A Comparative Study of Supervised Learning Techniques for the Radiative Transfer Equation Inversion

Esteban Garcia-Cuesta[1] *, Fernando de la Torre[2] and Antonio J. de Castro[1]
[1]Physics Department, Carlos III University Avenida de la Universidad 30, 28911 Leganes (Madrid), Spain
[2]Robotics Institute, Carnegie Mellon University Pittsburgh PA 15213, USA

Abstract— **Estimation of the constituents of a gas (e.g. temperature, concentration) from high resolution spectroscopic measurements is a fundamental step to control and improve the efficiency of combustion processes governed by the Radiative Transfer Equation (RTE). Typically such estimation is performed using thermocouples; however, these sensors are intrusive and must undergo the harsh furnace environment. In this paper, we follow a machine learning approach to learn the relation between the spectroscopic measurements and gas constituents such as temperature, concentration and length. This is a challenging problem due to the non-linear behavior of the RTE and the high dimensional data obtained from sensor measurements. We perform a comparative study of linear and neural network regression models, using canonical correlation analysis (CCA), principal component analysis (PCA), reduced rank regression(RRR), and kernel canonical correlation (KCCA) to reduce the dimensionality.**

Keywords: ***feature extraction, neural networks, dimensionality reduction, component analysis, radiative transfer equation.***

1 Introduction

The regulation of harmful substances is getting tighter in several commercial plants using combustion processes (e.g. gas turbines, boilers, incinerators). The control and retrieval of temperature is an important factor to understand the mechanism of the combustion, and thus minimize environmental disruption and improve the efficiency of combustors [1, 2, 3, 4]. Usually, thermocouples have been used to estimate temperature, however they have several drawbacks: they are intrusive and disturb the measurement, they must undergo the harsh furnace environment, and these measurements are taken at single points.

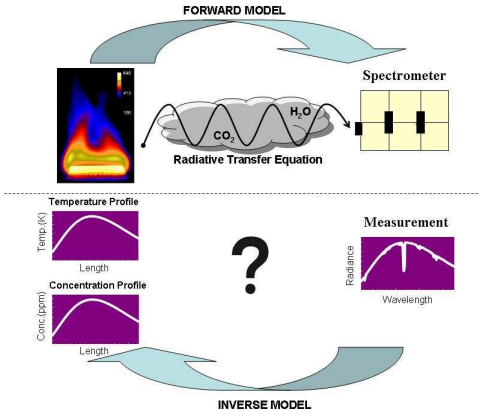

Figure 1: (Top) RTE measurement corresponding to the forward model. (Down) Retrieval of combustion constituents formulated as inverse problem.

Several efforts have been done in the literature to reconstruct the flame temperature. Some studies make use of different spectral regions (ultraviolet, visible, infrared) [1]. The infrared sensing appears to be very promising due to the fact that hot gases in a flame, mainly carbon dioxide CO_2 and H_2O, exhibit important emission bands in it. Others are based on spectroscopic data and use tunable infrared laser and optical fibre in order to measure the difference between the energy emitted and received [2]. This is considered an active technique because it uses an infrared source in addition to the sensor system. This method is high sensitive, but its high complexity and high cost make it not suitable for routine operations in industrial furnaces.

To overcome some of the limitations of previous approaches, we use remote sensing infrared measurements (i.e. spectrometer) to infer gas constituents at a distance. The remote sensing approach is a passive technique and

*E. Garcia-Cuesta is with the Physics Department, Carlos III University, Av. Universidad 30, Leganes 28911 (Madrid) Spain (phone/fax: +34 91-624-9962/9748; e-mail: egc@fis.uc3m.es).
F. de la Torre is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail:ftorre@cs.cmu.edu).
A. J. de Castro is with Physics Department, Carlos III University, Av. Universidad 30 Leganes 28911 (Madrid) Spain (e-mail: decastro@fis.uc3m.es).
Manuscript received August 10th, 2007.

it does not disturb the measurement. Figure 1 illustrates this approach. The spectral characteristics of the flame are recorded with a spectrometer at a given distance. The forward model is described with a set of well known integral equations (see section 2). The interaction of electromagnetic radiation with matter modifies to some extent the incident wave. The spectrometer measures the amount of energy emitted by the target over a range of wavenumbers, and these measurements are related with the different properties of the target.

