



<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper presented at *13th International FLINS Conference, Belfast, August 21-24, 2018.*

Citation for the original published paper:

Löfström, T., Johansson, U., Balkow, J., Sundell, H. (2018)

A data-driven approach to online fitting services

In: Jun Liu (Ulster University, UK), Jie Lu (University of Technology Sydney, Australia), Yang Xu (Southwest Jiaotong University, China), Luis Martinez (University of Jaén, Spain) and Etienne E Kerre (University of Ghent, Belgium) (ed.), *Data Science and Knowledge Engineering for Sensing Decision Support* (pp. 1559-1566).

<https://doi.org/10.1142/11069>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-15824>

A DATA-DRIVEN APPROACH TO ONLINE FITTING SERVICES

Tuwe Löfström*, Ulf Johansson*, Jenny Balkow⁺ and Håkan Sundell*

**Dept. of Computer Science and Informatics, Jönköping University and
Dept. of Information Technology, University of Borås, Sweden
E-mail: {tuwe.lofstrom, ulf.johansson, hakan.sundell}@ju.se*

*⁺Swedish School of Textiles, University of Borås, Sweden
E-mail: jenny.balkow@hb.se*

Being able to accurately predict several attributes related to size is vital for services supporting online fitting. In this paper, we investigate a data-driven approach, while comparing two different supervised modeling techniques for predictive regression; standard multiple linear regression and neural networks. Using a fairly large, publicly available, data set of high quality, the main results are somewhat discouraging. Specifically, it is questionable whether key attributes like sleeve length, neck size, waist and chest can be modeled accurately enough using easily accessible input variables as sex, weight and height. This is despite the fact that several services online offer exactly this functionality. For this specific task, the results show that standard linear regression was as accurate as the potentially more powerful neural networks. Most importantly, comparing the predictions to reasonable levels for acceptable errors, it was found that an overwhelming majority of all instances had at least one attribute with an unacceptably high prediction error. In fact, if requiring that all variables are predicted with an acceptable accuracy, less than 5% of all instances met that criterion. Specifically, for females, the success rate was as low as 1.8%.

Keywords: Predictive regression; Online fitting; Fashion

1. Introduction

Today, worldwide returns total more than US\$ 0.5 trillion annually^a. Furthermore, returns are especially troublesome for clothing retailers, where the return rate is over 10%, which is the highest among retail segments.

^awww.marketwatch.com/story/consumers-return-6426-billion-in-goods-each-year-2015-06-18

For clothing retailers, this rate of returns is extremely problematic since only 48% of all returns can be resold at full value. There is, however, an even larger implication of the ever-growing number of returns. i.e., the impact the resulting unnecessary transports have on the environment. So, reduced returns would benefit not only e-tailers and their customers, but also the public in general. One specific problem, leading to many returns, is the fact that a customer purchasing clothes online must somehow choose a specific size, without being able to try the item of clothing on. In fact, this has led to a behavior where a customer orders two or more identical items, of different sizes, and returns all but one.

With this in mind, the quality of the services provided for online fitting by the e-tailers in fashion, becomes very important. In this paper, we look at one specific type of fitting service, where the user inputs a few, but very important, measurements as sex, weight and height, before an algorithm predicts a number of other variables like sleeve length, neck size, waist and chest. One example of such a service is the plugin provided by the company SizeMe^b, as used by for instance TailorStore^c. Since the algorithm used by SizeMe is proprietary, we do not know exactly how the input data is used in order to produce the predicted sizes. Because of this, it is very important to recognize that this study in no way evaluates the specific service provided by SizeMe. Instead, we address the problem in a more general way; how can we use a data-driven approach to estimate some output variables (neck size, chest etc.) based on just a few input variables like height and weight? For this purpose we use a fairly large data set called ANSUR2, which is a public data set drawn from the 2012 U.S. Army Anthropometric Survey.¹ In the experiments, we compare using the standard technique multiple linear regression to neural networks, and we investigate the importance of adding a few readily available input measurements to the three basic variables sex, weight and height. Finally, we make a simulation where we look at the quality of the predictions, and determine whether these would be good enough if used as a guide when purchasing a shirt.

