Hybrid generative-discriminative Nucleus Classification of Renal Cell Carcinoma

Aydın Ulaş[†1⋆], Peter Schüffler[†2], Manuele Bicego[1], Umberto Castellani[1], and Vittorio Murino[1,3]

[1] University of Verona, Department of Computer Science, Verona, Italy
[2] ETH Zürich, Department of Computer Science, Zürich, Switzerland
[3] Istituto Italiano di Tecnologia (IIT), Genova, Italy

Abstract. *In this paper we propose to use advanced classification techniques with shape features for nuclei classification in tissue microarray images of renal cell carcinoma. Our aim is to improve the classification accuracy in distinguishing between healthy and cancerous cells. The approach is inspired by natural language processing. Several features are extracted from the automatically segmented nuclei and quantized to* visual words*, and their co-occurrences are encoded as* visual topics*. To do this, a generative model, the probabilistic Latent Semantic Analysis is learned from quantized shape descriptors (visual words). Finally, we extract from the learned models a generative score, that is used as input for new classifiers, defining a hybrid generative/discriminative classification algorithm. We compare our results with the same classifiers on the feature set to assess the increase in accuracy when we apply pLSA. We demonstrate that the feature space created using pLSA achieves better accuracies than the original feature space.*

Keywords: pLSA, renal cell carcinoma, SVM

1 Introduction

The computer-based detection and analysis of cancer tissues represents a challenging and yet unsolved task for researchers in both Medicine and Computer Science. The complexity of the data, as well as the intensive labor practice needed to obtain them, makes the development of such automatic tools very problematic. In this paper we consider the problem of classifying cancer tissues starting from a tissue microarray (TMA), a technology which enables studies associating molecular changes with clinical endpoints [15]. With this technique, $0.6mm$ tissue cylinders are extracted from primary tumor blocks of hundreds of different patients, and are subsequently embedded into a recipient paraffin block. Such array blocks can then be used for simultaneous analysis of primary tumors on DNA, RNA, and protein level.

Here we focus on the specific case of renal cell carcinoma (RCC): in order to analyse it, the tissue is transferred to an array and stained to make the

⋆ Corresponding author. †Equal contributors

morphology of cells and cell nuclei visible. Current image analysis software for TMAs requires extensive user interaction to properly identify cell populations on the TMA images, to select regions of interest for scoring, to optimize analysis parameters and to organize the resulting raw data. Because of these drawbacks, pathologists typically collect tissue microarray data by manually assigning a composite staining score for each spot. The manual rating and assessment of TMAs under the microscope by pathologists is quite unconsistent due to the high variability of cancerous tissue and the subjective experience of humans, as shown in [11]. Manual scoring also introduces a significant bottleneck that limits the use of tissue microarrays in high-throughput analysis. Therefore, decisions for grading and/or cancer therapy might be inconsistent among pathologists. With this work, we want to contribute to a more generalized and reproducible system that automatically processes TMA images and thus helps pathologists in their daily work. One keypoint in the automatic TMA analysis for renal

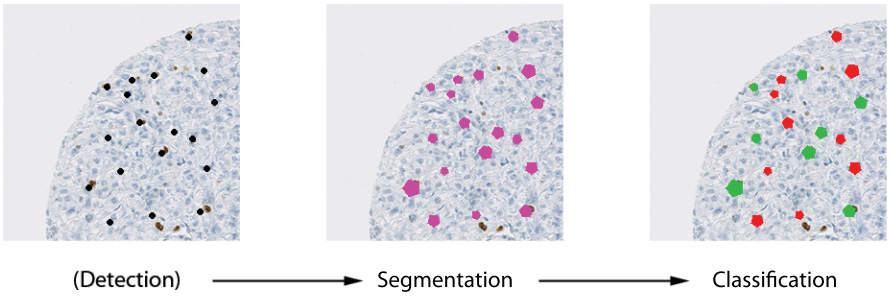

Fig. 1. The nuclei classification pipeline: detection, segmentation and classification into benign or cancerous.

cell carcinoma is the nucleus classification. In this context, the main goal is to automatically classify cell nuclei into cancerous or benign – this typically done by trained pathologists by eye. Clearly, prior to classification, the nucleus should be detected and segmented in the image.

