A Self-Learning Bayesian Classifier for Quality Evaluation in Automatic Sorting Systems of Fruit and Vegetables

Ch. Damyanov, V. Nachev, T. Titova, S. Yordanov

Key Words: *Bayesian classifier; self-learning; automatic sorting system; quality evaluation.*

Absract. The paper presents a self-learning algorithm for a quality-recognition (diagnostics) Bayesian classifier used in the AQS 602 automatic sorting system. The algorithm was created on the basis of a method for extraction of information from data in a regime of on-line classification and sorting of products. The new algorithm produces promising results providing the possibility for increasing the efficiency of the system which are likely to be of help for work with sorting equipment.

1. Introduction

With automatic sorting systems (ASS) of fruit and vegetables, the problem of recognizing the quality of products on the basis of their realizations amounts to the classification of random non-stationary signals in classes which cross each other by definition. That is why, for the purpose of ASS self-learning, the most often used methods are the statistical probabilistic models for image recognition. All these methods are grounded on the premises of Statistical Decision Theory [1,2,3,5,6]. The methods differ in that they resort to different criteria in order to reach a final solution.

Undoubtedly, the leading method in statistical decision theory is the Bayesian one which is based on a criterion in accordance with which the decision strategy is chosen in view of guaranteeing minimum average losses.

Briefly, the Bayesian classifier is a mechanism for classification error reduction. The minimum (Bayesian) error is the theoretical minimum which can be reached by a statistical classifier. In the case of a concrete classifier working with real data in real conditions, the method is optimal with respect to the average error risk but its classification capacities are restricted because of the requirement to know the exact prior probabilities for the different classes, i.e. the number of products from each class, which is unachievable in real conditions. This brings about a fundamental problem, especially concerning huge batches of products with great differences in quality.

The paper discusses a possibility for minimizing the errors during the quality recognition of fruit and vegetables with the help of a self-learning Bayesian classifier when the automatic sorting system works in an on-line regime.

2. Design of a Standard Bayesian Classifier

The strategy, based on the so-called Bayes' rule, requires the synthesis of a classifier using the Bayes' formula:

(1) $$P(D_k|S^*) = P(D_k).P(S^*|D_k) \Big/ \sum_{k=1}^{N} P(D_k).P(S^*|D_k)\,,$$

in which $P(D_k|S^*)$ is the posterior probability for the classification of a given object in class D_k, given the realization of a concrete complex of values of the features S^*. The solution is taken to be the evaluation determined by the maximization of $P(D_k|S^*)$ throughout D_k in the course of the entire observation interval

(2) $$S^* \in D_k \Leftrightarrow \arg\max_D \left[P(D_k|S^*)\right].$$

Out of practical considerations, it is expedient to use the logarithmic variant of (2):

(3) $$S^* \in D_k \Leftrightarrow \arg\max_D \left[\ell g P(D_k) + \ell g \sum_{p=1}^{N} P\left(S^*_{pq}|D_k\right)\right],$$

where the evaluation $\hat{P}(S^*_{pq}|D_k)$ is the conditional probability of the occurrence of the q-order of the p-feature for objects belonging to the class D_k. In (3), the denominator in the formula (1) has been ignored, which, if accurate to a known constant norm, is believed to be a negligibly small deviation from the strict optimal rule. It must also be paid attention to the fact that the denominator in the Bayes' formula is common for all classes D_k and most practical problems acknowledge the feature independence condition even in the presence of a correlation between some of the features.

In (3), the prior probability $P(D_k)$ for classes D_k and the conditional probabilities $P(S^*|D_k)$ are known from previous statistics and classifier training.

Formula (3) illustrates that the synthesis of the Bayesian classifier necessitates two probabilistic characteristics – the prior probabilities $P(D_k)$ and the conditional probabilities $P(S|D_k)$. It must be noted that the evaluation (3) is strongly dependent on the prior distribution $P(D_k)$, because of which the imprecise or random assignment of its type during the actual exploitation of the classifier would deprive us of the possibility for optimal evaluation [4,7].

