Induction of Optimal Semantic Semi-distances for Clausal Knowledge Bases

Claudia d'Amato, Nicola Fanizzi, Floriana Esposito

LACAM – Dipartimento di Informatica – Università degli studi di Bari
Campus Universitario, Via Orabona 4 – 70125 Bari, Italy
{claudia.damato|fanizzi|esposito}@di.uniba.it

Abstract. Several activities related to semantically annotated resources can be enabled by a notion of similarity, spanning from clustering to retrieval, matchmaking and other forms of inductive reasoning. We propose the definition of a family of semi-distances over the set of objects in a knowledge base which can be used in these activities. In the line of works on distance-induction on clausal spaces, the family is parameterized on a committee of concepts expressed with clauses. Hence, we also present a method based on the idea of simulated annealing to be used to optimize the choice of the best concept committee.

1 Introduction

Assessing semantic similarity between objects can support a wide variety of instance-based tasks spanning from *case-based reasoning* and *retrieval* to *inductive generalization* and *clustering*.

As pointed out in related surveys [16], initially, most of the proposed similarity measures for concept descriptions focus on the similarity of atomic concepts within simple concept hierarchies or are strongly based on the structure of the terms for specific FOL fragments [5]. Alternative approaches are based on related notions of *feature* similarity or *information content*. All these approaches have been specifically aimed at assessing similarity between concepts (see also [10]). In the perspective of exploiting similarity measures in inductive (instance-based) tasks like those mentioned above, the need for a definition of a semantic similarity measure for *instances* arises [1, 2, 12].

Recently, semantic dissimilarity measures for specific FOL fragments have been proposed which turned out to be practically effective for the targeted inductive tasks. Although these measures ultimately rely on the semantics of primitive concepts as elicited from the knowledge base, still they are partly based on structural criteria (a notion of normal form) which determine also their main weakness: they are hardly portable to deal with other FOL fragments.

Therefore, we have devised a new family of dissimilarity measures for semantically annotated resources, which can overcome the aforementioned limitations. Our measures are mainly based on Minkowski's measures for Euclidean spaces defined by means of the *hypothesis-driven* distance induction method [14]. Another source of inspiration was provided by the *indiscernibility* relationships investigated *rough sets* theory [11].

Namely, the proposed measures are based on the degree of discernibility of the input objects with respect to a committee of features, which are represented by concept descriptions. As such, these new measures are not absolute, since they depend on both the choice (and cardinality) of the features committee and the knowledge base they are applied to. Rather, they rely on statistics on objects that are likely to be maintained by the knowledge base management system, which can determine a potential speed-up in the measure computation during knowledge-intensive tasks. Differently from the original idea [14], we give a definition of the notion of projections which is based on model-theory in LP.

Furthermore, we also propose ways to extend the presented measures to the case of assessing concept similarity by considering concepts as represented by their extension, i.e. the set of their instances. Specifically, we recur to notions borrowed from clustering [6] such as the *medoid*, the most centrally located instance in a concept extension w.r.t. a given metric.

Experimentally[1], it may be shown that the measures induced by large committees (e.g. including all primitive and defined concepts) can be sufficiently accurate when employed for classification tasks even though the employed committee of features were not the optimal one or if the concepts therein were partially redundant. Nevertheless, this has led us to investigate on a method to optimize the committee of features that serve as dimensions for the computation of the measure. To this purpose, the employment of genetic programming and randomized search procedures was considered. Finally we opted for an optimization search procedure based on *simulated annealing* [7], a randomized approach that can overcome the problem of the search being caught in local minima.

The remainder of the paper is organized as follows. The definition of the family of measures is proposed in Sect. 2, where we prove them to be semi-distances and extend their applicability to the case of concept similarity. In Sect. 3, we illustrate and discuss the method for optimizing the choice of concepts for the committee of features which induces the measures. The effectiveness of the method is demonstrated in a preliminary experimentation (see Sect. 4) on the task of similarity search. Possible developments are finally examined in Sect. 5.

2 A Family of Semi-distances for Instances

In the following, we assume that objects (instances), concepts and relationships among them may be defined in terms of a function-free (yet not constant-free) clausal language such as DATALOG, endowed with the standard semantics (see [9] for reference).

