Feature selection for grasp classification

Lillian Chang
November 29, 2006
10-701 project final report

Abstract

Although the human hand is a complex biomechanical system, functional grasps may be described by small set of features. Supervised feature selection is used to evaluate the performance of reduced marker sets for grasp classification from motion capture data. Our reduced feature set maintains 85% grasp classification accuracy, compared to 90% accuracy from using the full 30 marker set. Using a linear classifier and as few as 5 surface markers allows for dramatic simplification of the experimental procedure and reduced computational cost for grasp classification.

1 Introduction

The human hand has several degrees of freedom, which provide it amazing flexibility as a manipulator, but pose challenges for measuring and modeling hand movement. One technique for measuring the human motion is motion capture, where optical markers attached to body segments are used to reconstruct joint movement. Motion capture of the hands is difficult because of marker occlusions due to the wide range of hand poses and close position of the fingers.

A reduced marker set would simplify the capture procedure and also describe hand configuration in a low-dimensional space. Furthermore, there is no standardized motion capture protocol for the hand. The aim of this study is to evaluate how the selected marker positions on the hand can affect the measured motion and investigate which markers may be the best to include in a reduced hand marker protocol for future experiments.

2 Related work

Previous literature in biomechanics has also investigated how to represent grasping and reach-to-grasping actions in a low-dimensional feature space using principal components analysis. For example, [1, 2] suggest that hand grasping motions can be represented by just a few principal components in the joint space. These studies used a data glove to capture the motion of the hand, and the experiments involved mimed hand motion or reach-to-grasp motion before the hand was in contact with the object.

Other work in the robotics community has investigated grasp classification. One study [3] used neural networks for predicting grasps, based on a taxonomy proposed in [4]. The experiments focused on only three object shapes and sizes. Hidden Markov Models have also been used for grasp recognition from data glove measurements [5]. Data gloves can be cumbersome to the user, may affect the natural grasping motion, and do not always fit individual subjects well. In our study, we focus on input data in the context of marker-based methods for motion capture and examine how to design an appropriate protocol which can simplify the data acquisition procedure.

3 Problem definition

This study investigates feature selection of the marker position inputs in conjunction with grasp classification. The classification goal is to predict the grasp at a single time frame given the measured marker positions representing the hand configuration. The purpose of feature selection is to evaluate which markers are the best predictors of the grasp class and determine which markers could be eliminated to simplify the marker protocol. We aim to explore questions such as: What is the minimum number of markers needed to represent the hand pose well, and where on the hand should those markers be placed? The redundancy in hand grasping motion found by [1, 2] suggests that the number of markers could be dramatically reduced without severely compromising grasp classification.

4 Proposed method

4.1 Grasp classification

Our approach uses linear classifiers for predicting grasp, which will then be combined with supervised feature selection. Although the fingers do exhibit nonlinear kinematics relative to the palm, the constraints on hand motion will limit each surface marker to a small set of clustered reachable positions. Different grasps are characterized by the relationships between marker positions, which will vary, but not in a severely nonlinear way. Thus, we believe that a classifier with linear decision boundaries can be successful for predicting grasp types from motion capture data. In addition, using linear classifiers can provide a simpler implementation and reduced computational cost compared to neural networks and Hidden Markov Models, as used in [3, 5].

We will evaluate three candidates for a baseline classifier which uses the full feature set for predicting grasps: Gaussian Naive Bayes (GNB) with class-independent variances, multi-class logistic regression (LR), and linear support vector machines (SVM). We expect that GNB will be less successful than LR and SVM due to the assumptions of conditional independence and Gaussian distribution of the marker coordinates, which are unlikely to be satisfied in our case of coordinated hand movement.

4.2 Supervised feature selection

Given a single baseline classifier, we wish to select a subset of features which simplifies the model. We use two standard approaches for supervised feature selection, as described in [6]. First, in the filter approach, single features are individually ranked by a scoring criterion, and the reduced set consists of the k best scoring features. The advantage of this method is that each feature need only be scored once, with the expense that the final selection does not consider possible interdependencies between the features. We evaluate two scoring criteria: (1) the mutual information between the target value and a single feature, and (2) the prediction accuracy of a single feature classifier on a validation test set. To compute the mutual information with the discrete target value, the continuous marker position features are discretized rather than fit to an assumed distribution.