Therefore, the conditional probability $P(S|D_k)$ requires continuous training with large-scale samples, which turns the re-adjustment of the classifier into a labour-consuming and rigid procedure.

This means that a trained machine based on a statistical probabilistic classifier, the Bayes' rule in this case, provides quite restricted possibilities for re-adjustment.

3. Modified Bayesian Classifier

The AQS 602 apparatus is a highly productive machine (3–5 tons/hr) for sorting potatoes (on the basis of exterior and interior defects) into 3 quality fractions (classes). *Figure 1* shows the functional block diagram of the apparatus. Each product goes through the data acquisition system, a PMC photometric camera. In a regime of free fall through the inspection area, the products are scanned and the information about their interior and exterior condition, acquired with the help of the products' spectral permeability, enters the computer in order to be processed so that a final classification solution can be obtained.

The data processing consists in the calculation of a given number of informative features, the complexes (S_p) for each product. The S features are determined on the basis of previous research and methods described in [1,7,8].

Figure 2 presents examples of the realizations of a signal *U(n)* which correspond to different product qualities (first quality – D_1 class, second quality – D_2 class, and third quality – D_3 class).

It must be pointed out that the realizations in *figure 2* may be not only spectral as is the case but also seismograms, cardiograms, accelerograms, etc. with other applications.

With AQS 602, solutions are grounded on the Bayes' formula in its logarithmic variant (3).

Figure 3 demonstrates a modified variant of the standard Bayesian classifier which is represented in the figure by the block for adaptation. As in the case with many other applications of the theory and practice of image recognition, the current image recognition problem is presented as a signal classification one, the signals representing the measured characteristics of the objects to undergo recognition [1,2]. The classification itself is reduced to the referral of each measured *n*-dimensional input vector U to one of the possibilities N of a number of classes of belonging (diagnoses) – D_k of the objects recognized (images).

Having in mind the above-mentioned disadvantages of the standard Bayesian classifier, the authors of this article propose that in an on-line regime the classifier undergoes self-learning correcting in specific time intervals, and after a given number of sorted products (batches), the algorithm-determined prior probabilities $P(D_k)$ with values approaching the real values of the prior probability distributions $\hat{P}(D_k)$ which really correspond to the concrete situation. Thus, the system undergoes self-learning and, by taking into account the new information, the algorithm adapts itself to the current state of the prior probability distributions of the recognition classes.

The prior probability distributions of the objects belonging to the respective classes are calculated by the formula $P(D_k) = m_k / m$ where m_k is the number of objects in the training sample from class k while m is the number of all objects on which the classifier has been trained.

After the block for concordance BC between $P(D_k)$ and the prior probability $\tilde{P}(D_k)$ measured for a given batch, the evaluation of the currently measured prior probability is:

$$\hat{P}(D_k) = \frac{m_k + \tilde{m}_k}{m + \tilde{m}}; \quad (\kappa=1,\ 2,\ 3), \tag{4}$$

where $\tilde{m}_k$ is the number of the objects in the batch classified in class k while $\tilde{m}$ is the number of all objects in the batch.

4. Results and Discussions

The experimental work related to the simulation and evaluation of the classifier outlined in *figure 3* was carried out in a MATLAB environment using a sample of 3,194 potato realizations belonging to three different grade classes. One part of them (700) were used for the training sample and the other realizations provided the basis for testing samples of equal sizes. The training of the classifier was preceded by a selection of the informative features via a genetic algorithm with the minimum classification error criterion for the testing sample. In the case of interest, the initial complex of 20 features was reduced to 3 (s1, s4 and $s9 = s7 - s8$). The distribution of the objects from the training sample into two solid angles is shown in *figure 4*.

The *table* presents the results of the classification.

The efficiency of the suggested algorithm is evident when the evaluation of the prior probability is corrected automatically after each object classified. This is shown clearly in *figure 5* which illustrates the trend of error reduction during a self-learning process of the kind. The procedure is in an „off-line" regime and cannot be presently incorporated in the AQS 602 system due to the limitations following from its highly accelerated work whereas self-learning for individual batches is really applicable.

