# Probabilistic Reasoning still the Shortest Path in Many Problems

Khalid Aboura
University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007
Fax: 61 2 9514 2633 - Email: kaboura@eng.uts.edu.au - http: www.uts.edu.au

**Abstract**
While the notion of chance may have been inherent in us from the beginning of the thinking age, it is a gambler's dispute in the 17th century that led to the creation by two famous French mathematicians of one of the most powerful mathematical theories, the theory of probability. It has since helped solve countless problems. In engineering, it is not the existence of uncertainty that is disputed, but how to deal with it. Probabilistic thinking becomes cumbersome when too many sources of variability exist. Engineers and physicists resort to "Black Box" models using data algorithmic approaches. Such is the case in some classification problems brought out by pattern recognition. We look at a character recognition problem, and present a probabilistic solution that is simple and easy to implement.
**Keywords** : Classification, Probability, Image Data Analysis.

## 1 Introduction

In 1654, letters written by Blaise Pascal and Pierre de Fermat formed the genesis of the theory of probability. A French nobleman, struggling with an apparent contradiction in a gambling problem, asked for Pascal's help. The fundamental principles of probability theory were laid down. In 1812, Pierre de Laplace introduced a formalism in his book, *Théorie Analytique des Probabilités*, and consequently applied probabilistic ideas to scientific problems. It took some time before arriving at a precise mathematical theory. The difficulty revolved around the definition of probability and the need to keep it applicable to a wide range of problems. In 1933, the Russian mathematician Andrey Nikolaevich Kolmogorov introduced an axiomatic approach, turning the probability concept into a mathematical theory. Probability theory is an undisputed modern theory, considered a subfield of measure theory. While no one questions the application of the theory as a mathematical tool for modeling random phenomena, until recently controversy still surrounded the meaning attributed to probability. Many schools of thoughts emerged, with the *Frequentists* and *Bayesians* representing the major ones. In 1763, a theorem by Reverend Thomas Bayes, a British mathematician, was published that eventually gave birth in the 20th century to the 'subjective' or 'personal' interpretation of probability. The Bayesian approach keeps the axiomatic laws of the theory of probability but enlarges the scope of application of the notion of probability. Today, the controversy

**International Journal of Computing Anticipatory Systems, Volume 20, 2008**
**Edited by D. M. Dubois, CHAOS, Liège, Belgium, ISSN 1373-5411 ISBN 2-930396-07-5**

between the frequentists and the Bayesians has all but faded, with many using the Bayesian approach in their modeling of uncertainty.

Most of us accept the inevitability of randomness in our daily lives and the reality of having to reason under uncertainty. A simple and convincing argument is the following: assume you are standing in front of a man who holds power of consequences over you. To obtain his graces, you must guess the number of coins in his pocket. The closer the answer is to the truth, the greater the benefit. No deterministic solution applies. Unless one can speed up the process of evolution and develop instantaneously X-rays powers, there is no way to deterministically know how many coins are in that pocket. And even so, the X-rays images obtained may not show clearly how many coins there are. At best one must 'guess', that is apply a probabilistic approach. One can see how probability evolved out of gambling problems.

Many uncertainty problems exist in all fields. While one can hope to explain phenomena in the universe through mathematical rules, dealing with decision problems under uncertainty requires probabilistic thinking. In many engineering problems, it is not the existence of uncertainty that is disputed, but how to deal with it. Probabilistic thinking becomes cumbersome when too many sources of variability influence the solution. Parsimonious probability models are hard to build, and engineers often turn to anything that can solve the problem. In adopting methodologies, they often guide themselves by the results. Most probabilists would acknowledge the difficulty of developing complicated statistical models in such cases, and may accept the use of data algorithmic methods. These are methods that emerged in the fields of data analysis, adopting the concept of 'no model' or a 'black box' approach. Neural networks [6] and support vector machines [8] are examples of such approaches. In some classification problems, the variability in the data is so large that it encourages to avoid the construction of a probability model. Neural networks have been used in these problems. Once reasonable results are obtained, then trends are started. Such is the case in the optical character recognition in the license plate recognition problem. We show, by the presentation of a probabilistic solution, that in such case there is no need for neural networks, a solution widely adopted. The probabilistic solution we introduce is simple, easy to build and easy to implement.

