The Eighth International Conference on Intelligent Technologies (InTech'07)
Sydney, Australia, 12-14 December 2007

Statistical Foundations of Intelligent Technologies

Khalid Aboura
Information and Communication Technologies
University of Technology, Sydney, Australia
Email: kaboura@eng.uts.edu.au

Abstract—**Intelligent Technologies problems have one common element; uncertainty. Randomness, natural or created, is an important part of the engineering problems in Intelligent Technologies. Modeling it properly and dealing with it in a coherent manner is essential to any solution. In this article, we address the issue of uncertainty in Intelligent Technologies through a classification problem. We look at the problems of pattern recognition and classification, an important area in automation, and use the example to highlight a simple but effective way to deal with uncertainty in complex environments.**

I. Introduction

Ever since the computing revolution was started by the Babylonians around the fourth century B.C. with the invention of the Abacus, and since its modern times take off in 1941 with the completion of the first general purpose programmable calculator by the German engineer Konrad Zuse, humans have been trying to automate functions and tasks. This pursuit has increased in recent times leading the way to the fields of Robotics and Autonomous Systems, Intelligent Transportation, Expert Systems, Pattern Recognition, and many others that fall within the area of Intelligent Technologies. Examples of applications abound, from the automation of simple tasks in engineering to the robotization of factories. A multitude of projects in research labs pursue new applications of intelligence in technologies such as the development of vision-guided helicopters, the automation of data search engines, vision based tracking and recognition, etc. In all these problems, there is one common element that cannot be escaped, and that is the existence of uncertainty. Uncertainty prevails in all aspects of the problems considered. Be it natural randomness, or uncertainty created by the problems, the lack of complete knowledge requires a proper logic to deal with the added complexity. In this article, we address the issue of uncertainty in problems of intelligent technologies, and how to manage it in a coherent way through the use of an example in pattern recognition and classification. Pattern recognition is an important area in automation, and the example highlights a simple and effective way to deal with uncertainty in complex environments.

II. Intelligent Technologies

Intelligent Technologies gather a large number of areas in many fields of applications. In this context, we view it as the application of intelligence to engineering tasks. A particular area would be robotics and autonomous systems. Another area of intelligent technologies is Intelligent Transportation. Machine Learning embodies a third area of application. We briefly introduce these areas of research to give a flavor of the richness of problems. We then focus on a classification problem to address the modeling of uncertainty in engineering.

1) Robotics and Autonomous Systems: The field of Robotics and Autonomous System is too wide to cover in an article. Robotics is an essential a tool in the automated industry. In the 19th century, the first remote control vehicles were built, and recently robots are successfully exploring the planet Mars. Robots are still not dotted with intelligent thinking, however many of their tasks involve the incorporation of intelligence in the algorithmic solutions of their behavior, particularly when uncertainty is involved in a decision problem. Uncertainty is found both within a robot and its environment. Sources of uncertainty are (i) randomness in the environment, (ii) imperfect assessment and (iii) imperfect completion of a task. Randomness will always exist in the environment of a robot, even in the most secure setup. The example of an electrical problem is a most basic random event, while a person stumbling unpredictably in front of the robot is another. Robots conduct many assessments, particularly that of geographical locations. Real-time object tracking is a critical task in many robotics applications and autonomous vehicles. If a path is visually obstructed by an obstacle, than there is uncertainty as to what is on that path. Finally, the issue of whether the robot will accomplish properly some task in the future introduces uncertainty at a decision moment in the present. These sources of uncertainty render the field of robotics and autonomous systems a research area that constantly addresses *decision problems under uncertainty*.

2) Intelligent Transportation: Since the early 1990s, the Intelligent Transportation Systems (ITS) program has been a worldwide initiative to introduce intelligence and information technology in transportation systems to reduce risks and increase efficiency. Particular attention was given to road transportation due to its volume and impact on society. The ITS programs are motivated by the problems caused by traffic congestion worldwide due to increased motorization and exploding population growth, the need to constantly reduce fatality rates and increase safety, the economic and environmental factors associated with fuel consumption and the recent drive to bring information to the potential customer through wireless internet and mobile technology. Research by some ITS programs focused on the development of warnings systems within the vehicles for obstacles detection and avoidance, lane departure, intersection collision and other hazards. Traffic management is another area of interest in ITS, where the aim is the development of Integrated Traffic Management Systems to monitor, control, and manage traffic on streets and highways. In transit systems, ITS programs work on traffic signal priority, automatic fare collection, automatic passenger counting, rider information, and a number of other services optimization in public transport. Key to such systems are control centers with communications capabilities, and sensors on the transportation roads. These systems facilitates rapid response to incidents, allows variable message signs and smart traffic signals. A more recent ITS research direction is for the development of on board systems that provide traffic information, navigation aids, local tourist and commercial adds, and internet access. Many research and development projects are conducted in different parts of the world. In the United States, the initiative is promoted by the U.S. Department of Transportation, that ensures a national ITS architecture and standards [1]. Diverse groups dedicated to developing and deploying intelligent transportation systems include private corporations, public agencies, academic institutions and research centers. The private sector includes auto manufacturers, consumer electronics manufacturers, telecommunications companies, consulting engineering firms, systems integrators and other companies (see the Intelligent Transportation Society of America [2]). In Europe, ERTICO [3] is the platform for about 100 partners working on providing Intelligent Transport Systems and Services. In Japan, the Vehicle, Road and Traffic Intelligence Society (VERTIS) [4] has similar goals as those of its counterparts in Europe and the US [5]. Intelligent Transport Systems Australia, ITS Australia, is Australia's organization, representing government, consumer organizations and academia, focused on facilitating the development and deployment of advanced technologies across all modes of transport; air, sea, road and rail [6]. Multiple intelligent transportation systems have been deployed in the US and elsewhere. All of these systems require the development of engineering solutions that include modeling uncertainties. An example is the License Plate Recognition problem. It is is a valued technology in some of the ITS programs. The goal is to use automatically the license plate information of a vehicle through the analysis of video images. Digitized images comprise a large number of data points. There is a sizeable random element that is included in the problem as noise and variation. This creates an interesting stochastic problem.

