Dimension Reduction of Anthropometric Measurements with Support Vector Machine for Regression: Application to a French Military Personnel Database

Pierre Puchaud [1,2], Simon Kirchhofer [3], Georges Dumont [2], Nicolas Bideau[4], Charles Pontonnier [1,2]

[1] Centre de Recherche des Ecoles de St-Cyr Coëtquidan, 56381 Guer, France
[2] Univ Rennes, CNRS, Inria, IRISA - UMR 6074, 35042 Rennes, France
[3] Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France
[4] Univ Rennes, Inria, M2S - EA 7470, F-35000 Rennes, France
pierre.puchaud@irisa.fr

Abstract. Collecting anthropometric data is a heavy and time-consuming procedure. The aim of this study was to find a reduced set of anthropometric measurements able to estimate the full-body dimensions of a given individual. The method was developed and applied on a database of 122 measurements carried out on 459 females and 771 males of the French military personnel. Among the 122, 26 key measurements were chosen. A regression method based on support vector machine was used to predict these key measurements in relation to each other. The designed "minimal measurement set selecting algorithm" chose 6 main inputs to predict the remained 20 measurements with mean correlation of 0.94 and 0.92, respectively on the training and on the testing data. This result tends to prove that the regression method can be used to predict the French military personnel anthropometrics.

Keywords: Anthropometrics · Military · Support Vector Machine · Machine Learning · Dimension Reduction · Scaling · Musculoskeletal Modelling

1 Introduction

From the early days of biomechanics to now on, anthropometric databases have been collected in armies (US, Australian and France) for product design. However, anthropometric protocols are heavy and time-consuming. They include numerous measurements which involves trained experimenters to be done. Thus, there is a need in simplification of such protocols to make them usable in a reduced time. Reducing the number of needed measurements and identifying the minimal set of measurements to correctly estimate unmeasured ones may be of great interest in such cases.

Thus, regression models have been designed to predict measurements as functions of other ones. Most anthropometric regression methods have been based on linear or multiple regression models to predict anthropometrics [1], [2], inertial parameters [3], [4] or muscle volumes [5] with a limited amount of data. These regressions have found an application in biomechanics to scale musculoskeletal (MSK) models. These MSK models combined with multi-body dynamic simulation provide useful insights and general guidelines to understand normal and pathological human movement.

With larger database, supervised machine learning can train more efficient algorithms to predict data with sophisticated algorithms such as support vector machine for regression (SVMR) [6]. These tools could make rapid anthropometric prediction with population-based statistical models [7]. These methods also have a great potential for dimension reduction in anthropometrics.

The aim of this study is to develop support vector machines for regressions to extract a minimal set of measurement able to estimate a complete set of measurements for MSK scaling. The method was applied to a French military personnel anthropometric database.

2 Materials and Methods

2.1 Anthropometric Database

Table 1. List of key measurements kept for musculoskeletal scaling.

N°	Name of the measure	Unit
1	Height of Left Anterior Superior iliac spine	mm
2	Height of Right Anterior Superior iliac spine	mm
3	Height of Left Great Trochanter	mm
4	Height of Right Great Trochanter	mm
5	Height of the middle Patella	mm
6	Distance from great trochanter to lateral femoral condyle	mm
7	Height of popliteal fossa	mm
8	Distance from lateral femoral condyle to lateral malleolus	mm
9	Bitrochanteric distance	mm
10	Maximal foot length	mm
11	Height vertex sat on a seat (sitting upright)	mm
12	C7 Height sat on a seat	mm
13	Acromial height sat on a seat	mm
14	Biacromial distance	mm
15	Pelvis width	mm
16	Anteroposterior thorax thickness	mm
17	Distance from elbow epicondyle to 3rd metacarpal head	mm
18	Distance from posterior plan to the tip of the hand	mm
19	Distance from acromion to epicondyle	mm
20	Distance from epicondyle to radial styloid	mm
21	Functional Upper Limb Reachable Height	mm
22	Height	mm
23	C7 Height	mm
24	Acromion Height	mm
25	Manubrium height	mm
26	Mass	kg

A scientifically sampled working data set from a French military personnel survey contains 1230 subjects, 459 females and 771 males respectively. In the database, the following information is stored for each subjects: demographics (eg. Age, sex, ethnicity) and 122 anthropometric measurements. These measurements were done on the torso, the upper limb, the lower limb, the head, the hand and the foot. Among the 122 measurements, 26 were selected as key measurements for their low ratio standard deviation over mean (<0.1) and their potential for MSK scaling. Indeed, these measurements are representative of the segments lengths and widths of the subject and may therefore be used for scaling. The subset of key measurements is listed Table 1.