## 2 Pattern Recognition and Classification

Ever since the computing revolution was started by the Babylonians around the fourth century B.C. with the invention of the Abacus, humans have tried to automate tasks. This pursuit has lead to the field of classification in pattern recognition. Classification is a daily human function. Many functions we perform start with a classification that we usually conduct without any great deal of effort. What seems easy for us to do is in fact a very difficult process to dissect and program into logical steps. Simply recognizing a character in an image can be a daunting task for a computer, particularly if the image comes each time with variation and noise. Classification is a field that has evolved significantly in recent years. While statisticians developed the field of *Statistical Classification*, engineers, physicists and computer scientists developed data algorithmic approaches. These are data analysis methods that do not require statistical modeling, such as neural networks and support vector machines. An example of a data algorithmic approach in classification is the use of acoustic emission for the investigation of local damage in materials and the application of neural networks to the study of the acoustic signals [12]. In the mid 1980's, neural networks, along with decision trees provided two new powerful algorithms for fitting data. Upon the successes of these approaches, a new direction in data analysis developed that is removed from the traditional statistical approach. Breiman [7], a professor of statistics, argues for the goodness of the data algorithmic approach. The statistical approach models uncertainty using probability, while data solutions chooses a 'black-box' approach. The classifiers from both schools have been compared on many problems and no single classifier outperforms the others [21]. The classifier performance depends greatly on the characteristics of the data.

### 2.1 Classification

Classification can be described as the task of building a function $\mathcal{F} : \mathcal{Z} \rightarrow \mathcal{C}$, known as a classifier, that maps an input or feature set $\mathcal{Z}$ into a class set $\mathcal{C}$. In one of its forms, supervised learning, the set $\mathcal{C}$ is known, and there exit a training set $\mathcal{D}$ of inputs for which the mapping is known. Using $\mathcal{D}$, the classifier $\mathcal{F}$ is built, estimated or approximated, then used for any incoming input $z \in \mathcal{Z}$. $\mathcal{F}$ often takes the form of a statistically estimated function, or an algorithm from the family of neural networks or support vector machines. A typical problem in classification for which the support vector machines in their original form apply, is the two classes classification or binary classification. In that case, $\mathcal{C}$ has only two elements. There are many methods for classification and they fall roughly into the two major categories mentioned: (i) statistical and (ii) data algorithmic. Examples of classification algorithms include: linear regression, quadratic regression, Fisher's linear discriminant, logistic regression, Naive Bayes method, k-nearest neighbor, decision trees, neural networks, Bayesian networks and support vector machines. A typical simple approach such as the k-nearest neighbor partition the feature space into regions, based on some metric and assumption of the distribution in the input space. A more optimized approach is the support vector machines which split the input space into two regions in the case of binary classification. Assuming such solution exists, that is the input space is separable, it maximizes a measure of that separability. Statistical methods proceed differently and use one of the following two general approaches; (i) estimate from the training data set $\mathcal{D}$ a parameter $\theta$ that is used to parameterize the probability of a class $C$ given a data point $Z$, i.e. $\text{Prob}(C|Z) = f(Z;\theta)$. Typically, $\theta$ is estimated from $\mathcal{D}$, or simply averaged over the values in $\mathcal{D}$, or (ii) use Bayes theorem and inverse the probabilities by computing

$$\text{Prob}(C|Z) = \frac{\text{Prob}(Z|C)\text{Prob}(C)}{\sum_{S \in \mathcal{C}} \text{Prob}(Z|S)\text{Prob}(S)}. \tag{1}$$

Using $\mathcal{D}$, Prob$(Z|C)$ is constructed (modeled). It is called the likelihood function and is used in Bayes theorem in eq. 1 to provide the class probabilities.

