Data-Driven Goal Recognition in Transhumeral Prostheses Using Process Mining Techniques

Zihang Su*, Tianshi Yu†, Nir Lipovetzky*, Alireza Mohammadi†, Denny Oetomo†,
Artem Polyvyanyy*, Sebastian Sardiña‡, Ying Tan†, Nick van Beest§

*School of Computing and Information Systems, The University of Melbourne, Australia
zihangs@student.unimelb.edu.au, {nir.lipovetzky,artem.polyvyanyy}@unimelb.edu.au
†Department of Mechanical Engineering, The University of Melbourne, Australia
tianshiy@student.unimelb.edu.au, {alireza.mohammadi,doetomo,yingt}@unimelb.edu.au
‡School of Computing Technologies, RMIT University, Australia
sebastian.sardina@rmit.edu.au
§Software Systems Group, Data61 | CSIRO, Australia
nick.vanbeest@data61.csiro.au

Abstract—A transhumeral prosthesis restores missing anatomical segments below the shoulder, including the hand. Active prostheses utilize real-valued, continuous sensor data to recognize patient target poses, or goals, and proactively move the artificial limb. Previous studies have examined how well the data collected in stationary poses, without considering the time steps, can help discriminate the goals. In this case study paper, we focus on using time series data from surface electromyography electrodes and kinematic sensors to sequentially recognize patients' goals. Our approach involves transforming the data into discrete events and training an existing process mining-based goal recognition system. Results from data collected in a virtual reality setting with ten subjects demonstrate the effectiveness of our proposed goal recognition approach, which achieves significantly better precision and recall than the state-of-the-art machine learning techniques and is less confident when wrong, which is beneficial when approximating smoother movements of prostheses.

Index Terms—goal recognition, active transhumeral prostheses, process mining

I. Introduction

Given a collection of candidate goals and observations of actions performed by an agent in an environment, a solution to the goal recognition (GR) problem suggests the true goal the agent strives to achieve in the environment [1]. Process mining (PM) techniques were recently used to implement a system for solving the GR problem [2], [3]. We refer to this system as the PM-based GR system. Assuming the actions the agent uses to achieve the various goals do not vary significantly over time, the PM-based GR system uses *process discovery* techniques [4] to construct process models, or behavior models, from the action sequences the agent performed in the past to achieve the goals. Each model is constructed from the historical observations of how the agent reached a specific goal and represents the exemplary behavior for achieving the goal. Subsequently, the PM-based GR system uses *conformance checking* techniques [5] to align a newly observed action sequence (a trace) with each discovered behavior model. Finally, the obtained conformance information regarding the commonalities and discrepancies between the trace and all the models is translated into a probability distribution over the candidate goals. This distribution describes the probabilities of the agent pursuing the goals.

Goal recognition techniques are used in many real-world scenarios, such as autonomous driving [6], [7], robotics [8], [9], and human-machine interaction [10]. This paper investigates the feasibility and potential benefits of using the PM-based GR system in active transhumeral prostheses designed to assist individuals with disabilities [11]–[13]. A transhumeral prosthesis replaces the function of missing anatomical segments below the shoulder of a patient, including the hand. An active prosthesis is equipped with sensors and motors that aim to recognize the intents of the patient and support them, for example, by automatically directing the artificial limb towards the target prosthetic pose, namely the goal.

This study uses the dataset collected for the development of active transhumeral prostheses. The data collection experiment involved ten non-disabled subjects in a virtual reality (VR) environment, aiming to emulate the behavior of patients using transhumeral prostheses [12]. Each subject was requested to accomplish forward-reaching tasks that involved achieving three elbow poses, namely the goals. They were required to extend their sound upper limb forward in a series of 30 iterations for each goal. The reaching targets are placed along the parasagittal plane. A small sphere in the VR environment communicated a reaching target to the subject, and the subject had to reach it with their hand. The subjects were requested to stay still for approximately one second after reaching the goal. During each iteration towards a goal, 47 features were extracted from the measurements of kinematic and surface electromyography (sEMG) sensors attached to the subject at regular intervals. These sequences of features constitute time series of continuous, real-valued data that characterize the behavior of the subject in achieving the goal.

This work is a natural extension of our previous research, which demonstrated that the measurements during the one-second holding period contain features suitable for discriminating between different fixed poses of the subjects, without considering the sequence data [12]. However, intuitively,

considering the time stamps in the measurements is expected to improve performance. In this study, we investigate the potential of utilizing sequences of these measurements to infer the subjects' goal poses as they strive to achieve them. By predicting the patient's goals, our aim is to inform the design of active prostheses that facilitate cooperation between the patient and the prosthesis, ultimately enhancing the overall user experience. To accomplish goal inference, we leverage our PM-based GR system. Additionally, since the GR system operates on discrete event data, we propose a two-stage approach for transforming the multi-dimensional, real-valued, continuous measurements into discrete events.

We compared the performance of the PM-based GR system with two state-of-the-art GR methods in the area of active prostheses: linear discriminant analysis (LDA) [13] and long short-term memory (LSTM) neural networks [14]. The results confirm that the PM-based GR system outperforms the LSTM-based and LDA-based machine learning methods. The artifacts constructed by the PM-based system (discovered models and alignments between new observations and the models) can be used to explain the inferences. Furthermore, the PM-based GR tends to provide less confident results when making mistakes, which is beneficial when approximating smoother movements of prostheses. The practical solution to the GR problem in the field of transhumeral prostheses can benefit from further developments in the GR techniques and their ensembling. Specifically, this paper contributes:

- An extension of the data-driven approach for goal recognition grounded in process mining techniques [2], [3] to repeated multi-dimensional, real-valued, continuous measurements that characterize the observed behavior of interest;
- The results of an evaluation based on a publicly available implementation[1] of the data-driven GR systems discussed in this paper, including the one grounded in process mining techniques, that compares their performance over a publicly available dataset[2] in the domain of transhumeral prostheses. These results confirm that the process mining approach achieves significantly better precision and recall in recognizing the target patients' poses and is less confident when wrong than the state-of-the-art machine learning techniques.