The obtained results make it possible to confirm the positive effect of the iterative adjustment calculations of the prior probabilities.

5. Conclusions

The stochastic nature of the classification problem under study is a prerequisite for the existence of an error source which cannot be eradicated in principle. Therefore, under conditions when errors occur along with the supposition that all processes of observation and data processing of the objects to be recognized are ideal, the outlined self-learning algorithm of the Baye

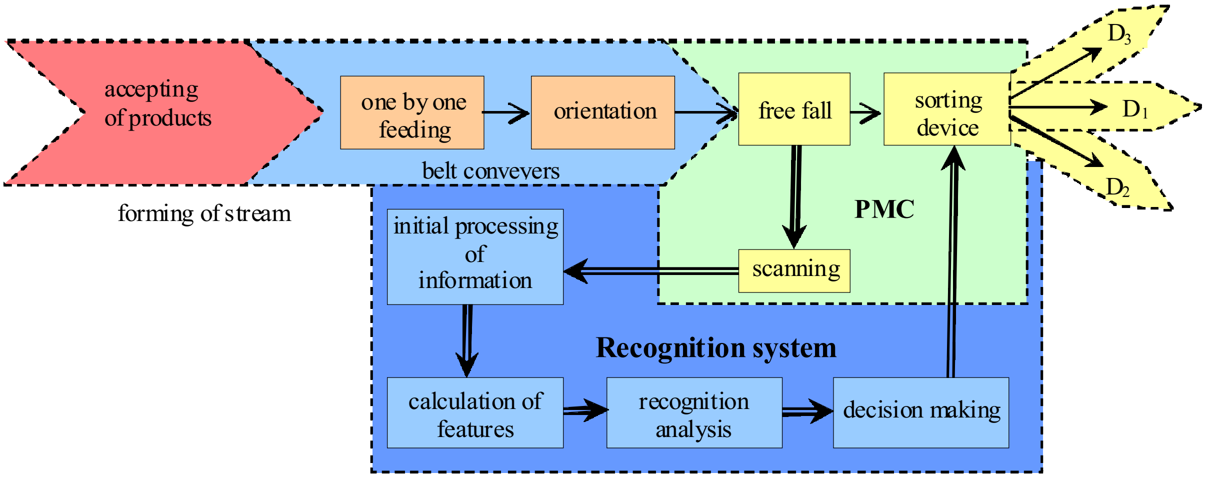

Figure 1. A block diagram of the AQS 602 Automatic Sorting System

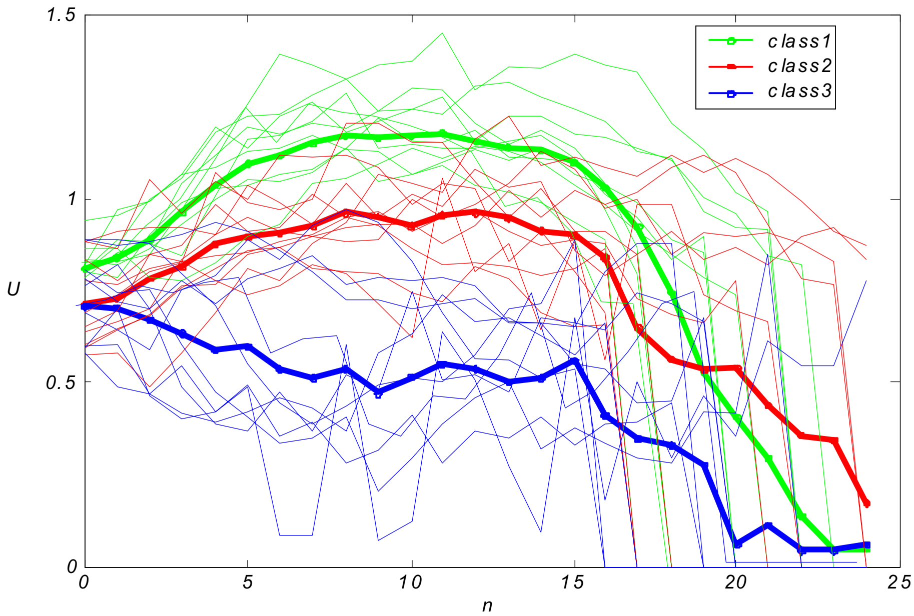

Figure 2. Images of potato tubers from three quality fractions - D_1, D_2 and D_3