### 2.2 A Pattern Recognition Problem

A pattern recognition problem that has been worked on extensively is the recognition of characters in license plates. A licence plate recognition system is designed for the automatic identification of a vehicle through its license plate. The system consists of a series of steps starting with the detection of a vehicle, the capture of images and the recognition of characters in the license plate. The last step involves image analysis in three parts; (i) the localization of the license plate in the image, (ii) the segmentation of characters from the localized license plate region, and (iii) recognition of those characters. These steps need to be performed automatically and require intelligent algorithms to achieve a high reliability. Many solutions have been applied and most of them use a learning approach. A historical data set is collected and used in the training and validation of the algorithm. The data set for the recognition of the characters consists of images of license plate characters that have been extracted from images of vehicles. They are binary images that have been treated using image analysis techniques. The characters are visually inspected one by one, and classified manually in the 36 possible classes {A,B,C, ...,X,Y,Z,0,1, ...,8,9}. Each set of characters is then split into a training set and a validation set. The problem is to develop an algorithm that can recognize an extracted character as one of the 36 possible characters. It is a classification problem and the solution is known as an optical character recognition (OCR) algorithm. While OCR applies in license plate recognition, it is a field that has a much wider stage of applications. It dates back to 1929 and is a method designed to translate automatically images of handwritten or typewritten text. There are software packages that can translate a faxed page or a scanned image into text. However, these OCR commercial packages do not yield good results in the case of the license plate problem. A series of approaches were developed to recognize license plate characters. The most common approaches are the correlation-based template matching [14], and neural networks [11]. Many other approaches are used, for example support vector machines or binary classifiers [16].

#### 2.2.1 Neural Networks

Neural networks (NN) are applied successfully in many prediction and recognition problems. In license plate recognition, they are used to localize the license plate in the image and to recognize the extracted characters of the plate. Neural networks grew out of research in artificial intelligence. A neural network is an artificial network made up of sets of interconnected nodes called neurons. In its simple form, there is a set of input nodes, features from the image being processed, that are connected through a network of nodes, hidden layers, to a set of output nodes, the classes to which the image belongs. The NN is trained to recognize by feeding it a set of inputs to which the outputs are known. The NN processes the inputs one by one and compares the resulting outputs against the desired outputs. Errors are calculated and weights which control the strength of network connections are adjusted at each iteration. The training is stopped once the NN reaches a satisfactory level of recognition. The set of final weights is used for processing new data.

### 2.2.2 Template Matching

The other approach most used in character recognition in license plate recognition is template matching. It is a technique in image analysis for scanning an image template until part of it matches an image at hand. There are many variants in the application of template matching to character recognition. In its simplest form, the image in a binary form is compared with same size parts of the template image using a suitable metric. The metric can be the euclidian distance or a correlation measure between the pixels of the image and the template. For example, the cross-correlation, a statistical measure used by Horowitz [13] and Pratt [18] for image recognition, can be a metric for template matching. The template matching approach is often combined with other methods in character recognition. However, it remains a method based on the minimization of a distance between two images.

## 3 The Probabilistic Approach

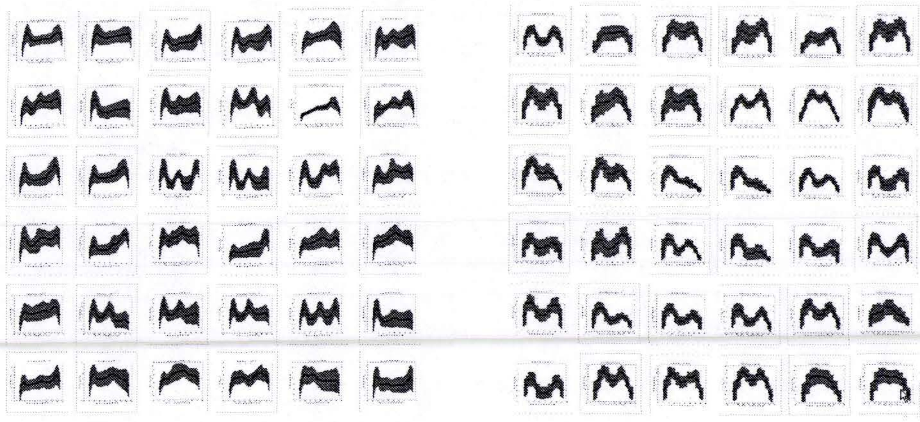
In the optical character recognition of the license plate problem, uncertainty prevails as to what class an input image belongs to. Given that it is a stochastic problem, one expects probability answers, if one adheres to the principle that probability is the only coherent way to address uncertainty [15]. Despite some attempts, the only noticeable probability based research direction is that of the probabilistic neural networks [4]. The probabilistic neural network was developed by Specht [20] and provides a solution to classification problems using Bayesian classifiers and the Parzen estimators. It is a class of neural networks which combine statistical pattern recognition and feed-forward neural networks technology. It is characterized as having very fast training times and it produces outputs with Bayes posterior probabilities. Probabilistic neural networks are effective in pattern recognition. However, the problem at hand is a simple OCR problem and can be addressed by a simple probabilistic approach.

