ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Feature Selection for Footwear Shape Estimation

Fernando Mateo[1], Mónica Millán-Giraldo[1], Juan J. Carrasco[1], Enrique Montiel[2], José A. Bernabeu[2] and José D. Martín-Guerrero[1] *

1- University of Valencia - Intelligent Data Analysis Laboratory
Avda Universitat S/N, 46100 Burjassot - Valencia, Spain

2- Instituto Tecnológico del Calzado y Conexas (INESCOP)
Polígono Industrial Campo Alto, 03600 - Elda, Spain

Abstract. This study proposes feature selection techniques to obtain a set of significant foot anthropometric measurements that can assist custumers in the choice of footwear size and width. The results given by a number of methods are averaged to provide a reliable set of features. Several machine learning methods are used to evaluate the classification (for the width) and regression (for the size) accuracies before and after feature selection. The results prove the benefits of carrying out feature selection, especially for the shoe width.

1 Introduction

In shoe industry, comfort is a crucial issue for the consumer. The comfort feel of a shoe is subjective for each person and depends on many different factors, like the design, fit and function of the shoe, the shock absorption qualities (padding), materials, weight, isolation and also the particular foot shape, sensitivity and kinetic and dynamic characteristics of the consumer (e.g. plantar pressure [1]).

There are many possible anatomic measurements that can be considered for an optimal choice of size and width. To make measurements collection feasible, it is necessary to minimize their amount. This paper is based on the application of a variety of feature selection techniques to extract the most correlated anthropometric measurements with the desired outputs related to the shoe shape, namely, size and width of the shoe. Different feature selection methods are considered for classification (width) and for regression (size). After the feature selection stage, several linear and non-linear machine learning techniques will be used to evaluate the classification and regression accuracy on the dataset before and after performing feature selection.

2 Data and methods

The data used for the experiments was provided by the *Instituto Tecnológico del Calzado y Conexas* (INESCOP). The dataset contains 43 variables and 621

*This work was financed by the Spanish Ministerio de Economía y Competitividad (MINECO) and FEDER funds in the framework of the project EFIS: Un Sistema Inteligente Adaptativo para la Gestión Eficiente de la Energía en Grandes Edificios (IPT-2011-0962-920000) and project Tecnologías Virtuales en Punto de Venta (TSI-020302-2011-37), funded by Ministerio de Industria, Energía y Turismo (MINETUR) and FEDER funds. The data for the experiments were provided by Instituto Tecnológico del Calzado y Conexas (INESCOP).

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

samples corresponding to foot anthropometric measurements of 311 individuals. The variables to predict will be the shoe size and shoe width. The width was coded with three labels (Narrow, Medium and Wide) and was addressed as a classification problem. On the other hand, as the size contained 11 possible numeric labels in steps of half a size, it was treated as a continuous function (i.e. a regression problem) and its estimated value was discretized *a posteriori*. In every case, the input data were standardized to zero mean and unit variance. A first correlation analysis was carried out to remove variables which were not correlated with the outputs or were linear combinations of other variables. The resulting dataset contains the 20 input variables listed in Table 1. These will be called the *pre-selected* variables.

Var.	Description	Var.	Description
8	Pro jected width	28	Standard toe medial width
10	Toe length	29	Standard toe lateral width
14	Heel width	34	Distal first phalanx girth
15	Heel medial width	35	Distal first phalanx height
16	Heel lateral width	37	Distal first phalanx medial width
17	High instep girth	38	Dista first phalanx lateral width
18	Medium instep girth	39	Arch height index
20	Mid metatarsal point (CL) distance from the back-most point of the heel	40	Arch displacement index
22	Standard Ball height	41	Egyptian 0-No, 1-Yes
26	Standard toe girth	43	With bunion 0-No, 1-Yes

Table 1: Used variables.

