



**HAL**  
open science

# Learning from partially labeled sequences for behavioral signal annotation

Anna Aniszewska-Stepień, Romain Hérault, Guillaume Hacques, Ludovic Seifert, Gilles Gasso

► **To cite this version:**

Anna Aniszewska-Stepień, Romain Hérault, Guillaume Hacques, Ludovic Seifert, Gilles Gasso. Learning from partially labeled sequences for behavioral signal annotation. International Workshop on Machine Learning and Data Mining for Sports Analytics, Sep 2020, Ghent, Belgium. pp.126-139, 10.1007/978-3-030-64912-8\_11 . hal-04465133

**HAL Id: hal-04465133**

**<https://hal.science/hal-04465133>**

Submitted on 19 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Learning from partially labeled sequences for behavioral signal annotation

Anna Aniszewska-Stepień<sup>1,2</sup>, Romain Hérault<sup>1</sup>, Guillaume Hacques<sup>2</sup>, Ludovic Seifert<sup>2</sup>, and Gilles Gasso<sup>1</sup>

<sup>1</sup> INSA Rouen Normandy, France

<sup>2</sup> Rouen University Normandy, France

**Abstract.** Herewith, we present a learning procedure that allows to deal with a partially labeled sequence dataset, i.e. when each sequence in the train dataset may contain labeled as well as unlabeled chunks. In our application case, this occurs when motor activity has been manually annotated (due to the recognition based on the video recording) and independently registered by the measuring system of high precision (touch sensors): human annotation misses some events that have been captured by the sensors. In the general setting, we aim at predicting the labels for a new fully unlabeled movement sequence, while the training has been performed on the partially labeled dataset. For this purpose we propose to use classical sequence model (hidden Markov model) that is furnished with a constrained Viterbi algorithm, which gives us a quick access to the *hard* approximation of the correct labeling sequences. We demonstrate, that this simple modification that constrained Viterbi provide, allows the HMM model to be trained on sparse data, and overall results in surprisingly high log-likelihood and accuracy level in annotating the partially labeled behavioral sequences in climbing. The same time we show the way to access correct labeling of the unannotated signal that can be helpful in various sport science studies for movement pattern sequential prediction.

**Keywords:** partially labeled data · hidden Markov model · constrained Viterbi algorithm · behavioral signal labeling · climbing pattern discovery

## 1 Introduction

In the real world, we very rarely possess the full information about the environment. For a machine learning prediction task, in the ideal setting, we shall dispose reliable and fully described data to train the model for pattern prediction in the newly observed data set; whereas this condition is hardly ever met in practice. Even for the experimental training data set, there can be missing information due to the errors or gaps in annotation, especially if the large information set is labeled by human and/or the labels are not trivially accessed. Depending on the type of data (discrete or continuous) and the suspected distributions of labels, there can be different approaches to solve this problem. In this type of tasks (e.g.

behavioral signal annotation, entity recognition, protein structure recognition), we want to label the observation sequences with a respective sequences that consist of the finite set of labels. Furthermore, the provided training sequences have some labels missing, while the missing ratio vary from sample to sample. Thus, all our training sequences may be partially labeled, but we still need to build a model that fully annotates the validation/test unlabeled sequences.

*Proposed approach* In the following article we will study a generative model that has been usually used in the context of sequential learning [8] adjusted to learn on sparsely labeled sequences. Namely, we will construct the procedure for the hidden Markov model to be learned from the sequences with voids in annotation, thanks to the Viterbi algorithm, that is modified in a way to find the optimal path passing by few constrained, known states. The proposed algorithm learns the distribution of the states with hidden Markov model and uses constrained Viterbi algorithm to predict the missing labels in the sequences; iteration over these two steps until convergence, emerges the stable sequence labeling close to the true one, which has been evaluated herewith in terms of log-likelihood and similarity (accuracy) ratio. We will show, how this relatively simple approach, resolves complex task, that is frequently encountered when we dispose only very little knowledge about the observation states. Our analysis aims at proving the efficiency of the proposed approach within the task of labeling the observation behavioral signal for which we never access the full information about the proper labels (not even for the training data sample). This situation is potentially frequently present in sport, when the access to the whole range of annotation is difficult and expensive (in terms of time and expert knowledge), while we need to draw generalized knowledge about the patterns present in large data sets, which as in our case, are registered time series. The motor task that produced the signal of our interest is the relative climber’s position on the wall in the moment of touching the hold. The state (label) that we would like to find, is the kind of limb used for the support. The so obtained designation is crucial to find climbers profiles in the skill acquisition patterns discovery [3]. We suppose however, that the proposed approach in the same way may be applied to design the patterns of subjects’ strategy during any other motor task.