2.2 Evaluation Criterion

In order to assess the performance of the developed prediction models, the Pearson's correlation coefficient r, the root mean square error (RMSE), the mean absolute error (MAE), the normalized RMSE (NRMSE) and the mean absolute percentage error (MAPE) were used as evaluation criteria. The criteria were applied to measure how close the real values are from the predicted values using the SVM models. They are given in equations (1), (2), (3), (4), (5) as:

$$r = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \ \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{y_{max} - y_{min}} \tag{4}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{5}$$

where y is the actual measurements, $\hat{y}$ is the estimated measurements using SVM and n is the total number of measurements. Criteria (2) and (3) indicated errors in the unit of the measurements and criteria (4) and (5) indicated relative errors in percentage to compare model performances between different magnitudes and units.

2.3 Minimal Measurement Set Selecting Algorithm using Support Vector Machine for Regression

Support vector machine is a supervised machine learning method used for classification and regression of data. The basic idea of SVMR is to find a function able to map input data into target data [6]. A linear epsilon-insensitive SVM (ε-SVM) regression has been used. Given a training dataset with inputs x and targets y, the ε-SVM regression try to find a function f(x) as the L1 loss $|y - f(x)| \leq \varepsilon$, while being as flat as possible.

Moreover, a gaussian function was chosen for the kernel of the SVM as we try to map statistic repartitions. In the training dataset, one or several measurements can be used as the input vector of the SVM to predict another measurement from the same dataset. The input must be chosen among the 122 measurement of the dataset. For example, we can use the Bitrochanteric distance and the C7 height to predict the Pelvis width.

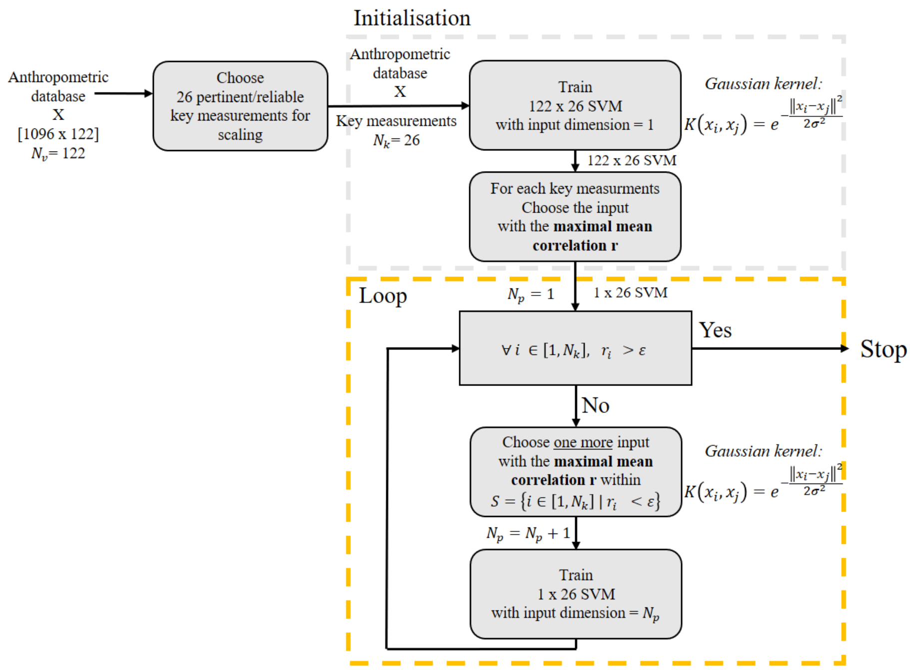

Fig. 1. Illustration of the minimal measurement set selecting algorithm.

The aim of the algorithm was to select a minimal set from the 122 measurement to predict the 26 measurements with a specified correlation using 26 SVMRs (Figure 1). The 122 measurements were all considered as potential input data for the 26 key measurements. Thus, 122 SVMRs with a Gaussian kernel were trained to predict each key measurement (122×26). The correlation r was used as the convergence criterion. The measurement, for which the SVMRs had the best correlations between real and predicted values in mean within the 26 key measurements, was selected as a common feature for 26 SVMRs.

