Abstract: The paper presents application of neural network in a hybrid, high speed, pattern recognition system. The feature extraction part is built as a grating based holographic ring wedge detector and the classifier is a probabilistic neural network. Since the feature extractor can be produced with relatively low costs from computer generated high resolution masks, such masks should be designed specifically to given recognition task. This requires automatic knowledge acquisition and processing with the goal of optimization of the feature space dedicated for subsequent use of neural network classifier. Appropriate methodology, proposed by the author in earlier works, has been enhanced by novel author’s modification of the notion of indiscernibility relation in theory of rough sets. New, generalized version of this relation, makes possible natural application of discrete type of rough knowledge representation into problems operating in continuous space and therefore using, like neural networks do, real valued data.

Key-Words: hybrid systems, neural networks, rough sets, holographic ring-wedge detectors, image processing, pattern recognition

1 Introduction

The automatic recognition of images constitutes one of the major areas in the pattern recognition problems, and is one of prominent application of machine learning methods. This field has focused attention of many researchers and companies trying to build system which is reliable and at the same time fast and easily adaptable. The prominent role among such systems play hybrid solutions. They are designed to perform heavy computations in optical mode, practically contributing no time delays. The post-processing of optical results, is made in computers, often with the use of machine learning methods, the artificial neural networks (ANN) being one of them.

Various hybrid solutions have different characteristics. Here we present enhancements in the system capable to recognize input images invariantly with respect to size, rotation and shift. The foundations of the system have been proposed by Casasent and Song, presenting the design of holographic ring wedge detectors (HRWD) [1], and George and Wang, who applied commercially available ring wedge-detector (RWD) and neural network to build the complete recognition system [2]. Despite the completeness of their solution the system was lacking possibility of adaptation due to application of uniformed and expansive RWD. Casasent’s HRWD, originally named by him as a computer generated hologram (CGH) seemed to have a lot of advantages over commercial RWD. One of more important were the lower cost and possibility to produce such elements fitted specifically to given application. According to optical characteristics the HRWD belongs to a wider class of grating based diffractive optical variable devices (DOVDs). The pioneering works proposing the methodology of optimization of HRWD masks has been proposed by Cyran and Mrózek [3] and by Jaroszewicz et al. [4]. This methodology was based on theory of rough sets, originated by Pawlak [5] and further developed by (among others) Mrózek [6, 7] Ziarko [8] and Skowron and Grzymala-Busse [9].

The initial method of optimization of HRWD was successfully applied to ANN-based system used in recognition of the type of subsurface stress in materials with embedded optical fiber [10 12]. Another examples of application of the HRWD-based feature extraction together with ANN-based classification include the systems designed by Podeszwa et al. [13] devoted for the monitoring of the engine condition, and by Jaroszewicz et al. [14] dedicated for airplane engines. There is also possible to implement purely optical version of this recognition system considered by Cyran and Jaroszewicz [15]. However, this fully concurrent system is limited by the development of technology of optical neural networks, due to difficulties in obtaining non linear activation function in an optically implemented artificial neuron.

One should also note that ANN is not the only classifier which could be applied for classification of patterns occurring in a feature space generated by HRWD. Moreover, the first version of optimization
procedure favored the rough set based classifiers, due to identical (and therefore fully compatible) discrete nature of knowledge representation in the theory of rough sets applied both to HRWD optimization and to subsequent rough set based classification. The application of general ideas of obtaining such rough classifier was presented by Cyran and Jaroszewicz [16] and the fast rough clas-sifier implemented as programming logic device (PLD) was considered and designed by Cyran [17].

Even, if the optimization was performed in a discrete feature space, as required by the application of a rough classifier, nevertheless, the resulting system was suboptimal. This suboptimality was a consequence of the fact that the feature space optically generated by HRWD is always continuous, and therefore it had to be discretized if rough set based machine learning methods were to be used. Natural enhancement would have been possible if both, continuous type of classifier and optimization procedure working in a continuous space, were applied. The first postulate have been easily materialized by application of a probabilistic neural network classifier, but the second demanded novel author’s modification of the indiscernibility relation in rough set theory, allowing natural processing of real valued attribute values. Such novel modification improved the results of evolutionary optimization of holographic RWD and equivalently, enhanced the optimization of the feature space generated by HRWD, and dedicated for continuous-valued ANN classifiers. This modification made possible to avoid, highly non linear transformation of separate features corresponding to conditional attributes in rough set theory. Such transformation, being non natural in our problem, was originally required by rough set based objective function used in the optimization procedure. The paper explains the novel author’s idea of generalization of indiscernibility relation notion in the section 2. In the section 3 the application of proposed modification into high speed HRWD-ANN based pattern recognition system is presented. Section 3 starts with foundations of the system considered, and it is followed by experimental results comparing the enhanced methodology, with that published before. The discussion and conclusions are included in section 4)

