ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Prior knowledge in an end-user trainable machine vision framework

Klaas Dijkstra[1] Walter Jansen[1] Jaap van de Loosdrecht[1] *

1- NHL University of Applied Sciences - Center of Expertise Computer Vision
P.O. Box 1080, 8900 CB Leeuwarden - Netherlands

Abstract. The increasing popularity of machine vision based solutions in common applications calls for a structured approach for incorporating the end user's domain knowledge and limiting the solution's dependency on expert knowledge. We propose a framework facilitating optimized classification results and will show several approaches in which prior knowledge of the solution is captured in a neural network or in a geometric pattern matcher. The methodology is applied to disc print reading for antibiotic susceptibility testing by disc diffusion. Results show that increased prior knowledge produces better classifiers, and that more thorough optimization is required to increase the accuracy of classifiers which use less prior knowledge.

1 Introduction and Related work

Antibiotic Susceptibility Testing by Disc Diffusion (DD AST) is a regular task for lab technicians [1]. DD AST is used to determine susceptibility of a bacteria to multiple antibiotics. Discs with printed abbreviations of the antibiotics they contain are placed on the inoculated agar to inhibit growth. A part of DD AST is reading of the disc prints, which is mostly done manually. A misread of a disc print could lead to a faulty diagnosis, making configuration of an automated disc print reading system a highly delicate task.

Each microbiology lab uses a different set of antibiotic disc prints and each prefer pictures taken with different illuminations for microbial analysis. Configuring each disc print classifier separately for each lab is a costly and time consuming task for the technology expert. There is a need for machine learning methods which are end-user trainable meaning they can be configured without the intervention of a technology expert.

Related to end-user trainability is end-user software engineering (EUSE) which deals with software engineering performed by end-users instead of professional programmers [2] [3].

EUSE uses an end-user centered approach and still relies on end-users who are programming applications by visual interfaces or recording of macros. Attempts with using an interactive instruction based and a machine learning [4] based method has been researched [5] [6].

*This research is part of a project for BD Kiestra. Experiments were performed using the vision operators, MLP, GA and BM from the software package VisionLab of Van de Loosdrecht Machine Vision BV.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

This approach requires the end-user to be a partial technical expert on either programming logic or machine learning methods. In our opinion a real end-user centered approach should use automatic optimization which relieves the end-user from having to know underlying principles and treat the solution more as a block box and focus on the application rather than on the technology used.

A short survey of automatic DD AST systems shows that disc print reading in these systems is limited or not implemented [7] [8] [9]. To our knowledge no end-user trainable machine system for reading disc prints exists.

1.1 Approach

Artificial Intelligence (AI) is used to achieve end-user trainability. Our framework uses a ground-truth, collected in a microbial lab, for input. Classifiers are tested against the ground-truth and an optimizer is used to reinforce the classifier.

Two levels of classifiers are researched, where at each level the amount of prior knowledge about the problem domain is reduced. At the first level, a geometric pattern matcher which uses a single disc print per class as a matching model is used. This choice assumes prior knowledge about the nature of disc print reading, because it is known that disc prints are geometric patterns with a low intra-class variance. A geometric pattern matcher called the Blob Matcher (BM) [10] is used, because it is suitable for handling these types of problems.

At the second classification level, less prior knowledge about the problem domain is used. This is achieved by using the more generic Multilayer Perceptron (MLP) [11]. This type of classifier can handle a greater diversity of problems and the model of the disc needs to be trained using the provided ground-truth.

Three levels of optimizers are researched, where at each level the solutions space is searched more thoroughly. At the first level only the manual configuration by an expert engineer is used. At the second level optimization is performed by a Genetic Algorithm (GA) [12]. At the third level an additional Single Parameter Exhaustive Search (SPES) optimization method is used for the BM. For the MLP the thoroughness of the GA search is increased at the third level.

2 End-user trainable framework

The MLP has five parameters + the number of features parameters which, have to be optimized: Neurons in the first hidden layer, Learn rate, Momentum, Epochs and which input features are enabled. The 7 Hu moments [13] and a circular summation of pixel values are used as input values. The BM has four parameters + the number of classes parameters to optimize: Number of rotations, fill sample size, perimeter fill ratio and which disc print to use as a model for each class (model choice).

In our end-user trainable framework the score function is calculated from the ground truth by the evaluator and is used to order classifiers for the optimizer to converge to higher scoring classifiers. The design of the score function is based on metrics from a modified confusion matrix and shown in table 2. The identifier

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

for the cth class is $class_x$. CT is a confidence threshold regulating the trade-off between True Positives (TP) and False Positives (FP). An additional Best True Positives(BTP) metric is calculated by increasing the confidence threshold to a level where FP decrease to zero: $CT_{btp} = \{min(CT)|FP = 0$.

