A Study on Model Selection from the *q*-Exponential Distribution for Constructing an Organ Point Distribution Model

Mitsunori Yamada$^{(\boxtimes)}$, Hidekata Hontani, and Hiroshi Matsuzoe

Nagoya Institute of Technology, Gokiso-cho, Showa-ku,
Nagoya-shi, Aichi 466-8555, Japan
`yamada@iu.nitech.ac.jp`

Abstract. A method is proposed that improves generalization performance of a point distribution model (PDM) of a target organ. Representing the PDM with a directed graphical model (DGM), the proposed method selects an appropriate model for each of the unary terms and of the pairwise terms of the DGM from a *q*-exponential distribution. The *q*-exponential distribution has a parameter, *q*, which controls the tail length, and its representation includes both a Gaussian distribution and a student's *t*-distribution: The distribution is identical with a Gaussian distribution when $q = 1$ and the distribution with a larger value of *q* has heavier tails. The proposed method selects a value of *q* for each of the distributions appeared in the DGM based on an Akaike's information criterion (AIC), which is employed for selecting a model that will minimize the generalization error. The proposed method is applied for the construction of a PDM of the liver and the results show that larger values of *q* are selected in the posterior region, which contacts with other soft organs.

Keywords: Point distribution model · Model selection · AIC · *q*-exponential distribution

1 Introduction

Point distribution models (PDMs) of target organs play important roles in the segmentation of 3D medical images (e.g. [1,2]). A PDM that represents statistical variety of the shape of a target organ is constructed from a set of training data and the model is registered to given new images for segmenting the target organ region in each of the images. In this study, employing a PDM for representing the statistical variety of the shapes, the authors focus on the improvement of the generalization performance of the model.

A PDM represents a target organ surface in an image with a set of N points on the surface. Let the three-dimensional coordinates of the *i*-th point in an image be denoted by $\boldsymbol{x}_j = [x_j, y_j, z_j]^T$ $(j = 1, 2, \ldots, N)$. Then an $3N$-vector,

$X = [x_1^T, x_2^T, \ldots x_N^T]^T$ denotes the location and the whole shape of the target organ. A statistical shape model of the target organ represents the prior probability distribution of these $3N$ variables $p(X) = p(x_1, x_2, \ldots, x_N)$ and a Gaussian distribution is widely employed implicitly and explicitly for this representation. A Gaussian distribution has two kinds of parameters, the mean μ and the covariance Σ, and you can represent the prior probability distribution with a Gaussian one by estimating the values of these parameters from a given training data set. You can estimate these parameter values straightforwardly from a set of training data but the estimated values are often degraded by outliers in the data, which are not well described by a Gaussian distribution, and often overfit to the training data when the number of the data is not large enough. These cause poor generalization performance for the resultant model. The objective of this study is *not* to estimate the true prior probability distribution X but to improve the generalization performance.

For example, an active shape model (ASM) [3,4], which is one of the most popular model used for the target organ segmentation in given medical images, implicitly employs a Gaussian distribution for representing the statistical variety of the shape of a target organ. Let a covariance matrix empirically estimated from a given training data set be denoted by $\hat{\Sigma}$ and let its eigenvectors and the corresponding eigenvalues be denoted by $u_1, u_2, \ldots$ and $\lambda_1, \lambda_2, \ldots$, respectively, where the eigenvalues are in decreasing order. Then, you can represent the statistical variety as follows:

$$X(\theta) = \hat{\mu} + U\theta, \tag{1}$$

where $\hat{\mu}$ denotes the estimated mean vector, $U = [u_1|u_2|\ldots|u_r]$ is a $3N \times r$ matrix where $r < 3N$, and θ is a r-vector that controls the shape described by the model. X in (1) obeys a Gaussian distribution, of which the mean is $\hat{\mu}$ and the covariance matrix is $\Sigma = U\Lambda U^T$ where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, when θ obeys a r-dimensional Gaussian distribution with zero mean and unit covariance. When you register the ASM to a given image, you firstly detect N points, $\tilde{x}_j$ ($j = 1, 2, \ldots, N$), from the image that correspond to x_j of the model and then estimate $\hat{\theta}$ in (1) so that $X(\hat{\theta})$ fits to the measured points, $\tilde{X} = [\tilde{x}_1^T, \tilde{x}_2^T, \ldots, \tilde{x}_N^T]^T$. You can estimate $\hat{\theta}$ by minimizing a cost function, $C(\theta)$, such that