### 3.1 Image Data

To solve the problem probabilistically is to treat features, or statistics, from the input image, using a probability model. The input character image to the OCR module is usually a binary image. *Thresholding* is the operation of converting the original image into a binary image. Thresholding is a mathematical operation performed on image data to translate information from a multi-dimensional space, in the case of color images, or a real valued interval, in the case of gray scale images, to the binary set $\{0,1\}$. It is an operation that divides the images into a background image and a foreground image. In the case of the license plate problem, the foreground includes the characters of the license plate. Depending on the method, thresholding manages to bring out desired features of the images while burying the rest into the background. There is a large variety of methods in thresholding. Many fields require thresholding such as medical imaging, visual surveillance and character recognition. In general, any technology that analyzes images finds it easier if not necessary to reduce the digitized image to a binary form. Often than not, thresholding determines the success of the technology that uses it. In license plate recognition, thresholding is essential in all steps. However, the issue is hardly ever mentioned, always assuming an effective or at the very least an adequate foreground separation. Aboura [1] introduces a new thresholding method that succeeds in highlighting license plate characters in an image. The method uses color-based histogram classification information and a linear regression model. In the character recognition problem at hand, using a thresholding method, the extracted character images are converted to binary image data, cleaned, cropped and normalized, then divided into two sets; training data and validation data. The characters are visually inspected one by one, and classified manually in the 36 possible classes {A,B,C, ..., X,Y,Z,0,1, ...,8,9}.

#### 3.1.1 Statistics of the Character Image

Examples of statistics of an extracted character are the fill percentage and the projected foreground. The fill percentage is the proportion of foreground pixels in the binary image. Let $F$ be the matrix of the digitized binary image, with $F(x,y) = 0$ or $1$ if pixel at location (x,y) is background or foreground, respectively. Then the fill percentage is $\sum_x \sum_y F(x,y)/(N_x N_y)$, where $N_x$ and $N_y$ are the height and width of the image in pixels. The projected foreground is the normalized projection of the foreground on the $x$ and $y$ axes of the image, $\sum_y F(x,y)/N_y$, $x = 1,\ldots,N_x$ and $\sum_x F(x,y)/N_x$, $y = 1,\ldots,N_y$, respectively. These projected histograms seem to provide distinguishing information, as shown in Figure 1, where the means of these statistics, taken over the training data set of images, is plotted, within one standard deviation. These statistics can be used for the recognition of characters. In a preliminary study, the *minimization of squared errors* was used in an attempt



(a) Projection on the x axis (b) Projection on the y axis

**Fig. 1**: Historical means of projected foreground for the 36 characters

to recognize the characters using the normalized projection foreground:

$$\textbf{Minimize}_{C=A,\ldots,9} \sum_x (\sum_y F(x,y)/N_y - \mu_x^C)^2 + \sum_y (\sum_x F(x,y)/N_x - \mu_y^C)^2 \quad (2)$$

where $\mu_x^C$ and $\mu_y^C$ in eq. 2, $C = A, B, \ldots, X, Y, Z, 0, 1, \ldots, 8, 9$, are the historical means from the training set (Fig. 1). The approach failed to return a good recognition. In an attempt to remedy to a possible lack of information from the statistics, the projected foreground was augmented with the distance of the pixels to the side of the image, thus incorporating the information about the locations of the foreground pixels. The minimum squared errors method failed again, an indication that these features of the image do not offer enough information to fully distinguish among the characters.

