Clustering with a Model of Sub-Mixtures of Different Distributions

Ivan Nagy*†, Evgenia Suzdaleva*, Matej Petrouš*
*Department of Signal Processing
The Institute of Information Theory and Automation of the Czech Academy of Sciences,
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
Email: suzdalev@utia.cas.cz
†Faculty of Transportation Sciences, Czech Technical University
Na Florenci 25, 11000 Prague, Czech Republic
Email: nagy@utia.cas.cz

***Abstract*—This paper deals with a modeling of data by several mixtures of different distributions within a task of clustering. This issue can be required from a practical point of view, e.g., for a multi-modal system, which generates measurements described by different distributions. The approach is based on the partition of the data on several parts, the factorization of the joint probability density function according to these parts and the estimation of each conditional mixture separately. Due to the data-based construction of the general model from the estimated components, the most suitable combination of the components is used at each time instant. The illustrative experiments are demonstrated.**

***Index Terms*—clustering, sub-mixtures, different distributions**

I. Introduction

This paper deals with a modeling of data by several mixtures of a different type of distributions. The task is considered within the cluster analysis. The clustering is often required in many application fields (medicine, industry, marketing, etc.), where the measurements should be grouped according to some similar attributes [1]. A variety of clustering approaches can be found in literature, such as, e.g., well-known centroid, density based methods, hierarchical clustering algorithms, etc. The overview of the clustering methods is available in [2], [3], etc.

The model-based clustering with the use of mixture models is one of approaches in this area. It describes the clusters in the data space by the components of a mixture model, see, e.g., [4], [5], etc. Normal components are probably most often used in this field.

The presented paper primarily focuses on the mixture of components, which have different distributions. This issue can be required from a practical point of view, e.g., for a multi-modal system, which generates measurements described by different distributions. The most significant application of the approach can be expected, for example, in modeling the traffic flow. In this case, the exponential distribution can be applied for modeling the traffic flow during congestions, the normal distribution for the free flow, the uniform distribution at night and the Bernoulli distribution for the queue existence. The air traffic as well as the train control can be also considered as the potential application fields.

The problem seems to be not sufficiently discussed in literature. Only several papers have been found. For instance, paper [6] deals with the mixture of the exponential, gamma and Weibull distributions. Paper [7] considers the estimation of the mixture of the normal-t distributions and the skew t-skew normal distributions. Paper [8] investigates the estimation of the t-, slash, contaminated normal and normal distributions. All of the papers found are based on the maximum likelihood estimation of parameters using the iterative expectation-maximization (EM) algorithm [9].

However, the EM algorithm is not suitable here, since the presented project avoids numerical computations as far as possible. The methodology applied within the project is based on the recursive Bayesian estimation algorithms [10], [11], [12], where the algebraic re-computation of the model statistics is used. In this area, the recursive estimation of the mixture with a different type of distributions has been discussed in [13]. One exponential and several normal components were considered and estimated separately with a common model of switching.

The presented paper takes the individual components as mixtures and constructs the general joint model from them. This opens a way for obtaining the estimation algorithm for a mixture of different distributions in the general form, which does not depend on the component type. However, the statistics of the probability density functions (pdf) should have the reproducible form. The proposed approach is based on (i) the partition of data on several parts, (ii) the factorization of the joint pdf according to these parts so that the mixture model of each part is conditioned by the previous one and (iii) the estimation of each conditional mixture separately. The switching models are conditional according to the factorization as well. Hence, the task of clustering with such a model consists in (i) the parameter estimation of all of the components and the switching models and (ii) the determination of the components,

This paper was supported by the project GAČR GA15-03564S.

978-1-5386-3855-2/17/$31.00 ©2017 IEEE

which generate data at the current time instant. Due to the data-based construction of the general model from the estimated components, the most suitable combination of the components is used at each time instant.

For the demonstration of the approach, an example with the model of three mixtures is presented in the paper. One of them is the mixture of exponential components and two others are the mixtures of normal components. The verification of the approach is performed with the help of simulated data and the comparison with theoretical counterparts.

The paper is organized in the following way. Section II describes the main idea, introduces the models and formulates the problem. Section III presents the general solution and the algorithm specified for the models, which are chosen for the illustrative example. Section IV demonstrates the validation results. Conclusions and some open problems can be found in Section V.