$$C(\theta) = \lambda \|X(\theta) - \tilde{X}\|^2 + \|\theta\|^2. \tag{2}$$

The cost function of θ in (2) comes from $-\log\{p(\tilde{X}|\theta)p(\theta)\}$, where the prior probability distribution, $p(\theta)$, is represented by the Gaussian with zero mean and unit variance and where the likelihood, $p(\tilde{X}|\theta)$, is also the Gaussian of which the mean is zero and the variance is $\lambda^{-1}I$. It is a MAP estimation to minimize $C(\theta)$ in (2). You need to estimate the covariance matrix, $\hat{\Sigma}$, for constructing the model in (1) and the estimated covariance matrix often overfits to training data especially when the number of the data is not large [5]. This overfitting worsens the generalization performance [6].

You have at least two approaches to improve the generalization performance: In one approach, some regularizers or priors are introduced to the estimation

of $p(\boldsymbol{X})$[6]. In other approach, a model function of which the generalization performance would be the best is selected from a set of functions prepared in advance. In this study, the latter approach is employed: The proposed method selects an appropriate probability distribution model from q-exponential distribution, which can represent both a Gaussian distribution and a student's t-distribution, so that the generalization performance is improved. For the model selection, the authors employ an Akaike's information criterion. The q-exponential distribution and the AIC are described in the next section.

2   Proposed Method

Let the surface of a target organ in the i-th training image be denoted by $\mathcal{S}^i$ ($i = 1, 2, \ldots, M$) and let the three-dimensional coordinates of the j-th corresponding point generated on $\mathcal{S}^i$ be denoted by $\boldsymbol{x}_j^i$ ($j = 1, 2, \ldots, N$). Given a set of the training data, $\{\boldsymbol{x}_j^i | i = 1, 2, \ldots, M; j = 1, 2, \ldots, N\}$, we estimate the prior probability distribution, $p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$, as a model.

2.1   Graphical Model Representation

In this study, the prior probability distribution is represented as shown in (3). You can represent the probability distribution model (3) with a directed graphical model (DGM).

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \prod_j p(\boldsymbol{x}_j) \prod_{(j,k) \in \mathcal{E}} p(\boldsymbol{x}_j | \boldsymbol{x}_k), \qquad (3)$$

where each node in the corresponding graph represents $\boldsymbol{x}_j$ ($j = 1, 2, \ldots, N$) and $\mathcal{E}$ denotes a set of directed edges in the graph. Using the model shown in (3), you can register the model to a given image by computing the posterior probability or by maximizing it. You can of course employ an undirected graphical model for the representation such that

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \prod_j p(\boldsymbol{x}_j) \prod_{(j,k) \in \mathcal{E}'} p(\boldsymbol{x}_j, \boldsymbol{x}_k), \qquad (4)$$

where $\mathcal{E}'$ denotes a set of undirected edges in the corresponding graph. A directed graphical model is often employed when there exist causal relationships between the variables and it looks natural to employ an undirected one for representing a PDM [6,7] because the points on an organ surface do not have any causality. The authors, though, employ a DGM because it is not straightforward to compute a conditional probability distribution, $p(\boldsymbol{x}_j | \boldsymbol{x}_k)$, from a simultaneous one, $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$ in case the simultaneous distribution is represented by a q-exponential distribution. You need to compute $p(\boldsymbol{x}_j | \boldsymbol{x}_k)$ when you infer the posterior probability distribution on the graphical model by using, e.g. a belief propagation or MCMC. If you employ a q-exponential distribution, the pairwise term, $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$,

in (4) is represented by a six-dimensional q-exponential distribution but the conditional probability distribution, $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$, computed from the pairwise term is *not* represented by a q-exponential distribution. When you employ q-exponential distribution, it is not easy to represent the prior with an undirected graphical model as shown in (4).