The next section introduces our transhumeral prosthetic study. Next, Section III discusses related work. Section IV presents the extended version of the PM-based GR system, which is evaluated against state-of-the art machine learning approaches in Section V. Finally, Section VI discusses the limitations of the current techniques and future work aiming to address these limitations, while Section VII concludes this paper.

II. STUDY DESCRIPTION

A. Background

The main objective of our study is to develop active transhumeral prostheses that replace missing limb segments below the shoulder, restoring upper limb function for achieving

specific prosthetic poses. Active prostheses are robotic devices with actuators, like electric motors, driving prosthetic joints for tasks such as reaching. In this work, these joint movements are controlled by synthesizing data from above-elbow sEMG and joint movement sensors on the residual limb (upper arm) and body [11], [12]. Given that the user-intended prosthetic joint pose is unknown to the prosthetic device, the precise identification of this goal becomes crucial. Failure to achieve this can lead to inefficient task execution, user dissatisfaction, and even potential abandonment of the device [15]. However, it is challenging to develop goal recognition to accurately detect the intended goal in prosthetic settings due to significant human variance in kinematic and muscle activity signals [16].

The goal recognition algorithm is required to be capable of accurately identifying the intended goals (target prosthetic poses) based solely on the signals accessible above the elbow. To this end, it is a common practice to first explore datasets from non-disabled human subjects [15]. These datasets capture the above-elbow sEMG and joint movement data for every sampling time instance while the subjects perform the task of forward-reaching. Each of these instances is labeled with its corresponding goal. The studied dataset herein was collected from ten non-disabled subjects performing forward-reaching tasks by extending their intact upper limb forward in a head-mounted display (HMD) VR environment. The data collection focused on transhumeral prostheses with a single prosthetic elbow joint. The aim was to attain distinct elbow poses that corresponded to specific spatial locations for forward-reaching.

B. Methods of Data Collection

This subsection elaborates on the dataset and data collection procedure. Ten healthy subjects (seven males and three females), all right-handed, were recruited. The experimental protocol was approved by the University of Melbourne Human Research Ethics Committee under project ID 11878. Informed consent was obtained from all the subjects.

The dataset contains 35 sEMG and 12 joint kinematic movement features of the subjects intending for three goals at three different shoulder poses. The features were sampled at a rate of 10 Hz (the measurements were taken every 0.1 seconds). The goals (target elbow poses) are denoted as T1, T2, T3 in Fig. 1a which shows the side-view schematic of the upper limb. The process of extracting the sEMG and joint pose features has been described in detail in our previous work [12].

The sensor setup and virtual avatar in VR are shown in Fig. 1b and Fig. 1c, respectively. In the experiment, the subjects stood naturally and unconstrained in an upright position to perform forward-reaching tasks by extending their upper limb towards the three goals at three shoulder poses. The goal is considered reached when the middle finger of the virtual avatar hits the target sphere, generated to elicit the target shoulder and elbow poses as shown in Fig. 1c. For each goal, the subjects completed 30 iterations of reaching. During each iteration, the subjects were instructed to maintain their final upper limb pose for one second upon reaching the goal. The
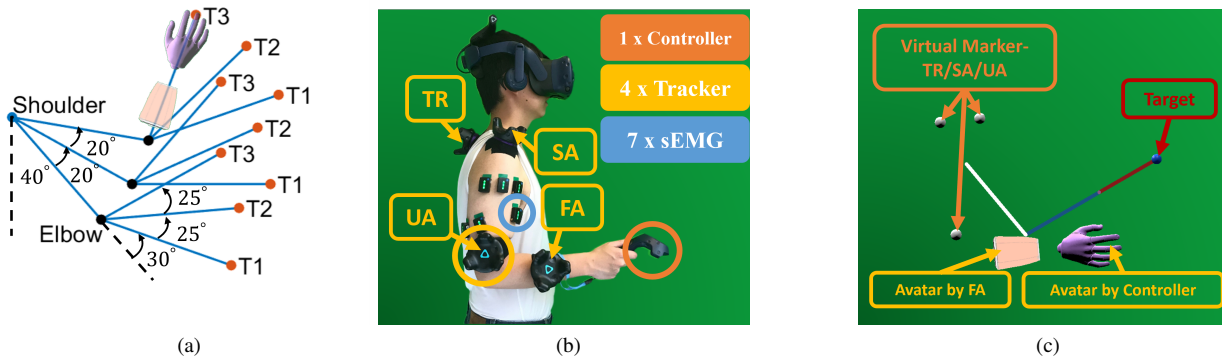
Fig. 1: (a) Target shoulder and elbow poses, T1–T3 denote the three goals, (b) Experimental setup and the placement of VIVE trackers and sEMG electrodes, (c) VR avatar and reaching target example (side view).

recorded data spanned from the start of the movement to the end of the holding period.

Given the focus on transhumeral prostheses, the kinematic movements of the above-elbow joints were recorded along with the muscle activity of the upper arm. To capture the kinematics, three HTC VIVE trackers were strategically positioned at the upper arm (UA), shoulder acromion (SA), and trunk (TR). An additional tracker was placed on the forearm (FA), and a controller was held in the hand solely for controlling the forearm and hand avatar within the VR environment, as shown in Fig. 1c. For monitoring the muscle activity, seven Delsys® Trigno™ wireless sEMG electrodes were attached to the muscles of the dominant upper arm of each subject.