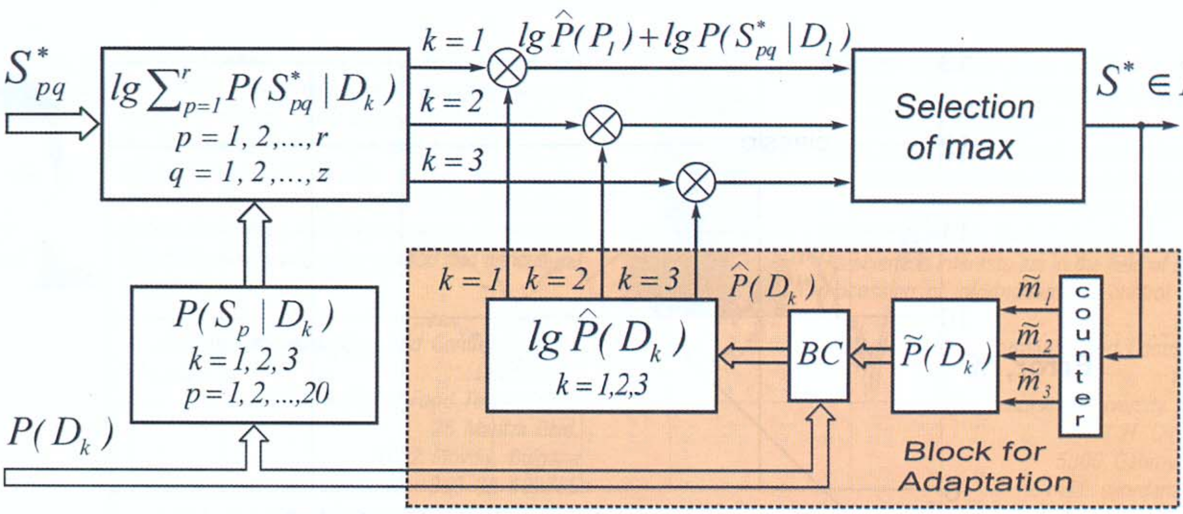

Figure 3. A functional diagram of the Self - Learning Bayesian Classifier

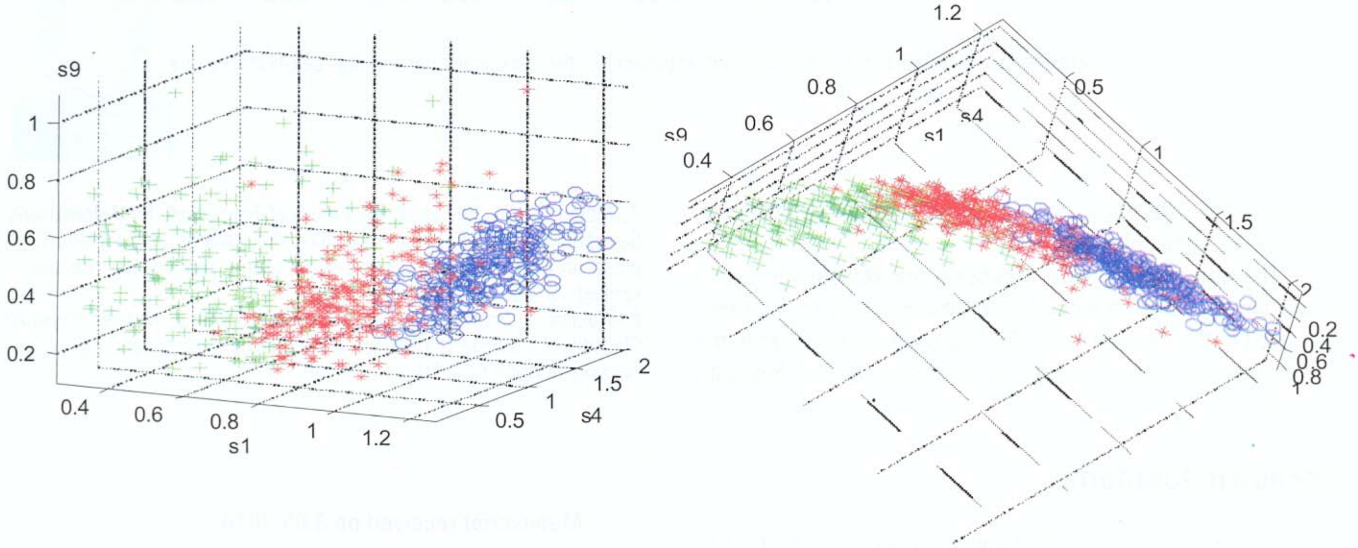

Figure 4. Distribution of the classified objects after the training

Table

	Training sample (1÷700)	Testing sample-1 (701÷1100)	Testing sample-2 (1101÷1500)	Testing sample-3 (1501÷1900)	Testing sample-4 (1901÷2300)	Testing sample-5 (2301÷2700)	Testing sample-6 (2701÷3194)
	$P(D_K)$	$\widetilde{P}(D_k)_{old} \rightarrow \widetilde{P}(D_k)_{new}$					