Each unary term, $p(\boldsymbol{x}_j)$ in (3), is estimated from a set of the training data, $\{\boldsymbol{x}_j^i|i = 1, 2, \ldots, M\}$. The pairwise term, $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$, is assumed to satisfy $p(\boldsymbol{x}_j|\boldsymbol{x}_k) = p(\boldsymbol{x}_j - \boldsymbol{x}_k)$ in this study and is estimated from a set of the data, $\{\boldsymbol{d}_{jk}^i|i = 1, 2, \ldots, M; (j, k) \in \mathcal{E}\}$, where $\boldsymbol{d}_{jk}^i = \boldsymbol{x}_j^i - \boldsymbol{x}_k^i$. The q-exponential distribution, which has three different kinds of parameters, is employed for representing these terms, and the parameters except q are estimated by using an EM algorithm. The last parameter, q, controls the shape of the distribution and is determined based on an AIC.

2.2 q-Exponential Distribution

A d-dimensional q-exponential distribution is represented as follows:

$$p_q(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = Z \left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]^{1/1-q}, \tag{5}$$

where

$$Z = \frac{\Gamma((\nu + d)/2)}{(\pi\nu)^{d/2}\Gamma(\nu/2)|\boldsymbol{\Sigma}|^{1/2}},$$

$\nu = -d - 2/(1 - q)$, $\boldsymbol{\mu}$ is a d-vector and denotes the mean, and $\boldsymbol{\Sigma}$ is a $d \times d$ symmetry matrix and corresponds to the covariance matrix of a Gaussian distribution. The distribution is integrable if $q < 1 + 2/d$ and converges to a Gaussian distribution when $q \to 1$. Figure 1 shows graphs of one-dimensional q-exponential distributions that have identical values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and have different values of q. As you can see, the distributions with larger values of q have heavier tails.

2.3 EM Algorithm for Parameter Estimation

Fixing the value of q, you can estimate the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that maximize the likelihood from a given training data by using an EM algorithm. The likelihood of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to a given set of training data, $\{\boldsymbol{x}^i|i = 1, 2, \ldots, M\}$, is represented as shown in (6)[8].

$$L_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = M \log(Z) + \frac{1}{1 - q}\sum_{i=1}^M \log\left(\frac{\nu + s^i}{\nu}\right), \tag{6}$$

where $s^i = (\boldsymbol{x}^i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}^i - \boldsymbol{\mu})$. Maximizing L_q in (6), you obtain the following equations:

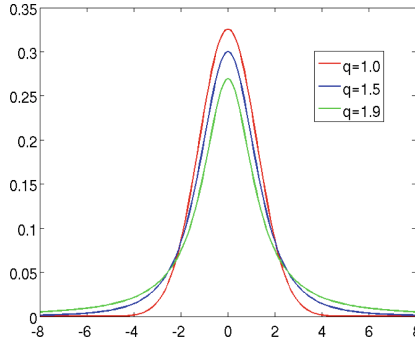$$\hat{\boldsymbol{\mu}}_{\text{ML}} = \frac{\sum_i w_i \boldsymbol{x}^i}{\sum_i w_i}, \tag{7}$$

Fig. 1. The shapes of the q-exponential distributions with different values of q. The distributions with larger values of q have heavier tails. The distribution is identical with a Gaussian one when $q = 1$.

$$\hat{\mathbf{\Sigma}}_{\mathrm{ML}} = \frac{1}{M} \sum_i w_i (\boldsymbol{x}^i - \hat{\boldsymbol{\mu}}_{\mathrm{ML}})(\boldsymbol{x}^i - \hat{\boldsymbol{\mu}}_{\mathrm{ML}})^T, \tag{8}$$

where $w_i = (\nu + d)/(\nu + s^i)$. As shown in the above equations, you need to compute the values of *weights*, w_i, in order for the estimation. Interpreting w_i as latent variables, you can estimate $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{ML}}$ with an EM algorithm [8]. The outline of the algorithm is described below.