In this paper the problem of the classification of nuclei in renal cancer cells is investigated with the use of hybrid generative-discriminative schemes, representing a quite recent and promising trend of classification methodologies [14, 16]. The underlying idea is to take advantage of the best of the generative and the discriminative paradigms – the former based on probabilistic class models and *a priori* class probabilities, learnt from training data and combined via Bayes law to yield posterior probabilities, the latter aimed at learning class boundaries or posterior class probabilities directly from data, without relying on generative

class models [17, 19]. In the hybrid generative-discriminative scheme, the typical pipeline is to learn a generative model from the data – suitable to properly describe the problem – through which project every object in a feature space (the so-called generative embedding space), where a discriminative classifier may be trained. This class of approaches have been successfully applied in many different scenarions, especially in the case of non-vectorial data (strings, trees, images) [21, 5, 18]. In this paper, their usefulness in the specific and challenging scenario of classification of cell nuclei of renal cell carcinoma is investigated.

In particular, as for the generative model, we choose to employ the probabilistic Latent Semantic Analysis (pLSA – [13]), a powerful methodology introduced in the text understanding community for unsupervised topic discovery in a corpus of documents, and subsequently largely applied in the computer vision community [9, 5] as well as in the medical informatics domain [2, 8, 3]. Referring to the linguistic scenario, where these models have been intially introduced, we can say that the basic idea is that a given document is characterized by the presence of one or more topics (e.g. sport, finance, politics), which may induce the presence of some particular words – the topic distribution may be learned by looking at word co-occurrence in the whole corpus. In our case, similarly to [5, 8], the documents are the cell nuclei images, whereas the words are visual features computed from the image – following the automated pipeline of TMA processing already proposed in [20]. Given a set of images, the visual features are quantized in order to define the so-called dictionary, and histograms of features describe the level of presence of the different visual words in every image. Then the pLSA model is learned to find local patterns of co-occurrences, by leading to the definition of the so called *visual topics*. Finally, the topic distributions of each image represent the new space (the generative embedding space), where any discriminative classifier may be employed.

The proposed approach has been tested in a dataset composed by 474 cell nuclei images, employing different visual features as well as different classifiers in the generative embedding final space. The results were compared to those obtained working directly with the visual features, encouraging us in going ahead along this direction.

The paper is organized as follows: in Section 2 we introduce PLSA and in Section 3, we introduce the data set and preprocessing used in this study. We explain the methods applied in Section 4, and show our experiments in Section 5. We conclude in Section 6.

2 Background: the probabilistic Latent Semantic Analysis

The probabilistic Latent Semantic Analysis (pLSA) [13] is a probabilisitc generative model firstly introduced in the linguistic scenario, to describe and model documents. The basic idea underlying this model – and in general under the class of the so called topic models (another excellent example is the the Latent Dirichlet Allocation LDA [4]) – is that each document is characterized by the presence of one or more topics (e.g. sport, finance, politics), which may induce

the presence of some particular words. From a probabilistic point of view, the document may be seen as a mixture of topics, each one providing a probability distribution over words. A topic model represents a generative model for documents, since a simple probabilistic procedure permits to specify how documents are generated. In particular, a new document may be generated in the following way: first choose a distribution over topics; then, for each word in that document, randomly select a topic according to its distribution, and draw a word from that topic. It is possible to invert the process, in order to infer the set of topics that were responsible for generating a collection of documents. The representation of documents and words with topic models has one clear advantage: each topic is individually interpretable, providing a probability distribution over words that picks out a coherent cluster of correlated terms. This may be really advantageous in the cancer detection context, since the final goal is to provide knowledge about complex systems, and provide possible hidden correlations.

