Trimming Prototypes of Handwritten Digit Images with Subset Infinite Relational Model

Tomonari Masada[1] and Atsuhiro Takasu[2]

[1] Nagasaki University, 1-14 Bunkyo-machi, Nagasaki-shi, Nagasaki, 852-8521 Japan,
masada@nagasaki-u.ac.jp

[2] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan,
takasu@nii.ac.jp

Abstract. This paper proposes a nonparametric Bayesian model for constructing a trimmed prototype representation of handwritten digit images. We assume that all images are resized to the same size. At each pixel point, we count the number of occurrences of grayscaled colors over multiple images. Then we obtain a color histogram at each pixel location. When we conduct this counting over images of the same category, e.g. images of handwritten digit "5", the obtained set of histograms can be regarded as a *prototype* of the category. After normalizing each histogram to a probability distribution over colors, we can calculate a likelihood of an unknown image by multiplying the probability of the color appearing at each pixel. We regard this method as the baseline and compare it with a method using a probabilistic model called Multinomialized Subset Infinite Relational Model (MSIRM), which constructs a prototype by clustering pixel columns and rows. While MSIRM can determine the number of clusters flexibly based on Chinese restaurant process, its interesting feature is that it can detect columns and rows *irrelevant* for constructing a prototype. In the experiment, we compared our method with the baseline and also with a histogram clustering by Dirichlet process mixture of multinomial. It was revealed that MSIRM could detect irrelevant columns and rows accurately at peripheral part of handwritten digit images. This means that MSIRM could provide trimmed prototypes. We could speed up testing processes by skipping irrelevant columns and rows with only a small degrade in accuracy.

Keywords: Bayesian nonparametrics, prototype, classification

1 Introduction

This paper considers image classification problem. While there are a vast variety of methods, we focus on *prototype-based* methods. We construct a prototype representation for each image category and classify an unknown image to the category whose prototype is the most similar to the image. Therefore, we need to propose a method for constructing a prototype and a method for evaluating similarity between images and prototypes. With respect to the former, we adopt a probabilistic approach and describe each prototype with probability distributions. With respect to the latter, we adopt a likelihood-based approach and classify an unknown image to the category giving the largest likelihood.

In this paper, we assume that all images are resized to the same size, say, N_1 by N_2 pixels and make image analysis focus only on color configuration. For example, Tiny Images Dataset[3] [7] is prepared for the experiments based on this intuition, where all images are resized to 32 by 32 pixels. When all images are of the same size, we can obtain a color histogram at each pixel location by counting the number of occurrences of colors over multiple images. That is, we have $N_1 N_2$ color histograms, each at different pixel locations. When we construct a set of histograms in this manner from the images of the same category, e.g. images of handwritten digit "5", it can be expected that the obtained set of histograms offers a color configuration specific to the category.

Formally, we construct a prototype as a set of parameters $\{g^h_{ijw}\}$, where g^h_{ijw} is the probability that the wth color appears at the 2D pixel location (i, j) for $w = 1, \ldots, W$, $i = 1, \ldots, N_1$, and $j = 1, \ldots, N_2$, where W is the number of colors. The superscript h is an index designating a particular category. For every pair of i and j, $\sum_{w=1}^{W} g^h_{ijw} = 1$ holds. We can calculate the log likelihood of an unknown image as $\sum_{i,j,w} n_{ijw} \log g^h_{ijw}$, where n_{ijw} is equal to 1 if the image has the wth color at the location (i, j) and is equal to 0 otherwise. Therefore, we can determine the category to which the image should be classified by calculating a log likelihood as $\arg\max_h \sum_{i,j,w} n_{ijw} \log g^h_{ijw}$ for each h. We regard this as the baseline method and would like to improve it with respect to efficiency.

Firstly, we reduce the complexity of prototypes. Any prototype given by the baseline has $N_1 N_2 W$ parameters, $\{g^h_{ijw}\}$. With respect to the number of colors, we simply quantize colors uniformly. A more intelligent quantization is reserved as a future work. For the rest of the paper, let W denote the number of quantized colors. We further reduce the number of probability distributions $N_1 N_2$ by clustering color histograms. This paper proposes a histogram clustering based on a new nonparametric Bayesian probabilistic model called Multinomialized Subset Infinite Relational Model (MSIRM), a modification of Subset Infinite Relational Model (SIRM) by Ishiguro et al.[2] The features of MSIRM are given below.