The challenge that we face in this paper is to invert the forward model. Analytical or numerical inversion of forward model is a challenging mathematical problem. In this paper, we make use of machine learning techniques to learn the relation between the flame energy spectrum and the temperature, concentration, and length.

The rest of the paper is organized as follows: in section 2 we review previous work in the physics of RTE model and related work in machine learning approaches. In section 3, we introduce the inverse problem as a supervised learning problem. Several techniques such as principal component analysis, reduced rank regression, canonical correlation, and kernel canonical correlation are summarized. Section 4 shows the results obtained applying these techniques. Conclusions and future work is given in section 5.

2 Previous work

The RTE explains how the propagation of radiation through a medium is affected by absorption, emission and scattering processes. RTE is common in different fields such as astrophysics, atmosphere, remote sensing, or geophysics, and has been studied in different ways over the last 40 years. In the next subsection, we make a physical review of the RTE and to previous machine learning approaches for dimensionality reduction and regression.

2.1 Physical Model

The most important by-products of combustion are carbon dioxide CO_2 and water vapor H_2O. These gases exhibit important emission bands in the infrared spectral region, and these types of emissions are governed by the RTE [5]. The RTE gives the spectral radiance L_i emitted by an inhomogeneous gas cloud at a given wavenumber i. The expression of the radiance L_i is given by:

$$
\begin{aligned}
L_i &= \int_{z_0}^{z_1} B_i\{T(z), C(z)\} K_i(z) dz \qquad (1)\\
K_i &= \frac{d\tau_i}{dz}\\
\tau_i &= \prod_{\Delta} \prod_{g} \tau_{ig}\\
\tau_{ig} &= exp^{-(\alpha_{ig} \cdot c \cdot z \cdot f_{i\Delta})}
\end{aligned}
$$

where B_i is Planck's law (standard black body emission), τ_i is the transmittance, $T(z)$ is the temperature profile, $C(z)$ is the concentration profile, and K_i is the so-called temperature weighting function that gives the relative contribution of the radiance coming from each region dz. In the transmittance formula, α_{ig} is the absorption coefficient of the gas g, c is its concentration, z is the path length and $f_{i\Delta}$ is a Lorentzian function that takes into account the spectral shape of the transmittance profile and takes values over the spectral range Δ.

During the last decade the progress in optoelectronic technologies has led to the fabrication of new sensors to measure radiance with unprecedented high resolution in spatial, spectral and temporal dimensions. For instance the spectrometer we used has a spectral range of $400 - 4500 cm^{-1}$ and a spectral resolution of $32 cm^{-1} - 0.5 cm^{-1}$. This allows new possibilities but also new challenges.

Work has been done in previous years to invert the RTE. In atmospheric sounding the objective is to retrieve the state of the atmosphere(temperature) and its constituents (e.g. water). In this context there are two main approaches to retrieval. The variational approach uses the forward model to calculate the radiance emitted by a specific atmospheric state. The measured radiance is compared with the estimated radiance and the state vector is adjusted to minimize the error [6]. Due to the fact that the forward problem complexity increases exponentially with the number of wavenumbers, this approach can only be used if either the number of wavenumbers or the number of iterations to converge is small. The second approach is based on the probability density of the pairwise [p(X), P(Y), P(Y, X)] which in practice is difficult to obtain.

2.2 Learning RTE inversion

There have been some approaches that use supervised learning techniques to learn the inverse of RTE [7, 8, 9]. Due to the high resolution of the spectral measurements, one of the challenges of this type of problem is to develop supervised learning algorithms that can efficiently (memory and time) learn from very high dimensional data. To solve the course of dimensionality, a common approach has been to apply dimensionality reduction techniques. There are three main approaches to perform dimensionality reduction: feature extraction (linear or nonlinear),transformation of data, and feature selection (searching for subsets of original variables) [10].

