Distance measures between atoms

Ramon J.
Bruynooghe M.
Van Laer W.

Report CW264, May 1998
Distance measures between atoms

Ramon J.
Bruynooghe M.
Van Laer W.

Report CW264, May 1998

Department of Computer Science, K.U.Leuven

Abstract

Many learning systems, e.g. systems based on clustering and instance based learning systems, need a measure for the distance between objects. Adequate measures are available for attribute value learners. In recent years there is a growing interest in first order learners, however existing proposals for distances between non-ground atoms have some drawbacks. In this paper we develop a new measure for the distance between non-ground atoms.

Keywords: Machine learning, distances, first order logic.
1 Introduction

In learning systems based on clustering (e.g. C0.5 [3], KBG [1]) and in instance based learning (e.g. [9, ch.4], RIBL [6]), a measure of the distance between objects is an essential component. Good measures exist for distances between objects in an attribute value representation (see e.g. [9, ch. 4]). 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 [2], ICL [5] and CLAUDIEN [4]). Some ad-hoc similarity measures exist for distances between first order objects [6], but they do not have all the desirable mathematical properties (e.g. the triangle inequality and the positive definiteness property)\(^1\), as a consequence their use may lead to sub-optimal/inconsistent results.

A first step in defining a good distance between first order objects is to define a distance between first order atoms. Some proposals exist for this, e.g. [10] and [8], but they do not handle variables in a satisfactory way. In this paper we propose a better measure for distances between non-ground atoms and prove that it has all the desirable properties of a metric. As in [8], the distance between two atoms is based on their distance to the least general generalisation. However, our distances are pairs \((F, V)\), where \(F\) accounts for the differences between the functors of both atoms and \(V\) for the difference due to the variables.

For example, the distance between the atoms \(p(x, x, y)\) and \(p(u, u, v)\) should be 0 as the two atoms are renamings of each other, and \(p(a)\) should be closer to \(p(x)\) than to \(p(b)\) because \(p(x)\) subsumes \(p(a)\) while \(p(a)\) and \(p(b)\) are distinct. None of the existing measures produce a distance which is in agreement with these intuitions.

In section 2, we recall some basic concepts about orders, distances and first order logic. In section 3 we prove that a distance can be derived from a semi-distance which is strictly order preserving and satisfies a so called diamond property. In section 4 we develop such a semi-distance for non-ground atoms. We summarize our achievements and discuss future work in Section 5.

2 Preliminaries

We recall some elementary definitions about order relations.

**Definition 1** \((N, \leq)\) (a set \(N\) equipped with a binary relation \(\leq\)) is a partial order iff \(\leq\) is reflexive, anti-symmetric and transitive. It is a total order iff it is a partial order and \(\forall a, b \in N: (a \leq b) \lor (b \leq a)\).

**Example 1** \(\mathbb{R}, \leq\) is a total ordered set.

Now we define a special case of string-ordering (only for strings of equal length)

**Definition 2** (Order on n-tuples) Let \(N^n\) be the set of n-tuples of elements of \(N\). On \(N^n\) we define an order \(\leq_n\) based on the order \(\leq\) on \(N\): \(\forall(u_1, \ldots, u_n),

\[d(x, y) \geq 0 \text{ and } (d(x, y) = 0 \text{ if } x = y)\]

\(^1\)positive definiteness
(v₁, ..., vₙ) ∈ Nⁿ : (u₁, ..., uₙ) ≤ₙ (v₁, ..., vₙ) ⇔ if n > 1 then (u₁ < v₁) ∨ [(u₁ = v₁) ∧ (u₂, ..., uₙ) ≤ₙ₋₁ (v₂, ..., vₙ)] else u₁ ≤ v₁.

**Definition 3 (monotonic)** Given two partially ordered sets N₁, ≤₁ and N₂, ≤₂ and a function f : N₁ → N₂. We say f is monotonic iff ∀a, b ∈ N₁ : a ≤₁ b ⇒ f(a) ≤₂ f(b), it is anti-monotonic if ∀a, b ∈ N₁ : a ≤₁ b ⇒ f(b) ≤₂ f(a). In addition, it is strict if a <₁ b implies f(a) <₂ f(b).

**Example 2** The function f : R → R with f(x) = 2x is strictly monotonic. The function g : R → R with g(x) = [x] ([x] is the greatest integer smaller than x) is monotonic but not strictly monotonic. The function h : R → R with h(x) = −x is strictly anti-monotonic.

**Definition 4 (convex)** A function N → R is convex iff ∀m, n, k, l : n ≥ k ∧ m ≥ l ⇒ v(n + m) − v(n) ≥ v(k + l) − v(k)

**Example 3** The function p(x) : R → R with p(x) = x² is convex because ∀m, n, k, l : n ≥ k ∧ m ≥ l ⇒ m(2n + m) ≥ l(2k + l) ⇒ n² + 2mn + m² − n² ≥ k² + 2kl + l² − k² ⇒ p(n + m) − p(n) ≥ p(k + l) − p(k)

A distance function is intended to quantify the difference between two objects. One distinguishes two kinds.