*Article structure* The paper is organized as follows. After introducing the state-of-the-art (Section 2), we illustrate the context in which the model has been tested and we provide the characteristics of the model (Section 3). Then, in Section 4, we detail the model and the adjustments for partial learning. The experiments and evaluation methods for the results are defined in Section 5. Next, we specify the type of evaluation used and finally we interpret the results and conclude (Section 6).

## 2 Related work

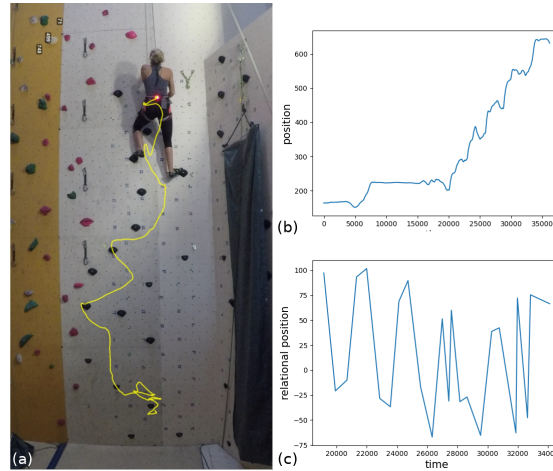
Research works using both labelled and unlabeled data for training have mostly considered the classical case of semi-supervised learning in which each sequence

example is either fully labeled or fully unlabeled [5]. The techniques that are used for semi-supervised sequential learning (within the broader area of structural learning) are multiple and deeply studied (mostly in the framework of NLP) over a number of recent years (e.g. deep learning models with pre-training word vectors [14] or self-training [6], [7], co-training [2], conditional random fields [19] and many other). We would like to however insist in the distinction of the above mentioned situation from the one we study in the present article: we here address observable sequences which are incompletely annotated, in which the presence of gaps and voids within each sequence reduce the labelling ratio (that is 1 for fully labelled sequence). Notwithstanding thus obtained *partially* labeled data sets are omnipresent, the situation of scarce access to the experimental data for sequential learning, has not yet been extensively explored in the literature.

One known discriminative approach to address this challenge has been proposed by Fernandes and Brefeld [9]. They used structured perceptron furnished with Viterbi algorithm to cope with the non-labeled parts of the sequences. This allowed them to train the perceptron on only the labelled parts of the sequences. Thus obtained model (simple transductive loss-augmented perceptron, STLAP), in spite of its simplicity, results in proper annotation for the initial labeling ratio larger than 0.5, which is comparable with standard supervised or semi-supervised methods. Also, in [10] the hidden Markov model with Baum-Welch algorithm has been compared by the same authors to the above mentioned STLAP in the partial training framework (see also [18]). The distinction with our approach however is that in our contribution we apply constrained Viterbi to even simpler model (hidden Markov model), for which we claim very good label annotation level. A separate branch of research employs conditional random fields, for example in Li *et al.* [13] different learning models are trained and subsequently combined for large scale sequential learning by adapting the idea of ensemble training. The result obtained on some types of partially labeled data were comparable to basic conditional random field approach, however the concept of many learning models may not be applicable to some type of data, especially when the transition between the tokens in the sequence is crucial as in our case. Conditional random fields combined with deep learning has been also recently studied in [22] in order to learn with annotation of various degrees (unsupervised, fully supervised and partially labeled sequences) in a unified way. Surprisingly, almost all the above mentioned papers claimed that the outcomes received for the partially labeled training data result in comparable performance or even outperform the ones received for the fully labeled data set (as they may avoid overfitting). Even though we do not study this property in the present article, some detailed studies of the phenomenon have been presented in [15] and [21]. Due to this fact and facing costly annotations of partially labeled data, there is a growing number of publications in the field, as for example the recent experiments in combining contradictory partial annotations from different datasets conducted by Huang *et al.* [12]. Additionally, and given that the above models have been tested the data only empirically, the required statistical consistency justification for the published discriminative models used for structural learning has been provided by Antoniuk *et al.* in [1].