3) Machine Learning: Machine Learning is an Artificial Intelligence field that attempts to design and implement algorithms for computers to learn. The learning focuses on recognition, classification and language processing mostly, due to the data oriented nature of the computers. The applications vary from detecting credit card fraud, to speech and handwriting recognition, to industrial robotization, to medical diagnosis, to web search engines. Pattern Recognition is an integral part of machine learning. Classification is an exercise in pattern recognition, and statistical classification is often associated with machine learning. License plate recognition is a pattern recognition problem that includes a classification exercise in the form of an Optical Character Recognition solution.

III. Pattern Recognition and Classification

Classification is a daily human function. We do it all the time, when checking the temperature of the room, when tasting food, when getting an estimate of the time, when navigating our way around. Many functions we perform start with a classification that we usually conduct without any great deal of effort. It is also a tool in the survival mechanism of animals in the wild. If one looks at animals in the wild, they are constantly screening their surroundings for smells, sounds or signs of danger. Each time instant, a classification is made into at least two classes, 'danger' and 'no danger'. In fact, some think that classification is the most basic part in the working of our brains, and that thinking in some cases can be reduced to a series of classifications. What seems so easy for us to do in a classification act, 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 slight variations and plenty of noise. Classification is a field that has evolved significantly in recent years, due to interest in pattern recognition and machine learning. While statisticians developed the field of *Statistical Classification*, engineers, physicists and computer scientists developed data algorithmic approaches that required no statistical modeling. These two approaches both lead to successes and failures in classification problems, without interfering too much with each other. The classification problems often being complex ones with a large number of variables, the researchers were pragmatic and didn't worry much about the foundations of the solutions, as long as they were solutions. In the last section of this article, we distinguish between these two approaches from the theoretical point of view. While one attempts to establish a theoretically sound, although axiomatic, approach to dealing with uncertainty, the other provides solutions that are often closer to mathematical optimization solutions than to stochastic solutions and that are supported because 'they work' on a number of problems. In this section, we introduce classification and point to a recognition problem.

Classification is found in many fields. In medical imaging, for example, statistical methods can be used to classify brain tissue in magnetic resonance (MR) images [7]. Manual recognition by a medical expert of the three brain tissue types, white matter, gray matter, and cerebrospinal fluid, is an extremely time consuming task. The images are three dimensional, and the volume of data involved is large. An automated approach is sought and statistical methods are used in this case. 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 [8]. In the mid 1980's, neural networks [9], along with decision trees provided two new powerful algorithms for fitting data [10]. Upon the successes of these approaches, a whole new direction in data analysis developed that is removed from the traditional statistical approach. Breiman [10], a professor of statistics, argues for the goodness of the data algorithmic approach, using many large real life problems he encountered as a consultant. A third approach is classification that relies solely on engineering knowledge and basic data fitting. Such is the case when acoustic emission are studied to monitor in real-time upper-ocean air-sea interface processes, and classify recorded noise into 'rain', 'heavy rain', 'drizzle' and 'no rain' [11]. In this case, understanding of the noise over the ocean is used in the classification, eliminating the need for heavy statistical modeling, or for the 'black-box' / 'no-modeling' data algorithmic approach. These three examples can be seen as three fundamental directions in classification. The third approach is surely preferred when possible. However, many classification problems do not lead to pure engineering or physics explanations. The uncertainty in the classification problem is the hardest part to deal with. The statistical approach models uncertainty using probability, while data solutions choose to use a 'black-box' approach, with no models. These classifiers from both schools have been compared on many problems, and no single classifier outperforms the others [12], [13]. All studies were based on experimental trials, and one cannot reach any theoretical conclusion. It has been concluded that the classifier performance depends greatly on the characteristics of the data.