**Definition 5 (semi-distance)** A semi-distance d over a set of objects O is a mapping O × O → (N', +, ≤) with (N', +) a commutative group with neutral element 0_N', and ≤ a total order on N' such that ”+” is order-preserving (a ≤ b ∧ c ≤ d ⇒ a + c ≤ b + d), such that ∀a, b, c ∈ O:

1. d(a, a) = 0_N' and d(a, b) ≥ 0_N'.
2. d(a, b) = d(b, a) (symmetry)
3. d(a, c) ≤ d(a, b) + d(b, c) (triangle inequality)

**Definition 6 (distance or metric)** A distance or metric d is a mapping O × O → (N', +, ≤) which is a semi-distance and ∀a, b ∈ O : d(a, b) = 0_N' ⇒ a = b.

**Example 4** In the n-dimensional Euclidean space Eₙ, well known metrics are the euclidean distance (dₑ : Rⁿ × Rⁿ → R) and the manhattan distance (dₘ : Rⁿ × Rⁿ → R). Let x = (x₁, ..., xₙ), y = (y₁, ..., yₙ) ∈ Eₙ.

\[dₑ(x, y) = \sqrt{(x₁ - y₁)² + \cdots + (xₙ - yₙ)²}\]

\[dₘ(x, y) = |x₁ - y₁| + \cdots + |xₙ - yₙ|\]

**Definition 7 (order preserving)** A (semi-)distance d : O × O → N on a partial ordered set O, ≤ is order preserving iff ∀a, b, c ∈ O : a ≤ b ≤ c ⇒ d(a, b) ≤ d(a, c) ∧ d(b, c) ≤ d(a, c).

**Definition 8 (strictly order preserving)** A semi-distance d is strictly order preserving iff it is order preserving and ∀a, b ∈ O : a < b ⇒ d(a, b) > 0_N'.

2
We also recall some terminology from logic. The set of terms \( T \) is built from the set of variables \( V \) and the set of functors \( F \): a variable is a term and, with \( f/n \) a functor and \( t_1, \ldots, t_n \) terms, \( f(t_1, \ldots, t_n) \) is a term. An atom is of the form \( p(t_1, \ldots, t_n) \), with \( p/n \) a predicate and \( t_1, \ldots, t_n \) terms. We denote the set of all atoms with \( \mathcal{A} \).

We will use the notion of position as defined in [7]. Positions are sequences of positive integers (e.g. \([2,3,2]\)), elements of \( \mathbb{N}_+^* \). We use the symbols \( u, v, w, \ldots \) to denote positions. \( \lambda \) denotes the empty position, and \( \cdot \) the concatenation operation on positions. The relation \( \subseteq \) in \( \mathbb{N}_+^* \) defined by \( u \subseteq v \iff \exists w, v = u \cdot w \) is the prefix order. With \( t \) a term or atom, the set of positions of \( t \), \( \text{Occ}(t) \) and the subterm of \( t \) at position \( u \), \( t/u \) are defined as follows:

- If \( t \) is a variable or a constant, then \( \text{Occ}(t) = \{ \lambda \} \) and \( t/\lambda = t \).
- If \( t = f(t_1, \ldots, t_n) \), then \( \text{Occ}(t) = \{ \lambda \} \cup \{ i \cdot u | 1 \leq i \leq n \wedge u \in \text{Occ}(t_i) \} \), \( t/\lambda = t \) and \( t/(i \cdot u) = t_i/u \).

The subset of positions selecting a subterm which is a variable is denoted \( \text{Occv} \), those selecting subterms which are non-variable terms are denoted \( \text{Occp} \).

Let \( S, G \in \mathcal{A} \) be logical objects. \( G \) is more general than \( S \) (\( G \triangleright S \)) iff there is a substitution \( \theta \) such that \( G\theta = S \). It is a preorder which can be used to define an equivalence relation (equivalent modulo renaming). The induced partial order over the quotient set is also denoted \( \triangleright \). The Least General Generalisation of two objects, \( \text{lgg}(A, B) \) is equal to \( G \) iff \( G \triangleright A \) and \( G \triangleright B \) and \( \forall L \in \mathcal{A} : L \triangleright A \land L \triangleright B \Rightarrow L \triangleright G \). The Least Specific Specialisation of two objects, \( \text{lss}(A, B) \) is equal to \( S \) iff \( A \triangleright S \) and \( B \triangleright S \) and \( \forall L \in \mathcal{A} : L \preceq A \land L \preceq B \Rightarrow L \preceq S \). We extend \( A \) with a special atom \( \top \) (treated as a variable) which is more general than all others, so that \( \text{lgg}(A, B) \) is always defined.

\( \text{Var}(t) \) is a predicate which is true iff \( t \) is a variable, \( \text{Vars}(t) \) is a function which returns the set of variables occurring in \( t \). \( \text{mgu}(t_1, t_2) \) denotes the most general unifier of \( t_1 \) and \( t_2 \). Components of substitutions are represented as \( x \rightarrow t \).