[7, 8] make use of PCA to reduce the dimensionality, and linear and Neural Networks, respectively, to learn a mapping. [11] uses feature selection methods to choose a few wavenumbers range, and show its effectiveness with synthetic data. The feature subset selection approach is interesting, because it allows the reduction of dimensionality and it also retains the semantic interpretation.

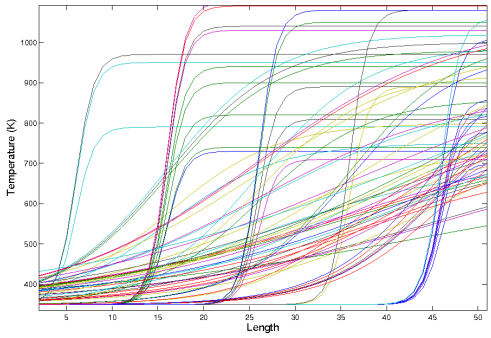

Figure 2: Temperature profiles included in the dataset used in the analysis.

Thus, a better understanding of the combustion properties is reached. Moreover, it allows for the design of specific sensors.

This paper differs from previous work in several aspects: firstly, we make a comparative study of several linear and non-linear dimensionality reduction and regression techniques. Secondly, we propose the use of KCCA and show how it outperforms previous methods. Thirdly, we apply these techniques in a gas retrieval ground remote sensing context where the length is unknown.

3 Supervised Learning for Temperature Retrieval

Equation 1 expresses the amount of energy emitted by each wavenumber i, and it is referred to as the forward problem. Typically, we would like to invert this equation, that is: from the radiance measure L_i provided by the spectrometer, we would like to know the temperature and concentration profiles. Given the measurements of energy represented by $\mathbf{x}_i \in \Re^{p\times 1}$ (see notation[1]), we would like to predict the flame constituents such as spatial temperature, length and concentration, represented by $\mathbf{y}_i \in \Re^{q\times 1}$. We would like to learn a mapping f such that $\mathbf{Y} = f(\mathbf{X})$, where $\mathbf{X} \in \Re^{p\times n}$ represents the energy emitted at different wavenumbers and $\mathbf{Y} \in \Re^{q\times n}$ indicates the temperature profiles and concentration. Recall that each column corresponds to a different observation. For instance, each sample $\mathbf{x}_i$ is a spectrum of radiance in the infrared range of $2110cm^{-1} - 2410cm^{-1}$ with $p = 6000$. Likewise, each sample $\mathbf{y}_i$ is the corresponding temperature, length, and concentration with $q = 153$.

The inversion of Equation 1 is non-linear and ill-posed since we are trying to retrieve $T(z)$, a continuous function, from finite measurements [12]. The non-linearity is due to the dependency of L with C and T. Also, recall that the mapping might not be unique; for instance, there could be more that one temperature/concentration profile that raises a particular radiance measurement.

In this section, a machine learning approach is adopted to learn the relation between the spectroscopist measurement and the gas constituents. Several techniques for a first dimensionality reduction have been compared: PCA, RRC, CCA and KCCA.

3.1 Principal Component Analysis (PCA)

PCA [15] is a standard linear dimensionality reduction technique, and it is optimal for Gaussian distributed classes. It is an energy-preserving transformation that decorrelates the data by projecting it into the first principal components. Following the notation of section 3, let $\mathbf{B} \in \Re^{p\times k}$ be the first k principal components which contains the directions of maximum variation of the data. The k principal components $\mathbf{B}$ maximize $\max_{\mathbf{B}} \sum_{i=1}^{n} ||\mathbf{B}^T\mathbf{x}_i||_2^2 = \max_{\mathbf{B}} ||\mathbf{B}^T\mathbf{\Gamma}\mathbf{B}||_F$ under the constraint $\mathbf{B}^T\mathbf{B} = \mathbf{I}$, where $\mathbf{\Gamma} = \mathbf{X}\mathbf{X}^T = \sum_i \mathbf{x}_i\mathbf{x}_i^T$ is the covariance matrix (zero mean data). The columns of $\mathbf{B}$ form orthonormal bases that spans the principal subspace. If the effective rank of $\mathbf{X}$ is much less than d, we can approximate the column space of $\mathbf{X}$ with $k << p$ principal components. The sample $\mathbf{x}_i$ can be approximated as a linear combination of the principal components as $\mathbf{x}_i \approx \mathbf{B}\mathbf{c}_i$ where $\mathbf{c}_i = \mathbf{B}^T\mathbf{x}_i$. PCA bases are energy preserving and do not necessarily provide a meaningful representation of the signal; that is, it does not correspond to physical quantities. To avoid this problem, [11] proposes to select a subset of features on the original raw data based in the coefficients of the first orthonormal basis.

3.2 Reduced Rank Regression (RRR)

One of the drawbacks of PCA for supervised learning is the lack of dependency between the principal components of $\mathbf{X}$ and $\mathbf{Y}$. That is, a small signal common to both sets (relevant for regression) will be lost if performing independent PCA in each set [16]. PCA reduces the dimensionality optimally in the sense of reconstruction error, but it does not assure a better coupling between the new features $\mathbf{c_i}$ and the data to estimate $\mathbf{y_i}$. That is, one could apply PCA separately to each set and then learn the mapping between them. However, this solution is suboptimal and assures the maximum variance within-set but not between-set.

A more direct approach of finding the direct mapping might be beneficial. A standard approach would be to perform direct regression between the variables $\mathbf{Y}$ and $\mathbf{X}$. For instance, finding the regression matrix $\mathbf{M}$ that

[1]Bold capital letters denote a matrix $\mathbf{D}$, bold lower-case letters a column vector $\mathbf{d}$. $\mathbf{d}_j$ represents the j^{th} column of the matrix $\mathbf{D}$. d_{ij} denotes the scalar in the row i and column j of the matrix $\mathbf{D}$ and the scalar i-th element of a column vector $\mathbf{d}_j$. All non-bold letters represent scalar variables. $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$ designates Euclidean norm of $\mathbf{x}$. $||\mathbf{A}||_F = tr(\mathbf{A}^T\mathbf{A}) = tr(\mathbf{A}\mathbf{A}^T)$ designates the Frobenious norm of a matrix.

minimizes $||\mathbf{Y} - \mathbf{MX}||_F$. The optimal $\mathbf{M}$ is given by $\mathbf{YX}^T(\mathbf{XX}^T)^{-1}$. For very high dimensional $\mathbf{X}$, it is likely that $(\mathbf{XX}^T)^{-1}$ is rank deficient. A common approach to solve this problem is to use reduced-rank regression (RRR). RRR minimizes $||\mathbf{Y}-\mathbf{MX}||_F$ subject to the constraint that $rank(\mathbf{M}) = k$.

A closed form solution for RRR [17] is accomplished by finding the principal directions of the augmented matrix $[\mathbf{X}_\mathbf{Y}]$, and use the QR-decompositions to extract subspaces for the rank-r (truncated). The optimal $\mathbf{M}$ is given by:

$$\mathbf{M} = \mathbf{Q}_Y \mathbf{F} \mathbf{Q}_X^T \tag{2}$$

where $[\mathbf{X}_\mathbf{Y}] = [\mathbf{U}_X{}_{\mathbf{U}_Y}]\mathbf{SV}^T$ is the singular value decomposition over the augmented matrix, $\mathbf{U}_X = \mathbf{Q}_X\mathbf{R}_X$ and $\mathbf{U}_Y = \mathbf{Q}_Y\mathbf{R}_Y$, are the QR-decomposition, and $\mathbf{F} = \mathbf{R}_Y\mathbf{R}_X^{-1}$ maps from $\mathbf{Q_X}$ projections onto $\mathbf{Q_Y}$, or vice versa. See [17] for more details.

3.3 Canonical Correlation Analysis (CCA)

Another common approach used to reduce the dimensionality between two or more datasets that preserve discriminative information is CCA. CCA finds directions of maximum correlation between two datasets. In particular, CCA finds a set of bases vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximized [18]. Consider the linear combinations $x = \mathbf{w}^T\mathbf{x}$ and $y = \mathbf{v}^T\mathbf{y}$ (x and y are called *canonical variates*). CCA finds the direction of $\mathbf{W}$ and $\mathbf{V}$ that maximizes:

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\mathbf{w}^T\mathbf{x}\mathbf{y}^T\mathbf{v}]}{\sqrt{E[\mathbf{w}^T\mathbf{x}\mathbf{x}^T\mathbf{w}]E[\mathbf{v}^T\mathbf{y}\mathbf{y}^T\mathbf{v}]}} = \frac{\mathbf{w}^T\mathbf{C}_{xy}\mathbf{v}}{\sqrt{\mathbf{w}^T\mathbf{C}_{xx}\mathbf{w}\mathbf{v}^T\mathbf{C}_{yy}\mathbf{v}}} \tag{3}$$

where $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ are the within-sets covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ respectively and $\mathbf{C}_{xy} = \mathbf{C}_{yx}^T$ is the between-sets covariance matrix. The pairwise $\mathbf{W}$ and $\mathbf{V}$ can be found solving a generalized eigenproblem, $(\mathbf{C}_{xx})^{-1}\mathbf{C}_{xy}(\mathbf{C}_{yy})^{-1}\mathbf{C}_{yx}\mathbf{W} = \lambda^2\mathbf{W}$ $(\mathbf{C}_{yy})^{-1}\mathbf{C}_{yx}(\mathbf{C}_{xx})^{-1}\mathbf{C}_{xy}\mathbf{V} = \lambda^2\mathbf{V}$ (see [19]).

3.4 Kernel Canonical Correlation Analysis (KCCA)

In the previous section, we have reviewed linear dimensionality reduction methods. However, eq. 1 describes a non-linear physical phenomenon. In this section, we introduce kernel canonical correlation analysis (KCCA), a non-linear extension of CCA. KCCA maps the data to a high dimensional space and performs linear CCA in this space. The mapping is implicitly defined with the kernel, and there is no need to explicitly compute the features in the high dimensional space (kernel trick). Therefore, let $\phi_x : X \rightarrow F_x$ and $\phi_y : Y \rightarrow F_y$ denote feature space mappings corresponding to possibly different kernel functions. The covariances in the feature space are represented by kernel matrices $\mathbf{K}_x = \mathbf{\Phi}_x\mathbf{\Phi}_x^T$ and $\mathbf{K}_y = \mathbf{\Phi}_y\mathbf{\Phi}_y^T$, and the spanned space is $\Im\{\mathbf{\Phi}_x\}$ and $\Im\{\mathbf{\Phi}_y\}$.

Since the canonical vectors $\mathbf{v}_j \in \Im\{\mathbf{\Phi}_x^T\}$ and $\mathbf{w}_j \in \Im\{\mathbf{\Phi}_y^T\}$ lie in the spaces spanned by the feature space mapped, we can represent them as linear combinations $\mathbf{v}_j = \mathbf{\Phi}_x^T\boldsymbol{\alpha}_j$ and $\mathbf{w}_j = \mathbf{\Phi}_y^T\boldsymbol{\beta}_j$ using $\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j \in \Re^n$ as expansion coefficients. Therefore, the canonical variates are $\mathbf{a}_j = \mathbf{\Phi}_x\mathbf{v}_j = \mathbf{K}_x\boldsymbol{\alpha}_j$ and likewise $\mathbf{b}_j = \mathbf{\Phi}_y\mathbf{w}_j = \mathbf{K}_y\boldsymbol{\beta}_j$. As in the linear case, we have to find the canonical vectors in terms of expansion coefficients $\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j \in \Re^n$. The solution can be reduced to an eigenproblem, where the objective is to find the canonical correlations between kernel feature spaces reducing the solution to linear CCA between kernel principal component scores