Wrapper methods are a second approach for feature selection. In contrast to the filter approach, wrapper algorithms consider the interaction between features in constructing the reduced set. By modeling how the set of features is related to the target attribute rather than only the relation between each single feature with the target, wrapper methods can potentially select a feature set of the same size which results in better prediction. However, this requires additional computational cost for training, as the features must be re-scored each time the current feature set changes. To avoid considering the exponential number of possible feature sets, the simplest wrapper algorithms consider a single feature at a time for locally-optimal feature selection. In this work, we will consider two versions of greedy wrappers. The forward method adds features incrementally to a reduced feature set, and the backward method discards features incrementally from a larger set of available features.

We make one modification to these standard algorithms for the motion capture application. Although the marker coordinates are represented by individual $x, y,$ and z features, which could be scored and added individually, our methods will add the features in subsets of three which correspond to the three coordinates of one marker. This is more useful for the practical application, where the goal is to reduce the number of markers in the protocol, instead of using one coordinate and ignoring the other two of the same marker. Thus the ranking method will be modified to score a single marker from its three features.

5 Experiments and results

The data set consists of labeled hand grasps of various objects, where the marker positions are described in local coordinates with respect to the back of the hand. Each example consists of a 90-dimensional vector which represents the 3D positions of 30 markers on a single subject's right hand at a single time frame of a grasp. Six grasp classes were considered, selected from the functional grasps for daily living [7]. Power grasps, characterized by large contact areas, included cylindrical grasp, spherical grasp, and lumbrical grasp. Precision grasps, for fine manipulation by the fingertips, included pinch grasp, tripod grasp, and handwriting grasp.

The available data is separated into three sets. The first data set has over 40,000 frames from 88 motion clips of grasps for 38 objects. Due to the volume of data and limited computational resources, training will generally only use the first data set. The second data set consists of frames from a separate set of motion clips where the demonstrator grasped the same 38 objects as the first data set. The second data set is used as a validation set in the early stage of model selection and for training the final selected classifier. The final data set represents grasps for 19 new objects and will be used as a held out test set for evaluating the final selected methods.

5.1 Baseline classifier

Our tests of the three linear classifiers aim to evaluate the grasp classification accuracy when the entire feature set is used. We implemented GNB with class-independent variances and multi-class LR with cross entropy error and softmax function for class probabilities. Parameters for LR were estimated from maximum conditional likelihood using gradient ascent. Software available from [8] was used for linear SVM. The training data in the first data set was split into two folds, and the selected model for each of three methods was chosen based on the cross-validation accuracy. Table 1 reports the cross-validation accuracy of the three classifiers from two-fold cross-validation. The classifiers were also evaluated on the second data set with grasps on the same objects but from different motion clips.

GNB had the worst performance, as expected due to its restrictive modeling assumptions. Although SVM had over 98% cross-validation accuracy on the first data set used for training, its performance was similar to that of LR for the second data set used as an additional

Table 1: Performance of three linear classifiers on the full feature set.

Evaluation metric	**GNB**	**multi-class LR**	**linear SVM**
Cross-validation accuracy on training data	0.7149	0.9120	0.9881
Accuracy on validation set with same objects	0.7251	0.9043	0.9066

validation test. The linear SVM model may have been overfit and only had high cross-validation accuracy because the training and validation examples in the first training data set came from the same motion clip sources. Because of the comparable performance on the second data set and shorter training time, multi-class LR was selected as the baseline classifier for the further investigations of feature selection.

5.1.1 Regularization parameter selection

To prevent over-fitting, the maximum conditional likelihood cost function for logistic regression can be modified by an additional term which penalizes large weights on the features. The regularization parameter λ determines the penalty on the weight vector magnitude relative to the data log likelihood. Several values of λ are evaluated by 2-fold cross validation accuracy on the first training data set and validation accuracy on the second data set. The results in Figure 1 suggest that regularization may not improve the performance of LR significantly, if at all. Based on these tests, the following experiments on feature selection will use multiclass LR without regularization, $\lambda = 0$.

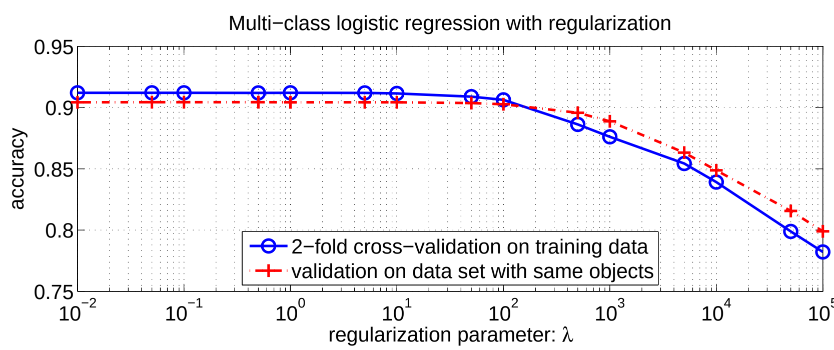


Figure 1: Classification accuracy for logistic regression with regularization. Penalty on the weights does not improve the classifier performance noticeably, if at all.

