Modeling of mixed data for Poisson prediction

Matej Petrouš
Faculty of Transportation Sciences
Czech Technical University
Na Florenci 25, 11000 Prague, Czech Republic
Department of Signal Processing
Czech Academy of Sciences
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
petromat@fd.cvut.cz

Evženie Uglickich
Department of Signal Processing
The Czech Academy of Sciences
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
https://orcid.org/0000-0003-1764-5924

***Abstract*—The paper deals with the task of modeling mixed continuous Gaussian and discrete Poisson data observed on a multimodal system. The proposed solution is based on recursive algorithms of Bayesian mixture estimation. The main contributions of the approach are: (i) the use of the discretized information of normal variables in the form of their clusters in order to keep the one-pass recursive estimation methodology and (ii) the prediction of the multimodal Poisson variable. Experiments with simulated and real data are presented.**

***Index Terms*—mixed data, Poisson distribution, mixture based clustering, passenger demand**

I. Introduction

The analysis of mixed continuous and discrete data is highly desired in many application fields, for example, medicine [1], [2], transportation control [3], economics and finance [4], etc. In the case of the multimodal behavior of an observed system, mixtures of distributions, see, e.g., [5], [6], can be suitable for the cluster analysis of the data. Various distributions can be used for the mixture model depending on the nature of modeled variables, e.g., normal [7]–[9], categorical [8], [10], uniform [11], exponential [12], etc.

This paper focuses on modeling mixed continuous and discrete data with a high number of possible nonnegative realizations, which can be required in specific application domains such as, e.g., queue theory [13], crash data [14], passenger demand [15], [16], etc. In this paper, a mixture of Gaussian distributions and Poisson models are used for the description of the data.

In the discussed area, a series of papers devoted to similar issues have been found. For example, the paper [17] proposed a Gaussian-Poisson mixture model to detect the specific crowd behavior. In the work [3], a Poisson inverse Gaussian model of Belgian census data was used to predict an origin-destination matrix. Poisson-gamma and Poisson log-normal models were also discussed. The study [14] presented a Poisson inverse Gaussian regression model for the crash data analysis in Texas. The paper [16] dealt with a passenger flow model in a transportation system with both continuous and discrete variables in a form of a Petri net.

Other nonlinear modeling approaches dealing with the problems close to the discussed field should be mentioned as well; for example, the neural network based feature extraction [18], the heuristic technique in the combination with Bayesian methods [19], the fuzzy logic approach [20], iterative learning control-based techniques [21], etc.

Despite the considerable amount of publications, none of them completely cover the task of the online prediction of the Poisson variable, which is the main emphasis of the presented paper. The previous work [22] used the Poisson regression for the prediction, which demonstrated a series of drawbacks regarding processing the explanatory data online. Here, the presented paper divides the considered task into the learning and testing phases, using continuous data available permanently and the discrete variable available only during the learning phase. The proposed solution is based on recursive algorithms of the mixture based clustering under Bayesian methodology [8], [9], [24], where the Poisson models are estimated on the detected clusters of the explanatory variable. This allows us to apply the learned models with the estimated parameters for finding clusters during the testing phase and predict the Poisson variable using them. The main contributions of the approach are: (i) the use of the discretized information of normal variables in the form of their clusters in order to avoid logistic regression and to keep the one-pass recursive estimation methodology and (ii) the prediction of the multimodal Poisson variable.

The practical application of the proposed solution is tailored to the task of modeling passenger demand in the public tram transportation, where the discrete variable represents the number of boarding passenger at a station and the continuous one represents the surroundings of the station.

The paper is organized in the following way. Section II formulates the problem and introduces the specified models with necessary assumptions. Section II-C describes the estimation of the models, while Section II-D is devoted to the prediction part. Experiments with simulated and real data from a tram network are demonstrated in Section III along with the

This work has been supported by the project SILENSE, project number ECSEL 737487 and MSMT 8A17006.

978-1-7281-7377-1/20/$31.00 ©2020 IEEE

discussion. Conclusions can be found in Section IV.

