Predicting Metabolic Cost During Human-in-the-Loop Optimization

Erez Krimsky
ekrimsky@stanford.edu

Eley Ng
eleyng@stanford.edu

***Abstract*— As humans have naturally optimized their movement towards a metabolic minimum, it is difficult to improve human walking economy or assist individuals with ambulatory deficiencies with simple control strategies. Human-in-the-Loop Optimization (HILO) is a technique using assistive robotic devices to augment human walking performance to overcome this challenge. A common metric used to determine human performance is metabolic cost, which is the amount of energy used by the human to perform a certain task. However, metabolic cost yields noisy measurements, is slow to respond, and is very difficult to measure outside of the laboratory environment. Therefore, we are interested in predicting metabolic cost using human walking data collected during human-in-the-loop optimization experiments. By approximating metabolic cost with other features that can be measured outside the laboratory environment, we may be able to design more robust and diverse assistance adaptation experiments in the future. Using a simple neural network implementation with a single hidden layer, trained on multiple days of walking data, we were able to achieve across-validated prediction accuracy with an MSE as low as 0.0089 on a data set with a variance of 0.061. The inherent noise in metabolic data sets a minimum achievable MSE of around 0.0025.**

I. INTRODUCTION

Humans are naturally skilled at walking, since evolution has shaped our physiology to naturally optimize motor control [1]. However, human performance can still be augmented or improved, especially for people with impairments that inhibit walking ability. Ankle exoskeleton research has made significant progress on reducing the weight of the hardware, as well as developing control algorithms for assistance. In particular, researchers have noticed significant improvements in human energy use when including human feedback in control algorithm design and when control algorithms are customized for individuals [2]. Recent developments in control algorithms for assistive devices use various optimization algorithms, including gradient descent [3], Covariance Matrix Adaptation-Evolutionary Strategy (CMA-ES) [2], and Bayesian Optimization [4] to find the best set of control law parameters for a particular individual and hardware combination. These techniques use online measurements of metabolic cost to define the cost function over which to optimize.

Metabolic cost is a measurement of rate of the energy required to perform a certain task at the tissue level. Energy cost is optimized naturally by animals, and is a useful metric when determining whether a motor skill is being performed optimally. Collecting metabolic cost data, however, is very difficult and limiting. Firstly, it cannot be easily measured outside of a laboratory setting and requires subjects to wear a restricting mask in during experiments. Subjects are also required to fast for a few hours prior to recording metabolic data which may be especially difficult for long trials. Metabolic cost is slow to change as the human adapts to a new control law, making it difficult to use as feedback for control parameters. It can take over a minute for metabolic data to reach a steady state value for any particular task [5]. Metabolic cost is also a very noisy measurement and typically many filters must be applied before the signal can be determinable.

Further complications of human-in-the-loop optimization involve the adaptation of the human to the hardware itself. Naïve users of the exoskeleton likely perform at a different level than expert users of the exoskeleton, and therefore different optimal control laws are applicable for individuals of varying expertise. Naïve exoskeleton users are simultaneously training and improving their expertise while the control parameters of the device are being optimized for their current level of expertise. During training, the optimal control parameters shift, framing the simultaneous training and optimization as a non-stationary process.

Due of the success of stochastic optimization algorithms in providing optimal control parameters and improving human energy consumption [2], [4], we would like to increase the effectiveness of Human-in-the-Loop optimization by improving the experimental process. In particular, we seek to use various metrics to predict metabolic cost such that human experimental studies can be made more feasible. The benefits of predicting metabolic cost extend to being able to run long experimental studies on humans, which could be taken outside of the lab and therefore inconvenience the human subject to a lesser degree.

