Modeling human expertise on a cheese ripening industrial process using GP

Olivier Barrière[1], Evelyne Lutton[1],
Cedric Baudrit[2], Mariette Sicard[2], Bruno Pinaud[2], and Nathalie Perrot[2]

[1] INRIA Saclay - Ile-de-France
Parc Orsay Université, 4, rue Jacques Monod
91893 ORSAY Cedex, France
{Olivier.Barriere, Evelyne.Lutton}@inria.fr
http://apis.saclay.inria.fr
[2] UMR782 Génie et Microbiologie des Procédés Alimentaires.
AgroParisTech, INRA, F-78850 Thiverval-Grignon, France
{cbaudrit, mariette.sicard, bruno.pinaud, nperrot}@grignon.inra.fr

Abstract. Industrial agrifood processes often strongly rely on human expertise, expressed as know-how and control procedures based on subjective measurements (color, smell, texture), which are very difficult to capture and model. We deal in this paper with a cheese ripening process (of french Camembert), for which experimental data have been collected within a cheese ripening laboratory chain. A global and a monopopulation cooperative/coevolutive GP scheme (Parisian approach) have been developed in order to simulate phase prediction (i.e. a subjective estimation of human experts) from microbial proportions and Ph measurements. These two GP approaches are compared to Bayesian network modeling and simple multilinear learning algorithms. Preliminary results show the effectiveness and robustness of the Parisian GP approach.

1 Introduction

This study is part of the large INCALIN research project, whose goal is the modeling of agrifood industrial processes[1]. The competitive challenge to which agrifood industries are facing is related to quality and sustainability of food products. The aim of the INCALIN project is to build decision support tools to manage the manufacturing process as a whole. Current knowledge on industrial agrifood processes are focussed on microbiological, mechanistic, sensorial or physicochemical changes, and are expressed in various ways: databases, mathematical models, and know-how of operators-experts in terms of formal or unformal reasoning. Among the fragmented knowledge available, the human-expert knowledge is certainly the most challenging to capture.

We focus in this paper on a cheese ripening process (section 2): The cheese, during ripening, is an ecosystem (a bio-reactor) that is extremely complex to be modeled as a whole, and where human experts operators have a decisive role. The modifications of substrate under the action of several populations of micro-organisms is only

[1] supported by the French ANR-PNRA fund.

partially known, and various macroscopic models have been experimented to embed expert knowledge, like expert systems [12, 13, 11], neural networks [14, 17], mechanistic models [1, 20], or dynamic Bayesian networks [4].

The major problem common to these techniques is related to the sparseness of available data: collecting experimental data is a long and difficult process, and resulting data sets are often uncertain or even erroneous. The precision of the resulting model is often limited by the small number of valid experimental data, and parameteter estimation procedures have to deal with incomplete, sparse and uncertain data. In this context we consider stochastic optimisation techniques, like evolutionary techniques, which have been proven successful on several complex agrifood problems [3, 8, 21].

The idea developed in this study is based on the following question: is it possible to capture (learn) in a satisfying way an expert knowledge with help of a model evolved by genetic programming, for a complex cheese ripening process ?

The first step in this direction aims at comparing a part of a reference dynamic Bayesian model whose structure is based on expert knowledge (section 2) with evolved GP estimators, using a global strategy (section 3) and a cooperative/coevolutive strategy (Parisian GP, section 4). Experimental results (section 5) prove the efficacy of GP approaches to estimate the phase parameter of the process (currently made “at hand” in industrial chains). Section 6 then sketches the next steps of the study in order to build an efficient model of the whole cheese ripening process.

2 The camembert-cheese ripening process

For soft-mould cheese the most important biochemical phenomena occur during ripening. Relationships between microbiological and physicochemical changes depend on environmental conditions (*e.g.* temperature, relative humidity ...) [15] and influence the quality of ripened cheeses [9, 16]. A ripening expert is able to estimate the current state of some of the complex reactions at a macroscopic level through its perceptions. Control decisions are then generally based on these subjective but robust expert measurements.

