Recursive Clustering Hematological Data Using Mixture of Exponential Components

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

***Abstract*—The paper deals with the mixture-based clustering of anonymized data of patients with leukemia. The presented clustering algorithm is based on the recursive Bayesian mixture estimation for the case of exponential components and the data-dependent dynamic pointer model. The main contribution of the paper is the online performance of clustering, which allows us to actualize the statistics of components and the pointer model with each new measurement. Results of the application of the algorithm to the clustering of hematological data are demonstrated and compared with theoretical counterparts.**

***Index Terms*—mixture-based clustering, recursive mixture estimation, exponential components**

I. Introduction

The cluster analysis [1] is required in many application areas (e.g., bioinformatics, marketing, social fields, transportation, fault detection, big data, etc). Its primary task is to find the groups of data with similar characteristics in the data space. In the area of bioinformatics it is relevant for e.g., gene expression analysis, drug discovery, cancer-related research and many others [2]–[4].

There are many approaches to the task of clustering described in literature, e.g., hierarchical methods [5], [6], partitioning methods categorized among centroid-based methods such as the famous k-means [7]–[9] as well as k-medoids [10] and density-based methods such as e.g., DBSCAN [11], etc. The overview of clustering methods can be found in many sources, e.g., [9].

One of the approaches is the cluster analysis based on the use of mixture models [12], [13], where clusters in the data space are described by distributions of mixture components. This approach is used in the presented paper. The focus is on clustering with the help of the mixture models, which consist of components in the form of probability density functions (pdfs) describing individual regimes of working a considered system and a model of their switching. The mixture-based clustering starts from some initial (mostly resulting from the prior data analysis) location of components and performs a search for density clusters in the data space with the aim of fitting component models to the data.

Distributions for the description of the mixture components are chosen in dependence on the nature of data to be modeled. Studies dealing with the investigation of various distributions for the tasks of mixture-based clustering can be found in literature, for example, [14]–[16]. Normal distributions are the most frequently used components, e.g., [17], [18], etc. However, the assumption of normality is not always applicable and can bring limitations in data modeling.

This paper deals with clustering with the mixture of exponential components. The similar issue is considered in e.g., [19], [20]. However, unlike most algorithms, the presented clustering is performed in the online mode, which can be the key point for diagnostics systems, where fast evaluation of the current state is necessary taking into account each new data item. The clustering algorithm is based on the recursive Bayesian mixture estimation [21]–[24], where the main subtasks to be solved are (i) the parameter estimation for components and the switching model and (ii) the classification of data to the active component. The algebraic recursions are used for updating the pdf statistics without using numerical computations such as the EM algorithm [25]. In this area the paper [26] is already published for the case of a mixture of a single exponential and several normal components.

In this paper, the clustering algorithm is applied to the data set of anonymized hematological measurements of patients with leukemia including such variable as blasts, neutrophils, platelets, overall survival, gene expression and death. Selected results of the clustering experiments are demonstrated. Their results are compared with k-means clustering, which successfully validates the location and shapes of detected clusters.

The remainder of the paper is organized in the following way. Section II introduces models used in the paper and formulates the problem. Section III presents the general solution to the formulated problem of clustering and provides the algorithmic summary. Section IV is devoted to the application of the approach to hematological data and the initialization of the clustering algorithm. Section V demonstrates results of the

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

978-1-5090-6664-3/17/$31.00 ©2017 IEEE

experiments and discusses them in the discussion. Conclusions and open problems can be found in Section VI.

II. Problem Formulation

A. Models

Let's consider a multi-modal system, which at each discrete time instant $t = 1, 2, ...$ generates the continuous data vector y_t of the dimension N and the discrete data z_t with the set of its possible values $\{1, 2, \ldots, m_z\}$. It is assumed that the observed system can generate the data y_t in m_c working modes. The working modes of the observed system are described by m_c components, which comprise a mixture model.

1) Components: The mixture components have the form of the following pdfs

$$f(y_t|\Theta, c_t = i),\ i \in \{1, 2, \ldots, m_c\}, \tag{1}$$

where $\Theta = \{\Theta_i\}_{i=1}^{m_c}$ is a collection of unknown parameters of all components and c_t is a discrete random variable, which is called the pointer [21]. Its values point to the active component which describes the data generated by the system at time t, i.e., $c_t \in \{1, 2, \ldots, m_c\}$. Θ_i includes parameters of the i-th component according to its distribution, i.e., $f\left(y_t|\Theta, c_t = i\right) = f\left(y_t|\Theta_i\right)$ for $c_t = i$.

A distribution of the component (1) is chosen according to the nature of data and assumptions made about modeled variables. In this paper, the pdfs (1) are the exponential distributions

$$\left(\prod_{l=1}^{N}(a_l)_i\right)\exp\left\{-a_i' y_t\right\}, \tag{2}$$

i.e., here $a_i \equiv \Theta_i$ and $(a_l)_i > 0$ are the l-th entries of the N-dimensional vector a_i with $l = \{1, 2 \ldots, N\}$. Currently, the independence of entries of the vector y_t is assumed.