II. Problem Formulation

A. Main Idea

Let the observed multi-modal system generate measurements denoted by y. The measurements are assumed to be divided among several groups, where each group exhibits the different behavior. It means that the groups of the data have the different distributions. The groups are denoted as follows:

$$\begin{aligned} y^1 &= \left\{y_1^1, \cdots, y_{n_1}^1\right\}, \\ y^2 &= \left\{y_1^2, \cdots, y_{n_2}^2\right\}, \\ &\cdots, \\ y^k &= \left\{y_1^k, \cdots, y_{n_k}^k\right\}, \end{aligned}$$

where n_k is the number of the data entries in the k-th group of the data. The general model can be written in the form of the joint pdf, which can be subsequently factorized according to the chain rule [11] as follows:

$$\begin{aligned} & f\left(y^1, y^2, \cdots, y^k\right) \\ &= f\left(y^1\right) f\left(y^2|y^1\right) \cdots f\left(y^k|y^1, y^2, \cdots, y^{k-1}\right), \quad (1) \end{aligned}$$

where the factorization is performed according to the defined groups. Each pdf in (1) represents one of the new sub-models, which will be used for the construction of the joint model. The pdfs in (1) describe the variables in the individual data groups. They are conditional, i.e., the variables modeled by them are mutually dependent via the conditions of the pdfs.

The idea is to consider each pdf in (1) as the mixture model and to represent the general joint model with the help of these individual sub-models (i.e., sub-mixtures).

Each sub-mixture is composed of several components with the corresponding distributions. Generally, the mixture model works so that at each time instant the most suitable component is selected to represent the whole model. Here, each sub-mixture operates similarly and the resulting model is produced by the optimal components from the sub-mixtures. Obviously, the combination of the optimal components can be different at each time instant. Thus, the variability in setting the joint model is much wider compared with the whole joint pdf presented by a single mixture with components of different distributions.

B. Model Composed of Sub-Mixtures

In this paper, three groups of variables measured at time instants $t = 1, 2, \ldots$ are considered with the following dimensions

$$y_t^1 = \left[y_{1;t}^1, y_{2;t}^1\right]', \; y_t^2 = \left[y_{1;t}^2, y_{2;t}^2, y_{3;t}^2\right]', \; y_t^3 = \left[y_{1;t}^3, y_{2;t}^3\right]'.$$

The whole data set is $y_t = \left\{y^1, y^2, y^3\right\}_t$. In this way, y_t^1 and y_t^3 are the two-dimensional data vectors and y_t^2 is the three-dimensional one. In this paper, the first group of variables y_t^1 is described by the multivariate exponential distribution. The rest of them are normally distributed.

The component pdf within each mixture has the following form (omitting the superscript for the sake of simplicity)

$$f\left(y_t|\Theta, c_t = i\right), \forall i \in \{1, 2, \ldots, m_c\}, \quad (2)$$

where Θ is the collection of the component parameters. Switching the components is modeled as the unmeasurable discrete random variable, which is called the pointer [10]. The pointer is denoted by $c_t = \{1, 2, \ldots, m_c\}$, where m_c is the number of components. The value of c_t at the time instant t points to the component, which generates the actual measurements, i.e., the so-called active component. Generally, the components are switching according to the pointer model

$$f\left(c_t = i|\alpha\right), \quad (3)$$

where α is the parameter of the pointer model. Since the data groups are modeled as mixtures in the factorized form (1), their pointer models are interrelated. It means that the pointer of each sub-mixture depends on the pointers of the previous sub-mixtures.

With the presented factorized model, the clustering problem is verbally formulated as follows:

- estimate the component parameters Θ of all of the sub-mixtures,
- estimate the parameters α of all of the pointer models,
- determine the active component within each sub-mixture, i.e., the value of the pointer c_t
- and use the resulting combination of the components for classifying the data.

III. Estimation of a Model of Sub-Mixtures

The estimation algorithm for the discussed model of the sub-mixtures is based on the recursive Bayesian estimation methods [10], [11], [12] and recent papers [14], [13]. The terminology used and the detailed derivations can be found in these sources.