II. Modeling of Mixed Data for Poisson Prediction

A. Problem Formulation

Let us observe the multimodal system, which generates the continuous generally multidimensional variable x_t and discrete variable $y_t \in \{1, 2, \ldots, N\}$ at discrete time instants denoted by t. Realizations of the variable x_t can be measured online in real time, while the values of y_t are available for a limited period of time $t \leq T$. We suppose that the modes of the system behavior can be expressed by the unknown discrete variable $c_t \in \{1, 2, \ldots, C\}$, which is called the pointer [8], [24].

Generally, the relationship of the variables y_t and x_t can be presented as the probability function

$$f(y_t = j|x_t, \lambda), \tag{1}$$

while the observed variable x_t is described by the probability density function (denoted by pdf, along with the probability function also) existing for each value of c_t, i.e.,

$$f(x_t|c_t = i, \Theta), \tag{2}$$

while the model of the pointer c_t has the form of the pdf

$$f(c_t = i|\alpha), \tag{3}$$

where $j \in \{1, 2, \ldots, N\}$ denotes realizations of the variable y_t, $i \in \{1, 2, \ldots, C\}$ is a value of the pointer indicating the active system mode at time t, and λ, Θ, α are mutually independent unknown parameters of the three introduced pdfs respectively.

The generally formulated task is (i) to estimate the pointer c_t at time t, which means detecting clusters on the space of the data x_t labeled by the pointer value and (ii) predict the value of y_t for the time $t > T$.

B. Model Specification

The main idea of the proposed solution is to estimate the model (1) on clusters detected on the space of the data x_t. This allows us to avoid using the logistic regression with the following assumption: the knowledge on the clusters of x_t is supposed to enter (1) via the pointer estimates only, i.e., the model (1) takes the form

$$f(y_t = j|c_t = i, \lambda) \sim \mathcal{P}oi(\lambda_i), \tag{4}$$

which is (in this paper) the Poisson distribution with the parameter λ_i under condition that $c_t = i$.

The pdf (2) is here represented as the multivariate normal distribution existing $\forall i$ in the form

$$f(x_t|c_t = i, \Theta) \sim \mathcal{N}(\theta_i, r_i), \tag{5}$$

where $\Theta = \{\Theta_i\}_{i=1}^C = \{\theta_i, r_i\}_{i=1}^C$ is the set of parameters of all the normal pdfs and $\Theta_i = \{\theta_i, r_i\}$ are the expectation and the covariance matrix of the i-th pdf corresponding to $c_t = i$.

The pointer model (3) is specified in this paper as follows

$$f(c_t = i|\alpha) \equiv \tag{6}$$

c_t	$c_t = 1$	$c_t = 2$	$\ldots$	$c_t = C$
$f(c_t = i\|\alpha)$	α_1	α_2	$\ldots$	α_C

where α_i are probabilities of the value i of the pointer c_t.

The formulated task of the clustering and prediction requires to have all of the unknown variables estimated, which includes the parameters of the normal pdfs $\{\theta_i, r_i\}_{i=1}^C \equiv \{\Theta_i\}_{i=1}^C$, the parameters of the Poisson pdfs $\{\lambda_i\}_{i=1}^C$ as well as of the pointer model $\alpha = \{\alpha_i\}_{i=1}^C$ along with the pointer value c_t.

The solution is proposed in the phases of *learning* and *testing*, where the first of them covers the task of clustering and constructing the predictive model, while the second one is aimed at predicting y_t based on the data x_t only.

C. Learning Phase

1) Clustering: Here, this part of the solution first focuses on clustering the data x_t, which requires the estimation of c_t along with Θ and α. The derivations of the estimation algorithm are based on the Bayes and chain rules [25], and the Bayesian recursive estimation methodology [8], [24] applied in the following way:

$$f(\Theta, \alpha, c_t = i|x(t)) \propto f(x_t, \Theta, \alpha, c_t = i|x(t-1)) \tag{7}$$
$$= \underbrace{f(x_t|c_t = i, \Theta)}_{(5)} \underbrace{f(\Theta|x(t-1))}_{prior\ pdf\ of\ \Theta} \underbrace{f(c_t = i|\alpha)}_{(6)} \underbrace{f(\alpha|x(t-1))}_{prior\ pdf\ of\ \alpha},$$

where the denotation $x(t)$ means all the data x_t measured up to the time t. Using (7), the posterior pdf for the pointer estimation is obtained via integrals of the joint pdf of the unknown variables on their entire definite space

$$f(c_t = i|x(t)) = \int\int f(\Theta, \alpha, c_t = i|x(t)) d\alpha d\Theta. \tag{8}$$