2) The Pointer Model: Switching the active components is described by the dynamic data-dependent model of the pointer c_t in the form of the following probability function (also denoted by pdf)

$$f\left(c_t = i|\alpha, c_{t-1} = j, z_t = k\right) = \tag{3}$$

	$c_t = 1$	$c_t = 2$	$\cdots$	$c_t = m_c$
$c_{t-1} = 1$	$(\alpha_{1\vert 1})_k$	$(\alpha_{2\vert 1})_k$	$\cdots$	$(\alpha_{m_c\vert 1})_k$
$c_{t-1} = 2$	$(\alpha_{1\vert 2})_k$		$\cdots$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$c_{t-1} = m_c$	$(\alpha_{1\vert m_c})_k$		$\cdots$	$(\alpha_{m_c\vert m_c})_k$

where the unknown parameter α is the $(m_c \times m_c)$-dimensional matrix, which exists for each value $k \in \{1, 2, \ldots, m_z\}$ of the discrete variable z_t. Its entries $(\alpha_{i|j})_k$ are non-negative probabilities of the pointer $c_t = i$ under condition that the previous pointer $c_{t-1} = j$ with $i, j \in \{1, 2, \ldots, m_c\}$ and the variable $z_t = k$.

B. Clustering Problem Specification

The main task of the recursive mixture-based clustering is to estimate online which component is active at each time t. It means that based on the data measured up to the time t, it is necessary to obtain the value of the pointer c_t and classify the actual data item to the active component indicated by the pointer.

With the introduced models, the clustering problem is comprised from the following sub-tasks:

- the estimation of the component parameters Θ,
- the estimation of the parameter α of the pointer model
- and the estimation of the value of the pointer c_t, which indicates the active component at time t.

III. Recursive Clustering with Exponential Components

Based on the Bayesian methodology on recursive estimation [21]–[24], the mixture estimation algorithm is obtained with the help of construction of the joint pdf of all of the unknown variables and application of the Bayes and chain rules, e.g., [23]. This enables us to use the algebraic recursive update of statistics of the used distributions.

Denoting the data collection available up to the time instant t as $D(t) = \{D_0, D_1, \ldots, D_t\}$, where D_0 denotes the prior knowledge and the data item D_t includes the pair $\{y_t, z_t\}$, the joint pdf of the unknown variables Θ, α and c_t has the following form

$$f(\Theta, c_t = i, c_{t-1} = j, \alpha|D(t))$$
$$\propto f(y_t, \Theta, c_t = i, z_t = k, c_{t-1} = j, \alpha|D(t-1))$$

obtained using the chain and Bayes rule and then decomposed

$$= \underbrace{f\left(y_t|\Theta, c_t = i\right)}_{(1)}\underbrace{f(\Theta|D(t-1))}_{prior\ pdf\ of\ \Theta}$$
$$\times \underbrace{f\left(c_t = i|\alpha, c_{t-1} = j, z_t = k\right)}_{(3)}\underbrace{f(\alpha|D(t-1))}_{prior\ pdf\ of\ \alpha}$$
$$\times \underbrace{f(c_{t-1} = j|D(t-1)),}_{prior\ pointer\ pdf} \tag{4}$$

$\forall i, j \in \{1, 2, \ldots, m_c\}$ and for $k \in \{1, 2, \ldots, m_z\}$. To obtain recursive formulas for the estimation of c_t, which is the main goal from a clustering point of view, it is necessary to marginalize (4) over the parameters Θ and α firstly and then over the values of the past pointer c_{t-1}.

The marginalization of (4) over parameters Θ provides the proximity (i.e., the closeness) of the current data item y_t to individual components at each time instant t [27]. It is evaluated similarly to the approximated likelihood [22] with the use of the point estimates of parameters. The proximity is the value of the component pdf obtained by substituting the point estimates of parameters of each i-th component from the previous time instant $t-1$ and the currently measured y_t into the pdf. According to [27], in the case of exponential

components, it is advantageous to use a rapidly decreasing function instead of the component pdf. In this paper, the polynomial in the form

$$\exp\{-(2\,\Delta_t)^5\} \tag{5}$$

is used as the proximity function, where Δ_t is the distance

$$\Delta_t = y_t - E[y_t], \tag{6}$$

with the expectation $E[y_t]$ computed for the exponential components as follows

$$\frac{1}{(\hat{a}_{l;t-1})_i} \tag{7}$$

using the point estimates $\hat{a}_{l;t-1}$ of the l-th entries of the parameter a from (2) from the previous time instant $t-1$, see, e.g., [28].

The proximities from all m_c components form the m_c-dimensional vector denoted by m.

Similarly, the integral of (4) over α provides the computation of its point estimate denoted by $(\hat{\alpha}_{t-1})_k$ with the help of the normalization of the previous-time statistics $(v_{t-1})_k$ of the conjugate prior Dirichlet pdf $f(\alpha|D(t-1))$ according to [22] for the actual value k of z_t.

