Neural Information
Processing Systems
Conference

Towards a Particle Swarm Model Selection algorithm

H. Jair Escalante

Computer Science Department
National Astrophysics, Optics and Electronics
Institute of Mexico

Contents

Model selection in supervised learning

- Model selection. *Estimating the performance of different models in order to choose the (approximate) best one* [Hastie et al 2001]

- Problem: "*there are many statistical and learning methods*", which one is the best for my data?

- Solution: Automated model selection (AMS)
  - Agnostic
  - Prior knowledge

The model selection game

- Organizers provide:
  - Data
    - ADA, GINA, HIVA, NOVA & SYLVA
  - Models
    - Challenge Learning Object Package (*CLOP*)
  - Fair evaluation
    - Large test set
    - Objective comparisons
  - *Travel support!*

Methods

- Naïve trial and error methodology
  - Empirical approach
  - Depends on prior knowledge on the domain
  - As well as on the methods to choose from

- Particle swarm model selection **(still on evaluation)**
  - Reliable models on a few iterations
  - Easy implementation
  - Computationally expensive

Trial and error model selection

Manually!

1.- *For each dataset D:*
Choose a model, with its respective parameters
Obtain the training (and if possible validation) error BER

Or bothered

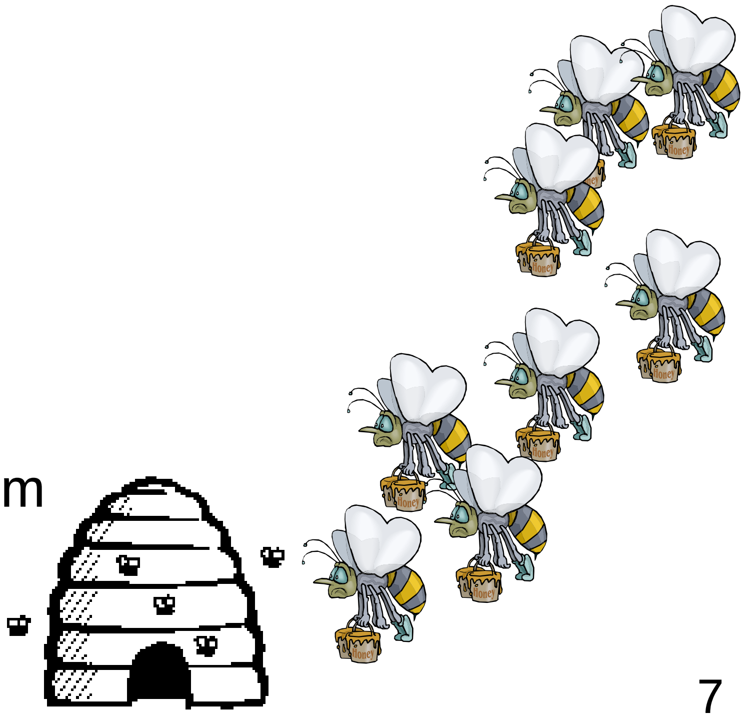
1.1 *repeat until satisfied*
Change model parameters (randomly, or heuristically)
train and test the model on dataset D, using a 10-fold CV approach
obtain the training (and if possible validation) error BER

2.- *save model and change dataset D*

Particle Swarm Model Selection

- Particle swarm optimization
  - Bio-inspired optimization method
  - Particles move towards the optimum
  - Updating positions and velocities
  - Similar to GAs, *though not the same*

- Simple Model
  - Swarm of Particles
  - Position in Solution Space
  - New Position by Random Steps
  - Direction towards current Optimum



Particle Swarm Model Selection

- Particle swarm optimization for model selection
  - Particles: a potential solution to the problem = a model
  - Swarm: population of particles
  - Fitness function: *How far are particles from the solution?* *BER*
  - (Classical) Position updating equations

$$v_{i,j} = w * v_{i,j} + c_1 * r_1 * (p_{i,j} - x_{i,j}) + c_2 * r_2 * (p_{g,j} - x_{i,j}) \quad (1)$$

$$x_{i,j} = x_{i,j} + v_{i,j} \quad (2)$$

  - Stop criteria: Minimum error value or a number of iterations

Particle Swarm Model Selection

- PSMS – Representation: *Models as particles*
  - Each model is represented as a numerical vector

Method l	Parameters For method l	Method 2	Parameters For method 2	. . . . .	Method n	Parameters For method n

$$R_i = [m_1, p_{m_1,1}, \ldots, p_{m_1,x}, \ldots\ldots m_n, p_{m_n,1}, \ldots, p_{m_n,z}]$$

  - Drawback: Method 1 is always performed before method 2

December 9th, 2006

Particle Swarm Model Selection

- PSMS – *fitness function = prediction error measure*
  - Several Options
    - MAE, SRE, Recall, Precision, AUC, …
  - Our choice
    - Balanced Error Rate (BER)

$$BER = \frac{E_+ + E_-}{2}$$

  - Extensions
    - Another prediction accuracy measures
    - Model complexity measures
    - Both of them (MOPSO)
  - Where does the fitness function should be computed?
    - Training set VS K-fold cross validation

Experiments with PSMS

- Experiments
  - Optimal hyper-parameter estimation
  - Method (s) selection and optimal hyper-parameter estimation
  - Full model selection

- PSO settings
  - Standard values for the PSO's parameters
  - Currently up to 200 iterations
  - Swarm size: 10 – 20 particles

