Some preliminary notes on extending the mimetic technique to Functional Logic Programs

José Hernández-Orallo
JORALLO@DSIC.UPV.ES
DSIC, Universitat Politècnica de València, Spain

April 2006

Abstract

In this paper we discuss the application of the Functional Logic Paradigm to express models using the mimetic technique developed.

1. Introduction

The mimetic technique (Domingos, 1997, 1998; Estruch, Hernández-Orallo & Ramírez, 2003; Blanco, Hernández-Orallo & Ramírez, 2004) generates a model which is similar to the initial model (oracle) but contextualized to the new cost. In order to do this, we propose at least six different ways to diminish the global cost of the mimetic model. Three criteria for adapting the classification threshold, as we have mentioned, and several different schemas for the mimetic technique are set out (without counting on the original data). We have centered our study on binary classification problems (only two classes).

The mimetic method is a technique for converting an incomprehensible model into one simple and comprehensible representation. Basically, it considers the incomprehensible model as an oracle, which is used for labelling an invented dataset. Then, a comprehensible model (for instance, a decision tree) is trained with the invented dataset. The mimetic technique has usually been used for obtaining comprehensible models. However, there is no reason for ignoring it as a cost-sensitive learning since it is in fact a model transformation technique.

Note that the mimetic technique is a transformation technique based on any learning technique since the mimetic model is induced from (invented) data.

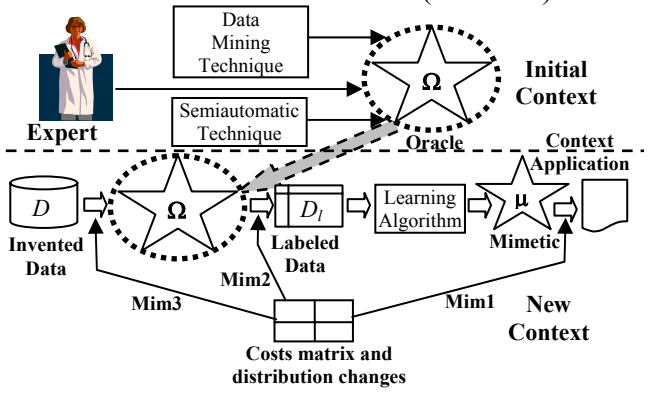

Figure 1. The mimetic context.

The mimetic context validation (see Figure 1) that we propose allows us to change the context of the initial model (oracle) so that it becomes sensible to the new cost.

The paper is organised as follows. ..

2. Inductive Functional-Logic Programming

Inductive Functional Logic Programming is an extension of inductive logic programming. It was developed by Hernández-Orallo & Ramírez 1998, Hernández-Orallo & Ramírez 1999,

provides good estimations of probabilities in order to calculate the threshold from them.

3. Experiments

In this section, we present the global results

3.1 Experimental Setting

For the experiments, we have employed 20 datasets (see Table 1) from the UCI repository (Black & Merz, 1998).

Table 1. Information about the datasets used in the experiments.

No.	Dataset	Balanced	Attributes		Size	Percentage	
			Num.	Nom.		Class 0	Class 1
1	credit-a	Almost	6	9	690	307	383
2	heart-statlog	Almost	13	0	270	150	120
3	monks1	yes	0	6	556	278	278
4	monks3	Almost	0	6	554	266	288
5	monks2	Yes	0	6	412	206	206
6	tic-tac	Yes	0	8	664	332	332
7	breast-cancer	Yes	0	9	170	85	85
8	labor	Yes	8	8	40	20	20
9	vote	Yes	0	16	336	168	168
10	diabetes	Yes	8	0	536	268	268
11	haberman-breast	No	3	0	306	81	225
12	monks2	No	0	6	601	206	395
13	abalone-morethan	No	7	1	4177	1447	2730
14	tic-tac	No	0	8	958	332	626
15	breast-cancer	No	0	9	286	85	201
16	labor	No	8	8	57	20	37
17	vote	No	0	16	435	168	267
18	credit-g	No	7	13	1000	300	700
19	diabetes	No	8	0	768	268	500
20	liver	No	6	0	345	145	200

3.2 General Results

4. Conclusions

In this paper, we have analysed the possible use of functional-logic programming for expressing the models of the mimetic technique...

References

Black C. L.; Merz C. J. (1998). UCI repository of machine learning databases.

Blanco Vega, Ricardo. (2007). Extraction and constextualisation of comprehensible rules from black-box models. Ph.D. Dissertation. Universidad Politécnica de Valencia.

Blanco-Vega R., Hernández-Orallo J., Ramírez-Quintana Ma. J. (2004). Analysing the Trade-off between Comprehensibility and Accuracy in Mimetic Models. 7th ICDS,

Domingos, P. (1999). MetaCost: A general method for making classifiers cost sensitive. In Proceedings of the Fifth Interna-tional Conference on Knowledge Discovery and Data Mining, pages 155–164. ACM Press.

Domingos, P. (1997). Knowledge Acquisition from Examples Via Multiple Models. Proc. of the 14th Int. Conf. on Machine Learning, pp: 98-106.

Domingos, P. (1998). Knowledge Discovery Via Multiple Models. IDA, 2(1-4): 187-202.

Estruch, V.; Ferri, C.; Hernandez-Orallo, J.; Ramirez-Quintana, M.J. (2003). Simple Mimetic Classifiers, Proc. of the Third Int. Conf. on ML & DM, Machine Learning and Data Mining in Pattern Recognition LNCS 2734, pp:156-171, Springer 2003.

Lachiche, Nicolas, Flach, Peter A. (2003) Improving Accuracy and Cost of Two-class and Multi-class Probabilistic Classifiers Using ROC Curves. ICML 2003: 416-423.

Witten, Ian H. and Frank, Eibe. (2005). Data Mining: Practical ML tools with Java implementations. (Second Edition). Morgan Kaufmann.

Hernandez-Orallo, J.; Ramirez-Quintana, MJ. "A Strong Complete Schema for Inductive Functional Logic Programming", 9th International Workshop, Inductive Logic Programming, 1999.

Hernandez-Orallo, J.; Ramirez-Quintana, MJ. "Inverse Narrowing for the Induction of Functional Logic Programs", APPIA-GULP-PRODE, 1998.