Denoting the data collection, which is available up to the time $t-1$ by $y(t-1) = \{y_0, y_1, y_2, \ldots, y_{t-1}\}$ and using the denotations from Section II, the joint pdf of the unknown variables to be estimated has the following form

$$f\left(y_t, c_t = i, \Theta, \alpha|y\left(t-1\right)\right) \quad (4)$$

$$= f\left(y_t^1, y_t^2, y_t^3, c_t^1, c_t^2, c_t^3, \Theta^1, \Theta^2, \Theta^3, \alpha^1, \alpha^2, \alpha^3 | y(t-1)\right),$$

where

- Θ^1, Θ^2, Θ^3 and α^1, α^2, α^3 are the parameters of the components and pointer models respectively of the corresponding data groups,
- c_t^1, c_t^2 and c_t^3 are the pointers of the corresponding sub-mixtures. Here, c_t^1 is described by the model (3), which is a vector. The pointer c_t^2 depends on c_t^1, while the pointer c_t^3 is conditioned by c_t^1 and c_t^2 . Both these models are matrices.

The joint pdf (4) is factorized under the assumption of the mutual independence of Θ and α as follows:

$$\underbrace{f\left(y_t^1|\Theta^1, c_t^1 = i,\right) f\left(c_t^1 = i|\alpha^1\right)}_{\text{for mixture 1}} \quad (5)$$
$$\times \underbrace{f\left(y_t^2|\Theta^2, y_t^1, c_t^2 = j\right) f\left(c_t^2 = j|c_t^1 = i, \alpha^2\right)}_{\text{for mixture 2}}$$
$$\times \underbrace{f\left(y_t^3|\Theta^3, y_t^1, y_t^2, c_t^3 = q\right) f\left(c_t^3 = q|c_t^1 = i, c_t^2 = j, \alpha^3\right)}_{\text{for mixture 3}}$$
$$\times \underbrace{f\left(\Theta^1|y(t-1)\right)}_{\text{prior exponential pdf}} \underbrace{f\left(\Theta^2|y(t-1)\right) f\left(\Theta^3|y(t-1)\right)}_{\text{prior GiW pdfs}}$$
$$\times \underbrace{f\left(\alpha^1|y(t-1)\right) f\left(\alpha^2|y(t-1)\right) f\left(\alpha^3|y(t-1)\right),}_{\text{prior Dirichlet pdfs}}$$

where the prior exponential pdf is used for the estimation of Θ^1 [15], [13]. The conjugate prior Gauss-inverse-Wishart (GiW) pdfs are used for Θ^2 and Θ^3, see [11], [10]. The conjugate prior Dirichlet pdfs are used for the estimation of α^1, α^2 and α^3 according to [12].

The pointer estimation is the key point of the discussed recursive clustering. The recursions can be derived with the help of the marginalization of (5) over the parameters Θ and α according to the following scheme (omitting the superscripts for the sake of simplicity)

$$f\left(y_t, c_t = i|y(t-1)\right) \quad (6)$$
$$= \int_{\Theta^*}\int_{\alpha^*} \left(y_t, c_t = i, \Theta, \alpha|y(t-1)\right) d\alpha d\Theta.$$

A. The 1st Sub-Mixture Estimation

For the mixture of the exponential components, the recursive estimation formulas are obtained from

$$\int_{\Theta^*}\int_{\alpha^*} f\left(y_t^1|\Theta^1, c_t^1 = i,\right) f\left(c_t^1 = i|\alpha^1\right) \quad (7)$$
$$\times f\left(\Theta^1|y(t-1)\right) f\left(\alpha^1|y(t-1)\right) d\alpha^1 d\Theta^1$$
$$= \int_{\Theta^*} f\left(y_t^1|\Theta^1, c_t^1 = i,\right) f\left(\Theta^1|y(t-1)\right) d\Theta^1$$
$$\times \int_{\alpha^*} f\left(c_t^1 = i|\alpha^1\right)) f\left(\alpha^1|y(t-1)\right) d\alpha^1,$$

where $f\left(y_t^1|\Theta^1, c_t^1 = i,\right)$ for $i = 1, 2, 3$ are the exponential components with the specified parameters $\Theta_i^1 = \{a, b\}_i$ for $c_t = i$, i.e.,

$$f\left(y_t^1|\Theta_i^1 = \{a, b\}_i\right) = a_i b_i \exp\left\{-a_i y_{1;t}^1 - b_i y_{2;t}^1\right\}. \quad (8)$$

