Coverage-based semi-distance between Horn clauses

Zdravko Markov[1] and Ivo Marinchev[2]

[1] Department of Computer Science, Central Connecticut State University
1615 Stanley Street, New Britain, CT 06050, U.S.A.
E-mail: markovz@ccsu.edu

[2] Faculty of Mathematics and Informatics, University of Sofia
5 James Bouchier Str., 1164 Sofia, Bulgaria
E-mail: ivo@fmi.uni-sofia.bg

Abstract. In the present paper we use the approach of *height functions* to defining a semi-distance measure between Horn clauses. This appraoch is already discussed elsewhere in the framework of propositional and simple first order languages (atoms). Hereafter we prove its applicability for Horn clauses. We use some basic results from lattice theory and introduce a family of language independent coverage-based height functions. Then we show how these results apply to Horn clauses. We also show an example of conceptual clustering of first order atoms, where the hypotheses are Horn clauses.

1 Introduction

Almost all approaches to inductive learning are based on generalization and/or specialization hierarchies. These hierarchies represent the hypothesis space which in most cases is a partially ordered set under some generality ordering. The properties of partially ordered sets are well studied in lattice theory. One concept from this theory is mostly used in inductive learning – this is the *least general generalization (lgg)* which given two hypotheses builds their most specific common generalization. The existence of an *lgg* in a hypothesis space directly implies that this space is a semi-lattice (the lgg plays the role of infimum). Thus the lgg-based approaches are theoretically well founded, simple and elegant.

Lgg's exist for most of the languages commonly used in machine learning. However all practically applicable (i.e. computable) lgg's are based on *syntactical* ordering relations. A relation over hypotheses is syntactical if it does not account for the background knowledge and for the coverage of positive/negative examples. For example, dropping condition for nominal attributes, instance relation for atomic formulae and θ-subsumption for clauses are all syntactical relations. On the other hand the evaluation of the hypotheses built by an lgg operator is based on their coverage of positive/negative examples with respect to the background knowledge, i.e. it is based on *semantic relations* (in the sense of the inductive task). This discrepancy is a source of many problems, where overgeneralization is the most serious one.

The idea behind the lgg is to make "cautious" (minimal) generalization. However this property of the lgg greatly depends on how similar are the hypotheses/examples used to build the lgg. For example there exist elements in the hypothesis space whose lgg is the top element (empty hypothesis). This is another source of overgeneralization.

An obvious solution of the latter problem is to use a distance (metric) over the hypothesis/examle space in order to evaluate the similarity between the hypotheses/examples. The basic idea is when building an lgg to choose the pair of hypotheses/examples with the minimal distance between them. Thus the lgg will be the minimal generalization over the whole set of hypotheses/examples. Various distance measures can be used for this purpose. The best choice however is a distance which corresponds to the lgg used, that is the pair of the *closest hypotheses* must produce the *minimal lgg*. To ensure this, the distance and the lgg must be well coupled. Ideally such a distance exists in semi-lattices, however it is based on syntactical relations and as we mentioned above the best way to evaluate the similarity between hypotheses is to use semantic relations. This is a typical problem in Inductive Logic Programming ([4]), where the hypotheses are usually Horn clauses which are generated by syntactical operators (e.g. θ-subsumption lgg) and evaluated by coverage-based functions.

In the present paper we use the approach of *height functions* to defining a semi-distance on a join semi-lattice. This appraoch was already discussed for propositional and simple first order languages (atoms) in [3]. Hereafter we prove its applicability for Horn clauses. For this purpose we repeat some of the basic results and further elaborate the notions introduced in [3].

The paper is organized as follows. The next section introduces some basic notions from lattice theory used throughout the paper. Section 3 describes the height-based approach to defining a semi-distance on a join semi-lattice. Section 4 proves the applicability of this appraoch to Horn clauses and Section 5 shows an example of this. Finally Section 6 concludes with a discussion of related approaches and directions for future work.

2 Preliminaries

The discussion in this section follows [3] with some modifications and elaborations (the proofs of the theorems are also skipped).

Definition 1 (Semi-distance, Quasi-metric). A *semi-distance (quasi-metric)* is a mapping $d : O \times O \rightarrow \Re$ on a set of objects O with the following properties $(a, b, c \in O)$:

1. $d(a, a) = 0$ and $d(a, b) \geq 0$.
2. $d(a, b) = d(b, a)$ (symmetry).
3. $d(a, b) \leq d(a, c) + d(c, b)$ (triangle inequality).

