
Utilizing automatic recognition and classification of images for pattern recognition

Mohammad Hadi Yousofi^{1,*}, Habib Yousofi², Sayyed Amir Mohammad Razavi³

¹Department of Mechatronics, Postgraduate School, Islamic Azad University of Kashan, Kashan, Iran

²School of Medicine, Kashan University of Medical Sciences, Kashan, Iran

³Department of Electrical and Computer, Islamic Azad University of Kashan, Kashan, Iran

Email address:

Mhu320@yahoo.com (M. H. Yousofi), Ha_usofi@yahoo.com (H. Yousofi), Mohammadrazavi@yahoo.com (S. A. M. Razavi)

To cite this article:

Mohammad Hadi Yousofi, Habib Yousofi, Sayyed Amir Mohammad Razavi. Utilizing Automatic Recognition and Classification of Images for Pattern Recognition. *International Journal of Intelligent Information Systems*. Special Issue: Research and Practices in Information Systems and Technologies in Developing Countries. Vol. 3, No. 6-1, 2014, pp. 80-83. doi: 10.11648/j.ijjis.s.2014030601.25

Abstract: Pattern recognition is a scientific approach for categorizing objects to class or subject numbers. These subjects need to be classified based on their applications (can be image, signal or any other type of measurements). Occupation, automation, military information, communication, industry and commercial applications and many other fields can benefit from Pattern recognition approaches. Perhaps, one the most important reasons that lead to pattern recognition prominent place in today's research studies, is the role of image auto-classifications. In this research, we investigate recent literatures about image auto-classification and image processing to identify patterns.

Keywords: Pattern Recognition, Images Auto-Classification, Image Processing, Support Vector Machine

1. Introduction

The simplicity that we recognize a portrait, understand words, read manuscripts, identify car keys with hands in our pocket leads to less attention to operation complexity which is done in these process as pattern recognition. Pattern recognition is a scientific method that aims to categorize objects to class or subject numbers. These subjects need to be classified depending on their applications (can be image, signal or any other type of measurements). Like other fields, progress in computer science led to increasing need of pattern recognition practical applications and as a result led to spread projects in this field. Atomization of works and role of pattern recognition in the research is one of the most important reasons that cause pattern recognition to have a prominent place in today's researches.

We can say that pattern recognition is receiving raw data and deciding based on data grouping. More research studies about pattern recognition are related to "supervised learning" or "non – supervised learning", which we explain in the next sections. Pattern recognition methods separate special patterns from data collection via previous knowledge about patterns or statistical information. Those patterns which are classified in this way are groups of measurements or

observations that form determined areas in a multidimensional space. This characterization is the main difference of pattern recognition with pattern matching, where patterns are identified base on a determined pattern and accurate real cases. Pattern recognition and matching pattern are main sections of image process discussion especially in machine vision.

Nowadays, there are more attentions to improved information systems because information is a main element in decision-making and the amount of information in the world is increasing in different forms with different complexity. One of the main issues in designing modern information systems is automated pattern recognition. Recognition is a leading characteristic of being human. A pattern is a description of an object [1]. Humans have high level informational system that is due to developed pattern recognition ability. According to nature of recognized patterns, recognition operation is divided into two main species.

2. Materials and Methods

2.1. Supervised and Un-Supervised Pattern Recognition

In most cases, representative pattern from every class are

available. In these situations, we can apply supervised pattern recognition, and the basis of this approach is a collection of identified practical patterns to classify and perform a suitable learning method. In some applications, maybe there is a collection of un-identified practical patterns to category. In these situations, we can apply un-supervised portrait recognition techniques. As mentioned before, supervised pattern recognition introduced by this matter are when correct group of every practical pattern is clear [2]. In un-supervised cases, however, we deal with learning matters in presented pattern classes. This issue is known as an un-supervised learning approach [3].

2.2. Classification Algorithms (Algorithms with a Predictor Supervisor)

The selected algorithm for recognizing patterns depend on the type of output, with or without supervision, and the static or dynamic nature of algorithms. Static algorithms are divided to two difference categories: differentiate and generative. In the following, we investigate different algorithms for recognizing patterns.

2.2.1. Decision Tree

Decision tree is a tool to support decision and employs various trees to make a model. Decision tree is usually used in research in operation. Especially it is used to determine strategy with a high probability to reach an aim in decision analysis. Description of conditional probability computations are the other applied area of decision trees [4].

Supervised decision tree, in usual statics method, is used more in data mining [5]. In decision analysis, a decision tree is used as a tool to draw and analyze decisions where the predicted amounts are computed from competitions frequently [6].

One decision tree has three types of groups:

- 1 Decision group; usually shown by square.
- 2 Random group; shown by circle.
- 3 Final group; shown with triangle.

