Knowledge Acquisition through Machine Learning: Minimising Expert's Effort

Ricardo Blanco-Vega
José Hernández-Orallo
María José Ramírez-Quintana

Agenda

- **Introduction**
- **Trade-off Analysis**
- **Optimisation of the Size of the training set using a Modified MML**
- **Experimental Evaluation**
- **Application Procedure**
- **Conclusions y Future Work**

Introduction

- One of the main problems in expert systems:

knowledge acquisition bottleneck

- Many experts are not able to write down their knowledge:
  - Clear
  - Unambiguous rules.

Introduction

- Expert write down all their knowledge:
  - a high effort
  - can be very time-consuming
  - difficult to maintain and
  - sometimes the result isn´t a model fully automated.

Introduction

Minimising Expert's Effort

- Training a model:
  - captures the expert's knowledge
  - high accuracy
  - high comprehensibility
  - with a minimum number of queries.

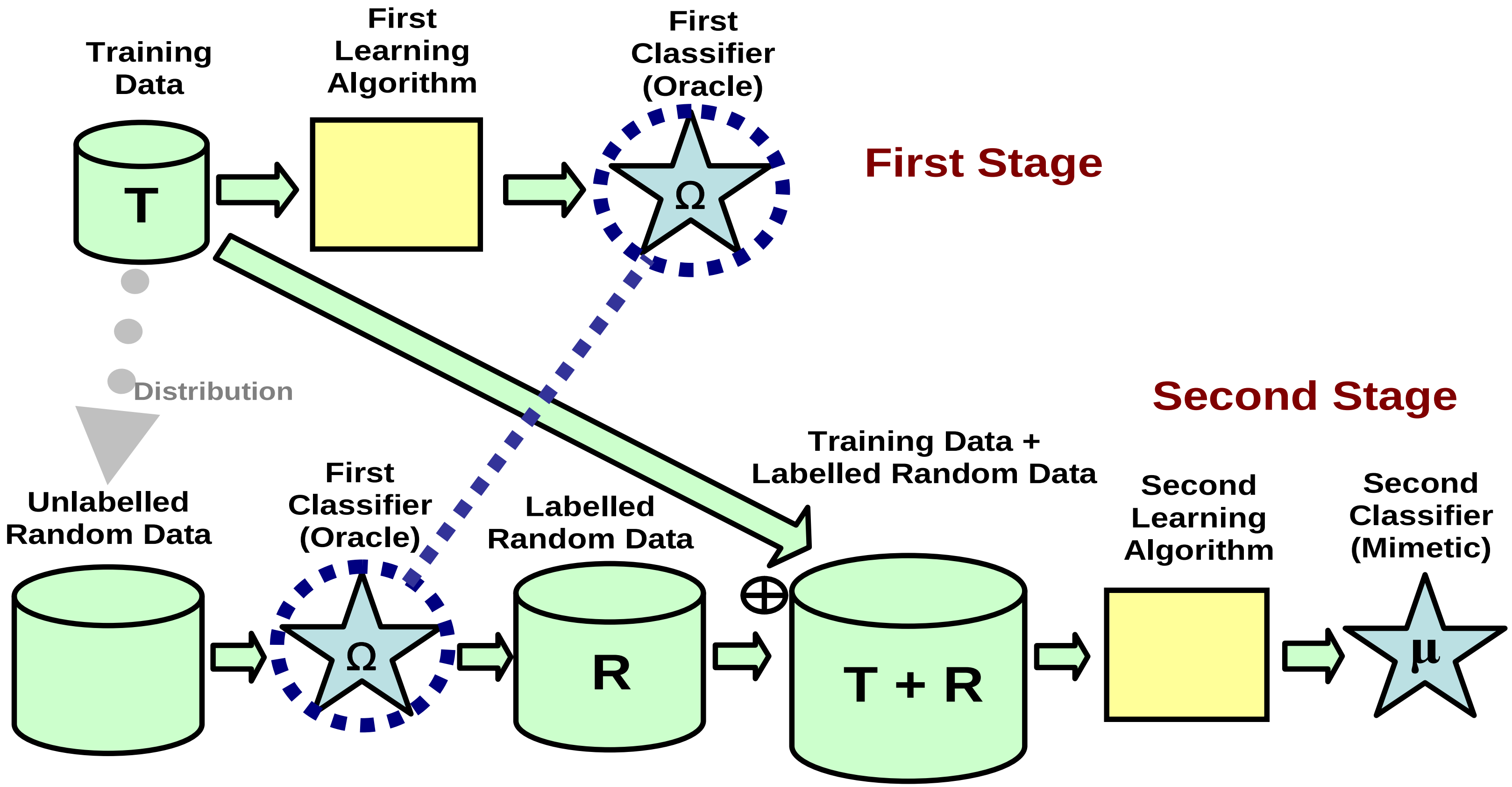
Introduction

- Applications:
  - Diagnosis
  - Estimation
  - Detection,
  - Selection, etc.