We will consider a *knowledge base* $\mathcal{K} = \langle \mathcal{P}, \mathcal{D} \rangle$, where $\mathcal{P}$ is a logic program representing the *schema*, with concepts (entities) and relationships defined through definite clauses, and the *database* $\mathcal{D}$ is a set of ground facts concerning the world state. In this context, without loss of generality, we will consider concepts as described by unary atoms. *Primitive* concepts are defined in $\mathcal{D}$ extensionally by means of ground facts only, whereas *defined* concepts will be defined in $\mathcal{P}$ by means of clauses. The set of the objects occurring in $\mathcal{K}$ is denoted with $\mathsf{const}(\mathcal{D})$.

[1] Such experiments, regarding a nearest neighbor search task, are not further commented here for the sake of brevity.

As regards the necessary inference services, our measures will require performing *instance-checking*, which amounts to determining whether an object belongs (is an instance) of a concept in a certain interpretation.

2.1 Basic Measure Definition

It can be observed that instances lack a syntactic structure that may be exploited for a comparison. However, on a semantic level, similar objects should *behave* similarly with respect to the same concepts, i.e. similar assertions (facts) should be shared. Conversely, dissimilar instances should likely instantiate disjoint concepts.

Therefore, we introduce novel dissimilarity measures for objects, whose rationale is the comparison of their semantics w.r.t. a fixed number of dimensions represented by concept descriptions (predicate definitions). Namely, instances are compared on the grounds of their behavior w.r.t. a reduced (yet not necessarily disjoint) committee of features, represented by a collection of concept descriptions, say $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$, which stands as a group of discriminating *features* expressed in the language taken into account. In this case, we will consider unary predicates which have a definition in the knowledge base.

Following [14], a family of totally semantic distance measures for objects can be defined for clausal representations. In its simplest formulation, inspired by Minkowski's metrics, these functions can be defined as follows:

Definition 2.1 (family of measures). *Let $\mathcal{K}$ be a knowledge base. Given a set of concept descriptions $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$, a family $\{d_p^{\mathsf{F}}\}_{p \in \mathbb{N}}$ of functions $d_p^{\mathsf{F}} : \mathsf{const}(\mathcal{D}) \times \mathsf{const}(\mathcal{D}) \mapsto [0, 1]$ is defined as follows:*

$$\forall a, b \in \mathsf{const}(\mathcal{D}) \quad d_p^{\mathsf{F}}(a, b) := \frac{1}{m} \left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) \mid^p \right]^{1/p}$$

where $\forall i \in \{1, \ldots, m\}$ the i-th projection function *π_i is defined by:*

$$\forall a \in \mathsf{const}(\mathcal{D}) \quad \pi_i(a) = \begin{cases} 1 & \mathcal{K} \vdash F_i(a) \\ 0 & \textit{otherwise} \end{cases}$$

The superscript F will be omitted when the set of features is fixed.

2.2 Discussion

We can prove that these functions have the standard properties for semi-distances:

Proposition 2.1 (semi-distance). *For a fixed feature set and $p \in \mathbb{N}$, function d_p is a semi-distance.*

Proof. In order to prove the thesis, given any three objects $a, b, c \in \mathsf{const}(\mathcal{D})$ it must hold that:

1. $d_p(a, b) \geq 0$ *positivity*
2. $d_p(a, b) = d_p(b, a)$ *symmetry*
3. $d_p(a, c) \leq d_p(a, b) + d_p(b, c)$ *triangular inequality*

Now, we observe that:

1. *trivial, by definition*
2. *trivial, for the commutativity of the operators involved*
3. *it follows from the properties of the power function:*

$$\begin{aligned}
d_p(a,c) &= \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(c) \mid^p\right]^{1/p} \\
&= \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) + \pi_i(b) - \pi_i(c) \mid^p\right]^{1/p} \\
&\leq \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) \mid^p + \mid \pi_i(b) - \pi_i(c) \mid^p\right]^{1/p} \\
&= \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) \mid^p + \sum_{i=1}^{m} \mid \pi_i(b) - \pi_i(c) \mid^p\right]^{1/p} \\
&\leq \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) \mid^p\right]^{1/p} + \frac{1}{m}\left[\sum_{i=1}^{m} \mid \pi_i(b) - \pi_i(c) \mid^p\right]^{1/p} \\
&= d_p(a,b) + d_p(b,c)
\end{aligned}$$

As such, these are only a semi-distances. Namely, it cannot be proved that $d_p(a,b) = 0$ iff $a = b$. This is the case of *indiscernible* instances with respect to the given set of features F [11].

