

Performance of Inductive Method with Unaccounted Interference (case study: polynomial, harmonic, noise)

Natalia Ponomareva¹, Mikhail Alexandrov²

¹University of Wolverhampton, Stafford Street, Wolverhampton, WV1 1SB, UK

²Autonomous University of Barcelona, Barcelona), 08193, Spain

nata.ponomareva@wlv.ac.uk, dyner1950@mail.ru

Abstract. *The noise immunity of the inductive method of model self-organization (IMMSO) has been studied in detail in many publications. The method demonstrates amazing possibilities in case of the correct choice of model class and presence of 'white' noise. But the IMMSO developers give a pessimistic prognosis in the case of using any non-random unaccounted interference. Nevertheless just this case is interested for new IMMSO applications where a priori information about model class is unknown (or almost unknown). In the paper we experimentally elaborate the performance of IMMSO on a simple example of polynomial model class in the presence of unaccounted harmonic component and white noise. Our aim is to obtain qualitative results and explain them using frequency presentations.*

Keywords

Inductive modeling, noise immunity, IWIM 2009

1. Introduction

Inductive method of model self-organization (IMMSO) is a method suggested and developed by acad. A. Ivakhnenko and his colleagues in 80s. It allows to select an optimal model from a given model class to describe a set of experimental data [1,2,4]. By a *model* we mean a formula, equation, algorithm, etc. By *optimal model* we mean a model constructed using the principles of stability for both training and unseen data.

IMMSO demonstrates high noise immunity working with very limited data set and with limited a priori information about model parameters [2, 4]. But this feature is manifested when the selected model class is close to the real model, the experimental data represents. But what to do if we have any unknown unaccounted regular interference in the data under consideration?

Obviously such a situation is equivalent to incompleteness of the model class and the common sense says that in this case the result can not be predicted: we may obtain either oversimplified or overcomplicated model. The developers of IMMSO recommend to reveal and to filter this interference from observation data before starting inductive modeling [2]. Evidently if at least the class of interference is known this class may be added to the main model class and inductive modeling can be accomplished without preliminary filtering unaccounted interference.

In the paper we elaborate the worst case when we do not know neither regular interference itself nor its class of functions. For the definiteness we consider the simplest example: polynomials as a model class, harmonic function as an interference and white noise. The real model hidden in the experimental data is a cubic parabola. In the terms

of communication theory it is a signal to be detected. We carry out series of experiments with distinct harmonics and different noise realizations and we use multicriterion decision-making [3] to find optimal solutions.

2. Experiments

2.1 Data description

As we mentioned above the model class is polynomials $F(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$ given on the interval $[-1,1]$. The experimental data include:

- hidden model (signal): $F_3(x) = 2x^3 - x$;
- interference (unaccounted component): $\varphi(x) = b \sin(k\pi x)$, $b=0.45$;
- white noise: $w(x)$ with the square root deviation $s=0.32$.

It is easy to see that the power of each component is equal to 0.1 approximately. Therefore, the complete ratio noise/signal is equal to 2. The experimental data are calculated in 100 points: the odd ones are taken for training data set, the even ones are taken for control set. Table 1 and Figure 1.1 show the energetic spectrum of signal $F_3(x)$. Figure 1.2 shows the full signal: $\psi(x) = F_3(x) + \varphi(x) + w(x)$ and its components for $k=5$.