### 3 A distance based on a semi-distance

As argued in the introduction, our interest is in a distance between equivalence classes (modulo renaming) of atoms. Following [8] we define a distance based on the notion of Least General Generalisation. In this section we prove a result which holds for any set of logical objects ordered by the more general relation. It shows that a strictly order preserving semi-distance which satisfies a so called diamond inequality can be the basis for a distance.

**Definition 9 (\( d_s \))** Given a mapping size that maps elements of \( \mathcal{O} \) to the \( n \)-dimensional space \( \mathbb{R}^n \). We then define \( d_s(A, B) = |\text{size}(A) - \text{size}(B)| = \max(|\text{size}(A) - \text{size}(B)|, |\text{size}(B) - \text{size}(A)|) \).

**Example 5** Let \( \text{size}(a) = (1, 2, 3) \) and \( \text{size}(b) = (1, 4, 1) \). We have \( (1, 2, 3) <_3 (1, 4, 1) \). \( d_s(a, b) = (1, 4, 1) - (1, 2, 3) = (0, 2, -2) \).
Lemma 1 \( d_s \) is a semi-distance. If size is strictly monotonic or strictly anti-monotonic, then \( d_s \) is strictly order preserving.

Definition 10 (Diamond inequality) Given a partial order \((O, \leq)\) such that \(\text{lss}(A, B)\) always exists. A semi-distance \(d_s : O \times O \to \mathbb{N}\), satisfies the diamond inequality iff the existence of \(\text{lss}(A, B)\) implies that \(d_s(A, \text{lgg}(A, B)) + d_s(\text{lss}(A, B), B) \leq d_s(A, \text{lss}(A, B)) + d_s(\text{ls}(A, B), B)\).

Definition 11 (\(d_l\)) Given a semi-distance \(d_s\) on logical objects. We define \(d_l(A, B) = d_s(A, \text{lgg}(A, B)) + d_s(\text{lgg}(A, B), B)\).

Theorem 1 Let \(d_s\) be a strictly order preserving semi-distance. If \(d_s\) satisfies the diamond inequality then the \(d_l\) based on this \(d_s\) is a distance.

Proof

We must prove that \(d_l\) satisfies the four properties of a distance as given in definitions 5 and 6.

- \(d_l(A, A) = d_s(A, A) + d_s(A, A) = 0\) and \(d_l(A, B) = d_s(A, \text{lgg}(A, B)) + d_s(\text{lgg}(A, B), B) \geq 0\).
- if \(d_l(A, B) = 0\) then \(d_s(A, \text{lgg}(A, B)) = d_s(\text{lgg}(A, B), B) = 0\). Since \(d_s\) is strictly order preserving, we must have \(\text{lgg}(A, B) = A = B\).
- \(d_l(A, B) = d_s(A, \text{lgg}(A, B)) + d_s(\text{lgg}(A, B), B) = d_s(B, \text{lgg}(B, A)) + d_s(\text{lgg}(B, A), A) = d_l(B, A)\).
- The triangle inequality remains to be proven:

\[
d_l(A_1, A_3) \leq d_l(A_1, A_2) + d_l(A_2, A_3)
\]  

Let \(L_{12} = \text{lgg}(A_1, A_2), L_{13} = \text{lgg}(A_1, A_3), L_{23} = \text{lgg}(A_2, A_3)\), and \(L = \text{lgg}(A_1, A_2, A_3)\) (see figure 1).

We start with applying the diamond inequality on \(L_{12}\) and \(L_{23}\):

\[
d_s(L_{12}, L) + d_s(L, L_{23}) \leq d_s(L_{12}, \text{ls}(L_{12}, L_{23})) + d_s(\text{ls}(L_{12}, L_{23}), L_{23})
\]  

We have \(L_{23} \geq \text{ls}(L_{12}, L_{23}) = \text{ls}(\text{lgg}(A_1, A_2), \text{lgg}(A_2, A_3)) \geq A_2\) and similarly, \(L_{12} \geq \text{ls}(L_{12}, L_{23}) \geq A_2\). As \(d_s\) is strictly order preserving, it
functor $f$ associating a set of positive weights

Lemma 2 Distance on a strictly anti-/monotonic function from the set of atoms

$\text{Distance between atoms}$. We will make use of Lemma 1 and base our semi-/distance/ on ground atoms. If we choose all weights $w_{f,j}$ in all examples, then $d_{s}(L, A_{3})$ is a semi-/distance/, so $d_{s}(L, A_{3}) \leq d_{s}(L, A_{1})$ and $d_{s}(L, A_{3}) \leq d_{s}(L, A_{1})$. This allows to simplify the left hand side of equation 3:

$$d_{s}(L, A_{1}) + d_{s}(L, A_{3}) \leq d_{s}(L, A_{1}) + d_{s}(L, A_{3}) + d_{s}(L, A_{2})$$