5.2 Supervised feature selection

5.2.1 Filter methods

Given the chosen baseline classifier, we then use supervised feature selection to examine the rate of trade-off between accuracy and using fewer markers. Filter methods are the simplest approach to investigating this trade-off, and also can be used to select an appropriate scoring criterion for use in the more computationally-expensive wrapper methods. Two popular scores used are the mutual information between a single feature and the target attribute, and the single feature classifier accuracy. We consider five methods of scoring an individual marker, which is a set of three x, y, z coordinate features:

- the maximum single mutual information of any of its three coordinates,
- the sum of the mutual information scores of all three coordinates,
- the maximum single-coordinate LR prediction accuracy of its three coordinates,
- the sum of the three single-coordinate LR prediction accuracies, and
- the single marker LR prediction accuracy from training on all three coordinates.

The markers are first ranked according to each of the scoring methods. For each feature set size k, a LR classifier is trained based on the top k scoring markers and evaluated using two-fold cross validation. Figure 2 shows that regardless of the scoring method, grasp classification accuracy generally increases steeply for a small number of markers but plateaus after about 15 markers. This suggests that the motion capture protocol could be simplified dramatically with small loss in classification accuracy. Scoring based on prediction accuracy was only slightly more successful than using the mutual information gain. The last scoring method is generally the most successful for small set sizes of 3-7 markers. This is probably because the single marker prediction accuracy does consider interaction between coordinates of the same marker.

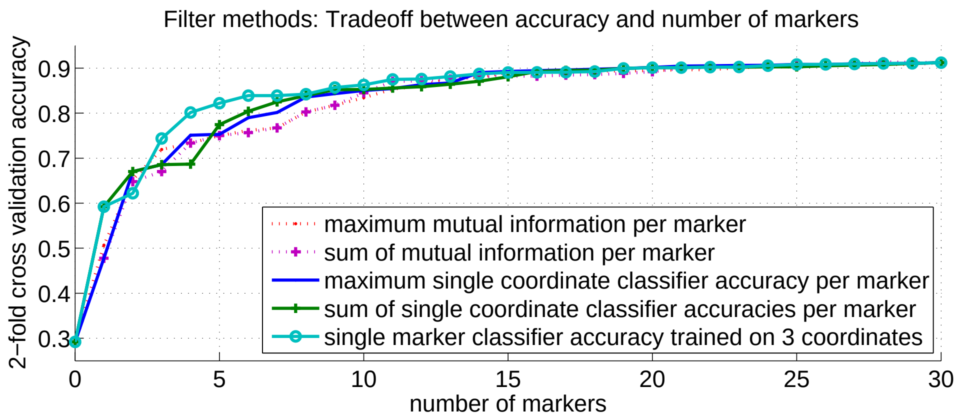


Figure 2: Filter methods: Feature selection based on individual marker scoring criterion.

5.2.2 Wrapper methods

Wrapper methods provide a framework for selecting a reduced feature set which can account for interaction not only between coordinates of the same marker, but between different markers as well. For either greedy addition or removal of features, each candidate feature is re-scored conditioned on the current selected feature set. Our implementation uses the classifier prediction accuracy as the scoring criterion for the wrapper method, analogous to the last scoring criterion investigated in the filter approach experiments. The classifier accuracy most directly relates to the end goal of selecting a locally-optimal subset for grasp classification, and the filter method results suggest it should be more successful than the other choices of scoring functions.

The forward wrapper starts with an empty set of markers, and each step of the algorithm augments the current feature set by the marker whose inclusion results in the best LR classifier accuracy. The backward wrapper starts with the full set of 30 markers, and each iteration removes the least informative marker. Both wrapper algorithms were evaluated using two-fold cross validation on the first training data set. Figure 3 shows the cross-validation accuracy of the two wrapper methods for different sizes of marker sets.