While all correlation of the 26 SVMRs remained inferior to a threshold, a new common feature was added to train the 26 SVMRs. This new feature was selected for its high correlation within the remained key measurements below the threshold. The threshold was set to 0.8 for the study, considered as a statistically strong correlation level for such sample size of normally distributed data [8]. Finally, a 10-fold cross validation was used as a testing data set to estimate the performance of the model on new data.

3 Results

For this threshold set to 0.8, the algorithm stopped with a minimal correlation of 0.85 for the SVMR predicting anteroposterior thorax thickness (measurement No 16). Six measurements were included in the 26 SVMRs: manubrium height, mass, distance of the great trochanter to the lateral femoral condyle, bitrochanteric distance, height of acromion sat on a seat. These measurements were included in the key measurements set.

For the 20 remaining predicted measurements, mean correlation r for the training data and the testing data were 0.94 ± 0.05, and 0.92 ±0.06. Mean RMSE were respectively 12.47 ± 7.48 mm and 14.71 ± 8.24 mm. Mean MAE were respectively 9.39 ± 5.39 mm and 10.77 ± 6.74 mm. Mean NRMSE for the training data and the testing data were respectively 0.05 ± 0.03 and 0.06 ± 0.03. Mean MAPE for the training data and the testing data were respectively 0.02 ± 0.01 and 0.02 ± 0.01. For more details, all results are presented in Table 2.

We observed that the smaller measurements are less well predicted compared to larger measurements. For example, RMSE and MAE are same order of magnitude for C7 Height and popliteus height (No. 23 and 7) but the correlation r, the NRMSE and MAPE are lower.

4 Discussion

In this paper, we explored the use of SVM to develop models to predict anthropometric quantities from a reduced number of measurements. A 26-measurement subset of the database was chosen for their potential use for musculoskeletal scaling. The trained SVMRs successfully reduced the needed set of measurements from 26 to 6. These new minimal set of measurements with the SVMRs can predict anthropometrics for various applications. This method is interesting for markerless motion capture systems, especially for inertial measurement units [9] which mainly rely on anthropometrics for the scaling of kinematic chains. Also, the measurements provided by the SVM could be used as a new way to geometrically scale a musculoskeletal model through optimization techniques [10], [11], diminishing the number of actual measures to be done to scale the complete model.

Table 2. Evaluation criteria for each key measurement with the training and testing data set.

No	Data type	R	RMSE (mm)	MAE (mm)	NRMSE (%)	MAPE (%)
1	Training	0.96	18.07	14.02	0.05	0.01
	Testing	0.94	20.65	15.98	0.06	0.02
2	Training	0.96	17.89	13.84	0.05	0.01
	Testing	0.94	20.51	15.81	0.06	0.02
3	Training	0.95	16.97	12.97	0.05	0.01
	Testing	0.94	19.41	14.72	0.05	0.02
4	Training	0.95	17.14	13.09	0.05	0.01
	Testing	0.93	19.95	15.17	0.06	0.02
5	Training	0.92	13.00	10.06	0.06	0.02
	Testing	0.89	14.74	11.53	0.07	0.02
7	Training	0.92	11.75	8.98	0.06	0.02
	Testing	0.90	13.29	10.35	0.07	0.02
8	Training	0.91	11.49	8.81	0.07	0.02
	Testing	0.89	12.90	9.98	0.07	0.02
10	Training	0.89	8.39	6.44	0.09	0.03
	Testing	0.85	9.37	7.35	0.10	0.03
11	Training	0.94	14.55	11.21	0.06	0.01
	Testing	0.92	16.40	12.67	0.07	0.01
12	Training	0.90	14.53	10.17	0.06	0.02
	Testing	0.88	15.99	11.56	0.07	0.02
15	Training	0.91	12.40	9.28	0.06	0.03
	Testing	0.88	14.23	10.72	0.07	0.04
16	Training	0.85	13.12	10.10	0.09	0.04
	Testing	0.82	14.32	11.17	0.09	0.05
17	Training	0.88	11.38	8.63	0.08	0.03
	Testing	0.86	12.58	9.77	0.09	0.03
18	Training	0.90	25.35	19.77	0.08	0.02
	Testing	0.87	28.28	22.50	0.09	0.03
19	Training	0.87	11.64	9.09	0.08	0.03
	Testing	0.83	12.88	10.23	0.08	0.03
20	Training	0.86	9.67	7.27	0.09	0.03
	Testing	0.83	10.57	8.15	0.10	0.03
21	Training	0.95	36.50	28.06	0.05	0.01
	Testing	0.94	41.86	31.92	0.06	0.02
22	Training	0.99	13.05	10.02	0.02	0.01
	Testing	0.98	17.69	11.89	0.03	0.01
23	Training	0.99	13.47	10.55	0.03	0.01
	Testing	0.98	17.39	12.37	0.03	0.01
24	Training	0.99	13.59	10.20	0.03	0.01
	Testing	0.98	17.46	11.94	0.03	0.01
Mean	*Training*		*12.47*	*9.39*	*0.05*	*0.02*
(std)		*0.94 (0.05)*	*(7.48)*	*(5.39)*	*(0.03)*	*(0.01)*
	Testing	*0.92*	*14.71*	*10.77*	*0.06*	*0.02*
		(0.06)	*(8.24)*	*(6.74)*	*(0.03)*	*(0.01)*