### 3.2 The Theoretical Solution

In general, let $Z$ be the random variable that represents the statistical feature of a character image. $Z$ need not be, and often isn't univariate. The probabilistic approach starts by building a probability model for that feature in the form of $\text{Prob}(Z|C)$, $C = A, B, \ldots, 9$. For each image in the training set, the value of $Z$ is computed. Data analysis tools are used, along with any engineering and prior knowledge to arrive at the probability model $\text{Prob}(Z|C)$. Seen as a function of the event $C$, that is the character is $C$, the probability model $\text{Prob}(Z|C)$ is known as the likelihood function $\mathcal{L}(C) = \text{Prob}(Z|C)$. This likelihood function is at the heart of the probabilistic approach. If this model is built properly, and the statistical feature $Z$ offers enough information about the character's class, the probabilistic approach will be effective. Given a likelihood model, the probabilistic approach proceeds as follows. Let $z$ be the value of $Z$ for an image being analyzed. Then the probability that the character is $C$, given the data $z$ is

$$\text{Prob}(C|Z = z) = \frac{\text{Prob}(Z = z|C)\text{Prob}(C)}{\sum_{S=A}^{S=9} \text{Prob}(Z = z|S)\text{Prob}(S)} \tag{3}$$

for $C = A, B, \ldots, 9$. Prob($C$) in eq. 3 is the model built with any prior knowledge about what $C$ might be. It is called the prior distribution. The denominator is a normalizing constant. Prob($C|Z = z$) is the posterior distribution of $C$. The solution is given in eq. 4 as the selected character $\hat{C}$, mode of the posterior distribution,

$$\text{Prob}(\hat{C}|Z = z) = \mathbf{Max}_{C=A,B,\ldots,9}\ \text{Prob}(C|Z = z). \tag{4}$$

If there is no prior knowledge as to what $C$ might be, then the discrete uniform distribution can be used where $\text{Prob}(C) = 1/36$ for $C = A, B, \ldots, 9$. This probabilistic approach is often referred to as a *Bayesian* approach. A non Bayesian approach would simply maximize the likelihood function, ignoring the prior component, and select $\hat{C}$ such that

$$\text{Prob}(Z = z|\hat{C}) = \mathbf{Max}_{C=A,B,\ldots,9}\ \text{Prob}(Z = z|C). \tag{5}$$

Both these statistical approaches rely on the likelihood function. These simple operations are at the heart of many probabilistic predictions, classifications and inferences. While seemingly simple in principle, their success depends on the proper selection of the random variable $Z$ and the probability model Prob($Z|C$).

### 3.3 The Probability Model

In the search for a statistic $Z$, the author arrived at the conclusion that the values of the pixels in the binary image hold all the information needed to recognize the character in an image. $Z$ is defined as the vector of the values of all the pixels in the image. The two-dimensional binary image data is converted into the one-dimensional array Z. Each of the Z entries is either 0 or 1, and corresponds to the value of a pixel in the binary image. For each pixel, the Bernoulli probability model $\theta_i^{Z_i}(1-\theta_i)^{1-Z_i}$ is applied, $Z_i$ being the value of $Z$ at pixel $i$. Making the assumption of conditional independence of the pixel values given an image, the likelihood function is constructed

$$\mathcal{L}(C) = \text{Prob}(Z|C) = \prod_{i=1}^{|Z|} \theta_i^{Z_i}(1 - \theta_i)^{1-Z_i} \tag{6}$$

where $|Z|$ is the cardinal, or vector size, of $Z$. To estimate the proportion $\theta_i$ for pixel $i$, a number of approaches are available. But given that the sizes of the historical sets are relatively large, the estimates converge to the average

$$\hat{\theta}_i = \sum_{j=1}^{N_C} x_{i,j}/N_C \tag{7}$$

where $N_C$ is the size of the training set for character $C$, and $x_{i,j} = \{0 \text{ or } 1\}$ is the value of pixel $i$ for image $j$ of the training set. This is done for each character $C$. For simplicity of notation, $\theta_i$ is used in eq. 6, when in fact it is a $\theta_i(C)$ that differs for each $C$. From a computational point of view, the assessment of the likelihood parameters is very simple. For each character $C$, all the images of the training set are added, then divided by the size of the set $N_C$, automatically providing a matrix of estimates $[\hat{\theta}_i]_{i=1}^{|Z|}$. This is a simple operation, inexpensive computationally, that replaces the training of a neural network. It needs to be done only once, and the estimates matrices are used subsequently to recognize the characters. Figure 2 shows