Table 1: Modified confusion matrix

		Label from Ground-truth $class_x$	Label from Ground-truth $!class_x$	
Class from Classifier	$class_x$	True Positive (TP)	False Positive (FP)	CT
	$!class_x$	False Negative (FN)	True Negative (TN)	

Because disc print reading has three objectives: Correctly read as many discs as possible (maximize TP), rather reject classification results than make a mistake (minimize FP) and be fast. The score function contains a general part with main objectives and a specific part with secondary objectives. The MLP score function is defined as: $score_{mlp} = TP + 10 \times FP + \frac{1}{n} + (1 - m)$ where n is the number of neurons used in the first hidden layer, and m the mean learn error of all output neurons. For the BM the score function is defined as: $score_{bm} = TP + 10 \times FP + \frac{1}{i} + \frac{1}{e}$ where i is the fill sample size and e is the number of rotations of the BM. The general part $(TP + 10 \times FP)$ is used to increase correctly read discs, but not allowing too many misreads. The weighing factor determines the trade off between these two objectives and determines how well the optimizer converges and which objective is favored in the end-result. For disc print reading, low FPs are favored meaning that the weighing should be above 1. It is empirically set to 10. The specific parts are used to reduce complexity of the classifier to make it faster. These values are all below one and above zero, so that these objectives are optimized when the main objectives stabilizes.

A Single Parameter Exhaustive Search (SPES) for each model choice of the BM is used. The disc print with the lowest aggregated error is used as a model for the class. For all remaining parameter of each classifier a GA is used.

We propose a method where two chromosomes with different mutation probabilities depending on the impact of the classifier parameter is used in the GA. The mutation ratio of each chromosome is the reciprocal of the number of genes in the chromosome.

The first chromosome of the MLP based systems contains settings for the number of neurons, learn rate, momentum and epochs, which impact the classifier as a whole. The second chromosome determines which features are enabled, because each class could be distinguished using a different set of features.

The first chromosome of the BM based systems determines the fill sample size, number of rotations and perimeter fill ratio, which impact the classifier as a whole. The second chromosome determines the model choices, which mostly impacts individual classes.

The value for population size and number of generations is mainly limited by available time and memory and should be as high as possible [14] and [15].

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

3 Experiments

Three ground-truth sets are used: Oxoid [16] containing 37 classes in 5620 discs, Rosco [17] containing 29 classes in 1148 discs and Mixed containing a total of 36 classes, 19 Oxoid and 17 Rosco with 390 images. The Mixed set has been selected by a microbiologist in a microbiology laboratory using specially designed software.

In table 2 a summary of the system configurations is shown.

Table 2: System configurations

Name	Classifier	Optimizer
MLP	Multilayer Perceptron	NA
BM	Blob Matcher	NA
GA_MLP	Multilayer Perceptron	Genetic Algorithm
GA_BM	Blob Matcher	Genetic Algorithm
SPESGA_BM	Single Parameter Exhaustive search as a preprocessor for the Genetic Algorithm	Blob Matcher

For each experiment the random fold is repeated five times and the GA optimizer is run five times for a fixed number of generations. The reported results are an aggregation of 25 results for the systems using a GA.

3.1 Comparing systems

Table 3: Best True Positives and match time on test set

System	Set	BTP mean (%)	Time mean (ms)
BM	Oxoid	96.0	56.2
GA_BM	Oxoid	99.0	13.7
SPESGA_BM	Oxoid	98.8	5.3
MLP	Oxoid	35.7	1.9
GA_MLP	Oxoid	27.3	1.9
BM	Rosco	93.8	84.5
GA_BM	Rosco	97.9	16.8
SPESGA_BM	Rosco	98.7	10.9
MLP	Rosco	95.4	0.3
GA_MLP	Rosco	90.7	0
BM	Mixed	87.2	74.7
GA_BM	Mixed	91.2	42.6
SPESGA_BM	Mixed	90.7	22.5
MLP	Mixed	49.9	0.3
GA_MLP	Mixed	69.7	0

Table 3 shows the BTP in percentages and the average classification speed on each of the three sets. The BM based systems, which use more prior knowledge, have higher accuracy than the MLP based systems. MLP based system produce faster classifiers. The automatically optimized classifiers are comparable or more accurate than manually configured classifiers. Further increasing the thoroughness of the search through the solution space by using SPES shows an increase

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Table 4: True and False Positives on set Oxoid

System	Validation	Subset	Gen.	Pop.	TP (%)	FP(%)
GA_MLP	two-fold	Training	30	25	61.0	0.03
GA_MLP	two-fold	Test	30	25	25.8	0.3
GA_MLP	two-fold	Training	60	36	99.9	0
GA_MLP	two-fold	Test	60	36	87.2	*6.1*
GA_MLP	three-fold	Training	60	36	94.6	0.02
GA_MLP	three-fold	Evaluation	60	36	75.7	0.6
GA_MLP	three-fold	Test	60	36	74.9	1.6

in classification speed of the resulting BMs. This is because the generic part of the score function favors less complex classifiers after optimizing accuracy.