After the marginalization of (4) over the parameters Θ and α the pdf $f(c_t = i, c_{t-1} = j|D(t))$ is obtained. It is joint for the actual pointer c_t and and past pointer c_{t-1}, which is also unknown.

To obtain the component weights expressing the probability of the component activity, the proximities are multiplied entry-wise by the previous-time point estimate of the parameter α and the prior m_c-dimensional weighting vector w_{t-1}, whose entries are the prior (initially chosen) pointer pdfs $(c_{t-1} = j|D(t-1))$, i.e.,

$$W_t \propto (w_{t-1}m') \,.*\, (\hat{\alpha}_{t-1})_k \tag{8}$$

where W_t denotes the square m_c-dimensional matrix comprised from pdfs $f(c_t = i, c_{t-1} = j|D(t))$ joint for c_t and c_{t-1}, and $.*$ is a "dot product" that multiplies the matrices entry by entry, see also [29].

The matrix W_t is normalized so that the overall sum of all its entries is equal to 1, and subsequently it is summed up over rows, which allows us to obtain the vector w_t with the updated component weights $w_{i;t}$ for all of the components. The maximum entry $w_{i;t}$ defines the currently active component, i.e., the point estimate of the pointer c_t at time t.

A. The Component Statistics Update

Using the obtained weights $w_{i;t}$ at time t, the component statistics are updated as follows. In (4), $f(\Theta|D(t-1))$ is the prior Gamma pdf proportional to

$$\left(\prod_{l=1}^{N}(a_l)_i\right)^{\kappa_{i;t}} \exp\left\{-a_i'(S_t)_i\right\}, \tag{9}$$

where the re-computable statistics (initially chosen) are updated in the following form

$$\kappa_{i;t} = \kappa_{i;t-1} + w_{i;t}, \tag{10}$$

$$(S_t)_i = (S_{t-1})_i + w_{i;t}y_t, \tag{11}$$

see, e.g., [28] or [26] based on [21]. The point estimates $\hat{a}_{i;t}$ of the parameters a_i are obtained as follows

$$(\hat{a}_{l;t})_i = \frac{\kappa_{i;t-1}}{(S_{l;t-1})_i}, \tag{12}$$

where $i \in \{1, 2, \ldots, m_c\}$ and $l = \{1, 2 \ldots, N\}$.

B. The Pointer Model Statistics Update

The statistics of the pointer model is re-computed similarly to the update of the individual categorical model [21], [22] using the joint weights $W_{i,j;t}$ [26], [29] from the matrix (8), where the row j corresponds to the value of c_{t-1} and the column i to the current pointer c_t

$$(v_{i|j;t})_k = (v_{i|j;t-1})_k + \delta(k; z_t)W_{j,i;t}, \tag{13}$$

and the Kronecker delta function $\delta(k; z_t) = 1$ for $z_t = k$ and 0 otherwise. The point estimate of the parameter α [22] of the pointer model (3) is then obtained by the normalization

$$(\hat{\alpha}_{i|j;t})_k = \frac{(v_{i|j;t})_k}{\sum_{l=1}^{m_c}(v_{l|j;t})_k}. \tag{14}$$

C. The Clustering Algorithm

Using the above derivations, the clustering algorithm can be summarized in the following form.

Initialization of the algorithm (for $t = 1$)

1) Set the number of components m_c (e.g., based on the prior data analysis or expert knowledge).
2) For all components, set the initial (expert-based or random) values of the statistics $\kappa_{i;0}$, which is a scalar and $(S_0)_i$, which is the N-dimensional vector.
3) For the pointer model, set the uniform initial values of the statistics $(v_0)_k$, which is the $(m_c \times m_c)$-dimensional matrix for each value $\forall k \in \{1, 2, \ldots, m_z\}$ of the variable z_t.
4) Using these initial statistics, compute the point estimates (12) and (14).
5) Set the initial m_c-dimensional weighting vector w_0 uniformly.

Online clustering (for $t = 2, 3, \ldots$)

1) Measure the data item y_t, z_t.
2) For all components, compute (7) and put it into (6) to obtain the distances for the polynomial proximity functions.
3) For all components, substitute (6) with y_t into (5) to obtain the proximities.
4) Using (14) for the actual value k of z_t, the obtained proximities and the weighting vector, compute W_t via (8).

5) Normalize W_t and sum up over rows to obtain the actual weighting vector w_t.
6) Declare the active component according the maximum entry of the vector w_t, which is the point estimate of the pointer c_t at time t and classify the data item y_t according to the pointer value.
7) For all components, using the actual weights, update the component statistics (10), (11) and the pointer statistics (13).
8) For all components, re-compute the point estimates (12) and (14) for the pointer model and go to Step 1 of the online clustering part of the algorithm.