  Cases are described by a fixed series of attributes and a dependent value.
- Expert's model predicts the dependent value according to the rest of attributes.
- This model structure is similar to predictive models in machine learning

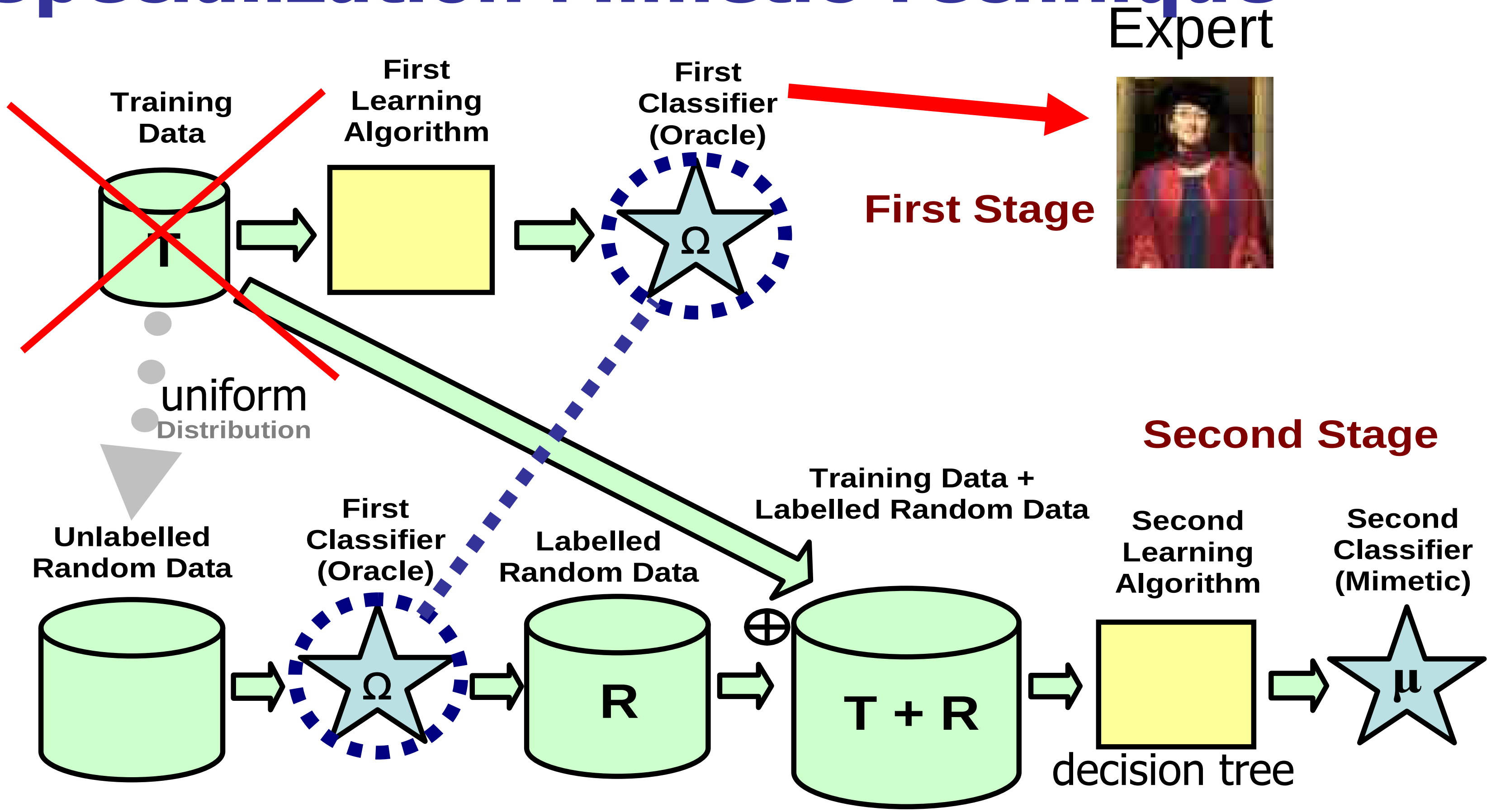
Introduction

Mimetic Technique



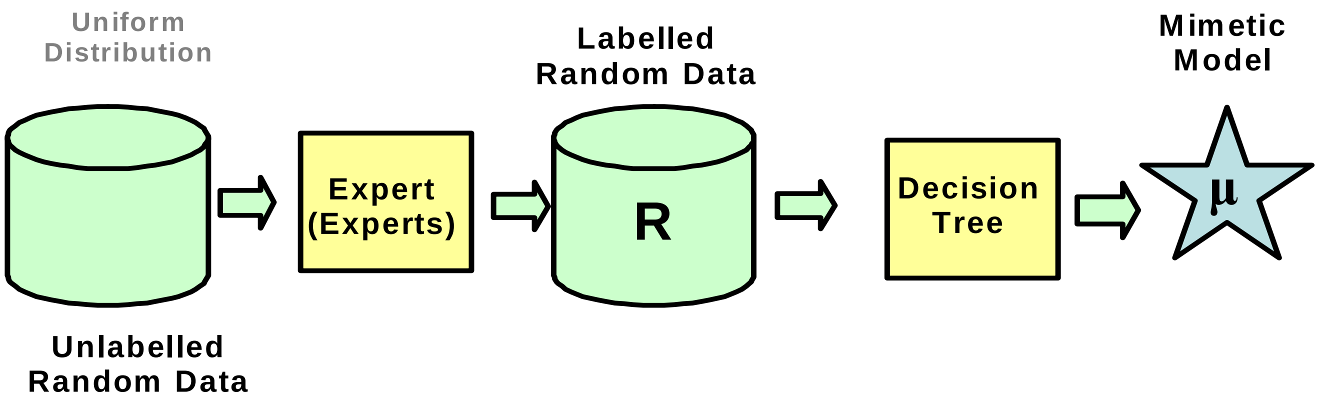
Introduction

Specialization Mimetic Technique

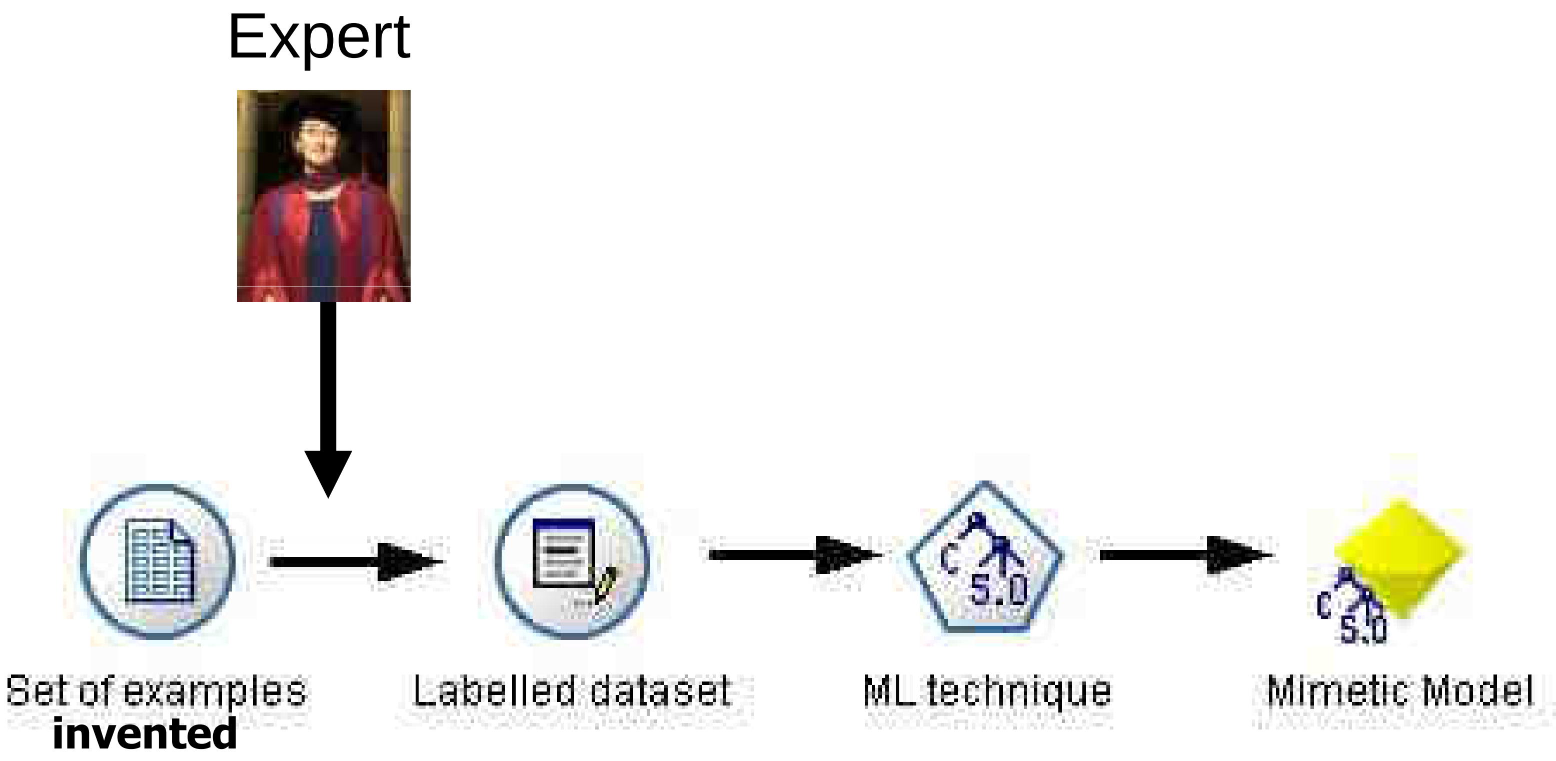


Introduction

Mimetic process with expert oracle



Introduction



Introduction

- Method based on learning curves of mimetic method.
- To predict the number of cases.
- Trade-off between accuracy and comprehensibility of the models.

Trade-off Analysis

size of the invented data increases → accuracy increases

smaller invented datasets → fewer rules (greater comprehensibility)

Trade-off Analysis

The minimum message length (MML) principle

$$\mathrm{MsgLen}(H \cap D) = \mathrm{MsgLen}(H) + \mathrm{MsgLen}(D \mid H)$$

Optimisation of the Size

- This problem can be seen as an optimisation process.
- The objective is to maximise the accuracy and the comprehensibility of the model given some constraints.
- The constraints are obtained by the learning curves for the mimetic model