A. *Classification*

Classification can be described as the task of building a function, known as a classifier $\mathcal{F}$, that maps an input or feature set $\mathcal{Z}$ into a label or class set $\mathcal{C}$:

$$\mathcal{F} : \mathcal{Z} \rightarrow \mathcal{C} \tag{1}$$

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 in (1) is known, that is for each input $Z \in D$, the corresponding class C is known. Using $\mathcal{D}$, the classifier $\mathcal{F}$ is built, then used for a new 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 suck 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:

- The first approach is to estimate from the training data set $\mathcal{D}$ a parameter θ that is used to parameterize the probability of a class C given a data point Z,

$$\text{Prob}(C|Z) = f(Z;\theta). \tag{2}$$

  Typically, θ is estimated from $\mathcal{D}$, or simply averaged over the values in $\mathcal{D}$, $\text{Prob}(C|Z) = \int(Z;\theta)\text{Prob}(\theta|\mathcal{D})d\theta$.
- The second approach is to use Bayes theorem and inverse probabilities by computing

$$\text{Prob}(C|Z) = \frac{\text{Prob}(Z|C)\text{Prob}(C)}{\sum_{S \in \mathcal{C}} \text{Prob}(Z|S)\text{Prob}(S)} \tag{3}$$

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

(2) and (3) provide a solution for $\mathcal{F}$ in (1) in a probabilistic form.

B. *Support Vector Machines*

While neural networks are typical of a data algorithmic solution and have been the major initial driving force behind that philosophy, support vector machines have become increasingly popular with data analyst. We introduce them here as an example of data driven methodology. Vapnik [14] provided strong theoretical foundations for support vector machines. They have proved to be successful in many problems, and from a mathematical point of view, there is nothing wrong with support vector machines. In classification, they offer one optimal solution, given that a particular objective function is maximized and that the optimal solution exist. Burgues [15] provides an excellent tutorial on support vector machines that lists among the many articles written on the subject. In short, consider the binary classification problem, and let the training data be $\mathcal{D} = \{(z_i, c_i), c_i = -1, +1, i = 1, \ldots, n\}$. The support vector machines are hyperplanes separating $\mathcal{D}$ such that a measure of that separation is optimal (see figure 1). "Optimal" is defined as meaning that the distance of the hyperplane to any prediction vector is maximal. The set of vectors in red points and blue points that achieve the minimum distance to the optimal separating hyperplane are called the support vectors. Their coordinates determine the equation of the hyperplane. The hyperplanes in figure 1 are moved around until settled on an optimal one that maximizes the separation. In a binary classification problem, separability does not always occur. However, if the dimensionality is increased by adding as variables all quadratic monomials in the original input variables Z_i; that is, all product terms of the form $Z_{ik}Z_{il}$, then the possibility of separation is greater. If no

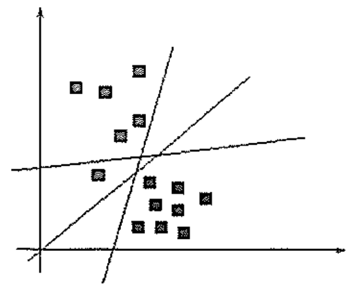


Fig. 1. A simple linear support vector machine

separation occurs, by adding cubic monomials as input features then one increases the chances for separability. If separability does not occur and increasing the dimensionality becomes infeasible, there is another version of support vector machines that provides a solution. The popularity of support vector machines is that they offer a computationally efficient quadratic programming solution. In addition, they offer the "Black-Box" effect, that is they are a mathematical solution that didn't require any modeling of the actual problem. It is all data driven.

C. *A Pattern Recognition Problem*

License plate recognition is a pattern recognition problem. It consists of a sensor that collects observations in the form of images, a feature extraction algorithms that detects characters in an image, and a classification solution that recognizes those characters. It is the use of video captured images for the automatic identification of a vehicle through its license plate. The applications are numerous and include parking lot attendance, toll roads, surveillance, identification of stolen vehicles and border crossing. A license plate recognition system consists of a series of steps for the detection of a vehicle, the capture of images and the recognition of characters in the license plates. 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 are performed automatically by software and require intelligent algorithms to achieve a high reliability. Many systems have been developed. Most of them apply a learning approach; a historical data set is collected and used to build the license plate solution. For the character recognition, the data set is made of character images extracted from the historical data set using the first two steps of the solution, and split into two basic sets, training data and validation data. These are images of license plate characters that have been extracted from images of vehicles. They are binary images that have been cleaned, cropped and normalized. 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 following section presents a solution described in [16].

IV. Example of a Simple Solution

Neural networks [9] have been applied successfully in many recognition and classification problems. The methodology 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 *feedforward* structure, there is a set of input nodes, for example the features and attributes of the image being processed for classification, that are connected through a network of nodes, hidden layers, to a set of output nodes, the classes to which the image belongs. The neural network processes the images 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 neural network reaches a satisfactory level of recognition. The set of final weights is used for processing new data. This method is applied successfully to the license plate recognition problem [17], [18]. The other approach most used in character recognition in the license plate problem 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 its 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 measure in (4), a statistical measure used by Horowitz [19] and Pratt [20] for image recognition, is a metric for template matching. If $f_I(x,y)$ and $f_T(x,y)$, $x = 1, \ldots, n_x$, $y = 1, \ldots, n_y$ are the pixel values of the digitized image and part of the template it is being compared to, respectively, both of size n_x x n_y, then a normalized cross-correlation measure, $\rho(u,v)$, is :

$$\frac{\sum_x \sum_y f_I(x,y) f_T(x-u, y-v)}{[\sum_x \sum_y f_I^2(x,y)]^{1/2} [\sum_x \sum_y f_I^2(x-u,y-v)]^{1/2}} \tag{4}$$

The template matching approach can be combined with other methods in character recognition. However, the approach remains a method based on the principle of the minimization of a distance between two images. The optical character recognition is a problem where 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 [21]. Neural networks provide satisfactory answers to the character recognition problem, but they have drawbacks in speed, complexity and training requirements. Template matching is a minimization of squared error approach that does not provide a probabilistic answer and can be computationally expensive. Probability is the theoretically sound framework in which an uncertainty problem is treated. In the probabilistic framework, a likelihood function is built to model the characteristics of the problem. Upon receipt of data, it is either maximized or inverted to provide the most likely solution. We introduce a simple probabilistic solution to the optical character recognition problem in the license plate problem. But before that, we review the theoretical foundations of the statistical approach.