This allows to simplify the left hand side of equation 3:

$$d_{s}(L, A_{1}) = d_{s}(L, A_{2}) + d_{s}(L, A_{3})$$

We have that $L \geq L_{3}$ and $L \geq L_{1}$ and $L \geq L_{3}$. As $d_{s}$ is strictly order preserving, it holds that $d_{s}(L, A_{1}) \geq d_{s}(L, A_{1})$ and $d_{s}(L, A_{3}) \geq d_{s}(L, A_{3})$. This allows to further reduce the left hand side:

$$d_{s}(L_{13}, A_{1}) + d_{s}(L_{13}, A_{3}) \leq d_{s}(A_{1}, L_{12}) + d_{s}(A_{1}, L_{23}) + d_{s}(A_{3}, A_{23})$$

Applying three times the definition of $d_{s}$, the equation reduces to equation 1 which was to be proven.

$\square$.

4 Distance between atoms

According to theorem 1, if we find a strictly order preserving semi-/distance/ between atoms which satisfies the diamond equality, then we can derive a distance between atoms. We will make use of Lemma 1 and base our semi-/distance/ on a strictly anti-/monotonic function from the set of atoms $A$ to the 2-dimensional space $\mathbb{R}^{2}$. The first component considers the functors and to allow the flexibility of giving different importance to components in different positions, we associate a set of positive weights $w_{f,0}, w_{f,1}, \ldots, w_{f,n}$ with each functor $f/n$ and a set of positive weights $w_{p,0}, \ldots, w_{p,n}$ with each predicate $p/n$.

The definition is parametrised by these weights.

Definition 12 (F-component) $F : A \cup T \rightarrow IR$ is defined as:

$F(t) =$ if $t = p(t_{1}, \ldots, t_{n})$ then $w_{p,0} + \sum_{i=1}^{n} w_{p,i}F(t_{i})$

else if $t = f(t_{1}, \ldots, t_{n})$ then $w_{f,0} + \sum_{i=1}^{n} w_{f,i}F(t_{i})$

else (a variable) 0.

Example 6 We use weights 1 in all examples. $F(f(g(a, x), h(x), y)) = w_{f,0} + w_{f,1}F(g(a, x)) + w_{f,2}F(h(x)) + w_{f,3}F(y) = 1 + w_{g,0} + w_{g,1}F(a) + w_{g,2}F(x) + w_{h,0} + w_{h,1}F(x) + 0 = 2 + w_{a,0} + 0 + 1 + 0 = 4$

Lemma 2 $F$ is strictly monotonic.

Note that we can define a $d_{s}(g(A, B)) = |f(A) - f(B)|$ and a $d_{g}(A, B) = d_{s}(A, \text{gg}(A, B)) + d(\text{gg}(A, B), B)$ on ground atoms. If we choose all weights as follows: $w_{p,0} = w_{f,0} = 1/2, w_{f,n,i} = w_{p,n,i} = 1/2n$, we can prove that $d_{g}$ is equal to the distance on ground atoms defined in $[10]$.

The second component takes into account the variables and their multiplicity. It is parametrised by a function $v : N \rightarrow \mathbb{R}$. The definition makes use of a function $frq(O, x) = \text{the number of positions of variable } x \text{ in the object } O$. 

5
Definition 13 (V-component) \(V : \mathcal{A} \times (\mathcal{A} \cup \mathcal{T}) \rightarrow \mathbb{R}\) is defined as:
\[
V(A, t) =
\begin{cases}
  \text{if } t = p(t_1, \ldots, t_n) \text{ then } \sum_{i=1}^{n} V(A, t_i) \\
  \text{else if } t = f(t_1, \ldots, t_n) \text{ then } \sum_{i=1}^{n} V(A, t_i) \\
  \text{else (a variable) } v(frq(A, t)).
\end{cases}
\]

Lemma 3 \(V(A, A) = \sum_{x \in \text{vars}(A)} frq(A, x).v(frq(A, x))\) with \(\text{vars}(A)\) the set of all variables occurring in \(A\).

Notation 1 In what follows we use \(V(A)\) as a shorthand for \(V(A, A)\).

Definition 14 (Size for atoms) \(\text{size} : \mathcal{A} \rightarrow \mathbb{R}^2\): \(\text{size}(A) = (F(A), V(A))\).

Example 7 Let \(A = f(g(a, x), h(x), y)\). \(\text{size}(A) = (4, v(1) + 2v(2))\) because \(F(A) = 4\) (see example 6) and \(V(A) = V(A, g(a, x)) + V(A, h(x)) + v(A, y) = V(A, a) + V(A, x) + V(A, y) = v(frq(A, x)) + v(frq(A, x)) + v(frq(A, y)) = v(2) + v(2) + v(1) = 2v(2) + v(1)\). We can also apply lemma 3 to calculate \(V(A)\): \(V(A) = frq(A, x).v(frq(A, x)) + frq(A, y).v(frq(A, y)) = v(1) + 2v(2)\).

Lemma 4 If \(v(n)\) is strictly monotonic for \(n \geq 1\) then \(\text{size}\) is strictly anti-monotonic.

