

Automated Classification of Fish Using Machine Learning and Pattern Recognition

Anis Sefidanoski ¹, Festim Halili ², Biljana R. Velkovska ³, Zorica Kaević ⁴

¹ UIST "St. Paul the Apostle", Partizanska bb, 6000 Ohrid, N. Macedonia

² State University in Tetovo, Ilindenska bb, 1200 Tetovo, N. Macedonia

³ SUGS "Orce Nikolov", Blvd Ilinden 101, 1000 Skopje, N. Macedonia

⁴ OUU "Aleksandar Makedonski", Ul. V. S. Bato bb, 1000 Skopje, N. Macedonia

Abstract— Automation in classification is a ubiquitous process in every modern industry. This process is especially critical in the seafood industry of raw fish, where nature produces magnitude of complex forms and patterns. This paper is presenting the various methods of pattern recognition to the fish classification problem, discusses the processing of images, decision-making methods for classification of patterns (Nearest Neighbor Algorithm and Bayesian decision theory), optimization of the classifiers for pattern recognition, ways of analysis of patterns by groups, and the basic concepts of K-means clustering algorithm.

Keywords -automation, image processing, pattern recognition, algorithms, artificial intelligence, machine learning.

1. Introduction

In our everyday life, we process and analyze a large amount of information of various types, with different meanings and qualities. Decisions are made on the basis of the analysis of that information. We receive information through the senses. More than 95% of them have an optical nature. [3]

DOI: 10.18421/SAR52-07

<https://doi.org/10.18421/SAR52-07>

Corresponding author: Anis Sefidanoski,
UIST "St. Paul the Apostle", Partizanska bb,
6000 Ohrid, N. Macedonia.

Email: anis.sefidanoski@uist.edu.mk

Received: 03 April 2022.

Revised: 08 June 2022.

Accepted: 14 June 2022.

Published: 27 June 2022.

 © 2022 Anis Sefidanoski et al; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

The article is published with Open Access at www.sarjournal.com

Images are a powerful medium for conveying information and a means of communication. Therefore, they are not only the primary source of information but also the basic way of communication between people, especially between humans and machines (e.g. Basic communication between a computer and a human is feasible through a monitor. [1]. The monitor serves as a means of transmitting an image or information, about what happens in the computer, to the person. On the basis of the information received, the person reacts appropriately. This images/information has a digital format). [2] [4] The digital image contains a large amount of information. This means that during the process of its analysis a large number of different information of different formats is processed. It is necessary to emphasize that the information received by other senses is transformed into pictorial information, and only then is it used to recognize patterns. In this regard, special methods for image analysis have been developed. The term "analysis" refers to complex image processing, which consists of two phases. In the first phase, the image is preprocessed; segmented and potential objects of interest are marked in it. In the second stage, the objects in the image are recognized, which means that they are mathematically described and classified as objects of a certain class, which is a segment of a set of predefined classes. [5]

2. The Process of Pattern Recognition

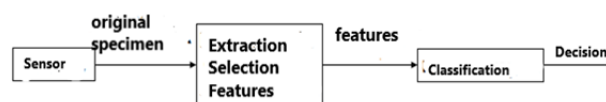


Figure 1. Block diagram of the sample recognition process

The Figure 1 shows a block diagram of the sample recognition process:

The sensor presents input sample, which should be classified (eg. face image, speech, RKG signal, set of

medical symptoms, and etc). The sample is usually used in practice as a vector.

The feature extraction and selection block define the smallest number of required features relevant for classification. [15]

The qualification block, based on the extracted characteristics, decides which class the input sample belongs to.

To exemplify the above-stated, qualification (sorting) of fish by species can be used.

The canned fish factory automatically sorts the fish on a conveyor belt by type (eg. salmon or European bass, also known as “branzino”). As shown on Figure 2, the camera captures the fish on the conveyor, and the software needs to project a classifier that will recognize whether it is salmon or European bass.

Question: On what characteristics of fish can we distinguish salmon from European bass?

Salmon and European bass differ based on their physical characteristics: length, light, width, number and shape of fins, the position of the mouth, etc.