Table 1. Energetic spectrum of cubic parabola $2x^3 - x$

Harmonics	0	1	2	3	4	5	6	7	8	9	10
Spectrum	0,000	0,067	0,114	0,094	0,075	0,062	0,053	0,046	0,040	0,036	0,032

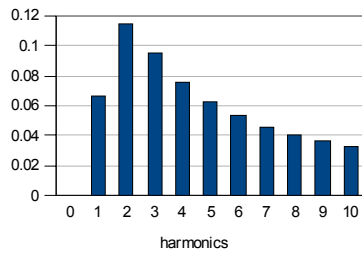


Fig.1.1 Energetic spectrum of signal

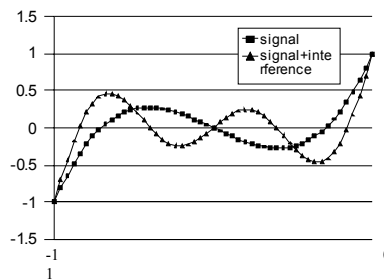


Fig.1.2 Signal and interference

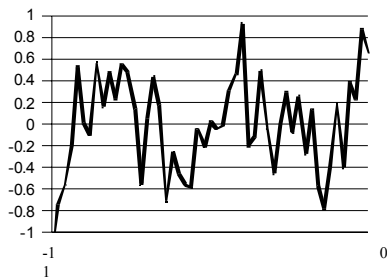


Fig. 1.3 Noisy signal

In our experiments we consider harmonics $k=\{1,2,5,10\}$. It should be pointed out that the first two of them interfere with the main spectral components of the signal and we expect worst results for these harmonics.

2.2. Inductive modeling with a noisy data

The experiments were carried for 5 realizations of white noise. The optimal model was found based on two external criteria: criterion of regularity (K_r) and criterion of unbiasedness (K_u). Tables 2,3,4, and 5 give average values (on 5 noise realizations) of these criteria for harmonics $k=\{1,2,5,10\}$ while Table 1 presents an example of criteria values for a harmonic $k=1$. Figures 2,3,4, and 5 illustrate criteria functions: the higher line refers to K_r , and the lower to K_u . For model selection we use multicriterion decision-making proposed in [3], which consists in consecutive application of two criteria: first, a set of models in the interval close to the global minimum of K_r is selected and, then, an optimal model is chosen according to a minimum value of K_u on the selected set. This strategy gives us the following optimal models: the model of the 1-st order for a harmonic $k=1$, the model of the 5-th order for a harmonic $k=2$, and the model of 3-rd order (the correct model) for harmonics $k=5,10$.

2.3 Inductive modeling without noise

It is interesting to analyze the results of inductive modeling in another extreme situation when white noise is very low, almost equal to 0. Figures 6,7,8 demonstrate the behavior of external criteria for harmonics $k=1,2,10$ in the absence of white noise. The cases $k=5$ and $k=10$ give similar results, so the former was not showed. Figure 9

presents a simplified qualitative dependency of the model complexity on the level of white noise, which was obtained by inductive modeling for different harmonics. In all experiments harmonic has the same amplitude.

Table 2. External criteria for $k=1$

Realizations	0	1	2	3	4	5	6	7	8	9	10
1	1,007	0,617	0,658	0,623	0,631	0,680	0,674	0,675	0,651	0,656	0,639
2	1,000	0,660	0,696	0,691	0,680	0,674	0,677	0,676	0,695	0,678	0,686
3	1,012	0,622	0,629	0,588	0,601	0,597	0,594	0,596	0,600	0,603	0,599
4	1,001	0,677	0,669	0,658	0,655	0,652	0,651	0,654	0,694	0,696	0,730
5	1,000	0,582	0,582	0,559	0,557	0,547	0,551	0,550	0,561	0,608	0,595
Regularity	1,004	0,632	0,647	0,624	0,625	0,630	0,630	0,630	0,640	0,648	0,650
1	0,121	0,095	0,259	0,245	0,268	0,424	0,418	0,421	0,396	0,420	0,408
2	0,020	0,047	0,273	0,272	0,290	0,296	0,303	0,304	0,393	0,377	0,401
3	0,163	0,157	0,188	0,169	0,209	0,203	0,203	0,206	0,211	0,239	0,236
4	0,046	0,075	0,163	0,158	0,167	0,206	0,232	0,238	0,423	0,448	0,486
5	0,020	0,046	0,050	0,063	0,159	0,150	0,174	0,176	0,222	0,339	0,318
Unbiasedness	0,074	0,084	0,187	0,181	0,219	0,256	0,266	0,269	0,329	0,365	0,370