### 3 Application context

The partially labeled learning has been studied mostly for the linguistic tasks (entity recognition, natural language processing) or biological sequence analysis (protein structure, DNA sequences). Our application in sport science requires to work on the behavioral signal that is difficult to be systematically processed due to its innate complexity and its spatio-temporal continuous nature. For a sport science application later purpose (climbers' patterns discovery, as in [3]), we study the following case. The climbing data (mass center position combined with handhold touch time of the climber while moving up the artificial wall, Fig. 1(a)) is recorded by a camera (for the mass center) and, separately, by the wall sensors (for hold touch time). Hold touches that are visible on the video recording have been manually annotated by which kind of foot or hand (left hand, right hand, left foot, right foot) performs the action. Naturally, so obtained labeling contains a lot of voids as part of the events recorded by the hold sensors are not present or hidden in the video used by human for annotating. Meanwhile, the manual annotation has been rendered only for a subset of the recordings.

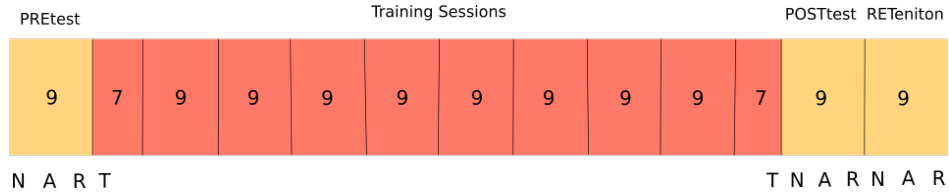


**Fig. 1.** (a) Experimental setting for climbing of the artificial wall (the mass-center trajectory of the climber is marked with the yellow line). The signal is composed by the time train of the distance between the climber's hip and limb position, when touching the hold. Different types of climbing routes and training paradigms have been used all along the data collection protocol. (b) Position of climber's body center collapsed to 1D by summation ( $x + y$ ) - evolution in time. (c) Observation signal as the relative position of the body centre vs the limb that is touching the hold (collapsed to 1D by summation).

Since human annotation is very costly (it requires expert knowledge and is time consuming) and at the same time - as being based on the visual attribution

- it is not precise, hence we face two challenges simultaneously: 1) the labeling does not cover the whole set of experimentally collected recordings and 2) annotations do carry the mistakes or gaps in the describing sequences. Thus, dealing with these both difficulties (as well as with possible noise) is not a trivial task. One another difficulty we encounter when annotating only subset of the recordings, is the supposed difference in the distributions between the annotated and the unannotated set. Due to the climbing learning protocol evolution, the sequences were not stationary, Fig. 2. The protocol (as explained in [11]) consisted in thirteen climbing sessions: three test sessions and ten training sessions. The test sessions aimed at scanning the behavioral repertoire of the participants. More specifically, to assess to what extent the participants could perform hand alternations (i.e., they used the two different hands on two subsequent movements) and hand repetitions (i.e., they used the same hand on two subsequent movements). In this purpose, they had to climb three different routes that either encouraged to perform alternations, repetitions or both coordination patterns (i.e., the Neutral route). Additionally on the test sessions, participants climbed the three routes in three different instruction conditions: a free condition (to observe the spontaneous behavior), an alternation condition (i.e., they were invited to perform as much alternations as they could on the routes) and a repetition condition (i.e., they were invited to perform as much repetitions as they could). Then, on the training sessions, participants climbed on more sophisticated routes where they were instructed to use all the handholds in a bottom-up order and to climb as fluently as possible, that is, avoiding saccades and stops during the ascents. That way, they had to search for the most adapted chain of movements for the route. According to their training group, participants climbed one to ten different routes during the training sessions. All groups had one route in common that they climbed at least three times per session (i.e., the Control route). Finally, the training protocol started (i.e., first ascent of the first training session) and ended (i.e., last ascent of the tenth training session) with one trial on a Transfer route for which they had no additional practice during the training sessions. This route was designed to assess whether the participants were able to find adapted chain of movements on new routes with training.

In many cases we dispose large registered data sets that can be labelled manually based on observations. Therefore in the whole process human mistakes in the labeling and the presence of artifacts in the recordings are inevitable - both error sources can prompt missing values in the labeling sequences. We would like to use the subset of partially labeled observations (recording of handhold touch time combined with the mass centre position, see Fig. 1 (b)) to predict the labeling (i.e left or right, hand or foot) of the whole dataset. As the data set consists of the pair sequences (observations and *incomplete* labels), thus we will apply the sequential learning with hidden Markov model. The challenging part is to manage the missing parts of the sequences and learn from the annotated parts to fill up the gaps with constrained Viterbi algorithm. Our task is similar to entity recognition with the average labeling ratio about one third of the sequence.