\textbf{Proof}
Consider atoms \(A_1\) and \(A_2\) such that \(A_1 \succ A_2\). Let \(\theta = mgd(A_1, A_2)\).

- If \(\theta\) has a component \(x \rightarrow f(t_1, \ldots, t_n)\) then there is at least one position \(u\) such that \(A_1/u = x\) and \(A_2/u = f(t_1, \ldots, t_n)\). For such positions, \(F(A_1/u) < F(A_2/u)\). As a consequence, \(F(A_1) < F(A_2)\) and \(\text{size}(A_1) < \text{size}(A_2)\).

- If \(\theta\) has no such component, then \(F(A_1) = F(A_2)\). However, then \(A_1\) has at least two variables, say \(x\) and \(y\) such that \(x\theta = y\theta\) as \(A_1\) is strictly more general than \(A_2\). So there is at least one positions \(u\) (e.g. the position of \(x\) in \(A_1\)) such that \(frq(A_1, A_1/u) < frq(A_2, A_2/u)\) and thus also \(V(A_1, A_1/u) = v(frq(A_1, A_1/u)) < V(A_2, A_2/u) = v(frq(A_2, A_2/u))\) because \(v\) is strictly monotonic. As a consequence, \(V(A_1) < V(A_2)\) and \(\text{size}(A_1) < \text{size}(A_2)\). \(\square\)

As a consequence of Lemma 1 and Lemma 4 we have also:

Corollary 1 If \(v(n)\) is strictly monotonic for \(n \geq 1\) then \(d_s(A, B) = |\text{size}(A) - \text{size}(B)|\) is a strictly order preserving semi-distance.

Next we analyse which property the function \(v\) must have to ensure that \(d_s\) also satisfies the diamond inequality. First we prove some lemmas. In these lemma’s, we assume \(A\) and \(B\) unifiable atoms and adopt the following notations: \(G = lgg(A, B), S = lss(A, B)\).

Lemma 5 \(d_s\) satisfies the diamond inequality iff \(\text{size}(A) + \text{size}(B) \leq \text{size}(G) + \text{size}(S)\).
**Proof**

Using the definition of $d_s$, the diamond inequality reduces to:

$$|\text{size}(A) - \text{size}(G)| + |\text{size}(B) - \text{size}(G)| \leq |\text{size}(A) - \text{size}(S)| + |\text{size}(B) - \text{size}(S)|$$

As $A \preceq G$ we have $\text{size}(A) \geq \text{size}(G)$ and similarly $\text{size}(B) \geq \text{size}(G)$, $\text{size}(S) \geq \text{size}(A)$, and $\text{size}(S) \geq \text{size}(B)$.

This allows to rewrite the inequality as:

$$\text{size}(A) - \text{size}(G) + \text{size}(B) - \text{size}(G) \leq \text{size}(S) - \text{size}(A) + \text{size}(S) - \text{size}(B)$$

or

$$\text{size}(A) + \text{size}(B) \leq \text{size}(G) + \text{size}(S)$$

$\square$

**Notation 2** With $Q$ an atom and $u$ a position: $U_Q(u) = \{v | v \in \text{Occ}(Q) \text{ and } Q/v = Q/u\}$, i.e. the set of all positions which select the same subterm as $u$.

**Lemma 6** Let $A$ and $B$ be unifiable atoms without common variables. Then $F(A) + F(B) \leq F(G) + F(S)$. Moreover, if $F(A) + F(B) = F(G) + F(S)$ then $\forall u : (u \in \text{Occ}(G) \Rightarrow (F(S/u) = F(A/u) \lor F(S/u) = F(B/u)))$.

**Proof**

Differences in value between $F(A), F(B), F(G)$, and $F(S)$ are due to the different values of $F$ at positions $v \in \text{Occ}(G)$. Now choose such a position $v \in \text{Occ}(G)$.

Let $U = U_S(v) \cap \text{Occ}(G)$. As for all $u \in \text{Occ}(G)$, $A/u$ and $B/u$ are unifiable, and their generalisation is a variable, at least one of them is a variable. Therefore, we can assume $U = \{u_1, \ldots, u_k, u_{k+1}, \ldots, u_l, u_{l+1}, \ldots, u_n\}$ such that $0 \leq k \leq l \leq n + 1$ and $\{u_1, \ldots, u_k\} \subseteq \text{Occ}(A) \cap \text{Occ}(B)$, $\{u_{k+1}, \ldots, u_l\} \subseteq \text{Occ}(B) \cap \text{Occ}(B)$, and $\{u_{l+1}, \ldots, u_n\} \subseteq \text{Occ}(A) \cap \text{Occ}(B)$.

We have $\sum_{i=1}^{n} (F(G/u_i) + F(S/u_i)) = \sum_{i=1}^{n} F(S/u_i)$ as $G/u_i$ is a variable. Also $\sum_{i=1}^{n} (F(A/u_i) + F(B/u_i)) = \sum_{i=1}^{k} F(A/u_i) + \sum_{i=l+1}^{n} F(B/u_i)$.

