Color Detection in Dermoscopy Images Based on Scarce Annotations

Catarina Barata[1(✉)], M. Emre Celebi[2], and Jorge S. Marques[1]

[1] Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal
ana.c.fidalgo.barata@ist.utl.pt
[2] Department of Computer Science, Louisiana State University, Shreveport, LA, USA

Abstract. Dermatologists often prefer clinically oriented Computer Aided Diagnosis (CAD) Systems. However, the development of such systems is not straightforward due to lack of detailed image annotations (medical labels and segmentation of their corresponding regions). Most of the times we only have access to medical labels that are not sufficient to learn proper models. In this study, we address this issue using the Correspondence-LDA algorithm. The algorithm is applied with success to the identification identification of relevant colors in dermoscopy images, obtaining a precision of 82.1 % and a recall of 90.4 %.

Keywords: Melanoma diagnosis · Correspondence-LDA · Image annotation · Color detection

1 Introduction

Computer Aided Diagnosis (CAD) systems for the analysis of dermoscopy images must fulfill a set of requirements, in order to be accepted by the medical community. For a system to be included in routine clinical practice it must be able to: (i) help dermatologists (instead of replacing them) and (ii) provide comprehensive information that justifies the diagnosis [1]. A way to solve the previous issues is to develop clinically oriented CAD systems [2] that are inspired by medical procedures [3]. These systems try to detect and characterize dermoscopic criteria (e.g., key colors or dermoscopic structures) [3], which are then used to diagnose the lesions.

One of the main problems of clinically oriented CAD systems it that they require detailed image annotations (i.e., medical labels and segmentation of relevant structures and colors) performed by experts. A dermatologist can easily provide medical labels stating whether or not a dermoscopic criterion is present or absent, but the segmentation of the corresponding regions is more difficult to obtain since this is a time consuming and highly subjective task. The absence of segmentations might result in incomplete systems. To prevent such issues, it is necessary to design systems that are capable of dealing with weakly annotated images (i.e., images for which there is only a medical label), and still provide

R. Paredes et al. (Eds.): IbPRIA 2015, LNCS 9117, pp. 309–316, 2015.
DOI: 10.1007/978-3-319-19390-8_35

information about the presence/absence of dermoscopic criteria as well as their location within the lesion.

This paper explores an approach to develop clinically oriented CAD systems, in which it is possible to learn a probabilistic representation of the dermoscopic criteria using only medical text labels. The probabilistic model used to learn the correlation between medical labels and image regions is a topical model called Correspondence-Latent Dirichlet Allocation (corr-LDA) [4]. This model is applied with success to the problem of color detection and to the best of our knowledge, this is the first time that such a model is applied to dermoscopy images. The remaining of the paper is organized as follows. Section 2 gives an overview of the investigated problem and identifies the main variables. Section 3 describes the general framework of the corr-LDA model and Sect. 4 describes the application of the model to dermoscopy images. Finally, Sect. 5 presents the experimental results and Sect. 6 concludes the paper.

2 Problem Statement

Dermatologists consider six clinically relevant colors (dark and light brown, blue-gray, black, red, and white [3]) that can be found in skin lesions. Some of these colors are more common in melanomas and the presence of 3 or more colors usually signals a suspicious lesion. A possible strategy to address the automatic detection of colors is to learn color models based on color region segmentations performed by dermatologists, as proposed in different works (e.g., [5,6]). Then, the learned models are applied to new images in order to identify the existing colors and their locations. These approaches can only be applied when color segmentations performed by experts are available, which is a considerable drawback. In this work, we do not have this information. Therefore, we investigate an alternative strategy that does not require medical segmentations.