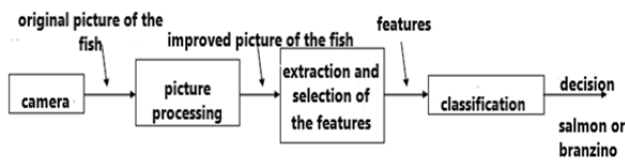


Figure 2. Block diagram of the fish qualification process

Length: A typical European bass is known to be longer than salmon. This means that we can classify fish according to this characteristic. The classifier classifies longer fish of a certain limit value as European bass and shorter fish like salmon.

We use a set of fish to train the classifier, giving a class distribution (salmon, European bass) as in the Figure 3.

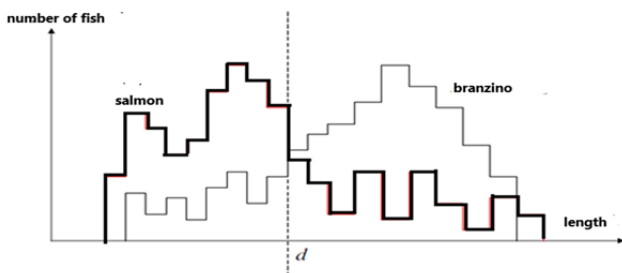


Figure 3. Distribution of fish species by length only

The distribution diagram shows that the two classes of fish overlap. Obviously, no matter which boundary we choose between classes, there will be a relatively large classification error. Thus, to reduce this error we will make a classification with some other characteristics such as the lightness value of the color (salmon is darker) [13], [14].

The lightness of the color is a much better feature because the distributions in the two classes of fish are much more strongly separated. However, there is a chance that we will make a mistake regardless of the color value limit we choose. [6]

There are two types of mistakes; we can put European bass in the salmon can and vice versa. Each of these mistakes has its price tag for the company's reputation:

- If we put European bass in the can of salmon, the buyers will be dissatisfied because the salmon is more expensive, they will hardly continue to buy our products. It is a big blow to the company's reputation.
- If we put salmon in the canned European bass, the buyers will not be dissatisfied, but the company will be at a loss. However, the mistake will have a lower price.

This suggests that to reduce customer dissatisfaction we can reduce the lightness limit value by reducing the error of selling European bass as salmon.

All things considered, every decision has its price, and the purpose of decision-making methods is to make a decision that will have the lowest cost. [7], [8]

It should be noted that to improve the accuracy of the classification we can try to use both characteristics: length and lightness of the fish. The vector of characteristics, in that case, is determined by the relation:

$$\mathbf{x} = [x_1 \ x_2]^T$$

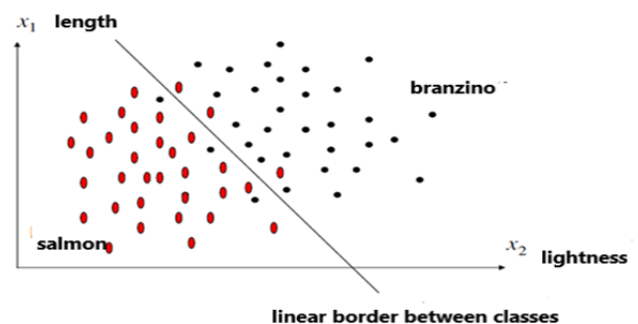


Figure 4. Linear classifier diagram

The last example displays a diagram of a linear classifier, because the space of the distribution of features for the two classes is divided by a linear function. In other words, a diagram of a linear classifier is given hence the space of the distribution of features for the two classes are divided by a linear function [9].

Principally, we can also use a nonlinear classifier (Figure 4). In that case, the boundary between the classes will be a nonlinear function.