We start by providing a background for our study in Section 2, where we also give an overview of predictive modeling. In Section 3, we present the method employed, including the experiments. The main results are presented in Section 4, before we conclude and suggest some future work in Section 5.

^bwww.sizeme.com

^cwww.tailorstore.com/mens-body-measurements

2. Background

Garments that fit remain one of the final frontiers for a seamless online shopping experience. Companies and researchers alike struggle to find the optimum solution to help customers, which has resulted in a surge of on-line service providers that offer customer guidance.² The service provided is usually based on one of two types of logic; *self measurements* or *body scanning*. Both solutions have their inherent problems. The body scanning solutions today are based on a 3D-scanned image which becomes very accurate in measurement, but this technology is not available to all.³ Service providers using ordinary computer cameras, such as UPcloud, are emerging, but is based on 2D measures of the body and requires standing in front of a camera.² The other technique, which is more frequently used is based on manual body measurements. The inherent problem with this solution is twofold; first, it takes time to measure and enter data into the system which discourages the consumer and, second, previous studies show that manual measures are not very accurate.⁴ The main exceptions are two measures, height and weight, that most consumers have accurate knowledge about without checking again. Thus it is very convenient to make predictions of size based on these measures only.

2.1. Predictive modeling

Most predictive techniques consist of a two-step process: first an *inductive step*, where a model is constructed from labeled data, and then a second, *deductive*, step where the model is applied to test instances. If the target variable is discrete, and restricted to a set of pre-defined labels, the task is called *classification*; if the target is numerical, it is instead called *regression*.

Predictive regression, consequently, maps each instance \mathbf{x} to a continuous target attribute T . More technically, an algorithm uses a set of labeled training instances, each consisting of an input vector x_i and a corresponding target value t_i to learn the function $T = f(\mathbf{x}; \theta)$. During training, the parameter values θ are optimized, based on a score function. When sufficiently trained, the regression model is able to predict a value t_j , when presented with a novel (test) instance \mathbf{x}_j .

In this study, we compare traditional multiple linear regression to using an artificial neural network for the modeling. With multiple linear regression, the model describes the dependent variable as a linear combination of the independent variables. Artificial neural networks (ANNs) have proved to be successful in numerous data mining applications. Specifically, *mul-*

tilayer perceptrons (MLPs), i.e., feed-forward networks with one or more hidden layers using non-linear activation functions, are general function approximators; in fact MLPs with just a single hidden layer are capable of representing any continuous function.

3. Method

3.1. *The data set*

The data is divided into two sets, one for males containing 4082 instances and one for females containing 1986 instances, which include the same set of variables. The data contains about 100 body measures and some additional variables representing sex, age, position etc. Most of the measures were deemed not relevant for this study and consequently excluded. The eight variables used in the study are presented in Table 1, with descriptions taken from the documentation accompanying the data.¹ The full names used in the source are presented, with parts in brackets excluded in our text.

Table 1: Description of the data used

Variable Name	Purpose	Description
Neck [Circumference Base]	Target	The circumference of the base of the neck encompassing the lateral neck and anterior neck landmarks
Sleeve [Outseam]	Target	The distance between the acromion landmark on the tip of the right shoulder and the stylium landmark
Chest [Circumference]	Target	The maximum circumference of the chest at the fullest part of the breast
Waist [Circumference]	Target	The horizontal circumference of the waist at the level of omphalion encompassing the waist (omphalion) landmarks
Stature	Input	The vertical distance from a standing surface to the top of the head
Weight	Input	Participant stands on the platform of the scale with weight distributed evenly on both legs
Span	Input	The distance between the tips of the middle fingers of horizontally outstretched arms
Age	Input	Age of the participant.

