Top-down design of chemical database classifiers based on oscillatory chemical reactions.

Jerzy Gorecki, Konrad Gizynski, Ludomir Zommer

Institute of Physical Chemistry,
Polish Academy of Sciences, Warsaw, Poland.

Abstract

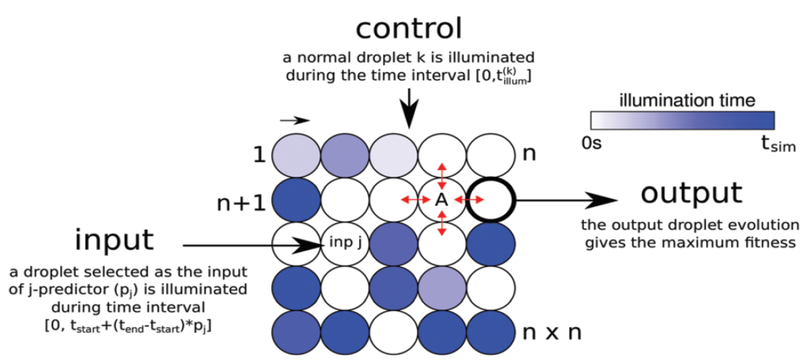
The top-down approach is applied to construct database classifiers performing information processing with a network of coupled chemical oscillators. Oscillators are described by a model of a photosensitive variant of Belousov-Zhabotinsky (BZ) reaction and their activity is controlled by external illumination. The considered model simulates time evolution of droplets containing reagents of BZ reaction surrounded by an oil phase. In the model, oscillations are inhibited at a high intensity of light. The input information is introduced by illumination of selected oscillators. The output information is extracted from the number of oscillations observed at a chosen network element. An evolutionary algorithm is used to determine the control (illumination time) of other droplets in the network to maximize the mutual information between the types of database records and the network outputs. The reliability of optimized classifiers is discussed for different numbers of oscillators involved in data processing and for different network geometries.

Problem

Given a dataset D as the set of records in the form of (n +1) tuples: (*p1,p2,....pn, q*) where *pi* are predictors and *q* is the record type. Solve the classification problem (i.e. find the record type if the values of predictors are known) using the time evolution of a network of oscillators.

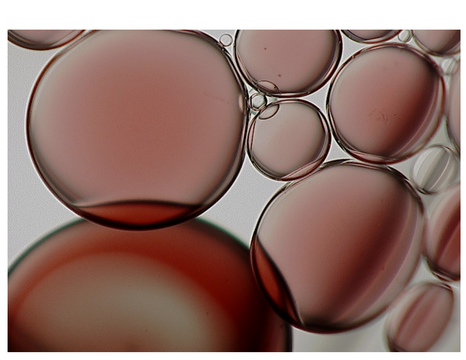
The idea of a classifier

Consider a network of oscillators that can be controlled by an external parameter ("illumination"). The control can be applied to individual oscillators and inhibit their activity. The output information is extracted as the number of oscillations of a selected droplet

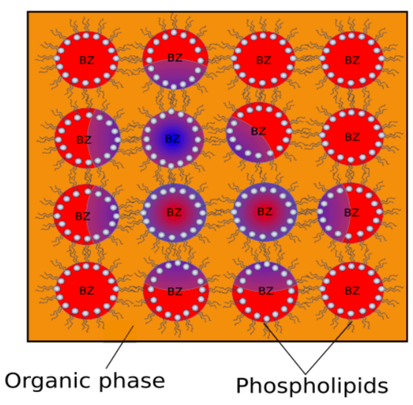


A classifier is characterized by a number of parameters:
A) the geometry of interactions between droplets,
B) times {t_start, t_end} defining the relationship between the value of predictor and the physical stimulus acting on a droplet,
C) location of input droplets,
D) times of illumination {t_ilum} for droplets that are not the input ones. They are fixed for a given network and define its activity.
E) location of the output droplet.
The parameters (B-D) are subject to optimization in order to maximize the mutual information between the number of output droplet oscillations and the types of processed database records (the network fitness).

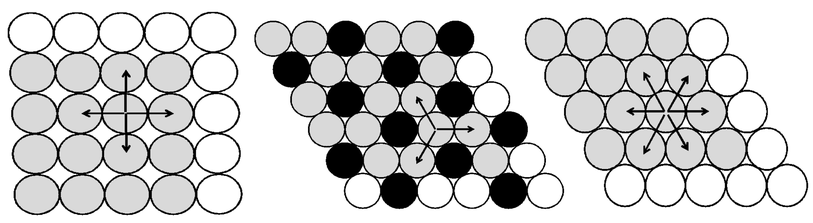
A physical prototype of an oscillatory network



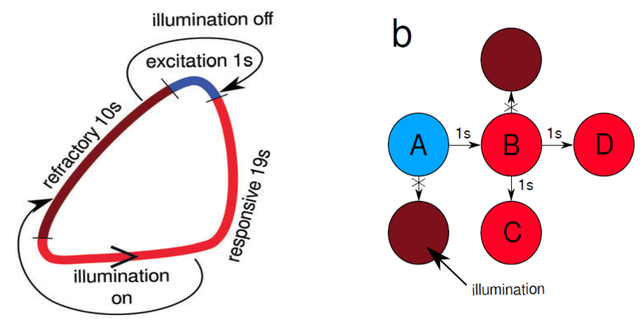
A physical prototype of the network are droplets containing reagents of photosensitive BZ reaction in the water phase. The droplets are surrounded by the oil phase containing lipids that stabilize them mechanically. Touching droplets communicate via diffusion of reaction activator (HBrO2) through the lipid bilayers.



We consider the following geometries of networks:

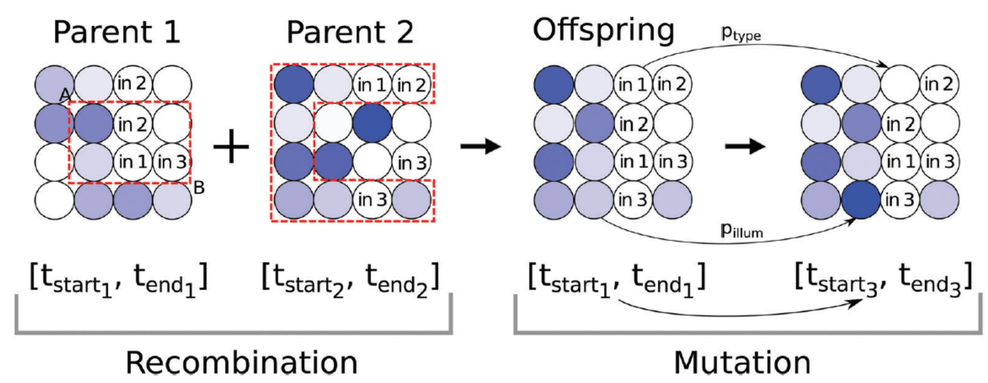


The model of droplet evolution includes 3 phases: excitation, inhibitory and responsive:



Classifier optimization using a genetic algorithm

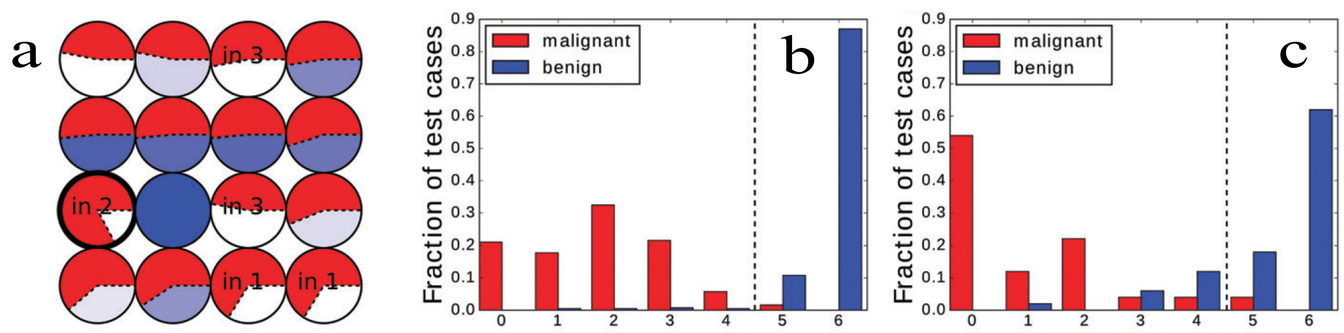
Optimization procedure starts with a population of N (from 50 to 5000) randomly generated networks. Next the following steps are repeated K times (from 200 to 50000):
1. identification of subpopulation of networks S characterized by maximum fitness (the upper 80 %).
2. recombination between networks belonging to S,
3. mutations of the obtained networks.