1. Set initial values of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ as $\boldsymbol{\mu}^{(0)}$ and $\mathbf{\Sigma}^{(0)}$, respectively and set $m = 0$.
2. E-step:
   Calculate the weight, $w_i^{(m)}$, as follows:

$$w_i^{(m)} = \frac{\nu + d}{\nu + (\boldsymbol{x}^i - \boldsymbol{\mu}^{(m)})^T (\mathbf{\Sigma}^{(m)})^{-1}(\boldsymbol{x}^i - \boldsymbol{\mu}^{(m)})}. \tag{9}$$

3. M-step:
   Update the values of the parameters as follows:

$$\boldsymbol{\mu}^{(m+1)} = \frac{\sum_i w_i^{(m)} \boldsymbol{x}^i}{\sum_i w_i^{(m)}}, \tag{10}$$

$$\mathbf{\Sigma}^{(m+1)} = \frac{1}{M} \sum_i w_i^{(m)} (\boldsymbol{x}^i - \boldsymbol{\mu}^{(m+1)})(\boldsymbol{x}^i - \boldsymbol{\mu}^{(m+1)})^T. \tag{11}$$

4. Check if the values of the parameters are converged. If not, increase m as $m \leftarrow m + 1$ and back to E-step. Otherwise output the final values as $\hat{\boldsymbol{\mu}}_{\mathrm{ML}} = \boldsymbol{\mu}^{(m)}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{ML}} = \mathbf{\Sigma}^{(m)}$.

Replacing $\boldsymbol{x}^i$ with $\boldsymbol{d}^i_{jk}$, you can estimate $p(\boldsymbol{x}_j | \boldsymbol{x}_k)$ in the same manner.

2.4   Model Selection with AIC

The shape of the q-exponential distribution changes with respect to q. Given an identical set of training data, you obtain different estimations, $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ depending on the value of q and the generalization performance of the resultant statistical model changes with respect to q. In the proposed method, an appropriate value of q, which will minimize the generalization error, is selected for each of the distributions, $p(\boldsymbol{x}_j)$ and $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$, based on an AIC.

An Akaike's Information Criterion (AIC) is derived from a KL-divergence between an employed probability distribution model and the true distribution and is used for selecting a model that will minimize the generalization error from a set of models [9]. The AIC of the q-exponential distribution is a function of q:

$$\mathrm{AIC}(q) = -2L_q(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}) + 2l, \qquad (12)$$

where l denotes the number of the parameters. The q-exponential distributions with different values of q have identical number of the parameters and the second term of the right hand side in (12) is constant with respect to q. You can hence select the best model that minimizes AIC by selecting the value of q at which the likelihood, $L_q(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$, is maximum with respect to q. An algorithm for the selection of q for each of the distributions, $p(\boldsymbol{x}_j)$ and $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$, is as follows:

1. Set $q = 1$ and set a small positive step size Δ_q.
2. Estimate $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ by using an EM-algorithm and estimate the likelihood $L_q(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$ by means of a cross-validation.
3. Update q as $q \leftarrow q + \Delta_q$. If $q < 1 + 2/d$, back to 2.
4. Select the value q^*, which maximizes L_q.

3   Experimental Results

Sets of artificial data were used for evaluating if the proposed method improved the generalization performance and a set of CT images was used for constructing a PDM of the liver and for studying what values of q were selected for the probability distributions in (3).

3.1   Experiments with Artificial Data

Two families of data were generated by drawing them from an identical d-dimensional Gaussian probability distribution for examining if the model selection improved the generalization performance. The former family consisted of sets of training data and the latter one consisted of a set of data for evaluating the constructed models. Multiple sets of training data were generated in order to investigate the relationships between the number of training data and the selected value of q.

Let $\mathcal{X}_{n_{\mathrm{T}}}^m$ denote a set of n_{T} training data: $\mathcal{X}_{n_{\mathrm{T}}}^m = \{\boldsymbol{x}_1^m, \boldsymbol{x}_2^m, \ldots, \boldsymbol{x}_{n_{\mathrm{T}}}^m\}$ where $\boldsymbol{x}_n^m$ $(n = 1, 2, \ldots, n_{\mathrm{T}})$, are d-vectors and denote the training data. For statistically evaluating the constructed models, we generated M_{T} sets of n_{T} training

data, $\mathcal{X}_{n_{\mathrm{T}}}^1, \mathcal{X}_{n_{\mathrm{T}}}^2, \ldots, \mathcal{X}_{n_{\mathrm{T}}}^{M_{\mathrm{T}}}$, by drawing the data from an identical Gaussian distribution. By changing the value of n_{T} as $n_{\mathrm{T}} = n_{\mathrm{T}}^1, n_{\mathrm{T}}^2, \ldots, n_{\mathrm{T}}^{N_{\mathrm{T}}}$, we generated $N_{\mathrm{T}} \times M_{\mathrm{T}}$ sets of training data. The proposed method was then applied to each of the training data sets: $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ were estimated at each q and then q^* at which the AIC was minimum was selected.