According to [24], [25], the conjugate Gauss-inverse-Wishart (GiW) prior pdf is used for the normal models (5), while the Dirichlet prior pdf is taken for the pointer model (6), see [8].

2) Estimation of the Poisson Predictive Model: Due to the introduced assumptions in Section II-B, the Poisson model (4) is estimated for each i-th cluster indicated by $c_t = i$. The estimation is derived similarly using the corresponding joint pdf of the variables unknown in this phase, i.e.,

$$f(\lambda|c_t = i, y(t)) \propto f(y_t = j, \lambda|c_t = i, y(t-1)) \tag{9}$$
$$= f(y_t = j|\lambda, c_t = i, y(t-1)) f(\lambda|c_t = i, y(t-1)).$$

Here, the Gamma prior pdf can be used for the parameter estimation [26]. However, it is not necessary, and the maximum likelihood (ML) estimation of the Poisson distribution [27] can be used directly for the current value of c_t.

Finally, the learning phase is summarized as the following algorithm, which runs recursively at each time t after the initialization of the pdf statistics and the number of clusters.

Algorithm 1 (Learning phase):
For the time $t = 1, 2, \ldots, T$
1) Measure data x_t, y_t.
2) Substitute the data item x_t and previous-time parameter point estimates $\hat{\theta}_{i;t-1}$ and $\hat{r}_{i;t-1}$ into the i-th normal pdf (5)

$\forall i$, which gives the proximity of the data item to the clusters [9], [11]

$$m_i = \mathcal{N}_{x_t}(\hat{\theta}_{i;t-1}, \hat{r}_{i;t-1}). \quad (10)$$

3) Obtain the weighting vector $w_t = [w_{1;t}, \ldots, w_{C;t}]'$ multiplying the vector of all proximities m and the previous point estimate of the parameter α [8]

$$w_t = m. * \hat{\alpha}_{t-1}, \quad (11)$$

where $.*$ means multiplying by entries.
4) Update the statistics of the GiW pdfs for the normal models [24], [25]

$$(V_t)_i = (V_{t-1})_i + w_{i;t}[x_t \ 1]'[x_t \ 1], \quad (12)$$
$$(\kappa_t)_i = (\kappa_{t-1})_i + w_{i;t} \quad (13)$$

as well as of the Dirichlet pdf the pointer model [8]

$$\nu'_t = \nu'_{t-1} + w_t. \quad (14)$$

5) $\forall i$, re-compute the point estimates $\hat{\theta}_{i;t}, \hat{r}_{i;t}$ with the help of partitioning the information matrix $(V_t)_i$ according to [24], [25] and the point estimates of the pointer model $\hat{\alpha}_t$ by the normalization of the statistics ν_t [8].
6) Obtain the point estimate of c_t according to the index of the maximum entry of the weighting vector w_t, see, e.g., [24].
7) For the i-th cluster indicated by the point estimate of the pointer c_t, update the statistics of the Poisson pdf (4) that can be easily derived from the ML estimation, see, e.g., [27], [28]

$$(S_t)_i = (S_{t-1})_i + w_{i;t} y_t, \quad (15)$$
$$(K_t)_i = (K_{t-1})_i + w_{i;t}. \quad (16)$$

8) Re-compute the point estimate of the parameter λ_i of the Poisson pdf (4) [27], [28]

$$(\hat{\lambda}_t)_i = \frac{(S_t)_i}{(K_t)_i}. \quad (17)$$

In this way, the parameters of all of the models (4), (5), and (6) are estimated. The constructed predictive model (4) can be used for the prediction in the testing phase.

D. Testing Phase

The learned models (4) and (5) are used for the testing phase, where y_t is no longer measured for the time $t > T$ and should be predicted based on the permanently available data x_t. The data x_t can still be used for clustering with the estimated model (5). Recalling the assumptions from Section II-B, the model (4) with the substituted parameter point estimate provides the required prediction on the detected clusters. The prediction algorithm includes the following steps.

Algorithm 2 (Testing phase):
For the time $t = T+1, T+2, \ldots$
1) Measure data x_t.
2) $\forall i$, substitute the data item x_t along with $\theta_{i;T}, \hat{r}_{i;T}$ into the i-th normal estimated pdf and obtain the proximity of x_t to each cluster via (10).
3) Obtain the actual weighting vector according to (11) using this proximity and the last point estimate $\hat{\alpha}_T$.
4) Determine the cluster i corresponding to the data item according to the index of the maximum entry of the weighting vector, i.e., using the point estimate of c_t.
5) Use the last point estimate $(\hat{\lambda}_T)_i$ of the model (4) from (17) to obtain the predicted expectation of y_t.
6) Use the i-th model (4) with the substituted point estimate $(\hat{\lambda}_T)_i$ for generating the predicted values of y_t.

The validation of the approach in a free and open source environment Scilab (www.scilab.org) is discussed below.

III. Experiments

A. Experiments with Simulated Data

The aim of this part of the study was to verify whether the proposed clustering and prediction work correctly with the help of simulated values of the pointer c_t along with the variables x_t and y_t. Three possible realizations of the pointer, i.e., $C = 3$, have been used for the simulation. Using them, 3000 values of y_t and x_t were generated from the Poisson and normal distributions respectively. Here, the scalar explanatory variable x_t was used. The following figures demonstrate the results of testing the proposed approach.

The clusters detected on the data space of the variable x_t according to Algorithm 1 are shown in Figure 1, where the values of x_t and y_t are plotted against each other. Three clusters, which represent the discretized information about the data x_t to enter the Poisson model (4), can be seen.

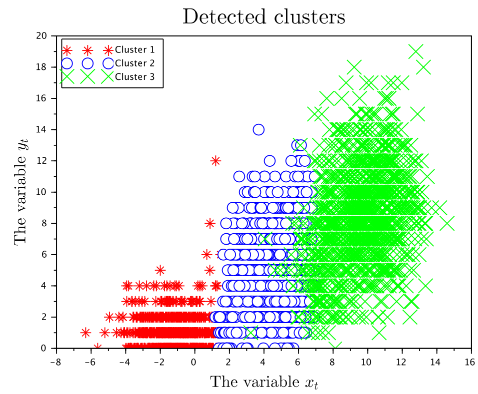

Fig. 1. Clusters of simulated data used for learning

Results of Algorithm 2, where the estimated models are used for predicting y_t under condition of the availability of x_t are provided in subsequent figures. Figure 2 compares histograms of the original values of y_t and their predictions. Three Poisson distributions can be seen in both the histograms. The predicted values correspond to the original ones.

The mean values of the prediction, which are the point estimates of the parameters λ_i are compared with simulated data used for the testing phase in Figure 3. This plot shows that with the exception of fragments around 70, 132, and 194 time periods, the testing data items were classified unambiguously.

The evolution of the pointer estimates during the testing phase is displayed in Figure 4. The three pointer realizations are regularly switching, which confirms the model fitting.

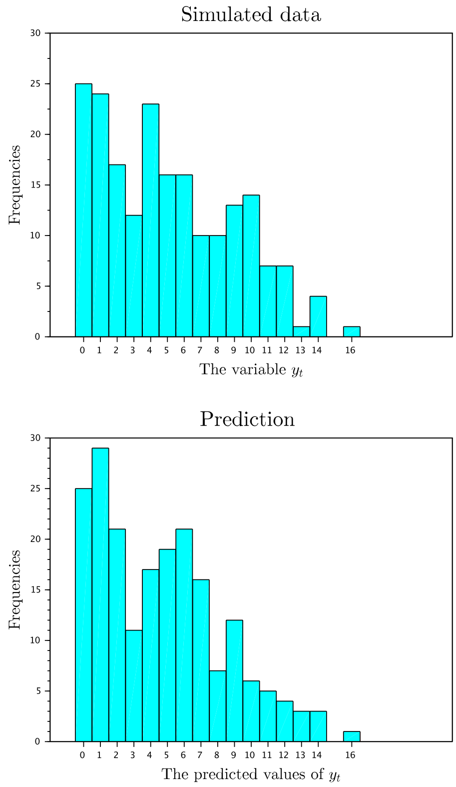

Fig. 2. The comparison of histograms of simulated data and predictions

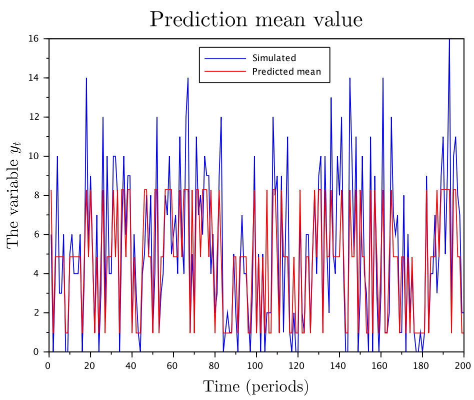

Fig. 3. The predicted mean value of the classified Poisson distributions

B. Experiments with Real Data

One of the application fields of the presented approach is the passenger demand prediction for the tram transportation. In this area, the critical task is to predict the number of boarding and/or disembarking passengers at each station, which can only be measured for a limited period of time until $t = T$, i.e., it is the variable y_t. Using the identical approach for modeling both of them, we suppose that they can be described by the Poisson distributions, which change at each station depending on the time of day [22]. Each station has the surroundings x_t, which comprise, e.g., tram delay, time between lines, number

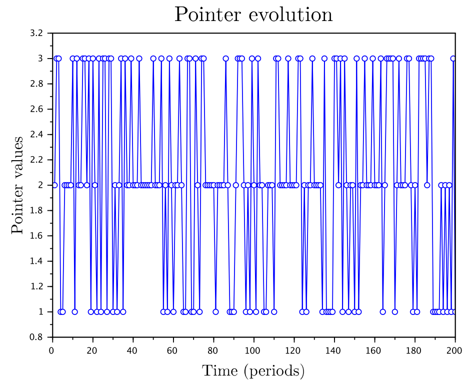

Fig. 4. The pointer evolution during the testing phase

of transfers, distance to a subway station, etc. The surroundings variable x_t is available in real time. Using the model of y_t depending on x_t, the passenger demand at a station of a tram line can be recursively calculated from its value at the previous station by subtracting the number of disembarking passengers and adding the number of boarding passengers at the current station [22]. However, working directly with the point estimates of y_t brings limitations primarily from the prediction accuracy point of view. To take into account the uncertainty influencing the passenger behavior, the mixture of pdfs represented by histograms can be used for the passenger demand prediction via, e.g., the interval analysis [29]. Thus, the predictive pdfs are the main focus of this part of the work.

Here, a data set from a real tram network was used. 2300 data items were used for learning the models and 200 randomly chosen values were taken for the testing phase. For this paper, values of the number of boarding passengers only were used as realizations of the variable y_t. Results of its prediction for a single station and a station belonging to a tram line of two stations are compared. For one station, the surroundings x_t are the delay of the tram at the station. For a line of two stations, the surroundings of the previous station naturally influence the behavior of passengers at the neighboring station, i.e., the vector $[x_{t;s},\ x_{t;s-1}]'$ should be considered, where the subscript s corresponds to a station. In the case of more stations, the surroundings of previous stations can be used with a forgetting factor. Three values of the pointer c_t initialized from the data set express the morning rush hour, midday calm traffic, and afternoon rush hour.