Experimental procedures in laboratories (“model cheeses”) use pasteurized milk inoculated with *Kluyveromyces marxianus* (*Km*), *Geotrichum candidum* (*Gc*), *Penicillium camemberti* (*Pc*) and *Brevibacterium auriantiacum* (*Ba*) under aseptic conditions (detailed in [16]).

Experts use their senses to follow cheese ripening and they probably aggregate in a complex way these information to regulate the evolution of the process. An important information for parameter regulation is the subjective estimation of the current state of the ripening process, discretised in four phases:

- **Phase 1** is characterized by the surface humidity evolution of cheese (drying process). At the beginning, the surface of cheese is very wet and evolves until it presents a rather dry aspect. The cheese is white with an odor of fresh cheese.
- **Phase 2** begins with the apparition of a *P. camemberti*-coat (*i.e* the white-coat at the surface of cheese), it is characterized by a first change of color and a "mushroom" odor development.
- **Phase 3** is characterized by the thickening of the creamy under-rind. *P. camemberti* cover all the surface of cheeses and the color is light brown.

– **Phase 4** is defined by strong ammoniac odor perception and the dark brown aspect of the rind of cheese.

These four steps are representative of the main evolution of the cheese during ripening. The expert's knowledge is obviously not limited to these four stages. But these stages help to evaluate the whole dynamics of ripening and to detect drift from the standard evolution.

3 Phase estimation using GP

The interest of evolutionary optimisation methods for the resolution of complex problems related to agrifood has been proved by various recent publications. For example [3] uses genetic algorithms to identify the smallest discriminant set of variables to be used in certification process for an italian cheese (validation of origin labels). [8] used GP to select the most significant wavenumbers produced by a Fourier transform infrared spectroscopy measurement device, in order to build a rapid detector of bacterial spoilage on beef. And a recent overview on optimisation tools in food industries [21] mentions works based on evolutionary approaches.

In a previous work on cheese ripening modeling [4, 19], a dynamic bayesian network has been built, using human expert knowledge, to represent the macroscopic dynamic of each variable. The phase the network is in at time t plays a determinant role for the prediction of the variables at time $t+1$. Moreover, four relevant variables have been identified, the derivative of pH, la, *Km* and *Ba* at time t, allowing to predict phase at time $t+1$.

In this paper, we will focus on the phase estimation process: a genetic programming approach is used to search for a convenient formula that links the four derivatives of micro-organisms proportions to the phase at each time step t (static model), without *a priori* knowledge of the phase at $t-1$. This problem is a symbolic regression one, however, it has to be noted that the small number of samples and their irregular distribution make it difficult. Results will be compared with the performances of a static Bayesian network, extracted from the DBN of [4], and with a very simple learning algorithms (multilinear prediction, see section 5).

3.1 Search space

The derivatives of four variables will be considered, i.e. the derivative of pH (acidity), la (lactose proportion), *Km* and *Ba* (lactic acid bacteria proportions, see section 2), for the estimation of the phase (static problem). The GP will search for a phase estimator $\widehat{Phase}(t)$, i.e. a function defined as follows:

$$\widehat{Phase}(t) = f(\frac{\partial pH}{\partial t}, \frac{\partial la}{\partial t}, \frac{\partial Km}{\partial t}, \frac{\partial Ba}{\partial t})$$

The function set is made of arithmetic operators: $\{+, -, *, /, \hat{}, log\}$, with protected $/$ and log, and logical operators $\{if, >, <, =, and, or, xor, not\}$ in order to allow complex estimation formula.

The terminal set is made of the four partial derivatives plus real constants. The constant's values are not limited, but randomly initialised using one of the following laws $\mathcal{U}\ [0,1]$, $-\mathcal{U}\ [0,1]$, $\mathcal{N}\ (0,1)$ ($\mathcal{U}$ is the uniform law, and $\mathcal{N}$ the normal law).

3.2 Fitness function

Available data are shared in two sets: learning set and validation set, that are randomly chosen within the available data set for each run. The 16 available experiences are thus randomly shared between learning and validation sets. The size of the learning set vary from 10 to 15 experiments, while the size of the corresponding validation set vary from 6 to 1 experiment (see section 5).