Here, we make the assumption that the feature-set F may represent a sufficient number of (possibly redundant) features that are able to discriminate really different objects. As hinted in [14], redundancy may help appreciate the relative differences in similarity.

Compared to other proposed distance (or dissimilarity) measures, the presented functions are not based on structural (syntactical) criteria; namely, they require only deciding whether an object can be an instance of the concepts in the committee.

Note that the computation of projection functions can be performed in advance (with the support of suitable DBMSs) thus determining a speed-up in the actual computation of the distance measure. This is very important for the integration of these measures in instance-based methods which massively use distances, such as in case-based reasoning and clustering.

2.3 Extensions

The definition above might be further refined and extended by recurring to model theory. Namely, the set of Herbrand models of the knowledge base $\mathcal{M}_\mathcal{K} \subseteq 2^{|\mathcal{B}_\mathcal{K}|}$ may be considered, where $\mathcal{B}_\mathcal{K}$ stands for its Herbrand base.

Now, given two instances a and b to be compared w.r.t. a certain feature F_i, $i = 1, \ldots, m$, we might check whether they can be distinguished in the world represented by a Herbrand interpretation $\mathcal{I} \in \mathcal{M}_\mathcal{K}$: $\mathcal{I} \models F_i(a)$ and $\mathcal{I} \models F_i(b)$. Hence, a distance measure should count the cases of disagreement, varying the Herbrand models of the knowledge base: The resulting definition for a dissimilarity measure is the following:

$$\forall a, b \in \mathsf{const}(\mathcal{D}) \quad d_p^{\mathsf{F}}(a,b) := \frac{1}{m \cdot |\mathcal{M}_{\mathcal{K}}|} \left[\sum_{\mathcal{I} \in \mathcal{M}_{\mathcal{K}}} \sum_{i=1}^{m} \mid \pi_i^{\mathcal{I}}(a) - \pi_i^{\mathcal{I}}(b) \mid^p \right]^{1/p}$$

where the projections are computed for a specific world state as encoded by a Herbrand interpretation $\mathcal{I}$:

$$\forall a \in \mathsf{const}(\mathcal{D}) \quad \pi_i^I(a) = \begin{cases} 1 & F_i(a) \in \mathcal{I} \\ 0 & \textit{otherwise} \end{cases}$$

Following the rationale of the average link criterion used in clustering [6], the measures can be extended to the case of concepts, by recurring to the notion of medoids. The *medoid* of a group of objects is the object that has the highest similarity w.r.t. the others. Formally. given a group $G = \{a_1, a_2, \ldots, a_n\}$, the medoid is defined:

$$m = \mathrm{medoid}(G) = \operatorname*{argmin}_{a \in G} \sum_{j=1}^{n} d_p^{\mathsf{F}}(a, a_j)$$

Now, given two concepts C_1, C_2, we can consider the two corresponding groups of objects obtained by retrieval $R_i = \{a \in \mathsf{const}(\mathcal{D}) \mid \mathcal{K} \models C_i(a)\}$, and their resp. medoids $m_i = \mathrm{medoid}(R_i)$ for $i = 1, 2$ w.r.t. a given measure d_p^{F} (for some $p > 0$ and committee F). Then we can define the function for concepts as follows:

$$d_p^{\mathsf{F}}(C_1, C_2) := d_p^{\mathsf{F}}(m_1, m_2)$$

Alternatively, a metric can be defined based on the single-link and complete-link principles [6]:

$$d_p^{\mathsf{F}}(C_1, C_2) = \frac{\min\{d_p^{\mathsf{F}}(a,b) \mid \mathcal{K} \models C_1(a) \wedge C_2(b)\}}{\max\{d_p^{\mathsf{F}}(a,b) \mid \mathcal{K} \models C_1(a) \wedge C_2(b)\}}$$

3 Optimization

Although the measures could be implemented according to the definitions, their effectiveness and also the efficiency of their computation strongly depends on the choice of the feature committee (*feature selection*). Indeed, various optimizations of the measures can be foreseen as concerns their parametric definition.

Among the possible committees, those that are able to better discriminate the objects in the ABox ought to be preferred:

Definition 3.1 (good feature set). *Let* $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ *be a set of concept descriptions. We call* F *a* good feature set *for the knowledge base* $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ *iff* $\forall a, b \in \mathsf{const}(\mathcal{D})\ \exists i \in \{1, \ldots, m\} :\ \pi_i(a) \neq \pi_i(b)$.

