Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Forensic Science International 148 (2005) 163–167
Forensic Science International
www.elsevier.com/locate/forsciint

Aging process variability on the human skeleton: artificial network as an appropriate tool for age at death assessment

Marc-Michel Corsini[a,*], Aurore Schmitt[b], Jaroslav Bruzek[b]

[a]*Laboratoire des Sciences Cognitives, EA 487, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33 076 Bordeaux Cedex, France*
[b]*PACEA, Laboratoire d'Anthropologie des Populations du Passé, UMR 5199 Université Bordeaux 1 Avenue des Facultés, 33 405 Talence Cedex, France*

Received 5 May 2003; received in revised form 26 May 2004; accepted 29 May 2004
Available online 17 July 2004

Abstract

Adult age-at-death assessment is one of the most difficult problem encountered in paleoanthropology. Many procedures have been proposed using either skeletal remains or dental records, but most show systematic bias. Data processing of current methods are a source of error because they neglect that process of biological ageing is very variable between individuals and populations. The aim of this study is to test the potentiality of artificial neural networks (ANN) as a prediction tool. ANN have been used for a wide variety of applications where statistical methods are traditionally employed. But it performs better to solve linearly non separable patterns. We applied this technique after observation of several features' aging changes of the pubic symphysis and the auricular surface of the ilium.

Although we failed to reduce the size of the intermediate class (30–59 years), the neural network identifies, with better reliability than previous works, the youngest (20–29 years) and the oldest (above 60 years) individuals.

Keywords: Age-at-death; Artificial neural network; Adult; Pubic symphysis; Auricular surface

1. Introduction

Estimation of adult age at death is fundamental for all the studies of human skeletal remains: research on fossils, individual identification in the forensic field, paleobiological studies (health and disease, paleodemography). The validity of age-at-death assessment, one of the most difficult problems encountered by biological anthropologists, is necessary for interpretations to be reliable [1]. Biological anthropologists and forensic specialist frequently ignore that there is no linear relationships between chronological age and indicators or disregard the extremely complex variability of the ageing process. As a consequence, data processing is a great source of error, since predictions are elaborated with inappropriate statistical methods. For instance, most of ageing techniques use linear regression to correlate morphological score of one indicator and chronological age. The equation of this regression line is used to convert unknown values of the age indicators into predicted ages. But, the poorer the correlation, the greater is the bias [2]. There is a systematic trend towards overestimating the age of the young adults and underestimating that of older individuals. Given that the correlation between biological data and age is low [3,4], it represents a fundamental limitation to this predictive technique. We argue that artificial neural networks (ANN) increase the reliability of adult age-at-death assessment, since this tool is appropriate when relationship between variables is difficult to model. We applied ANN on the scoring system elaborated by [5] on the pubic symphysis and the sacropelvic surface of the ilium on 677 individuals from identified dry bone collections. As the use of multiple indicators gives better results than individual criterion [6], we pooled both indicators. We used ANN as a classifier in age range categories.

* Corresponding author. Tel.: +33 5 57 57 14 48; fax: +33 5 56 98 57 36.
E-mail address: marc.corsini@u-bordeaux2.fr (M.-M. Corsini).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2004.05.008

2. Materials and methods

We observed 677 individuals, 327 males and 350 females. The mean age is 53.24 years with a minimum of 19 and a maximum of 96, the standard deviation (S.D.) is 18.09, all individuals are within three S.D. from the mean selected from five collections of known age at death. The skeletal sample was selected from five identified collections: the collection of identified skeletons (Departamento de Antropologia, Universita Coimbra, Portugal); the Spitafields collection (Museum of Natural History of London, England); the Gemmerich collection (Département d'Anthropologie et d'Ecologie, Université de Genève, Switzerland); the Alcione collection (Departamento de Medicina Legal, Facultad de Medicina Madrid, Spain) and the Hammann-Todd collection (Museum of Natural History of Cleveland, Ohio, USA). Table 1 shows the number of individuals selected in each collection.

Table 1
Number of individual observed by sex and collections

Collections	Male	Female	*N*
Coimbra	64	73	174
Londres	79	95	137
Geneve	27	19	45
Madrid	33	34	67
USAa[a]	41	44	85
USAe[b]	83	85	168

[a] African origin.
[b] European origin.

