Markov-based Modeling of Respiratory Patterns for Prediction of Extubation Readiness in Preterm Newborns

Charles C Onu charles.onu@mail.mcgill.ca
School of Computer Science
McGill University

Abstract

In this work, we applied Markov chain modeling to gain insight into the respiratory patterns of extremely preterm newborns being considered for weaning from breathing support (*extubation*). We show that semi-Markov chains provide more robust modeling capability than Markov chains, especially when dwell time durations in states are very long and are not distributed exponentially. We also introduce the use of multi-chain models for exposing non-stationary behaviour across time segments. The developed models revealed interesting characteristics between newborns who were ready to be extubated and those who were not. They were then applied to the predictive task of determining extubation readiness.

1. Introduction

Infants who are born preterm (gestational age $\leq$ 28 weeks) usually require respiratory support. A breathing tube is inserted into the infant's trachea and connected to a mechanical ventilator - *intubation*. The newborn is monitored under this system until an attending clinician deems it ready for *extubation* (removal of the tube). Extubation is a critical decision for clinicians. Extubation done too early could lead to the suffocation and death (or disability) of the newborn. On the other hand, delayed extubation could result to the development of broncho-pulmonary dysplasia (BPD) or chronic lung disease.

In this work, we approach the task of predicting whether a patient would succeed or fail extubation based on respiratory patterns extracted as a time-series during a 5-minute period of spontaneous breathing trial (SBT) conducted upon the decision of an attending clinician to extubate. 5 discrete breathing patterns/states were considered: Pause, Asynchronous breathing, Movement artifact, Synchronous breathing and Unknown.

The respiratory pattern transitions were first modeled as a *Discrete-time Markov chain*. The extremely long dwell times observed, motivated the use of semi-Markov chain models which extracted interesting information about cross-pattern transitions of the two groups, as well as sojourn (or dwell) time behaviour. The predictive ability of the generative models were evaluated by maximising the posterior probability with respect to models conditioned on each class. The relatively low performance obtained suggests that newborns who succeed or fail extubation may not be distinguishable on the sole basis of their respiratory pattern sequences.

The sections of this report are broken down as follows: Section 2 provides more detail about the data used in the work. Section 3 describes related work. Section 4 describes our methodology. We present results in Section 5 and provide a discussion on the significance of our results in Section 6, as well as future directions of the work.

2. Cohort

Data from a total of 186 infants was available for this study. The data was collected from sites in Canada and the USA: Royal Victoria Hospital, Montreal Childrens Hospital, Glen site of the Montreal Childrens Hospital, Quebec, Detroit Medical Center, MI, and Women and Infants Hospital of Rhode Island, RI. Ethical approval was obtained from the boards at all institutions, and informed parental consent was obtained before recruiting each patient.

In the following sub-sections we provide brief details about the patient and data collected that is relevant to this present work. Readers interested in full data collection procedure for this study are referred to Precup et al. (2012) and Robles-Rubio et al. (2015).

2.1 Cohort Selection

Eligible infants were of gestational age $\leq$ 28 weeks, birth weight $\leq$ 1250g, and undergoing endo-tracheal tube with mechanical ventilation (ETT-MV) at time of recruitment. Infants were excluded if they had any major congenital anomalies such as heart disease, or were receiving any vasopressor or sedative drugs at the time of extubation.

2.2 Data Acquisition

Respiratory signals were measured using Respiratory Inductance Plethysmography (RIP) bands placed around the infant's ribcage and abdomen. Signals were acquired for a 5-minute period of spontaneous breathing trial prior to extubation at a sampling frequency of 1000Hz.

2.3 Respiratory Patterns

RIP signals were analyzed (to 50Hz) using Automated Unsupervised Respiratory Event Analysis (AUREA), which extracts sample-by-sample metrics of respiratory power, synchrony between the ribcage (RCG) and abdomen (ABD), and movement artifact (Robles-Rubio et al., 2011). AUREA uses k-means clustering to assign each sample to one of the following respiratory patterns:

- Pause (PAU): A cessation of breathing.
- Synchronous Breathing (SYB): RCG and ABD are in phase.
- Asynchronous Breathing (ASB): RCG and ABD are out of phase.
- Movement Artifact (MVT): Associated with infant moving or nurse handling.
- Unknown (UNK): Ambiguous patterns not belonging to any other pattern category.

AUREA provides repeatable results with no human intervention. An example of RIP signals and corresponding patterns assigned by AUREA to the different samples is shown in Fig. 1.

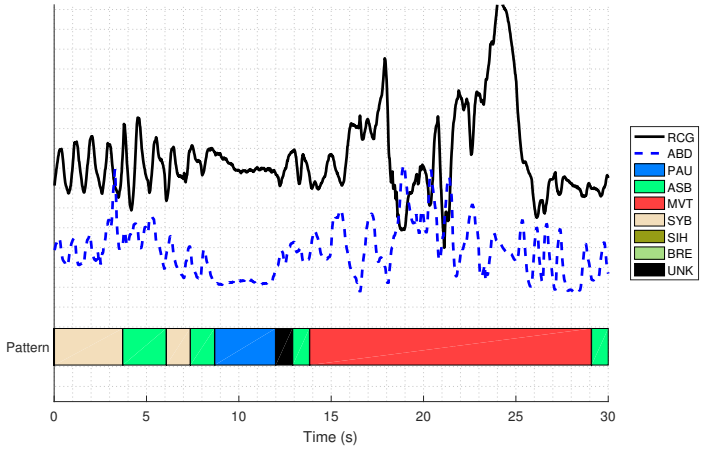

Figure 1: Example of a RCG and ABD signal segment and the corresponding respiratory patterns computed by AUREA.

3. Related Work

The use of Markov chain models for understanding sequence data and making predictions about outcomes has been popular across many domains. Ye et al. (2000) applied Markov chain modeling to detect anomalous activities in computer networks. In Gabriel and Neumann (1962) Markov chain modeling was successfully used to track rainfall occurence patterns. More recently Alinovi et al. (2017) applied semi-Markov chains to model and simulate breathing patterns of infant patients with respiratory disorders. Their results showed that as a statistical model, Markov chains provide a "reliable and realistic method to simulate breathing patterns and respiratory pauses/apneas".

4. Methods

We applied three main approaches to the task of predicting extubation readiness from the respiratory pattern data: the use of 1) Markov chains and 2) Semi-Markov chains 3) Multi-semi-Markov chain models. We explain the formulation of each in the following subsections.

4.1 Discrete-time Markov Chain

A Markov chain is a specific configuration of a probabilistic graphical model. It is represented as a sequence of nodes, each of which is conditioned only on the one before it (Figure 2). We consider a discrete-time Markov chain. That is, one for which there is a discrete set of states S from which each node x_t takes its value.

Given a time-series of observations $x_1, x_2, ..., x_T$ up to a time step T, the joint likelihood of the Markov chain representation is given as:

$$P(x) = P(x_1)\prod_{t=2}^{T} P(x_t|x_{t-1}) = \pi_{x_1}\prod_{t=2}^{T} A_{x_{t-1},x_t} \tag{1}$$

where π_{x_1} is the probability starting in state at x_1 and A_{x_{t-1},x_t} is the probability of transitioning from the state at x_{t-1} to the state at x_t. Similarly, we define the transition matrix A, an $|S|$ x $|S|$ matrix containing the transition probabilities for all possible combinations of states.

Fitting or learning the model of a Markov chain involves estimating these 2 parameters, π and A, from data. Given a set of data sequences, each corresponding to one example, the maximum likelihood estimates for the parameters are given as (Barbu and Limnios, 2008):

$$\pi_j = \frac{\#\ of\ sequences\ starting\ in\ state\ j}{\#\ of\ sequences} \quad \forall j \in S \tag{2}$$

$$A_{i,j} = \frac{\#\ of\ transitions\ from\ state\ i\ to\ j\ in\ all\ sequences}{\#\ of\ times\ state\ i\ was\ visited\ in\ all\ sequences} \quad \forall i,j \in S \tag{3}$$

In our problem, we explore the hypothesis that the respiratory patterns of infants who succeed extubation follows a different Markov chain from those who fail. Thus, conditional models are learned by fitting one Markov chain (π^s_j, $A^s_{i,j}$) to the respiratory pattern sequences of success patients, and another (π^f_j, $A^f_{i,j}$) to the sequences of failure patients. The prediction for a new example sequence x is made by assigning to it the class (success or failure) of maximum posterior probability. To do this, the likelihood of the sequence is computed with respect to the success and failure models using equation (1). The class which gives higher likelihood is selected.

4.2 Discrete-time Semi-Markov Chain

Semi-Markov models are different from standard Markov chains in that self-transitions, i.e. transitions from a state to itself, are collapsed. Instead, each state has a separate *dwell or sojourn time distribution*, which models the duration spent in the state until a transition out of the state occurs, coupled with a probability distribution of transitioning to *other* states. The latter results in a chain where every transition results in a state that differs from its predecessor.

Using this framework is useful for a number of reasons. First, Markov chains implicitly model dwell times as an exponential distribution (Barbu and Limnios, 2008) whereas semi-Markov chain models allow for the fitting of distributions of choice to the dwell times. Secondly, in data characterised by very long dwell times the transition matrix of the Markov chain would have very large diagonal elements, reducing the numerical resolution in off-diagonal elements(cross-state transitions). The use of semi-Markov chain addresses this.

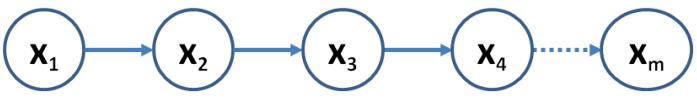

Figure 2: Example of a Markov chain

Thirdly, if the sampling rate of the data were to change, the Markov chain model could drastically change, whereas the semi-Markov chain model would not be significantly affected.

A semi-Markov chain model is thus characterised by 3 parameters:

- The start state distribution, $\pi_j, \quad \forall j \in S$
- The transition probabilities, $A_{i,j}, \quad \forall i, j \in S$ where $i \neq j$
- The conditional dwell time distribution, $F_{i,j}, \quad \forall i, j \in S$ where $i \neq j$

The joint likelihood (also called the approached likelihood) of a semi-Markov chain of observation is given as:

$$P(x) = \pi_{x_1} \prod_{t=2}^{T} A_{x_{t-1},x_t} F_{x_{t-1},x_t}(|x_t|) \tag{4}$$

where $F_{x_{t-1},x_t}(|x_t|)$ is defined as the probability of sojourning in the state at x_t for a given duration, after transitioning to x_t from x_{t-1} (Barbu and Limnios, 2008).

To learn the parameters π and A, equations 2 and 3 were employed. To model the dwell time distributions, the following process was repeated for the success and then for the failure patients: all dwell times in each state were collected across the patients in the population. A probability density function (PDF) estimate was obtained by plotting a histogram of the data normalised by number of observations in a bin and the bin width. Several known probability distributions were then fit to the data, selecting the one that minimised the Bayesian Information Criterion (BIC) (Schwarz et al., 1978).

As in the case of Markov chains, prediction for a new sequence x can be done by fitting conditional models (based on success and failure examples) to the data, estimating the likelihood of the new sequence given the models, and selecting the class with the maximum posterior probability for that sequence.

4.3 Multi-chain Models

The Markov and semi-Markov chains described in the previous 2 sections are first-order, stationary models because the transition probabilities considers only one previous state and are independent of time. The stationarity assumption may be inadequate for many real world scenarios. One popular way of handling this is the use of higher-order Markov chains which conditions on more than one previous state. Higher-order Markov chain models however have the down side that the number of transition probabilities to estimate scales exponentially with the order of the chain, consequently requiring more computation and more data to get good estimate. Multi-chain models approximate non-stationarity by fitting several stationary models to different segments of time. The probabilities to estimate only increases linearly with the number of chains.

4.4 Symmetric KL Divergence

The Kullback-Leibler (KL) divergence is a measure of how well one distribution Q is approximating another P. KL divergence is 0 if the two distributions are identical, and it goes to ∞ if Q does not have support for the whole domain of P. It is given as:

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \tag{5}$$

The KL-divergence is non-symmetric: $D_{KL}(P||Q) \neq D_{KL}(Q||P)$, which is not desirable in our application. Hence, we apply symmetrized KL-divergence to compare distributions over pattern transitions, defined as:

$$D_{KLS}(P||Q) = D_{KL}(P||Q) + D_{KL}(Q||P)$$

$$D_{KLS}(P||Q) = \sum_i (P(i) - Q(i)) log \frac{P(i)}{Q(i)} \tag{6}$$

4.5 Model Evaluation

Due to the small size of our dataset and high imbalance in class, it was necessary to give particular thought to appropriate evaluation methods for our predictive models.

First, the relatively small number of examples made k-Fold cross-validation, though a popular choice for model evaluation (Kohavi et al. (1995), Refaeilzadeh et al. (2008)), not suitable for our problem. In particular, k-Fold cross-validation is only a good estimator when $\frac{n}{k}$ is fairly larger than k (Wong, 2015) where n is the number of examples. For example, in our problem, using 10-fold cross validation would mean only about 18 examples in each fold, leading to high variance in estimates of the error. We thus applied Leave-one-out cross validation which is more suited for small datasets (Wong, 2015). Every example in the dataset was left out once, and the model fit to the remaining data.

Secondly, the proportion of success examples to failure was approximately 73:27. This high imbalance meant that optimising for classification accuracy could be misleading, and could result to the selection of sub-optimal (or even degenerate) models. For example a majority classifier (one that simply predicts the majority class) would have an accuracy of 73% on this dataset. To account for class imbalance, we optimised for *re-balanced accuracy* while also tracking the sensitivity (fraction of failure examples correctly identified) and specificity(fraction of success examples correctly identified).

Here we first show that the accuracy, *acc* measure is essentially a weighted sum of the sensitivity and specificity, where the weights are the fraction of failure and success examples, respectively in the dataset. We then extend that solution to obtain the re-balanced accuracy, acc_w, which is an equally-weighted sum of the sensitivity and specificity.

$$acc = \frac{tp + tn}{tp + fn + tn + fp} = \frac{tp + tn}{p + n} = \frac{tp}{p + n} + \frac{tn}{p + n}$$

$$acc = \frac{tp}{p}\left(\frac{p}{p + n}\right) + \frac{tn}{n}\left(\frac{n}{p + n}\right)$$

$$acc = sensitivity\left(\frac{p}{p + n}\right) + specificity\left(\frac{n}{p + n}\right)$$

Re-balanced accuracy:

$$acc_w = sensitivity * 0.5 + specificity * 0.5$$

Notation: tp - true positives, tn - true negatives, fp - false positives, fn - false negatives, p - number of positive examples, n - number of negative examples

5. Results

The objective of our experiments was to build predictive Markov models of the breathing patterns (or respiratory states) extracted during a 5 minute period of spontaneous breathing trial (SBT) prior to extubation of a preterm newborn. Data from a total of 186 patients (136 successes and 50 failures) were available. We show results of modeling experiments as well as of classification.

5.1 Analysis of Respiratory State Durations

As a starting point in modeling, we first obtained the total duration of time spent in each state as a fraction of the total SBT time for the patients in the 2 groups (Figure 3). The results showed that for at least half of the time both groups of patients experienced Synchronous breathing. The success patients spent slightly more time than the failures in synchronous breathing. The fraction of time spent in the other states were essentially equal (confirmed with error estimates obtained by bootstrapping), except for the Pause state, where failure patients had higher duration.

