Bone Age Assessment Using Support Vector Machine Regression

Daniel Haak, Hendrik Simon, Jing Yu, Markus Harmsen, Thomas M. Deserno

Department of Medical Informatics, RWTH Aachen University
dhaak@mi.rwth-aachen.de

Abstract. Bone age assessment on hand radiographs is a costly and time consuming task in radiology. Recently, an automatic approach combining content-based image retrieval and support vector machines (SVM) has been developed. In this paper, we we apply support vector regression (SVR) as a novel method, yielding a gain in performance. Our methods are designed to cope with the age range 0-18 years as compared to the age range 2-17 of the commercial product BoneXpert. On a standard data set from University of South Carolina, our approaches reach a root-mean-square error of 0.95 and 0.80 years for SVM and SVR, respectively. This is slightly below the performance of the commercial product using an active shape approach.

1 Introduction

Bone age assessment (BAA) is a frequently and time consuming method for growth disturbances determination in the human body. Usually this task is done manually by radiologists on hand radiographs, requiring domain knowledge and experience [1, 2]. Previous work on BAA [3, 4, 5] presented first approaches in automation of this process with content-based image retrieval (CBIR) methods. In [4, 5] new results have been published by combining a support vector machine (SVM) for classification and cross-correlation similarity to a prototype image as feature vector. Although this is a promising method the SVM needs to be extended by multiclass adaption for BAA. The age is discretized in 30 classes and the prediction is calculated by one-against-one voting.

In this work, we apply with support vector regression (SVR) [6, 7, 8] a regression model as a novel approach that naturally supports multiclass problems.

2 Materials and methods

Fig. 1 illustrates the processing pipeline of our approach. At first, the epiphysal centers are located on the radiographs. Then eROIs are extracted, and rotated into a reference position. Cross correlation features are extracted. Classification is done by SVM and SVR.

H.-P. Meinzer et al. (Hrsg.), *Bildverarbeitung für die Medizin 2013*, Informatik aktuell, DOI: 10.1007/978-3-642-36480-8_30, © Springer-Verlag Berlin Heidelberg 2013

2.1 Evaluation data

The standard data available for research is the USC hand atlas that is composed of 1,097 images from different age classes, gender and ethnics [9]. In previous work an automatic approach for eROI extraction, presented by Fischer et al. [10], has been used. However, the automatic method can suffer from artefact, e.g. noise and misplacement, in the images. This leads to errors in feature extraction and affects the classication performance. To avoid those errors, here a semi-automatic approach for eROI extraction is used, where eROI location is performed manually. To be comparable to [4, 5], the USC data has been reprocessed. In total, 14 available epiphyseal regions are used dismissing the five regions close to the wrist. Fig. 2 shows the epiphysal centers and their corresponding region numbers. eROIs are extracted and normalized with respect to vertical alignment (rotation). Based on 512×512 pixel images, an eROI is of size 60×50. Hence, our experiments are based on 29,050 eROIs.

In [11], 15 images have been removed from USC data. This data is denoted USC-15.

2.2 SVM and SVR

Prototypes for each class are determined and the cross-correlation function (CCF) is used for feature extraction to represent similarities between the images [4, 5]. The resulting feature vectors include the gender of the patient and CCF values for each class and epiphyseal region. Data management is performed using the Image Retrieval in Medical Applications (IRMA) framework (http://irma-project.org).

To cope with the high number of classes, the one-against-one approach is used as multi-class extension for SVM. Contrarly, the SVR inherently handles

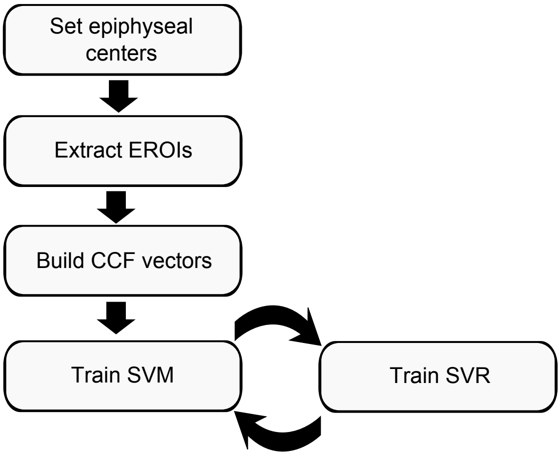

Fig. 1. Integration of SVR into the processing pipeline.