A variety of probabilistic topic models have been used to analyze the content of documents and the meaning of words. These models all use the same fundamental idea – that a document is a mixture of topics – but make slightly different statistical assumptions; here we employed the probabilistic Latent Semantic Analysis, briefly presented in the following. Let us introduce the pLSA model from the original and most intuitive point of view, namely from the linguistic community perspective. As a starting point, the pLSA model takes as input a data set of N documents $\{d_i\}, i=1,...,N$, encoded by a set of words. Before applying pLSA, the data set is summarized by a co-occurrence matrix of size $M \times N$, where the entry $n(w_j, d_i)$ indicates the number of occurrences of the word w_j in the document d_i. The presence of a word w_j in the document d_i is mediated by a latent *topic* variable, $z \in T = \{z_1,..., z_Z\}$, also called *aspect* class, *i.e.*,

$$P(w_j, d_i) = \sum_{k=1}^{Z} P(w_j|z_k)P(z_k|d_i)P(d_i). \tag{1}$$

In practice, the topic z_k is a probabilistic co-occurrence of words encoded by the distribution $P(w|z_k)$, $w = \{w_1,..., w_M\}$, and each document d_i is compactly $(Z < M)$[4] modeled as a probability distribution over the topics, *i.e.*, $P(z|d_i)$, $z = \{z_1,..., z_Z\}$; $P(d_i)$ accounts for varying number of words. The hidden distributions of the model, $P(w|z)$ and $P(z|d)$, are learnt using Expectation-Maximization (EM), maximizing the model data-likelihood L:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{M} P(w_j, d_i)^{n(w_j, d_i)} \tag{2}$$

The E-step computes the posterior over the topics, $P(z|w, d)$, and the M-step updates the parameters, $P(w|z)$ which identifies the model. Once the model has been learnt, the goal of inference is to estimate the topic distribution of a novel

[4] Both Z and M are constants to be a-priori set.

document. To do this, one can use the standard learning algorithm keeping fixed the parameters $P(w|z)$.

The typical classification scheme with pLSA is a standard generative approach, where one has to learn a model per-class and assign a new sample to the category whose model fits the point best, i.e., the model with highest likelihood (see Equation 2). Recently, other approaches successfully used meaningful distributions or other by-products coming from a generative model, as feature for a discriminative classifier. The intuition is that generative models like pLSA are built to understand how samples were generated, and they haven't any notion of discrimination; on the other hand, discriminative classifiers are built to separate the data and they are highly more effective if the data has been previously "explained" by a generative model. In this paper pLSA has been used in such a hybrid generative-discriminative context.

3 The Tissue Microarray (TMA) pipeline

In this section the tissue microarray pipeline is briefly summarized. For a full description please refer to [20]. In particular, first we describe how TMA are determined, followed by the image normalization and patching (how to segment nucleui). Finally, the image features we employed are described.

3.1 Tissue Micro Arrays

Small round tissue spots of cancerous tissue are attached to TMA glass plate. The diameter of the spots is 1mm and the thickness corresponds to one cell layer. Eosin staining made the morphological structure of the cells visible, so that cell nuclei appear bluish in the TMAs. Immunohistochemical staining for the proliferation protein MIB-1 (Ki-67 antigen) makes nuclei in cell division status appear brown.

For computer processing, the TMA slides were scanned with a magnification of 40x, resulting in a per pixel resolution of $0.23\mu m$. The final spots of single patients are separately extracted as three channel color images of size 3000x3000px.

The data set we employed in the evaluation consists of the top left quarter of eight tissue spots from eight patients. Therefore, each image shows a quarter of the whole spot, i.e. 100-200 cells per image (see Figure 2). In order to have a ground truth, the TMA images were independently labeled by two pathologists [11], retaining only those nuclei on which the two pathologists agree on the label.