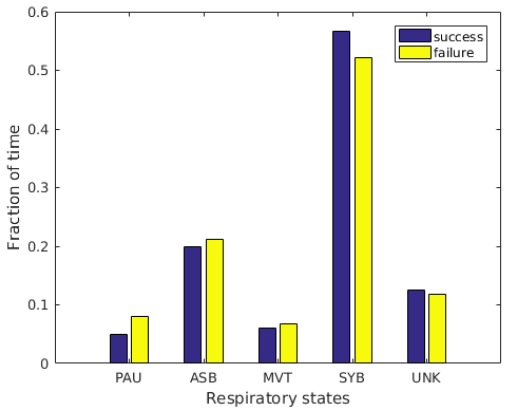

Figure 3: Fraction of time spent in each state for success and failure patients

5.2 Modeling of Dwell Times

The dwell times in each state were modeled using the process described in section 4.2. To avoid overfitting the data, marginal not conditional dwell times were fit. The probability density functions (PDF) were obtained using a bin width of 1 second. It was found that for every state, the distribution type which best fit the dwell times for the success population was the same as that the best fit for the failure population. However, the values for the parameters of the distributions varied (see Table 1). As an example, the PDF of the dwell times in Asynchronous breathing state are shown in Figure 4.

Table 1: Dwell (or sojourn) time distributions in each state for success and failure patients

	Success	**Failure**
Pause	Exponential $\mu = 2.51$	Exponential $\mu = 2.94$
Asynchrony	GeneralizedExtremeValue k=0.63, $\sigma = 1.30$, μ=1.85	GeneralizedExtremeValue k=0.63, $\sigma = 1.30$, μ=1.85
Movement Artifact	GeneralizedPareto k=-0.22, $\sigma = 3.62$	GeneralizedPareto k=-0.11, $\sigma = 3.31$
Synchrony	InverseGaussian μ =8.61, $\lambda = 3.61$	InverseGaussian μ =7.83, $\lambda = 3.41$
Unknown	GeneralizedPareto k=-0.22, $\sigma = 3.62$	GeneralizedPareto k=-0.11, $\sigma = 3.31$

5.3 Modeling Transitions

The transition probabilities were first modeled as a Markov chain. Due to the relatively long dwell times, all self-state transition probabilities (diagonal elements of the transition matrix) were at least 0.99 (not shown). This led to near-zeros values in the off-diagonal elements and consequently little differentiating factors between the transition matrices of the success and failure groups. The symmetric KL divergence between the transition distribution of the success and failure populations estimated was 0.0019.

The transition distribution was modeled as a semi-Markov chain i.e., focusing on only cross-states transitions. The obtained transition matrices for success and failure groups are shown in Table 2. It was observed that the transition behaviour between both groups differed mainly given a start state of Pause or Movement Artifact, both of which are shown graphically in Figures 5. The symmetric KL divergence between the 2 distributions was 0.27, a significant increase from the Markov chain case.

5.4 Multi-chain Modeling

The 5-minute SBT data was split into 2 halves for each patient. One semi-Markov chain was fit to the first 2.5minutes and another to the second. The symmetric KL divergence

Table 2: Transition matrices for a semi-Markov chain model fit to success (left) and failure(right) patients

	PAU	**ASB**	**MVT**	**SYB**	**UNK**
PAU	0	0.27	0.09	0.26	0.38
ASB	0.10	0	0.16	0.29	0.45
MVT	0.12	0.32	0.42	0.43	0.14
SYB	0.06	0.25	0.15	0	0.54
UNK	0.13	0.28	0.04	0.55	0

	PAU	**ASB**	**MVT**	**SYB**	**UNK**
PAU	0	0.28	0.06	0.39	0.28
ASB	0.12	0	0.21	0.28	0.40
MVT	0.17	0.41	0	0.32	0.09
SYB	0.14	0.21	0.14	0	0.52
UNK	0.15	0.30	0.03	0.52	0