The fitness function, *to be minimised*, is made of a factor that measures the quality of the fitting on the learning set, plus a "parsimony" penalisation factor in order to minimize the size (the number of nodes, actually) of the evolved structures (to avoid bloat). It is divided by the number of variables involved in the evaluated tree in order to favour structures that embed all four variables of the problem (this is a requirement of biologists ; experiments also show that recognition results are better with this constraint):

$$fitness = \frac{\sum_{learning\_set} \left| f(\frac{\partial pH}{\partial t}, \frac{\partial la}{\partial t}, \frac{\partial Km}{\partial t}, \frac{\partial Ba}{\partial t}) - Phase(t) \right| + W\#Nodes}{\#Variables + 1}$$

The parameter W has been experimentally tuned, the optimal value ($W = 1$) favours evolution of structures with 30 to 40 nodes.

3.3 Genetic operators

A classical tree crossover (exchange of subtrees from a randomly chosen node) has been used with probability p_c (defined per tree), as a means of evolving the structure of the tree. Two types of mutations have been used:

- **a subtree mutation** (mutation of the structure), that randomly rebuilt a new subtree from a randomly chosen node, applied with probability p_{sm} (defined per tree),
- **a point mutation** (mutation of nodes content), applied with probability p_{cm} (also defined per tree) that does not modify the structure, but randomly changes the content of each node of the tree within the set of compatible functions or terminals (arity constraints). The probabilities (defined per node) are detailed in table 1. Real values are considered separately and undergo a real mutation with probability p_{rm} as a multiplicative perturbation according to a χ^2 law of parameter N:

  $$x' = x \frac{\sum_{i=1}^{N} \mathcal{N}(0,1)^2}{N}$$

  p_{rm} and N vary linearly according to generations, from 0.1 to 0.5 for p_{rm}, and from 1 to 1000 for N, in order to start with rather unfrequent large radius mutations and finish with more frequent mutations with smaller radius.

Table 1. Probabilities of point mutation operators

From	to	probability
operator	operator	0.1
variable	variable	0.1
variable	constant	0.05
constant	variable	0.05
constant	constant	p_{rm}: 0.1 to 0.5
		N: 1 to 1000

Crossover, subtree and point mutation probabilities vary along evolution according to the adapting scheme[6] available in the GPLAB toolbox[10]. p_c, p_{sm} and pcm are initially fixed to $\frac{1}{3}$, and are updated according statistics of success of the various operators computed on a tuneable window of past generations.

4 Phase estimation using a Parisian GP

Cooperative co-evolution techniques rely on the imitation of cooperative capabilities of natural populations, and their ability to build solutions via a cooperation process. These techniques are starting to be used with success on learning problems, see [2] for a recent reference on the topic. The large majority of these approaches deals with a coevolution process that happens between a fixed number of separated populations. We experiment here a different implementation of cooperative coevolution principles, known as the Parisian approach [5, 18], that uses cooperation mechanisms within a *single* population. It is based on a two-level representation of an optimization problem, in the sense that an individual of a Parisian population represents only a part of the problem solution. An aggregation of multiple individuals must be built in order to obtain a solution at hand. In this way, the co-evolution of the whole population (or a major part of it) is favoured instead of the emergence of a single best individual, as in classical evolutionary schemes. The motivation is to make a more efficient use of the genetic search process, and reduce the computational expense. Successful applications of such a scheme usually rely on a lower cost evaluation of the partial solutions (i.e. the individuals of the population), while computing the full evaluation only once at each generation.

Phase estimation can actually be split into 4 combined (and simpler) phase detection trees. The structures searched are then binary output functions (or binarised functions) that are able to characterize one of the four phases. A global solution being made of at least one individual of each phase.