Recent work has shown that many biological gait parameters can be robustly measured outside of the laboratory environment. These include, but are not limited to, joint angles [10], ground reaction forces [7], gait patterns [6], and surface electromyography (EMG) [11]. Using other outputs as a proxy for metabolic data could allow for faster convergence of laboratory experiments as metabolic data is slow to change, but could also allow for studies to be carried out for significantly longer by not constraining all experimentation to a laboratory setting. Furthermore, removing experiments from a laboratory setting allows assistive devices to be tested in the environments where they would likely be used, such as uneven terrain or inclines.

II. METHODS

An ongoing study in the Stanford Biomechatronics Laboratory examining the effects of simultaneous training and

optimization using bilateral ankle exoskeletons is the source of all the data presented here. Only naive subjects are used in order to study the effects of training. The control strategy, which was used in the human-in-the-loop optimization study by Zhang, et. al. [2], is dictated by 4 parameters that describe the dynamic torque applied at each ankle by the exoskeleton hardware. These control parameters include peak time, rise time, peak magnitude, and settling time of the torque applied at each ankle. The current study includes data collection of metabolic rate, ground reaction force, ankle angle, and 8 surface EMG sensors per leg giving muscle activations for both upper and lower leg muscle groups, all collected over several optimization trials, which are 72 minutes in length, as well as over several static (constant) control trials and zero-torque trials. The 72 minute trial consist of 36 different control schemes where the control scheme changes every 2 minutes. This data is collected over 6 different days of walking in the exoskeleton for each subject. During these optimization trials, the human subject learns to walk in the exoskeleton while a CMA-ES optimizer determines the next generation's set of optimal control laws to apply using feedback with the data collected. Up to 8 generations of optimal control laws are identified per subject.

A. Data Processing

1) Metabolic Data: Baseline metabolic measurements are different for every person and can shift significantly from day to day. For this reason, metabolic data is recorded for one 6 minute standing trial and two 6 minute normal walking trials. The mean metabolic rate from the standing trial is considered to be baseline for that day of experiments and is subtracted off from all other metabolic data from that day. The metabolic measurements from the first half of the normal walking trial are discarded as participants metabolic rate is still increasing from the prior inactivity and the average metabolic rate from the remainder of the normal walking trials is used as a normalizing factor for the metabolic data taken from that day. Measurements are only taken from the last 30 seconds of each 2 minute sub-trial to allow for the subject time to adjust to the new control parameters.

2) Step Data: Using signals from heel switches mounted in the shoes and ground reaction forces from the treadmill the gait phase can be easily determined. For each 30 second interval of data recorded, the following features were extracted from the raw data on a per-step basis:

- Peak vertical reaction force before toe-off
- Maximum/minimum ankle angle
- Stride Width
- Stride Time

As a very short or very long stride may significantly shift the mean for these parameters, the median of each step parameter of the 30 second period is chosen instead. Ground reaction forces are normalized by subject's weight on the day of the experiment to account for small changes in weight between days.

3) EMG: As EMG signals are noisy AC signals, the signal mean is not a well correlated to the muscle activations. A standard approach to processing EMG signals is to run the data through a high-pass butterworth filter followed by rectifying the signal and filtering with a secondary low-pass butterworth filter. The RMS of the resulting signal can then be correlated with the magnitude of muscle activations.

Although EMG sensors are placed in approximately the same locations on every day in the experiment, small deviations in sensor location can lead to large changes in the magnitude of the electrical signal. To account for this, EMG signals are recorded for two 6 minute periods of normal walking on each day. After initial processing as described above, the peak value for each signal is calculated. These peak values are then used as normalizing factors for each signal during the optimization trials.

B. Algorithms

1) Curve Fitting: To set a baseline prediction value to compare against a neural network implementation, we investigate a simple linear regression model for predicting the metabolic cost. As the total number of features (29) is on a similar order of magnitude to the number of data points per subject (≈ 180) standard linear regression is prone to overfitting. To account for this we use lasso regularization and apply *k-fold* cross-validation with $k = 10$ to determine the proper value for the regularization term.

