Expert Systems with Applications 40 (2013) 811–819

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Ensemble-based regression analysis of multimodal medical data for osteopenia diagnosis

Wei-Liang Tay [a,*], Chee-Kong Chui [b,1], Sim-Heng Ong [a,c], Alvin Choong-Meng Ng [d]

[a] Department of Electrical and Computer Engineering, National University of Singapore, Block E4, Level 8, Room 25, 4 Engineering Drive 3, Singapore 117583, Singapore
[b] Department of Mechanical Engineering, National University of Singapore, Block E1, Level 5, Room 17, Engineering Drive 2, Singapore 117576, Singapore
[c] Department of Bioengineering, National University of Singapore, Block E4, Level 5, Room 14, 4 Engineering Drive 3, Singapore 117583, Singapore
[d] Department of Endocrinology, Singapore General Hospital, Block 6, Level 6, Outram Road, Singapore 169608, Singapore

ARTICLE INFO

Keywords:
Ensemble-based systems
Regression
Osteoporosis screening
Diagnostic CT
Areal bone mineral density

ABSTRACT

Areal bone mineral density (aBMD) is used in clinical practice to diagnose osteoporosis. In previous studies, aBMD was estimated from diagnostic computed tomography (dCT) images, but a battery of medical tests was also taken that can be used to improve the regression performance. However, it is difficult to exploit the multimodal data as the additional features have poor informativeness and may lead to overfitting. An ensemble-based framework is proposed to improve the regression accuracy and robustness on multimodal medical data with a high relative dimensionality. Instead of case-wise bootstrap aggregating, a filtering-based metalearner scheme was employed to build feature-wise ensembles. The proposed approach was evaluated on clinical data and was found to be superior to bagging and other ensemble methods. The feature-wise ensembling approach can also be used to automatically determine if any multimodal features are related to bone mineral density. Several blood measurements were identified to be linked with bone mineral density, and a literature search supported the automatic identification results.

1. Introduction

In clinical studies, besides the main modalities being studied, other medical measurements are often taken. For a radiological study, it is common to also take blood and hormone measurements for control purposes. These multimodal data are often left unstudied as they are not the focus of the investigation. However, there may be hidden relationships between the disease symptoms and these multimodal data. Although it is likely that any hidden relationships are weaker than the primary modality, there is potential for the primary relationship to be improved by exploiting the hidden information contained in multimodal data. In this study, we are interested in using blood, hormone, and physical measurements to improve the areal bone mineral density (aBMD) estimated from diagnostic computed tomography (dCT). It is not feasible to solve the problem by directly applying multivariate regression, as the additional multimodal features are less informative. The increased ratio of features to training cases also introduces the problem of high relative dimensionality, which may lead to overfitting.

Ensemble method have favorable properties that make them suitable for datasets with high dimensionality (Moon et al., 2007), high class imbalance (Lo et al., 2008), or missing features (Nanni, Lumini, & Braham, 2012). Data from medical studies typically suffer from one or more of the above conditions, due to the difficulty and cost of acquiring clinical data. Ensemble methods are therefore suitable to be applied to medical datasets. By modifying the ensemble method (Wall, Cunningham, Walsh, & Byrne, 2003), it is also possible to explain the ensemble results for decision support systems, which may provide insights into the disease condition and improve operator confidence in the ensemble decision. One limitation, however, is that most work on ensemble method have largely been carried out on classification rather than regression problems.

In this paper, we study how ensemble regression methods can be applied to solve a regression problem on a multimodal medical dataset with high relative dimensionality. Based on insights obtained by using several feature selection and data transformation techniques with linear regression, an ensemble regression method using filtering is proposed. The filtering-based ensemble technique chooses a set of regressors from several candidate regressors such that the component regressors are diverse and uncorrelated. The proposed method generates the best results on the multimodal medical data and can be used to mine the most informative features of the dataset.

* Corresponding author. Tel.: +65 94370323.
E-mail address: tayweiliang@nus.edu.sg (W.-L. Tay).
[1] Tel.: +65 65161336.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.031

2. Related work

In clinical practice, dual-energy X-ray absorptiometry (DXA) is a dedicated imaging modality that generates an aBMD score by which osteoporosis and osteopenia can be diagnosed (World Health Organization, 1994). To facilitate opportunistic bone screening, recent studies (Link et al., 2004; Tay, Chui, Ong, & Ng, 2012) have tried to estimate an DXA-equivalent aBMD score using other imaging modalities that are commonly used in surgical planning or diagnosis. Diagnostic computed tomography (dCT) is a promising modality for opportunistic screening as it is performed frequently and contains densitometric information correlated to BMD (Revilla, Cardenas, Hernndez, Villa, & Rico, 1995; Schreiber, Anderson, Humberto, Buchholz, & Au, 2011). However, while it is feasible to use dCT scans to estimate DXA-equivalent aBMD, several factors inherent to dCT imaging, such as beam hardening (Zhang, Yan, Chui, & Ong, 2010), can adversely affect the reliability of the estimation results. Radiological modalities may also require machine-wise calibration to account for differences in beam and source properties. One way to increase the robustness of aBMD estimation is to incorporate additional features to the prediction model (Carrino & Ohno-Machado, 2005). These additional features can be diagnostic factors (Mantzaris, Anastassopoulos, Iliadis, Kazakos, & Papadopoulos, 2010) that are unrelated and independent of dCT, or describe other aspects, such as the topological, morphological, and mechanical properties (Akgundogdu, Jennane, Aufort, Benhamou, & Ucan, 2010), of the dCT information. In this work, we generate two additional sets of features to improve the aBMD estimation. The first set of additional features describe the HU distributions and morphological features of the bone, and is drawn from dCT data. For the second set of features, we exploit the physical, blood, and hormone data that was also recorded during the clinical experiments. This second set of features provides a multimodal dataset that is independent of dCT, and may be helpful in increasing the robustness of the regression.