1) Boarding Number Prediction: Figure 5 compares histograms of the original and predicted number of boarding passengers at a single station and a tram line of two stations. The top plot shows the histogram of original data, while the middle and bottom plots demonstrate the histograms of the prediction. It can be seen that the mixture of pdfs in the Figure 5 (bottom) obtained for a station of a tram line is closer to the original one in the top plot than that for a single station in the middle plot. The mean values of the three distributions about 4, 16, and 22 passengers can be guessed in both these figures. Unlike them, the histogram in the middle plot is closer to a mixture of normal pdfs, which indicates that the information from the surroundings of the previous station is beneficial for the model and the stations should be described in the connection with other stations as a line.

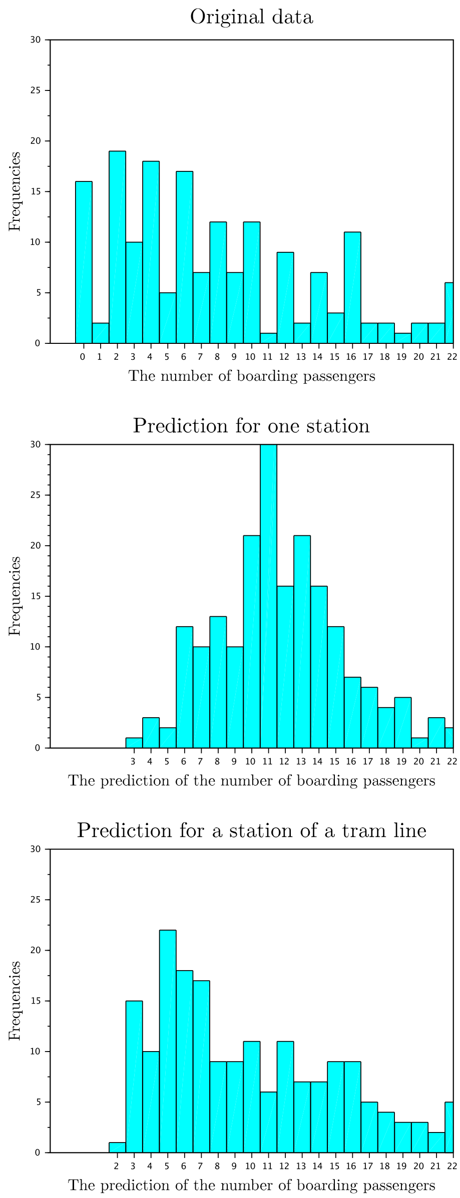

Fig. 5. Histograms of the original data (top) and their prediction for a single station (middle) and a tram line of two stations (bottom)

The data prediction comparison presented in Figure 6 confirms the advantages of modeling the station as a part of a tram line, where the top plot shows a worse quality of the prediction than the bottom plot. However, the outliers were not caught by either of them.

2) Clusters: The influence of the time of the day on the passenger behavior expressed by the pointer estimation can be presented by means of clusters detected on the data space of the surroundings. Figure 7 shows two-dimensional clusters of

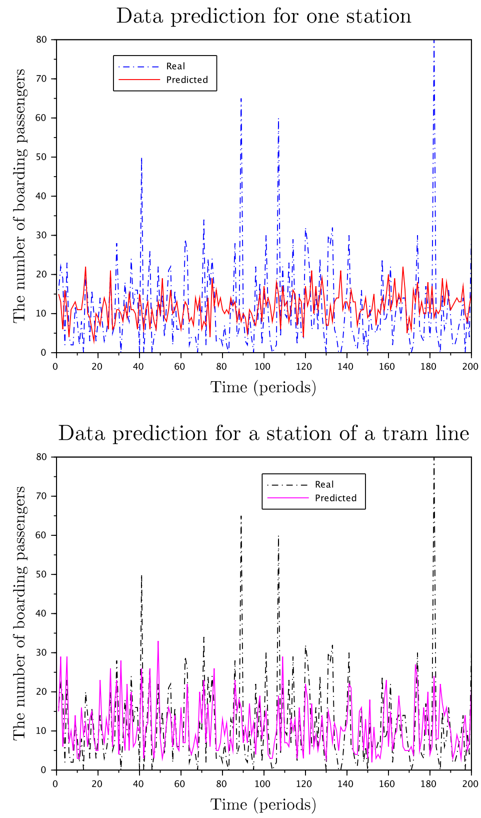

Fig. 6. The data prediction for one station (top) and a tram line (bottom)