Definition 2 (Order preserving semi-distance). A semi-distance $d : O \times O \to \Re$ on a partially ordered set $(O, \preceq)$ is *order preserving* iff for all $a, b, c \in O$, such that $a \preceq b \preceq c$ it follows that $d(a, b) \leq d(a, c)$ and $d(b, c) \leq d(a, c)$

Definition 3 (Join/Meet semi-lattice). A *join/meet semi-lattice* is a partially ordered set $(A, \preceq)$ in which every two elements $a, b \in A$ have an infimum/supremum.

Definition 4 (Diamond inequality). Let $(A, \preceq)$ be a join semi-lattice. A semi-distance $d : A \times A \to \Re$ satisfies the *diamond inequality* iff the existence of $sup\{a, b\}$ implies the following inequality: $d(inf\{a, b\}, a) + d(inf\{a, b\}, b) \leq d(a, sup\{a, b\}) + d(b, sup\{a, b\})$.

Definition 5 (Size function). Let $(A, \preceq)$ be a join semi-lattice. A mapping $s : A \times A \to \Re$ is called a *size function* if it satisfies the following properties:

S1. $s(a, b) \geq 0, \forall a, b \in A$ and $a \preceq b$.
S2. $s(a, a) = 0, \forall a \in A$.
S3. $\forall a, b, c \in A$, such that $a \preceq c$ and $c \preceq b$ it follows that $s(a, b) \leq s(a, c) + s(c, b)$ and $s(c, b) \leq s(a, b)$.
S4. Let $c = inf\{a, b\}$, where $a, b \in A$. For any $d \in A$, such that $a \preceq d$ and $b \preceq d$ it follows that $s(c, a) + s(c, b) \leq s(a, d) + s(b, d)$.

Consider for example the partially ordered set of first order atoms under θ-subsumption. A size function $s(a, b)$ on this set can be defined as the number of different functional symbols (a constant is considered a functional symbol of arity zero) occurring in the substitution θ mapping a onto b ($a\theta = b$). A family of similar size functions is introduced in [1], where they are called a *size of substitution*. Although well defined these functions do not account properly for the variables in the atoms and consequently cannot be used with non-ground atoms.

Theorem 1. Let $(A, \preceq)$ be a join semi-lattice and s – a size function. Let also $d(a, b) = s(inf\{a, b\}, a) + s(inf\{a, b\}, b)$. Then d is a *semi-distance* on $(A, \preceq)$.

A widely used approach to define a semi-distance is based on an order preserving size function and the diamond inequality instead of property $S4$. The use of property $S4$ however is more general because otherwise we must assume that (1) all intervals in the lattice are finite and (2) if two elements have an upper bound they must have a least upper bound (supremum) too. An illustration of this problem is shown in Figure 1, where a_3 is an upper bound of b_1 and b_2 and $e = sup\{b1, b2\}$. Generally the interval $[e, a_3]$ may be infinite or e may not exists. This however does not affect our definition of semi-distance.

Further, a size function can be defined by using the so called *height functions*. The approach of height functions have the advantage that it is based on estimating the object itself rather than its relations to other objects.

Definition 6 (Height function). A function h is called *height* of the elements of a partially ordered set $(A, \preceq)$ if it satisfies the following two properties:

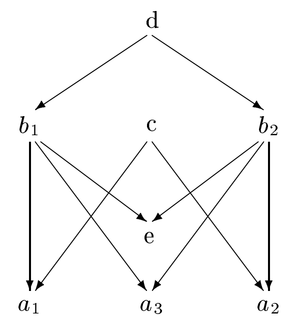

Fig. 1. A semi-lattice structure

H1. For every $a, b \in A$ if $a \preceq b$ then $h(a) \leq h(b)$ (isotone).
H2. For every $a, b \in A$ if $c = inf\{a, b\}$ and $d \in A$ such that $a \preceq d$ and $b \preceq d$ then $h(a) + h(b) \leq h(c) + h(d)$.

Theorem 2. Let $(A, \preceq)$ be a join semi-lattice and h be a height function. Let $s(a, b) = h(b) - h(a), \forall a \preceq b \in A$. Then s is a *size function* on $(A, \preceq)$.

Corollary 1. Let $(A, \preceq)$ be a join semi-lattice and h be a height function. Then the function $d(a, b) = h(a) + h(b) - 2h(inf\{a, b\}), \forall a, b \in A$ is a *semi-distance* on $(A, \preceq)$.

3 Semantic semi-distance on join semi-lattices

In this section we briefly outline the approach to defining a semantic semi-distance on join semi-lattices originally introduced in [3].