multi-class prediction. Given a training data set $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ of size n with corresponding target values, SVR tries to approximate a target function $f(x)$. This function has maximum deviation of ϵ from target values y_i in the training data. The parameter ϵ is critical and needs to be chosen carefully. A too high ϵ might result in an arbitrarily bad approximation of the target value function, where a too small ϵ leads to overfitting. Optimizing the approximation with respect to a ϵ-intensive loss function results in a hyperplane regression model. The ϵ-intensive loss function penalizes data points with higher deviation than ϵ.

Like in SVM, this hyperplane is only affected by the support vectors, which are those data points having a loss greater than 0 or lying directly on the ϵ-border. The main difference to SVM is that the resulting hyperplane function produces a continuous output, which can be interpreted as a soft class assignment. Unlike SVM, the regression function is already multi-class capable, since it approximates the target values, i.e. the class labels. For our method we use the SVR library coming with libSVM, which is also used by [4, 5]. Further, the ν-SVR method is used, since it indirectly controls ϵ via a parameter ν. This new parameter trades off ϵ against the model complexity and individual error tolerances of the training data, also known as slack variables [7]. Since the regression function is trained on the class labels, the output represents a class and needs to be discretized, which is done by a mapping to discrete values. In this first approach, the mapping is done by simple truncation of the SVR output.

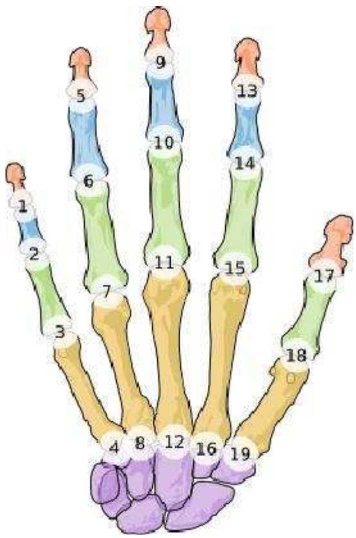

Fig. 2. EROIs and corresponding region numbers.

Table 1. SVM & SVR optimal parameters and regions.

Method	C	γ	ν	Regions
SVM	2048	0.0078125	0.5	2,6,11,13,15,18
SVR	128	0.003125	0.5	6,7,9,10,11,13,14,15,18

2.3 Evaluation

To produce comparable results SVM experiments done by [4, 5] were repeated with reprocessed USC images. SVM and SVR were evaluated using five-fold cross-validation. Class prototypes were chosen randomly from both datasets, but fixed for all experiments. Optimal parameters for SVM and SVR were determined using grid search and grid regression, respectively. The optimal subsets of regions for SVM and SVR were analyzed by calculating experiments on each possible subset of the 14 regions. The determined parameters and best regions for both methods are shown in Tab. 1.

For all experiments, the mean error (ME) was computed

$$\mathrm{ME} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| r^i_{\mathrm{est}} - r^i_{\mathrm{rad}} \right| \tag{1}$$

where n denotes the size of the dataset, r_{est} and r_{rad} denote the estimated age and the radiologist reading respectively.

3 Results

The recomputation of the experiments done by [4, 5] with semi-automatic feature extraction can be found in Tab. 2, where the same data set and SVM parameters were used here. Although we used a different processing pipeline, resulting in different EROIs, the results are very similiar. So we can conclude that the semi-automatic approach shows the same performance as the automatic method. For all other experiments on the USC atlas, new optimal parameters and subsets were used, resulting in slightly deviating results.

Comparing SVR to SVM (Tab. 3) it turns out that the mean error decreases from 0.832 to 0.768 by 7% for the age range 0-18 years and drops even more significantly from 0.798 to 0.692 by 13% in the age range 2-17 years. Comparing with BoneXpert the RMS error decreases from 0.95 to 0.80 by 15%. The percentage of correctly classified hands is reduced from 35.42% to 32.52% (Tab. 4),

Age range (years)	USC atlas	USC atlas (new EROIs)
$0-18$	0.832	0.835
$2-17$	0.826	0.819

Table 2. Mean classification error in years, using the setup of [4, 5].

