Штучна інтелектуальна підтримка для виявлення захворювань на зображеннях з бездротової капсульної ендоскопії товстої кишки людини

Ян Цихнерский

Департамент архітектури обчислювальних систем, Гданській технічний університет, ПОЛШЦА, Гданськ, Г. Нарutowича 11/12
E-mail: Jan.Cychnerski @ eti.pg.gda.pl

У даній роботі були описані обрані алгоритми розпізнавання захворювань на зображеннях з бездротової капсульної ендоскопії і були порівняні за їх ефективністю. Алгоритми були названі іменами їх авторів: Кодоґіанніс [2] (на основі статистичних особливостей зображень), Магоулас (перший генератор, заснований також на статистичних особливостей картинки [3], а другий – на основі вейвлет перетворень [4]) і Баопулі [5] (на основі вейвлет-перетворення). У роботі був також описаний простий, тестовий генератор. Алгоритми були оцінені, зважаючи на корисність для застосування в підтримці комп'ютерної системи для діагностики травної системи. Оцінки були отримані від передового середовища тестування, яке було сформоване з використанням великої колекції фільмів на тему ендоскопії, отриманих від медичного університету в Гданську. Для класифікації зображень ендоскопії були використані нейронні мережі (мульти шар перцептрона) і SVM класифікатори (опорні вектори). Ці два найпопулярніші класифікатори, що викорис- товуються в аналогічних роботах, були описані і порівняні у своїй ефективності. Специфіка процедури тестування покликана робити спеціальні метрики. У даній роботі така метрика була розроблена (яка базується на F-відмітці), пояснена і описана (Ел1). Для досягнення кращих результатів ефективності при розпізнаванні хвороби, всі алгоритми "вхідних" параметрів (табл. 1) були порівняні і оптимізовані для розпізнавання захворювань (табл. 2). Два з алгоритмів дали хороші результати щодо розпіз- навання хворої тканини (Магоулас 2 і Кодоґіанніс – обидві з нейронними мережами як класифікаторами, див. таблицю 3), але жоден з алгоритмів не зміг би відрізняти два захворювань. Оптимальні знайдені поєднання захворювань, алгоритмів і параметрів були описані в таблиці 4.

---

Artificial intelligence support for disease detection in wireless capsule endoscopy images of human large bowel.

Jan Cychnerski

Department of Computer Architecture, Gdańsk University of Technology, POLAND, Gdańsk, G. Narutowicza 11/12
E-mail: Jan.Cychnerski@eti.pg.gda.pl

In the work the chosen algorithms of disease recognition in endoscopy images were described and compared for their efficiency. The algorithms were estimated with regard to utility for application in computer system’s support for digestive system’s diagnostics. Estimations were achieved in an advanced testing environment, which was built with use of the large collection of endoscopy movies received from Medical University in Gdańsk. For classification of the endoscopy images the neural networks and SVM classifiers were used. Efficiency of classifiers was also compared. To achieve the best scores of efficiency in disease recognition, all of the algorithms’ input parameters were optimized. In the summary scores of tests and the bests of algorithms were described.

Keywords – artificial intelligence, wireless capsule endoscopy, large bowel diagnostics, disease recognition algorithms, artificial neural networks, support vector machines

I. Introduction

Endoscopy is the one of the most important medical examinations. It make it possible to detect many dangerous diseases (e.g. cancer) as early as it is necessary to undergo the effective treatment. One of the most progressing branch in endoscopy examinations is the Wireless Capsule Endoscopy (WCE). The examination lies in swallowing by patient the capsule, which includes the camera and the source of light. After the examination the consultant receives the movie ready to analysis. Example images from WCE are presented in Fig. 1. Unfortunately, specificity of these movies causes that analyze is time-consuming (lasts about 1-2 hours with 8-hours movie). It makes the task more difficult, because the consultant has to be very focused on his work.

Fig.1. Images in Wireless Capsule Endoscopy

Because of this, there exists a demand for the system which could help in doctor’s work and cut down the analysis time. The servo system should make first analysis and point where the suspicion of degenerative disorders could occur.

Building of such system needs to render assistance from medical consultants from suitable branches of medicine. In this work scores of endoscopy examinations prepared by consultant were utilized. Collection of standardized photos (showing diseased and healthy cases) was made for both training and testing the algorithms.
II. Artificial intelligence algorithms used

In the literature it is possible to find a lot of publications about various methods of disease recognition on the endoscopy movies [1]. But the most popular algorithm type in this medical branch is classification of signatures generated from pictures [2]. Building the signature consists in converting the picture into a tuple of numbers (in process called features extraction). This process might be preceded by some initial activities called preprocessing (not included in this work). General scheme of algorithms’ operations is presented in Fig. 2.