The generalization performances of the constructed models, $p_q(\boldsymbol{x}|\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$, were evaluated by using the set of the data for the evaluation. Let the set for the evaluation be denoted by $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{N_{\mathrm{E}}}\}$, where $\boldsymbol{y}_i$ $(i = 1, 2, \ldots, N_{\mathrm{E}})$ are d-vectors and are drawn from the Gaussian distribution. The generalization error is evaluated as follows:

$$E(q) = -\sum_{i=1}^{N_{\mathrm{E}}} \log p_q(\boldsymbol{y}_i|\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}). \tag{13}$$

The error shown in (13) comes from a KL-divergence between a true probability distribution from which the data are derived and a model probability distribution. Let the true probability distribution be denoted by $\bar{p}_{\mathrm{T}}(\boldsymbol{x})$, which is assumed to be a Gaussian distribution in the experiments, and let the model probability distribution be denoted by $p_q(\boldsymbol{x})$. The KL-divergence between $\bar{p}_{\mathrm{T}}(\boldsymbol{x})$ and $p_q(\boldsymbol{x})$ is given as

$$\mathrm{KL}[\bar{p}_{\mathrm{T}}(\boldsymbol{x})||p_q(\boldsymbol{x})] = \int \bar{p}_{\mathrm{T}}(\boldsymbol{x}) \log \bar{p}_{\mathrm{T}}(\boldsymbol{x}) d\boldsymbol{x} - \int \bar{p}_{\mathrm{T}}(\boldsymbol{x}) \log p_q(\boldsymbol{x}) d\boldsymbol{x}. \tag{14}$$

The first term in the right hand side in (14) is independent of the model and cannot be evaluated in general. The second term can be approximately evaluated by using a set of observed data, $\{\boldsymbol{y}_i | i = 1, 2, \ldots, N_{\mathrm{E}}\}$, which are derived from $\bar{p}_{\mathrm{T}}(\boldsymbol{x})$, as $-\sum_i \log p_q(\boldsymbol{y}_i)$. The generalization error shown in (13) hence decreases when the KL divergence between the true distribution, $\bar{p}_{\mathrm{T}}(\boldsymbol{x})$, and the model $p_q(\boldsymbol{x})$ is smaller.

The experimental results demonstrated in this subsection were obtained when $d = 3$ and when the mean and the covariance matrix of the Gaussian were $\bar{\boldsymbol{\mu}} = [0, 0, 0]^T$ and $\bar{\boldsymbol{\Sigma}} = \mathrm{diag}(10, 10, 5)$, respectively. The number of the data used for the evaluation of (13) was set as $N_{\mathrm{E}} = 1000$ and the step size of q was set as $\Delta_q = 1/100$. The graphs in Fig. 2 show the relationships between the value of q and the generalization error, $E(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$, obtained when $n_{\mathrm{T}} = 10, 20, 40$, and 50. The bars in the graphs demonstrate one-sigma of the generalization errors evaluated by using $N_{\mathrm{E}} = 1000$ evaluation data. $E(q = 1)$ indicates the generalization errors of a (classical) Gaussian distribution model. The green dots indicate the value of $q = \bar{q}$ at which the generalization error, E, is minimum. The blue dots indicate the average value of q^*s that were selected from $M_{\mathrm{T}} = 100$ training data sets based on the AIC. The average value was not identical with $\bar{q}$ but the selected model ($q = q^*$) had better generalization performance than the Gaussian one ($q = 1$) even though all of the data obeyed the Gaussian probability distribution. As shown in Fig. 2, $\bar{q}$ and q^* decreased to one as the number of the training data, n_{T}, increased. The proposed method successfully selected a model that had the generalization performance better than a Gaussian one.