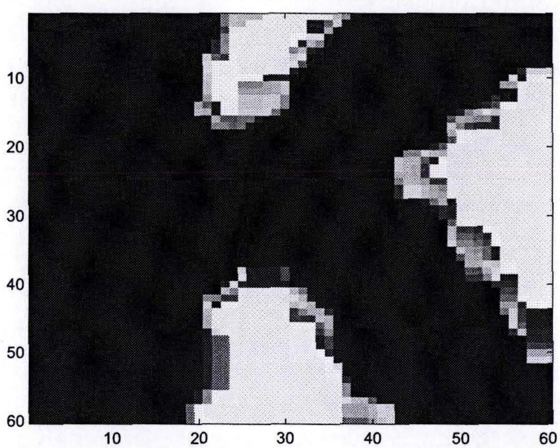



**Fig. 2**: Likelihood image of character K

the matrix $[\hat{\theta}_i]_{i=1}^{|Z|}$ for character K for the particular training set used. Each value of the pixel $i$ of the image in Figure 2 is the estimate $0 \leq \hat{\theta}_i \leq 1$ for character K. It is the estimate of the parameter of the Bernoulli model for pixel $i$, for character K. In essence, this method builds a likelihood function value for each character $C$, that results in a matrix, where the values of 1 signify that the corresponding pixels are always present in the foreground of $C$ (in dark in figure 2), the values 0 mean that the corresponding pixels are always background in $C$ (in light in figure 2), and the values in between correspond to pixels $i$ that are present in $C$ with a probability

$\hat{\theta}_i$. One observes that the K in the image is not a perfect one, as the extracted images of characters in LPR are most often taken at angles and subject to many sources of noise and deformation. In addition, the colors of the image do not show all the nuances in the values. But the matrix $[\hat{\theta}_i]_{i=1}^{|Z|}$ is computed with accuracy and provides good estimates of the probabilities of the foreground existence.

Once the likelihood function is constructed, it is used to recognize characters in a simple operation. Let $z$ be the realization of the statistic $Z$ for a binary image that has received similar cleaning, cropping and resizing as have the images of the training set. Then

$$\mathrm{Prob}(C|z) = \frac{\mathrm{Prob}(z|C)\mathrm{Prob}(C)}{\sum_{S=A}^{S=9} \mathrm{Prob}(z|S)\mathrm{Prob}(S)} \tag{8}$$

where

$$\mathrm{Prob}(z|C) = \prod_{i=1}^{|Z|} \hat{\theta}_i^{z_i}(1-\hat{\theta}_i)^{1-z_i} \tag{9}$$

noting that the appropriate $\hat{\theta}_i$'s correspond to the given $C$. This posterior probability distribution ranks the characters $A, B, \ldots, X, Y, Z, 0, 1, \ldots, 8, 9$ for their likelihood of being the character in the image being treated. This method is a Bayesian probabilistic approach. One introduces prior knowledge in the above in the form of $\mathrm{Prob}(C)$ for each character $C$. Such knowledge can vary from country to country for example, where the position of a character in the license plate may indicate whether it is a letter or a number, for example. If such knowledge does not exist, or is hard to embody in a formal model, then the characters are declared to be equally probable apriori, letting $\mathrm{Prob}(C) = 1/36$ in this case. This reduces, from a solution point of view, to maximizing the likelihood function.

This method yielded excellent results. As most methods in this category, it fails to distinguish, as such, fully between characters 2 and Z, 5 and S, 1 and I, B and 8, and O,0,D and Q. However, using the same logic and applying it exclusively to parts of the image, the authors in [3] reach a 97% reliability. This is a high reliability for a pure approach that doesn't use heuristics. In practice, one would supplement with redundancy checks and use multiple images per vehicle to obtain a full reliability. As such, the method is highly accurate. In essence, the approach is a Naive Bayes approach, as known in the classification literature, and provides a probabilistic solution to the optical character recognition. We show that a simple probabilistic solution is still the easiest way to solve a particular problem. Aboura [2] does the same by presenting a license plate localization solution using a statistical analysis of the Discrete Fourier Transform of a signal. This approach resulted in a more reliable localization of the license plate, at a faster speed.