To attempt improve on the prediction accuracy of the linear regression we use a standard neural network with one hidden layer, a *tanh* activation function between the hidden layer and the output layer and a linear activation on the output. As the data set is small and inherently noisy from metabolic measurements, we found that we were not able to create statistically meaningful train, validation and testing splits from the data and instead ran a *k-fold* cross validation to tune the network and choose an appropriate value for the number of neurons in the hidden layer. We performed our network training using a Bayesian Regularization algorithm. Although significantly slower than other training algorithms, Bayesian Regularization has been shown to be more robust at avoiding overfitting for noisy and small data sets [13].

The features can be broken up into three categories: 1) step data, 2) EMG data, and 3) exoskeleton control parameters. To investigate which category of features are most important for making accurate predictions, the cross validation described above for linear regression and network tuning was performed four times on using 1) all features, 2) step & EMG features, 3) step only features and 4) EMG only features.

2) Dimensionality Reduction: The number of input features characterizing our dataset is quite large compared to the amount of samples. In total, we have 29 features (including controls), and 180 data points. Therefore, it would make sense to reduce the dimensionality of our features to prevent overfitting. We used forward step-wise selection to select features in a reduced model, and perform cross-validation with the new models to select the best subset of features

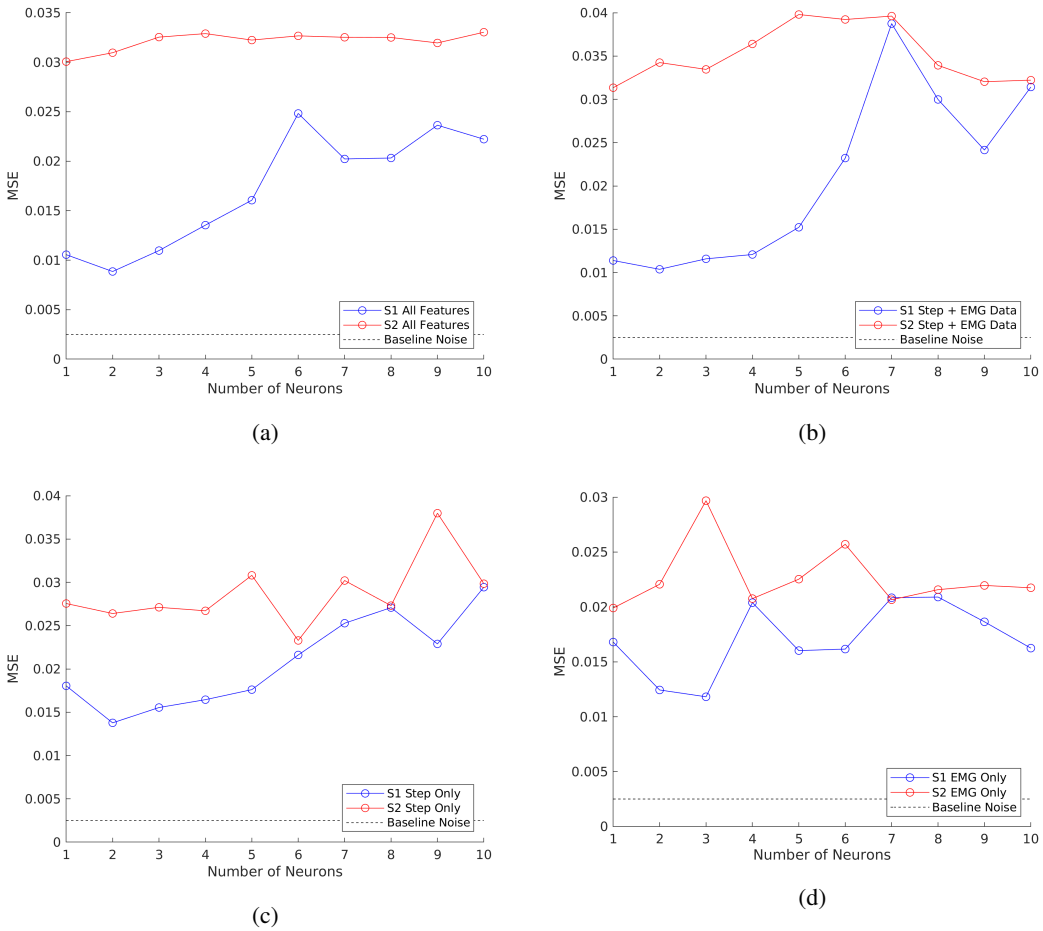

Fig. 1: The black dashed line above represents the minimum achievable MSE based off inherent noise in metabolic data. Comparisons of training networks on data from two subjects using (a) all features, (b) step and EMG features, (c) step features only and (d) EMG features only.

using the network tuned in the k-fold cross validation step. Due to the variability of forward step-wise selection, we ran several trials to determine the features that are consistently significant across trials. For more stable results, we also used Principal Components Analysis (PCA) after normalizing our dataset to compare model selection methods.

III. RESULTS