1. MSIRM clusters histograms in column and row wise. Denote the number of column clusters and that of row clusters as K_1 and K_2, respectively. Each pixel location is now indexed by a pair (k_1, k_2) of a column cluster ID k_1 and a row cluster ID k_2, where $k_1 = 1, \ldots, K_1$ and $k_2 = 1, \ldots, K_2$. MSIRM associates each pair (k_1, k_2) of column and row cluster IDs with different probability distributions over colors. Therefore, we can reduce the complexity of prototypes from $O(N_1 N_2 W)$ to $O(K_1 K_2 W)$.
2. MSIRM determines K_1 and K_2 flexibly by using Chinese restaurant process (CRP) [6] for calculating posterior probabilities of cluster assignments.
3. MSIRM detects *irrelevant* columns and rows based on the statistics found in a given image set, and clustering is conducted only on relevant columns and rows. This feature is inherited from SIRM.

[3] http://groups.csail.mit.edu/vision/TinyImages/

The main feature of MSIRM is the third one, because we can reduce the complexity of prototypes and can determine the number of clusters flexibly with any nonparametric Bayesian clustering, e.g. Dirichlet process mixture of multinomial [5]. In fact, the third feature can lead to an efficiency in time complexity. So, *secondly*, we reduce the time complexity of testing processes with MSIRM. The third feature of MSIRM can lead to a speeding up of testing processes, because we can reduce the execution time required for classifying unknown images by skipping irrelevant columns and rows in log likelihood calculation. Technically speaking, we assign probability one to the pixels of irrelevant columns and rows in log likelihood calculation. We will later show that skipping irrelevant columns and rows in log likelihood calculation only leads to a small degrade in accuracy.

The rest of the paper is organized as follows. Section 2 gives preceding proposals important for us. Section 3 provides details of MSIRM. Section 4 includes the results of our experiment. Section 5 concludes the paper with discussions.

2 Preceding Works

This paper proposes a nonparametric Bayesian model MSIRM as an extension of Subset Infinite Relational Model (SIRM) by Ishiguro et al. [2]. SIRM is, in turn, an improvement of Infinite Relational Model (IRM) by Kemp et al. [3].

Assume that we have a set of N_1 entities of type T_A and a set of N_2 entities of type T_B. Further, assume that we have a binary relation R defined over the domain $T_A \times T_B$. The relation R can be represented by an $N_1 \times N_2$ binary matrix, as is depicted in the left panel of Fig. 1. IRM discovers a bidirectional (horizontal-vertical) clustering of the entities so that the $N_1 \times N_2$ binary matrix takes on a relatively clean block structure when sorted according to the clustering, as is shown in the center panel of Fig. 1.

Let K_1 and K_2 be the number of column clusters and that of row clusters, respectively. IRM employs Chinese restaurant process (CPR) for determining K_1 and K_2 flexibly. Further, IRM associates each block, enclosed by thick lines in the center panel of Fig. 1, with a binomial distribution determining the probability that the pairs of entities in the corresponding block fall under the relation R. Obviously, each such block can be indexed by a pair of column cluster ID $k_1 \in \{1, \dots, K_1\}$ and row cluster ID $k_2 \in \{1, \dots, K_2\}$. Therefore, IRM associates each pair (k_1, k_2) of column cluster ID k_1 and row cluster ID k_2 with a binomial distribution. IRM can be extended to the cases where we have more than two types of entity, though such cases are not considered here.

When constructing a prototype from images of the same category, we consider relations between pixel columns and pixel rows by taking an IRM-like approach, because contiguous pixel locations are likely to give similar color distributions. However, IRM has the following two problems:

1. IRM is vulnerable to noisy data. We need a mechanism for detecting *irrelevant* columns or rows, which should be excluded from clustering process. This problem is addressed by SIRM [2].

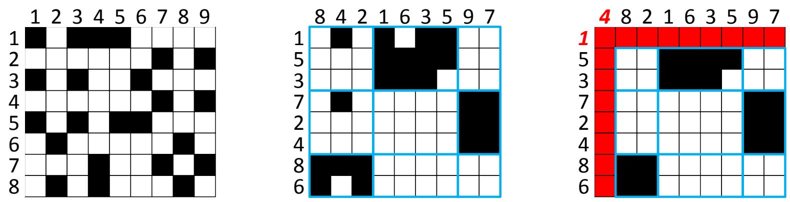

Fig. 1. Clustering of a binary relation (left) by IRM (center) and SIRM (right).