Our dataset contains D dermoscopy images in which the lesions were divided into small patches, each patch being characterized by a feature vector r_n^d, $n = 1, ..., N^d$, where N^d is the number of patches of image d. Besides the patches, we have for each image a set of global texts labels provided by a group

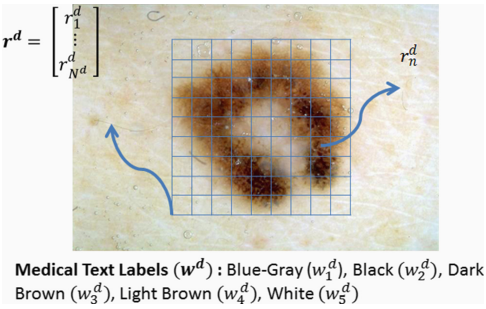


Fig. 1. Available information given image d: r_n^d is the feature vector of the $n-th$ patch, $\mathbf{r}^d$ is the set of feature vectors, w_m^d is one of the text labels, and $\mathbf{w}^d$ is the set of labels.

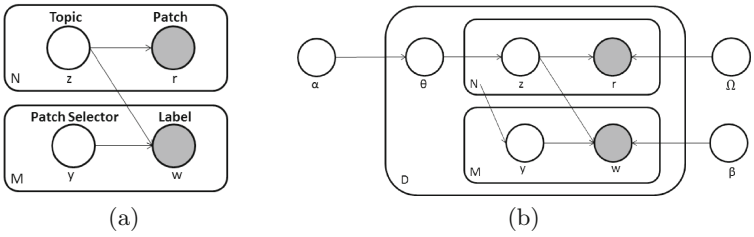


Fig. 2. Graphical representations of Corr-LDA: simplified (left) and complete representation (right). Each of the boxes represents an image, a patch or a label replication. The filled circles represent the variables observed in the training set.

of dermatologists w_m^d $m = 1, ..., M^d$, stating which are the M^d colors that can be found in the lesion d. Figure 1 exemplifies the used notation. Our goal is to be able to estimate a probabilistic model to represent the data, such that is not only possible to predict global labels for the lesion but also to be able to associate a color to each of the patches. An algorithm that suits this kind of formulation is corr-LDA [7]. This method allows the estimation of two important probabilities: (i) the label distribution given the image information $p(w_m^d|\mathbf{r}^d)$ and (ii) the distribution of the label given a single patch $p(w_m^d|r_n^d)$. The first probability can be used to predict global labels for the lesion, while the second probability can be used to fulfill the goal of associating a label to each patch. In the following section we address the main characteristics of corr-LDA and show how to compute the relevant probabilities.

3 Corr-LDA

corr-LDA is a probabilistic topic models that generates images using latent variables called topics [4]. This model is an extension of the popular LDA [4], which was proposed for text analysis and retrieval and later used with success in image analysis tasks such as scene recognition [8]. Without modifications, LDA can not be used for image annotation. Thus Blei and Jordan proposed the altered version corr-LDA that is capable of estimating a joint distribution between an image ($\mathbf{r}$) and its labels ($\mathbf{w}$), using the topics as hidden variables.

In order to obtain a better understanding of how the observed and latent variables relate to each other, take a look at Fig. 2(a). This figure depicts the generative process of corr-LDA, which can be summarized as follows. First, each of the N patch feature vectors (r_n) is generated conditioned on a topic z_n. This allows the creation of an image $\mathbf{r} = \{r_1, ..., r_N\}$ characterized by a set of latent topics $\mathbf{z} = \{z_1, ..z_N\}$. Then, each of the M labels w_m is generated. To generate a label it is necessary to first select an image patch and then condition the choice of the label on the topic that was used to create the patch. The selection of the image patches to be used in the labeling process is performed using the latent indexing variable y_m that takes values between 1 and N.

Each of these variables is generated by parametric distribution depending on a set of parameters. Figure 2(b) shows the parameters involved in this model, as well

as the generative process for a set of D images. The generative process involved in the creation of the pair image patches/labels $(\mathbf{r}^d, \mathbf{w}^d)$ is as follows [7]

1. For each image d, from a set of D images, sample a topic distribution $\theta \sim$ Dirichlet(α).
2. For each of the N image patches r_n
   (a) Sample $z_n \sim$ Multinomial(θ).
   (b) Sample $r_n \sim p(r|z_n, \Omega)$ from a multivariate Gaussian distribution conditioned on z_n.
3. For each of the M labels w_m
   (a) Sample $y_m \sim$ Uniform$(1, ..., N)$.
   (b) Sample $w_m \sim p(w|y_m, \mathbf{z}, \beta)$ from a multinomial distribution conditioned on the z_{y_m} topic.