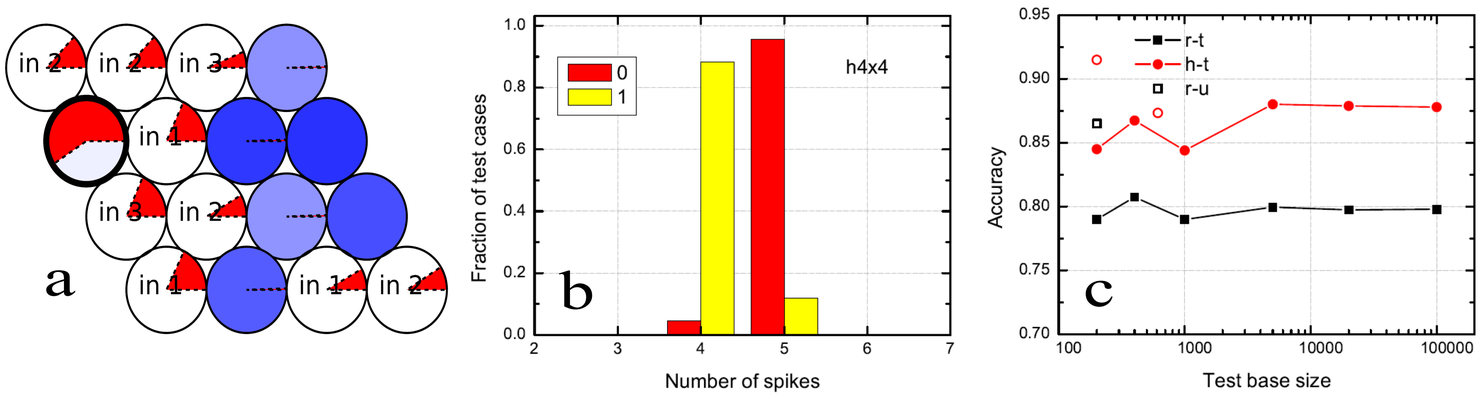
The predictive power of chemical classifiers:

We use two databases: one, of the order of 500 records, is used for network training,the other, usually larger, for testing the network accuracy.

A classifier of the Wisconsin Breast Cancer Dataset based on the regular lattice. (a) The structure of a classifier optimized using the training dataset within 500 generations. The intensity of blue color increases with {t_illum}. The output signal of the network obtained for the records from (b) the training (599 records) and (c) the test dataset (100 records). The dashed line indicates a threshold that leads to the classification rule with an accuracy of 98% for the training dataset. The same threshold value applied for the test dataset gives an accuracy of 88%.



A classifier of the sphere inscribed in the unit cube. Records are made of three random numbers in the range [0,1] describing a point in the cube [0,1]x[0,1]x[0,1]. The record type is 1 if the point is located inside the inscribed sphere. (a) The structure of a classifier optimized using a population of 100 networks for 5000 generations. (b) The output signal of the network for the training dataset (200 records). (c) The comparison between regular and hexagonal network structures for test datasets of different sizes.



Conclusions

- A network of coupled chemical oscillators can work as a database classifier with the reliability exceeding 90%.
- The geometry of connections seems to play a secondary role. The classifier with the hexagonal geometry is only slightly better than the one formed by the square lattice.
- Chemical classfiers do have predictive power. They can classify correctly the records that were not included into a training dataset.
- The classifier functionality does not depend on the oscillatory process that governs the time evolution of droplets. We can use enzymatic, thermochemical or electrochemical oscillators instead.

Acknowledgment: The work was supported by the Polish National Science Centre grant UMO-2014/15/B/ST4/04954.

References

[1] G. Gruenert, K. Gizynski, G. Escuela, B. Ibrahim, J. Gorecki, P. Dittrich, Int. J. Neur. Syst. (2015) Vol. 25(7), 1450032.
[2] K. Giżyński, J. Górecki, Computational Methods in Science and Technology (2016) Vol. 22, pp. 167-177.
[3] K. Gizynski, G. Gruenert, P. Dittrich, J. Gorecki, Evolutionary Computation 2017, Vol. 25, pp. 643–671.
[4] K. Gizynski, J. Gorecki, Cancer classification with a network of chemical oscillators, Phys. Chem. Chem. Phys. (2017) Vol. 19, pp.28808-28819.