2. IRM can only handle binary data, though image colors are in general not binary. We need a mechanism for representing observed data with *multinomial* distributions. This problem is addressed by MSIRM we propose.

Ishiguro et al. [2] improved IRM by introducing a mechanisim for detecting *irrelevant* columns and rows. The proposed model is called *Subset* Infinite Relational Model (SIRM), because a nonparametric Bayesian clustering is conducted only on a subset of columns and rows after excluding irrelevant columns and rows. We flip a coin for each column and each row to determine whether relevant or not and then apply a clustering only to relevant columns and rows. The pixels belonging to an irrelevant column or an irrelevant row are bundled into a single cluster, as is shown in the right panel of Fig. 1 with red colored cells, where column 4 and row 1 are detected as irrelevant by SIRM. However, both IRM and SIRM can only handle binary data. Therefore, we modify SIRM to obtain our probabilistic model, MSIRM.

Needless to say, any existing clustering method can be applied to color histograms after ignoring 2D location information. The problem is then turned into an ordinary histogram clustering. Therefore, in the experiment presented in Section 4, we compared our proposal, not only with the baseline, but also with Dirichlet process mixture of multinomial distributions (DP-multinomial) [5].

3 MSIRM

This section contains the details of Multinomialized Subset Infinite Relational Model (MSIRM). We give a generative description of MSIRM below.

1. Draw a parameter λ_1 of the binomial distribution $\mathrm{Bi}(\lambda_1)$ for column-wise coin flips from the Beta prior $\mathrm{Be}(a_1, b_1)$. Then, for each of the N_1 columns, draw a 0/1 value from $\mathrm{Bi}(\lambda_1)$. Let r_{1i} be the value for the ith column, which is regarded as irrelevant if $r_{1i} = 0$, and as relevant otherwise.
2. Draw a parameter λ_2 of the binomial $\mathrm{Bi}(\lambda_2)$ for row-wise coin flips from the Beta prior $\mathrm{Be}(a_2, b_2)$. Then, for each of the N_2 rows, draw a 0/1 value from $\mathrm{Bi}(\lambda_2)$. Let r_{2j} be the value for the jth row, which is regarded as irrelevant if $r_{2j} = 0$, and as relevant otherwise.
3. Draw a set of parameters $\phi_{kl} = (\phi_{kl1}, \ldots, \phi_{klW})$ of the multinomial distribution $\mathrm{Mul}(\phi_{kl})$ from the Dirichlet prior distribution $\mathrm{Dir}(\beta)$. $\mathrm{Mul}(\phi_{kl})$ is the

multinomial for generating the color histograms belonging to the kth column cluster and, at the same time, to the lth row cluster. Each histogram is defined over W different colors. ϕ_{klw} is the probability that the wth color appears at a pixel location belonging to the kth column cluster and, at the same time, to the lth row cluster.

4. Draw a set of parameters $\psi = (\psi_1, \ldots, \psi_W)$ of the multinomial $\text{Mul}(\psi)$ from the Dirichlet prior $\text{Dir}(\gamma)$. $\text{Mul}(\psi_{kl})$ is the multinomial for generating irrelevant histograms, i.e., the histograms belonging to an irrelevant column or to an irrelevant row. ψ_w is the probability that the wth color appears at a pixel location belonging to an irrelevant column or to an irrelevant row.
5. For each relevant column, draw a cluster ID based on the Chinese restaurant process $\text{CRP}(\alpha_1)$. We introduce a latent variable z_{1i}, which is equal to k if the ith column is relevant and belongs to the kth column cluster. Let m_{1k} be the number of columns that are relevant and belong to the kth column cluster. $\text{CRP}(\alpha_1)$ assigns the ith column to the kth column cluster with the probability $\frac{m_{1k}}{\alpha_1+M_1}$ and to a new column cluster with the probability $\frac{\alpha_1}{\alpha_1+M_1}$, where $M_1 \equiv \sum_k m_{1k}$, i.e., the number of relevant columns.
6. For each relevant row, draw a cluster ID based on the Chinese restaurant process $\text{CRP}(\alpha_2)$. We introduce a latent variable z_{2j}, which is equal to l if the jth row is relevant and belongs to the lth row cluster. Let m_{2l} be the number of rows that are relevant and belong to the lth row cluster. Based on $\text{CRP}(\alpha_2)$, we assign the jth row to the lth row cluster with the probability $\frac{m_{2l}}{\alpha_2+M_2}$ and to a new row cluster with the probability $\frac{\alpha_2}{\alpha_2+M_2}$, where $M_2 \equiv \sum_l m_{2l}$, i.e., the number of relevant rows.
7. For each pixel location (i, j), $i = 1, \ldots, N_1$ and $j = 1, \ldots, N_2$, draw a color from $\text{Mul}(\psi)$ if $r_{1i}r_{2j} = 0$, and from $\text{Mul}(\phi_{z_{1i}z_{2j}})$ otherwise.