**Fig. 2.** The protocol of data collection with session profile (number of ascends in each session is provided in the box). The PREtest, POSTtest and RETention sessions (yellow) stand for the partially labeled sequences in our training dataset. They consist of simple routes (Neutral, Alternation and Repetition type); Transfer route opens and closes the Training Session. Each of the 10 Training Sessions (the unannotated data set, for which we predict the labels with our model) starts with the Control Route.

## 4 Methods

### 4.1 HMM Viterbi training on fully labeled sequences

The sequential learning task [8] is to find the relation between the observation sequence  $o = (o_1, o_2, \dots, o_T)$  and its pair label sequence  $l = (l_1, l_2, \dots, l_T)$ , where each  $l_t \in \{L_1, L_2, \dots, L_n\} = L$  (assuming the sequences  $o$  are some continuous trains of measurements of length  $T$  and labels are the limbs: LH, RH, LF, RF). For HMM supervised learning framework [17], with fully labeled pairs  $(o, l)$ , this relationship is described with a generative model  $m$ . It is subsequently applied, with the use of Viterbi algorithm [20], on new sequences of observations in order to find most likely corresponding label sequences (Fig. 3(a)). We notify, that by applying the Viterbi algorithm we obtain *hard* label assignation, whereas by using Baum-Welsh algorithm we would get probabilistic one. The former hard assignation is necessary for the next step of the processing in sport science: climb profiling by clustering (not presented here). Seemingly, there is a relation within each pair  $(o_t, l_t)$  all along the sequences, which we describe with probability of observation emission  $e_s(o_t) = P(o_t | l_t = s)$ . In case of a Markov process, there is no dependence between non-adjacent labels, whereas, there is one between the adjacent labels in the sequence  $l$ , that is described with the probability of transition  $f_{rs} = P(l_{t+1} = r | l_t = s)$ . Both mentioned types of conditional probabilities, that in case of sequences apply to every element of sequences  $o$  and  $l$ , in form of matrices  $E = (e_s(o_t))_{s \in L, 1 \leq t \leq T}$  and  $F = (f_{rs})_{r, s \in L}$ , as well as initial probabilities  $p_s = P(l_1 = s)$ , adjusted to data pairs  $(o, l)$  are the parameters of our model  $m = (p, E, F)$ . We find the optimal parameters by maximizing the joint probability  $P(o, l)$ . Once the model parameters are found, we predict new labels for unlabeled observations  $o$  with Viterbi algorithm. The algorithm, starting from  $p$  and propagating through the observation sequence  $o$ , stores the probabilities of most likely path of labels  $l$  that generated  $o$  and same time, the most likely label sequence. The resulting optimal  $\hat{l}$  is the sequence of argmax, once we have found all the probabilities. At first, in the framework

of hidden Markov model, we calculate the joint probability for the sequence of observations  $o$  and the sequence of labels  $l$  as

$$P(o, l) = p_s \prod_{t=1}^T P(o_t | l_t) \prod_{t=1}^{T-1} P(l_{t+1} | l_t) \quad (1)$$

and we estimate the parameters of model  $m$  by maximizing log-likelihood

$$\hat{m} = \arg \max_m \sum_{i=1}^n \log P(o, l | m). \quad (2)$$

Subsequently, with the Viterbi algorithm, the parameters, in a recursive manner and through  $T$  steps, serve to compute the maximum-likelihood label sequence  $\hat{l}$  given the observation sequence  $o$ .

The recurrence is used to track the  $v$  and  $w$  intermittent variables:

$$v_{0,s} = p_s \quad (3)$$

$$v_{t,s} = \max_r v_{t-1,r} f_{rs} e_s(o_t) \quad (4)$$

$$w_{t,s} = \arg \max_r v_{t-1,r} f_{rs} \quad (5)$$

and that for the last estimate label element

$$\hat{l}_T = \arg \max_r v_{T,r} \quad (6)$$

by backtracking we find the whole estimate sequence

$$\hat{l}_t = w_{t+1, \hat{l}_{t+1}}. \quad (7)$$

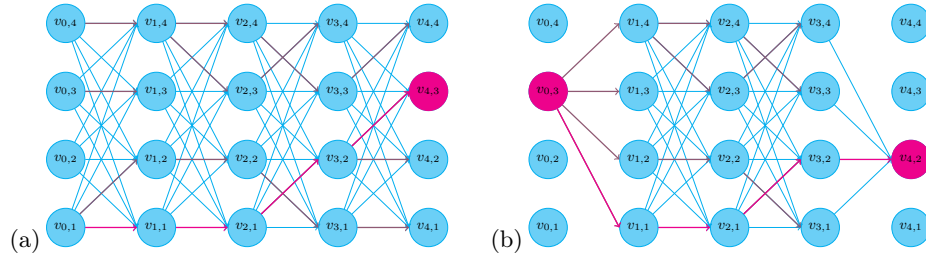
Thus, the ultimate sequence  $\hat{l}$  is the optimal sequence of labels. The optimality of the Viterbi algorithm can be shown as in [16].