Note that, when the function defined in the previous section adopts a good feature set, it has the properties of a metric on the related instance-space.

Since the function strongly depends on the choice of concepts included in the committee of features F, two immediate heuristics can be derived:

1. controlling the number of concepts of the committee (which has an impact also on efficiency), including especially those that are endowed with a real discriminating power;
2. finding optimal sets of discriminating features of a given cardinality, by allowing also their composition employing the specific refinement operators.

Both these heuristics can be enforced by means of suitable ILP techniques especially when knowledge bases with large sets of instances are available. Namely, part of the entire data can be drawn in order to induce optimal F sets, in advance with respect to the application of the measure for other specific purposes as those mentioned above. The adoption of genetic programming has been considered for constructing optimal sets of features. Yet these algorithms are known to suffer from being possibly caught in local minima. An alternative may consist in employing a different probabilistic search procedure which aims at a global optimization. Thus a method based on simulated annealing [7] has been devised, whose algorithm is reported in Fig. 1.

Essentially the algorithm searches the space of all possible feature committees starting from an initial guess (determined by MAKEINITIALFS($\mathcal{K}$)) based on the concepts (both primitive and defined) currently referenced in the knowledge base. The loop controlling the search is repeated for a number of times that depends on the temperature which gradually decays to 0, when the current committee can be returned. The current feature set is iteratively refined calling a suitable procedure RANDOMSUCCESSOR(). Then the fitness of the new feature set is compared to that of the current one determining the increment of energy ΔE. If this is positive then the candidate committee replaces the current one. Otherwise it will be replaced with a probability that depends on ΔE.

As regards the heuristic FITNESSVALUE(F), it can be computed as the average *discernibility factor* [11] of the objects w.r.t. the feature set. For example, given a set of objects $IS = \{a_1, \ldots, a_n\} \subseteq \mathsf{const}(\mathcal{D})$ the fitness function may be defined:

$$\textsc{fitnessValue}(\mathsf{F}) = k \cdot \sum_{1 \leq i < j \leq n} \sum_{h=1}^{m} \mid \pi_h(a_i) - \pi_h(a_j) \mid$$

where k is a normalization factor which may be set to: $(1/m)\,(n \cdot (n-1)/4 - n)$, depending on the number of couples of different instances that really determine the fitness measure.

As concerns finding candidates to replace the current committee (RANDOMSUCCESSOR()), the function was implemented by recurring to simple transformations of a feature set:

- adding (resp. removing) a concept C: $\mathsf{nextFS} \leftarrow \mathsf{currentFS} \cup \{C\}$ (resp. $\mathsf{nextFS} \leftarrow \mathsf{currentFS} \setminus \{C\}$)
- randomly choosing one of the current concepts from $\mathsf{currentFS}$, say C, and replacing it with one of its refinements $C' \in \textsc{ref}(C)$

Refining concept descriptions is language-dependent. For the adopted clausal logic, various refinement operators have been proposed in the literature [9]. *Complete* operators are to be preferred to ensure exploring the whole search-space.

FeatureSet OPTIMIZEFEATURESET($\mathcal{K}$, ΔT)
input $\mathcal{K}$: Knowledge base
ΔT: function controlling the decrease of temperature
output FeatureSet
local currentFS: current Feature Set
nextFS: next Feature Set
Temperature: controlling the probability of downward steps
begin
currentFS $\leftarrow$ MAKEINITIALFS($\mathcal{K}$)
for $t \leftarrow 1$ **to** ∞ **do**
Temperature $\leftarrow$ Temperature $- \Delta T(t)$
if (Temperature = 0)
return currentFS
nextFS $\leftarrow$ RANDOMSUCCESSOR(currentFS,$\mathcal{K}$)
$\Delta E \leftarrow$ FITNESSVALUE(nextFS) $-$ FITNESSVALUE(currentFS)
if ($\Delta E > 0$)
currentFS $\leftarrow$ nextFS
else *// replace FS with given probability*
REPLACE(currentFS, nextFS, $e^{\Delta E/\mathsf{Temperature}}$)
end

Fig. 1. Feature Set optimization based on a Simulated Annealing procedure.