## 4 Conclusion

We introduced a simple and inexpensive method to solve a relatively important problem, using a probabilistic approach. Neural networks are routinely used in this problem. Debates about how neural networks fit into classification, and how they compare and relate to statistical methods have been conducted ([19], [22]). While some have adopted an outright stance against neural networks [10], most accept that there are two philosophically different approaches in data analysis [7]. Recently, more use is made of combining both approaches [17], particularly by engineers. In a way, the statistical community failed to address the need of the data analysis community. Data analysts and engineers often require answers to problems with large data sets. But in a search for a solution, there is also danger in thinking that any answer is a good answer. The laws of probability are not a truth. They form only a model for thinking about uncertainty, based on axioms. But they are the result of centuries of building, bit by bit, a theory that has been tested by many and is well laid out ([9], [5]). It can be said that it is universally accepted that probability is the only coherent way to solve decision problems. But what are data analysis problems but decision problems.

## References

[1] Aboura Khalid (2008), Automatic Thresholding of License Plate. International Journal of Automation and Control. In print.

[2] Aboura Khalid (2007) Stochastic Analysis of Discrete Fourier Transform Data for License Plate Localization. Technical Report, University of Technology of Sydney.

[3] Aboura Khalid and Al-Hmouz Rami (2007) Probabilistic License Plate Optical Character Recognition. Technical Report, University of Technology of Sydney.

[4] Anagnostopoulos Christos Nikolaos E., Anagnostopoulos Ioannis E., Loumos Vassili and Kayafas Eleftherios (2006) A License Plate-Recognition Algorithm for Intelligent Transportation System Applications. IEEE Transactions on Intelligent Transportation Systems. Volume 7, No. 3, pp. 377-392.

[5] Bernardo José M. and Smith Adrian F. M. (2000) Bayesian Theory. John Wiley and Sons.

[6] Bishop Chistopher M. (1995) Neural Networks for Pattern Recognition. Oxford University Press.

[7] Breiman Leo (2001) Statistical Modeling: The Two Cultures. Statistical Science. Volume 16, No. 3, pp. 199-231.

[8] Burgues Christopher J.C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. Vol. 2, pp. 121-167.

[9] Degroot Morris H. (1970) Optimal Statistical Decisions. McGraw-Hill, New York.

[10] Dewdney Alexander K. (1997) Yes, We Have No Neutrons: An Eye-Opening Tour Through the Twists and Turns of Bad Science. John Wiley and Sons.

[11] Fahmy Maged M. M. (1994) Automatic Number-Plate Recognition: Neural Network Approach. Proceedings of the Vehicle Navigation and Information System Conference. pp. 99-101.

[12] Godin Nathalie, Huguet Stéphane and Gaertner Roger (2005) Integration of the Kohonen's Self-Organising Map and k-means Algorithm for the Segmentation of the AE Data Collected during Tensile Tests on Cross-Ply Composites. NDT & E International. Vol. 38, pp. 299-309.

[13] Horowitz M. (1957) Efficient Use of a Picture Correlator. Journal of the Optical Society of America. Vol. 47, No. 4, pp. 327-327.

[14] Ko Mi-Ae and Kim Young-Mo (2003) License Plate Surveillance System Using Weighted Template Matching. Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop. pp. 269-274.

[15] Lindley Dennis V. (1987) The probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems. Statistical Science. Vol. 2, No. 1, pp. 17-24.

[16] Odone Francesca (2006) Experiments on a License Plate Recognition System. technical Report DISI-TR-06-17, DISI, Universita degli Studi di Genova, Italy.

[17] Osl Melanie, Baumgartner Christian, Tilg Bernhard and Dreiseitl Stephan (2008) On the Combination of Logistic Regression and Local Probability Estimates. Proceedings of the Third International Conference on Broadband Communications, Information Technology and Biomedical Applications, Pretoria, South Africa, pp. 124-128.

[18] Pratt William K. (1974) Correlation Techniques of Image Registration. IEEE Transactions on Aerospace and Electronic Systems, Vol. 10, No. 3, pp. 353-358.

[19] Ripley B. D. (1993) Neural Networks and Related Methods for Classification. Journal of the Royal Statistical Society B. Vol. 2, pp. 121-167.

[20] Specht Donald F. (1990) Probabilistic Neural Networks. Neural Networks. Vol. 3, pp. 109-118.

[21] van der Walt Christiaan and Barnard Etienne (2006) Data Characteristics that Determine Classifier Performance. Proceedings of the Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa. pp. 160-165.

[22] Zhang Guoqiang P. (2000) Neural Networks for Classification: A Survey. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 30, No. 4, pp. 451-462.