3.2 Image Normalization and Patching

The images were adjusted in contrast to minimize illumination variances among the scans. To classify the nuclei individually, we extracted patches, of dimension 80x80 pixels, from the whole image, such that each patch has one nucleus in

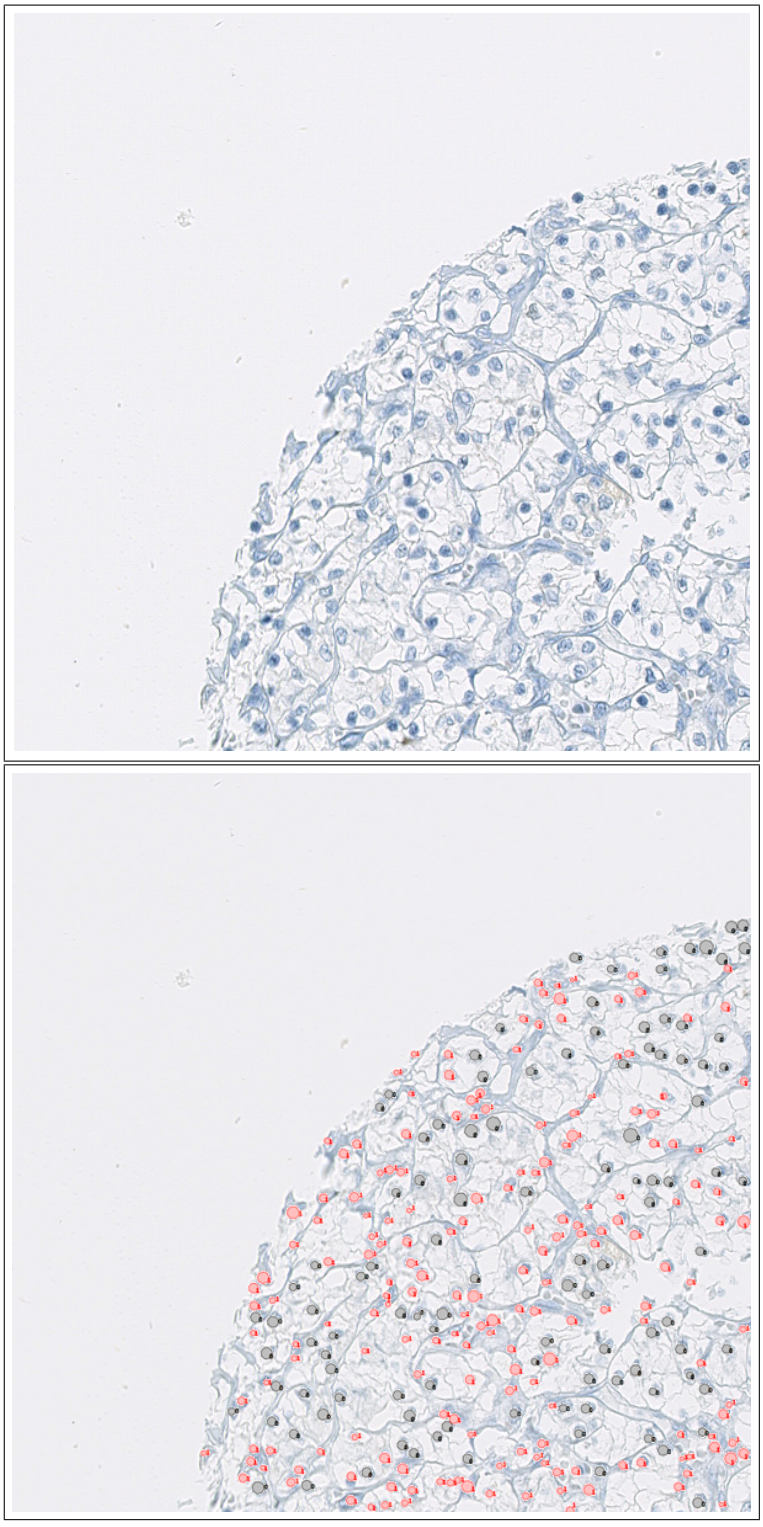

Fig. 2. Top: One 1500x1500px quadrant of a TMA spot from a RCC patient. **Bottom:** A pathologist exhaustively labeled all cell nuclei and classified them into malignant (black) and benign (red).