![General scheme of image recognition algorithms](image)

Fig.2. General scheme of image recognition algorithms

In this work, there is investigated efficiency of algorithms which work according to above-mentioned scheme. Testing were carried out by means of different generators of signatures and different classifiers in special combinations. Also, very important is fact that transformations and the classifiers which were used, usually have theirs special parameters, which have major influence on theirs operations. Selection of these parameters is not trivial, therefore many combinations of their values were also tested.

A. Signature generation algorithms

As the second block of processing (called signature generator), 4 advanced transformations were based on methods proposed by authors of other publications. Moreover, one trivial algorithm was created for verification purposes. All of the algorithms are described below (named by author's surname):

1. **Kodogiannis** – signature generator based on statistical features of picture. First, the picture is divided into 6 canals: R, G, B and H, S, V. In each canal the NTu transformation [2], which imitates the neighborhood, is carried out. Then, the histogram is set up. Finally, 9 statistical features are pointed for each histogram, giving a 54 numbers long signature. In this algorithm, the only parameter is the capacity of histogram’s bin.

2. **Magoulas 1** – generator also based on statistical features. Firstly, the picture is converted into grayscale. Subsequently the co-occurrence matrix [3] is created, and then, statistical features are pointed from this matrix, which builds 16-numbers signature. In this generator there are no parameters.

3. **Magoulas 2** – generator based on wavelet transformations and statistical features. This is an extended version of generator Magoulas 1 (mentioned above). At first, the picture is converted into grayscale and subsequently treated with two-dimensional wavelet transformation. Then, the new-created picture is divided into 4 equal-sized pieces. The left-upper piece (canal LL) is rejected. Co-occurrence matrices are calculated for the rest of pieces, and finally, statistical features are pointed from matrices, which builds the signature 48-numbers long[4]. In this generator there are 2 parameters: type of wavelet used (Haar, Daubechies or CDF 9/7) and the standard of picture’s contrast after transformation.

4. **BaopuLi** – generator based on wavelet transformations and statistical features. Firstly, the picture is divided into canals: R, G, B. Secondly, for each canal wavelet transform is performed. Canal LL is rejected. Then, LBP Transformation (which imitates the neighborhood) [5], is carried out. Finally, 10-bin histogram is made for each of 9 canals, which build the signature 90 numbers long. This generator does not have parameters to test.

5. **Test** – trivial, testing generator based on color features of picture. The generator consists in pointing average pixel value for R, G, B and H, S, V canals. Purpose of this generator is to be a reference to others generators, as a representation of a low quality algorithm.

B. Classification algorithms

As a classification block, two most popular in literature classifiers were examined (their implementation was taken from OpenCV library [6]):

1. **Neural Network (Multi Layer Perceptron)** – simple perceptron network with many hidden layers and one output neuron. The activating function implemented in each neuron is sigmoid function. For training the network the RPROP algorithm [7] was used. The classifier’s parameters, which were tested, were: neural network parameters (number of layers and number of hidden neurons in each hidden layer) and training algorithm's parameters (the highest number of training iterations, epsilon, \( \alpha \) and \( \Delta_{\min} \)).

2. **Support Vector Machines Classifier** – working of this classifier lies in researching for an optimal hyperplane, which divides 2 sets of points (what ensures the biggest possible margin between set points and plane). Basic version of the classifier, called SVC, was used. Examined parameters were: type of used kernel function (linear, or Radial Base Function), parameters of kernel function (C and gamma) and training parameter – the highest number of iterations.

III. Testing procedure

To estimate the efficiency of algorithms, the special testing environment was used, which adjust algorithms’ parameters’ values, trains the classifiers and tests them with pointed data set. Because of tests’ character, to compare quality of classification, especially prepared metric was used.

A. Metric

Classification of data set gives 4 components of result: True Positive, True Negative, False Positive and False Negative. However, those numbers give a few information about efficiency of recognition. Because of that, in medical examinations two measures: sensitivity and specificity are used. This two-component metric is useful for man, but not enough for computer, because the linear order relation is not defined for it.

In document classification often the F-score, which takes recall and precision into consideration, is used. By analogy with this F-score, its equivalent for medical examinations was created:
\[ F_2 = \left(1 + \beta^2\right) \frac{\text{sensitivity} \cdot \text{specificity}}{\beta \cdot \text{specificity} + \text{sensitivity}} \]  

- created metric: [0,1], where 0 is the worst
- \( \beta \) - comparison coefficient. Means how many times sensitivity is more important than specificity

**B. Testing environment**

Because of number of algorithms, various input parameters and time-consuming calculations – it was demanded to create a dedicated testing environment, which allows to automatic adjustment of parameters and carry out the tests on real medical pictures.

The built testing environment takes 3 inputs: a set of algorithms to be tested, a set of their parameters with assigned values' ranges, and sets of medical images, on which classifiers are trained and tested with cross-validation [8].

Cross-validation was found necessary, when overfitting of the classifiers was observed in the tests. It allows for better use of training data (there is no need to waste some data for testing purposes). The main idea is to train and test the classifiers \( k \) times, every time on a different subset of input data. One defect of this method is that it has computational complexity proportional to \( k \).