Table 3. Average criteria for $k=2$

	0	1	2	3	4	5	6	7	8	9	10
Regularity	1,002	1,012	1,013	0,927	0,934	0,776	0,790	0,777	0,802	0,797	0,808
Unbiasedness	0,129	0,187	0,197	0,282	0,345	0,390	0,414	0,423	0,474	0,484	0,519

Table 4. Average criteria for $k=5$

	0	1	2	3	4	5	6	7	8	9	10
Regularity	1,003	0,956	0,960	0,756	0,760	0,760	0,763	0,763	0,769	0,747	0,752
Unbiasedness	0,100	0,120	0,150	0,179	0,227	0,264	0,286	0,310	0,323	0,353	0,374

Table 5. Average criteria for $k=10$

	0	1	2	3	4	5	6	7	8	9	10
Regularity	1,000	0,991	1,002	0,861	0,863	0,871	0,884	0,891	0,907	0,911	0,939
Unbiasedness	0,074	0,147	0,229	0,212	0,221	0,263	0,309	0,355	0,396	0,423	0,477

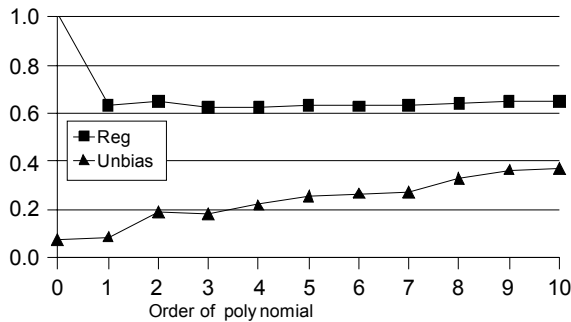


Fig. 2. Criteria for $k=1$ (noisy data)

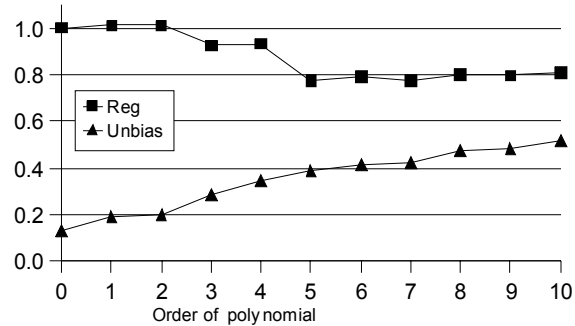


Fig. 3. Criteria for $k=2$ (noisy data)

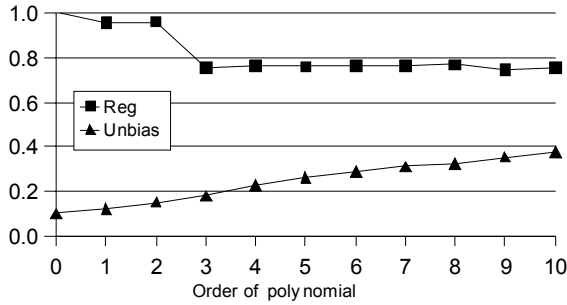


Fig. 4. Criteria for $k=5$ (noisy data)

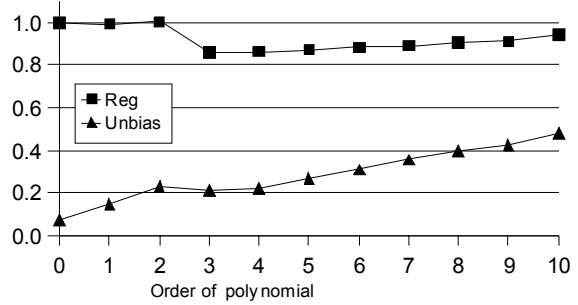


Fig. 5. Criteria for $k=10$ (noisy data)