the number of boarding passengers measured at a single station for the learning phase and its prediction at the previous station that can be also used in the surroundings vector. Three clusters are clearly outlined in the figure. The first cluster denoted by '$*$' corresponds to a calm time of a day, while the second denoted by 'o' and the third one with 'x' express the number of boardings for rush hours. The clustering at each time instant of the testing phase can be also demonstrated by plotting the pointer evolution, but it is not shown here to save space. All of the three pointer values are regularly switching, which means that the initialized number of clusters is adequate.

C. Discussion

To summarize the obtained results, the main aim of the study, i.e., modeling the discrete variable based on continuous explanatory variables using the information from their clustering, was successfully achieved. Due to the adopted methodology of the one-pass estimation, the recursive algorithms free of iterative computations have been used for clustering the explanatory data and predicting the Poisson variable y_t.

The data prediction should be obviously improved. However, the predictive pdfs represented by the histograms are suitable for the subsequent passenger demand prediction as it was mentioned in the beginning of Section III-B.

Due to two phases of the algorithm, the efficiency and precision evaluation can be treated twofold. As regards the precision

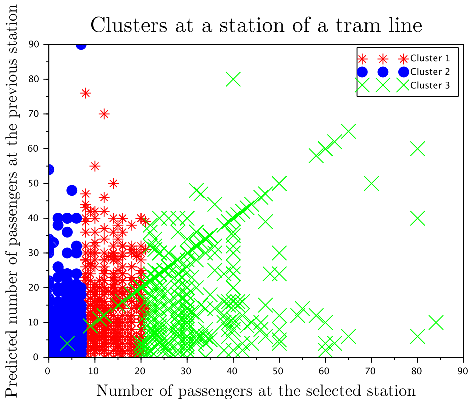

Fig. 7. Clusters at a station of a tram line

of the clustering phase, the clusters can be validated using the cluster validity indices or by verifying the results with the help of reliable theoretical counterparts. The prediction quality can be evaluated with the help of the prediction error using the available measurements.

IV. Conclusion

The paper described the approach of predicting the discrete variable using the discretized knowledge from clusters of continuous variables, which are obtained with the help of the recursive Bayesian mixture estimation theory. The learned models are used for the prediction based on continuous variables available in real time. The validation results obtained on simulations and real data look promising.

The potential application of the approach is not limited by the discussed specific case. Learning the dependent variable on clusters of the explanatory ones for the aim of its prediction can be considered for other suitable distributions under condition of the multimodality of the observed system. In the case of nonnegative data, one of the candidates is the Rayleigh distribution, which has a shape of the pdf close to the Poisson distribution. Their comparison will be demonstrated elsewhere.

References

[1] F. N. Z. Abidin and D. R. Westhead, "Flexible model-based clustering of mixed binary and continuous data: application to genetic regulation and cancer," Nucleic Acids Research, vol. 45, April 2017, pp. e53.

[2] F. Jiryaie, N. Withanage, B. Wu and A.R. de Leon, "Gaussian copula distributions for mixed data, with application in discrimination", Journal of Statistical Computation and Simulation, vol. 86, 2016, pp. 1643-1659.

[3] K. Perrakis, D. Karlis, M. Cools and D. Janssens, "Bayesian inference for transportation origin-destination matrices: the Poisson-inverse Gaussian and other Poisson mixtures," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 178, 2015, pp. 271-296.

[4] C. Foroni and M. Marcellino, "Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials," Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 178, 2015, pp. 57-82.

[5] P. D. McNicholas, Mixture model-based classification, CRC Press, 2017.

[6] S. Frühwirth-Schnatter. Finite Mixture and Markov Switching Models, 2nd Edition, Springer New York, 2006.

[7] G. Malsiner-Walli, S. Frühwirth-Schnatter and B. Grün, "Model-based clustering based on sparse finite Gaussian mixtures," Statistics and Computing, vol. 26, 2016, pp. 303-324.