In the first integral in (7), $f\left(\Theta^1|y(t-1)\right)$ is the prior exponential pdf

$$\left(a_i b_i\right)^{\kappa_{i;t}^1} \exp\left\{-\left[a_i, b_i\right]\left(S_t\right)_i\right\} \quad (9)$$

with the re-computable statistics (initially chosen) in the form

$$\kappa_{i;t}^1 = \kappa_{i;t-1}^1 + w_{i;t}^1, \quad (10)$$
$$\left(S_t\right)_i = \left(S_{t-1}\right)_i + w_{i;t}^1 \begin{bmatrix} y_{1;t}^1 \\ y_{2;t}^1 \end{bmatrix}, \quad (11)$$

where $w_t^1 = [w_{1;t}^1, w_{2;t}^1, w_{3;t}^1]$ is the weighting vector for the first sub-mixture, i.e., the first part of the general model, see [15], [13] based on [10]. This integral is evaluated with the help of the substitution of the point estimates of the component parameters from the previous time instant $t-1$

$$\hat{a}_{i;t-1} = \frac{\kappa_{i;t-1}^1}{\left(S_{1;t-1}\right)_i}, \ \hat{b}_{i;t-1} = \frac{\kappa_{i;t-1}^1}{\left(S_{2;t-1}\right)_i} \quad (12)$$

and the actual measurements y_t^1 into the model pdf, see the mentioned sources for the details.

The second integral in (7) uses the conjugate prior Dirichlet pdf [12] with the re-computable statistics in the form of the three-dimensional vector and the pointer model (3). The result of their multiplication is normalized.

For the first group of the data, the weighting vector w_t^1 is obtained with the help of the multiplication of the results of the first integral (after the substitution of the point estimates of the parameters and the actual data item y_t^1 into the component exponential pdf) and the point estimate of the parameter α^1 for the time instant $t-1$.

The index of the maximum weight in the weighting vector w_t^1 is the point estimate of the pointer c_t^1, which points to the active component within this sub-mixture.

B. The 2nd Sub-Mixture Estimation

The recursions for the second data group are derived in a similar way, i.e., using

$$\int_{\Theta^*} f\left(y_t^2|\Theta^2, y_t^1, c_t^2 = j\right) f\left(\Theta^2|y(t-1)\right) d\Theta^2 \quad (13)$$
$$\times \int_{\alpha^*} f\left(c_t^2 = j|c_t^1 = i, \alpha^2\right) f\left(\alpha^2|y(t-1)\right) d\alpha^2.$$

The differences are as follows. Here, the pdfs $f\left(y_t^2|\Theta^2, y_t^1, c_t^2 = j\right)$ for $j = 1, 2, 3$ are the normal components with the expectations

$$\theta_j^2 \begin{bmatrix} y_t^1 \\ 1 \end{bmatrix}$$

and the covariance matrices r_j^2. It means that within the second sub-mixture, the parameters Θ_j^2 include $\{\theta^2, r^2\}_j$ for $c_t^2 = j$. The conjugate GiW pdfs $f\left(\Theta^2|y(t-1)\right)$ used in the first integral of (13) have the re-computable statistics in the form of the counter $\kappa^2_{j;t}$ and the information matrix $(V_t)^2_j$. They are recursively updated (starting from the values chosen initially) in the following way [10]

$$\kappa^2_{j;t} = \kappa^2_{j;t-1} + w^2_{j;t}, \quad (14)$$

$$(V_t)^2_j = (V_{t-1})^2_j + w^2_{j;t} \begin{bmatrix} y^2_t \\ y^1_t \\ 1 \end{bmatrix} \left[y^2_t,\, y^1_t,\, 1 \right], \quad (15)$$

where $w^2_t = [w^2_{1;t}, w^2_{2;t}, w^2_{3;t}]$ is the weighting vector of the second sub-mixture. The first integral in (13) is evaluated by substituting the previous-time point estimates of the parameters Θ^2 and the actual data item y^2_t into the normal component pdf, see [13] based on [10]. The point estimates for each component are obtained according to [11] using the partition

$$(V_t)^2_j = \begin{bmatrix} V_{yy} & V'_y \\ V_y & V_1 \end{bmatrix}, \quad (16)$$

where V_{yy}, V'_y and V_1 are matrices and vectors of the appropriate dimensions depending on the dimensions of the vectors y^2_t and y^1_t. The point estimates are [11]

$$(\hat{\theta}^2_t)_j = V_1^{-1} V_y, \quad (\hat{r}^2_t)_j = \frac{V_{yy} - V'_y V_1^{-1} V_y}{\kappa^2_{j;t}}. \quad (17)$$