2.1 Feature selection methods

A variety of feature selection methods have been proposed to obtain a reliable subset of features. Some methods will be used for regression, some for classification and some for both purposes. These methods are:

- Sequential Feature Selection (SeqFS): It chooses a subset of features by sequentially selecting them until there is no improvement in the prediction [2]. The model used to evaluate the performance at each iteration was a neural network with a hidden layer composed of 10 nodes.
- Stepwise fit (SWfit): It uses a stepwise method to perform a multilinear regression of the response values between the output data and the input data. Predictive terms are added sequentially to the model according to their importance in terms of p value [3].
- Genetic algorithm with delta test (GA-DT): Genetic algorithm that uses the delta test as fitness function [4]. Delta test is an estimation of the variance of the noise at the output of an unknown continuous function, based on a nearest neighbor formulation [5].
- ReliefF: A general attribute estimator for classification that is able to effectively provide quality estimates of attributes in problems with dependencies between them [6]. The regression version is called RReliefF.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

- Least Angle Regression (LARS): It adds the variables sequentially to the model according to their correlation with the residual of the response [7].
- Least Squares Feature Selection (LSFS): Solves a least squares system and orders input features according to their dependency on the outputs [8].
- Information Gain (IG): Evaluates the worth of an attribute by measuring the information gain with respect to the class [9].
- Minimum Redundancy Maximum Relevance (mRMR): It selects variables that have the highest relevance with the target class and are also maximally dissimilar to each other [10].
- Sparse Multinomial Logistic Regression Method (SBMLR): Logistic regression is used to build a generalized regression model that can handle multiclass data. A constraint is used to shrink the model [11].
- Plus-L-takeaway-R feature selection (LR): Each iteration is divided in two steps. First, sequential forward search is used to include L new variables. Second, sequential backward search removes R variables [12].

3 Experimental setup and results

The feature selection methods used in this work were implemented in Matlab (version R2012a). Some methods were adapted from PRTools Toolbox and Weka. The data were randomly split into training (2/3) and test (1/3) subsets. The feature selection techniques were evaluated on the training data and the test data was left aside for testing the models. The feature selection methods were applied to 20 different dataset splittings in order to obtain reliable results. In each repetition, each variable was given a linear score between 0 and 1 according to its position in the ranking. The scores were averaged for all repetitions and, afterwards, they were averaged for all feature selection methods. The results, both for size and width, are listed in Tables 2 and 3, respectively.

The next step is to prune out the unnecessary variables. To achieve this, we have kept the minimum number of variables that explain 80% of the variance of the outputs. These variables are highlighted in bold in Tables 2 and 3.

3.1 Results

Once the number of variables has been reduced, linear and non-linear predictive models can now be applied to obtain the optimal size and width for a particular customer. k-Nearest Neighbor (KNN) [13] (with $k = 1$) and Least Squares-Support Vector Machine (LS-SVM) [14] (with radial basis function kernels) models are used both for regression and classification. In addition, Robust Multiple Linear Regression (RMLR) [7] and Linear Discriminant Analysis (LDA) [13] are employed for regression and classification, respectively.

Fig. 1 shows the classification accuracy obtained by the KNN, LDA and LS-SVM models for the width prediction, when data correspond to the pre-selected

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Var.	SeqFS	SWfit	GA-DT	RReliefF	LARS	LSFS	Mean	% O.V.*	Rank
8	0.37	0.61	0.70	0.90	0.91	0.89	0.73	**9.70**	3
10	0.62	0.48	0.92	0.80	0.60	0.80	0.70	**9.34**	4
14	0.08	0.66	0.32	0.56	0.80	0.74	0.53	**7.00**	6
15	0.09	0.28	0.22	0.32	0.30	0.37	0.26	**3.50**	11
16	0.00	0.08	0.09	0.26	0.00	0.63	0.17	2.32	16
17	0.32	0.04	0.62	0.94	0.68	0.95	0.59	**7.86**	5
18	0.69	0.81	0.61	0.86	0.92	0.86	0.79	**10.52**	2
20	1.00	1.00	0.81	1.00	1.00	1.00	0.97	**12.88**	1
22	0.12	0.15	0.45	0.46	0.78	0.66	0.44	**5.80**	7
26	0.08	0.00	0.21	0.50	0.00	0.65	0.24	3.20	12
28	0.13	0.00	0.04	0.50	0.10	0.56	0.22	2.95	13
29	0.08	0.11	0.04	0.30	0.00	0.39	0.15	2.02	17
34	0.08	0.00	0.11	0.16	0.00	0.29	0.11	1.41	20
35	0.16	0.22	0.39	0.55	0.35	0.51	0.36	**4.80**	8
37	0.08	0.07	0.14	0.36	0.25	0.44	0.22	2.95	14
38	0.12	0.15	0.10	0.50	0.15	0.16	0.19	2.58	15
39	0.08	0.04	0.15	0.35	0.04	0.06	0.12	1.59	18
40	0.04	0.74	0.07	0.47	0.42	0.19	0.32	**4.27**	9
41	0.04	0.00	0.18	0.31	0.02	0.11	0.11	1.46	19
43	0.05	0.27	0.32	0.42	0.43	0.26	0.29	**3.86**	10