2 Modification of indiscernibility relation in rough set theory

In order to avoid the need of independent discretization of individual features when calculating the rough set based objective function, defined as the consistency measure of decision table, we propose to modify the meaning of the indiscernibility relation in rough set theory. The indiscernibility relation plays crucial role in methodology of machine learning described by rough set formalism. Our detailed analysis of notions present in the theory of rough sets proved that vast majority of them do not require the specific, classical form of this relation. The only real demand is that this relation should be an equivalence relation, i.e. it should be reflexive, symmetric and transitive.

In particular this is true for the coefficient referred to as the consistency measure of a decision table. This coefficient has been used as the objective function in evolutionary optimization of HRWD for multimodal distribution of classes in a feature space. The motivations supporting such criterion have been considered in [3]. They seem to be reasonable, both from theoretical and experimental perspective, but assuming classical definition of indiscernibility relation, the result was always sub-optimal. Therefore we propose to change this definition to a more general version, which defines two objects as indiscernible in the sense of rough sets theory, if they belong to the same clusters in a continuous space (or subspaces) of real-valued attributes.

Formally, Let $S=\langle U, Q, v, f,\rangle$ be the information system composed of universe $U$, set of attributes $Q$, information function $f$, and a mapping $v$. This latter mapping associates each attribute $q \in Q$ with its domain $V_q$. The information function $f: U \times Q \rightarrow V$ is defined in such a way, that $f(x,q)$ denotes the value of attribute $q$ for the element $x \in U$, and $V$ is a domain of all attributes $q \in Q$, defined as a union of all domains of single attributes, i.e. $V = U_q \in Q, V_q$. Then each nonempty set of attributes $C \subseteq Q$ defines the classical indiscernibility relation $I_0(C) \subseteq U \times U$ of discrete attributes $q \in C$, as

$$x \ I_0(C) \ y \iff \forall q \in C, \ f(x,q) = f(y,q)$$

(1)

where $x, y \in U$. If we originally have real valued attributes, then before application of rough set theory some clustering and discretization of continuous values of attributes has to be performed. Let this process be denoted as a function $A: \mathbb{R} \rightarrow \{1, 2, \ldots, \xi\}$

Where $\xi$ is the number of clusters covering the domain of the individual attributes $q \in C$. Furthermore, let discretization of any individual attribute $q \in C$ be denoted as a scalar function

$$A: \mathbb{R} \rightarrow \{1, 2, \ldots, \xi\}$$

In this case we obtain the classical form of indiscernibility relation defined as:

$$x \ I_0(A[c]) \ y \iff \forall q \in C, \ f(x,A[q]) = f(y,A[q])$$

(2)

As was already stated, the above form of the indiscernibility relation, proposed by classical theory of rough sets, as well as by its generalization named variable precision model, is not actually required for rough set based machine learning. To introduce formally
the modification, let us change the notation of indiscernibility relation from classical form dependent on unstructured set of attributes to generalized version dependent on a family (set) of sets of attributes. This allows to introduce hierarchy of sets into originally unstructured set of attributes that the relation depends on.

Let $\mathcal{C} = \{ C_1, C_2, \ldots, C_N \}$ be a family of disjoint sets of attributes $C_n \subseteq \mathcal{Q}$ such that unstructured set of attributes $\mathcal{C} \subseteq \mathcal{Q}$ is equal to the union of elements (each element is a set of attributes) of the family $\mathcal{C}$, i.e. $\mathcal{C} = \bigcup_{C \in \mathcal{C}} C_n$. Then, let the indiscernibility relation be dependent on $\mathcal{C}$ instead of being depent on $C$. Observe that both $\mathcal{C}$ and $C$ contain the same collection of single attributes, however $\mathcal{C}$ includes additional structure as compared to $C$. If this structure is irrelevant for the problem considered, it can be simply ignored and we can obtain, as a special case, the classical version of indiscernibility relation $I_0$. However we can also obtain other versions of this modified relation for which the introduced structure is meaningful.