the center (see Figure 3). The locations of the nuclei were known from the labels of the pathologists. Both procedures drastically improved the following segmentation of cell nuclei.

3.3 Segmentation

The segmentation of cell nuclei was performed with graphcut [7]. The gray intensities were used as unary potentials. The binary potentials were linearly weighted based on their distance to the center to prefer roundish objects lying in the center of the patch (see Figure 3). The contour of the segmented object was used to calculate several shape features as described in the following section.

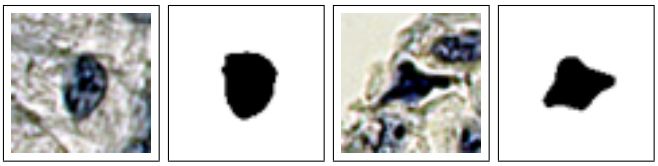

Fig. 3. Two examples of nucleus segmentation. The original 80x80 pixel patch are shown, each with the corresponding nucleus shape found with graphcut.

3.4 Feature extraction

Given the patch image, several features have been extracted, on the basis of several intuitive guidelines used by pathologists to classify nuclei. They are based on pixel intensities as well as on shape descriptors. Such features have been then summarized in histograms, representing the starting point of our algorithm. In [20], histograms have been directly used for classification: we will show in this paper that a significant benefit may be gained when an intermediate generative step is introduced before the final classification. In Table 1) the full list of histograms are described.

4 Nuclei classification

The hybrid generative discriminative approach employed to classify the nuclei can be summarized as follows:

1. **Nucleus image characterization via feature extraction and summarization**: in this step each image is analysed following the pipeline described in the previous section, giving as output features, histograms.

Table 1. Features extracted from patch images for training and testing. All features are histograms.

Shortcut	Feature Description
ALL	**Patch Intensity**: A 16-bin histogram of gray scaled patch
FG	**Foreground Intensity**: A 16-bin histogram of nucleus
BG	**Background Intensity**: A 16-bin histogram of background
LBP	**Local Binary Patterns**: This local feature has been shown to bring considerable performance in face recognition tasks. It benefits from the fact that it is illumination invariant.
COL	**Color feature**: The only feature comprising color information. The colored patch (RGB) is rescaled to size 5x5. The 3x25 channel intensities are then concatenated to feature vector of size 75.
FCC	**Freeman Chain Code**: The FCC describes the nucleus' boundary as a string of numbers from 1 to 8, representing the direction of the boundary line at that point ([12]). The boundary is discretized by subsampling with grid size 2. For rotational invariance, the first difference of the FCC with minimum magnitude is used. The FCC is represented in a 8-bin histogram.
SIG	**1D-signature**: Lines are considered from the object center to each boundary pixel. The angles between these lines form the signature of the shape ([12]). As feature, a 16-bin histogram of the signature is generated.
PHOG	**Pyramid histograms of oriented gradients**: PHOGs are calculated over a level 2 pyramid on the gray-scaled patches ([6]).