α is the Dirichlet parameter and has the same size as the number of topics (K). Ω is the set of parameters of one of the $k = 1, ..., K$ multivariate Gaussian distributions that characterize the image patches, and β is the distribution of the possible labels over each of the k topics. These are model parameters, while θ is an image specific parameter of size K that is sampled once per image.

The application of corr-LDA can be divided into two phases. The first one is the learning phase, where the parameters of the model are estimated. In the second phase, the estimated parameters and distribution are applied to new images (test phase). Let us now focus on the learning stage. This phase depends on the estimation of the posterior distribution of the latent variables $(\theta, \mathbf{z}, \mathbf{y})$ given the patches' features and the global labels

$$p(\theta, \mathbf{z}, \mathbf{y}|\mathbf{w}, \mathbf{r}, \alpha, \beta, \Omega) = \frac{p(\theta, \mathbf{z}, \mathbf{r}, \mathbf{y}, \mathbf{w}|\alpha, \beta, \Omega)}{p(\mathbf{r}, \mathbf{w}|\alpha, \beta, \Omega)}. \tag{1}$$

However, the exact computation of the posterior distribution is intractable. To tackle this issue, Variational Inference is used to compute an approximation for the posterior [4,7]. This approach consists of using Jensen's inequality to obtain a family of lower bounds of the log-likelihood [9]. The lower bounds are defined using a set of variational parameters (γ, ϕ, λ). Based on this, it is possible to formulate a factorized distribution of the latent variables as follows

$$q(\theta, \mathbf{z}, \mathbf{y}) = q(\theta|\gamma) \left(\prod_{n=1}^{N} q(z_n|\phi_n) \right) \cdot \left(\prod_{m=1}^{M} q(y_m|\lambda_m) \right), \tag{2}$$

where γ is a K-dimensional Dirichlet parameter, ϕ_n are N K-dimensional multinomial parameters and λ_m are M N-dimensional multinomial parameters. These parameters are chosen during an optimization process that tries to compute the tightest possible lower bound (for details refer to [4,7]). The estimation of the parameters is performed using a variational EM. This method consists of iteratively applying the following two steps until convergence

– **E-Step:** The variational parameters $(\gamma^d, \phi^d, \lambda^d)$ are estimated for each image in the dataset and the lower bound is computed.

– **M-Step:** The model parameters α, β, and Ω are estimated by maximizing the lower bound obtained in the E-step.

The update equations for all the parameters will be introduced in the following section. Given an approximation for the posterior, it is also possible to compute the conditional distributions of interest $p(w|\mathbf{r})$ and $p(w|r_n)$ (please refer to [7]). After parameter estimation, corr-LDA can be applied to new images. First, the E-step must be applied to all the images using the estimated model parameters (α, β, Ω) in order to obtain the image specific variational parameters and distributions. Then, the joint probabilities $p(w|\mathbf{r})$ and $p(w|r_n)$ are computed in order to obtain global and patch labels.

4 Application to Dermoscopy Images

This section describes the application of corr-LDA to dermoscopy images as well as some modifications introduced to the original algorithm.

In order to apply corr-LDA, we divide the dermoscopy images into small non-overlapping patches of size 12×12 (same sized used in [6]) and characterize each of them using the mean color vector in the HSV color space. Patches which area is more than 50 % outside the lesion are discarded. Then, the features extracted from the training set were used to estimate the parameters of a corr-LDA model. Finally, the estimated parameters were applied to the test images as described in the previous section. To label the image patches we apply $p(w|r_n)$ to compute the probability of the six colors and then label the patch according to the color that has the highest probability. Global labeling is performed as follows. First, we compute $p(w|\mathbf{r})$ for each of the six colors and then we compare the outputs with an empirically determined threshold. Each color is considered present if $p(w|\mathbf{r})$ is above the threshold.

4.1 Inclusion of von-Mises Distribution

The HSV color space is a mixture of angular (Hue) and linear (Saturation and Value) information. The original corr-LDA assumes that the observations are modeled using Gaussian distributions. This is not appropriate in the case of the Hue channel, since this is a periodic angular measure. In order to describe the HSV color content, we modified the patch distribution $p(r_n|z_n, \Omega)$. In our approach, the H channel is modeled by a von-Mises distribution, while S and V channels are modeled using a multivariate Gaussian [7], as before. The von-Mises distribution is periodic, so it can be used to describe angular data such as H. Assuming independence between H and the remaining channels it is possible to obtain the following formulation

$$p(r_n|z_n, \Omega) = \nu(\mathrm{H}_n|z_n, \tau, \varepsilon).\, G(\mathrm{S}_n, \mathrm{V}_n|z_n, \mu, \Sigma), \tag{3}$$

where G is 2-dimensional Gaussian and ν is a von-Mises distribution

$$\nu(\mathrm{H}_n|z_n, \tau, \varepsilon) = \frac{1}{2\pi I_0(\varepsilon)} e^{\varepsilon \cos(\mathrm{H}_n - \tau)}, \tag{4}$$

where I_0 is the modified zero-order Bessel function of the first kind and $\varepsilon \geq 0$ denotes the concentration of the distribution around the mean τ.

Considering the new formulation of $p(r|z_n, \Omega)$ it is now possible to define the update equations for all the parameters. These equations are obtained by taking derivatives of the lower bound with respect to each of the parameters (refer to [4,7] for more details). The update equations are as follows

– **E-step** (performed for each lesion d)

$$\gamma_k^d = \alpha_k + \sum_{n=1}^{N_d} \phi_{nk}^d, \quad \lambda_{mn}^d \propto \exp\{\sum_{k=1}^{K} \phi_{nk}^d \log p(w_m|y_m = n, z_m = i, \beta)\}, \tag{5}$$

$$\begin{aligned} \phi_{nk}^d \propto p(r_n|z_n = k, \tau, \varepsilon, \theta, \Sigma) \exp\{E_q[\log q(\theta_k|\gamma^d)]\}. \\ . \exp\{\sum_{m=1}^{M_d} \lambda_{mn}^d \log p(w_m|y_m = n, z_m = i, \beta)\}, \end{aligned} \tag{6}$$

– **M-step** (performed using all the training set)

$$\alpha \text{ - use the Newton Raphson method [4]}, \beta \propto \sum_{d=1}^{D} \sum_{l=1}^{M_d} w_{lm}^d \sum_{n=1}^{N_d} \phi_{nk}^d \lambda_{ln}^d, \tag{7}$$

$$\mu_k = \frac{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d [\mathrm{S,V}]_n^d}{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d}, \tau_k = \tan^{-1}\left(\frac{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d \sin \mathrm{H}_n^d}{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d \cos \mathrm{H}_n^d}\right), \tag{8}$$

$$\varepsilon_k = \frac{\overline{R} - \overline{R}^3}{1 - \overline{R}^2}, \quad \text{where} \quad \overline{R} = \frac{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d \cos([\mathrm{H}]_n^d - \tau_k)}{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d}, [10] \tag{9}$$

$$\Sigma_k = \frac{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d ([\mathrm{S,V}]_n^d - \mu_k)([\mathrm{S,V}]_n^d - \mu_k)^T}{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{nk}^d}. \tag{10}$$

5 Results

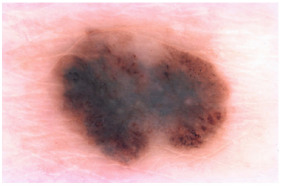
The proposed approach was evaluated on a set of 142 images randomly selected from the EDRA database [3]. This is a multisource dataset that contains dermoscopy images form three different university hospitals: University Federico II of Naples (Italy), University of Graz (Austria), and University of Florence (Italy). Each of the images was analyzed by a group of dermatologists who provided text labels regarding the presence or absence of the six clinically relevant colors (dark and light browns, blue-gray, black, red, and white).