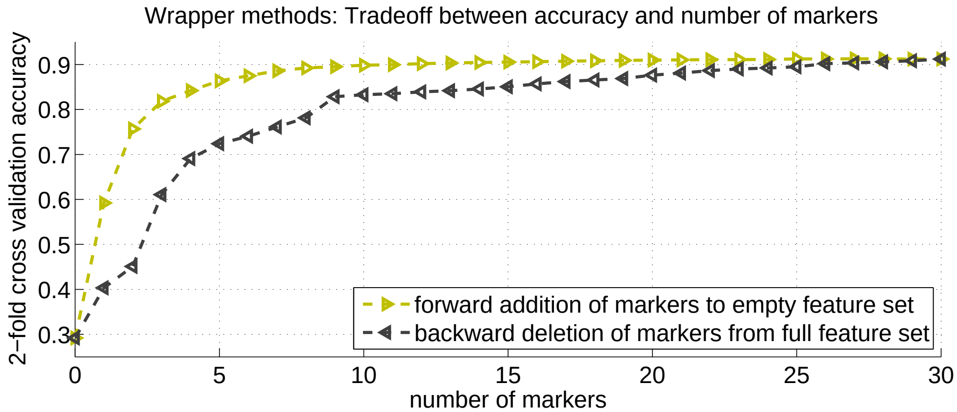


Figure 3: Wrapper methods: Feature selection based on classifier prediction accuracy scored on feature subsets.

The forward wrapper algorithm was the most successful in selecting a reduced set of features with small loss in the classifier accuracy. In other applications, the backward wrapper could be better because it may not remove features which have higher order interactions that would be missed by a greedy forward wrapper. However, for our application, the results suggest that distinguishing the worst feature is more difficult than selecting the best feature. The backward algorithm may have removed potentially strong features early on in an almost random fashion since the classification accuracies of different large feature sets usually differed by less than 1%. In contrast, the classification accuracies for the forward wrapper differed by up to 20% for different choices of small marker sets, so the best features were quickly added to build strong classifiers from a limited number of markers.

5.3 Final classifiers based on selected feature set

The final classifier is chosen from the marker ranking order found from the forward wrapper algorithm, which was the most successful of the feature selection methods tested. Two feature sets are chosen based on the size for which the forward wrapper cross-validation accuracy was at least 85% and 90%, resulting in a small set of 5 markers and a medium set of 12 markers, respectively. For these two feature sets and the full feature set, a multi-class logistic regression classifier is trained on a total example set consisting of the first data set and second data set of grasp examples from the same objects. These classifiers are then evaluated on the final held out test set with grasps of new objects. The results in Table 2 confirm that the marker protocol can be dramatically simplified to only 5 or 12 markers for a negligible loss of grasp classification accuracy.

Table 2: Classification performance of three feature sets chosen from forward wrapper results.

Number of markers	**small: 5**	**medium: 12**	**full: 30**
Accuracy on total training set	0.8801	0.9210	0.9381
Accuracy on test set with new objects	0.8543	0.8867	0.8993

5.4 Analysis and extensions

The test set prediction rates shown in Figure 4 show that spherical and handwriting grasps were the most difficult to classify for all three model sizes. The spherical grasp examples in the test set were all from the same, small object, and it is not surprising that these hand poses were mistaken for pinch, tripod, and handwriting precision grasps usually used for small objects. Similarly, the error of classifying the handwriting grasp as tripod grasp underscores the resemblance between the two grasps which both contact objects with the first three finger tips.