$$n_T = 10, \ \bar{q} = 1.18, \ q^* = 1.25$$

$$n_T = 20, \ \bar{q} = 1.19, \ q^* = 1.14$$

$$n_T = 40, \ \bar{q} = 1.07, \ q^* = 1.06$$
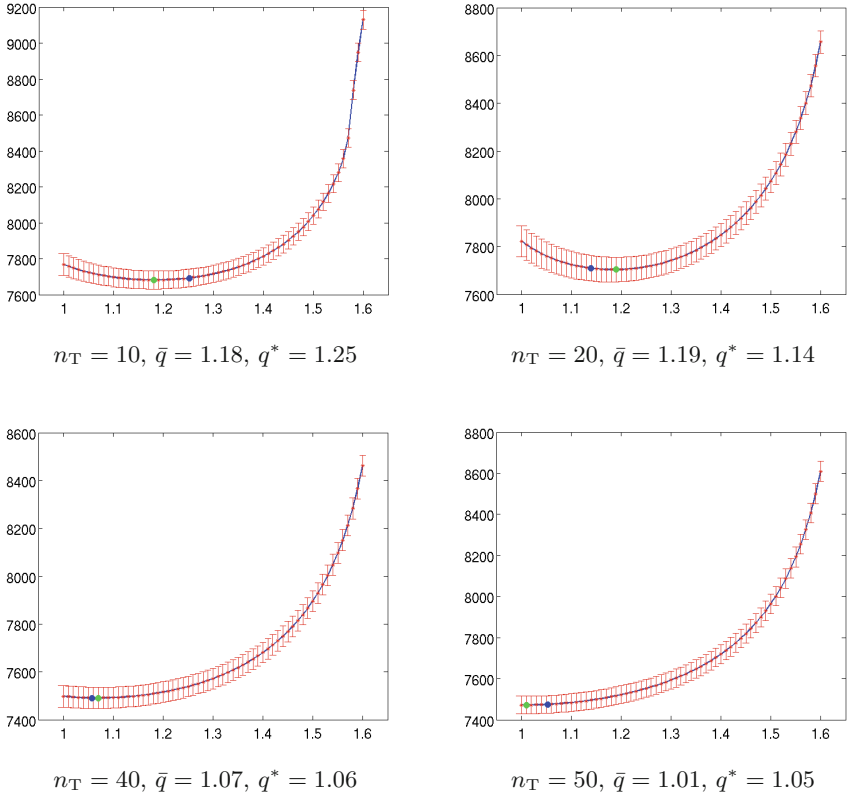
$$n_T = 50, \ \bar{q} = 1.01, \ q^* = 1.05$$

Fig. 2. Results of the simulation experiments. In each graph, the horizontal axis shows the value of q and the vertical one shows the evaluated generalization error, E in (13). The green dot indicates $\bar{q}$ where the error is minimum and the blue dot indicates the average value of q^*.

3.2   Experiments with CT Images

A PDM of the liver in a CT image was constructed from a set of 40 training CT images each in which the liver region was manually labeled. The statistical variety of the liver was represented by a directed graphical model as shown in (3) and q^* was selected for each of the unary terms and the pairwise ones, $p(\boldsymbol{x}_j)$ and $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$ based on the AIC. The step size of q was set as $\Delta_q = 1/1000$.

The locations and the shapes of the patients' bodies in the training images were normalized in advance by firstly automatically detecting 198 landmarks from each training image by a MCMC-based method [10,11] and by secondly deforming each image so that the detected landmarks were registered to their average locations. Figure 3 shows an example of the locations of the landmarks. Then, a set of 1300 corresponding points was generated on the surface of each labeled liver region in the deformed images by using a generalized multi-dimensional scaling (GMDS)[12] (see Fig. 4). Let $\boldsymbol{x}_j^i$ denote the j-th
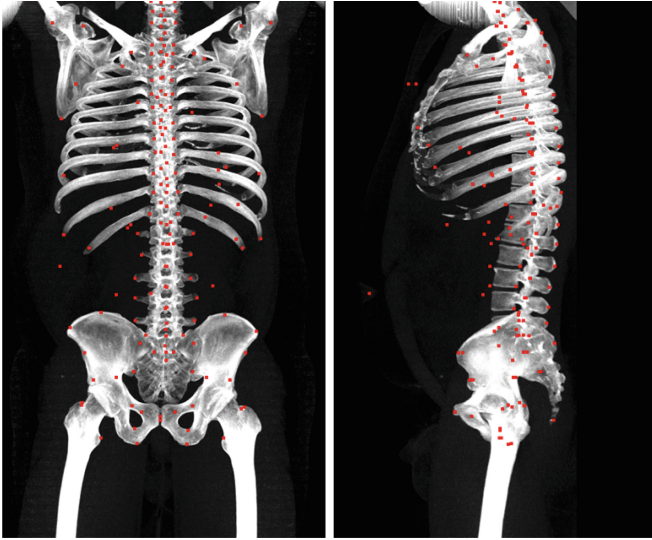
Fig. 3. The locations of the landmarks used for the image normalization (red dots)

corresponding point generated on the surface in the i-th training image. It should be noted that you will obtain a different statistical model from an identical set of the training images if you normalize the images in a different way or if you employ a different method for the generation of the corresponding points.