The mean MSE values from the *k-fold* cross validation can be seen in Figure 1. For all 4 parameter sets the network was trained on, there did not seem to be any significant benefit in using more than 3 neurons. For some, the minimum average MSE value was achieved with only 4 neurons. The MSE values for the "best" network for each set are summarized in Table I. As both the linear regression fitting and network training were performed with cross-validation and regularization, we feel confident that we have avoided overfitting the data.

It is note worthy that the predictions on Subject 1 consistently had lower MSE values than the predictions on Subject 2. One explanation for this is that the data for Subject 1 exhibits much higher variance than the data for Subject 2, as can be seen in the "Constant Prediction MSE" column of Table I. As the changes in metabolic data over the course of experiments were much more drastic for Subject 1 than for Subject 2, the data for Subject 1 may have been able to better capture the complex relationship between the features and metabolic data.

Over several runs, we selected features which were present in 60% of the trials. Results from running the forward stepwise selection algorithm using our tuned network on the no-controls data for Subject 1 resulted in the selection of 23 features for a new model (out of the 29 total features). This new model predicted with an MSE of 0.0761. We also ran PCA using our tuned network on our data for Subject 1, which gave us more stable results. PCA with 23 features captured 98.8% of the variance of the data. The new features found using PCA predicted with an MSE of 0.0700.

IV. DISCUSSION

Our initial approach to the neural network fitting consisted of a network with same basic architecture as the one described above but with more neurons in the hidden layer. We initially experimented with 2 to 10 neurons in the hidden

TABLE I: Summary of prediction accuracies for *k-fold* cross validation on network training and linear regression. *Constant Prediction MSE* refers to the MSE value that would be achieved by simply predicting the mean of all the observations.

	Constant Prediction MSE	All Features		Step & EMG		Step Only		EMG Only	
		Lin. Reg	NN	Lin. Reg.	NN	Lin. Reg	NN	Lin. Reg.	NN
Subject 1	0.0657	0.0089	0.0089	0.0089	0.0104	0.0200	0.0138	0.0158	0.0118
Subject 2	0.0465	0.0176	0.0301	0.0168	0.0313	0.0217	0.0233	0.0203	0.0199

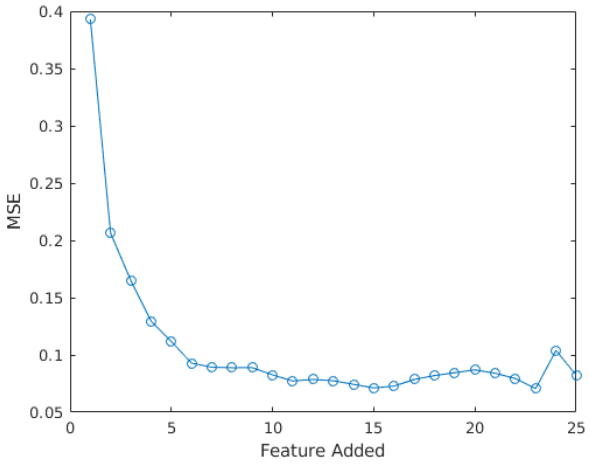

Fig. 2: MSE over the feature added during forward stepwise selection.