#### 4.2 Our contribution: HMM Viterbi training on partially labeled sequences

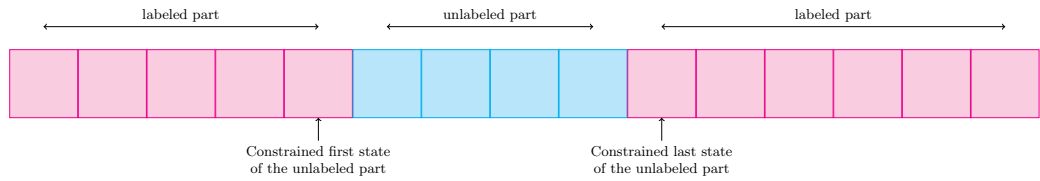
In case of partially labeled sequences in training set, we must adapt the basic supervised HMM procedure in order to train only from the labeled chunks of the sequences. For this purpose we will use the constrained Viterbi algorithm [4], [9]. Firstly, we can notify that for unsupervised learning (when all the observations are unlabeled), we can randomly initialize the model parameters  $m$  and use the Expectation-Maximization algorithm to find both the  $m$  and the sequence of labels in fully unsupervised framework. The herewith technique combines the supervised and unsupervised learning so that initially and in each EM iteration, we take advantage of the labeled chunks of the sequences and constrain the unsupervised scenario.

Namely, we initialize the model with random parameters or with pre-training held on only labeled chunks of the sequences  $(l_t, o_t)$ , for which we get interim





**Fig. 3.** (a) Standard Viterbi algorithm. The magenta path is the optimal one, ending with the maximum  $v$  value (magenta circle). All the possible paths (in cyan) are explored. At each tracking/forward step the most likely path to each state is recorded (dark magenta). At the end, only the path that leads to the most likely ending state is retained (magenta). (b) Constrained Viterbi algorithm. Here, only the paths that contain the first and last pre-fixed states are explored during the forward stage.



**Fig. 4.** Example of partially annotated sequence. Constrained Viterbi is applied on blue chunk accounting for the last magenta token of the first labelled chunk and first magenta token of the second labelled chunk.

parameters  $\hat{m}$ . In the next step, we find the interim labels  $\hat{l}$  for the partially labeled sequence (Fig. 4) with the use of constrained Viterbi, that is enforced to move through the already known label chunks, by maximizing the probability of the most likely label path. Indeed, in the constrained scheme, the traditional Viterbi algorithm operates on the unlabeled segments of sequences, but with the first and the last states labeled (which are the fixed adjacent tokens), treating each chunk sequence independently (Fig. 3(b)). Subsequently, we use the fully labeled sequences  $\hat{l}$  to generate new model parameters  $\hat{m}$  and further adjust the labeling on the initially unlabeled parts of the sequences. While EM finds the local optimum, we iterate until convergence to find the best representation in terms of log-likelihood of the label sequence joint probability (Algorithm 1).

---

**Algorithm 1** HMM with constrained Viterbi

---

- 1: *Initialization*: train model  $\hat{m}$  with *only* fully labeled chunks of sequences  $(o, l)$
  - 2: **repeat**
  - 3:   *Step 1 (E)*: complete the gaps in the sequences of labels with constrained Viterbi algorithm (parameters calculated with model  $\hat{m}$ ), to get fully labeled estimates  $(o, \hat{l})$
  - 4:   *Step 2 (M)*: update model  $\hat{m}$  due to new labels  $(o, \hat{l})$
  - 5: **until** *End condition*: iteration number
-

## 5 Results

### 5.1 Experimental setting

We test the model on the following data:

1. synthetic data set from known model parameters where some annotations are randomly discarded,
2. experimental climbing data - chunks of full sequences with some annotations randomly discarded,
3. experimental climbing data - original partially labelled sequences.