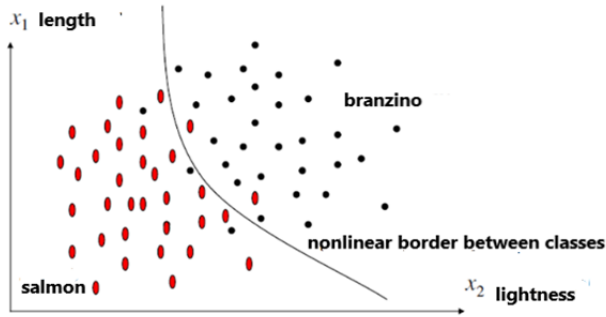


Figure 5. The nonlinear classifier diagram

The Figure 5 presents a diagram of a nonlinear classifier, because the space of the distribution of features for the two classes is divided by a nonlinear function.

3. The Nearest Neighbor Algorithm

There are two basic methods of deciding pattern classification: Nearest neighbor algorithm; Bayes' theory of decision-making. [11]

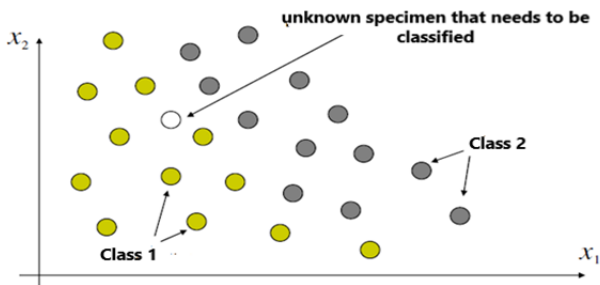


Figure 6. Nearest neighbor algorithm-NN

During the nearest neighbor algorithm (NN rule) the unknown sample is compared to known samples (training samples). The distance from the unknown sample to all training samples (all neighbors) is measured. The training sample that is closest to the unknown sample in the feature space is found. The unknown sample is classified in the class to which the nearest neighbor belongs (Figure 6).

When applying this algorithm, problems arise and they need to be solved additionally.

A sample that needs to be classified has two closest neighbors belonging to different classes. Therefore, it is not possible to determine which class it belongs to.

Two samples belonging to different classes have the same closest neighbor. The consequence is that it is not possible to determine which class they belong to [12].

In both cases, it is necessary to choose another combination of features in which the closest neighbors of the two unknown samples will not belong to the same class.

Image noise can cause changes in feature vectors, and thus classification errors. This is excluded by improving the generalization property with the help of a larger number of training samples. [10]

The classification by the nearest neighbor method can be interpreted as follows:

- In the learning phase we use a larger number of training samples that belong to different classes.
- We define the density of training samples $D_i(x)$ as the number of samples belonging to class C_i in a constant environment at place x in the feature space.

If $D_k(x)$ is large for a class C_k in place of x , it means that there are a large number of training samples belonging to the class C_k , so it is likely that the unknown sample x belongs to the class C_k

According to the NN rule, it is worth that

$$D_k(x) = \max_i D_i(x)$$

And then the vector x belongs to the class C_k .

The density of the training vector $D_i(x)$ is related to the conditional probability $p(x|C_i)$

$$p(x|C_i) = \frac{D_i(x)}{\int D_i(x) dx}$$

number of vectors that belong to class i on place x
 total number of vectors that belong to class i

$p(x|C_i)$ is normalized density $D_i(x)$.

We can conclude that the NN rule reaches a decision based on conditional probabilities $p(x|C_i)$ (likelihood).

The NN rule does not make a MAP classification, but a Maximum Likelihood (ML) classification.

The NN rule does not take into account a priori probabilities which in some cases make a biased classification (bias).

The NN rule can be modified in a way to make a MAP classification.

The NN rule can be corrected directly by multiplying the density of training samples $D_i(x)$ by $P(C_i)$ in order to find the highest density class before making a comparison.

The NN rule can be corrected as well, indirectly by the fact that it is necessary to select the sets of training vectors of the individual classes so that their magnitudes are proportional to the a priori probabilities of those classes.

The following conclusions can be drawn about the NN rule:

- The NN rule does not take into account the a priori probabilities of the classes and can therefore produce a biased classification.
- It is wrong to train an NN classifier using the same number of samples for each class.
- Proper training of an NN classifier is when the structure of the training set is statistically equal to the vectors that appear in normal operation.