III. Related Work

Existing GR methods can be broadly classified into three categories: plan library [17], planning-based [18], and data-driven [2], [3], [19] methods. A plan library GR method relies on exemplary plans for accomplishing candidate goals, usually hand-crafted by domain experts. The method proceeds by comparing the observed agent actions with the available example plans to infer the true goal of the agent. The planning-based GR techniques use pre-defined domain models that describe the environment and possible actions agents can take in the environment. Once a new observed sequence of actions in the environment is available, these techniques rely on the domain models to construct and compare plans toward the candidate goals to infer the likelihoods of the goals. Finally, the data-driven techniques use historical data collected from agents and the environment to perform goal inference.

In the field of transhumeral prostheses, previous studies have demonstrated that employing diverse features tailored to individual patients results in improved performance when classifying target poses [13]. This personalized approach to handling patients complicates crafting exemplary plans and domain models, as each patient requires dedicated plans and models. In this light, data-driven GR techniques appear appropriate for the individual user, as they can learn personalized artifacts from the historical behavior of a patient for subsequent goal inferences tuned for the patient. In this work, we apply the data-driven GR at the level of individual subjects.

The data-driven GR approach based on LSTM neural networks [19] learns a model from a set of action sequences. It then identifies the most likely goal of the agent based on a newly observed sequence of actions. Based on a collection of historical sequences of actions towards a candidate goal, the PM-based GR system [2], [3] uses process discovery techniques to learn a process model that describes the skill for accomplishing the goal. Subsequently, it relies on conformance checking techniques [5] to study commonalities and discrepancies between a newly observed sequence of actions and all the learned models, one model for each candidate goal. These commonalities and discrepancies are then translated into a probability distribution over the candidate goals, capturing the likelihoods that the actions aim to reach these goals.

Machine learning techniques can be used to implement accurate prosthesis control [16]. Due to their robustness, machine learning classifiers are commonly implemented in upper-limb prostheses [20]. Given an input signal, for example, from sensors, a classifier predicts the output signal and, the intended movement of the patient. Linear discriminant analysis (LDA) is the most commonly used classifier algorithm in prosthesis control. It can provide high control accuracy using small training and processing times [21]. Our recent work [13] confirmed that LDA can discriminate poses reliably. Recent works explore the use of artificial neural networks in prosthesis control [16]. Neural networks can achieve high goal recognition accuracy by learning non-linear dependencies between signal input and control outputs but require extensive training over large datasets [16]. For instance, Huang et al. [14] successfully used LSTM neural networks to predict target poses based on time series of electromyography signals.

This paper aims to understand how the PM-based GR system can contribute to developing active transhumeral prostheses. To this end, we extend the system to work with repeated, real-valued, continuous data captured by typical sensors and propose dedicated training techniques. We use LDA [13] and LSTM [14], [19] models as performance baselines.

IV. Approach

This section presents the PM-based GR system [2], [3] and its new training approach based on real-valued, continuous

data. The approach starts by reducing the data dimensionality by selecting relevant features (cf. Section IV-A). Next, the data points of reduced dimensionality are clustered to define discrete events (cf. Section IV-B). Section IV-C explains our approach to discover process models from discrete events. Finally, Section IV-D summarizes our technique for goal inference based on conformance diagnostics between the discovered models and new observations.

To illustrate the approach, we use six traces in which events are characterized by 30 real-valued, continuous features f_1 to f_{30}; the dataset and tool for replicating the running example are publicly available at https://doi.org/10.26188/24131493. The first three traces (1–3) represent the signal sequences collected when reaching the target pose T1, while the last three traces (4–6) were recorded when reaching the target pose T2. Table I shows an extract of the dataset. In the table, each row holds feature values that were collected simultaneously and the rows are ordered by their time stamp of data collection.

Fig. 2: Correlation matrix.

Trace	Goal	f_1	f_2	f_3	$\dots$	f_{29}	f_{30}
1	T1	5.19727337	7.02395793	0.00254431	…	5.39759498	-0.3722619
1	T1	7.76278776	8.08816201	0.00472689	…	1.01557531	1.37592798
1	T1	13.4185557	8.87159453	0.00821896	…	-4.0004147	1.65328609
1	T1	22.0916619	9.04377674	0.01015369	…	-5.5399488	-1.7805512
1	T1	31.3641039	9.3586209	0.009165	…	-3.5156837	1.36367015
1	T1	38.2312577	10.139119	0.00616715	…	-1.4720033	5.87820456
1	T1	42.0592085	10.8827908	0.00315491	…	-0.3338844	4.29640897
2	T1	7.39110795	6.07336937	0.00064332	…	2.92403705	1.46698529
2	T1	10.5229866	7.44734189	0.00194998	…	1.60034347	2.94734496
…	…	…	…	…	…	…	…
6	T2	64.1830578	25.2975943	-0.0003433	…	-1.1970367	0.92412363
6	T2	66.8916142	27.5304609	-0.0017204	…	0.31022101	0.95595258

TABLE I: Extract of the running example dataset.

A. Feature Selection

To reduce the dimensionality of the data, we eliminate highly correlated features from the analysis. We use the agglomerative hierarchical clustering method [22] to cluster the features into N_f groups based on the distances defined by the correlations between the features. Figure 2 shows the matrix of absolute values of the Pearson correlation between the features, while Fig. 3 depicts the dendrogram of the clusters. The dendrogram represents the hierarchical structure of clusters, which supports the flexible selection of the desired features using a similarity threshold. A similarity threshold value defines the desired distance between the constructed clusters. For instance, Fig. 3 shows the selection of 15 clusters using the similarity threshold value of 1.23. A cluster is defined by all the leaf features in a dendrogram branch cut horizontally by the threshold line. The systematic approach for selecting the features is discussed in Section V. Here, we use N_f of 15 for demonstration.