Table 4 shows that a more thorough search through the solution space, by increasing the number of generations and the population size, results in more accurate MLPs. A side effect of the more thorough search is that the GA tends to over-fit the MLP. This is shown by comparing two validation methods. In a two-fold cross validation the training set of the MLP is also used to calculate the GA score function. In the three-fold cross validation both use separate sets (training and evaluation). The test set is used to assess generalization. The TP and FP for the evaluation and test set are close for the three-fold cross validation, while for the two-fold cross validation the TP and FP for the training and test set are more different. This shows that a three-fold cross validation produces better generalizing MLPs.

Disc print classification can be further improved by adding more prior knowledge. Usually the configuration of discs in a Petri-dish is known in advance. This means that only one disc needs to be read correctly, making the final probability of rejection of a Petri-dish containing 5 discs for the Mixed set: $(TN + FN)^5 = (2.58\% + 3.35\%)^5 = 7.33 \times 10^{-5}$.

4 Conclusions

Classifiers with different levels of prior knowledge produced by this framework are configured automatically and directly from the ground-truth provided by an end-user. The resulting classifiers are in general more accurate and faster than their manually configured counterparts. These facts show that end-user trainability is achieved using the proposed framework.

With MLP based classifiers a more thorough search through the solution space by increasing population size and the number of generations shows increased accuracy. For BM based systems a more thorough search using SPES produces faster classifiers. The best overall accuracy and speed is achieved by a combination of SPES, GA and BM.

5 Future Work

The framework is currently being extended to make regression analysis end-user trainable. The pilot application for this is DD AST, and preliminary results are

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

encouraging.

References

[1] European Commitee on Antimicrobial Susceptibility Testing. http://www.eucast.org [Accessed 5 December 2012], 2012.

[2] Andrew J. Ko, Robin Abraham, Laura Beckwith, Alan Blackwell, Margaret Burnett, Martin Erwig, Chris Scaffidi, Joseph Lawrance, Henry Lieberman, Brad Myers, Mary Beth Rosson, Gregg Rothermel, Mary Shaw, and Susan Wiedenbeck. The state of the art in end-user software engineering. *Computing Surveys*, 43, 2011.

[3] Howie Goodell. End-user programming. http://www.cs.uml.edu/ hgoodell/EndUser/ [Accessed 8 june 2012], 2012.

[4] E Alpaydin. *Introduction to machine learning*. The MIT Press, 1 edition, 2004.

[5] Michael Freed, Daniel Bryce, Jiaying Shen, and Ciaran O'Reilly. Interactive bootstrapped learning for end-user programming. In *Artificial Intelligence and Smarter Living The Conquest of Complexity: Papers from the 2011 AAAI Workshop*, 2011.

[6] Margaret Burnett. What is end-user software engineering and why does it matter? In V. Pipek et. al, editor, *International Symposium on End-User Development*, volume 2009, pages 15–28, 2009.

[7] J M Andrews, F J Boswell, and R Wise. Evaluation of the oxoid aura image system for measuring zones of inhibition with the disc diffusion technique. *Journal of Antimicrobial Chemotherapy*, 46(4):pp.535–540, 2000.

[8] Biomic. How does biomic v3 recognize each antibiotic disk or etest on the plate? Available at: http://www.biomic.com/BIOMIC_V3_common.html [Accessed 2 June 2011], 2011.

[9] BioLogics. Omnicon zone reader. http://www.biologics-inc.com/antibiotic-zone-reader.html [Accessed 28 Februari 2012], 2012.

[10] Jaap van de Loosdrecht. Computer vision course material. Available online at http://webserv.nhl.nl/ loosdrec/vision_course/ [Accessed on 30 november 2012], 2012.

[11] Simon Haykin. *Neural Networks and Learning Machines*. Pearson Education Inc., 2009.

[12] A E Eiben and J E Smith. *Introduction to Evolutionary Computing*. Springer, 2nd edition, 2007.

[13] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.

[14] S Gotshall and B Rylander. Optimal population size and the genetic algorithm. In *Proc On Genetic And Evolutionary Computation Conference*, 2000.

[15] D Vrajitoru. Large population or many generations for genetic algorithms ? implications in information retrieval. *Soft Computing in Information Retrieval. Techniques and Applications*, 2000:pp199–220, 2000.

[16] Oxoid. Product list 2011/2012 (antimicrobial susceptibility testing). http://www.oxoid.com/pdf/prodlist/ProdList2011-12.pdf [Accessed 20 Januari 2012], 2012.

[17] Rosco. Susceptibility testing. http://www.rosco.dk/default.asp?mainmenu=3 [Accessed 20 Januari 2012], 2012.