Assume that MSIRM finds K_1 column clusters and K_2 row clusters. Since MSIRM prepares one multinomial for all irrelevant columns and rows, we have $K_1K_2 + 1$ multinomials in total. Therefore, the complexity of prototypes is $O(K_1K_2W + N_1 + N_2)$, where $N_1 + N_2$ is the number of latent variables.

We adopt Gibbs sampling for inferring the posterior distributions of MSIRM. Here we introduce some notations. Let n_{klw} denote the number of times the wth color occurs in total at the pixel locations belonging to the kth column cluster and to the lth row cluster. Further, we define $N_{kl} \equiv \sum_w n_{klw}$, $q_w \equiv \sum_{i,j} n_{ijw}(1 - r_{1i}r_{2j})$, and $Q \equiv \sum_w q_w$, where n_{ijw} is the number of times the wth color appears at pixel location (i, j).

We consider an update of the cluster assignments for the ith column. Let z_{1i} denote the ID of the column cluster to which the ith column belongs. Then the conditional probability $p(z_{1i}, r_{1i}|\boldsymbol{X}, \boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i})$, given the cluster assignments $\boldsymbol{Z}^{\backslash 1i}$ and the coin flips $\boldsymbol{R}^{\backslash 1i}$ of the rest columns, can be written as follows:

$$p(z_{1i}, r_{1i}|\boldsymbol{X}, \boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) \propto p(\boldsymbol{X}^{+1i}|z_{1i}, r_{1i}, \boldsymbol{X}^{\backslash 1i}, \boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i})p(z_{1i}, r_{1i}|\boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) \ . \tag{1}$$

The superscript notation $\backslash 1i$ means that only the ith column is removed. The first half of the right hand side of Eq. (1) can be obtained as follows:

$$
\begin{aligned}
&p(\boldsymbol{X}^{+1i}|z_{1i}, r_{1i}=0, \boldsymbol{X}^{\backslash 1i}, \boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) \\
&= \frac{\prod_w \Gamma(q_w^{\backslash 1i} + \sum_{j=1}^{N_2} n_{ijw} + \gamma_w)}{\Gamma(Q^{\backslash 1i} + \sum_w \sum_{j=1}^{N_2} n_{ijw} + \sum_w \gamma_w)} \cdot \frac{\Gamma(Q^{\backslash 1i} + \sum_w \gamma_w)}{\prod_w \Gamma(q_w^{\backslash 1i} + \gamma_w)} \text{ and} \\
&p(\boldsymbol{X}^{+1i}|z_{1i}=k, r_{1i}=1, \boldsymbol{X}^{\backslash 1i}, \boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) \\
&= \prod_{l=1}^{K_2} \frac{\prod_w \Gamma(n_{klw}^{\backslash 1i} + \sum_{j=1}^{N_2} n_{ijw} r_{2j} z_{2jl} + \beta_w)}{\Gamma(N_{kl}^{\backslash 1i} + \sum_w \sum_{j=1}^{N_2} n_{ijw} r_{2j} z_{2jl} + \sum_w \beta_w)} \frac{\Gamma(N_{kl}^{\backslash 1i} + \sum_w \beta_w)}{\prod_w \Gamma(n_{klw}^{\backslash 1i} + \beta_w)} \\
&\quad \cdot \frac{\prod_w \Gamma(q_w^{\backslash 1i} + \sum_{j=1}^{N_2} n_{ijw}(1 - r_{2j}) + \gamma_w)}{\Gamma(Q^{\backslash 1i} + \sum_w \sum_{j=1}^{N_2} n_{ijw}(1 - r_{2j}) + \sum_w \gamma_w)} \frac{\Gamma(Q^{\backslash 1i} + \sum_w \gamma_w)}{\prod_w \Gamma(q_w^{\backslash 1i} + \gamma_w)} \ . \qquad (2)
\end{aligned}
$$