As the features within a cluster are highly correlated, we select the medoid of the cluster, the feature with the smallest average dissimilarity to all other features in the cluster, as the representative of the cluster. This way, we obtain N_f features that characterize the source data. In the running example, we reduce the original 30 features to $N_f = 15$ representative features highlighted in Fig. 3 in red and listed in Table II.
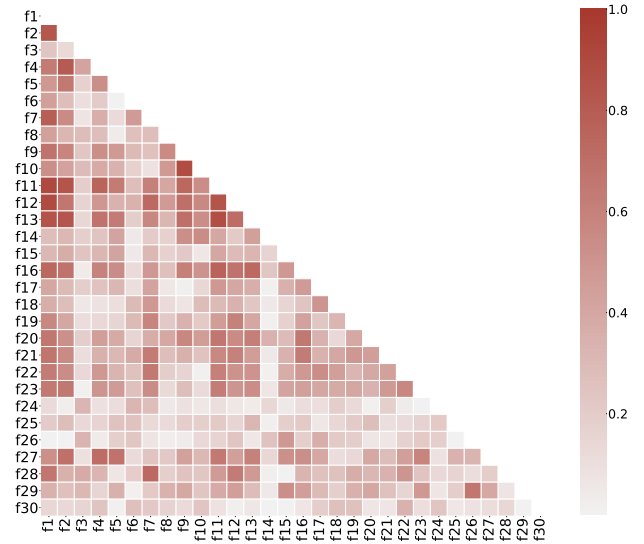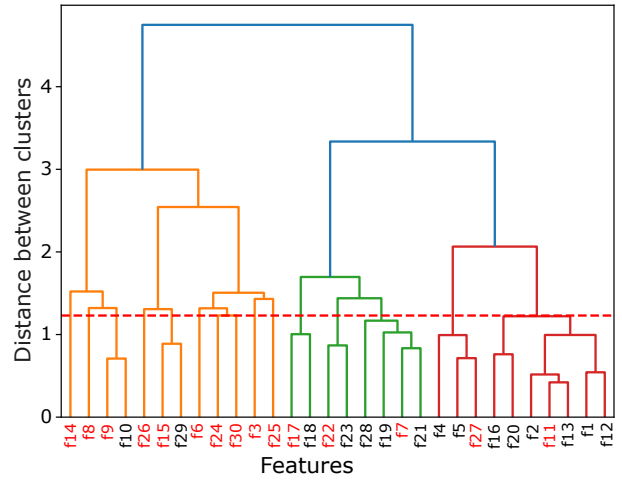

Fig. 3: Dendrogram and selection of clusters. The red dotted line represents the threshold used to cut the dendrogram to form 15 clusters.

B. Event Discretization

We transform the N_f-dimensional time series of real-valued, continuous measurements obtained in the previous step into traces of events amenable to process mining. To this end, we cluster the N_f-dimensional data points from the series into N_c clusters. The data points in a cluster are accepted as the same event. We use k-means algorithm [23] due to the performance considerations for clustering high-dimensional data.

In the example, we clustered the 15-dimensional data points defined by the 15 selected in the previous step features into 10 clusters, represented by events e_0 to e_9, see the "Event" column in Table II. Again, the systematic approach for selecting the number of clusters is discussed in Section V. Here, we use $N_c = 10$ for demonstration only. The traces of obtained events can be split into two event logs, containing the traces towards goals T1 and T2, denoted by L_1 and L_2, respectively.

Trace	Goal	f_3	f_6	...	f_{27}	f_{30}	Event
1	T1	0.002544311	0.121301538	...	9.057075924	-0.372261852	e_0
1	T1	0.004726894	0.210557727	...	10.55266625	1.375927982	e_6
1	T1	0.008218956	0.391712037	...	4.416704571	1.653286092	e_5
1	T1	0.010153689	0.327711654	...	1.372325318	-1.780551234	e_2
1	T1	0.009165	0.311548058	...	5.526959903	1.363670147	e_5
1	T1	0.00616715	0.734098175	...	8.869853729	5.87820456	e_6
1	T1	0.003154906	1.227944362	...	4.916643199	4.296408971	e_8
2	T1	0.000643317	0.103164637	...	12.86040763	1.466985286	e_8
2	T1	0.001949978	0.336630569	...	13.45204634	2.94734496	e_6
...	...	...	...	...	...	...	...
6	T2	-0.000343322	0.410761081	...	23.78469914	0.924123631	e_4
6	T2	-0.001720367	0.504684583	...	21.08908217	0.955952575	e_3

TABLE II: Extract of the reduced running example dataset.

C. Process Discovery

Given an event log of discrete events towards a target pose, we use process discovery to construct a process model that describes and generalizes the historical traces that were followed in the past to achieve the corresponding goal. For example, we use the Directly Follows Miner algorithm [4] to construct Petri nets M_1 and M_2 shown in Figs. 4 and 5 from event logs L_1 and L_2 obtained in the previous step.

Since event logs L_1 and L_2 solely consist of traces related to reaching target poses T1 and T2, respectively, process models M_1 and M_2 depict the potential processes involved in reaching corresponding final target poses T1 and T2. Models M_1 and M_2 are stored in our GR system as "knowledge," which assists in predicting the target pose of the prosthesis when the GR system receives new sequences of signals.
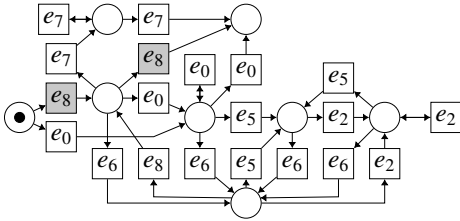

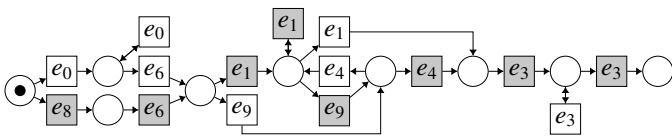Fig. 4: Process model M_1 discovered from even log L_1.

Fig. 5: Process model M_2 discovered from even log L_2.

