Neurocomputing 147 (2015) 387–394

Contents lists available at ScienceDirect

Neurocomputing

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

Combining visual and textual features for medical image modality classification with ℓ_p-norm multiple kernel learning

Hong Wu*, Lele He

School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, PR China

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form
26 April 2014
Accepted 22 June 2014
Communicated by Qingshan Liu
Available online 28 June 2014

Keywords:
Modality classification
Medical image
Multiple kernel learning
Feature combination

ABSTRACT

Automatic modality classification of medical images is an important tool for medical image retrieval. In this paper, we combine visual and textual information for modality classification. The visual features used are SIFT feature, LBP feature, Gabor texture feature and Tamura texture feature. And the textual feature is a tf–idf feature vector drawn from image description text. We combine these features by ℓ_p-norm multiple kernel learning (ℓ_p-norm MKL), and use One-vs-All approach for this multi-class problem. ℓ_p-norm MKL is explored with different norm value ($p \geq 1$). These MKL based methods are compared with several other feature combination methods and evaluated on the dataset of modality classification task in ImageCLEFmed 2010. The experimental results indicate that multiple kernel learning is a promising approach to combine visual and textual features for modality classification, and outperforms other simple kernel combination methods and the traditional early fusion method.

1. Introduction

In the last decades, medical imaging has played a more and more important role in diagnosis and medical education, and the amount of medical images has grown increasingly rapidly. The huge amount of data is stored in PACSs (Picture Archive and Communication Systems), patient health records, or on the web as the content of online documents, and has led to an increasing demand for effective and efficient medical image retrieval technology. From 2004, ImageCLEF, a well-known multi-modal image retrieval campaign, has begun to provide a medical retrieval task (ImageCLEFmed) to evaluate medical image retrieval techniques.

Text-based medical image retrieval methods are well-studied and widely used, but it has many limitations. Firstly, manually annotating image is tedious, subjective and error prone. Secondly, annotations which are automatically extracted from the surrounding text are always noisy. Furthermore, there is always far more information in an image than can be described with words. Advances in computer vision have led to content-based image retrieval (CBIR) [1], which retrieves images by their visual features. CBIR research has made some success in very constrained medical domains, such as pathology, head MRIs, lung CTs, and mammograms [2]. However, current CBIR systems have poor performances when applied to a diverse collection of medical images [3,4].

It has been cognised that some attributes of medical images can be used as filters to improve medical image retrieval. Such attributes include imaging modality, anatomical region, view angle and pathological finding. Some of these attributes may have been associated with images in the form of DICOM headers, but DICOM headers have proven to contain errors [5]. In addition, this kind of information is often lost when images are published online, since most online images use compressed image file like JPEG. Fortunately, some of these attributes can be inferred based on visual appearance. In 2005, an automatic medical image annotation subtask was added to ImageCLEFmed, but it only focused on X-ray images.

Among the aforementioned attributes, imaging modality is the most fundamental one upon which other attributes, like anatomical region and pathology, are highly dependent. Previous studies have also indicated that imaging modality is an important attribute by which clinicians would like to limit the search results [6,7]. Popular radiology image search engines such as ARRS GoldMiner and Yottalook have allowed search results to be filtered by imaging modality. However, the modality information, which are usually extracted from the text associated with images, are sometimes incorrect or absent. On the other hand, imaging modality is most easily identified using visual features among the aforementioned attributes. Therefore, it is promising to combine visual and textual information to identify image modality.

* Corresponding author.
E-mail address: hwu@uestc.edu.cn (H. Wu).

http://dx.doi.org/10.1016/j.neucom.2014.06.046
0925-2312/

Automatic modality classification was first studied in [8], where the authors proposed the so-called required and frequently occurring visual features for each modality and used them with an image categorization metric for classification. Florea et al. [9] compared the efficiency of two different systems, MedIC, a medical image categorization system, and MedGIFT, a content-based medical image retrieval system, on classifying an image database of six standard modalities in radiology and nuclear medicine. Authors in [10] used color/intensity histograms and gray level correlation matrices based texture features to train neural network classifiers for modality classification of color and gray-scale medical images. In [11], the authors used textons and patch-based descriptors as visual features and a Naive Bayes Nearest Neighbor classifier for modality classification.