Optimisation of the Size

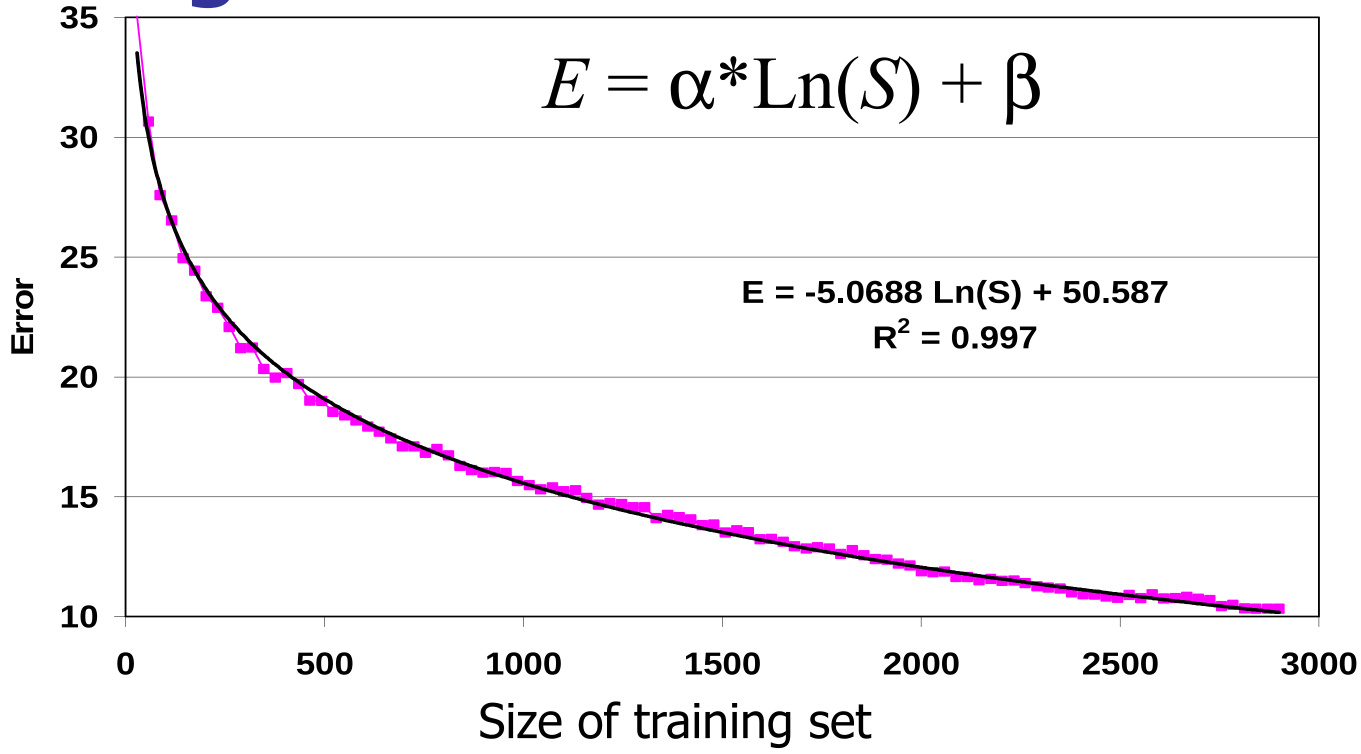
Modified MML

Cost of mimetic model:

$$\mathrm{Cost}(M) = \mathrm{MsgLen}(M) + \mathrm{MsgLen}(D|M) + \mathrm{Query}(D)$$