The pointer model used in the second integral of (13) is conditional due to the factorization (1). It is represented by the following table:

$$f\left(c^2_t = j | c^1_t = i, \alpha^2\right) = \quad (18)$$

	$c^2_t = 1$	$c^2_t = 2$	$c^2_t = 3$
$c^1_t = 1$	$\alpha^2_{1\|1}$	$\alpha^2_{2\|1}$	$\alpha^2_{3\|1}$
$c^1_t = 2$	$\alpha^2_{1\|2}$	$\alpha^2_{2\|2}$	$\alpha^2_{3\|2}$
$c^1_t = 3$	$\alpha^2_{1\|3}$	$\alpha^2_{2\|3}$	$\alpha^2_{3\|3}$

The second integral in (13) is evaluated again with the prior Dirichlet pdf according to [12] with the re-computable statistics. However, here the statistics is the (3×3)-dimensional matrix, which is updated similarly to the dynamic pointer model in [13] but with c^1_t in the condition. The point estimate of the parameter α^2 is obtained by the normalization of the corresponding statistics of the pointer model.

To obtain the weighting vector w^2_t of the second sub-mixture, it is necessary to multiply the results of the first integral, the weighting vector w^1_t and the point estimate of the parameter α^2 from the previous time instant. The point estimate of the pointer c^2_t as well as the active normal component are determined similarly to the previous case.

C. The 3rd Sub-Mixture Estimation

Here, the situation is practically identical with the previous normal mixture. However, due to the factorization (1), the conditions of the pdfs of the third sub-mixture are enriched by the measurements and the pointers from the first and second mixtures. It means that the recursions are obtained using

$$\int_{\Theta^*} f\left(y^3_t | \Theta^3, y^1_t, y^2_t, c^3_t = q\right) f\left(\Theta^3 | y(t-1)\right) d\Theta^3 \quad (19)$$

$$\times \int_{\alpha^*} f\left(c^3_t = q | c^1_t = i, c^2_t = j, \alpha^3\right) f\left(\alpha^3 | y(t-1)\right) d\alpha^3.$$

The normal components $f\left(y^3_t | \Theta^3, y^1_t, y^2_t, c^3_t = q\right)$ for $q = 1, 2, 3$ have the expectations

$$\theta^3_q \begin{bmatrix} y^2_t \\ y^1_t \\ 1 \end{bmatrix}$$

and the covariance matrices r^3_q, i.e., the parameters Θ^3_q include $\{\theta^3, r^3\}_q$ for the pointer $c^3_t = q$. The statistics of the corresponding GiW pdfs are re-computed similarly to (14), see [10], i.e.,

$$\kappa^3_{q;t} = \kappa^3_{q;t-1} + w^3_{q;t}, \quad (20)$$

$$(V_t)^3_q = (V_{t-1})^3_q + w^3_{q;t} \begin{bmatrix} y^3_t \\ y^2_t \\ y^1_t \\ 1 \end{bmatrix} \left[y^3_t,\, y^2_t,\, y^1_t,\, 1 \right], \quad (21)$$

where $w^3_t = [w^3_{1;t}, w^3_{2;t}, w^3_{3;t}]$ is the weighting vector for this data group. The point estimates of the component parameters are computed according to (17), see [11].