Mean classification error	**8.6**	**11.6**	**10.6**	**10.8**	**10.4**	**9.6**	**9.5**

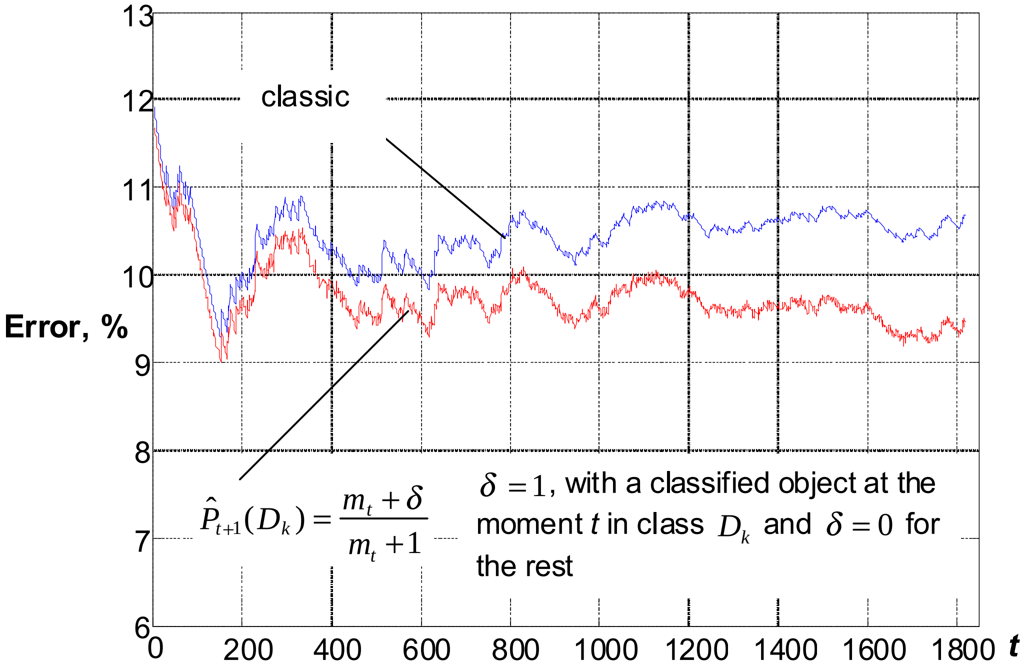

Figure 5. An error trend chart for both approaches, the classical one and the self-learning one

sian classifier differs from the standard Bayesian classifier because of its undisputed efficiency.

The proposed idea for improving the standard Bayesian classifier with a built-in block for self-learning may be successfully used both in the existing AQS 602 system as well as in the development of new industrial applications for the Bayesian classifier.