4 Experiments on Similarity Search

In order to prove the effectiveness of the distance coupled with the optimization procedure, an experimentation was performed on the task of *similarity search* [16], i.e. searching instances that can be answers to relational queries by means of a notion of distance. We intended to evaluate both the effectiveness of the distance and the impact of its optimization phase.

To this purpose, a relational kNN algorithm was devised, similar to RIBL [2], with a voting procedure weighted by the distance of the query instance from its neighbors. The Java implementation exploits external Prolog libraries[2] for the reasoning services required for determining the distance between individuals.

Four relational datasets from very different domains were selected: a small one was artificially generated for the PHASE TRANSITION [3], (problem `pt4444`), the University of Washington CSE dept. dataset (UW-CSE) [13], one from the *Mutagenesis* datasets [15], and one concerned the layout structure of scientific papers (SCI-DOCS) [4]. The details[3] about these datasets are reported in Tab. 1. A simple discretization had to be preliminary operated on the numerical attributes, if present. for the measure currently does not handle these cases. Hence, the number of concepts was increased w.r.t. the original dataset.

[2] JPL 3. See `http://www.swi-prolog.org`
[3] As stated in Sect.2, *concepts* correspond to unary predicates while predicates with larger arity are generically referred to as *relations*. *Individuals* correspond to the objects denoted by the constant names, i.e. the resources to be searched.

Table 1. Details about the datasets that were employed in the experiments.

dataset	#concepts	#relations	#individuals
PHASE TRANSITION	1	4	400
UW-CSE	9	20	2208
SCI-DOCS	30	9	4585
MUTAGENESIS	68	2	9292

Table 2. Experimental results: cardinality of the induced feature set and average outcomes ($\pm$ standard deviation).

dataset	$\vert F \vert$	%correct	%false pos.	%false neg.
PHASE TRANSITION	6	99.97 ± 0.13	0.00 ± 0.00	0.03 ± 0.13
UW-CSE	9	99.01 ± 1.92	0.05 ± 0.08	0.94 ± 1.94
SCI-DOCS	5	85.49 ± 9.06	1.66 ± 1.87	12.85 ± 8.96
MUTAGENESIS	11	98.68 ± 1.92	0.08 ± 0.12	1.24 ± 1.94

We intended to assess the accuracy of the answers obtained inductively from the kNN procedure compared to the correct (deductive) ones. Preliminarily, an optimal distance was obtained using the procedure described in the previous section, to be employed both for selecting the nearest neighbors and for determining their weights. A 5% sample of instances was drawn from the dataset for performing the distance optimization task finding a proper feature set.

In the successive phase, a number of 20 queries (clauses whose head defines a new concept) were randomly generated provided that they had non-empty answer sets for the head variable. Then, search was simulated by testing class-membership w.r.t. the query concepts employing the kNN procedure based on the distance. The experiment was repeated applying a 10-fold cross-validation setting.

In all of the runs the number of nearest neighbors k that determine the classification of the test instance was set to $\sqrt{|\mathsf{TrSet}|}$, where TrSet is the training set related to the given fold. The cardinality of the committees determined in the first phase and the average results of the classification are reported in Table 2.

We note that the performance is quite good with a decay for the case of the SCI-DOCS dataset, which is determined by the larger variance: some queries were perfectly answered while some yielded poorer results. In terms of retrieval measures, we can say that the procedure suffers more in terms of recall rather than precision. the good results were probably due to the regularity of the information in the various datasets: for each individual the same amount of information is known, which helps to discern among them. More sparsity (incomplete information) would certainly decrease the distance acuity and, hence, the overall performance to the task.

The good performance on such datasets, despite some of them are known to be particularly difficult for learning methods, is due to the fact that for the considered task a characterization of an unknown concept is not to be learnt. Rather, it is an input of the inductive procedure based on discriminative features that help to discern between member and non-member instances.

Table 3. Experimental results (no optimization phase): average outcomes of the experiment.

dataset	%correct	%false pos.	%false neg.
PHASE TRANSITION	N/A	N/A	N/A
UW-CSE	94.88	0.7	4.42
SCI-DOCS	81,29	2,65	16,06
MUTAGENESIS	94.76	0.55	4,69

It is also possible to compare the number of new features induced for the distance measure and the overall number of (primitive or defined) concepts in the KB. In the case of the MUTAGENESIS dataset, it needed only about 15% of the available concepts. However it is to be admitted that many of them had been added to the original dataset during the discretization process.