The pointer model $f\left(c^3_t = q | c^1_t = i, c^2_t = j, \alpha^3\right)$ is the (3×3)-dimensional table, which exists for each value of the pointer c^1_t. For each $i = 1, 2, 3$, it has the following form

$$f\left(c^3_t = q | c^1_t = i, c^2_t = j, \alpha^3\right) = \quad (22)$$

	$c^3_t = 1$	$c^3_t = 2$	$c^3_t = 3$
$c^2_t = 1$	$(\alpha^3_{1\|1})_i$	$(\alpha^3_{2\|1})_i$	$(\alpha^3_{3\|1})_i$
$c^2_t = 2$	$(\alpha^3_{1\|2})_i$	$(\alpha^3_{2\|2})_i$	$(\alpha^3_{3\|2})_i$
$c^2_t = 3$	$(\alpha^3_{1\|3})_i$	$(\alpha^3_{2\|3})_i$	$(\alpha^3_{3\|3})_i$

The evaluation of the integrals in (19) is performed identically to Section III-B. The weighting vector w^3_t is also obtained similarly using w^2_t and α^3 for the multiplication.

D. Clustering with the Model of Sub-Mixtures

In this way, the sub-mixtures are estimated in a standard way of the recursive Bayesian estimation [10], [12], [14], [13]. The substantial difference is the dependence of each pointer on the pointers of the previous sub-mixtures used in the factorized form (1) of the general model.

The resulting general model is composed of the active components of the sub-mixtures corresponding to the maximum weights in the actual weighting vectors w^1_t, w^2_t and w^3_t at the time instant t. The indices of these maximum weights are the point estimates of the pointers c^1_t, c^2_t and c^3_t respectively.

The clustering algorithm with the resulting model can be summarized as follows.

Initialization for $t = 1$

1) Set the number of sub-mixtures and components.
2) Set the initial statistics of the prior pdfs for components and pointers.
3) Compute the point estimates of all of the parameters.
4) Set the initial weighting vectors. Details about the initialization can be found, e.g., in [16].

On-line clustering for $t = 2, 3, \ldots$

1) Measure the data items y_t^1, y_t^2 and y_t^3.
2) For each sub-mixture, substitute the previous point estimates of the parameters and the actual measurements into the component pdfs.
3) For the first sub-mixture, multiply the result of Step 2 and the previous point estimate of the pointer parameter α^1 (i.e., at time $t-1$), see Section III-A. Normalize the result of this multiplication and obtain the weighting vector w_t^1 actualized by the data.
4) For the second sub-mixture, multiply the result of Step 2, the weighting vector w_t^1 and the previous point estimate of the pointer parameter α^2, see Section III-B. Obtain w_t^2 with the help of the normalization.
5) For the third sub-mixture, multiply the result of Step 2, the weighting vector w_t^2 and the previous point estimate of the pointer parameter α^3, see Section III-C. Obtain w_t^3 with the help of the normalization.
6) For each sub-mixture, update the statistics of all of the components and pointers using the new weights.
7) Re-compute the point estimates of the parameters.
8) Obtain the point estimates of the pointers as indices of the maximum weights in the weighting vectors.
9) Classify the data items according to the point estimates of the pointers, which give the active components.
10) Go to Step 1 of the on-line part of the algorithm and use the re-computed point estimates of the parameters.

The details about the estimation can be found in [10], [12], [14], [13], etc.

IV. Validation Experiments

The presented approach was verified with the help of the comparison with the k-means method [2] known as a successful classifier.

For the experiments, 20 data sets with 300 values of the vectors y_t^1, y_t^2 and y_t^3 (see Section II-B) were simulated in Scilab (www.scilab.org) using different random generators. The obtained clustering results were of the similar quality for all of the data sets. The typical results are presented below.

A. Clusters

In the case of the seven-dimensional data space, the visualization of the clusters is possible only for the data pairs, i.e., by plotting two variables against each other. For the exponential data y_t^1, the results of the proposed clustering are shown in Figure 1 (top), where the variables $y_{1;t}^1$ and $y_{2;t}^1$ are plotted against each other. The clusters are compared with the results of the k-means algorithm [2] in Figure 1 (bottom). The clusters in both the figures have the similar shapes and locations. The only difference can be found near value 25 of the variable $y_{1;t}^1$, where the data item belongs to cluster 2 in Figure 1 (top) and to cluster 1 in Figure 1 (bottom).

The results for the normal data y_t^2 are presented in Figure 2 (top) and compared with the k-means clustering in Figure 2 (bottom). The variables $y_{2;t}^2$ and $y_{3;t}^2$ from the vector y_t^2 are chosen for the cluster visualization. Here, the clusters of both the methods have the similar shapes and locations with the exception of several data items around value 10 of the variable $y_{3;t}^2$. They were classified into cluster 1 by the proposed method and into clusters 2 and 3 by the k-means algorithm.