In this study, the structure of the directed graphical model was determined after q^* for each unary term, $p(\boldsymbol{x}_j)$, was selected. A pairwise term, $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$, is represented by a directed edge that links from $\boldsymbol{x}_k$ to $\boldsymbol{x}_j$ and the set of the edges, $\mathcal{E}$ in (3), was determined by the Delaunay diagram of the points on the average surface. You can obtain the average surface by computing the average of the training data of each point, $\hat{\boldsymbol{x}}_j = (\sum_{i=1}^{N} \boldsymbol{x}_j^i)/N$ and two nodes in the graphical model were linked by an edge if the corresponding two points in the PDM were linked by an edge in the Delaunay diagram. The direction of each edge was in the increasing order of q^*. Then, you obtain an acyclic directed graph. Once you determine $\mathcal{E}$, you can estimate $p(\boldsymbol{x}_j|\boldsymbol{x}_k)$ ($[j,k] \in \mathcal{E}$) from a set of training data, $\{\boldsymbol{d}_{jk}^i | i = 1, 2, \ldots, M\}$, where $\boldsymbol{d}_{jk}^i = \boldsymbol{x}_j^i - \boldsymbol{x}_k^i$.

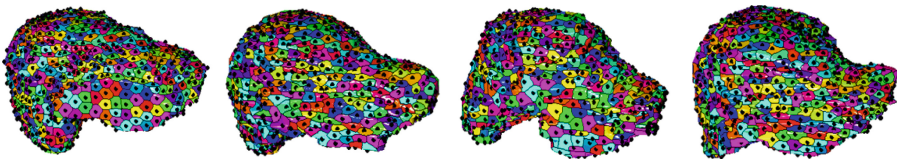**Fig. 4.** Examples of the corresponding points generated by GMDS [12]. The colors indicate the correspondence between different surfaces.

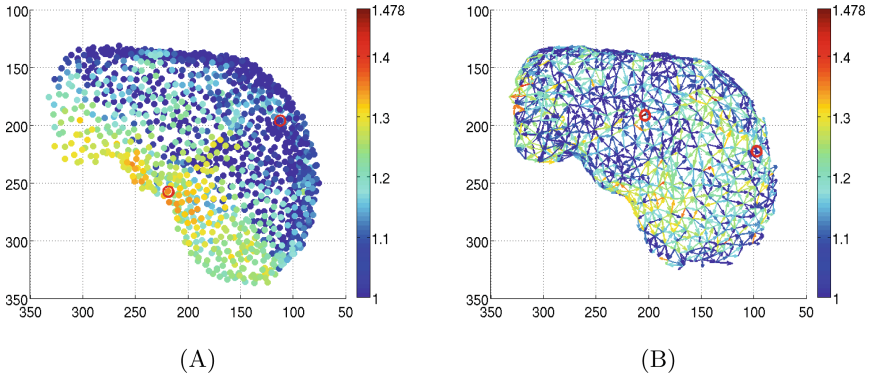(A)   (B)

Fig. 5. q^* selected for the unary terms (A) and for the pairwise terms (B). The training data set for each point marked by a circle is shown in Figs. 6 and 7
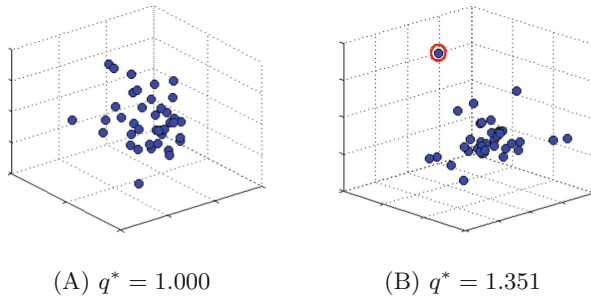

(A) $q^* = 1.000$ (B) $q^* = 1.351$

Fig. 6. Examples of the training data sets for the two points marked in Fig. 5. $q^* = 1.000$ was selected from (A) and $q^* = 1.351$ from (B).