In the tests, \( k=10 \) was assumed, as compromise between speed and accuracy. Data subsets was constructed in a way, that in each subset proportion of diseases was preserved, and all images from one examination belonged to the same subset.

After choosing best algorithms' parameters' values, classifiers were trained on all of the subsets together, and then, the validation test was performed on external, additional data sets.

Testing environment is a parallel application run on a computer cluster. It has master-slave architecture. Splitting on \( n \) cluster nodes speeded up the testing almost \( n \) times.

**TABLE 1**

<table>
<thead>
<tr>
<th>ELEMENT</th>
<th>PARAMETER</th>
<th>VALUE RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magoulas2</td>
<td>wavelet</td>
<td>Haar, Daubechies, CDF 9/7</td>
</tr>
<tr>
<td></td>
<td>contrast level</td>
<td>normal, high</td>
</tr>
<tr>
<td>Kodogiannis</td>
<td>bin size</td>
<td>1, 25, 50, 100</td>
</tr>
<tr>
<td>Magoulas1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BaopuLi</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Test</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SVM</th>
<th>kernel function</th>
<th>linear, RBF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iterations</td>
<td>5, 10, 50, 200</td>
</tr>
<tr>
<td></td>
<td>gamma</td>
<td>0.0001, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.1, 1.0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.01, 0.5, 1.5, 10000</td>
</tr>
<tr>
<td>Neural network</td>
<td>layers</td>
<td>3, 4, 5</td>
</tr>
<tr>
<td></td>
<td>hidden neurons</td>
<td>2, 3, 4, 5, 6, 10, 16, 24</td>
</tr>
<tr>
<td></td>
<td>max. iterations</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>epsilon</td>
<td>0.001, 0.0005, 0.0001, 0.00001</td>
</tr>
<tr>
<td></td>
<td>( \Delta )</td>
<td>0.01, 0.1, 1</td>
</tr>
<tr>
<td></td>
<td>( \Delta_{\text{min}} )</td>
<td>0</td>
</tr>
</tbody>
</table>

**IV. Obtained results**

With presented testing environment many detailed test of described algorithms were performed. The tests were meant to attain maximal efficiency in disease recognition by adjusting algorithms' parameters.

**A. Generators and classifiers tested**

In Table 1, tested algorithms, their parameters and ranges of values are described. Recognition efficiency was examined for 15 data sets. 5 of them was meant to distinguish between diseased and healthy tissue, and the other 10 to distinguish between two different diseases. Characteristics of data sets is described in Table 2.

**B. Results**

Validation tests results and cross-validation tests results (in brackets) for the best found parameters’ values are presented in Table 3.

Signature generators’ results show that two algorithms (Magoulas 2 and Kodogiannis) are very efficient at recognizing diseased tissue. Their efficiency reaches in some cases almost 100%. The worst algorithms, according to predictions, were Test and Magoulas 1, which efficiency was not much greater than random.

As predicted, data sets “disease/disease” were much more difficult for all of the algorithms than sets “disease/healthy”. None of tested algorithms was good enough to distinguish between two different diseases. The only algorithm, that in some cases gave promising results, was (again) Magoulas 2.

Classification algorithms’ comparison gave more unambiguous conclusions. Artificial neural network, in all “disease/healthy” data sets, worked much better than SVM classifier. However, in the rest of data sets, situation often reverted.

**Conclusion**

According to cross-validation tests’ results, optimal found conjunctions data set - signature generator – classifier – parameters are described in table 4.
TABLE 3

<table>
<thead>
<tr>
<th>Tests’ Results</th>
<th>ANN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>0.94 (0.97)</td>
<td>0.28 (0.77)</td>
</tr>
<tr>
<td>Colitis</td>
<td>0.94 (0.96)</td>
<td>0.19 (0.68)</td>
</tr>
<tr>
<td>Crohn</td>
<td>0.96 (0.91)</td>
<td>0.72 (0.60)</td>
</tr>
<tr>
<td>Polyph</td>
<td>0.87 (0.98)</td>
<td>0.52 (0.72)</td>
</tr>
<tr>
<td>Ulcer</td>
<td>0.88 (0.96)</td>
<td>0.19 (0.66)</td>
</tr>
</tbody>
</table>

(Cancer - both with neural networks - gave good results at recognizing diseased tissue, but not at distinguishing between diseases. Very high sensitivity and specificity (reaching 100% in some cases) signals possibility to use in medicine as an element of diagnostics support system.

Two of tested algorithms (Magoulas 2 and Kodogiannis) – both with neural networks – gave good results at recognizing diseased tissue, but not at distinguishing between diseases. Very high sensitivity and specificity (reaching 100% in some cases) signals possibility to use in medicine as an element of diagnostics support system.

References


(special acknowledgements to Miss Karolina Kaczyńska)