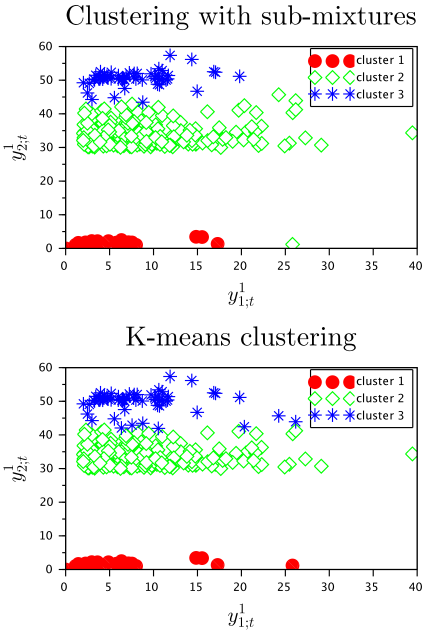

Fig. 1. The clustering with sub-mixtures (top) and the k-means (bottom) for the exponential variables $y_{1;t}^1$ and $y_{2;t}^1$

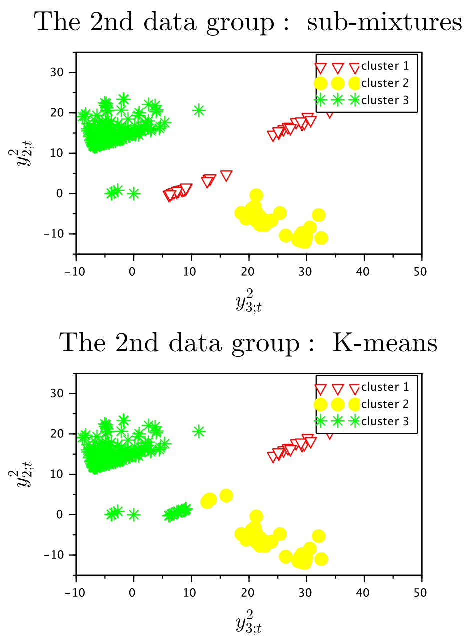

Fig. 2. The clustering with sub-mixtures (top) and the k-means (bottom) for the normal variables $y_{2;t}^2$ and $y_{3;t}^2$

The clusters detected in the third group of the normal data y_t^3 are demonstrated in Figure 3, where the variables $y_{1;t}^3$ and $y_{2;t}^3$ are plotted. Similarly to the previous case, the sub-mixture clustering (top) and the k-means results (bottom) are close to each other. The insignificant difference can be seen around the values 0 and 70 of the variable $y^3_{1;t}$.

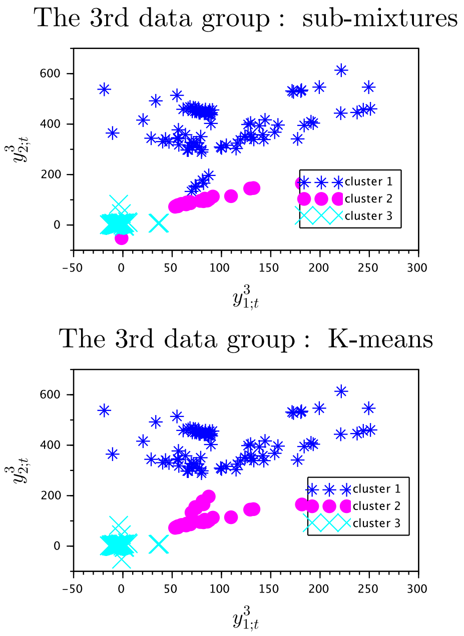

Fig. 3. The clustering with sub-mixtures (top) and the k-means (bottom) for the normal variables $y^3_{1;t}$ and $y^3_{2;t}$

The clusters were shown for each data group separately using the point estimates of the pointers of the corresponding sub-mixtures. In the case of plotting the variables from different data groups, the number of clusters grows because of the conditional models of the pointers. In this case, the cluster visualization is not suitable. However, the quality of the clustering was validated using the data prediction from the corresponding active components. To save space, the prediction is not shown here.