D. Goal Recognition

Once a new (prefix of a) trace of multi-dimensional, real-valued, continuous measurements is observed, it is used to infer the goal of the patient. The trace is reduced to the representative features and transformed into an event trace. The selection of representative features is described in Section IV-A. A data point is attributed to the closest event cluster out of the clusters obtained as described in Section IV-B. For example, suppose we observe this new trace of events:

$$\tau = \langle e_8, e_6, e_2, e_1, e_1, e_9 \rangle.$$

The PM-based GR system uses conformance checking to diagnose discrepancies and commonalities between the event trace and all the discovered process models. Specifically, it constructs optimal alignments [5] between the trace and the models. Below, we show optimal alignments σ_1 and σ_2 between example trace τ and process models M_1 and M_2. The transitions of Petri nets M_1 and M_2 that participate in the alignments with τ are highlighted in gray in the figures.

$$\sigma_1 = \begin{array}{|c|c|c|c|c|c|c|}\hline \tau & e_8 & \gg & e_6 & e_2 & e_1 & e_1 & e_9 \\ \hline\end{array}$$
$$\begin{array}{|c|c|c|c|c|c|c|}\hline M_1 & e_8 & e_8 & \gg & \gg & \gg & \gg & \gg \\ \hline\end{array}$$

$$\sigma_2 = \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline \tau & e_8 & e_6 & e_2 & e_1 & e_1 & e_9 & \gg & \gg & \gg \\ \hline\end{array}$$
$$\begin{array}{|c|c|c|c|c|c|c|c|c|}\hline M_2 & e_8 & e_6 & \gg & e_1 & e_1 & e_9 & e_4 & e_3 & e_3 \\ \hline\end{array}$$

An alignment is a sequence of moves. In a table that encodes an alignment, columns represent moves. A move with the special "$\gg$" skip symbol is an asynchronous move. Otherwise, a move is a synchronous move. A synchronous move represents that both the trace and model can proceed synchronously by executing the same action. A move with the skip in the bottom (top) row is a move on trace (on model), denoting that the model (trace) cannot mimic the action in the trace (model). An optimal alignment is an alignment with the "cheapest" asynchronous moves according to some predefined positive costs of asynchronous moves that capture the minimal (as per the costs) deviations between the trace and model. For example, optimal alignment σ_1 starts with a synchronous move, denoting that both the trace and the model can start with event e_8, followed by five asynchronous moves on trace, capturing that it is cheaper for the model not to match these occurrences of events in the trace.

The more synchronous moves between trace τ and model M_G in the alignment, the higher the likelihood that the trace is following M_G, indicating an intention to reach goal G. In the running example, observed trace τ is more likely to reach target pose T2. The detailed probability calculation uses *alignment weight* between τ and M_G, as defined below [2].

$$\omega(\tau, M_G) = \phi + \lambda^m \times \sum_{i=1}^{n} \left(i^\delta \times c(\tau, M_G, i) \right) \quad (1)$$

In Eq. (1), $c(\tau, M_G, i)$ is the cost of the move in the alignment between τ and M_G at position i (we use a cost of one for all asynchronous moves on trace and a cost of zero for all other moves), n is the length of the alignment, ϕ is a constant, δ is a discount factor that emphasizes later asynchronous moves in the alignment, $\lambda \geq 1$ is a penalty for asynchronous moves on trace, and m is the number of consecutive asynchronous moves on trace occurring at the end of the alignment. Intuitively, the more asynchronous moves at the end of the alignment, the greater the alignment weight.

The alignment weight is used to compute the probability of reaching the corresponding goal, as captured below [2].

$$\Pr(G \mid \tau) = \frac{e^{-\beta \times \omega(\tau, M_G)}}{\sum\limits_{G' \in \mathcal{G}} e^{-\beta \times \omega(\tau, M_{G'})}} \quad (2)$$

In Eq. (2), $\mathcal{G}$ is the set of candidate goals, G' denotes each goal candidate in the set $\mathcal{G}$, and $\beta \in (0, 1]$ is the level of trust over the discovered process models. Based on the two example alignments, the probabilities of trace τ reaching T1 and T2 are 0.06 and 0.94, respectively, using the default parameter setting [2]. Consequently, the PM-based GR system infers that trace τ intends to reach T2.

V. EVALUATION

A. Experiment

We collected data on ten subjects, each executing 30 traces to each of the three target poses, a relatively small dataset. Thus, we use cross-validation to appropriately split the available experimental data into training and testing sets. In the cross-validation, we conduct experiments for 30 iterations. In each iteration, three traces, one trace per target pose, are selected as testing traces, while the remaining 87 traces, 3 goals $\times$ (30-1) traces, are used for training. Hence, we test each trace once, resulting in 90 training sets for each subject.

The recognition performance is measured by precision and recall on a micro level. Since the PM-based GR system may infer multiple possible poses, we follow the method from our previous work [3] to calculate the precision (p) and recall (r) for each problem instance and then compute the averages of these individual precision and recall measurements over all the instances. In each GR instance, where the number of true target poses is one, precision (p) is defined as the fraction of correctly inferred target poses (either zero or one) among all the inferred poses (T1, T2, or T3). Recall (r) is defined as the fraction of the correctly inferred true poses. Thus, recall is one if the unique true pose is inferred, regardless of the number of inferred false target poses, and zero otherwise.

The configurable parameters in the PM-based target pose recognition system, mentioned in Section IV, are optimized through a combination of brute force search method and the PRIM algorithm [24] with Latin hypercube sampling [25]. The number of selected features N_f is chosen from a set $N_f \in \{1, 2, \ldots, 47\}$, and the number of clusters N_c for event discretization is selected from a set $N_c \in \{10, 20, ..., 200\}$. We conduct experiments on every combination of N_f and N_c, and then select the N_f and N_c values that result in the highest recognition performance (F1-score defined by $\frac{(2p \times r)}{(p+r)}$). For parameters such as ϕ, δ, and λ in Eq. (1), we followed the method described in [3] to utilize the PRIM algorithm and explored 100 different parameter configurations, and selected the configuration that yielded the highest F1-score.