4. Bayesian Theory of Decision Making

Bayesian theory of decision making is a fundamental statistical tool for solving the problem of sample recognition. This approach quantifies the compromises between different decisions in terms of price and probability. [16]

Problem: The unknown example $x \in$ needs to be classified into one of N classes C_i , $i = 1, \dots, N$

We assume that we know the conditional probability of belonging to $p(x|C_i)$, that is the a priori probability of the class $p(C_i)$.

Posterior probability (probability after easurement) of class C_i if x is measured according to the Bayes formula is equal to:

$$P(C_i | \mathbf{x}) = \frac{p(\mathbf{x} | C_i)P(C_i)}{p(\mathbf{x})}$$

where:

$$p(\mathbf{x}) = \sum_j p(\mathbf{x} | C_j)P(C_j)$$

In Bayesian decision rules, the unknown sample x is classified in class C_j if it has the highest posterior probability:

$$P(C_j | \mathbf{x}) = \max_i P(C_i | \mathbf{x})$$

This method of classification is called MAP (maximum a posteriori) classification.

5. Optimal Number of Features

One of the most important issues when designing a classifier is choosing the dimensions of the feature vector. It is intuitively clear that by increasing the number of features we can reduce classification errors. [18], [19] Nonetheless, increasing the number of features improves the quality of the classification to a certain extent. After that limit, the quality of the classification decreases. The optimal number of features depends on the size of the training set. The

larger the set of training vectors, the more characteristics we can have (larger dimension of the vectors). The optimal number of features needs to be determined experimentally. [20] If the vector dimension is too large in relation to the number of training vectors, then we do not get a good classification because we have too small a training set. If the size of the vector is too large in relation to the number of training vectors, then it is unlikely that an unknown vector will have a training vector of its class nearby, because the feature space will be overcrowded, so there will be vectors of another class between the same class vectors. [21]

6. Cluster Analysis

Sample group analysis algorithms are used in sample recognition processes to detect areas in the feature space where objects are grouped. Objects organized in groups belong to one class. Clustering algorithms actually perform a kind of classification of vectors [17]. There are two types of classifiers that use clustering algorithms:

- Supervised learning classifiers;
- Unsupervised learning classifiers.

7. K-means Clustering Algorithm

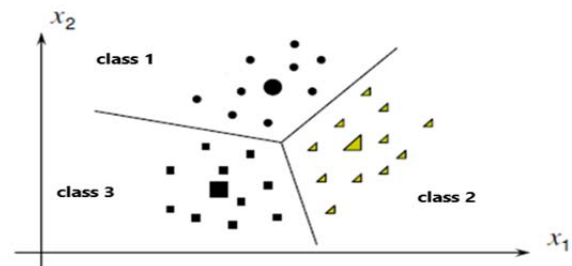


Figure 7. Linear classifier diagram

There are several steps in the developing process of the K-means clustering algorithm, which is illustrated in Figure 7:

- Select the initial centers of the group;
- Calculate the quality of the grouping;
- Repeat;
- Assign a new vector closest to the defined center of the group;
- Calculate the new group center;
- Calculate the new grouping quality;
- Repeat until there is no change in center or grouping quality;
- The centers of the groups are in the center of gravity of the individual classes;

- The classification of the unknown sample is determined by the method of the nearest centre of the class, as shown in Figure 8.

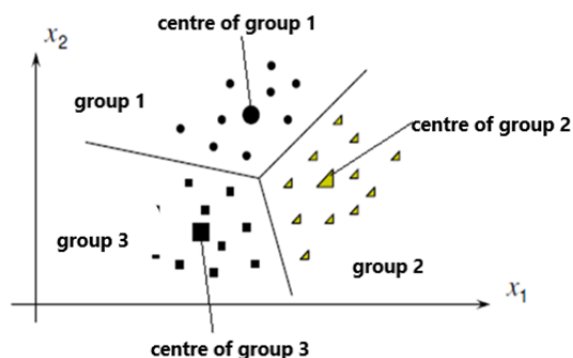


Figure 8. Starting centers

8. Conclusion

One of the basic problems when creating a pattern recognition system is the selection of appropriate numerical attributes, to describe the characteristics of the objects, which are extracted from the classification image. [22] A special mathematical concept is used in that direction. This concept is used in other sciences likewise, such as mechanics and statistics. The image is described at the time of the event, hence global image properties are used instead of local ones. Therefore, in this paper, emphasis is placed on the application of optical information in the processes of recognizing unknown patterns.