B. Discussion

The main aim of this study was the model-based clustering of the data with different distributions, where the data groups were described by sub-mixtures. It should be noticed that the clustering of the exponential and normal measurements was successfully validated by the comparison with the theoretical counterpart. However, the successful results were obtained so far for the non-overlapping exponential components. In the case of exponential components located about 0 the obtained clusters were mixed.

The potential application of the approach can be found in the fields of the multi-dimensional data analysis, where multivariate variables with different distributions measured on the multi-modal system should be modeled (e.g., transportation, fault detection, smart city, big data, etc.).

The limitation of the approach is the assumption of distributions with the re-computable statistics.

V. Conclusion

The paper proposed the algorithm of the recursive clustering with the model of sub-mixtures of a different type of distributions. The main contribution of the approach is to combine the general model from the active components of the used sub-mixtures. The proposed algorithm is expected to bring advantages of the more flexible "combined" model for the detection of the active components.

The application of the algorithm to the other types of distributions (for instance, uniform and categorical) is planned within the presented project. It can be beneficial for the case of dependent variables of the uniformly distributed data vector.

Acknowledgment

This paper was supported by the project GAČR GA15-03564S.

References

[1] M.J. Zaki, Jr. W. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, 2014.
[2] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, vol.31, 8 (2010), p. 651–666.
[3] C. Bouveyron, B. Hammer, T. Villmann, Recent developments in clustering algorithms. In: *Verleysen M*, ed. ESANN, 2012, p. 447–458.
[4] G. Malsiner-Walli, S. Frühwirth-Schnatter, B. Grün, Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and computing*, 26(1–2), 2016, p.303–324.
[5] A. O'Hagan, T.B. Murphy, I.C. Gormley, P.D. McNicholas, D. Karlis, Clustering with the multivariate normal inverse Gaussian distribution. *Computational Statistics & Data Analysis*, 93, 2016, p.18–30.
[6] Yusuf Abbakar Mohammed, Bidin Yatim, Suzilah Ismail, A Simulation Study of a Parametric Mixture Model of Three Different Distributions to Analyze Heterogeneous Survival Data. *Modern Applied Science*, vol. 7(7), 2013.
[7] F. Z. Doğru, O. Arslan, Robust Mixture Regression Using Mixture of Different Distributions. In *Recent Advances in Robust Statistics: Theory and Applications*, 2016, p. 57-79, Springer India.
[8] Runqing Yang, Xin Wang, Jian Li, Hongwen Deng, Bayesian robust analysis for genetic architecture of quantitative traits, *Bioinformatics*, (2009) 25 (8): 1033-1039.
[9] M.R. Gupta, Y. Chen, Theory and use of the EM method. In: *Foundations and Trends in Signal Processing*, vol. 4, 3, 2011, p. 223–296.
[10] M. Kárný, J. Kadlec, J., E.L Sutanto, Quasi-Bayes estimation applied to normal mixture. In: *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, eds. J. Rojíček, M. Valečková, M. Kárný, K. Warwick, CMP'98 /3./, Prague, CZ, 07.09.1998–09.09.1998, p. 77–82.
[11] V. Peterka, Bayesian system identification. In: *Trends and Progress in System Identification,* ed. P. Eykhoff, Oxford, Pergamon Press, 1981, p. 239–304.
[12] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer-Verlag London, 2006.
[13] E. Suzdaleva, I. Nagy, T. Mlynářová, Recursive Estimation of Mixtures of Exponential and Normal Distributions. In: *Proceedings of the 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Warsaw, Poland, September 24–26, 2015, p.137–142.
[14] I. Nagy, E. Suzdaleva, M. Kárný, T. Mlynářová, Bayesian estimation of dynamic finite mixtures. *Int. Journal of Adaptive Control and Signal Processing*, vol.25, 9 (2011), p. 765–787.
[15] L.Yang, H. Zhou and S.Yuan, "Bayes estimation of parameter of exponential distribution under a bounded loss function," *Research Journal of Mathematics and Statistics*, vol. 5, no. 4, pp. 28–31, 2013.
[16] E. Suzdaleva, I. Nagy, T. Mlynářová, Expert-based initialization of recursive mixture estimation. In: *Proceedings of the IEEE International conference Intelligent systems IS'16*, 2016, p. 308–315.