Because $A/u_i \succeq S/u_i$, we have $F(A/u_i) \leq F(S/u_i)$ and similarly $F(B/u_i) \leq F(S/u_i)$. As this holds for all $v$ and corresponding $U$, we have $F(A) + F(B) \leq F(G) + F(S)$.

Moreover, if the equality holds, then for $1 \leq i \leq k : F(A/u_i) = F(S/u_i) \land F(B/u_i) = 0$, $k + 1 \leq i \leq l : F(A/u_i) = F(B/u_i) = F(S/u_i) = 0$, $l + 1 \leq i \leq n : F(A/u_i) = 0 \land F(B/u_i) = F(S/u_i) \lor F(A/u_i) = F(S/u_i) \lor F(B/u_i) = F(S/u_i)$.

All choices of $v$ and $u_i$, either $F(A/u_i) = F(B/u_i) = F(S/u_i) = 0$ (if $k = 0$ and $l = n + 1$) or $F(A/u_i)$ or $F(B/u_i) = F(S/u_i) \lor F(A/u_i) = F(S/u_i)$.

$\square$

**Lemma 7** Let $A$ and $B$ be unifiable atoms without common variables. Given: $\text{Occ}(G) = \text{Occ}(S)$. If $v$ is convex, then $V(A) + V(B) \leq V(G) + V(S)$.

**Proof**

Set $A_G = A$, $S_B = S$ and observe that the following properties hold:

$$\text{Occ}(A) = \text{Occ}(S) = \text{Occ}(S_B) = \text{Occ}(A_G).$$

$$V(A) + V(S_B) \leq V(A_G) + V(S).$$

(0)

(1)
\[ S \preceq A \preceq AG \preceq G, \quad S \preceq S_B \preceq B \preceq G, \quad AG = \text{lgg}(A, S_B), \quad \text{and} \quad S = \text{ls}(A, S_B). \]  

(2) \quad \forall u \in \text{OccV}(G) : U_{A_G}(u) \subseteq U_{S_B}(u)

If \( S_B \) is a renaming of \( B \) then \( A_G \) is a renaming of \( G \) and the lemma trivially holds. Otherwise, consider the following piece of code:

\[
\text{while} \quad S_B \prec B \quad \text{do} \\
\quad S_{B_o} = S_B; \quad A_{G,o} = A_G; \\
\quad \text{Select positions } u_0, v_0 \in \text{OccV}(G) \text{ such that} \\
\quad S_{B_o}/u_0 = S_{B/o}/u_0 \text{ and } B/u_0 \neq B/v_0; \\
\quad \text{Let } x, y \text{ be fresh variables;} \\
\quad S_B = S_{B,o} \text{ except at positions } u \in U_{S_{B,o}}(u_0) \text{ where} \\
\quad S_B/u = (\text{if } B/u = B/u_0 \text{ then } x \text{ else } y); \\
\quad A_G = \text{lgg}(A, S_B)
\]

Figure 2: diagram for lemma 7

Observe that \( u_0 \) and \( v_0 \) always exist as \( S_B \prec B \). We show that the while loop preserves the properties (0) - (3). This is fairly straightforward for (0), (2), and (3). Concerning (1), the differences in the \( V \)-value of \( S_B \) and \( S_{B,o} \), \( A_G \) and \( A_{G,o} \), are due to the \( \text{V-values at the positions of } U_{S_{B,o}}(u_0) \). Let \( frq(S_{B,o}, S_{B,o}/u_0) = f \), \( frq(S_B, x) = f_x \), and \( frq(S_B, y) = f_y \) with \( f = f_x + f_y \). Because of (3), there exists sets \( U_i \), \( i = 1, \ldots, k \) which are a partition of \( U_{S_{B,o}}(u_0) \) and such that for each \( i \): \( \exists u_i \in U_{S_{B,o}}(u_0) : U_i = U_{A_{G,o}}(u_i) \). Let \( frq(A_G, A_G/u_i) = f_i (\sum_{i=1}^k f_i = f) \). In \( A_G \), the lgg of \( A \) and \( S_B \) each set of positions \( U_{A_{G,o}}(u_i) \) is split in two sets, a set with size \( f_x,i \geq 0 \) where the term is a fresh variable \( x_i \) and a set with size \( f_y,i \geq 0 \) where the term is a fresh variable \( y_i \) \((f_x,i + f_y,i = f_i)\). (1) holds at the beginning of the loop, i.e. \( V(A) + V(S_{B,o}) \leq V(A_{G,o}) + V(S) \). To prove that (1) is preserved, it suffices to show that \( V(S_B) - V(S_{B,o}) \leq V(A_G) - V(A_{G,o}) \) or

\[
\sum_{u \in U_{S_{B,o}}(u_0)} V(S_B, S_B/u) - \sum_{u \in U_{S_{B,o}}(u_0)} V(S_{B,o}, S_B/o/u) \leq \sum_{u \in U_{S_{B,o}}(u_0)} V(A_G, S_B/o/u) - \sum_{u \in U_{S_{B,o}}(u_0)} V(A_{G,o}, S_B/o/u)
\]