In this paper we pointed to:

- The process of pattern recognition;
- Decision methods for specimen classification (Nearest Neighbor Algorithm and Bayes theory of decision making);
- How pattern recognition classifiers are optimized;
- The manner of analysis of the specimens by groups;
- Basic concepts of K-means clustering algorithm.

References

- [1]. Daniel Jurafsky & James H. Martin. (2021). The hidden Markov model. Retrieved from: <https://web.stanford.edu/~jurafsky/slp3/A.pdf> [accessed: 01 April 2022].
- [2]. Home page for David B. Leake. (2021). CS Dept., Indiana U., Retrieved from: <http://www.cs.indiana.edu/~leake> [accessed: 15 September 2021].
- [3]. Koza, J. R., & Poli, R. (2005). Genetic programming. In *Search methodologies* (pp. 127-164). Springer, Boston, MA.
- [4]. Digital Ink Signatures – Concepts and Technologies. (2021). Retrieved from: <http://msdn.microsoft.com/library/default.asp?url=/library/enus/dntablet/html/tbconinksig.asp> [accessed: 10 April 2022].
- [5]. Oliver, S. (2020). *Biometrijski sustav* (Doctoral dissertation, Virovitica College).
- [6]. Crowd Dynamics Ltd. (2021). Retrieved from: <https://www.crowddynamics.com/>, [accessed: 10 April 2022].
- [7]. Hancer, E., Xue, B., Karaboga, D., & Zhang, M. (2015). A binary ABC algorithm based on advanced similarity scheme for feature selection. *Applied Soft Computing*, 36, 334-348.
- [8]. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- [9]. A. Sefidanis, Z. Kaevik, B. R. Velkovska (2020) "Pattern Recognition", Kosta Abras, Ohrid North Macedonia, 1(1), 31-41
- [10]. Minsky, M., & Papert, S. (1969). *Perceptrons* Cambridge, MA: MIT Press. *zbMATH*, 3(2), 99-120
- [11]. Čupić, M., & Šnajder, J. (2001). Web-orijentirani simulator modela neuronskih mreža s primjerom raspoznavanja novčanica i generatorom uzoraka, 4(2), 60-71
- [12]. Mitchell, T. M. (1997). *Machine Learning*, 1(2), 33-39
- [13]. Duda, R. O., Peter, E. H., & Stork, D. G. (2001). *Pattern Classification*, 4(5), 101-120
- [14]. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.
- [15]. Baughman, D. R., & Liu, Y. A. (1995). *Neural Networks in Bioprocessing and Chemical Engineering: With Disk*, 3(1), 90-109
- [16]. Galinec, G. (2010). *Razvoj softverskih senzora za identificiranje i inferencijsko vođenje rafinerijskih procesa* (Doctoral dissertation, doktorska disertacija, Fakultet kemijskog inženjerstva i tehnologije Sveučilišta u Zagrebu, Zagreb).
- [17]. Bolf, N., & Jerbić, I. (2006). Primjena umjetnih neuronskih mreža pri identificiranju i vođenju procesa. *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske*, 55(11), 457-468.
- [18]. Jiang, X., & Xu, W. (2006). Contactless hand recognition. *Proceedings of Dostupno*.
- [19]. Divjak, B., Lovrenčić, A. (2005). Diskretna matematika s teorijom grafova. TIVA Varazdin, Varazdin, (1)1, 20-40
- [20]. Sedgewick, R. (2002). *Algorithms in C++, Part 5: Graph algorithms*. Addison-Wesley.
- [21]. Veljan, D. (2001). *Kombinatorna i diskretna matematika [Combinatorial and Discrete Mathematics]*. Croat., Algoritam, Zagreb.
- [22]. Wu, B.Y., Chao, K., (2004). *Spanning trees and optimization problems*. Schapman Hall/CRC.