4.1 Search space and local fitness measurements

We now search for formulas of type: $I(\frac{\partial pH}{\partial t}, \frac{\partial la}{\partial t}, \frac{\partial Km}{\partial t}, \frac{\partial Ba}{\partial t})$ with real outputs mapped to binary outputs, via a sign filtering: $(I() > 0) \rightarrow 1$ and $(I() \leq 0) \rightarrow 0$. The functions (except logical ones) and terminal sets, as well as the genetic operators, are the same as in the global approach above.

Using the available samples of the learning set, four values can be computed, in order to measure the capability of an individual I to characterize each phase:

$$k \in \{1,2,3,4\} \quad F_k(I) = 3 \sum_{i, phase=k} \frac{I(sample(i))}{\#Samples_{phase=k}} - \sum_{i, phase \neq k} \frac{I(sample(i))}{\#Samples_{phase \neq k}}$$

i.e. if I is good for representing phase k, then $F_k(I) > 0$ and $F_{\neq k} < 0$

The local fitness value, *to be maximised*, is a combination of three factors:

$$LocalFit = \max\{F_1, F_2, F_3, F_4\} \times \frac{\#Ind}{\#IndPhaseMax} \times \left.\frac{NbMaxNodes}{NbNodes}\right|_{\text{if } NbNodes > NbMaxNodes}$$

The first factor is aimed at characterising if individual I is able to distinguish one of the four phases, the second factor tends to balance the individuals between the four phases ($\#IndPhaseMax$ is the number of individuals representing the phase corresponding to the $argmax$ of the first factor and $\#Ind$ is the total number of different individuals in the population) and the third factor is a parsimony factor in order to avoid large structures. $NbMaxNodes$ has been experimentally tuned, and is currently fixed to 15.

4.2 Sharing distance

The set of measurements $\{F_1, F_2, F_3, F_4\}$ provides a simplified representation in $\mathbb{R}^4$ of the discriminant capabilities of each individual. As the aim of a Parisian evolution is to evolve distinct subpopulations, each being adated to one of the four subtasks (i.e. characterize one of the four phases), it is natural to use an euclidean distance in this four dimensional phenotype space, as a basis of a simple fitness sharing scheme [7].

4.3 Aggregation of partial solutions and global fitness measurement

At each generation, the population is shared in 4 classes corresponding to the phase each individual characterises the best (i.e. the argmax of $\max\{F_1, F_2, F_3, F_4\}$ for each individual). The 5% best of each class are used via a voting scheme to decide the phase of each tested sample[2]. The global fitness measures the proportion of correctly classified samples:

$$GlobalFit = \frac{\sum_{i=1}^{learningset} CorrectEstimations}{\#Samples}$$

The global fitness is then distributed as a multiplicative bonus on the individuals who participated in the vote:

$$LocalFit' = LocalFit \times (GlobalFit + 0.5)^{\alpha}$$

As $GlobalFit \in [0, 1]$, multiplying by $(GlobalFit + 0.5) > 1$ corresponds to a bonus. The parameter α varies along generations, for the first generations (a third of the total number of generations) $\alpha = 0$ (no bonus), and then α linearly increases from 0.1 to 1, in order to help the population to focus on the four peaks of the search space.

Two sets of indicators are computed at each generation (see section 5, third line of figure 2):

[2] This scheme may also yield a confidence level of the estimation. This measurement is not yet exploited but can be used in future developments of the method.

– the sizes of each class, that show if each phase is equally characterised by the individuals of the population.
– the discrimination capability of each phase, computed on the 5% best individuals of each class as the minimum of:

$$\Delta = \max_{i \in [1,2,3,4]} \{F_i\} - \frac{\sum_{k \neq argmax\{F_i\}} \{F_k\}}{3}$$

5 Experimental analysis

Available data have been collected from 16 experiments during 40 days each, yielding 575 valid measurements.[3] The derivatives of pH, la, *Km* and *Ba* have been averaged and interpolated (spline interpolation) for some missing days. Logarithms of these quantities are considered.