Let A be a set of objects and let $\preceq_1$ and $\preceq_2$ be two binary relations on A. Let also $\preceq_1$ be a partial ordering and $(A, \preceq_1)$ – a join semi-lattice.

Definition 7 (Ground elements of a join semi-lattice (GA)). GA is the set of all maximal elements of A w.r.t. $\preceq_1$, i.e. $GA = \{a | a \in A \text{ and } \neg\exists b \in A : a \preceq_1 b\}$.

Definition 8 (Ground coverage). For every $a \in A$ the *ground coverage* of a w.r.t $\preceq_2$ is $S_a = \{b | b \in GA \text{ and } a \preceq_2 b\}$.

The ground coverage S_a can be considered as a definition of the semantics of a. Therefore we call $\preceq_2$ a *semantic relation* by analogy to the Herbrand interpretation in first order logic used to define the semantics of a given term. The other relation involved, $\preceq_1$ is called *constructive (or syntactic) relation* because it is used to build the lattice from a given set of ground elements GA.

The basic idea of our approach is to use these two relations, $\preceq_1$ and $\preceq_2$ to define the semi-distance. According to Corollary 1 we use the syntactic relation $\preceq_1$ to find the infimum and the semantic relation $\preceq_2$ to define the height function

h. The advantage of this approach is that in many cases there exists a proper semantic relation however it is intractable, computationally expensive or even not a partial order, which makes impossible to use it as a constructive relation too (an example of such a relation is logical implication). Then we can use another, simpler relation as a constructive one (to find the infimum) and still make use of the semantic relation (in the height function).

Not any two relations however can be used for this purpose. The following teorem states the necessary conditions for two relations to form a correct height function.

Theorem 3. Let A be a set of objects and let $\preceq_2$ and $\preceq_1$ be two binary relations in A such that:

1. $\preceq_1$ is a partial order and $(A, \preceq_1)$ is a join semi-lattice.
2. For every $a, b \in A$ if $a \preceq_1 b$ then $|S_a| \geq |S_b|$[1].
3. For every $a, b \in A$ and $c = inf\{a, b\}$ such that there exists $d = sup\{a, b\}$ one of the following must hold:
   C1. $|S_d| < |S_a|$ and $|S_d| < |S_b|$
   C2. $|S_d| = |S_a|$ and $|S_c| = |S_b|$
   C3. $|S_d| = |S_b|$ and $|S_c| = |S_a|$

Then there exists a family of *height functions* $h(a) = x^{-|S_a|}$, where $a \in A$, $x \in \Re$ and $x \geq 2$.

Proof.

1. Let $a, b \in A$, $a \preceq_1 b$. Then by the assumptions $|S_a| \geq |S_b|$ and hence $h(a) \leq h(b)$.
2. Let $a, b \in A$, $c = inf\{a, b\}$ and $d = sup\{a, b\}$.
   (a) Assume that C1 is true. Then $|S_d| < |S_a|$ and $|S_d| < |S_b| \Rightarrow |S_a| \geq |S_d| + 1$ and $|S_b| \geq |S_d| + 1 \Rightarrow -|S_a| \leq -|S_d| - 1$ and $-|S_b| \leq -|S_d| - 1$. Hence $h(a) + h(b) = x^{-|S_a|} + x^{-|S_b|} \leq x^{-|S_d|-1} + x^{-|S_d|-1} = 2x^{-|S_d|-1} \leq x.x^{-|S_d|-1} = x^{-|S_d|} = h(d) \leq h(c) + h(d)$.
   (b) Assume that C2 is true. Then $|S_d| = |S_a|$ and $|S_c| = |S_b|$. Hence $h(a) + h(b) = h(c) + h(d)$.
   (c) Assume that C3 is true. Then $|S_d| = |S_b|$ and $|S_c| = |S_a|$. Hence $h(a) + h(b) = h(c) + h(d)$.

4 Coverage-based semi-distance between Horn clauses

Within the language of Horn clauses we use *θ-subsumption* for the constructive relation $\preceq_1$ and *logical implication* (semantic entailment) for the semantic relation $\preceq_2$.

Definition 9 (θ-subsumption). Let a and b be Horn clauses. Then a θ-subsumes b denoted $a \preceq_\theta b$, iff there exist a substitution θ, such that $a\theta \subseteq b$ (the clauses are considered as sets of literals).

[1] Generally an isotone property is required here. However we skip the other case, $|S_a| \leq |S_b|$ since it is analogous.

Under θ-subsumption a set of Horn clauses with same predicates at their heads (same functors and arity) forms a join semi-lattice, where the join operator is the θ-subsumption-based least general generalization (lgg_θ). Further, we will show that θ-subsumption and logical implication can be used to define a correct height function on this semi-latice which in turn implies the existence of a coverage-based semi-distance between Horn clauses.