For the synthetic data generation we used the Gaussian bi-variate distribution for the four label states. All the four states have the same covariance  $\mathbf{S}$ , they differ on their mean  $\mathbf{m}$ . For each of the two dimension of the mean, we used either  $-\mu$  or  $\mu$ , leading to 4 possible states as shown in Equation 8:

$$\mathbf{m}_1 = \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \quad \mathbf{m}_2 = \begin{pmatrix} \mu \\ -\mu \end{pmatrix}, \quad \mathbf{m}_3 = \begin{pmatrix} -\mu \\ \mu \end{pmatrix}, \quad \mathbf{m}_4 = \begin{pmatrix} -\mu \\ -\mu \end{pmatrix}, \quad \mathbf{S} = \sigma \mathbf{I}. \quad (8)$$

So defined synthetic data distribution was to approximate the statistics of the climbing data. To apply random label discarding in fully labeled sequences, the level of gap ratio has been fixed (0.25). For the experimental sport data set (points 2 and 3), we dispose the 497 partially annotated sequences of observations (simple sessions), which are divided into train and test sets. The labeling with four types of limbs, reaches the annotation level of about 0.3 for the sequences of the average length of 20 tokens. Finally, the established model will be used for the annotation of overall number of about 3700 sequences recorded with touch sensors. In order to evaluate model performance in case of the first two data sets (where ground true labeling has been initially known), we calculated accuracy measure which is similarity ratio (averaged sum over whole sequence, scoring 0 for mismatch and 1 for match on each token in the sequence) employed on the fully labeled initial sequence  $l$  (before label removal) vs the sequence estimated by the model  $\hat{l}$ . Hence, the evaluation score for a whole sequence  $l$  and its prediction  $\hat{l}$  reads

$$d(l, \hat{l}) = \frac{1}{T} \sum_{t=1}^T d_t(l_t, \hat{l}_t) \quad (9)$$

with the loss function  $d_t$  defined over two tokens  $l_t$  and  $\hat{l}_t$  as

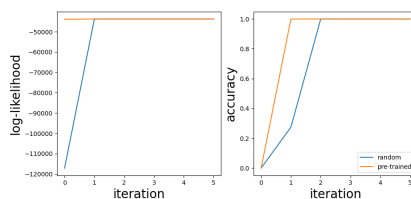
$$d_t(l_t, \hat{l}_t) = \begin{cases} 1 & \text{for } l_t = \hat{l}_t, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

To test the hidden Markov model with constrained Viterbi we used synthetically generated data with known parameters. The missing labels was to best resemble the true annotation present in real data (the gap ratio of about 0.3). We compared the predicted model parameters with the original counterpart as well as the scoring in two sequences: label sequence generated with original model

and label sequence estimated with predicted model parameters. The same way of evaluation has been employed for experimental data with full annotations. In these cases we are able to compare the resulting fully labelled sequences with the initial true labels, before label removal. For all the dataset however, the likelihood of the observation sequences knowing the model parameters is always accessible whether the true labels are fully known, partially known or totally unknown. To sum-up, in case of fully labelled sequences altered artificially to partially labelled ones (dataset 1 and 2), we can compute a similarity ratio comparing the estimated labels to the true ones. Otherwise (dataset 3), only the likelihood of the HMM is accessible as an evaluation measure.

## 5.2 Evaluation

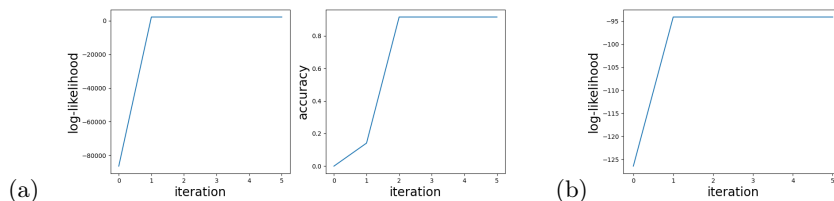
*Synthetic data set* As explained in point 1, firstly we have tested the model on artificially generated data (with either random initialization or pre-trained on labeled chunks of sequences). The number of sequences in trained and test sets was 1000, the average length of sequences was about 20 tokens, with label gap ratio equal to 0.25. We analyse the log-likelihood over the joint probability of observation and labeling sequences that evolves across the iterations for label distribution with mean 1.5 and covariance 0.2 (Fig. 5). This value determines how well the label sequence describes their pair observation sequence, based on the label distribution. As expected, we observe monotonic increase of log-likelihood over the iterations, as well as for the similarity score computed for unannotated chunks. The stability is reached after 2nd iteration. The log-likelihood evolution proves that the model fits well with the experimental data and at the same time leads to the generation of optimal labels. Likewise, score analysis confirm increasing recognition of labels if compared to the ground true annotations in the subsequent steps.



**Fig. 5.** Synthetic data set (generated with mean parameter  $\mu = 1.5$  and covariance parameter  $\sigma = 0.2$ ): log-likelihood value evolution (left) and similarity ratio (right). The labels have been discarded with ratio 0.25. The applied initialization was either random (blue) or pre-trained (orange) on the labeled chunks.

*Climbing data set* We first explore the real experimental dataset with artificial deletion of the annotations (with an alteration rate of 0.3). On the artificially

deleted slots, we do have knowledge of the true labeling, thus we can compute the similarity ratio. After alternation, first, we pre-trained the HMM on the fully labeled chunks of sequences, bypassing the voids. Then, the HMM was trained using the constrained Viterbi in the manner exposed in the Methods 4, with all the partially labelled sequences. In the Figure 6(a), we observe learning through iterations (growth of the log-likelihood), and at the same time - the annotations of the labels are mostly correct (the maximum similarity ratio is over 0.9).



**Fig. 6.** Sport data set. (a) Sequences originally fully labeled but with artificial label deletion: log-likelihood evolution (left) and similarity ratio (right). In order to apply the model, the labels have been synthetically discarded with ratio of about 0.25. (b) original data, partially labeled with gap ratio about 0.3: log-likelihood evolution.

Based on this fact, we studied the case of partially labeled original sequences registered by touch sensors (Point 3). Similarly, for experimental sport data we observe monotonic growth of the log-probability function, which is - as in synthetic data case and the artificially discarded labelings in sport data - depicting learning across the iterations. While the ratio of unannotated labels in all three experimental data cases was about 0.3, we may expect that the results are comparable. Like in the previous data sets, in the original partially labeled sport data, the model log-likelihood converges also very fast, within two iterations (Fig. 6(b)). In this case, unlike the previous two cases, since the true labels remain unknown, the similarity ratio is not accessible. The end-values of log-likelihood function however, are lower than in the artificially label discarded sport data case. We may suppose that this fact could have an impact on the quality of the label prediction. Although we do not dispose any other objective evaluation method, the convergence of the model may be a proof of leading towards the correct solution. We assume that the mislabelling, if applicable, might have occurred in case the artifacts appear in the collected data (e.g. when the sensor registered a touch made not by the hand or foot).

To complete the analysis, we pursued the machine learning evaluation method in which for synthetic data train set to test set ratio was 1000:1000 and for the sport data annotated sequences we split the data set with the ratio 400:97. Namely, we verified the accuracy of the model trained on one subset of partially labeled data (train set), when applied on the previously unused subset of data with no labeling (test set). For this setting, the similarity score evaluated on the

new test sequences were not substantially different from the previously observed training evaluation (Tab.1), and confirmed correct label attribution.

Data set	training	testing
synthetic	0.99	0.99
climbing	0.91	0.90

**Table 1.** Similarity score for synthetic data set vs climbing data set.

## 6 Conclusion and perspective

In the present article, we have been studying how the generative model (HMM) with a simple adjustment (constrained Viterbi) deals with the learning the partial labelling in the training sequences. It is a predictive model, that allows to access the distribution of the limbs used while climbing the artificial wall. We tested the approach on the real climbing data as well as on the synthetic data, that was alike. We claim that, based on the log-likelihood value, which specifies how good is the description that we find for the measured data, due to our experiments we can judge on model performance. We must admit however, that the additional challenge we face in our application is the discrete set of states (labels). This situation renders difficult to measure the effective labelling ratio based only on log-likelihood value. We have compared the log-likelihood ratio with similarity score computed for unannotated chunks in cases, when the full labelling has been known in order to assert patterns of likelihood function. We observed that it proves apt attribution of labels in case of the data set, when the prior full labelling is not accessible, as in the sport data set. We have demonstrated, that the log-likelihood matches scoring and can be used for model evaluation in case of training on only partially labeled data sets.

*Perspective* We can show experimentally that the log-likelihood convergence with a general tendency of monotonic increase (and not decrease) approximates correct label attribution. In a broader setting though, HMM with constrained Viterbi may not be a perfect label predictor for the continuous behavioral signal, but is a simple way to approximate labelling of unknown sequence, with scarce prior knowledge about training sample. Therefore, to outperform our algorithm, some more sophisticated method, as for instance neural networks would have to be applied. Likewise, the convergence of optimality function for the leading model (not only for the Viterbi constrained, but its implementation into the framework of EM-like full Algorithm 1), requires more systematic research and theoretically correct proof. Although the model is simplistic and its output may not be faultless, it is still sufficient for the first approximation of the behavioral signal labelling in sport application and it can be the baseline for the further research on the subject, that could be tested in the contexts of different paradigms in human movement science.