2.1. Scoring system

The current methods based on the pubic symphysis and the sacropelvic surface of the ilium have been evaluated on various skeletal collections of known age [7–9]. It has been established that they have poor reliability; moreover the scoring methods they used are rather complex and subjective.

As a consequence, we developed a gradual scale of visual scoring for each character of each criterion [5,10]. We observed three features on the pubic symphysis: posterior plate (billowings, modification, complete); ventral rampart (absent, formation, complete); dorsal lip (absent, present) and four features on the auricular surface of ilium: the transverse organization (striae, absent); the texture and porosity (fine, granulation, coarsening, porosities), the apical activity (absent, present) and the retroarticular activity (absent, present). Short description of the scoring system is given in Table 2. Notice that the label absent does not mean that the feature is missing (i.e. destroyed) but that the feature is not present. The repeatability of this visual scoring between observers is good. Morphological individuals' observations are independent from the observer [5].

2.2. Artificial neural networks (ANN)

ANN is a collective term referring to a variety of highly distributed computational models. Such models are well

Table 2
Brief description of the scoring system for the pubic symphysis and the auricular surface of the ilium

Features	Scoring			
	Score 1	Score 2	Score 3	Score 4
Pubic symphysis				
Posterior plate	Ridges and furrows	Flattening of the surface, dorsal margin begins to appear	Complete	
Ventral plate	Ridges and furrows	Flattening of the surface, rampart expanding	Ventral rampart presents a complete surface	
Dorsal lip	Regular dorsal rim	Lipping of the dorsal rim		
Auricular surface of the ilium				
Transverse organization	Undulations or striae following a transverse organisation	Absent		
Texture and porosity	Dense surface	Uniform or partially coarsening granular texture	Coarse granular texture, distinct signs of bone destruction, porosities	Irregular surface, large bone destruction, deep porosities
Apical activity	Apex sharp and distinct, regular margins	The apex becomes broader, formation of a rim		
Retroauricular activity	Smooth surface	Irregular surface, osteophytes		

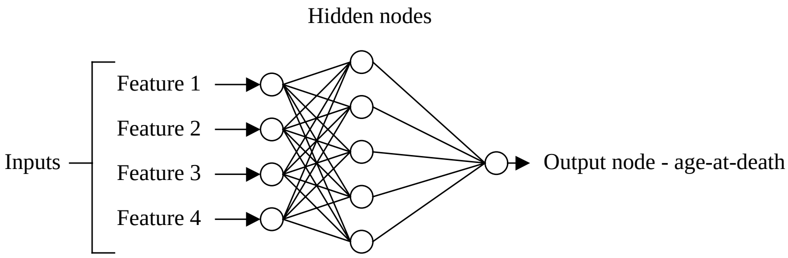

Fig. 1. A multi-layer perceptron.

suited to solve nonlinear problems. The particular type of ANN used in this study is the multi-layer perceptron (mlp) with an architecture comprising input nodes (observations of indicator), hidden nodes and output nodes (age-at-death). Those processing units are interconnected via a set of ''weights''. The weights reflect the strength of the connections between the nodes. Fig. 1 illustrates mlp architecture. When an mlp is initialized the connections are given small arbitrary weights. The network is then trained. Training consists of changing the weights in order to minimize the output error over all the examples. Once the network has been trained, new input items can be presented to it and the predictive output value can be calculated. Detailed explanations can be found elsewhere [11,12].

In this study, the inputs of the mlp are discrete values (observation of the indicator, sex and collection). Since it is well known [13], the way in which the data is represented affects the way in which the network learns, we have experimented three different encoding strategies. (1) Place encoding: one value for each dimension, all but one are off; (2) binary encoding: for P values, one needs $\log_2 P$ units, each possible value is arbitrary assigned to a discrete value; (3) Gray encoding: the only difference with binary encoding is that two contiguous values will only have one bit different.

The target is a continuous numeric function of the age at death of individuals that we intent to map into age categories. We have used the softmax scaling, this encoding is known to be useful for squashing (un)evenly distributed data into an appropriate range. Let $v_1 \ldots v_n$ be the possible values of the variable to be squashed, v denotes the mean, and σ_v be the S.D., λ is the number of S.D. from the mean which are to fall in the linear section of the squashing function.

$$x_i = \frac{(v_i - v)}{(\lambda\sigma_v/2\pi)} \quad \& S_i = \frac{1}{(1 + e^{-xi})}$$

Thus s_i is in the range $0\ldots1$. For the experiment, we set λ to 3 in order to cover 99% of the age distribution (assuming age distribution respects normal distribution).