Formally, let modified indiscernibility relation $I_1(\mathcal{C}) \in U \times U$ be defined as

$$x \ I_1(\mathcal{C}) \ y \iff \forall C_n \in \mathcal{C}, \ Clus(x, C_n) = Clus(y, C_n)$$

where $x, y \in U$, and $Clus(x, C_n)$ denotes the number of the cluster, that the element $x$ belongs to. The cluster analysis is therefore required to be performed in a continuous vector spaces defined by sets of real valued conditional attributes $C_n \in \mathcal{C}$. There are two extreme cases of this relation, obtained when family $\mathcal{C}$ is composed of exactly one set of conditional attributes $C$, and when family $\mathcal{C}$ is composed of $\text{card}(C)$ sets, each containing exactly one conditional attribute $q \in C$.

3 Application into Optimization of HRWD-ANN based System

The first subsection of this application-oriented section deals with definition of the objective function in the optimization of holographic ring wedge detector. The evolutionary optimization and experimental results are finally included in a following two subsections.

3.1 Definition of Objective Function

The choice of the objective function in optimization of feature extraction is simple when all classes can be represented by single clusters. The distance between clusters can be used as an objective function in such cases. In more complex problems however, one class can be represented by points, which not necessarily form single cluster in a feature space. These are so called multimodal problems, and we consider an example of such problem. Then the definition of the good objective function is not a trivial task. In our previous work [3, 4] we addressed the problem by proposing the objective function as coefficient defined in the theory of rough sets. More specifically it was the consistency measure $\gamma_c(D^*)$ of the decision table with conditional attributes corresponding to rings and wedges of HRWD and the decision attribute being the class of the image.

Such defined objective, when maximized, led to the feature space optimal for the rough classifiers, however for ANN-based classifiers the feature space was suboptimal. The reason for this was that to calculate the objective function one had discrete light intensities independently on each other, as demanded by original definition of indiscernibility relation in rough set theory. Now, we propose to use the same criterion, however computed for novel version of indiscernibility relation defined by formula (3).

3.2 Evolutionary Optimization of HRWD

Since the defined above enhanced objective function is not differentiable, gradient-based search method should be excluded. However the HRWD can be optimized in a framework of evolutionary algorithm. As genetic operations classical one-point recombination and uniform mutation have been used. The selection was proportional, however in elitist model propagating best solution from generation to generation with probability 1.

The algorithm has two flow control parameters: MaxGenNum (specifying maximum number of epochs for evolution) and MaxValue, indicating the maximum required value of the objective function. Normally MaxValue should be set to 1 – to obtain fully consistent decision table but sometimes this could be to strong demand to fulfill – then one should reduce this parameter.

This algorithm is very similar to that applied in the case of objective function calculated from classical definition of discernibility relation. The difference is in the meaning of $\xi$ parameter. Previously it was the discretization factor required by rough set theory, now it is the number of clusters in clustering procedure. This change influences the initial value of $\xi$ and the termination of presented program. The initial value of $\xi$ now is calculated as $2^Q$ for such minimum $Q$ for which $\xi \geq \text{Card}(U)$. The program is terminated after achieving the maximum value of $\gamma_c(D^*) = \text{MaxValue}$ for $\xi = \text{NumOfClasses}$ (NumOfClasses denotes the number of classes to be recognized), as opposed to previous version, terminating when $\gamma_c(D^*) = \text{MaxValue}$, for $\xi = 2$. Another difference is that in the above algorithm
the function $\chi$ denotes the feature extraction, while in previous version it denoted the feature extraction with discretization, and explicitly the clusterization is invoked. As the result of operation of the algorithm the parameters describing optimized HRWD are obtained (they are encoded in chromosome $x_{opt}$).

### 3.3 Results

We verified the recognition abilities of presented system using PNN for classification of speckle structure images coming from the output of the optical fiber.