$$(\mathbf{C}_x^T\mathbf{C}_x)^{-1}\mathbf{C}_x^T\mathbf{C}_y(\mathbf{C}_y^T\mathbf{C}_y)^{-1}\mathbf{C}_y^T\mathbf{C}_x\psi_j = \lambda_j^2\boldsymbol{\psi}_j \tag{4}$$

$$(\mathbf{C}_y^T\mathbf{C}_y)^{-1}\mathbf{C}_y^T\mathbf{C}_x(\mathbf{C}_x^T\mathbf{C}_x)^{-1}\mathbf{C}_x^T\mathbf{C}_y\psi_j = \lambda_j^2\boldsymbol{\xi}_j \tag{5}$$

where $\mathbf{C}_x = \mathbf{\Phi}_x\mathbf{U}_x = \mathbf{K}_x\mathbf{A}_x$ and $\mathbf{U}_x = \mathbf{\Phi}_x^T\mathbf{A}_x$ being $\mathbf{U}_x$ the principal components of $\mathbf{\Phi}_x$, and $\mathbf{a}_j = \mathbf{K}_x\mathbf{A}_x\boldsymbol{\psi}_j$ and $\mathbf{b}_j = \mathbf{K}_y\mathbf{A}_y\boldsymbol{\xi}_j$ are the kernel canonical variates. Likewise, we can obtain $\mathbf{C}_y$ (for a detailed explanation see [20]).

4 Experiments

In this section, we describe the experimental design and the results for synthetic data.

4.1 Experimental Design

The synthetic dataset used in this study has been generated with a simulator called CASIMIR [13], based on the well known experimental database HITRAN [14]. The parameter ranges used to generate this dataset are based on typical combustion environment conditions. The temperature ranges from 296 K to 1100 K with several different profiles (see figure 2), the length range covers from 0.05 to 0.85 meters, and the concentration values for CO_2 and H_2O have been selected as typical ones from combustion of fossil fuels at different temperatures.

4.2 Results

In this section, we report results from applying the techniques described above for dimensionality reduction and regression to the combustion retrieval problem.

Let $\mathbf{Y} = f(\mathbf{C})$ be the mapping between the the projected coefficients of the spectrum $\mathbf{C} \in \Re^{r\times n}$ and the temperature and concentration profiles, represented by $\mathbf{Y}$. $\mathbf{C}$ is obtained by any of the dimensionality reduction techniques explained in section 3 (e.g. PCA, RRR, CCA, KCCA). To make a fair comparison in terms of parameters of the models, we have chosen $r = 80$ for each of the projections. We also fix the multilayer perceptron

Table 1: Mean Absolute Error per sample(MAEs) of temperature and concentration, and its standard deviation (SD).

Melthod	Temperature (MAEs/SD)K	Concentration (MAEs/SD)ppm.
PCA+MLP	2.38/17.33	1.6E-4/1.2E-3
KCCA+MLP	2.27/ 4.65	1.5E-4/3.1E-4
CCA+MLP	8.81/25.18	6.0E-4/1.7E-3
RRC	2.30/ 9.57	1.2E-4/6.0E-4
KCCA+linear	35.80/54.86	2.0E-3/3.8E-3
CCA+linear	32.09/75.84	2.0E-3/5.0E-3

(MLP) architecture to a unique hidden layer with 80 hidden neurons. We split the dataset in 70% training and 30% testing.

