

Renal Cancer Cell Classification Using Generative Embeddings and Information Theoretic Kernels

Manuele Bicego¹, Aydın Ulaş¹, Peter Schüffler², Umberto Castellani¹
Vittorio Murino^{1,3}, André Martins^{4,6}, Pedro Aguiar^{5,6}, and Mario Figueiredo^{4,6}

¹ University of Verona, Department of Computer Science, Verona, Italy

² ETH Zürich, Department of Computer Science, Zürich, Switzerland

³ Istituto Italiano di Tecnologia, Genova, Italy

⁴ Instituto de Telecomunicações, Lisboa, Portugal

⁵ Instituto de Sistemas e Robótica, Lisboa, Portugal

⁶ Instituto Superior Técnico, Technical University of Lisbon, Portugal

Abstract. In this paper, we propose a hybrid generative/discriminative classification scheme and apply it to the detection of renal cell carcinoma (RCC) on tissue microarray (TMA) images. In particular we use *probabilistic latent semantic analysis* (pLSA) as a generative model to perform generative embedding onto the *free energy score space* (FESS). Subsequently, we use information theoretic kernels on these embeddings to build a kernel based classifier on the FESS. We compare our results with support vector machines based on standard linear kernels and RBF kernels; and with the nearest neighbor (NN) classifier based on the Mahalanobis distance using a diagonal covariance matrix. We conclude that the proposed hybrid approach achieves higher accuracy, revealing itself as a promising approach for this class of problems.

1 Introduction

The computer-based detection and analysis of cancer tissues represents a challenging, yet unsolved, task for researchers in Medicine, Computer Science and Bioinformatics. The complexity of the data, as well as the intensive labor needed to obtain them, makes the development of such automatic tools very problematic. In this paper, we consider the problem of classifying cancer tissues from tissue microarray (TMA) data, a technology which enables studies associating molecular changes with clinical endpoints [13]. In particular we focus on the specific case of renal cell carcinoma (RCC). One keypoint in the automatic TMA analysis for renal cell carcinoma is the nucleus classification. In this context, the main goal is to automatically classify cell nuclei into cancerous or benign, which is typically done by trained pathologists by visual inspection. Clearly, prior to classification, the nucleus needs to be detected and segmented in the image, as illustrated in Fig. 1.

In this paper, the classification problem described in the previous paragraph is addressed by using hybrid generative-discriminative schemes [12,14]. The underlying idea is to combine the best of the generative and discriminative paradigms

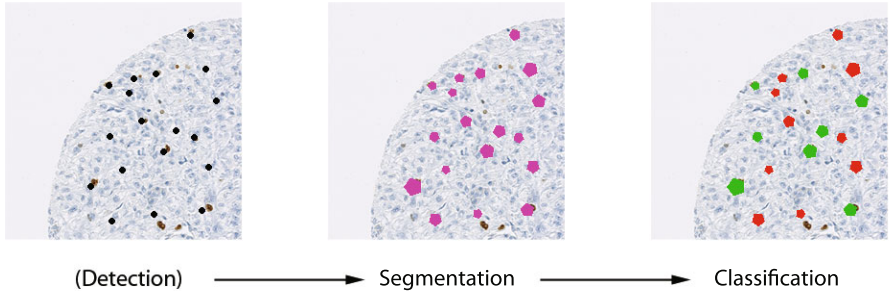


Fig. 1. The nuclei classification pipeline: detection, segmentation and classification into benign or cancerous

– the former being 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 aiming at learning class boundaries or posterior class probabilities directly from data, without relying on probabilistic class models [17,20]. In the hybrid generative-discriminative scheme, the typical pipeline is to learn a generative model – able to properly model the data in hand – from the data, and then use it to project every object onto a feature space (the so-called generative embedding space), where a discriminative classifier may be trained. This class of approaches has been successfully applied in many different scenarios, especially with non-vectorial data (strings, trees, images) [24,5,19].

In particular, concerning the generative model, we adopt the *probabilistic latent semantic analysis* (pLSA) [11], a powerful methodology introduced in the text understanding community for unsupervised topic discovery in a corpus of documents, and subsequently largely applied by the computer vision community [9,5] as well as in medical informatics [1,8,2]. Given the trained generative model, two generative embedding spaces have been considered: the posterior distribution over topics (as in [5,8]); the very recently proposed *free energy score space* (FESS) [19,18]. The latter has been shown to outperform other generative embeddings (including those in [12] and [24]) in several applications [19,18].