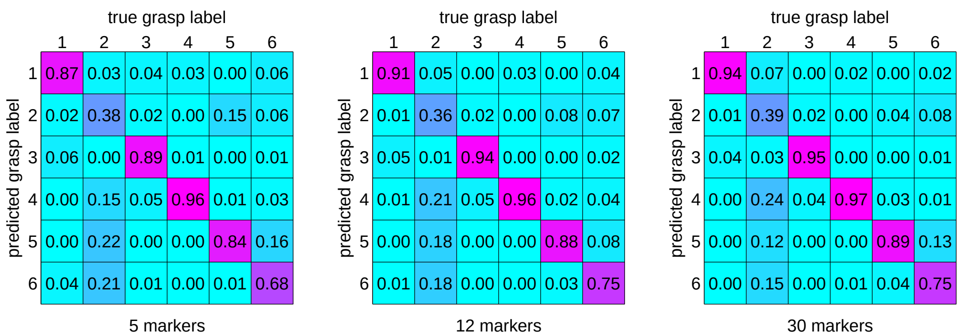


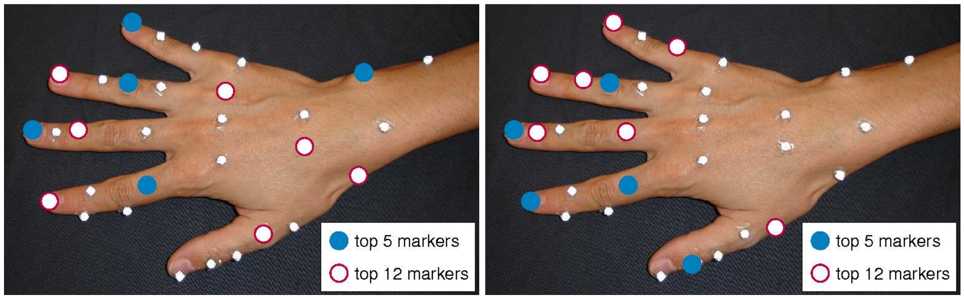
Figure 4: Test set predictions of different grasp classes for different size feature sets. The values in each column denote the percentage of examples predicted to be in each class, such that high values on the diagonal denote an accurate classifier. The six grasp types are: (1) cylindrical, (2) spherical, (3) lumbrical, (4) pinch, (5) tripod, and (6) handwriting.

The feature selection experiments not only suggest the size of the reduced marker set, but also where the markers should be placed on the hand. One interesting result of the tests is that the top 5 and top 12 markers selected include markers on the lower arm and on the back of the hand (Figure 5(a)), which intuitively should not be able to predict the shape of the hand. This is most likely due to an artifact of the available data, where the motion clips are from a single hand with exactly the same marker placement. The hand and arm markers may have been correlated with grasp due to systematic skin movement and systematic wrist angles due to the controlled position of objects.

To address the practical concern of grasp classification for different users and different object locations, the forward wrapper method was tested on a limited feature set with only the 21 finger markers. Using this modification, the top 5 finger markers (Figure 5(b)) suggest that a reduced marker set should include one marker on each of the first four digits and an additional marker on the index finger, which is reasonable given the index finger role in precision grips. Using the selected finger markers results in comparable classification accuracy for the same number of markers (Table 3) compared to the original method which allowed markers on the arm and back of the hand (Table 2).

Table 3: Classification performance on three feature sets chosen from finger markers only.

Number of markers	**small: 5**	**medium: 12**	**full finger set: 21**
Accuracy on total training set	0.8744	0.9064	0.9147
Accuracy on test set with new objects	0.8620	0.8882	0.9021



(a) full marker set (b) finger markers only

Figure 5: Selected markers from forward wrapper method on (a) full set of 30 markers and (b) only the 21 finger markers, excluding markers on the back of the hand or arm.

6 Discussion

In summary, supervised feature selection has been used to methodically design a reduced marker protocol for motion capture of the hand. Using as few as 5 markers as input features still maintains above 85% grasp classification accuracy, compared to about 90% accuracy when using a full set of 30 markers. In addition, the dramatic reduction of the number of features with small loss in accuracy was possible using a linear classifier which is computationally less expensive than the more complex models used in previous studies.

The main limitation of the study is the controlled conditions of the motion capture data used for training and evaluation. More exploration is needed to validate the performance of the reduced marker protocol in practical applications, in particular with respect to other subjects whose hand geometry and method of grasping will differ. Other extensions of the work might address the restrictions of the six grasps considered by developing methods to handle a larger grasp taxonomy, automatically learn the grasp classes, or reduce the marker set size using unsupervised feature selection.

References

[1] M. Santello, M. Flanders, and J.F. Soechting. Postural hand synergies for tool use. *Journal of Neuroscience*, (18):10105–15, 1998.

[2] C.R. Mason, J.E. Gomez, and T.J. Ebner. Hand synergies during reach-to-grasp. *Journal of Neurophysiology*, (86):2896–2910, 2001.

[3] M. Moussa and M. Kamel. A connectionist model of human grasps and its application to robot grasping. In *IEEE International Conf. on Neural Networks*, 2555–2559, 1995.

[4] M. R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *Robotics and Automation, IEEE Transactions on*, 5(3):269–279, 1989.

[5] S. Ekvall and D. Kragic. Grasp recognition for programming by demonstration. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 748–753, 2005.

[6] E. P. Xing. Feature selection in microarray analysis. In D.P. Berrar, W. Dubitzky, and M. Granzow, editors, *A Practical Approach to Microarray Data Analysis*. Kluwer Academic Publishers, 2003.

[7] S. J. Edwards, D. J. Buckland, and J. D. McCoy-Powlen. *Developmental & Functional Hand Grasps*. Slack Incorporated, Thorofare, New Jersey, 2002.

[8] C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.