A. The Laws of Probability

There have been many approaches and tribulations in modeling uncertainty. While *Probability Theory* is regarded as a sound theoretical approach in dealing with uncertainty, there have been other models. *Fuzzy Logic Theory* [22] and Shafer-Dempster Belief Functions Theory [23], [24] are the most notable other directions. Let E be a random event or uncertain event, then probability theory assigns a number between 0 and 1 to that event and denotes it $P(E|\mathcal{H})$. $\mathcal{H}$ stands for 'history' and denotes all the background history and information held by the assessor of that probability. $\mathcal{H}$ is often assumed and not shown in the notation, and we write $P(E)$ for the probability of event E. E can be any possible or imagined event, such as 'rain tomorrow', or the 'image belonging to class C'. E represents the truth of a proposition and $P(E)$ the probability of that proposition being true. There are three laws upon which probability theory is built:

- Convexity: $0 \leq P(E) \leq 1$
- Addition: if E_1 and E_2 are mutually exclusive, that is they both cannot occur together, then
$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$$
- Multiplication: $P(E_1 \text{ and } E_2) = P(E_1|E_2)P(E_2)$, where $P(E_1|E_2)$ is known as the *conditional probability* of event $E1$ 'given' event E_2 has occurred.

Based on these simple laws, a whole body of explanatory science, inference and prediction using statistics is built. While probability theory is laid down with mathematical rigor and require definitions upon which the three laws are built, in essence, it is these 3 simple laws that are at the heart of any probabilistic operation, and any statistical study. Two other laws are derived from the three basic ones, and form the engines for inference and prediction. They are the 'Law of Total Probability' and 'Bayes' Law [32]. With a proper interpretation of the meaning of probability $P(E)$, this is all that is needed to conduct a complex data analysis.

B. The Probabilistic Approach

Returning to the optical character recognition, let an extracted character image be at hand. We want to recognize the character C in that image, C being a letter or a one digit number. Let Z be a random variable that represents a selected statistical feature of the 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 along with engineering knowledge are used 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

$$\text{Prob}(C|Z = z) = \frac{\text{Prob}(Z = z|C)\text{Prob}(C)}{\sum_{S=A}^{S=9} \text{Prob}(Z = z|S)\text{Prob}(S)} \quad (5)$$

for $C = A, B, \ldots, 9$. $\text{Prob}(C)$ in (5) is the prior probability that C is the actual character in the analyzed image. The denominator $\delta = \sum_{S=A}^{S=9} \text{Prob}(Z = z|S)\text{Prob}(S)$ is a normalizing constant. $\text{Prob}(C|Z = z)$ is the resulting posterior distribution of C. The probabilistic solution chooses that character $\hat{C}$, that maximizes the posterior distribution:

$$\text{Prob}(\hat{C}|Z = z) = \mathbf{Max}_{C=A,B,\ldots,9}\ \text{Prob}(C|Z = z). \quad (6)$$

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:

$$\text{Prob}(Z = z|\hat{C}) = \mathbf{Max}_{C=A,B,\ldots,9}\ \text{Prob}(Z = z|C). \quad (7)$$

Both these statistical solutions, (6) and (7), rely on the likelihood function. These simple operations are at the heart of many probabilistic classifications, predictions and inferences. While seemingly simple in principle, their success depends on the proper selection of the random variable Z and the probability model $\text{Prob}(Z|C)$ in (5).

C. Recognition of Characters

For the license plate recognition problem, Z is defined in [16] as the vector of the values of all the pixels in the extracted image, after such image has been cleaned, cropped and normalized. The author in [16] arrived at the conclusion that the values of the pixels in the binary image hold all the information that can be used to recognize the character in the extracted image. Each of these pixel values is either 0 or 1, the input image being binary. 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 follows as

$$\mathcal{L}(C) = \text{Prob}(Z|C) = \prod_{i=1}^{|Z|} \theta_i^{Z_i}(1 - \theta_i)^{1 - Z_i}, \quad (8)$$

where $|Z|$ is the cardinal, or vector size, of Z. To estimate the proportion θ_i for pixel i, a number of approaches can be

used. But given that the sizes of the training sets are relatively large, the average is used as an estimate

$$\hat{\theta}_i = \sum_{j=1}^{N_C} x_{i,j}/N_C \quad (9)$$

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, θ_i is used in (8), when in fact it is $\theta_i(C)$. It differs for each character C. The same holds for Eq. (9). 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. 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

$$\text{Prob}(C|z) = \frac{\text{Prob}(z|C)\text{Prob}(C)}{\sum_{S=A}^{S=9} \text{Prob}(z|S)\text{Prob}(S)} \quad (10)$$