[8] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma and L. Tesař, Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer-Verlag, London, 2006.

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

[10] M. Marbac, C. Biernacki and V. Vandewalle, "Model-based clustering for conditionally correlated categorical data," Journal of Classification, vol. 32, 2015, pp. 145-175.

[11] 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, pp. 527-534.

[12] I. Nagy, E. Suzdaleva and M. Petrouš, "Clustering with a model of sub-mixtures of different distributions," in Proceedings of IEEE 15th International Symposium on Intelligent Systems and Informatics SISY 2017, Subotica, p. 315-320.

[13] S. K. Kumaran, D. P. Dogra and P. P. Roy, "Queuing theory guided intelligent traffic scheduling through video analysis using Dirichlet process mixture model," Expert Systems with Applications, vol. 118, 2019, pp. 169-181.

[14] L. Zha, D. Lord and Y. Zou, "The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data," Journal of Transportation Safety & Security, vol. 8, 2016, pp. 18-35.

[15] A. S. Briand, E. Come, M. K. El Mahrsi and L. Oukhellou, "A mixture model clustering approach for temporal passenger pattern characterization in public transport," International Journal of Data Science and Analytics, vol. 1, 2016, pp. 37-50.

[16] S. Haar and S. Theissing, "Forecasting passenger loads in transportation networks," Electronic Notes in Theoretical Computer Science, vol. 327, 2016, pp.49-69.

[17] J. Yu, J. Gwak and M. Jeon, "Gaussian-Poisson mixture model for anomaly detection of crowd behaviour," in 2016 International Conference on Control, Automation and Information Sciences (ICCAIS), 2016, pp. 106-111.

[18] K. Sarma, "Neural network based feature extraction for Assamese character and numeral recognition", International Journal of Artificial Intelligence, vol. 2, no. S9, 2009, pp. 37-56.

[19] C. Pozna, R. Precup, J. K. Tar, I. Škrjanc and S. Preitl, "New results in modelling derived from Bayesian filtering", Knowledge-Based Systems, vol. 23, no. 2, 2010, pp. 182-194.

[20] R.P. Alvarez Gil, Z.C. Johanyák and T. Kovács, "Surrogate model based optimization of traffic lights cycles and green period ratios using microscopic simulation and fuzzy rule interpolation", International Journal of Artificial Intelligence, vol. 16, no. 1, 2018, pp. 20-40 .

[21] A.Albu, R. Precup and T. Teban, "Results and challenges of artificial neural networks used for decision-making in medical applications", Facta Universitatis, Series: Mechanical Engineering, vol. 17, no 4, 2019, pp. 285-308.

[22] M. Petrouš, E. Suzdaleva and I. Nagy, "Modeling of passenger demand using mixture of Poisson components," in Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019), Prague, 2019, p. 617-624.

[23] J. Roos, G. Gavin and S. Bonnevay, "A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data," Transportation Research Procedia, vol. 26, 2017, pp. 53-61.

[24] M. Kárný, J. Kadlec and E.L. Sutanto, "Quasi-Bayes estimation applied to normal mixture," in 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Rojíček, J., Valečková, M., Kárný, M., Warwick, K. (Eds.), 1998, September, Prague, Czech Republic, pp. 77-82.

[25] V. Peterka, Bayesian system identification, in Eykhoff, P. (Ed.), Trends and Progress in System Identification. Oxford, Pergamon Press, 1981, pp. 239-304.

[26] B. Puza, Bayesian Methods for Statistical Analysis, ANU Press, 2017.

[27] S. Yang and G. Berdine, "Poisson regression," The Southwest Respiratory and Critical Care Chronicles. 3(9), 2015, p. 61-64.

[28] N. L. Johnson, A. W. Kemp and S. Kotz, Univariate Discrete Distributions, 3rd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2005.

[29] H. Dawood, Theories of Interval Arithmetic: Mathematical Foundations and Applications, Saarbrücken: LAP LAMBERT Academic Publishing, 2011.