Optimisation of the Size

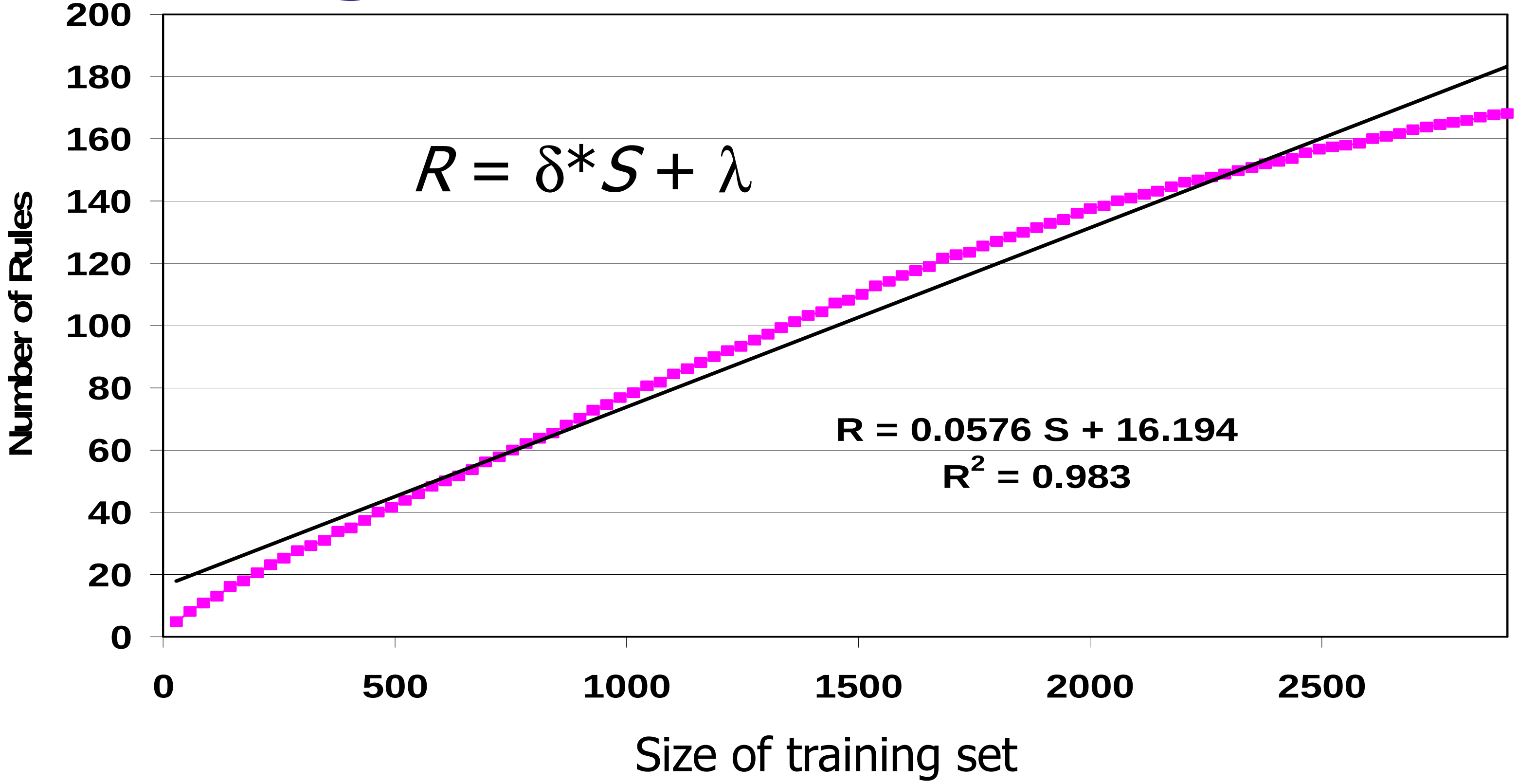
Learning Curves for the Mimetic Models



Error vs. size for the balance-scale dataset

Optimisation of the Size

Learning Curves for the Mimetic Models



Number of rules vs. size for the balance-scale dataset

Optimisation of the Size

Calculating

MsgLen(M) ≈ R* cr

MsgLen(D|M) ≈ E* ce

Query(D) ≈ |D|* cq

Cost(Model) ≈ R* cr + E* ce + |D|* cq

Cost(Model) ≈ R* cr' + E* ce

the second derivative

$Sopt = -K^* \alpha / \delta$

$-K^*\alpha / S^2$

the $Sopt$ value corresponds to a minimum

Experimental Evaluation

No.	Data	Num. Atr.	Nom. Atr.	Classes	Size
1	anneal	6	32	6	898
2	audiology	0	69	24	226
3	balance-scale	4	0	3	625
4	breast-cancer	0	9	2	286
5	cmc	2	7	3	1,473
6	colic	7	15	2	368
7	diabetes	8	0	2	768
8	hayes-roth	0	4	3	132
9	hepatitis	6	13	2	155
10	iris	4	0	3	150
11	monks1	0	6	2	556
12	monks2	0	6	2	601
13	monks3	0	6	2	554
14	sick	7	22	2	3,772
15	vote	0	16	2	435
16	vowel	10	3	11	990
17	waveform-5000	40	0	3	5,000
18	zoo	1	16	7	101

Experimental Evaluation

Parameters and determination coefficients (R^2) for the learning curves with three points (n=3)

Dataset	n=3			
	Error vs Size		Rules vs Size	
	α	R^2	δ	R^2
1	-4.5711	0.97	0.0776	0.98
2	-7.9902	0.99	0.1808	0.99
3	-5.4376	1.00	0.0591	0.99
4	-0.4968	0.74	0.0949	1.00
5	-1.5936	0.93	0.0956	1.00
6	-2.66	0.95	0.0097	0.96
7	-1.2611	0.98	0.0445	1.00
8	-7.197	0.96	0.0699	0.93
9	-3.628	0.98	0.0456	0.99
10	-8.9288	0.97	0.0498	0.98
11	-7.6452	0.95	0.0081	0.49