* % of the output variance

Table 2: Averaged scores and ranking of pre-selected features for size (regression). The selected features (accumulating 80% of the output variance) are highlighted in bold.

Var.	IG	mRMR	ReliefF	SBMLR	LR	GA-DT	LSFS	Mean	% O.V.*	Rank
8	0.99	0.82	0.95	0.96	0.75	0.95	0.99	0.92	**9.82**	1
10	0.56	0.07	0.06	0.27	0.56	0.62	0.17	0.33	3.53	16
14	0.58	0.06	0.44	0.00	0.37	0.36	0.61	0.35	3.70	15
15	0.50	0.80	0.68	0.68	0.34	0.50	0.59	0.58	**6.25**	6
16	0.45	0.00	0.41	0.19	0.31	0.57	0.36	0.32	3.47	17
17	0.81	0.60	0.47	0.49	0.75	0.63	0.86	0.66	**7.06**	4
18	0.90	0.38	0.83	0.00	0.79	0.68	0.93	0.64	**6.89**	5
20	0.41	0.00	0.62	0.42	0.56	0.48	0.34	0.41	**4.34**	12
22	0.91	0.94	0.80	0.99	0.65	0.66	0.89	0.83	**8.94**	2
26	0.86	0.54	0.89	0.51	0.60	0.39	0.84	0.66	**7.07**	3
28	0.73	0.09	0.65	0.10	0.33	0.07	0.72	0.38	**4.12**	13
29	0.41	0.59	0.56	0.08	0.40	0.33	0.53	0.41	**4.43**	11
34	0.32	0.00	0.28	0.14	0.15	0.00	0.36	0.18	1.92	19
35	0.70	0.79	0.55	0.53	0.33	0.33	0.68	0.56	**5.98**	7
37	0.63	0.63	0.47	0.44	0.25	0.31	0.59	0.47	**5.07**	8
38	0.25	0.30	0.24	0.84	0.29	0.30	0.31	0.36	3.87	14
39	0.20	0.00	0.13	0.08	0.33	0.20	0.09	0.15	1.57	20
40	0.15	0.21	0.18	0.18	0.31	0.22	0.29	0.22	2.36	18
41	0.10	0.17	0.97	0.52	0.74	0.42	0.13	0.44	**4.67**	10
43	0.05	0.78	0.35	0.67	0.76	0.37	0.25	0.46	**4.92**	9

* % of the output variance

Table 3: Averaged scores and ranking of pre-selected features for width (classification). The selected features (accumulating 80% of the output variance) are highlighted in bold.

and selected sets. As it can be observed, the accuracy rate is improved when the most relevant variables are used to build the classification models. When classifiers are trained with the pre-selected set, KNN outperforms the remaining classifiers. However, when the selected set is used for their training, LS-SVM improves the rest of models. It is for this classifier that the feature selection effect

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

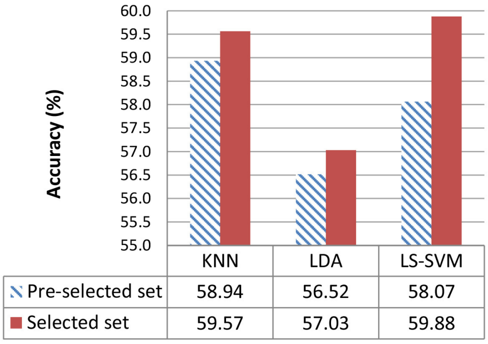

Fig. 1: Overall accuracy obtained by the three classification models: KNN, LDA and LS-SVM in the width prediction.