The advantages of decision trees are as follows:

- 1 Simple understanding: every human can learn the work style of decision tree with little study and practice.
- 2 works with big and complex data: in spite of simplicity, decision tree can work with huge data easily and make a decision base on them.
- 3 Simple re-use: when a decision tree is made for one problem, we can compute different samples of that problem with the same tree.
- 4 Combination with other methods: we can combine the results of decision tree with other decision techniques and gain better results.

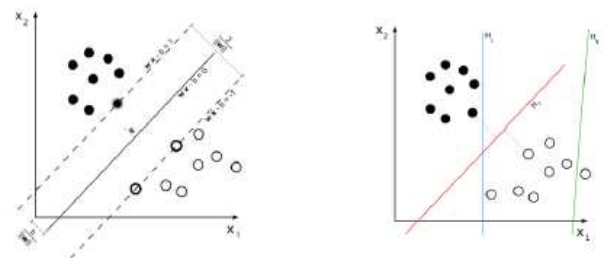
The disadvantage of decision tree is that it becomes big exponentially.

2.2.1.1. Support Vector Machine

Support Vector Machine (SVM) is one the relatively modern methods that has showed high efficiently in recent years compared to older methods for classification including

perceptron nervous systems. Work basis of SVM classification is data linearly classification and in linear division, we try to select a line that has more reliable edge. To solve this equation, we should find optimum line for data by quadratic programming QP methods that are determined methods to solve limited problems. Before line division, we take data to higher dimensions via phi function to classify data with higher complexity in machine. We apply Lagrange Duality Theorem to transform minimization problem to duality form where we use simple function called core function that is vector coefficient of Phi function to solve a problem with very high dimensions. We can use different core functions including exponential cores, vector and Sigmund.

Therefore, we define learning data collection 'D' including 'n' member as follow:



$$D = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^n$$

Figure 1. Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

Where y is equal to 1 or -1 and every X_i is equal to real vector of $-p$ dimensional. The aim is to find separator hyper plane with a most distance from marginal areas that separate areas with $y_i=1$ from $y_i=-1$. Every cloud page can be a set of points that satisfies the following condition wrote:

$$\mathbf{W} \cdot \mathbf{X} - b = 0$$

Where (\cdot) is the multiply symbol. \mathbf{W} is a normal vector that is vertical to hyper plane. We want to select \mathbf{W} and b in a way that provide the most distance among parallel hyper plane which separate data from each other. These hyper planes are defined as follow (Equation 1):

$$\mathbf{w} \cdot \mathbf{x} - b = 1 \quad (1)$$

And

$$\mathbf{w} \cdot \mathbf{x} - b = -1. \quad (2)$$

If the separative supervised data \mathbf{W} is linear, then we can have two hyper planes in point's edge so that they do not have any common point and try to maximize their distance. Given the geometry, the distance of these two plane is $\frac{2}{\|\mathbf{w}\|}$. So we should minimize $\|\mathbf{w}\|$. In order to avoid entering the margin, we add the following conditions. For every 'i':

$$\begin{aligned}
& \text{of the first class} \\
& \mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \quad \text{for } \mathbf{x}_i \\
& \text{of the second class} \\
& \mathbf{w} \cdot \mathbf{x}_i - b \leq -1 \quad \text{for } \mathbf{x}_i
\end{aligned} \tag{3}$$

It can be written as follows:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad \text{for all } 1 \leq i \leq n. \tag{4}$$

2.2.1.2. Nonlinear Support Vector Machine

The optimum separator was first presented by Vapnik in 1963 that was a linear category. In 1992, Bernhard, Boser, Guyon, Isabelle and Vapnik suggested a way to provide non-linear grouper via placing core to find hyper plane with more edge [7, 8]. The achieved algorithm look same, except all of points multiplies was placed with a nonlinear core function. This led the algorithm to be suitable for hyper plane with the most edge in a transformed space [9].

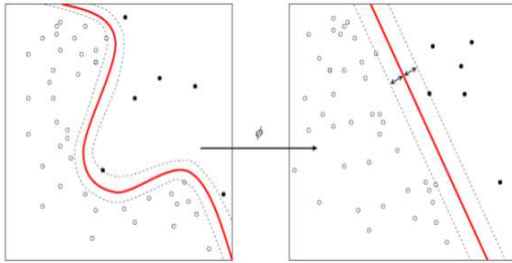


Figure 2. Kernel machine

Perhaps if the transform is nonlinear, the transformed spaces have a higher dimension. However, the grouper is a hyper plane in characterize space with high dimensions that it may be nonlinear in input space. If we use core with Gusin function, corresponding property space is a infinite Hilbert space. The rouper of the most edge, is in good – arrange, so infinite dimensions do not destroy the result. The common cores are as follow:

- The polynomial (homogeneous):

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d \tag{5}$$

- The polynomial (inhomogeneous):

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d \tag{6}$$

- Gusin:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \text{ for } \gamma > 0. \text{ Some } \tag{7}$$

Times parametrized using $\gamma = 1/2\sigma^2$

- Hyperbolic tangent

$$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa \mathbf{x}_i \cdot \mathbf{x}_j + c), \text{ for some (not every) } \kappa > 0 \text{ and } c < 0 \tag{8}$$

2.2.1.3. Multiclass Support Vector Machine

Basically, SVM is a binary separator. Therefore, we explain theory basis of SVM for grouping of two classes. One multi-class pattern recognition can be reached via multi-classes SVM combination. Usually, there are two

visions for this aim. One is the strategy of "one against all" for classifying every pair and remained classes. The other strategy is "one against one" to classify every pair. When first classification leads to vague classification, the general approach for multi-classes is minimized to multi-classes problems and several binaural problems. Every problem is solved with one binary separator. Then, SVM binary separator output combines together and in this way the multiclass problem can be solved [10, 11].

The advantages of this approach are:

- 1 The supervision is relatively simple.
- 2 Unlike the neural network, it does not get stuck in a local maximum.
- 3 For high dimensional data, it has relatively good answer.
- 4 Reconciliation between grouper complexity and error amount can be controlled clearly.

The disadvantage is that we need to Select a good kernel function parameter C.

2.2.2. Perceptron

Perceptron is a binary separator that corresponds X self input (a vector from real numbers) to output amount $f(x)$, (one scalar from real numbers) that compute as follow:

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Where w is a vector from weights with real amounts and \cdot is a point coefficient (that compute the total weight). "b" is a bias ; a fix sentence that do not relate to input.

$f(x)$ sign to classify x to a positive or negative sample and it is used in binary classification problems. Bias can be considered as an action function adjustor or every amount for basis for action of output neuron "b" is negative, inputs weighted combination should be a positive amount higher than, to place a grouper neuron in a condition higher than zero threshold .^[12]

3. Discussion and Conclusion

In this article, we have considered different group of pattern recognition algorithms. Considering the advantages and disadvantages of the discussed, we cannot rely to specific algorithm permanently. The decision about which algorithm with less computation, less time and best results could be different based on the situation and applications.

Based on our analysis from different methods, we should try to use combination system with a combination of pattern recognition algorithms to reach minimum computation and less time, as possible.

References

- [1] Milewski, Robert; Govindaraju, Venu (31 March 2008). "Binarization and cleanup of handwritten text from carbon copy medical form images". *Pattern Recognition* 41 (4): 1308–1315. doi:10.1016/j.patcog.2007.08.018.

- [2] Richard O. Duda, Peter E. Hart, David G. Stork (2001) Pattern classification (2nd edition), Wiley, New York, ISBN 0-471-05669-3
- [3] R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice, Wiley, ISBN 978-0-470-51706-2, 2009.
- [4] Quinlan, J. R., (1986). Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers
- [5] Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.
- [6] Barros R. C., Cerri R., Jaskowiak P. A., Carvalho, A. C. P. L. F., A bottom-up oblique decision tree induction algorithm. Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011).Conference Location :Cordoba. DOI:10.1109/ISDA.2011.6121697
- [7] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995.
- [8] Aizerman, Mark A.; Braverman, Emmanuel M.; and Rozonoer, Lev I. (1964). "Theoretical foundations of the potential function method in pattern recognition learning". Automation and Remote Control25: 821–837.
- [9] Boser, Bernhard E.; Guyon, Isabelle M.; and Vapnik, Vladimir N.; A training algorithm for optimal margin classifiers. In Haussler, David (editor); 5th Annual ACM Workshop on COLT, pages 144–152, Pittsburgh, PA, 1992. ACM Press. doi:10.1145/130385.130401. ISBN 089791497X.
- [10] Duan, Kai-Bo; and Keerthi, S. Sathya (2005). "Which Is the Best Multiclass SVM Method? An Empirical Study". Proceedings of the Sixth International Workshop on Multiple Classifier Systems. Lecture Notes in Computer Science 3541: 278. doi:10.1007/11494683_28. ISBN 978-3-540-26306-7.
- [11] Hsu, Chih-Wei; and Lin, Chih-Jen (2002). "A Comparison of Methods for Multiclass Support Vector Machines". Neural Networks, IEEE Transactions on (Volume:13 , Issue: 2). DOI:10.1109/72.991427.
- [12] Liou, D.-R.; Liou, J.-W.; Liou, C.-Y. (2013). "Learning Behaviors of Perceptron". ISBN 978-1-477554-73-9. iConcept Press.