layer using a standard Levenberg-Marquardt algorithm to update the weights. In our initial configuration we split the data randomly in separate train, validation, and testing sets. The validation sets were initially used to prevent overfitting and halt the training at the epoch that led to increased error on the validation set. Although this approach at times led to tantalizingly low MSE values on the validation and test set, we found that retraining with slightly different splits lead to vastly different outcomes. This led us to the approach described in the methods section of using both regularization and cross-validation to ensure consistency of the outcomes and report values that are actually indicative of the predictive power of the algorithm. Data from all days of the study was pooled and split randomly for cross-validation due to the limited number of points. Ideally, a better test of the predictive power of the algorithms would be to make predictions on a a full day of data that the algorithm has not seen.

In order for these results to be generalizable and useful for other exoskeleton optimization studies, it would be ideal if we could achieve high accuracy predictions without any knowledge of the control parameters, since the structure of the control parameters and an individual's response to them may differ between devices. Table I shows that including control parameters does slightly reduce MSE, but very similar values can be achieved using only step and EMG data. For both subjects the EMG data is a significantly better predictor of metabolic cost than step data. Significantly cleaner EMG signals are typically achieved with individuals who have greater muscle definition, which may be the reason for the discrepancy in EMG predictions seen between subjects. When the EMG signal is particularly noisy, we may simply require much more training data than was available for this study.

The dimensionality of our features were reduced using PCA and forward stepwise selection methods. Running the forward stepwise selection algorithm many times allowed us to select the features which were deemed significant more consistently. However, testing these new models resulted in relatively high MSE values. The model built using forward stepwise selection varied over many trials due to the greedy nature of the algorithm; the one we selected to produce Figure 2 was built from selecting features that showed importance across trials and is most likely not an optimal model, as shown by its poor predictive performance. PCA separates the data according to the number of components, attempting to maximize the variance in each; however, it does not take the prediction targets into account, which may explain why the predictions using the newly defined features were poor.

After testing different datasets, we found that all of the models considered perform with relatively the same accuracy using data with controls compared to data without controls as features. It can also be determined that the EMG data can be significantly better predictors than the step features alone. Therefore, we could potentially do well in predicting metabolic cost using EMG alone.

The results of the predictions and model selection process show that it is difficult to use the data to predict the metabolic cost of different individuals. In the feature reduction stage, some features were more significant for an individual compared to others and therefore resulted in different features selected. For instance, for Subject 1, step width was an important feature that showed up much more often in the trials, and for Subject 2, left and right peak force were consistently significant. Another observation from the feature selection step was that for certain individuals and for certain features which consider both the left and right sides of the body (e.g. right and left ground force), features from mostly one side appeared as significant features. This may imply that capturing the same type of data from both legs may be redundant and unneccesary. A potential improvement to the dataset could therefore be individual-specific features, which, in addition to collecting data on more individuals, could improve generalization of the predictions.

The variations in predictions between the two subjects indicates that trying to create a network to generalize these

predictions for multiple individuals would likely yield poor predictions. However, with significantly more subjects and the inclusion of subject specific parameters such as height, age, weight, and general fitness, it may be possible to create a generalized predictor model. Additional kinematic data could also be collected during experiments using wearable Inertial Measurement Units (IMUs), which would augment the feature space and potentially improve accuracy.

V. CONCLUSIONS & FUTURE WORK

In the best case, our algorithms were able to achieve an MSE values on the normalized metabolic rate of 0.0089 corresponding to approximately 10% error. This is a significant improvement over a constant prediction, which would yield an error of approximately 25%. It is also a significant step towards the goal of 5% accuracy, which can be considered as the baseline noise level for metabolic data.