Definition 10 (Model). A set of ground literals which does not contain a complementary pair is called a *model*. Let M be a model, c – a clause, and C – the set of all ground clauses obtained by replacing the variables in c by ground terms. M is a model of c iff each clause in C contains at least one literal from M.

Definition 11 (Semantic entailment). Let f_1 and f_2 be well-formed formulae. f_1 *semantically entails* f_2, denoted $f_1 \models f_2$ (or $f_1 \preceq_{\models} f_2$) iff every model of f_1 is a model of f_2.

Corollary 2. Let a and b be clauses such that $a \preceq_\theta b$. Then $S_a \supseteq S_b$ and $|S_a| \geq |S_b|$.

Proof. Let a and b be clauses and let a θ-subsumes b. According to Definitions 9 and 10 a semantically entails b, i.e. $a \preceq_{\models} b$. Then according to Definition 8 $S_a \supseteq S_b$ and $|S_a| \geq |S_b|$.

Now we will show that the two assumptions of Theorem 3 hold:

1. Let a and b be clauses and let $a \preceq_1 b$. Then by Corollary 2 $|S_a| \geq |S_b|$.
2. Let $d = sup\{a, b\}$ w.r.t. $\preceq_\theta$. Then $a \preceq_\theta d$, $b \preceq_\theta d$, and by Corollary 2 $|S_d| \leq |S_a|$ and $|S_d| \leq |S_b|$. Further, we will show that actually $|S_d| < |S_a|$ and $|S_d| < |S_b|$. First, we assume that for any two clauses c_1 and c_2 if $S_{c_1} \equiv S_{c_2}$ then $c_1 \equiv c_2$. Thus, in fact instead of clauses we use *equivalence classes* of clauses w.r.t. $\preceq_{\models}$. Let $x \in S_a \triangle S_b$ (symmetric difference). Assume now that $x \in S_d$. Then by Corollary 2 $S_d \subseteq S_a$ and $S_d \subseteq S_b$, that is $x \in S_a \cap S_b$ which is a contradiction. Hence $x \notin S_d$, i.e. $S_d \subset S_a$ and $S_d \subset S_b$, i.e. $|S_d| < |S_a|$ and $|S_d| < |S_b|$.

Then according to Corollary 1 the following function is a semi-distance

$$d(a, b) = x^{-|S_a|} + x^{-|S_b|} - 2x^{-|S_{lgg_\theta(a,b)}|},$$

where a and b are Horn clauses and S_a, S_b and $S_{lgg_\theta(a,b)}$ are models of a, b and $lgg_\theta(a, b)$.

5 Example

To illustrate the semi-distance between Horn clauses we use the inductive algorithm described in [3, 2]. The algorithm starts with a given set of examples (ground atoms) GA and builds a hierarchy of Horn clauses covering this examples (i.e. a partial lattice, where GA is the set of maximal elements of the lattice). The algorithgm is as follows:

1. Initialization: $G = GA$, $C = GA$;
2. If $|C| = 1$ then exit;
3. $T = \{h | h = lgg_\theta(a, b), (a, b) = argmin_{a,b \in C} d(a, b)\}$;
4. $DC = \{h | h \in C$ and $\exists h_{min} \in T : h_{min} \preceq_2 h\}$;
5. $C = C \setminus DC$;
6. $G = G \cup T$, $C = C \cup T$, go to step 2.

We use 10 instances of the `member` predicate and supply them as a GA set to our algorithm. Figure 2 shows the lattice structure built upon this set of examples. The two successors of the top element form the well known definition of *member* (the recursive clause contains a redundant literal). The generated tree structure can be seen as an example of *conceptual clustering of first order atoms*, where the hypotheses are Horn clauses.