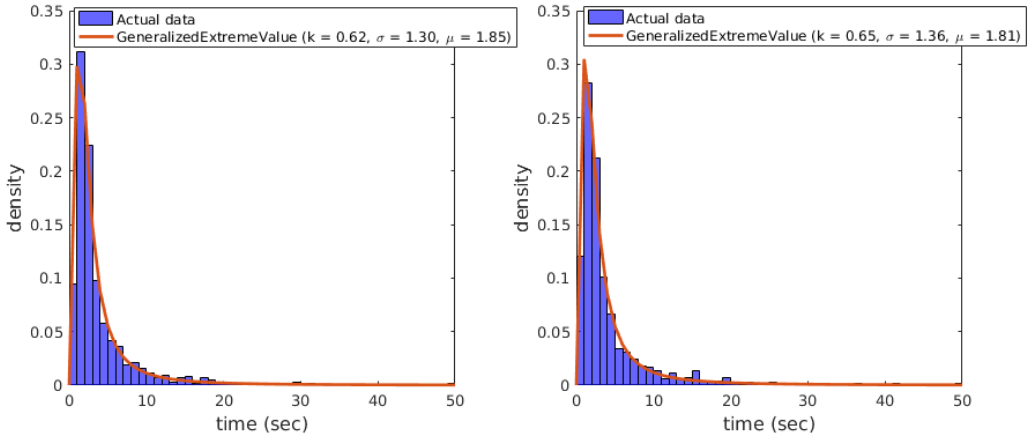

Figure 4: Probability density function (PDF) of the dwell time distributions in Asynchronous breathing state for success (left) and failure (right) populations

between the models for the success and failure groups was found to be 0.30 in the first half and 0.37 in the second half.

5.5 Classification with Markov and Semi-Markov Chain Models

We tested the predictive ability of the learned Markov and semi-Markov chain models. Prediction was done by maximising the posterior probability of a test example. Leave-one-out cross validation was applied. Performance of the models on sensitivity, specificity and re-balanced accuracy measures are shown in Table 3. Both models identify infants that succeed extubation at a rate of 73%. Whereas the semi-Markov model does slightly better on identifying infants that fail.

Table 3: Performance using Markov and Semi-Markov chain models for prediction

	Sensitivity	Specificity	Re-balanced accuracy
Markov chain model	0.46	0.73	0.60
Semi-Markov chain model	0.50	0.73	0.62

5.6 Classification with Multi-chain Models

The predictive capability of the multi-chain semi-Markov chain model was evaluated using leave-one-out cross validation. For each example, prediction was made by maximising the posterior probability given the 1st chain only, the 2nd chain only and with the likelihood from both chains summed. The results are summarised in Table 4.

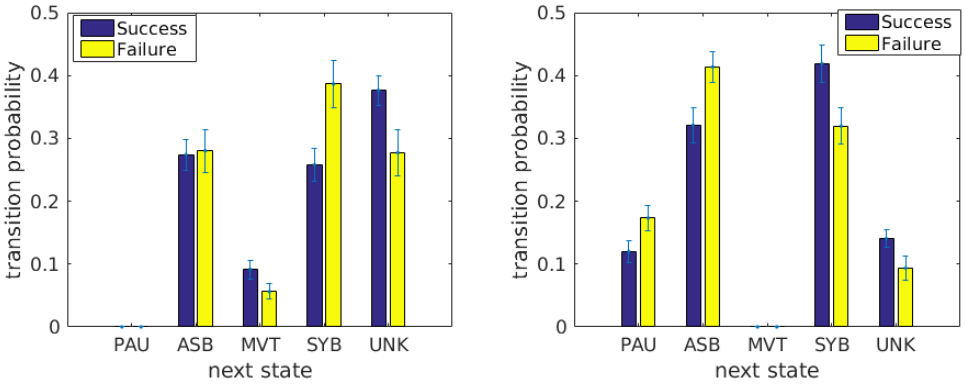

Figure 5: Transition probabilities given the start state of Pause (left) and a start state of Asynchronous breathing(right)

6. Discussion

In this work, we have applied Markov-based models to model respiratory pattern behaviour of preterm newborns in the period prior to extubation, and used same generative models to make predictions. We showed that Markov chain models blur cross-transition behaviour in the presence of long dwell times, and that semi-Markov chain models provide more robust mechanism for separately modeling dwell time and transitions between different states. The use of multi-semi-Markov chain models suggests that the transition behaviour of the two groups of patients diverges as they go from the first to second half of spontaneous breathing trial (SBT). This was seen in an increase in symmetric KL divergence from 0.30 to 0.37.