B. Baselines

We compare the performance of the PM-based GR system with two state-of-the-art target pose recognition techniques based on LSTM and LDA (cf. Section III for details). The LSTM neural network is specifically designed to capture dependencies and patterns in sequential data, making it suitable for classifying multi-dimensional, real-valued, continuous measurements (sEMG and kinematic signals) to identify the target pose of the prosthesis. We implement the LSTM network using the same settings and hyperparameters outlined in [14] as a benchmark for comparison. LDA is a trainable classification function that linearly separates multi-dimensional, real-valued, continuous measurements (data points) into a specified number of clusters. However, it is typically applied to classify single data points, such as the sEMG and kinematic signals at a specific timestamp, rather than a sequence of signals. In this transhumeral prosthesis scenario, where the instantaneous sEMG and kinematic signals at different target poses are distinct, the LDA classifier is trained with sets of signals collected during the period when the arm is held at three different target poses. When testing the recognition performance of LDA, as the prosthesis moves closer and closer to a specific target pose, the last observed instantaneous signals become increasingly recognizable to the trained LDA classifier. Therefore, we implement an existing LDA-based target pose recognition approach [13] based on the last instantaneous signal of an observed trace, rather than a sequence of signals like the PM or LSTM approaches.

All recognition approaches (PM, LSTM, LDA) were executed on a single core of an Intel Xeon Processor@2.0GHz, using the same selected features and employing the same cross-validation strategy to split the experimental dataset into training and testing sets. The PM and LSTM approaches were trained using the sequences of signals collected during the arm movement towards target poses, while the LDA approach was trained using 10 data points collected after the arm reaches the target pose and holds at the final position. These signals used to train the LDA approach are not used to train the PM and LSTM approaches. As GR techniques aim to identify goals before the full sequences of signals are observed, we evaluate the approaches using prefixes of the full traces that are cut off at the first 10%, 30%, 50%, and 70% of the full sequences.

C. Results

The average precisions (p) and recalls (r), obtained by experimenting on each individual subject to recognize the target poses using different lengths of prefixes (10%, 30%, 50%, and 70% of the full trace), are presented in Table III. The two columns "Features" and "Clusters" show the number of selected features, N_f, and the number of discrete event clusters, N_c, for each subject. Since the LSTM- and LDA-based approaches only infer one possible pose for each problem instance, the precision and recall values for these approaches remain the same in all cases. The last row shows the average across all subjects and all levels of observation.

We performed t-tests to examine the statistical significance of the differences in average precisions and recalls between the PM-based approach and the other two benchmarks. The null hypothesis of the t-tests is that there is no significant difference between the average precisions and recalls. The t-tests comparing the average precision and recall between the PM- and LSTM-based approaches yield p-values of 1.185e-03 and 4.173e-34, respectively. Similarly, the p-values between the PM- and LDA-based approaches are 5.153e-10 and 1.188e-34.

Subject	Features	Clusters	Obs%	PM		LSTM	LDA
				p	r	$p = r$	$p = r$
1	29	70	10%	0.365	**0.967**	0.389	**0.411**
			30%	**0.431**	**0.722**	0.422	0.378
			50%	0.491	**0.744**	**0.500**	0.456
			70%	**0.628**	**0.822**	0.589	0.611
2	1	10	10%	0.328	**0.967**	0.300	0.289
			30%	**0.361**	**0.911**	0.333	0.333
			50%	**0.396**	**0.967**	0.333	0.244
			70%	**0.435**	**0.956**	0.333	0.311
3	2	150	10%	0.328	**0.978**	**0.333**	0.322
			30%	**0.446**	**0.878**	0.322	0.344
			50%	**0.524**	**0.778**	0.311	0.333
			70%	**0.606**	**0.744**	0.333	0.333
4	34	50	10%	0.344	**0.956**	0.322	**0.356**
			30%	0.352	**0.911**	0.422	**0.433**
			50%	0.387	**0.856**	0.433	**0.444**
			70%	0.480	**0.833**	0.533	**0.556**
5	32	90	10%	**0.356**	**0.956**	**0.356**	0.244
			30%	**0.509**	**0.856**	0.478	0.278
			50%	**0.572**	**0.756**	0.456	0.211
			70%	**0.789**	**0.856**	0.489	0.422
6	28	160	10%	0.339	**0.911**	0.367	0.367
			30%	**0.406**	**0.722**	0.289	0.389
			50%	**0.454**	**0.600**	0.333	0.344
			70%	0.569	**0.667**	0.467	**0.611**
7	22	80	10%	**0.387**	**0.911**	0.356	0.356
			30%	**0.391**	**0.744**	0.344	0.322
			50%	0.443	**0.700**	0.444	**0.467**
			70%	0.480	**0.733**	0.489	**0.544**
8	34	100	10%	0.322	**0.822**	**0.378**	**0.378**
			30%	**0.409**	**0.678**	0.389	0.400
			50%	0.522	**0.678**	**0.533**	0.411
			70%	0.689	**0.744**	0.522	**0.700**
9	23	100	10%	0.333	**0.989**	0.333	**0.378**
			30%	**0.393**	**0.878**	0.389	0.344
			50%	0.383	**0.633**	**0.411**	0.344
			70%	0.398	**0.622**	0.478	0.556
10	28	170	10%	0.333	**1.000**	**0.344**	0.256
			30%	**0.567**	**0.833**	0.511	0.233
			50%	**0.685**	**0.833**	0.511	0.278
			70%	**0.844**	**0.911**	0.611	0.589
Average				**0.462** ± 0.041	**0.826** ± 0.037	0.412 ± 0.027	0.389 ± 0.037

TABLE III: Average precision (p) and recall (r) for individual subjects at different levels of observation (highest in bold).