IV. Application to Hematological Data

This section is devoted to the application of the presented algorithm to the clustering of anonymized hematological data. The primary aim of the algorithm application is the search for clusters in the data space which can indicate some specific groups of the leukemia patients characterized by similar features. The measured variables comprising the vector y_t are as follows:

- $y_{1;t}$ – blasts [%],
- $y_{2;t}$ – neutrophils [50^9/l],
- $y_{3;t}$ – platelets [10^9/l)],
- $y_{4;t}$ – overall survival [month],
- $y_{5;t}$ – gene expression [in a log scale].

The discrete variable z_t has two possible values: 1 denotes that the patient died and 2 – didn't die.

A. Choice of Component Distributions

The components with the exponential distribution can be chosen based on the analysis of histograms of the available prior measurements. Fig. 1 and Fig. 2 show histograms for prior data of blasts, neutrophils, platelets as well as overall survival. It can be clearly seen in these figures that the measured values in the histograms demonstrate the exponential nature of the variables. The histograms of blasts, platelets and overall survival more correspond to a mixture of exponential distributions, while the neutrophils not. However, mixture models are known to be universal approximations for modeling the data variables [30]. A histogram of gene expression plotted in Fig. 3 (top) is also close to the exponential course, although not so significantly as the rest of the variables from the vector y_t. The last histogram in Fig. 3 (bottom) belongs to the discrete variable z_t, which is used in the condition of the pointer model.

B. Initialization of Component Number

The number of components for the online clustering algorithm can be initialized using the visualization of the data pairs from a smaller set of prior data against each others. Fig. 4 (top) shows the prior values of platelets plotted against gene expression. Two clusters can be seen around values of 100 and 200 of the platelets and the third one can be guessed between values 500 and 600 [$10^9/l$]. Fig. 4 (bottom) plots the prior values of platelets against the values of blasts, where the similar location of three clusters can be visually guessed. This initial guess also corresponds to the frequencies in the histograms in Fig. 1 (top) and Fig. 2. It allows us to initialize the number of components as 3.

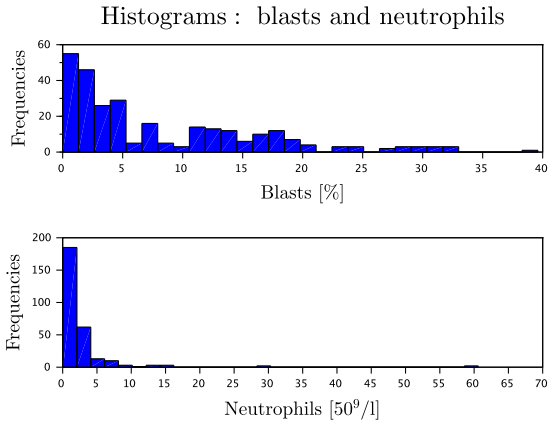

Fig. 1. Histograms of blasts and neutrophils

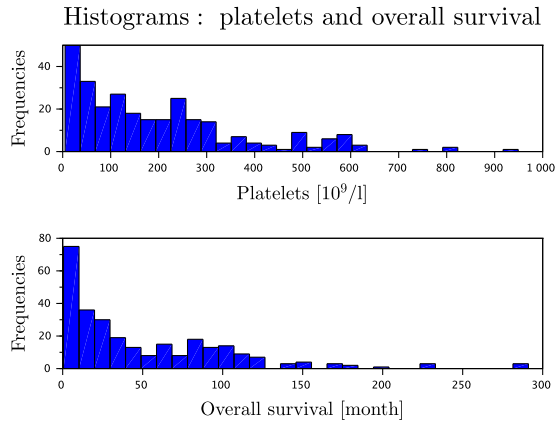

Fig. 2. Histograms of platelets and overall survival

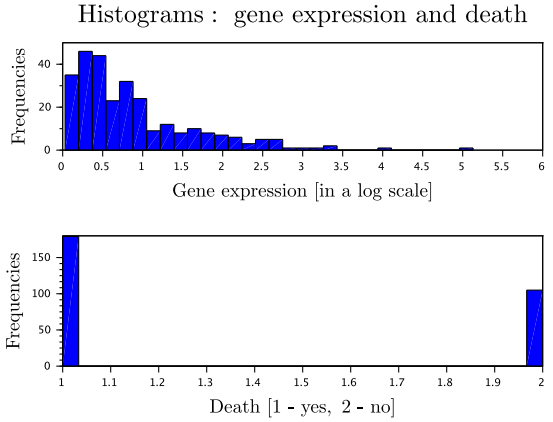

Fig. 3. Histograms of gene expression and death

It would be also appropriate to apply some classifier in the offline mode to the prior data set to pre-estimate the number of components. Here, only the prior data visual analysis is done for this purpose. The suitability of the model construction is validated during the online estimation via the monitoring of the component weights (it will be explained later).