3.2. Experiments

In the first experiment, we compare using different modeling techniques (ANNs and multiple linear regression) as well as different sets of input variables. More specifically, we try the following four sets of input variables:

- **Basic:** Just the weight and the length. This represents the baseline, and is a typical choice for most online fitting services.
- **Age:** Weight, length and age. The reason for including age is that it is a variable that all users know the value for, i.e., no measurement is necessary.
- **Span:** Weight, length and span. Obviously several different variables could be used instead, but we settled for using span since it is fairly easy to measure with high precision.
- **All:** Weight, length, age and span.

Obviously we look at men and women separately. For the modeling, the tool KNIME⁵ was used. After some experimentation with different parameter settings, the default settings were used for both ANNs and multiple linear regression. Min-max normalization in the interval (0,1) was used with the ANNs. The actual evaluation uses 5-fold cross-validation, i.e., all reported results are the average results from the five folds.

In the second experiment, we simulate using the predictions when ordering a shirt. This is done by, for each person in the data set, comparing the predicted value of the four key variables (Neck, Sleeve, Chest and Waist) to the true values. In order to determine whether the predictions are accurate enough, we set a threshold for every variable. It must be noted that there is no universally accepted method for setting this threshold, and the exact definition will of course impact the results significantly. For this preliminary study, we decided to base the threshold on the typical differences between two sizes, i.e., a prediction is correct if the absolute error is smaller than half the range of one size. Using this definition, we get the thresholds in Table 2 below.

Table 2: Thresholds for the variables given as maximum absolute error

	Men	Women
Neck	10 mm	10 mm
Sleeve	5 mm	5 mm
Chest	35 mm	20 mm
Waist	30 mm	20 mm

4. Results

Starting with the results for the multiple linear regression, Table 3 below shows the accuracies for the different input variable sets.

Table 3: Accuracy using multiple linear regression

	MAE				R^2			
	Basic	Age	Span	All	Basic	Age	Span	All
Men								
Neck	12.9	12.8	12.9	12.8	0.59	0.60	0.60	0.60
Sleeve	14.5	14.5	10.8	10.8	0.64	0.64	0.81	0.81
Chest	25.5	24.4	25.5	24.4	0.86	0.87	0.86	0.87
Waist	35.1	32.2	34.8	31.8	0.84	0.87	0.85	0.87
Women								
Neck	10.6	10.6	10.6	10.6	0.53	0.53	0.53	0.53
Sleeve	14.2	14.2	9.6	9.6	0.63	0.63	0.83	0.83
Chest	33.9	32.7	33.7	32.6	0.73	0.75	0.74	0.75
Waist	38.4	37.4	38.3	37.4	0.76	0.77	0.76	0.77

Looking at the mean absolute errors, which of course are given in millimeters, and the R^2 -values, we see that the variable age adds some information regarding Waist and Chest; both for men and women. Span, as expected, is extremely valuable for Sleeve. The best results overall are achieved when all variables are used, but to be honest the differences compared to using only weight and length must be considered marginal, with the exception of Sleeve. Comparing the results for women and men, it may be observed that the R^2 -values are generally substantially higher for men; with the exception of Sleeve. Turning to the accuracy for the ANNs, we see in Table 4 below the picture is very similar.

Table 4: Accuracy using artificial neural networks

	MAE				R^2			
	Basic	Age	Span	All	Basic	Age	Span	All
Men								
Neck	12.9	12.8	12.9	12.8	0.60	0.60	0.60	0.60
Sleeve	14.6	14.7	11.1	11.2	0.64	0.63	0.80	0.79
Chest	25.3	24.2	25.6	24.7	0.86	0.88	0.86	0.87
Waist	35.1	32.4	34.7	32.0	0.84	0.86	0.85	0.87
Women								
Neck	10.6	10.6	10.6	10.6	0.53	0.53	0.53	0.53
Sleeve	14.3	14.3	9.9	10.1	0.63	0.63	0.82	0.81
Chest	34.0	32.8	33.9	32.7	0.73	0.75	0.74	0.75
Waist	38.6	37.6	38.8	37.6	0.76	0.77	0.76	0.77

Somewhat discouraging, when comparing the in Table 3 and 4 above, we see that there is nothing to gain from using the more advanced ANN models. If anything, the multiple linear regression models are actually marginally more accurate than the ANNs.

