Support Vector Machine Classification using Correlation Prototypes for Bone Age Assessment

Markus Harmsen[1], Benedikt Fischer[1], Hauke Schramm[2], Thomas M. Deserno[1]

[1]Department of Medical Informatics, RWTH Aachen University, Aachen, Germany
[2]Department of Applied Computer Science, Fachhochschule Kiel, Kiel, Germany
markus.harmsen@rwth-aachen.de

Abstract. Bone age assessment (BAA) on hand radiographs is a frequent and time consuming task in radiology. Our method for automatic BAA is done in several steps: (i) extract of 14 epiphyseal regions from the radiographs, (ii) for each region, retain image features using the IRMA framework, (iii) use these features to build a classifier model, (iv) classify unknown hand images. In this paper, we combine a support vector machine (SVM) with cross-correlation to a prototype image for each class. These prototypes are obtained choosing the most similar image in each class according to mean cross-correlation. Comparing SVM with k nearest neighbor (kNN) classification, a systematic evaluation is presented using 1,097 images of 30 diagnostic classes. Mean error in age prediction is reduced from 1.0 to 0.9 years for 5-NN and SVM, respectively.

1 Introduction

Bone age assessment (BAA) on hand radiographs is a frequent and time consuming task in diagnostic radiology. In the method developed by Greulich & Pyle [1], the radiologist compares all bones of the hand (Fig. 1) with a standard atlas, whereas according to Tanner & Whitehouse [2], only a certain subset of bones is considered. Different approaches have been developed to automate this

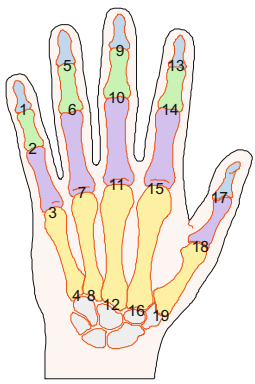


Fig. 1. Epiphysial regions of interest (eROIs) and their corresponding numbers

T. Tolxdorff et al. (Hrsg.), *Bildverarbeitung für die Medizin 2012,* Informatik aktuell,
DOI 10.1007/978-3-642-28502-8_75,

process, including the approach of Fischer et al. [3], in which (i) the epiphyseal/metaphyseal regions (eROIs) are extracted, (ii) using content-based image retrieval (CBIR), similar images to these regions are queried from a database, (iii) classified using the k nearest neighbor (kNN) method and (iv) algebraically combined to propose a bone age.

Although this approach is promising, some weak points have been identified: (i) a fixed amount of k neighbors is used for classification, which – depending on the dataset – may not be optimal; (ii) the classification considers the eROIs to be independent and uses only a (weighted) age average of these regions for age determination; (iii) the gender is disregarded completely, although male and female growth spurts differ significantly; (iv) BAA is performed numerically, based on similar images, and does not use any medical useful classification in respect of growth spurts; and (v) computation is expensive, since the cross-correlation between all existing reference images is determined.

In the past few years, the support vector machine (SVM) has been introduced into many classification fields and have demonstrated the state of the art performance. For example a combination of SVM with CBIR has been successfully applied to detect malign structures in mammography [4]. Despite of their broad applicability, some essential problems have to be addressed when using SVM. Besides the fundamental choice of attributes, the SVM only classifies binary problems, i.e., a classification into more than two classes is only possible with several SVMs, and the class size has to be chosen carefully.

In this work, our method on automatic CBIR-based BAA is extended by SVM and evaluated critically with respect to the standard kNN classifier.

2 Materials and methods

Our approach is integrated with the image retrieval in medical applications (IRMA) framework (http://irma-project.org), which supports content-based access to large medical image repositories [5]. Global, local, and structural features are supported to describe the image, an eROI, or a constellation of eROIs, respectively [6]. Furthermore, our work is based on Fischer et al. [3].

2.1 ROI extraction

Extracting the eROI has been presented previously [7, 8, 9]. Essentially, a structural prototype is trained where the phalanges and metacarpal bones are represented by nodes, and location, shape as well as texture parameters are modeled with Gaussians. The centers of eROIs are located automatically. Fourteen eROIs are extracted, geometrically oriented into an upright position, and inserted into the IRMA database with reference to the according hand radiography.

2.2 Class prototypes

To address the problem of class size, we have grouped the data related to the growth spurts. According to the ontology defined by Gilsanz & Ratib [10],

reference ages can be quantized in steps of 2 m, 4 m, 6 m, and 12 m for the intervals [8 m ... 20 m), [20 m ... 28 m), [2.5 y ... 6 y), and [6 y ... 18 y], respectively, where m and y denotes months and years, respectively. This creates a set of 29 classes with four different ranges. A 30^{th} class for bone ages > 18 years was added. Notice that gender information is so far not used for class building. All eROIs and their according attributes are loaded from the IRMA database and scaled to the range $[-1, +1]$ to avoid attributes in greater numeric ranges dominating those in smaller ranges.