Table 2. Parameters of the GP methods

	GP	**Parisian GP**
Population size	1000	1000
Number of generations	100	50
Function set	arithmetic and logical functions	arithmetic functions only
Sharing	no sharing	$\sigma_{share} = 1$ on the first third of generations, then linear decrease from 1 to 0.1 $\alpha_{share} = 1$ (constant)

The parameters of both GP methods are detailed in table 2. The code has been developed in Matlab, using the GPLAB toolbox[10]. Comparative results of the four considered methods (multilinear regression, Bayesian network, GP and Parisian GP) are displayed in figure 1, and a typical GP run is analysed in figure 2.

The multilinear regression algorithm used for comparison works as follows: the data are modeled as a linear combination of the 4 variables:

$$\widehat{Phase}(t) = \beta_1 + \beta_2 \frac{\partial pH}{\partial t} + \beta_3 \frac{\partial la}{\partial t} + \beta_4 \frac{\partial Km}{\partial t} + \beta_5 \frac{\partial Ba}{\partial t}$$

The 5 coefficients $\{\beta_1, \ldots, \beta_5\}$ are estimated using a simple least square scheme.

Experiments show that both GP outperform multilinear regression and Bayesian network approaches in terms of recognition rates. Additionally the analysis of a typical GP run (figure 2) shows that much simpler structures are evolved: The average size of evolved structures is around 30 nodes for the classical GP approach and between 10 an 15 for the Parisian GP.

It has also to be noted in figure 2 that co-evolution is balanced between the four phases, even if the third phase is the most difficult to characterize (this is in accordance with human experts' judgement, for which this phase is also the most ambiguous to characterize).

[3] The data samples are relatively balanced except for phase 3, which has a longer duration, thus a larger number of samples: We got 57 representatives of phase 1, 78 of phase 2, 247 of phase 3 and 93 of phase 4.

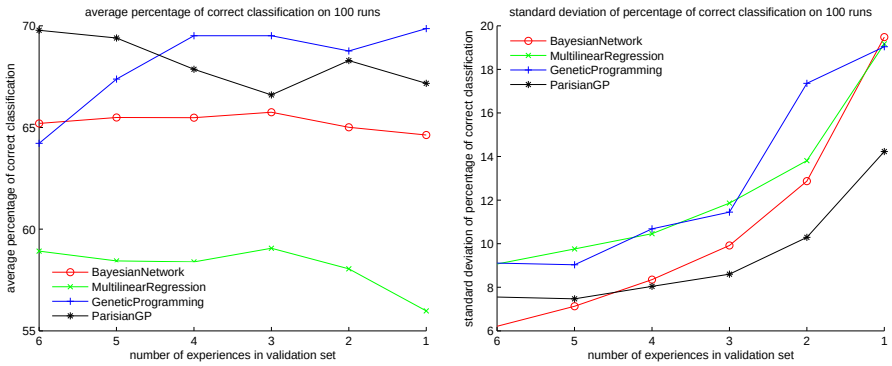

Fig. 1. Average (left) and standard-deviation (right) of recognition percentage on 100 runs for the 4 tested methods, the abscissa represent the size of the test-set

6 Conclusion and future work

This work is a first step toward the use of GP to model complex interactions within a cheese ripening industrial chain. Preliminary results presented in this paper show the effectiveness of GP schemes to capture subjective mechanisms related to human expertise. This point is extremely important for the automation of industrial process as well as for the transmission of expert knowledge.

Additionally, the developement of a cooperative-coevolution GP scheme (Parisian evolution) seems very attractive, as it allows to evolve simpler structure during less generations, and yield results that are easier to interpret. There are however some difficulties to overcome in future developments. First, the computation time is almost equivalent between both presented methods (100 generations of a classical GP against 50 generations of a Parisian one), as one "Parisian" generation necessitates more complex operations, all in all). One can expect a more favourable behaviour of the Parisian scheme on more complex issues than the phase prediction problem, as the benefit of splitting the global solutions into smaller components may be higher and may yield computational shortcuts (see for example [5]). The second difficulty comes from the fact that the Parisian sheme has to be adapted to the problem, it is not obvious for the moment that a convenient sub-problem splitting can be built for other, more complex, prediction problems.