The second half of the right hand side of Eq. (1) can be obtained as follows:

$$
\begin{aligned}
p(z_{1i}, r_{1i}=0|\boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) &= \frac{b_1 + \sum_{i' \neq i}(1 - r_{1i'})}{a_1 + b_1 + N - 1} \ , \\
p(z_{1i}=k, r_{1i}=1|\boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) &\propto m_{1k}^{\backslash 1i} \cdot \frac{a_1 + \sum_{i' \neq i} r_{1i'}}{a_1 + b_1 + N - 1} \ , \text{ and} \\
p(z_{1i}=K_1+1, r_{1i}=1|\boldsymbol{Z}^{\backslash 1i}, \boldsymbol{R}^{\backslash 1i}) &\propto \alpha_1 \cdot \frac{a_1 + \sum_{i' \neq i} r_{1i'}}{a_1 + b_1 + N - 1} \ . \qquad (3)
\end{aligned}
$$

A similar argument can be repeated for image rows. The derivation is almost the same with that for SIRM. Further, we update the hyperparameters of Dirichlet priors and Beta priors based on a method proposed by Minka [4]. The number of Gibbs sampling iterations was set to 100 in our experiment.

4 Experiment

In the experiment, we compared the prototypes obtained by MSIRM with those obtained by the baseline method and also with those obtained by DP-multinomial, which was trained by Gibbs sampling [5].

Our target is MNIST dataset[4], consisting of 60,000 training images and 10,000 test images of handwritten digits. All images are 28 by 28 pixels in size, i.e., $N_1 = N_2 = 28$. We quantize 8 bit grayscale into 4 bit grayscale uniformly. Therefore, $W = 16$. Let n_{ijw}^h be the number of times the wth color appears at the pixel location (i, j) in the training images belonging to the hth category. Then, the color histogram placed at the location (i, j) is determined by $n_{ij1}, \ldots, n_{ijW}$. Therefore, each of the ten categories (i.e., from "0" to "9") induces a set of $28 \times 28 = 784$ histograms defined over 16 colors.

We obtain prototypes of the baseline method by calculating the probability of the wth color at the location (i, j) for each category h as $g_{ijw}^h = \frac{n_{ijw}^h + \eta_w}{\sum_w (n_{ijw}^h + \eta_w)}$.

[4] http://yann.lecun.com/exdb/mnist/

This is a MAP estimation of the probability when a Dirichlet prior with hyperparameters $\eta_1, \ldots, \eta_W$ is placed. We set $\eta_w = 0.1$ for all w. We classify 10,000 test images by calculating the log likelihood of each image as $\sum_i \sum_j \hat{n}_{ijw} \log g^h_{ijw}$, where $\hat{n}_{ijw}$ is equal to 1 if the wth color appears at the location (i, j) of the test image and is equal to 0 otherwise. Then we determine a candidate category for each test image as $\arg\max_h \sum_i \sum_j \hat{n}_{ijw} \log g^h_{ijw}$. For DP-multinomial, we set g^h_{ijw}s to the posterior probabilities estimated by Gibbs sampling. For MSIRM, we obtain a candidate category as $\arg\max_h \sum_i \sum_j \hat{n}_{ijw} \{ r_{1i} r_{1j} \log \frac{n_{z_{1i} z_{2j} w} + \beta_w}{\sum_w (n_{z_{1i} z_{2j} w} + \beta_w)} + (1 - r_{1i} r_{1j}) \log \frac{q_w + \gamma_w}{\sum_w (q_w + \gamma_w)} \}$.

We give the results of the experiment. With respect to the complexity of prototypes, DP-multinomial was superior to MSIRM for MNIST dataset. Let K be the number of clusters given by DP-multinomial. Then, the complexity of prototypes given by DP-multinomial is $O(KW)$, where $N_1 N_2$ is the number of latent variables for cluster assignments of pixel locations. Recall that the complexity of prototypes given by MSIRM is $O(K_1 K_2 W)$ as is discussed in Section 1. For MNIST dataset, $KW < K_1 K_2 W$, because K was around 85 for all digits, and K_1 and K_2 were both around 20 for all digits.