12	-7.5818	0.95	0.1045	0.97
13	-4.1454	0.94	0.0043	0.73
14	-0.2817	1.00	0.0068	0.99
15	-1.5628	0.99	0.0267	0.99
16	-4.9703	1.00	0.2144	1.00
17	-3.1991	0.99	0.1099	1.00
18	-10.865	0.99	0.1083	0.99
Avg		0.97		0.94

Application Procedure

```
size_set= {10, 20}; // initial number of examples
margin= 0.1; // percentage of error wrt. the optimum size.
i= 20;
while(true) {
  Ask_Expert_Until(i);
  opt= Estimate_Opt_Value(size_set);
   if ((opt < i) || (i/opt > 1 - margin))
     break;
   else {
     i= opt;
     size_set= size_set ∪ { i };
   }
}
```

Application Procedure

An example of the trace

Iteration	*i*	*opt*	*i/opt*
1	20	207	0.1
2	207	340	0.6
3	340	353	0.97 (STOP)

Conclusions

- We have analysed a scenario where knowledge acquisition is made through simple queries to one or more experts.
- Our approach is:
  - practical,
  - easy-to-implement and
  - general (in many situations).

Conclusions

- We need:
  - The expert,
  - some unlabelled data and
  - any machine learning technique.
- We propose a methodology to estimate the number of cases needed to obtain the “optimal” model.

Conclusions

- A step forward in making knowledge acquisition through machine learning much more practical and easy.
- Which can help to solve the knowledge acquisition bottleneck.

Future work

- Grouping similar cases by clustering techniques and then ask the expert to label the clusters.
- Applied for other machine learning methods and other machine learning tasks.