2. **Generative model training**: given the training set, the pLSA generative model is trained. In particular, we straightforwardly assume that the visual features previously presented represent the words w_j, while the nuclei are the documents d. With such a point of view, the extracted histograms represent the counting vectors, able to describe how much a visual feature (namely a word) is present in a given image (namely a document). Given the histograms, pLSA is trained following the procedure described in Section 2. Only one model is trained for both classes, disregarding labels. Despite its simplicity – many other schemes may be used to fit the generative model in a classification task [1] – this option yielded promising results.
3. **Generative embedding**: within this step, all the objects involved in the problem (namely training and testing patterns) are projected, through the learned model, to a vector space. In particular, for a given nucleus d, the representation $\phi(d)$ in the generative embedding space is defined as the estimated pLSA posteriors distribution (namely the mixture of topics characterizing the nuclei). In formulae we have that the

$$\phi(d) = [P(z|d)] = [P(z_1|d), \cdots P(z_Z|d)] \tag{3}$$

Our intuition is that the co-occurrence of visual features is different between healthy and cancerous cells. Since the co-occurrences are captured by the topic distribution $P(z|d)$, we are defining a meaningful score for discrimination. This representation with the topic posteriors has been already

successfully used in computer vision tasks [9, 5] as well as in the medical informatics domain [8, 3].

4. **Discriminative classification**: In the resulting generative embedding space any discriminative vector-based classifier may be employed. In this fashion, according to the generative/discriminative classification paradigm, we use the information coming from the generative process as discriminative features of a discriminative classifier.

5 Experiments

In this section the presented approach has been evaluated. In particular we give details about the experimental setup, together with the results and a discussion.

The classification experiments have been carried using a subset of the data presented in [20]. We selected a three patient subset preserving the cancerous/benign cell ratio. In particular, we employed three patients: from the labeled TMA images, we extracted 600 nuclei-patches of size 80x80 pixels. Each patch shows a cell nucleus in the center (see Figure 3). 474 (79 %) from the nuclei form our data set (as said before, we retain only those where the two pathologists agree on the label): 321 (67 %) benign and 153 (33 %) malignant nuclei.

The data of 474 nuclei samples is divided into ten folds (with stratification). We have eight representations (All, Bg, Col, Fcc, Fg, Lbp, Phog, and Sig); for each representation and each fold, we train pLSA on the training set and apply it to the test set. The number of topics has been chosen using leave-another-fold-out (of the nine training folds, we used 9-fold cross validation to estimate the best number of topics) cross validation procedure on the training set. In the obtained space, different classifiers have been tried. The obtained results have been compared with those obtained with the same classifier working on the original histograms (namely without the intermediate generative coding). In particular we employed the following classifiers (where not explicitly reported, all parameters have been tuned via cross validation on the training set)

- (svl): support vector machines with linear kernel (this represents the most widely employed solution with hybrid generative-discriminative approaches).
- (svp): support vector machines with polynomial kernel: after a preliminary evaluation, the degree p was set to 2.
- (svr): support vector machines with radial basis function kernel.
- (ldc): linear discriminant classifier
- (qdc): quadratic discriminant classifier
- (knn): k-nearest neighbor classifier
- (tree): decision tree

All results were computed by using PRTools [10] MATLAB toolbox. They are reported in tables 2 and 3, for the SVM family and for the other classifiers, respectively. The feature representations where the proposed approach overperforms the original space are marked with bold face (statistically significant difference with paired t-test, $p = 0.05$). In particular, results are averaged over ten

runs. In all experiments the standard errors of the mean were inferior to 0.01 for support vector machines and 0.017 for other classifiers.

Table 2. Accuracies with SVM. ORIG is the original histogram based feature space, whereas PLSA stands for the proposed approach.

	svl		*svp*		*svr*	
	ORIG	PLSA	ORIG	PLSA	ORIG	PLSA
All	68.36	**74.26**	65.40	**75.06**	74.47	75.11
Bg	72.88	70.82	66.79	**71.50**	**74.22**	71.92
Col	66.90	69.03	56.93	**70.32**	68.98	68.82
Fcc	67.30	67.72	66.89	67.92	67.95	68.57
Fg	70.68	71.97	64.12	**72.62**	70.49	71.09
Lbp	68.61	69.43	42.36	**70.70**	68.79	70.47
Phog	75.45	**79.67**	63.92	**79.22**	76.55	76.80
Sig	67.72	**68.34**	58.64	**67.69**	67.72	67.72