However, MSIRM could accurately detect irrelevant columns and rows at peripheral part of handwritten digit images as Figure 2 shows. We visualized prototypes by mixing grayscale colors linearly by multiplying their probabilities at each pixel location. In Fig. 2, the red colored area in each visualized prototype corresponds to irrelevant columns and rows. Astonishingly, MSIRM accurately detected the area irrelevant for identifying each digit. We consider this "trimming" feature with respect to classification accuracy. The accuracies of the baseline method, DP-multinomial, and MSIRM were 0.840, 0.839, and 0.837, respectively. While the accuracies were far from the best reported at the Web site of the dataset, they were good enough for a meaningful comparison. It can be said that our method and DP-multinomial gave almost the same classification accuracies with the baseline. However, for MSIRM, we can try another method for test image classification by utilizing irrelevant columns and rows. That is, we simply skip irrelevant columns and rows in calculation of log likelihoods. That is, we obtain a candidate category as $\arg\max_h \sum_{\{(i,j): r_{1i} r_{1j} = 1\}} \sum_w \hat{n}_{ijw} \log \frac{n_{z_{1i} z_{2j} w} + \beta_w}{\sum_w (n_{z_{1i} z_{2j} w} + \beta_w)}$. In case of the prototypes in Figure 2, we could skip 32.2% (2,523 pixels) of the $28 \times 28 \times 10 = 7,840$ pixels, and this led to a speeding up of testing processes. The achieved accuracy was 0.819, a small degrade from 0.837. This means that we can speed up testing by utilizing the trimming effect brought by MSIRM.

5 Conclusion

This paper proposes a prototype-based image classification method using a nonparametric Bayesian model called MSIRM. While MSIRM is a slight extension of SIRM, it is our novel idea to apply an infinite relational model to image analysis, because we usually do not view an image as a collection of "multicolored" relations between a column and a row. MSIRM could detect irrelevant columns

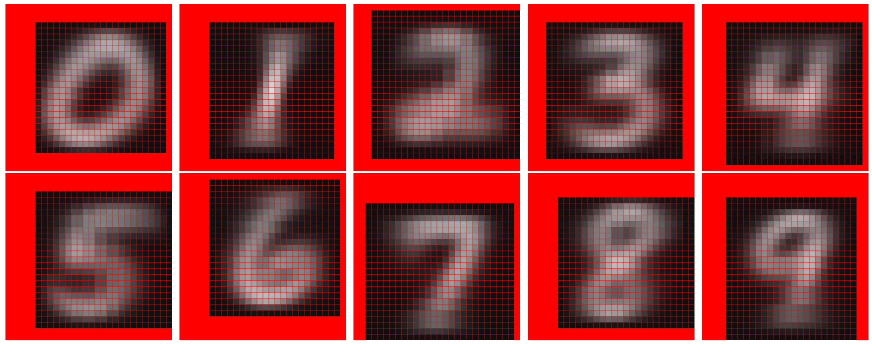

Fig. 2. Irrelevant part trimming achieved by MSIRM. The original image size is 28×28. The size of the relevant part of each prototype is: 22×22 for "0", 21×23 for "1", 25×25 for "2", 23×23 for "3", 23×24 for "4", 23×23 for "5", 22×23 for "6", 25×23 for "7", 23×22 for "8", and 22×24 for "9".

and rows accurately at peripheral part of handwritten digit images and could speed up testing processes by skipping irrelevant columns and rows with only a small degrade in accuracy. While DP-multinomial is comparable with MSIRM in its classification accuracy and is even superior to MSIRM in reduction of prototype complexity, it cannot detect any irrelevant part of digit images and thus cannot reduce the time required for test image classification.

We have a future plan to extend the proposed method so as to quantize colors in a nonparametric Bayesian manner by introducing an additional axis aside from the column and the row axes. Further, we also have a plan to incorporate a mechanism of clustering training images and to give more than one prototypes for each category.

References

1. Asuncion, A., Welling, M., Smyth, P., Teh, Y. W.: On smoothing and inference for topic models. in Proc. of UAI'09, pp. 27–34 (2009)
2. Ishiguro, K., Ueda, N., Sawada, H.: Subset infinite relational models. JMLR W&CP 22 (AISTATS 2012), 547–555 (2012)
3. Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., Ueda, N.: Learning systems of concepts with an infinite relational model. in Proc. of AAAI'06, pp. 381–388 (2006)
4. Minka., T. P.: Estimating a dirichlet distribution. http://research.microsoft.com/en-us/um/people/minka/papers/dirichlet/ (2000)
5. Neal, R. M.: Markov Chain Sampling Methods for Dirichlet Process Mixture Models. J. Comput. Graph. Statist., 9(2), 249–265 (2000)
6. Pitman, J.: Combinatorial stochastic processes. Notes for Saint Flour Summer School (2002)
7. Torralba, A., Fergus, R., Freeman, W. T.: 80 million tiny images: a large dataset for non-parametric object and scene recognition. IEEE Trans. Pattern Anal. Mach., 30(11), 1958–1970 (2008)