The experiments were conducted for a set of 128 images of speckle patterns generated by intermodal interference occurring in optical fiber and belonging to eight classes taken in 16 sessions $S_l$ ($l = 1, \ldots, 16$). The Fraunhofer diffraction patterns of input images were obtained by calculating the intensity patterns from discrete Fourier transform. The training set consisted of 120 images taken out in 15 sessions and testing set contained 8 images belonging to different classes, representing one session $S_l$. The process of training and testing was performed 16 times, according to jack-knife method, i.e. for each iteration another session was used for testing set, and all but one sessions were used for training set. That gave the base for reliable cross-validation with still reasonable number of images used for training. The time course of evolutionary optimization is given in Fig. 1. The results of testing of PNN to classify images in feature space obtained from standard, optimized and optimized with modified indiscernibility relation HRWDs, are presented in table 1. More detailed results of jack-knife tests are presented in table 2.

#### Table 1. Results of testing the classification abilities of the system. The classifier is a PNN having Gaussian radial function with standard deviation $\sigma = 0.125$.

<table>
<thead>
<tr>
<th></th>
<th>Correct decisions [%]</th>
<th>Normalized decision error [%]</th>
<th>Improvement [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard HRWD</td>
<td>84.4</td>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>HRWD optimized with standard indiscernibility relation</td>
<td>87.5</td>
<td>1.6</td>
<td>20</td>
</tr>
<tr>
<td>HRWD optimized with modified indiscernibility relation</td>
<td>88.2</td>
<td>1.5</td>
<td>25</td>
</tr>
</tbody>
</table>

#### Table 2. Results of PNN testing for different HRWD-generated feature spaces

<table>
<thead>
<tr>
<th>session number:</th>
<th>0</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bad decisions</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Standard HRWD</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRWD optimized with standard indiscernibility relation</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>HRWD optimized with modified indiscernibility relation</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

The normalized decision errors ranging from 1.5 to 2 percent indicate overall good recognition abilities of the system considered. The 20% reduction of this error is obtained by optimization of HRWD. Further 5% error reduction is caused by modification of indiscernibility relation according to (3).

#### 4 Conclusions

The paper presents author’s original modification of the indiscernibility relation used in the theory of rough sets. This theory (and the generalization called variable precision model) has been successfully applied in many machine learning problems. However, it is well known drawback of this theory, that it deals with continuous attributes in an unnatural way. To overcome this...
disadvantage, the generalization of indiscernibility relation has been proposed (3). This generalization introduce the structure into originally unstructured collection of attributes that the relation depends on. The classical relation is the special case of the modified version, therefore this modification can be viewed as more general. What is particularly important, the generalization is equivalently valid for classical theory of rough sets, as well as for the variable precision model, predominantly used in machine learning of huge data sets.

Proposed in the paper generalization of indiscernibility relation, introduces the flexibility in applying particular special case of it to given application. In the case of real-valued attributes, our modification allows for performing multidimensional cluster analysis, contrary to one-dimensional analyses required by classical form. In majority of cases the cluster analysis should be performed in a space generated by all attributes. This corresponds to a family $C$ composed of one set containing all conditional attributes, and is the opposite extreme case compared to classical relation, assuming that family $C$ is composed of $\text{card}(C)$ one-element sets. However, other less extreme cases are allowed as well, and in fact in an experimental study we use a family $C = \{C_B, C_W\}$ composed of two sets containing 8 elements each ($\text{card}(C) = 16$).

Proposed modification has been applied in optimization procedure of the hybrid opto-electronic pattern recognition system composed of HRWD and PNN. It allowed to improve the recognition abilities by reducing the normalized decision error by 5% compared to system optimized with classical indiscernibility relation. One should notice that this improvement is achieved on already optimized solution, which made any further improvement difficult. Obtained results confirmed our suggestions about suboptimality of earlier solutions, supplying with tools which allowed to obtain better classification ability of PNN. This experiment is an illustration of application of proposed methodology into hybrid pattern recognizer. But we think, presented generalization of indiscernibility relation will find many more applications for rough set based machine learning, since it gives natural way of processing real valued attributes within a rough set formalism.

5 Acknowledgements

The author would like to acknowledge here the financial support of this work by Institute of Informatics, at Silesian University of Technology in the framework of the BW2006 activities. Special acknowledgments are also for L. R. Jaroszewicz for supporting the author with the images of speckle structures used here as the illustrative example.

References:


