#31

BICA VPS 2023

Using Electronic Nose in Forensic Odor Analysis.

Alexander N. Shtanko*[1], Sergey D. Kulik[2].

[1]National Research Nuclear University MEPhI, Kashirskoe shosse 31, Moscow 115409, Russia, shtanko-mephi@yandex.ru. [2]National Research Nuclear University MEPhI, Kashirskoe shosse 31, Moscow 115409, Russia.

Alexander N. Shtanko

Sergey D. Kulik

ABSTRACT

Forensic odor analysis often uses biodetectors (usually dogs) for odor matching. Biodetectors can be expensive to train and maintain. The task at hand is finding a sample corresponding to a given one out of several samples. It can be prohibitively expensive using biodetectors with large enough number of samples. An alternative is using artificial odor analyzer (e-nose). We propose a two-step method: first, picking few candidates using e-nose, then applying biodetector to find the required sample out of the already chosen candidates. We also calculate theoretical performance of this method as well as generalize the method into an abstract cascade classifier and calculate its theoretical performance.

INTRODUCTION

Generally, animals (biodetectors) are unrivaled in odor analysis. Criminal forensics agencies use trained dogs and have established procedures for odor analysis. Given these procedures the results of biodetectors can be considered perfect for the purposes of this article. However live animals can be expensive to train and maintain thus the scale of their application is limited. One potential way to apply odor on a larger scale is using gas analyzer (artificial nose, e-nose), it is already used in various fields, such as food quality control or medical diagnostics. It is possible that at some point e-nose would be accurate enough to be used in forensics. It should be expected that these types of devices will have lower accuracy than animals. However, if the cost of their use is low enough, they can be used at least together with biodetectors. In this article we propose two-step method for using e-nose together with biodetectors, we also generalize the approach to a cascade classifier similar to the one used in Viola-Jones algorithm.

TWO-STEP METHODS

The task at hand. There are several examined odor samples and a given query sample. The task is to find among the examined odor samples the only one (target sample) corresponding to the query sample. This task can be viewed as binary classification task since every sample is either target or not. We suppose that the biodetector can perform classification perfectly. E-nose classify a positive sample correctly with probability p_{11}, incorrectly with p_{10}; a negative sample correctly with $p00$, incorrectly with p_{01}. These can be changed varying e-nose positive classification threshold. E-nose also has ROC-curve, we interpret it as a function $p_{11} = ROC(p_{01})$.

Let there be *N* examined samples including one target sample. The proposed method consists of two steps: all examined sampled are classified by e-nose; chosen by e-nose candidates are classified by biodetectors.

On average e-nose will positively classify $(N - 1)p_{01} + p_{11} \approx Np_{01}$. The probability that the target sample will be among the chosen candidates equals p_{11}. Since the resources of biodetectors are limited, the number of chosen samples as candidates should be small. Suppose the number of candidates can not be larger than *M*. Then, according to the formula above $(p_{01}N) \le M$ meaning the threshold of the e-nose should be chosen so $p_{01} \le M/N$. As ROC-curve is monotonic, $p_{11} \le ROC(M/N)$.

CASCADE CLASSIFIER

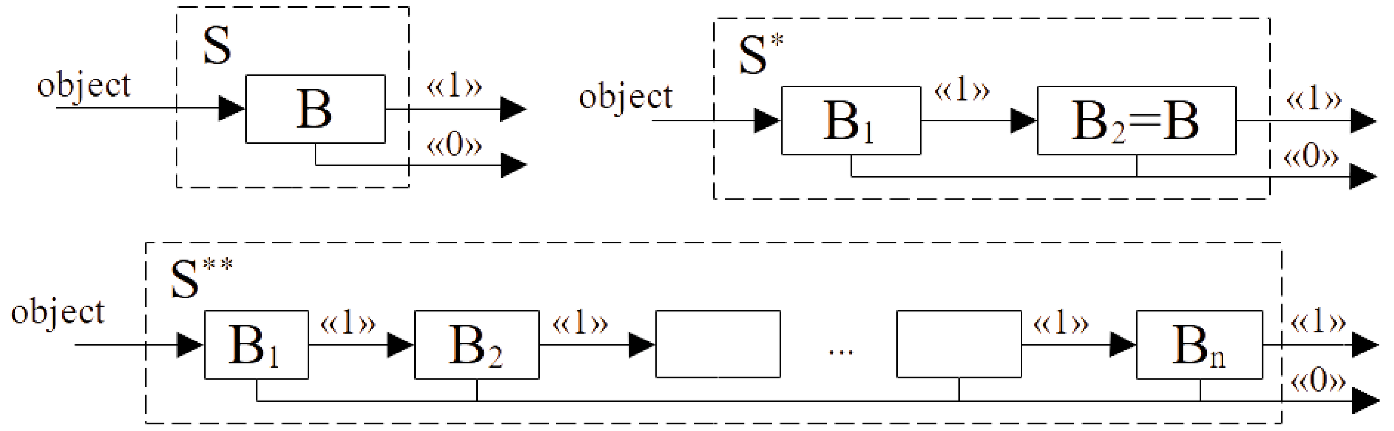
The proposed method could be generalized to a cascade classified like the one used in Viola-Jones face detection algorithm. Suppose there is a stream of objects to be classified. Objects can be either positive ("1") with probability p_1 or negative ("0") with p_0 ($p_1 \ll p_0$ in this case).

Consider the initial system S (see images below) consisting of a single block B. This system has average per object cost of running C and associated probabilities of correct classifications and mistakes $(p_{00}, p_{01}, p_{10}, p_{11})$.

To reduce the costs, it can be replaced with a two-block system S*. Another (faster) block is added at the front to form a two-block cascade classifier. New block has its own error probabilities $(p_{00}^1, p_{01}^1, p_{10}^1, p_{11}^1)$ and costs C^1. In practice the block rarely reject positive objects $p_{10}^1 \ll 1$, $p_{11}^1 \approx 1$. The costs equal $C^* = C^1 + (p_0 p_{01}^1 + p_1 p_{11}^1)C$. Let the efficiency measure be $E = C/C^*$ then $E = 1/\left(\frac{C^1}{C^2} + p_0 p_{01}^1 + p_1 p_{11}^1\right)$. It makes sense to use improved system if $E \ge 1$, it can be rewritten to $\alpha(1 - p_0 p_{01}^1 - p_1 p_{11}^1) \ge 1$, $\alpha = C^2/C^1$. Given $p_0 \approx 1$, $p_1 \approx 0$ it is roughly the same $\alpha p_{00}^1 \ge 1$.

Probability of a true positive classification $p_{11}^* = p_{11}^1 p_{11}^2$, false positive classification $p_{01}^* = p_{01}^1 p_{01}^2$. Probability of the correct negative classification $p_{00}^* = p_{00}^1 + p_{01}^1 p_{00}^2$, false negative classification $p_{10}^* = p_{10}^1 + p_{11}^1 p_{10}^2$.

The system can be further generalized to use multiple blocks (system S**). Probabilities and costs can be calculated as above.



CONCLUSIONS

Thus, odor analysis (especially in forensics) is accurately performed by biodetectors (trained dogs). Artificial odor analysis devices (e-nose) exist and are applied in various fields; however, we should expect them to perform less accurately.

We proposed a two-step method for the task of searching for a sample corresponding to a given one among a number of samples. This method makes use of both e-nose and biodetectors. True positive rate of the system is calculated to be no greater than *ROC*(*M*/*N*).

This approach is further generalized to an abstract cascade classifier. Resource consumption of this method is calculated. It is shown that such an approach can be more resource efficient than using a one-step classifier. The criterion of the efficient application is introduced. Error probabilities are calculated.

As such artificial odor analysis devices can be applied to practical tasks even if they are not as accurate as biodetectors

ACKNOWLEDGMENTS

This work was supported by the MEPhI Program Priority 2030.