The results from the t-tests indicate that the average precisions and recalls for different approaches are significantly different from each other at a 95% confidence level (Šidák correction considered). Therefore, on average, the PM-based GR approach significantly outperforms the LSTM and LDA-based approaches in terms of precision and recall. Furthermore, in situations where the system makes mistakes, we analyze the probability gaps, which are defined as the difference between the maximum probability associated with the inferred goals and the probability associated with the true goal. We observe the following average probability gaps: 0.064 for the PM-based system, 0.277 for the LSTM-based approach, and 0.623 for the LDA-based approach. These results indicate that the PM-based GR system exhibits less confidence when making mistakes than the other two benchmarks.

Table IV shows the average precisions and recalls, along with the 95% confidence intervals, across all levels of observation for each subject. Differences in performance are subject-dependent. Subjects 5 and 10 lead to a broader performance gap between all approaches. Still, the PM-based approach outperforms other methods in eight out of ten subjects.

Table V shows the average precisions and recalls across all ten subjects for different levels of observation (10%, 30%, 50%, and 70%), along with the corresponding 95% confidence intervals. The precision of all approaches shows a gradual

Subject	PM		LSTM	LDA
	p	r	$p = r$	$p = r$
1	**0.479** ± 0.035	**0.814** ± 0.040	0.475 ± 0.052	0.464 ± 0.052
2	**0.380** ± 0.020	**0.950** ± 0.023	0.325 ± 0.049	0.294 ± 0.047
3	**0.476** ± 0.034	**0.844** ± 0.038	0.325 ± 0.049	0.333 ± 0.049
4	0.391 ± 0.023	**0.889** ± 0.033	0.428 ± 0.051	**0.447** ± 0.052
5	**0.556** ± 0.037	**0.856** ± 0.036	0.444 ± 0.052	0.289 ± 0.047
6	**0.442** ± 0.039	**0.725** ± 0.046	0.364 ± 0.050	0.428 ± 0.051
7	**0.425** ± 0.034	**0.772** ± 0.044	0.408 ± 0.051	0.422 ± 0.051
8	**0.486** ± 0.040	**0.731** ± 0.046	0.456 ± 0.052	0.472 ± 0.052
9	0.377 ± 0.030	**0.781** ± 0.043	0.403 ± 0.051	**0.406** ± 0.051
10	**0.607** ± 0.037	**0.894** ± 0.032	0.494 ± 0.052	0.339 ± 0.049

TABLE IV: Average precision (p) and recall (r) at all levels of observation for ten individual subjects (highest in bold).

increase as more signals are observed. However, for the PM-based approach, the recall initially starts high, then slightly drops, and finally increases slightly again. The PM-based approach exhibits significantly higher recall than the other approaches at all levels of observation. For precision, the PM-based approach is comparable to the other approaches at 10% and 30%, and significantly higher at 50% and 70% levels of observation. Note that LDA performs well when the full trace is observed, achieving a precision (p) and recall (r) of $p = r = 0.709 \pm 0.030$. In comparison, the PM approach achieves a precision of $p = 0.645 \pm 0.026$ and a recall of $r = 0.793 \pm 0.027$. This is due to the high similarity of LDA's training data with the observed signal at the end of the trace, when the subject already achieved the target pose.

Obs%	PM		LSTM	LDA
	p	r	$p = r$	$p = r$
10%	0.344 ± 0.009	**0.946** ± 0.015	**0.348** ± 0.031	0.336 ± 0.031
30%	**0.426** ± 0.020	**0.813** ± 0.026	0.390 ± 0.032	0.346 ± 0.031
50%	**0.486** ± 0.024	**0.754** ± 0.028	0.427 ± 0.032	0.353 ± 0.031
70%	**0.592** ± 0.026	**0.789** ± 0.027	0.484 ± 0.033	0.523 ± 0.033

TABLE V: Average precision (p) and recall (r) for all subjects at 10%, 30%, 50% and 70% observation (highest in bold).

VI. LIMITATIONS & FUTURE WORK

While our work provides a novel and promising approach to goal recognition for transhumeral prostheses, several aspects can be further elaborated. Although we used existing datasets for comparison with the state-of-the-art, real-world deployment may require more extensive experiments, including a larger number of subjects (currently 10) and a larger number of (and more complex) target poses and goals (currently three).

Our experiments confirmed that feature selection and discretization have a significant impact on the accuracy of the system (in our case, of the GR outcome). Our current approach is arguably simple and we would like to explore elaborate feature selection and event discretization. For example, it would be interesting to identify feature conditions capturing *meaningful* prosthetic postures/configurations, and test whether sequences of these yield good predictor for the goal being pursued. Doing so will also allow us to take advantage of one of the key features of our PM-based GR approach, namely, the possibility to explain the outcome of the system based on the process model and misalignment of the observed behavior.

Yet another area for further exploration is the testing of alternatives alignment approaches. While process discovery can

be done in linear time [4], alignment techniques—including the one we used in this work [5]—are often exponential in the worst case. Since our approach does require alignments to be computed at recognition time, extracting them as fast as possible is important for real-time applicability. However, the worst cases often do not come up in practice, and our alignment system took around 0.04 seconds per goal (note alignments across goals can be parallelized). Nonetheless, we would like to experiment with techniques that are specifically designed for *online* conformance checking [26] as well as those that seek *approximate* alignments, but *fast* [27]. Indeed, we conjecture online conformance checking approaches may be a good fit for our GR setting, since observations are incrementally extended and such approaches extract alignments incrementally by reusing previously computed ones.

The used dataset captured the features from forward-reaching tasks using sound limbs rather than real-time control of a prosthesis. The difference between real-time prosthetic states and sound limbs introduces variations in visual feedback, which could affect the feature patterns collected during movement. The feature patterns can have an impact on the real-time recognition accuracy of machine-learning-based techniques [28], [29]. Therefore, collecting data from real-time control experiments and pre-defining standard traces toward the goal with distinguishable feature patterns for training the PM-based GR system has the potential to improve accuracy.