Figure 3 shows the temperature profile Mean Square Error (MSE) for the different types of dimensionality reduction and regression techniques. The MSE of temperature profile is computed as $MSE_z = \frac{1}{n}\sum_{j=1}^{n}(||\mathbf{y}_{zj} - \hat{\mathbf{y}}_{zj}||)^2$ being $\hat{\mathbf{y}}$ the estimated value. Recall that linear refers to a standard linear regression. As we can observe in figure 3, KCCA+MLP (solid line) achieves the best results in terms of MSE. This is not surprising, since PCA does not guarantee that the projection $\mathbf{C}$ has correlation with the predicted variable $\mathbf{Y}$. Surprisingly, CCA+MLP achieves a worse performance than PCA+MLP, this can be due to several factors such as local minima in the MLP or not getting the optimal number of components for CCA. However, KCCA achieves better performance for the non-linear component. Table 1 shows the Mean Absolute Error profile per sample (MAEs). The MAEs in temperature and concentration is computed as $MAEs = \frac{1}{z}\frac{1}{n}\sum_{k=1}^{z}\sum_{j=1}^{n}|\mathbf{y}_{kj} - \hat{\mathbf{y}}_{kj}|$ where z is the discretized length, and n the number of samples. The MAEs gives an idea of the physical error. In synthetic data experiments, the MAEs is below 1% relative error (2.27 K.) which is an acceptable level of accuracy for most practical applications [2].

We also compare the MLP technique with linear regression techniques such as RRR and linear mapping. As can be observed in fig. 3, RRR has a slight improvement over PCA+MLP or CCA+MLP. This indicates that the mapping, although non-linear, is not far from being linear. RRR has advantages with respect to the other techniques: firstly, it is extremely efficient (in space and time) to compute; secondly it is not prone to local minima problems and the model for different dimensions can be computed recursively.

5 Conclusions

In this paper, we have presented a comparative study of different dimensionality reduction and regression techniques to invert the radiative transfer equation. KCCA+MLP has outperformed other techniques in terms of MSE and MAEs, which confirms the non-

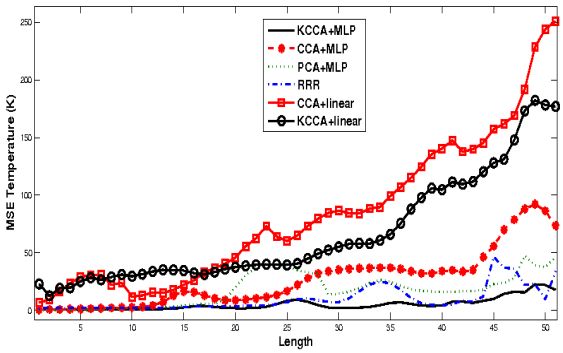

Figure 3: Temperature MSE. X axis is the length discretized in arbitrary units. Best performance is for KCCA+MLP model in solid line.

linearity of the RTE. Not surprisingly, KCCA overcomes the limitations of CCA and PCA to model non-linear structure between the the spectrum values and the temperature profiles. The relative error in temperature achieved by KCCA+MLP is around 1%, which in the literature is considered to be "extremely accurate" [2]. Furthermore, using MLP for regression allows the the modelling of complicated dependencies.

On the other hand, RRR has worked surprisingly well and it is a very computationally efficient method. RRR works especially well when the length is known a priori. This suggests a method of splitting the input data into several lengths and computing a local model for each of them. We are currently investigating extensions of this idea.

Although we have reported promising results in synthetic data, there are still several issues to solve in the future. We plan to gather real data and test the robustness of the algorithm and the validity of the RTE inversion in real environments. Moreover, we have fixed the number of projected components and the number of hidden units in the MLP, to make a fair comparison in terms of parameters of the model. Each model may have its optimal number of projections and hidden units. More research needs to be done to address this problem. On the other hand, the performance of KCCA greatly depends on the choice of the kernels and their parameters and we are currently working on automatic methods to learn the kernels following prior work in computer vision [22].

Acknowledgments

The author E. Garcia-Cuesta wishes to acknowledge the Spanish Ministry of Education and Science for financial support under the project TRA2005-08892-C02-01.

References

[1] Romero, C., Xianchang L., Shahla K., Rodney R., "Spectrometer-based combustion monitoring for flame stoichiometry and temperature control", *Appl. Therm. Eng.* 25 (2005) pp. 659-676.