To evaluate the performance of corr-LDA in the color detection problem we compute two metrics for each color: precision and recall. Precision corresponds to the proportion of images where a specific color was correctly identified among all the images where that color was detected. Recall is the percentage of images where the color was correctly identified. Both of these metrics are computed using a leave-one-out approach, where one of the images is kept for testing and the remaining 141 are used to estimate the parameters of corr-LDA. This is procedure is repeated for the 142 images and the final scores are the average of the individual scores.

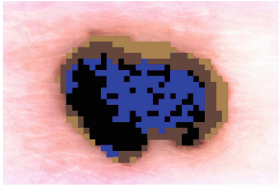
Figure 3 shows some examples of the output of corr-LDA as well as the labels provided by the experts. Table 1 shows the performance scores for each of the colors and the average performance of the model. Color identification is a challenging problem, especially if one is using only text labels, but corr-LDA achieves good results. The worse scores are obtained for the red and white colors, although good recall scores are obtained. This performance was expected since the num-

Table 1. Color detection results.

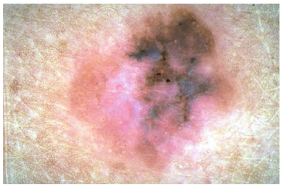
	Precision	Recall	#Images
Blue-Gray	91.7 %	86.4 %	103
Dark-Brown	96.3 %	98.5 %	134
Light-Brown	91.6 %	91.3 %	115
Black	84.3 %	100 %	100
Red	68.4 %	76.5 %	17
White	61.3 %	90.0 %	10
Average	82.1 %	90.4 %	-



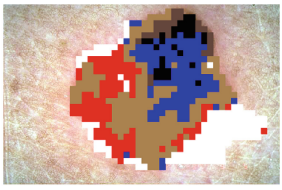
Dark and light browns, blue, black



Dark and light browns, blue, black



Dark brown, red, blue, white, black



Dark and light browns, red, blue, white, black.

Fig. 3. Original image medical labels (left) and output of corr-LDA (right).

ber of examples for each of these colors is significantly smaller than those of the remaining colors. We plan to address this problem by including more examples of these two colors.

6 Conclusions

This paper describes a methodology to detect relevant colors in dermoscopy images using weakly annotated data, i.e., we only have access to medical text labels. In order to tackle this issue, we use corr-LDA to obtain a probabilistic model that relates text labels and image features. This allows us to not only obtain global image labels but also to label image patches. The original corr-LDA was altered in order to include a von-Mises-Gaussian distribution, which was more suitable to describe the Hue data.

The obtained results were promising, with an average precision of 82.1 % and a recall of 90.4 %.

Acknowledgments. This work was funded by grant SFRH/BD/84658/2012 and by the FCT project FCT [UID/EEA/50009/2013].

References

1. Dreiseitl, S., Binder, M.: Do physicians value decision support? a look at the effect of decision support systems on physician opinion. Artif. Intel. Med. **33**(1), 25–30 (2005)
2. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: a review. Artif. Intel. Med. **56**(2), 69–90 (2012)
3. Argenziano, G., Soyer, H.P., De Giorgi, V., et al.: Interactive atlas of dermoscopy. In: EDRA (2000)
4. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
5. Seidenari, S., Pellacani, G., Grana, C.: Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment. Br. J. Dermatol. **149**(3), 523–529 (2003)
6. Barata, C., Figueiredo, M.A.T., Celebi, M.E., Marques, J.S.: Color identification in dermoscopy images using gaussian mixture models. In: ICASSP 2014, pp. 3611–3615 (2014)
7. Blei, D., Jordan, M.: Modeling annotated data. In: 26th ACM SIGIR, pp. 127–134. ACM (2003)
8. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR 2005, vol. 2, pp. 524–531. IEEE (2005)
9. Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. Mach. Learn. **37**(2), 183–233 (1999)
10. Sra, S.: A short note on parameter approximation for von mises-fisher distributions: and a fast implementation of $I_s(x)$. Comput. Stat. **27**(1), 177–190 (2012)