DIMENSIONALITY REDUCTION BY PHONOTACTIC INTERSESSION VARIATION COMPENSATION MODEL FOR IMAGE CLASSIFICATION

Jiří Král
Master Degree Programme (2), FIT BUT
E-mail: xkralj04@stud.fit.vutbr.cz

Supervised by: Michal Hradiš
E-mail: ihradis@fit.vutbr.cz

Abstract: This contribution deals with analysis and testing of algorithms and models for image classification. I ivestigated the suitability of Phonotactic Intersession Variation Compensation model for reduction of random effects in image representation in the context of image classification. In this work it is used for dimensionality reduction. The proposed approach was evaluated on Pascal VOC 2007 data and show supperior performance over standard the mean average precision is 41.2 % over all 20 classes.

Keywords: Machine, Learning, Classification, Image

1 INTRODUCTION

The aim of classification process is to categorize objects into classes. In the field of image processing, classification deals mainly with assignment of semantic classes to images with object detection. In this context, machine learning is used as a tool for creation of classification model or for tuning up the model parameters. The performance of image classification methods is restricted by presence of disruptive elements such as background noise or viewing conditions.

2 PHONOTACTIC INTERSESSION VARIATION COMPENSATION – PIVCO

Model PIVCO was developed for channel adaptation of phonetic data [2]. It adopts the idea of Joint factor analysis (JFA) probabilistic modeling and adapts this technique to multinomial models.

A multinomial probabilistic model is defined by parameters μ which correspond to probabilities with which an event may occur. Probability of some datapoint $\mathbf{z}$ which is defined as count of occurences of events is then in multinomial model [1] defined as:

$$p(\mathbf{z}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{z_k} \tag{1}$$

where K is dimension of data.

The PIVCO model tries to adapt the parameters of multinomial model to every datapoint by searching parameter subspace. This subspace can model different channels or background colors and is represented by loading matrix $\mathbf{U}$.

To prevent problems with numerical underflow in PIVCO, all probabilities of multinomial model μ_i are transfered to logarithms $\log \mu_i$. Weithted row of loading matrix U is then added to these log-probabilities:

$$q_i = \log \mu_i + \mathbf{u_i x} \tag{2}$$

where $\mathbf{x}$ is vector of factors (weights). To follow restriction that probabilities of all events must sum to one, linear normalization of q_i is applied:

$$\omega_i = \frac{q_i}{\sum_{j=1}^{K} q_j} \tag{3}$$

where K is number of probabilities. Using normalization (3) a logaritmical PIVCO-multinomial model can be writen as:

$$\log p(\mathbf{z}|\boldsymbol{\mu}) = \sum_{i=1}^{K} z_i \omega_i, \tag{4}$$

or alternatively rewriten as:

$$\log p(\mathbf{z}|\boldsymbol{\mu}) = \sum_{i=1}^{K} z_i \log \frac{e^{\log \mu_i + u_i \mathbf{x_i}}}{\sum_{j=1}^{K} e^{\log \mu_j + u_j \mathbf{x_i}}}. \tag{5}$$

Matrix $\mathbf{U}$ is trained on whole dataset target class, so it can model variability of all datapoints, but every datapoint has its own vector of factors $\mathbf{x}$ to maximize response of multinomial model. Width of loading matrix $\mathbf{U}$ defines how many variabilities we want to model.

2.1 DIMENSION REDUCTION

Originally, model PIVCO was designed to compensate noise elements in phonnetic data by creating factor vector $\mathbf{x}$ for each data element. Alternatively, PIVCO can be used for preprocessing as data reduction element. In this case, factor vector $\mathbf{x}$ is new data element. Loading matrix U and multinomial model is then trained on whole training dataset across all classes. And with this model is both training and test dataset transformed to a reduced form. After this step standard learning algorithms can be used.

For example, Support vector machines (SVM) are trained faster on data of lower dimension. There is a linear dependency between the ammount of data and time necessary to train. Eg: only one tenth of time is necessary for ten-fold reduced data. Also memory demands are diminished for dimensionally reduced dataset.

3 EXPERIMENTAL RESULTS

Pascal VOC 2007 dataset[1] was used for experiments. For each image in the dataset, a feature vector of 2048 elemets was extracted. Feature extraction method consisted of dense sampling, CSIFT descriptor and hard assignment to bag of visual words [3]. The dimensional reduction with PIVCO was applied to these feature vectors.

Multiple target dimensions were tested and in addition to comparison to baseline data, Principal component analysis (PCA) with same target dimension were applied. Target dimensions were: 10, 20, 50, 100 and 300. Linear SVM classifier with χ^2 kernel was used for baseline data. Standard RBF

[1] http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/index.html

kernel was applied for PCA and PIVCO reduced data. PCA data was normalized by substracting the means.

Table (1) shows mean average precision over 20 classes of Pascal VOC 2007 after dimensionality reduction both for PIVCO and PCA. Its obvious that there is positive corelation between target number of dimensions and mean average precision for PCA reduced data, however for PIVCO, the best results were achieved when reducing 2048 features to 100 target features.

	Mean average precision	
Baseline data	0.461898	
Target dimension	**PIVCO**	**PCA**
10	0.272256	0.216747
20	0.335075	0.268401
50	0.382145	0.308091
100	0.412	0.344118
300	0.406311	0.369327

Table 1: Mean average precisions for PCA and PIVCO

It is obvious that PIVCO dimensionality reduction achieved better results for all target dimensions then PCA did. Results shows that dimensionality reduction to one-twentyth of original dimension with PIVCO was the most succesful. This means that twenty times more data can be used for training and the SVM model will be trained in the same amount of time and memory space. Or whole learning process can be twenty times more faster.

Since there is positive corelation between results of PCA and target dimension, one can assume that PCA would achieve better results for higher target dimensions. What we need in dimensionality reduction is the best (highest) compression ratio as possible. For this high compression ratios is PIVCO better by 4 to 7 % then PCA (1).

4 CONCLUSION

In this paper I have shown that PIVCO model can be used as an preprocessing element for image bag of words representation. Comparison between PIVCO and PCA was presented, and PIVCO – dimensionality reduction proves to be more suitable for reliable reduction of data.

PIVCO achieved better results especially for very small target dimensions where dimension reduction to one fortyth or one hundredth of original dimension was applied.

REFERENCES

[1] Bishop, C. M.: Pattern Recognition and Machine Learning. Springer, 2006, ISBN 0-387-31073-8

[2] Glembek, O.; Matějka, P.; Burget, L.; aj.: Advances in Phonotactic Language Recognition

[3] Hradiš, M.; Řezníček, I.; Bařina, D.; aj.: Semantic Indexing. Brno University of Technology at TRECVid 2010