2.3 Feature extraction

The cross-correlation function (CCF) is easy to compute, robust regarding the radiation dose, and has already been used successfully in BAA tasks [3]. The similarity between two eROIs q and p is computed by

$$S_{\text{CCF}}(p,q) = \max_{|m|,|n| \le d} \left\{ \frac{\sum_{x=1}^{X} \sum_{y=1}^{Y} (p(x-m, y-n) - \overline{p})\,(q(x,y) - \overline{q})}{\sqrt{\left(\sum_{x=1}^{X} \sum_{y=1}^{Y} (p(x-m, y-n) - \overline{p})^2 \right) \left(\sum_{x=1}^{X} \sum_{y=1}^{Y} (q(x,y) - \overline{q})^2 \right)}} \right. \tag{1}$$

where $\overline{p}, \overline{q}$ denote the mean gray values of p, q respectively, and d the warp range for maximum correlation. In our experiments, we set $d = 2$ and use 32×32 pixel scaled version of the eROIS.

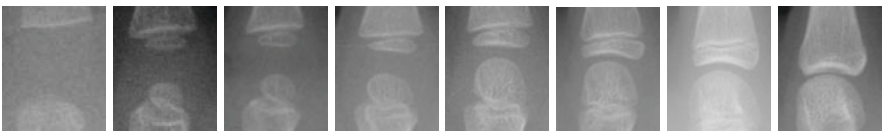
For each region R and class C, we select the optimal image I according to mean CCF similarity as our feature prototype

$$F_{R,C} = \max_{I_x \in R,C} \left\{ \frac{1}{n} \cdot \sum_{i=1}^{n} (S_{CCF}(I_x, I_i)) \; | I_i \in R, C, i \neq x \right. \tag{2}$$

We use the IRMA framework to index eROI features extracted from the 30 class prototypes (Fig. 2).

2.4 Creating a classification model

For SVM classification, binary classes are usually presented as two features ranging between zero and one. Hence, the data instance vector used for classification with kNN as well as SVM yields



0.0 – 0.7 1.2 – 1.3 2.0 – 2.3 3.5 – 4.0 5.5 – 6.0 9 – 10 13 – 14 17 – 18

Fig. 2. Subset of prototypes for eROI 11 with maximum average similarity in the responding age class clearly indicating the development of epiphyses. The y-translation on age class $5.5 - 6.0$ will be normalized by CCF

$$\boldsymbol{v}_{\text{hand}} = \begin{pmatrix} g_m \\ g_f \\ E_{01} \\ \vdots \\ E_{14} \end{pmatrix}, \quad E_i = \begin{pmatrix} F_{i1} \\ \vdots \\ F_{i30} \end{pmatrix} \tag{3}$$

where g denotes the gender (g_m is 1 if gender is male, 0 otherwise. g_f in analogous), E_i donates the loaded feature vector for the currently referenced hand eROI i with feature parameters F_{i1} till F_{i30} for each class similarity. Notice that all feature parameters F consist of a vector with further numeric values, describing specific image characteristics.

2.5 Age computation

For a hand radiograph with unknown bone age, the eROIs are extracted (cf. Sect. 2.1). For each eROI, the features are extracted (cf. Sect. 2.3), and the data instance vector v_x is built (cf. Sect. 2.4). The SVM and the trained classificatory model is used to classify the new radiograph. For the SVM-based classifier, we used the "one-against-one" approach instead of "one-against-all" due to its good performance and short training time. A comparison of these methods can be found in [11]. The bone age is determined by

$$\text{age} = \frac{1}{2}\Big(\text{upper}\big(\text{class}(v_x)\big) + \text{lower}\big(\text{class}(v_x)\big)\Big) \tag{4}$$

where upper(class(x)) donates the upper age bound of class x and lower(class(x)) the lower age bound of class x analogous.

2.6 Validation experiments

The class prototypes were determined from the 1,097 images of the USC hand atlas. Afterwards, for each hand, the feature vector is generated by measuring the similarity to corresponding region class-prototypes as well as the gender (cf. Sect. 2.4). Now we apply the same feature vectors for kNN and SVM, using k-fold cross-validation for SVM with $k = 5$ and leave-one-out cross-validation for kNN. For SVM, optimal classification parameters will be computed by grid search.

3 Results

In terms of accuracy, the results for kNN range from $12\% - 22\%$ and $13\% - 27\%$. In terms of mean age error they range from $1.38 - 2.65$ years and $1.02 - 2.32$ years for $k = 1$ and $k = 5$, respectively for individual regions. The SVM results in terms of accuracy $21\% - 32\%$ and mean age error $0.99 - 2.34$ years (Tab. 1).

Table 1. Experiment outcome for kNN/SVM and single regions. The second experiment with $k = 5$, as well as SVM, also use the gender. SVM parameters $C = 2048$ and $\gamma = 0.0078125$ have been determined via grid search. Max and min values are bold, some regions have been omitted.

	kNN $k = 1$		kNN $k = 5$		SVM	
Region	Accuracy	Mean error	Accuracy	Mean error	Accuracy	Mean error
1	14.36 %	**2.65**	**13.00 %**	**2.32**	21.36 %	**2.34**
2	**12.18 %**	2.43	13.55 %	2.00	**20.36 %**	2.32
7	20.09 %	1.63	24.18 %	1.18	31.27 %	1.07
11	**22.00 %**	**1.38**	**27.64 %**	**1.02**	**32.45 %**	**0.99**
15	19.64 %	1.44	25.36 %	1.11	30.09 %	1.11

Table 2. Classification results for multiple regions. A distance of 0 denotes a correct labeled class, whereas a distance of 1 indicates the classifier has labeled a wrong class directly one before or after the actual age class.