Table 3. Accuracies using different classifiers. ORIG is the original histogram based feature space, whereas PLSA stands for the proposed approach.

	ldc		*qdc*		*knn*		*tree*	
	ORIG	PLSA	ORIG	PLSA	ORIG	PLSA	ORIG	PLSA
All	71.71	70.21	69.55	69.01	72.35	**73.44**	*71.97	70.30
Bg	**70.79**	68.31	68.48	67.52	**74.25**	71.29	62.25	**67.29**
Col	69.42	69.86	67.55	67.94	69.41	68.62	60.62	62.44
Fcc	66.68	65.25	60.76	**65.19**	66.66	67.71		
Fg	70.24	70.70	68.59	68.78	69.79	70.48	63.07	63.46
Lbp	71.55	71.98	70.71	68.37	71.13	70.29	60.14	63.97
Phog	75.29	*77.57	67.93	***74.62**	70.71	***74.69**	63.51	66.49
Sig	67.73	66.87	64.74	**68.95**	63.50	67.72	58.04	61.85

Observing the Table 2, we can see that the best accuracy using a SVM is 75.45% whereas the best accuracy on the pLSA space is 79.22 %. For most representations (except Lbp, Phog and Col), the accuracies of different kernels on the original space do not have large differences. We also observe that the data set is a difficult data set because there are some classifiers which have accuracy equal to the prior class distribution of the data set (67 per cent). We see that except the support vector machine with rbf kernel, the space constructed by pLSA always supercedes the original space (except Bg on *svl*) in terms of average accuracy. The bold face in the table shows feature sets where pLSA space is more accurate than the original space using 10-fold CV paired t-test at $p = 0.05$.

By looking at the result with other classifiers (Table 3), we can again see that when we transform to the space with pLSA, we get higher accuracies with other classifiers but this time the difference is not strong as in support vector machines. The values with a "*" shows the best classification accuracy using that classifier and again bold face shows feature sets where pLSA space is more accurate than the original space. We can see that, although the number of feature sets where pLSA is better than the original space decreases, except for the decision tree, pLSA space gives the best results for all the classifiers.

5.1 Discussion

We have seen that by using the generative abilities of pLSA and applying the idea of natural language processing to shape features, we can project our data to another space where discriminative classifiers work better. We see that our algorithm automatically finds number of topics and on the space created by pLSA, we have the best results. We observe this behavior with support vector machine variants and also other classifiers.

In this preliminary work, we used a subset of all available subjects to test if the new space created by pLSA has advantages. We have seen that with the new space we have higher accuracy than applying on the original feature space. This promising result encourages us to use more data and apply other kernels to get better classification accuracies. In this work, we use the outputs of pLSA as features in a new space. Another approach would be to directly compute kernels after pLSA and use them for classification. We will explore this option as a future work.

6 Conclusion

In this paper, we propose the use of pLSA to transfer the given shape features into another space to get better classification accuracy for the classification of nuclei in TMA images of renal clear cell carcinoma. Our results show that the features computed by pLSA are more discriminative and achieves higher classification accuracies.

This study extends our previous works by using pLSA to project the data into another space which is more discriminative. We have used the outputs of pLSA as features in a new space but for future work, we plan to compute kernels from the outputs of pLSA and directly use them in kernel based classification. Since the outputs of pLSA are actually probability density functions, we believe that by computing the kernel directly and applying them in a kernel learning paradigm, we may achieve better classification accuracies.

In this work, we used image based feature sets for creating multiple features. In a further application of this scenario, the use of other modalities or other features (e.g. SIFT) extracted from these images, as well as the incorporation of complementary information of different modalities to achieve better classification accuracy is possible.