Two training strategies have been used: batch mode (one backpropagation per epoch) and online mode (one backpropagation per pattern presentation). This makes a total of six classifiers: three encoding strategies * two training strategies.

As we are interested to map individuals age-at-death into range of years, and not, as usually, find the most precise age at death for each individual we have defined a decision procedure as follows. For each input pattern, we get the actual output, map the output to age (in years), and compare the network class output versus the network class target.

Two categories sets were chosen: one is by decennial class (20–29, 30–39, 40–49, 50–59, +60) since mortality profiles of preindustrial societies are available in tables are organized as such, second is a three classes' set (young, middle-age, old), where young stands for less than 32-years-old, old above 59-years-old, and middle-age for the remainder.

The effectiveness of each analysis was evaluated on three criteria: the mean square error (M.S.E.), the weights histograms and rate of good classification which was performed on a validation set process by drawing at random 50% of the data as training data, the remaining as validation set.

3. Results

The mlp has one hidden layer of nine cells; the size of the output layer is one, while the size of the input layer depends on the kind of encoding strategy selected (20 for the place encoding, and 17 for the binary & gray encoding). The activation function for all units is the logistic function: $1/(1 + e^{-xi})$. The learning rate ($\eta = 0.25$) and the momentum ($\alpha = 0.8$) have been set as suggested by the work of Luo [14] in order to avoid oscillations as much as possible. All results were obtained by averaging over a ten times run. The networks were trained during 15 000 epochs, whatever the strategy we used the M.S.E. was 0.02 ± 0.003 for the training set and 0.04 ± 0.005 for the validation set. Considering the Weights histograms, the weights of the net are within $-4\ldots+4$, most of them are concentrated near zero, but they do not show the theoretical hump around 0. Further studies are needed to decide if this behavior is due to structural errors.

Table 3 gives the global rate of good classification for both the 5 and 3 age categories on the validation set. At first glance, it seems that batch learning is a bit better than online learning and that binary encoding gives worst result.

Table 3
Global restitution rate on validation set

Mode	Encoding strategy		
	Place	Binary	Gray
5 Categories classification			
Batch	46.03	37.14	45.08
On line	40.32	42.86	45.71
3 Categories classification			
Batch	66.67	64.76	66.35
On Line	58.41	61.59	62.86

Table 4
Performance (a posteriori) on validation set

Mode	Encoding	20–29	30–39	40–49	50–59	>60
Batch	Place	90.00	28.57	25.00	24.41	73.58
Online	Place	72.41	30.95	23.73	19.04	62.38
Batch	Binary	80.00	25.00	22.98	28.57	76.82
Online	Binary	59.45	31.57	25.45	19.67	60.48
Batch	Gray	0.00	26.98	23.53	26.67	66.13
Online	Gray	60.53	26.09	28.36	18.87	64.18

If we now, pay attention to the performance by age category, i.e. a posteriori rate for the validation set with 5 age categories (Table 4), Batch mode performs better than Online mode at identifying the youngest and oldest individuals, except the gray encoding. However, both modes fail in the identification of the intermediate categories. Another surprising tendency is that gray encoding does not perform well on any categories compare to the other encoding systems. This is surprising since there is no real order on the values, note also that the gray encoding differs from the binary encoding only for features with more than 3 possible values, and that this happens only for the ‘‘posterior plate’’, the ‘‘ventral rampart’’ and the ‘‘texture and porosity’’.

4. Discussion

It has been reported [15] that 75–85% of individuals are misclassified with the pubic symphyseal stages, and that only 35 among 125 individuals were well classified with the method based on the sacropelvic surface of the ilium. Considering the estimated age-at-death, the average error is of 12 years, but it might be up to 22 years for the oldest skeletons [16,17] with the method based on pubic symphysis. For the method based on sacropelvic surface of the ilium, [16,18] reported an average age error of 8–10 years, but of 16–24 for the oldest. These results led us to reconsider the methods, by proposing a new gradual scale of visual scoring for each feature and by using a multilayer perceptron as classifier. We have reported that the network performance is almost insensible to the encoding strategy we choose. The results we obtained show a real improvement in adult age at death assessment, even though the perceptron architecture has not been tuned as much as it could be. We obtained reliable classification that distinguishes three age intervals: 20–29, 32–59, and over 60 years old. Identifying individuals aged over 60 is particularly interesting. Indeed, archaeological samples that have been aged using skeletal morphological indicators show an absence of or at least very few of those individuals. Test of methods on known age-at-death sample indicate the same trend. Our methodology enables to encounter this systematic bias. Because of the variability of ageing process, age estimation must focus on reliability instead of accuracy.