where

$$\text{Prob}(z|C) = \prod_{i=1}^{|Z|} \hat{\theta}_i^{z_i}(1 - \hat{\theta}_i)^{1 - z_i} \quad (11)$$

noting that the appropriate $\hat{\theta}_i$'s in (10) correspond to a given C. The posterior probability distribution in (11) 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 yielded excellent results in [16]. As such, the method fails to distinguish fully between the 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 [25] reach a 97% reliability . This is a high reliability for a pure approach that doesn't use heuristics and ad-hoc techniques. In practice, one would supplement this approach with redundancy checks in the form of different solution methodologies, and use multiple images per vehicle to obtain close to 100% accuracy. As such, the method is highly accurate. In essence, the approach of [16] is a Naive Bayes approach, as known in the classification literature, and provides a probabilistic solution to the optical character recognition in the license plate problem.

V. Statistical Foundations of Intelligent Technologies

Debates about how neural networks fit into classification, and how they compare and relate to statistical methods have been conducted (see for example [26], [27]). While some have adopted an outright stance against neural networks [28], most accept that there are two philosophically different approaches in data analysis, one based on the laws of probability and the other algorithmic and data driven [10]. Many use both approaches in their dealing with data, particularly engineers who are pragmatic and care that a method works and be robust. While support vector machines are optimization tools used to separate data, and as such aren't subject to criticism, neural networks on the other hand are data analysis tools that cannot be justified on the ground of any theory. They were loosely designed to mimic the functioning of the human brain, and do not represent any logical or axiomatic approach in dealing with the data, besides the empirical facts that they have been successful in some problems. Their outputs are not probabilities, only estimates of probabilities. As such, neural networks cannot be used in decision theoretic problems. They achieve their full potential in problems only under ideal conditions. Their training is often complicated and time consuming. There is the issue of 'over fitting'. By over fitting the neural network to the training data, one risks to extrapolate badly, that is to have a poor neural network performance on the new data. The consistency of the estimators in neural networks can be proven only numerically, as posterior probabilities are only estimated. Finally, neural networks assume equal costs for misclassification, which may not be true in practice. Neural networks are nevertheless universal approximators. They are model free, data driven, which explains a large part in their success. They help avoid the curse of dimensionality in many problems and are robust to noise. Although often processed by serial computers, neural networks do offer, in principle, the ability to use parallel processing.

If any criticism is to be directed, it should be to the statistical community who failed to address the need of the data analysis community. One can state that some statisticians produce work where they do not justify their models, focusing rather on the use of such models in sophisticated statistical solutions, while on the other hand engineers, physicist and computer scientists struggle with large data sets to explain phenomena observed in the real world. The data algorithmic approach may be seen as a way by some communities to eliminate their needs for statisticians and their models. Data analysts and engineers often require answers to large data sets without dwelling indefinitely on the explanation of such data. However, there is danger in thinking that any answer is a good answer. The laws of probability are not a truth. They form only a model of thinking, based on axioms. But it is the result of centuries, of building bit by bit a theory for dealing with uncertainty. While there is absolutely no guarantee that it is the optimal way to think, it certainly has been tested by many great minds over centuries. It is not in the scope of this article to survey the foundations of probability, but many authors have provided excellent texts on the subject [29]–[32]. And it is certainly accepted universally that probability is the only coherent way to solve decision problems. But what are data analysis problems but decision problems.

