Filtering Multi-instance Problems for Feature Selection in ILP (Extended Abstract)

Erick Alphonse & Nicolas Stroppa
LRI – Bâtiment 490
Université Paris-Sud,
91405 ORSAY CEDEX, France

A recently proposed paradigm [Alphonse and Matwin, 2002] shows how ILP tasks can be filtered in a manner similar to Feature Subset Selection (FSS) in classical Attribute-value learning (AVL). Let us recall that, due to the expressivity of ILP, the need to reduce the dimensionality of its tasks is even more acute than in AVL. In the proposed paradigm, an ILP task is converted into a set of so called Multi-instance Problems (MIPs) [Dietterich et al., 1997], which are a particular setting of AVL problems. This conversion process introduces inherent attribute and class noise.

Here we present an FSS method which is designed specifically to deal with these particular noisy MIP problems. We first show that, in a MI context, the filtering task is equivalent to the learning one. To address the learning task, we use a Bayesian approach based on the idea of Diverse Density proposed by Maron [1998]. In this approach, the quantity we need to maximize is the likelihood function, given by:

$$P(D|h) = \prod_i P(B_i|h) = \prod_i \frac{P(h|B_i) \times P(B_i)}{P(h)} \sim \prod_i P(h|B_i)$$

where the data D consists of the bags B_i and h is the target concept. Note that the distributions $P(h|B_i)$ are chosen to handle attribute noise. An EM algorithm [Zhang and Goldman, 2001, Neal and Hinton, 1998] is also proposed to reduce the needed computation.

The obtained FSS method is used to reduce MIP problems corresponding to the original ILP task, and the results are converted back into the ILP representation, to be given to an ILP learner. Preliminary experiments have been performed on artificial trains problems (100 training examples and 100 test examples), as summarized in table , in which we compare the performance of selected ILP learners given original ILP tasks and their reduced versions. Filtering these problems yields noisy MIPs with up to 30 attributes and 100 instances

Table 1: Compared Performances

Problem	Propal	F. + Propal	#hyp for P.	#hyp for F. + P.	filtering ratio
Conj. Trains	99%	100%	112	2	26.8%
Disj. Trains	91%	94%	1230	245	61.7%

per bag, and therefore is considered a non trivial task in the FSS community. As we can see, filtering disjunctive problems, which involve both class-noise and attribute-noise, is well-handled by our technique. We plan to further investigate the filtering of ILP tasks involving high amount of class-noise, like Mutagenesis [Alphonse and Matwin, 2002].

References

E. Alphonse and S. Matwin. Feature subset selection and inductive logic programming. In *ICML'02 (to appear)*, 2002.

T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1–2):31–71, 1997.

Oded Maron. Learning from ambiguity. Technical Report AITR-1639, 1998.

R. Neal and G. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*. Kluwer, 1998.

Qi Zhang and Sally A. Goldman. Em-dd: An improved multiple-instance learning technique (preliminary version), 2001.