[2] Lu, G., Yan, Y., Colechin, M., *A digital imaging based multifuncional flame monitoring system*, IEEE T. Instrum. Meas. 53 (2004) 1152-1158.

[3] Deguchi, Y., Noda1, M., Fukuda, Y., Ichinose, Y., Endo, Y., Inada, M., Abe, Y., Iwasaki, S., "Industrial Applications of Temperature and Species Concentration Monitoring Using Laser Diagnostics", *Meas. Sci. Technol.*, V13, , pp. R103–R115, 9/02.

[4] Liu, L.H., Jiang, J., "Inverse radiation problem for reconstruction of temperature profile in axisymmetric free flames", *Journal of Quantitative Spectroscopy & Radiative Transfer*, V70, pp. 207–215, 2001.

[5] Goody, R.M. and Yung, Y.L, *Atmospheric Radiation. Theoretical basis (Chap.2)*", Oxford University Press, New York, 1989.

[6] Rodgers, C.D., "Retrieval of Atmospheric Temperature and Composition from Remote Measurements of Thermal Radiation", *J. Geophys. Res.*, V14 (7), pp. 609-624, 1976.

[7] Aires, F. and Chedin, A. and Scott, N. A. and William B. Rossow, "A Regularized Neural Net Approach for retrieval of Atmospheric and Surface Temperatures with the IASI Instrument", *Journal of Applied Meteorology*, V41, pp. 144-159, 2001.

[8] Blackwell, W.J., "A Neural-Network Technique for Retrieval of Atmospheric Temperature and Moisture Profiles from High Spectral Resolution Sounding Data", *IEEE Trans. Geosci. Remote Sens.*, V43(11), pp. 2535-2546, 2005.

[9] Huang, H.L. and Antonelli, P., "Application of Principal Component Analysis to High-Resolution Infrared Measurement, Compression and Retrieval", *J. Appl. Meteorol.*, V40 (3), pp.365-388, 2001.

[10] Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press, 1999.

[11] Garcia-Cuesta, E., de Castro, A. and Galvan, Ines M., "Spectral High Resolution Feature Selection for Retrieval of Combustion Temperature Profiles", *Lect. Notes Comput. SC* , V4224, pp. 754-762, 2006.

[12] McCornick, N J, "Inverse Radiative Transfer Problems: a review, *Nuclear Science and Engineering*, V112 (3), pp. 185-198, 1992.

[13] Garcia-Cuesta, E., "CASIMIR: Calculos Atmosfericos y Simulacion de la Transmitancia en el Infrarrojo", Technical Project Dept. of Physics University Carlos III, Madrid, 2003.

[14] Rothman, L. S. et al., "The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001", *J. Quant. Spectrosc. Radiat. Transfer*, 2003.

[15] Jollife, I. T., "Principal Component Analysis (2nd Ed.)", *Springer Series in Statistics Springer-Verlag*, New York", 2002.

[16] de la Torre Frade, F. and Black, M. J. "Dynamic Coupled Component Analysis" *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 643-650, June 2001.

[17] Brand, M.E., "Subspace Mappings for Image Sequences", *Statistical Methods in Video Processing*, June 2002.

[18] Shawe-Taylor, J., Cristianini, N., *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[19] Melzer, T., Reiter, M., Beschof, H., "Appearance Models Based on Kernel Canonical Correlation Analysis", *Pattern Recognition*, 36(9):1961-1973, 2003.

[20] Kuss, M., Graepel, T., "The Geometry of Kernel Canonical Correlation Analysis" *Technical Report No.108 May.2003 Available in: http://www.kyb.tuebingen.mpg.de/techreports.html*

[21] Schölkopf, B., A. J. Samola and K.-R. Müller, "Nonlinear component analysis as kernel eigenvalue problem". *Neural Computation*, 10, pp. 1299-1319, 1998.

[22] de la Torre Frade, F. and O. Vinyals, "Learning Kernel Expansions for Image Classification" *IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.