This can be written as:

\[
\begin{align*}
\sum_{i=1}^{k} f_x,i v(f_x,i) + f_y,i v(f_y,i) - f_v(f) & \leq \sum_{i=1}^{k} [f_x,i v(f_x,i) + f_y,i v(f_y,i)] - \sum_{i=1}^{k} f_i v(f_i) \\
\sum_{i=1}^{k} f_x,i v((\sum_{j=1}^{k} f_x,i) + \sum_{j=1}^{k} f_y,i) v((\sum_{j=1}^{k} f_x,i) + (\sum_{j=1}^{k} f_y,i)) - (\sum_{j=1}^{k} f_x,i + \sum_{j=1}^{k} f_y,i) v((\sum_{j=1}^{k} f_x,i) + (\sum_{j=1}^{k} f_y,i)) \\
\sum_{i=1}^{k} f_y,i v((\sum_{j=1}^{k} f_x,i) + f_y,i) v(f_x,i) & \leq \sum_{i=1}^{k} f_x,i v(f_x,i) + f_y,i v(f_y,i) - \sum_{i=1}^{k} (f_x,i + f_y,i) v(f_x,i + f_y,i) \\
\sum_{i=1}^{k} f_x,i v((\sum_{j=1}^{k} f_x,i) + \sum_{j=1}^{k} f_y,i) - v((\sum_{j=1}^{k} f_x,i) + (\sum_{j=1}^{k} f_y,i)) & \leq 0
\end{align*}
\]

which is true when

\[
\begin{align*}
v(\sum_{i=1}^{k} f_x,i + \sum_{i=1}^{k} f_y,i) - v(\sum_{i=1}^{k} f_x,i) & \geq v(f_x,i + f_y,i) - v(f_x,i) \\
v(\sum_{i=1}^{k} f_x,i + \sum_{i=1}^{k} f_y,i) - v(\sum_{i=1}^{k} f_x,i) & \geq v(f_x,i + f_y,i) - v(f_y,i) 
\end{align*}
\]

Both inequalities follow from the fact that \( v \) is convex. As a term can only have a finite number of
generalisations, at some point, \( S_B = B \) and, from (2), \( A_G = G \). This completes the proof.

\[ \]

\textbf{Notation 3} \( t[x_1, \ldots, x_i] \) represents a term \( t \) with \( \text{Vars}(t) = \{x_1, \ldots, x_i\} \).

\textbf{Lemma 8} Let \( A, B, \) and \( G = \text{lglg}(A, B) \) be atoms which differ only at a finite set of positions \( U \) in such a way that for \( i \in \{1, \ldots, k\}, j \in \{1, \ldots, n_i\} \):
\[ (1) \ A/u_i,j = t[x_1, \ldots, x_i] \theta_i \] where \( \theta_i \) are variable to variable substitutions, \( (2) \) \( B/u_i,j = y_i \) if \( i = 1 \), \( (3) \) \( G/u_i,j = y_i \) for \( i \neq j \), \( (4) \) \( x_i \theta_i \neq t[x_1, \ldots, x_i] \theta_j \) and that for \( i, j \in \{1, \ldots, k\}, p, q \in \{1, \ldots, l\}, p \neq q \).

\[ \text{The positions relevant for the difference in } n \text{ are } i \in \{1, \ldots, k\}, j \in \{1, \ldots, n_i\} : S/u_i,j = y_i \text{ also variable to variable substitutions. If } v \text{ satisfies} \]
\[ v(n + m) - v(n) \geq v(k + l) - v(k) \]

for natural numbers \( n, m, k, \) and \( l \) such that \( n \geq k \) and \( m \geq l \) (i.e. \( v \) is convex), then \( V(B) - V(S) \leq V(G) - V(A) \).

\textit{Proof} As differences in \( V \)-value between \( A, B, G, \) and \( S \) are due to differences in the positions \( U \), it suffices to prove \( \sum_{u \in U} V(B, B/u) - \sum_{u \in U} V(S, S/u) \leq \sum_{u \in U} V(G, G/u) - \sum_{u \in U} V(S, S/u) \) or \( \sum_{u \in U} V(A, A/u) \geq \sum_{u \in U} V(B, B/u) - \sum_{u \in U} V(G, G/u) \).

The positions relevant for the difference in \( V \)-value between \( B \) and \( G \) are:
\[ u_1, \ldots, u_{n_1}, u_2, u_{n_2}. \] We have:
\[ \sum_{u \in U} V(B, B/u) - \sum_{u \in U} V(G, G/u) = (n_1 + n_2) v(n_1 + n_2) - n_1 v(n_1) - n_2 v(n_2). \]

\textbf{Theorem 2} Let \( A \) and \( B \) be unifiable atoms without common variables. If \( v \) is convex, then \( d_v \) satisfies the diamond inequality.

\[ \]

\[ ^3 \text{E.g. the atoms } A_G, G, G_1, \text{ and } A_G_1 \text{ in Figure 3.} \]
We have that size(A) + size(B) ≤ size(G).