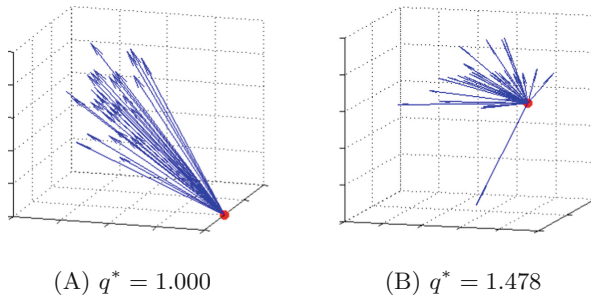(A) $q^* = 1.000$ (B) $q^* = 1.478$

Fig. 7. Examples of the training data sets for the two edges marked in Fig. 5. $q^* = 1.000$ was selected from (A) and $q^* = 1.478$ from (B).

Figure 5 shows the experimental results: The panel (A) demonstrates q^* selected for each single point on the liver surface and the panel (B) demonstrates q^* for each edge in the directed graph. As shown in the panel (A), larger values of q were selected at the points in the posterior portion of the liver: The shape of the probability distribution selected at each point in the posterior liver portion had heavier tails. Figure 6 shows two examples of the training data sets. The panel (A) shows a training data set of a point in the anterior portion of the liver at which $q^* = 1.000$ and the panel (B) shows a training data set in the posterior portion at which $q^* = 1.351$. As shown, some *outliers* (circled in the figure) were found in the training data set when larger q^* was selected. It should be noted that the location of the anterior portion of the liver is constrained more strongly than that of the posterior portion because of the locations of the landmarks used for the image normalization. The anterior portion of the liver is located near the anterior portion of the abdominal cavity, of which location in a normalized image is more strongly constrained by a set of the landmarks detected at the rib cage and the navel. As shown in Fig. 5(B), probability distributions with heavier tails were selected for representing the pairwise terms in the right portion of the liver. Two examples of the training data for the pairwise terms are shown in Fig. 7. Larger values of q were selected when the training data sets include outliers, of which the Mahalanobis distance from the mean are large.

4   Conclusion

The authors propose to represent the statistical variety of a target organ with a directed graphical model and to select an appropriate model from q-exponential distribution for representing the unary terms and the pairwise ones in order to improve the generalization performance. For the model selection, the AIC is employed. Simulation experiments demonstrated that the model selection improved the generalization performance and the experiments with clinical images showed that the distributions with heavier tails were selected for the representation of the unary terms in the posterior portion of the liver. The future works include to implement a method for registering the PDM constructed in this study and to compare it with a Gaussian model in regard to the registration performance.

References

1. Cootes, T.F., Taylor, C.J.: Statistical models of appearance for medical image analysis and computer vision. In: SPIE Medical Imaging, pp. 236–248. International Society for Optics and Photonics (2001)
2. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)

4. Blake, A., Isard, M.: Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. Springer Science & Business Media, New York (2012)
5. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
6. Sawada, Y., Hontani, H.: A study on graphical model structure for representing statistical shape model of point distribution model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 470–477. Springer, Heidelberg (2012)
7. Hontani, H., Tsunekawa, Y., Sawada, Y.: Accurate and robust registration of nonrigid surface using hierarchical statistical shape model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2977–2984. IEEE (2013)
8. Kotz, S., Nadarajah, S.: Multivariate t-Distributions and Their Applications. Cambridge University Press, Cambridge (2004)
9. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**(6), 716–723 (1974)
10. Nemoto, M., Masutani, Y., Hanaoka, S., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. In: SPIE Medical Imaging, p. 79623E. International Society for Optics and Photonics (2011)
11. Hanaoka, S., Masutani, Y., Nemoto, M., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: Probabilistic modeling of landmark distances and structure for anomaly-proof landmark detection. In: Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability, pp. 159–169 (2011)
12. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical Geometry of Non-rigid Shapes. Springer Science & Business Media, New York (2008)