Results: our best entry

- Official results (up to December 1st)
  - *BRun2311062*
    - Trial and error: GINA, HIVA, NOVA (Extensive experimentation)
    - PSMS: ADA and SYLVA (100 & 5 iterations, respectively)

Data	**Prep.**	**Classifier**	**Parameters**
Ada^*	slng, std, nmlz	NN	$u = 5, s = 0.008, its = 373$
$Gina^{**}$	nmlz	SVC	$c = 0.1, K = p, d = 5, s = 0.01$
$Hiva^{**}$	std, nmlz	$kridge^{***}$	$c = 1, s = 1, K = 1$
$Nova^{**}$	nmlz	NN^{***}	$u = 1, s = 0.2, its = 50$
$Sylva^*$	std, nmlz	NN	$u = 6, s = 0.028, its = 359$

Results: our best entry

- Official results (up to December 1st)
  - *BRun2311062*
    - Trial and error: GINA, HIVA, NOVA
    - PSMS: ADA and SYLVA

3rd

Dataset	Train BER	Valid BER	Train AUC	Valid AUC
Ada	0.1626	0.1918	0.9167	0.8768
Gina	0	0.0285	1	0.9953
Hiva	0.1153	0.2537	0.959	0.8145
Nova	0.0004	0.044	1	0.997
Sylva	0.0038	0.0045	0.9997	0.9987
Overall	0.0564	0.1045	0.9751	0.9364

Results: our best entry

- Positions

	Dataset	Rank on validation	Rank on test
psms	ADA	11th	7th
	GINA	17th	3rd
	HIVA	14th	2nd
	NOVA	9th	3rd
psms	SYLVA	22nd	7th
	Overall	7th	3rd

Results: our best entry with PSMS

- Current (*best*) results with PSMS **[up to December 7th]**

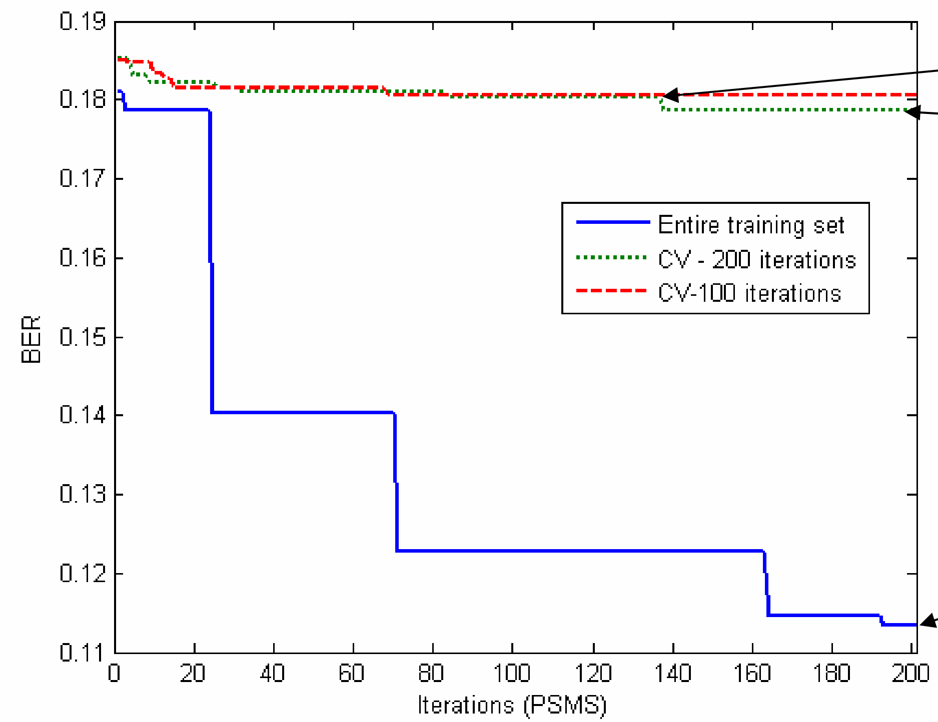
Dataset	Train BER	Valid BER	Iterations	Min. fitness	Best particle	Rank
Ada	0.1626	0.1918	100	0.1806	5;0.081;1;373	11^{th}
Hiva	0.1105	0.2523	85	0.2571	7;0.78;0;675	10^{th}
Sylva	0.0037	0.0033	50	0.025	14;0.29;1;350	1^{st}

- *Rank on test set ?*

Overfitting

- Fitness function on the entire training data or using a k-cross fold validation approach?

Validation error



0.1918

0.2013

0.2562

PSMS for the ADA dataset

Conlusions

- PSMS an alternative to the T&E method
  - Similar results?
  - Agnostic method
  - Computationally expensive, but tractable

- PSMS on development
  - Still on evaluation
  - Here we considered the classical PSO method
  - Further possible extensions

- T&E not a good option, we need:
  - Some knowledge of both: methods and domain
  - A lot of time available
  - Good luck !

Current & future work

- More experiments for evaluating PSMS
  - More iterations
  - Full model selection
  - What k on CV for the fitness function
- Improvements
  - Making efficient PSMS
  - Other error measures
  - Include model complexity in the fitness function
  - MOPSO

Acknowledgments

- Dr. Isabelle Guyon (***Clopinet***), Professor L. Schomaker, and all of the organizers

Thank you!