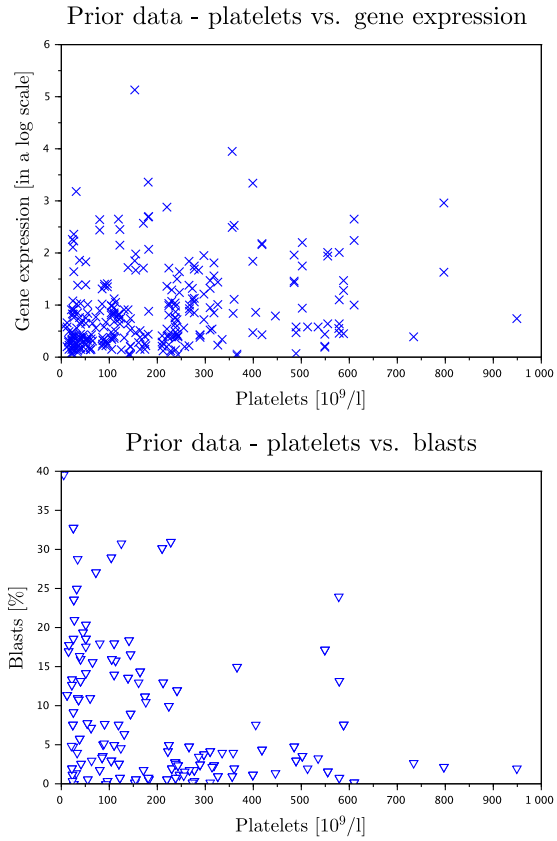

Fig. 4. Prior values of platelets plotted against gene expression (top) and blasts (bottom)

The rest of initial settings are chosen according to the initialization part of the algorithm in Section III-C. After the initialization, the online part of the clustering algorithm can be started. The obtained results are presented in the subsequent section.

V. Results

The whole capacity of the data set available for the research is 284 data items. This limited amount is explained by a small number of patients with rare forms of leukemia. This can complicate the search for clusters, e.g., some of the clusters can be formed by several data items only unlike the extensive data set, where they would have been much larger.

The results of the application of the presented algorithm were evaluated according to the following criteria:

- The regular activity of all of the components is observed by monitoring the weights of the components during the online part of the algorithm. It means that in the case of the properly initialized number of components, all of the weights are regularly approaching to 0 or 1 and none of of the components remains non-active for a longer time or is non-active at all.
- The evolution of the switching of the components is worth observing as well.
- Using real data, the pointer values indicating the active components cannot be compared with its true values, because the pointer variable is not measured. The location and shape of the clusters found can be validated by the comparison with theoretical counterparts, i.e., other successful clustering methods.

A. Weights of Components

The evolution of the component weights expressing the probability of the component activities is demonstrated in Fig. 5. The results for all of the three components during the online clustering are shown in the top, middle and bottom plots respectively. It can be seen that all of the weights are regularly close to the values of 1 and 0. It brings the unambiguous decision about the current component activity at each time t.

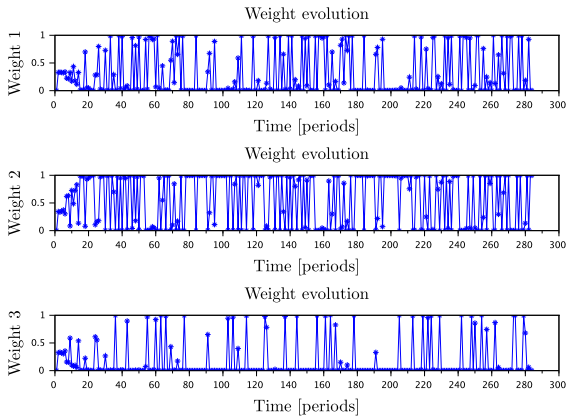

Fig. 5. The evolution of component weights during the online clustering

B. Switching the Components

Switching the components during the online part of the algorithm can be found in Fig. 6. The pointer estimates indicating the active components are placed in axis y. Three components show the regular activity during the estimation. In the case of a rare activity of any of the components it would be necessary to reduce the initialized number of components. However, here the results shown in Fig. 5 and Fig. 6 confirm that the model was initialized correctly.

C. Clusters

In the multi-dimensional data space such as in the case of the five-dimensional data vector y_t, the clusters can be shown by plotting the variables from the vector against each other. The most illustrative results from a clustering point of view