References

1. M. Aldarf, F. Fourcade, A. Amrane and Y. Prigent, Substrate and metabolite diffusion within model medium for soft cheese in relation to growth of Penicillium camembertii. *J. Ind. Microbiol. Biotechnol.*, vol. 33, 685–692, 2006.
2. J. Bongard, H. Lipson, Active Coevolutionary Learning of Deterministic Finite Automata, Journal of Machine Learning Research 6, pp 1651-1678, 2005.
3. D. Barile, J.D. Coisson, M. Arlorio and M. Rinaldi, Identification of production area of Ossolano Italian cheese with chemometric complex aproach, *Food Control*, Volume 17, Issue 3, March 2006, Pages 197-206.

4. C. Baudrit, P-H. Wuillemin, M. Sicard, and N. Perrot, A Dynamic Bayesian Metwork to represent a ripening process of a solf mould cheese. *submitted*.
5. P. Collet, E. Lutton, F. Raynal and M. Schoenauer, Polar IFS + Parisian Genetic Programming = Efficient IFS Inverse Problem Solving, *In Genetic Programming and Evolvable Machines Journal*, Volume 1, Issue 4, pp. 339-361, October, 2000.
6. L. Davis, Adapting Operators Probabilities in Genetic Algorithms. 3rd ICGA conference, 1989, Morgan-Kaufmann pp 61-69.
7. K. Deb and D. E. Goldberg, An investigation of niche and species formation in genetic function optimization, in *Proceedings of the third Conference on Genetic Algorithms*, pages 42–50, 1989.
8. D. I. Ellis, D. Broadhurst and R. Goodacre, Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform infrared spectroscopy and machine learning, *Analytica Chimica Acta*, Volume 514, Issue 2, 1 July 2004, Pages 193-201.
9. A. Gripon, Mould-ripened cheeses. In *Cheese: Chemistry, Physics and Microbiology*, 2, 111-136, 1993 (Ed. PF Fox). London, United Kingdom: Chapman & Hall.
10. S. Silva, GPLAB A Genetic Programming Toolbox for MATLAB, `http://gplab.sourceforge.net/`.
11. I. Ioannou, G. Mauris, G. Trystram, and N. Perrot, Back-propagation of imprecision in a cheese ripening fuzzy model based on human sensory evaluations. *Fuzzy Sets And Systems*, vol. 157, 1179–1187, 2006.
12. I. Ioannou, N. Perrot, C. Curt, G. Mauris, and G. Trystram, Development of a control system using the fuzzy set theory applied to a browning process - a fuzzy symbolic approach for the measurement of product browning: development of a diagnosis model - part I. *Journal Of Food Engineering*, vol. 64, 497–506, 2004.
13. I. Ioannou, N. Perrot, G. Mauris, and G. Trystram, Development of a control system using the fuzzy set theory applied to a browning process - towards a control system of the browning process combining a diagnosis model and a decision model - part II. *J. .Food Eng.*, (64), 507-514, 2004.
14. S. A. Jimenez-Marquez, J. Thibault, and C. Lacroix, Prediction of moisture in cheese of commercial production using neural networks. *Int. Dairy J.*, vol. 15, 1156–1174, 2005.
15. M. N. Leclercq-Perlat, D. Picque, H. Riahi, and G. Corrieu, Microbiological and Biochemical Aspects of Camembert-type Cheeses Depend on Atmospheric Composition in the Ripening Chamber. *J. Dairy Sci*, (89), pp. 3260-3273, 2006.
16. M. N. Leclercq-Perlat, F. Buono, D. Lambert, E. Latrille, H. E. Spinnler, and G. Corrieu, Controlled production of Camembert-type cheeses. Part I: Microbiological and physicochemical evolutions. *J. Dairy Res.*, (71), pp. 346-354, 2004.
17. H. X. Ni and S. Gunasekaran, Food quality prediction with neural networks. *Food Technology*, vol. 52, 60–65, 1998.
18. G. Ochoa, E. Lutton and E. Burke, Cooperative Royal Road Functions, *In Evolution Artificielle*, Tours, France, October 29-31, 2007.
19. B. Pinaud, C. Baudrit, M. Sicard, P-H. Wuillemin, and N. Perrot, Validation et enrichissement interactifs d'un apprentissage automatique des paramètres d'un réseau bayésien dynamique appliqué aux procédés alimentaires, *Journées Francophone sur les Réseaux Bayésiens*, Lyon, France (2008).
20. M. H. Riahi, I. C. Trelea, M. N. Leclercq-Perlat, D. Picque and G. Corrieu, Model for changes in weight and dry matter during the ripening of a smear soft cheese under controlled temperature and relative humidity. *International Dairy Journal*, 17, 946–953, 2007.
21. C.D. Tarantilis and C.T. Kiranoudis, Operational research and food logistics, *Journal of Food Engineering*, Volume 70, Issue 3, October 2005, Pages 253-255.