Future work would include significantly more subjects in the study and perhaps the inclusion of more kinematic data either from IMUs or from a motion capture system. As further increasing the feature space will increase our chances of overfitting, a significantly larger volume of data would need to be recorded. Currently, the large number of sensors used in this experiment may be prohibitive for future work. Given the quality of EMG based predictions, it would be value to identify which EMG signals contribute the most to the predictions.

ACKNOWLEDGEMENTS

We would like to acknowledge Zhang et al, [2] as well as Thu Nguyen, Katherine Poggensee, and Guan Rong Tan [9] for providing the inspiration for this study. We would specifically like to thank Katherine Poggensee for providing experimental data.

CONTRIBUTIONS

Eley Ng performed the model selection work. Erez Krimsky processed the data and worked on curve fitting. Both contributed equally to the network tuning, ideation, and writing of this report.

GITHUB REPOSITORY

The code used in this project can be found at https://github.com/ngeley/cs229project.

REFERENCES

[1] C. O. Lovejoy, The natural history of human gait and posture: Part 1.spine and pelvis,Gait & posture, vol. 21, no. 1, pp. 95112, 2005.

[2] J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee,C. G. Atkeson, and S. H. Collins, "Human-in-the-loop optimization of exoskeleton assistance during walking," Science, vol. 356, no. 6344,pp. 12801284, 2017.

[3] Felt W, Selinger JC, Donelan JM, Remy CD. Body-In-The-Loop: Optimizing Device Parameters Using Measures of Instantaneous Energetic Cost. PloS one. 2015;10(8):e0135342. pmid:26288361

[4] Y. Ding, M. Kim, S. Kuindersma, and C. J. Walsh, Human-in-the-loop optimization of hip assistance with a soft exosuit during walking, Science Robotics, vol. 3, no. 15, pp. eaar5438, 2018.

[5] Felt W, Selinger JC, Donelan JM, Remy CD. Body-In-The-Loop: Optimizing Device Parameters Using Measures of Instantaneous Energetic Cost. PloS one. 2015;10(8):e0135342. pmid:26288361

[6] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs and J. A. Paradiso,"Gait Analysis Using a Shoe-Integrated Wireless Sensor System," in IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 4, pp. 413-423, July 2008. doi: 10.1109/TITB.2007.899493

[7] P. H. Veltink, C. Liedtke, E. Droog and H. van der Kooij, "Ambulatory measurement of ground reaction forces," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 13, no. 3, pp. 423-427, Sept. 2005. doi: 10.1109/TNSRE.2005.847359

[8] Tibor Hortobgyi, Adria Finch, Stanislaw Solnik, Patrick Rider, Paul DeVita, "Association Between Muscle Activation and Metabolic Cost of Walking in Young and Old Adults," The Journals of Gerontology: Series A, Volume 66A, Issue 5, 1 May 2011, Pages 541547. https://doi.org/10.1093/gerona/glr008

[9] T. Nguyen, K. Poggensee, G.R., "Improving Human-in-the-Loop Optimization Algorithms using Machine Learning Technique", poster, CS229 final project, December 2017.

[10] Asli Atalay, Vanessa Sanchez, Ozgur Atalay, Daniel M. Vogt, Florian Haufe, Robert J. Wood, and Conor J. Walsh, "Batch Fabrication of Customizable Silicone-Textile Composite Capacitive Strain Sensors for Human Motion Tracking," Advanced Materials Technologies, 2018.

[11] Albert Chi. Sawyer Smith, Isaac Womack, Robert Armiger, "The Evolution of Man and Machinea Review of Current Surgical Techniques and Cutting Technologies After Upper Extremity Amputation", The Military Perspective, 2018.

[12] Young, Aaron Hargrove, Levi. (2015). A Classification Method for User-Independent Intent Recognition for Transfemoral Amputees Using Powered Lower Limb Prostheses. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society. 24. 10.1109/TNSRE.2015.2412461.

[13] David J.C. MacKay, "Bayesian Interpolation", Neural Computation, 1991