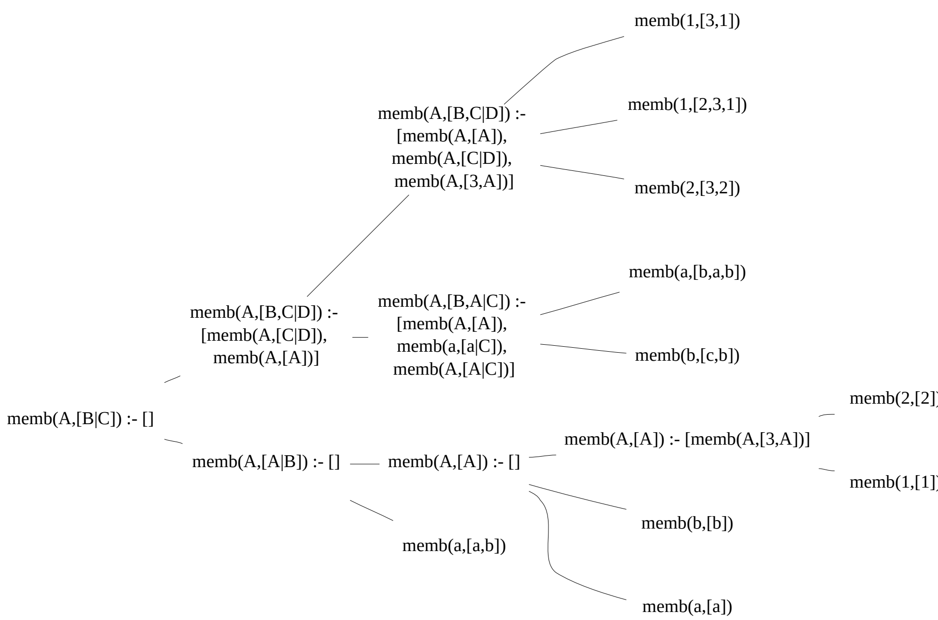

Fig. 2. Hypothesis space for the instances of the `member` predicate.

A major problem in applying our algorithm is the *clause reduction*. This is because although finite the length of the lgg_θ of n clauses can grow exponentially with n. Some well-known techniques of avoiding this problem are discussed in [4]. By placing certain restrictions on the hypothesis language the number of literals in the lgg_θ clause can be limited by a polynomial function independent on n. Currently we use *ij-determined* clauses in our experiments (actually 22-determinated).

6 Conclusion

Distance measures are widely used in machine learning, pattern recognition, statistics and other related areas. Most of the distances in these areas are based on attribute-value (or feature-value) languages and further elaborate well known distances in feature spaces (e.g. Euclidean distance, Hamming distance etc.). Recently a lot of attention has been paid to studying distance measures in first order languages. The basic idea is to apply the highly successful instance based algorithms to relational data described in the much more expressive language of first order logic. Various approaches have been proposed in this area. Some of the most recent ones are [1, 5–7]. These approaches as well as most of the others define a simple metric on atoms and then extend it to sets of atoms (clauses or models) using the Hausdorff metric or other similarity functions. Because of the complexity of the functions involved and problems with the computability of the models these approaches are usually computationally hard. Compared to the other approaches our approach has two basic advantages:

- It is language independent, i.e. it can be applied both within propositional (attribute-value) languages and within first order languages.
- It allows consistent integration of generalization operators with a semantic distance measure. This makes the approach particularly suitable for inductive algorithms, such as the one discussed in Section 5.

We see the following directions for future work:

- Particular attention should be paid to the clause reduction problem when using the language of Horn clauses. Other lgg operators, not based on θ-subsumption should be considered too.
- The practical learning data often involve numeric attributes. In this respect proper relations, lgg's and covering functions should be investigated in order to extend the approach for handling numeric data.
- More experimental work should be done to investigate the applicability of the proposed algorithm in real domains.

References

1. A. Hutchinson. Metrics on terms and clauses. In M. van Someren and G. Widmer, editors, *Machine Learning: ECML-97*, volume 1224 of *Lecture Notes in Artificial Intelligence*, pages 138–145. Springer-Verlag, 1997.
2. Z. Markov. An algebraic approach to inductive learning. In *Proceedings of 13th International FLAIRS Conference*, pages 197–201, Orlando, Florida, May 22-25, 2000. AAAI Press.
3. Z. Markov and I. Marinchev. Metric-based inductive learning using semantic height functions. In R. L. de Mantaras and E. Plaza, editors, *Machine Learning: ECML 2000*, volume 1810 of *Lecture Notes in Artificial Intelligence*, page in print. Springer, 2000.

4. S. Muggleton. Inductive logic programming. In S. Muggleton, editor, *Inductive Logic Programming*, pages 3–28. Academic Press, 1992.
5. S.-H. Nienhuys-Cheng. Distance between herbrand interpretations: a measure for approximations to a target concept. Technical Report EUR-FEW-CS-97-05, Erasmus University, 1997.
6. J. Ramon, M. Bruynooghe, and W. V. Laer. Distance measure between atoms. Technical Report CW 264, Katholieke Universiteit Leuven, 1998.
7. J. Ramon, M. Bruynooghe, and W. V. Laer. A framework for defining a distance between first-order logic objects. Technical Report CW 263, Katholieke Universiteit Leuven, 1998.