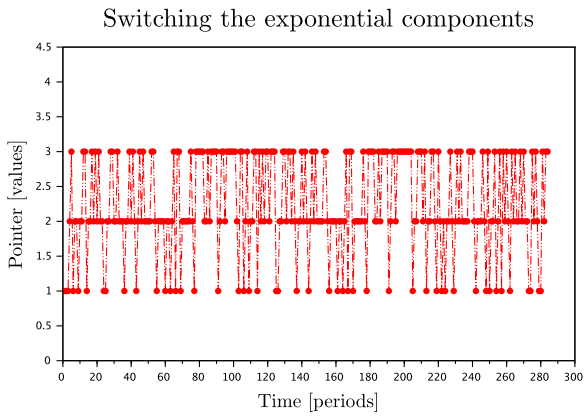

Fig. 6. The pointer estimation during the online clustering

were obtained for the variable $y_{3;t}$, which is the platelets. It creates the clusters of the clearly visible form with all of the modeled variables. Fig. 7 (top) demonstrates the detected clusters for the platelets and the blasts and validates them via comparison with the k-means method [9], see Fig. 7 (bottom). The location and shape of the clusters are compared. The insignificant difference can be seen in clusters 1 (denoted by '$\diamond$') and 2 (denoted by 'v') near the value of 15% of blasts. However, the rest of the measurements are distributed among the clusters similarly for both the compared algorithms, i.e., clusters 1, 2 and 3 have the very similar location and shapes.

Fig. 8 compares clusters of the platelets and the neutrophils obtained with the presented algorithm (top) and k-means (bottom). Again, with the insignificant difference in cluster 3 near the value of 100 of the platelets, the location and shapes of the clusters are almost identical. Two data items corresponding to the values of 30 and 60 of the neutrophils are outliers.

Fig. 9 and Fig. 10 provide the results of clustering obtained for the overall survival against the platelets and the gene expression against the platelets respectively. The location and shapes of the detected clusters are verified by the k-means method, i.e., the quality of the clustering is similar to the results in Fig. 7 and Fig. 8.

D. Discussion

The main goal of the presented study was to apply the mixture-based clustering algorithm to hematological data and find the clusters in the available data set. The clusters are expected to express groups existing in the patients' data which are characterized by similar, but non-specified features.

It can be said that the application of the algorithm was successful. Three clusters were detected and their location and shapes were confirmed by another well-known classifier such as k-means, which does not model the data vector, but iteratively looks for the data groups. This makes the comparison in some sense independent. However, it should be noticed that the presented algorithm performs the clustering online, i.e., it

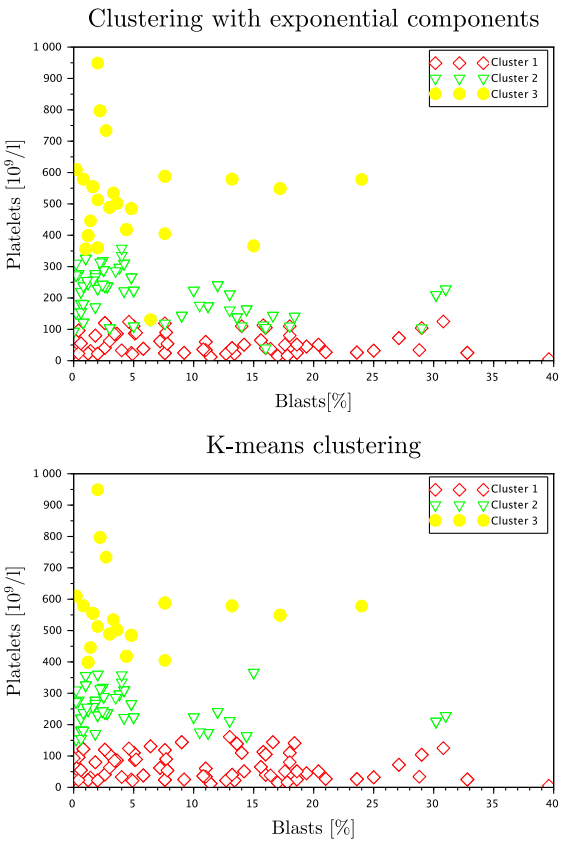

Fig. 7. Comparison of clusters of platelets and blasts

updates the estimates recursively by each new data item, which was measured. The k-means looks for clusters in the offline mode, which means it works with the whole data set and in order to make the clustering with the new measurement, it is necessary to run computations again with all of the data.

The clearly visible, practically non-overlapping clusters were detected for the data pair of platelets with the rest of the modeled variables, including the overall survival, which is the most interesting relationship in the data. The detected clusters of these two variables can be of a potential practical interest and should be investigated in further studies.

However, clusters detected for the rest of data pairs were much more overlapping (not shown here simply to save space) or even mixed. Their clustering with the help of k-means gave the same results, which means that in the available data set the groups of data with similar features should be searched only in the relationship of platelets with other variables.