Turning to Experiment 2, Table 5 below shows the proportion of all instances that are deemed to be sufficiently accurate, i.e., the absolute error is smaller than the maximum absolute error, as defined in Table 2. The last column shows the proportion of all instances where all individual predictions were inside the acceptable intervals.

Table 5: Proportion of predictions that are correct, i.e., with an absolute error smaller than the threshold

	Neck	Sleeve	Chest	Waist	Total
Men	0.476	0.227	0.730	0.503	0.045
Women	0.560	0.216	0.364	0.320	0.018

Looking first at the results for men, we see that the hardest attribute to predict with an acceptable accuracy is Sleeve. Less than one quarter of all predictions are within the acceptable interval. Waist and Neck are both also quite hard; approximately 50% of all predictions obtained a too high mean absolute error. All-in-all, less than 5% of all instances had acceptable predictions for all four output variables.

Actually, the picture for females is even worse. Here, for both Waist and (in particular) Chest, the proportions of poor predictions are much higher than for men. One key reason is of course that the acceptable intervals are smaller; but it must be noted that even a direct comparison with regard to mean absolute errors, shows that these attributes are in fact harder to predict for women. In total, only 1.8% of all females in the data set, had all four predictions inside the acceptable intervals.

While we do not claim that all instances with at least one prediction having an unacceptably high error would necessarily lead to a return, we still find the results remarkable. At the very least, it is questionable if systems based on techniques like the ones tried here really can be useful. In fact, showing predictions with a precision of 1 cm, given the magnitude of the prediction errors we observe here, might even be considered misleading.

5. Concluding remarks

We have in this paper investigated the possibility to predict some key attributes related to fitting of shirts, based on some very basic variables like weight and height. From the experiments, we see that while the models often are fairly accurate, as seen by low mean absolute errors and high R^2 values, it remains questionable, whether the precision is of acceptable quality. For this specific data set, standard linear regression was as accurate as the more complex and potentially more powerful neural networks. Generally, the effect of adding age to the input variables was marginal, while the measure span, as expected, could lower the errors for Sleeve. Comparing the predictions to pre-defined levels for acceptable errors, it was found that less than 5% of all instances were predicted with the necessary accuracy. In fact, for females the success rate was as low as 1.8%.

Regarding future work, there are many different possibilities to pursue. It would be interesting to look for other measures that are easy to take with good precision, but with a higher predictive power. From a more technical perspective, it must be noted that both techniques used here, i.e., multiple linear regression and neural networks, are *eager* learners, meaning that they use all training instances to produce a model that is later used for the predictions. A different approach would be to use a *lazy* learner, typically some variant of k-nearest neighbors. Using k-nearest neighbors, the prediction is more specialized since it is based only on the most similar examples, i.e., here people with very similar weight, height, age etc.

Acknowledgement

Funded by The Knowledge Foundation, grant nr. 20160035.

References

1. C. C. Gordon and et al., *2012 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics - Final Report*, tech. rep., Army NATICK Soldier Research Development and Engineering Center MA (2012).
2. A. Vecchi, F. Peng and M. Al-Sayegh, Size recommendations in online fashion retail: Opportunities and challenges, in *Advanced Fashion Technology and Operations Management*, (IGI Global, 2017) pp. 248–362.
3. Y.-A. Lee, M. L. Damhorst, M.-S. Lee, J. M. Kozar and P. Martin, *Clothing and Textiles Research Journal* **30**, 102 (2012).
4. P. E. Zwane, M. Sithole and L. Hunter, *International Journal of Consumer Studies* **34**, 265 (2010).
5. M. R. Berthold and et al., KNIME: The Konstanz Information Miner, in *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*, (Springer, 2007).