Acknowledgements

We thank Dr. Cheng Soon Ong very much for helpful discussions. Also, we acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250).

References

1. Bicego, M., Cristani, M., Murino, V., Pekalska, E., Duin, R.P.: Clustering-based construction of hidden markov models for generative kernels. In: Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR '09. pp. 466–479. Berlin, Heidelberg (2009)
2. Bicego, M., Lovato, P., Ferrarini, A., Delledonne, M.: Biclustering of expression microarray data with topic models. In: Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR '10. pp. 2728–2731. Washington, DC, USA (2010)
3. Bicego, M., Lovato, P., Oliboni, B., Perina, A.: Expression microarray classification using topic models. In: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC '10. pp. 1516–1520. New York, NY, USA (2010)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (March 2003)
5. Bosch, A., Zisserman, A., Munoz, X.: Scene classification via plsa. In: Proceedings of the European Conference on Computer Vision, ECCV '06. pp. 517–530 (2006)
6. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval. pp. 401–408. ACM, New York, NY, USA (2007)
7. Boykov, Y., Veksler, O., Zabih, R.: Efficient approximate energy minimization via graph cuts. IEEE transactions on Pattern Analysis and Machine Intelligence 20(12), 1222–1239 (November 2001)
8. Castellani, U., Perina, A., Murino, V., Bellani, M., Rambaldelli, G., Tansella, M., Brambilla, P.: Brain morphometry by probabilistic latent semantic analysis. In: Proceedings of Workshop on Probabilistic Models for Medical Image Analysis, MICCAI '10. pp. 177–184 (2010)
9. Cristani, M., Perina, A., Castellani, U., Murino, V.: Geo-located image analysis using latent representations. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR '08. pp. 1–8 (2008)
10. Duin, R.P.W.: Prtools, a matlab toolbox for pattern recognition version 4.0.14. available at http://www.prtools.org/ (2005), `http://www.prtools.org/`
11. Fuchs, T.J., Wild, P.J., Moch, H., Buhmann, J.M.: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. MICCAI (2008)
12. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital image processing using matlab (2003), 993475
13. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42(1–2), 177–196 (2001)
14. Jaakkola, T.S., Haussler, D.: Exploiting generative models in discriminative classifiers. In: Proceedings of the 1998 conference on Advances in neural information processing systems, NIPS '98. pp. 487–493. Cambridge, MA, USA (1999)

15. Kononen J, Bubendorf L, e.a.: Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med. Jul;4(7), 844–7 (1998)
16. Lasserre, J.A., Bishop, C.M., Minka, T.P.: Principled hybrids of generative and discriminative models. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '06. vol. 1, pp. 87–94. Washington, DC, USA (2006)
17. Ng, A.Y., Jordan, M.I.: On discriminative vs generative classifiers: A comparison of logistic regression and naive Bayes. In: Advances in Neural Information Processing Systems, NIPS 2002. pp. 841–848 (2002)
18. Perina, A., Cristani, M., Castellani, U., Murino, V., Jojic, N.: A hybrid generative/discriminative classification framework based on free-energy terms. In: IEEE 12th International Conference on Computer Vision, ICCV '09. pp. 2058–2065 (29 2009-oct 2 2009)
19. Rubinstein, Y.D., Hastie, T.: Discriminative vs informative learning. In: Proceedings of International Conference on Knowledge Discovery and Data Mining. pp. 49–53. AAAI Press (1997)
20. Schüffler, P.J., Fuchs, T.J., Ong, C.S., Roth, V., Buhmann, J.M.: Computational tma analysis and cell nucleus classification of renal cell carcinoma. In: Proceedings of the 32nd DAGM conference on Pattern recognition, DAGM '10. pp. 202–211. Springer-Verlag, Berlin, Heidelberg (2010)
21. Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S., Müller, K.R.: A new discriminative kernel from probabilistic models. Neural Computation 14, 2397–2414 (October 2002)