A framework for defining distances between first-order logic objects [1]

Jan Ramon, Maurice Bruynooghe
Katholieke Universiteit Leuven, Department of Computer Science
Celestijnenlaan 200A, B-3001 Heverlee, Belgium
email: {janr,maurice}@cs.kuleuven.ac.be

In learning systems based on clustering (e.g. TIC, KBG) and in instance based learning (e.g. RIBL), a measure of the distance between objects is an essential component. Good measures exist for distances between objects in an attribute value representation . Recently there is a growing interest in using more expressive first order representations of objects and in upgrading propositional learning systems into first order learning systems (e.g. TILDE, ICL, and CLAUDIEN). The upgrading of clustering and instance based learning systems requires to develop a measure for the distance between first order objects.

Some proposals for distance measures between atoms and clauses exists . They use Hausdorff metrics to extend distances between atoms into distances between sets of atoms (clauses or models). This has two drawbacks. Firstly, the value of the Hausdorff metric depends very much on the most extreme value in both sets. Secondly, the similarity due to occurrences of the same subterm (constant, variables, ...) in different atoms of the same clause has no influence on the value. Other authors use rather ad-hoc measures of similarity which do not comply with all axioms of a distance, in particular with the triangle axiom ($d(x,z) \leq d(x,y) + d(y,z)$).

In this paper we develop a framework for distances between clauses and distances between models. The framework can be parametrised by a measure for the distance between atoms. It takes into account subterms common to distinct atoms of a set of atoms in the measurement of the distance between sets. Moreover, for a constant number of variables, the complexity of the distance computation is polynomially bounded by the size of the objects. Initial experiments show that the framework can be the basis of good clustering algorithms.

The framework consists of three levels: At the first level one chooses a distance between atoms .

The second level upgrades this distance to a distance between sets of atoms. We propose a framework that is a generalisation of three polynomial time computable similarity measures proposed by Eiter and Mannila, and an instance which is a real distance function, computable in polynomial time. We develop also a binary prototype function for sets of points. Prototype functions for atoms, are described in [1].

[1]This paper appeared as J. Ramon and M. Bruynooghe. A framework for defining distances between first-order logic object *In* Proceedings of the 8th International Conference on Inductive Logic Programming, *Lecture Notes in Artificial Intelligence, pages 271–28* Springer-Verlag, 1998.

At the third level, we develop a distance function $d_0^{\Theta_r}(A, B)$ which takes into account the "similarity" of the atoms in the sets by computing the minimal distance between a variant of A and a variant of B.

We also did clustering experiments using instances of this scheme, i.e. distance functions between models. We obtained promising results, much better than when data are represented in an attribute value setting, and comparable to other first-order learners.

[1] is a full version of this paper including all proofs.

Acknowledgements

Maurice Bruynooghe is supported by the Fund of Scientific Research, Flanders. This work is supported by the European community Esprit project no. 20237, Inductive Logic Programming 2.

References

[1] J. Ramon and M. Bruynooghe. A framework for defining distances between first-order logic objects. Technical Report CW 263, Department of Computer Science, Katholieke Universiteit Leuven, 1998. `http://www.cs.kuleuven.ac.be/-publicaties/rapporten/CW1998.html`.