand for hotel spending by visitors to Hong Kong: A study of various forecasting techniques. *Journal of Hospitality and Leisure Marketing, 6*(4), 17–29.
Law, R. (2000b). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management, 21*(3), 331–340.
Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management, 20*(1), 89–97.
Law, R., Goh, C., & Pine, R. (2004). Modeling tourism demand: A decision rules based approach. *Journal of Travel and Tourism Marketing, 16*(2/3), 61–69.
Law, R., To, T., & Goh, C. (2008). How do mainland Chinese travelers choose restaurants in Hong Kong? An exploratory study of Individual Visit Scheme travelers and packaged travelers. *International Journal of Hospitality Management, 27*(3), 346–354.
Lawrence, S., & Giles, C. (2000). Overfitting and neural networks: Conjugate gradient and back-propagation. In *Proceedings of the IEEE international conference on neural networks* (pp. 114–119).
Lee, C. K., Var, T., & Blaine, T. W. (1996). Determinants of inbound tourist expenditures. *Annals of Tourism Research, 23*(3), 527–542.
Lim, C. (1997). An econometric classification and review of international tourism demand models. *Tourism Economics, 3*(1), 69–81.
Lin, C. J. (2001). Formulations of support vector machines: A note from an optimization point of view. *Neural Computation, 13*(2), 307–317.
Mok, H. M. K. (1990). A quasi-experimental measure of the arc elasticities of destinational advertising: Some evidence from Hawaii. *Journal of Travel Research*, 30–34.
Neal, R. M. (1996). *Bayesian learning for neural networks. Lecture notes in statistics*. Springer.
Pai, P. F., & Hong, W. C. (2005). An improved neural network model in forecasting arrivals. *Annals of Tourism Research, 32*(4), 1138–1141.
Perez, E. A., & Sampol, C. J. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research, 27*(3), 624–637.
Quinonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research, 6*, 1939–1959.
Rasmussen, C. E. (1996). *Evaluation of gaussian processes and other methods for non-linear regression*. PhD thesis, Department of Computer Science, University of Toronto, Toronto, Ontario.
Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, Massachusetts: MIT Press.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management, 29*(2), 203–220.
Song, H., & Witt, S. (2000). *Tourism demand modelling and forecasting: Modern econometric approaches*. Amsterdam: Pergamon.
Uysal, M., & Roubi, M. S. E. (1999). Artificial neural networks versus multiple regression in tourism demand analysis. *Journal of Travel Research, 38*, 111–118.
Vapnik, V. N. (2000). *The nature of statistical learning*. New York: Springer-Verlag.

Walle, A. H. (1997). Quantitative versus qualitative tourism research. *Annals of Tourism Research, 24*(3), 524–536.
Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting, 11*, 447–475.
Wu, Q. (2009). The forecasting model based on wavelet v-support vector machine. *Expert Systems with Applications, 36*(4), 7604–7610.
Wu, Q. (2010). A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization. *Expert Systems with Applications, 37*(3), 2388–2394.
Xu, X., Law, R., & Wu, T. (2007). Classification of business travelers using SVMs combined with kernel principal component analysis. *Lecture Notes in Artificial Intelligence, 4632*, 524–532.
Xu, X., Law, R., & Wu, T. (2009). Support vector machines with manifold learning and probabilistic space projection for tourist expenditure analysis. *International Journal of Computational Intelligence Systems, 2*(1), 17–26.
Xu, X., & Wang, X. N. (2005). An adaptive network intrusion detection method based on PCA and support vector machines. *Lecture Notes in Artificial Intelligence, 3584*, 696–703.