The potential application of the presented algorithm can be found in the online diagnostics systems using patients' measurements for evaluating their state from a required point of view. However, the area of the potential application is not restricted by the biostatistical field and could be found elsewhere with relevant requirements for the online cluster analysis (for example, driver assistance systems in the

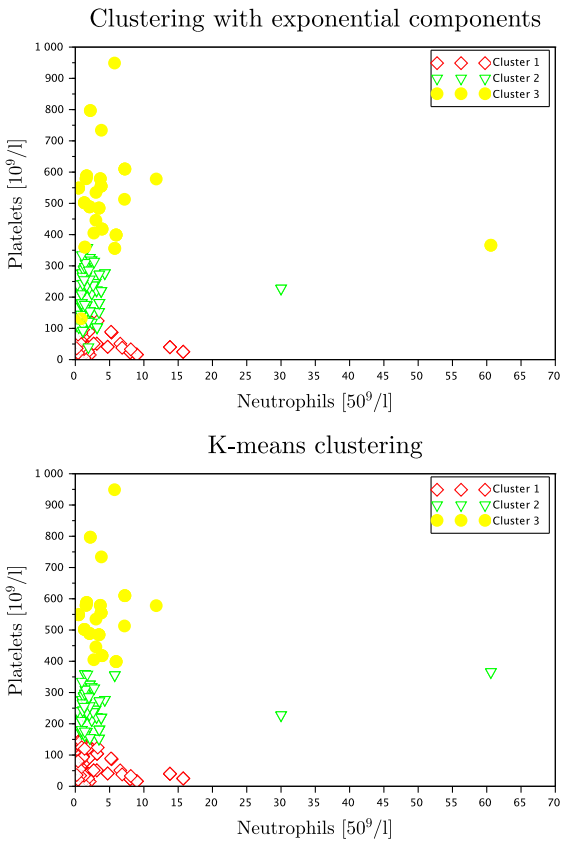

Fig. 8. Comparison of clusters of platelets and neutrophils

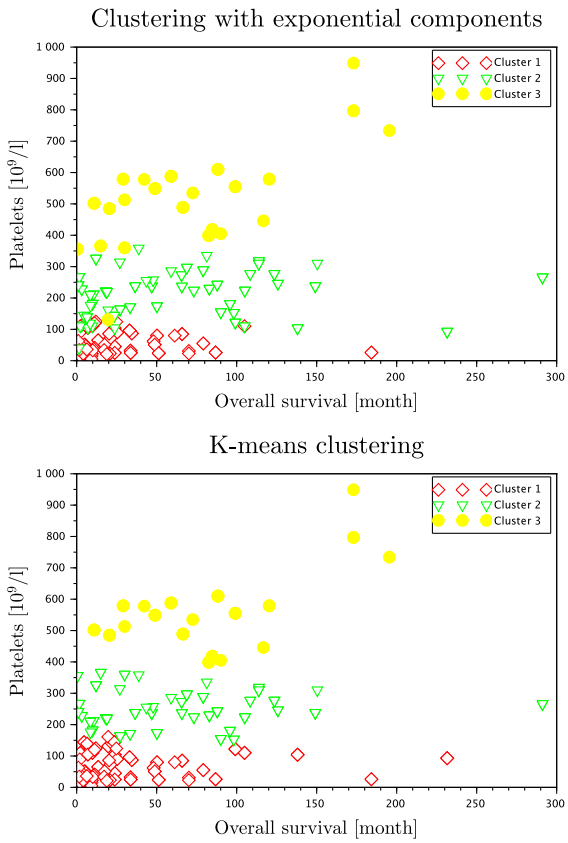

Fig. 9. Clusters of overall survival and platelets

transportation field, online state prediction and fault detection systems in the field of industrial processing plants, smart city and big data areas, etc.).

A limitation of the approach is the use of components described by distributions, which have the re-producible statistics to be used in the recursive update.

VI. Conclusion

The paper presented the online clustering algorithm based on the recursive Bayesian mixture estimation and its application to the anonymized data of leukemia patients. The main tasks solved in the paper are the parameter estimation of exponential components as well as the data-dependent dynamic pointer model and the online clustering of data according to the pointer value. The performed experiments with hematological data report the detection of clusters between platelets and the rest of the modeled variables.

The further issues planned to be considered within the presented study include: (i) modeling and prediction of the discrete variable z_t, which is here the death of the patient, (ii) prediction of the overall survival.

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] J.D. MacCuish, N.E. MacCuish. Clustering in bioinformatics and drug discovery. CRC Press, 2010.

[3] D. de Ridder, J. de Ridder, J. and M.J. Reinders. Pattern recognition in bioinformatics. Briefings in bioinformatics, 2013, 14(5), pp.633-647.

[4] J.D. MacCuish, N.E. MacCuish. Chemoinformatics applications of cluster analysis. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014, 4(1), pp.34-48.

[5] G. Sudipto, R. Rastogi, K. Shim. ROCK: A robust clustering algorithm for categorical attributes. *Inform. Systems*, 25(5), 2000, p. 345–366

[6] G. Karypis, E.-H. Han, V. Kumar. CHAMELEON: A hierarchical clustering algorithm using dynamic modeling. *Computer*, 32, 1999, p. 68–75.

[7] H. Steinhaus. Sur la division des corp matériels en parties *Bull. Acad. Polon. Sci.* (in French), 4 (12): 801-804, 1957.

[8] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematics. Statistics and Probability*, University of California Press, 1967, p. 281–297.

[9] Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, vol.31, 8 (2010), p. 651–666.

[10] H. Jiawei, M. Kamber, J. Pei. *Data Mining: Concepts and Techniques, Third Edition*, Morgan Kaufmann, 2011.