From 2010, ImageCLEFmed [12] has organized a modality classification sub-task to foster the research of medical image modality classification. The participants used visual and/or textual features for modality classification. Various color, texture, edge, and shape features as well as local features were explored, and the classifiers ranged from kNN to Ada-Boost, MLP and SVM. These different features were combined with early fusion or late fusion methods such as product, sum, maximum and averaging rules. Since then, many research works have combined visual and textual features for modality classification, and used the Image-CLEFmed test collection for evaluation. In Han and Chen's work [13], images were represented by SIFT feature and some global features, such as histograms of edge, gray/color intensity and their block-based variations, and image caption was represented by a binary histogram of some predefined vocabulary words. Different features were combined by averaging normalized kernels for SVM classification. Csurka et al. [14] used orientation Histograms (ORH) and local color means and standard deviations as visual features, and compared BOV (bag of visual words) and FV (Fisher Vector) image representations. Both visual and textual features were classified by Logistic regression classifiers, and all features were combined by averaging their classification scores. Lu et al. [15] proposed very light image features, such as image entropy and image energy, and combined visual and textual features with early fusion in SVM. Sun et al. [16] explored a large variety of image descriptors to learn ensemble SVM classifiers for individual modality and fused them based on confidence indicator measurements. For the likely misclassified samples, image-based classifiers and text-based classifier were combined based on a matrix of closeness and a set of additional fusion rules. In [17], a simple voting and a weighted voting ensemble classifiers were tested to combine the 17 classification results of the participants of Image-CLEF 2010, and both yielded superior results. Recently, multiple kernel learning which learns an optimized convex combination of kernels provides a new approach for feature combination. In this paper, we use ℓ_p-norm MKL ($p \geq 1$) to combine visual and textual features for modality classification.

The rest of the paper is organized as follows: Section 2 describes the related works like kernel methods, SVM and a brief introduction to feature combination methods. Section 3 gives the details on multiple kernel learning and its application to modality classification. Section 4 introduces the data collection and the features used in this study. Section 5 reports the experiment setup, the results obtained and the analysis on them. Conclusions are drawn in the last section.

2. Kernel method and feature combination

2.1. Kernel method

Although linear algorithms are efficient and have been well studied and developed, real world data analysis problems are often nonlinear. To deal with nonlinear problem, kernel methods are introduced to embed data from the input space $\mathbb{R}^d$ into a higher (possibly infinite) dimensional Hilbert space $\mathcal{H}$, or feature space, and the mapping function is denoted as $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$. The linear algorithms will benefit from this mapping because the transformed samples are more likely to be linearly separable in the higher dimensional feature space. But the computational load would dramatically increase because of the higher dimensional samples. Fortunately, such a computation can be avoided by using the kernel trick. The kernel trick [18] is to develop learning algorithms where $\Phi(\mathbf{x})$ is only used in the form of dot products, and choose the mapping such that these high-dimensional dot products $\langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ can be computed within the original input space by a kernel function $k(\mathbf{x}, \mathbf{x}')$. In this way, we do not need to know the mapping $\Phi(\mathbf{x})$ but only the kernel function. Consequently, the important thing is to find kernel functions that correspond to an inner product in some feature space. Mercer's theorem [19] states that valid kernel functions must be symmetric, continuous, and positive semi-definite. For example, RBF, polynomial and sigmoid kernels are widely used kernels. This kind of kernels is called Mercer's kernel which also has some good properties related to our study. Assuming that k_1 and k_2 are two Mercer's kernels on $\mathbb{R}^d \times \mathbb{R}^d$, and constant $\mu > 0$, the following new kernels are also valid:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mu k_1(\mathbf{x}_i, \mathbf{x}_j) \tag{1}$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j) + k_2(\mathbf{x}_i, \mathbf{x}_j) \tag{2}$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j) k_2(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

Based on these properties, new kernels can be designed by summing up (weighted) or multiplying valid kernels.

2.2. Support vector machine

Support vector machine (SVM) [20] is a powerful machine learning method and has been successfully used in many applications. It achieves good generalization by performing structural risk minimization. Given N training samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i$ is a d-dimension input vector and label $y_i \subseteq \{+1, -1\}$. Support vector machine performs classification by finding the hyperplane that maximizes the margin between the two classes. The use of kernel allows the algorithm to find the optimal hyperplane in the feature space induced by the mapping function $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$. The resulting linear discriminant function in the feature space is

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b \tag{4}$$

The classifier can be trained by solving the following quadratic optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \boldsymbol{\xi}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \forall i, \quad y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \ 0 \leq i \leq N \end{aligned} \tag{5}$$

where $\mathbf{w}$ is the normal vector to the hyperplane, ξ_i is the slack variable which measures the degree of misclassification of $\mathbf{x}_i$, the cost parameter $C > 0$ controls the trade-off between model simplicity and classification error, and b is the bias term. The parameters $(\mathbf{w}, b)$ are determined by solving an equivalent dual optimization:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \\ \text{s.t.} \quad & \forall i, \quad 0 \leq \alpha_i \leq C, \quad \sum_i \alpha_i y_i = 0 \end{aligned} \tag{6}$$

The dual optimization depends only on inner products of inputs which can be alternatively computed by means of kernel function