VII. Conclusion

In this paper, we proposed a novel PM-based GR method to recognize a patient's intended target pose when moving an artificial limb, utilizing continuous sensor data. Our approach transforms the data into discrete events to discover a behavior model and subsequently predict the goal by aligning the observed behavior with the model. Experiments on a real-life dataset demonstrate the effectiveness of our proposed GR approach, which significantly outperforms state-of-the-art LSTM-based and LDA-based machine learning methods on both precision and recall. In addition, our GR approach is less confident when wrong, which results in a much smoother movement of the prostheses in less certain scenarios.

References

[1] R. Mirsky, S. Keren, and C. Geib, "Introduction to symbolic plan and goal recognition," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 16, no. 1, 2021.

[2] A. Polyvyanyy, Z. Su, N. Lipovetzky, and S. Sardiña, "Goal recognition using off-the-shelf process mining techniques," in *AAMAS*, 2020.

[3] Z. Su, A. Polyvyanyy, N. Lipovetzky, S. Sardiña, and N. van Beest, "Fast and accurate data-driven goal recognition using process mining techniques," *Artificial Intelligence*, vol. 323, 2023.

[4] S. Leemans, E. Poppe, and M. T. Wynn, "Directly follows-based process mining: Exploration & a case study," in *ICPM*, 2019.

[5] W. van der Aalst, A. Adriansyah, and B. van Dongen, "Replaying history on process models for conformance checking and performance analysis," *WIREs Data Mining & Knowledge Discovery*, vol. 2, no. 2, 2012.

[6] C. Brewitt, B. Gyevnar, S. Garcin, and S. V. Albrecht, "GRIT: Fast, interpretable, and verifiable goal recognition with learned decision trees for autonomous driving," in *IROS*, 2021.

[7] S. V. Albrecht, C. Brewitt, J. Wilhelm, B. Gyevnar, F. Eiras, M. Dobre, and S. Ramamoorthy, "Interpretable goal-based prediction and planning for autonomous driving," in *ICRA*, 2021.

[8] Y. Demiris, "Prediction of intent in robotics and multi-agent systems," *Cognitive Processing*, vol. 8, 2007.

[9] A. Kulkarni, S. Sreedharan, S. Keren, T. Chakraborti, D. E. Smith, and S. Kambhampati, "Designing environments conducive to interpretable robot behavior," in *IROS*, 2020.

[10] C. Rich, C. L. Sidner, and N. Lesh, "COLLAGEN: Applying collaborative discourse theory to human-computer interaction," *AI Magazine*, vol. 22, no. 4, 2001.

[11] M. Legrand, C. Marchand, F. Richer, A. Touillet, N. Martinet, J. Paysant, G. Morel, and N. Jarrassé, "Simultaneous control of 2DOF upper-limb prosthesis with body compensations-based control: A multiple cases study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, 2022.

[12] T. Yu, A. Mohammadi, Y. Tan, P. Choong, and D. Oetomo, "Sensor selection with composite features in identifying user-intended poses for human-prosthetic interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, 2023.

[13] T. Yu, R. Garcia-Rosas, A. Mohammadi, Y. Tan, P. Choong, and D. Oetomo, "Comparing the outcomes of population-averaged and personalised input feature selection for transhumeral prosthetic interfaces," in *SMC*, 2021.

[14] J. Huang, G. Li, H. Su, and Z. Li, "Development and continuous control of an intelligent upper-limb neuroprosthesis for reach and grasp motions using biological signals," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 6, 2021.

[15] L. J. Hargrove, E. J. Scheme, K. B. Englehart, and B. S. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 1, 2010.

[16] A. W. Shehata, H. E. Williams, J. S. Hebert, and P. M. Pilarski, "Machine learning for the control of prosthetic arms: Using electromyographic signals for improved performance," *IEEE Signal Processing Magazine*, vol. 38, no. 4, 2021.

[17] H. A. Kautz and J. F. Allen, "Generalized plan recognition," in *AAAI*. Morgan Kaufmann, 1986.

[18] M. Ramírez and H. Geffner, "Probabilistic plan recognition using off-the-shelf classical planners," in *AAAI*, 2010.

[19] W. Min, B. W. Mott, J. P. Rowe, B. Liu, and J. C. Lester, "Player goal recognition in open-world digital games with long short-term memory networks," in *IJCAI*, 2016.

[20] P. Geethanjali, "Myoelectric control of prosthetic hands: State-of-the-art review," *Medical Devices: Evidence and Research*, vol. 9, 2016.

[21] N. Parajuli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. J. Hamilton, G. R. Naik, U. Gunawardana, and G. D. Gargiulo, "Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Sensors*, vol. 19, no. 20, 2019.

[22] W. H. E. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of Classification*, vol. 1, 1984.

[23] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, 1979.

[24] J. H. Friedman and N. I. Fisher, "Bump hunting in high-dimensional data," *Statistics and Computing*, vol. 9, 1999.

[25] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, 2003.

[26] S. van Zelst, A. Bolt, M. Hassani, B. van Dongen, and W. van der Aalst, "Online conformance checking: Relating event streams to process models using prefix-alignments," *International Journal of Data Science and Analytics*, vol. 8, 2019.

[27] B. van Dongen, J. Carmona, T. Chatain, and F. Taymouri, "Aligning modeled and observed behavior: A compromise between computation complexity and quality," in *CAiSE*. Springer, 2017.

[28] R. B. Woodward and L. J. Hargrove, "Adapting myoelectric control in real-time using a virtual environment," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 11, 2019.

[29] M. Ortiz-Catalan, F. Rouhani, R. Brånemark, and B. Håkansson, "Offline accuracy: A potentially misleading metric in myoelectric pattern recognition for prosthetic control," in *EMBC*, 2015.