(1)

It follows from Lemma 6 that F(A) + F(B) ≤ F(G) + F(S). If F(A) + F(B) < F(G) + F(S) then size(A) + size(B) ≤ size(G) + size(S) trivially follows. Otherwise, F(A) + F(B) = F(G) + F(S). In this case, we have to prove V(A) + V(B) ≤ V(G) + V(S)

(2)

From lemma 6 then also follows that u ∈ Ocv(G) ⇒ (F(S/u) = F(A/u) ∨ F(S/u) = F(B/u)).

Now let p = #(Ocv(G) \ Ocv(S)). We prove the theorem by induction on p. If p = 0, Ocv(G) ⊂ Ocv(S) from which Ocv(G) = Ocv(S), so we can apply lemma 7 and obtain (2).

Now we prove (2) for some value of p > 0, assuming that the theorem is proved for all values smaller than p.

Let u be such a position. Without loss of generality, we can assume B/u1 = y and A/u1 is a non-variable term. Let U_B(B/u1) = {u_{i,1}, . . . , u_{i,n_i}, . . . , u_{k,1}, . . . , u_{k,n_k}} such that for each i, {u_{i,1}, . . . , u_{i,n_i}} = U_B(G/u_{i,1}) and G/u_{i,1} = z_i (see Figure 3). Because S is the lgs of A and B, their lgg, and because of (2) and (3), there exists a term t[x_1, . . . , x_l] with x_1, . . . , x_l fresh variables and variable to variable substitutions σ and σ_i such that ∀u ∈ U_B(B/u1) : S/u = t[x_1, . . . , x_l]σ and ∀u ∈ U_B(G/u_{i,1}) : A/u = t[x_1, . . . , x_l]σ_i. Now define S_B as B{y ← t[x_1, . . . , x_l]}. Because the x_i are fresh variables, their only positions in S_B are inside the subterms at positions in U_B(B/u1) and differences between S_B and S_B are due to differences at the positions in U_B(B/u1).

Define A_G as lgg(A, S_B). For j ∈ {1, . . . , n_i}, it is possible to write A_G/u_{i,j} as t[x_1, . . . , x_l]θ_i where the θ_i are also variable to variable substitutions. Moreover, the only difference between G and A_G are at the positions in U_B(B/u1) and the variables x_i/θ_i cannot occur in A_G outside the subterms A_G/u_i (as the variables x_i do not occur outside the subterms at the positions in U_B(B/u1) of S_B). So also here, the differences between V-value of G and A_G are due to differences at the positions in U_B(B/u1).

The number of positions u where Var(A_G/u) ∧ not(Var(B/u)) hold is, compared to G and B, reduced by ∑_{i=1}^{n} n_i ≥ 1 so we can apply the induction hypotheses (4): V(A) + V(S_B) ≤ V(A_G) + V(S), so, to prove V(A) + V(B) ≤ V(G) + V(S), it suffices to prove: V(B) − V(S_B) ≤ V(G) − V(A_G).

(5)

If k, the number of different variables in the positions u ∈ U_B(B/u1) of G is 1 then G is a renaming of B and A_G is a renaming of S_B, thus V(B) − V(S_B) = V(G) − V(A_G) and (7) trivially holds. This is the base case for an induction step. Assuming (5) holds up to value k − 1, we prove it holds for k ≥ 2. We define A_{G_i} = A_{G_i} guru(t[x_1, . . . , x_l]θ_1, t[x_1, . . . , x_l]θ_2). For i ∈ {2, k}, j ∈ {1, n_i} we have A_{G_i}/u_{i,j} = t[x_1, . . . , x_l]θ_i and for i = 1, j ∈ {1, n_1} we have A_{G_i}/u_{i,j} = t[x_1, . . . , x_l]θ_2 for θ_2, . . . , θ_k variable to variable substitutions. We define G_1 = lgg(A_{G_1}, B). For the positions in U_B(B/u1), G_1 has only k − 1 different variables (see Figure 3) and, from the induction hypotheses on G_1, A_{G_1}, B, and S_B, it follows that V(B) − V(S_B) ≤ V(G_1) − V(A_{G_1}).

(6)

Lemma 8 is applicable on the atoms G, A_G, G_i, and A_{G_i}, so V(G_1) − V(A_{G_1}) ≤ V(G) − V(A_G). Combining this with (6) one obtains (5). □
Figure 3: positions $u_{1,1}, \ldots, u_{k,n_k}$ where the atoms differ and the subterms at these positions
The example shows that the difference between the arguments of $p(x)$ and $q(x)$ in distances for sets of atoms which are parameterised with a distance between atoms. We have described such distances in [11] and [12].

Acknowledgements

We thank Luc De Raedt for the many interesting discussions. We thank also Shan Hwei Nienhuys-Cheng for reading the draft and the useful comments.
Wim Van Laer and Maurice Bruynooghe are 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