Table 3. Mean error and standard deviation in years. RMS denotes the root mean squared error.

Dataset	Age range (years)	SVM	SVR	BoneXpert [11]
USC	$0-18$	0.832 ± 0.775	0.768 ± 0.657	–
	$2-17$	0.798 ± 0.658	0.692 ± 0.572	–
USC-15	$2-17$	0.950 (RMS)	0.799 (RMS)	0.61 (RMS)

but the overall class distance is significantly lower. Most hands have a class distance of at most two, which indicates an error of approximately one year.

4 Discussion

USC data has been re-processed, resulting in new training data. The performance of SVM comparing the old and new USC dataset is stable and yields negligibly different results. So the semi-automatic and automatic approach show the same performance. A reason for this might be that the standard USC data set with carfeully selected images only sparsely contains radiographs with artefacts. Probably this image set is not optimal for showing improvement by manual eROI center location. Further experiments with additional BAA image sets including noisy images from daily routine can result in further knowledge. Anyway the SVM produces stable results and seems to be robust against small shifts in the data.

With SVR a novel method was introduced and evaluated on USC data. It turned out that this approach yields significantly better results than SVM, especially on the age range 2-17 years, but still does not yet reach the performance of BoneXpert. A smarter class mapping from SVR output could even further decrease the error rate. Another idea is the interpretation of BAA as natural regression problem, avoiding detour over (artificial) age classes. Here, the SVR can directly work on data with age readings as labels and can predict the age without classification.

The fixed random prototype selection also offers potential for improvement. Generation of prototypes by mean values or optimizied selection of prototype images can yield in better results.

Class distance	0	1	2	3	4	5	6	7	8	9	10	≥ 11
SVM hits(%)	35.42	42.33	14.35	5.18	1.27	0.54	0.54	0.18	0.00	0.00	0.00	0.18
		$\leftarrow$ 92.1 $\rightarrow$						$\leftarrow$ 7.9 $\rightarrow$				
SVR hits(%)	32.52	46.23	15.53	3.36	1.09	0.73	0.09	0.27	0.09	0.00	0.00	0.09
		$\leftarrow$ 94.28 $\rightarrow$						$\leftarrow$ 5.72 $\rightarrow$				

Table 4. Class distances of SVM & SVR in the range of $0-18$ years on the USC data.

Furthermore, the best region selection can be optimized. An option is training several SVM classifiers working on different region sets. Preprocessing of data can then be used to assign each image to the most applicable classifier.

Anyway, comprehensive evaluation of hand radiograph data has shown that BAA prediction in computer-aided diagnostics (CAD) offers promising results for further investigations. In special, the suitability of this methods in practical routine should be evaluated.

References

1. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. Am J Med Sci. 1959;238(3):393.
2. Tanner J, Healy M, Goldstein H, et al. Assessment of skeletal maturity and prediction of adult height (TW3). WB Saunders, London. 2001.
3. Martin M, Martin-Fernandez M, Alberola-Lopez C. Automatic Bone Age Assessment: A Registration Approach. Proc SPIE. 2003;5032:1765–76.
4. Harmsen M, Fischer B, Schramm H, et al. Support vector machine classification using correlation prototypes for bone age assessment. Proc BVM. 2012; p. 434–9.
5. Harmsen M, Fischer B, Schramm H, et al. Support vector machine classification based on correlation prototypes applied to bone age assessment. IEEE Trans Inf Technol Biomed. 2012;PP(99):1.
6. Basak D, Pal S, Patranabis DC. Support vector regression. Neural Inform Process Lett Rev. 2007;11(10):203–24.
7. Schölkopf B, Smola AJ, Williamson RC, et al. New support vector algorithms. Neural Comput. 2000;12(5):1207–45.
8. Schölkopf B, Burges CJC, Smola AJ. Advances in kernel methods: support vector learning. MIT Press; 1998.
9. Gertych A, Zhang A, Sayre J, et al. Bone age assessment of children using a digital hand atlas. Comput Med Imaging Graph. 2007;31(4-5):322.
10. Fischer B, Brosig A, Welter P, et al. Content-based image retrieval applied to bone age assessment. Proc SPIE. 2010;7624:762412.
11. Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol. 2010;17(11):1425–32.