The method could also be applied to the prediction of non-palpated landmarks such as hip joint centers [12]. Last, such an approach can be useful in product design such as exoskeleton design [13], enabling the development of anthropometrics based scaling methods to characterize populations of subjects instead of individuals.

As a perspective, this result should be extended to the time cost and the reliability of each measure. Indeed, the intra and inter experimenter variability is a measurement bias that can diminish the power of the method. From a methodological perspective, the support vector machine method could be compared to other approaches such as multi-linear regression or artificial neural networks. Finally, this study opens the perspective of defining an ideal minimal set of measurements which would be able to predict the parameters of a complete musculoskeletal model such as anthropometrics, body segment inertial parameters and muscles parameters (volumes, pennation angles, maximal isometric forces…).

References

1. R. Drillis, R. Contini, and M. Bluestein, “Body Segment Parameters, A Survey of Measurement Techniques,” Artif. Limbs, vol. 25, pp. 44–66, Nov. 1964.
2. W. T. Dempster, “Space Requirements of the Seated Operator,” Am. J. Phys. Anthro, vol. 22, no. March, pp. 1–254, 1955.
3. R. Dumas, L. Chèze, and J. P. Verriest, “Adjustments to McConville et al. and Young et al. body segment inertial parameters,” J. Biomech., vol. 40, no. 3, pp. 543–553, Jan. 2007.
4. Z. Merrill, S. Perera, and R. Cham, “Predictive regression modeling of body segment parameters using individual-based anthropometric measurements,” J. Biomech., vol. 96, Nov. 2019.
5. G. G. Handsfield, C. H. Meyer, J. M. Hart, M. F. Abel, and S. S. Blemker, “Relationships of 35 lower limb muscles to height and body mass quantified using MRI,” J. Biomech., vol. 47, no. 3, pp. 631–638, Feb. 2014.
6. A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” Stat. Comput., vol. 14, no. 3, pp. 199–222, 2004.
7. S. Yeung, J. W. Fernandez, G. G. Handsfield, C. Walker, T. F. Besier, and J. Zhang, “Rapid muscle volume prediction using anthropometric measurements and population-derived statistical models,” Biomech. Model. Mechanobiol., Oct. 2019.
8. S. Hulley, S. Cummings, W. Browner, D. Grady, and T. Newman, “Estimating Sample Size and Power: Applications and Examples,” in Designing Clinical Research, 3rd ed., Lippincott Williams & Wilkins, Ed. Philadelphia, PA, 2007, pp. 65–94.
9. D. Roetenberg, H. Luinge, and P. Slycke, “Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors,” 2009.
10. M. E. Lund, M. S. Andersen, M. de Zee, and J. Rasmussen, “Scaling of musculoskeletal models from static and dynamic trials,” Int. Biomech., vol. 2, no. 1, pp. 1–11, 2015.
11. P. Puchaud, C. Sauret, A. Muller, N. Bideau, G. Dumont, H. Pillet et al., Accuracy and kinematics consistency of marker-based scaling approaches on a lower limb model: a comparative study with imagery data, Comput. Methods Biomech. Biomed. Engin. (2019), pp. 1–12.
12. R. Hara, J. McGinley, C. Briggs, R. Baker, and M. Sangeux, “Predicting the location of the hip joint centres, impact of age group and sex,” Sci. Rep., vol. 6, 2016.
13. A. B. Zoss, H. Kazerooni, and A. Chu, “Biomechanical design of the Berkeley Lower Extremity Exoskeleton (BLEEX),” IEEE/ASME Trans. Mechatronics, vol. 11, no. 2, pp. 128–138, 2006.