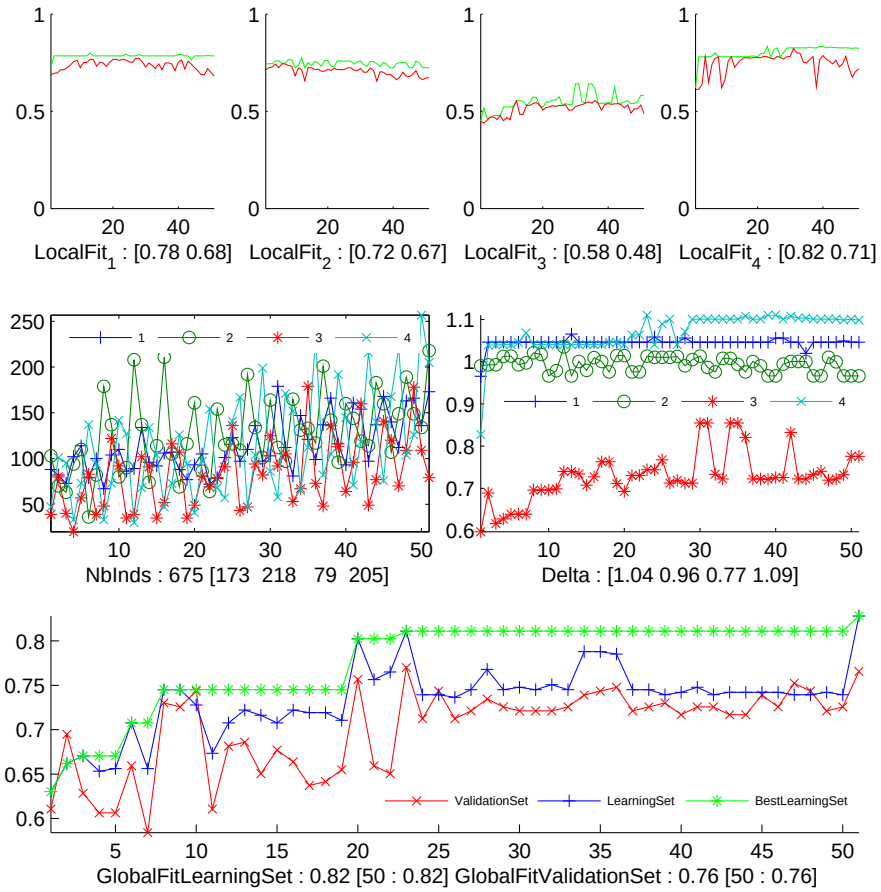

Fig. 2. A typical run of the Parisian GP:
- **First line:** the evolution with respect to generation number of the 5% best individuals for each phase: the upper curve of each of the four graphs is for the best individual, the lower curve is for the "worst of 5% best" individuals.
- **Second line left:** the distribution of individuals for each phase: the curves are very irregular but numbers of representatives of each phases are balanced.
- **Second line right:** discrimination indicator, which shows that the third phase is the most difficult to characterize.
- **Third line:** evolution of the recognition rates of learning and verification set. The best-so-far recognition rate on learning set is tagged with a star.