[11] M. Ester, H.-P. Kriegel, J. Sander, X. Xu. A density-based algorithm for discovering clusters in large spatial databases. *Proc. 1996 Int. Conf. Knowledge Discovery and Data Mining (KDD'96)*, Portland, OR, 1996, August, p. 226–231.

[12] A. Roy, A. Pal, U. Garain. JCLMM: A Finite Mixture Model for Clustering of Circular-Linear data and its application to Psoriatic Plaque Segmentation, *Pattern Recognition*, 2017, doi 10.1016/j.patcog.2016.12.016.

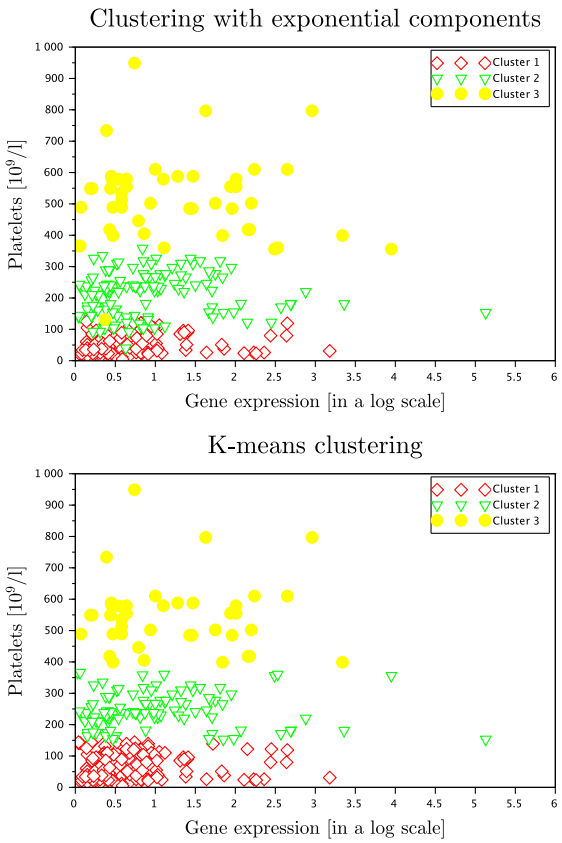

Fig. 10. Clusters of gene expression and platelets

[13] L.Scrucca. Genetic algorithms for subset selection in model-based clustering. *Unsupervised Learning Algorithms*, p. 55–70, 2016, Springer International Publishing.

[14] D. Fernández, R. Arnold, S. Pledger. Mixture-based clustering for the ordered stereotype model. *Computational Statistics & Data Analysis*, 2016, 93, p.46–75.

[15] R.P. Browne and P.D. McNicholas. A mixture of generalized hyperbolic distributions. *Canadian Journal of Statistics*, 2015, 43(2), p.176–198.

[16] K. Morris and P.D. McNicholas. Clustering, classification, discriminant analysis, and dimension reduction via generalized hyperbolic mixtures. *Computational Statistics & Data Analysis*, 2016, 97, p.133–150.

[17] G. Malsiner-Walli, S. Frühwirth-Schnatter, B. Grün. Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and computing*, 2016, 26(1–2), p.303–324.

[18] A. O'Hagan, T.B. Murphy, I.C. Gormley, P.D. McNicholas, D. Karlis, D. Clustering with the multivariate normal inverse Gaussian distribution. *Computational Statistics & Data Analysis*, 2016, 93, p.18–30.

[19] M. Tahir, M. Aslam, Z. Hussain, A. Khan. On finite 3-component mixture of exponential distributions: Properties and estimation. Cogent Mathematics, 2016, 3(1), 1275414.

[20] M. Tahir, M. Aslam, Z. Hussain, N. Abbas. On the finite mixture of exponential, Rayleigh and Burr Type-XII Distributions: Estimation of Parameters in Bayesian framework. Electronic Journal of Applied Statistical Analysis, 2017, 10(1), pp.271-293.

[21] 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.

[22] 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.

[23] V. Peterka, Bayesian system identification. In: *Trends and Progress in System Identification,* ed. P. Eykhoff, Oxford, Pergamon Press, 1981, p. 239–304.

[24] 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.

[25] M. R. Gupta and Y. Chen. Theory and use of the EM method. In: *Foundations and Trends in Signal Processing*, 2011, vol. 4, 3, p. 223–296.

[26] 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.

[27] I. Nagy, E. Suzdaleva, P. Pecherková. Comparison of Various Definitions of Proximity in Mixture Estimation. In: *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, Lisbon, Portugal, July, 29 – 31, 2016, p. 527–534.

[28] 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.

[29] I. Nagy, E. Suzdaleva. *Algorithms and Programs of Dynamic Mixture Estimation. Unified Approach to Different Types of Components*, SpringerBriefs in Statistics. Springer International Publishing, 2017.

[30] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan, 1994, New York.