References

[1] U.S. Department of Transportation, Federal Highway Administration, "National ITS Architecture and Standards Final Rule", http://ops.fhwa.dot.gov/its_arch_imp/index.htm
[2] Intelligent Transportation Society of America, http://www.itsa.org/
[3] ERTICO - ITS Europe, http://www.ertico.com/
[4] ITS Japan, http://www.its-jp.org/english/index.htm
[5] ITS Strategy Committee, "ITS Strategy in Japan", Report of the ITS Strategy Committee, ITS Japan, Summary version, http://www.its-jp.org/english/topics_e/doc/strategy_e.pdf
[6] ITS Australia, http://www.its-australia.com.au/kmxserver3/
[7] M. Bach Cuadra, L. Cammoun, T. Butz, O. Cuisenaire and J. Thiran, "Comparison and Validation of Tissue Modelization and Statistical Classification Methods in T1-weighted MR Brain Images", IEEE Transactions on Medical Imaging, Vol. 24(12), 1548-1565, 2005.
[8] N. Godin, S. Huguet and R. Gaertner, "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 38 (4): 299-309, 2005.
[9] D. Patterson, *Artificial Neural Networks*. Singapore: Prentice Hall, 1996.
[10] L. Breiman, "Statistical Modeling: The Two Cultures", Statistical Science, Vol. 16, No. 3, 199231, 2001.
[11] J.A. Nystuen and H.D. Selsor, "Weather classification using passive acoustic drifters", J. Atmos. and Oceanic Tech., 14, 656-666, 1997.
[12] P. Brazdil, J. Gama and B. Henery, "Characterizing the applicability of classification algorithms using metalevel learning", in Proceedings of the European Conference on Machine Learning, pp. 83-102, 1994.
[13] C.M. van der Walt and E. Barnard,"Data characteristics that determine classifier performance", in Proceedings of the Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa, pp.160-165, 2006.
[14] V. Vapnik, *Estimation of Dependences Based on Empirical Data* [in Russian]. Nauka, Moscow, 1979. (English translation: Springer Verlag, New York, 1982).
[15] C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and knowledge Discovery, 2:121-167, 1998.
[16] K. Aboura, "Probabilistic optical character recognition for license plate identification", Computer Systems Engineering Department, University of Technology, Sydney, Australia, Tech. rep., 2007.
[17] M.M.M. Fahmy, "Automatic number-plate recognition : neural network approach", in *Proc. of VNIS94 Vehicle Navigation and Information System Conference*, Aug-Sept 1994, pp. 99-101.
[18] W. Wei, Y. Li, M. Wang and Z. Huang, "Research on number-plate recognition based on neural networks", in *Proc. 2001 IEEE Signal Processing Society Workshop, Neural Networks for Signal Processing XI*, 2001, pp. 529-538.
[19] M.Horowitz, "Efficient use of a picture correlator", *J. Opt. Soc. Am*, vol. 47, pp. 327, 1957.
[20] W. K. Pratt, "Correlation Techniques of image Registration", *IEEE Trans.*, vol AES-10, pp. 353-358, 1974.
[21] D.V. Lindley, "The probability approach to the treatment of uncertainty in artificial intelligence and expert systems", *Statistical Science*, vol. 2, no. 1, pp. 17-24, Feb. 1987.
[22] L. Zadeh, "Fuzzy Sets", Information and Control, vol. 8, pp. 338-353, 1965
[23] G. Shafer, *A mathematical theorey of evidence*, Princeton, University Press, Princeton, NJ, 1976
[24] A.P. Dempster, "A generalization of Bayesian Inference", Journal of the Royal Statistical Society B, vol. 30, pp.- 205-247, 1968
[25] K. Aboura and R. Al-Hmouz, "Probabilistic License Plate Optical Character Recognition". Technical Report, Computer Systems Engineering Department, University of Technology Sydney, Australia, 2007.
[26] B. D. Ripley, "Neural Networks and Related Methods for Classification", JRSS, Series B, 1993.
[27] Guoqiang Peter Zhang, "Neural Networks for Classsification: A Survey", IEEE Trans. Cybernetics, 2000.
[28] A.K. Dewdney, *Yes, we have no neutrons: An eye-opening tour through the twists and turns of bad science*, NY:Wiley, 1997.
[29] L. J. Savage, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954.
[30] H. M. Degroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.
[31] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, 2 edition, Mar 25 1993.
[32] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, Wiley, April 24, 2000
[33] D.V. Lindley, *The Bayesian Approach to Statistics*, In Some Recent Advances in Statisticsp, eds. J.T. de Oliveria and B. Epstein, Acdemic Press, London, pp. 65-87
[34] D.V. Lindley, "Scoring rules and the inevitability of probability", Internattional Statistical Review, vol. 50, pp. 1-26

The Eighth International Conference on Intelligent Technologies (InTech'07)
Sydney, Australia, 12-14 December 2007

A Hybrid Evolutionary Approach for Heterogeneous Multiprocessor scheduling

K. C. Tan, C. K. Goh, E. J. Teoh and D. K. Liu

***Abstract*—This paper considers the assignment of tasks with interdependencies in a heterogeneous multiprocessor environment where task execution time varies with task as well as the processing element processing it. The solution to this heterogeneous multiprocessor scheduling problem involves the optimization of complete task assignments and processing order within the assigned processors with minimum makespan, subject to the precedence constraint. To solve such a NP-hard combinatorial optimization problem, this paper presents a hybrid evolutionary algorithm that incorporates two local search heuristics that exploits the intrinsic structure of the solution as well as specialized genetic operators to encourage exploration of the search space. The effectiveness and contribution of the proposed features are validated on a set of benchmark problems characterized by different degrees of communication times, task and processor heterogeneities. Simulation results demonstrate the algorithm is capable of finding useful schedules on the set of new benchmark problems.**

***Index Terms*—Multiprocessor scheduling, heterogeneous, hybrid evolutionary algorithm, local search, precedence**

I. Introduction

THE emergence of computer programs with increasingly higher computational requirements and algorithmic complexity has necessitated the need for parallel processing elements in a multi-computer environment, which in turn has seen the increasing need for task allocation to be optimally distributed in a suitable manner to these individual processing units. The multiprocessor scheduling problem is a class of NP-hard combinatorial optimization problems [13], [17], [21], [26] and it can be categorized into different classes based on the characteristics of the program, the tasks to be scheduled, the multiprocessor system, as well as the availability of information [19], [20], [9]. Typically the processing elements constituting the multi-computer environment can be of the same capability or of a different capability. This paper is focused on the latter. The problem becomes even more challenging when communication delays are accounted for.