Acknowledgements

This study was carried out in the framework of the DO 02-143/16.12.2008 project titled „Development of Intelligent Technologies for Assessment of Quality and Safety of Food and Agricultural Products", financed by the National Science Fund of the Bulgarian Ministry of Education and Science.

References

1. Damyanov, Ch. Iv. Non-destructive Quality Evaluation in Automatic Sorting Systems (in Bulgarian). UFT Academic Publishing House, Plovdiv, 2006.
2. Duda, R. O., P. E. Hart, D. G. Strork. Pattern Classification. 2nd Ed., John Wiley&Sons, Inc., N.Y., 2001.
3. Howson, C., P. Urbach. Scientific Reasoning: The Bayesian Approach. 2nd Ed., Open Court, La Salle, IL, 1993.
4. Huang, K., Zh. Zhou, Ir King, M. R. Lyu. Improving Naive Bayesian Classifier by Discriminative Training. Proceedings International Conference on Neural Information Processing (ICONIP' 05), Taipei, Taiwan, 2005.
5. Korb K. B., A. E. Nicholson. Bayesian Artificial Intelligence. Chapman & Hall/CRC Pres, London, 2004.
6. Mackay, D. Y. C. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 6th Ed., 2004.
7. Moran, S., Y. He, et al. An Empirical Framework for Automatically Selecting the Best Bayesian Classifier. Proceedings of the World Congress on Engineering, 2009. Imperial College, London, UK, Newswood Limited. 1: 289-294, 2009.
8. Weldon, T. Improved Image Segmentation with a modified Bayesian Classifier. Intonation Conference on Acoustics, Speech, and Signal Processing (ICASSP'06), May 14-19, 2006.

Manuscript received on 3.05.2010

Chavdar Damyanov *(born 1946), has an M.Sc. and PhD degree from the Technical University of Sofia. He is a D.Sc. and professor at the University of Food Technologies of Plovdiv and he is a Head of Automatics, Information and Control Systems Department in the same university. His scientific and research interests are in the areas of automatic control systems, pattern recognition, application of artificial intelligence and quality evaluation and nondestructive sorting of food products. He has over 180 publications and several patents.*

Contacts:
Automatics, Information and Control Systems Department
University of Food Technologies
26 Maritza Blvd.
4002 Plovdiv, Bulgaria
Phone: +359 32 603897
e-mail: chavdam@yahoo.com

Veselin Nachev *(born 1975), has a doctoral degree (PhD) in automatics and computing components and devices. He is a Principal Assist. Prof. at the Automatics, Information and Control Systems Department at the University of Food Technologies of Plovdiv. His scientific interests are control system, pattern recognition, artificial intelligence systems and automatic quality evaluation of food and agricultural products.*

Contacts:
Automatics, Information and Control Systems Department
University of Food Technologies
26 Maritza Blvd.
4002 Plovdiv, Bulgaria
Phone: +359 32 603765
e-mail: v_nachevbg@yahoo.com

Tanya Titova *(born 1982), has an M.Sc. degree in Automatics, Information and Control Systems. She is an extramural PhD student. Assist. Prof. Titova works at the Automatics, Information and Control Systems Department at the University of Food Technologies of Plovdiv. Her scientific interests are control system, pattern recognition, artificial intelligence systems and automatic quality evaluation of food and agricultural products.*

Contacts:
Automatics, Information and Control Systems Department
University of Food Technologies
26 Maritza Blvd.
4002 Plovdiv, Bulgaria
Phone: +359 32 603765
e-mail: t_titova@abv.bg

Stanimir Yordanov *(born 1967), has M.Sc. from the Technical University of Gabrovo (1992). In 2005 he received a PhD degree in Computer Sciences. Since 1993 till the present moment he is working as a Assistant Professor and Assoc. Professor (2009) in the Department of Automation, Information and Control Systems in Technical University of Gabrovo. His scientific interests are in the field of systems for processing of information and control.*

Contacts:
Automatics, Information and Control Systems Department
Technical University of Gabrovo
4 H. Dimitur Street
5300 Gabrovo, Bulgaria
e-mail: sjjordanov@mail.bg