Typically, the feature vectors resulting from the generative embedding are used to feed some kernel-based classifier, namely, a *support vector machine* (SVM) with linear or radial basis function (RBF) kernels. In this paper, we follow an alternative route. Instead on relying on standard kernels, we investigate the use of the recently introduced information theoretic (IT) kernels [15]. The rationale behind this choice is that these kernels can exploit the probabilistic nature of the generative embeddings, possibly improving the classification results of hybrid approaches. In particular, we investigate a particular class of IT kernels, based on a non-extensive generalization of the classical Shannon information theory, and defined on normalized (probability) or unnormalized

measures. In [15], these IT kernels were successfully used for text categorization, based on multinomial (bag-of-words type) text representations. Here, the idea is to consider the points of the generative embedding as multinomial distributions, thus valid arguments for the information theoretic kernels.

The proposed approach has been tested on a dataset composed by 474 cell nuclei images, employing different features as well as different IT kernels, in comparison with standard kernels and nearest neighbor classifiers on the original feature space and the score spaces created by the generative embeddings. The results are encouraging, showing that this is a promising research direction.

The remainder of the paper is organized as follows: in Section 2, we explain the tissue micro array pipeline and how the features are extracted; Section 3 introduces our methods, while the experimental results are reported in Section 4. Section 5 concludes the paper.

2 The Tissue Microarray (TMA) Pipeline

In this section, the TMA pipeline is briefly summarized; for more details, please refer to [21]. In particular, we first describe how TMA are obtained, followed by the image normalization and patching (how to segment the nuclei). Finally, the image features that we employed are described.

2.1 Tissue Micro Arrays

A TMA is a microscope slide containing a set of small round tissue spots of (possibly cancerous) tissue, adequate for microscopic histological analysis. The diameter of the spots is of the order of 1mm and the thickness corresponds to one cell layer. Eosin staining is used to make the morphological structure of the cells visible, so that the cell nuclei appear. Immunohistochemical staining for the proliferation protein MIB-1 (Ki-67 antigen) makes the nuclei of cells in division status appear brown.

For subsequent computer analysis, the TMA slides are scanned into three-channel color images at a resolution of $0.23 \mu\text{m}/\text{pixel}$. The spots of single patients are collected into images of size 3000×3000 pixels.

The data set used in this paper consists of the top left quarter of eight tissue spots from eight patients, therefore, each image shows a quarter of a spot, with $100 \sim 200$ cells per image (see Figure 2). In order to have a ground truth, these TMA images were independently labeled by two pathologists [10], retaining only those nuclei on which the two pathologists agree on the label.

2.2 Image Normalization and Patching

The images were adjusted to minimize illumination variations among the scans. To classify the nuclei individually, patches of dimension 80×80 pixels were extracted from the whole image, such that each patch contains one nucleus approximately in 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.

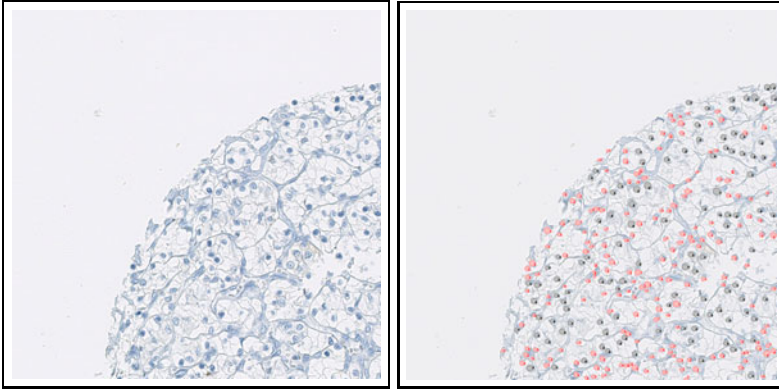


Fig. 2. Left: One 1500×1500 pixel quadrant of a TMA spot from an RCC patient. **Right:** A pathologist exhaustively labeled all cell nuclei and classified them into malignant (black) and benign (red).

2.3 Segmentation

Segmentation of cell nuclei was performed using the *graph cuts* approach [7], with the gray levels used in the unary potentials. The binary potentials were linearly weighted based on their distance to the center, to encourage 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 using the graph cuts algorithm with the potentials described in the text (the size of the patches in 80×80 pixels)

2.4 Feature Extraction

Given the patch image, several features are extracted, inspired by several intuitive guidelines used by pathologists to visually classify the nuclei [21]. In this work, we employed *pyramid histograms of oriented gradients* (PHOG, see [6] for details) – calculated over a 2-level pyramid on gray-scaled patches – which have been previously proven to be the most informative [22].

3 The Proposed Nuclei Classification Scheme

In this section, the proposed hybrid generative-discriminative approach to classify the nuclei is presented. After a brief overview, each step is thoroughly described.

3.1 Overview

Given the characterization of each nucleus by the features described in the previous section, the general scheme may be summarized as follows:

1. **Generative model training:** given the training set, a generative model is trained. In particular we employ the pLSA model, for the reasons explained below.
2. **Generative embedding:** in this step, all the objects involved in the problem (namely training and testing patterns) are embedded, using the learned model, in a vector space. Here we use two types of embedding: the posterior distribution over topics (of pLSA) and the FESS embedding.
3. **Discriminative classification:** in this step, the objects in the generative embedding space are classified. In particular, we consider information theoretic kernels, to be used in SVM and nearest neighbor techniques.

The following subsections describe each of these step in detail.

3.2 Generative Model Training

The generative model adopted is based on pLSA [11], which was introduced in the text understanding community for unsupervised topic discovery in a corpus of documents, and subsequently largely applied by the computer vision community [9,5], as well as in bioinformatics [1,8,2].

The basic idea underlying pLSA – and in general the class of the so-called topic models (of which another well-known example is the *latent Dirichlet allocation* model [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 generative probabilistic point of view, pLSA generates a set of co-occurrences of the form (d, w) , where each of these pairs specifies the presence of a given word w in a document d (as in *bag-of-words* descriptions of documents). The generative model underlying these co-occurrence pairs is as follows: (i) obtain a sample z from the distribution over the topics $P(z)$; (ii) given that topic sample, obtain a word sample from the conditional distribution of words given topics $P(w|z)$; (iii) given that topic sample, obtain a document sample (independently from the word sample) from the conditional distribution of documents given topics $P(d|z)$. The resulting distribution is

$$P(d, z) = \sum_z P(z)P(d|z)P(w|z),$$

where the sum ranges over the set of topics in the model. The parameters of this generative model may be obtained from a dataset using an expectation-maximization (EM) algorithm; for more details, the reader is referred to [11].

In our approach, we simply assume that the visual features previously described are the words in the pLSA model, while the nuclei are the documents. The pLSA model learned from this data can be seen as defining *visual topics*. 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 advantageous in the cancer detection context, since the final goal is to provide knowledge about complex systems, and provide possible hidden correlations.

3.3 Generative Embedding

In this step, all the objects involved in the problem (namely training and testing patterns) are projected, through the learned model, onto a vector space. Different approaches have been proposed in the past, each one with different characteristics, in terms of interpretability, efficacy, efficiency, and others. Here we employ two schemes: the posterior distribution $P(z|d)$ – which was the first generative embedding based on pLSA models that was considered – and the *free energy score space* (FESS) – a novel method whose efficacy has been shown in different contexts [19,18].

In the posterior distribution embedding, a given nucleus (or document) d is represented by the vector of posterior topic probabilities, obtained via the function ψ defined as

$$\psi(d) = [P(z = 1|d), \dots, P(z = T|d)] \in \mathbb{R}^T, \quad (1)$$

where we are assuming that the set of topics is indexed from 1 to T (the total number of topics). The intuition is that the co-occurrence of visual features is different between healthy and cancerous cells and that these co-occurrences are captured by the topic distribution $P(z|d)$, which should thus contain meaningful information for discrimination. This representation with the topic posteriors has been already successfully used in computer vision tasks [9,5] as well as in medical informatics [8,2].

The FESS embedding [19,18] has been shown to outperform other generative embeddings (including those in [12] and [24]) in several applications. This embedding expresses how well each data point fits different parts of the generative model, using the variational free energy as a lower bound on the negative log-likelihood. It has been shown that the FESS embedding yields highly informative for discriminative representations that lead to state-of-the-art results in several computational biology and computer vision problems (namely, scene/object recognition) [19,18]. Due to lack of space, the details of the FESS embedding are not reported here – please refer to [19,18] for a detailed presentation. The only important fact that needs to be pointed out here is that (as the posterior distribution embedding), the components of the FESS embedding of any object are all non-negative.

3.4 Discriminative Classification

In a typical hybrid generative-discriminative classification scenario, the feature vectors resulting from the generative embedding are used to feed some kernel-based classifier, namely, a *support vector machine* (SVM) with simple linear or radial basis function (RBF) kernels. Here, we take a different approach. Instead of relying on standard kernels, we investigate the use of the recently introduced information theoretic (IT) kernels [15] as a similarity measure between objects in the generative embedding space. The main idea is that, with such kernels, we can exploit the probabilistic nature of the generative embeddings, improving even more the classification results of the hybrid approaches – this has been already shown in other classification contexts [3,16].

More in details, given two probability measures p_1 and p_2 , representing two objects, several information theoretic kernels (ITKs) can be defined [15]. The Jensen-Shannon kernel (will be referred to as JS) is defined as

$$k^{\text{JS}}(p_1, p_2) = \ln(2) - \text{JS}(p_1, p_2), \quad (2)$$

with $\text{JS}(p_1, p_2)$ being the Jensen-Shannon divergence

$$\text{JS}(p_1, p_2) = H\left(\frac{p_1 + p_2}{2}\right) - \frac{H(p_1) + H(p_2)}{2}, \quad (3)$$

where $H(p)$ is the usual Shannon entropy.

The Jensen-Tsallis (JT) kernel (will be referred to as JT) is given by

$$k_q^{\text{JT}}(p_1, p_2) = \ln_q(2) - T_q(p_1, p_2), \quad (4)$$

where $\ln_q(x) = (x^{1-q} - 1)/(1 - q)$ is a function called the q -logarithm,

$$T_q(p_1, p_2) = S_q\left(\frac{p_1 + p_2}{2}\right) - \frac{S_q(p_1) + S_q(p_2)}{2^q} \quad (5)$$

is the Jensen-Tsallis q -difference, and $S_q(r)$ is the Jensen-Tsallis entropy, defined, for a multinomial $r = (r_1, \dots, r_L)$, with $r_i \geq 0$ and $\sum_i r_i = 1$, as

$$S_q(r_1, \dots, r_L) = \frac{1}{q-1} \left(1 - \sum_{i=1}^L r_i^q\right).$$

In [15], versions of these kernels applicable to unnormalized measures were also defined as follows. Let $\mu_1 = \omega_1 p_1$ and $\mu_2 = \omega_2 p_2$ be two unnormalized measures, where p_1 and p_2 are the normalized counterparts (probability measures), and ω_1 and ω_2 arbitrary positive real numbers (weights). The weighted versions of the JT kernels are defined as follows:

- The weighted JT kernel (version A, will be referred to as JT-W1) is given by

$$k_q^A(\mu_1, \mu_2) = S_q(\pi) - T_q^\pi(p_1, p_2), \quad (6)$$

where $\pi = (\pi_1, \pi_2) = \left(\frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2}\right)$ and

$$T_q^\pi(p_1, p_2) = S_q(\pi_1 p_1 + \pi_2 p_2) - (\pi_1^q S_q(p_1) + \pi_2^q S_q(p_2)).$$

- The weighted JT kernel (version B, will be referred to as JT-w2) is defined as

$$k_q^B(\mu_1, \mu_2) = (S_q(\pi) - T_q^\pi(p_1, p_2)) (\omega_1 + \omega_2)^q. \quad (7)$$

The approach herein proposed consists in defining a kernel between two observed objects x and x' as the composition of the generative embedding function ψ (the posterior embedding or the FESS embedding) with one of the JT kernels presented above. Formally,

$$k(x, x') = k_q^i(\psi(x), \psi(x')), \quad (8)$$

where $i \in \{\text{JT}, \text{A}, \text{B}\}$ indexes one of the Jensen-Tsallis kernels (4), (6), or (7), and $\psi(x)$ is the generative embedding of object x . Notice that this kernel is well defined because all the components of ψ are non-negative, as is clear from (1) for the posterior probability embedding and was mentioned above for the FESS embedding. Once the kernel is defined, SVM learning can be applied. Recall that positive definiteness is a key condition for the applicability of a kernel in SVM learning. It was shown in [15] that k_q^A is a positive definite kernel for $q \in [0, 1]$, while k_q^B is a positive definite kernel for $q \in [0, 2]$. Standard results from kernel theory [23, Proposition 3.22] guarantee that the kernel k defined in (8) inherits the positive definiteness of k_q^i , thus can be safely used in SVM learning algorithms. Moreover, we also employ nearest neighbor (NN) classifiers, in order to clearly assess the suitability of the derived kernels.

4 Experiments

In this section, we give details about the experimental setup and present the results obtained.

The classification experiments have been carried using a subset of the data presented in [21]. We selected a subset of three patients preserving the cancerous/benign cell ratio. From the labeled TMA images, we extracted 600 nuclei-patches of size 80×80 pixels. Each patch shows a cell nucleus in the center (see Figure 3). In 474 (79 %) of the 600 nuclei, the two pathologists agree on the label, with the following proportions: 321 (67 %) benign and 153 (33 %) malignant; all the experiments are performed on this set of 474 nuclei images, which is divided into ten folds (with stratification). For each fold, we learn a pLSA model from 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. We applied the same partitioning scheme also to choose the q parameter in JT kernels. All reported accuracies are percentual accuracies and are the averages over 10 folds. In all experiments the standard errors around the mean were inferior to 0.02.

4.1 One Model for Both Classes

In this setup, pLSA is trained in an unsupervised way, *i.e.*, we learn the pLSA model ignoring the class labels. Table 1 presents the results using the posterior

distribution (referred to as PLSA) and the FESS embedding with SVM classification; these results show that in the proposed hybrid generative-discriminative approach, the IT kernels outperform linear and RBF kernels. The first and second columns show the classification results of ψ and FESS scores classified using linear and RBF kernels which allows us to show the contribution of the IT kernels.

The results of the NN classifier are shown in Table 2. Although NN is not a good choice for this experiment (baseline NN accuracy using Mahalanobis distance on the original data is 64.57%), we still see the advantage of the IT kernels on the generative approach. We can achieve 72.74% and 72.53% using pLSA and FESS embeddings, respectively, when we use the similarities computed by the IT kernels in the NN classifier.

Table 1. Average accuracies (in percentage) using pLSA and FESS embeddings with SVMs. ORIG shows the baseline accuracies on the original feature space.

	LIN	RBF	JS	JT	JT-W1	JT-W2
PLSA	76.78	76.99	79.31	80.17	74.22	80.17
FESS	77.41	76.17	73.21	78.87	72.31	79.96
ORIG	75.45	76.55			N/A	

4.2 One Model Per Class

In our second experimental setup, we apply pLSA in a supervised manner, *i.e.* the training set is split into the two classes and one pLSA model is trained for each class. The final feature space embedding is formed by the concatenation of the embeddings based on each of the two models. The accuracies obtained with SVM classification using this embeddings are shown in Table 3. Again the first column is the result of classifying the score spaces without the IT kernels. Although the accuracies obtained in this case with the linear kernel are better than those obtained with a single pLSA model, the IT kernels yield a smaller improvement of this linear kernel. We believe that this may be due to curse of dimensionality; when we use pLSA in a supervised way, we concatenate the outputs of each pLSA doubling the number of features. We can still achieve 78.48% accuracy with posterior distribution. With NN classification, comparing Table 4 with Table 2, we see that the accuracies increase except for FESS with JT-W2.

Table 2. Average accuracies (in percentage) using pLSA and FESS embeddings with NN classifiers. ORIG shows the baseline accuracies on the original feature space with Mahalanobis distances.

	MB	JS	JT	JT-W1	JT-W2
PLSA	66.41	68.97	72.53	72.74	68.75
FESS	67.11	67.08	72.53	71.27	71.08
ORIG	64.57				N/A

Table 3. Average accuracies (in percentage) using pLSA and FESS embeddings with SVMs in the supervised pLSA learning setup

	LIN	RBF	JS	JT	JT-W1	JT-W2
PLSA	78.20	74.47	78.48	78.48	74.47	78.26
FESS	78.38	73.20	79.73	79.73	73.41	79.73
ORIG	75.45	76.55			N/A	

Table 4. Average accuracies (in percentage) using pLSA and FESS embeddings with NN classification in the supervised pLSA learning setup

	MB	JS	JT	JT-W1	JT-W2
PLSA	67.29	68.97	73.61	73.38	68.97
FESS	68.26	68.33	72.76	71.69	68.33
ORIG	64.57				N/A

5 Conclusions

In this paper, we have presented a hybrid generative-discriminative classification approach that combines generative embeddings based on probabilistic latent semantic analysis (pLSA) with kernel-based discriminative learning based on a class of recently proposed information theoretic kernels. We applied the proposed approach to the diagnosis of Renal Cell Carcinoma on tissue micro array (TMA) images. We have seen that coupling the generative capabilities of pLSA with the discriminative capabilities of the information theoretic kernels yields higher classification accuracies than previous approaches based on linear and RBF kernels.

Acknowledgements. We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250).

References

1. Bicego, M., Lovato, P., Ferrarini, A., Delledonne, M.: Biclustering of expression microarray data with topic models. In: Proceedings of the International Conference on Pattern Recognition, pp. 2728–2731 (2010)
2. Bicego, M., Lovato, P., Oliboni, B., Perina, A.: Expression microarray classification using topic models. In: ACM Symposium on Applied Computing (Bioinformatics and Computational Biology track) (2010)
3. Bicego, M., Perina, A., Murino, V., Martins, A., Aguiar, P., Figueiredo, M.: Combining free energy score spaces with information theoretic kernels: Application to scene classification. In: Proceedings of the IEEE International Conference on Image Processing, pp. 2661–2664 (2010)
4. Blei, D., Ng, A., Jordan, M.: Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022 (2003)

5. Bosch, A., Zisserman, A., Muñoz, X.: Scene Classification Via pLSA. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 517–530. Springer, Heidelberg (2006)
6. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, pp. 401–408 (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 (2001)
8. Castellani, U., Perina, A., Murino, V., Bellani, M., Brambilla, P.: Brain morphometry by probabilistic latent semantic analysis. In: International Conference on Medical Image Computing and Computer Assisted Intervention (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, pp. 1–8 (2008)
10. Fuchs, T., Wild, P., Moch, H., Buhmann, J.: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In: International Conference on Medical Image Computing and Computer Assisted Intervention (2008)
11. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 42(1-2), 177–196 (2001)
12. Jaakkola, T., Haussler, D.: Exploiting generative models in discriminative classifiers. In: Advances in Neural Information Processing Systems, pp. 487–493 (1999)
13. Kononen, J., Bubendorf, L., Kallionimeni, A., Bärklund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M., Sauter, G., Kallionimeni, O.: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Medicine* 4, 844–847 (1998)
14. Lasserre, J., Bishop, C., Minka, T.: Principled hybrids of generative and discriminative models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York (2006)
15. Martins, A., Smith, N., Xing, E., Aguiar, P., Figueiredo, M.: Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research* 10, 935–975 (2009)
16. Martins, A.F.T., Bicego, M., Murino, V., Aguiar, P.M.Q., Figueiredo, M.A.T.: Information Theoretical Kernels for Generative Embeddings Based on Hidden Markov Models. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escalano, F. (eds.) SSPR&SPR 2010. LNCS, vol. 6218, pp. 463–472. Springer, Heidelberg (2010)
17. Ng, A., Jordan, M.: On discriminative vs generative classifiers: A comparison of logistic regression and naive Bayes. In: Advances in Neural Information Processing Systems (2002)
18. Perina, A., Cristani, M., Castellani, U., Murino, V., Jovic, N.: Free energy score space. In: Advances in Neural Information Processing Systems (2009)
19. Perina, A., Cristani, M., Castellani, U., Murino, V., Jovic, N.: A hybrid generative/discriminative classification framework based on free-energy terms. In: Proceedings of the International Conference on Computer Vision (2009)

20. Rubinstein, Y.D., Hastie, T.: Discriminative vs informative learning. In: Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pp. 49–53. AAAI Press (1997)
21. 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: Goelese, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) Pattern Recognition. LNCS, vol. 6376, pp. 202–211. Springer, Heidelberg (2010)
22. Schüffler, P., Ulaş, A., Castellani, U., Murino, V.: A multiple kernel learning algorithm for cell nucleus classification of renal cell carcinoma. In: Proceedings of the 16th International Conference on Image Analysis and Processing (2011)
23. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press (2004)
24. Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S., Müller, K.R.: A new discriminative kernel from probabilistic models. *Neural Computation* 14, 2397–2414 (2002)