Classification based on these generative models were in general not impressive. The accuracy in detecting patients ready for extubation was 73% for the best model, while the accuracy in detecting patients who were not ready for extubation was 60% for the best model. In order to gain insight into the relatively low performance, the posterior probability of the training examples were obtained given the learned models. Displayed as a histogram in Figure 6, it was found that the likelihoods of both success and failure examples were distributed at roughly the same range of values. This explained the difficulty of the classifier in effectively separating the test examples.

Results from this work suggest that infants who succeed or fail exubation are not distinguishable on the sole basis of the respiratory pattern sequences they follow. The models were also tested using respiratory pattern data (of the same patients) which was manually labeled by clinical experts. The results were consistent with the above. More variables such

Table 4: Performance using multi-semi-Markov chain models for prediction

	Sensitivity	Specificity	Re-balanced accuracy
1st chain	0.40	0.69	0.55
2nd chain	0.60	0.64	0.62
Both chains	0.52	0.71	0.62

as metrics of cardiorespiratory variability from which the patterns were originally derived may need to be considered to develop effective models for this task.

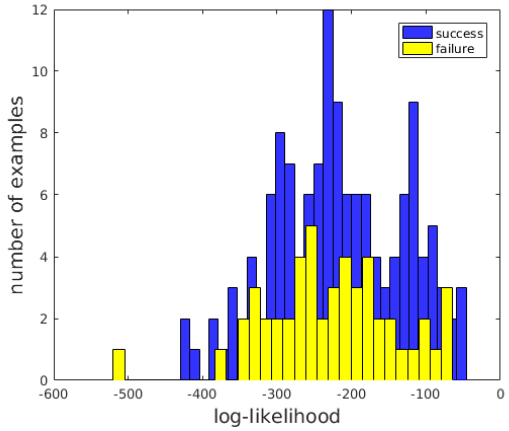

Figure 6: Distribution of posterior likelihoods on training data using learned semi-Markov chain model

References

Davide Alinovi, Gianluigi Ferrari, Francesco Pisani, and Riccardo Raheli. Markov chain modeling and simulation of breathing patterns. *Biomedical Signal Processing and Control*, 33:245 – 254, 2017. ISSN 1746-8094. doi: http://doi.org/10.1016/j.bspc.2016.12.002.

Vlad Barbu and Nikolaos Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis*. Springer Publishing Company, Incorporated, 1 edition, 2008. ISBN 0387731717, 9780387731711.

KR Gabriel and J Neumann. A markov chain model for daily rainfall occurrence at tel aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375):90–95, 1962.

Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA, 1995.

D. Precup, C. A. Robles-Rubio, K. A. Brown, L. Kanbar, J. Kaczmarek, S. Chawla, G. M. Sant'Anna, and R. E. Kearney. Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5630–5633, Aug 2012. doi: 10.1109/EMBC.2012.6347271.

P Refaeilzadeh, L Tang, and H Liu. Cross-validation (2008). *URL: http://www. public. asu. edu/˜ ltang9/papers/ency-cross-validation. pdf*, 2008.

C. A. Robles-Rubio, K. A. Brown, and R. E. Kearney. Automated unsupervised respiratory event analysis. In *2011 Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society, pages 3201–3204, Aug 2011. doi: 10.1109/IEMBS.2011.6090871.

CA Robles-Rubio, J Kaczmarek, S Chawla, L Kovacs, KA Brown, RE Kearney, and GM Sant Anna. Automated analysis of respiratory behavior in extremely preterm infants and extubation readiness. *Pediatric Pulmonology*, 50(5):479–486, 2015. ISSN 1099-0496. doi: 10.1002/ppul.23151. URL `http://dx.doi.org/10.1002/ppul.23151`.

Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464, 1978.

Tzu-Tsung Wong. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9):2839–2846, 2015.

Nong Ye et al. A markov chain model of temporal behavior for anomaly detection. In *Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop*, volume 166, page 169. West Point, NY, 2000.