In a nutshell, the goal of a scheduler is to assign partitioned tasks to available processors in such a manner that not only are the requirements of precedence between these tasks are met, but also in addition to the objective in obtaining as minimal as possible a makespan [29]. Presently, there are numerous methods and approaches which have been developed and subsequently applied to the multiprocessor scheduling problem, typically using a deterministic approach. El-Rewini *et al* [9] provides a fairly comprehensive taxonomy of how scheduling problems can be categorized, and highlights the key differences that distinguishes one class from the next. Further to this, in [19], [20], Kwok and Ahmad present a wide-ranging overview and classification of scheduling algorithms, particularly focusing on deterministic and static scheduling problems. Most of the present techniques are based on heuristics [18], [22] that are not only greedy in nature but also capable of solving certain instances of the scheduling problem efficiently.

Evolutionary algorithm (EA) is a class of stochastic global optimization technique has been applied to solve the heterogenous multi-processor scheduling (HMPS) optimization problem. EAs are excellent global search algorithms but they can take a relatively long time to locate the local optimum in the region of convergence [25]. On the other hand, local search heuristics are capable of locating the optimum quickly but are prone to local optimal traps. Therefore, researcher often hybridized EAs with local search heuristics to maintain a balance between exploration and exploitation to improve the optimization processes [6], [12], [16], [23], [24].

This paper presents a new hybrid evolutionary algorithm (HEA) for the HMPS optimization problem. The proposed algorithm incorporates two local search operators, based on list scheduling and task duplication, to exploit the intrinsic structure of the scheduling problem. Unlike existing evolutionary approaches to the HMPS problem, the proposed HEA also implements a variable length chromosome which preserves the precedence relations, a PE schedule crossover which facilitates the exchange of good schedules assigned to the individual processors as well as specialized mutation operators to improve the diversity of the evolving population.

This paper is organized as follows: Section II gives an overview of existing works as well as the problem formulation of the HMPS. Section III presents the various features of the proposed HEA including the local search heuristics and specialized genetic operators as well as the algorithmic flow. Section IV presents the extensive simulation results and analysis of the proposed algorithm. Conclusions are drawn in Section V.

II. Heterogenous Multi-processor Scheduling Problem

A. Problem Formulation

The multiprocessor scheduling problem can be simply stated as follows:

> Assuming there are n tasks that have to be executed on m processors - where and when should each task

K. C. Tan, C. K. Goh and E. J. Teoh are with the Department of Electrical and Computer Engineering, National University of Singapore, 4, Engineering Drive 3, Singapore 117576 (e-mail: {ckgoh,eletankc}@nus.edu.sg). D. K. Liu is with the ARC Centre of Excellence for Autonomous Systems (CAS), Faculty of Engineering, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia

W2A-4	Strong Laws of Large Numbers of Exchangeable Fuzzy Set-Valued Random Variables *Xiaoqin Han, Shoumei Li* Beijing University of Technology, China	104
W2B	**Intelligent Sensing**	
W2B-1	Locating and Docking Method in Self-reconfiguration of Microrobots Based on Infrared Sensor *Jianzhong Wang, Jiadong Shi, Jingjing Liu* Beijing Institute of Technology, China	111
W2B-2	A Probabilistic Image Sensor Model for Recursive Bayesian Search and Tracking *Shen Hin Lim, Tomonari Furukawa* University of New South Wales, Australia	116
W2B-3	Simultaneous Material Type Classification and Mapping Data Acquisition using a Laser Range Finder *N. Kirchner, T. Taha, D. K. Liu, G.Paul* University of Technology Sydney, Australia	124
W2B-4	Capacitive Object Ranging and Material Type Classifying Sensor *N. Kirchner, D.K. Liu, T. Taha, G.Paul* University of Technology Sydney, Australia	130
T1A	**Mathematical Foundations of Intelligent Technologies**	
T1A-1	Asymmetric Information Measures: How to Extract Knowledge From an Expert So That the Expert's Effort is Minimal *Hung T. Nguyen,* New Mexico State University, USA *Vladik Kreinovich, Elizabeth Kamoroff,* University of Texas at El Paso, USA	136
T1A-2	How to Avoid Gerrymandering: A New Algorithmic Solution *Gregory B. Lush, Esteban Gamez, Vladik Kreinovich* University of Texas at El Paso, USA	146
T1A-3	Spiral Curriculum: Towards Mathematical Foundations *Olga Kosheleva,* University of Texas at El Paso, USA	152

T1A-4	Finite Clause-Length Theorem for Deterministic-Rule Programs on Specialization Systems *Kiyoshi Akama,* Hokkaido University, Japan *Ekawit Nantajeewarawat,* Thammasat University, Thailand *Toshihiro Wakatsuki*, Hokkaido University, Japan	158
T1B	**Intelligent Technologies in Automation**	
T1B-1	Wind Farm Power Curve Modeling Using Adaptive Neuro-Fuzzy Inference Systems *P. Johnson, M. Negnevitsky* The University of Tasmania, Australia	168
T1B-2	Takagi-Sugeno Fuzzy Modelling of Multivariable Nonlinear System via Genetic Algorithms *Haiping Du, Nong Zhang* University of Technology Sydney, Australia	176
T1B-3	Asset Management of Stormwater System using Fuzzy Logic *A. Kannapiran[1], J. Jeyakumaran[1], A. Chanan[2], J. Kandasamy[1], G. Singh[2], P. Tambosis[2], A. Al-Jumaily[1]* [1] University of Technology Sydney, Australia [2] Assets and Services, Kogarah Council, Sydney Australia	182
T1B-4	Artificial Intelligent Approach to Predict Arc Furnace Response *M. O. Haruni, Kashem M. Muttaqi, M. Negnevitsky* The University of Tasmania, Australia	189
T2A	**Business and Finance Management**	
T2A-1	Optimizing Vendor Selection for Information Systems Outsourcing under Uncertainty *Hepu Deng, Alemayehu Molla* RMIT University, Australia	195
T2A-2	Exploring the use of RFID in Emergency Management *Ashir Ahmed, Ly-Fie Sugianto* Monash University, Australia	202
T2A-3	A Comprehensive Analytical Model for Video over IP in Telecollaboration Business System Environments *V. Mahadevan, Z. Chaczko, J. Agbinya* University of Technology Sydney, Australia	212