Distance	0	1	2	3	4	5	6	7	8	9	10	...	29
kNN hits(%)	26.34	44.76	18.51	6.74	1.82	1.0	0.46	0.27	0.18	0.18	0	...	0
		⟵ 89.61 ⟶						⟵ 10.65 ⟶					
SVM hits(%)	33.36	38.29	17.59	5.9	1.73	1.82	0.64	0.36	0.18	0.09	0	...	0
		⟵ 89.24 ⟶						⟵ 10.72 ⟶					

Table 3. Best results from kNN and SVM using different eROIS and parameters. The Clopper-Pearson intervals donate the propotion of labeled hands with an error in the interval $[-2; 2]$ for $\beta = 0.95$.

Classifyer	Accuracy	Mean error	Conf. int.	Used regions	Parameter
kNN	26.34 %	1.00	[0.886; 0.922]	7, 11, 15	$k = 5$
SVM	33.36 %	0.90	[0.903; 0.936]	2, 5, 7, 8, 10, 11, 13	$C = 2048$
					$\gamma = 0.0078125$

The best experimental result for kNN (accuracy 26.27%, mean error 1.00 years) is optained by a subset of 3 regions and $k = 5$. Best results for SVM (accuracy 33.36%, mean error 0.90 years) use a subset of 7 regions (Tab. 3) and therefore outperform kNN as well as the method by Fischer et al., which has a mean error of 0.97 years. Both perform best only on a subset of regions, whereas the SVM can use larger data.

4 Discussion

The overall classification accuracy seems to be low, but most mislabeled classes only have a class distance of 1 or 2 (Tab. 2). Using a larger amount of eROIS for classification, the SVM reaches a mean error of 0.90 years. Therefore SVM

outperforms kNN, which has a mean error of 1.00 years (Table 3). The true propotion for our svm conficence intervall is also slighly higher as for the kNN one (Tab. 3, column 4). For this reasons, SVM clearly improves our previous method as it can handle a larger amount of attributes. It's remarkable that our SVM only uses gender and CCF as features and may easily enriched by further features. Computation time for SVM is even lower than for kNN, since the classification model is only build once and then used for all unknown hands. Furthermore using correlation prototypes remarkably improves the performance.

So far, gender is used only as an attribute for classification and not for prototype building, since the classifier should implicit model the fact of different growth spurts for male and female subjects. This should be verified by using the gender to build twice as many prototypes – and therefore increasing the feature space – and repeating the experiments. Another possible improvement might be using prototypes from a standard reference atlas – which is also used in radiology and therefore perfectly match our age classes – instead of computing own prototypes.

References

1. Greulich WW, Pyle SI. Radiographic Atlas of Skeletal Development of Hand Wrist. Stanford CA.: Stanford University Press; 1971.
2. Tanner JM, Healy MJR, Goldstein H, et al. Assessment of Skeletal Maturity and Prediction of Adult Height (TW3) Method. London: WBSaunders; 2001.
3. Fischer B, Welter P, Grouls C, et al. Bone age assessment by content-based image retrieval and case-based reasoning. Proc SPIE. 2011;7963.
4. de Oliveira JEE, Machado A, Chavez G, et al. A. MammoSys: a content-based image retrieval system using breast density patterns. Comput Methods Programs Biomed. 2010;99(3):289–297.
5. Lehmann TM, Gueld MO, Thies C, et al. Content-based image retrieval in medical applications. Methods Inf Med. 2004;43(4):354–61.
6. Deserno TM, Antani S, Long R. Ontology of gaps in content-based image retrieval. J Digit Imaging. 2008;online-first, DOI 10.1007/s10278-007-9092-x.
7. Deserno TM, Beier D, Thies C, et al. Segmentation of medical images combining local, regional, global, and hierarchical distances into a bottom-up region merging scheme. Proc SPIE. 2005;5747(1):546–555.
8. Thies C, Schmidt-Borreda M, Seidl T, et al. A classification framework for content-based extraction of biomedical objects from hierarchically decomposed images. Proc SPIE. 2006;6144:559–568.
9. Fischer B, Sauren M, Gueld MO, et al. Scene analysis with structural prototypes for content-based image retrieval in medicine. Proc SPIE. 2008;6914; online first, DOI 10.1117/12.770541.
10. Gilsanz V, Ratib O. Hand Bone Age: A Digital Atlas of Skeletal Maturity. Berlin: Springer-Verlag; 2005.
11. Hsu CW, Lin CL. A comparison of methods for multi-class support vector machines. IEEE Trans Neural Netw. 2002;13:415–425.