The use of ANN as a classifier appears as a useful tool to progress in age-at-death assessment. Identification of reliable intervals for each individual represents a crucial information for both forensic and physical anthropology topics. As this is a work in progress, a lot of improvements still remain to be done: varying the learning and momentum rates, exploring other architectures (the number of hidden units, the connections) and including new skeletal samples.

References

[1] M. Jackes, Building the bases for paleodemographic analysis, in: M.A. Katzenberg, S.R. Saunders (Eds.), Biological Anthropology of the Human Skeleton, Wiley-Liss, 2000, pp. 417–466.
[2] R.G. Aykroyd, D. Lucy, A.M. Pollard, C.A. Roberts, Nasty, brutish, but not necessarily short:a reconsideration of the statistical method used to calculate age at death from adult human skeletal and dental age indicators, Am. Antiq. 64 (1999) 55–70.
[3] C. Masset, Estimation de l'âge au décès par les sutures crâniennes. Thèse de Doctorat d'Etat: Université Paris I, 1982. p. 301.
[4] F.W. Rösing, S.I. Kvaal, Dental age in adults—a review of estimation methods, in: K.W. Alt, F.W. Rösing, M. Teschler-Nicola(Eds.), Dental Anthropology. Fundamentals, Limits, and Prospects, Springer Wien, New-York, 1998, pp. 443–468.
[5] A. Schmitt, Variabilité de la sénescence du squelette humain. Réflexions sur les indicateurs osseux: à la recherche d'un outil performant pour estimer l'âge au décès des adultes, Thèse de doctorat, Laboratoire d'Anthropologie des Populations du Passé – Université Bordeaux 1, 2001, p. 347.
[6] R.S. Meindl, K.F. Russel, Recent advances in method and theory in paleodemography, Annu. Rev. Anthropol. 27 (1997) 375–399.
[7] D. Katz, J.M. Suchey, Age determination of the male os pubis, Am. J. Phys. Anthrop. 69 (1986) 427–435.
[8] J.M. Suchey, D.V. Wiseley, D. Katz, Evaluation of the Todd and McKern Stewart methods for aging the male os pubis, in: K. Reichs (Ed.), Forensic Osteology: Advances in the Identification on Human Remains, Charles Thomas, Springfield, 1986, pp. 33–67.
[9] C.O. Lovejoy, R.S. Meindl, T.R. Prysbeck, R.P. Mensforth, Chronological metamorphosis of the auricular surface of the

ilium: a new method for the determination of adult skeletal age at death, Am. J. Phys. Anthrop. 68 (1985) 15–28.

[10] A. Schmitt, C. Broqua, Approche probabiliste pour estimer l'âge au décès a partir de la surface auriculaire de l'ilium, Bull. et Mém Soc. Anthropol. Paris 12 (2000) 279–301.

[11] J. Hertz, A. Krogh, R. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Reading, MA, 1995.

[12] G. Hinton, Connectionist learning procedures, Artif. Intell. 40 (1989) 185–234.

[13] K. Swingler, Applying Neural Networks: A practical Guide, Academic Press, Londres, 1996.

[14] Z. Luo, On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks, Neural. Comput. 3 (1991) 226–245.

[15] A.L. Santos, How old is this pelvis? a comparison of age of death estimation using the auricular surface of the ilium and os pubis, in: Papers of the 10th Congress of the Pan African Association for Prehistory and Related Studies, Aspects of African Archaeology, 1996, pp. 29–36.

[16] S.R. Saunders, C. Fitzgerald, T. Rogers, C. Dudar, H. Mc Killop, A test of several methods of skeletal age estimation using a documented archaeological sample, Can. Soc. Forensic Sci. 25 (1992) 97–118.

[17] R.D. Hoppa, Population variation in osteological aging criteria: An example from the pubic symphysis, Am. J. Phys. Anthrop. 111 (2000) 185–191.

[18] K.A. Murray, T. Murray, A test of the auricular surface aging technique, J. Forensic Sci. 36 (1991) 1162–1169.