Foreword

The Eighth International Conference on Intelligent Technologies (InTech'07), to be held in Sydney from 12 to 14 December 2007, is hosted by University of Technology Sydney (UTS) and the ARC Centre of Excellence for Autonomous Systems (CAS). It is the first time this annual conference takes place in Australia.

Previous InTech conferences have been the venue for internationally-recognised researchers in applied mathematics and artificial intelligence producing excellent work in the areas of Fuzzy Logic, Knowledge-Based Systems, Neural Networks, Learning/Adaptive Systems, Data Mining, Uncertainty Processing and Soft Computing, Intelligent Data Analysis, Control and Decision Science.

InTech'07 merges its tradition in laying mathematical foundations of intelligent technologies with the conference theme on "*Intelligent Technologies in Robotics and Automation*", reflecting an increasing interest in real-world applications. The conference's technical programme covers also recent advances in information technologies, business and finance management, biomedical engineering, control and power engineering, sensing, vision and image processing.

The International Programme Committee and invited reviewers were working very hard to return all paper reviews in time, with each paper being peer-reviewed at least by two specialists in the relevant area. For this, we are indebted to all colleagues involved. Out of over 70 paper submissions and interest registrations, the Committee has accepted 40 regular papers for presentation and 3 poster papers. In addition to ten regular sessions, InTech'07 features also four plenary talks, two invited papers, a Gala dinner speech, and an expert panel discussion.

The generous sponsorship obtained from the UTS Faculty of Engineering and the UTS node of the Centre for Autonomous Systems is gratefully acknowledged. We would like to sincerely thank our colleagues from University of Texas at El Paso, USA, and from the InTech founding institution, Assumption University, Thailand, for their invaluable cooperation. Many thanks go also to staff and students at the host institution, who have helped and supported the conference in one way or another. Lastly, we would like to express our heartfelt thanks to all the authors, participants and esteemed guests for their solid contributions and making for this unique event.

We hope InTech'07 will be successful, fruitful to all of you, and would like to wish you a joyful stay in Sydney.

Assoc. Prof. Q. P. Ha
InTech'07 General Chair
UTS-CAS, Nov. 2007

Panel Discussion

The International Conference on Intelligent Technologies (InTech) has traditionally had a strong focus on the development of mathematical foundations for Intelligent Technologies. Recent InTech conferences have been the venue for internationally-recognised researchers in applied mathematics and artificial intelligence producing excellent work in the areas of Fuzzy Logic, Knowledge-Based Systems, Neural Networks, Learning/Adaptive Systems/Data Mining, Uncertainty Processing and Soft Computing, Intelligent Data Analysis, Control and Decision Science.

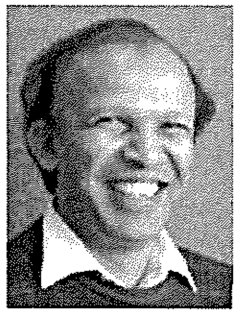
InTech'07 is complementing these with intelligent technology developments for applications in information technologies, business and finance management, biomedical engineering, control and power engineering, sensing, image processing and the conference mainstream of robotics/mechatronics and automation. Towards productive amalgamation of sciences and applications in real world deployment of intelligent technologies, a Panel Discussion will be organised on the theme:

Intelligent Technologies: Bridging the Gap between Sciences and Applications

Facilitator:

Professor Gamini Dissanayake, University of Technology Sydney, Australia
Centre for Autonomous Systems, UTS Node Director

Biography:



Gamini Dissanayake is the James N Kirby Professor of Mechanical and Mechatronic Engineering at University of Technology, Sydney (UTS). He has expertise in a broad range of topics in robotics including; robot localisation, mapping and simultaneous localization and mapping (SLAM) using sensors such as laser, radar, vision and inertial measurement units; terrain mapping; multi-robot coordination for SLAM, target tracking and probabilistic search; motion planning for single and multiple robot manipulators, legged robots, and cranes; and application of robotic systems in urban search and rescue. He leads the UTS node of the ARC Centre of Excellence for Autonomous Systems. He graduated in Mechanical/Production Engineering from the University of Peradeniya, Sri Lanka in 1977. He received his M.Sc. in Machine Tool Technology and Ph.D. in Mechanical Engineering (Robotics) from the University of Birmingham, England in 1981 and 1985 respectively.

Proceedings of the Eighth International Conference on Intelligent Systems (InTech'07)

Sydney Australia 12-14 December 2007

University of Technology Sydney

ISBN 978-1-86365-718-1