Employing smaller committees (with comparable performance results) is certainly desirable for the sake of an efficient computation of the measure. In order to asses the potential of the measure when employing basically the concepts already contained in the knowledge base, the same experiment with the same settings (10-fold cross validation) was repeated with no preliminary optimization phase; instead, for the comparison, we randomly selected the same amount of pre-defined concepts in the knowledge base as the size of the optimal set (indicated in the second column of Tab. 2). The average results obtained are reported in Tab. 3. The PHASE TRANSITION dataset had only one predefined concept all predicates are relations with arity ≥ 2) which excluded it from the possibility of a comparison.

The performance in terms of time was quite satisfactory, however various optimizations can be implemented for this specific search-task such as, for instance, computing and storing the distances in appropriate data structures [16] (e.g. *kD-trees* or *ball-trees*) that may speed-up the overall retrieval process.

5 Conclusions and Ongoing Work

In the line of past works on distance-induction, we have proposed the definition of a family of semi-distances over the instances in a clausal knowledge base. The measures are parameterized on a committee of concepts that can be selected by the proposed randomized search method.

Possible subsumption relationships between clauses in the committee may be explicitly exploited in the measure for making the relative distances more accurate. The extension to the case of concept distance may also be improved. Particularly, the measure should be extended to cope with numeric information which abounds in biological/chemical datasets.

The measures may have a wide range of application in distance-based methods to knowledge bases. Currently we are exploiting the measures in conceptual clustering algorithms where clusters will be formed by grouping instances on the grounds of their similarity assessed through the measure, triggering the induction of new emerging concepts.

Another possibility is also the extension to learning relational kernels which encode a notion of similarity, as in kFOIL [8], where measure induction and performance evaluation are intertwined.

References

[1] G. Bisson. Learning in fol with a similarity measure. In W. Swartout, editor, *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 82–87. MIT Press, 1992.

[2] W. Emde and D. Wettschereck. Relational instance-based learning. In L. Saitta, editor, *Proceedings of the 13th International Conference on Machine Learning, ICML96*, pages 122–130. Morgan Kaufmann, 1996.

[3] F. Esposito, S. Ferilli, N. Fanizzi, T. Basile, and N. Di Mauro. An exhaustive matching procedure for the improvement of learning efficiency. In *Proceedings of the 13th International Conference on Inductive Logic Programming, ILP2003*, volume 2835 of *LNCS*, pages 112–129. Springer, 2003.

[4] F. Esposito, S. Ferilli, N. Fanizzi, T. Basile, and N. Di Mauro. Incremental learning and concept drift in INTHELEX. *Journal of Intelligent Data Analysis*, 8(1/2):133–156, 2004.

[5] A. Hutchinson. Metrics on terms and clauses. In M. van Someren and G. Widmer, editors, *Proceedings of the 9th European Conference on Machine Learning, ECML97*, volume 1224 of *LNAI*, pages 138–145. Springer, 1997.

[6] A. Jain, M. Murty, and P. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[8] N. Landwehr, A. Passerini, L. De Raedt, and P. Frasconi. kFOIL: Learning simple relational kernels. In *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI06*. AAAI Press, 2006.

[9] S. Nienhuys-Cheng and R. de Wolf. *Foundations of Inductive Logic Programming*, volume 1228 of *LNAI*. Springer, 1997.

[10] S.-H. Nienhuys-Cheng. Distances and limits on herbrand interpretations. In D. Page, editor, *Proceedings of the 8th International Workshop on Inductive Logic Programming, ILP98*, volume 1446 of *LNAI*, pages 250–260. Springer, 1998.

[11] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, 1991.

[12] 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.

[13] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62:107–136, 2006.

[14] M. Sebag. Distance induction in first order logic. In S. Džeroski and N. Lavrač, editors, *Proceedings of the 7th International Workshop on Inductive Logic Programming, ILP97*, volume 1297 of *LNAI*, pages 264–272. Springer, 1997.

[15] A. Srinivasan, S. Muggleton, R. King, and M. Sternberg. Mutagenesis: ILP experiments in a non-determinate biological domain. In S. Wrobel, editor, *Proceedings of the 4th International Workshop on Inductive Logic Programming, ILP94*, number 237 in GMD-Studien, pages 217–232, 1994.

[16] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach*. Springer, 2007.