Machine learning is a popular approach for computer-aided diagnosis, and was previously used to diagnose fractures (Li et al., 2009) and osteoporotic diseases (Tay, Chui, Ong, & Ng, 2011; Valentinitsch, Patsch, Mueller, Kainberger, & Langs, 2010) based on QCT images. These methods are capable of achieving good detection rates, but typically involve the use of black boxes, which makes it difficult to evaluate their reliability and generality without more extensive clinical validation. Also, most classification algorithms return only an outcome value, or a bias value at best, which makes it difficult to estimate the severity of the diagnosed condition. Therefore, in this work, rather than focusing on the classification outcome of osteoporosis, we are interested in the aBMD value, from which the risk of osteoporosis is known based on previous studies (World Health Organization, 1994).

One recurring problem with constructing diagnosis systems for medical applications is the lack of training data (Serrano, Tomeckova, & Zvarova, 2006), which occurs because of the cost of acquiring patient data and the low prevalence rates of diseases (Chan, Sahiner, & Hadjiiski, 2004; Mazurowski et al., 2008). In general, the lack of training data results in an undersampling of the problem space which tends to lead to poor classification performance (Brumen et al., 2007; Raudys & Jain, 1991). The problem is further compounded by the imbalanced nature of the class samples; typically the number of positive class instances (diseased cases) is much less than the number of negative class instances (normal cases) (Gu, Cai, Zhu, & Huang, 2008; Mazurowski, Habas, Tourassi, & Zurada, 2007). Lastly, clinical data may have missing or incomplete features. These problems impair the performance of machine learning methods, but some ensemble techniques have been found to be robust to high dimensionality (Moon et al., 2007), high class imbalance (Lo et al., 2008), or missing features (Nanni et al., 2011). Ensemble methods are also known to improve the accuracy over single learners, and have been previously studied for use in medical diagnosis (Antal et al., 2010). Ensemble methods work by combining the contributions of several weak component learners, which reduces the variance of errors.

3. Ensemble regression methods

Ensemble methods can be applied to regression problems to obtain better robustness and accuracy. In this section, we describe the bootstrap aggregating method before introducing a feature-wise modification which is more helpful for datasets with high relative dimensionality. Building upon the bootstrap aggregating approach, the use of metalearners for improving ensemble performance is discussed. We review two basic metalearner ensembling schemes before presenting our correlation-based filtering technique for metalearner ensembling. The new technique is designed to form ensembles that are both diverse and robust.

Let the DXA-derived aBMD values be denoted as the target variable matrix Y. The data matrix X is then obtained by feature-wise concatenation of the dCT-derived aBMD values, the dCT-derived HU features, and the additional multimodal features from blood and physical measurements. The regression problem is defined as regressing the target Y based on the data X such that unknown future samples can be predicted.

3.1. Bootstrap aggregating

Bootstrap aggregating (Breiman, 1996), also known as bagging, may be capable of overcoming the high dimensionality of the data relative to the number of training samples. Bagging can improve classification/regression accuracy and stability, and any learning model may be used with bagging. In this work, several linear regression models are bagged to form a regression ensemble.

In bagging, the ensemble is composed of several component classifiers, each of which is trained on a different subset of the training data, and the ensemble decision is obtained by taking an average of the individual ensemble regressors. The subsets are randomly drawn with resampling from the training set, and the subsets are traditionally drawn in a case-wise fashion. In *case-wise bagging*, each ensemble component is trained on a different resampled training set. The resampled training sets are formed by randomly drawing training cases (with resampling). To reduce large instabilities in the regression and to better constrain the regression, the resampled training sets are resampled to contain more cases than there are features. For an input training set consisting of n data and target pairs $\{x_i, y_i\}$, where $i = 1:n$, the case-wise bagging algorithm for a k-component ensemble with a case oversampling factor of s_c is given by:

Algorithm 1. Case-wise bagging

procedure CB_TRAIN(X,Y,k,s_c)
  for $j = 1:k$ **do**
    $S_j \leftarrow \{\emptyset\}$
    while $numel(S_j) < (s_c \times n)$ **do**
      $randNo \leftarrow rand(1:n)$
      $S_{temp} \leftarrow \{x_{randNo}, y_{randNo}\}$
      $S_j \leftarrow \{S_j, S_{temp}\}$
    end while
    $R_j(x) = LR(S_j)$ ▷ Linear regression of S_j
  end for