## 7 Acknowledgement

This work was supported by a grant from the French National Agency of Research (reference: ANR-17-CE38-0006 DynACEV).

## References

1. Antoniuk, V., Franc, V., Hlavac, V.: Consistency of structured output learning with missing labels. In: Holmes, G., Liu, T.Y. (eds.) *JMLR: Workshop and Conference Proceedings*, pp. 45:81–95. ACML 2015 (2015)
2. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. *Proceedings of the Workshop on Computational Learning Theory* pp. 92–100 (1998)
3. Boulanger, J., Seifert, L., Hérault, R., Coeurjolly, J.F.: Automatic sensor-based detection and classification of climbing activities. *IEEE Sensors Journal* **16**(3), 742–749 (2016)
4. Cao, L., Chen, C.W.: A novel product coding and recurrent alternate decoding scheme for image transmission over noisy channels. *IEEE Transactions on Communications* **51**, 1426 – 1431 (10 2003)
5. Chapelle, O., Scholkopf, B., Zien, A.: *Semi-supervised Learning*. MIT Press, Cambridge (2006)
6. Clark, K., Luong, M.T., Manning, C., Le, Q.: Semi-supervised sequence modeling with cross-view training. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. pp. 1914–1925. Association for Computational Linguistics, Brussels, Belgium (10 2018). <https://doi.org/10.18653/v1/D18-1217>, <https://www.aclweb.org/anthology/D18-1217>
7. Dai, A., Le, Q.: Semi-supervised sequence learning. *Proceedings of Neural Information Processing Systems Conference* pp. 1–10 (2015)
8. Dietterich, T.: Machine learning for sequential data: A review. *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition* **1**, 1–15 (2002)
9. Fernandes, E., Brefeld, U.: *Learning from Partially Annotated Sequences*. Springer-Verlag Berlin Heidelberg 2011 (2011)
10. Fernandes, E., Brefeld, U., Blanco, R., Atserias, J.: Using Wikipedia for cross-language named entity recognition. In: Atzmueller, M., Chin, A., Janssen, F., Schweizer, I., Trattner, C. (eds.) *Big Data Analytics in the Social and Ubiquitous Context*. pp. 1–25. Springer International Publishing, Cham (2016)
11. Hacques, G., Komar, J., Bourbousson, J., Seifert, L.: Climbers’ learning dynamics : an exploratory study. 4th International rock climbing congress (IRCRA). 12-15th July, Chamonix, France (2018)
12. Huang, X., Dong, L., Boschee, E., Peng, N.: Learning a unified named entity tagger from multiple partially annotated corpora for efficient adaptation. In: *Proceedings of the 23rd Conference on Computational Natural Language Learning*. pp. 515–527 (2019)
13. Li, J., Liu, C., Liu, B.: Large scale sequential learning from partially labeled data. 2013 IEEE Seventh International Conference on Semantic Computing pp. 176–183 (2013)

14. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. NIPS (2013)
15. Ning, Q., He, H., Fan, C., Roth, D.: Partial or complete, that's the question. In: Proceedings of NAACL conference. pp. 2190–2200 (01 2019). <https://doi.org/10.18653/v1/N19-1227>
16. Omura, J.: On the Viterbi decoding algorithm. IEEE Transactions on Information Theory **15**(1), 177–179 (1969)
17. Rabiner, R.: A tutorial on hidden Markov models and selected applications on speech recognition. Proceedings of the IEEE **77**(2), 257–286 (1989)
18. Scheffer, T., Decomain, C., Wrobel, S.: Active hidden Markov models for information extraction. In: Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis. pp. 309–318 (09 2001)
19. Veeramachaneni, W., Liao, V.: A simple semi-supervised algorithm for named entity recognition. Workshop on Semi-supervised Learning for Natural Language Processing pp. 58–65 (2009)
20. Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory **13**(2), 260–269 (1967). <https://doi.org/10.1109/TIT.1967.1054010>
21. Zhi, S., Liu, L., Zhang, Y., Wang, S., Li, Q., Zhang, C., Han, J.: Partially-typed NER datasets integration: Connecting practice to theory. (in press) pp. 1–13 (May 2020)
22. Zhou, J., Fang, M., Zhang, H., Gong, C., Peng, X., Cao, Z.G., Goh, R.: Learning with annotation of various degrees. IEEE Transactions on Neural Networks and Learning Systems **PP**, 1–11 (01 2019). <https://doi.org/10.1109/TNNLS.2018.2885854>