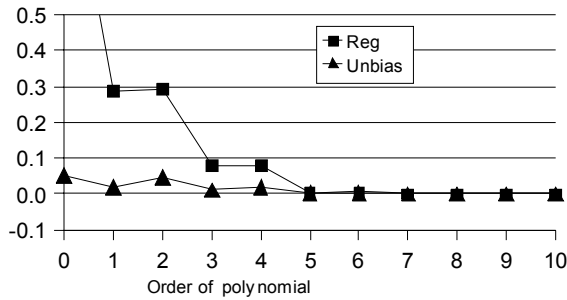


Fig. 6. Criteria for $k=1$ (no noise)

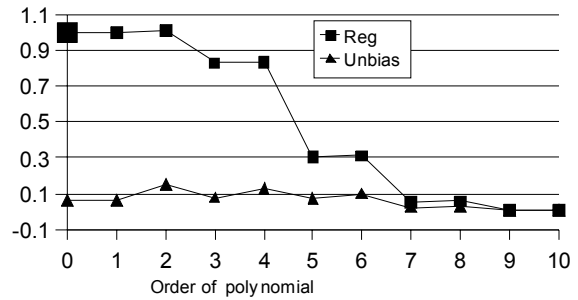


Fig. 7. Criteria for $k=2$ (no noise)

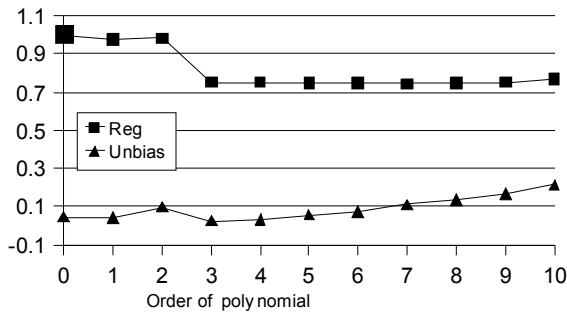


Fig. 8. Criteria for $k=10$ (no noise)

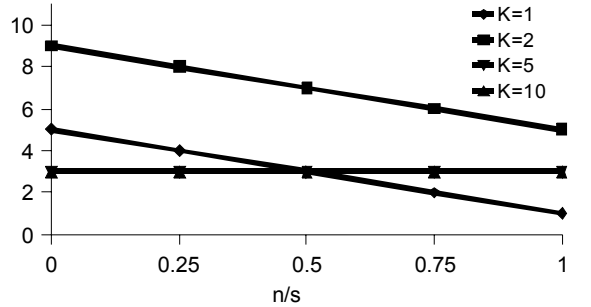


Fig. 9. Change of the model complexity with increase of n/s ratio

3 Conclusions

A non-random unaccounted component of a given model class affects the results very differently.

- If its spectrum belongs to the principal part of the spectrum of the model (signal) to be detected then: a) the results are unpredictable; b) noise simplifies the model essentially stronger than in the case of the full model class.
- Otherwise this component has a role of white noise. So, its influence is well predictable [2,4].

References

- [1] Ivakhnenko, A.: Inductive Method of Model Self-Organization of Complex Systems. *Kiev, Naukova Dumka*, 1982 (rus.)
- [2] Ivakhnenko, A., Stepashko, V.: Noise Immunity of Modeling. *Kiev, Naukova Dumka*, 1985 (rus.)
- [3] Ivakhnenko, A., Savchenko, E.: Investigation of Efficiency of Additional Determination Method of the Model Choice in the Modelling Problems with Application of GMDH Algorithms. *Journ. of Inform. and Automat. Sciences, Begell House Inc*, 2008, Vol. 40, N_3, pp. 4-22
- [4] Stepashko, V.: Method of Critical Variances as an Analytical Tool of the Inductive Modeling Theory. *Journ. of Inform. and Automat. Sciences, Begell House Inc*, 2008, Vol. 40, N_3, pp. 47-58