	KNN			LDA			LS-SVM		
Class	1	2	3	1	2	3	1	2	3
PS	**44.78**	**49.53**	67.63	**11.88**	**36.51**	78.87	18.67	30.46	**83.43**
S	42.84	49.05	**69.88**	9.35	35.40	**81.01**	**21.94**	**40.86**	79.67

Table 4: Accuracy obtained by the three classification models: KNN, LDA and LS-SVM, for each class with the pre-selected (PS) and selected (S) sets.

	RMLR		KNN		LS-SVM	
	Original	Flexible	Original	Flexible	Original	Flexible
PS	30.39	58.26	**41.11**	64.66	**31.88**	60.12
S	**30.72**	**59.28**	39.93	**67.20**	31.55	**61.57**

Table 5: Accuracy obtained by the three regression models: RMLR, KNN and LS-SVM in the size prediction, for pre-selected (PS) and selected (S) sets.

becomes more noticeable. Table 4 shows the percentage of correctly classified instances for each class (Narrow (1), Medium (2) and Wide (3)). The best accuracy values are highlighted in bold. For KNN and LDA models, Table 4 shows that the accuracy decreases for classes 1 and 2, when the number of variables is reduced, whereas it increases for class 3. However, the opposite behavior is observed for the LS-SVM, where accuracy increases for classes 1 and 2, and decreases for class 3.

Regarding the size, Table 5 shows the accuracy rate estimated without (*Original*) and with a tolerance of half a size up and a half a size down (*Flexible*). As it can be observed, an improvement in the *Flexible* classification accuracy is obtained by doing feature selection for the three models. However, in relation with the *Original* accuracy this behavior is only observed for the RMLR model. Furthermore, *Flexible* accuracy rates are higher than the *Original* ones as expected. The KNN classification model shows the best performance in both *Original* and *Flexible* accuracy rates, whereas RMLR appears to be the worst model in both accuracy measurements for the size prediction.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

4 Conclusions

This paper has analyzed how feature selection of anthropometric measurements can affect the prediction of footwear size (regression problem) and width (classification problem). For this purpose, ten different feature selection methods have been used in order to identify the most representative variables in both cases. The accuracy rate has been measured by different linear and non-linear models, including KNN, LDA, LS-SVM and RMLR. The results have suggested that feature selection may improve the prediction accuracy, especially in the width prediction case. In the case of size prediction, this improvement happens when half a size tolerance has been considered. Without this tolerance, only the prediction by RMLR improves the results.

References

[1] C. Jordan and R. Bartlett. Pressure distribution and perceived comfort in casual footwear. *Gait and Posture*, 3(4):215–220, 1995.

[2] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[3] N. R. Draper and H. Smith. *Applied Regression Analysis*, chapter pp. 307–312. Wiley-Interscience, Hoboken, NJ, 1998.

[4] F. Mateo, D. Sovilj, and R. Gadea. Approximate k-NN delta test minimization method using genetic algorithms: Application to time series. *Neurocomputing*, 73(10-12):2017–2029, June 2010.

[5] E. Eirola, E. Liitiäinen, A. Lendasse, F. Corona, and M. Verleysen. Using the delta test for variable selection. In *Proceedings of the European Symposium on Artificial Neural Networks, ESANN 2008*, pages 25–30, 2008.

[6] I. Kononenko. Estimating attributes: Analysis and extensions of relief. In *Proceedings of the European Conference on Machine Learning, ECML 2004*, pages 171–182. Springer Verlag, 1994.

[7] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, Second Edition: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd edition, 2009.

[8] T. Kanamori, S. Hido, and M. Sugiyama. A least-squares approach to direct importance estimation. *J. Mach. Learn. Res.*, 10:1391–1445, December 2009.

[9] I. Guyon and A. Elisseeff. *Feature Extraction Foundations and Applications*. Springer, 2006.

[10] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226 –1238, aug. 2005.

[11] G. C. Cawley, N. L. C. Talbot, and M. Girolami. Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation. *Advances in Neural Information Processing Systems*, 19, 2007.

[12] S. D. Stearns. On selecting features for pattern classifiers. In *Proceedings of the 3rd International Conference on Pattern Recognition (ICPR 1976)*, pages 71–75, Coronado, CA, 1976.

[13] K. Fukunaga. *Introduction to Statistical Pattern Recognition, Second Edition (Computer Science & Scientific Computing)*. Academic Press, 2 edition, 1990.

[14] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.